Python Worksheet #4

Matplotlib graphics, Python Diffusion of Innovation application, LaTex final paper prep

Dr. Scott

CSI 500

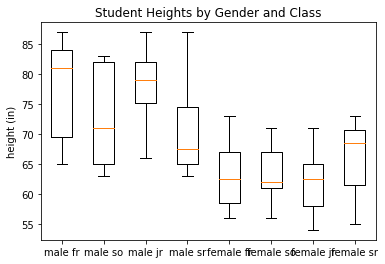
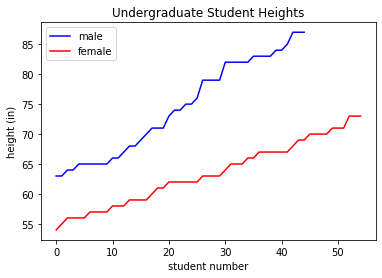
Spring 2020

# Problem #1: Matplotlib graphics

In the lectures, we discussed how to read and write data using files. Use your Python skills to accomplish the following.

## 1.1 Practice with graphics

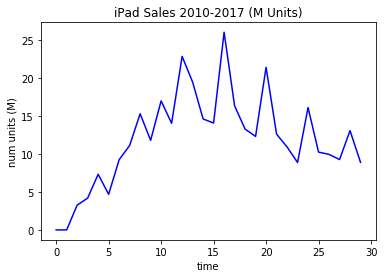
* Using the 'student\_data.csv' file we discussed in the lectures, create a 2-D X-Y graph for student heights by gender. Sort the data for each gender in ascending order.
* Using the 'student\_data.csv' file we discussed in lectures, create a multiple boxplot graph showing student heights by gender and class year.
* Your output should look something like this.



**[1] Provide Python code to generate these graphics. Also, turn in a copy of your graphics results as image files (JPEG or PNG).**

## 1.2 Diffusion of Innovation analysis

Make a 2-D plot of iPad sales data over time using MatplotLib Pyplot tools. Include X and Y axis labels and a title. It should look something like this:



**[2] Provide Python code to generate these graphics. Also, turn in a copy of your results as an image file (JPEG or PNG). Save the image, as it will be used later in Problem #3.**

# Problem #2: Using Python to model diffusion of innovation.

In the lectures, we developed a Python model of the diffusion of innovation in a population. Enter the code we discussed in lectures into you Python Spyder IDE, and use this model to experiment with the diffusion phenomenon.

* Does the population size *N* seem to make any difference in the model behavior?
* Does the adoption rate *beta* seem to make any difference in the model behavior?
* Does the disposal rate *gamma* seem to make any difference in the model behavior?
* Does the time period *max\_time* seem to make any difference in the model behavior?
* Are there sets of parameters (*N, beta, gamma, max\_time*) that seem to align well to the real world diffusion of innovation for iPad sales? If so, what are these parameters?

**[3] Provide short answers to these questions. Note that these answers will be used later in Problem #3.**

# Problem #3: Course Final Paper Prep

As we are wrapping up the Python section of our course, let's document what we've found so far and prepare portions of our final LaTex technical paper.

* Create a LaTex Article document for your final project report. This will be a working template that we add material to throughout the rest of the semester.
* Use the fullpage package to set margins at 1" on the top, bottom, and sides.
* Use the graphicx package to allow graphics to be used in the article.
* Choose a title for your paper
* Add your name and affiliation (you may use GMU and CSI 500 if you wish)
* Leave the Abstract blank for now
* Don't do any table of contents, table of figures, or table of tables.
* Create sections for the following
  + Introduction
  + Background
  + Methods (will have a Data subsection)
  + Results
  + Discussion
  + Conclusions
  + References
* Leave the Introduction blank for now
* Under Background, cut and paste the material you previously created for your article literature review of the Rogers and Roberson articles. Include citations (which means you need to create a .bib file with the BibTex code for your citations).
* Under the Methods section, describe in a couple of paragraphs what we did to study diffusion of innovation using Python. You may wish to discuss the analysis, the classes used, and a general overview of the modeling approach.
* Add a subsection in Methods for Data, and cut and paste the table that you created earlier in the semester containing the iPad sales data. Include a short paragraph explaining what this data is and how it is important to our study.
* In the Data subsection, Include the 2D XY plot you created in Problem 1, and provide appropriate title for your figure. Include a short description of the graphic.
* In the Results section, add a couple of paragraphs explaining what you found using the Python model of diffusion of innovation. You may wish to include the narrative you developed for Problem #2. Did your findings from the Python model seem to align with the real-world iPad sales data?
* Leave Discussion blank for now
* Leave Conclusions blank for now
* Generate a bibliography under the References section. It should have at least the two papers you reviewed so far.

**[4] Turn in a PDF of your draft LaTex Article with the above material included.**

# Problem #4. Using Python Notebooks

In the lectures, we touched briefly on how to create and use Python notebooks with simple calculations and graphics. As an example of a more sophisticated notebook, download the file "**CSI500.ipynb**" and store it in folder on your desktop workspace. Using the **juptyer-notebook** executable at the command prompt, open the notebook. This notebook demonstrates how to load external libraries that you will use in later Data Science and Machine Learning courses. Click on the various cells, and see how the imbedded Python operations process the built-in Iris flower data set.

Using the notebook and your best judgement, answer the following questions about the Iris flower data set (note: this data set is considered a classic in the Machine-Learning community).

4.1. what is the mean petal size of each species of Iris?

4.2. what is the mean sepal size of each species of Iris?

4.3. could you distinguish species using only sepal length and width? Why or why not?

4.4 could you distinguish species using only petal length and width? Why or why not?

**[5] Provide answers to questions 4.1, 4.2, 4.3, and 4.4.**

Enjoy your Spring Break!