

IMD0033 - Probabilidade

Aula 12 - Análise Exploratória de Dados II

Ivanovitch Silva
Abril, 2018



Agenda

- Estudo de caso: analisando filmes
- Motivação: enviesamento de dados
- Gráfico de barra & personalização
- Gráficos de dispersão & personalização

Atualizar o repositório

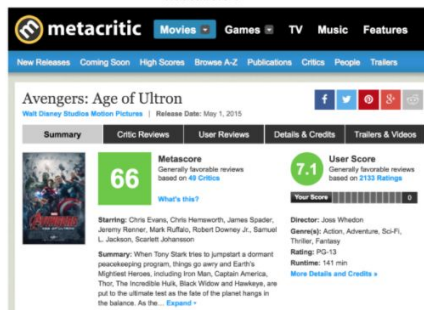
```
git clone https://github.com/ivanovitchm/imd0033_2018_1.git
```

Ou

```
git pull
```

Estudo de caso: avaliando filmes

Metacritic



metacritic Movies Games TV Music Features

New Releases Coming Soon High Scores Browse A-Z Publications Critics People Trailers

Avengers: Age of Ultron

Walt Disney Studios Motion Pictures | Release Date: May 1, 2015

Summary Critic Reviews User Reviews Details & Credits Trailers & Videos

66 Metascore
Generally favorable reviews based on 49 Critics

7.1 User Score
Generally favorable reviews based on 2133 Ratings

What's this?

Starring: Chris Evans, Chris Hemsworth, James Spader, Jeremy Renner, Mark Ruffalo, Robert Downey Jr., Samuel L. Jackson, Scarlett Johansson

Director: Josh Whedon

Genre(s): Action, Adventure, Sci-Fi, Thriller, Fantasy

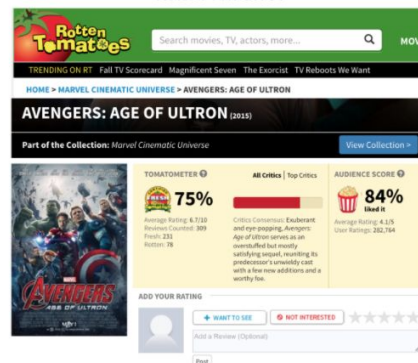
Rating: PG-13

Runtime: 141 min

[More Details and Credits](#)

Summary: When Tony Stark tries to jumpstart a dormant peacekeeping program, things go awry and Earth's Mightiest Heroes, including Iron Man, Captain America, Thor, The Incredible Hulk, Black Widow and Hawkeye, are put to the ultimate test as the fate of the planet hangs in the balance. [Expand](#)

Rotten Tomatoes



Rotten Tomatoes Search movies, TV, actors, more... MOVIE

TRENDING ON RT Fall TV Scorecard Magnificent Seven The Exorcist TV Reboots We Want

HOME > MARVEL CINEMATIC UNIVERSE > AVENGERS: AGE OF ULTRON

AVENGERS: AGE OF ULTRON (2015)

Part of the Collection: Marvel Cinematic Universe [View Collection](#)

TOMATOMETER **75%**
Average Rating: 6.7/10
Reviews Counted: 369
Fresh: 253
Rotten: 76

AUDIENCE SCORE **84%**
Average Rating: 4.2/5
User Ratings: 282,794

Critic Consensus: Exuberant and eye-popping, *Avengers: Age of Ultron* serves as an overwrought but moodily satisfying sequel, rewarding its predecessor's unwieldy cast with a few new additions and a worthy foe.

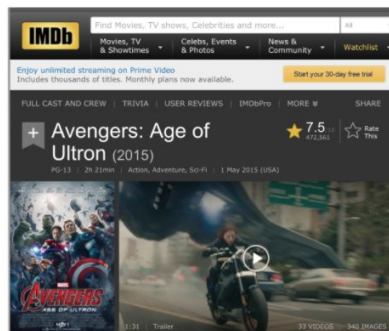
ADD YOUR RATING

[WANT TO SEE](#) [NOT INTERESTED](#) ★★★★★

Add a Review (Optional)

Post

IMDB



IMDb First Movies, TV shows, celebrities and more... all

Movies, TV & Showtimes Critics, Events & Photos News & Community Watchlist

Enjoy unlimited streaming on Prime Video
Includes thousands of titles. Monthly plans now available. [Start your 30-day free trial](#)

FULL CAST AND CREW TRIVIA USER REVIEWS | IMDbPro | MORE W SHARE

Avengers: Age of Ultron (2015)

PG-13 | 2h 23min | Action, Adventure, Sci-Fi | 1 May 2015 (USA)

7.5 Rate This
472,361

Trailer

33 VIDEOS 141 PHOTOS

Fandango



FANDANGO Enter City + State, ZIP Code, or Movie GO

AVENGERS: AGE OF ULTRON (2015)

OVERVIEW MOVIE TIMES + TICKETS SYNOPSIS MOVIE REVIEWS TRAILER

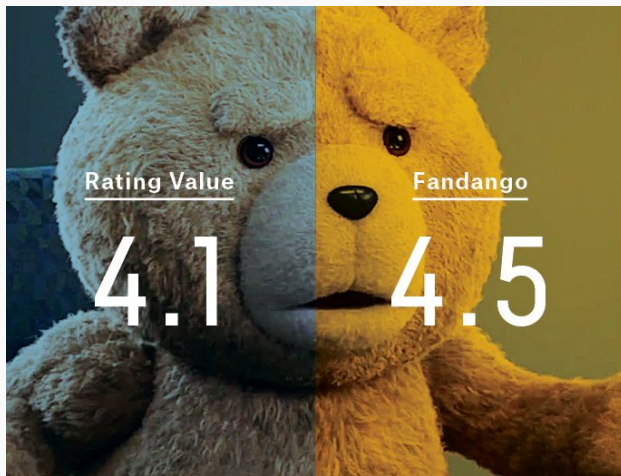
Released MAY 1, 2015

PG-13 - 2 hr 21 min
Action/Adventure
Family

★★★★★
15,861 Fan Ratings

GLOBAL A

Enviesamiento de dados

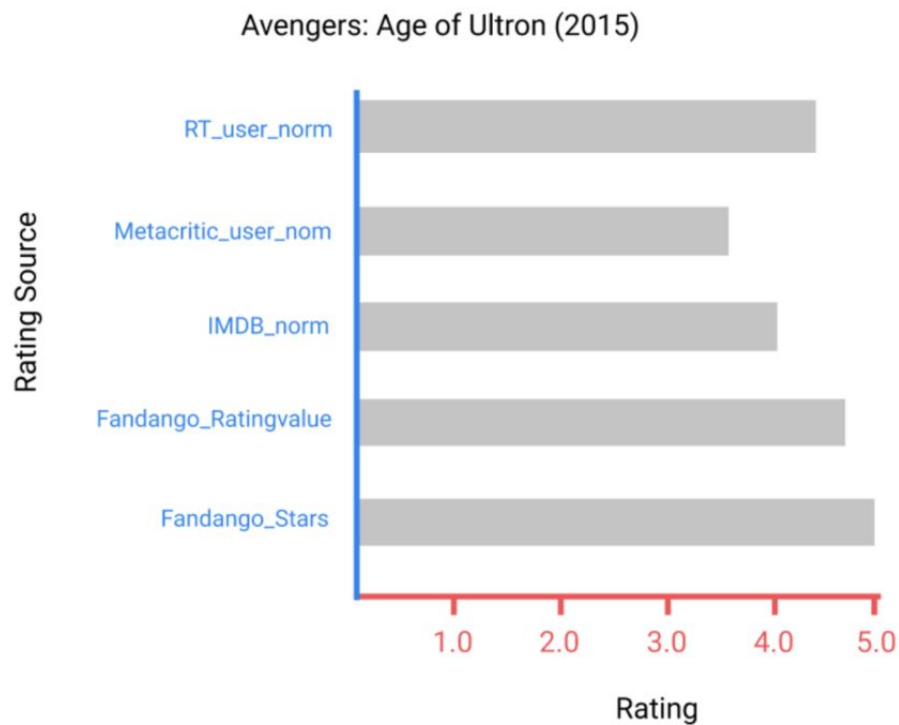
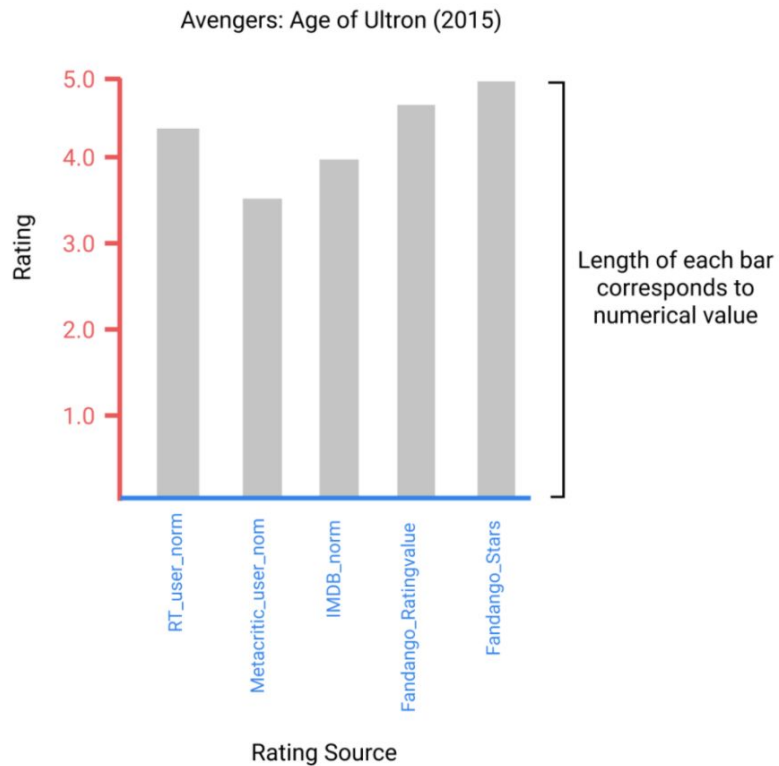


Conjunto de dados

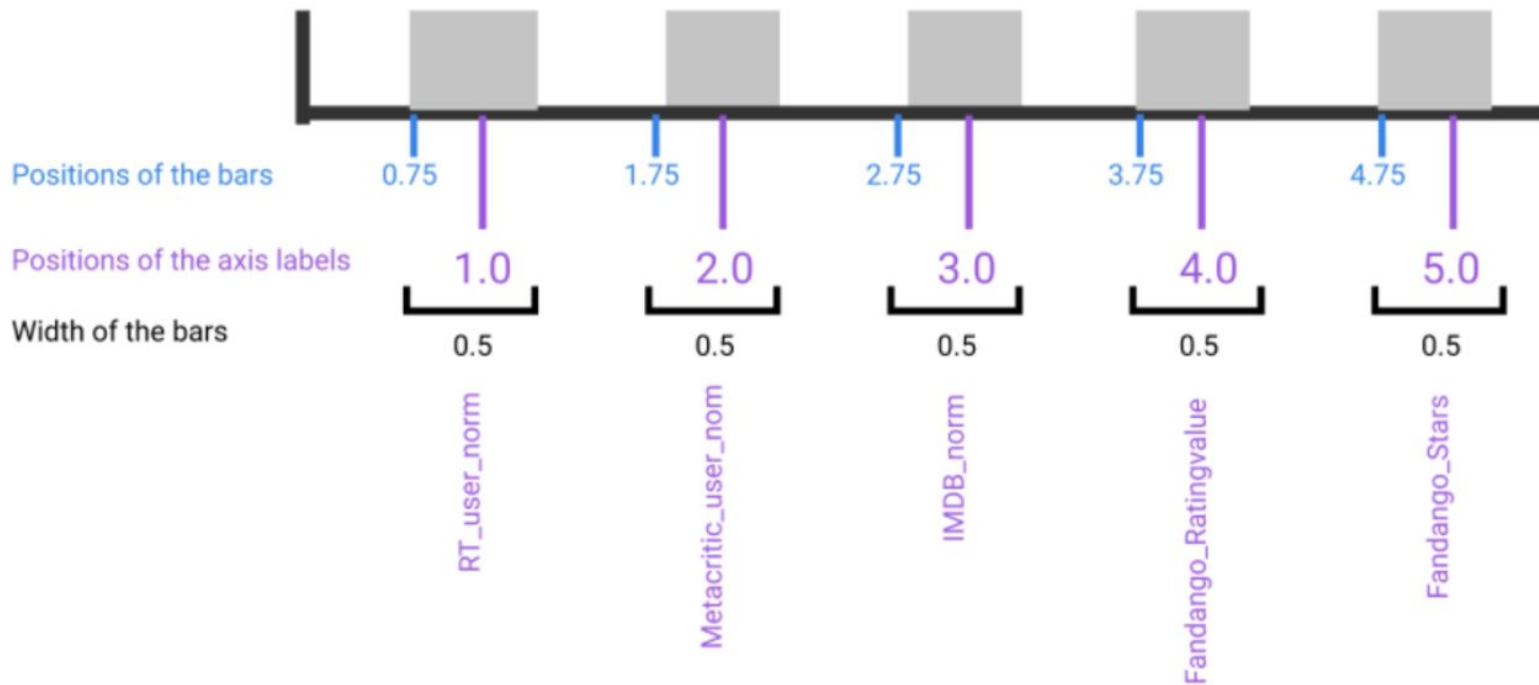
	FILM	RT_user_norm	Metacritic_user_nom	IMDB_norm	Fandango_Ratingvalue	Fandango_Stars
0	Avengers: Age of Ultron (2015)	4.3	3.55	3.90	4.5	5.0
1	Cinderella (2015)	4.0	3.75	3.55	4.5	5.0
2	Ant-Man (2015)	4.5	4.05	3.90	4.5	5.0
3	Do You Believe? (2015)	4.2	2.35	2.70	4.5	5.0
4	Hot Tub Time Machine 2 (2015)	1.4	1.70	2.55	3.0	3.5

<https://github.com/fivethirtyeight/data/tree/master/fandango>

Gráfico de barra



Criando as barras I



Criando as barras II

```
fig, ax = plt.subplots()
```

```
# Positions of the left sides of the 5 bars. [0.75, 1.75, 2.75, 3.75, 4.75]
```

```
from numpy import arange
```

```
bar_positions = arange(5) + 0.75
```

```
# Heights of the bars. In our case, the average rating for the first movie  
in the dataset.
```

```
num_cols = ['RT_user_norm', 'Metacritic_user_nom', 'IMDB_norm', 'Fandango_Ra  
tingvalue', 'Fandango_Stars']
```

```
bar_heights = norm_reviews[num_cols].iloc[0].values
```

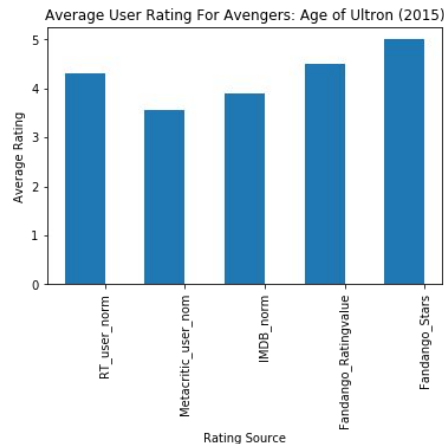
```
ax.bar(bar_positions, bar_heights)
```

Configurando os eixos

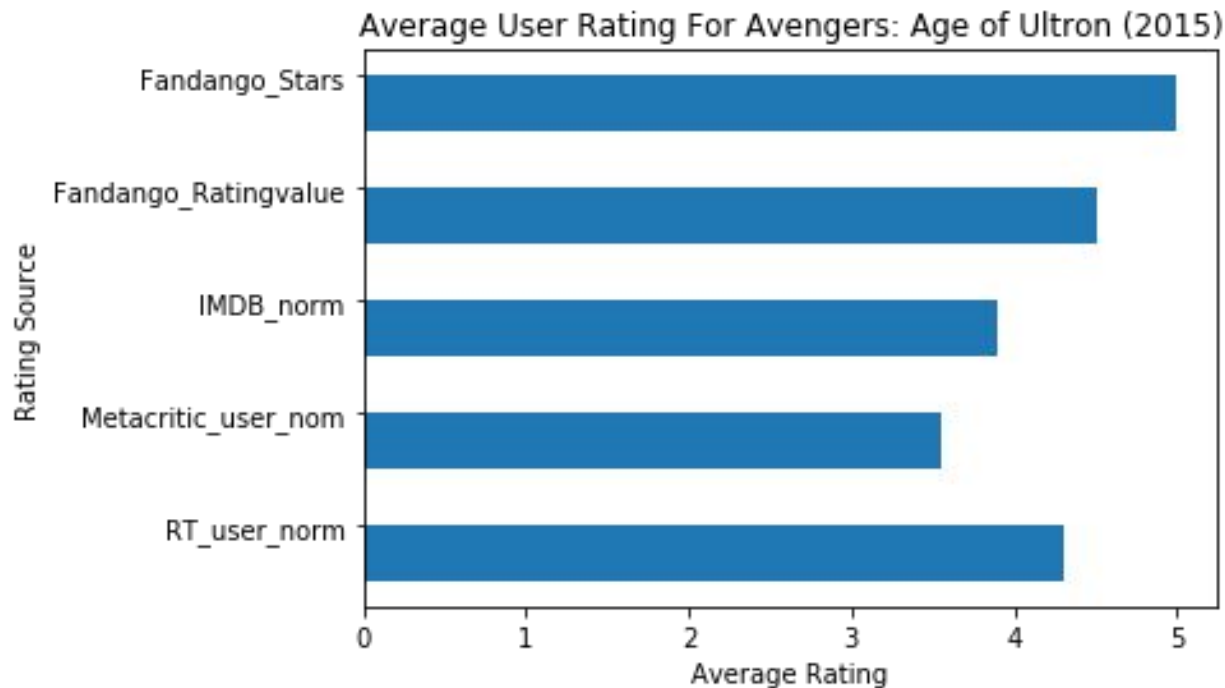
```
tick_positions = range(1,6)  
ax.set_xticks(tick_positions)
```

```
num_cols = ['RT_user_norm', 'Metacritic_user_nom', 'IMDB_norm', 'Fandango_Ra  
tingvalue', 'Fandango_Stars']  
ax.set_xticklabels(num_cols)
```

```
ax.set_xticklabels(num_cols, rotation=90)
```



Barras horizontais



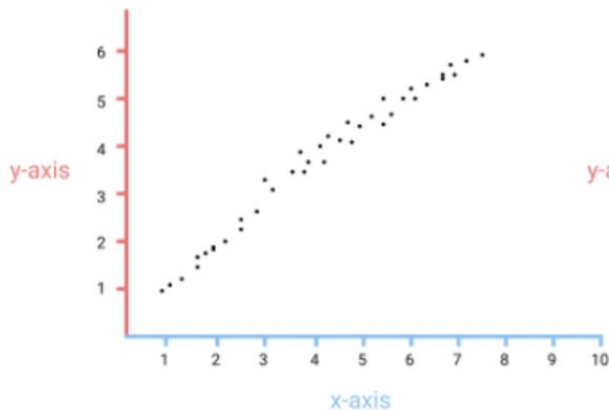
Barras horizontais

```
bar_widths = norm_reviews[num_cols].iloc[0].values
bar_positions = arange(5) + 0.75
ax.barh(bar_positions, bar_widths, 0.5)
```

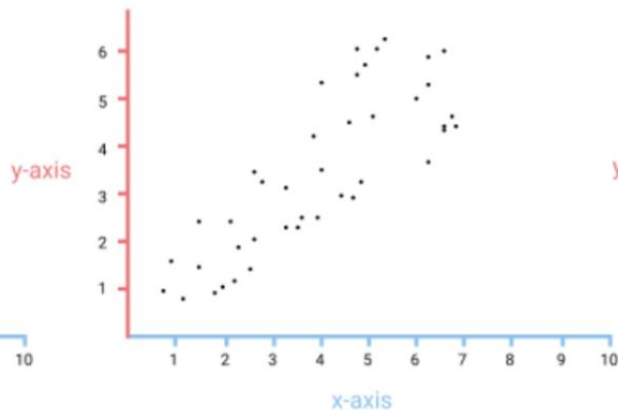
```
tick_positions = range(5) + 1
num_cols = ['RT_user_norm', 'Metacritic_user_nom', 'IMDB_norm', 'Fandango_Ra
tingvalue', 'Fandango_Stars']
ax.set_yticks(tick_positions)
ax.set_yticklabels(num_cols)
```

Gráficos de dispersão

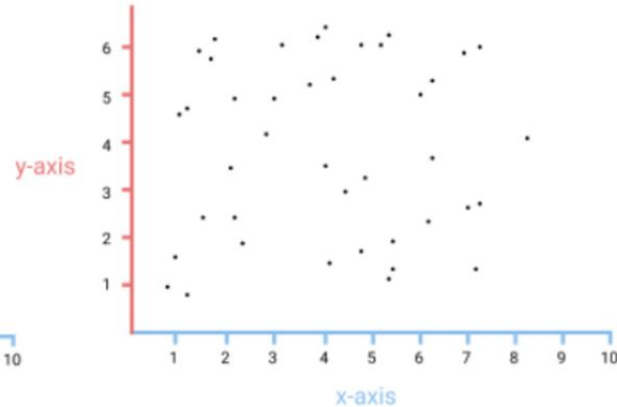
Strong correlation



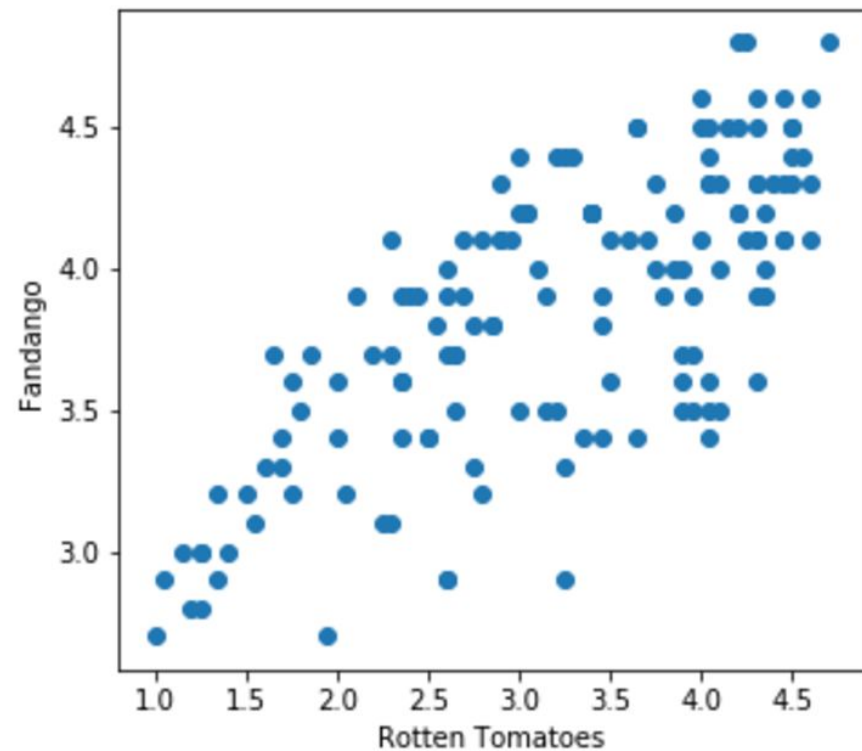
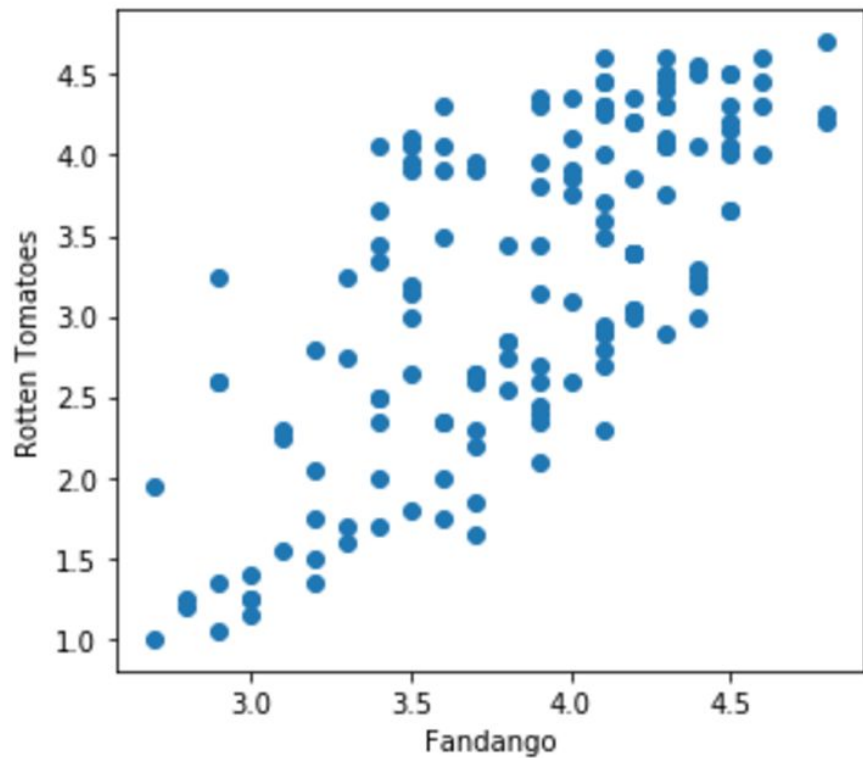
Weak correlation



No correlation



Trocando os eixos



Vocês concordam com a conclusão final descrita no notebook?

