

Inteligência Artificial e Cases de Negócio

Data Visualization

Prof. Fábio Buiati
buiati@gmail.com

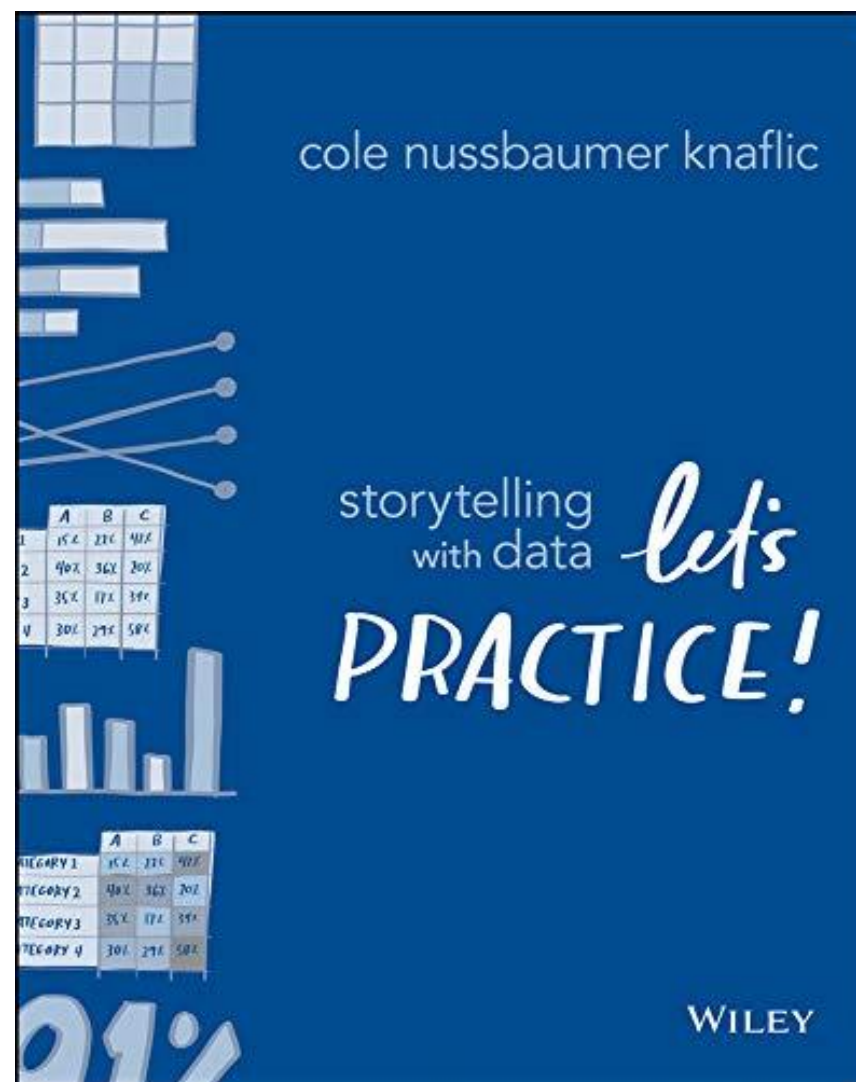


BIBLIOGRAFIA



Kindle
R\$45,43

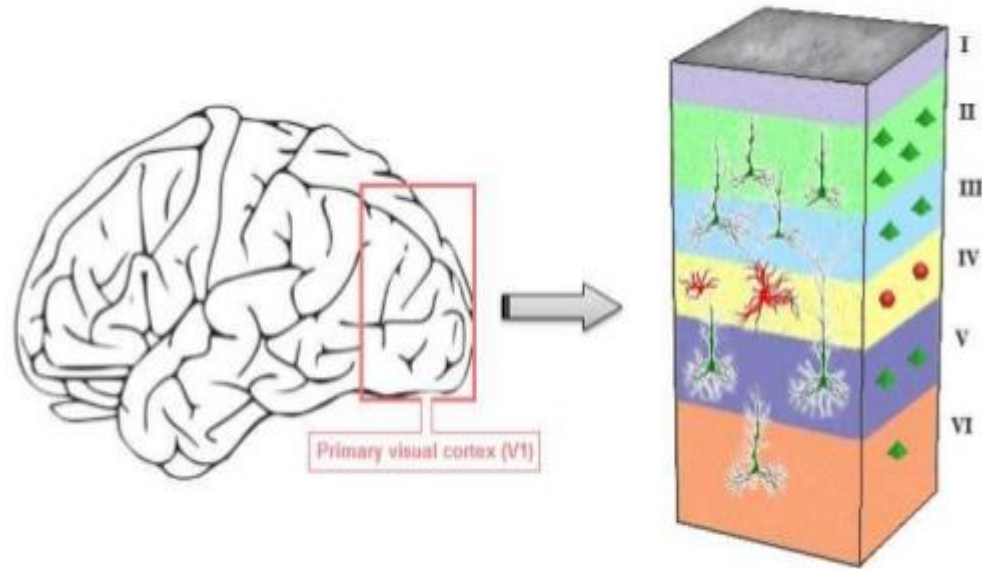
Capa Comum
a partir de R\$67,92



Kindle
R\$146,70

Capa Comum
R\$157,22

The Hardware



The Software

“Top-down”



- High-level concepts: objects, symbols
- Involves working memory
- **Slow, sequential, conscious**

Visual Perception

“Bottom-up”



- Sensory input
- Low-level features: orientation, shape, color, movement
- **Rapid, parallel, automatic**

City	State	Population
Baton Rouge	Louisiana	191,741
Birmingham	Alabama	220,927
Broken Arrow	Oklahoma	58,018
Eugene	Oregon	115,890
Glendale	Arizona	245,868
Huntsville	Alabama	55,741
Lafayette	Louisiana	87,737
Mobile	Alabama	98,147
Montgomery	Alabama	126,250
New Orleans	Louisiana	322,172
Norman	Oklahoma	101,590
Peoria	Arizona	167,868
Portland	Oregon	514,108
Salem	Oregon	147,631
Scottsdale	Arizona	134,335
Shreveport	Louisiana	68,756
Surprise	Arizona	90,548
Tempe	Arizona	143,369
Tulsa	Oklahoma	392,138



Qual a cidade **mais populosa** da lista?



Qual estado da lista tem **mais cidades**?

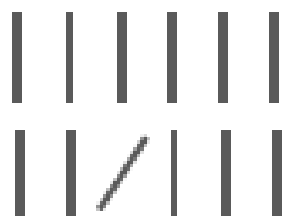


Qual a cidade **mais populosa** da lista?

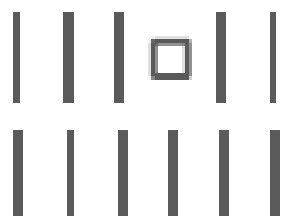


Qual estado da lista tem **mais cidades**?

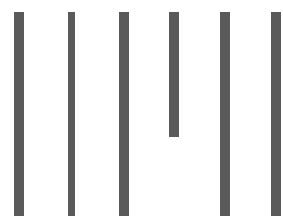
Atributos que mexem com nossa mente



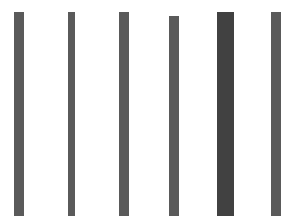
Orientation



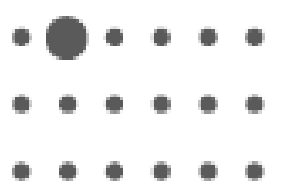
Shape



Line length



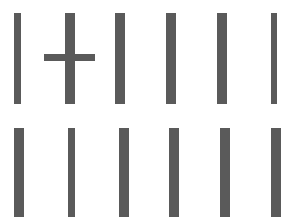
Line width



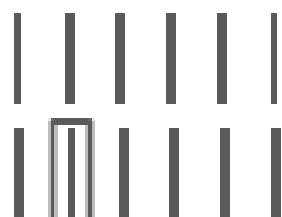
Size



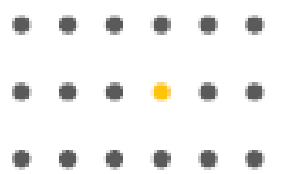
Curvature



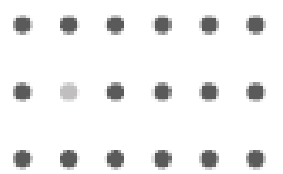
Added marks



Enclosure



Hue



Intensity



Spatial position



Motion





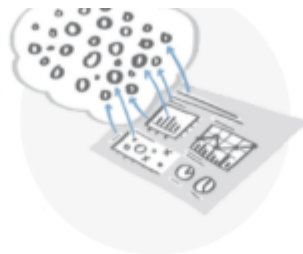
understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4



tell a
story

5

storytellingwithdata.com



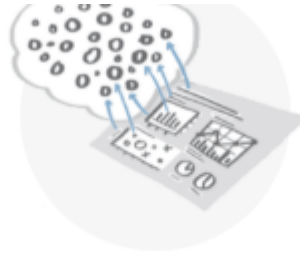
understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4



tell a
story

5

ENTENDENDO O CONTEXTO

ENTENDENDO O CONTEXTO



ANÁLISE EXPLORATÓRIA

- É a que você faz para compreender os dados
- Destacar o que pode ser interessante para outras pessoas
- Aluguel das bicicletas é um exemplo. Vários insights gerados



ANÁLISE EXPLANATÓRIA

- É o que deve ser “comunicado”
- Transformar os dados em informações e conhecimento
- Informações que podem ser consumidas facilmente
- Erro comum: apresentar todos os dados, depois de fazer toda uma análise



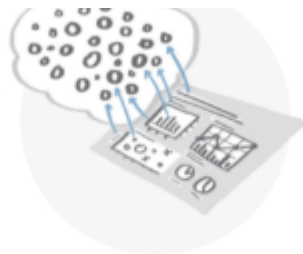
understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4



tell a
story

5

ESCOLHENDO UM VISUAL

ESCOLHENDO UM VISUAL

Você está...	Então use...		
Comparando dados	 Gráfico de barra	 Gráfico linear	 Gráfico de área circular
Mostrando a composição geral dos dados	 Gráfico de área sobreposta	 Gráfico de pizza	 Gráfico de barras sobrepostas
Mostrando a distribuição dos dados	 Histograma de barra	 Histograma linear	 Gráfico de dispersão
Mostrando a relação entre pontos de dados	 Gráfico de dispersão		



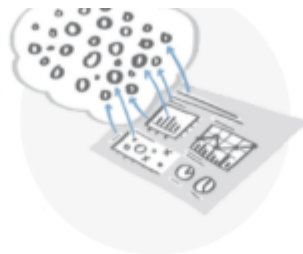
understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4



tell a
story

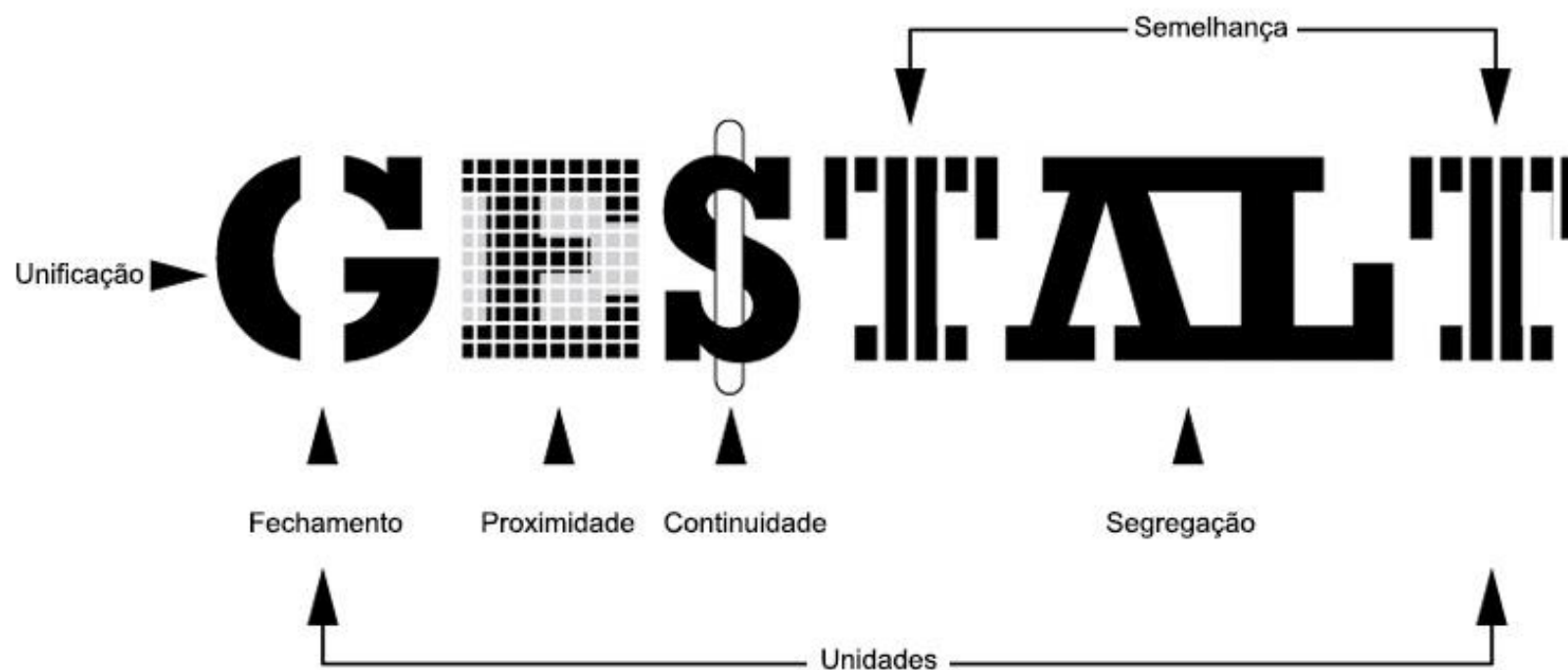
5

ELIMINANDO A SATURAÇÃO

- São elementos visuais que ocupam espaço, mas não aumentam o entendimento
- Ela faz com que os visuais pareçam mais complicados que o necessário
- **Princípio de Gestalt** ajuda a identificar elementos do visual:
 - sinal (informação que queremos passar)
 - ruído (saturação)
- Dar valor ao espaço em branco, alinhamento, eixos, legendas, etc.

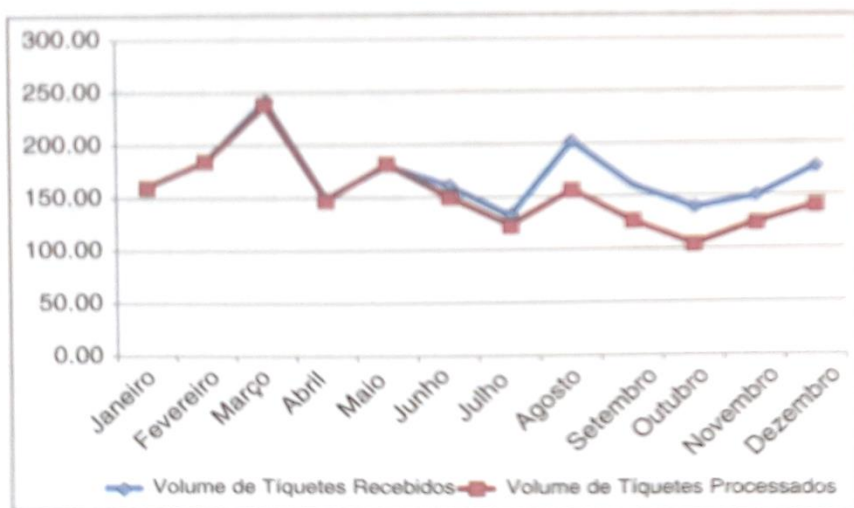
3 ELIMINANDO A SATURAÇÃO

PRINCÍPIO DE GELSTALT DA PERCEPÇÃO VISUAL



3 ELIMINANDO A SATURAÇÃO

DESSATURANDO: PASSO A PASSO



- 1 – Remova borda do gráfico
- 2 – Remova as linhas de grade
- 3 – Remova marcadores de dados
- 4 – Limpe as legendas de eixo (os 0 e abreviação dos meses do ano)
- 5 – Legenda de dados diretamente nas linhas
- 6- Potencialize as cores (legendas da mesma cor que os dados)



understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4

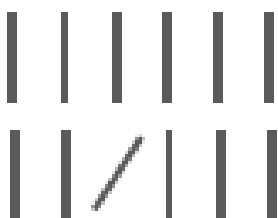


tell a
story

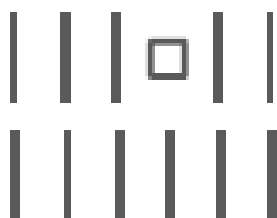
5

CHAME A ATENÇÃO

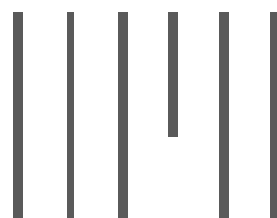
Atributos “pré-atentivos” chamam a atenção



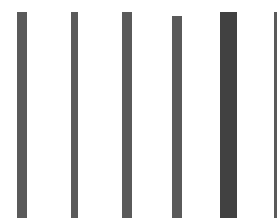
Orientation



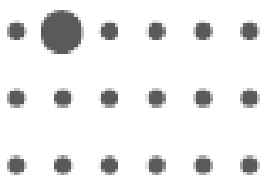
Shape



Line length



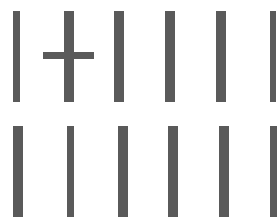
Line width



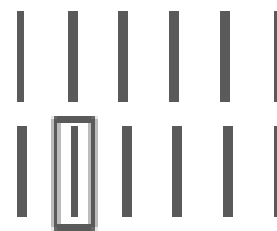
Size



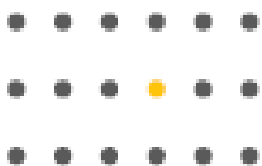
Curvature



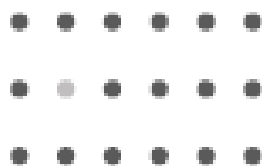
Added marks



Enclosure



Hue



Intensity



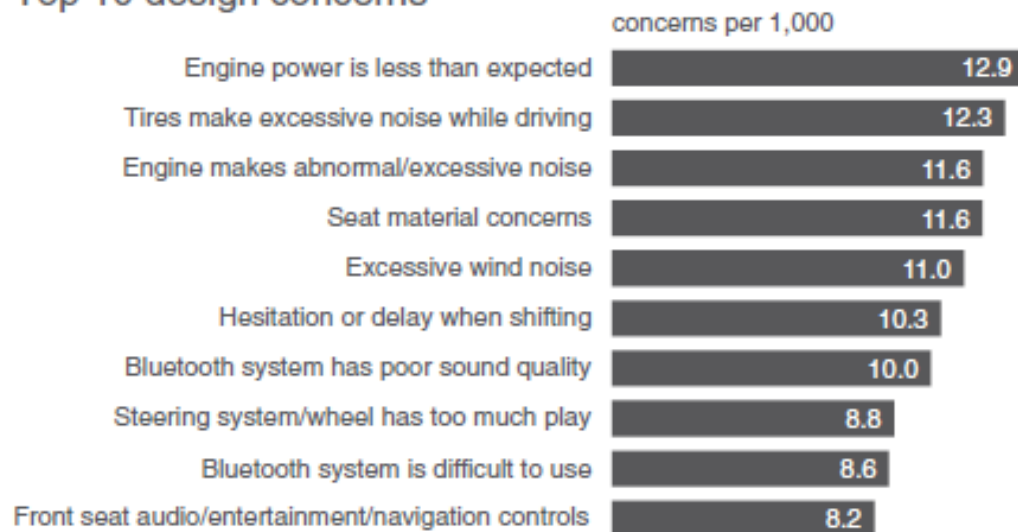
Spatial position



Motion

Atributos chamam a atenção

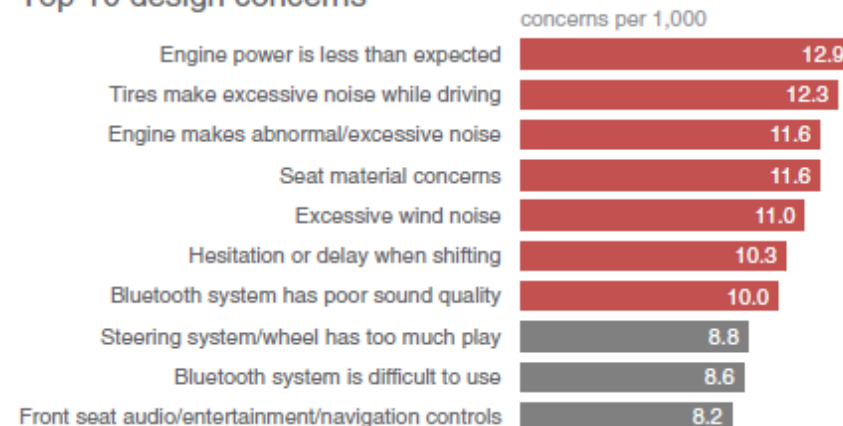
Top 10 design concerns



7 of the top 10 design concerns have 10 or more concerns per 1,000.

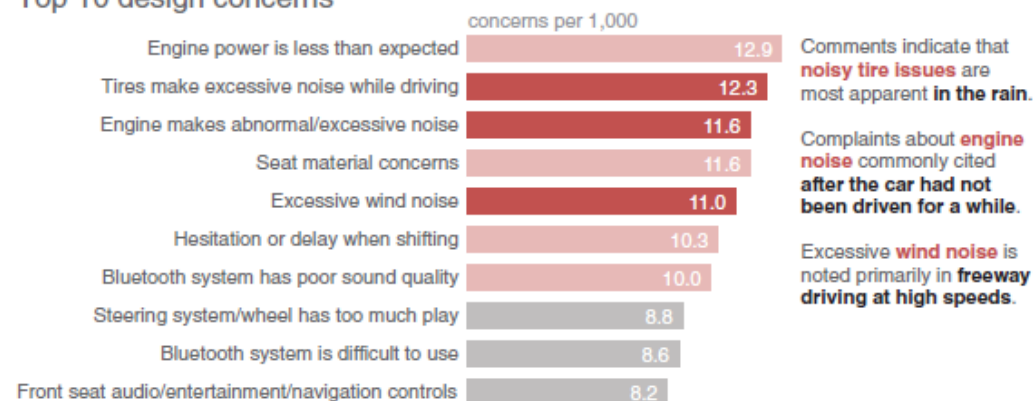
Discussion: is this an acceptable default rate?

Top 10 design concerns



Of the top design concerns, three are noise-related.

Top 10 design concerns

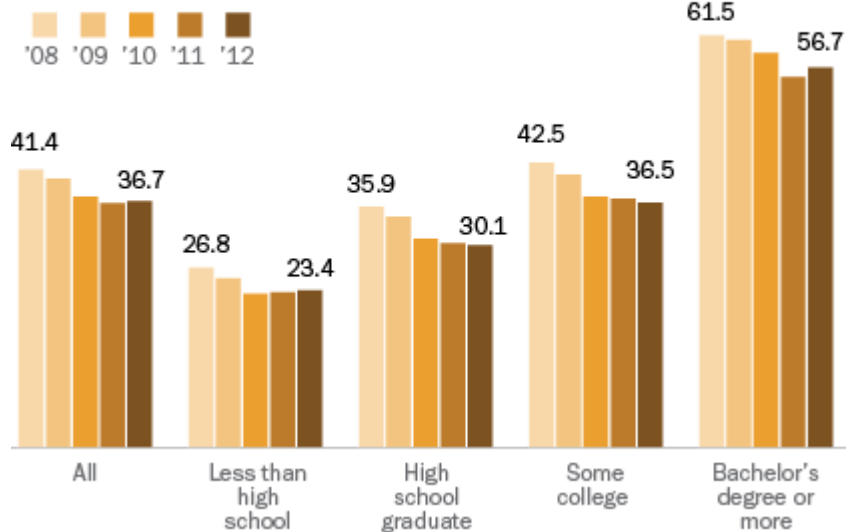


CHAME A ATENÇÃO

Dados do censo mostram que mais americanos estão se casando, especialmente os de nível superior

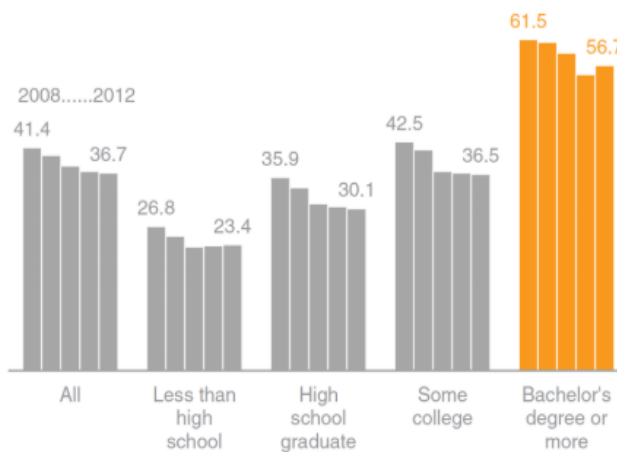
New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



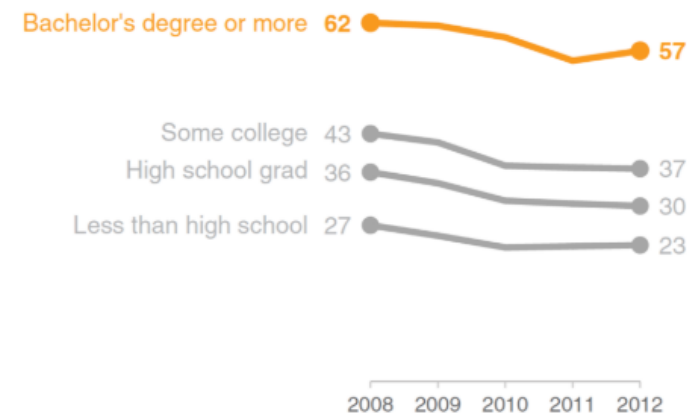
New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults





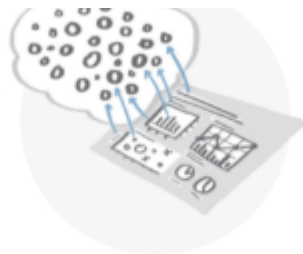
understand the
context

1



choose an
effective visual

2



eliminate
clutter

3



focus
attention

4



tell a
story

5

CONTE UMA HISTÓRIA

TEXTO SIMPLES

After the pilot program,

68%

of kids expressed interest towards science,
compared to 44% going into the program.

Quando você tem 1 ou 2 números para compartilhar, um simples texto pode ser uma ótima maneira de comunicar

EXEMPLOS DE STORYTELLING

SEM STORYTELLING

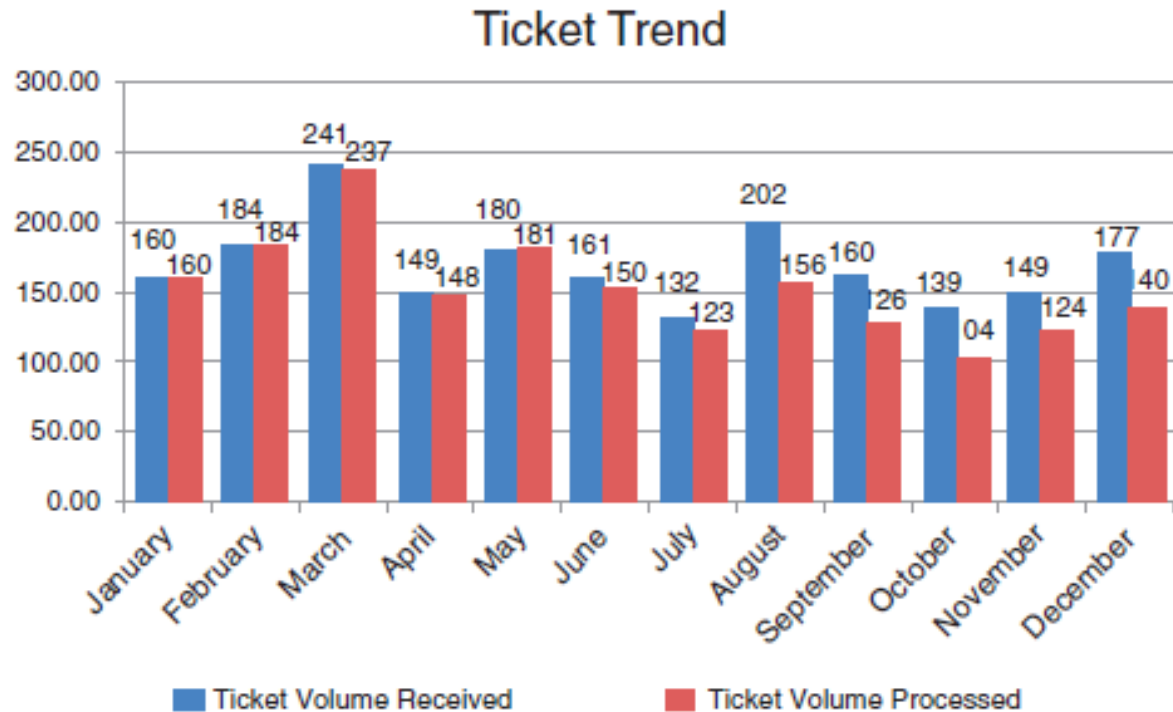


FIGURE 0.2 Example 1 (before): showing data

COM STORYTELLING

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

FIGURE 0.3 Example 1 (after): storytelling with data

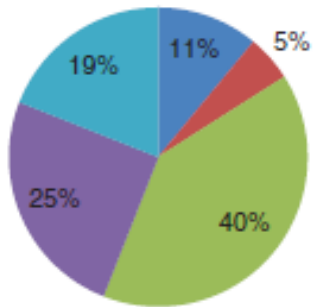
EXEMPLOS DE STORYTELLING

SEM STORYTELLING

Survey Results

PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

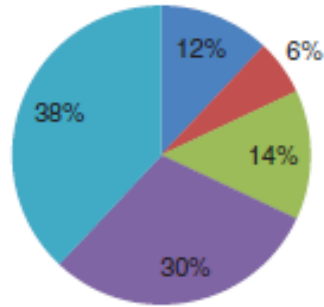


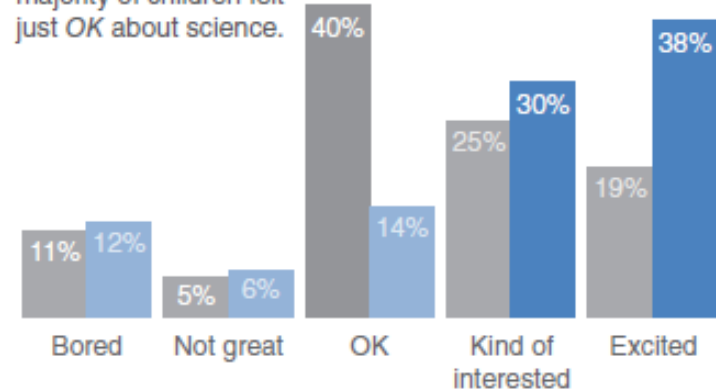
FIGURE 0.4 Example 2 (before): showing data

COM STORYTELLING

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just OK about science.



AFTER program, more children were *Kind of interested* & *Excited* about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 0.5 Example 2 (after): storytelling with data

EXEMPLOS DE STORYTELLING

SEM STORYTELLING

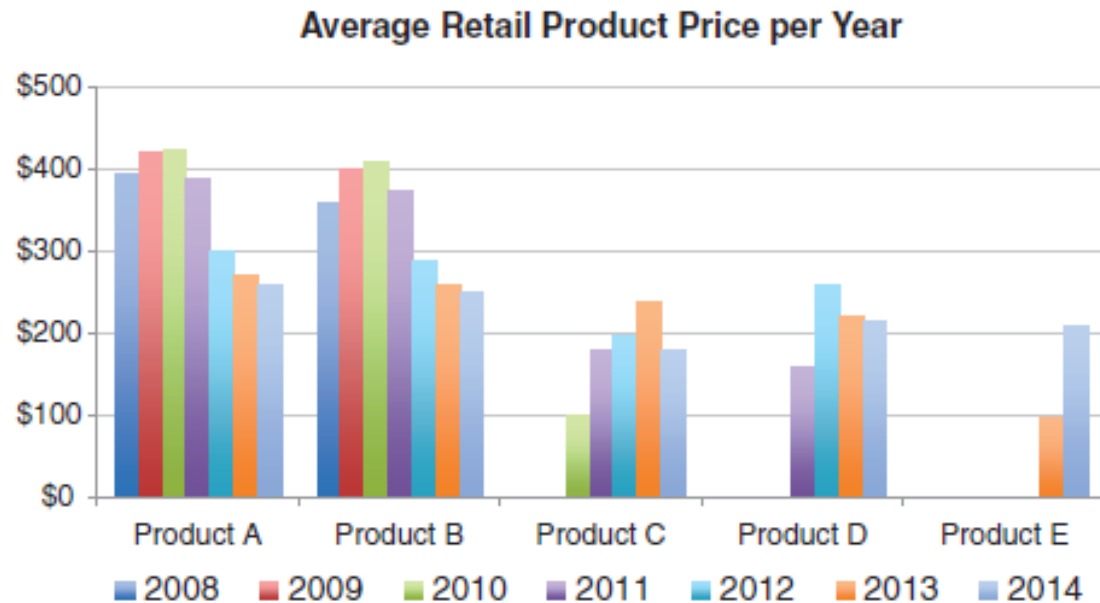


FIGURE 0.6 Example 3 (before): showing data

COM STORYTELLING

To be competitive, we recommend introducing our product *below the \$223 average price point in the \$150–\$200 range*

Retail price over time by product

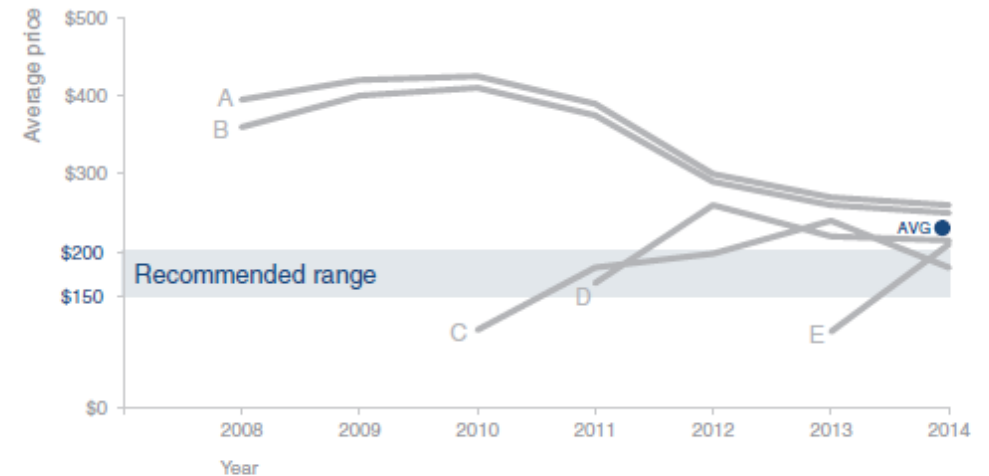
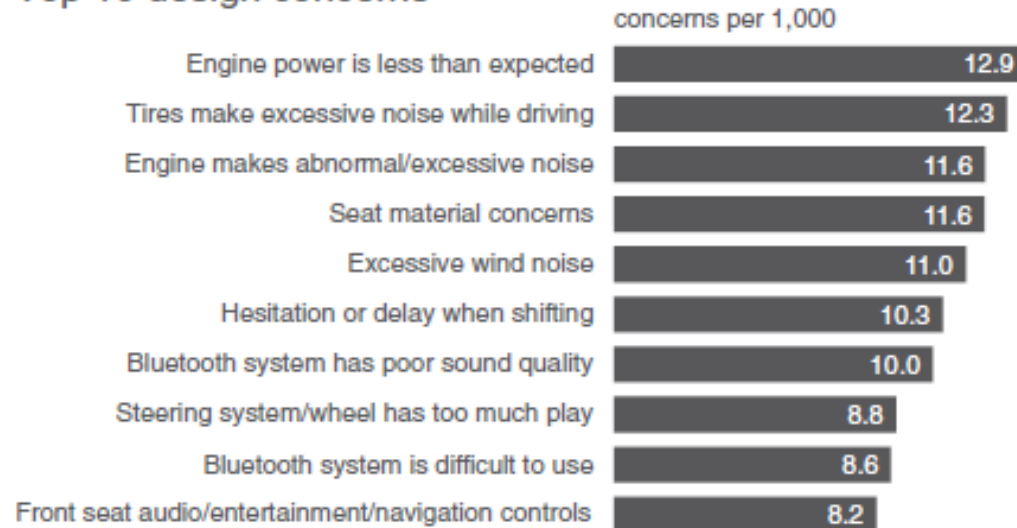


FIGURE 0.7 Example 3 (after): storytelling with data

EXEMPLOS DE STORYTELLING

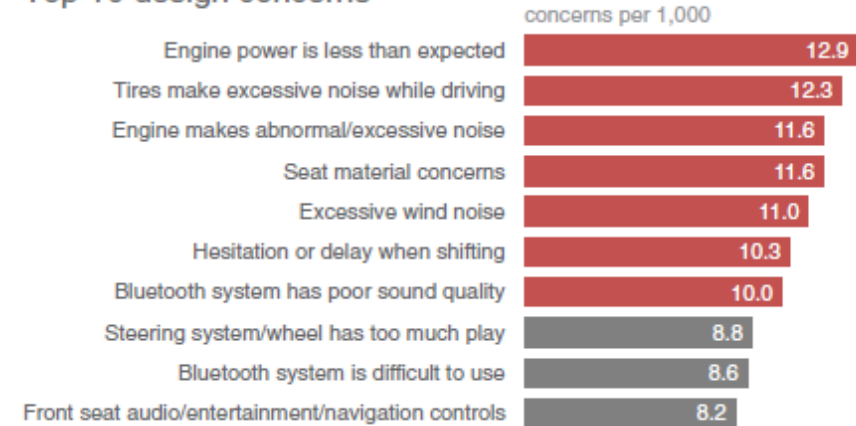
Top 10 design concerns



7 of the top 10 design concerns have 10 or more concerns per 1,000.

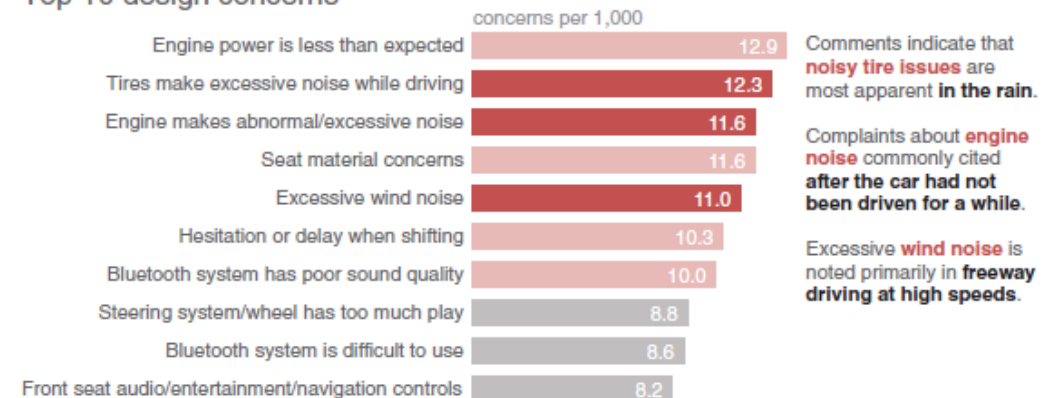
Discussion: is this an acceptable default rate?

Top 10 design concerns

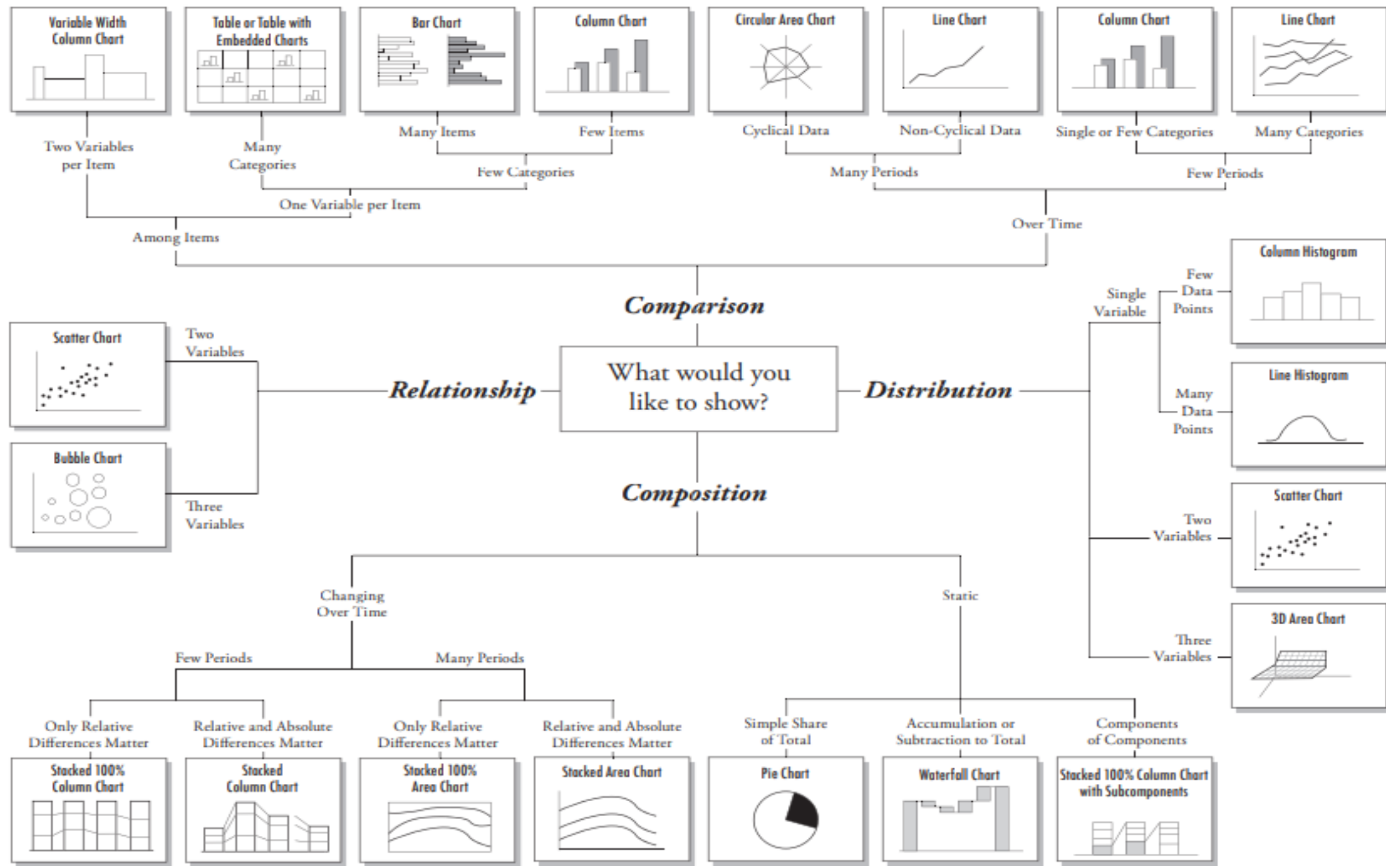


Of the top design concerns, three are noise-related.

Top 10 design concerns







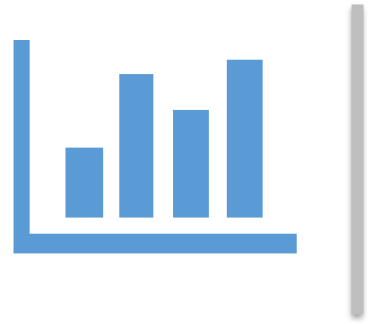
Caso Prático



Inteligência Artificial e Cases de Negócio

**MATERIAL ADICIONAL
DATA VISUALIZATION**

Prof. Fábio Buiati
buiati@gmail.com



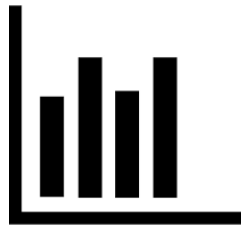
TOP 9 GRÁFICOS BÁSICOS



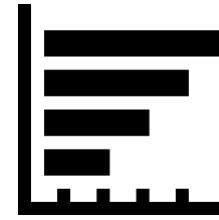
TOP 9 GRÁFICOS BÁSICOS



Linha



Colunas



Barras horizontal



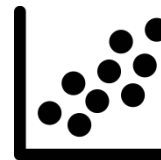
Bolhas



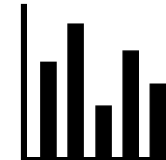
Pizza



Área



Dispersão

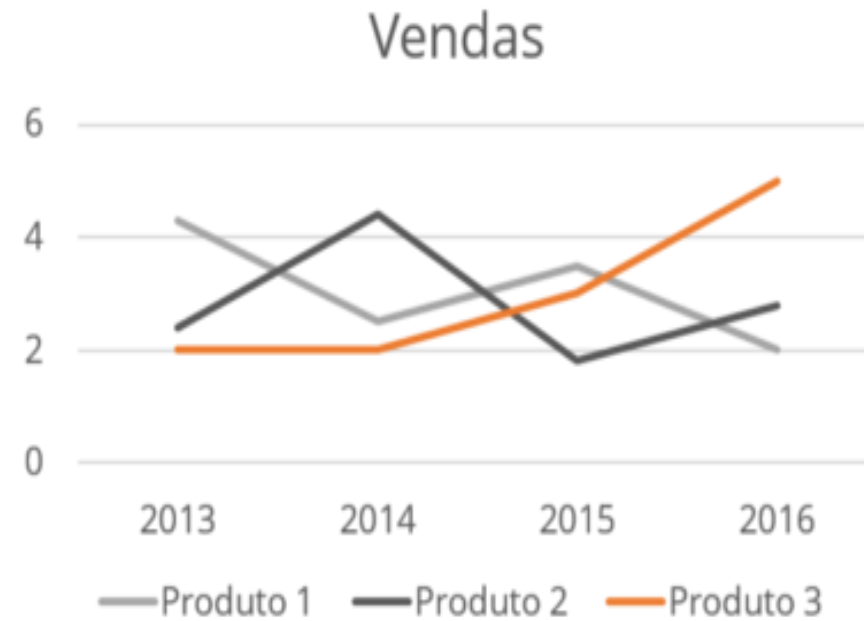
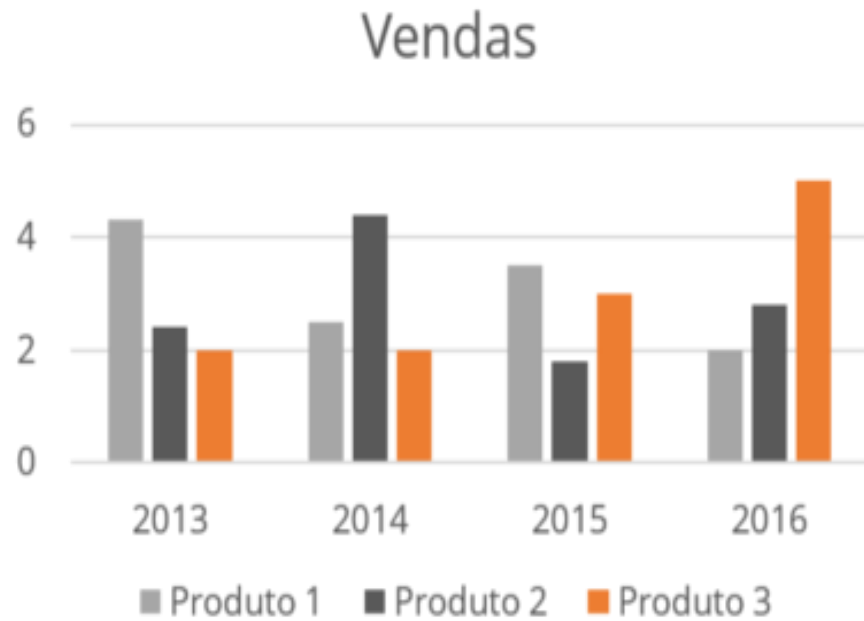


Histograma



Radar

GRÁFICO DE LINHA



Legibilidade: fácil

Função: comparação. Dados em muitos períodos e não cíclicos ou dados em poucos períodos e muitas categorias

Exemplos: Visitas, vendas.

GRÁFICO DE COLUNAS



Legibilidade: fácil

Função: comparação.

Exemplos: frequência relativa, contagem, percentuais.

Atenção: em alguns casos, o gráfico de barras horizontais pode ser mais fácil de ler do que a versão de barras verticais (colunas)

GRÁFICO DE BARRAS HORIZONTAIS



Legibilidade: fácil

Função: comparação.

Exemplos: visitas, vendas, investimento, etc.

GRÁFICO DE PIZZA



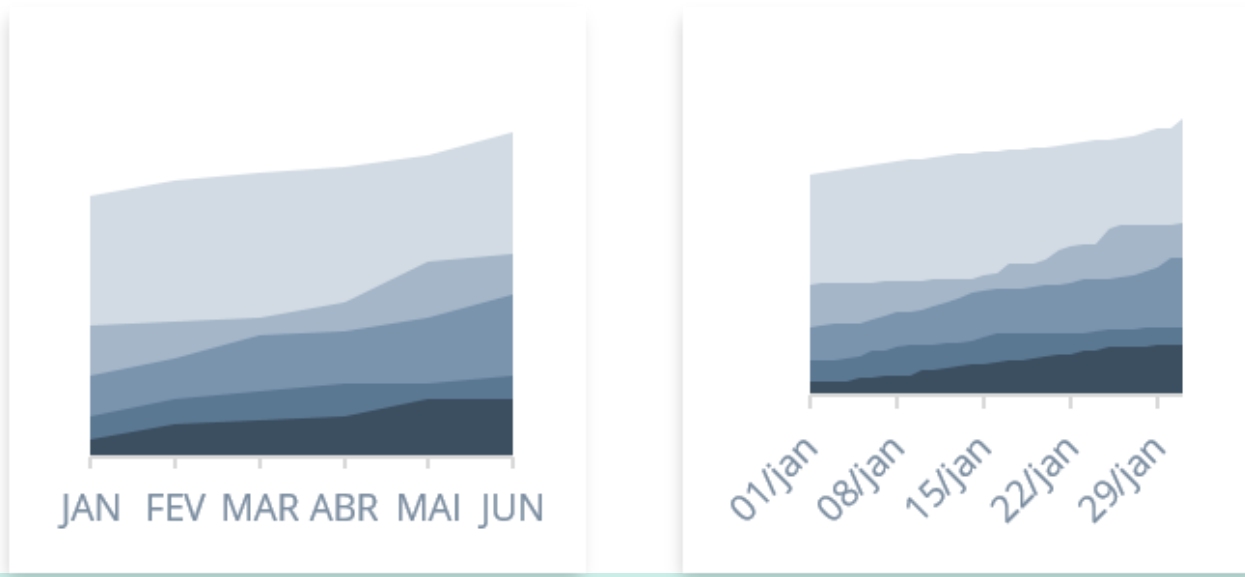
Legibilidade: fácil

Função: composição

Exemplos: visitantes novos, % de homens/mulheres, etc.

Atenção: Nunca use gráfico de pizza se a soma das fatias não for igual a 100% dos dados. Também não é recomendável usar gráfico de pizza com mais de seis fatias e nem com efeito tridimensional (3D).

GRÁFICO DE ÁREA



Legibilidade: média

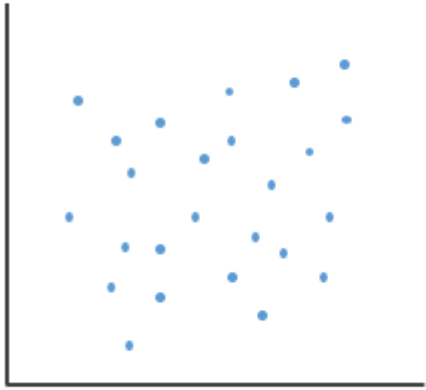
Função: composição

Exemplos: visitas, vendas, receitas.

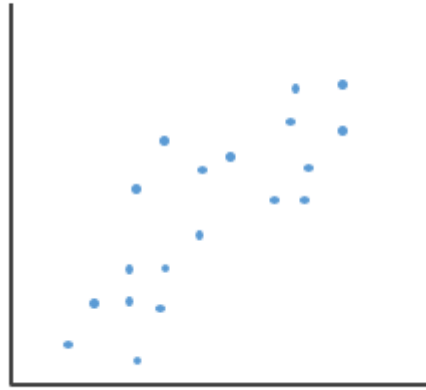
Atenção: áreas empilhadas podem distorcer a visão de cada série, que se apoia na base variável da série que está embaixo, o que atrapalha para identificar padrões nas séries individuais.

GRÁFICO DE DISPERSÃO

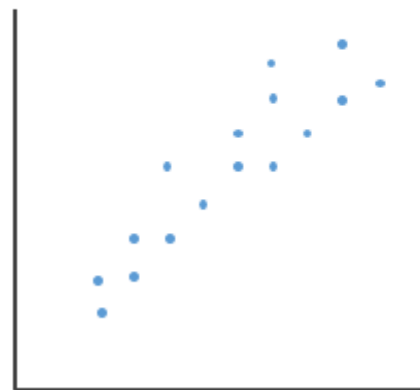
Nenhuma correlação



Possível correlação positiva



Correlação positiva



Correlação negativa



Legibilidade: média

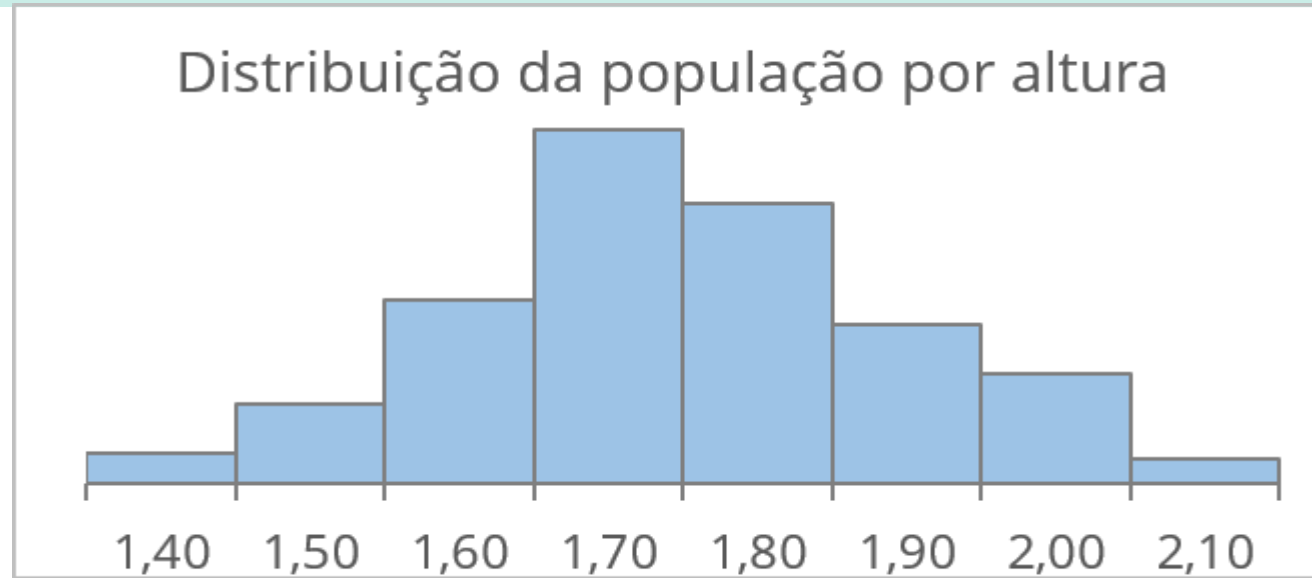
Função: distribuição de duas variáveis ou interdependência entre duas variáveis

Exemplos: investimento em mídia e vendas.

Atenção: o gráfico de dispersão serve para observar o que acontece com uma variável quando outra variável se altera, podendo identificar possível relação de causa e efeito.

No entanto, a correlação nem sempre indica relação direta de causalidade.

GRÁFICO DE HISTOGRAMA



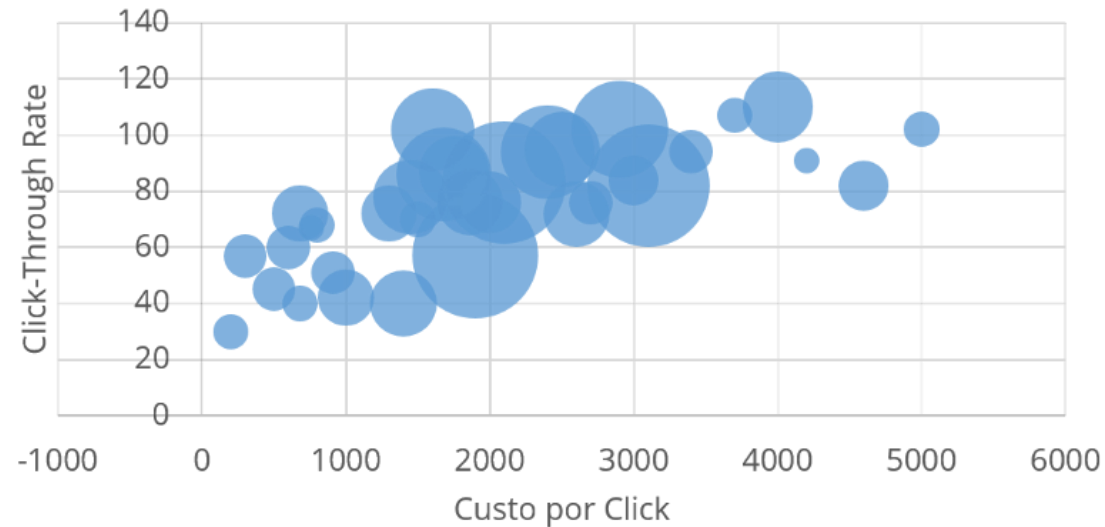
Legibilidade: média

Função: distribuição de frequências ou densidade de probabilidades, com variável única

Exemplos: distribuição de renda, demografia.

Atenção: Os dados são simétricos? Existem dados que devem ser desconsiderados por estarem distante dos demais dentro do conjunto? Como os dados estão dispersos?

GRÁFICO DE BOLHA



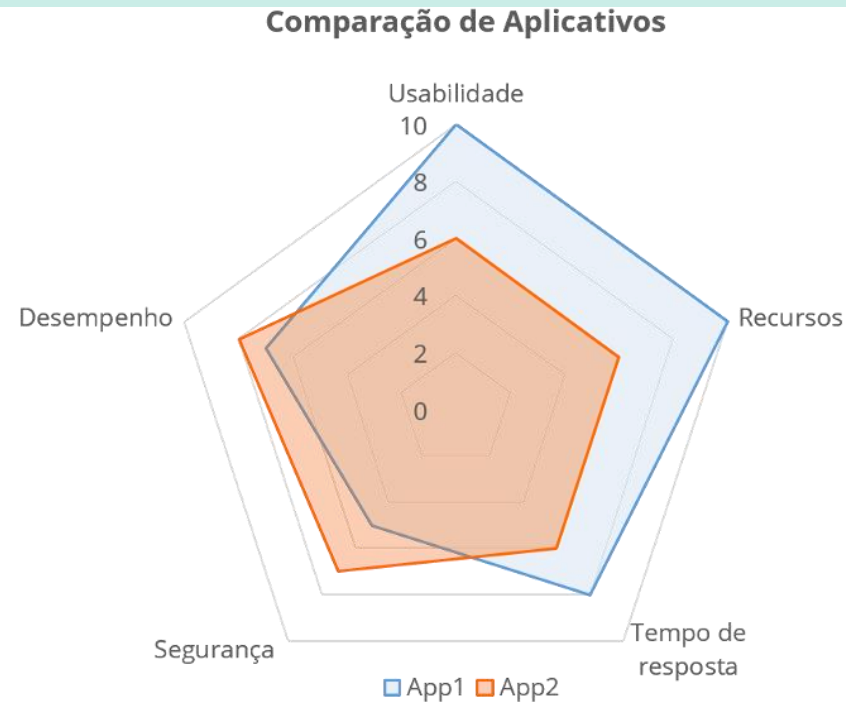
Legibilidade: difícil

Função: relação – interdependência entre variáveis

Exemplos: contagem, frequência e proporção; CPC, clicks e conversões.

Atenção: apesar de o gráfico de bolhas ser ótimo para uma comparação rápida das variáveis, não é muito adequado para análises de precisão. Além disso, alguns softwares usam o diâmetro e não a área (volume) para a variável que altera o tamanho da bolha, o que torna mais difícil a comparação.

GRÁFICO DE RADAR

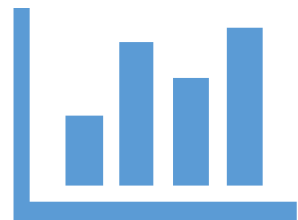


Legibilidade: média

Função: comparação

Exemplos: comparação entre características de produtos, cidades, etc.

Atenção: a ordem dos eixos altera a figura e pode influenciar a interpretação do gráfico



TOP 10 GRÁFICOS AVANÇADOS



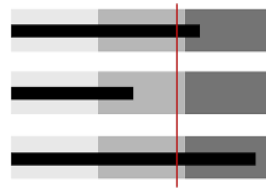
TOP 10 GRÁFICOS AVANÇADOS



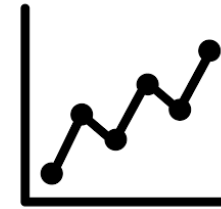
Geográfico



Gauge



Bullet



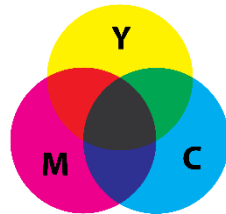
Sazonalidade



Sankey



SunBurst



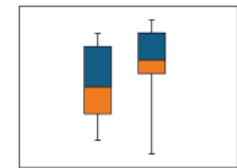
Venn



Funil

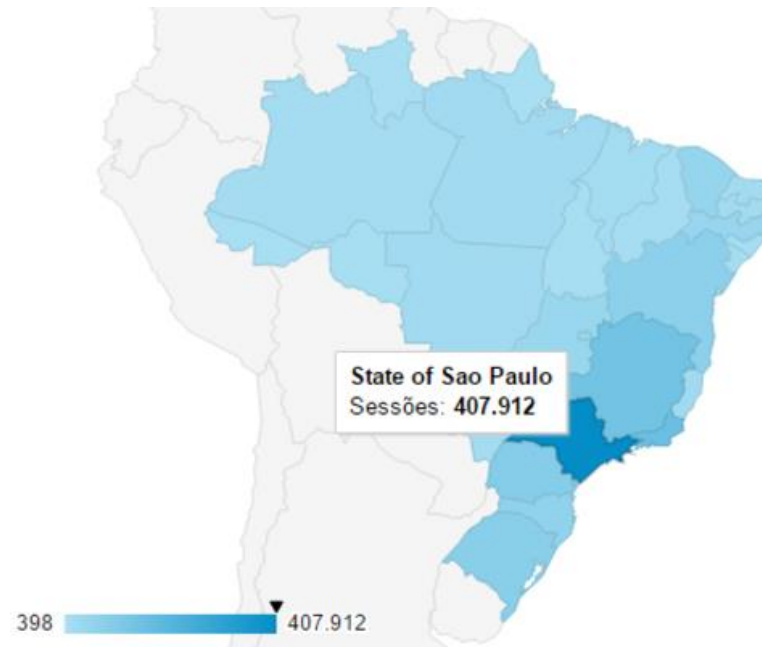


Treemap



BoxPlot

MAPA GEOGRÁFICO



Legibilidade: fácil

Função: distribuição.

Exemplos: sessões por cidade, vendas por estado, receita por país.

Os gráficos regionais podem usar uma escala de cor gradual, como no exemplo acima, ou com diferentes cores para indicar cada segmento dos dados.

GAUGE

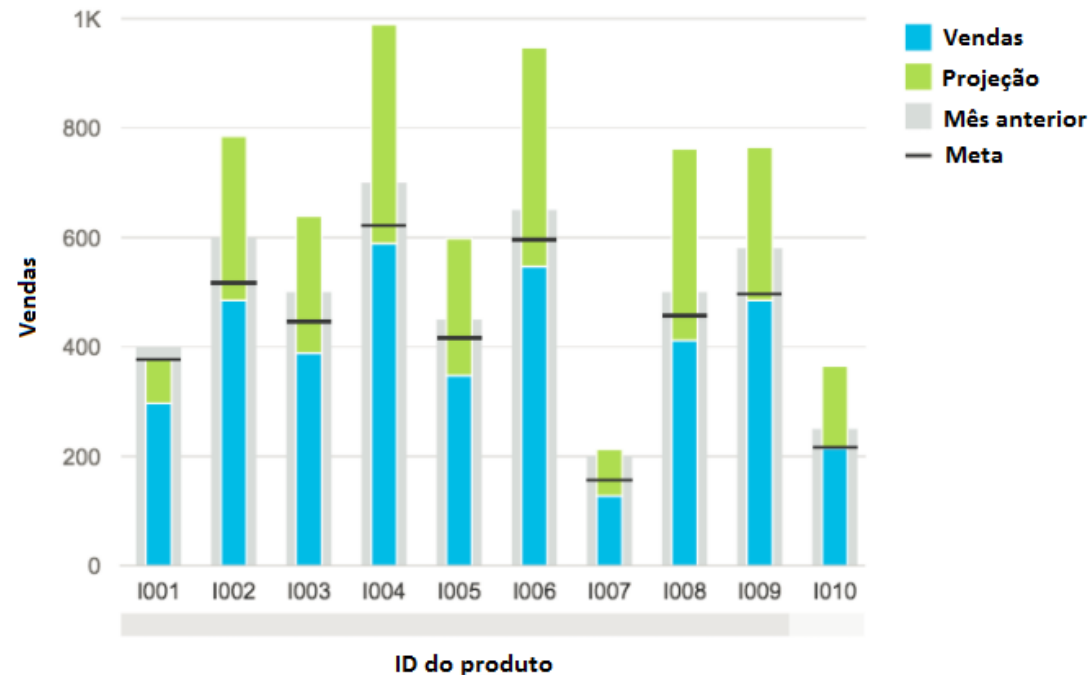


Legibilidade: fácil

Função: comparação.

Exemplos: acompanhamento de metas, alerta para instabilidades.

BULLET



Legibilidade: média

Função: comparação.

Exemplos: meta de leads ou de vendas.

Atenção: O gráfico bullet não é tão comum quanto o gauge. Visualmente, mostra o cumprimento das metas de forma mais precisa que o gauge.

SAZONALIDADE



Legibilidade: média

Função: distribuição

Exemplos: evolução das vendas

Atenção: Na prática, os gráficos de sazonalidade podem ser feitos como um gráfico de colunas agrupadas. O exemplo acima agrupa por mês os dados de 2015, 2016 e a projeção de 2017, além de destacar com cores diferentes os meses mais intensos e menos intensos.

SUNBURST



Legibilidade: média

Função: composição

Exemplos: vendas por subcategorias, caminhos de navegação

Descrição: os círculos internos representam os níveis primários, que se desdobram nos círculos externos conforme subcategorias ou dados subsequentes (como na navegação em landing pages e nas páginas seguintes de um site).

SUNBURST

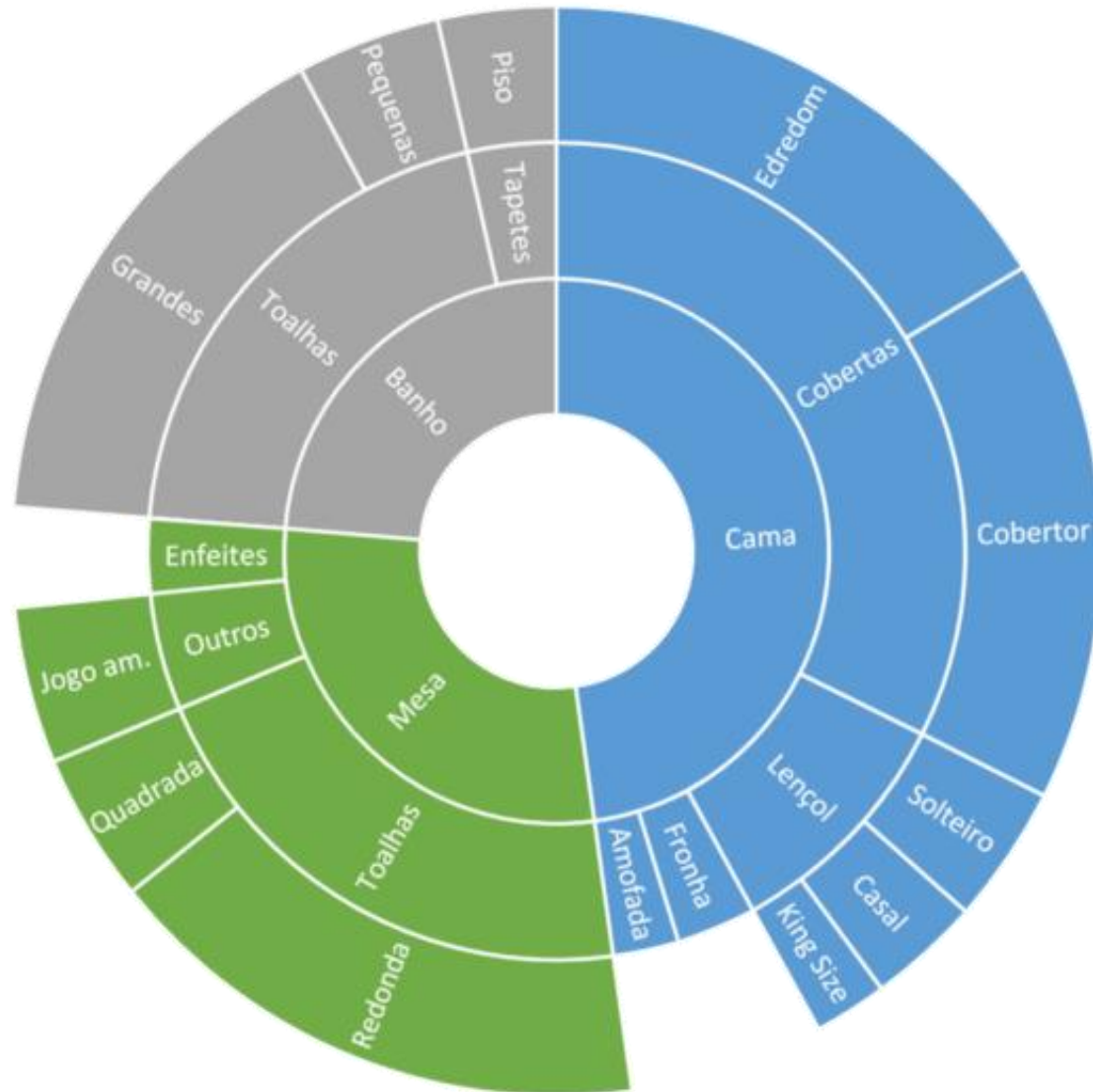
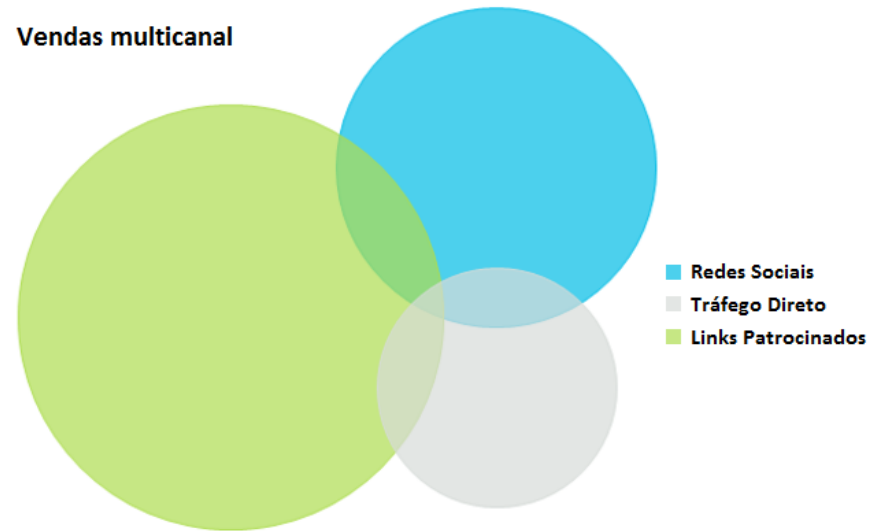


DIAGRAMA DE VENN



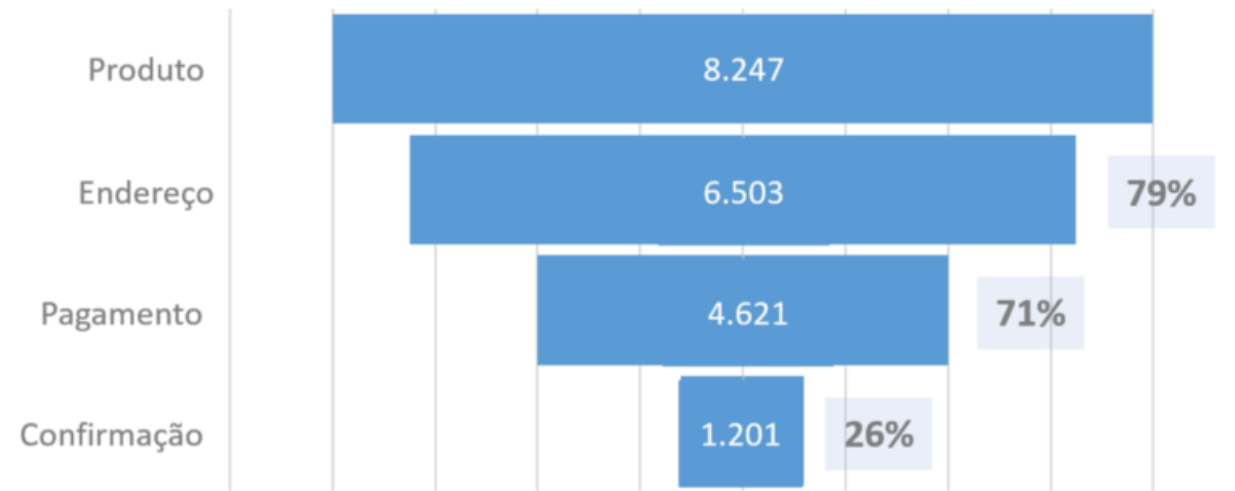
Legibilidade: média

Função: composição

Exemplos: participação de cada canal nas vendas

Atenção: é usado na matemática para mostrar a composição de elementos de cada conjunto. Em uma perspectiva multicanal, o diagrama de Venn exhibe a sobreposição dos diferentes canais de mídia, mostrando que algumas vendas podem ter influência de mais de um canal.

FUNIL



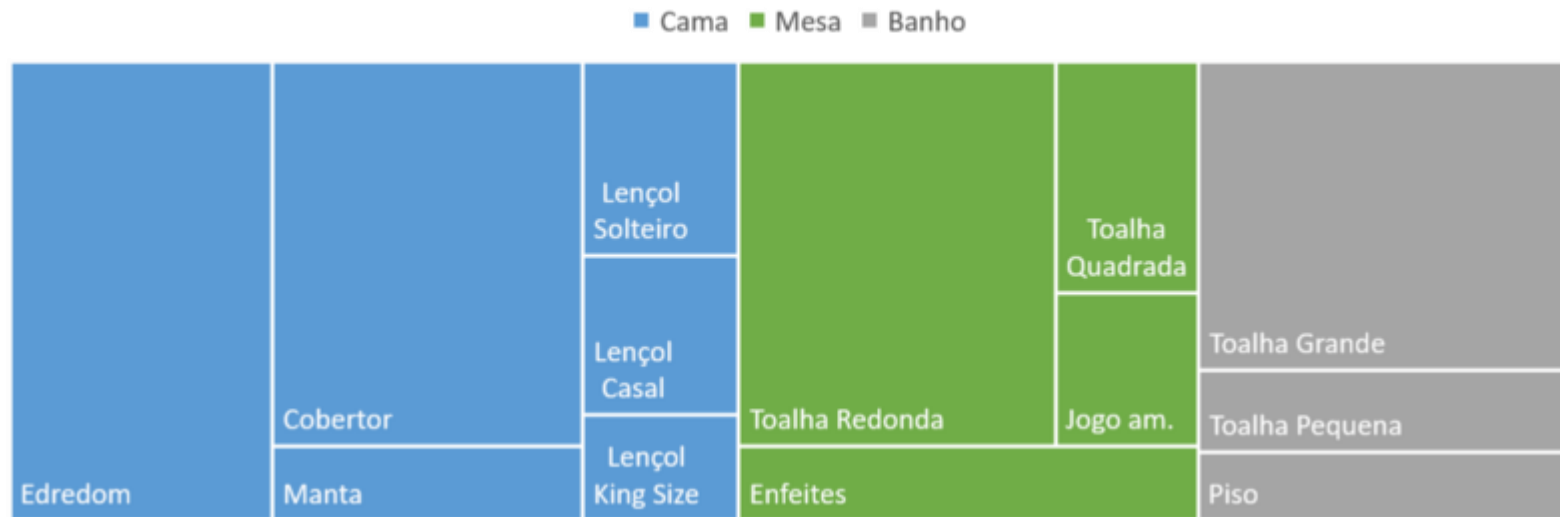
Legibilidade: média

Função: composição

Exemplos: formulários com várias etapas, fases de um carrinho de compras

Como ler: é usado para mostrar que, em processos de várias etapas, sempre há uma perda, mostrando quais etapas tiveram maior perda (de usuários, de receita, entre outras métricas), para que se possa investigar problemas e oportunidades de melhoria.

TREEMAP



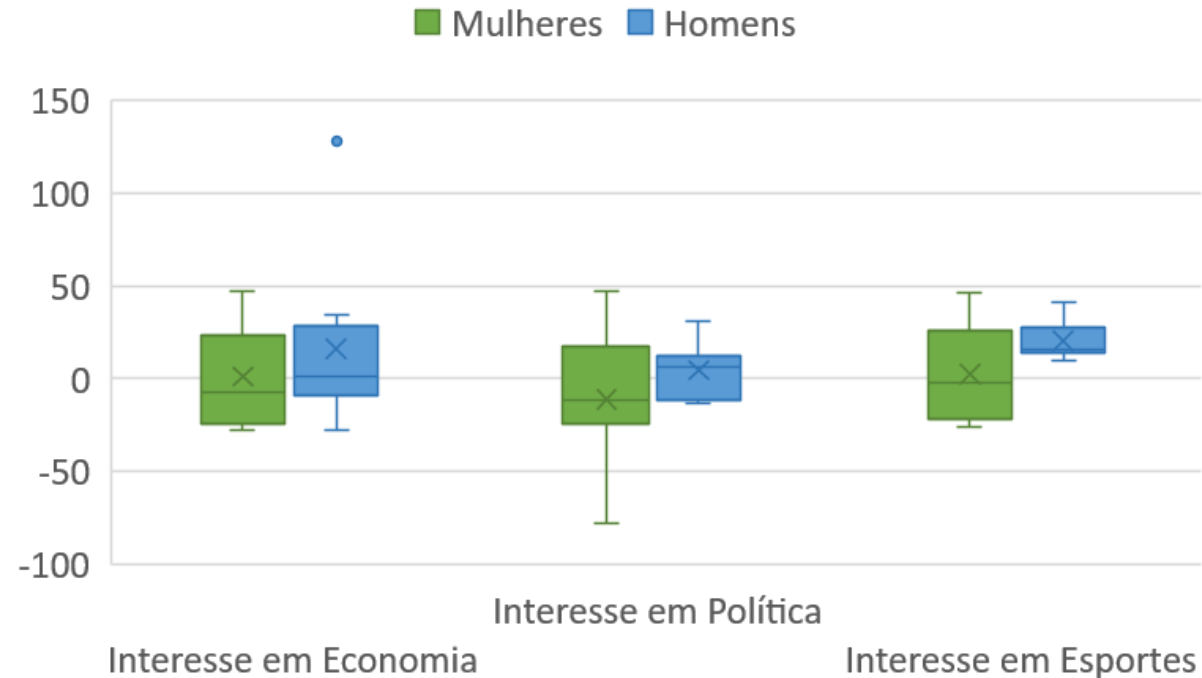
Legibilidade: difícil

Função: composição

Exemplos: contagem de valores por categorias

Como ler: é a representação visual de uma árvore de dados. Assim como em uma árvore, há um único tronco, que se desdobra em vários galhos, e cada galho pode ter zero ou mais folhas.

BOXPLOT



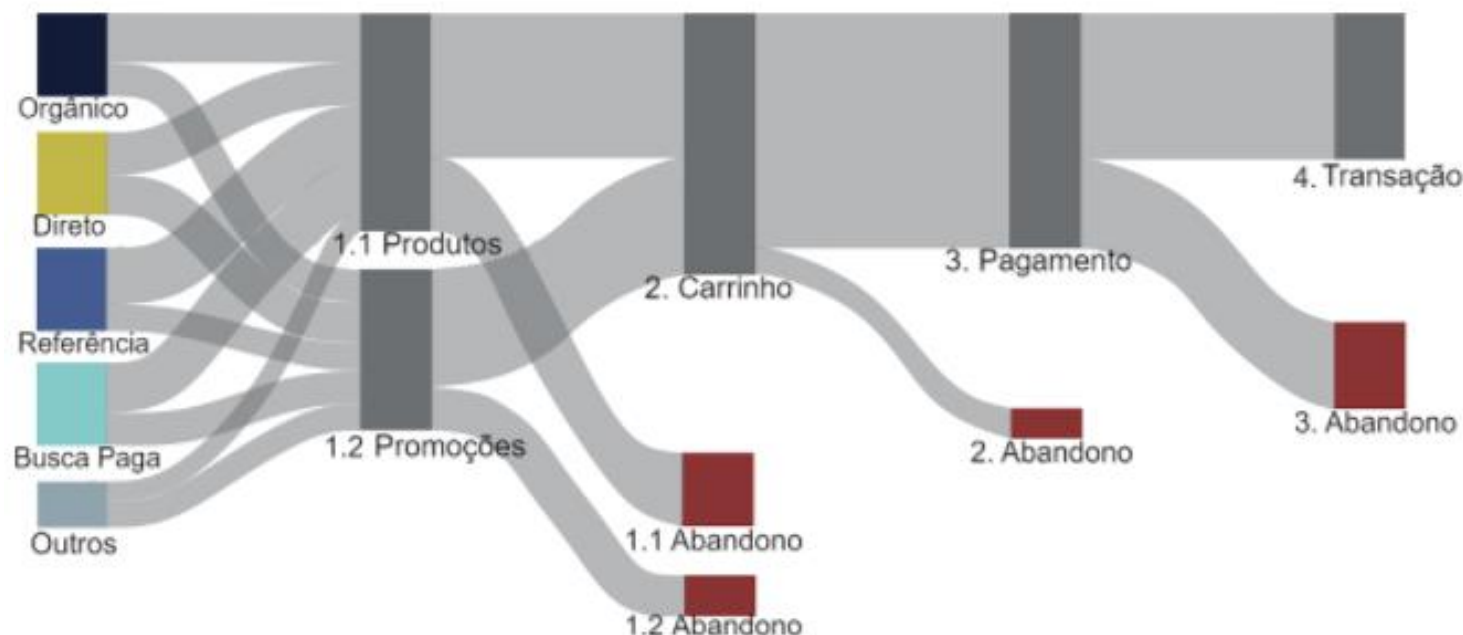
Legibilidade: péssima (para quem não é da área)

Função: distribuição

Exemplos: análises demográficas de grupos de interesse em determinados tópicos ou produtos.

Atenção: Para entender um gráfico de vela é importante entender os conceitos estatísticos de mediana e quartis.

SANKEY

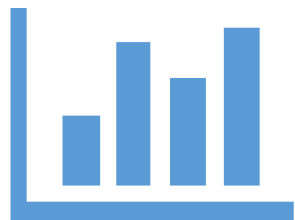


Legibilidade: difícil

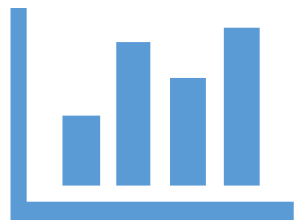
Função: composição

Exemplos: fluxos (cadastro, vendas, formulários, navegação) com muitas ramificações.

Atenção: mostra os caminhos que levam a diferentes etapas, indicando o volume do fluxo entre elas. De certo modo, é uma versão mais complexa de um funil, em que há ramificações dos caminhos.



5 DICAS IMPORTANTES

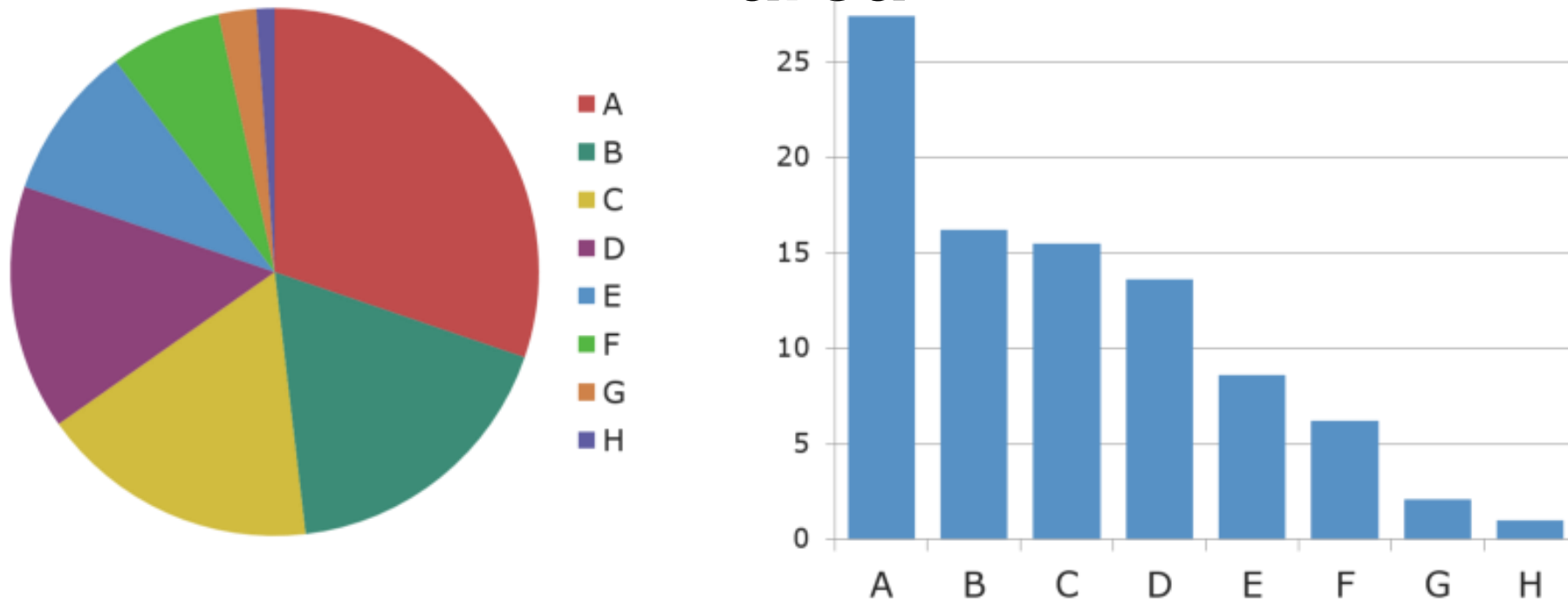


DICA 1

É mais fácil entender comprimento que área

DICA 1

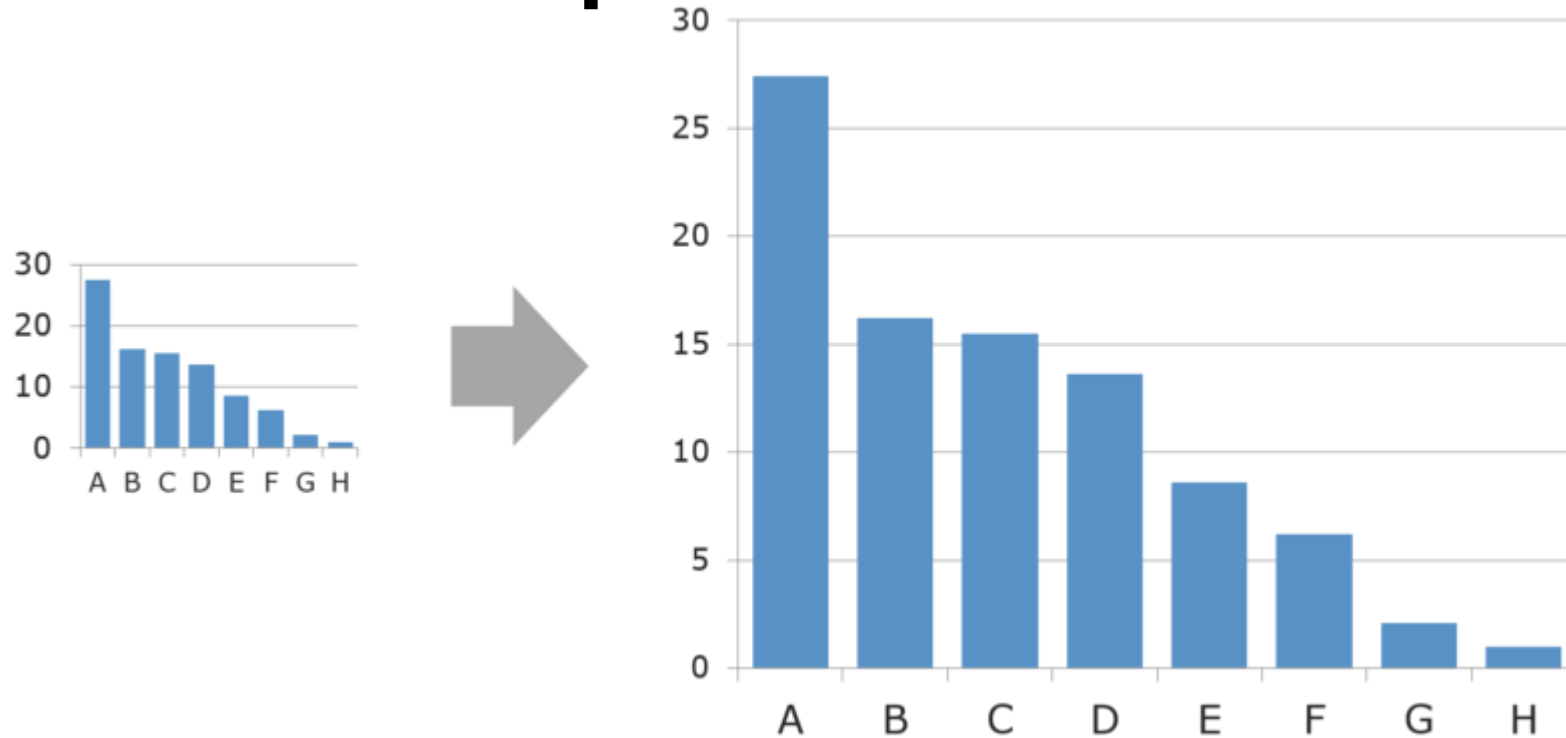
É mais fácil entender comprimento que área



Quando é necessário **comparar valores** individuais, o gráfico de colunas funciona melhor que o de pizza, pois nossa **percepção visual** de comprimentos (como em barras ou colunas) é melhor do que a de áreas (como na pizza).

DICA 1

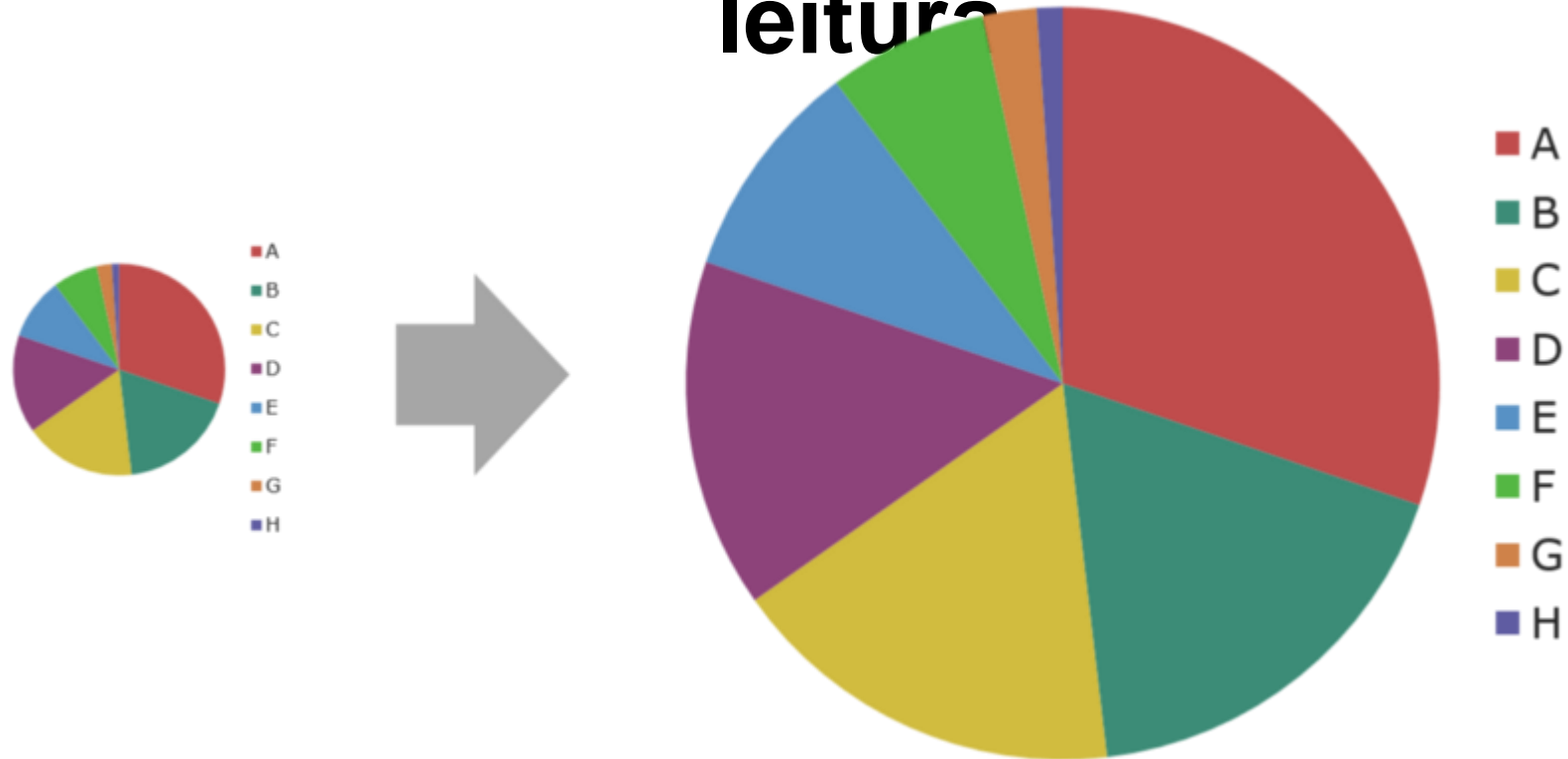
É mais fácil comparar valores combinados



O gráfico de barras também torna as informações mais claras para comparar valores combinados: o exemplo acima mostra votos em uma eleição e é difícil ver no gráfico de pizza se uma **aliança entre C e D** supera os votos de A, mas no gráfico de colunas a comparação é possível.

DICA 1

Aumentar gráfico pizza não traduz em melhor leitura



Um gráfico de colunas de tamanho pequeno mostra melhor as diferenças entre cada elemento do que um gráfico de pizza. Se aumentarmos o tamanho do gráfico de colunas, as diferenças ficam ainda mais claras. Por outro lado, **aumentar o tamanho de um gráfico de pizza não facilita a leitura nem traz mais informações.**

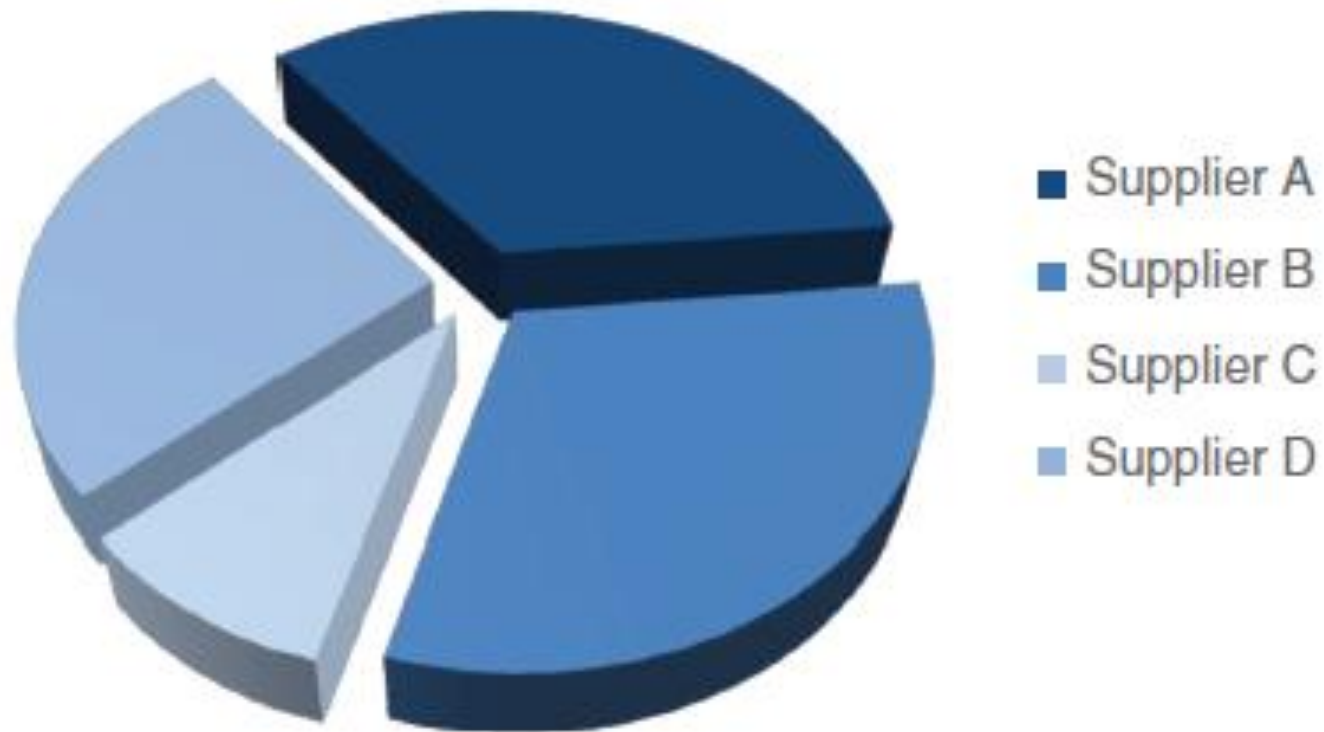


DICA 2

Gráficos 3D podem distorcer os dados

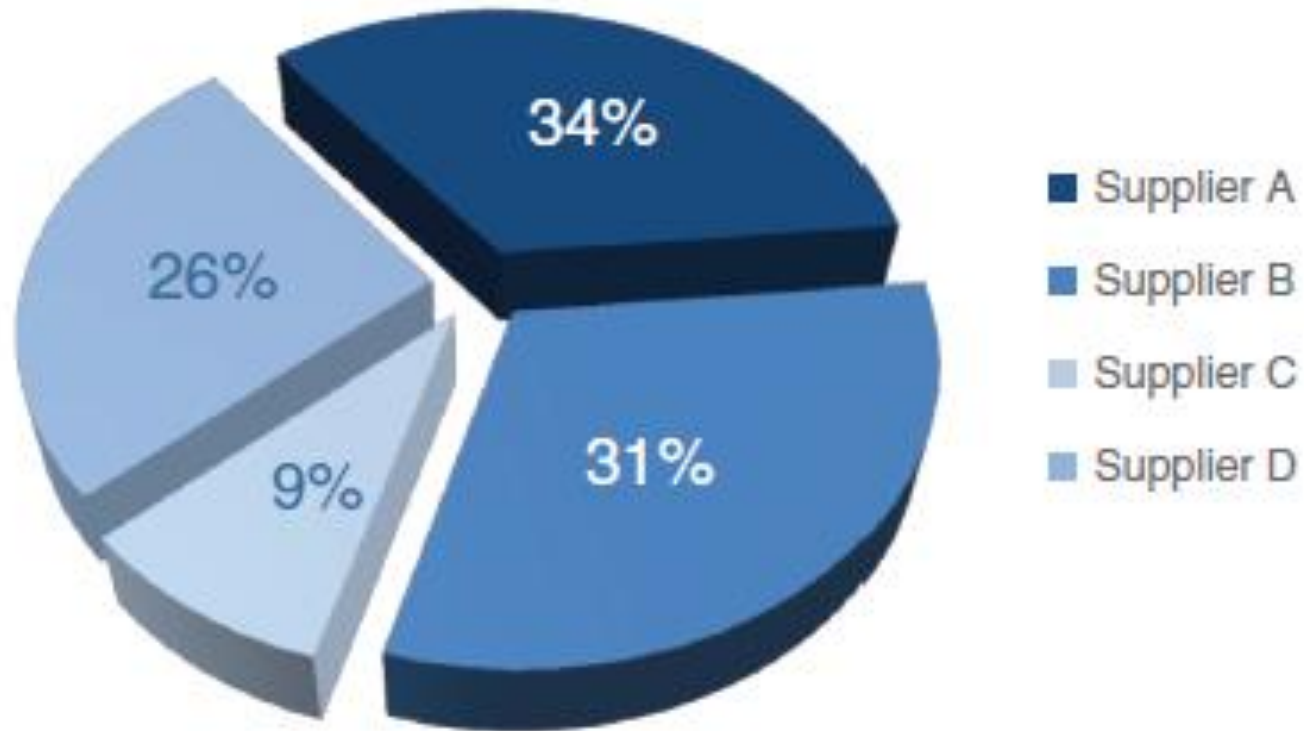
DICA 2

Qual “supplier” do gráfico tem o maior valor?



DICA 2

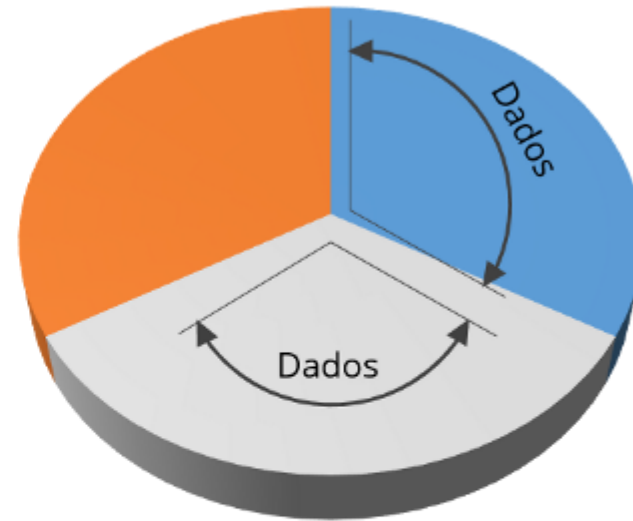
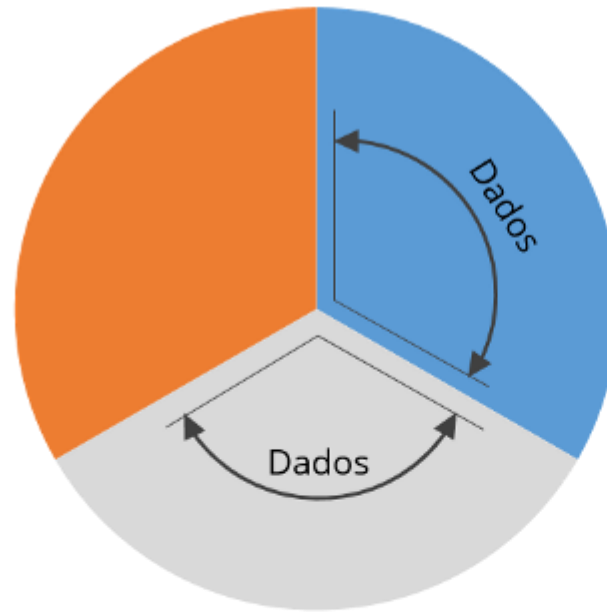
Qual “supplier” do gráfico tem o maior valor?



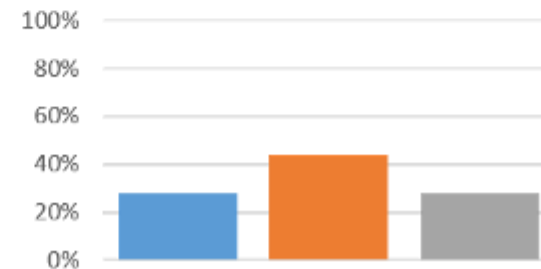
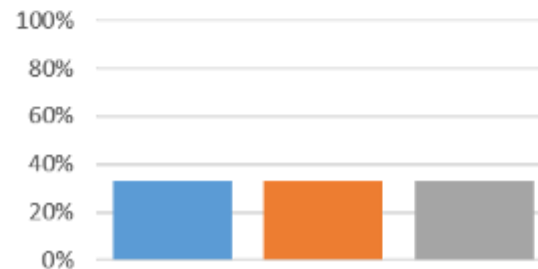
DICA 2

Gráficos 3D podem distorcer os dados

Ângulo
das fatias



Área das
fatias



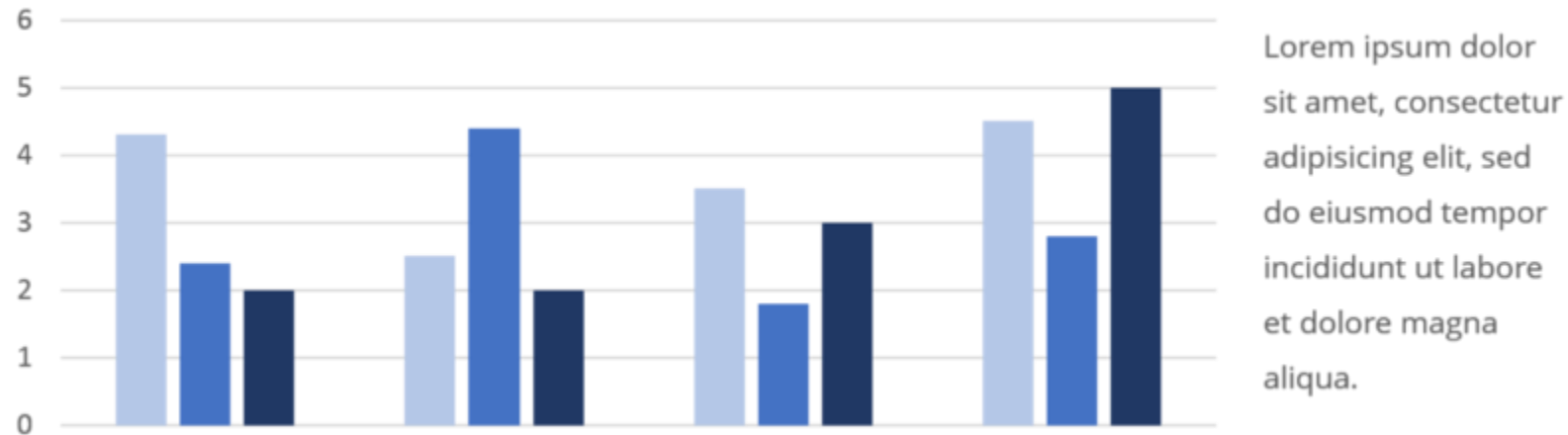


DICA 3

Informações visuais devem ser apresentadas primeiro

DICA 3

Informações visuais devem ser apresentadas primeiro



A **ordem de leitura** em português é da esquerda para a direita, portanto se temos informações visuais e um texto de apoio, é razoável que as informações visuais sejam colocadas primeiro, pois em geral são mais fáceis de entender.



DICA 4

A imagem importa, mas as letras também

DICA 4

A imagem importa, mas as letras também



COM SERIFA (serif)
Ex: Times New Roman

Melhor para versão impressa



SEM SERIFA (sans serif)
Ex: Arial

Melhor para versão digital

Usar fontes compatíveis com várias plataformas ou sistemas evita que o arquivo perca a formatação quando for enviado para outras pessoas.

Exemplos dessas fontes são: **Verdana, Arial, Georgia e Times New Roman.**



DICA 5

Pizza 3D não, por favor



6 FERRAMENTAS DE VISUALIZAÇÃO

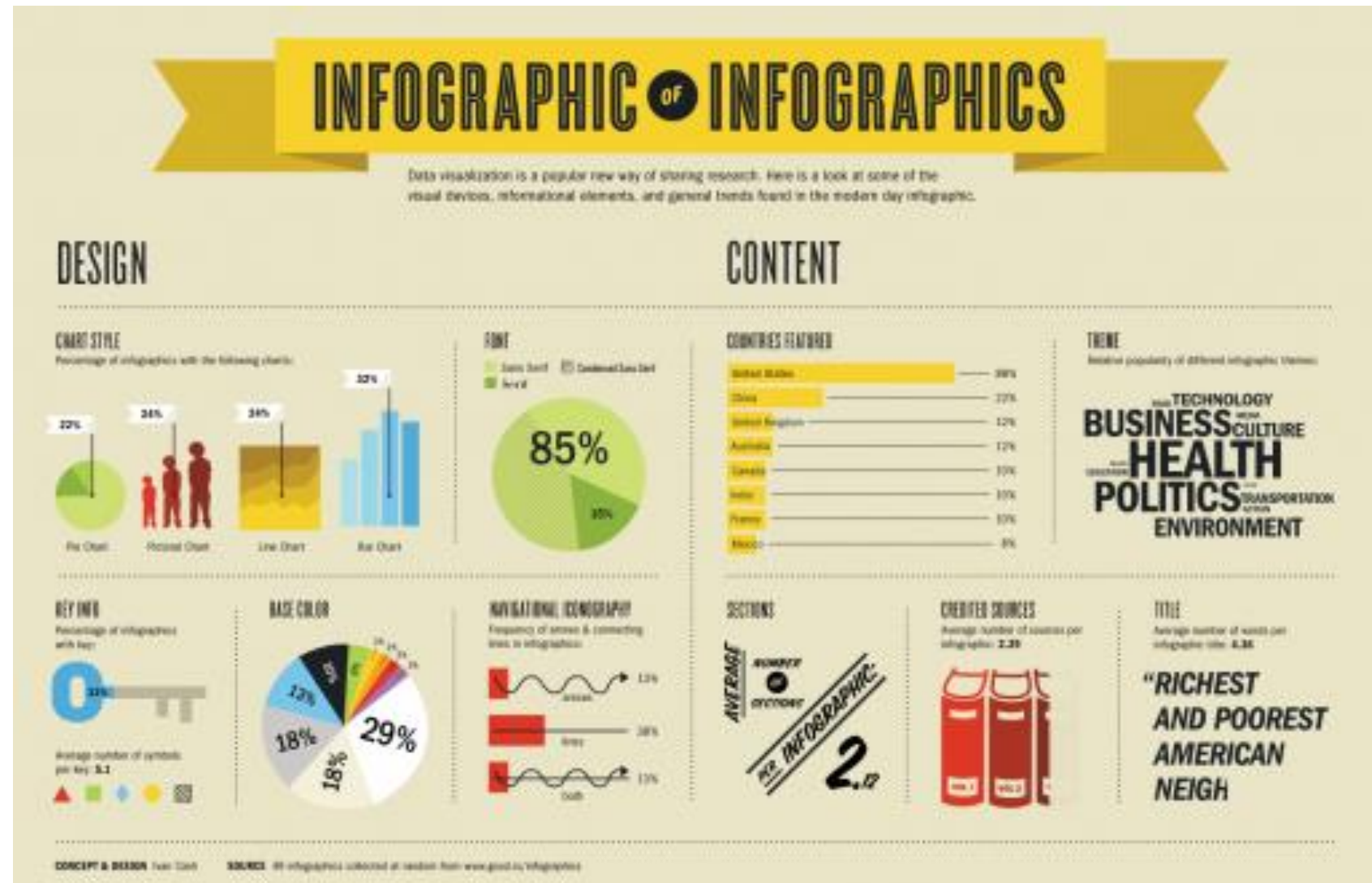


6 FERRAMENTAS DE VISUALIZAÇÃO

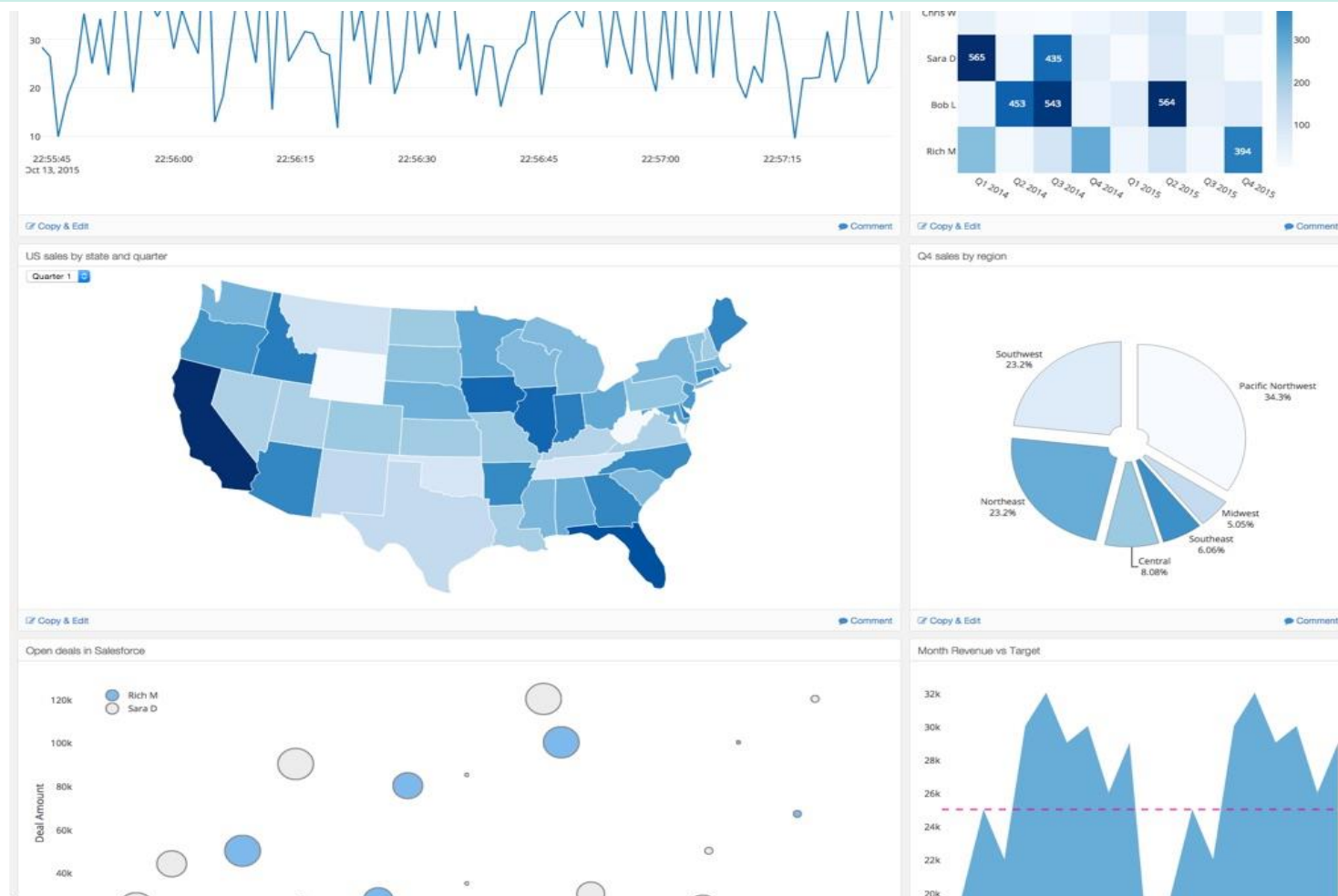


6 FERRAMENTAS DE VISUALIZAÇÃO

infogram



6 FERRAMENTAS DE VISUALIZAÇÃO

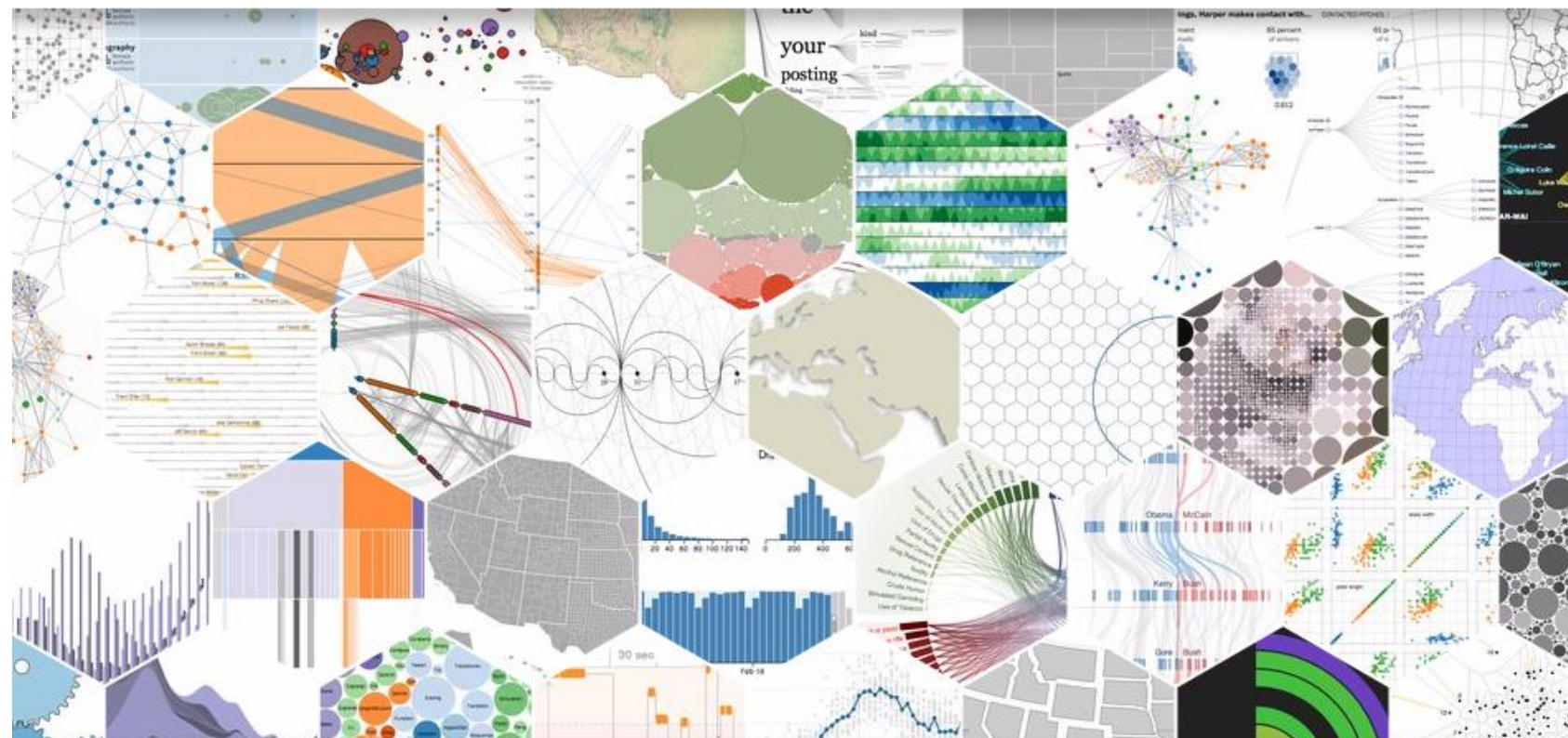


Plotly é uma biblioteca de visualização de dados para Python, Javascript e R

6 FERRAMENTAS DE VISUALIZAÇÃO

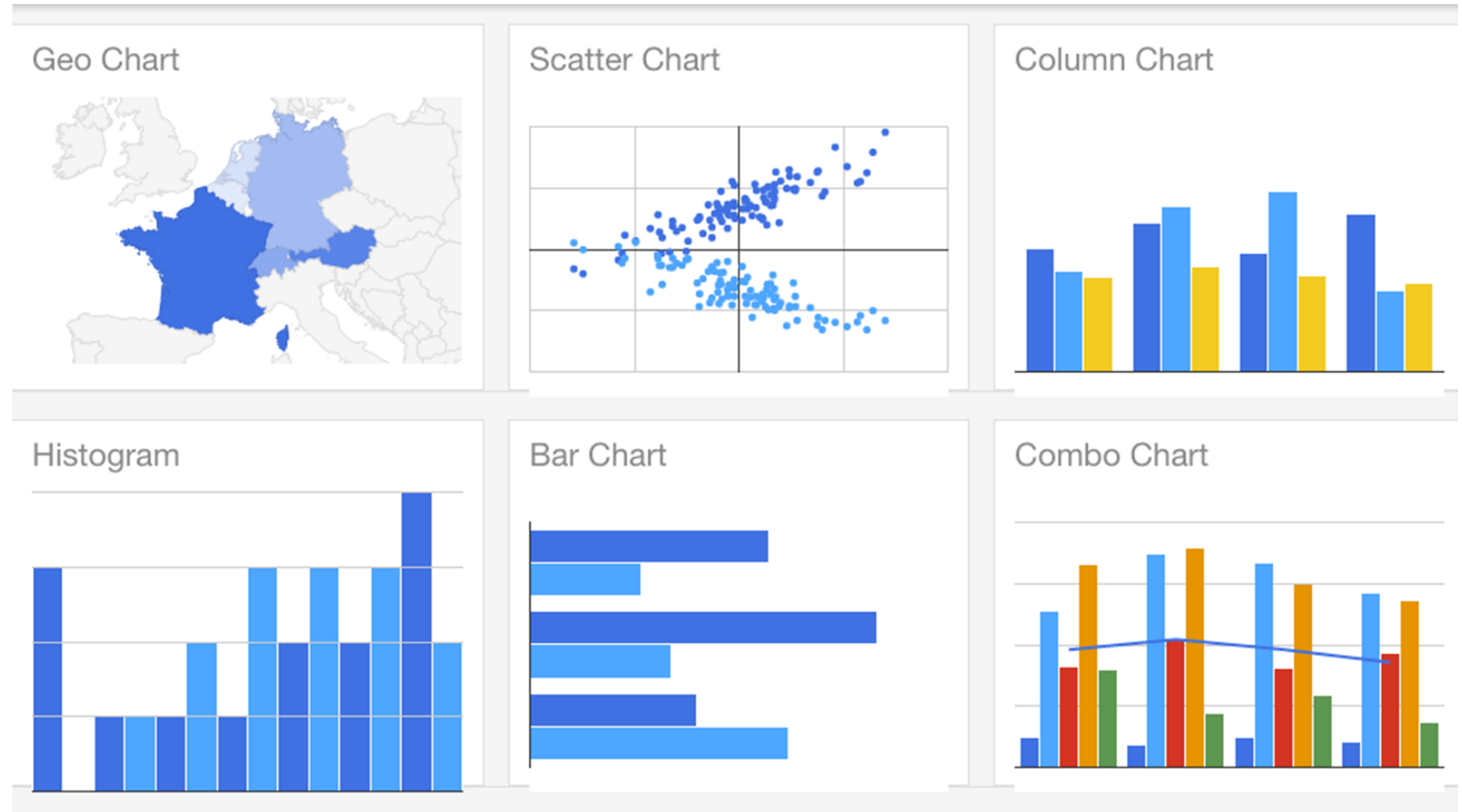


Data-Driven Documents



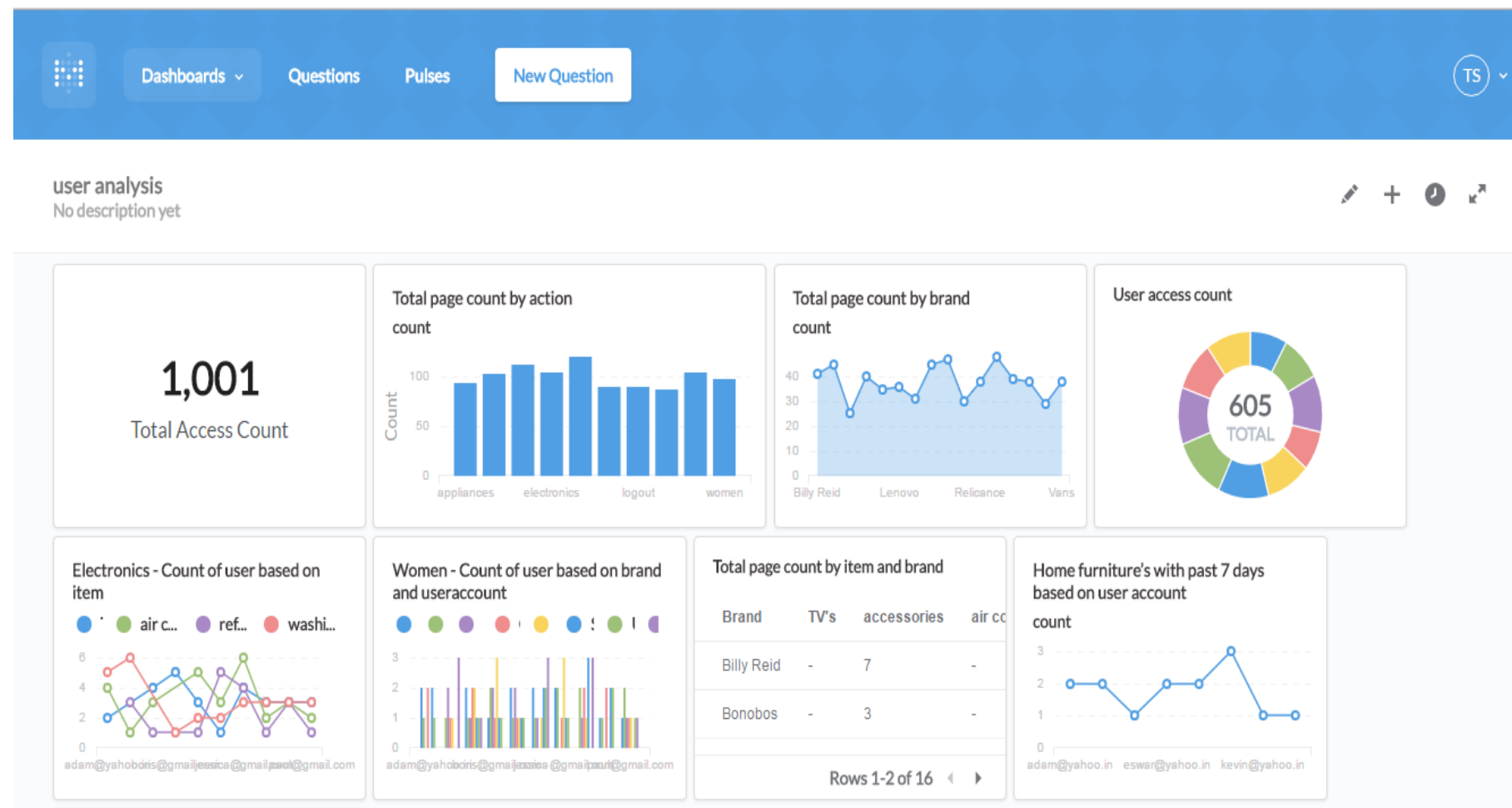
D3.js é gratuita e de código aberto. Para renderizar os gráficos, ela usa HTML, CSS e SVG

6 FERRAMENTAS DE VISUALIZAÇÃO



Compatibilidade com os sistemas iOS e Android.

6 FERRAMENTAS DE VISUALIZAÇÃO



Melhor ferramenta de BI open source