

Estatística Descritiva

Sumário

- ❑ Iniciativas de sucesso
- ❑ Recursos e exemplos
- ❑ Dicas sobre gráficos

Iniciativas de sucesso





Storytelling

É a arte de construir e transmitir narrativas que organizam informações de forma coerente, envolvente e significativa.

Em lugar de apresentar dados de maneira isolada, o storytelling busca conectar esses dados a um contexto, criando uma história que destaca padrões, tendências e insights relevantes para o público-alvo. Alguns aspectos são:

- ❑ **Integração:** garantir a qualidade e a consistência dos dados, bem como o alinhamento da análise às necessidades e interesses da audiência;
- ❑ **Engajamento:** capturar e manter a atenção do público, incentivando-o a interagir e a se conectar com a mensagem;
- ❑ **Memorabilidade:** garantir que os *insights* gerados a partir dos dados não apenas sejam compreendidos, mas também permaneçam na memória do público após a apresentação, de forma a impulsionar ou orientar sua ação;



Hans Rosling e o Gapminder

Histórico e Impacto:

Hans Rosling revolucionou a maneira de comunicar dados globais ao apresentar, de forma dinâmica e interativa, estatísticas sobre saúde, renda e desenvolvimento. Com seus vídeos e palestras – muitos deles disponíveis no TED – Rosling mostrou que, ao animar gráficos de dispersão que ilustram o progresso de países ao longo do tempo, é possível derrubar preconceitos e estimular debates embasados em evidências.

Como o Data Storytelling Contribuiu:

- ❑ **Visualizações Dinâmicas:** Os gráficos animados permitiram visualizar o progresso em indicadores complexos de forma intuitiva.
- ❑ **Engajamento do Público:** A clareza e a narrativa envolvente motivaram tanto o público em geral quanto formuladores de políticas públicas a reavaliarem suposições sobre países em desenvolvimento.

https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?subtitle=pt-br



Our World in Data

Histórico e Impacto:

Our World in Data é uma plataforma que compila dados sobre temas como saúde, educação, meio ambiente e desigualdade, apresentando-os de forma interativa e narrativa. Suas visualizações têm sido amplamente utilizadas por pesquisadores, formuladores de políticas e educadores para fundamentar debates sobre desenvolvimento sustentável e políticas públicas..

Como o Data Storytelling Contribuiu:

- ❑ **Narrativa Integrada:** Cada conjunto de dados é contextualizado com explicações que relacionam os números a tendências históricas e desafios atuais, facilitando a compreensão dos problemas complexos.
- ❑ **Referência para Decisões:** Organizações internacionais e governos frequentemente utilizam essas visualizações para embasar relatórios e orientar estratégias de desenvolvimento.



Tchiluanda/Tesouro Transparente

Histórico e Impacto:

A iniciativa Tchiluanda é uma equipe da Secretaria do Tesouro Nacional que usa técnicas de data storytelling para comunicar informações sobre políticas públicas.

Como o Data Storytelling Contribuiu:

- ❑ **Contribuições para a cultura de dados:** Desenvolvimento de pacotes em R para recuperar dados do SICONFI (Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro).
- ❑ **Promoção da transparência pública:**
 - ❑ Criação de visualizações de dados com gráficos animados.
 - ❑ Publicação de artigos sobre políticas públicas no Medium.

<https://tchiluanda.github.io/>

<https://www.tesourotransparente.gov.br/descubra-explore-crie/explore>

Recursos e exemplos





Recursos

❑ **Gapminder**

Tools:

O Gapminder apresenta dados globais em gráficos de bolhas interativos, permitindo explorar tendências históricas e comparações entre países. O design limpo, o uso adequado de cores e a narrativa visual tornam este exemplo um paradigma em data storytelling.

<https://www.gapminder.org/tools/>

❑ **Our**

World

in

Data:

Excelente exemplo de como organizar e apresentar grandes volumes de dados de forma interativa e narrativa. Cada tema é explorado por meio de gráficos interativos, acompanhados de textos explicativos que contextualizam as informações e facilitam a compreensão.

<https://ourworldindata.org/>

❑ **Information**

is

Beautiful:

Canal dedicado à apresentação de dados complexos por meio de visualizações estáticas e interativas de alta qualidade, usando elementos visuais e narrativas que ajudam a desvendar histórias contidas nos dados.

<https://informationisbeautiful.net/>



Recursos

- ❑ **Storytelling with Data:**

Guias, artigos, livros, eventos e outros materiais relacionados à disciplina de storytelling com dados.

<https://www.storytellingwithdata.com/>

- ❑ **FlowingData:**

Artigos, guias e outros materiais acerca de elementos de visualização estáticas, usando princípios de design e narrativa.

<https://flowingdata.com/>

- ❑ **Datawrapper**

Serviço de Software-as-a-Service com uma extensa de galeria do gráficos e outros recursos de comunicação interativa Datawrapper É uma ótima referência para ver como escolhas de cores, tipografia e layout podem facilitar a compreensão dos dados.

<https://academy.datawrapper.de/category/178-example-charts-maps>



Plataformas de Análise

Qlik Sense (<https://www.qlik.com/>):

Plataforma de análise e visualização de dados que se destaca pela sua abordagem associativa. Ele permite que os usuários explorem os dados de maneira não-linear, revelando conexões que podem passar despercebidas em abordagens tradicionais.

- ❑ **Modelo Associativo:** Permite a exploração interativa dos dados, possibilitando descobrir relações e insights ocultos por meio de buscas dinâmicas.
- ❑ **Escalabilidade:** Adequado tanto para pequenas empresas quanto para grandes corporações, suportando altos volumes de dados com performance robusta.
- ❑ **Self-Service BI:** Oferece ferramentas que capacitam os usuários a criar suas próprias visualizações e análises, sem depender exclusivamente da área de TI.
- ❑ **Segurança e Governança de Dados:** Recursos avançados para controlar o acesso e garantir a integridade dos dados.



Plataformas de Análise

Metabase (<https://www.metabase.com/>):

Ferramenta open source de business intelligence que visa tornar a exploração de dados acessível a todos os usuários, mesmo aqueles sem conhecimentos técnicos avançados.

- ❑ **Fácil Configuração e Uso:** Com uma interface simples, permite que usuários criem consultas e dashboards rapidamente, sem necessidade de escrever SQL (embora essa opção esteja disponível para usuários avançados).
- ❑ **Código Aberto:** A comunidade ativa e a possibilidade de customização fazem do Metabase uma opção flexível para organizações que desejam adaptar a ferramenta às suas necessidades.
- ❑ **Custo-Benefício:** Ideal para startups e pequenas empresas que buscam uma solução de BI sem altos custos de licenciamento.
- ❑ **Visualizações Interativas:** Embora mais simples que outras ferramentas comerciais, oferece dashboards interativos e relatórios que facilitam a análise dos dados.



Plataformas de Análise

Power BI:

Desenvolvido pela Microsoft, o Power BI é uma plataforma robusta de business intelligence que integra perfeitamente com o ecossistema Office e Azure, oferecendo uma ampla gama de recursos para análise e visualização de dados.

- ❑ Integração com Ferramentas Microsoft: Fácil integração com Excel, SharePoint e outras soluções da Microsoft, tornando a importação e análise de dados mais simples.
- ❑ Recursos Avançados de Análise: Oferece funcionalidades para análise preditiva, machine learning e modelagem de dados, além de uma vasta biblioteca de visualizações.
- ❑ Custo internalizado: Disponível para toda a APF por meio de convênio com a Microsoft.
- ❑ Interface Intuitiva e Customizável: Permite a criação de dashboards interativos com um design atraente e personalizável, ideal para apresentações executivas.



Plataformas de Análise

Google Data Studio [Looker Studio] (<https://lookerstudio.google.com/>):

Uma ferramenta gratuita e baseada na web, ideal para transformar dados de diversas fontes (como Google Analytics, Planilhas Google, BigQuery, entre outros) em dashboards interativos e relatórios dinâmicos.

- ❑ Integração com o Ecossistema Google: Facilita a conexão com outros produtos Google.
- ❑ Interface Intuitiva: Permite a criação rápida de relatórios sem a necessidade de conhecimentos avançados em programação.
- ❑ Colaboração em Tempo Real: Possibilita que equipes trabalhem juntas na criação e edição dos dashboards.
- ❑ Casos de Uso: Relatórios de marketing digital, análises de desempenho de sites e monitoramento de campanhas online.



Plataformas de Análise

Apache Superset (<https://superset.apache.org/>):

Ferramenta open source de visualização de dados originalmente desenvolvida pelo Airbnb, que permite criar dashboards interativos e relatórios customizados com facilidade.

Características:

- ❑ Flexibilidade e Personalização: Altamente customizável por ser de código aberto.
- ❑ Escalabilidade: Adequado para organizações que desejam manter controle total sobre a infraestrutura de BI.
- ❑ Comunidade Ativa: Amplo suporte da comunidade open source, com constantes atualizações e melhorias.
- ❑ Casos de Uso: Soluções de BI internas, dashboards de monitoramento e análises exploratórias em ambientes que demandam soluções personalizadas.



Plataformas de Análise

Tableau (<https://www.tableau.com/>):

Uma das ferramentas líderes de mercado para visualização e análise de dados interativos. Conhecido por sua interface intuitiva, o Tableau permite que usuários criem dashboards sofisticados sem necessidade de programação profunda.

- ❑ Interface Drag & Drop: Permite que usuários criem visualizações de forma rápida e intuitiva, facilitando a exploração de dados.
- ❑ Integração com Múltiplas Fontes de Dados: Conecta-se a diversos tipos de bases de dados, desde arquivos Excel até grandes bancos de dados relacionais e nuvens de dados.
- ❑ Visualizações Avançadas: Oferece uma ampla variedade de gráficos e mapas, que podem ser personalizados para atender a necessidades específicas.
- ❑ Comunidade Ativa: Uma comunidade robusta e fóruns online que ajudam a solucionar dúvidas e compartilhar boas práticas.



Bibliotecas e frameworks

Dash by Plotly (<https://plotly.com/dash/>):

Dash é um framework Python desenvolvido pela Plotly para criar aplicativos analíticos e dashboards interativos. Ele combina componentes React com callbacks em Python, facilitando a integração de visualizações complexas e interatividade.

- ❑ Forte integração com a biblioteca Plotly para visualizações de alta qualidade.
- ❑ Suporte para layouts responsivos e customizáveis.
- ❑ Excelente para criar dashboards corporativos e aplicações analíticas robustas.
- ❑ Casos de Uso: Aplicativos financeiros, monitoramento de KPIs e análises em tempo real.



Bibliotecas e frameworks

Streamlit (<https://streamlit.io/>):

Biblioteca open source que permite transformar scripts Python em aplicativos web interativos de forma extremamente rápida e simples.

- ❑ Curva de aprendizado baixa, com uma sintaxe simples para criação de interfaces.
- ❑ Excelente para prototipagem rápida e para projetos de ciência de dados.
- ❑ Integração nativa com bibliotecas de visualização como Matplotlib, Plotly, Seaborn, entre outras.
- ❑ Casos de Uso:
- ❑ Demonstrações de modelos de machine learning, painéis interativos e exploração de dados.



Bibliotecas e frameworks

Panel (<https://panel.holoviz.org/>):

Panel é uma biblioteca da iniciativa HoloViz que facilita a criação de dashboards interativos e aplicações web com Python. Ela é construída sobre Bokeh, mas também integra outras bibliotecas como Matplotlib, Plotly e Altair.

- ❑ Permite combinar visualizações de diversas bibliotecas em um único dashboard.
- ❑ Flexibilidade para criar layouts complexos e interativos.
- ❑ Fácil integração com widgets interativos para exploração de dados.
- ❑ Casos de Uso:
- ❑ Aplicações empresariais, painéis de monitoramento e explorações de dados que requerem múltiplas fontes de visualização.



Exemplos

- ❑ **The New York Times – Interactive Graphics:**
Conhecido por suas reportagens com dados, o New York Times possui uma seção dedicada a gráficos interativos que contam histórias complexas de maneira acessível e visualmente atraente.
<https://www.nytimes.com/interactive>
- ❑ **The Pudding:**
Canal de publicação de dissertações visuais/interativas com diversos exemplos de boa utilização das técnicas de storytelling na comunicação e discussão de informações baseadas em dados
<https://pudding.cool/>
- ❑ **Mastering your productivity:**
Análise exploratória de dados acerca do sentimento de produtividade, demonstra a aplicação dos princípios de comunicação visual
<https://community.fabric.microsoft.com/t5/Data-Stories-Gallery/Mastering-Your-Productivity-in-2017-Survey-by-Levo-and-Microsoft/td-p/115410>



Exemplos (data-storytelling):

- ❑ **We feel fine:**

Representações artísticas de estatísticas acerca de expressões relacionadas a sentimento humano utilizadas na internet

<https://jjh.org/we-feel-fine>

- ❑ **Oxfam Impact:**

Relatório estatístico anual da atuação da Oxfam Ireland, demonstra a utilização de uma estrutura narrativa simples e engajante na apresentação dos dados.

<https://stories.oxfamireland.org/impact/index.html>

- ❑ **Gendered Descriptions:**

Apresentação de uma pesquisa quantitativa/analítica sobre representação de gênero na literatura.

<https://pudding.cool/2020/07/gendered-descriptions/>

- ❑ **Jews of Lebanon:**

Iniciativa do grupo Arab News, apresentando a história dos movimentos migratórios da população judia do Líbano.

<https://www.arabnews.com/JewsOfLebanon>

Dicas sobre gráficos





Boas práticas

- ❑ **Defina um Tema Consistente:**
Utilize `sns.set_theme()` ou `sns.set_style()` para estabelecer um estilo visual uniforme para todos os gráficos.

```
import seaborn as sns

sns.set_theme(style="whitegrid") # Opções: "darkgrid", "ticks", etc.
```

- ❑ **Adapte o Contexto do Gráfico:**
Configure o contexto (tamanho dos elementos gráficos) com `sns.set_context()`, ideal para apresentações, publicações ou relatórios.

```
sns.set_context("talk") # Outras opções: "paper", "notebook", "poster"
```

Boas práticas

- ❑ **Escolha Paletas de Cores Apropriadas:**
Seaborn oferece paletas embutidas (como "deep", "muted", "pastel", "bright", "dark" e "colorblind"). Use paletas que facilitam a distinção entre categorias.

```
#https://seaborn.pydata.org/tutorial/color\_palettes.html  
sns.set_palette("light:b", as_cmap=True)
```

- ❑ **Ajuste os Limites e Escalas:**
Configure os limites dos eixos com `plt.xlim()` e `plt.ylim()` e considere usar escalas logarítmicas ou outros ajustes quando os dados exigirem.
- ❑ **Controle de Legendas e Ordens:**
Utilize parâmetros como `hue`, `hue_order` e `dodge` para organizar a informação e evitar sobreposição de dados.

```
sns.barplot(  
    data=df, x='categoria', y='valor',  
    hue='grupo', hue_order=["A", "B", "C"], dodge=True  
)
```




Boas práticas

- ❑ **Ajuste de Tamanhos e Fontes:**
Use `sns.set_context()` em conjunto com `font_scale` para melhorar a legibilidade.

```
sns.set_context("paper", font_scale=1.1)
```

- ❑ **Remoção de linhas desnecessárias:**
Remova spines desnecessários com `sns.despine()`, deixando o gráfico mais limpo e focado nos dados.

```
sns.despine(left=True, bottom=True)
```

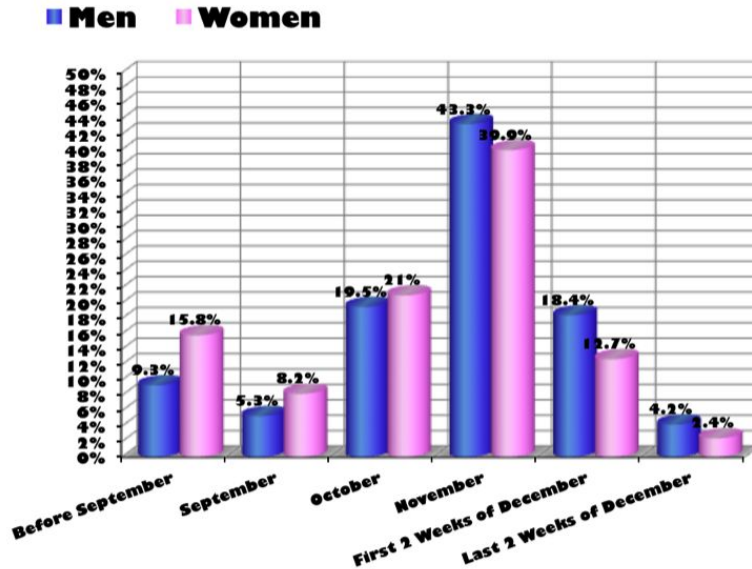
- ❑ **Controle de Legendas e Ordens:**
Utilize parâmetros como `hue`, `hue_order` e `dodge` para organizar a informação e evitar sobreposição de dados.

```
sns.barplot(  
    data=df, x='categoria', y='valor',  
    hue='grupo', hue_order=["A", "B", "C"], dodge=True  
)
```

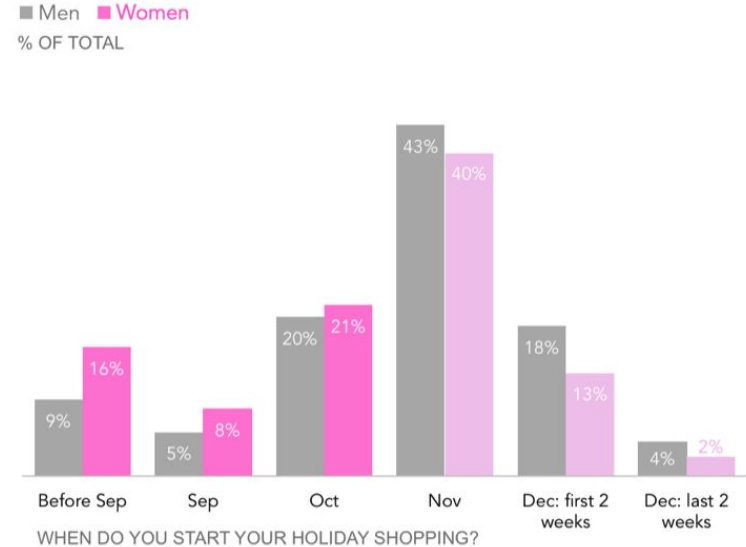
Exemplos

Before & after

Shoppers Begins Shopping for Holidays



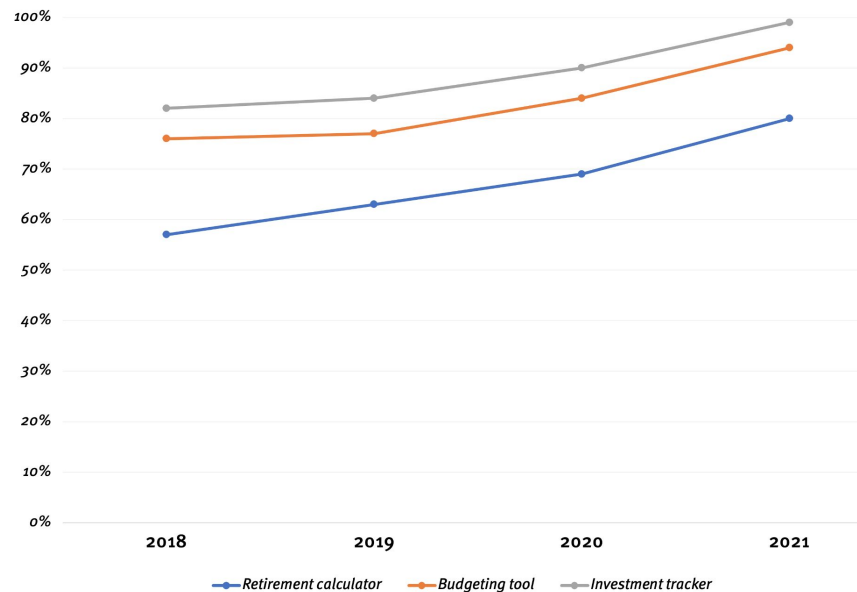
More women start their holiday shopping early



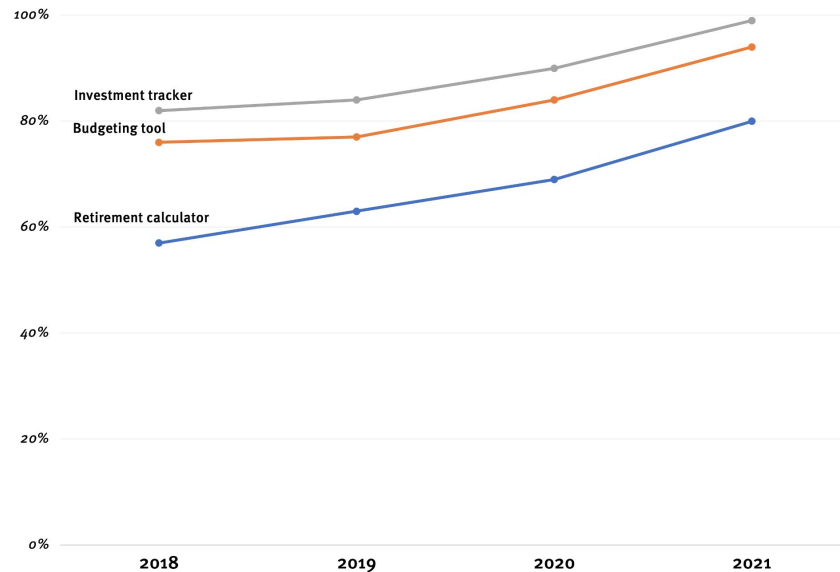
Exemplos



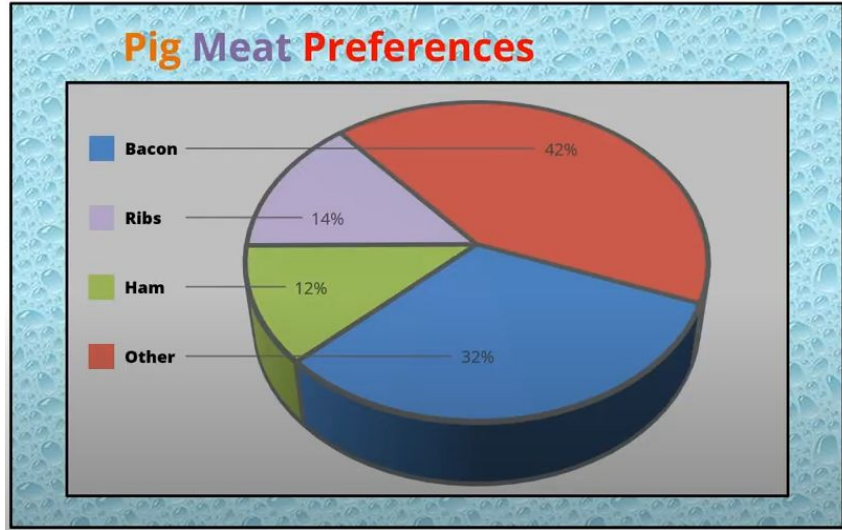
Success rates for top features continue to improve



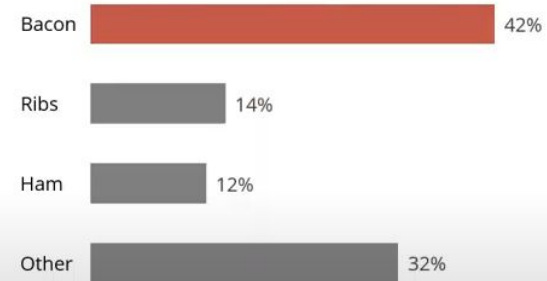
Success rates for top features continue to improve



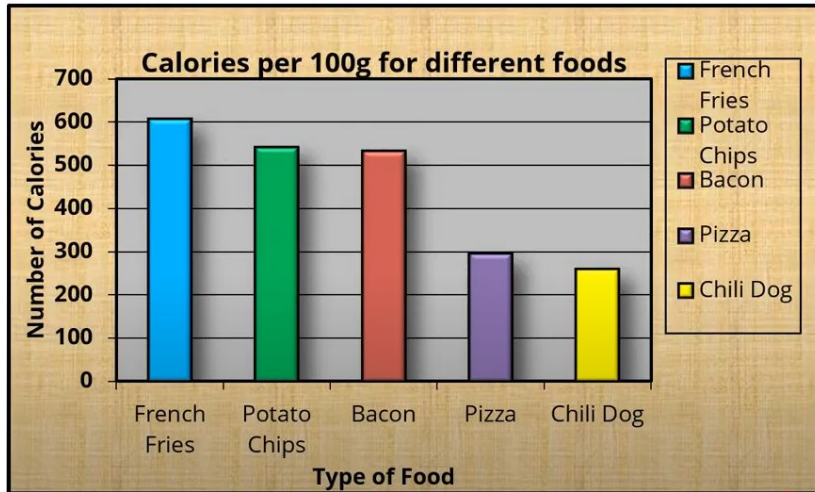
Exemplos



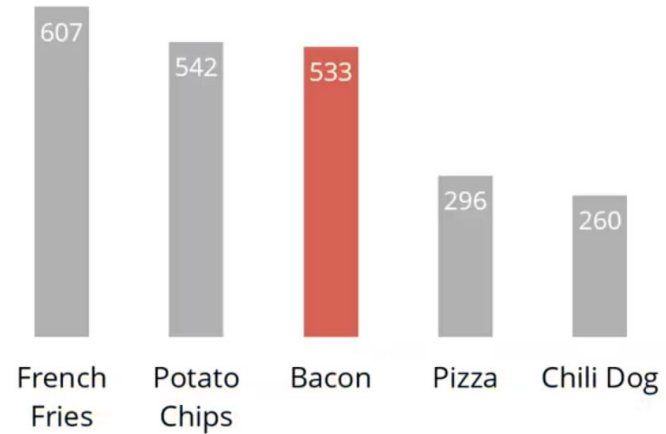
Pig Meat Preferences



Exemplos



Calories per 100g



Prática





Exercícios:

1. Atualize os gráficos no seu caderno de análise de forma a atender aos princípios de storytelling;
2. Revise a ordem da análise, das medidas estatísticas e gráficos apresentados, de forma a construir uma narrativa coerente para seus achados de análise;

Obrigado

Stefano Mozart

linkedin.com/in/stefano-mozart/

github.com/stefanomozart

