

The Truthful Art: Data, Charts and Maps for Communication

Alberto Cairo

Sérgio Seabra Moreira e Mirtes Cristina Marins de Oliveira

1st Edition. Berkeley: New Riders, 2016. 382p.

1 O autor

Alberto Cairo é um contador de histórias que conversa com o leitor com a mesma atenção dedicada à filha de oito anos quando esta lhe pergunta porque os planetas não param de girar – a história de sua resposta, construída após algum estudo e muitas imagens produzidas, ilustra a introdução de *The Truthful Art* (2016). Com farto uso de notas de rodapé, apresentando os autores com quem dialoga no trajeto em sua compreensão dos conceitos abordados e se utilizando de variados exemplos, Cairo possui um didatismo que respeita a inteligência de seu interlocutor.

À experiência da prática profissional como consultor de projetos especiais ou diretor de infografia em muitos periódicos – inclusive com uma passagem pelo Brasil, na Editora Globo, Revista Época, entre 2009 e 2012 –, soma-se a atuação como professor. Atualmente, leciona Infografia e Visualização na Universidade de Miami, onde também é diretor do programa de Visualização no Centro de Ciências Computacionais. Cairo atuou em mais de uma oportunidade como orientador e tutor de cursos online dedicados a temas relacionados a Visualização de Dados e Infografia pelo Centro Knight para Jornalismo nas Américas.

Não obstante tenha editado outros trabalhos e anuncie em seu site¹ o lançamento, ainda em 2017, de um livro baseado em sua tese de doutorado, a presente obra é tratada como o segundo livro de uma trilogia que deseja encerrar em 2019 com *The Insightful Art*, anunciado no epílogo com o objetivo de cobrir os tópicos do último capítulo. O primeiro da série, *The Functional Art* (2012) é dedicado à Infografia e às narrativas que esta pode realizar, sobretudo aquelas ligadas ao jornalismo, um campo de atuação de onde o autor parece ter partido rumo a territórios mais abrangentes e afeitos à sua curiosidade, pode-se dizer, científica.

¹ <http://www.thefunctionalart.com/2016/01/the-truthful-art-is-finished-what-comes.html>

2 A obra

A utilização de gráficos de modo equivocado é um fato consumado. Se não a cada esquina, em alguma mesa de escritório ou mesmo em um periódico estabelecido. O acesso a *softwares* ou aplicações web permite a mais gente sintetizar dados em formas visuais, não bastasse a existência de programas como o Excel e todas as suas opções de gráficos em 3D, o que têm ajudado a deseducar grande quantidade de profissionais administrativos. O trabalho de Alberto Cairo sobrevoa, bem do alto, mas com mergulhos certos, este ambiente.

The Truthful Art – data, charts, and maps for communication pode ser lido por iniciantes e iniciados na lida de interpretar e apresentar dados por meio de visualizações. Quem se achega ao campo será recebido com uma boa síntese de conhecimentos distribuídos de forma cuidadosa e encadeada, com base em princípios estatísticos – que compõem o conteúdo mais denso e extenso (a Parte III, *Functional*, constituída por sete capítulos) do trabalho. Aqueles que já tem uma prática estabelecida de atuação profissional também têm portas de entrada na obra, seja como um repositório de fontes para aprofundamento, seja pelo posicionamento crítico de Cairo frente a trabalhos descuidados. Uma estratégia recorrente é a do redesign, ora de gráficos ruins, ora de gráficos que podem ser muito melhores, ou, também, para atender a diferentes finalidades.

Além da já citada terceira parte dedicada aos aspectos funcionais, o livro possui outras três. A primeira – *Foundations* – e a segunda – *Truthful* –, com dois capítulos cada, apresentam as bases e, pode-se dizer, o ponto de vista do autor sobre o que são as visualizações de dados e o que estas devem conter e evitar, substancialmente. A quarta parte – *Practice* –, mais curta, é dedicada a apresentar trabalhos que, na visão de Cairo, propõem-se a aperfeiçoar e expandir o vocabulário e a gramática visual dos gráficos ou, em outro sentido, a popularizar a sua leitura, educando o público a ler as formas existentes de visualização – papel que as *news applications* têm cumprido especialmente. Em síntese, o livro defende princípios para o trabalho de escolher, descartar e visualizar dados de um modo que estes tornem-se significativos.

3 Fundamentos

Cairo aborda a visualização como um termo guarda-chuva que abriga qualquer tipo de representação visual projetada para permitir comunicação, análise, descoberta e exploração mas, logo de início, informa que não abordará os variados ramos desta extensa árvore, apenas aqueles que objetivam comunicar-se efetivamente com o público em geral. Deste recorte, já bastante extenso, elege como foco o gráfico – uma exibição em que dados são codificados com símbolos de diferentes formas, cores e proporções. De passagem, cita

2 O autor define infográfico como uma representação visual de informação composta de múltiplas seções, incorporando não apenas gráficos, mas mapas, ilustrações e texto com a intenção de comunicar uma ou mais mensagens específicas proporcionando explanação e contexto.

o infográfico², mas tratará sobretudo de gráficos isolados ou inseridos em interfaces de visualização de dados ou de aplicações de notícias (*news applications*), que permitem ao usuário buscar a informação que mais lhe diz respeito em meio a uma grande quantidade de possibilidades.

O alicerce do trabalho do designer de visualização com o objetivo de comunicar informações deve ser construído sobre quatro atributos que, se atendidos, geram um quinto: o esclarecimento (*enlightening*). O primeiro deles, que dá nome à obra e merecerá toda a segunda parte do livro é *truthful*. A visualização precisa ser baseada em uma completa e honesta pesquisa sobre os dados, portanto verdadeira, confiável. A segunda qualidade é ser funcional, construída de um modo que permita às pessoas realizarem operações significativas de compreensão a partir do que veem. O terceiro atributo: *beautiful*, no sentido de ser atrativa, intrigante, esteticamente agradável. A quarta qualidade, *insightful*, é aquela que imprime ao trabalho a capacidade de revelar uma evidência que leve a uma compreensão súbita, que forneça um atalho ao caminho do entendimento. Ou, em outra medida, muito frequente em visualizações interativas, que permita a construção do conhecimento por meio da exploração gradual e deliberada de um conjunto de informações. Uma visualização, logo se vê, precisa ter nascido de uma legítima ação projetual de design, este sendo uma “expressão de propósito”, na expressão de Charles Eames, recuperada pelo autor.

Afora as situações em que se quer, de modo consciente, ludibriar as pessoas com um gráfico deliberadamente viciado, verdade e inverdade não são conceitos absolutos, mas estão nas extremidades de um mesmo espectro. Há, entre as pontas, muitos descuidos e autoenganos que ocorrem entre os modelos mentais que têm tanto o designer quanto cada um daqueles que compõem a audiência de seu trabalho. Os equívocos podem surgir de armadilhas mentais – *mind bugs*, no original – como a de reconhecer padrões mesmo onde estes não existem, um fenômeno cognitivo de percepção de conexões em dados aleatórios, ou partir de histórias pré-definidas, por ativismo, ideologia ou fé. Outra cilada é a tentativa de reduzir dissonâncias, atendendo à tendência que temos a buscar a confirmação de nossas crenças. A “verdade” também sai prejudicada por escolhas técnicas ruins – tais como a adoção de um eixo duplo que distorça as linhas ou a escolha de uma forma inadequada para o tipo, ou a quantidade, de dados que se quer apresentar.

Para superar estes obstáculos na oferta de informações visuais verdadeiras, o autor sugere a busca por profundidade e alargamento da extensão dos dados com os quais se quer trabalhar. Trata-se, no primeiro caso, de operar com a mesma informação distendida no tempo ou por mais categorias e, no segundo, de trazer mais variáveis que permitam ampliar a compreensão do fenômeno que se quer apresentar. No exemplo do livro, Cairo reconstrói a notícia de que “mais de ¼ dos jornalistas graduados desejavam ter escolhido outra

carreira”, para melhor contextualizá-la, de modo a subverter sua intenção original.

Neste contexto, cita o frequente erro de interpretação de uma máxima de John Maeda, pelo qual deixa-se de lado a segunda frase da oração “Simplicidade é sobre subtrair o óbvio e adicionar o significativo”. Para o autor, que diz ter aprendido a diferença entre *simplifying* e *clarifying* com o designer Nigel Holmes, não se deve simplificar em demasia a visualização, mas purificá-la, por vezes aumentando a quantidade de informação, não reduzindo. Sem descuidar do conteúdo, Cairo defende que o modo como se apresentam os dados importa tanto quanto a solidez da informação. E complementa que, neste sentido, deve-se adotar como prática a divulgação da fonte dos dados e a metodologia utilizada.

A estratégia que defende no quarto capítulo é a postura científica, nascida da curiosidade do designer, algo que Cairo demonstra possuir. Neste contexto, apresenta os tipos de variáveis, cuja codificação permite alçar a conjectura ao nível de uma hipótese: 1) nominal ou categórica, distintas por sua identidade, sem peso quantitativo; 2) ordinal, cujos valores indicam intensidade ou magnitude; 3) de intervalo, em que a medida é baseada em incrementos de mesmo tamanho sem que haja um ponto zero; e 4) de proporção, que possuem um zero significativo e, por isso, podem ser comparadas diretamente. Também apresenta as diferenças entre variáveis discretas e contínuas e estudos transversais, com dados coletados em um momento específico, e longitudinais, coletados ao longo do tempo.

Dentre as “regras de ouro”, pondera, deve-se sempre perguntar o que é necessário comparar com o que, quem, quando e onde, além de procurar pelas peças que estão faltando no modelo, isto é, aumentar a profundidade e a largura da pesquisa até um ponto razoável. Uma preocupação constante do autor é problematizar os dados com os quais se está trabalhando, sejam estes exemplos pessoais, ligados à mídia, ou estudos que circulam na comunidade científica. Ao final do quarto capítulo, Cairo destaca a incerteza, o ruído e a aleatoriedade que rondam os dados – além de sua frequente natureza incompleta, com erros e redundâncias, para algumas das quais recomenda ferramentas de limpeza como o *Open Refine*³.

4 A melhor representação visual possível

Na terceira parte do trabalho, antes de entrar no terreno mais denso da estatística, Cairo busca compartilhar algumas pistas que levem à melhor resposta para a pergunta que mais lhe fazem os estudantes: “Como você sabe qual é o gráfico correto para representar os dados?”. São apresentados alguns catálogos de formas de visualização, com seus respectivos endereços web, pelos quais pode-se navegar em busca daquela que possibilite, para os dados que se têm em mãos, a melhor solução. A ferramenta favorita do autor é a hierarquia de

tarefas perceptivas elementares ou métodos de codificação, feita pelos estatísticos William S. Cleveland e Robert McGill – mas que não deve ser adotada de forma acrítica. A resposta à questão recorrente, afinal, é que os dados precisam ser vistos por diferentes ângulos para deles tirar sua melhor forma.

Repetidas vezes, com muitos exemplos, o autor expõe que não se deve perder de vista a tarefa destinada ao gráfico e o que ele deveria revelar: o melhor modo de se chegar à solução é testar várias possibilidades. Seleccionada a melhor forma gráfica, deve-se cuidar de organizar as variáveis, da maior para a menor, alfabeticamente ou qualquer outro critério, tendo em conta que as regras de visualização valem tanto quanto o resultado dos testes que devem ser feitos com os leitores.

O sexto capítulo dá início à escalada que o autor afirma ser apenas um vislumbre de como a estatística pode ser surpreendentemente prazerosa na busca por padrões e tendências que estão à espreita, tanto quanto os desvios destes padrões. As histórias interessantes podem surgir de ambos: da norma ou de suas exceções. Todo o capítulo é construído sobre uma planilha com dados do IDEB⁴ brasileiro, pela qual explicará em detalhes a definição e os usos da média, moda, mediana, intervalo e o nível de distribuição, por meio da utilização de diferentes gráficos – histogramas e *strip plots*, *box plots* – de modo a encontrar diferentes histórias nos dados. No capítulo seguinte, são apresentados os conceitos de variância e desvio padrão. Não obstante estes cálculos possam ser entregues prontos por diferentes softwares, o autor lança mão de apresentar um punhado de fórmulas, indicando que o designer responsável deve tê-las em mente ao trabalhar com visualização de dados. A mesma estratégia do capítulo anterior é utilizada, agora com uma planilha com salários de trabalhadores de TI da Nigéria e dos Estados Unidos, de modo a identificar em que país eles são melhor remunerados, considerando, evidentemente, os contextos de cada um dos dois países.

Revelando Mudanças, o capítulo 8, parte da análise da taxa de desemprego na Espanha e das muito frequentes análises de séries temporais para se observar a tendência das variáveis e se há algum elemento sazonal ou ruído. O autor observa que explorar múltiplos níveis de agregação é primordial para evitar paradoxos e compara os valores da série temporal a partir de índices criados especialmente para cada abordagem: uma média de vários anos e outra do ano inicial, buscando demonstrar as peculiaridades e pertinências de cada uma delas, destacando quando é necessário utilizar a escala logarítmica ou como as séries temporais podem ser enganosas, quando há uma lacuna nos dados. Ao apresentar o gráfico de horizonte, concebido por Jorge Camões, Cairo rende-se às soluções elegantes do colega ao dizer que “a estética, a brincadeira e os cuidados com tipografia, cor e composição são tão importantes na visualização artística como na apresentação de resultados analíticos” (p.221).

3 <http://openrefine.org/>. Ferramenta de código aberto e fácil utilização que permite realizar transformação de dados textuais em larga escala de modo a sintetizar diferentes entradas de um dado em uma variável comum. Por exemplo: “SAO PAULO, São paulo e Sao palo” como “São Paulo”.

4 Índice de Desenvolvimento da Educação Básica.

Em *Vendo Relações*, capítulo 9, o autor inicia destacando quão frequentes são os exemplos na mídia de dados correlacionados que não têm, de fato, nenhuma relação, muito menos uma de causa e efeito. Neste capítulo, são abordados de modo mais detalhado os gráficos de dispersão (*scatter plots*), abordando os conceitos de associação e correlação. Na obra, Cairo aborda apenas um tipo de regressão, a dos mínimos quadrados lineares e indica a bibliografia para as demais. Para a exploração de dados multivariados, comuns no campo científico, apresenta o uso de matrizes de dispersão, mapas de calor e as coordenadas paralelas.

Em sentido amplo, pondera o autor, todo o livro trata de mapeamento de dados, uma vez que lida com a representação espacial de informações. No capítulo 10, *Mapping Data*, contudo, restringe o uso do termo para abordar visualizações feitas sobre áreas geográficas, que têm entre seus principais atributos a escala, a projeção e os símbolos usados para retratar as informações. Após apresentar as questões e preocupações envolvidas com os diferentes processos de projeção, o modo pelo qual transforma-se o globo em uma imagem plana, Cairo apresenta as formas como os dados podem ser expostos sobre mapas – com pontos, linhas, área e volume. Linhas e volumes são tratados de modo passageiro, assim como os cartogramas, em que a área geográfica assume a magnitude de uma variável. Na representação mais simples, por pontos, atenta para os frequentes erros de proporção ao se lidar com sua escala e uma falha frequente relacionada à ilusão de Ebbinghaus, pela qual círculos de mesma área são percebidos de forma diferente de acordo com o que está em seu entorno. Dedicar-se com mais detalhe aos mapas de *choropleth*, que codificam informações pela atribuição de cores ou tons de cor a áreas definidas, podendo representar diferentes tipos de dado.

Instigado por uma manchete infeliz – “Opinião pública da Catalunha vai para o ‘não’ pela independência, diz pesquisa” – a respeito de uma candente questão em seu país, Cairo pede algum esforço do leitor para retornar ao conceito de distribuição e aprofundá-lo abordando as formas de cálculo do erro padrão e do intervalo de confiança, que é um modo de visualizar a incerteza nos dados de que se dispõe. Neste capítulo que encerra a terceira parte, o autor aponta a necessidade de se expor esta incerteza, tanto quanto factível, no próprio gráfico – e não apenas em uma nota de rodapé.

Um último conselho do autor a quem se aventura na construção de visualizações de dados – de modo amador ou profissional – é o de se aproximar de especialistas nos dados com os quais se está lidando, de modo a compreender como eles foram obtidos, processados e testados. Resultado direto de sua prática profissional e acadêmica, o livro entrega o que as credenciais do autor e o prefácio anunciam: um trabalho denso para a introdução e o aperfeiçoamento em técnicas e estratégias para que possamos encontrar, no escrutínio dos dados, as questões adequadas – e suas melhores respostas.

Sobre os autores

Sérgio Francisco Seabra Moreira

<sergiofco@gmail.com>

Mestrando no Programa de Pós-Graduação em Design (PPGDesign-UAM) Universidade Anhembi Morumbi. São Paulo/SP

Dra. Mirtes Cristina Marins de Oliveira

<mirtescmoliveira@gmail.com>,

Professora no Programa de Pós-Graduação em Design (PPGDesign-UAM) Universidade Anhembi Morumbi. São Paulo/SP