Kevin R. Clemmons
(oduprogrammer16)
CS432 — Spring 2016   **Assignment 4**   February 26, 2016

All files mentioned in this file are uploaded into the *github* repository.

The LATEXcode invovled in the generation of this document was aided by the example code provided in the links that Dr. Nelson sent out on January 17, 2016[1].

This document was compiled on `www.sharelatex.com`

Problem 1

In order to extract the data located in *mln.graphml* a script *extractFacebookFriend-Data.py* was written. Since the given input file was written in xml, the xml-dom parser played a role in extracting the data [2] . After extracting of the xml-data is complete, an r-script is generated called *FacebookPlot.r*. This script generates the graph titled *NelsonPlot.pdf* and will display the mean, standard deviation, median,number of friends for Dr. Nelson, number of friends who have more friends than Dr. Nelson and the number of friends who have less friends than Dr. Nelson.
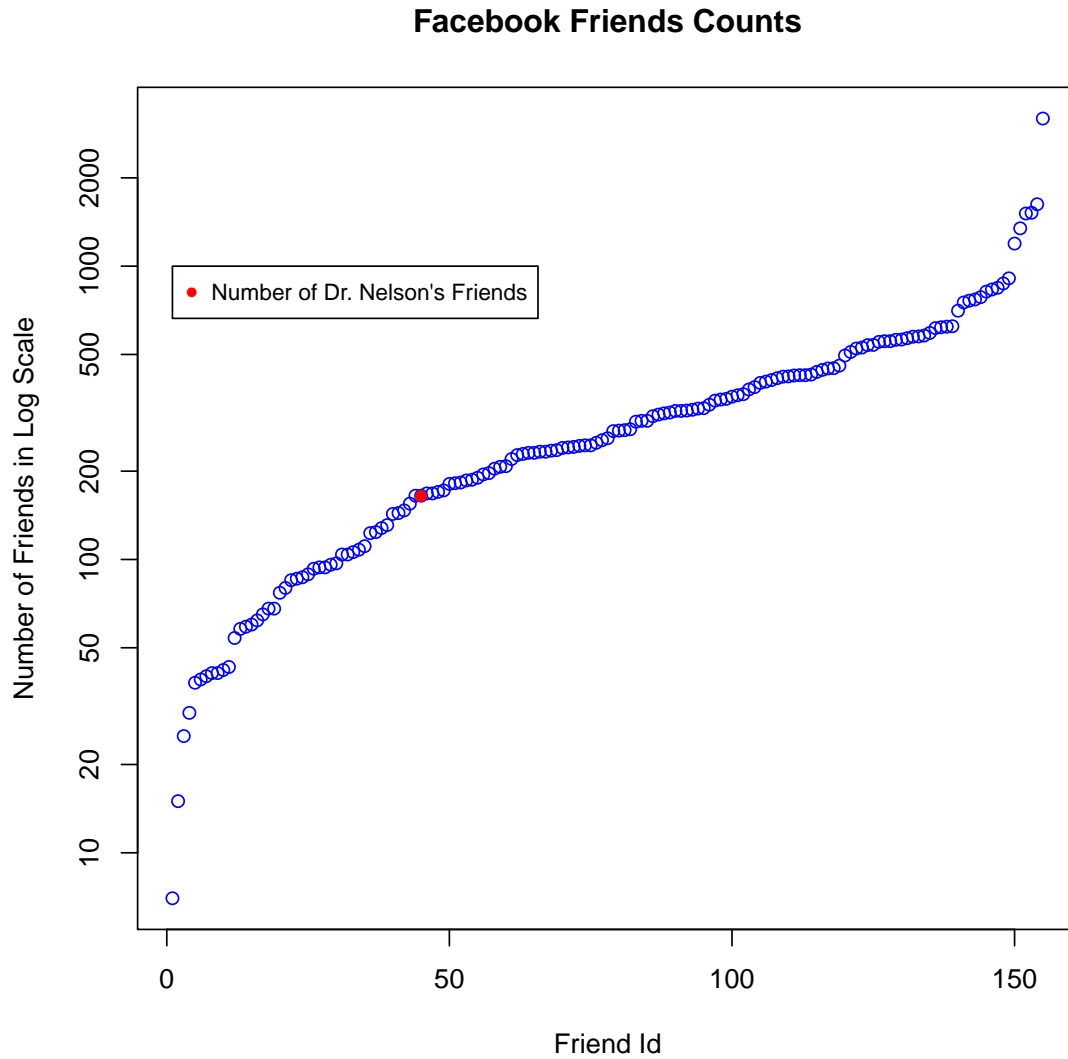
Facebook Friends Counts

Figure 1: Number of Friends that Dr. Nelson and His Friends Have

Assignment 4

Measure	Value
Mean	357.73
Standard Deviation	370.70
Median	259
Number of Friends for Dr. Nelson	165
Number of Friends Who have More Friends than Dr. Nelson	110
Number of Friends Who have Less Friends than Dr. Nelson	43

Table 1: Statistics from Figure-1, computed by the R-script

Based on the statistics in Table-1, it is concluded that for Dr. Nelson's facebook account in October 2013, the friendship paradox holds true due to the fact that 110 friends have more friends than Dr. Nelson.

Assignment 4

Problem 2

In order to extract the followers and the number of their followers a script called *twitterFriendData* has been created. The script takes advantage of the Tweepy library to acquire a list of followers and their counts [3]. Once the information on the followers and the number of their followers have been obtained, a json file called *twitteFollowerCounts.json* will then be generated.

Once the json file is created, another script called *createRScript.py* parses the json file and creates an r script called *twitterData.r* which produces a graph called *KevinTwitterPlot.pdf*. After the graph is generated, script will display the mean, median, standard deviation, number of followers, number of followers who have more followers than KevinClemmons and number of followers who have less followers than KevinClemmons .
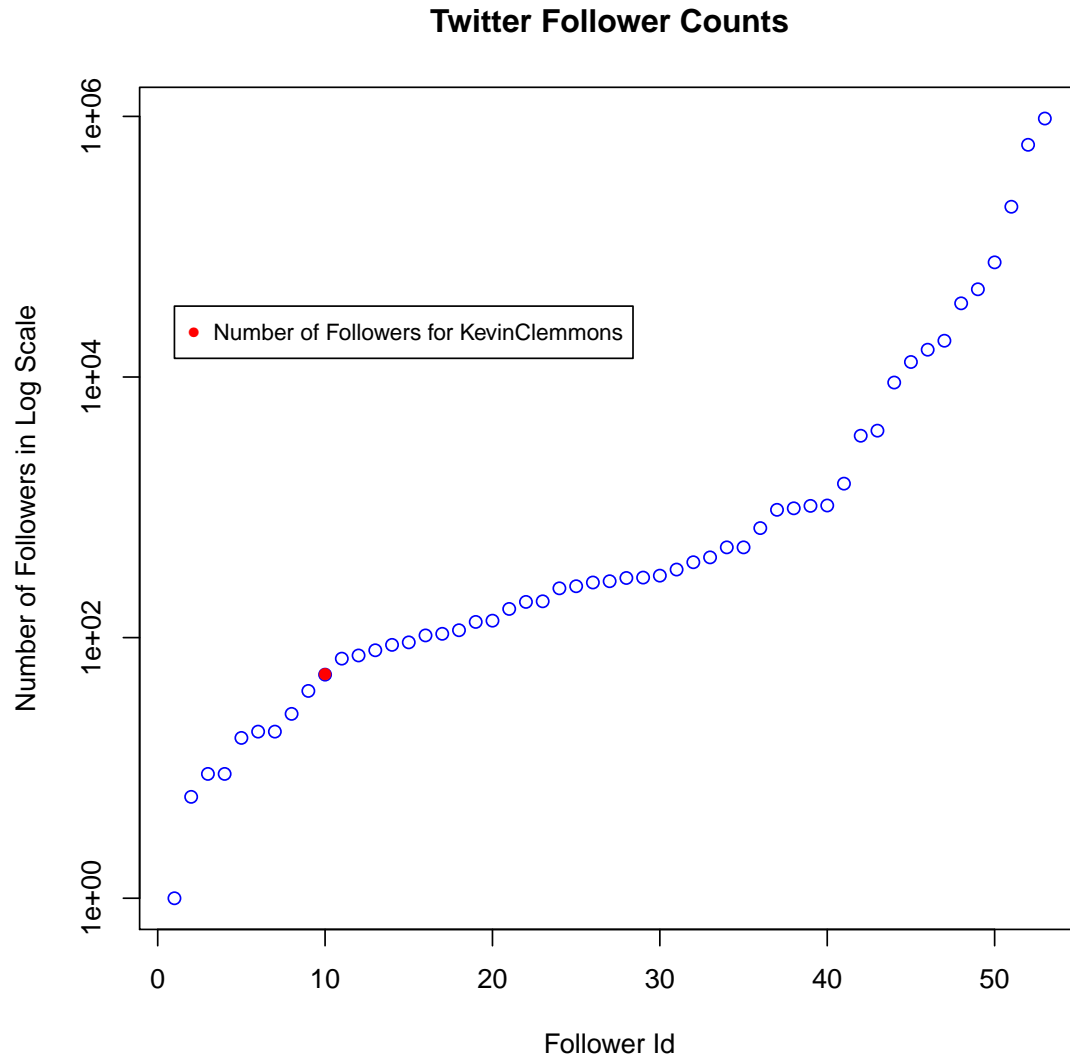
Figure 2: Number of followers that KevinClemmons and his friends have

Measure	Value
Mean	37888.301
Standard Deviation	156282.478
Median	271
Number of Followers for KevinClemmons	51
Number of Followers Who have More Followers than KevinClemmons	43
Number of Followers Who have Less Followers than KevinClemmons	9

Table 2: Statistics from Figure-2 which computed in the R-script

Based on the the statistics in Table-2, and the graph in Figure-2, it is concluded that for the twitter account KevinClemmons, the friendship paradox does indeed exist due to the fact that 43 followers have more followers than the account has.

Final Note

Lastly, an interesting thing to note is that the number of Followers who have More Followers than KevinClemmons is equal to the number of friends who have less friends than Dr. Nelson. Perhaps one may want to investigate the possibility of some form of relation exists between the two findings.

References

[1] Mohammed Aturban. *Assigment 9*. 2013. URL: https://github.com/maturban/cs595-f13/blob/master/assignment9/latex/Mohamed_Aturban_Assignment_9.tex.

[2] Python Software Foundation. *The Document Object Model API*. Jan. 2016. URL: https://docs.python.org/2/library/xml.dom.html.

[3] Joshua Roesslein. *Cursor Tutorial*. 2009. URL: http://docs.tweepy.org/en/latest/cursor_tutorial.html.