Homework Assignment 1

I'd like for each of you to complete the assignment as a jupyter notebook that you will turn in and that I can then evaluate to see if it works. You will create your own notebook on the bio399cloud.org server, and download it directly from there, to be emailed to me by class on Wednesday. A good looking notebook contains a mixture of formatted text, code, and results such that the reader is easily guided through your train of thought.

Question 1: Write a function that will compute the reverse complement of a DNA sequence which has been stored as a string. If you are unsure of what that even means take a look at this wikipedia page on DNA complimentarity. **Hint:** while the reversal is simple enough, the complimentation might not be. Try using a Dictionary for this. To show that your function is working reverse complement the sequence "TACAGAT".

Question 2: In class we downloaded COVID case data from http://covidtracking.com aggregated across the entire country (i.e., all states) and used that to plot the cummulative number of positives over time. For the problem set you will plot data at the state level and compare growth trajectories among states.

First download the data as you did above by running the following either in a terminal window (without the preceding exclamation point '!') or below in a jupyter notebook cell with the exclamation point

```
! wget http://covidtracking.com/api/states/daily.csv
```

next let's check out the structure of this file by printing out the first line (I'm assuming you're using a jupyter notebook cell for the code below)

```
# let's print the first line of this file to see what the column names are
f = open("daily.csv.1",'r')
#print out the first line to see what things are called
print(f.readline())
```

okay so state is a new column, but the rest are basically the same as before. Here is your job– use that new column of information to plot the following.

1. The number of positives over time in New York versus California. You can do this in two separate plots but it's easy enough to compare two lines in the same plot by repeatedly calling `plt.plot()`

2. The number of deaths in New York versus California

3. Finally can you create a plot with a three way comparison of California, Oregon, and Washington?

Hint: take the following code cell and adjust it by adding a test to ask if a given line pertains to the state of interest

```
#make an array for my numbers
```

```
pos = []
my_file = open("daily.csv","r") #open a handle to the file
for line in my_file: #go through every line in file
    tokens = line.split(",") #split every line by the commas
    if tokens[2] != 'positive': #test if equal to header
        pos.append(int(tokens[2]))#append value to list as integer

print(pos)
```

Extra style moves: we haven't yet been able to cover the crucial aesthetic considerations that need to be considered when creating visual displays of quantitative information. These include proper labeling of axes, color pallete choices, and for the data that we are dealing with, data-time formats. Feel free to explore with any or all of these in creating your notebooks!