**Assignment Four**

Jacob Berlin

CS432 – Spring 2016

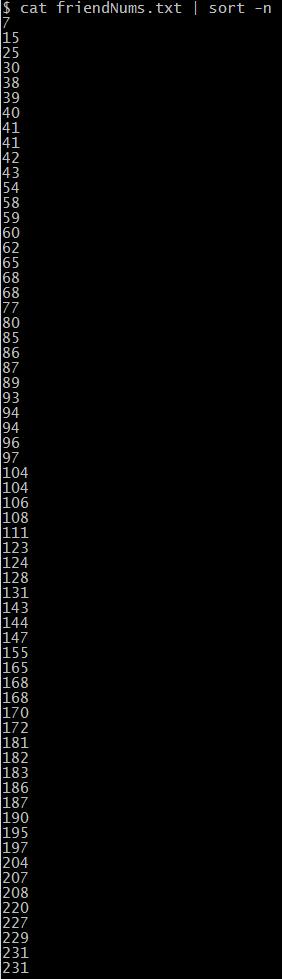
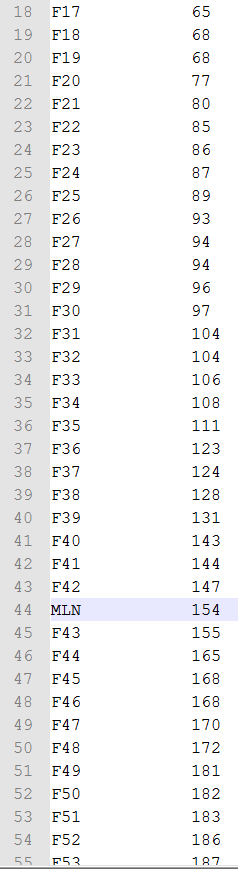
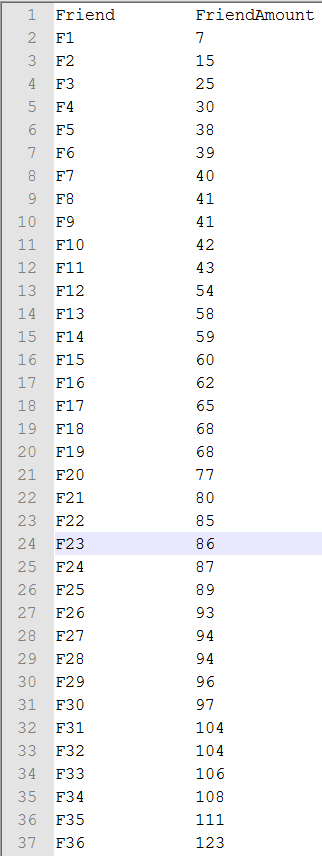
1. 1. Determine if the friendship paradox holds for my Facebook account.\* Compute the mean, standard deviation, and median of the number of friends that my friends have. Create a graph of the number of friends (y-axis) and the friends themselves, sorted by number of friends (x-axis). (The friends don't need to be labeled on the x-axis: just f1, f2, f3, ... fn.) Do include me in the graph and label me accordingly.



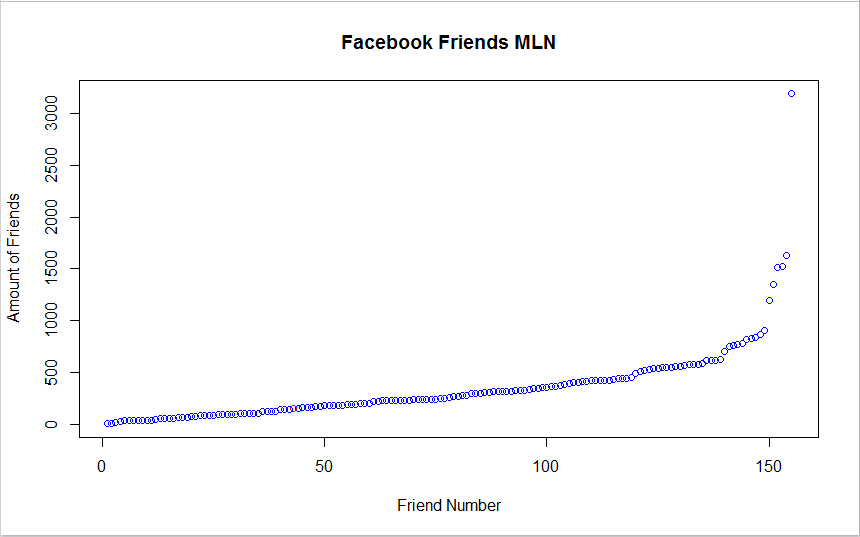
Initially for this problem I started with the raw XML format that was retrieved from the ‘mln.graphml’ file and then created a python program to parse the data.



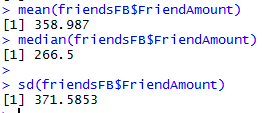
This file exclusively grabbed the amount of friends that each of the nodes in ‘rawFile.xml’ had and added it to a file called ‘friendNums.txt.’ Once I had retrieved all of the friends and the number of friends that each of them have, I sorted the list through the command-line using the sort –n command. After retrieving all of the data required, I parsed through the data and created a file with all of the friends along with adding the ‘MLN’ value to show you in the data.



Finally, after retrieving all of this data and formatting it as needed I went to RStudio and put the file into a plot generator. The graph I made shows each of the friends’ amount of friends alongside your own value from least to greatest. The ‘MLN’ value is shown in red while all of the other values are shown in blue.



To calculate the standard deviation, mean, and median of the list was just as simple as putting the data into R’s built in math functions. I made sure to take the ‘MLN’ value out of the dataset as to not skew the data.



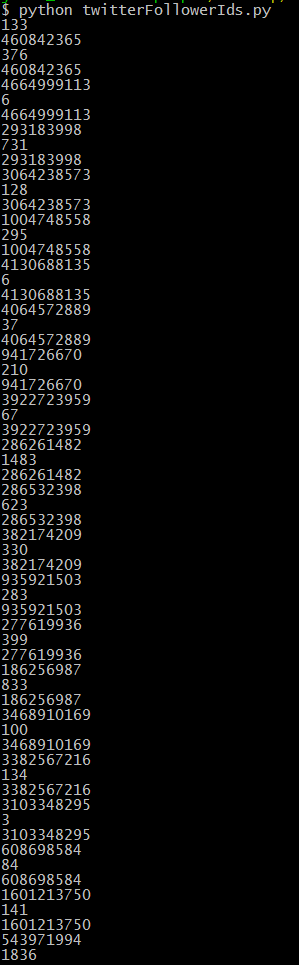
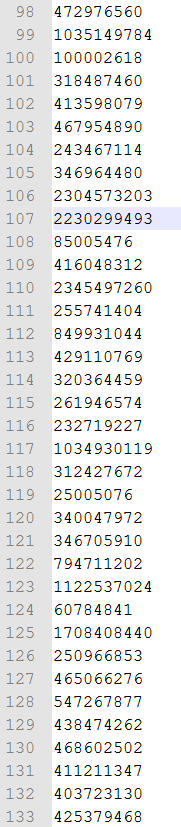
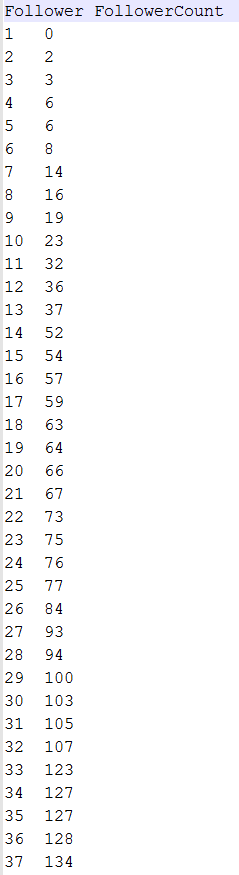
With your total of 154 friends, the friendship paradox holds for your Facebook account due to the fact that the majority of your friends have more friends on Facebook than you do. This is shown by the fact that your friend amount is smaller than the mean, median, and mode of the dataset.

2. Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use "followers" as value you measure (i.e., "do your followers have more followers than you?").

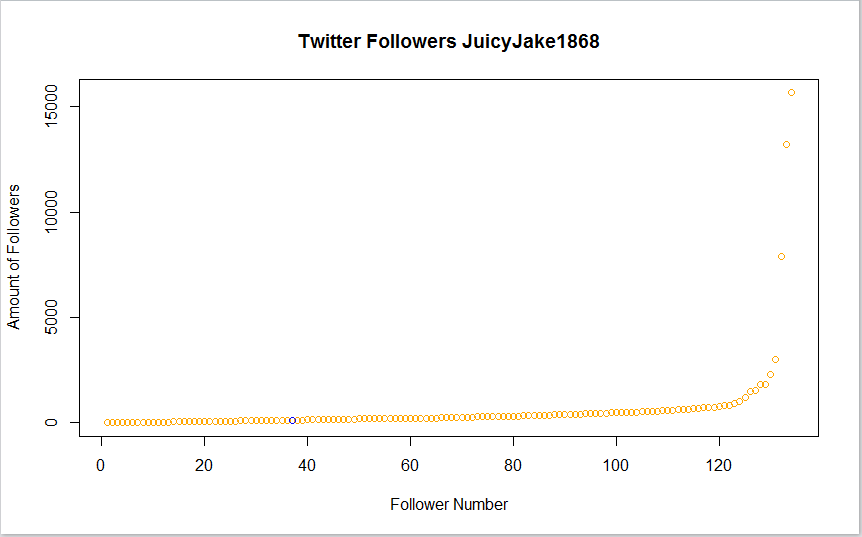
Generate the same graph as in question #1, and calcuate the same mean, standard deviation, and median values.



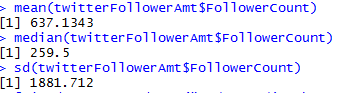
For question two, I started in the same fashion of question whereas I created a python file to parse through each of my personal Twitter account. Using the ‘Tweepy’ library, I was able to grab a list of both the follower’s ids along with the amount of followers that each of my followers have. Only the ids were stored in the ‘twitterFollowersIds.txt’. Just the follower amount values were then sorted using the sort –n command and then the sorted follower numbers were stored in the ‘twitterFollowerAmt.txt’



Along with in the first problem, I now had the entire list of data needed to put into RStudio to generate the graph. The specific value for my amount of followers is detailed in blue while my follower’s followers are detailed in orange.



To compute the mean, median, and standard deviation I plugged the values into R after making sure to take the ‘JWB’ value out.



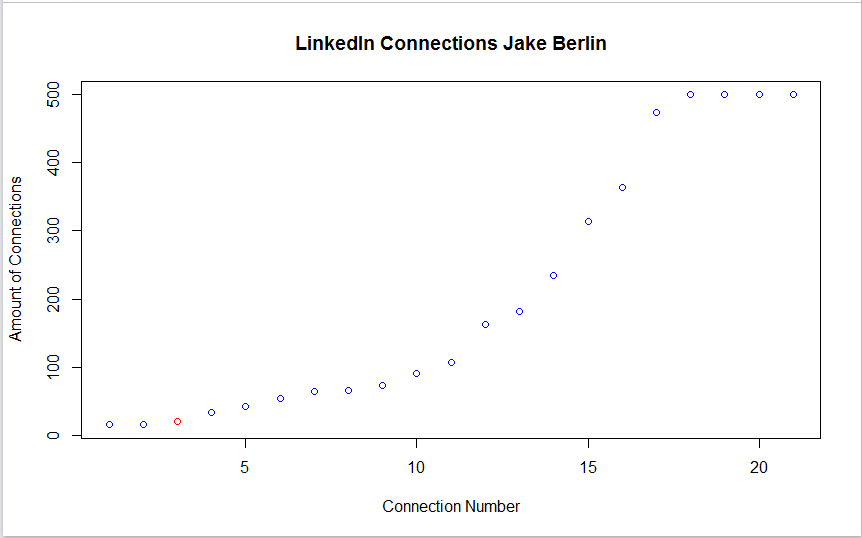
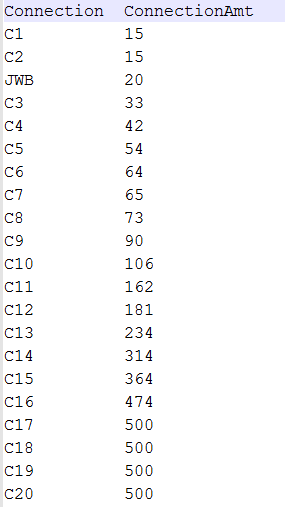
Therefore, by the rule of the friendship paradox and with my follower count being only at 133, the friendship paradox still holds for my account. This is due to my follower count being significantly less than all three of the calculated values.

3. E.C. Repeat question #1, but with your LinkedIn profile.

On my LinkedIn profile I only have 20 connections. (Account name is Jake Berlin)

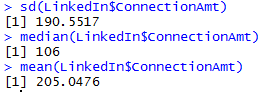
Due to the LinkedIn API not working as of 2/20/2016, I have manually gone through each of my connections and have grabbed all of their data.

Data:



As in question 1, I created a plot graph based on the amount of connections I had and the connections that each of my connections had. The connection in red is the ‘JWB’ connection, showing myself alongside the rest of the data.

I then calculated the standard deviation, mean, and median in R after making sure to take out the ‘JWB’ value.



Finally after retrieving all of the data needed and with only 20 connections the friendship paradox holds for my LinkedIn account due to the fact that my connection amount is marginally lower than the three values I calculated.