

Uma Investigação Sobre as Técnicas de Visualização de Informação Quantitativas e Qualitativas

Diego Barros, Raquel Prates, Raquel Melo-Minardi

Departamento de Ciência da Computação (DCC)

Universidade Federal de Minas Gerais (UFMG)

Belo Horizonte, Brasil

{diego.barros, rprates, raquelcm}@dcc.ufmg.br

RESUMO

Nos últimos anos tem havido um grande crescimento na área de visualização de informação e diferentes técnicas e ferramentas estão sendo criadas. Muitas das diferentes técnicas de visualização de informação são construídas a partir de dados quantitativos e qualitativos. Contudo, muitas vezes técnicas de visualização de informação classificadas como qualitativas estão sendo usadas em uma abordagem quantitativa. Por exemplo, uma nuvem de palavras pode ser analisada qualitativamente mas também quantitativamente através da frequência dos termos. Em meio a este cenário, temos muitos arcabouços e técnicas que visam classificar diferentes visualizações de informação a fim de apoiar os usuários durante a escolha da visualização que melhor adapta às suas necessidades. Este artigo apresenta um estudo em andamento que envolve as áreas de Interação Humano-computador e Visualização de Informação, cujo objetivo é investigar as técnicas de visualização de informação qualitativa e quantitativa a fim de avaliar a viabilidade e importância em se fazer tal classificação, através de um estudo comparativo envolvendo diferentes técnicas.

Palavras-Chave

Visualização de Informação; Pesquisa Qualitativa; Pesquisa Quantitativa; Arcabouços

ABSTRACT

In the last years a big growth in Information Visualization field has happened and different techniques and tools are being created. Many different information visualization techniques are built from qualitative and quantitative data. However, sometimes information visualization techniques classified as qualitative are being used in a quantitative approach, for example, a word cloud can be analyzed qualitatively but also quantitatively through the frequency of the terms.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. IHC'14, Brazilian Symposium on Human Factors in Computing Systems, October 27-31, 2014, Foz do Iguaçu, PR, Brazil. Copyright 2014 SBC. ISSN 2316-5138 (pendrive). ISBN 978-85-7669-291-1 (online).

Amid this scenery we have several frameworks and techniques which aims to classify different information visualization to support users when choosing visualizations that best fits their needs. This paper presents an ongoing study involving Human-computer interaction and information visualization field whose goal is to investigate quantitative and qualitative information visualization techniques to assess the viability and importance in doing such classification, through a comparative study involving different techniques.

Author Keywords

Information Visualization; Qualitative Research; Quantitative Research; Frameworks

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Design, Human Factors

INTRODUÇÃO

Nos últimos anos tem havido um grande crescimento na área de visualização de informação. Henderson e Segal [8] afirmam que o desenvolvimento nesta área tem alcançado níveis exponenciais. Juntamente com este crescimento, surgem várias técnicas e sistemas de visualização que buscam apoiar os usuários durante a escolha, análise e criação de visualizações. Por exemplo, Abdalla e Barbosa [1] em seu trabalho apresentam e avaliam o ViSC, um sistema de recomendação de visualizações, cujo objetivo é oferecer suporte na construção de gráficos interativos para a visualização de dados estatísticos, através de um conjunto de recomendações baseadas nos dados selecionados e na interação dos usuários com a ferramenta.

Visualização de dados é um processo baseado em dados quantitativos e qualitativos que resultam em uma imagem com a representação de dados brutos que são lidos por visualizadores e dão suporte à exploração, verificação e comunicação dos dados [2]. As técnicas de visualização estão fundamentadas nos tipos de dados que elas expressam, são eles: quantitativos e qualitativos.

Dados quantitativos são definidos como valores baseados em dados numéricos que representam algum tipo de informação,

tais como: itens de escala likert, coordenadas geográficas, codificações quantificadas e etc. Os dados qualitativos por sua vez possuem o escopo mais amplo e estão na forma de entrevistas transcritas, gravações de conversas, imagens, vídeos, desenhos que capturam eventos, processos e resultados. Cada um destes tipos de dados requer abordagens e métodos de visualização únicos para comunicar de forma eficaz o conhecimento que eles contém [2].

Em meio a este cenário, muitos arcabouços e técnicas têm sido propostos a fim de classificarem os diversos tipos de visualizações com o objetivo de além de tornar o seu desenvolvimento mais fácil e efetivo, apoiarem os usuários durante a escolha das visualizações que melhor atendam às suas necessidades. Porém, apesar de haver classificações voltadas para os tipos de dados, conforme afirmam Henderson e Segal [8] em última análise, metodologicamente a maioria das técnicas de visualização transformam os dados qualitativos em segmentos quantificáveis.

Sendo assim, este trabalho propõe uma investigação acerca das diferenças entre os diversos tipos de arcabouços e técnicas voltados para visualizações qualitativas e quantitativas. Através de um estudo comparativo entre diferentes arcabouços, a fim de se fazer uma apreciação sobre a viabilidade e importância em se fazer tal tipo de classificação. Acredita-se que obter tal entendimento sobre as visualizações é algo importante tanto para projetistas de sistemas durante a implementação e proposta de novas técnicas e sistemas de visualizações, quanto para usuários uma vez que possibilitará que eles tenham um melhor entendimento sobre o uso das visualizações e *frameworks* existentes.

Na próxima seção será apresentado o referencial teórico que fundamenta este trabalho, contendo uma visão geral sobre visualização de informação e os tipos de visualizações existentes, além de alguns arcabouços cujos objetivos são classificar as diferentes técnicas de visualizações. Finalmente é descrita a metodologia usada neste estudo, a análise dos resultados obtidos e a conclusão do trabalho.

REFERENCIAL TEÓRICO

Abordaremos 3 diferentes arcabouços teóricos cujos objetivos são classificar as técnicas de visualizações de informações que serão utilizados na realização deste estudo. O processo de anotação de Franciscani Júnior et al [5] cujo foco é quantitativo, o arcabouço de Bhowmick [3] voltado estudos puramente qualitativos e o de Henderson e Segal [8] desenvolvido com foco nas visualizações textuais.

Classificação de Visualizações

Acredita-se que a área de visualização de dados tenha alcançado um nível onde tem se tornado crítico para a comunidade de avaliadores, aprimorarem como a mesma projeta e comunica as informações [2].

Sendo assim, observa-se o surgimento de diversos arcabouços que buscam classificar as visualizações, a fim de oferecer suporte aos avaliadores durante a escolha de qual técnica de visualização melhor se adapta para determinados tipos de dados, metodologias, contextos e problemas.

Neste sentido, Franciscani Júnior et al [5] afirmam que a área de visualização de informação requer um processo de anotação unificado que permita que sua comunidade anote ou associe os termos das tradicionais e as novas técnicas de visualizações que estão sendo desenvolvidas. Sendo assim, é proposto um processo de anotação para as técnicas de visualização, baseados em uma coleção de termos e conceitos que cobrem os componentes visuais das técnicas de visualização, por exemplo, objetos usados na composição visual, tipos de *displays* e atributos pré-atentivos; e capacidades que são os recursos mais amplos que englobam os relacionamentos quantitativos descritos, os padrões visuais revelados, assim como, as técnicas de navegação e interação analíticas que possam ser usadas.

Dentro do campo da pesquisa qualitativa, Bhowmick [3] apresenta um arcabouço para pesquisadores qualitativos. Segundo o autor, tal *framework* dá um passo em direção à construção de uma ferramenta exploratória que possibilitará pesquisadores qualitativos obterem ideias e conectarem estruturas conceituais com representações concretas, durante a solução de complexos problemas de pesquisa qualitativas de várias tradições metodológicas. Baseado nesta metodologia o arcabouço realiza a seguinte classificação:

1. *Tradição Metodológica*: Etnografia, Biografia, Fenomenologia, Grounded Theory, Estudo de Caso;
2. *Funções da Visualização*: Exploração, Análise, Síntese, Apresentação;
3. *Tipos de Dados*: Espacial, Temporal, Espaço-Temporal, Atributos;
4. *Análise de Dados*: Análise de documentos, análise de conteúdo, análise de redes, etc ...
5. *Métodos de Coleta de Dados*: Entrevistas, observações, métodos participatórios, etc ...
6. *Teorias Sociais*: Teoria discursiva, teoria de redes sociais, etc ...

A fim de explorar as várias visualizações de dados para texto, Henderson e Segal [8] propõe um arcabouço que visa organizar e trazer o entendimento de diferentes segmentos de dados textuais, são eles: palavras, frases, temas e narrativas. Referenciados pelos autores como “Níveis de *Displays*”, tais segmentos possibilitam a contagem, catalogação e conexão com outros segmentos. Além do nível de *display*, as visualizações são classificadas de simples à complexas conforme o nível de conhecimento ou habilidade necessária para o desenvolvimento da visualização.

METODOLOGIA

Para a realização deste estudo foram recuperados todos os artigos que tratam sobre a área de Visualização de Informação das últimas 5 edições dos seguintes eventos: Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC) e da Conference on Human Factors in Computing Systems (CHI). O critério para a escolha dos artigos foi a leitura por título, abstract e leitura diagonal do conteúdo. A fim de se obter: objetivos, motivação, metodologia e abordagem

de pesquisa (qualitativa, quantitativa ou mista) além do nome das técnicas de visualização que foram abordadas em cada um dos artigos.

Posteriormente as visualizações identificadas nos artigos que fazem o seu uso na pesquisa, foram reunidas e classificadas conforme os arcabouços teóricos dos autores Franciscani Júnior et al [5], Bhowmick [3] e, Henderson e Segal [8]. Com o objetivo de realizar um comparativo entre a classificação feita por cada arcabouço e a abordagem empregada no contexto de pesquisa dos artigos recuperados.

ANÁLISE E CONSOLIDAÇÃO

Ao todo foram recuperados um total de 57 artigos, sendo 8 referentes ao IHC, 49 a CHI e 68 diferentes técnicas e sistemas de visualizações. Destas, 38 aparecem em pesquisas cuja abordagem é mista e 15 foram utilizadas respectivamente em pesquisas cuja abordagem são qualitativa e quantitativa. A partir da leitura dos trabalhos foram coletados os objetivos, motivação, metodologia, abordagem de pesquisa (qualitativa, quantitativa ou mista) e o nome de sistemas e visualizações de informação que foram abordados em cada um dos artigos. Após a leitura e análise destes dados, foram reunidos os trabalhos que fizeram uso de técnicas de visualizações, ou seja, aqueles em que os autores utilizaram as técnicas de visualização para apoiar o processo de análise em seu estudo. Destes, foram identificados apenas 4 artigos com as seguintes abordagens de pesquisa apresentados na tabela 1.

Autor	Abordagem	Visualização
Gasparini [6]	Mista	Gráfico de Barras Gráfico de Dispersão Nuvem de Termos Mapas Gráfico de Setores
Gilbert [7]	Quantitativa	Word Tree Density Plot Mapa de Calor
Buchdid [4]	Mista	Nuvem de Termos Gráfico de Barras
Petrillo [9]	Quantitativa	Gráfico de Barras Empilhado Gráfico de Barras

Tabela 1. Artigos recuperados com suas respectivas abordagens de pesquisa e técnicas de visualizações utilizadas

Não foram identificados trabalhos cuja abordagem seja puramente qualitativa. Todos os trabalhos ou são quantitativos ou fazem uso de uma abordagem mista, ou seja, apresentam uma análise quantitativa juntamente com uma qualitativa. Com base nestes dados as visualizações contidas em cada um dos trabalhos foram classificadas utilizando as técnicas dos autores: Franciscani Júnior et al [5], Bhowmick [3] e, Henderson e Segal [8]. Onde o processo de anotação de Franciscani Júnior et al [5] tem o seu foco para uma abordagem quantitativa e os arcabouços dos autores Bhowmick [3] e, Henderson e Segal [8] para uma abordagem qualitativa, sendo o último voltado para visualizações textuais.

Gasparini et al [6] através de uma abordagem mista fazem uso das seguintes técnicas de visualizações: Gráfico de Barras,

Gráfico de Dispersão, Nuvem de Termos, Mapas e Gráfico de Setores. Com o objetivo de apresentar uma exploração visual do campo de IHC no Brasil, através de uma análise dos últimos 15 anos de produções de trabalhos no Simpósio Brasileiro de Fatores Humanos e Computacionais. As técnicas de visualização Gráfico de Barras, Gráfico de Dispersão e Gráfico de Setores, foram totalmente mapeadas pelo processo de anotação de Franciscani Júnior et al [5]. Porém, nenhum dos outros dois arcabouços conseguiram classificar as visualizações. Algo esperado devido ao fato de tais visualizações representarem dados quantitativos e estarem sendo aplicadas em arcabouços voltados para visualizações qualitativas.

Além destas visualizações foram identificadas duas técnicas de visualizações dadas como qualitativas, são elas: Nuvem de Termos e Mapas. A Nuvem de Termos foi classificada perfeitamente pelos arcabouços de Bhowmick [3] e, Henderson e Segal [8] que a classificou como nível de complexidade sendo simples e nível de *display* como palavra. A respeito do mapa (Figura 1) o trabalho de Franciscani Júnior et al [5] é o que melhor consegue classificar tal visualização devido os valores quantitativos que ele apresenta.

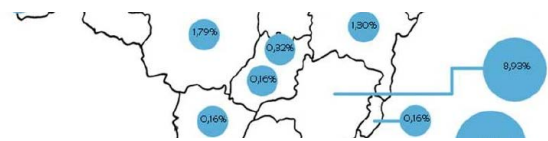


Figura 1. Representatividade de cada estado Brasileiro nos IHCs - Adaptado. Gasparini et al (2013)

Gilbert e Bakhshi [7] em seu estudo quantitativo da Rede Social Pinterest, apresentam um mapa de calor em formato de mapa-múndi representando a quantidade de usuários da Rede Social por sexo ao redor do mundo. Tanto o trabalho de Franciscani Júnior et al [5] quanto o de Bhowmick [3] conseguem classificar esta visualização. O primeiro devido os valores quantitativos referentes à frequência de usuários por sexo ao redor do mundo. O segundo no entanto consegue classificar em alguns pontos a visualização devido a função da visualização ser de apresentação, a análise como análise de redes e o tipo de dados como atributos.

Além dessa visualização foi utilizada a técnica de visualização *Density Plot*. Devido ao fato de lidar com dados quantitativos somente o trabalho de Franciscani Júnior et al [5] pôde mapear essa visualização. É apresentado também um word tree, técnica de visualização textual que é completamente classificada pelos *frameworks* de Bhowmick [3] e principalmente Henderson e Segal [8] cujo foco são visualizações textuais, com média complexidade e nível de *display* como frase. Apesar disso, a mesma foi utilizada em uma análise puramente quantitativa onde foi levado em consideração apenas a frequência de aparição das palavras.

Buchdid e Baranauskas [4] fazem uma análise do programa completo das 3 maiores conferências de IHC, com o objetivo de identificar o foco principal das contribuições. Através de uma abordagem mista, foram utilizadas as técnicas de Nuvem de Termos e Gráfico de Barras em sua análise. Acerca

do Gráfico de Barras, o processo de anotação de Franciscani Júnior et al [5] foi a única das 3 técnicas que pôde classificar essa visualização, algo esperado por se tratar de uma visualização que representa dados quantitativos.

Já a Nuvem de Termos por se tratar de uma visualização textual, os arcabouços de Bhowmick [3] e Henderson e Segal [8] conseguem classificá-la. Porém, no trabalho de Buchdid e Baranauskas [4] ela é utilizada tanto qualitativamente através da análise do conteúdo dos termos, ou seja, do significado transmitido pelas palavras, quanto quantitativamente através da análise das frequências de aparição dos termos no texto.

Finalmente, Petrillo et al [9] em uma abordagem quantitativa apresentam uma técnica de visualização interativa que auxilia na análise dos dados obtidos em uma escala Likert de muitas fontes alternativas. Em seu trabalho são utilizados no processo de análise o Gráfico de Barras e o Gráfico de Barras Empilhado, duas técnicas de visualização voltadas para dados quantitativos, com isso, somente a técnica de Franciscani Júnior et al [5] consegue mapear tais visualizações.

CONCLUSÃO

Este trabalho apresentou os resultados parciais de um estudo em andamento que envolve visualização de informação, métodos qualitativos e interação humano-computador. Onde foi feita a análise de diferentes arcabouços que buscam classificar as diferentes técnicas de visualização com o objetivo de apoiar os usuários durante a escolha das visualizações que melhor atendam às suas necessidades.

A partir da análise dos resultados foi possível observar que todos os arcabouços cumprem o seu papel em classificar as visualizações com suas respectivas abordagens de pesquisas qualitativas e quantitativas. Porém, nos trabalhos analisados, em alguns casos visualizações dadas como qualitativas estão sendo utilizadas quantitativamente, por exemplo, como foi visto através da análise quantitativa utilizando a técnica *Word Tree* no trabalho de Gilbert e Bakhshi [7] e a Nuvem de Termos que foi utilizada em uma abordagem mista no trabalho de Buchdid e Baranauskas [4] que avaliou não somente o conteúdo dos termos que a visualização expressa mas também a frequência de aparecimento das palavras.

Das técnicas avaliadas observa-se que o arcabouço de Bhowmick [3] classifica as visualizações qualitativas de maneira mais geral e está voltada para pesquisas cuja abordagem é puramente qualitativa. Já o arcabouço de Henderson e Segal [8], apesar de tratar de visualizações qualitativas, possui o seu foco voltado para visualizações de dados textuais o que o torna um pouco restrito. O processo de anotação de Franciscani Júnior et al [5] demonstra ser mais amplo em classificar as visualizações. Porém, o seu foco é para visualizações de dados quantitativas o que faz que ele ofereça pouco suporte para visualizações de dados qualitativos, dando assim abertura para trabalhos futuros acerca de possíveis adaptações ou extensões do processo a fim de abranger também visualizações qualitativas.

Apesar de haver casos dentro do contexto dos trabalhos analisados onde visualizações de informação qualitativas estão sendo utilizadas quantitativamente, não foram encontrados

casos onde visualizações quantitativas foram utilizadas em estudos qualitativos. Observa-se também que os arcabouços tem um importante papel de guia ao fornecer diretrizes para pesquisadores e usuários, sobre quais visualizações melhor se adequam dado o seu contexto e problema.

AGRADECIMENTOS

Agradecemos aos órgãos de fomento à pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Financiadora de Estudos e Projetos (FINEP), INCT-Web (INWeb) (MCT/CNPq nº 57.3871/2008-6) e também aos colegas do Núcleo de Pesquisa em Engenharia Semiótica e Interação (PENSí) e DCC/UFMG pelo apoio.

REFERÊNCIAS

1. Abdalla, T. F. D. S., and Barbosa, D. J. S. Sistema de recomendação para apoiar a construção de gráficos com dados estatísticos. In *IHC - Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, vol. 5138 (Manaus, 2013), 168–177.
2. Azzam, T., Evergreen, S., Germuth, A. A., and Kistler, S. J. Data visualization and evaluation. *New Directions for Evaluation 2013*, 139 (2013), 7–32.
3. Bhowmick, T. Building an exploratory visual analysis tool for qualitative researchers. In *AutoCarto International Symposium on Automated Cartography* (Vancouver, 2006), 1–13.
4. Buchdid, S. B., and Baranauskas, M. C. C. IHC em contexto: o que as palavras relevam sobre ela. In *IHC - Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, vol. 5138 (Manaus, 2012), 199–208.
5. Franciscani Jr., G., Santos, R. L. T., Ottoni, R., Pesce, J. P., Meira Jr., W., and Melo-Minardi, R. An annotation process for data visualization techniques. In *Proceedings of the 3rd International Conference on Data Analytics, IARIA* (Rome, Italy, 2014).
6. Gasparini, I., Kimura, M. H., Pimenta, M. S., and Alegre, P. Visualizando 15 anos de IHC. In *IHC - Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais* (Manaus, 2013), 238–247.
7. Gilbert, E., and Bakhshi, S. I need to try this: A statistical overview of pinterest. In *SIGCHI - Conference on Human Factors in Computing Systems* (Paris, 2013), 2427–2436.
8. Henderson, S., and Segal, E. H. Visualizing qualitative data in evaluation research. *New Directions for Evaluation 2013*, 139 (2013), 53–71.
9. Petrillo, F., André Suslik Spritzer, Freitas, C. D. S., and Pimenta, M. Interactive Analysis of Likert Scale Data using a Multichart Visualization Tool. In *IHC - Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais* (Porto de Galinhas, 2011), 358–365.