



UNIVERSIDADE FEDERAL DO PIAUÍ
CAMPUS SENADOR HELVÍDIO NUNES DE BARROS
BACHARELADO EM ADMINISTRAÇÃO
Administração de Sistemas de Informação

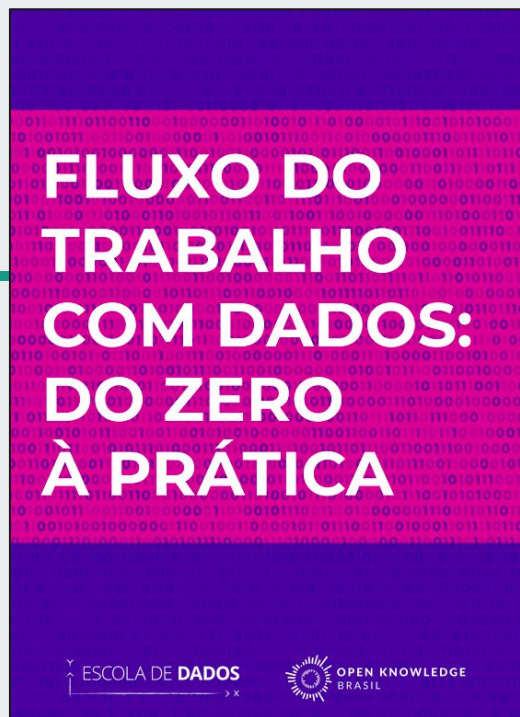
Contando histórias a partir de dados - Introdução

Prof. Me. Alexandre Ribeiro Cajazeira Ramos
E-mail: alexandre.cajamos@gmail.com
<https://cajazeiramos.github.io/>

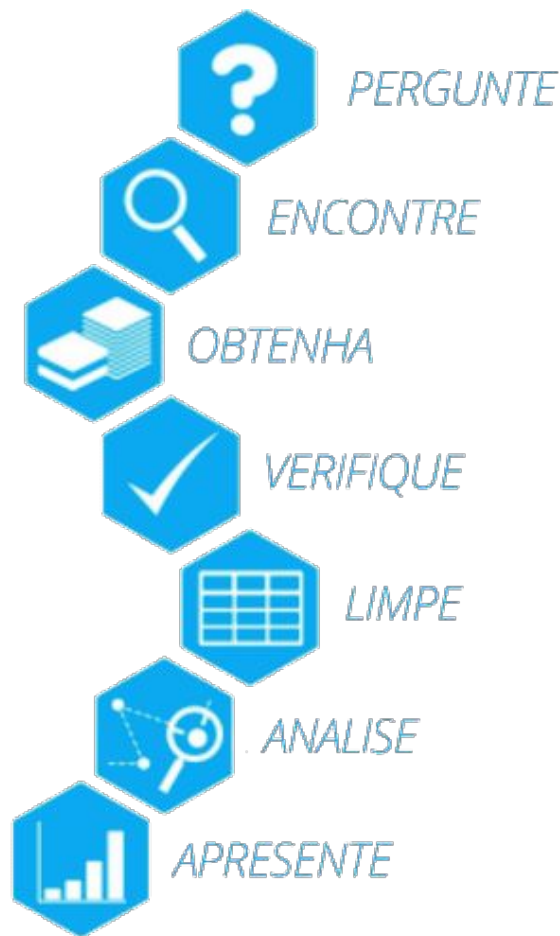


Cronograma de Aula

- Contando histórias a partir de dados (*data storytelling*)
- Principais etapas do Fluxo de trabalho com dados
- Como tudo isso se aplica à gestão de negócios?



Fluxo de Trabalho com Dados: Do Zero À Prática



A humanidade sempre produziu dados





Contando histórias com dados

- O uso, análise e visualização de dados são praticamente indissociáveis à experiência humana como conhecemos.
- Existem observações pré-históricas sobre o uso de dados
(*vale a leitura deste [trabalho detalhado sobre a história do campo](#))*
- Entretanto, a partir da consolidação de diferentes áreas do saber relacionadas à mensuração de fenômenos físicos, no século XVII, pode ser observado um ponto de inflexão da área.

O gráfico de Florence Nightingale (em 1858) sobre as mortes durante a Guerra da Crimeia

Hoje considerada mãe da enfermagem moderna, também compilou estatísticas sobre a causa de mortalidade dos combatentes

Através de um diagrama de área polar, comunica diretamente a informação, de que mais pessoas estavam morrendo por doenças que nos combates;

Demonstra visualmente, também, como a proporção de mortes por causas evitáveis era maior ao longo dos meses

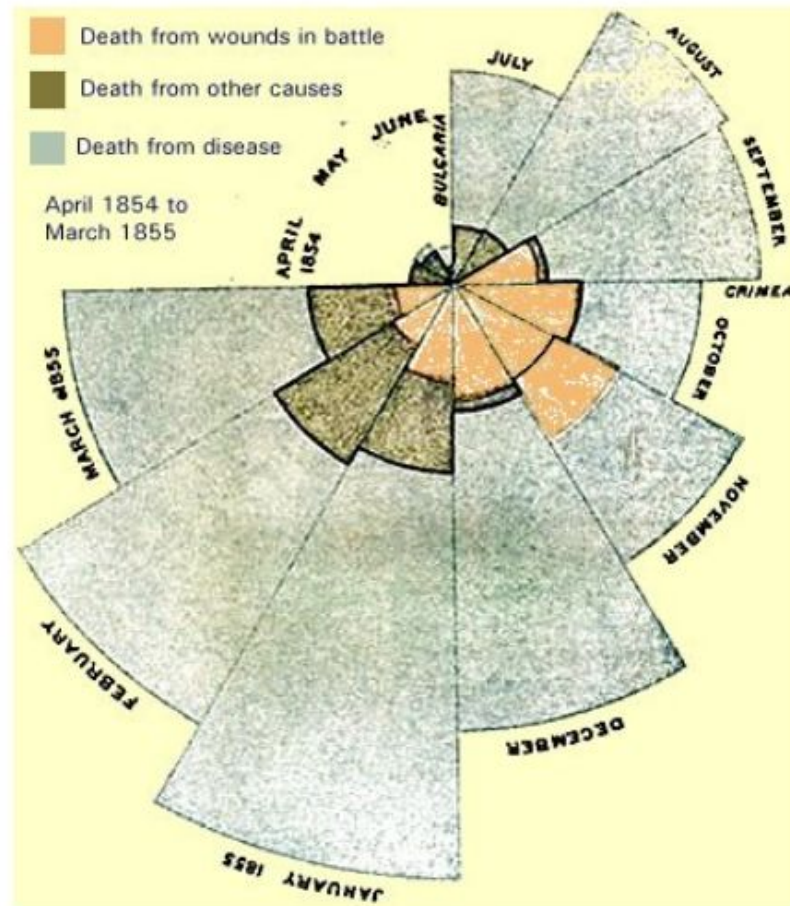


Figura 3: No gráfico, cada fatia representa um mês e a área coberta pelas cores indica as causas de mortalidade: em laranja, mortes em batalha; em azul, por doenças; em verde, outras causas. Fonte: University of Houston.

O Mapa da Cólera, de John Snow Inglaterra 1850

John Snow decidiu **levantar dados “geolocalizados”** sobre a doença, mapeando os casos e óbitos

Representou pontos no mapa a partir de casos da doença, descobrindo uma bomba d'água como vetor de disseminação de cólera no bairro de Soho.

A partir do estudo com dados, **Snow reforçou a teoria** de que a **água poderia ser um dos vetores de contaminação**, em contraposição à teoria dominante à época que supunha a transmissão por meio do ar.



Figura 4: Mapa da cólera de John Snow na Inglaterra (1854) - Fonte: Wikimedia Commons



Contando histórias com dados

- Com a pandemia da Covid, tornou-se ainda mais comum a comunicação através de gráficos, mapas e outras visualizações.
- A análise e visualização de dados passou à ordem do dia nas notícias e informativos relacionados à pandemia
- Mas *Data Storytelling* está presente nos ambientes de pesquisa e de gestão de negócios há muito tempo.



O que há de novo?

- Internet e a sociedade do Big Data
 - Parte significativa das relações humanas moldadas por dados, incluindo as relações de consumo
- Crescimento do debate sobre transparência pública, ética no uso de dados e os impactos da Inteligência Artificial



O que há de novo?

- “Oceanos” de bases de dados disponíveis
- Um ecossistema de ferramentas e tecnologias robustas e acessíveis;
- Demanda crescente por uma comunicação guiada por dados, em diferentes áreas de atuação.



Visualização de dados moderna

- **Transparência** - As bases de dados utilizadas estão disponíveis?
- **Reprodutibilidade** - Outros jornalistas poderão, facilmente, reproduzir o seu método? Terão o mesmo resultado?
- **Credibilidade** - Qual a fonte do dado? Como foram coletados? Quais análises foram realizadas?

Etapas do Fluxo de Trabalho com Dados





Considerações iniciais

- O fluxo de trabalho com dados não é, necessariamente, sequencial
 - A visualização pode nos remeter à definição do problema, por exemplo
- Vamos conhecer, nesta ordem, a Definição, Obtenção, Checagem, Limpeza, Análise e Visualização de dados

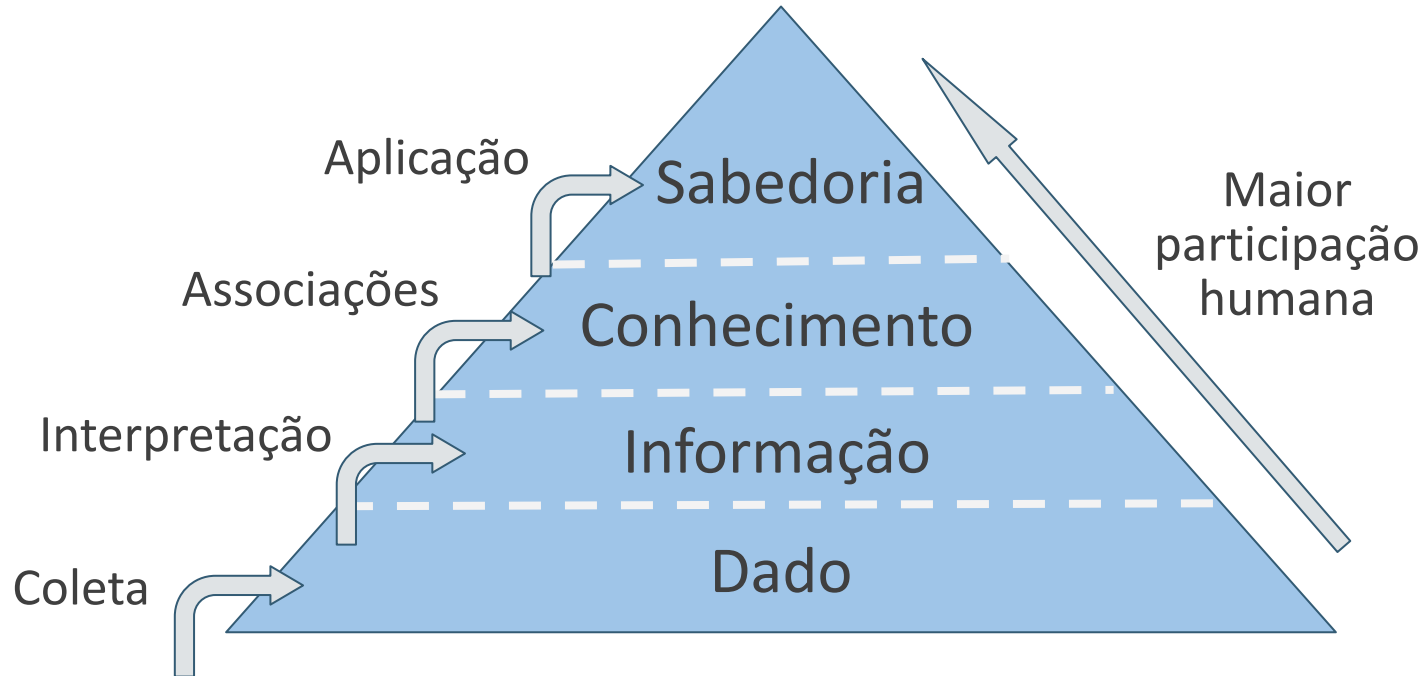


Considerações iniciais

- Antes de começar...
 - O que é um **dado**?
 - Qual a diferença entre Dado e Informação? E entre Informação e Conhecimento?
 - Como isso se aplica aos problemas reais?

Pirâmide DKIW

(Data-Information-Knowledge-Wisdom)





Definição

- “Na prática, tudo começa com os dados ou um tema”
 - Qual será a nossa história guiada por dados?
- É possível abordar qualquer assunto com dados
 - Boas perguntas
 - Escopo bem delimitado
 - Conhecimento dos principais tipos de conjuntos de dados



Definição

- Quantas casas em Picos ficaram sem água na última semana?
- Qual produto tem o melhor rendimento em minha rede de lojas?
- Qual o público-alvo correto para nossa próxima campanha?



Definição

- **Objetividade** -> Clareza e concisão ajuda muito!
- **Pode ser abordado com dados** -> Se não se baseia em dados...
- **Complexidade** -> Objetividade não implica redução da complexidade do tema;
- **Viabilidade** -> Será possível levantar, analisar e comunicar dados deste tema?
- **Originalidade** -> Refazer análises é legal para aprender, mas para apresentar...
- **Relevância** -> Qual o seu público? Por que é relevante para ele?



Obtenção

- Nesta etapa buscamos **ter em mãos um ou mais conjuntos de dados** para dar sustentação à narrativa.
- Existem, basicamente, três maneiras de obtenção de dados:
 - Coleta primária (você constrói o seu dataset ou recebe da empresa)
 - Bases disponíveis na web
 - Dados públicos via Lei de Acesso à Informação (LAI)



Obtenção - Principais tipos de dados

- **Dados tabulares** - Linhas e colunas como abstrações da realidade;
- **Dados espaciais** - Mapas, coordenadas e suas peculiaridades;
- **Redes ou grafos** - Representam, em geral, hierarquias e conexões;
- **Textos** - Nuvens de palavras e NLP ;
- **Imagens** - Extração de características, análises de emoções e outros.



Checagem

- Quem coletou os dados e por que os coletou?
 - Existe conflito de interesse? Você consegue mapear algum viés?
- Como os dados foram coletados?
 - A metodologia é fundamental para compreender o resultado.
- Quando os dados foram coletados ou atualizados?
 - Os dados retratam que momento da realidade?



Checagem

- Entender o que significa cada linha e cada coluna do *dataset* é fundamental:
 - **Compleitude e coerência** - A falta de registros deve ser imediatamente verificada; Pode um dataset do orçamento/UF sem o PI?
 - **Valores anormais** - Outliers são muito especiais ou erros de digitação!
 - **Unidades de medida** - Já pensou comparar Real com Dólar hoje?



Limpeza

- Com dados devidamente verificados, vamos agora deixá-los prontos para Operações matemáticas, agrupamentos, verificações automatizadas e outras formas de análises.
 - **Lembrete:** Tabelas são frutos do trabalho humano e estão repletas de erros humanos. A limpeza envolve remoção de inconsistências, duplicação, erros de digitação, dentre outros.
 - **Vale pesquisar:** Dados Organizados (tidy data) / Hadley Wickham



Análise de Dados

- A análise é a etapa em que vamos **explorar o conjunto de dados**, conhecer seus fenômenos e abstrações, extrair informações relevantes e definir uma narrativa.
- “O maior risco de erro vem de onde menos prestamos atenção: nós mesmos. Nossos próprios vieses e eventuais falhas nas premissas que adotamos na hora de analisar um banco de dados são um estoque inesgotável de cascas de banana.”



Análise de Dados - Como fazer?

- **Desenvolva um método:** “O que exatamente você espera encontrar? De que maneira vai sistematizar essa busca? Que evidências seriam suficientes para corroborar suas hipóteses?”
- **Reporte contra suas convicções:** Procure, nos dados, evidências que possam derrubar sua matéria e suas hipóteses.
- **Seja transparente:** Um método claro não só te ajuda a entender seus processos, como também garante a reprodutibilidade da sua análise.



Análise de Dados - Ferramentas

Interfaces gráficas



Google Sheets



Programação





Visualização de Dados

- Visualização é uma área dedicada à geração de imagens que auxiliam seus usuários na compreensão de dados e processos. // [Tamara MUNZNER](#)
- Visualização é qualquer tipo de representação visual de informação projetada para permitir comunicação, análise, descoberta, exploração, etc. // [Alberto CAIRO](#)



Por que Visualizar?

- Informações visuais são assimiladas, mesmo inconscientemente, a uma velocidade impressionante;
- Nossos olhos são extremamente sensíveis a variações de cor, formato, padrão, dentre outras;
- Utilizando técnicas e ferramentas de visualizações, podemos tornar nossa comunicação verdadeiramente efetiva.

Por que Visualizar?



Gráficos também são usados para desinformação!

Esse material do Nexa ([Como mentir com gráficos](#)) vale muito a leitura!

Por que Visualizar?

Tente, rapidamente identificar os números “5” no texto



12267079846923893248154688687921977136690980879183
12347930860471823335206799160482347872697703325234
81641397066517642409397139019919668245212413643027
70469324621326845296437186289608878982849136307346
09384758302094857398539483820203987970921582928392
92020948928345877829388379209374582892938294978398
01039384759938382038429028757892084702034928347937
01928374639973481383477374598348932983492827636021

Por que Visualizar?

Tente, rapidamente identificar os números “5” no texto



12267079846923893248154688687921977136690980879183
12347930860471823335206799160482347872697703325234
81641397066517642409397139019919668245212413643027
70469324621326845296437186289608878982849136307346
09384758302094857398539483820203987970921582928392
92020948928345877829388379209374582892938294978398
01039384759938382038429028757892084702034928347937
01928374639973481383477374598348932983492827636021

Por que Visualizar?

Tente, rapidamente identificar os números “5” no texto

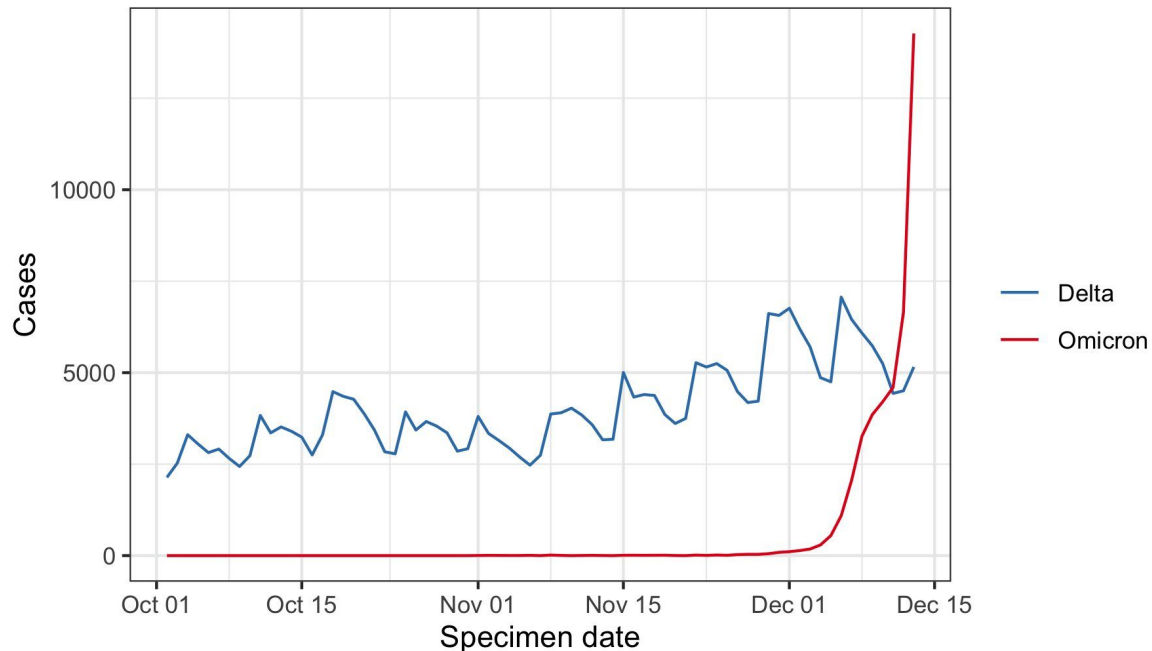


12267079846923893248154688687921977136690980879183
12347930860471823335206799160482347872697703325234
81641397066517642409397139019919668245212413643027
70469324621326845296437186289608878982849136307346
09384758302094857398539483820203987970921582928392
92020948928345877829388379209374582892938294978398
01039384759938382038429028757892084702034928347937
01928374639973481383477374598348932983492827636021

Por que Visualizar?



Cases in London by specimen date and variant



Data: UKHSA. SGTF proportions from daily update combined with case numbers from coronavirus.data.gov.uk.

Por que Visualizar?

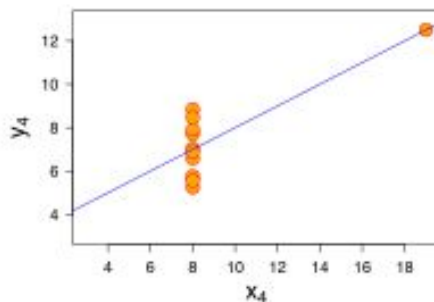
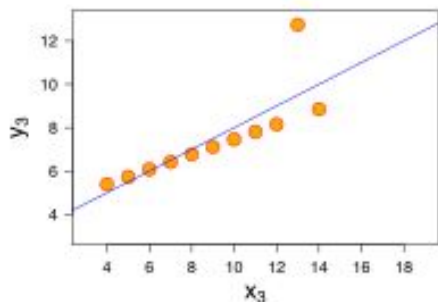
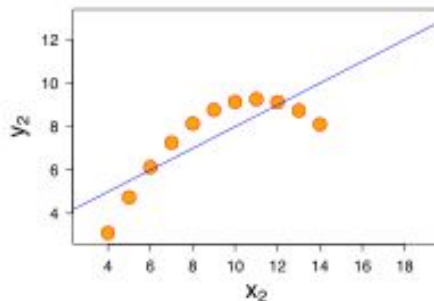
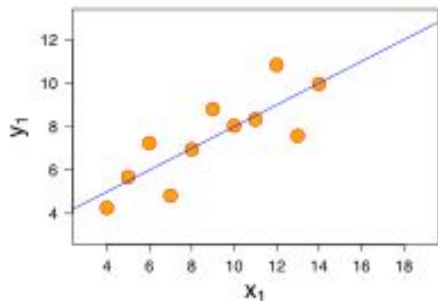


___ O que você consegue inferir sobre estes dados?

x1	y1	x2	y2	x3	y3	x4	y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Quarteto de Anscombe -> Quatro conjuntos com **estatísticas descritivas quase idênticas** (Média, variância, Correlação, etc).

Por que Visualizar?



Quarteto de Anscombe - >

Quatro conjuntos com **estatísticas descritivas quase idênticas** (Média, variância, Correlação, etc).

Etapas do Fluxo de Trabalho com Dados

Dúvidas

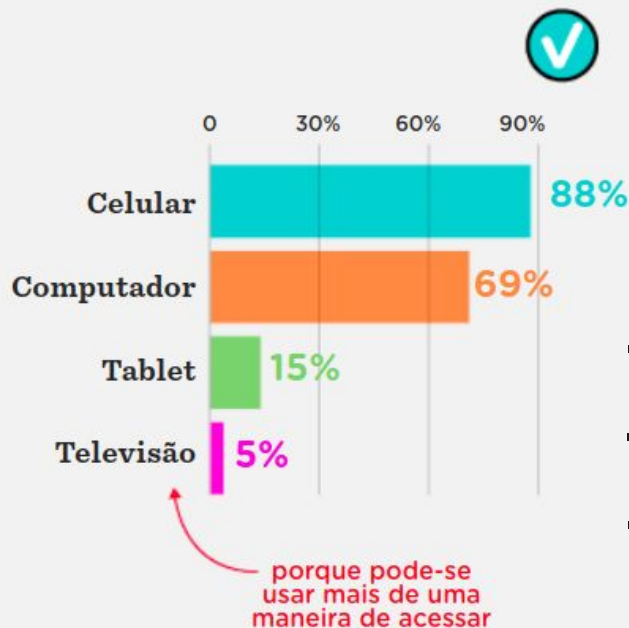
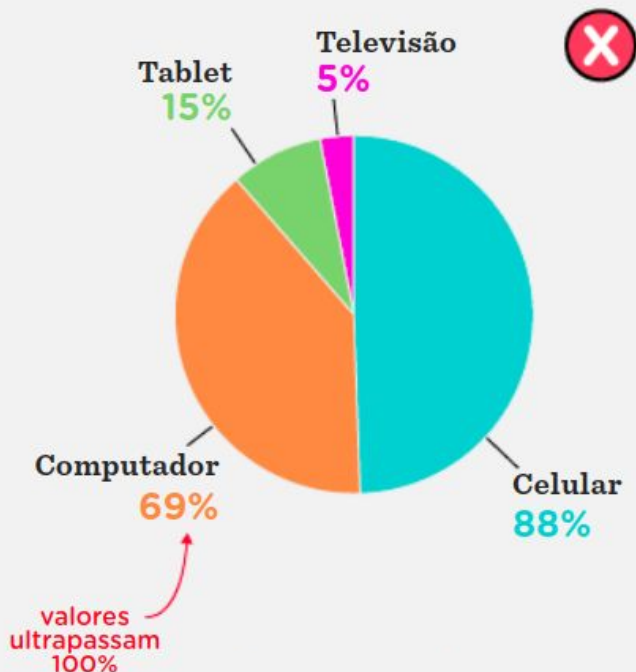


Por que estudar análise e visualização de dados?



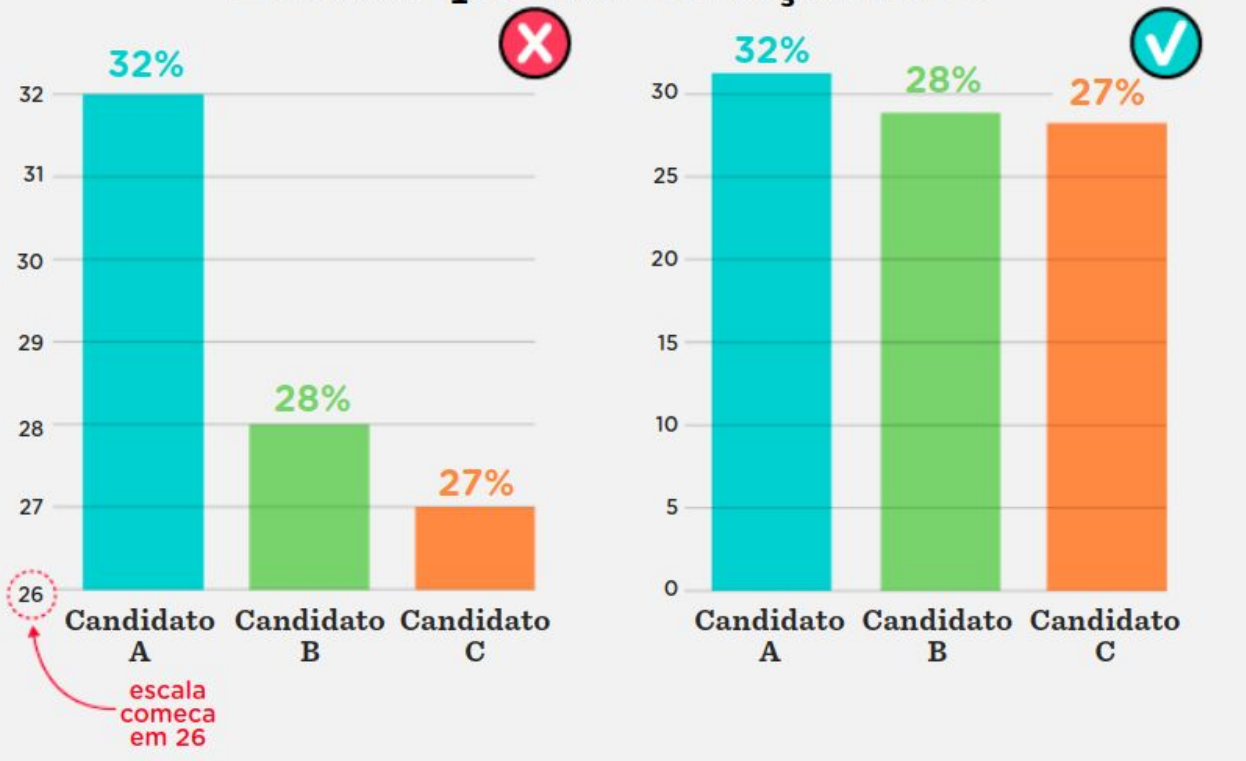
- Compreender melhor o mundo ao nosso redor
- Não ser enganado com gráficos e insights inconsistentes
- Conseguir produzir conhecimento a partir de dados
- Se destacar profissionalmente

Contas que não fecham



Como mentir
com gráficos
(Nexo)

Escalas que não começam no 0



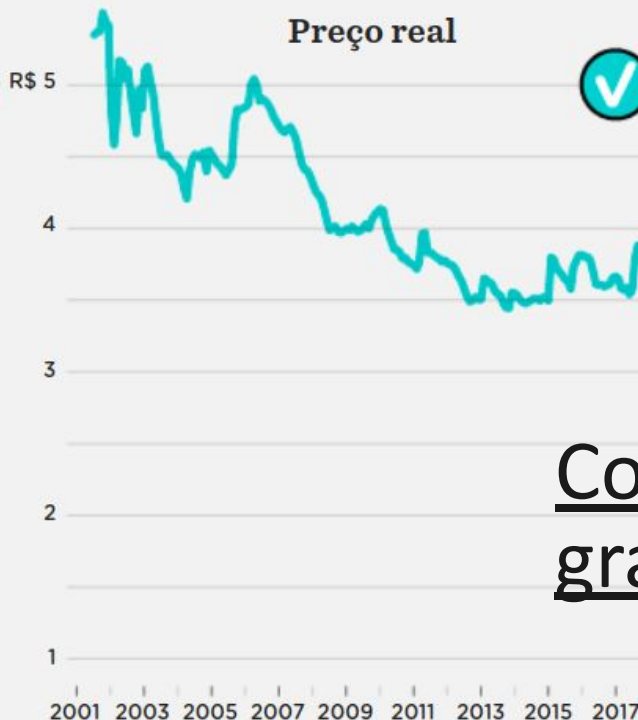
Como mentir com gráficos (Nexo)

Valores não deflacionados

Preço nominal



Preço real

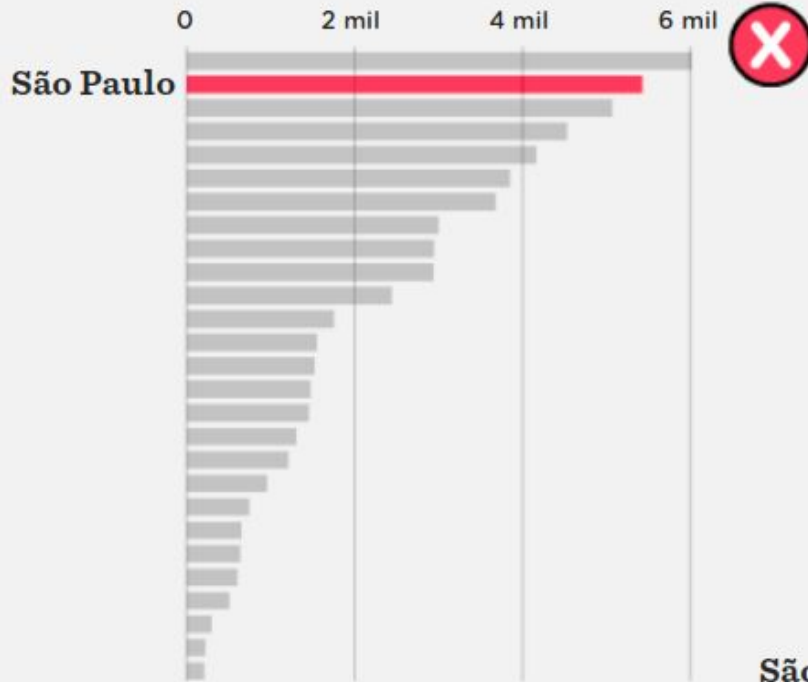


Como mentir com gráficos (Nexo)

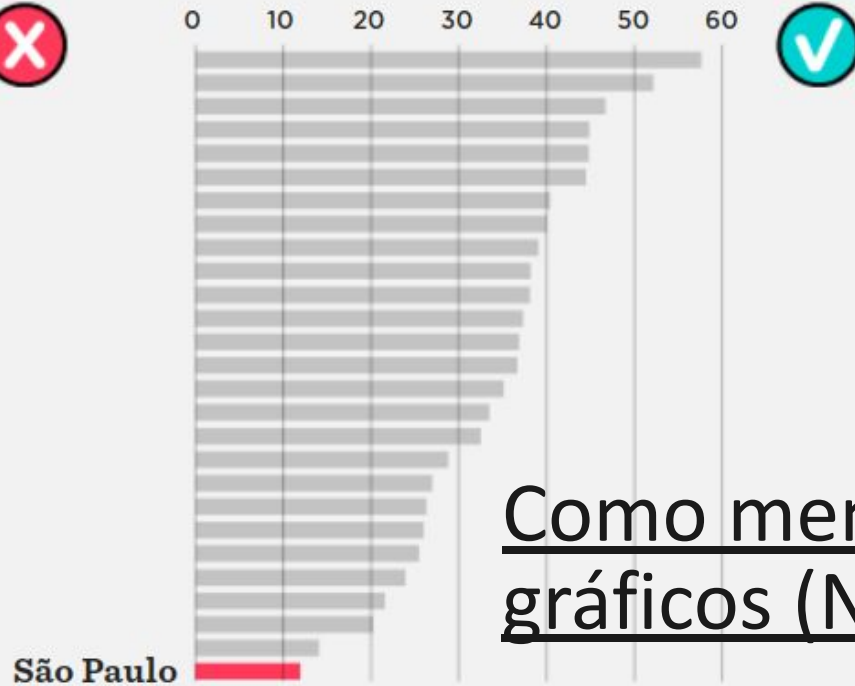
Dados absolutos x Taxas



“São Paulo é o segundo mais violento”



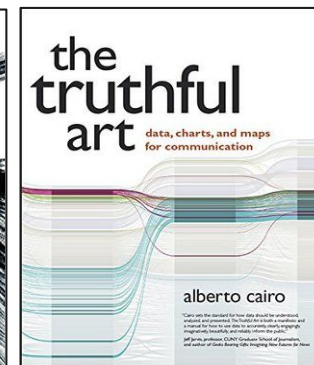
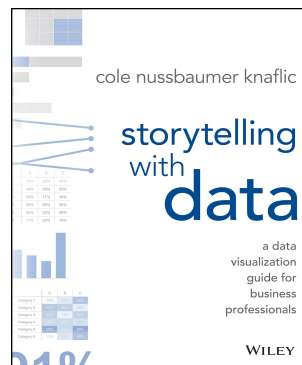
“São Paulo é o menos violento”



Como mentir com gráficos (Nexo)

Material complementar

- [Guia prático da visualização de dados: definição, exemplos e recursos de aprendizado](#)
- [Tutorial STORYTELLING COM DADOS: ESTRUTURAS, TÉCNICAS E MÃO NA MASSA COM FLOURISH](#)
- [Como mentir com gráficos: 7 detalhes que podem te enganar](#)
- [Conceitos básicos de Visualização](#)
[\(Prof. Adolfo Guimarães\)](#)

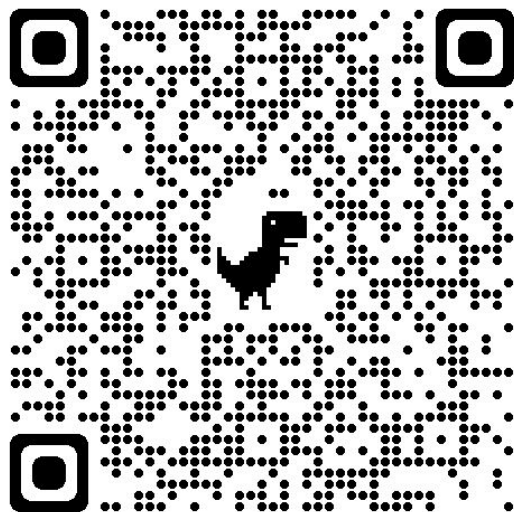




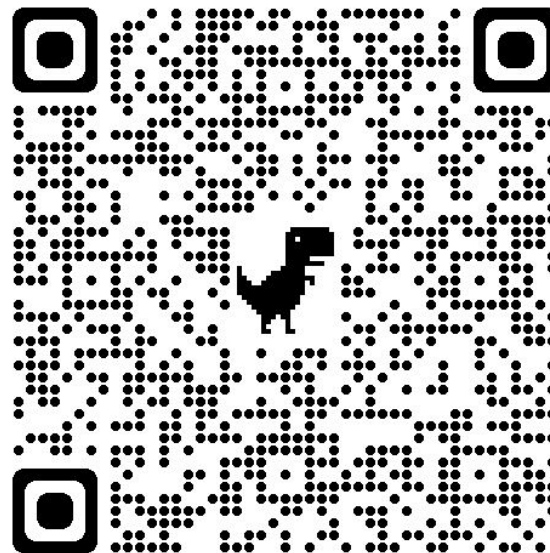
Atividade (para próxima aula)

- Relacione os conceitos explorados com a gestão de negócios em um texto argumentativo-dissertativo, que considere questões como:
 - Como o universo dos dados interfere em nossas relações em sociedade, em especial nos negócios?
 - Como Data Storytelling se aplica à gestão de negócios? Quais os seus benefícios?
 - Por que conhecer e explorar o fluxo de trabalho com dados na atualidade?

Acesso ao material



Este slide



Atividade