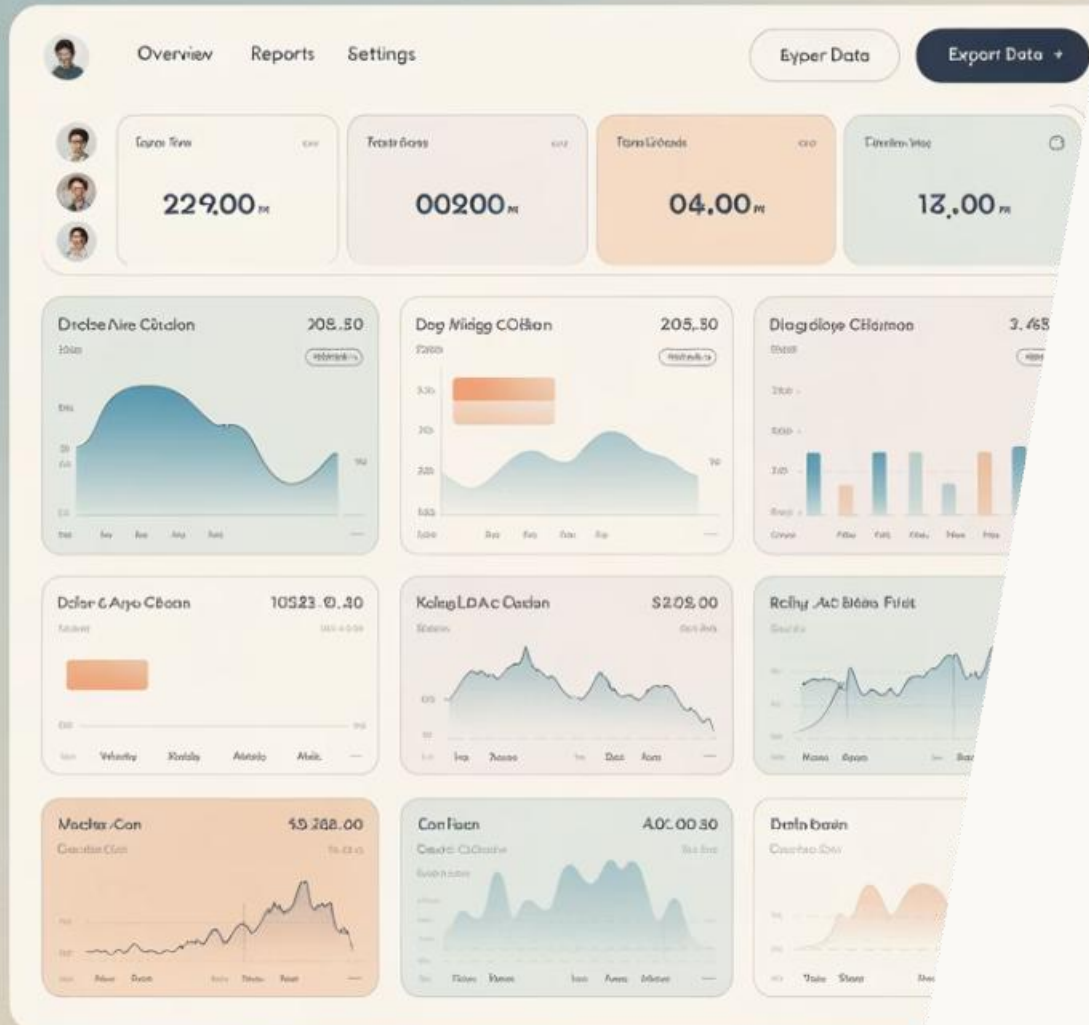


Waimata luyuyaatiarl



Visualização de Dados em Ciência de Dados

Prof. Daniel Otávio Tambasco Bruno

Nesta apresentação, exploraremos a importância dos gráficos na análise de dados e como extrair informações valiosas através de diferentes técnicas de visualização.



Importância dos Gráficos na Análise de Dados



Obtenção de Informações

Gráficos revelam padrões e tendências ocultas nos dados.



Análise de Processos

Visualizações ajudam a entender comportamentos específicos.



Identificação de Anomalias Anomalias

Facilitam a detecção de valores discrepantes em conjuntos de dados.

Conjunto de Dados: Fish.csv

Conteúdo

Espécies de peixes e suas medidas corporais.

Atributos

Espécie, peso, comprimentos, altura e largura.

Objetivo

Explorar diferentes visualizações para análise dos dados.



SPECIES IDENTIFICATION CHART

Preparação dos Dados

Carregamento

Importação do arquivo Fish.csv para análise.

Renomeação

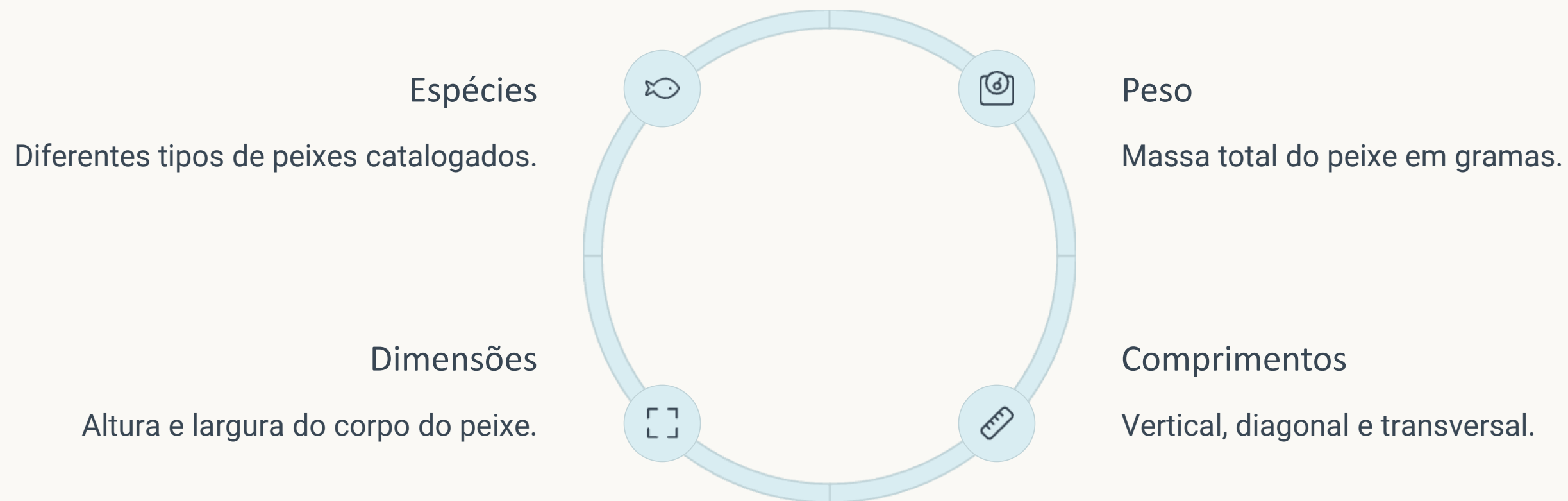
Alteração dos nomes das colunas para facilitar a seleção.

Verificação

Confirmação dos tipos de dados e ausência de valores nulos.



Atributos do Conjunto de Dados



Histogramas: Conceito

Definição

Gráficos de barras que mostram a distribuição de frequências de valores em um conjunto de dados.

Os dados são agrupados em intervalos uniformes chamados "bins".

Utilidade

Permitem visualizar quantos pontos estão dentro de cada intervalo.

Ajudam a entender a distribuição e concentração dos dados.

Criando Histogramas com Python



Numpy

Use `np.histogram` para obter a divisão dos grupos e frequências.



Seaborn

Utilize `sns.histplot()` para criar o gráfico visual.



Parâmetros

Configure `bins`, `kde` e `color` para personalizar a visualização.



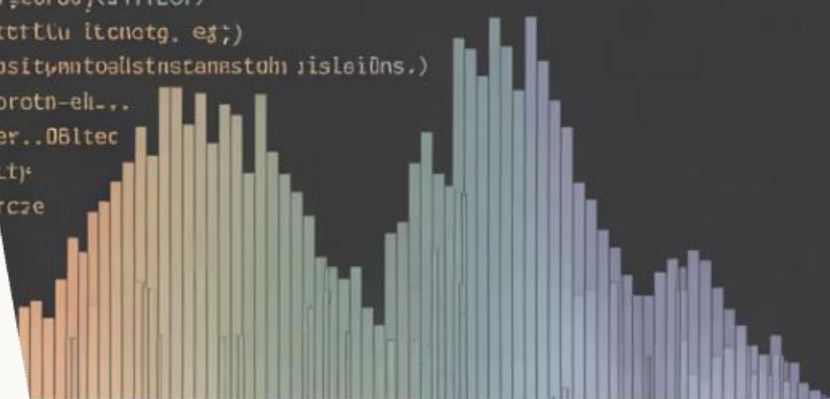
Título

Adicione `set(title='texto')` para nomear o gráfico.

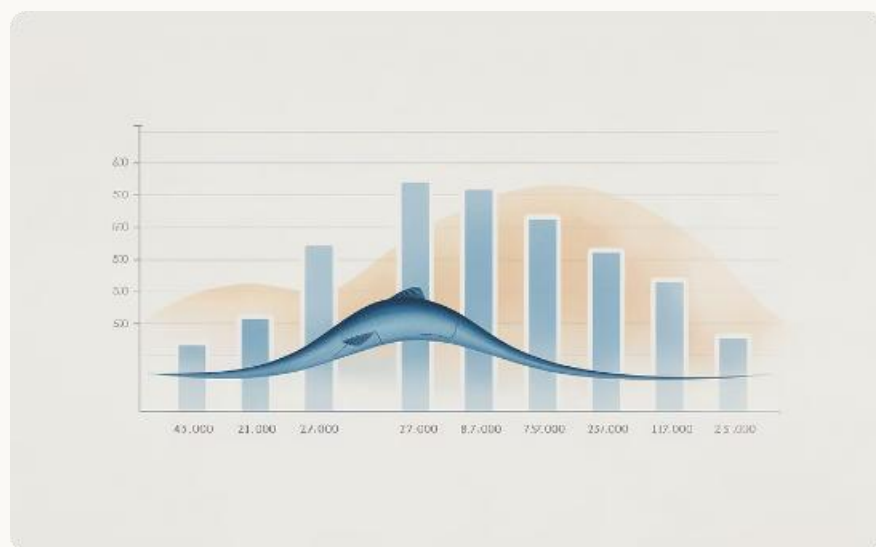
Elegant data visualization

[DOCUMENTATION](#)[EXAMPLES](#)[COMMUNITY](#)

```
SEScabornHistogram
x tiboscttentomnatugucola,nt, oelcto, ciean a.mya cüoctti ill, eelsita,)
F
[yobh,)
tLlhc .b3C0lw
ucbi, ch. e3,11
(11, eor6u7(JYTYEOt)
ycttttllu itcnotg. e3;)
iiositymntoalstnstanestoh iislei0ns.)
torotn-el...
ter..06lter
oltj
rcze
urnileentisyriastig; vohy coopred.searstunyuurfelc)
ton, se. eol' eoga,cardtnentg,)
natev'lito .enetgo14..ac)
```

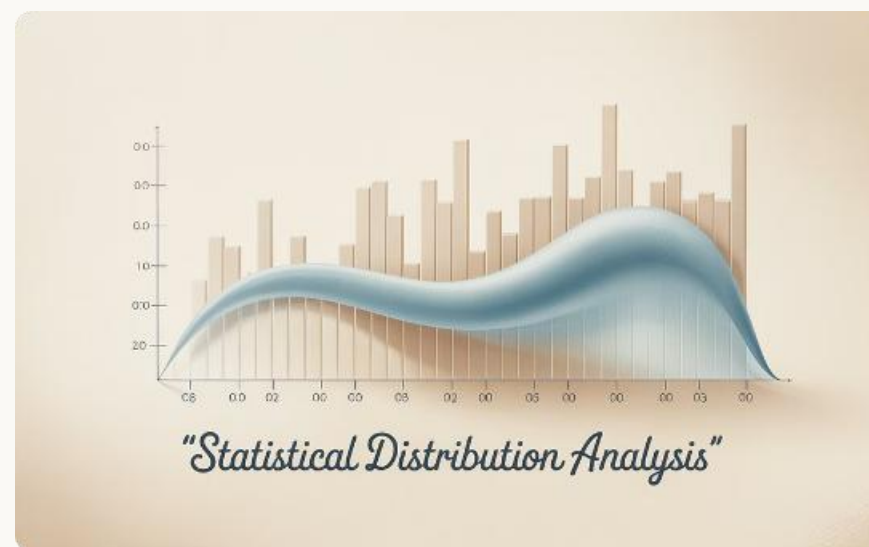


Exemplo de Histograma: Peso dos Peixes



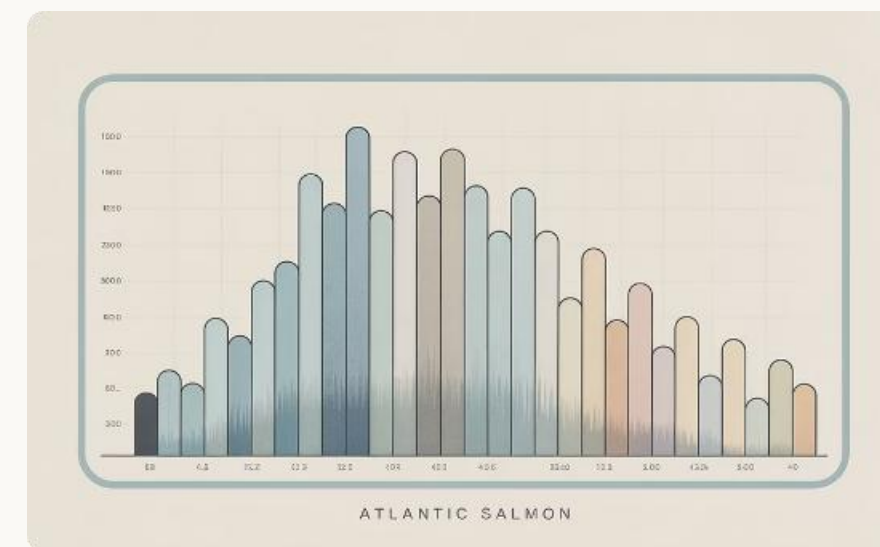
Histograma Básico

Distribuição do peso de todos os peixes no conjunto de dados.



Com Curva de Densidade

Histograma com kde=True mostrando a curva de densidade de probabilidade.



Histograma Filtrado

Distribuição de peso apenas para peixes da espécie Bream.

Gráficos de Densidade

Definição

Estimativa de uma distribuição contínua de probabilidades que pode ter gerado os dados observados.

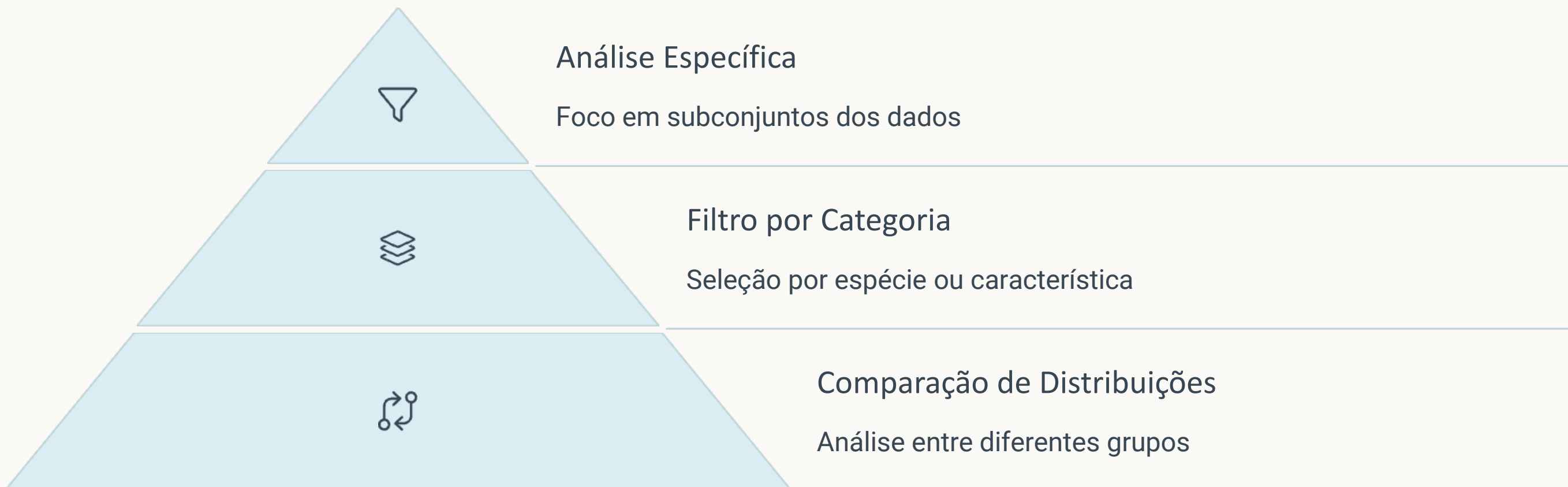
Suaviza a representação da distribuição dos dados.

Implementação

Use `sns.kdeplot()` do Seaborn para criar gráficos de densidade.

Pode ser combinado com histogramas usando `kde=True` no `histplot`.

Histogramas com Filtros



Os histogramas filtrados permitem analisar a distribuição de uma variável dentro de uma categoria específica. Por exemplo, podemos visualizar apenas o peso dos peixes da espécie Bream para entender melhor suas características.



Gráficos de Dispersão: Conceito

Definição

Gráficos que mostram a relação entre duas variáveis numéricas (x e y).

Objetivo

Identificar se existe dependência ou correlação entre as variáveis.

Aplicação

Descobrir padrões e relações matemáticas para possíveis previsões.

Seaborn Scatter Plots: a visual guide

Installation

Quickstart

API Reference

Examples

```
SSACSAAXIC PP1SC1}
festafoes tcail},--lorellul},ctoulietor},tlinnti,eucl},
accnoeeinoeetr--tustane.(),
atininethloegel,-iecco:
ui luns_ccafiton0:
scitcunaatorttuttonolascstenli,()
a sturesdeerr.(i,)*taliduneged>()
a uitunhctstec--(at_ordcem.atucoatontiatuicction),)
0 moti,esalidehociteli:
5 dohcstti,(itces=ttctoidd),(stteytilon(),
1 ancatoxtettin}te.lasainistoi.dli:
2 suateo,} Catei:au.-cte-recst.()-(),
a lttuocokilianlie-ll--filiatto()ctattt
a jll, fietstiyrekstetutcotestcellis,te;
a scattutotcotettotc,}, tcontoftocdcotenc,,tcftt:
1 dwtthii,(utafllstcatactnflstcttorccutid-De..),
a erectoloverlals_cutto-(),
I teoñotintel-ll},
m destokttetaisghisttabldlsterlio).()
```



X

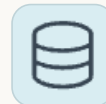


Y



HUE

Criando Gráficos de Dispersão



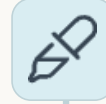
Seleção de Variáveis

Escolha duas variáveis numéricas para comparar (ex: altura e peso).



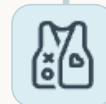
Plotagem Básica

Use `sns.scatterplot()` para criar o gráfico inicial.



Adição de Categorias

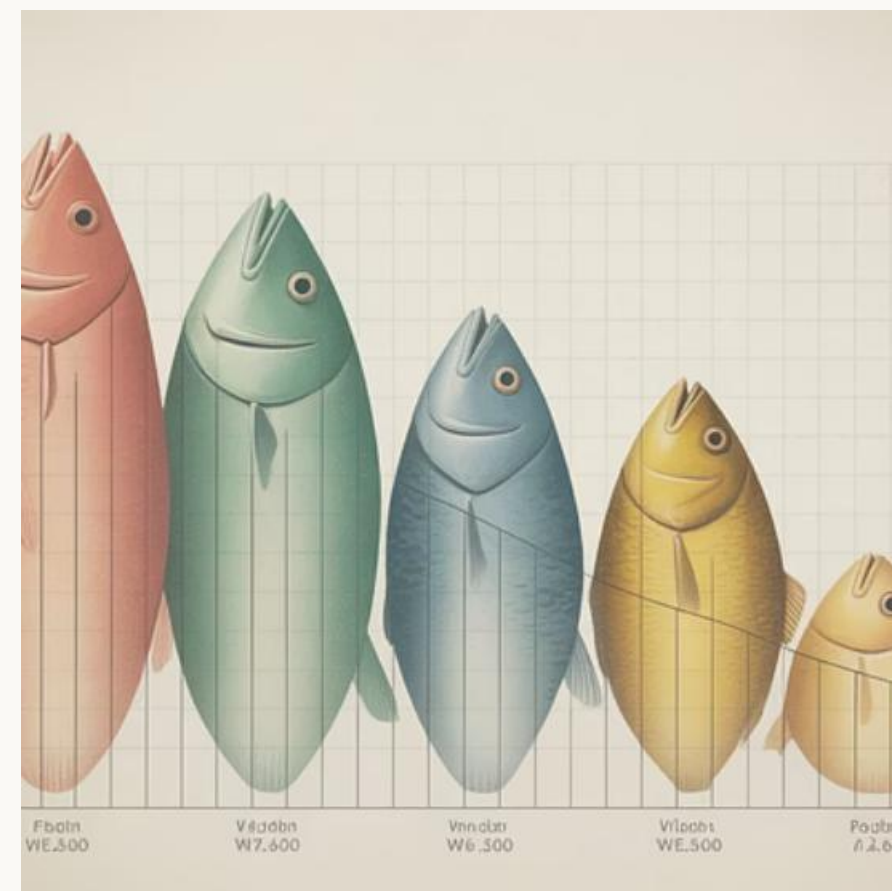
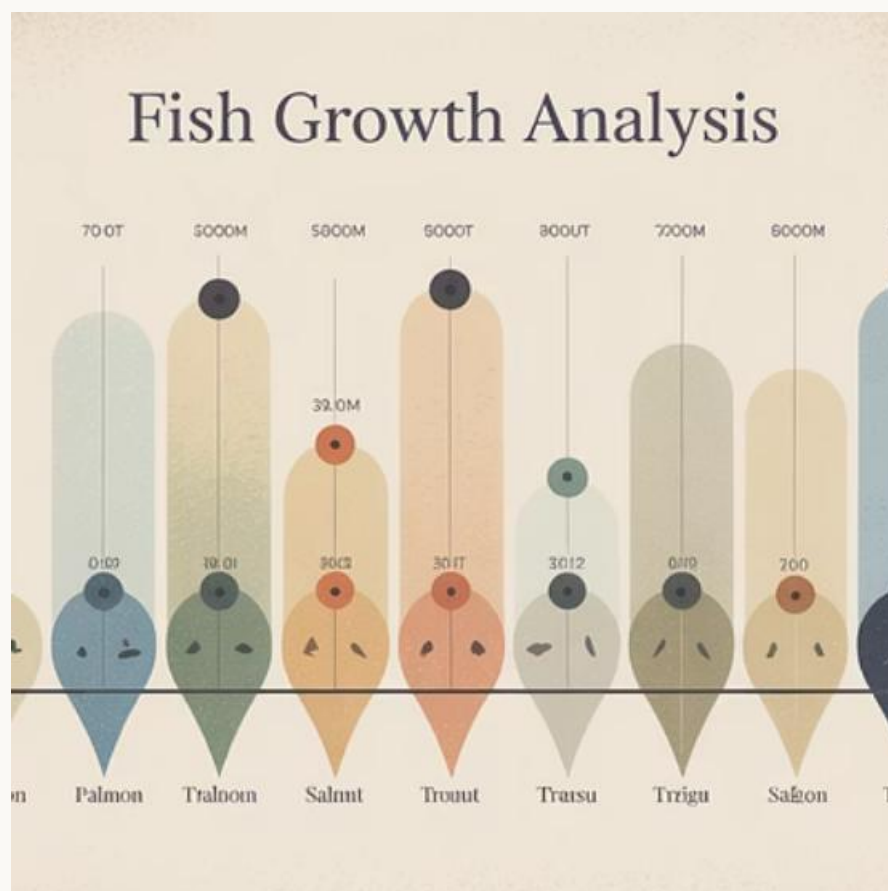
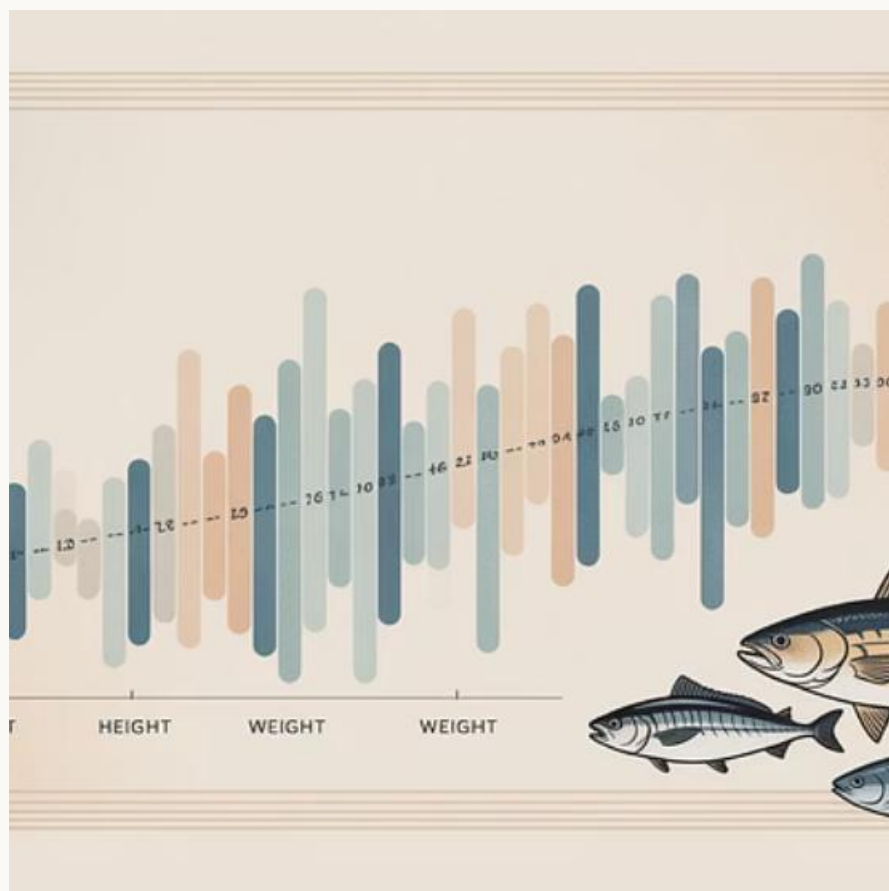
Utilize o parâmetro `hue` para diferenciar categorias por cores.



Diferenciação por Marcadores

Aplique o parâmetro `style` para usar diferentes símbolos por categoria.

Exemplo: Dispersão Altura vs Peso



Os gráficos de dispersão revelam uma clara relação positiva entre altura e peso dos peixes. Algumas espécies se destacam por formarem grupos distintos no gráfico.

Análise da Relação Altura vs Peso



Correlação Positiva

Peixes mais altos tendem a ser mais pesados



Agrupamento por Espécie

Cada espécie forma um grupo distinto no gráfico



Identificação de Outliers

Pontos fora do padrão são facilmente identificáveis

FRESHWATER
TRIO

Espécies que se Destacam

1

Perch

Maior variabilidade em peso e altura

2

Pike

Relação altura/peso diferenciada

3

Bream

Grupo bem definido com maior
altura média

Pairplot: Visualização Multivariada

Definição

Matriz de gráficos que compara todas as combinações possíveis de variáveis numéricas.

Permite visualizar múltiplas relações simultaneamente.

Componentes

Diagonal: Gráficos de densidade para cada variável.

Demais posições: Gráficos de dispersão entre pares de variáveis.

Criando um Pairplot

Implementação Simples

Use `sns.pairplot(dataframe)` para criar toda a matriz de gráficos.



Personalização

Adicione `hue` para diferenciar categorias por cores.



Eficiência

Evita criar múltiplos gráficos individualmente.

```
z* *""ptyhlon)""pusalorparp)lo)""
(  " suabor" * (oeo.Jocon.o)oh "()"
os  snaberoqurn ""
ie  sotole"()ueasso) ""
dn  *yate.)e + *nc""
t    ( yeyute")<
du  sns.7 "ouu7soded dtotoolelo)"))
oo  do+otaberplo))"
t    ""solortcdeurplot")")ci
t    dos.ceE. "paion.oipicele"
nn  s.)pte queso)""= , t"
de  (odelo) dus+
t    (, ")oloe)" fa )*
    (.ios "oualee .doioncatunletereo(
    (tlos. doseacocd.)" , ,
    "Γ stout "cin])c" { olofoopl lo(
    !    ts. =(ac
d
```


Vantagens do Pairplot



Visão Abrangente

Visualiza todas as relações entre variáveis de uma só vez.



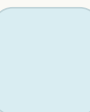
Identificação de Padrões

Facilita a descoberta de correlações não óbvias.



Economia de Tempo

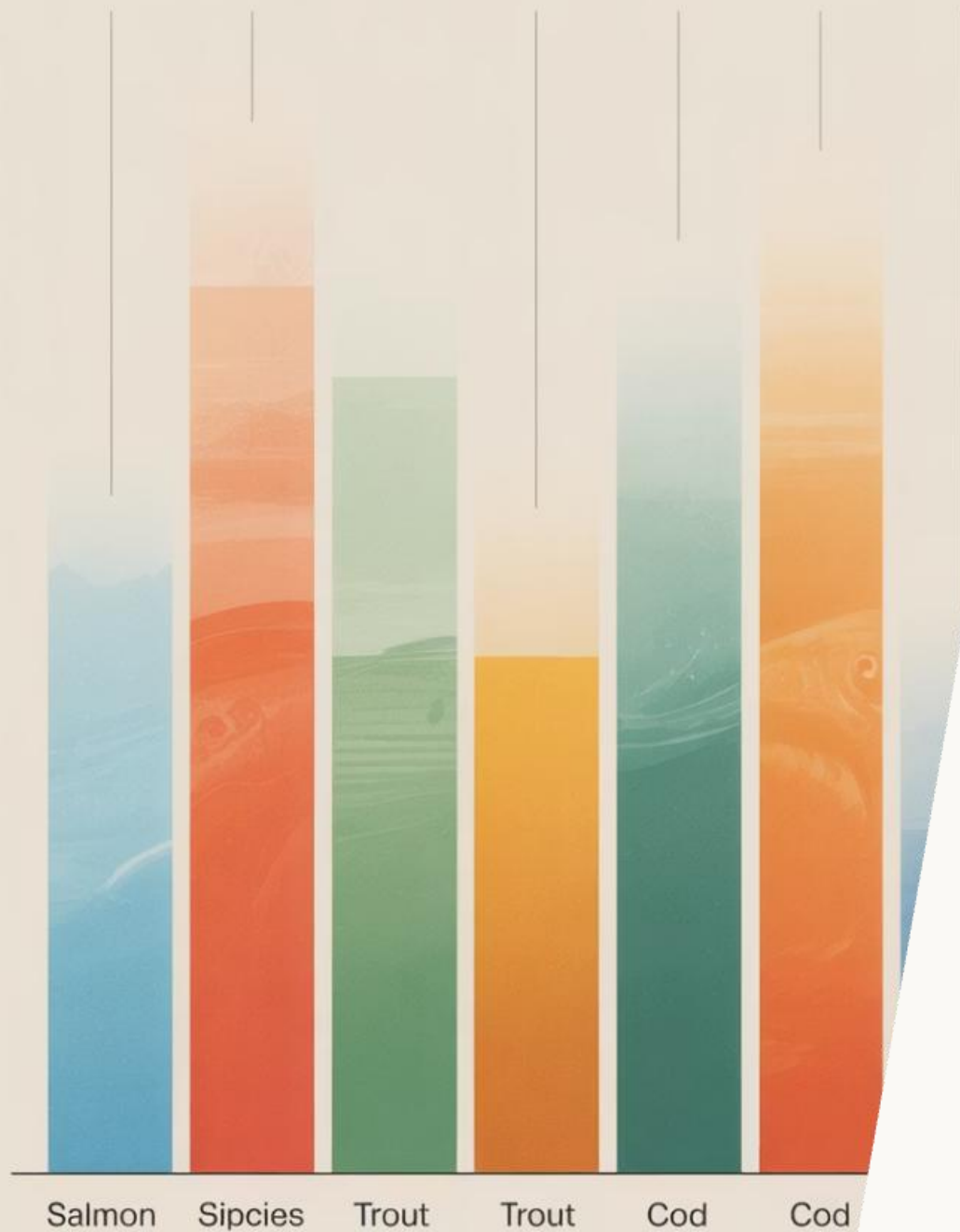
Evita a necessidade de criar múltiplos gráficos separados.



Insights Rápidos

Permite uma análise exploratória inicial eficiente.





Gráficos de Barras: Conceito

Definição

Representação visual que usa barras retangulares para comparar categorias.

Aplicação

Ideal para visualizar a frequência de ocorrência de categorias.

Implementação

Use `sns.countplot(data=dados, x='coluna')` para criar gráficos de barras.

Gráficos de Setores (Pizza)

Definição

Círculo dividido em fatias proporcionais aos valores

Limitação

Menos eficaz com muitas categorias



Propósito

Mostrar proporções relativas entre categorias

Aplicação

Ideal para visualizar partes de um todo

Criando Categorias para Análise



Peixes Leves

Até 550g



Peso Médio

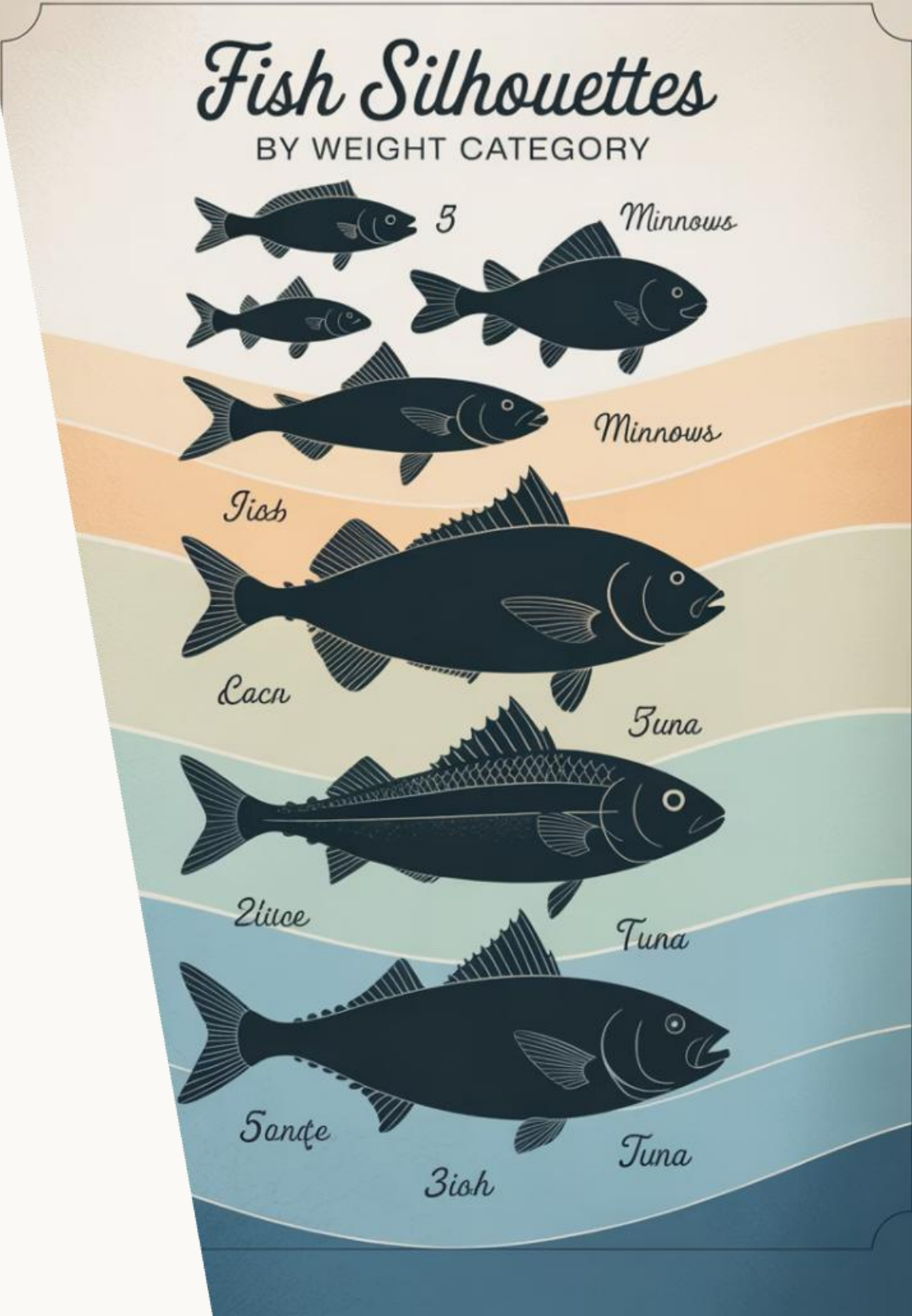
Entre 550g e 825g



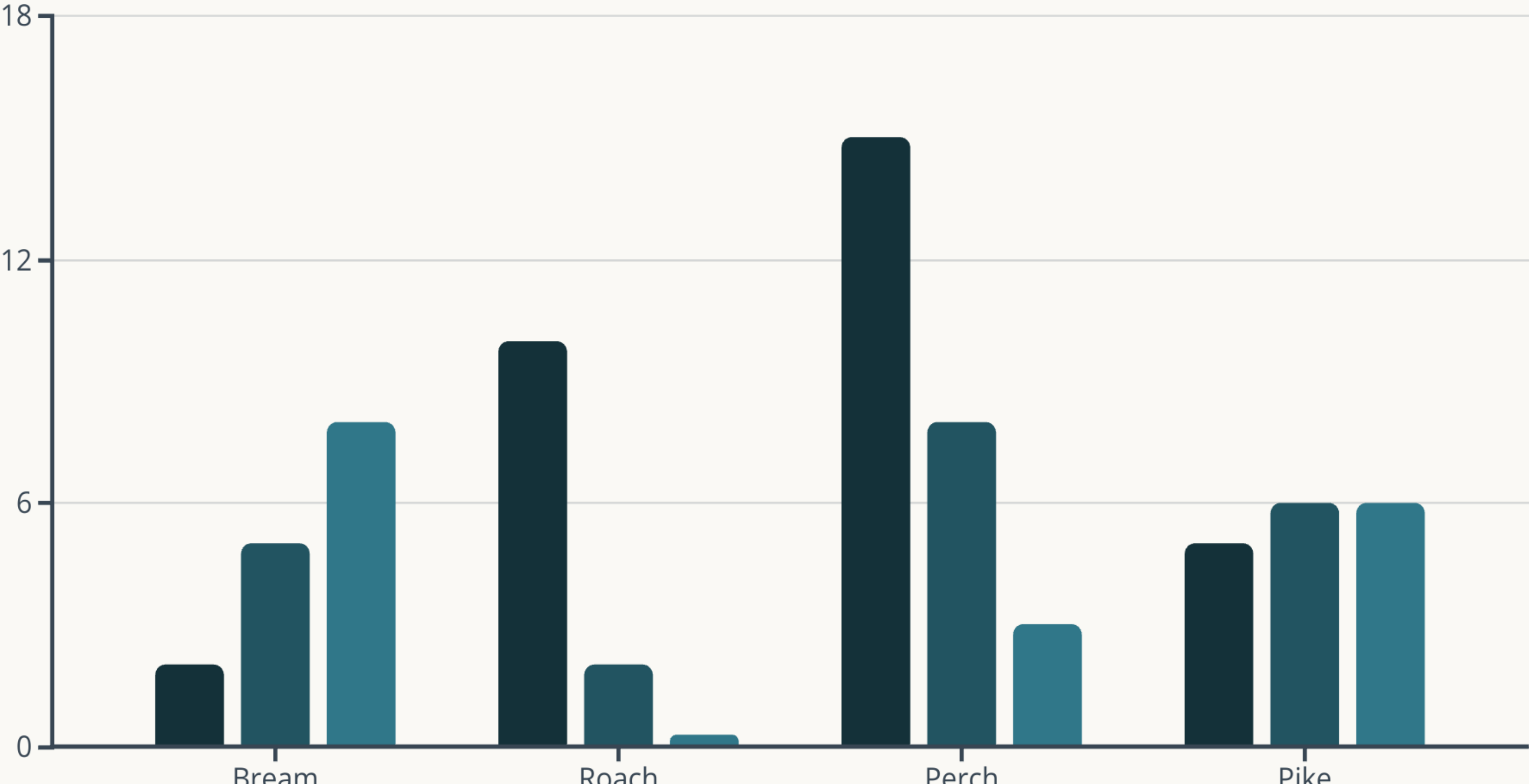
Peixes Pesados

Acima de 825g

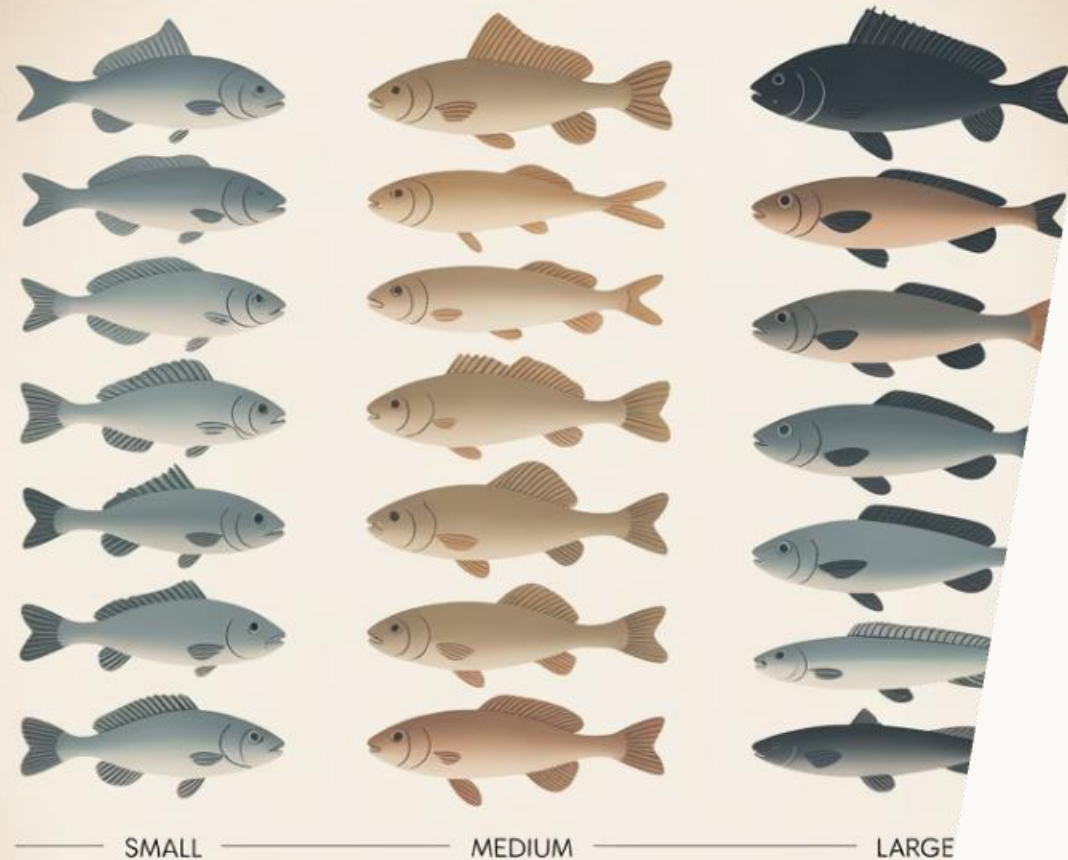
Para enriquecer a análise, criamos uma nova coluna categórica baseada no peso dos peixes. Esta categorização permite explorar relações entre o peso e outras características.



Gráficos de Barras com Subgrupos



Fish Weight Categories



Análise do Gráfico de Barras com Subgrupos

1

Bream

Predominância de peixes pesados, indicando maior massa corporal média.

2

Roach

Maioria dos peixes são leves, sem representantes na categoria pesados.

3

Perch

Distribuição concentrada em peixes leves, com poucos exemplares pesados.

4

Pike

Distribuição mais equilibrada entre as três categorias de peso.

Subplots: Múltiplos Gráficos em Uma Imagem

Definição

Técnica para criar vários gráficos em uma única figura.

Permite comparação direta entre diferentes visualizações.

Implementação

Use `plt.subplots(linhas, colunas, figsize=(x,y))`

Cada subplot pode conter um tipo diferente de gráfico.

Criando Subplots em Python



Definir Grade

`fig, axes = plt.subplots(linhas, colunas, figsize=(15, 15))`



Criar Gráficos

Adicione visualizações a cada posição da grade.



Adicionar Títulos

Nomeie cada subplot para facilitar a interpretação.



Ajustar Layout

Use `plt.tight_layout()` para melhorar o espaçamento.

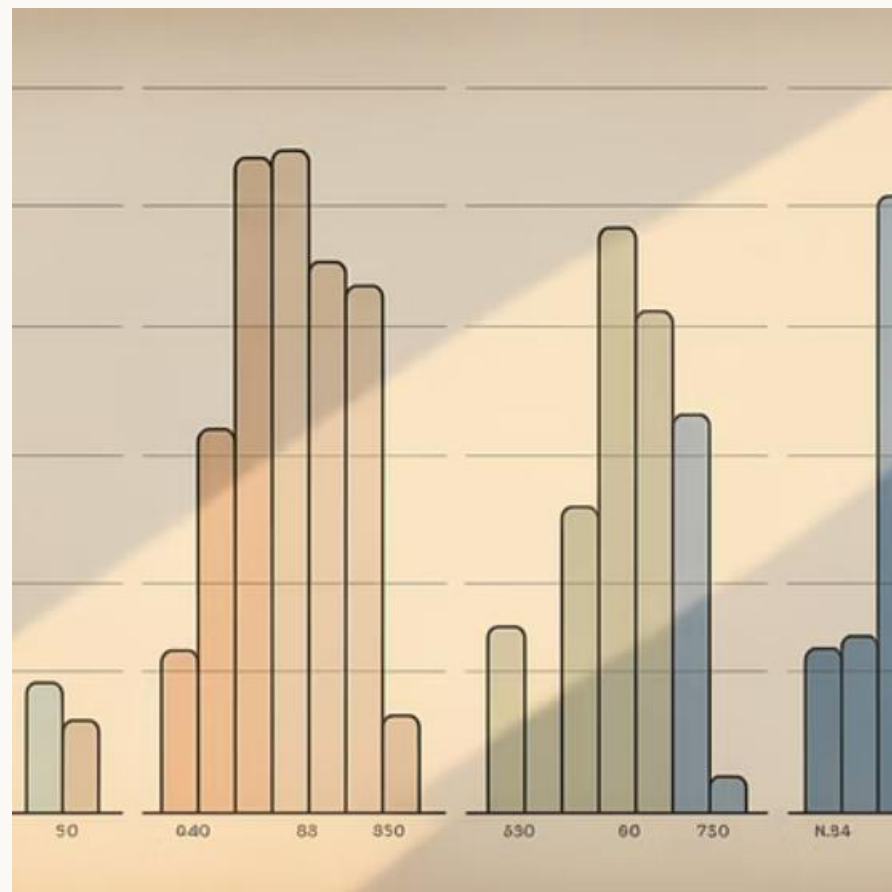
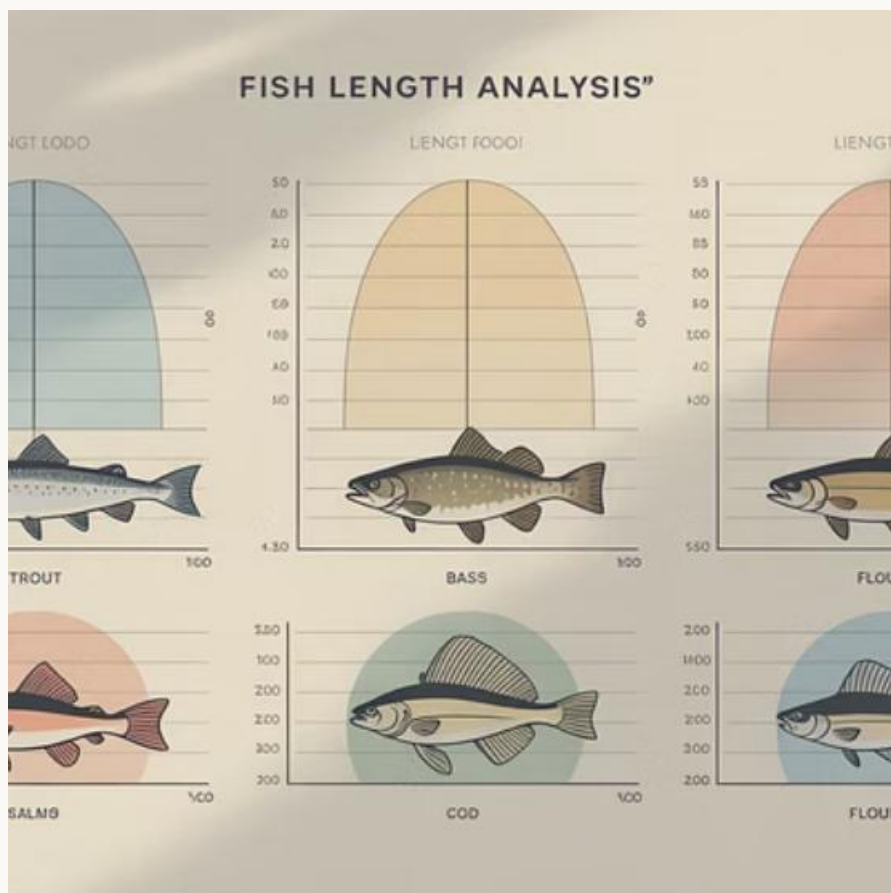
Matplotlib*

```
(CornpelloJel(Γ{"agltol"})(  
    un cacartsezn be to  
    vn acabnh;() ty lq  
    wil noraune((ny ab|neolol)),  
    ro navo deΓ"o,  
    pr Thot,-fccun ealt; (T"█:-█:-█)  
    ra ac|;  
    a tnenljld e"-corplole|l; " men'Ze ..TIn-Γ"  
    a Thne derlille|rk)  
    Γ"coper atyoe|elsl,tt|p|T  
    aus;..|re de cupr(-?toon nt|olorl,l|elz()(h)  
    us unyoe derfere, ploerc("e) snal(Γ"l  
    ndoyt.
```

Explore the versatility of
matplotlib's subplot function

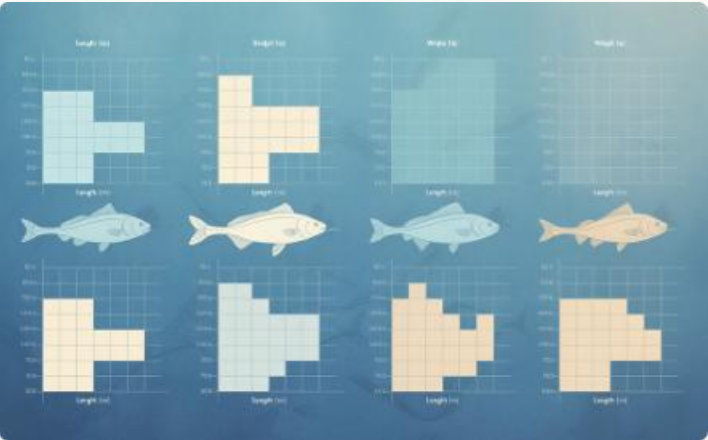
Try it now

Exemplo: Histogramas por Espécie



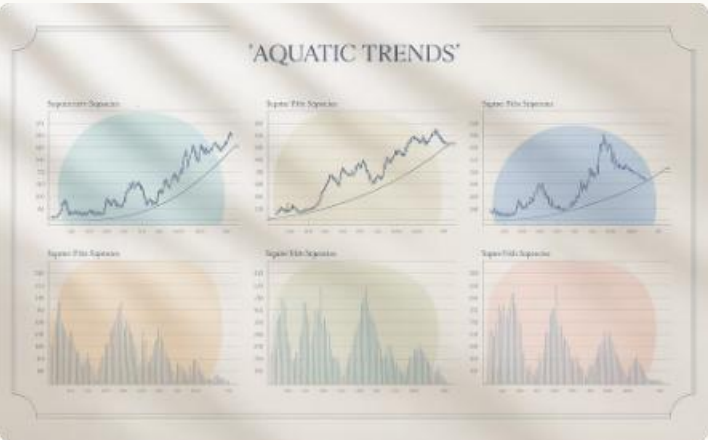
Usando subplots, podemos criar histogramas separados para cada espécie de peixe, facilitando a comparação das distribuições de peso entre diferentes tipos.

Exemplo: Gráficos de Dispersão Comparativos



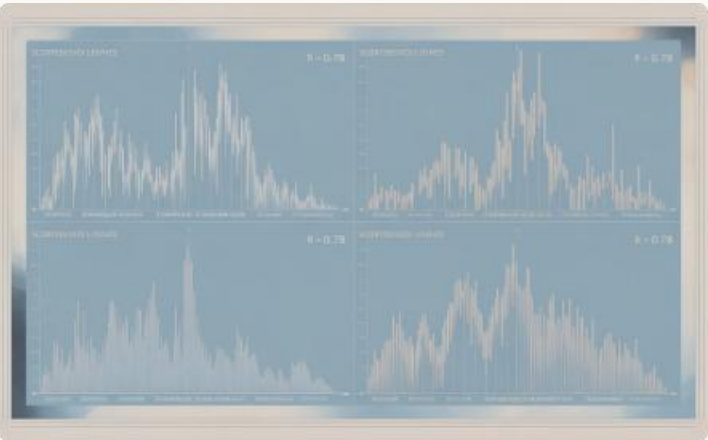
Múltiplas Relações

Comparação de diferentes pares de variáveis lado a lado.



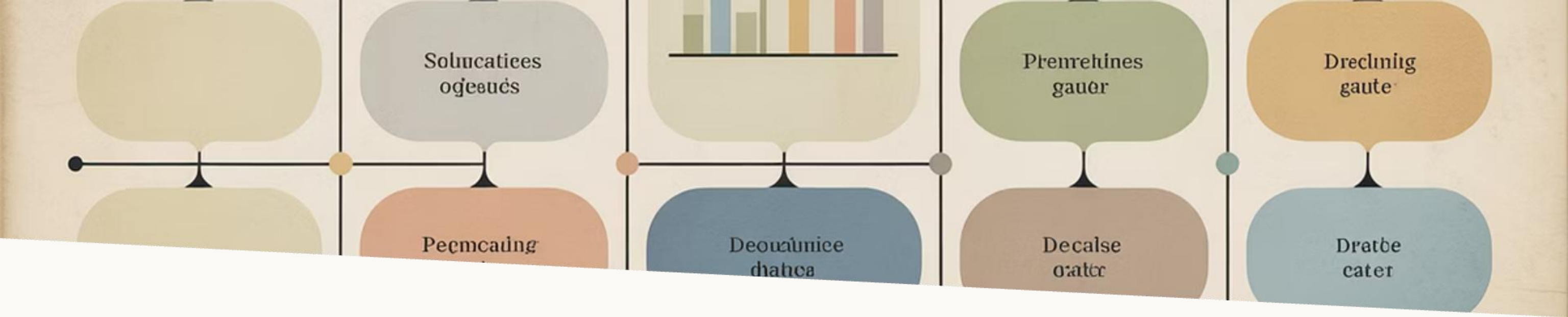
Análise por Espécie

Visualização da mesma relação para diferentes espécies.



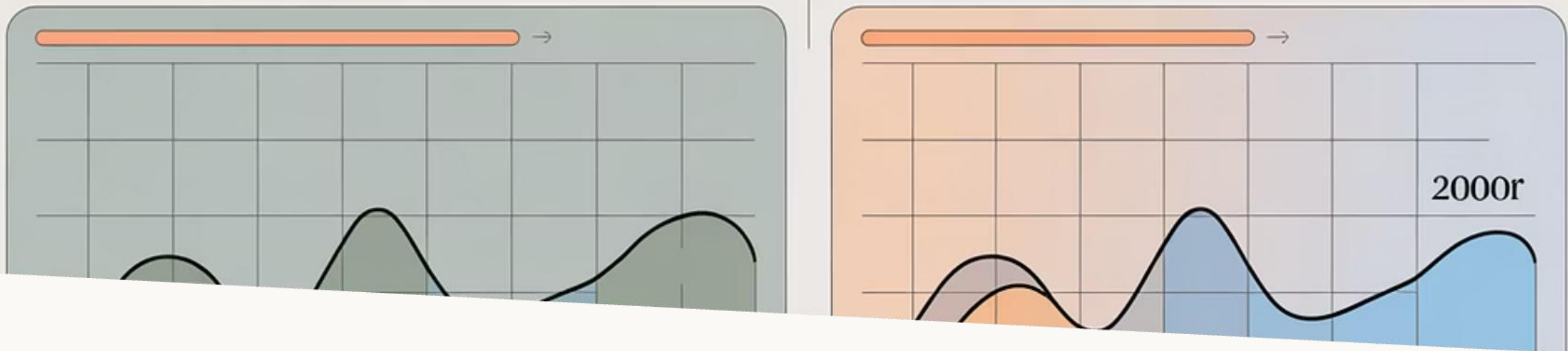
Com Estatísticas

Adição de linhas de tendência e coeficientes de correlação.



Escolhendo o Gráfico Adequado

Tipo de Dado	Objetivo	Gráfico Recomendado
Uma variável numérica	Distribuição	Histograma ou Densidade
Duas variáveis numéricas	Relação/Correlação	Dispersão
Variável categórica	Frequência	Barras ou Setores
Múltiplas variáveis	Visão geral	Pairplot ou Subplots

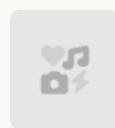


Boas Práticas em Visualização de Dados



Objetivo Claro

Defina o que deseja comunicar antes de escolher o gráfico.



Simplicidade

Evite elementos desnecessários que distraem da mensagem principal.



Rotulagem Adequada

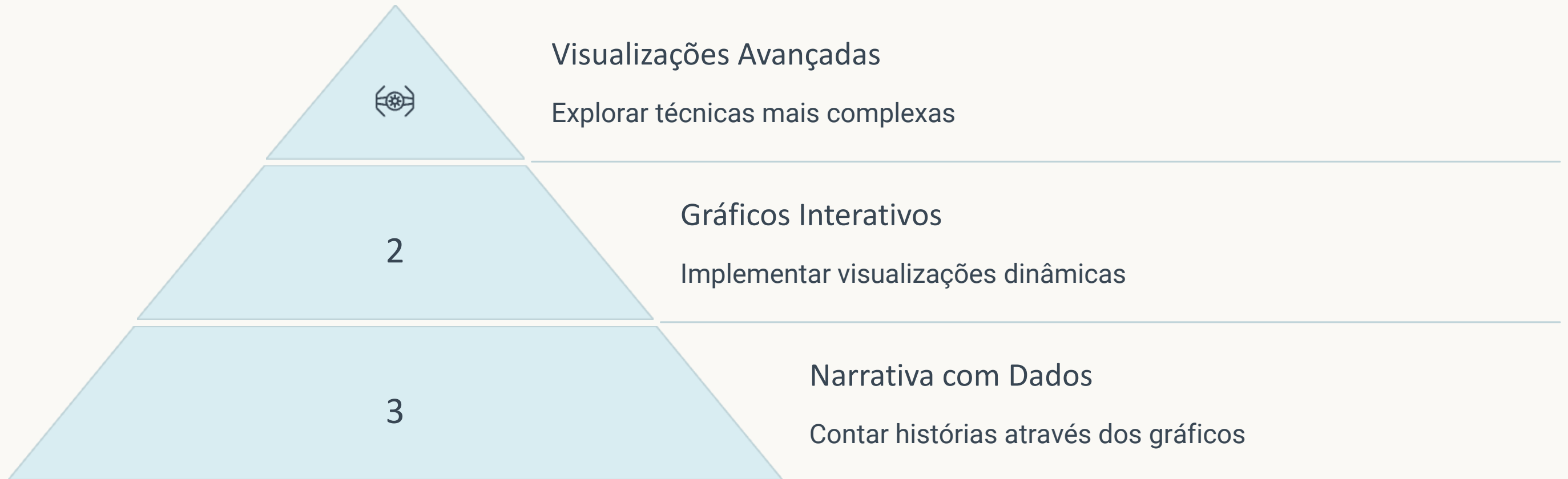
Inclua títulos, legendas e unidades para facilitar a interpretação.



Cores Consistentes Consistentes

Use paletas de cores que facilitem a distinção entre categorias.

Conclusão e Próximos Passos



Nesta apresentação, exploramos diversos tipos de gráficos essenciais para análise de dados. Dominando estas técnicas de visualização, você estará preparado para extrair insights valiosos de qualquer conjunto de dados e comunicá-los efetivamente.