

An abstract graphic on the left side of the slide. It features a dark background with several glowing, curved lines in shades of orange, yellow, and purple. These lines intersect at several points, some of which are marked with small, glowing white circles, resembling a network or data visualization.

# ADMINISTRAÇÃO

IBM0112 DATA MINING

## Tipos de visualizações de dados

Cassius Figueiredo



Visualizando  
distribuições

The background features several abstract geometric elements: a large teal circle on the right, a medium grey circle in the upper left, a blue square outline on the left, a blue line segment at the top, and several grey dashed lines in the lower left. The word "Histograma" is written in white on the teal circle.

Histograma

# O histograma

---

- É um gráfico de barras que apresenta uma distribuição de frequências (quantidade de ocorrências) de dados em intervalos distintos, chamados de caixas.

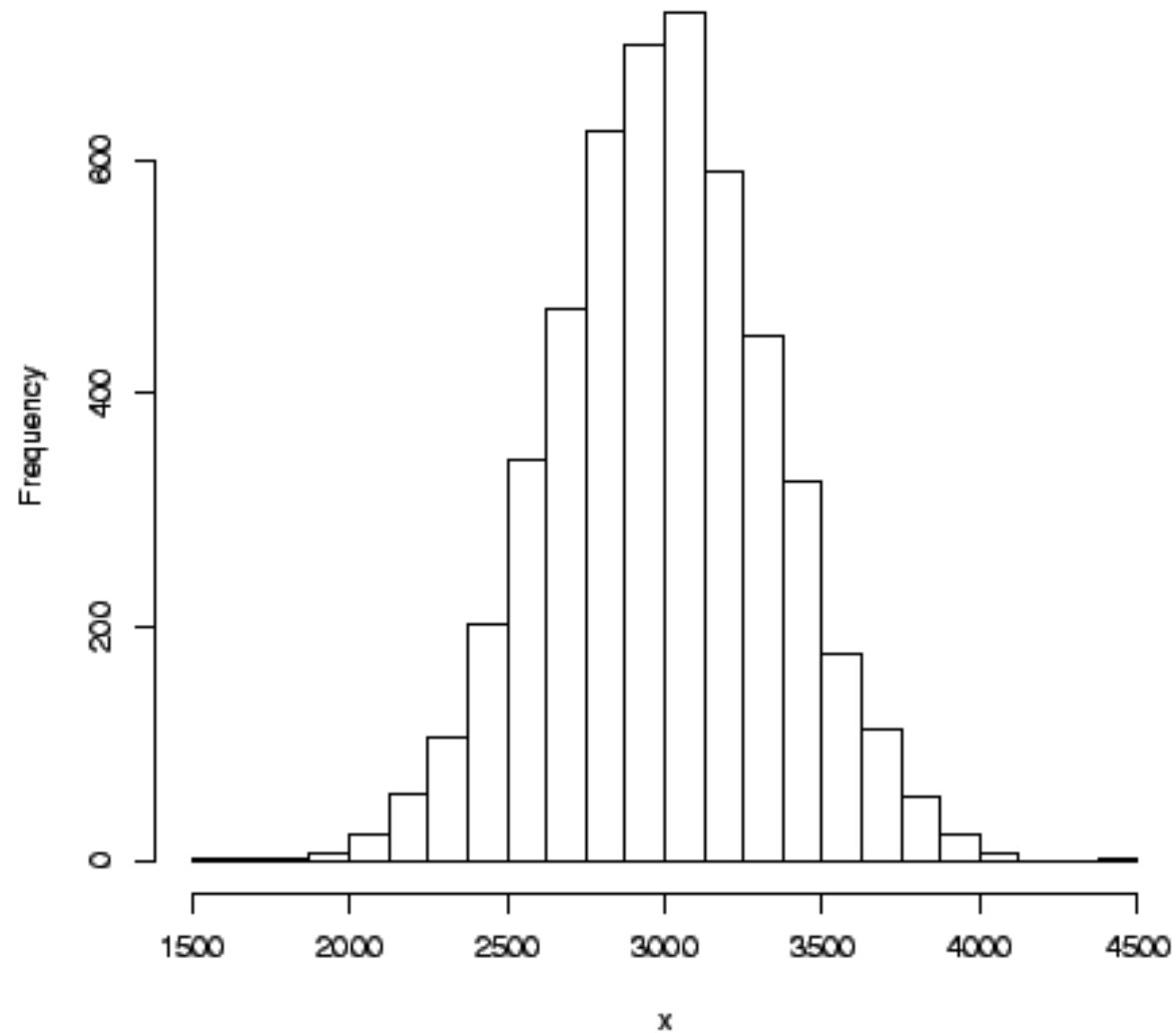
# Como criamos o histograma

---

- Definimos intervalos de valores que irão delimitar nossas “caixas”, onde classificaremos cada um dos dados da variável em estudo.
- Pegamos cada um dos valores e colocamos na caixa apropriada.
- Ao final, teremos uma visualização assim...

# O histograma

---

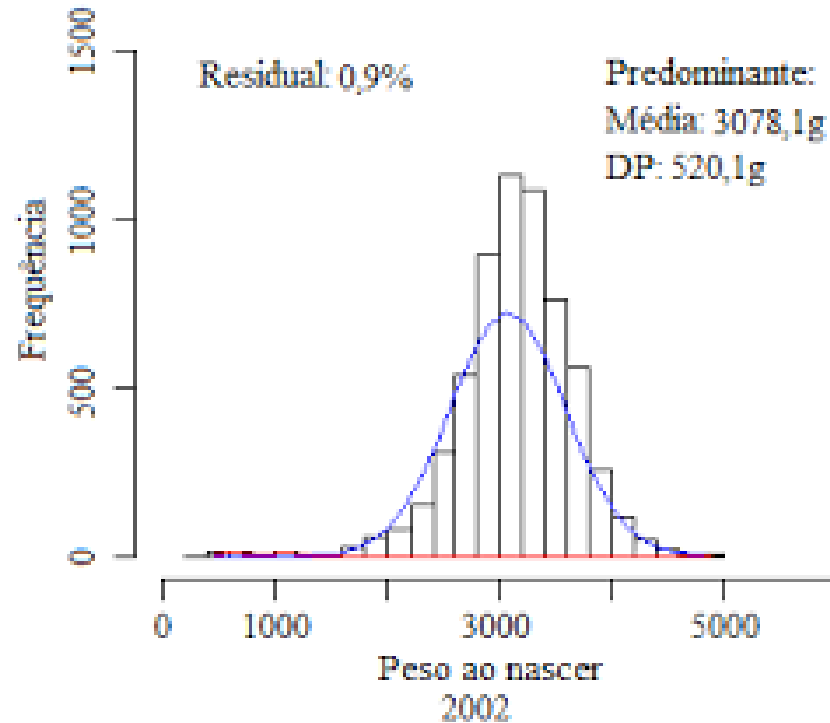


# A importância do histograma

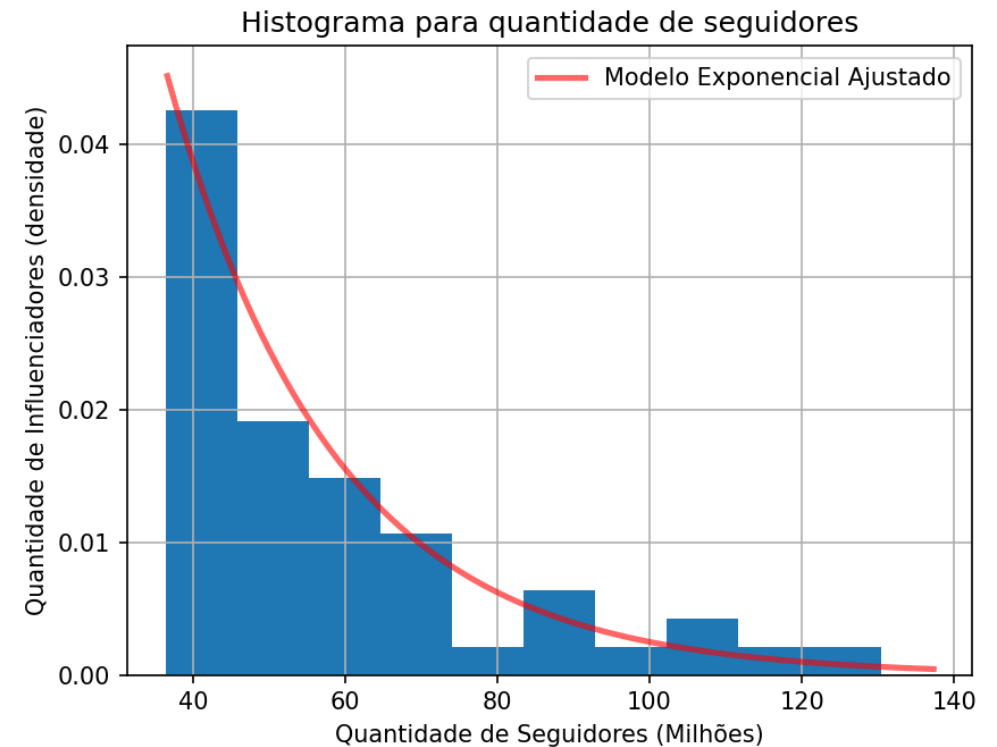
---

- Visualização que nos ajuda a fazer uma avaliação do comportamento estatístico do dado em análise.

# A importância do histograma



Fontes: Associação Brasileira de Estudos Populacionais e  
<https://teobcalvo.wordpress.com/2021/12/23/porque-usar-distribuicao-exponencial/>







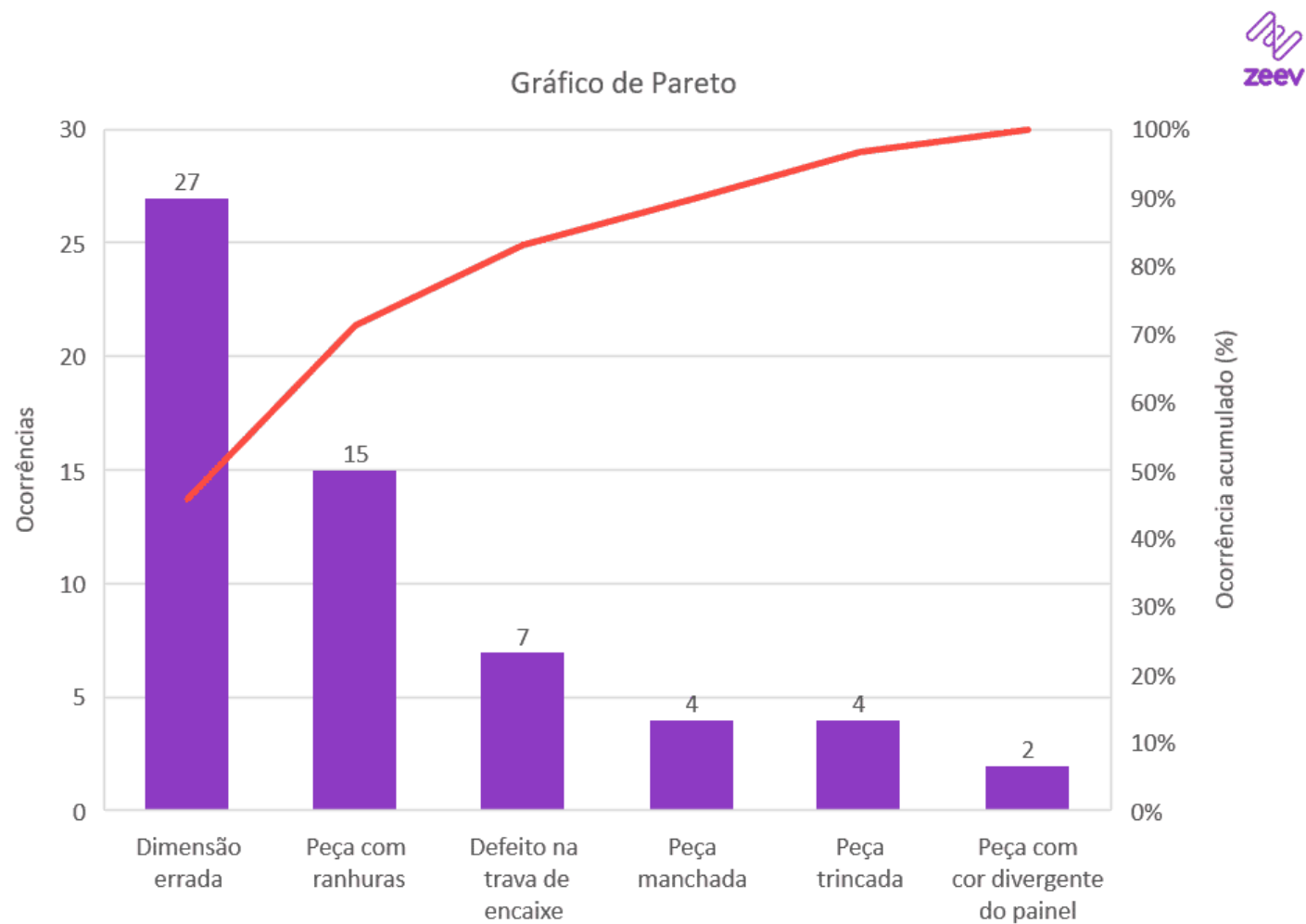
# Gráfico de Pareto

# O gráfico de Pareto


---

- É um histograma adaptado para que possamos avaliar os impactos de determinados aspectos de um problema e definirmos abordagens de solução.

# O histograma



Fonte: <https://blog.zeev.it/diagrama-de-pareto/>



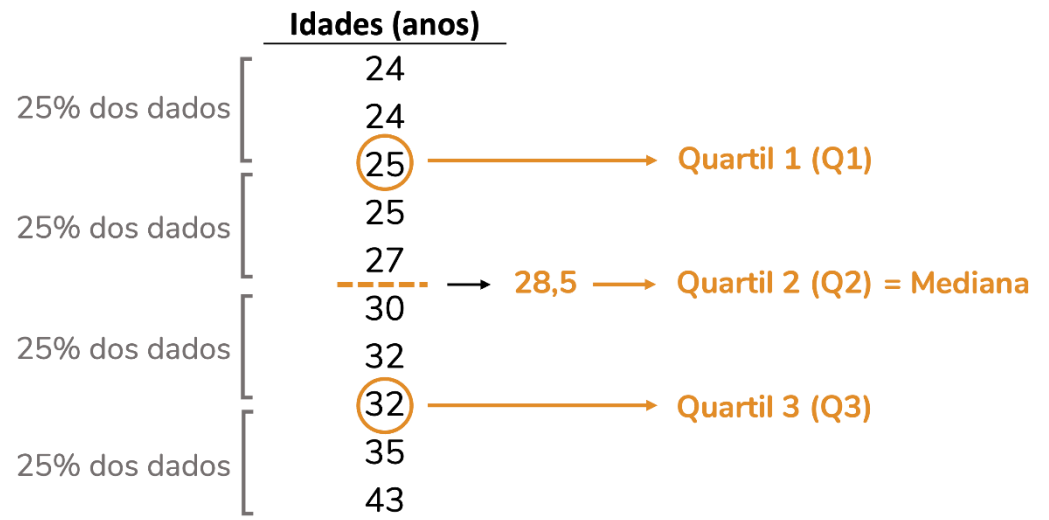
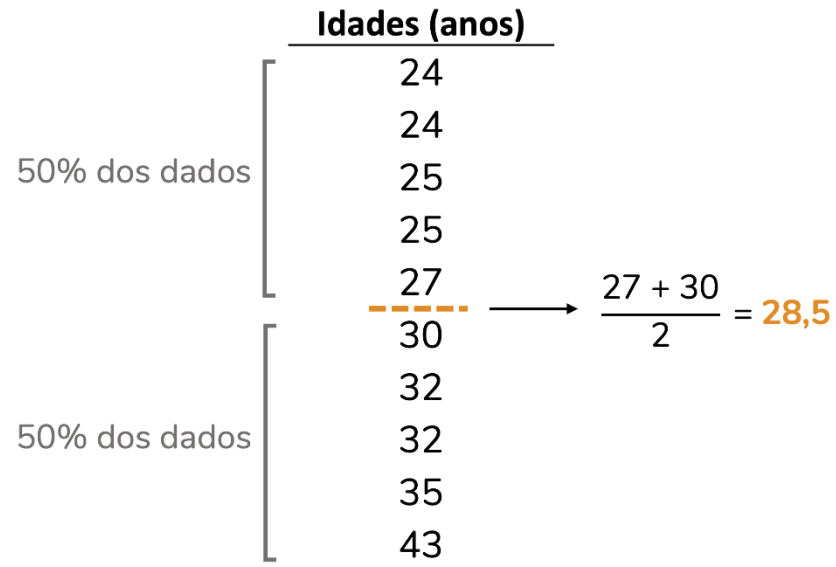
Boxplot ou  
Box and Whiskers

# Boxplot ou Box and Whiskers

---

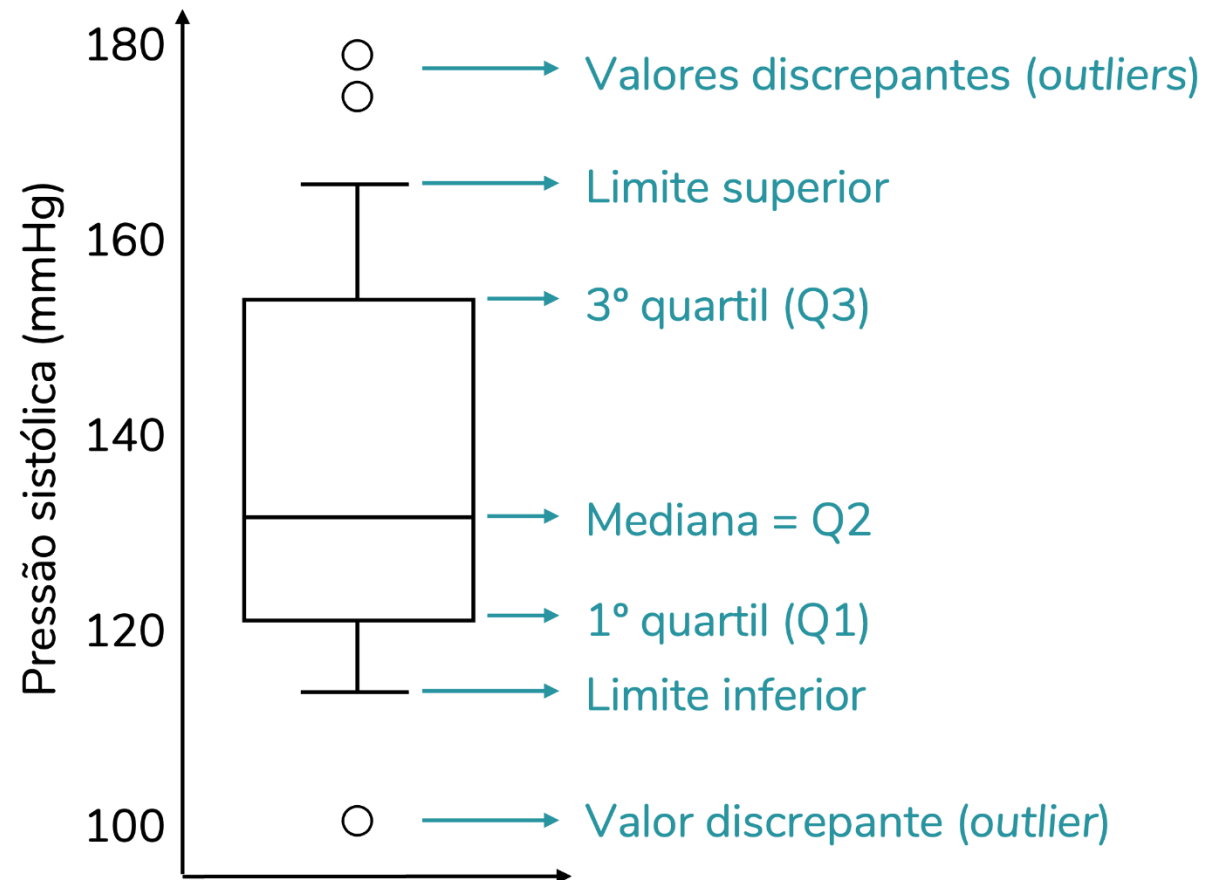
- São visualizações que trazem em uma única imagem diversas características estatísticas:
  - Mediana
  - Valores extremos (máximo e mínimo)
  - Concentração de valores
  - Distribuição dos dados
  - Outliers

# Boxplot ou Box and Whiskers



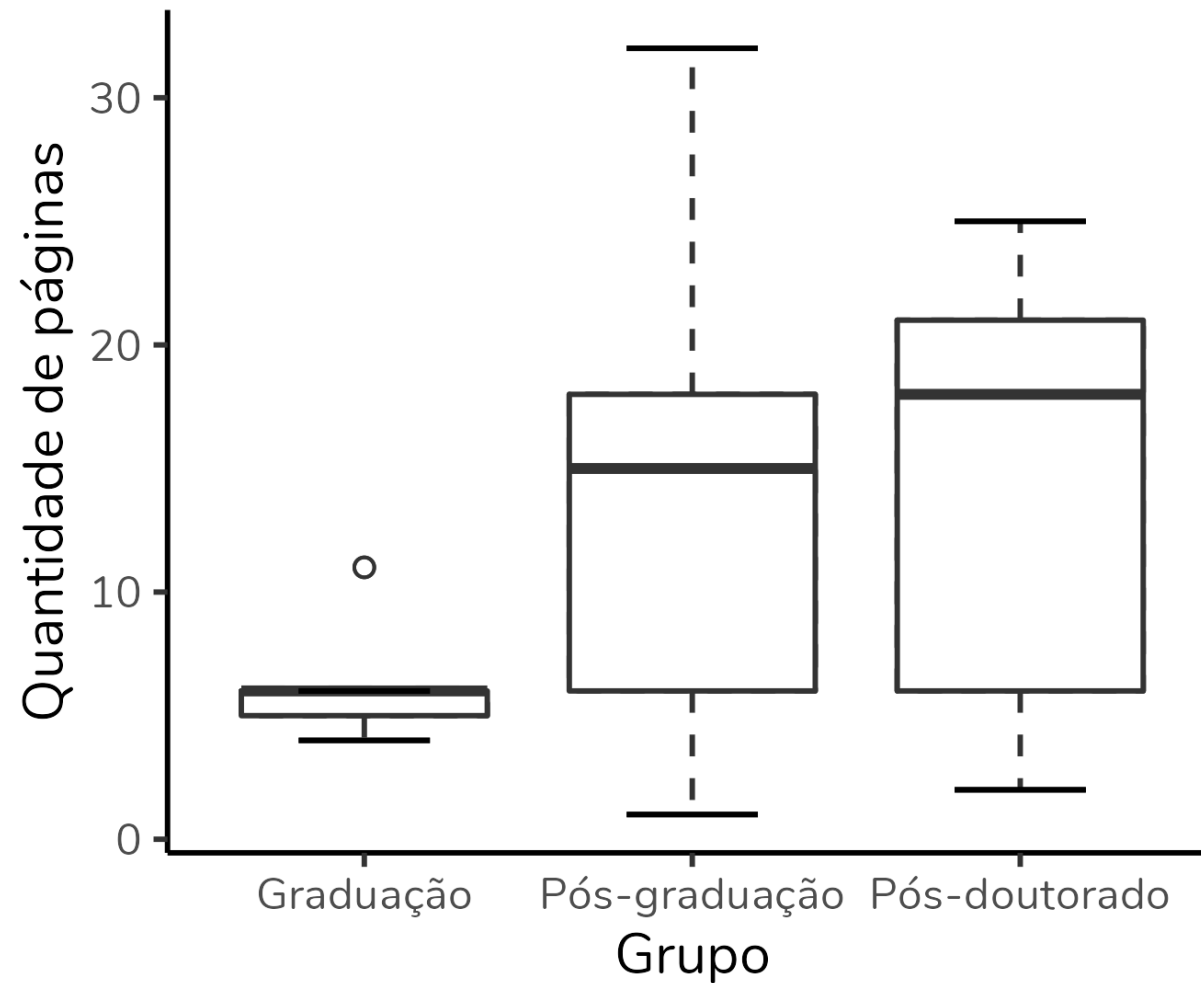
# Boxplot ou Box and Whiskers

---



# Boxplot ou Box and Whiskers

---







Comparando  
informações



# Gráficos de barras

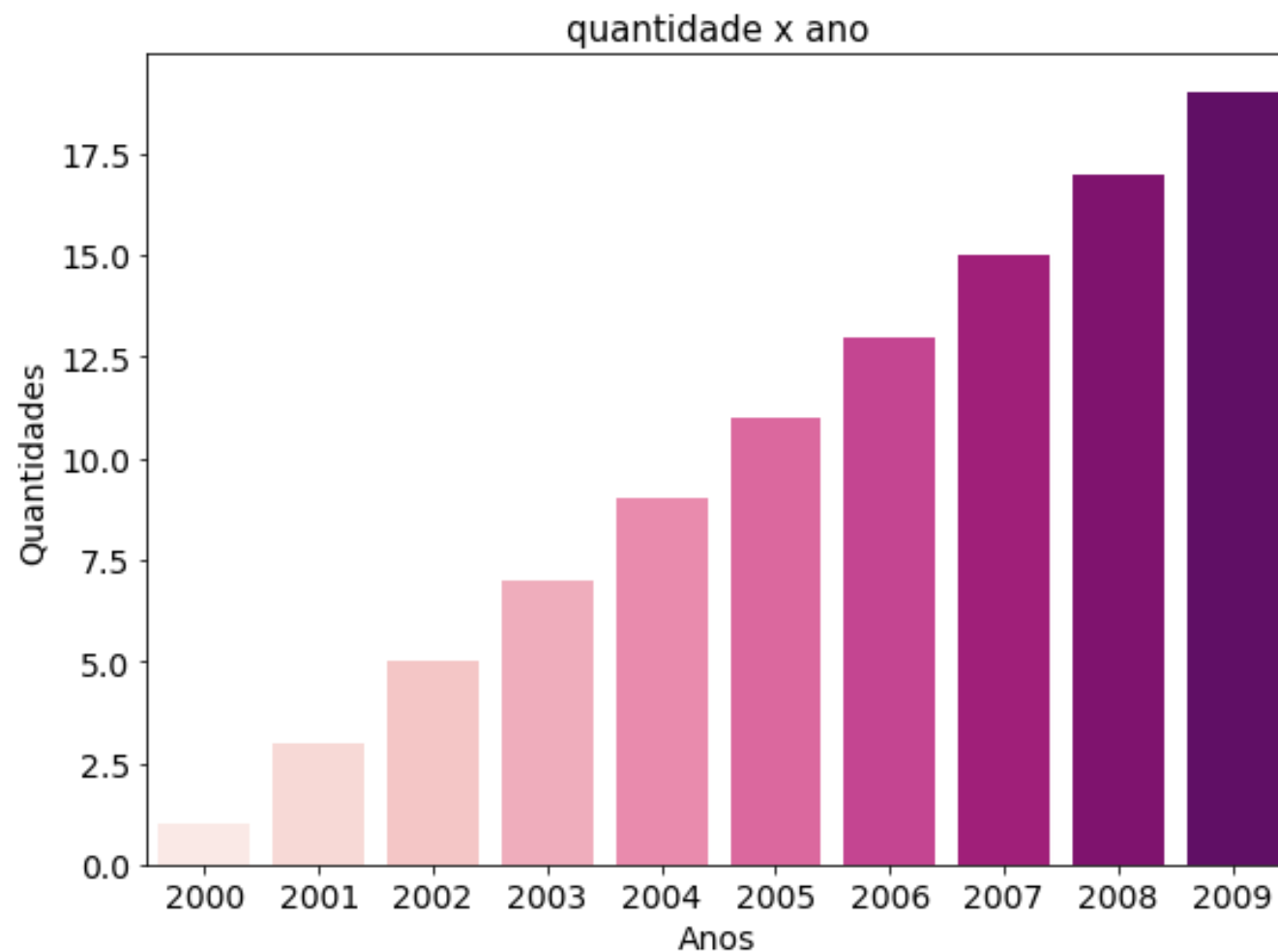
# O gráfico de barras

---

- A forma mais simples e direta de comparar dados/informações é por meio de um gráfico de barras.
- É importante identificar o(s) elemento(s) chave que queremos abordar em nossa visualização.
- Imaginem a situação onde queremos apresentar a evolução de vendas de um determinado produto no decorrer de um período de tempo.

# Gráfico de barras

---



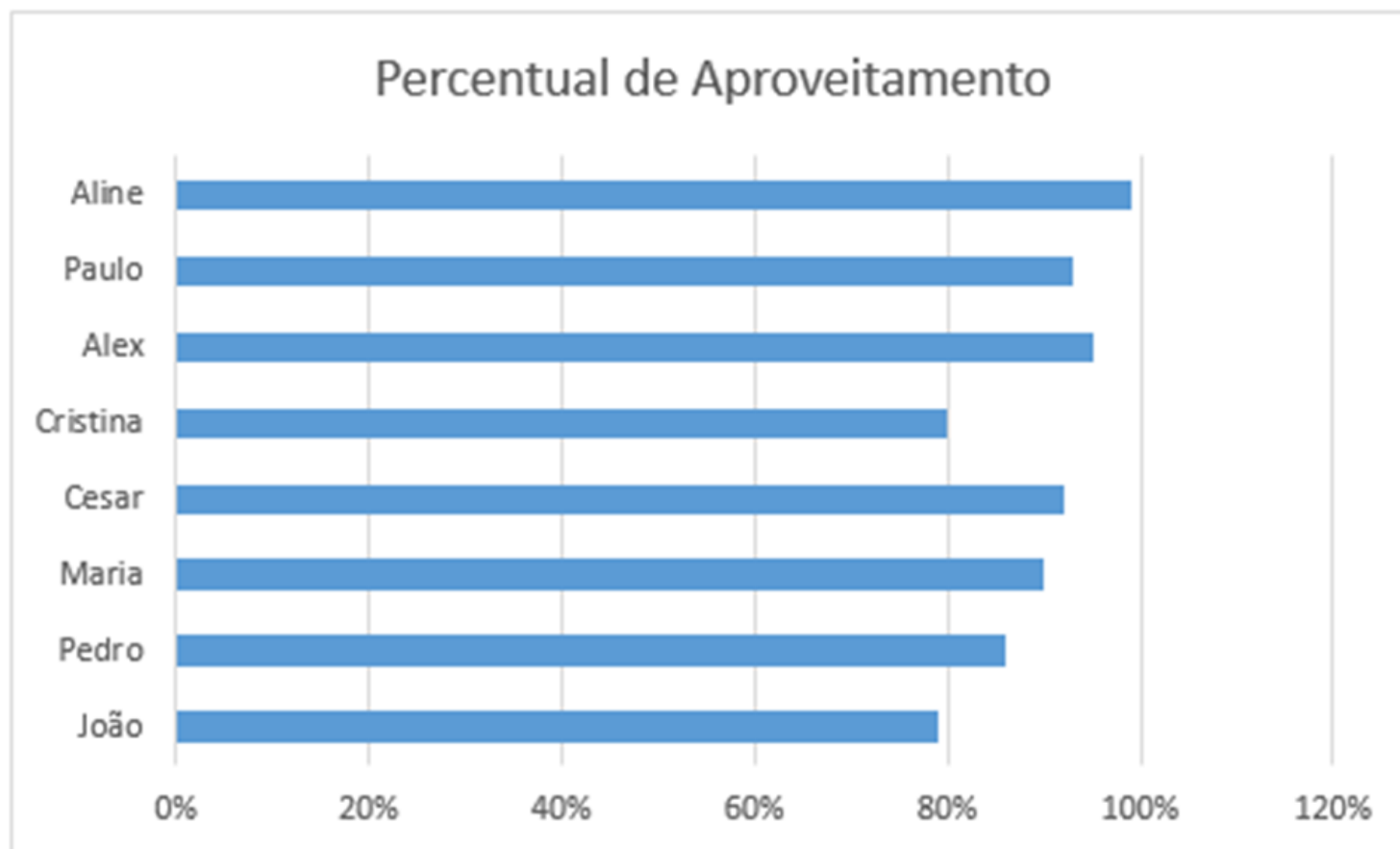
Fonte: autor desconhecido

# Variações do gráfico de barras

---

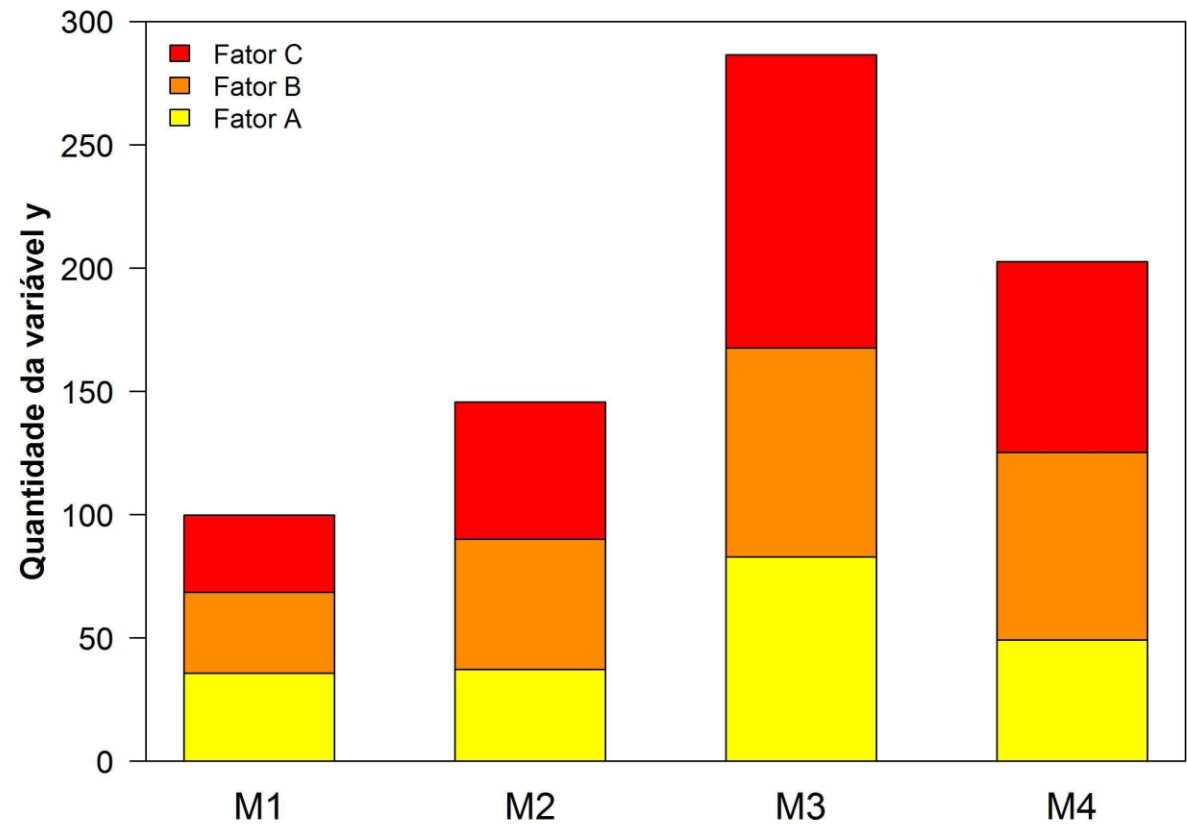
- E se quisermos indicar o aproveitamento de alunos em uma determinada avaliação?
- Como estamos identificando os alunos por seus nomes pode ser mais apropriado apresentar as barras horizontalmente.

# O gráfico de barras horizontais

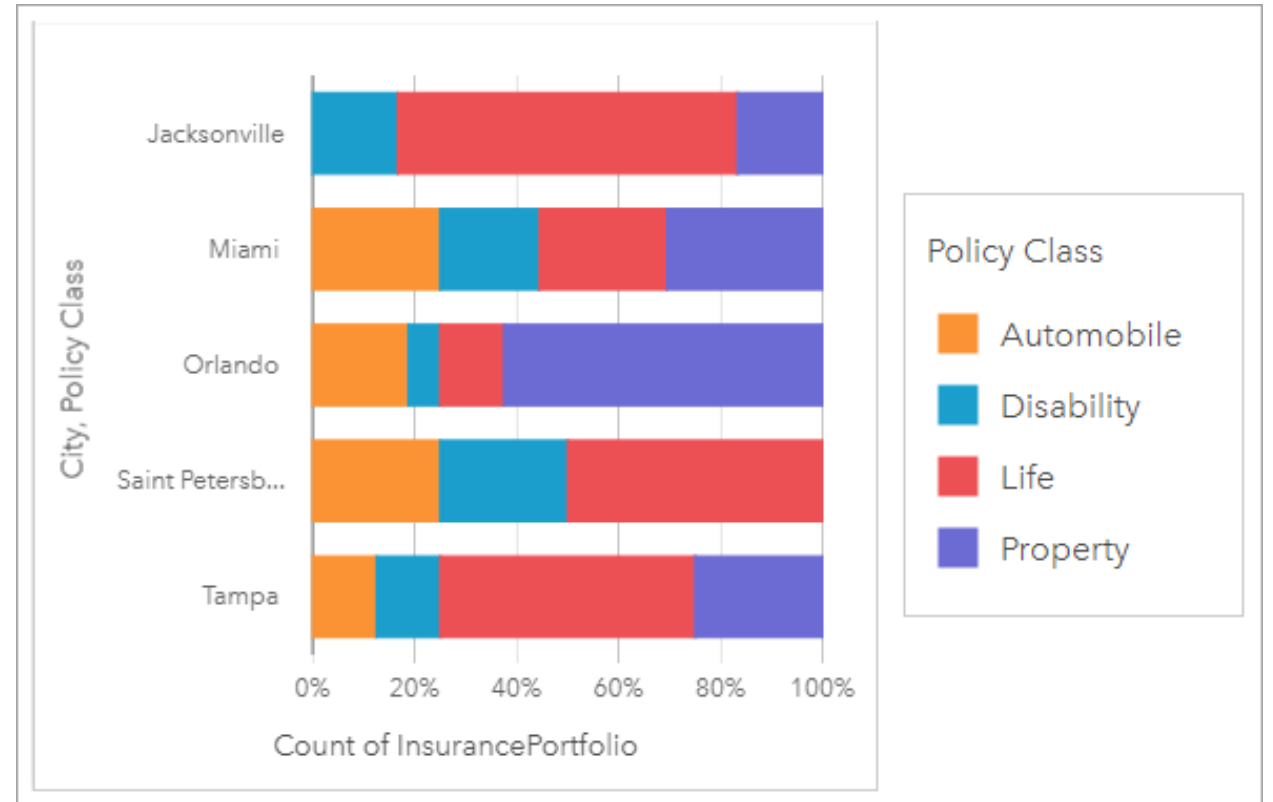


# Gráfico de barras empilhadas

- Um outro cenário é termos que passar mais de uma informação por período/barra.



# Gráfico de barras empilhadas 100%



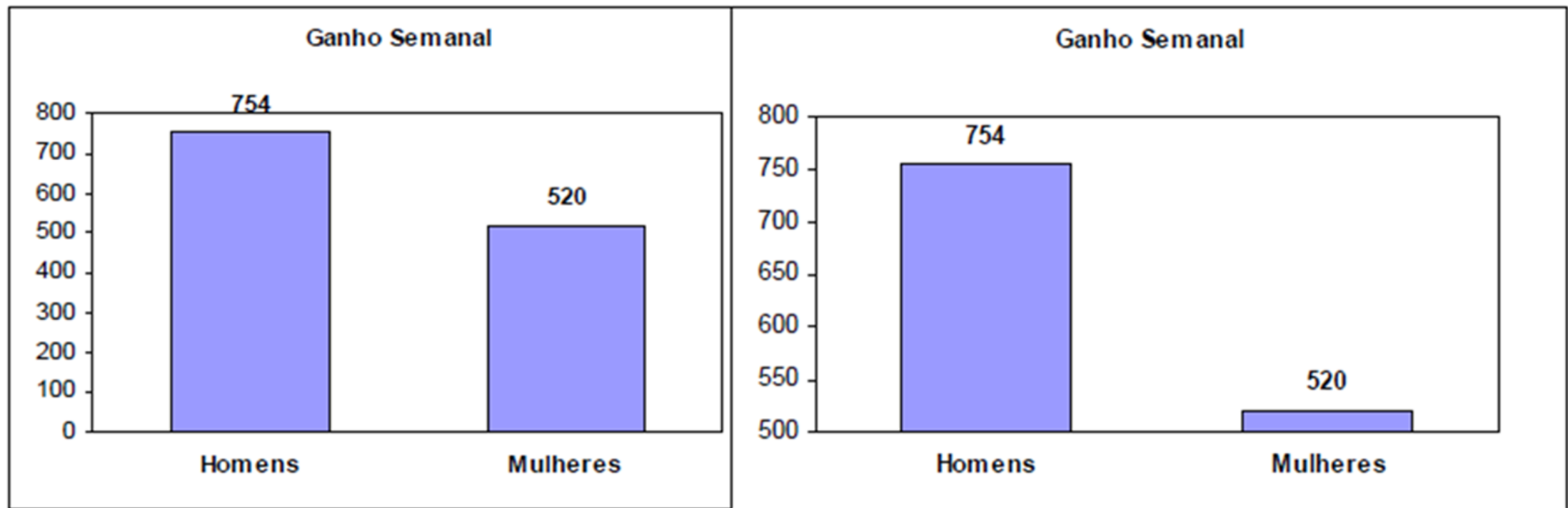


# ATENÇÃO!

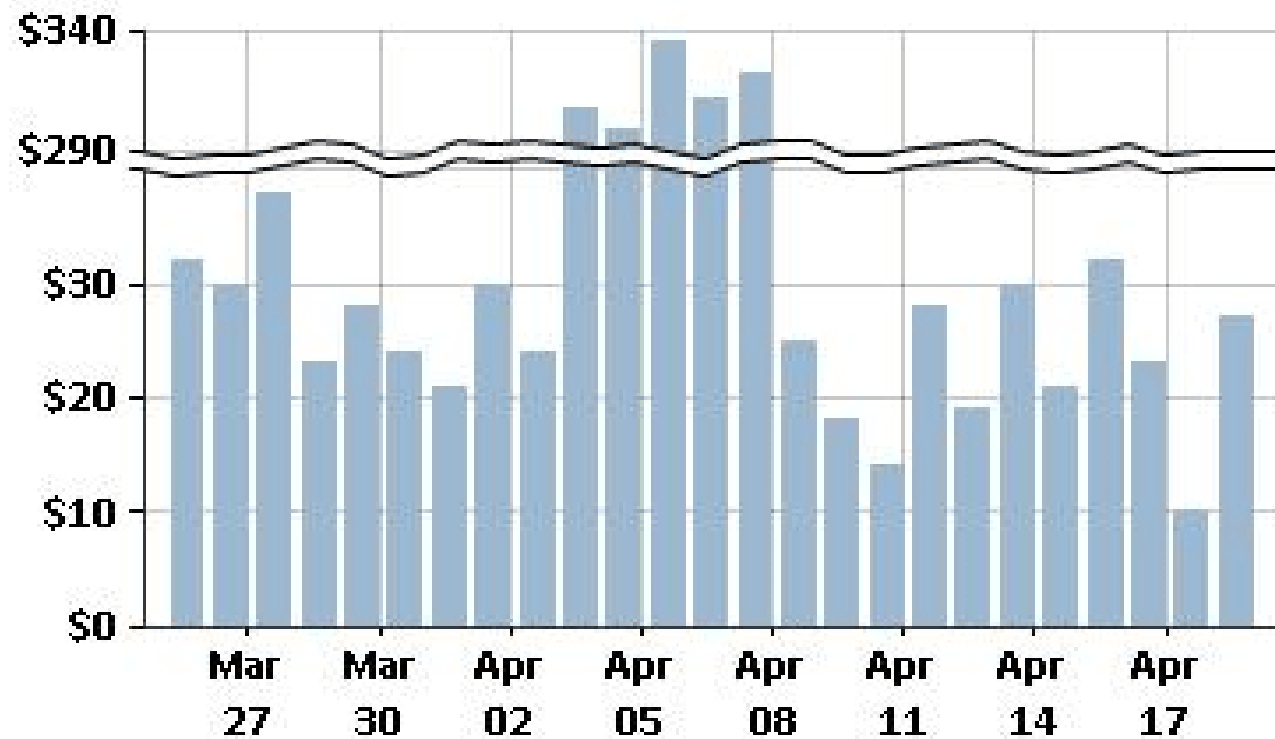
---

- Devemos sempre apresentar a informação de forma correta e justa!

# Atenção!

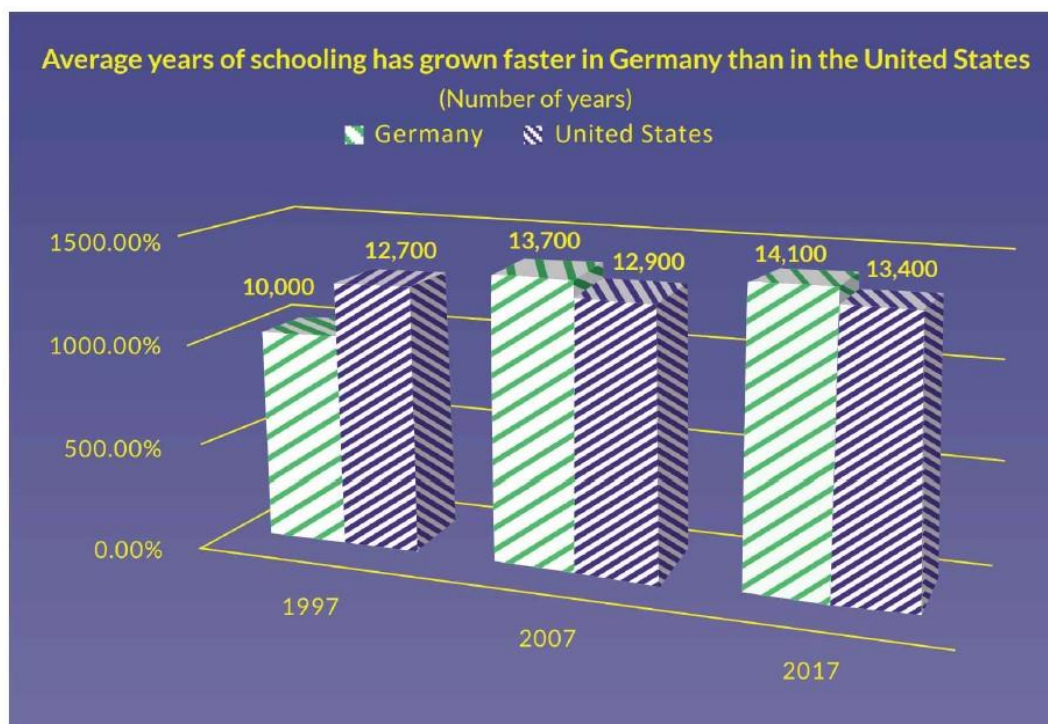


# Atenção!

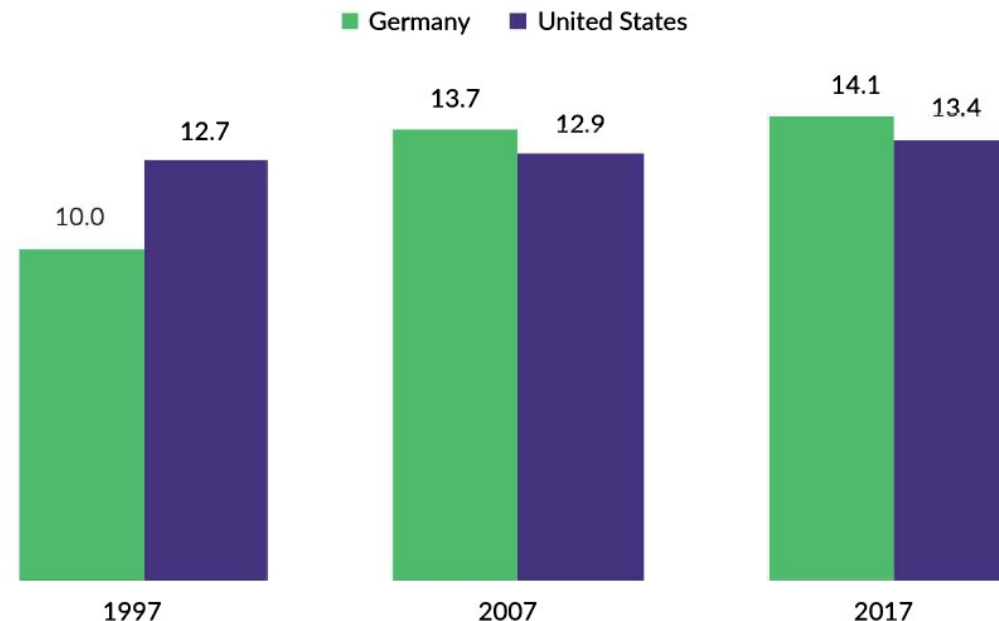


Fonte: autor desconhecido

# Evitar a poluição visual e excesso de informações



**Average years of schooling has grown faster in Germany than in the United States**  
(Number of years)





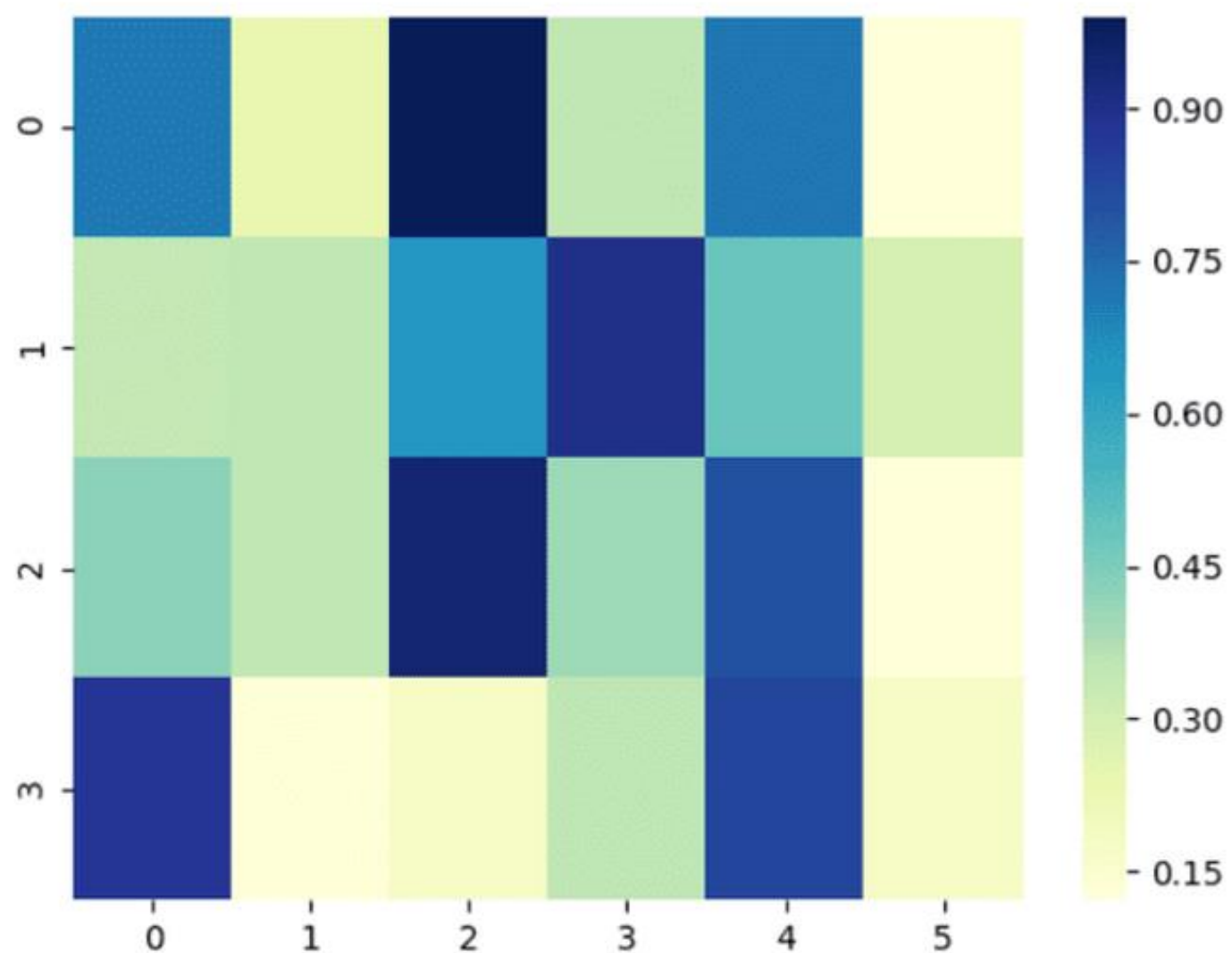
Mapas de calor

# Mapas de calor

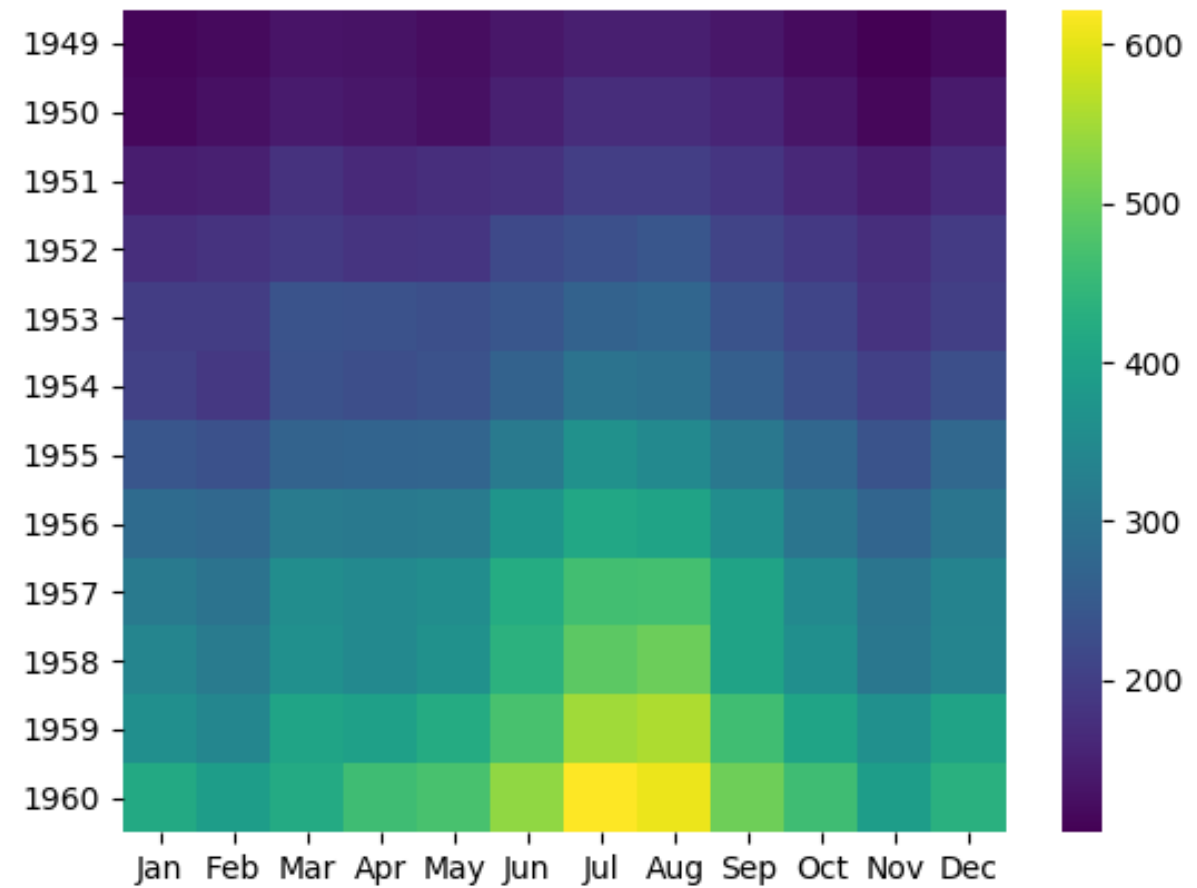
---

- Os mapas de calor utilizam cores para indicar comportamento dos dados.
- São muito utilizados quando queremos dar ênfase a padrões dos dados, no lugar do próprio dado em si.

# Mapas de calor

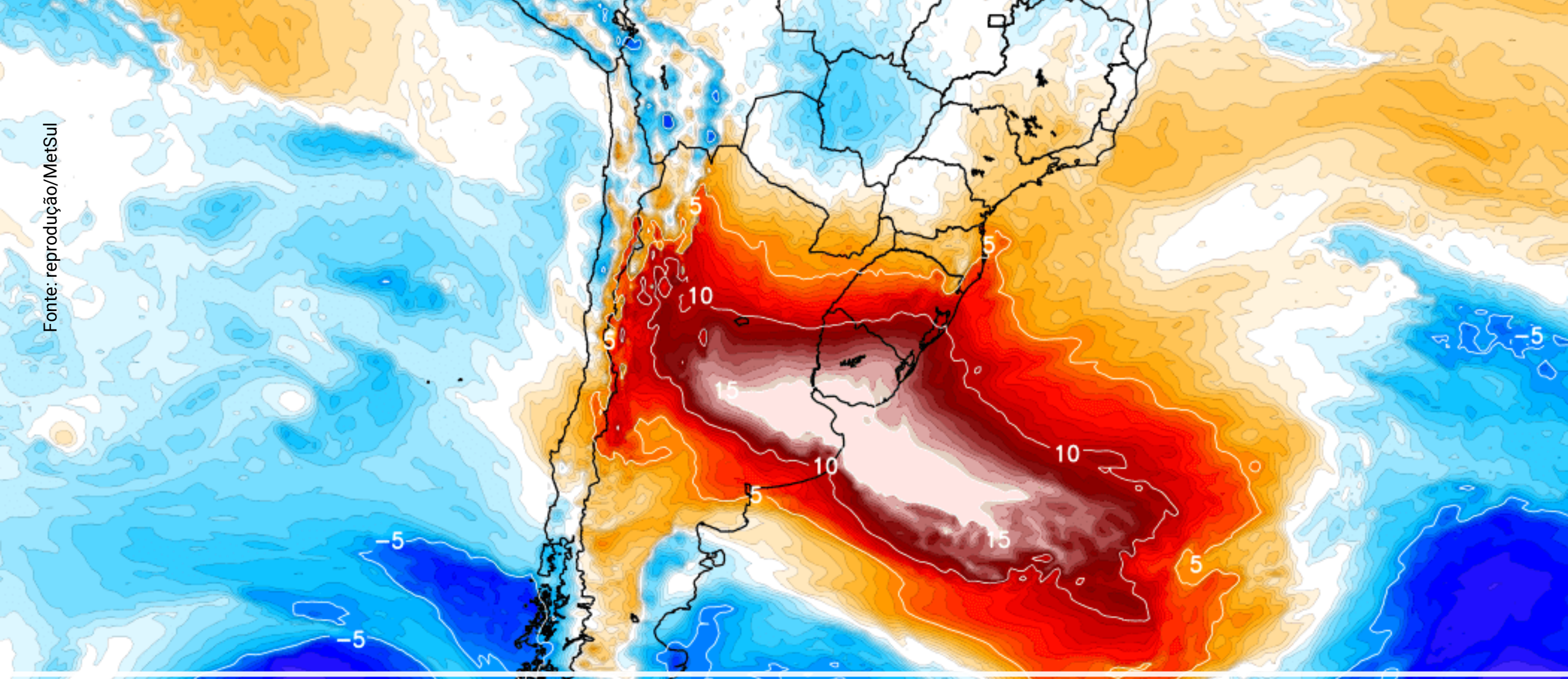


# Mapas de calor



Fonte: autor desconhecido





## Mapas de calor





# Gráficos de bolhas

# Gráficos de bolhas

---

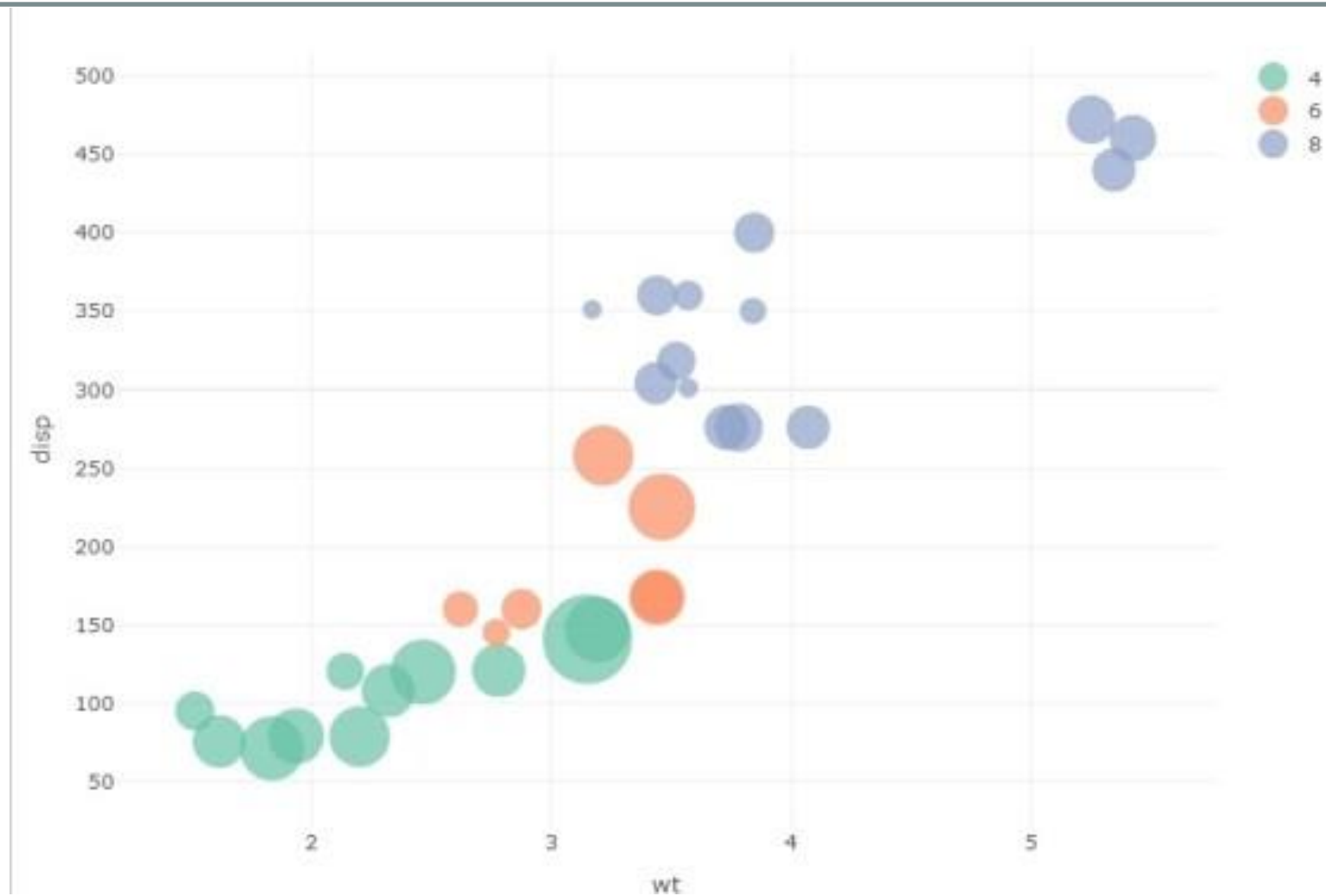
- Utilizam tamanhos de círculos e cores para representar valores e agregar informação útil a uma visualização única.
- São particularmente importantes quando não precisamos nos importar com a apresentação de um valor exato, e sim de uma escala de valor.



## Gráficos de bolhas

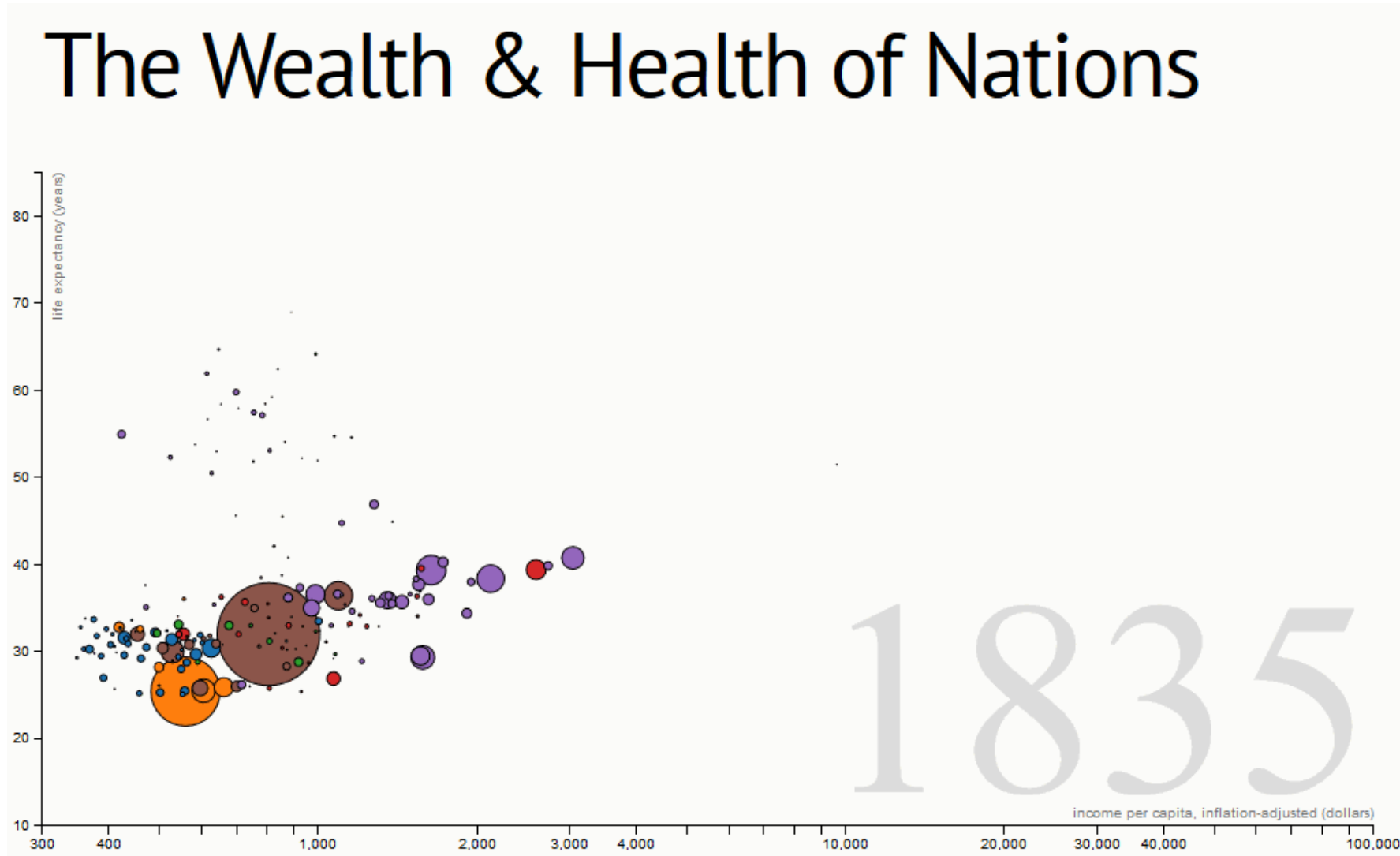
Fonte: autor desconhecido

# Gráficos de bolhas



Fonte: autor desconhecido

# Gráficos de bolhas animado





The image features a large orange semi-circle on the right side, which serves as a background for the text. To the left of this semi-circle is a solid blue circle. Further left, there is a green square outline. In the top left corner, there are two short vertical yellow dashed lines. In the top right corner, there is a yellow circle. A green line starts from the top center and extends diagonally upwards to the right. In the bottom left area, there are several yellow dashed lines of varying lengths and orientations, including one that is nearly vertical and others that are diagonal.

Visualizando dados  
no tempo



# Gráficos de linhas



# O gráfico de linhas

---

- É uma visualização bem conhecida, que facilita a apresentação de informações que variam no decorrer de um período.

# Gráfico de linhas

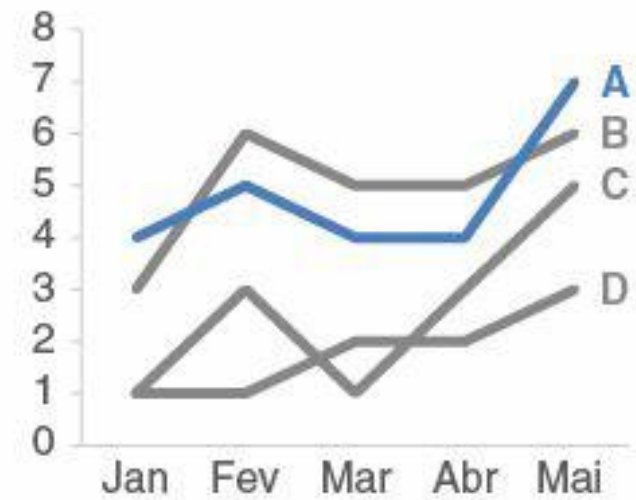
Uma série



Duas séries



Várias séries

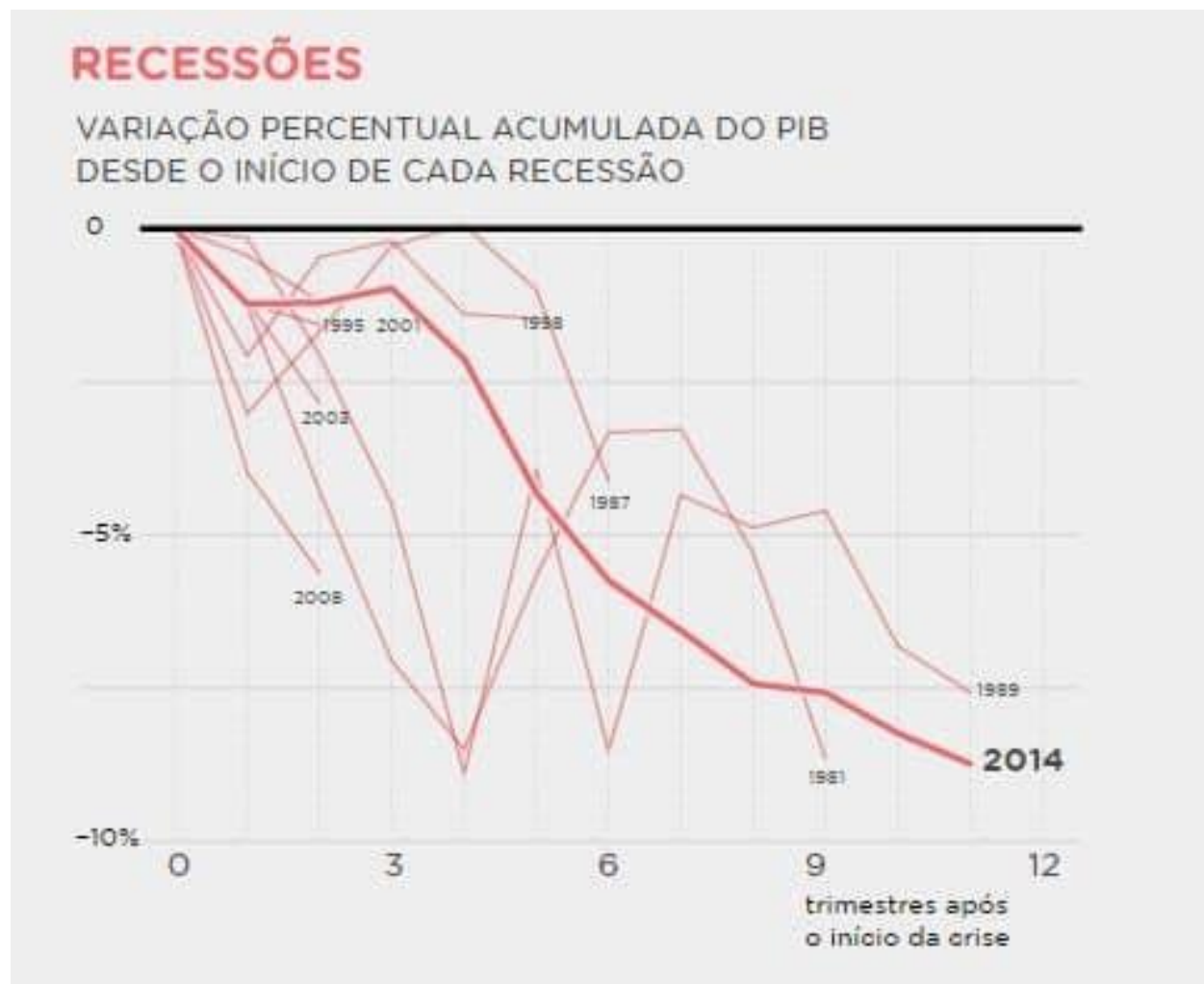


# O gráfico de linhas

---

- E se quisermos chamar atenção para uma série de dados em especial, dentre outras?
- Podemos fazer uso do processamento pré-atentivo.

# O gráfico de linhas

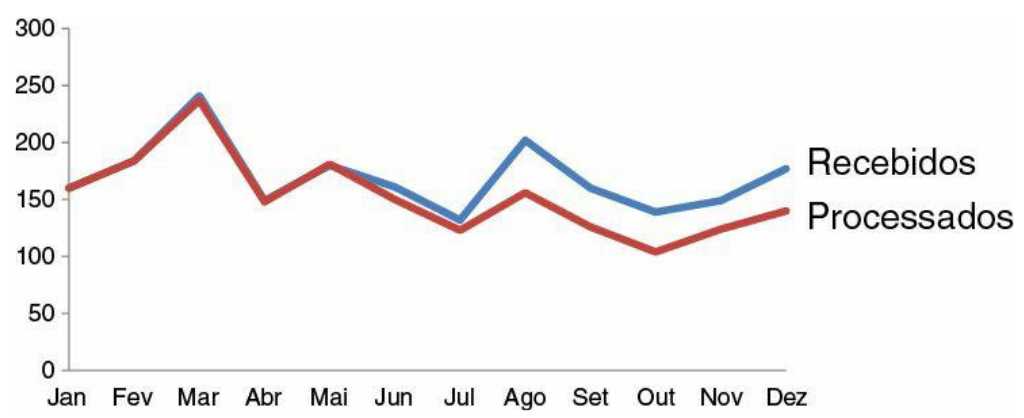


# O gráfico de linhas

---

- Podemos e devemos incluir informações textuais em nossa visualização para facilitar o entendimento do público-alvo.
- Porém, devemos ter cuidado para não poluir demais a visualização!

# O gráfico de linhas



Principais épocas de terminos de relacionamento  
de Acordo com as Atualizações de Status no Facebook



Fonte: Cole Nussbaumer Knaflic. Storytelling com Dados



# Gráficos de áreas

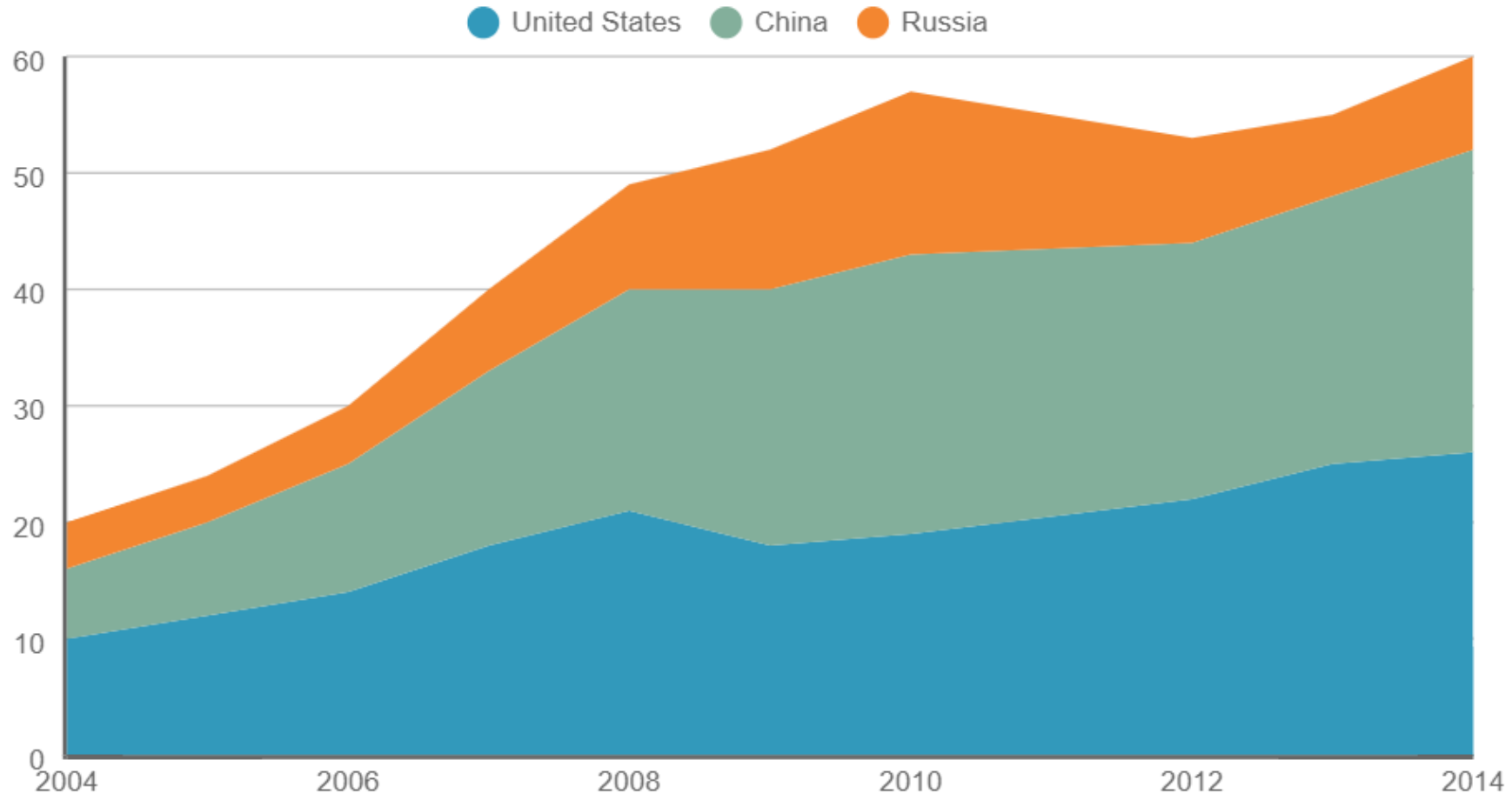
# O gráfico de áreas

---

- São que gráficos de linha onde preenchemos toda a área abaixo da linha. Este tipo de visualização é muito útil quando queremos chamar a atenção para comportamentos diferenciados entre os dados.



# O gráfico de áreas



The background features several abstract geometric elements: a large teal semi-circle on the right, a smaller grey circle in the upper left, a blue L-shaped line at the top, a blue square outline on the left, and several short, dark grey dashed lines scattered in the lower left area.

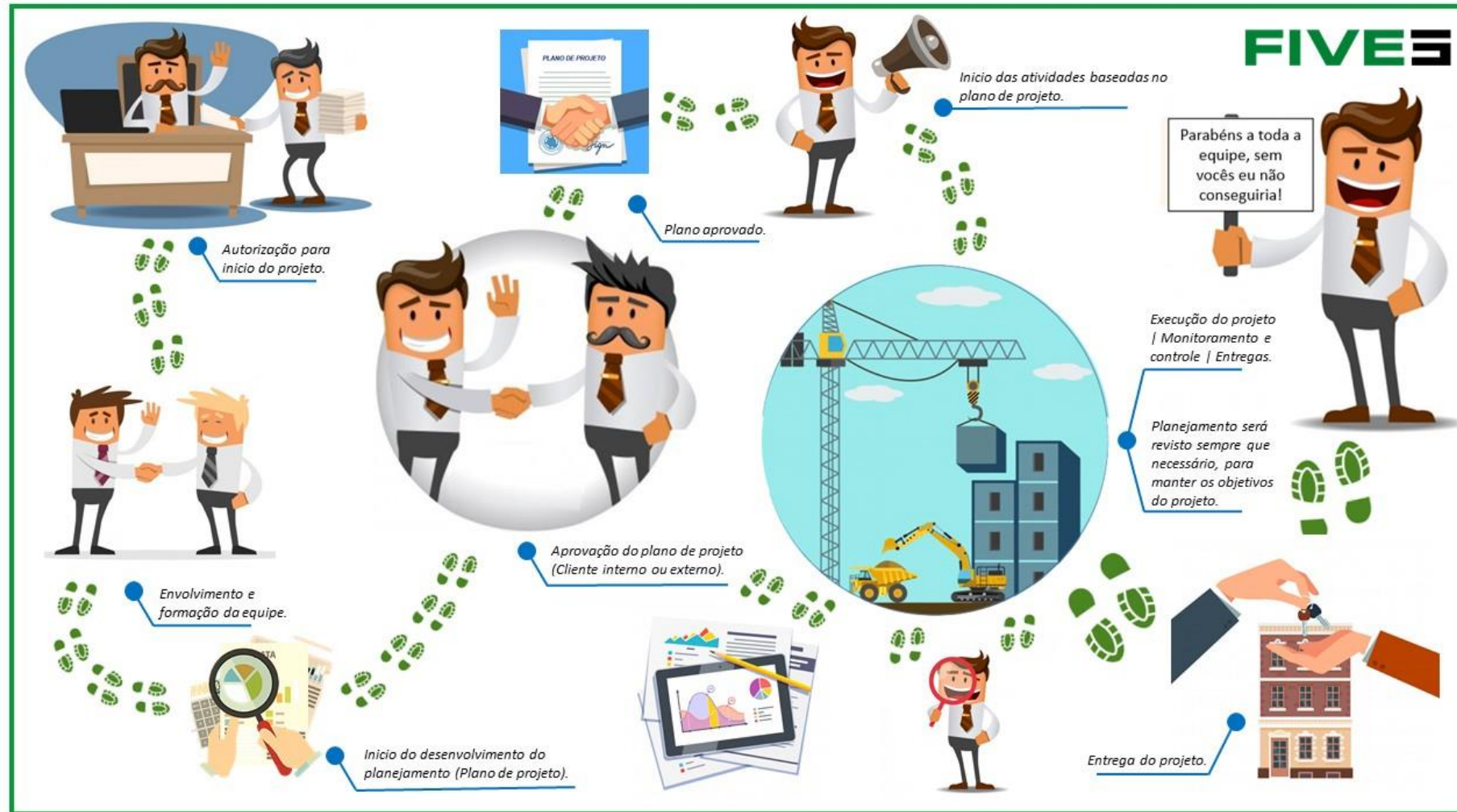
Fluxos e  
linhas do tempo

# Fluxos e linhas do tempo

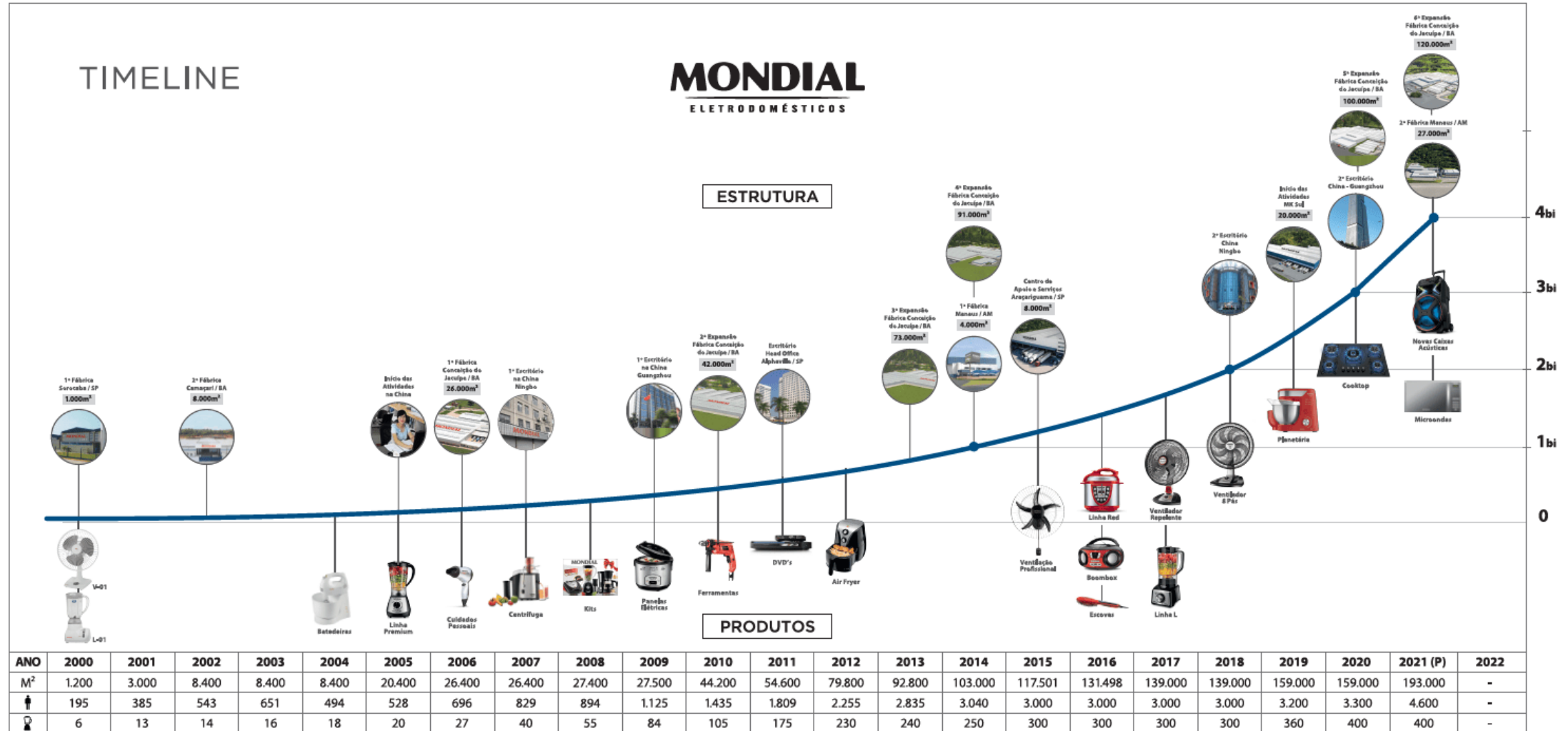
---

- São visualizações mais complexas, muitas vezes criadas por especialistas e com a utilização de ferramentas específicas.
- Porém são recursos poderosos para visualizarmos processos e fluxos no tempo mais complexos, além de ter um grande apelo visual.

# Fluxos e linhas do tempo



# Fluxos e linhas do tempo



Fonte: <https://www.emondial.com.br/linha-do-tempo>



Visualizando  
relações



# Diagrama de dispersão

# O diagrama de dispersão

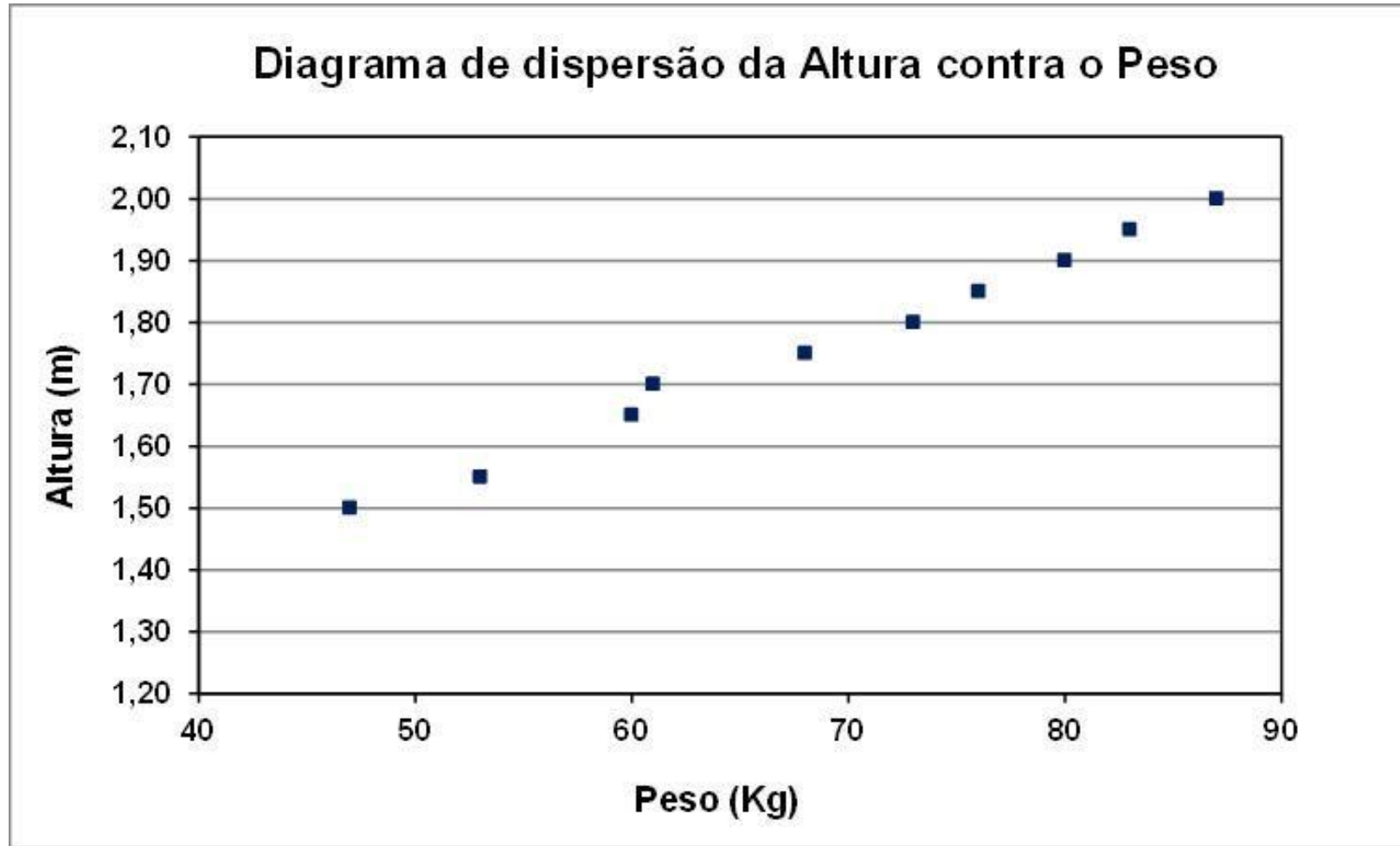
---

- É uma visualização que nos ajuda a avaliar como duas variáveis se relacionam entre si e perceber padrões de comportamento comuns
- Apesar de muitas vezes não serem percepções definitivas, nos ajudam a melhorar nosso entendimento sobre a dinâmica do comportamento das variáveis.
- Colocamos os valores de cada uma das duas variáveis em estudo em um dos eixos (X e Y) e marcamos o ponto de cruzamento entre estes valores.



# O diagrama de dispersão

---



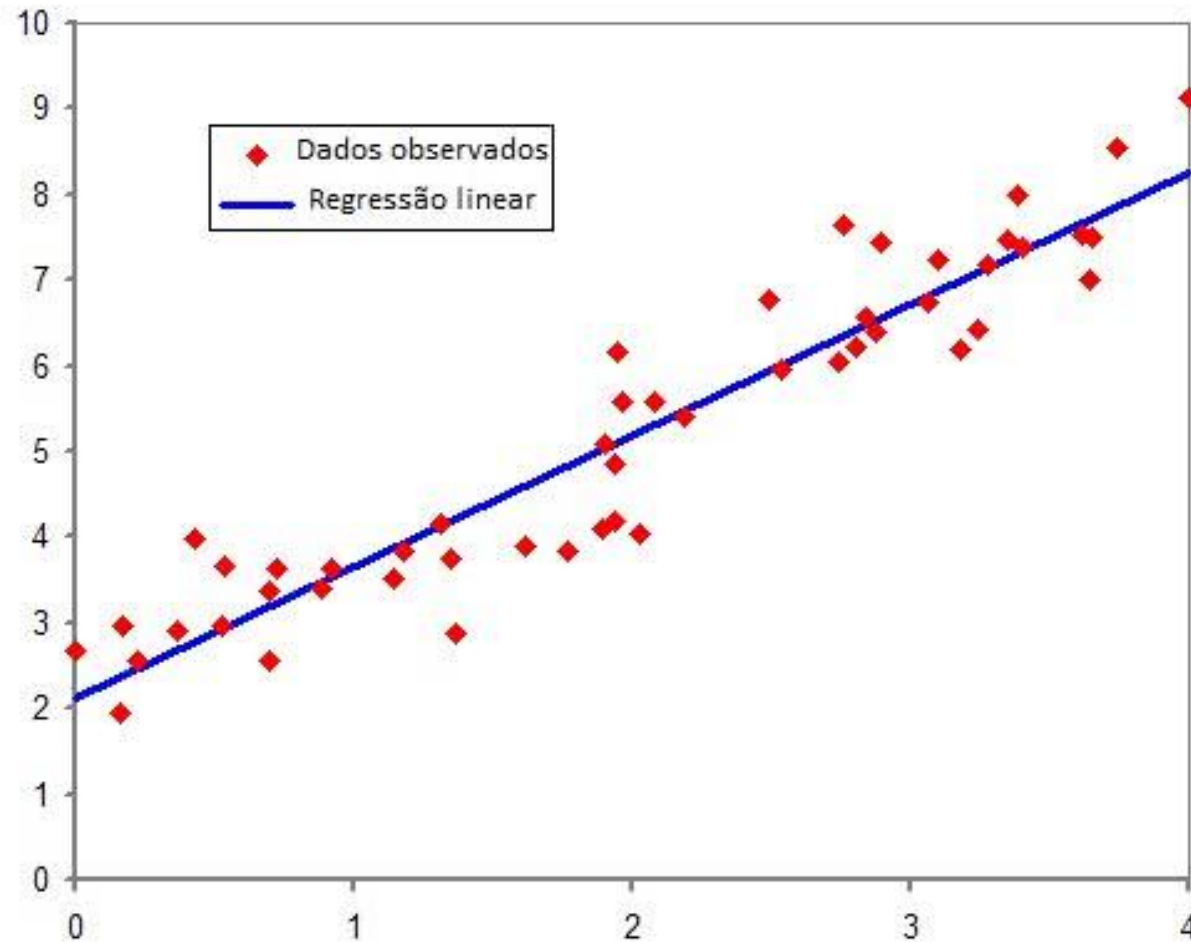
# Reta de tendência ou regressão

---

- Podemos incluir uma reta que nos ajudará a ter uma ideia melhor sobre o “comportamento médio” das variáveis.

# Reta de tendência ou regressão

---



# Correlação a partir do diagrama de dispersão

---

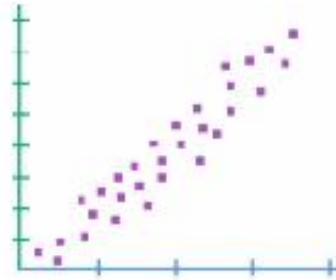
- A correlação é uma medida estatística importante para avaliarmos o comportamento conjunto de duas variáveis numéricas.

# Correlação a partir do diagrama de dispersão

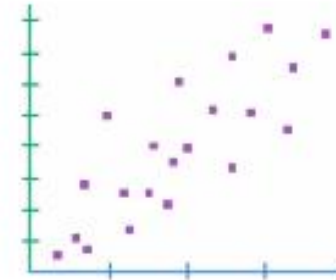
---



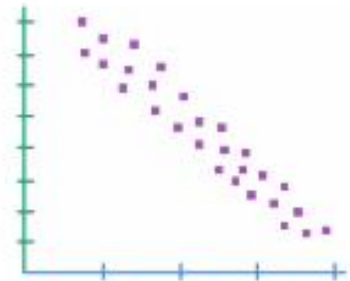
**Sem correlação**



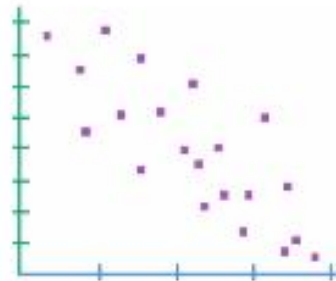
**Correlação  
positiva forte**



**Correlação  
positiva média**



**Correlação  
negativa forte**



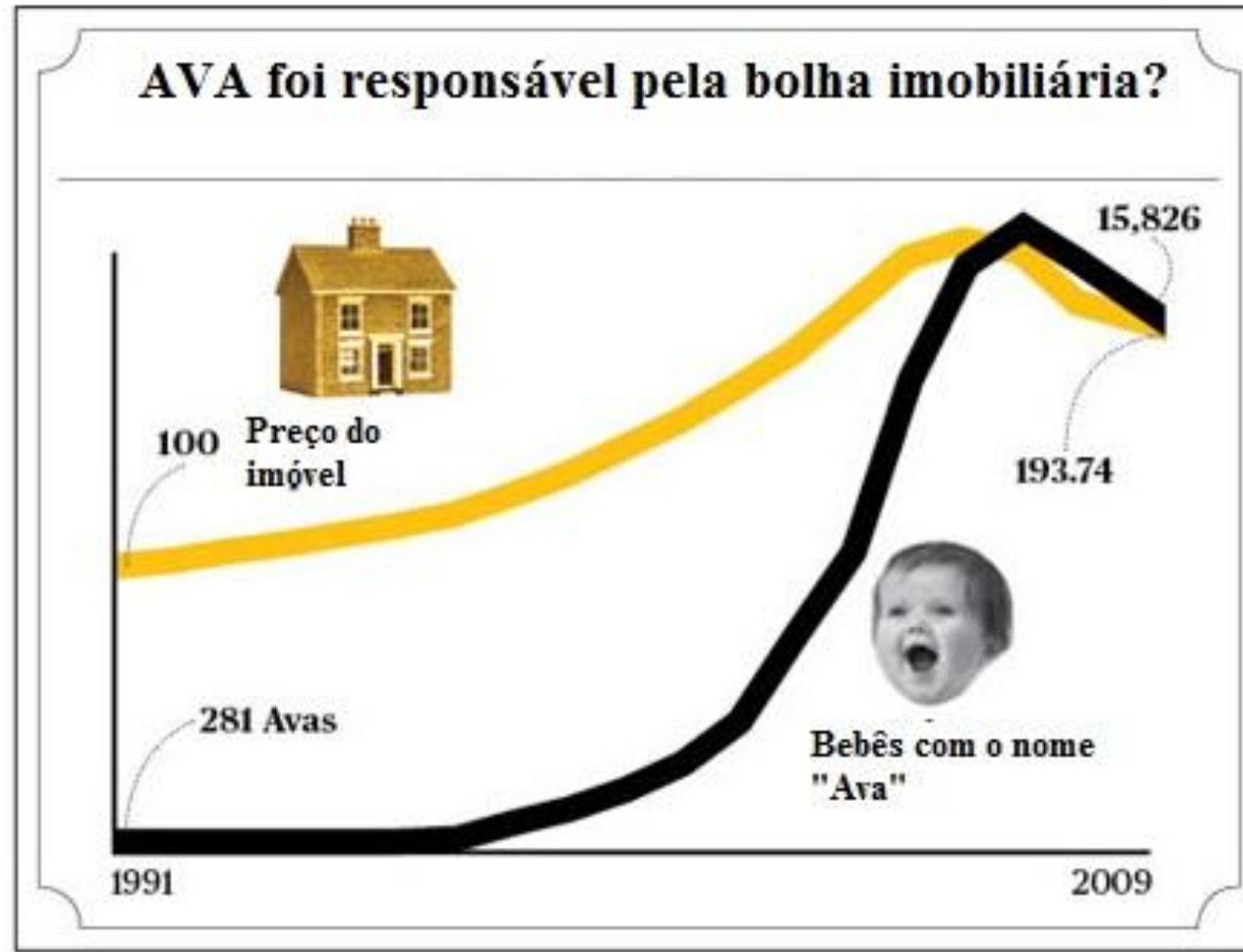
**Correlação  
negativa média**

# Atenção! Correlação não indica causalidade!

---

- Apesar da correlação ser uma medida estatística muito relevante, nunca devemos partir do princípio que uma correlação alta seja indicativo de relação de causa-efeito entre as variáveis em estudo.
- Vale uma visita ao site [Spurious Correlations \(em inglês\)](#).

# Atenção! Correlação não indica causalidade!





# Gráficos radar



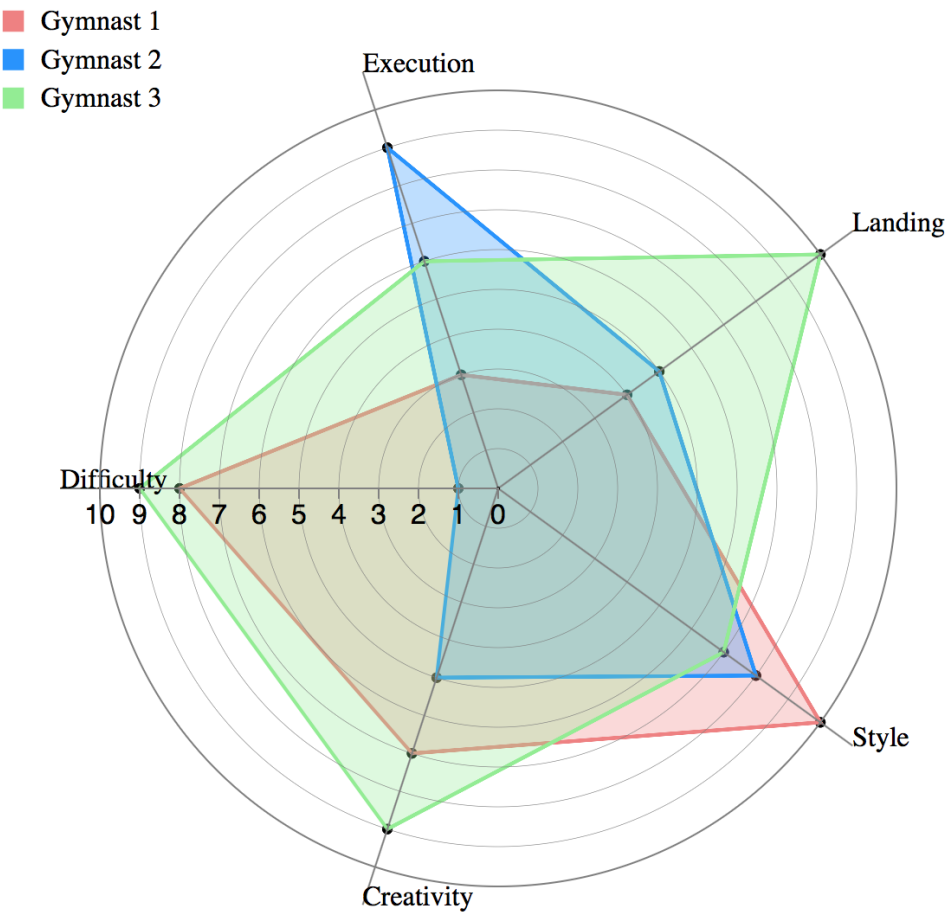
# O gráfico radar

---

- Utilizado para apresentar características dos dados que favorecem a comparação dos dados de forma simples e compacta.
- Muito usado em jogos eletrônicos!

# O gráfico radar

Gymnast Scoring Radar Chart



# O gráfico radar

## TOP 3 JOGADORES

### L. Messi



Nacionalidade: Argentina  
Idade: 31  
Posição: RF  
Valor: €110.5MM  
Clube: FC Barcelona  
Score: 94



### Cristiano Ronaldo



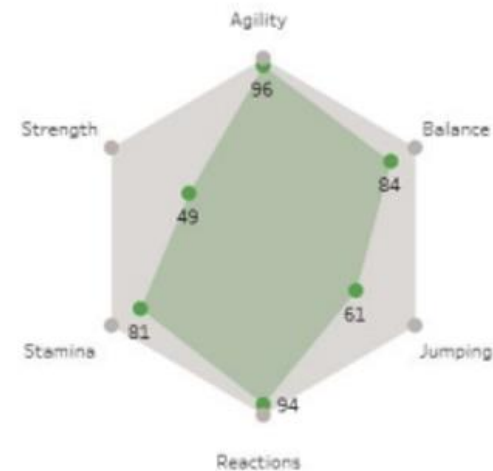
Nacionalidade: Portugal  
Idade: 33 anos  
Posição: ST  
Valor: €77MM  
Clube: Juventus  
Score: 94



### Neymar Jr



Nacionalidade: Brasil  
Idade: 26  
Posição: LW  
Valor: €118.5MM  
Clube: PSG  
Score: 92



#### Habilidade

- ☐ (Tudo)
- ☐ Acceleration
- ☐ Aggression
- ☒ Agility
- ☒ Balance
- ☐ BallControl
- ☐ Composure
- ☐ Crossing
- ☐ Curve
- ☐ Dribbling
- ☐ Finishing
- ☐ FkAccuracy
- ☐ HeadingAccuracy
- ☐ Interceptions
- ☒ Jumping
- ☐ LongPassing
- ☐ LongShots
- ☐ Marking
- ☐ Penalties
- ☐ Positioning
- ☒ Reactions
- ☐ ShortPassing
- ☐ ShotPower
- ☐ SlidingTackle
- ☐ SprintSpeed
- ☒ Stamina
- ☐ StandingTackle
- ☒ Strength
- ☐ Vision
- ☐ Volleys



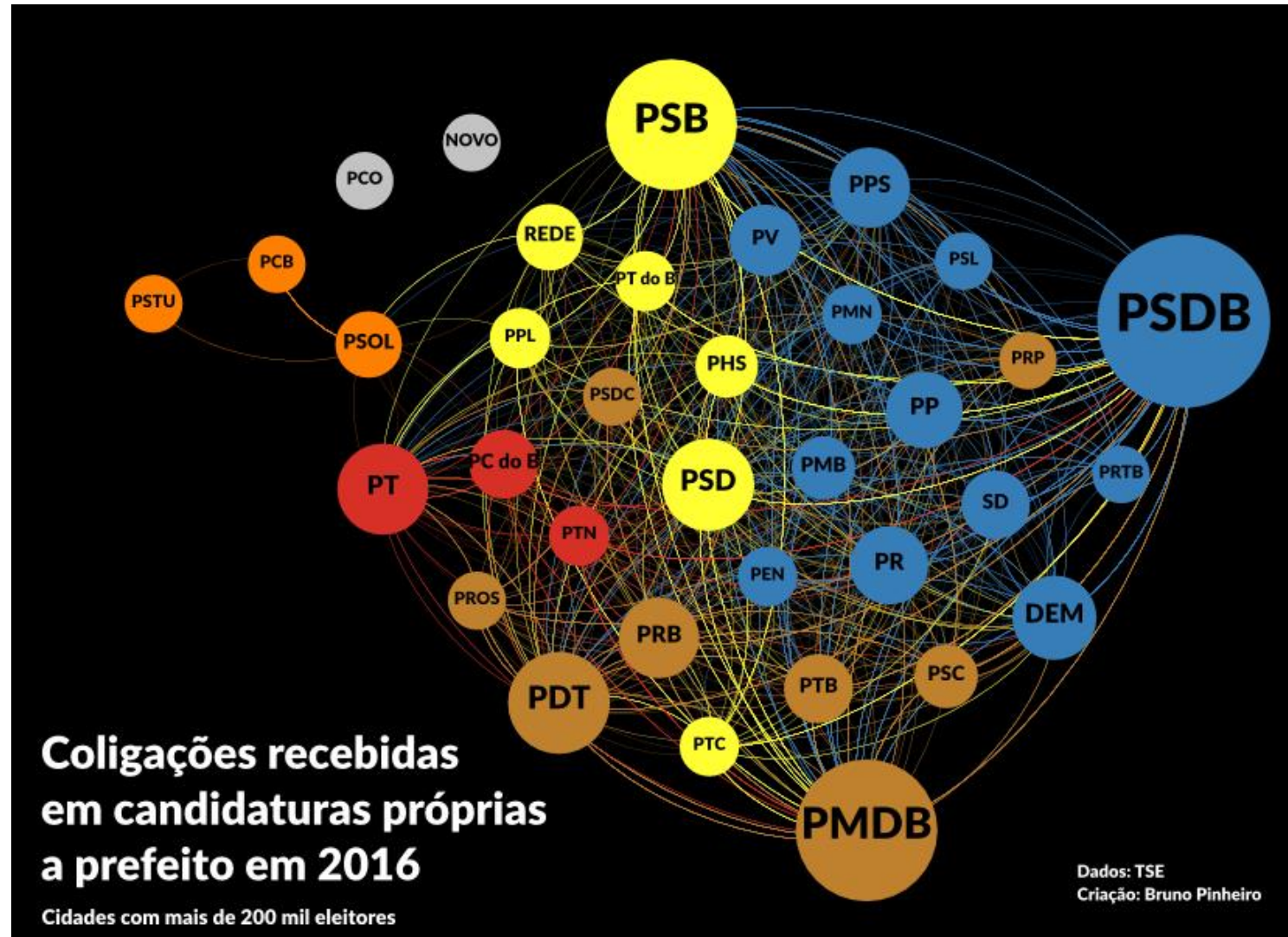
# Diagramas de rede

# O diagrama de rede

---

- Visualizações baseadas em grafos que nos indicam relações entre elementos. Formada por nós e arestas.

# O diagrama de rede





Visualizando Partes  
do todo



Gráfico de setores



# O gráfico de setores

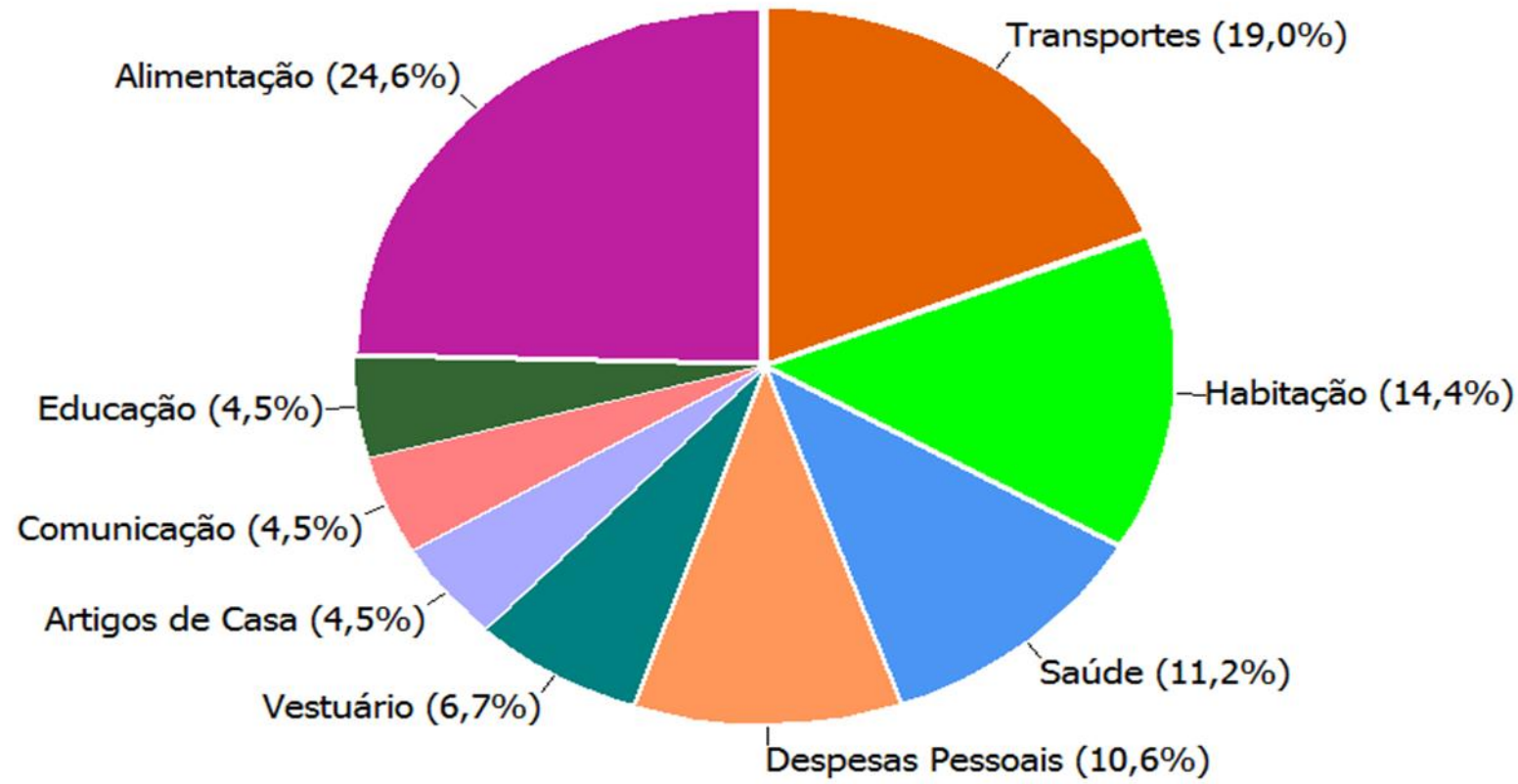
---

- Popularmente conhecidos por gráficos de pizza são, provavelmente, as visualizações mais conhecidas e mais utilizadas, principalmente na área de negócios de forma geral.
- Apesar de muito usada, esta visualização apresenta algumas deficiências e casos de uso um pouco mais específicos.
- De forma simples, basta apresentar as partes como “fatias” de uma pizza (ou, se preferir, uma torta) onde a soma das proporções das partes deverá ser sempre igual a 100%.

# O gráfico de setores

---

**Peso dos itens no IPCA (em jan/14)**



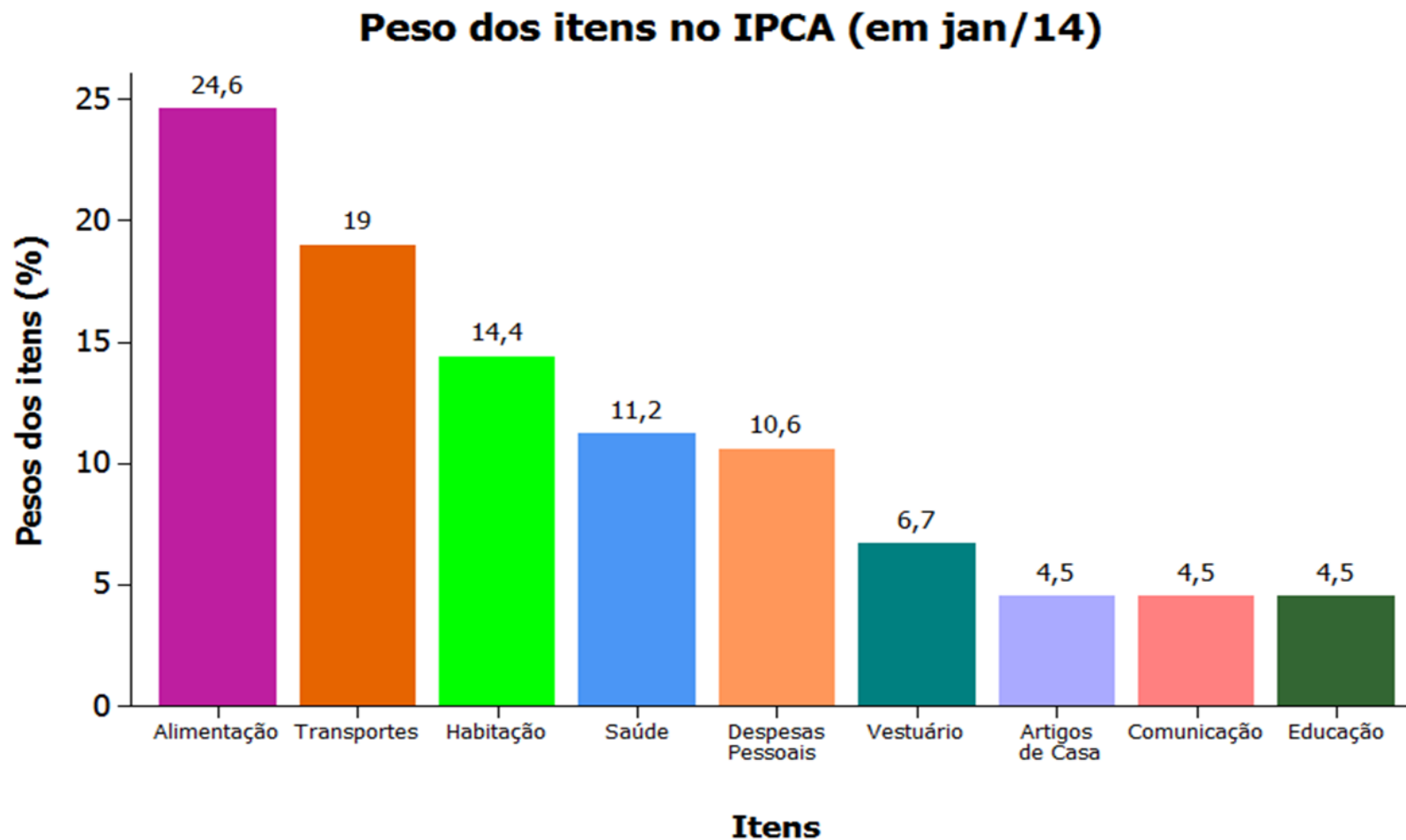
# O gráfico de setores

---

- Muitos autores defendem que um gráfico de setores com muitas partes, como vimos no slide anterior, pode e deve ser substituído por um gráfico de barras.

# O gráfico de setores

---



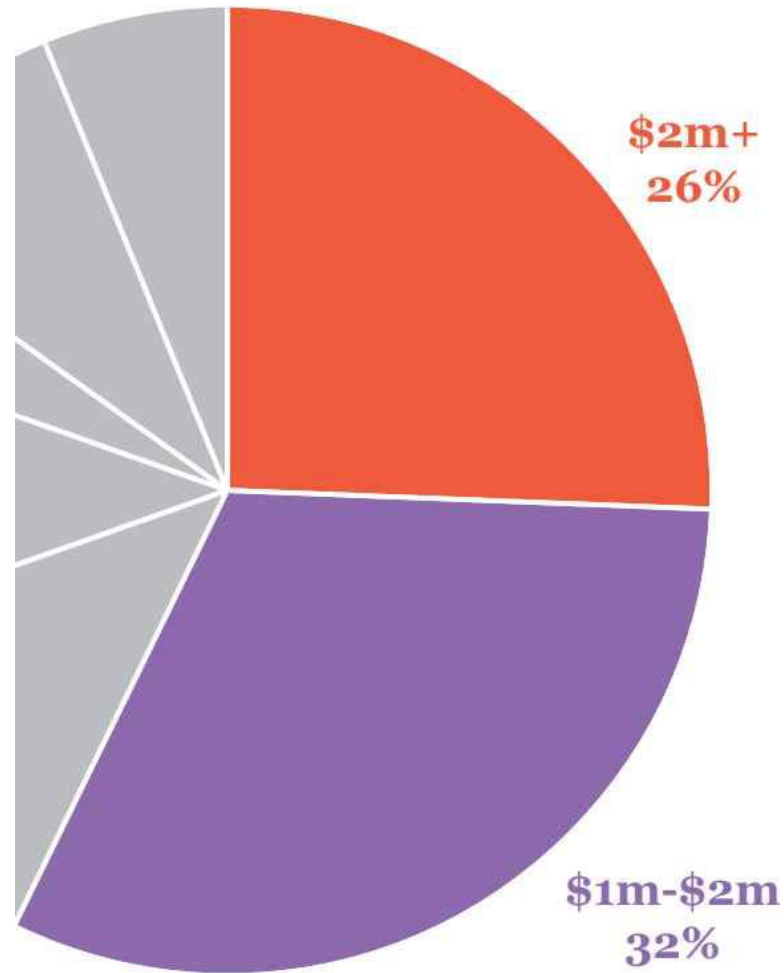
# O gráfico de setores

---

- Outra abordagem interessante é reduzir o número de partes e destacar as partes para as quais queremos atenção, seja com algum recurso visual ou agregando partes menores.
- Um gráfico de setores ideal não deveria ter mais de três partes!

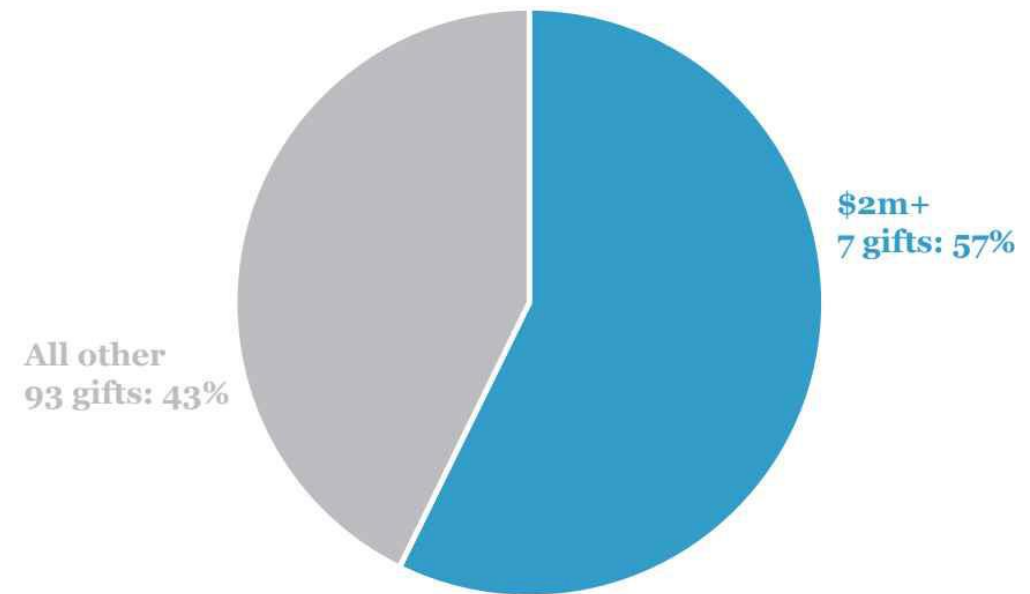
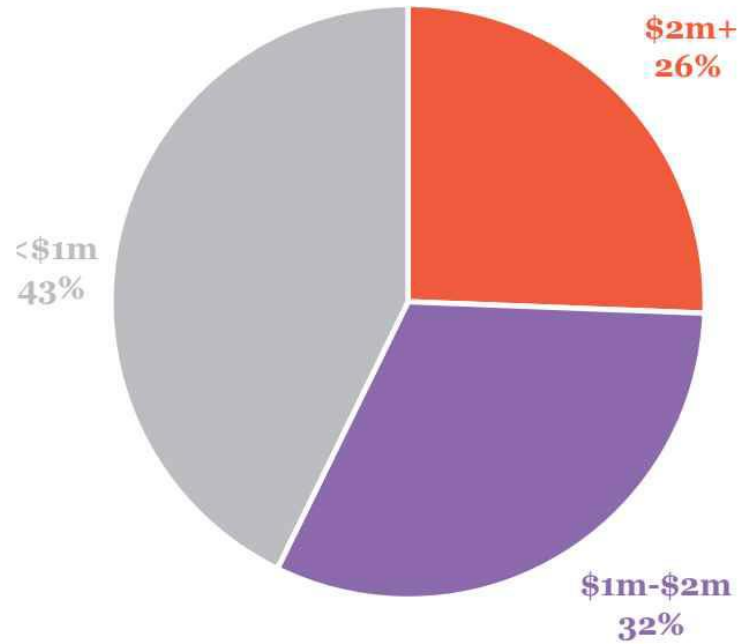
# O gráfico de setores

---



# O gráfico de setores

---





Treemap



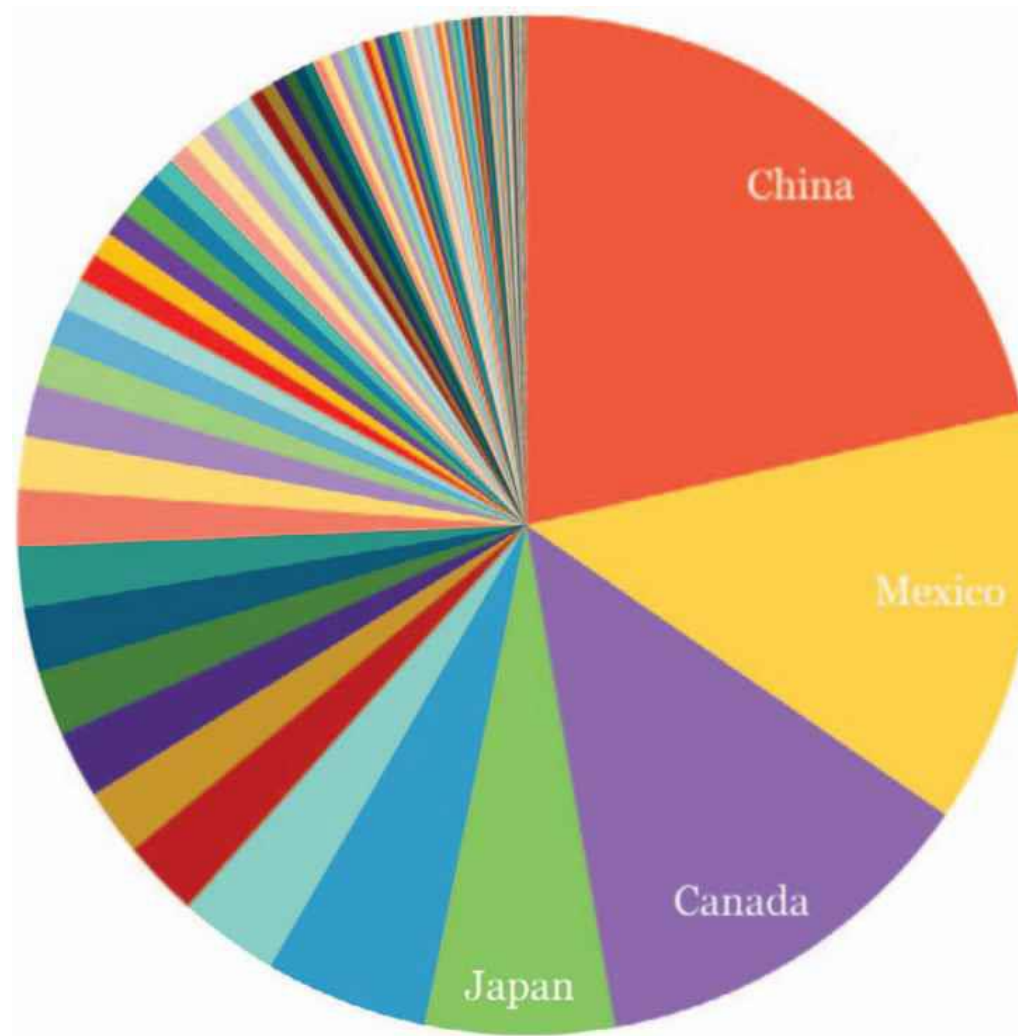
# Treemap

---

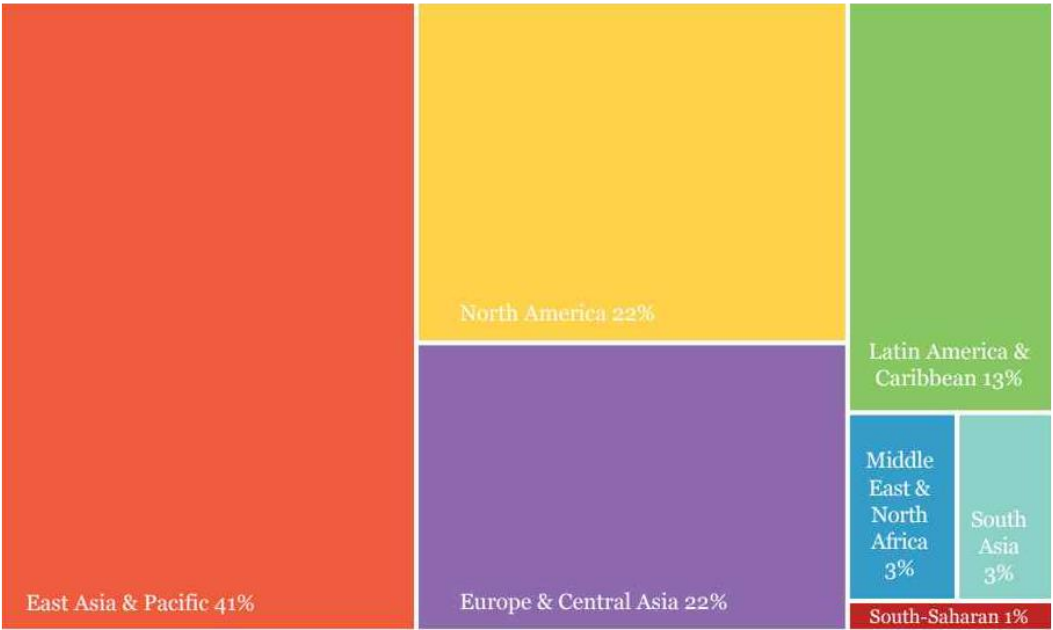
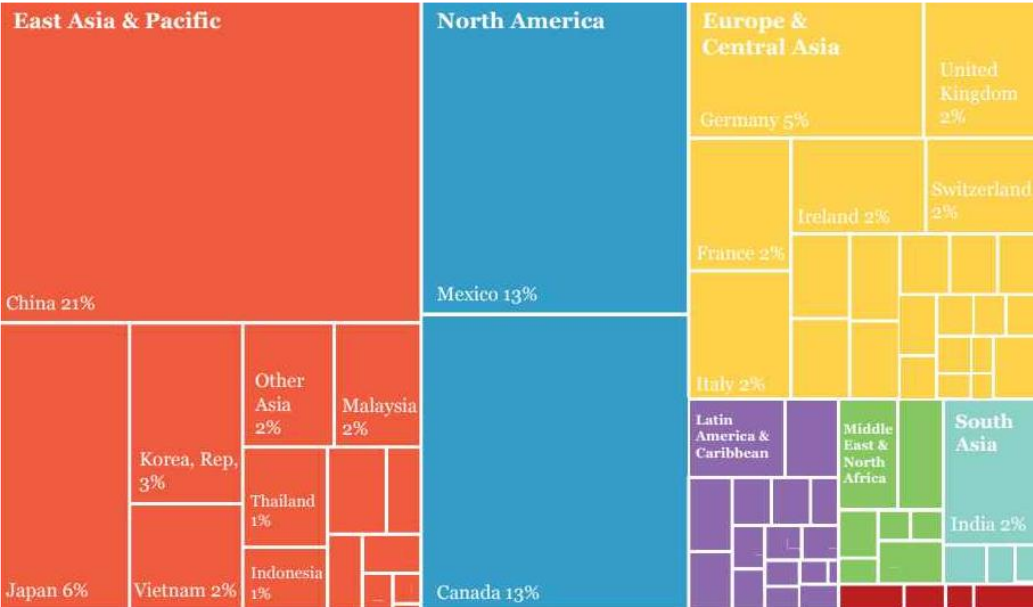
- É uma boa opção ao gráfico de setores. Divide um quadrado ou retângulo em setores, ilustrando algum grau de hierarquia ou relação de parte com o todo.
- As hierarquias podem ser identificadas, via rótulos, diretamente na visualização.
- Vamos ver como fica, comparando com um gráfico de setores.

# O gráfico de setores vs. Treemap

---



# O gráfico de setores vs. Treemap



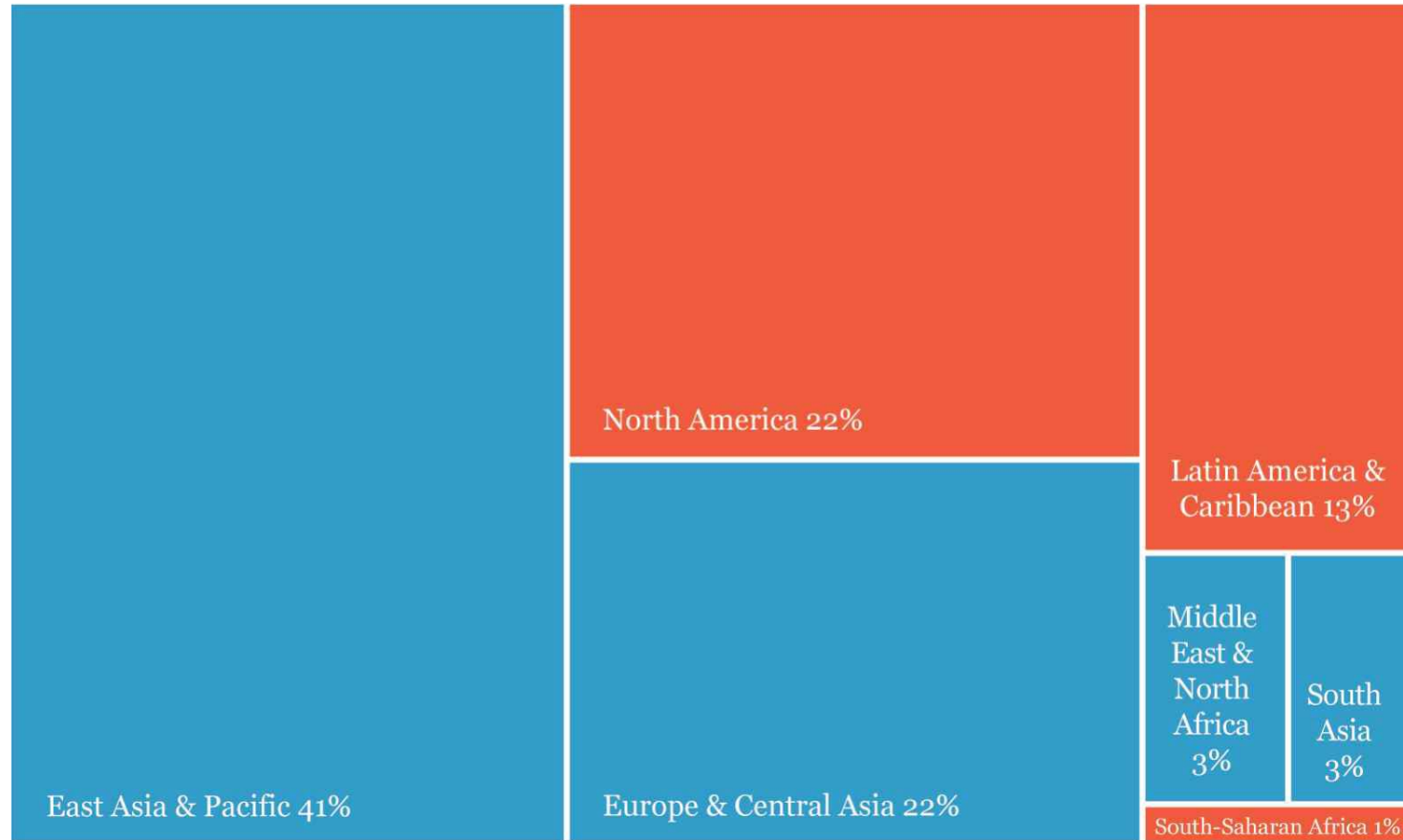
# Treemap

---

- Podemos ainda usar cores para indicar comportamentos que consideramos relevantes e queremos destacar.
- Na próxima visualização, azul indica o crescimento em relação ao período anterior e o vermelho queda.

# Treemap

---





Sunburst

# Sunburst

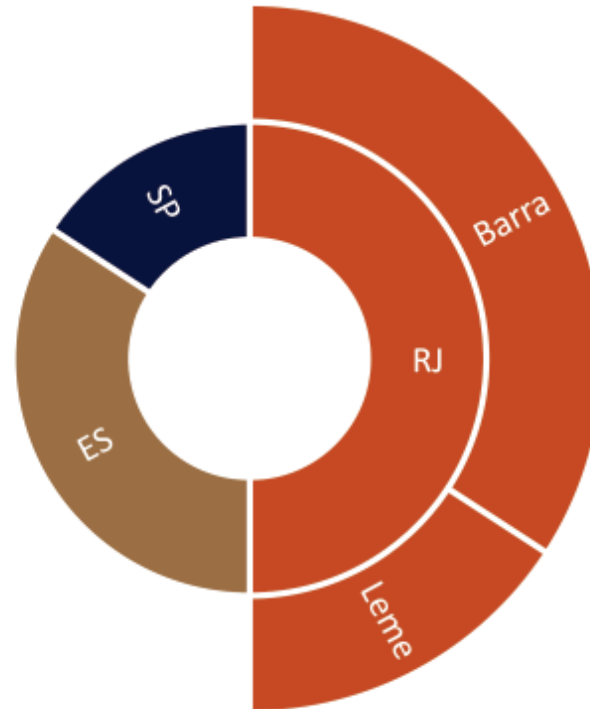
---

- Visualização que lembra um gráfico de setores, porém com características de um Treemap e esteticamente mais bonita.
- Apresenta hierarquia de informações de dentro para fora, com o nível mais alto sendo representado no anel interior da visualização e aumentando a granularidade da informação de dentro para fora.

# Sunburst

---

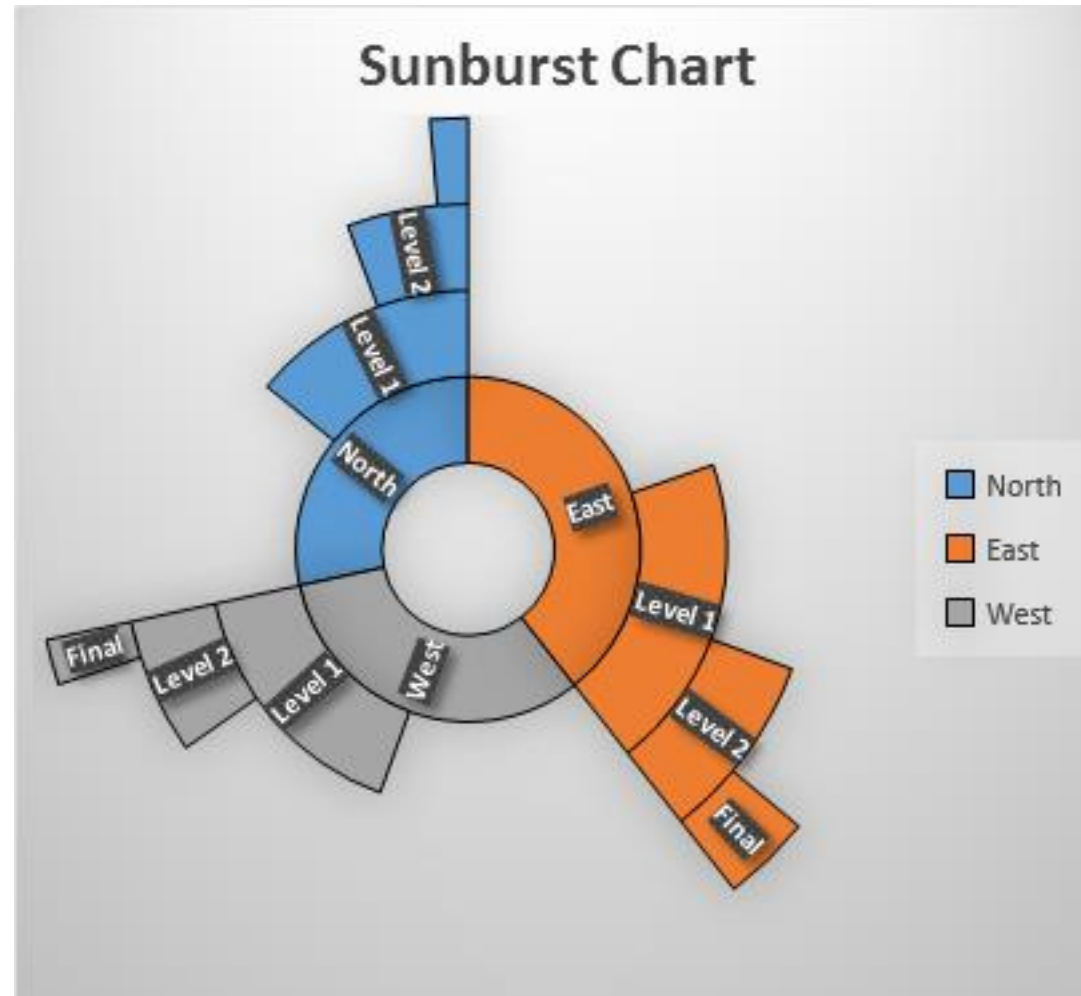
**Vendas 2021 RJ (Barra e Leme), ES e SP**





# Sunburst

---



The background features several abstract elements: a large orange circle on the right, a blue circle in the upper left, a yellow circle in the top right, a green L-shaped line in the top center, a green square outline on the left, and several yellow dashed lines scattered on the left side.

Outras visualizações  
úteis



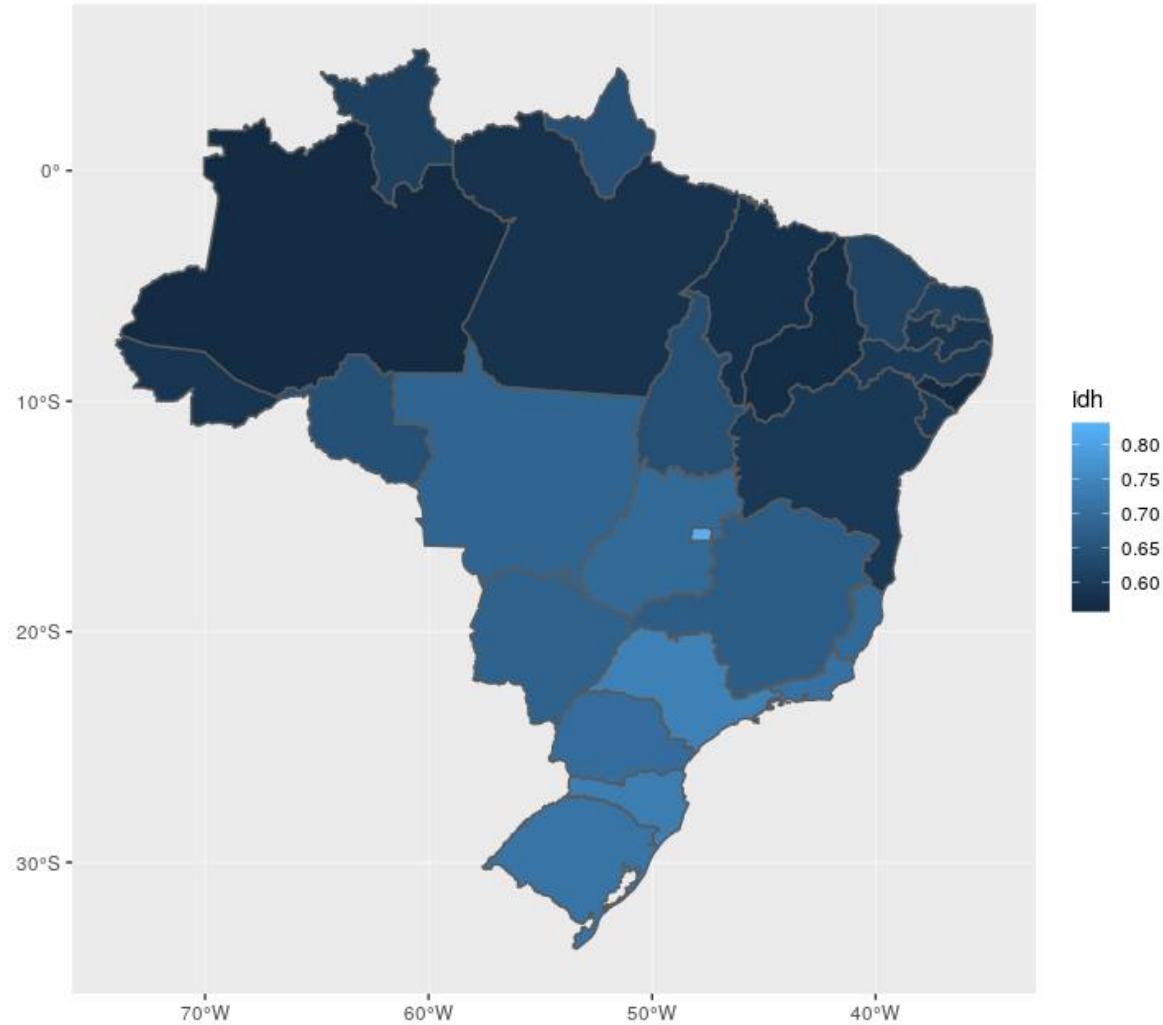
Utilizando mapas  
para visualizações

# Uso de mapas em visualizações

---

- A utilização de mapas como base de uma visualização é mais um recurso muito importante e necessário na “caixa de ferramentas” de qualquer cientista de dados.
- Os mapas trazem para as visualizações uma imagem conhecida para as pessoas e podem agregar um senso de pertencimento e inclusão ao público-alvo.

# Uso de mapas em visualizações



Fonte: <https://blog.curso-r.com/posts/2019-02-10-sf-miojo/>

# Uso de mapas em visualizações

---





Nuvem de palavras

# Nuvem de palavras

---

- Para apresentar a frequência de ocorrência e/ou relevância de palavras específicas em um conjunto de palavras podemos utilizar a visualização chamada Nuvem de palavras.
- É um recurso de visualização cada vez mais utilizado pela mídia pois é de simples entendimento e bastante informativo, quando utilizada corretamente.
- Podemos agregar informações a uma nuvem de palavras por meio de cores.



# Nuvem de palavras



Fonte: Mayara Lopes de Freitas Lima

# Nuvem de palavras



Fonte: Francisco Giovanni David Vieira



Tabelas

# Tabelas

---

- Um recurso que deve ser utilizado com cuidado, mas que pode ser sempre considerado uma opção válida, é a utilização de tabelas para apresentação de informações.
- Neste caso, o uso do processamento pré-atentivo é uma ferramenta muito importante na indicação dos dados relevantes para o público-alvo.

# Tabelas

	2010	2011	2012	2013	2014	2015	2016
China	10.10	9.01	7.33	7.23	6.76	6.36	6.12
India	8.76	5.25	4.13	5.10	6.14	6.90	5.89
United States	1.68	0.85	1.46	0.96	1.80	2.09	0.74
Indonesia	4.83	4.79	4.68	4.24	3.73	3.65	3.85
Brazil	6.50	3.00	0.98	2.07	-0.38	-4.37	-4.25
Pakistan	-0.48	0.61	1.34	2.21	2.51	2.61	3.44
Nigeria	5.00	2.12	1.52	2.61	3.52	-0.02	-4.16
Bangladesh	4.40	5.25	5.28	4.77	4.84	5.37	5.96
Russia	4.46	5.20	3.48	1.57	-1.04	-3.04	-0.41
Mexico	3.49	2.12	2.15	-0.06	1.45	1.90	1.58

	2010	2011	2012	2013	2014	2015	2016
China	10.10	9.01	7.33	7.23	6.76	6.36	6.12
India	8.76	5.25	4.13	5.10	6.14	6.90	5.89
United States	1.68	0.85	1.46	0.96	1.80	2.09	0.74
Indonesia	4.83	4.79	4.68	4.24	3.73	3.65	3.85
Brazil	6.50	3.00	0.98	2.07	-0.38	-4.37	-4.25
Pakistan	-0.48	0.61	1.34	2.21	2.51	2.61	3.44
Nigeria	5.00	2.12	1.52	2.61	3.52	-0.02	-4.16
Bangladesh	4.40	5.25	5.28	4.77	4.84	5.37	5.96
Russia	4.46	5.20	3.48	1.57	-1.04	-3.04	-0.41
Mexico	3.49	2.12	2.15	-0.06	1.45	1.90	1.58

Fonte: Schwabish , Jonathan; Better Data Visualizations A Guide for Scholars, Researchers, and Wonks