

Café com Dados

O Workshop do DataLab

DataLab

[Back to Agenda](#)

O DataLab

**Subárea da UGIP, focada em construir
produtos baseados em Dados.**

Time:



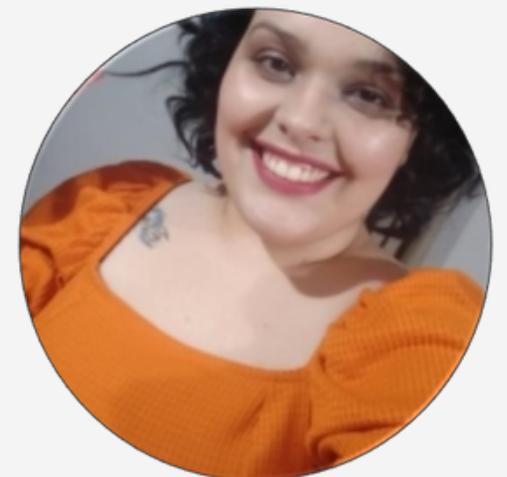
João



Maurício



Ray



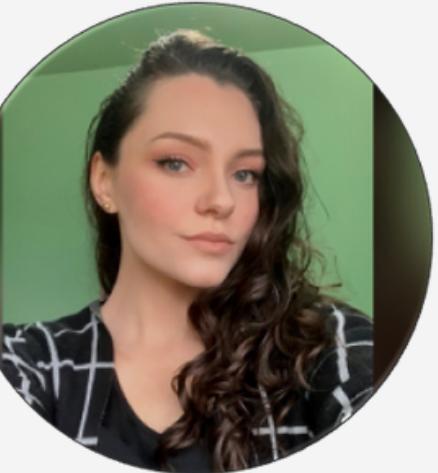
Carol Selis

Cientista de Dados

Formada em Estatística pela UFPR

MBA Data Science e Analytics pela USP/ESALQ

Fã do Pedro Pascal



Bianca

Product Designer

Forma(n)da em Design Gráfico pela UTFPR

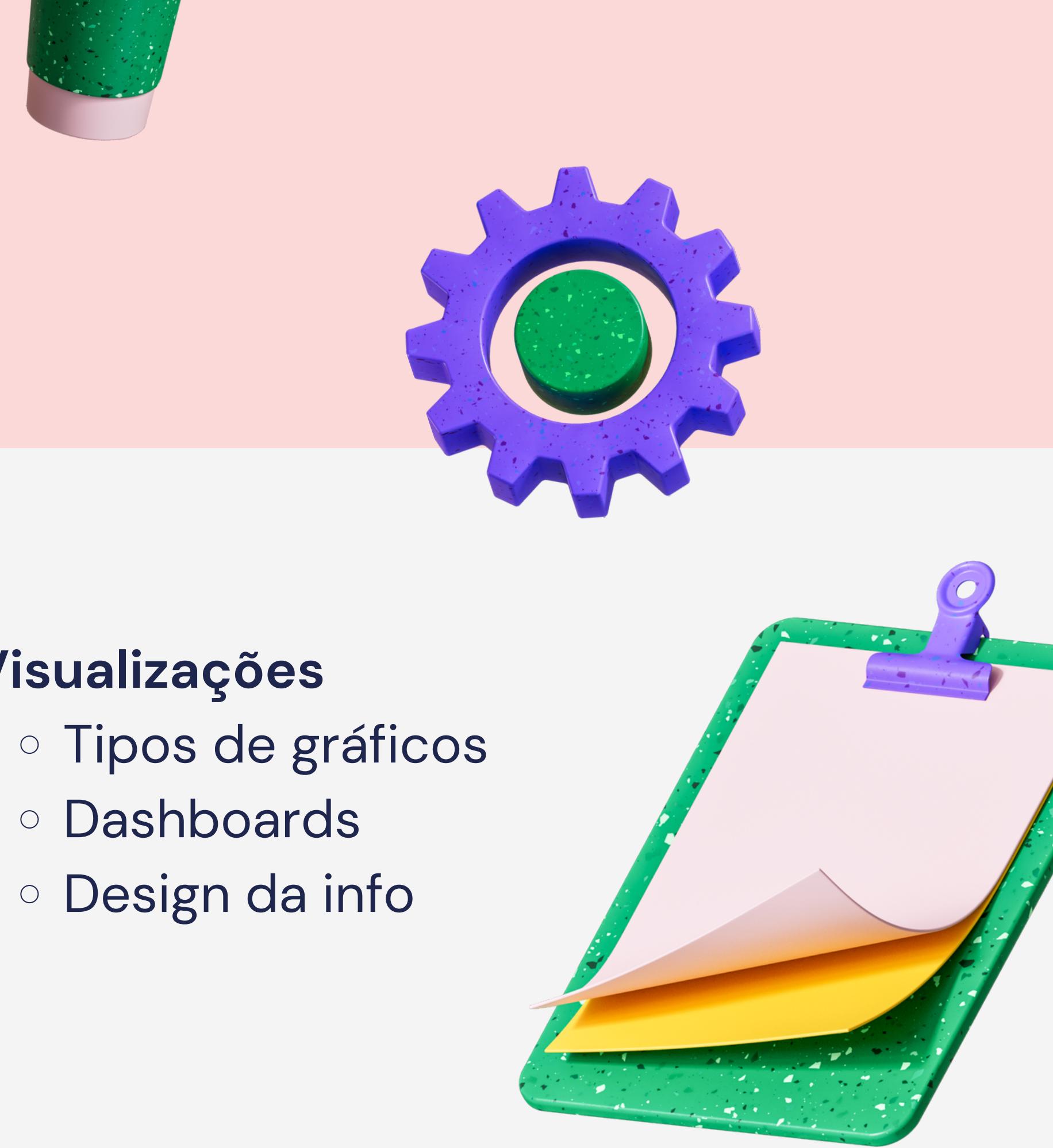
Formação em desenvolvimento Web

não sabe quem é Pedro Pascal



Agenda

- Dados
 - O que são Dados?
 - Etapas dos Dados
- Análises
 - Descritiva
 - Correlação
 - Modelos
- Visualizações
 - Tipos de gráficos
 - Dashboards
 - Design da info





O que são Dados?

Dados são uma coleção de fatos, assim como números, palavras, medições, observações, descrições...

[Back to Agenda](#)

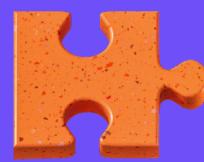
O que são Dados?

[Back to Agenda](#)



Dados Quantitativos

São as informações numéricas e contáveis, e são divididos em duas categorias:



Discreto

São valores que podem ser contados



Contínuo

São valores de mensuração



Dados Qualitativos

Atribuem uma característica a um determinado objeto



Quantitativo

Qualitativo

Discreto	Contínuo	Qualitativa
<ul style="list-style-type: none">• Número de pessoas em uma sala• Quantidade de computadores no SEBRAE• Resultados ao jogar um dado	<ul style="list-style-type: none">• Tempo entre casa e trabalho• Área construída do SEBRAE• Peso de um elefante	<ul style="list-style-type: none">• Características de um país• Cores do arco-íris• Material de uma mochila



[Back to Agenda](#)

Atividade 1



5 minutos

Escolha em objeto qualquer e liste até 5 dados do mesmo.
Depois classifique seus dados em Qualitativo, Quantitativo
Discreto, ou Quantitativo Contínuo.



[Back to Agenda](#)

Etapas do Dado

Antes de realizar uma análise, o dado passa por algumas etapas até estar pronto para ser estudado.



0

Problema e Hipótese

Objetivo e solução do problema.

Qualquer pesquisa/Estudo se inicia com a colocação de um problema, ou uma dúvida. A hipótese é criada para solucionar o problema ou a dúvida. E os dados são utilizados para testar essas hipóteses.

1

Coleta de Dados

Processo que visa reunir os dados para uso secundário

2

Limpeza

Processo para tirar erros, dados repetidos...

3

Organização

Processo que formata a base para melhor uso





Análise Descritiva

A análise descritiva de dados é usada para analisar e investigar conjuntos de dados e resumir suas principais características.

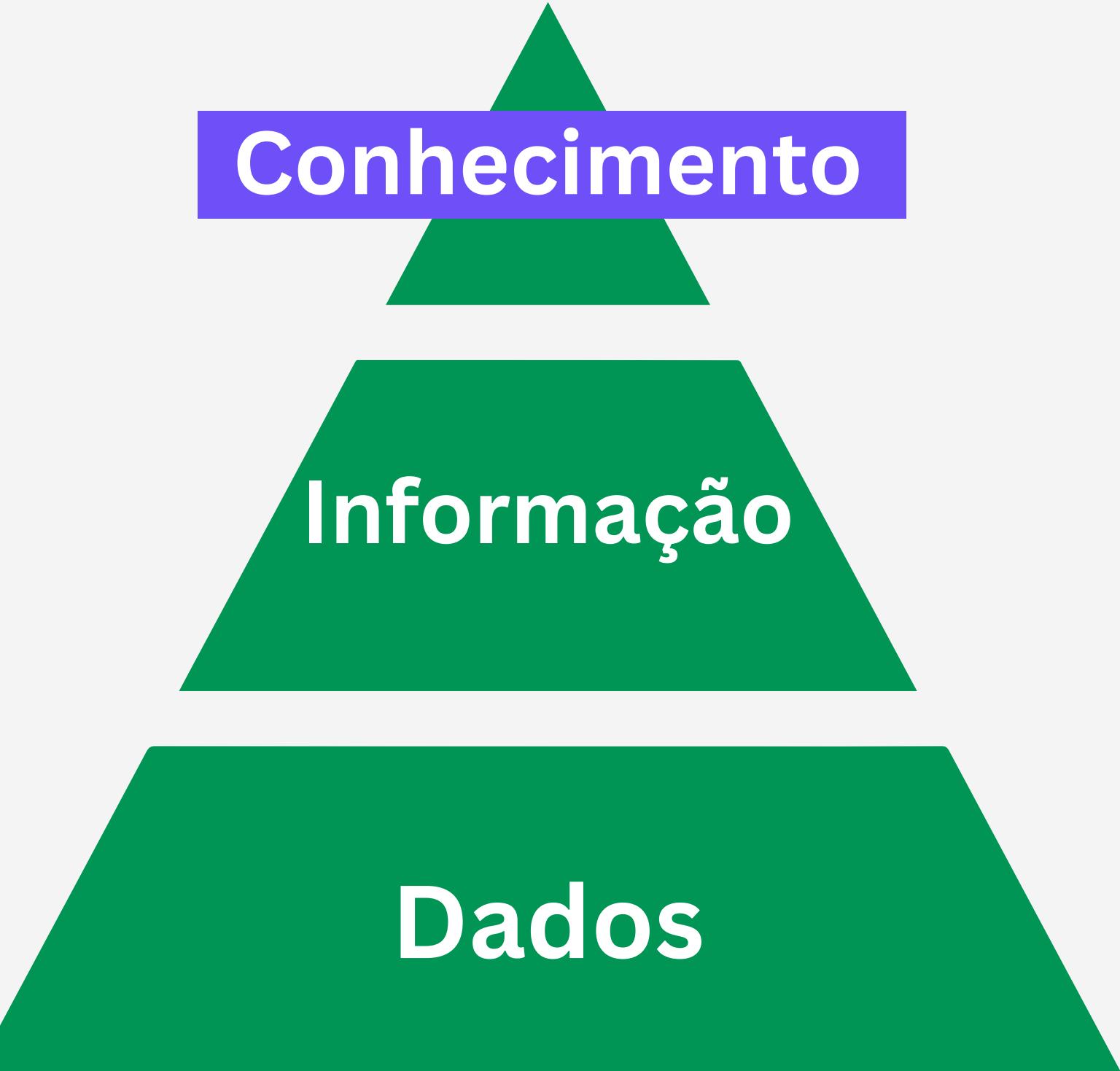
[Back to Agenda](#)

Análise Descritiva

Geralmente a análises descritivas são as primeiras manipulações realizadas em um estudo e tem como principal objetivo resumir, summarizar e explorar o comportamento dos dados. Isso pode ser feito através de tabelas de frequências, gráficos e medidas de resumo numérico.

O que não pode faltar:

- ✓ Valores resumidos em Média, Máximo, Mínimo, e demais estatísticas
- ✓ Leitura de tabelas e Gráficos
- ✓ Ferramentas: Excel, Power BI e etc...



Conhecimento

Informação

Dados

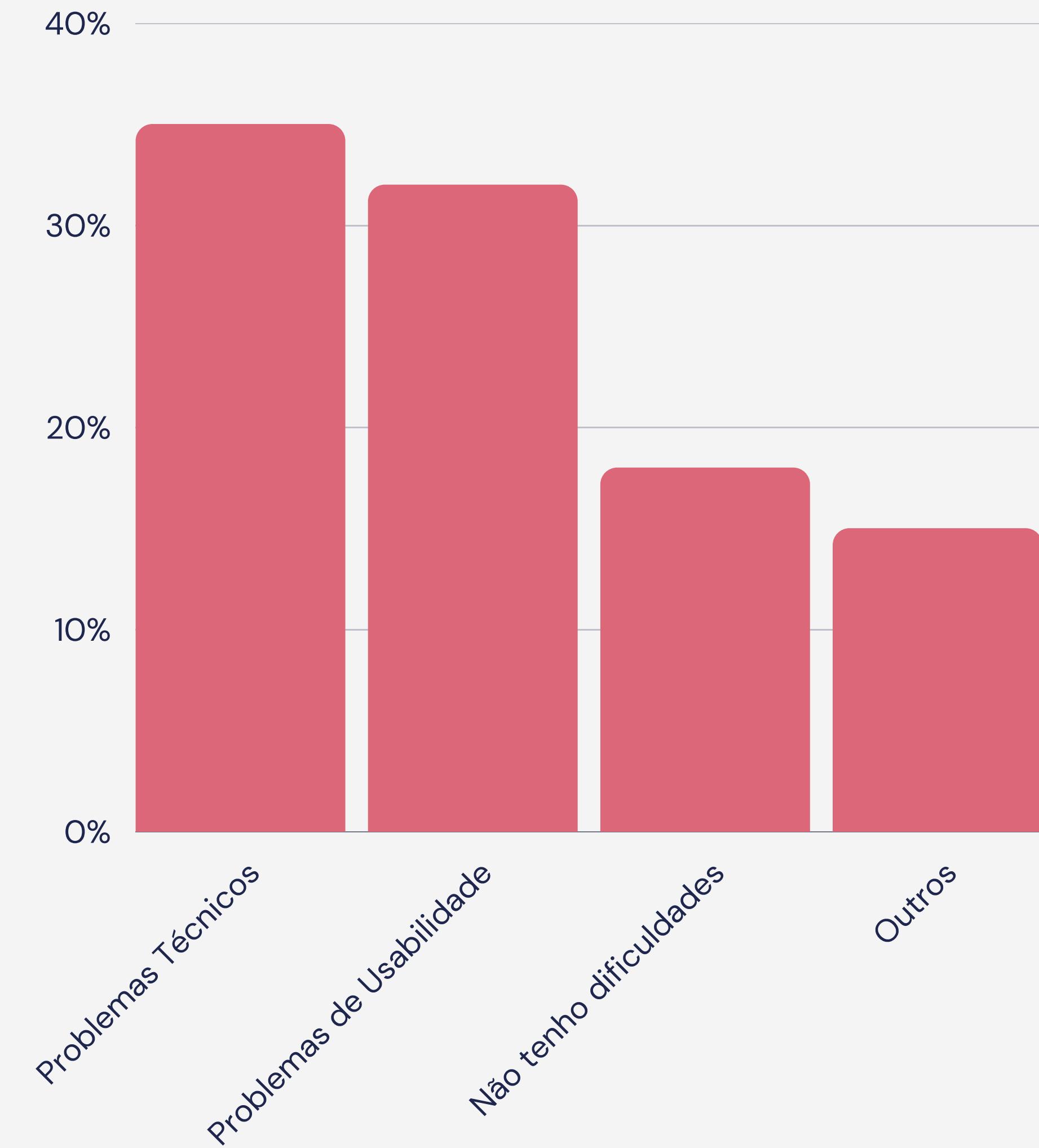
[Back to Agenda](#)

Análise Descritiva

Quais as principais dificuldades e problemas em relação ao sistema Portal de Empresas Credenciadas?

- 35% dos respondentes enfrentam problemas Técnicos no Portal, tais problemas incluem lentidão ao carregar, portal fora do ar, links que travam, e etc.
- Já 32% dos credenciados respondentes têm problemas caracterizados como de usabilidade, tais como dificuldades em encontrar links, plataforma não ser user friendly, entre outros.

Dados simulados





Análise de Correlação

Avalia se o valor de uma variável impacta no valor de outra variável, isto é, se existe um relacionamento linear entre ambas as variáveis.

[Back to Agenda](#)

Análise Correlação

Existem maneiras distintas de estudar o comportamento conjunto entre duas variáveis. Uma forma bastante funcional é a plotagem de um diagrama de dispersão, onde visualmente se observa a relação entre as variáveis, assim como utilizar o Coeficiente de Correlação de Pearson.

O que não pode faltar:

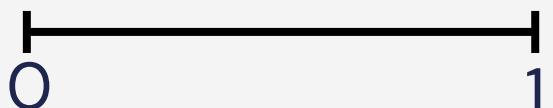
-  Diagrama de Dispersão
-  Coeficiente de Correlação de Pearson
-  Ferramentas: Excel, Python, R, etc...

[Back to Agenda](#)

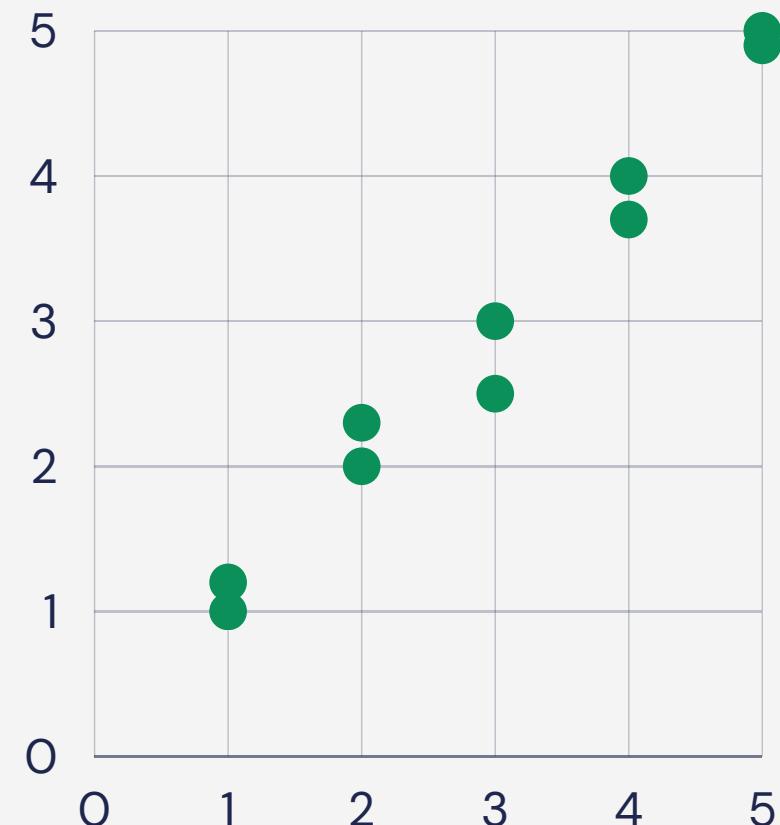
Correlação de Pearson

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]}}$$

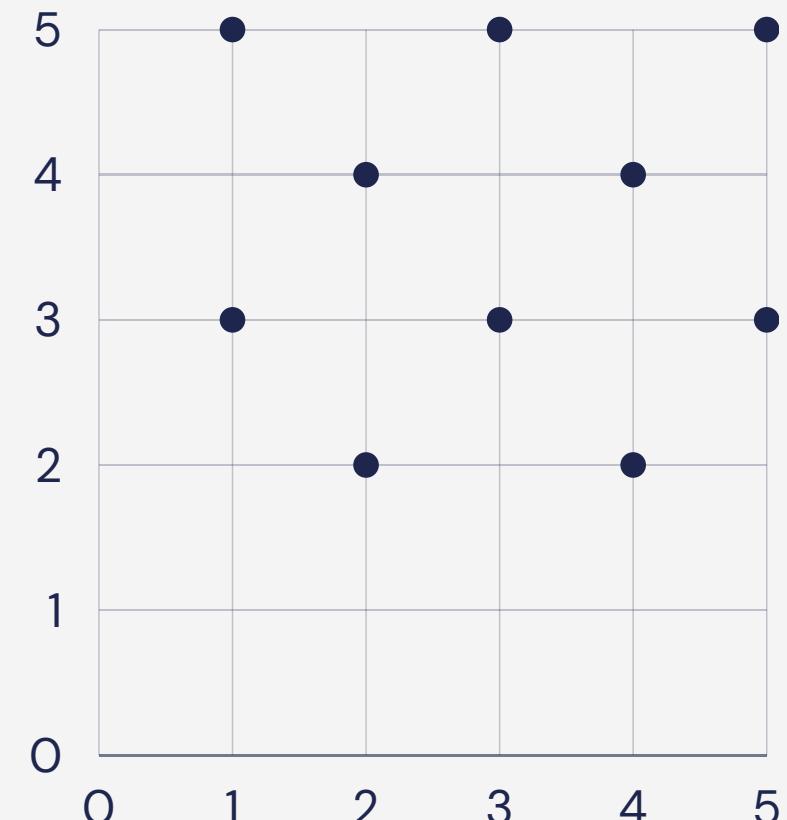
No Excel: CORREL()



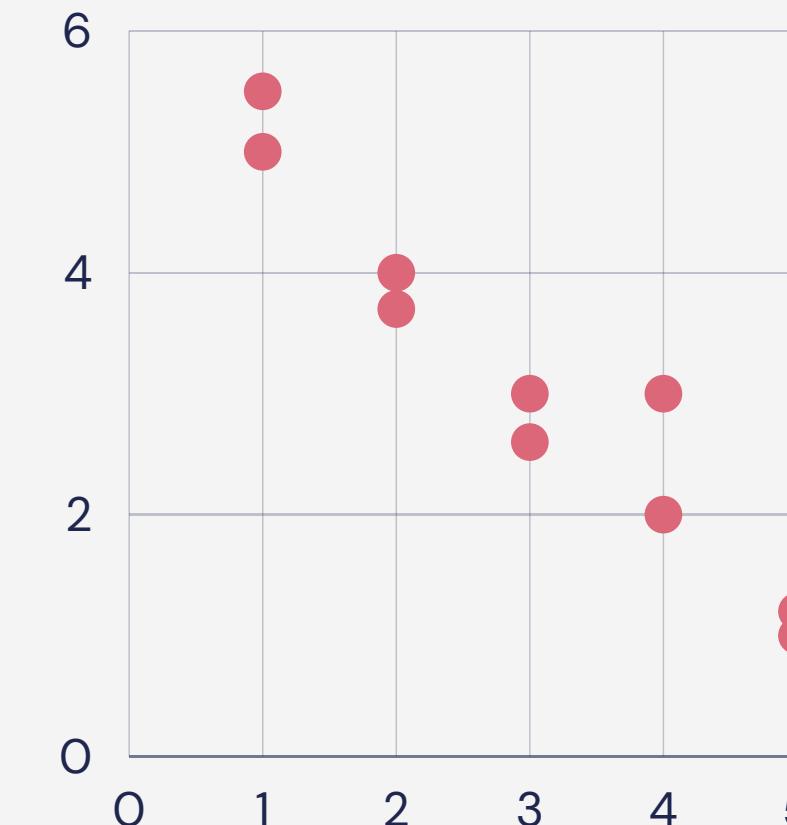
Correlação Positiva



Sem correlação

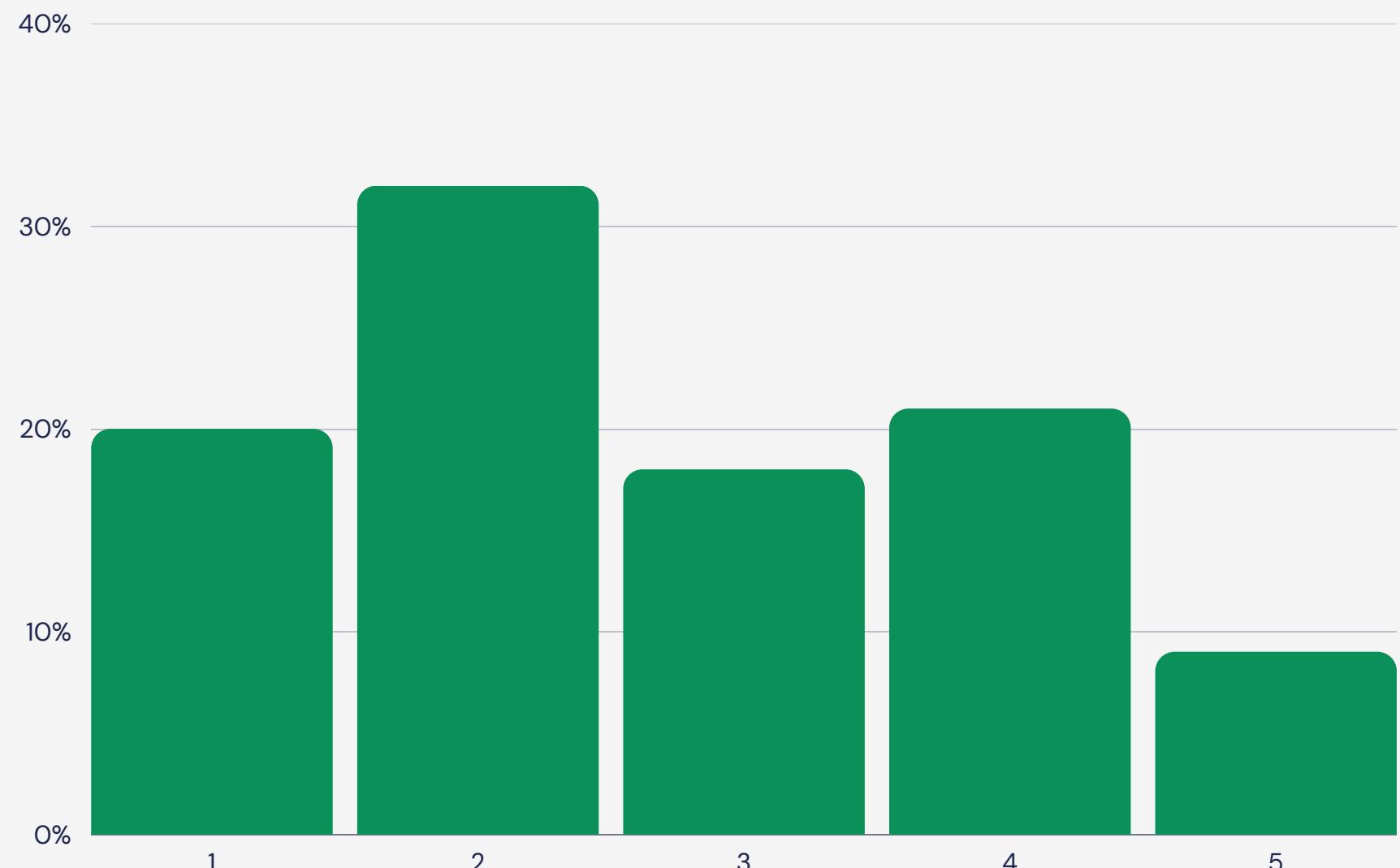


Correlação Negativa

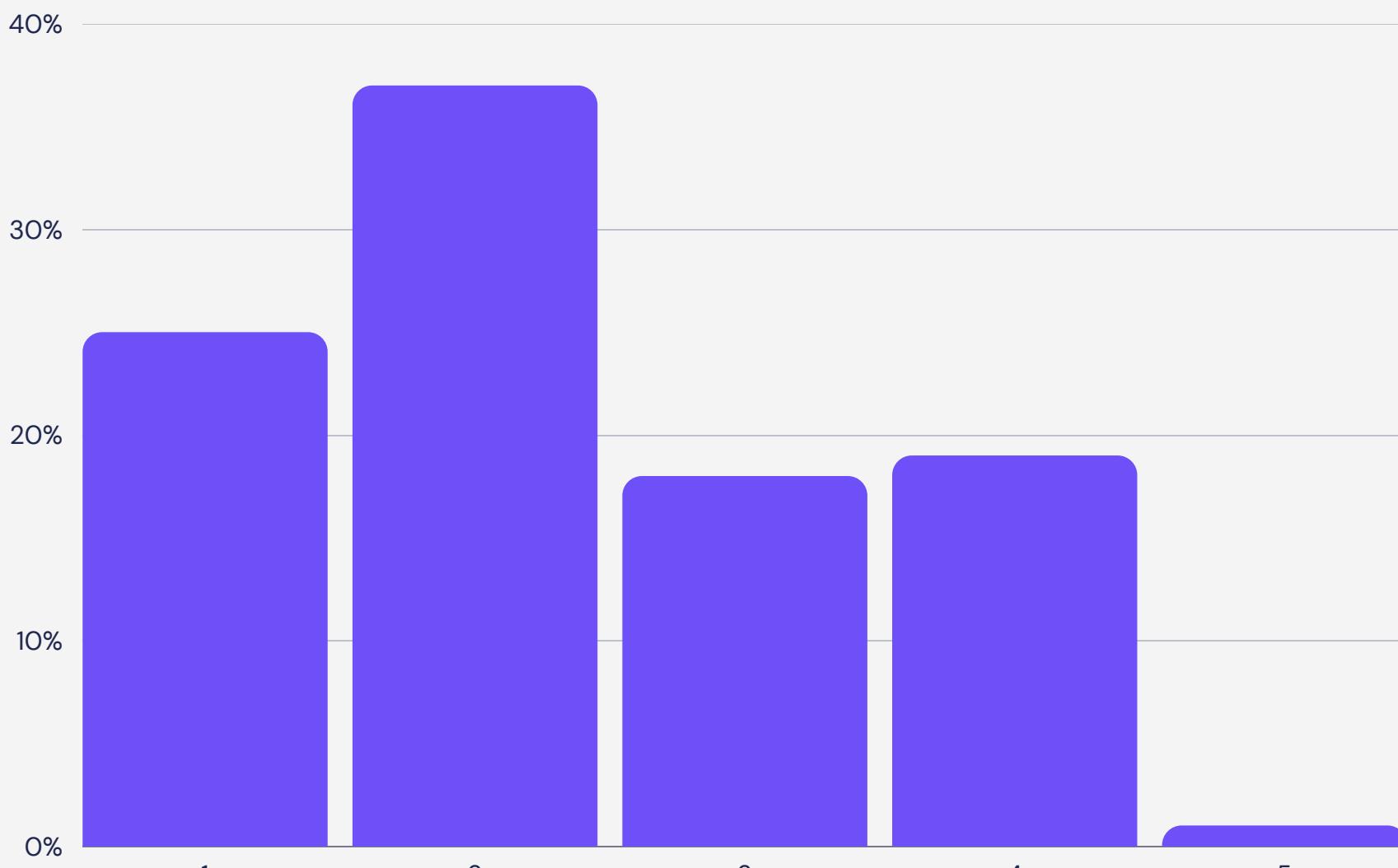


Análise Descritiva vs Análise Correlação

Dado 1: Nota de 1 a 5 : Há clareza na definição e descrição das minhas entregas



Dado 2: Nota de 1 a 5 : Quão frequentes são realizadas reuniões de alinhamento sobre as entregas



[Back to Agenda](#)

Dados simulados

Coeficiente de Pearson: 0,89

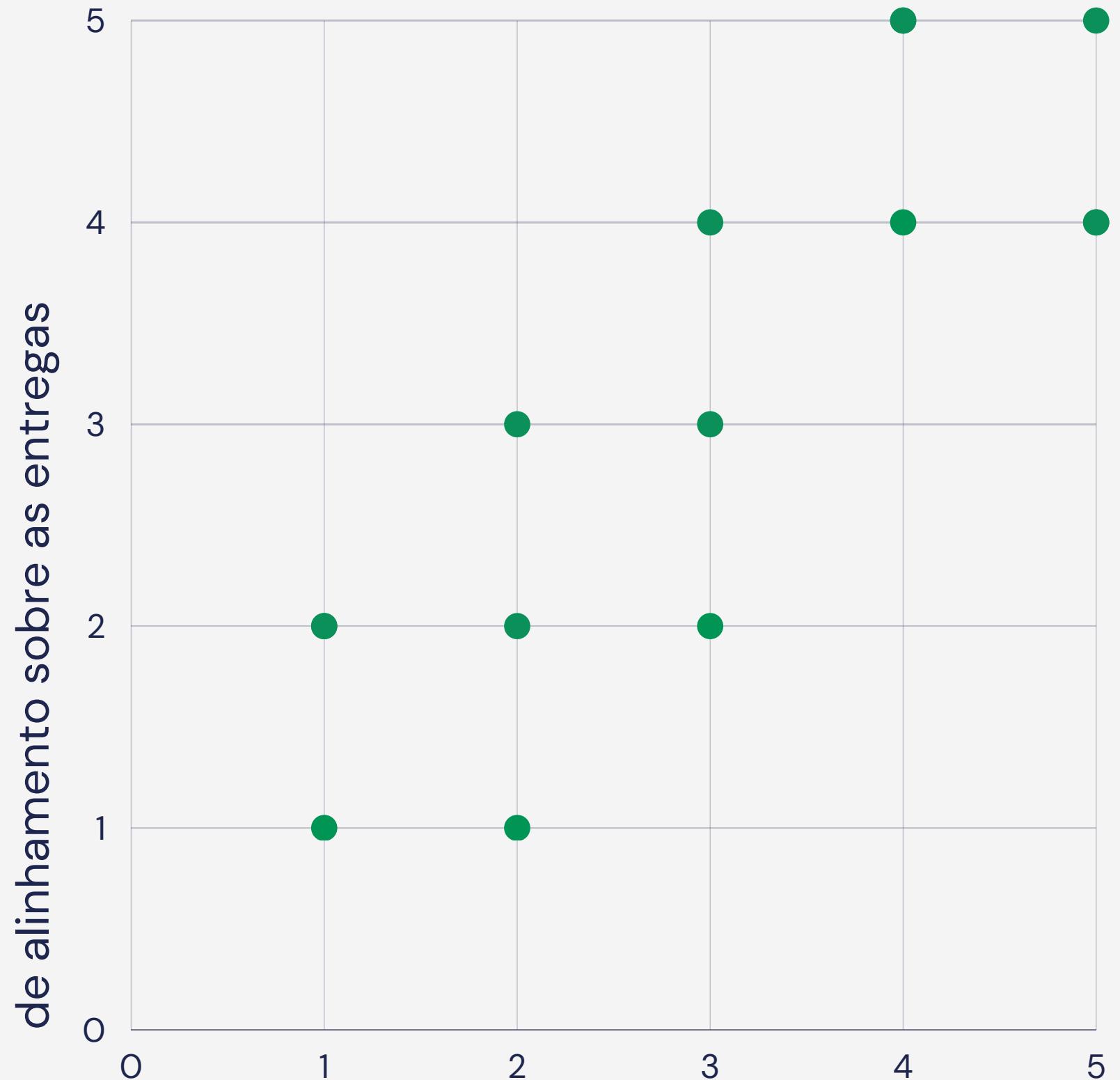
[Back to Agenda](#)

Análise de Correlação

Dado 1: Nota de 1 a 5 : Há clareza na definição e descrição das minhas entregas

Dado 2: Nota de 1 a 5 : São realizadas reuniões de alinhamento sobre as entregas

Dados simulados



Dado 1: Nota de 1 a 5 : Há clareza na definição e descrição das minhas entregas



Modelos Estatísticos

É uma representação da realidade na qual definimos a relação entre variáveis para entender e prever o comportamento de um fenômeno.

[Back to Agenda](#)

Modelos Estatísticos

A ideia por trás dessa metodologia é criar um modelo que consiga descrever os elementos mais importantes para uma análise.

Trata-se, portanto, de utilizar métodos matemáticos para ajudar a encontrar possíveis resultados e melhor compreender o mundo.

O que não pode faltar:

- ✓ Variável resposta e variável (s) explicativa
- ✓ Ferramentas: Python, R, etc...

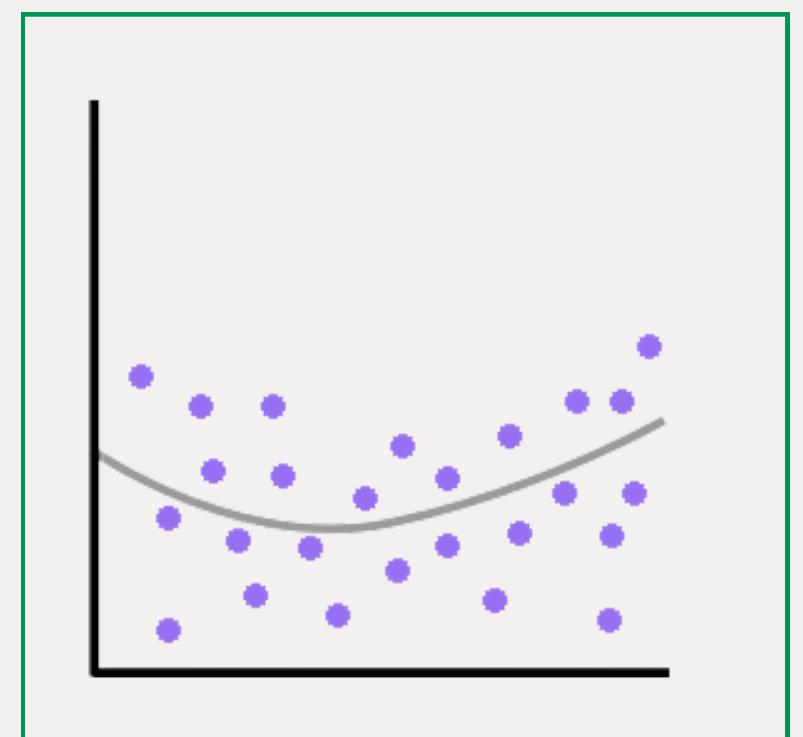
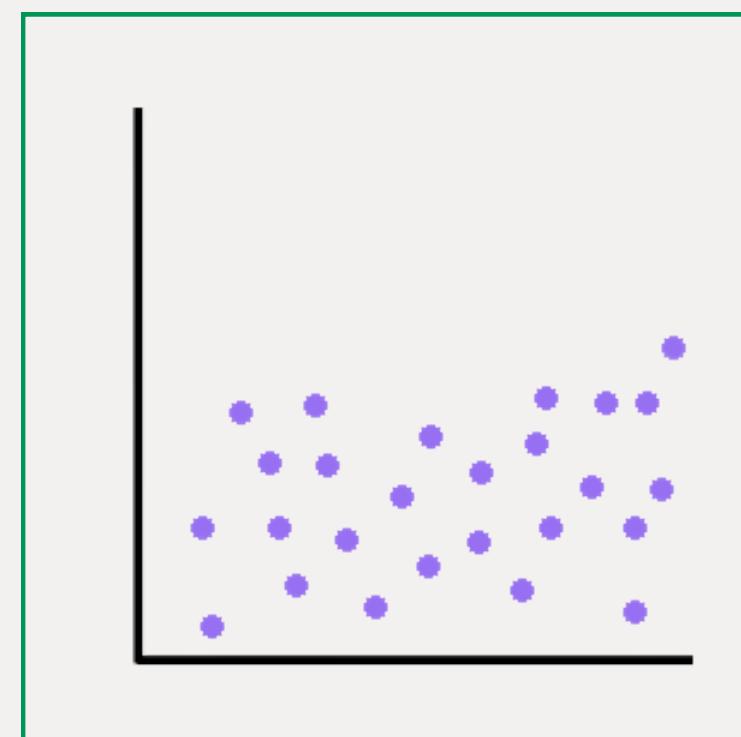
[Back to Agenda](#)

Variável resposta: é aquela que queremos entender, essa variável também é chamada de varável dependente

Variável explicativa: são aquelas que ajudam a explicar a variação da variável resposta, essa variável também é chamada de independente ou preditora.

Função Modelo:

$$Y = \beta_0 + \beta_1 X + e$$



[Back to Agenda](#)

Modelos Estatísticos

Exemplo:

Silvia possui uma loja de roupas, e gostaria de entender o motivo do seu número de vendas variar tanto ao longo da semana.

Variável resposta (Y): Número de Vendas

Variáveis Explicativas (X): Fluxo de pessoas, Data Comemorativa, Marketing Digital.

Dados simulados

Cabeçalho de Dados:

Nº Vendas	Pessoas	Data Comemorativa?	Nº de Campanhas de Marketing D.
78	120	Sim	2
37	70	Não	5
12	22	Não	0

Modelo:

Var. Explicativas significativas:

- Fluxo de Pessoas
- Marketing Digital

Var. Explicativas não significativas:

- Data Comemorativa

$$Y = \beta_1 x_1 + \beta_2 x_2 + e$$

$$Nº\ de\ Vendas = \beta_1(\text{Fluxo\ de\ Pessoas}) + \beta_2(\text{Marketing\ Digital}) + e$$

$$Nº\ de\ Vendas = 0,5\ (\text{Fluxo\ de\ Pessoas}) + 3\ (\text{Marketing\ Digital}) + e$$

Ação:

$$Nº\ de\ Vendas = 0,5\ (300) + 3\ (10) + e$$

$$Nº\ de\ Vendas = 180 + e$$

Atividade 2



30 minutos

- Crie uma hipótese baseada no problema imposto, e na base de dados que você recebeu.
- Quais Dados da base você utilizaria?
- Quais análises você faria?
- Se for necessário um modelo estatístico, qual é a sua variável resposta, e quais são suas variáveis explicativas?





Visualizações

Por que gerar visualizações é importante?

Design dando forma para os dados



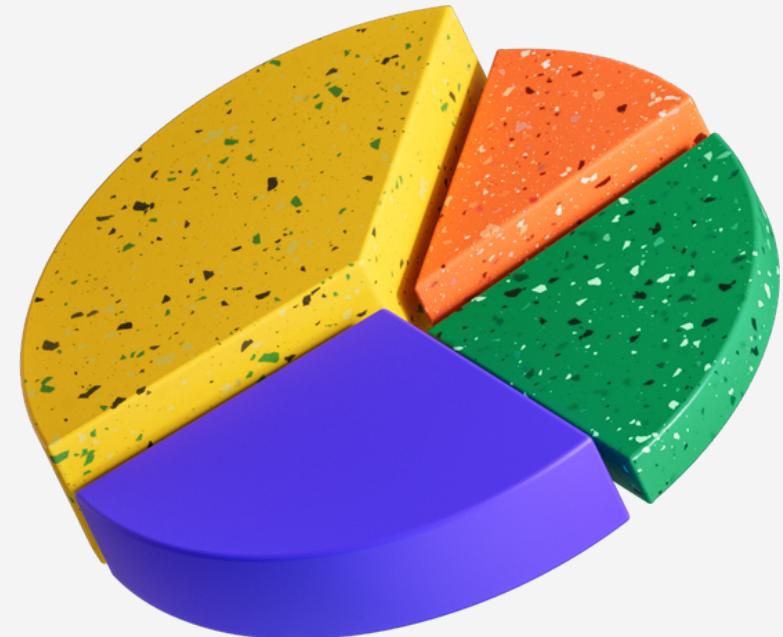
Porque fazer visualizações

-  Interpretação em outro formato
-  Compreensão mais rápida
-  Identificação de novos padrões

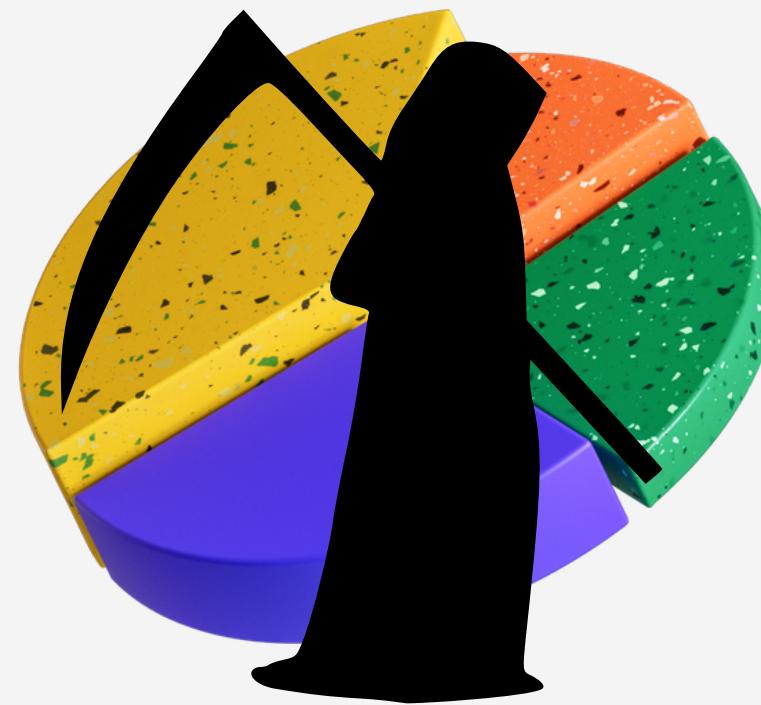
**Quando falamos em
representação de dados é
possível que você lembre
primeiro dele...**

O gráfico de pizza

**Eu entendo... ele é um dos primeiros que
aprendemos na escola e tem um nome
muito mais convidativo não é mesmo?**



10 opções para você parar de usar SÓ gráfico de pizza



Tipos de Gráficos

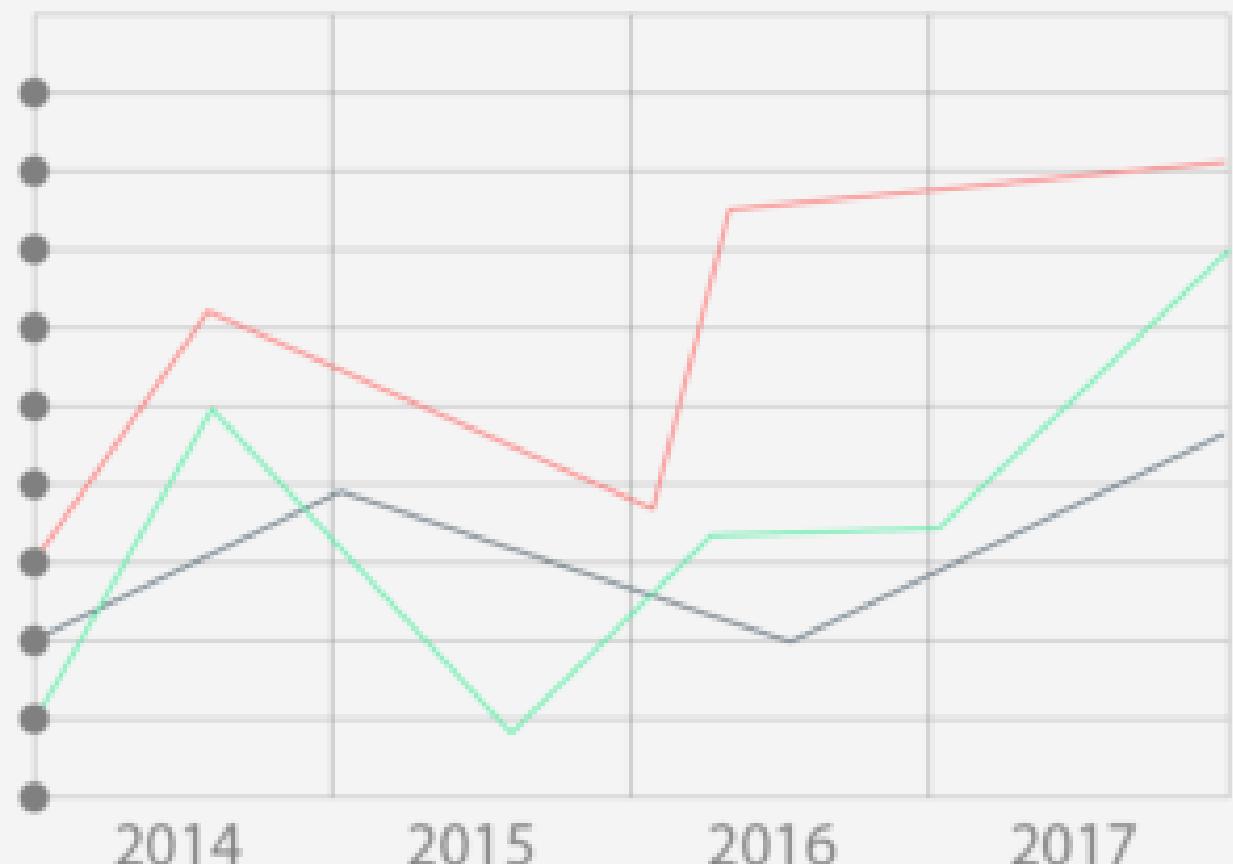


Tipos:

- Comparação
- Composição
- Distribuição
- Relação

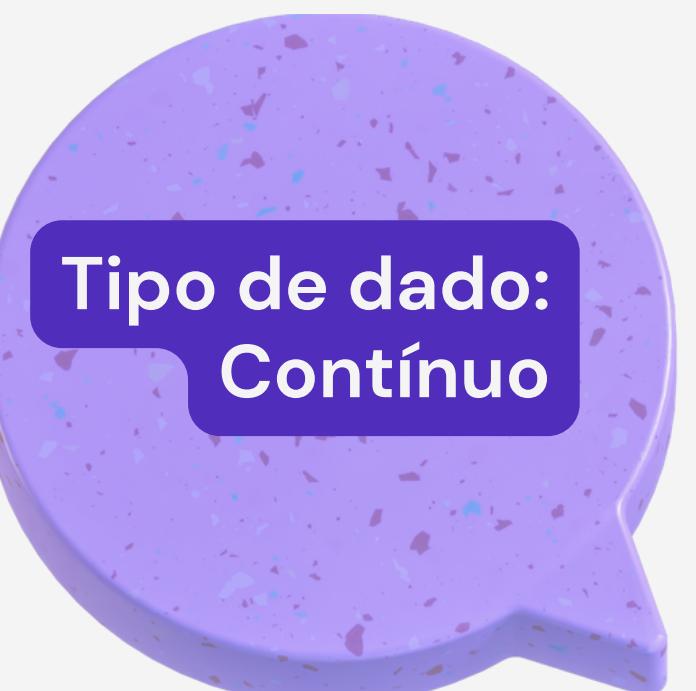
COMPARAÇÃO

Gráfico de linha



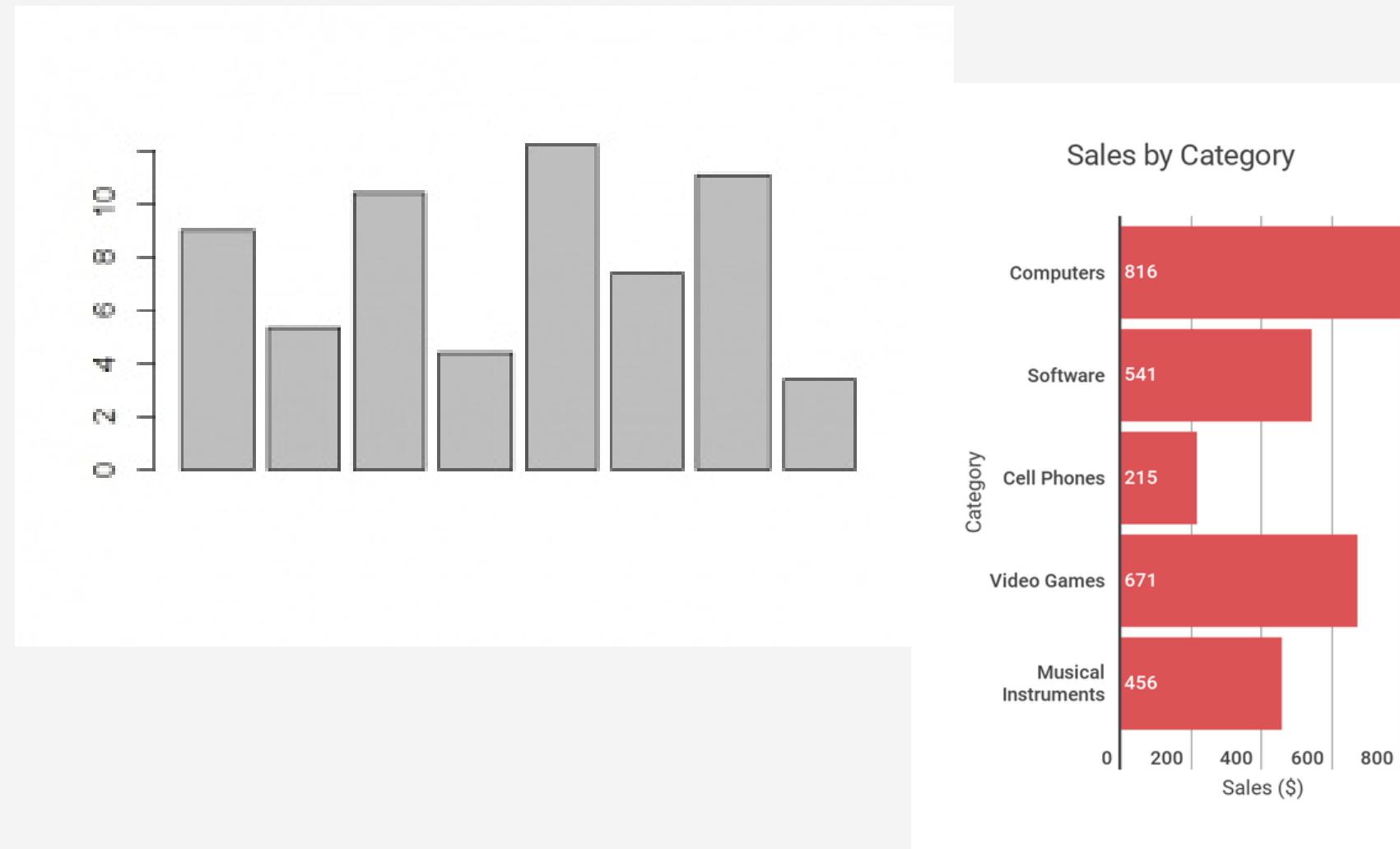
Aplicações:

- Indicam evolução;
- Tendências e progressos em um **tempo**;
- Funcionam bem com maior número de variáveis



COMPARAÇÃO

Gráfico de barras/coluna



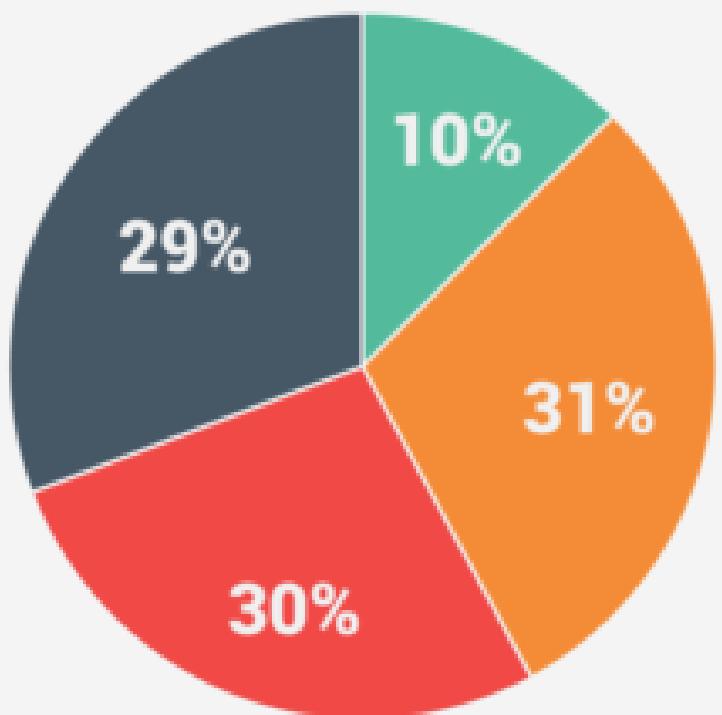
Aplicações:

- Comparações de categorias
- Ranking

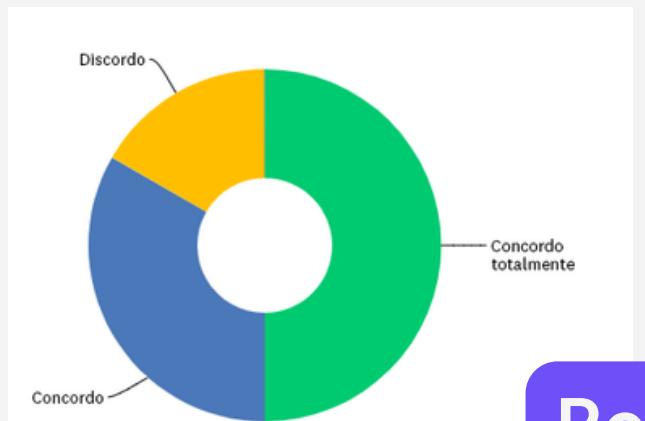
Tipo de dado:
Quantitativo
discreto e
qualitativo

COMPOSIÇÃO

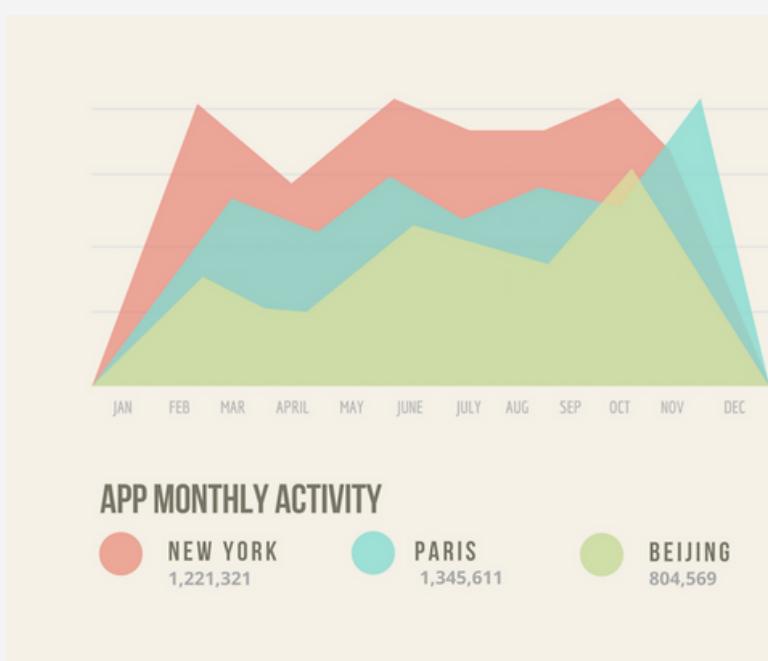
Gráficos de composição



Pizza



Rosca



Área

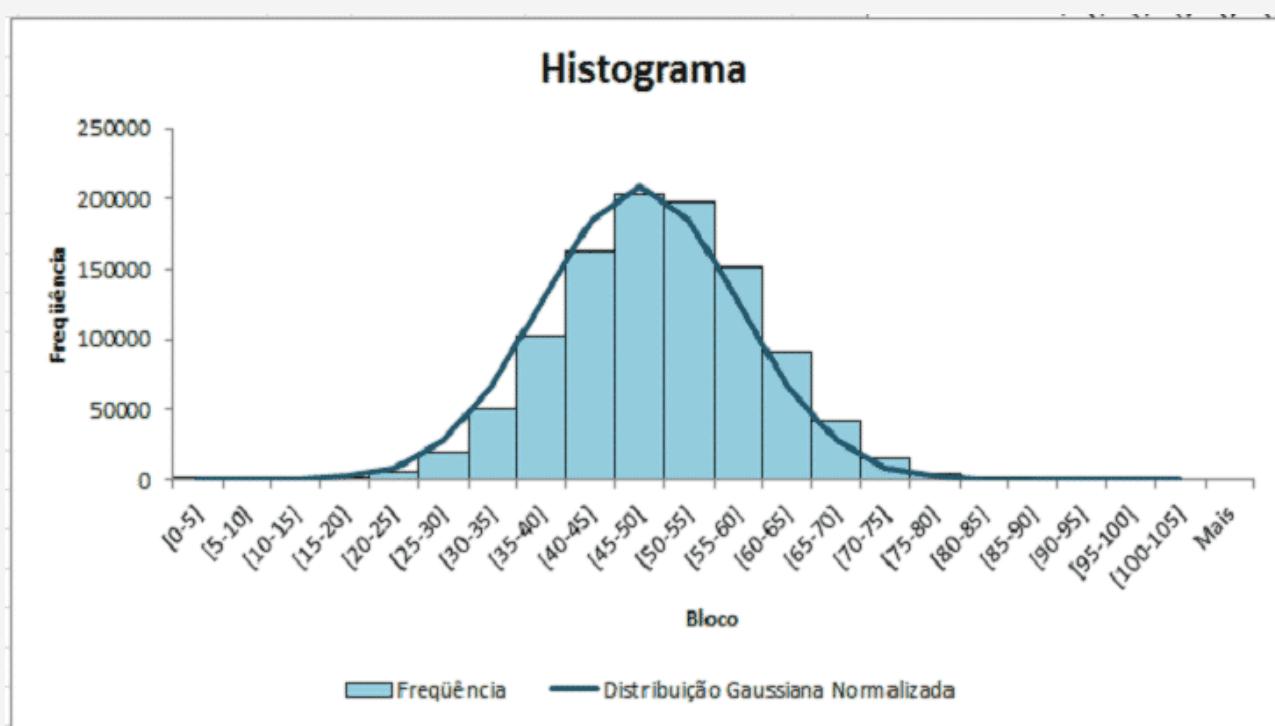
Aplicações:

- Número estático
- Percentuais
- São formas familiares e conhecidas



DISTRIBUIÇÃO

Histograma



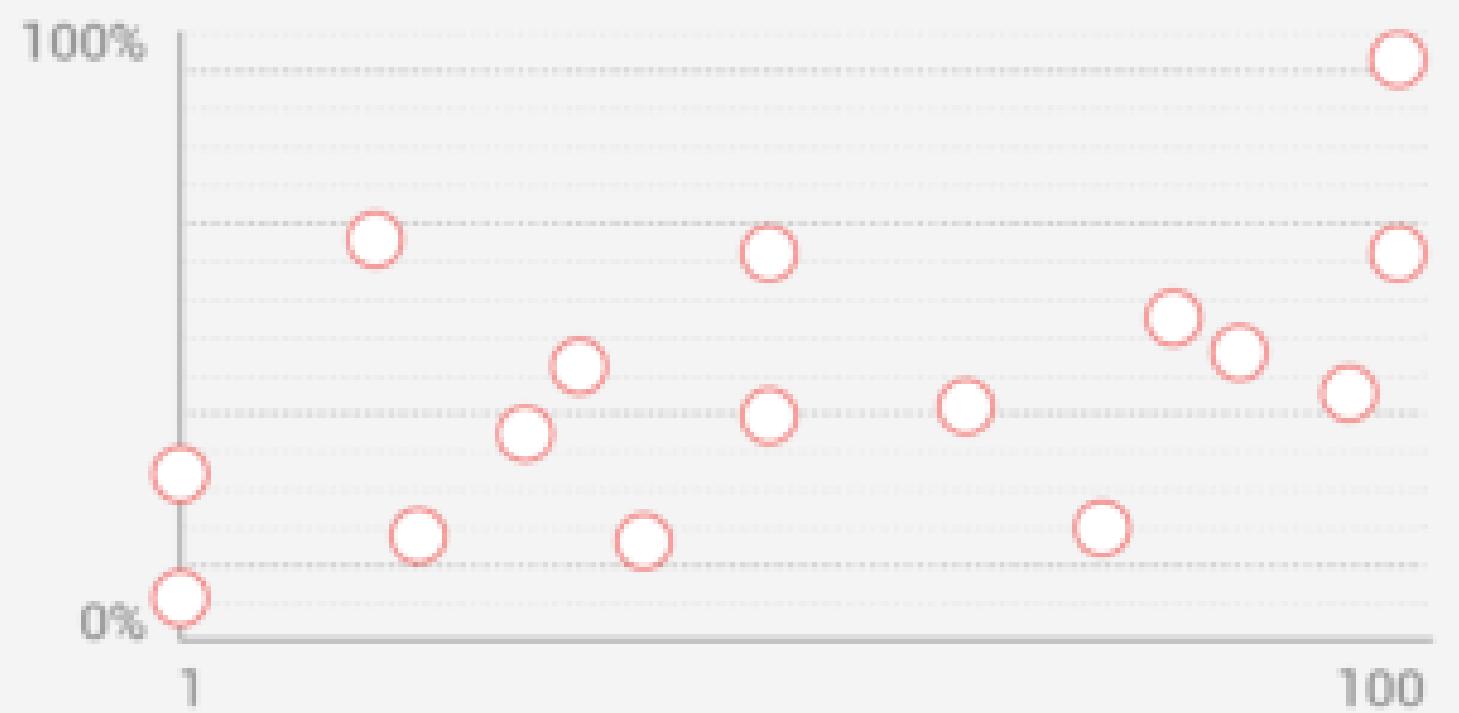
Aplicações:

- Distribuição
- Quais dados aparecem mais

Tipo de dado:
Descritivo e
qualitativo

DISTRIBUIÇÃO E RELAÇÃO

Gráfico de Dispersão



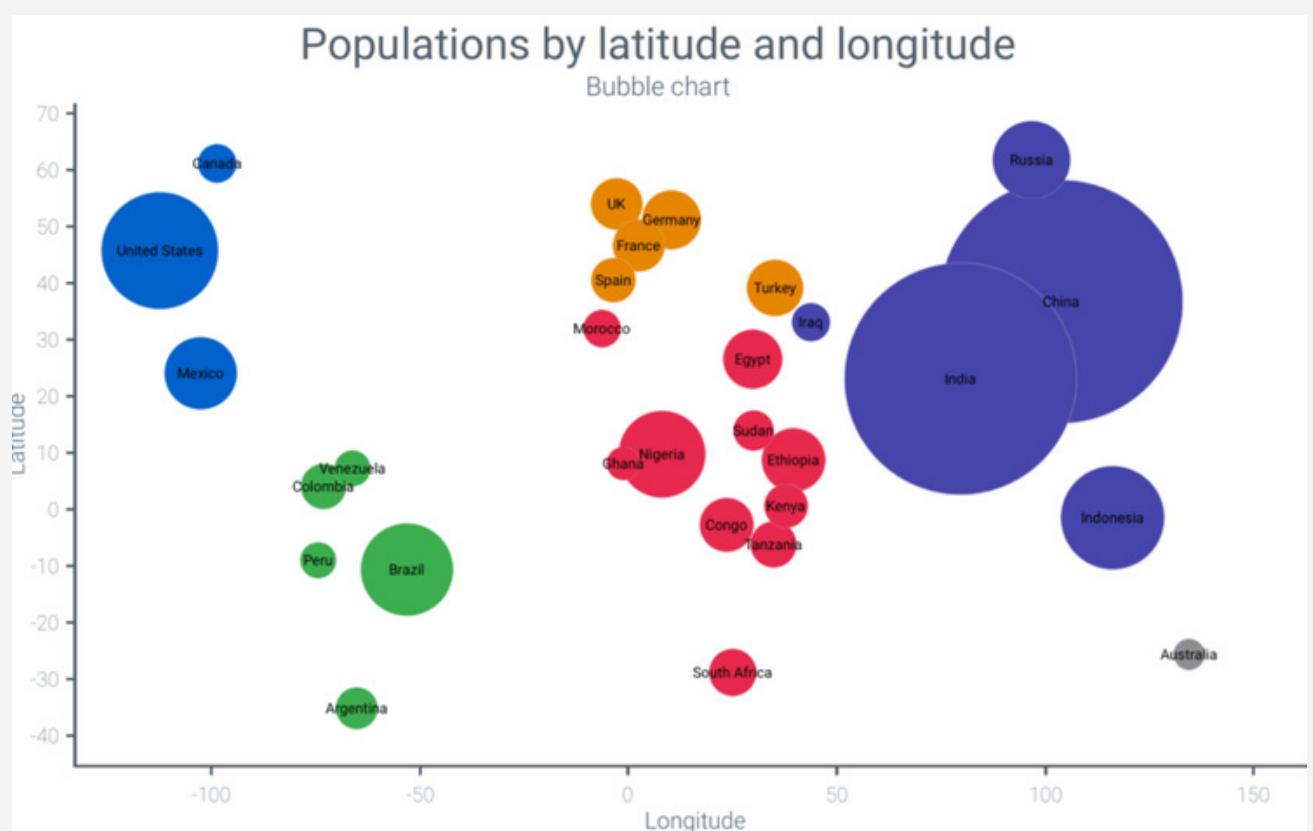
Aplicações:

- Conexão entre duas ou mais variáveis
- Causa e efeito



RELAÇÃO

Gráfico de Bolhas



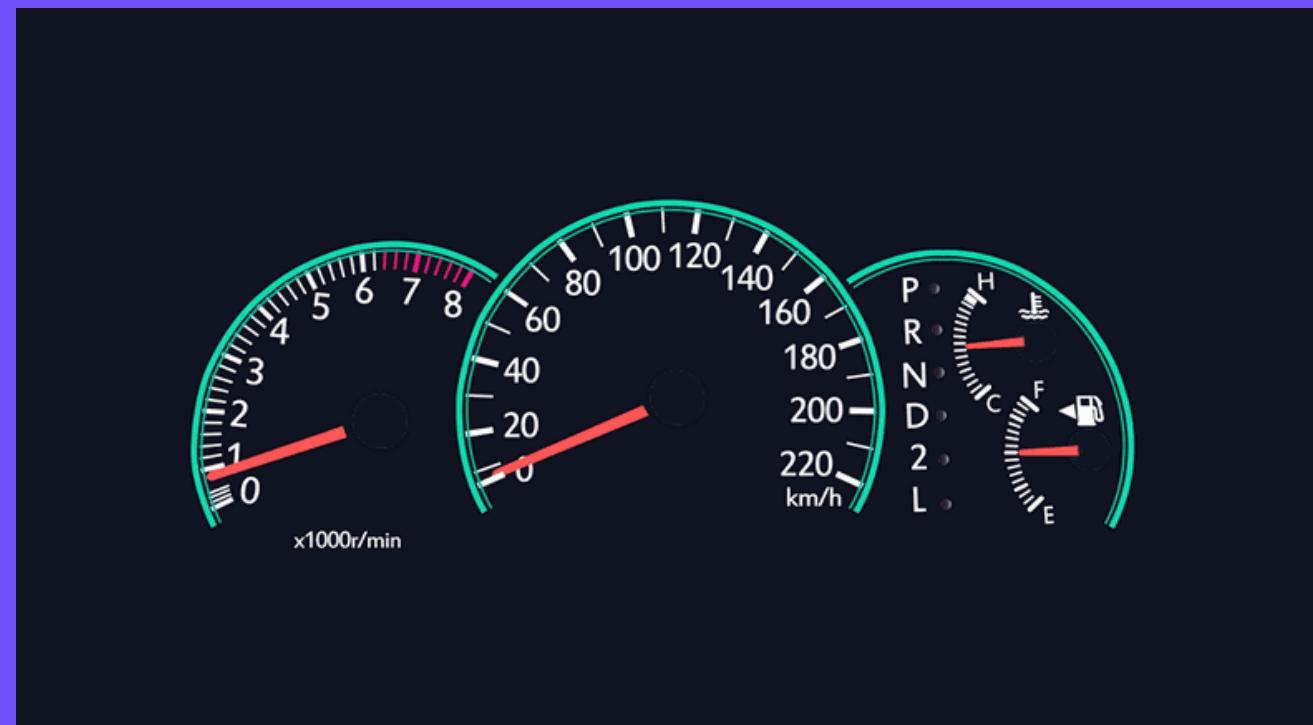
Aplicações:

- Visualizações de padrões

Tipo de dado:
Quantitativo
discreto e
continuo e
qualitativo

Dashboards

ex: painel do carro



Já pensou se tivesse um gráfico de pizza para mostrar a velocidade? ou um gráfico de dispersão para mostrar a temperatura?

Daí a importância de escolher o tipo de gráfico certo para cada informação.

Passo a passo de como criar suas visualizações

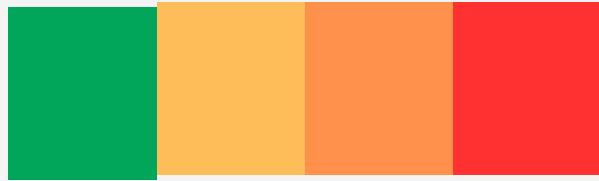
Um dashboard ajuda seus usuários a compreender e analisar, apresenta os pontos principais da análise de forma rápida e prática.

O que o seu cliente quer olhando para sua representação?

Objetivos:

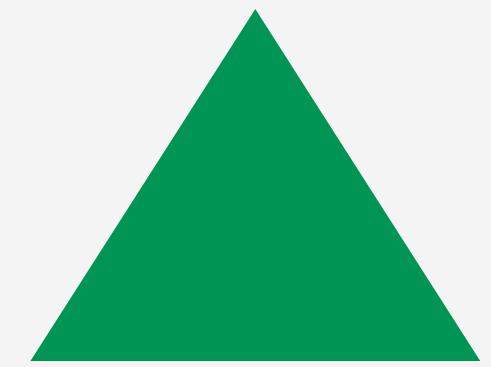
- OPERACIONAL – Situação atual Ex. painel do carro
- ANALÍTICO – Performance, evolução Ex. desempenho de uma campanha.
- ESTRATÉGICO – Indicadores, tendências

Diretrizes

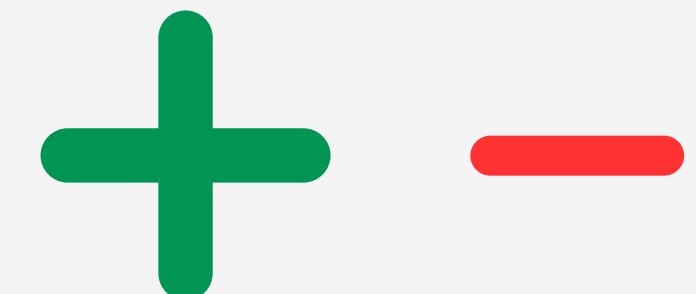


Temperatura

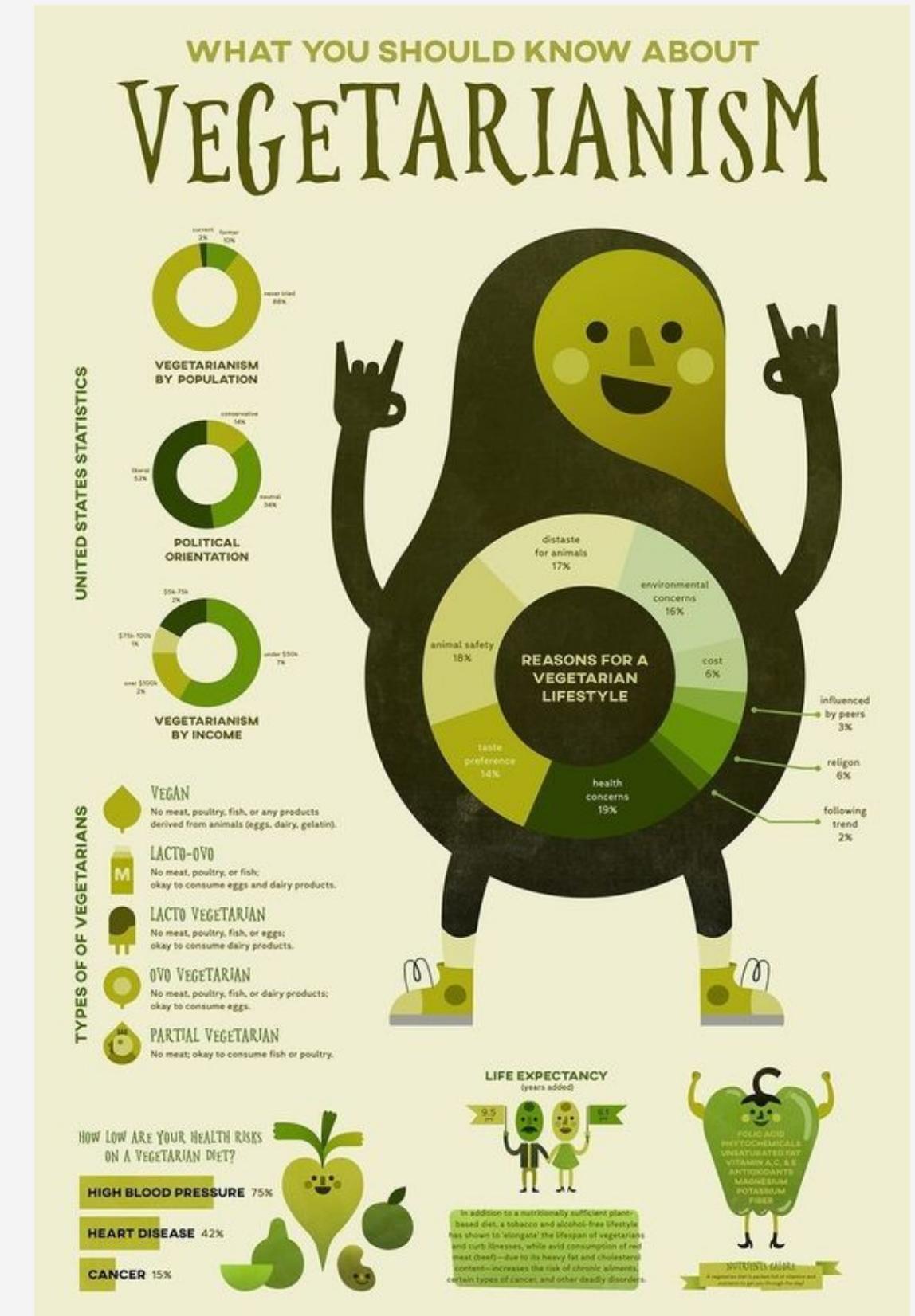
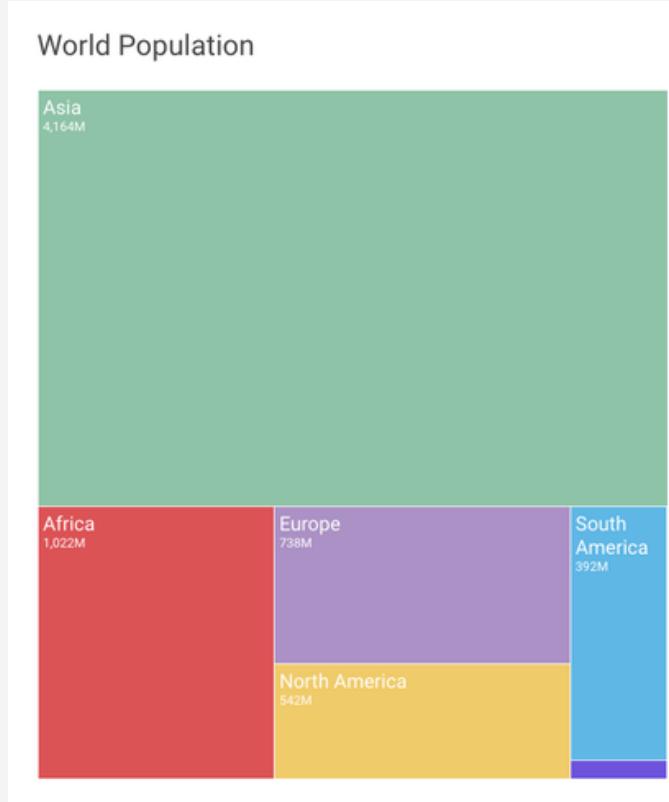
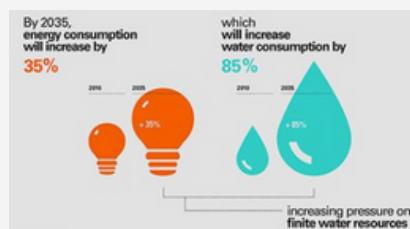
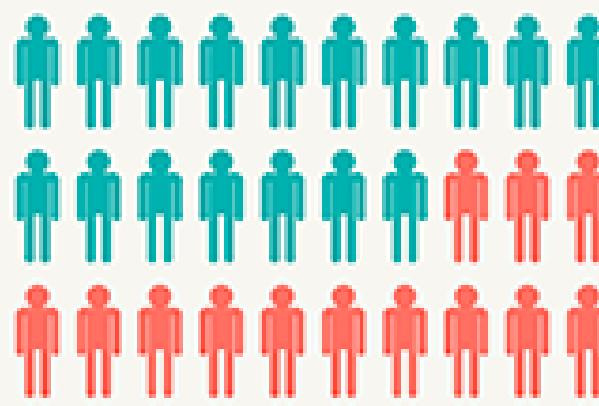
1. Defina o objetivo
2. Escolha os melhores gráficos para o contexto
3. Contextualize e reforce com ícones/desenhos
4. Utilize hierarquia: as informações mais importantes vem primeiro (priorize)
5. Proximidade: o que tem relação fica próximo, alinhe informações que tem relação
6. Pense no storytelling, por onde começar?
7. Use cores intuitivas



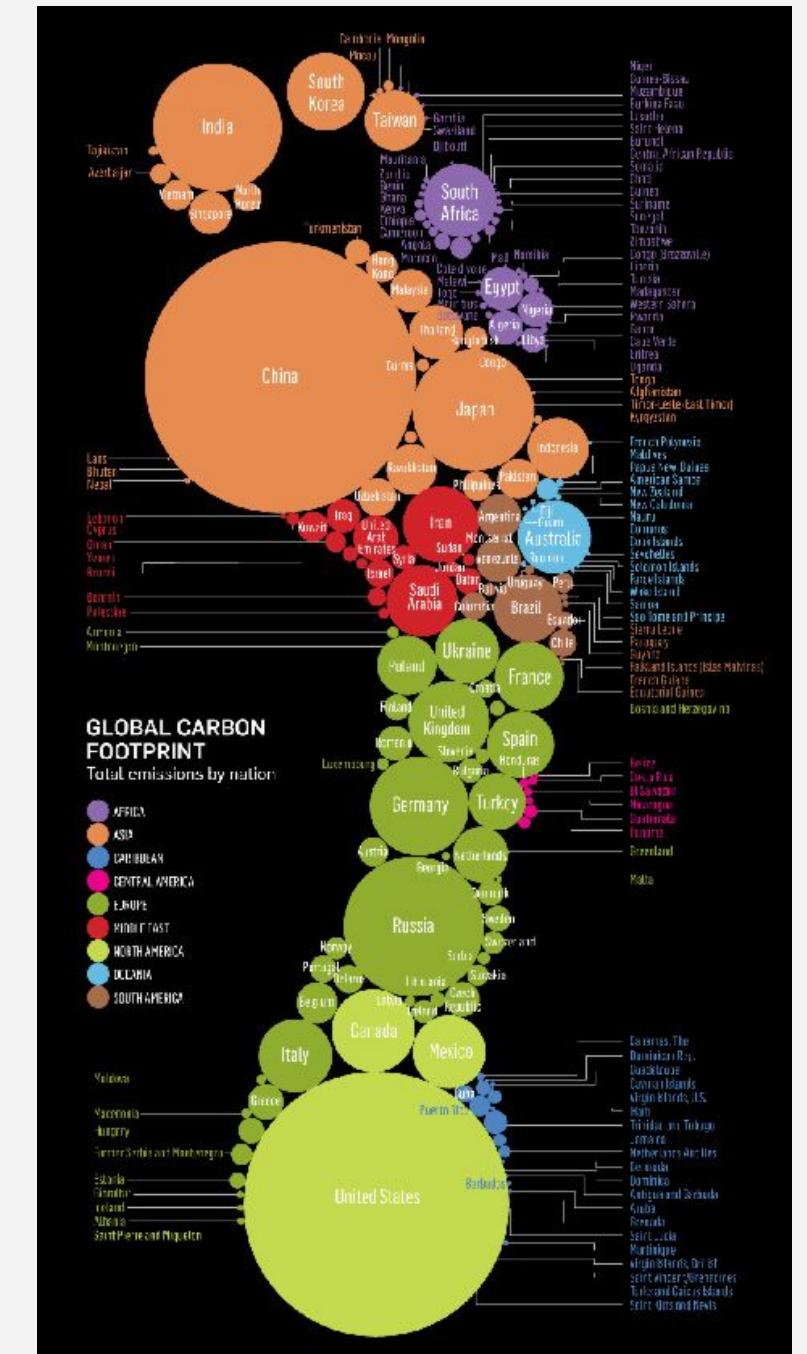
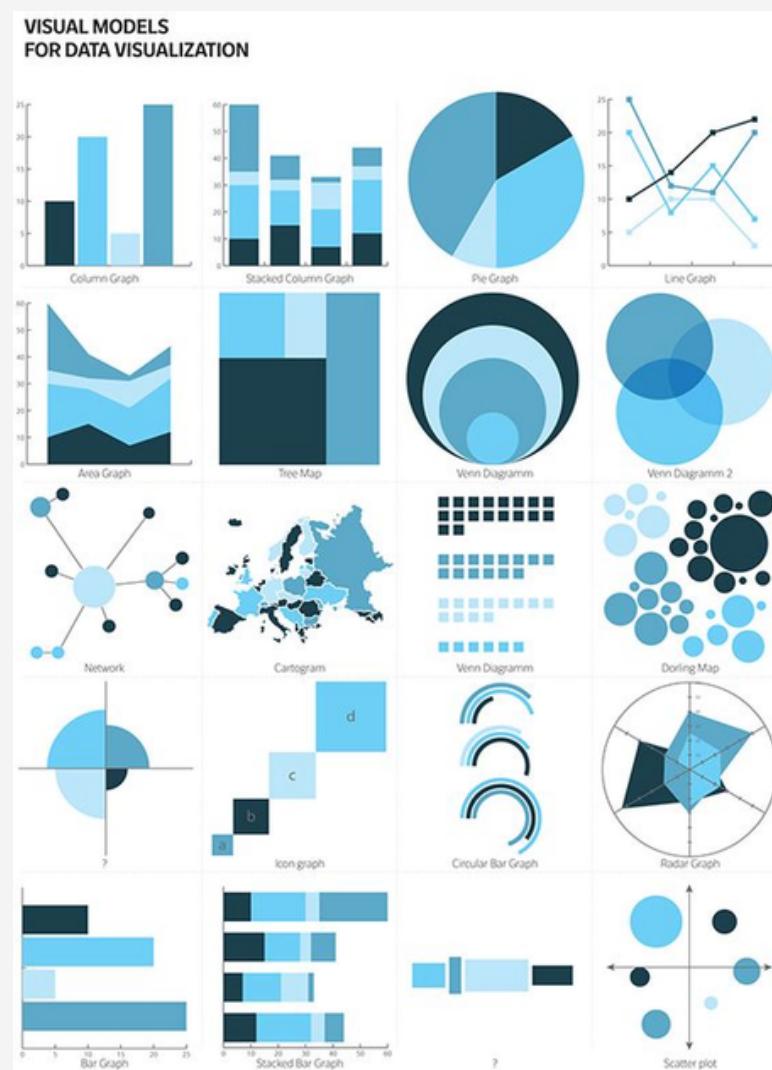
Aumento/crescimento



Recursos para deixar relatórios de dados mais interessantes



Recursos para deixar relatórios de dados mais interessantes



Atividade 5

Criar um dashboard com as visualizações das hipóteses

