

# Visualizing The Gender Gap Across College Degrees

***Author : Kenneth Lucas Kusima***

***Attributor: DataQuest***

## Project Origin:

This project was made per instructions from:

**The DataQuest Course: StoryTelling Through Data Visualization**

## Objectives:

The aim is to visualize the gender gap across the plethora of college degrees. the degree majors are grouped in three and are represented as follows:

### I. STEM (Science, Technology, Engineering and Mathematics):

- Psychology
- Biology
- Math and Statistics
- Physical Sciences
- Computer Science
- Engineering

### II. Liberal Arts

- Foreign Languages
- English
- Communications and Journalism
- Art and Performance
- Social Sciences and History

### III. Others

- Health Professions
- Public Administration
- Education
- Agriculture
- Business
- Architecture

In [38]:

```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt

#Reading the dataset
women_degrees = pd.read_csv('percent-bachelors-degrees-women-usa.csv')
#Color blindness 10 scheme
cb_dark_blue = (0/255, 107/255, 164/255)
cb_orange = (255/255, 128/255, 14/255)

#All Degree Categories
stem_cats = ['Psychology', 'Biology', 'Math and Statistics', 'Physical Sciences', 'Computer Science', 'Engineering']
lib_arts_cats = ['Foreign Languages', 'English', 'Communications and Journalism', 'Humanities']
other_cats = ['Health Professions', 'Public Administration', 'Education', 'Agriculture']
```

## Plotting the STEM degree Gender gap

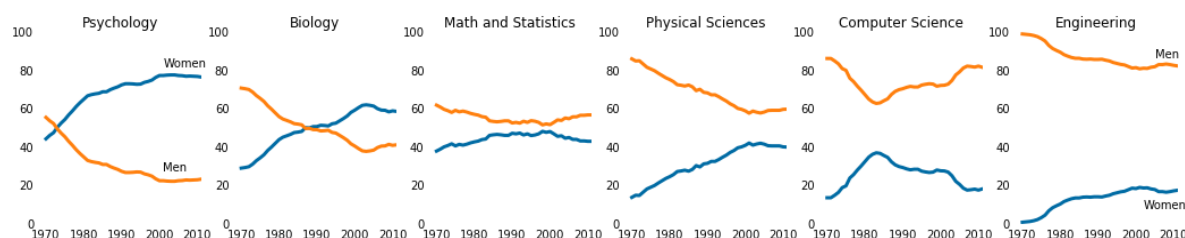
In [39]:

```
fig = plt.figure(figsize=(18, 3))

for sp in range(0,6):
    ax = fig.add_subplot(1,6,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[sp]], c=cb_dark_blue, label='Women')
    ax.plot(women_degrees['Year'], 100-women_degrees[stem_cats[sp]], c=cb_orange, label='Men')
    ax.spines["right"].set_visible(False)
    ax.spines["left"].set_visible(False)
    ax.spines["top"].set_visible(False)
    ax.spines["bottom"].set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(stem_cats[sp])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if sp == 5:
        ax.text(2005, 87, 'Men')
        ax.text(2002, 8, 'Women')
    elif sp == 0:
        ax.text(2001, 82, 'Women')
        ax.text(2001, 28, 'Men')

plt.show()
```



## Plotting the Liberal Arts degree Gender gap

In [40]:

```

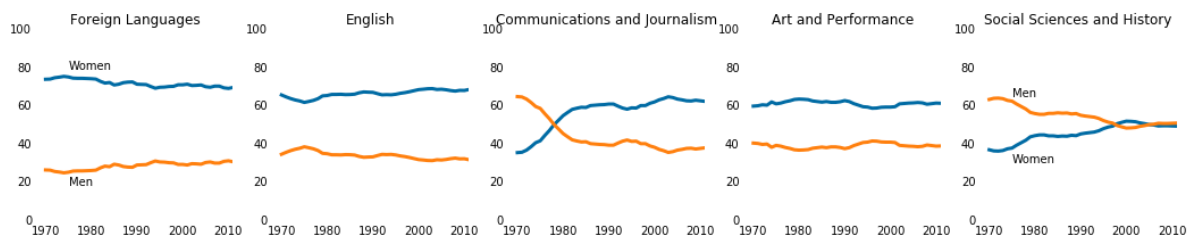
fig = plt.figure(figsize=(18, 3))

for sp in range(0,5):
    ax = fig.add_subplot(1,5,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[sp]], c=cb_dark_blue)
    ax.plot(women_degrees['Year'], 100-women_degrees[lib_arts_cats[sp]], c=cb_orange)
    ax.spines["right"].set_visible(False)
    ax.spines["left"].set_visible(False)
    ax.spines["top"].set_visible(False)
    ax.spines["bottom"].set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(lib_arts_cats[sp])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if sp == 4:
        ax.text(1975, 65, 'Men')
        ax.text(1975, 30, 'Women')
    elif sp == 0:
        ax.text(1975, 79, 'Women')
        ax.text(1975, 18, 'Men')

plt.show()

```



## Plotting the Other degrees Gender gap

In [41]:

```

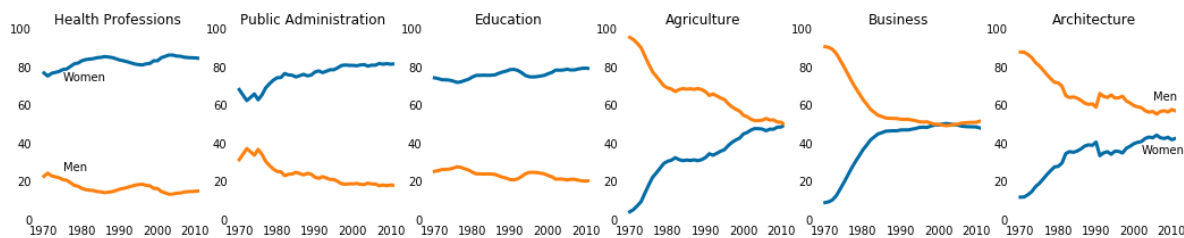
fig = plt.figure(figsize=(18, 3))

for sp in range(0,6):
    ax = fig.add_subplot(1,6,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[sp]], c=cb_dark_blue, label='Women')
    ax.plot(women_degrees['Year'], 100-women_degrees[other_cats[sp]], c=cb_orange, label='Men')
    ax.spines["right"].set_visible(False)
    ax.spines["left"].set_visible(False)
    ax.spines["top"].set_visible(False)
    ax.spines["bottom"].set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(other_cats[sp])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if sp == 5:
        ax.text(2005, 63, 'Men')
        ax.text(2002, 35, 'Women')
    elif sp == 0:
        ax.text(1975, 73, 'Women')
        ax.text(1975, 26, 'Men')

plt.show()

```



## Plotting the Gender gap for all degrees

In [42]:

```

fig = plt.figure(figsize=(16, 20))

## Generate first column of line charts. STEM degrees.
for sp in range(0,18,3):
    cat_index = int(sp/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[cat_index]], c=cb_dark_b)
    ax.plot(women_degrees['Year'], 100-women_degrees[stem_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(stem_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if cat_index == 0:
        ax.text(2003, 85, 'Women')
        ax.text(2005, 10, 'Men')
    elif cat_index == 5:
        ax.text(2005, 87, 'Men')
        ax.text(2003, 7, 'Women')

## Generate second column of line charts. Liberal arts degrees.
for sp in range(1,16,3):
    cat_index = int((sp-1)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[cat_index]], c=cb_dar)
    ax.plot(women_degrees['Year'], 100-women_degrees[lib_arts_cats[cat_index]], c=cb)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(lib_arts_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if cat_index == 0:
        ax.text(2003, 78, 'Women')
        ax.text(2005, 18, 'Men')
    elif cat_index == 4:
        ax.text(1975, 65, 'Men')
        ax.text(1975, 30, 'Women')

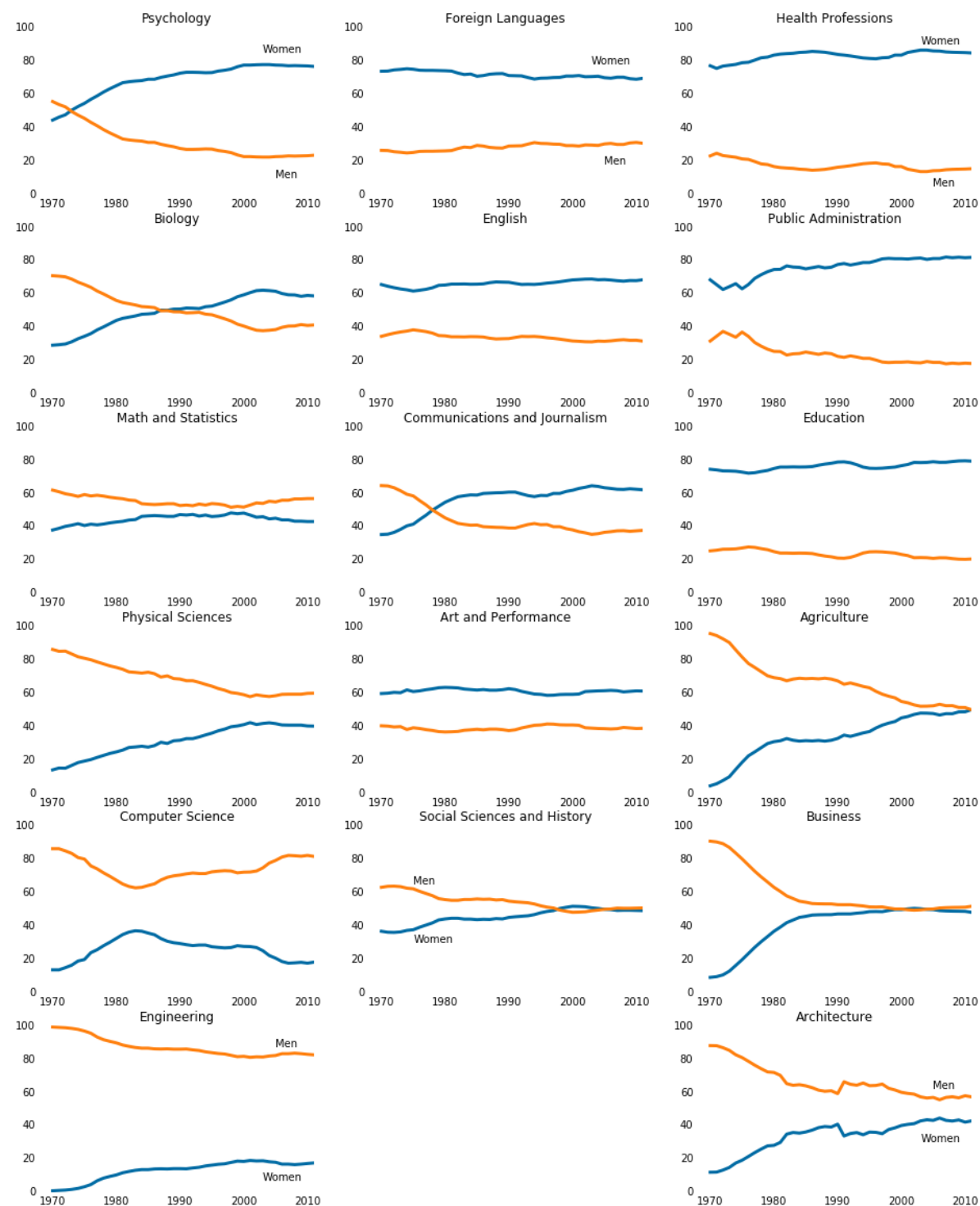
## Generate third column of line charts. Other degrees.
for sp in range(2,20,3):
    cat_index = int((sp-2)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[cat_index]], c=cb_dark_k)
    ax.plot(women_degrees['Year'], 100-women_degrees[other_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(other_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off")

    if cat_index == 0:
        ax.text(2003, 90, 'Women')
        ax.text(2005, 5, 'Men')

```

```
elif cat_index == 5:
    ax.text(2005, 62, 'Men')
    ax.text(2003, 30, 'Women')
```

```
plt.show()
```



## Removing the x-axis

In order to maximize space and avoiding overlapping text, we can remove the x-axis labels for all line charts except the bottommost line charts in each column.

In [43]:

```

fig = plt.figure(figsize=(16, 20))

## Generate first column of line charts. STEM degrees.
for sp in range(0,18,3):
    cat_index = int(sp/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[cat_index]], c=cb_dark_b)
    ax.plot(women_degrees['Year'], 100-women_degrees[stem_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(stem_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 85, 'Women')
        ax.text(2005, 10, 'Men')
    elif cat_index == 5:
        ax.text(2005, 87, 'Men')
        ax.text(2003, 7, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate second column of line charts. Liberal arts degrees.
for sp in range(1,16,3):
    cat_index = int((sp-1)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[cat_index]], c=cb_dar)
    ax.plot(women_degrees['Year'], 100-women_degrees[lib_arts_cats[cat_index]], c=cb)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(lib_arts_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 78, 'Women')
        ax.text(2005, 18, 'Men')
    elif cat_index == 4:
        ax.text(1975, 65, 'Men')
        ax.text(1975, 30, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate third column of line charts. Other degrees.
for sp in range(2,20,3):
    cat_index = int((sp-2)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[cat_index]], c=cb_dark_k)
    ax.plot(women_degrees['Year'], 100-women_degrees[other_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_title(other_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:

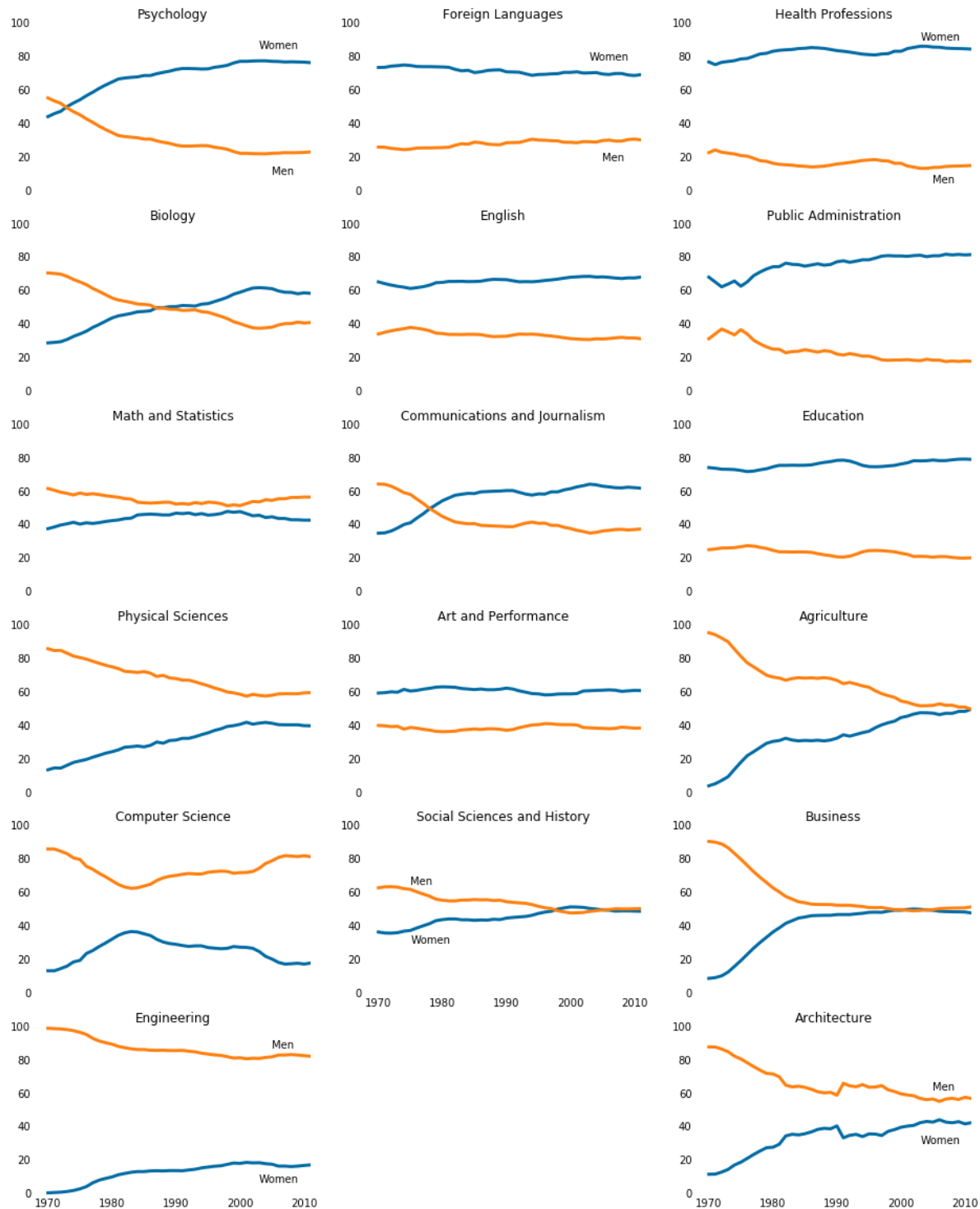
```



```

ax.text(2003, 90, 'Women')
ax.text(2005, 5, 'Men')
elif cat_index == 5:
    ax.text(2005, 62, 'Men')
    ax.text(2003, 30, 'Women')
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom="off", labeltop="off",
plt.show()

```



## **Simplifying the y axis**

We can in turn keep only the starting and ending labels (0 and 100) so as to reduce the clutter for each plot.

In [44]:

```

fig = plt.figure(figsize=(16, 20))

## Generate first column of line charts. STEM degrees.
for sp in range(0,18,3):
    cat_index = int(sp/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[cat_index]], c=cb_dark_b)
    ax.plot(women_degrees['Year'], 100-women_degrees[stem_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_yticks([0,100])
    ax.set_title(stem_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 85, 'Women')
        ax.text(2005, 10, 'Men')
    elif cat_index == 5:
        ax.text(2005, 87, 'Men')
        ax.text(2003, 7, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate second column of line charts. Liberal arts degrees.
for sp in range(1,16,3):
    cat_index = int((sp-1)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[cat_index]], c=cb_dar)
    ax.plot(women_degrees['Year'], 100-women_degrees[lib_arts_cats[cat_index]], c=cb)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_yticks([0,100])
    ax.set_title(lib_arts_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 78, 'Women')
        ax.text(2005, 18, 'Men')
    elif cat_index == 4:
        ax.text(1975, 65, 'Men')
        ax.text(1975, 30, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate third column of line charts. Other degrees.
for sp in range(2,20,3):
    cat_index = int((sp-2)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[cat_index]], c=cb_dark_k)
    ax.plot(women_degrees['Year'], 100-women_degrees[other_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_yticks([0,100])
    ax.set_title(other_cats[cat_index])

```

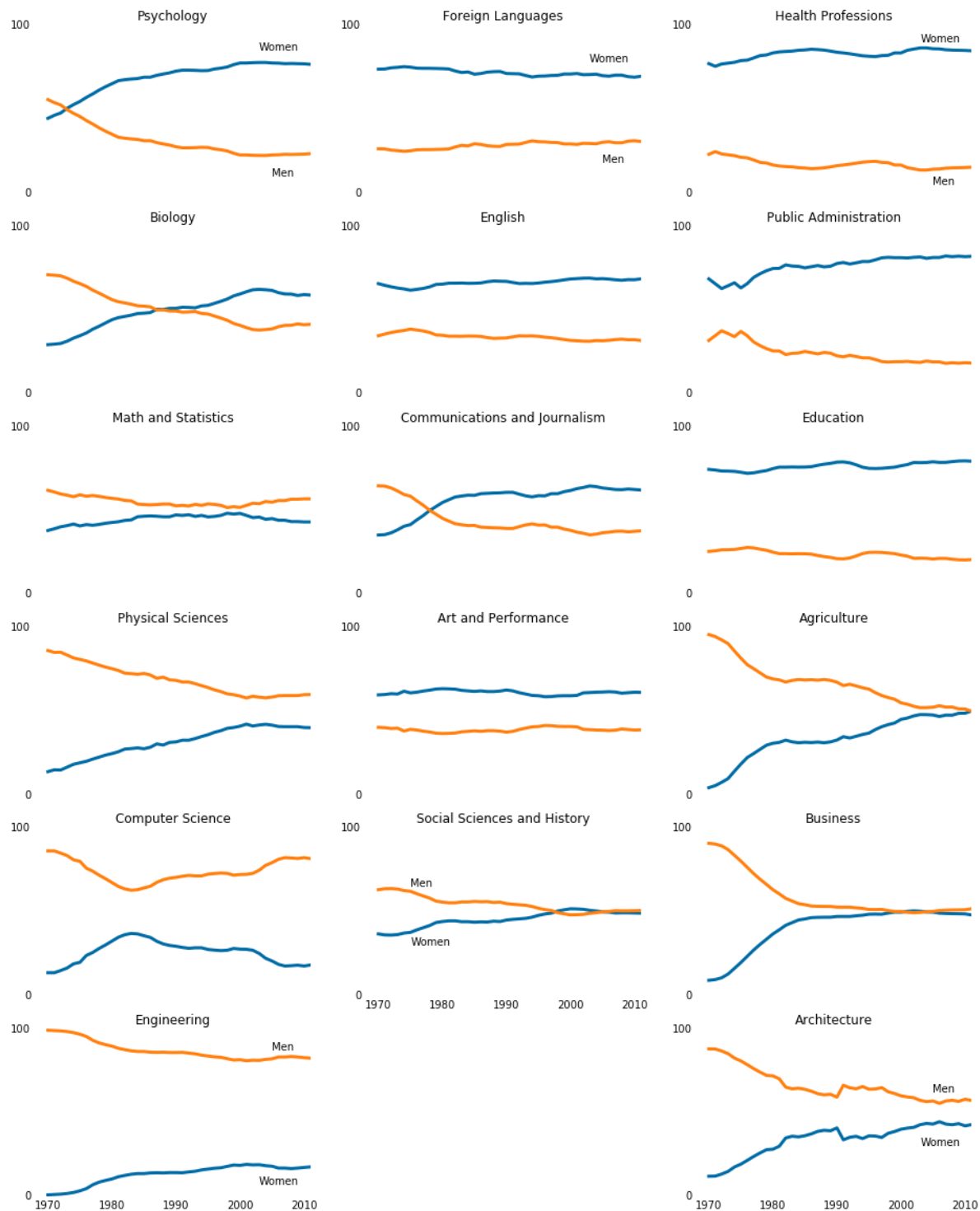
```

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

if cat_index == 0:
    ax.text(2003, 90, 'Women')
    ax.text(2005, 5, 'Men')
elif cat_index == 5:
    ax.text(2005, 62, 'Men')
    ax.text(2003, 30, 'Women')
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

plt.show()

```



## Adding a horizontal line

This horizontal line would signify zero gender gap and would act as a reference point for comparison.

In [45]:

```

fig = plt.figure(figsize=(16, 20))

## Generate first column of line charts. STEM degrees.
for sp in range(0,18,3):
    cat_index = int(sp/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[stem_cats[cat_index]], c=cb_dark_b)
    ax.plot(women_degrees['Year'], 100-women_degrees[stem_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_yticks([0,100])
    ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3)
    ax.set_title(stem_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 85, 'Women')
        ax.text(2005, 10, 'Men')
    elif cat_index == 5:
        ax.text(2005, 87, 'Men')
        ax.text(2003, 7, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate second column of line charts. Liberal arts degrees.
for sp in range(1,16,3):
    cat_index = int((sp-1)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[lib_arts_cats[cat_index]], c=cb_dar)
    ax.plot(women_degrees['Year'], 100-women_degrees[lib_arts_cats[cat_index]], c=cb)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)
    ax.set_yticks([0,100])
    ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3)
    ax.set_title(lib_arts_cats[cat_index])
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

    if cat_index == 0:
        ax.text(2003, 78, 'Women')
        ax.text(2005, 18, 'Men')
    elif cat_index == 4:
        ax.text(1975, 65, 'Men')
        ax.text(1975, 30, 'Women')
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

## Generate third column of line charts. Other degrees.
for sp in range(2,20,3):
    cat_index = int((sp-2)/3)
    ax = fig.add_subplot(6,3,sp+1)
    ax.plot(women_degrees['Year'], women_degrees[other_cats[cat_index]], c=cb_dark_k)
    ax.plot(women_degrees['Year'], 100-women_degrees[other_cats[cat_index]], c=cb_or)
    for key,spine in ax.spines.items():
        spine.set_visible(False)
    ax.set_xlim(1968, 2011)
    ax.set_ylim(0,100)

```

```

ax.set_yticks([0,100])
ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3)
ax.set_title(other_cats[cat_index])
ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

if cat_index == 0:
    ax.text(2003, 90, 'Women')
    ax.text(2005, 5, 'Men')
elif cat_index == 5:
    ax.text(2005, 62, 'Men')
    ax.text(2003, 30, 'Women')
    ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom

plt.savefig('plots.png')

plt.show()

```

