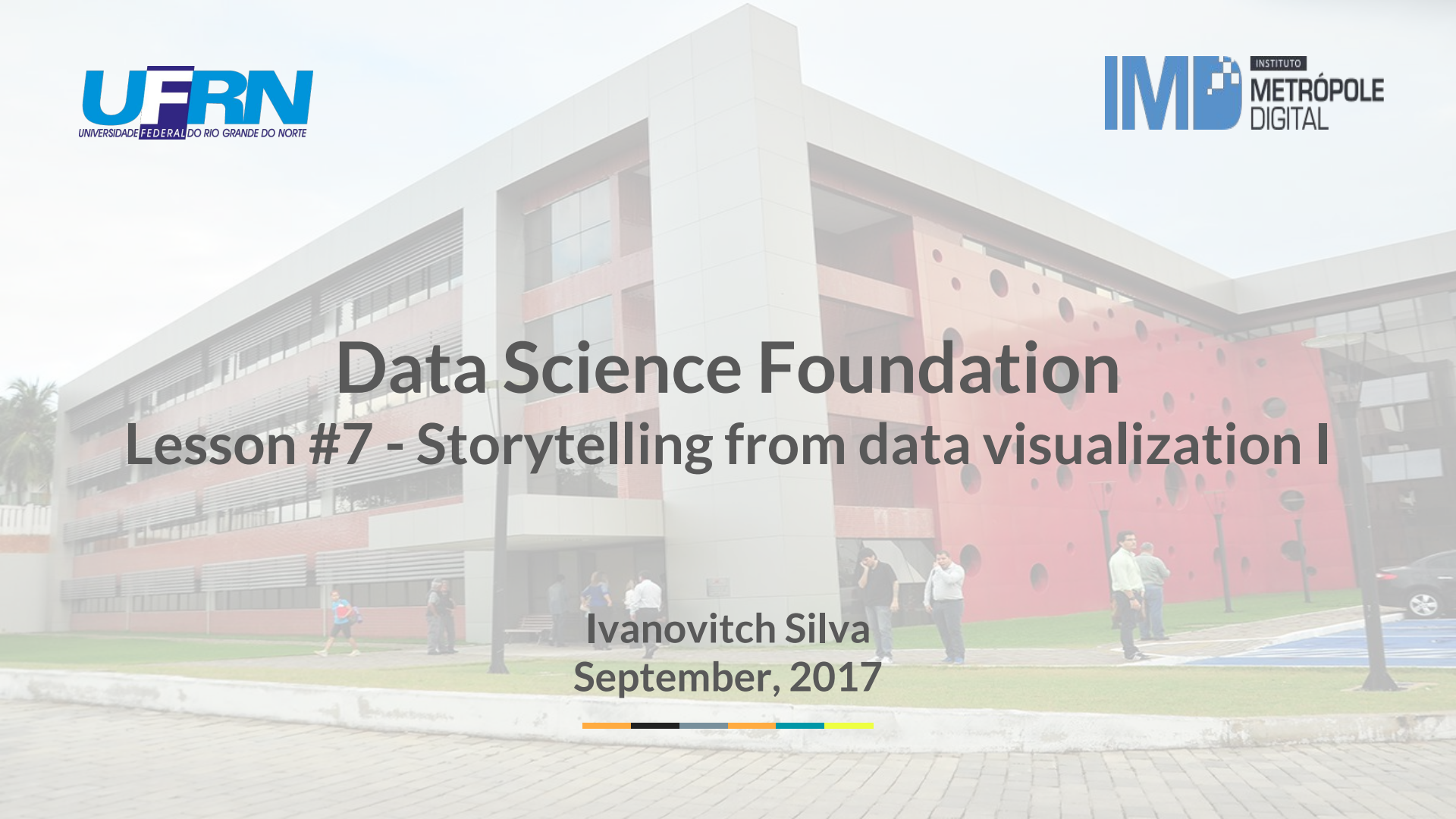


Data Science Foundation

Lesson #7 - Storytelling from data visualization I

Ivanovitch Silva

September, 2017

Agenda

- Case study: visualizing gender gap
- Data Ink Ratio
- Hiding tick marks, spines
- Color, layout, line width
- Legend and annotations

Update the repository

```
git clone https://github.com/ivanovitchm/EEC2006.git
```

Or ....

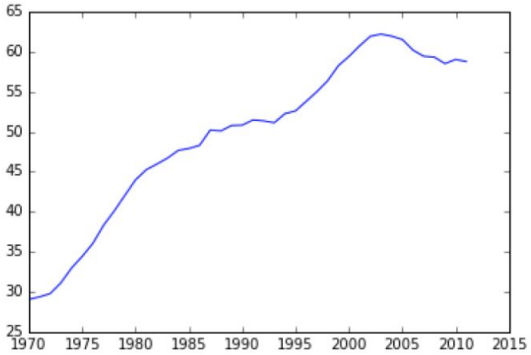
```
git pull
```

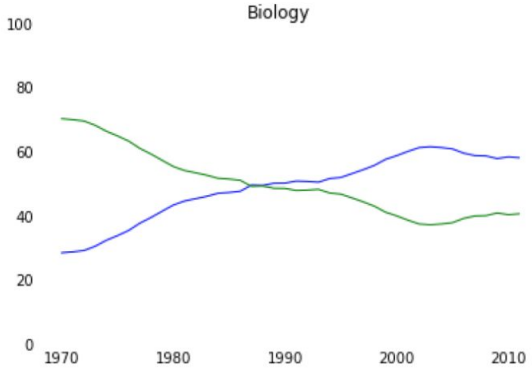
Motivation

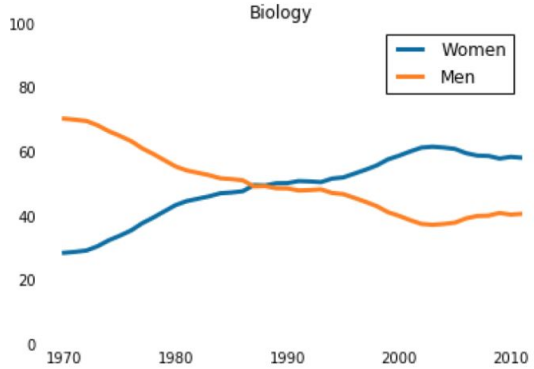
Year	Agriculture	Architecture	Art and Performance	Biology	Business	Communications and Journalism	Computer Science	Education	Engineering
1970	4.229798	11.921005	59.7	29.088363	9.064439	35.3	13.6	74.535328	0.8
1971	5.452797	12.003106	59.9	29.394403	9.503187	35.5	13.6	74.149204	1.0
1972	7.420710	13.214594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

The Department of Education Statistics releases a data set annually containing the percentage of bachelor's degrees granted to women from 1970 to 2012.

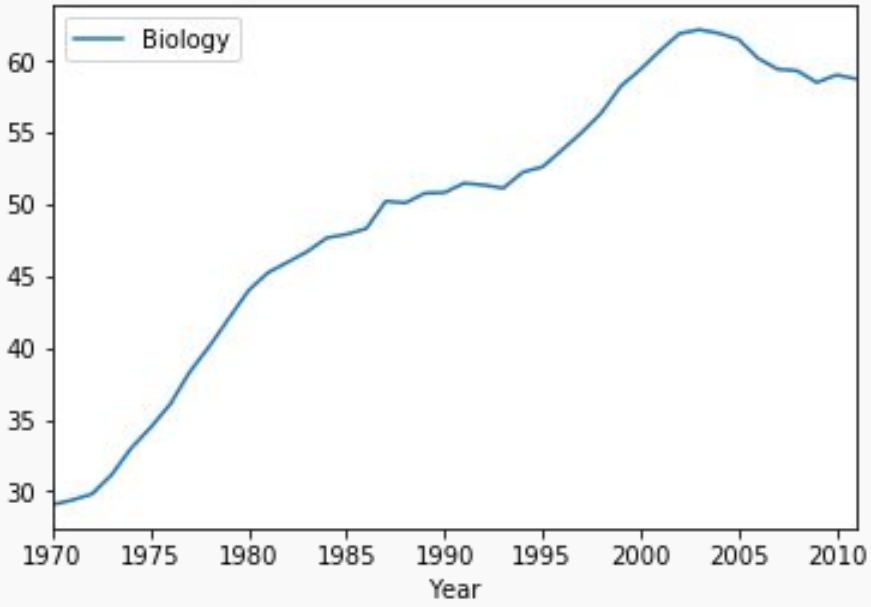
Improving Plot Aesthetics





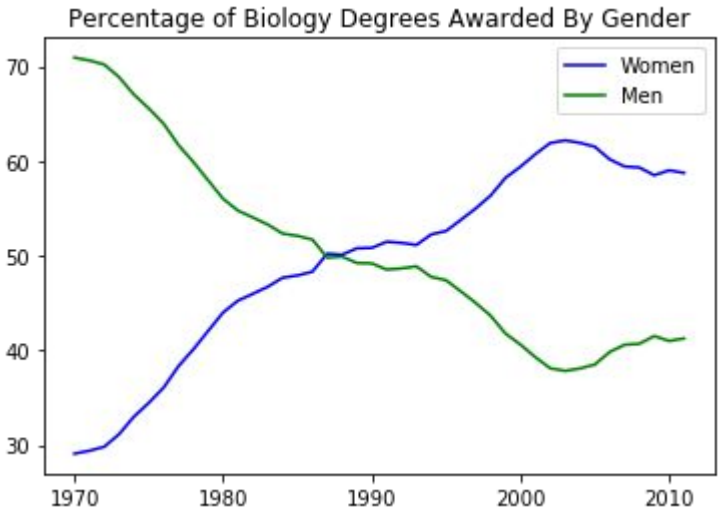


Visualizing the gender gap



- Biology degrees increased steadily from 1970 and peaked in the early 2000's.
- It only tells half the story.

Visualizing the gender gap




```
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()
```

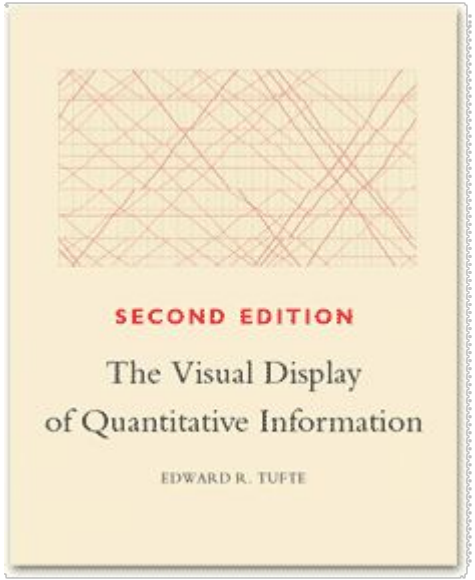
Visualizing the gender gap

```
%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'])
```

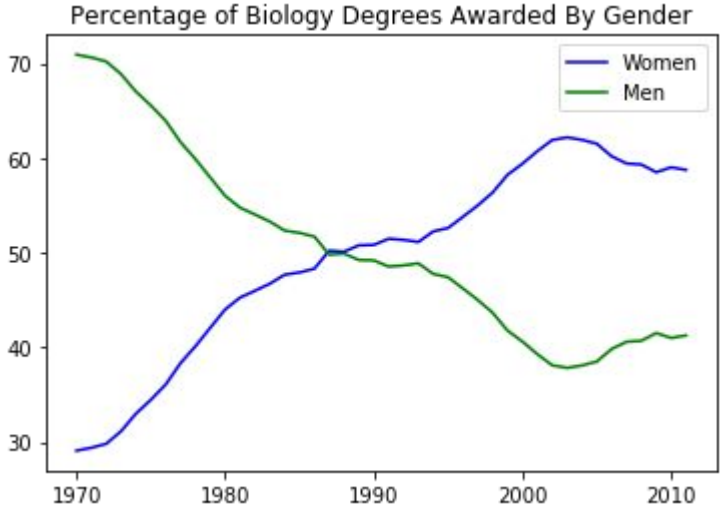
```
ax = women_degrees.plot(kind='line',x='Year',y=['Biology','men_bio'],
                   title='Percentage of Biology Degrees Awarded By Gender',
                   color=['blue','green'])
ax.legend(loc='best',labels=['Women','Men'])
```

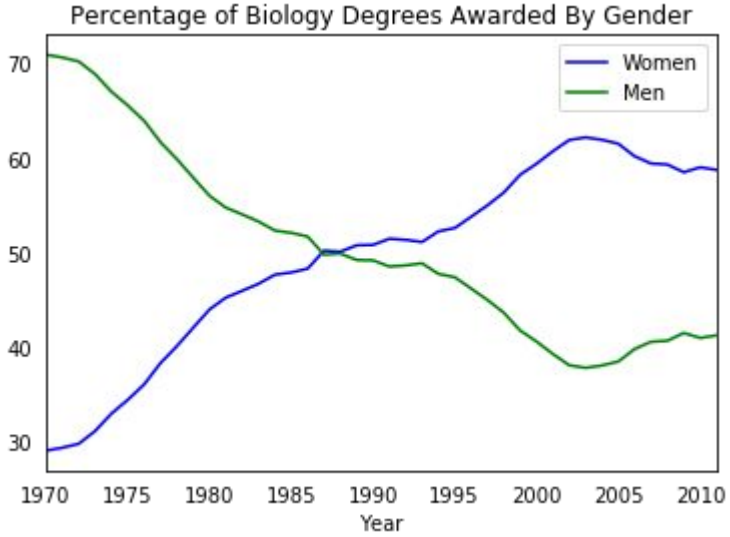
Data Ink Ratio

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



Hiding tick marks

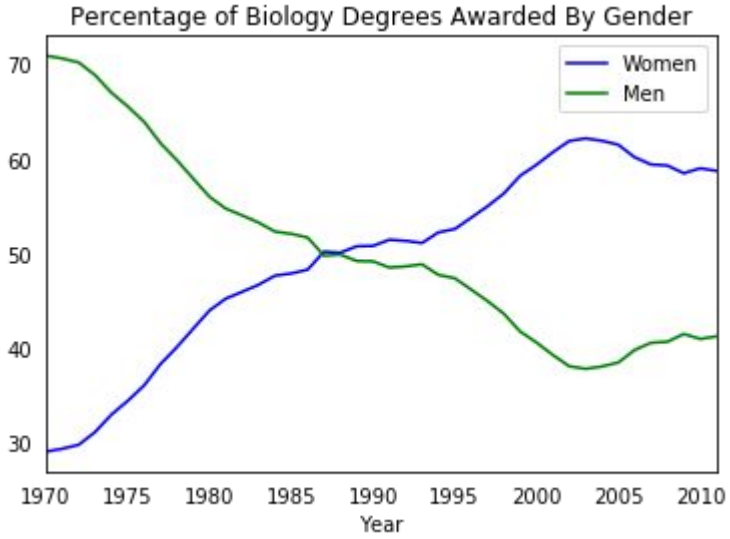


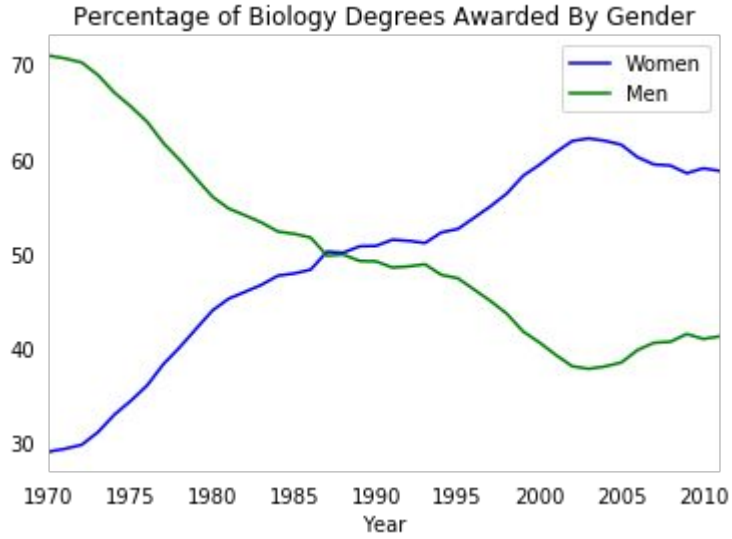



```
ax.tick_params(bottom="off", top="off", left="off", right="off")
```

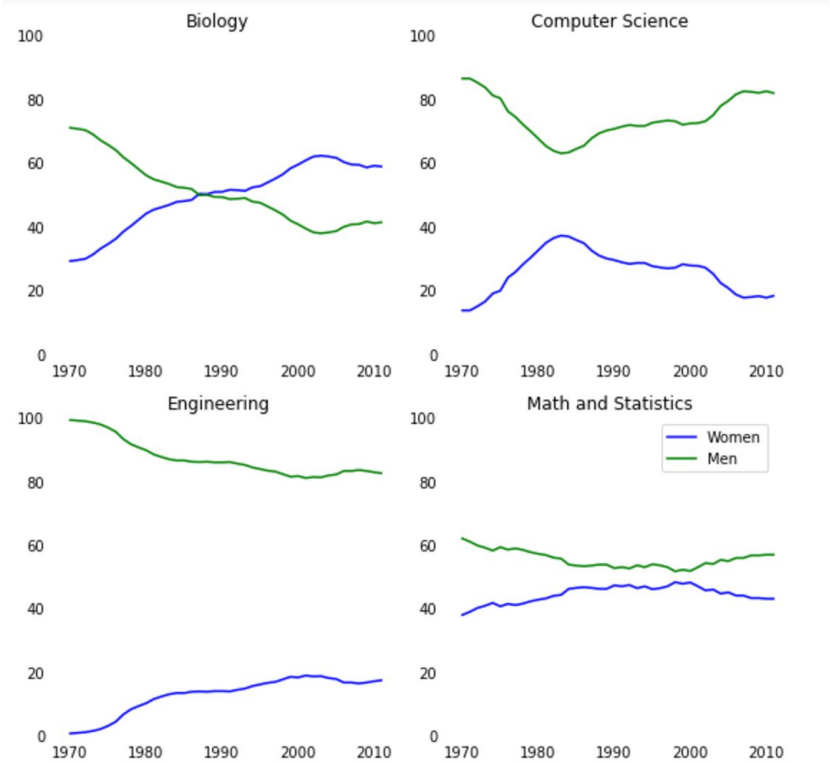
Hiding spines

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



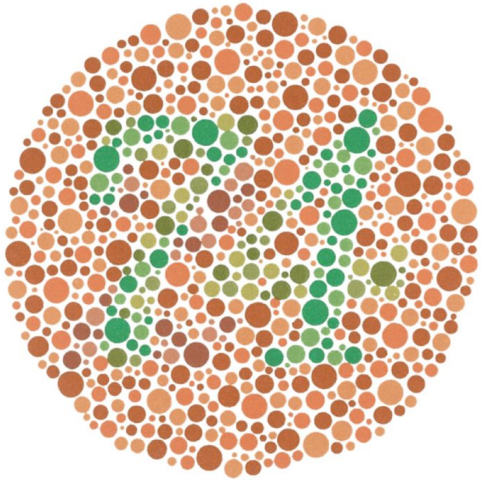


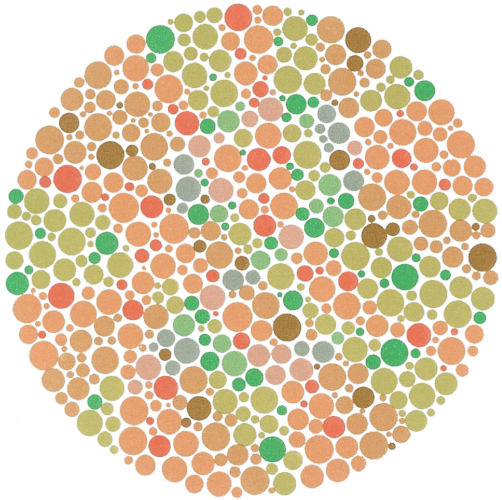
Comparing gender gap across degree categories

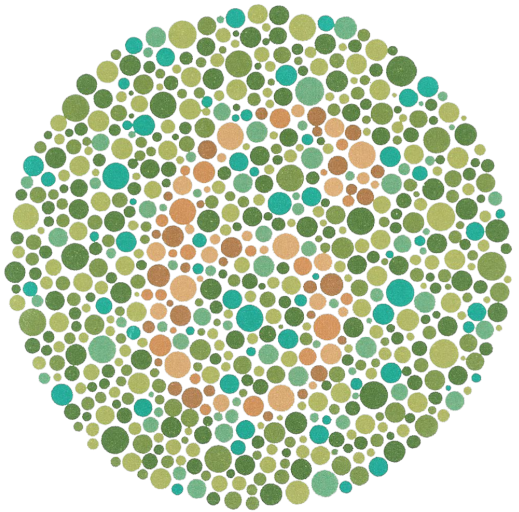


Storytelling I.ipynb

Color

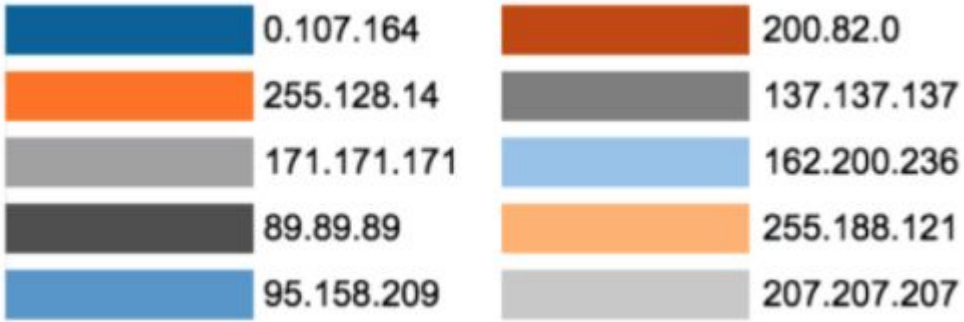






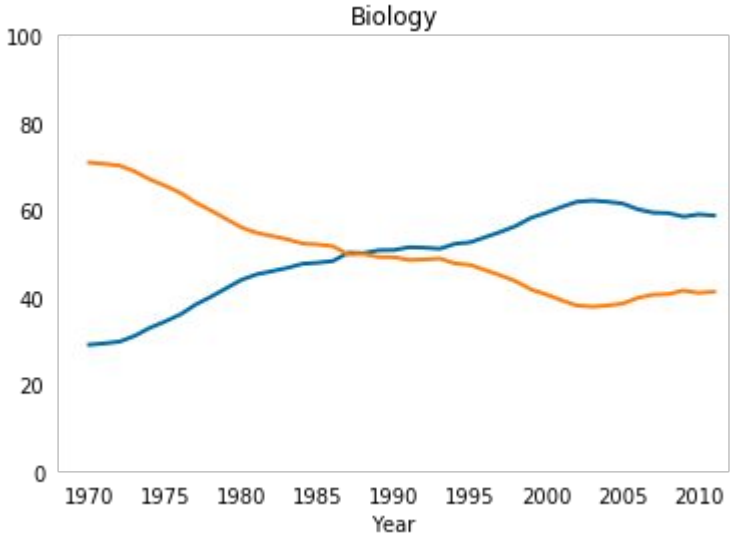
https://en.wikipedia.org/wiki/Ishihara_test

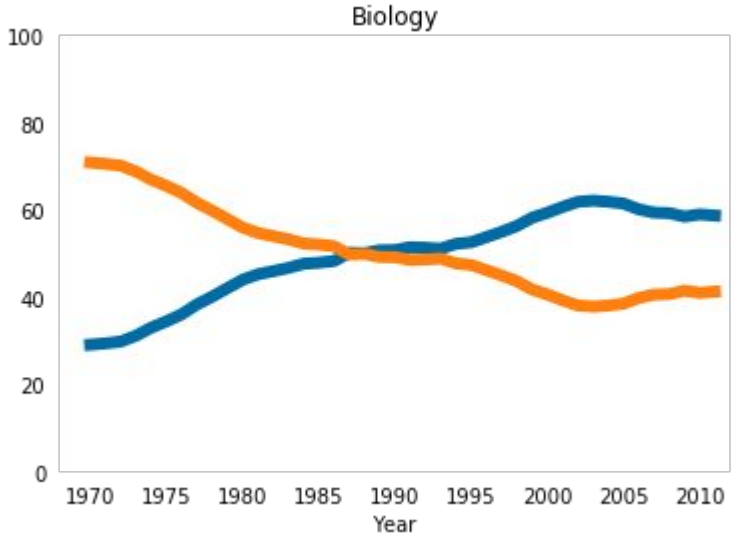
Setting line color using RGB




```
cb_dark_blue = (0/255,107/255,164/255)
women_degrees.plot(kind='line',x='Year',y='Biology', color=cb_dark_blue)
```

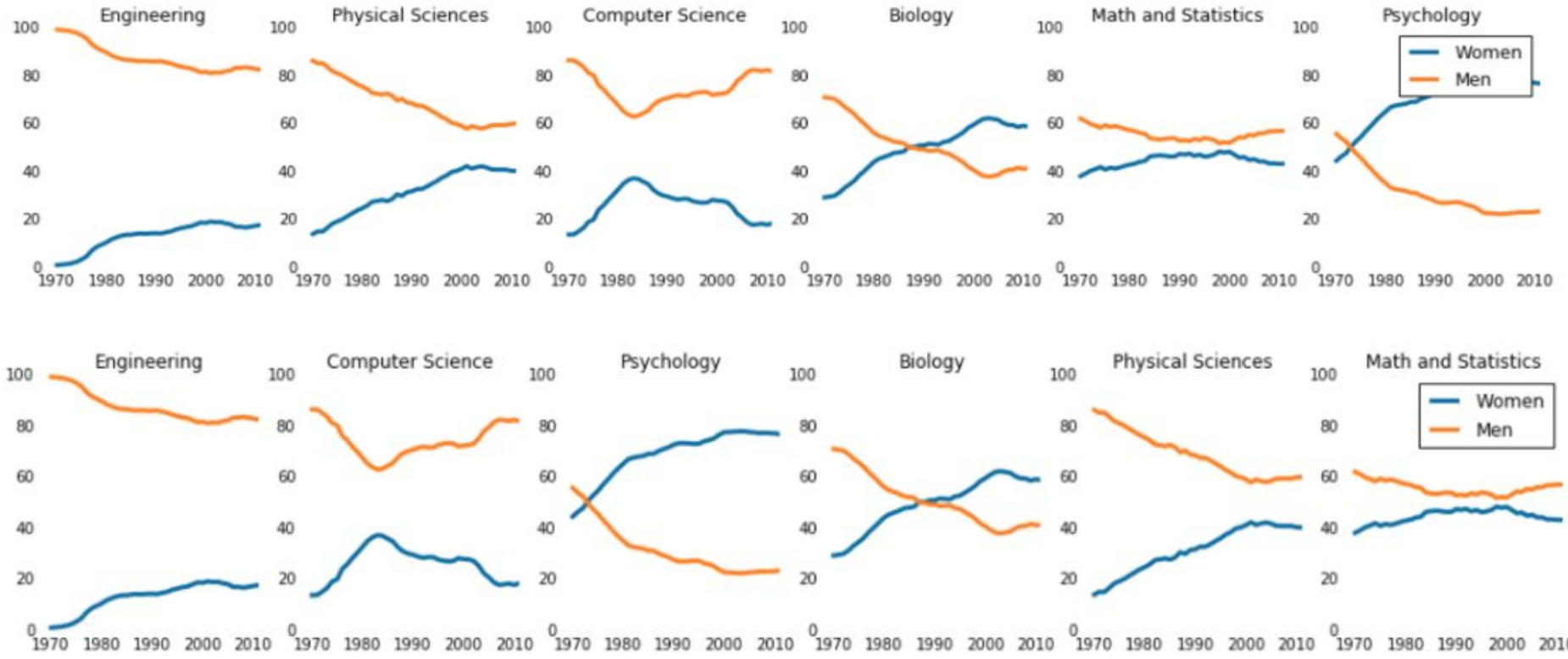
Setting line width



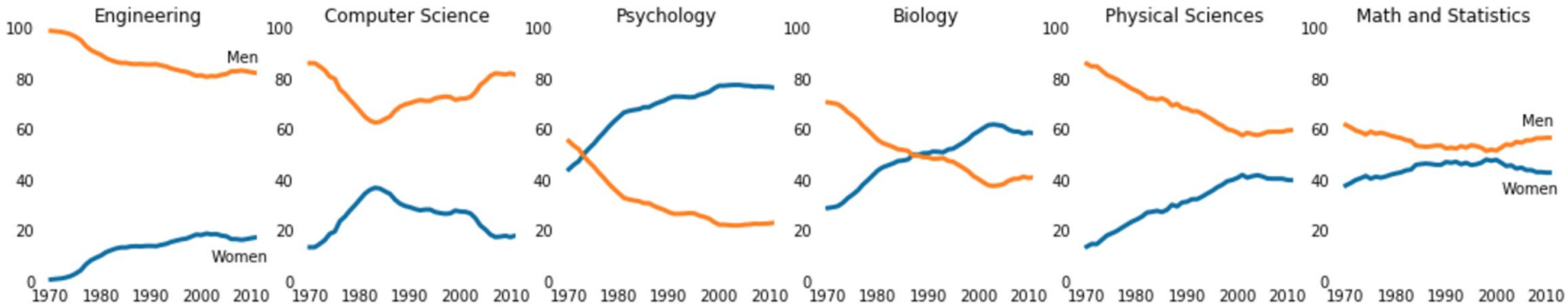



```
women_degrees.plot(kind='line',x='Year',y='Biology', linewidth=2)
```

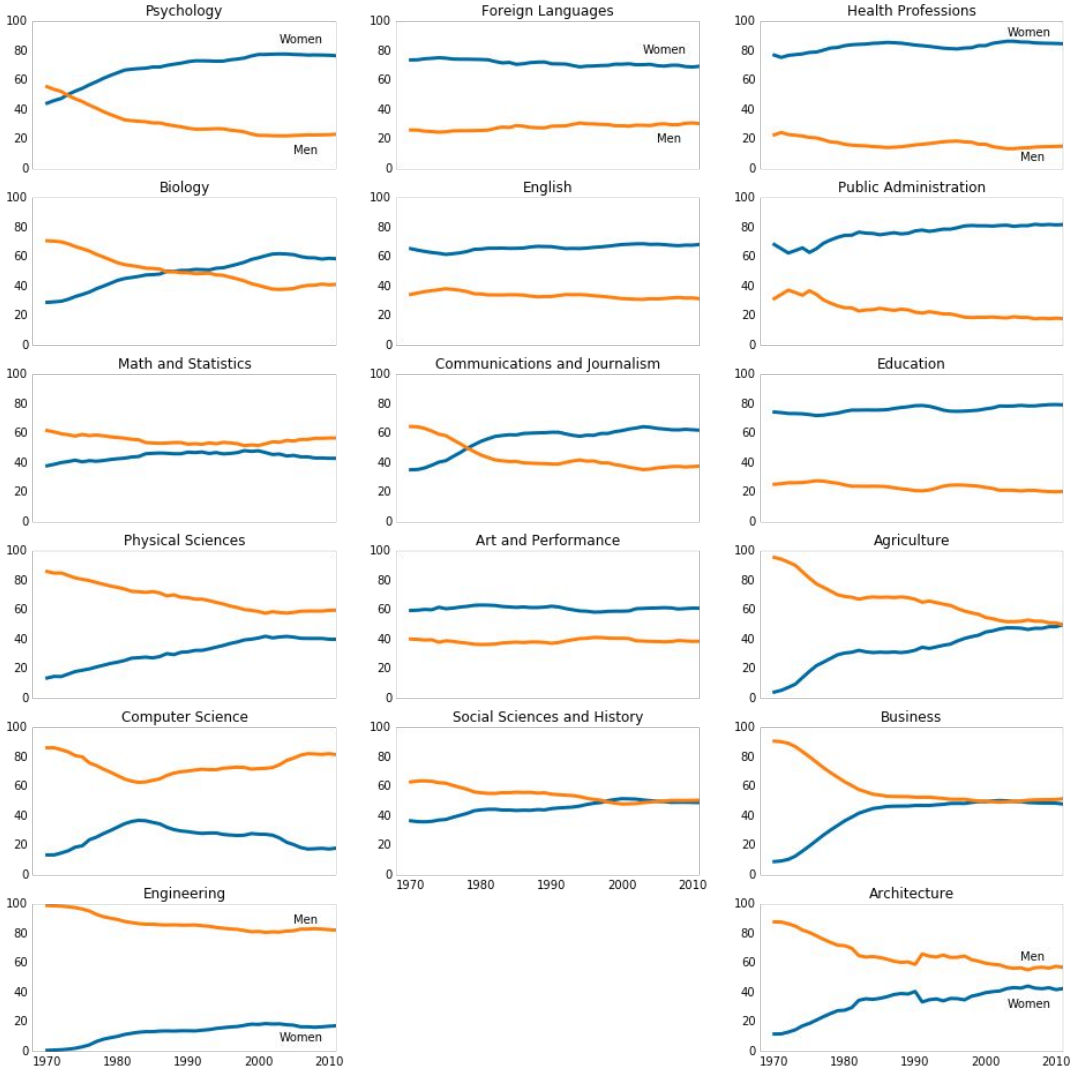
Improve the layout and ordering



Replace the legend with annotations




```
ax.text(1970, 0, "starting point")
```


Psychology
Women
Men
Foreign Languages
Women
Men
Health Professions
Women
Men
Biology
English
Public Administration
Math and Statistics
Communications and Journalism
Education
Physical Sciences
Art and Performance
Agriculture
Computer Science
Social Sciences and History
Business
Engineering
Men
Women
Architecture
Men
Women
0
20
40
60
80
100
1970
1980
1990
2000
2010

Storytelling II.ipynb