ASSIGNMENT 10

CS 432 Spring 2017

MAY 1, 2017

Michelle Graham

Q1:

Most of the code for this question was provided in the slides and the PCI book. The cosine function was easily found on Stack Overflow. I ran the urls.txt file through genfeedvector.py to generate a blog matrix for 1000 terms. Once that is completed, I used that to find the distance between vector word counts by using a stop list filter and knnestimatedata.

Code Snippet**:**

```
import math
import operator
count=0
outfile = open ("kValues.txt",'wb')
def dot_product(v1, v2):
    return sum(map(operator.mul, v1, v2))


def vector_cos(v1, v2):
    prod = dot_product(v1, v2)
    len1 = math.sqrt(dot_product(v1, v1))
    len2 = math.sqrt(dot_product(v2, v2))
    return prod / (len1 * len2)


def getdistances(data, vec1):
    distancelist = []

    # Loop over every item in the dataset
    for i in range(len(data)):
        vec2 = data[i]

        # Add the distance and the index
        distancelist.append((vector_cos(vec1, vec2), i))
```

Knn Estimates:

fmeasure
19.0
57.5
51.6
52.4
57.45
Web Science
22.0
25.0
38.0
42.5
43.6

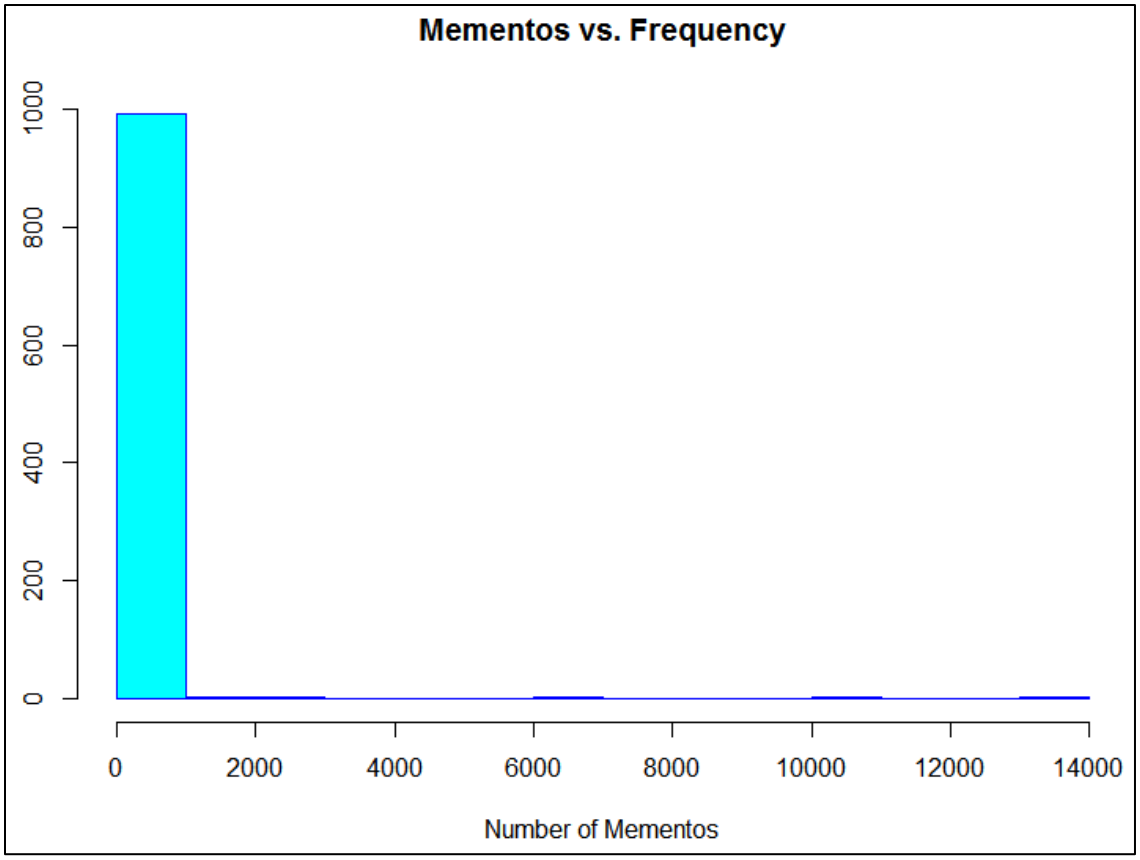
Q2:

I spent several days trying to install the libsvm library installed properly on my laptop. I tried to go by recommendations by other students but just didn't have any luck.

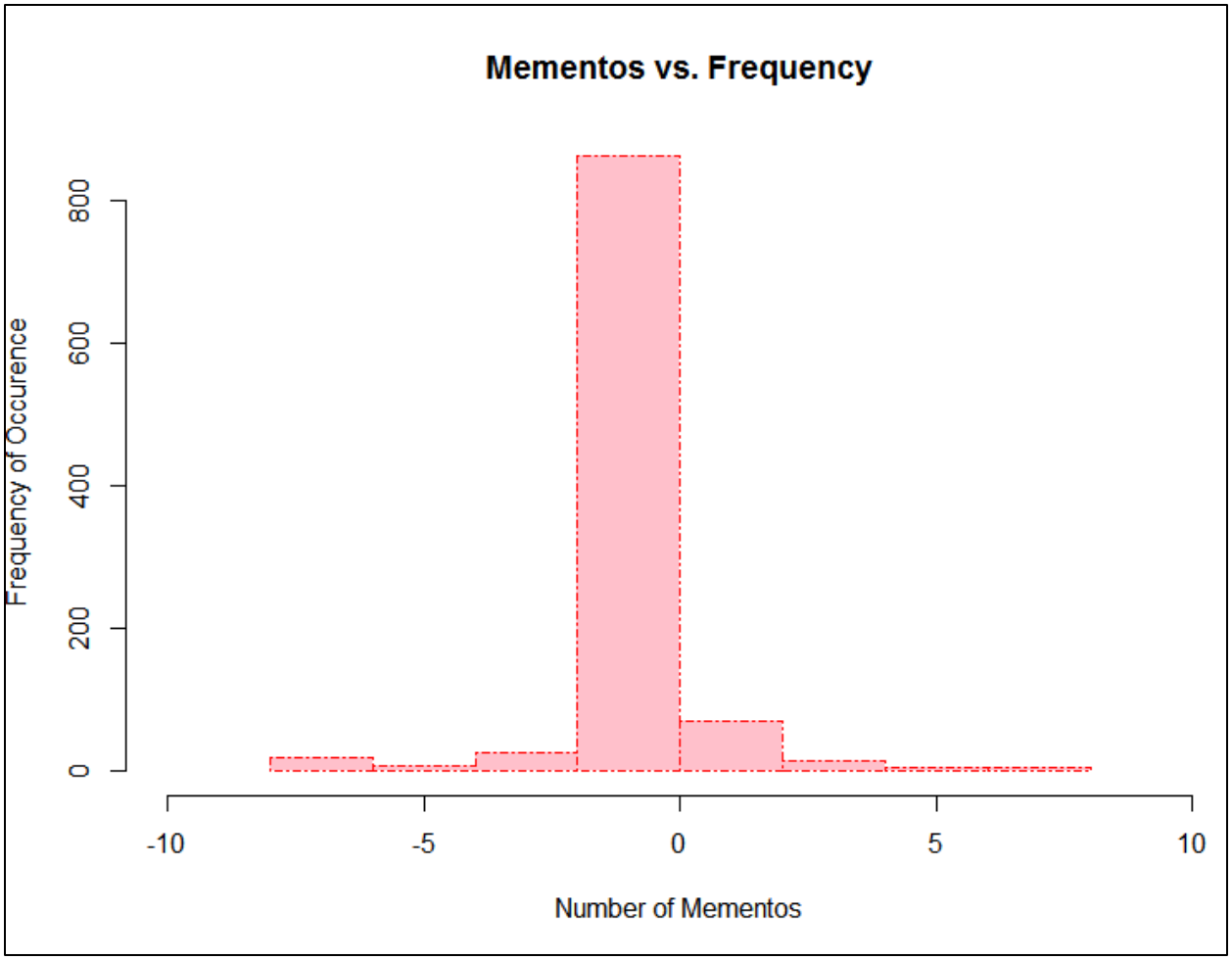
Q3:

Because I was unable to complete this assignment earlier in the semester, I got help from a classmate so that I could evaluate their code. I used their data to generate mementoes of each 1000 unique links. I compared the data to their original graphed data from assignment 2. This was done by use of the histogram function in r studio.

```
oldData <- read.table("C:/CS432/A10/Q3/finalMementoList10")
hist(histData$V1, xlab='Number of Mementos',
     ylab='Frequency of Occurence', main='Mementos vs. Frequency',
     ty =4, col ='cyan', border = 'blue')
```




```
histData <- read.table("C:/CS432/A10/Q3/differenceTimeMaps")
hist(histData$V1, xlab='Number of Mementos', xlim = range(c(-10,10)),
     lab='Frequency of Occurence', main='Mementos vs. Frequency',
     ty =4, col ='pink', border = 'red')
```


Mementos vs. Frequency
Frequency of Occurence
0
200
400
600
800
-10
-5
0
5
10
Number of Mementos

Q4:

Resources:

https://github.com/rreelachart/cs532-s17/tree/master/submissions/assignment%202