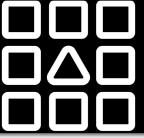


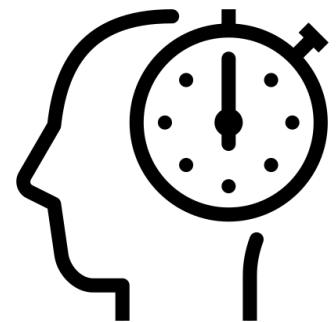


VISUALIZAÇÃO DE DADOS

Data Science & Business Analytics | [VD]



AVALIAÇÃO DO MÓDULO



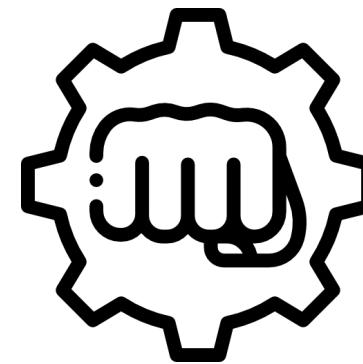
Assiduidade e
Pontualidade
15%



Motivação e
Participação
20%



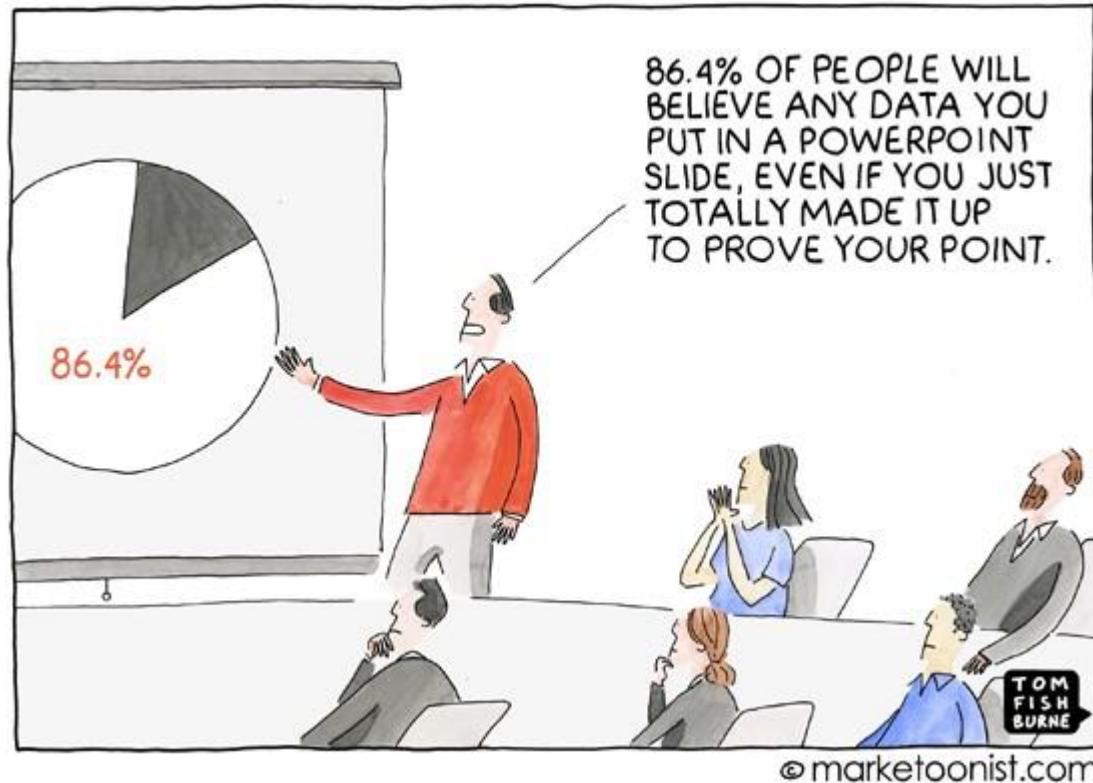
Domínio dos
Conteúdos das Aulas
15%



Aplicação
Técnica
20%

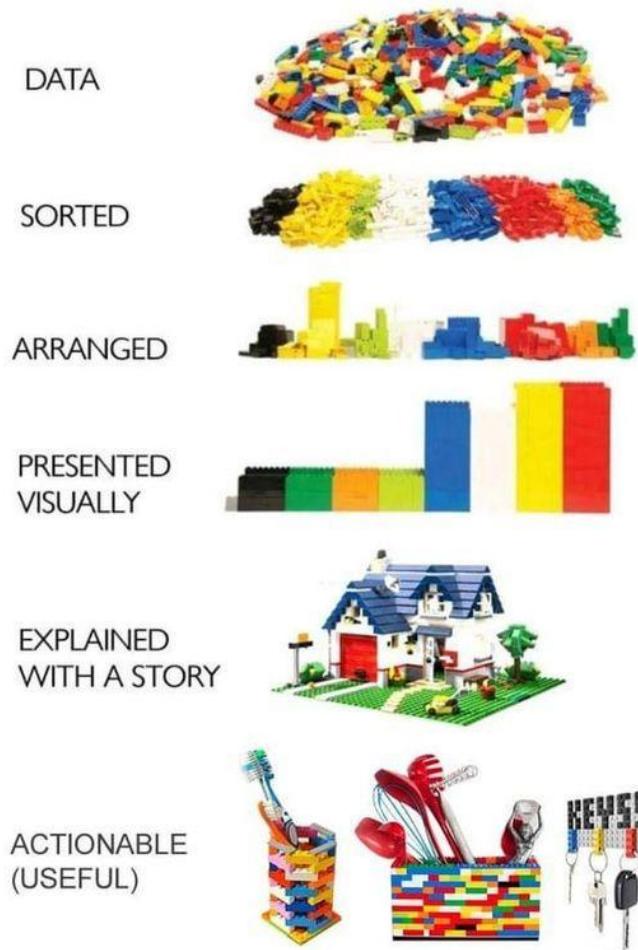


Trabalho Final
30%



“Visualization gives you answers to questions you didn't know you had.”

Ben Schneiderman



Mais que Visualizar,
queremos comunicar.

UM POUCO DE HISTÓRIA

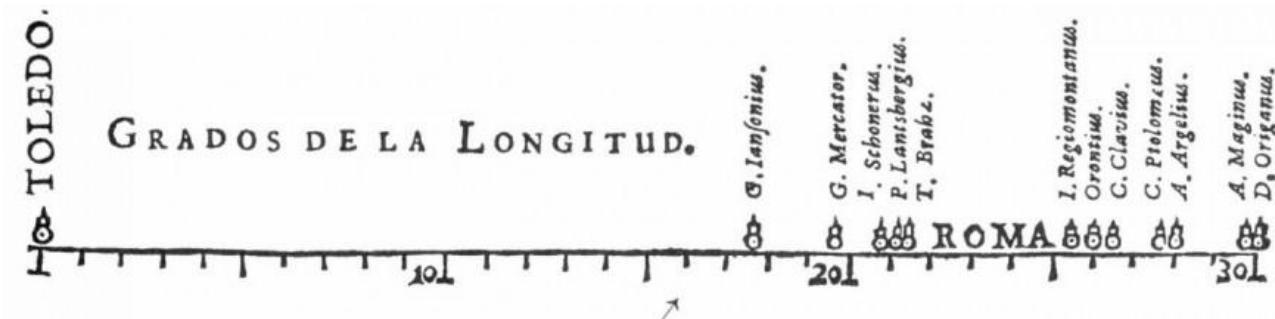
Alguns nomes importantes na visualização
(não é, de todo, uma lista exaustiva)

**MICHAEL FLORENT VAN
LANGREN**

(1598–1675)

Astrónomo e cartógrafo da região da Flandres (Países Baixos/Bélgica).

Uma das suas maiores contribuições foi para o cálculo da longitude. Criou um dos primeiros gráficos estatísticos (o primeiro, segundo a [Wikipédia](#)) mostrando a relação entre as estimativas de distância em longitude entre Toledo e Roma.

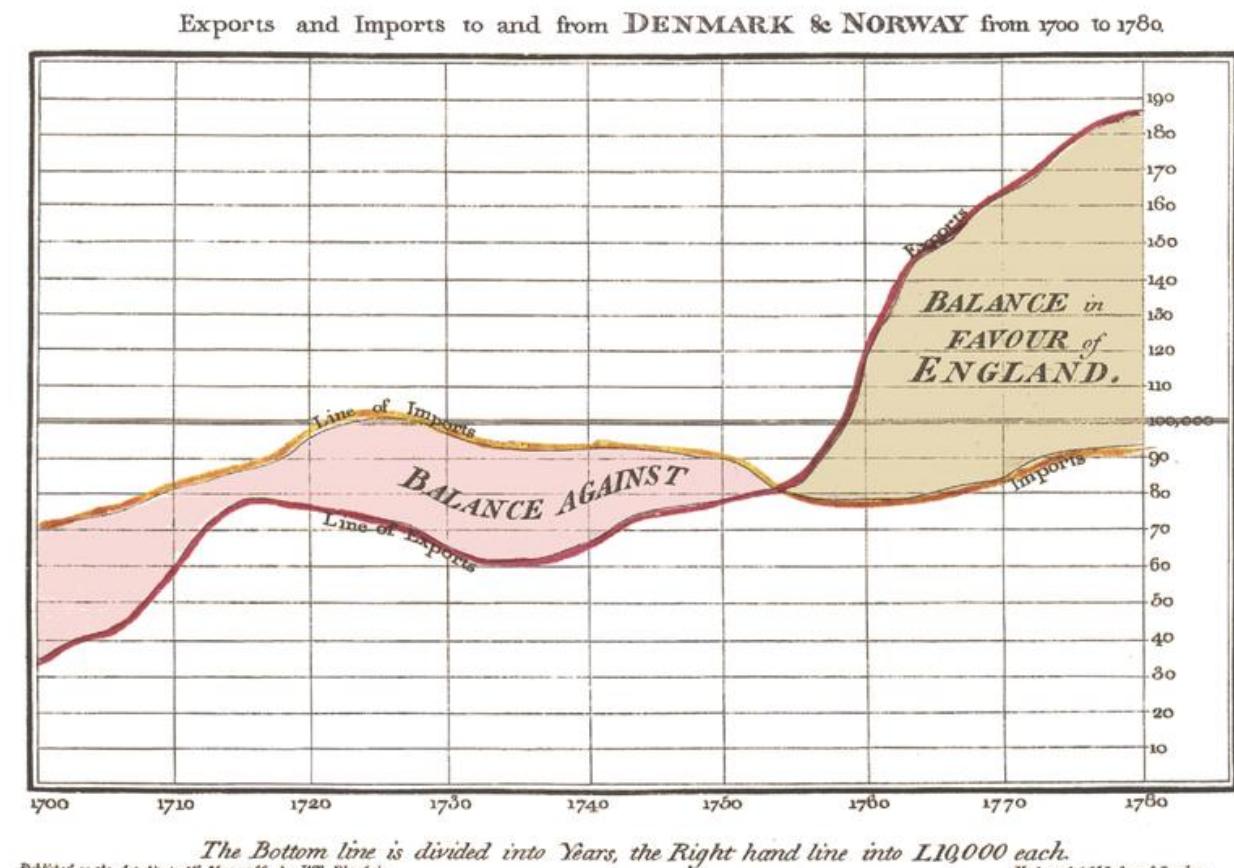


WILLIAM PLAYFAIR

(1759–1823)

Engenheiro e economista político Escocês, publicou “The Commercial and Political Atlas” em 1786.

Foi um dos primeiros a utilizar gráficos de linhas e de barras como os conhecemos hoje.

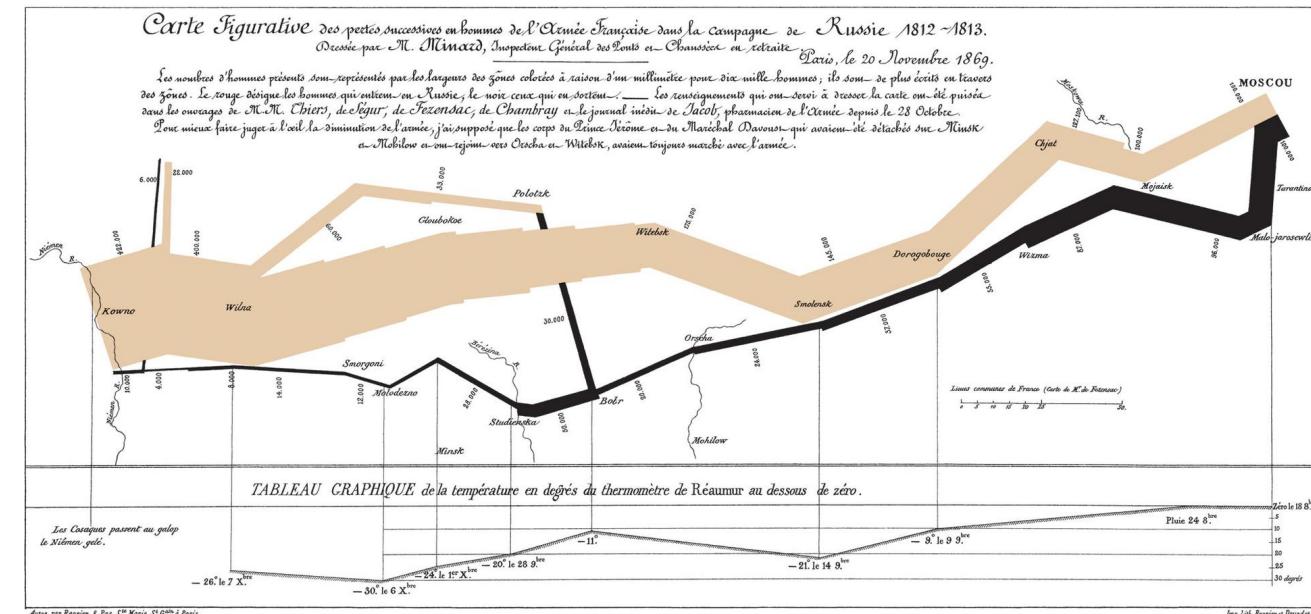


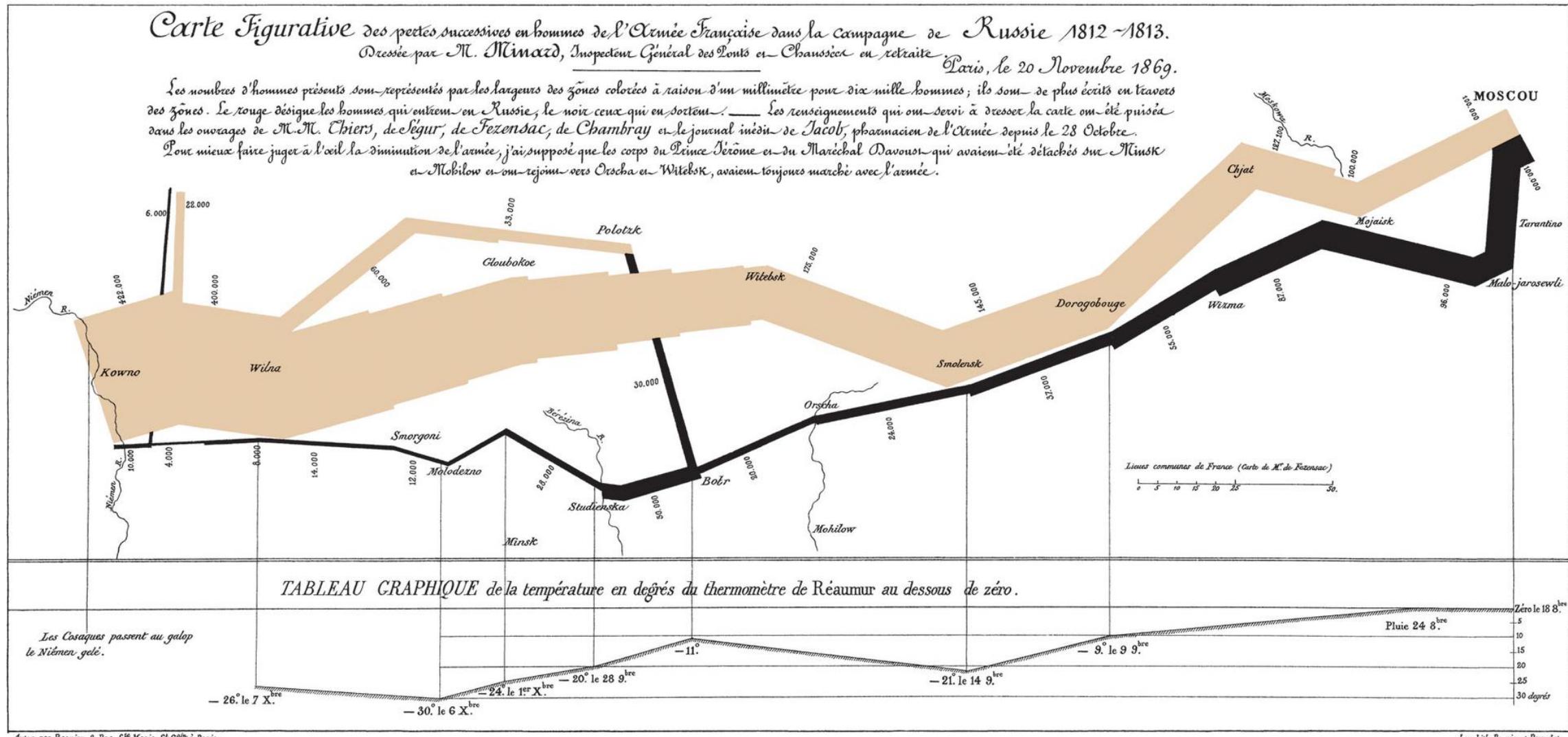
CHARLES JOSEPH MINARD

(1781–1870)

Engenheiro civil francês, ficou famoso pela representação de dados numéricos em mapas.

O seu trabalho mais famoso é a imagem ao lado, da campanha russa das tropas de Napoleão em 1812. É um trabalho considerado por muitos o melhor gráfico estatístico alguma vez desenhado. Representa o início o ‘jornalismo de dados’.





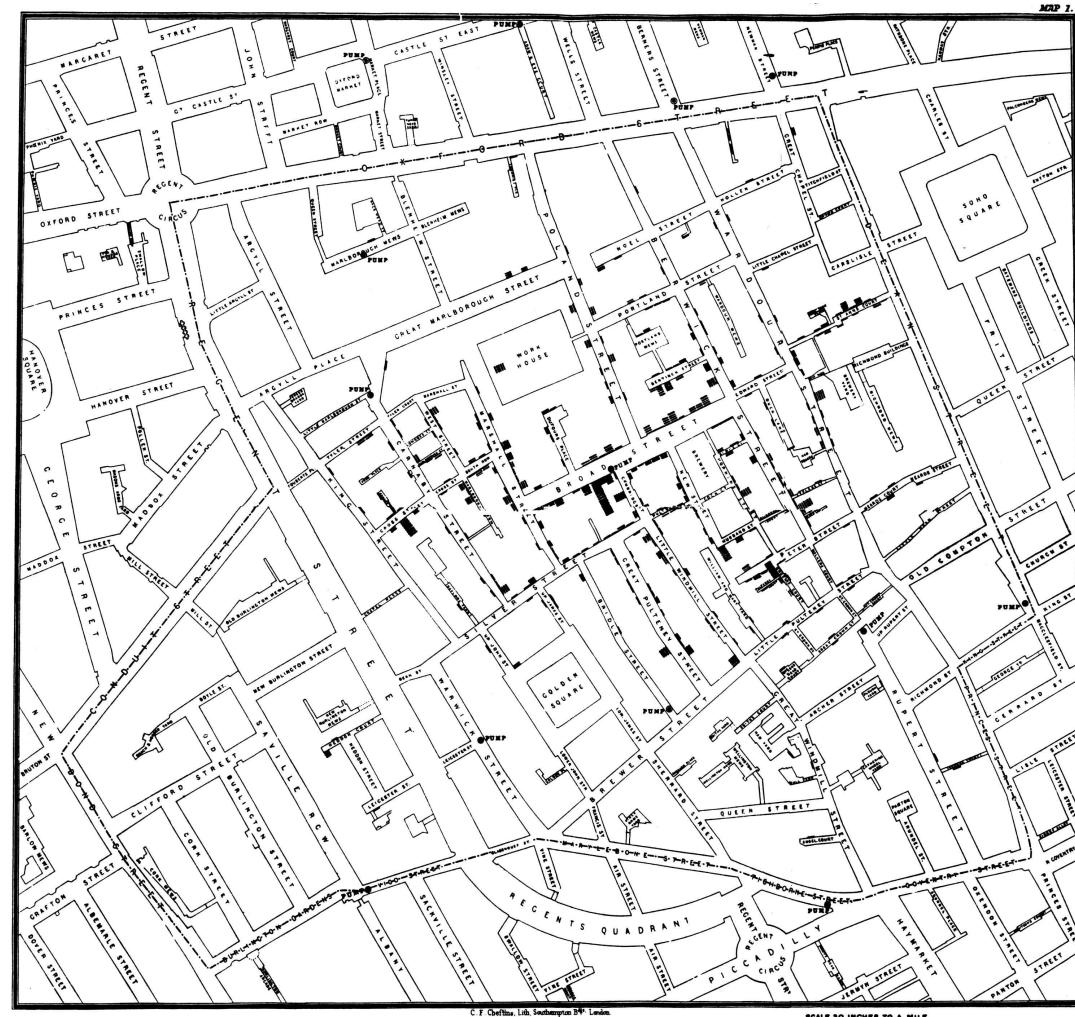
By Charles Minard (1781-1870) - see upload log, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=297925>

JOHN SNOW

(1813–1858)

Médico Inglês, ‘pai’ da epidemiologia moderna.

Em 1854, utilizou o mapa anexo para identificar a fonte de contaminação num caso de cólera na cidade de Londres.

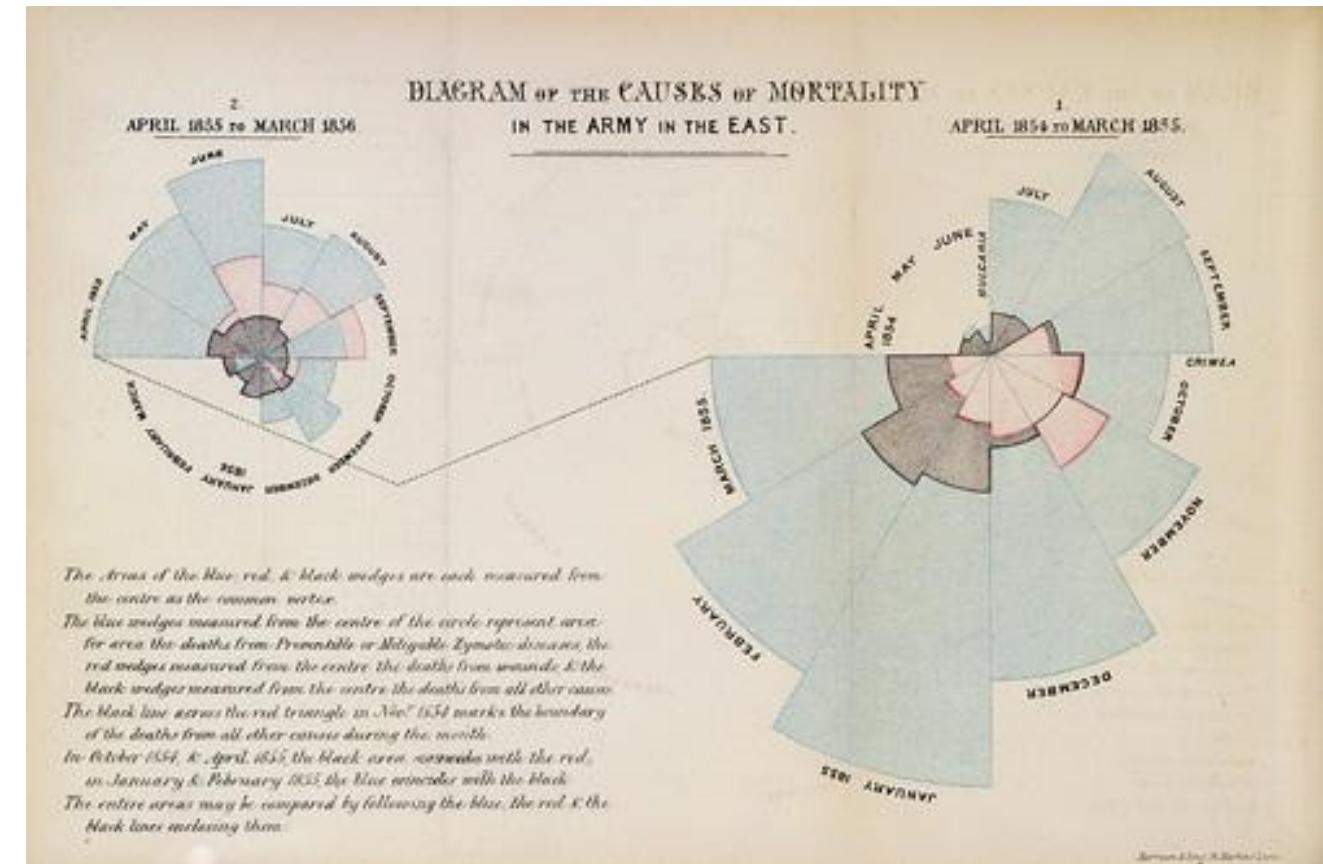


FLORENCE NIGHTINGALE

(1820–1910)

Enfermeira, estatística, escritora. Ficou famosa por ser pioneira no tratamento dos feridos de guerra durante a guerra da Crimeia.

Pioneira no uso de infografias, um dos seus gráficos mais conhecidos é o gráfico sectorial (Rose chart ou coxcomb) das causas de mortalidade no exército na referida guerra.

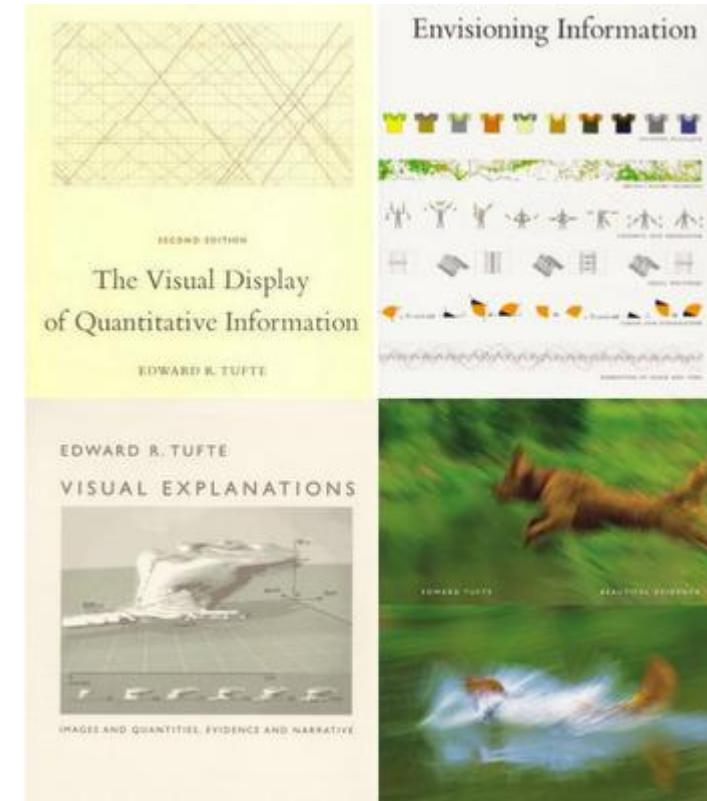


EDWARD TUFTE

(1942-)

Professor de estatística, design gráfico e economia política.

Autor de vários livros sobre visualização de dados, é um nome de referência na visualização moderna.



<https://www.edwardtufte.com/tufte/>

APRESENTAR INFORMAÇÃO

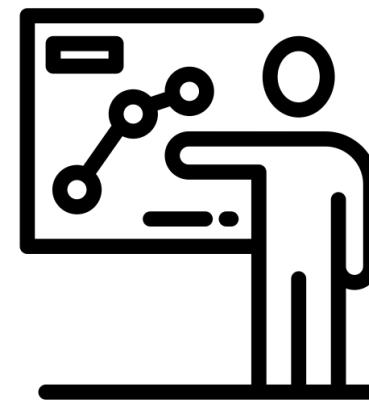
Diferentes meios usados para comunicar informação



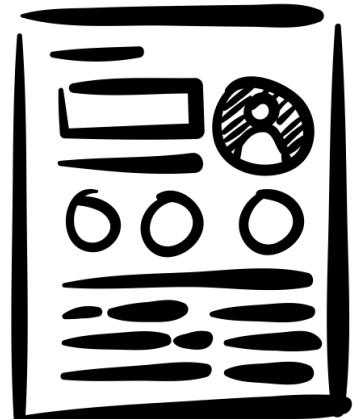
Dashboards



Infografias



Apresentações



Outros
Documentos
Folhetos
Artigos
Posters
...



EXECUTIVO / ESTRATÉGICO

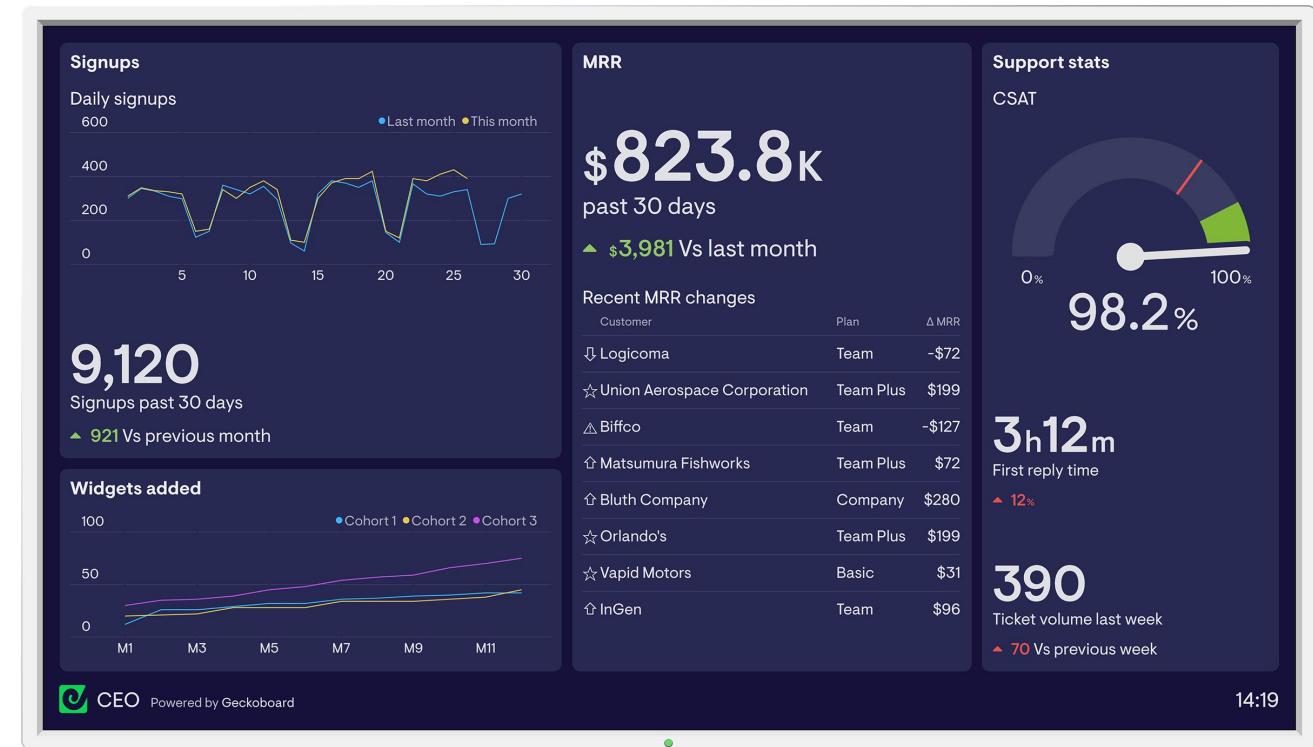
Público - Alvo: CEO/Executivos de topo

Visão geral rápida

Complexo – métricas simples devem refletir a performance geral

Desenhado para decisões estratégicas a longo prazo

Focado em métricas agregadas a um nível macro (KPI), muitas vezes comparadas com objectivos estratégicos





TÁCTICO

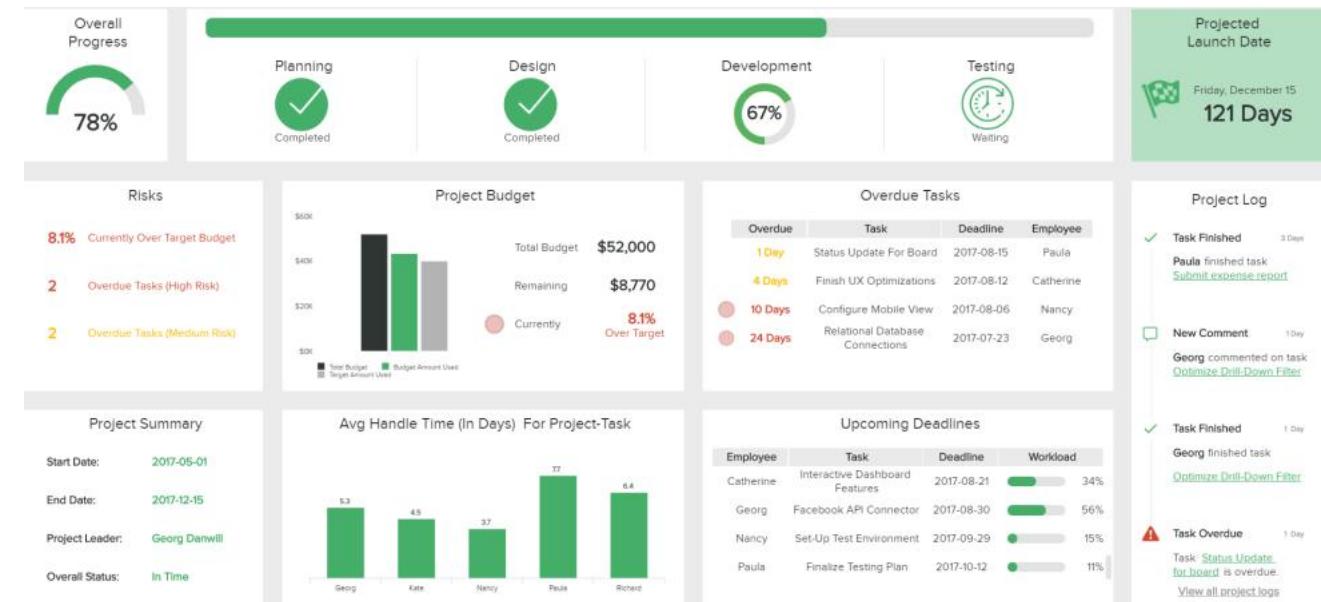
Público - Alvo: Gestores | Analistas

Monitorizar os processos que apoiam as iniciativas estratégicas da empresas (dados operacionais, estratégicos entre outros)

Identificar tendências e ajudar e ajudar a tomar decisões inteligentes

Pode conter dados evolutivos que permitem uma análise de planeamento longitudinal

O âmbito de dados situa-se habitualmente entre o Operacional e o Executivo





ANALÍTICO

Público - Alvo: Analistas

Por vezes equiparado ao táctico, contém normalmente bastante informação

Criado por analistas para uso próprio, na sua função de apoio aos executivos

Orientado para conhecimento/monitorização aprofundado/a do negócio ou de um problema em particular





OPERACIONAL

Público - Alvo: Supervisores | Especialistas

Métricas sobre operações específicas das empresas

Simples e detalhado – métricas directas de negócio

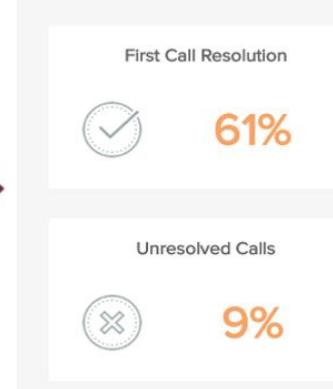
Pretendem ser um ‘snapshot’ do negócio

Indicadores que permitem a identificação de problemas e falhas nos processos

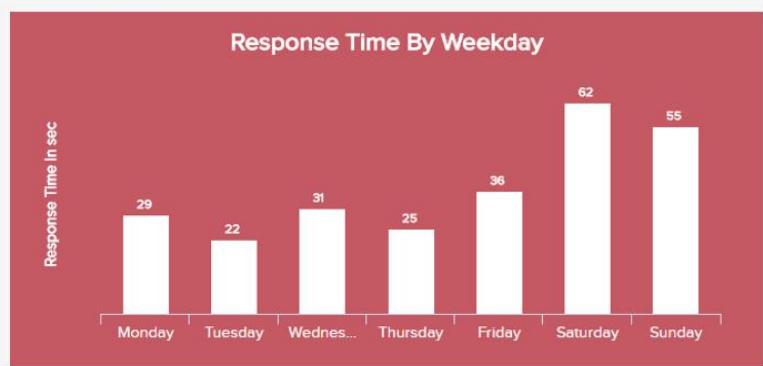
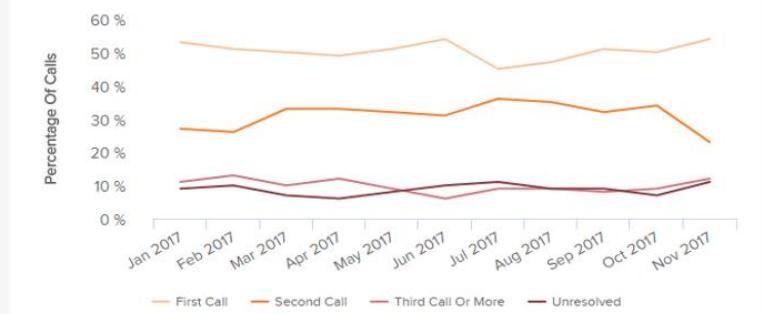
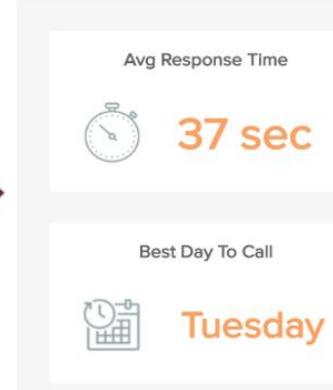
Informação relativas a decisões de curto prazo e que sejam executadas de acordo com o planeamento previsto

Customer Service Team Dashboard

R E S O L U T I O N S



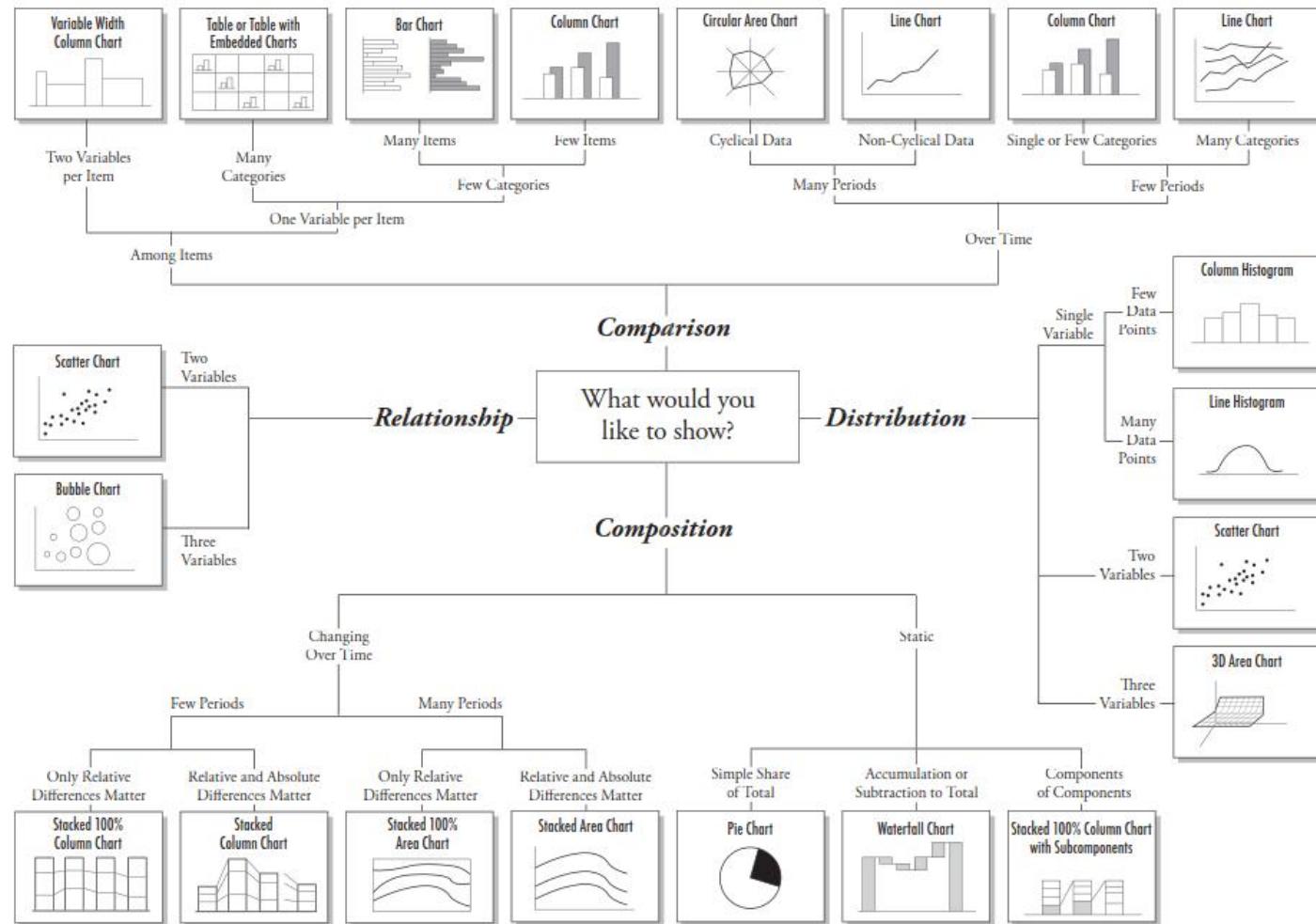
R E S P O N S E T I M E



ESCOLHA DO GRÁFICO

Como escolher o gráfico, tendo em conta a informação e propósito

Chart Suggestions—A Thought-Starter



A selecção ao lado pode ser vista como “clássica” e não incorpora alguns tipos mais recentes de visualizações existentes.

O capítulo 5 e seguintes do livro **Fundamentals of Data Visualization – de Claus O. Wilke** sugerem também alguns tipos de gráfico para cada tipo de dados. Essas não é, também, uma lista exaustiva.

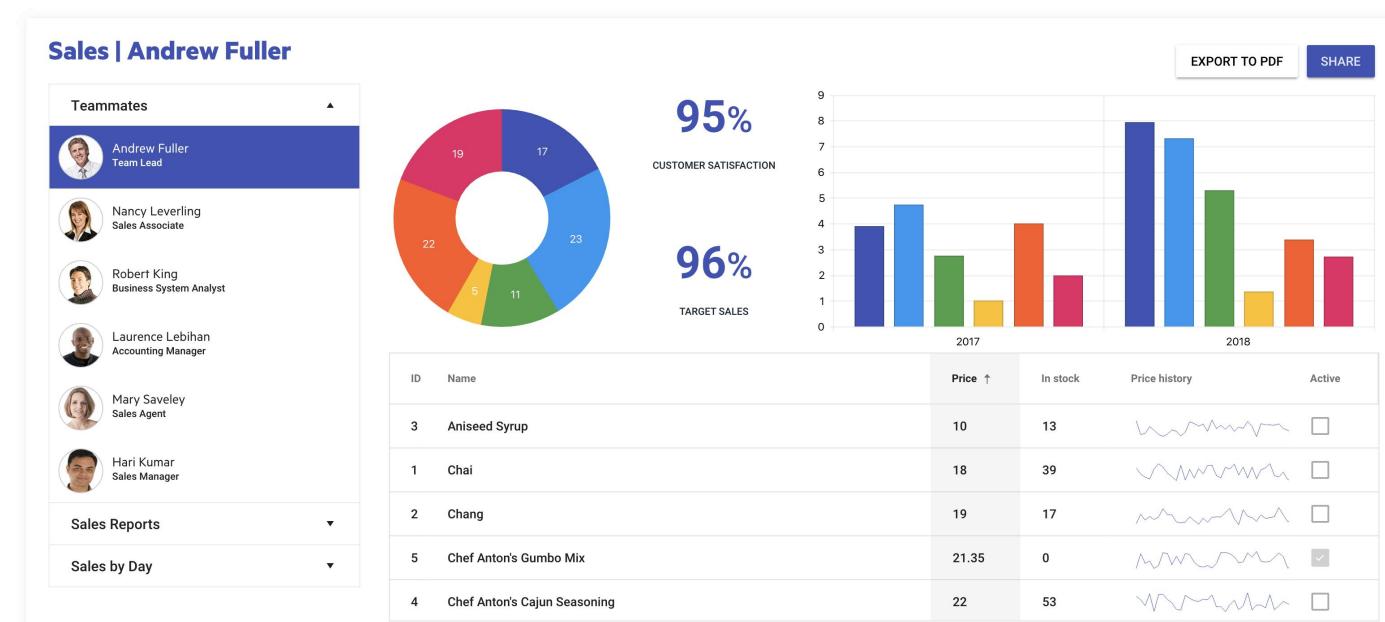
(preview: <https://clauswilke.com/dataviz>)

Mais opções:
<https://www.data-to-viz.com>

NÚMEROS



TABELAS



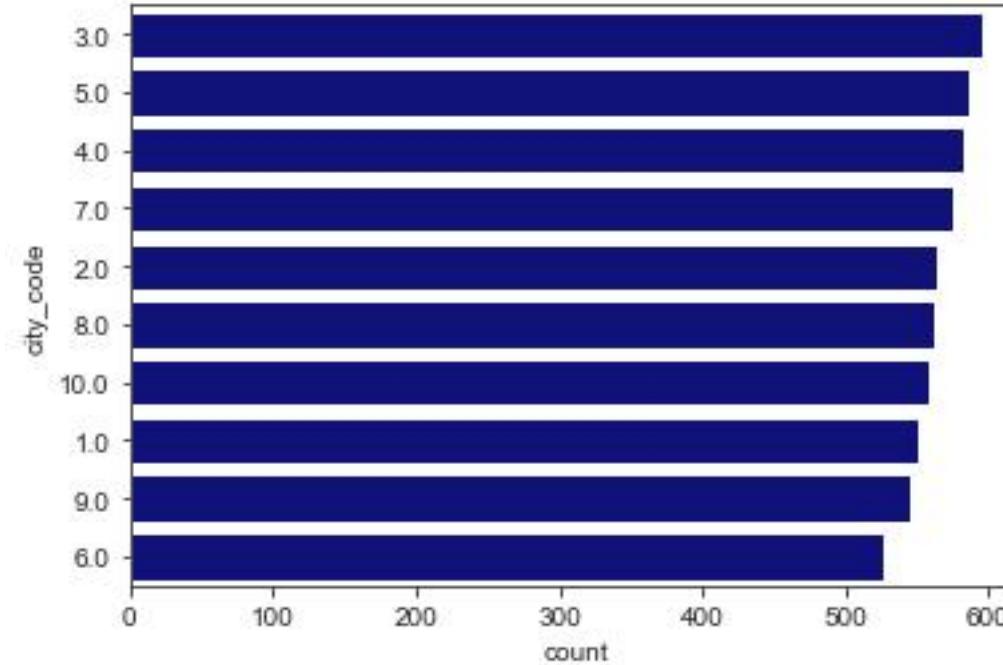
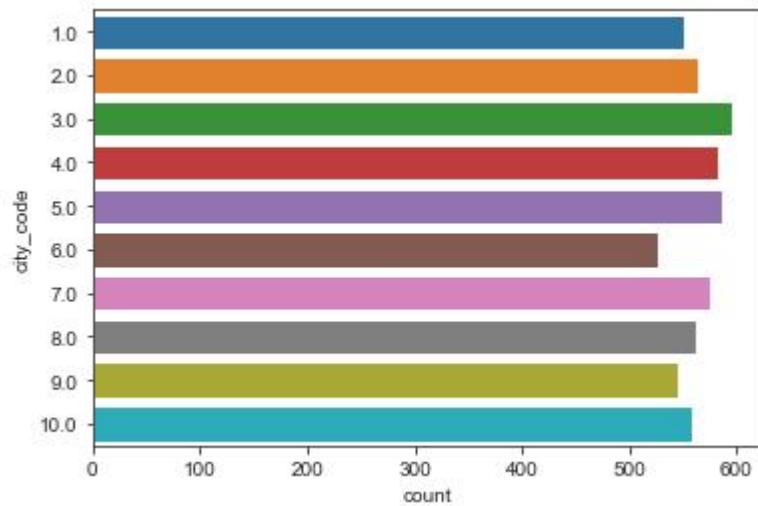
PRINCÍPIOS DE PERCEPÇÃO VISUAL

Aplicando alguns princípios da *Gestalt* à visualização de dados

PREGNÂNCIA

(PRÄGNANZ - SIMPLICITY)

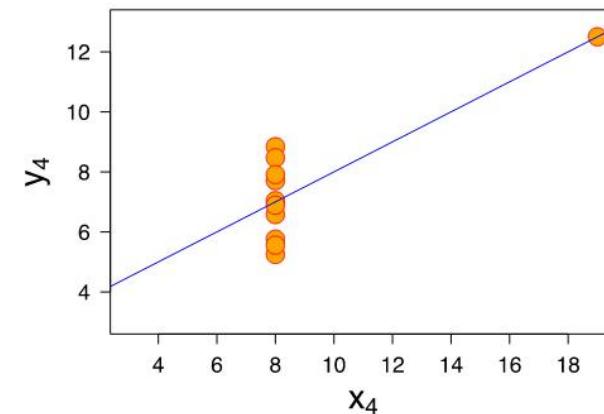
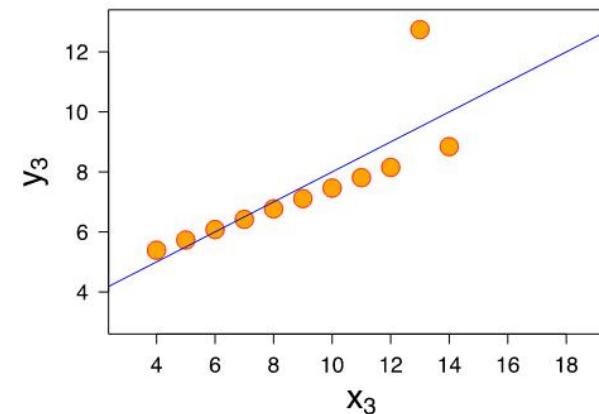
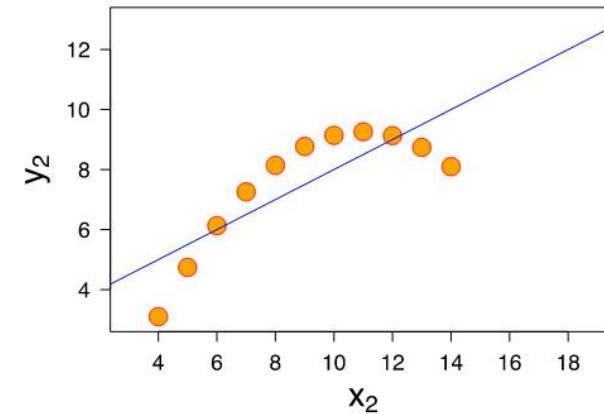
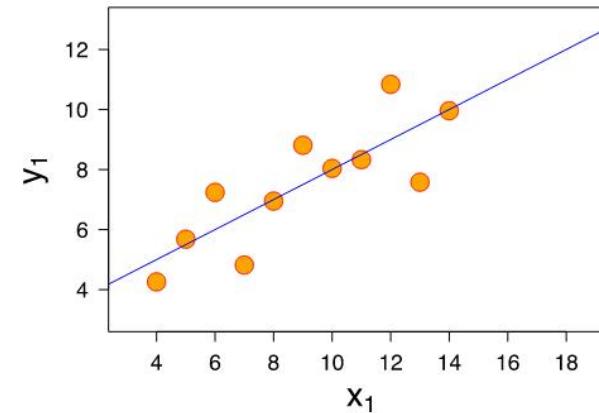
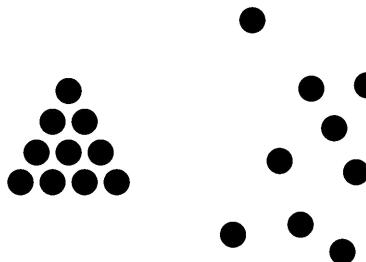
O cérebro prefere simplicidade. Se elementos adicionais (como a cor ou [des]ordenação), é preferível não incluir.



PROXIMIDADE (PROXIMITY)

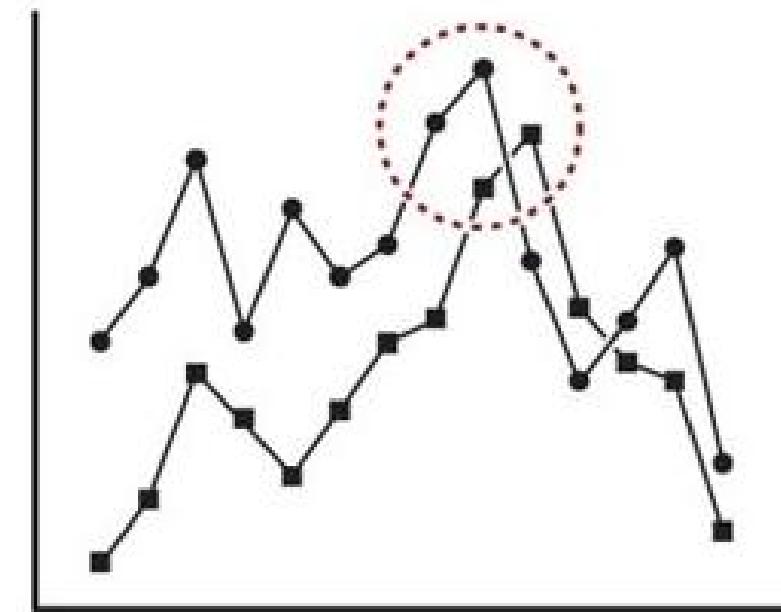
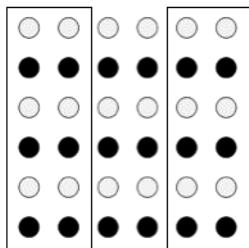
Objectos mais próximos são percebidos como pertencentes ao mesmo grupo

Objectos afastados/isolados são percebidos como diferentes dos demais



“PERTENÇA” (ENCLOSURE)

Agrupamos objectos contidos num mesmo ‘espaço’ (por contorno ou sombreado, por exemplo)

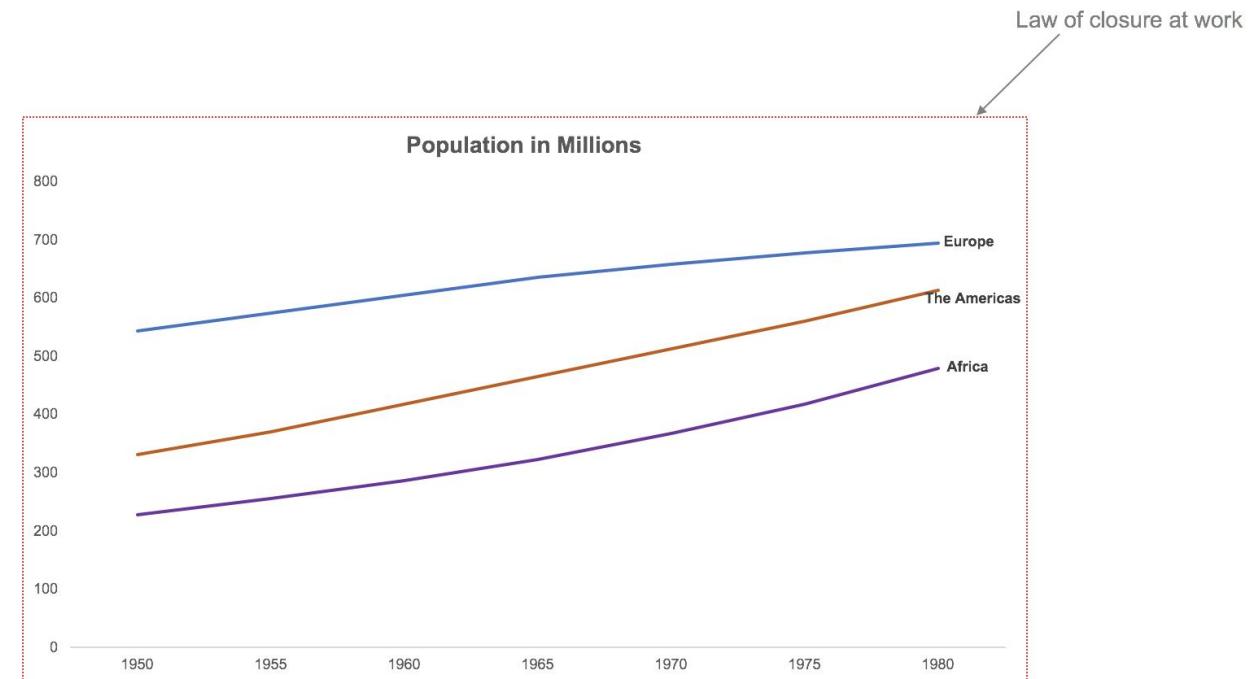
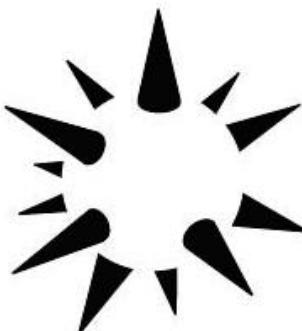


FECHAMENTO

(CLOSURE)

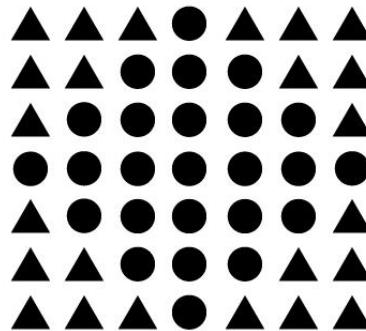
Vemos objectos como completos mesmo quando não o são.

Particularmente útil para retirar elementos para destacar o que é mais importante.

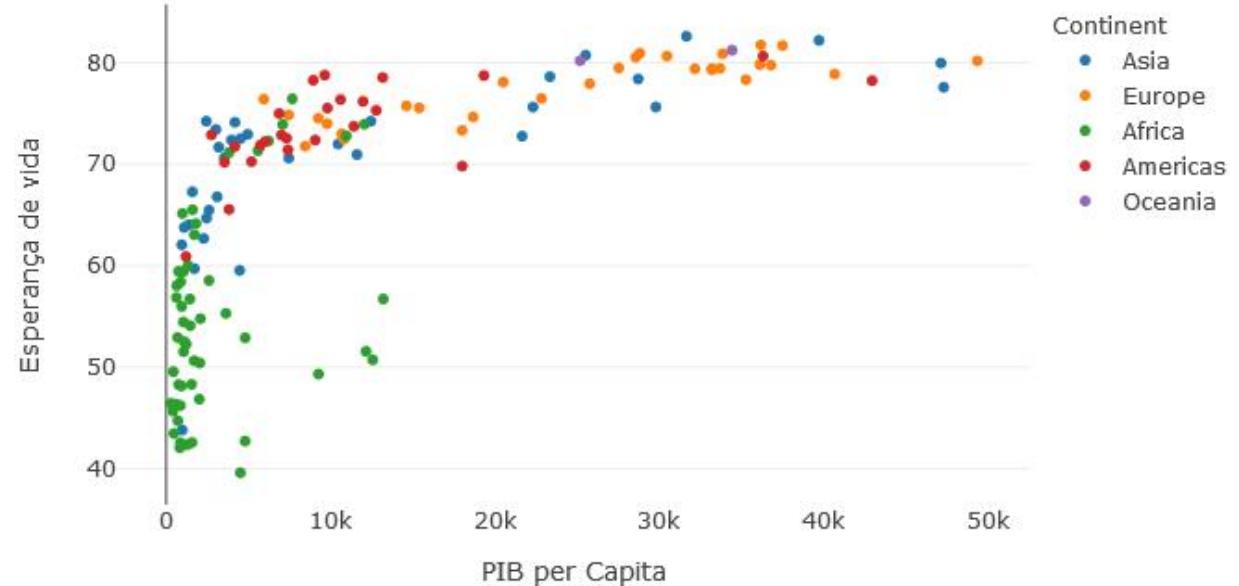


SEMELHANÇA (SIMILARITY)

Objectos de cor ou forma semelhante são percebidos como pertencendo ao mesmo grupo

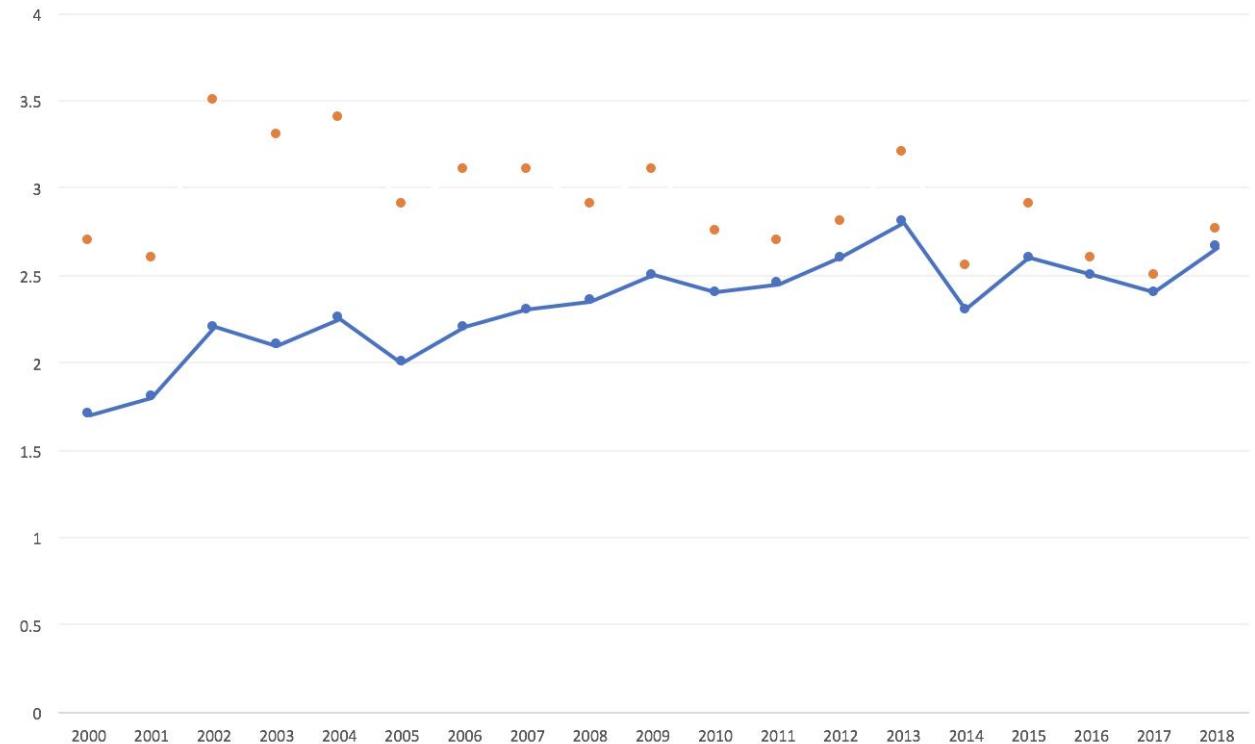
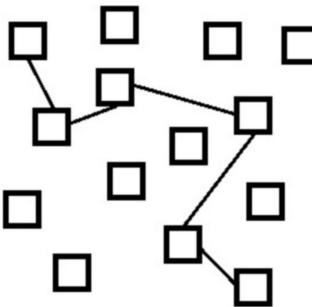


PIB per capita vs Esperança de vida



CONECTIVIDADE (CONNECTEDNESS)

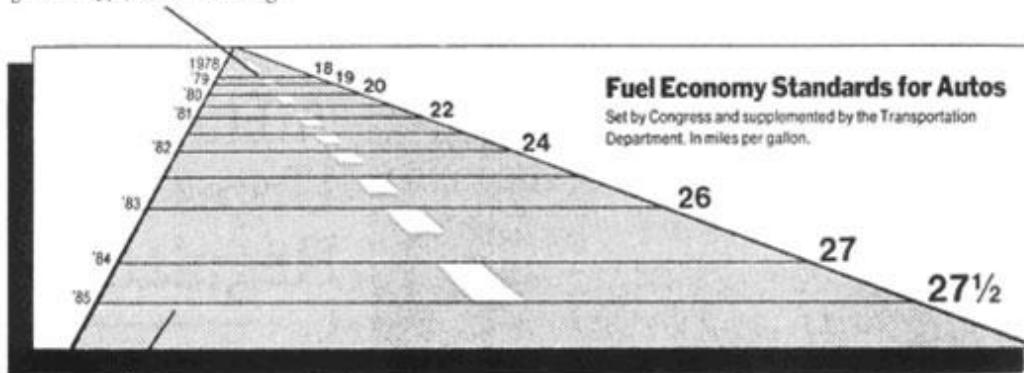
Objectos conectados fisicamente tendem a ser vistos como sendo do mesmo grupo



THE LIE FACTOR

$$\text{LIE FACTOR} = \frac{\text{SIZE OF EFFECT SHOWN IN THE GRAPHIC}}{\text{SIZE OF EFFECT IN DATA}}$$

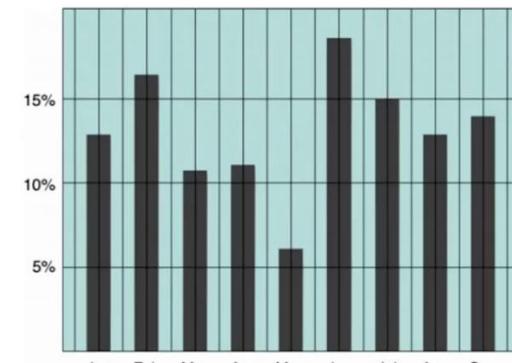
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



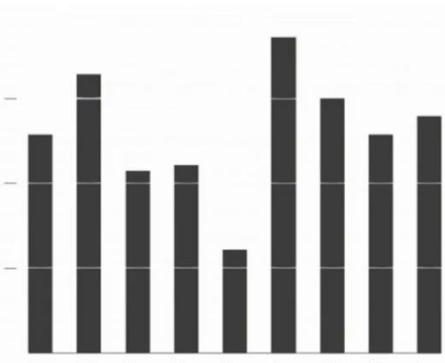
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

DATA INK RATIO

- Data-ink ratio = $\frac{\text{data-ink}}{\text{total ink used to print the graphic}}$
- = proportion of a graphic's ink devoted to the non-redundant display of data-information
 - = $1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$

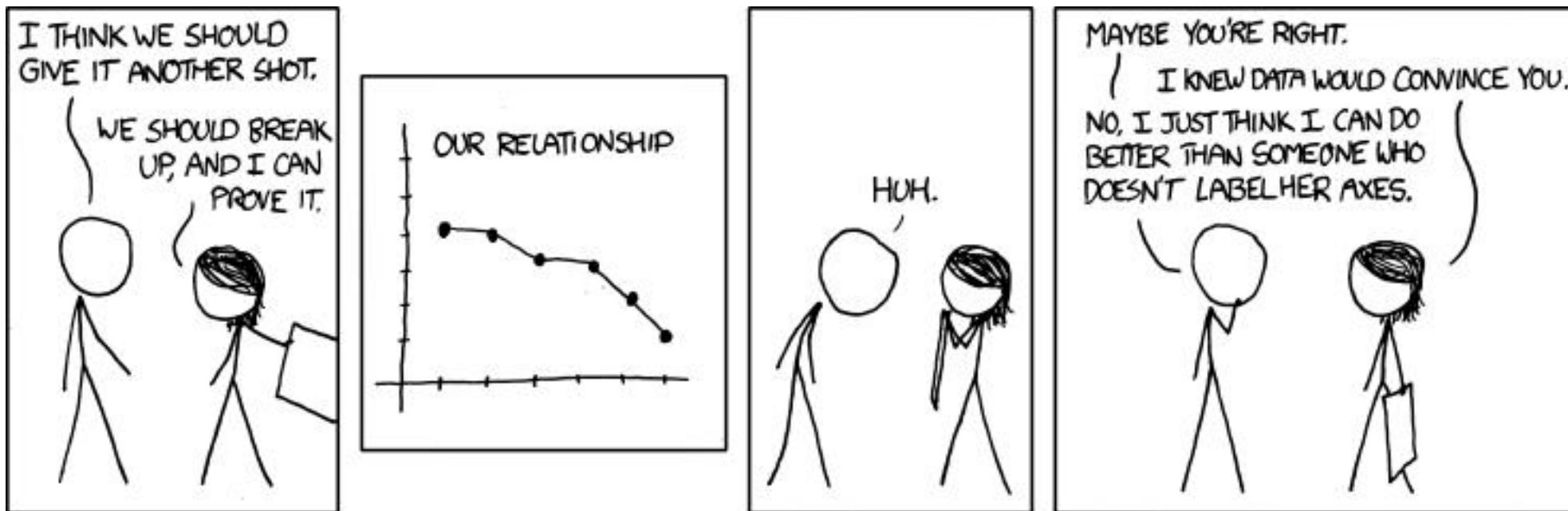


low data-ink ratio



high data-ink ratio

Não directamente relacionados com a *Gestalt*, mas estes conceitos introduzidos por Tufte ajudam à percepção visual



"And if you labeled your axes, I could tell you exactly how MUCH better."

Convincing from xkcd - : <https://xkcd.com/833/>

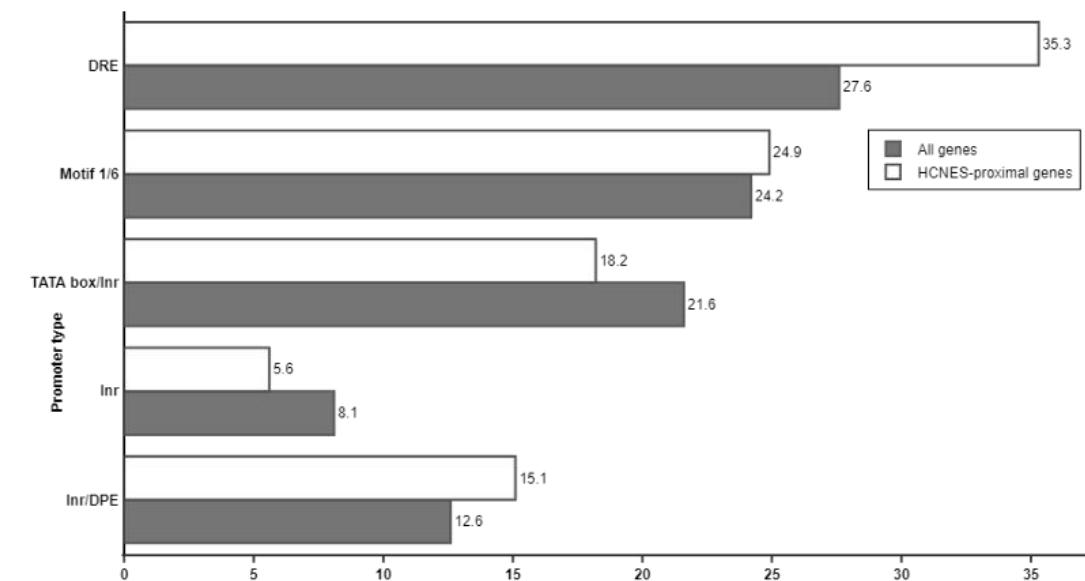
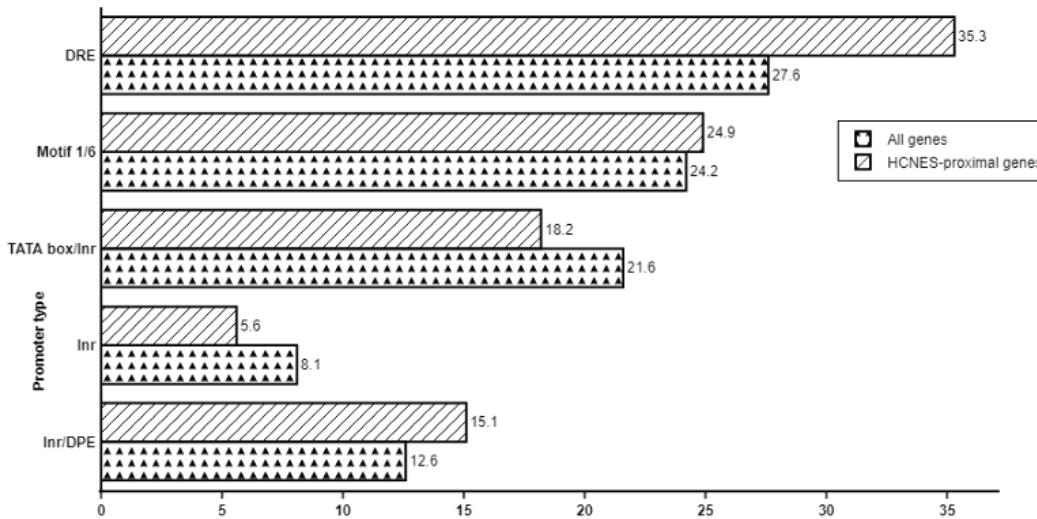
Erros a evitar ao construir uma visualização

- Tipo e tamanho da letra pouco legíveis
 - Texto com letra trabalhada*
- Gráficos inadequados – má escolha do gráfico para os dados
- Gráficos sem contexto
 - Falta de título
 - Sem descrição nos eixos
 - Sem *labels*
- Muitas cores diferentes sem propósito – dificulta a leitura
- Demasiados tipos de visualização – dispersa a atenção
- Tamanho muito pequeno
- Pouco ou demasiado espaço em branco
- 3D: leitura e interpretabilidade mais difícil
- Dados em bruto (excepto quando o valor é o mais importante)
- Falta de consistência
- Nenhuns ou demasiados items destacados

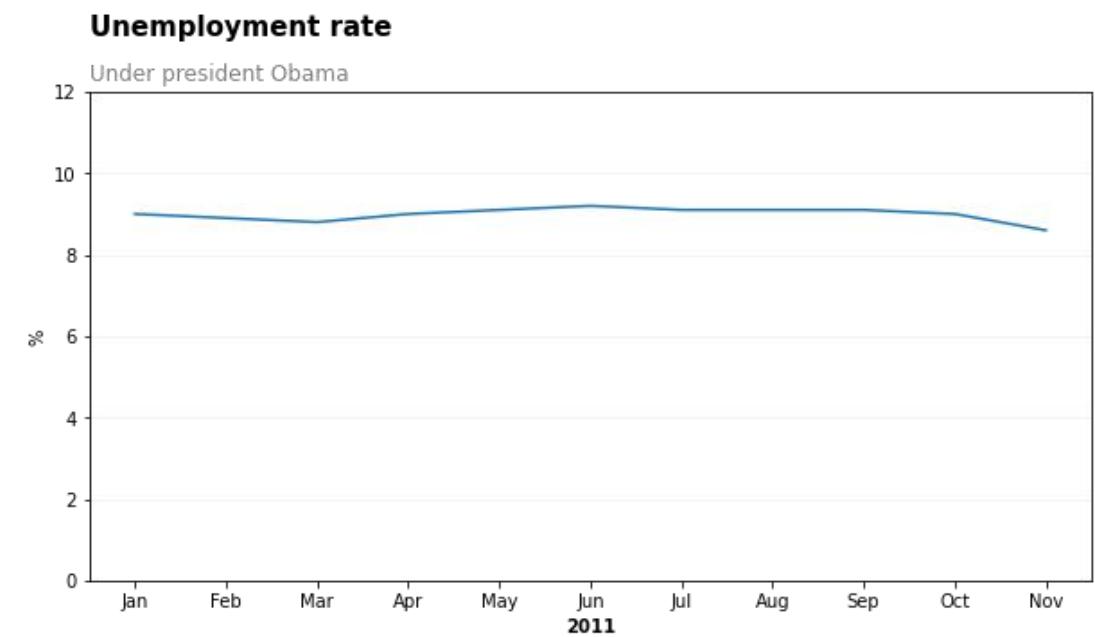
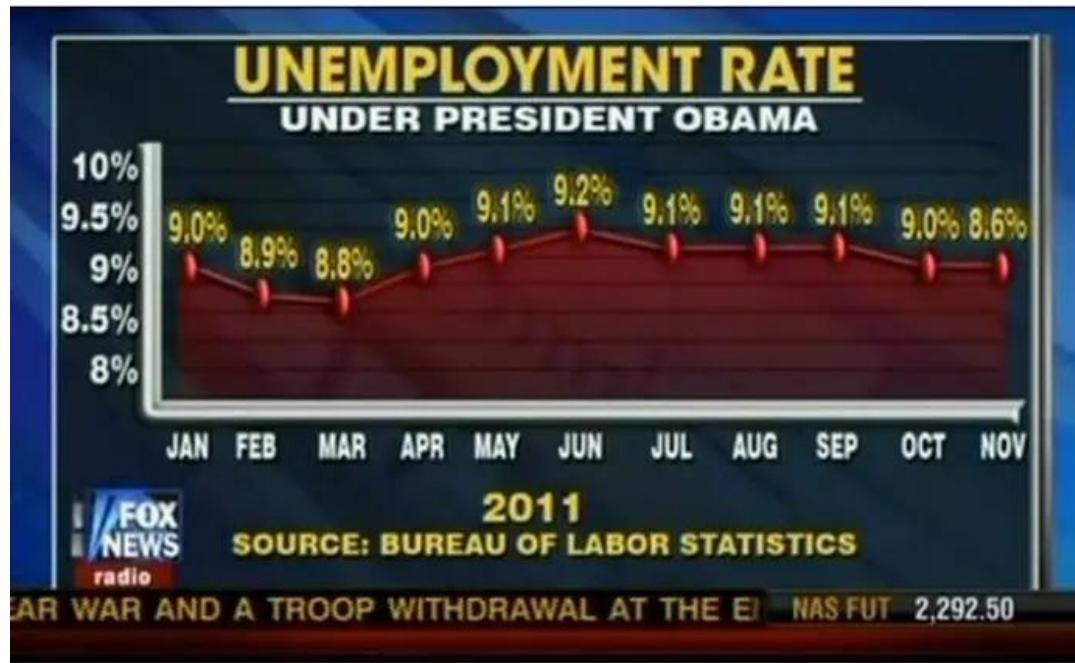
ANÁLISE CRÍTICA

Bons ou maus exemplos (ou algo no meio)

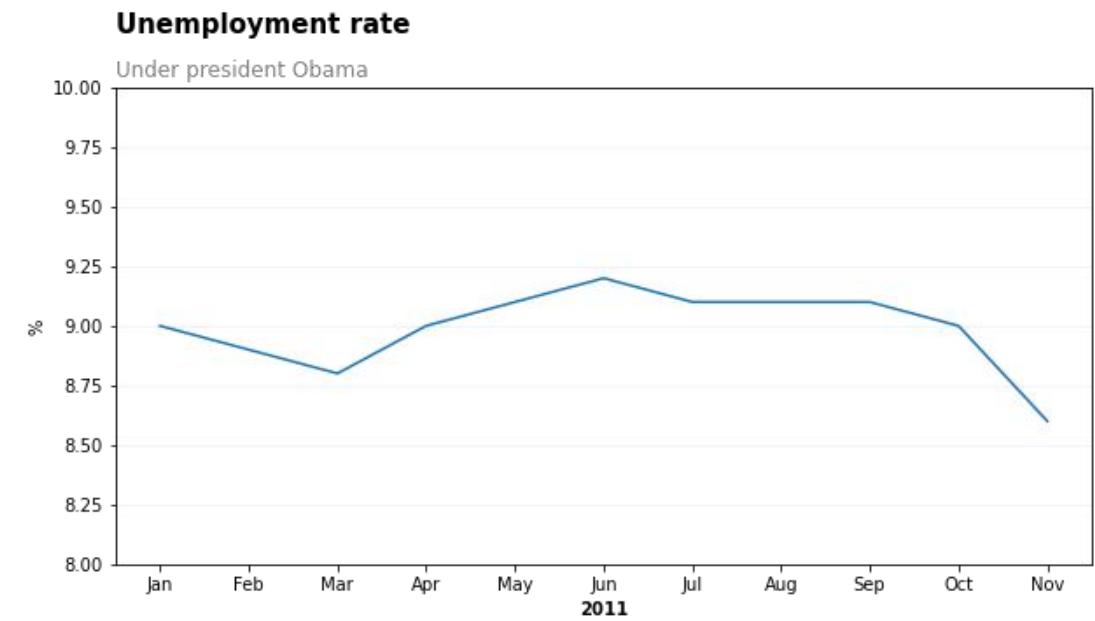
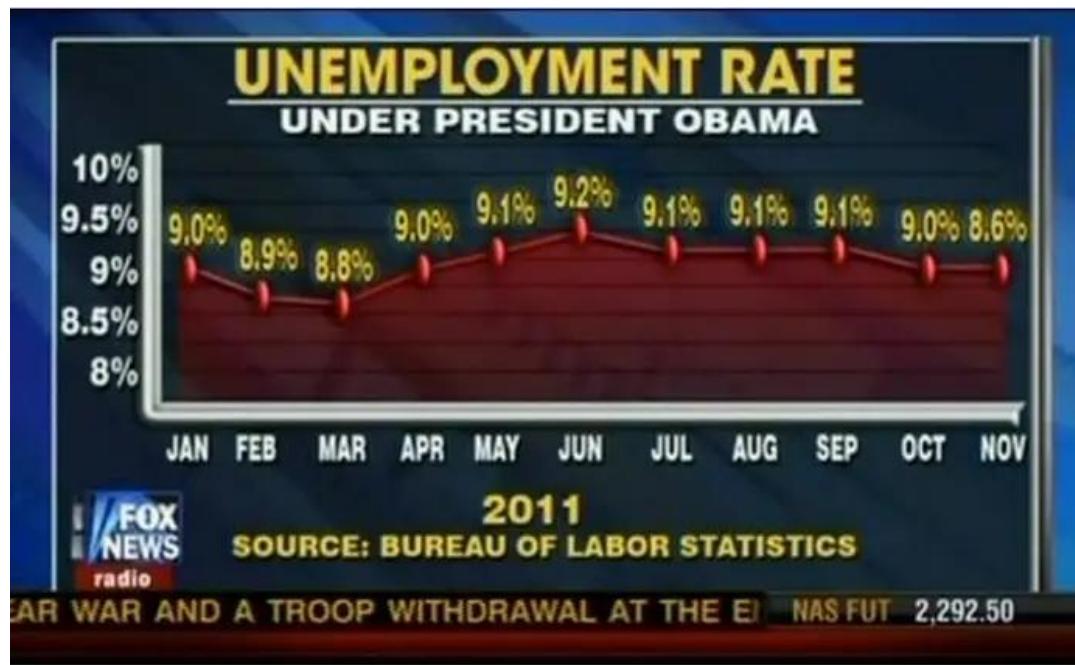
O QUE MUDAVAM?



O QUE MUDAVAM?



UTILIZANDO A MESMA ESCALA NO EIXO DOS YY

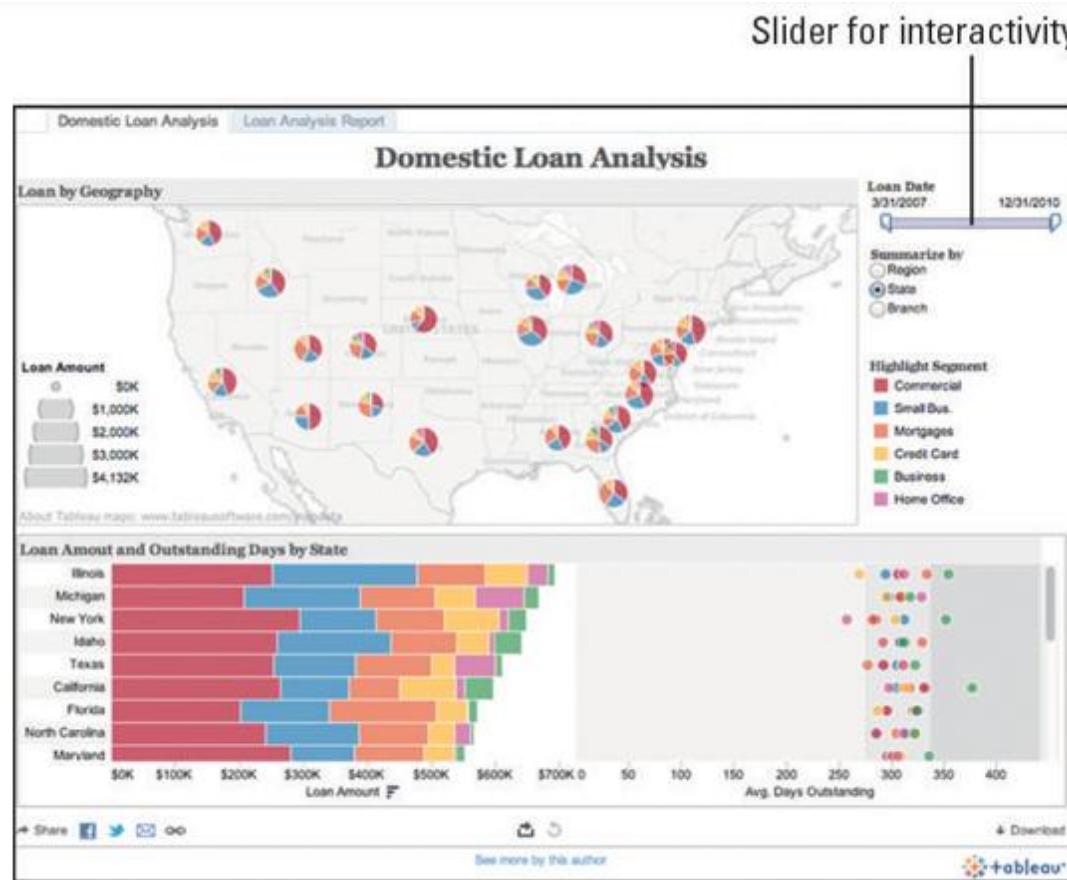


O QUE MUDAVAM?



- Pouco espaço para ‘respirar’
- Cores (aparentemente) inconsistentes
- Demasiada informação
- Propósito parece indefinido

O QUE MUDAVAM?



Funciona

- Consistência nas cores
- Simplicidade
- Interactividade

Não funciona

- Tipo de gráficos
- Demasiada informação no ecrã, sem destacar pontos relevantes para o utilizador

WHEN PLATES COLLIDE

PREDICTING THE NEXT BIG EARTHQUAKE

A year after the devastating earthquake in Haiti, people everywhere have been speculating about when and where the next cataclysmic one will hit our shores. If major quakes occur approximately 15 times a year around the world, where will the next one hit?

HOW OFTEN DO MAJOR EARTHQUAKES OCCUR?

MAGNITUDE	AVERAGE FREQUENCY PER YEAR
8.0+	1
7-7.9	15
6-6.9	134
5-5.9	1,319
4-4.9	13,000
3-3.9	130,000
2-2.9	1,300,000

On average, the vast majority of detectable tremors range between a magnitude of 2.0 to 4.0 on the Richter scale. Earthquakes comparable to the ones that rocked Haiti and Chile last year occur about 17 times a year. Those greater than an 8.0 magnitude happen approximately once a year.

The largest recorded earthquake in the world:
MAGNITUDE 9.5
CHILE, MAY 22, 1960

*Based on observations since 1900. Earthquake magnitudes before this year are estimates based on historical geological evidence.

DETECTED VS. UNDETECTED

Several million earthquakes shock the world each year but many go undetected because they hit remote areas or have small magnitudes.

There are an estimated

500,000

earthquakes detected in the world each year

100,000

of those can be felt.

100

of them cause damage.



Sources: U.S. Geological Survey | Discover | Time | Live Science

WHERE WILL THE NEXT BIG EARTHQUAKE STRIKE?

Although it's impossible to predict with a precise date and time, there are earthquake "hot spots" that scientists have identified, using historical information and sensitive tectonic plates as guides. Cataclysmic earthquakes can occur anywhere, but based on past notable earthquakes, these areas are often cited as some of the most susceptible.

CALIFORNIA

The San Andreas fault zone that forms the tectonic boundary between the Pacific Plate and the North American Plate cuts through a large length of California and has created many notable earthquakes.

THE PROBABILITY* OF A MAJOR EARTHQUAKE OVER THE NEXT 30 YEARS, OCCURRING IN:

SAN FRANCISCO BAY AREA

SOUTHERN CALIFORNIA



*Based on scientific estimates



The average rate of motion across the San Andreas Fault Zone is about **TWO INCHES A YEAR**. At this rate, Los Angeles and San Francisco could be pushed together as neighbors in approximately **15 MILLION YEARS**.



JAPAN

Roughly the same size as the state of California, Japan also has very active plate tectonics. The Tokai Gap of Japan is expected to rupture roughly every 150 years, causing 8.0 magnitude earthquakes.

FUKU 6/28/1948 7.3 Magnitude 3,769 Fatalities

MINO-OWARI 10/27/1891 8 Magnitude 7,237 Fatalities

NANKAIODA 1/20/1946 8.1 Magnitude 1,330 Fatalities

KOBE 1/16/1995 6.9 Magnitude 5,502 Fatalities

TONANKAI 12/7/1944 8.1 Magnitude 1,223 Fatalities

SANRIKU 3/2/1933 8.4 Magnitude 2,990 Fatalities

KANTO 9/1/1923 7.9 Magnitude 143,000 Fatalities

SANRIKU 6/15/1896 8.5 Magnitude 27,000 Fatalities

While California has roughly 37 million inhabitants, Japan has a population more than 3 times as large, with many packed into dense cities.

CHILE

Claiming the highest magnitude earthquake in recorded history, Chile is no stranger to calamitous earthquakes. Many of the country's deadliest quakes are due to offshore plate collisions that cause large tsunamis that threaten coastal communities.

SALAMANCA 4/6/1943 8.2 Magnitude 25 Fatalities

TARAPACA 6/13/2005 7.8 Magnitude 11 Fatalities

VALPARAISO 8/17/1906 8.2 Magnitude 3,882 Fatalities

VALDIVIA 5/22/1960 9.5 Magnitude 5,700 Fatalities

MAULE 2/27/2010 8.8 Magnitude 486 Fatalities

ARICA 8/13/1868 9 Magnitude 25,000 Fatalities

LA UGUA 3/28/1965 7.4 Magnitude 400 Fatalities

SUMATRA-ANDAMAN ISLANDS 12/26/2004 9.1 Magnitude 227,898 Fatalities

NORTHERN SUMATRA 3/28/2005 8.6 Magnitude 1,313 Fatalities

SOUTHERN SUMATRA 9/30/2009 7.5 Magnitude 1,117 Fatalities

SOUTHERN SUMATRA 9/13/2007 8.5 Magnitude 25 Fatalities

INDONESIA

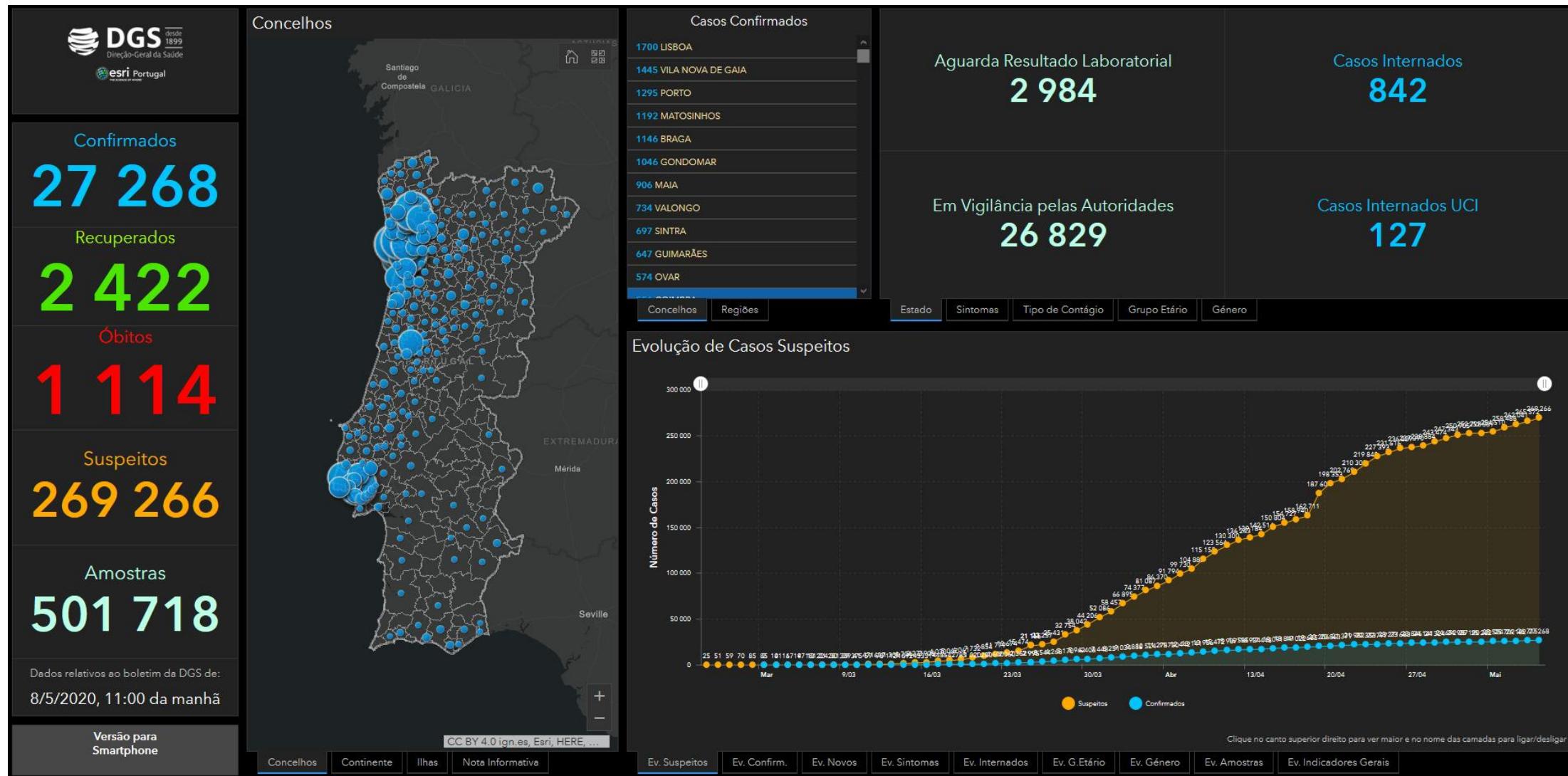
Indonesia sits at the center of turbulent fault lines that have been pressed against each other for thousands of years. Called the "Ring of Fire," it's a semicircle of volcanoes and shifting plates surrounding the edge of the Pacific Ocean. This volatile area has triggered countless quakes, volcanoes, and tsunamis in the Southeast Asia region.

PAPUA 6/25/1976 7.1 Magnitude 5,000 Fatalities

BANDA SEA 2/1/1998 8.5 Magnitude 8,500 Fatalities

FLORES REGION 12/15/1992 7.8 Magnitude 2,500 Fatalities

A COLLABORATION BETWEEN GOOD AND COLUMN FIVE



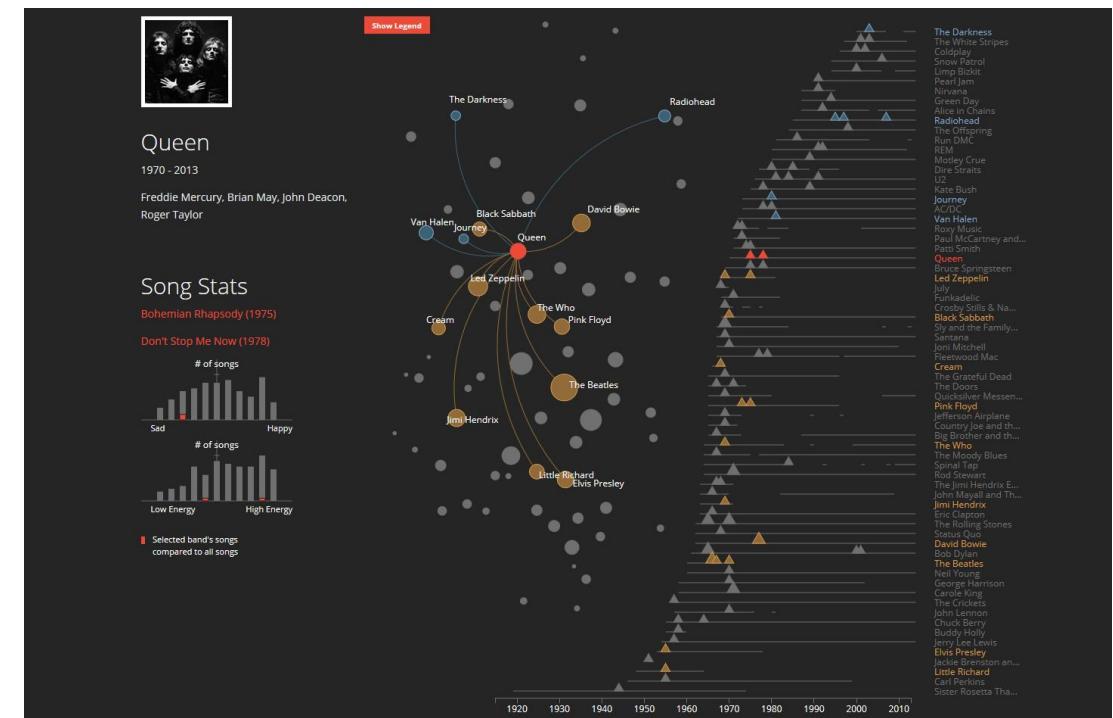
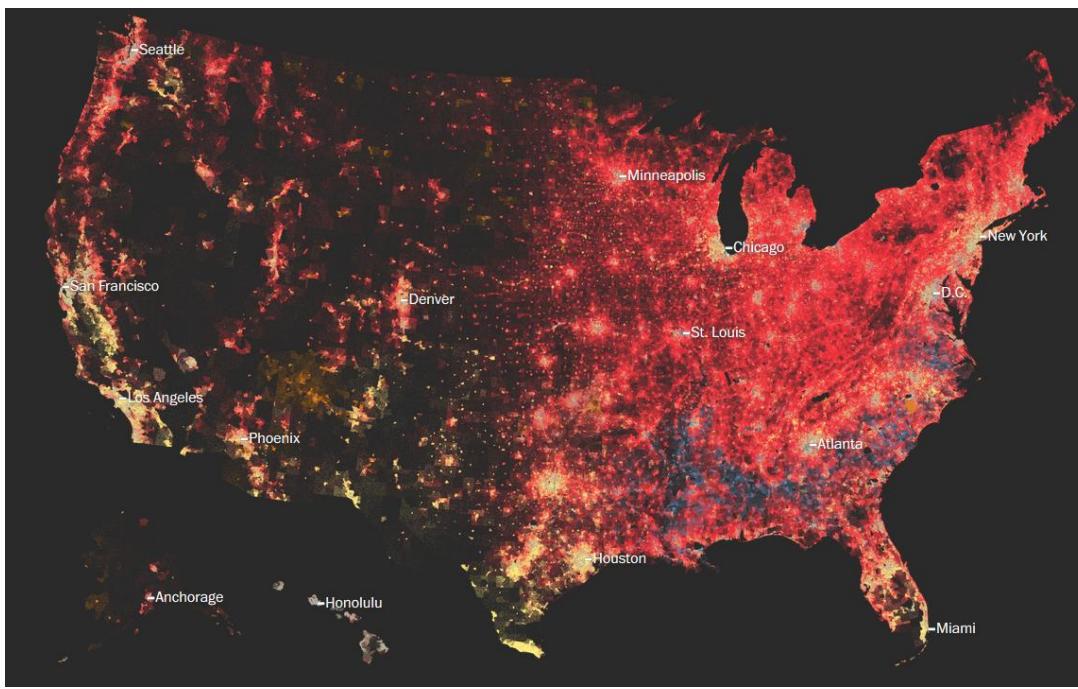
<https://esriportugal.maps.arcgis.com/apps/opsdashboard/index.html#/acf023da9a0b4f9dbb2332c13f635829>



<https://pplware.sapo.pt/informacao/como-e-o-mundo-das-amizades-no-facebook/>

<https://www.facebook.com/notes/10158791468612200/> - original

BONS EXEMPLOS



<https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/>

<https://svds.com/rockandroll/>

Para ler

- _ **Telling Human Stories with Data:** <https://medium.com/@ODSC/telling-human-stories-with-data-aef76c7ca3ae>
- _ **10 rules for better dashboard design:** <https://uxplanet.org/10-rules-for-better-dashboard-design-ef68189d734c>
- _ **What is an Infographic?** <https://visme.co/blog/what-is-an-infographic/>
- _ **20 ideas for better data visualization:** <https://uxdesign.cc/20-ideas-for-better-data-visualization-73f7e3c2782d>

Para ver

- _ **Visualization throughout the data science... - Lindsay Brin, PhD | ODSC East 2018:**
<https://www.youtube.com/watch?v=FnkOnlbamPU>
- _ **Effective Data Storytelling: How to turn Insights into Action - Brent Dykes:** https://youtu.be/jq4i1g0Ch_M?t=212

Consultar

- _ **Fundamentals of Data Visualization - Claus O.Wilke:** <https://clauswilke.com/dataviz/>
- _ **Bons exemplos de visualizações:** <https://informationisbeautiful.net/>
- _ **Livros recomendados (apenas alguns exemplos)**
Storytelling with data – Cole Nussbaumer Knaflic – [Wook](#) | [Bertrand](#) | [Amazon UK](#) | [Amazon \(versão em Espanhol\)](#)
Effective Data Visualization: The Right Chart for the Right Data – Stephanie Evergreen – [Amazon UK](#)
Information Dashboard Design – Stephen Few – [Wook](#) | [Bertrand](#) | [Amazon UK](#)

E
D
I.
T.



QUESTÕES