CS 495 - Introduction to Web Science

Fall 2014

Assignment 5

by

Eric Littley
UIN: 00821698

October 16, 2014

Instructor

Dr. Michael Nelson

Department of Computer Science
Old Dominion University

Honor Pledge

I pledge to support the Honor System of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagiarism. I am aware that as a member of the academic community it is my responsibility to turn in all suspected violations of the Honor Code. I will report to a hearing if summoned.

Signed: Eric Littley

Contents

1 Introduction

The purpose of assignment 5 was to explore the friendship paradox using twitter. The friendship paradox is essentially the rule that people have fewer friends than their friends have (on average). Several people were analyzed using the twitter API to determine if the friendship paradox held for them. These were Dr. Michael Nelson, Kim Kardashian, and myself.

2 Design

The assignment was split into three parts: getting the data from twitter, formatting the data, and visualizing the data. Python was used to extract data from twitter, a bash script was used to format the data, and R was used to graph the data.

The twitter API has a function that pulls down all the friends of a user (using their user name) along with information on each friend. One of the pieces of information is how many friends that friend has. A python script was written to pull down the list of friends and corresponding information for the three people mentioned above. A number, starting from one, was used to represent a friend. That number along with the corresponding number of friends for that friend was written to a file. A bash script was then used to sort the file from lowest number of friends to highest number of friends. An R script was written to graph the data, and to report the median, mean, and standard deviation. The graphs and results are shown below, the person being analyzed is shown in blue on the graph.

3 Graphs and Results

Nelson's friends:	148
Median:	191
Mean:	405.23
Standard dev:	546.41

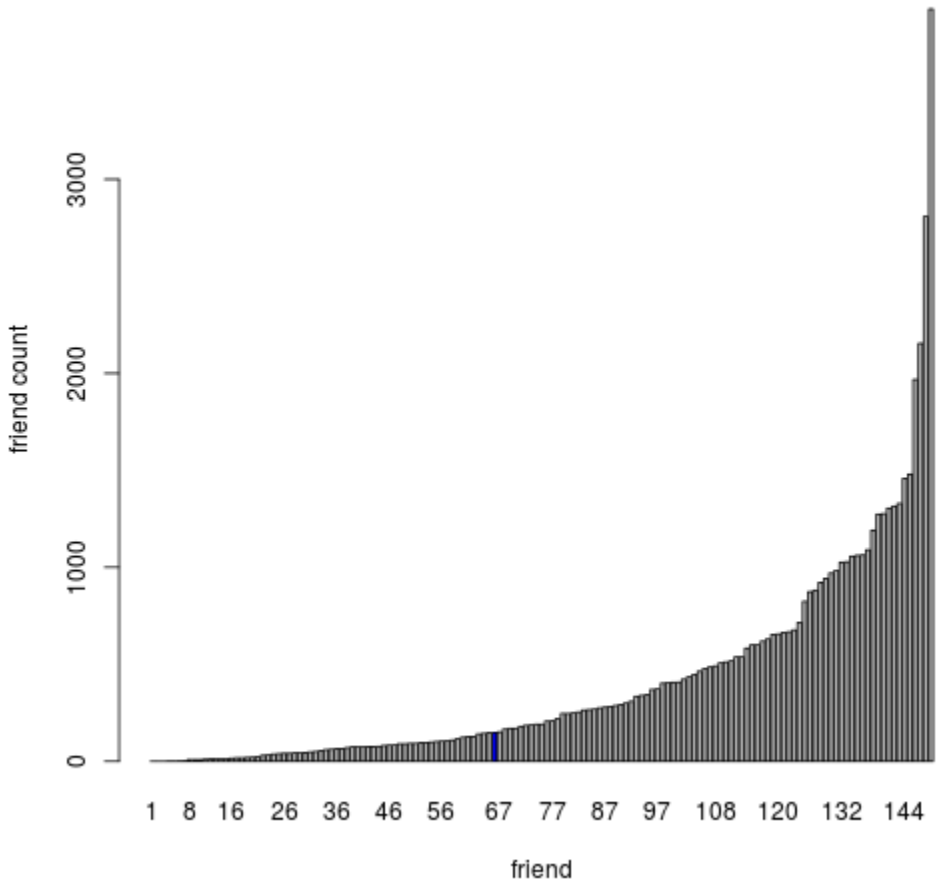

Figure 1: Nelson's Friend Graph

My friends:	14
Median:	53
Mean:	943.53
Standard dev:	2357.8

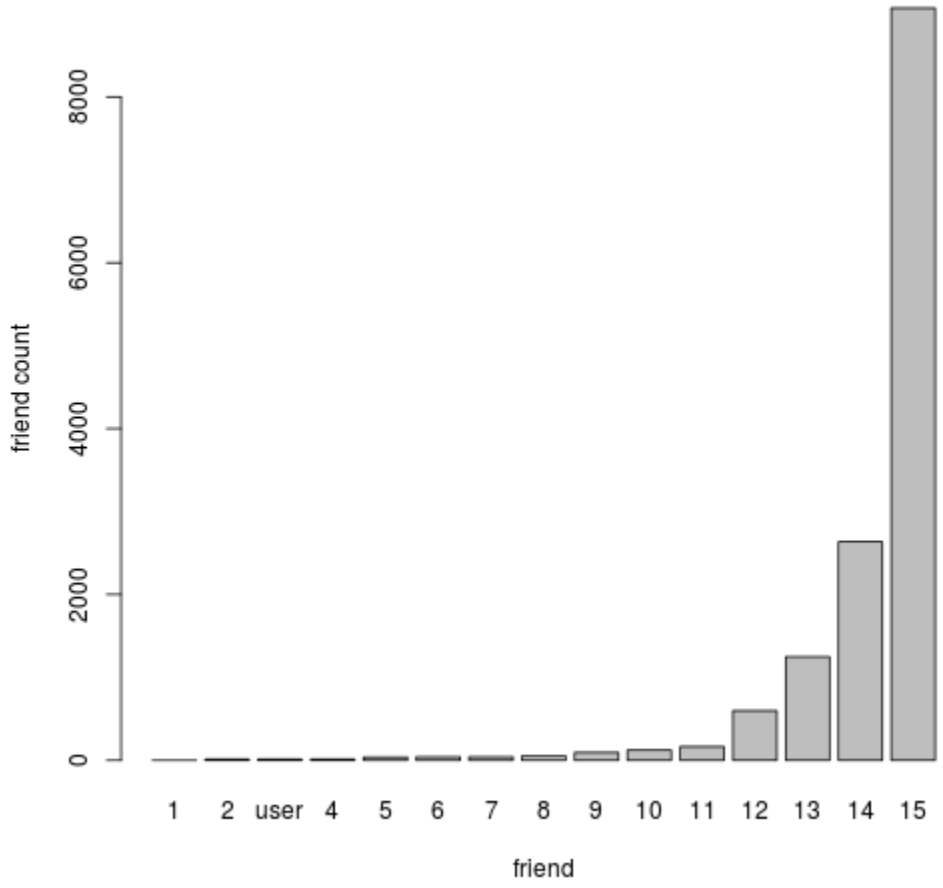

Figure 2: My Friend Graph

Kim's friends:	139
Median:	233.5
Mean:	816.66
Standard dev:	3282.4

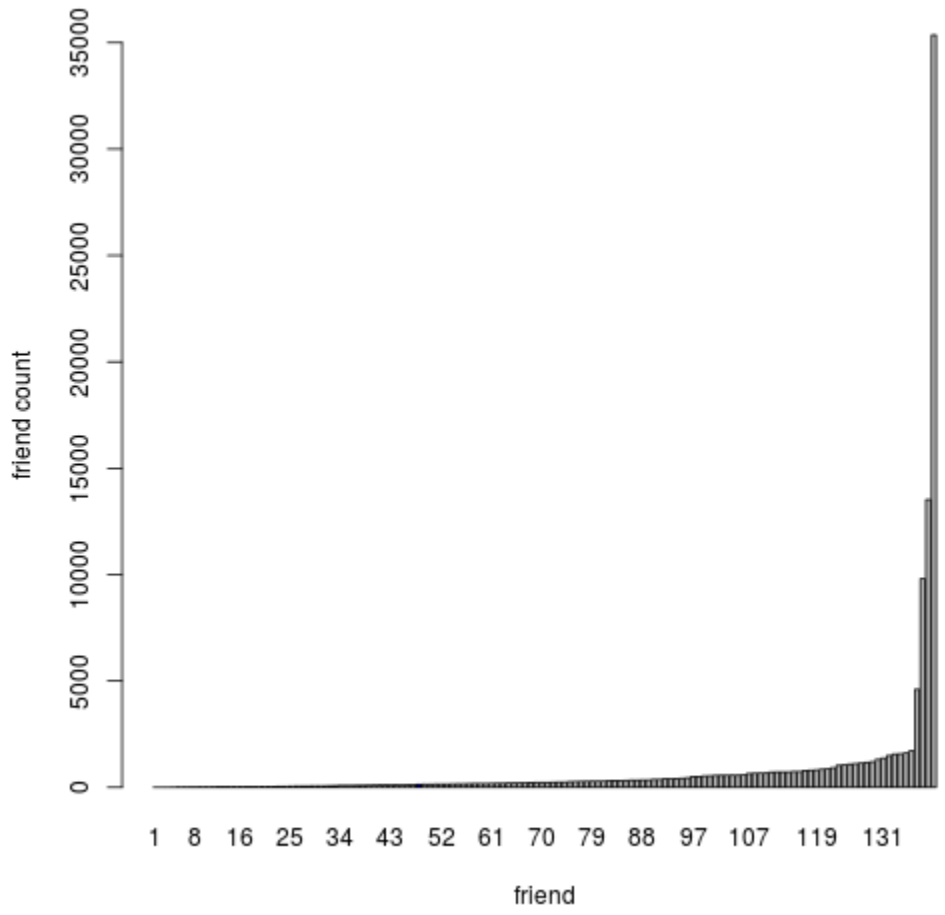

Figure 3: Kim Kardashian's Friend Graph

4 Discusion

The friendship paradox holds for everyone above. Since I only have 14 friends then it is not interesting that most of my friends have more friends than me. It may also be expected that Dr. Michael Nelson's friends, on average, have more friends than him since he is a relatively normal person (probably a little bit more famous than a normal person due to his professorship and research). However, Kim Kardashian may seem like an ideal candidate to break the friendship paradox because of her fame, but it holds. This is probably because she mostly follows people as famous as herself or more famous, and it does not appear that she

follows many people that follow her (139 friends vs millions of followers). She would undoubtedly break the paradox if we used followers as "friends" instead of people she follows.