

Visualização de Dados Científicos

Caio Lente – 07/05/2019

Introdução

Exibir dados científicos da melhor forma possível garante¹ que ideias complexas sejam comunicadas com **clareza, precisão e eficiência**. Um bom gráfico deve tornar coerentes grandes conjuntos de dados, encorajar o leitor a comparar porções dos dados e revelar a estrutura latente dos dados, ao mesmo tempo em que não chama a atenção para si próprio.

Excelência Gráfica

A “apresentação bem-feita de dados interessantes” depende de **substância, estatística e design**. Excelência gráfica implica em bons conjuntos de dados, é quase sempre multivariada, traz ao leitor o maior número de ideias no menor tempo possível e, necessariamente, requer que seja transmitida a verdade sobre os dados.

As grandezas físicas medidas na superfície do gráfico devem ser proporcionais às quantidades numéricas. Rotulação cuidadosa e detalhada dos dados, uniformidade de design e utilização do número correto de dimensões para representar os dados são princípios que ajudam a manter a integridade de um gráfico. É importante também nunca usar dados fora do **contexto** (unidades monetárias padronizadas, exibir o antes e o depois quando possível, etc.).

Gráficos não precisam conter desenhos ou texturas desnecessárias. A audiência não é menos inteligente e, se os dados forem interessantes, um gráfico nunca será desinteressante. Se um conjunto de dados for pequeno, então a melhor solução é usar uma **tabela**.

Teoria Gráfica

No mundo onde cada arquivo pode ser compartilhado inúmeras vezes, um gráfico mal-feito ou mentiroso estará confundindo ou iludindo uma vez para cada cópia sua que for lida.

Alguns princípios devem guiar qualquer visualização de dados, ainda mais as científicas. Eles não são absolutos, mas antes de fazer cada novo gráfico vale a pena ter elas em mente.

- Acima de tudo, mostre os dados;
- Maximize (enquanto for aceitável) a razão entre dados e **tinta**;
- Descarte elementos que não representem dados²;
- Sempre revise e edite um gráfico múltiplas vezes;
- Abandone o *chartjunk* (grades, texturas, “patos”, etc.);
- Faça com que os gráficos sejam **auto-evidentes**;
- Tome cuidado com as paletas de cores (opte por P&B quando possível) e
- Prefira alta **densidade** de dados e, quando necessário, *small multiples*.

¹ Como demonstrado em Anscombe em 1973, medidas resumo são insuficientes para caracterizar um conjunto de dados.

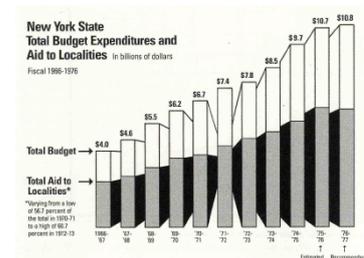


Figure 1: Antes

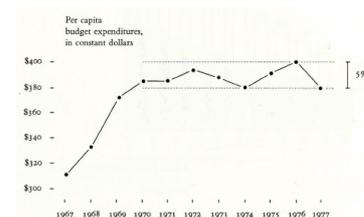


Figure 2: Depois

² “Todos falavam sobre uma sobrecarga de informação, mas na verdade havia uma sobrecarga de não-informação” – Richard Wurman.

Princípios Gráficos

Tão importante quanto o modo de criar um gráfico é que tipo de informação deveria estar em destaque. Mais relevante do que apenas mostrar dados, um gráfico precisa mostrar **comparações**, contrastes e diferenças, causas³, mecanismos, estrutura e **explicações**.

Um gráfico é como um parágrafo de números e deve, portanto, estar completamente integrado com palavras e documentação. Ele é acima de tudo um instrumento de argumentação, uma *beautiful evidence*. Falar a verdade, documentar as fontes e argumentar com honestidade devem ser os princípios que guiam o planejamento de uma visualização.

Apresentações Científicas

Nem sempre dados são visualizados através de gráficos. Grande parte da divulgação científica moderna se faz através de *workshops* e congressos, onde o principal meio de visualização é o PowerPoint (ou produtos equivalentes).

Por causa de seu formato extremamente restritivo, apresentações PP carregam consigo um **estilo cognitivo**. Uma linha de raciocínio linear, fragmentação de evidência, argumentação rasa e apresentações que soam mais como *pitches* do que como divulgação científica são consequências da filosofia PP. Slides PP são orientados a apresentador, não orientados a conteúdo ou plateia.

Em uma pesquisa, Tufte relata que livros médicos possuem aproximadamente 26 caracteres por cm² e que sites de notícia chegam a 6,8 caracteres por cm². Já PPs não passam de 0,5 caracteres por cm²! Em virtude desta aridez de conteúdo, Tufte afirma que o melhor software de apresentações é o processador de texto.

Além de rasas, apresentações PP incentivam conclusões sem nuances porque cada slide precisa de um título conciso. Tanto slides profundos e bem embasados quanto slides de pura propaganda acabam tendo o mesmo visual e quantidade de material. Isso é nocivo para a Ciência porque a coloca no mesmo nível que *pitches* de marketing⁴.

Conclusão

As visualizações de dados científico são um dos pontos mais importantes do processo de se fazer Ciência. Sem elas, poucas pesquisas poderiam ser bem divulgadas ou adequadamente interpretadas e, portanto, devem ser elaboradas com atenção e precisão.

As principais regras para gráficos tem por objetivo torná-los visualmente mais simples e conceitualmente mais complexos. Contexto, **evidência**, densidade e análise multivariada são os princípios-guia desta ciência e arte.

Para apresentações, o mais importante é tentar fugir da imposição cognitiva das ferramentas utilizadas. É melhor distribuir um panfleto no início da palestra para que a plateia possa ter um **resumo** que use frases e parágrafos.

A regra de ouro é buscar a **verdade** e colocá-la à frente do resto.

³ Em *Beautiful Evidence*, Tufte conta de um livro que pretendia encontrar uma relação entre o preço de uma obra e idade do pintor. A única evidência fornecida são duas curvas ajustadas sem os pontos subjacentes.

	Use Tables	For Your Numbers	But Not too Many
This row	10	90	100
This row	0.6	0.4	1
This row	1	2	3
That row	1	2	3

Figure 3: Tabela PP

⁴ “Para uma tecnologia bem-sucedida, a realidade deve ter precedência sobre as relações públicas, pois a Natureza não pode ser enganada” – Richard Feynman