

Visualização de Dados e Data Storytelling

AULA 2

Gráficos e boas práticas

Slides, Artigos, Materiais...



DATA STORYTELLING E A DADIFICAÇÃO DE TUDO: UM GÊNERO BASTARDO DE MÃE NARRATIVA E PAI BANCO DE DADOS

Marcelo El Khouri Buzato

GÊNERO RELOADING

BUZATO, Marcelo. Data Storytelling e a dadificação de tudo: um gênero bastardo de mãe narrativa e pai banco de dados. In: LIMA-LOPES, R. E. DE; BUZATO, M. E. K. (Org.). Gênero Reloading. Campinas, SP: Potes Editores, 2018. p. 95–123.

DATA STORYTELLING E A DADIFICAÇÃO DE TUDO: UM GÊNERO BASTARDO DE MÃE NARRATIVA E PAI BANCO DE DADOS

Marcelo El Khouri Buzato
Universidade Estadual de Campinas

Resumo

A dadificação é o processo pelo qual volumes extraordinários de dados têm sido obtidos por empresas e governos via monitoramento e sensoramento do cotidiano das pessoas. Essa massa heterogênea ganha significado (vira informação) a partir do trabalho estatístico convertido em representações visuais chamadas visualizações e/ou narrativas de dados. Neste capítulo, caracterizo as visualizações e narrativas de dados por referência aos métodos e pressupostos ontológicos e epistemológicos do quantitativismo. Após mostrar que o enquadre desses objetos semióticos nas teorias de gêneros discursivos é problemático, a despeito do seu caráter retórico e enunciativo-discursivo vir paulatinamente ganhando reconhecimento, proponho que uma forma de estudar visualizações e narrativas de dados como gêneros discursivos é caracterizá-las como gêneros híbridos de narrativa e banco de dados.



Sobre o artigo

- A **dadificação** do mundo
 - operacionalizada por um tipo de computação **ubíqua, móvel, locativa, em nuvem e cognitiva**
- Visualização de dados e narrativas de dados como **gêneros discursivos**

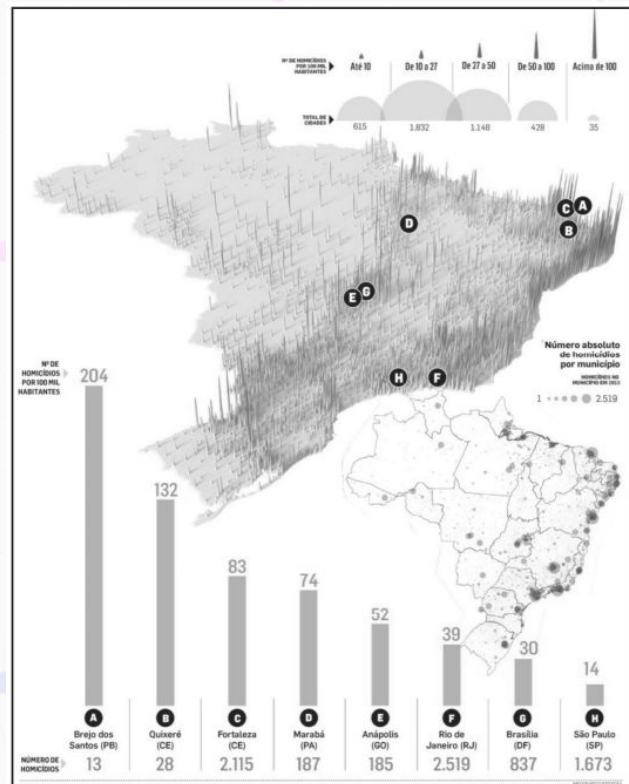


Figura 1 - Mapa da violência no Brasil por volta de 2010
Fonte: Adaptado de Estadão Infográficos (2010)

Definição de Data Storytelling?

- narrativas digitais em que “a análise ou a mensagem não se concentra nos argumentos, opiniões, atos ou eventos, mas na **interpretação contextual** de uma informação e suas relações com outros dados” (Rocha, 2016)
- “um método de transmitir mensagens derivadas da análise de dados complexos de forma a permitir que o público **assimile** rápida e facilmente determinado material (...)” (Gilbert, 2017)
- “processos sequenciais de **definição do contexto, seleção de informação, escolha de modalidade e definição de uma ordem** que possa transmitir a narrativa desejada de forma efetiva, onde quer seja ficcional ou para comprovação dos fatos (trial evidence), a sequência e forma de agrupamento utilizados numa narrativa afetam o significado que é construído” (Hullman et al, 2013)

Definição de Data Storytelling?

- relatos de eventos reais ou imaginários que envolvam o ouvinte em **comunicação interativa (Gilbert, 2017)**

Interface de dados \neq Data Storytelling

Definição de Data Storytelling?

Todo artefato digital carrega duas camadas simultâneas:

- A camada **humana/cultural**: imagens, histórias, metáforas.
- A camada **computacional**: códigos, algoritmos, dados.

Narrativas → sequência, causa e efeito

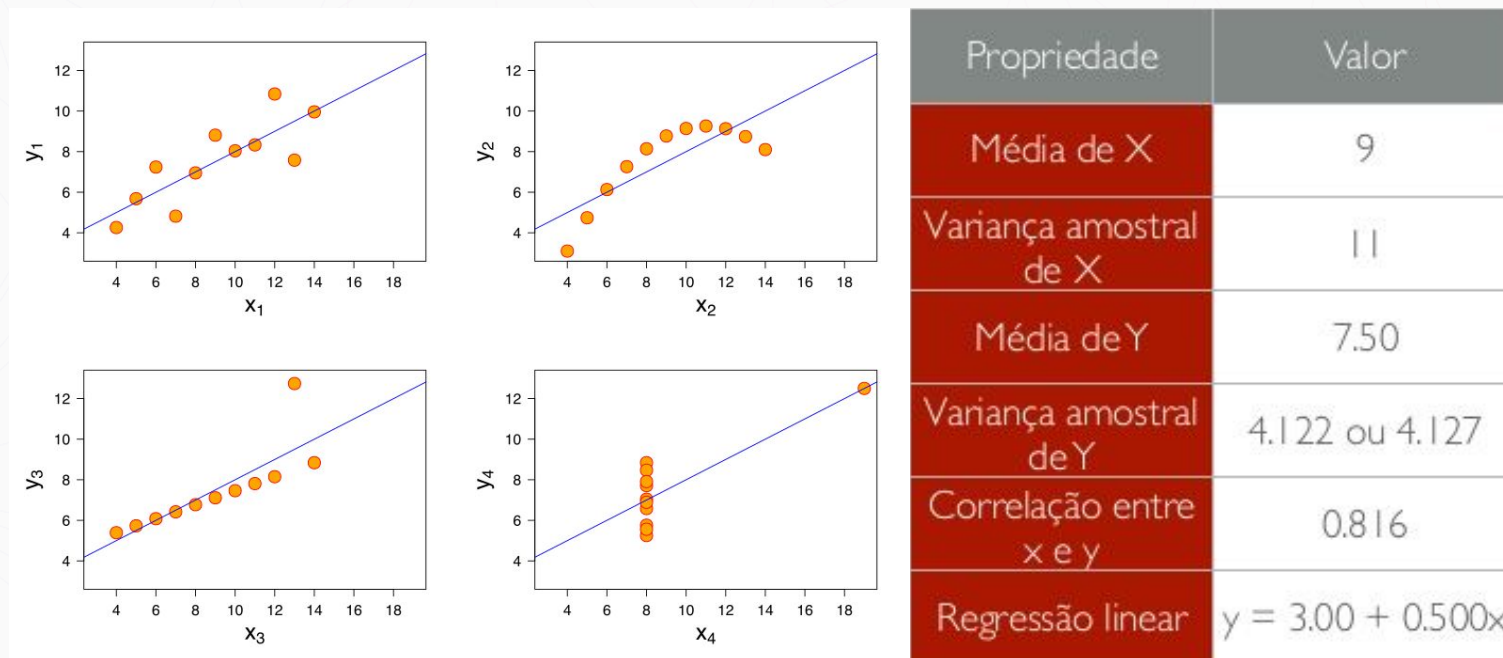
Bancos de dados → relações, padrões, exploração

Ambos **coexistem** (ex: Netflix, Spotify, Facebook)

O desafio do data storytelling é criar histórias com sentido usando uma estrutura associativa de dados.

Sobre o artigo

- o Quarteto de Anscombe



A escolha de um visual eficaz

91%

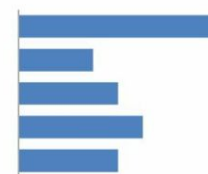
Texto simples



Gráfico de dispersão



Barras verticais



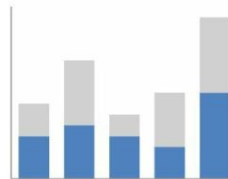
Barras horizontais

	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

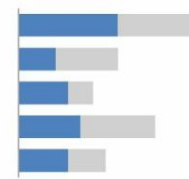
Tabela



Linha



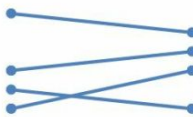
Barras verticais empilhadas



Barras horizontais empilhadas

	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

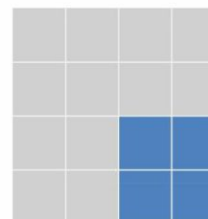
Mapa de calor



Mapa de inclinação



Cascata



Área quadrada

Texto simples

Use números isolados para máximo impacto.

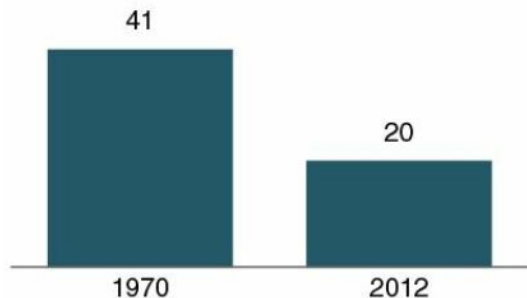
- Ideal para **1 ou 2 números-chave**
- Destaque visual direto (ex: número em destaque)
- Use **palavras mínimas** de apoio
- Evite tabelas ou gráficos nesses casos
- Gráficos podem diluir o impacto
- Comunicação **rápida e direta**

Quando um número já comunica tudo (ex: “93% de aprovação”).

Texto simples

Filhos com mães "tradicionais" que ficam em casa

*% dos filhos com mães
casadas que ficam em casa,
enquanto maridos trabalham*



Observação: baseado em filhos menores de 18 anos. Suas mães foram classificadas de acordo com a situação de emprego de 1970 e 2012.

Fonte: análise do March Current Population Surveys Integrated Public Use Microdata Series (IPUMS - CPS) feita pelo Pew Research Center, 1971 e 2013.

Adaptado do PEW RESEARCH CENTER



20%

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

Tabelas

Comunique seus dados de forma detalhada

- Leitura **linha a linha**
- Boa para **múltiplas medidas**
- Útil para **públicos diversos**
- Evite em falas ao vivo
- Design discreto valoriza os dados

Relatórios, anexos ou dashboards interativos — não como foco em slides falados.

Bordas grossas

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$X.X	Y%	Z,ZZZ
Grupo 2	\$X.X	Y%	Z,ZZZ
Grupo 3	\$X.X	Y%	Z,ZZZ
Grupo 4	\$X.X	Y%	Z,ZZZ
Grupo 5	\$X.X	Y%	Z,ZZZ

Bordas claras

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$X.X	Y%	Z,ZZZ
Grupo 2	\$X.X	Y%	Z,ZZZ
Grupo 3	\$X.X	Y%	Z,ZZZ
Grupo 4	\$X.X	Y%	Z,ZZZ
Grupo 5	\$X.X	Y%	Z,ZZZ

Bordas mínimas

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$X.X	Y%	Z,ZZZ
Grupo 2	\$X.X	Y%	Z,ZZZ
Grupo 3	\$X.X	Y%	Z,ZZZ
Grupo 4	\$X.X	Y%	Z,ZZZ
Grupo 5	\$X.X	Y%	Z,ZZZ

Mapa de calor

Realça padrões em tabelas usando cores.

- Células coloridas indicam magnitude
- Facilita identificação de extremos
- Menos esforço mental que números puros
- Use saturação com legenda clara
- Ideal com formatação condicional

Ótimo para destacar padrões em tabelas densas (ex: desempenho por região).

Mapa de calor

Tabela

	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

Mapa de Calor

BAIXO-ALTO

	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

Gráfico de Dispersão

Revela relações entre duas variáveis.

- Eixos X e Y simultâneos
- Mostra correlação ou padrões
- Comum em ciência, menos em negócios
- Pode destacar exceções ou médias
- Design pode guiar foco visual

Análise de custo vs. uso, desempenho vs. tempo, correlações em dados quantitativos.

Gráfico de Dispersão

Custo por milha por milhas dirigidas

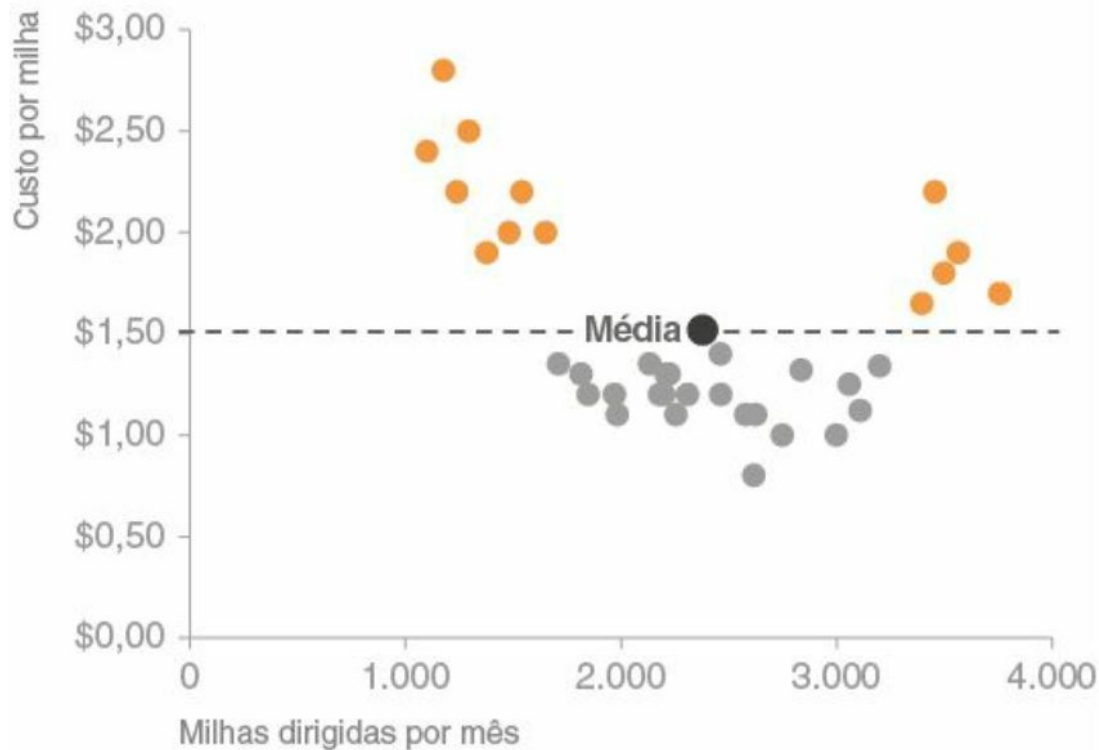


Gráfico de Linhas

Mostra variações ao longo do tempo.

- Ideal para séries temporais
- Pode ter 1 ou várias linhas
- Eixo X exige intervalos regulares
- Pode incluir médias e intervalos
- Reforça tendências e flutuações

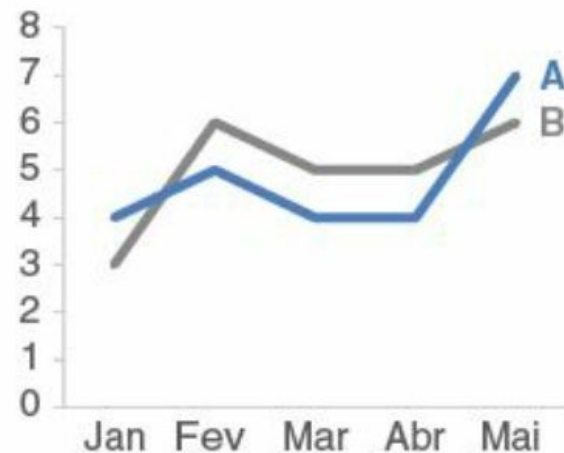
Evolução de indicadores (ex: visitas mensais, média de espera, previsão de demanda).

Gráfico de Linhas

Uma série



Duas séries



Várias séries

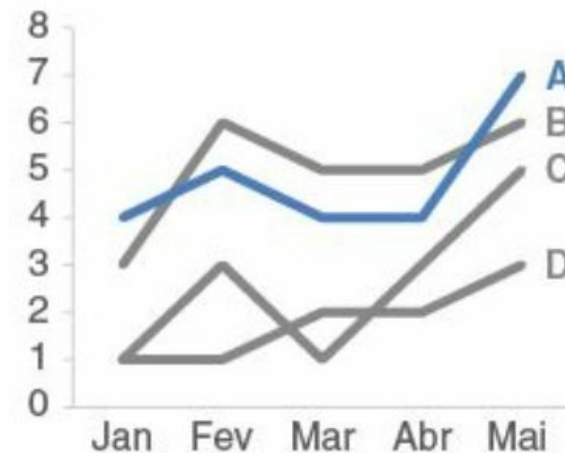


Gráfico de Linhas

Tempo de espera no controle de passaportes

Últimos 13 meses

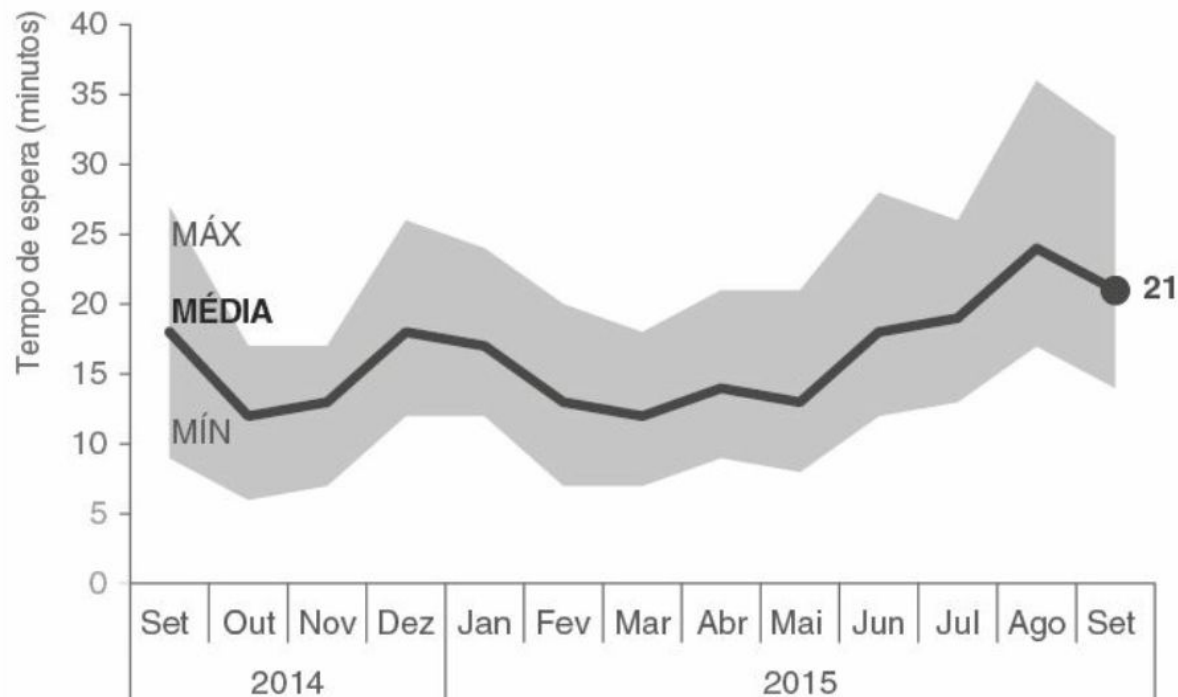


Gráfico de Inclinação

Compara mudanças entre dois pontos no tempo.

- Mostra aumentos ou quedas visuais
- Usa dois pontos por categoria
- Inclinação indica direção da mudança
- Intuitivo e direto
- Exibe valores absolutos e variação

Comparações antes/depois (ex: resultados de pesquisa em 2 anos)

Gráfico de Inclinação

Opinião dos funcionários ao longo do tempo

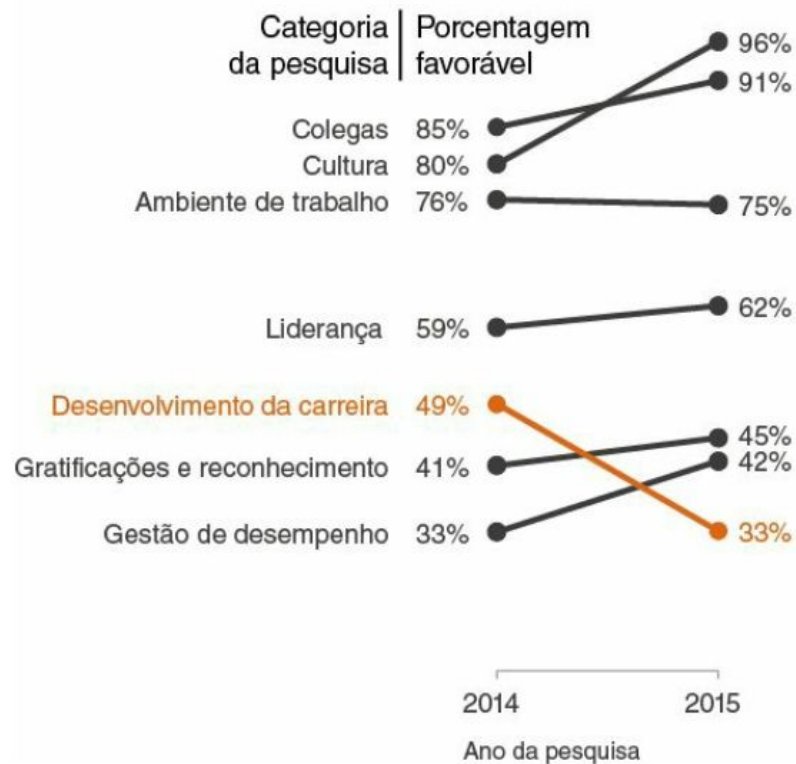


Gráfico de Barras

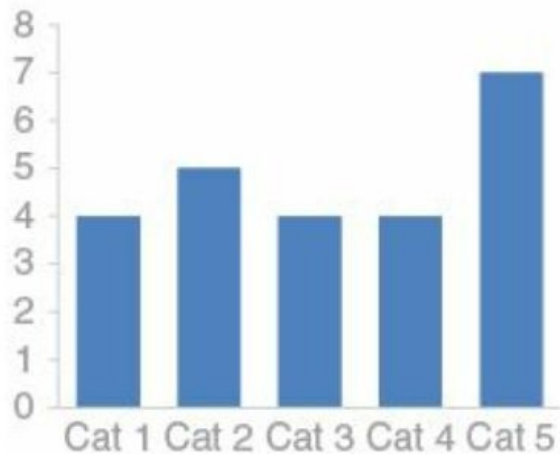
Clareza visual para comparar categorias.

- Fácil leitura e comparação
- Público já familiarizado
- Destaque em extremos das barras
- Sempre usar linha de base zero
- Evite distorções visuais

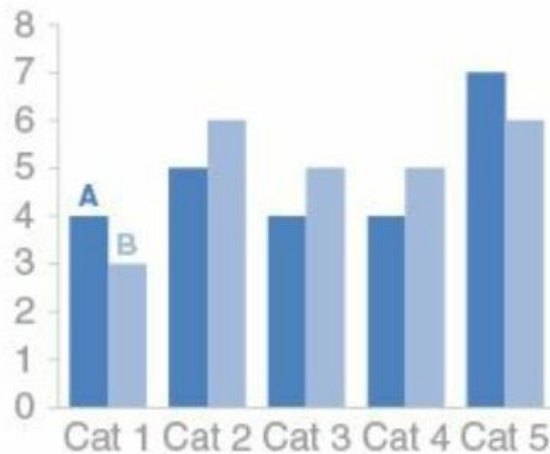
Comparar categorias (ex: número de visitantes por cidade).

Gráfico de Barras

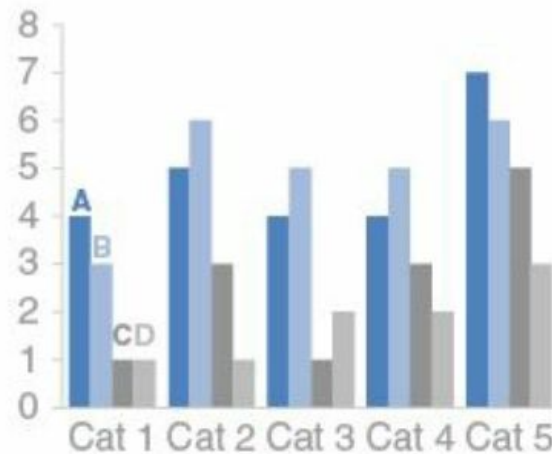
Uma série



Duas séries



Várias séries



Barras Horizontais ou Verticais?

Barras horizontais facilitam a leitura de categorias.

- Leitura natural da esquerda para a direita
- Melhores para nomes longos
- O público lê categoria antes do valor
- Pode ter 1 ou várias séries
- Ordenação lógica importa

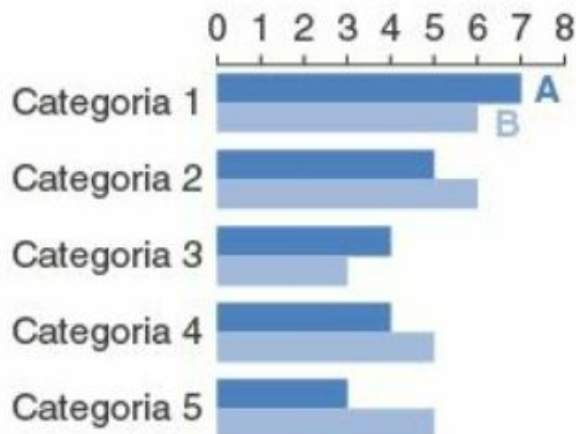
Listas categóricas com rótulos extensos (ex: políticas públicas por região).

Barras Horizontais ou Verticais?

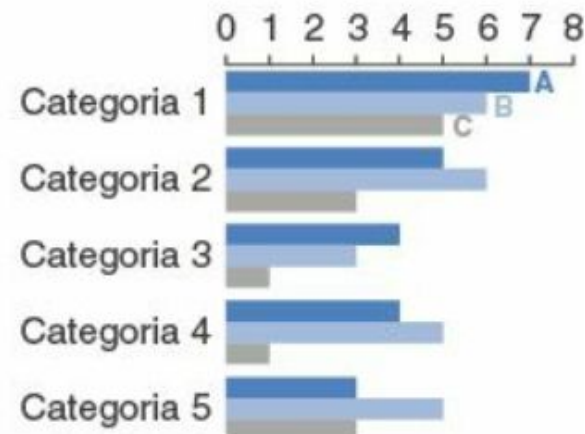
Uma série



Duas séries



Várias séries



Barras Empilhadas

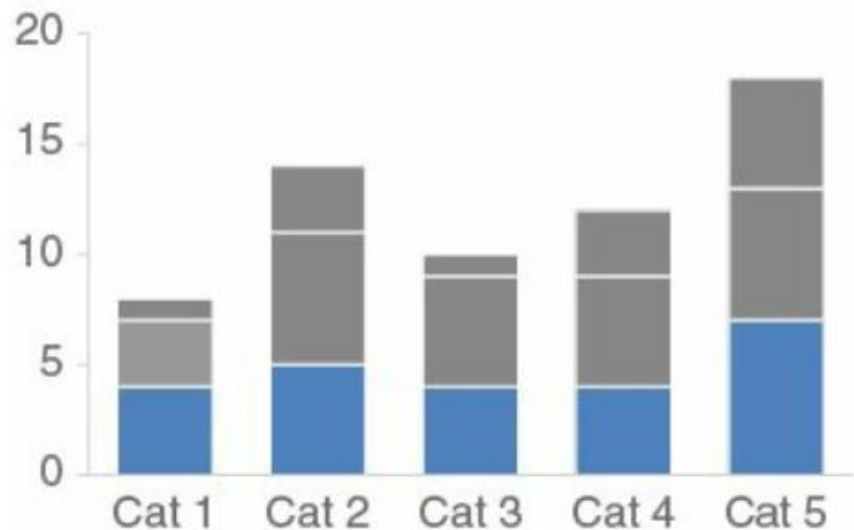
Compara totais e subcomponentes por categoria.

- Mostra partes e o todo
- Útil para comparar totais
- Difícil comparar subpartes
- Linha de base só para a base
- Pode usar valores ou percentuais

Distribuição interna por categoria (ex: tipo de gasto dentro de cada setor).

Barras Empilhadas

Comparar *estas* é fácil



Comparar *estas* é difícil

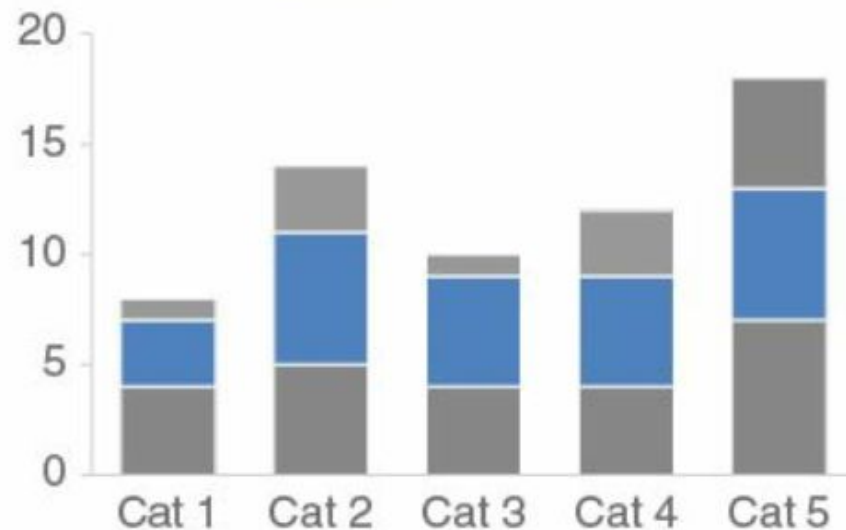


Gráfico de Cascata

Mostra variações acumuladas entre início e fim.

- Destaca aumentos e reduções
- Indica ponto inicial e final
- Ótimo para mudanças acumuladas
- Pode ser simulado com barras empilhadas
- Exige ajuste manual em alguns softwares

Evolução do total de funcionários, fluxo de receitas e despesas, variação de estoque.

Gráfico de Cascata

Número de funcionários em 2014

Embora mais funcionários tenham entrado do que saído da equipe, a contratação agressiva significa que o número de funcionários (NF) aumentou em 16% no decorrer do ano.

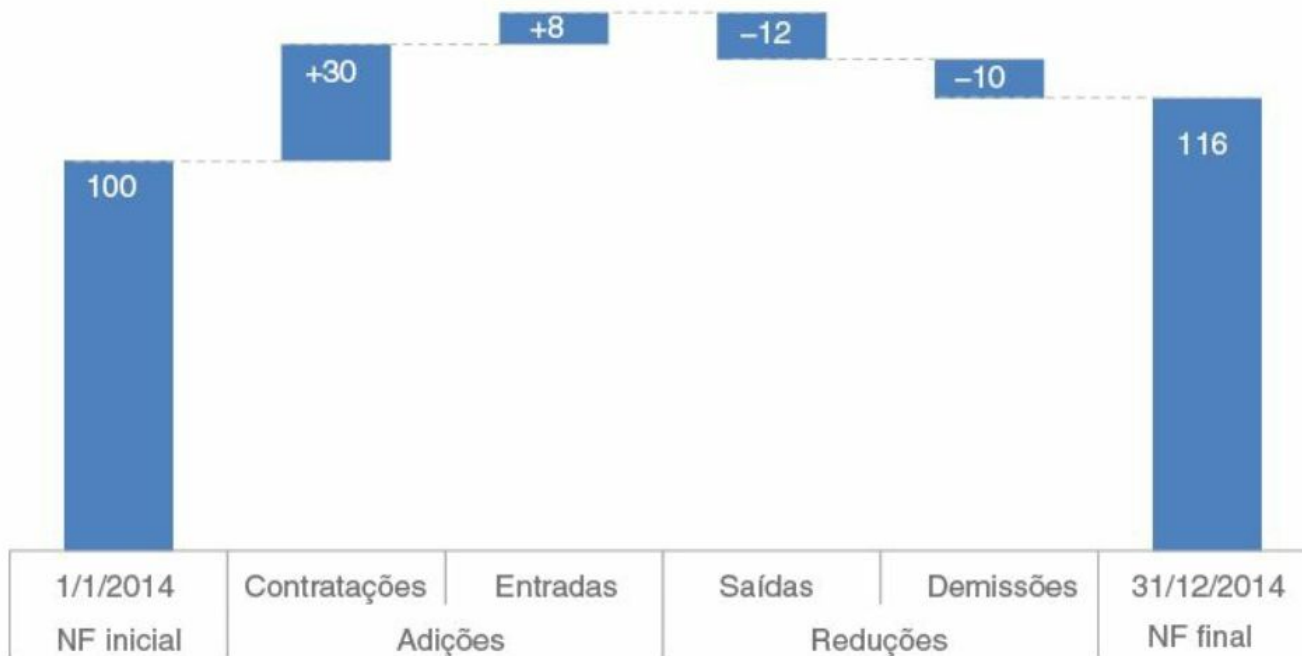


Gráfico de Área

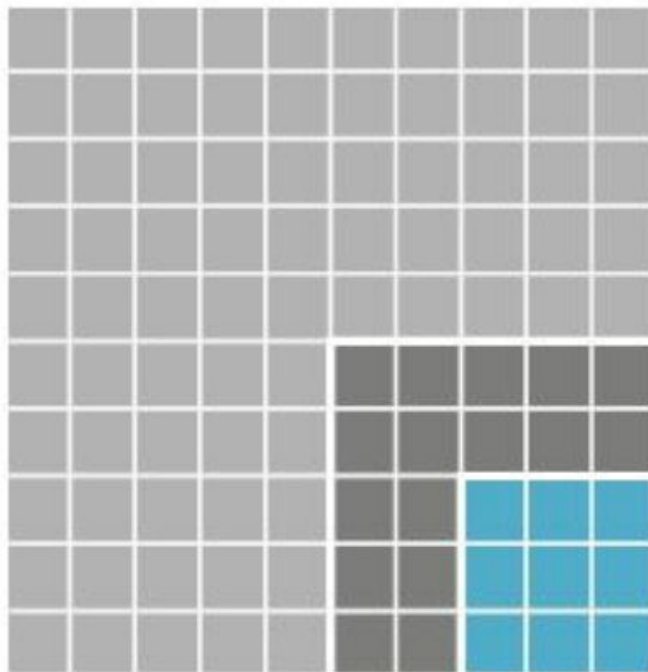
Evite, salvo em casos de magnitudes muito distintas.

- Leitura quantitativa imprecisa
- Espaço 2D confunde proporções
- Pouco intuitivo visualmente
- Útil para mostrar grandes diferenças
- Mais compacto que barras em certos casos

Comparar categorias com escalas muito diferentes (ex: população de países grandes vs. pequenos).

Gráfico de Área

Decomposição da entrevista



De cada 100 ligações telefônicas...

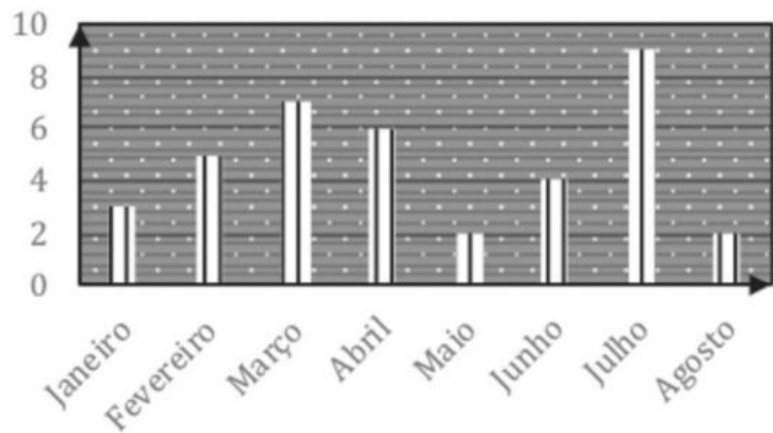
trouxemos 25 candidatos para entrevista...

e 9 propostas foram oferecidas.

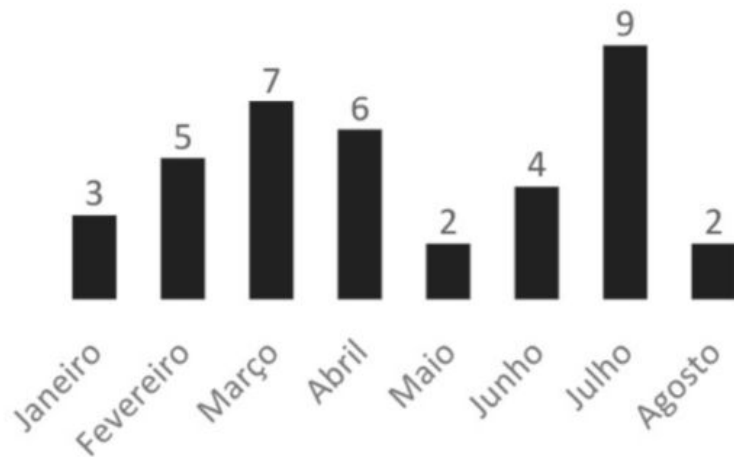
Boas práticas em visualização de dados

- Princípios e métricas no “design da informação”

Razão dados/tinta baixa

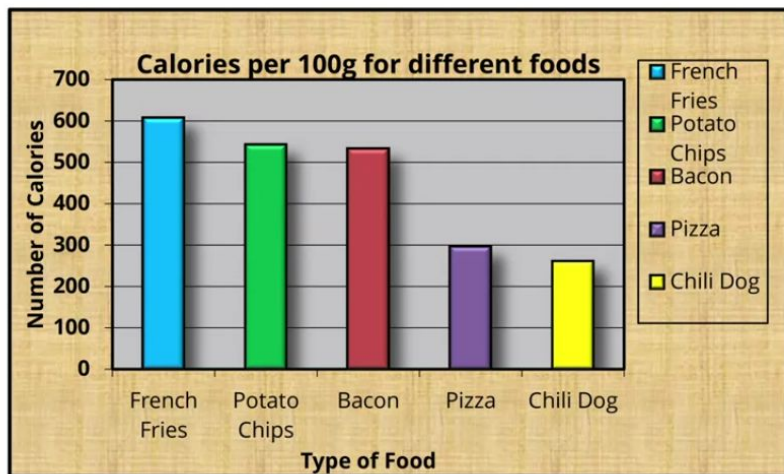


Razão dados/tinta alta

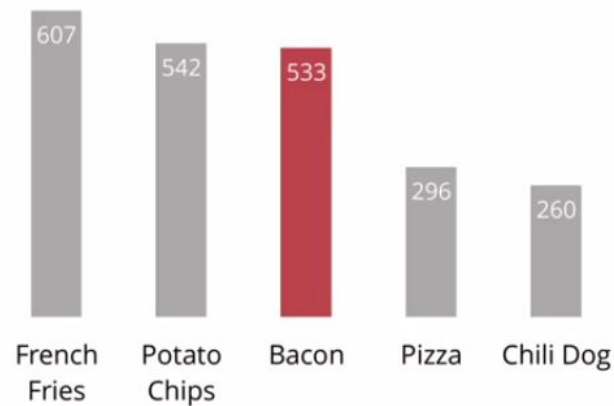


Boas práticas em visualização de dados

- Princípios e métricas no “design da informação”

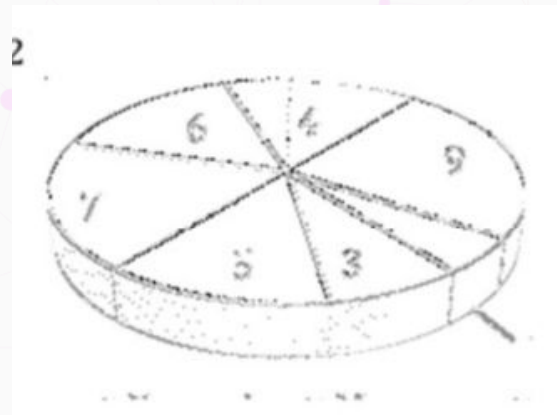
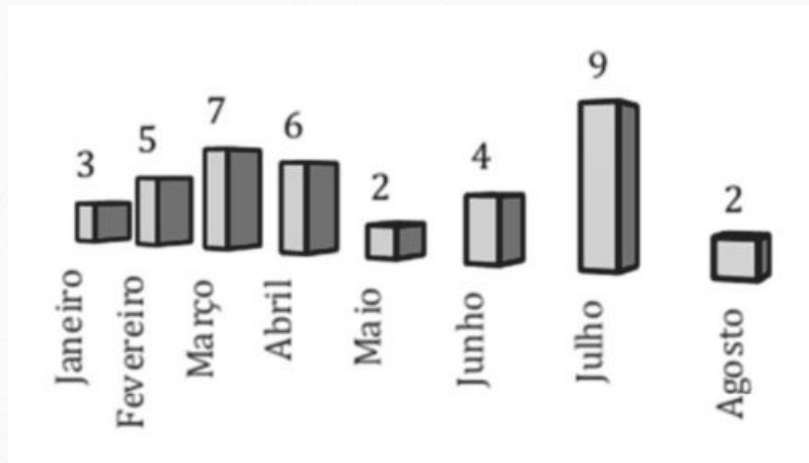


Calories per 100g



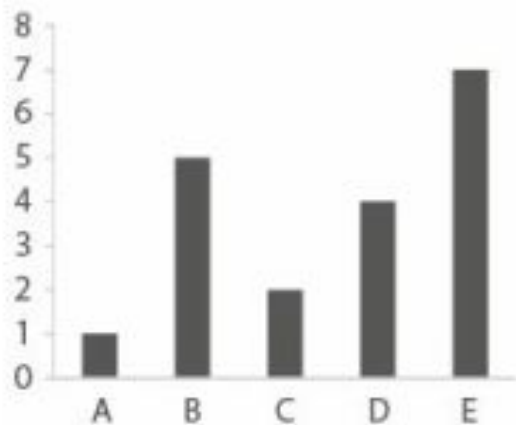
Boas práticas em visualização de dados

- O “fator mentira”

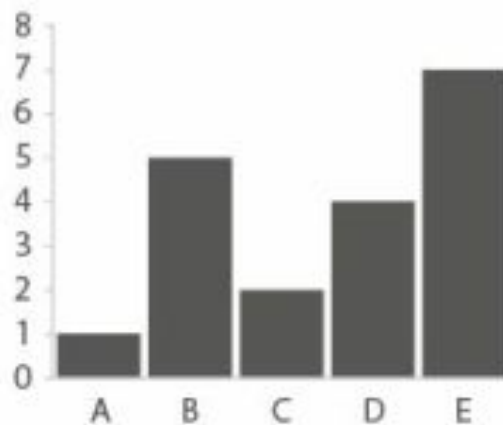


Boas práticas em visualização de dados

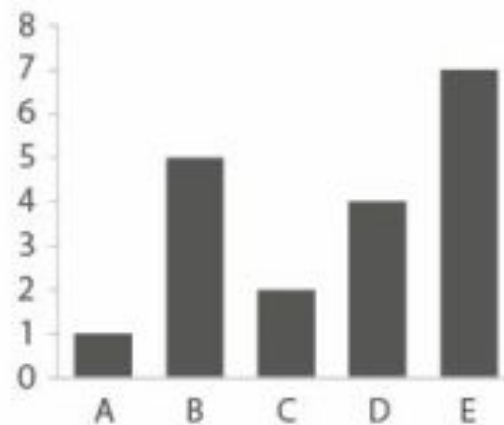
Finas demais



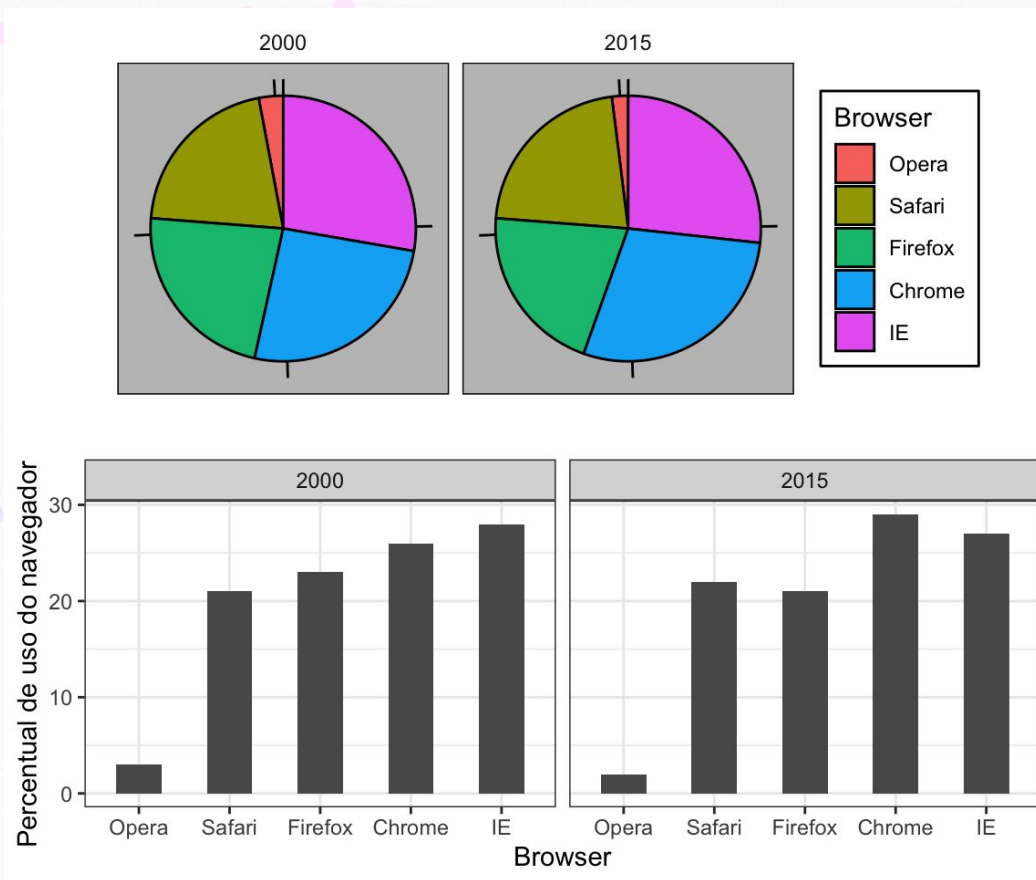
Grossas demais



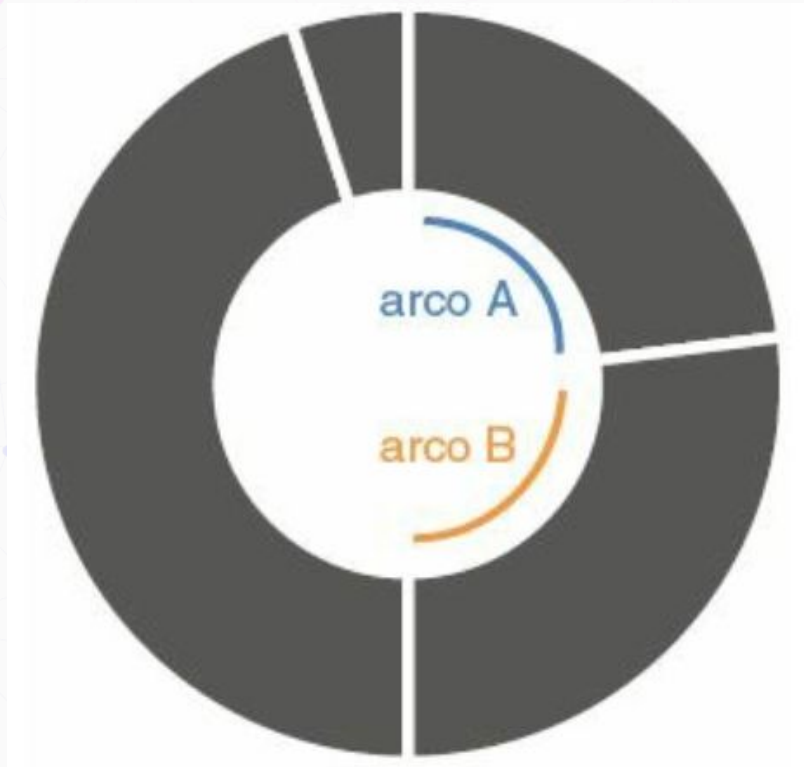
Corretas



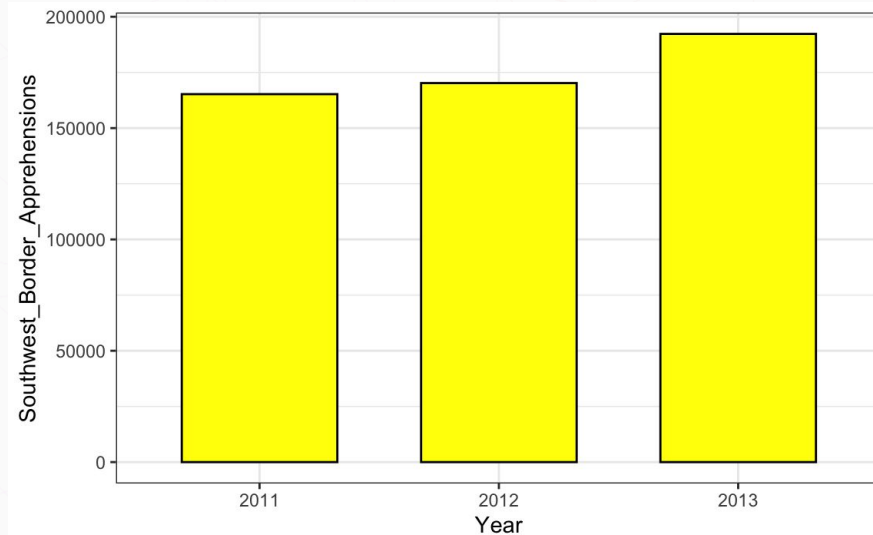
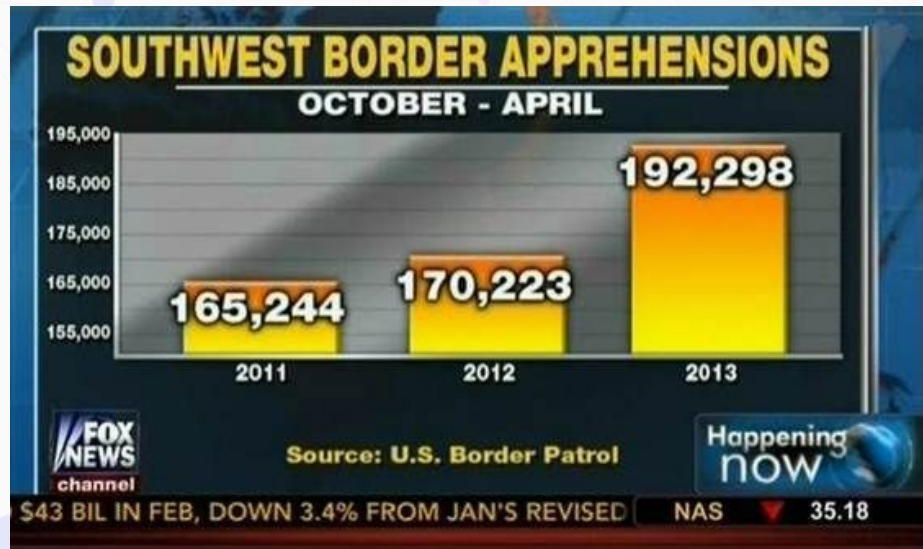
Lembram da briga do "gráfico de pizza"?



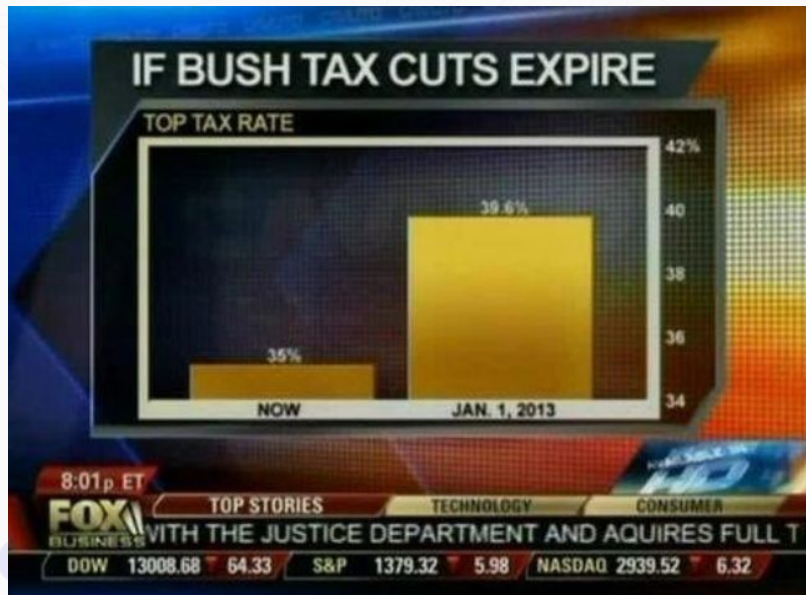
Também vale para o gráfico de rosca



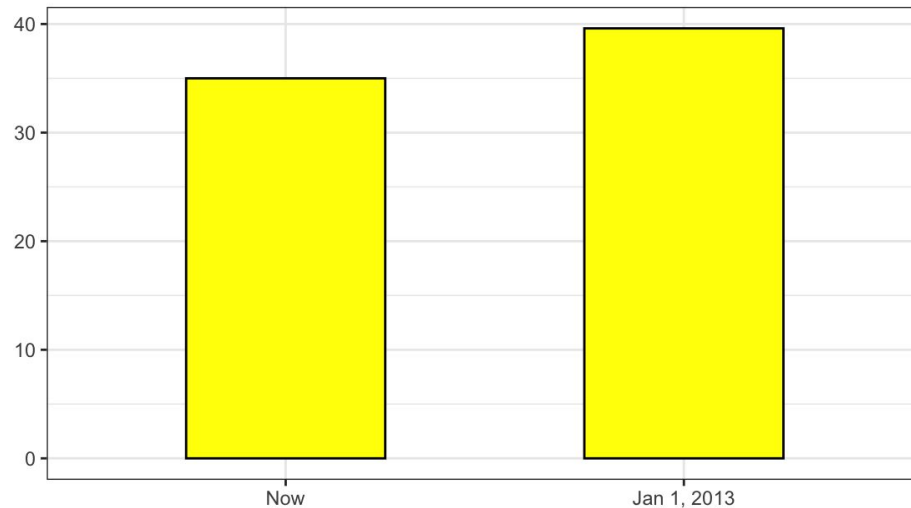
A escala importa



A escala importa



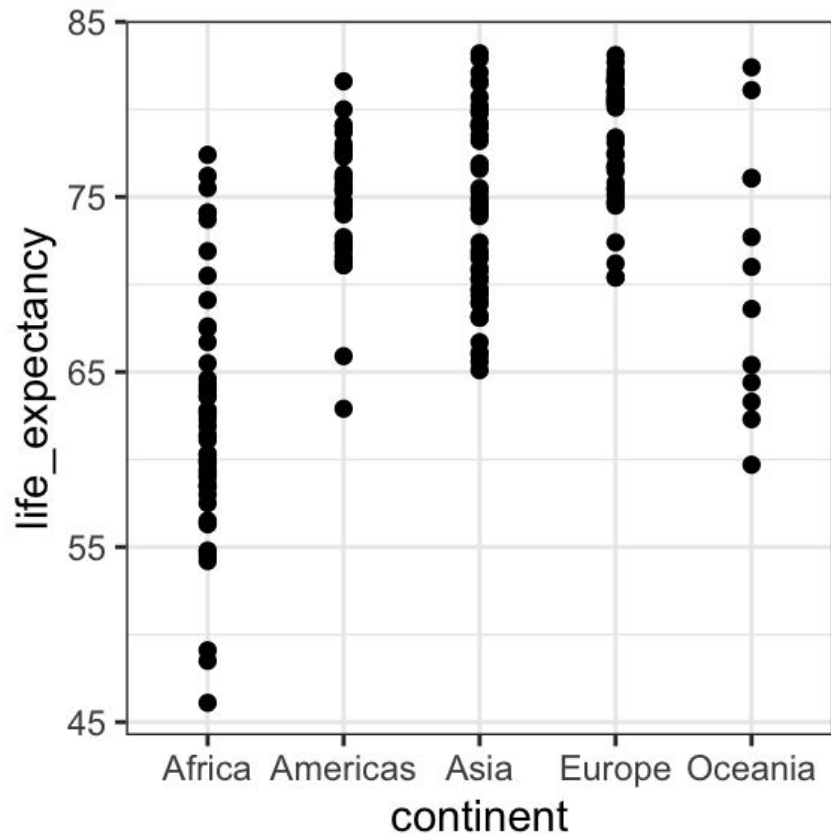
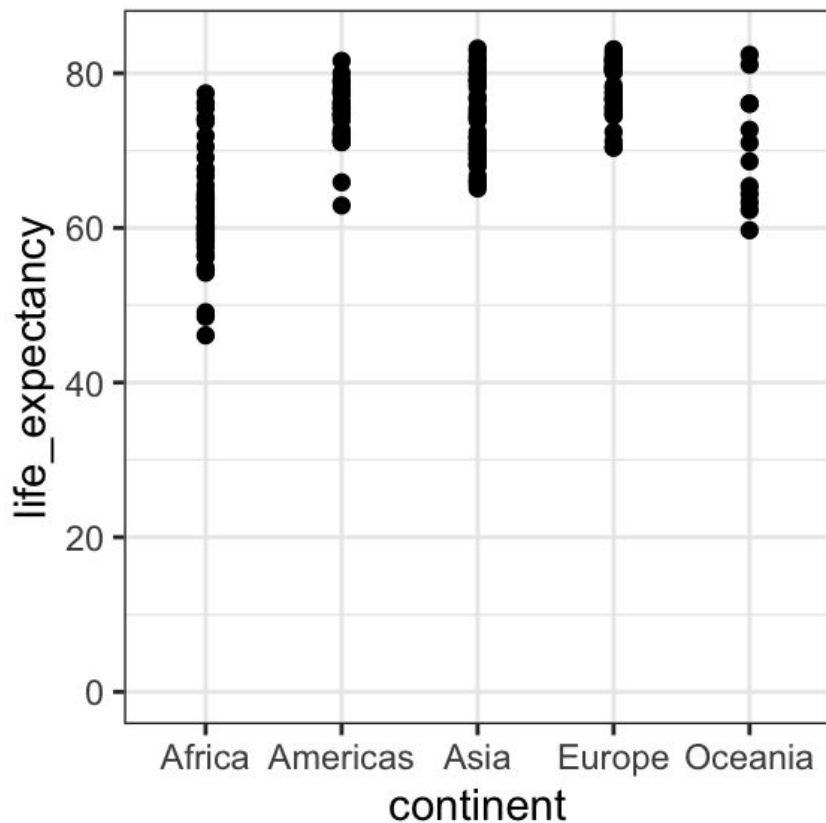
Aumento nas taxas se o corte de impostos de Bush expirar



A escala importa



A escala importa



Proporção também importa

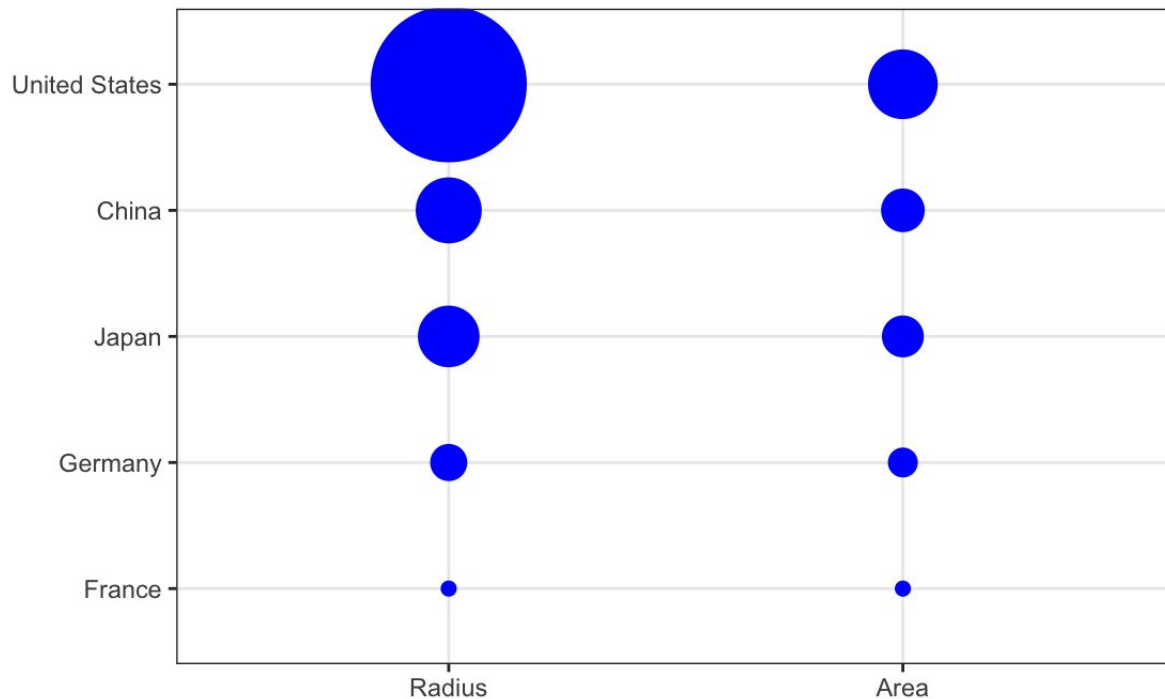


AMERICA'S ECONOMY

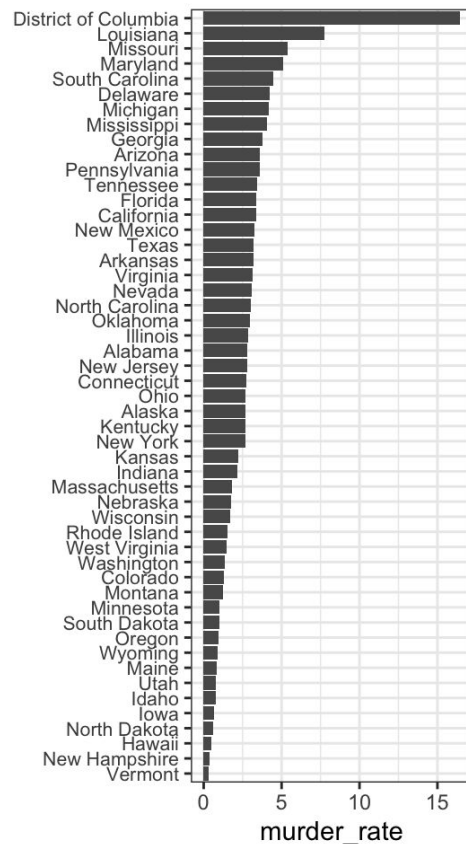
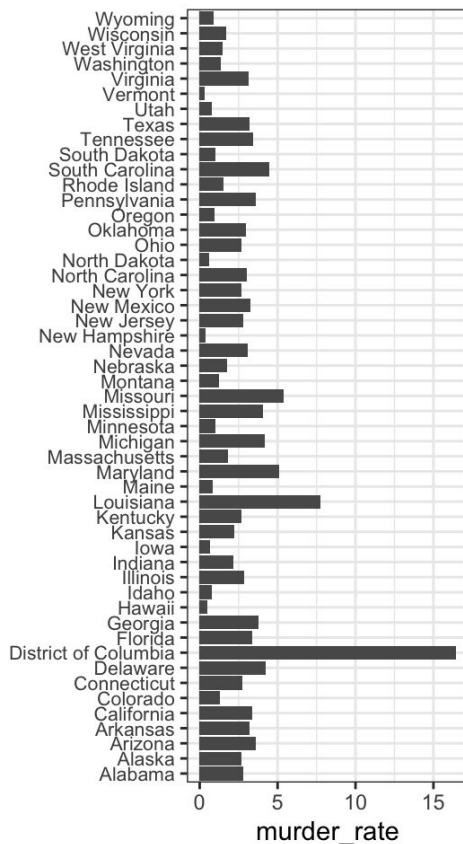
2010 GROSS DOMESTIC PRODUCT



Proporção também importa

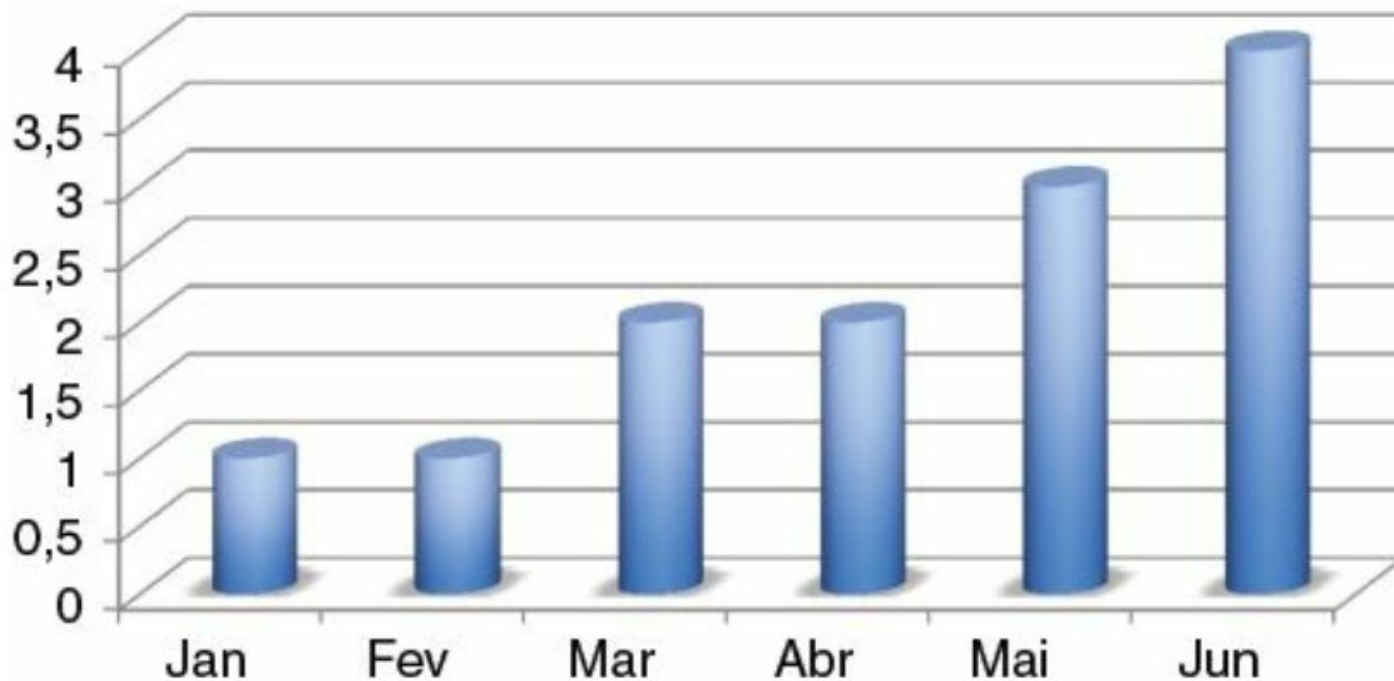


Ordem também importa!



Nunca use 3D

Numero de publicações



Evite eixo y secundário



Evite eixo y secundário

Alternativa 1: legendar diretamente



Alternativa 2: separar verticalmente



Falta de ordem visual

Demonstrando que a eficiência é a consideração mais importante ao se selecionar um provedor

De modo geral, quais atributos são os mais importantes para
você ao selecionar um provedor de serviços?
(Escolha até 3)



A pesquisa mostra que
demonstrar resultados é a dimensão
mais importante ao se escolher
um provedor de serviços.

Acessibilidade dos serviços
e experiência anterior
trabalhando em conjunto,
considerados muito importantes
no processo de tomada de decisão,
foram menos citados como
atributos importantes.

Fonte de dados: xyz; inclui número N de entrevistados.
Note que os entrevistados podiam escolher até 3 opções

Falta de ordem visual

Demonstrando que a eficiência é a consideração mais importante ao se selecionar um provedor

De modo geral, quais atributos são os mais importantes para você ao selecionar um provedor de serviços?

(Escolha até 3)



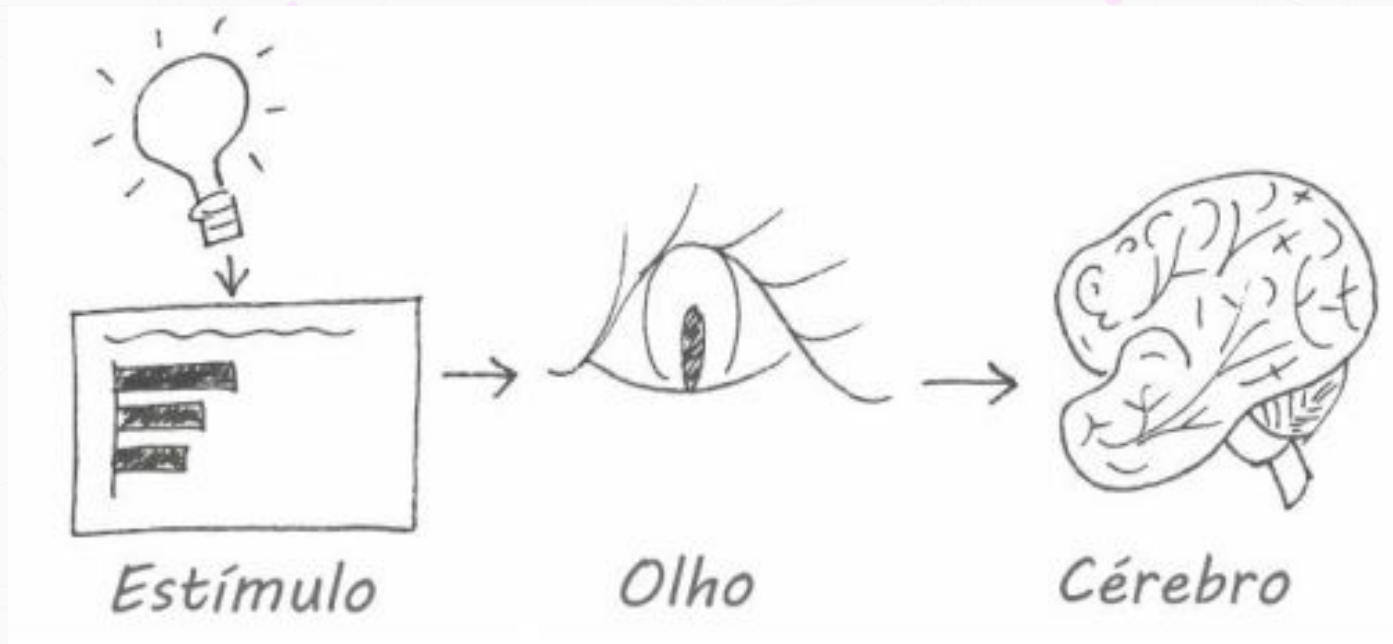
A pesquisa mostra que demonstrar resultados é a dimensão mais importante ao se escolher um provedor de serviços.

Acessibilidade do Serviço e experiência anterior trabalhando em conjunto, considerados muito importantes no processo de tomada de decisão, foram menos citados como atributos importantes.

Fonte de dados: xyz; inclui número N de entrevistados.

Note que os entrevistados podiam escolher até 3 opções

Focalize a atenção do público no que importa!



Use o contraste de forma estratégica

Análise de Desempenho

■ Nossa empresa

■ Concorrente A

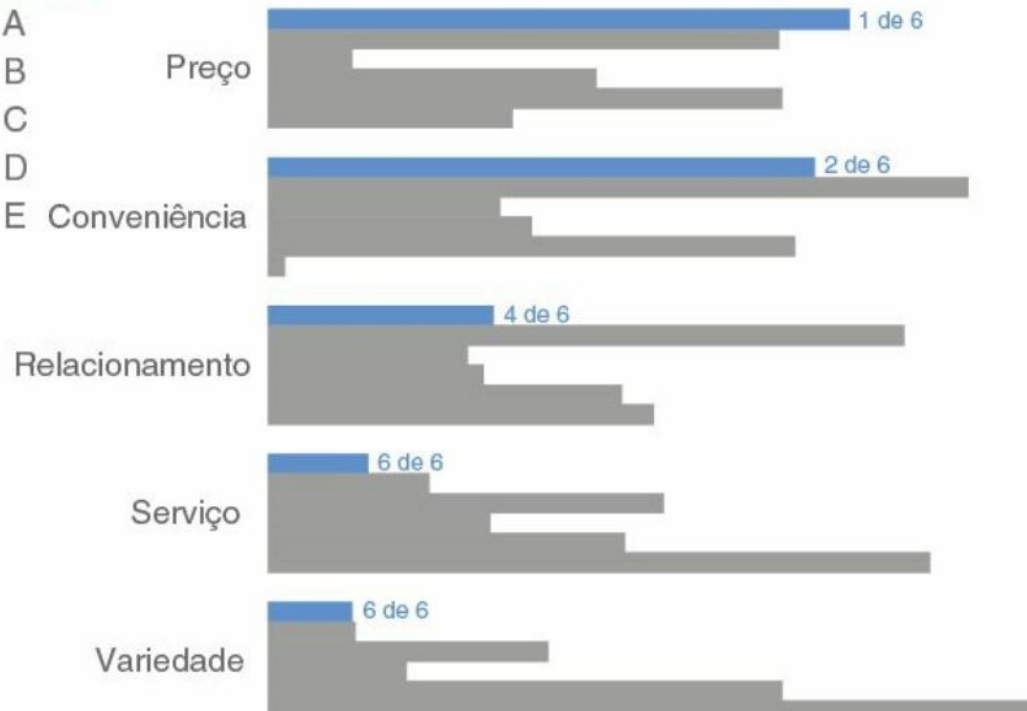
■ Concorrente B

■ Concorrente C

■ Concorrente D

■ Concorrente E

Índice de desempenho ponderado | Classificação relativa



Use cores moderadamente e em tons

Cinco Remédios Mais Vendidos por País

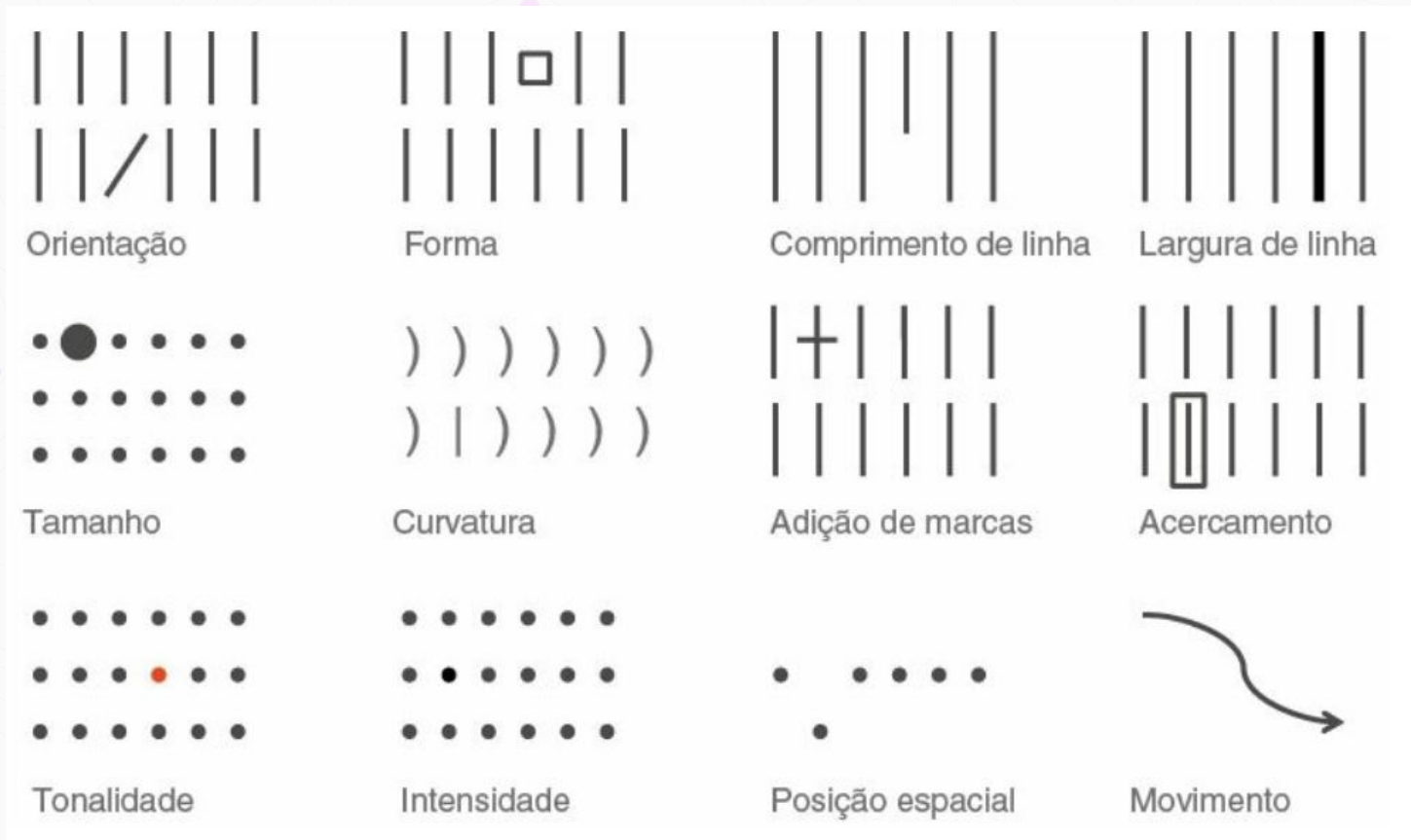
A distribuição pelas cores do arco-íris indica a classificação das vendas no país dado, de 1 (vermelho) até 10 ou mais (roxo escuro)

País	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
ALE	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
ESP	2	3	4	5	11
TUR	7	2	3	4	8
RU	1	2	3	6	7
EUA	1	2	4	3	5

5 remédios mais vendidos: classificação das vendas por país

CLASSIFICAÇÃO	1	2	3	4	5+
PAÍS REMÉDIO	A	B	C	D	E
Austrália	1	2	3	6	7
Brasil	1	3	4	5	6
Canadá	2	3	6	12	8
China	1	2	8	4	7
França	3	2	4	8	10
Alemanha	3	1	6	5	4
Índia	4	1	8	10	5
Itália	2	4	10	9	8
México	1	5	4	6	3
Rússia	4	3	7	9	12
Espanha	2	3	4	5	11
Turquia	7	2	3	4	8
Reino Unido	1	2	3	6	7
Estados Unidos	1	2	4	3	5

Use o contraste de forma estratégica



Atributos pré-atentivos em texto

O que estamos fazendo bem?

Temas e exemplos de comentários

- **Ótimos produtos:** "Esses produtos são claramente os melhores de sua classe."
- **Peças substitutas são enviadas quando necessário:**
"Vocês me enviam peças sem que eu precise pedir,
e eu preciso mesmo delas!"
- **Os problemas são resolvidos prontamente:** "O escritório de cobranças resolveu rapidamente um problema que eu tinha."
- **O serviço ao cliente supera as expectativas:**
"O gerente de contas até fez contato fora do horário comercial.
Vocês têm uma excelente empresa — continuem o bom trabalho!"

Atributos pré-atentivos em gráficos

10 principais preocupações com o design

preocupações por 1.000

A potência do motor é menor que o esperado

12.9

Os pneus fazem barulho excessivo ao dirigir

12.3

O motor faz barulho anormal/excessivo

11.6

O material do assento preocupa

11.6

Barulho excessivo do vento

11.0

Exitação ou atraso na mudança de marcha

10.3

O sistema Bluetooth tem baixa qualidade de som

10.0

O sistema de direção/volante tem muita folga

8.8

O sistema bluetooth é difícil de usar

8.6

Controles do assento dianteiro de áudio/entretenimento/navegação

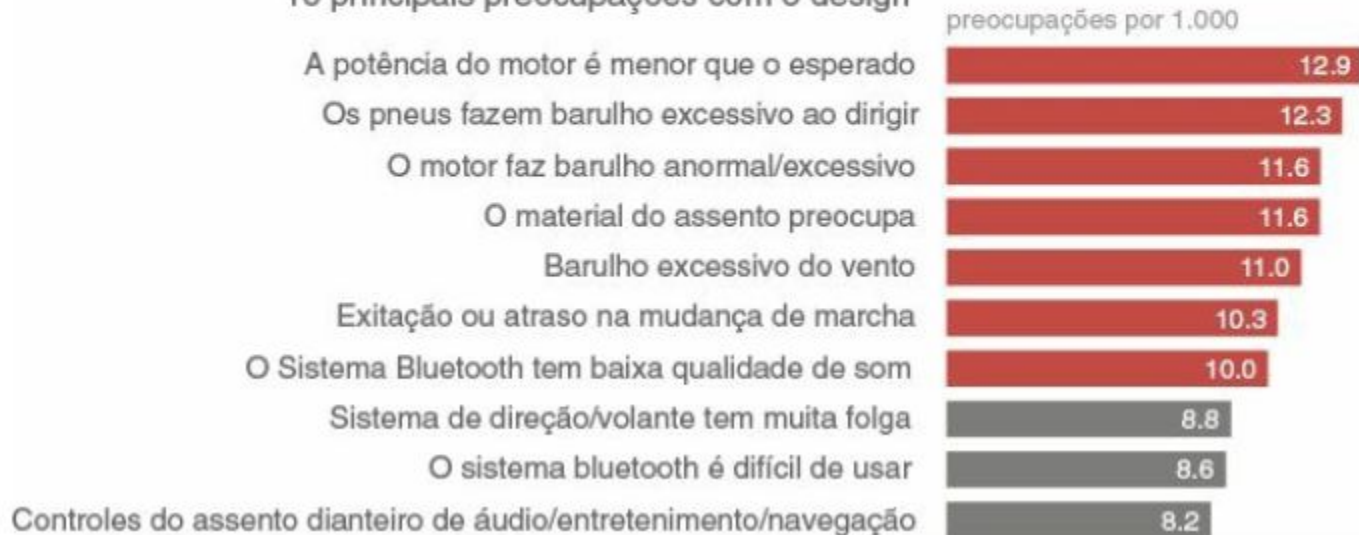
8.2

Atributos pré-atentivos em gráficos

7 das 10 principais preocupações com o design têm 10 ou mais preocupações por 1.000

Discussão: Essa é uma taxa padrão aceitável?

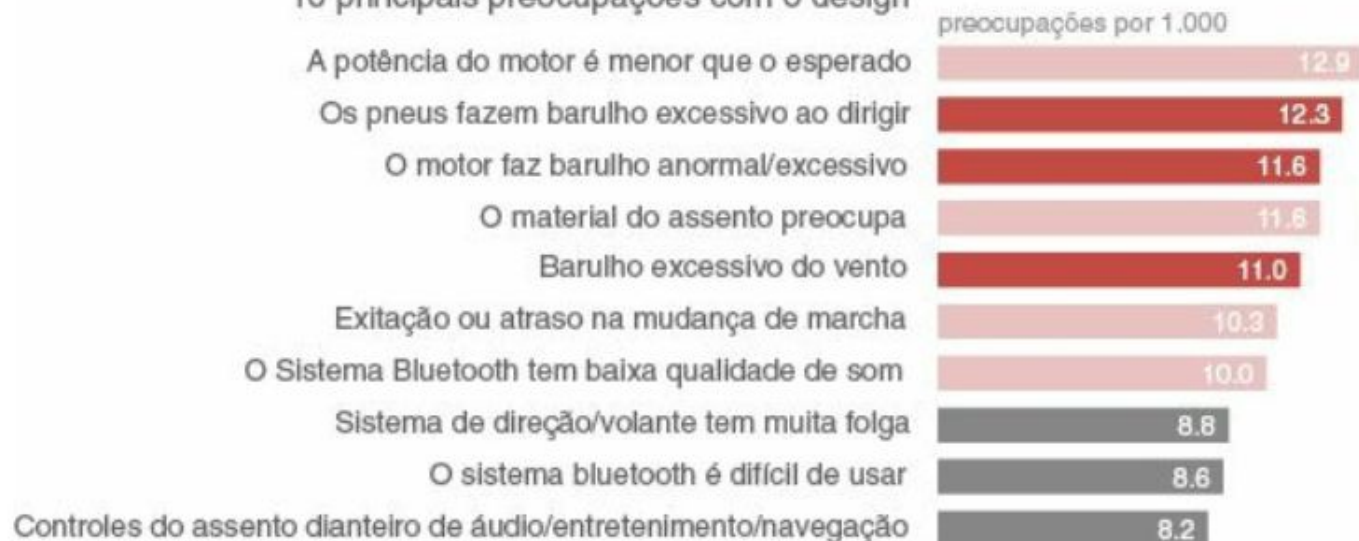
10 principais preocupações com o design



Atributos pré-atentivos em gráficos

Das 10 preocupações com o design, 7 estão relacionadas ao ruído

10 principais preocupações com o design



Comentários indicam que os problemas de **barulho dos pneus** aparecem mais **na chuva**.

Reclamações sobre **ruído do motor** são mais citadas **depois** que o carro esteve parado por algum tempo.

O **barulho do vento** é observado principalmente em autoestradas, em alta velocidade



Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Brainstorm



Bubble Chart



Bubble Map



Bullet Graph



Calendar



Candlestick Chart



Chord Diagram



Choropleth Map



Circle Packing



Connection Map



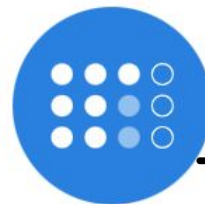
Density Plot



Donut Chart



Dot Map



Dot Matrix Chart



OBRIGADO!

Até a próxima aula!



Hora da pausa! Voltamos em:

<<20:00->>