**Practical 8**

**Source Code 1**

install.packages("tm")

require("tm")

install.packages("ggplot2")

install.packages("textreuse")

install.packages("devtools")

my.corpus<-Corpus(DirSource("/cloud/project/InputFile"))

my.corpus<-tm\_map(my.corpus, removeWords,stopwords("english"))

my.tdm<-TermDocumentMatrix(my.corpus)

#inspect(my.tdm)

my.dtm<-DocumentTermMatrix(my.corpus,control = list(weighting = weightTfIdf,stopWords = TRUE))

#inspect(my.dtm)

my.df<-as.data.frame(inspect(my.tdm))

my.df.scale<-scale(my.df)

d<-dist(my.df.scale,method = "euclidean")

fit<-hclust(d,method="ward")

plot(fit)

))

**Output:**

Terms File1.txt File2.txt File3.txt

afternoon 0 1 0

are 0 0 1

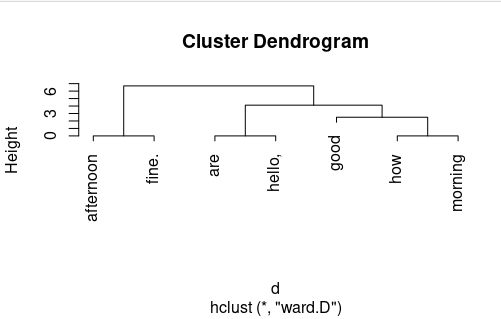
fine. 0 1 0

good 1 1 1

hello, 0 0 1

how 1 0 1

morning 1 0 1



**Source Code 2:**

install.packages("tm")

require("tm")

install.packages("ggplot2")

install.packages("textreuse")

install.packages("devtools")

my.corpus<-Corpus(DirSource("/cloud/project/InputFile"))

my.corpus<-tm\_map(my.corpus, removeWords,stopwords("english"))

my.tdm<-TermDocumentMatrix(my.corpus)

inspect(my.tdm)

#inspect(my.dtm)

my.df<-as.data.frame(inspect(my.tdm))

barplot(as.matrix(my.tdm))

barplot(as.matrix(my.tdm),col=blues9)

**Output :**

Terms File1.txt File2.txt File3.txt

afternoon 0 1 0

are 0 0 1

fine. 0 1 0

good 1 1 1

hello, 0 0 1

how 1 0 1

morning 1 0 1

