

Visualização de Dados e Storytelling

"Um gráfico vale mais que mil números"

Transformando dados em histórias visuais impactantes



O que é Visualização de Dados?

A visualização é o processo de transformar dados em representações visuais.

É a prática de traduzir informações numéricas ou textuais em um formato gráfico que facilita a compreensão e análise. Através de elementos visuais como gráficos, mapas e diagramas, podemos identificar padrões, tendências e relações que seriam difíceis de perceber em formato bruto.

Objetivos da Visualização de Dados:

- Explorar:** Descobrir padrões e insights ocultos nos dados
- Analisar:** Compreender relações e tendências
- Comunicar:** Transmitir informações de forma clara e acessível
- Tomar decisões:** Fundamentar escolhas com base em evidências visuais



Tipos de Gráficos e Quando Usá-los

Escolher o tipo certo de gráfico é essencial para comunicar efetivamente sua mensagem. Cada tipo de visualização tem pontos fortes específicos e casos de uso ideais.

Gráfico de Barras

Compara valores entre diferentes categorias usando barras retangulares.

Exemplo: Comparar vendas por produto ou região.

Gráfico de Dispersão

Exibe a relação entre duas variáveis numéricas como pontos em um plano.

Exemplo: Relação entre altura e peso de uma população.

Gráfico de Linha

Mostra dados contínuos ao longo do tempo, conectados por linhas.

Exemplo: Evolução de temperatura ou crescimento populacional.

Gráfico de Pizza

Mostra a proporção de partes de um todo em formato circular.

Exemplo: Porcentagem de votos para cada candidato.

Boxplot

Resume a distribuição de um conjunto de dados mostrando quartis e outliers.

Exemplo: Distribuição de notas em diferentes turmas.

Mapa de Calor

Usa cores para representar valores em uma matriz bidimensional.

Exemplo: Visualizar correlações entre variáveis.

Gráfico de Barras

O gráfico de barras é uma das formas mais comuns e eficazes de visualizar dados categóricos. Ele usa barras retangulares com comprimentos proporcionais aos valores que representam.

Quando usar:

- ✓ **Comparação entre categorias:** Ideal para comparar valores entre diferentes grupos ou categorias
- ✓ **Dados discretos:** Perfeito para visualizar dados que podem ser contados em unidades distintas
- ✓ **Ranking:** Excelente para mostrar classificações e ordenações
- ✓ **Distribuição:** Útil para visualizar a distribuição de dados em categorias

Variações:

- **Barras verticais:** Mais comum, ideal quando há muitas categorias
- **Barras horizontais:** Melhor para nomes de categorias longos ou muitas categorias
- **Barras agrupadas:** Para comparar múltiplas variáveis entre categorias
- **Barras empilhadas:** Para mostrar partes de um todo em cada categoria

Exemplo: Vendas por Produto (em milhares)

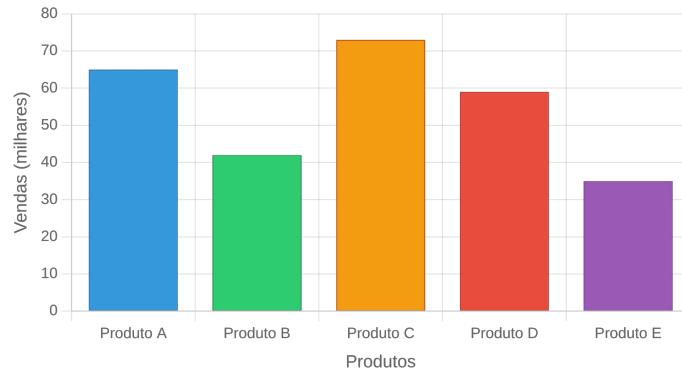
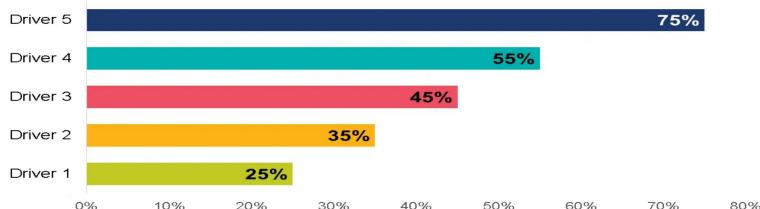


Gráfico de barras verticais mostrando comparação entre produtos

Drivers de aumento de eficiência com gráfico de barras horizontais



This graph/chart is linked to excel, and changes automatically based on data. Just left click on it and select "Edit Data".

Gráfico de Linha

O gráfico de linha conecta pontos de dados individuais com linhas retas, sendo ideal para visualizar tendências ao longo do tempo ou mostrar a relação contínua entre pontos de dados.

Quando usar:

- ✓ **Dados temporais:** Perfeito para mostrar tendências ao longo do tempo
- ✓ **Dados contínuos:** Ideal para valores que mudam continuamente
- ✓ **Comparação de tendências:** Excelente para comparar múltiplas séries temporais
- ✓ **Previsões:** Útil para projetar tendências futuras

Boas práticas:

- **Escala apropriada:** Use escalas que não distorçam a percepção das tendências
- **Limite de linhas:** Evite mais de 4-5 linhas para não sobrecarregar o gráfico
- **Cores distintas:** Use cores contrastantes para diferenciar múltiplas linhas
- **Marcadores:** Considere usar marcadores para destacar pontos de dados importantes

Exemplo: Temperatura Média Mensal (°C)

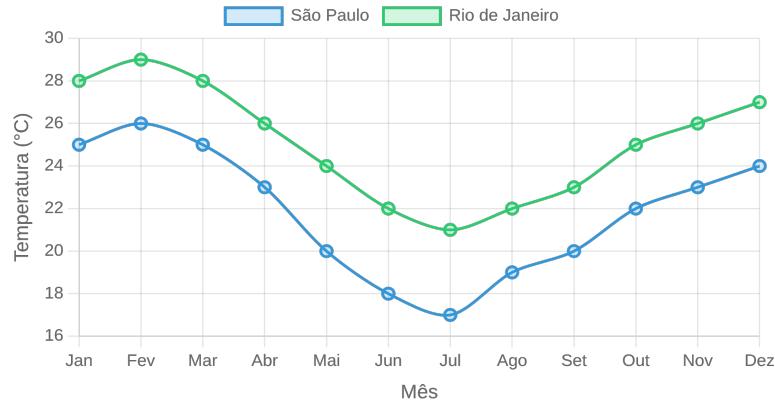


Gráfico de linha mostrando a variação de temperatura ao longo do ano

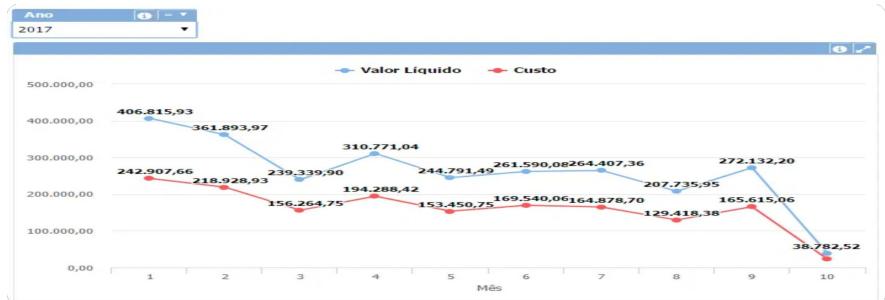


Gráfico de Pizza

O gráfico de pizza divide um círculo em fatias para mostrar a proporção de partes de um todo. Cada fatia representa uma categoria, e o tamanho da fatia corresponde à sua porcentagem do total.

Quando usar:

- ✓ **Proporções:** Para mostrar como diferentes partes contribuem para um todo
- ✓ **Poucas categorias:** Ideal quando há no máximo 5-7 categorias
- ✓ **Valores percentuais:** Quando a soma de todos os valores é 100%
- ✓ **Composição:** Para visualizar a composição de um conjunto de dados

Limitações:

- ⚠ **Difícil comparação:** Comparar o tamanho de fatias não adjacentes é difícil
- ⚠ **Muitas categorias:** Torna-se confuso com mais de 7 categorias
- ⚠ **Fatias pequenas:** Difícil visualizar categorias com valores muito pequenos
- ⚠ **Alternativas:** Considere gráficos de barras para comparações mais precisas

Exemplo: Distribuição de Vendas por Região

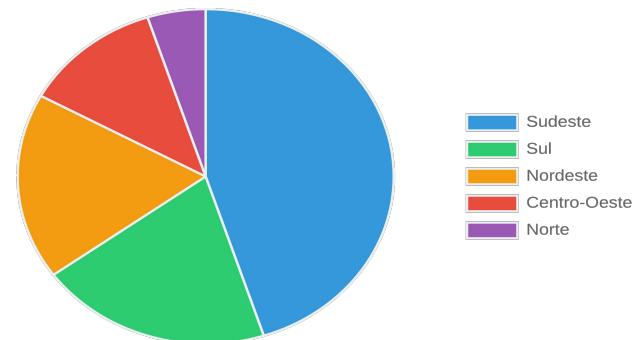


Gráfico de pizza mostrando a proporção de vendas em diferentes regiões

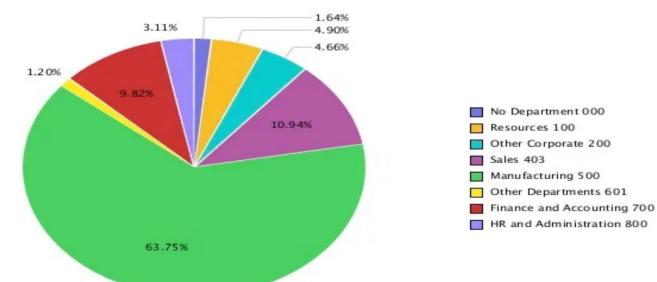


Gráfico de Dispersão

O gráfico de dispersão (ou scatter plot) exibe a relação entre duas variáveis numéricas, representando cada observação como um ponto em um plano cartesiano.

Quando usar:

- ✓ **Correlações:** Para identificar relações entre duas variáveis
- ✓ **Padrões:** Para detectar agrupamentos, tendências ou valores atípicos
- ✓ **Distribuição:** Para visualizar como os dados estão distribuídos em duas dimensões
- ✓ **Regressão:** Como base para análises de regressão linear ou não-linear

Interpretando correlações:

Correlação positiva: Pontos formam uma linha ascendente (quando X aumenta, Y tende a aumentar)

Correlação negativa: Pontos formam uma linha descendente (quando X aumenta, Y tende a diminuir)

Sem correlação: Pontos dispersos sem padrão claro (X e Y não têm relação aparente)

Exemplo: Relação entre Altura (cm) e Peso (kg)

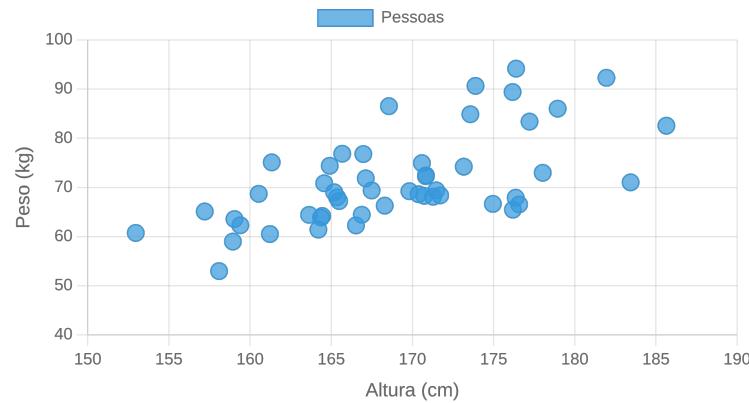
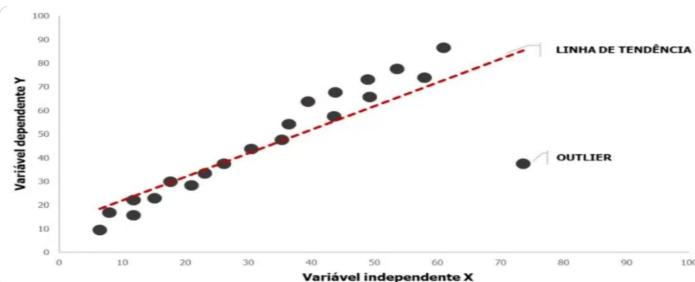


Gráfico de dispersão mostrando correlação positiva entre altura e peso



Boxplot (Diagrama de Caixa)

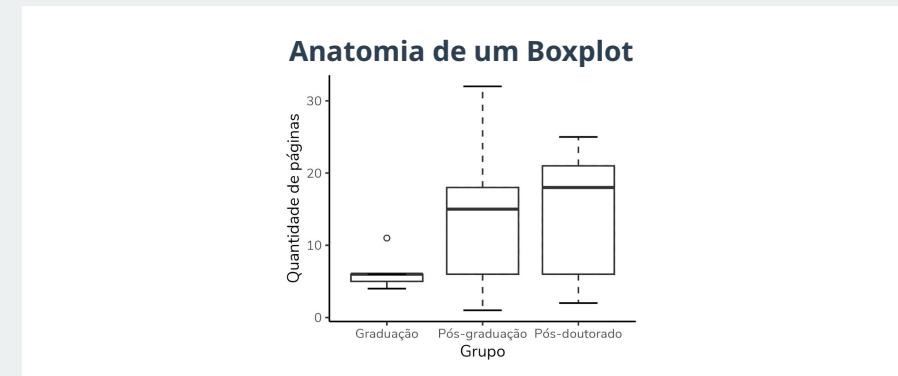
O boxplot é uma representação gráfica da distribuição de dados numéricos através de seus quartis, permitindo identificar outliers e avaliar a dispersão dos dados.

Quando usar:

- ✓ **Comparar distribuições:** Para diferentes grupos
- ✓ **Identificar outliers:** Detectar valores atípicos
- ✓ **Analisar dispersão:** Visualizar variabilidade
- ✓ **Resumir dados:** Apresentar estatísticas resumidas

Como interpretar:

- **Linha central:** Mediana (Q2) - valor central
- **Caixa:** Intervalo interquartil (IQR) - entre Q1 e Q3
- **Bigodes:** Valores mínimo e máximo dentro de $1,5 \times \text{IQR}$
- **Pontos isolados:** Outliers - valores atípicos



Boas Práticas em Visualização de Dados

H Título e Legendas

Sempre inclua um título claro e descriptivo. Use legendas para explicar elementos do gráfico e facilitar a interpretação.

↔ Eixos

Os eixos X e Y devem ser nomeados e ter a escala correta. Evite cortar eixos para não distorcer a percepção dos dados.

Cores

Use cores de forma consciente e com propósito. Escolha paletas que sejam acessíveis para daltônicos e mantenha consistência.

🚫 Evite Poluição Visual

Remova elementos desnecessários que não contribuem para a compreensão dos dados. Simplifique para destacar a mensagem principal.

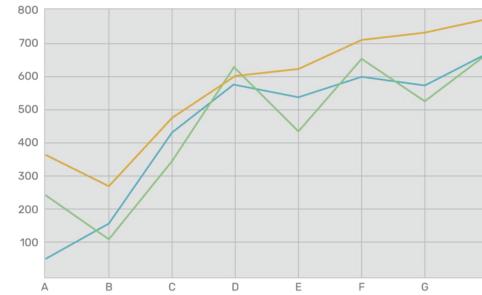
🎯 Foco na Mensagem

Destaque os insights mais importantes. Cada visualização deve responder a uma pergunta específica ou comunicar uma mensagem clara.

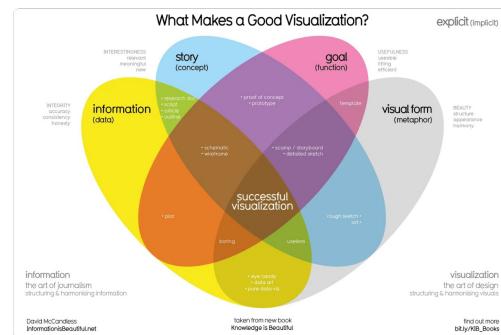
💻 Fonte dos Dados

Sempre cite a fonte dos dados utilizados. Isso aumenta a credibilidade da visualização e permite que outros verifiquem as informações.

Exemplo de Boa Prática



O que faz uma boa visualização?



Prática: Python para Visualizar

```
# Importando as bibliotecas necessárias
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Criando um DataFrame simples
dados = {
    'cidade': ['São Paulo', 'Rio de Janeiro', 'Belo Horizonte', 'Salvador', 'Brasília'],
    'populacao': [12.33, 6.75, 2.72, 2.88, 3.05], # em milhões
    'area': [1521, 1200, 331, 693, 5802] # em km²
}

df = pd.DataFrame(dados)

# Criando um gráfico de barras
plt.figure(figsize=(10, 6))
sns.barplot(x='cidade', y='populacao', data=df)
plt.title('População das Principais Cidades Brasileiras')
plt.xlabel('Cidade')
plt.ylabel('População (milhões)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('grafico_barras.png')

# Criando um gráfico de dispersão
plt.figure(figsize=(10, 6))
sns.scatterplot(x='area', y='populacao', data=df)
plt.title('Relação entre Área e População')
plt.xlabel('Área (km²)')
plt.ylabel('População (milhões)')
for i, txt in enumerate(df.cidade):
    plt.annotate(txt, (df.area[i], df.populacao[i]))
plt.tight_layout()
plt.savefig('grafico_dispersao.png')
```

- **Passo 1:** Importar as bibliotecas (pandas, matplotlib, seaborn)
- **Passo 2:** Criar ou carregar os dados em um DataFrame
- **Passo 3:** Escolher o tipo de gráfico adequado para os dados
- **Passo 4:** Personalizar títulos, eixos e outros elementos visuais

Atividades Práticas

Atividade B: Prática Individual - "Do Dado ao Gráfico"

Cada aluno trabalhará individualmente com um dataset público para criar visualizações que comuniqueem insights relevantes.

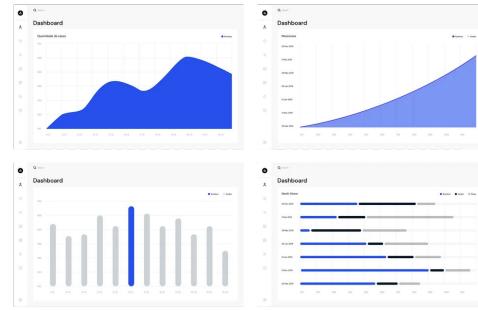
Passo a Passo da Atividade Individual:

-  **Escolha do Dataset:** Será fornecido um link para um dataset limpo (ex: dados de vendas, indicadores de saúde)
-  **Gráfico de Barras/Pizza:** Contar ocorrências de categorias e criar visualização comparativa
-  **Gráfico de Linha/Dispersão:** Analisar evolução temporal ou relação entre variáveis
-  **Análise:** Identificar e documentar insights encontrados nas visualizações

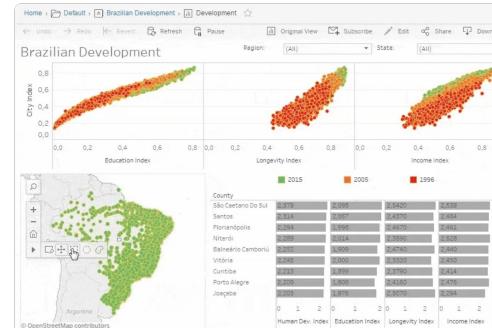
Atividade C: Produção de Dashboards em Grupos

Em grupos, os alunos criarão dashboards interativos usando o Google Sheets para responder perguntas específicas sobre os dados.

Exemplo de Dashboard Individual



Exemplo de Dashboard em Grupo



Checklist de Boas Práticas

Para futuros projetos de visualização de dados

Antes de finalizar sua visualização, verifique:

- Título claro e descritivo que comunica a mensagem principal
- Eixos nomeados corretamente e com escalas apropriadas
- Legendas para cada elemento do gráfico
- Cores escolhidas com propósito e acessíveis para daltônicos
- Fonte dos dados claramente indicada
- Elementos visuais desnecessários removidos
- Tipo de gráfico adequado para os dados apresentados
- Insights principais destacados visualmente
- Contexto suficiente para entendimento da visualização



"A visualização de dados é uma das habilidades mais valiosas para um cientista de dados, pois é através dela que a análise ganha sentido e se torna acessível a todos."