



Visualização de dados: representações gráficas e padrões de dados

Data visualization: graphic representation and data patterns

Amanda Pereira, Carolina Riff, Barbara Castro, Doris Kosminsky

ferramentas, visualização, dados, padrão

A visualização de dados é uma área interdisciplinar em expansão que vem nas últimas décadas e está diretamente relacionada diretamente ao aumento na produção de dados e aos avanços das tecnologias digitais. Nesse contexto, tem-se observado um aumento no número de ferramentas automáticas de produção de visualizações. Este artigo tem como objetivo expor a primeira etapa de uma investigação. Este artigo tem como objetivo expor a primeira etapa de uma investigação sobre estas ferramentas, que consiste na identificação dos tipos de visualização existentes e os conjuntos de dados aplicáveis a cada uma delas, visando uma futura aplicação deste conhecimento na análise das ferramentas automáticas a serem estudadas na continuidade do projeto.

tools, visualization, data, pattern

Data visualization is an interdisciplinary field that has been expanding in recent decades and is related directly with the increase in the production of data and the advances of technology itself. Thus, it has been seen an increase in the number of automatic visualization tools. This article aims to present the first stage of an investigation into automatic data visualization tools that consists in identify the known types of visualization and the data applicable in each of them. This knowledge will be used in the future analysis of automatic data visualization tools.

1 Introdução

A visualização de dados é uma área interdisciplinar em expansão desde o início do século XXI com o avanço no processo de informatização trazido pelas mídias digitais e que se propõe a contribuir com a organização do pensamento frente à avalanche de dados da contemporaneidade. Trata-se de situações em que dados quantificáveis, mas não necessariamente numéricos ou visualizáveis, configuram uma representação visual (MANOVICH, 2004), visando amplificar a cognição (CARD, 2004).

Como qualquer nova disciplina, a visualização de dados tem passado por diversas fases (CARD, 2004) não necessariamente sucessivas. A primeira foi uma fase exploratória, onde investigaram-se os limites e as capacidades da disciplina. O segundo momento foi de caracterização, onde criaram-se taxonomias, desenvolveram-se e empregaram-se métodos de organização próprios e específicos, como a linguagem de programação D3, voltada para a criação de representações gráficas para WEB. Alguns dos métodos empregados apontam para o surgimento ou variação dos tipos de representação, assim como de novas formas de organização dos padrões de dados observados. Neste contexto, também chama a atenção o advento de ferramentas automáticas onde é possível fazer o upload de uma planilha de dados e

escolher o formato desejado dentre um cardápio de representações sugeridas, podendo assim dispensar a programação, total ou parcialmente.

Esta pesquisa consiste, inicialmente, em identificar os formatos de representação e sua relação com os padrões de dados. Realizaremos uma pesquisa exploratória das ferramentas existentes, aplicando as informações aprendidas de forma a desenvolver uma classificação das ferramentas quanto às suas características. Na continuidade, identificaremos um grupo de ferramentas a serem analisadas em maior profundidade através da realização de testes com conjuntos de dados, com o objetivo de investigar a sua aplicabilidade por leigos e profissionais.

Este trabalho consiste na primeira etapa desta pesquisa que tem por objetivo responder as seguintes perguntas: Quais são os tipos de visualização de dados observados com maior frequência e que conjunto de dados se aplicam à essas representações? As visualizações de dados podem ser agrupadas de acordo com o padrão de organização dos dados utilizados?

Nesse artigo, apresentaremos os diversos tipos de visualização utilizados, relacionando-os aos conjuntos de dados possíveis de serem empregados em cada uma delas. Acreditamos que, de posse do conhecimento obtido com esta investigação inicial, poderemos, no futuro, investigar as características e potencialidades das ferramentas automáticas de visualização de dados de forma a identificar a sua utilidade e também a localizar o papel do designer nesta nova etapa de desenvolvimento da área.

2 Método

O avanço no volume de dados e a sua digitalização gerou o processo de mapeamento, onde um mesmo código numérico pode ser representado por diversas formas (MANOVICH, 2004, 150). A plasticidade dos dados digitalizados propiciou a criação de novos formatos de visualização, em contraponto aos formatos tradicionais de representação de informação surgidos no final do século XIX (MANOVICH, 2010).

Iniciamos este trabalho com uma pesquisa bibliográfica, consultando o artigo *A Tour Through the Visualization Zoo* e a dissertação *The Language of Graphics* que analisa os tipos de representação de informação de acordo com diversos autores tais como Bertin, Tufte e Bounford, dentre outros (apud ENGELHARD). Em seguida, comparamos as informações obtidas com a encontrada no site *Data Visualization Catalogue*, que relaciona a organização dos conjuntos de dados com tipos de visualização.

3 Representações gráficas

A seguir, apresentaremos exemplos de nove grupos de representações gráficas (ENGELHARD 2002:137) e os tipos de dados empregados por elas.

Mapas

São os tipos de visualização em que a estrutura sintática baseia-se em um espaço físico ou em áreas geográficas, podendo ser uma representação literal ou figurada do espaço em questão. Um exemplo de representações deste tipo seriam as cartografias, como a figura 1.

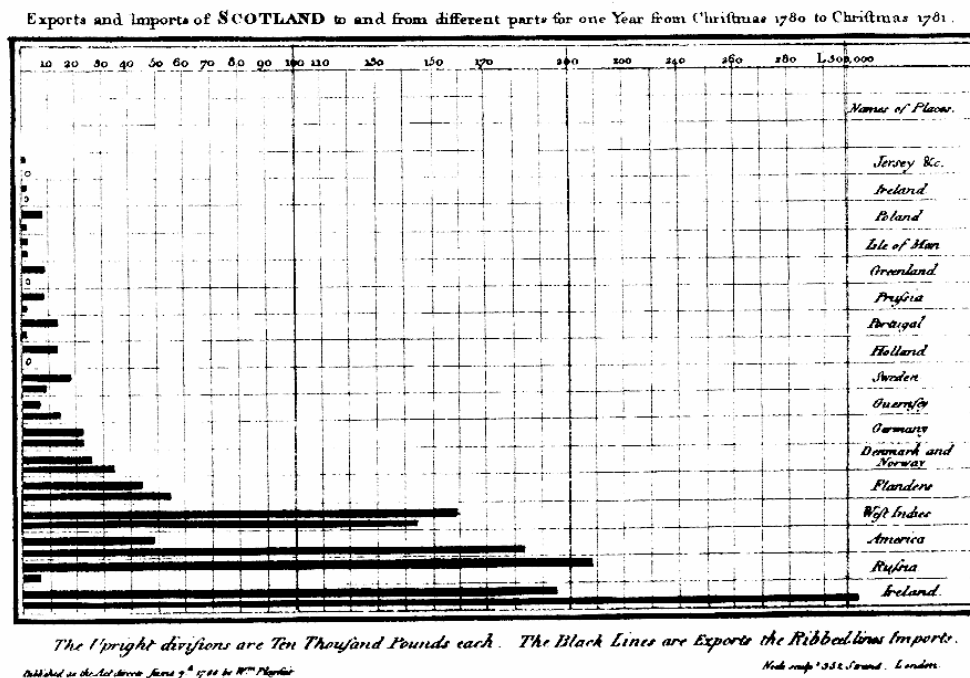
Figura 1: Cartografia representando a superfície do planeta Terra. Disponível em: http://upload.wikimedia.org/wikipedia/commons/c/cf/Worldmap_LandAndPolitical.jpg - acesso em: 4 de Abr. 2015



Gráficos estatísticos

A estrutura sintática busca a representação de quantidades. Na figura 2 temos uma visualização do tipo gráfico de barras onde as barras horizontais são utilizadas para representar o número de importações e exportações da Escócia entre os natais de 1780 e 1781. A largura da barra representa as quantidades, quanto mais largas forem as barras, maior serão os números de importação e exportação para o país.

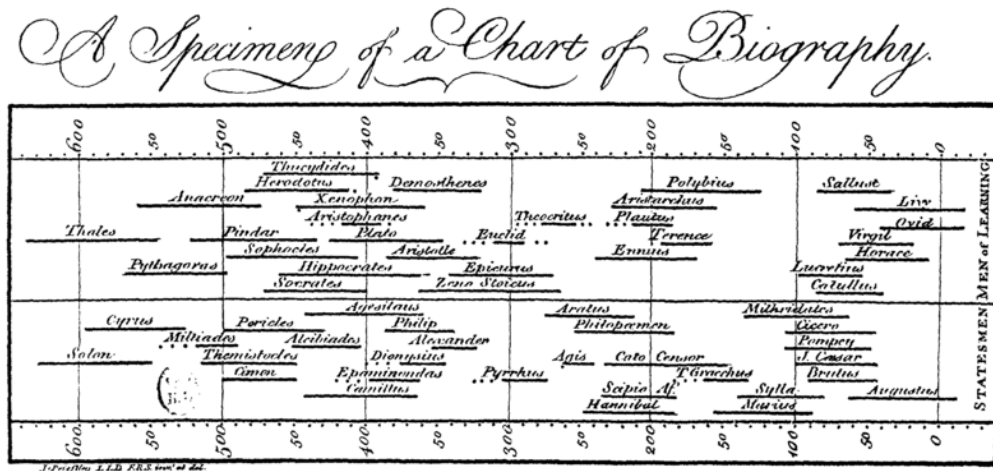
Figura 2: Gráfico de barras produzido por William Playfair em 1786. Disponível em: http://en.wikipedia.org/wiki/File:Bar_chart#/media/File:Playfair_Barchart.gif - acesso em: 17 de Mar. 2015.



Gráficos temporais

A estruturação sintática tem como objetivo evidenciar a passagem de tempo (dados temporais) geralmente de forma metafórica: a passagem de tempo é representada por uma distância física. A forma mais comum deste tipo de visualização são as linhas do tempo. O diagrama de Priestley apresenta uma comparação do tempo de vida de figuras importantes e as época em que viveram.

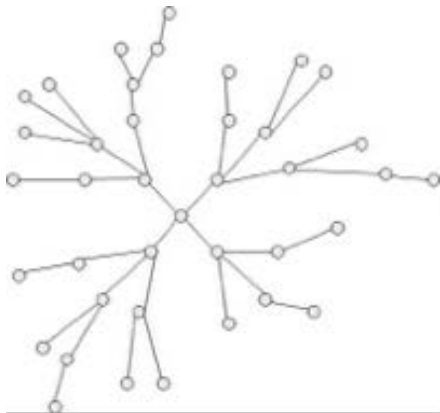
Figura 3: Considerada a primeira linha do tempo. Disponível em: <http://en.wikipedia.org/wiki/File: PriestleyChart.gif> Acesso em: 4 de Abr. 2015.



Diagramas de rede

Este tipo de representação gráfica tem como objetivo evidenciar as relações existentes entre dados através de ligações. Podem se dar através de um formato linear, circular, de árvore ou rede.

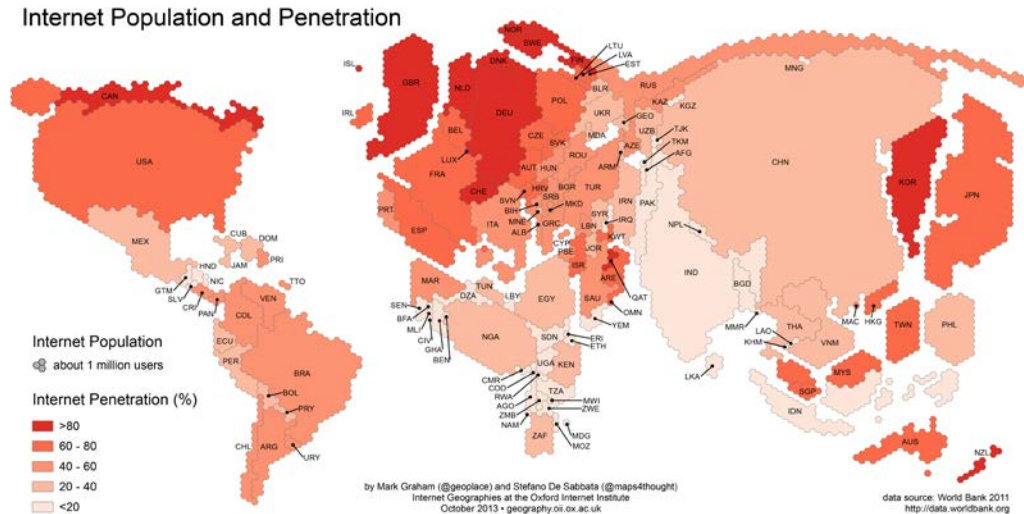
Figura 4: Exemplo de diagrama de rede <http://commons.wikimedia.org/wiki/File:Network_Tree_diagram.png> Acesso em: 17 de Mar. 2015.



Mapas estatísticos

Representação que cruza dados numéricos e geográficos em uma única visualização. Na figura 5 foram utilizados pequenos círculos, cada um representando 1 milhão de pessoas que possuem acesso a internet no país em questão. Os círculos são posicionados no mapa de forma a ocuparem localizações semelhantes às suas localizações geográficas. As cores representam a densidade demográfica de pessoas no país que possuem acesso, sendo os mais saturados os com a maior concentração de internautas.

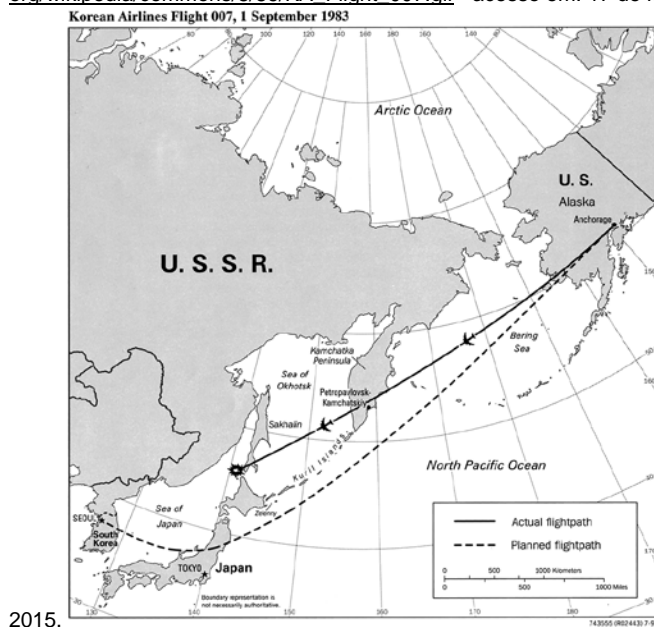
Figura 5: Mapa estatístico representando o acesso a internet no mundo. Disponível em: http://upload.wikimedia.org/wikipedia/commons/9/9f/InternetPopulation2011_DeSabbata_Graham_Oll.png - acesso em 17 de Mar. 2015.



Mapas de rede

Representação que trabalha com dados geográficos que possuem algum tipo de relação entre eles. Na figura 6 as ligações representam o trajeto do voo 007 que em 1983 saiu do Alaska em direção a Coreia do Sul mas acabou se desviando de sua rota e foi derrubado por jatos soviéticos ao Norte do Japão.

Figura 6 : Diagrama do desvio de rota sofrido pelo voo 007. Disponível em: http://upload.wikimedia.org/wikipedia/commons/8/86/KA_Flight_007.gif - acesso em: 17 de Mar.

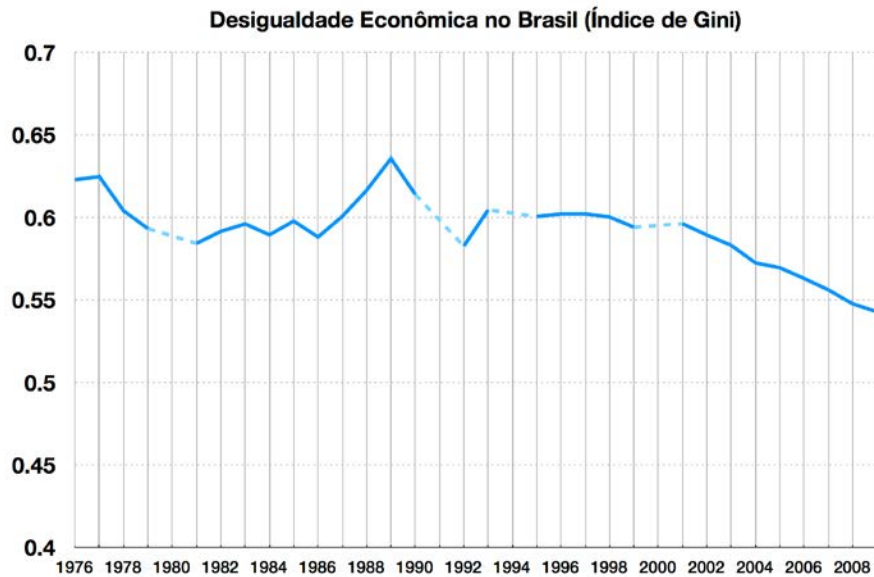


2015.

Gráficos temporais estatísticos

Representação que atende ao mesmo tempo dados temporais e numéricos. Na figura 7 é representado a variação do índice de desigualdade econômica no Brasil ao longo dos anos. A altura da linha simboliza o valor do índice, enquanto o valor horizontal indica o período em que este dado ocorreu.

Figura 7: Gráfico representando a variação no índice de desigualdade econômica no Brasil ao decorrer dos anos. Disponível em: http://upload.wikimedia.org/wikipedia/commons/0/0e/Indice_Gini_Brasil.png - acesso em: 17 de Mar.

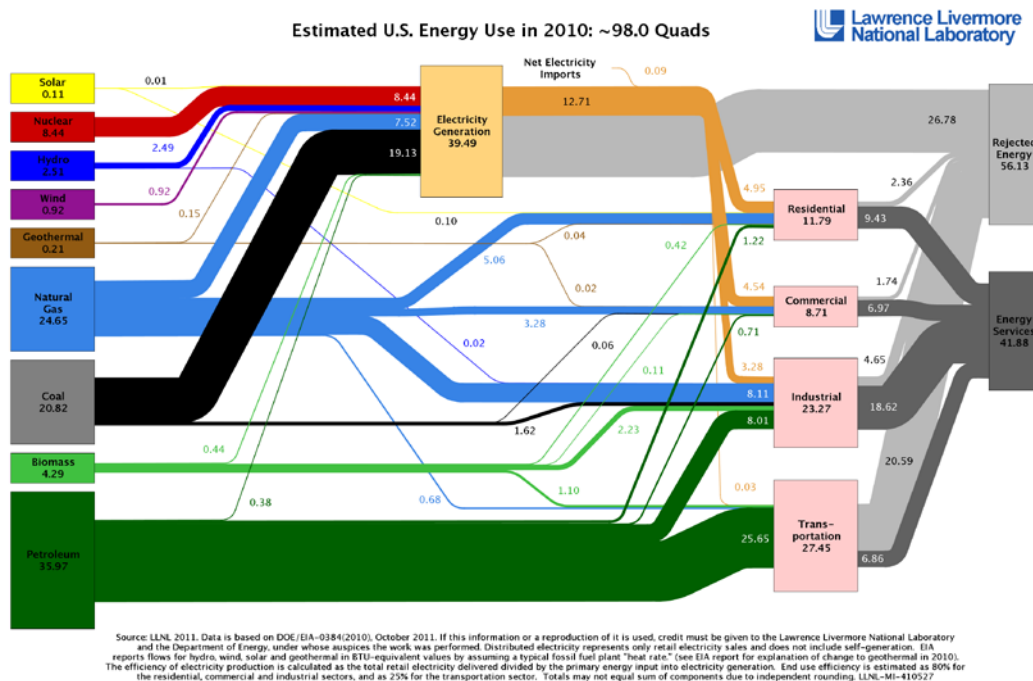


2015.

Diagrama de rede estatístico

Atende a dados de teor numérico que também possuem algum tipo de relação entre si. O gráfico de estilo Sankey na figura 8 representa as transformações de energia ocorridas em 2010 nos Estados Unidos. A largura das linhas representa a quantidade de energia sendo transformada por cada setor em cada tipo de processo. As linhas podem se dividir, demonstrando que a energia foi utilizada para diversos fins e/ou apresentou perdas energéticas.

Figura 8 : Gráfico de transformação de energia nos Estados Unidos em 2010. Disponível em: <http://upload.wikimedia.org/wikipedia/commons/d/dc/LLNLUSEnergy2010.png> - acesso em: 17 de Mar. 2015.



Referências

- CARD, S. 2004. Foreword. In: WARE, Collin. *Information Visualization*. Perception for Design. CA: Elsevier.
- RIBECCA, S. 2015. *Data Visualization Catalogue*. Disponível em: <http://www.datavizcatalogue.com> - acesso 15/3/2015.
- ENGELHARD, J. 2002. *Language of Graphics: A framework for the analysis of syntax and meaning in maps, charts and diagrams*. 197 f. Tese. Faculdade de Ciências, Universidade de Amsterdã, Amsterdã.
- HEER, J.; BOSTOCK, M.; OGIEVETSKY, V. 2010. *A Tour Through the Visualization Zoo*. Communications of the ACM, 53(6).
- MANOVICH, Lev. 2010. *What is visualization?* Disponível em: http://manovich.net/blog/wpcontent/uploads/2010/10/manovich_visualization_2010.doc - acesso 26/1/2011.
- MANOVICH, L. 2004. *Visualização de dados como uma nova abstração e anti-sublime*. In: LEÃO, Lucia. *Derivas: cartografias do ciberespaço*. São Paulo: Annablume, Senac.

Sobre os autores

- Amanda Pereira, Universidade Federal do Rio de Janeiro. Brasil. <aspd.contato@gmail.com>
- Barbara Castro, Universidade Federal do Rio de Janeiro. Brasil.
<barbarap.castro@yahoo.com.br>
- Carolina Riff, Universidade Federal do Rio de Janeiro. Brasil. <carolinariff@gmail.com>
- Doris Kosminsky. Universidade Federal do Rio de Janeiro. Brasil. <doriskos@gmail.com>