

Apoiando a construção de gráficos por usuários inexperientes através de recomendações

Taissa Abdalla Filgueiras de Sousa
Departamento de Informática
Pontifícia Universidade Católica, PUC-Rio
22451-900 Rio de Janeiro, Brasil
+55 (21) 35271510
tsousa@inf.puc-rio.br

Simone Diniz Junqueira Barbosa
Departamento de Informática
Pontifícia Universidade Católica, PUC-Rio
22451-900 Rio de Janeiro, Brasil
+55 (21) 35271500 Ramal: 4353
simone@inf.puc-rio.br

ABSTRACT

Visualization research emphasizes the need for systems that assist in decision-making and visual analysis. This paper outlines a recommender system that supports the interactive construction of charts using statistical data. Based on a sequence of simple user interactions, the interface automatically generates an elementary view. This process yields knowledge in an implicit manner, which enables the system to analyze the selected dataset, suggest visualization options and provide guidelines for selecting the most appropriate representation for a given problem.

Categories and Subject Descriptors

H.5.0 [Information Systems]: Information Interfaces And Presentation - General

General Terms

Design

Keywords

information visualization, recommender systems, human-computer-interaction

1. INTRODUÇÃO

Essa pesquisa visa contribuir para a área de interação-humano-computador (IHC) através de um sistema de visualização de informação que agrega recomendações de tarefas para apoiar a construção de gráficos para usuários inexperientes.

Gráficos são um meio de comunicação eficiente que fazem uso da percepção. No entanto, é possível confundir a percepção do leitor através da distorção das representações. Nesse sentido, Tufte [12] descreve as técnicas mais comuns. Assim, para uma interpretação correta, pesquisas demonstram que a interpretação de gráficos requer conhecimentos específicos que não são facilmente aprendidos [11][13][14]. Considerando ainda as elevadas taxas de analfabetismo funcional, problemas no ensino médio e o consumo de informações sem questionamento podemos verificar que existem problemas na área de visualização de informações, especialmente por usuários inexperientes. Motivados por esse problema, nossa proposta é criar uma ferramenta que auxilie na construção das visualizações com a utilização de dados estatísticos.

Este artigo está dividido da seguinte forma: A seção 2 cita ferramentas e *toolkits* de visualização disponíveis e descreve suas limitações. Além disso faz uma análise de técnicas utilizadas em

projetos de interfaces de visualização de dados. A seção 3 descreve o problema que motivou a pesquisa. A seção 4 apresenta o objetivo da proposta. A seção 5 listas as contribuições esperadas para a área. A seção 6 descreve as metodologias. A seção 7 faz uma breve descrição de pesquisas já realizadas a fim de atingir o objetivo proposto e em que estágio o trabalho se encontra. A seção 8 compara a proposta com projetos recentes. Por último, a seção 9 conclui o artigo.

2. CONTEÚDO TEÓRICO

Diversos sistemas têm sido desenvolvidos com o objetivo de melhorar a experiência do usuário na visualização de dados e motivar o interesse por dados estatísticos. Entre eles, podemos citar: Manyeyes,¹ GapMinder,² Worldmapper,³ Statplanet,⁴ Google Public Data,⁵ diversos atlas multimídia, SIDRA,⁶ Séries estatísticas,⁷ etc. Elias e Bezeriano [1] citam algumas *toolkits* (Flare,⁸ Silverlighth,⁹ JavaScript InfoVis tk,¹⁰ ivtk¹¹) que permitem tanto a criação de gráficos quanto de combinações deles. Afirmam que elas 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. Elias e Bezeriano [1] afirmam ainda que as ferramentas para usuários novatos tais como ManyEyes, Sense.us e Polstar, normalmente restringem usuários a uma única visualização por vez.

Heer e Shneiderman [4] apresentam três alternativas utilizadas em projetos de interfaces de visualização de dados. A primeira é através de *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 segunda alternativa apresentada é através do uso de *data-flow graphs*, no

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

² <http://www.gapminder.org/>

³ <http://www.worldmapper.org/>

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

⁵ <http://www.google.com/publicdata/directory>

⁶ <http://www.sidra.ibge.gov.br/>

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

⁸ <http://flare.prefuse.org/>

⁹ <http://www.silverlighth.net/>

¹⁰ <http://thjit.org/>

¹¹ <http://ivtk.sourceforge.net>

qual o processo de visualização é desconstruído em conjuntos refinados de operadores para importação, transformação, layout, etc. Através das combinações flexíveis, eles permitem maior variação de designs, porém requerem um maior esforço de entrada que o método anterior e ainda são limitados pelo conjunto de operadores. A terceira alternativa é baseada em *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*,¹² *Protovis for HTML5* [5] e *Google Chart Tools*.¹³ No entanto, todas elas requerem um mínimo de habilidade em programação. Ainda segundo Heer e Shneiderman [4], essas três alternativas não são mutuamente exclusivas. Pode-se aplicar um sistema de *data-flow* ou *formal grammars* para definir novos componentes para incluir dentro de uma *chart typology*, aproveitando melhor a expressividade da primeira e da facilidade de utilização da última.

Pesquisas sobre visualização ainda demonstraram a necessidade de criar significados formais sobre as formas de visualizações de dados [6][7][8]. Tory e Torsten [6] 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. Amar et al. [23] apresentam um conjunto de dez tipos de tarefas de análise primitiva que representam os tipos de questões específicas que uma pessoa pode perguntar ao trabalhar com um conjunto de dados. Duke et al. [7][8] destaca a necessidade de uma ontologia para visualização que suporta trabalhos colaborativos e educação.

3. IDENTIFICAÇÃO DO PROBLEMA

Verificamos que existem problemas na interpretação e construção de gráficos [11][13][14][3] e necessidade de ferramentas que apoiem essas tarefas por usuários novatos [1][9]. Mackinlay et al. [9] confirmam a necessidade quando afirmam que todos os analistas têm conhecimento sobre o domínio de seus problemas, mas poucos têm habilidade para construir apresentações gráficas eficientes da informação. Acrescenta ainda que as pessoas precisam de sistemas de análises visuais que automaticamente apresentam dados usando as melhores praticas do design gráfico.

4. OBJETIVO

Pelo motivo apresentado, buscamos uma solução que auxilie usuários inexperientes na construção de visualizações. Nossa pesquisa visa o desenvolvimento de um sistema de visualização que auxilie usuários na construção de gráficos eficientes com dados estatísticos, evitando construções que levam a interpretações equivocadas. A eficiência da representação é definida por Bertin [16] pela seguinte proposição: "se, para tentar obter 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" [16, p.139]. Bertin completa ainda que "as construções mais eficientes são aquelas nas quais quaisquer questões, não importando tipo ou nível,

podem ser respondidas num simples instante da percepção, ou seja, em uma só imagem" [16, p. 146].

5. CONTRIBUIÇÕES ESPERADAS

Como descrito na seção 2, existem diversas ferramentas de visualização. No entanto, as ferramentas que possibilitam maior interação são voltadas para usuários mais experientes e muitas vezes até necessitam de conhecimento de programação por parte deles. As ferramentas voltadas para usuários inexperientes possuem muitas restrições, como limitação de interação com dados e limitação nas construções de gráficos.

Portanto, nossa contribuição direta será: i) na resolução de problemas de visualização e construção de gráficos por usuários inexperientes com uma ferramenta que permite diversos tipos de construções e operações avançadas através das recomendações; e indiretamente: ii) na avaliação da ontologia de visualização que define relações entre dados, modelos de visualização e perguntas dos usuários; iii) no sentido de estimular usuários inexperientes a uma análise das representações gráficas.

6. METODOLOGIA ADOTADA

O sistema de visualização proposto integra o conceito de *content-based recommender system*. O processo básico executado por esse sistema consiste na combinação de atributos do perfil do usuário com atributos do conteúdo do objeto para recomendar ao usuário novos itens de interesse [25]. Assim, a partir da seleção de um dado pelo usuário, o sistema apresentará perguntas que podem ser respondidas com a visualização apresentada e recomendará outras visualizações eficientes. As visualizações possíveis e as perguntas têm relação direta com o dado selecionado. O conhecimento dessa relação foi definido através de uma ontologia de visualização [2], explicada mais adiante.

As perguntas estão sendo elaboradas com base nas dez tarefas de [23]. Elas visam criar um diálogo com o usuário em linguagem natural para que ele identifique com mais facilidade as visualizações relacionadas ao dado desejado. A cada interação do usuário, as perguntas exibidas podem ser alteradas ou receber um destaque diferente na interface devido à sua relevância para o dado selecionado. Dessa forma, o sistema permitirá que o usuário encontre com mais facilidade visualizações eficientes para responder sua pergunta.

Como mecanismo de *relevance feedback*, o usuário retorna um *feedback* implícito para o sistema a cada interação e demonstra a finalização da realização de uma tarefa ao exportar uma visualização. Nesta hora pretendemos incluir ainda um tipo de *feedback* explícito para verificar se a opção selecionada tem relação com as perguntas mais relevantes sugeridas.

Para verificar se objetivo foi alcançado, realizaremos testes de IHC junto a usuários utilizando os métodos utilizados em pesquisa anterior [3]: Método de Avaliação de Comunicabilidade (MAC) [18, p.344] e *think aloud* [19] com co-participação [20]. Os testes de IHC realizados avaliaram as ferramentas Statplanet, SIDRA e Séries estatísticas e estão descritos na seção 7. A avaliação da ferramenta será através da utilização de um cenário que possibilite o uso de ViSC 2 (Visualization with Smart Charts 2) e das demais ferramentas avaliadas anteriormente [3] para a obter uma comparação entre as ferramentas e o novo sistema. O usuário, conforme avaliações anteriores, poderá construir um gráfico em mais de um sistema e optar pelo uso de um deles. Isso,

¹² <http://had.co.nz/ggplot2/>

¹³ <https://developers.google.com/chart/interactive/docs/index>

porque além de avaliar a comunicabilidade de ViSC 2, identificar melhorias e avaliar se evoluímos no sentido possibilitar a construção de visualizações eficientes, temos o objetivo também de fazer uma análise comparativa entre as ferramentas.

7. ESTÁGIO ATUAL DO TRABALHO

Esta proposta foi fundamentada em duas pesquisas anteriores e no desenvolvimento de uma primeira versão da ferramenta ViSC como proposta preliminar.

7.1 Avaliação das ferramentas de visualização

Para avaliar como ferramentas de visualização interativas podem influenciar a interpretação dos dados [3], utilizamos dois métodos de avaliação com usuário e escolhemos as ferramentas Statplanet, SIDRA e Séries estatísticas. O MAC foi fundamental para encontrar signos que atrapalhavam na construção dos gráficos verificados através das rupturas de comunicação entre a metamsagem do designer e a recebida pelo usuário. O método *Think Aloud* com co-participação, possibilitou entender sobre os processos abducativos na escolha das visualizações pelos usuários, seus modelos mentais e problemas de compreensão na leitura de gráficos.

Para Santaella [22], 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” [22, 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” [22, p. 92].

Observamos que temos maior complexidade dos dados e flexibilidade de construções, maior foi a dificuldade encontrada pelos usuários. As ferramentas mais simples, no entanto, permitiram pouca interação e limitaram 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. Encontramos diversos problemas de IHC em ferramentas de visualização que podem atrapalhar a interpretação dos dados e nas construções gráficas pelo usuário inexperiente.

7.2 Criação da Ontologia de Visualização

Esta pesquisa visava melhorar a comunicação entre usuários e sistemas de visualização. Vimos que a formulação da pergunta era um dos critérios para uma visualização eficiente [17]. Sobre o número de questões em cada visualização, Bertin [16] afirma que existem tantas questões quanto o número de dimensões. Cada dimensão produz um tipo de questão. Assim, cada visualização responde a determinadas perguntas. A pergunta que o usuário quer responder tem relação com a tarefa e com a eficiência da representação tem relação.

Visko [24] e a ontologia de visualização da UK National e-Science Center [7][8] serviram de base para a criação da nova ontologia. Nosso esforço foi no sentido de criar uma ontologia

que definisse as relações entre dados e modelos de visualização a fim de responder perguntas de usuários através de visualizações eficientes. Definimos então 5 classes de nível superior: dado, atributo de exibição, visualização, transformação e tarefa (definida por perguntas).

7.3 ViSC

Com a primeira versão da ferramenta ViSC buscamos resolver problemas apresentados na primeira pesquisa e avaliar a ontologia desenvolvida na segunda. ViSC já agregou um sistema de recomendação. As recomendações no ViSC são apresentadas através de perguntas pertinentes às visualizações e ao dado selecionado de acordo com a ontologia [2]. As perguntas são exibidas segundo a relevância para a visualização apresentada. Elas foram cadastradas e classificadas com pesos de acordo com uma condição de seleção. Enquanto as recomendações das visualizações dependem das características semânticas do dado, as perguntas verificam, além das visualizações disponíveis, o número de elementos selecionados em cada atributo de exibição e as visualizações selecionadas pelo usuário, atribuindo mais peso para perguntas que respondem a estas visualizações. Na interface, o grau de relevância das perguntas é comunicado pela localização na tela e pela gradação da cor. Quanto mais no topo da página e mais escuro é o verde do fundo da pergunta, mais relevante.

A Figura 1 apresenta a tela de construção de visualizações de ViSC. A área 1 corresponde ao menu de cada uma das dimensões do dado. A área 2 apresenta o menu de visualizações da ferramenta. Os itens desse menu são habilitados ou desabilitados de acordo com o dado selecionado. A área 3 exibe os botões de ordenação, a área 4 é onde se localizam os gráficos e a área 5 onde as perguntas estão localizadas.

Essa ferramenta foi avaliada através do Método de Inspeção Semiótica (MIS) [18, p.330] e identificamos necessidades a serem corrigidas antes de fazer avaliações com usuários.

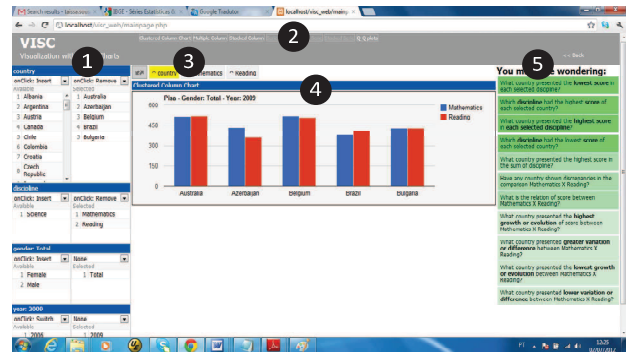


Figura 1: Tela de construção de visualizações da ferramenta ViSC

Para a proposta final ViSC 2, estamos ainda mapeando construções improdutivas a fim de incluir recomendações que indiquem problemas na construção gráfica com objetivo de evitar construções ineficientes. Precisamos também de implementações complementares para permitir que o usuário tenha o nível de interação esperado tais como zoom, pan, alteração de configurações visuais e inclusão de botões de exportação dos resultados. Além disso, desejamos melhorar como as perguntas estão sendo exibidas na interface, de forma que se torne mais claro para o usuário a relação com as visualizações.

8. COMPARAÇÃO COM TRABALHOS RELACIONADOS

Na busca por automatização de apresentação de informação e sistemas de recomendação em sistemas de visualizações, encontramos *Show me* [9] e *Explorations Views* (EV) [1]. *Show me* é um conjunto integrado de comandos de interface que incorpora apresentações automáticas no Tableau, um sistema comercial projetado para ser usado tanto por usuários novatos quanto experientes e para construir visualizações eficientes baseado na semiologia de gráficos de Bertin [16]. *Show me* é um diálogo de escolhas com *tooltips* que descrevem condições para uma escolha estar ativa. EV também sugere gráficos e *templates* para a criação de *dashboards* por usuários novatos. Essa proposta partiu de pesquisas sobre a necessidade de usuários novatos em análises visuais. Assim, determinaram que a sequência de passos para a criação de *dashboards* deveria ser simples, com um contínuo *feedback* visual e além disso prover sugestões de gráficos e *templates* baseado nas tarefas do usuário.

O sistema proposto também parte da definição de valores *defaults* para algumas dimensões e gera automaticamente uma visualização elementar. A inclusão de um sistema de recomendação visa sugerir não apenas outras visualizações mas também estabelecer uma conversa com o usuário e compreender a tarefa a ser executada através da pergunta que o usuário deseja responder. Com isso, nossa proposta pretende estimular o interesse do usuário inexperiente pela busca de outras formas de visualização com custo baixo de experimentação. Assim esse sistema de recomendação pretende guiar o usuários para visualizações possíveis baseado não apenas nas características do dado como também na tarefa do usuário. Utiliza os métodos como *feedback* visual e definição de valores *defaults* para automatizar parte do processo de construção gráfica.

9. CONCLUSÕES

Este trabalho consiste no desenvolvimento da ferramenta Web para auxiliar na construção de gráficos. A fim de alcançar esse objetivo, projetaremos uma interface que visa oferecer apoio através de um sistema de recomendações com base em perguntas adequadas às visualizações e sinalização de problemas de construção. As avaliações com usuários serão realizadas após a implementação das melhorias para verificar se o objetivo foi alcançado.

10. REFERÊNCIAS

[1] Elias, Micheline; Bezerianos, Anastasia. 2011. Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices. Interact 2011, Part IV, pp. 274-291.

[2] Sousa, Taissa. 2011. Caracterização semântica de mecanismos de visualização. Monografia de fim de curso da disciplina Top. Hipertexto/Multimídia II

[3] Sousa, Taissa. 2011. Como sistemas de significação e comunicação influenciam na interpretação de dados estatísticos por usuários interessados em responder determinadas perguntas. Monografia de fim de curso da disciplina Engenharia Semiótica.

[4] Heer, Jeffrey; Shneiderman, Ben. 2012. Interactive Dynamics for Visual Analysis. A taxonomy of tools that support the fluent and flexible use of visualizations. ACMqueue.

[5] Bostock, Michael; Heer, Jeffrey. 2009. Protovis: A Graphical Toolkit for Visualization. InfoVis 2009.

[6] Tory, Melanie; Möller, Torsten. Rethinking Visualization: A High-Level Taxonomy

[7] Duke, D.J.; Brodli K.W.; Duce, David.A. Building an Ontology of Visualization

[8] National e-Science Center. Visualization Ontologies. Último acesso em 13/11/2011. Disponível em http://www.nesc.ac.uk/talks/393/vis_ontology_report.pdf

[9] Mackinlay, Jock D.; Hanrahan, Pat; Stolte, Chris. Show Me: Automatic Presentation for Visual Analysis. IEEE Transactions on Visualizations and Computer Graphics, vol. 13, número. 6, november/ december 2007.

[10] Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future testing (pp. 73-126). Hillsdale, NJ: Erlbaum.

[11] Goldenberg, E. Paul. (1988) 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), 135-173.

[12] Tufte, Edward R. The Visual Display of Quantitative Information. 2001

[13] Clement, J. "Misconceptions in Graphing". Proceeding 9nd Annual Meeting of the International Group for the Psychology of Mathematics Education, 1985. p. 369-375.

[14] Gomes Ferreira, V. G. Exploring Mathematical Functions Through Dynamic Microworlds. 1997, 353 f. Tese (Doutorado em Educação). Institute Education, Universidade de Londres. Londres, 1997.

[15] Carzola, Irene. A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos. Campinas, 2002.

[16] Bertin, Jacques. Semiology of Graphics: Diagrams, Networks, Maps. 1918. esri first edition 2011, usa

[17] Steele Julie; Lliinsky, Noah. Beautiful Visualization. Looking at Data through the Eye of Experts. O'Reilly, 2010.

[18] Barbosa, S.D.J. & Silva, B.S. (2010) Interação Humano-Computador. Campus/Elsevier.

[19] Ericsson, K. Anders; Simon, Herbert A. Protocol Analysis. MIT, 1993

[20] Wilson, C. (1998). Usability techniques: Pros and cons of co-participation in usability studies. Usability interface, 4(4).

[21] Duke, D.J.; Brodli K.W.; Duce, David.A; Herman, Ivan. 2005. Visualizations Viewpoints - Do you see what I mean? IEEE Computer Society 2005.

[22] Santaella, Lucia. O método anticartesiano de C. S. Peirce. Editora UNESP, 2004.

[23] Amar, Robert, Eagan, James, Stasko, John. 2005. Low-Level Components of Analytic Activity in Information Visualization. IEEE Symposium on Information Visualization 2005. USA

[24] Del Rio, N; da Silva, Paulo Pinheiro. Visualization Queries. Último acesso em 12/11/2011. Disponível em <http://trust.utep.edu/visko/docs/docs/visquery.pdf>

[25] Ricci, Francesco; Rocarch, Lior; Shapira, Bracha; Kantor, Paul B. Recommender System handbook. Springer 2011.