Visualizing The Gender Gap in College Degrees

May 11, 2018

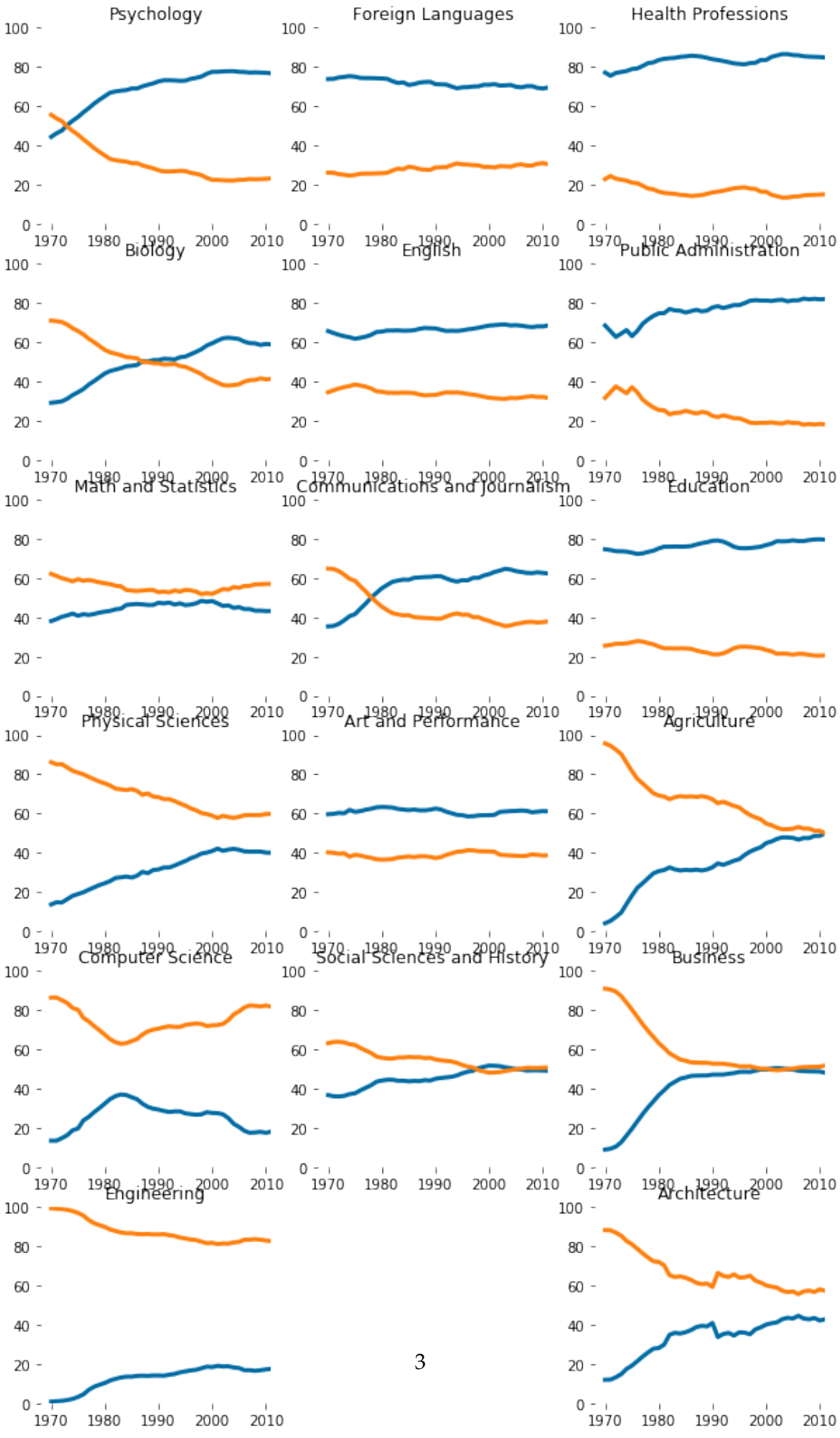
0.1 Introduction

In [15]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

women_degrees = pd.read_csv('.\databank\percent-bachelors-degrees-women-usa.csv')
cb_dark_blue = (0/255,107/255,164/255)
cb_orange = (255/255, 128/255, 14/255)
```

0.1.1 Visualize Gender Participation in Tabular Chart for Major Degree

In [16]:
```
stem_cats = ['Psychology', 'Biology', 'Math and Statistics', 'Physical Sciences', 'Co
lib_arts_cats = ['Foreign Languages', 'English', 'Communications and Journalism', 'Ar
other_cats = ['Health Professions', 'Public Administration', 'Education', 'Agricultur

# stem_count positions: 1,4,7,10,13,16
# lib_art_count positions: 2,5,8,11,14
# other_count positions 3,6,9,12,15,18

fig = plt.figure(figsize=(10, 18))
lbl = 'on'

# Generaing the plots for all three degree Series
for count in range(0,3):
    for sp in range(0,6):
        if (count == 0):
            ax = fig.add_subplot(6,3,3*sp+1)
            item = stem_cats
        elif ((count == 1) & (sp == 5)):
            continue
        elif ((count == 1) & (sp < 5)):
            ax = fig.add_subplot(6,3, 3*sp+2)
            item = lib_arts_cats
        elif (count == 2):
            ax = fig.add_subplot(6,3,3*sp+3)
            item = other_cats
        ax.plot(women_degrees['Year'], women_degrees[item[sp]], c=cb_dark_blue, label=
```

```
            ax.plot(women_degrees['Year'], 100-women_degrees[item[sp]], c=cb_orange, label
            ax.set_xlim(1968, 2011)
            ax.set_ylim(0,100)
            ax.set_title(item[sp])
            for key, spine in ax.spines.items():
                spine.set_visible(False)

plt.show()
```

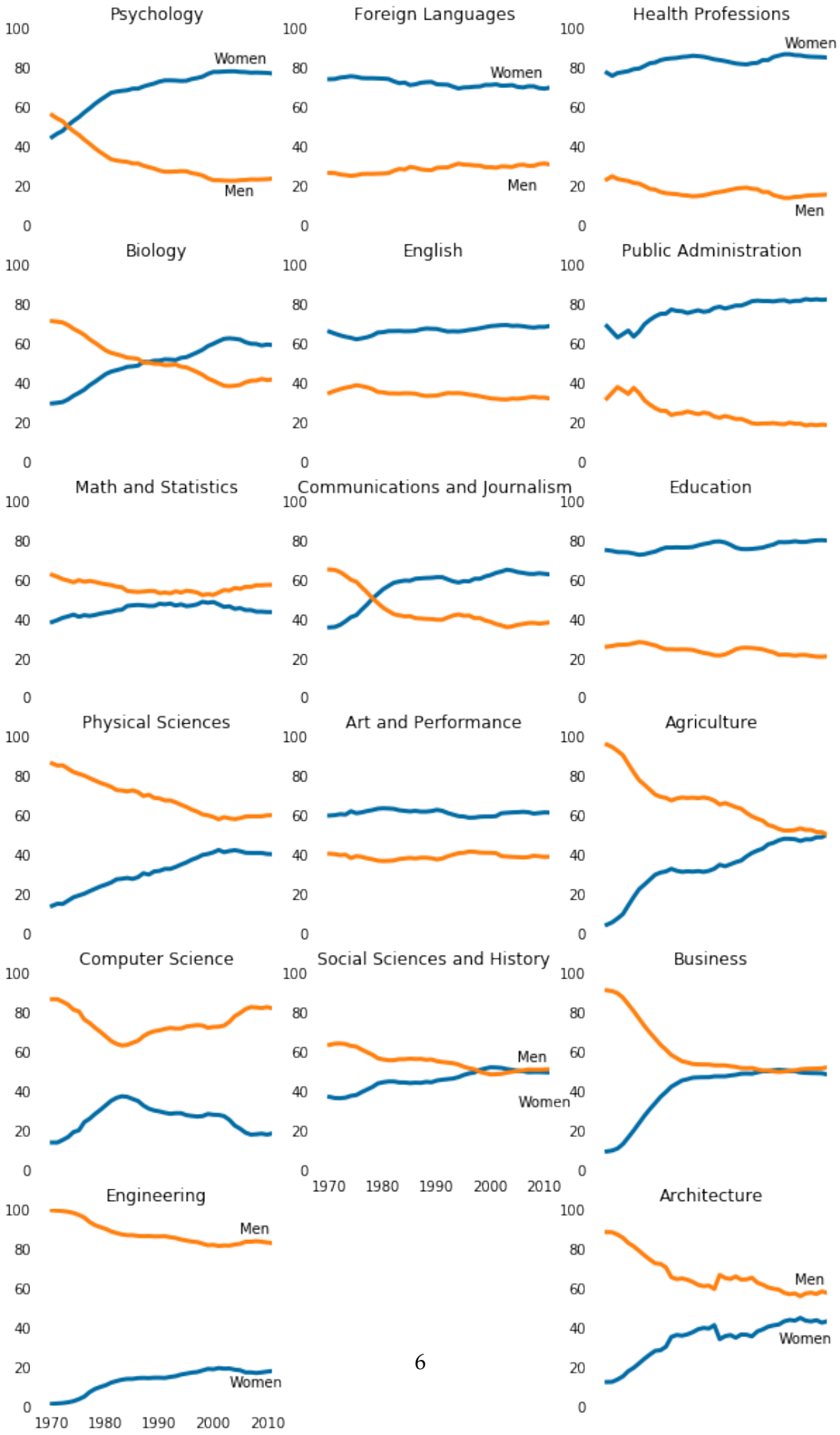

Psychology
Foreign Languages
Health Professions
Biology
English
Public Administration
Math and Statistics
Communications and Journalism
Education
Physical Sciences
Art and Performance
Agriculture
Computer Science
Social Sciences and History
Business
Engineering
Architecture
0
20
40
60
80
100
1970
1980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

0.2 Hiding X-axis Labels Except the Bottommost Line

In [17]:
```
fig = plt.figure(figsize=(10, 18))
lbl = 'on'

# Generaing the plots for all three degree Series
for count in range(0,3):
    for sp in range(0,6):
        if (count == 0):
            ax = fig.add_subplot(6,3,3*sp+1)
            item = stem_cats
        elif ((count == 1) & (sp == 5)):
            continue
        elif ((count == 1) & (sp < 5)):
            ax = fig.add_subplot(6,3, 3*sp+2)
            item = lib_arts_cats
        elif (count == 2):
            ax = fig.add_subplot(6,3,3*sp+3)
            item = other_cats
        ax.plot(women_degrees['Year'], women_degrees[item[sp]], c=cb_dark_blue, label=
        ax.plot(women_degrees['Year'], 100-women_degrees[item[sp]], c=cb_orange, label
        ax.set_xlim(1968, 2011)
        ax.set_ylim(0,100)
        ax.set_title(item[sp])
        for key, spine in ax.spines.items():
            spine.set_visible(False)

# Cut off the x-axis lables for all subplots but the bottom most one
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

# Keeping the x-axis 'Year' label for bottom most plots only
        if (count == 0):
            if (sp == 0):
                ax.text(2000, 82, 'Women')
                ax.text(2002, 15, 'Men')
            if (sp == 5):
                ax.text(2005, 88, 'Men')
                ax.text(2003, 10, 'Women')
                ax.tick_params(labelbottom = lbl)
        elif (count == 1):
            if (sp == 0):
                ax.text(2000, 75, 'Women')
                ax.text(2003, 18, 'Men')
            if (sp == 4):
                ax.text(2005, 55, 'Men')
```

```
                    ax.text(2005, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)
            elif (count == 2):
                if (sp == 0):
                    ax.text(2003, 90, 'Women')
                    ax.text(2005, 5, 'Men')
                if (sp == 5):
                    ax.text(2005, 62, 'Men')
                    ax.text(2002, 32, 'Women')

    plt.show()
```

```
C:\Users\Yogi_Ashwast\Anaconda3\lib\site-packages\matplotlib\cbook\deprecation.py:107: Matplotl
  warnings.warn(message, mplDeprecation, stacklevel=1)
```

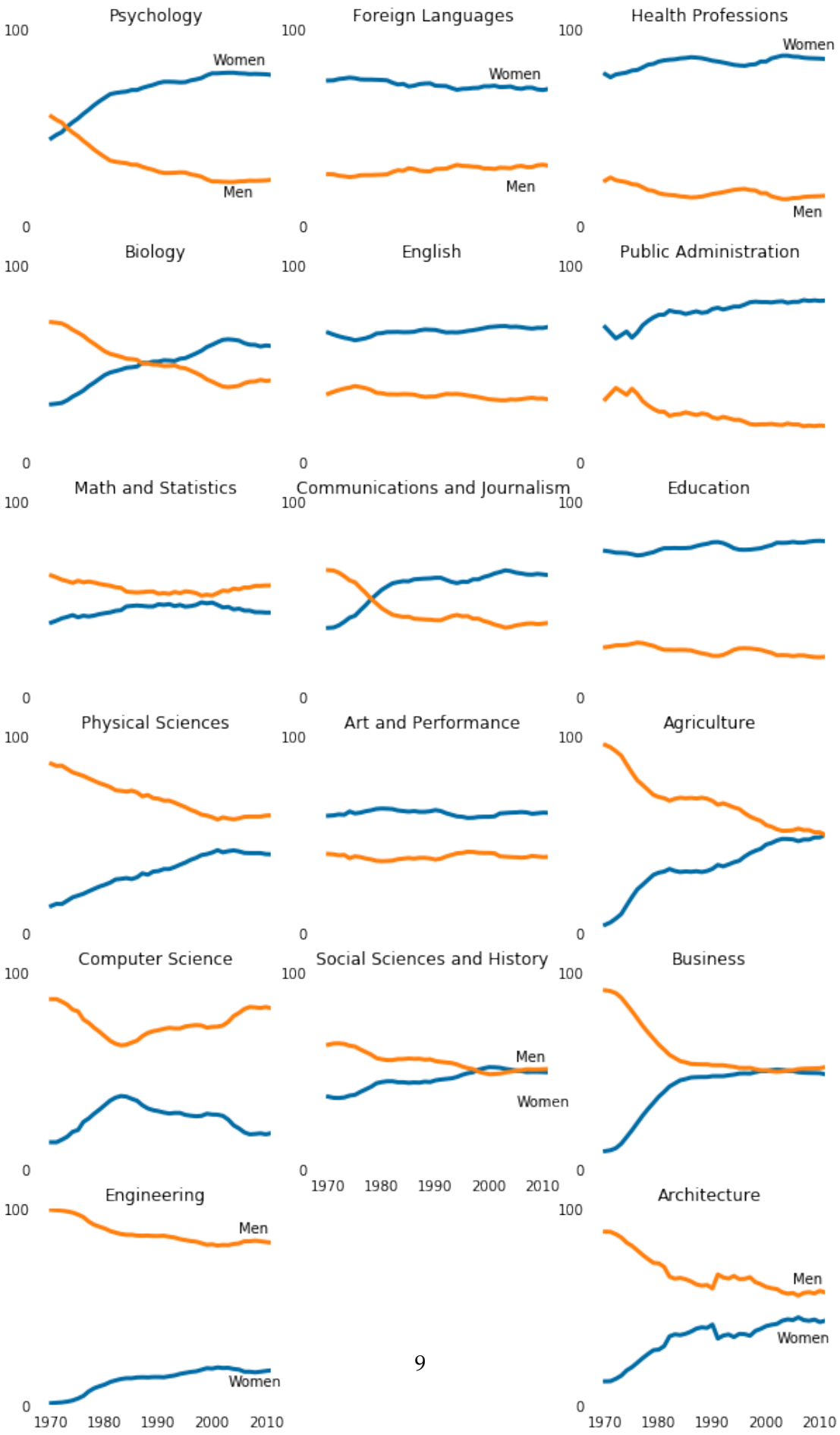

Psychology
Women
Men
Foreign Languages
Women
Men
Health Professions
Women
Men
Biology
English
Public Administration
Math and Statistics
Communications and Journalism
Education
Physical Sciences
Art and Performance
Agriculture
Computer Science
Social Sciences and History
Men
Women
Business
Engineering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

0.3 Keeping Only Starting and Ending Y-Axis Labels

In [18]:
```
fig = plt.figure(figsize=(10, 18))
lbl = 'on'

# Generaing the plots for all three degree Series
for count in range(0,3):
    for sp in range(0,6):
        if (count == 0):
            ax = fig.add_subplot(6,3,3*sp+1)
            item = stem_cats
        elif ((count == 1) & (sp == 5)):
            continue
        elif ((count == 1) & (sp < 5)):
            ax = fig.add_subplot(6,3, 3*sp+2)
            item = lib_arts_cats
        elif (count == 2):
            ax = fig.add_subplot(6,3,3*sp+3)
            item = other_cats
        ax.plot(women_degrees['Year'], women_degrees[item[sp]], c=cb_dark_blue, label=
        ax.plot(women_degrees['Year'], 100-women_degrees[item[sp]], c=cb_orange, label
        ax.set_xlim(1968, 2011)
        ax.set_ylim(0,100)
        ax.set_title(item[sp])
        for key, spine in ax.spines.items():
            spine.set_visible(False)

# Cut off the x-axis lables for all subplots but the bottom most one
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

# Keeping the x-axis 'Year' label for bottom most plots only
        if (count == 0):
            if (sp == 0):
                ax.text(2000, 82, 'Women')
                ax.text(2002, 15, 'Men')
            if (sp == 5):
                ax.text(2005, 88, 'Men')
                ax.text(2003, 10, 'Women')
                ax.tick_params(labelbottom = lbl)
        elif (count == 1):
            if (sp == 0):
                ax.text(2000, 75, 'Women')
                ax.text(2003, 18, 'Men')
            if (sp == 4):
                ax.text(2005, 55, 'Men')
```

```
                ax.text(2005, 32, 'Women')
                ax.tick_params(labelbottom = lbl)
            elif (count == 2):
                if (sp == 0):
                    ax.text(2003, 90, 'Women')
                    ax.text(2005, 5, 'Men')
                if (sp == 5):
                    ax.text(2005, 62, 'Men')
                    ax.text(2002, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)

        # Keeping just the starting and ending labels for the y-axis
                ax.set_yticks([0,100])

        plt.show()
```

```
C:\Users\Yogi_Ashwast\Anaconda3\lib\site-packages\matplotlib\cbook\deprecation.py:107: Matplotl
  warnings.warn(message, mplDeprecation, stacklevel=1)
```

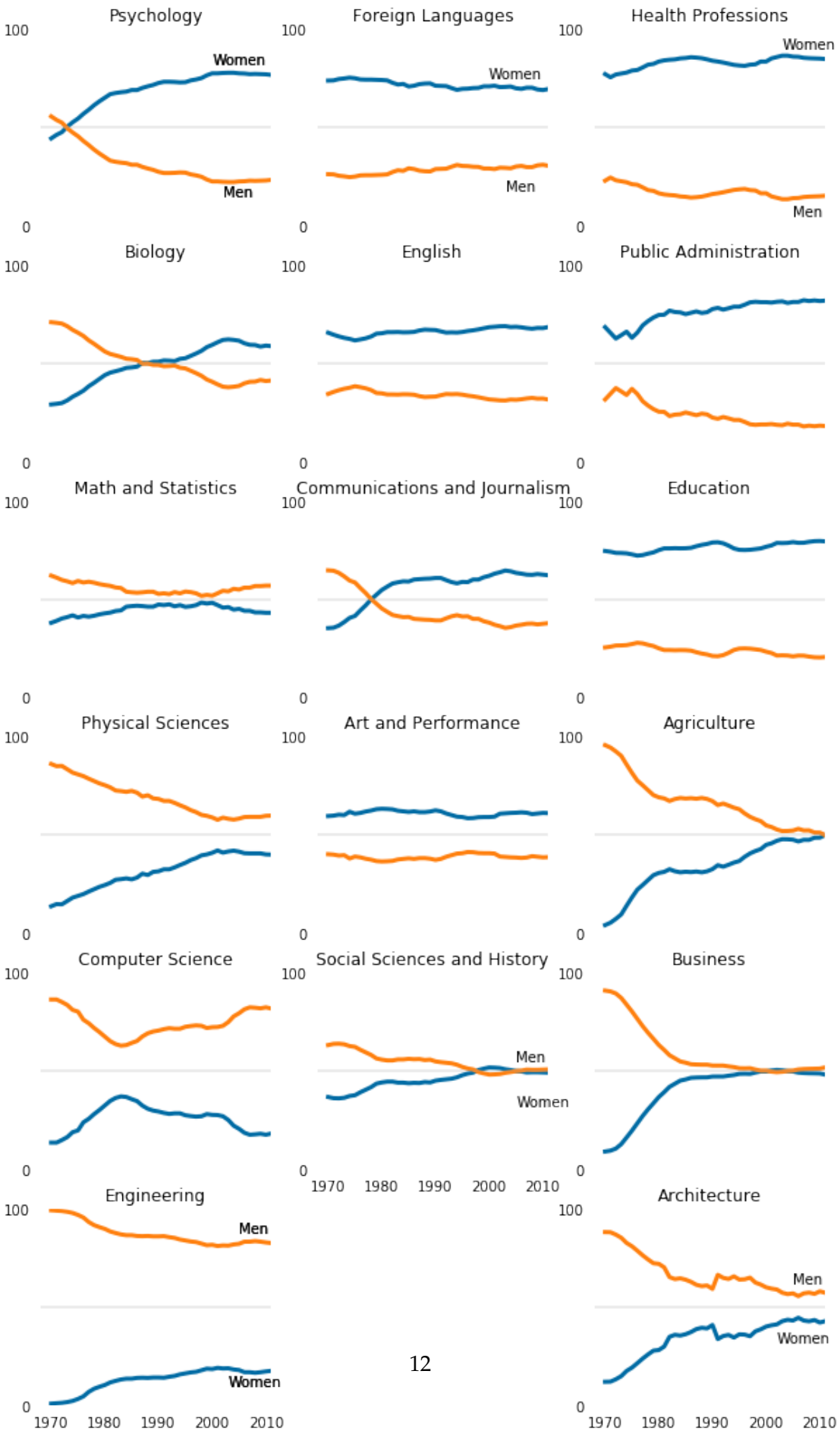

Psychology
Women
Men
Foreign Languages
Women
Men
Health Professions
Women
Men
Biology
English
Public Administration
Math and Statistics
Communications and Journalism
Education
Physical Sciences
Art and Performance
Agriculture
Computer Science
Social Sciences and History
Men
Women
Business
Engineering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

0.4 Insert Horizontal Line At 50% Gender Population

In [19]:
```
fig = plt.figure(figsize=(10, 18))
lbl = 'on'

# Generaing the plots for all three degree Series
for count in range(0,3):
    for sp in range(0,6):
        if (count == 0):
            ax = fig.add_subplot(6,3,3*sp+1)
            item = stem_cats
        elif ((count == 1) & (sp == 5)):
            continue
        elif ((count == 1) & (sp < 5)):
            ax = fig.add_subplot(6,3, 3*sp+2)
            item = lib_arts_cats
        elif (count == 2):
            ax = fig.add_subplot(6,3,3*sp+3)
            item = other_cats
        ax.plot(women_degrees['Year'], women_degrees[item[sp]], c=cb_dark_blue, label=
        ax.plot(women_degrees['Year'], 100-women_degrees[item[sp]], c=cb_orange, label
        ax.set_xlim(1968, 2011)
        ax.set_ylim(0,100)
        ax.set_title(item[sp])
        for key, spine in ax.spines.items():
            spine.set_visible(False)

# Cut off the x-axis lables for all subplots but the bottom most one
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

# Keeping the x-axis 'Year' label for bottom most plots only
        if (count == 0):
            if (sp == 0):
                ax.text(2000, 82, 'Women')
                ax.text(2002, 15, 'Men')
            if (sp == 5):
                ax.text(2005, 88, 'Men')
                ax.text(2003, 10, 'Women')
                ax.tick_params(labelbottom = lbl)
        elif (count == 1):
            if (sp == 0):
                ax.text(2000, 75, 'Women')
                ax.text(2003, 18, 'Men')
            if (sp == 4):
                ax.text(2005, 55, 'Men')
```

```
                    ax.text(2005, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)
            elif (count == 2):
                if (sp == 0):
                    ax.text(2003, 90, 'Women')
                    ax.text(2005, 5, 'Men')
                if (sp == 5):
                    ax.text(2005, 62, 'Men')
                    ax.text(2002, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)

        # Keeping just the starting and ending labels for the y-axis
            ax.set_yticks([0,100])

        # Inserting horizontal line at 50% of gender population for better clarity
            ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3)
            if (count == 0):
                if (sp == 0):
                    ax.text(2000, 82, 'Women')
                    ax.text(2002, 15, 'Men')
                if (sp == 5):
                    ax.text(2005, 88, 'Men')
                    ax.text(2003, 10, 'Women')

        plt.show()
```

```
C:\Users\Yogi_Ashwast\Anaconda3\lib\site-packages\matplotlib\cbook\deprecation.py:107: Matplotl
  warnings.warn(message, mplDeprecation, stacklevel=1)
```

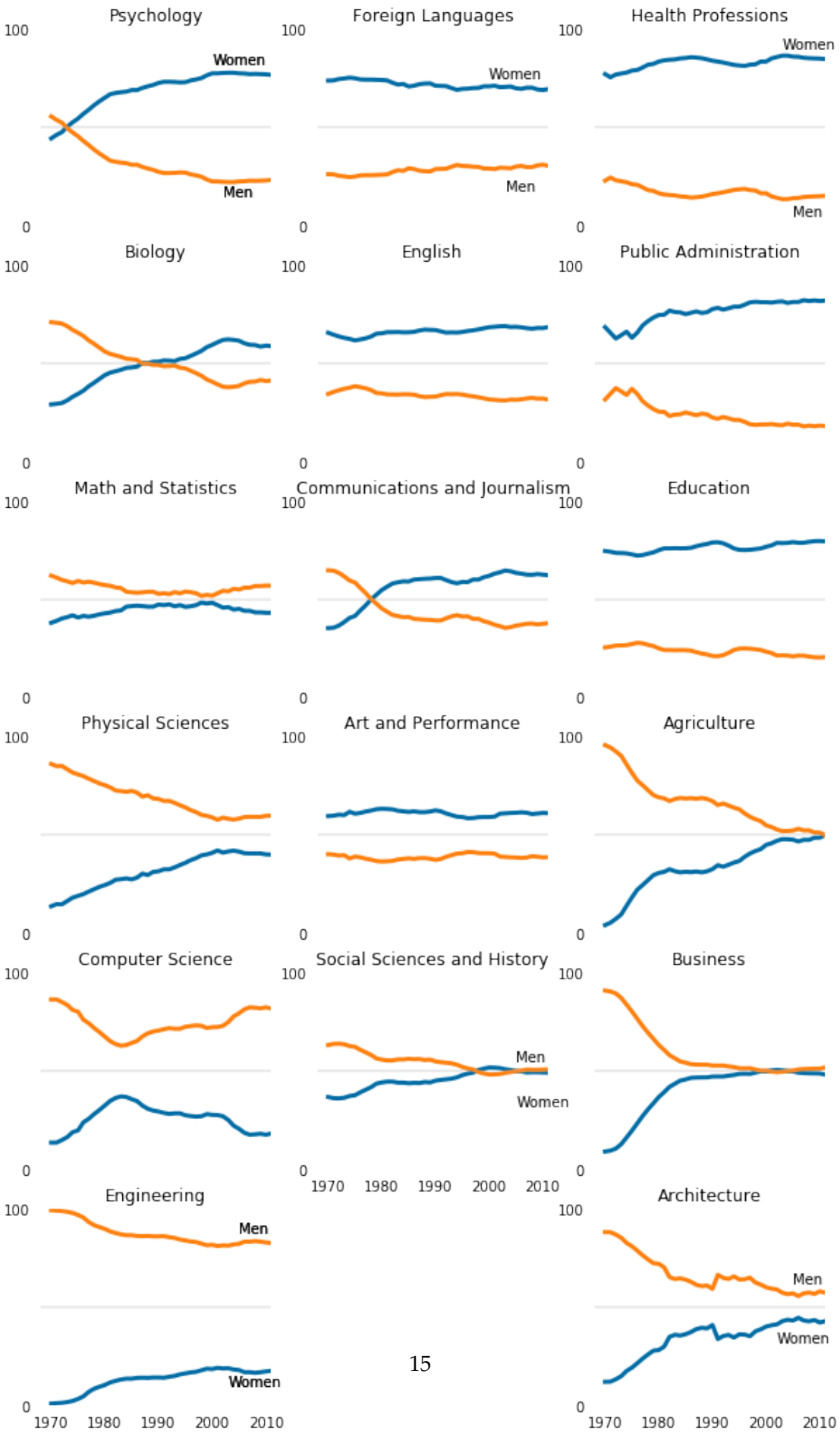

Psychology
Women
Men
Foreign Languages
Women
Men
Health Professions
Women
Men
Biology
English
Public Administration
Math and Statistics
Communications and Journalism
Education
Physical Sciences
Art and Performance
Agriculture
Computer Science
Social Sciences and History
Men
Women
Business
Engineering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

0.5 Export the Tablar Chart To a File

In [20]:
```
fig = plt.figure(figsize=(10, 18))
lbl = 'on'

# Generaing the plots for all three degree Series
for count in range(0,3):
    for sp in range(0,6):
        if (count == 0):
            ax = fig.add_subplot(6,3,3*sp+1)
            item = stem_cats
        elif ((count == 1) & (sp == 5)):
            continue
        elif ((count == 1) & (sp < 5)):
            ax = fig.add_subplot(6,3, 3*sp+2)
            item = lib_arts_cats
        elif (count == 2):
            ax = fig.add_subplot(6,3,3*sp+3)
            item = other_cats
        ax.plot(women_degrees['Year'], women_degrees[item[sp]], c=cb_dark_blue, label=
        ax.plot(women_degrees['Year'], 100-women_degrees[item[sp]], c=cb_orange, label
        ax.set_xlim(1968, 2011)
        ax.set_ylim(0,100)
        ax.set_title(item[sp])
        for key, spine in ax.spines.items():
            spine.set_visible(False)

# Cut off the x-axis lables for all subplots but the bottom most one
        ax.tick_params(bottom="off", top="off", left="off", right="off", labelbottom =

# Keeping the x-axis 'Year' label for bottom most plots only
        if (count == 0):
            if (sp == 0):
                ax.text(2000, 82, 'Women')
                ax.text(2002, 15, 'Men')
            if (sp == 5):
                ax.text(2005, 88, 'Men')
                ax.text(2003, 10, 'Women')
                ax.tick_params(labelbottom = lbl)
        elif (count == 1):
            if (sp == 0):
                ax.text(2000, 75, 'Women')
                ax.text(2003, 18, 'Men')
            if (sp == 4):
                ax.text(2005, 55, 'Men')
```

```
                    ax.text(2005, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)
            elif (count == 2):
                if (sp == 0):
                    ax.text(2003, 90, 'Women')
                    ax.text(2005, 5, 'Men')
                if (sp == 5):
                    ax.text(2005, 62, 'Men')
                    ax.text(2002, 32, 'Women')
                    ax.tick_params(labelbottom = lbl)

        # Keeping just the starting and ending labels for the y-axis
            ax.set_yticks([0,100])

        # Inserting horizontal line at 50% of gender population for better clarity
            ax.axhline(50, c=(171/255, 171/255, 171/255), alpha=0.3)
            if (count == 0):
                if (sp == 0):
                    ax.text(2000, 82, 'Women')
                    ax.text(2002, 15, 'Men')
                if (sp == 5):
                    ax.text(2005, 88, 'Men')
                    ax.text(2003, 10, 'Women')

        # Exporting the plots from the Jupyter Notbook
        fig.savefig("gender_degrees.png")

        plt.show()
```

```
C:\Users\Yogi_Ashwast\Anaconda3\lib\site-packages\matplotlib\cbook\deprecation.py:107: Matplotl
  warnings.warn(message, mplDeprecation, stacklevel=1)
```


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
1980
1990
2000
2010