Visualizing the Gender Gap in Various College Majors

February 8, 2021

1 Visualizing the Gender Gap in Various College Majors

In this project, I will be exploring the gender gap across college majors through various visualizations in Matplotlib. My aim is to illustrate gender discrepancies between majors – especially those in STEM fields. The dataset I will be using has been cleaned and provided by University of Pennslyvania professor Randal Olson, and can be found on his personal website here. It explores the percentages of men and women from 1968 to 2011 who have comprised different majors.

First, we'll begin by reading in the data.

[1]:
```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt

women_degrees = pd.read_csv('/Users/natasharavinand/Downloads/datasets/Projects/
↪percent-bachelors-degrees-women-usa.csv')
cb_dark_blue = (0/255,107/255,164/255)
cb_orange = (255/255, 128/255, 14/255)
```

There are 17 majors included in the dataset. In order to visualize this majors effectively, we will group them into one of three categories:

- **STEM** (Science, Technology, Engineering, Mathematics)
  - *Psychology, Biology, Math and Statistics, Physical Sciences, Computer Science, Engineering* ***
- **Liberal Arts**
  - *Foreign Languages, English, Communications and Journalism, Art and Performance, Social Sciences and History* ***
- **Other**
  - *Health Professions, Public Administration, Education, Agriculture, Business, Architecture*

We'll group these categories into three lists to use as reference.

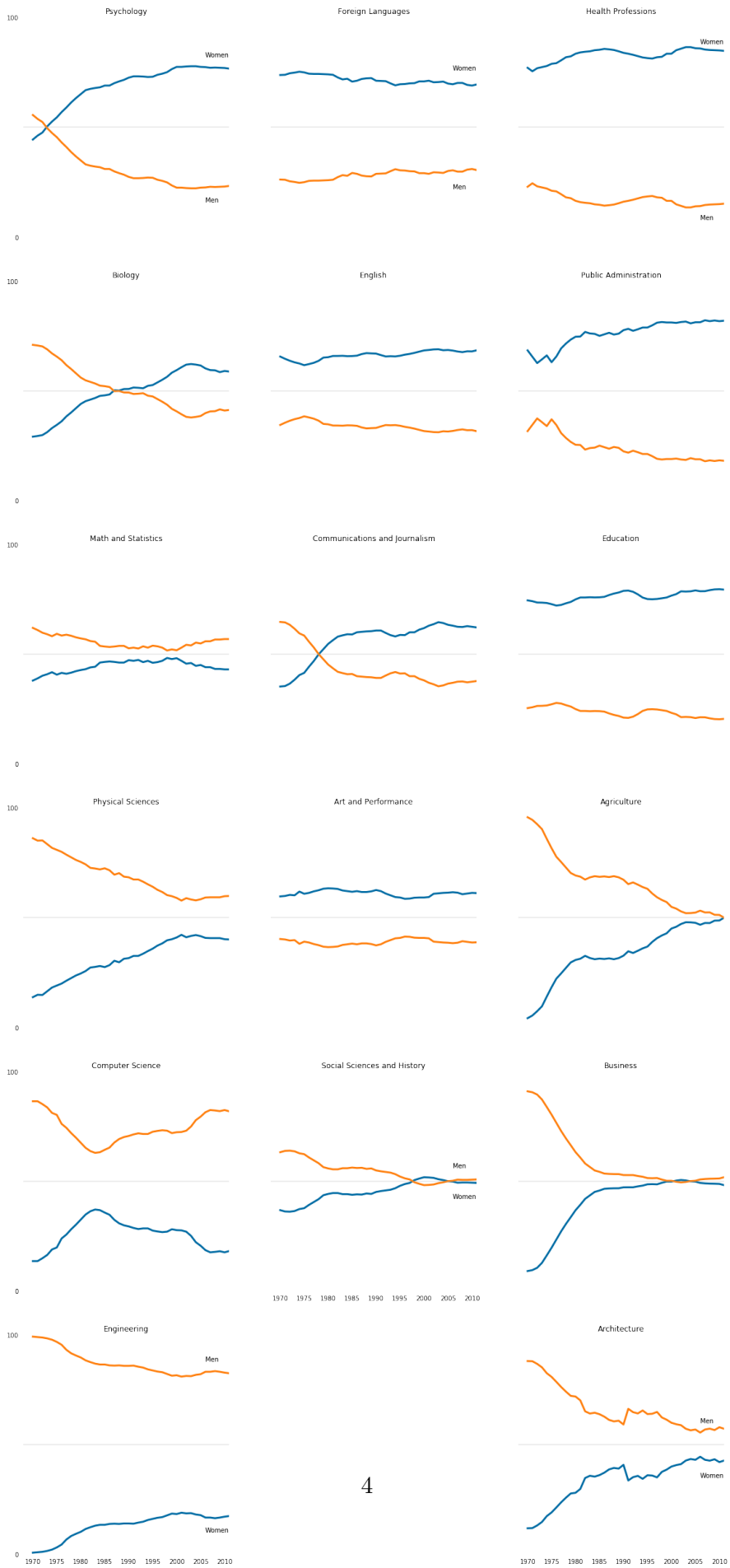
[2]:
```
stem_cats = ['Psychology', 'Biology', 'Math and Statistics', 'Physical␣
↪Sciences', 'Computer Science', 'Engineering']

lib_arts_cats = ['Foreign Languages', 'English', 'Communications and␣
↪Journalism', 'Art and Performance', \
```

```
'Social Sciences and History']

other_cats = ['Health Professions', 'Public Administration', 'Education',␣
↪'Agriculture','Business', 'Architecture']
```

Now, we'll create a plot, with multiple subplots, that displays the trends of all these majors.

[7]:
```
# Graphing STEM degrees

fig = plt.figure(figsize=(20, 45))

for i in range(0, 18, 3):
    cat_index = int(i/3)
    ax = fig.add_subplot(6, 3, i+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[cat_index]],␣
↪color=cb_dark_blue, linewidth=3)
    ax.plot(women_degrees['Year'], 100 - women_degrees[stem_cats[cat_index]],␣
↪color=cb_orange, linewidth=3)
    ax.set_title(stem_cats[cat_index])
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0, 100)
    ax.tick_params(bottom=False, top=False, left=False, right=False)
    for key, spine in ax.spines.items():
        spine.set_visible(False)
    if i == 0:
        ax.text(2006, 82, "Women")
        ax.text(2006, 16, "Men")
    if i == 15:
        ax.text(2006, 88, "Men")
        ax.text(2006, 10, "Women")
        ax.tick_params(labelbottom=True)
    if i != 15:
        ax.tick_params(labelbottom=False)
    ax.set_yticks([0,100])
    ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3) #shows the 50% mark


# Graphing liberal arts degrees

for i in range(1, 16, 3):
    cat_index = int(i/3)
    ax = fig.add_subplot(6, 3, i+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[cat_index]],␣
↪color=cb_dark_blue, linewidth=3)
    ax.plot(women_degrees['Year'], 100 -␣
↪women_degrees[lib_arts_cats[cat_index]], color=cb_orange, linewidth=3)
    ax.set_title(lib_arts_cats[cat_index])
```

```
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0, 100)
    ax.tick_params(bottom=False, top=False, left=False, right=False)
    for key, spine in ax.spines.items():
        spine.set_visible(False)
    if i == 1:
        ax.text(2006, 76, "Women")
        ax.text(2006, 22, "Men")
    if i == 13:
        ax.text(2006, 56, "Men")
        ax.text(2006, 42, "Women")
        ax.tick_params(labelbottom=True)
    if i != 13:
        ax.tick_params(labelbottom=False)
    ax.set_yticks([])
    ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3) #shows the 50% mark


# Graphing other degrees

for i in range(2, 20, 3):
    cat_index = int(i/3)
    ax = fig.add_subplot(6, 3, i+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[cat_index]],␣
 ↪color=cb_dark_blue, linewidth=3)
    ax.plot(women_degrees['Year'], 100 - women_degrees[other_cats[cat_index]],␣
 ↪color=cb_orange, linewidth=3)
    ax.set_title(other_cats[cat_index])
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0, 100)
    ax.tick_params(bottom=False, top=False, left=False, right=False)
    for key, spine in ax.spines.items():
        spine.set_visible(False)
    if i == 2:
        ax.text(2006, 88, "Women")
        ax.text(2006, 8, "Men")
    if i == 17:
        ax.text(2006, 60, "Men")
        ax.text(2006, 35, "Women")
        ax.tick_params(labelbottom=True)
    if i != 17:
        ax.tick_params(labelbottom=False)
    ax.set_yticks([])
    ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3) #shows the 50% mark

plt.show()
```


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
Men
Women
Business
Engineering
Men
Women
Architecture
Men
Women
100
0
1970 1975 1980 1985 1990 1995 2000 2005 2010

1.1 Trends

We see some clear trends in the graphs pictured above. They include:

- Computer science and engineering majors tend to show the greatest gender discrepancy

---

- Biology and Psychology are regarded as STEM majors that have become majority women through the years

---

- Almost an equal percentage of men and women study social science and history are the collegiate level

---

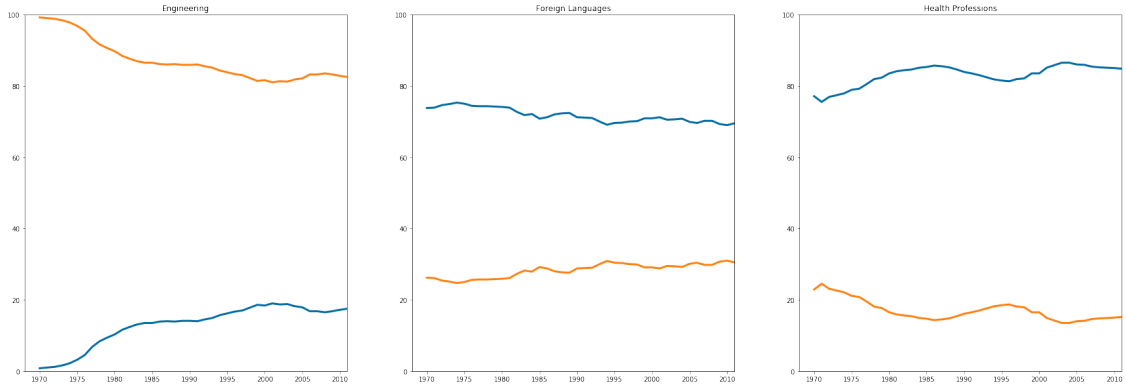
- Fields such as business, architecture, and agriculture have evolved over time, with the percentage of men decreasing from nearly 90% to nearly 50%

We can see clearly, however, that the largest gender gaps tend to be within the engineering and health professions majors. To illustrate, we will display the major with the highest gender gap from each of the three categories.

[4]:
```
fig = plt.figure(figsize=(30, 10))

ax = fig.add_subplot(1, 3,1)
ax.plot(women_degrees['Year'], women_degrees['Engineering'], c=cb_dark_blue,␣
 ↪linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Engineering'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Engineering")

ax = fig.add_subplot(1, 3,2)
ax.plot(women_degrees['Year'], women_degrees['Foreign Languages'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Foreign Languages'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Foreign Languages")

ax = fig.add_subplot(1, 3, 3)
```

```
ax.plot(women_degrees['Year'], women_degrees['Health Professions'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Health Professions'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Health Professions")
```

[4]: `Text(0.5, 1.0, 'Health Professions')`



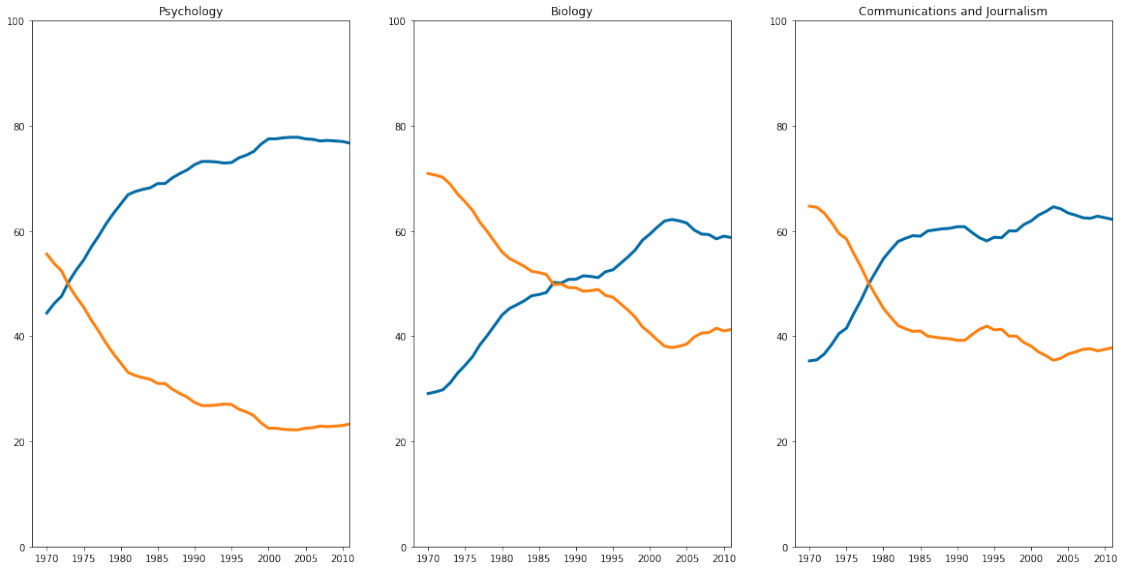
We see clearly that the Engineering and Health Professions majors have the largest gaps. This could be the result of societal bias against women or men, respectively, entering these fields, encouraging less of them to study at the college-level. We see a historical gap with both these majors – since 1968, they has been a fairly significant gender gap.

However, not all majors exhibited this characteristic. Some majors evolved from male-dominated to female dominated. Some are shown below:

[5]:
```
fig = plt.figure(figsize=(20, 10))

ax = fig.add_subplot(1, 3,1)
ax.plot(women_degrees['Year'], women_degrees['Psychology'], c=cb_dark_blue,␣
 ↪linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Psychology'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Psychology")

ax = fig.add_subplot(1, 3,2)
ax.plot(women_degrees['Year'], women_degrees['Biology'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Biology'],␣
 ↪c=cb_orange,linewidth=3)
```

```
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Biology")

ax = fig.add_subplot(1, 3, 3)
ax.plot(women_degrees['Year'], women_degrees['Communications and Journalism'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Communications and␣
 ↪Journalism'], c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Communications and Journalism")
```

[5]:
```
Text(0.5, 1.0, 'Communications and Journalism')
```

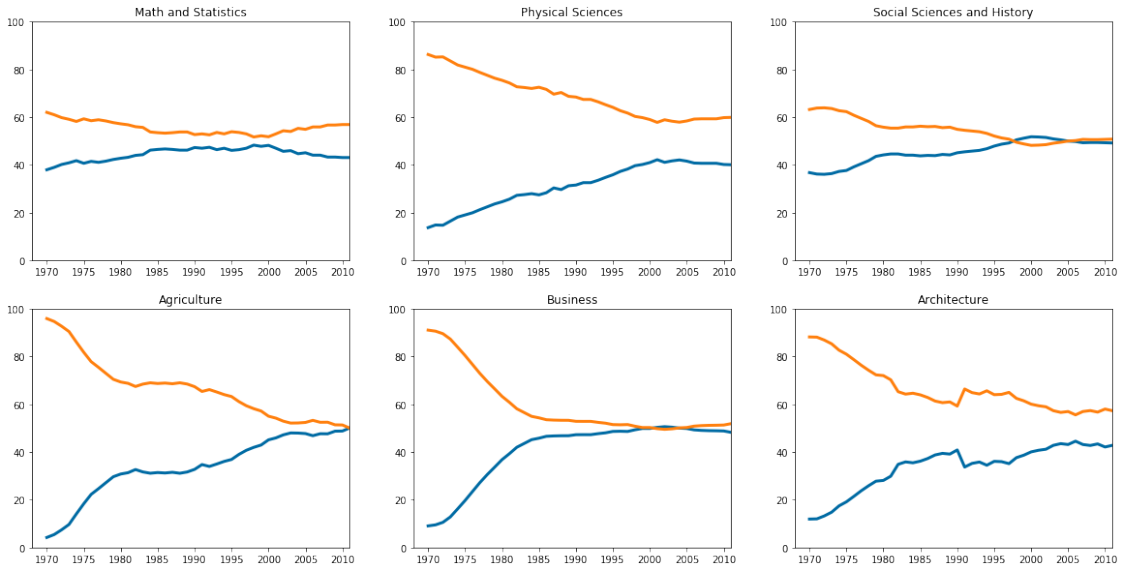


Other majors have evolved from male dominated to nearly equal percentages:

[6]:
```
fig = plt.figure(figsize=(20, 10))

ax = fig.add_subplot(2,3,1)
ax.plot(women_degrees['Year'], women_degrees['Math and Statistics'],␣
 ↪c=cb_dark_blue, linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Math and Statistics'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Math and Statistics")
```

```
ax = fig.add_subplot(2, 3,2)
ax.plot(women_degrees['Year'], women_degrees['Physical Sciences'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Physical Sciences'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Physical Sciences")

ax = fig.add_subplot(2, 3, 3)
ax.plot(women_degrees['Year'], women_degrees['Social Sciences and History'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Social Sciences and␣
 ↪History'], c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Social Sciences and History")

ax = fig.add_subplot(2, 3, 4)
ax.plot(women_degrees['Year'], women_degrees['Agriculture'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Agriculture'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Agriculture")

ax = fig.add_subplot(2, 3, 5)
ax.plot(women_degrees['Year'], women_degrees['Business'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Business'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Business")

ax = fig.add_subplot(2, 3, 6)
ax.plot(women_degrees['Year'], women_degrees['Architecture'],␣
 ↪c=cb_dark_blue,linewidth=3)
ax.plot(women_degrees['Year'], 100 - women_degrees['Architecture'],␣
 ↪c=cb_orange,linewidth=3)
ax.set_xlim(1968, 2011)
ax.set_ylim(0, 100)
ax.set_title("Architecture")
```

[6]:
```
Text(0.5, 1.0, 'Architecture')
```



However, none of the majors that were female dominated became male dominated, or even reached equal percentages. This could be the result of societal bias and gender labeling of these majors or fields.