Web Science Assignment 4 Report

Dr. Nelson

James Pindell

2/25/2018

Content

Problem 1

The "friendship paradox" (http://en.wikipedia.org/wiki/Friendship_paradox) says that your friends have more friends than you do.

Determine if the friendship paradox holds for Alex Nwala's 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 Alex Nwala in the graph and label me accordingly.

A csv file of Alex Nwala's 2014 friends list will be uploaded on the #assignment-4 slack channel

Solution

The solution for this problem is outlined by the following steps:

1. **Transfer Data from CSV to TXT**: Due to some technical difficulties, the csv file couldn't be used to determine if the friendship paradox holds. So instead, all the data was copied and pasted into a txt file. Also, all of the names of Alex Nwala's friends were replaced with simple counting numbers (i.e., 1, 2, 3…n). In addition, the data was manually rearranged from lowest friend count to highest friend count. You can view both the full csv and txt files, called acnwala-friendscount.csv and acnwala-friendscount.txt respectively, from within the package, but a snippet of the txt file is shown below:

```
USER FRIENDCOUNT
1 1
2 20
3 40
4 48
5 51
6 55
7 68
8 76
9 85
10 94
11 98
12 108
13 122
14 127
15 135
16 139
17 146
18 155
19 167
20 191
21 195
22 205
23 210
24 225
25 238
26 253
27 256
28 261
```

2. **Calculate Median, Mean, and Standard Deviation**: Once all of the data was transferred and rearranged, the median, mean and standard deviation of the data was calculated by using the corresponding functions in R. You can find the functions used to calculate these in the txt file, called Friend-Follower Plot.txt, from within the package, but a snippet of the code will be shown in the next step.

3. **Plot Data into Graph**: Once the median, mean, and standard deviation was calculated, all of the data was plotted into a graph. You can view both the code and the graph, called Friend-Follower Plot.txt and Facebook Friend Count Plot.pdf respectively, from within the package, but you can also view this graph at the bottom of the report (Graph 1), and a snippet of the code is shown below:

```
# Read table from acnwala-friendscount.txt (Problem 1) or jdpindell-followerscount.txt (Problem 2) and store info in data
data <- read.table("jdpindell-friendscount.txt", header = TRUE, sep = " ")

# Calcualte median, mean and standard deviation
follower <- (data$FOLLOWERS)
median(follower)
mean(follower)
sd(follower)

# Mark and identify Alex Nwala (Problem 1) or James Pindell (Problem 2), Median, Mean, and St. Dev.
jdp.follower <- with(data, data[follower == 60, ])
median.follower <- with(data, data[follower == 466, ])
mean.follower <- with(data, data[follower == 219969, ])
sd.follower <- with(data, data[follower == 597366, ])

# Plot data and mark Alex Nwala (Problem 1) or James Pindell (Problem 2), Median, Mean, and St. Dev. with different colors
plot(data, main = "Plot of Users and Follower Count", xlim = c(0, 65), ylim = c(0, 3500000), xlab = "Users", ylab = "Follower Count", pch = 20)
points(jdp.follower, col="red", pch=20)
points(median.follower, col="green", pch=20)
points(mean.follower, col="blue", pch=20)
points(sd.follower, col="purple", pch=20)

#Add Legend (Alex Nwala for Problem 1, James Pindell for Problem 2)
legend(1, 3500000, legend=c("James Pindell", "Median", "Mean", "St. Dev", "Data"), col=c("red", "green", "blue", "purple", "black"), pch=20, cex=1.0)
```

Note: The snippet of code shown here was the snippet of code used for Problem 2. For Problem 1, different key terms were used in the code to create the corresponding graph.

Problem 2

Determine if the friendship paradox holds for my (James Pindell) 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

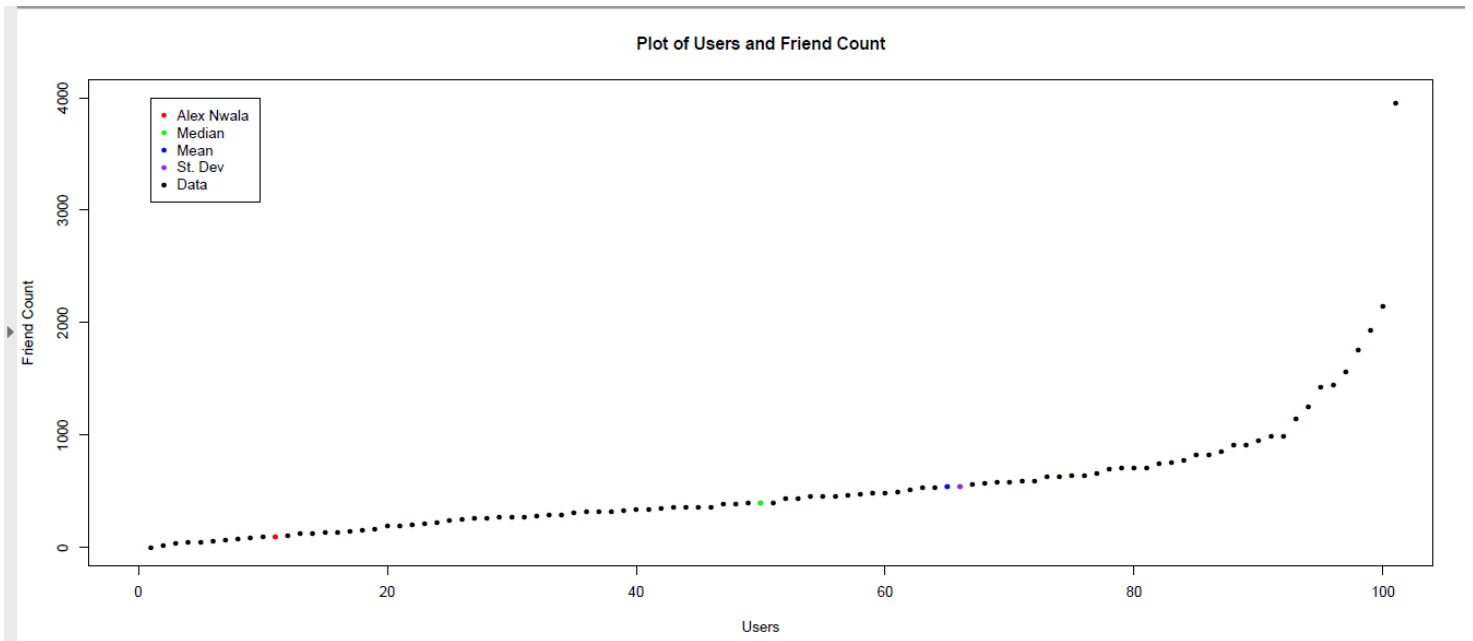If you do not have followers on Twitter (or don't have more than 50), then use Alex Nwala's twitter account "acnwala".

Solution

The solution for this problem is outlined by the following steps:

1. **Rinse and Repeat**: Simply put, the steps taken to solve this problem are the exact same steps taken as explained in Problem 1. The only difference is that instead of using Alex Nwala's Facebook friend count, as specified in the csv file, James Pindell's Twitter account was used instead. You can view all of the corresponding files from within the package, but you can view the corresponding graph at the bottom of this report (Graph 2), and a snippet of the data file, called jdpindell-followerscount.txt, is shown below:

```
USER FOLLOWERS
1 4
2 9
3 18
4 23
5 29
6 39
7 40
8 46
9 60
10 62
11 66
12 83
13 87
14 87
15 106
16 108
17 135
18 148
19 159
20 182
21 205
22 238
23 238
24 252
25 275
26 306
27 350
28 355
```

(Graph 1)

Plot of Users and Friend Count

Legend:
- Alex Nwala
- Median
- Mean
- St. Dev
- Data

(Graph 2)

Plot of Users and Follower Count