

Como apresentar informações

Fazendo gráficos que as pessoas possam entender

Cássia Sampaio Sanctos
CONE 2021



A close-up, slightly blurred image of a pile of light-colored wooden blocks. The blocks are of various sizes and are scattered haphazardly. Many of the blocks have black letters or numbers printed on them. Visible characters include 'M', 'R', 'C', 'N', 'J', 'P', 'Z', 'W', 'S', 'A', '2', '1', '3', '8', and '4'. The lighting is soft and even, highlighting the natural wood grain of the blocks.

Primeira situação

Estamos felizes & contentes fazendo uma modelagem e chegamos nestes resultados

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k-vizinhos mais próximos	0.84815

Qual acurácia é maior?

3 segundos

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k-vizinhos mais próximos	0.84815

Será que a gente consegue em 1 segundo?

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k-vizinhos mais próximos	0.84815

Sabe como fica mais direto ainda?

O modelo de **floresta aleatória**
teve a **maior acurácia**,
de **86%**,
entre os 7 modelos testados.

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
mvs	0.82318
rede neural simples	0.84532
k vizinhos mais próximos	0.84815

Tudo depende de qual é a nossa
informação principal



O que queremos mostrar?

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k vizinhos mais próximos	0.84815

O que queremos mostrar?

Queremos **comparar** valores

modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k vizinhos mais próximos	0.84815

Como podemos fazer isso?

Como podemos fazer isso?

podemos usar a própria tabela
(já vimos como pode ser)

Como podemos fazer isso?

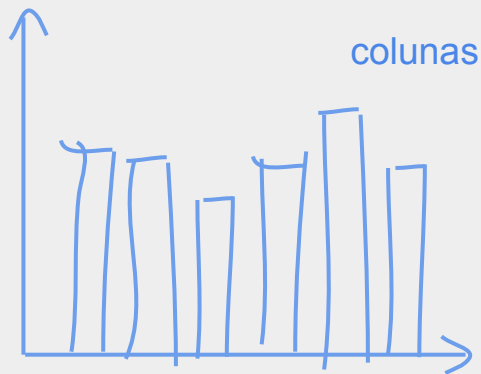
podemos usar a própria tabela
(já vimos como pode ser)

podemos usar algum gráfico que
coloque quantidades lado a lado para
comparamos

Como podemos fazer isso?

podemos usar a própria tabela
(já vimos como pode ser)

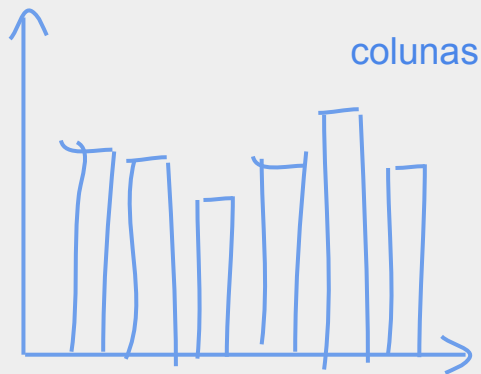
podemos usar algum gráfico que
coloque quantidades lado a lado para
comparamos



Como podemos fazer isso?

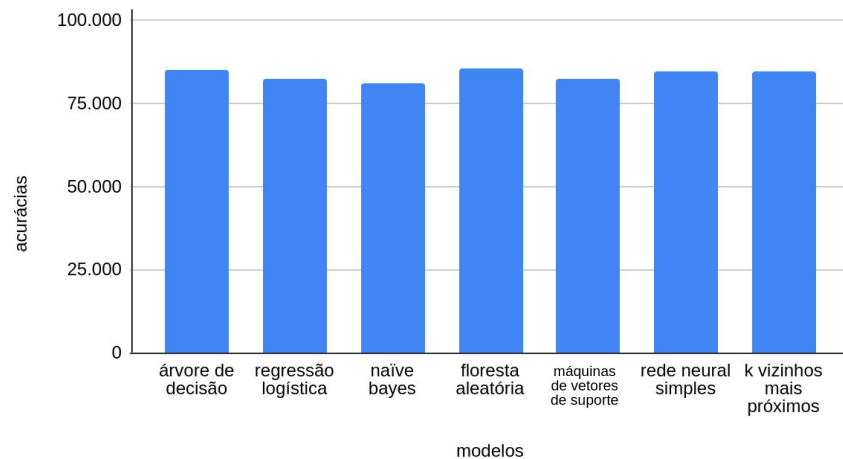
podemos usar a própria tabela
(já vimos como pode ser)

podemos usar algum gráfico que
coloque quantidades lado a lado para
comparamos

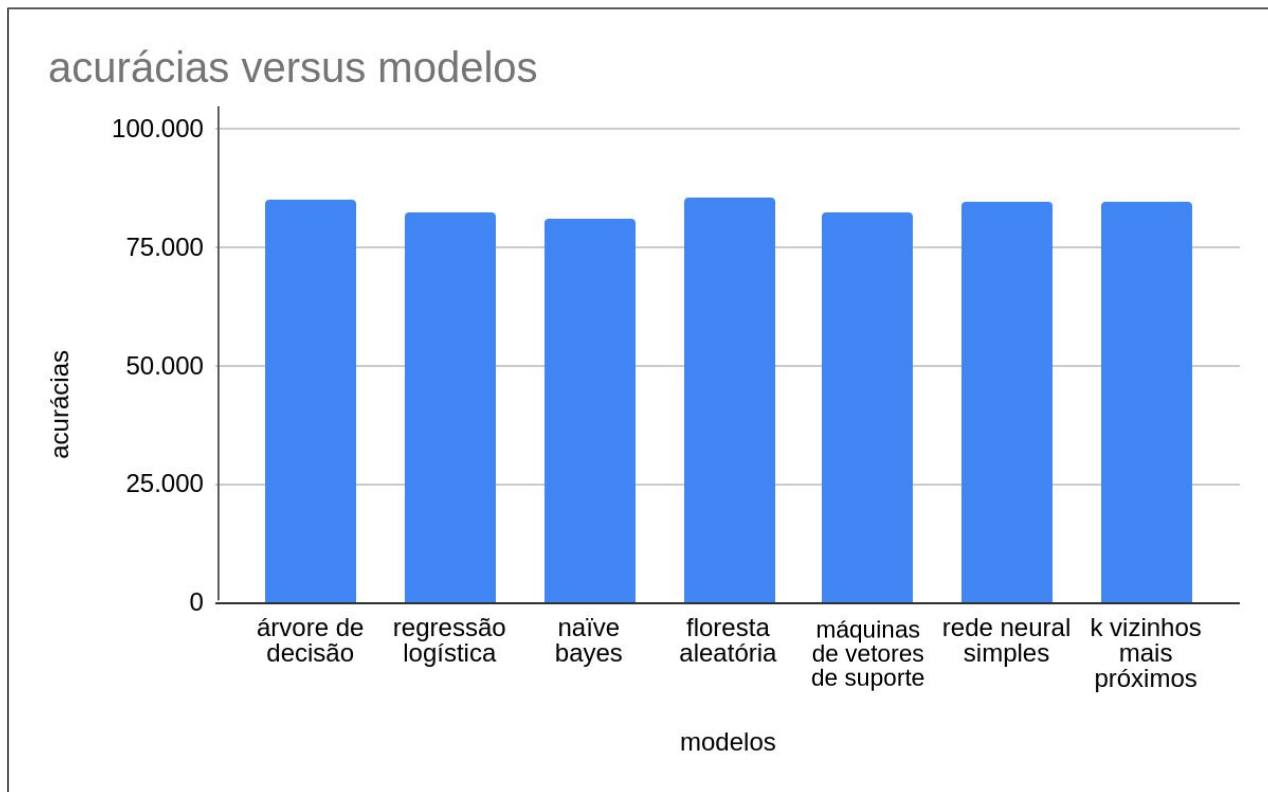


modelos	acurácias
árvore de decisão	0.85090
regressão logística	0.82308
naïve bayes	0.81010
floresta aleatória	0.85558
máquinas de vetores de suporte	0.82318
rede neural simples	0.84532
k vizinhos mais próximos	0.84815

acurácias versus modelos

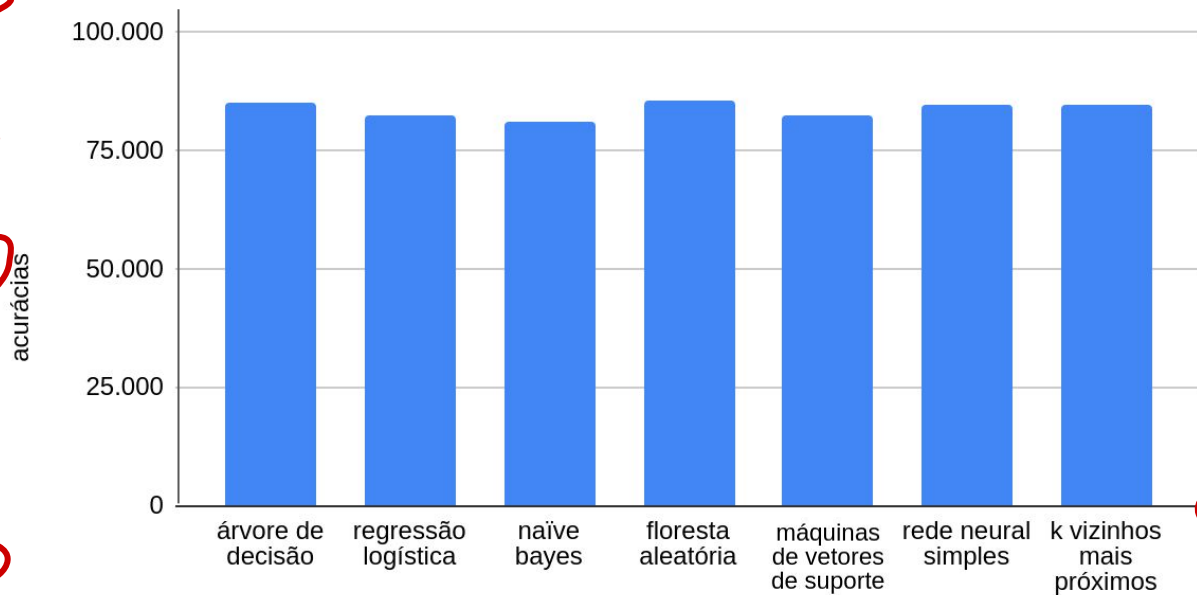


o problema da borda



o problema da borda

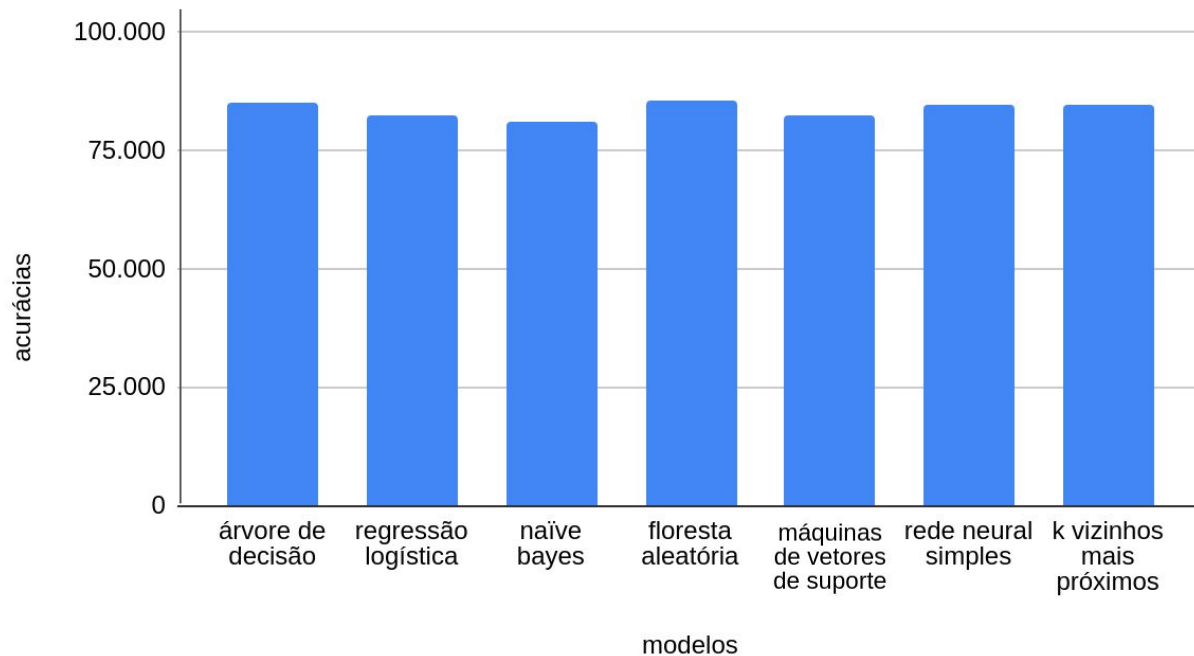
acurácias versus modelos



modelos

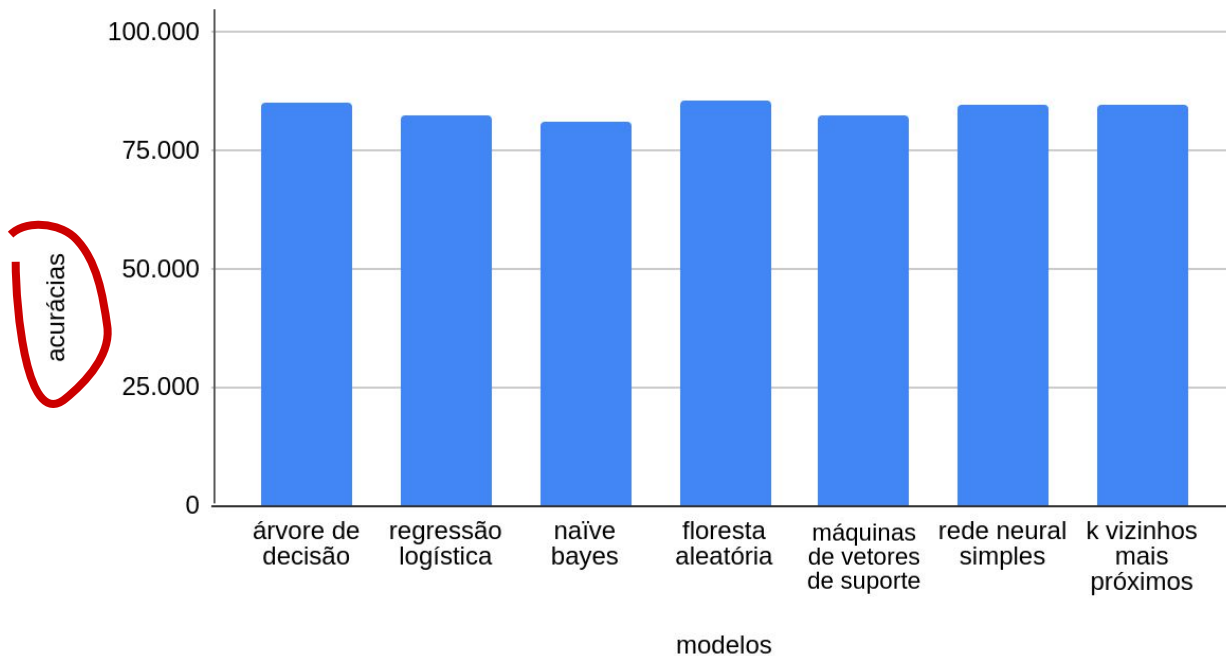
o problema de virar o pescoço
na acurácia

acurácias versus modelos

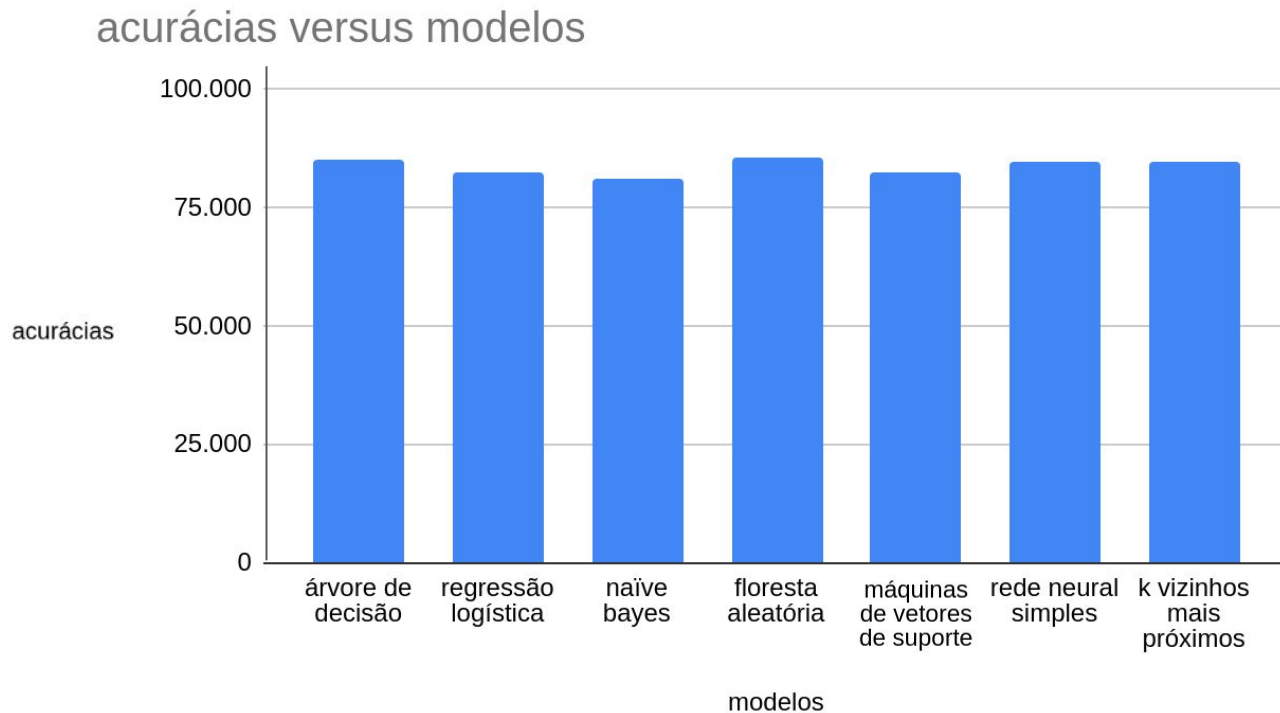


o problema de virar o pescoço
na acurácia

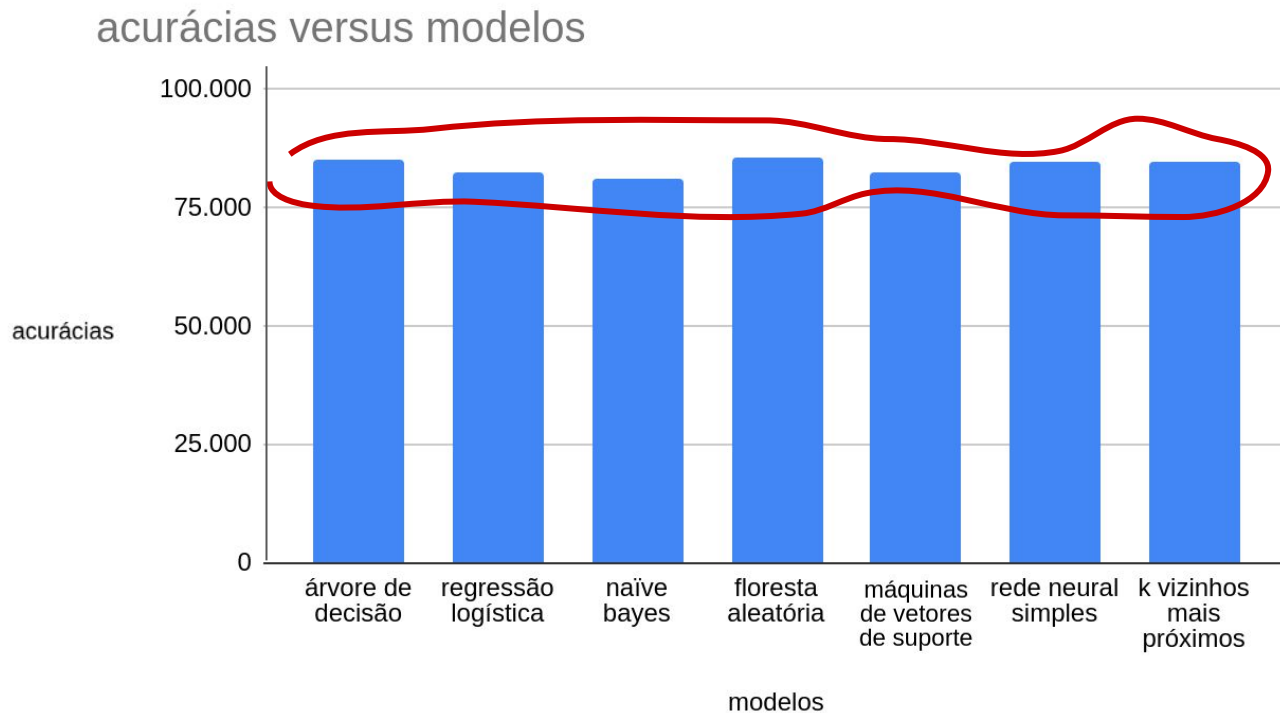
acurácias versus modelos



o problema de qual o número das colunas

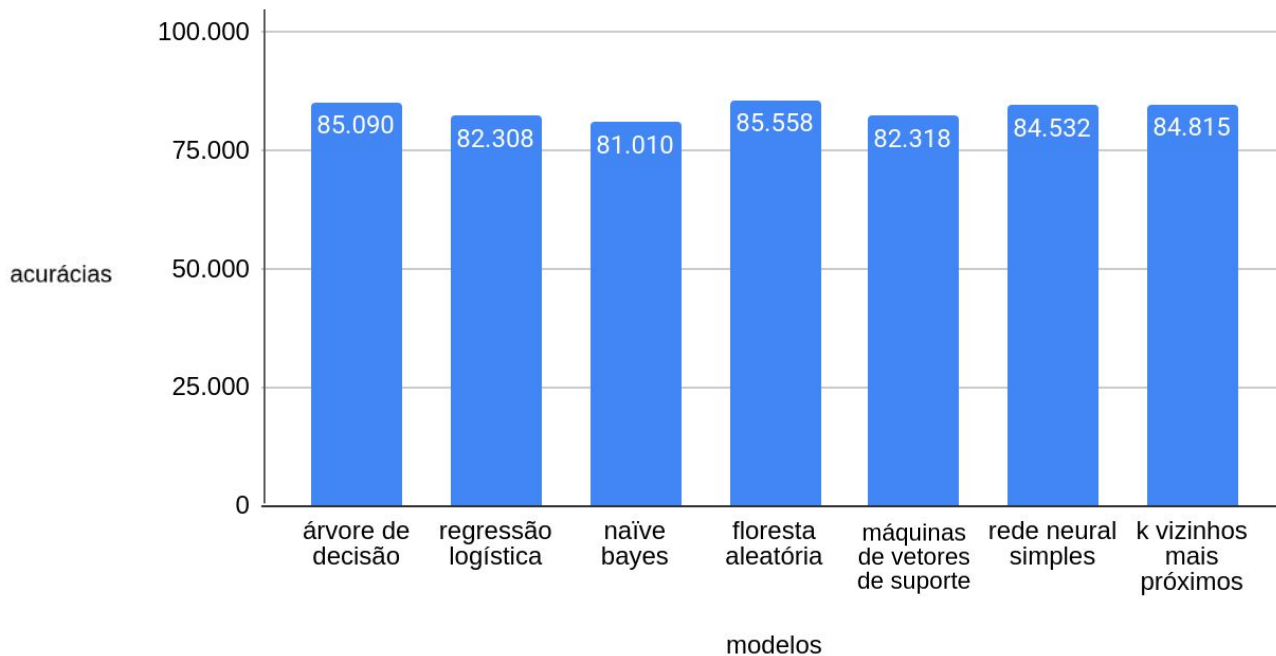


o problema de qual o número das colunas



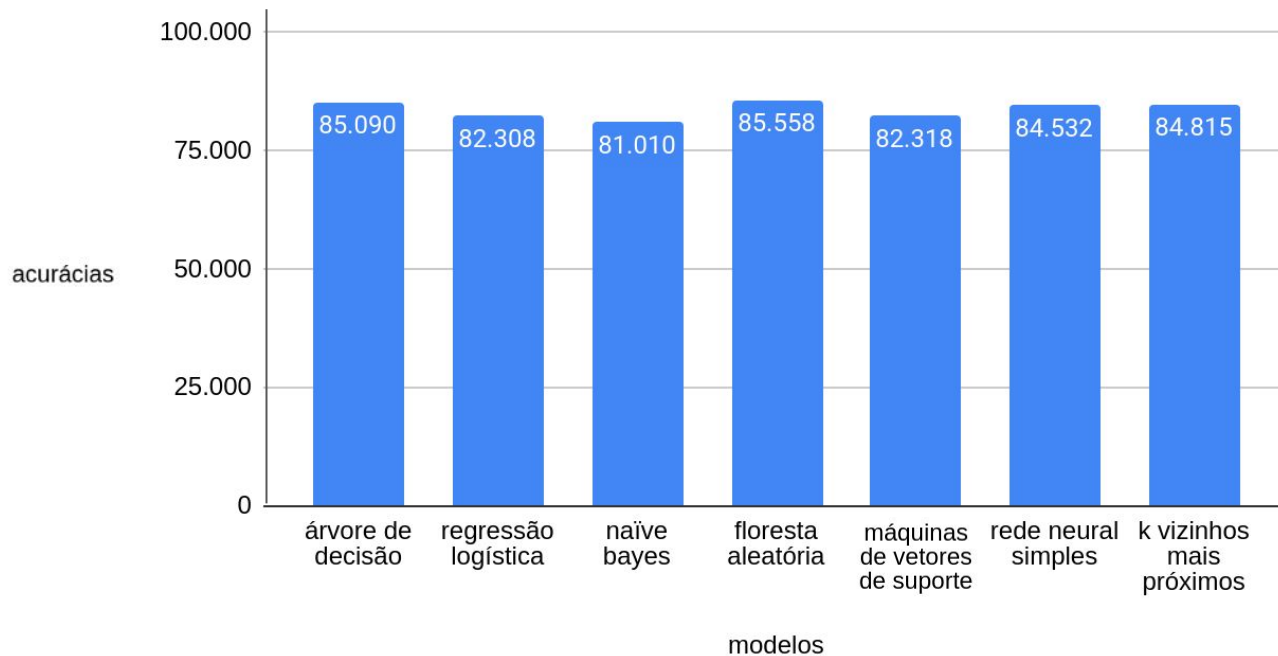
o problema das linhas de grade e das
marcações nos eixos

acurácias versus modelos

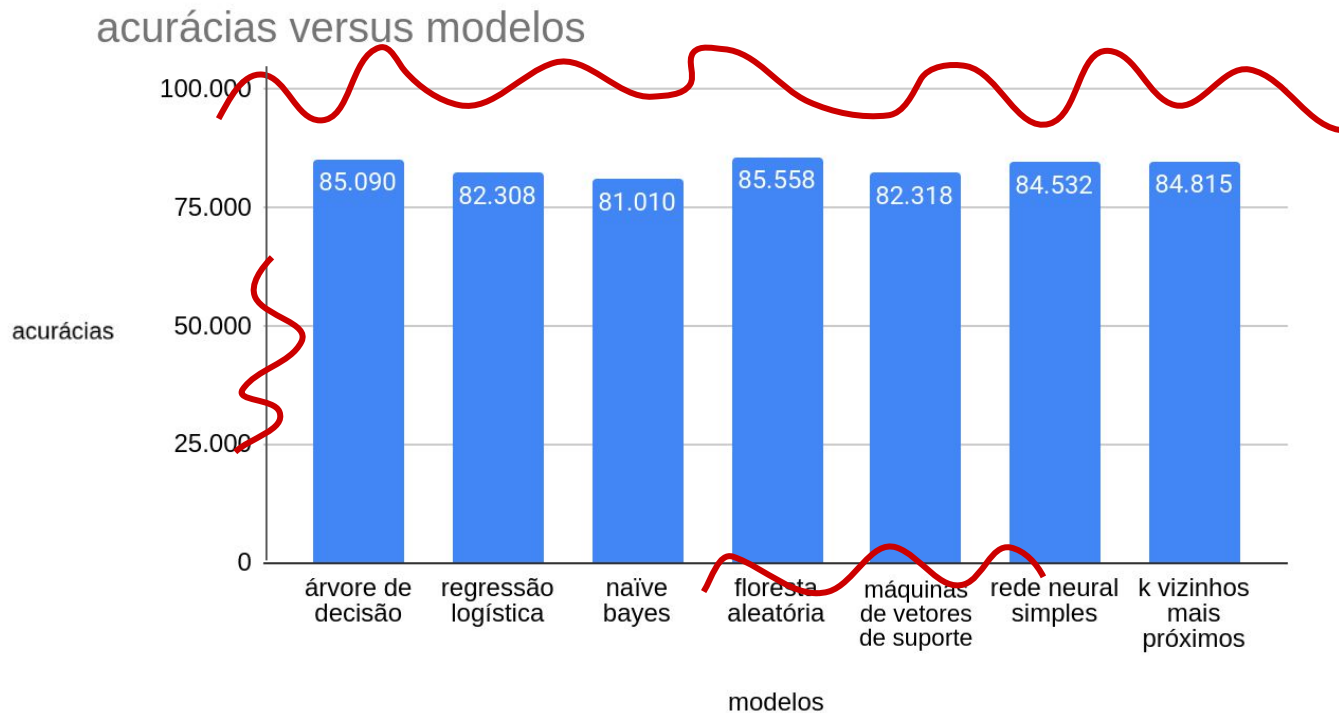


o problema das linhas de grade e das
marcações nos eixos

acurácias versus modelos

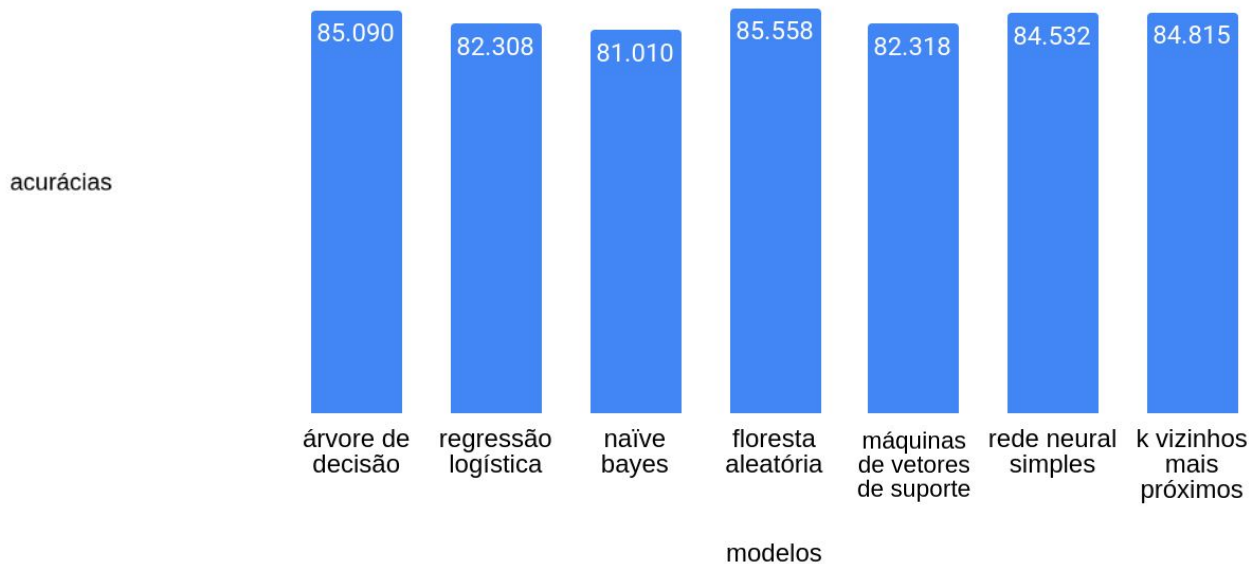


o problema das linhas de grade e das
marcações nos eixos



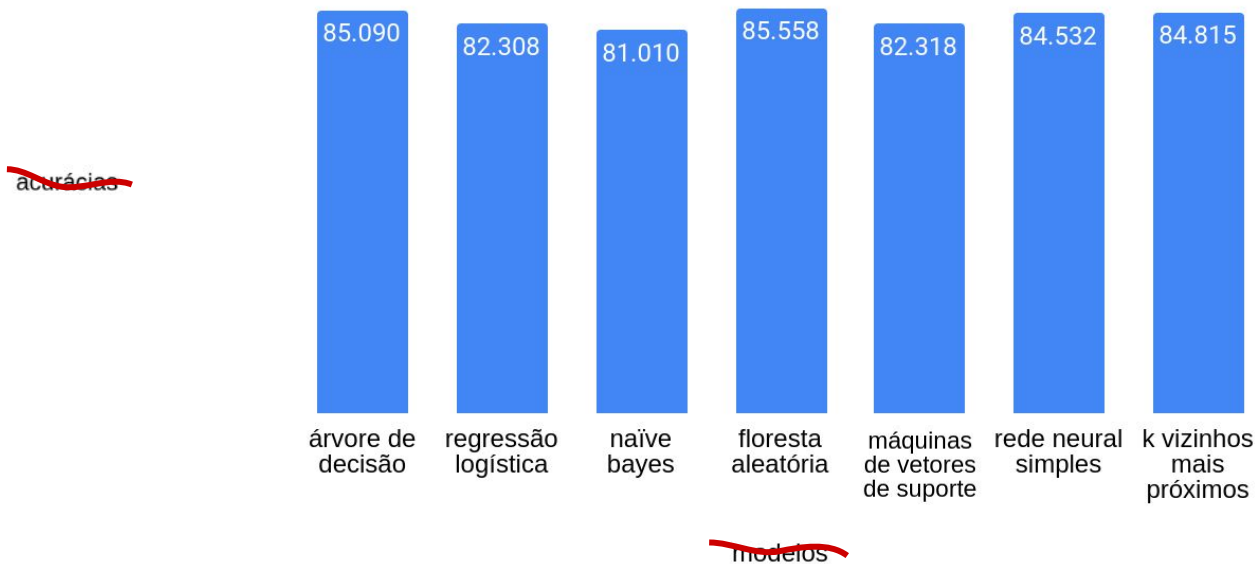
o problema dos títulos dos eixos

acurácias versus modelos



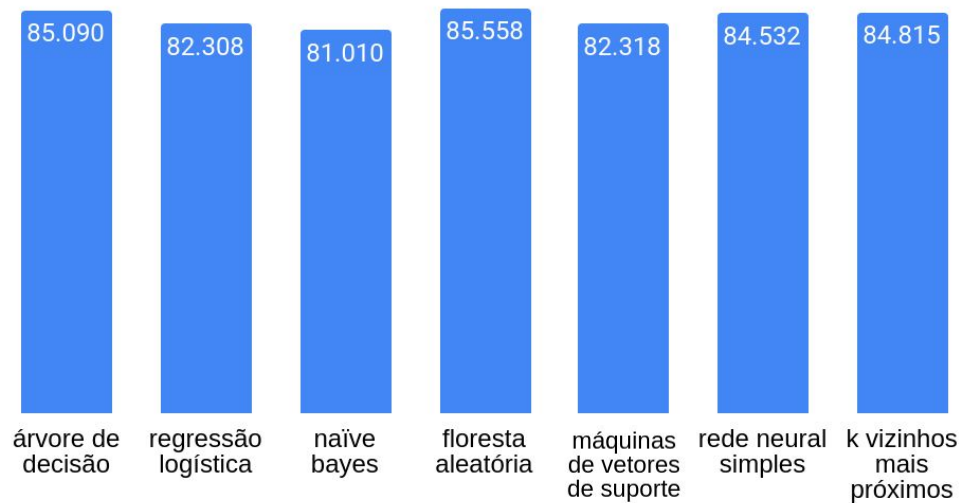
o problema dos títulos nos eixos

acurácias versus modelos



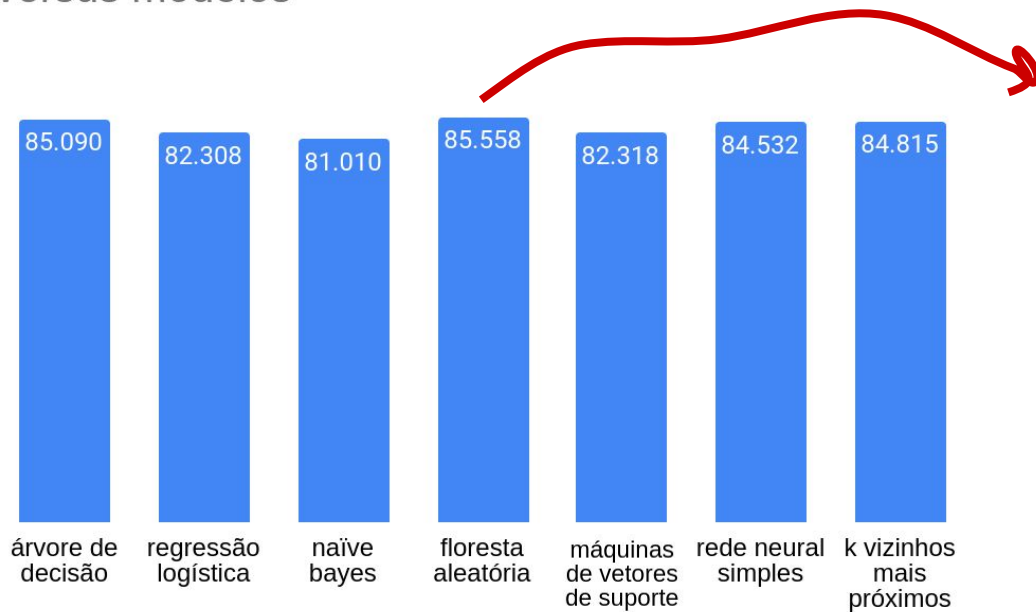
o problema da ordem das colunas

acurácias versus modelos



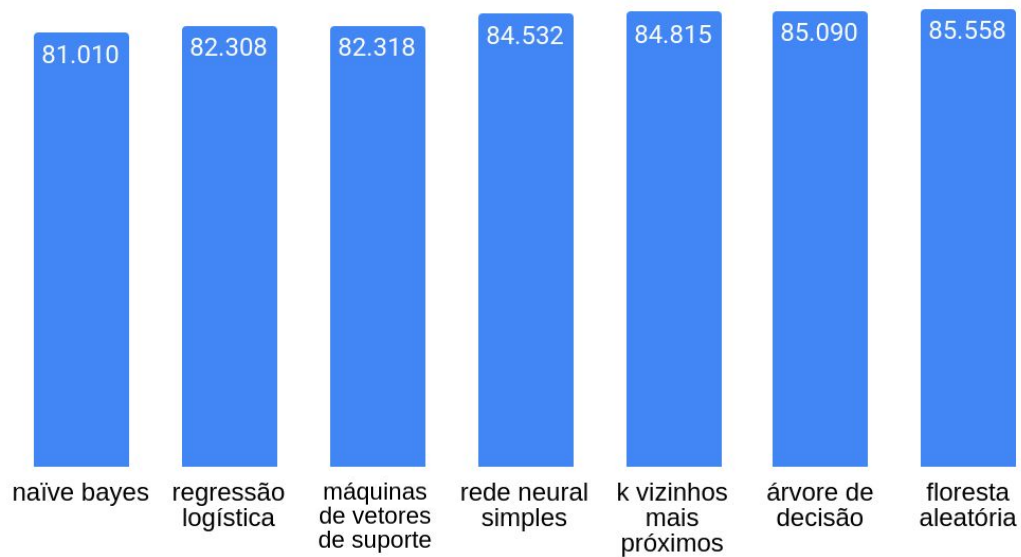
o problema da ordem das colunas

acurácias versus modelos



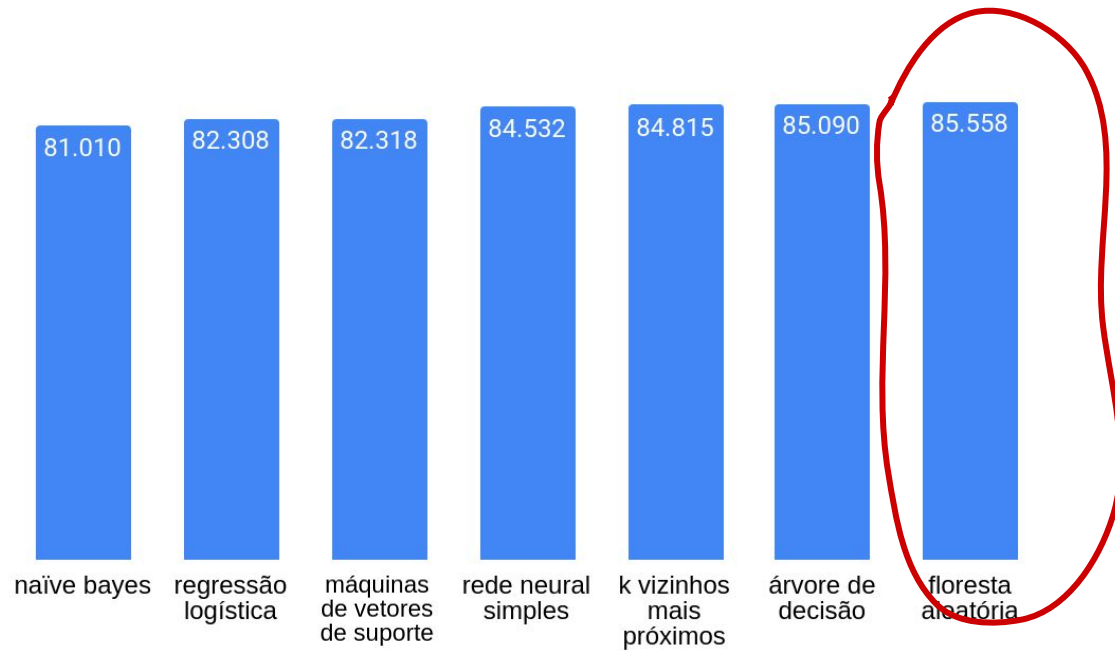
o problema do destaque das colunas

acurácias versus modelos



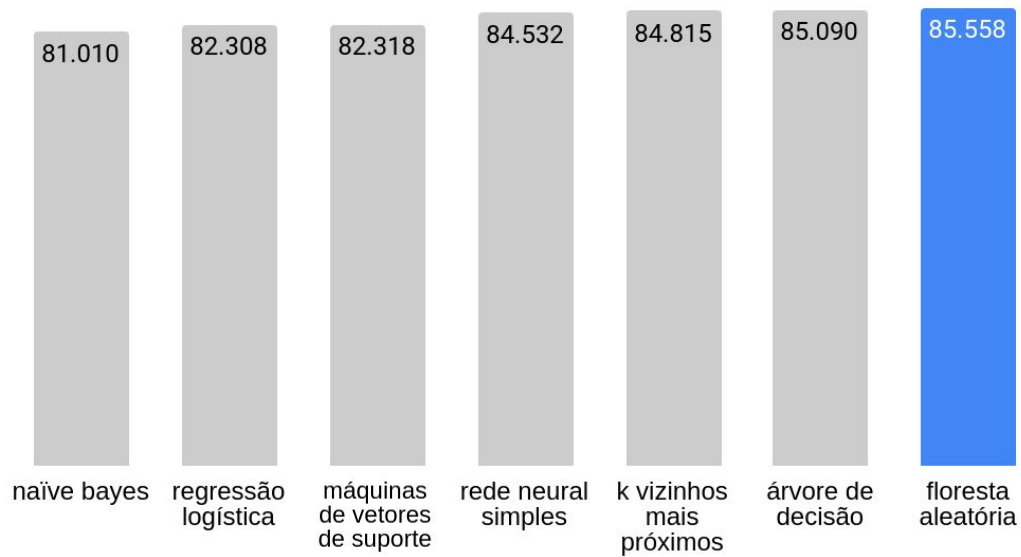
o problema do destaque das colunas

acurácias versus modelos



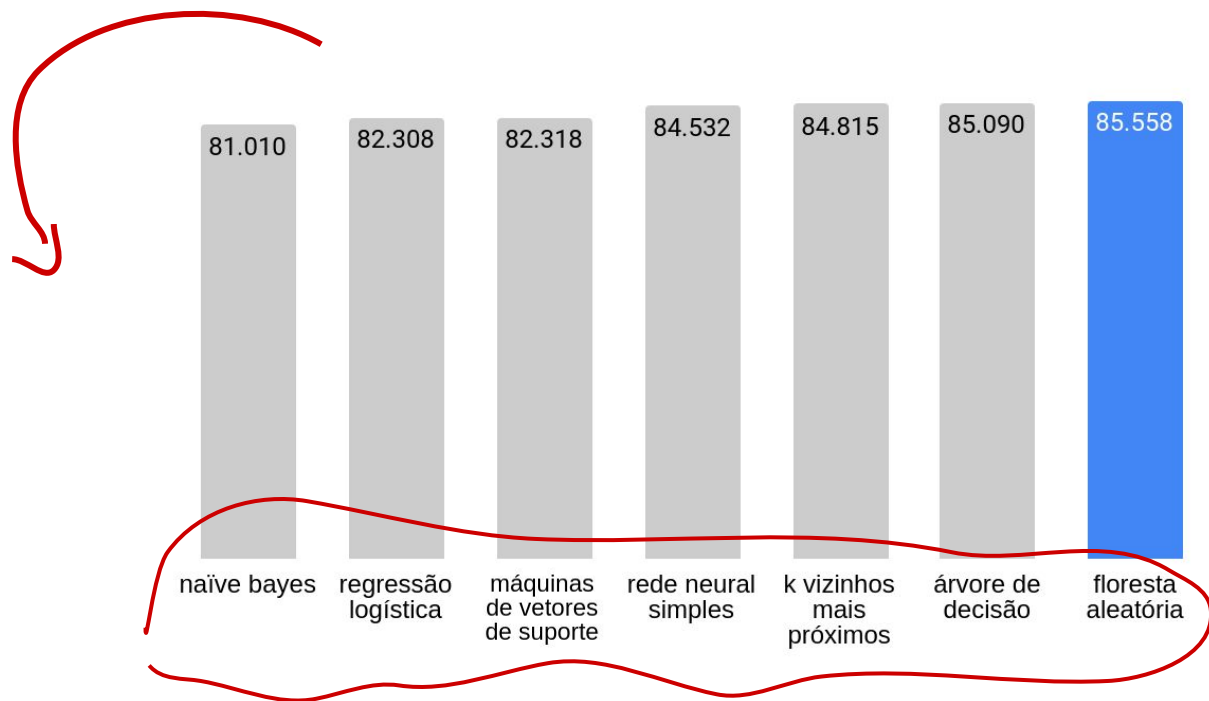
o problema de ler os nomes dos modelos

acurácias versus modelos



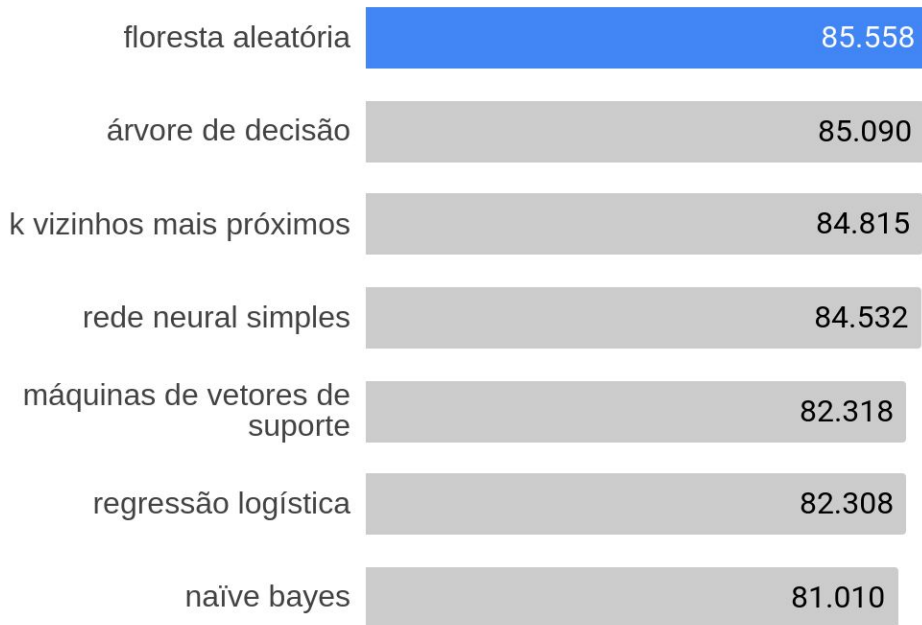
o problema de ler os nomes dos modelos

acurácias versus modelos



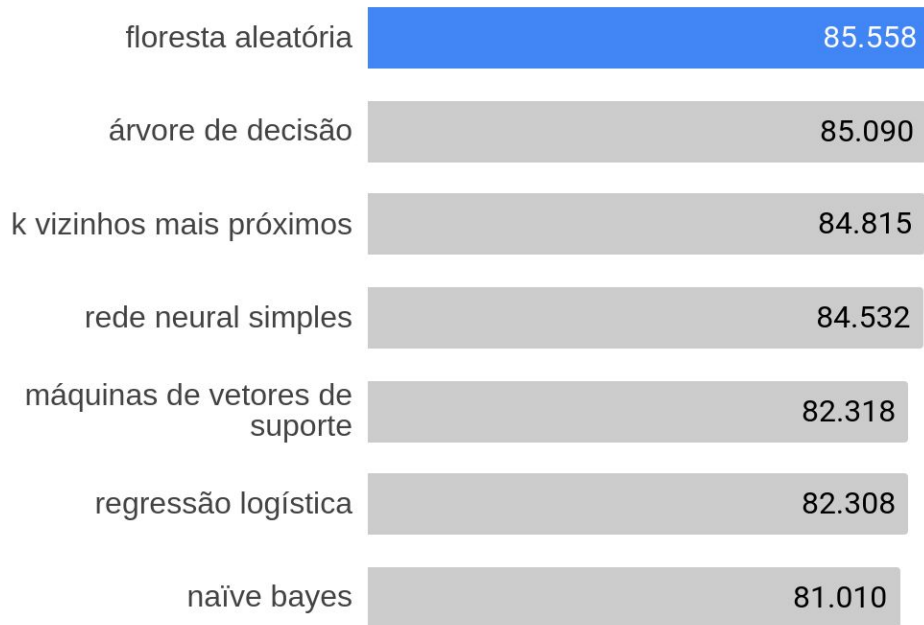
o problema do título

acurácias versus modelos



o problema do título

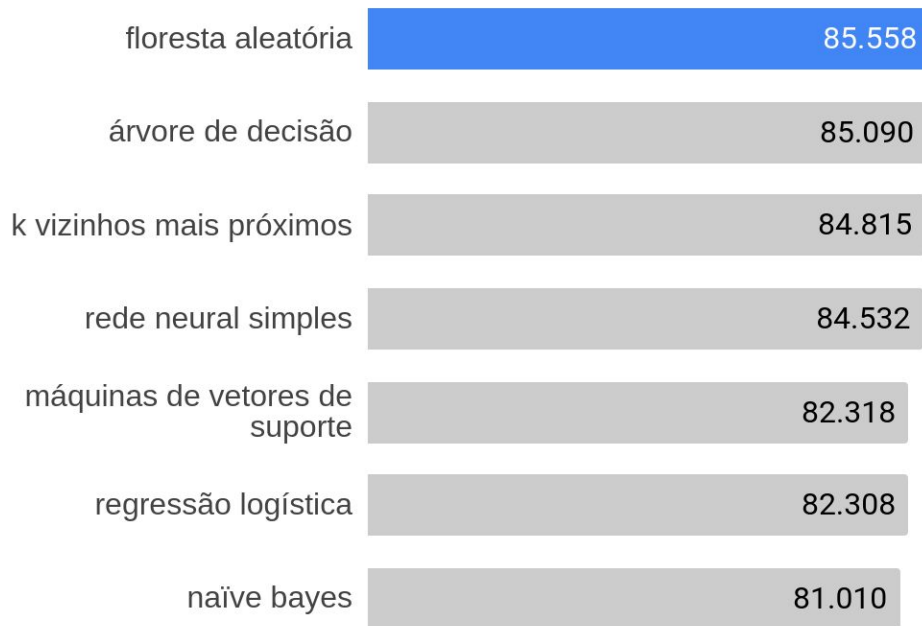
~~acurácias versus modelos~~

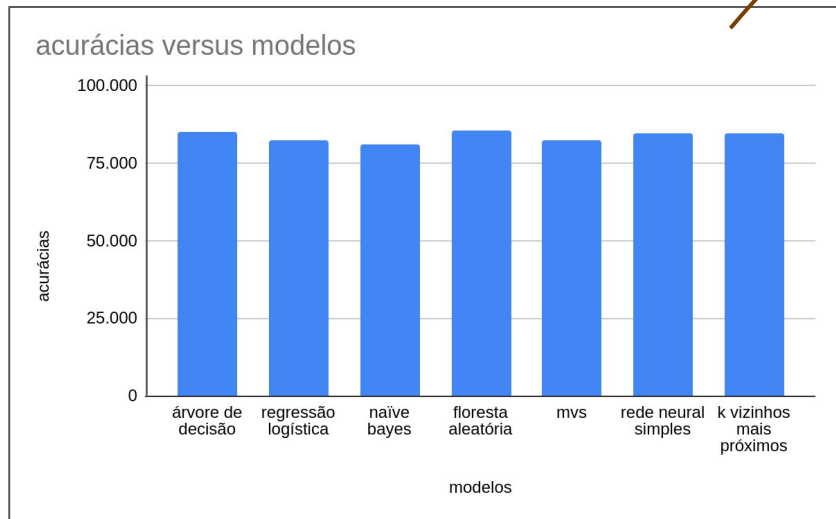


floresta aleatória	85.558
árvore de decisão	85.090
k vizinhos mais próximos	84.815
rede neural simples	84.532
máquinas de vetores de suporte	82.318
regressão logística	82.308
naïve bayes	81.010

o problema do título

A acurácia mais alta, de **86%**, foi do modelo de floresta aleatória





A acurácia mais alta, de **86%**, foi do modelo de floresta aleatória

floresta aleatória	85.558
árvore de decisão	85.090
k vizinhos mais próximos	84.815
rede neural simples	84.532
máquinas de vetores de suporte	82.318
regressão logística	82.308
naïve bayes	81.010

A background image of a pile of wooden blocks, some with letters and numbers, creating a textured, light-colored backdrop.

Segunda situação

Analizamos dados da rede elétrica &
queremos mostrar o que está acontecendo com o
consumo de energia (kwh)

meses	consumo (kwh)
março	80
abril	40
maio	80
junho	110
julho	180
agosto	120
setembro	240

Nesta situação, a gente também quer
comparar quantidades

Nesta situação, a gente também quer
comparar quantidades

Neste caso, também faríamos um gráfico de barras?





meses	consumo (kwh)
março	80
abril	40
maio	80
junho	110
julho	180
agosto	120
setembro	240

tempo

meses	consumo (kwh)
março	80
abril	40
maio	80
junho	110
julho	180
agosto	120
setembro	240

Como geralmente pensamos o tempo?

Como geralmente pensamos o tempo?

Na vertical?



Como geralmente pensamos o tempo?

Na vertical?



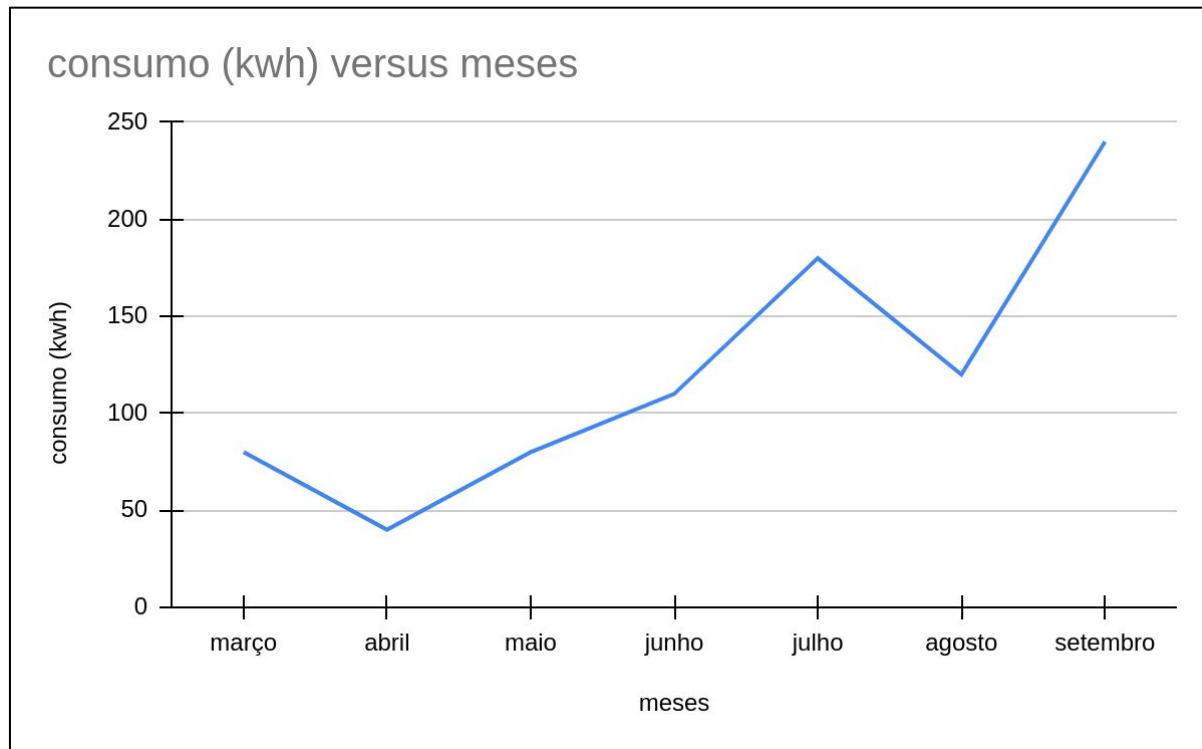
Ou na horizontal?



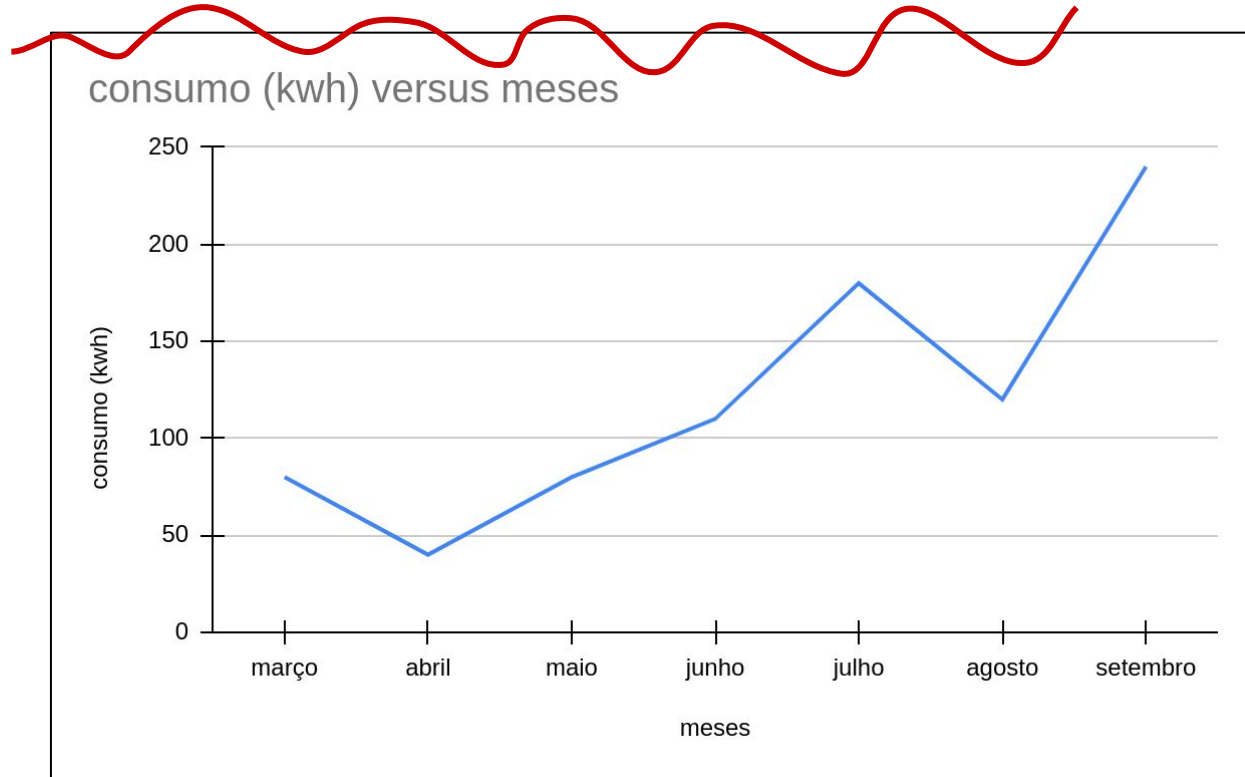
Queremos um gráfico de linhas



Assim está bom?

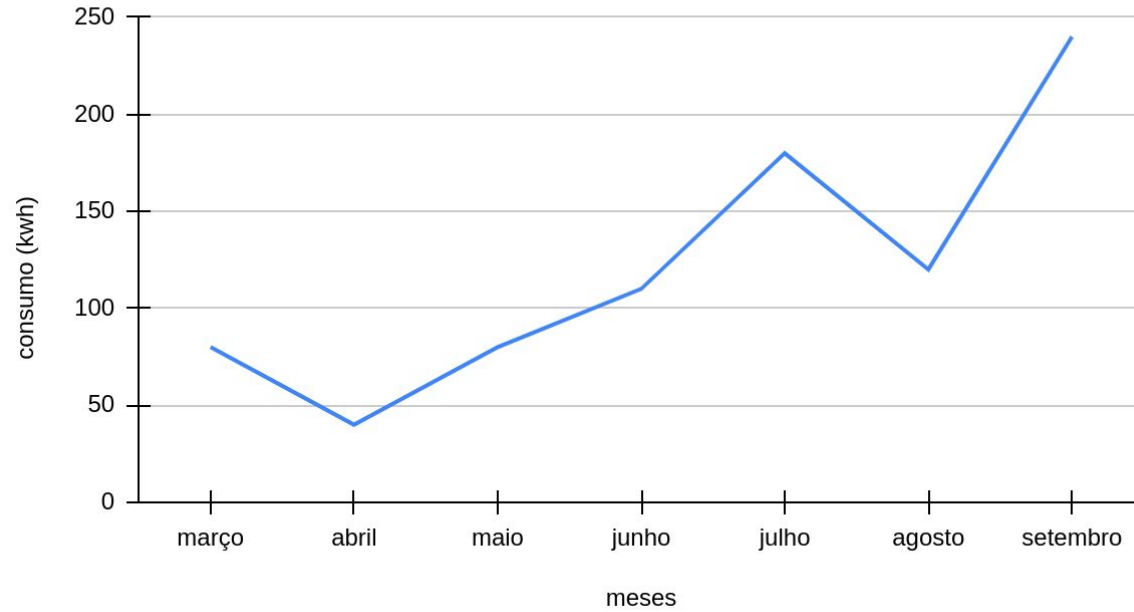


removendo a borda



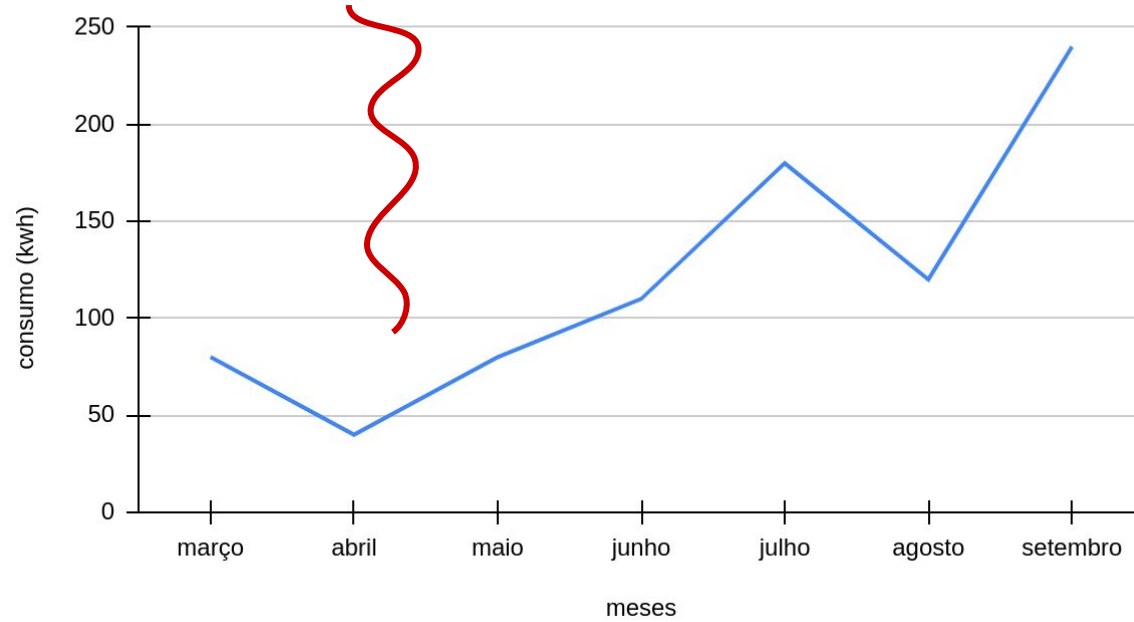
removendo as linhas de grade

consumo (kwh) versus meses



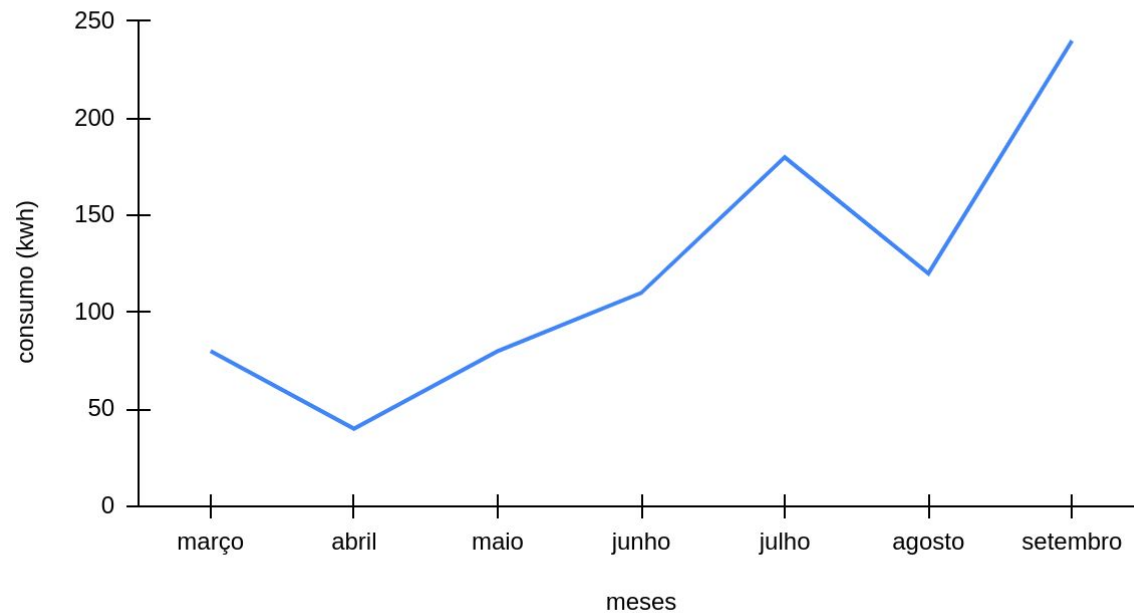
removendo as linhas de grade

consumo (kwh) versus meses



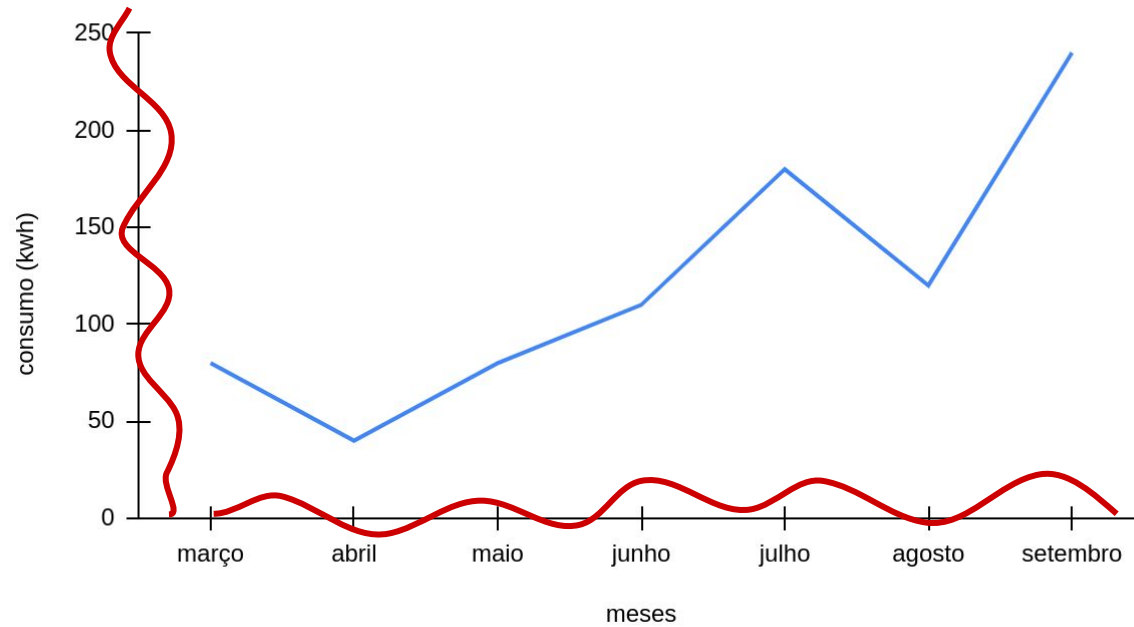
removendo as marcações dos eixos

consumo (kwh) versus meses



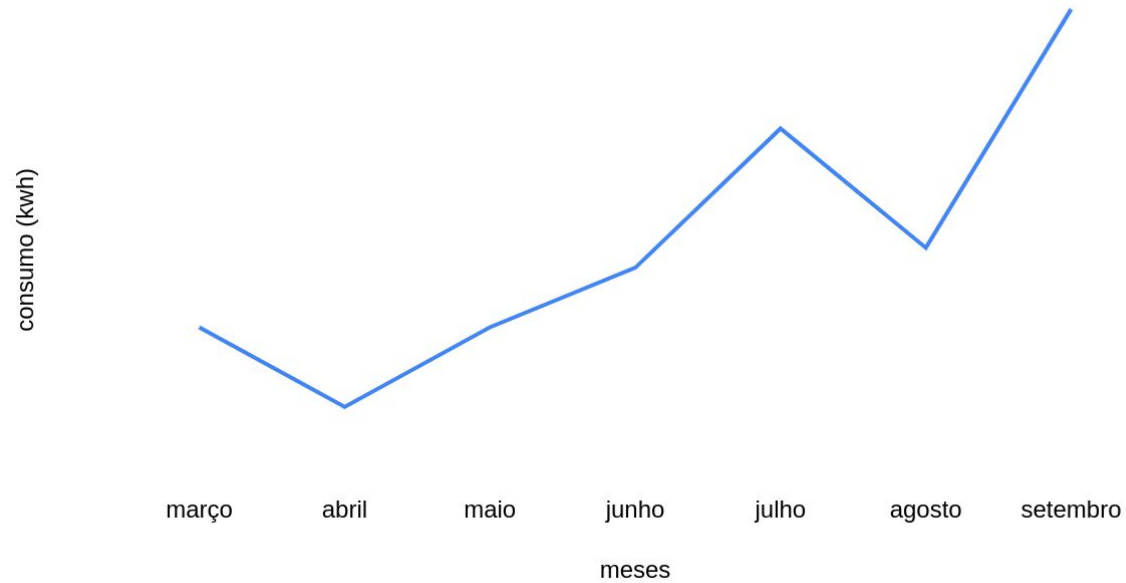
removendo as marcações dos eixos

consumo (kwh) versus meses



colocando valores na linha

consumo (kwh) versus meses



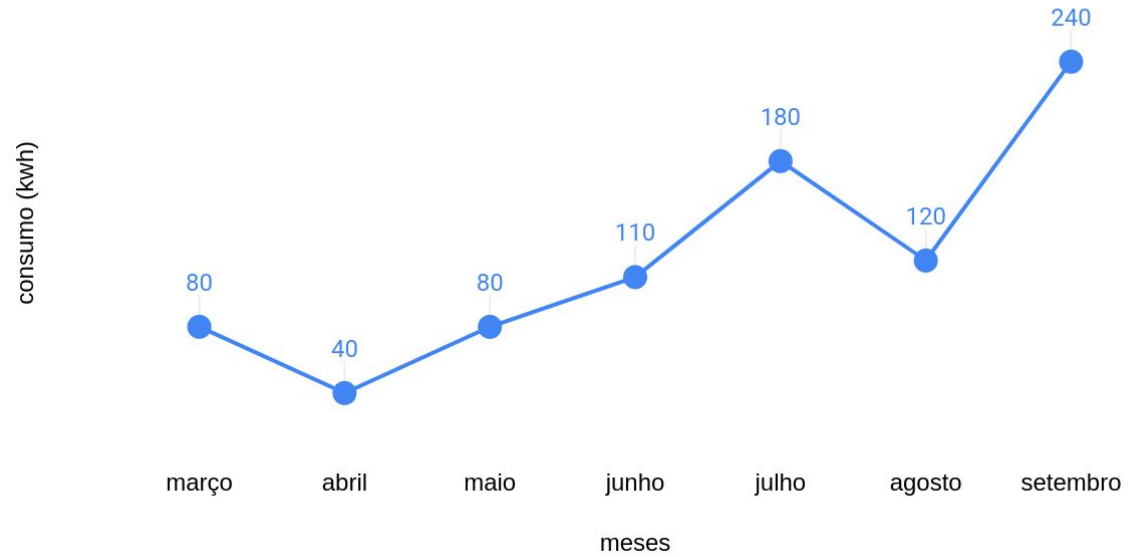
colocando valores na linha

consumo (kwh) versus meses



arrumando o consumo vira pescoço, bolinhas e os meses

consumo (kwh) versus meses

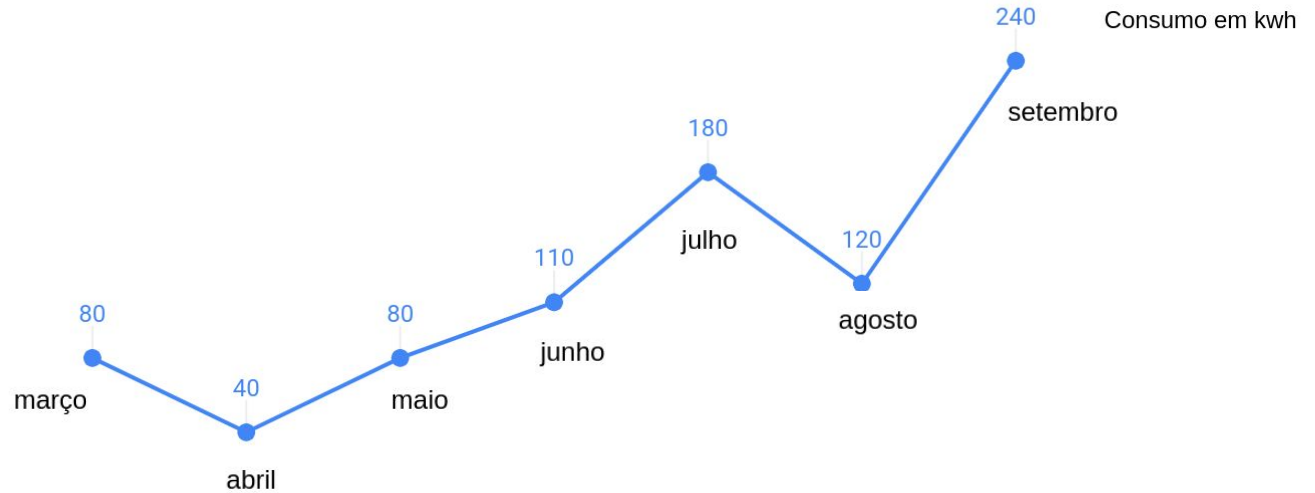


arrumando o consumo vira pescoço, bolinhas e os meses



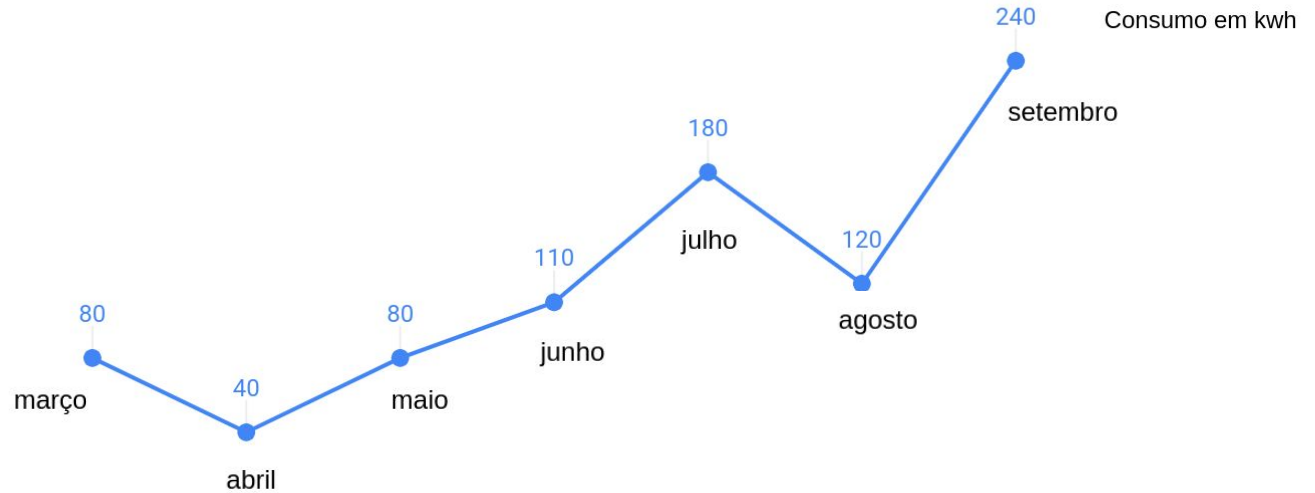
arrumando o título

consumo (kwh) versus meses



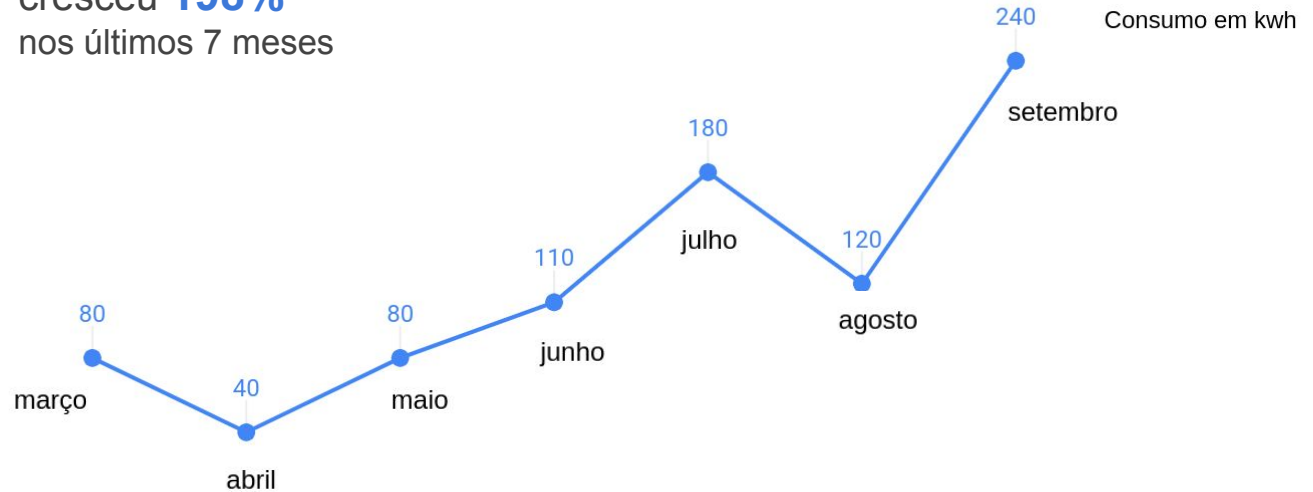
arrumando o título

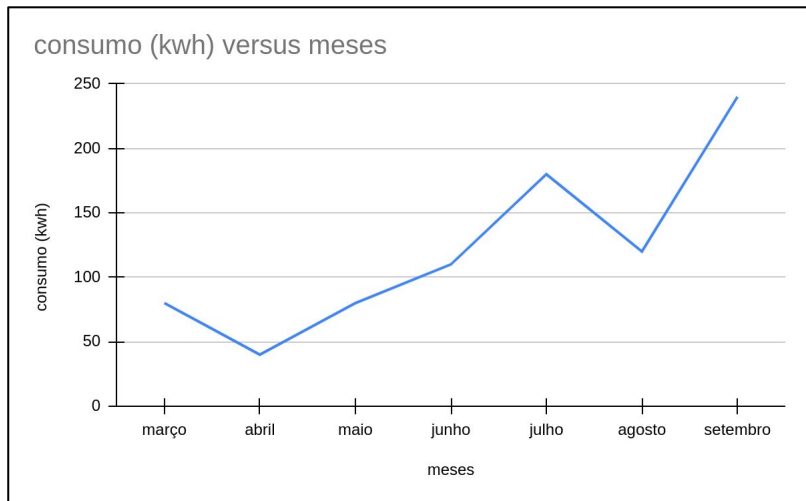
~~consumo (kwh) versus meses~~



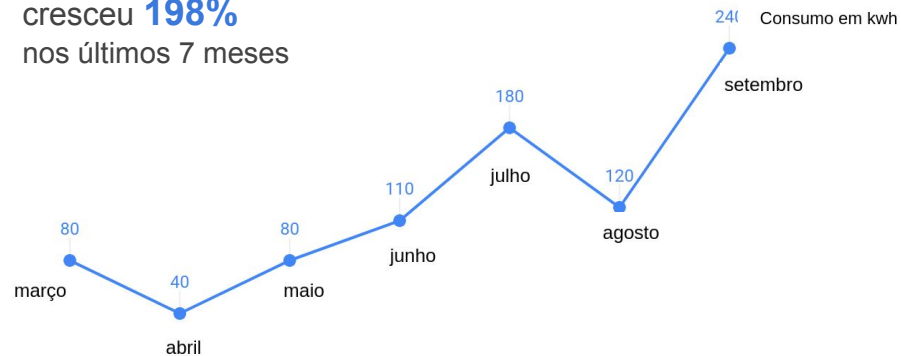
arrumando o título

O consumo de energia (kwh)
cresceu **198%**
nos últimos 7 meses





O consumo de energia (kwh)
cresceu **198%**
nos últimos 7 meses





Terceira situação

Vamos analisar dados de séries &
queremos saber qual a proporção de categorias
que o pessoal dá mais like



categorias	likes
ação	349
ficção	257
drama	100
comédia	125

O que queremos mostrar?

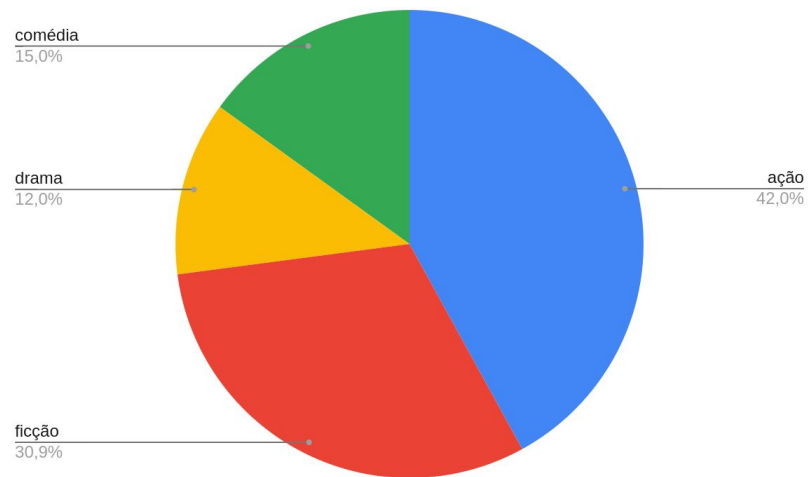
categorias	likes
ação	349
ficção	257
drama	100
comédia	125

O que queremos mostrar?

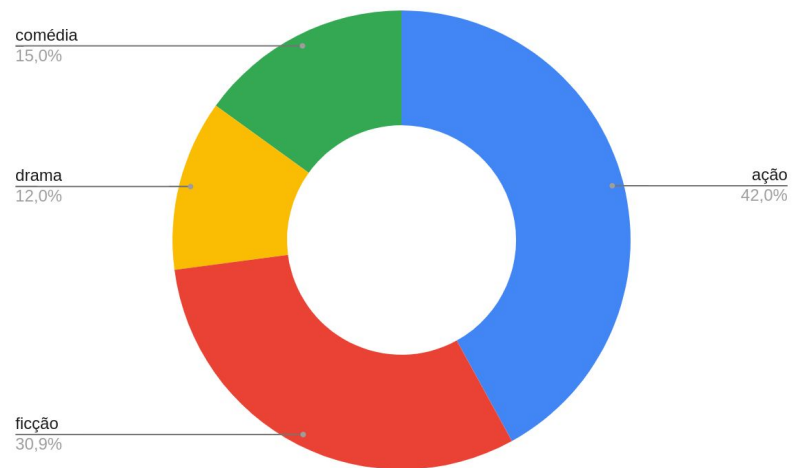
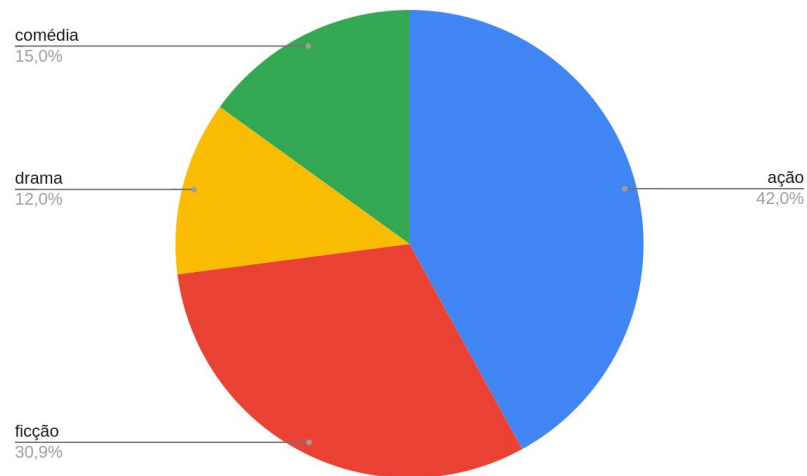
Como cada categoria **compõe**
um total

categorias	likes
ação	349
ficção	257
drama	100
comédia	125

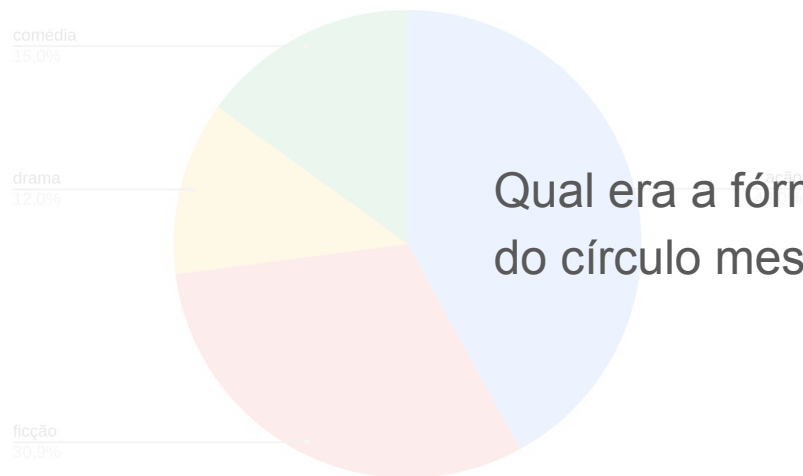
Gráficos mais usados



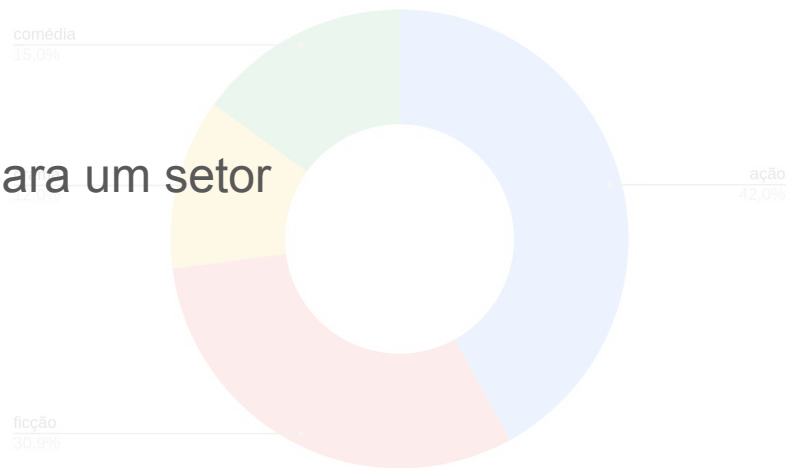
Gráficos mais usados



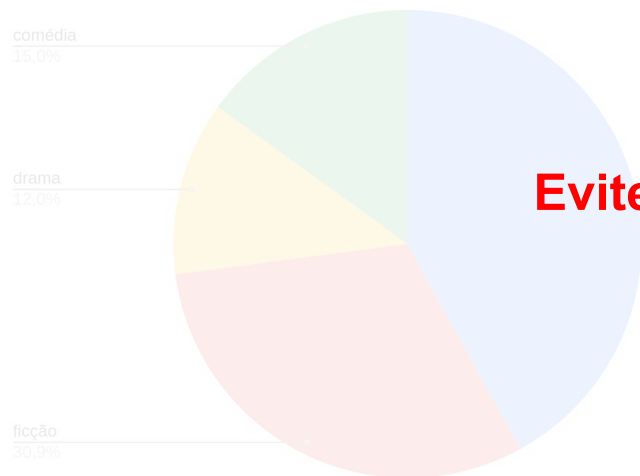
Gráficos mais usados



Qual era a fórmula para um setor do círculo mesmo?



Gráficos mais usados



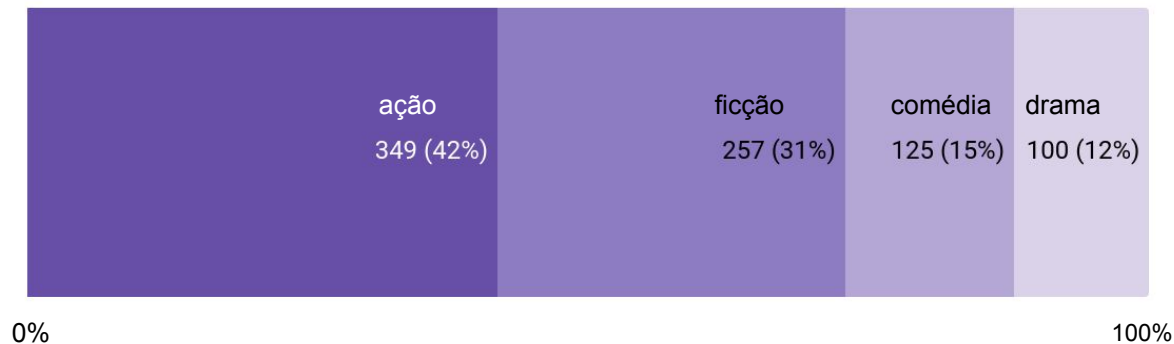
Evite estes gráficos sempre
que puder



Como saímos dessa?

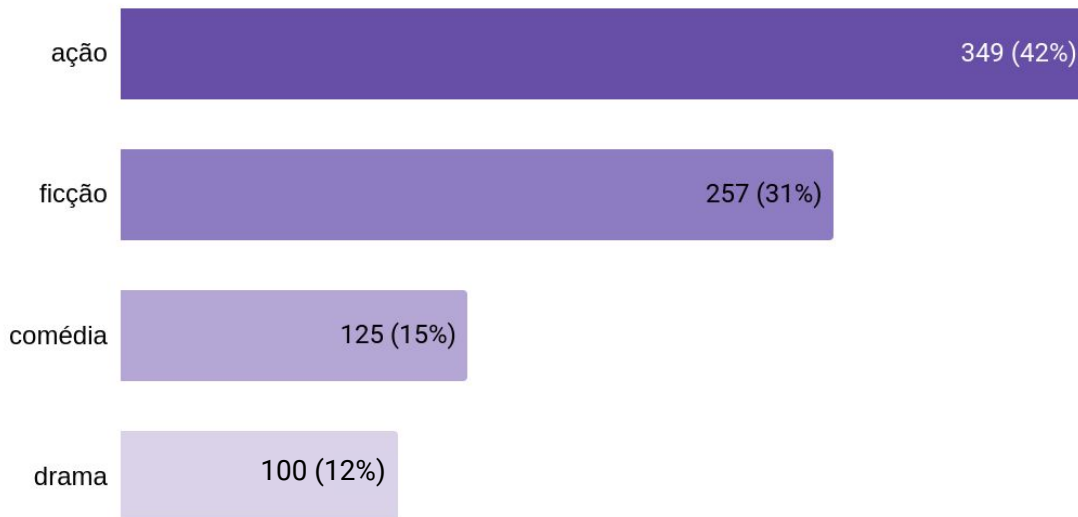
Como talvez saíamos dessa

Proporção de likes de cada gênero de série



Como realmente saímos dessa

Ação é a categoria com mais likes, compondo **42%** do total



A close-up photograph of a large pile of light-colored wooden blocks, similar to Scrabble tiles. The blocks are scattered and overlapping, with some showing letters (M, R, C, N, J, P, Z) and others showing numbers (1, 2, 3, 8). The word "Mantras" is overlaid in a brown, serif font on the left side of the image.

Mantras

Pontos que conversamos hoje

Pontos que conversamos hoje

- Deixe a principal informação destacada, explícita para quem vai visualizar

Pontos que conversamos hoje

- Deixe a principal informação destacada, explícita para quem vai visualizar
- **Simples** é melhor do que complexo (se a tabela funciona, use a tabela)

Pontos que conversamos hoje

- Deixe a **principal informação destacada, explícita** para quem vai visualizar
- **Simples** é melhor do que complexo (se a tabela funciona, use a tabela)
- Escolha gráficos de acordo com o que quer **mostrar** com os dados

Pontos que conversamos hoje

- Deixe a **principal informação destacada, explícita** para quem vai visualizar
- **Simples** é melhor do que complexo (se a tabela funciona, use a tabela)
- Escolha gráficos de acordo com o que quer **mostrar** com os dados
- **Liberte** os gráficos das informações excessivas

Pontos que conversamos hoje

- Deixe a **principal informação destacada, explícita** para quem vai visualizar
- **Simples** é melhor do que complexo (se a tabela funciona, use a tabela)
- Escolha gráficos de acordo com o que quer **mostrar** com os dados
- **Liberte** os gráficos das informações excessivas
- **Evite** gráficos de comidas

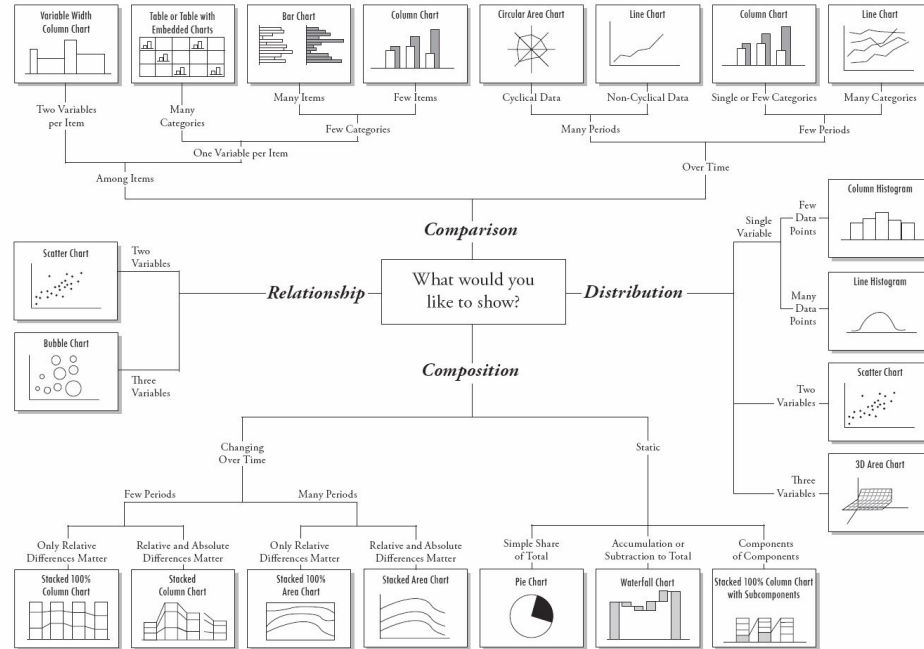
Pontos que conversamos hoje

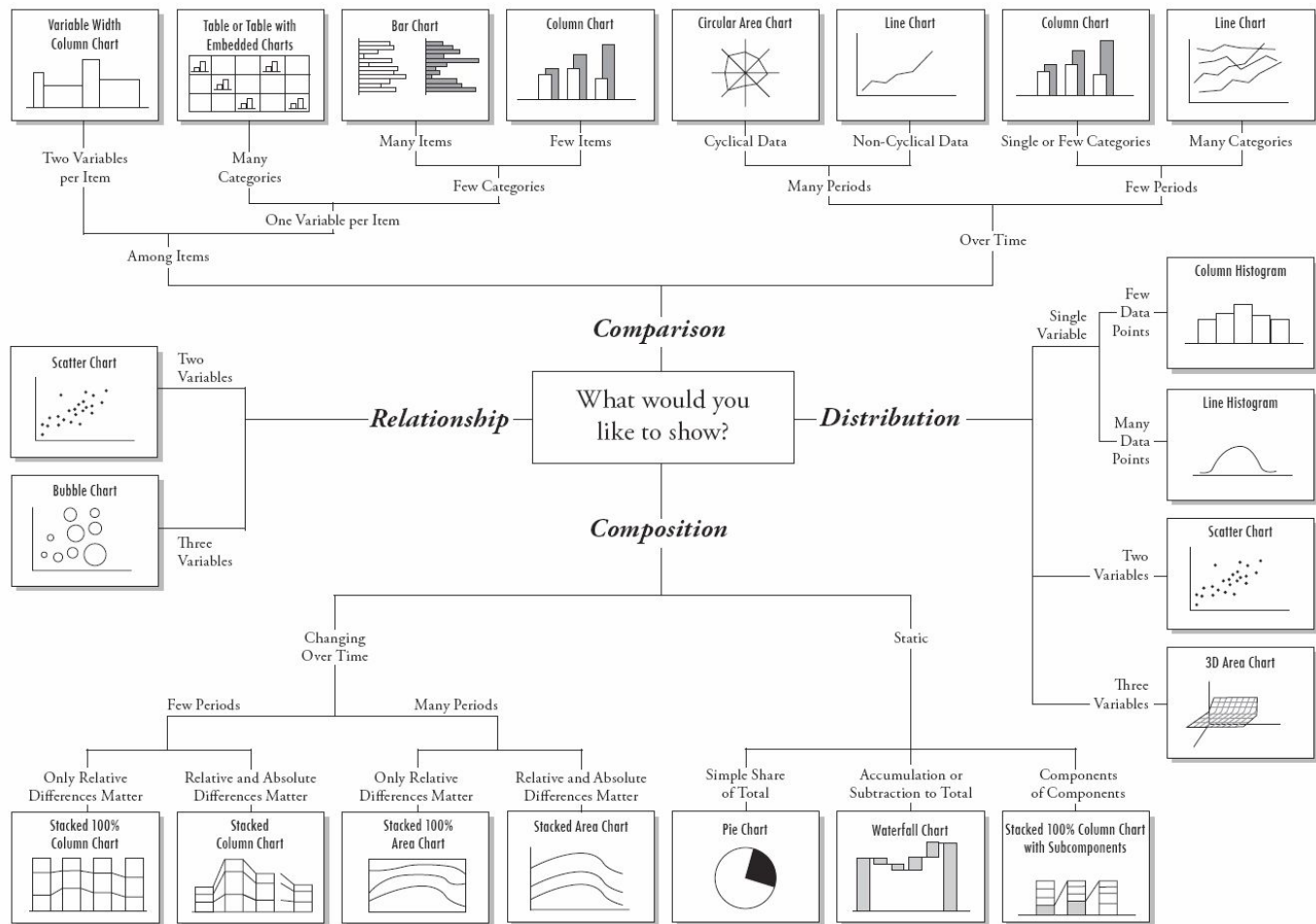
- Deixe a **principal informação destacada, explícita** para quem vai visualizar
- **Simples** é melhor do que complexo (se a tabela funciona, use a tabela)
- Escolha gráficos de acordo com o que quer **mostrar** com os dados
- **Liberte** os gráficos das informações excessivas
- **Evite** gráficos de comidas
- Use **cores** para **destacar, unir ou diferenciar** elementos

A background image of a pile of wooden blocks, similar to Scrabble tiles, with various letters and numbers printed on them. The blocks are light-colored wood and are scattered in a disorganized manner. The text "Como escolher um gráfico?" is overlaid on the image in a brown, serif font.

Como escolher um
gráfico?

Guia para a escolha de gráficos







Como escolher as
cores?

Guia de cores opostas e complementares



Cores opostas



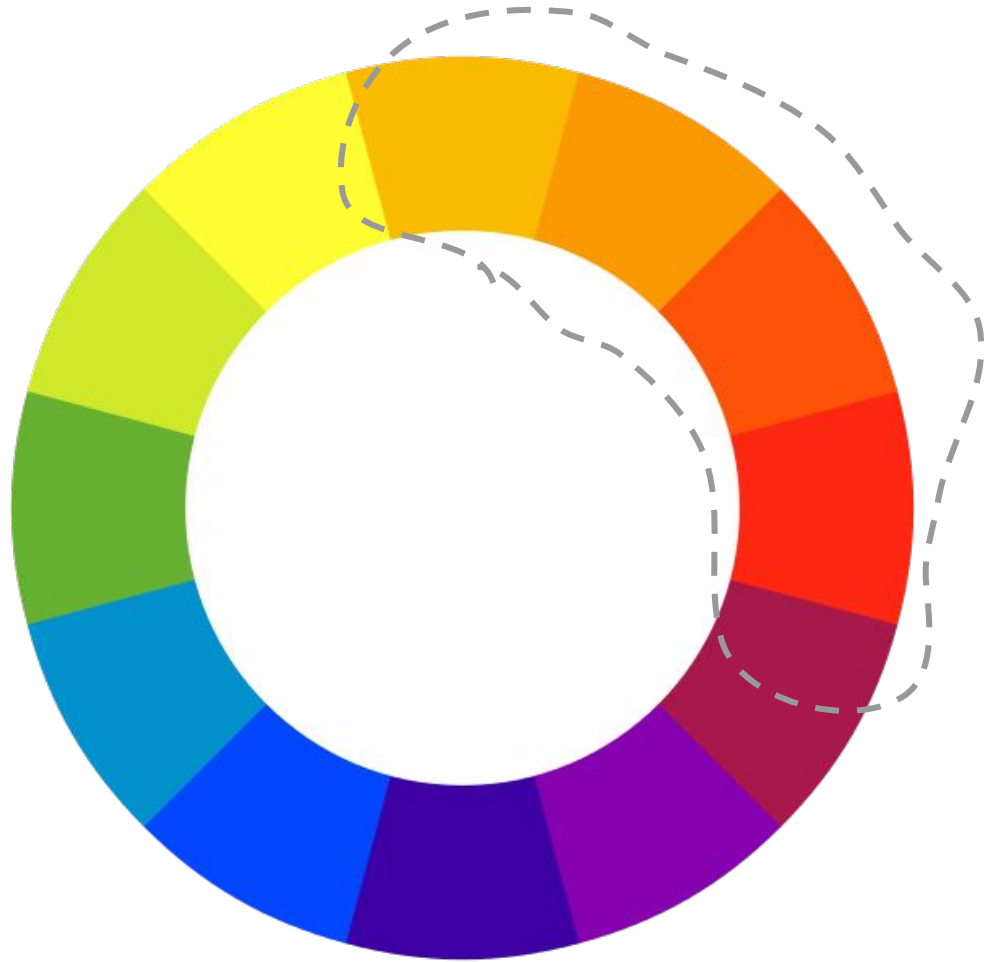
Cores opostas



Cores complementares



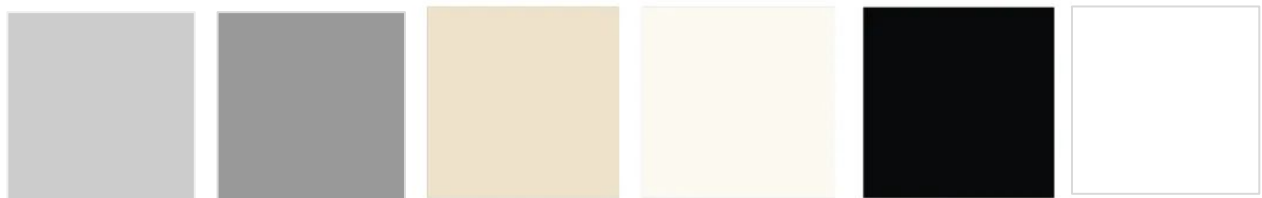
Cores complementares



Cores complementares



Cores neutras



Fontes dos dados & gráficos

- gráficos de barras e pizza: <https://www.maisbolsas.com.br/enem/matematica/analise-de-graficos>
- gráficos de pizza, linha e barras: <https://escola.britannica.com.br/artigo/gr%C3%A1fico/481401/recursos/134386>
- diagrama de caixa: <https://operdata.com.br/blog/como-interpretar-um-boxplot/>
- gráfico complexo: https://businessq-software.com/2019/02/28/chartjunk_and_why_to_avoid_them/
- gráfico de dispersão: <https://estatsite.com.br/2018/03/04/grafico-de-dispersao-no-python/>
- logo linkedin: <https://logodix.com/logos/91001>
- logo github e logo twitter: http://www.newdesignfile.com/post_linkedin-icons-black-circle_372056/
- dados de pesos por idade: <https://www.unimed.coop.br/viver-bem/pais-e-filhos/estatura-por-idade>
- círculo cromático: <https://www.significados.com/circulo-cromatico/>
- guia de gráficos: <https://www.labnol.org/software/find-right-chart-type-for-your-data/6523/>
- guia de gráficos colorida:
https://looker-elearning-resources.s3.amazonaws.com/DataVisualizationInfographic_PPT1.jpg
- tabela de acurácias:
https://www.researchgate.net/figure/Comparison-of-classification-accuracy-of-13-models_tbl2_284136695
- imagens dos slides: <https://www.pexels.com>

Contato

@cassiasamp

