Assignment 4

CS 432

Spring 2017

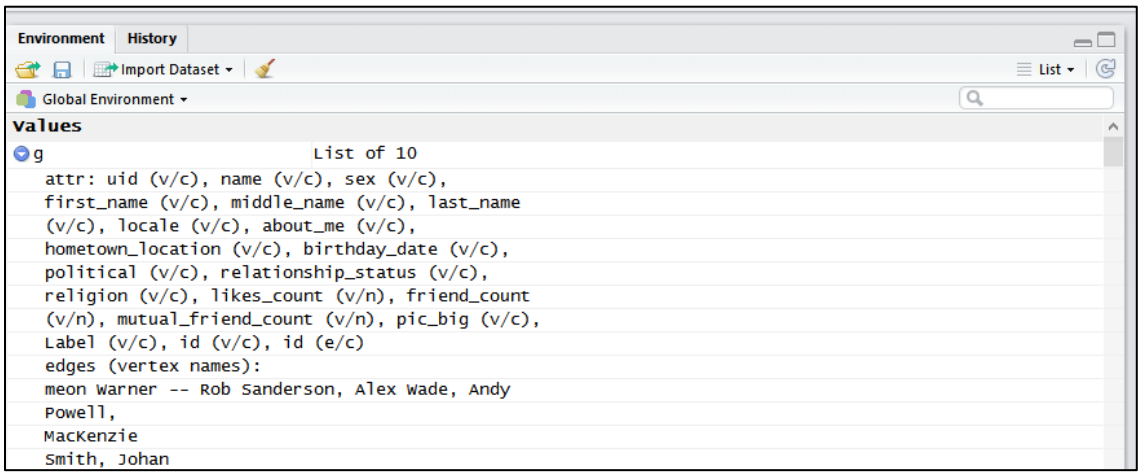
Michelle Graham

Question 1:

In order to process the given graphml file, I used RStudio to read the file and look at the data by using the igraph library.

```
import igraph

g=read.graph('/Users/mgrah/mln.graphml',format='graphml')
```

Summary(g) shows me the list of values or attributes associated with graph g, or it can be seen from the environment window under, "values."



Then, I created a data frame to look at the information from the graphml file. I played around with this a few times to look at different attributes and how they listed in the df values.

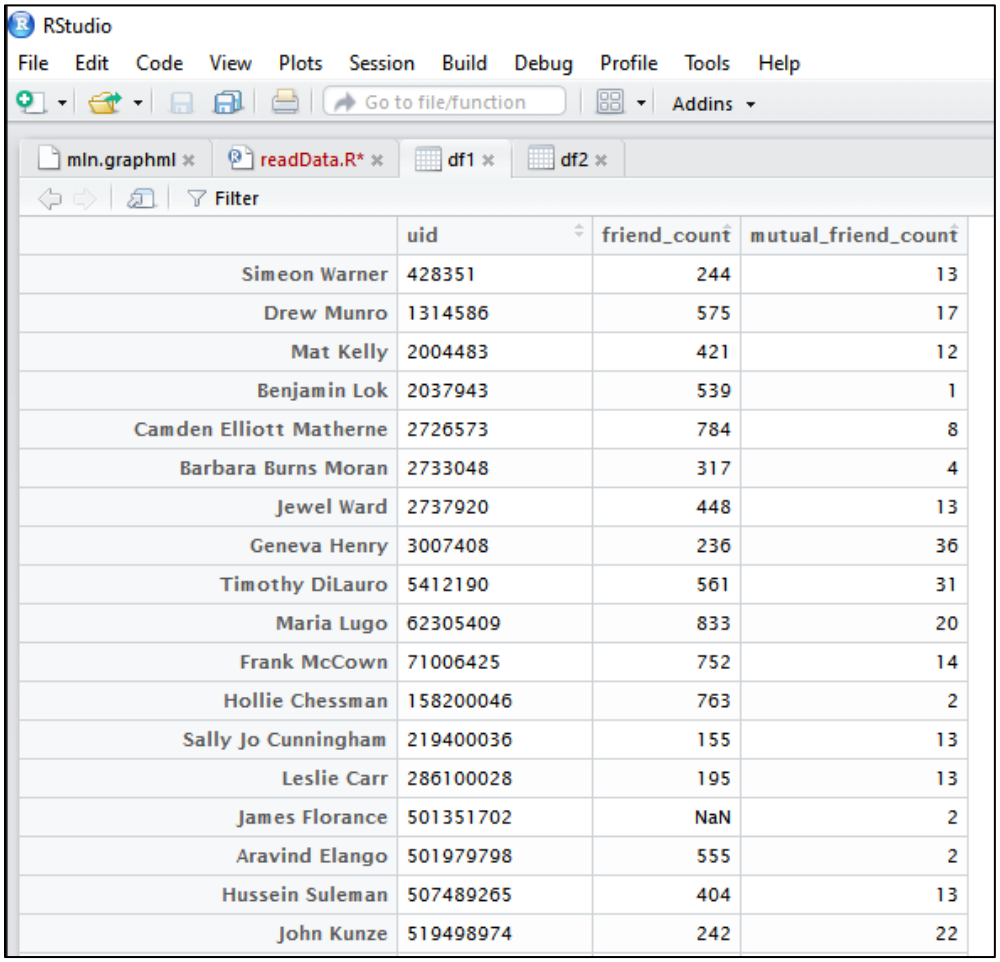
```
df <- get.data.frame(g, what='vertices')
df$id <- as.numeric(gsub("[A-Za-z]+", "", df$id) #if you need only the `numeric` part
row(df) <- NULL
df1 <- df[,c(1:2,15,16)]
df1

df2 <-df[15]
```

In order to do this, I looked at the vertices and edges and how the data was organized. After successfully isolating the desired data (name, id, friend_count, and mutual_friend_count), I deleted the nodes with empty friend lists manually in Notepad++ (in the graphml file) and saved it as a new file- mln_update.graphml.

```
> V(g)
+ 165/165 vertices, named:
  [1] Simeon Warner                Drew Munro                   Mat Kelly
  [4] Benjamin Lok                 Camden Elliott Matherne      Barbara Burns Moran
  [7] Jewel Ward                   Geneva Henry                 Timothy DiLauro
 [10] Maria Lugo                   Frank McCown                 Hollie Chessman
 [13] Sally Jo Cunningham          Leslie Carr                  James Florance
 [16] Aravind Elango               Hussein Suleman              John Kunze
 [19] Carlton Northern             Kat Hagedorn                 Jeffery Shipman
 [22] Hany SalahEldeen             Gregory Crane                Terry Harrison
 [25] Lillian Riley Cassel         Gary Marchionini             Mariana Rocha Biojone
 [28] Leslie Johnston              Sharon Spencer Stilwell      Brian E McCallum
+ ... omitted several vertices
```

```
+ ... omitted several vertices
> E(g)
+ 745/745 edges (vertex names):
 [1] Simeon Warner--Rob Sanderson                  Simeon Warner--Alex Wade
 [3] Simeon Warner--Andy Powell                    Simeon Warner--MacKenzie Smith
 [5] Simeon Warner--Johan Bollen                   Simeon Warner--Tyler Walters
 [7] Drew Munro   --Maria Lugo                     Drew Munro   --Hollie Chessman
 [9] Drew Munro   --Terry Harrison                 Drew Munro   --Mister-Photon Haptic-Interface Furr
[11] Drew Munro   --Kim Beveridge                  Drew Munro   --Fred Moore
[13] Drew Munro   --Cynthia Vaona                  Drew Munro   --Art Everett
[15] Drew Munro   --Daryl Schoolar                 Drew Munro   --Ellen Stein Beares
[17] Drew Munro   --Gregg Brooks                   Drew Munro   --Yuichiro Russell
[19] Drew Munro   --Scott Kinkade                  Drew Munro   --Johan Bollen
+ ... omitted several edges
>
```

Sample Data List:



	uid	friend_count	mutual_friend_count
Simeon Warner	428351	244	13
Drew Munro	1314586	575	17
Mat Kelly	2004483	421	12
Benjamin Lok	2037943	539	1
Camden Elliott Matherne	2726573	784	8
Barbara Burns Moran	2733048	317	4
Jewel Ward	2737920	448	13
Geneva Henry	3007408	236	36
Timothy DiLauro	5412190	561	31
Maria Lugo	62305409	833	20
Frank McCown	71006425	752	14
Hollie Chessman	158200046	763	2
Sally Jo Cunningham	219400036	155	13
Leslie Carr	286100028	195	13
James Florance	501351702	NaN	2
Aravind Elango	501979798	555	2
Hussein Suleman	507489265	404	13
John Kunze	519498974	242	22

Omitted from List:

James Florance	NaN
Joy Gooden	NaN
Kim Beveridge	NaN
Alfredo SÃ¡nchez	NaN
Sarah Shreeves	NaN
Sally Mauck	NaN
Dan Swaney	NaN
Robert Gordeaux	NaN
Joseph Kaplan	NaN
Michael Milner	NaN
Catherine Kemble Cronin	NaN

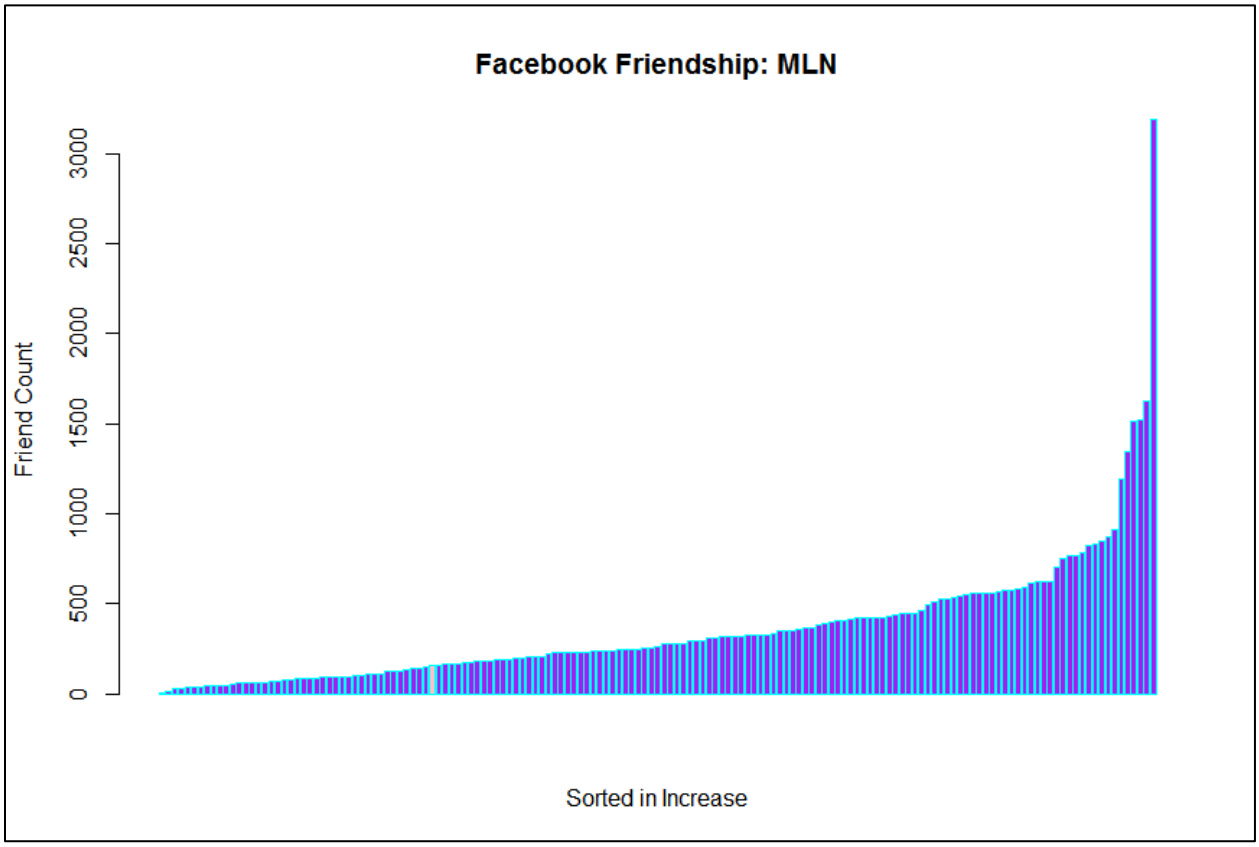
Sort:

```
> sort(x=c(244, 575, 421, 539, 784, 317, 448, 236, 561, 833,
 752, 763, 155, 195, 555, 404, 242, 425, 366, 321, 1194, 259
, 427, 297, 400, 592, 424, 555, 97, 387, 622,  337, 496, 705
, 819, 229, 1521, 324, 208, 619, 227, 3187, 351, 181, 104, 2
95, 233, 348, 87, 1512, 190, 363, 62, 315,  449, 186, 380, 2
97, 359, 274, 245, 425, 562, 275, 240,  510, 409, 245, 1626,
 58, 89, 278, 30, 580, 197, 321, 276,  68, 168, 182, 124, 23
3, 552, 131, 615, 1346, 436, 770,  322, 93, 94, 106, 568, 17
0, 143, 128, 220, 312, 844, 255,  420, 624, 204, 576, 524, 1
68, 873, 231, 68, 443, 65, 241, 86, 144, 54, 172, 60, 250, 4
3, 183, 909, 94, 38, 528, 40, 80, 108, 231, 458, 41, 42, 235
, 327, 15, 187, 207, 165, 7, 77, 308, 415, 111, 328, 123, 10
4, 538, 147, 353, 59, 41, 96, 85, 25, 39),decreasing=FALSE)
  [1]    7   15   25   30   38   39   40   41   41   42
 [11]   43   54   58   59   60   62   65   68   68   77
 [21]   80   85   86   87   89   93   94   94   96   97
 [31]  104  104  106  108  111  123  124  128  131  143
 [41]  144  147  155  165  168  168  170  172  181  182
 [51]  183  186  187  190  195  197  204  207  208  220
 [61]  227  229  231  231  233  233  235  236  240  241
 [71]  242  244  245  245  250  255  259  274  275  276
 [81]  278  295  297  297  308  312  315  317  321  321
 [91]  322  324  327  328  337  348  351  353  359  363
[101]  366  380  387  400  404  409  415  420  421  424
[111]  425  425  427  436  443  448  449  458  496  510
[121]  524  528  538  539  552  555  555  561  562  568
[131]  575  576  580  592  615  619  622  624  705  752
[141]  763  770  784  819  833  844  873  909 1194 1346
[151] 1512 1521 1626 3187
```

Plot:

I had trouble learning to sort data properly in R, so I used a roundabout method of adding and editing data. Instead of reading from a text file, graphml file, or table, I just added all of the friend_counts to the "height" value in a barplot in R. I was also able to use a sort function in R to line them up correctly.

```
> barplot(height = c(7, 15, 25, 30, 38, 39, 40, 41, 41, 42, 43, 54, 58, 59, 60, 62, 65, 68, 68, 77, 80, 85, 86, 87, 89, 93, 94,
 94, 96, 97, 104, 104, 106, 108, 111, 123, 124, 128, 131, 143, 144, 147, 154, 155, 165, 168, 168, 170, 172, 181, 182, 183, 186,
 187, 190, 195, 197, 204, 207, 208, 220, 227, 229, 231, 231, 233, 233, 235, 236, 240, 241, 242, 244, 245, 245, 250, 255, 259, 2
74, 275, 276, 278, 295, 297, 297, 308, 312, 315, 317, 321, 321, 322, 324, 327, 328, 337, 348, 351, 353, 359, 363, 366, 380, 387
, 400, 404, 409, 415, 420, 421, 424, 425, 425, 427, 436, 443, 448, 449, 458, 496, 510, 524, 528, 538, 539, 552, 555, 555, 561,
562, 568, 575, 576, 580, 592, 615, 619, 622, 624, 705, 752, 763, 770, 784, 819, 833, 844, 873, 909, 1194, 1346, 1512, 1521, 162
6, 3187), col=col, border='cyan',main="Facebook Friendship: MLN",xlab="Sorted in Increase",ylab="Friend Count")
```



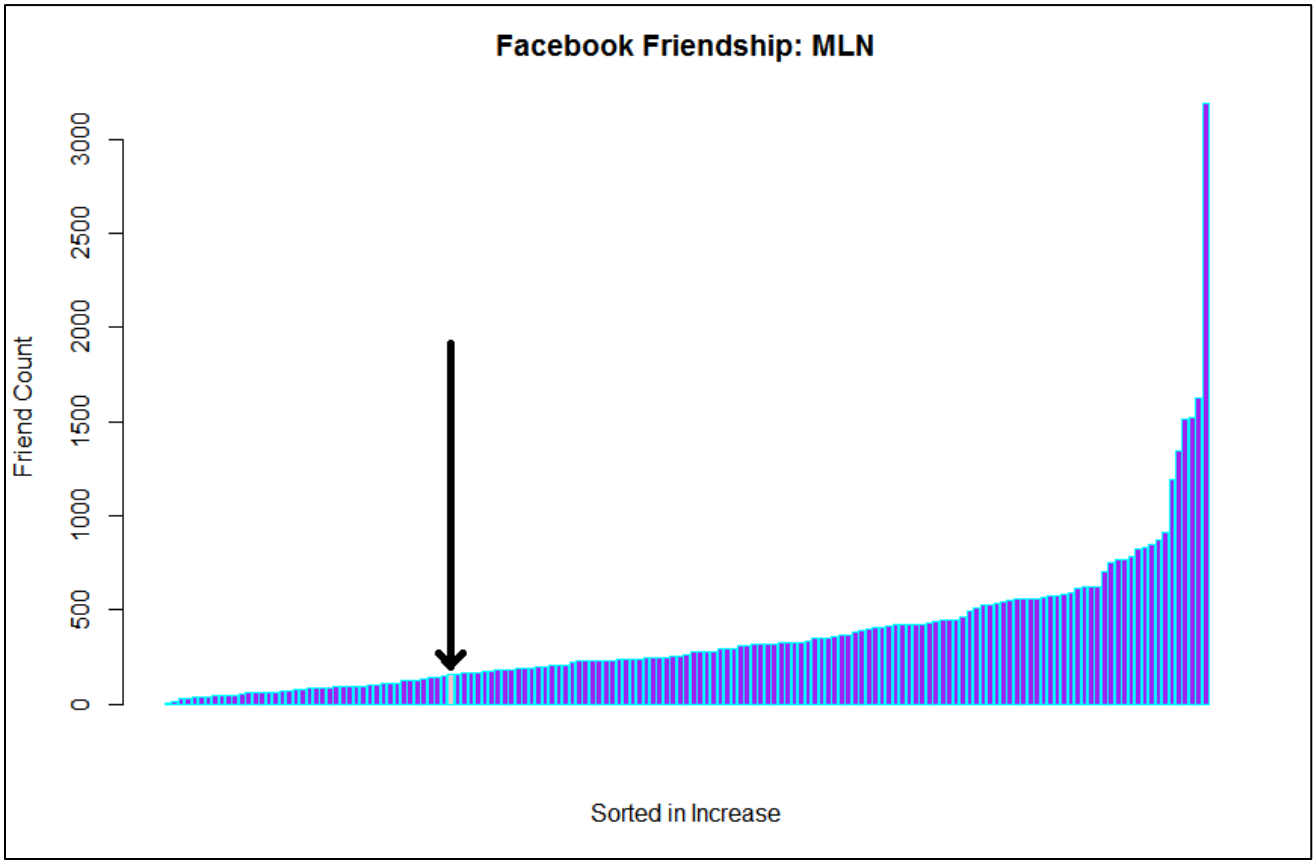
For Nelson's Bar:

For every bar in my list of values, each bar would be purple up until bar number forty-three, which is Nelson's bar! In order to differentiate between his bar and the others, I created a "col" value to hold three different instances of color based on the bar's number in the list. For every bar up to forty-two is purple, forty-three is pink, and the remaining one hundred and eleven are purple.

```
> col <- c(rep(c("purple"),42),rep(c("pink"),1),rep(c("purple"),111))
```

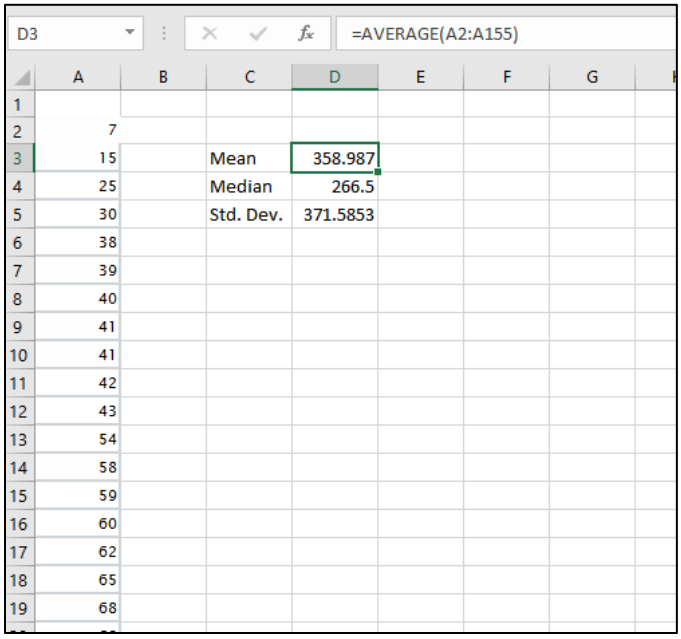
Since it is a little hard to see the Nelson bar, so I added an arrow with Paint.

Edited Photo:

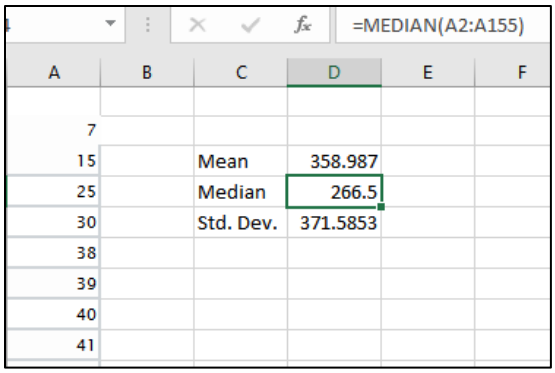


I used Microsoft Excel to calculate the mean, median, and standard deviation for the data excluding the Nelson bar.

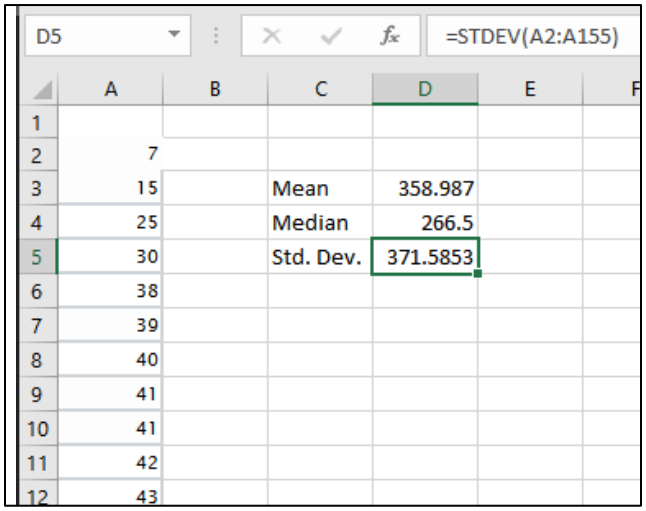
Mean:



Median:



Standard Deviation:



The friendship paradox holds true!

Question 2:

I used Tweepy to access my Twitter data to see how many followers I have and list their names and number of followers. This is very closely related to how assignment two was administered. For a while, I was stuck on a rate limit error, however, I found an example that used a counter to resolve the issue.

Code:

```
#coding: utf-8
import tweepy
from tweepy import OAuthHandler
import sys

def init():
    global api
consumer_key = "uIHgcJNDmmbeunAQ3PtwnB7aW"
consumer_secret = "3E15kDsBLyfIFncgxQ3Cnh6Vu6TTVsVuTa0y5nt6oIx2JVV921"
access_key="721752096169705476-M8XjTgLwx23Dv5wFn7QwSyQJywvMTFI"
access_secret="aiaYQEY5Tg9KIu17OmJFt9ikZmpjmwv69Zh13CfobJSMy"
auth=tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_key, access_secret)

api = tweepy.API(auth)

file = open('TwitterFollowers.txt', 'w')

user = api.get_user("koidumpling")
print user.screen_name
print user.followers_count

lFollowers = []
for user in tweepy.Cursor(api.followers, count = 50).items():
    lFollowers.append((user.followers_count, user.screen_name))

lFollowers.sort()
for follower in lFollowers:
    print "%d - %s" % (follower[0], follower[1])
```

Output:

```
mgrah@DESKTOP-3OIR4AC MINGW64 /c/CS432/A4
$ python getFollowers.py
koidumpling
49
15 - piti_1234
31 - holaaagirl
42 - AprilsHair101
45 - LohGrandeB
69 - itsknodtluke
80 - bryancarey432
91 - sysoevakira3
113 - Wayne_King1980
135 - scotchmonkey118
178 - TylerRacing
196 - tinyasiannigger
209 - Strictly_Memes
288 - MrMeeseeks187
305 - NotClarence1
411 - EasyRizzy
436 - ExiledArchitect
534 - PikminGuts92
556 - YaBoyStanny
618 - TheGodDamnDM
908 - Kev__Jackson
1012 - D_Buirst
1020 - SeanG_Baker
1273 - renfid
1818 - potatefact
1916 - sandipan__sarma
2000 - bonafidedamien
3122 - Freddygzone
3322 - Srizlys
3408 - marketingdoctor
3442 - ClydeAlexanderu
4545 - DexBurnomII
6610 - BlakeLemay2
10796 - Sunni_Tzu
13886 - heavensrecipes
14429 - hiDeb_wutzDope
15400 - itslifethOught
18008 - Gam3Wrld
19336 - cynicalhumor
21777 - BobsFunnyJokes
25516 - thefunnyteeng
27385 - ernest6words
79004 - billbelew_com
100807 - Dystify
126107 - relatable_poem
140125 - ExclusiveGems
140330 - BallinCribs
179488 - darrynzewalk
246554 - ThisFoodTho
406140 - attacksn
```

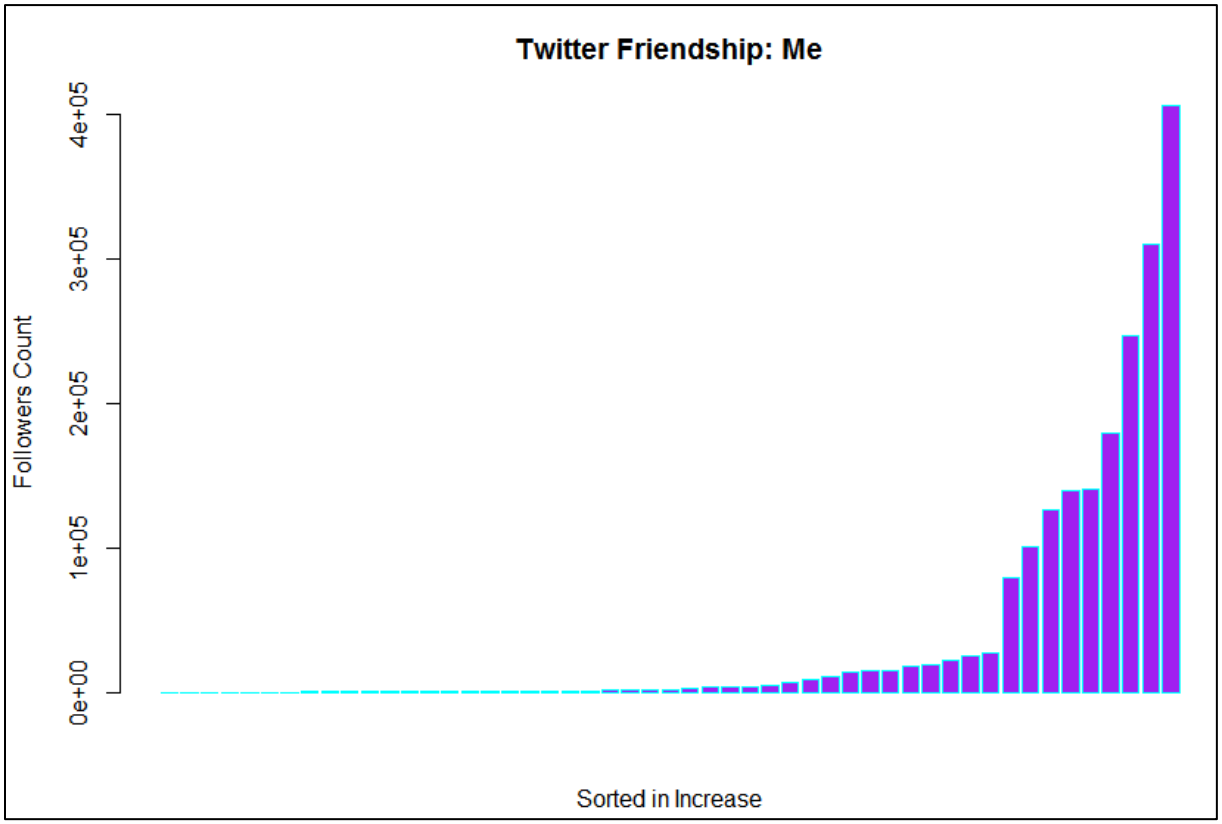
After collecting the desired data from Twitter and storing it in a text file, I loaded it into R to generate a graph.

```
> read.table('C:/CS432/A4/TwitterFollowers.txt')
       V1 V2              V3
1      15  -       piti_1234
2      31  -      holaaagirl
3      42  -   AprilsHair101
4      45  -      LohGrandeB
5      69  -    itsknodtluke
6      80  -   bryancarey432
7      91  -    sysoevakira3
8     113  -  Wayne_King1980
9     135  - scotchmonkey118
10    178  -     TylerRacing
11    198  - tinyasiannigger
12    211  -  Strictly_Memes
13    288  -   MrMeeseeks187
14    305  -    NotClarence1
15    412  -       EasyRizzy
16    437  - ExiledArchitect
17    534  -    PikminGuts92
18    556  -     YaBoyStanny
19    618  -    TheGodDamnDM
20    908  -    Kev__Jackson
21   1012  -        D_Buirst
22   1021  -     SeanG_Baker
23   1273  -          renfid
24   1820  -      potatefact
25   1916  - sandipan__sarma
26   2000  -  bonafidedamien
27   3122  -     Freddygzone
28   3321  -         Srizlys
29   3408  - marketingdoctor
30   3441  - ClydeAlexanderu
31   4544  -     DexBurnomII
32   6611  -     BlakeLemay2
33  10796  -       Sunni_Tzu
```

```
> barplot(height = c(15, 31, 42, 45, 69, 80, 91, 113, 135, 178, 199, 214, 289, 305, 414, 437, 534, 556, 620, 908, 1011, 1021,
 1273, 1820, 1917, 2003, 3122, 3323, 3408, 3441, 4545, 6611, 8716, 10797, 13886, 14429, 15408, 18008, 19336, 21775, 25519, 27
388, 79011, 100807, 126108, 140128, 140335, 179489, 246560, 310115, 406148), col=col, border='cyan',main="Twitter Friendship:
 Me",xlab="Sorted in Increase",ylab="Followers Count")
> barplot(height = c(15, 31, 42, 45, 69, 80, 91, 113, 135, 178, 199, 214, 289, 305, 414, 437, 534, 556, 620, 908, 1011, 1021,
 1273, 1820, 1917, 2003, 3122, 3323, 3408, 3441, 4545, 6611, 8716, 10797, 13886, 14429, 15408, 18008, 19336, 21775, 25519, 27
388, 79011, 100807, 126108, 140128, 140335, 179489, 246560, 310115, 406148), col="purple", border='cyan',main="Twitter Friend
ship: Me",xlab="Sorted in Increase",ylab="Followers Count")
>
```

For some reason the y-axis displayed some weird number formatting. I did not have time to fix this, however, the graph still displays the data accurately.

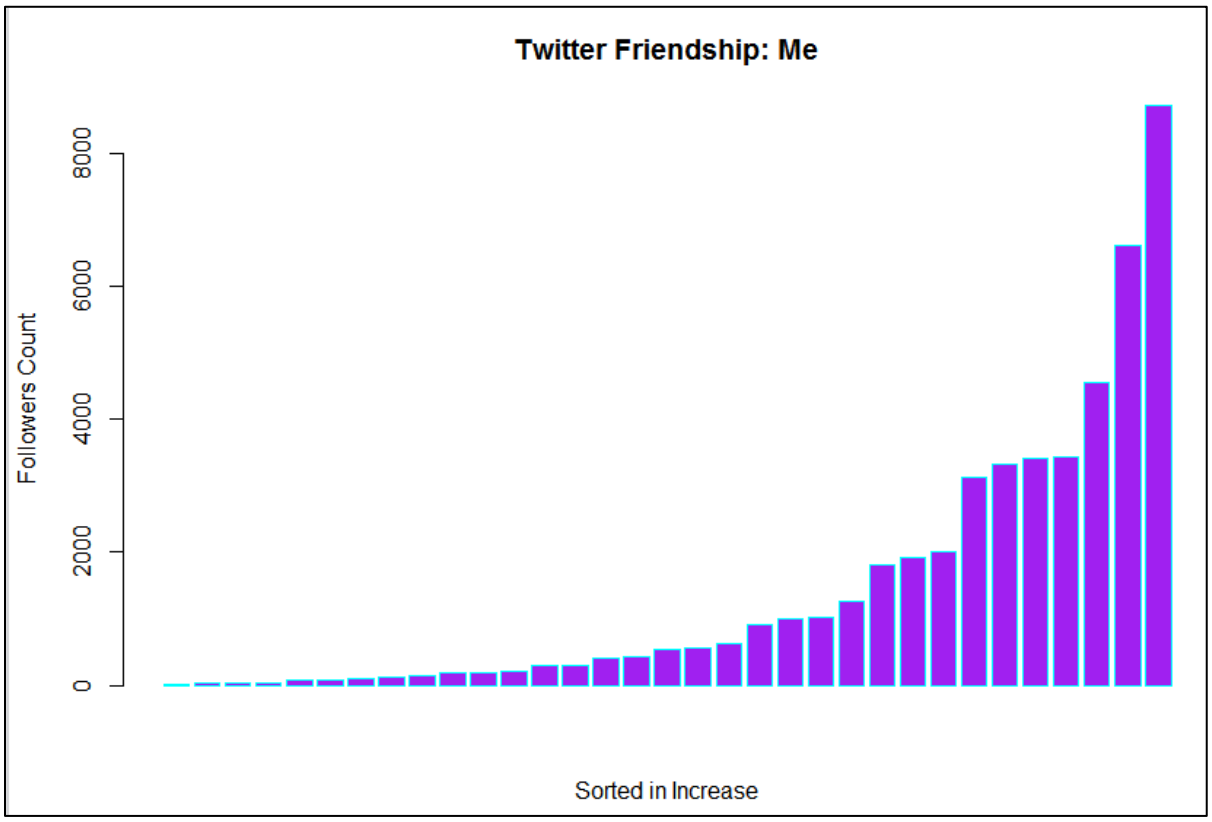
Plot:



The friendship paradox holds true!

Update:

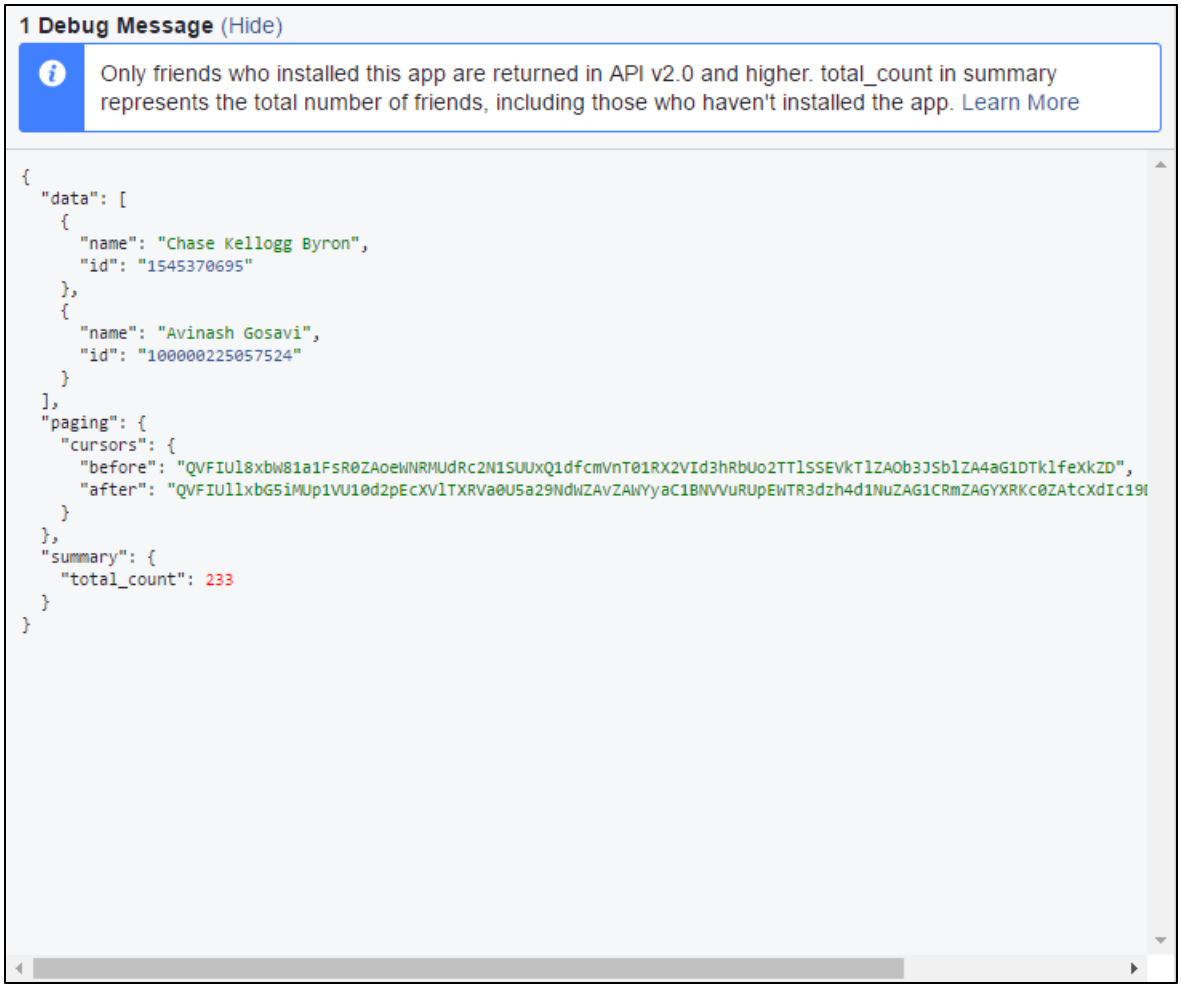
The y-axis was not displaying properly because the number of followers was so high! Here is an example with a reduced list:



Question 3:

Question 4:

To begin with, I registered with Facebook Developers in order to gain access to my user data. There is a tool called "Graph API Explorer," which enabled me to look at the data for my friends. Unfortunately, it only allows me to see the data for my friends who also have the app installed with version 2.0 and up. Interestingly enough, two of my friends do in fact have the app! The summary at the end shows the total amount of friends that I have, regardless if they have the app or not.



cURL code:

```
curl -i -X GET \


"https://graph.facebook.com/v2.8/me/friends?access_token=EAACEdEose0cBADzj3Uw3ZB5XT4ugpS8shm9DaKvzG17JOnMsg1rV8kcuVW8XYWGDTh9q2Fpn9PSRZCkdY8RUTTi6QZBtOQ0lZBtJvaEg3oxbtlpGI58MCd4IUbnivZCMMU6ZCpTsYT0ZBFGXL7rQWZBp18ZAkjnfidPpCMuMZAY6Wp7TjrZAytZA3cZB7Vm6gOGghHKAZD"
```

Chase's Friends:

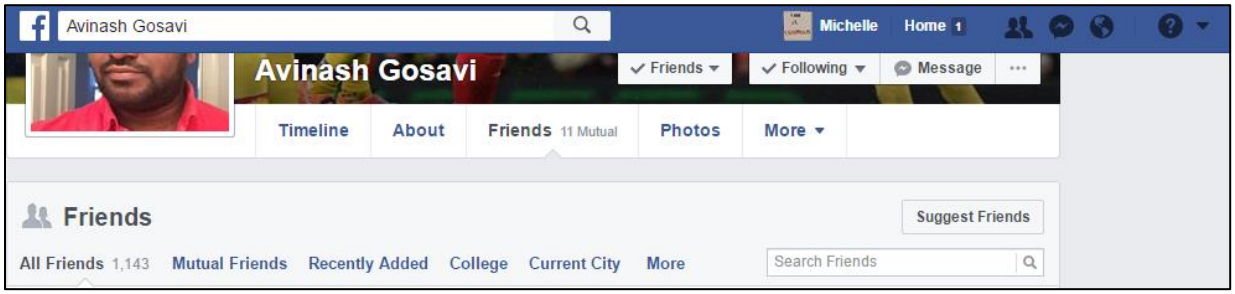
1 Debug Message (Show)


```
{
  "data": [
    {
      "friends": {
        "data": [
          {
            "name": "Michelle Graham",
            "id": "157845274550682"
          }
        ],
        "paging": {
          "cursors": {
            "before": "QVFIUlJWOEZATT214LUhmOUZAVazktTHFzSm04alUwM2JuNWFqWUlXcG9QTmNIbXBackhmbjQ3MlZAFX19IUndYeVdSQ3I
            "after": "QVFIUlJWOEZATT214LUhmOUZAVazktTHFzSm04alUwM2JuNWFqWUlXcG9QTmNIbXBackhmbjQ3MlZAFX19IUndYeVdSQ3B
          }
        },
        "summary": {
          "total_count": 377
        }
      },
      "id": "1545370695"
    },
    {
```


Verification:

Avinash's Friends:

Verification:



Resources:

https://stat.ethz.ch/R-manual/R-devel/library/base/html/rep.html

http://stackoverflow.com/questions/22781685/different-colors-for-each-bar-in-stacked-bar-graph-base-graphics

http://stackoverflow.com/questions/26269665/how-to-convert-data-extracted-from-graphml-to-desired-multiple-columns-in-r

http://hadim.fr/pygraphml/usage.html

http://drumcoder.co.uk/blog/2012/sep/28/tweepy-followers-count/

http://stackoverflow.com/questions/29332259/tweepy-tweeperror-rate-limit-exceeded

https://www.youtube.com/watch?v=WteK95AppF4&feature=youtu.be