Instructions (please follow to get full credit):

- **After completing the assignment, please submit your .ipnyb file to NYU Classes with the following naming convention: Lastname_Firstname_NetID_ProblemSet# (ex. Smith_John_js123_ProblemSet2)**
- Submit your answers in a **Jupyter notebook** with **proper markdowns to indicate problem numbers.**
- **Write the questions in markdown before you provide your answers.**
- **When copying the dictionary or any values directly from this file, make sure that all the quotations and brackets are in the right form in Jupyter Notebook. (Especially for string quotations – sometimes if you copy directly from a pdf file, the quotation breaks and it won't show up properly as a string in Jupyter Notebook)**
- See Grading Guidelines under Announcements on NYU Classes.
- Tuning the parameters of the axes object is key in this problem set
- All plots should be displayed with a title, legend, suitably named x and y labels, properly distributed x and y ticks (carries credit)
- Plot should be of medium to large size so that the visualization doesn't look overcrowded. Very small plots are unreadable (carries credit)
- Position any legends neatly in or outside the plot without overlapping with plot lines (carries credit)
- For every sub-question on plotting, show the plot (or whatever output) before going on to the next sub-question
- In-line comments are preferred for this assignment but not mandatory
- No explanations are expected at the end of answers, unless requested

Problems:

1. Consider the data at:

   - url1 = 'https://raw.githubusercontent.com/fivethirtyeight/data/master/'

   - url2 = 'college-majors/recent-grads.csv'

   - url = url1 + url2

   The variables are described at:

   https://github.com/fivethirtyeight/data/tree/master/college-majors

   a) Read the csv file at URL into a data frame named df538. What are its dimensions?
   b) Construct horizontal bar charts of the top ten majors sorted by median salary
   c) Plot the bar charts again this time sorted by the salary of the 25th percentile.
      In each case plot just the variable you sorted on.

2. Using the csv file at location http://pages.stern.nyu.edu/~dbackus/Data/debt.csv, create a data frame called debt and do the following with Year as index:

   a) Plot each country's debt against Year using a suitable plot method. Change the line width to 2.

b) First create figure and axis objects (empty) with plt.subplots(). Then graph public indebtness over the years using our debt data and the axis object we just created. Color the lines red, blue and green. By calling a method on the axis object, change the lower limit on the y axis to zero.

c) Make the line for Argentina thicker than the others. Hint: Do this by plotting a separate line applied to the same axis object.

3. data = {'Size': ['1 to 4', '5 to 9', '10 to 19', '20 to 49', '50 to 99', '100 to 249', '250 to 499', '500 to 999', '1000 to 2499', '2500 to 4999', '5000 to 9999', '10000+'],
    'Firms': [2846416, 1020772, 598153, 373345, 115544, 63845, 19389, 9588, 6088, 2287, 1250, 1357],
    'Emp': [5998912, 6714924, 8151891, 11425545, 8055535, 9788341, 6611734, 6340775, 8321486, 6738218, 6559020, 32556671]}
    bds = pd.DataFrame(data)
    bds = bds.set_index('Size')

a) Create a bar chart for the number of employees in each size category.

b) Create figure and axis objects. Add a horizontal bar chart of the number of firms in each category to the axis object you created. Color the bars red. Use the style 'seaborn-pastel'.

4. data = {'BRA': [13.37, 13.30, 14.34, 15.07, 15.46, 15.98, 16.10],
    'JPN': [33.43, 31.83, 33.71, 34.29, 35.60, 36.79, 37.39],
    'USA': [48.30, 46.91, 48.31, 49.72, 51.41, 52.94, 54.60],
    'Year': [2008, 2009, 2010, 2011, 2012, 2013, 2014]}
    weo = pd.DataFrame(data)
    These numbers are GDP per person in thousands of US dollars, 2008 to 2014, variable PPPPC in the IMF's World Economic Outlook Database.

a) For every year, plot the GDP of each country in a stacked bar chart. Use 'red' for Brazil, 'blue' for USA, 'yellow' for Japan.

b) Change the background color of the plot, make the title larger and in bold face.

c) Reverse the axes, i.e. let year be y axis and GDP be x. Apply a different color for x and y ticks. Place the legend outside the plot.

5. Using the given csv file('company.csv'), complete the following tasks.

a) Create figure and axis objects. For this, you will want two side-by-side plots with 1 row and 2 columns.

b) On the left chart, plot the number of employees at each company over time. In the right chart, plot the percent female over time.

c) Change the size of the plot and create titles for each plot to make it look presentable.

d) Change the style to 'fivethirtyeight'.