Assignment 4

CS 432

Spring 2017
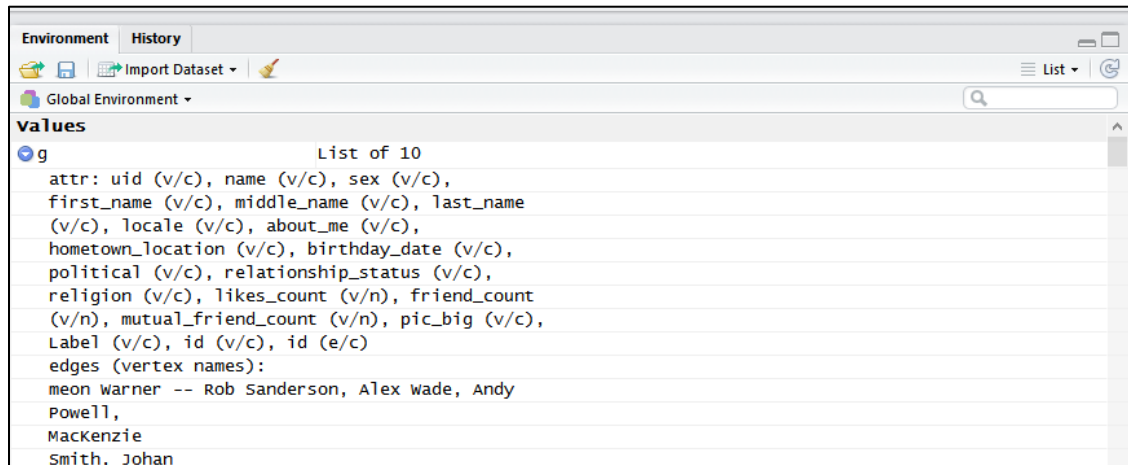
Michelle Graham

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."

```
Environment   History

  Import Dataset                                    List    G
  Global Environment

Values
 g                         List of 10
    attr: uid (v/c), name (v/c), sex (v/c),
    first_name (v/c), middle_name (v/c), last_name
    (v/c), locale (v/c), about_me (v/c),
    hometown_location (v/c), birthday_date (v/c),
    political (v/c), relationship_status (v/c),
    religion (v/c), likes_count (v/n), friend_count
    (v/n), mutual_friend_count (v/n), pic_big (v/c),
    Label (v/c), id (v/c), id (e/c)
    edges (vertex names):
    meon warner -- Rob Sanderson, Alex Wade, Andy
    Powell,
    MacKenzie
    Smith, Johan
```

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")
```


Facebook Friendship: MLN
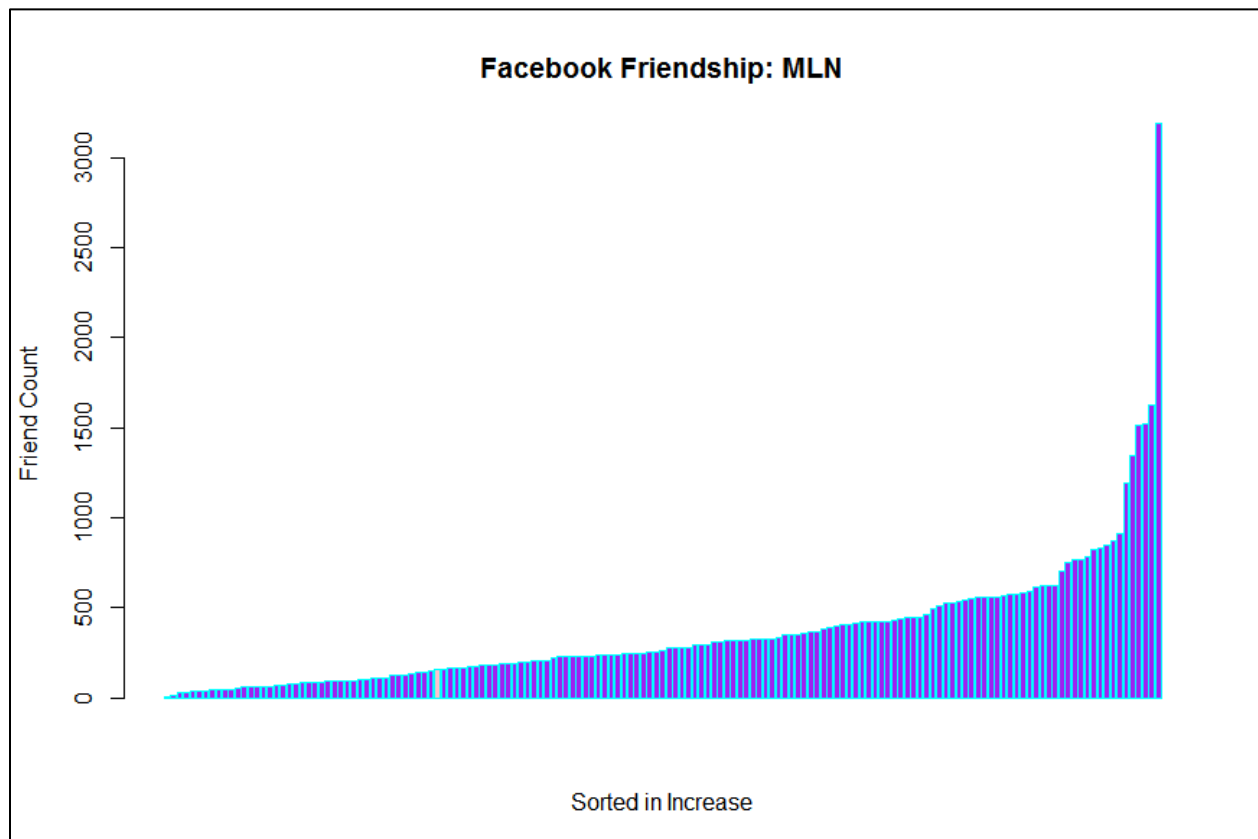
For Nelson's Bar:

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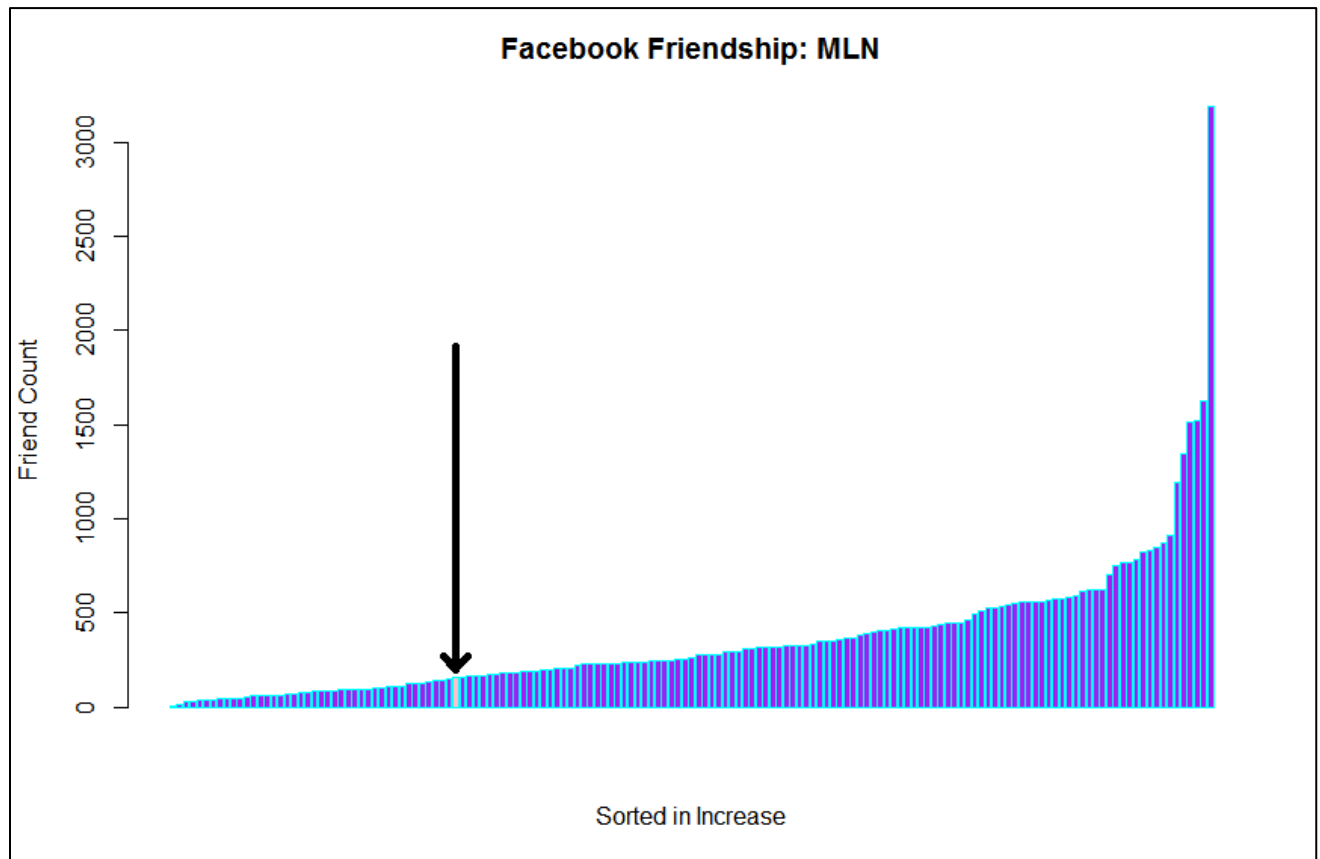
I used Microsoft Excel to calculate the mean, median, and standard deviation for the data excluding the Nelson bar.

Mean:

		fx	=AVERAGE(A2:A155)

	A	B	C	D	E	F	G
1							
2	7						
3	15		Mean	358.987			
4	25		Median	266.5			
5	30		Std. Dev.	371.5853			
6	38						
7	39						
8	40						
9	41						
10	41						
11	42						
12	43						
13	54						
14	58						
15	59						
16	60						
17	62						
18	65						
19	68						

Median:

		fx	=MEDIAN(A2:A155)

A	B	C	D	E	F
7					
15		Mean	358.987		
25		Median	266.5		
30		Std. Dev.	371.5853		
38					
39					
40					
41					

Standard Deviation:

	A	B	C	D	E	F
				=STDEV(A2:A155)		
1						
2	7					
3	15		Mean	358.987		
4	25		Median	266.5		
5	30		Std. Dev.	371.5853		
6	38					
7	39					
8	40					
9	41					
10	41					
11	42					
12	43					

The friendship paradox holds true!

Question 2:

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.

1 Debug Message (Hide)

ℹ️ Only friends who installed this app are returned in API v2.0 and higher. total_count in summary represents the total number of friends, including those who haven't installed the app. Learn More

```
{
  "data": [
    {
      "name": "Chase Kellogg Byron",
      "id": "1545370695"
    },
    {
      "name": "Avinash Gosavi",
      "id": "100000225057524"
    }
  ],
  "paging": {
    "cursors": {
      "before": "QVFIUl8xbW81a1FsR0ZAoeWNRMUdRc2N1SUUxQ1dfcmVnT01RX2VId3hRbUo2TTlSSEVkTlZAOb3JSblZA4aG1DTklfeXkZD",
      "after": "QVFIUllxbG5iMUp1VU10d2pEcXVlTXRVa0U5a29NdWZAvZAWYyaC1BNVVuRUpEWTR3dzh4d1NuZAG1CRmZAGYXRKc0ZAtcXdIc19
    }
  },
  "summary": {
    "total_count": 233
  }
}
```

cURL code:

```
curl -i -X GET \

"https://graph.facebook.com/v2.8/me/friends?access_token=EAACEdEose0cBADzj3Uw3ZB5
XT4ugpS8shm9DaKvzG17JOnMsg1rV8kcuVW8XYWGDTh9q2Fpn9PSRZCkdY8RUTTi6QZBtOQ0lZBtJvaEg
3oxbtlpGI58MCd4IUbnivZCMMU6ZCpTsYT0ZBFGXL7rQWZBp18ZAkjnfidPpCMuMZAY6Wp7TjrZAytZA3
cZB7Vm6gOGghHKAZD"
```

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:

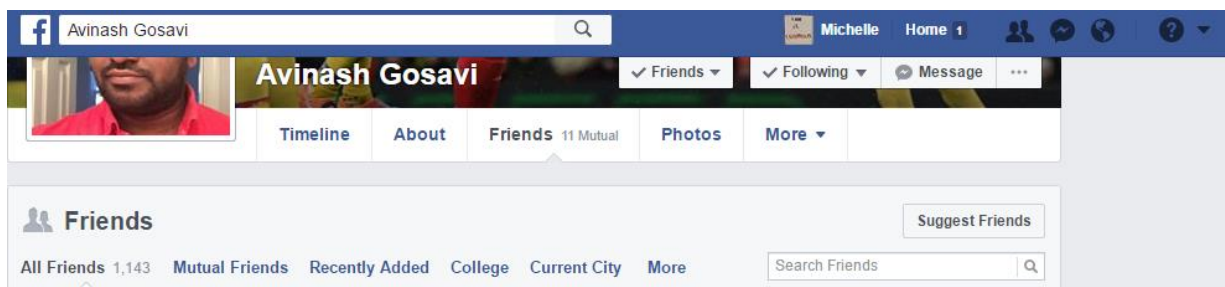## 1 Debug Message (Show)

```
        },
        {
          "name": "Alok Gupta",
          "id": "10202745432349116"
        },
        {
          "name": "Harisankar PS",
          "id": "10204097250784477"
        },
        {
          "name": "Prakash Sejwani",
          "id": "1068181556528232"
        },
        {
          "name": "Madhu Sudhan",
          "id": "848753668471185"
        },
        {
          "name": "Pushpalata Chavhan Patle",
          "id": "1022769107742172"
        }
      ],
      "paging": {
        "cursors": {
          "before": "QVFIUlhWTUg0WTlyWTMyNkZACNnowbjloN0hMcktsaGpfMHdfcnltMERtdF9GZAUdQZA2ZAqWDQ2d2dEaUhMZAEVPZA1B:
          "after": "QVFIUlNlb3YwUmV2Y3AwZAXdSM3pNb08zUGNkQ3lEX2VOTThYd3dNYzJvSDVBRkpDUzBmU1ZAaTEpXWHFNRDJPLVQ0aloz'
        },
        "next": "https://graph.facebook.com/v2.8/100000225057524/friends?access_token=EAACEdEose0cBADzj3Uw3ZB5XT4u|
      },
      "summary": {
        "total_count": 1143
      }
    },
    "id": "100000225057524"
  }
],
```

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

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