



Visualização de dados: passado, presente e futuro

Data visualization: past, present and future

Fabiano Couto Corrêa da Silva*

RESUMO

São expostos os princípios fundamentais da ciência de dados e as generalidades de uma de suas áreas de estudo: a Visualização de dados. O artigo aborda como os dados multivariados tem sido representados por meio de imagens e gráficos ilustrados que relacionam os elementos de sintaxe e semântica que podem contemplar o pensamento analítico nas margens visuais. Analisa como a Visualização de Dados foi desenvolvida ao longo do tempo, utilizando exemplos reconhecidos como de vanguarda neste campo, validando a pesquisa com análise cognitivas básicas em princípios de apresentação de evidências nos displays de informação.

Palavras-chave: Visualização de Dados; Infografias; Dados Científicos; Storytelling, Big Data.

ABSTRACT

The fundamental principles of data science and the generalities of one of its areas of study are exposed: Data Visualization. The article discusses how multivariate data has been represented through illustrated images and graphs that relate the elements of syntax and semantics that can include analytical thinking in visual margins. It analyzes how Data Visualization has been developed over time, using examples recognized as cutting edge in this field, validating research with basic cognitive analysis on principles of evidence presentation in information displays.

Keywords: Data Visualization; Infographics; Scientific Data; Storytelling, Big Data.

INTRODUÇÃO

Existem diversas maneiras pelas quais podemos mostrar que os dados passaram a ser utilizados para a construção de infográficos e gráficos em centros de tecnologia, análise social, de mercado e em praticamente todas as camadas sociais. Antes de iniciar qualquer estudo sobre técnicas de processamento e/ou análise de imagens digitais, devemos perguntar qual é a base da formação dessas imagens. De uma maneira genérica, podemos associar o conceito de imagem com o de um "mapa espacial" ou espaço-tempo que um tipo de sensor produz em uma dada informação. Exemplos desse tipo de "mapas" são nossas percepções de uma determinada situação ou cena de nossos sistemas sensoriais (visão, audição, tato, paladar e olfato). Restringiremos o presente estudo ao campo da percepção visual.

Para focar ainda mais a presente análise, questionaremos o significado real do que significa ver ou adquirir informações através de um sistema sensorial visual. Se nos

* Doutor em Información y documentación en la sociedad del conocimiento (Universitat de Barcelona). Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS). Endereço: Ramiro Barcelos, 2705 - Campus Saúde, CEP 90035-007. Porto Alegre, RS. Telefone: 51 3308-5067.

focarmos no modelo visual humano, associaremos o conceito de ver com o de perceber um sinal de luz com uma intensidade mínima e em uma dada faixa de frequência espectral. No entanto, as possibilidades de obter imagens de sensores que funcionam em condições de iluminação muito diferentes às quais o olho humano é sensível são bem conhecidas, tais como infravermelho, raios-x, etc. Assim, a possibilidade de formar imagens deve estar associada ao tipo de sensor utilizado e às possibilidades de o referido sensor capturar e decodificar a informação que chega. Essa visão nos coloca em uma situação mais geral do que seria o simples estudo de imagens coloridas capturadas por nossos olhos.

A interpretação dos dados, proporcionada pelo avanço da tecnologia da informação, está possibilitando que predomine uma cultura visual no conjunto social de todas as áreas do conhecimento, incidindo profundamente nas formas tradicionais de comunicação científica, em especial na elaboração e representação da informação visual.

A representação gráfica da linguagem humana possibilitou um acúmulo de múltiplas possibilidades frente às limitações do texto ou hipertexto, fazendo surgir um novo conceito, a “infografia”, que rapidamente se tornou independente das suas raízes. Conceitualmente, engloba mais do que a soma da terminologia, que definem a ferramenta -informática- e o produto - *representação gráfica*, que criaram o neologismo *infografia* (MENDONÇA, 2010). Neologismo que, quando utilizado no marco da *infografia* científica, deve ampliar seu significado numa dupla vertente: 1) desde o ponto de vista do processo de elaboração de produtos informativos predominantemente gráficos; 2) desde a definição conceitual destes produtos, de tal maneira que quando se usa o termo infográfico, com significado de informação que inclui componente textual e gráfico, se trata de uma nova maneira de apresentar uma pesquisa, conceito ou notícia.

A diferença essencial entre a Infografia e a Visualização de dados reside principalmente na dificuldade para alterar os dados. Enquanto a Infografia é estática, representando os dados por meio de uma imagem, a Visualização de dados apresenta maior facilidade de manipulação, uma vez que os dados podem ser manipulados mediante a instalação de um software (SATO, 2017).

Esta nova maneira de representar uma informação, embora antiga, como poderemos verificar no presente artigo, tem sofrido diversos tipos de representações. A base desta transformação ocorreu essencialmente por duas circunstâncias: o avanço tecnológico e a utilização de recursos visuais em periódicos. Porém, somente a gestão técnica de uma infografia não basta para produzir imagens eficazes, é necessário integrar o texto para obter resultados lógicos. Enquanto a televisão apresenta imagens que representam as palavras, a infografia deve utilizar as imagens como se fossem palavras e estruturar mensagens que narrem informações com a lógica da pesquisa científica.

Neste contexto a visualização de dados representa o imaginário, reproduz a realidade e suas simulações e trazem apoio para informações textuais. As capacidades comunicacional e técnica se juntam e se reforçam mutuamente, tornando possível a tríade Informação/Comunicação/Técnica. Atua midiaticamente, ou seja, é representada pela palavra e a imagem recorre a teatralização e a uma construção da realidade baseada no conteúdo textual. Representa o imaginário, reproduz a realidade e suas simulações e trazem apoio para informações textuais. Vai além das palavras, buscando apoio semiótico para obter a máxima clareza com a linguagem icônica.

A visualização de dados é vista por muitas disciplinas como um equivalente moderno da comunicação visual. Trata-se da criação e estudo da representação visual dos dados, ou seja, “as informações que foram abstraídas de forma esquemática, incluindo atributos ou variáveis das unidades de informação” (CARD, et. al. 1999). Seu principal objetivo é comunicar informações de forma clara e eficiente aos usuários através de gráficos estatísticos, gráficos de informação, tabelas e gráficos selecionados. A visualização eficaz ajuda os usuários na análise e no raciocínio sobre dados e testes. Torna os dados complexos mais acessíveis, compreensíveis e utilizáveis. Os usuários podem ter certas tarefas analíticas, como fazer comparações ou entender a causa e o princípio do design gráfico (SATO, 2017).

A velocidade com que os dados são gerados aumentou, impulsionada por uma economia crescente baseada em informações. Os dados criados pela atividade da Internet e um número crescente de sensores no ambiente, como satélites e câmeras de tráfego, são conhecidos como "Big Data". Processamento, análise e comunicação desses dados apresentam uma variedade de desafios éticos e analíticos para a visualização de dados (MENDONÇA, 2010).

A visualização gráfica de dados constitui uma disciplina própria dentro do universo da ciência de dados. Essa prática acarretou marcos importantes ao longo da história. Neste artigo buscaremos entender, ainda que brevemente, a importância e impacto das representações visuais, desde a pré-história até o fenômeno *big data*. Desse modo, situamos a visualização de dados e infografia como áreas de estudo. É nesse contexto que analisamos o percurso da Visualização de dados como uma prática da Ciência de dados. Em seu aspecto funcional, se trata de um recurso para garantir maior compreensão de pesquisas e análises empíricas por meio da compilação de um conjunto de soluções infográficas, ao passo que busca identificar aspectos que qualificam uma comunicação visual eficiente.

DADOS, DADOS, DADOS.

Eles estão por toda parte. Em essência, tudo que fazemos ou dizemos pode ser entendido como o ponto de partida de um dado.

Com tantos dados no mundo, como podemos interpretá-los de maneira eficiente e eficaz?

Não somos os primeiros a fazer esta pergunta. Durante milhares de anos as pessoas têm questionado como interpretar toda a informações em seu entorno. Suas contribuições ao longo da história da humanidade possibilitaram a visualização de dados como conhecemos hoje.

O principal objetivo da visualização de dados é comunicar informações ou ideias complexas de forma clara, precisa e eficiente, de uma forma que ajude os usuários a analisar e raciocinar sobre dados e evidências. A ideia de visualizar dados existe há muito tempo, mas, como conceito, tornou-se mais relevante à medida que a sobrecarga de informação aumenta e as necessidades de comunicação evoluem. A atual sociedade da informação tem enormes quantidades de dados. No entanto, não se trata apenas de ter a tecnologia para obtê-los e analisá-los, mas sim de poder dar sentido a esses números e estatísticas para explicar uma história com eles. É isso que os infográficos e mapas interativos são responsáveis, uma modalidade de visualização do chamado *Big Data* cada vez mais utilizada pela mídia, instituições financeiras, iniciativas privadas ou centros de pesquisa científica, entre outros atores que viram o valor da comunicação escondida em conclusões atrás dos números. Sua

capacidade de construir narrativas visuais consumíveis a partir de uma enorme quantidade de dados tornou-a uma ferramenta vital que foi integrada em ambientes e disciplinas diferentes. Isso favorece ainda mais seu desenvolvimento, estimulando novos usos impensáveis até o momento.

A visualização dos dados é tanto uma arte quanto uma ciência (APARÍCIO; COSTA, 2014), como diz o ditado "uma imagem vale mais que mil palavras", ou adaptada aos nossos tempos "uma imagem vale mais que mil linhas de dados".

De acordo com McInerney et. al. (2014), quando um tópico é intangível, devido à sua escala, complexidade ou abstração, as visualizações ajudam a explorar as informações e torná-las compreensíveis. Por isso, seria desejável que o investimento e a inovação na comunicação visual fossem simultâneos aos avanços na pesquisa e na política científica.

As visualizações não são apenas o que mostram, mas também o que ocultam. Ocasionalmente, a ciência pode ter um relacionamento desconfortável com a beleza e o estilo, a ponto de as apresentações mais sugestivas e persuasivas poderem ser interpretadas incorretamente, criando até uma impressão de verdade conhecida como 'cartohypnosis' (PICKLES, 2004). Um exemplo de como uma imagem pode ser representada por meio de cartohypnosis e, ao mesmo tempo, ocultar informações são os mapas. Embora um mapa seja o recurso mais óbvio para exibir informações geoespaciais, nem sempre é a melhor opção para explicar os recursos quantitativos de uma análise. Qual é o melhor mapa? Como não existe uma resposta única, é possível procurar o mais adequado para os dados disponíveis e o objetivo pretendido.

Outro aspecto a considerar é que as visualizações não são realidade, mas a representação de dados derivados de uma série de transformações e filtros que geraram o estilo e o argumento da visualização. Qualquer visualização pode ser reutilizada em comparações equivalentes com outros dados. A seguir veremos como esse processo ocorreu ao longo do tempo.

PRÉ-HISTÓRIA

O início da visualização de dados começou na pré-história.

É possível verificar manuais de caça desenhados em paredes de cavernas e ilustrações sobre a vida espiritual. Por milhares de anos os povos antigos demonstraram manifestações artísticas. Embora não soubessem a escrita na época, foram capazes de produzir obras de arte autênticas, que consistiam em representações gráficas (desenhos, símbolos, sinais) feitas nas paredes das cavernas por homens pré-históricos.

Imagem 1: Arte Rupestre em rocha nos EUA.



Fonte: Mark Herreid / shutterstock.com

Estes desenhos foram usados como guias para entender e sobreviver neste mundo. Isso, em certo sentido, permanece atual até o momento.

À medida que nossa civilização evolui, o mesmo acontece com a maneira como nos comunicamos. As antigas civilizações grega e egípcia tinham um profundo conhecimento sobre o espaço. Eles criaram mapas das constelações e os movimentos do sol. Estes são os primeiros exemplos da ciência moderna, preservada em barro por séculos.

MAPAS, ONDE ESTARÍAMOS SEM ELES?

Os mapas são um dos exemplos mais óbvios de visualização de dados. Sem eles estaríamos literalmente perdidos.

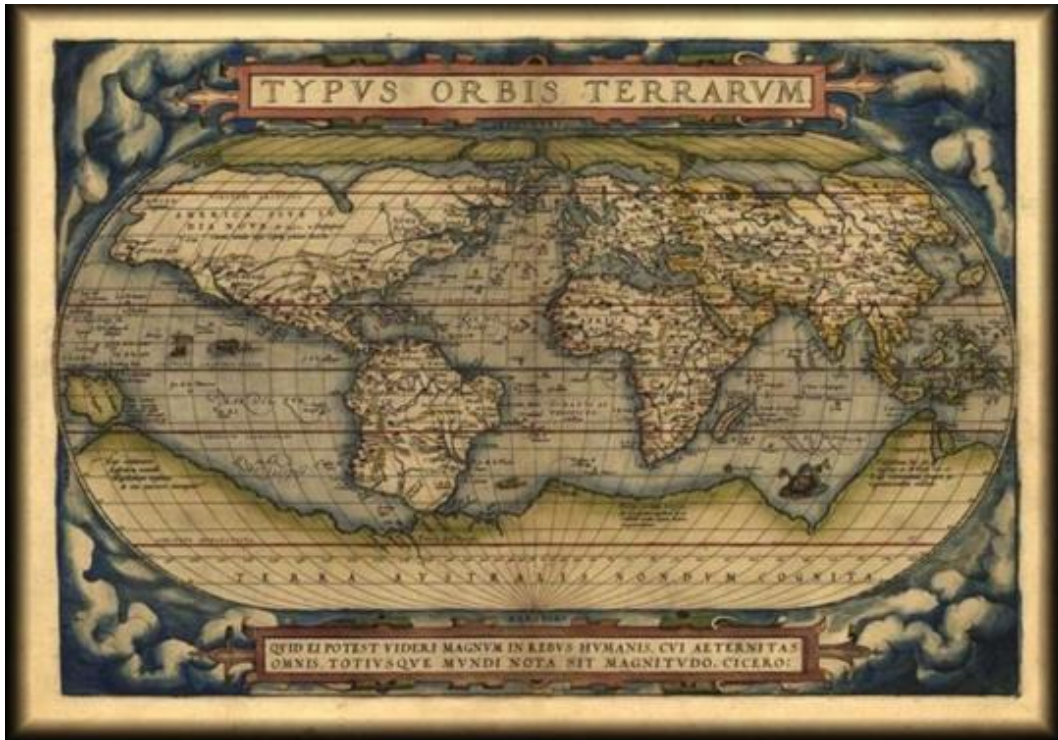
A ciência cartográfica cresceu no final do século XII porque os textos islâmicos começaram a ser traduzidos, renascendo o interesse pela cartografia. Entre as primeiras conquistas está o mapa da Inglaterra, de Matthew Paris, em 1250, com indicação de cidades, portos e mosteiros. Também a partir de meados do século XIV o mapa anônimo da ilha de Gough, no Oceano Atlântico, com a localização de estradas, rios e distâncias entre diferentes pontos. Mas até o relógio e a bússola começarem a ser usados na expressão de cartas ou mapas, havia sido realizado pouco progresso de forma conceitual. Em termos materiais, a gravura em cobre italiana melhorou a qualidade em meados do século XV. Na sequência dos apresentados o mapa do mundo, do alemão Henricus Martellus Germanus (1489), cuja semelhança com o que utilizamos atualmente é marcante (MARTINELLI, 2011).

Na época do Império Romano, os mapas eram usados para entender e planejar o movimento dos exércitos. Estratégias competitivas nasceram desses mapas. Uma das contribuições mais importantes desses impérios.

Somente em 1570 na Bélgica que surgiu o primeiro mapa do mundo, criado por Abraham Ortelius. De fato, Ortelius publicou um atlas composto de 53 mapas, como o

primeiro esforço para reunir todo o conhecimento da cartografia mundial (HISTORY OF MAPPING, 2019).

Figura 2: Mapa Mundi



Fonte: Magnus Mundi. <https://www.magnusmundi.com/a-evolucao-do-mapa-mundi/>

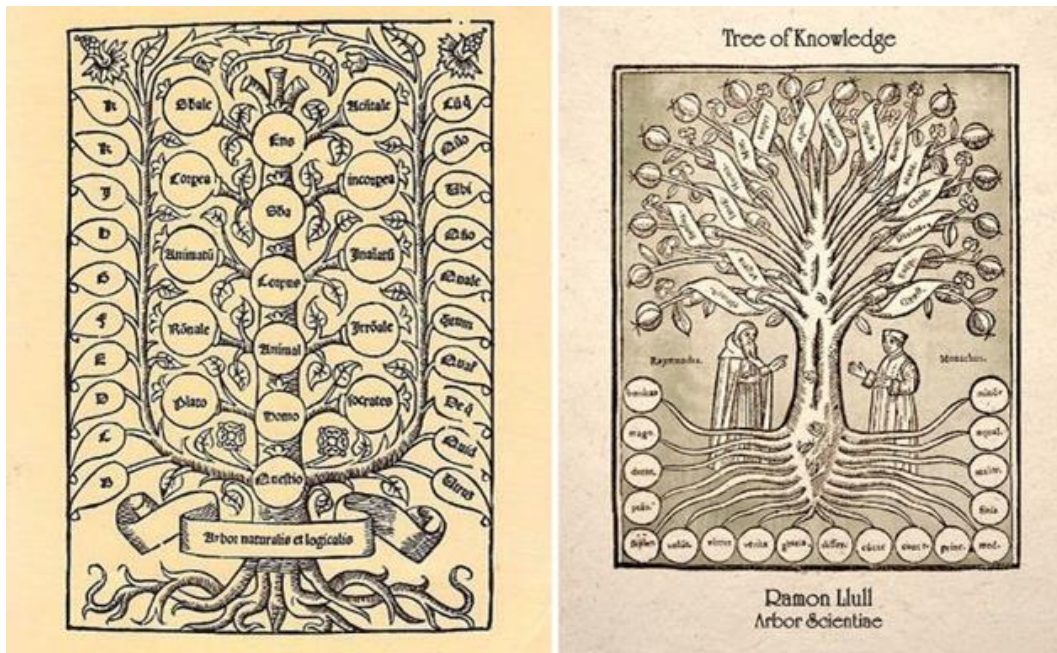
AS ÁRVORES DE RAMON LLULL (1303)

Ramón Llull foi um filósofo alquimista que viveu entre a iluminação cristã e a tecnologia árabe, com resultados fascinantes. Seus diagramas mecânicos de conhecimento serviram como inspiração para o desenvolvimento de sua lógica simbólica (também conhecida como o Antigo Testamento da Computação). Nesta entrada podemos ver duas de suas catorze árvores científicas, onde Llull usa pela primeira vez a estrutura orgânica de uma árvore para representar a origem e o desenvolvimento das grandes áreas do conhecimento.

O livro é um bom exemplo da abordagem metódica e contemplativa de Llull. O Arbor Scientiae é estruturado de acordo com um simbolismo de árvore especial usando a árvore e suas partes como um símbolo da vida e relacionando-as de uma maneira clara e compreensível com a ciência. Uma árvore com suas raízes, tronco, galhos, folhas e frutos representa simbolicamente as diferentes disciplinas da ciência. As raízes representam os princípios básicos de cada ciência; o tronco é a estrutura; os ramos, os gêneros; as folhas, as espécies; e os frutos, o indivíduo, seus atos e seus propósitos. Desta forma, a filosofia natural e moral converge. O objetivo dessa comparação serviu para permitir a compreensão do conhecimento universal, da arte e da ciência (MONTEIRO; GIRALDES, 2011).

Se a analogia parece antiga, é porque ela é - especificamente, 1303 -, mas sua influência tem sido tão decisiva que não é apenas um lugar comum, mas está registrada no DNA de nossa linguagem abstrata, incluindo conceitos básicos e como "a raiz do problema" ou "o assunto central". Na ilustração, as raízes representam os grandes princípios básicos; o tronco é a estrutura; os ramos são os gêneros, as horas são as espécies e os frutos são nós, a última partícula indivisível da grande árvore do conhecimento.

Imagem 3: As árvores da ciência de Ramón Llull



Fonte: <https://www.alamy.com/stock-photo-llull-ramon-12331235-13151316-majorcan-writer-and-philosopher-arbor-111381440.html>

VISUALIZAÇÃO DE DADOS NO SÉCULO XVIII

No século XVIII ocorreram muitos marcos importantes na história universal. Entre eles, podemos citar a Revolução Francesa, a colonização da América e o começo da revolução industrial. Depois da primeira metade do século, um teólogo inglês começou a trabalhar em uma das mais impressionantes visualizações de dados de todos os tempos: a Carta de Biografia de Joseph Priestley, um cronograma que organiza vários séculos de história: marca o começo e a queda dos impérios, as figuras mais proeminentes e os momentos mais relevantes.

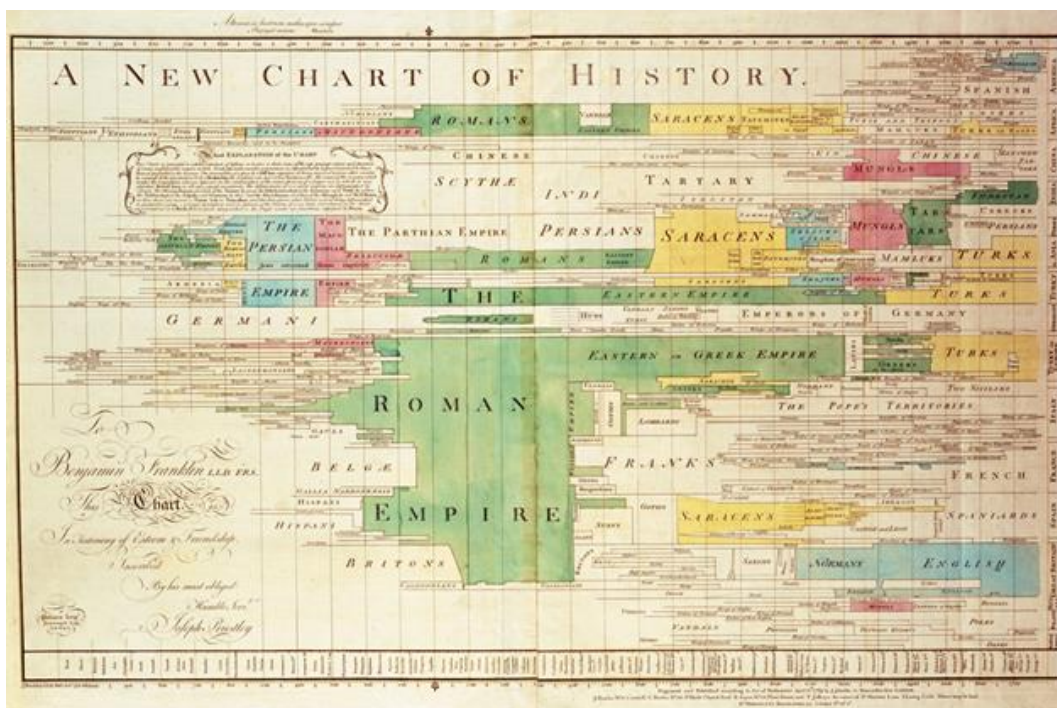
Pela primeira vez na história, foram utilizadas linhas para marcar a duração da vida das pessoas. Priestley revolucionou o modo como a história foi registrada.

Priestley acreditava que o gráfico *A New Chart of History* (1769) permitiria aos alunos "traçar distintamente a dependência de eventos para distribuí-los em tais períodos e divisões, como estabelecerá toda a reivindicação de transações passadas de uma maneira justa e ordenada" (PRIESTLEY, 1786, p. 18).

A visualização de dados desenvolvida por Priestley abrange um vasto período, de 1200 A.C. a 1800 D.C., e inclui dois mil nomes. Priestley organizou sua lista em seis categorias: estadista e guerreiro; Divinas e metafísicos; Matemáticos e Médicos

(filósofos naturais foram colocados aqui); Poetas e Artistas; Oradores e Críticos (autores de ficção em prosa foram colocados aqui); e historiadores e antiquários (advogados foram colocados aqui). O "princípio da seleção" de Priestley era fama, não mérito; portanto, como ele menciona, "o gráfico é um reflexo da opinião atual" (Priestley, 1786). Ele também buscou garantir que seus leitores reconhecessem os desejos no gráfico. Priestley teve dificuldade em designar todas as pessoas listadas para categorias individuais; ele tentou listá-las na categoria sob a qual seu trabalho mais importante havia sido feito. Maquiavel é, portanto, listado como um historiador, em vez de um estadista, e Cícero é listado como um estadista em vez de um orador. O gráfico também foi organizado em ordem de importância; os estadistas são colocados na margem inferior, onde são mais fáceis de visualizar.

Figura 4: A New Chart of History



Fonte: <https://gallica.bnf.fr/ark:/12148/btv1b53093488f>

A MARCHA NAPOLEÔNICA DE CHARLES MINARD (1869)

Quando se trata de documentar os horrores da guerra, há um exemplo de visualização de dados que o torna melhor do que qualquer outro.

50 anos depois da trágica marcha de Napoleão em Moscou, Charles Minard decidiu criar um registro da perda de vidas colossal com a visualização de dados.

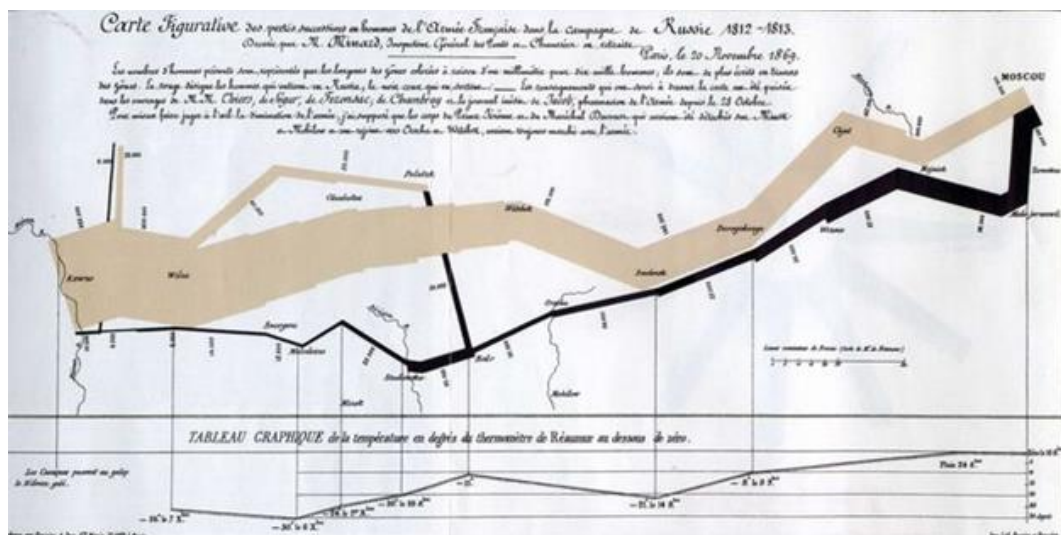
Em junho de 1812, a Grande Armée de Napoleão cruzou o rio Niemen na direção de Moscou. Não foi uma visita de cortesia. O exército maior e melhor preparado do mundo invadiu a Rússia enquanto as tropas do czar Alexandre I recuavam deixando um rastro de fumaça e destruição, sua famosa tática de terra queimada. Quando os franceses chegaram a Moscou, os generais russos haviam partido, mas não antes de incendiá-lo.

Em 19 de outubro, após seis semanas de espera por uma capitulação que não chegou a uma cidade defenestrada, onde não havia comida para os soldados, nenhum pasto para os cavalos, nenhum abrigo contra o inverno, Napoleão começou a voltar para casa. Lá eles encontraram o comandante Mikhail Kutuzov, que estava esperando por eles calmamente no rio Bérézina (LENTZ, 2008).

Dos 422 mil homens que iniciaram a campanha, apenas 100 mil chegaram a Moscou. Daqueles que se voltaram, apenas 10.000 chegaram vivos na França. A "Carta figurativa das perdas sucessivas de homens do Exército francês na campanha russa de Napoleão em 1812" famoso o venerável engenheiro francês Charles Joseph Minard (ele já tinha 80 anos de idade) retrata com precisão e aliviar a dolorosa história desta campanha com um linha dourada que sangra e cinco variáveis dolorosas: a data, seu endereço e localização geográfica, a diminuição das temperaturas, a direção do Exército e a perda de vidas humanas (LENTZ, 2008).

O fluxo de homens foi registrado da seguinte maneira: em cor dourada, aqueles que iniciaram a cruzada e em preto aqueles que retornaram. O gráfico demonstra com detalhes as mortes que ocorreram (PISSETTI, 2015).

Figura 5: Mapa figurativo das perdas sucessivas de homens da Marinha francesa na campanha da Rússia de Napoleão em 1812 por Charles Minard (1869)



Fonte: <http://awcolley.com/EdwardTufte/MarchToMoscow>

MAPA DA CÓLERA DE JOHN SNOW (1855)

A cólera é uma doença explosiva e fulminante: começa com espasmos dolorosos e segue com diarreias, vômitos e coágulos sanguíneos nas veias de desidratação pura, deixando cadáveres de uma cor de pele azulada característica. Embora os primeiros casos documentados na Europa sejam de 1831, a cólera teve o seu ápice com a industrialização das primeiras cidades, com maior ênfase na doentia Londres vitoriana, onde matou 14.137 pessoas apenas em 1848 e outros 10.738 em outro surto em 1853 (MINISTÉRIO DA SAÚDE, 2019).

As grandes instituições científicas estavam convencidas de que a infecção era causada pela inalação de ar contaminado de águas estagnadas ou residuais. Seu principal argumento era que a epidemia atingiu especialmente as áreas mais baixas

da geografia urbana, áreas onde os "sucos" diários da cidade estavam concentrados e onde os setores mais pobres da população viviam (Johnson, 2008).

Entre os dissidentes estava o epidemiologista John Snow, cuja hipótese era de que a doença era causada por um germe e que se espalhava não pelo ar, mas pela água. Quando um surto especialmente violento matou 578 pessoas em uma semana, Snow começou um mapa onde cada morte em cada casa era marcada com um ponto preto e os 13 pontos de abastecimento de água, com um círculo vermelho.

Tufte (2000) conta em *The Visual Display of Quantative Information* que Snow percebeu que a incidência de cólera ocorria quase que exclusivamente entre os que viviam perto do suprimento de água da Broad Street. Mais tarde soube-se que o suprimento da Broad Street havia sido contaminado com as fezes de um bebê doente. "Ele os obrigou a mudar o acesso contaminado", continua Tufte (p.5), "encerrando uma epidemia que custou mais de 500 vidas". O que é verdade até certo ponto.

Conforme Alberto Cairo (2011), professor de infografia e multimídia da Universidade de Miami, aponta que Snow não foi o primeiro a usar mapas epidemiológicos, nem mesmo cólera; Isso seria o mapa de Leeds Dr. Robert Baker no primeiro surto de 1833. Mais importante ainda, o seu impacto não foi tão grande como sugerido por Tufte ou pinta Steven Johnson em *The Ghost Map*, divertido e mistificar sua história ficcionada mapa.

A genialidade de Snow não foi adicionar, mas remover elementos do mapa; enquanto outros afogavam a visualização com muitas referências, sua decisão de limitar os dados a três tornava visível o que não era antes: que a origem de todos os males era uma fonte específica. A verdade é que o establishment médico-sanitário não aceitou sua teoria até os anos 60, quando Snow cultivava malvas há mais de uma década.

Figura 6: O mapa da cólera da neve por John Snow (1854)



Fonte: <https://webtic.eu/late-bytes/highlights/2008/05/subjectieve-bewijsvoering/>

AS FLORES DE FLORENCE NIGHTINGALE (1858)

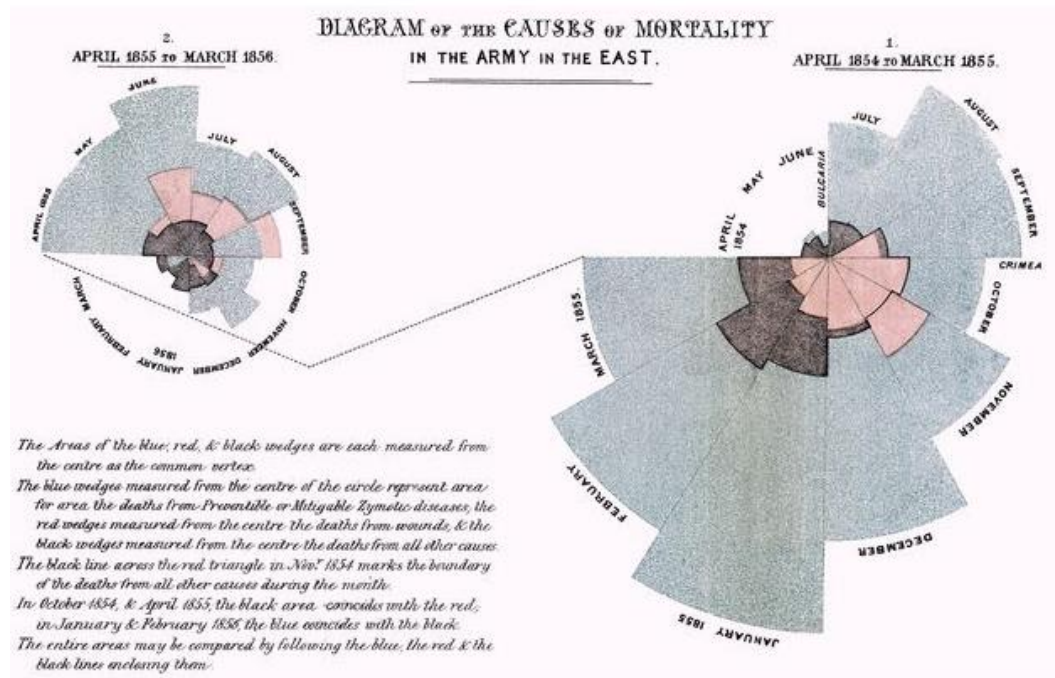
A mãe da enfermagem moderna não foi a primeira a usar gráficos estatísticos para provar uma teoria; seria William Playfair em 1801. Mas poderia ter sido o primeiro a usá-los para mudar a política de saúde de todo um país e a primeira mulher membro da Sociedade Estatística de Londres. Como Minard, Nightingale usou motivos gráficos para narrar a perda de um Exército, neste caso o que o Império Britânico enviou à Guerra da Crimeia (COSTA et. al., 2009).

Em 21 de outubro de 1854, Nightingale partiu para a Crimeia com uma equipe de 38 enfermeiras voluntárias. Relatórios anteriores à viagem antecipou uma cena horrível, mas, uma vez lá, Florence descobriu que, mesmo no pior momento da guerra grande parte dos soldados morreram de feridas recebidas no campo de batalha, devido as condições de higiene que estavam expostos no quartel. Em seu relatório *Notes on matters affecting the health, efficiency, and hospital administration of the British Army* (Nightingale, 1858), ela dividiu as perdas em três classes: em vermelho, morte por feridas; em azul, "doenças de zimoses evitáveis ou mitigáveis" (doenças infecciosas como cólera e disenteria); em preto, todas as outras causas.

Cada parte do gráfico é relativa ao número de incidentes; quanto mais mortes, maior a pétala correspondente. Flores Nightingale salvou mais vidas britânicas do que todas as mães no Reino Unido rezando o terço juntos, porque, apesar de não ter nada de novo, a representação das mortes ocorreu de forma tão dramática, que

imediatamente a rainha Victoria compreendeu suas implicações, levando o trabalho realizado por Nightingale para todos os hospitais ingleses (LOPES; SANTOS, 2010).

Figura 5: As rosas por Florence Nightingale



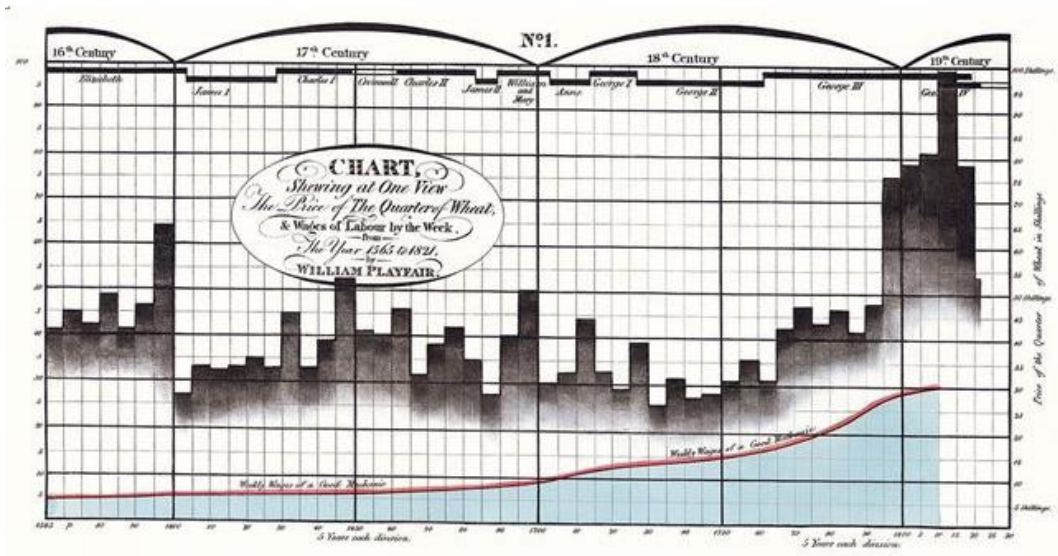
Fonte: <https://robsonmami.blogspot.com/2018/05/13-diagrama-de-rosa-de-florence.html>

O PREÇO DO TRIGO DE WILLIAM PLAYFAIR (1821)

O economista escocês William Playfair é considerado por muitos como o pai do gráfico estatístico: seu livro *The Commercial and Political Atlas*, que enviou a Louis XVI pouco antes de o monarca perder a cabeça (1787), já continha gráficos de barras. Com o *Statistical Breviary* (1801), inventou três padrões de visualização que são utilizados até hoje: o gráfico de linha, o gráfico de barra e o gráfico de pizza (PEREIRA, 2015).

A economia foi sua maior vocação, especialmente o contraste: suas impressões mais notáveis mostram a cobrança de impostos em diferentes países da Europa (mostrando que a Grã-Bretanha estava pagando demais) e o fluxo de bens comerciais entre as nações, levando a temperatura para o mercado de importação e exportação no continente. O gráfico que selecionamos cruza "o salário semanal de um bom mecânico" com "o preço de um quarto de trigo" sob diferentes monarcas (PEREIRA, 2015).

Figura 7: História da Visualização de Dados: o preço do trigo de William Playfair (1822)



Fonte: <https://pantheon.ufrj.br/bitstream/11422/6889/1/CPinto.pdf>

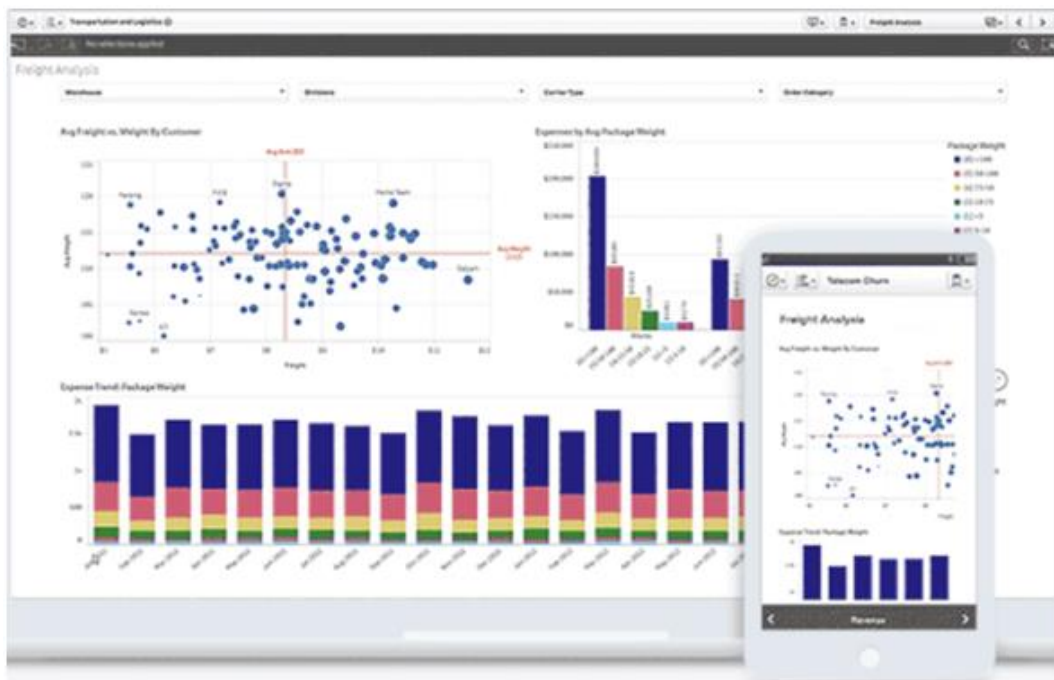
VISUALIZAÇÃO DE DADOS MODERNOS

Ao longo da década de 1990 e início dos anos 2000, a mídia e as empresas incorporaram a infografia impressa como uma maneira mais esclarecedora de representar números. A novidade foi que os usuários passaram a interagir com tais visualizações, fato que possibilitou tanto a internet quanto os avanços tecnológicos relacionados à programação de computadores.

As maneiras pelas quais podemos mostrar dados avançaram exponencialmente: de infográficos e gráficos em rede a centros de comando e salas de guerra.

A internet e demais mídias digitais permitiram que informações e ideias fluíssem livremente, facilitando o aprendizado de outras pessoas.

Figura 8: Dados estruturados



As novas formas de apresentar dados possibilitaram encontrar *insights* instantaneamente, que agora podem ser exibidos em qualquer lugar do mundo. Assim, as novas ferramentas e técnicas de visualizações de dados resolveram um antigo problema, em que conhecimento é visto como sinônimo de poder, mas ter muita informação pode ser contraproducente.

Com o advento do *big data* e a democratização da visualização de dados, tornou-se fundamental que os executivos e analistas de negócios tenham o poder de escolher as informações relevantes para eles, de acordo com o projeto em que estão trabalhando.

O ritmo de criação de dados está aumentando ano após ano de uma maneira notável e espera-se que aumente ainda mais graças ao grande número de sensores e dispositivos que enviarão informações em todos os momentos. Todos esses dados devem ser coletados, explorados, processados, armazenados, analisados e finalmente convertidos em informações valiosas. Temos que saber como mostrar essa informação valiosa de forma adequada e compreensível, para que todos os destinatários possam entendê-la e ajudar na tomada de decisões (MCINERNY, 2014).

Em resumo, poderíamos indicar quatro pontos principais para fazer uma visualização de dados:

- Ter um conjunto de dados limpo: significa um conjunto de dados no formato apropriado para a maioria das ferramentas de visualização.
- Única mensagem de comunicação: selecionar uma única mensagem, que será destacada em cada gráfico. Dependendo do público, o tipo de gráfico e o nível de precisão necessário.
- Escolher o gráfico adequado: A forma de um gráfico depende do que queremos transmitir: por exemplo, comparar, distribuir, etc.
- Design e cor: destacar o que é essencial, usando cores.

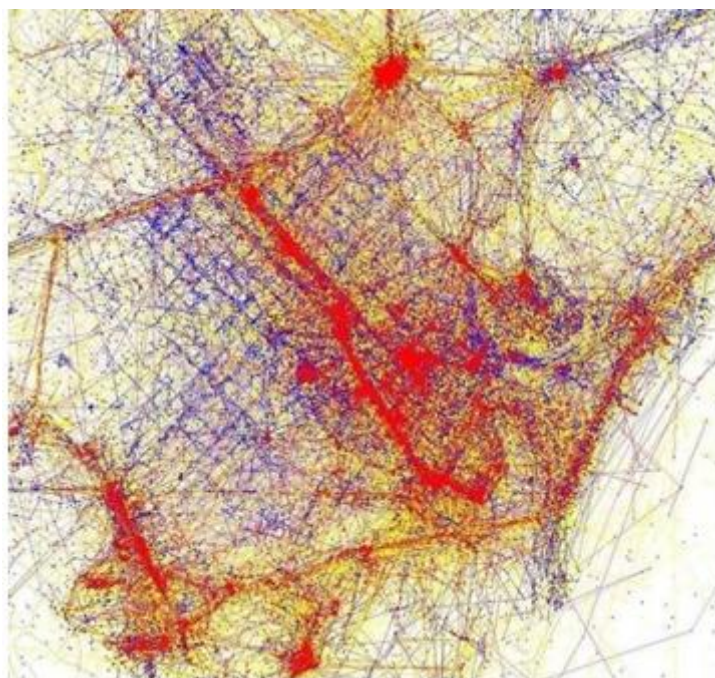
Sempre que possível, é recomendável adicionar interação em nossas visualizações de dados, esta ação facilita e amplia a compreensão das informações que queremos transmitir.

Os objetivos que estabelecemos ao fazer uma visualização interativa são os seguintes:

- Colocar os dados em contexto;
- Que o usuário seja capaz de explorar os dados;
- Que o usuário tenha a possibilidade de encontrar padrões.

Como exemplo para encontrar padrões, Fisher (2019) criou mapas das cidades mais turísticas do mundo com base na geolocalização de *Tweets* em um trabalho chamado *Locals and Tourists*¹ que nos dá uma ideia muito precisa de onde os turistas se movem (fazendo *Tweets*).

Figura 7: Visualização de habitantes x turistas



Fonte: <https://brilliantmaps.com/tourists-vs-locals/>

Desse modo, o processo de comunicação contempla uma série de elementos para cumprir sua finalidade: facilitar que uma mensagem chegue a uma pessoa ou grupo de pessoas. Assim, os dados se tornam o elemento chave da comunicação, desde que esteja devidamente estruturado para ser transmitido. Nesse processo, os fatores fundamentais da comunicação são o autor, o receptor, o tema da mensagem e os dados em que a mensagem é emitida.

Os gráficos têm uma certa magia. O perfil de uma curva revela uma história completa em segundos. A evolução de uma epidemia, uma crise ou uma era de

¹ <https://brilliantmaps.com/tourists-vs-locals/>

prosperidade. A curva dá informação à nossa mente, desperta a nossa imaginação e convence. Este processo tem ocorrido ao longo da história até os dias atuais.

BIG DATA FUNCIONA POUCO SEM VISUALIZAÇÃO DE DADOS ADEQUADA

Temos muitos exemplos de como o Big Data transforma nossa relação com o cotidiano. De algo tão visual quanto a redução de engarrafamentos, através de seu uso para a personalização de produtos e serviços, ou para melhorar a experiência de videogame *multiplayer*, a coleta e análise de enormes quantidades de dados não estruturados é um grande passo à frente. Mas vamos falar sobre visualização de dados:

Os dados são inúteis se não formos capazes de visualizar o que estamos procurando. Já fizemos progressos em uma ocasião, o desafio para as empresas é extrair o valor dos dados e, para isso, é necessário ter as melhores ferramentas de visualização. Com o tempo, é verdade que os modelos de inteligência artificial e *Machine Learning* serão aqueles que processam essas enormes quantidades de dados (já é feito, de fato, em alguns campos), mas sempre haverá um fator humano decisivo.

A visualização de dados crescerá em importância no curto prazo. Se trata de um termo geral que descreve qualquer esforço para ajudar as pessoas a entender a importância dos dados, colocando-os em um contexto visual. Padrões, tendências e correlações que podem passar despercebidas em dados baseados em texto podem ser mais facilmente expostos e reconhecidos com o software de visualização de dados. Se falamos de *Big Data*, a visualização de dados é crucial para direcionar a tomada de decisões em alto nível com maior sucesso. A análise do *Big Data* tem imenso potencial para ajudar as empresas a tomar decisões e colocar a empresa diante de um futuro realista (SATO, 2017).

Apesar de todos os benefícios que o *big data* oferece, o processamento de dados dessa magnitude não é necessário para todas as empresas e só crescerá à medida que tivermos novas aplicações em nossas mãos para usá-las. Agora, o desafio não se baseia em gerar e capturar mais dados, mas baseia-se na descoberta de novas maneiras de condensar, interpretar e tomar decisões sobre esses dados. A chave está na visualização de dados, que é um dos métodos mais novos e mais conhecidos para reduzir ou ilustrar dados de maneira simplificada e visual (MCINERNEY, 2014).

Nós não somos "projetados" para processar números complexos e associar esses números a conceitos abstratos, mas reconhecer padrões visuais rapidamente. A capacidade humana de reconhecimento de padrões visuais é inata. Portanto, representar números complexos como padrões visuais nos permite aproveitar nossas habilidades analíticas naturais.

É evidente que o interesse na visualização de dados está aumentando, é uma tendência clara e, ao mesmo tempo, uma ferramenta poderosa para mostrar tendências e representar dados simples, etc. Mas como os resultados da análise complexa de grandes volumes de dados ou análise preditiva ou prescritiva podem ser explicados? A visualização de dados resolve todos os nossos problemas de compreensão de *big data*? Quais são os limites?

Estes são quatro possíveis limites ou "problemas" do conceito de visualização de dados:

- a) Ferramentas de visualização de dados mostram, mas não explicam

Se analisarmos o processo pelo qual as empresas obtêm informações, ela não mudou muito nos últimos 30 anos, pois os analistas analisam os dados e depois escrevem relatórios. Por outro lado, embora tenhamos poderosas ferramentas de visualização que esperam que o usuário seja especialista em todos os tipos de dados e na seleção das melhores técnicas para visualizar e representar os dados.

b) A simplificação excessiva dos dados

Uma das maiores atrações da visualização é a capacidade de obter grandes volumes de dados e simplificá-los de maneira sintética e compreensível. No entanto, isso pode nos levar a tirar conclusões infundadas, se tentarmos apenas pegar milhões de registros de dados e limitar suas conclusões a alguns gráficos de um painel, já que podemos negligenciar completamente certos modificadores significativos que poderiam mudar completamente as suposições.

c) As limitações humanas dos algoritmos

Este é o maior problema em potencial e também o mais complicado. Qualquer algoritmo usado para reduzir dados a ilustrações visuais é baseado em contribuições humanas, e as contribuições humanas podem estar fundamentalmente erradas. Por exemplo, um programador que desenvolve um algoritmo pode destacar os diferentes dados que são “mais” importantes a considerar e descartar completamente os outros, o que não seria ideal para todas as empresas e situações, especialmente se houver exceções ou situações únicas que exijam um abordagem alternativa. O problema é agravado pelo fato de que a maioria dos sistemas de visualização de dados são implementados em uma escala global e se tornam algoritmos que se adaptam a todos e não atendem às necessidades específicas de cada usuário.

d) Da visualização de dados aos dados *Storytelling*²

Os gráficos são excelentes para transmitir ideias simples rapidamente, mas às vezes não são suficientes. Se palavras pudessem ser substituídas por pictogramas, elas teriam sido há muito tempo. Para expressar uma situação complexa, frases e frases são necessárias, juntamente com um sistema que é inteligente o suficiente para articular seu processo de raciocínio. Além do painel automático, o que é realmente valioso na tomada de decisão que leva à ação é a narrativa dos dados com o componente de conhecimento contextual. É importante ressaltar que a linguagem também garante que o usuário final realmente entenda.

Talvez a inteligência artificial ofereça algum tipo de solução ao escrever automaticamente uma narrativa que explique as idéias e os fatores por trás das mudanças nos dados. É evidente que o futuro do treinamento será direcionado ao *storytelling*. Como vimos Big Data serve pouco sem uma visualização de dados adequada.

CONSIDERAÇÕES FINAIS

A visualização dos dados pode ser obtida através de um gráfico, tabela ou mapa com uma avaliação comparativa dos resultados obtidos através de uma investigação baseada em dados. A representação gráfica pode ser uma variedade de formas de

² De acordo com o National Storytelling Network (2019), dos Estados Unidos, “Storytelling é uma forma de arte muito antiga e uma valiosa forma de expressão humana usada de diversas maneiras.”

representar os dados e também pode contribuir com mais propósitos para a análise e apresentação de informações. Estes gráficos de dados proporcionam uma maneira de simplificar a comunicação de uma informação complexa e pode nos ajudar a identificar padrões e relacionamentos que se desenvolvem em um determinado período.

Os dados exibidos em visualizações não contêm a profundidade da informação de um conjunto completo de dados contemplados numa informação descritiva, portanto existe o risco que o contexto da informação possa ser perdido ou comunicado incorretamente. Os gráficos de dados podem sugerir relações ou tendências dentro e entre as variáveis estudadas, mas não se pode presumir que eles sejam ilustrativos de uma condição até que ela seja verificada cientificamente por análise estatística ou certificada a sua autenticidade de acordo com as fontes levantadas. (MENDONÇA 2010; SATO, 2017).

A visualização de dados crescerá em importância no curto prazo. Se trata de um termo geral para ajudar as pessoas a entenderem a importância dos dados ao colocá-los em um contexto visual. Padrões, tendências e correlações que podem passar despercebidas em dados baseados em texto passam a ser expostos e reconhecidos mais facilmente com o uso de técnicas e ferramentas de visualização.

Artigo recebido em 09/07/2019 e aprovado em 15/10/2019.

REFERÊNCIAS

- CAIRO, Alberto. *El arte funcional*. Madrid: Alamut. 2011.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. *Focus+Context*. In: Card, S. K.; Mackinlay, J. D., Shneiderman, B. *Readings in information visualization: Using visualization to think*. San Francisco, Morgan Kaufmann Publ., 1999, p. 307-309.
- COASTING-EAVES, Patricia; MACDONALD-ROSS, Michael. William Playfair (1759-1823). *Statistical Science*. V. 5, n.3, 1990. p. 318-326. Disponível em: <https://projecteuclid.org/download/pdf_1/euclid.ss/1177012100>. Acesso em: 26 jul. 2019.
- COSTA, Roberta et al. O legado de Florence Nightingale: uma viagem no tempo. *Texto contexto - enferm.*, Florianópolis, v. 18, n. 4, p. 661-669, Dec. 2009.
- FISCHER, Eric. Tourists Vs Locals: 20 Cities Based On Where People Take Photos. Disponível em: <<https://brilliantmaps.com/tourists-vs-locals/>>. Acesso em: 26 jul. 2019.
- HABER, R. B.; McNabb, D. A. Visualization idioms: a conceptual model for scientific visualization systems. *Visualization in scientific computing*, 1990. p. 74-93.
- HISTORY OF MAPPING. Disponível em: <<https://www.icsm.gov.au/education/fundamentals-mapping/history-mapping>>. Acesso em: 26 jul. 2019.
- MARTINELLI, Marcelo. *Mapas da Geografia e Cartografia Temática*. São Paulo: Contexto. 2011.
- MENDONÇA, Ítalo de Oliveira. *O design da informação aplicado à infografia*. 2010. Monografia (Graduação em Design — Habilitação Gráfico) — Universidade Federal de Santa Catarina — UFSC.

MINISTÉRIO DA SAÚDE. *Cólera: causas, sintomas, transmissão, tratamento e diagnóstico*. Disponível em: <http://www.saude.gov.br/saude-de-a-z/colera>. Acesso em: 26 jul. 2019.

MONTEIRO, Silvana Drumond; GIRALDES, Maria Júlia Carneiro. Aspectos lógico-filosóficos da organização do conhecimento na esfera da ciência da informação. *Inf. & Soc.:Est.*, João Pessoa, v.18, n.3, p. 13-27, set./dez. 2008.

NATIONAL STORYTELLING NETWORK. Disponível em: <<https://storynet.org>>. Acesso em: 15 set. 2019.

NIGHTINGALE, Florence. Notes on matters affecting the health, efficiency, and hospital administration of the British Army. *Internet Archive*. Londres. 1858. Disponível em: <<https://archive.org/details/b20387118/page/n8>>. Acesso em: 27 jul. 2019.

PEREIRA, Augusto Bernardi. *Visualização das informações: o site encyclopædia metallum: the metal archives*. Curso de Gestão da Informação. Curitiba. 2015. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/41085/2015-2%20TCC_Augusto%20Bernardi%20Pereira.pdf?sequence=1&isAllowed=y>. Acesso em: 26 jul. 2019.

PICKLES, John. *A History of Spaces: Cartographic Reason, Mapping, and the Geo-coded World*. Londres: Routledge Taylor & Francis Group. 2004.

STEVEN, Johnson. *O mapa fantasma*. Rio de Janeiro: Zahar, 2008.

LENTZ, Thierry. *Napoleão*. Tradução C. Egrejas. São Paulo: Editora Unesp, 2008.

LOPES, Lúcia Marlene Macário; SANTOS, Sandra Maria Pereira dos. Florence Nightingale – Apontamentos sobre a fundadora da Enfermagem Moderna. *Revista de Enfermagem Referência*. n. 2 - Dez. 2010. pp.181-189.

MCLNERNY, et. al. Visualización de información para ciencia y política: involucrar a los usuarios y evitar sesgos. *Trends in Ecology & Evolution*. Disponível em: <[https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347\(14\)00016-0](https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347(14)00016-0)>. Acesso em: 15 set. 2019.

PISSETTI, Rodrigo Fernandes. Princípios fundamentais do design analítico. *Revista Imagem*. V. 5, N. 1, 2015.

PRIESTLEY, Joseph, 1733-1804. *A description of a new chart of history, containing a view of the principal revolutions of empire that have taken place in the world*. London, Printed for J. Johnson, 1786. Disponível em: <https://catalog.hathitrust.org/Record/000438914>. Acesso em: 26 jul. 2019.

SATO, Susana Narimatsu. *A infografia na divulgação científica: um estudo de caso da revista Pesquisa FAPESP*. 2017. 155f. Dissertação (Mestrado) - Escola de Comunicação e Artes, Universidade de São Paulo, 2017.

TUFTE, EDWARD R. The Visual Display of Quantitative Information. *American Journal of Physics*, 2000. v. 31, n. 11.