**CMSC 113: Computer Science I**
**Project: Data Visualization**
**sketch to be handed in during lab, March 27 or 29, 2018**
**due on Gradescope by the beginning of class on April 3, 2018**

**Collaboration Policy:** You may choose to work with a partner for this project. If you work with a partner, let me know as soon as you make your decision. If you choose not to work with a partner, you must follow the guideline "If you're talking in Java, you've gone too far." Remember, I take a breach of this policy seriously – if you're uncertain, ask before making any assumptions.

**Your task:** Write a program that enables a user to visualize some data. Additionally, your program must allow for some form of interaction between the user and the data (some suggestions are given below).

Your program should be able to read in the data from a file, parse it (e.g., using `split`), and then display the data clearly. Depending on your data and the story you wish to tell about it, you may wish to use a line chart, a bar chart, a pie chart, or some other visualization of your choice. Above all, the choice of visualization format should be informed by the story that you want to tell about your particular data.

Don't be afraid to be creative! Visual enhancements that don't necessarily convey additional information but make the presentation of the data more interesting are encouraged. Although the examples there are much more intricate than I expect for this project, you might be inspired by this list of visualizations.

Here are a few suggestions for how you might add an interactive element to your visualization:

1. Use text to display the exact value of a datapoint when the mouse hovers over the relevant portion of the visualization.
2. Zoom in on a particular segment of the data by clicking the mouse on a segment of the visualization.
3. Toggle the appearance of a moving average line (in a time series plot) with the click of a button.

You should select a dataset that lends itself to telling an interesting story through visualization. You are welcome to find your own data from any source you'd like as long as it contains a reasonable number of datapoints (a rough rule of thumb would be at least 50 or more).

Here are some places you might find data of interest:
1. US Government data on a range of subjects: data.gov
2. Product usage data: web browser popularity (also see here for more data).
3. Economic data: U.S. GDP (consult here for more economic time series).
4. Weather data: for Philadelphia consult the Franklin Institute Page.
5. Stock market data: data from the book *Irrational Exuberance* (see here for more detail).
6. A bunch of datasets: https://vincentarelbundock.github.io/Rdatasets/datasets.html

7. Federal Uniform Crime Reporting Statistics: [You can customize the kind of dataset you want.](#)

**Sketch to be handed in during lab:** In lab on Nov. 6 or 8, you will hand in a paper sketch of your visualization. This need not be very pretty (a free-hand drawing on a piece of paper is just fine), but it should show me what your final project will look like. In order to do this, you will have to identify your dataset and consider what kind of visualization would be best. I'll discuss your choices with you, and will have a chance to work on your visualization during lab. Note that handing in this sketch is a requirement of this assignment, and I will be comparing your final submission against this sketch. You may want to make a copy or picture of this sketch so you can use it as a guide to your work.

**Program requirements:**

1. Your program must read in data from a file.
2. The file must contain real-world data, with at least 50 data points.
3. You must display this data, choosing a visualization method that makes the data easy for a human to understand. This visualization must include all the information necessary to understand the data, including labels on the axes (if necessary), a legend (if necessary), a title, etc. If you present numerical data, you must include a scale (say, by including number labels on your axes).
4. Your program must contain an interactive feature, allowing the user to interact with the data.
5. Your program must uses classes and methods as necessary. For example, if each data point has a latitude, longitude, and color, these values should be stored in class fields of a new class `DataPoint` (or some more descriptive name). You would then have one array of `DataPoints` in your main program instead of three arrays.
6. Your code should avoid magic numbers related to your data. For example, if your data file has information on lines 17-63, instead of writing the numbers 17 and 63 in your code, write logic to detect the beginning and end of the data. This makes your program more robust if the data file is updated.
7. Your program must display (in a small font, likely) the source of your data.

It may be easiest to place the data files in the project folder within Eclipse. That way, they will get submitted when you submit your project on Gradescope. If your data is not included with your submission, I will be unable to grade the project; you will lose 20% of the project grade if I have to hunt you down to get the data.

Include in your submission a text file *reflections.txt* containing your answers to the following questions:

1. Does your final program meet up with your sketch?
2. If not, which parts were you unable to complete and why?
3. How long did this assignment take you?
4. What was the most challenging part?
5. Do you have any other questions or experiences to share?