

Data Science & ML Course

Lesson #8 Exploratory Data Analysis III

Ivanovitch Silva
October, 2018



Agenda

- Case study: gender gap
- Aesthetics
- Colors, Lines width
- Annotations

Update from repository

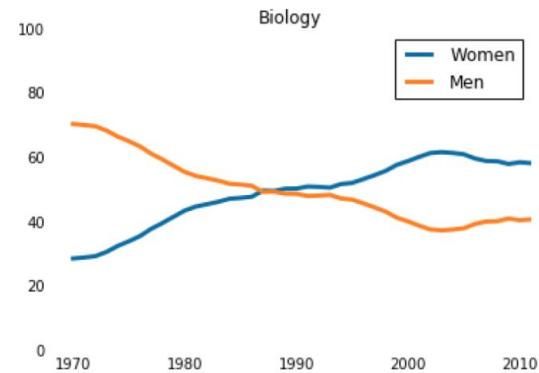
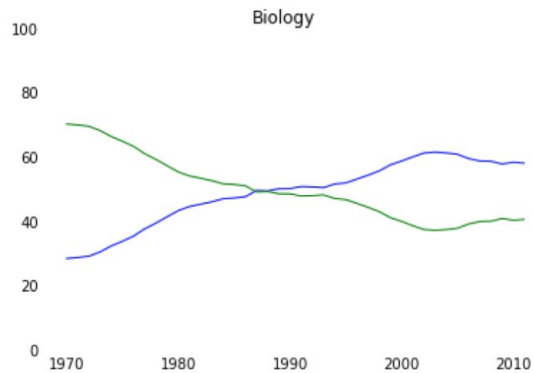
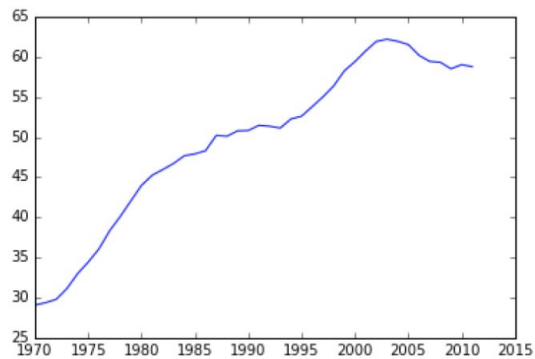
```
git clone https://github.com/ivanovitchm/datascience2machinelearning.git
```

Or

```
git pull
```



Aesthetics

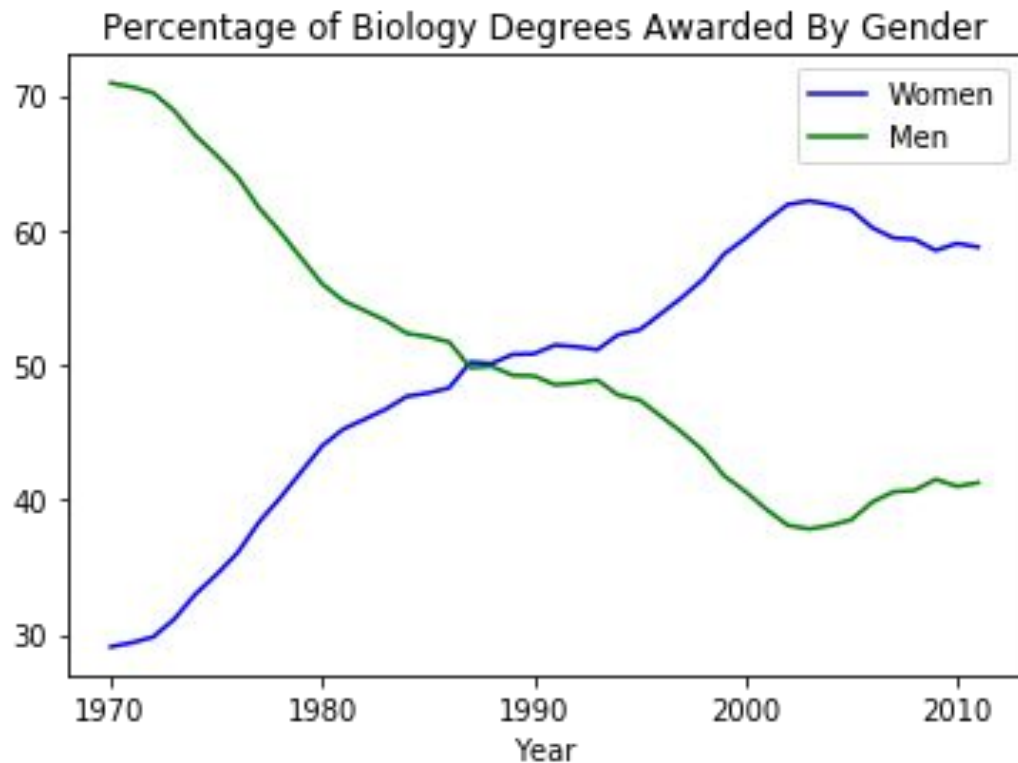


Introduction to the data

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.

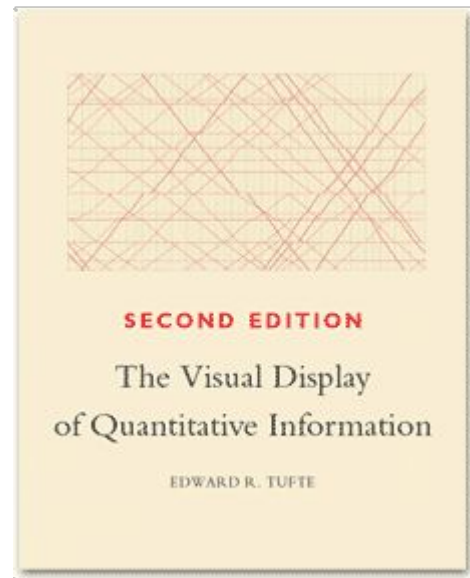
The dataset contains the percentage of bachelor's degrees granted to women from 1970 to 2012



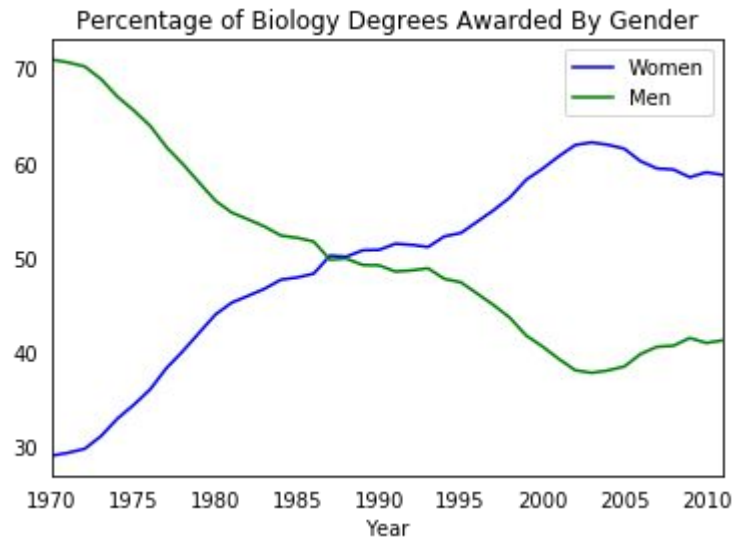
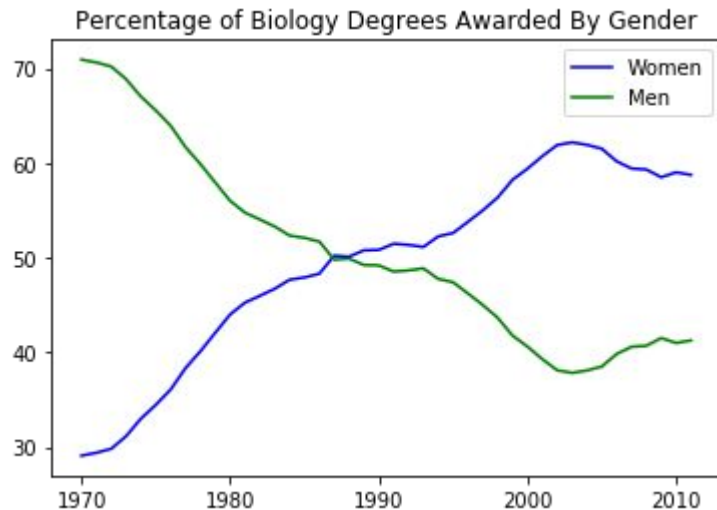
```
ax = women_degrees.plot(kind='line',x='Year',y=['Biology','men_bio'],  
                        title='Percentage of Biology Degrees Awarded By Gender',  
                        color=['blue','green'])  
ax.legend(loc='best',labels=['Women','Men'])
```

Data Link Ratio

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



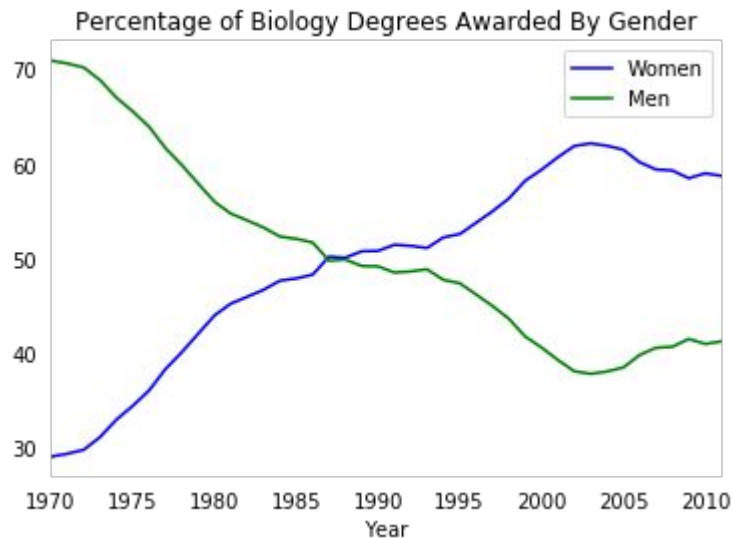
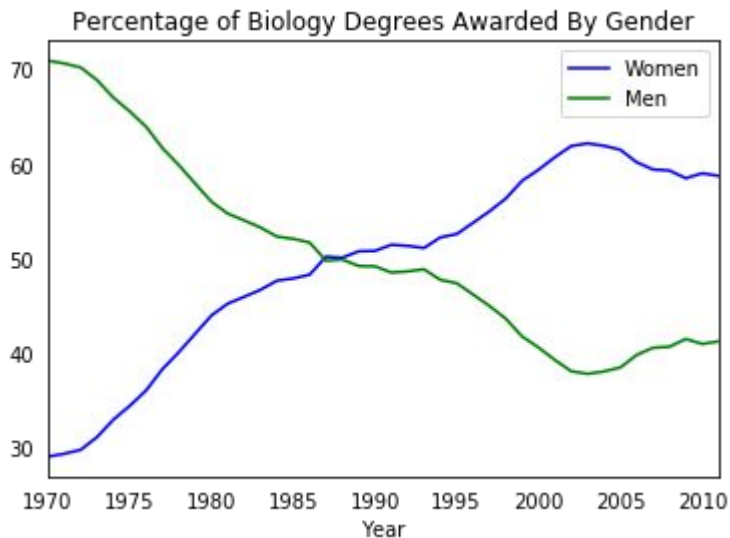
Hiding tick marks

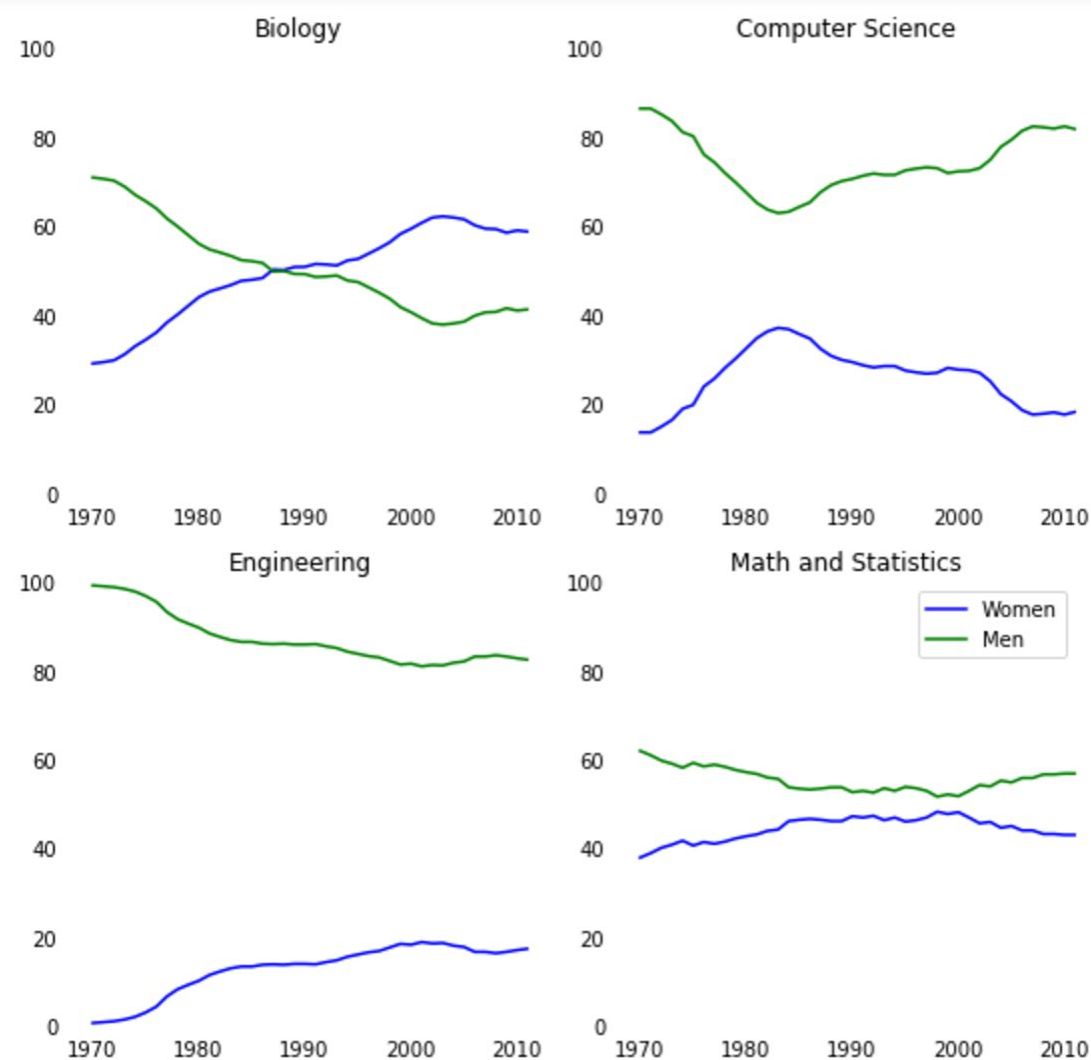


```
ax.tick_params(bottom=False, top=False, left=False, right=False)
```



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





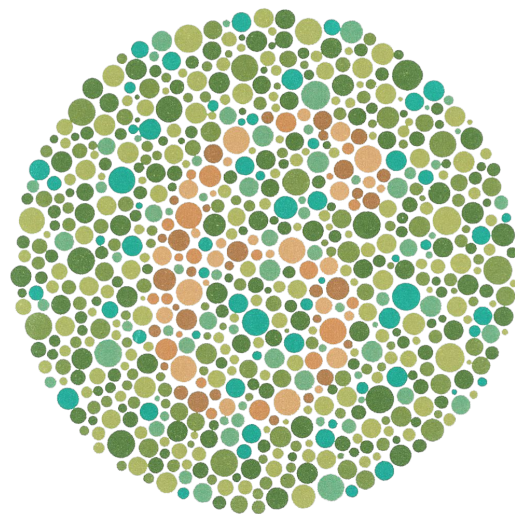
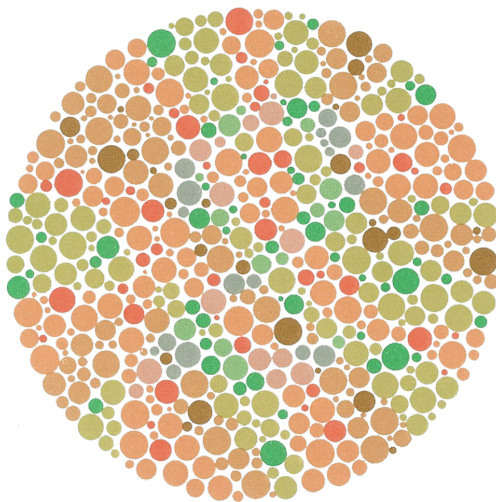
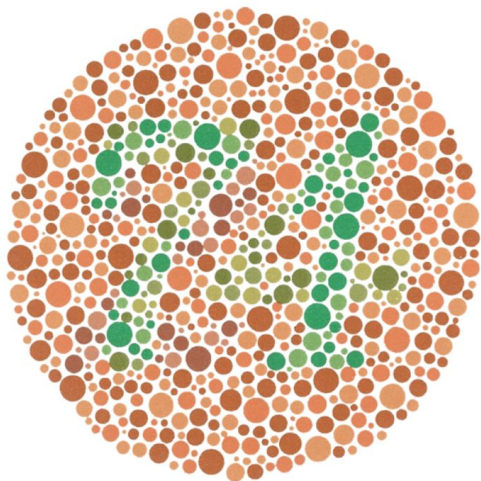
Comparing Gender Gap Across Degree Categories

Lesson #8 - Exploratory Data Analysis III.ipynb











Section 1



Colors



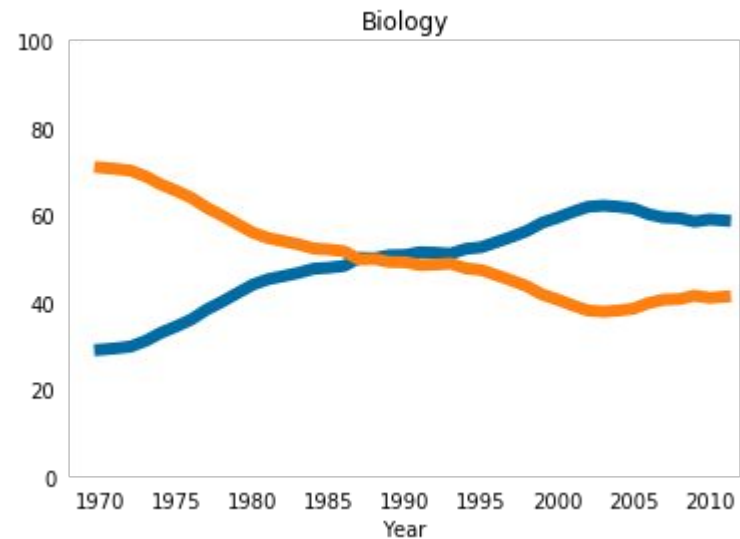
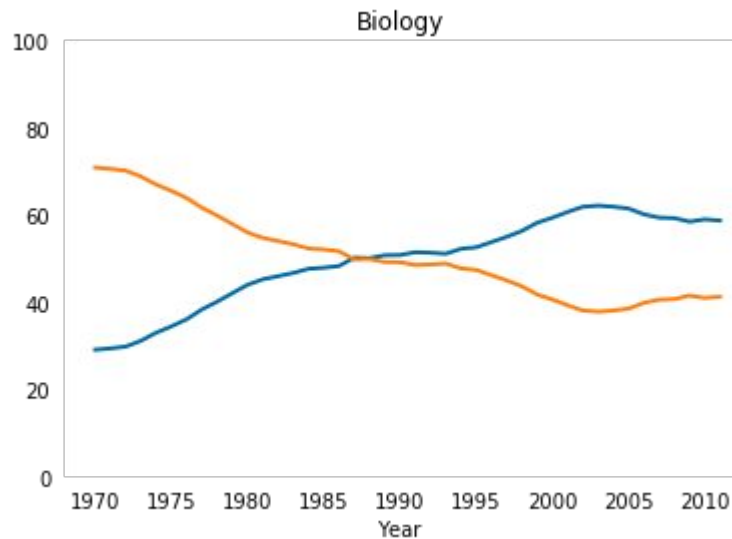
Setting Line Color Using RGB

	0.107.164		200.82.0
	255.128.14		137.137.137
	171.171.171		162.200.236
	89.89.89		255.188.121
	95.158.209		207.207.207

```
cb_dark_blue = (0/255,107/255,164/255)
```

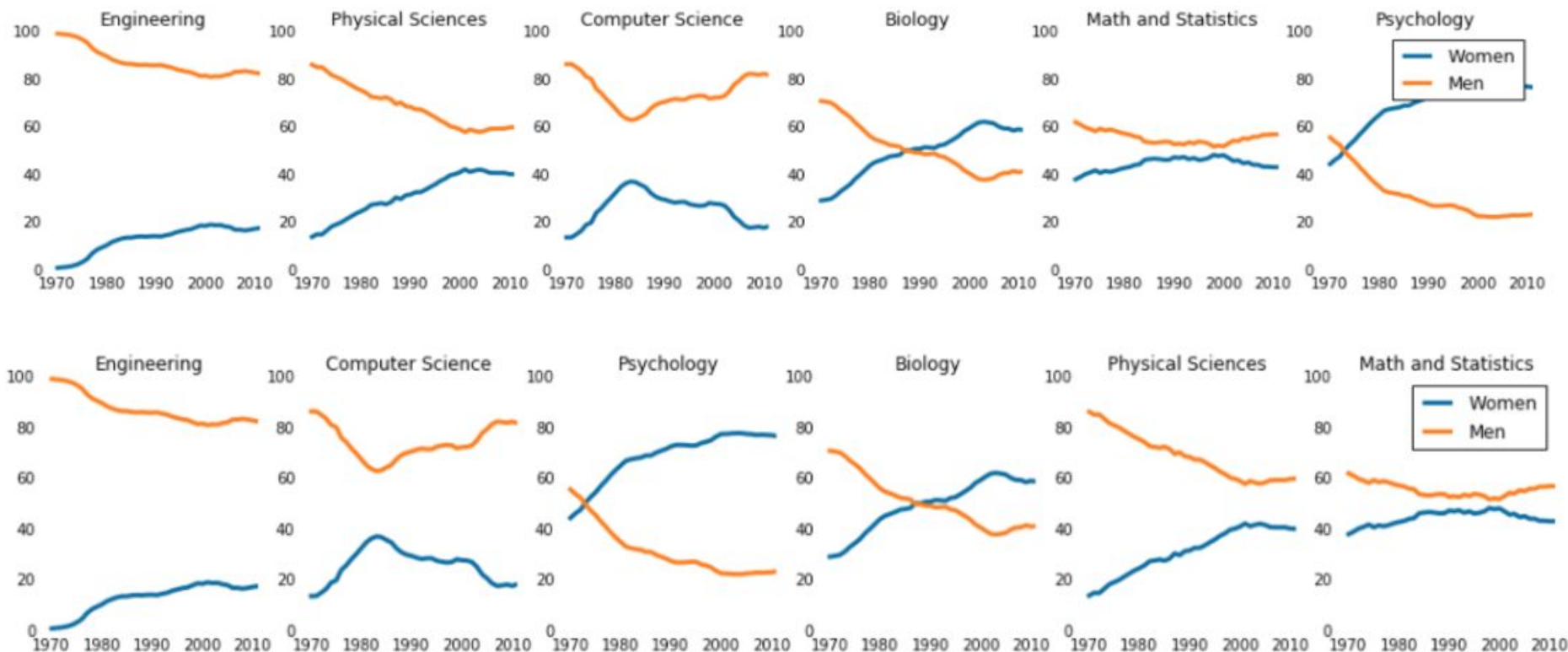
```
women_degrees.plot(kind='line',x='Year',y='Biology', color=cb_dark_blue)
```

Setting Lines Width

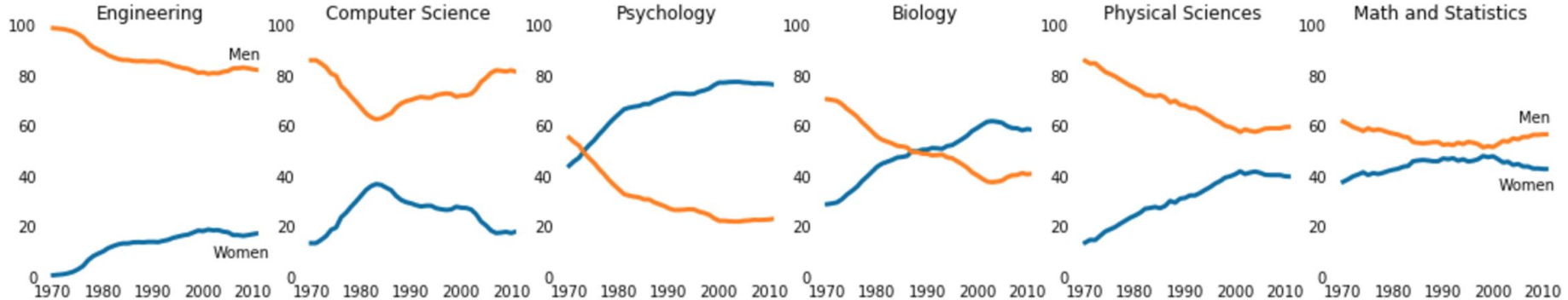


```
women_degrees.plot(kind='line',x='Year',y='Biology', linewidth=2)
```

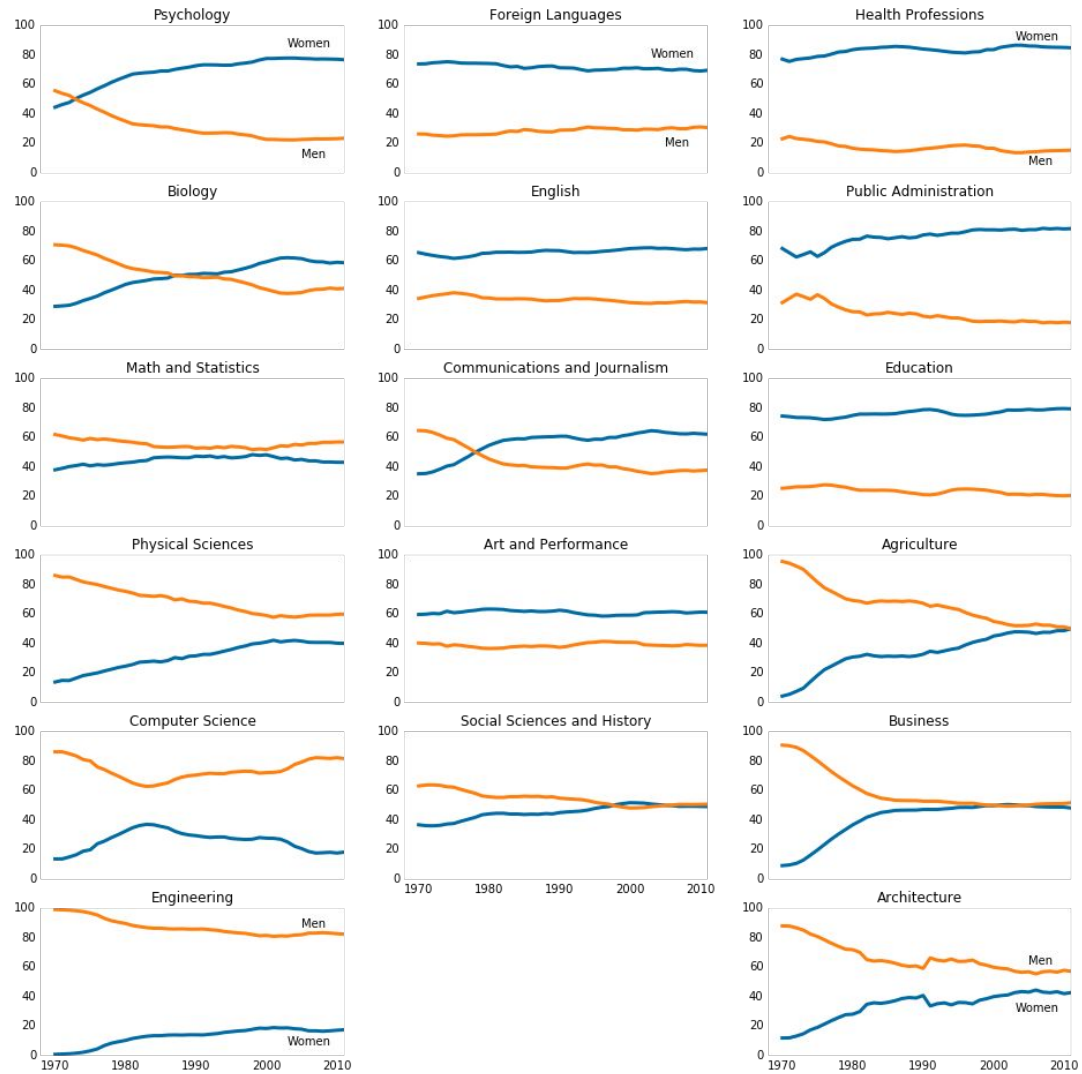
Improve the Layout and Ordering



Replacing the Legend with Annotations



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

Lesson #8 - Exploratory Data Analysis III.ipynb

Section 2

