Homework 6

Yiwei He (yh9vhg), Da Lin (dl2de), Ziyue Jin (zj5qj)

Question 1:

**2003-2005**  **2006-2008**

Text, letter

Description automatically generated Text, letter

Description automatically generated

**2009-2012 From the world cloud we generated,**

Text, letter

Description automatically generated **we found words like “model”, “method”, “data”, “function”, and “use” are always popular from 2003 to 2012. However, there are more words that are popular only for a period of time, such as “estim”. It didn’t show up during 2003-2005 but consistently being popular from 2006-2012. Also, the word “distribut” was popular from 2003-2008, but become less popular during 2009-2012.**

Question 2

**Peter J Bickel has 8 collaborators**

**Jianqing Fan has 38 collaborators**

**David Dunson has 32 collaborators**

**Therefore, Jianqing Fan has the most collaborators.**

Chart, line chart

Description automatically generatedQuestion 3: **We pick Peter J Bickel as the author to analyze**

a) **We discovered that the Peter J Bickel’s network grow with time. More people work with him in 2009-2012 than 2003-2005.**

Text, letter

Description automatically generatedText, letter

Description automatically generatedText, letter

Description automatically generatedb)  **2003-2005 2006-2008 2009-2012**

**We discovered that the world cloud from Peter J Bickel’s articles is similar to the global trend as both of them has word such as “estim”, “model”, “method”. However, the trend is not same in details. For example, the word “use” from Peter’s articles became less popular during 2009-2012 but globally its popularity is consistent for all time.**

c) **2003-2012 whole career**

**No, the period from 2003-2012 does not match his whole academic career. For example, the word “ test” and “estim” are very popular in his whole academic career, but during 2003-2012, these two words seems not have similar popularity. However, words like “statist” and “analysi” have consistent popularity.**

Text

Description automatically generatedText

Description automatically generated with low confidence

**Appendix**

papers1 <- scan("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/paperList.txt",sep="\n",what="")

papers <- scan("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/paperList\_Abstracts\_Keyword.txt",sep="\n",what="")

#### notice that items in the two data sets match in their order

papers <- papers[-1]

#### So papers1 contains the papers by title and time, papers contains papers by abstract and time

dt <- scan("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/paperList\_Abstracts\_Keyword.txt",what="",sep="\n")

dt <- dt[-1]

### now extract year

tmp1 <- lapply(dt,function(s)unlist(strsplit(s,'"')))

years <- unlist(lapply(tmp1,function(x) return(x[3])))

abss <- unlist(lapply(tmp1,function(x) return(x[4])))

year1=list()

year2=list()

year3=list()

for(k in 1:length(tmp1)){

if(tmp1[[k]][3]==",2003," | tmp1[[k]][3]==",2004," | tmp1[[k]][3]==",2005,")

{year1=append(year1,list(lapply(tmp1[[k]],function(s)unlist(strsplit(s,'"')))))}

else if(tmp1[[k]][3]==",2006," | tmp1[[k]][3]==",2007," | tmp1[[k]][3]==",2008,")

{year2=append(year2,list(lapply(tmp1[[k]],function(s)unlist(strsplit(s,'"')))))}

else{year3=append(year3,list(lapply(tmp1[[k]],function(s)unlist(strsplit(s,'"')))))}}

abss1 <- unlist(lapply(year1,function(x) return(x[4])))

abss2 <- unlist(lapply(year2,function(x) return(x[4])))

abss3 <- unlist(lapply(year3,function(x) return(x[4])))

abss1 <- gsub('\\{','',abss1)

abss1 <- gsub("}","",abss1)

abss1 <- gsub("#","",abss1)

abss1 <- gsub("&","",abss1)

abss1 <- gsub("<","",abss1)

abss1 <- gsub(">","",abss1)

abss1 <- gsub("/","",abss1)

abss1 <- gsub(")","",abss1)

abss1 <- gsub("\\(","",abss1)

abss1 <- gsub("amp","",abss1)

abss2 <- gsub('\\{','',abss2)

abss2 <- gsub("}","",abss2)

abss2 <- gsub("#","",abss2)

abss2 <- gsub("&","",abss2)

abss2 <- gsub("<","",abss2)

abss2 <- gsub(">","",abss2)

abss2 <- gsub("/","",abss2)

abss2 <- gsub(")","",abss2)

abss2 <- gsub("\\(","",abss2)

abss2 <- gsub("amp","",abss2)

abss3 <- gsub('\\{','',abss3)

abss3 <- gsub("}","",abss3)

abss3 <- gsub("#","",abss3)

abss3 <- gsub("&","",abss3)

abss3 <- gsub("<","",abss3)

abss3 <- gsub(">","",abss3)

abss3 <- gsub("/","",abss3)

abss3 <- gsub(")","",abss3)

abss3 <- gsub("\\(","",abss3)

abss3 <- gsub("amp","",abss3)

for(k in 1:length(abss1)){

write(abss1[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder1/",k,sep=""))

}

for(k in 1:length(abss2)){

write(abss2[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder2/",k,sep=""))

}

for(k in 1:length(abss3)){

write(abss3[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder3/",k,sep=""))

}

library(tm)

library(SnowballC)

cname1 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder1"

cname2 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder2"

cname3 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder3"

docs1 <- Corpus(DirSource(cname1))

meta(docs1,"id")

document.names1 <- unlist(meta(docs1,"id"))

docs2 <- Corpus(DirSource(cname2))

meta(docs2,"id")

document.names2 <- unlist(meta(docs2,"id"))

docs3 <- Corpus(DirSource(cname3))

meta(docs3,"id")

document.names3 <- unlist(meta(docs3,"id"))

docs1=tm\_map(docs1, removePunctuation)

docs2=tm\_map(docs2, removePunctuation)

docs3=tm\_map(docs3, removePunctuation)

docs1 <- tm\_map(docs1, removeNumbers)

docs2 <- tm\_map(docs2, removeNumbers)

docs3 <- tm\_map(docs3, removeNumbers)

docs1 <- tm\_map(docs1, tolower)

docs2 <- tm\_map(docs2, tolower)

docs3 <- tm\_map(docs3, tolower)

docs1 <- tm\_map(docs1, removeWords, stopwords("english"))

docs2 <- tm\_map(docs2, removeWords, stopwords("english"))

docs3 <- tm\_map(docs3, removeWords, stopwords("english"))

docs1 <- tm\_map(docs1, stemDocument)

docs2 <- tm\_map(docs2, stemDocument)

docs3 <- tm\_map(docs3, stemDocument)

docs1 <- tm\_map(docs1, stripWhitespace)

docs2 <- tm\_map(docs2, stripWhitespace)

docs3 <- tm\_map(docs3, stripWhitespace)

dtm1 <- DocumentTermMatrix(docs1)

dtm2 <- DocumentTermMatrix(docs2)

dtm3 <- DocumentTermMatrix(docs3)

tdm1 <- TermDocumentMatrix(docs1)

tdm2 <- TermDocumentMatrix(docs2)

tdm3 <- TermDocumentMatrix(docs3)

freq1 <- colSums(as.matrix(dtm1))

freq2 <- colSums(as.matrix(dtm2))

freq3 <- colSums(as.matrix(dtm3))

dtms1 <- removeSparseTerms(dtm1, 0.995)

dtms2 <- removeSparseTerms(dtm2, 0.995)

dtms3 <- removeSparseTerms(dtm3, 0.995)

DTM1 <- as.matrix(dtms1)

DTM2 <- as.matrix(dtms2)

DTM3 <- as.matrix(dtms3)

doc.index1 <- as.numeric(document.names1)

doc.index2 <- as.numeric(document.names2)

doc.index3 <- as.numeric(document.names3)

final.DTM1 <- matrix(0,nrow=nrow(DTM1),ncol=ncol(DTM1))

final.DTM1[doc.index1,] <- DTM1

final.DTM2 <- matrix(0,nrow=nrow(DTM2),ncol=ncol(DTM2))

final.DTM2[doc.index2,] <- DTM2

final.DTM3 <- matrix(0,nrow=nrow(DTM3),ncol=ncol(DTM3))

final.DTM3[doc.index3,] <- DTM3

colnames(final.DTM1) <- colnames(DTM1)

colnames(final.DTM2) <- colnames(DTM2)

colnames(final.DTM3) <- colnames(DTM3)

library(wordcloud)

TF1 <- final.DTM1/(0.5+rowSums(final.DTM1))

iDF1 <- log(nrow(final.DTM1)/colSums(final.DTM1>0))

TFiDF1 <- t(t(TF1) \* iDF1)

term.TFiDF1 <- colSums(TFiDF1)

ix1 <- sort(term.TFiDF1,decreasing=TRUE,index.return=TRUE)$ix

TF2 <- final.DTM2/(0.5+rowSums(final.DTM2))

iDF2 <- log(nrow(final.DTM2)/colSums(final.DTM2>0))

TFiDF2 <- t(t(TF2) \* iDF2)

term.TFiDF2 <- colSums(TFiDF2)

ix2 <- sort(term.TFiDF2,decreasing=TRUE,index.return=TRUE)$ix

TF3 <- final.DTM3/(0.5+rowSums(final.DTM3))

iDF3 <- log(nrow(final.DTM3)/colSums(final.DTM3>0))

TFiDF3 <- t(t(TF3) \* iDF3)

term.TFiDF3 <- colSums(TFiDF3)

ix3 <- sort(term.TFiDF3,decreasing=TRUE,index.return=TRUE)$ix

final.DTM1 <- final.DTM1[,ix1[2:501]]

final.DTM2 <- final.DTM2[,ix2[2:501]]

final.DTM3 <- final.DTM3[,ix3[2:501]]

wordcloud(

words = colnames(final.DTM1),

freq = colSums(final.DTM1),

scale = c(3, 0.25),min.freq = 200)

wordcloud(

words = colnames(final.DTM2),

freq = colSums(final.DTM2),

scale = c(3, 0.25))

wordcloud(

words = colnames(final.DTM3),

freq = colSums(final.DTM3),

scale = c(3, 0.25))

################################################################################

A2P <- read.table("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/authorPaperBiadj.txt")

authors <- read.table("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/authorList.txt",stringsAsFactors=FALSE)

PJBindex=which(authors$V1=="Peter J Bickel")

JQFindex=which(authors$V1=="Jianqing Fan")

DDindex=which(authors$V1=="David Dunson")

PJB=A2P[,which(A2P[PJBindex,] == 1)]

PJBnum=rowSums(PJB)

PJBnum[PJBnum>0]=1

PJBco=sum(PJBnum)-1

JQF=A2P[,which(A2P[JQFindex,] == 1)]

JQFnum=rowSums(JQF)

JQFnum[JQFnum>0]=1

JQFco=sum(JQFnum)-1

DD=A2P[,which(A2P[DDindex,] == 1)]

DDnum=rowSums(DD)

DDnum[DDnum>0]=1

DDco=sum(DDnum)-1

PJBco

JQFco

DDco

################################################################################

#3a)

library(igraph)

PJB\_papers\_id=which(A2P[2490,] >0)

PJB\_papers\_id=names(unlist(A2P[2490, PJB\_papers\_id]))

PJB\_papers\_id=gsub('V','',PJB\_papers\_id)

PJB\_papers\_id=as.numeric(PJB\_papers\_id)

PJB\_papers=dt[PJB\_papers\_id]

tmpPJB <- lapply(PJB\_papers,function(s)unlist(strsplit(s,'"')))

yearsPJB <- unlist(lapply(tmpPJB,function(x) return(x[3])))

yearsPJB <- gsub(",","",yearsPJB)

yearsPJB <- as.numeric(yearsPJB)

A2P=as.matrix(A2P)

par(mfrow = c(1, 3))

year\_list=c(2003, 2006, 2009, 2013)

label=c('2003-2005', '2006-2008', '2009-2012')

for (i in c(1:3)) {

year=PJB\_papers\_id[which(yearsPJB >= year\_list[i] & yearsPJB < year\_list[i+1])]

A2P\_sub=A2P[, year]

Coauthor.Adj <- A2P\_sub%\*%t(A2P\_sub)

diag(Coauthor.Adj) <- 0

g <- graph.adjacency(Coauthor.Adj, mode="undirected")

cl <- clusters(g)

lcc <- which(cl$membership==which.max(cl$csize))

A.lcc <- Coauthor.Adj[lcc, lcc]

g.lcc <- graph.adjacency(A.lcc, mode="undirected")

deg <- degree(g.lcc, mode="all")

Coauthor.Adj[Coauthor.Adj > 0] <- 1

g <- graph.adjacency(Coauthor.Adj, mode="undirected")

cl <- clusters(g)

lcc <- which(cl$membership==which.max(cl$csize))

A.lcc <- Coauthor.Adj[lcc, lcc]

g.lcc <- graph.adjacency(A.lcc, mode="undirected")

inc.edges <- incident(g.lcc, V(g.lcc)[names(V(g.lcc)) == "Peter\_J\_Bickel"], mode = 'all')

ecol <- rep("black", ecount(g.lcc))

ecol[as.numeric(inc.edges)] <- "green"

elty <- rep(2, ecount(g.lcc))

elty[as.numeric(inc.edges)] <- 1

V(g.lcc)$size <- 2.5 + deg\*3

V(g.lcc)$frame.color <- "blue"

vcol <- rep("yellow", vcount(g.lcc))

vcol[names(V(g.lcc)) == "Peter\_J\_Bickel"] <- "red"

lo <- layout\_with\_graphopt(g.lcc)

plot(g.lcc, edge.width = 1, layout = lo, vertex.color=vcol,

edge.color=ecol, edge.lty = elty,

main = label[i])

}

################################################################################

#3b)

PJB\_papers1=year1[PJB\_papers\_id]

PJB\_papers2=year2[PJB\_papers\_id]

PJB\_papers3=year3[PJB\_papers\_id]

abss4 <- unlist(lapply(PJB\_papers1,function(x) return(x[4])))

abss5 <- unlist(lapply(PJB\_papers2,function(x) return(x[4])))

abss6 <- unlist(lapply(PJB\_papers3,function(x) return(x[4])))

abss4 <- gsub('\\{','',abss4)

abss4 <- gsub("}","",abss4)

abss4 <- gsub("#","",abss4)

abss4 <- gsub("&","",abss4)

abss4 <- gsub("<","",abss4)

abss4 <- gsub(">","",abss4)

abss4 <- gsub("/","",abss4)

abss4 <- gsub(")","",abss4)

abss4 <- gsub("\\(","",abss4)

abss4 <- gsub("amp","",abss4)

abss5 <- gsub('\\{','',abss5)

abss5 <- gsub("}","",abss5)

abss5 <- gsub("#","",abss5)

abss5 <- gsub("&","",abss5)

abss5 <- gsub("<","",abss5)

abss5 <- gsub(">","",abss5)

abss5 <- gsub("/","",abss5)

abss5 <- gsub(")","",abss5)

abss5 <- gsub("\\(","",abss5)

abss5 <- gsub("amp","",abss5)

abss6 <- gsub('\\{','',abss6)

abss6 <- gsub("}","",abss6)

abss6 <- gsub("#","",abss6)

abss6 <- gsub("&","",abss6)

abss6 <- gsub("<","",abss6)

abss6 <- gsub(">","",abss6)

abss6 <- gsub("/","",abss6)

abss6 <- gsub(")","",abss6)

abss6 <- gsub("\\(","",abss6)

abss6 <- gsub("amp","",abss6)

for(k in 1:length(abss4)){

write(abss4[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder4/",k,sep=""))

}

for(k in 1:length(abss5)){

write(abss5[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder5/",k,sep=""))

}

for(k in 1:length(abss6)){

write(abss6[k],file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder6/",k,sep=""))

}

library(tm)

library(SnowballC)

cname4 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder4"

cname5 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder5"

cname6 <- "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/TextFolder6"

docs4 <- Corpus(DirSource(cname4))

meta(docs4,"id")

document.names4 <- unlist(meta(docs4,"id"))

docs5 <- Corpus(DirSource(cname5))

meta(docs5,"id")

document.names5 <- unlist(meta(docs5,"id"))

docs6 <- Corpus(DirSource(cname6))

meta(docs6,"id")

document.names6 <- unlist(meta(docs6,"id"))

docs4=tm\_map(docs4, removePunctuation)

docs5=tm\_map(docs5, removePunctuation)

docs6=tm\_map(docs6, removePunctuation)

docs4 <- tm\_map(docs4, removeNumbers)

docs5 <- tm\_map(docs5, removeNumbers)

docs6 <- tm\_map(docs6, removeNumbers)

docs4 <- tm\_map(docs4, tolower)

docs5 <- tm\_map(docs5, tolower)

docs6 <- tm\_map(docs6, tolower)

docs4 <- tm\_map(docs4, removeWords, stopwords("english"))

docs5 <- tm\_map(docs5, removeWords, stopwords("english"))

docs6 <- tm\_map(docs6, removeWords, stopwords("english"))

docs4 <- tm\_map(docs4, stemDocument)

docs5 <- tm\_map(docs5, stemDocument)

docs6 <- tm\_map(docs6, stemDocument)

docs4 <- tm\_map(docs4, stripWhitespace)

docs5 <- tm\_map(docs5, stripWhitespace)

docs6 <- tm\_map(docs6, stripWhitespace)

dtm4 <- DocumentTermMatrix(docs4)

dtm5 <- DocumentTermMatrix(docs5)

dtm6 <- DocumentTermMatrix(docs6)

tdm4 <- TermDocumentMatrix(docs4)

tdm5 <- TermDocumentMatrix(docs5)

tdm6 <- TermDocumentMatrix(docs6)

freq4 <- colSums(as.matrix(dtm4))

freq5 <- colSums(as.matrix(dtm5))

freq6 <- colSums(as.matrix(dtm6))

dtms4 <- removeSparseTerms(dtm4, 0.995)

dtms5 <- removeSparseTerms(dtm5, 0.995)

dtms6 <- removeSparseTerms(dtm6, 0.995)

DTM4 <- as.matrix(dtms4)

DTM5 <- as.matrix(dtms5)

DTM6 <- as.matrix(dtms6)

doc.index4 <- as.numeric(document.names4)

doc.index5 <- as.numeric(document.names5)

doc.index6 <- as.numeric(document.names6)

final.DTM4 <- matrix(0,nrow=nrow(DTM4),ncol=ncol(DTM4))

final.DTM4[doc.index4,] <- DTM4

final.DTM5 <- matrix(0,nrow=nrow(DTM5),ncol=ncol(DTM5))

final.DTM5[doc.index5,] <- DTM5

final.DTM6 <- matrix(0,nrow=nrow(DTM6),ncol=ncol(DTM6))

final.DTM6[doc.index6,] <- DTM6

colnames(final.DTM4) <- colnames(DTM4)

colnames(final.DTM5) <- colnames(DTM5)

colnames(final.DTM6) <- colnames(DTM6)

library(wordcloud)

TF4 <- final.DTM4/(0.5+rowSums(final.DTM4))

iDF4 <- log(nrow(final.DTM4)/colSums(final.DTM4>0))

TFiDF4 <- t(t(TF4) \* iDF4)

term.TFiDF4 <- colSums(TFiDF4)

ix4 <- sort(term.TFiDF4,decreasing=TRUE,index.return=TRUE)$ix

TF5 <- final.DTM5/(0.5+rowSums(final.DTM5))

iDF5 <- log(nrow(final.DTM5)/colSums(final.DTM5>0))

TFiDF5 <- t(t(TF5) \* iDF5)

term.TFiDF5 <- colSums(TFiDF5)

ix5 <- sort(term.TFiDF5,decreasing=TRUE,index.return=TRUE)$ix

TF6 <- final.DTM6/(0.5+rowSums(final.DTM6))

iDF6 <- log(nrow(final.DTM6)/colSums(final.DTM6>0))

TFiDF6 <- t(t(TF6) \* iDF6)

term.TFiDF6 <- colSums(TFiDF6)

ix6 <- sort(term.TFiDF6,decreasing=TRUE,index.return=TRUE)$ix

final.DTM4 <- final.DTM4[,ix4]

final.DTM5 <- final.DTM5[,ix5]

final.DTM6 <- final.DTM6[,ix6]

wordcloud(

words = colnames(final.DTM4),

freq = colSums(final.DTM4),

scale = c(3, 0.25))

wordcloud(

words = colnames(final.DTM5),

freq = colSums(final.DTM5),

scale = c(3, 0.25))

wordcloud(

words = colnames(final.DTM6),

freq = colSums(final.DTM6),

scale = c(3, 0.25))

################################################################################

#3c

library(scholar)

PJBindex=get\_scholar\_id(first\_name = "Peter", last\_name = "Bickel")

get\_profile(PJBindex)

pubPJB=get\_publications(

PJBindex,

flush = FALSE,

sortby = "year"

)

pubPJB=pubPJB[which(pubPJB$cites > 0 | pubPJB$year ==2021),]

before2003 <- pubPJB$title[which(pubPJB$year < 2003)]

from2003to2012 <- pubPJB$title[which(pubPJB$year >= 2003 & pubPJB$year < 2013)]

after2012 <- pubPJB$title[which(pubPJB$year >= 2013)]

for(k in 1:length(before2003)){

write(before2003[k], file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder7/", k, sep=""))

}

for(k in 1:length(from2003to2012)){

write(from2003to2012[k], file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder8/", k, sep=""))

}

for(k in 1:length(after2012)){

write(after2012[k], file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder9/", k, sep=""))

}

for(k in 1:length(pubPJB$title)){

write(pubPJB$title[k], file=paste("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder10/", k, sep=""))

}

folders = c("/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder7/", "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder8/", "/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder9/","/Users/a/Desktop/STAT 3280/HW6/statisticians/Data/Textfolder10/")

for (i in c(1:4)) {

cname = folders[i]

docs <- Corpus(DirSource(cname))

meta(docs,"id")

document.names <- unlist(meta(docs,"id"))

docs=tm\_map(docs, removePunctuation)

docs <- tm\_map(docs, removeNumbers)

docs <- tm\_map(docs, tolower)

docs <- tm\_map(docs, removeWords, stopwords("english"))

docs <- tm\_map(docs, stemDocument)

docs <- tm\_map(docs, stripWhitespace)

dtm <- DocumentTermMatrix(docs)

tdm <- TermDocumentMatrix(docs)

freq <- colSums(as.matrix(dtm))

dtms <- removeSparseTerms(dtm, 0.995)

DTM <- as.matrix(dtms)

doc.index <- as.numeric(document.names)

final.DTM <- matrix(0,nrow=nrow(DTM),ncol=ncol(DTM))

final.DTM[doc.index,] <- DTM

colnames(final.DTM) <- colnames(DTM)

TF <- final.DTM/(0.5+rowSums(final.DTM))

iDF <- log(nrow(final.DTM)/colSums(final.DTM>0))

TFiDF <- t(t(TF) \* iDF)

term.TFiDF <- colSums(TFiDF)

ix <- sort(term.TFiDF,decreasing=TRUE,index.return=TRUE)$ix

final.DTM <- final.DTM[,ix]

wordcloud(

words = colnames(final.DTM),

freq = colSums(final.DTM),

scale = c(3, 0.25))

}

#######