Problem Set 3 - Part 2 (Reading)

Last Updated: 7 Feb 2019

Visualization with Python

Line plot and Scatter Plot

The following is an example of how you may produce a simple line plot using data stored in two lists. You first begin by importing the pyplot module of the matplotlib library. Type it out in your IDE and see what happens.

```
import matplotlib.pyplot as plt
x = [1, 2, 3, 4, 5]
y = [1, 4, 9, 16, 25]
plt.plot(x,y)
plt.xlabel('x values')
plt.ylabel('y values')
plt.title('y = x squared')
plt.show()
```

This produces a line plot. This set of code has the following steps that you need to plot any graph.

- Get your x and y data ready
- Plot the type of graph you want
- Label the x-axis and y-axis
- Put a title

If you want your graph to be represented by points, use a scatter plot. Replace the fourth line in the code above with the following statement.

```
plt.scatter(x,y)
```

References
More details on these functions can be found in the matplotlib documentation:

- **`plot`** function https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html
- **`scatter`** function https://matplotlib.org/api/_as_gen/matplotlib.pyplot.scatter.html

Statistical Plot - Box plot & histogram

Given a set of numerical data, a box plot (or box-and-whisker plot) shows you information about the minimum, 25th percentile, median, 75th percentile and maximum.
The following python code shows you how to generate a box plot.

```
some_scores = [49, 21, 22, 39, 7, 41, 66, 43, 27, 9, 14, 48, 33 ]
plt.boxplot(some_scores)
plt.show()
```

The median is the middle number. You could sort the list

```
some_scores.sort()
```

And see for yourself that the median is indeed what it is.

You could use a histogram to see the distribution of the data as well. In the code below, we specify that we want 5 bins.

```
plt.hist(some_scores, bins = 5)
plt.show()
```

Statistical Plot - Bar plot

Let's say you do a count and there are 20 male students and 32 female students in your class. You could display this information in a bar plot.

```
categories = ['female', 'male']
count = [32, 20]
plt.bar(categories, count)
plt.show()
```

References
More details on these functions can be found in the matplotlib documentation:

- **`hist`** function https://matplotlib.org/api/_as_gen/matplotlib.pyplot.hist.html
- **`boxplot`** function https://matplotlib.org/api/_as_gen/matplotlib.pyplot.boxplot.html
- **`bar`** function https://matplotlib.org/api/_as_gen/matplotlib.pyplot.bar.html

The range function

You may use the range function to generate a list of integer values.
For example, to get a list with contents [-3, -2, -1, 0, 1, 2], use the following code

```
x = list( range(-3, 3, 1) )
```

All inputs must be integers, and can be positive or negative.
The first input is the starting value.
The second input decides the last value you want to generate.
In this case we want to generate values *up to but not including* 3.
The third input is the increment.