

IMD0033 - Probabilidade

Aula 14 - Visualização Exploratória de Dados I

Ivanovitch Silva
Abril, 2018



Agenda

- Interface entre Pandas & Matplotlib
- Motivação
- Estudo de caso: diferença entre gêneros para cursos STEM

Atualizar o repositório

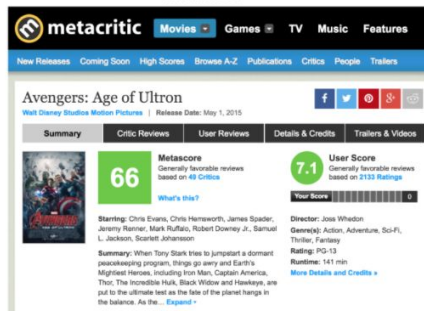
```
git clone https://github.com/ivanovitchm/imd0033_2018_1.git
```

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

Metascore
66
Generally favorable reviews based on 49 Critics

User Score
7.1
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

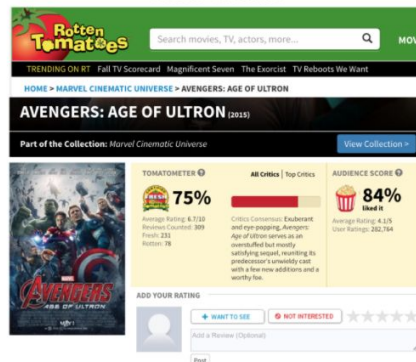
Rating: PG-13

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: 309
Fresh: 233
Rotten: 76

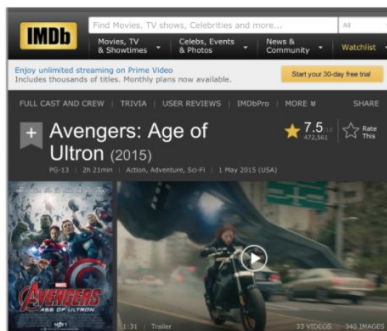
AUDIENCE SCORE 84%
Average Rating: 4.2/5
User Ratings: 282,794

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

ADD YOUR RATING

★★★★★

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)

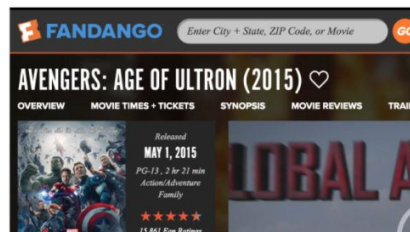
PG-13 | 2h 23min | Action, Adventure, Sci-Fi | 1 May 2015 (USA)

7.5 472,361 Rate This

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

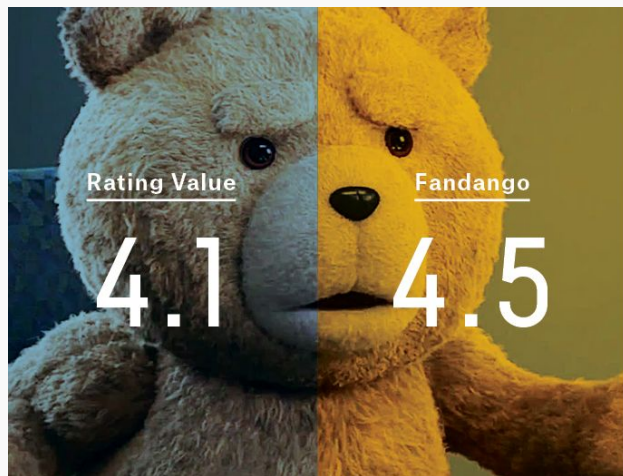
Released MAY 1, 2015

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

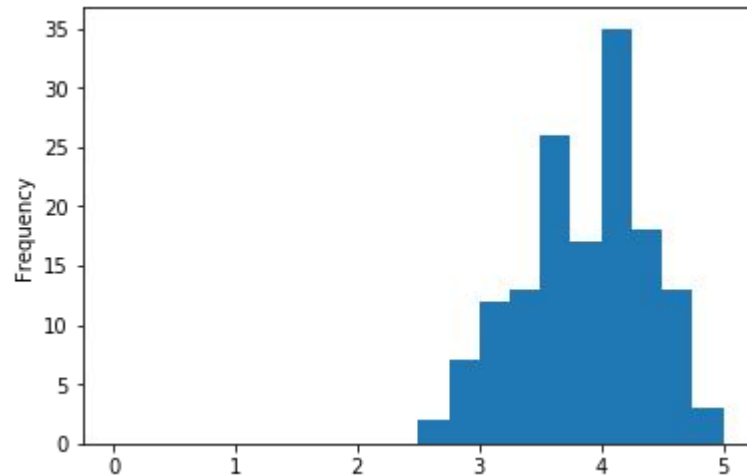
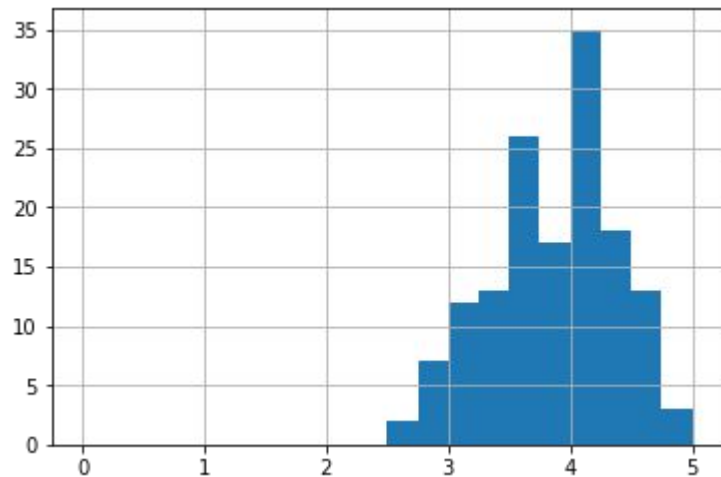
<https://github.com/fivethirtyeight/data/tree/master/fandango>

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

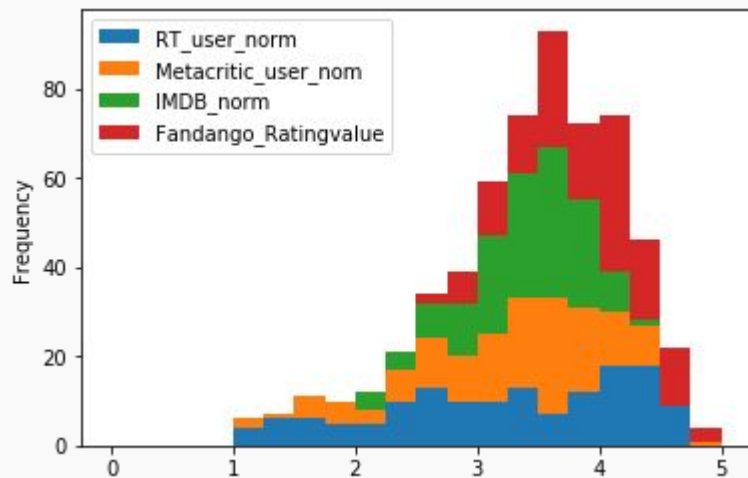
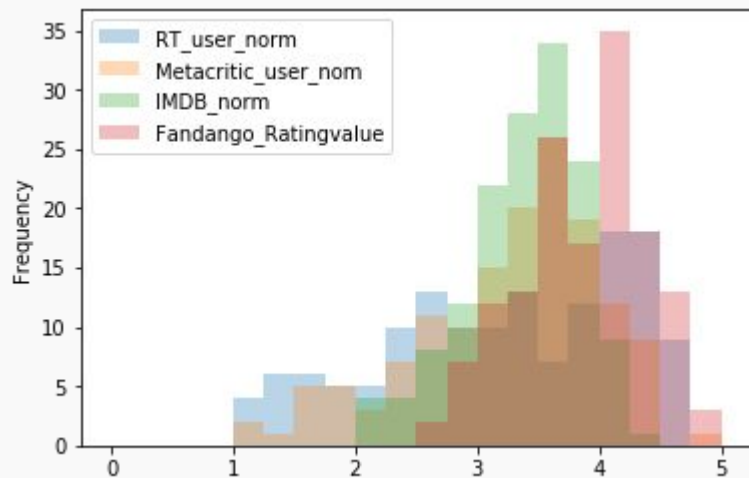


matplotlib



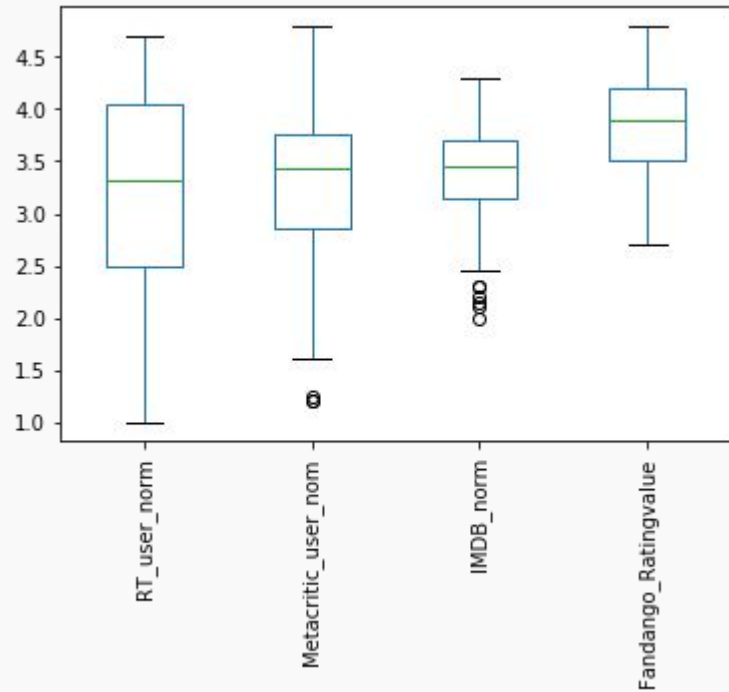
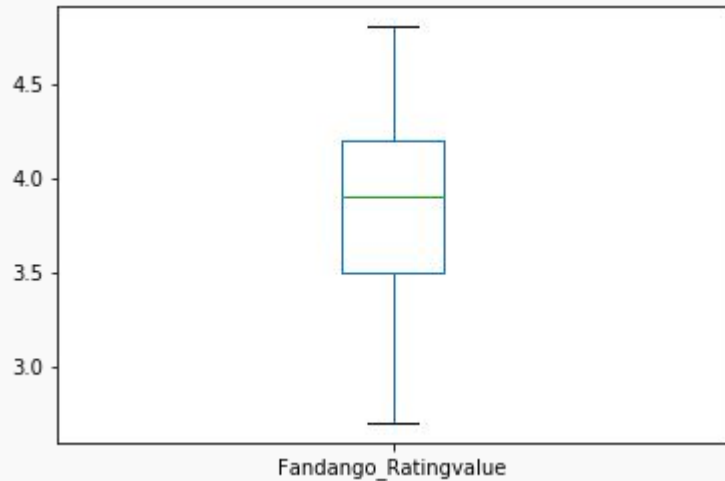
```
# Enable matplotlib plot inline
%matplotlib inline
norm_reviews.Fandango_Ratingvalue.hist(bins=20, range=(0,5))
```

```
# other way to do the same thing
norm_reviews.Fandango_Ratingvalue.plot(kind='hist', bins=20, range=(0,5));
```

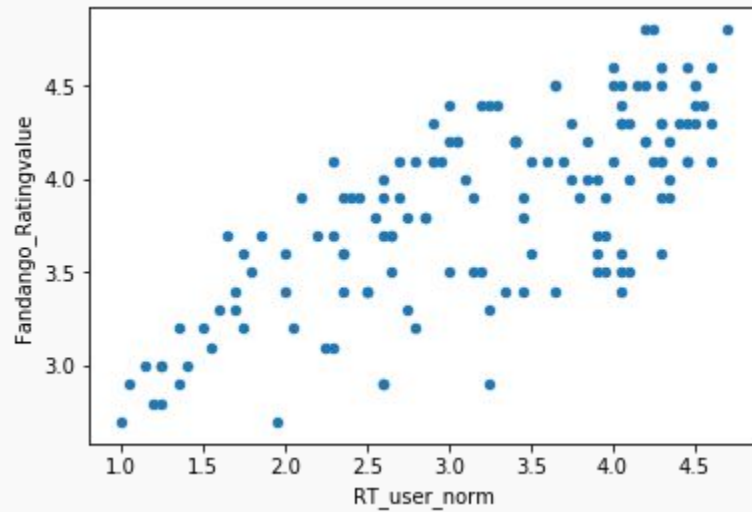
```
norm_reviews.plot(kind='hist', bins=20, range=(0,5), alpha=0.3);
```

```
norm_reviews.plot(kind='hist', bins=20, range=(0,5), stacked=True);
```



```
norm_reviews.Fandango_Ratingvalue.plot(kind='box')
```

```
norm_reviews.plot(kind='box', rot=90)
```

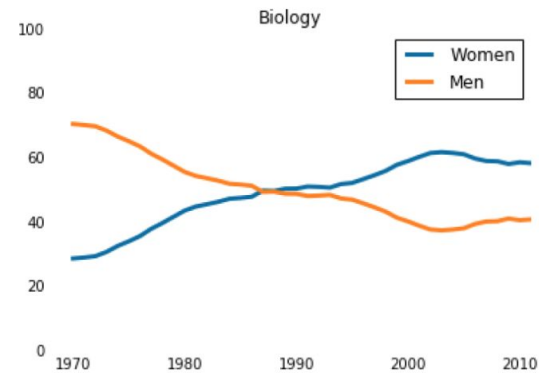
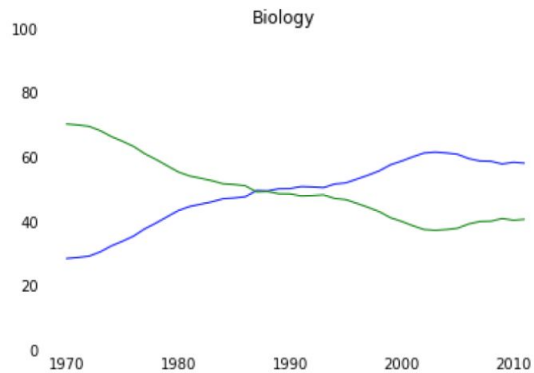
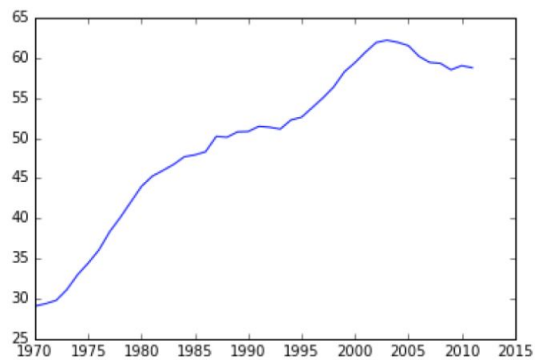


```
norm_reviews.plot(kind='scatter',x='RT_user_norm', y='Fandango_Ratingvalue')
```



Exploratory Data Analysis V - Plotting with Pandas.ipynb

Estética

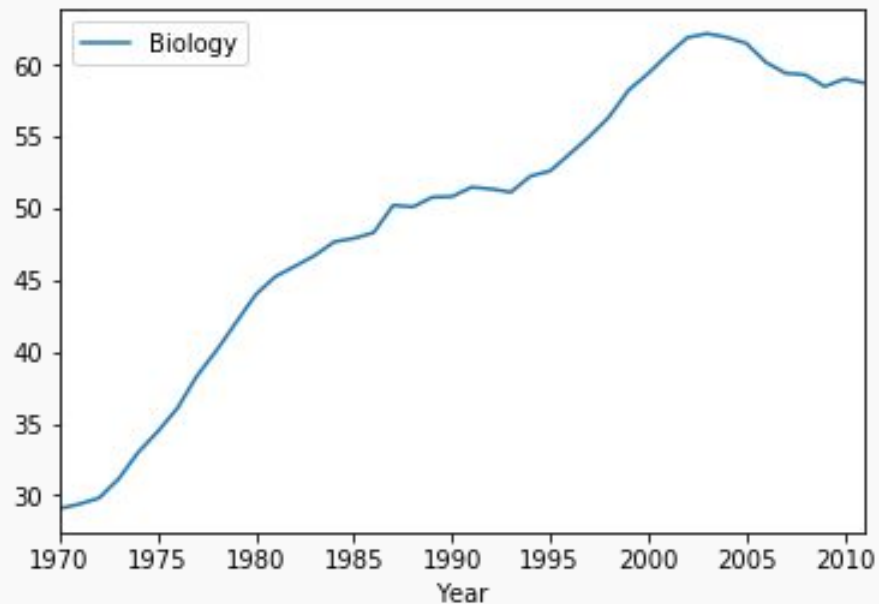


Introdução ao dataset

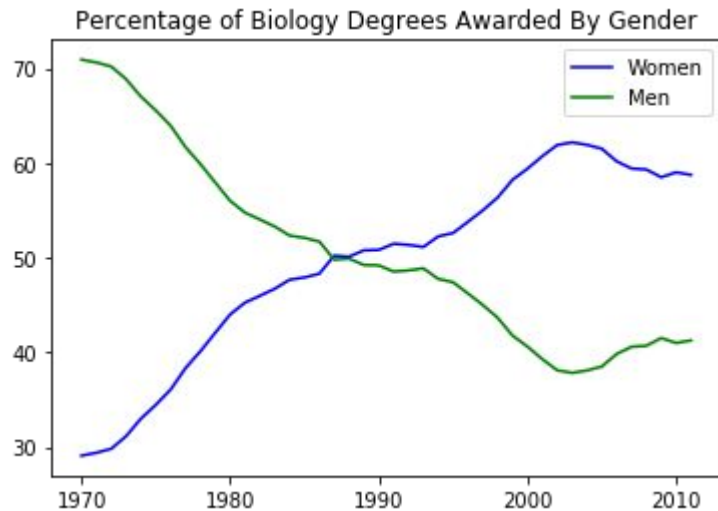
Year	Agriculture	Architecture	Art and Performance	Biology	Business	Communications and Journalism	Computer Science	Education	Engineering
1970	4.229798	11.921005	59.7	29.088363	9.064439	35.3	13.6	74.535328	0.8
1971	5.452797	12.003106	59.9	29.394403	9.503187	35.5	13.6	74.149204	1.0
1972	7.420710	13.214594	60.4	29.810221	10.558962	36.6	14.9	73.554520	1.2

- Porcentagem de mulheres que se formaram entre 1970 a 2012
- Departamento Americano para Estatísticas Educacionais

Visualizando a diferença de gênero



Visualizando a diferença de gênero



```
import matplotlib.pyplot as plt

plt.plot(women_degrees['Year'],
         women_degrees['Biology'], c='blue', label='Women')
plt.plot(women_degrees['Year'],
         100-women_degrees['Biology'], c='green', label='Men')
plt.legend(loc='upper right')
plt.title('Percentage of Biology Degrees Awarded By Gender')
plt.show()
```

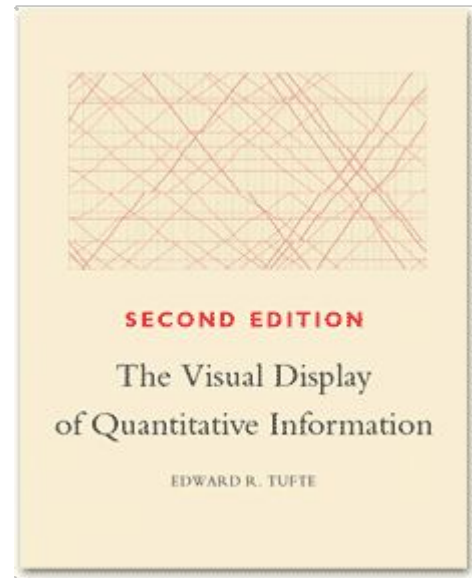
Visualizando a diferença de gênero

```
%matplotlib inline
women_degrees['men_bio'] = 100-women_degrees['Biology']
women_degrees.plot(kind='line',x='Year',y=['Biology','men_bio'],
                    title='Percentage of Biology Degrees Awarded By Gender',
                    color=['blue','green']).\
                    legend(loc='best',
                           labels=['Women','Men'])
```

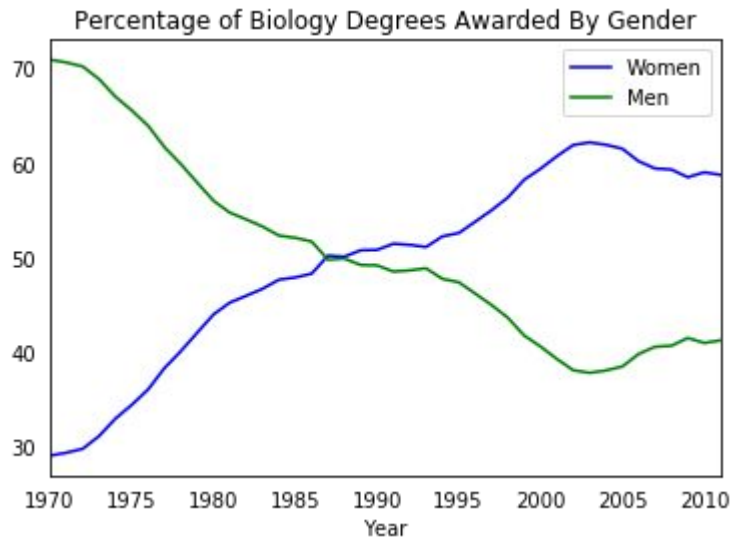
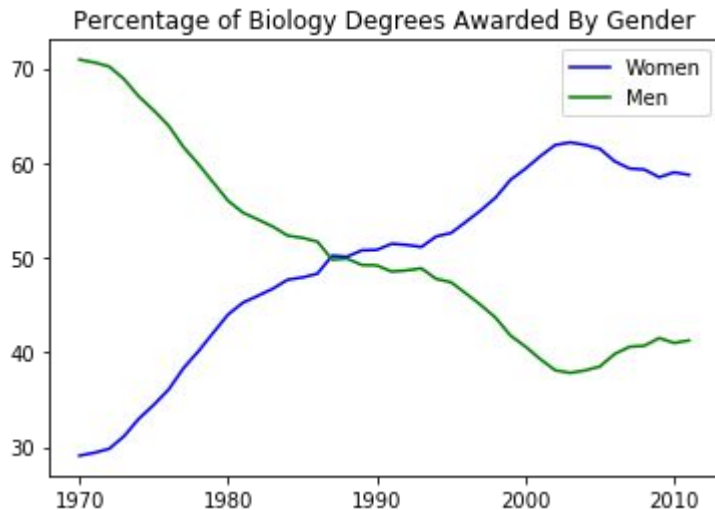
```
ax = women_degrees.plot(kind='line',x='Year',y=['Biology','men_bio'],
                        title='Percentage of Biology Degrees Awarded By Gender',
                        color=['blue','green'])
ax.legend(loc='best',labels=['Women','Men'])
```

Menos é mais

Remove
to improve
(the **data-ink** ratio)



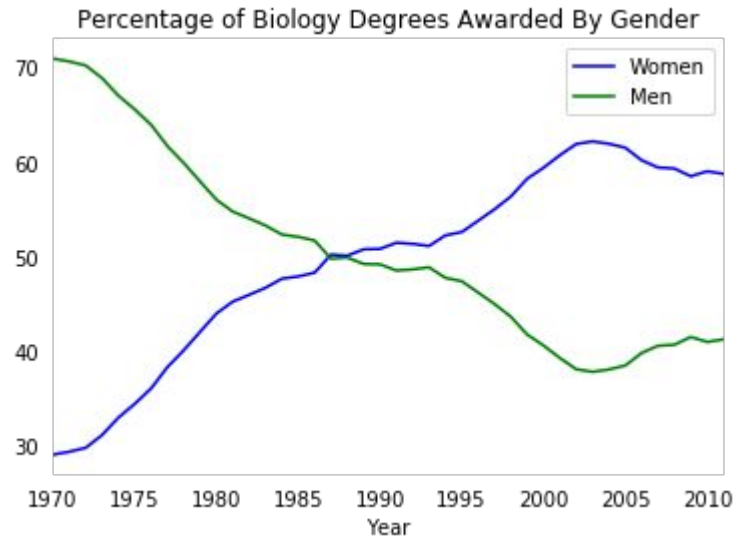
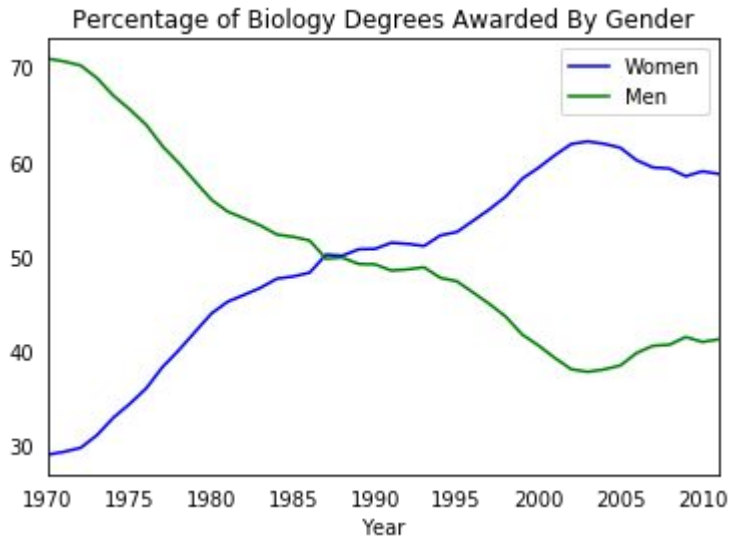
Ocultar as marcas dos eixos



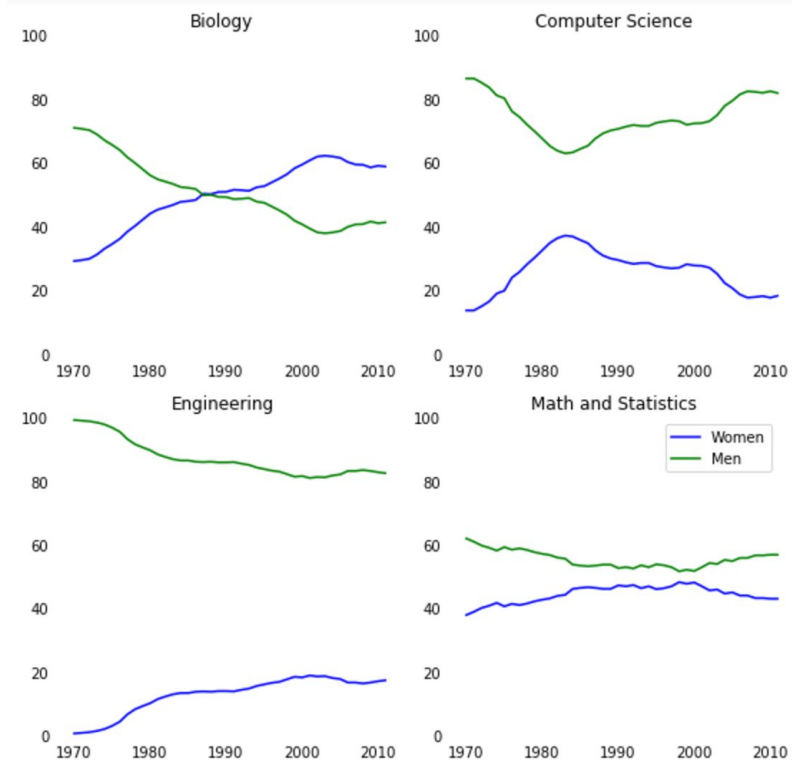
```
ax.tick_params(bottom="off", top="off", left="off", right="off")
```

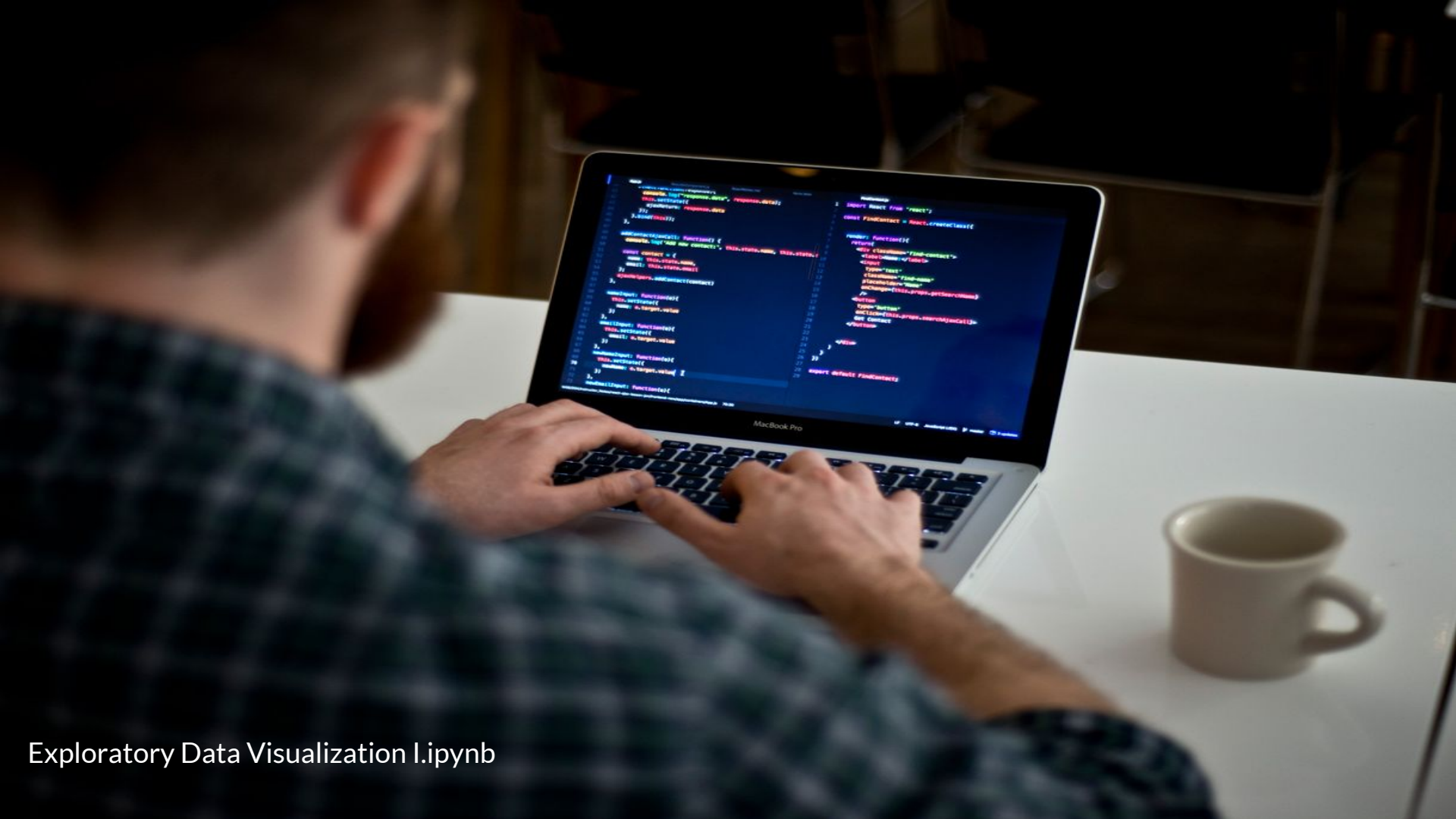
Ocultar os contornos

```
ax.spines["right"].set_visible(False)  
ax.spines["left"].set_visible(False)  
ax.spines["bottom"].set_visible(False)  
ax.spines["top"].set_visible(False)
```



Comparação final





Exploratory Data Visualization I.ipynb