



Taissa Abdalla Filgueiras de Sousa

**Sistema de recomendação para apoiar a construção de
gráficos com dados estatísticos**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do título de Mestre pelo Programa de Pós-
Graduação em Informática da PUC-Rio.

Orientadora Prof.^a Simone Diniz Junqueira Barbosa

Rio de Janeiro
Março de 2013



Taissa Abdalla Filgueiras de Sousa

**Sistema de recomendação para apoiar a construção de
gráficos com dados estatísticos**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof.^a Simone Diniz Junqueira Barbosa
Orientadora
Departamento de Informática – PUC-Rio

Prof.^a Clarisse Sieckenius de Souza
Departamento de Informática – PUC-Rio

Prof. Bruno Feijó
Departamento de Informática – PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de Janeiro, 22 de março de 2013

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Taissa Abdalla Figueiras de Sousa

Bacharel em Desenho Industrial pela Pontifícia Universidade Católica do Rio de Janeiro em julho de 1999 e Pós-graduação em Análise, Projeto e Gerência de Sistemas também pela PUC-Rio em 2003.

Ficha Catalográfica

Sousa, Taissa Abdalla Figueiras de

Sistema de recomendação para apoiar a construção de gráficos com dados estatísticos / Taissa Abdalla Figueiras de Sousa ; orientadora: Simone D. J. Barbosa. – 2013.

180 f. : il. (color.) ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2013.

Inclui bibliografia

1. Informática – Teses. 2. Visualização de informação. 3. Sistemas de recomendação. 4. Engenharia semiótica. 5. Interação humano-computador. I. Barbosa, Simone D. J. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Aos meus pais

Agradecimentos

Agradeço primeiramente a Deus por todas as bênçãos que tive ao longo da vida.

Aos meus pais Bosco e Marilú e aos meus irmãos Rodrigo e Thiago por todo amor, carinho e apoio durante o mestrado.

A todos os meus amigos e familiares que compreenderam as minhas ausências em virtude dos estudos.

À minha orientadora Simone pela paciência, confiança, dedicação e entusiasmo com a minha pesquisa, pela sua competência, pelos valiosos conhecimentos transmitidos durante todo o processo de orientação e ensinamentos que levarei ao longo da vida.

Aos professores Clarisse, Daniel, Poggi e Casanova pelas brilhantes aulas e profícuas discussões que foram importantes para a minha formação.

Aos amigos da PUC e do Ideias que estiveram sempre junto comigo, apoiando-me nos momentos mais difíceis do mestrado ou comemorando alguma etapa cumprida. Em especial: Jefferson Rômulo, Vitor Pinheiro, Bernardo Nunes, Aline Saettler, Fabiana Simões, Alexandre Silva, Heraldo Borges, Gabriel Sotero, José Antônio Motta, Malu Seixas, Bruno Santana, Rodrigo Maués, Henrique Sousa, Sofia Manso e Adriano Branco.

Aos amigos do IBGE, principalmente Edna Campello, David Wu Tai, Sérgio Cortes, Paulo Quintslr, Luiz Agner e Danielle Macedo que me incentivaram para a realização desse curso.

A todos os outros amigos que estiveram presente ao longo dessa pesquisa e tornaram esse período da minha vida ainda mais especial e divertido. Em especial: Marcela Percú, Sabrina Aleixo, Regina Reznik, Eduardo Milson, Gabriella Vaccari, Renata Vilanova, Carolina Trapani, Viviana Rangel, Ilana

Novikov,,Daniela Moura, Izabel Leal, Paula Veloso, Magalí Gacek, Pedro Ávila, Luiz Flávio Oswald, Marcelo Prista, Adriene Freitas, Carolina Campos, Adriana Tolentino e Kaká Boa Morte.

Aos participantes e colaboradores dessa pesquisa pela gentil e voluntária participação.

Mais uma vez agradeço à Adriana Tolentino pela voz emprestada para o vídeo de apresentação da ferramenta apresentada nesta dissertação.

Finalmente, ao CNPq, à VRAc/PUC-Rio e ao IBGE pela concessão do afastamento e pelo suporte financeiro oferecido à execução dessa pesquisa.

Resumo

Sousa, Taissa Abdalla Filgueiras; Barbosa, Simone Diniz Junqueira. **Sistema de recomendação para apoiar a construção de gráficos com dados estatísticos**. Rio de Janeiro, 2013. 180p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Pesquisas sobre visualização de dados estatísticos demonstram a necessidade de sistemas que apoiem tomadas de decisões e análises visuais. Constatando problemas de construção de visualizações por usuários inexperientes, nossa questão de pesquisa foi: *Como apoiar usuários inexperientes na construção de visualizações eficientes com dados estatísticos?* Assim, desenvolvemos ViSC, um sistema de recomendações que apóia a construção interativa de gráficos para visualizar dados estatísticos, através de uma série de recomendações baseadas nos dados selecionados e na interação do usuário com a ferramenta. O sistema explora uma ontologia de visualização para oferecer um conjunto de gráficos que ajudam a responder questões baseadas em informação relacionadas aos dados exibidos no gráfico. Percorrendo os gráficos recomendados através de suas questões relacionadas, o usuário implicitamente adquire conhecimento tanto do domínio quanto dos recursos de visualização que melhor representam os conceitos do domínio de interesse. Esta dissertação apresenta os problemas que motivaram a pesquisa, descreve a ferramenta ViSC e apresenta os resultados de uma pesquisa qualitativa realizada para avaliar ViSC. Para a avaliação, utilizamos o Método de Inspeção Semiótica (MIS) e o "*Retrospective Communicability Evaluation*" (RCE) — uma combinação do Método de Avaliação da Comunicabilidade (MAC) e *Retrospective Think Aloud Protocol*. Concentramo-nos em verificar como as recomendações influenciam na realização de uma tarefa e nas visualizações geradas para então endereçar nossa questão mais ampla.

Palavras-chave

Visualização de informação; sistemas de recomendação; engenharia semiótica; interação humano-computador.

Abstract

Sousa, Taissa Abdalla Filgueiras; Barbosa, Simone Diniz Junqueira (Advisor). **Recommender System to support chart constructions with statistical data**. Rio de Janeiro, 2013. 180p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Research on statistical data visualization emphasizes the need for systems that assist in decision-making and visual analysis. Having found problems in chart construction by novice users, we decided to research the following question: *How can we support novice users to create efficient visualizations with statistical data?* Thus we, created ViSC, a recommender system that supports the interactive construction of charts to visualize statistical data by offering a series of recommendations based on the selected data and the user interaction with the tool. The system explores a visualization ontology to offer a set of graphs that help to answer information-based questions related to the current graph data. By traversing the recommended graphs through their related questions, the user implicitly acquires knowledge both on the domain and on visualization resources that better represent the domain concepts of interest. This dissertation presents the problems that motivated the research, describes the ViSC tool and presents the results of a qualitative study conducted to evaluate ViSC. We used two methods in our evaluation: the Semiotic Inspection Method (SIM) and the Retrospective Communicability Evaluation (RCE) — a combination of the Communicability Evaluation Method (CEM) and Retrospective Think Aloud Protocol. We first analyze how the questions influence the users' traversal through the graph and, then, we address the broader question.

Keywords

Information visualization; recommender systems; semiotic engineering; human-computer-interaction.

Sumário

1	Introdução	17
1.1.	Problema, questão, objetivos e escopo da pesquisa	18
1.2.	Métodos e resultados	21
1.3.	Organização do trabalho	24
2	Trabalhos relacionados	26
2.1.	As propriedades e as regras do sistema gráfico	26
2.2.	Sistemas de significação e comunicação e ferramentas de visualização	30
2.3.	Conceito de abdução	31
2.4.	Técnicas de visualização em sistemas interativos	33
2.5.	Sistemas de recomendação	35
3	Estudos preliminares	40
3.1.	Como as ferramentas de visualização influenciam nas visualizações geradas	40
3.1.1.	Discussão	46
3.2.	Caracterização semântica de mecanismos de visualização	48
3.2.1.	Discussão	55
4	A ferramenta ViSC	56
4.1.	Requisitos	56
4.2.	Metamensagem	59
4.3.	A gramática de ViSC	60
4.4.	O sistema de recomendação de ViSC	61
4.5.	A interface de ViSC	65
4.6.	A interação com ViSC	67
4.7.	Comparação entre ViSC e outras ferramentas relacionadas	69

5 Avaliação	72
5.1. Métodos	72
5.2. Preparação	76
5.3. Método de Inspeção Semiótica (MIS)	80
5.3.1. Inspeção da ferramenta VISC	80
5.3.2. Inspeção da ferramenta Tableau Public	88
5.4. Retrospective Communicability Evaluation	99
5.4.1. Tarefa 1	100
5.4.2. Tarefa 2	127
5.4.3. Discussão sobre as ferramentas	153
5.4.4. Elaboração do perfil semiótico	156
5.5. Discussão final	157
6 Conclusões	160
6.1.1. Contribuições	162
6.1.2. Trabalhos futuros	163
7 Referências bibliográficas	165
APÊNDICE A	169
Outros resultados obtidos no primeiro estudo preliminar (seção 3.1)	169
APÊNDICE B	170
Formulário pós-entrevista	170
Roteiro de entrevista	172
APÊNDICE C	174
Conjunto de modelos que representam as condições necessárias para cada pergunta ser ou não apresentada.	175

Lista de Figuras

Figura 1: O problema da escolha da representação mais eficiente	28
Figura 2: Exemplos de problemas de separação angular	29
Figura 3: Assistente de gráfico do Excel 2003. Exemplo de chart Typology	33
Figura 4: Workspace da ferramenta Impure	34
Figura 5: Código e visualização gerada a partir do código pelo <i>toolkit</i> Protovis	35
Figura 6: Resultado obtido pelo usuário A01 do grupo 1 para a tarefa 1	42
Figura 7: Resultado obtido por A02, A03 (grupo 1) e D03 (grupo 2) para a tarefa 1.	42
Figura 8: Gráfico de dispersão com diagonal	42
Figura 9: Gráfico de dispersão com reta de regressão linear. Statplanet	42
Figura 10: Gráfico de colunas agrupadas	43
Figura 11: Resultado da dupla D01 (grupo 2) para a tarefa 1: 3 gráficos de barras com seleção do Brasil	44
Figura 12: Resultado da dupla D02 (grupo 2): 3 gráficos de barras com os 32 países selecionados: Solução mais eficientes entre as apresentadas.	44
Figura 13: Resultado para a tarefa 2	45
Figura 14: Classe dado	49
Figura 15: Classe atributos de exibição	50
Figura 16: Exemplo de uma sub-classe de visualização	51
Figura 17: Séries do tipo <i>index</i> ou agrupadas	52
Figura 18: Séries empilhadas	52
Figura 19: Séries múltiplas	53
Figura 20: Série horizontal	53
Figura 21: Exemplo de funções da classe transformação	54
Figura 22: Exemplo dos 8 tipos de visualização disponíveis em ViSC	61
Figura 23: Pré-condições que definem quais visualizações podem ser utilizadas de acordo com características do dado	63
Figura 24: Condições para a tarefa calcular valor derivado.	64
Figura 25: Interface de ViSC	65

Figura 26: Menu de países em Statplanet	66
Figura 27: Menu de seleção de elementos de uma dimensão em ViSC	66
Figura 28: Interface inicial	67
Figura 29: Interface com gráfico vazio	68
Figura 30: Detalhe de questões relacionadas ao gráfico aberto.	69
Figura 31: Detalhe de questões não relacionadas ao gráfico aberto mas sim ao dado selecionado	69
Figura 32: Interface do Tableau. (1) dados selecionados ou importados. (2) lista de dimensões disponíveis. (3) lista de medidas. (4) painéis com filtro, marcas e legenda. (5) menu e barra de ferramentas. (6) colunas e linhas onde dimensões e medidas são selecionadas (7) painel <i>Show me</i> .	70
Figura 33: Exploration Views consiste em: (a) o <i>Dashboard</i> principal e seus componentes; (b) o <i>Clipboard drawer widget</i> que inclui <i>templates</i> para gráficos (parte superior) e dados de exemplos de <i>widget</i> de dados (parte inferior) que o usuário tenha criado como uma paleta; (c) uma representação visual dos filtros de dados globais; e (d) um conjunto de outras funções, como buscar, salvar, compartilhar, etc.	70
Figura 34: Alerta exibido na inserção de valores a uma terceira dimensão	81
Figura 35: Passo 1 da interação de seleção em ViSC	82
Figura 36: Passo 2 da interação da seleção em ViSC	82
Figura 37: Passo 3 da interação de seleção em ViSC	83
Figura 38: Gráfico de colunas empilhadas (<i>stacked column</i>) no ViSC	85
Figura 39: Séries empilhadas (<i>stacked series</i>) no ViSC	85
Figura 40: Resultado para tarefa 2 sem utilização das perguntas	86
Figura 41: Solução 1 para a tarefa 2	86
Figura 42: Solução 2 para a tarefa 2	87
Figura 43: Solução 3	88
Figura 44: Detalhe da área do gráfico antes da inserção dos dados	89
Figura 45: Detalhe do final da janela Show me.	89
Figura 46: Janelas do Tableau	91

Figura 47: Janela <i>Show me</i> em 4 momentos da esquerda para a direita: sem seleção, na seleção de uma dimensão geográfica, na seleção de uma <i>string</i> e na seleção de uma medida	92
Figura 48: Linhas e colunas preenchidas com dimensões	92
Figura 49: Linhas e colunas preenchidas com uma medida	92
Figura 50: Menu de contexto e janela filtro para uma variável discreta ou uma dimensão	93
Figura 51: Menu de contexto e janela filtro para uma variável contínua ou uma medida	94
Figura 52: Resultado 1 obtido para a tarefa 1 com Tableau	97
Figura 53: Resultado 2 obtido para a tarefa 1 com Tableau	97
Figura 54: Resultado 3 obtido para a tarefa 1 com Tableau	97
Figura 55: Resultado 4 obtido para a tarefa 1 com Tableau	97
Figura 56: Solução para tarefa 2 com Tableau	98
Figura 57: Tentativa de inclusão da variação no Tableau	98
Figura 58: Resultado de U03 para a tarefa 1	101
Figura 59: Resultado de U05 para a tarefa 1	101
Figura 60: Resultado de U07 para a tarefa 1	101
Figura 61: Log de uso do ViSC por U03	102
Figura 62: Log de uso do ViSC por U05	103
Figura 63: Log de uso de ViSC por U07	103
Figura 64: Resultado de U02 para a tarefa 1	104
Figura 65: Resultado de U04 para a tarefa 1	104
Figura 66: Resultado de U08 para a tarefa 1	104
Figura 67: Gráfico de barras horizontais gerado por U02	115
Figura 68: Passo 1 da interação de U02	120
Figura 69: Passo 2 da interação de U02	121
Figura 70: Passo 3 da interação de U02	121
Figura 71: Resultado de U02 para a tarefa 2	127
Figura 72: Resultado de U04 para a tarefa 2	128
Figura 73: Resultado de U08 para a tarefa 2	128
Figura 74: Log de uso de U02 no ViSC	129
Figura 75: Log de uso de U04 no ViSC	129
Figura 76: Log de uso de U08 no ViSC	130

Figura 77: Resultado de U03	131
Figura 78: Resultado de U05	131
Figura 79: Resultado de U07	131
Figura 80: Resultado de A01 para tarefa 2 na pesquisa [17] com apoio do Excel.	169
Figura 81: Resultado de A02 para tarefa 2 na pesquisa [17]. Usuário não conseguiu unir as 3 colunas em um único gráfico	169
Figura 82: Pré-condições que definem que visualizações podem ser mapeadas de acordo com características do dado	175
Figura 83: Condições para a tarefa1: retornar valor.	176
Figura 84: Condições para a tarefa 2: filtrar.	176
Figura 85: Condições para a tarefa 3: calcular valor derivado.	177
Figura 86: Condições para a tarefa 4: achar extremo.	177
Figura 87: Condições para a tarefa 5. ordenar	178
Figura 88: Condições para a tarefa 6: determinar intervalo.	178
Figura 89: Condições para tarefa 7: caracterizar distribuição.	179
Figura 90: Condições para a tarefa 8: encontrar anomalia.	179
Figura 91: Condições para tarefa 9: agrupar.	180
Figura 92: Condições para tarefa 10: correlacionar.	180

Lista de Tabelas

Tabela 1: Técnicas de recomendação [36]	36
Tabela 2: Exemplos de perguntas e sua relação com a visualização e a tarefa	54
Tabela 3: Caracterização das etiquetas de comunicabilidade [12]	75
Tabela 4: Método do quadrado latino utilizado nas avaliações com usuários.....	76
Tabela 5: Perfis dos usuários	77
Tabela 6: Brasil - Valor do rendimento médio mensal por Pessoas de 10 anos ou mais de idade (Mil pessoas)	79
Tabela 7: Sequência de etapas da avaliação com os usuários e tempo estimado.....	79
Tabela 8: Respostas sobre dificuldade encontrada na tarefa 1 e na ferramenta que apoiou a realização da tarefa.....	105
Tabela 9: Respostas sobre a satisfação do resultado	106
Tabela 10: O que o usuário pretendia construir, o que foi desenhado no papel e o resultado obtido	106
Tabela 11: Respostas dos usuários do grupo 2 sobre o grau de interação com Show me.....	107
Tabela 12: Respostas dos usuários do grupo 2 sobre a influência da janela Show me.....	107
Tabela 13: Respostas dos usuários do grupo 1 sobre o grau de interação com as perguntas	107
Tabela 14: Respostas dos usuários do grupo 1 sobre a influência das perguntas para o resultado obtido.....	108
Tabela 15: Problemas nas Funcionalidades da ferramenta ViSC ou no sistema gráfico	108
Tabela 16: Problemas nas funcionalidades da ferramenta Tableau ou no sistema gráfico.....	112
Tabela 17: Etiquetagem das rupturas encontradas na tarefa 1 com Tableau.....	116
Tabela 18: Etiquetagem das rupturas encontradas na tarefa 1 com Tableau Public.....	119
Tabela 19: Respostas sobre dificuldade encontrada na tarefa 2 e na ferramenta que apoiou a realização da tarefa.	133

Tabela 20: Respostas sobre a satisfação do resultado	133
Tabela 21: O que o usuário pretendia construir, o que foi desenhado no papel e o resultado obtido	134
Tabela 22: Respostas dos usuários do grupo 1 sobre o grau de interação com Show me.....	135
Tabela 23: Respostas dos usuários do grupo 1 sobre a influência da janela Show me.....	135
Tabela 24: Respostas dos usuários do grupo 2 sobre o grau de interação com as perguntas	135
Tabela 25: Respostas dos usuários do grupo 2 sobre a influência das perguntas para o resultado obtido.....	135
Tabela 26: Problemas nas Funcionalidades da ferramenta ViSC ou no sistema gráfico.....	136
Tabela 27: Problemas nas Funcionalidades da ferramenta Tableau ou no sistema gráfico.....	139
Tabela 28: Etiquetagem das rupturas encontradas na tarefa 1 com Tableau.....	144
Tabela 29: Etiquetagem das rupturas encontradas na tarefa 2 com Tableau.....	146
Tabela 30: Relação entre questões, tarefas, e representações visuais	174

1 Introdução

Essa pesquisa visa contribuir para a área de interação-humano-computador (IHC) com discussões sobre interação em ferramentas de visualização de informação em gráficos para construção de gráficos com dados estatísticos, através do desenvolvimento e avaliação da ferramenta ViSC — *Visualization with Smart Charts* —, que agrega sugestões para apoiar a construção de gráficos para usuários inexperientes utilizando dados estatísticos. Esta dissertação apresenta os problemas que motivaram a pesquisa, descreve a ferramenta ViSC e apresenta os resultados de uma pesquisa qualitativa feita com a ferramenta criada.

Visualização é um processo que envolve psicologia, cognição e semiótica e pode ser classificada em dois grupos: científica e de informações. O primeiro envolve dados científicos com um componente espacial (3D) enquanto o segundo envolve dados abstratos e não espaciais [1], tais como dados estatísticos, foco dessa pesquisa.

O objetivo das visualizações é ajudar no entendimento do dado alavancando a habilidade do sistema visual humano de identificar padrões, detectar tendências e discrepâncias [2]. Carzola [3], com base em Pinker [4], afirma que os gráficos são um meio efetivo de comunicação porque tiram partido efetivo dos mecanismos cognitivos, particularmente da percepção. Baseada em Shah e Carpenter [5], descreve ainda que os gráficos são utilizados na apresentação de dados numéricos porque é conhecida a capacidade humana de reconhecer padrões visuais. Para Manovich [6], de maneira semelhante à arte modernista, que reduz o particular a um esquema platônico, a visualização de dados nos permite enxergar padrões e estruturas por trás de vastos conjuntos de dados, podendo ser classificada como uma nova abstração.

Tirando partido das características da percepção humana, no entanto, recursos podem ser utilizados na linguagem gráfica também para confundir o leitor. Tufte [7] descreve, além de métodos para criar gráficos bem desenhados, técnicas comuns utilizadas para atrapalhar o entendimento do leitor, tais como: a

ausência ou esquecimento da escala, a ocultação do valor inicial da escala, que deveria ser sempre zero (sob pena de causar uma desproporção entre os valores comparados) e a comparação de um elemento inteiro com parte dele (por exemplo: comparar valores do ano anterior, completo, com valores do ano corrente, ainda em curso). Assim, a visualização adequada pode ajudar na compreensão, memória e tomada de decisões, enquanto que, se indevida, pode atrapalhar na interpretação pelo leitor.

Para interpretar gráficos corretamente, Goldenberg [8] afirma que estudantes precisam de conhecimentos matemáticos, não apenas experiência perceptual. Goldenberg [8], Clement [9] e Gomes [10] afirmam que a interpretação de gráficos exige um conhecimento do sistema gráfico e que a sua dificuldade se deve ao fato de o sistema de representação não ser tão trivial, envolvendo regras que não são tão facilmente apreendidas pelos leitores. Considerando ainda as altas taxas de analfabetismo funcional em certos países como o Brasil [11], problemas na educação da estatística no ensino médio [3], culturas que favorecem a absorção de informação sem questionamento [3] e dificuldades na compreensão de abstrações [12] é importante tratar os problemas existentes em visualização de informação e construção de visualizações, especialmente por usuários novatos.

1.1.Problema, questão, objetivos e escopo da pesquisa

Sistemas computacionais que permitem a construção e a interação com gráficos pelo usuário, no entanto, podem influenciar na interpretação dos dados e na compreensão do próprio sistema gráfico. Tais ferramentas têm o objetivo de melhorar a experiência do usuário na visualização de dados e motivar seu interesse. Entre as ferramentas de visualização de dados estatísticos, podemos citar algumas que estão disponíveis na Web, tais como: Manyeyes,¹ GapMinder,² Worldmapper,³ Statplanet,⁴ Google Public Data,⁵ diversos atlas multimídia,

¹<http://www-958.ibm.com/software/analytics/manyeyes/>

²<http://www.gapminder.org/>

³<http://www.worldmapper.org/>

⁴<http://www.sacmeq.org/interactive-maps/statplanet/StatPlanet.html>

⁵http://www.google.com/publicdata/directory?hl=en_US&dl=en_US#!

SIDRA,⁶ Séries estatísticas⁷. Existem ainda diversos *toolkits* disponíveis, por exemplo: Flare,⁸ Silverlight,⁹ JavaScript InfoVis toolkit,¹⁰ ivtk,¹¹ que permitem a construção de gráficos. No entanto, Elias e Bezerianos [13] alertam que as ferramentas para usuários inexperientes, tais como ManyEyes, Sense.us [14] e Polstar [15], normalmente restringem usuários a uma única visualização por vez. Eles também afirmam que os *toolkits* não visam usuários novatos e normalmente necessitam de programação adicional para realizar as operações para a criação de visualizações. Neste caso, a curva de aprendizado é normalmente alta e visa usuários mais experientes.

Mackinlay *et al.* [16] apresenta a necessidade de pesquisas nessa área quando afirma que "*todos os analistas têm conhecimento sobre seus problemas, mas apenas poucos têm habilidades para construir representações gráficas eficientes da informação*". Ainda segundo eles, "*pessoas precisam de sistemas de análise visual que automaticamente apresentem dados usando as melhores práticas do design gráfico*".

Com base em resultados de uma pesquisa qualitativa realizada com as ferramentas SIDRA, Statplanet e Séries estatísticas [17], também observamos problemas nas construções de visualizações. Observamos que quanto mais opções de visualização são oferecidas para o usuário, mais problemas de entendimento foram encontrados. No entanto, as ferramentas mais simples eram muito limitadas em relação às construções gráficas. Em ambos os casos, a construção de gráficos foi influenciada pela ferramenta, acarretando resultados inadequados ou ineficientes para a tarefa solicitada.

Assim, definimos nossa questão de pesquisa da seguinte forma: ***Como apoiar usuários inexperientes na construção de visualizações eficientes com dados estatísticos?***

No sentido de respondê-la, primeiramente realizamos um estudo, fundamentado na teoria da Engenharia Semiótica [18], que visava investigar a seguinte sub-questão:

⁶ <http://www.sidra.ibge.gov.br/>

⁷ <http://seriesestatisticas.ibge.gov.br/>

⁸ <http://flare.prefuse.org/>

⁹ <http://www.silverlight.net/>

¹⁰ <http://thejit.org/>

¹¹ <http://ivtk.sourceforge.net>

SQ1. Como sistemas de significação e comunicação influenciam na interpretação de dados estatísticos por usuários interessados em responder determinadas perguntas?

Paralelamente, verificamos que a formulação da pergunta sobre a informação a ser obtida em um gráfico é uma etapa fundamental para a leitura e construção de gráficos [19]. Assim realizamos um estudo para responder nossa segunda sub-questão:

SQ2. Como criar significados formais que relacionem perguntas dos usuários sobre diferentes tipos de dado com visualizações eficientes?

Esta pesquisa nos levou a construir uma ontologia para definir relações entre dados e modelos de visualização a fim de responder perguntas de usuários através de visualizações eficientes.

Com a ontologia definida e com o objetivo de utilizá-la como recurso em ferramentas de visualização para oferecer ajuda ao usuário inexperiente, desenvolvemos a ferramenta ViSC (*Visualization with Smart Charts*). ViSC é uma ferramenta para visualização e construção de dados estatísticos que inclui um sistema de recomendação, cuja base de conhecimento é a ontologia desenvolvida, que associa as perguntas mais comuns que usuários costumam querer responder com gráficos. Ao apresentar uma gama de perguntas relacionadas aos dados correntes para o usuário, acreditamos que isso possa ajudá-lo na formulação da sua pergunta e, dessa forma, guiá-lo para uma análise de visualizações eficientes para respondê-la e assim, construir um gráfico eficiente. Para avaliar como essas recomendações apóiam o usuário, elaboramos a seguinte sub-questão:

SQ3. Como as perguntas influenciam a realização da tarefa e as visualizações geradas?

Para entender a influência das perguntas, devemos observar, nessa avaliação, a interação do usuário com as perguntas, gerando novas sub-questões:

SQ4. De que forma o usuário interage com as perguntas da interface?

Com esta pergunta visamos descobrir se ele interagiu com a perguntas e em caso positivo, como e com que objetivo ele interagiu com as perguntas.

SQ5. Como as perguntas são entendidas pelo usuário?

Com esta pergunta visamos verificar se o usuário entende como o sistema sugere, altera e ordena as perguntas. Desejamos verificar ainda se o usuário

observa a classificação de relevância das perguntas e neste caso, se concorda com essa classificação.

Para entender como seriam os resultados dos usuários em uma interface sem a as perguntas e com a utilização de outros recursos, fizemos também uma avaliação com outra ferramenta de visualização, a Tableau Public 7¹², uma vez que abrange o mesmo público alvo, é baseada em apresentações automáticas, possui um sistema recomendação de visualizações e é uma ferramenta comercial que possui diversos recursos para construção e gráficos e para análises visuais. A partir daí surgiu uma nova sub-questão para ser respondida na avaliação:

SQ6. Quais foram as facilidades e dificuldades dos usuários em cada ferramenta?

Com essa sub-questão pretendíamos analisar, além das soluções de IHC dadas pelas duas ferramentas, se **as perguntas de ViSC tiveram influência na realização da tarefa** e nas visualizações geradas e se essa influência foi positiva ou negativa. Indicadores para esta análise, foram dados pela observação da interação do usuário com as perguntas de ViSC, dos resultados obtidos através dos métodos utilizados e de suas respostas no questionário e nas entrevistas.

A pesquisa guiada por essas sub-questões teve como objetivo:

- Identificar a influência das perguntas na realização da tarefa e das visualizações geradas.

E com isso, alcançar o objetivo mais amplo:

- Verificar se perguntas são ou não uma solução para apoiar os usuários inexperientes na construção de gráficos eficientes

Dentre os objetivos secundários podemos citar:

- Identificar signos das duas interfaces que atrapalharam o usuário na realização da tarefa.

- Validar a ontologia utilizada como base para as recomendações.

1.2. Métodos e resultados

O primeiro passo foi pesquisar os problemas que existiam na área de visualização de informações com dados estatísticos. Isso foi feito através de leitura de publicações de psicologia e educação. Depois, realizamos o primeiro

¹² <http://www.tableausoftware.com/public>

estudo [17], que revelou problemas de interpretação e construção de gráficos por usuários inexperientes. Assim, escolhemos um problema e restringimos o escopo do trabalho a buscar uma solução apenas para o problema de construção de gráficos por usuários inexperientes.

Para essa primeira pesquisa [17], procuramos diversas ferramentas de visualização e selecionamos três para fazer uma avaliação com usuários: Statplanet, SIDRA e Séries Estatísticas. As três ferramentas possuem características diferentes se destinam à públicos alvo diferentes. Criamos um cenário que pudesse ser resolvido com o apoio dessas ferramentas. Utilizamos os seguintes métodos: o Método de Avaliação de Comunicabilidade (MAC) [20] e *Think Aloud* [21] com co-participação [22]. O primeiro tinha como objetivo avaliar falhas de recepção da metagemagem designer-usuário pelo usuário e revelar características dos signos computacionais que atrapalham essa metacomunicação. O segundo foi usado para investigar os processos de construção do raciocínio dos usuários, ao estimular a verbalização de seus pensamentos e motivar a conversa entre eles. Essa pesquisa nos retornou problemas de IHC na interface das ferramentas e a necessidade de uma interfaces simples com maior flexibilidade para a construção de gráficos voltada para usuários inexperientes.

Paralelamente, estudamos regras do sistema gráfico encontradas na literatura. As fontes mais importantes foram *Semiology of Graphics*, de Bertin [19], *Show me the numbers*, de Stephen Few [23] e as obras de Tufte [7]. Vimos que a formulação da pergunta era um requisito importante para a leitura de gráficos [19]. Pesquisamos também ontologias de visualização existentes. A partir dessas pesquisas criamos uma ontologia para definir as relações entre dados e modelos de visualização a fim de responder perguntas de usuários através de visualizações eficientes [24].

Essas pesquisas sobre visualização e a criação da ontologias foram as bases para a construção ViSC, *Visualization with Smart Charts*. Os requisitos para a construção da nossa ferramenta de visualização — ViSC — foram definidos conforme a lista abaixo:

1. Gerar gráficos eficientes, claros e precisos;
2. Motivar usuários inexperientes a analisar dados graficamente;
3. Permitir vários tipos de construção e operações;

4. Elaborar significados precisos da visualização em ontologias;
5. Prover *feedback* visual, visualizações automáticas e valores *default*;
6. Possuir um recurso de ajuda interativo.

Estudamos ainda técnicas de interfaces inteligentes e sistemas de recomendação ao mesmo tempo que pesquisamos também novas publicações e ferramentas de visualização que utilizam apresentações automáticas e visam o usuário inexperiente. Nesse contexto, encontramos Explorations Views [13] e Show Me [16], um conjunto de comandos integrados da ferramenta Tableau.

Analizamos vários *toolkits* de visualização e escolhemos um para desenvolvermos a primeira versão da ferramenta. Através do Método de Inspeção Semiótica (MIS) [25] identificamos algumas oportunidades de melhorias:

1. Comunicar melhor como as questões estão relacionadas às visualizações.
2. Mapear construções improdutivas para melhorar as recomendações.
3. Desenvolver mecanismos de filtro e zoom para os gráficos.

Como nosso objetivo era observar as questões da interface, alteramos a ferramenta criando uma nova solução para apresentar as questões e incluímos um nível de zoom. Ainda tínhamos poucas tarefas definidas. Aproveitamos para incluir novas tarefas e perguntas. Com base no trabalho de Amar *et al.* [26], as tarefas foram alteradas para 10 tarefas de níveis mais baixo, entre as quais incluía a tarefa de filtrar, um requisito encontrado na inspeção.

Para verificar então como as perguntas de ViSC influenciaram no resultado e poder responder a nossa questão mais ampla, selecionamos uma segunda ferramenta com outros recursos de interação, a Tableau Public 7, como descrito na seção anterior e utilizamos a triangulação de dois métodos qualitativos. Primeiramente realizamos uma inspeção através do MIS [25] nas duas ferramentas, depois realizamos uma avaliação com os usuários que chamamos de “*Retrospective Communicability Evaluation*” (RCE), técnica criada a partir da observação com usuários, com *Retrospective Think Aloud* (RTA) [27] e posterior etiquetagem e definição do perfil semiótico conforme no Método de Avaliação de Comunicabilidade (MAC) [28]. A escolha do MIS visava avaliar falhas na emissão da mensagem em relação às perguntas oferecidas na interface de ViSC e dos recursos disponíveis no Tableau, revelando características dos signos que poderiam não ser bem transmitidos na metacomunicação. A aplicação do RCE

tinha como objetivo revelar os processos de raciocínio e geração de hipóteses na experiência dos usuários com o uso das ferramentas e na construção de gráficos, bem como comparar as metagensagens recebidas com as emitidas.

Como optamos por uma pesquisa qualitativa em profundidade, envolvemos um número reduzido de participantes, definindo um total de 6 usuários. Cada um deles teve que cumprir todas as etapas da avaliação que consistiam: a) no preenchimento do questionário pré-teste e realização de uma tarefa de construção de gráficos em papel a fim de avaliar os requisitos necessários para realização do teste e o perfil do usuário; b) realização de duas tarefas, uma em cada ferramenta e de acordo com o cenário. c) *Retrospective Think Aloud*: observação do vídeo de sua própria avaliação realizado previamente narrando os pensamentos e ações que pretendia executar durante o teste; e d) preenchimento de questionário pós-teste estruturado e entrevista semi-estruturada.

Os resultados dos testes realizados com os usuários foram interpretados, analisados e posteriormente comparados com o resultado do MIS. Com a obtenção dos resultados na avaliação, respondemos as sub-questões de pesquisa e por último a nossa questão mais ampla.

1.3. Organização do trabalho

O Capítulo 2 descreve os trabalhos relacionados, incluindo conceitos e técnicas de semiótica que são importantes para a compreensão deste trabalho. A Seção 2.1 descreve resumidamente algumas regras do sistema gráfico que foram pesquisadas para a análise das ferramentas, para a criação da ontologia e da ferramenta ViSC. A Seção 2.2 relaciona o conceitos de sistemas de significação e comunicação com ferramentas de visualização. A Seção 2.3 explica o conceito de abdução e sua importância no contexto da nossa análise, a Seção 2.4 compara algumas técnicas de interface sistemas de visualização. Por último, a Seção 2.5 resume os tipos de sistema de recomendação estudados e explica o porquê da escolha do tipo utilizado em ViSC.

O Capítulo 3 descreve os estudos que levaram ao desenvolvimento da ferramenta. A Seção 3.1 resume a pesquisa realizada anteriormente que investigou como as ferramentas de visualização influenciam o usuário na interpretação de

visualizações. A Seção 3.2 apresenta a ontologia de visualização desenvolvida e utilizada em ViSC.

O Capítulo 4 explica a ferramenta proposta. Primeiramente, a Seção 4.1 descreve os requisitos, a Seção 4.2 descreve a metamsagem obtida a partir dos requisitos, a Seção 4.3 explica sua gramática, a Seção 4.4 descreve o funcionamento do sistema de recomendação adotado. Em seguida, na Seção 4.5, apresentamos a interface e, na Seção 4.6 descrevemos como é feita a interação usuário-sistema. Por último, na Seção 4.7 fazemos comparações com as ferramentas relacionadas e apresentamos as contribuições esperadas.

O Capítulo 5 apresenta todas as etapas das avaliações realizadas. Primeiramente, a Seção 5.1 explica os métodos utilizados. Depois, a Seção 5.2 descreve a etapa de preparação incluindo os cenários criados. 2. A Seção 5.3 descreve todas as etapas da inspeção das duas ferramentas realizadas com utilização do Método de Inspeção Semiótica (MIS). Em seguida, a Seção 5.4 apresenta os dados coletados e a interpretação dos resultados através do método "*Retrospective Communicability Evaluation*" (RCE), separados por tarefa. Ao final de cada tarefa também apresentamos uma discussão sobre cada tarefa e, ao final da seção, realizamos uma discussão sobre as ferramentas e finalizamos com a elaboração do perfil semiótico. Por último, a seção 5.5 consolida os resultados das avaliações através de uma discussão final respondendo as sub-questões dessa pesquisa.

O Capítulo 6 finaliza a dissertação, trazendo uma conclusão geral e respondendo a questão principal deste trabalho. Apresenta ainda as contribuições realizadas e as necessidade de trabalhos futuros.

2 Trabalhos relacionados

Nosso primeiro trabalho relacionado foi no sentido de entender as regras e problemas do sistema gráfico (Seção 2.1). Depois, estudamos dois conceitos da Engenharia Semiótica — sistemas de significação e comunicação (Seção 2.2); abdução (Seção 2.3) — a fim de realizar uma pesquisa qualitativa com usuários de algumas ferramentas de visualização. Em uma fase posterior estudamos técnicas de visualização em sistemas de interação (Seção 2.4) para basear o desenvolvimento da ferramenta e por último estudamos tipos de sistema de recomendação (Seção 2.5) para criar nosso recurso de ajuda interativo através de perguntas.

2.1.As propriedades e as regras do sistema gráfico

Ignatius e Senay [29] explicam que marca é o bloco de construção mais primitiva que pode codificar algumas informações úteis na visualização de dados. Explicam ainda que as marcas possuem propriedades. Essas propriedades são classificadas em posicional, temporal e de retina. Os dados são geralmente codificados através da variação das propriedades das marcas em uma técnica de visualização. A codificação posicional da informação é uma variação das posições das marcas na imagem. Uma marca pode então expressar uma correspondência entre duas séries constituídas por duas dimensões planares, horizontal e vertical. A codificação temporal da informação é uma variação das propriedades da marca ao longo do tempo. A codificação de informação de retina é qualquer variação nas propriedades das marcas de retina que é sensível pela retina do olho independente da posição das marcas. As propriedades das marcas de retina incluem tamanho, textura, orientação, forma, e os três parâmetros de cor (matiz, saturação e brilho).

É importante entender os tipos de percepção humana para entender quando e como utilizar às variáveis de um gráfico em cada marca de retina. Assim, listamos os quatro tipos de percepção segundo Bertin [19]:

1. Associativa: É útil quando se está buscando equalizar uma variação e para agrupar correspondências com todas as categorias desta variação combinada. Variações na forma, orientação, matiz e textura são associativas. Variações na forma e na orientação são utilizadas para revelar elementos similares e diferentes e facilitar a identificação externa através do simbolismo forma.

2. Seletiva: Utilizada para obter uma resposta para uma questão: “*Onde está a categoria?*” O olho deve ser capaz de isolar todos os elementos desta categoria, desconsiderar todos os outros sinais e perceber a imagem através da categoria dada. Tamanho, brilho, matiz, textura e orientação são percebidas imediatamente. Porém, Bertin [20] afirma ainda que a percepção seletiva é muito limitada e não aconselha mais que 5 passos para definir categorias no tamanho, e 7, no brilho (em escala de branco, cinza e preto).

3. Ordenada: Usada em comparações de duas ou mais ordens. Textura, brilho e tamanho estabelecem ordens que são universais e imediatamente perceptíveis. Não se pode reordenar o brilho.

4. Quantitativa: usada quando se busca definir numericamente a relação entre dois signos e quando se busca agrupar signos homogeneamente, isto é, que envolvem pequenas distâncias quantitativas e, portanto, definir os passos naturais resultantes de um estudo estatístico. Somente tamanho é quantitativo.

Para compreender sobre o problema da compreensão do sistema gráfico e da construção de gráficos, resumimos as principais regras segundo a semiologia de Bertin [19].

Segundo Bertin [19, p. 100], o problema básico da construção de gráficos é a definição de que representação se deve utilizar. Sobre o uso apropriado de um gráfico, afirma que depende de vários fatores. A decisão de transcrever a informação graficamente deve ser baseada na avaliação de propriedades específicas e eficiência de cada linguagem, ou seja, cada sistema de signo. Tal decisão depende de hábitos adquiridos, aptidões pessoais ou até mesmo de moda. A Figura 1 ilustra o problema da escolha e o conjunto dos principais tipos de gráficos de representações possíveis para representar um conjunto de informação.

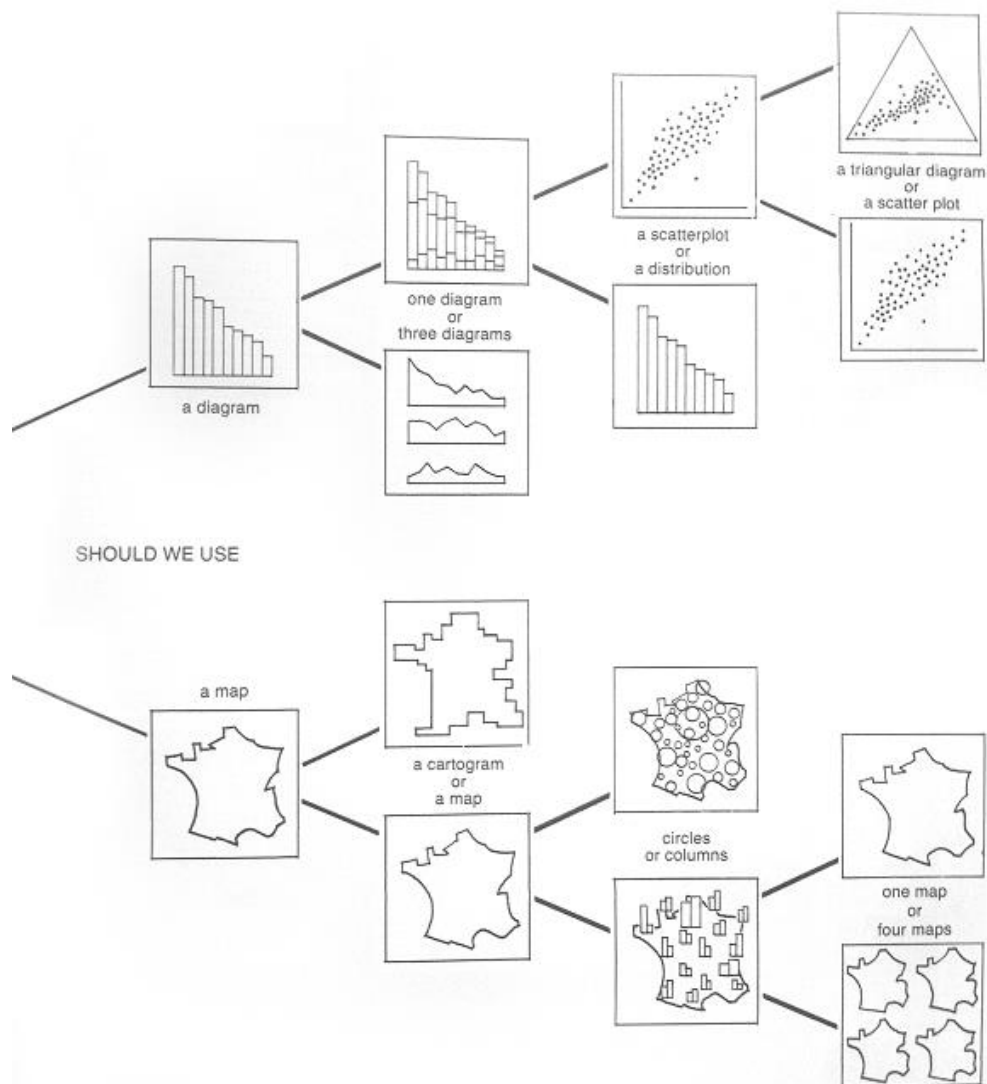


Figura 1: O problema da escolha da representação mais eficiente

Explica ainda que ler um gráfico é seguir mais ou menos rapidamente as sucessivas operações:

1. Identificação externa: Esta operação depende de hábitos adquiridos no reconhecimento de palavras, formas ou cores. Isso nos permite isolar do conhecimento humano o domínio preciso tratado pela figura.
2. Identificação interna: Nessa operação, o leitor deve reconhecer que variáveis visuais cada um dos componentes é representado no gráfico.

3. Percepção de correspondências pertinentes: Após a identificação, o leitor está pronto para a percepção das correspondências que o desenho isola a partir de um vasto número de correspondências. **Para isso, devemos formular uma questão, o que é feito conscientemente ou não. As questões podem ser muitas, no entanto, para cada conjunto de dados, elas são um número finito e podem ser rigorosamente definidas.**

As regras para a construção de um gráfico podem ser resumidas da seguinte forma:

1. Representar a informação em uma imagem única ou em um número mínimo de imagens necessárias.
2. Simplificar a imagem sem reduzir o número de correspondências.
3. Simplificar a imagem por redução e assim criar uma mensagem clara e eficiente

Sobre as regras de legibilidade, Bertin explica que um gráfico deve utilizar o intervalo de diferenciação perceptível proporcionada pelas variáveis visuais de forma fique com uma densidade gráfica ótima e o olho possa:

1. Separar as duas dimensões planares e evitar o "esmagamento" do plano, o que limita uma diferenciação angular. Isso é chamado de separação angular. A Figura 2 exemplifica problemas de separação angular.

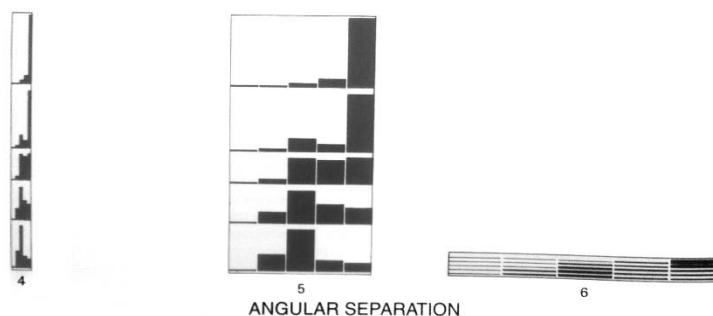


Figura 2: Exemplos de problemas de separação angular

2. Separar as marcas significativas daquelas sem significado, ou seja, separar a figura do fundo e conteúdo da forma. E, quando se utiliza

variáveis de retina (tamanho, textura, orientação, forma, matiz e brilho), separá-las de forma que a imagem não fique muito fraca ou muito densa.

2.2. Sistemas de significação e comunicação e ferramentas de visualização

Segundo de Souza [18], "*sistemas de significação são o resultado de funções sígnicas culturalmente convencionalizadas. Em outras palavras, quando uma cultura produz uma convenção que correlaciona certo conjunto de elementos expressivos para um determinado conjunto de conteúdos, um sistema de significação está em vigor. Sistemas de significação constituem os códigos disponíveis para o processo de comunicação. Esses processos, no entanto, não estão restritos a expressões que pertencem a um sistema de significação. Processos de comunicação exploram e trabalham em sistemas de significação existentes, mas comunicadores podem escolher usar funções sígnicas latentes ou inovadoras permitidas por sinais do plano expressivo que tem potencial de participar em funções sígnicas*".

Barbosa e Silva [30] explicam que, em um processo de comunicação, sistemas de significação correspondem às formas de representar (expressão) seus significados pretendidos (conteúdo), de modo a alcançar uma variedade de objetivos (intenção).

Dessa forma, podemos entender que o sistema gráfico (representado por gráficos cartesianos, mapas, cartogramas ou até mesmo outras formas de representação gráfica) constitui sistemas de significação, pois é definido a partir de convenções matemáticas, estatísticas e geográficas que nos permitem interpretar informações, ou seja, os códigos para o processo de comunicação.

Ferramentas de visualização, por sua vez, são sistemas computacionais interativos que utilizam representações específicas do sistema gráfico. A interação permite a criação de novos signos que são compatíveis com as funções sígnicas desse sistema gráfico. Cada uma dessas ferramentas possui também uma linguagem própria, seu sistema de significação, que é uma linguagem artificial constituída por signos presentes na interface e signos computacionais processados

através da interação entre usuário-sistema. Tais signos comunicam a mensagem do designer para o usuário sobre como eles podem e devem utilizar o sistema, por que e com que efeitos, iniciando um processo de comunicação entre o usuário e o designer mediada pelo sistema (metacomunicação designer-usuário).

As ferramentas de visualização compõem, portanto, sistemas de significação e comunicação em diferentes níveis: pelo conjunto de visualizações geradas e pelo conjunto de signos produzidos pelo designer para interação com a ferramenta. O processo de comunicação utiliza como forma de representar (expressão) esses dois sistemas de significação, seu significado (conteúdo) é composto pelo conjunto de dados disponíveis e seus objetivos (intenção) correspondem à intenção de responder perguntas de usuários sobre diversos dados simultaneamente e estimular o aprendizado do sistema gráfico. A utilização dos novos signos criados pela interação permite influenciar no entendimento do sistema gráfico e na compreensão da mensagem pelo usuário.

Pela Engenharia Semiótica [18] que é uma teoria de IHC centrada na comunicação e caracterizada como um caso de comunicação humana mediada por sistemas, uma interface deve comunicar ao usuário a lógica do design (intenções do designer que criou a interface e os princípios de interação definidos no processo de design). Comunicabilidade pode ser definida então, como um conceito de qualidade para avaliar se essa lógica é comunicada de forma eficiente e efetiva para os usuários. O entendimento do conceito de sistemas de significação e comunicação foi importante para analisar a comunicabilidade nas ferramentas estudadas e para a criação da ferramenta ViSC.

2.3. Conceito de abdução

Abdução é normalmente considerada como sendo um dos três tipos principais de inferência, sendo as outras duas dedução e indução. Em inferências dedutivas, o que se infere é necessariamente verdade se as premissas a partir das quais se infere são verdadeiras, isto é, a verdade das premissas garante a verdade da conclusão [31].

Para Santaella [32], o conceito de abdução é o raciocínio que leva à adoção de uma hipótese para ser comprovada e indução é o teste da hipótese. Segundo a mesma autora, em 1901, Peirce definiu abdução como “*aceitação ou criação de*

uma premissa menor como uma solução hipotética para um silogismo cuja premissa maior é conhecida e cuja conclusão descobrimos ser um fato” [32, p. 92]. Este conceito foi logo depois ampliado para a visão de que ela “*consiste no exame de uma massa de fatos permitindo que estes fatos sugiram uma teoria*” [32, p. 92]. Ainda segundo a autora, sobre abdução, Peirce afirmou: “*A explicação deve ser uma proposição tal que levaria à predição dos fatos observados, quer como conseqüências necessárias, quer, pelo menos, como muito prováveis sob certas circunstâncias. Uma hipótese, então, tem de ser adotada como plausível nela mesma e tornando os fatos plausíveis. Esse passo de adotar uma hipótese como sugerida pelos fatos é que chamo de abdução*” [32, p. 93].

Segundo de Souza [18, p. 43], “*o interessante para a ciência da computação é que o raciocínio abduativo não depende da existência do conhecido, de primitivas estáveis ou de axiomas. Os elementos que funcionam como primitivas e axiomas são hipotetizados, produzidos por demanda, e testados para adequação contra evidências coletadas pelo intérprete humano*”.

Para a nossa análise é interessante observar que existem dois níveis de raciocínio abduativo. No primeiro nível, as técnicas de visualização de informações que representam graficamente dados de um determinado domínio e, no segundo nível, as ferramentas de visualização com sua linguagem artificial constituída de seus signos de comunicação e metacomunicação. No primeiro nível, o usuário deve desenvolver hipóteses para a melhor representação gráfica dos dados. No segundo nível, a linguagem artificial da ferramenta deve ser aprendida pelo usuário através de um processo de aquisição, que envolve o reconhecimento de padrões simbólicos. No entanto, os processos de ambos os níveis envolvem correções de suas hipóteses feitas com base no entendimento da linguagem, que é adquirido e refinado através da interação do usuário com a ferramenta.

Sistemas computacionais interativos podem apoiar o raciocínio abduativo do usuário na busca de obter representações eficientes e compreender as informações com mais facilidade. Para isso, no entanto, o usuário deve compreender a linguagem da ferramenta. Se as hipóteses formuladas pelos usuários convergirem para construções gráficas eficientes e análises corretas, mais eficiente podemos considerar a comunicação entre o designer da ferramenta e o usuário.

2.4. Técnicas de visualização em sistemas interativos

Conforme afirma Manovich [6], o uso de meios eletrônicos muda as representações, sendo possível visualizar maiores conjuntos de dados, criar visualizações dinâmicas, alimentar dados em tempo real, fundamentar representações gráficas de dados sobre análises matemáticas, entre outras operações. Heer e Shneiderman [33] auxiliam *designers*, pesquisadores, analistas profissionais, agentes de compras, educadores e estudantes na avaliação e criação de ferramentas de análises visuais apresentando uma taxonomia de interação dinâmica que contribui para o sucesso dos diálogos analíticos em ferramentas de análises visuais, estudo que foi importante para definirmos as técnicas a serem utilizadas no sistema ViSC.

Nesta pesquisa, Heer e Shneiderman [33] descrevem três alternativas utilizadas em projetos de interfaces de visualização de dados. A primeira é através de tipologia de gráficos (*chart typology*), uma paleta de *templates* de visualizações disponíveis na qual analistas podem apresentar seus dados. Segundo ele, esse método é familiar aos usuários de planilhas e possui benefícios como a simplicidade e familiaridade, mas limita os tipos de visualização e torna custoso o experimento de diferentes visualizações com o mesmo dado. A Figura 3 apresenta um exemplo de uso de tipologia de gráficos na interface do assistente gráfico do Excel 2003.

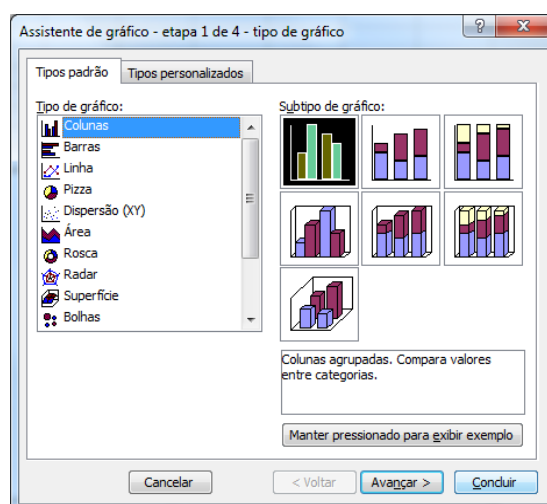


Figura 3: Assistente de gráfico do Excel 2003. Exemplo de chart Typology

A segunda alternativa apresentada é através do uso de grafos de fluxo de dados (*data-flow graphs*), no qual o processo de visualização é desconstruído em conjuntos refinados de operadores para importação, transformação, layout, etc. Através das flexíveis combinações permitem maior variação de *designs*, porém requer um maior esforço de entrada que o método anterior e ainda é limitado pelo conjunto de operadores. A Figura 4 apresenta um exemplo de uso de grafos de fluxo de dados na interface da aplicação *Impure*.¹³

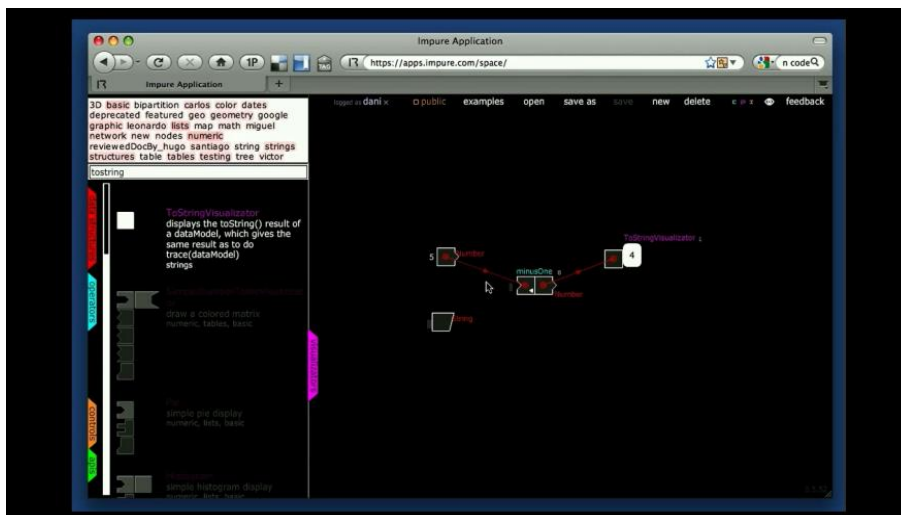


Figura 4: Workspace da ferramenta Impure

A terceira alternativa é baseada em gramáticas formais (*formal grammars*) para a visualização de construções, constituídas por linguagens de alto nível que descrevem como o dado deve ser mapeado nas características visuais. Exemplos: *ggplot for the R statistical analysis platform*,¹⁴ *Protovis para HTML* [34] e *Google Chart Tools*.¹⁵ No entanto, todas elas requerem um mínimo de habilidade em programação. A Figura 5 exemplifica essa alternativa apresentando um trecho de código desenvolvido com o *toolkit* Protovis e o resultado da visualização obtida a partir desse código.

¹³ <http://www.impure.com/>

¹⁴ <http://had.co.nz/ggplot2/>

¹⁵ <https://developers.google.com/chart/interactive/docs/index>

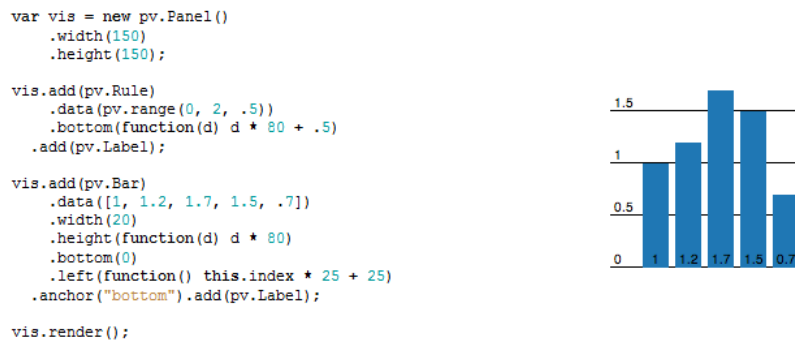


Figura 5: Código e visualização gerada a partir do código pelo toolkit Protovis

Ainda segundo Heer e Shneiderman [33], essas três alternativas não são mutuamente exclusivas. Pode-se aplicar um sistema de grafo de fluxo de dados ou gramáticas formais para definir novos componentes para incluir dentro de uma tipologia de gráficos, aproveitando melhor a expressividade das primeiras e da facilidade de utilização da última. A partir dessas técnicas selecionamos uma combinação de tipologia de gráficos e gramática formal para aplicar no desenvolvimento de nossa ferramenta. No entanto, a nossa gramática formal é usada internamente e, portanto, não requer habilidades de programação do utilizador.

2.5. Sistemas de recomendação

Segundo Ricci *et al.* [35], sistemas de recomendação são ferramentas de *software* e técnicas que fornecem sugestões de itens úteis para o usuário. As sugestões se referem a vários processos decisórios, tais como: itens para comprar, músicas para ouvir, ou notícias online para ler. Burke [36] afirma que a definição atual do termo tem uma conotação ampla, descrevendo qualquer sistema que produza recomendações individualizadas como saída ou que tem o efeito de guiar o usuário de forma personalizada a objetos úteis ou interessantes em um grande espaço de opções possíveis. Estes sistemas têm como apelo um ambiente onde a quantidade de informação on-line supera amplamente capacidade de qualquer indivíduo para examiná-lo.

As pesquisas sobre sistemas de recomendação tiveram como base os artigos de Robin Burke [36][37] e o livro *Recommender System Handbook* de Ricci,

Rocarch, Shapira, e Kantor [35]. Nosso objetivo com essa pesquisa era escolher a técnica mais adequada para, com o uso da ontologia, criar recomendações de visualizações através da de perguntas que usuários poderiam querer responder através da interação com o sistema.

Para resumir as técnicas de sistemas de recomendação na tabela a seguir, Burke [36] descreve inicialmente as partes que os compõem: (i) *background data*: a informação que o sistema tem antes do processo de recomendação começar; (ii) *input data*: a informação que o usuário deve comunicar para o sistema a fim de gerar uma recomendação; e (iii) um algoritmo que combina *background data* e *input data* para se chegar às sugestões.

Ele distingue cinco técnicas que são apresentadas na Tabela 1. Para compreensão, assumir que I é o conjunto de itens sobre os quais as recomendações são feitas, U é o conjunto de usuários de quem as preferências são conhecidas, u é o usuário para quem as recomendações precisam ser geradas e i é algum item para o qual nós gostaríamos de prever a preferência de u .

Tabela 1: Técnicas de recomendação [36]

Técnica	Background data	Input data	Processo
<i>Collaborative recommender systems</i>	Classificações de U dos itens de I .	Classificações de u dos itens de I .	Identificar os usuários de U semelhantes a u , e extrapolar de suas classificações em i .
<i>Content-based recommender systems</i>	Características dos itens de I	Classificações de u dos itens de I .	Gerar um classificador que encaixa a classificação de comportamento de u e usá-lo em i .
<i>Demographic recommender systems</i>	Informações demográficas sobre U e suas avaliações de itens de I .	Informação demográfica de u	Identificar os usuários que são demograficamente semelhante ao u , e extrapolar de suas classificações de i .
<i>Utility-based recommender systems</i>	Características dos itens de I	A função de utilidade sobre itens em I que descreve as preferências de u .	Aplicar a função para os itens e determinar a classificação de i .

<i>Knowledge-based recommender systems</i>	Características de itens de I . Conhecimento de como esses itens atendem às necessidades de um usuário.	A descrição das necessidades ou interesses de u .	Inferir uma correspondência entre i e a necessidade de u .
--	---	---	--

De acordo com Burke, *collaborative recommender systems* são o tipo mais conhecido, o mais amplamente implementado e com a mais madura das tecnologias. Sistemas de recomendação colaborativos agregam classificações ou recomendações dos objetos, reconhece semelhanças entre usuários com base em suas classificações e gera uma nova recomendação baseada nas comparações entre usuários. Um perfil típico de usuário em um sistema colaborativo consiste em um vetor com itens e suas classificações continuamente aumentado de acordo com a interação do usuário com o sistema ao longo do tempo.

Demographic recommender systems visam a categorizar o usuário com base nos atributos pessoais e fazer recomendações com base nas classes demográficas. Um exemplo deste tipo de sistema citado por Burke é o Grundy [38**Erro! Fonte de referência não encontrada.**], que recomendava livros baseado nas informações pessoais recolhidas através de um diálogo interativo. As respostas dos usuários foram comparadas com uma biblioteca de estereótipos usuário manualmente montados. Alguns sistemas de recomendação recentes também têm usado este método. Em alguns, utiliza-se uma pequena pesquisa que é usada para coletar dados do usuário, em outros, aprendizado de máquina é usado para se chegar nas classificações baseadas no dado demográfico. A representação da informação demográfica em um modelo de usuário pode variar muito. Técnicas de recomendações demográficas formam correlações "pessoa-a-pessoa" como método colaborativo, mas com o uso de dados diferentes. A vantagem de uma abordagem demográfica é que ela não exige uma história de classificação do usuário como em técnicas colaborativas (*collaborative*) e baseadas em conteúdo (*content-based*).

Content-based recommendation system é um desdobramento e continuação da pesquisa de filtragem de informação. Nesses sistemas, os objetos de interesse são definidos pelas características associadas. Por exemplo, sistemas de recomendação de texto podem utilizar as palavras dos textos como características. Um sistema de recomendação baseado em conteúdo aprende o perfil do interesse

do usuário nas características presentes nos objetos que o usuário classificou. O tipo de perfil do usuário derivado de um sistema baseado em conteúdo depende do método de aprendizado empregado.

Utility-based e *knowledge-based recommenders* não tentam construir generalizações a longo prazo sobre seus usuários, mas sim basear os seus pareceres sobre a avaliação da partida entre a necessidade de um usuário e um conjunto de opções disponíveis. *Utility-based recommenders* fazem sugestões com base em um cálculo da utilidade de cada objeto para o usuário. O problema central é como criar a função de utilidade para o usuário. A vantagem desse tipo de sistema é que ele pode fatorar atributos que não são produtos, tais como a confiabilidade do fornecedor e a disponibilidade do produto para o cálculo utilitário, tornando possível, por exemplo, para a troca de preços contra cronograma de entrega para um usuário que tem uma necessidade imediata.

Por último, *knowledge-based recommender systems* tentam sugerir objetos com base em inferências sobre as necessidades de um usuário e preferências. De certo modo, todas as técnicas de recomendação poderiam ser descritas como algum tipo de inferência. Abordagens baseadas em conhecimento se distinguem por terem conhecimento funcional: eles têm conhecimento sobre como um item em particular atende a uma necessidade do usuário em particular, e pode, portanto, concluir sobre a relação entre uma necessidade e uma possível recomendação. O perfil do usuário pode ser qualquer estrutura de conhecimento que suporta esta inferência. No caso do Google, ele pode simplesmente ser a consulta que o usuário tenha formulado.

Em [37], Burke acrescenta que uma vantagem do *knowledge-based recommender systems* é de não possuir os problemas de arranque, uma vez que não necessitam de classificações dos usuários. Além disso, não têm que coletar informações sobre um usuário em particular porque suas decisões não dependem de gostos individuais.

De acordo com as técnicas estudadas, a abordagem *knowledge-based recommender systems* foi considerada mais apropriada para recomendar visualizações, uma vez que tais recomendações não dependem muito do perfil do usuário, mas sim da tarefa que ele deseja realizar. Observamos que, através dos dados selecionados e visualizações pelo usuário, poderíamos inferir as perguntas que ele possa querer responder com base na sua seleção. À medida que o usuário

interage, as perguntas devem se modificar de acordo com o novo dado. O conhecimento dos itens foi dado pelo ontologia de visualização que relaciona as perguntas às visualizações. Apesar da escolha dessa abordagem para essa versão, sabemos que uma combinação com a abordagem colaborativa poderia contribuir para obtermos feedback do usuário e verificarmos como essas perguntas estão atendendo às tarefas.

3 Estudos preliminares

O primeiro estudo preliminar realizado tinha como objetivo compreender como as ferramentas de visualização influenciam nas visualizações geradas pelo usuário e responder a nossa sub-questão 1 (SQ1). As conclusões deste trabalho foram importantes para definirmos o problema que iríamos abordar nesta pesquisa. O segundo estudo visava buscar vocabulários formais para visualizações. A partir dele, verificamos a necessidade de desenvolver um vocabulário próprio de visualização que para relacionar perguntas dos usuários sobre diferentes tipos de dado com visualizações eficientes (SQ2). Esse estudo foi importante para definir a ontologia de conhecimento utilizada pela ferramenta ViSC.

3.1. Como as ferramentas de visualização influenciam nas visualizações geradas

Confirmamos as lacunas nas pesquisas sobre visualização quando analisamos problemas na construção e interpretação de gráficos [17] através de métodos de avaliações qualitativas com algumas ferramentas de visualização Web. Ao avaliar as ferramentas Statplanet, SIDRA e Séries estatísticas, definimos os requisitos para a criação da nova ferramenta.

O Statplanet é uma aplicação de visualização de dados interativa, que fornece dados estatísticos de países, para criar infográficos e cartogramas. Também tem como objetivo ser um software educacional. SIDRA e Séries estatísticas são ferramentas do IBGE que fornecem dados do Brasil em diversos níveis geográficos. A primeira é um aplicativo para recuperação de dados agregados e a segunda recupera e exibe informações estatísticas de pesquisas que têm uma periodicidade definida.

Para essa pesquisa, utilizamos dois métodos de IHC de avaliação com usuários: Método de Avaliação de Comunicabilidade (MAC) [28], pertencente à Engenharia Semiótica e *Think Aloud* [21] com co-participação [22]. A triangulação das duas formas de avaliação tinha como objetivo avaliar as falhas na

recepção da mensagem pelo usuário, identificar signos computacionais que ajudam ou atrapalham a comunicação e separar os dois processos abduativos de modo a permitir a análise dos dois níveis de interpretação: da ferramenta e do sistema gráfico a fim de respondermos nossa questão de pesquisa.

O perfil do usuário escolhido foi de pessoas com habilidade na leitura de gráficos. Por isso, preferencialmente selecionamos pessoas que fizeram vestibular há pouco tempo, como estudantes nos primeiros períodos da graduação. Como opção, estudantes e profissionais de áreas exatas também estariam dentro do perfil. Selecionamos 3 participantes para serem observados através do método MAC (grupo 1) e 3 duplas para realizarem o *Think Aloud* com co-participação (grupo 2). O tempo destinado para a tarefa foi de 30 minutos.

Analisando os resultados, podemos entender os problemas mencionados na construção de gráficos.

As duas primeiras tarefas foram realizadas no Statplanet. Ao serem solicitados para construir um gráfico para analisar que país, entre os 31 que fizeram parte do PISA de 2000, teve a evolução mais positiva, os usuários construíram os gráficos ilustrados na e na Figura 7. Note que a Figura 6 foi formada por uma composição de gráficos de colunas com diferentes números de colunas que não permite uma comparação. A Figura 7 também não é uma visualização eficiente porque não permite a visualização em um instante da percepção, conforme a definição de eficiência de Bertin [19] mas foi o resultado obtido por 3 entre as 6 avaliações feitas (duas do grupo 1 e uma do grupo 2). Seguindo a definição de eficiência de Bertin [19], a forma de representação mais eficiente do gráfico pedido no item 1 possível era o gráfico de dispersão (Figura 8), mas os usuários não estavam familiarizados com esse signo e não compreenderam sua mensagem. O gráfico de colunas agrupadas (Figura 10) não existia na ferramenta e seria uma opção para responder a pergunta.

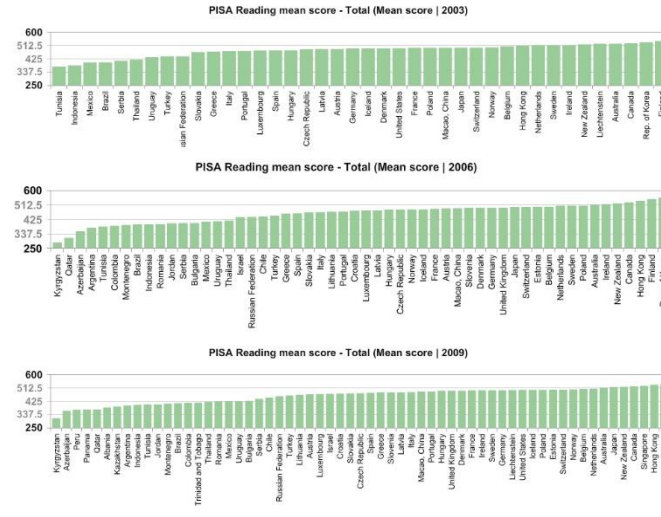


Figura 6: Resultado obtido pelo usuário A01 do grupo 1 para a tarefa 1

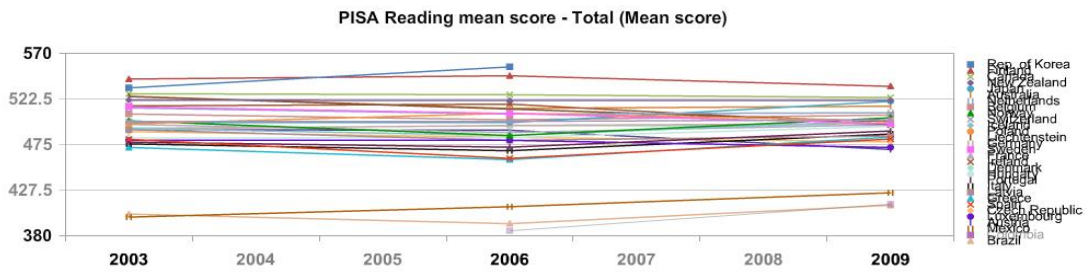


Figura 7: Resultado obtido por A02, A03 (grupo 1) e D03 (grupo 2) para a tarefa 1.

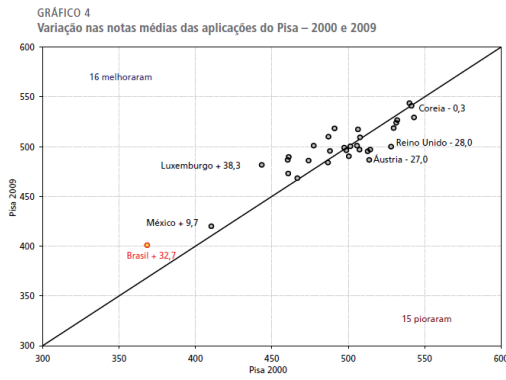


Figura 8: Gráfico de dispersão com diagonal

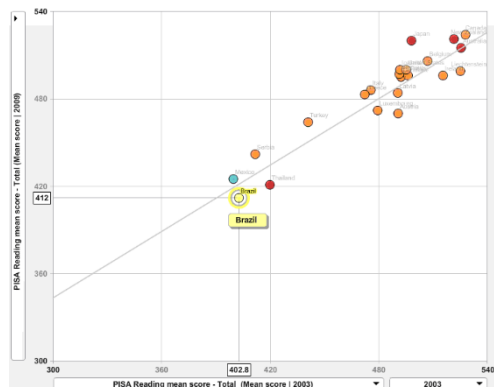


Figura 9: Gráfico de dispersão com reta de regressão linear. Statplanet

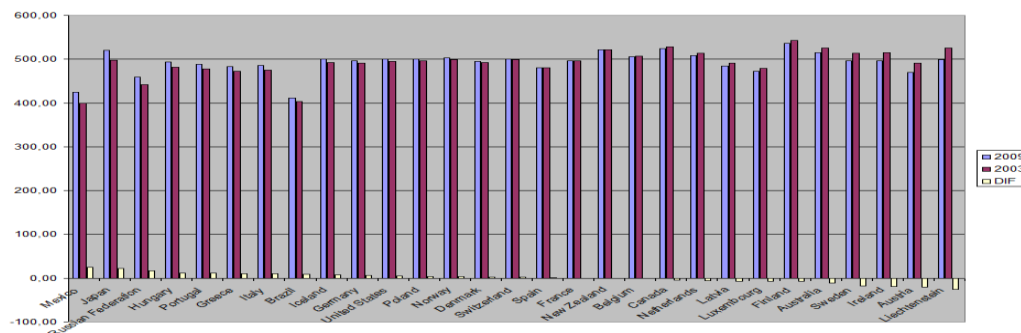


Figura 10: Gráfico de colunas agrupadas

Entre os signos computacionais que apresentaram problemas de comunicação encontramos problemas no gráfico de dispersão que não permitia interação de modo a traçar a diagonal. Note que o Brasil, representado pelo ponto vermelho está acima da reta no gráfico da esquerda (Figura 8) e abaixo da reta no gráfico da direita (Figura 9).

Entre os usuários do segundo grupo, os outros resultados (D01 e D02) estão apresentados na Figura 11 e na Figura 12. Note que em todos os resultados de colunas múltiplas, o número de colunas se torna mais denso a cada ano. Isso porque novos países passaram a realizar o exame. Para uma comparação eficiente, o número de colunas deveria ser igual, mas nenhum deles observou isso.

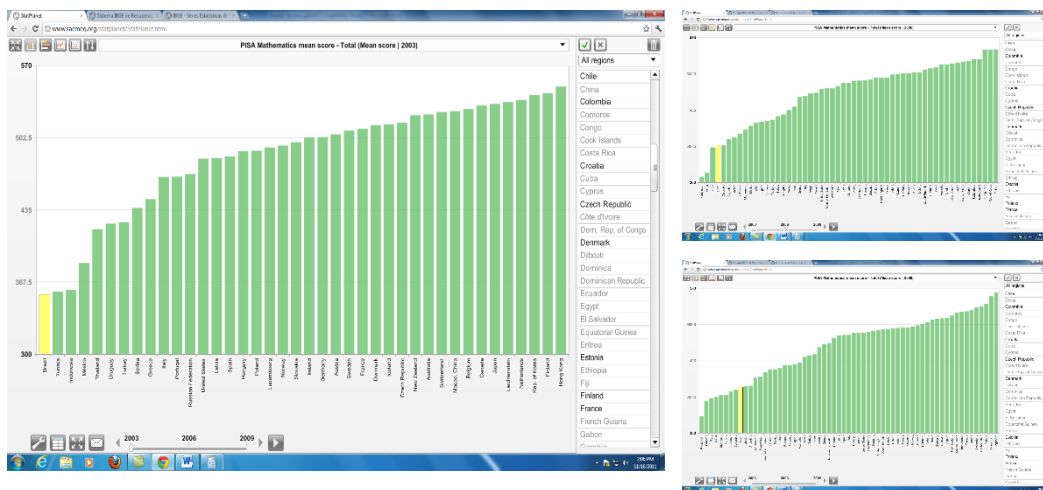


Figura 11: Resultado da dupla D01 (grupo 2) para a tarefa 1: 3 gráficos de barras com seleção do Brasil

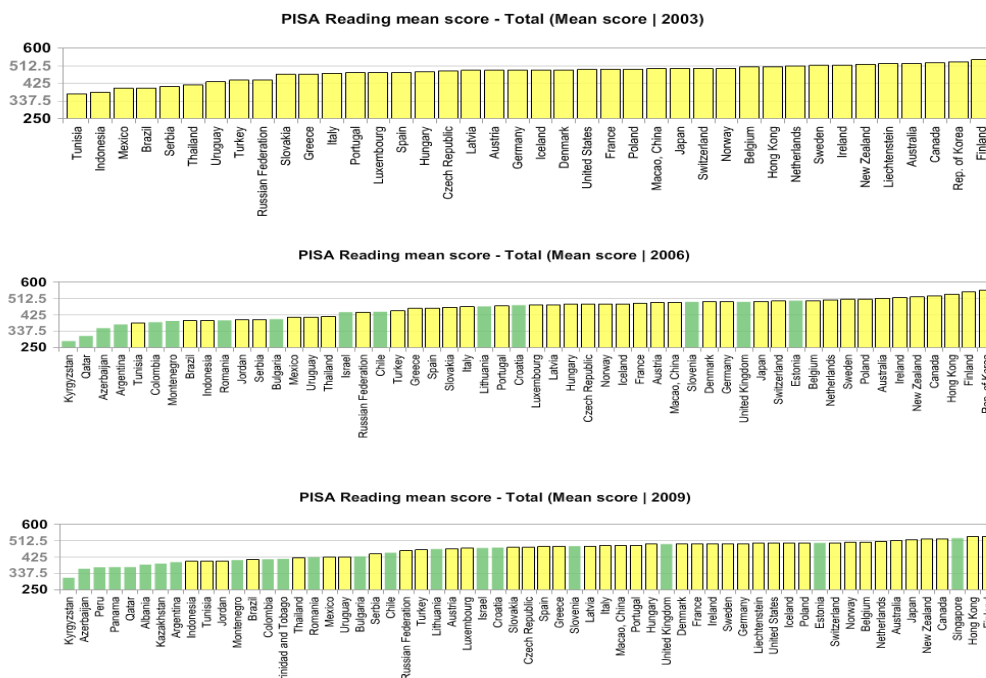


Figura 12: Resultado da dupla D02 (grupo 2): 3 gráficos de barras com os 32 países selecionados: Solução mais eficientes entre as apresentadas.

Considerando os gráficos realizados no item 1 da tarefa, observamos que a ferramenta Statplanet não levou à construção de gráficos eficientes. A interação com a ferramenta tornou possível algumas respostas corretas sobre países que tiveram evolução mais positiva que o Brasil, porém de maneira pouco eficiente.

A segunda tarefa pedia para apresentar a evolução das notas do PISA no Brasil e bastava apresentar um gráfico de linhas conforme exibido na Figura 13. Este resultado foi alcançado pelas 3 duplas (grupo 2) e por A03 (grupo 1). Mesmo

assim, entre os resultados individuais, os outros dois não chegaram a esse resultado. A01 compôs um gráfico de colunas no Excel com os 3 anos no eixo horizontal (Figura 80 — Apêndice A), o que a ferramenta não permitia e A02, criou uma composição com 3 gráficos de colunas também (Figura 81 — Apêndice A).

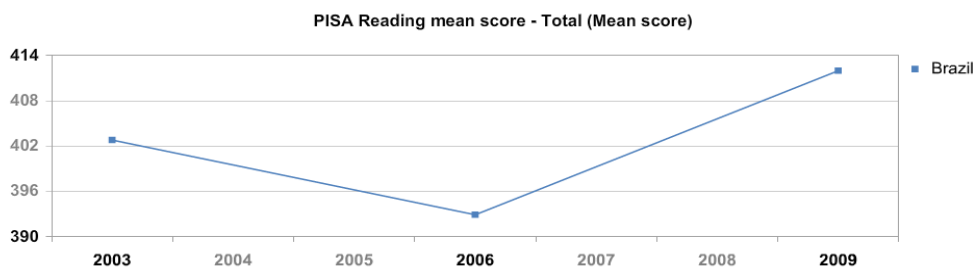


Figura 13: Resultado para a tarefa 2

O segundo item da tarefa foi mais simples que o primeiro e nesse momento os usuários já tinham formado explicações sobre o uso da ferramenta através de seu processo abduativo. A maioria dos usuários conseguiu gerar os gráficos da maneira esperada.

As explicações geradas no processo abduativo dos usuários sobre o funcionamento da ferramenta Statplanet foram, em muitos casos, inferidas através da repetição das mesmas rupturas de comunicação ou por opções redundantes que a ferramenta disponibilizava para a realização das tarefas. Ou seja, apesar dos problemas encontrados na customização e seleção pelo mapa, a ferramenta era flexível e de alguma forma os usuários conseguiram realizar as tarefas, entendendo seu funcionamento depois de repetidas rupturas ou de outra forma, porém rupturas no entendimento da gramática de construção dos gráficos foi mais grave.

A tarefa 3 pedia um gráfico para mostrar que o total de jovens com idade entre 15 e 16 aumentou no Brasil. Essa tarefa poderia ser feita tanto no SIDRA quanto no Séries estatísticas. O SIDRA possui diversos problemas de IHC em sua interface e visa usuários experientes. Possui 25 possibilidades de gráficos e permite manipular as dimensões desde que não haja mais que duas com múltiplos elementos selecionados. Séries estatísticas criam apenas um tipo de gráfico, de linhas e por isso só possuem dados que apresentam uma periodicidade.

Entre os usuários do grupo 1, nenhum conseguiu entregar um gráfico. A01 exportou uma tabela pelo Séries estatísticas. A02 tentou o SIDRA, mas logo desistiu e também exportou uma tabela. A03 começou pelo sistema Séries estatísticas, desistiu e tentou o SIDRA, tendo recorrido a manuais mas não realizou a tarefa. Entre as duplas do grupo 2, nenhuma conseguiu entregar a tarefa.

3.1.1. Discussão

Concluimos que, no SIDRA, onde temos maior complexidade dos dados e flexibilidade de construções, maior foi a dificuldade encontrada pelos usuários. As ferramentas mais simples (Statplanet e Séries estatísticas) também apresentaram problemas uma vez que permitem pouca interação e limitam as construções. Além disso, as ferramentas oferecem pouco ou nenhum apoio ao usuário inexperiente na construção de visualizações eficientes para responder a sua pergunta e ainda apresentam problemas em construções geradas que podem levar a interpretações equivocadas. Observamos ainda que a linguagem e a gramática das ferramentas de visualização tem grande influência sobre as visualizações geradas pelos usuários.

Além disso, foram encontrados diversos problemas de IHC em ferramentas de visualização que podem atrapalhar a interpretação dos dados e nas construções gráficas pelo usuário inexperiente. Entre os problemas de IHC da interface do Statplanet encontramos problemas de seleção, uma vez que o mapa era muito maior que a área de seleção através dos nomes dos países, dificuldade na criação de uma região com países personalizados devido à falta de *feedback* e falta de *wizard* para explicar a sequência. Sobre o SIDRA, apesar de ser uma ferramenta voltada para especialistas da área, usuários desistiram porque não conseguiram formular explicações sobre a sua linguagem em seu processo abutivo. Houve falhas completas de comunicação nas avaliações com usuários e comentários expressivos que evidenciaram a dificuldade dos usuários. Além disso, foram encontrados também outros problemas relatados na inspeção do MAC. Dessa forma, podemos concluir que essa ferramenta possui problemas de comunicabilidade e que necessitam de mais investigações. Novos testes com tarefas que mantenham o foco do usuário no SIDRA podem trazer resultados mais significativos sobre problemas dessa ferramenta. A outra ferramenta do IBGE,

Séries Estatísticas, foi mais bem compreendida pelos usuários que a utilizaram. No entanto, mesmo com signos estáticos em linguagem natural, encontramos rupturas de comunicação para encontrar os temas devido aos títulos longos, técnicos e muito semelhantes. Com isso, os participantes recorreram à busca, ignorando os menus de classificação. Também foram encontrados problemas de comunicabilidade nos gráficos gerados. O espaçamento entre os anos era o mesmo apesar de em alguns anos a pesquisa não ter sido realizada.

Sobre os métodos utilizados concluímos que foram complementares para os resultados dessa pesquisa. O MAC possibilitou encontrar rupturas de comunicação entre designer e usuário e divergências entre as metagensagens emitida e recebida, bem como identificar os signos computacionais que atrapalharam na geração dos gráficos. O *think aloud*, por sua vez, reforçou alguns pontos encontrados pelo MAC e ainda possibilitou identificar os signos que influenciaram usuários na escolha das construções gráficas.

Sobre o processo abduutivo de realização dos gráficos do item 1, concluímos ainda que os signos computacionais representados pelos gráficos foram determinantes para a escolha da representação pelo usuário. Como a gramática da linguagem da ferramenta era diferente para cada gráfico (barras e colunas eram preenchidos com todos os países por padrão enquanto o padrão de linhas era vazio e gráfico de dispersão precisava selecionar outro tema para o eixo y) alguns usuários encontraram essa facilidade (comportamento oportunista) do gráfico de colunas e optaram por essa opção compondo 3 gráficos, que apesar de não ser a resposta mais eficiente levou a construções mais claras. No entanto, outros, que tinham em seu modelo mental que tendências são construídas com gráficos de linhas, não mudaram seu modelo, resultando em construções confusas e ineficientes. Outro fato que influenciou para a escolha do gráfico de linhas foi ser o gráfico padrão da interface. Sobre os usuários que recorreram ao Excel podemos concluir que não compreenderam a gramática da linguagem da ferramenta.

Através da análise desses sistemas de significação concluímos que existem problemas de comunicação que necessitam de mais pesquisas de IHC na interpretação e construção de visualizações para apoiar à tomada de decisões por usuários inexperientes. Vimos ainda que a linguagem e a gramática das ferramentas de visualização têm grande influência sobre as visualizações geradas

pelos usuários. Se existem muitas rupturas na recepção da mensagem pelo usuário, isto é refletido para as construções gráficas e na interpretação dos dados.

Nesta seção apresentamos os resultados obtidos dos usuários e as conclusões obtidas. Mais detalhes sobre o estudo podem ser encontrados em [17].

3.2.Caracterização semântica de mecanismos de visualização

Com o objetivo de criar um sistema com uma interface inteligente que possa auxiliar o usuário na construção de visualizações eficientes, pesquisamos ontologias que poderiam agregar semântica à nossa ferramenta, melhorando a comunicação entre os usuários e o sistema de visualização. Durante a pesquisa, encontramos diversos estudos que demonstravam a necessidade de criar significados formais sobre as formas de visualizações de dados [39, 40, 41,42,1]. Tory e Möller [1] apresentam uma taxonomia que organiza as técnicas de visualização com base em modelos de dados e categoriza esses modelos de acordo com a natureza do domínio do dado (contínuo ou discreto) e em suas restrições, além de destacar o papel do usuário nos modelos conceituais. Duke *et al.* [41] destaca a necessidade de uma ontologia para visualização que suporta trabalhos colaborativos e educação. As ontologias têm como objetivo tornar as visualizações mais compreensíveis na Web, agregando semântica a essas representações de forma a melhorar a comunicação entre usuários e sistemas de visualização em todas as etapas do ciclo de comunicação, de forma que, tanto o usuário quanto o sistema compartilhem a mesma linguagem com significados comuns.

O projeto Visko [43] [44], que possui sua própria ontologia, a ontologia de visualização criada pelo UK National e-Science Center [41], a semiologia de gráficos de Bertin [19], a bibliografia de Few [23] e literatura pesquisada serviram de base para a nossa ontologia descrita de forma completa em [24].

Com base nas duas ontologias, vimos a necessidade de algumas classes que são comuns às duas: visualização, dado e transformação. As outras fontes foram importantes para o detalhamento das classes. A ontologia de Visko ainda possui uma classe atributos de exibição para definir as características de visualização e a de UK National e-Science Center incluiu o conceito de tarefa.

Definimos, então, em 5 as classes de nível superior:

1. Dado
2. Atributo de exibição
3. Visualização
4. Tarefa
5. Transformação

Baseado principalmente nos estudos de Tory *et al.*[1] e Ignatius e Senay. [29] e nas ontologias Scovo¹⁶ e Data Cube Vocabulary,¹⁷ construímos a ontologia da classe dado (Figura 14).

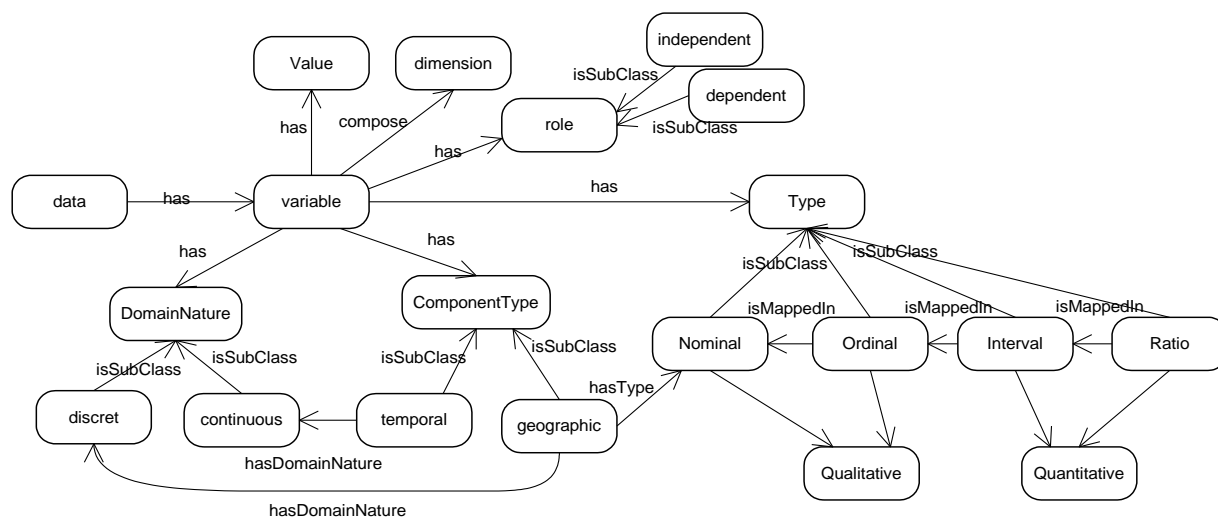


Figura 14: Classe dado

O número de dimensões de um dado é obtido pelo número de variáveis. Variáveis independentes são aquelas que são manipuladas, enquanto variáveis dependentes são apenas medidas ou registradas. As variáveis ainda devem ser classificadas quanto à sua estrutura. Estruturas de dados escalares possuem uma magnitude, mas nenhuma informação direcional além de um sinal. São, portanto definidos com um único número. Para visualização da informação, não iremos trabalhar com vetores e sensores, mas apenas com dados escalares, e por isso essa classificação não entrou no diagrama. Outro critério de classificação é quanto à natureza do dado. Discretos são dados que só se pode tomar um número finito ou infinito enumerável de valores distintos. Contínuos são os que se podem tomar todos os valores de um intervalo real. Junto à outra dimensão, o dado pode ser

¹⁶ <http://sw.joanneum.at/scovo/schema.html>

¹⁷ <http://publishing-statistical-ata.googlecode.com/svn/trunk/specs/src/main/html/cube.html>

classificado como intervalar, razão, nominal ou ordinal. Escalas de dados intervalares preservam a diferença real entre valores quantitativos (como os graus Fahrenheit), mas não têm um ponto natural zero. Escalas de razão de dados são como escalas intervalares, mas elas têm um zero natural e podem ser definidas em termos de unidades arbitrárias. Por exemplo, 200 dólares são o dobro de 100 dólares. Escalas nominais não têm uma ordem natural e ordinais são as que têm ordem natural, por exemplo: conjuntos como pequeno, médio e grande. Incluímos ainda uma classificação que chamamos de tipo do componente, seguindo a definição de Bertin [19] que pode ser temporal, geográfico ou nenhuma das duas subclasses.

A classe atributos de exibição apresentada na Figura 15 foi baseada nas propriedades das marcas e se relaciona com os atributos de dado através dos quatro tipos de percepção.

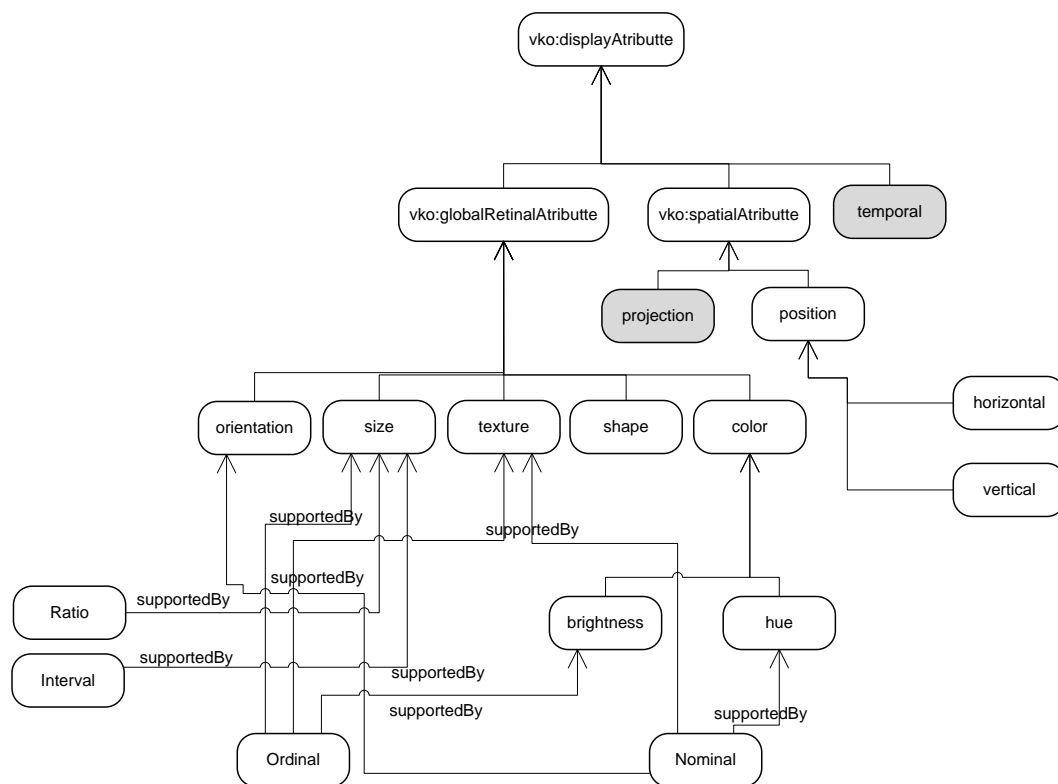


Figura 15: Classe atributos de exibição

Como as marcas ainda podem ser composta, ou seja, uma coleção de marcas simples formando uma unidade perceptual, associamos os dados nominais às classes associativas (matiz, orientação e textura), dados ordinais aos tipos

ordenados (tamanho, brilho e textura) e dados intervalares e de razão aos quantitativos (tamanho).

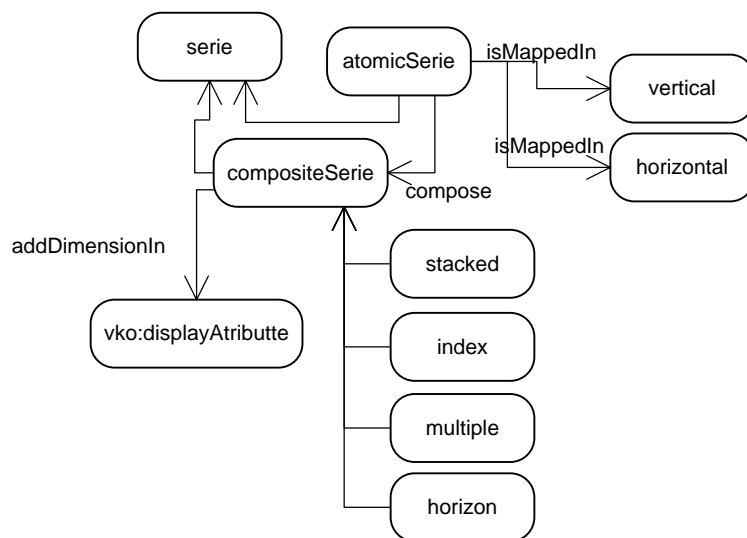


Figura 16: Exemplo de uma sub-classe de visualização

A Figura 16 exemplifica uma sub-classe da classe visualização, a classe dos gráficos de série. Uma série, assim como outras visualizações, pode ser atômica ou composta. Entre as visualizações compostas que formam uma série temos a série empilhada — *index* (Figura 17), *stacked* (Figura 18), múltiplas (Figura 19) e horizontais (Figura 20). Uma série simples possui apenas uma dimensão multivalorada que é mapeada nas posições verticais e horizontais. Duas ou mais dimensões multivaloradas compõem uma série composta. A adição de uma dimensão em uma visualização é feita através da inclusão de um atributo de exibição. Os gráficos de dispersão e os gráficos de colunas ou barras seguem o modelo apresentado no exemplo da Figura 16. A ontologia contemplou também gráficos de hierarquia e mapas que possuem características diferentes, porém não vamos entrar em detalhes sobre eles pois ainda não fazem parte do conjunto de visualizações da ferramenta que iremos apresentar. É importante lembrar que as visualizações em 3D não foram contempladas na ontologia, uma vez que, segundo Few [23], devem ser evitadas no uso de visualização de informação pois são ineficientes.

Para exemplificar o uso do atributo de exibição, apresentamos as figuras correspondentes às quatro sub-classes de séries compostas.

Gráficos de séries temporais mostram valores quantitativos em relação a pontos em seqüência no tempo [23]. Todos expressam o senso de continuidade que é requerido para expressar o tempo (dado contínuo). Na Figura 17 temos uma composição de múltiplas séries simultaneamente. Apesar de apenas dois eixos, temos uma terceira dimensão representada na cor, que é um atributo de exibição. Esta série é do tipo agrupada.

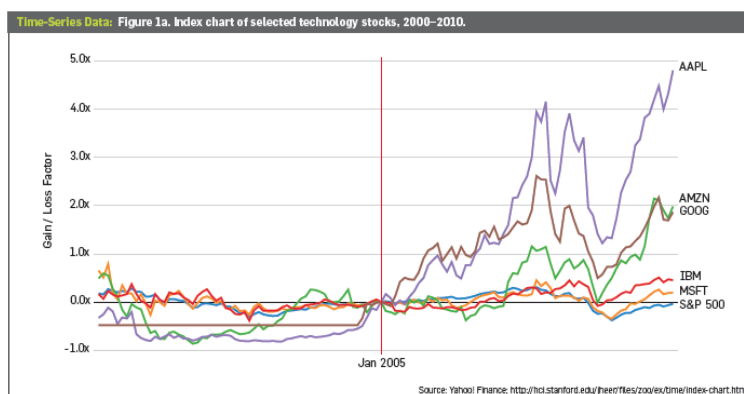


Figura 17: Séries do tipo *index* ou agrupadas

O gráfico da Figura 18 é um gráfico empilhado (*stacked*). Segundo Heer *et al.* [2], esse gráfico não suporta números negativos, mas o fato é que ele suporta somente números com mesmo sinal, positivos ou negativos. Como é fácil identificar uma unidade apenas com números positivos (exemplo: número de pessoas) mas não o contrário, esse gráfico foi limitado para valores positivos. Nele, também temos uma terceira dimensão representada pela cor.

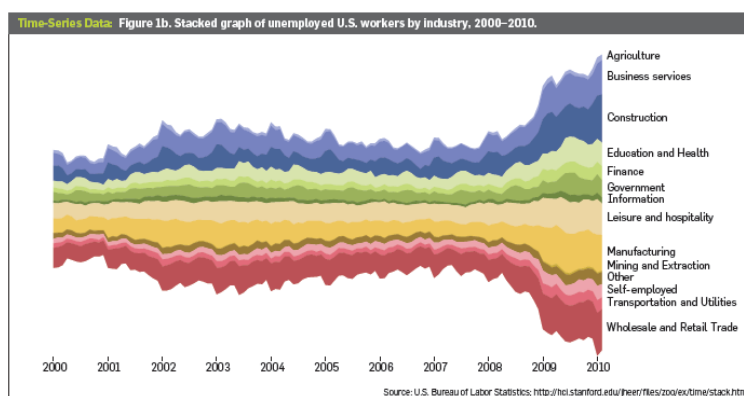


Figura 18: Séries empilhadas

A Figura 19 já é uma composição diferente. A cor não é a variável que define a terceira dimensão que é dada pela posição, criada a partir de novos eixos.

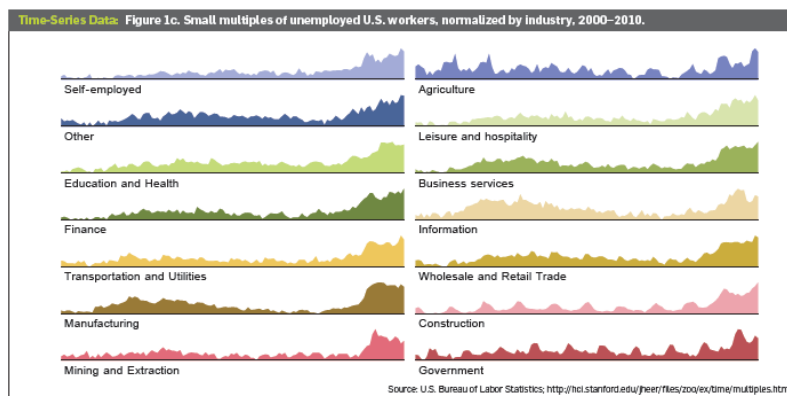


Figura 19: Séries múltiplas

O gráfico da Figura 20 mostra apenas duas dimensões e a cor está sendo utilizada para diferenciar números positivos e negativos, já que os valores são todos posicionados acima do eixo horizontal. O recurso de brilho também é utilizado para, em uma área de altura menor, exibir as formas sobrepostas.

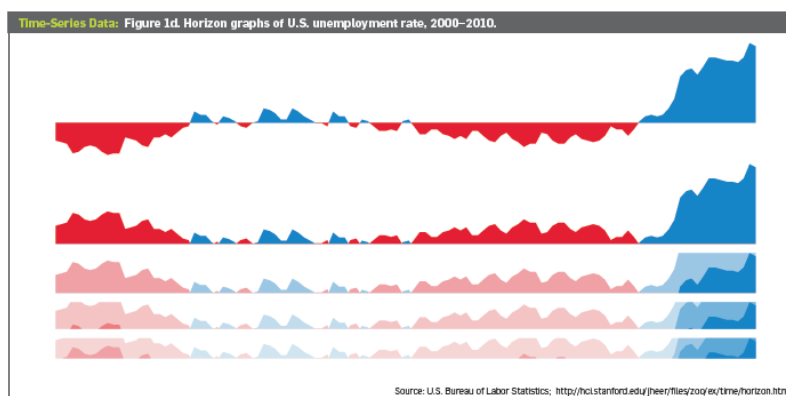


Figura 20: Série horizontal

A tarefa é o objetivo do usuário com a visualização dos dados e o que deve ser feito com a visualização. A tarefa tem relação com a pergunta que o usuário quer responder e com a eficiência da representação. Assim como Bertin [19], Steele and Lliinsky [45] também afirmam que a formulação da questão é parte importante para um processo de visualização eficiente. A Tabela 2 exemplifica alguns modelos de perguntas e tarefas definidas nessa ontologia. A limitação dos números da tabela leva em consideração os limites do tipo de percepção visual, definida na classe atributo de exibição.

Tabela 2: Exemplos de perguntas e sua relação com a visualização e a tarefa¹⁸

Pergunta	Visualização	Tarefa
Existe alguma relação entre os valores v1 e v2 da dimensão D?	Gráfico de dispersão	Identificar correlações entre um número ilimitado N de pares de variáveis quantitativas
Qual é a categoria com maior ou menor valor de determinada variável?	Gráfico de colunas agrupadas	Obter ranking entre um número limitado M de variáveis quantitativas de um número ilimitado N de categorias
Qual é o valor da variável v1 da classe m1 no instante t?	Séries agrupadas	Visualizar tendências, mudanças, aumento, flutuação, crescimento, declínio de uma dimensão D quantitativa de um número limitado de M classes

Finalmente as funções da classe transformação é que definem que dimensões podem ir para cada atributo de exibição de modo que a visualização seja eficiente para a tarefa ou pergunta.

A propriedade *ismappedOn* define que os dados quantitativos podem ser mapeados em qualitativos. Essa transformação irá acontecer sempre que for necessário exibir tais dados de forma categórica. Uma transformação que pode ser feita é a divisão do conjunto de categorias em partes iguais o que é denominado *quartis*. Outros dados também devem ser mapeados nos atributos de exibição através de prioridades. Por exemplo: Apenas barras horizontais devem ser usadas para exibir valores temporais. Gráficos de séries temporais sempre usam o eixo horizontal para a escala de tempo e o vertical para escalas quantitativas [23].

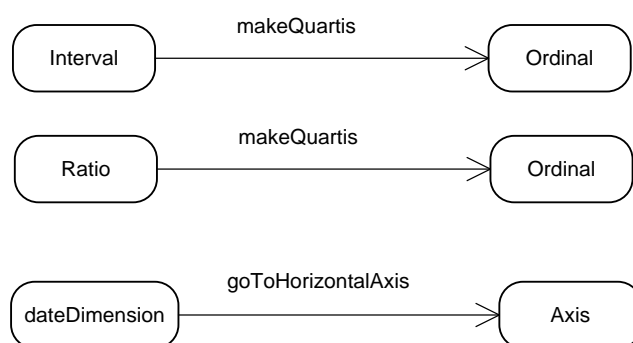


Figura 21: Exemplo de funções da classe transformação

¹⁸ A tabela completa pode ser consultada no apêndice C desta pesquisa

3.2.1. Discussão

A dificuldade em criar ontologias para visualização está ligada a dois principais motivos. Primeiramente ao fato da sua compreensão depender de fatores cognitivos e, em segundo porque existem inúmeras possibilidades de representar um dado com as oito variáveis que os gráficos oferecem (as duas dimensões planares, cor, brilho, textura, orientação, tamanho e forma).

Pesquisamos aqui os componentes que devem ser levados em consideração no estudo para a criação de uma ontologia de visualização, através de um número limitado de representações gráficas. Uma ontologia provê um vocabulário pelo qual usuários e sistemas podem se comunicar, mas só é útil se refletir o consenso de uma comunidade. Por isso, essa ontologia deveria ser testada com novas representações a fim de verificar seu grau de interoperabilidade. A visão é que uma ontologia deve dizer o que é um mapeamento inválido, mas como vimos, a questão é cognitiva.

Para mais detalhes sobre a pesquisa e para visualização o diagrama completo da ontologia, pode-se consultar a pesquisa completa [24]. É importante lembrar que, na ferramenta ViSC, as tarefas apresentadas aqui foram substituídas por tarefas de mais baixo nível estudadas em [26] para melhor formularmos as perguntas. Muitas outras transformações também foram necessárias para possibilitar as transformações dos dados nas visualizações contidas na ferramenta.

4 A ferramenta ViSC

A ferramenta ViSC foi concebida a partir dos resultados obtidos nos trabalhos anteriores. Concluímos que era necessária uma ferramenta de visualização que apoiasse os usuários inexperientes na construção de gráficos através da pergunta que ele queria responder. Este capítulo descreve os requisitos, a ferramenta (gramática que gerou o conhecimento do sistema, funcionamento do sistema de recomendação, interface e interação) e faz uma comparação com outras ferramentas relacionadas.

4.1. Requisitos

Com base nas pesquisas descritas chegamos a conclusões sobre os requisitos que a ferramenta deveria conter na tentativa de responder nossa questão de pesquisa: *Como apoiar usuários inexperientes na construção de visualizações eficientes com dados estatísticos?*

Com base nesses estudos, listamos alguns requisitos para a criação de uma ferramenta de visualização para construção de gráficos por usuários inexperientes:

1. Gerar gráficos eficientes, claros e precisos.

A eficiência da representação é definida por Bertin [19] pela seguinte proposição: “*se, para tentar obter um uma resposta correta e completa para uma questão dada, sendo todas outras coisas iguais, uma construção requer um tempo de observação mais curto que outra construção, nós podemos dizer que é mais eficiente para a questão*” [19, p.139]. Bertin completa ainda que “*as construções mais eficientes são aquelas nas quais quaisquer questões, não importando tipo ou nível, podem ser respondidas em um simples instante da percepção, ou seja, em uma só imagem*” [19, p. 146].

Segundo Tufte [7], a excelência nos gráficos estatísticos consiste na comunicação de ideias complexas com clareza, precisão e eficiência. Segundo o mesmo autor, para isso, os gráficos devem: mostrar os dados;

induzir o espectador a pensar sobre a essência mais do que sobre a metodologia, design gráfico, a tecnologia de produção gráfica, ou qualquer outra coisa; evitar a distorção dos dados; apresentar muitos números em um espaço pequeno; fazer grande conjuntos de dados coerentes; encorajar o olho a comparar diferentes pedaços de dados; revelar o dado em vários níveis de detalhes, desde uma visão ampla até uma estrutura refinada; servir para uma finalidade clara (descrição, exploração, tabulação ou decoração); estar estreitamente integrada com descrições estatísticas e verbais do conjunto de dados.

2. Motivar usuários inexperientes a analisar dados graficamente

Conforme já discutimos, o sistema gráfico exige conhecimento matemático e percepção que envolve a capacidade de abstração. Assim, desenvolver uma ferramenta que motive a análise dos dados poderia ajudar no aprendizado do sistema gráfico.

3. Permitir vários tipos de construção e operações

Ao analisar como as ferramentas de visualização influenciavam na interpretação de gráficos [17], verificamos que as ferramentas com maior variedade de gráficos e funcionalidades eram muito complexas e voltadas para usuários mais experientes. Quando utilizadas por usuários inexperientes geraram rupturas graves de comunicação entre usuário e sistema. No entanto, ferramentas mais simples possuíam limitações que não permitiam que o usuário construísse visualizações desejadas. Assim, precisávamos de uma solução que permitisse mais possibilidades de construção para usuários novatos.

4. Usar ontologia para elaborar significados precisos da visualização

Pesquisas na literatura nos levaram a compreender que a elaboração da pergunta poderia ser uma alternativa melhorar a comunicação entre usuários e sistemas de visualização. Para tornar as visualizações mais compreensíveis na Web, o usuário e o sistema devem compartilhar da mesma linguagem. Para isso nossa segunda pesquisa buscou agregar semântica às visualizações. Assim, investigamos taxonomias e ontologias de visualização e desenvolvemos um estudo com objetivo de criar uma ontologia para definir relações entre dados e modelos de

visualização a fim de responder perguntas de usuários através de visualizações eficientes [24].

5. Prover *feedback* visual, visualizações automáticas e valores *default*.

Pesquisas na literatura demonstram que esses três recursos são necessários em um processo de criação fácil, o que é essencial para usuários inexperientes. Além disso, as alterações entre cada uma das visualizações deveriam ser igualmente simples para evitar a escolha de um gráfico equivocado pelo comportamento oportunista do usuário, o que foi observado em [17].

6. Possuir um recurso de ajuda interativo.

Recomendações são muito utilizadas em sites de comércio eletrônico. A possibilidade de recomendação de visualizações foi pensada como uma solução para ajudar o usuário a criar visualizações mais eficientes de acordo com o dado selecionado e com a pergunta que ele queria responder. Assim os conceitos de sistema de recomendação ajudaram a definir como utilizar recomendações em um sistema de visualização.

Além dos requisitos listados, buscamos construir uma interface para solucionar problemas de IHC que geraram rupturas leves na avaliação com as ferramentas pesquisadas em [17].

Buscando satisfazer esses requisitos, construímos a ferramenta ViSC, uma ferramenta de visualização para construção gráficos com dados estatísticos que ajuda usuários inexperientes a explorar visualizações através de recomendações de visualizações relacionadas baseadas em perguntas comuns de busca de informação (*common information seeking questions*).

O Sistema ViSC foi desenvolvido para ser uma aplicação Web e escolhemos portanto a linguagem PHP. Como o maior objetivo era fornecer um sistema interativo, fizemos uso também de jQuery e AJAX. Para a construção dos gráficos, algumas bibliotecas foram pesquisadas, porém, pela ampla variedade de representações e métodos de interação, optamos pela *API Google Chart Tools*. Como no SIDRA, ViSC suporta dados estatísticos que possuem até 6 dimensões.

4.2. Metamensagem

Com base na Engenharia semiótica, a metamensagem é a mensagem do designer para o usuário sobre como ele pode e deve utilizar o sistema, por que e com que efeitos [25].

Para elaborar essa metamensagem, o especialista de Engenharia Semiótica deve responder as seguintes questões:

1. Quem eu entendo que são os usuários do produto do meu *design*?
2. O que eu aprendi sobre o que esses usuários querem ou precisam?
3. Quais são as preferências dos usuários em relação ao que eles querem e precisam e por quê?
4. Portanto, que sistema eu projetei para esses usuários e como eles podem ou devem usá-lo?
5. Qual é a minha visão de *design*?

Com essas respostas, a paráfrase da metamensagem é então usada como modelo a ser preenchido:

*Este é o meu entendimento, como designer, de **quem você, usuário, é**, do que aprendi sobre você **quer ou precisa fazer**, de que maneiras prefere fazer, e por quê. Este, portanto, é o sistema que projetei para você, e esta é **a forma como você pode ou deve utilizá-lo** para alcançar uma gama de objetivos que se encaixam nessa visão.*

Com base nos requisitos listados anteriormente, podemos construir a metamensagem da comunicação de ViSC da seguinte forma:

Penso que seja um estudante ou um profissional de áreas ligadas ou não a estatística que precisa construir gráficos eficientes. Como você pode não ter domínio sobre as visualizações, você prefere uma interface que, com pouca interação, gere um gráfico e partir dele você consiga facilmente alterar os valores e visualizar outras opções de gráfico. Você também gostaria de poder comparar visualizações. Assim, projetamos um sistema que, a partir da seleção do tema e duas variáveis, apresenta um gráfico em que os valores possuem um valor selecionado por padrão automaticamente. Então, basta incluir ou alterar os valores e selecionar a forma de visualização. Para lhe ajudar a decidir que visualização utilizar, a interface recomenda visualizações com base na pergunta que você deseja responder. Você pode selecionar a visualização através das

opções do menu ou das perguntas. Neste caso, você só precisa encontrar a pergunta e escolher uma ou mais entre as visualizações recomendadas. Para lhe ajudar, a interface explica como obter a resposta e a apresenta no gráfico. As recomendações ainda foram classificadas com uma nota para ajudar na sua escolha. Você também pode alternar os eixos, exibir a variação, mudar escalas, eliminar o zero do gráfico ou reordenar para tornar sua visualização mais eficiente.

4.3. A gramática de ViSC

Para cumprir o requisito 4, a ontologia desenvolvida e apresentada na seção 3.2 foi utilizada em ViSC. As triplas da ontologia foram modeladas em banco de dados relacional e algumas alterações foram realizadas.

Para compor a classe de visualização, entre as visualizações estudadas, selecionamos oito tipos para usar em ViSC:

1. colunas agregadas (*clustered column chart*)
2. colunas múltiplas (*multiple columns*)
3. colunas empilhadas (*stacked columns*)
4. linhas para representar séries (*series chart*)
5. séries múltiplas (*multiple series*)
6. séries empilhadas (*stacked series*) através do gráfico de área
7. gráfico de dispersão (*scatterplot*)
8. tabela (*table*).

Todas elas são apresentadas na Figura 22 e são compostas de até 3 dimensões: uma dependente no eixo vertical e duas independentes, que em (1), (3), (4) e (6) que são representadas no eixo horizontal e na cor (ou na legenda). Já em (2) e (5), as duas independentes são representadas em dois eixos horizontais (ou nas legendas). Em (7) temos duas dependentes representadas pelos eixos e uma independente, pela cor (ou na legenda).



Figura 22: Exemplo dos 8 tipos de visualização disponíveis em ViSC

Para definir melhor as perguntas, substituímos as tarefas da ontologia desenvolvida anteriormente pelas tarefas de mais baixo nível propostas pela taxonomia de Amar *et al.* [26]. Essa taxonomia apresenta um conjunto de dez tarefas que são relacionadas a questões específicas que uma pessoa pode perguntar enquanto trabalha em um conjunto de dados. As dez tarefas são: retornar valor, filtrar, calcular valor derivado, achar extremo, ordenar, determinar intervalo, caracterizar distribuição, achar anomalias, agrupar e correlacionar.

Assim, a gramática de ViSC se baseia na ontologia anteriormente desenvolvida [17] (e seção 3.2), no *toolkit* de visualização *Google Chart Tools* e na taxonomia de Amar *et al.* [26]. Os dados são mapeados dentro das características apresentadas na pesquisa [24], utilizando as classes atributo de exibição e visualização do *toolkit*. As novas tarefas e questões foram mapeadas dentro das oito visualizações exemplificadas na seção anterior.

4.4.O sistema de recomendação de ViSC

Buscando criar uma forma de ajuda para o usuário interagir com a ferramenta (requisito 6), usamos uma técnica de recomendação. ViSC usa um sistema de recomendação baseado em conhecimento (*knowledge-based*

recommender system), segundo a definição de Burke [37]. Esse tipo de sistemas de recomendação tem conhecimento sobre como um item em particular atende a uma necessidade do usuário em particular, e pode, portanto, concluir sobre a relação entre uma necessidade e uma possível recomendação. Em outras palavras, não depende muito do perfil do usuário, mas sim da tarefa que ele deseja realizar.

Em ViSC, a partir da seleção de um dado pelo usuário, **o sistema não só apresenta questões que podem ser respondidas pela visualização exibida no momento, como também recomenda outras visualizações eficientes para um diferente conjunto de questões relacionadas ao dado selecionado.** O conhecimento que sustenta essas relações foi definido pela composição de gramáticas formais descritas na seção anterior.

As questões visam a habilitar um diálogo entre sistema e usuário em uma linguagem próxima a natural de modo que o usuário percorra outras visualizações relacionadas até encontrar a resposta que ele procura. Para cada interação, as questões exibidas podem mudar ou se tornarem mais ou menos destacadas na interface do usuário, a fim de refletir sua relevância para o dado selecionado. Assim, o sistema guia o usuário na procura por uma visualização eficiente para responder sua questão. As questões são dinamicamente geradas baseadas em *templates* armazenados em um banco de dados. O dado selecionado e suas características são parametrizados para criar as questões. Abaixo apresentamos dois exemplos de *templates* de questões:

- *How many <persons> with <10 years of study> are in <Rio de Janeiro>?*
(tarefa: Retornar valor)
- *What is the average of <grade point average> in <disciplines> of the selected set?* (tarefa: Calcular valor derivado)

As questões também são classificadas com pontuação entre um e cinco, de acordo com a condição de seleção. Além das características semânticas do dado, tais como o tipo do componente (temporal, geográfico ou nulo), tipo de domínio (discreto ou contínuo) entre outros que caracterizam as pré-condições (Figura 23), as recomendações das questões também precisam de uma condição de seleção, a qual inclui o número de elementos selecionados de cada dimensão, como no exemplo apresentado na Figura 24, que apresenta as relações de uma tarefa, suas questões, condições e gráficos. Conforme descrito na seção anterior, nosso

processo semântico seguiu a literatura de visualização de dados baseada em tarefas [26], na semiologia de Bertin [19] e nos trabalhos de Few [23].

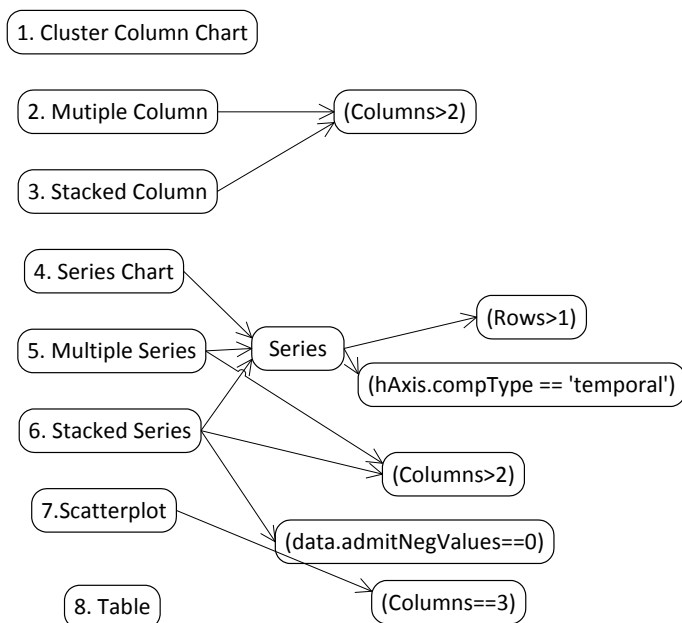


Figura 23: Pré-condições que definem quais visualizações podem ser utilizadas de acordo com características do dado

A Figura 23 ilustra as triplas que definem quando uma visualização pode estar ativa. Verifica-se que o gráfico de colunas agregadas (*clustered column chart*) e a tabela (*table*) estão sempre ativos. **Assim, optamos por tornar o gráfico de colunas agregadas o gráfico *default* do sistema, uma vez que ele atende a qualquer condição.** No entanto, colunas múltiplas (*multiple columns*) e empilhadas (*stacked column*) dependem da seleção de dois elementos da dimensão selecionada na legenda (ou coluna). Para montar uma série, é preciso mais de um elemento no eixo horizontal (ou linha) e ter selecionado para este eixo um componente temporal (exemplo: horas, dias, meses ou anos). Assim como nos gráficos de colunas, as séries múltiplas (*multiple series*) e as empilhadas (*stacked series*) também dependem da seleção de dois ou mais elementos na legenda (ou coluna). Este último também não permite números negativos, conforme descrito na ontologia. O gráfico de dispersão (*scatterplot*) deve necessariamente conter dois valores selecionados. Na Figura 23, o gráfico de dispersão mostra que o

número de colunas deve ser igual a 3 porque, quando tratados como tabela, a primeira coluna corresponde aos elementos da linha.

Entre as dez tarefas listadas, no cenário avaliado utilizamos apenas calcular o valor derivado, cujo modelo está ilustrado pela Figura 24. Os modelos de todas as tarefas estão representados no Apêndice C.

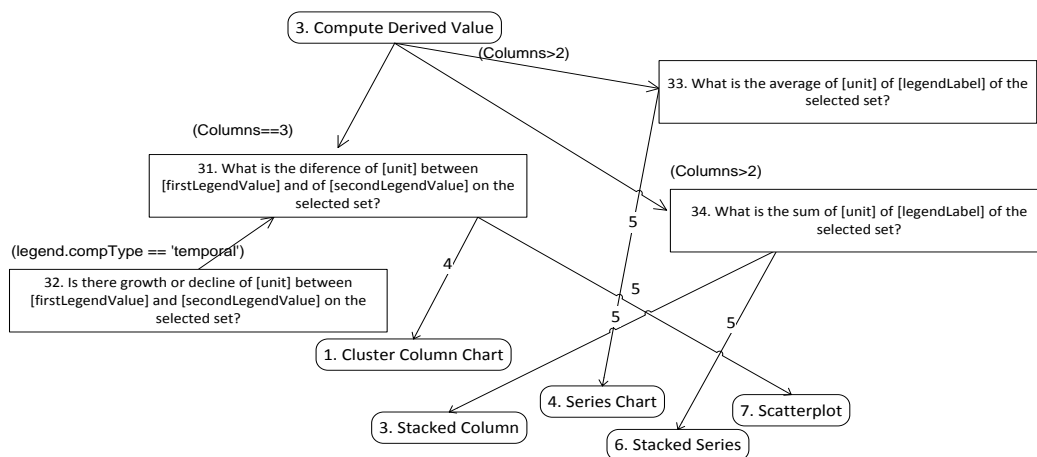


Figura 24: Condições para a tarefa calcular valor derivado.

A Figura 24 apresenta um exemplo das relações da tarefa calcular valor derivado, suas questões, condições e visualizações. Neste caso as somas retornam nos gráficos empilhados (*stacked*). Se é um dado já validado pela pré-condição como temporal, exibe os dois; caso contrário, exibe apenas o de colunas agrupadas, onde o componente temporal será transformado em discreto. Observa-se que, para a pergunta 31, que retorna a diferença entre dois valores quantitativos, a condição de seleção é conter 3 colunas entre os dados selecionados. Para esta pergunta, a visualização com melhor classificação é o gráfico de dispersão (*scatterplot*), classificado com cinco pontos. Da mesma forma, observamos as condições e classificações das outras perguntas apresentadas na figura. A pergunta de diferença é ainda uma generalização da pergunta sobre crescimento ou declínio. Isto significa que, se na pré-condição tivermos um gráfico com componente temporal, o texto exibido trocará a palavra diferença por crescimento ou declínio.

4.5. A interface de ViSC

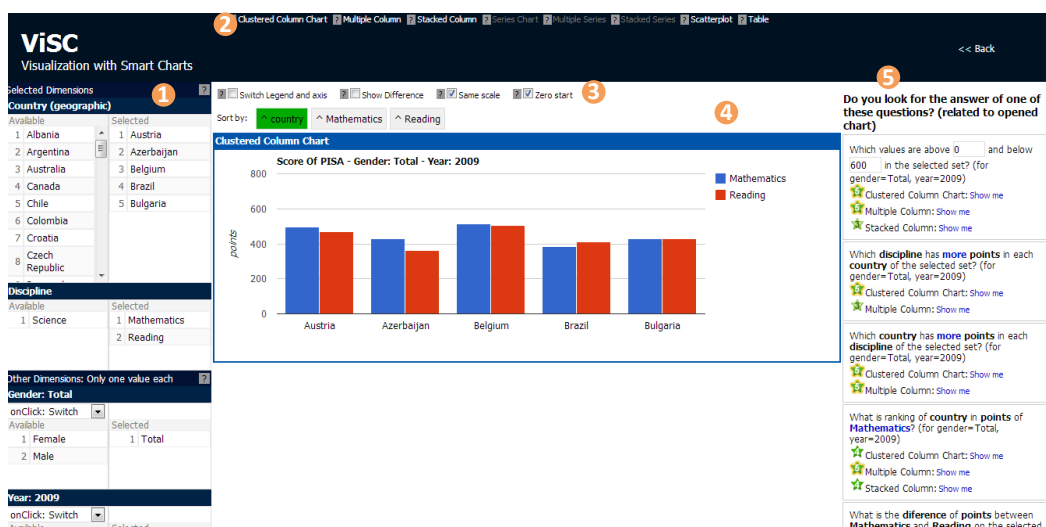


Figura 25: Interface de ViSC

A Figura 25 apresenta a tela de construção do gráfico. A área 1 corresponde à opção de menu de cada dimensão. A área 2 exibe o menu das visualizações, no qual os itens são habilitados ou desabilitados conforme as características semânticas dos dados selecionados (pré-condições). A área 3 apresenta botões para ordenar o dado no gráfico exibido. A área 4 é a área principal do gráfico e a área 5 exibe as questões relacionadas que vinculam para diferentes gráficos.

ViSC foi desenhada principalmente após a observação das rupturas de comunicação ocorridas nas ferramentas avaliadas [17] (e seção 3.1) e no artigo de Heer e Shneiderman [33].

Para a criação de gráficos eficientes, claros e precisos (requisito 1), aplicamos algumas das técnicas estudadas nas fontes pesquisadas. Comparado ao Statplanet, a área é maior e a proporção entre os lados do retângulo também busca uma melhor separação angular [19, p. 175]. Além disso, o uso da escala iniciando em zero por *default* em todos os gráficos, o uso da mesma escala em gráficos múltiplos, a presença de descrição textual em todos os componentes (eixos e legenda) e a alternância entre dados discretos e contínuos de acordo com o que é mais adequado visavam a aumentar a legibilidade e a facilitar a compreensão.

Para motivar uma melhor análise por parte do usuário (requisito 2), na área 4 o sistema permite visualizar diversos gráficos simultaneamente. Através de

funções de expandir, colapsar ou fechar, o usuário pode manter abertos apenas os gráficos que deseja comparar e facilmente trocar as visualizações exibidas.

Para cumprir o requisito 3, de criar mais opções de visualizações em sistemas para usuários inexperientes, selecionamos os oito tipos já descritos anteriormente (Figura 23), o que permite que o usuário selecione qualquer classe do dado selecionado para qualquer um dos dois atributos de exibição (eixo horizontal ou legenda). Comparando com Statplanet, que é limitado em cinco visualizações e previamente seleciona as classes do dado para os atributos de visualização sem permitir alteração do usuário, ViSC incluiu novas possibilidades.

Na avaliação com as ferramentas de visualização [17] observamos também que, no Statplanet, usuários não utilizaram a seleção pelo mapa (preferencial do designer) ou quando a utilizaram apresentaram rupturas de comunicação. Como os países eram marcados em amarelo na lista, como mostra a Figura 26, além de confundir com itens selecionados como itens onde o *mouse* estava posicionado, o usuário não tinha uma visão geral da sua seleção. Assim, em ViSC, optamos pela criação de duas colunas, uma com os itens disponíveis e outra com os itens selecionados e numerados, como mostra a Figura 27, o que fornece um melhor *feedback* visual para o usuário (requisito 5).

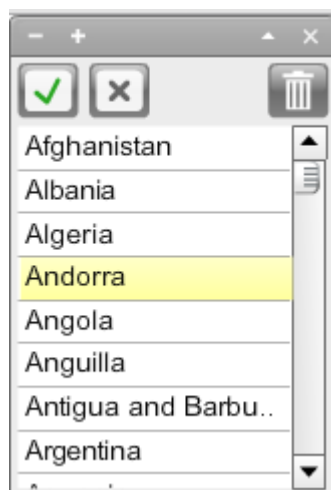


Figura 26: Menu de países em Statplanet

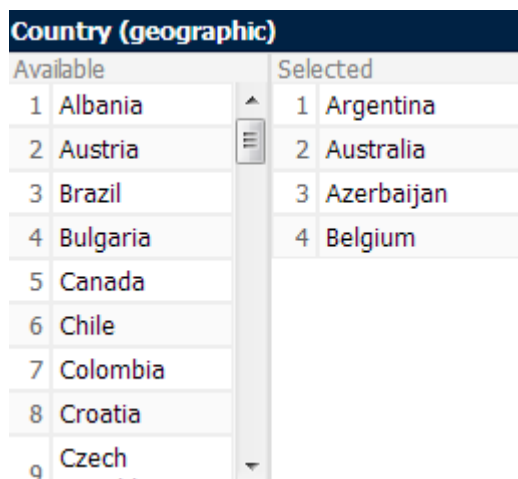


Figura 27: Menu de seleção de elementos de uma dimensão em ViSC

4.6. A interação com ViSC

Ainda com o objetivo de tornar o processo de construção de gráficos simples (requisito 5), em ViSC, o usuário deve selecionar primeiramente apenas o tema e duas dimensões: uma para o eixo horizontal e outra para a legenda nos campos de seleção mostrados na Figura 28. Então, o sistema cria um gráfico vazio (Figura 29) automaticamente. Nesse momento, não existem questões para serem selecionadas, apenas uma instrução, presente na barra à direita informando que o usuário deve selecionar valores para os eixos, isto é, o dado. No entanto, com exceção da primeira dimensão, que o usuário deve selecionar os dados, para as outras dimensões o sistema atribui um valor *default* fixado.



Figura 28: Interface inicial

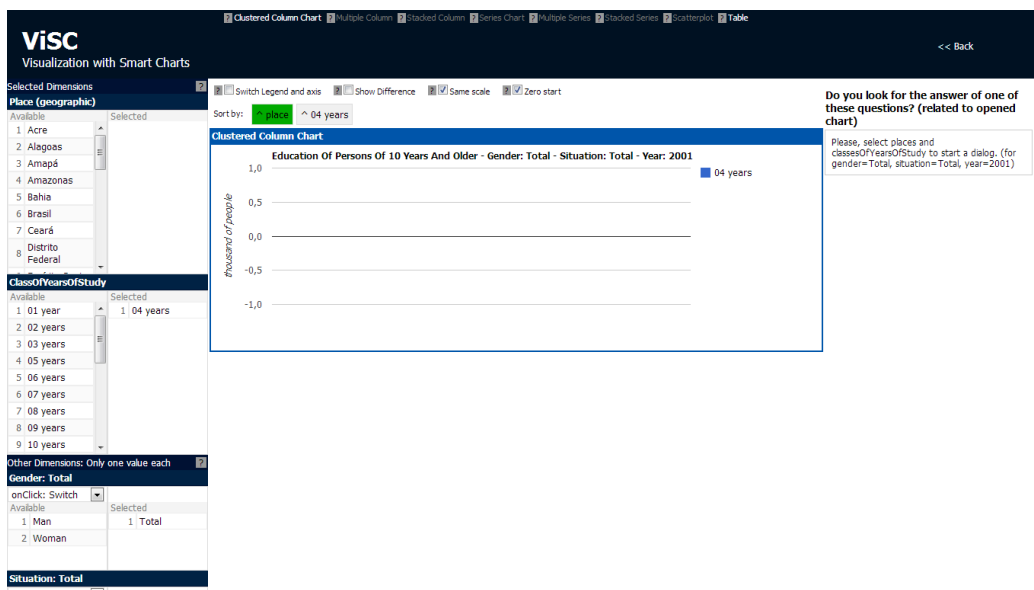


Figura 29: Interface com gráfico vazio

Após selecionar as variáveis, a tela apresentada se parece com a Figura 25. As questões são mostradas e podem mudar depois de cada interação entre usuário e sistema. A Figura 30 mostra em detalhe as questões apresentadas na Figura 25. No exemplo, temos quatro valores selecionados no eixo horizontal (ou nas linhas: "Alagoas", "Amazonas", "Ceará" e "Goiás") e dois valores na legenda (ou nas colunas: "02 anos" e "03 anos"). A Figura 25 mostra o gráfico de colunas agrupadas e todas as questões relacionadas ao gráfico em ordem decrescente de peso da recomendação. Essas questões podem ser melhor visualizadas na Figura 30. Essas mesmas questões também possuem links para outras visualizações que elas também respondem. O gráfico aberto é sempre mostrado como primeira opção.

As estrelas numeradas ao lado do nome das visualizações indicam o peso, isto é, a relevância do gráfico selecionado para responder a questão, conforme o modelo apresentado. Incluímos ainda questões adicionais relacionadas ao dado que o gráfico aberto não permite responder. Essas questões são separadas por outro título, como mostrado na Figura 31. Note que, como temos dois elementos selecionados em uma das dimensões, questões relacionadas às tarefas de anomalia e correlação, que são melhor visualizadas com o gráfico de dispersão, também são exibidas para o dado selecionado na área de perguntas relacionadas a outros gráficos.

Do you look for the answer of one of these questions? (related to opened chart)

Which values are above 0 and below 700 in the selected set? (for gender=Total, situation=Total, year=2001)

- ★ Clusted Column Chart: [Show me](#)
- ★ Multiple Column: [Show me](#)
- ★ Stacked Column: [Show me](#)

Which **classOfYearsOfStudy** has **more thousand of people** in each **place** of the selected set? (for gender=Total, situation=Total, year=2001)

- ★ Clusted Column Chart: [Show me](#)
- ★ Multiple Column: [Show me](#)

Which **place** has **more thousand of people** in each **classOfYearsOfStudy** of the selected set? (for gender=Total, situation=Total, year=2001)

- ★ Clusted Column Chart: [Show me](#)
- ★ Multiple Column: [Show me](#)

What is ranking of **place** in **thousand of people of 03 years**? (for gender=Total, situation=Total, year=2001)

- ★ Clusted Column Chart: [Show me](#)
- ★ Multiple Column: [Show me](#)
- ★ Stacked Column: [Show me](#)

Figura 30: Detalhe de questões relacionadas ao gráfico aberto.

Or you may look for one of these? (related to closed charts)

Is there any **anomaly** between **thousand of people of 01 year and 03 years**? (for gender=Total, situation=Total, year=2001)

★ Q-Q plots: [Show me](#)

Is there any **correlation** between **thousand of people of 01 year and of 03 years**? (for gender=Total, situation=Total, year=2001)

★ Q-Q plots: [Show me](#)

What is the **sum** of **classesOfYearsOfStudy** of **each place** on the selected set? (for gender=Total, situation=Total, year=2001)

★ Stacked Column: [Show me](#)

Figura 31: Detalhe de questões não relacionadas ao gráfico aberto mas sim ao dado selecionado

4.7. Comparação entre ViSC e outras ferramentas relacionadas

Tableau (Figura 32) e Explorations Views (Figura 33) [13] são ferramentas de visualização recentes que foram desenvolvidas para serem utilizadas tanto por usuários inexperientes quanto experientes. Para isso, ambas possuem recursos que facilitam a interação do usuário menos experiente. O Tableau possui uma paleta denominada Show me [16], que apresenta um conjunto de comandos integrados que adicionam apresentações automáticas à ferramenta para criar visualizações eficientes baseada na semiologia dos gráficos de Bertin [19] e na álgebra usada por APT [46]. O painel *Show me* (Figura 32.7) consiste em um diálogo de escolhas com *tooltips* que descrevem condições para uma opção estar ativa.

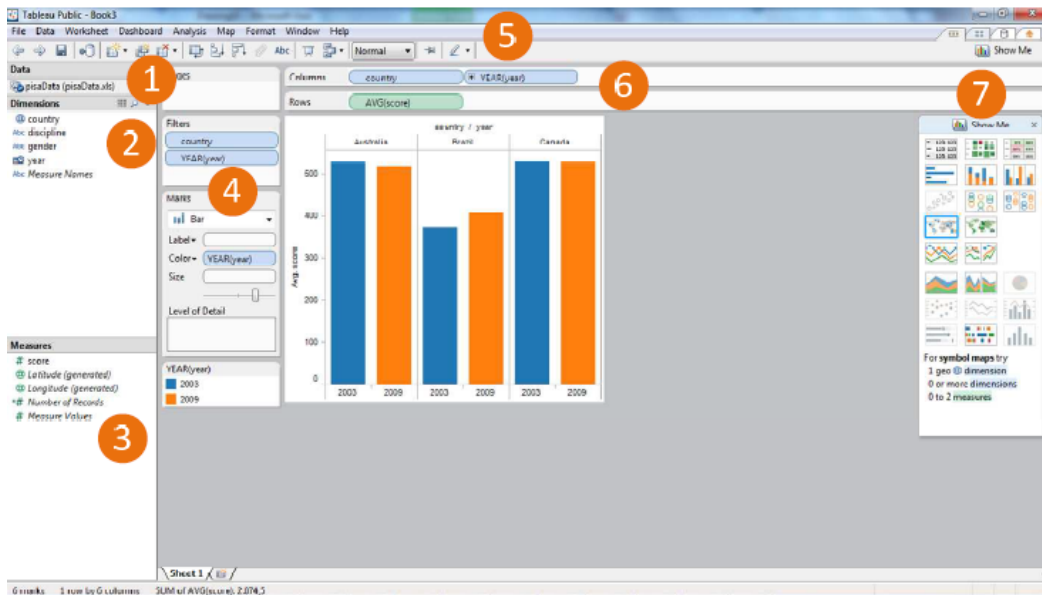


Figura 32: Interface do Tableau. (1) dados selecionados ou importados. (2) lista de dimensões disponíveis. (3) lista de medidas. (4) painéis com filtro, marcas e legenda. (5) menu e barra de ferramentas. (6) colunas e linhas onde dimensões e medidas são selecionadas (7) painel *Show me*.



Figura 33: Exploration Views consiste em: (a) o *Dashboard* principal e seus componentes; (b) o *Clipboard drawer widget* que inclui *templates* para gráficos (parte superior) e dados de exemplos de *widget* de dados (parte inferior) que o usuário tenha criado como uma *palette*; (c) uma representação visual dos filtros de dados globais; e (d) um conjunto de outras funções, como buscar, salvar, compartilhar, etc.

Exploration Views (EV) é um protótipo que também sugere gráficos e *templates* para criar *dashboards* por usuários inexperientes. Sua proposta começou por uma pesquisa sobre o que era necessário para usuários novatos em análises visuais. Desse modo, EV foi desenvolvida seguindo alguns requisitos: um assistente (*wizard*) simples para criar gráficos, *feedback* visual contínuo e ainda fornece gráficos e *templates* de sugestões baseados em tarefas (comparação, tendência e contribuição).

Assim como em EV, em ViSC definimos alguns valores *default* e *feedback* visual para gerar visualizações automáticas como parte de um processo de busca de informação. No entanto, o sistema de recomendação visa não apenas a possibilitar visualizações, mas também a estabelecer um diálogo com o usuário através da questão que ele quer responder. Esta é a principal diferença e a principal contribuição de ViSC em relação a outras ferramentas. Fazendo isso, pretendemos promover o interesse do usuário na busca de outras representações com baixo custo de experimentação.

Assim, avaliamos ViSC para verificar as seguintes contribuições em visualização de dados:

i) solucionar problemas de construção e interpretação de gráficos por usuários inexperientes com uma ferramenta que permite vários tipos de construção e com operadores avançados através de recomendações.

Indiretamente pretendíamos também:

ii) avaliar a ontologia de visualização utilizada, que define as relações entre dado, modelos de visualização e questões do usuário;

iii) motivar usuários inexperientes a analisar dados graficamente.

5 Avaliação

A avaliação com os usuários visava responder as sub-questão de pesquisa deste trabalho, entre as quais a principal delas é:

SQ3. Como as perguntas influenciam na realização da tarefa e nas visualizações geradas?

5.1.Métodos

Para responder essa sub-questão, então, utilizamos a triangulação de dois métodos qualitativos: Método de Inspeção Semiótica (MIS) [25] e um método que chamamos de "*Retrospective Communicability Evaluation (RCE)*", criado a partir de uma observação com usuários, *Retrospective Think Aloud (RTA)* [27] e posterior etiquetagem do Método de Avaliação de Comunicabilidade (MAC) [20]. A escolha do MIS visava avaliar falhas na emissão da mensagem pelo designer em relação às perguntas contidas na interface, revelando características dos signos que poderiam não ser bem transmitidos na metacomunicação. A aplicação do RCE tinha como objetivo revelar sobre os processos abduativos na experiência dos usuários durante a leitura e construção de gráficos, buscando uma resposta para uma pergunta específica e comparar a metamensagem recebida com a emitida.

O MIS é um método proposto pela Engenharia Semiótica [25] e examina a grande diversidade de signos pelos quais os usuários podem interagir com artefatos computacionais. A mensagem que o designer deseja passar para os usuários é expressada através desses signos de um ou mais sistemas de significação (seção 2.2). Após a inspeção dos signos metalinguísticos (presentes na documentação da ferramenta e em mensagens em linguagem natural), estáticos (compostos por imagens, ícones, cores, etc) e dinâmicos (animações e comportamentos gerados a partir de eventos), compara-se as metamensagens do designer para o usuário geradas em cada inspeção nos passos anteriores e por fim, analisa-se a qualidade da metacomunicação geral.

O MAC e *Think Aloud* ou *Concurrent Think Aloud* (CTA) com co-participação se apresentaram eficazes em conjunto na investigação anterior feita com ferramentas visualizações da Web [17] e assim decidimos utilizar características dos dois métodos para compor o novo método — "*Retrospective Communicability Evaluation*" (RCE). No entanto, o CTA foi substituído pelo RTA.

Para entender o uso do RTA em vez do CTA como parte do método, buscamos as vantagens e desvantagens desses métodos. No protocolo CTA, a parte principal da instrução é geralmente muito curta, fazendo referência a um procedimento que se presume já ser familiar dos usuários. Por exemplo: "*Tente pensar em voz alta, você deve fazer isso frequentemente quando você está sozinho pensando num problema*" [21]. Segundo Haak *et al.*, [27] o RTA difere desse método porque em vez de pensar em voz alta enquanto trabalham, os usuários inicialmente realizam suas tarefas de trabalho em silêncio e só verbalizam seus pensamentos baseado no vídeo gravado com a sua tarefa. Teoricamente, existem vantagens e desvantagens de usar RTA em vez do CTA. Uma vantagem do RTA envolve uma possível diminuição na reatividade: os usuários são totalmente motivados a executar uma tarefa em seu ritmo e, portanto, não terão um desempenho melhor ou pior do que o habitual. No CTA, por outro lado, é mais propenso a reatividade: os participantes podem ter um melhor desempenho do que o habitual como resultado de um processo mais estruturado de trabalho, ou podem executar pior como resultado da sua carga de trabalho dupla [27]. A segunda vantagem refere-se aos tempos do RTA, mas que não seria útil no caso de CTA, uma vez que o requisito de pensar alto retarda o processo de execução da tarefa em graus variáveis. Uma terceira vantagem seria a de que os usuários têm a possibilidade de refletir sobre seu processo de utilização do sistema, o que pode fazer com que eles destaquem causas de alto nível para problemas individuais de usabilidade. Pelos motivos apresentados optamos pelo RTA e, para reduzir o número de participantes, realizamos avaliações individuais com os participantes em vez de duplas.

O "*Retrospective Communicability Evaluation*" consistiu em três etapas principais: 1) observação, 2) *Retrospective think aloud* 3) etiquetagem; 4) elaboração do perfil semiótico

A primeira etapa, observação, consistiu na observação e gravação do uso das ferramentas pelo usuário e a realização anotações importantes observadas nesta etapa.

A segunda etapa, *Retrospective think aloud*, consistiu na observação e anotação das falas em voz alta dos usuários enquanto assistiam a gravação realizada na etapa anterior com posterior transcrição das falas dos usuários e interpretação dessas falas. A interpretação consistiu na separação das frases de acordo com a funcionalidade onde ocorreu o problema de IHC. Para facilitar a visualização dos problemas ocorridos, para cada usuário marcamos, em uma tabela, as funcionalidades cujos comentários indicaram que houve problemas.

A terceira etapa, etiquetagem, conforme no MAC, foi realizada para identificar os pontos das rupturas de comunicação, mapeando essas rupturas em problemas de IHC. De Souza [20] explica que, na etapa de etiquetagem, assistimos ao vídeo sessão do teste do usuário e analisamos para encontrar evidências de falhas de comunicação. Para cada uma das falhas encontradas, atribuímos uma entre as treze etiquetas propostas pela engenharia semiótica. Essas etiquetas são expressões comuns na comunicação humana. São elas: "Cadê?", "E agora?", "O que é isto?", "Epa!", "Onde estou?", "Ué, o que houve?", "Por que não funciona?", "Assim não dá.", "Vai de outro jeito.", "Não, obrigado!", "Pra mim está bom.", "Socorro!" e "Desisto." Na etapa de interpretação, com o material etiquetado, o avaliador busca identificar os principais problemas com a metacomunicação analisando a frequência e contexto de ocorrência de cada tipo de etiqueta, a existência de sequência padrões de cada tipo de etiqueta, o nível de problema assinalado pela ocorrência dos tipos de etiqueta e sequência e os problemas de comunicabilidade que causaram as rupturas observadas. Esses problemas podem ser classificados com a ajuda da categorização teórica das etiquetas que define três categorias de falhas: completas parciais e temporárias. Essa caracterização está ilustrada na Tabela 3¹⁹ assim como o resumo dos sintomas e característica de cada etiqueta.

Por fim, na quarta etapa, perfil semiótico, produzimos uma caracterização da mensagem geral transmitida pelo sistema [20], quando o especialista em

¹⁹ Tradução de [Erro! Fonte de referência não encontrada.] feita por [Erro! Fonte de referência não encontrada.].

engenharia semiótica reconstrói a metagemagem do designer (explicada na seção 4.2) como ela foi recebida pelo usuário.

Tabela 3: Caracterização das etiquetas de comunicabilidade [12]

Categoria	Característica	Etiqueta	Sintoma ilustrativo
Falhas temporárias 1. A interpretação do usuário é temporariamente interrompida	porque o usuário não consegue encontrar a expressão apropriada para a ação que deseja realizar.	1. Cadê?	O usuário sabe o que está tentando fazer mas não consegue encontrar um elemento da interface que irá dizer ao sistema o que fazer. Ele navega pelos menus, abre e fecha caixas de diálogo, etc., procurando por um determinado signo.
	Porque o usuário não consegue ver ou compreender a comunicação do preposto do designer.	2. Ué, o que houve?	O usuário não entende a resposta do sistema sobre o que ele disse para fazer. Frequentemente, ele repete a operação cujo efeito é ausente, ou não percebido.
	Porque o usuário não consegue encontrar uma estratégia de interação apropriada	3. E agora?	O usuário não sabe o que fazer a seguir. Ele vagueia pela interface procurando por pistas para recuperar uma comunicação produtiva com o sistema. Ele inspeciona menus, caixas de diálogo, etc., sem saber exatamente o que quer encontrar ou fazer. O avaliador deve confirmar se o usuário sabe o que está procurando ("Cadê?"), ou não ("E agora?").
Falhas temporárias 2. O usuário percebe que a intenção pretendida está errada.	Porque é dita no contexto errado	4. Onde estou?	O usuário está dizendo coisas para o sistema que seriam apropriadas em outros contextos de comunicação. Ele pode tentar selecionar objetos que não estão ativos ou interagir com signos que são apenas de saída.
	Porque sua expressão está errada	5. Epa!	O usuário comete um engano rápido mas imediatamente o corrige. A operação de "undo" é um exemplo típico dessa etiqueta.
	Porque uma conversa de vários passos não causou os efeitos desejados	6. Assim não dá	O usuário está envolvido em uma longa sequência de operações, mas de repente percebe que não é o caminho correto. Então, ele abandona aquela sequência de ações enquanto "Epa!" caracteriza uma única ação.
Falhas temporárias 3. O usuário pretende esclarecer a significação desejada pelo preposto do designer.	Através da metacomunicação implícita	7. O que é isso?	O usuário não entende um signo da interface e procura esclarecimentos lendo um <i>tooltip</i>
	através da metacomunicação explícita	8. Socorro!	O usuário pede ajuda explicitamente acessando a ajuda <i>online</i> , procurando pela documentação do sistema, ou mesmo chamando o avaliador como um "ajudante pessoal"
	através da interpretação autônoma	9. Por que não funciona?	O usuário insiste na repetição de uma operação que não produz os efeitos esperados. Ele percebe que os efeitos não são produzidos, mas acredita

			fortemente que o que está fazendo é a coisa certa a se fazer. De fato, ele não entende porque a interação não está correta.
Falhas parciais	usuário não entende a solução de design	10. Vai de outro jeito.	Usuário comunica sua intenção de design com signos inesperados porque não consegue ver ou entender o que o sistema está dizendo sobre as melhores soluções para atingir seu objetivo.
	usuário entende a solução de design	11. Não obrigado	O usuário escolhe deliberadamente por comunicar sua intenção com signos inesperados, embora tenha compreendido quais soluções preferenciais do designer foram promovidas.
Falhas Completas	usuário não tem consciência da falha	12. Pra mim está bom.	O usuário acredita que atingiu seu objetivo, embora não tenha atingido.
	usuário tem consciência da falha	13. Desisto.	O usuário acredita que não pode atingir seu objetivo e interrompe sua interação

5.2.Preparação

Definimos um cenário com duas tarefas. A fim de reduzir o efeito de aprendizado nos resultados da avaliação, utilizamos o quadrado latino para distribuir as tarefas e ferramentas entre dois grupos de usuários como ilustra a Tabela 4. As tarefas 1 e 2 foram definidas para apresentar o mesmo nível de dificuldade.

Tabela 4: Método do quadrado latino utilizado nas avaliações com usuários

Usuários	Tarefa 1	Tarefa 2
Ímpares (grupo 1)	ViSC	Tableau Public
Pares (grupo 2)	Tableau Public	ViSC

Definimos o perfil dos usuários como pessoas com habilidades na leitura de gráficos mas não deveriam ser profissionais da área como estatísticos, economistas, jornalistas ou trabalhar com análise de dados. Para garantir isso, escolhemos estudantes de graduação ou mestrado de áreas exatas. Como optamos por uma pesquisa qualitativa em profundidade, envolvemos um número reduzido de participantes. Assim, definimos que a amostra da nossa análise qualitativa seria composta de 6 participantes, 3 em cada grupo.

Para compor os 6 participantes desejados na avaliação: 3 para o grupo 1 e 3 para o grupo 2, realizamos testes com 8 usuários uma vez que 2 deles tiveram de ser descartados, conforme será explicado mais adiante. A Tabela 5 apresenta todos os perfis selecionados. As linhas sombreadas correspondem aos perfis descartados enquanto as demais referem-se aos testes que geraram os resultados apresentados.

Tabela 5: Perfis dos usuários

Usuário	Grupo	Perfil
U01	1	Idade: 20 anos Escolaridade: Graduação incompleta. Conclusão em 2015 Curso: Engenharia de Computação / PUC-Rio Nunca utiliza MS Office ou Libre Office e nunca utilizou para geração de gráficos.
U02	2	Idade: 22 anos Escolaridade: Graduação incompleta. Conclusão em 2014 Curso: Ciência da Computação / PUC-Rio Frequentemente (aprox. 2 a 3x/semana) utiliza MS Office ou Libre Office porém nunca as utilizou para geração de gráficos.
U03	1	Idade: 22 anos Escolaridade: Graduação incompleta. Conclusão em 2013 Curso: Sistemas de Informação / PUC-Rio Sempre (quase todo dia) utiliza MS Office ou Libre Office e já utilizou algumas vezes (5 a 10x) para geração de gráficos.
U04	2	Idade: 23 anos Escolaridade: Graduação incompleta. Conclusão em 2013 Curso: Engenharia de Computação / PUC-Rio Frequentemente (aprox. 2 a 3x/semana) utiliza MS Office ou Libre Office e utilizou poucas vezes (até 4x) para geração de gráficos.
U05	1	Idade: 22 anos Escolaridade: Graduação incompleta. Conclusão em 2014 Curso: Engenharia de Computação / PUC-Rio Sempre (quase todo dia) utiliza MS Office ou Libre Office e utilizou poucas vezes (até 4x) para geração de gráficos.
U06	2	Idade: 22 anos Escolaridade: Graduação incompleta. Conclusão em 2013 Curso: Engenharia de Computação / PUC-Rio Frequentemente (aprox. 2 a 3x/semana) utiliza MS Office ou Libre Office e já utilizou algumas vezes (5 a 10x) para geração de gráficos.
U07	1	Idade: 26 anos Escolaridade: Graduação completa. Conclusão: 2012 Curso: Mestrado / PUC-Rio Raramente (até 1x/mês) utiliza MS Office ou Libre Office e utilizou poucas vezes (até 4x) para geração de gráficos.

U08	2	Idade: 35 anos Escolaridade: Graduação incompleta. Conclusão em 2017 Curso: Sistemas de Informação / PUC-Rio Frequentemente (aprox. 2 a 3x/semana) utiliza MS Office ou Libre Office porém nunca as utilizou para geração de gráficos.
-----	---	---

Criamos um cenário (Cenário 1) com base em pesquisa anterior [17] fazendo ajustes no texto para torná-lo mais objetivo e reduzindo para apenas duas tarefas porque cada tarefa deveria ser realizada com uma ferramenta de acordo com o método do quadrado latino apresentado anteriormente na Tabela 4.

Cenário 1: Para tarefa a ser realizada com apoio das ferramentas

Paulo é um pesquisador do governo. Ele analisa dados estatísticos para fornecer suporte técnico e institucional às ações governamentais para a formulação e reformulação de políticas públicas e programas de desenvolvimento brasileiro. Ele trabalha basicamente com dados de educação. Certo dia, teria que participar de diversas reuniões e deixou que João, seu estagiário recém-contratado, reunisse dados em forma de gráficos para que analisassem depois. Para isso enviou um e-mail passando instruções para João.

Caro João,

Precisamos verificar como evoluíram as habilidades cognitivas dos jovens nos últimos 10 anos no Brasil e como fica o Brasil em comparações internacionais. Para essa pesquisa, preciso que gere gráficos para:

1. Mostrar se houve aumento do total de alunos com escolaridade entre 7 e 9 anos de estudo em 2003, 2006 e 2009. (**use a ferramenta** _____)
2. Observar a média das notas no exame do PISA do Brasil, Canadá e Austrália em 2003 e em 2009 e identificar o país que teve maior ascensão. (**use a ferramenta** _____)

Após a geração de cada gráfico, favor imprimir a tela e salvar em um documento Word. Se necessário, anotar observações no documento. Obrigado!

Para cada usuário, ao final da descrição de cada tarefa, acrescentamos, nas lacunas, o nome da ferramenta que deveria ser utilizada para o usuário de acordo com o seu grupo.

Nesta etapa também definimos as perguntas do questionário pré-teste no qual também incluímos uma tarefa para ser realizada no próprio papel cujo cenário foi descrito abaixo (Cenário 2). Esta tarefa visava confirmar a habilidade prévia em construção de gráficos e verificar como o usuário constrói o gráfico

sem o apoio de ferramentas. Um requisito necessário era nunca ter utilizado a ferramenta Tableau, o que também foi perguntado no questionário.

Cenário 2: Para tarefa a ser realizada em papel

Com base nos dados abaixo:

Tabela 6: Brasil - Valor do rendimento médio mensal por Pessoas de 10 anos ou mais de idade (Mil pessoas)

Classe de rendimento	2001	2003	2005	2007	2009	2011
Até 1/2 salário mínimo	6923	11238	12000	11302	13075	11588
Mais de 1/2 a 2 salários mínimos	42586	46719	55775	59293	63515	65330
Mais de 2 a 5 salários mínimos	21868	22641	21467	22749	23436	25032
Mais de 5 a 10 salários mínimos	7.665	6.587	6.773	7.107	6.389	6.810
Mais de 10 salários mínimos	4.998	4.327	3.576	3.593	3.319	3.119

Fonte 1: IBGE, PNAD.

Desenhe gráficos para:

1. Mostrar se houve crescimento ou queda do total de pessoas com rendimento maior que 5 salários mínimos na comparação de 2001 com 2011.
2. Visualizar classes de rendimento que tiveram diminuição e aumento do número de pessoas entre os anos 2009 e 2011.

Além disso, definimos as perguntas do questionário pós-testes e o roteiro de entrevista²⁰. Em ambos os questionários (pré e pós), para medir os atributos foi utilizada a escala de Likert de 5 pontos. As perguntas dos questionários e da entrevista estão no Apêndice B e serão apresentadas de forma consolidada na seção de consolidação dos resultados.

Instalamos as duas ferramentas na máquina onde os testes seriam realizados com os dados necessários igualmente carregados e a ferramenta Camtasia²¹ para a captura da gravação. Criamos um vídeo de apresentação de ViSC semelhante ao vídeo de apresentação do Tableau. Para criar este vídeo, criamos um roteiro e convidamos uma atriz para a gravação da voz, eliminando qualquer viés que poderia surgir com o uso de uma voz conhecida.

Determinamos que as etapas do testes seriam realizadas de acordo com a Tabela 7:

Tabela 7: Sequência de etapas da avaliação com os usuários e tempo estimado

Etapa	Tempo estimado
-------	----------------

²⁰ Os questionários utilizados na avaliação e o roteiro da entrevista estão no Apêndice B

²¹ <http://www.techsmith.com/camtasia.html>

1. Preenchimento do questionário pré-teste e realização da tarefa no papel.	5 minutos
2. Visualização do vídeo tutorial do Tableau e do vídeo tutorial de ViSC.	10 minutos
3. Familiarização de 5 minutos com a ferramenta Tableau e com a ferramenta ViSC.	10 minutos
4. Leitura do cenário e realização das tarefas com apoio das ferramentas.	10 minutos
5. Visualização do vídeo realizando o RTA.	10 minutos
6. Preenchimento dos questionário pós-teste.	5 minutos
7. Entrevista	10 minutos

O tempo estimado para a realização de todas essas etapas foi de uma hora conforme o somatório dos tempos estimados. Para chegar nesta definição final realizamos dois testes pilotos. Com o primeiro encontramos *bugs* que foram corrigidos na ferramenta, ajustamos questionários e incluímos a tarefa do papel. O segundo mostrou a necessidade de adequar os cenários para reduzir o tempo de teste e nivelar as tarefas.

5.3. Método de Inspeção Semiótica (MIS)

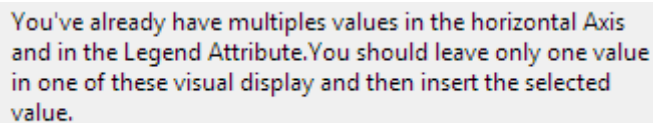
Primeiramente, realizamos uma inspeção dos signos metalinguísticos, estáticos e dinâmicos, reconstruímos a paráfrase da metamsagem do designer e verificamos se existem inconsistências, ambigüidades e contradições. Realizamos essa inspeção para cada uma das ferramentas com as duas tarefas do cenário.

5.3.1. Inspeção da ferramenta VISC

A ferramenta VISC possui poucos signos metalinguísticos. Na primeira tela encontramos a seguinte explicação:

"ViSC is a visualization tool that provides charts through a smart way. You just need to select the theme and the two dimensions you may want to compare. Thus, through picking a question you want to answer, you will have efficient visualizations."

O trecho acima descreve que a ferramenta ViSC é uma ferramenta de visualização que fornece gráficos de uma maneira inteligente. Logo depois, instrui que os usuários só precisam selecionar o tema e duas dimensões para comparar, considerando que os usuários não necessitam experiência com análises gráficas. E por fim, explica que devem selecionar uma pergunta que deseja responder para ter visualizações eficientes, o que faz uma introdução ao modo de funcionamento da ferramenta. Ao lado desta instrução, encontramos mais um signo metalinguístico que diz "*visualizações disponíveis*", apresentando ao usuário as visualizações que a ferramenta permite construir. Ao selecionar as variáveis, aparece um botão com etiqueta "*Start ViSC Now!*", onde o usuário deve clicar para iniciar a construção dos gráficos. Na tela seguinte, encontramos outros signos metalinguísticos ao posicionar o cursor sobre os ícones de ajuda ("?") presentes na tela. Existem explicações para as dimensões permitirem múltiplos elementos selecionados ou um, para as *checkbox* funcionarem ou não quando marcadas e para cada opção do menu de gráficos estar ou não habilitada. Outro signo desta natureza pode ser encontrado ao tentar inserir múltiplos valores ao gráfico para uma terceira dimensão, um alerta exibe a seguinte mensagem (Figura 34):



You've already have multiples values in the horizontal Axis and in the Legend Attribute.You should leave only one value in one of these visual display and then insert the selected value.

Figura 34: Alerta exibido na inserção de valores a uma terceira dimensão

A mensagem Figura 34 alerta o usuário que para inserir um valor a partir da terceira variável, deve-se excluir valores de alguma outra dimensão. Ainda temos uma instrução para incluir valores ao gráfico que a partir da interação é substituída pelas perguntas.

Quanto aos signos estáticos, na primeira tela vemos os campos de seleção indicando os elementos selecionados e uma imagem para cada uma das visualizações que a ferramenta constrói. Na tela de construção de gráficos, temos os menus que exibem valores disponíveis e selecionados para cada dimensão comunicando que os valores podem ser incluídos ao gráfico. Abaixo das duas dimensões selecionadas (eixo horizontal e legenda), as dimensões seguintes já possuem um valor selecionado pelo sistema e uma caixa de seleção. Há também ícones de ajuda, as opções do menu de visualizações, os *checkbox* com operações

(inversão de eixos, inclusão de variação, exclusão do valor zero e ajuste das escalas), os botões de ordenação, a área dos gráficos e a área com as perguntas, onde o link *show me* indica que é uma facilidade para o usuário.

O signos dinâmicos correspondem àqueles que tem uma duração, ou seja, ocorrem na durante a interação do usuário com a ferramenta. A Figura 35 e a Figura 36 mostram a sequência de passos na seleção de elementos para a primeira dimensão, a que foi selecionada para o eixo horizontal. Na análise dos signos dinâmicos, observamos que, na inserção dos valores ao gráfico, clicando em um elemento da coluna da esquerda (área 1) da dimensão selecionada para o eixo horizontal, esse valor aparece no gráfico (área 3) e passa para a coluna da direita (área 2).

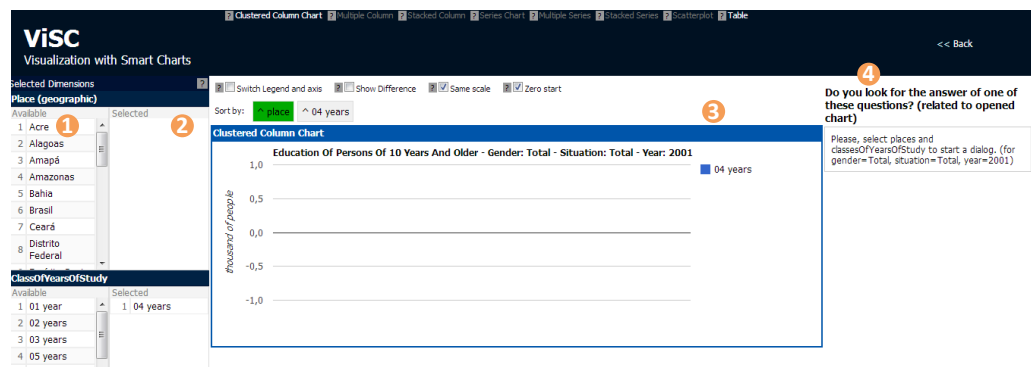


Figura 35: Passo 1 da interação de seleção em ViSC

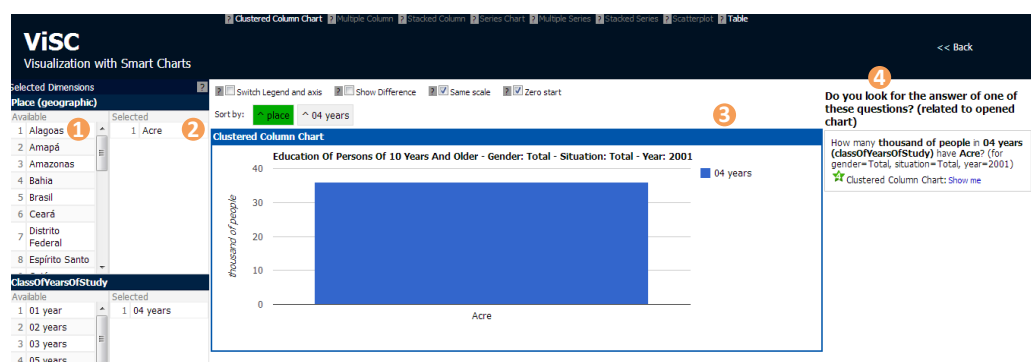


Figura 36: Passo 2 da interação da seleção em ViSC

Além disso, a primeira instrução na coluna da direita some e perguntas são apresentadas (área 4). Na Figura 37 adicionamos uma legenda. Na seleção de valores da legenda (área 6), vemos que outras opções de visualização (área 5) são habilitadas e mais perguntas (área 4) aparecem, além da remoção da coluna de

elementos disponíveis (área 6) e inclusão em elemento selecionados (área 7) e no gráfico (área 3).



Figura 37: Passo 3 da interação de seleção em ViSC

A partir da terceira coluna, o comportamento do clique já é diferente, fazendo a troca do elemento selecionado na coluna da esquerda com o elemento da coluna da direita. À medida que o usuário seleciona outro gráfico, as perguntas são reorganizadas. Para cada pergunta são apresentadas algumas opções de visualizações como resposta e cada uma dessas opções possui uma classificação. No clique da barra dos gráficos é possível expandir ou colapsar, deixando mais de uma visualização visível na tela.

Reconstruímos então a metamensagem que ficou de acordo com a descrita na seção 4.1 e verificamos as contradições, inconsistências e ambiguidades.

Na primeira tela percebemos uma **ambiguidade** pois o usuário poderia querer selecionar o gráfico diretamente através do clique das imagens de gráficos exibidas, consistindo em uma ruptura temporária que não seria um problema grave.

Para a realização da tarefa 1, seria necessária a escolha do tema "*Education of persons of 10 years and older*" e das variáveis "*ClassOfYearsOfStudy*" e "*Years*" para que pudéssemos somar as 3 classes de estudo solicitadas e selecionar os 3 anos pedidos: 2003, 2006 e 2009. As demais classes — "*place*" (lugar), "*situation*" (situação rural ou urbana) e "*gender*" (sexo) — deveriam permanecer com seus valores padrões: "Brazil", "Total" e "Total". O usuário poderia não entender que variáveis selecionar, já que a tarefa também pede dados do Brasil, e, dessa forma, selecionar a variável "*place*". Neste caso, teria dificuldade, uma vez que há uma **inconsistência** em relação a funcionalidade das duas primeiras dimensões. Para entender o funcionamento, ele deveria esperar a mensagem de

tratamento de erro (Figura 34) ou tentar entender a partir dos ícones de ajuda presentes na interface.

O gráfico de colunas empilhadas (Figura 38) e o gráfico de séries empilhadas (Figura 39) seriam as duas visualizações esperadas e possíveis de realizar tanto através do meu quanto da recomendação da pergunta: "*What is the sum of classesOfYearsOfStudy of each year on the selected set? (for gender=Total, place=Brasil, situation=Total)*"²². Ambos foram classificados com a pontuação máxima, 5, sendo os dois igualmente eficientes de acordo com a ferramenta. Segundo Few [23, p.75], gráficos de colunas empilhadas possuem desvantagens mas seu uso é justificado quando deseja-se mostrar o todo de uma unidade de medida mas também fornecer algum senso do tamanho relativo de cada parte. A mensagem dominante é o total mas o gráfico também transmite como mensagem secundária do relativo. Sobre o gráfico de séries empilhadas, Heer *et al.* [2] afirma que empilhando gráficos de área um sobre o outro chega-se a um somatório visual das séries temporais. Acrescenta ainda que esse tipo de gráfico descreve os padrões globais e, muitas vezes suporta o detalhamento em subconjuntos de séries individuais.

Um fator que influencia na recomendação é o tipo de dado. Como neste caso a variável ano pode ser tratada tanto como um dado discreto quanto contínuo, ambas as visualizações seriam possíveis. Outro fator que tem influencia é a forma como o usuário faz a seleção. Como as recomendações são dinâmicas, por se tratar de uma série, o link para o gráfico de séries empilhadas (Figura 39) só aparece se a dimensão ano tiver sido selecionada no eixo horizontal ou tiver sido selecionada na legenda e a opção de inversão de eixos for marcada pelo usuário.

Uma operação que ainda não é suportada pela ferramenta é a soma desses subconjuntos unindo em um único grupo. Isso poderia trazer problemas para o usuário que poderia considerar a tarefa **inconsistente** com a ferramenta. No entanto, um dos propósitos da tarefa era estimular a análise das visualizações.

²² Qual é a soma das classes de anos de estudo de cada ano no conjunto selecionado? (para sexo=Total, lugar=Brasil, situação=Total)

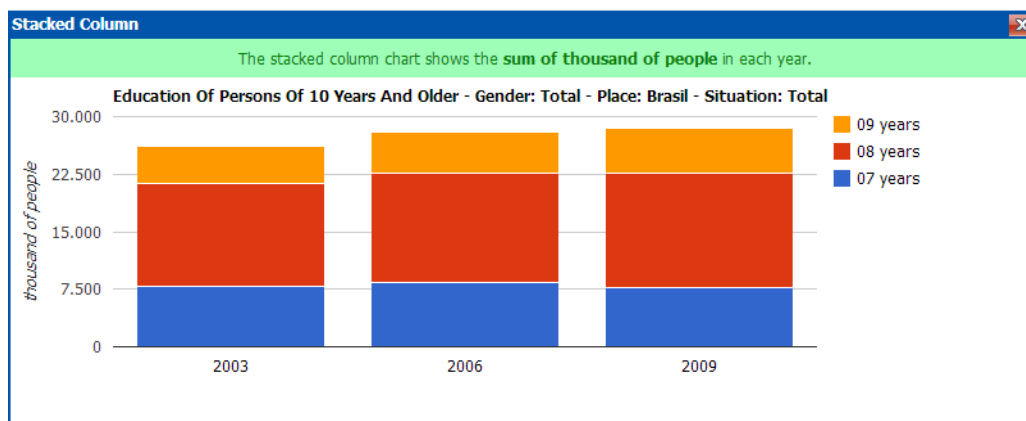


Figura 38: Gráfico de colunas empilhadas (*stacked column*) no ViSC

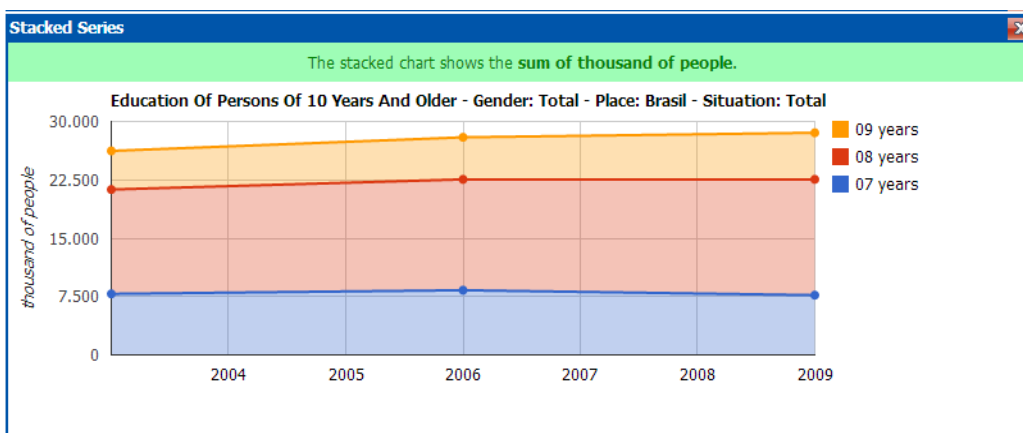


Figura 39: Séries empilhadas (*stacked series*) no ViSC

Para a realização da tarefa 2, seria necessária a escolha do tema "Score of Pisa" (Pontuação no PISA) e das variáveis "Country" (Países) e "Years" (Anos) para que pudéssemos selecionar os 3 países ("Brazil", "Canada" e "Australia") e os 2 anos pedidos: 2003 e 2009. As demais classes - *discipline*" (disciplina) e "gender" (sexo) — deveriam permanecer com seus valores padrões: "reading" (Leitura) e "Total". Conforme a tarefa anterior o usuário poderia não entender as variáveis que deveria selecionar, uma vez que a tarefa pede dados do leitura e ele poderia pensar em selecionar a variável "discipline". Assim, da mesma forma que na tarefa 1, encontraria a **inconsistência** na inserção dos valores ao gráfico. O gráfico inicialmente construído, colunas agregadas (*clustered column chart*), já mostrava a variação, mas devido à escala, ficava um pouco difícil tirar alguma conclusão. Assim, como solicitávamos a variação, o usuário poderia selecionar a opção "show difference" e ordenar pela variação de forma decrescente obtendo o gráfico da Figura 40, que apresenta o Brasil em primeiro lugar.

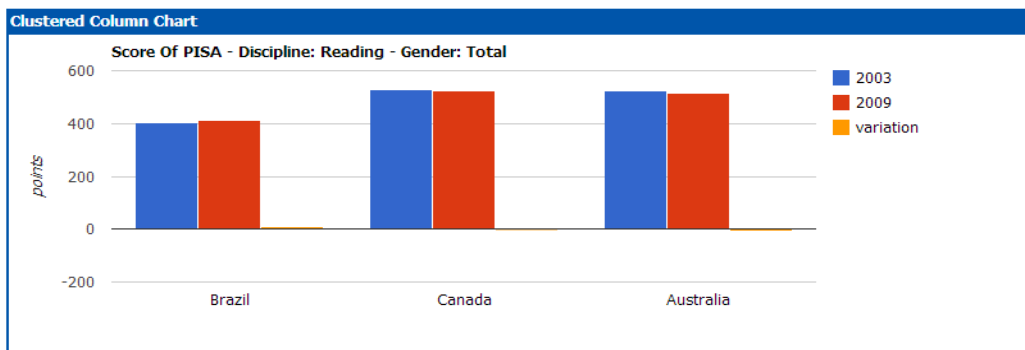


Figura 40: Resultado para tarefa 2 sem utilização das perguntas

Outra opção de chegar na resposta seria através da pergunta: "*What is the difference of points between 2003 and 2009 on the selected set? (for discipline=Reading, gender=Total)*" que recomendava 3 opções. A primeira o gráfico de dispersão (*scatterplot*) com 5 pontos, em segundo o gráfico de colunas múltiplas (*multiple column*) com 4 pontos e por último o gráfico de colunas agrupadas (*clustered column chart*) com 3 pontos representados pela Figura 41, Figura 42 e Figura 43 respectivamente.

Segundo Few [23], gráficos de pontos, tais como gráfico de dispersão ou de bolhas são usados para mostrar correlações entre pares de dados quantitativos. Quando uma das dimensões selecionadas (eixo horizontal ou legenda) possuem dois valores selecionados, ViSC habilita essa opção de visualização e transforma os dados, colocando cada valor em um eixo. A diferença é visualizada através da distância entre o ponto e até a diagonal. Os pontos que ficaram abaixo da reta tiveram queda, os pontos localizados na reta não tiveram variação e os que estão acima da reta, ascensão.

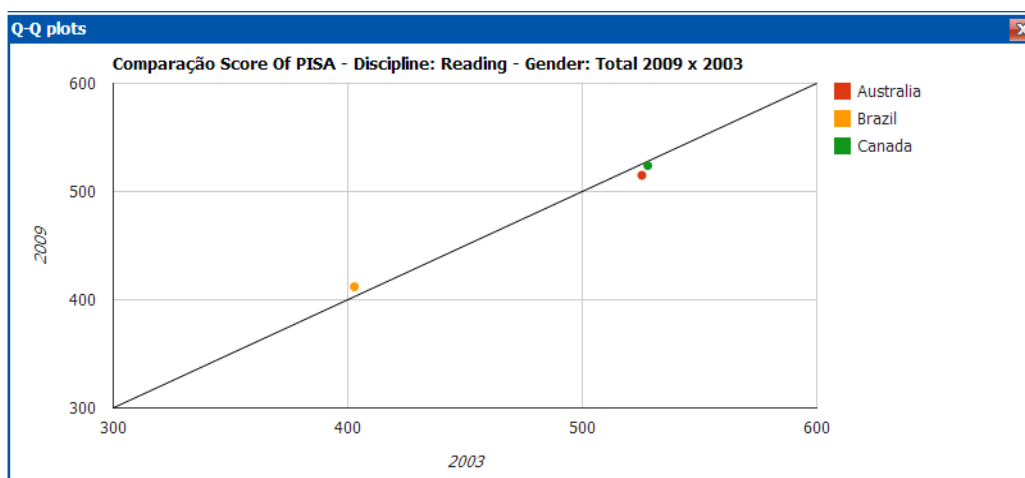


Figura 41: Solução 1 para a tarefa 2

Para o resultado exibido na Figura 41, desmarcamos a opção "zero start" para obtenção de uma imagem mais aproximada e facilitar a visualização. Na Figura 41 percebe-se que o Brasil, representado pelo ponto laranja, foi o único país com variação positiva e, portanto o país com maior ascensão.

Para Heer *et al.* [2], gráficos múltiplos são uma alternativa para evitar a sobreposição de elementos em um gráfico o que reduz a legibilidade. Esta foi considerada uma alternativa melhor que o gráfico de colunas agrupadas (Figura 42) pois permite a alteração de escalas, permitindo a melhor visualização da variação. Tanto o gráfico de colunas agrupadas quanto o de colunas múltiplas, mostram a diferença mas mantém os valores absolutos dos valores geradores desse cálculo.

Para o resultado exibido na Figura 42, desmarcamos a opção "same scale" para facilitar a visualização do terceiro gráfico com a variação. Nesta visualização também pode-se notar que o Brasil foi o único país com variação positiva.

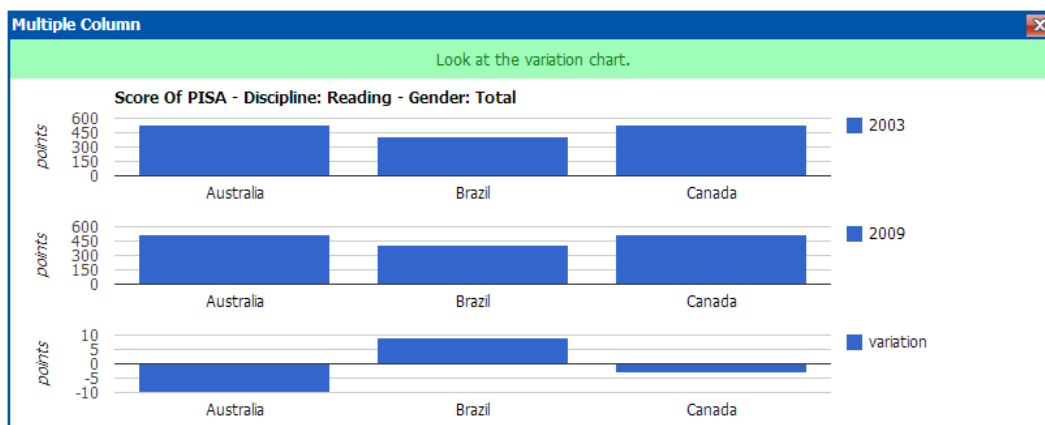


Figura 42: Solução 2 para a tarefa 2

O gráfico de colunas agrupadas (Figura 43), apesar de ser uma boa opção para exibir dados dessa natureza, não fica muito legível para diferenças tão pequenas. Este caso apresenta o problema de separação angular da regra de legibilidade de gráficos. Note na Figura 43 que as barras das variações estão tão pequenas que fica difícil perceber até mesmo quais são os valores positivos e quais são os negativos. No entanto, ele permite a ordenação pela variação conforme foi feito na Figura 40.

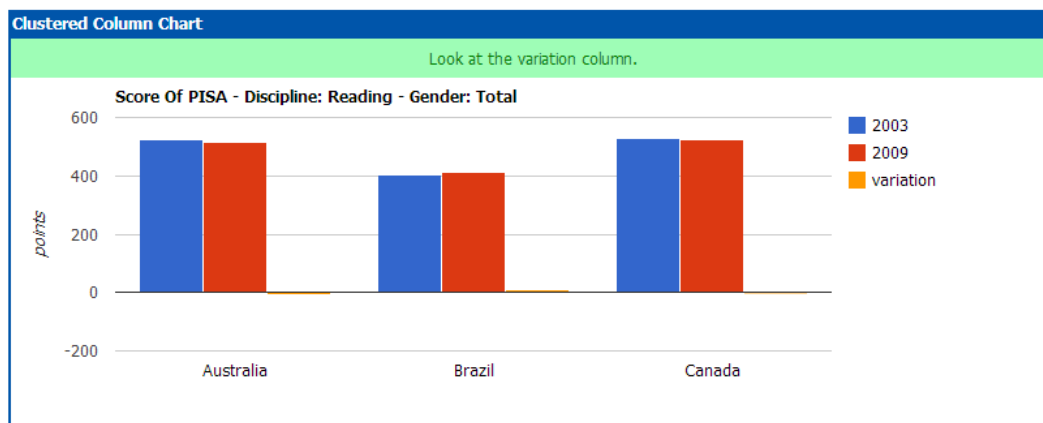


Figura 43: Solução 3

Devido às lacunas observadas na metamensagem formada por signos estáticos na troca de variáveis selecionadas, consideramos que se observássemos que o usuário tivesse problema nesta funcionalidade e isso estivesse impactando para a continuidade do teste iríamos oferecer ajuda, uma vez que nosso objetivo não era investigar a funcionalidade de seleção e sim como foi sua interação com as perguntas.

Em relação as imagens dos gráficos na tela inicial são importantes para comunicar os tipos de gráfico que a ferramenta permite. Não há *link* para eles pois conforme mencionado na seção 4.4 (sistema de recomendação de ViSC), a seleção de gráficos depende do número de elementos selecionados. Sobre as operações não disponíveis na ferramenta e às opções de sugestões apresentadas e suas pontuações, entendemos que o usuário pode achar que são inconsistentes uma vez que não está familiarizado com todas as opções de visualização fornecidas. No entanto, isso faz parte do processo abduativo que queremos investigar.

5.3.2. Inspeção da ferramenta Tableau Public

Com base na documentação presente no link de ajuda (signos metalinguísticos), vimos que Tableau Public é uma ferramenta gratuita que permite que qualquer pessoa possa publicar dados interativos na Web, interagir com os dados publicados e criar suas próprias visualizações. Enfatizam também que não é necessário conhecimento em programação. Ainda segundo a documentação, suas perguntas de negócio podem ser respondidas através de manipulação direta em uma interface visual, arrastando seus dados de forma livre entre as janelas da interface. A ferramenta permite a criação de gráficos,

relatórios e *dashboard* e o compartilhamento dos resultados com poucos cliques. As visualizações lhe permitem consultar, exibir, analisar, filtrar, ordenar, agrupar, detalhar, calcular, organizar, resumir e apresentar dados de forma rápida e eficiente. A ferramenta ainda lhe permite compartilhar visualizações interativas, relatórios e *dashboards* para que qualquer pessoa interagir, personalizar ou monitorar. Para a análise visual, Tableau Public fornece tabelas interativas, exibição de dados com imagens eficientes, comparações lado a lado e gráficos codificados com uso de cor, tamanho e forma. O passo fundamental do Tableau para o processo de análise é começar com uma visão geral para depois focar nos detalhes de interesse. Tableau Public ainda possui uma interface flexível que permite uma exploração de forma livre.

Os signos metalinguísticos encontrados na interface foram o texto "*Drop field here*" que elucida que os dados devem ser inseridos na área retangular no meio da tela conforme detalhe mostrado na Figura 44. Além disso, no final da janela *Show me* (Figura 45), o usuário é informado que deve selecionar ou arrastar o dado e, para isso pode usar as teclas *Shift* ou *Ctrl* para selecionar mais de um campo.



Figura 44: Detalhe da área do gráfico antes da inserção dos dados

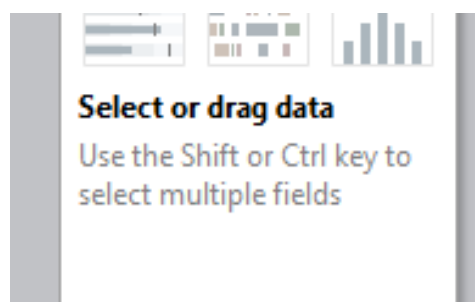


Figura 45: Detalhe do final da janela Show me.

Com base na análise dos signos estáticos vimos que Tableau Public possui diversas funcionalidades que estão presentes no menu superior. Analisando esses menus vemos como abrir, salvar, fechar ou publicar na Web um novo documento. Encontramos também como abrir ou colar um banco de dados, como criar um novo *worksheet* ou limpar o existente, como criar um *dashboard*. Observamos ainda que é possível realizar operações ligada à análises de dados (operações que não parecem fáceis à primeira vista) e formatar fontes, sombreamentos, alinhamentos, bordas, linhas, cores, anotações, legendas, entre outras configurações. Esse menu permite ainda ocultar ou exibir janelas e alternar entre o modo de apresentação ou de desenvolvimento e contém um link para a ajuda (*help*). Existe ainda uma barra de ferramentas com atalhos para algumas dessas funcionalidades.

Os signos estáticos também são constituídos pelas janelas abertas por *default*. A Figura 46 mostra essas janelas: dados (*data*), dimensões (*dimensions*), medidas (*measures*), páginas (*pages*), filtros (*filters*), marcas (*marks*), colunas (*columns*), linhas (*rows*), o espaço do gráfico. A janela *Show me* foi destacada separadamente na Figura 47.

O usuário então deve estar familiarizado com esses termos. Na coluna da esquerda vemos dados, dimensões e medidas. Em dados vemos o nome dos dados importados. Em dimensões, cada um dos campos de categoria do dados e, em medidas encontramos os valores quantitativos do dado. O sistema cria também, por *default*, como dado quantitativo, o número de registros (*number of records*) e os valores de medidas (*measures values*). Este último não deixa um significado muito claro. Note que ao lado de cada valor, existe um símbolo. Ao lado de cada dimensão, o signo "Abc" representa que a variável é do tipo *string*. O ícone de calendário representa que é do tipo *data*, e a tralha ("#") representa que é um tipo numérico.

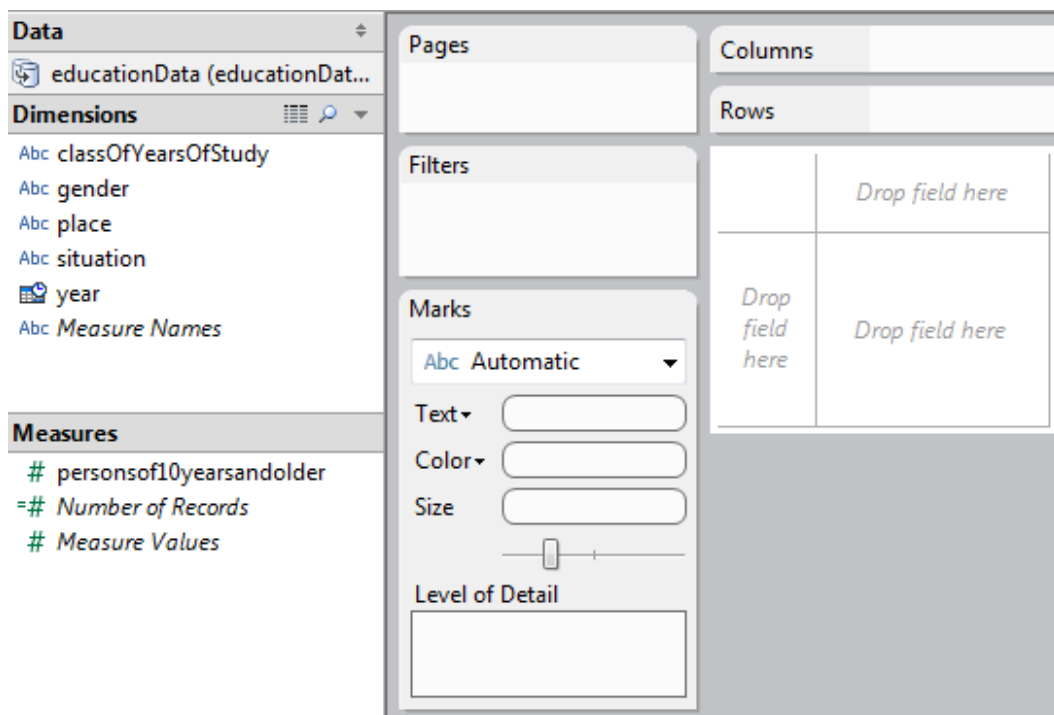


Figura 46: Janelas do Tableau

Na coluna do meio, existe a janela páginas que está vazia, a janela filtros, vazia também mas que indica que ali pode-se criar filtros e as marcas. O campo de seleção com o item automático não é claro na sua função mas percebe-se que ali é possível alterar texto, cor e tamanho. A coluna da direita possui as janelas colunas, linhas e uma área retangular onde é possível assumir que é a área do gráfico. Linhas e colunas são termos utilizados para construção de tabelas e isso é familiar para o usuário que já utilizou um programa de planilhas por exemplo. Na construção de gráficos, no entanto, é mais difícil de relacionar esses termos.

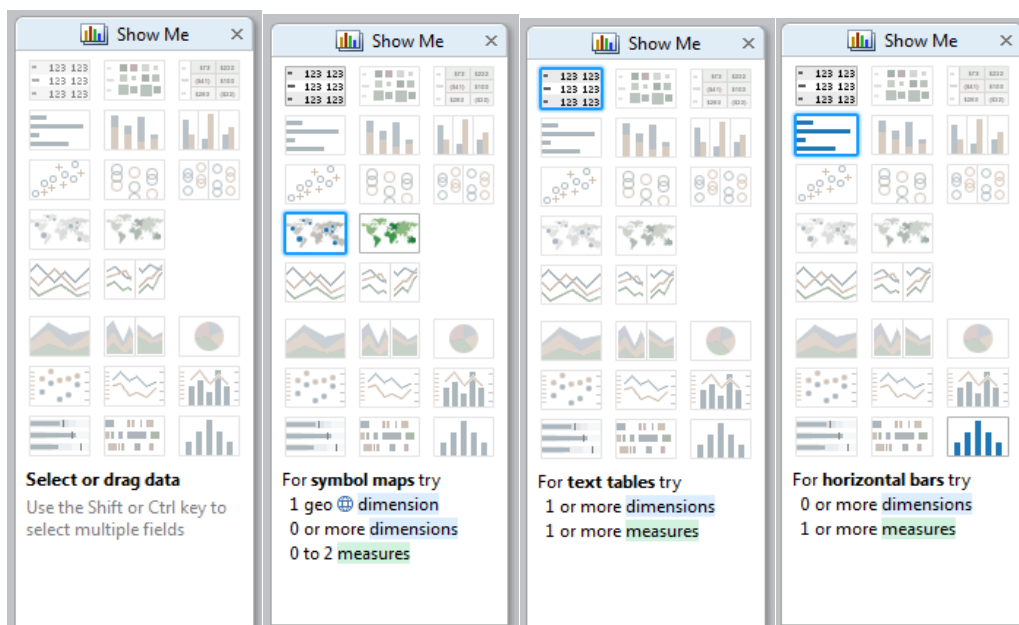


Figura 47: Janela *Show me* em 4 momentos da esquerda para a direita: sem seleção, na seleção de uma dimensão geográfica, na seleção de uma *string* e na seleção de uma medida

Observando os signos da janela *Show me* (Figura 47) percebemos os gráficos que podem ser construídos com a ferramenta. No entanto, os ícones parecidos não permitem identificar a diferença entre eles sem a interação. Note que as explicações ao final da janela estão destacadas da seguinte forma: *dimensions* (dimensões) em azul e *measures* (medidas) em verde. Da mesma forma, quando uma dimensão é inserida no gráfico, ela fica em azul nas janelas de linhas e colunas (Figura 48). E, quando uma medida é inserida, fica na cor verde (Figura 49).



Figura 48: Linhas e colunas preenchidas com dimensões

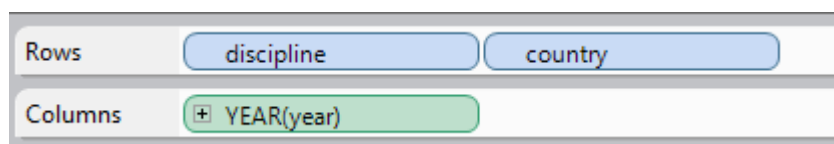


Figura 49: Linhas e colunas preenchidas com uma medida

As janelas de filtros também serão utilizadas pelo nosso cenário. A Figura 50 apresenta a janela de filtros (à direita) que é selecionada através do menu de contexto (à esquerda) aberto ao clicar sobre uma **dimensão** do gráfico. A Figura

51 apresenta a janela de filtros (à direita) que é selecionada através do menu de contexto (à esquerda) exibido ao clicar sobre uma **medida** do gráfico. Note que ambas as janelas e os menus apresentam muitas diferenças.

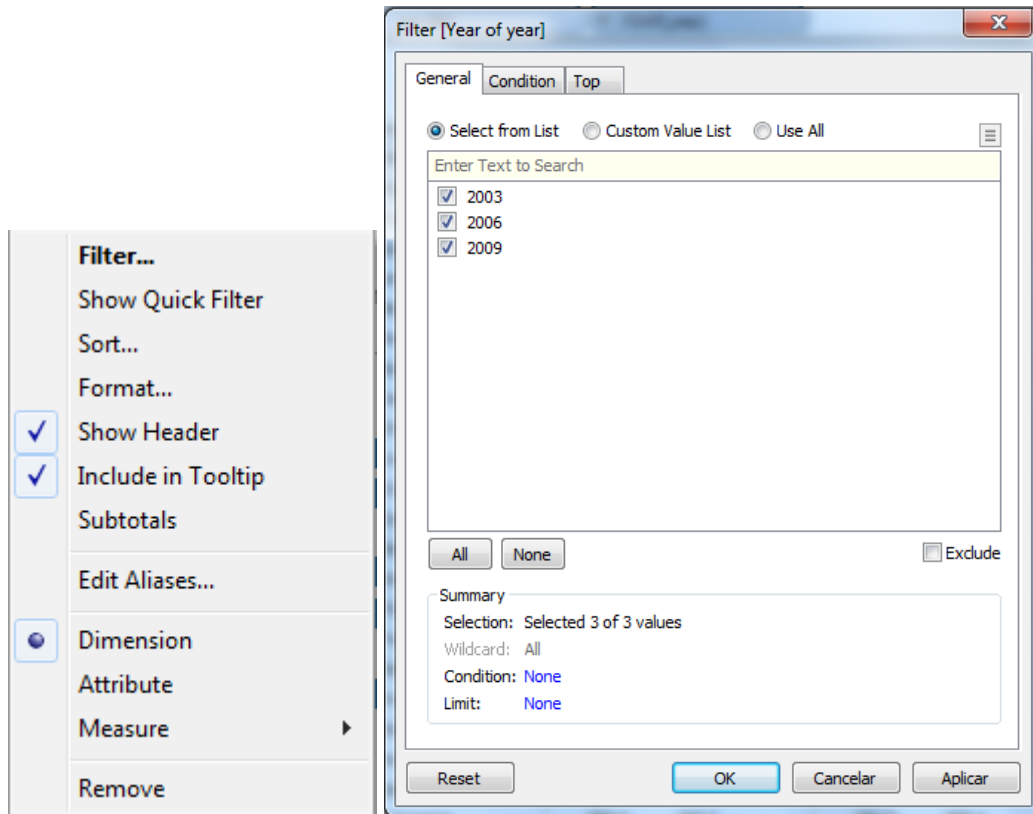


Figura 50: Menu de contexto e janela filtro para uma variável discreta ou uma dimensão

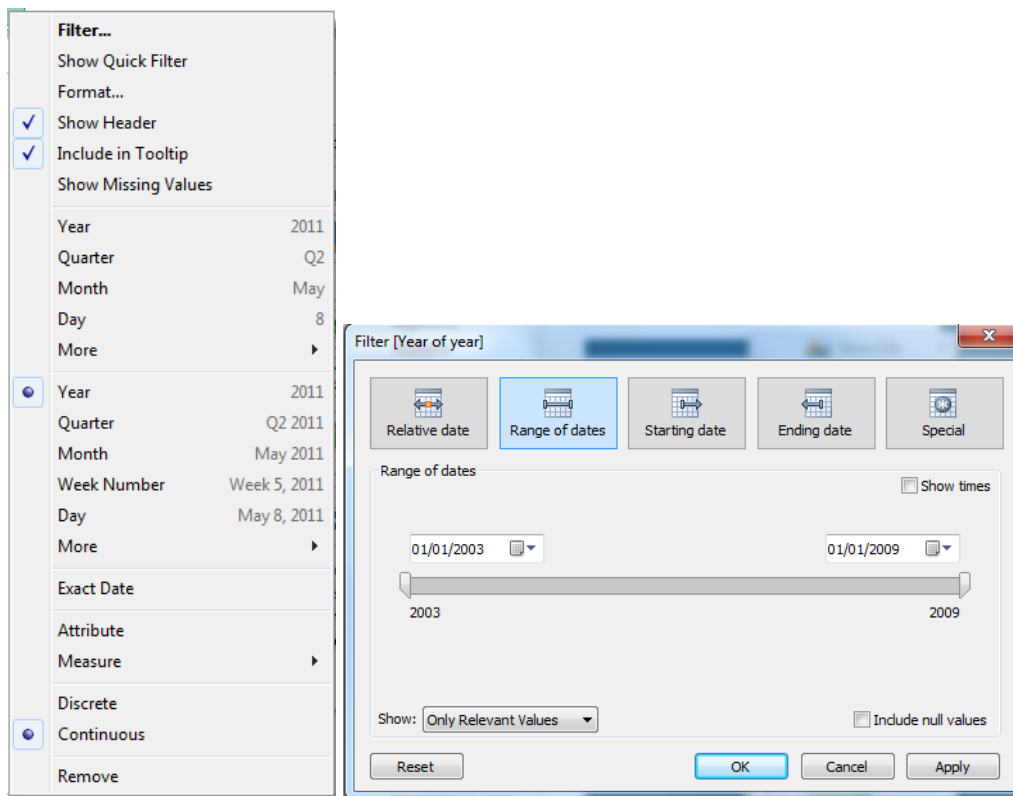


Figura 51: Menu de contexto e janela filtro para uma variável contínua ou uma medida

Na observação dos signos dinâmicos nota-se, na Janela *Show me* (Figura 47) que os gráficos se tornam ativos ou inativos de acordo com as variáveis selecionadas ou incluídas no gráfico. A Figura 47 mostra a janela *Show me* em vários momentos, por exemplo: a) inicialmente a janela está toda desabilitada; b) depois quando uma classe geográfica é selecionada em medidas as visualizações dos mapas se tornam ativas; c) quando a classe é do tipo *string*, a única opção de visualização é a tabela e; d) no caso de uma medida, alguns diagramas se tornam ativos.

É possível arrastar os valores (dimensões ou medidas) tanto para o gráfico quanto para as janelas de colunas e linhas. O duplo clique também insere o valor no gráfico mas o critério que relaciona com as marcas não fica claro. Ao incluir uma dimensão, tais valores ficam coloridos em azul nas janelas de linhas e colunas (Figura 48), conforme descrito anteriormente. No entanto, por exemplo, ao substituir a tabela pelo gráfico de Gantt²³ através da janela *Show me*, o sistema executa uma transformação no dado de forma que a dimensão ano (*year*) tem sua

²³ Gráficos de Gantt são eficientes para mostrar duração em relação a uma variável quantitativa independente. No Tableau, Gantts são também uma forma de mostrar a distribuição da medida [16].

cor alterada para verde. Neste caso, o dado ano passa a ser tratado como contínuo e não mais como discreto como as dimensões são tratadas. Assim, as cores azul e verde também se referem ao tipo do dado: contínuo ou discreto.

Na troca entre outras visualizações, os dados selecionados também sofrem transformações que podem ou não retirar dados selecionados, passando para a janela de filtros ou para uma marca, janela de marcas, tal como o campo texto. Essa transformações não são tão facilmente observadas e as alterações entre os dados pode confundir o usuário principalmente pelas funcionalidades que são diferentes.

Assim, após a análise dos signos, desconsideramos as mensagens do *help*, uma vez que encontramos inconsistências entre o perfil do usuário desse sistema no *help* e na inspeção dos signos estáticos. Como o *help* não seria acessado, utilizamos preferencialmente as informações obtidas na inspeção da interface. Assim, reconstruímos a metagemagem do designer da seguinte forma:

Penso que você já tenha alguma experiência com análise de dados, ou pelo menos entende como os dados e suas variáveis podem ser classificadas. Você também pode ser um especialista nesse assunto. Você prefere uma interface que seja flexível na interação pois você pode ter preferência por interagir através de drag and drop, com clique do mouse ou com a ajuda do teclado. Você também gostaria de ter muitas opções de visualização e que a experiência na troca de visualizações fosse simples. Além disso, você gostaria de poder realizar operações matemáticas com os dados, mudar as preferências de visualização, construir dashboards e compartilhar seus gráficos. Assim projetamos uma interface onde basta arrastar os dados para a área do gráfico ou clicar sobre eles que eles são automaticamente incluídos na sua visualização. Você também pode usar o teclado para ajudar na seleção. Uma janela chamada Show me exibe quando os gráficos estão ou não ativos e lhe explica o porquê de cada visualização. Para simplificar a troca das visualizações, o sistema transforma o dado de acordo com a visualização selecionada, reorganizando as marcas das variáveis selecionadas, podendo até mudar características das variáveis. Outra solução dada pelo sistema foi a visão geral dos dados como forma de análise inicial. Os detalhes são obtidos através da inclusão e filtragem dos valores.

Para realizar o gráfico da tarefa 1, a primeira opção seria o gráfico denominado *stacked bar* — em português, barras empilhadas — *stacked columns*

no ViSC. Como alguns passos a mais que a ferramenta anterior era possível desenhar o gráfico com Tableau da seguinte forma: a) incluir a dimensão "*year*" (ano) e a dimensão "*classesOfYearsOfStudy*" (classes de anos de estudo) na coluna e, na linha, incluir a unidade de medida do dado que era "*peopleOf10YearsAndOlder*" (pessoas de 10 ou mais anos de idade). b) filtrar as duas dimensões selecionando apenas as classes de anos pedidos (7, 8 e 9 anos) e escolher o gráfico de barras empilhadas. Como a ferramenta mostrava a visão global, o que implica na soma de todos os valores das dimensões não selecionadas, e os dados possuíam um valor totalizador em cada uma das outras dimensões — "*gender*" (sexo), "*situation*" (situação) e "*place*" (lugar) —, essas também deveriam ser incluídas no gráfico e filtradas de forma que "*gender*" e "*situation*" estivessem somente com o total selecionado e "*place*" com "Brazil". O resultado é apresentado na Figura 52.

Uma forma encontrada para unir as cores no Tableau seria através da alteração das configurações de cores da legenda, colocando todas elas da mesma cor. A Figura 53 apresenta como o gráfico ficaria.

O Tableau também possibilita o gráfico de séries empilhadas (Figura 54), aqui chamado de *area charts* (gráfico de área). Para desenhá-lo, após a seleção dos dados alterar o gráfico para *area charts*. Do mesmo modo, era possível unir as cores através das alterações nas configurações de cor (Figura 55), selecionando a mesma cor para cada elemento da legenda.

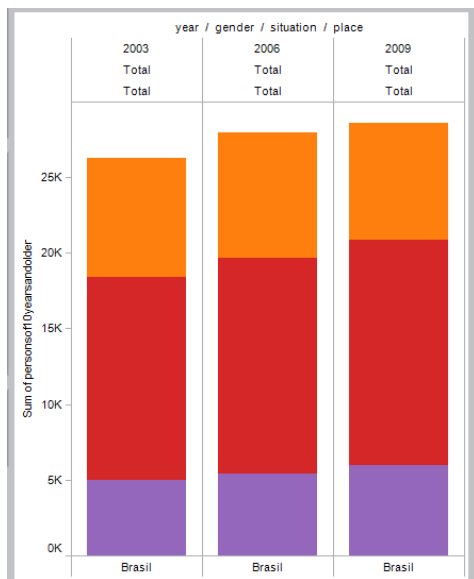


Figura 52: Resultado 1 obtido para a tarefa 1 com Tableau

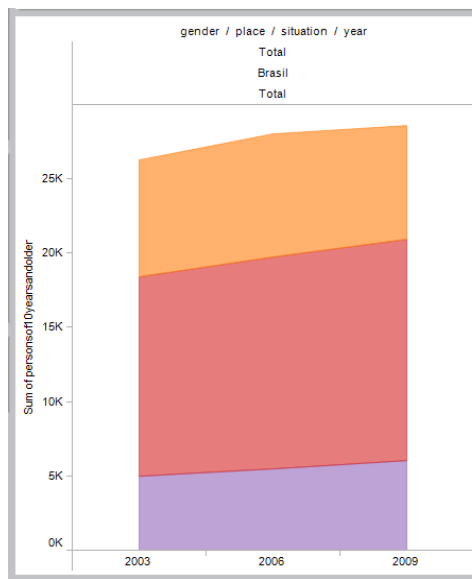


Figura 54: Resultado 3 obtido para a tarefa 1 com Tableau

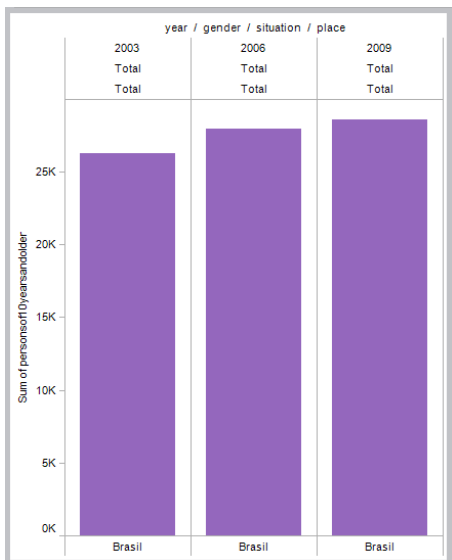


Figura 53: Resultado 2 obtido para a tarefa 1 com Tableau

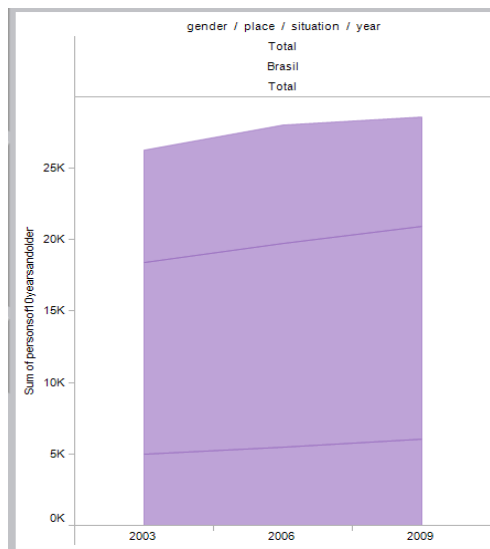


Figura 55: Resultado 4 obtido para a tarefa 1 com Tableau

Para realizar o gráfico da tarefa 2, a opção gráfico de dispersão não foi possível porque a ferramenta não permite criar dois filtros para uma dimensão, o que seria necessário para incluir apenas 2003 na linha e apenas 2009 na coluna.

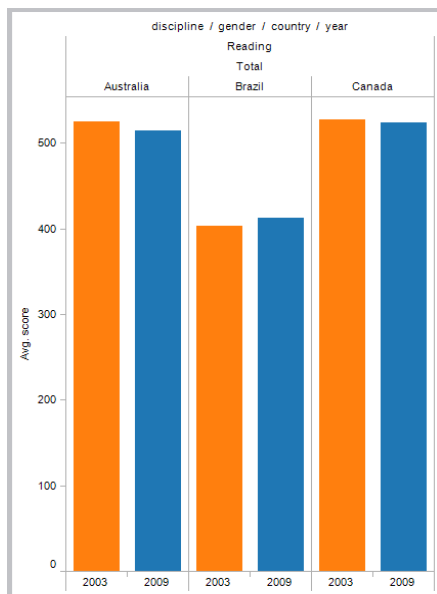


Figura 56: Solução para tarefa 2 com Tableau

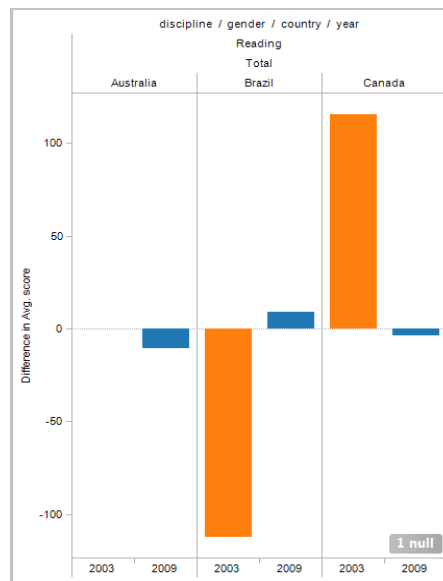


Figura 57: Tentativa de inclusão da variação no Tableau

A solução encontrada foi então o gráfico de colunas agregadas chamado de "*side-by-side bars*" (barras lado a lado) pelo Tableau. Os passos para montar o gráfico da Figura 56 eram: a) incluir a dimensão "*discipline*" e "*country*" e "*year*" na coluna e, na linha, incluir a unidade de medida do dado que era "*score*". b) filtrar as três dimensões: "*discipline*" (disciplina) pelo valor "*reading*" (leitura), "*countries*" (países) pelo valores "Brazil", "Canada" e "Australia" e "*years*" (anos) pelos valores 2003 e 2009. Depois escolher o gráfico "*side-by-side bars*", o mesmo chamado de colunas agrupadas em ViSC. Conforme explicado anteriormente, era preciso incluir também a dimensão "*gender*" (sexo), e selecionar somente o valor "*Total*". Essa foi o resultado mais eficiente obtido. Comparando com o resultado obtido na ViSC com as colunas agregadas que tinha pontuação 3, este gráfico é mais eficiente uma vez que está em uma escala diferente e assim, não apresentou problema de separação angular, facilitando a análise.

Uma tentativa de incluir a diferença entre as barras gerou o resultado da Figura 57. As barras azuis parecem exibir o valor da diferença entre 2009 e 2003, mas as barras laranjas não são facilmente entendidas.

Verificamos que haviam **inconsistências** entre o que diz o *help* e os termos utilizados pela ferramenta que sugerem que o usuário pelo menos entenda a diferença entre dimensão e medida, entre um dado discreto e um contínuo. Outro

problema encontrado no cenário foi a dificuldade em encontrar operações simples como soma na tarefa 1 ou a média na tarefa 2. Pode ficar **ambíguo** para o usuário que esteja procurando tais operações. No menu de *Analysis* são fornecidas várias operações com nomes semelhantes que podem levá-lo ao erro. As operações solicitadas são encontradas, no entanto, no menu de contexto que se abre ao clicar com o botão direito sobre o dado quantitativo presente no gráfico.

5.4. Retrospective Communicability Evaluation

O "*Restrospective Communicability Evaualiton*" foi o método utilizado para a avaliação com usuários. Para cada tarefa descrevemos as três etapas do método com cada uma das duas ferramentas: ViSC e Tableau.

Na etapa de observação, os dois vídeos das ferramentas foram apresentados para cada usuário. Após o vídeo, cada usuário teve cinco minutos para interagir com cada ferramenta já com os dados importados. Só então passamos o cenário com as tarefas a serem realizadas. Orientamos que eles realizassem as tarefas na ordem e que se houvesse alguma dificuldade com inglês poderia ser perguntado. Além disso, explicamos cada variável dos temas carregados. O tempo total dos testes duraram entre uma hora e uma hora e vinte minutos, conforme o tempo estimado. O único que apresentou discrepância em relação ao tempo estimado foi o do usuário U08 que durou quase duas horas.

Conforme mencionado, os usuários U01 e U06 tiveram seus testes descartados. U01 finalizou as duas tarefas com sucesso apesar de não ter utilizado as perguntas. Declarou não ter utilizado porque achou improvável encontrar a pergunta que queria no meio delas. O problema do teste foi um erro encontrado na ontologia cadastrada na ferramenta ViSC o que, segundo U01 atrapalhou um pouco na realização da tarefa. Apesar de, ainda assim, U01 ter classificado a ferramenta ViSC como fácil, o erro foi corrigido posteriormente e o teste teve que ser descartado. No caso de U06, seu teste foi descartado por ter interpretado de forma errada a tarefa realizada em ViSC. Ele utilizou as perguntas para auxiliá-lo na resposta e encontrou um gráfico eficiente para a sua interpretação da tarefa, mas a tarefa realizada foi diferente do que foi pedido.

Outro problema ocorrido com os dados coletados foi a perda da entrevista de U08 por problemas técnicos. No entanto, extraímos o resultado de seu teste a

partir dos questionários preenchidos, das duas gravações — o teste realizado e a narração do teste feita por ele durante a observação do vídeo — e das anotações realizadas durante a entrevista.

Na interpretação, primeiramente realizamos o RTA, verificando quais funcionalidades não foram bem compreendidas. Depois realizamos a etiquetagem dos vídeos de interação procurando rupturas de comunicação, ou seja, classificamos os momentos da interação em que o usuário demonstra dificuldades em compreender a metamensagem do designer ou em transmitir a sua intenção com a interface. Esse método possibilitou encontrar tanto as falhas na comunicação como também pontos positivos que os usuários encontraram nas ferramentas, além de revelar sobre seus processos abduativos.

5.4.1. Tarefa 1

Esta seção apresenta os dados coletados na etapa de observação da tarefa 1 — *Mostrar se houve aumento do total de alunos com escolaridade entre 7 e 9 anos de estudo em 2003, 2006 e 2009* — e a interpretação realizada com estes dados.

5.4.1.1. Observação de uso

Nesta etapa apresentamos os fatos anotados durante a observação, os resultados entregues por cada usuário e os dados preenchidos nos questionários na tarefa 1.

5.4.1.1.1. Com apoio de ViSC

Entre os usuários do grupo 1, que realizaram com o apoio de ViSC, obtivemos os resultados esperados, conforme exibidos nas Figura 58, Figura 59 e Figura 60.



Figura 58: Resultado de U03 para a tarefa 1

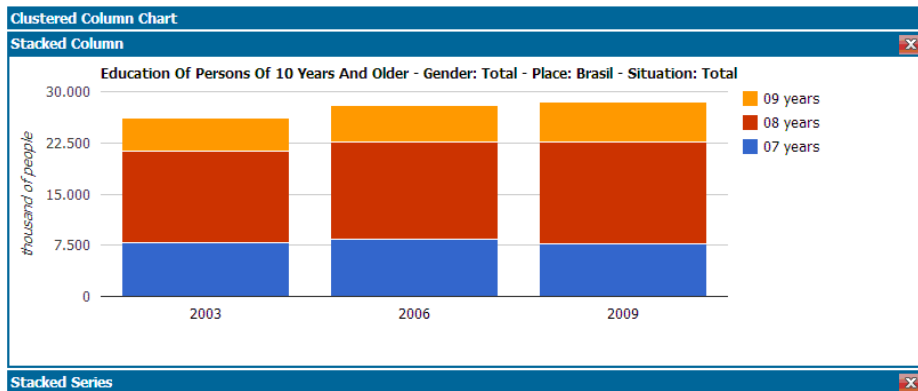


Figura 59: Resultado de U05 para a tarefa 1

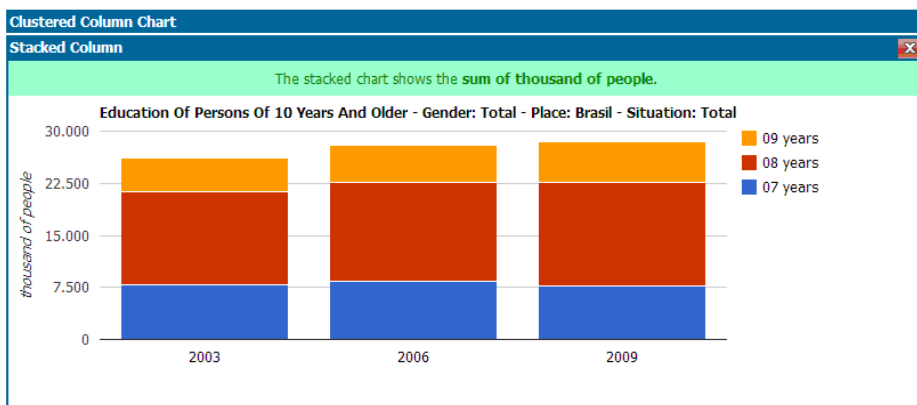


Figura 60: Resultado de U07 para a tarefa 1

Todos os três usuários desse grupo utilizaram as perguntas e todos eles analisaram as opções recomendadas para decidir sua resposta.

U03 inicialmente entregou o gráfico de colunas agrupadas (*Clustered Column Chart*) com dados errados após a inserção do ano 2007 (linha 7 - Figura 61), sem tentar outra alternativa. Assim, **oferecemos ajuda chamando atenção para que lesse a tarefa novamente**. Ele corrigiu os dados que, por falta de atenção havia incluído de forma errada e recorreu diretamente à pergunta correta, encontrando com facilidade o gráfico séries empilhadas (*Stacked Series*) como pode ser visto no log (cod. 36_6 - Figura 61), uma das respostas para a pergunta que buscava responder. Após encontrar, optou por entregar o gráfico inicial junto com a recomendação, *Stacked Series* (Figura 58).

rckjbeeamnoqt7dlaamgeuldh0	5956	click	back			2012-11-28 08:31:06
rckjbeeamnoqt7dlaamgeuldh0	5957	click	addValueChart	addRow	07 years	2012-11-28 08:31:26
rckjbeeamnoqt7dlaamgeuldh0	5958	click	addValueChart	addRow	08 years	2012-11-28 08:31:28
rckjbeeamnoqt7dlaamgeuldh0	5959	click	addValueChart	addRow	09 years	2012-11-28 08:31:29
rckjbeeamnoqt7dlaamgeuldh0	5960	click	addValueChart	addCol	2006	2012-11-28 08:31:32
rckjbeeamnoqt7dlaamgeuldh0	5961	click	addValueChart	addCol	2003	2012-11-28 08:31:37
rckjbeeamnoqt7dlaamgeuldh0	5962	click	addValueChart	addCol	2007	2012-11-28 08:31:38
rckjbeeamnoqt7dlaamgeuldh0	5963	click	remValueChart	remCol	2001	2012-11-28 08:32:23
rckjbeeamnoqt7dlaamgeuldh0	5964	click	remValueChart	remCol	2007	2012-11-28 08:32:40
rckjbeeamnoqt7dlaamgeuldh0	5965	click	addValueChart	addCol	2009	2012-11-28 08:32:41
rckjbeeamnoqt7dlaamgeuldh0	5966	click	operation	variation	1	2012-11-28 08:32:56
rckjbeeamnoqt7dlaamgeuldh0	5967	click	showme	36_6	What is the sum	2012-11-28 08:33:18

Figura 61: Log de uso do ViSC por U03

U05 também encontrou o gráfico séries empilhadas (*Stacked Series*) através das perguntas mas optou pelo gráfico que já havia encontrado, o de colunas empilhadas (*Stacked Column*). Observe na Figura 59 que U05 minimizou os gráficos *Stacked Series* e *Clustered Column Chart*. Chegou à conclusão usando o menu de gráficos e mesmo após encontrar o gráfico que respondia, quis olhar os outros pra tentar reunir as partes em uma só coluna. Nessa busca, encontrou a pergunta que exibia a soma de pessoas das três classes solicitadas e clicou nas duas opções: *Stacked Column* (cod. 36_3 - Figura 62) e depois *Stacked Series* (cod. 36_6 - Figura 62), ambas com pontuação máxima e optou pelo gráfico selecionado inicialmente.

7k7toq99k5q0hh8jg9shv82g8q4	6184	click	select	independentVar2	classOfYearsOfStudy	2012-12-10 11:56:57
7k7toq99k5q0hh8jg9shv82g8q4	6185	click	send			2012-12-10 11:56:58
7k7toq99k5q0hh8jg9shv82g8q4	6186	click	back			2012-12-10 11:56:58
7k7toq99k5q0hh8jg9shv82g8q4	6187	click	addValueChart	addRow	2003	2012-12-10 11:57:12
7k7toq99k5q0hh8jg9shv82g8q4	6188	click	addValueChart	addRow	2006	2012-12-10 11:57:14
7k7toq99k5q0hh8jg9shv82g8q4	6189	click	addValueChart	addRow	2009	2012-12-10 11:57:15
7k7toq99k5q0hh8jg9shv82g8q4	6190	click	addValueChart	addCol	07 years	2012-12-10 11:57:23
7k7toq99k5q0hh8jg9shv82g8q4	6191	click	remValueChart	remCol	04 years	2012-12-10 11:57:24
7k7toq99k5q0hh8jg9shv82g8q4	6192	click	addValueChart	addCol	08 years	2012-12-10 11:57:25
7k7toq99k5q0hh8jg9shv82g8q4	6193	click	addValueChart	addCol	09 years	2012-12-10 11:57:26
7k7toq99k5q0hh8jg9shv82g8q4	6194	click	topmenu	2	ColumnChart	2012-12-10 11:58:01
7k7toq99k5q0hh8jg9shv82g8q4	6195	click	topmenu	3	ColumnChart	2012-12-10 11:58:09
7k7toq99k5q0hh8jg9shv82g8q4	6196	click	topmenu	5	AreaChart	2012-12-10 11:58:29
7k7toq99k5q0hh8jg9shv82g8q4	6197	click	topmenu	4	LineChart	2012-12-10 11:58:34
7k7toq99k5q0hh8jg9shv82g8q4	6198	click	operation	sort	^ 07 years	2012-12-10 11:58:51
7k7toq99k5q0hh8jg9shv82g8q4	6199	click	operation	sort	^ year	2012-12-10 11:58:53
7k7toq99k5q0hh8jg9shv82g8q4	6200	click	operation	sort	^ 07 years	2012-12-10 11:58:54
7k7toq99k5q0hh8jg9shv82g8q4	6201	click	operation	sort	^ 08 years	2012-12-10 11:58:57
7k7toq99k5q0hh8jg9shv82g8q4	6202	click	operation	sort	^ 09 years	2012-12-10 11:58:58
7k7toq99k5q0hh8jg9shv82g8q4	6203	click	operation	sort	^ year	2012-12-10 11:59:00
7k7toq99k5q0hh8jg9shv82g8q4	6204	click	showme	34_3	What is the sum of classesOfYearsOfStudy of year	2012-12-10 11:59:10
7k7toq99k5q0hh8jg9shv82g8q4	6205	click	showme	34_6	What is the sum of classesOfYearsOfStudy of year	2012-12-10 11:59:14

Figura 62: Log de uso do ViSC por U05

U07 selecionou as variáveis corretas nos eixos e encontrou a solução a partir do clique no gráfico de colunas empilhadas (*Stacked Column*) da pergunta de ordenação. Após o clique, encontrou a pergunta de soma e clicou nela, obtendo o mesmo resultado. Percebeu que os anos estavam fora de ordem e tentou retirá-los e inseri-los em ordem cronológica, sem sucesso. Até que conseguiu reordená-los pelos botões de ordenação e salvou o resultado.

id_session	order	event	category	id_value	value	date_time
r3e02c7a9kmro120pc1j9ka840	6351	click	select	independentVar1	year	2012-12-10 07:32:52
r3e02c7a9kmro120pc1j9ka840	6352	click	select	independentVar2	classOfYearsOfStudy	2012-12-10 07:33:07
r3e02c7a9kmro120pc1j9ka840	6353	click	send			2012-12-10 07:33:08
r3e02c7a9kmro120pc1j9ka840	6354	click	back			2012-12-10 07:33:08
r3e02c7a9kmro120pc1j9ka840	6355	click	addValueChart	addRow	2003	2012-12-10 07:33:19
r3e02c7a9kmro120pc1j9ka840	6356	click	addValueChart	addRow	2006	2012-12-10 07:33:22
r3e02c7a9kmro120pc1j9ka840	6357	click	addValueChart	addRow	2009	2012-12-10 07:33:25
r3e02c7a9kmro120pc1j9ka840	6358	click	addValueChart	addCol	07 years	2012-12-10 07:33:31
r3e02c7a9kmro120pc1j9ka840	6359	click	addValueChart	addCol	08 years	2012-12-10 07:33:33
r3e02c7a9kmro120pc1j9ka840	6360	click	addValueChart	addCol	09 years	2012-12-10 07:33:36
r3e02c7a9kmro120pc1j9ka840	6361	click	remValueChart	remCol	04 years	2012-12-10 07:33:38
r3e02c7a9kmro120pc1j9ka840	6362	click	topmenu	1	ColumnChart	2012-12-10 07:34:13
r3e02c7a9kmro120pc1j9ka840	6363	click	showme	5_3	What is ranking of year in 	2012-12-10 07:34:38
r3e02c7a9kmro120pc1j9ka840	6364	click	showme	34_3	What is the sum of classes	2012-12-10 07:34:59
r3e02c7a9kmro120pc1j9ka840	6365	click	remValueChart	remRow	2009	2012-12-10 07:35:05
r3e02c7a9kmro120pc1j9ka840	6366	click	remValueChart	remRow	2003	2012-12-10 07:35:08
r3e02c7a9kmro120pc1j9ka840	6367	click	remValueChart	remRow	2006	2012-12-10 07:35:09
r3e02c7a9kmro120pc1j9ka840	6368	click	addValueChart	addRow	2003	2012-12-10 07:35:10
r3e02c7a9kmro120pc1j9ka840	6369	click	addValueChart	addRow	2006	2012-12-10 07:35:12
r3e02c7a9kmro120pc1j9ka840	6370	click	addValueChart	addRow	2009	2012-12-10 07:35:14
r3e02c7a9kmro120pc1j9ka840	6371	click	remValueChart	remRow	2009	2012-12-10 07:35:16
r3e02c7a9kmro120pc1j9ka840	6372	click	addValueChart	addRow	2009	2012-12-10 07:35:17
r3e02c7a9kmro120pc1j9ka840	6373	click	remValueChart	remRow	2009	2012-12-10 07:35:20
r3e02c7a9kmro120pc1j9ka840	6374	click	remValueChart	remRow	2003	2012-12-10 07:35:20
r3e02c7a9kmro120pc1j9ka840	6375	click	remValueChart	remRow	2006	2012-12-10 07:35:21
r3e02c7a9kmro120pc1j9ka840	6376	click	addValueChart	addRow	2009	2012-12-10 07:35:22
r3e02c7a9kmro120pc1j9ka840	6377	click	addValueChart	addRow	2006	2012-12-10 07:35:23
r3e02c7a9kmro120pc1j9ka840	6378	click	addValueChart	addRow	2003	2012-12-10 07:35:24
r3e02c7a9kmro120pc1j9ka840	6379	click	remValueChart	remRow	2003	2012-12-10 07:35:32
r3e02c7a9kmro120pc1j9ka840	6380	click	addValueChart	addRow	2003	2012-12-10 07:35:37
r3e02c7a9kmro120pc1j9ka840	6381	click	remValueChart	remRow	2003	2012-12-10 07:35:39
r3e02c7a9kmro120pc1j9ka840	6382	click	addValueChart	addRow	2003	2012-12-10 07:35:40
r3e02c7a9kmro120pc1j9ka840	6383	click	operation	sort	^ 08 years	2012-12-10 07:35:43
r3e02c7a9kmro120pc1j9ka840	6384	click	operation	sort	^ 09 years	2012-12-10 07:35:45
r3e02c7a9kmro120pc1j9ka840	6385	click	operation	sort	^ 07 years	2012-12-10 07:35:46
r3e02c7a9kmro120pc1j9ka840	6386	click	operation	sort	^ year	2012-12-10 07:35:47

Figura 63: Log de uso de ViSC por U07

5.4.1.1.2. Com apoio do Tableau

Entre os usuários do grupo 2, a tarefa 1 foi realizada com o apoio do Tableau Public. Os resultados obtidos estão apresentados em Figura 64, Figura 65

e Figura 66. Os tempos para a execução das tarefas foram superiores aos tempos com ViSC, em torno de 8 a 10 minutos com o uso desta.

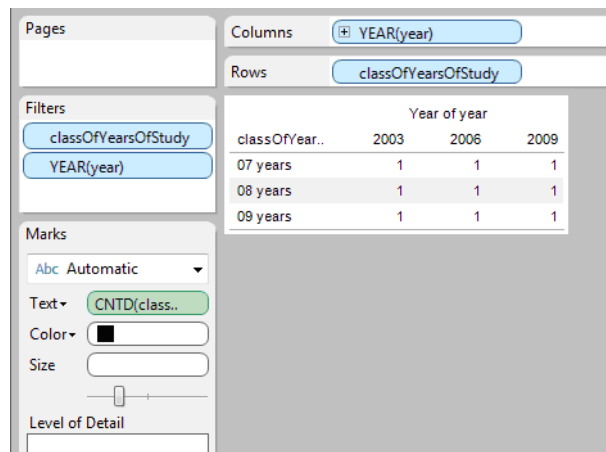


Figura 64: Resultado de U02 para a tarefa 1

A Figura 64 corresponde ao resultado de U02. Ele demonstrou dificuldade em entender os conceitos de medidas, dimensões, colunas e linhas, o que levou a uma solução com dados errados. Na tentativa de encontrar os valores, arrastou a dimensão "classOfYearsOfStudy" diversas vezes para a janela "measure" e de volta para a "dimensions". Clicou nos elementos das medidas mas não arrastou para a área de construção do gráfico e acabou gerando uma tabela com valores errados.

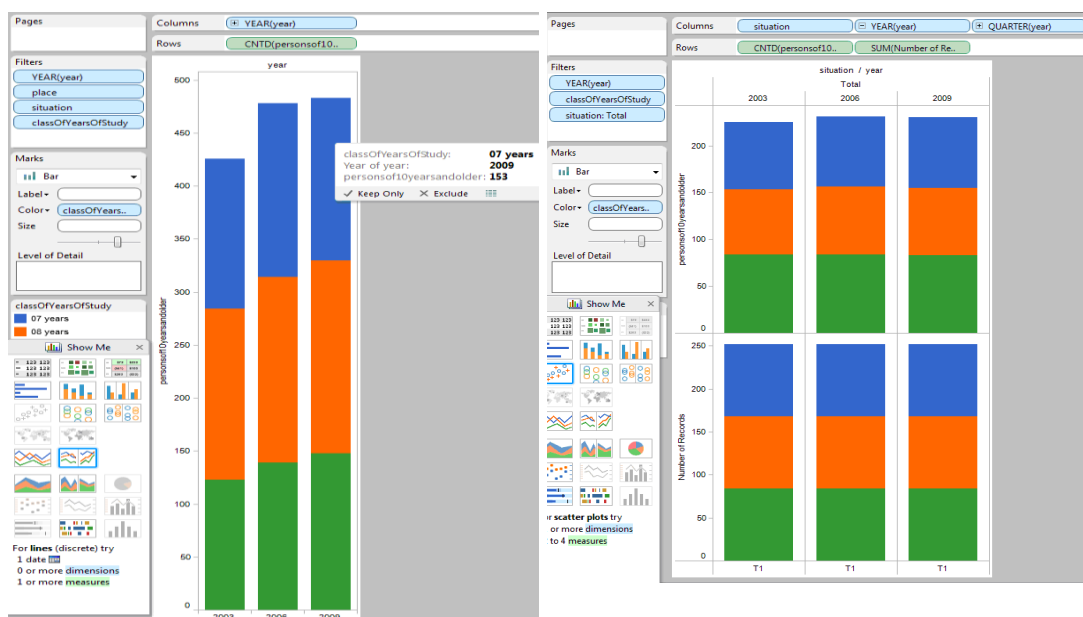


Figura 65: Resultado de U04 para a tarefa 1

Figura 66: Resultado de U08 para a tarefa 1

A Figura 65 mostra o resultado de U04. Sua dificuldade foi em entender os filtros. Demorou um tempo para criar um gráfico com dados apenas do Brasil. Gerou um gráfico correto, onde as cores representam as classes de ano de estudo e o eixo horizontal, os anos. No entanto, os dados do eixo vertical estavam errados: o dado não estava totalmente correto pois algumas classes ainda deveriam ser filtradas uma vez que a ferramenta exibe, por *default*, a soma dos valores. De acordo com os dados carregados, era necessário filtrar "*place*" pelo valor "*Brazil*", "*situation*" por "*Total*" e "*gender*" por "*Total*" para obter o valor correto no eixo vertical.

U08 também chegou no resultado de colunas agrupadas (Figura 66), porém após receber ajuda duas vezes do avaliador: a primeira vez para atentar para as medidas e a segunda, para que ele atentasse para o *feedback* do *show quick filter* que ele já tinha clicado diversas vezes e não tinha visto a janela que abria. Demonstrou muita confusão no entendimento da ferramenta, mas em relação ao dado houve compreensão. Conseguiu filtrar pela situação total mas não filtrou o local nem conseguiu retirar o segundo gráfico e nem sequer havia compreendido o significado das colunas de mesmo tamanho.

5.4.1.1.3. Dados coletados no questionário

A primeira pergunta do questionário era sobre a dificuldade encontrada na tarefa e na ferramenta utilizada pra desenvolvê-la. As respostas obtidas estão descritas na Tabela 8.

Tabela 8: Respostas sobre dificuldade encontrada na tarefa 1 e na ferramenta que apoiou a realização da tarefa.

Respostas	Grupo 1 - ViSC			Grupo 2 - Tableau		
	U03	U05	U07	U02	U04	U08
1. Muito difícil						
2. Difícil				Ta		T1/Ta
3. Regular		Vi		T1		
4. Fácil	T1/Vi	T1	T1/Vi		T1/Ta	
5. Muito fácil						

T1 - Tarefa 1 / Vi - Visc / T - Tableau

Verificamos que os usuários que realizaram a tarefa no ViSC, classificaram a tarefa como fácil assim como a ferramenta utilizada. Já os usuários que

utilizaram o Tableau, com exceção de U04, que não percebeu que haviam problemas nos dados, atribuíram uma nota mais difícil.

Nas respostas sobre a satisfação do resultado, apresentadas na Tabela 9, observamos que os usuários de ViSC ficaram satisfeitos com as respostas e os de Tableau um pouco menos.

Tabela 9: Respostas sobre a satisfação do resultado

Respostas	Grupo 1 - ViSC			Grupo 2 - Tableau		
	U03	U05	U07	U02	U04	U08
1. Não concluí a tarefa						
2. Desenhei um gráfico que não achei muito adequado mas era o que a ferramenta permitia				X		X
3. Desenhei um gráfico adequado porque a ferramenta sugeriu.		X				
4. Desenhei o gráfico igual ao que tinha em mente após analisar as opções sugeridas porque achei mais claro.	X				X	
5. Desenhei o gráfico recomendado e achei melhor.			X			

Os usuários também foram perguntados se o gráfico gerado correspondia ao que inicialmente desejavam construir. A Tabela 10 apresenta a comparação entre o que o usuário pretendia, o que fez no papel e com a ferramenta.

Tabela 10: O que o usuário pretendia construir, o que foi desenhado no papel e o resultado obtido

Grupo	Usuário	Era o gráfico que desejava construir?	Tarefa papel	Resultado obtido
Grupo 1 - ViSC	U03	Sim	Linhas	Stacked series (Linhas)
	U05	Não, foi próximo, apenas queria desenhar as colunas da mesma cor.	Linhas	<i>Stacked Columns</i>
	U07	Não, pretendia criar o somatório, não o <i>stacked</i>	Colunas mostrando apenas o somatório	Stacked Columns
Grupo 2 - Tableau	U02	Não, eu pretendia fazer um gráfico de linhas	Linhas	Tabela
	U04	Sim	Colunas mostrando apenas o somatório	Stacked Columns
	U08	Não, imaginada um gráfico com colunas paralelas (<i>clustered column chart</i>)	Linhas	Stacked Columns

Na tabela acima, vemos que, com exceção de U05 e U08, os outros quatro foram coerentes em relação ao que desenharam no papel e o que imaginaram fazer

com a ferramenta. Sobre a diferença entre os gráfico desenhado no papel e com ajuda da ferramenta, U05 justificou que era mais simples: *"Eu desenhei linha ali porque é mais fácil de desenhar do que em barra utilizando papel e lápis. (...) Imagina, eu vou desenhar as colunas ia fazer tudo torto. Se você tivesse me passado essa tarefa com a ferramenta, eu teria feito em barra, ou então pizza. (...) Quando você tem uma ferramenta que permite fazer outros tipos de gráfico, acho que é legal procurar os gráficos melhores de visualizar."* No entanto, tivemos dois resultados negativos: U02 que não conseguiu desenhar um gráfico e U08 que não demonstrou coerência no raciocínio e nem satisfação com o resultado. Ambos pertencem ao grupo 2.

Os usuários também responderam questões sobre a interação (Tabela 11) e a influência (Tabela 12) da janela Show me do Tableau na realização da tarefa 1.

Tabela 11: Respostas dos usuários do grupo 2 sobre o grau de interação com Show me

Respostas	U02	U04	U08
1. Nem vi			
2. Vi que existia mas não interagi			
3. Passei o mouse sobre as opções mas não senti necessidade de clicar			
4. Cliquei nas opções mas não entendi seu funcionamento	X		
5. Cliquei nas opções e entendi seu funcionamento		X	X

Tabela 12: Respostas dos usuários do grupo 2 sobre a influência da janela Show me

Respostas	U02	U04	U08
1. Atrapalharam muito			
2. Atrapalharam pouco			
3. Nem atrapalhou nem ajudou			
4. Ajudou um pouco	X	X	X
5. Ajudou muito			

Em resposta à pergunta sobre o grau de interação com as perguntas de ViSC, a Tabela 13 apresenta os resultados. Quando perguntados sobre a influência das perguntas, os resultados foram apresentados na Tabela 14.

Tabela 13: Respostas dos usuários do grupo 1 sobre o grau de interação com as perguntas

Respostas	U03	U05	U07
1. Nem vi			

2. Vi que existia mas não interagi			
3. Passei o mouse sobre as opções mas não senti necessidade de clicar			
4. Cliquei nas opções mas não entendi seu funcionamento			
5. Cliquei nas opções e entendi seu funcionamento	X	X	X

Tabela 14: Respostas dos usuários do grupo 1 sobre a influência das perguntas para o resultado obtido.

Respostas	U03	U05	U07
1. Atrapalharam muito			
2. Atrapalharam pouco			
3. Nem atrapalhou nem ajudou		X	
4. Ajudou um pouco			
5. Ajudou muito	X		X

5.4.1.2. Retrospective Think Aloud

Nesta etapa, para cada ferramenta, apresentamos a interpretação dos dados coletados na tarefa 1 através do *Retrospective Think Aloud*.

5.4.1.2.1. ViSC

Selecionamos as funcionalidades onde encontramos rupturas de comunicação no MIS ou problemas durante a observação e montamos a Tabela 15. Os resultados apresentados são referentes aos testes dos usuários do grupo 1 (U03,U05,U07), que realizaram a tarefa 1 com apoio de ViSC. O asterisco (*) indica que houve problemas na funcionalidade.

Tabela 15: Problemas nas Funcionalidades da ferramenta ViSC ou no sistema gráfico

	U03	U05	U07
1. Seleção do tema e variáveis			
2. Entendimento dos dados			*
3. Inclusão e remoção de valores nos eixos			*
4. Troca de dimensões			
5. Menu de gráficos			*
6. Compreensão das perguntas ou não utilização			

7. Interação com os gráficos			*
8. <i>Checkbox</i>	*		
9. Ordenação		*	*
10. Função de expandir e colapsar gráfico			

1. Seleção das variáveis

U03 selecionou as variáveis corretas. Declarou que verificou qual variável era melhor no eixo horizontal e assim escolheu classes de estudo. "*Selecionei os anos de estudo e os anos que eram pra fazer a comparação.*"

U05 também pensou sobre as variáveis que iria usar. Definiu o tema com "*Education of persons of 10 years and older*" (Educação de pessoas com 10 anos de idade ou mais de idade) e chegou à conclusão que deveria selecionar "*years*" (anos) no eixo horizontal e "*classOfyearsOfStudy*" (classe de anos de estudo), na legenda.

2. Entendimento dos dados

U07 afirmou "*Eu tava selecionado o ano da classificação das pessoas, depois qual é a idade, e aí eu peguei esse 'class of years of study' que era o que mais pareceu (...)*". A partir dessa afirmação entendemos que, apesar de ter selecionado as variáveis corretas, ele buscava a idade e encontrou classes de anos de estudo, apresentando ruptura no entendimento dos dados.

3. Inclusão e remoção dos valores

Na segunda tela, U03 verificou quais anos eram interessantes para o gráfico: "*Haviam dois anos errados. Foi distração. Tirei o ano que estava a mais e mudei o ano de 2007 para 2009.*" Como ele afirmou que foi distração, não consideramos como uma ruptura. Ele confirmou que entendeu o funcionamento.

U05 declarou que filtrou facilmente os dados que precisava: "*Aí eu fui filtrar pelos dados que eu precisava. Só pedia 2003, 2006 e 2009. Depois filtrar as classes de 7, 8 e 9 anos de estudo.*"

U07 também não encontrou problemas após a seleção dos dados: "*(...) aí depois disso, fui selecionar os dados que eu queria mostrar. No caso de cada ano e as classes de anos de escolaridade, 6, 7 e 9. Depois eu removi a classe de 4*

anos." Porém, na tentativa de ordenar, incluiu várias vezes e lembrou disso na visualização do vídeo: *"Algum momento vou ficar removendo e inserindo"*.

5. Menu de gráficos

U05 afirmou ter lembrado do menu superior e buscou o gráfico de colunas empilhadas. No entanto U07 clicou exatamente no gráfico exibido e apresentou uma ruptura, comentando: *"Aquele botão cluster column chart clicando nele, não foi."*

6. Perguntas

U03 comentou: *"Estava lendo a tarefa e resolvi ver se tinha alguma pergunta que podia me ajudar. E teve! Foi exatamente o que eu tava pensando em fazer."*

Essa afirmação mostrou que as perguntas foram importantes para encontrar a resposta. Segundo U03, ele resolveu mostrar os dois gráficos (Figura 58) porque achou que ficaria mais completo em vez de mostrar apenas a evolução. Comentou ainda: *"(...) mas esse (gráfico exibido pelo clique na pergunta) sozinho já seria o resultado certo"*, o que reforça a importância das perguntas.

Passando o *mouse* pelas perguntas, U05 afirmou: *"Vou dar uma olhada aqui, ver se tem alguma coisa dessas me ajuda a fazer um gráfico, juntando eles três."* Ele já sabia que tinha uma resposta eficiente porém, queria a soma: *"Já tinha o total mas será que tem um total consolidado? Poder juntar os três numa cor só. Eu não tava achando, de qualquer forma, esse satisfazia mas eu tava procurando talvez algum outro gráfico que ficasse melhor que ele. Mas eu já tinha uma resposta. Eu tava agora só vendo se tinha alguma coisa melhor pra complementar esse ou então colocar ele com uma cor só. Aí no final das contas acabei deixando."* Clicou no primeiro *show me* da pergunta de soma: *"E aí, esse que eu cliquei show me, era exatamente o que eu já tinha feito. Já tava lá."* Clicou na segunda opção (séries empilhadas), comparou os dois e afirmou: *"Não, esse aqui (colunas empilhadas) é melhor mesmo, vou usar esse aqui. Acho que é mais fácil ver o total."*

Sobre as perguntas U07 afirmou: *"E aí depois eu vi as opções à direita que falei: "Bom, isso aí são as opções, respostas que ele me dá, as perguntas que ele sugere então aqui deve ter a pergunta que eu quero responder. Então eu bati o*

olho ali e falei: bom, aqui são as perguntas mais comuns. Eu selecionei uma pergunta aí depois eu mudei pra sum (soma) e achei o certo."

7. Interação com os gráficos

Na tentativa de unir as cores do gráfico de colunas empilhadas, U07 afirmou que ficou pensando: *"Bom, como é que eu faço pra somar todo mundo? Eu tentei marcar eles e ver se posso somar. Ver o que podia fazer com o botão direito."*

8. Checkbox

U03 passou o mouse sobre as outras variáveis, clicou no *checkbox* da variação e afirmou que estava verificando o que seriam os elementos da interface: *"Pra eu chegar no resultado que eu esperava do gráfico, queria ver o que seria."* **O problema foi que o *checkbox* não retornou nenhum *feedback* uma vez que haviam mais de dois elementos selecionados.**

9. Ordenação

Os dois usuários que utilizaram as funcionalidade de ordenação não compreenderam seu funcionamento. U05 buscou compreender os botões de ordenação: *"Vi as diferenças do sort by (botão de ordenação). Tava tentando ver qual era a diferença de um pro outro e não tinha entendido muito bem. Na verdade, até agora não entendi mas eu sei que ano, year, era exatamente o que eu queria. Total por ano."*

U07 também precisou ordenar, uma vez que clicou na pergunta que alterou a ordenação: *"Mas eu não entendia porque ele tava botando 2003 no meio, sendo que 2003 por ordem deveria vir antes (de 2006 e 2009) e daqui eu consegui em algum momento fazer essa ordenação. Não lembro como. Aí eu tentei com o botão direito e pensei: Pô, retirei 2003. Caraca, não para de entrar e sair. Aí eu fiz alguma alteração. Ah, acho que tava 7 anos aí eu coloquei 'years' (anos) inteiro e foi direitinho. Foi no chute aquilo ali. Cliquei ali, do nada organizou. Pensei: Ótimo! Aí, beleza."*

10. Função de expandir e colapsar gráficos

Não houve comentários positivos nem negativos sobre esta funcionalidade.

5.4.1.2.2. Tableau

No Tableau, uma vez que a interface é diferente, as funcionalidades que apresentaram rupturas no MIS e problemas na observação com usuários foram diferentes das encontradas em ViSC. Os resultados apresentados são referentes aos testes dos usuários do grupo 2 (U02,U04,U08), que realizaram a tarefa 1 com apoio do Tableau. Assim, montamos a Tabela 16 para exibir pontos de problemas de comunicação com a ferramenta ou com o sistema gráfico. O asterisco (*) indica que houve problemas na funcionalidade.

Tabela 16: Problemas nas funcionalidades da ferramenta Tableau ou no sistema gráfico

	U02	U04	U08
1. Compreensão dos dados	*	*	*
2. Compreensão do filtro		*	*
3. Compreensão de dimensões e medidas	*		*
4. Compreensão de dados contínuos e discretos	*	*	*
5. Compreensão dos gráficos exibidos	*		*
6. Compreensão de linhas e colunas			*
7. Compreensão da visão geral		*	*
8. Execução da tarefa			

1. Compreensão dos dados

Percebemos que o usuário U02 teve problemas no entendimento dos dados. No início, ele mesmo afirmou que **confundiu a classe anos de escolaridade com a variável temporal ano**, criando uma tabela de linhas e colunas formada pela dimensão ano. Comentou: "*Demorei um tempinho pra perceber isso.*"

U04 também apresentou rupturas nessa funcionalidade. No gráfico de colunas empilhadas, clicou nas divisões com o botão direito: "*Eu tava observando que não era isso que eu queria. Estava aparecendo os estados (UFs) aí.*" Clicou com o botão direito sobre a medida (Números de registros de pessoas de 10 anos ou mais de idade) incluída na linha e selecionou o filtro: "*Me confundi, achei que aqui selecionava a idade. Depois que fui ver não era isso que precisava fazer.*"

U08 também demonstrou problemas na compreensão dos dados: *"Aí tava lendo que ele pedia também a situação. Eu tava tentando montar o gráfico. A situação aqui é mostrar se houve aumento dos alunos."* Situação era a dimensão que classificava a situação do domicílio em rural ou urbana. Depois demonstrou confusão na hora de excluir as classes de anos de estudo excedente, dizendo que queria a idade e mantendo classes desnecessárias: *"Aí eu queria a idade, e queria também conseguir colocar o ano. Aí eu comecei a excluir os anos porque tinham anos demais. Ela tava pedindo só 7 e 9. Aí eu to excluindo, comecei a excluir todos os anos. Eu não sabia aí se precisava o restante da informação tipo 'not determinated' e 'no statment'. Eu não sabia, até deixei mas depois eu arranquei, exclui eles. Eu deixei o total, eu achei que fosse a soma (de 7 e 9 anos), né?"*.

2. Compreensão do filtro

U02 afirmou que depois de incluir a dimensão ano na coluna e classe de ano de estudo na linha afirmou que foi meio intuitivo clicar com o botão direito. Assim encontrou o botão *"keep only"* e conseguiu filtrar pelos anos selecionados na tarefa.

U04 apresentou uma ruptura no filtro mas conseguiu corrigir: *"Estava tentando achar primeiro as unidades certas. Me enrolei muito com essa coluna 'place' (lugar) que eu coloquei porque eu achava que deveria ter. Não era o que eu queria. Eu queria o Brasil inteiro mas ele ficava dividido entre todos os estados. Demorei muito tempo pra tirar isso da visualização. O resto fui só filtrando, primeiro os anos que eu queria. Depois qual era o melhor tipo de gráfico pra poder visualizar a resposta."* Em todas as vezes que precisou filtrar alguma classe, desmarcou todas as opções que não eram necessárias. Sobre isso comentou: *"Eu fiz uma besteira, devia ter selecionado nenhum e depois só colocar os que eu queria e eu fiquei desmarcando todos. Depois de clicar em aplicar que eu olhei. Sempre faço o trabalho mais burro possível."* Sobre a sua dificuldade em unir os estados do Brasil, comentou: ***"Não acredito que levei tanto tempo pra fazer uma coisa tão simples."*** Clicou diversas vezes com o botão direito sobre a dimensão *'place'*: *"De novo estava tentando procurar o lugar (pra unir os estados)."*

U08 inicialmente conseguiu excluir anos desnecessários: *"Tava procurando na parte de cima como excluir porque eu queria só 3 anos, né? Achei e comecei a*

excluir pra selecionar só os anos necessários. Estou excluindo um por um. Deixei os três". No entanto, demonstrou problemas para filtrar a situação, dimensão que ele incluiu no gráfico. Por duas vezes ele clicou em "*quick filter*" e não viu a janela que abriu no canto direito.

3. Compreensão de dimensões e medida

Após excluir valores desnecessários, U02 ainda tinha um problema: "*Só que estava com esses 'Abcs' aí que até agora eu não sei o que é pra falar a verdade.*" Tentou encontrar os valores na janela de medidas. "*Aí eu lembrei que no vídeo ele selecionava alguma coisa ali embaixo que eram os valores. Ele clicava ali e depois no Show Me*". U02 clicou em todas as medidas e arrastou a dimensão classe de anos de estudo para a janela de medidas. "*A informação que eu queira era a classe anos de estudo. Então eu fiquei pensando que não era coluna, era informação. Aí eu joguei para informação* (se referindo à janela medidas)." Ao incluir a dimensão como uma medida, o *software* informa que está exibindo o número de registros daquela classe. Ao aparecer números na tabela, o usuário, mesmo sem entender, ficou satisfeito. "*Também não sei.* (referindo-se aos números que apareceram na tabela que ele não havia entendido). *Fiquei mexendo, apareceu e fiquei feliz.*"

U08 recebeu ajuda nesta funcionalidade para que o teste pudesse prosseguir.

4. Compreensão de dados contínuos e discretos

As medidas eram dados contínuos e as dimensões, dados discretos. No entanto, existem dois tipos de gráfico de linhas disponíveis. O primeiro trata os anos dados contínuos e o segundo como discretos. Isso era apresentado na diferença entre as cores selecionadas para as linhas e as colunas. Nenhum deles observou a diferença entre as propriedades dos dados, o que foi considerado um problema de IHC neste método.

5. Compreensão dos gráficos exibidos

Na janela *Show Me*, U02 encontrou problemas na compreensão do gráfico. Clicou no gráfico de linhas e o resultado foi uma linha reta já que os valores são todos iguais. Tentou então o gráfico de barras horizontais (Figura 67). Ficou olhando por algum tempo e afirmou "*eu achava que estava bom à princípio mas depois,*

sei lá, eu mudei. Percebi que tinham anos demais. Era só 2003, 2006 e 2009. Ainda não sabia o que estava acontecendo aqui em cima (referindo-se aos gráficos, colunas e linhas)".

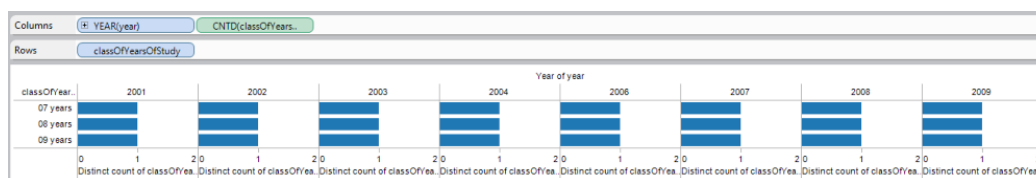


Figura 67: Gráfico de barras horizontais gerado por U02

U04 demonstrou que entendeu. Testou diversas possibilidades de gráficos e comentou: "*Só pra ver qual ficava melhor*".

U08 demonstrou que também teve dificuldades com o gráfico: "(...) *Aí eu tava tentando montar. Eu tava procurando ali como é que eu fazia. A minha intenção aí era colocar só três barrinhas. As três barrinhas que eram 2003, 2006 e 2009. Aí eu consegui incluir o gráfico (stacked bars porém com todas as barras da mesma altura porque selecionou o número de registro como medida e não a unidade) só que eu também não entendi esse gráfico. Tava tentando montar um outro de barras diferente (selecionou clustered column chart ainda com os valores citados (...))* *Aí eu lembrei de uma outra parte do vídeo (...)* *Aí já mudou o gráfico. Ainda não achei que ficou legal (ao clicar sobre 'measure values' nas medidas).* *Se eu não me engano eu mudo de novo. Depois eu volto. Aí eu tava tentando lembrar do vídeo. (...)* *Eu tava comparando os valores, tava vendo as barras de cima todas iguais. Eu não entendi o gráfico. Não entendia o que tinha sido formado apesar de ter sido eu que fiz. Eu tava procurando alguma coisa pra tentar tirar que eu não entendi essa parte aí (posicionando o mouse sobre as dimensões)".*

6. Compreensão das linhas e colunas

U08 teve problemas na manipulação dos dados no gráfico: "*Eu tava procurando como inverter (no menu superior). Eu queria colocar os anos na parte de baixo. Aí eu acabei desistindo. Aí eu voltei pro gráfico*". Esse comentário demonstrou que não entendeu que precisava inverter as linhas com as colunas.

7. Compreensão da visão geral

U04 passou um tempo sem perceber que, por padrão, as dimensões não selecionadas para o gráfico eram exibidas com a soma dos valores. Ao visualizar a parte em que exibiu o gráfico de colunas empilhadas com os 27 estados disse: "Esse (gráfico) seria o melhor mas no caso tirando o 'place'." Ao verificar a seleção pelo gráfico de colunas agregadas e ter clicado com o botão direito sobre a coluna 'place', afirmou estar procurando algo que unisse os estados: "Eu achei que ele (place) deveria estar aí mas que tivesse uma opção para juntar tudo num só lugar." Os outros não chegaram a afirmar nada sobre isso mas também não corrigiram os valores no gráfico. Para U02 não consideramos isso como um problema porque ele não chegou a inserir as medidas.

8. Execução da tarefa

Não houve problemas de desvio da tarefa na execução desta.

5.4.1.3. Etiquetagem

Nesta etapa, para cada ferramenta, apresentamos a interpretação dos dados coletados na tarefa 1 através da etiquetagem.

5.4.1.3.1. ViSC

A Tabela 17 mostra como foram classificadas as rupturas encontradas nos testes dos mesmos usuários (grupo 1 – U03, U05 e U07). Cada círculo preenchido (●) representa uma ruptura.

Tabela 17: Etiquetagem das rupturas encontradas na tarefa 1 com Tableau

		U03	U05	U07
Falhas temporárias 1. Interrompem a sessão do usuário	1. Cadê?	●	●●	●●●
	2. Ué, o que houve?		●●●●●	
	3. E agora?	●		●●
Falhas temporárias 2. Usuário percebe que seu ato comunicativo não foi bem sucedido	4. Onde estou?			
	5. Epa!			●
	6. Assim não dá			●●

Falhas temporárias 3. Usuário tenta compreender o ato comunicativo do sist.	7. O que é isso?		●●●●●●●●	●●●●●●●●
	8. Socorro			
	9. Por que não funciona?	●		●●
Falhas parciais	10. Vai de outro jeito.	●		
	11. Não obrigado			
Falhas Completas	12. Pra mim está bom.	●		
	13. Desisto.			

Como foi explicado, o usuário U03 foi o mais rápido na execução da tarefa, tendo apresentado menos rupturas de comunicação. No entanto, foi o único que apresentou rupturas mais graves: falhas parciais e completas.

A falha completa (**12. Pra mim está bom**) ocorreu devido à falta de compreensão da tarefa. Verificando que ocorreu no nível estratégico, após o usuário entregar a tarefa como resolvida oferecemos ajuda ao pedir para observar o que a tarefa realmente solicitava.

A falha parcial (**10. Vai de outro jeito**) ocorreu logo após o usuário desistir do *checkbox* de variação e buscar resposta através das perguntas.

No entanto, sete entre as nove falhas temporárias foram encontrada nos testes dos usuários:

1. Cadê? No teste de U03, esta etiqueta ocorreu uma vez no início da tarefa ao verificarmos que o usuário posicionou o *mouse* sobre todos os menus procurando verificar se os dados solicitados pela tarefa estavam corretos.

No teste de U05, ocorreu duas vezes: a primeira quando o usuário buscava o gráfico para solucionar a tarefa e a segunda após encontrar o gráfico, buscando uma solução melhor. No teste de U07, as ocorrências de "Cadê?" aconteceram no início quando o usuário procurava uma maneira de somar as três classes. Ele clicou sobre cada uma das colunas buscando alguma função que fizesse isso.

2. Ué, o que houve? Foi encontrada apenas no teste de U05. Tivemos cinco ocorrências dessa etiqueta na sua tentativa de entender o funcionamento dos

botões de ordenação. Usuário clicou em todos os botões, voltando ao botão inicialmente selecionado buscando entender a diferença entre eles.

3. E agora? No teste de U03 classificamos com essa etiqueta quando o usuário, ao encontrar a pergunta e o gráfico correto, ficou na dúvida entre entregar apenas o segundo ou ambos. Ele rapidamente optou por entregar os dois.

No teste de U07 essa etiqueta ocorreu após a seleção dos elementos do gráfico e o usuário parou pra entender como iria juntar as colunas. Depois, ao montar o gráfico, ele pensa um pouco até salvá-lo como resposta.

5. Epa! No teste de U07 ocorreu durante a tentativa de inverter a ordem das linhas que compunham o gráfico. Após a reordenação através do clique na pergunta de ordenação, usuário tentou três vezes retirar os elementos e incluir novamente. A primeira vez pareceu que ele não estava muito convicto de como havia inserido consideramos como "Epa!".

6. Assim não dá. Ocorreu apenas no teste de U07, sua primeira ocorrência foi verificada após o usuário compreender que não era através da seleção das colunas que iria juntá-las. A segunda vez ocorreu quando ele percebeu que não adiantava retirar os elementos do gráfico e inserir em outra ordem que não mudaria a ordem dos elementos no gráfico.

7. O que é isso? Para U05, esta etiqueta apareceu pela primeira vez na seleção das variáveis. O usuário abriu a caixa de seleção para ler as variáveis disponíveis para o eixo horizontal. Depois, ocorreu mais duas vezes quando buscava o gráfico solicitado. Primeiro abriu o gráfico de colunas múltiplas e depois o gráfico de colunas empilhadas. Após encontrar, ainda quis verificar se encontrava um gráfico melhor e abriu o gráfico de séries múltiplas e depois séries, evidenciando mais duas ocorrências desse etiqueta. Verificou ainda o que era o botão de ordenação e as duas visualizações apresentadas para a pergunta que ele buscava.

Para U07, esta etiqueta apareceu seis vezes sendo que em dois momentos. O primeiro momento foi na busca do gráfico, ao passar o *mouse* sobre alguns links sem clicar (gráfico de colunas agrupadas no topo da página e depois no link "*show me*" de uma das perguntas) até clicar no gráfico de colunas agrupadas da pergunta

da tarefa ordenação. Depois quando tentava entender a funcionalidade dos botões de ordenação.

9. Por que não funciona? No teste de U03 encontramos apenas a segunda quando o usuário clicou no *checkbox* de variação, que não retornou nenhum *feedback* ao usuário, uma vez que tinha 3 anos e 3 classes de anos de estudo selecionados e a operação de variação só é possível quando uma das dimensões possui apenas 2 elementos selecionados.

As duas repetições mal sucedidas de U07 na tentativa de reordenar as linhas que compunham o gráfico também foram classificadas com essa etiqueta.

5.4.1.3.2. Tableau

A Tabela 18 mostra como foram classificadas as rupturas encontradas nos testes dos mesmos usuários (grupo 2 – U02, U04 e U08). Cada círculo preenchido (●) representa uma ruptura.

Tabela 18: Etiquetagem das rupturas encontradas na tarefa 1 com Tableau Public

		U02	U04	U08
Falhas temporárias 1. Interrompem a semiose do usuário	1. Cadê?		●●●●●●●● ●	●●●●●●●● ●●●●●●●● ●●●●
	2. Ué, o que houve?	●		
	3. E agora?	●●●●●●●● ●●●●●	●●	●●●●●●●●
Falhas temporárias 2. Usuário percebe que seu ato comunicativo não foi bem sucedido	4. Onde estou?			
	5. Epa!		●●	●●●●●●
	6. Assim não dá	●●●●	●●●●	●●●●●●
Falhas temporárias 3. Usuário tenta compreender o ato comunicativo do sist.	7. O que é isso?	●●●●●●●● ●●●●●●●● ●●●●●	●●●●●●●● ●●●●●●●● ●	●●●●●●●● ●●●●●●●● ●●●●
	8. Socorro			
	9. Por que não funciona?			

Falhas parciais	10. Vai de outro jeito.	●●	●	
	11. Não obrigado			
Falhas Completas	12. Pra mim está bom.	●		●
	13. Desisto.			

Foram encontradas apenas nove das treze etiquetas.

1. Cadê? No teste de U04, após a rápida seleção dos elementos do gráfico, a maior dificuldade do usuário foi no **filtro dos lugares**. Ele deveria escolher somente o Brasil. Essa etiqueta ocorreu basicamente na sua busca por unir todos os estados. Ele não viu que existia um elemento Brasil entre os lugares disponíveis e inferiu que os valores disponíveis eram apenas as Unidades da Federação (UFs) e que deveria somar todos. Buscou no menu de lugares, em seus filtros, no filtros do elemento de medida no filtro da classe de anos de estudo e por algum tempo passando o *mouse* pelos elementos da tela.

O usuário U08, no entanto apresentou 15 repetições dessa etiqueta. Sete ocorrências aconteceram devido à **falta de medidas ao gráfico** o que fez com que ele buscasse em todos os itens do menu superior e no botão direito dos elementos algo que pudesse alterar os valores das barras. Já tendo conseguido incluir valores de medidas, sua busca foi como retirar as colunas que estavam a mais (no gráfico) e o segundo gráfico. Ele conseguiu retirar as colunas mas entregou o resultado com dois gráficos, sem observar que havia adicionado o número de registro como medida.

2. Ué, o que houve? No teste de U02 foi encontrada apenas uma vez quando o sistema, automaticamente incluiu a **medida** "*classOfYearsOfStudy*" na marca "*text*" (texto) após uma interação do usuário e substituiu as *strings* "*Abc*" (Figura 68) pelo número de registro da classe selecionada (Figura 69 e Figura 70). O usuário não entendeu o que estava acontecendo mas não buscou mais explicações o que levou à falha completa com a entrega da tarefa com dados errados.

classOfYear..	2001	2002	2003	2004	2006	2007	2008	2009
07 years	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
08 years	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
09 years	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc

Figura 68: Passo 1 da interação de U02

		Year of year							
		2001	2002	2003	2004	2006	2007	2008	2009
Drop		3	3	3	3	3	3	3	3

Figura 69: Passo 2 da interação de U02

		Year of year							
		2001	2002	2003	2004	2006	2007	2008	2009
classOfYear..									
07 years		1	1	1	1	1	1	1	1
08 years		1	1	1	1	1	1	1	1
09 years		1	1	1	1	1	1	1	1

Figura 70: Passo 3 da interação de U02

3. E agora? Uma falha temporária recorrente no teste de U02, tendo ocorrido 10 vezes. A cada interação o usuário permanecia um tempo sem realizar nenhuma ação, parecendo estar formulando novos planos de ação.

U04 verificou que escolheu um gráfico correto (colunas empilhadas), parou para analisar a próxima ação, que seria **como unir as UFs do Brasil**. A segunda vez, após diversas tentativas de unir as UFs, passou o *mouse* pela tela a procura de alguma dica de como continuar.

U08, assim como U02, também ficou diversas vezes perdidos após realizar determinadas ações.

5. Epa! No teste de U02, classificamos essa etiqueta na divisão dos anos por trimestres (na coluna) e no clique errado dos anos que deveria ser filtrado.

No teste de U08 ocorreram mais vezes. Logo no início quando ele clicou no nome do banco de dados duas vezes sem obter retorno no sistema. Depois, quando o usuário verificou que, ao clicar no ícone da tabela no menu "Show me" que sumira uma das dimensões incluídas.

6. Assim não dá. Após a compreensão dos gráficos e a exibição de números na tabela por U02, passamos a considerar "Assim não dá." o clique nos gráficos já visualizados com esta etiqueta. Nesse momento o usuário estava buscando uma solução para a entrega da tarefa e ao ver um gráfico inadequado, selecionava outro. O usuário queria construir um gráfico e testou as opções, descartando as que não achou adequadas.

U04 foi etiquetado dessa forma quando não encontrou o botão de soma das UFs do Brasil nos filtros onde procurou e decidiu fechar. No teste de U08, consideramos esta etiqueta quando verificamos a seleção de um gráfico, realizada algumas operações e depois clicava *control+z* voltando para o gráfico anterior. Fez isso diversas vezes.

7. O que é isto? Encontramos 16 repetições no teste de U2. Considerando que a ferramenta era novidade para o usuário, essas falhas temporárias não foram consideradas problemas graves porque ele estava tentando compreender o sistema. Utilizamos essa etiqueta tanto nas buscas por explicações ao posicionar o *mouse* sobre um elemento quanto nos cliques em objetos que para entender o sistema através do *feedback* exibido com a interação.

Como exemplos dessa etiqueta no teste de U02, podemos citar a busca pela compreensão do preenchimento com "Abc" na tabela, das medidas e dos gráficos desabilitados e habilitados. Na compreensão dos gráficos habilitados quando o *feedback* da interação trazia as explicações para o usuário.

No teste de U04, essa etiqueta foi verificada ao posicionar o *mouse* sobre os elementos do gráfico para verificar o que ele estava mostrando em cada alteração de gráfico ou de elementos selecionados na coluna. Também foi verificada na janela de filtro de medidas para entender o seu funcionamento. E, por último, diversas vezes quando o usuário posicionou o *mouse* sobre os gráficos da janela "Show me" ou clicou para visualizar o gráfico que seria gerado.

No teste de U08, o maior índice desta etiqueta foi no posicionamento do *mouse* sobre as dimensões.

As falhas parciais foram representadas pela etiqueta:

10. Vai de outro jeito. U02 não entendeu porque o sistema retornava um gráfico de linhas ao incluir a medida "*classOfYearOfStudy*" na janela de coluna e moveu novamente a classe para a janela de dimensões, conseguindo obter uma tabela.

U04, no entanto, não encontrou como unir as UFs do Brasil e decidiu remover a coluna "*place*" do gráfico, encontrando uma resposta mais satisfatória e salvando o resultado.

Houve uma ruptura grave nos testes dos usuários U02 e U08:

12. Pra mim está bom. No teste de U02, esta etiqueta aconteceu quando ele decidiu entregar uma tabela como resultado da tarefa 1 que solicitava a construção de um gráfico. U08 também decidiu entregar o resultado sem conseguir retirar o segundo gráfico. Como ambos declaram não ter ficado satisfeitos com o resultado entregue mas era o que a ferramenta permitia interpretamos com essa etiqueta.

5.4.1.4. Discussão sobre a tarefa 1

Com a etiquetagem, verificamos que ambas as ferramentas apresentaram as mesmas nove rupturas de comunicação.

No Tableau, a interpretação através da etiquetagem possibilitou encontrar as rupturas mais comuns de comunicação e as que levaram a rupturas mais graves. Foram estas: **a dificuldade em perceber a diferença entre dimensão e medida e a dificuldade em perceber que uma dimensão não incluída no gráfico, por padrão, era incluída com a soma dos valores o que dava uma visão geral dos dados. Como U08 recebeu ajuda para prosseguir, podemos considerar que ambos, U02 e U08 não perceberam a diferença entre medida e dimensão e, não conseguiram formular hipótese sobre o uso da ferramenta**, ou seja, não conseguiram entender a ferramenta e por isso não conseguiram construir um gráfico adequado. Eles têm consciência de que não fizeram o gráfico mais eficiente mas consideram a tarefa finalizada. Isso é verificado na resposta ao quesito sobre a satisfação. Suas respostas foram: desenharam um gráfico não muito adequado porque a ferramenta não permitia. Isso foi evidenciado também pela etiqueta encontrada como "*Pra mim está bom*" e não "*Desisto*".

A ruptura na comunicação das dimensões e medidas acabou levando U02 a entregar um resultado não adequado, o que foi confirmado com sua resposta sobre a satisfação do gráfico (Figura 52) e em entrevista "*entendi bem o 'drag and drop' mas não entendi bem o significado daquilo. O que era coluna, o que era informação. Ficou um pouco confuso isso.*" Ele chamou dimensão de coluna e medida de informação. Para U4 a dificuldade maior foram os filtros e a diferença entre dimensões e medidas: "*Nunca tinha visto algo parecido então comecei a me enrolar um pouco com os filtros das dimensões e as medidas. Não sabia exatamente qual que tinha que usar*". Ele achou os filtros com facilidade só não percebeu que para somar as UFs bastava retirá-las da seleção. Isso foi percebido

no final. No caso, sua resposta ficou errada porque os dados continham além das UFs, os dados consolidados do Brasil, o que quando retirado da seleção duplicou os valores. U08 apresentou rupturas de comunicação tanto na compreensão das medidas quanto na visualização dos filtros. Para que pudéssemos dar prosseguimento ao teste, ele recebeu duas ajudas e, ainda assim, entregou um resultado ineficiente.

Pelos comentários dos usuários durante o *Retrospective Think Aloud* verificamos ainda problemas no entendimento dos dados. Observamos que **U02 e U08 não tinham qualquer planos de ação formulado para resolver a tarefa. Em diversos momentos afirmaram que buscavam lembrar o que tinha sido feito no vídeo apresentado. O único comentário positivo foi de U04 quando disse que não sabia como tinha gastado tanto tempo com algo tão simples.**

Apesar de U04 e U08 terem afirmado que entenderam seu funcionamento, na verdade só haviam compreendido que era uma paleta de gráficos. Suas respostas na entrevista confirmaram isso. Ao ser perguntado sobre o entendimento, U04 respondeu: "*A única coisa que eu queria que ela fizesse era justamente montar o gráfico que você quer de acordo com os valores que você já tem aqui na coluna e na linha. Nem reparei que ela desabilita mas tinha que ter alguma coisa aqui na esquerda (referindo-se às medidas) pra você poder gerar o gráfico. Não reparei essa parte de baixo não. Ah, sim! Essa parte do mapa que tinha que ter alguma latitude e longitude sim, mas a parte de baixo não.*" A parte de baixo a qual ele se refere e foi perguntado é a explicação das características das variáveis para tornar uma visualização habilitada que mostra, por exemplo, quantos dados discretos e contínuos devem estar selecionados para a opção estar ativa.

No ViSC, com observamos que a **falha parcial ocorrida foi decorrente a um problema que deve ser corrigido na ferramenta: a seleção do campo de variação (checkbox) não demonstrou nenhum feedback.** A ferramenta deve retornar um *feedback* em qualquer situação. Vimos ainda que **muitas rupturas ocorreram porque os usuários buscavam uma funcionalidade que a ferramenta ainda não permitia.** No caso do cenário utilizado: a soma para unir as cores das colunas empilhadas por U05 e U07. A soma, assim como a média deverão ser incluídas como projetos futuros para a ferramenta. **A ordenação também foi mal compreendida pelos usuários que a utilizaram.** U07 tentou

reordenar as variáveis através da remoção e inclusão dos valores em uma nova ordem e os botões não emitiram a mensagem de forma clara.

Esse método possibilitou encontrar também outros problemas de IHC e funcionalidades que ajudaram o usuário. **Entre os problemas, vimos: falta de entendimento dos dados mesmo com a explicação no início do teste; a necessidade de relacionar os gráficos exibidos na tela com o item no menu de gráficos para não deixar dúvidas para o usuário e; uma curiosidade de que mesmo sendo um programa online, o usuário U07 declarou que tentou encontrar a funcionalidade através do clique no botão direito. Entre as funcionalidade que ajudaram, podemos observar na Tabela 15 que não encontramos rupturas em relação à utilização das perguntas e que os comentários foram consistentes e demonstraram que elas tiveram influência importante e positiva sobre os resultados.**

Na entrevista pedimos para que eles explicassem o entendimento das perguntas. U03 explicou seu entendimento na entrevista da seguinte forma: *"Sinceramente não sei como elas são feitas mas ela parece que vê a informação disponível pra você e sugere algum cruzamento de informação. Tal coisa pode ser interessante. Ou baseado no gráfico ou baseado somente nas informações. Isso pode ser legal de você querer mostrar"*.

U05 explicou: *"(...) Então eu não me toquei a princípio que aquilo lá já era um facilitador. Com essas variáveis aqui, você pode desenhar esses tipos de gráficos aqui. Eu só fui perceber isso depois, tanto é que eu cheguei a clicar no gráfico que eu já tinha feito."* U07 afirmou: *"Entendimento! Bom eu sabia que ele ia mostrar as perguntas relativas ao que eu tinha, tipo o que seria mais como ser extraído dali. Então foi isso que me ajudou. Elas mudavam, mas eu não lembro quando aconteceu essas alterações. Acho que quando eu mudava qual era o domínio, a imagem, ali, tinham essas alterações."*

Quando perguntados se haviam observado as estrelas que pontuavam as visualizações, U07 afirmou: *"Isso! Isso daí também influenciou. Tipo bati o olho primeiro por exemplo na 5 estrelas. Bom, é isso aqui o que eu quero? Dependendo da descriçãozinha ali."* Para U05 isso não foi perceptível: *"Vi mas eu não cheguei a entender o que que tinha a estrelinha ou não. Assim, de cara, né? Foi rápido meu contato."*

Segundo U03 as perguntas foram cruciais para encontrar a resposta. A resposta dada por U03 sobre a satisfação com o resultado é consistente com a explicação dada durante o *Restrospective Think Aloud* que mostra que ele realmente analisou as opções: "(...) *resolvi mostrar os dois gráficos (Figura 58) que eu achei que ficaria mais completo em vez de mostrar apenas a evolução, mas esse sozinho já seria o resultado certo.*" **Para U05 e U07 sua importância foi para conferir que o gráfico realizado estava correto.** Apesar de U05 e U07 terem desenvolvido a mesma linha de raciocínio, pretendendo construir um gráfico que a ferramenta não permitia (com o total da soma), U07 preferiu a solução recomendada. **Para U07, além de agilizar também retornou o resultado.** Segundo ele elas ajudaram da seguinte forma: "*Primeiro que ele já basicamente fazia meu trabalho. Tipo: Meu trabalho era responder aquela pergunta: Quantas pessoas blá blá blá? E ele ali já dava a pergunta. Então, falei: Bom, você clica aqui, ele já gera o que eu preciso, já faz o treinamento todo necessário. Só preciso da pergunta.*" **Podemos supor aqui que a ferramenta pode ter lhe trazido aprendizado, uma vez que lhe trouxe uma visualização nova.** U05, no entanto, considerou as perguntas neutras para o resultado. Avaliamos que elas tiveram importância na obtenção do resultado uma vez que, ao encontrar um gráfico que já tinha obtido parou de buscar a soma consolidada que não era possível na ferramenta, o que agilizou a realização da tarefa. Ele afirmou em entrevista que as perguntas foram "*importantes sim, fundamentais não*". Segundo ele "*meu primeiro ímpeto seria sempre de tentar desenhar o gráfico*" e sentiu falta das miniaturas de gráficos no ViSC conforme a existência no Tableau. Além disso, ao ser perguntado sobre a outra alternativa recomendada, ele afirmou: "*Achei interessante até porque eu não sei se eu ia pensar em desenhar aquele de linha.*" **Essa afirmação de U05 mostra que pode ter havido aprendizado para ele também.**

Na tarefa em questão, vimos que as perguntas influenciaram estimularam a análise dos dados, podendo ter trazido aprendizado para U05 e U07. Os 3 usuários de ViSC interagiram e entenderam qual era a sua proposta. Apesar de como isso ser feito não tenha ficado claro, eles observaram mudanças e consideraram fundamental para o resultado apresentado e a maior facilidade da ferramenta. No Tableau, apesar da ferramenta permitir mais interação e ser aparentemente atrativa

para os usuários, as respostas foram consideradas menos satisfatórias e a ferramenta mais difícil.

5.4.2. Tarefa 2

Esta seção apresenta os dados coletados na etapa de observação da tarefa 2 — *Observar a média das notas no exame do PISA do Brasil, Canadá e Austrália em 2003 e em 2009 e identificar o país que teve maior ascensão* — e a interpretação realizada com estes dados.

5.4.2.1. Observação de uso

Nesta etapa apresentamos os fatos anotados durante a observação, os resultados entregues por cada usuário e os dados preenchidos nos questionários na tarefa 2.

5.4.2.1.1. Com o apoio de ViSC

Entre os usuários do grupo 2, que realizaram com o apoio de ViSC, obtivemos os resultados esperados, conforme exibidos nas figuras abaixo (Figura 71, Figura 72 e Figura 73).

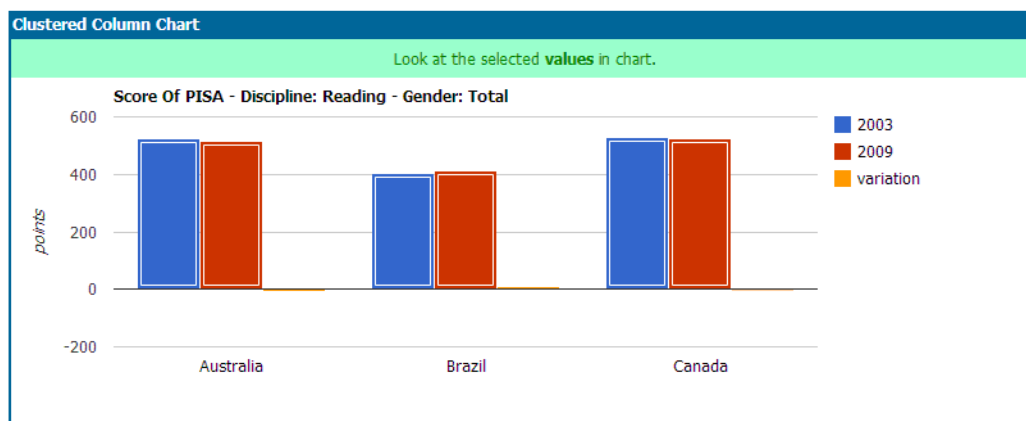


Figura 71: Resultado de U02 para a tarefa 2

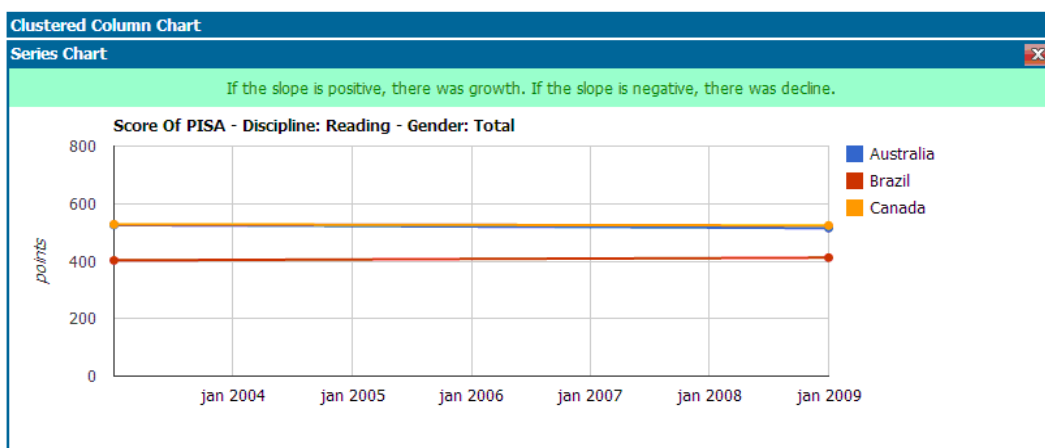


Figura 72: Resultado de U04 para a tarefa 2

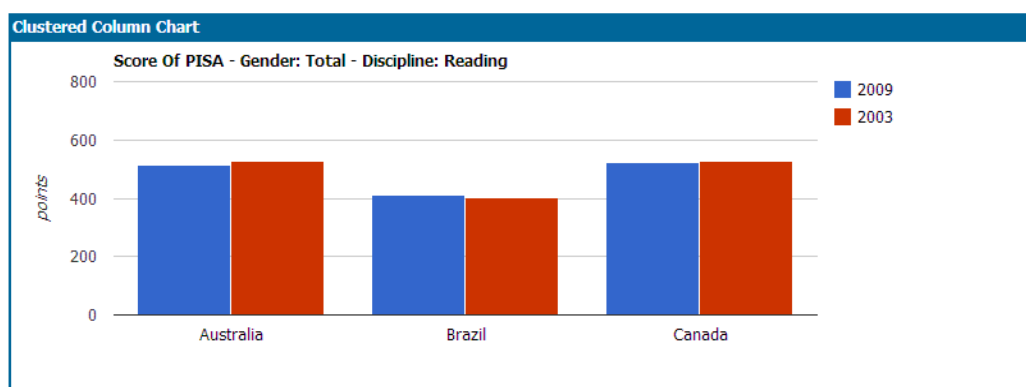


Figura 73: Resultado de U08 para a tarefa 2

Em VISC, U02 chegou a abrir a visualização mais pontuada, o gráfico de dispersão, através do clique na pergunta correta (cod. 38_7 da Figura 74) mas fechou sem ter lido a pergunta clicada nem a resposta exibida na faixa acima do gráfico. Disse que clicou porque buscava um gráfico de linhas em inglês, mas também não usou o menu superior. Acabou chegando no resultado correto de menor pontuação, gráfico de colunas agregadas, através do clique em uma pergunta aleatória que exibiu esse gráfico.. O clique na pergunta foi uma forma de abrir um gráfico quando estavam fechados. Ele usou as perguntas para navegar entre as opções de visualização sem lê-las.

oukg713m5r88epekvmcjhagbd0	5850	click	back			2012-11-28 04:17:57
oukg713m5r88epekvmcjhagbd0	5851	click	addValueChart	addRow	2003	2012-11-28 04:18:03
oukg713m5r88epekvmcjhagbd0	5852	click	addValueChart	addRow	2009	2012-11-28 04:18:04
oukg713m5r88epekvmcjhagbd0	5853	click	addValueChart	addRow	2006	2012-11-28 04:18:05
oukg713m5r88epekvmcjhagbd0	5854	click	addValueChart	addCol	Canada	2012-11-28 04:18:14
oukg713m5r88epekvmcjhagbd0	5855	click	addValueChart	addCol	Australia	2012-11-28 04:18:15
oukg713m5r88epekvmcjhagbd0	5856	click	remValueChart	remRow	2006	2012-11-28 04:18:28
oukg713m5r88epekvmcjhagbd0	5857	click	showme	38_7	What is the diferen	2012-11-28 04:19:19
oukg713m5r88epekvmcjhagbd0	5858	click	showme	2_1	Which values are above	2012-11-28 04:20:44
oukg713m5r88epekvmcjhagbd0	5859	click	operation	variation	1	2012-11-28 04:21:08

Figura 74: Log de uso de U02 no ViSC

U04 selecionou com facilidade as variáveis e elementos do cenário. Observou o gráfico gerado e optou pela questão que exibía a evolução ou crescimento de um dado (cod. 33_4 na Figura 75), o que exibiu o gráfico de série. Apesar de verificar que a linha não tinha inclinação suficiente para determinar a maior ascensão, observou os números posicionando a *mouse* sobre os pontos e identificou que foi o Brasil. Respondeu a questão com o gráfico de série e com a seguinte observação por escrito: "*Brasil foi o que obteve maior ascensão*".

rvhe06hign89jmotrvglftk3k0	6090	click	select	theme	pisa	2012-12-05 03:07:20
rvhe06hign89jmotrvglftk3k0	6091	click	select	independentVar2	year	2012-12-05 03:07:43
rvhe06hign89jmotrvglftk3k0	6092	click	send			2012-12-05 03:07:45
rvhe06hign89jmotrvglftk3k0	6093	click	back			2012-12-05 03:07:46
rvhe06hign89jmotrvglftk3k0	6094	click	addValueChart	addCol	2003	2012-12-05 03:07:52
rvhe06hign89jmotrvglftk3k0	6095	click	addValueChart	addRow	Brazil	2012-12-05 03:07:57
rvhe06hign89jmotrvglftk3k0	6096	click	addValueChart	addRow	Canada	2012-12-05 03:07:58
rvhe06hign89jmotrvglftk3k0	6097	click	addValueChart	addRow	Australia	2012-12-05 03:08:01
rvhe06hign89jmotrvglftk3k0	6098	click	showme	33_4	Was there growth or	2012-12-05 03:08:34

Figura 75: Log de uso de U04 no ViSC

U08 selecionou inicialmente os dados errados e apresentou dificuldade em alterar as variáveis. Como os anos estavam em uma dimensão que não possibilitava a inclusão de múltiplos valores, substituiu várias vezes até verificar o *combobox* com o valor *insert*. Quando o usuário parecia não ter entendido o alerta exibido no clique desta funcionalidade, o avaliador ofereceu ajuda, perguntando se ele havia entendido a mensagem. Como a resposta foi negativa, o avaliador leu a mensagem de forma traduzida para que ele compreendesse o texto e desse prosseguimento ao teste. Ele selecionou o ano que faltava e entregou o resultado sem clicar em pergunta alguma.

id_session	order	event	category	id_value	value	date_time
47v5pnkbn2ouaj0cq8pude9i155	6470	click	back			2012-12-13 07:56:40
47v5pnkbn2ouaj0cq8pude9i155	6471	click	addValueChart	addRow	Brazil	2012-12-13 07:56:47
47v5pnkbn2ouaj0cq8pude9i155	6472	click	addValueChart	addRow	Canada	2012-12-13 07:56:54
47v5pnkbn2ouaj0cq8pude9i155	6473	click	addValueChart	addRow	Australia	2012-12-13 07:56:58
47v5pnkbn2ouaj0cq8pude9i155	6474	click	addValueChart	addCol	Mathematics	2012-12-13 07:57:04
47v5pnkbn2ouaj0cq8pude9i155	6475	click	addValueChart	addCol	Science	2012-12-13 07:57:17
47v5pnkbn2ouaj0cq8pude9i155	6476	click	switchTitle	remove	2009	2012-12-13 07:57:34
47v5pnkbn2ouaj0cq8pude9i155	6477	click	switchTitle	insert	2006	2012-12-13 07:57:35
47v5pnkbn2ouaj0cq8pude9i155	6478	click	switchTitle	remove	2006	2012-12-13 07:57:36
47v5pnkbn2ouaj0cq8pude9i155	6479	click	switchTitle	insert	2003	2012-12-13 07:57:36
47v5pnkbn2ouaj0cq8pude9i155	6480	click	switchTitle	remove	2003	2012-12-13 07:57:37
47v5pnkbn2ouaj0cq8pude9i155	6481	click	switchTitle	insert	2009	2012-12-13 07:57:37
47v5pnkbn2ouaj0cq8pude9i155	6482	click	switchTitle	remove	2009	2012-12-13 07:57:42
47v5pnkbn2ouaj0cq8pude9i155	6483	click	switchTitle	insert	2003	2012-12-13 07:57:42
47v5pnkbn2ouaj0cq8pude9i155	6484	click	switchTitle	remove	2003	2012-12-13 07:57:44
47v5pnkbn2ouaj0cq8pude9i155	6485	click	switchTitle	insert	2009	2012-12-13 07:57:44
47v5pnkbn2ouaj0cq8pude9i155	6486	click	switchTitle	insert	2003	2012-12-13 07:58:18
47v5pnkbn2ouaj0cq8pude9i155	6487	click	switchTitle	remove	2009	2012-12-13 07:58:18
47v5pnkbn2ouaj0cq8pude9i155	6488	click	switchTitle	insert	2009	2012-12-13 07:58:39
47v5pnkbn2ouaj0cq8pude9i155	6489	click	switchTitle	remove	2003	2012-12-13 07:58:39
47v5pnkbn2ouaj0cq8pude9i155	6490	click	remValueChart	remCol	Reading	2012-12-13 07:59:40
47v5pnkbn2ouaj0cq8pude9i155	6491	click	remValueChart	remCol	Science	2012-12-13 07:59:43
47v5pnkbn2ouaj0cq8pude9i155	6492	click	addValueChart	addCol	Reading	2012-12-13 07:59:50
47v5pnkbn2ouaj0cq8pude9i155	6493	click	remValueChart	remCol	Mathematics	2012-12-13 07:59:51
47v5pnkbn2ouaj0cq8pude9i155	6494	click	switchVars	remove	discipline	2012-12-13 07:59:59
47v5pnkbn2ouaj0cq8pude9i155	6495	click	switchTitle	remove	2003	2012-12-13 07:59:59
47v5pnkbn2ouaj0cq8pude9i155	6498	click	switchTitle	remove	2003	2012-12-13 08:00:00
47v5pnkbn2ouaj0cq8pude9i155	6499	click	switchVars	insert	year	2012-12-13 08:00:00
47v5pnkbn2ouaj0cq8pude9i155	6500	click	switchVars	remove	discipline	2012-12-13 08:00:00
47v5pnkbn2ouaj0cq8pude9i155	6501	click	switchTitle	insert	2009	2012-12-13 08:00:00

Figura 76: Log de uso de U08 no ViSC

5.4.2.1.2. Com o apoio do Tableau

A tarefa 2 foi realizada com o apoio do Tableau Public pelos usuários do grupo 1. Os resultados obtidos estão apresentados na Figura 77, na Figura 78 e na Figura 79.

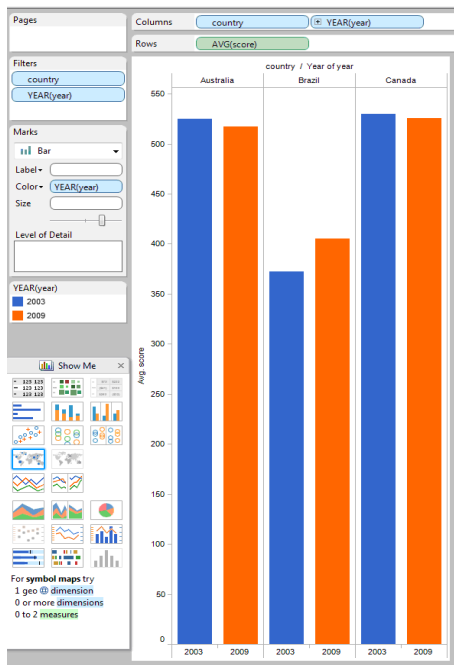


Figura 77: Resultado de U03

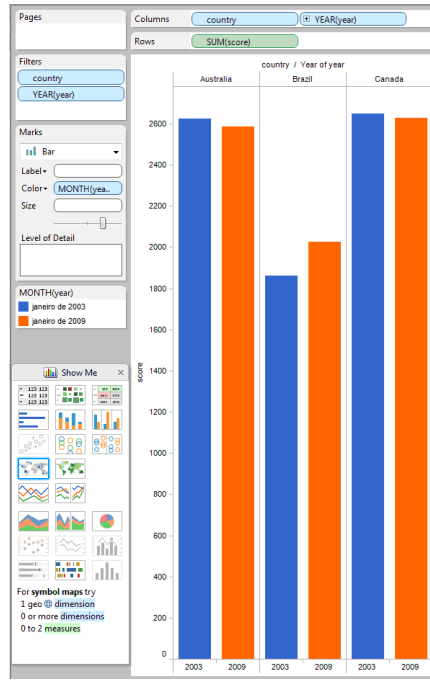


Figura 78: Resultado de U05

A diferença entre os resultados de U03 e U05 foi a forma de agregação dos dados. O primeiro entendeu, após receber ajuda, que, por *default*, a ferramenta exibia a soma dos pontos e conseguiu alterar para média — $AVG(score)$ — conforme a Figura 77, enquanto o segundo manteve os dados agregados pela soma — $SUM(score)$, conforme Figura 78

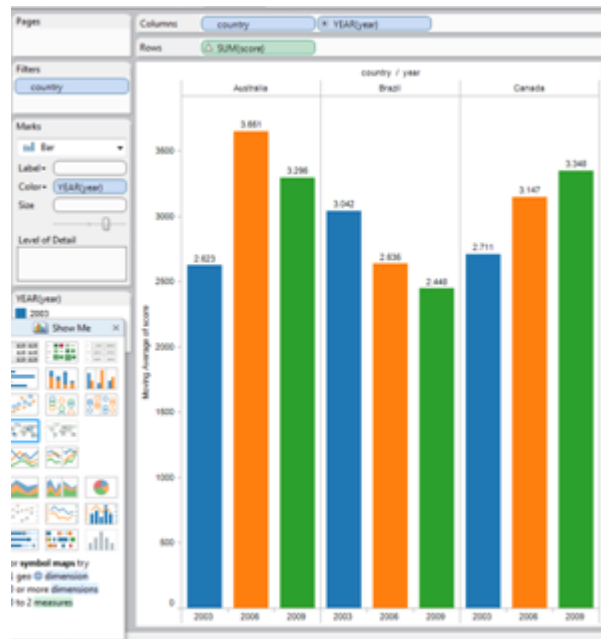


Figura 79: Resultado de U07

Inicialmente, U03 teve dificuldade em encontrar a variável de medida e depois teve problemas em encontrar um gráfico que procurava já que estavam desativados. Com dificuldade de encontrar como remover elementos excessivos do eixo horizontal (filtrar), buscou outras opções de visualização. Desviou da tarefa na tentativa de entender melhor a ferramenta. Ao entregar o resultado, o avaliador perguntou se ele ia entregar com a soma mesmo e U03 buscou rapidamente a média dos valores.

U05 teve uma certa dificuldade em encontrar os filtros, mas sua maior dificuldade foi em entender porque os valores eram exibidos com "Abc", ou seja entender que faltou incluir uma medida. Em relação aos filtros, na primeira vez que precisou, encontrou. No entanto, na segunda vez, já não sabia mais onde estavam. Após realização das filtragens, como o usuário não incluiu uma medida para formar o gráfico, a única opção disponível, além da tabela, era o gráfico de Gantt, o que fez com que se perdesse ainda mais. Esta opção mostrou uma janela diferente para os filtros porque transformou a dimensão "year" para um dado contínuo (compare a janela da Figura 50 e a da Figura 51) e incluiu os anos que já haviam sido retirados pelo usuário na linha de tempo. **Após receber ajuda para dar prosseguimento ao teste, U05 foi lembrado das medidas e incluiu o número de pontos (score) no gráfico, chegando a um gráfico correto porém sem filtrar leitura para a dimensão disciplina, total para a dimensão sexo e mantendo a soma dos pontos em vez da média.**

U07 não teve dificuldades em filtrar nem de entender a diferença entre dimensões e medidas. Selecionou o gráfico de colunas agregadas e ainda incrementou incluindo o valor sobre as colunas. No entanto, manteve a coluna de 2006 que atrapalhava a visualização e não entendeu como trocar a soma dos valores pela média. Acabou selecionando "moving average" em vez de *average* o que gerou dados errados, que é um cálculo normalmente usado para suavizar flutuações a curto prazo nos dados para se verificar tendências de longo prazo.²⁴

²⁴ A explicação sobre *moving average* pode ser encontrada no *help on line* do Tableau na página: http://onlinehelp.tableausoftware.com/v7.0/public/online/en-us/calculations_tablecalculations_definebasic_moving.html

5.4.2.1.3. Dados coletados no questionário

Sobre a dificuldade encontrada na tarefa e na ferramenta utilizada pra desenvolvê-la, as respostas foram agrupadas na Tabela 19.

Tabela 19: Respostas sobre dificuldade encontrada na tarefa 2 e na ferramenta que apoiou a realização da tarefa.

Respostas	Grupo 1			Grupo 2		
	U03	U05	U07	U02	U04	U08
1. Muito difícil						
2. Difícil		Ta	Ta			
3. Regular	T2/Ta			T2/Vi		
4. Fácil		T2	T2		T2/Vi	
5. Muito fácil						T2/Vi

T2 - Tarefa 2 / V - ViSC / Ta - Tableau

Os usuários do grupo 1 conseguiram desvincular a tarefa da ferramenta, diferentemente do grupo 2. No entanto, fica claro que, também na tarefa 2, os usuários que utilizaram ViSC consideraram a ferramenta mais fácil. A Tabela 19 mostra que o Tableau foi classificado como uma ferramenta de nível regular a difícil enquanto que ViSC entre regular e muito fácil.

A Tabela 20 apresenta as respostas sobre a satisfação do resultado obtido na tarefa 2. Observamos que os usuários do Tableau ficaram mais satisfeitos com as respostas.

Tabela 20: Respostas sobre a satisfação do resultado

Respostas	Grupo 1 - Tableau			Grupo 2 - ViSC		
	U03	U05	U07	U02	U04	U08
1. Não concluí a tarefa						
2. Desenhei um gráfico que não achei muito adequado mas era o que a ferramenta permitia						
3. Desenhei um gráfico adequado porque a ferramenta sugeriu.					X	
4. Desenhei o gráfico igual ao que tinha em mente após analisar as opções sugeridas porque achei mais claro.	X	X				X
5. Desenhei o gráfico recomendado e achei melhor.			X	X		

Em relação à satisfação dos resultados com o uso Tableau, U07 (do grupo 1), que marcou 5 (muito satisfeito), entregou um gráfico que seria eficiente porém com dados inadequados. Primeiramente manteve o ano de 2006 como uma informação desnecessária que atrapalhava na leitura e em segundo porque utilizou uma função de cálculo errada. U02 (do grupo 2) que marcou a mesma opção, no entanto entregou um resultado recomendado, selecionando um gráfico eficiente e os dados corretos para a resposta. Outros usuários do grupo 1 também apresentaram problemas. Tanto U03 quanto U05 não observaram que o valor *default* era a soma dos valores. O primeiro conseguiu após receber ajuda e o segundo entregou um resultado com erros também, o que mostra que entre os 3 usuários que utilizaram a Tableau, nenhum deles percebeu os erros. Entre U04 e U08, que utilizaram ViSC, ambos encontraram a resposta, apesar da pouca eficiência dos gráficos apresentados.

Quando perguntados se o gráfico gerado correspondia ao que inicialmente desejavam construir, o resultado é mostrado na Tabela 21 e somente U05, do grupo 1 e U02, do grupo 2 mantiveram uma coerência. U02 foi consistente com o que havia desenhado, utilizando colunas.

Tabela 21: O que o usuário pretendia construir, o que foi desenhado no papel e o resultado obtido

Grupo	Usuário	Era o gráfico que desejava construir?	Tarefa papel	Resultado obtido
Grupo 1 - Tableau	U03	Não, porque cometi um erro na seleção dos anos	Linhas	<i>Clustered Colum Chart</i> (colunas)
	U05	Sim	Linhas	<i>Clustered Colum Chart</i> (colunas)
	U07	Não, não tinha nada em mente, mas sabia que teria apoio da ferramenta.	Linhas	<i>Clustered Colum Chart</i> (colunas)
Grupo 2 - ViSC	U02	Não, eu pretendia fazer um gráfico de linhas	Colunas múltiplas	<i>Clusted Column Chart</i> com variação (Colunas)
	U04	Sim	Colunas mostrando apenas a variação	Series chart (Linhas)
	U08	Sim	Linhas	Clustered Column Chart (Colunas)

Em resposta à pergunta sobre o grau de interação com a janela Show me do Tableau, a Tabela 22 apresenta os resultados. Quando perguntados sobre a influência da janela *Show me*, os resultados foram apresentados na Tabela 23.

Tabela 22: Respostas dos usuários do grupo 1 sobre o grau de interação com Show me

Respostas	U03	U05	U07
1. Nem vi			
2. Vi que existia mas não interagi			
3. Passei o mouse sobre as opções mas não senti necessidade de clicar			
4. Cliquei nas opções mas não entendi seu funcionamento			
5. Cliquei nas opções e entendi seu funcionamento	X	X	X

Tabela 23: Respostas dos usuários do grupo 1 sobre a influência da janela Show me

Respostas	U03	U05	U07
1. Atrapalharam muito			
2. Atrapalharam pouco			
3. Nem atrapalhou nem ajudou			
4. Ajudou um pouco	X		X
5. Ajudou muito		X	

Em resposta à pergunta sobre o grau de interação com as perguntas de ViSC, a Tabela 24 apresenta os resultados. Quando perguntados sobre a influência das perguntas, os resultados foram apresentados na Tabela 25.

Tabela 24: Respostas dos usuários do grupo 2 sobre o grau de interação com as perguntas

Respostas	U02	U04	U08
1. Nem vi			
2. Vi que existia mas não interagi			X
3. Passei o mouse sobre as opções mas não senti necessidade de clicar			
4. Cliquei nas opções mas não entendi seu funcionamento			
5. Cliquei nas opções e entendi seu funcionamento	X	X	

Tabela 25: Respostas dos usuários do grupo 2 sobre a influência das perguntas para o resultado obtido.

Respostas	U02	U04	U08*
1. Atrapalharam muito			
2. Atrapalharam pouco			
3. Nem atrapalhou nem ajudou			

4. Ajudou um pouco	X		
5. Ajudou muito		X	

* U08 não usou as perguntas

5.4.2.2. Retrospective Think Aloud

Nesta etapa, para cada ferramenta, realizamos a interpretação dos dados coletados na tarefa 2 através do *Retrospective Think Aloud*.

5.4.2.2.1. ViSC

Os resultados apresentados são referentes aos testes dos usuários do grupo 2 (U02, U04 e U08). O asterisco (*) indica que houve problemas na funcionalidade.

Tabela 26: Problemas nas Funcionalidades da ferramenta ViSC ou no sistema gráfico

	U02	U04	U08
1. Seleção do tema e variáveis	*	*	*
2. Entendimento dos dados			
3. Inclusão e remoção de valores nos eixos			
4. Troca de dimensões			*
5. Menu de gráficos	*		
6. Compreensão das perguntas ou não utilização	*		*
7. Interação com os gráficos			*
8. <i>Checkbox</i>			
9. Ordenação			
10. Função de expandir e colapsar gráfico	*		
11. Desvio da tarefa			

1. Seleção das variáveis

U02 escolheu as variáveis "anos" e "países" no eixo horizontal e na legenda respectivamente, que foi adequado à tarefa. Ele demorou um pouco e explicou: "*Eu achava que era meio inútil. Apesar de eu ter que escolher as variáveis aqui, eu achei que poderia mudar na outra tela. Não faria muita diferença.*"

U04 teve dúvidas na seleção, mas selecionou as variáveis corretas: países e anos. Buscou incluir as 3 dimensões pedidas: anos, países e disciplinas. Ele

comentou: "***Pensei que seria legal se desse pra escolher mais de um desses que aparece nos eixos. Era pedido os anos, países e a classe de leitura. Não sei se precisava mas se pudesse selecionar antes.***"

U08 apresentou problemas na seleção. Escolheu países e disciplinas. "***Tinha que ter 3 países no caso, Brasil, Canadá e Austrália. E o gráfico seria de 2003 e 2009. Então eu escolhi a disciplina e inicializei. Inicializei o 'country' e inicializei o ViSC.***"

3. Inclusão e remoção dos valores

U08 achou a inclusão de valores fácil. Segundo ele: "***(...) achei esse bem mais fácil. Quando eu cliquei já apareceu o gráfico lá. É só jogar os valores. Coloquei os três países: Brasil, depois Canadá e Austrália. Aí eu coloquei as duas matérias (disciplinas). Não prestei atenção na leitura, acabei incluindo as duas matérias.***"

4. Troca de dimensões

Apesar da facilidade na inclusão de valores, a troca de dimensões não foi tão fácil para U08: "***(...) eu tinha colocado três matérias (disciplinas) e era só pra colocar a leitura. Eu botei as três matérias e aí eu não conseguia fazer a inclusão dos outros dois anos. Eu cliquei no 'insert' só que não funcionou, apareceu a mensagem.***"

5. Menu de gráficos

U02 afirmou que em algum momento testou se dava pra selecionar o gráfico na página inicial.

6. Perguntas

U02, apesar de ter utilizado as perguntas, usou-as como forma de navegação entre os gráficos. Na segunda página, mantendo o gráfico inicial, colunas agregadas, selecionou os dois anos e os três países: "***Olhei, aí vi que já tava bom nessa hora. Quer dizer, o ano tava errado mas já dava pra identificar quem teve o maior aumento de notas. Mas eu ainda quis ver se dava pra melhorar o gráfico.***"

U02 passou o *mouse* sobre as perguntas e clicou no gráfico mais pontuado da pergunta correta: "***Eu tava procurando gráfico de linha em inglês. E, achei esse aí (gráfico de dispersão). Eu entendi o gráfico mas ficou meio esquisito. Ficou***

2003 X 2009 e achei que não servia pra tarefa." Em outro momento, clicou no gráfico de colunas agregadas da primeira pergunta que era relacionada a tarefa de filtro, ou seja, não era adequada para a tarefa dada. Como ele já havia clicado na pergunta variação, o novo gráfico abriu com a coluna de variação aberta: *"Aí abriu outro e gostei, achei que a variação mostrava o que a tarefa pedia."* Neste momento, U02 clicou no botão variação: *"Lembrei que vi essa variação no vídeo."* Salvou o gráfico.

Sobre as perguntas, ao selecionar os valores, U04 viu o gráfico de colunas agrupadas e não achou legal: *"Fui direto por aí (perguntas) pra ver se tinha alguma coisa que me ajudava. Cliquei nessa pergunta que faz sentido (Houve crescimento ou declínio em algum país?) mas o gráfico (séries) acabou não adiantando muito, ficou bem ruinzinho."* Apesar de não gostar, não testou outros: *"Eu achei que esse fosse o melhor que ele (sistema) pudesse me informar. Fui na dele, achei que esse seria o ótimo."*

U08 não recorreu às perguntas.

7. Interação com os gráficos

U08 gostaria de ter mais interação com o gráfico: *"Eu queria tirar uma parte do gráfico ali que não tava sendo usado, no caso o número 800 (do eixo vertical). Não tava sendo usado. Os valores estavam chegando só até quase 600."*

10. Expandir e colapsar gráfico

U02 fechou o gráfico de dispersão e o gráfico de colunas ficou minimizado: *"Fechei o gráfico, só que aí eu reparei que fechei os dois e fiquei me perguntando o que eu fiz?"*

5.4.2.2.2. Tableau

Os resultados apresentados são referentes aos testes dos usuário do grupo 1 (U03, U05 e U07). O asterisco (*) indica que houve problemas na funcionalidade.

Tabela 27: Problemas nas Funcionalidades da ferramenta Tableau ou no sistema gráfico

	U03	U05	U07
1. Compreensão dos dados			
2. Compreensão do filtro	*	*	
3. Compreensão de dimensões e medidas	*	*	
4. Compreensão de dados contínuos e discretos	*	*	*
5. Compreensão dos gráficos exibidos	*	*	
6. Compreensão de linhas e colunas		*	
7. Compreensão da visão geral	*	*	
8. Execução da tarefa	*		
9. Compreensão das funções			*

1. Compreensão dos dados

Não houve comentários positivos ou negativos em relação ao entendimento dos dados por nenhum dos 3 usuários.

2. Compreensão do filtro

U03 não sabia como começar e afirmou: "*Como (a tarefa) compara país, achei que começar pelo país poderia ser uma boa (...) Resolvi pegar o mapa mesmo e ver o que ele me dava. Me deu um mapa. Aí eu fui mexendo, pra ver o que aconteceria se eu mexesse nele. Achei como selecionar os países que eu queria. Selecionei o Brasil e vi que não era bem o que eu queria. Ele só me dava um resultado pelo país. Aí eu desisti e nem cliquei nos outros dois.*" Ele retirou longitude e latitude das janelas "colunas" e "linhas" e afirmou que **foi bom porque trouxe a janela de países para a sua vista, uma coisa que ele queria achar**, tendo apresentado um comportamento oportunista. Depois de encontrar a medida, U3 comentou: "*Aí eu quis começar a filtrar por ano.*" Arrastou a dimensão "ano" para as colunas e o resultado gerou um gráfico de linhas com a soma das notas de cada país em cada ano. Segundo ele: "*Piorou um pouco a situação. Eu fiquei procurando como separar por mês. Na hora eu achei que faria sentido. Aí foram umas idéias muito doidas. Tudo o que eu queria fazer era tirar 2006.*" Abriu uma janela para filtrar meses do ano e comentou: "***Eu vi que isso***

tava errado (...) Meti a mão na massa. Só queria 2003 e 2009. Tinha informação demais. Aí fiquei pensando como é que poderia resultar. Ver o que descobro e o que acontece. Qualquer coisa, control z." Mais uma vez U03 apresentou comportamento oportunista, buscando em todas as funcionalidades algo que pudesse ajudá-lo a excluir o ano 2006.

U05 também encontrou dificuldade em filtrar, inclusive após já ter feito uma vez: Ao clicar no sinal de mais da dimensão "year", ela quebrou em trimestres: *"Aí foi minha dificuldade de filtrar, na verdade, os anos. Não lembrava onde filtrava. Tentei botão direito e isso aqui (mostrando o botão '+'). Tinha que filtrar país e ano."* Ao clicar na janela filtros comentou: *"Querida criar um filtro como no gmail. Criar um filtro pra email. Isso que eu pensei. E aí, cacei por aí alguma coisa."* Quando achou os filtros, durante o teste comentou: *"Vou ter que desmarcar um por um?"* Tentou clicar no 'all', e explicou: *"Eu cliquei no 'all' duas vezes, porque quando você clica no 'all' do gmail você coloca todo mundo e tira todo mundo."* Encontrou o 'none' rapidamente, desmarcou todos e selecionou os três países. Ainda precisava tirar um dos anos: *"Eu não quero 2003 e agora filtrar o ano. Eu tinha esquecido o que eu tinha feito. Aí eu tive que caçar o filtro do ano agora. Não é muito evidente onde clica o filtrar. Aí eu fiquei meia hora procurando o filtro pra filtrar o ano."* Passou pela barra de menu superior do sistema. *"Será que aí em cima tem como colocar o filtro?"* Acabou encontrando como excluir uma coluna da tabela e comentou: *"Aí eu excluí o ano errado. Aí tentei voltar com todos os anos. Tentei tirar e colocar o ano de novo pra ver se o ano voltava. Aí eu sem querer achei isso aí (filtro) também e tirei o certo."*

U07 achou o filtro rapidamente, porém através do 'quick filter' que não apresentava todas as opções de seleção: *"Ah, agora eu vou fazer o filtro do país. Ah! 'Quick filter' tava ali do lado mas não vi as modificações (abriu a janela de filtros a partir do quick filter). Já tinha tanta coisa na tela que era complicado. É, ali não tinha o botão para desselecionar todo mundo."*

3. Compreensão de dimensões e medidas

Sobre a dificuldade em encontrar um dado quantitativo, U03 comentou: *"Aí o 'score' que foi pedido. Eu demorei pra vê-lo."*

U05 iniciou comentando: *"Bem, aí eu já olhei e falei: ferrou! O que eu vou fazer agora? Eu tinha esquecido que eu tinha que ir jogando os valores. (...) E eu*

fiquei só preso às dimensões. Eu esqueci que lá em baixo tinha os 'measures' (...)" Em outro momento comentou: *"Ainda tava tentando ver se funcionava.(...) Eu não sabia nem o que fazer, porque que tava dando esse gráfico assim. Primeiro eu achei que eu tava errado, que eu tava fazendo linha e coluna errada. É que eu troquei pra barra e tal. (...) Aí você falou. Lembra que tem o 'score'_. Aí eu caçei o 'score' lá no 'measure', mas eu tava preso às 'dimensions' só, ignorei aquele 'measure' ali. Aí sim, ficou o gráfico que eu queria."*

U07 não teve problemas em encontrar as medidas: *"Eu selecionei o país e ano e usei o 'score' ali. Só que o 'score' eu não sabia se ele significava a média ou se era o total. Por isso que aí depois (não terminou de explicar)."*

4. Compreensão de dados contínuos e discretos

U03 não entendeu que o gráfico de linhas selecionado alterou o tipo da variável ano de discreta para contínua e por isso inseriu os anos retirados. Para U03: *"Na minha cabeça, por mais errado que podia parecer se eu selecionasse os dois anos e o intervalo, ele ia mostrar apenas o que eu queria. Aí eu voltei (gráfico de colunas agregadas). Quase lá."*

U05 se atrapalhou no filtro de anos por não perceber que estava trabalhando com o ano do tipo contínuo: *"(...) E os anos reapareceram. Eu tentei filtrar só 2003 e 2009. Fiquei perdido aí. Não sei se era isso aí. E não era bem isso. Perdi muito tempo tentando filtrar só 2003 e 2009. Aí pensei: Ah, deve ser aqui (analisando a janela de filtros), antes e depois. Droga, não dava pra encontrar antes e depois aqui. Vou escrever. Tentei escrever 2003 e não foi. Não da pra escrever. Vou voltar aqui. Aí ele voltou por mês. Vou demorar um século aqui. Aí vi que dava pra arrastar, mas não adiantou nada porque eu não conseguia selecionar agora o 'Ending date' (data final). E eu pegaria um 'range' (intervalo). Não pegaria uma data e outra. Só percebi no final das contas que só conseguia ver 'range' (intervalo) ali. (...) Aí eu pensei: Deixa eu sair daqui! Mas ainda continuei tentando mexer, aí eu cancelo. Vou sair daqui, não é isso. (...) Aí eu lembrei que eu tinha mexido ali (marks) quando comecei a fazer a experiência com esse trecho aí. Mas também não era isso."*

U07 também não percebeu a transformação automática nos dados entre os dois gráficos de linhas: *"Acho que cheguei a mudar o gráfico não lembro. Cheguei. Pra ver se tinha uma visualização melhor. Aí eu falei, hum, não,*

estranho (observando o gráfico de linhas contínuas). Esse aí, por exemplo, a escala dele ficou muito ruim. "

5. Compreensão dos gráficos exibidos

Sobre a janela *Show me*, U03 afirmou: "*Não me deixava clicar. Tentei (referindo-se aos gráficos da janela show me que estavam desabilitados).*" U03 passou pelo gráfico de colunas empilhadas na segunda tarefa: "*Eu achei esse gráfico bonitinho por isso eu cliquei nele. Escolhi os países que eram pedidos na tarefa.*" Porém, ao ver o gráfico de colunas empilhadas com apenas uma coluna, disse: "*O gráfico tava longe de ser o que eu imaginava (...)*". Mais tarde, ao conseguir selecionar os dados solicitados: "*(...) Obtive o resultado que imaginava, só a escala ficou errada.*" Tentou o gráfico de linhas novamente: "*Fui ver se ficava melhor ou não. Não me agradou muito. São só três linhas. Não faz muito sentido. Não dá nem pra ver a evolução direito.*" Voltou ao gráfico de colunas. *Era visualmente claro que o do meio (Brasil) tinha a maior diferença. A linha tava muito reta. A do Canadá tava muito reta. Fui lembrado que 'line' não era uma boa escolha (ao selecionar outra vez o de linhas múltiplas). Eu queria fazer o de barras.*"

U05 não compreendeu porque somente o gráfico de Gantt foi habilitado para a sua seleção e afirmou: "*Beleza, agora eu tinha país e ano mas eu tinha esquecido do 'measure'. Aí eu já tava querendo partir pro gráfico. Que na minha cabeça o 'measure' já tava ali. Eu ignorei a coluna de 'measure'. Aí começou a gerar esse gráfico estranho (Gannt). Eu pensei: Que gráfico é esse?*" Ao conseguir incluir as medidas: "*Mas ficou com mais anos que eu queria ainda. Aí agora eu tava procurando um jeito de deixar só com os anos que eu queria. Aí finalmente eu consegui. Aí ficou um país em cima do outro. Aí não, eu fui ajustando o gráfico. Ficou mais fácil. Ajeitei pra um do lado do outro. Aí teve um problema que ficou por mês. Eu não queria por mês do ano. Eu queria por ano. Aí eu troquei ali pra ano do ano, 'year of year'. E aí, pronto. Aí deu certo. Eu copieie e coleie.*"

6. Compreensão das linhas e colunas

U05 afirmou: "*Demorei porque tava lendo o problema e tentando montar na minha cabeça quem seriam as minhas colunas e as minhas linhas. Aí você já não pensa no gráfico, você tá pensando mais como se fosse uma tabela "*

7. Compreensão da visão geral

U03 recebeu ajuda para entender a visão geral. Ele afirmou: "*Só falta responder pergunta que era a média e não a soma como **você lembrou**, mas eu ainda não tinha me tocado disso.*"

U05 não mencionou nada sobre isso, entregando um resultado que manteve a soma dos dados.

U07 percebeu a visão geral: "*Depois que eu vi que o 'measure' tava com 'sum' aí eu descii para mudar.*"

8. Execução da tarefa

Durante a parte em que U03 ficou procurando os filtros, ele desviou completamente da tarefa: "*Aí eu perdi completamente a linha. Porque mostrava os intervalos. Não percebi que pra tirar os intervalos era só voltar pro gráfico de colunas. Aí aqueles minutos entediante que eu falei que ia acontecer.*" Os minutos entediante que ele se refere são aos minutos que percorreu janelas de todo o site, com ou sem relação com a tarefa. Sobre isso afirmou: "*Uma grande parte foi curiosidade. Nos cinco minutos anteriores não deu pra ver muita coisa. Sou uma pessoa um pouco curiosa.*" A curiosidade de U03 o distanciou da tarefa, porém demonstrou muito sobre seu processo abduativo. Em outro momento, ao reposicionar a janela *Show Me* na tela e aumentar a largura das linhas, expressou: "*Aí, o meu TOC falou mais alto. Eu sou cheio de frescura.*"

9. Compreensão das funções de análise

U07 não encontrou onde trocar a soma por média e buscou uma função diferente: "*No caso, vi que era somatório e como eu podia transformar isso aí na média. Aí eu fiquei um tempão tentando ver com o botão direito que informação ele me dava pra poder trabalhar em cima. **Não tendo**. Aí eu vi, bom tem '**moving average**'. Deve ser o que eu quero. Aí eu falei. Beleza.*" Existia a função dentro do menu de contexto '*measure*' (medida) mas ele foi no menu de contexto '*quick table calculation*'.

5.4.2.3. Etiketagem

Nesta etapa, para cada ferramenta, apresentamos a interpretação dos dados coletados na tarefa 2 através da etiquetagem.

5.4.2.3.1. ViSC

A Tabela 28 exibe os resultados obtidos na etapa da etiquetagem pelos mesmos usuários (grupo 2 – U02, U04 e U08). Cada círculo preenchido (●) representa uma ruptura.

Tabela 28: Etiketagem das rupturas encontradas na tarefa 1 com Tableau

		U02	U04	U08
Falhas temporárias 1. Interrompem a semiose do usuário	1. Cadê?	●●●	●	
	2. Ué, o que houve?			●●
	3. E agora?	●●●	●	●
Falhas temporárias 2. Usuário percebe que seu ato comunicativo não foi bem sucedido	4. Onde estou?			
	5. Epa!	●●	●	
	6. Assim não dá			
Falhas temporárias 3. Usuário tenta compreender o ato comunicativo do sist.	7. O que é isso?	●●●	●●●●●●●● ●●●	●●●●●●●● ●●●●
	8. Socorro			●
	9. Por que não funciona?			●
Falhas parciais	10. Vai de outro jeito.	●		
	11. Não obrigado	●		
Falhas Completas	12. Pra mim está bom.			
	13. Desisto.			

1. Cadê? No teste de U02, as duas primeiras ocorrências dessa etiqueta ocorreram durante a busca por um gráfico que ele não sabia com que nome procurar. Isso ficou mais claro na entrevista quando afirmou que buscava o nome do gráfico de linhas em inglês. A terceira ocorrência dessa etiqueta foi quando ele procurou a operação de soma na terceira dimensão, se perdendo com o objetivo da tarefa.

No teste de U04, essa etiqueta ocorreu durante a busca pela pergunta. Ele realmente verificou as perguntas existentes e escolheu a pergunta que apresentava se houve ou não crescimento dos pontos entre os anos e para os países selecionados.

2. Ué, o que houve? No teste de U08, essa etiqueta ocorreu duas vezes na tentativa do usuário em entender o que estava acontecendo ao ter clicado nos valores da dimensão ano, que não havia sido escolhida e por isso, não era possível incluir mais de um valor. U08 alternou duas vezes até perceber que o gráfico estava sendo alterado.

3. E agora? No teste de U02, esta etiqueta ocorreu após a seleção dos anos, depois, após a seleção dos países e por último, após o fechamento de um gráfico mantendo minimizado o gráfico *default*. **Ele ficou sem saber o que fazer para abri-lo novamente. U04 demonstrou não gostar muito do gráfico apresentado e parou, o que entendemos como se estivesse definindo um novo plano.**

No teste de U08, esta etiqueta aconteceu após o usuário ter selecionado variáveis para o eixo horizontal e para a legenda quando ficou sem saber como prosseguir.

5. Epa! No teste de U02, a primeira ocorrência foi quando o usuário percebeu que havia incluído anos a mais e retirou da seleção. A segunda foi quando ele fechou o gráfico e minimizou o gráfico *default*.

No início do teste de U04, este escolheu como segunda variável "disciplina" mas logo depois alterou para a variável certa: anos.

7. O que é isso? No teste de U02, esta ocorreu primeiramente quando ele observou o gráfico de dispersão (*scatterplot*) buscando compreendê-lo. As outras vezes foram ao posicionar o *mouse* sobre dois botões de ordenação procurando alguma informação.

Das 10 repetições encontradas no teste de U04, a primeira foi quando ele verificou o que era a caixa de seleção da terceira dimensão e as duas seguintes após montar o gráfico de colunas agrupadas quando ele posicionou o *mouse* sobre duas colunas do gráfico para verificar os valores apresentados. Após encontrar o

gráfico entregue, consideramos uma nova ocorrência quando tentou entender o gráfico e as próximas seis ocorreram quando ele passou o *mouse* sobre os pontos do gráfico para obter os seis valores apresentados. No teste de U08, nove delas foram para verificar os valores das barras e, por último quando recebeu o alerta informando como trocar a variável.

8. Socorro! Após receber a mensagem para retirar valores de uma das variáveis, U08 perguntou ao avaliador se estava deixando o valor correto.

9. Por que não funciona? U08 voltou a **tentar incluir um novo ano no gráfico, sendo que ano não era nem a variável escolhida para o eixo horizontal nem para a legenda.**

Entre as falhas parciais, encontramos:

10. Vai de outro jeito! Ocorreu no teste de U02 quando ele buscou abrir o gráfico **através de uma pergunta aleatória, que não estava de acordo com a tarefa.** Ele não entendeu o uso das perguntas e usou como forma de abrir o gráfico fechado.

11. Não obrigado. U02 recusou a solução recomendada (gráfico de dispersão) pelo sistema por achar que o gráfico que buscava era mais eficiente.

5.4.2.3.2. Tableau

A Tabela 29 exhibe os resultados obtidos na etapa da etiquetagem pelos mesmos usuários (grupo 1 — U03, U05 e U07). Cada círculo preenchido (●) representa uma ruptura.

Tabela 29: Etiquetagem das rupturas encontradas na tarefa 2 com Tableau

		U03	U05	U07
Falhas temporárias	1. Cadê?			
1. Interrompem a semiose do usuário		●●●●●●●● ●●●	●●●●●●●● ●●●●●●●● ●●●●●●●● ●	●●●●●

	2. Ué, o que houve?		●	
	3. E agora?	●●●●	●●●●●	●
Falhas temporárias	4. Onde estou?			
2. Usuário percebe que seu ato comunicativo não foi bem sucedido	5. Epa!	●●●●	●●●	
	6. Assim não dá	●●●		●●
Falhas temporárias	7. O que é isso?	●●●●●●●● ●●●●●●	●●●●●●●● ●●●●●●●● ●	●●●●●●●● ●●●●
3. Usuário tenta compreender o ato comunicativo do sist.	8. Socorro	●	●	
	9. Por que não funciona?	●●	●●●	
Falhas parciais	10. Vai de outro jeito.		●	
	11. Não obrigado			
Falhas Completas	12. Pra mim está bom.		●	●
	13. Desisto.			

1. Cadê? No teste de U03, essa etiqueta, primeiramente, ocorreu na busca **dos filtros de países** com três repetições. Depois, **mais seis vezes na busca do filtro dos anos**. Por uma formulação errada do plano de ação, ocorrido no nível estratégico, houve um maior índice dessa etiqueta na busca desse filtro. E, por último, uma vez, no final na tentativa de alterar a linha com a função soma dos pontos pela média.

Encontramos 18 ocorrências no teste de U05. O usuário passou um bom tempo do teste **procurando onde estava localizada a opção de filtrar**, clicando em diversos elementos e abrindo janelas aleatórias em busca do filtro. Na primeira vez encontrou a opção, mas já na segunda vez que precisou, já não sabia mais como encontrar. Após a seleção do gráfico de Gannt o qual alterou as características da variável ano, fez com que se perdesse novamente.

U07 teve quatro ocorrências que foram na **busca pelo filtro de países** e a última na busca de como inserir os valores acima das colunas no gráfico, o que resolveu com certa facilidade.

2. Ué, o que houve? Ocorreu quando U05 selecionou o gráfico de Gantt e fez com que os anos já filtrados aparecessem novamente, uma vez que ele **transformou os anos para dados contínuos**, porém o usuário não entendeu isso.

3. E agora? ocorreu quatro vezes no teste de U03. Basicamente, logo após a etiqueta "Assim não dá!". O usuário parava um tempo para elaborar um novo plano de ação após verificar que não chegaria a resposta através da solução anterior em situações diferentes.

No teste de U05, ocorreu após a inclusão dos valores nas linhas e colunas quando tentava formular um plano para filtrar os elementos. A última ocorrência foi ao conseguir trocar o gráfico de Gantt pelo de barras e **não entender os valores**. Como estava impactando o teste, ele aceitou receber ajuda e após interferência do avaliador observou que ainda faltava incluir a medida.

U07 obteve esta etiqueta após testar diferentes visualizações, voltou ao gráfico de colunas agrupadas buscando algo que pudesse **melhorar o gráfico**.

5. Epa! Apareceu quando U03 rapidamente percebeu uma escolha de gráfico errada por duas vezes, quando **quebrou o ano em trimestres sem querer** e quando verificou que abriu uma janela que não era adequada ao plano.

No teste de U05, ocorreu quando o usuário **quebrou o ano em trimestres** e rapidamente descartou a opção. Isso aconteceu duas vezes. A terceira vez com alto custo foi quando ele percebeu que excluiu a coluna errada, mas não sabia consertar.

6. Assim não dá! No teste de U03 ocorreu três vezes. Inicialmente após observar as opções dadas pelo mapa por algum tempo, **verificou que o mapa não era o caminho para a realização da tarefa**. Em outros dois momentos, **na busca pelo filtro de datas**, o usuário passou por caminhos longos, por exemplo: abrindo janela de filtros de meses do ano, sendo que ele queria filtrar os anos exibidos. Após verificar todos os elementos disponíveis na janela, percebeu que não era a estratégia correta e fechou. Com o mesmo plano abriu também a janela '*Edit Axis*' (*year of year*) e após analisá-la também fechou.

No teste de U07, a primeira ocorrência desta etiqueta foi após ele parar um tempo lendo o *menu* que abriu no clique de um elemento da coluna. O usuário

percorreu todo ele com o *mouse*, não encontrou o que queria e fechou. A segunda vez foi ao perceber que os gráficos de linha não mostravam uma solução eficiente para a tarefa, **descartando a visualização**.

7. O que é isso? Sempre que mudava de planos, U03 percorria o *mouse* pela tela posicionando ou clicando sobre os elementos. **Durante repetidas vezes, o usuário parou o *mouse* sobre os elementos da dimensão, das medidas e sobre os gráficos da janela *Show me***. Consideramos esta etiqueta também nos momentos em que U03 se desligou da tarefa para buscar mais informações sobre o que a ferramenta oferecia, tais como: alterar a espessura da linha do gráfico através do *slider* tamanho e clique em todos os signos da janela '*edit Axis*'.

No teste de U05 foram 13 ocorrências, principalmente no início e na **busca do filtro** de anos aberta com o gráfico de Gannt.

U07 passou o *mouse* sobre todas as **colunas do gráfico** a fim de identificar o que estava sendo exibido. Consideramos esta etiqueta também quando ele clicou sobre os *links* dos gráficos para verificar como seriam apresentados. No final, após substituir a soma por '*moving average*', verificou novamente cada coluna.

8. Socorro! U03 não observou a **visão geral** que atribuía a soma e aceitou a recomendação do avaliador. U05 teve problemas em observar a falta das **medidas** e aceitou ajuda do avaliador.

9. Por que não funciona? ocorreu quando U03 tentou clicar nos ícones de **gráficos inativos**. Ele repetiu duas vezes para se certificar que nada acontecia.

Dois ocorrências foram encontradas no teste de U05 durante **a busca pelo filtro de anos**. Primeiro quando o usuário repetidamente clicou na janela '*filters*' e depois no cabeçalho da tabela esperando uma resposta que não aconteceu. A outra ocorrência foi no clique duplo do botão '*All*' da janela de '*filters*' esperando que desmarcasse todos no segundo clique.

Entre as etiquetas de falhas parciais encontramos:

10. Vai de outro jeito. Ocorreu quando U05 não encontrou como filtrava os anos e **encontrou uma outra solução, através da exclusão da** coluna que não era solicitada.

Entre as etiquetas mais graves, de falhas completas encontramos:

12. Pra mim está bom. U05 entregou um gráfico que continha erros para a tarefa solicitada. Consideramos esta etiqueta também para U07 porque ele alterou os valores consideravelmente, distorcendo os resultados após a operação *'moving average'* selecionada, por engano, quando ele queria mesmo a média.

5.4.2.4. Discussão sobre a tarefa 2

Através da etiquetagem, nesta tarefa verificamos que, no uso do Tableau, os problemas que levaram às rupturas mais graves se repetiram e **ocorreram pela falta de entendimento entre a diferença entre medidas e dimensões, pelo não entendimento da visão geral e pela dificuldade de localização dos filtros que levou até mesmo ao desvio da tarefa por parte de U03**. Este problema partiu de um outro problema grave que foi a falta de percepção da mudança na interface pela transformação do dado discreto em contínuo. U05 explicou em entrevista que: *"O que mais me dificultou no Tableau foi essa parte de escolher que dados, quais seriam as variáveis do meu gráfico. Primeiro eu tive dificuldade de perceber aquele negócio de 'rows' e 'columns' e depois que eu me toquei daquilo, eu não percebi que ele tava dividindo em 'dimensions' e 'measures' também. Não me toquei dessa divisão. (...) O que me ajudou foi essa parte do 'show me' pra escolher qual o formato do gráfico mas tirando isso foi bem penoso (...)".* U03 que teve problemas com os filtros e declarou: *"Algumas opções achei, não digo escondidas, mas não tão na sua cara quanto na ViSC. Até porque ela tem mais opções, não dava pra fazer um menu cheio de botões que ninguém ia achar nada."* O único usuário que conseguiu entender esses conceitos foi U07 que gerou a resposta que apresentava mais problemas pois não entendeu as funções de cálculos da ferramenta. Mesmo tendo apresentado um resultado que achou eficiente, segundo ele, a ferramenta não ajuda muito: *"O que ajudou? Pouca coisa! (...) A apresentação da ferramenta, o tutorial é muito bonitinho. É muito melhor que o da ViSC. É um tutorial que você fala: Caraca! Eu consigo extrair muita coisa daqui de forma fácil. Mas aí depois você vai meter a mão na massa. Você fala, cara, onde é que tá o que eu preciso? O que que vai me levar a fazer tão bonitinho quanto a apresentação, sabe? Tão fácil! Tão direto ali. Aí eu fiquei*

meio perdido. Dificultou na forma de como eu ia manipular esses dados. Tipo, eu tenho que clicar aqui mas eu tenho que segurar o control também. Isso em nenhum momento foi falado no vídeo. O vídeo só falava: você seleciona essa, seleciona essa (...) Então eu vou clicar no show me. Nada! Hum, tá. Agora eu tenho que selecionar esse dois dados. Lembra que era, acho que era 'gender' e ano, aí depois batia com 'score', né? Alguma coisa assim. Aí, não falava que eu tinha que fazer isso. Clicou, selecionou. Não! Clicou num, desselecionou o outro."

Rupturas menos graves como a quebra do ano em meses, a visualização de um gráfico inadequado e a repetição da visualização dos *tooltips* dos elementos foram facilmente contornadas pelos usuários.

O *retrospective think aloud* revelou ainda seus processos abduativos que levaram a determinados comportamentos. U03, por exemplo, se demonstrou mais curioso que perdido em relação à ferramenta e à tarefa. U03 resolveu investigar caminhos na tentativa de encontrar algo que pudesse lhe ajudar, sabendo que poderia retroceder de qualquer ponto pelo comando *Ctrl+Z*. U05 demonstrou relacionar interface da ferramenta com a do Google e também comentou sobre o problema de formar um gráfico pensando como tabela e seus planos na tentativa de filtrar os anos (contínuos). Esses dois usuários também comentaram como chegaram na solução do gráfico de barras, mostrando mais uma vez que foi por influência da ferramenta que não mostrou o gráfico de linhas em uma escala eficiente e ainda por incluir valores que não haviam sido pedidos (devido à transformação do tipo de dado).

Conforme na tarefa 1, todos eles mostraram que seu entendimento sobre a janela *Show me* não foi completo, embora todos tenham afirmado que a entenderam quando perguntados. U03 disse apenas que foi crucial tanto para ir para o lugar errado quanto para achar o certo porque ela gera o gráfico sozinha, mas não observou a parte de baixo com as explicações das variáveis necessárias para ativar as visualizações. Já U05, respondeu: "*Ela foi intuitiva pra escolher qual era o formato que eu queria mostrar meu gráfico. Não, eu percebi que quando algumas variáveis entravam em jogo, ele habilitava algumas opções. Por exemplo, a parte do mapa era bem intuitiva. Você tinha uma variável única de 'country' que permitia você ter aquela visualização pelo mapa. Isso eu já tinha percebido desde a familiarização que você me deu de 5 minutos. Mas assim, quando uma variável me permitia fazer um gráfico em linha outro em barra, isso*

não ficou claro. Sei que eu coloquei as variáveis que eu queria, ele me deu a opção de barra e eu usei." U07 compreendeu melhor: "Pra mim é a mesma coisa do Excel. Mostrava quais os gráficos disponíveis que você pode fazer." E sobre quando estavam habilitadas respondeu: "Aí era quando eu trocava as dimensões do dado mas eu não mexi muito com isso não." Mas também declarou não ter visto as explicações no final da janela.

Na ferramenta ViSC, as rupturas mais graves encontradas com a etiquetagem foram: **a utilização das perguntas para navegação sem a leitura das questões por U02 que chegou a abrir o gráfico mais pontuado para a tarefa sem mesmo analisá-lo e um problema na ontologia, verificado no teste de U04, que selecionou uma pergunta satisfatória que não lhe deu um resultado eficiente.** Foram encontrados também outros problemas de IHC menos graves tais como: a dificuldade na substituição de dimensões e a funcionalidade de colapsar e expandir um gráfico que não ficou evidente.

O *Retrospective Think Aloud* **revelou facilidade em selecionar os valores, dificuldade no raciocínio durante a seleção das dimensões e confirmou o problema das substituições das dimensões selecionadas. Confirmou ainda problemas identificados na tarefa anterior: falta de entendimento dos dados. No entanto, sua maior contribuição foi verificar como e porque as perguntas foram utilizadas.** Para U02, que recorreu as perguntas verificamos que ele já sabia que tinha uma resposta mas quis verificar se encontrava algo melhor. U04 já afirmou ter ido diretamente e não testou outras possibilidades pelo fato de ter sido a recomendação do sistema.

Ao serem perguntados sobre o entendimento das perguntas, U02 demonstrou não ter entendido muito bem e que não fez nenhuma análise: "*Entendi que eram várias maneiras de mostrar a informação e dava pra filtrar algumas vezes. Só que eu não queria filtrar. Eu queria só achar a melhor visualização. Primeiro eu procurei um gráfico de linhas. Achei um gráfico que também não queria. Igual na outra.*" Afirmou ainda não ter percebido as pontuações das visualizações e o fato de ajudar um pouco foi no sentido de ter aberto o gráfico pra ele.

U04, no entanto, mesmo tendo não ficado muito satisfeito mostrou que entendeu as perguntas e que para ele, elas foram importantes: "*Só seria talvez um atalho se você tem alguma coisa em mente que você quer gerar, ele te ajuda. Eu*

*vi que ele é meio inteligente pra ver os dados que eu to querendo olhar e relacionar as perguntas com esses dados. No meu caso foi útil. Eu cliquei aqui e foi direto pra resposta. Só faltou melhorar um pouco a escala mas era justamente o que eu queria. **Eu achei importante, me salvou bastante tempo.**"*

Segundo U08, ele não interagiu devido à quantidade de texto. Essa pergunta nos levou a perguntá-lo se o fato das perguntas estarem em inglês foi determinante para não utilizá-las e eles afirmaram que se tivesse em português talvez tivesse lido.

Na tarefa 2, vimos que as perguntas foram importantes para U04 mesmo exibindo um resultado não tão eficiente. Ele foi o único usuário que realmente compreendeu seu funcionamento. No entanto, U08 não utilizou devido à grande massa de texto e U02 utilizou de forma inadequada. O Tableau, novamente foi aparentemente atrativo para os usuários, mas as respostas possuíam problemas que não foram percebidos pelos usuários.

5.4.3. Discussão sobre as ferramentas

Como na tarefa 1, embora todos tenham afirmado que entenderam, suas respostas quando perguntados sobre como foi seu entendimento, mostraram que o entendimento não foi completo.

Quando perguntados sobre o que gostou mais nas duas ferramentas, U02 respondeu: *A primeira parecia que tinha mais opções. Vendo o Show me, existiam mais opções de gráficos, só que eu não conseguia usar, mas parecia que dava. **O outro (ViSC) eu gostei que era mais simples. Já começou e já tinha um gráfico na minha frente. A primeira começa, a tela está em branco pra você mexer pra formar o gráfico. A segunda já tem um gráfico e você vai modificando ele. (...)** Achei o outro (ViSC) mais fácil de usar. Talvez tenha menos funcionalidades. **Esse Tableau parece ser mais completo. Tem mais coisas pra fazer.** " Para ele a facilidade maior foi na seleção das variáveis e no fato de ter um gráfico que modifica. Em relação ao que ajudou e dificultou, ele respondeu: *"No Tableau, eu acho que mais dificultou. **Pelo vídeo, eu fiquei mais interessado pelo Tableau. Se eu fosse ver o vídeo e comprar uma, acho que eu compraria o Tableau. Pareceu mais atrativo. Mas usando, eu achei o outro mais fácil. É, não sei explicar exatamente o que que ajudou e o que dificultou.**"**

Na mesma pergunta, U03 respondeu: " *Na Visc, a **simplicidade** e a disponibilidade por ser uma ferramenta web. No Tableau, as possibilidades. Tendo mais tempo para conhecê-la, acho que seria melhor.*" **No ViSC, o que ajudou foi:** "*Ela sugeriu o que eu tava procurando que são essas perguntas aqui do lado.*" e o que ela dificultou foi: "*O esquema de clicar pra remover, clicar para adicionar. Aquele click insert. Só nesse caso. Na seleção errada foi distração minha, não foi culpa da ferramenta.*" Ele se refere a alteração de dimensões selecionadas. Sobre o Tableau, ela ajudou: "*Na seleção do ano. Colocar o ano foi fácil. Na seleção de país por 'checkbox'. Arrastar as classes. A interação com o usuário nela achei melhor. Saber de onde você tá tirando e levando. Mais simples.*" Dificultou em: "*Algumas opções achei, não digo escondidas, mas não tão na sua cara quanto na Visc. (...) Só que a Visc conseguiu lidar melhor com a interface até por ter menos opções. Eu lembro a dificuldade que eu tive pra retirar o ano. Tudo bem, em parte foi erro meu de tentar retirar o ano em gráficos de linha, mas foi um pouco mais difícil mexer com o ano no Tableau que na Visc. Ainda mais que era uma tabela simples.*"

Para U04: "*No Tableau, essa **variedade** bem grande dos gráficos que você pode montar. Parece que esse aqui se você quiser montar gráficos com bastante complexidade seja melhor usar este que o outro. O ViSC foi bem mais rápido de pegar a resposta que o outro. **Simplicidade** ajudou bastante. Pra esse problema específico ficou mais fácil.*" Para ele, o que ajudou no Tableau: "*foi exatamente essa rapidez de você ver os valores todos dados e conseguir agrupar e montar o gráfico bem rápido. É fácil de responder a pergunta.*" Já a dificuldade foi por ter muita informação. Já no ViSC, ele afirmou: "*Esse **filtro aqui fica bem fácil**, acho que fica mais fácil de filtrar do que no Tableau. **Essas perguntas também são legais porque ajudam já montar mais ou menos o gráfico que você quer.** Te dá mais ou menos a resposta que você quer. É mais automático. Não tive muita dificuldade não. Só se talvez ele me desse mais opções ali no 'show me' na pergunta.*"

Para U05 o que ele mais gostou em cada um foi: "*A **ViSC eu achei muito simples de filtrar.** Na Visc, o que eu mais gostei foi exatamente essa parte inicial de quais vão ser as variáveis do meu gráfico, achei super prático. **E no Tableau foi a parte do 'Show me'** ali, de eu clicar em qual formatozinho do gráfico que eu quero.*" Para ele, o principal problema do ViSC foi a troca de visualizações por

não possuir as miniaturas como o Tableau, mas ainda assim, não foi uma dificuldade enquanto que, a filtragem do Tableau já foi um problema mais grave. Acrescentou ainda *"A ViSC me ajudou na parte de filtrar os dados. Já tem um problema na interpretação dos textos que muitas vezes o texto pode vir um pouco ambíguo, você faz uma leitura rápida "*

O usuário U07 gostou mais das perguntas: *"No ViSC, as perguntas. Aquilo lá já te dá uma mão, já prevendo o que você quer fazer. Então, é um 'shortcut'. Te dá um atalhinho ali pra você ver e não perder tanto tempo manipulando o que você precisa. E o Tableau, eu achei que ela tem uma **apresentação** legal mas não correspondeu ao que esperava vendo o vídeo."* Sobre o Tableau, respondeu: *"O que ajudou? Pouca coisa! (...) o tutorial é muito bonitinho. É muito melhor que o da ViSC (...). Dificultou na forma de como eu ia manipular esses dados. Tipo, eu tenho que clicar aqui mas eu tenho que segurar o control também. Isso em nenhum momento foi falado no vídeo (...)".* Sobre o que ajudou em ViSC, respondeu: *"Essas perguntas, depois que eu elaborei os gráficos, a relação entre os dados que eu queria, ela me deu perguntas que eu falei: Bom, eu sei mais ou menos qual é a cara da pergunta que eu quero responder. Vamos ver quais as perguntas que ele tá me oferecendo ali que bate com isso. Até que finalmente eu encontrei a que correspondia. Ah! Então beleza. Isso aqui gerou ali. Tá até melhor do que eu planejava fazer. Vamos usar esse aqui mesmo."* Sobre as dificuldades afirmou que teve a mesma dificuldade nas duas que foi saber como usar porque nunca tinha usado antes.

O usuário U08 recebeu ajuda em ambas as tarefas mas mesmo assim declarou *"O Tableau eu achei ele mais difícil que o outro."*

Na comparação das ferramentas feitas pelo usuário, vimos alguns pontos positivos que se repetiram na fala dos usuários de ambas as tarefas, quando perguntados sobre o que mais gostaram, sobre o que mais ajudou e o que mais atrapalhou em cada ferramenta. Primeiramente, a simplicidade de ViSC foi comentada por 4 usuários. Três usuários comentaram que as perguntas foi uma das funcionalidades que mais gostaram ou que mais ajudou. Outros dois demonstraram que gostaram da forma de filtragem dos dados. Os pontos considerados positivos no Tableau foram as miniaturas presentes na janela Show me citadas por dois usuários, a variedade de gráficos disponíveis (também citado

por dois usuários), o vídeo de apresentação (também citado por dois usuários), a facilidade de interação através do *drag and drop*.

Sobre os pontos negativos de ViSC, nenhum deles considerou as perguntas como ponto negativo, porém um usuário fez uma observação de que os textos ficavam ambíguos apesar de ter considerado elas importantes. A falta de miniaturas no menu dos gráficos foi considerada um problema em ViSC, assim como a inclusão das variáveis citada por um usuário.

5.4.4. Elaboração do perfil semiótico

Abaixo descrevemos as metagensagens de emissão de cada ferramenta reconstruída na inspeção (seções 4.2 e 5.3.2) dos signos de acordo com o MIS e destacamos os pontos que não foram bem recebidos pelos usuários observados na etapa de etiquetagem.

Metagemensagem designer-usuário de ViSC:

*Penso que seja um estudante ou um profissional de áreas ligadas ou não a estatística que precisa construir gráficos eficientes. Como você pode não ter domínio sobre as visualizações, você prefere uma interface que, com pouca interação, gere um gráfico e partir dele você consiga facilmente alterar os valores e visualizar outras opções de gráfico. Você **também gostaria de poder comparar visualizações**. Assim, projetamos um sistema que, a partir da seleção do tema e duas variáveis, apresenta um gráfico em que os valores possuem um valor selecionado por padrão automaticamente. Então, basta incluir ou alterar os valores e selecionar a forma de visualização. Para lhe ajudar a decidir que visualização utilizar, a interface recomenda visualizações com base na pergunta que deseja responder. Você pode selecionar a visualização através das opções do menu ou das perguntas. Neste caso, você só precisa encontrar a pergunta e escolher uma ou mais entre as visualizações recomendadas. **Para lhe ajudar, a interface explica como obter a resposta e a apresenta no gráfico**. As recomendações ainda foram classificadas com uma nota para ajudar na sua escolha. Você também pode alternar os eixos, exibir a variação, mudar escalas, eliminar o zero do gráfico ou reordenar para tornar sua visualização mais eficiente.*

Metamensagem designer-usuário de Tableau:

*Penso que você já tenha alguma experiência com análise de dados, ou pelo menos entende como os dados e suas variáveis podem ser classificadas. Você também pode ser um especialista nesse assunto. Você prefere uma interface que seja flexível na interação pois **você pode ter preferência por interagir através de drag and drop, com clique do mouse ou com a ajuda do teclado**. Você também gostaria de ter muitas opções de visualização e que experiência na troca de visualizações fosse simples. Além disso, você gostaria de poder realizar operações matemáticas com os dados, mudar as preferências de visualização, construir dashboards e compartilhar seus gráficos. Assim projetamos uma interface onde basta arrastar os dados para a área do gráfico ou clicar sobre eles que eles são automaticamente incluídos na sua visualização. Você também pode usar o teclado para ajudar na seleção. Uma janela chamada Show me exibe quando os gráficos estão ou não ativos e **lhe explica o porquê de cada visualização**. Para simplificar a troca das visualizações, o sistema transforma o dado de acordo com a visualização selecionada, **reorganizando as marcas das variáveis selecionadas, podendo até mudar características das variáveis**. Outra solução dada pelo sistema foi a **visão geral dos dados como forma de análise inicial**. Os detalhes são obtidos através da inclusão e filtragem dos valores.*

5.5. Discussão final

Através das análises realizadas, podemos responder as sub-questões de pesquisa relacionadas à avaliação, em ordem inversa.

SQ6. Quais foram as facilidades e dificuldades dos usuários em cada ferramenta?

As facilidades de cada ferramenta foram informadas nas entrevistas e entre as facilidades de ViSC encontramos respostas de que as perguntas facilitaram muito e a forma de filtrar também. No Tableau as respostas foram mais variadas e obtivemos seleção das dimensões, filtro, rapidez da visualização dos valores selecionados e a janela "Show me" que facilitou apresentando as miniaturas dos gráficos.

As dificuldades foram verificadas na interpretação da observação do uso das ferramentas e nas respostas informadas nas entrevistas. Entre as dificuldades informadas em entrevista no uso de ViSC, identificamos: a troca de dimensões feitas através do campo de seleção *click-insert* e a troca de visualizações devido à falta de imagens. Entre as dificuldades observadas, verificamos: a seleção do campo de variação (*checkbox*) que não demonstrou *feedback* em algumas situações; a falta de entendimento dos dados por alguns usuários; o não entendimento da função de colapsar e expandir um gráfico que não ficou evidente; e a dificuldade em relacionar gráficos abertos com os itens clicados do menu. Já no Tableau, os usuários informaram em entrevista que as dificuldades ocorreram devido ao grande volume de informações que dificultava encontrar as opções, entre elas o filtro. Entre as dificuldades observadas encontramos: a dificuldade em perceber a diferença entre dimensão e medida e a dificuldade em perceber que uma dimensão não incluída no gráfico, por padrão, era incluída com a soma dos valores o que dava uma visão geral dos dados. Isso gerou problemas de compreensão da ferramenta como um todo. Outros problemas observados foram as dificuldades de localização dos filtros e de entendimento da janela "*Show me*". Nenhum usuário viu as explicações que tornam uma opção ativa.

SQ5. De que forma o usuário interage com as perguntas da interface de ViSC?

Podemos observar quatro comportamentos diferentes em relação à interação com as perguntas. Primeiramente, usuários (U03 e U04) que viram as perguntas como uma oportunidade de realizar a tarefa de forma mais rápida, em segundo, usuários que não estavam tão satisfeitos ou tinham dúvidas em relação à solução dada (U05 e U07) e recorreram às perguntas como forma de validar ou melhorar o resultado obtido. Em terceiro, um usuário (U02) que utilizou como navegação sem ler a pergunta e por último um usuário (U08) que não interagiu com as perguntas.

SQ4. Como as perguntas estão sendo entendidas pelo usuário?

Basicamente eles entenderam que elas são geradas dinamicamente e ajudam na construção de visualizações eficientes, no entanto as classificações não foram compreendidas por todos.

Assim, podemos responder a pergunta principal da avaliação com os usuários:

SQ3. Como as perguntas influenciam na realização da tarefa e nas visualizações geradas?

Observamos que o resultado foi positivo ou neutro em todas as formas de interação. Em relação à tarefa, em alguns casos as perguntas agilizaram o processo, servindo como facilitador e em outros, foram importantes para conferir o resultado. Entre as visualizações geradas, os usuários que utilizaram as perguntas realizaram, em sua maioria, gráficos eficientes, com exceção de um caso no qual percebemos a necessidade de melhoria da ontologia. Em um caso percebemos que a resposta foi corrigida com o uso das perguntas e em dois outros, apenas reforçou o que já haviam feito. Para os usuários que chegaram a analisar as opções recomendadas para escolher, observamos que gerou confiança na resposta obtida, além de haver possibilidade de aprendizado nesta análise. Entre usuários que não utilizaram ou utilizaram como navegação, a influência foi neutra.

6 Conclusões

O objetivo desta pesquisa foi criar e avaliar uma solução **para verificar como ela apóia usuários inexperientes na construção de gráficos eficientes com dados estatísticos.**

Estudos acadêmicos sobre visualização de informação através de gráficos interativos e uma avaliação de ferramentas de visualização nos serviu de base para construir uma nova ferramenta que chamamos de ViSC, *Visualization with Smart Charts*. Em ViSC, procuramos cobrir pontos de IHC nos quais verificamos rupturas de comunicação na avaliação feita com usuários. Também utilizamos técnicas de apresentação automática, apresentação visual de *feedbacks* e tipologia de gráficos e gramáticas formais. O ponto mais importante, no entanto, foi a inclusão de recomendações através de perguntas mais comuns que o usuário possa querer responder sobre o dado relacionando às visualizações eficientes. Essas recomendações funcionam como uma ajuda, guiando o usuário na escolha de visualizações relacionadas ao dado no momento da sua interação com ferramenta. A técnica de recomendações utilizadas foi baseada em conhecimento, *knowledge-based recommender system*. O conhecimento do sistema foi então baseado em uma ontologia também desenvolvida durante a pesquisa.

A ferramenta ViSC foi avaliada com usuários através do Método de Inspeção Semiótica (MIS) e do *Retrospective Communicability Evaluation* (RCE). Exploramos como as perguntas recomendadas estavam sendo entendidas pelos usuários e como elas influenciavam os usuários na realização da tarefa e dos resultados gerados. Também fizemos uma avaliação com a ferramenta Tableau Public. Cada usuário que participou do teste realizou duas tarefas, uma em cada ferramenta de forma intercalada.

Os dois métodos nos ajudaram a identificar falhas no entendimento da ferramenta. Tanto com o MIS quanto com o RCE observamos problemas de IHC que podem ser tratados em ViSC. O MIS possibilitou encontrar a metagemensagem designer-usuário pelas ferramentas e revelou algumas inconsistências e

ambiguidades que poderiam trazer dificuldades ao usuário durante a interação. O *Retrospective Communicability Evaluation* possibilitou entender mais sobre os processos abduativos dos usuários de construção dos gráficos e no uso das ferramentas. Através dele encontramos diversas rupturas de comunicação, a partir das quais conseguimos identificar signos que influenciaram na interpretação dos usuários e falhas nas metagensagens emitidas por ambas as ferramentas. Identificamos ainda problemas de entendimento de dados e até mesmo comportamentos de usuários que desviaram por completo da tarefa. Além desses problemas de comunicabilidade, esse método, que integrou *Retrospective Think Aloud* e etiquetagem, possibilitou ainda encontrar pontos positivos das ferramentas que foram observados através de comentários espontâneos dos usuários.

Sobre as melhorias de IHC realizadas, em relação a seleção dos dados, observamos que foram satisfatórias. Comentários sobre a simplicidade de ViSC e a forma de seleção ou filtragem apareceram em respostas de quatro entre os seis usuários que realizaram a avaliação.

No nível da ferramenta, entre os signos do sistema de significação e comunicação de ViSC que atrapalharam o usuário encontramos: a troca de dimensões selecionadas feitas através do campo de seleção *click-insert*; a função de ordenação; a troca de visualizações devido à falta de miniaturas dos gráficos; o campo variação (*checkbox*) que não demonstrou *feedback* em algumas situações; a falta de entendimento dos dados por alguns usuários, o não entendimento da função de colapsar e expandir um gráfico que não ficou evidente e a falta de relacionamento entre os gráficos abertos com os itens clicados do menu.

Em relação ao sistema de significação e comunicação do segundo nível, representado pelo sistema gráfico, vimos mais uma vez que o único usuário que selecionou o gráfico de dispersão não compreendeu seu significado, tendo buscado uma solução mais elementar e mais parecida ao seu modelo mental em seu processo abduativo. Em relação aos gráficos empilhados (séries ou colunas), dois usuários, entre os três que construíram esse tipo de gráfico no ViSC, afirmaram ter preferido. No entanto nenhum dos três teriam pensado nessa solução sem apoio da ViSC.

Em relação à influência das perguntas na realização da tarefa e das visualizações geradas, chegamos a conclusões positivas sobre o seu uso. Dos seis

usuários que realizaram o teste, cinco utilizaram as perguntas e, destes, quatro consideraram as perguntas importantes e demonstraram isso através dos seus resultados e de seus comentários durante o *Retrospective Think Aloud* que revelou o motivo do uso das perguntas mesmo já quando já tinham uma solução. Tivemos dois casos onde a influência foi neutra. Observamos ainda que em dois casos, os usuários mantiveram o gráfico *default* do sistema sem procurar por outras soluções, ou seja, o gráfico inicial também teve influência.

A solução dada pelas perguntas tornou possível a realização da tarefa 1 com resultados eficientes. Os três usuários do grupo 1 utilizaram esse recurso. Os resultados obtidos na tarefa 1 com ViSC foram melhores que os obtidos pelos usuários do grupo 2 que utilizaram o Tableau para realizar a mesma tarefa. Entre os usuários do grupo 1, a tarefa 1, realizada com ViSC demonstrou melhores resultados. Apesar de, na tarefa 2 os resultados com ViSC não terem sido tão eficientes quanto na primeira ainda foram melhores que os resultados obtidos na tarefa 1 com Tableau por esses mesmos usuários. Ainda sim, um dos resultados que não foi tão eficiente foi exibido em uma pergunta relacionada, reforçando que alguns usuários ainda preferem recorrer ao caminho mais rápido, através das perguntas, que construir ou analisar opções. Esse caso demonstrou que a ontologia apresenta problemas que devem ser corrigidos.

Verificamos então, que as perguntas foram uma solução eficiente para apoiar os usuários inexperientes na construção de gráficos eficientes mas que ainda precisam de melhorias na ontologia para exibir melhores resultados.

6.1.1. Contribuições

Retornando às contribuições esperadas, nossa expectativa era no sentido de:

i) solucionar problemas de construção de gráficos por usuários inexperientes com uma ferramenta que permite vários tipos de construção e com operadores avançados através de recomendações.

Indiretamente pretendíamos também:

ii) avaliar a ontologia de visualização utilizada que define as relações entre dado, modelos de visualização e questões do usuário; e

iii) motivar usuários inexperientes a analisar dados graficamente.

Concluimos que a principal contribuição foi alcançada visto que os usuários utilizaram as perguntas e conseguiram obter resultados eficientes através de seu uso. Isso mostra que as perguntas foram uma solução eficiente para a construção de gráficos para os usuários inexperientes. As questões tiveram uma influência positiva e importante no resultado e para a avaliação da ferramenta como fácil de usar pelos usuários.

Além disso, alguns usuários realmente compararam as recomendações de visualização entre elas ou com uma construção previamente realizada por eles mostrando que, nesses casos, conseguimos motivá-los na análise dos dados através de gráficos em seu processo abduativo. Observamos ainda que pode ter havido aprendizado por parte desses usuários que analisaram através dos comentários feitos por eles.

Em relação à ontologia, vimos um caso em que a recomendação não se mostrou a mais eficiente. Concluimos que esta ainda deve ser melhorada e estendida.

6.1.2.Trabalhos futuros

A partir da análise dos resultados da avaliação, observamos a necessidade de ampliar as avaliações, incluindo outros grupos de usuários, e investigar as hipóteses que surgiram nessa avaliação. Primeiramente, surgiu a necessidade de avaliar usuários inexperientes e que não possuem uma pergunta bem definida. Todos os usuários avaliados tinham uma pergunta que foi dada pelo cenário, o que pressupõe que os usuários conheciam a base de dados e queriam construir visualizações para comunicar algo que já sabiam que existia na base. A partir dessa avaliação, identificou-se a oportunidade de avaliar usuários que navegam pelos gráficos de forma epistêmica, ou seja, que não têm uma pergunta *a priori* e querem buscar por novas informações a partir de um processo de extração do conhecimento (*knowledge discovery*). Assim, pretendemos avaliar se o sistema de recomendação de visualizações através de perguntas também é uma boa solução para esse segundo grupo de usuários. Com relação à ontologia, verificamos que a classificação das visualizações não foi observada por quatro entre os cinco participantes que utilizaram as perguntas. Entendemos que, apesar da tarefa dada, esses usuários estavam explorando as visualizações no processo de *knowledge*

discovery da ferramenta. Dessa forma, para as avaliações futuras que pretendemos, essas classificações talvez devam ser retiradas ou apresentadas de forma diferente.

Ainda planejamos aprofundar a investigação anterior com o objetivo de verificar, através de uma avaliação longitudinal, se e como a ferramenta pode ajudar no aprendizado da construção de visualizações. E, ainda, investigar quando e por que os usuários não usaram as perguntas.

Sobre a perspectiva de análise visual, podemos utilizar as visualizações para analisar a base de dados, ou seja, para verificar a qualidade dos dados

Em relação à ferramenta, incluiremos novas funcionalidades e utilizaremos mais recursos de interação para permitir uma melhor visualização e análise dos dados, como o uso de animações que podem ser utilizadas tanto para a inclusão de mais uma dimensão na visualização quanto para destacar parte da informação exibida. Quanto às recomendações, pretendemos implementar um mecanismo de *relevance feedback* para que possamos medir a relevância das questões para a eficiência da visualização apresentada. Para aprimorar as recomendações, uma opção seria criar formas híbridas de recomendações, através do conhecimento do perfil do usuário ou de filtragem colaborativa. Quanto à ontologia, pretendemos mapear construções improdutivas para retornar recomendações mais eficientes, considerando o número máximo de elementos selecionados em cada gráfico. Além disso, planejamos sua expansão com a inclusão de questões de tarefas compostas, tais como ordenação pela média o que seria a combinação das tarefas ordenação com valor derivado e encontrar o maior valor depois de um dado elemento, uma combinação de retornar extremo e retornar valor.

7 Referências bibliográficas

- 1 TORY, M; MÖLLER, T. **Rethinking Visualization: A High-Level Taxonomy.** IEEE Symposium on. INFOVIS 2004. Visualization, 2004.
- 2 HEER, J.; BOSTOCK, M; OGIEVETSKY, V. **A Tour Through the Visualization Zoo.** Communications of the ACM, 2010.
- 3 CARZOLA, I. **A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos.** SP, Campinas: Unicamp, 2002.
- 4 PINKER, S. **A theory of graph comprehension.** In R. Freedle (Ed.), Artificial intelligence and the future testing. Hillsdale, NJ: Erlbaum, 1990. pp. 73-126.
- 5 SHAH, P. e CARPENTER, P. A. **Conceptual limitations in comprehending line graphs.** Journal of Experimental Psychology: General, 124 (1), pp. 43-61. 1995.
- 6 MANOVICH, L. **Data Visualisation as New Abstraction and Anti-Sublime.** Berlin, 2002. Disponível em: <http://manovich.net/DOCS/data_art_2.doc>. Acesso em: 08/03/2013.
- 7 TUFTE, E.R. **The Visual Display of Quantitative Information.** Second Edition. USA. Graphics Press. 2001.
- 8 GOLDENBERG, E. P. **Mathematics, metaphors, and human factors: Mathematical, technical, and pedagogical challenges in the educational use of graphical representation of functions.** The Journal of Mathematical Behavior, Vol 7(2), pp. 135-173. 1998.
- 9 CLEMENT, J. **Misconceptions in Graphing.** Proceeding 9nd Annual Meeting of the International Group for the Psychology of Mathematics Education, pp. 369-375, 1985.
- 10 GOMES FERREIRA, V. G. **Exploring Mathematical Functions Through Dynamic Microworlds.** 1997, 353 f. Thesis (Education). Institute Education, Universidade de Londres. Londres, 1997.

- 11 Portal do Professor. Disponível em:
 <<http://portaldoprofessor.mec.gov.br/fichaTecnicaAula.html?aula=23488>>
 Acesso em: 10/01/2012.
- 12 BIM, S. **Obstáculos ao Ensino dos Métodos de Avaliação da Engenharia Semiótica**. PUC-Rio. Tese. 2009.
- 13 ELIAS, M.; BEZERIANOS, A. **Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices**. Interact 2011. Part IV, pp. 274-291. 2011.
- 14 HEER, J., VIÉGAS, F. B., WATTENBERG, M. **Voyagers and voyeurs: Supporting asynchronous collaborative information visualization**. In: Proc. of the Conference on Human Factors in Computing Systems (CHI), 2007, pp. 1029–1038.
- 15 CUBRARNIC, D. **Polstar: Assisted navigation for exploring multi-dimensional information spaces**. In: Human-Computer Information Retrieval (HCIR), 2008.
- 16 MACKINLAY, J.D.; HANRAHAN, P.; STOLTE, C. **Show Me: Automatic Presentation for Visual Analysis**. IEEE. Transactions on Visualizations and Computer Graphics, vol. 13, number. 6, november/ december 2007.
- 17 SOUSA, T. 2012. **How signification and communication systems influence the interpretation of statistical data by users with specific information needs**. PUC-Rio. Monographs in Computer Science, number 15/12. 2012.
- 18 DE SOUZA, Clarisse S. **The semiotic engineering of human-computer interaction**. Cambridge: Mass. The MIT Press. 2005. Capítulos 1, 2 e 4.
- 19 BERTIN, J. **Semiology of Graphics: Diagrams, Networks, Maps**. ESRI first edition 2011. 1918.
- 20 DE SOUZA, C. S.; LEITÃO, C. F. **Semiotic Engineering Methods for Scientific Research in HCI**. Morgan & Claypool Publishers. 2009.
- 21 ERICSSON, K. A.; SIMON, H.A. **Protocol Analysis**. MIT. 1993.
- 22 WILSON, C. **Usability techniques: Pros and cons of co-participation in usability studies**. Usability interface. 1998.
- 23 FEW, Stephen. **Show me the numbers. Designing tables and graphs to enlighten**. Analytics Press, Oakland, California. 2004.
- 24 SOUSA, T. 2012. **Semantic characterization of visualization mechanisms**. PUC-Rio. Monographs in Computer Science, number 16/12. 2012.

- 25 DE SOUZA et al. **The Semiotic Inspection Method**. Proceedings of the VII Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais, IHC. pp. 148-157. 2006.
- 26 AMAR, R., EAGAN, J., STASKO, J. **Low-Level Components of Analytic Activity in Information Visualization**. IEEE Symposium on Information Visualization, 2005.
- 27 HAAK, M. J. V. V.; JONG, M. D. T.; SCHELLENS, P. J. **Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue**. Behaviour & Information Technology, 2003.
- 28 DE SOUZA, C. S.; LEITÃO, C. F. **A method for evaluating the communicability of Users Interface**. Morgan & Claypool Publishers, 2009.
- 29 IGNATIUS, E.; SENAY, H. **Rules and Principles of Scientific Data Visualization**. Technical Report GWU-IIST-90-13. Institute for Information Science and Technology, The George Washington University, 1990.
- 30 BARBOSA, S. D. J. & SILVA, B. S. **Interação Humano-Computador**. Campus/Elsevier, 2010.
- 31 Stanford Encyclopedia of Philosophy. **Abduction**. Disponível em: <<http://plato.stanford.edu/entries/abduction/>> Acesso em: 23/10/2011.
- 32 SANTAELLA, L. **O método anticartesiano de C. S. Peirce**. Editora UNESP, 2004.
- 33 HEER, J.; SHNEIDERMAN, B. **Interactive Dynamics for Visual Analysis. A taxonomy of tools that support the fluent and flexible use of visualizations**. ACM queue, 2012.
- 34 BOSTOCK, M.; HEER, J. **Protovis: A Graphical Toolkit for Visualization**. InfoVis 2009, Visualization and Computer Graphics, IEEE Transactions on, 2009.
- 35 RICCI, F. et al. **Recommender System Handbook**. Springer, 2011.
- 36 BURKE, R. **Hybrid Recommender Systems: Survey and Experiments**. User Modeling and User-Adapted Interaction 12: pp. 331-370, 2002.
- 37 BURKE, R. **Knowledge-based recommender systems**. University of California, Irvine. Encyclopedia of Library and Information Systems, volume 69. Marcel Dekker, New York, 2000.
- 38 RICH, E. **User Modeling via Stereotypes**. Cognitive Science 3, pp. 329-354. 1979.

- 39 BRODLIE K.W.; NOOR, N. M. **Visualization Notations, Models and Taxonomies**. EG UK Theory and Practice of Computer Graphics, 2007.
- 40 DUKE, D. J.; BRODLIE K.W.; DUCE, D..A; HERMAN, I. **Visualizations Viewpoints - Do you see what I mean?** Computer Graphics and Applications, IEEE Computer Society, 2005.
- 41 DUKE, D. J.; BRODLIE, K. W.; DUCE, D. A. **Building an Ontology of Visualization**. Proceedings of the conference on Visualization 04, IEEE Computer Society, 2004.
- 42 National e-Science Center. **Visualization Ontologies**. Disponível em: <http://www.nesc.ac.uk/talks/393/vis_ontology_report.pdf>. Acesso em: 13/11/2011.
- 43 **Visko Visualization Knowledge**. Disponível em: <<http://trust.utep.edu/visko/dl/>>. Acesso em: 13/11/2011.
- 44 DEL RIO, N; DA SILVA, P. P. **Visualization Queries**. Disponível em: <<http://trust.utep.edu/visko/docs/docs/visquery.pdf>>. Acesso em: 12/11/2011.
- 45 STEELE, J; LLIINSKY, N. **Beautiful Visualization. Looking at Data through the Eye of Experts**. O'Reilly, 2010.
- 46 MACKINLAY, J. D. **Automating the design of graphical presentations of relational information**. ACM Trans. on Graphics, 5 (2): 110-141. Apr. 1986.

APÊNDICE A

Nesta seção se encontram outros resultados do estudo preliminar descrito na seção 3.1

Outros resultados obtidos no primeiro estudo preliminar (seção 3.1)

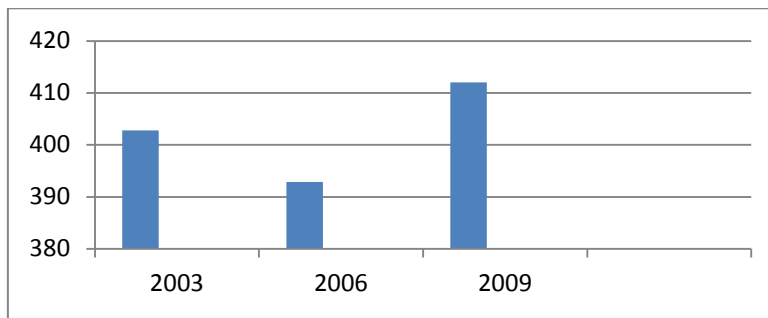


Figura 80: Resultado de A01 para tarefa 2 na pesquisa [17] com apoio do Excel.

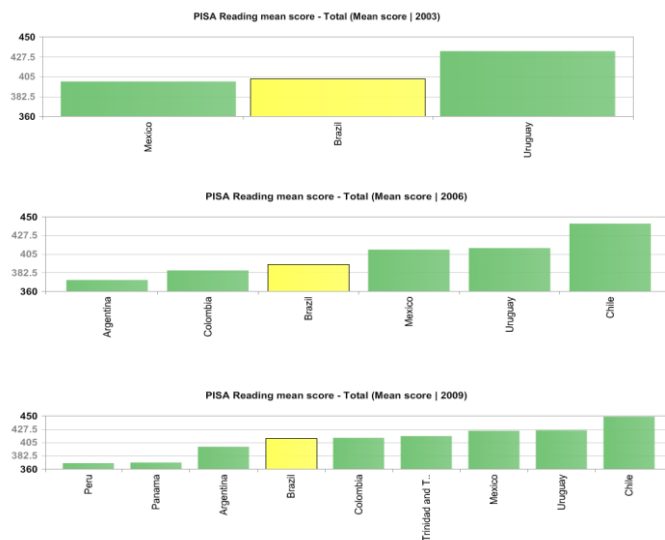


Figura 81: Resultado de A02 para tarefa 2 na pesquisa [17]. Usuário não conseguiu unir as 3 colunas em um único gráfico

APÊNDICE B

Nesta seção se encontram o questionário pós-teste e o roteiro da entrevista.

Formulário pós-entrevista

Perguntas referentes à tarefa 1: () Tableau/ () ViSC

1) Como você classifica o grau de dificuldade para a realização da tarefa?

1. () Muito difícil
2. () Difícil
3. () Regular
4. () Fácil
5. () Muito fácil

2) Interagiu com _____?

1. () Nem vi essa janela
2. () Vi que ela existia mas não interagi
3. () Passei o mouse sobre as opções mas não senti necessidade de clicar
4. () Cliquei nas opções mas não entendi seu funcionamento
5. () Cliquei nas opções e entendi o funcionamento

3) Se marcou 3, 4 ou 5 em pelo menos uma das colunas da questão acima, responda: Como elas influenciaram na realização das tarefas?

1. () Atrapalharam muito
2. () Atrapalharam pouco
3. () Nem atrapalhou nem ajudou
4. () Ajudou um pouco
5. () Ajudou muito

4) O gráfico gerado foi o mesmo que tinha em mente inicialmente?

1. Sim
2. Não (Justifique)_____

5) Como você classifica a qualidade do gráfico gerado?

1. Não concluí a tarefa.
2. Desenhei um gráfico que não achei muito adequado mas era o que a ferramenta permitia.
3. Desenhei o gráfico recomendado porque a ferramenta sugeriu mas prefiro o que tinha pensado.
4. Desenhei o gráfico igual ao que tinha em mente após analisar as opções sugeridas porque achei mais claro.
5. Desenhei o gráfico recomendado e achei melhor.

Perguntas referentes à tarefa 2: Tableau/ ViSC

1) Como você classifica o grau de dificuldade para a realização da tarefa?

1. Muito difícil
2. Difícil
3. Regular
4. Fácil
5. Muito fácil

2) Interagiu com _____?

1. Nem vi essa janela
2. Vi que ela existia mas não interagi
3. Passei o mouse sobre as opções mas não senti necessidade de clicar
4. Cliquei nas opções mas não entendi seu funcionamento
5. Cliquei nas opções e entendi o funcionamento

3) Se marcou 3, 4 ou 5 em pelo menos uma das colunas da questão acima, responda: Como elas influenciaram na realização das tarefas?

1. Atrapalharam muito
2. Atrapalharam pouco
3. Nem atrapalhou nem ajudou
4. Ajudou um pouco

5. () Ajudou muito
- 4) O gráfico gerado foi o gráfico que tinha em mente inicialmente?
1. () Sim
 2. () Não (Justifique)_____
- 5) Como você classifica a qualidade do gráfico gerado?
1. () Não concluí a tarefa.
 2. () Desenhei um gráfico que não achei muito adequado mas era o que a ferramenta permitia.
 3. () Desenhei o gráfico recomendado porque a ferramenta sugeriu mas prefiro o que tinha pensado.
 4. () Desenhei o gráfico igual ao que tinha em mente após analisar as opções sugeridas porque achei mais claro.
 5. () Desenhei o gráfico recomendado e achei melhor.

Roteiro de entrevista

- 1) Em que a ferramenta X te ajudou na realização da tarefa? Em que ela mais dificultou?
 - 2) O que você mais gostou em cada ferramenta?
 - 3) Que gráficos você pensou em construir no início de cada tarefa?
 - 4) O que mais lhe influenciou para a escolha dos gráficos?
 - 5) Qual foi o seu entendimento sobre a janela "Show me" da ferramenta Tableau? Pode explicar em que ela ajudou ou em que atrapalhou?
- As próximas perguntas se referem à ferramenta ViSC
- 6) Qual foi o seu entendimento sobre as perguntas exibidas da ferramenta ViSC? E o que achou delas?
 - 8) Se não interagiu com as perguntas, explique o porquê?
 - 9) Se interagiu, como considera a importância das perguntas para encontrar o gráfico? Pode explicar em que elas ajudaram ou em que atrapalharam?
- (Se algum gráfico foi diferente do gráfico desenhado)

10) O gráfico gerado pela ferramenta para a tarefa X foi diferente do gráfico desenhado antes do teste para uma tarefa similar. Por que você não desenhou gráfico semelhante?

(Se algum gráfico foi similar do gráfico desenhado e não escolheu o mais pontuado no caso de ViSC)

12) O gráfico gerado pela ferramenta para a tarefa X foi similar ao gráfico desenhado antes do teste para uma tarefa similar. Por que você não selecionou uma das visualizações recomendadas pela ferramenta?

APÊNDICE C

Nesta seção se encontram a tabela com as perguntas listadas no segundo estudo preliminar e o conjunto de modelos que relaciona as perguntas e as tarefas utilizados na ferramenta ViSC.

Tabela 30: Relação entre questões, tarefas, e representações visuais

Questions	Visualization	Task
Is there any correlation between v_1 and v_2 ?	Scatterplot	Correlate N pairs of quantitative variables.
Is there any correlation between v_1 , v_2 and v_3 ?	Bubble chart	Correlate N triples of quantitative variables.
What element N had the highest (or the lowest) value of a category M?	Clutstered Column Chart	Find extreme between N quantitative variables of M categories
Which category M had the highest (or lowest) value in time t	Index timeseries	
Which category N (place) had the highest (or lowest) value?	Cartogram	Find extreme between N quantitative values
What element has the highest (or the lowest) perceptual value of a category?	100% stacked column chart	Find extreme in perceptual value between N quantitative variables of M categories
What element has the highest (or the lowest) sum of its M categories?	Stacked column Chart	Derived value of sum from the M quantitative values of categories and find extreme between N qualitative variables.
What is the value of a variable v_1 of class m_1 in time t?	Index time series	Retrieve a quantitative value
When did variable v_1 of M_1 class have value X?	Index time series	Retrieve index of N values
	Horizon time series	
Which categories had the highest (or lowest) variation of a category in the interval t_1 to t_n ?	Index time series	Visualize trends, changes, increase, fluctuation, growth, decline of a quantitative variable V of M classes
	Multiple time series	
	Stacked time series	
Which places belong to a category M?	choroplethMap	Compare M categories of a qualitative variable between N categories of a geographic component
In which category, place X is classified?	choroplethMap	
What is the perceptual of the category X in the place Y?	graduated Symbol Map	Visualize part of a whole of M quantitative values of N categories that is a geographic component
How many level has the hierarchy?	Radial Icicle Sunburst	Find steps
Who are the children of category N?	Radial Cartesian Indented Icicle Sunburst Treemap Nested	Find subclasses
What are the leaves of the hierarchy?	Cartesian Treemap	Find leaves

Conjunto de modelos que representam as condições necessárias para cada pergunta ser ou não apresentada.

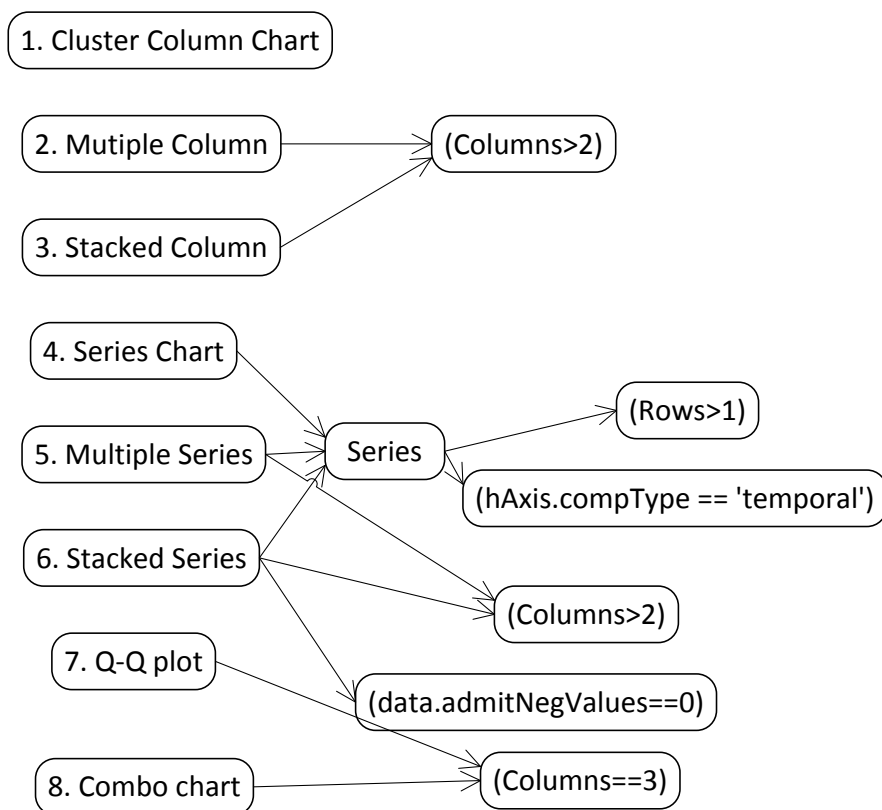


Figura 82: Pré-condições que definem que visualizações podem ser mapeadas de acordo com características do dado

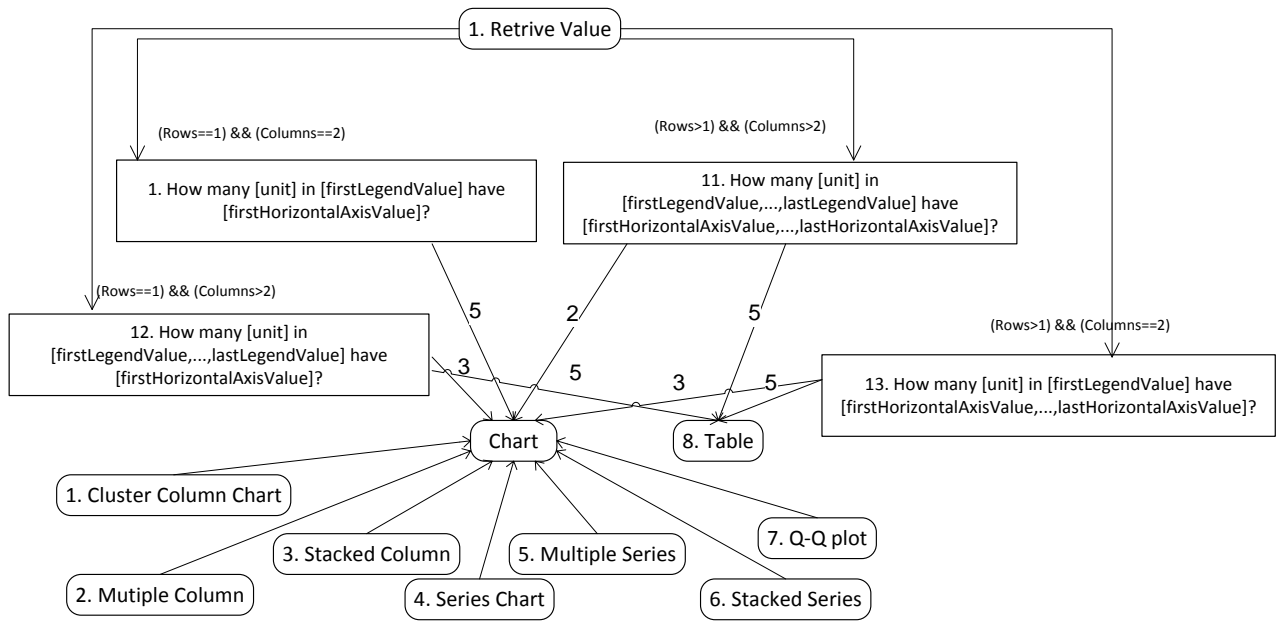


Figura 83: Condições para a tarefa1: retornar valor.

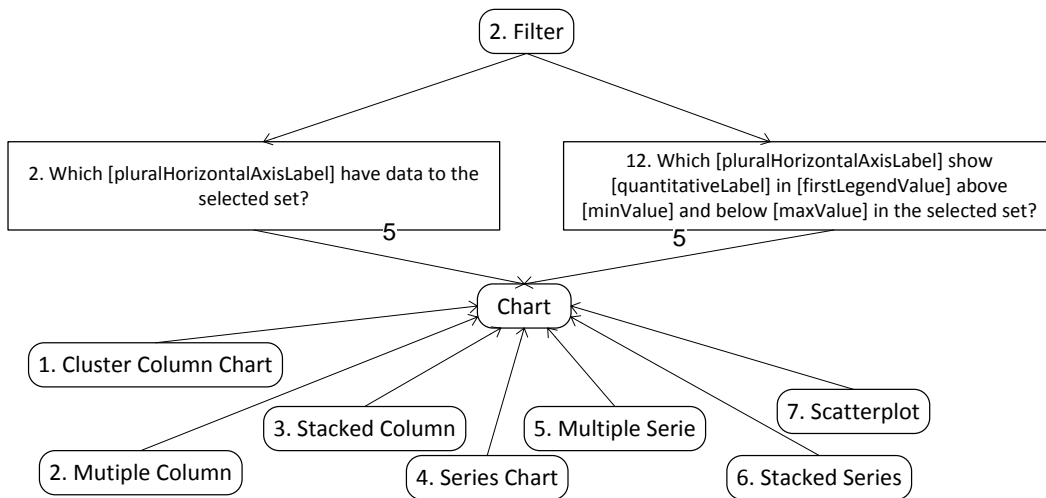


Figura 84: Condições para a tarefa 2: filtrar.

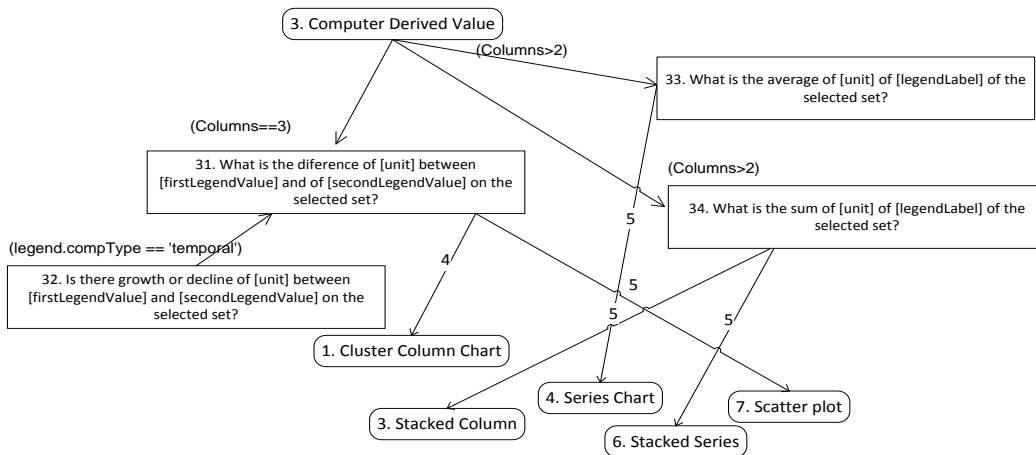


Figura 85: Condições para a tarefa 3: calcular valor derivado.

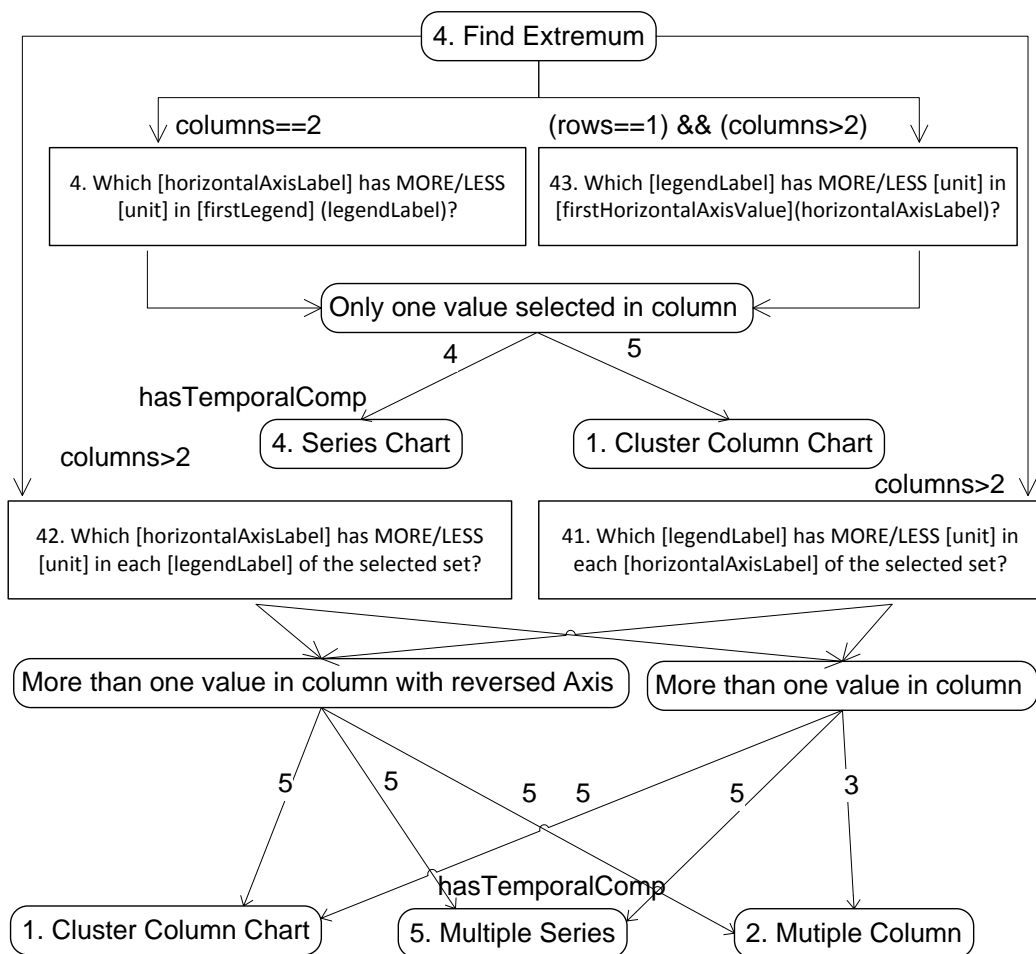


Figura 86: Condições para a tarefa 4: achar extremo.

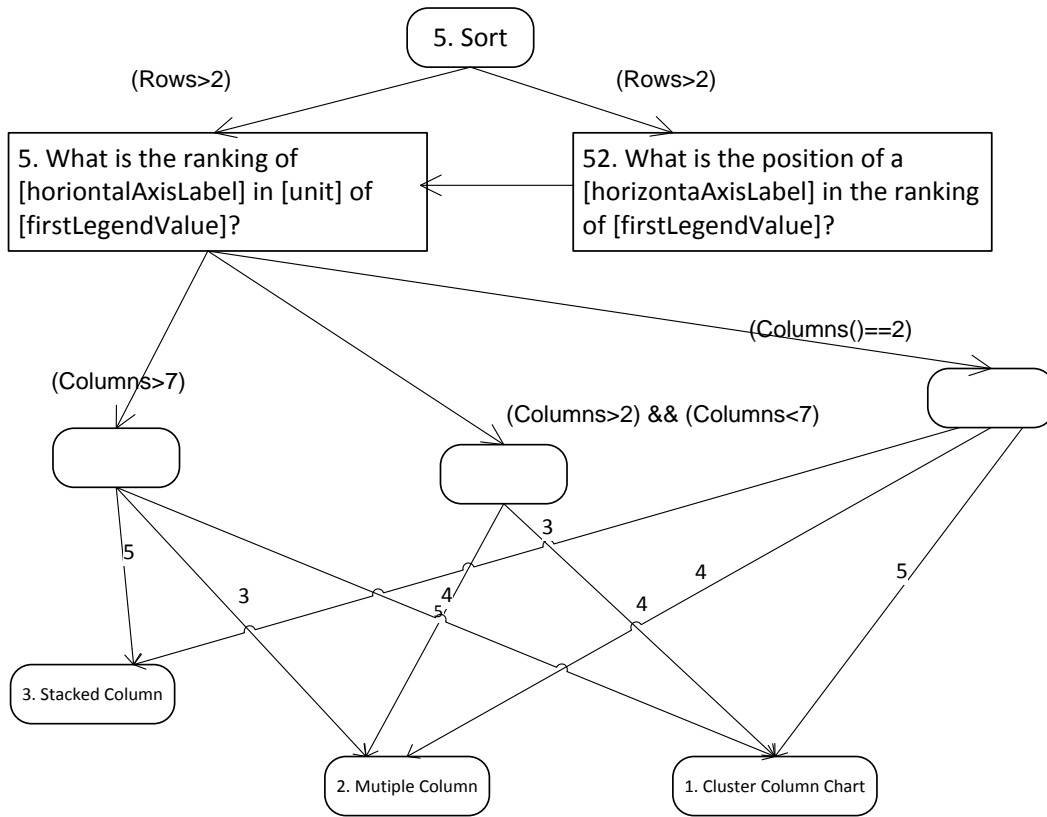


Figura 87: Condições para a tarefa 5. ordenar

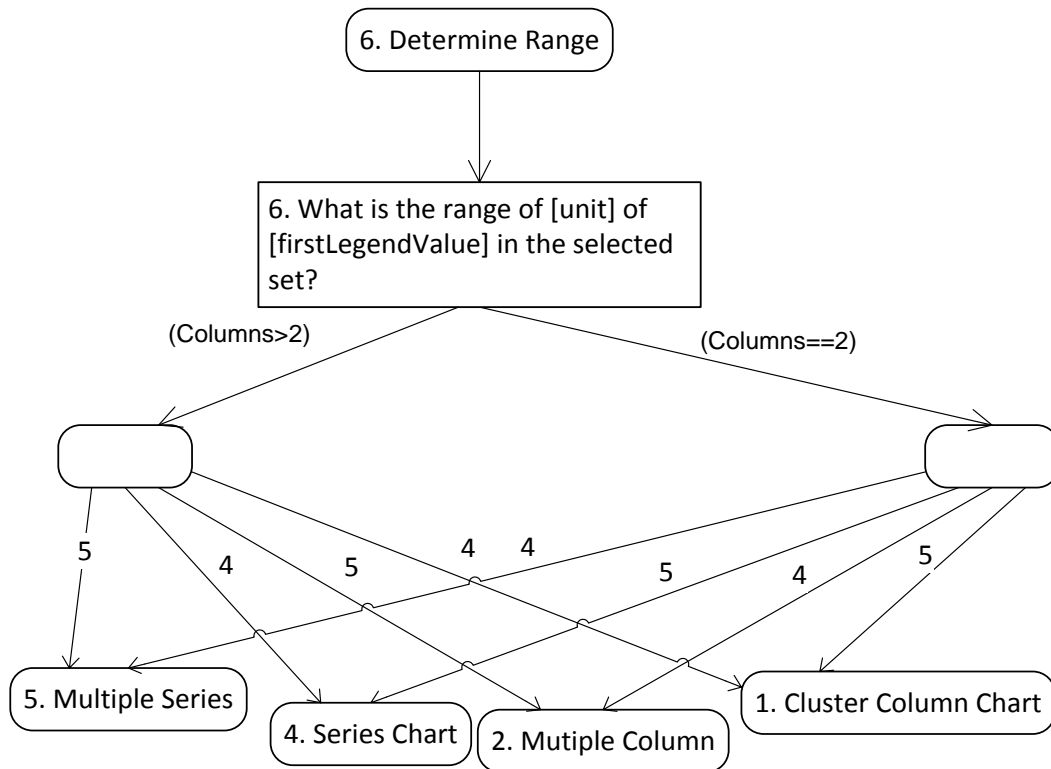


Figura 88: Condições para a tarefa 6: determinar intervalo.

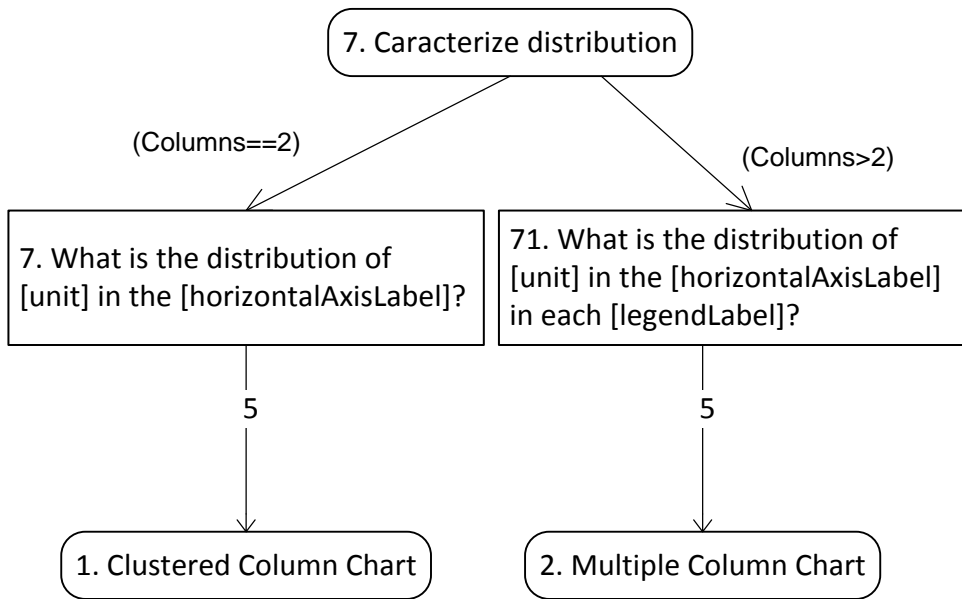


Figura 89: Condições para tarefa 7: caracterizar distribuição.

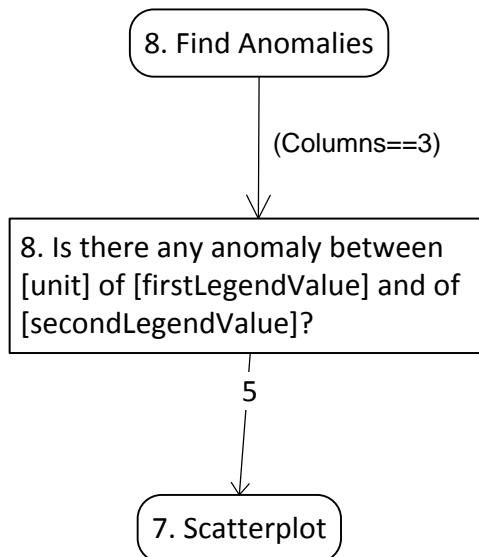


Figura 90: Condições para a tarefa 8: encontrar anomalia.

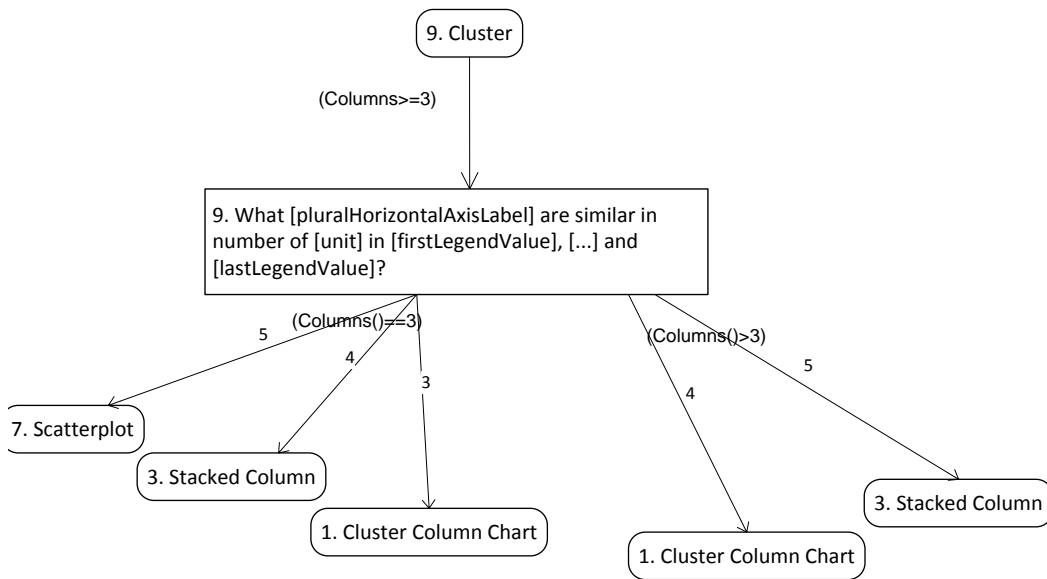


Figura 91: Condições para tarefa 9: agrupar.

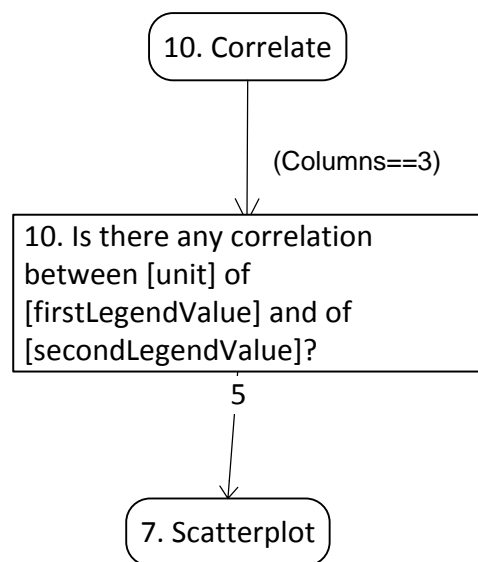


Figura 92: Condições para tarefa 10: correlacionar.