

Assignment 4

Tarek Fouda

2016-02-25

1 Introduction

This report explains how I managed to solve Assignment 4 in Web Science class which is due 02/25/2016. It is mainly consisted of Two Questions and two extra credit points questions. Will show my approaches and implementation in each of the two Questions.

2 Problem 1 Getting the Median/Standard Deviation/Mean

2.1 friendship paradox

Basically, there is something called the “friendship paradox” which checks if your friends have more friends than you have. First, I implemented a program that takes the xml file given in the assignment, and extract all the data tags that has a key equals to “Friends count” and put it in a text file named “friendscount.txt”. We were looking to the following tag:

```
<data key="friend_count">
```

The following program is the code that performs this extraction from the xml file:

```
1 def readingxml():
2     i = 0
3     with open('xml.txt', 'r') as file:
4         with open('freindscount.txt', 'w') as f:
5             for line in file:
6                 if '"friend_count"' in line:
7                     i = i + 1
8                     counter = str(int(i))
9                     f.write(line)
```

Listing 1: python code to get friend counts from XML file

After that we get the text file named “friendscount.txt” and we loop through all it’s lines and get the number that exists in each line. That will be the number of friends. The following is the program that performs such a task:

```

1 def extractingnumber():
2     with open('freindscount.txt','r') as f ,open('numbers_count
      .txt ','w') as n:
3         for line in f:
4             temp="".join(re.findall(r'\d+',line
              )) +"\n"
5             n.write(temp)

```

Listing 2: python code to get number in each line

The next step after getting the number of friends is to perform the calculations to get the mean, standard deviation, and median of the number of friends that the professor's friends have.

```

1 import re
2 import math
3 import statistics
4
5
6 def readingxml():
7     i = 0
8     with open('xml.txt', 'r') as file:
9         with open('freindscount.txt','w') as f:
10             for line in file:
11                 if '"friend_count"' in line:
12                     i = i +1
13                     counter = str(int(i))
14                     f.write(line)
15
16
17 def extractingnumber():
18     with open('freindscount.txt','r') as f ,open('numbers_count
      .txt ','w') as n:
19         for line in f:
20             temp="".join(re.findall(r'\d+',line
              )) +"\n"
21             n.write(temp)
22
23
24 def calculateMean():
25     lis=[]
26     with open('numbers_count.txt','r') as nc:
27         total = sum(int(x)
28             for line in nc
29             for x in line.split())
30         #print ("Total = ",total)
31     mean = total /154
32     print ("Mean = ",mean)
33     return mean
34
35
36
37 def calculateSD(var):
38     ls=[]
39     with open('numbers_count.txt','r') as nc:
40         for line in nc:
41             no = int(line) - var

```

```

42         no2= no *no
43         ls.append(no2)
44     sqTotal = sum(ls)
45     Mean= sqTotal/ var
46     sd= math.sqrt(sqTotal)
47     print ("STD = ",sd)
48
49
50
51
52
53
54 def calculateMedian():
55     ls=[]
56     with open ('numbers_count.txt ','r') as n:
57         for line in n:
58             ls.append(line.strip('\n'+'))
59     ls =list(map(int, ls))
60     median = statistics.median(ls)
61     print ("Median = ",median)
62 readingxml()
63 extractingnumber()
64 mean = calculateMean()
65 calculateSD(mean)
66 calculateMedian()

```

Listing 3: python code to get number in each line

The following is the graph, the x-co-ordinates represents the friends number, and the y-co-ordinates represent the number of friends the professor's friends have.

The calculations which were used in the above graph are shown below upon executing the python code:

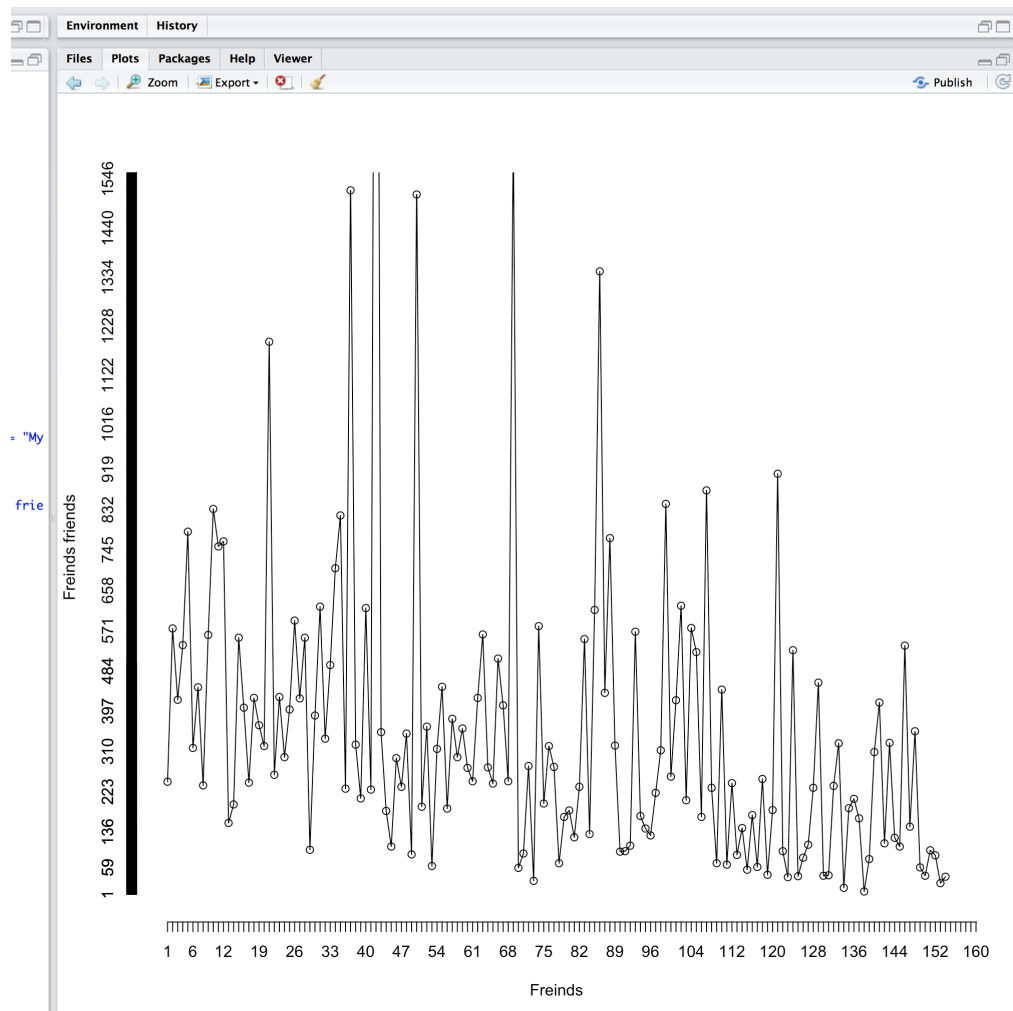


Figure 1: TheGraph

```

Mean = 358.987012987013
STD = 4596.256299862528
Median = 266.5

```

Figure 2: TheGraph

3 Problem number 2

In thisPart of the assignment we were required to get the number of followers my twitter account has and see if the “friendship paradox” will hold, or in other words if my followers have more followers than I have.

So we need to check the number of followers I have, then loop on each one of these followers, and check the number of followers they have and do the calculations the same as part 1.

I used Tweepy Api to get the number of followers.

```
1 import time
2 import tweepy
3
4 consumer_key = "KEtYeXDYJwgdX0IHuwUfIHsw"
5 consumer_secret = "
   xkErPgMDfParcniEkbbhRLgf8T6EXGOfhNP1bAbKSbog5rqCKxi"
6 access_key = "308651543-bWKmgqe2AP3xTx85jyHPBUrovjdMtNej2SOqOjZd"
7 access_secret = "PgKsQJxjvqaocAZmNG2D5t2Q7ZkAcDoPTHvLfOEe2ghj9"
8
9 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
10 auth.set_access_token(access_key, access_secret)
11
12 api = tweepy.API(auth)
13
14 user = api.get_user('amrali70')
15
16 print "Number of Followers I have is " + "{0}".format(user.
   followers_count)
17
18 ids = []
19 i = 0
20 for user in tweepy.Cursor(api.followers, screen_name="amrali70",
   count=1000).items():
21     try:
22         i = i+1
23         name = api.get_user(user.screen_name)
24         with open('detailsfollowers.txt', 'a') as f:
25             line = "Number of Followers " + user.screen_name + " have is
   " + "{0}".format(name.followers_count) + '\n'
26             f.write(line)
27             print "Number of Followers " + user.screen_name + " have is =
   " + "{0}".format(name.followers_count)
28             if ((i%50) == 0):
29                 print i
30                 if (i==299):
31                     time.sleep(60*15)
32             except:
33                 print "waiting"
34                 time.sleep(10)
35 print "Number of Followers" + "{0}".format(i)
```

Listing 4: python code to get number of followers my twitter username and my follower's usernames have

And this is a sample of the output I get on running the above code:

```

Select C:\Windows\system32\cmd.exe - python followers.py

umber of Followers ElwadZEE have is = 24
umber of Followers FyonkaBamby have is = 140
umber of Followers RawdaHisham have is = 2073
umber of Followers SaraMshmsi have is = 1396
umber of Followers moeazab86 have is = 129
umber of Followers YahiaElGamal have is = 350
umber of Followers Mariamh1996 have is = 177
umber of Followers Mennziie have is = 933
umber of Followers yarahassan22 have is = 154
umber of Followers kuliokkkk have is = 12
umber of Followers soahah have is = 357
umber of Followers Monaamr have is = 301
umber of Followers merihanamr have is = 93
umber of Followers NourElDebeiky have is = 1117
umber of Followers RaMoneim have is = 842
umber of Followers CokoCoko1 have is = 967
umber of Followers linasheta have is = 111
umber of Followers tareksamyziad have is = 0
umber of Followers Farrah_Muhamed have is = 2802
umber of Followers yasmina_mb have is = 24
umber of Followers DonnaMGouda have is = 384
umber of Followers MernaEmad9 have is = 1233
umber of Followers NadaMuhmed have is = 292
umber of Followers NourMohamed_ have is = 347
umber of Followers MaramAmgad have is = 133
umber of Followers lsamar7041 have is = 168
umber of Followers Shrouk_ElMaghry have is = 765
umber of Followers y_yarra have is = 658
umber of Followers SadiyahFazal have is = 2383
umber of Followers RanaDarwiish have is = 208
umber of Followers amraraby52 have is = 402
umber of Followers antoninojagger have is = 319
umber of Followers cherryelgende have is = 910
umber of Followers Dewiidar have is = 292
umber of Followers no_tmimi have is = 319
umber of Followers _MariamNour have is = 909
umber of Followers SamaaSalamaa have is = 157
umber of Followers ShahdAbaza have is = 860
umber of Followers mahmoudgallall have is = 18
umber of Followers Nadeenazazy have is = 822
umber of Followers Klyzuuuhh have is = 640
umber of Followers jasminexxos have is = 1094
umber of Followers romanamoussa have is = 140
umber of Followers Nadaaomarr have is = 761
umber of Followers MollieMcGuirex have is = 587
umber of Followers SheryhanKilany_ have is = 910
umber of Followers sherinesameeh have is = 85
umber of Followers NurhanTurky have is = 534
umber of Followers Hosam_Mansour have is = 383
umber of Followers Dina_fadaly have is = 1754
umber of Followers LostSoul91 have is = 158
umber of Followers JosephFathy have is = 31
umber of Followers SaRaHaHmEdMoUsA have is = 266
umber of Followers MirnaOsama7 have is = 675
umber of Followers TukaTarek_ have is = 937
umber of Followers Farhota have is = 1057
umber of Followers Nada_Saad have is = 833

```

Figure 3: The followers and the number of followers they have

```
Mean = 85428122.98051947
STD = 9459837280.32109
Median = 563.0
```

Figure 4: The followers and the number of followers they have

I printed the output in a text file and by executing the same methods and techniques in 2, I managed to calculate the mean, median and standard deviation. These were the calculations and the graph:

So the x-axis has the number of friends I have, and the Y-axis has the number of friends each follower have.

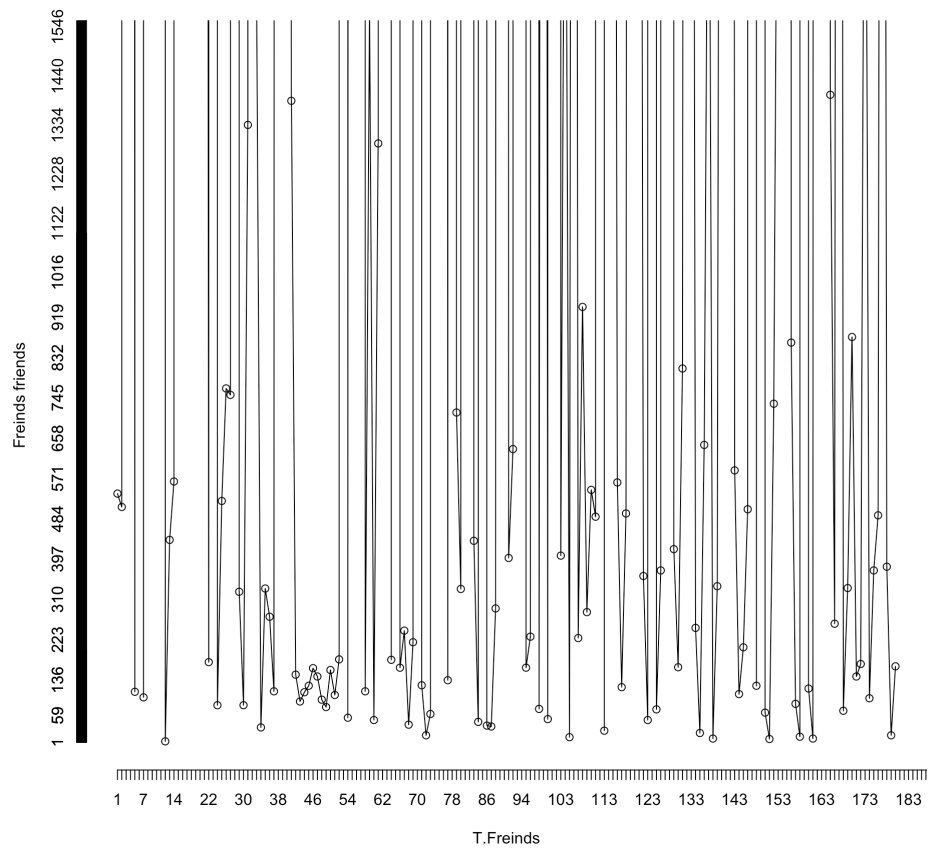


Figure 5: The followers and the number of followers they have