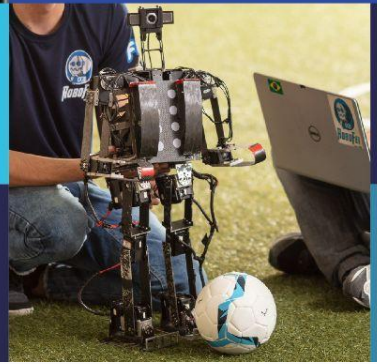




CCP010

Digital Experience




Visualização de Dados

Visualização de Dados

- A visualização de dados consiste na apresentação de informações através de elementos visuais, como ***gráficos***, ***diagramas***, ***mapas***, ***tabelas***, entre outros
  - Por meio dela, fica muito mais fácil analisar os resultados, auxiliando o processo de identificar padrões e tendências e tomar decisões
  - Boas visualizações de dados ajudam a descomplicar as informações, transmitindo uma mensagem clara e objetiva

Fonte: https://www.neomove.com.br/data-visualization/visualizacao-de-dados-entenda-o-que-e-e-qual-a-importancia/

Visualização de Dados

Benefícios

- Com o enorme volume de informações disponíveis, as ferramentas de visualização de dados são cada vez mais fundamentais para as empresas
- Benefícios:
  - Apresentação mais envolvente / ***Absorção rápida das informações***
  - Facilita a tomada de decisão
  - Auxilia a encontrar conexões importantes para solucionar os problemas da sua empresa

Fonte: https://www.neomove.com.br/data-visualization/visualizacao-de-dados-entenda-o-que-e-e-qual-a-importancia/

Gráficos enganosos

Gráfico errado

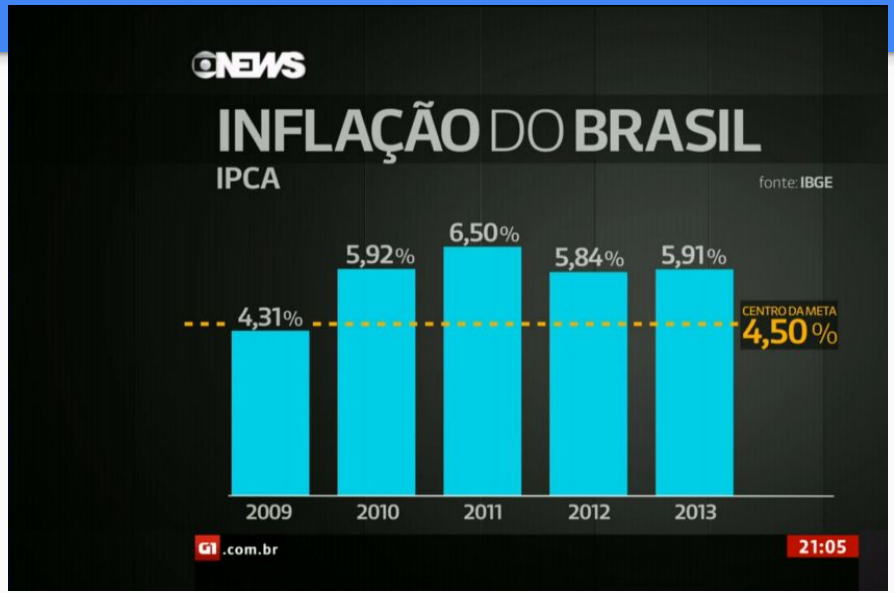
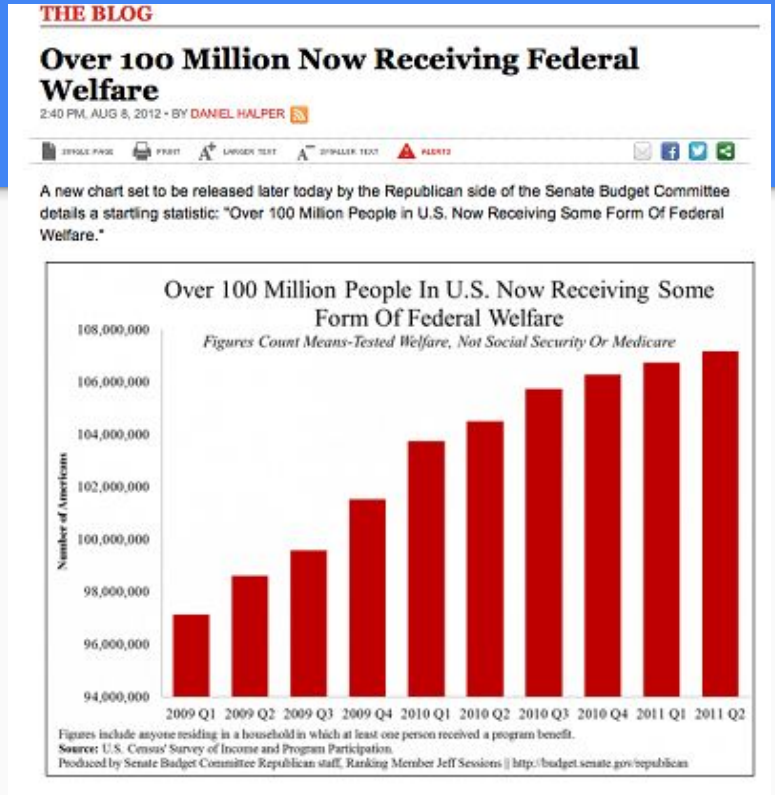


Gráfico corrigido

Fonte: https://gizmodo.uol.com.br/mentir-visualizacao-dados/

Gráficos enganosos

- USA Today - Este gráfico permite a interpretação de que a assistência social federal está fora de controle



THE BLOG

Over 100 Million Now Receiving Federal Welfare

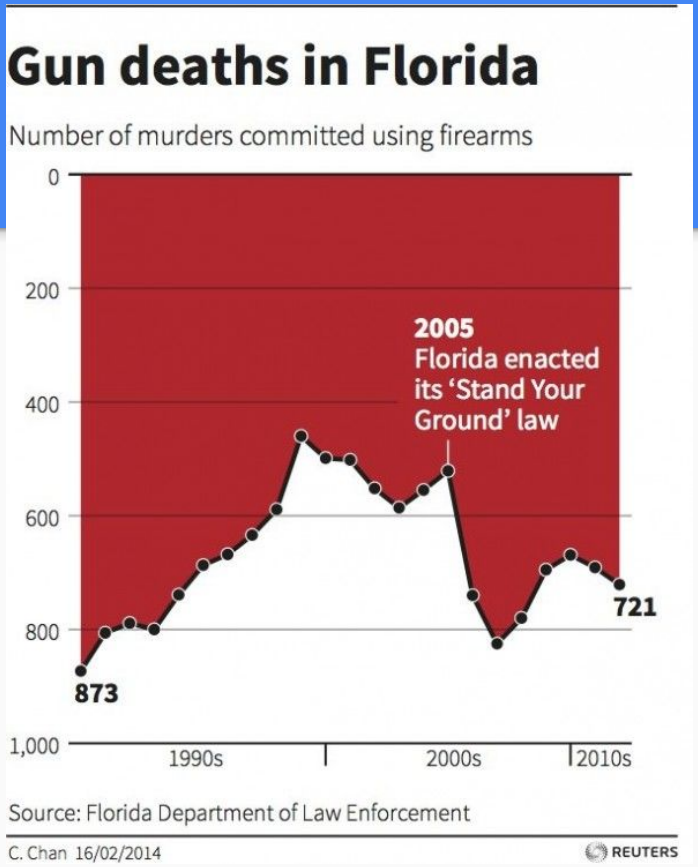
2:40 PM, AUG 8, 2012 • BY DANIEL HALPER

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."

Fonte: https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/misleading-graphs/

Gráficos enganosos

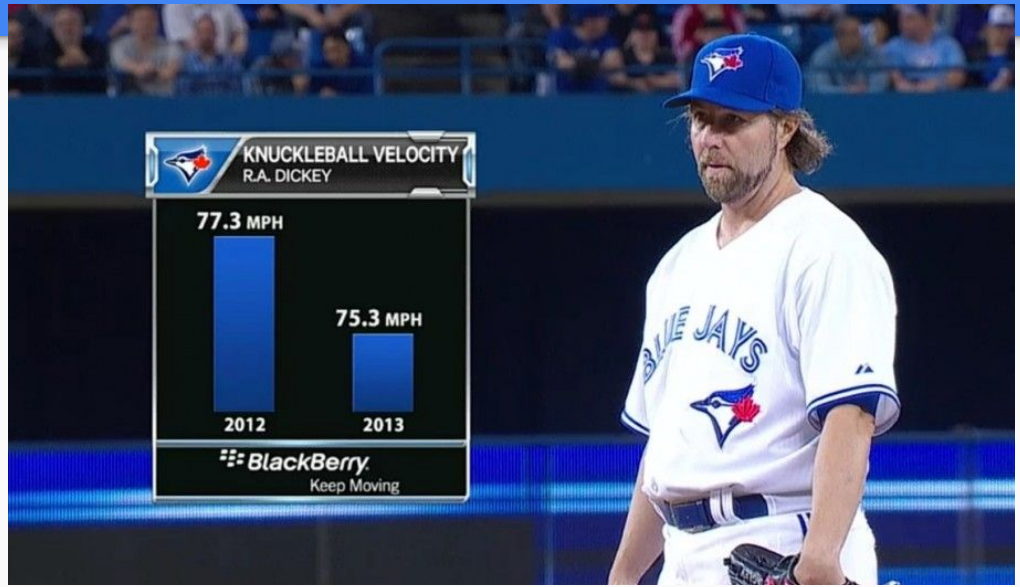
- Reuters - O número de mortes por arma de fogo teve um aumento em 2005 na Florida?



Fonte: https://gizmodo.uol.com.br/mentir-visualizacao-dados/

Gráficos enganosos

- Dickey perdeu muita potência de lançamento de 2012 para 2013?



Fonte: https://gizmodo.uol.com.br/mentir-visualizacao-dados/

Tipos Principais de Gráficos

Gráfico de barras (colunas)

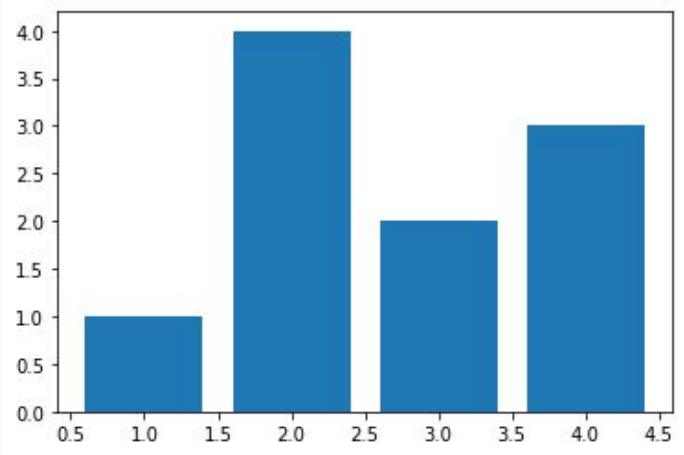
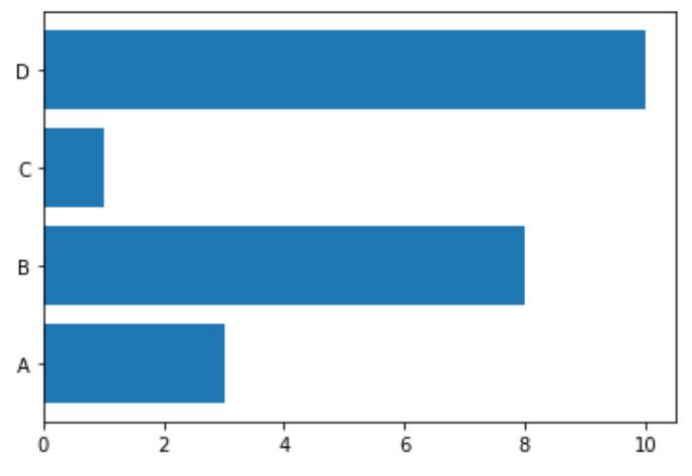


Gráfico de barras horizontais



Tipos Principais de Gráficos

Gráfico de barras empilhadas

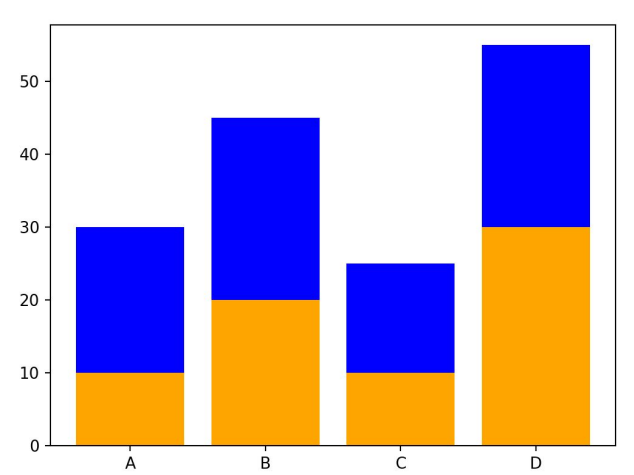
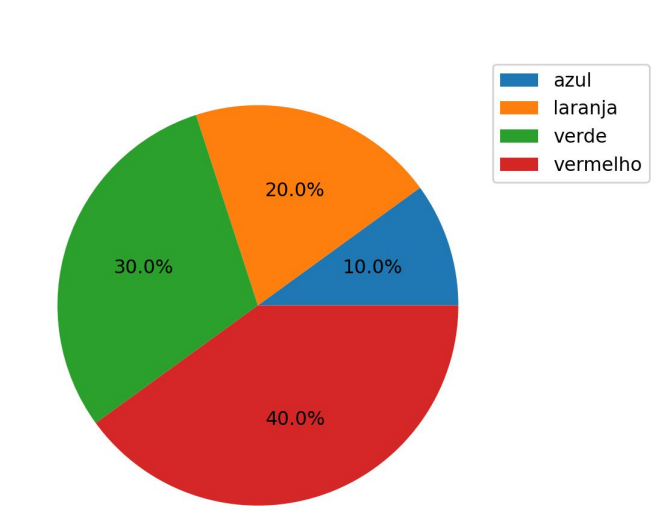


Gráfico de pizza



Tipos Principais de Gráficos

Gráfico de dispersão (pontos)

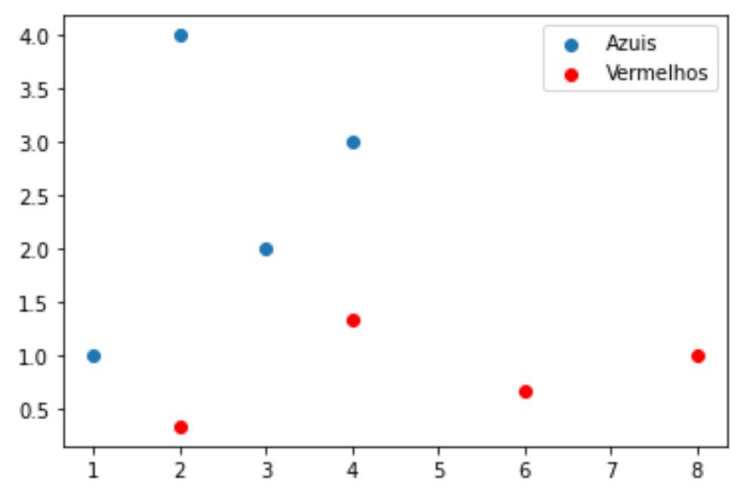
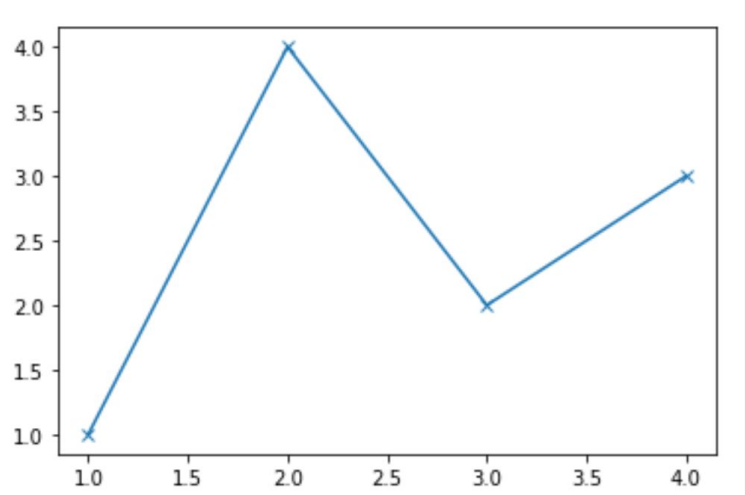


Gráfico de linha



Limpeza de Dados

Por que realizar limpeza?

- Dados incompletos / faltantes
- Dados redundantes / duplicados
- Dados inconsistentes
- Dados com ruídos

Dados incompletos / faltantes

Limpeza de Dados

Dados incompletos / faltantes

Dados que deveriam estar disponíveis mas não estão presentes para certas observações ou variáveis

Razão	Gravidade	Significado
MCAR: *missing completely at random*	Baixa	A probabilidade de um dado estar ausente é **igual** para todos os indivíduos na amostra e não está relacionado com o valor real da variável
MAR*: missing at random*	Média	A probabilidade de um dado estar faltando pode estar relacionada **com outras** variáveis observadas na amostra.
MNAR: *missing not at random*	Grave	A probabilidade de um dado estar faltando pode estar relacionada **com o valor** da variável observada além da relação **com outras** variáveis

Limpeza de Dados

Dados incompletos / faltantes

Estratégia de solução

Estratégia	Observações
Exclusão de registro	Para casos MCAR e MAR. E por que não MNAR?
Exclusão de coluna	Todos os casos. Quando a coluna não é importante
Média, Mediana e Moda	Para casos MCAR e MAR
Imputação por regressão ou múltipla	Para MNAR.

Dados redundantes / duplicados

Limpeza de Dados

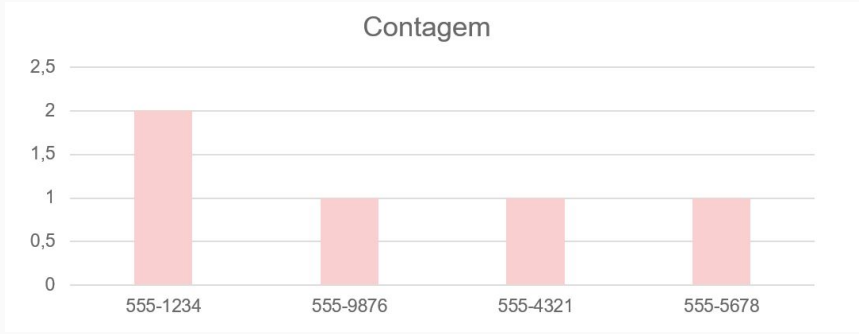
Dados duplicados / redundantes

Ocorrência de registros idênticos ou muito semelhantes em um conjunto de dados. Isso pode ser um problema comum em conjuntos de dados e requer tratamento adequado

Limpeza de Dados

Dados duplicados / redundantes

ID	Nome	Email	Telefone
1	João Silva	joao@example.com	555-1234
2	Maria Santos	maria@example.com	555-9876
3	Pedro Alves	pedro@example.com	555-4321
4	João Silva	joao@example.com	555-1234
5	Ana Oliveira	ana@example.com	555-5678



Limpeza de Dados

Dados duplicados / redundantes

- ***Tipos***
  - erros de medição
  - interferência de fontes externas ou internas
- ***Identificação***
  - Identificação manual
  - Análise de frequência
  - Clustering
- ***Tratamento***
  - Ignorar
  - Eliminar registro/coluna
  - Consolidar/fundir

Dados inconsistentes

Limpeza de Dados

Dados inconsistentes

- ***Tipos***
  - Valores inválidos: dados que **não estão** dentro dos limites permitidos ou esperados para uma determinada variável
  - Condições lógicas: dados que apresentam **informações conflitantes** dentro do conjunto de dados
- ***Identificação***
  - Análise estatística: outliers
  - Identificação manual: conhecimento de domínio
- ***Tratamento***
  - Ignorar
  - Eliminar registro ou coluna
  - Correção manual

Ruído

Limpeza de Dados

Dados inconsistentes

Erro aleatório, variabilidade ou interferência aleatória que possa estar presente nos dados e afetar a precisão das medidas e das análises

- ***Tipos***
  - erros de medição
  - interferência de fontes externas ou internas
  - outliers
  - aleatórios
- ***Identificação***
  - Identificação manual
  - Análise descritiva*
- ***Tratamento***
  - Ignorar
  - Eliminar registro
  - Normalização/padronização
  - Suavização

Revisão Estatística

Estatística

Média

- ***Média:*** demonstra a **concentração dos dados** de uma distribuição (variáveis numéricas)

Professor	Altura (cm)	Peso (kg)
1	165	68
2	172	*NaN*
3	*NaN*	60
4	180	75
5	158	*NaN*

Altura: 168.75

Peso: 67.67

Estatística

Média

- Podemos usar a média como ***estratégia de imputação***

Professor	Altura (cm)	Peso (kg)
1	165	68
2	172	*NaN*
3	*NaN*	60
4	180	75
5	158	*NaN*

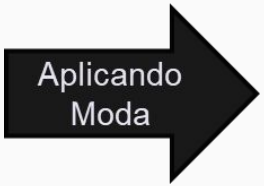
Professor	Altura (cm)	Peso (kg)
1	165	68
2	172	**67.67**
3	**168.75**	60
4	180	75
5	158	**67.67**

Estatística

Moda

- ***Moda:*** **dado mais frequente** (variáveis categóricas)

Professor	Estado civil
1	Casado
2	Casado
3	*NaN*
4	Solteiro
5	Casado



Professor	Estado civil
1	Casado
2	Casado
3	**Casado**
4	Solteiro
5	Casado

Estatística

Mediana

- ***Mediana:*** indica qual é o valor que está **exatamente no meio** de um conjunto de dados (variáveis numéricas)

Aluno	Nota
1	1
2	10
3	*NaN*
4	8
5	7
6	6

{1,6,**7**,8,10}

Aplicando Mediana

Aluno	Nota
1	1
2	10
3	**7**
4	8
5	7
6	6

Estatística

Desvio Padrão

- ***Desvio Padrão:*** medida de **dispersão em torno da média**
- Como os dados estão distribuídos ao redor da média?

Aluno	Nota
1	4.5
2	5.5
3	5.0
4	6.5
5	7.0
6	6.0

Média: 5.75

Desvio Padrão: 0.854

Estatística

Desvio Padrão

- ***Desvio Padrão:*** medida de **dispersão em torno da média**
- Como os dados estão distribuídos ao redor da média?

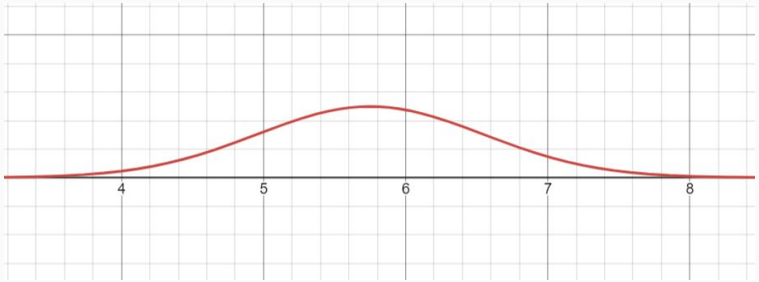
Aluno	Nota
1	1.5
2	9.0
3	9.5
4	2.5
5	10.0
6	2.0

Média: 5.75

Desvio Padrão: 3.772

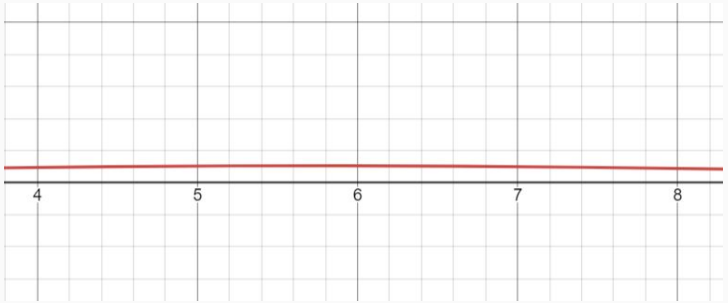
Estatística

Desvio Padrão



Média: 5.75

Desvio Padrão: 0.854



Média: 5.75

Desvio Padrão: 3.772

https://www.desmos.com/calculator/0x3rpqtgrx?lang=pt-BR

Análise de *Dataset:*
Sustainable Development Goals (SDG)

Indicadores relevantes extraídos dos Indicadores de Desenvolvimento Mundial, reorganizados de acordo com os objetivos e metas dos ***Objetivos de Desenvolvimento Sustentável (ODS)***. Esses indicadores podem ajudar a monitorar os ODS *(SDGs)*, mas nem sempre são os indicadores oficiais para monitoramento dos ODS.

Fonte: https://www.kaggle.com/datasets/theworldbank/sustainable-development-goals

Objetivos de Desenvolvimento Sustentável (ODS)

17 objetivos ambiciosos e interconectados que abordam os principais desafios de desenvolvimento enfrentados por pessoas no Brasil e no mundo

Os Objetivos de Desenvolvimento Sustentável são um apelo global à ação para acabar com a pobreza, proteger o meio ambiente e o clima e garantir que as pessoas, em todos os lugares, possam desfrutar de paz e de prosperidade.

https://brasil.un.org/pt-br/sdgs

Para próxima aula - Apresentação Projeto1

- Atividade em dupla
- Buscar um dataset ou reutilizar o já utilizado com dados relacionados à alguma ODS
- Apresentar para sala o dataset encontrado, explicitando:
  - sobre o que é o dataset
  - a quantidade de amostras
  - quantidade de atributos (*features*)
  - apresentar os dados quantitativos e qualitativos
  - **Apresentar gráficos e resumos estatísticos**
  - **Apresentar conclusões obtidas pelo estudo dos dados (gráficos e estatística)**
  - **Mostrar possíveis soluções sobre os problemas relatados na conclusão**
  - Fazer tudo utilizando Excel e Powerpoint
  - **Apresentação de 5 minutos**