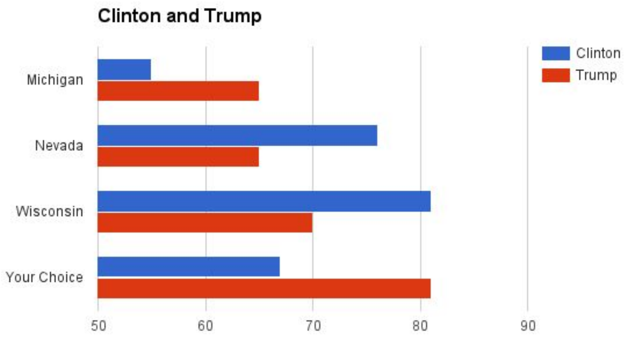
Assignment 4

Your task (to be completed as a group assignment) is based on using different twitter modules available for Python. You are required to perform sentiment analysis for four swing states: Michigan, Nevada and Wisconsin, and one swing state of your choosing. . You need to accomplish this using three modules: `tweepy`, `twython`, and `textblob`.

For tasks 1 and 2 you need to provide a graphical output which looks like the following (the x axis represents sentiment score):



Notes and clarifications

1. **`textblob`** provides a sentiment score between -1 and 1. Transform that to a score between 0 and 100, where 0 maps to 50.
2. Before calculating the sentiment score you need to "clean" each tweet -- by getting rid of stop words, making everything lower case and getting rid of special characters (as we had done in class). Use the cleaned tweet to create the **`textblob`** object from which you calculate the sentiment score.
3. Collect a maximum of 1000 tweets for each category (candidate and state). You can use any mechanism to impose this limit - a common approach being to use the `while` loop and breaking out of it once the tweet count is satisfied.
4. One way to define the geography for tweets is to use the geolocation in your search. It is equally acceptable to choose a city in a state (e.g. Detroit in Michigan) to obtain your tweets.

Task 1

To complete task 1, you first need to install `tweepy`. To do that you need to have pip for Python 3 installed.

```
 sudo apt-get update
 sudo apt-get -y install python3-pip
```

Once you have done that, you install `tweepy` with the following command

```
 pip install tweepy
```

Use `tweepy` and the twitter search terms "Clinton" and "Trump" and the respective states (or the geolocation search option) to provide the required output.

Task 2

To complete task 2, you need to have installed `twython`. I am assuming that you already have pip for Python 3 installed. Install `twython` with the following command

```
 pip install twython
```

Use `twython` and the twitter search terms “Clinton” and “Trump” and the respective states (or the geolocation search option) to provide the required output.

Task 3

Compare the packages twitter, tweepy, and twython in a Table format, as follows. Please limit your response to 5 or fewer criteria. Provide proper attributions and references for your responses.

	Twitter module		
	twitter	tweepy	twython
Criterion 1			
Criterion 2			
...			

Task 4 (bonus)

Explain what a streaming framework is and how it works (150 words max). Demonstrate the process / framework and provide the python code.

Useful resources

Tweepy: Link 1, Link 2, Link 3, Link 4, Link 5
Twython: Link 1, Link 2, Link 3
Textblob: Link 1, Link 2

How to submit your assignment

1. You need to submit three Python notebooks (for Tasks 1, 2 and 4) and one word or pdf file (for Task 3) to Blackboard to the Assignment 04 link.
2. The programs should be commented well enough so that the TA or I should not have to struggle with understanding variable names and codes and what the program does.
3. The grading rubric is shown below.

Rubric for Grading the Programming Assignment

	Unsatisfactory	Satisfactory	Good	Excellent
Delivery	● Completed less than 70% of the requirements. ● Not delivered on time or not in correct format (Blackboard or git)	● Completed between 70-80% of the requirements. ● Delivered on time, and in correct format (Blackboard or git)	● Completed between 80-90% of the requirements. ● Delivered on time, and in correct format (Blackboard or git)	● Completed between 90-100% of the requirements. ● Delivered on time, and in correct format (Blackboard or git)
Coding Standards	● No name, date, or assignment title included ● Poor use of white space (indentation, blank lines). ● Disorganized and messy ● Poor use of variables (many global variables, ambiguous naming).	● Includes name, date, and assignment title. ● White space makes program fairly easy to read. ● Organized work. ● Good use of variables (few global variables, unambiguous naming).	● Includes name, date, and assignment title. ● Good use of white space. ● Organized work. ● Good use of variables (no global variables, unambiguous naming)	● Includes name, date, and assignment title. ● Excellent use of white space. ● Creatively organized work. ● Excellent use of variables (no global variables, unambiguous naming).
Documentation	● No documentation included.	● Basic documentation has been completed including descriptions of all variables. ● Purpose is noted for each function.	● Clearly documented including descriptions of all variables. ● Specific purpose is noted for each function and control structure.	● Clearly and effectively documented including descriptions of all variables. ● Specific purpose is noted for each function, control structure, input requirements, and output results.
Runtime	● Does not execute due to errors. ● User prompts are misleading or non-existent. ● No testing has been completed.	● Executes without errors. ● User prompts contain little information, poor design. ● Some testing has been completed.	● Executes without errors. ● User prompts are understandable, minimum use of symbols or spacing in output. ● Thorough testing has been completed	● Executes without errors excellent user prompts, good use of symbols, spacing in output. ● Thorough and organized testing has been completed and output from test cases is included.
Efficiency	● A difficult and inefficient solution.	● A logical solution that is easy to follow but it is not the most efficient.	● Solution is efficient and easy to follow (i.e. no confusing tricks).	● Solution is efficient, easy to understand, and maintain.