

# Homework Assignment #5

CS 3753/5163

## Individual work

Submit one Jupyter notebook named as yourLastName\_HW05 with appropriate code cell to solve the following questions. Make sure that each question is included in your notebook as a markdown cell above your answer. You must use only the basic Python and no external module or library is allowed other than math. Include any specific direction/instruction to run your script in comments. The plotting questions to be included in your Jupyter notebook.

**Note:** if you want to include your script as separate .py files, you may submit all source files with a Jupyter notebook in a zipped file (compressed). In the notebook, inside the code cell of a question, write %run Qx. Where x is the number of the question.

**Undergraduate Students:** Do Q1 only (100 points).

**Graduate Students:** Do Q1 (50 points) and Q2 (50 points).

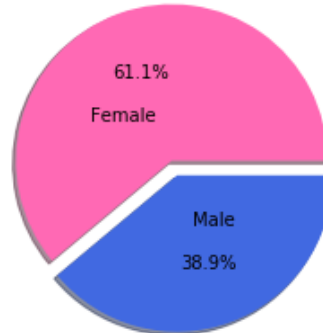
**Bonus:** (10 points)

**Due date:** 04/13/2019 at 11:59PM

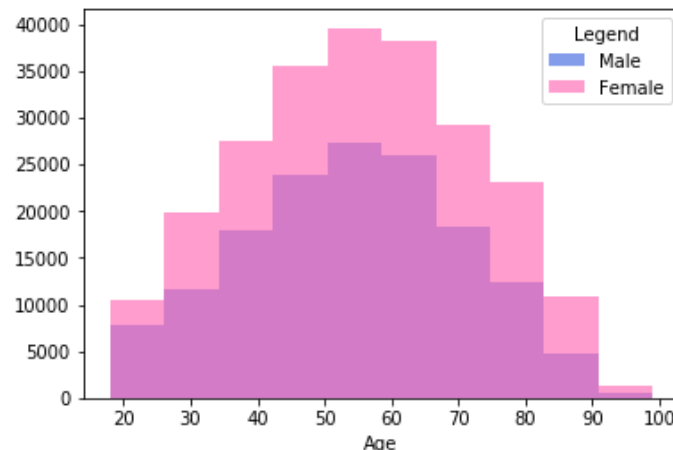
**Q1)** From HW4Q2, use numpy and matplotlib.pyplot to reproduce the following charts. The charts are for data after cleaning and conversion of variables (to Kilogram and Centimeters).

- a) (10 points/5 points) Pie chart for the number of males and females participating in the study. Colors used are hotpink and royalblue. Use parameter labeldistance=0.3 to move the label inside.

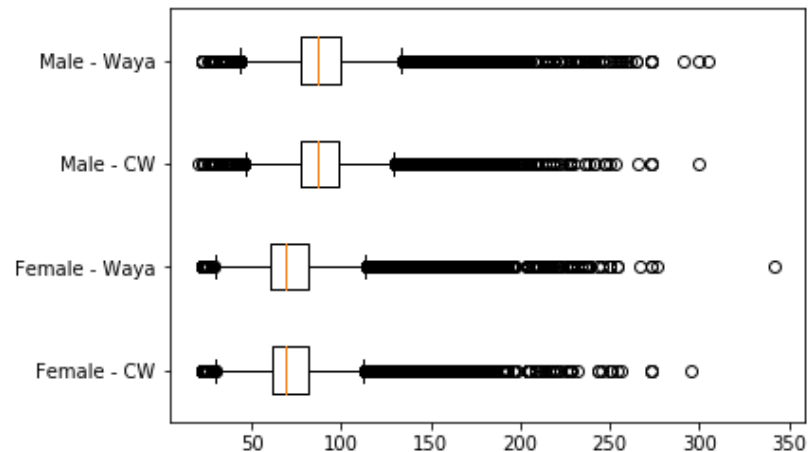
CDC BRFSS 2008



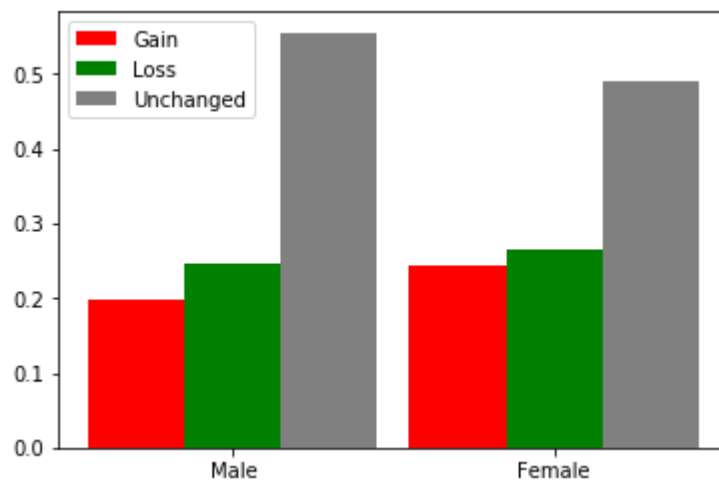
- b) (10 points/5 points) Histograms for age distribution.



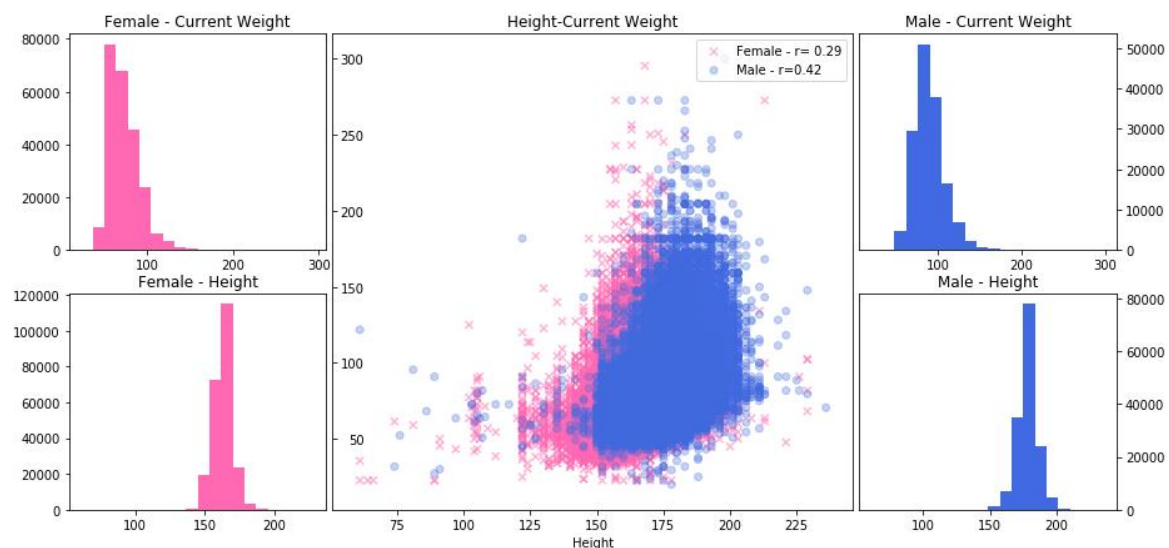
c) (10 points/ 5 points) Box plots for CW and WAYA.



d) (25 points/ 12.5 points) Bar chart to show the percentage of people who gained, lost, or had unchanged weight from last year to current year.



e) (45 points/ 22.5 points) A grid consists of height and current weight distributions and the scatter chart. To move yticks to the right, use `ax.yaxis.tick_right()` and to move the ticks inside the plot, use `ax.tick_params(axis="y",direction="in", pad=-25)`.

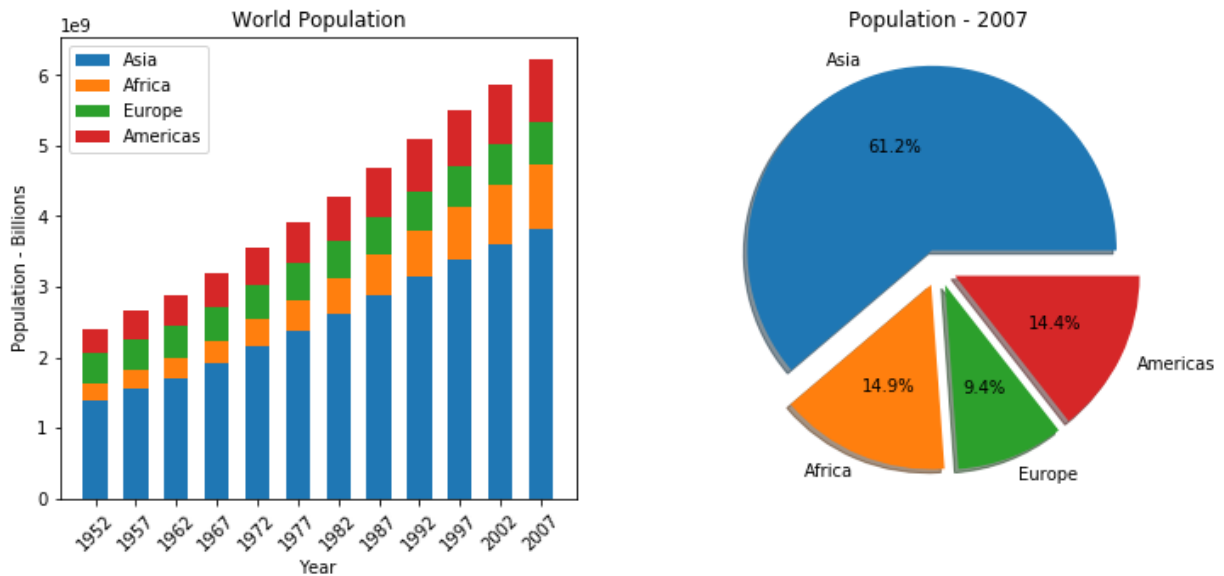


### Graduate Students Only

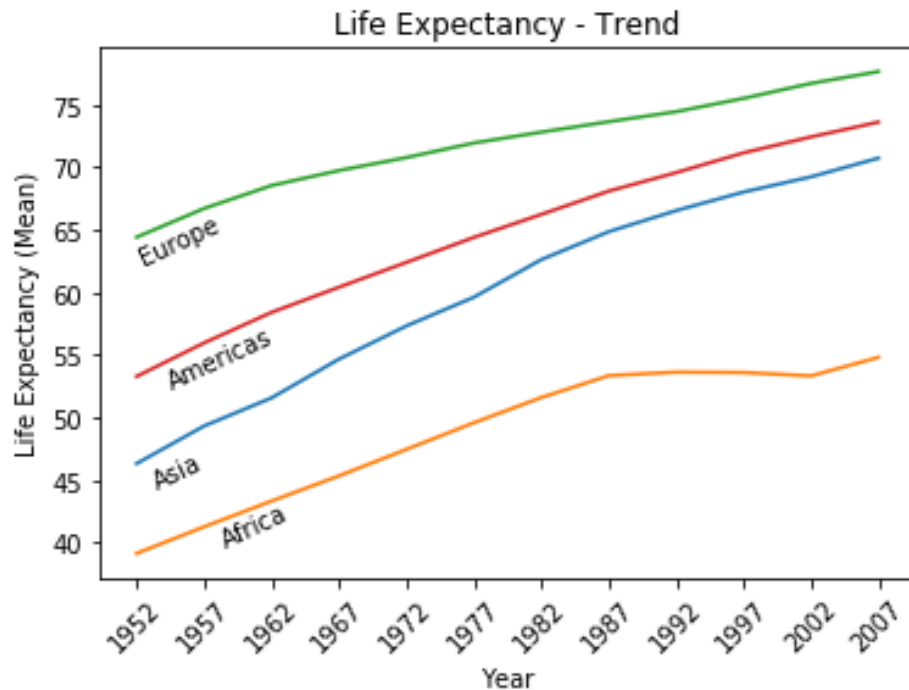
Use the dataset provided (gapminder\_lifeExp\_gdp) to reproduce the following plots. Use numpy and matplotlib.pyplot. First, remove all entries for Oceania. You may use the following structured data type:  
dt = {'names':('country', 'year', 'pop', 'continent', 'lifeExp', 'gdp'), 'formats':('U25',np.int16, np.float, 'U10', np.float, np.float)}

#### Q2) (50 points)

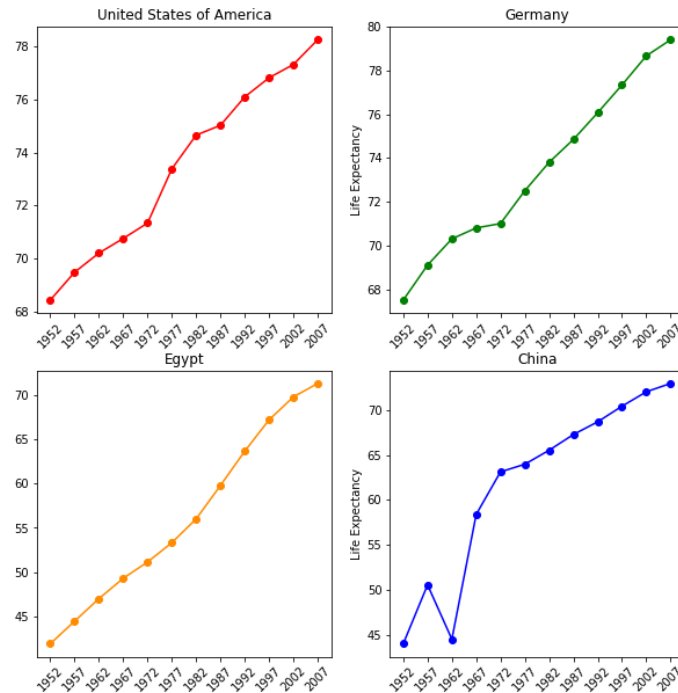
- a) (20 points) Stacked bar chart and a pie chart (as subplots) for world population.



- b) (15 points) Mean of life expectancy over years in each continent



c) **(15 points)** The trend of life expectancy over years for four countries (USA, Germany, Egypt, and China).



**Bonus for UG and G**

**(10 points)** The grid of scatter plots show the correlation between gdp per capita and life expectancy. To move the label of y axis to right use `ax.yaxis.set_label_position("right")`. Set the scale of x axis for each subplot to log using `ax.set_xscale('log')`.

