
CS432 Spring 2018

Assignment 4

Jonathan Kruszewski



Part 1:

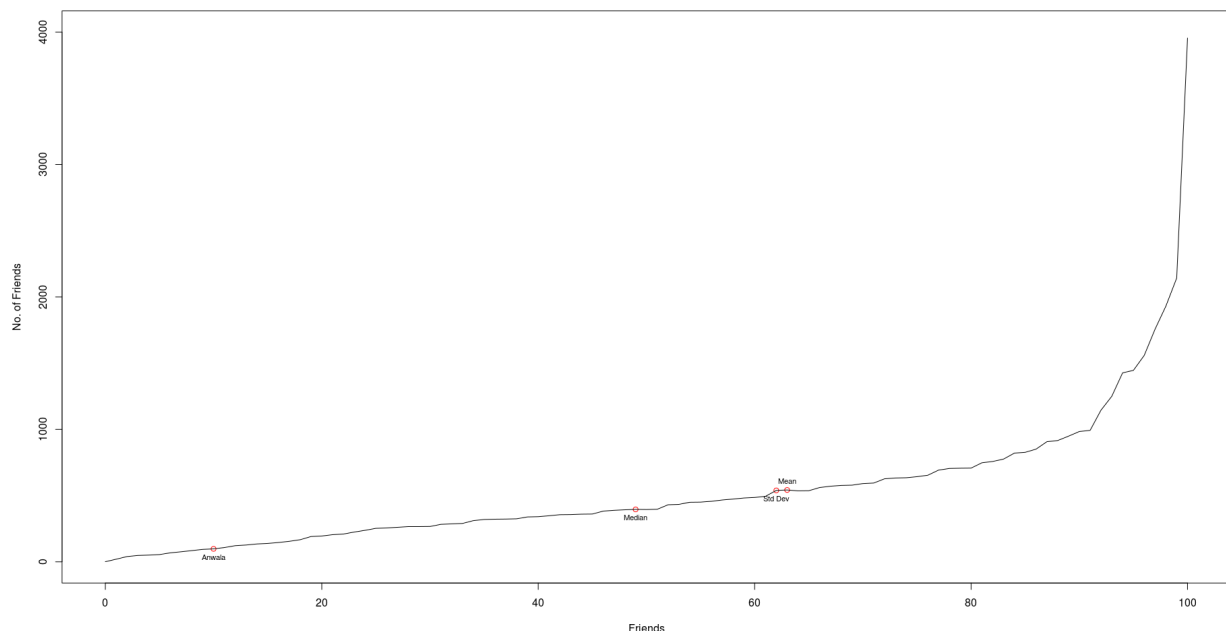
1. Determine if the friendship paradox holds for my Facebook account.* Compute the mean, standard deviation, and median of the number of friends that my friends have. Create a graph of the number of friends (y-axis) and the friends themselves, sorted by number of friends (y-axis). (The friends don't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

* = This used to be more interesting when you could more easily download your friend's friends data from Facebook. Facebook now requires each friend to approve this operation, effectively making it impossible.

I will upload a csv file of my 2014 friends list on the #assignment-4 slack channel

To accomplish this I have first copied the contents of the csv file shared in Slack into a text file named "friendCount.txt" I then used a python program titled "ReadNames.py" to read the list, sort it by number of friends, computer mean, median, and number of friends and inserted those along with the standard deviation into the list of friends. The sorted list can be found at "SortedFriendList.txt" I then executed the R code found below to make the graph.

```
d1 = read.table("SortedFriendList.txt")
d2 = read.table("Data2.txt")
{
plot(d1, type = "n", xlim=c(0,100), ylim=c(0,4000))
plot(d2, col="red", type = "p", xlim=c(0,100), ylim=c(0,4000), xlab = "Friends", ylab="No. of Friends")
lines(d1, type = "l", xlim=c(0,100), ylim=c(0,4000))
text(10, 98, labels="Anwala", cex= 0.7, pos = 1)
text(49, 396, labels="Median", cex= 0.7, pos = 1)
text(62, 539.43373852397, labels="Std Dev", cex= 0.7, pos = 1)
text(63, 542.67346938776, labels="Mean", cex= 0.7, pos = 3)
}
```



The friendship paradox appears to be true based on the provided list of friends and number of friends, as Anwala has less friends than the median and mean number of friends.

Part 2:

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 calculate the same mean, standard deviation, and median values.

For the Twitter 1.1 API to help gather this data, see:

<https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-list>

If you do not have followers on Twitter (or don't have more than 50), then use my twitter account "acnwala".

To accomplish this task the program "GetFollowers.py" was utilized.