

FUNDAMENTOS DE VISUALIZAÇÃO DE DADOS

Leila Rodrigues



ANÁLISE EXPLORATÓRIA X ANÁLISE EXPLANATÓRIA

- **Análise Exploratória:**
 - Identificar os comportamentos médios e discrepantes; comparar esses comportamentos; investigar a interdependência entre as variáveis; procurar e identificar tendências.
 - As técnicas da análise **exploratória** de dados nos ajudam a extrair informações relevantes de um conjunto de dados.

ANÁLISE EXPLORATÓRIA X ANÁLISE EXPLANATÓRIA

- **Análise **Explanatória**:**
 - Apresentar ao público alvo as informações obtidas a partir da análise exploratória dos dados.
 - Apesar de tentador mostrar toda a análise feita, todos os gráficos, tabelas, é interessante filtrar e transmitir apenas a informação necessária para o público.
 - Foco deste material.

QUEM

Identifique o público-alvo.


O QUÊ

Deve sempre querer que o público **saiba** ou **faça algo**.

COMO


Quais dados serão usados e como serão apresentados.

QUEM, O QUÊ E COMO: EXEMPLO



Professor que encerrou um curso experimental de férias.

Você avaliou os alunos no início e no fim do curso



A partir da análise de dados é possível provar o sucesso do curso.

Quais os potenciais públicos interessados nessa informação?

QUEM, O QUÊ E COMO: EXEMPLO

QUEM

Comitê de orçamento que aprova os recursos financeiros para continuação do curso.

O QUÊ

Conforme análise, solicitamos a aprovação do orçamento de R\$X para continuidade do programa.

COMO

Ilustrar o sucesso com dados coletados por meio da avaliação realizada antes e depois do curso.

- Imagine uma tela em branco, cada elemento adicionado a essa tela absorve carga cognitiva por parte do público.
- É importante minimizar a carga cognitiva percebida e ainda assim permitir a transmissão claramente da informação ao público.
- Ruído ou saturação: elementos visuais que ocupam espaço, mas não aumentam o entendimento.

- Os **Princípios da Gestalt de Percepção Visual** ajuda a identificar o que é sinal (informação útil) e que é ruído.
- Princípios serão discutidos:
 - proximidade, similaridade, acercamento, fechamento e conexão.

- Tendemos a pensar que objetos fisicamente próximos fazem parte de um grupo.



Princípios da proximidade da Gestalt



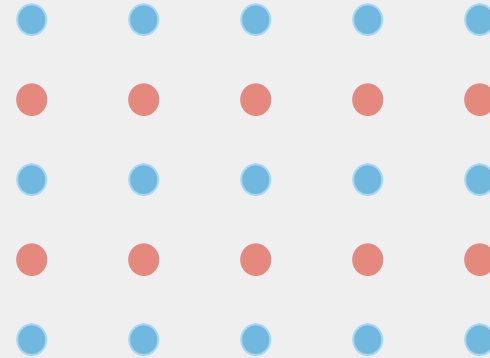
Pontos visualizados como colunas e linhas

SIMILARIDADE

- Objetos que tem cor, forma, tamanho ou orientações semelhantes são percebidos como relacionados ou pertencentes a parte de um grupo.

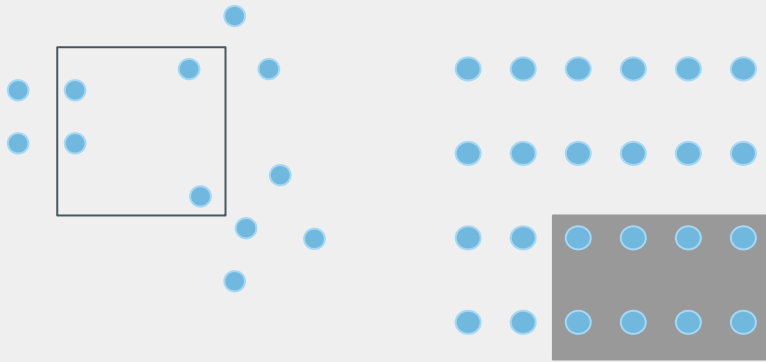


Princípios da similaridade da Gestalt

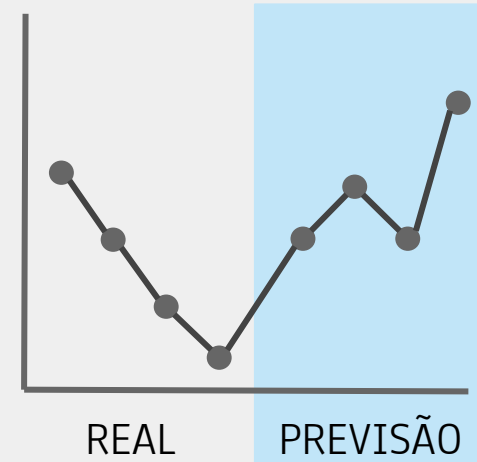


Linhas são visualizadas devido à similaridade da cor

- Tendemos a pensar que objetos fisicamente delimitados fazem parte de um grupo.



Princípios da acercamento da Gestalt



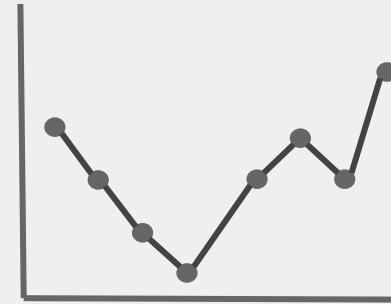
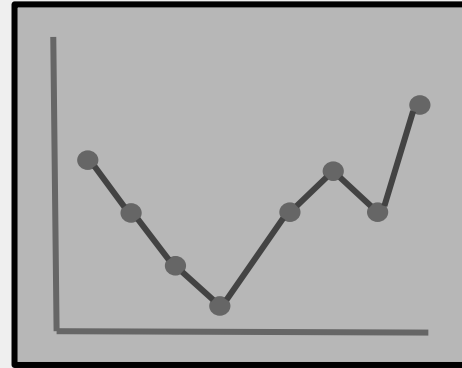
Área sombreada separa a previsão dos dados reais

FECHAMENTO

- Tendemos a perceber um conjunto de elementos individuais como uma única forma reconhecível.
- Nossos olhos tendem a preencher as lacunas.

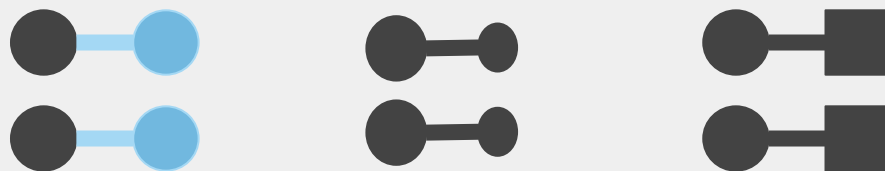


Princípios da fechamento da Gestalt



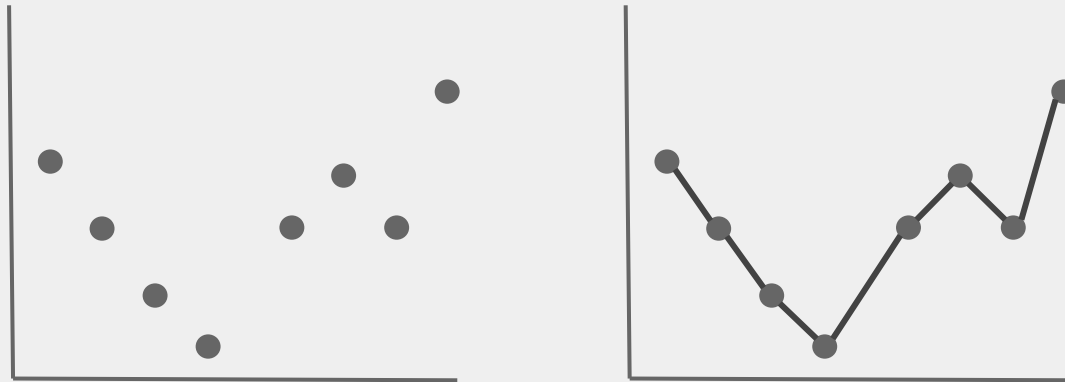
Antes e depois da aplicação do fechamento

- Tendemos a considerar objetos fisicamente conectados como parte de um grupo.
- A propriedade conectiva normalmente tem um valor associativo mais forte do que cor, tamanho ou forma similar.



Princípios da conexão da Gestalt

- Aplicação frequente em gráficos de linhas.

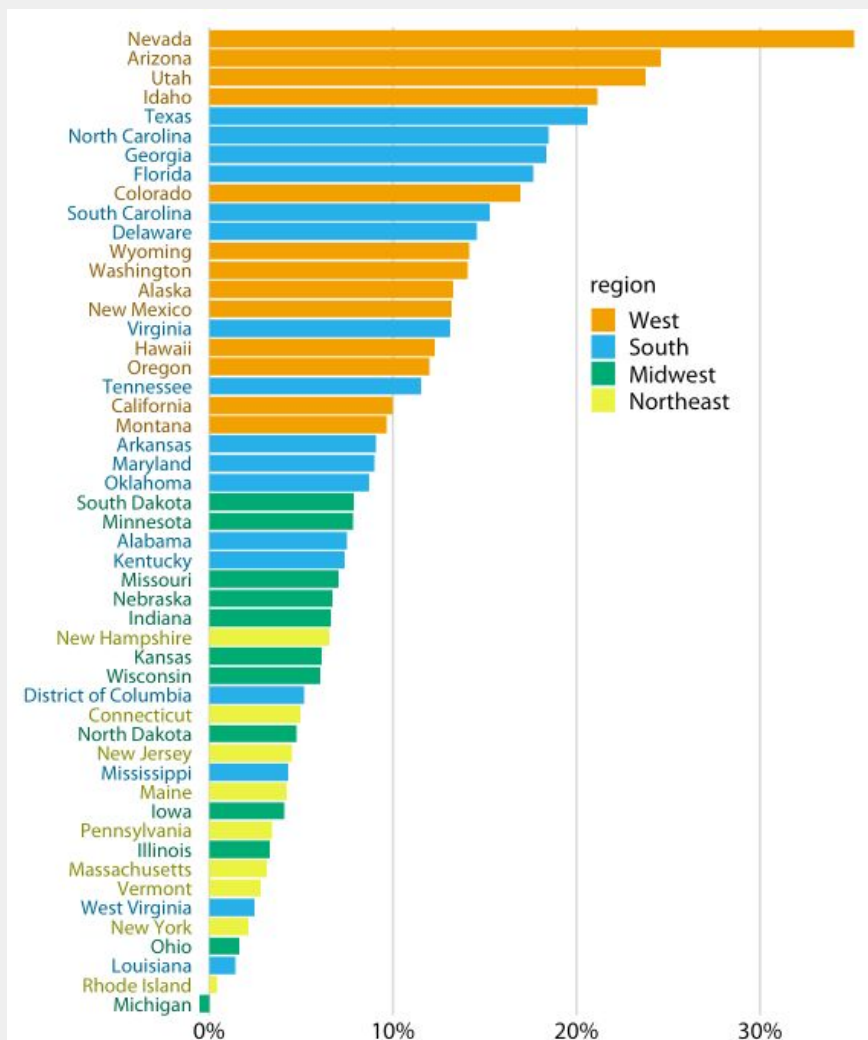


Linhas conectam pontos

- As cores são ferramentas importantes para visualização de dados e podem ser usadas com objetivos distintos:
 - usar cores para distinguir grupos de dados
 - usar cores para representar valores de dados
 - usar cores para realçar

COR COMO UMA FERRAMENTA PARA DISTINGUIR

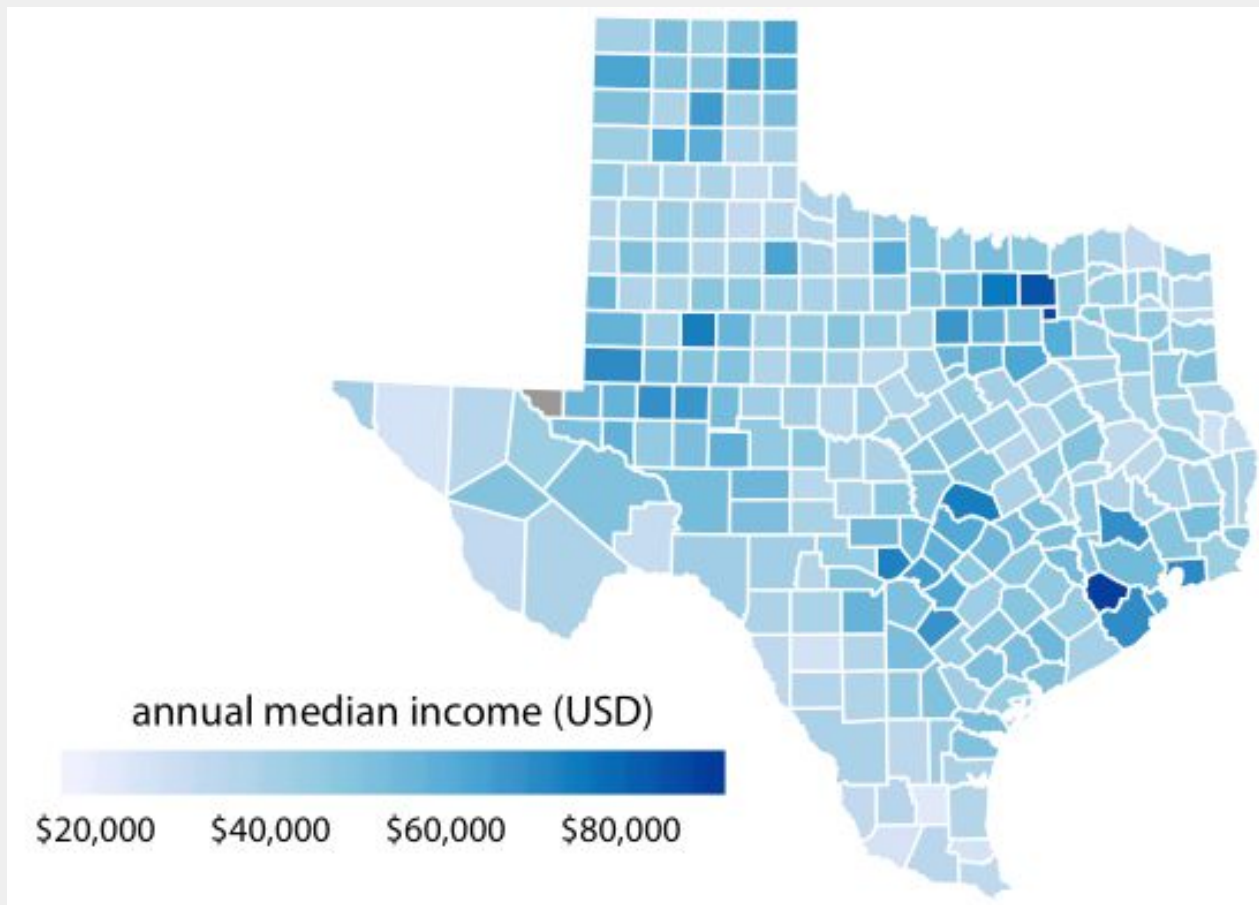
- Frequentemente cores são usadas como um meio de distinguir itens ou grupos distintos que não têm uma ordem intrínseca.
- Escala qualitativa de cores.



Crescimento populacional nos EUA de 2000 a 2010.

COR PARA REPRESENTAR VALORES DE DADOS

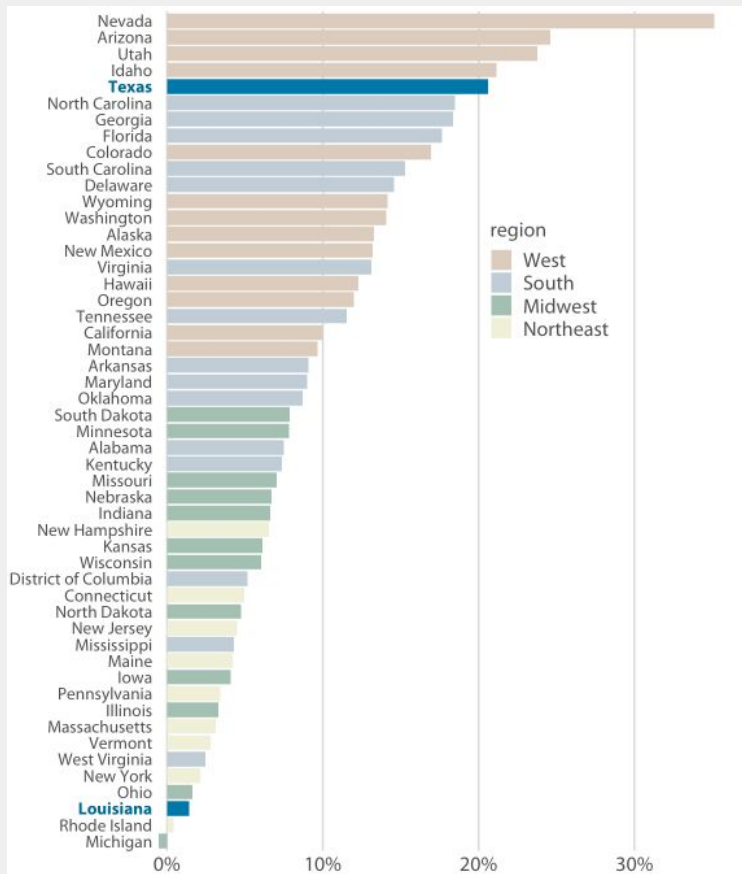
- Escala de cores sequencial.
- Sequência de cores que indica claramente:
 - quais valores são maiores ou menores que quais outros.
 - quão distantes dois valores específicos estão um do outro.



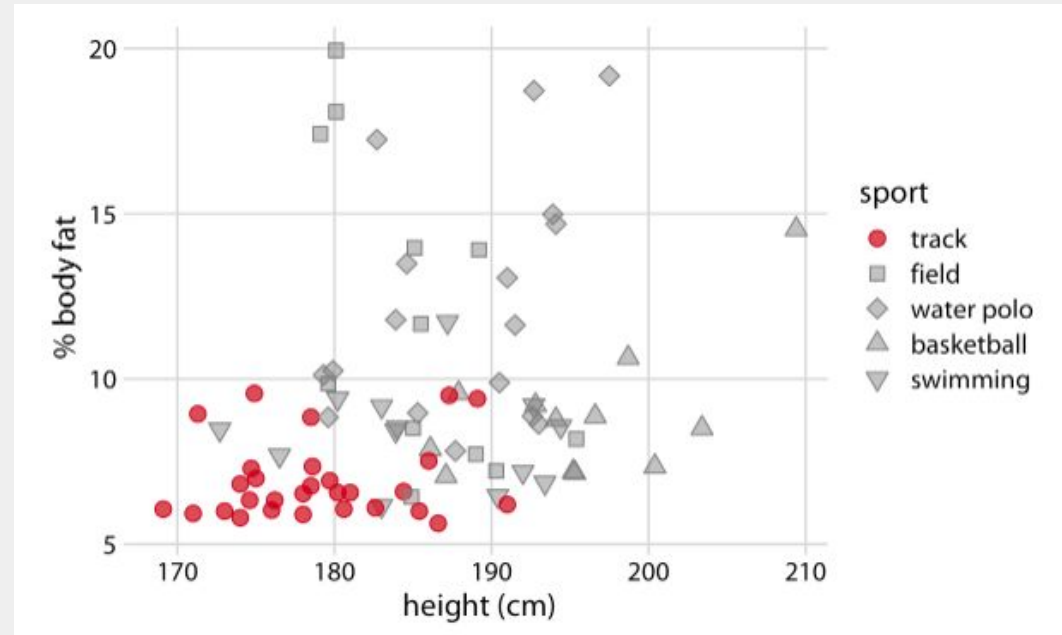
Renda anual mediana dos municípios Texas

COR COMO FERRAMENTA PARA DESTACAR

- Pode haver categorias ou valores específicos no conjunto de dados que carregam informações importantes.
- Escalas de cores acentuadas.

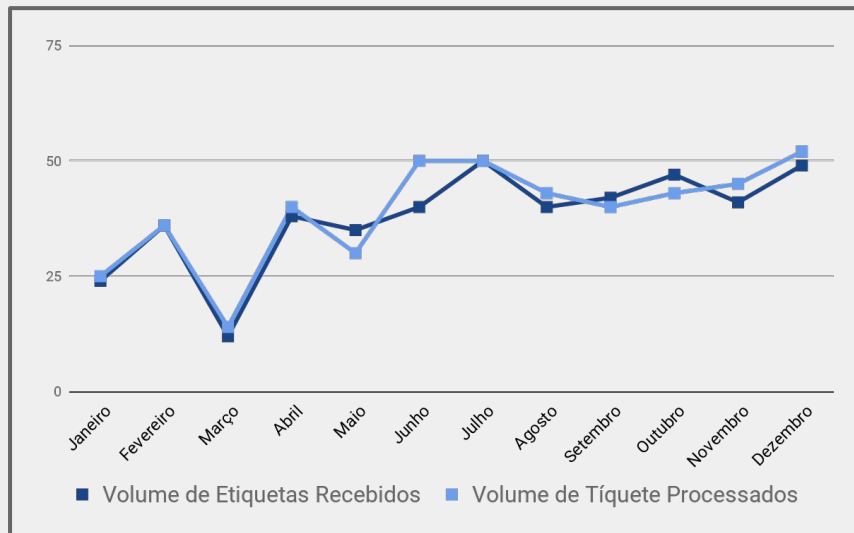


Estados maior e menor crescimento populacional nos EUA



Atletas de esportes populares e sua porcentagem de gordura em relação a altura

ESTUDO DE CASO



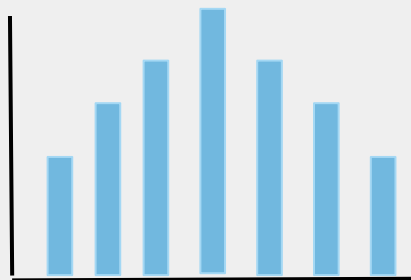
Antes

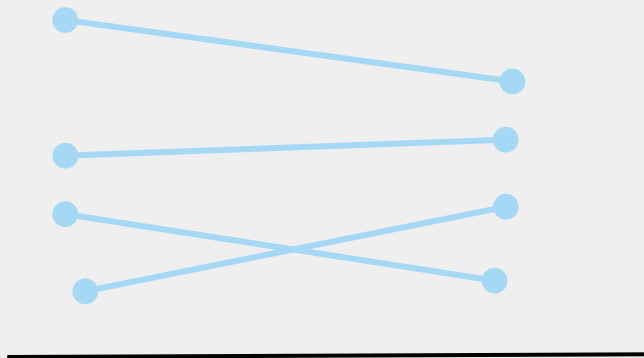


Depois

92%

VISUALIZANDO DADOS

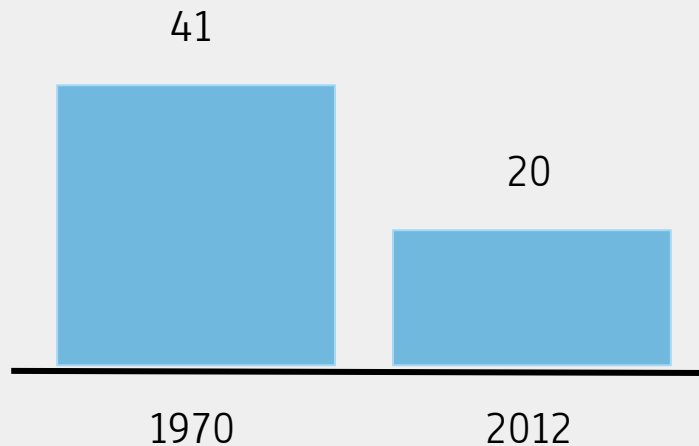




- Quando se tem apenas um número ou dois para compartilhar.
- O fato de você ter alguns números não significa a necessidade de um gráfico.

FILHOS COM MÃES “TRADICIONAIS” QUE FICAM EM CASA

% dos filhos de mulheres casadas que ficam em casa, enquanto os maridos trabalham



20%

dos filhos tinham **mães que trabalhavam em casa** em 2012, comparados com 41% em 1970.

- Interação bem com nosso sistema verbal, ou seja, as tabelas são lidas.
- Comunicar para um público misto, cujos membros procurarão linha de interesse particular.
- Interessante para comunicar várias unidades de medidas diferentes.

Em apresentações **ao vivo**:

- Não é uma boa ideia.
- Enquanto o público lê, você não é ouvido.

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$ X.X	Y%	Z.Z cm
Grupo 2	\$ X.X	Y%	Z.Z cm
Grupo 3	\$ X.X	Y%	Z.Z cm

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$ X.X	Y%	Z.Z cm
Grupo 2	\$ X.X	Y%	Z.Z cm
Grupo 3	\$ X.X	Y%	Z.Z cm

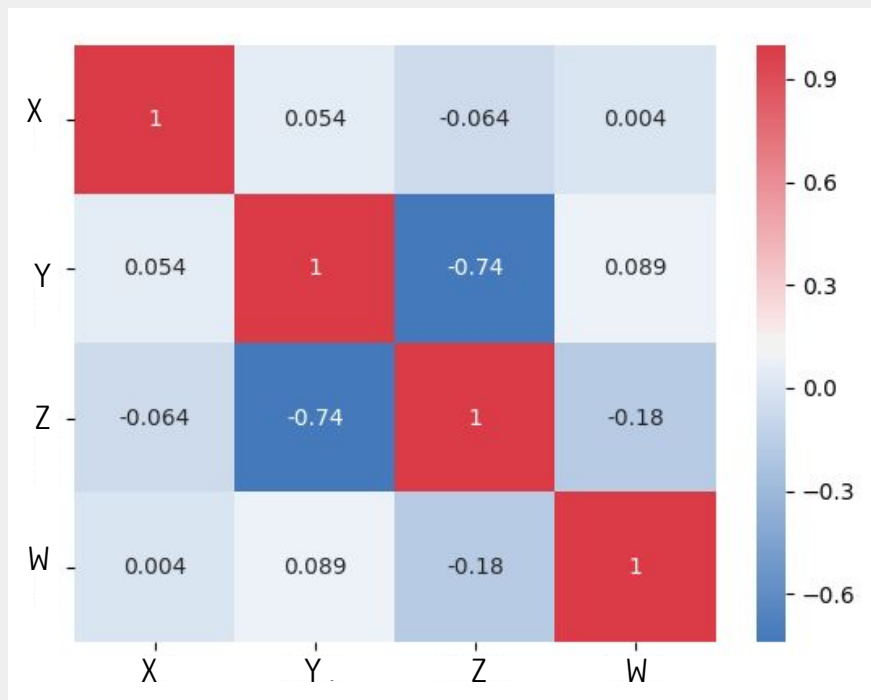
TABELAS

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$ X.X	Y%	Z.Z cm
Grupo 2	\$ X.X	Y%	Z.Z cm
Grupo 3	\$ X.X	Y%	Z.Z cm

- Modo de visualizar dados em formato tabular, no qual, as células coloridas transmitem a magnitude relativa dos números.
- Uso da saturação de cor para indicar pistas visuais.
- Útil para mostrar tendências.

MAPA DE CALOR

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	15%	22%	43%
Grupo 2	36%	50%	17%
Grupo 3	11%	55%	29%



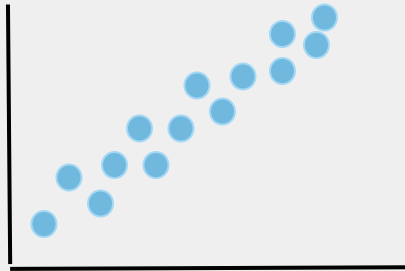
- Interação bem com nosso sistema visual.
- Gráficos bem projetados comunicam mais rápido que tabelas.
- Quatro categorias principais: pontos, linhas, barras e área.

GRÁFICO DE DISPERSÃO

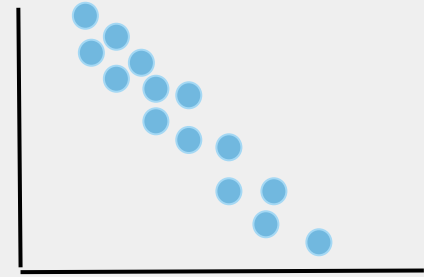
- Úteis para mostrar a relação entre duas variáveis.
- Cada par (x,y) é um ponto.
- Mais utilizado no meio acadêmico.



Sem correlação



Correlação positiva



Correlação negativa

GRÁFICO DE DISPERSÃO

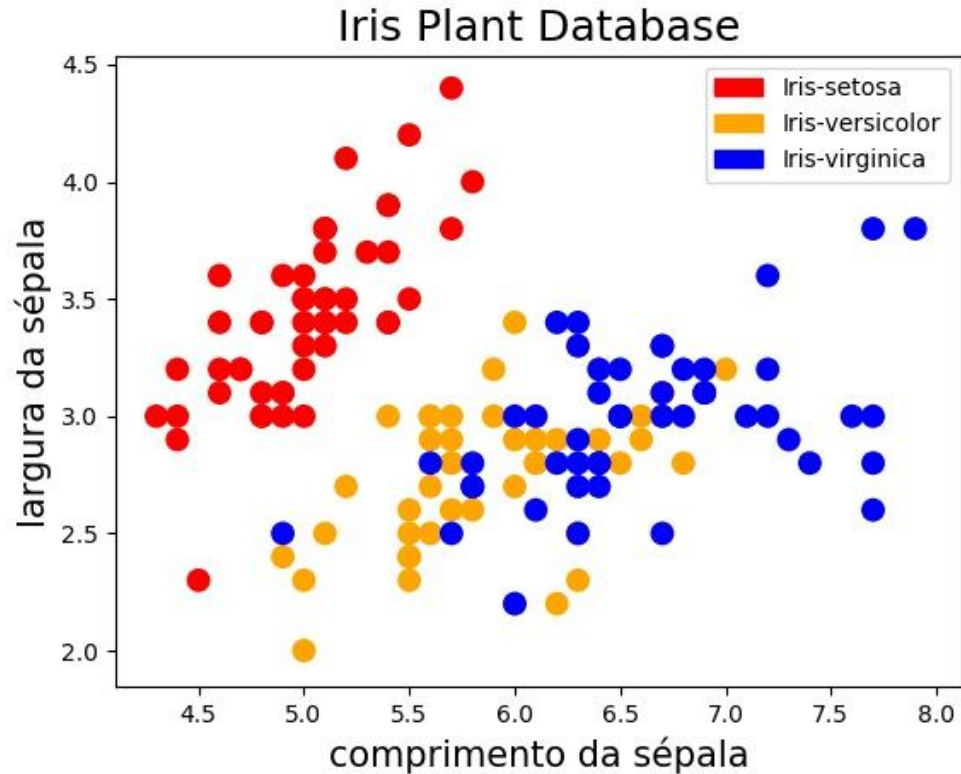


GRÁFICO DE LINHAS

- Mais usados para registrar dados contínuos.
- Uma, duas ou várias séries de dados.

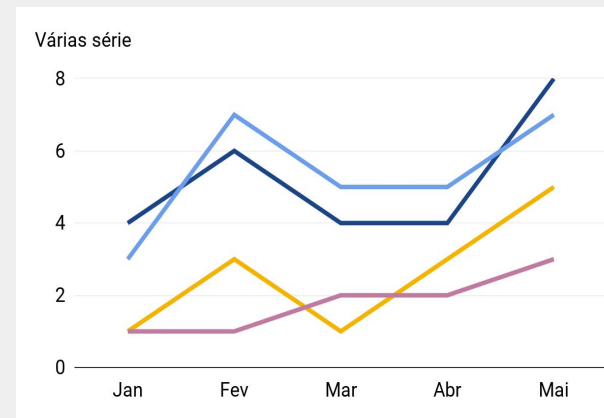
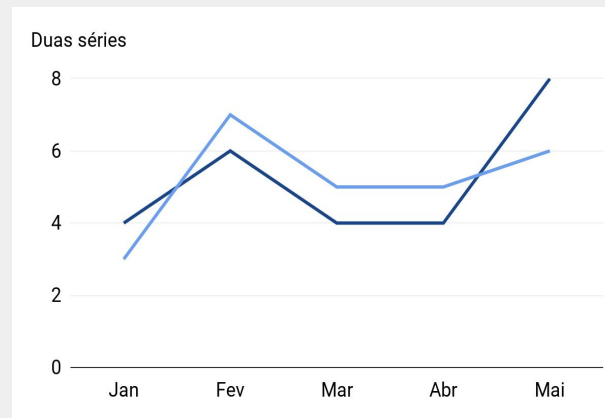
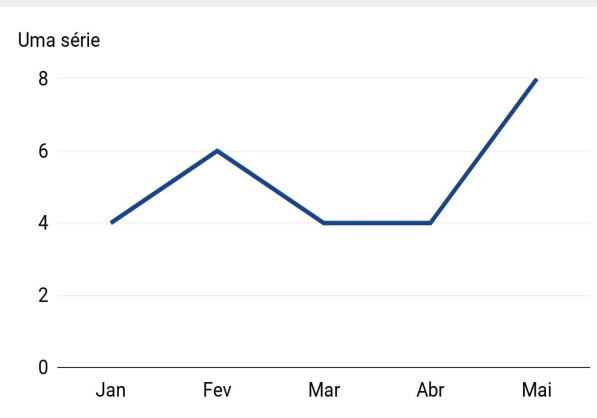
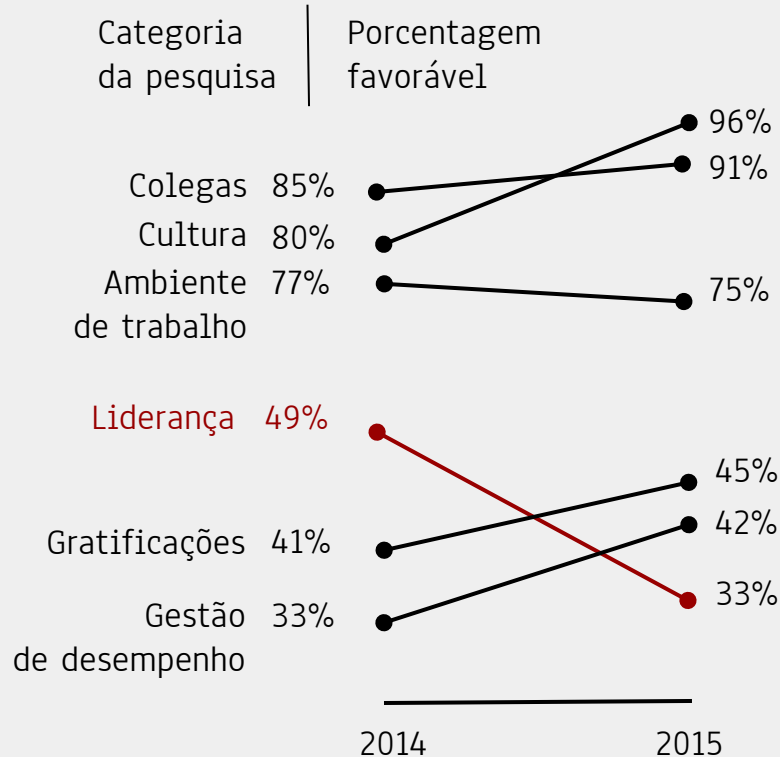


GRÁFICO DE INCLINAÇÃO

- Úteis quando há dois períodos de tempo ou pontos de comparação e existe a necessidade de se mostrar rapidamente aumentos e diminuições.

Opinião dos funcionários ao longo do tempo



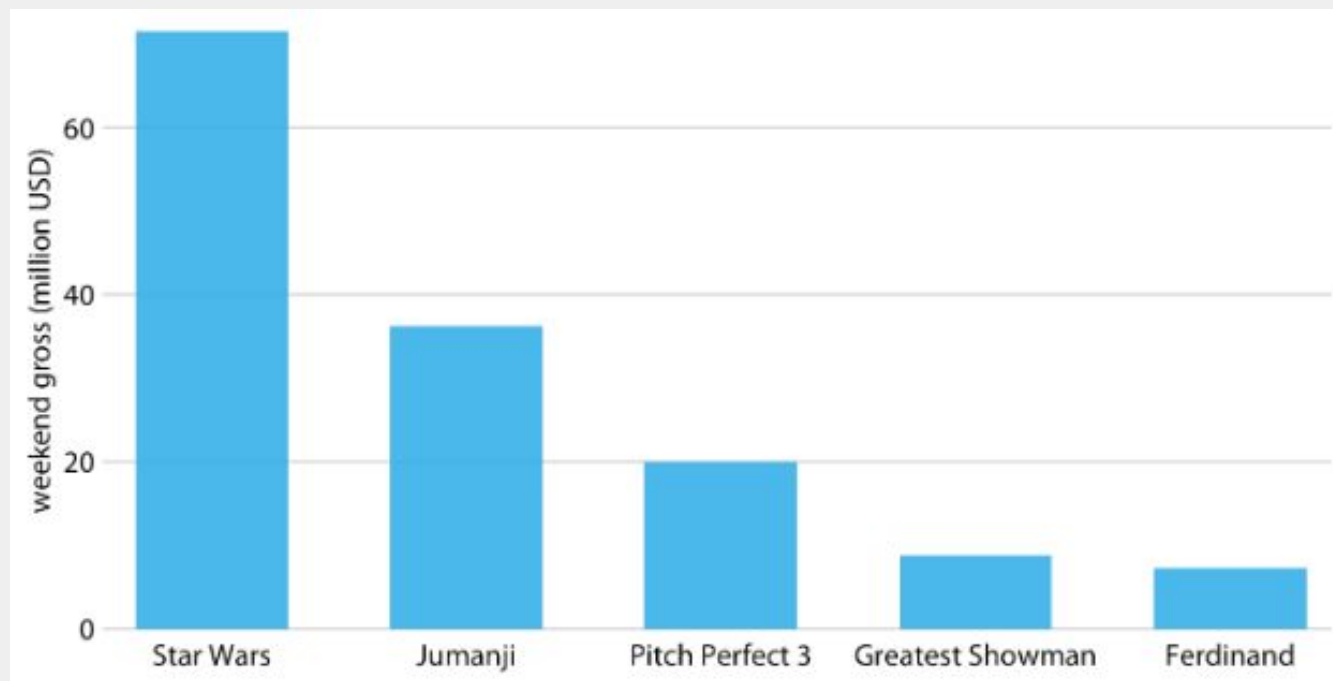
- Nossos olhos leem facilmente gráficos desse tipo.
- Importante que tenham uma linha de base zero, caso contrário comparação visual falsa.
- Fique atento a largura das barras.

GRÁFICO DE BARRAS

Classificação	Título	Fim de semana bruto
1	Guerra nas Estrelas: Os Últimos Jedi	\$ 71.565.498
2	Jumanji: Bem-vindo à selva	\$ 36.169.328
3	Passo perfeito 3	\$ 19.928.525
4	O Maior Showman	\$ 8.805.843
5	Ferdinand	\$ 7.316.746

Tabela 1. Filmes com maior bilheteria no fim de semana de 22 a 24 de dezembro de 2017

GRÁFICO DE BARRAS



Filmes com maior bilheteria no fim de semana
de 22 a 24 de dezembro de 2017

GRÁFICO DE BARRAS

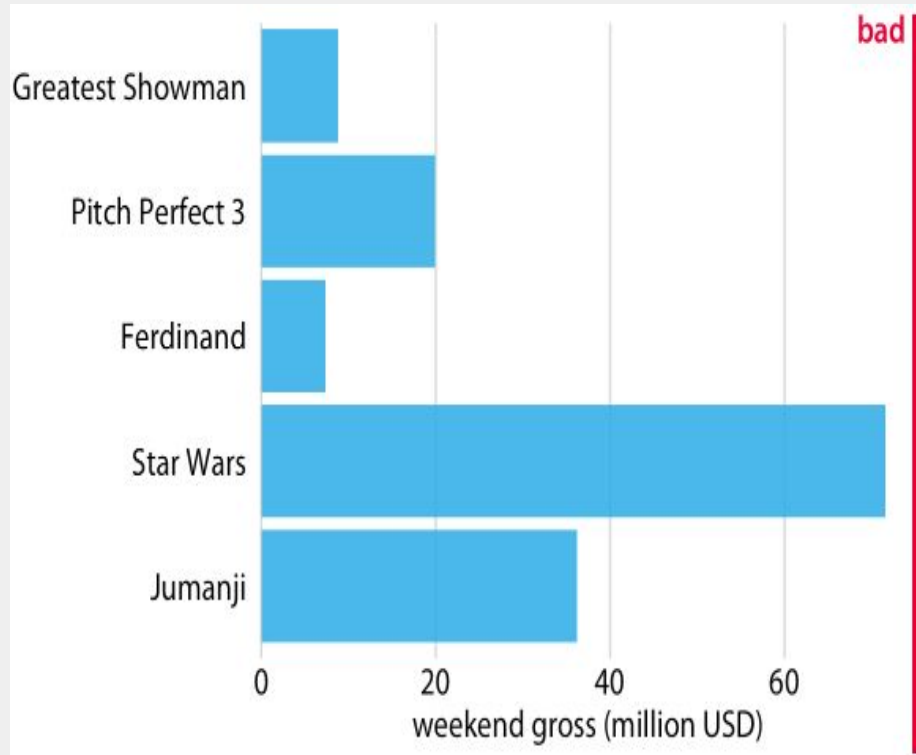
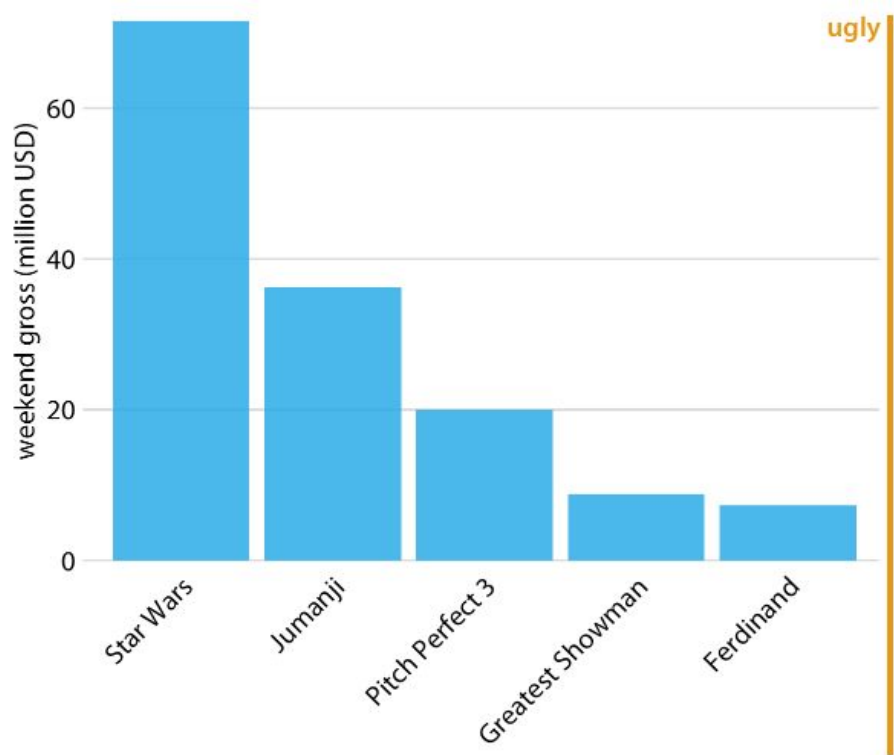
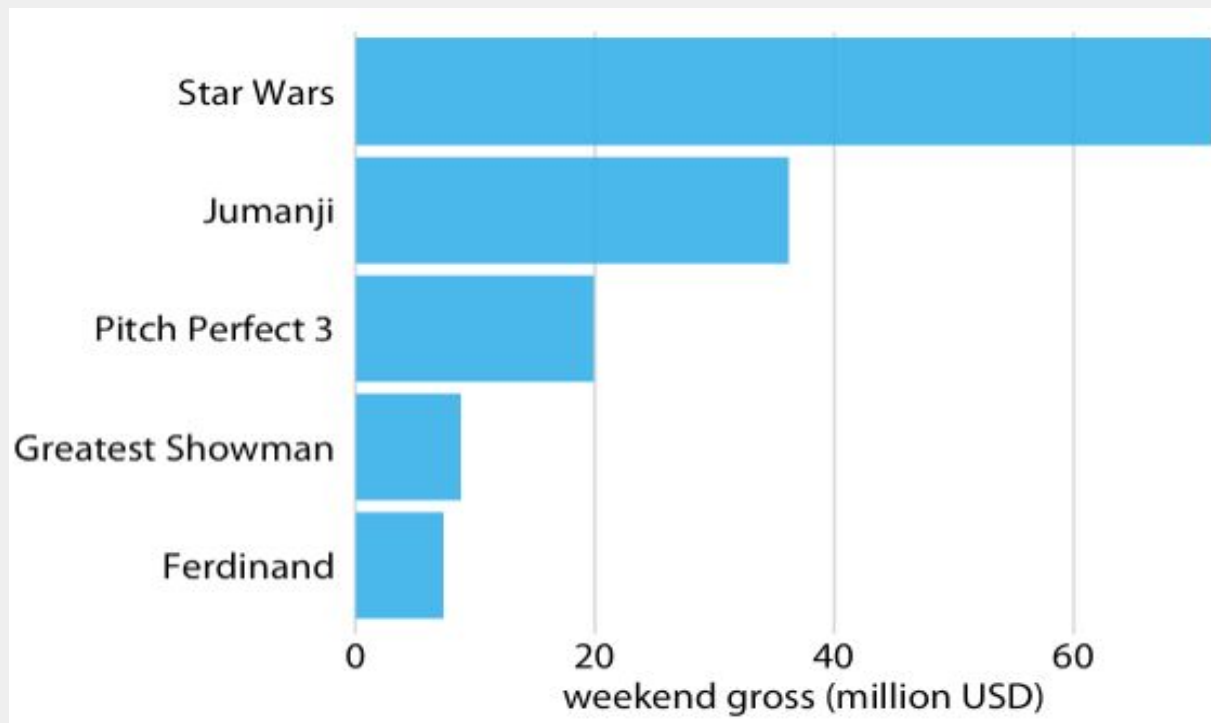


GRÁFICO DE BARRAS



Filmes com maior bilheteria no fim de semana
de 22 a 24 de dezembro de 2017

GRÁFICO DE BARRAS HORIZONTAIS EMPILHADAS

Estados com mais casos de covid-19 a cada 100 mil habitantes



Fonte:
Nexo Jornal
(09/04/2020)

	Gráfico de pizza	Barras empilhadas	Barras lado a lado
Visualiza claramente os dados como proporções de um todo	✓	✓	✗
Permite fácil comparação visual das proporções relativas	✗	✗	✓
Enfatiza visualmente frações simples, como 1/2, 1/3, 1/4	✓	✗	✗
Parece visualmente atraente, mesmo para conjuntos de dados muito pequenos	✓	✗	✓
Funciona bem quando o todo é dividido em vários pedaços	✗	✗	✓
Funciona bem para a visualização de muitos conjuntos de proporções ou séries temporais de proporções	✗	✓	✗

Coronavírus: aumento de casos e evolução das mortes no mundo



GRÁFICO

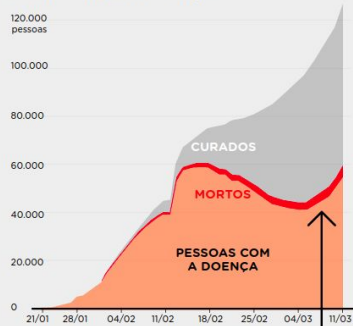
Coronavírus: aumento de casos e evolução das mortes no mundo

Gabriel Zanlorenssi e Lucas Gomes 12 de mar de 2020 (atualizado 13/03/2020 às 16h41)

Desde o dia 3 de março, a maioria das novas mortes por Covid-19 é registrada fora da China. A queda da mortalidade no país asiático está associada à restrição de circulação

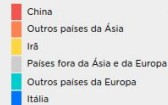


Relação entre número de infectados por coronavírus, curados e óbitos causados pela doença



Percentuais de novas mortes por coronavírus por região

A CADA RELATÓRIO DA OMS



MAIS RECENTES

EXPRESSO **Bolsonaro: entre notas de repúdio, deposição em debate e tutela** Estêvão Bertoni

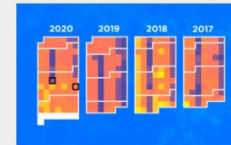
EXPRESSO **O acervo de imagens que 'tornam a internet incrível'** Juliana Domingos de Lima

EXPRESSO **Os vídeos que colocam em evidência os sons da culinária** Natan Novelli Tu

GRÁFICO **A trajetória de audiência do Big Brother. E o pico atual** Caroline Souza, Gabriel Zanlorenssi e Lucas Gomes

EXPRESSO **Os países comandados por mulheres que se destacam na crise** Juliana Domingos de Lima

GRÁFICOS



GRÁFICO

A trajetória de

RESUMINDO

- Entenda o contexto:
 - **Quem, O quê e Como**
- Escolha um visual adequado
- Elimine os ruídos no gráfico, tabelas e imagens
- Chame a atenção do público para onde você quer que o público foque
- Conte uma história



- Fundamentals of Data Visualization - Claus O. Wilke
 - <https://serialmentor.com/dataviz/>
- Storytelling com Dados: Um guia sobre visualização de dados para profissionais de negócios - Cole Nussbaumer Knaflitz
- Data Science do Zero - Primeiras Regras Com o Python - Joel Grus

