

# PTC 3467– Ciência dos Dados

## Visualização de Dados

**Prof. Dr. Pedro Luiz Pizzigatti Corrêa**

**Prof. Dr. André Filipe Batista**



ESCOLA  
POLITÉCNICA  
DA USP

# Agenda

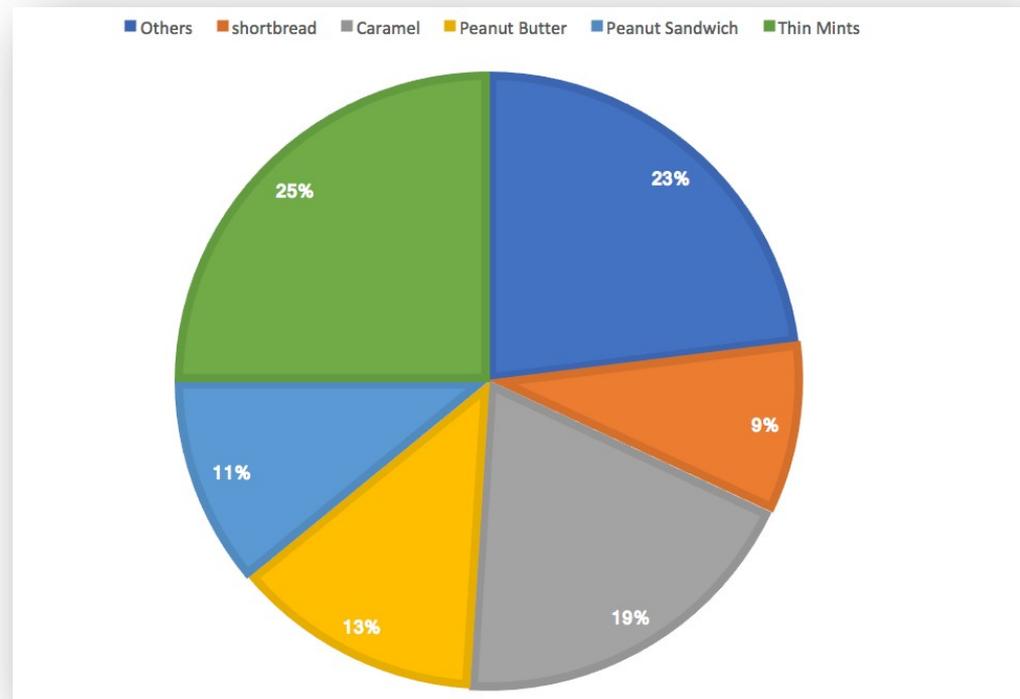


- Motivação
- Visualização de Dados e Visualização Analítica
- Escolhendo a visualização mais adequada para seus dados
- Visualizando seus dados com Tableau
- Parte Prática: Tableau 10.2
- Tableau e Dados do Projeto GoAmazon

# A Importância da Visualização

- Qual das seguintes imagens permite uma melhor compreensão?

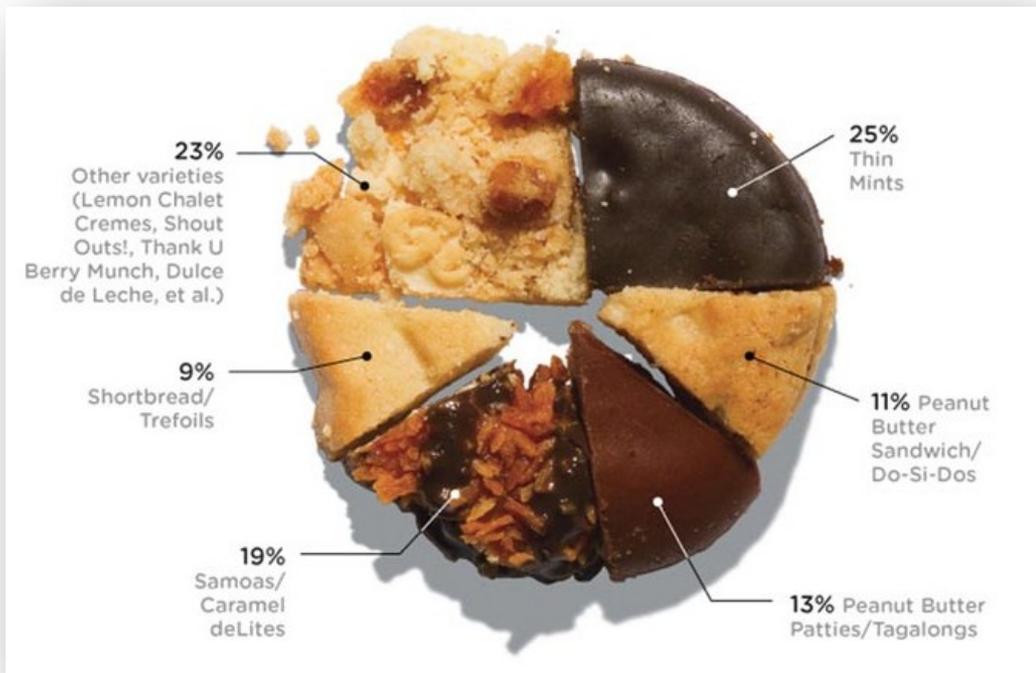
Opção 1:



# A Importância da Visualização

- Qual das seguintes imagens permite uma melhor compreensão?

Opção 2:

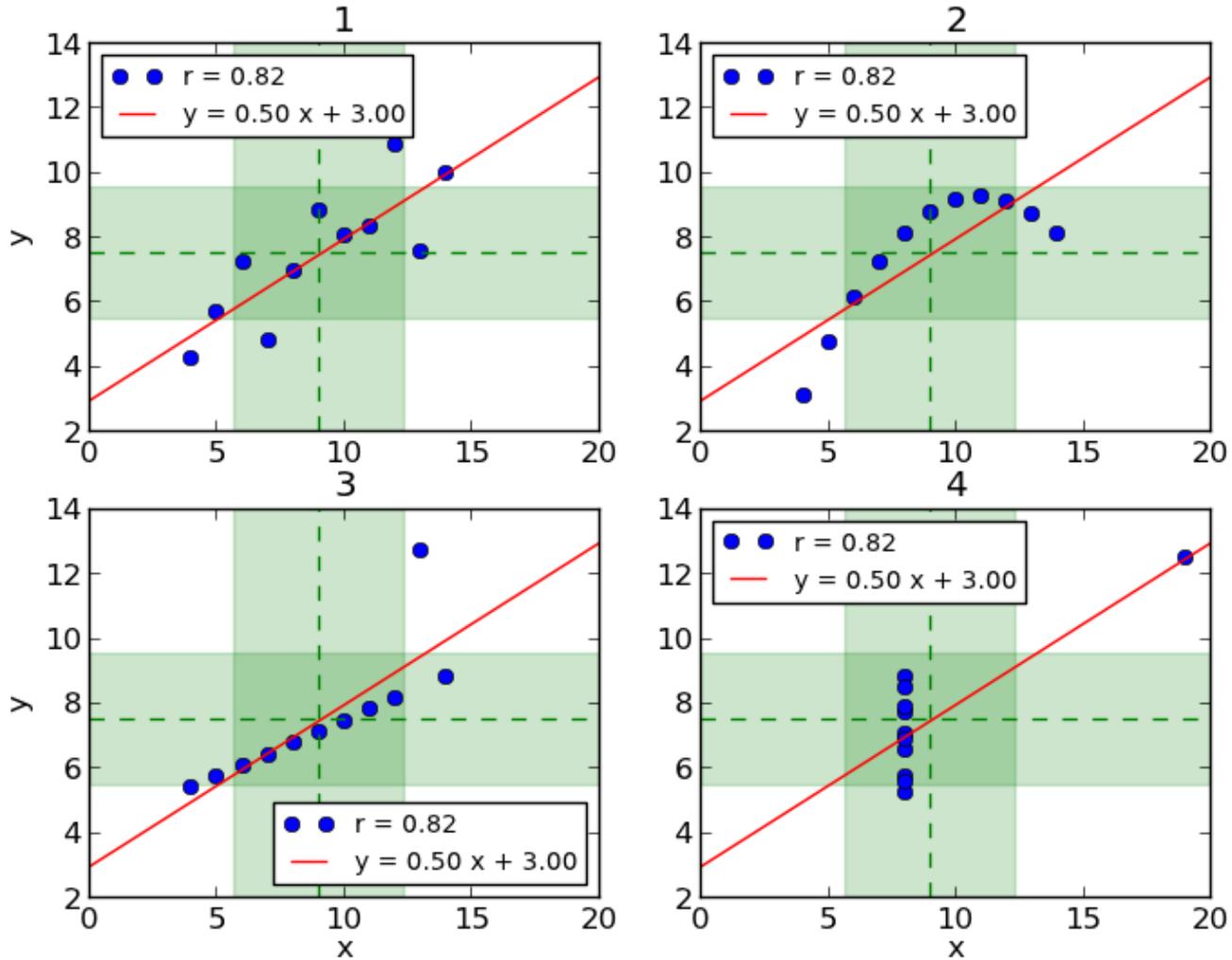


Observe estes dados... eles parecem similares

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

# Mas não são ...

## Anscombe's Quartet



# Mais um exemplo: A Guerra da Crimeia

- Conflito que se estendeu de 1854 a 1856
- 25k britânicos, 100k franceses e 1Mde russos morreram

Combatentes			
	Império Francês		Império Russo
	Império Otomano		
	Reino Unido		
	Reino da Sardenha		

British Casualties in the Crimean War

Data	Lesão	Em campo	Doença
05/1854	0	95	105
06/1854	0	40	95
07/1854	0	140	520
08/1854	20	150	800
09/1854	220	230	740
10/1854	305	310	600
11/1854	480	290	820
12/1854	295	310	1100
01/1855	230	460	1440
02/1855	180	520	1270
03/1855	155	350	935
04/1855	195	195	560
05/1855	180	155	550
06/1855	330	130	650
07/1855	260	130	430
08/1855	290	110	490
09/1855	355	100	290
10/1855	135	95	245
11/1855	100	140	325
12/1855	40	120	215
01/1856	0	160	160
02/1856	0	100	100
03/1856	0	125	90

Qual foi o mês com maior número de mortes em campo de batalha?

Em quais meses o número de mortes por lesão excede morte em campo?

Qual o mês com o maior número de mortes no total?

Meses em que a % de mortes por doença foi menor que 50%?

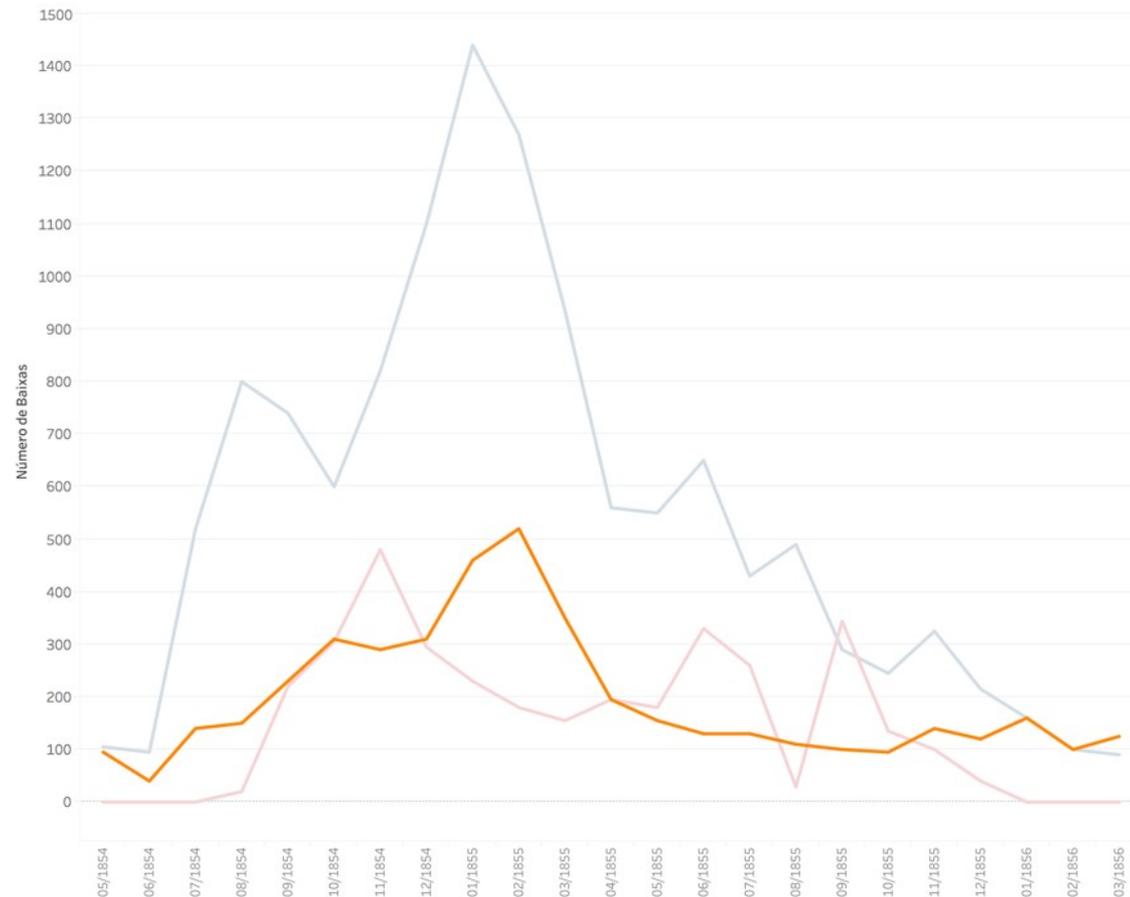
Adaptado de (Sultanum, 2015)

# Mais um exemplo: A Guerra da Crimeia

Qual foi o mês com maior número de mortes em campo de batalha?

Data	Lesão	Em campo	Doença
05/1854	0	95	105
06/1854	0	40	95
07/1854	0	140	520
08/1854	20	150	800
09/1854	220	230	740
10/1854	305	310	600
11/1854	480	290	820
12/1854	295	310	1100
01/1855	230	460	1440
02/1855	180	520	1270
03/1855	155	350	935
04/1855	195	195	560
05/1855	180	155	550
06/1855	330	130	650
07/1855	260	130	430
08/1855	290	110	490
09/1855	355	100	290
10/1855	135	95	245
11/1855	100	140	325
12/1855	40	120	215
01/1856	0	160	160
02/1856	0	100	100
03/1856	0	125	90

British Casualties in the Crimean War



Adaptado de (Sultanum, 2015)

# Mais um exemplo: A Guerra da Crimeia

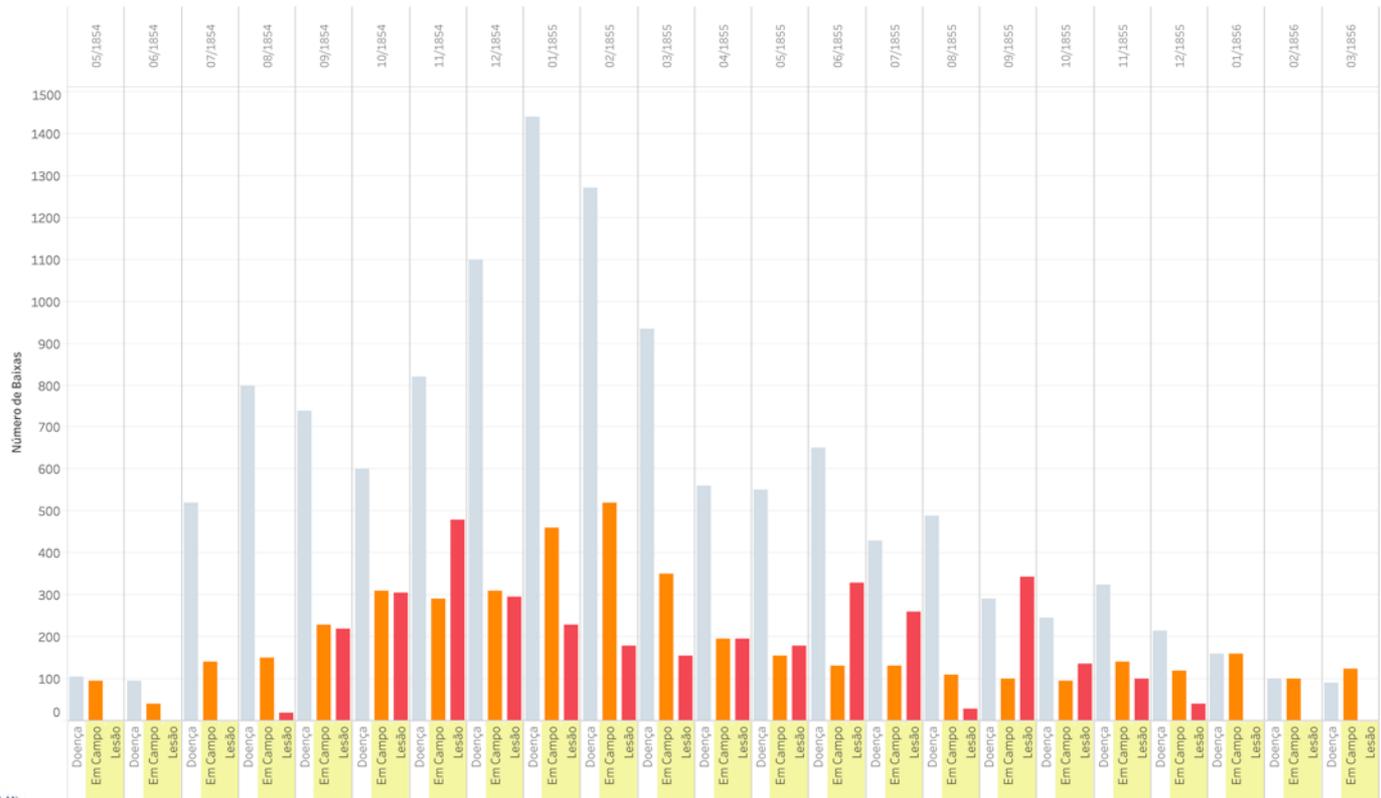
Em quais meses o número de mortes por lesão excede morte em campo?

Nomes de medida

- Lesão
- Doença
- Em Campo

Data	Lesão	Em campo	Doença
05/1854	0	95	105
06/1854	0	40	95
07/1854	0	140	520
08/1854	20	150	800
09/1854	220	230	740
10/1854	305	310	600
11/1854	480	290	820
12/1854	295	310	1100
01/1855	230	460	1440
02/1855	180	520	1270
03/1855	155	350	935
04/1855	195	195	560
05/1855	180	155	550
06/1855	330	130	650
07/1855	260	130	430
08/1855	290	110	490
09/1855	355	100	290
10/1855	135	95	245
11/1855	100	140	325
12/1855	40	120	215
01/1856	0	160	160
02/1856	0	100	100
03/1856	0	125	90

British Casualties in the Crimean War



Adaptado de (Sultanum, 2015)

# Mais um exemplo: A Guerra da Crimeia

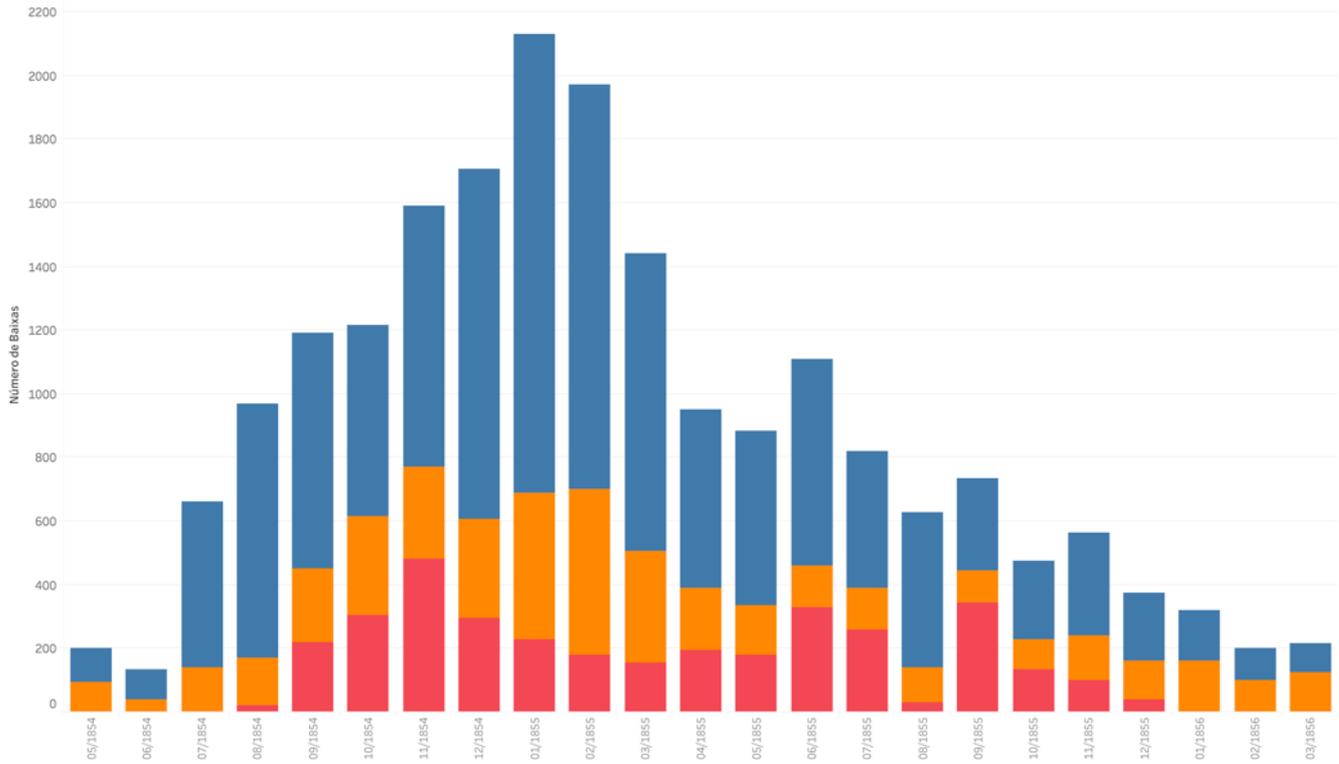
Qual o mês com o maior número de mortes no total?

Nomes de medida

- Lesão
- Doença
- Em Campo

Data	Lesão	Em campo	Doença
05/1854	0	95	105
06/1854	0	40	95
07/1854	0	140	520
08/1854	20	150	800
09/1854	220	230	740
10/1854	305	310	600
11/1854	480	290	820
12/1854	295	310	1100
01/1855	230	460	1440
02/1855	180	520	1270
03/1855	155	350	935
04/1855	195	195	560
05/1855	180	155	550
06/1855	330	130	650
07/1855	260	130	430
08/1855	290	110	490
09/1855	355	100	290
10/1855	135	95	245
11/1855	100	140	325
12/1855	40	120	215
01/1856	0	160	160
02/1856	0	100	100
03/1856	0	125	90

British Casualties in the Crimean War



Adaptado de (Sultanum, 2015)

# Mais um exemplo: A Guerra da Crimeia

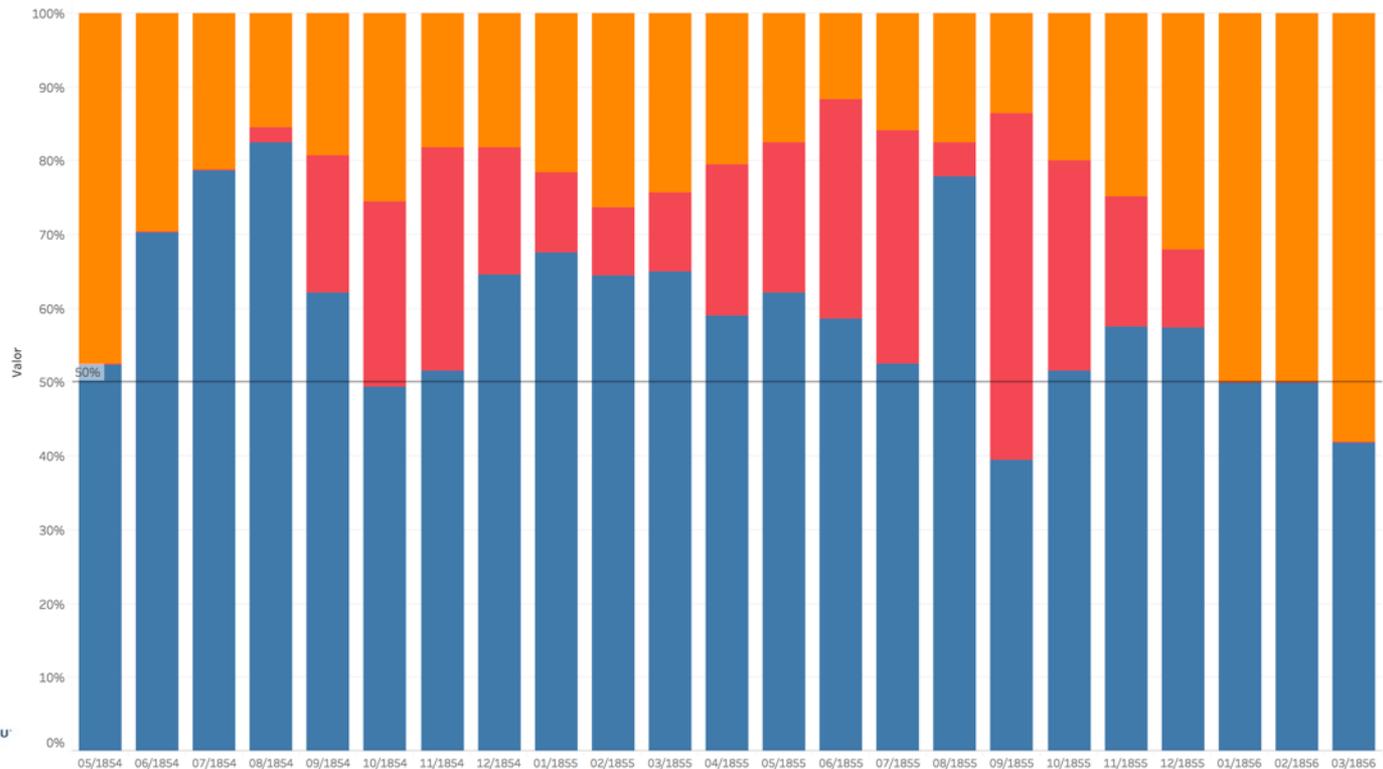
Meses em que a % de mortes por doença foi menor que 50%?

Nomes de medida

- Lesão
- Doença
- Em Campo

Data	Lesão	Em campo	Doença
05/1854	0	95	105
06/1854	0	40	95
07/1854	0	140	520
08/1854	20	150	800
09/1854	220	230	740
10/1854	305	310	600
11/1854	480	290	820
12/1854	295	310	1100
01/1855	230	460	1440
02/1855	180	520	1270
03/1855	155	350	935
04/1855	195	195	560
05/1855	180	155	550
06/1855	330	130	650
07/1855	260	130	430
08/1855	290	110	490
09/1855	355	100	290
10/1855	135	95	245
11/1855	100	140	325
12/1855	40	120	215
01/1856	0	160	160
02/1856	0	100	100
03/1856	0	125	90

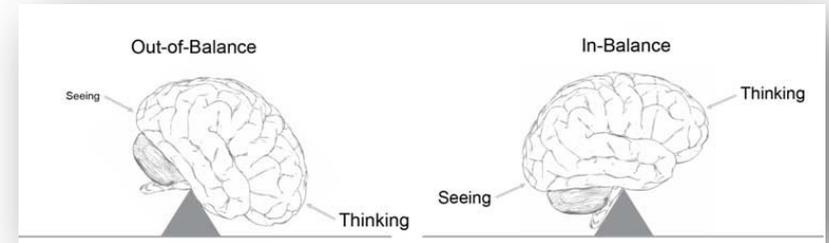
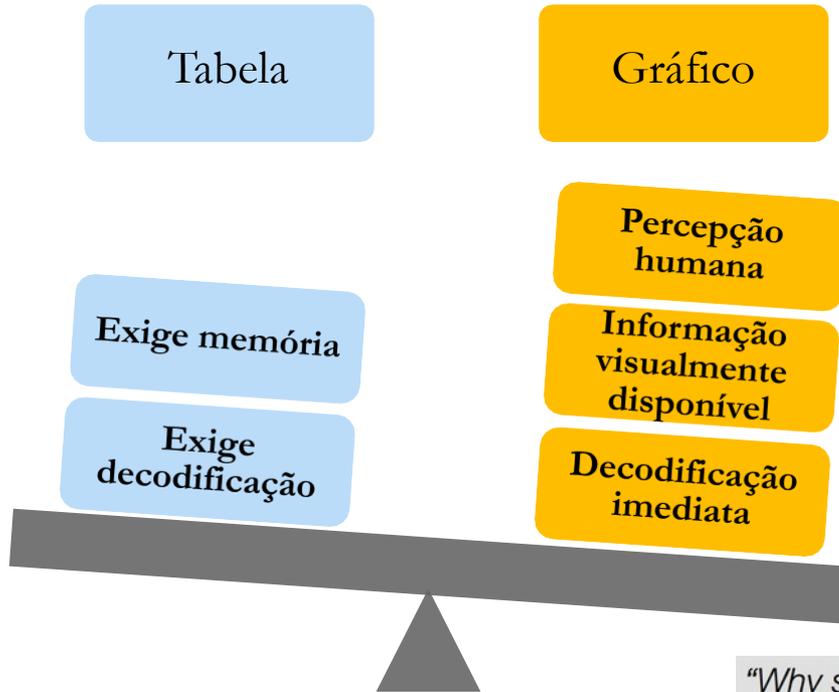
British Casualties in the Crimean War



tableau

Adaptado de (Sultanum, 2015)

# Visão → Alta Performance



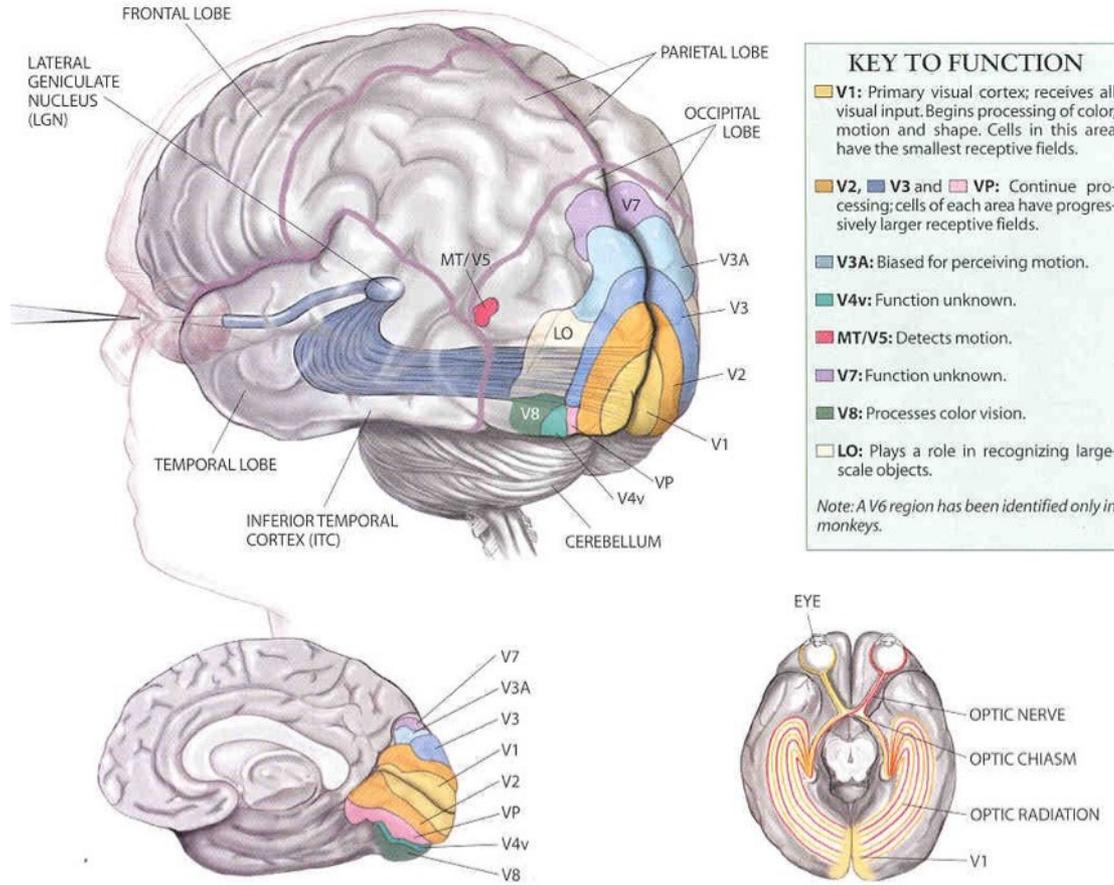
“An estimated 50 percent of the brain's neurons are associated with vision. Visualisation <...> aims to put that neurological machinery to work.”

- B. McCormick, T. DeFanti, and M. Brown. Definition of Visualization. *ACM SIGGRAPH Computer Graphics*, 21(6), November 1987, p.3

*“Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers.”*

Ware, Colin. Information visualization: perception for design. Elsevier, 2013.

# Somos Visuais



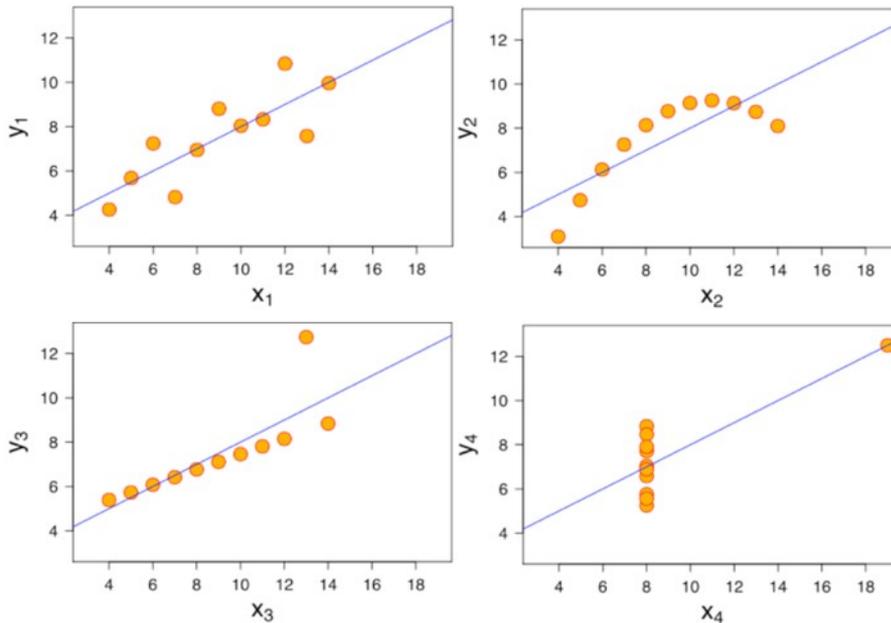
TERESE WINSLOW, WITH ASSISTANCE FROM MOUNICHRNE, HAOJIKHANI AND ROBERT TOOTELL, Harvard Medical School

**HUMAN VISUAL PATHWAY** begins with the eyes and extends through several interior brain structures before ascending to the various regions of the visual cortex (V1, and so on). At the optic chiasm, the optic nerves cross over partially so that each hemisphere of the brain receives input from both eyes. The information

is filtered by the lateral geniculate nucleus, which consists of layers of nerve cells that each respond only to stimuli from one eye. The inferior temporal cortex is important for seeing forms. Researchers have found that some cells from each area are active only when a person or monkey becomes conscious of a given stimulus.

# Visualização expõe detalhes

- Reduzir/Sumarizar os dados remove detalhes que podem ser cruciais. Por isso é importante visualizar os dados



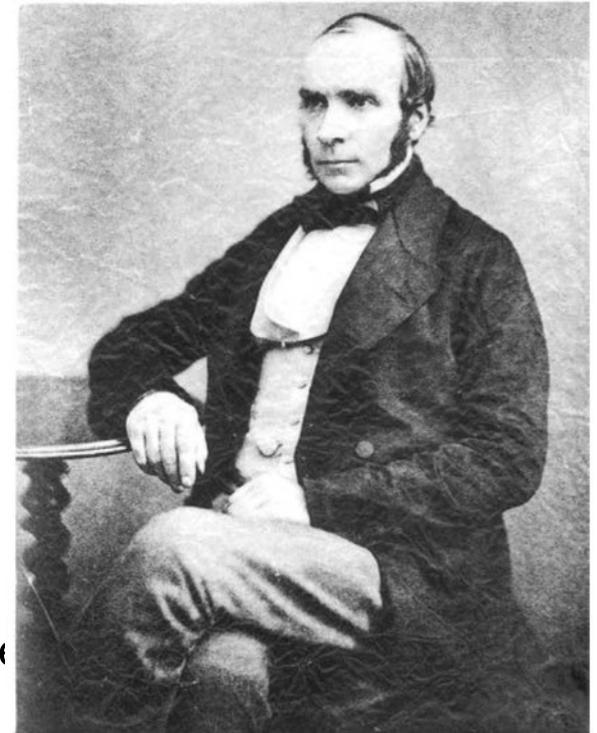
➤ Média, variância, correlação e linha de regressão linear são idênticas...

➤ ... mas estruturas diversas!

➤ Distribuição, outliers, continuidade...

# Exemplo histórico

- Em 1854, um surto de cólera atingiu Londres
  - 127 pessoas próximas a Broad St. morreram em 3 dias
  - 616 pessoas morreram nos primeiros 30 dias
- Explicação inicial:
  - *“Miasma na atmosfera”*
- Dr. John Snow foi o primeiro a associar o surto de cólera à **contaminação na água**
- Como ele fez isto?
  - Conversou com os residentes locais
  - Identificou uma fonte de água como fonte provável
  - Utilizou mapas para ilustrar a sua teoria
  - Convenceu autoridades a desativarem a fonte de água



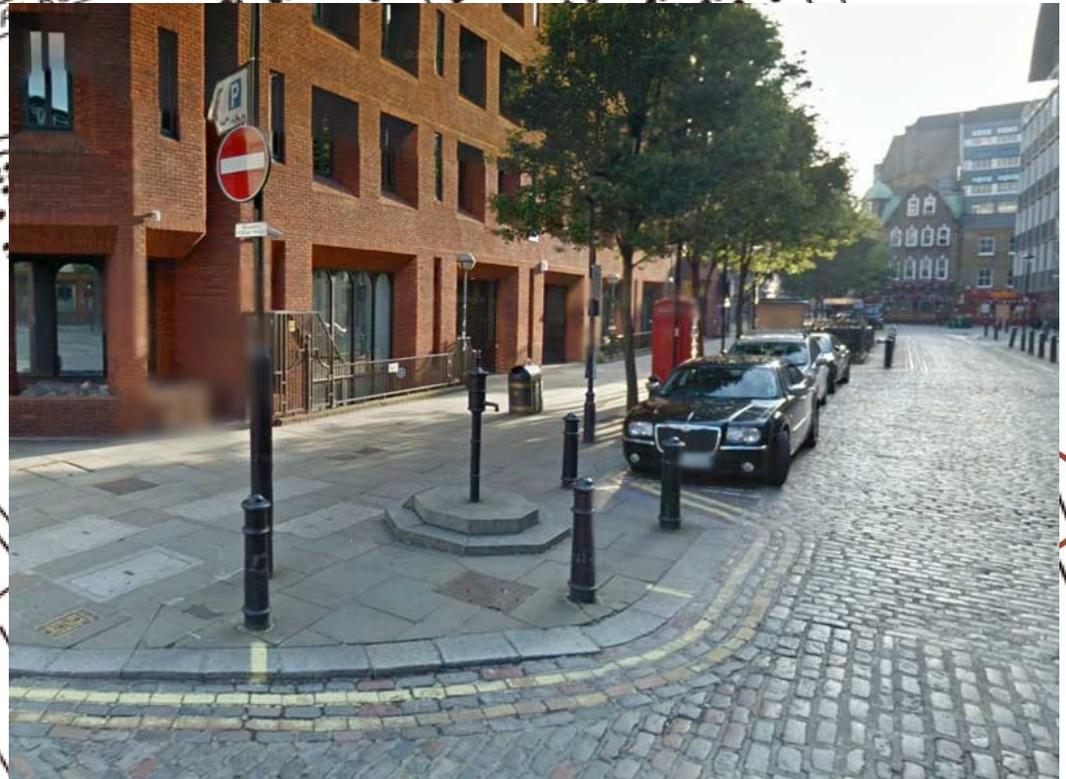
# Exemplo histórico



# Exemplo histórico



The Red Granite kerbstone  
marks the site of the historic  
**BROAD STREET PUMP**  
associated with Dr. John Snow's  
discovery in 1854  
that Cholera is conveyed by water



# Hans Rosling

- Médico, acadêmico, estatístico e orador público sueco
- Foi professor de Saúde Internacional no Karolinska Institutet
- Promoveu o uso de dados para explorar problemas de desenvolvimento
- Faleceu em 7/2/2017



Imagem: Wikipedia (2017).



YouTube

200 anos países, 200 anos, 4 minutos

<https://www.youtube.com/watch?v=fJ6y8ZJ>

[MoqM](#)

# O Poder da Visualização Science, 27/4/2017



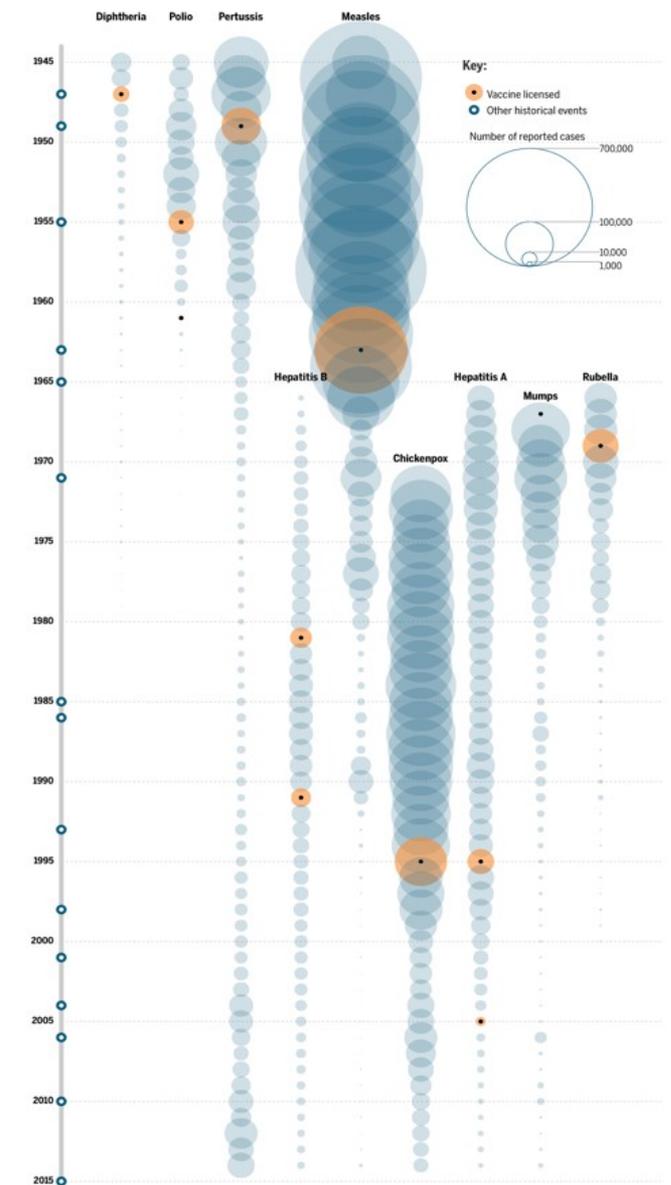
Here's the visual proof of why vaccines do more good than harm

By Jia You | Apr. 27, 2017, 1:15 PM

Acesse o link da publicação para ver o Gráfico interativo

<http://www.sciencemag.org/news/2017/04/heres-visual-proof-why-vaccines-do-more-good-harm>

\*Chickenpox data were not reported by all states between 1981 and 2003.



(Graphic) J. You/Science (Data) Centers for Disease Control and Prevention

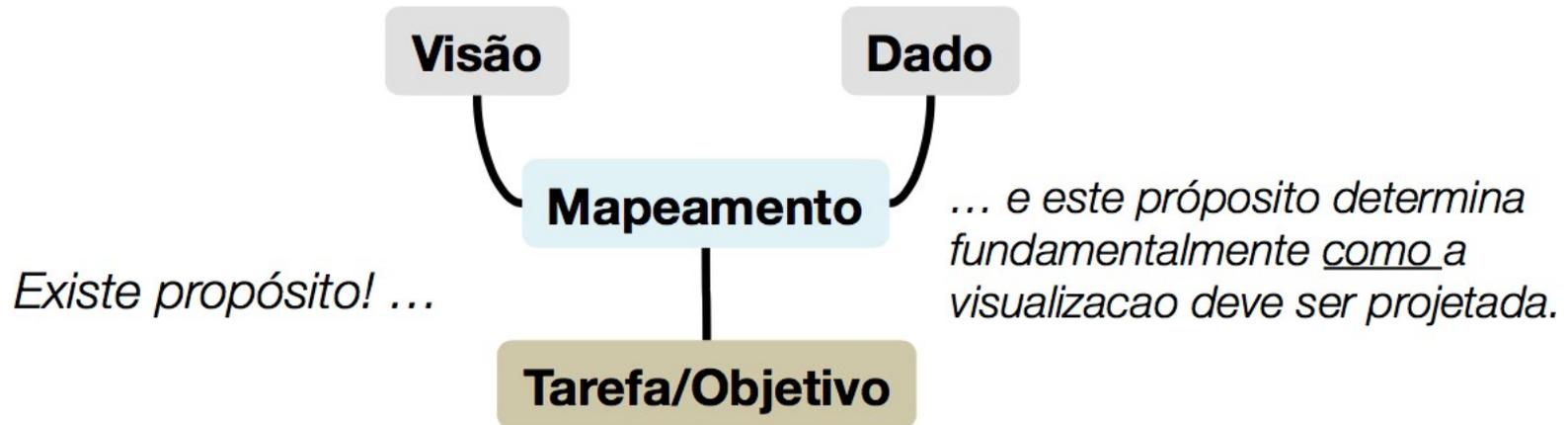
# Definição de Visualização de Dados

*“The use of computer-supported, interactive, visual representations of data to amplify cognition”.*

*Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think  
Morgan Kaufmann Publishers, 1999.*

*“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”*

*Munzner, Tamara. Visualization Analysis and Design. CRC Press, 2014.*



Adaptado de (Sultanum, 2015)

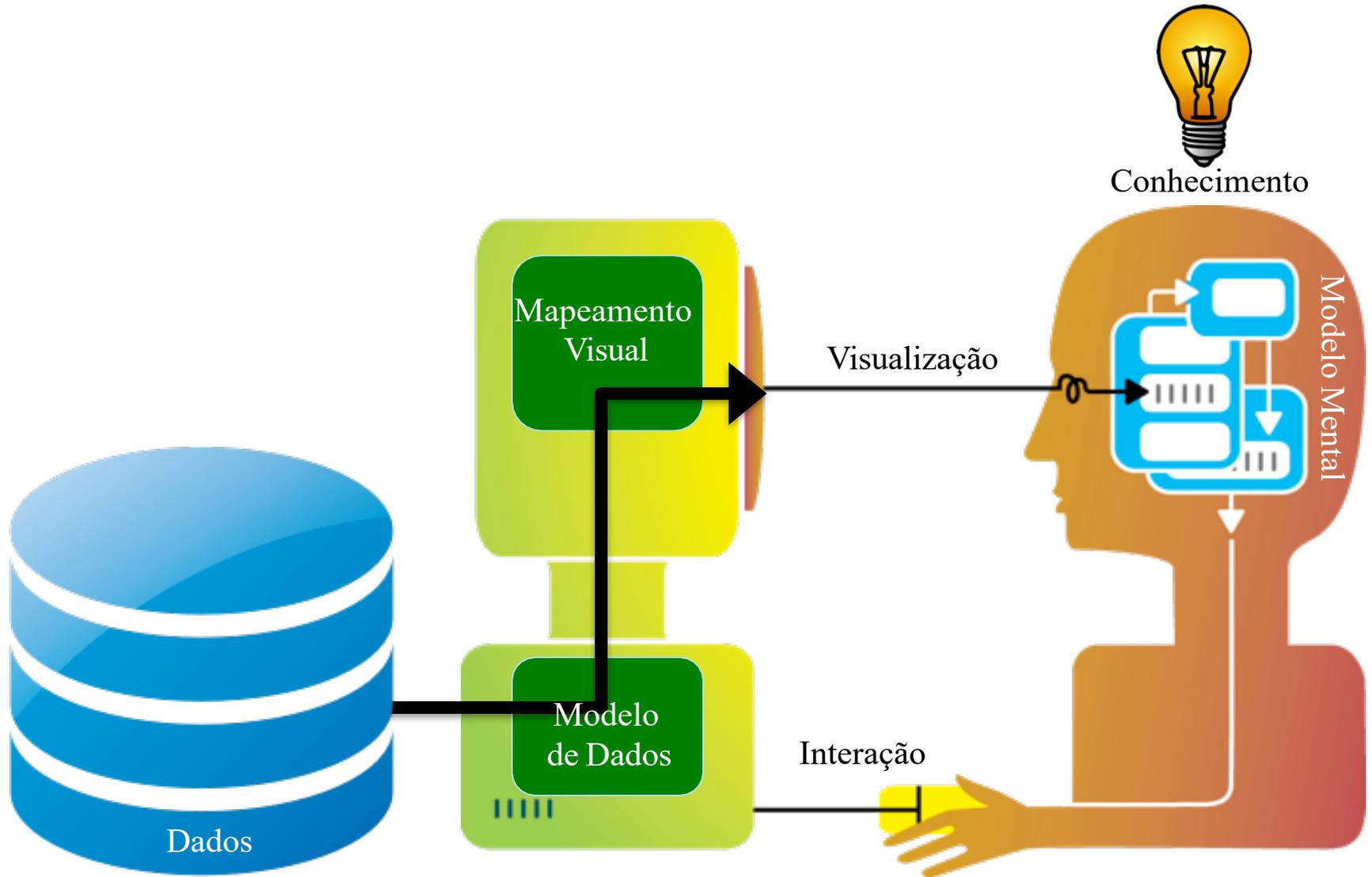
# A área de Visualização de Dados

- Multidisciplinar
  - Diversos domínios de aplicação
  - Áreas do conhecimento:
    - Ciências Cognitivas
    - Design
    - Interação Humano-Computador
    - Computação
    - Estatística

# Sub-áreas de Visualização

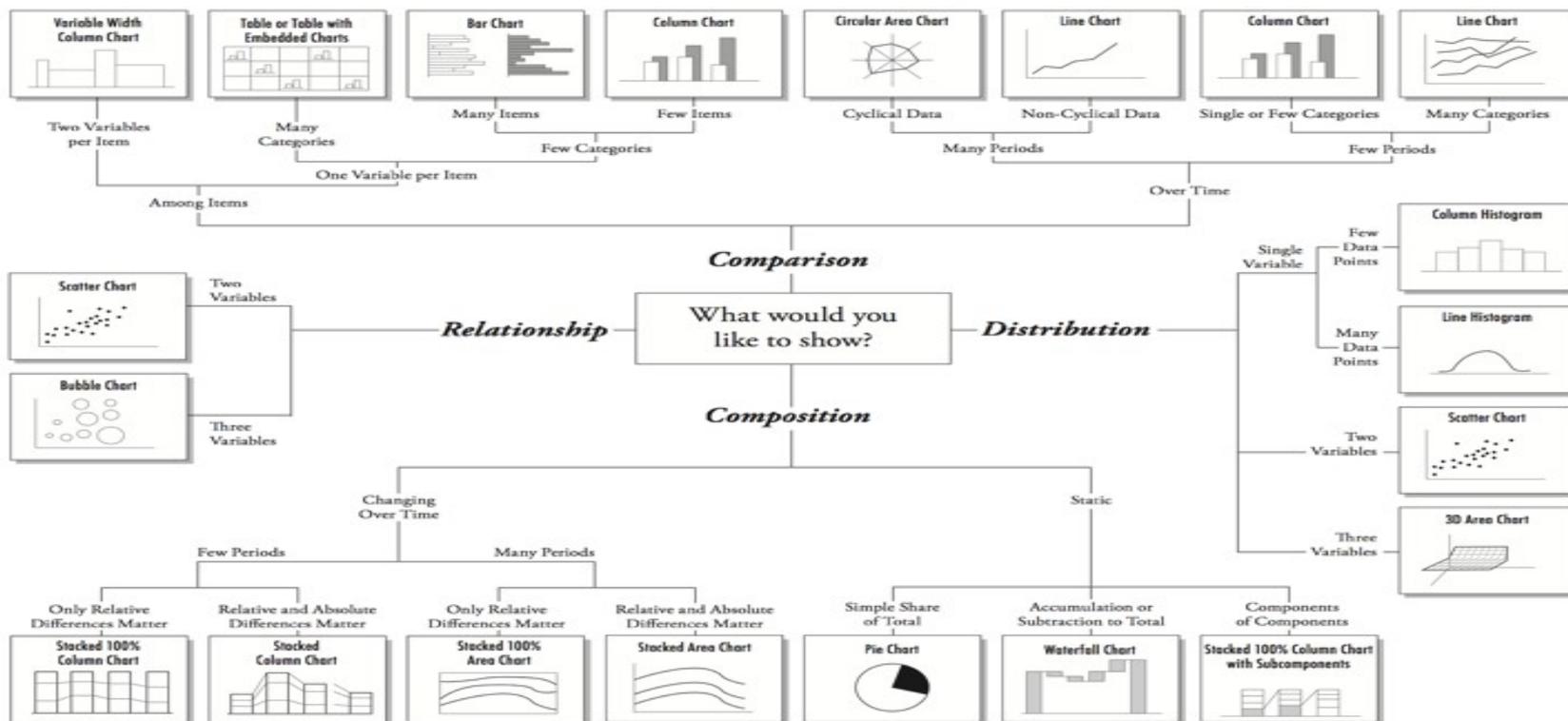
- Visualização Científica
  - Dados de aplicação científica
  - Simulações físicas, dados de scanners médicos
- Visualização da Informação
  - Transações financeiras, tweets, conexões em redes sociais, dados de censo, etc...
- Visualização Analítica
  - Uso de ferramentas analíticas para “amplificação da cognição humana”
  - Datasets multidimensionais, aplicações em big data, correlações
  - Dados + Visualização + Interatividade

# Visualização Analítica



# Escolhendo a visualização mais adequada para os seus dados

## Chart Suggestions—A Thought- Starter



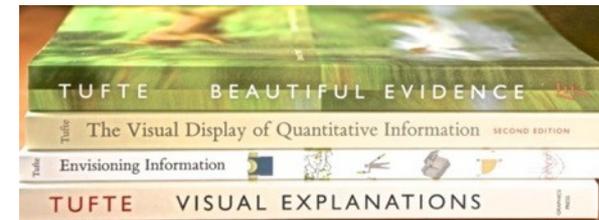
www.ExtremePresentation.com  
© 2009 A. Abela — a.v.abela@gmail.com

# Edward Tufte

- Professor da Yale University
- Definiu um conjunto de “princípios” sobre visualização de dados



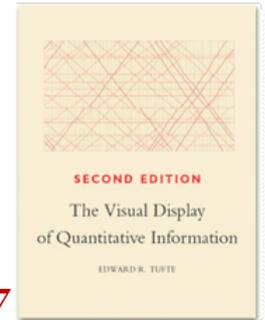
- “Tufte’s first principles”: - graphical displays should:
  - Show the data
  - Have viewer think about substance, rather than methodology
  - Avoid distorting what lies in data
  - Present numbers in a small space
  - Make large data sets coherent
  - Encourage the eye to compare different pieces of data
  - Reveal data at several levels of detail
  - Serve clear purpose: description, exploration, tabulation or decoration
  - Be closely integrated with statistical data and verbal description



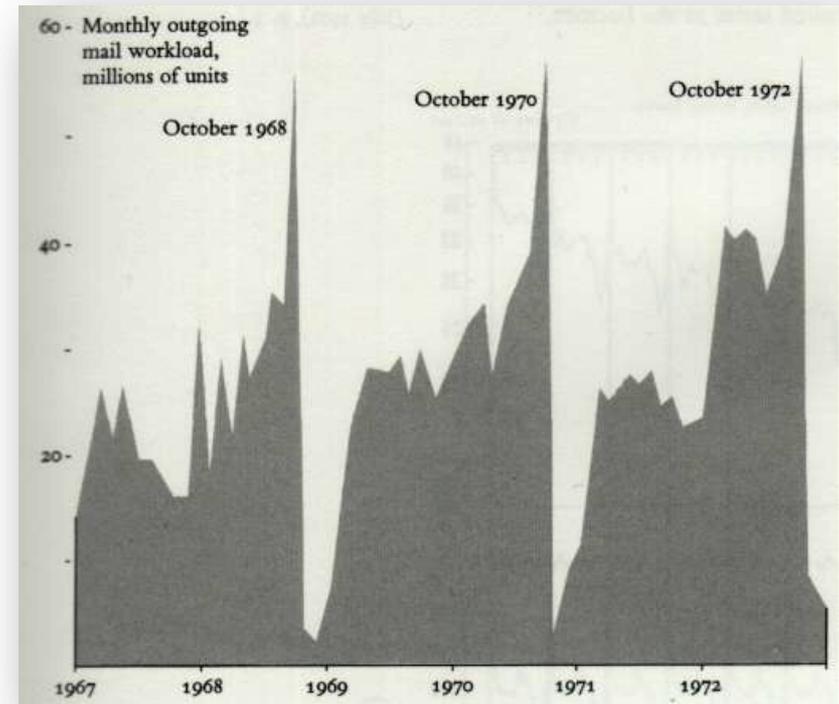
# Algumas recomendações de Tufte

- Cuidado com gráficos de séries temporais
- A simples passagem do tempo não é uma boa variável explanatória de um fenômeno

*“descriptive chronology is not causal explanation”, pg. 37*

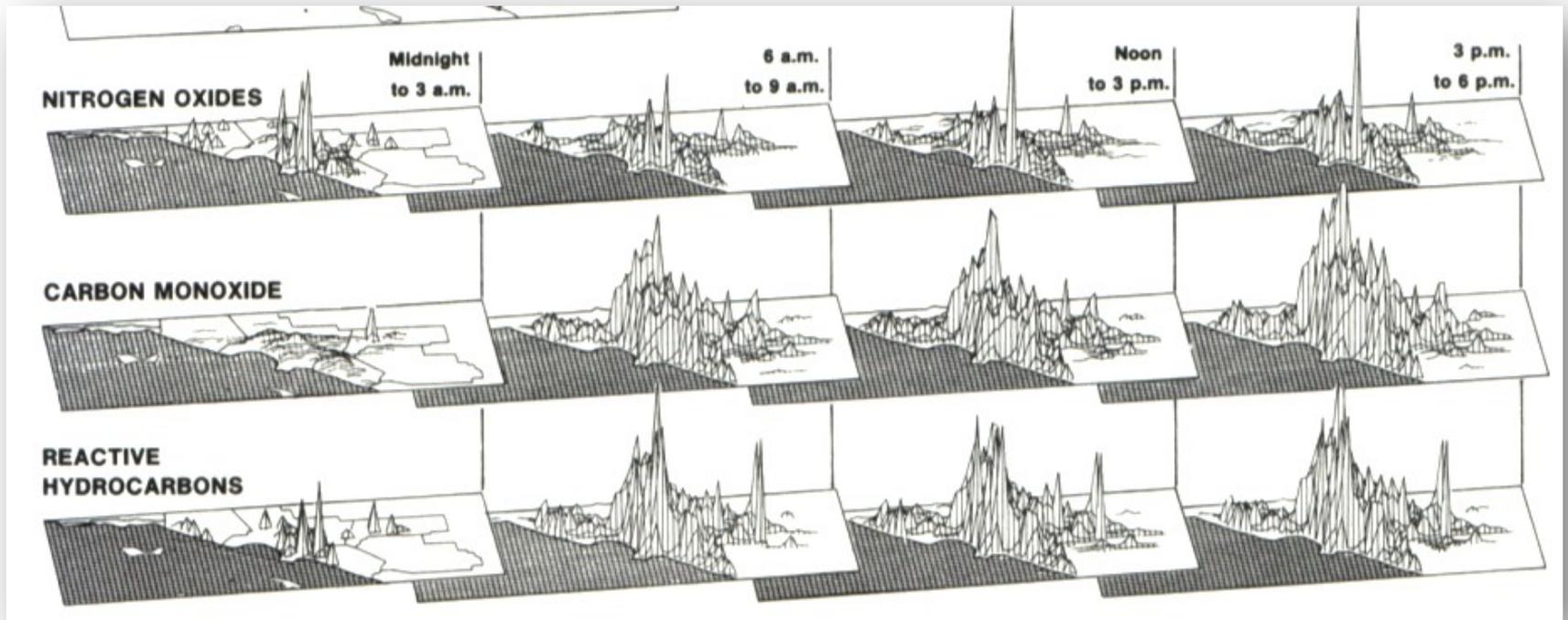


Visualização de correspondência  
enviadas pelo congresso  
Eleições americanas → Novembro



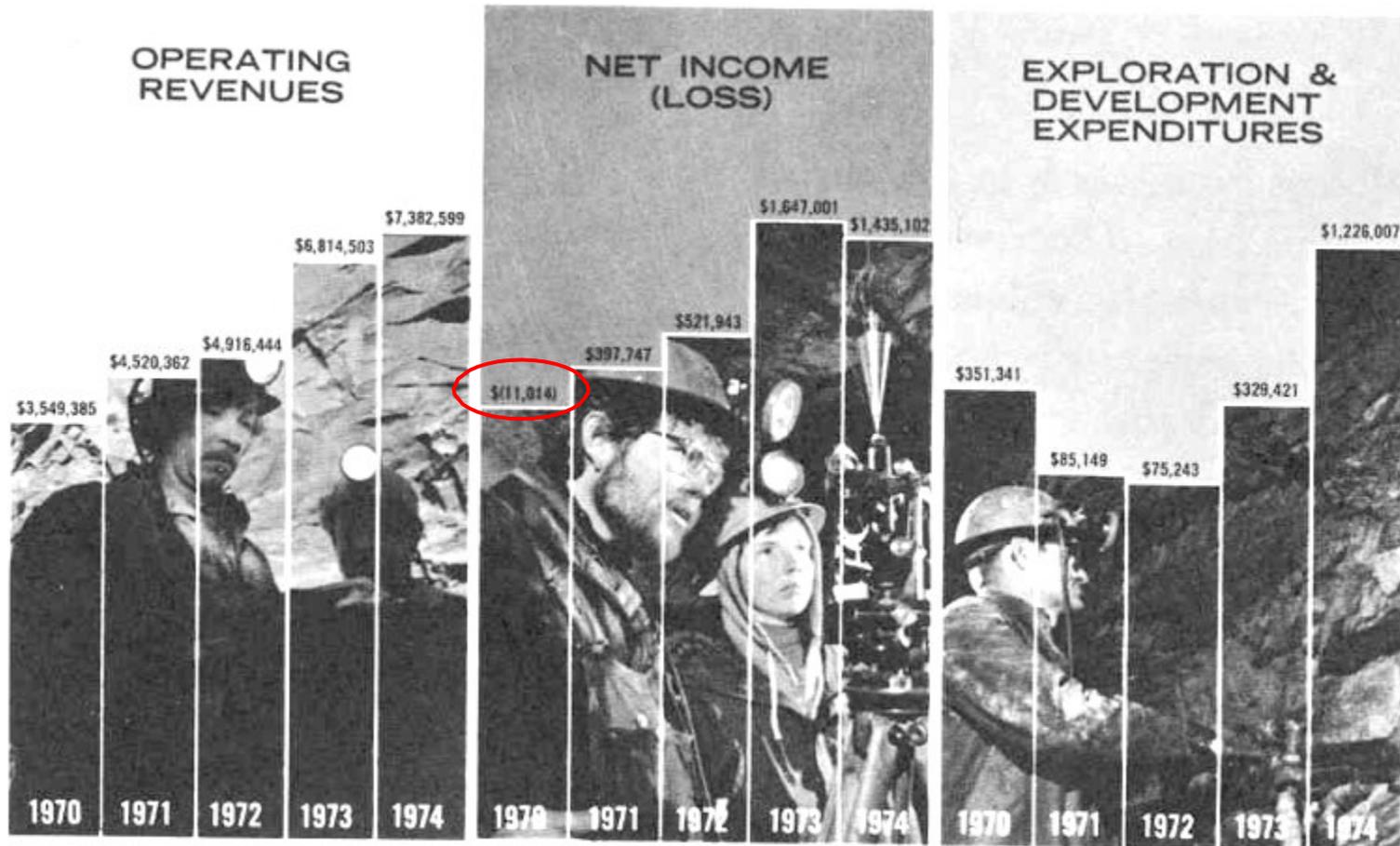
# Algumas recomendações de Tufte

- *Small multiple – more later*
- *Poluentes na bacia de LA*



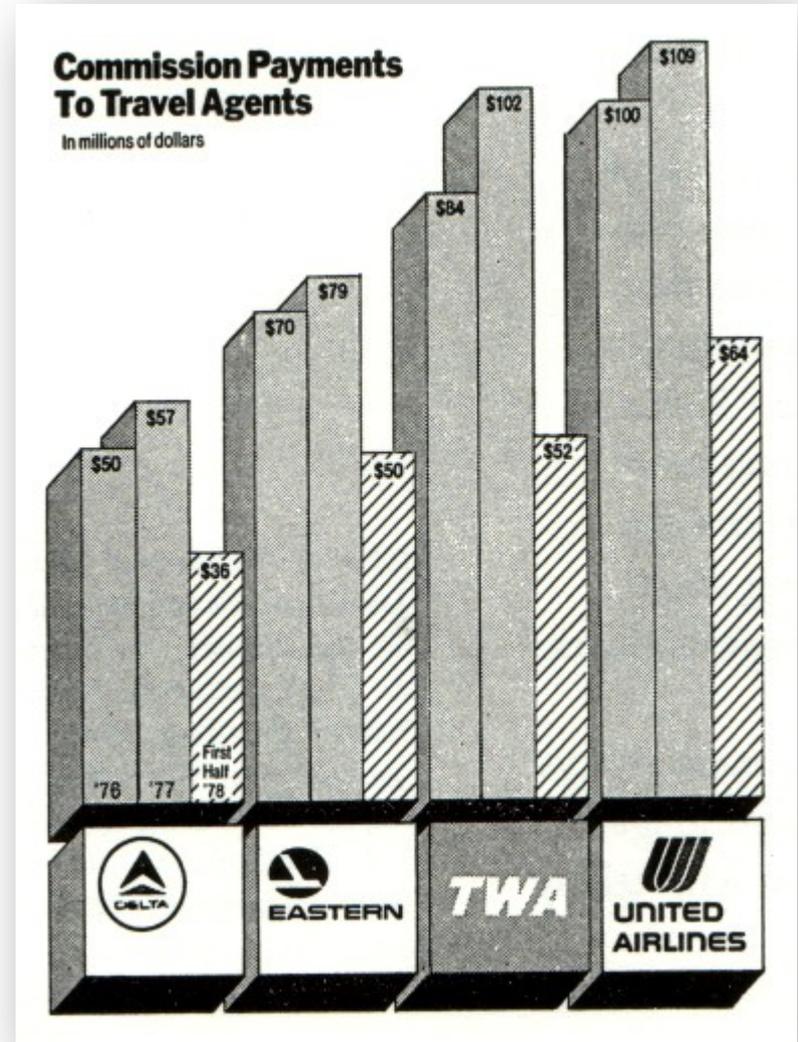
# Algumas recomendações de Tufte

- *Mantenha a integridade visual*



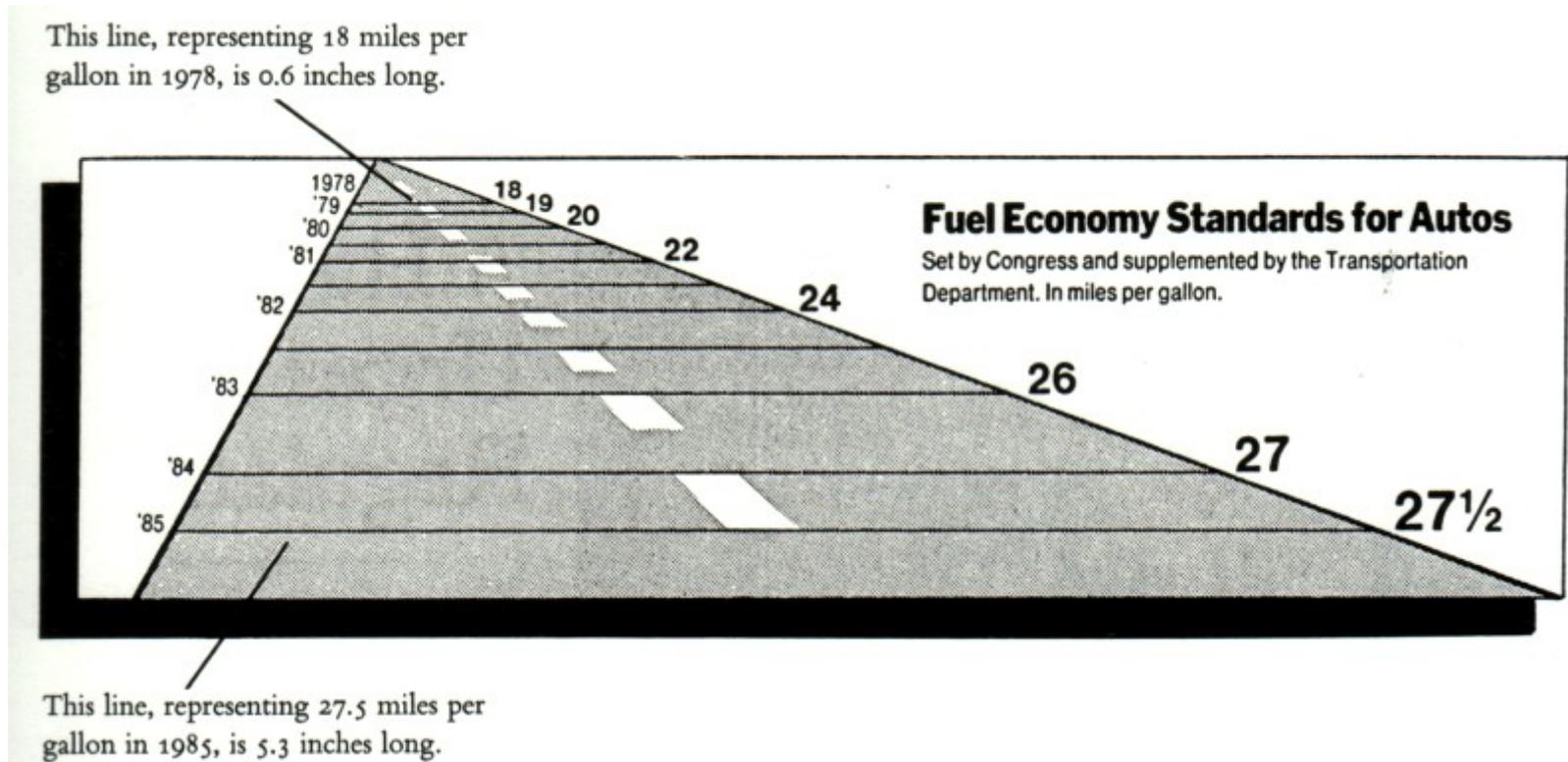
# Algumas recomendações de Tufte

- *Mantenha a integridade visual*
- *1978 aparenta ser um ano ruim, mas o gráfico foi feito apenas com dados do primeiro semestre de 78.*



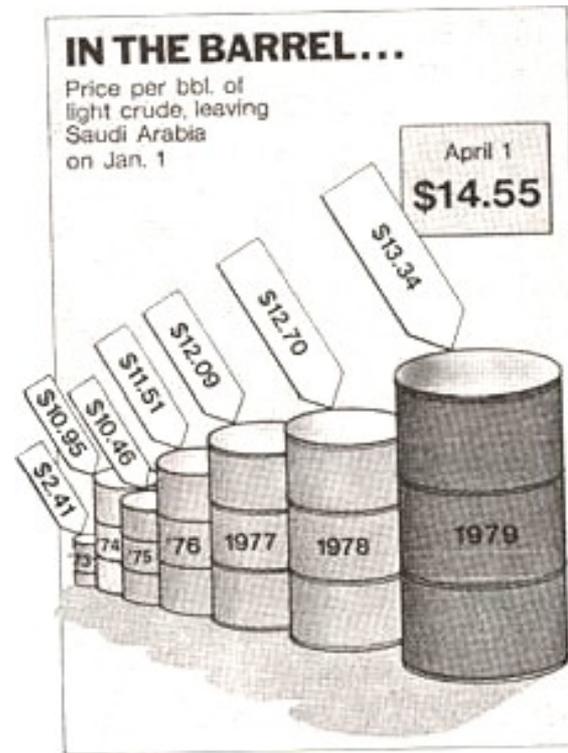
# Algumas recomendações de Tufte

- **Lie Factor**
- Size of effect in graphic / Size of effect in graphic, should = 1 = “integrity”



# Algumas recomendações de Tufte

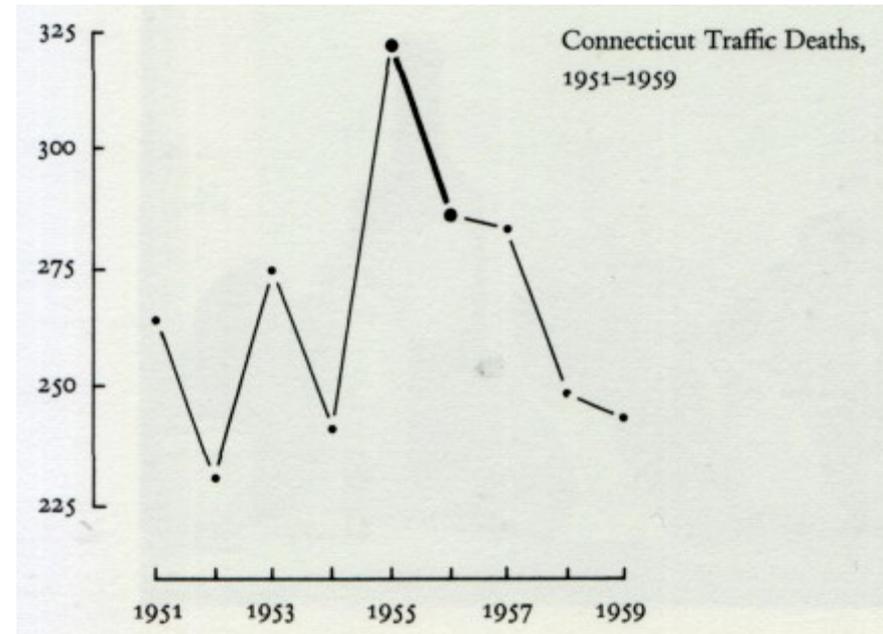
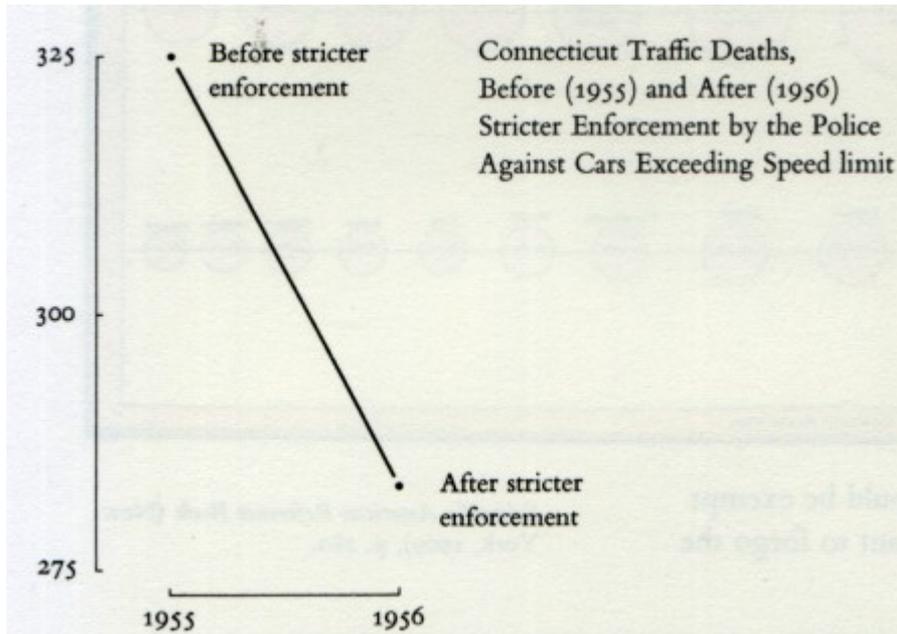
- **Lie Factor**
- Size of effect in graphic / Size of effect in graphic, should = 1 = “integrity”



Lie Factor = 9.4

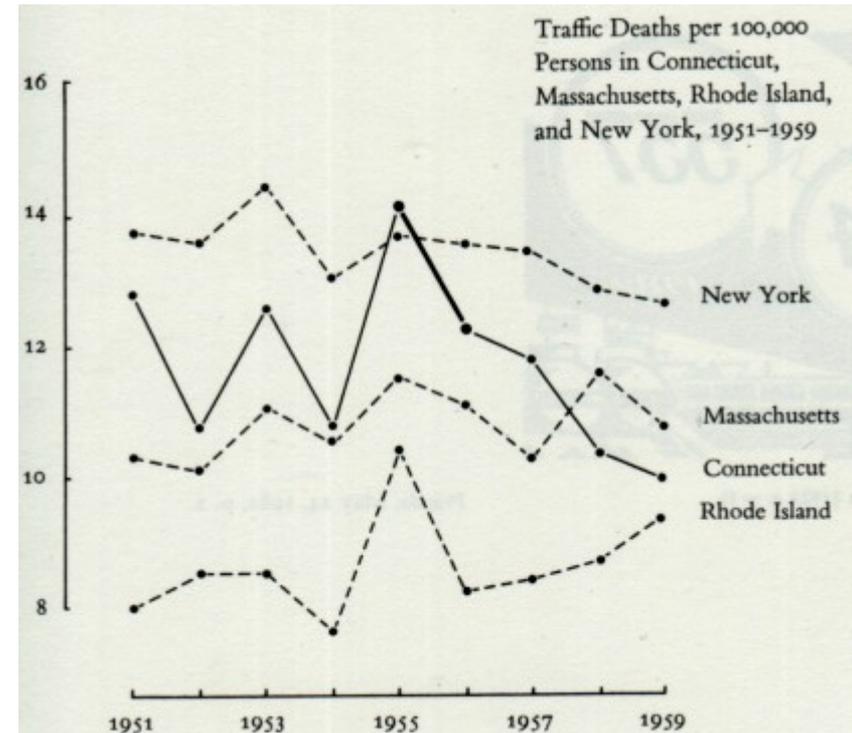
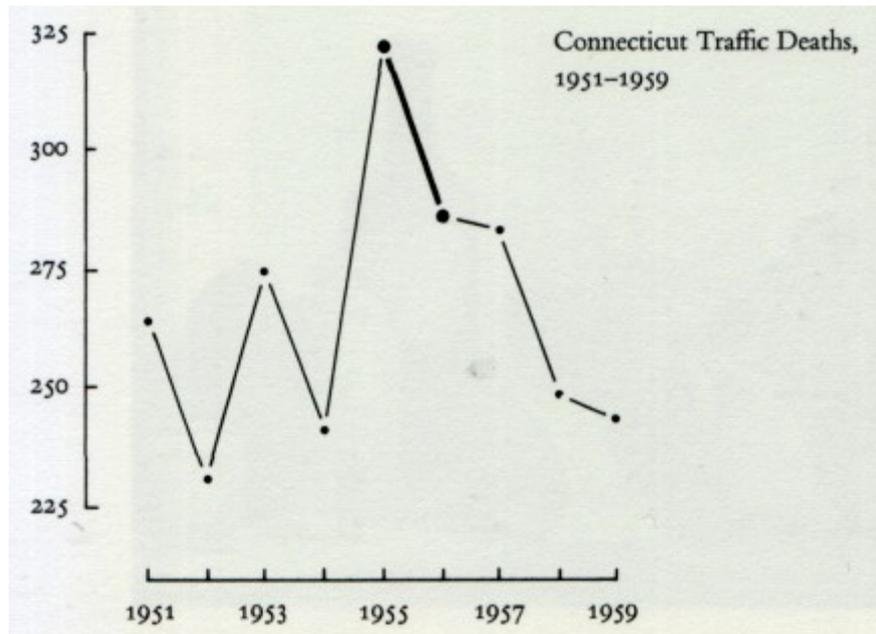
# Algumas recomendações de Tufte

- *Contextualize*



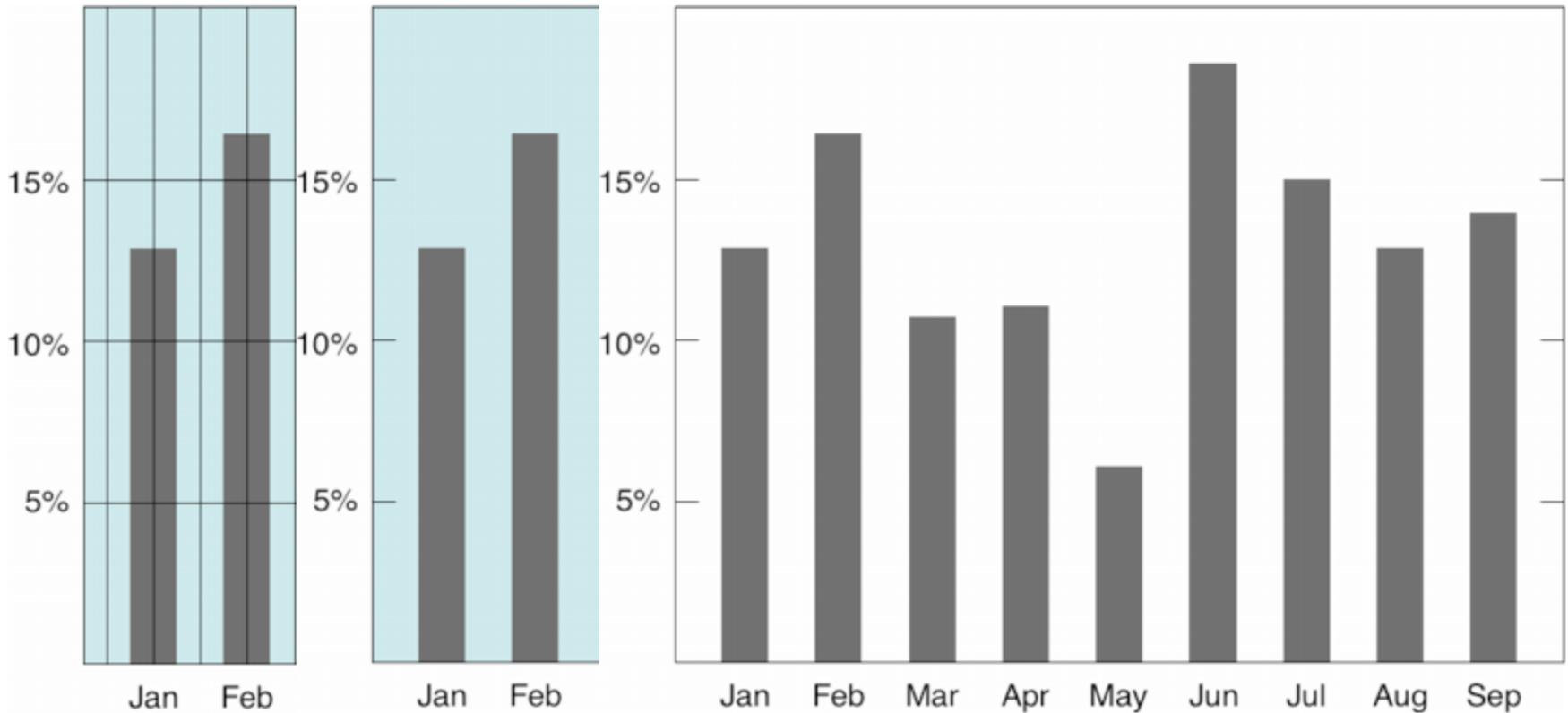
# Algumas recomendações de Tufte

- *Contextualize*



# Algumas recomendações de Tufte

- Elementos visuais desnecessários ofuscam os dados



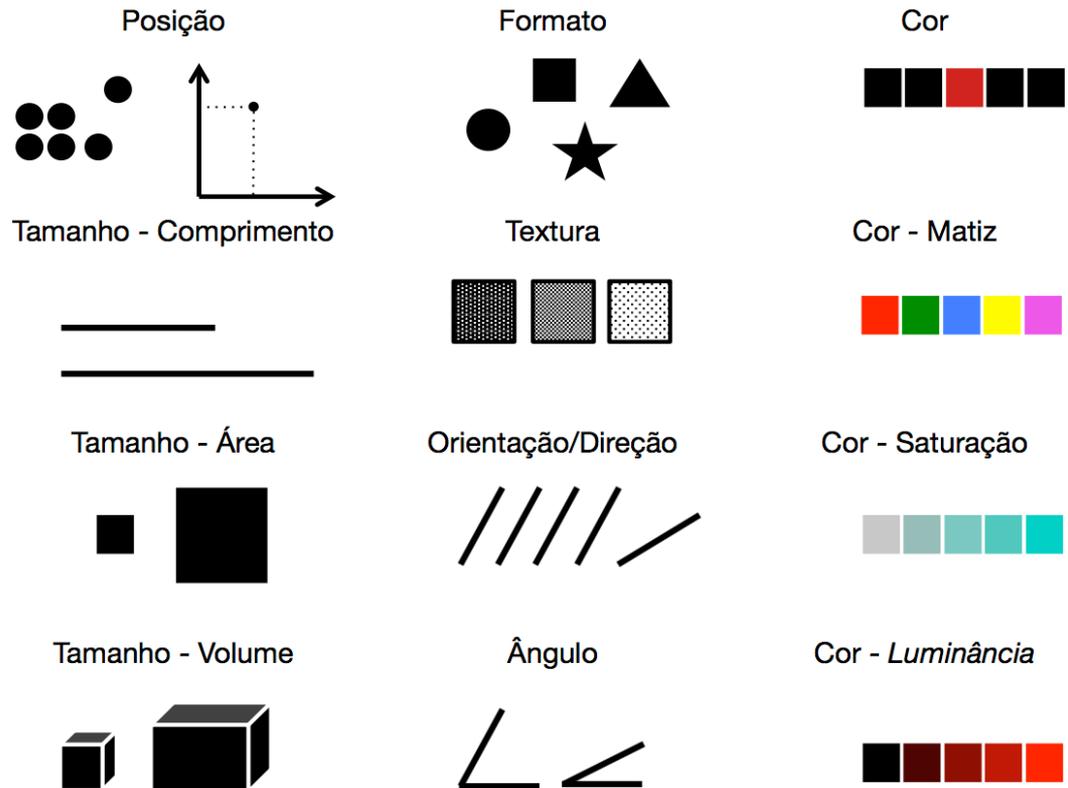
# Algumas recomendações de Tufte

- Elementos visuais desnecessários ofuscam os dados



# Atributos Visuais

- Variáveis visuais baseadas em características perceptuais



Adaptado de (Sultanum, 2015)

# Tipos de Dados

Modelo	Tipo	CV	Preço	Airbag
Volkswagen Fox	Pequeno	81	9100	0 (sem airbag)
Hyundai Elantra	Pequeno	124	10000	0 (sem airbag)
Pontiac Firebird	Esporte	160	17000	2 (mot. & pass.)
Ford Mustang	Esporte	105	15900	1 (motorista)

## Quantitativos

## Ordinais

## Categóricos (ou nominais)

- Ordem definida
- Valor numérico, passível de operações matemáticas
- Discreto ou contínuo

*Altura, peso, comprimento, preço, número de ocorrências*

...

- Ordem definida
- Operações aritméticas não definidas, mas comparação é possível

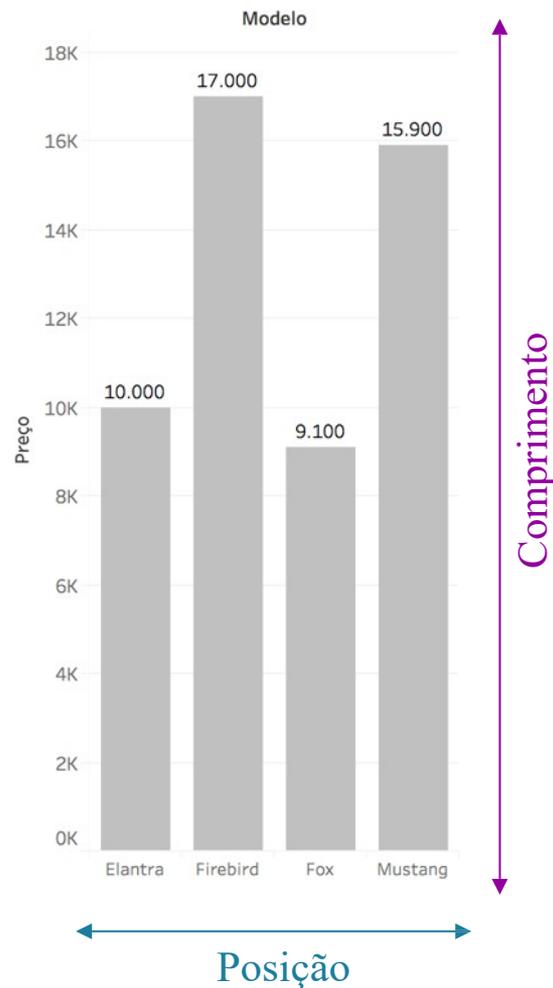
*Tamanho de camisas (P,M,G), rankings..*

- Sem ordem definida
- Possível dizer se igual ou diferente; pode ter hierarquia associada

*Nomes de cidades, gêneros...*

# Mapeamento Tipo de Dados vs. Variável Visual

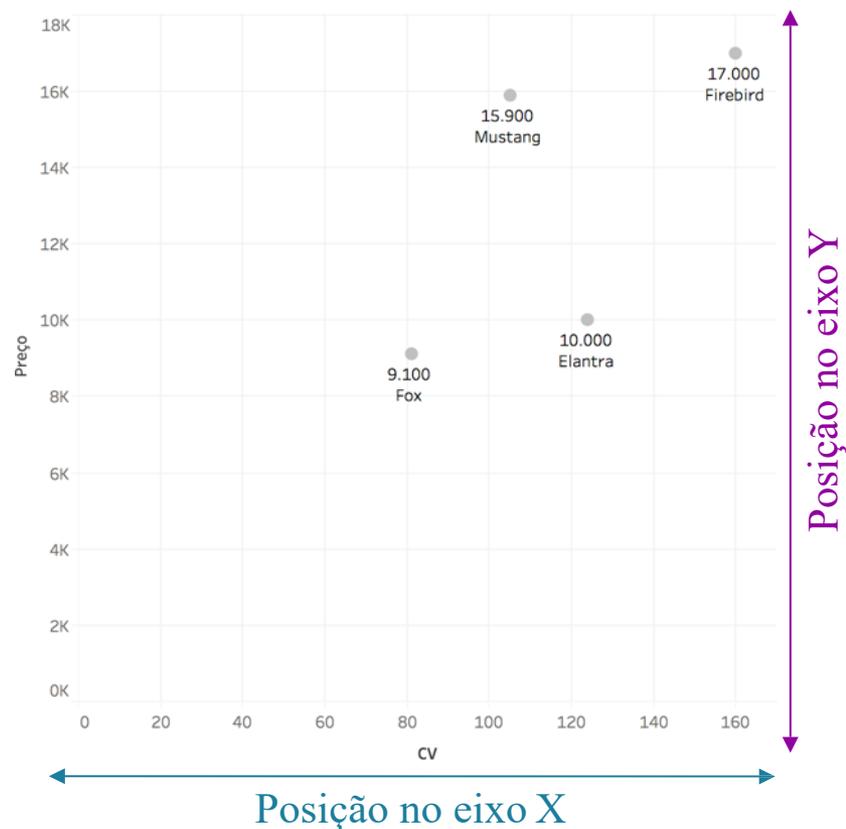
Modelo	Tipo	CV	Preço	Airbag
Volkswagen Fox	Pequeno	81	9100	0 (sem airbag)
Hyundai Elantra	Pequeno	124	10000	0 (sem airbag)
Pontiac Firebird	Esporte	160	17000	2 (mot. & pass.)
Ford Mustang	Esporte	105	15900	1 (motorista)



Adaptado de (Sultanum, 2015)

# Mapeamento Tipo de Dados vs. Variável Visual

Modelo	Tipo	CV	Preço	Airbag
Volkswagen Fox	Pequeno	81	9100	0 (sem airbag)
Hyundai Elantra	Pequeno	124	10000	0 (sem airbag)
Pontiac Firebird	Esporte	160	17000	2 (mot. & pass.)
Ford Mustang	Esporte	105	15900	1 (motorista)



Adaptado de (Sultanum, 2015)

# Mapeamento Tipo de Dados vs. Variável Visual

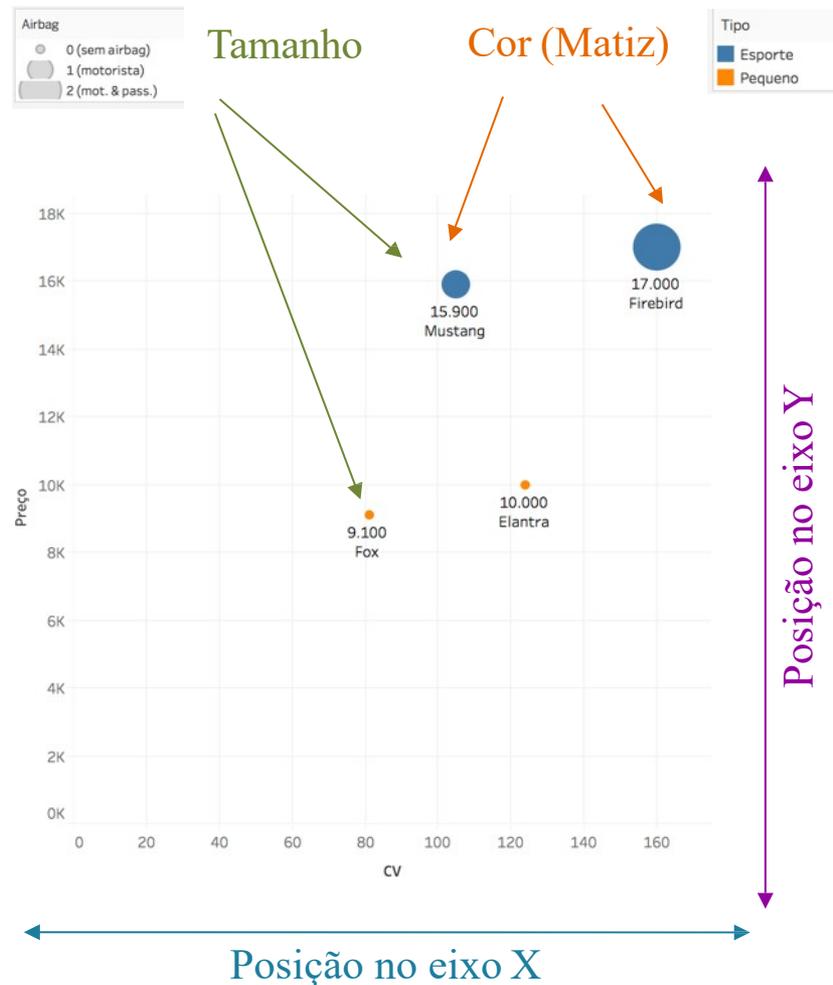
Modelo	Tipo	CV	Preço	Airbag
Volkswagen Fox	Pequeno	81	9100	0 (sem airbag)
Hyundai Elantra	Pequeno	124	10000	0 (sem airbag)
Pontiac Firebird	Esporte	160	17000	2 (mot. & pass.)
Ford Mustang	Esporte	105	15900	1 (motorista)



Adaptado de (Sultanum, 2015)

# Mapeamento Tipo de Dados vs. Variável Visual

Modelo	Tipo	CV	Preço	Airbag
Volkswagen Fox	Pequeno	81	9100	0 (sem airbag)
Hyundai Elantra	Pequeno	124	10000	0 (sem airbag)
Pontiac Firebird	Esporte	160	17000	2 (mot. & pass.)
Ford Mustang	Esporte	105	15900	1 (motorista)



Adaptado de (Sultanum, 2015)

# Tenha cautela na associação das variáveis visuais!

- Há dois princípios que uma visualização deve buscar atender:

- **Princípio da Expressividade**

- Representações devem expressar toda, e somente, a informação dos atributos associados

- Dados ordenados devem ser mostrados de forma a nosso sistema perceptual reconhecê-los intrinsecamente como ordenado;
- Dados não-ordenados não devem ser codificados a dar impressão de ordem

Quais representações posso usar para codificar os dados?

- **Princípio da Eficácia**

- Atributos importantes devem ser representados com as codificações mais importantes, de forma a se destacarem;
- Atributos de menor importância podem ser associados a codificações menos eficazes

Que representações são mais adequadas para garantir que estou passando a mensagem correta?

Mackinlay, Jock. "Automating the design of graphical presentations of relational information." ACM Transactions on Graphics (TOG) 5.2 (1986): 110-141.

# Tipos de Dados e sua expressividade

- Escalas de Medidas \*

\* Stevens, S.S. (1946). "On the Theory of Scales of Measurement". *Science* **103** (2684): 677–680.

- Dados Nominais (👎 ordem, 👎 distância)
  - Gênero, nacionalidade
- Dados Ordinais (👍 ordem, 👎 distância)
  - Avaliação: ruim, regular, bom, excelente
- Dados Intervalares (👍 ordem, 👍 distância, 👎 proporções, 👎 zero absoluto)
  - Temperatura, tempo
- Dados Razão (👍 ordem, 👍 distância, 👍 proporções, 👍 zero absoluto)
  - Distâncias, duração

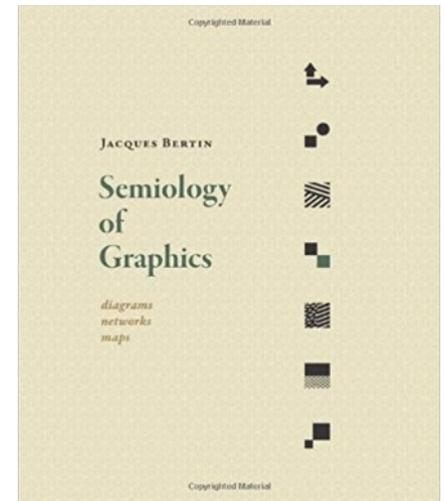
	Ordinal	Invervalar		Razão	Norminal
Name	Birth date	School grade	Address	Distance to school, m	Getting to school
Peter	17/05/2005	3	12, Pine street	850	by bus
Julia	23/08/2004	4	9, Oak avenue	400	on foot
Paul	10/12/2005	2	56, Maple road	1500	by car
Mary	06/10/2003	5	71, Linden lane	900	on foot

Adaptado de (Andrienko e Andrienko, 2017)

# Escolhendo a Variável Visual

- **Semiologia Gráfica**

- Conjunto de regras que permite o melhor uso de elementos visuais para a elaboração mais eficaz de gráficos, diagramas e mapas
- Concebida baseando-se no processo de percepção do sistema visual humano
- O trabalho seminal é de **Jacques Bertin**
  - “Semiologia gráfica: gráficos, diagramas e mapas”, 1962.



[https://images-na.ssl-images-amazon.com/images/I/518z2BqNTDL.\\_SX414\\_BO1,204,203,200\\_.jpg](https://images-na.ssl-images-amazon.com/images/I/518z2BqNTDL._SX414_BO1,204,203,200_.jpg)

# Semiologia Gráfica – Variáveis Visuais

- Para Jacques Bertin há **oito variáveis visuais**
  - **Dois** dimensões do plano (X e Y)
  - **Seis** variáveis retínicas: forma, tamanho, orientação, valor (brilho), granulação, cor
- **FORMA**
  - Pode ser geométrica (círculo, quadrado, triângulo) ou irregular
  - É ideal para diferenciar dados qualitativos (nominais)

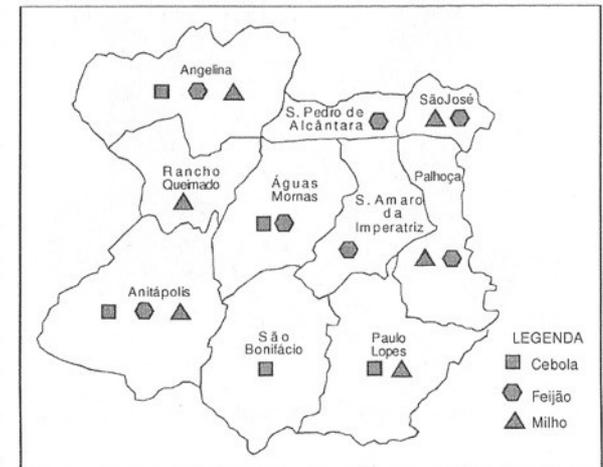


Figura 5.13 – Variável visual *forma* no modo de implantação pontual

# Semiologia Gráfica – Variáveis Visuais

- **TAMANHO**

- Indicada para representar dados quantitativos absolutos
- Esta variável visual transcreve a quantidade
- Número de habitantes, renda bruta, área cultivada, etc → são bons exemplos de representação pelo tamanho

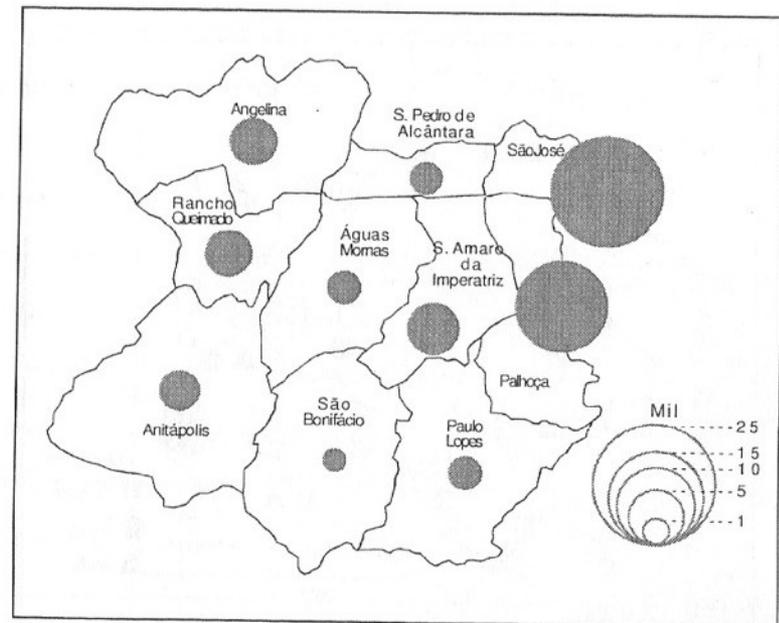


Figura 5.14 – Variável visual *tamanho* no modo de implantação pontual

# Semiologia Gráfica – Variáveis Visuais

- **ORIENTAÇÃO**

- Aplicada para linhas e formas alongadas
- Esta variável visual é utilizada para dados nominais

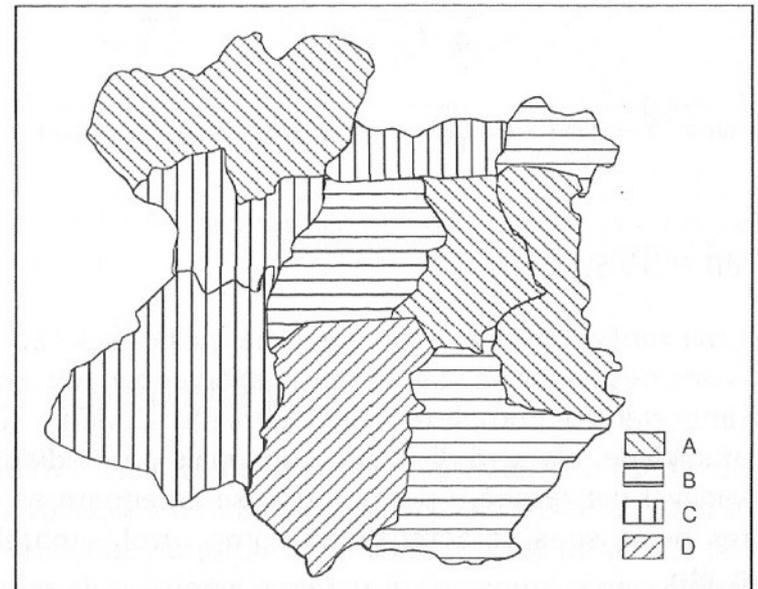
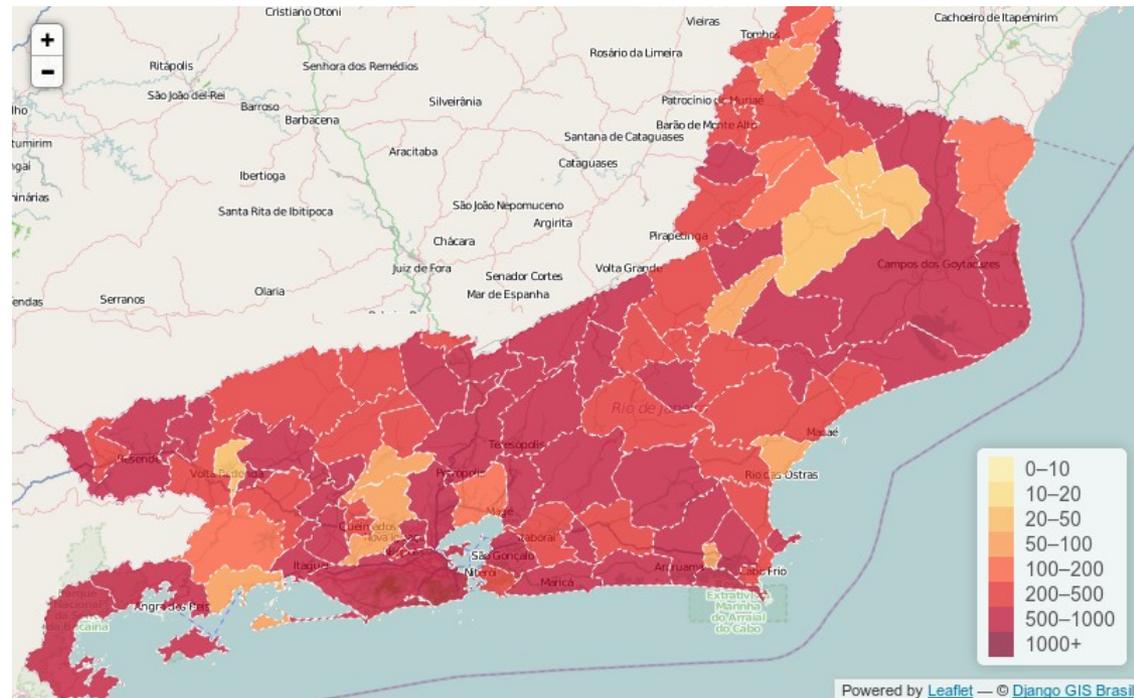


Figura 5.16 – Emprego da variável visual orientação

# Semiologia Gráfica – Variáveis Visuais

- **GRANULAÇÃO**

- Textura padronizada
- São utilizadas em mapas **coropléticos** (valores relativos)



# Semiologia Gráfica – Variáveis Visuais

- **COR**

- A cor possui três variáveis visuais:

- **MATIZ**

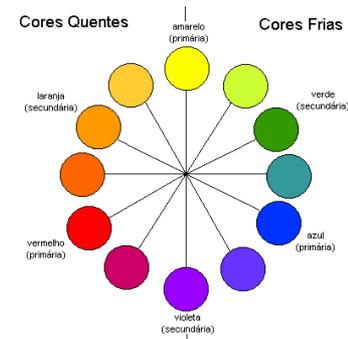
- Diferentes comprimentos de onda de luz visível do espectro eletromagnético
    - É uma variável visual seletiva (**dados nominais**)
    - Mas pode ser utilizada de forma **ordenada** se adotarmos o **círculo das cores**

- **SATURAÇÃO**

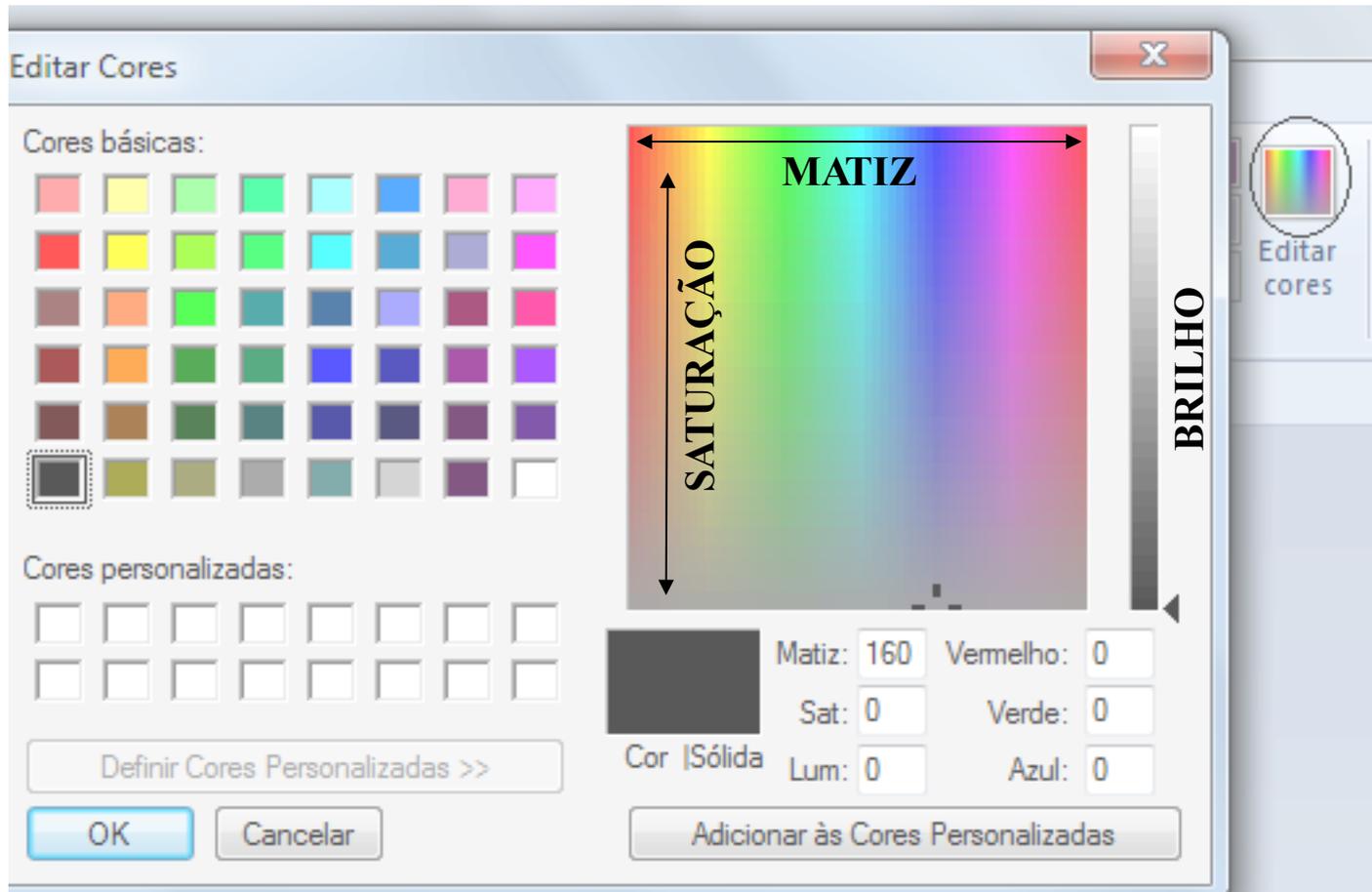
- Variação entre o cinza e a cor pura
    - Utilizada para valores ordenados

- **BRILHO**

- Contraste de luminosidade aplicado à cor, que varia do preto (ausência de brilho) ao branco (brilho total)
    - Permite ordenação



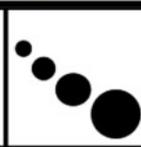
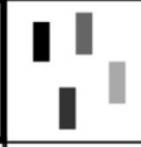
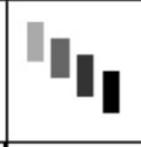
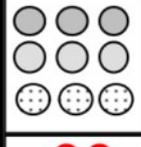
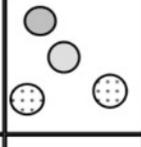
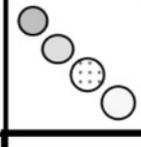
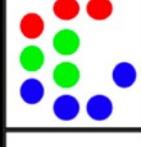
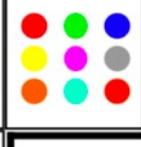
# Como?! Esclarecendo ....



# Classificação das Variáveis

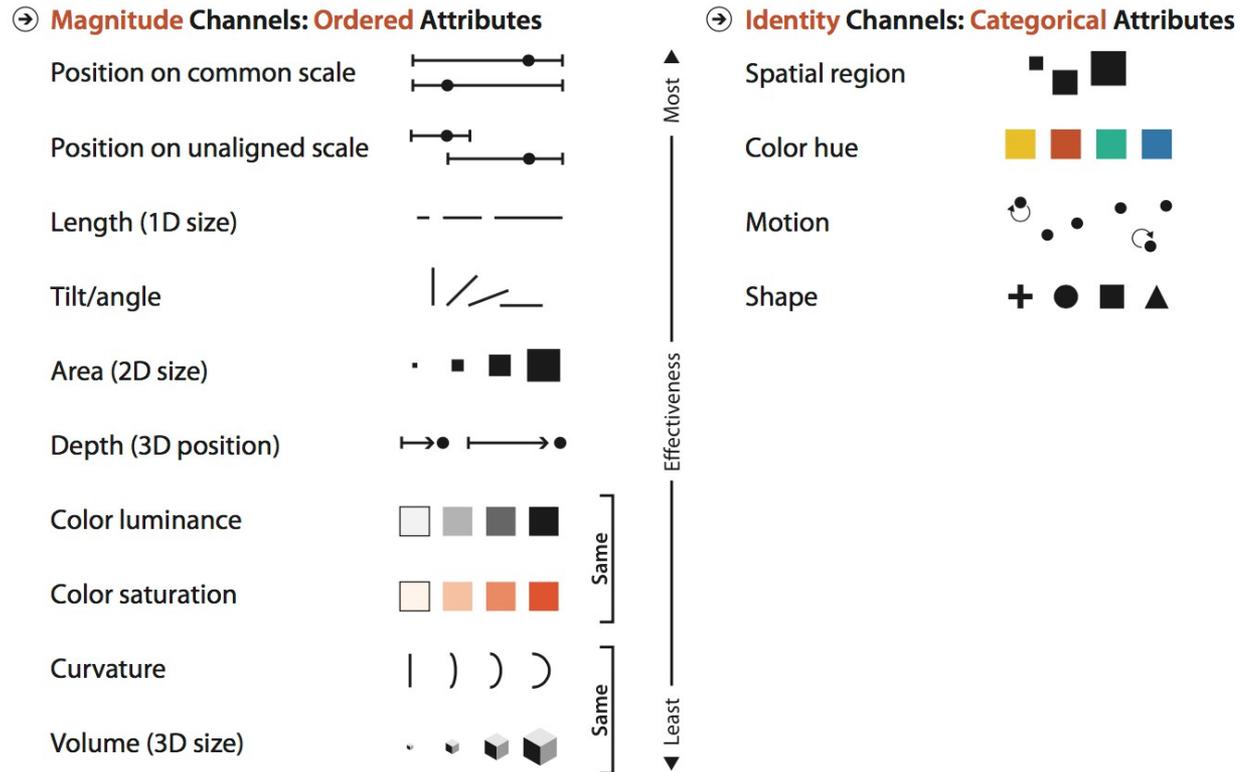
- Para Bertin, as variáveis são classificadas em:
- **Seletiva**
  - Permite diferenciar os elementos representados por elas;
  - Usada para representar dados no nível de medida nominal
- **Associativa**
  - Permite que os elementos representados por elas sejam percebidos como um grupo;
  - Usada para representar dados no nível de medida nominal
- **Ordenada**
  - Permite que os elementos por elas representados sejam ordenados
  - Representa dados no nível de medida ordinal, intervalar e razão
- **Quantitativa**
  - Transcreve quantidades, dando uma aproximação da ordem fornecida pelos dados
  - São utilizadas para representar os dados no nível de medida proporcional

# Classificação das Variáveis

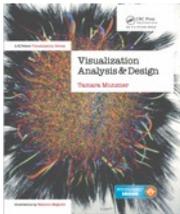
	Association	Selection	Order	Quantity
Size				
Value				
Texture				
Colour				
Orientation				
Shape				

# Trabalhos atuais

## Channels: Expressiveness Types and Effectiveness Ranks

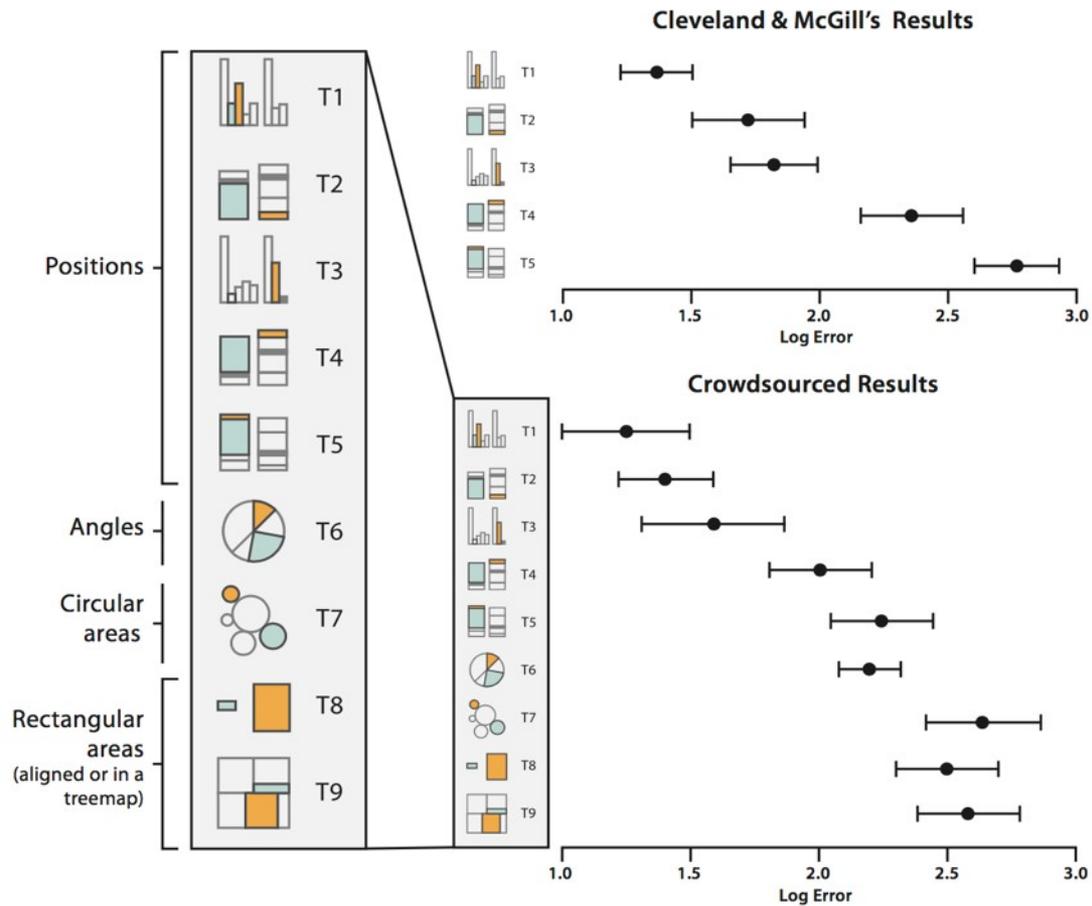


**Figure 5.6.** Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels.

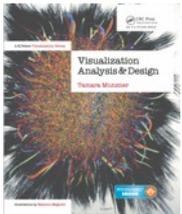


Munzner, Tamara. Visualization Analysis and Design. CRC Press, 2014.

# Trabalhos atuais



**Figure 5.8.** Error rates across visual channels, with recent crowdsourced results replicating and extending seminal work from Cleveland and McGill [Cleveland and McGill 84a]. After [Heer and Bostock 10, Figure 4].



Munzner, Tamara. Visualization Analysis and Design. CRC Press, 2014.

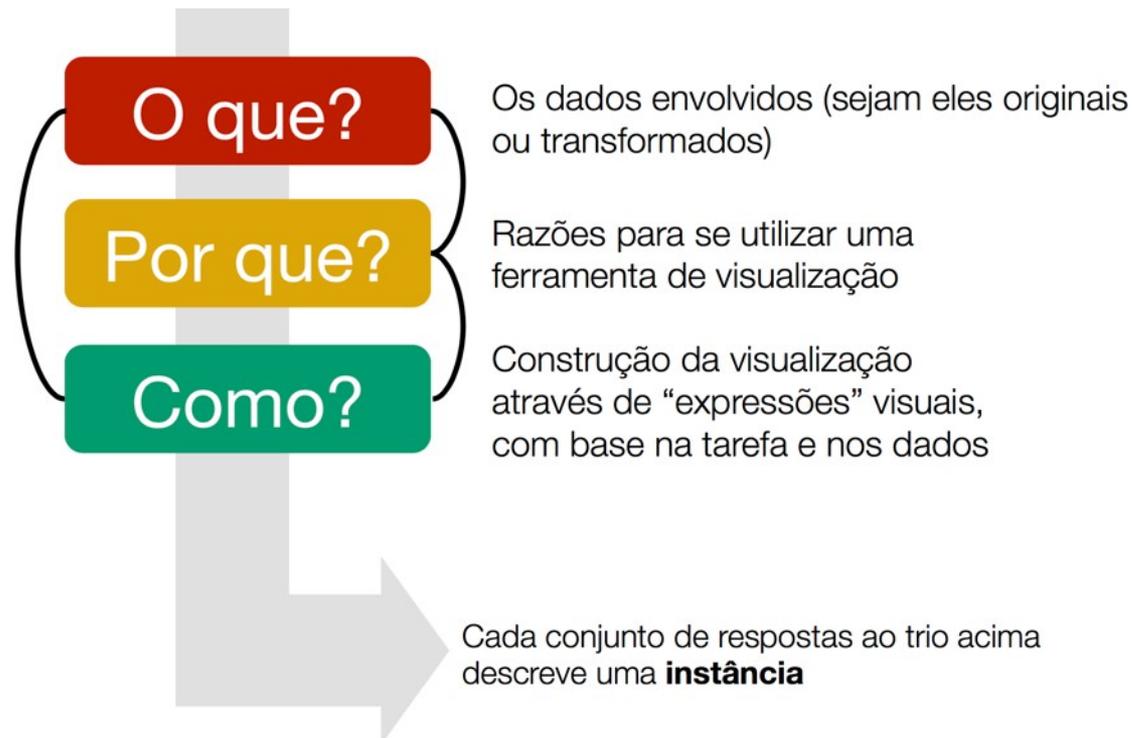
Cleveland, William S., and Robert McGill. "Graphical perception: Theory, experimentation, and application to the development of graphical methods." *Journal of the American statistical association* 79.387 (1984): 531-554.

Heer, Jeffrey, and Michael Bostock. "Crowdsourcing graphical perception: using mechanical turk to assess visualization design." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010.

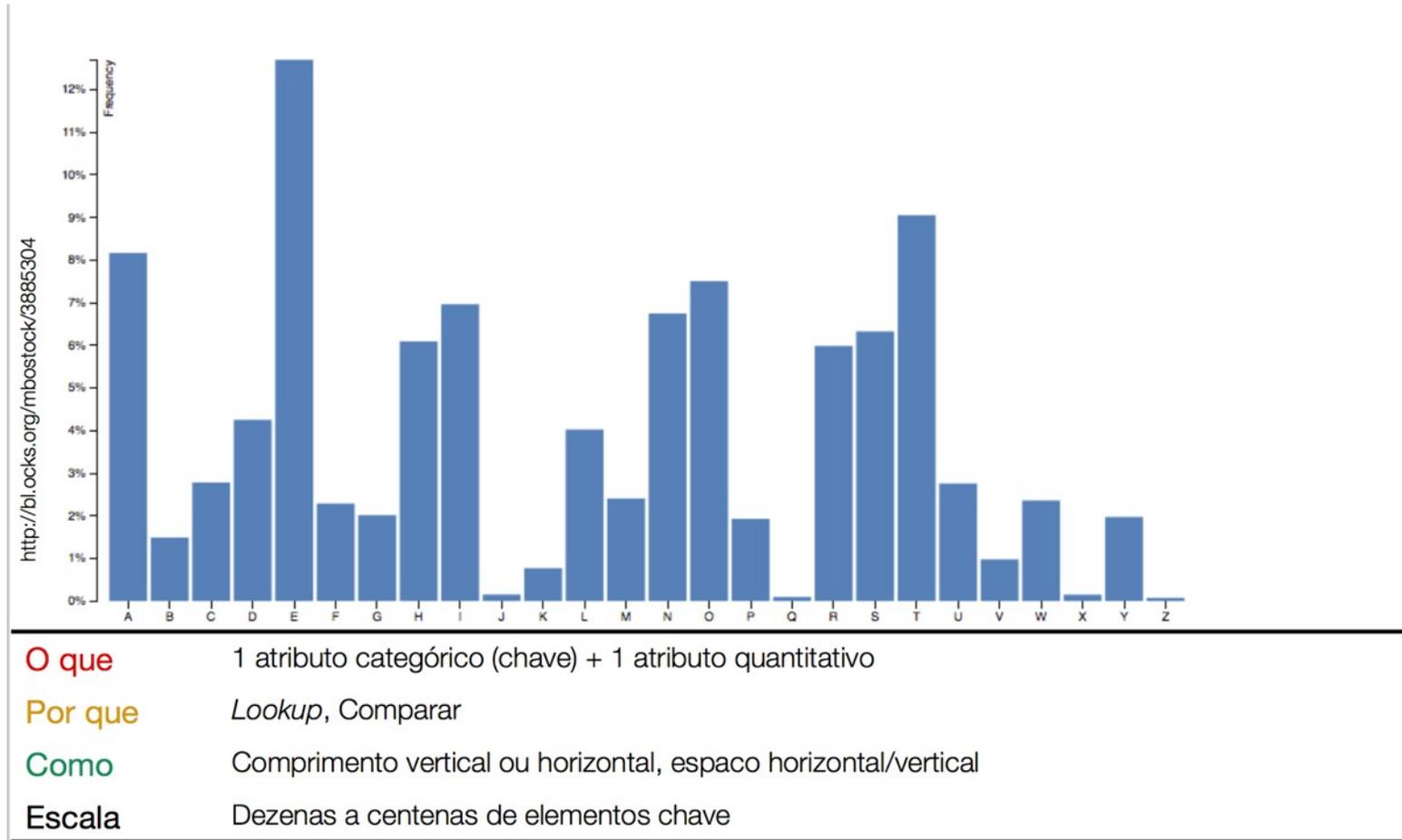


# Qual a melhor visualização para meus dados?

- Munzner (2014) apresenta uma proposta de aprender por exemplos
- Cada gráfico foi classificado em:

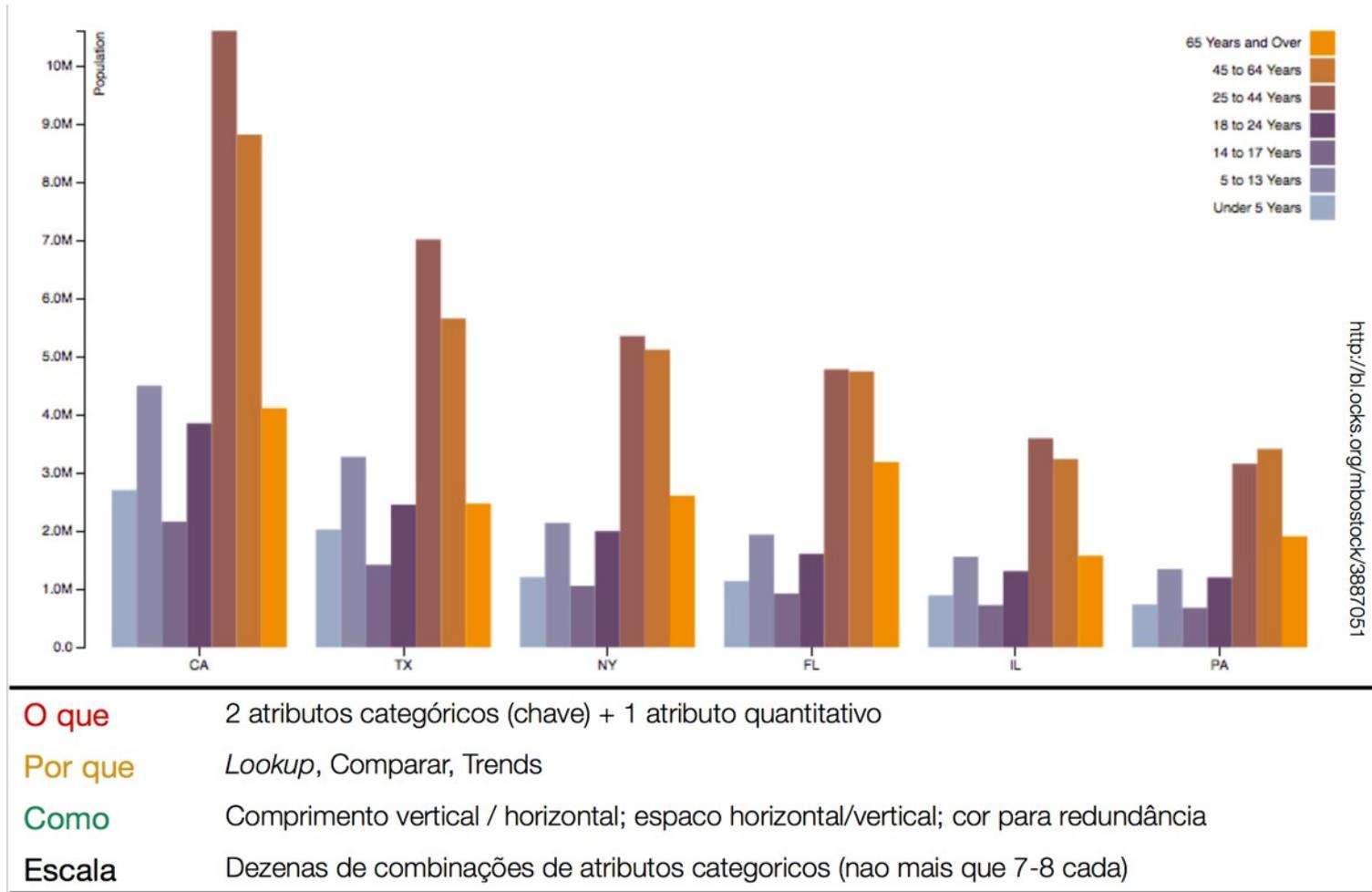


# Diagrama de Barras



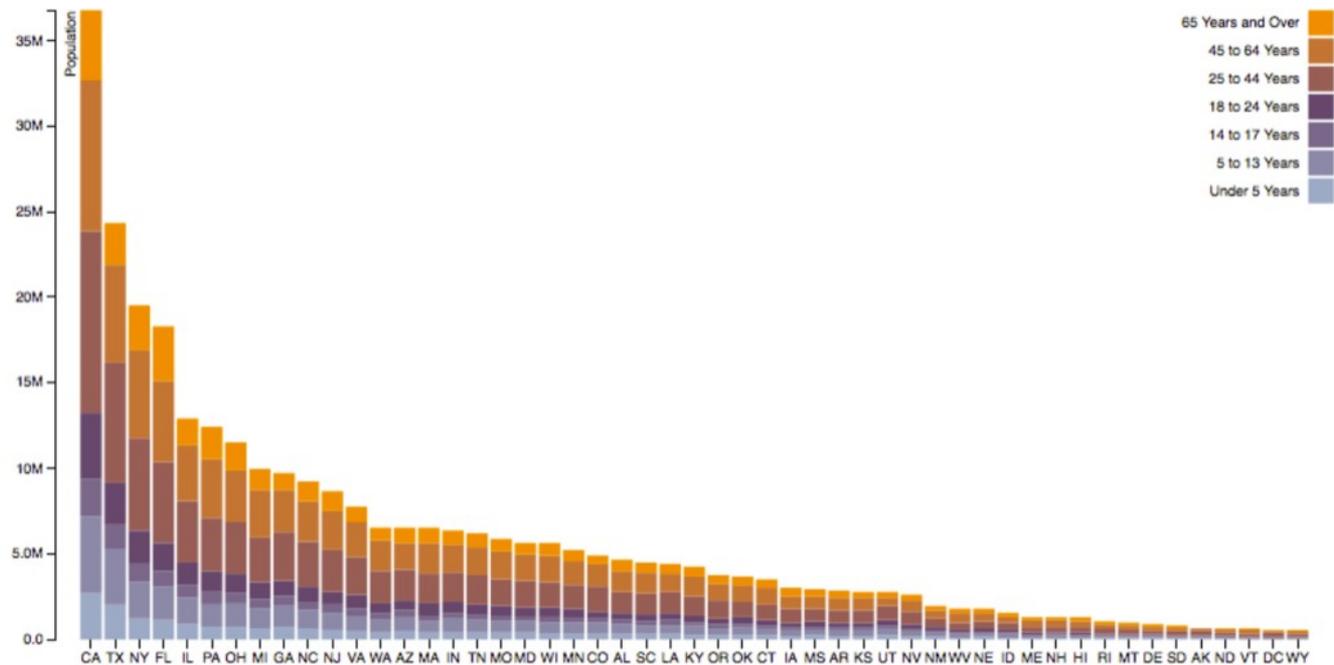
Adaptado de (Sultanum, 2015)

# Barras Agrupadas



Adaptado de (Sultanum, 2015)

# Barras em Pilha



<http://bl.ocks.org/mbostock/3886208>

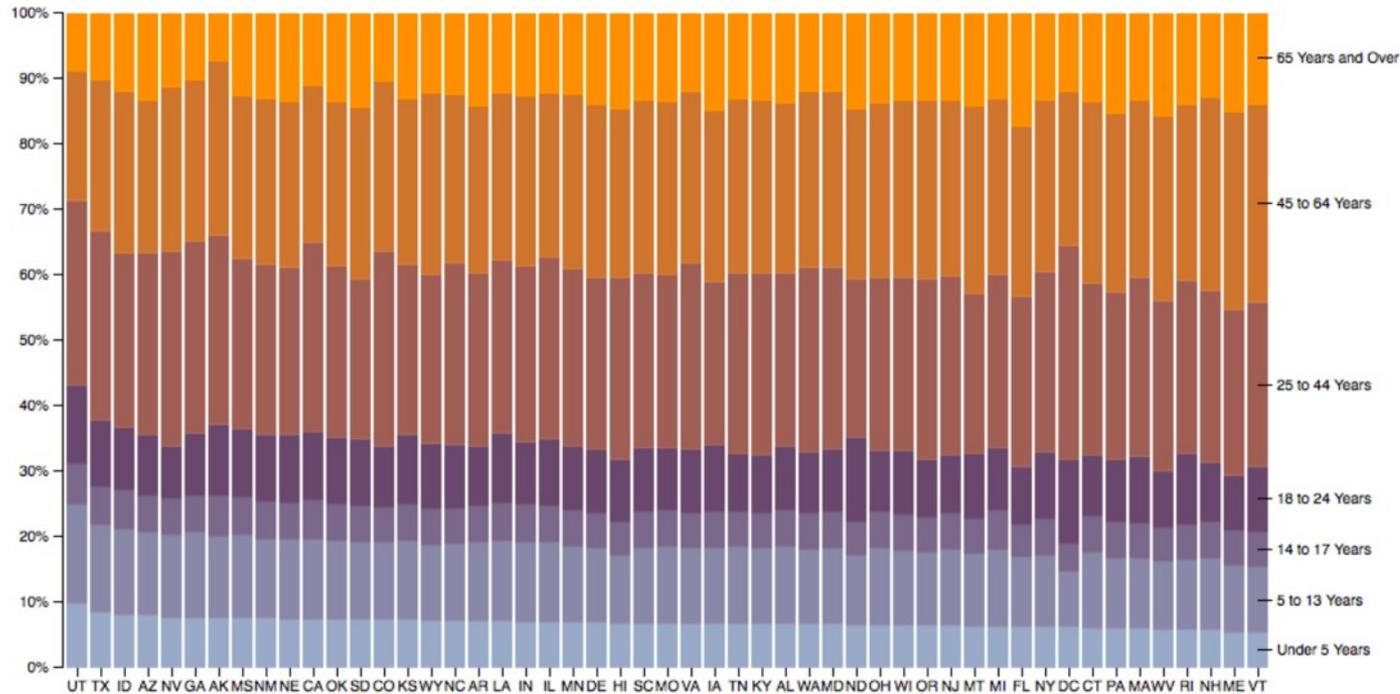
---

O que	2 atributos categóricos (chave) + 1 atributo quantitativo
Por que	Comparar: Partes-do-todo, Lookup, Trends
Como	Comprimento vertical/Horizontal
Escala	1ª chave: dezenas a centenas; 2ª chave: +- 10

---

Adaptado de (Sultanum, 2015)

# Diagrama de barras normalizado

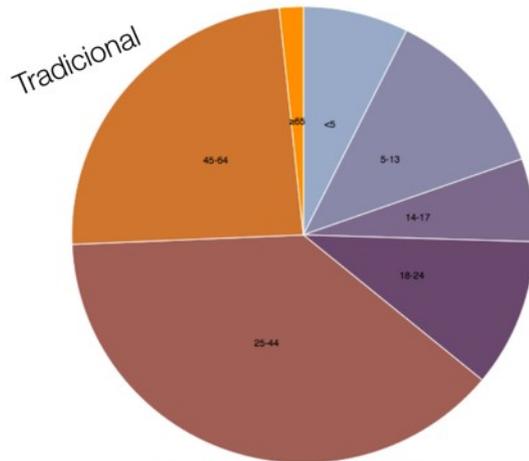


<http://bl.ocks.org/mbostock/3986394>

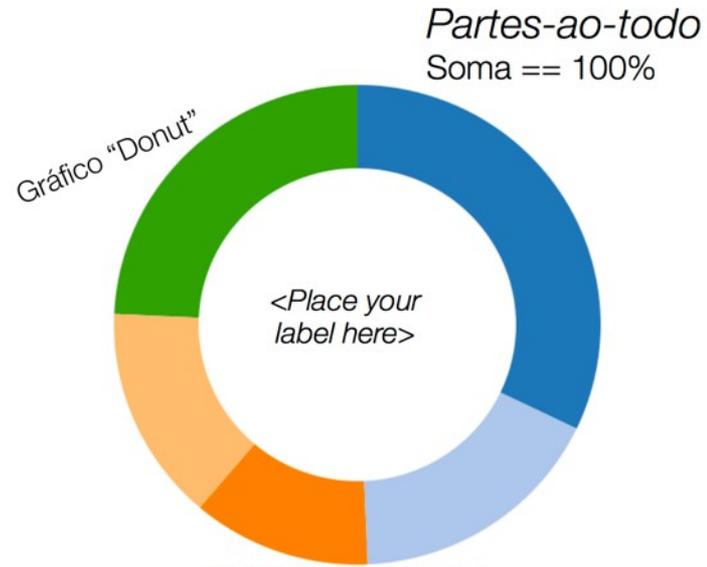
- 
- O que** 2 atributos categóricos, 1 atributo quantitativo (como %)
  - Por que** Partes-ao-todo
  - Como** Separação do espaço de barra entre categorias, codificando %.
  - Escala** ~12 categorias para a pilha; dezenas de categorias para o eixo
- 

Adaptado de (Sultanum, 2015)

# Gráficos Pizza e Donut



<http://bl.ocks.org/mbostock/3887235>



<http://bl.ocks.org/mbostock/5681974>

---

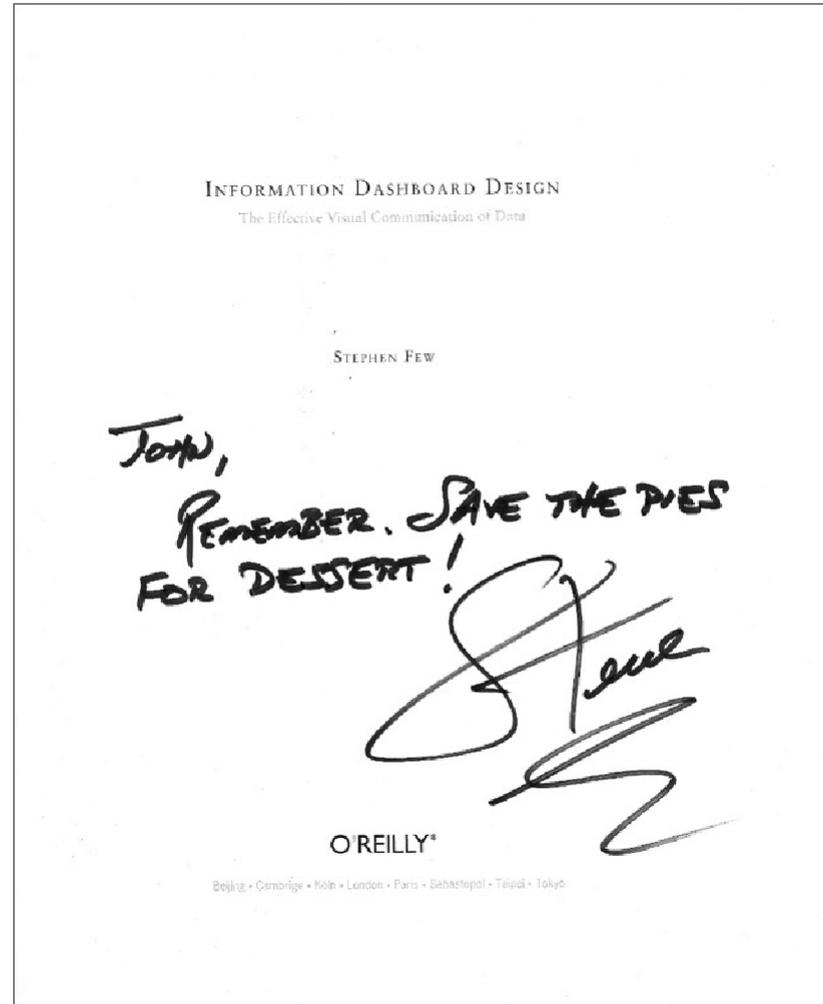
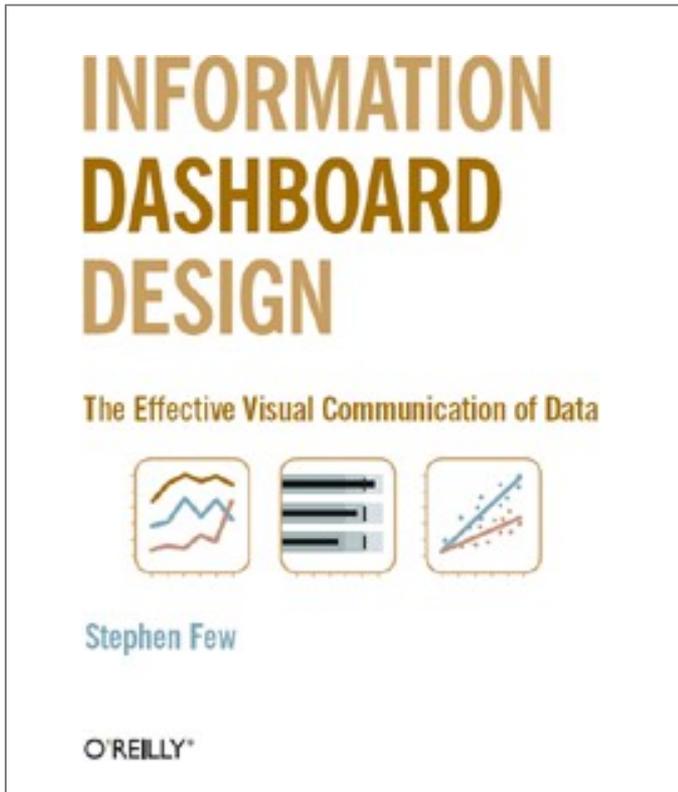
<b>O que</b>	1 atributo quantitativo, 1 categórico
<b>Por que</b>	Relação Partes-ao-todo
<b>Como</b>	Porcentagem por áreas angulares; <i>layout</i> radial; frequentemente cor para categorias
<b>Escala</b>	~12 categorias

---

Adaptado de (Sultanum, 2015)

# Gráfico de Pizza ... Cuidado!

Slide de John Rome, Deputy CIO, Arizona State University.



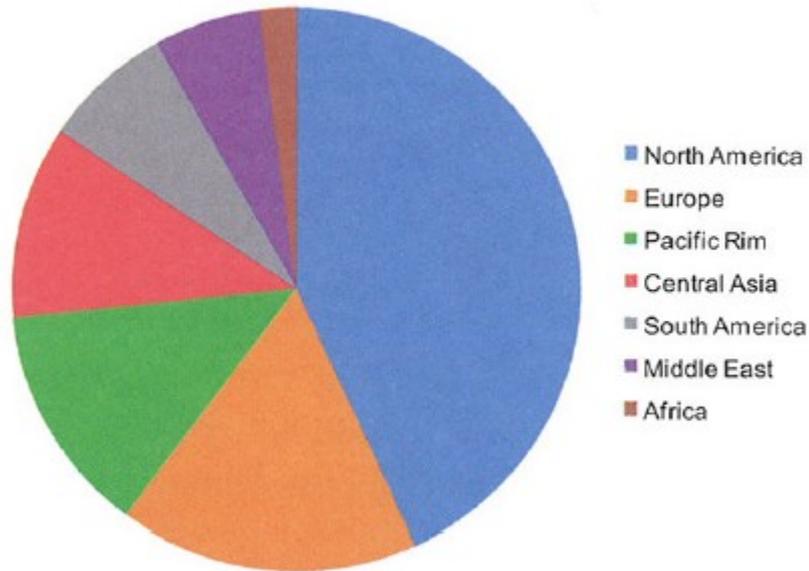
Leia: [http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/save\\_the\\_pies\\_for\\_dessert.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf)

# Gráfico de Pizza ... Cuidado!

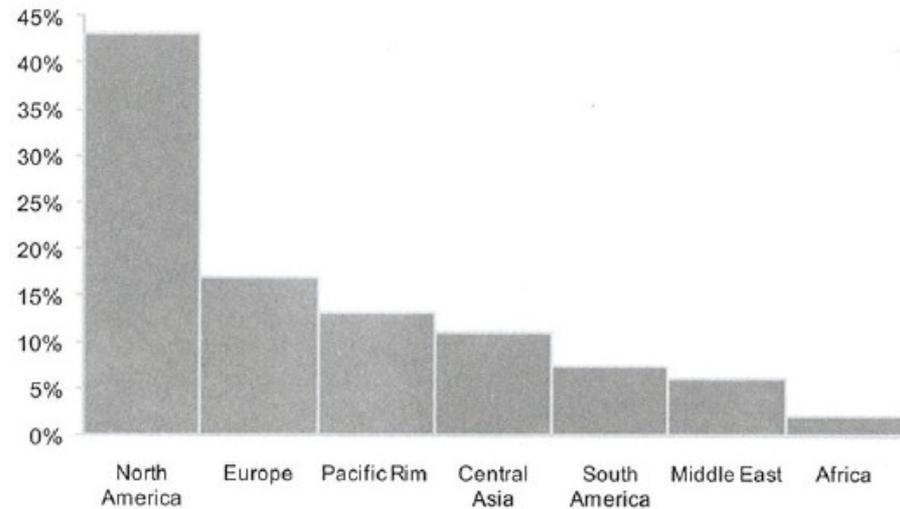
- Comparar seções da “pizza” é difícil
  - Comparação por ângulo é baixa no ranking perceptual
  - Difícil comparar valores similares
- Comparar vários gráficos entre si é ainda pior
- Não escala
  - Muitas categorias → seções menores → dificuldade de comparação
- Porém,
  - São fáceis de compreender
  - Podem ser utilizadas quando precisão não é um problema

# Gráfico de Pizza vs Gráfico de Barras

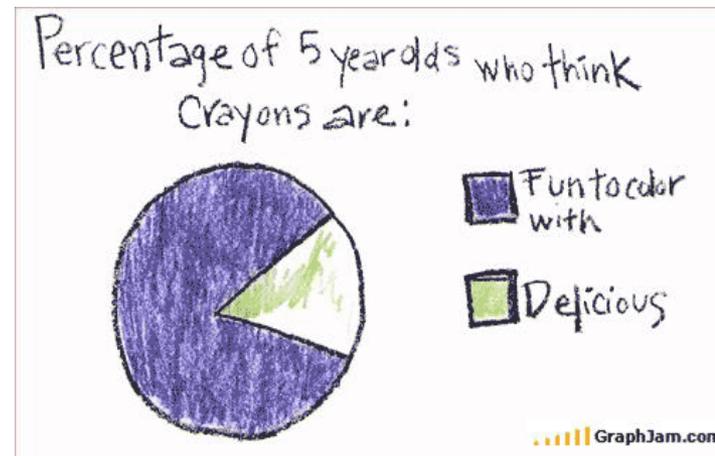
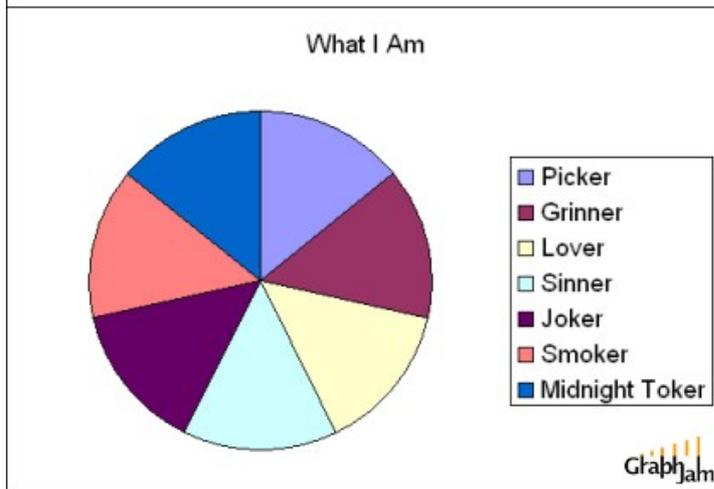
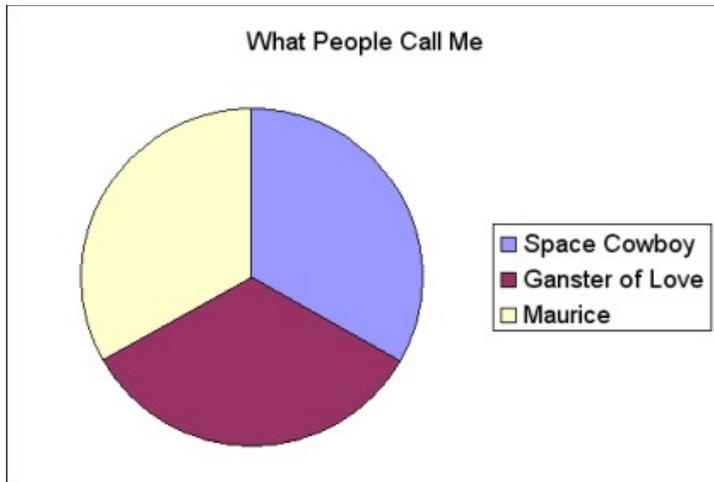
Total Sales by Region



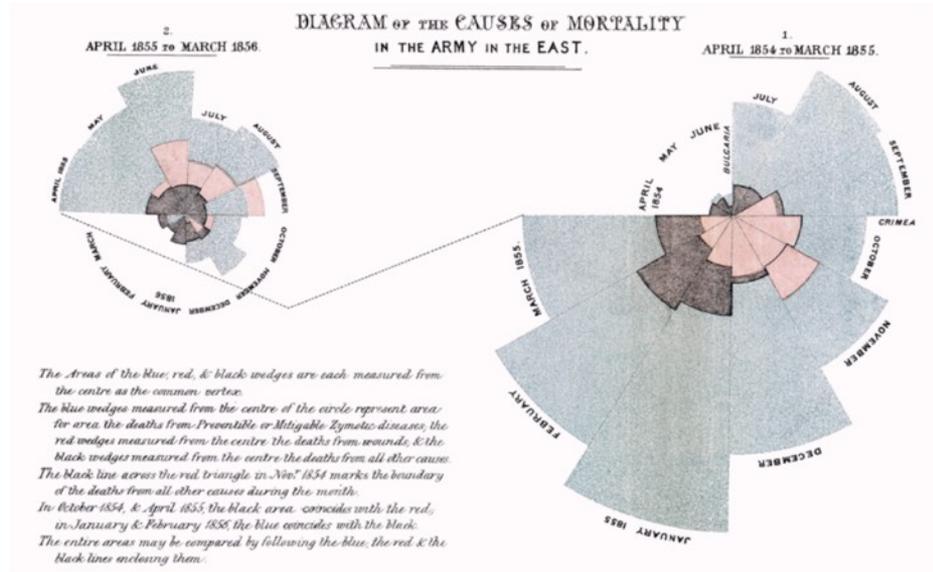
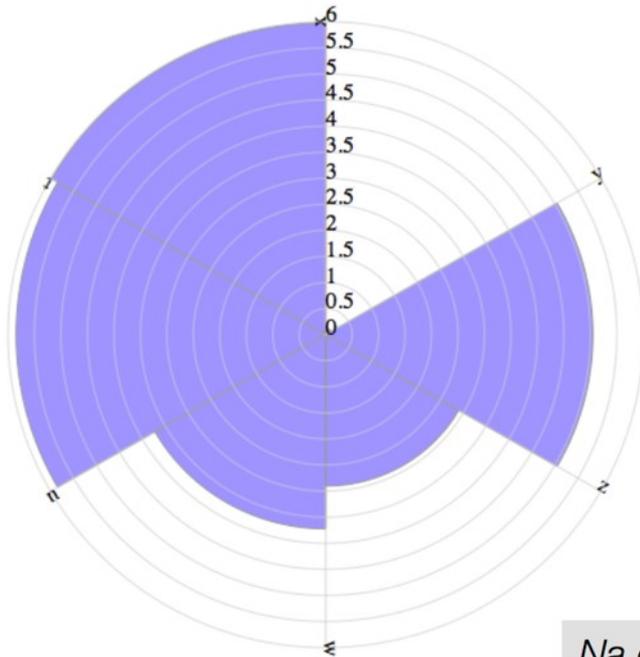
Total Sales by Region



# Gráfico de Pizza ... Mais exemplos



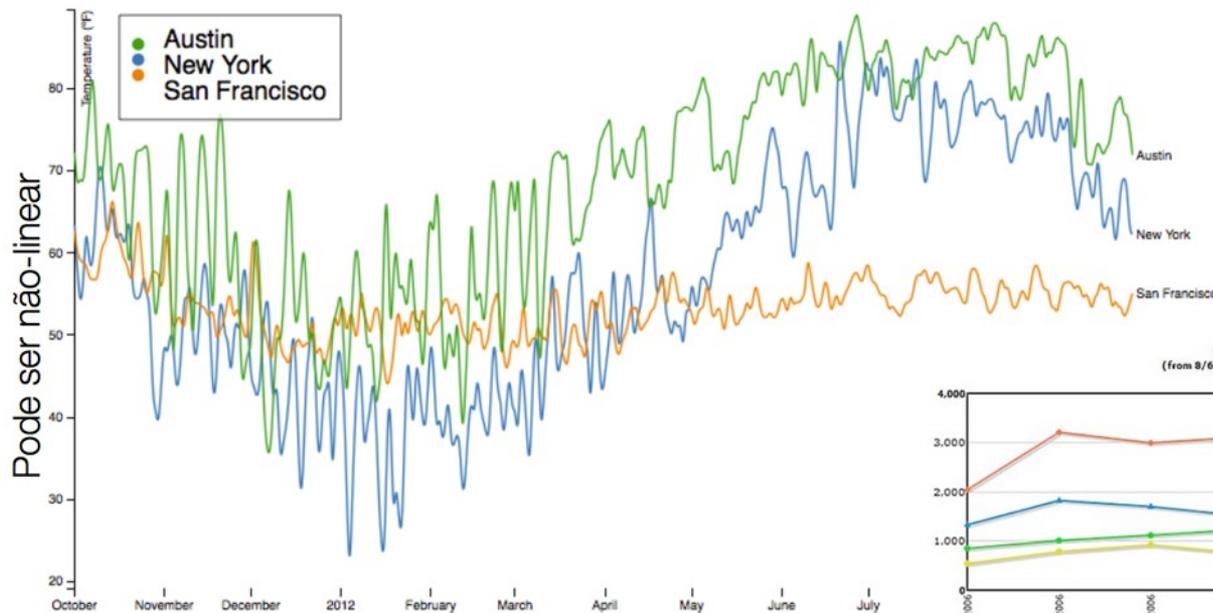
# Diagrama de Área Circular



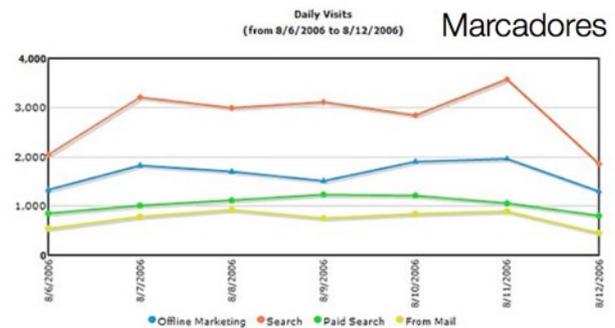
Na grande maioria das vezes, é melhor usar bar charts...

O que	1 atributo quantitativo, 1 atributo categórico
Por que	Partes-ao-todo, comparação
Como	Divisão do espaço radial entre categorias; comprimento da seção denota valor quantitativo
Escala	~12 categorias

# Gráfico de Linha



<http://bl.ocks.org/ZJONSSON/3918369>



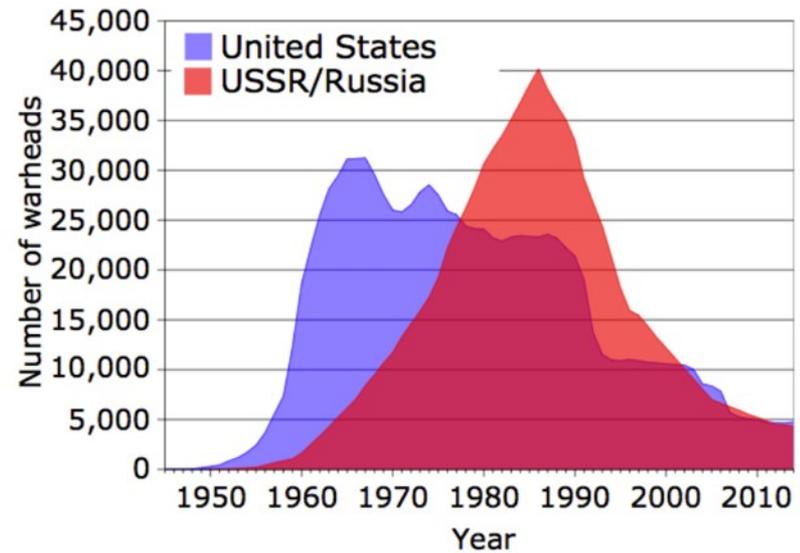
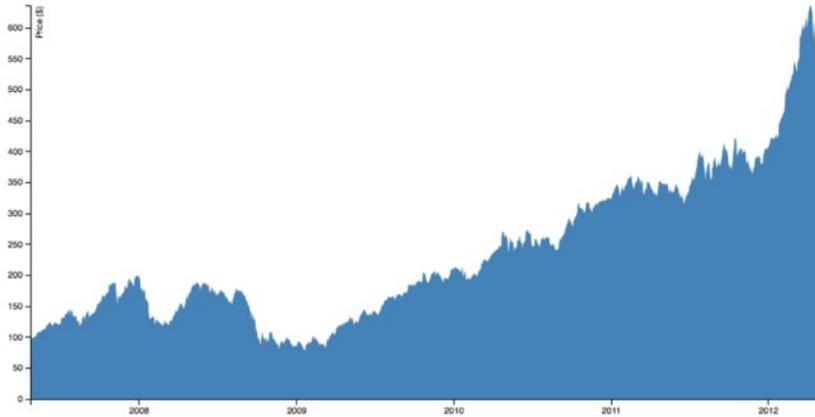
---

O que	1 atributo quantitativo, 1 atributo ordenado (chave), potencialmente 1 atributo categorico
Por que	Tendências, extremos, características, correlação
Como	Posição no plano XY, com linhas. Chave usualmente no eixo X. Cores/marcadores p cat.
Escala	Centenas de níveis para os atributos no eixo XY. ~12 categorias.

---

Adaptado de (Sultanum, 2015)

# Gráfico de Área



[http://en.wikipedia.org/wiki/Area\\_chart](http://en.wikipedia.org/wiki/Area_chart)

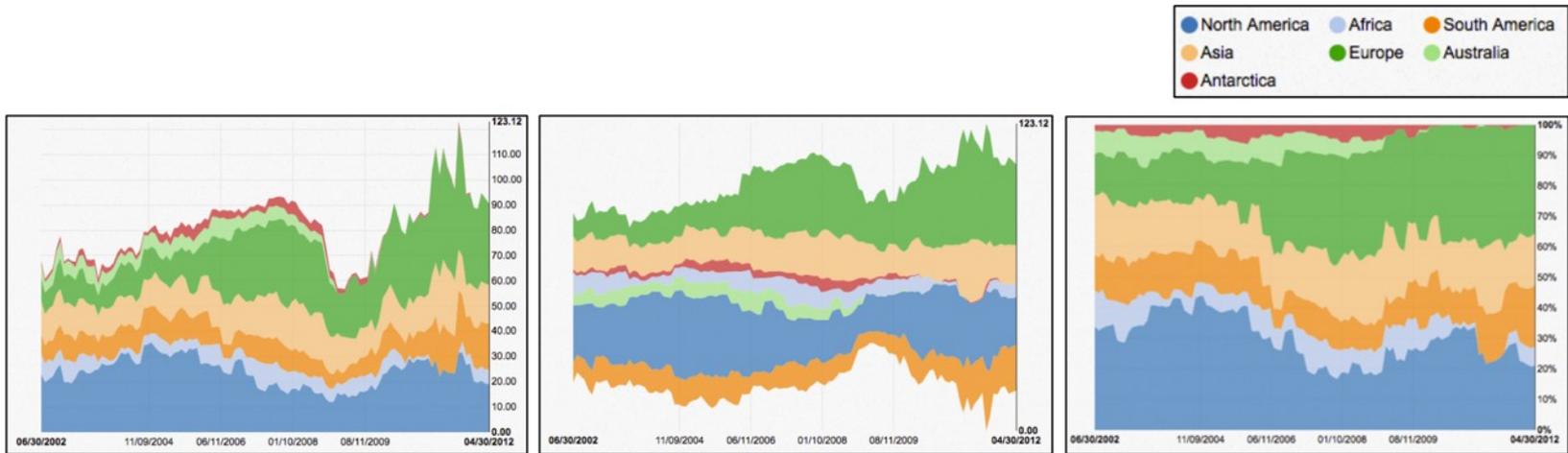
---

<b>O que</b>	1 atributo quantitativo, 1 atributo ordenado (chave)
<b>Por que</b>	Tendências, extremos, correlação
<b>Como</b>	Posição no plano X (ord) Y (quant), com linhas; área preenchida.
<b>Escala</b>	Centenas de níveis, ambos os atributos.

---

Adaptado de (Sultanum, 2015)

# Gráfico de Área em Pilha



<http://nv3.org/examples/stackedArea.html>

Gráfico de Área em pilha

Gráfico de Stream (*Streamgraph*)

Gráfico de Área Normalizado

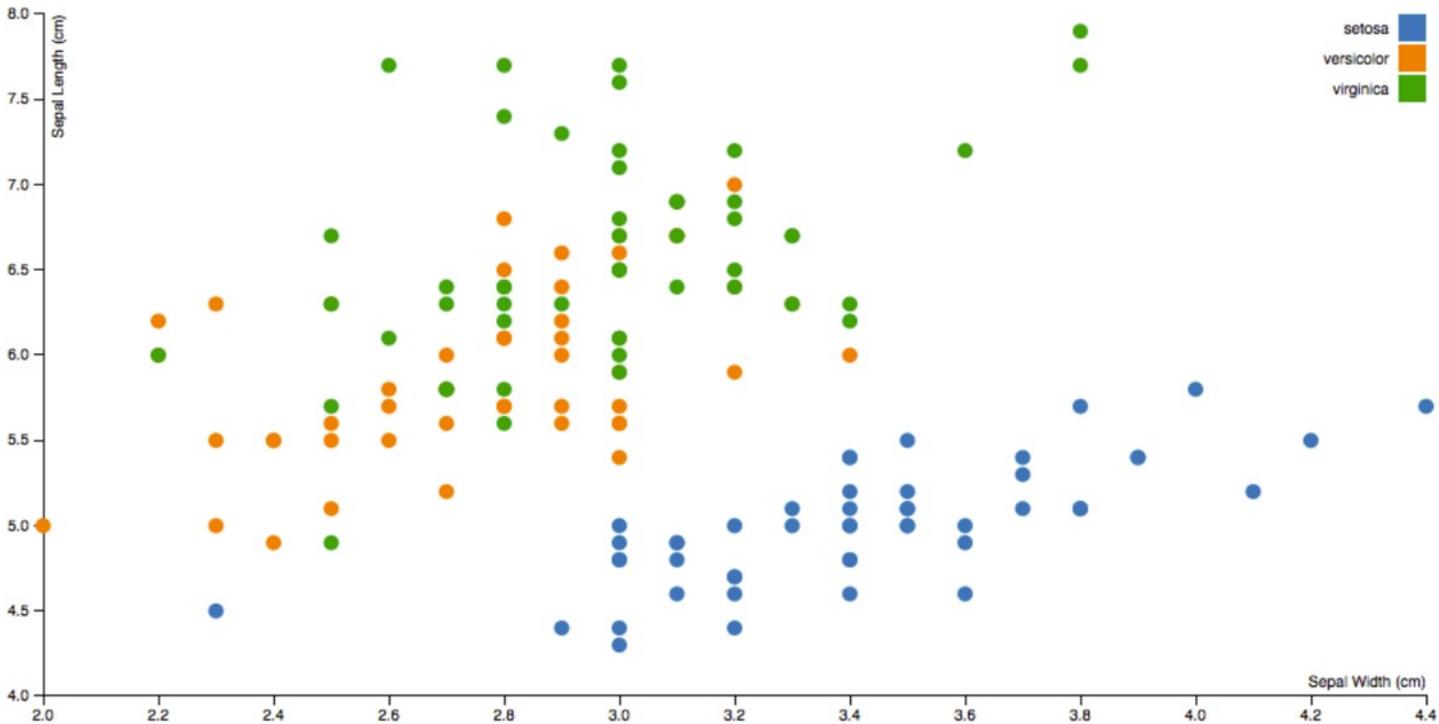
---

O que	1 atributo quantitativo, 1 atributo ordenado, 1 atributo categórico (+ 1 quant. p ordenação)
Por que	Partes-ao-todo, Tendências
Como	Organização do espaço vertical entre categorias, distancia por atr. quantitativo.
Escala	Centenas de pontos ao longo do eixo ordenado; ~12 categorias

---

Adaptado de (Sultanum, 2015)

# Gráfico de Dispersão



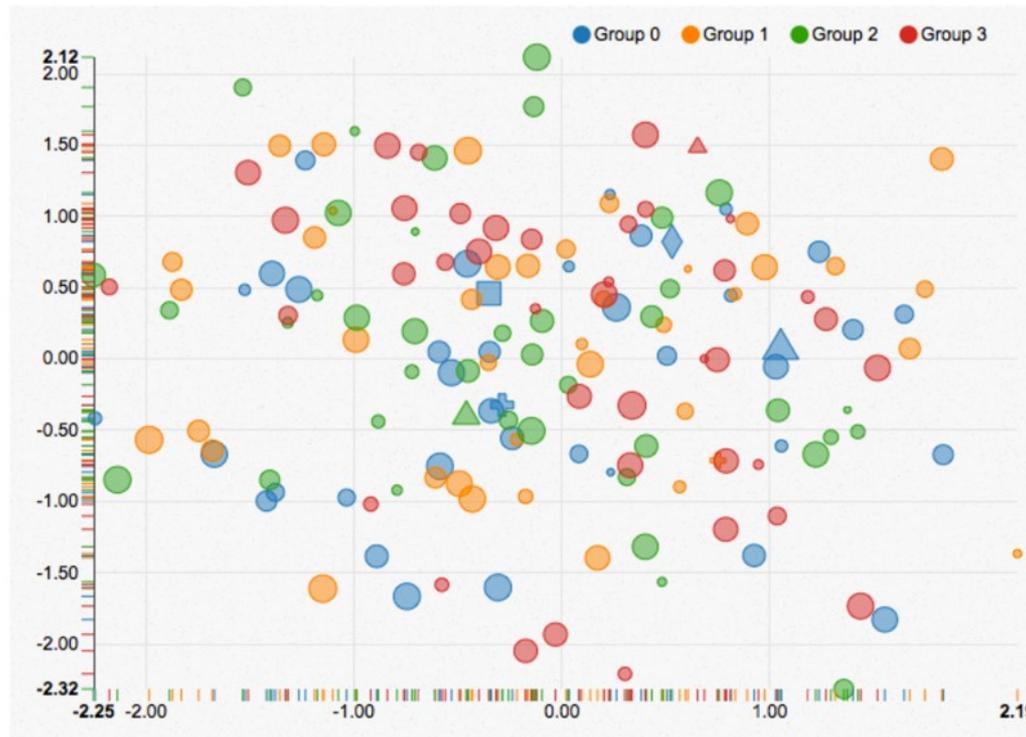
---

<b>O que</b>	2 atributos quantitativos (possivelmente +1 atributo categórico)
<b>Por que</b>	Tendências, extremos, distribuição, correlação, localizar clusters
<b>Como</b>	Posicao no eixo XY; cores/forma para atributo categórico
<b>Escala</b>	Centenas de itens

---

Adaptado de (Sultanum, 2015)

# Gráfico de Bolhas



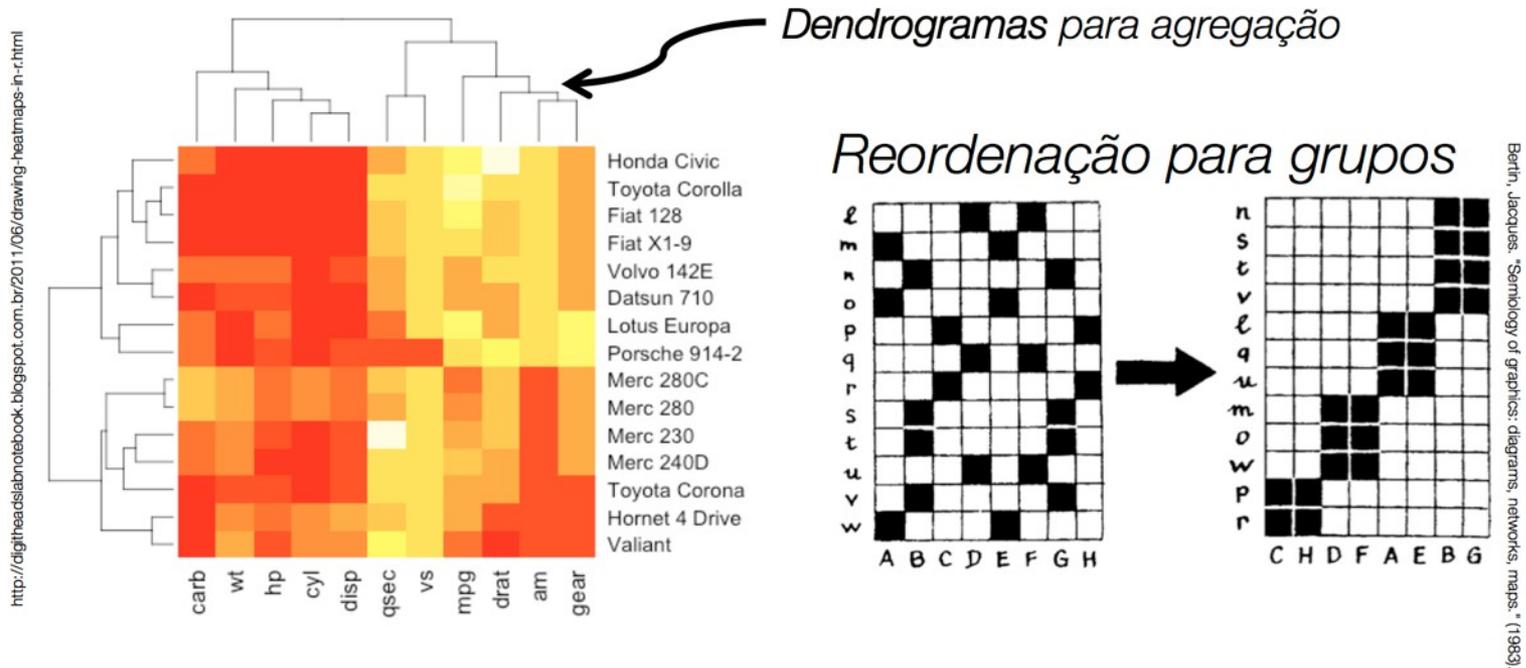
---

O que	2 atributos quantitativos(XY) + 1 atributo ordinal (possivelmente +2 atributos categóricos)
Por que	Tendências, extremos, distribuição, correlação, localizar clusters
Como	Posicao no eixo XY; atr. ordinal mapeado a tamanho do marcador; cor/forma atrs. cats.
Escala	Dezenas de itens

---

Adaptado de (Sultanum, 2015)

# Mapa de Calor Agrupado (Heatmap)



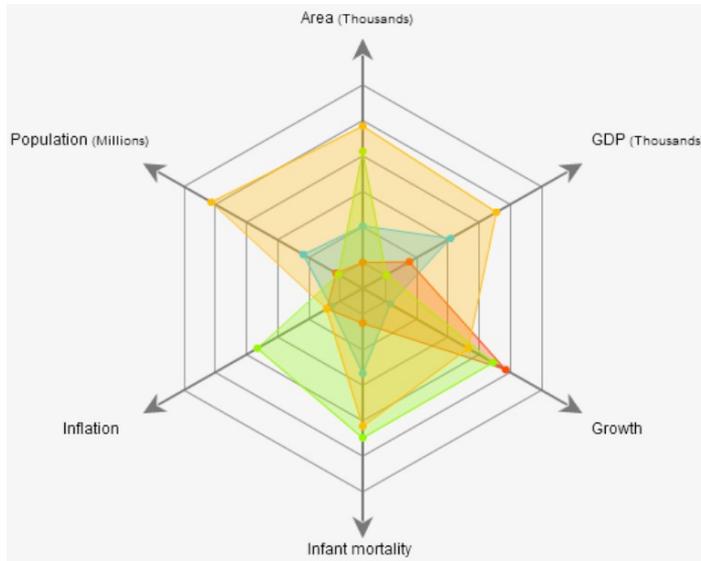
O que	2 atributos categóricos + 1 atributo quantitativo
Por que	Identificar <i>clusters</i> , extremos, resumir
Como	Espaço retangular dividido entre categorias; mapeamento de cores a atr. Quantitativo.
Escala	> Milhares itens. Centenas de atributos categoricos; 3-11 níveis p atr. quantitativo.

# Heatmap - Exemplo

Year (dimension)	Region (dimension)	Sales (measure)
2009	North	0
2009	South	10
2009	East	70
2009	West	2
2010	North	4
2010	South	22
2010	East	71
2010	West	6
2011	North	15
2011	South	68
2011	East	79
2011	West	11
2012	North	27
2012	South	80
2012	East	90
2012	West	50



# Diagrama Estrela (Radar)



<http://www.scottlogic.com/blog/2011/09/23/a-critique-of-radar-charts.html>

- Permite comparação de áreas...
  - Mas claro, não muito preciso.
- Não escala para muitos itens simultâneos
- Comparação entre eixos opostos é mais difícil
- Pode ser usado para caracterizar classes diferentes de elementos de acordo com seus atributos

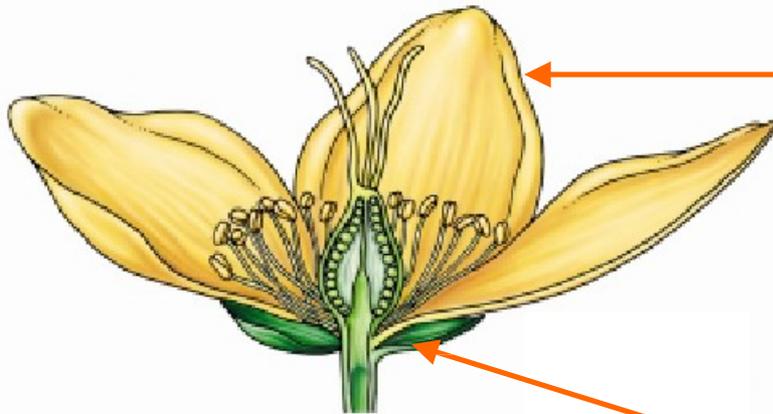
---

<b>O que</b>	Vários atributos (principalmente quantitativos)
<b>Por que</b>	Características, Comparação
<b>Como</b>	Espaço radial → eixo atributos; comprimento → valores; itens representados como linhas/área
<b>Escala</b>	Dezenas de eixos; dezenas de itens

---

Adaptado de (Sultanum, 2015)

# Coordenadas Paralelas



**Petála**, parte não-reprodutiva da flor

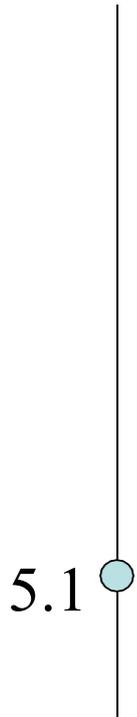
**Sepála**, parte não-reprodutiva da flor



**Dataset Iris**

# Coordenadas Paralelas

Comprimento  
da Sépala

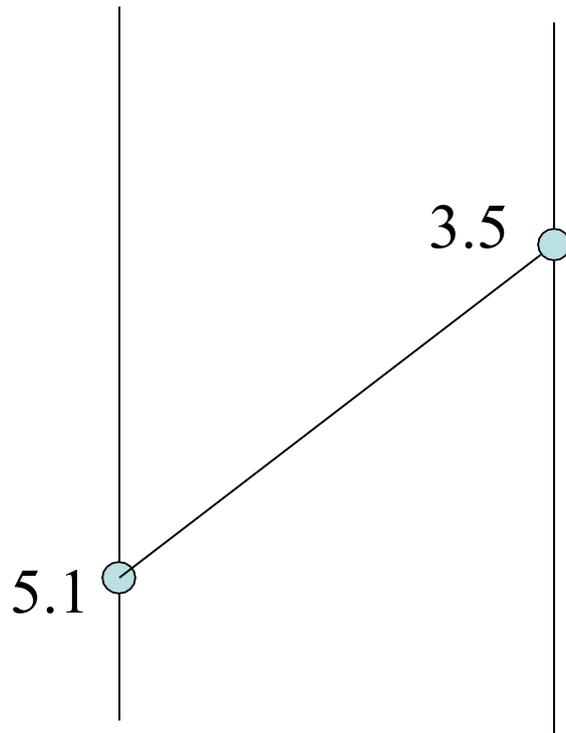


Comprimento Sépala	Largura Sépala	Comprimento Pétala	Largura Pétala
5.1	3.5	1.4	0.2

# Coordenadas Paralelas: 2 dimensões

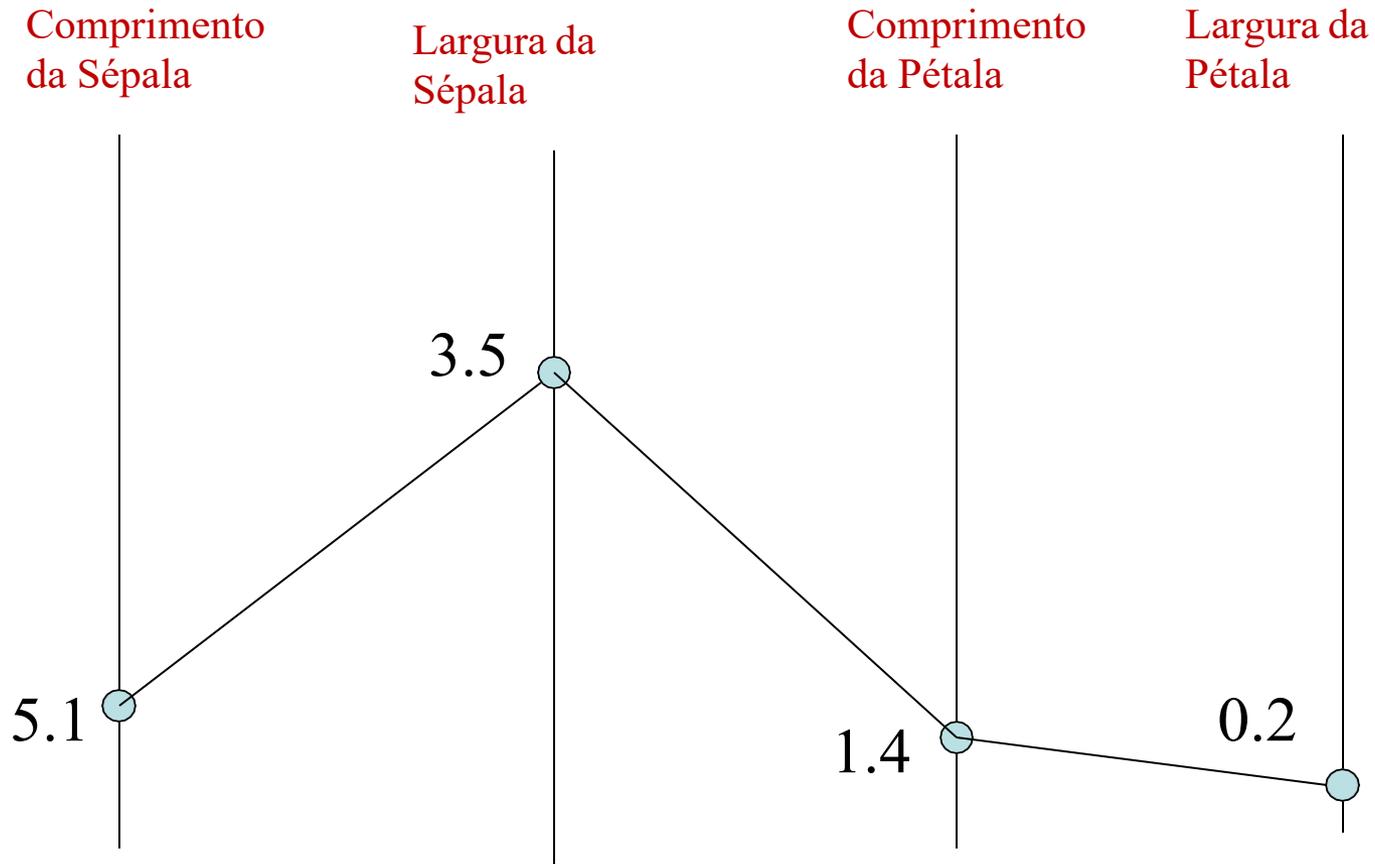
Comprimento  
da Sépala

Largura da  
Sépala



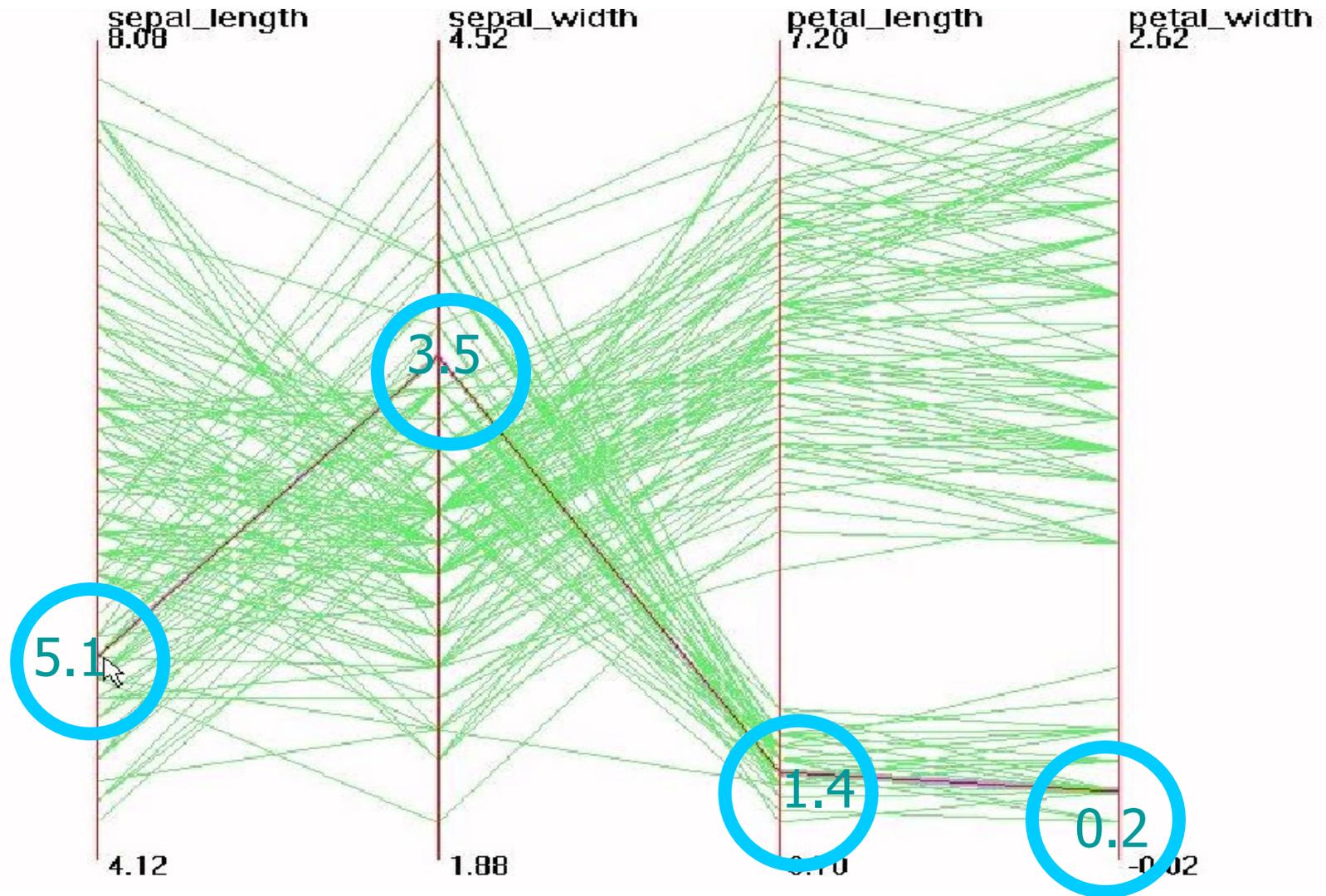
Comprimento Sépala	Largura Sépala	Comprimento Pétala	Largura Pétala
5.1	3.5	1.4	0.2

# Coordenadas Paralelas: 4 dimensões



<b>Comprimento Sépala</b>	<b>Largura Sépala</b>	<b>Comprimento Pétala</b>	<b>Largura Pétala</b>
5.1	3.5	1.4	0.2

# Coordenadas Paralelas: Iris Dataset

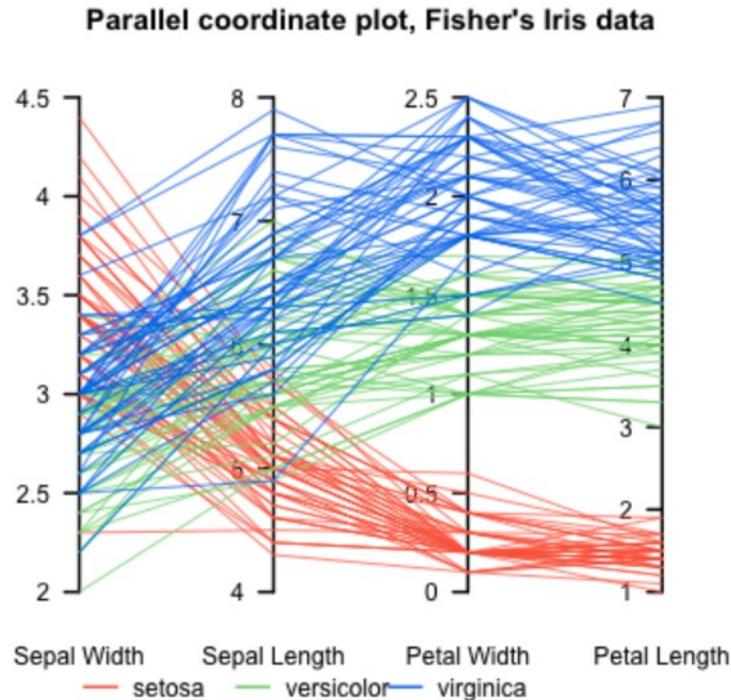


# Coordenadas Paralelas

> Cada elemento é uma linha;

> Cada eixo é uma dimensão;

Permite dezenas de eixos; mas datasets grandes ficam ilegíveis rapidamente



Interatividade é quase sempre necessária:

- Reordenar eixos
- Filtragem por intervalo

Explorações:

- Reorganizar;
- Agrupar;
- Suavizar linhas redundante;

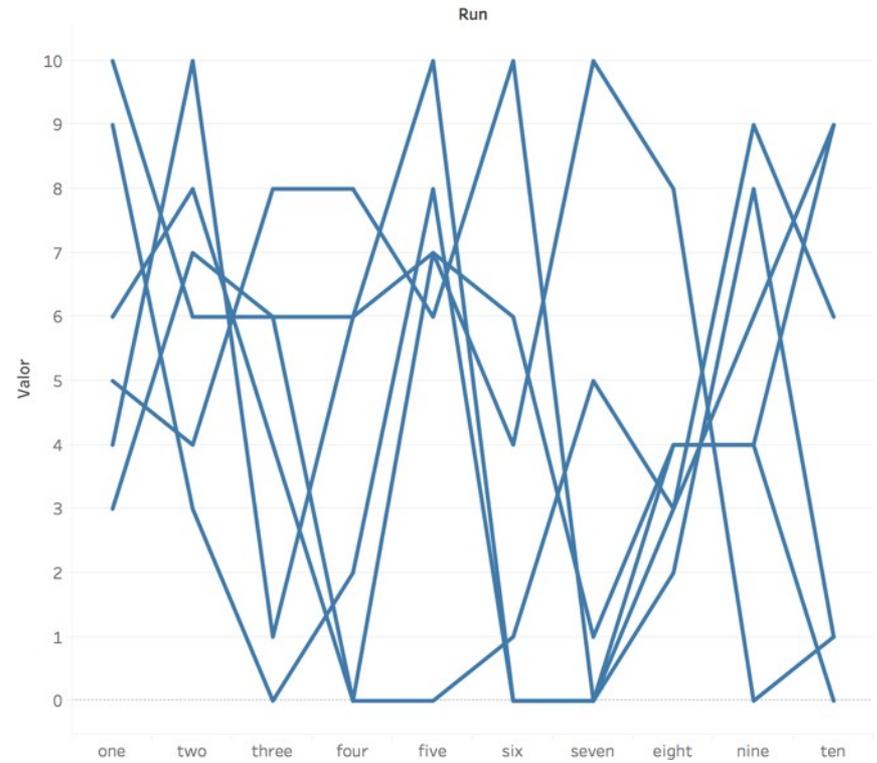
---

<b>O que</b>	Vários atributos (quantitativos, categóricos)
<b>Por que</b>	Tendências, extremos, distribuição, correlação, localizar clusters
<b>Como</b>	Espaço horizontal → atributos; vertical → valores; itens representados como linhas.
<b>Escala</b>	Dezenas de eixos; centenas de itens.

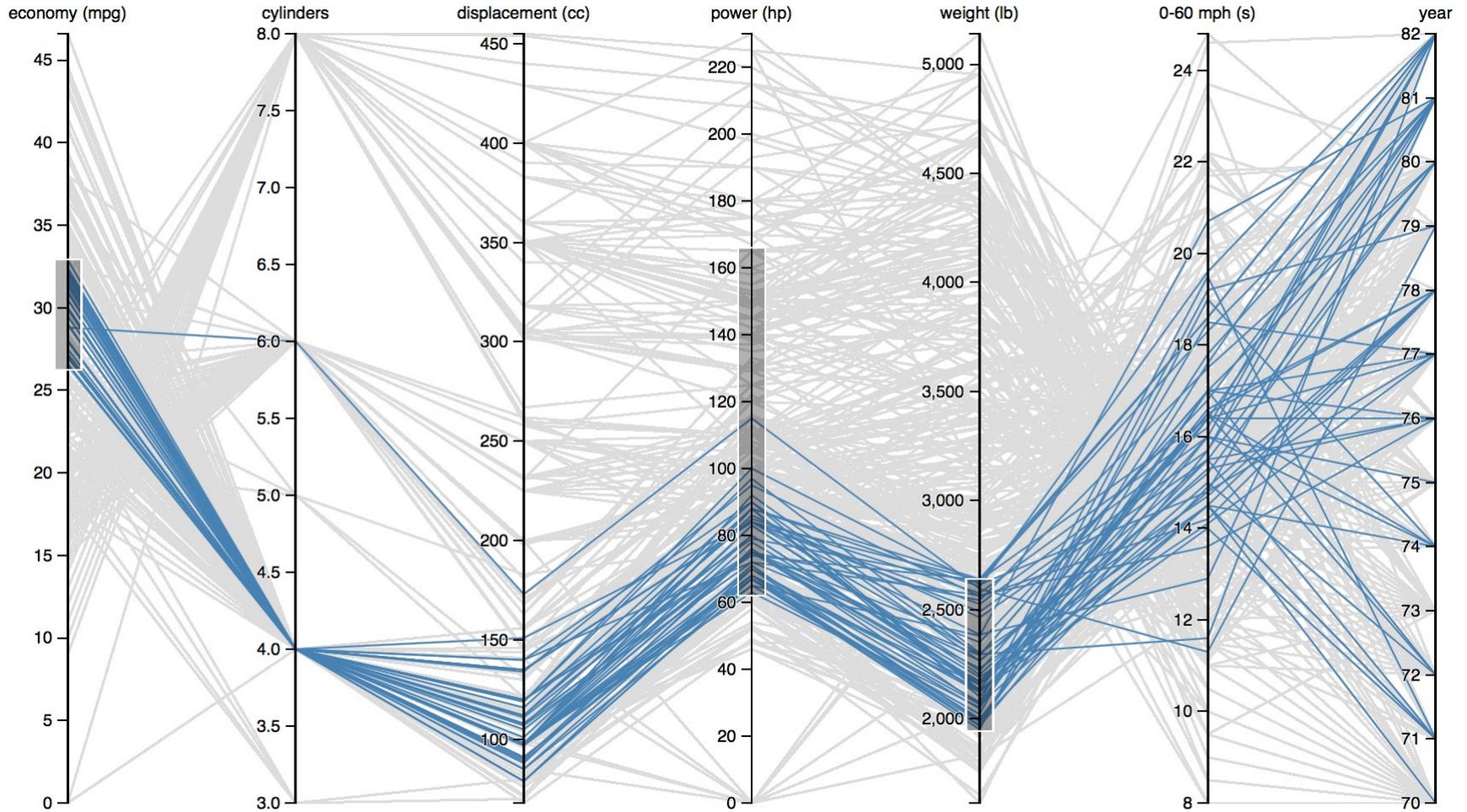
---

# Coordenadas Paralelas

Abc Sheet1 Run	# Sheet1 D1	# Sheet1 D2	# Sheet1 D3	# Sheet1 D4	# Sheet1 D5	# Sheet1 D6
one	4	6	3	9	5	10
two	10	8	7	3	4	6
three	1	4	6	0	8	6
four	6	0	0	2	8	6
five	7	0	7	8	6	10
six	6	1	4	0	10	0
seven	1	5	10	0	0	0
eight	4	3	8	3	2	4
nine	4	9	0	6	8	4
ten	9	6	1	9	1	0



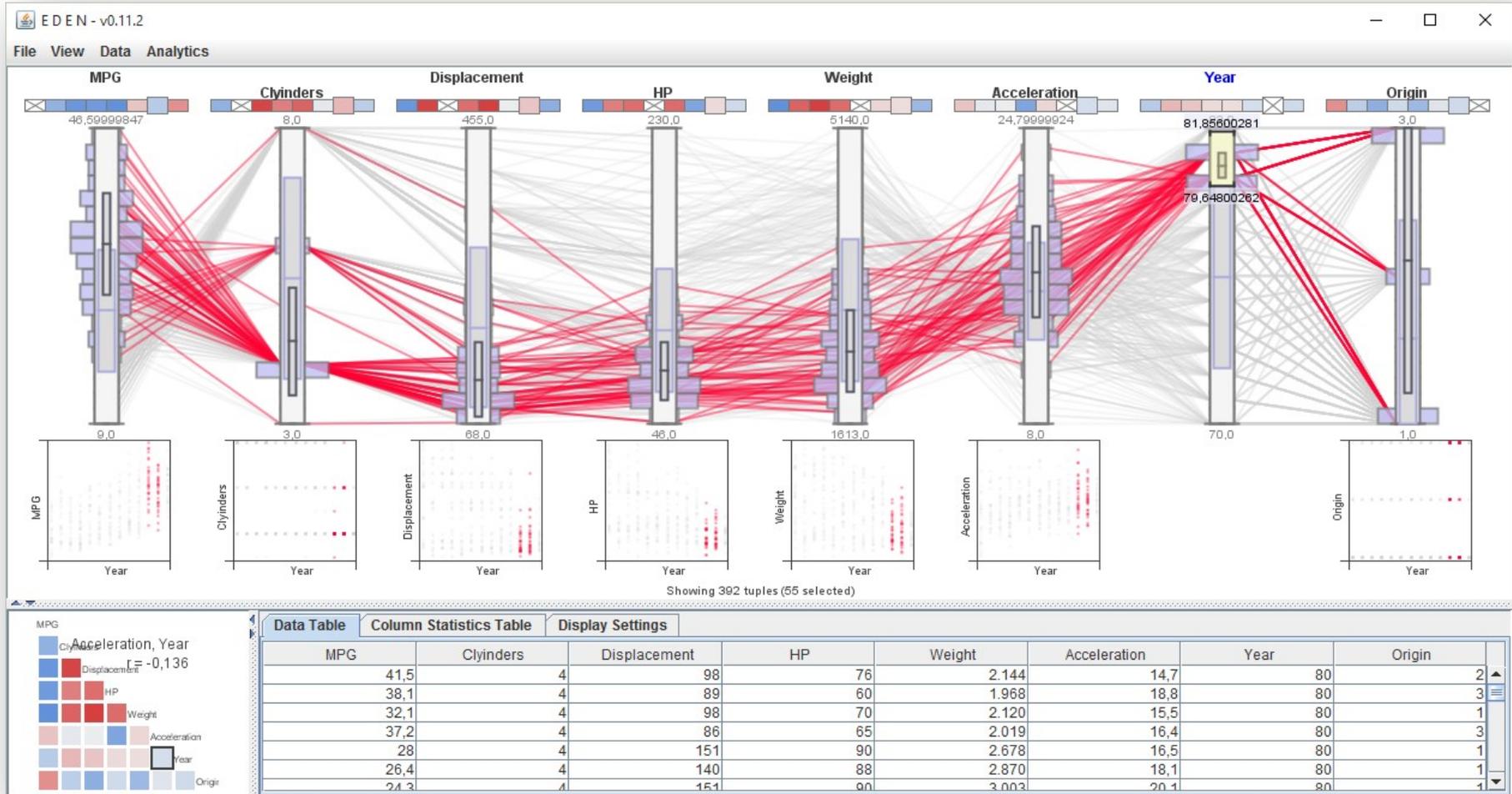
# Coordenadas Paralelas → Brush



<https://bl.ocks.org/jasondavies/1341281>

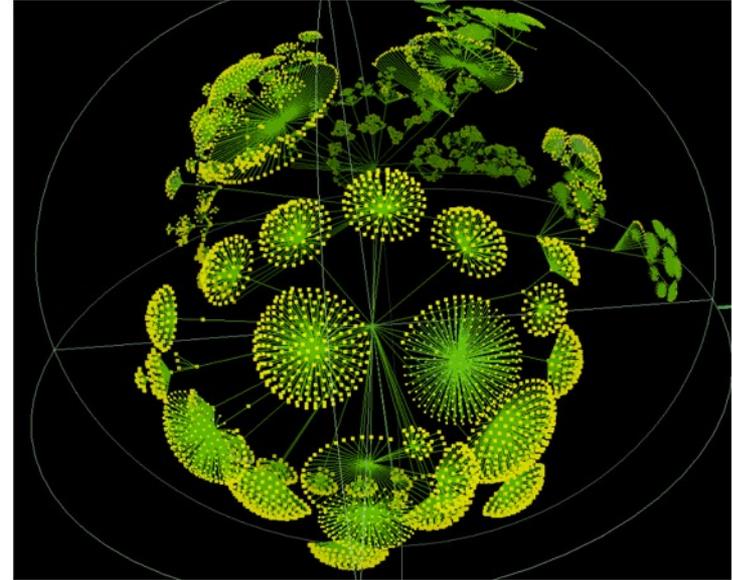
# Por Exemplo ...

<http://cda.ornl.gov/projects/eden/>



# Redes Nó-Link

<http://bl.ocks.org/mbostock/4062045>



<http://www.caida.org/tools/visualizations/malrus>

---

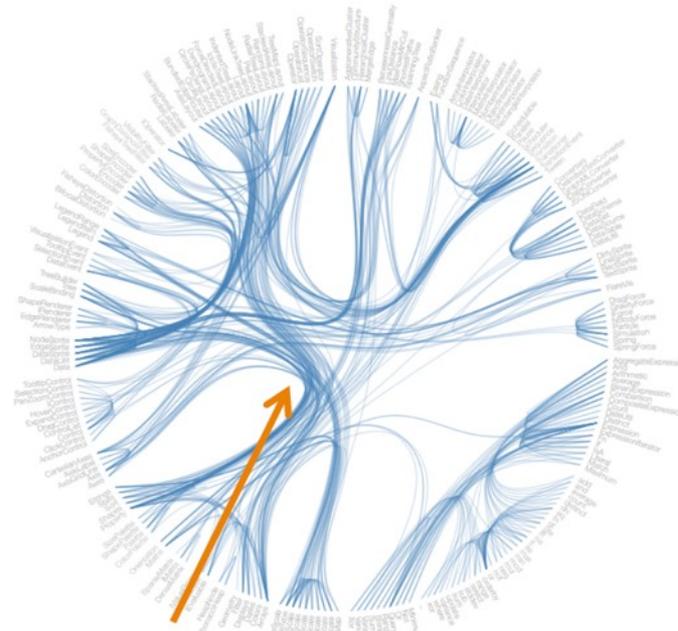
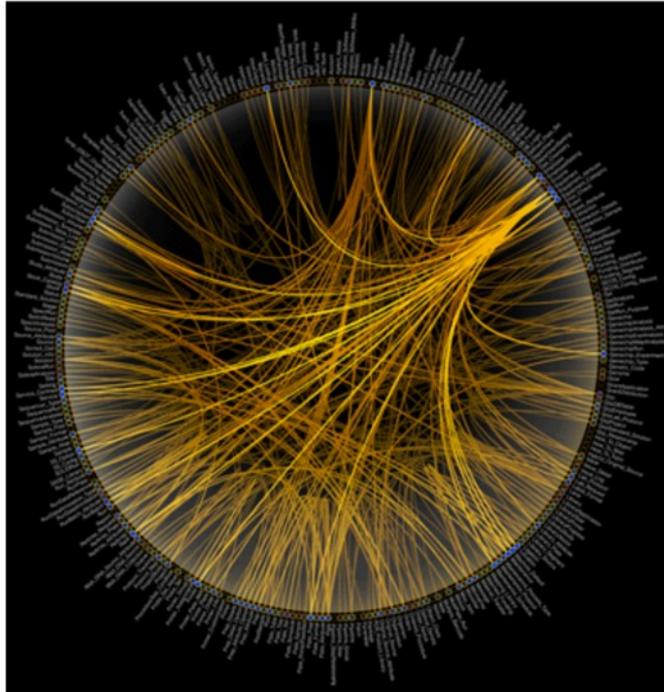
O que	Grafo
Por que	Topologia, Caminhos
Como	Pontos para nós, linhas para conexões; posicionamento no plano
Escala	Nós; até centenas; conexões: centenas. Densidade nó/conexão: $C < 4N$ .

---

Adaptado de (Sultanum, 2015)

# Diagramas de Rede Radiais

<http://mkweb.bcgsc.ca/schemaball/>



<http://bl.ocks.org/mbostock/7607999>

*Bundling - Feixes*

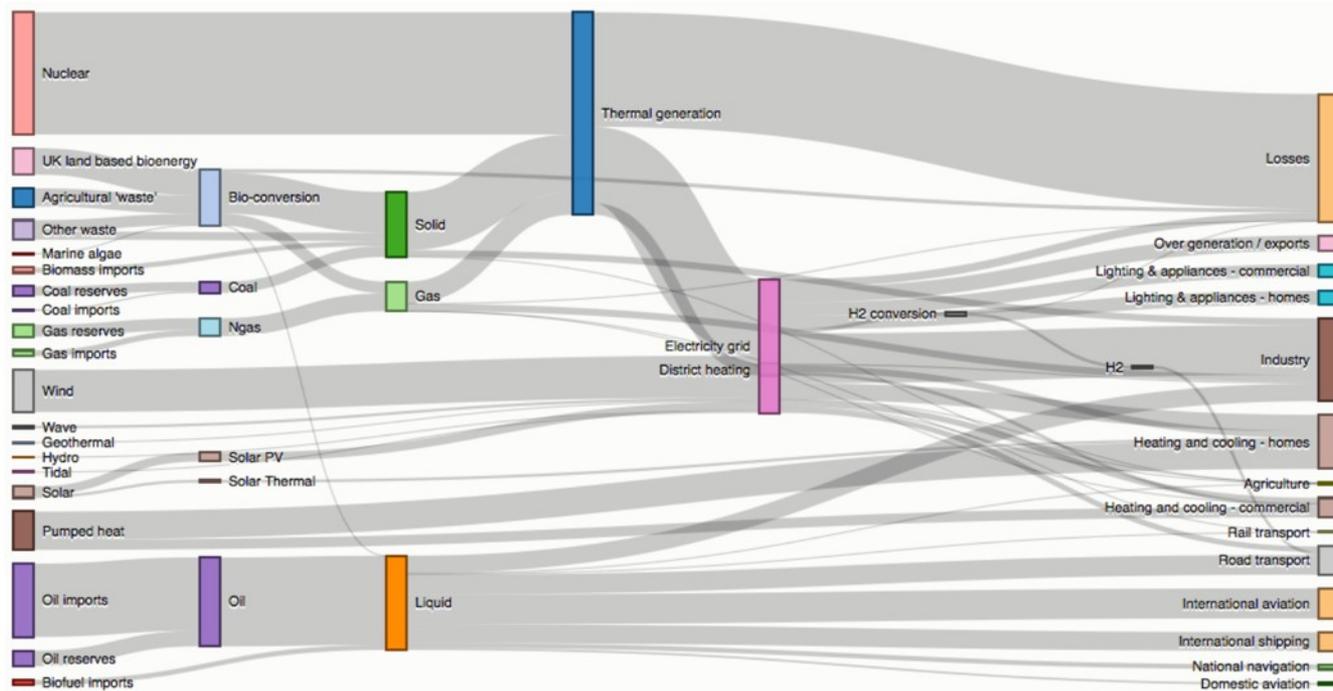
---

O que	Grafo
Por que	Conexões; Dependência
Como	Elementos em layout radial (agrupados por cluster); conexões entre elementos
Escala	Centenas de itens

---

Adaptado de (Sultanum, 2015)

# Diagramas de Sankey



---

O que	Grafo direcionado (atributo quantitativo/ordinal)
Por que	Topologia; caminhos; magnitude do fluxo; conexões
Como	Largura nós é mapeada a magnitude do fluxo; posicionamento dependente de ordem no fluxo
Escala	Dezenas de elementos

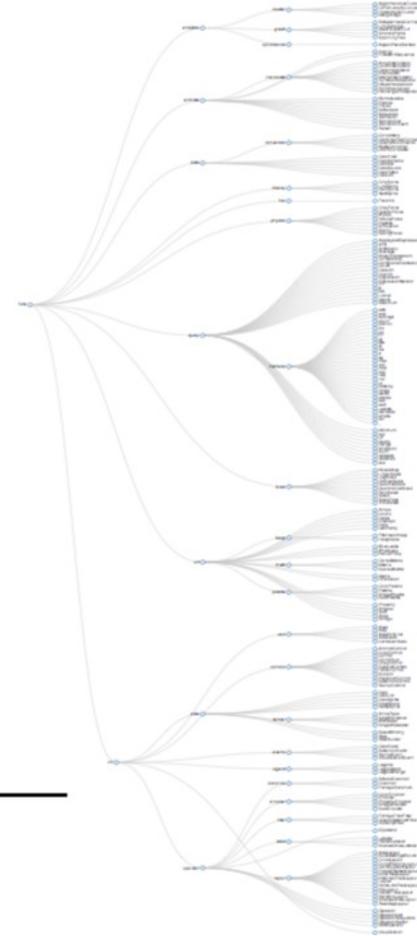
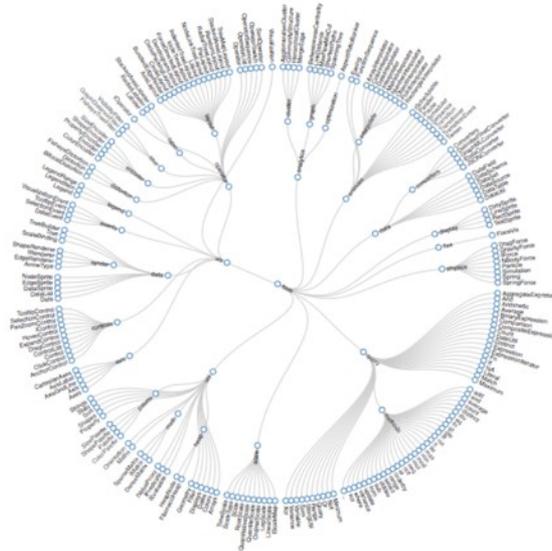
---

Adaptado de (Sultanum, 2015)

# Dendrograma

## Dendrograma dendrogram

<http://bl.ocks.org/mbostock/4339607>



<http://bl.ocks.org/mbostock/4063570>

---

<b>O que</b>	Árvore
<b>Por que</b>	Explorar, localizar, dependência
<b>Como</b>	Layout radial ou linear; nós alinhados por profundidade
<b>Escala</b>	Dezenas de nós

---

Adaptado de (Sultanum, 2015)

# Tree Maps



- 
- O que**      Árvore, atributo quantitativo (opc. + 1 atributo quantitativo/categórico)
  - Por que**    Localizar, explorar, dependência
  - Como**      Divisão recursiva do espaço por hierarquia; atr. Quant mapeado a espaço. Cores podem mapear categorias, ou valores.
  - Escala**     > Milhares folhas, > milhares de links.
-

# Tensão com a Coreia do Norte: O mundo pode estar próximo da 3ª Guerra Mundial?

# Marinherdeira da extrema direita rumo à conquista da França

**Tensão com a Coreia do Norte: O mundo pode estar próximo da 3ª Guerra Mundial?**

Analistas afirmam que tensão entre EUA e Coreia do Norte não será resolvida com mísseis, mas na mesa de negociações. Facebook & Twitter; British Broadcasting Corporation © BBC 2016 Todos os direitos reservados. Por BBC. 23/04/2017 11

Apr 22 23:46 | 287 related articles | Globo.com



Membro da OSCE morre no leste da Ucrânia

Participação em alta Eleições na França têm participação estável

Nuas, ativistas tentam invadir colégio eleitoral contra candidata Marine Le Pen

Trump quer financiar muro na fronteira com México custe o que custar

Acidente deixa três mortos na Avenida Presidente Vargas, Rio

Vídeos: Terrorista esfaqueia quatro em hotéis de Tel Aviv

Colégio eleitoral na França é evacuado por causa de veículo abandonado

Incêndio de grandes proporções atinge galpão na Pampulha

Silval nega acordo de delação, mas diz que vai confessar crimes

Avião com 400 kg de pasta base de cocaína cai em município de MT e cinco pessoas são presas

Policial militar de folga é morto a tiros dentro de lanchonete na Bahia

PM morre ao ser atropelado por moto em fuga no Rio

Jovem é assassinado e tem o corpo atirado em Centro Socioeducativo de Valadares

X-Men: Fênix Negra estreia nos cinemas em 2018

Palco de Luan Santana desaba e deixa feridos em rodeio de Catanduva

Em jogoço, Rio vence Osasco no tie-break e conquista Superliga pela 12ª vez

Emoção e virada: na prorrogação, Arsenal vence City e avança à final da Copa da Inglaterra

Reunião do IMFC Foi conclusiva sobre protecionismo

BR-163 deve ser fechada em protesto contra demissões e paralisação de obras

Jerry Adriani morre aos 70 anos no Rio

Paolla Oliveira define romance de Jeiza e Zeca: 'Amor de gato e rato'

Erin Moran, atriz da série 'Happy days', morre aos 56 anos

United vence mais uma e volta a brigar por uma vaga na Champions

Corinthians lança novos uniformes antes do clássico; veja os modelos

'Querem acabar com o futebol', diz Felipe Melo sobre debate do fair play

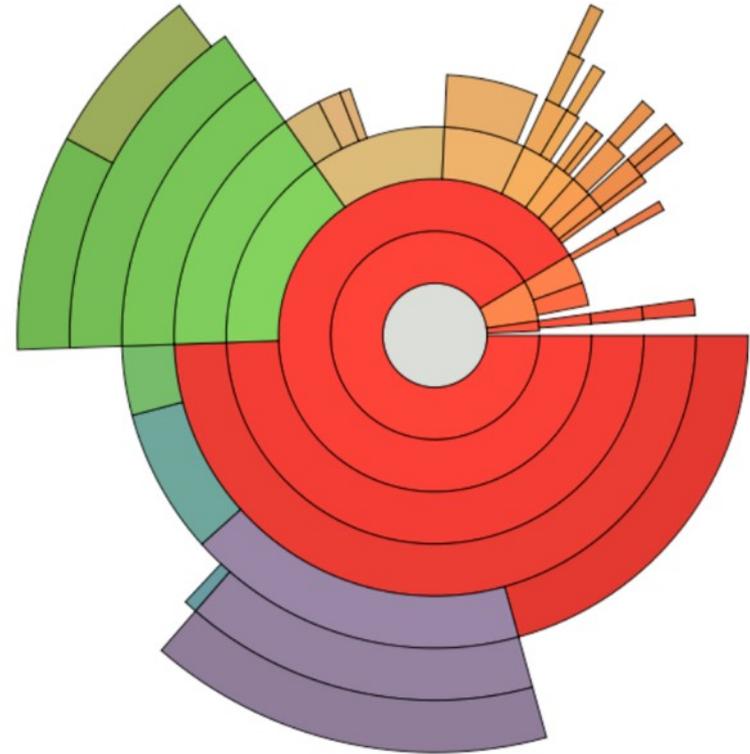
Irmã da jovem que morreu com sarampo está internada

Por que é tão difícil frear a escalada da obesidade infantil?

Grupos de risco Segunda etapa de vacinação contra gripe começa amanhã

# Gráfico Sunburst

<http://bl.ocks.org/mbostock/4063423>



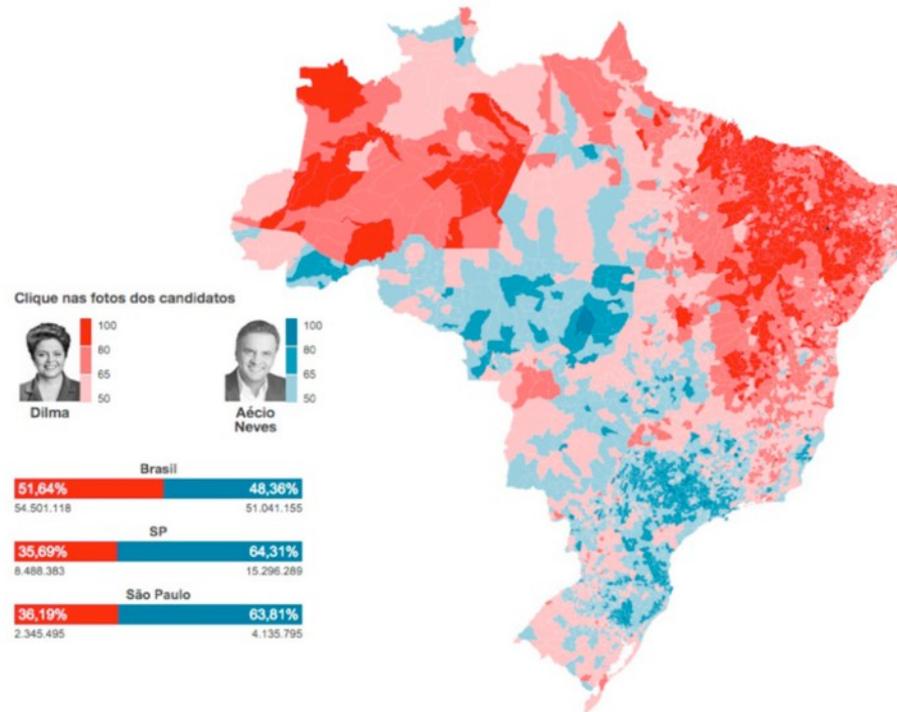
[http://en.wikipedia.org/wiki/Pie\\_chart](http://en.wikipedia.org/wiki/Pie_chart)

---

O que	Árvore + atributo quantitativo (dimensao nó) + atributo categórico (tipo)
Por que	Localizar, explorar, dependência
Como	Divisão do espaço em layout radial. Angulo mapeado a dimensão; cor a categoria do nó.
Escala	Dezenas de níveis; > milhares folhas; > milhares links

---

# Mapa Coroplético



---

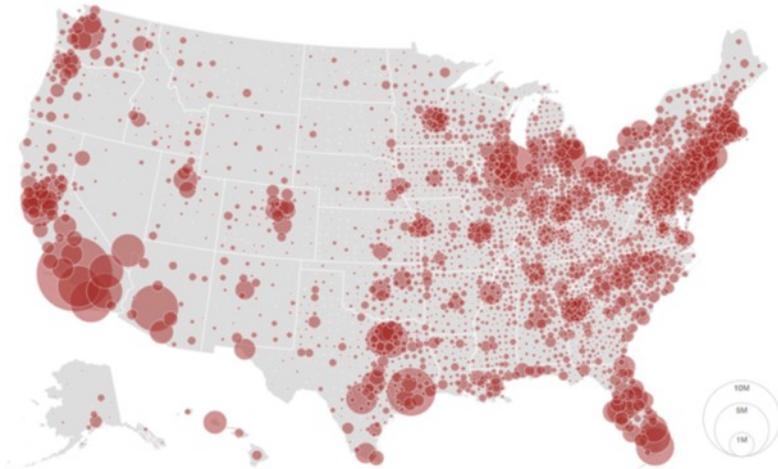
<b>O que</b>	Dado geográfico + 1 atributo quantitativo (ou categórico)
<b>Por que</b>	Distribuição, extremos, características
<b>Como</b>	Uso do espaço geográfico dado, pintura da área com mapa de cores sequencial/segmentado
<b>Escala</b>	< 12 categorias/segmentos

---

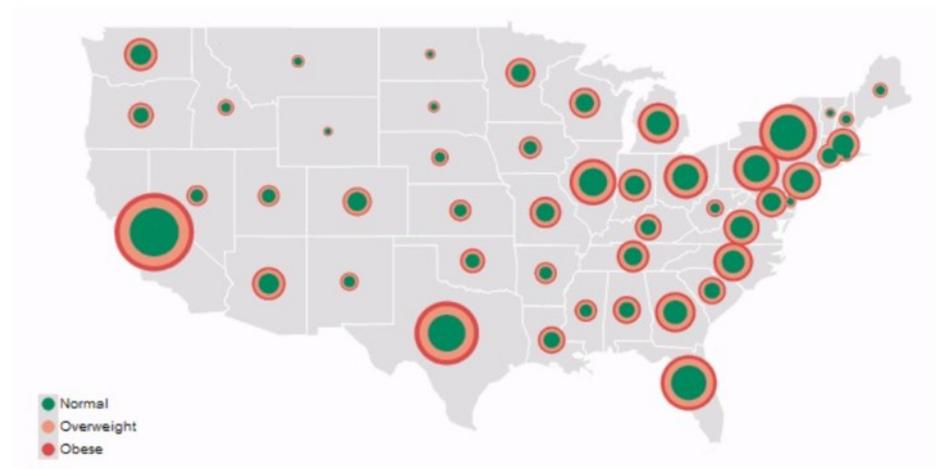
Adaptado de (Sultanum, 2015)

# Símbolos georeferenciados

<http://bl.ocks.org/mbostock/9943478>



<http://homes.cs.washington.edu/~jheer/files/zoo/>



---

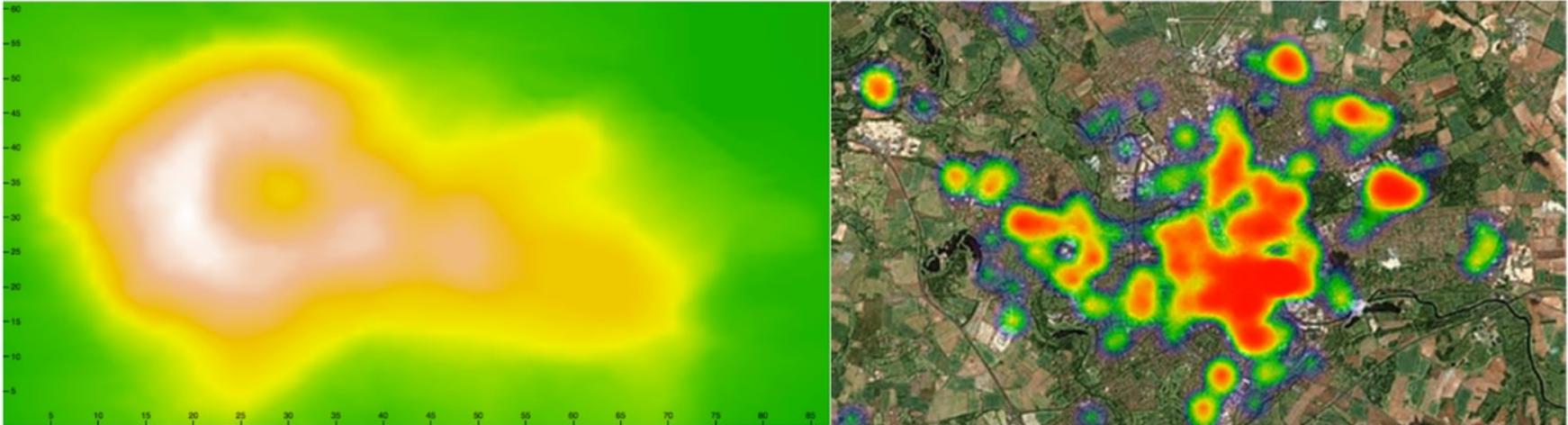
<b>O que</b>	Dado espacial + Atributos quantitativos/qualitativos (associados a posicoes georeferenciadas)
<b>Por que</b>	Distribuição, extremos, características, dependência
<b>Como</b>	Em cada posicao espacial dada é posto marcador (mapeado a cor, tamanho, forma...)
<b>Escala</b>	Dezenas a centenas de marcadores

---

Adaptado de (Sultanum, 2015)

# Heatmaps (contínuos)

<http://bl.ocks.org/mbostock/3074470>



---

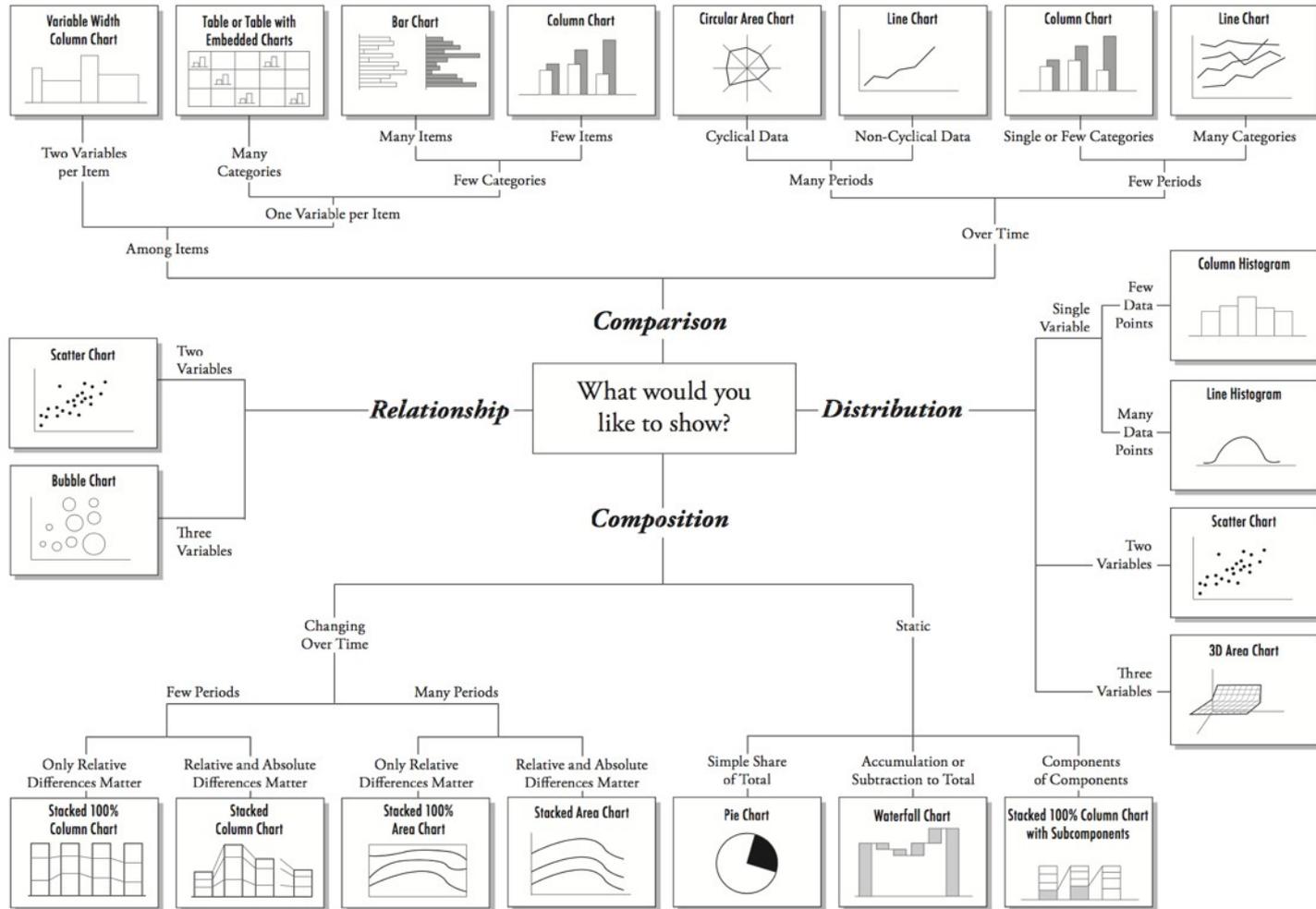
<b>O que</b>	Dado espacial + atributo quantitativo
<b>Por que</b>	Distribuição, extremos, características, resumir
<b>Como</b>	Pixel XY a cada posicao espacial mapeado a valor de atributo quantitativo; mapeado a cores
<b>Escala</b>	> Milhares ocorrências

---

Adaptado de (Sultanum, 2015)

# Chart Suggestion

## Chart Suggestions—A Thought-Starter



# II Workshop em Ciência dos Dados

São Paulo, 11e 12 de maio de 2017

## Visualizando seus dados com Tableau

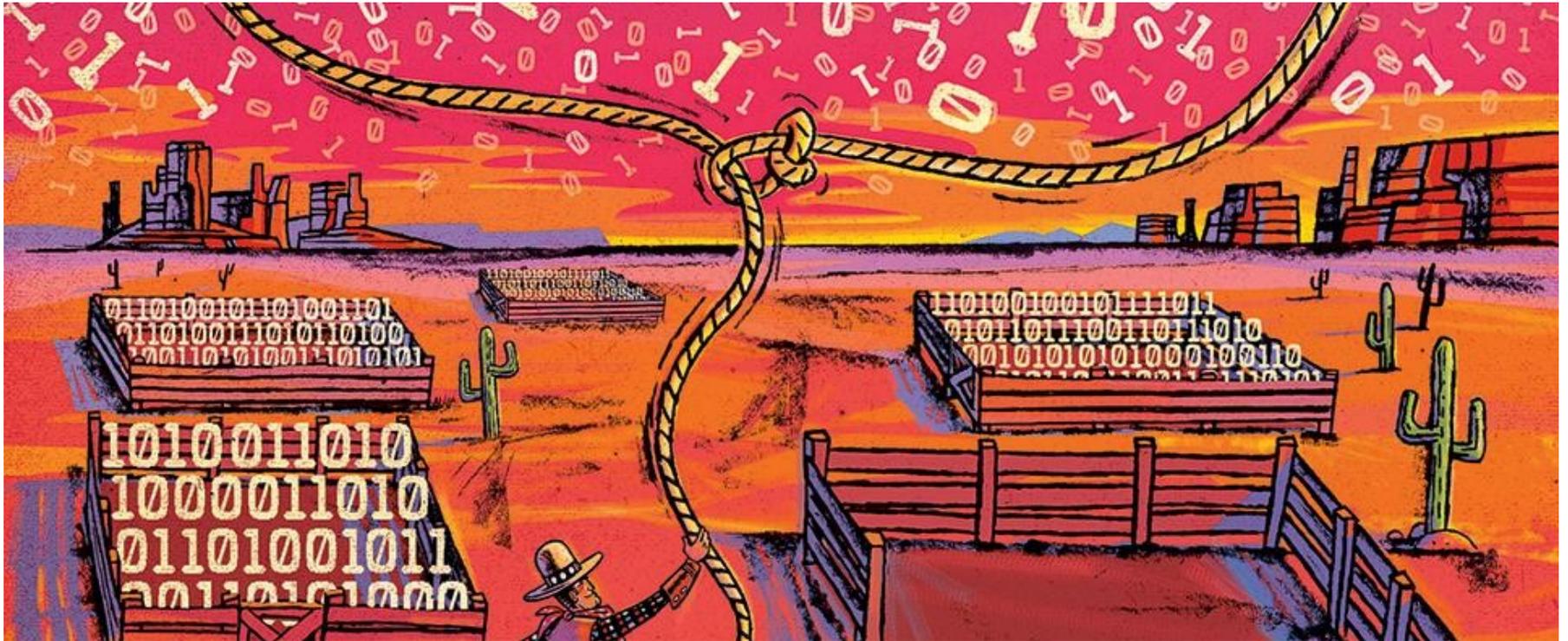
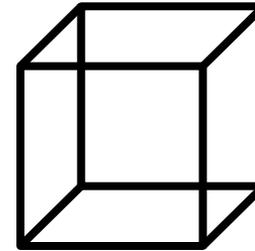
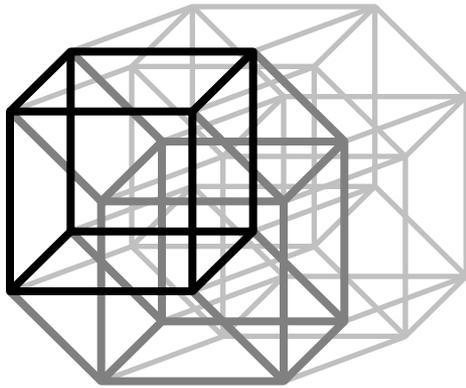


Imagem: <https://alumni.berkeley.edu/sites/default/files/styles/960x400/public/wranglingbigdata.jpg?itok=k0fK1fJQ>

# Mapeamento



Dados de entrada:  
um determinado número de  
dimensões e medidas

Saída gráfica:  
3 dimensões (frequentemente 2), e 1  
dimensão temporal

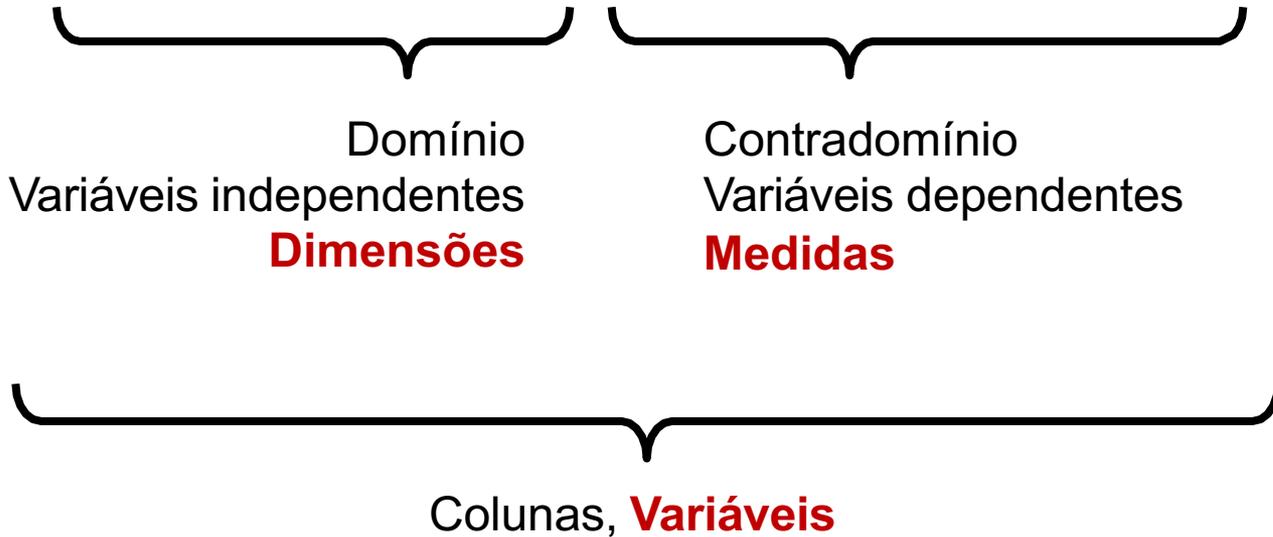
# Terminologia

Um vídeo:

x	y	tempo	red	green	blue
0	0	0	255	0	0
0	1	0	200	10	6
...					
0	0	0.1	255	50	100
0	1	0.1	255	200	190
...					



**Tupla**,  
ponto multidimensional,  
Vetor,  
Linha  
Instância



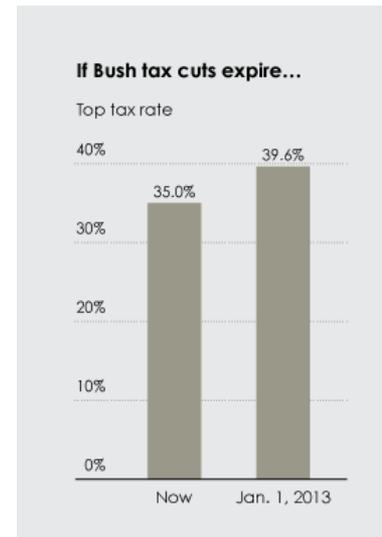
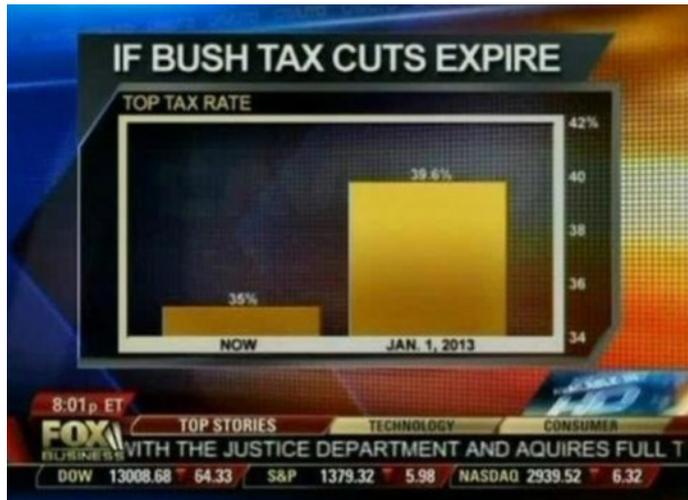
# Dados com mais de uma dimensão

year	id	State	Population	Index offenses	Violent crime	Murder	Forcible rape	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft
1960	1	Alabama	3266740	39920	6097	406	281	898	4512	33823	11626	19344	2853
1960	2	Alaska	226167	3730	236	23	47	64	102	3494	751	2195	548
1960	4	Arizona	1302161	39243	2704	78	209	706	1711	36539	8926	23207	4406
1960	5	Arkansas	1786272	18472	1924	152	159	443	1170	16548	5399	10250	899
1960	6	California	15717204	546069	37558	616	2859	15287	18796	508511	143102	311956	53453
1960	8	Colorado	1753947	38103	2408	73	229	1362	744	35695	9996	21949	3750
1960	9	Connecticut	2535234	29321	928	41	103	236	548	28393	8452	16653	3288
1960	10	Delaware	446292	9642	375	33	41	157	144	9267	2661	5867	739
1960	11	District of Co	763956	20725	4230	81	111	1072	2966	16495	4587	9905	2003
1960	12	Florida	4951560	133919	11061	527	403	4005	6126	122858	39966	73603	9289
...													
1972	54	West Virginia	1781000	25584	2299	109	146	562	1482	23285	7356	13976	1953
1972	55	Wisconsin	4520000	133382	4358	126	376	1661	2195	129024	28862	89642	10520
1972	56	Wyoming	345000	10461	511	14	48	117	332	9950	2057	7190	703
1973	1	Alabama	3539000	91389	12390	468	751	2809	8362	78999	31754	39206	8039
1973	2	Alaska	330000	16313	1269	33	147	221	868	15044	3852	9456	1736
1973	4	Arizona	2058000	137966	9877	167	637	3031	6042	128089	40301	76560	11228
1973	5	Arkansas	2037000	56149	5905	180	398	1456	3871	50244	18088	29204	2952
1973	6	California	20601000	1298872	116563	1862	8357	49531	56813	1182309	407824	643488	130997
1973	8	Colorado	2437000	133933	10088	193	944	3970	4981	123845	38963	70931	13951
1973	9	Connecticut	3076000	112717	6421	102	342	2589	3388	106296	31661	58742	15893
...													
2000	44	Rhode Island	1048319	36444	3121	45	412	922	1742	33323	6620	22038	4665
2000	45	South Carolina	4012012	209482	32293	233	1511	5883	24666	177189	38888	123094	15207
2000	46	South Dakota	754844	17511	1259	7	305	131	816	16252	2896	12558	798
2000	47	Tennessee	5689283	278218	40233	410	2186	9465	28172	237985	56344	154111	27530
2000	48	Texas	20851820	1033311	113653	1238	7856	30257	74302	919658	188975	637522	93161
2000	49	Utah	2233169	99958	5711	43	863	1242	3563	94247	14348	73438	6461
2000	50	Vermont	608827	18185	691	9	140	117	425	17494	3501	13184	809
2000	51	Virginia	7078515	214348	19943	401	1616	6295	11631	194405	30434	146158	17813
2000	53	Washington	5894121	300932	21788	196	2737	5812	13043	279144	53476	190650	35018
2000	54	West Virginia	1808344	47067	5723	46	331	749	4597	41344	9890	28139	3315
2000	55	Wisconsin	5363675	172124	12700	169	1165	4537	6829	159424	25183	119605	14636
2000	56	Wyoming	493782	16285	1316	12	160	70	1074	14969	2078	12318	573

# Dicas

## (Do and Don'ts – Duke University)

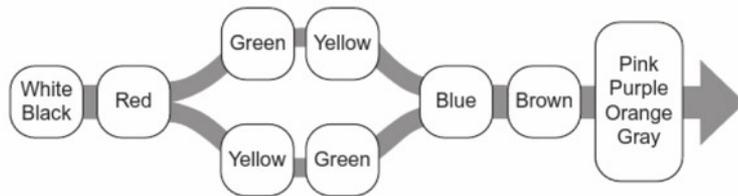
- **Posição** é tudo
  - Textos rotacionados são mais difíceis de ler
  - Pessoas têm dificuldade de ler textos em determinados ângulos
  - Evite sobreposições
- Use todo o eixo, evitando assim distorções



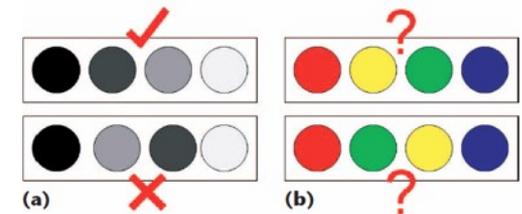
# Dicas

## (Do and Don'ts – Duke University)

- **Cores** não são tão fáceis de distinguir
  - Pessoas têm dificuldade em diferenciar mais de 5 – 7 cores
  - É difícil distinguir mais do que 7 tons de uma cor
  - Evite *rainbows* e faça uso de uma variação de cor razoável



**Figure 4.11** This is the order of appearance of color names in languages around the world, according to the research of Berlin and Kay (1969). The order is fixed, with the exception that sometimes yellow is present before green and sometimes the reverse is the case.



1 Perceptual ordering. (a) We can easily place the gray paint chips in order based on perception, (b) but cannot do this with the colored chips.



# Tableau – Visualizações Disponíveis

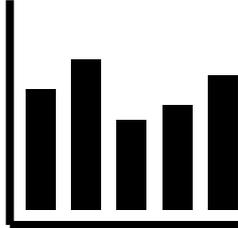
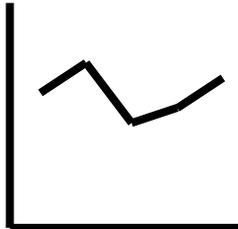
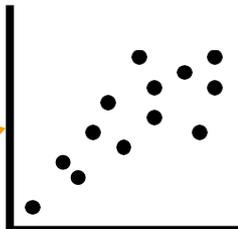
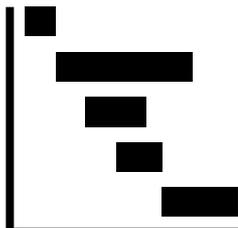
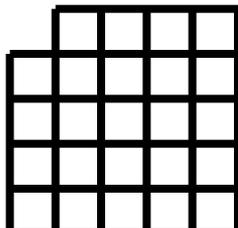
- Text tables  
*a grid representing variables by size and color*
- Heat maps
- Highlight tables  
*a grid representing variables by text and color*
- **Maps (symbol, filled)**
- Pie charts
- Horizontal bars
- Stacked bars
- Side-by-side bars
- **Treemap**  
*a grid representing variables by size*

- Circle views
- Side-by-side circles
- Lines/Area charts
- Lines/Area charts (discrete)
- Dual lines
- Dual combination
- Scatter plots
- Histogram
- Box-and-whisker
- Gantt
- Bullet graphs
- Packed bubbles/  
Word cloud



# Tableau

## Heurísticas do Tableau

Quantitative variable as a function of a categorical variable	Bar chart	
Quantitative variable as a function of a quantitative variable	Line graph	
Quantitative variable as a function of (ordinal) time		
Two dependent quantitative variables	Scatter plot	
Categorical variable as a function of a quantitative variable	Gantt chart	
Categorical independent variable + quantitative independent variable		
Two independent categorical variables	Cross tabulation ("cross tab")	

# II Workshop em Ciência dos Dados

São Paulo, 11e 12 de maio de 2017

## Visualização de Dados na Prática: Tableau 10.2

**André Filipe M. Batista**

Daniel Lins da Silva

Suelane Garcia

Thiago Adriano Coleti



ESCOLA  
POLITÉCNICA  
DA USP

# Boas referências

- <http://www.dataplusscience.com/TableauReferenceGuide/index.html>

data + science  
= transforming data to insight

Tableau Reference Guide

SEARCH TERM

Graphing	Color	Calculations	Mapping	Data	Level of Detail
Training Videos	JavaScript API	Sound and Video	Tooltips	Server	Misc.

## Graphing

- [Creating a Lollipop Chart](#) by Andy Kriebel
- [Creating a Lollipop Chart- How to \(with Viz and Video\)](#) by Andy Kriebel
- [Dual Axis Chart with Overlapping Bars and a Line](#) by Andy Kriebel
- [Moving Reference Line](#) by Andy Kriebel
- [Dashed Lines](#) by Michael Kane
- [Panel Charts](#) by Andy Cotgreave
- [Panel Charts Improved](#) by Joe Mako
- [How to Make Donut Charts](#) by Ryan Sleeper
- [KPI donut charts](#) by Andy Kriebel
- [Radial Bar Charts](#) by Dave Hart
- [Radial Trees](#) by Chris DeMartini
- [Radar Chart](#) by Andrew Ball
- [Sparkbars with Information at the End](#) by Andy Kriebel
- [Arrow Chart](#) by Joe Mako
- [Marimekko Charts](#) by Rob Austin
- [Network Graphs with NodeXL](#) by Allan Walker
- [Parallel Coordinates](#) by Joe Mako
- [Slopegraphs Tutorial](#) by Ben Jones
- [Slopegraphs](#) by Alex Kerin
- [Slopegraphs in 5 Steps](#) by Andy Kriebel
- [Slopegraphs \(Video\)](#) by Andy Kriebel
- [Sankey Style Slopegraphs Part 1](#) by Jeffrey Shaffer
- [Sankey Style Slopegraphs Part 2 - "How To"](#) by Jeffrey Shaffer
- [Sankey Style Slopegraphs Part 3 - "Multiple Graphs"](#) by Jeffrey Shaffer
- [Sankey Diagram - Original Post](#) by Jeffrey Shaffer
- [Sankey Diagram with Polygons](#) by Olivier Catherin
- [Creating Sigmoid with 2 Rows of Data \(for Sankey and Slopegraphs\)](#) by Alexander Mou
- [Sankey Diagram with Data Densification](#) by Chris Love

# Dataset



## San Francisco Crime Classification

Predict the category of crimes that occurred in the city by the bay

2,335 teams · a year ago

This dataset contains incidents derived from SFPD Crime Incident Reporting system. The data ranges from 1/1/2003 to 5/13/2015. The training set and test set rotate every week, meaning week 1,3,5,7... belong to test set, week 2,4,6,8 belong to training set.

2003-01-07 07:52:00	WARRANTS	WARRANT ARREST	Tuesday	SOUTHERN	ARREST, BOOKED	5TH ST / SHIPLEY ST	-122.402843	37.779829
2003-01-07 04:49:00	WARRANTS	ENROUTE TO OUTSIDE JURISDICTION	Tuesday	TENDERLOIN	ARREST, BOOKED	CYRIL MAGNIN STORH ST / EDDY ST	-122.408495	37.784452
2003-01-07 03:52:00	WARRANTS	WARRANT ARREST	Tuesday	NORTHERN	ARREST, BOOKED	OFARRELL ST / LARKIN ST	-122.417904	37.785167
2003-01-07 03:34:00	WARRANTS	WARRANT ARREST	Tuesday	NORTHERN	ARREST, BOOKED	DIVISADERO ST / LOMBARD ST	-122.442650	37.798999
2003-01-07 01:22:00	WARRANTS	WARRANT ARREST	Tuesday	SOUTHERN	ARREST, BOOKED	900 Block of MARKET ST	-122.409537	37.782691
2003-01-06 23:30:00	WARRANTS	ENROUTE TO OUTSIDE JURISDICTION	Monday	BAYVIEW	ARREST, BOOKED	REVERE AV / INGALLS ST	-122.384557	37.728487
2003-01-06 23:14:00	WARRANTS	WARRANT ARREST	Monday	CENTRAL	ARREST, BOOKED	BUSH ST / HYDE ST	-122.417019	37.789110
2003-01-06 22:45:00	WARRANTS	WARRANT ARREST	Monday	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421
2003-01-06 22:45:00	WARRANTS	ENROUTE TO OUTSIDE JURISDICTION	Monday	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421
2003-01-06 22:19:00	WARRANTS	ENROUTE TO OUTSIDE JURISDICTION	Monday	NORTHERN	ARREST, BOOKED	GEARY ST / POLK ST	-122.419740	37.785693
2003-01-06 21:54:00	WARRANTS	ENROUTE TO OUTSIDE JURISDICTION	Monday	NORTHERN	ARREST, BOOKED	SUTTER ST / POLK ST	-122.420120	37.787757

### Data fields

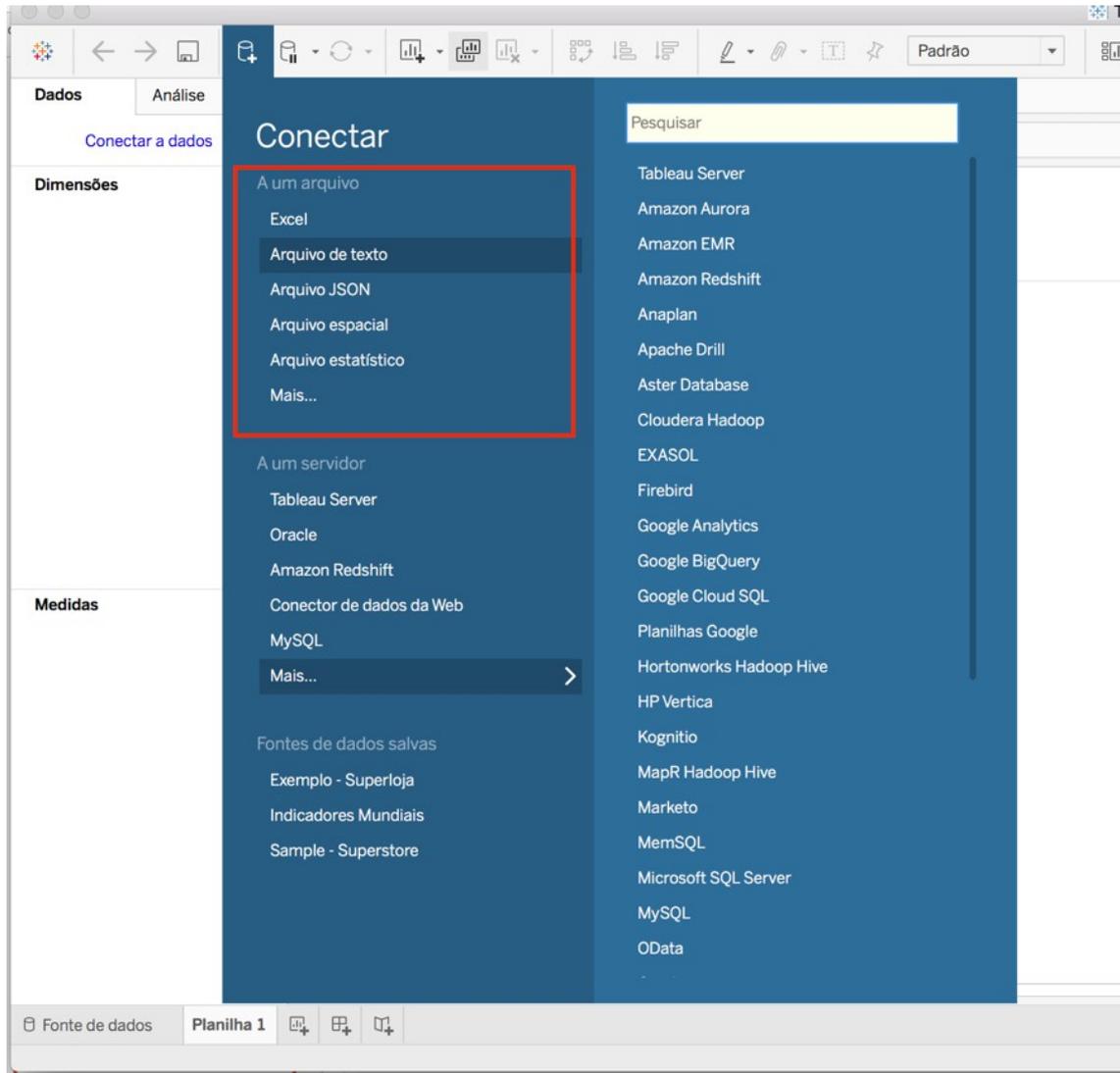
- **Dates** - timestamp of the crime incident
- **Category** - category of the crime incident (only in train.csv). **This is the target variable you are going to predict.**
- **Descript** - detailed description of the crime incident (only in train.csv)
- **DayOfWeek** - the day of the week
- **PdDistrict** - name of the Police Department District
- **Resolution** - how the crime incident was resolved (only in train.csv)
- **Address** - the approximate street address of the crime incident
- **X** - Longitude
- **Y** - Latitude

<https://www.kaggle.com/c/sf-crime/data>

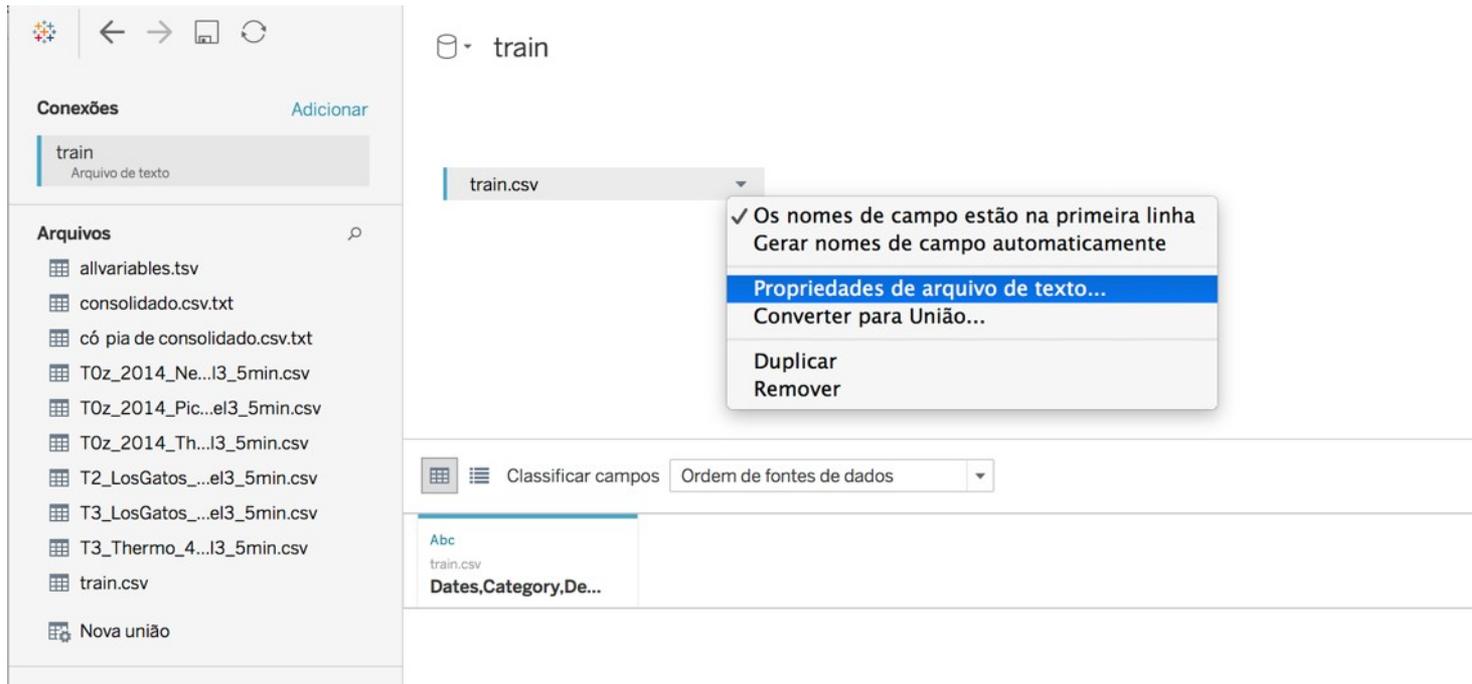
# Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Dates,Category,Descript,DayOfWeek,PdDistrict,Resolution,Address,X,Y															
2	2015-05-13 23:53:00,WARRANTS,WARRANT ARREST,Wednesday,NORTHERN,"ARREST, BOOKED",OAK ST / LAGUNA ST,-122.425891675136,37.7745985956747															
3	2015-05-13 23:53:00,OTHER OFFENSES,TRAFFIC VIOLATION ARREST,Wednesday,NORTHERN,"ARREST, BOOKED",OAK ST / LAGUNA ST,-122.425891675136,37.7745985956747															
4	2015-05-13 23:33:00,OTHER OFFENSES,TRAFFIC VIOLATION ARREST,Wednesday,NORTHERN,"ARREST, BOOKED",VANNESS AV / GREENWICH ST,-122.42436302145,37.8004143219856															
5	2015-05-13 23:30:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,NORTHERN,NONE,1500 Block of LOMBARD ST,-122.42699532676599,37.80087263276921															
6	2015-05-13 23:30:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,PARK,NONE,100 Block of BRODERICK ST,-122.438737622757,37.771541172057795															
7	2015-05-13 23:30:00,LARCENY/THEFT,GRAND THEFT FROM UNLOCKED AUTO,Wednesday,INGLESIDE,NONE,0 Block of TEDDY AV,-122.40325236121201,37.713430704116															
8	2015-05-13 23:30:00,VEHICLE THEFT,STOLEN AUTOMOBILE,Wednesday,INGLESIDE,NONE,AVALON AV / PERU AV,-122.423326976668,37.7251380403778															
9	2015-05-13 23:30:00,VEHICLE THEFT,STOLEN AUTOMOBILE,Wednesday,BAYVIEW,NONE,KIRKWOOD AV / DONAHUE ST,-122.371274317441,37.7275640719518															
10	2015-05-13 23:00:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,RICHMOND,NONE,600 Block of 47TH AV,-122.508194031117,37.776601260681204															
11	2015-05-13 23:00:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,CENTRAL,NONE,JEFFERSON ST / LEAVENWORTH ST,-122.419087676747,37.8078015516515															
12	2015-05-13 22:58:00,LARCENY/THEFT,PETTY THEFT FROM LOCKED AUTO,Wednesday,CENTRAL,NONE,JEFFERSON ST / LEAVENWORTH ST,-122.419087676747,37.8078015516515															
13	2015-05-13 22:30:00,OTHER OFFENSES,MISCELLANEOUS INVESTIGATION,Wednesday,TARAVAL,NONE,0 Block of ESCOLTA WY,-122.487983072777,37.737666654332706															
14	2015-05-13 22:30:00,VANDALISM,"MALICIOUS MISCHIEF, VANDALISM OF VEHICLES",Wednesday,TENDERLOIN,NONE,TURK ST / JONES ST,-122.41241426358101,37.7830037964534															
15	2015-05-13 22:06:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,NORTHERN,NONE,FILLMORE ST / GEARY BL,-122.432914603494,37.7843533426568															
16	2015-05-13 22:00:00,NON-CRIMINAL,FOUND PROPERTY,Wednesday,BAYVIEW,NONE,200 Block of WILLIAMS AV,-122.397744427103,37.7299346936044															
17	2015-05-13 22:00:00,NON-CRIMINAL,FOUND PROPERTY,Wednesday,BAYVIEW,NONE,0 Block of MENDELL ST,-122.38369150395901,37.7431890419965															
18	2015-05-13 22:00:00,ROBBERY,"ROBBERY, ARMED WITH A KNIFE",Wednesday,TENDERLOIN,NONE,EDDY ST / JONES ST,-122.412597377187,37.783932027727296															
19	2015-05-13 21:55:00,ASSAULT,AGGRAVATED ASSAULT WITH BODILY FORCE,Wednesday,INGLESIDE,NONE,GODEUS ST / MISSION ST,-122.421681531572,37.7428222004845															
20	2015-05-13 21:40:00,OTHER OFFENSES,TRAFFIC VIOLATION,Wednesday,BAYVIEW,"ARREST, BOOKED",MENDELL ST / HUDSON AV,-122.38640086995301,37.738983491072															
21	2015-05-13 21:30:00,NON-CRIMINAL,FOUND PROPERTY,Wednesday,TENDERLOIN,NONE,100 Block of JONES ST,-122.412249767634,37.782556330202															
22	2015-05-13 21:30:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,INGLESIDE,NONE,200 Block of EVELYN WY,-122.449389111284,37.7426688025766															
23	2015-05-13 21:17:00,ROBBERY,"ROBBERY, BODILY FORCE",Wednesday,INGLESIDE,NONE,1600 Block of VALENCIA ST,-122.420272135283,37.7473316298785															
24	2015-05-13 21:11:00,WARRANTS,WARRANT ARREST,Wednesday,TENDERLOIN,NONE,100 Block of JONES ST,-122.412249767634,37.782556330202															
25	2015-05-13 21:11:00,NON-CRIMINAL,"STAY AWAY OR COURT ORDER, NON-DV RELATED",Wednesday,TENDERLOIN,NONE,100 Block of JONES ST,-122.412249767634,37.782556330202															
26	2015-05-13 21:10:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,NORTHERN,NONE,FILLMORE ST / LOMBARD ST,-122.43604920358601,37.79984122288229															
27	2015-05-13 21:00:00,NON-CRIMINAL,LOST PROPERTY,Wednesday,TENDERLOIN,NONE,300 Block of OFARRELL ST,-122.41050925879499,37.786043222299206															
28	2015-05-13 21:00:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,NORTHERN,NONE,2000 Block of BUSH ST,-122.43101755702699,37.7873880712241															
29	2015-05-13 21:00:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,INGLESIDE,NONE,500 Block of COLLEGE AV,-122.42365634294501,37.7325564882065															
30	2015-05-13 21:00:00,LARCENY/THEFT,ATTEMPTED THEFT FROM LOCKED VEHICLE,Wednesday,TARAVAL,NONE,19TH AV / SANTIAGO ST,-122.475773497852,37.744919069591															
31	2015-05-13 20:56:00,OTHER OFFENSES,MISCELLANEOUS INVESTIGATION,Wednesday,TARAVAL,NONE,2000 Block of 41ST AV,-122.49978712069,37.74851760347229															
32	2015-05-13 20:54:00,LARCENY/THEFT,GRAND THEFT FROM LOCKED AUTO,Wednesday,NORTHERN,NONE,1300 Block of WEBSTER ST,-122.431046366089,37.7830295716044															
33	2015-05-13 20:50:00,NON-CRIMINAL,"CIVIL SIDEWALKS, CITATION",Wednesday,MISSION,"ARREST, BOOKED",400 Block of CASTRO ST,-122.43515009981199,37.7617597243598															
34	2015-05-13 20:45:00,VANDALISM,"MALICIOUS MISCHIEF, VANDALISM",Wednesday,NORTHERN,NONE,1500 Block of FILLMORE ST,-122.432743822617,37.7838424505847															

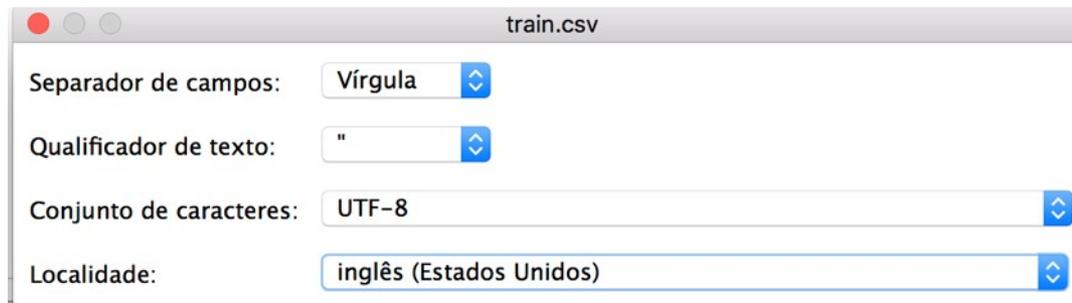
# Importando Dataset no Tableau



# Ajustando o separador do arquivo



The screenshot shows a file manager window titled 'train'. On the left, there is a sidebar with 'Conexões' (Connections) and 'Arquivos' (Files). The 'Arquivos' list includes files like 'allvariables.tsv', 'consolidado.csv.txt', and 'train.csv'. The main area shows 'train.csv' selected, with a context menu open. The menu options are: 'Os nomes de campo estão na primeira linha Gerar nomes de campo automaticamente' (checked), 'Propriedades de arquivo de texto...' (highlighted), 'Converter para União...', 'Duplicar', and 'Remover'. Below the menu, there is a 'Classificar campos' (Sort fields) section with a dropdown set to 'Ordem de fontes de dados' (Order of data sources).



The screenshot shows the 'train.csv' file properties dialog box. It has four settings:

- Separador de campos:** Vírgula
- Qualificador de texto:** "
- Conjunto de caracteres:** UTF-8
- Localidade:** inglês (Estados Unidos)

# Dados lidos pelo Tableau

 train.csv <b>Dates</b>	Abc train.csv <b>Category</b>	Abc train.csv <b>Descript</b>	Abc train.csv <b>DayOfWeek</b>	Abc train.csv <b>PdDistrict</b>	Abc train.csv <b>Resolution</b>	Abc train.csv <b>Address</b>	# train.csv <b>X</b>	# train.csv <b>Y</b>
13/05/15 23:53:00	WARRANTS	WARRANT ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122,425892	37,774599
13/05/15 23:53:00	OTHER OFFENSES	TRAFFIC VIOLATION ...	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122,425892	37,774599
13/05/15 23:33:00	OTHER OFFENSES	TRAFFIC VIOLATION ...	Wednesday	NORTHERN	ARREST, BOOKED	VANNESS AV / GREE...	-122,424363	37,800414
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	NORTHERN	NONE	1500 Block of LOMBA...	-122,426995	37,800873
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	PARK	NONE	100 Block of BRODER...	-122,438738	37,771541
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	INGLESIDE	NONE	0 Block of TEDDY AV	-122,403252	37,713431
13/05/15 23:30:00	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	INGLESIDE	NONE	AVALON AV / PERU AV	-122,423327	37,725138
13/05/15 23:30:00	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	BAYVIEW	NONE	KIRKWOOD AV / DON...	-122,371274	37,727564
13/05/15 23:00:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	RICHMOND	NONE	600 Block of 47TH AV	-122,508194	37,776601
13/05/15 23:00:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEA...	-122,419088	37,807802
13/05/15 22:58:00	LARCENY/THEFT	PETTY THEFT FROM L...	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEA...	-122,419088	37,807802

# Ajustando coordenadas

	Abc
	NTS
	FFENSES
	FFENSES
	Y/THEFT
13/05/15 23:30:00	LARCENY/THEFT
13/05/15 23:30:00	LARCENY/THEFT
13/05/15 23:30:00	VEHICLE THEFT
13/05/15 23:30:00	VEHICLE THEFT
13/05/15 23:00:00	LARCENY/THEFT
13/05/15 23:00:00	LARCENY/THEFT
13/05/15 22:58:00	LARCENY/THEFT

#	#	
		AGUNA ST
		AGUNA ST
		AV / GREE...
		o of LOMBA...
		of BRODER...
		TEDDY AV
		V / PERU AV
		D AV / DON...
		of 47TH AV
		N ST / LEA...
		N ST / LEA...

train.csv	train.csv
X	Y
-122,425892	37,774599
-122,425892	37,774599
-122,424363	37,800414
-122,426995	37,800873
-122,438738	37,771541
-122,403252	37,713431
-122,423327	37,725138
-122,371274	37,727564
-122,508194	37,776601
-122,419088	37,807802
-122,419088	37,807802

# Criando gráficos

Conexões Adicionar

train  
Arquivo de texto

train.csv

Arquivos 🔍

- allvariables.tsv
- consolidado.csv.txt
- có pia de consolidado.csv.txt
- T0z\_2014\_Ne...l3\_5min.csv
- T0z\_2014\_Pic...el3\_5min.csv
- T0z\_2014\_Th...l3\_5min.csv
- T2\_LosGatos\_...el3\_5min.csv
- T3\_LosGatos\_...el3\_5min.csv
- T3\_Thermo\_4...l3\_5min.csv
- train.csv

🔗 Nova união

Classificar campos

train.csv Dates	Abc train.csv Category	Abc train.csv Descript	Abc train.csv DayOfWeek	Abc train.csv PdDistrict	Abc train.csv Resolution	Abc train.csv Address	train.csv X	train.csv Y
13/05/15 23:53:00	WARRANTS	WARRANT ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122,425892	37,774599
13/05/15 23:53:00	OTHER OFFENSES	TRAFFIC VIOLATION ...	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122,425892	37,774599
13/05/15 23:33:00	OTHER OFFENSES	TRAFFIC VIOLATION ...	Wednesday	NORTHERN	ARREST, BOOKED	VANNESS AV / GREE...	-122,424363	37,800414
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	NORTHERN	NONE	1500 Block of LOMBA...	-122,426995	37,800873
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	PARK	NONE	100 Block of BRODER...	-122,438738	37,771541
13/05/15 23:30:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	INGLESIDE	NONE	0 Block of TEDDY AV	-122,403252	37,713431
13/05/15 23:30:00	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	INGLESIDE	NONE	AVALON AV / PERU AV	-122,423327	37,725138
13/05/15 23:30:00	VEHICLE THEFT	STOLEN AUTOMOBILE	Wednesday	BAYVIEW	NONE	KIRKWOOD AV / DON...	-122,371274	37,727564
13/05/15 23:00:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	RICHMOND	NONE	600 Block of 47TH AV	-122,508194	37,776601
13/05/15 23:00:00	LARCENY/THEFT	GRAND THEFT FROM ...	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEA...	-122,419088	37,807802
13/05/15 22:58:00	LARCENY/THEFT	PETTY THEFT FROM L...	Wednesday	CENTRAL	NONE	JEFFERSON ST / LEA...	-122,419088	37,807802

Fonte de dados Planilha 1 + - 🔍

# Planilha

The image shows a Power BI interface with a blank worksheet titled "Planilha 1". The interface is divided into several sections:

- Dados:** Shows a table named "train".
- Análise:** A dropdown menu.
- Páginas:** A section for managing pages.
- Colunas:** A section for managing columns.
- Linhas:** A section for managing rows.
- Dimensões:** A list of dimensions including Address, Category, Dates, DayOfWeek, Descript, PdDistrict, Resolution, and Nomes de medida.
- Medidas:** A list of measures including X, Y, Número de registros, and Valores de medida.
- Filtros:** A section for adding filters.
- Marcas:** A section for adding marks, including options for Cor (Color), Tamanho (Size), Texto (Text), and Dica de ferram. (Field mark).

The main area of the worksheet is currently blank, with the text "Soltar campo aqui" (Drop field here) appearing in two locations, indicating that no fields have been added to the visualization yet.

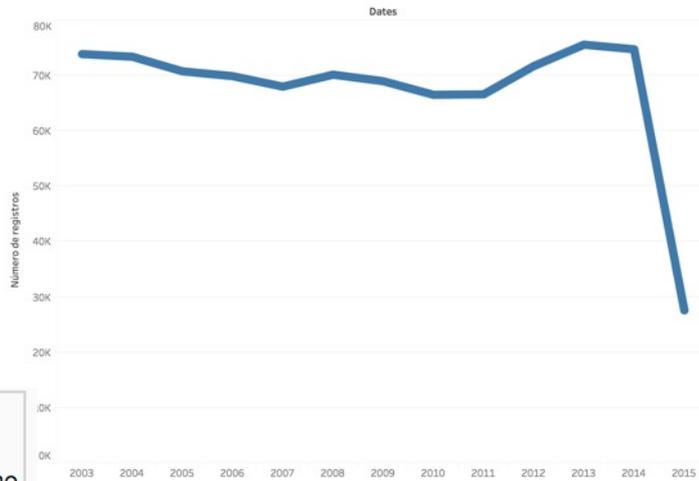
# Criando visualizações simples

This dataset contains incidents derived from SFPD Crime Incident Reporting system. The data ranges from 1/1/2003 to 5/13/2015. The training set and test set rotate every week, meaning week 1,3,5,7... belong to test set, week 2,4,6,8 belong to training set.

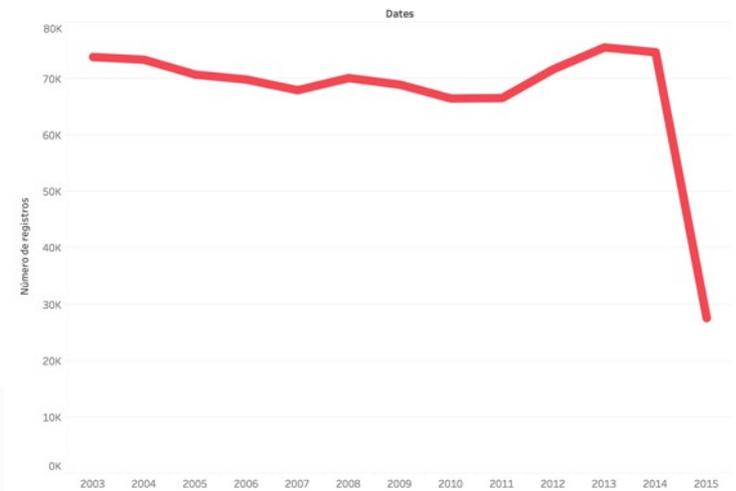


# Alterando características

Número de Crimes por Ano



Número de Crimes por Ano

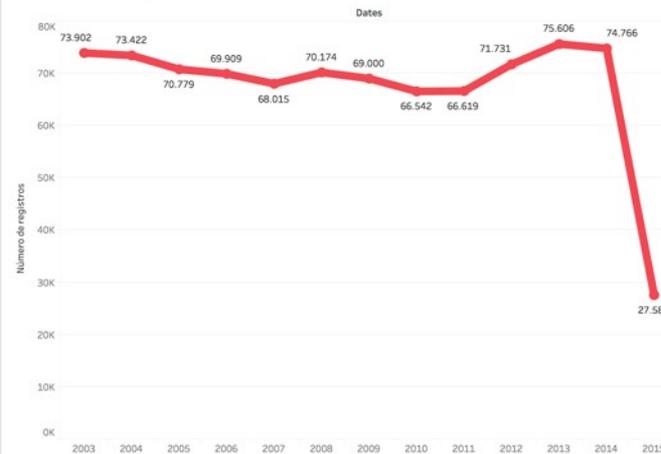


Tamanho



Cor

Número de Crimes por Ano



Rótulo

# Usando o “Mostre-me”

Mostre-me

Para **linhas** (discreto), tente

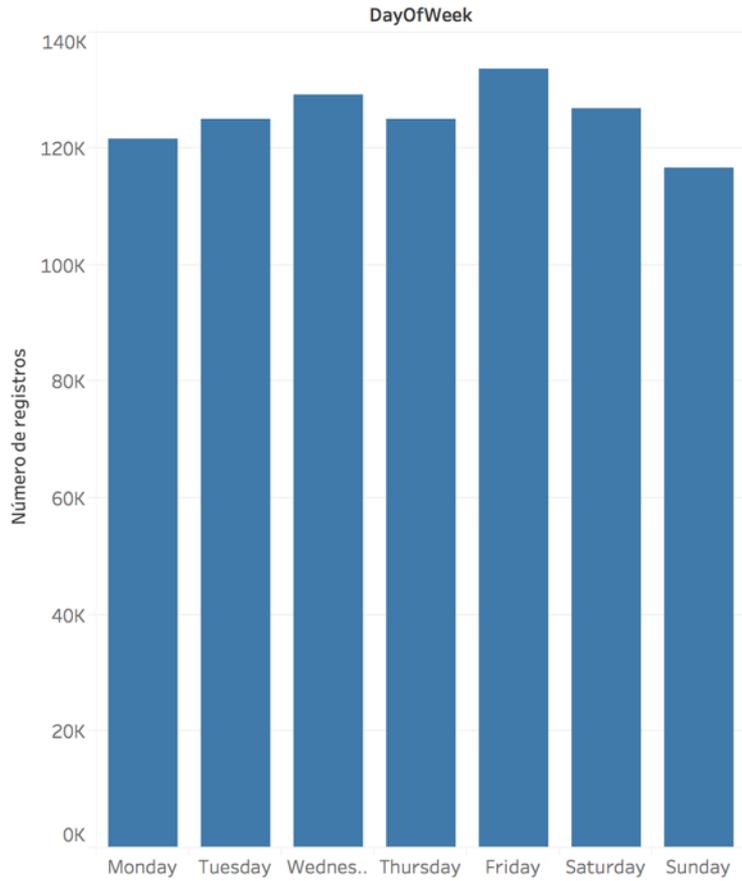
1 data 📅

0 ou mais **Dimensões**

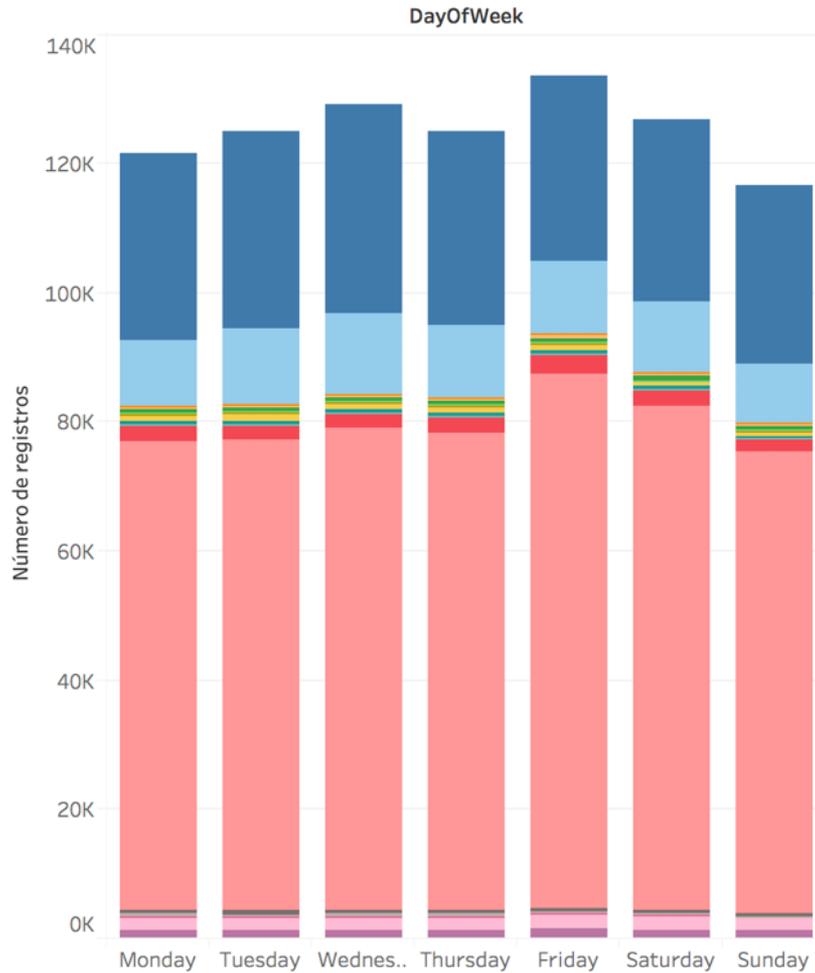
1 ou mais **Medidas**

iii Colunas	DayOfWeek
☰ Linhas	SOMA(Número de reg..)

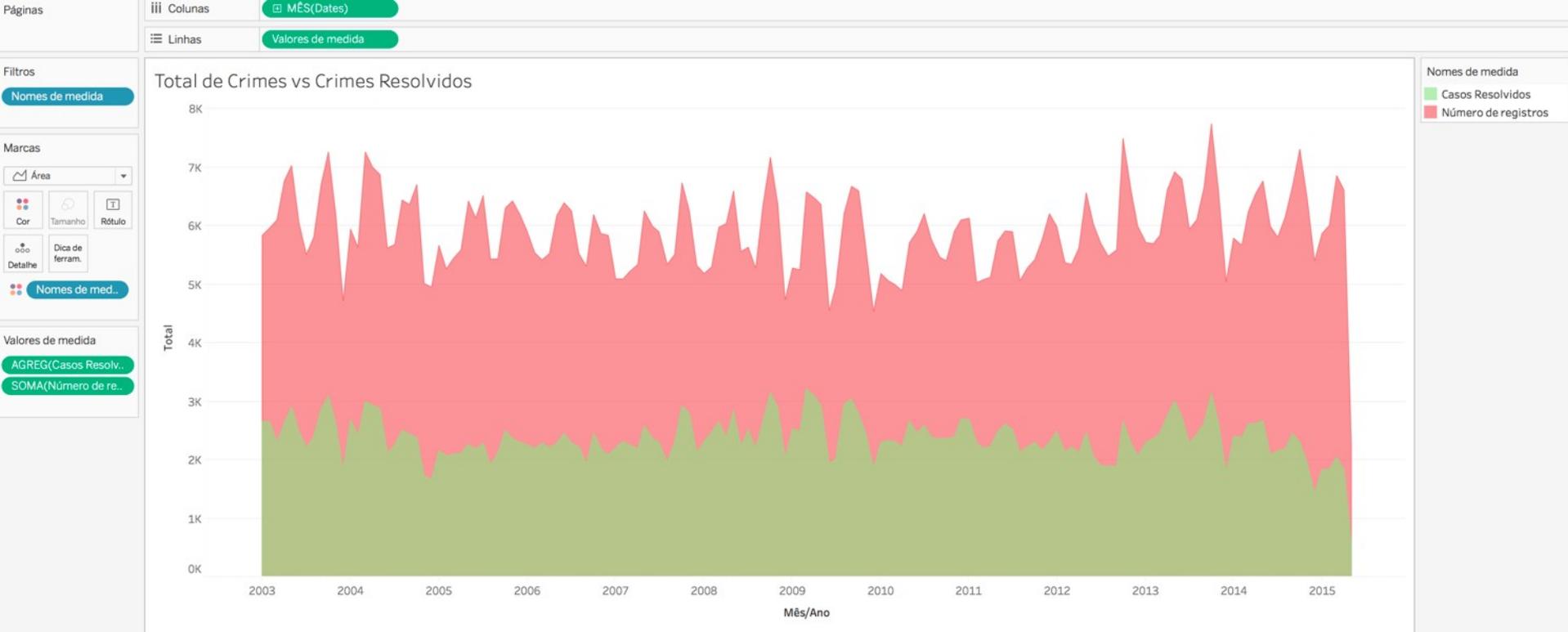
### Número crimes / dia da semana



# Número crimes / dia da semana



- Resolution
- ARREST, BOOKED
  - ARREST, CITED
  - CLEARED-CONTACT JUVENILE FOR MORE INFO
  - COMPLAINANT REFUSES TO PROSECUTE
  - DISTRICT ATTORNEY REFUSES TO PROSECUTE
  - EXCEPTIONAL CLEARANCE
  - JUVENILE ADMONISHED
  - JUVENILE BOOKED
  - JUVENILE CITED
  - JUVENILE DIVERTED
  - LOCATED
  - NONE
  - NOT PROSECUTED
  - PROSECUTED BY OUTSIDE AGENCY
  - PROSECUTED FOR LESSER OFFENSE
  - PSYCHOPATHIC CASE
  - UNFOUNDED



- Medidas**
- Casos Resolvidos
  - ⊕ X
  - ⊕ Y
  - Número de registros
  - Valores de medida

Casos Resolvidos

SUM(IF[Resolution] <> "NONE" THEN 1 ELSE 0 END)

O cálculo é válido.

Planilhas afetadas

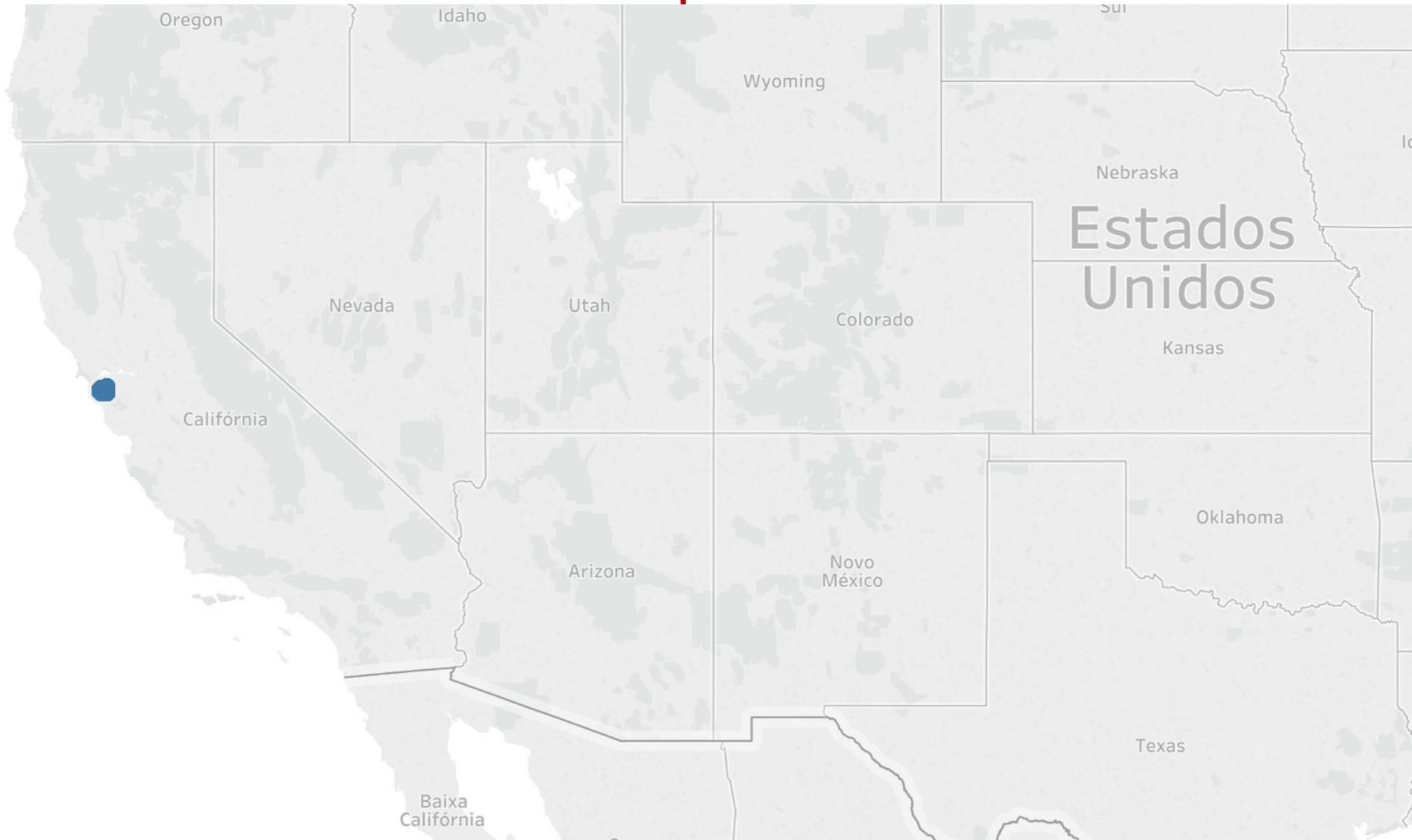
Aplicar OK

Análise Mapa Formatar Servidor Janela Aj

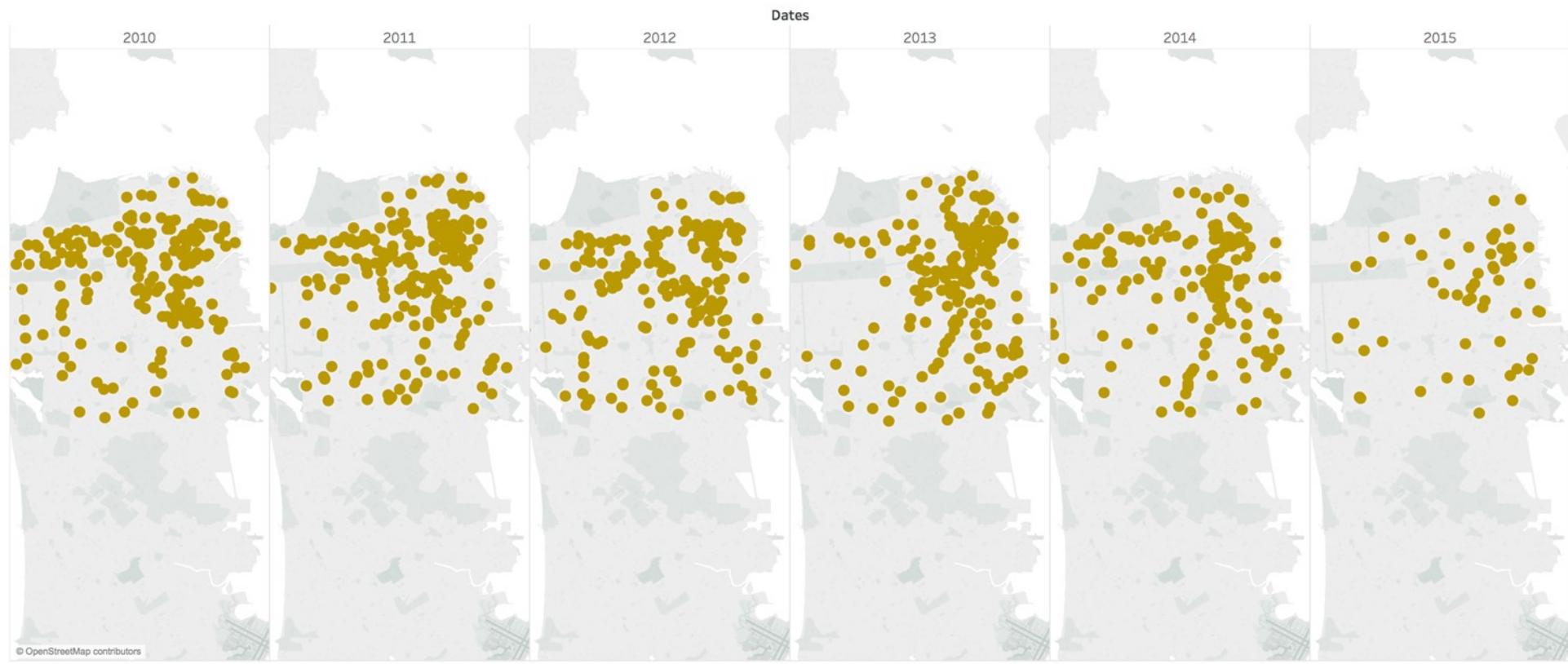
- Mostrar rótulos de marcas
- ✓ Agregar medidas
- Empilhar marcas Automático
- Visualizar dados... Ativado
- Revelar dados ocultos Desativado
- Porcentagem de
- Totais
- Previsão
- Linhas de tendência
- Valores especiais
- Layout da tabela
- Legendas
- Filtros
- Destaque
- Parâmetros
- Crear campo calculado...
- Editar campo calculado
- Girar campos
- Alternar linhas e columnas ^%W



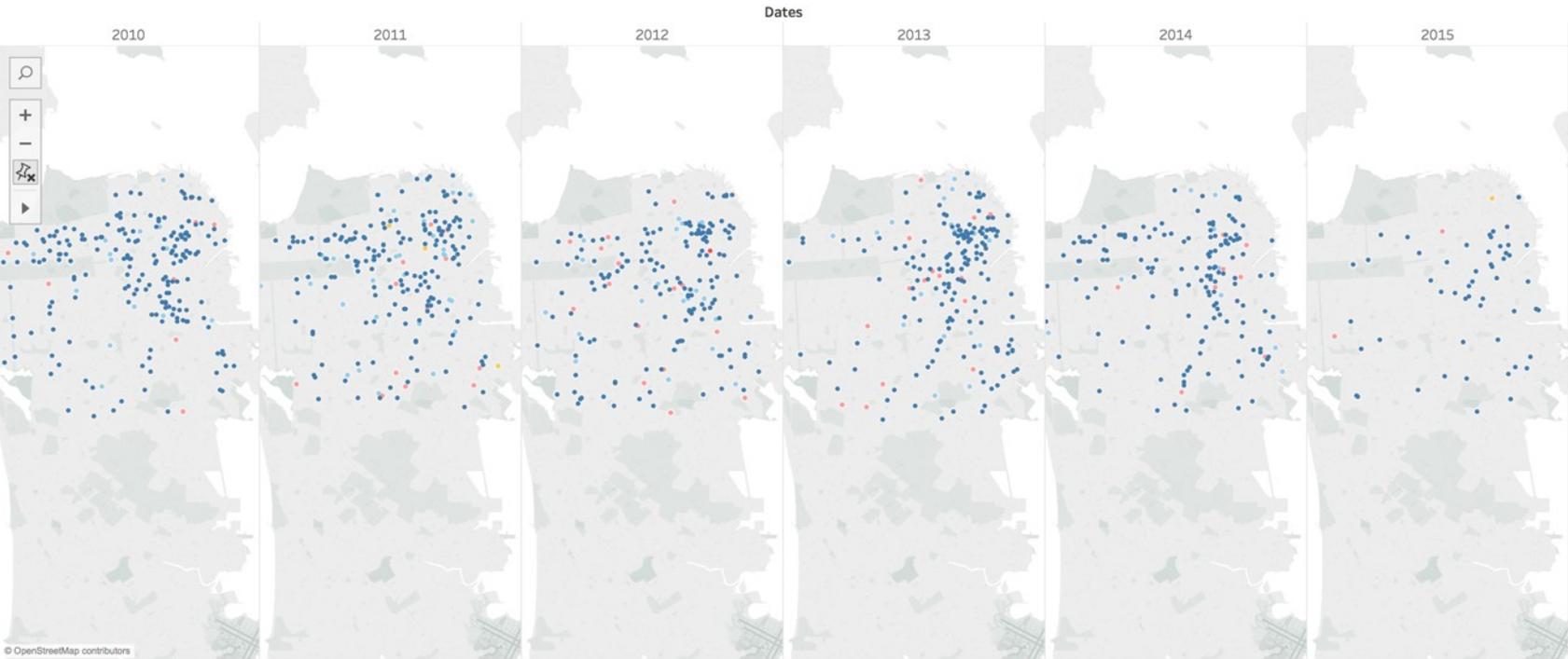
# Mapas



# Driver under the influence of drugs - 2010 to 2015



# Driver under the influence of drugs - 2010 to 2015



**ANO(Dates)**

- (Tudo)
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015

**Category**

- DRIVING UNDER THE I..

**Resolution**

- ARREST, BOOKED
- ARREST, CITED
- JUVENILE BOOKED
- LOCATED
- NONE
- UNFOUNDED

Páginas

Colunas ANO(Dates) X

Linhas Y

Filtros

ANO(Dates)

Marcas

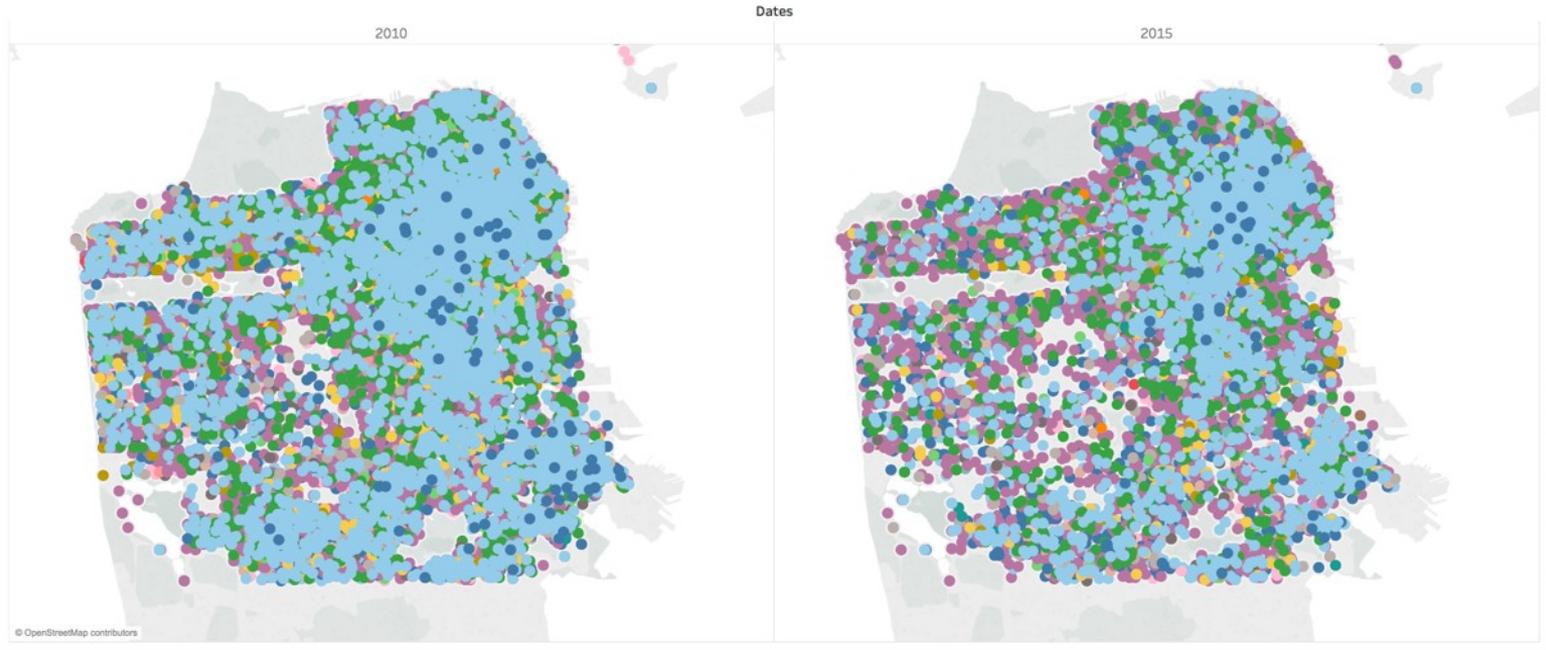
Automático

Cor Tamanho Rótulo

Detalhe

Category

### Category of Crimes - 2010 and 2015



- ANO(Dates)
- (Tudo)
  - 2003
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2011
  - 2012
  - 2013
  - 2014
  - 2015

- Category
- ARSON
  - ASSAULT
  - BAD CHECKS
  - BRIBERY
  - BURGLARY
  - DISORDERLY COND..
  - DRIVING UNDER TH..
  - DRUG/NARCOTIC
  - DRUNKENNESS
  - EMBEZZLEMENT
  - EXTORTION
  - FAMILY OFFENSES
  - FORGERY/COUNTE..
  - FRAUD
  - GAMBLING

Manilha 2

Páginas

Colunas ANO(Dates) X

Linhas Arrested Y

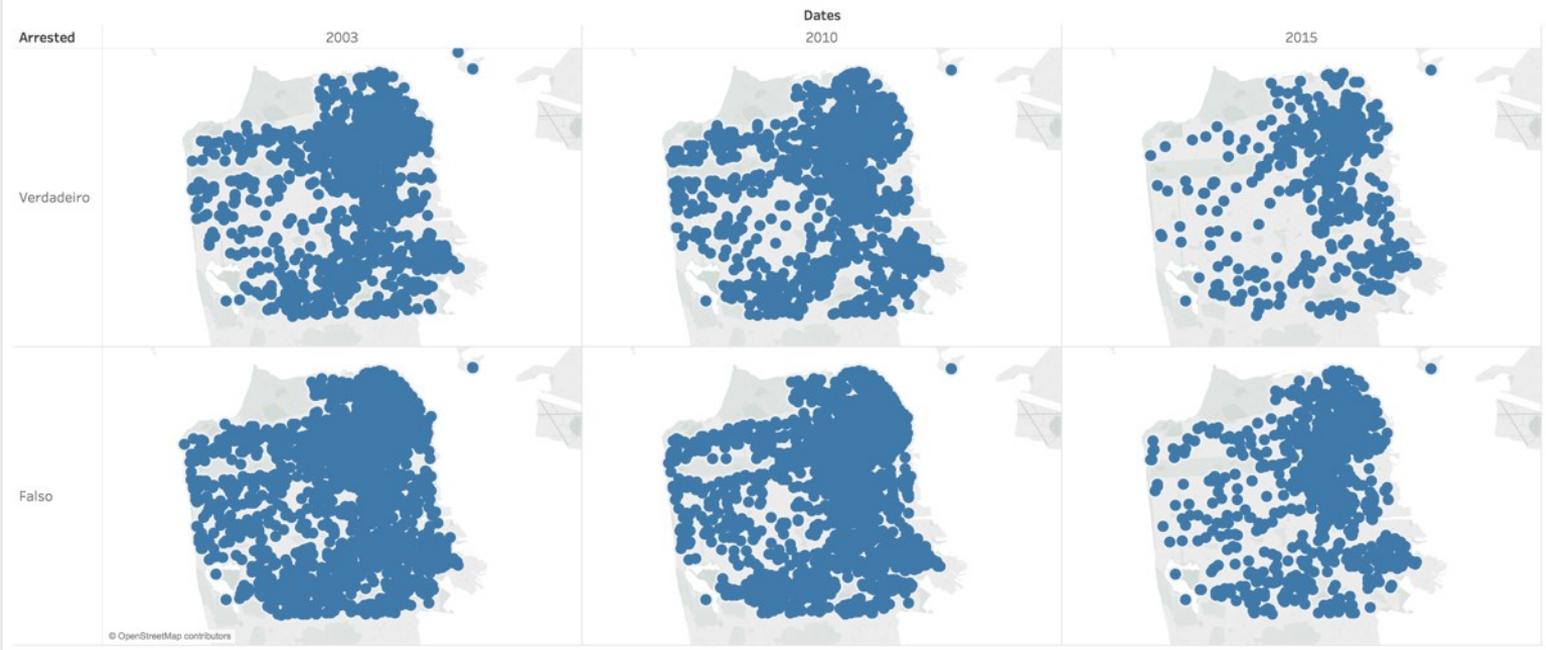
Filtros

ANO(Dates)  
Category: ASSAULT

Marcas

Automático  
Cor Tamanho Rótulo  
Detalhe Dica de ferram.  
Category  
Resolution

### Category of Crimes - 2010 and 2015



ANO(Dates)

- (Tudo)
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015

Category

- ASSAULT

Planilha 2

# Referências

- Sultanum, Nicole. Introdução a Visualização de Dados: Teoria e Prática. Jornada de Informática na Biodiversidade. - LNCC – 2015.
- Zoss, Angela. Practical Data Visualuization. COMPSCI 216: Everything Data, 2015.
- Hearts, Marti. Information Visualization: Principles, Promise, and Pragmatics
- Chen et al. Visualizing Data.
- Volinsky, Chris. Exploratory data Analysis and Data Visualization
- McGuffin, Michael. An Overview of Information Visualization.
- Christoph, K. Tableau Tutorial.

# Libraries and Software

- InfoVis Toolkit (Java library, 2004/2005) <http://ivtk.sourceforge.net/>
- Prefuse (Java library, 2004/2005) <http://prefuse.org/> and <http://vimeo.com/19278481> and Flare (ActionScript library, 2008) <http://flare.prefuse.org/>
- Protovis (JavaScript library, 2009) <http://vis.stanford.edu/protovis/> and <http://eagereyes.org/tutorials/protovis-primer-part-1>
- **D3** (JavaScript library, 2011) <http://d3js.org/> and <http://vimeo.com/29862153>
- Many Eyes (website for sharing visualizations of data) <http://maneyeyes.alphaworks.ibm.com/>
- For visualizing graphs
  - **Tulip** (software application) <http://www.tulip-software.org/>
  - **Gephi** (software application) <http://gephi.org/>
  - **WebCoLa** / cola.js (JavaScript library) <http://marvl.infotech.monash.edu/webcola/>
  - Pajek (software application) <http://pajek.imfm.si/>
  - Walrus (software application) <http://www.caida.org/tools/visualization/walrus/>
  - **NetworkX** (Python library) <http://networkx.lanl.gov/>
  - OGDF (C++ library) <http://www.ogdf.net>
  - Boost Graph Library (BGL) (C++ library) <http://www.boost.org/doc/libs/release/libs/graph/>
  - **Graphviz** (applications and libraries) <http://www.graphviz.org/>
  - **JUNG** (Java library) <http://jung.sourceforge.net/>
  - JGraph (Java library) <http://www.jgraph.com/>
  - yFiles (Java library) <http://www.yworks.com>
- For visualizing multidimensional data
  - **Tableau** (software application) <http://www.tableausoftware.com>
  - **plot.ly** (JavaScript library) <https://plot.ly/>
  - **ggplot2** (R library) <http://ggplot2.org/>
  - **matplotlib** (Python library) <http://matplotlib.org/>
  - **Google Charts** (JavaScript library) <https://developers.google.com/chart/>
  - NVD3 (JavaScript library) <http://nvd3.org/>
  - XmdvTool (software application) <http://davis.wpi.edu/~xmdv/>
  - Spotfire (software application) <http://www.cs.umd.edu/hcil/spotfire/>
  - Polaris (software application) <http://window.stanford.edu/projects/polaris/>
  - Mondrian (software application) <http://rosuda.org/Mondrian>

<https://github.com/amycesal/dataviz-style-guide/blob/master/Sunlight-StyleGuide-DataViz.pdf>