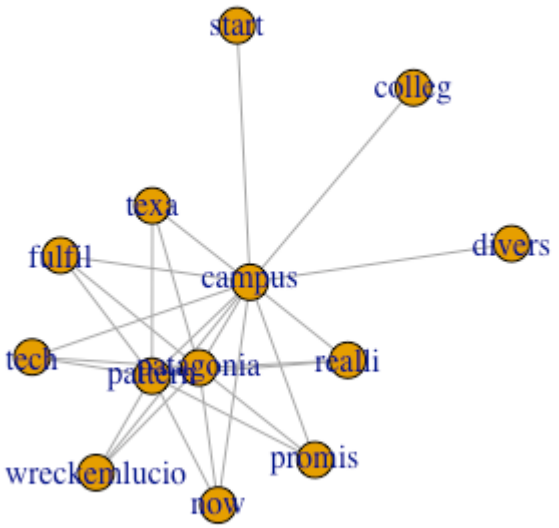
Association Mining of Twitter Data- Campus Diversity Data Set

Question #1

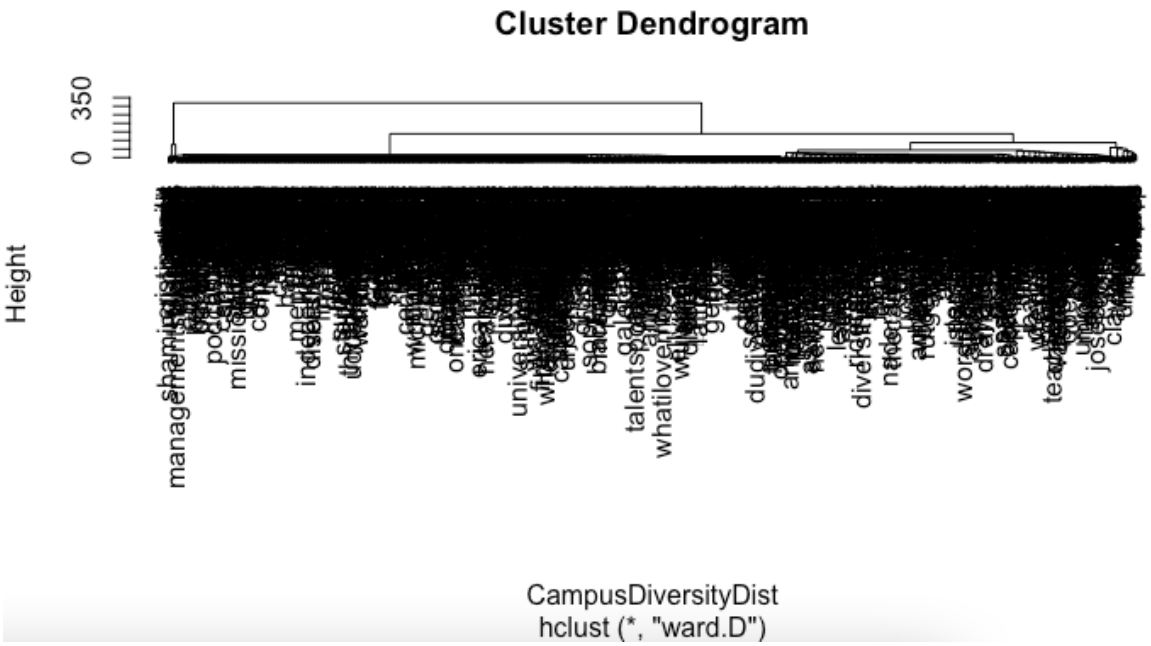
Plot1




```
> findAssocs(CampusDiversityTDM, 'campus', 0.20)
$campus
      fulfil    patagonia      pattern       promis         tech
        0.27         0.27         0.27         0.27         0.27
        texa wreckemlucio       divers        start      realli
        0.27         0.26         0.25         0.23         0.23
         now       colleg
        0.20         0.20
```

Question #2

Plot 2



Code:

```
library(readxl)
library(cluster)
library(igraph)
library(tm)
library(topicmodels)
library(RCA)
library*SnowballC::
library(SnowballC)
library(syuzhet)

CampusDiversity <- read_excel("Exercise Analyzing the Structure of a Twitter
Conversation/CampusDiversity.xlsx")
#View(CampusDiversity)
#making dataframe
CampusDiversityCorp=Corpus(VectorSource(CampusDiversity$...17))

#preprocesscorpus
CampusDiversityCorp=tm_map(CampusDiversityCorp,content_transformer(tolower))
CampusDiversityCorp=tm_map(CampusDiversityCorp, removeWords, stopwords('english'))
CampusDiversityCorp=tm_map(CampusDiversityCorp, removePunctuation)
CampusDiversityCorp=tm_map(CampusDiversityCorp, removeNumbers)
CampusDiversityCorp=tm_map(CampusDiversityCorp, stripWhitespace)
```

```
View(as.matrix(TermDocumentMatrix(CampusDiversityCorp)))
StripString=content_transformer(function(x,pattern) gsub(pattern,'',x))
CampusDiversityCorp=tm_map(CampusDiversityCorp, StripString, 'http[[:alnum:]]*')
CampusDiversityCorp=tm_map(CampusDiversityCorp,StripString,'[\r\n]')
CampusDiversityCorp=tm_map(CampusDiversityCorp,StripString,'[\t]')



CampusDiversityCorp=tm_map(CampusDiversityCorp, stemDocument)
CampusDiversityTDM=TermDocumentMatrix(CampusDiversityCorp)
CampusDiversityDTM=DocumentTermMatrix(CampusDiversityCorp)

View(data.frame(as.matrix(CampusDiversityTDM)))

CampusDiversityfreq=colSums(as.matrix(CampusDiversityDTM))
CampusDiversityfreq=sort(CampusDiversityfreq, decreasing=T)
cbind(CampusDiversityfreq)


findAssocs(CampusDiversityTDM, 'campus', 0.20)


#not sure
term='campus'
anet1=as.data.frame(findAssocs(CampusDiversityTDM, term, 0.20))
anet1=cbind(term,rownames(anet1),anet1)
colnames(anet1)=c('word1','word2','freq')
rownames(anet1)=NULL
anet2=anet1

term='pattern'
anet1=as.data.frame(findAssocs(CampusDiversityTDM, term, 0.20))
anet1=cbind(term,rownames(anet1),anet1)
colnames(anet1)=c('word1','word2','freq')
rownames(anet1)=NULL

anet2=as.data.frame(rbind(anet1,anet2))

term='patagonia'
anet1=as.data.frame(findAssocs(CampusDiversityTDM, term, 0.20))
anet1=cbind(term,rownames(anet1),anet1)
colnames(anet1)=c('word1','word2','freq')
rownames(anet1)=NULL
anet2=as.data.frame(rbind(anet1,anet2))

#thisgavemeplot1
CampusDiversityGraph=graph_from_data_frame(anet2[1:2], directed=F)
```

```
CampusDiversityGraph=simplify(CampusDiversityGraph)
plot(CampusDiversityGraph)


CampusDiversityDist=dist(CampusDiversityTDM,method='euclidean')
CampusDiversityClust=hclust(d=CampusDiversityDist, method='ward.D')

findFreqTerms(CampusDiversityTDM, 200)

library(Rgraphviz)
plot(CampusDiversityClust)
rect.hclust(CampusDiversityClust, k=2, border='red2')

CampusDiversityTopics=cutree(CampusDiversityClust,k=6)


#plot(CampusDiversityClust, yaxt='n', xlab='', ylab='', hang=1,main='', sub='', cex=1.75)

TermDocumentMatrix()


#findAssocs(CampusDiversity, 'Portland', 0.20)

#datacleansing

#creating TDM
CampusDiversityCooccurMatrix=as.matrix(CampusDiversity) %*%
  t(as.matrix(CampusDiversity))
View(as.matrix(TermDocumentMatrix(CampusDiversity)))
CampusDiversityTDM=TermDocumentMatrix(CampusDiversity)
View(data.frame(as.matrix(CampusDiversityTDM)))
#ReyDTM=DocumentTermMatrix(ReyCorp)

#creating edge list
term='portland'
anet1=as.data.frame(findAssocs(ATDM, term, 0.20))
anet1=cbind(term,rownames(anet1),anet1)
colnames(anet1)=c('word1','word2','freq')
rownames(anet1)=NULL


CampusDiversityDist=dist(CampusDiversity,method='euclidean')
```