George_Smith_HW4_IST772

George Smith

8/1/2021

Introduction

The Federalist Papers were a series of eighty-five essays urging the citiznes on Ney York to ratify the new United States Constiution. The essays originally appeard anonymously in New York newspapers in 1787 and 1788 under the pen names "Publius". It was not until 1818 that the authors Alexander Hamilton, James Madison, and John Jay were identified by name. Using clustering algorithms, k-Means, EM, and HAC I am going to solve the mystery of who wrote each of the Federalist Papers.

installs

```
# install.packages('wordcloud')
# install.packages('tm')
# install.packages('slam')
# install.packages('quanteda')
# install.packages('SnowballC')
# install.packages('arules')
# install.packages('proxy')
# install.packages('cluster')
# install.packages('stringi')
# install.packages('Matrix')
# install.packages('tidytext')
# install.packages('plyr')
# install.packages('ggplot2')
# install.packages('factoextra')
# install.packages('mclust')
# install.packages('dplyr')
# install.packages('rdwplus')
# install.packages('corpus')
# install.packages('quanteda')
# install.packages('tm')
# install.packages('Rcpp')
```

```
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(tm)
```

```
## Loading required package: NLP
```

```
library(slam)
library(quanteda)
```

```
## Package version: 3.0.0
## Unicode version: 10.0
## ICU version: 61.1
```

```
## Parallel computing: 12 of 12 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
##
## Attaching package: 'quanteda'
```

```
## The following object is masked from 'package:tm':
##
##     stopwords
```

```
## The following objects are masked from 'package:NLP':
##
##     meta, meta<-
```

```
library(SnowballC)
library(arules)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'arules'
```

```
## The following object is masked from 'package:tm':
##
##     inspect
```

```
## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
library(proxy)
```

```
##
## Attaching package: 'proxy'
```

```
## The following object is masked from 'package:Matrix':
##
##     as.matrix
## The following objects are masked from ’package:stats’:
##
##     as.dist, dist

## The following object is masked from ’package:base’:
##
##     as.matrix
```

```
library(cluster)
library(stringi)
library(Matrix)
library(tidytext)
library(plyr)
library(ggplot2)
```

```
##
## Attaching package: ’ggplot2’

## The following object is masked from ’package:NLP’:
##
##     annotate
```

```
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(mclust)
```

```
## Package ’mclust’ version 5.4.7
## Type ’citation("mclust")’ for citing this R package in publications.
```

```
library(dplyr)
```

```
##
## Attaching package: ’dplyr’

## The following objects are masked from ’package:plyr’:
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from ’package:arules’:
##
##     intersect, recode, setdiff, setequal, union

## The following objects are masked from ’package:stats’:
##
##     filter, lag

## The following objects are masked from ’package:base’:
##
##     intersect, setdiff, setequal, union
```

```r
library(rdwplus)
```

```
## Loading required package: raster

## Loading required package: sp

##
## Attaching package: 'raster'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: rgrass7

## Loading required package: XML

## GRASS GIS interface loaded with GRASS version: (GRASS not running)
```

```r
library(corpus)
library(tm)
library(Rcpp)
```

read in file

```r
FederalistPapers <- read.csv("C:/Users/GeorgeSmith/Documents/fedPapers85.csv", row.names = 2, na.string
```

Create backup of FederalistPapers in case it's needed

```r
FederalistPapers_Orig <- FederalistPapers
```

Check for NAs and missing values

```r
sum(is.na(FederalistPapers))
```

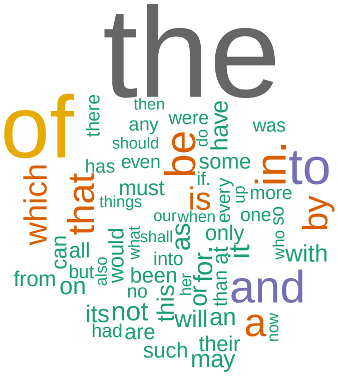
```
## [1] 0
```

```r
FederalistPapers <- FederalistPapers[,-1]
```

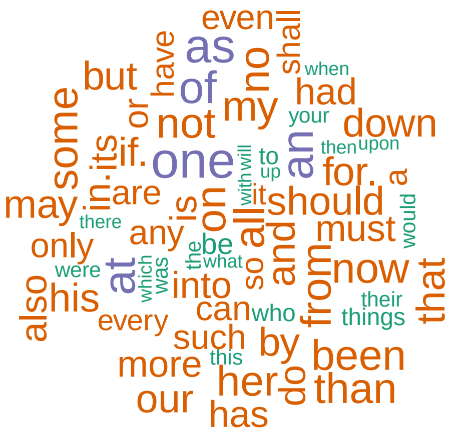
There are no NAS in this data set

first, remove the file and author names for a word cloud gallery

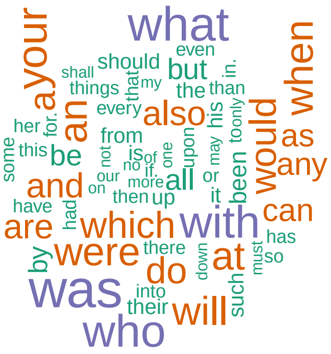
```
FedPapers_wc <- as.matrix(as.dfm(FederalistPapers)) #FederalistPapers[,3:72]
hamPapers = FedPapers_wc[12:62,]
DisputedPapersWC <- wordcloud(colnames(FedPapers_wc), FedPapers_wc[11,],
                              rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



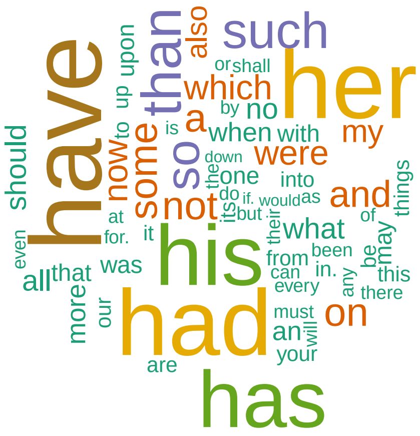
```
HamiltonPapersWC <- wordcloud(colnames(FedPapers_wc), FedPapers_wc[12:62,],
                              rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



```
MadisonPapersWC <- wordcloud(colnames(FedPapers_wc), FedPapers_wc[71:85,],
                             rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



```
JayPapersWC <- wordcloud(colnames(FedPapers_wc), FedPapers_wc [66:70,],
                         rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



K means

Need to clean the data by removing the labels and determining the optimal numbers of clusters for the clustering algorithm.

Remove author names from dataset for clustering purposes

```
FederalistPapers <- read.csv("fedPapers85.csv", na.strings = c(""))
```

Make the file names the row names. Need a dataframe of numerical values for k-means

```
FedPapers_km <- FederalistPapers[,2:72]
```

Make the file names the row names. Need a dataframe of numerical values for k-means

```
rownames(FedPapers_km) <- FedPapers_km[,1]
FedPapers_km[,1] <- NULL
```

Set seed for fixed random seed

```
set.seed(20)
```

run k-means

```
Clusters <- kmeans(FedPapers_km, 6)
FedPapers_km$Clusters <- as.factor(Clusters$cluster)
str(Clusters)
```

```
## List of 9
##  $ cluster     : Named int [1:85] 1 6 1 6 6 1 6 5 1 6 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_f
##  $ centers     : num [1:6, 1:70] 0.297 0.216 0.16 0.299 0.363 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:6] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:70] "a" "all" "also" "an" ...
##  $ totss       : num 12.6
##  $ withinss    : num [1:6] 2.039 0.423 0.599 1.791 1.12 ...
##  $ tot.withinss: num 7.29
##  $ betweenss   : num 5.28
##  $ size        : int [1:6] 20 5 5 24 13 18
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```
Clusters$centers[]
```

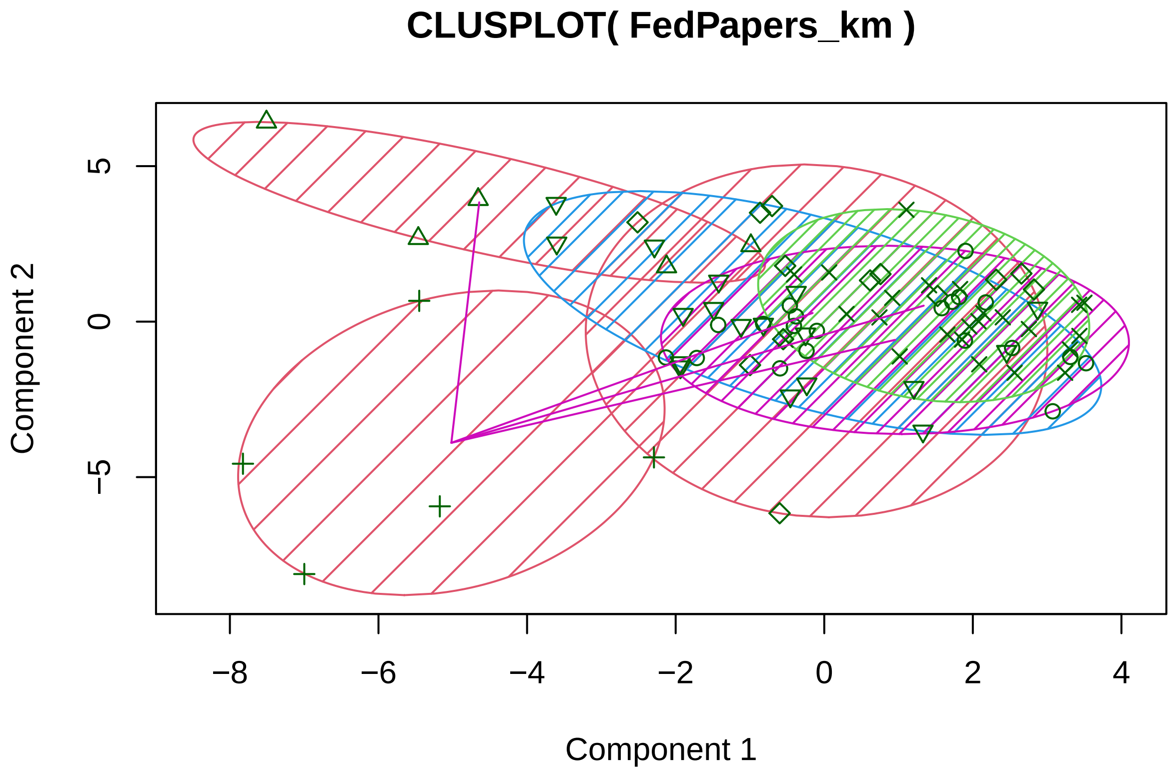
```
##           a        all        also         an       and        any        are
## 1 0.2971500 0.05520000 0.008900000 0.06480000 0.3287000 0.04300000 0.07375000
## 2 0.2156000 0.05760000 0.013000000 0.05200000 0.4990000 0.01960000 0.08540000
## 3 0.1598000 0.03600000 0.019800000 0.02520000 0.7152000 0.03760000 0.08520000
## 4 0.2992083 0.05291667 0.003333333 0.08833333 0.3406667 0.05016667 0.07466667
## 5 0.3633846 0.05938462 0.005923077 0.07476923 0.3697692 0.03261538 0.07800000
## 6 0.2888889 0.04872222 0.008444444 0.05772222 0.3925000 0.04238889 0.07872222
##          as         at        be       been        but        by        can
## 1 0.1350000 0.03375000 0.3415000 0.05160000 0.03130000 0.1271000 0.03205000
## 2 0.0700000 0.04640000 0.1196000 0.03280000 0.02400000 0.1648000 0.01620000
## 3 0.1568000 0.03600000 0.2754000 0.02680000 0.04920000 0.1362000 0.03300000
## 4 0.1300417 0.04579167 0.3185000 0.06358333 0.03237500 0.0992500 0.03883333
## 5 0.1078462 0.05784615 0.2678462 0.06276923 0.03176923 0.1120769 0.03284615
## 6 0.1222222 0.04583333 0.3148333 0.07777778 0.03138889 0.1623333 0.04322222
##            do         down       even      every      for.       from
## 1 0.005200000 0.0016000000 0.00685000 0.02845000 0.0929500 0.06675000
## 2 0.002400000 0.0020000000 0.00560000 0.01060000 0.0784000 0.08560000
## 3 0.008200000 0.0000000000 0.00760000 0.00600000 0.0960000 0.09100000
## 4 0.006916667 0.0032916667 0.01600000 0.02170833 0.0907500 0.08166667
## 5 0.009230769 0.0003846154 0.01307692 0.02253846 0.0750000 0.08923077
## 6 0.004944444 0.0002222222 0.01177778 0.03144444 0.1158889 0.08016667
##          had        has       have         her        his        if.       in.
## 1 0.01495000 0.03380000 0.08525000 0.001900000 0.02135000 0.02500000 0.3358500
## 2 0.05560000 0.05240000 0.06180000 0.012200000 0.07520000 0.01140000 0.2538000
## 3 0.01640000 0.02880000 0.08680000 0.014800000 0.00900000 0.05260000 0.2714000
## 4 0.01895833 0.04387500 0.10241667 0.002333333 0.04329167 0.02833333 0.3377083
## 5 0.01953846 0.05815385 0.10607692 0.022384615 0.01684615 0.02730769 0.3194615
## 6 0.02394444 0.04916667 0.09822222 0.009333333 0.01816667 0.02600000 0.2985556
##         into        is        it        its        may       more       must
## 1 0.02330000 0.1675500 0.1681500 0.04660000 0.06170000 0.04545000 0.02920000
## 2 0.02420000 0.1258000 0.1008000 0.05360000 0.02600000 0.05080000 0.01100000
## 3 0.04460000 0.0936000 0.2048000 0.03340000 0.05680000 0.08680000 0.02120000
## 4 0.01712500 0.1724167 0.1709167 0.05666667 0.06895833 0.03566667 0.03720833
## 5 0.02984615 0.1249231 0.1183846 0.04084615 0.05069231 0.04392308 0.03238462
## 6 0.02438889 0.1707222 0.1550556 0.04738889 0.07177778 0.04738889 0.04166667
##            my         no        not         now        of         on        one
## 1 0.002150000 0.03830000 0.09590000 0.006000000 0.9746500 0.08920000 0.03815000
## 2 0.005000000 0.02900000 0.04040000 0.007600000 0.8950000 0.07960000 0.04460000
## 3 0.001800000 0.01500000 0.10800000 0.006600000 0.6390000 0.07460000 0.08140000
## 4 0.003208333 0.03362500 0.09483333 0.005333333 0.9127917 0.04333333 0.03512500
## 5 0.002538462 0.02423077 0.08361538 0.008307692 1.0096154 0.05407692 0.04130769
## 6 0.005000000 0.03572222 0.10211111 0.004777778 0.8388889 0.08827778 0.03855556
##         only         or        our      shall     should         so       some
## 1 0.02600000 0.09765000 0.00715000 0.02180000 0.02600000 0.02510000 0.01570000
## 2 0.01100000 0.07320000 0.00720000 0.01180000 0.00700000 0.02180000 0.01780000
## 3 0.04340000 0.16080000 0.06600000 0.01740000 0.04140000 0.04460000 0.02140000
## 4 0.02125000 0.10008333 0.01804167 0.02175000 0.03425000 0.03000000 0.01683333
## 5 0.01792308 0.08992308 0.04330769 0.01476923 0.02300000 0.03038462 0.01984615
## 6 0.02277778 0.08494444 0.02500000 0.01655556 0.02083333 0.03255556 0.02883333
##         such       than      that      the      their        then      there
## 1 0.02825000 0.04755000 0.2320000 1.476150 0.07090000 0.005800000 0.02600000
## 2 0.02060000 0.03680000 0.1330000 1.337800 0.09840000 0.007800000 0.00820000
## 3 0.05120000 0.06280000 0.2434000 0.854400 0.14160000 0.008000000 0.01400000
## 4 0.03233333 0.03804167 0.2109583 1.332833 0.07741667 0.006708333 0.03520833
## 5 0.02192308 0.04346154 0.1880000 1.123000 0.07807692 0.002461538 0.03800000
## 6 0.02772222 0.04500000 0.2218889 1.210833 0.09883333 0.007166667 0.01511111
##        things       this        to          up       upon        was       were
## 1 0.003200000 0.09230000 0.5008500 0.000500000 0.02400000 0.02060000 0.01340000
## 2 0.001800000 0.06880000 0.4004000 0.005400000 0.01220000 0.08340000 0.03980000
## 3 0.001400000 0.05320000 0.4834000 0.000000000 0.00180000 0.02480000 0.02880000
## 4 0.002708333 0.08795833 0.6470000 0.005875000 0.04745833 0.02062500 0.01879167
## 5 0.004615385 0.09692308 0.5103077 0.006769231 0.03961538 0.02084615 0.02076923
## 6 0.001166667 0.08716667 0.4968889 0.001666667 0.01555556 0.02650000 0.02150000
##         what        when     which        who       will       with      would
## 1 0.01340000 0.013850000 0.1522000 0.02625000 0.12565000 0.07495000 0.11200000
## 2 0.00520000 0.011400000 0.1484000 0.03820000 0.03660000 0.09600000 0.04120000
## 3 0.01840000 0.021000000 0.0986000 0.05160000 0.12600000 0.09500000 0.12520000
## 4 0.01475000 0.011916667 0.1616667 0.03833333 0.09187500 0.07570833 0.10587500
## 5 0.01284615 0.009769231 0.1603846 0.02130769 0.08169231 0.09515385 0.14823077
## 6 0.01033333 0.008111111 0.1758333 0.03300000 0.09955556 0.07027778 0.06144444
##           your
## 1 0.0000000000
## 2 0.0000000000
## 3 0.0064000000
## 4 0.0009166667
## 5 0.0007692308
## 6 0.0060000000
```

Add clusters to dataframe original dataframe with author name

```
FedPapers_km2 <- FederalistPapers
FedPapers_km2$Clusters <- as.factor(Clusters$cluster)
```

Plot results

```
clusplot(FedPapers_km, FedPapers_km$Clusters, color=TRUE, shade=TRUE, labels=0, lines=0)

clusplot(FedPapers_km, FedPapers_km$Clusters, color=TRUE, shade=TRUE, labels=0, lines=T)
```



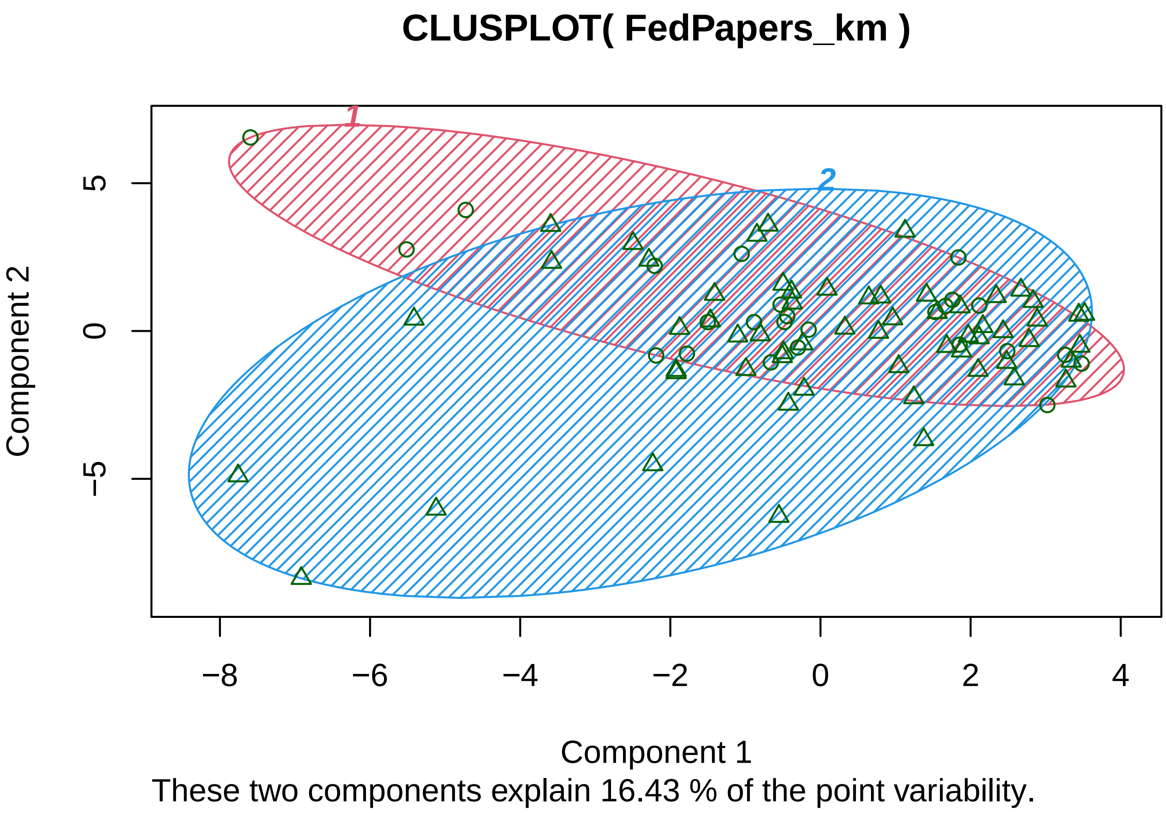
These two components explain 16.39 % of the point variability.

# word clouds based on authorship

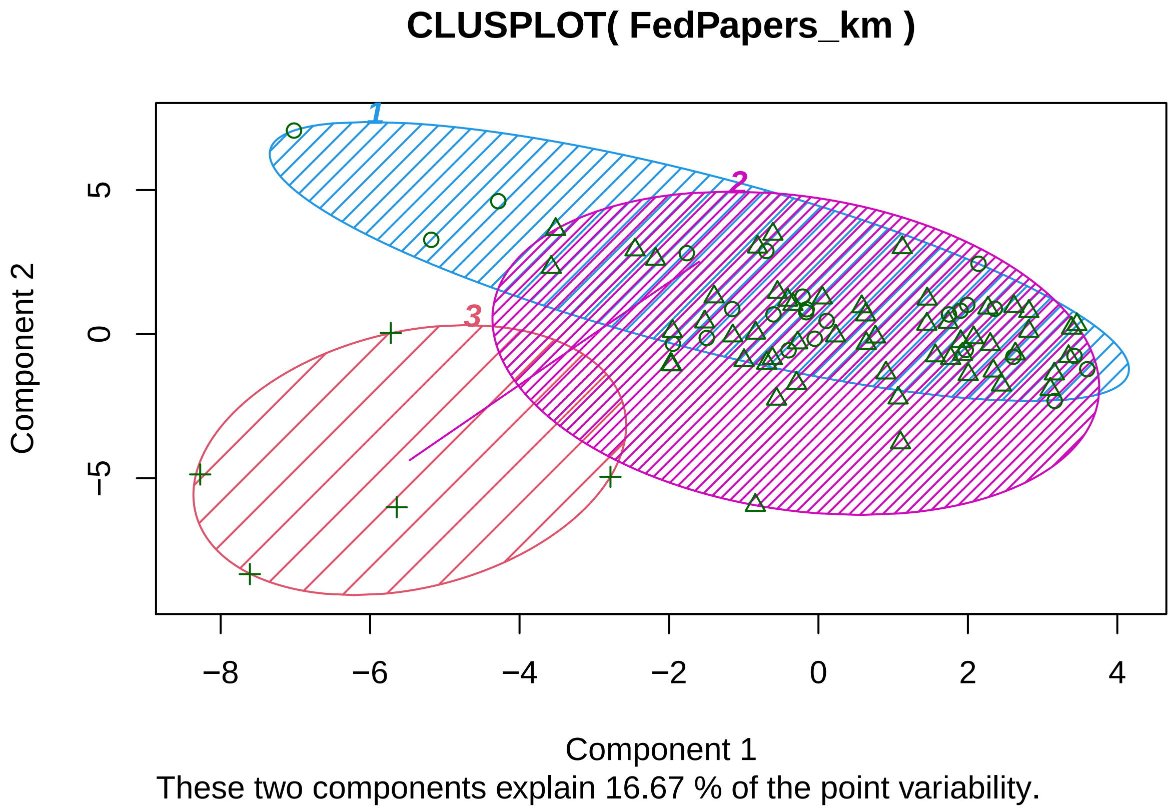
#Loop

```
cluster_loop <- c(2,3,4,5,6,7,8,9)
set.seed(20)
for (x in cluster_loop){
  print(x)
  # run k-means
  Clusters <- kmeans(FedPapers_km, x)
  FedPapers_km$Clusters <- as.factor(Clusters$cluster)
  str(Clusters)
  #print(Clusters$centers)
  # Plot results
  clusplot(FedPapers_km, FedPapers_km$Clusters, color=T, shade=T, labels=4, lines=T)
}
```

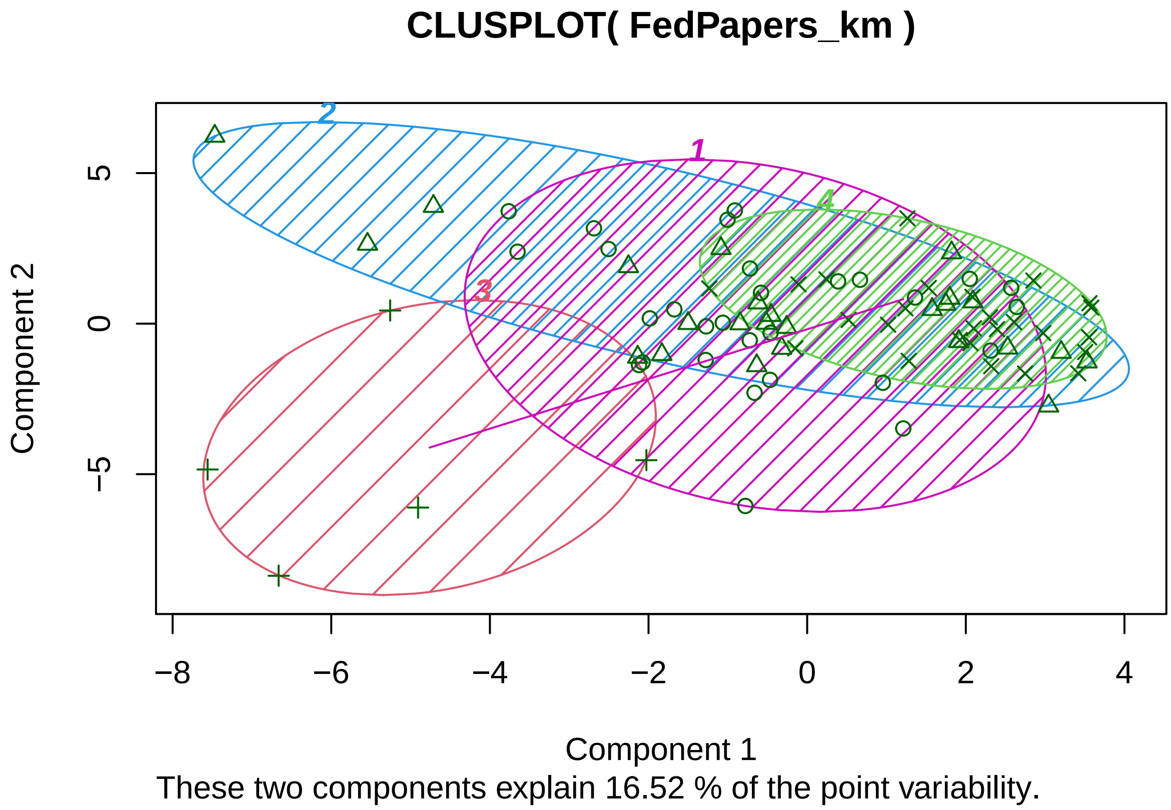
```
## [1] 2
## List of 9
##  $ cluster     : Named int [1:85] 1 2 1 2 2 1 2 2 1 2 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:2, 1:71] 0.28084 0.2984 0.05568 0.05165 0.00972 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "1" "2"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 295
##  $ withinss    : num [1:2] 7.2 65.8
##  $ tot.withinss: num 73
##  $ betweenss   : num 222
##  $ size        : int [1:2] 25 60
##  $ iter        : int 1
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



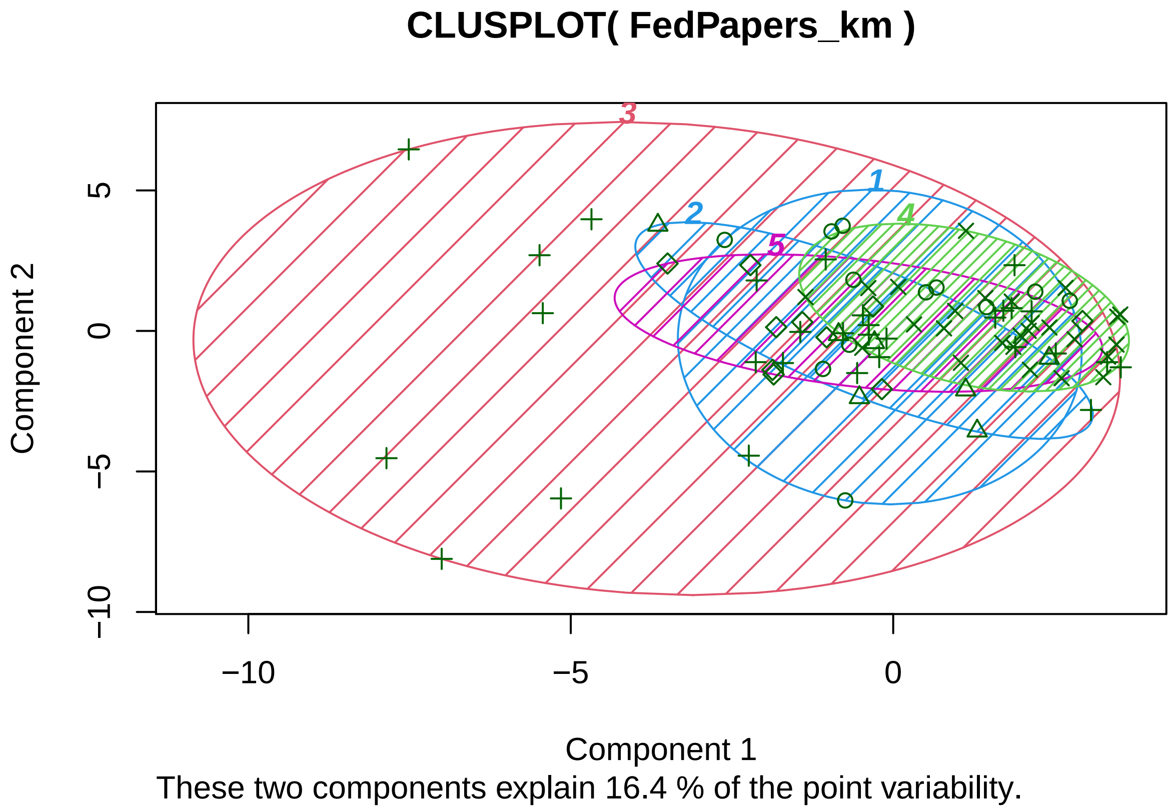
```
## [1] 3
## List of 9
##  $ cluster     : Named int [1:85] 1 2 1 2 2 1 2 2 1 2 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:3, 1:71] 0.2808 0.311 0.1598 0.0557 0.0531 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3] "1" "2" "3"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 30.2
##  $ withinss    : num [1:3] 3.197 5.543 0.599
##  $ tot.withinss: num 9.34
##  $ betweenss   : num 20.9
##  $ size        : int [1:3] 25 55 5
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



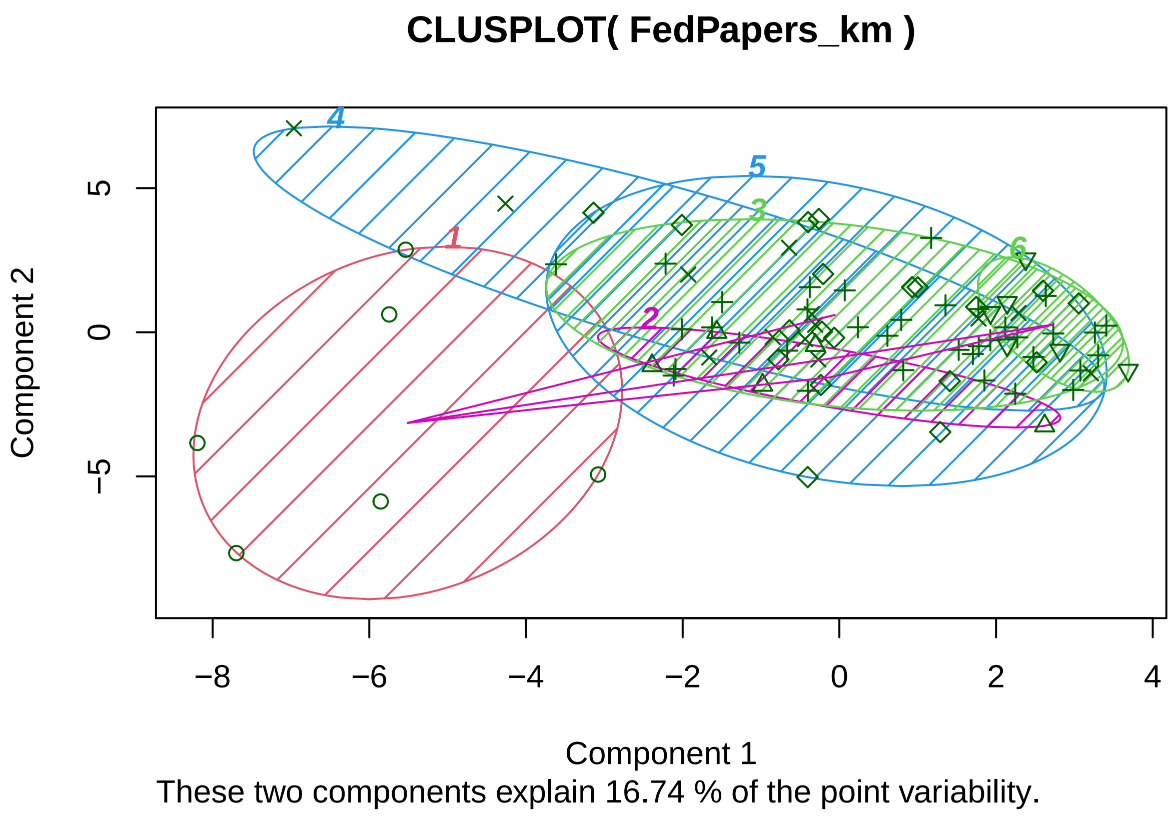
```
## [1] 4
## List of 9
##  $ cluster     : Named int [1:85] 2 1 2 4 1 2 1 1 2 1 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_f
##  $ centers     : num [1:4, 1:71] 0.3168 0.2808 0.1598 0.3045 0.0537 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "1" "2" "3" "4"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 37.9
##  $ withinss    : num [1:4] 2.713 3.197 0.599 2.017
##  $ tot.withinss: num 8.53
##  $ betweenss   : num 29.3
##  $ size        : int [1:4] 29 25 5 26
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



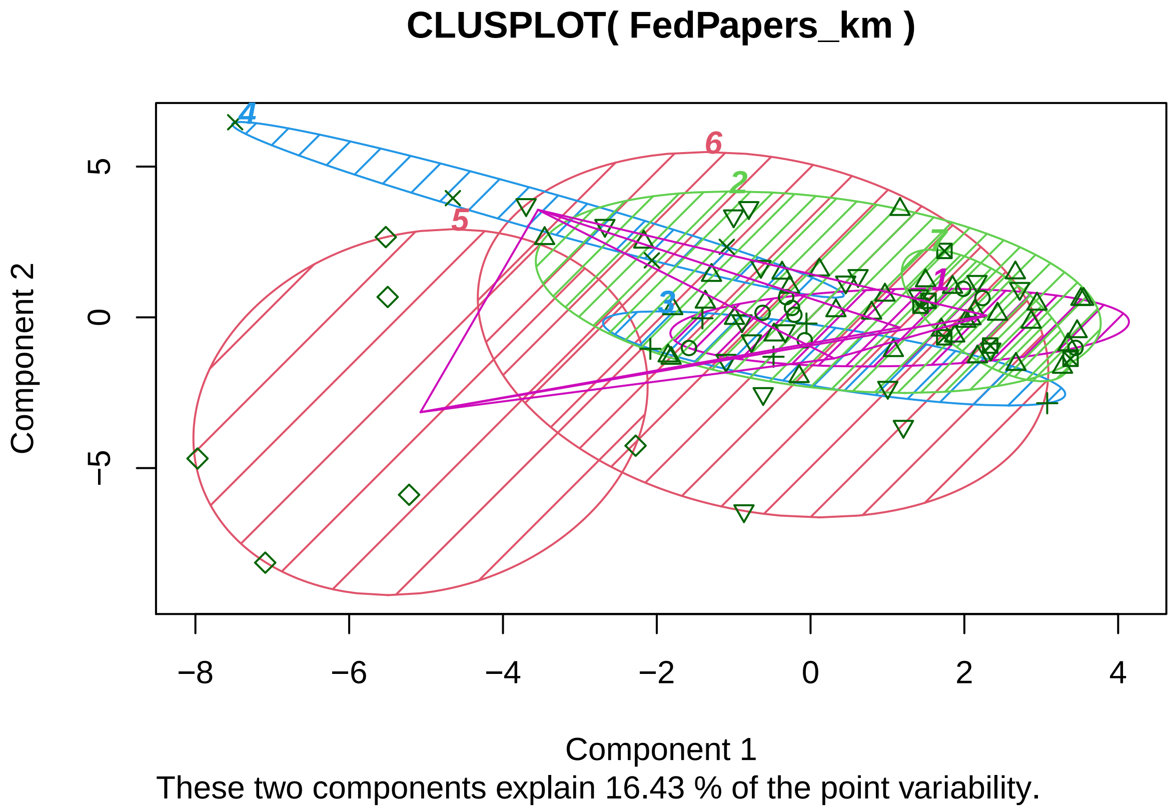
```
## [1] 5
## List of 9
##  $ cluster     : Named int [1:85] 3 2 3 4 5 3 2 1 3 5 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:5, 1:71] 0.355 0.256 0.261 0.305 0.314 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 141
##  $ withinss    : num [1:5] 1.008 0.511 10.619 2.017 0.562
##  $ tot.withinss: num 14.7
##  $ betweenss   : num 127
##  $ size        : int [1:5] 12 7 30 26 10
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



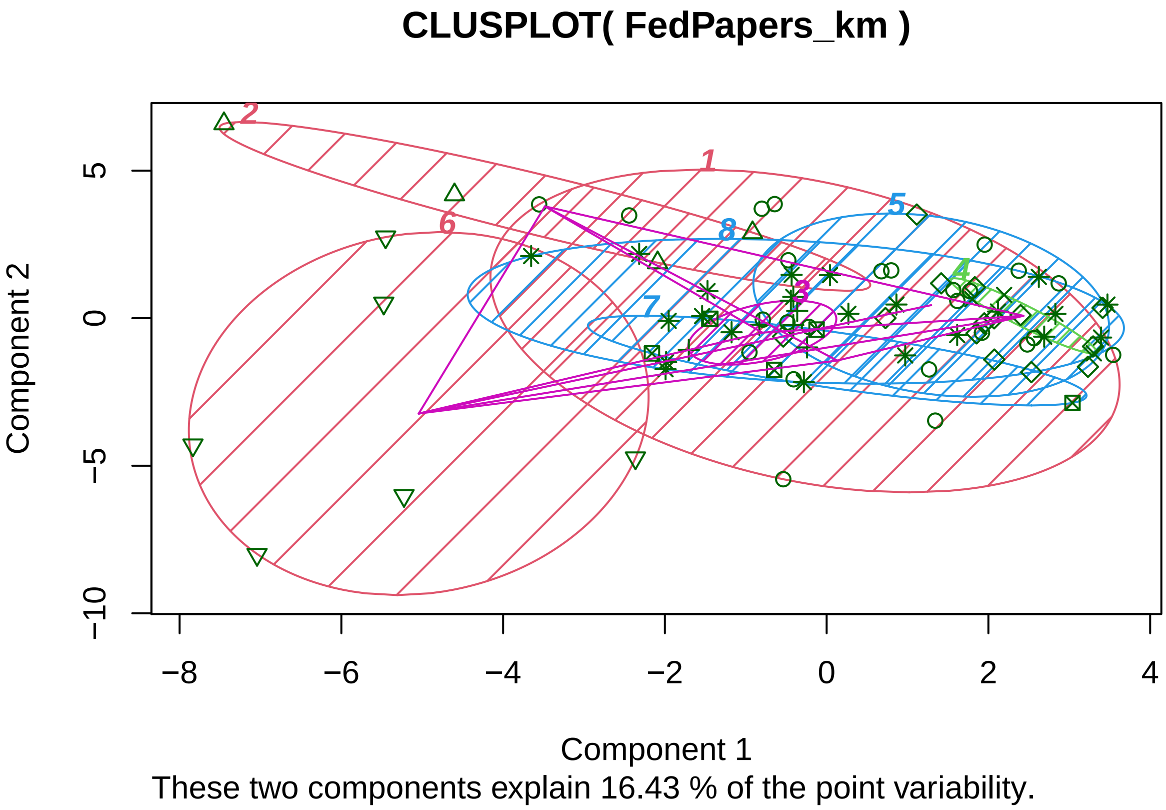
```
## [1] 6
## List of 9
##  $ cluster     : Named int [1:85] 4 5 2 3 3 4 5 5 4 3 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:6, 1:71] 0.171 0.261 0.307 0.264 0.319 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:6] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 131
##  $ withinss    : num [1:6] 0.858 0.416 10.386 1.234 6.293 ...
##  $ tot.withinss: num 19.6
##  $ betweenss   : num 111
##  $ size        : int [1:6] 6 5 36 13 19 6
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



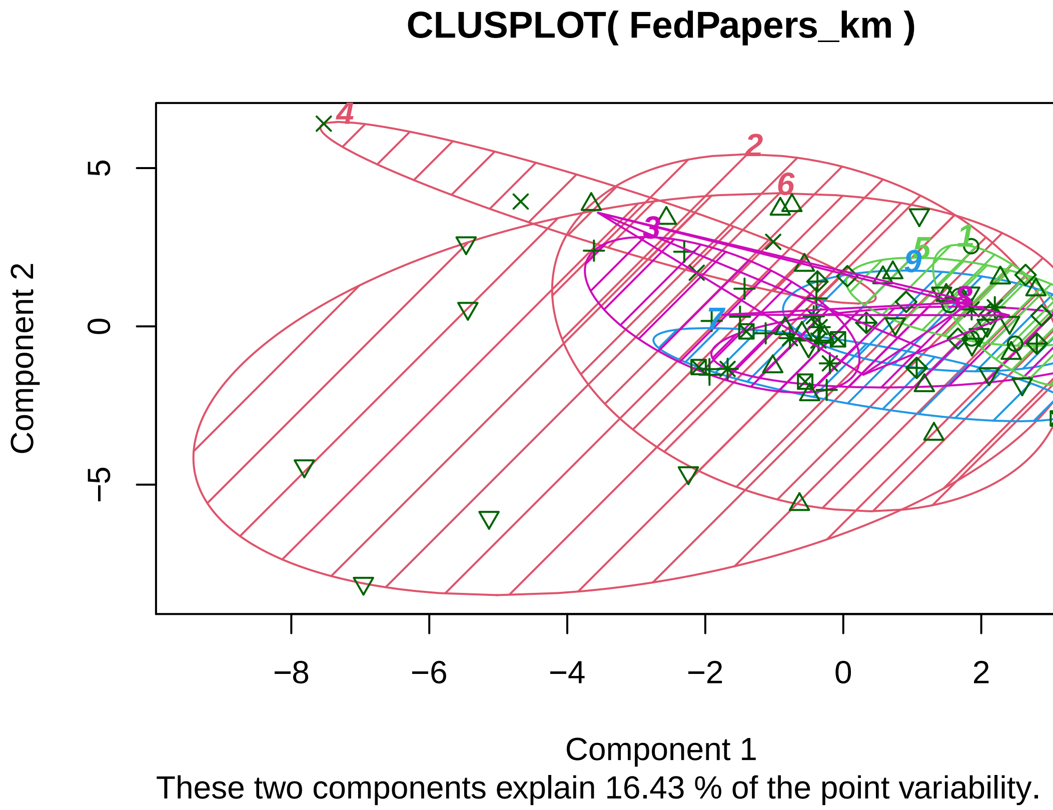
```
## [1] 7
## List of 9
##  $ cluster     : Named int [1:85] 1 6 3 2 2 1 6 6 1 2 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:7, 1:71] 0.287 0.307 0.261 0.213 0.171 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:7] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 153
##  $ withinss    : num [1:7] 0.651 3.164 0.416 0.264 0.858 ...
##  $ tot.withinss: num 7.61
##  $ betweenss   : num 145
##  $ size        : int [1:7] 9 36 5 4 6 19 6
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



```
## [1] 8
## List of 9
##  $ cluster     : Named int [1:85] 3 1 7 8 8 3 1 1 3 8 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:8, 1:71] 0.324 0.213 0.277 0.305 0.263 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:8] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 358
##  $ withinss    : num [1:8] 7.64 0.264 0.325 0.18 0.925 ...
##  $ tot.withinss: num 12.3
##  $ betweenss   : num 346
##  $ size        : int [1:8] 25 4 6 3 14 6 5 22
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



```
## [1] 9
## List of 9
##  $ cluster     : Named int [1:85] 8 2 7 3 3 8 2 2 8 3 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:9, 1:71] 0.343 0.319 0.305 0.213 0.37 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:9] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 676
##  $ withinss    : num [1:9] 0.388 1.872 0.537 0.264 0.386 ...
##  $ tot.withinss: num 15
##  $ betweenss   : num 661
##  $ size        : int [1:9] 6 19 10 4 7 20 5 9 5
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



## Hierachical Clustering Algorithms (HAC)

Remove author names from dataset

```
FedPapers_HAC <- FederalistPapers[,c(2:72)]
```

Make the file names the row names. Need a dataframe of numerical values for HAC

```
rownames(FedPapers_HAC) <- FedPapers_HAC[,1]
FedPapers_HAC[,1] <- NULL
View(FedPapers_HAC)
```

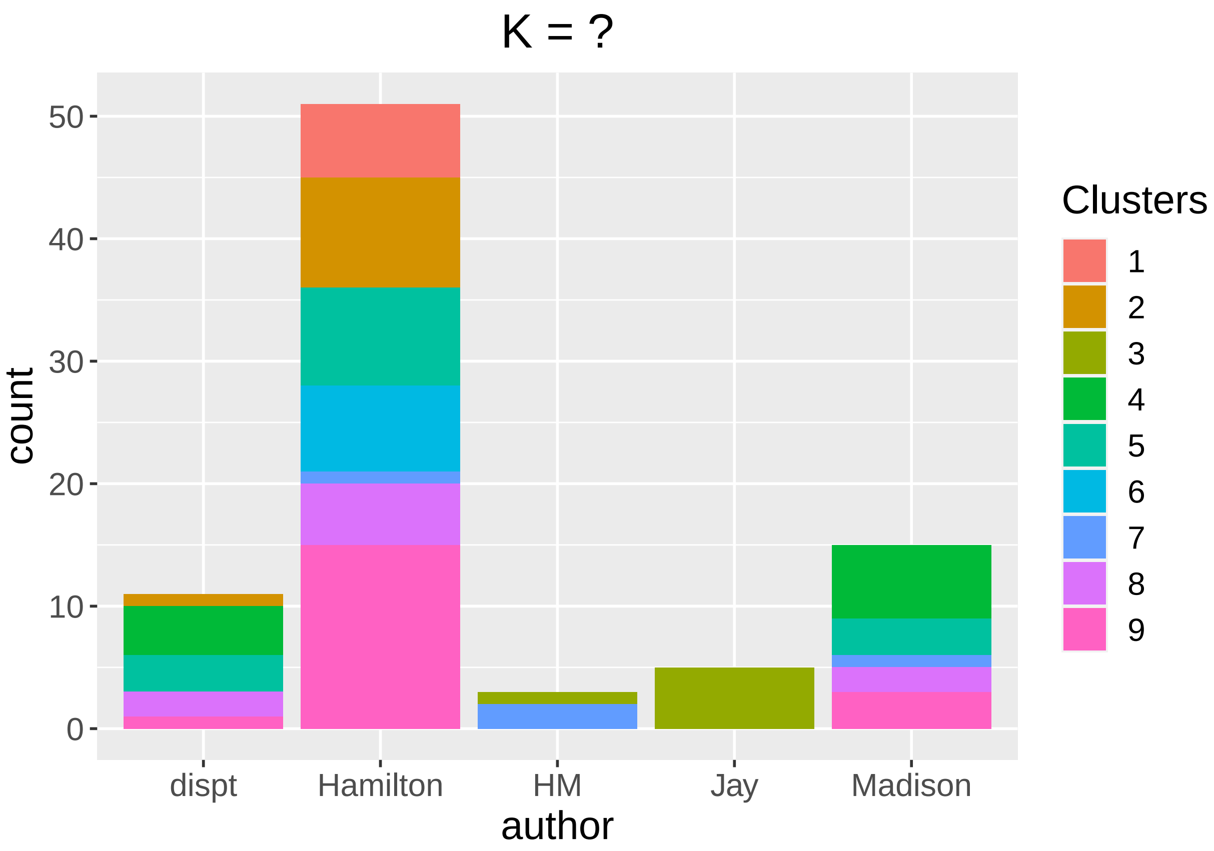
Calculate distance in a variety of ways

```
distance <- dist(FedPapers_HAC, method = "euclidean")
distance2 <- dist(FedPapers_HAC, method = "maximum")
distance3 <- dist(FedPapers_HAC, method = "manhattan")
```

```r
distance4 <- dist(FedPapers_HAC, method = "canberra")
distance5 <- dist(FedPapers_HAC, method = "binary")
distance6 <- dist(FedPapers_HAC, method = "minkowski", p = 3)

Clusters1 <- kmeans(FedPapers_km, 9)
FedPapers_km2$Clusters <- as.factor(Clusters1$cluster)
str(Clusters)
```

```
## List of 9
##  $ cluster     : Named int [1:85] 8 2 7 3 3 8 2 2 8 3 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:9, 1:71] 0.343 0.319 0.305 0.213 0.37 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:9] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 676
##  $ withinss    : num [1:9] 0.388 1.872 0.537 0.264 0.386 ...
##  $ tot.withinss: num 15
##  $ betweenss   : num 661
##  $ size        : int [1:9] 6 19 10 4 7 20 5 9 5
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```r
ggplot(data=FedPapers_km2, aes(x=author, fill=Clusters))+
  geom_bar(stat="count") +
  labs(title = "K = ?") +
  theme(plot.title = element_text(hjust=0.5), text=element_text(size=15))
```



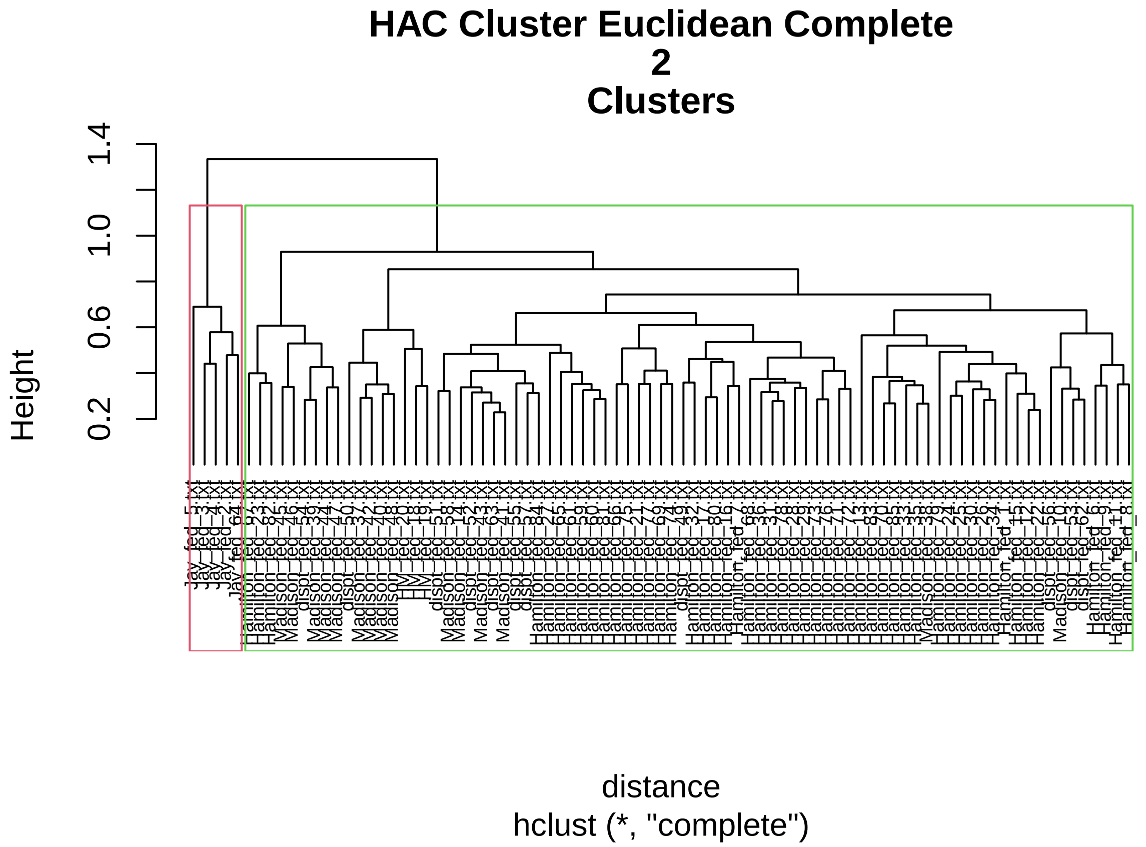
# Madison essays

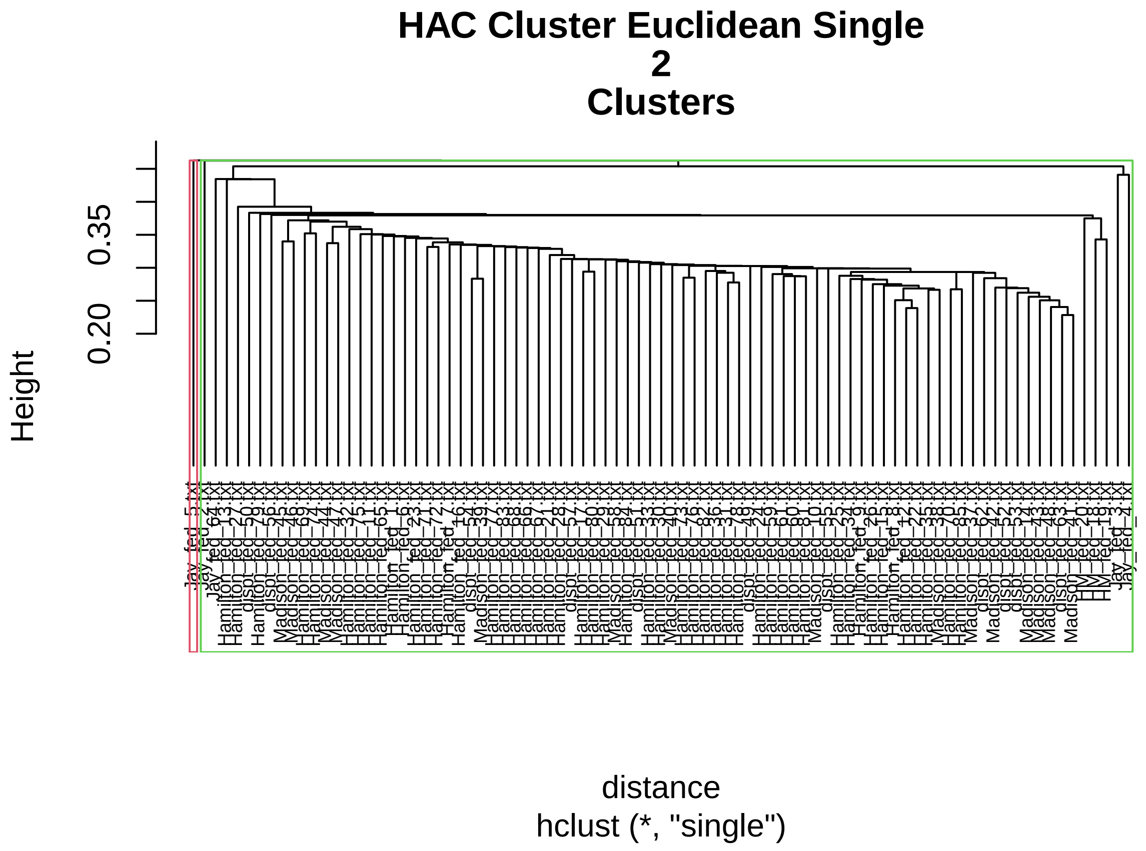
```
Madison_Leaning <- FederalistPapers[which(FedPapers_km2$Clusters[c(1:11)]== 8 | FedPapers_km$Clusters[c
Madison_Leaning
```

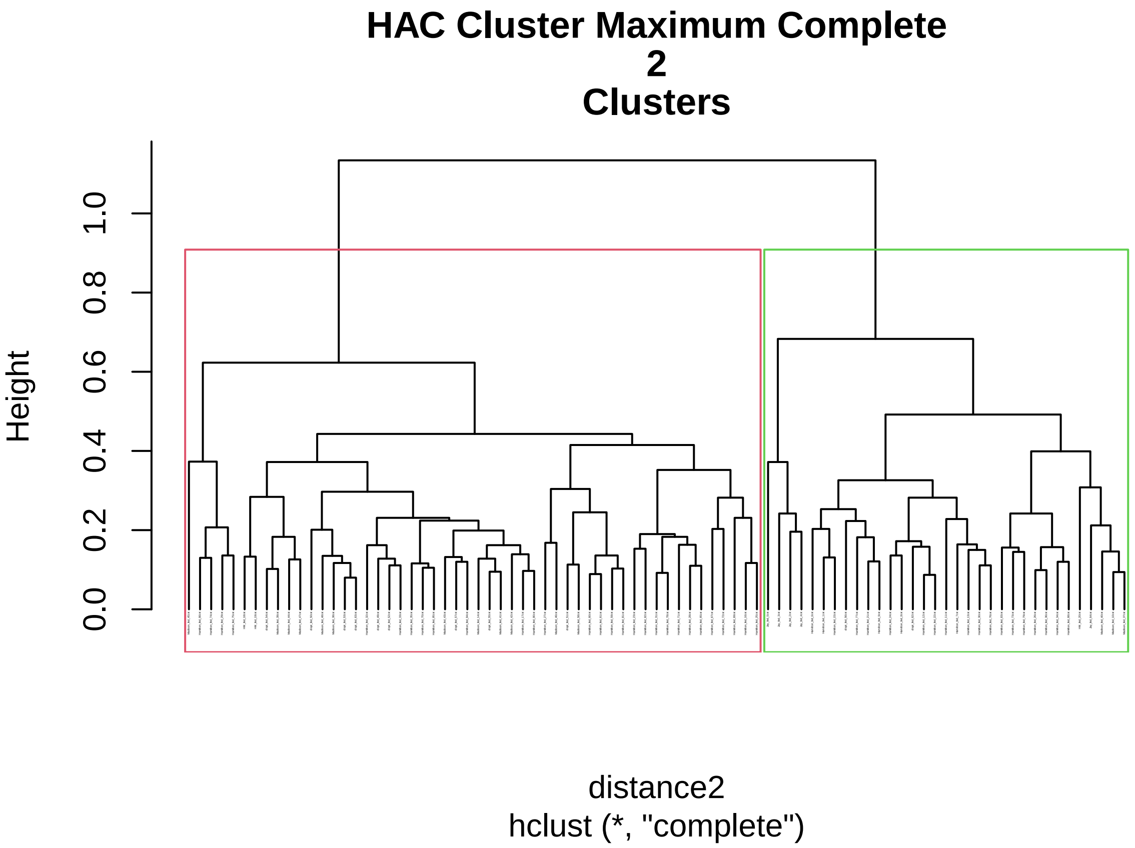
```
## [1] "dispt_fed_50.txt" "dispt_fed_52.txt" "dispt_fed_53.txt" "dispt_fed_55.txt"
## [5] "dispt_fed_62.txt" "dispt_fed_63.txt"
```

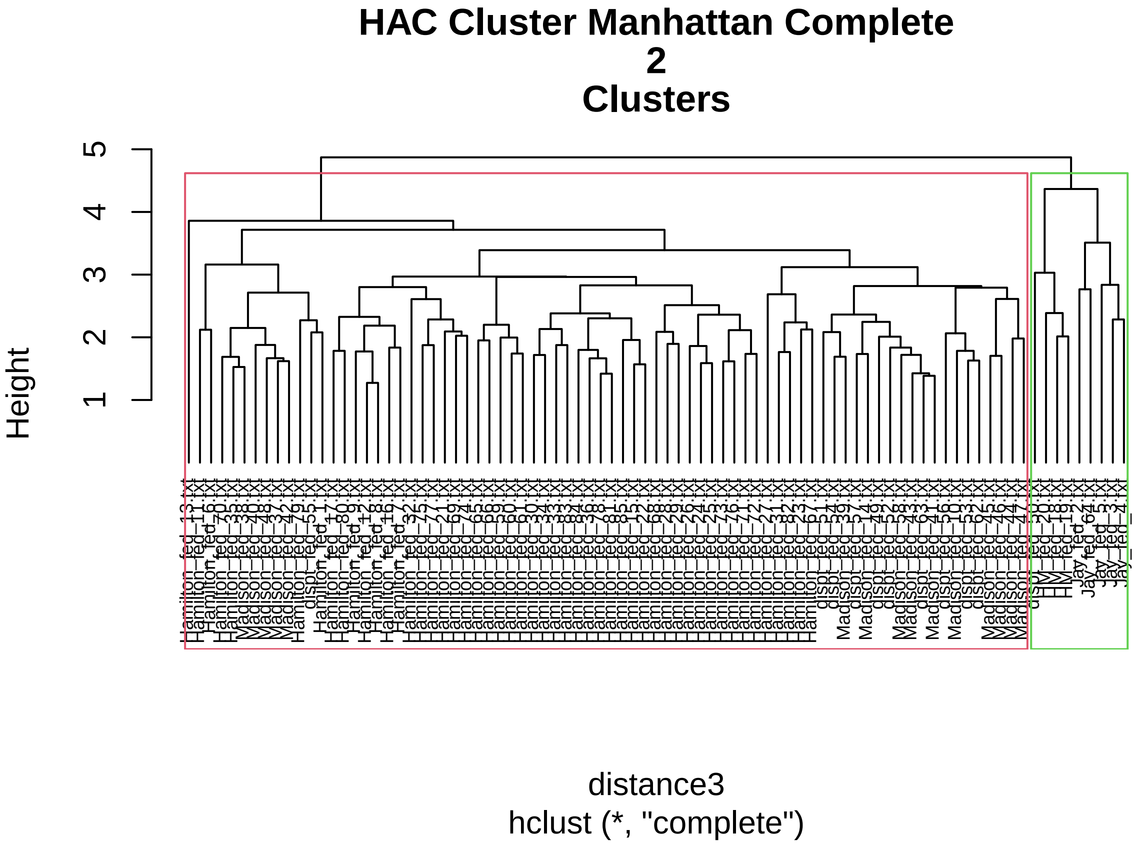
A loop to plot multiple HACs

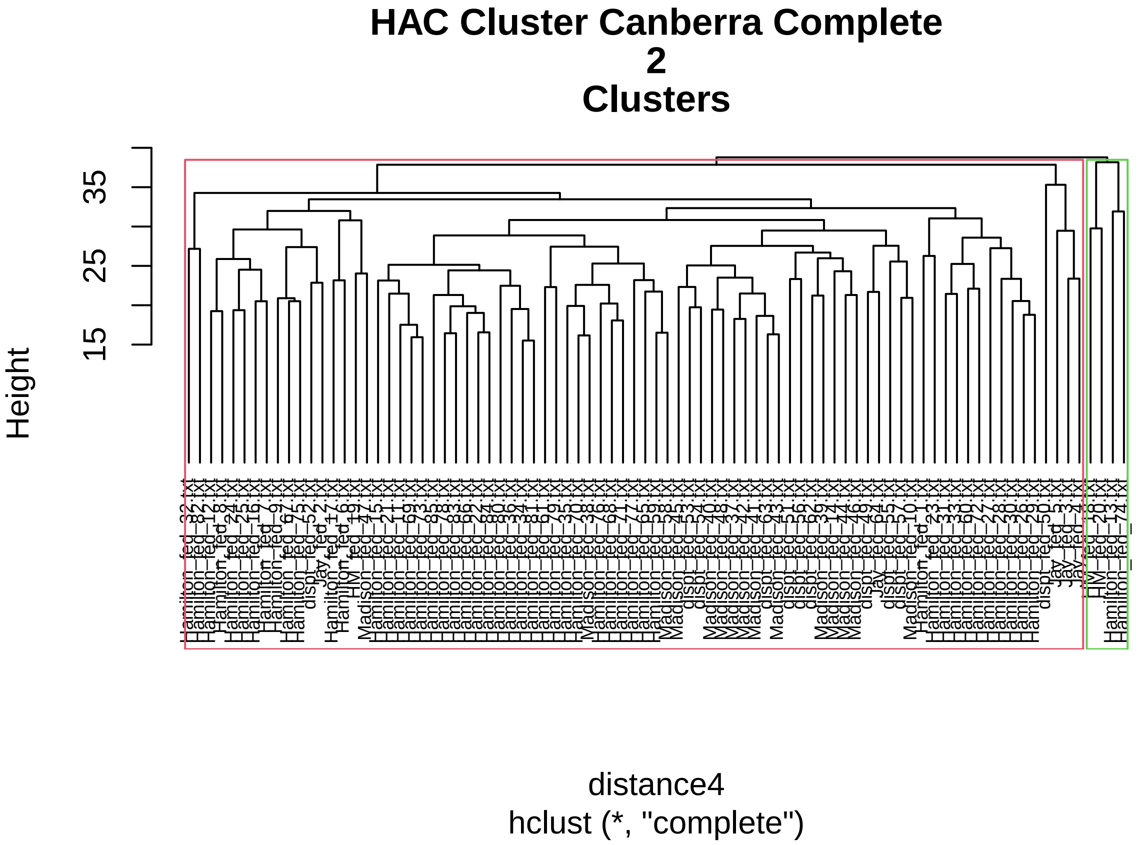
```
hac_loop <- c(2,3,4,5,6,7,8,9)
for (y in hac_loop) {
  HAC <- hclust(distance, method="complete")
  plot(HAC, cex=0.6, hang=-1, main = c("HAC Cluster Euclidean Complete", y, "Clusters"))
  rect.hclust(HAC, k = y, border=2:5)

  HACSingle <- hclust(distance, method="single")
  plot(HACSingle, cex=0.6, hang=-1, main = c("HAC Cluster Euclidean Single", y, "Clusters"))
  rect.hclust(HACSingle, k = y, border=2:5)

  HAC2 <- hclust(distance2, method="complete")
  plot(HAC2, cex=.1, hang=-1, main = c("HAC Cluster Maximum Complete", y, "Clusters"))
  rect.hclust(HAC2, k =y, border=2:5)
  HAC3 <- hclust(distance3, method="complete")
  plot(HAC3, cex=0.6, hang=-1, main = c("HAC Cluster Manhattan Complete", y, "Clusters"))
  rect.hclust(HAC3, k =y, border=2:5)

  HAC4 <- hclust(distance4, method="complete")
  plot(HAC4, cex=0.6, hang=-1, main = c("HAC Cluster Canberra Complete", y, "Clusters"))
  rect.hclust(HAC4, k =y, border=2:5)

  HAC5 <- hclust(distance5, method="complete")
  plot(HAC5, cex=0.6, hang=-1, main = c("HAC Cluster Minkowski Complete", y, "Clusters"))
  rect.hclust(HAC5, k =y, border=2:5)

  HAC6 <- hclust(distance6, method="complete")
  plot(HAC6, cex=0.6, hang=-1, main = c("HAC Cluster Maximum Complete", y, "Clusters"))
  rect.hclust(HAC6, k =y, border=2:5)
}
```

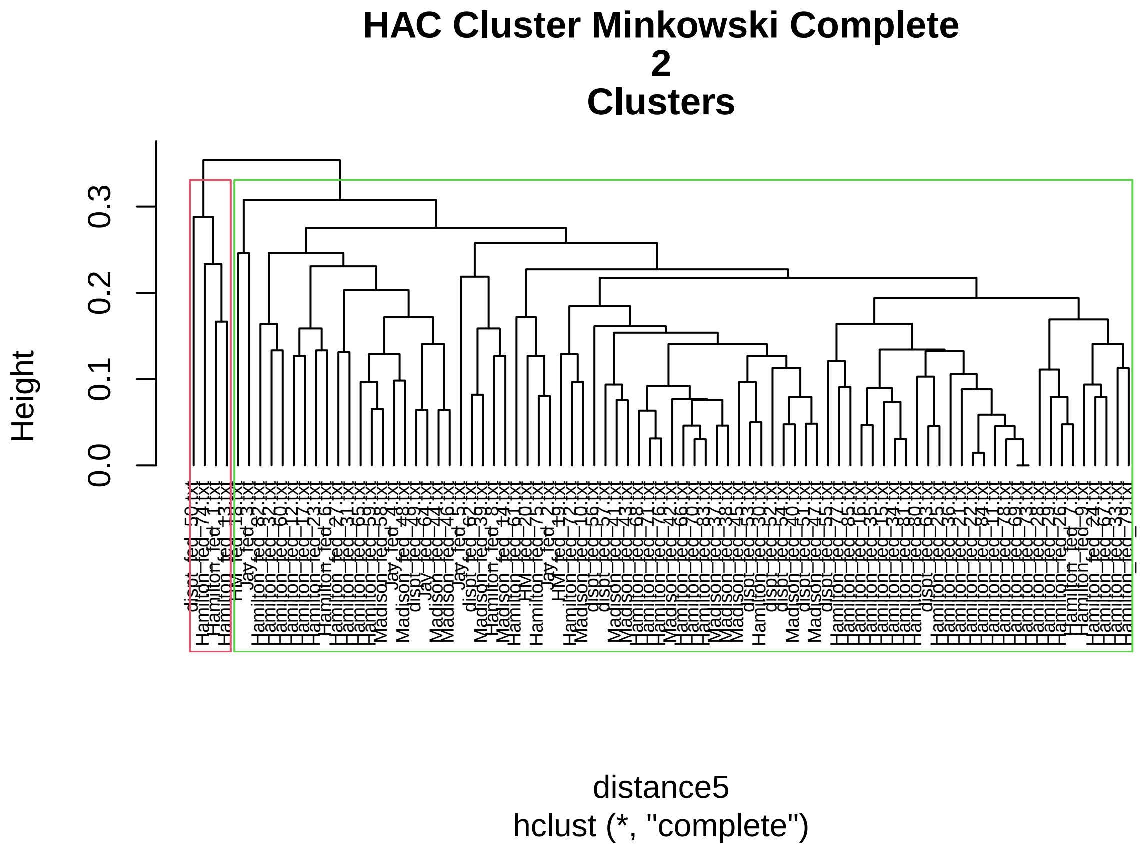


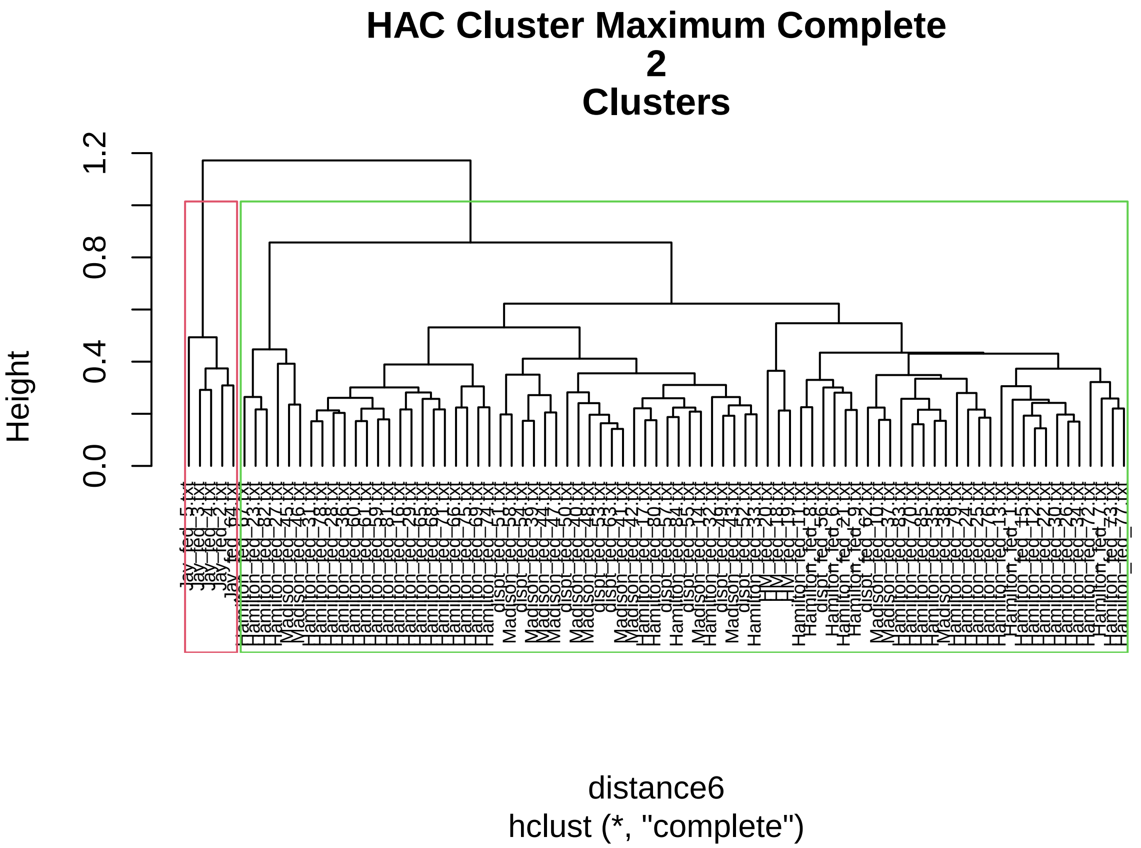

HAC Cluster Euclidean Single
2
Clusters
Height
0.20
0.35
Jay_fed_5.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_13.txt
Hamilton_fed_27.txt
dispt_fed_50.txt
Hamilton_fed_79.txt
dispt_fed_56.txt
Madison_fed_45.txt
Madison_fed_46.txt
Hamilton_fed_69.txt
Hamilton_fed_74.txt
Madison_fed_44.txt
Madison_fed_47.txt
Hamilton_fed_32.txt
Hamilton_fed_75.txt
Hamilton_fed_11.txt
Hamilton_fed_65.txt
Hamilton_fed_1.txt
Hamilton_fed_33.txt
Hamilton_fed_21.txt
Hamilton_fed_72.txt
Hamilton_fed_7.txt
Hamilton_fed_16.txt
dispt_fed_39.txt
Madison_fed_37.txt
Hamilton_fed_83.txt
Hamilton_fed_68.txt
Hamilton_fed_69.txt
Hamilton_fed_61.txt
Hamilton_fed_28.txt
dispt_fed_57.txt
Hamilton_fed_70.txt
Hamilton_fed_80.txt
Madison_fed_58.txt
Hamilton_fed_84.txt
dispt_fed_51.txt
Hamilton_fed_31.txt
Hamilton_fed_30.txt
Madison_fed_43.txt
Hamilton_fed_76.txt
Hamilton_fed_36.txt
Hamilton_fed_78.txt
dispt_fed_49.txt
Hamilton_fed_24.txt
Hamilton_fed_59.txt
Hamilton_fed_60.txt
Hamilton_fed_81.txt
Madison_fed_10.txt
dispt_fed_55.txt
Hamilton_fed_32.txt
Hamilton_fed_9.txt
Hamilton_fed_26.txt
Hamilton_fed_15.txt
Hamilton_fed_8.txt
Hamilton_fed_22.txt
Hamilton_fed_35.txt
Madison_fed_38.txt
Hamilton_fed_70.txt
Hamilton_fed_85.txt
Madison_fed_62.txt
dispt_fed_62.txt
Madison_fed_42.txt
dispt_fed_52.txt
dispt_fed_53.txt
Madison_fed_14.txt
Madison_fed_43.txt
Madison_fed_48.txt
dispt_fed_63.txt
Madison_fed_41.txt
HM_fed_20.txt
HM_fed_18.txt
HM_fed_19.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance
hclust (*, "single")

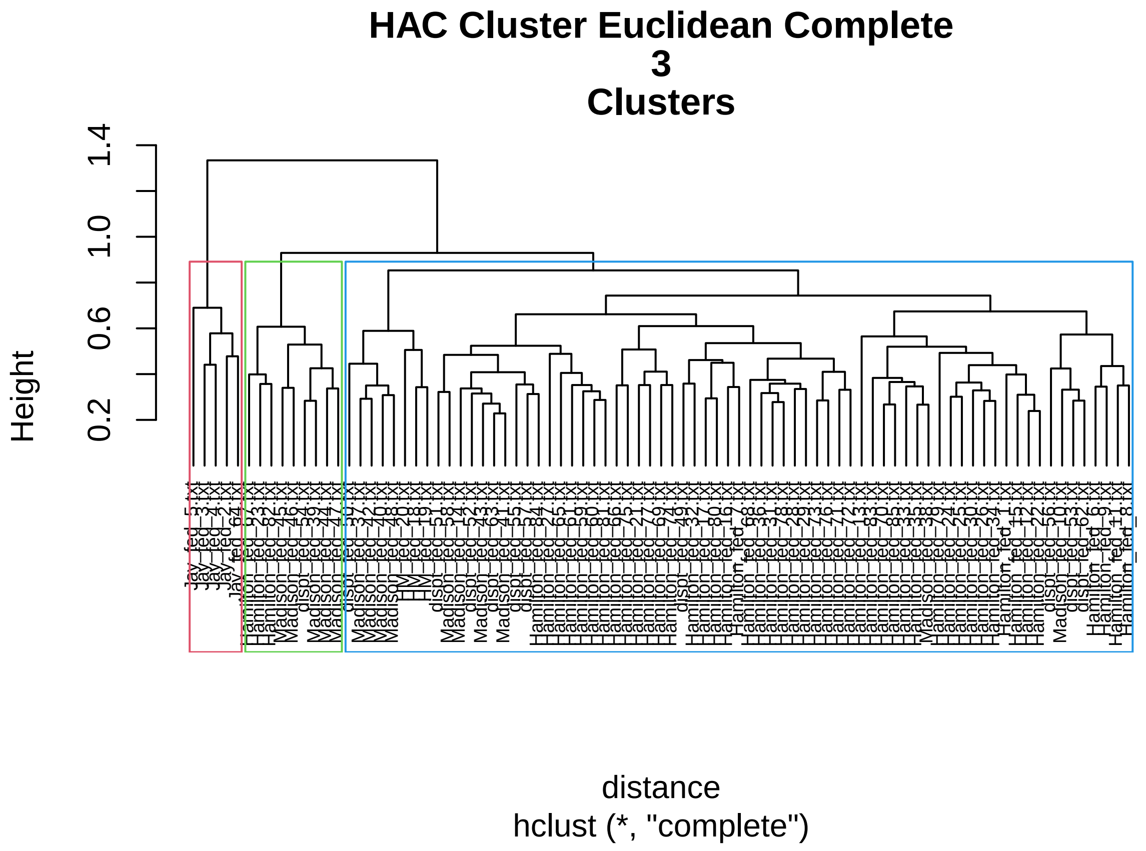

HAC Cluster Maximum Complete
2
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

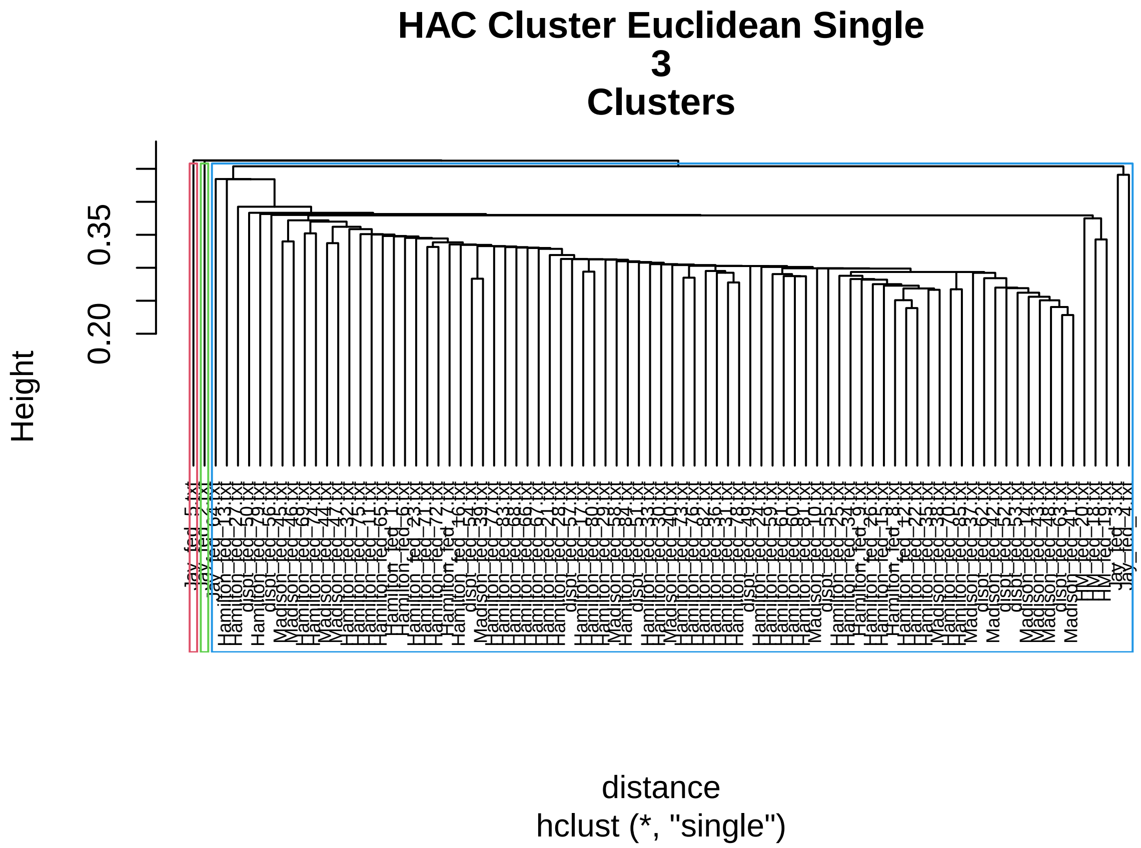

HAC Cluster Manhattan Complete
2
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")

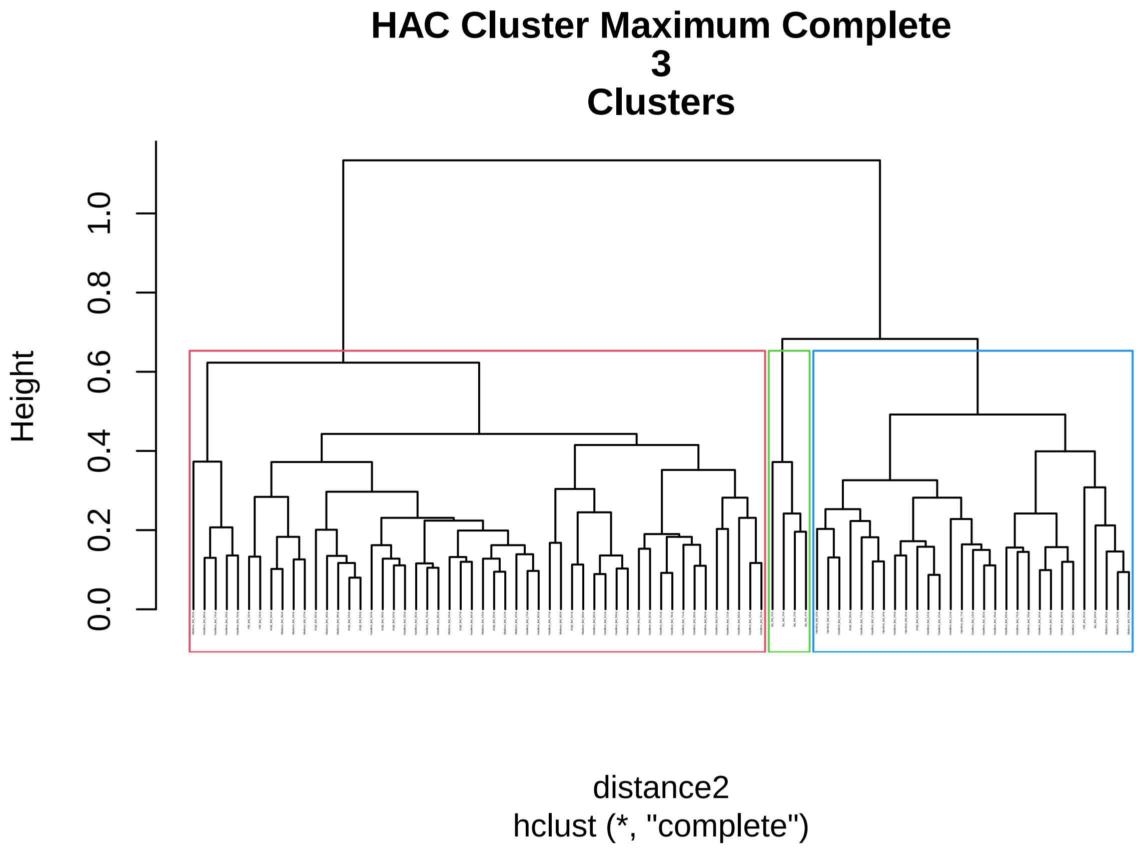

HAC Cluster Canberra Complete
2
Clusters
Height
15
25
35
distance4
hclust (*, "complete")

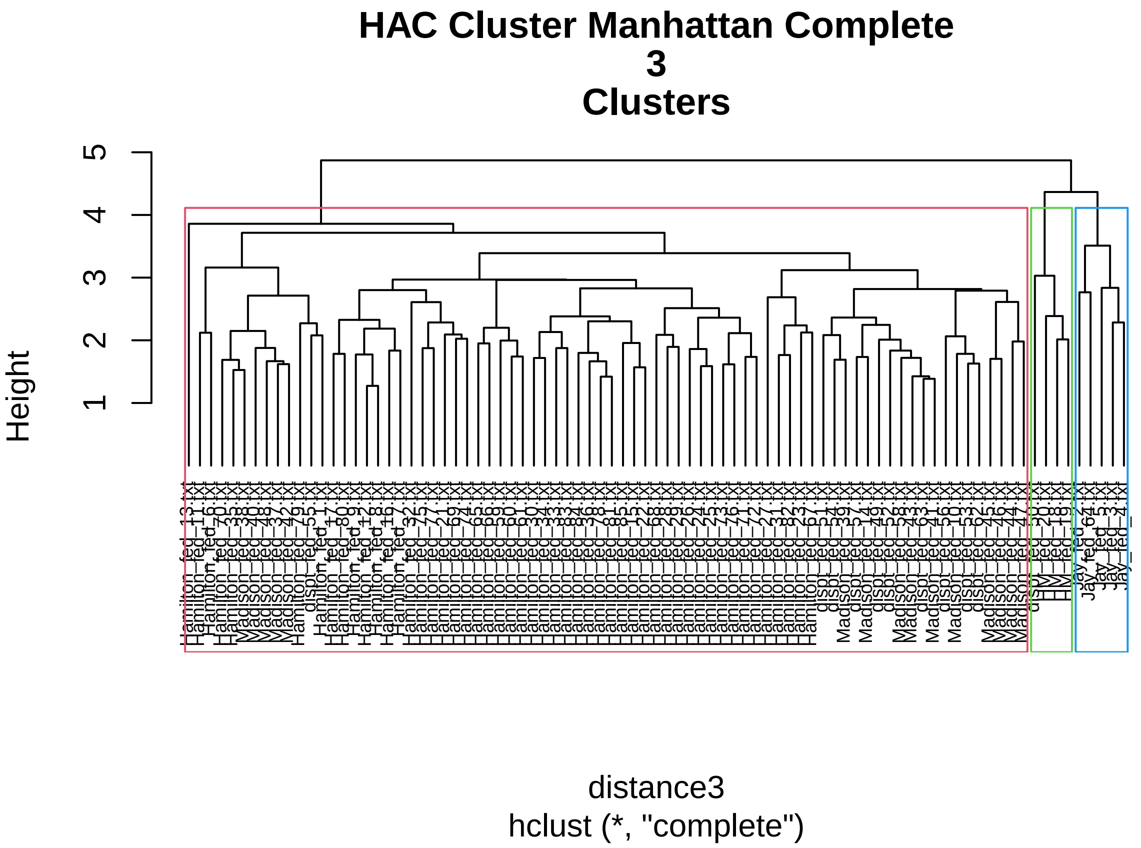

HAC Cluster Minkowski Complete
2
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")

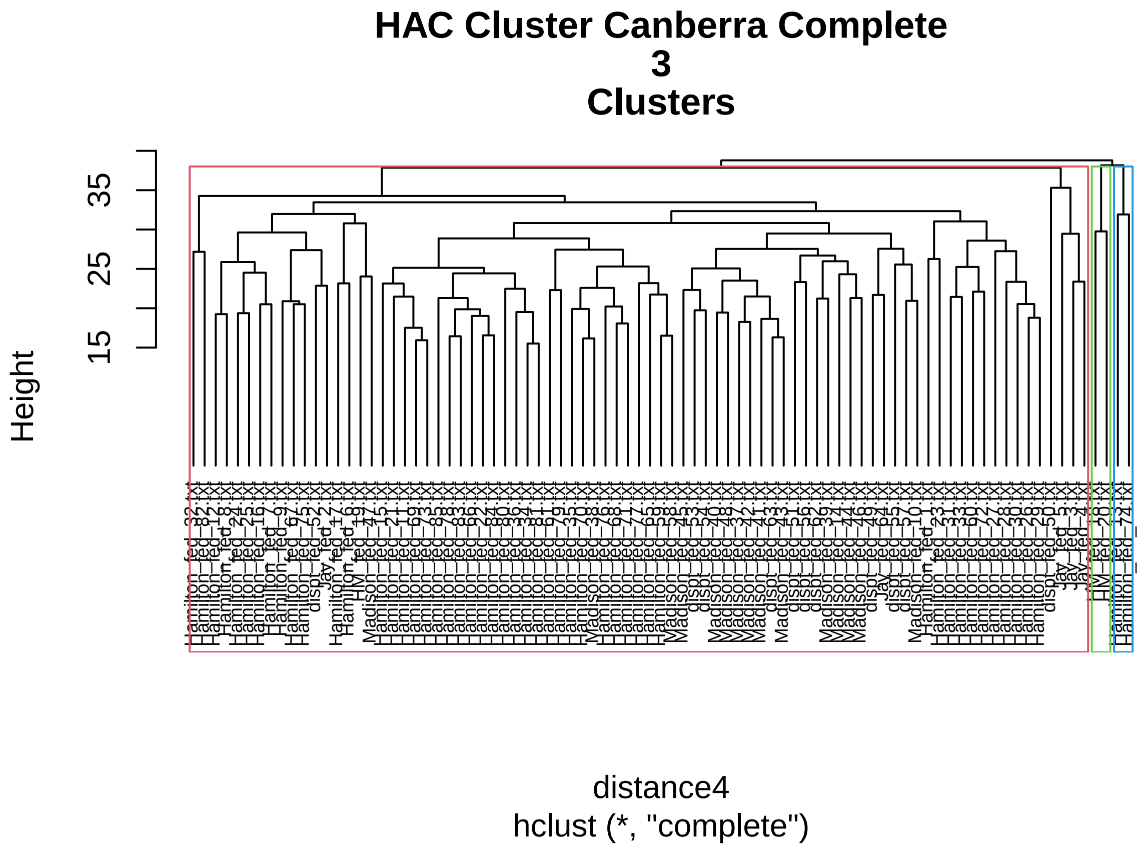

HAC Cluster Maximum Complete
2
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance6
hclust (*, "complete")

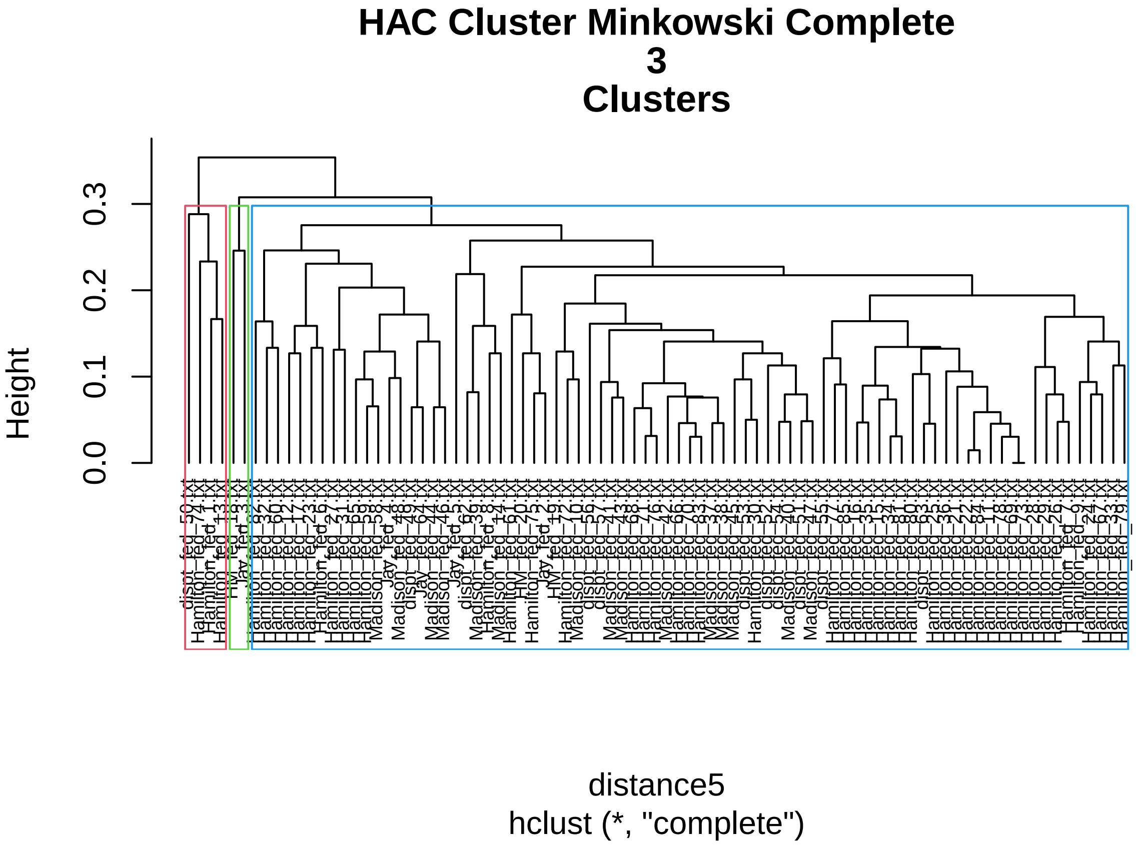

HAC Cluster Euclidean Complete
3
Clusters
Height
0.2
0.6
1.0
1.4
distance
hclust (*, "complete")

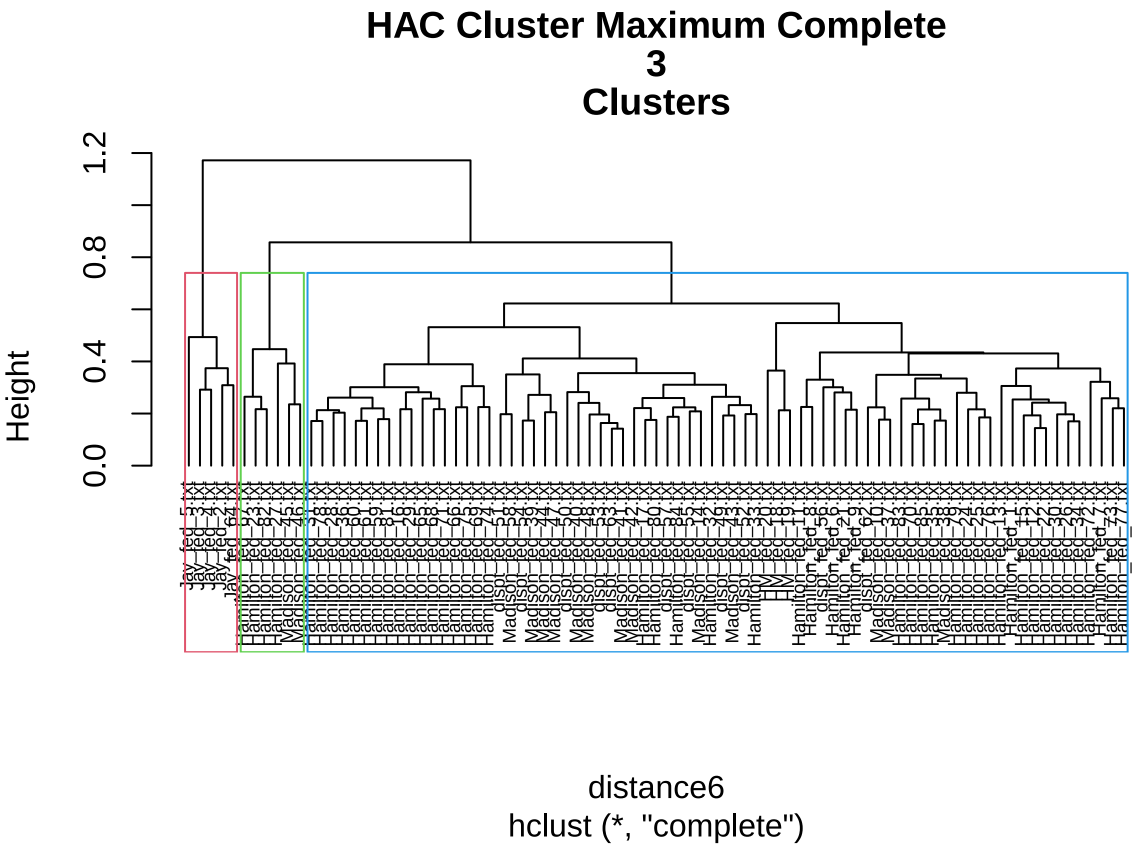

HAC Cluster Euclidean Single
3
Clusters
Height
0.20
0.35
distance
hclust (*, "single")

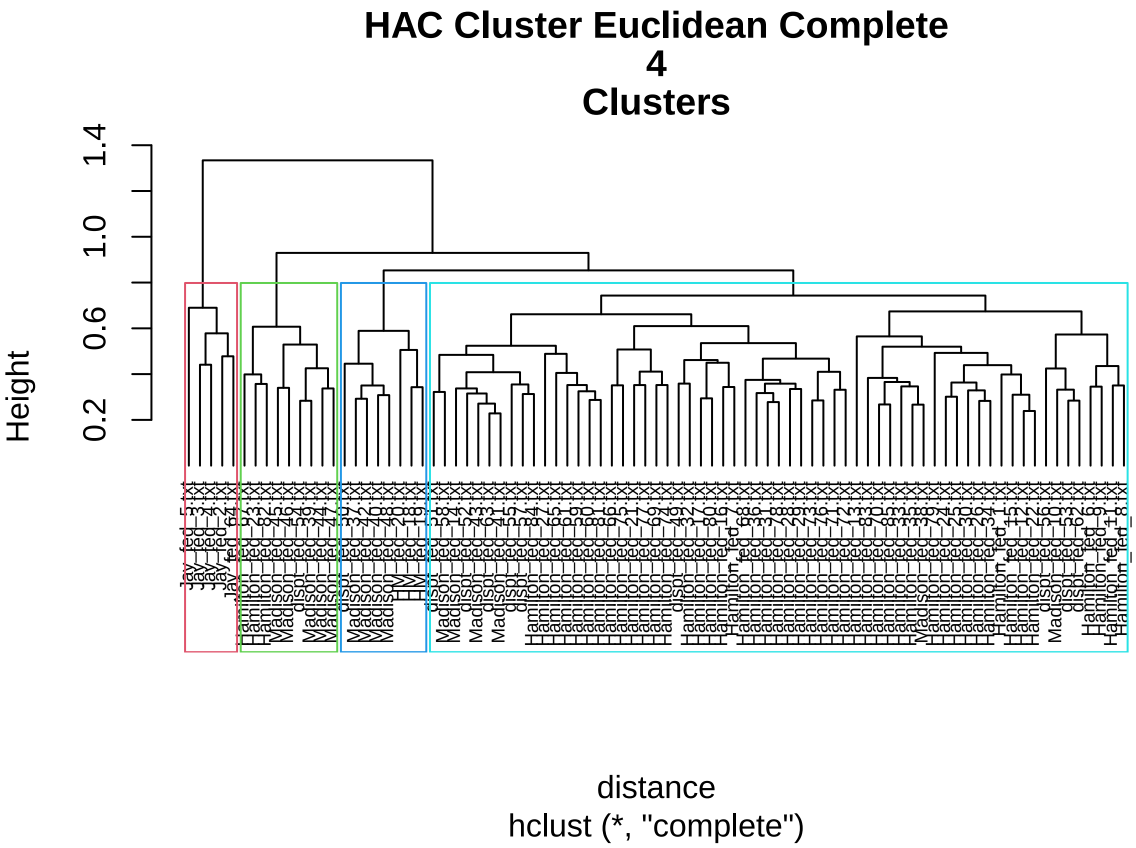

HAC Cluster Maximum Complete
3
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

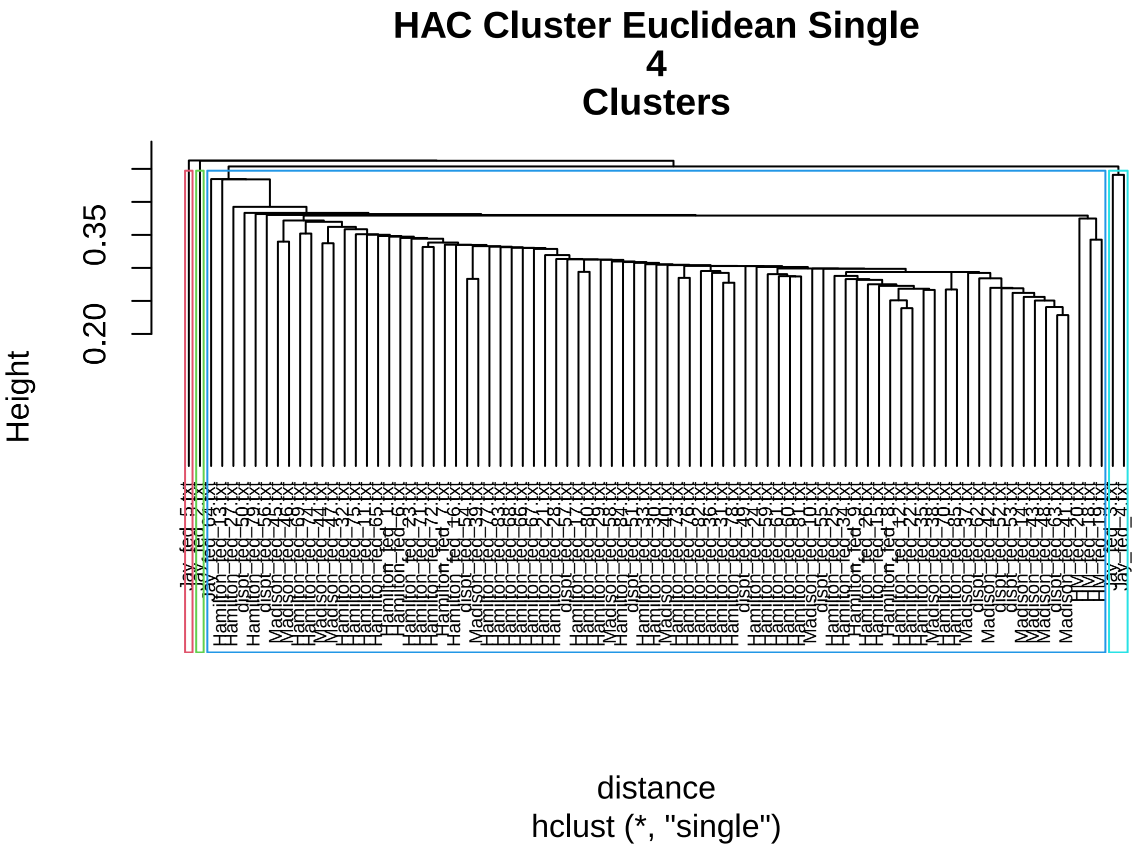

HAC Cluster Manhattan Complete
3
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")

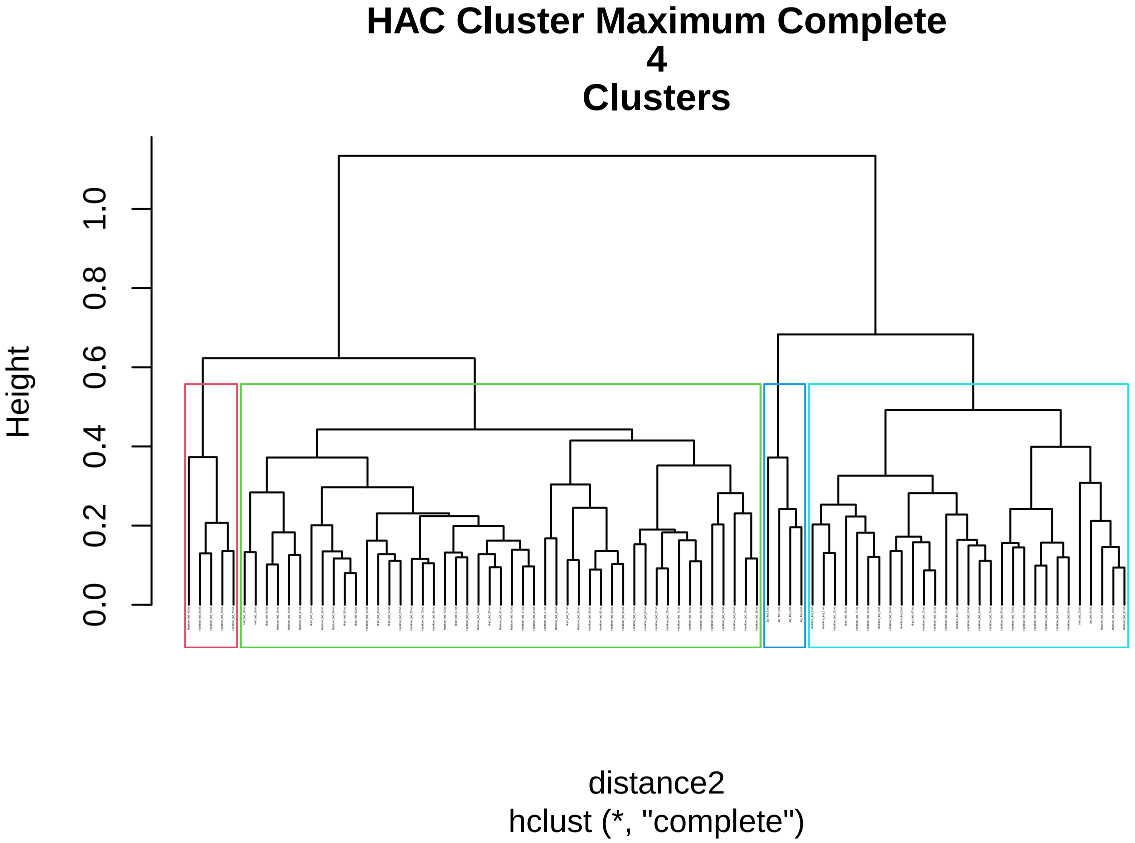

HAC Cluster Canberra Complete
3
Clusters
Height
15
25
35
distance4
hclust (*, "complete")

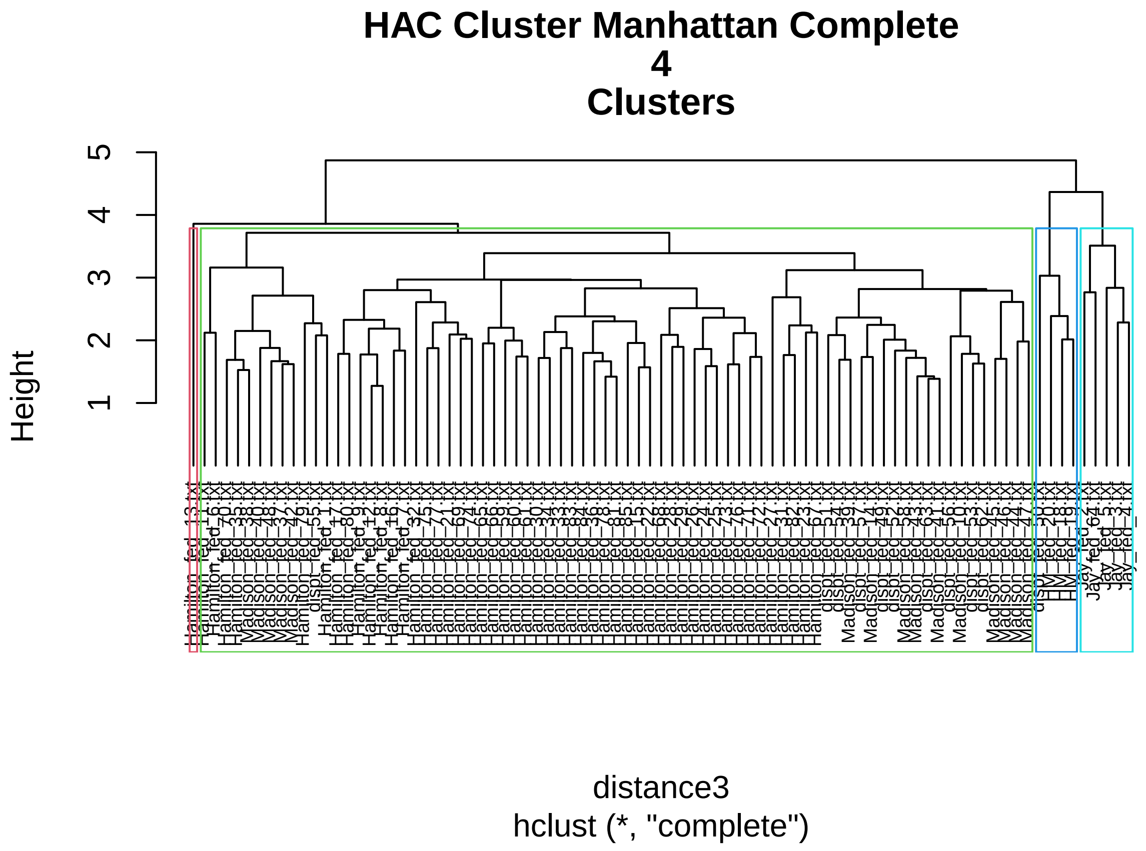
HAC Cluster Minkowski Complete
3
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")

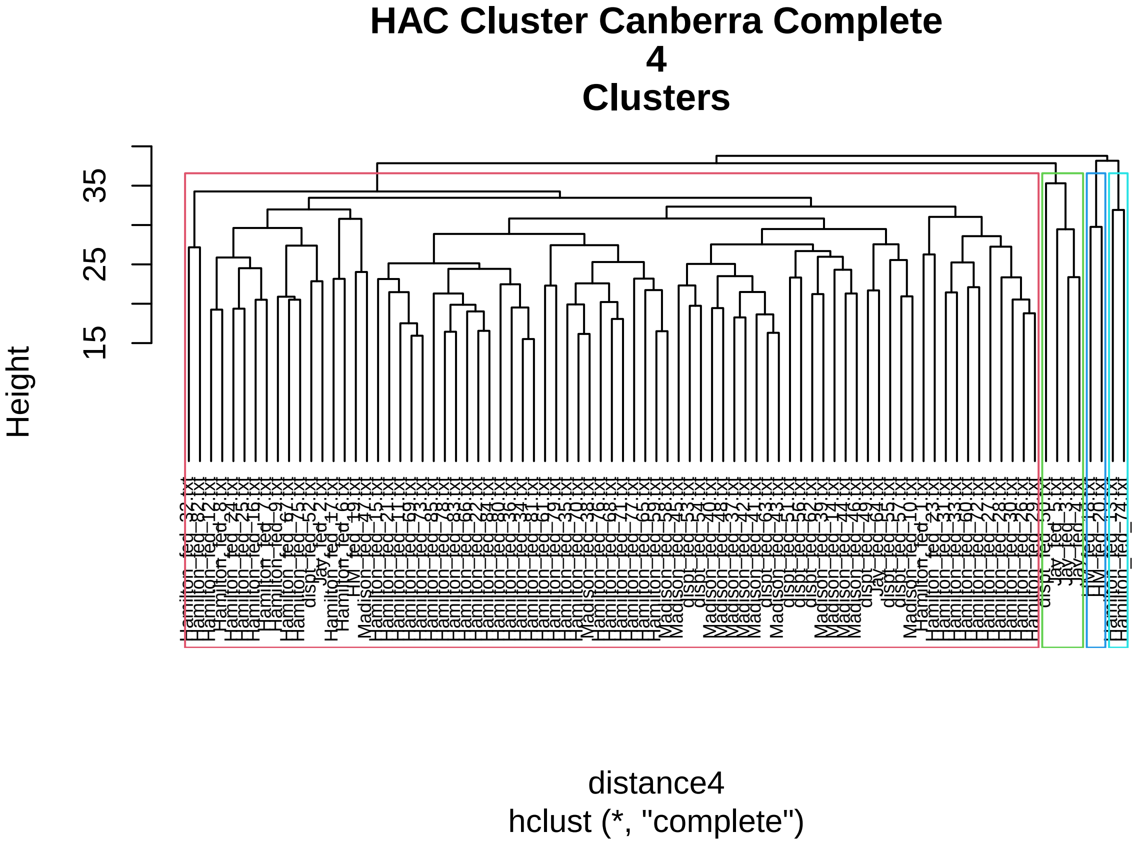

HAC Cluster Maximum Complete
3
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance6
hclust (*, "complete")

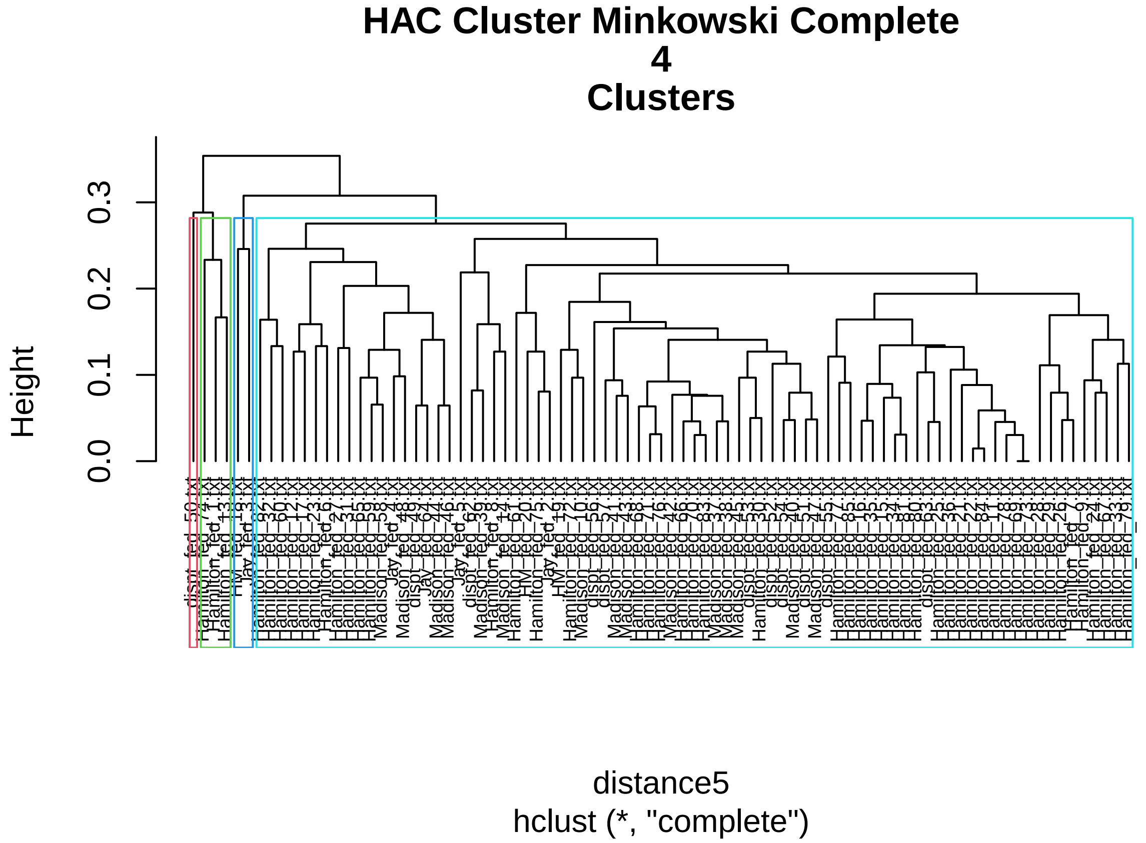

HAC Cluster Euclidean Complete
4
Clusters
Height
1.4
1.0
0.6
0.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance
hclust (*, "complete")

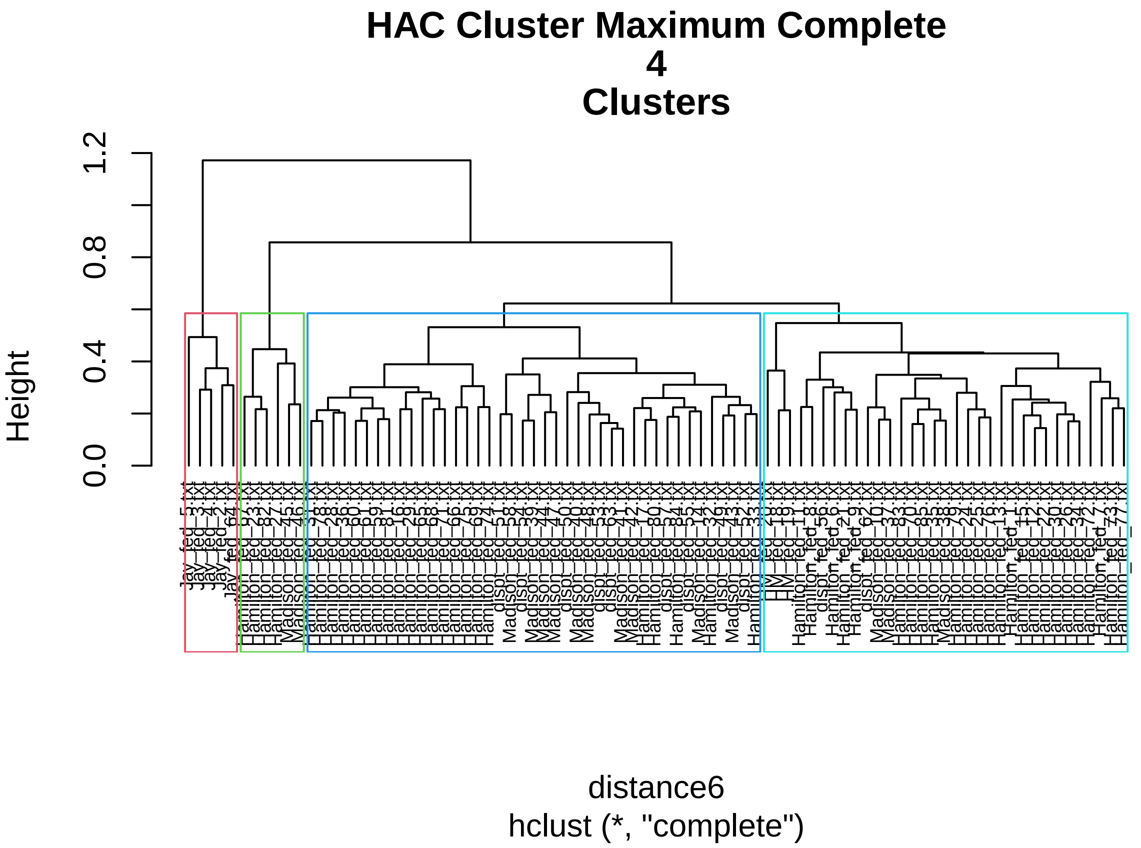

HAC Cluster Euclidean Single
4
Clusters
Height
0.20
0.35
Jay_fed_5.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_13.txt
Hamilton_fed_27.txt
dispt_fed_50.txt
Hamilton_fed_79.txt
dispt_fed_56.txt
Madison_fed_45.txt
Madison_fed_46.txt
Hamilton_fed_69.txt
Hamilton_fed_74.txt
Madison_fed_44.txt
Madison_fed_47.txt
Hamilton_fed_32.txt
Hamilton_fed_75.txt
Hamilton_fed_11.txt
Hamilton_fed_65.txt
Hamilton_fed_1.txt
Hamilton_fed_33.txt
Hamilton_fed_21.txt
Hamilton_fed_72.txt
Hamilton_fed_7.txt
Hamilton_fed_16.txt
dispt_fed_39.txt
Madison_fed_37.txt
Hamilton_fed_83.txt
Hamilton_fed_68.txt
Hamilton_fed_60.txt
Hamilton_fed_61.txt
Hamilton_fed_28.txt
dispt_fed_57.txt
Hamilton_fed_17.txt
Hamilton_fed_80.txt
Madison_fed_58.txt
Hamilton_fed_84.txt
dispt_fed_51.txt
Hamilton_fed_30.txt
Hamilton_fed_30.txt
Madison_fed_40.txt
Hamilton_fed_73.txt
Hamilton_fed_76.txt
Hamilton_fed_86.txt
Hamilton_fed_36.txt
Hamilton_fed_78.txt
dispt_fed_49.txt
Hamilton_fed_24.txt
Hamilton_fed_59.txt
Hamilton_fed_60.txt
Hamilton_fed_81.txt
Madison_fed_10.txt
dispt_fed_55.txt
Hamilton_fed_32.txt
Hamilton_fed_9.txt
Hamilton_fed_26.txt
Hamilton_fed_15.txt
Hamilton_fed_8.txt
Hamilton_fed_22.txt
Hamilton_fed_35.txt
Madison_fed_38.txt
Hamilton_fed_70.txt
Hamilton_fed_85.txt
Madison_fed_14.txt
dispt_fed_62.txt
Madison_fed_42.txt
dispt_fed_52.txt
dispt_fed_53.txt
Madison_fed_43.txt
Madison_fed_48.txt
dispt_fed_63.txt
Madison_fed_41.txt
HM_fed_20.txt
HM_fed_19.txt
HM_fed_18.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance
hclust (*, "single")

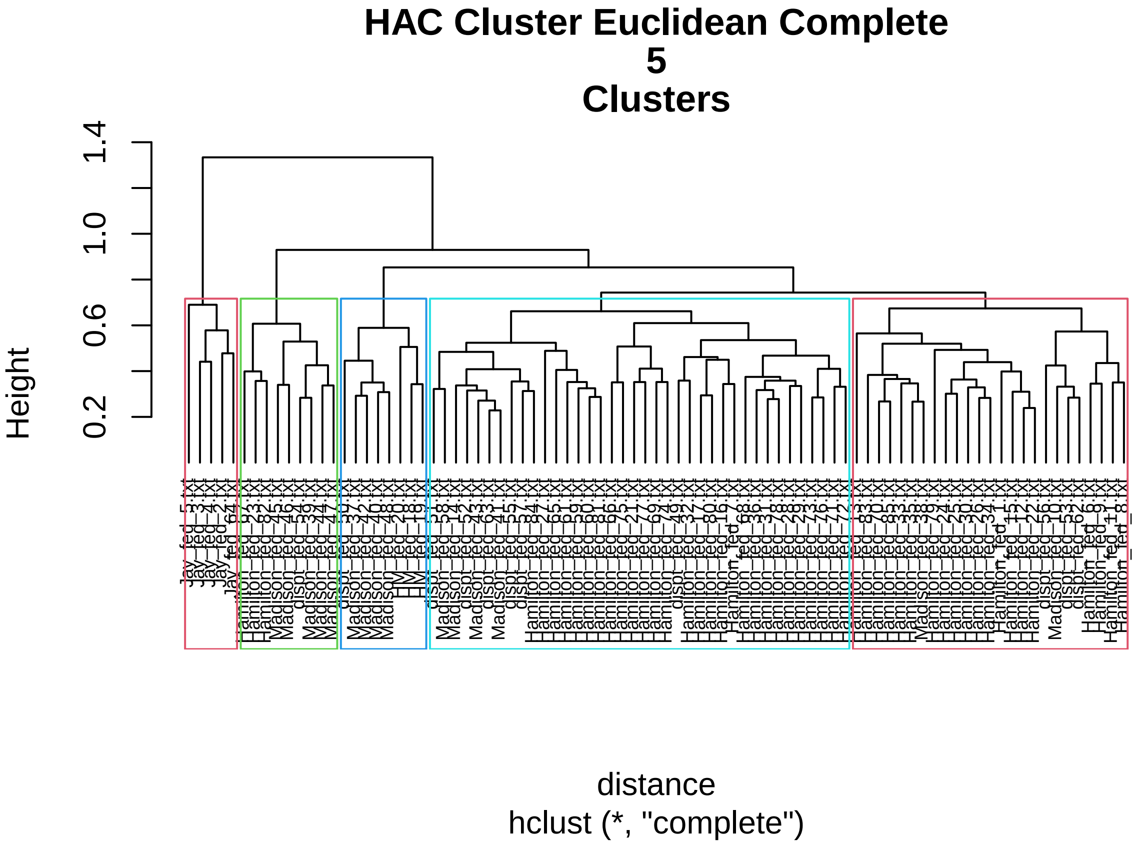

HAC Cluster Maximum Complete
4
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

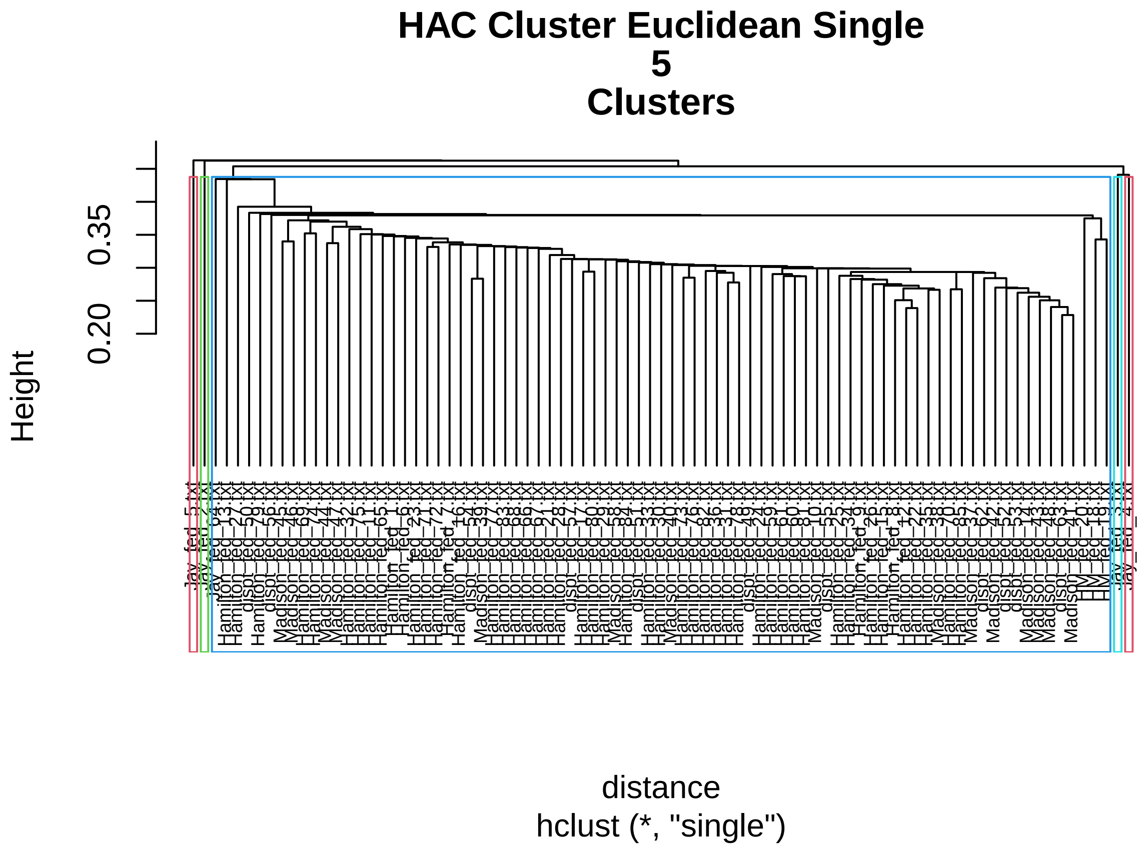

HAC Cluster Manhattan Complete
4
Clusters
Height
1
2
3
4
5
Jay_fed_2.txt
Jay_fed_64.txt
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance3
hclust (*, "complete")

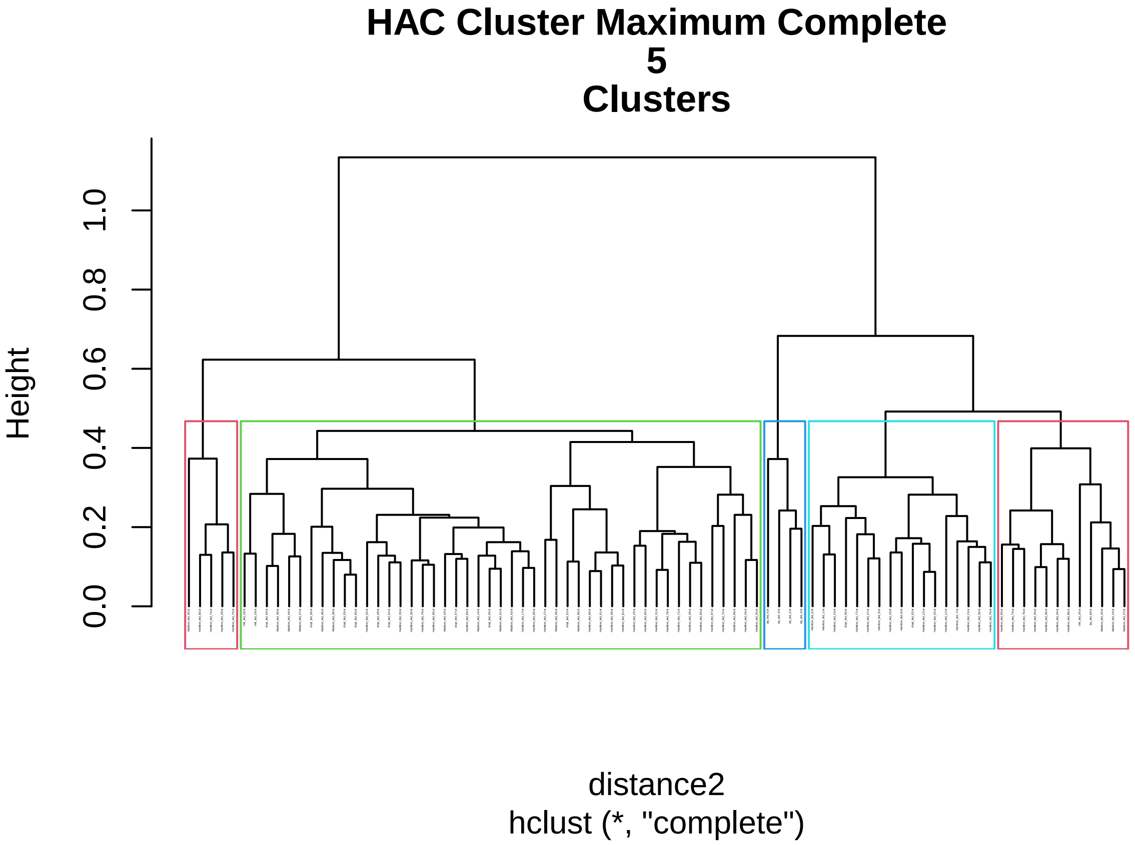

HAC Cluster Canberra Complete
4
Clusters
Height
15
25
35
Hamilton_fed_32.txt
Hamilton_fed_82.txt
Hamilton_fed_12.txt
Hamilton_fed_8.txt
Hamilton_fed_24.txt
Hamilton_fed_25.txt
Hamilton_fed_16.txt
Hamilton_fed_7.txt
Hamilton_fed_9.txt
Hamilton_fed_67.txt
Hamilton_fed_75.txt
dispt_fed_52.txt
Jay_fed_2.txt
Hamilton_fed_17.txt
Hamilton_fed_18.txt
HM_fed_19.txt
Madison_fed_47.txt
Hamilton_fed_15.txt
Hamilton_fed_21.txt
Hamilton_fed_69.txt
Hamilton_fed_73.txt
Hamilton_fed_85.txt
Hamilton_fed_78.txt
Hamilton_fed_83.txt
Hamilton_fed_59.txt
Hamilton_fed_64.txt
Hamilton_fed_80.txt
Hamilton_fed_36.txt
Hamilton_fed_84.txt
Hamilton_fed_81.txt
Hamilton_fed_79.txt
Hamilton_fed_35.txt
Madison_fed_70.txt
Hamilton_fed_38.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_77.txt
Hamilton_fed_65.txt
Hamilton_fed_58.txt
Madison_fed_43.txt
Madison_fed_45.txt
dispt_fed_53.txt
dispt_fed_54.txt
Madison_fed_40.txt
Madison_fed_37.txt
Madison_fed_42.txt
Madison_fed_41.txt
dispt_fed_63.txt
Madison_fed_51.txt
dispt_fed_56.txt
dispt_fed_62.txt
dispt_fed_39.txt
Madison_fed_14.txt
Madison_fed_44.txt
Madison_fed_46.txt
Madison_fed_49.txt
dispt_fed_64.txt
Jay_fed_64.txt
dispt_fed_55.txt
dispt_fed_57.txt
Madison_fed_10.txt
Hamilton_fed_1.txt
Hamilton_fed_23.txt
Hamilton_fed_31.txt
Hamilton_fed_33.txt
Hamilton_fed_30.txt
Hamilton_fed_29.txt
Hamilton_fed_27.txt
Hamilton_fed_28.txt
Hamilton_fed_30.txt
Hamilton_fed_26.txt
Hamilton_fed_28.txt
dispt_fed_50.txt
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
HM_fed_18.txt
HM_fed_20.txt
Hamilton_fed_13.txt
Hamilton_fed_74.txt
distance4
hclust (*, "complete")

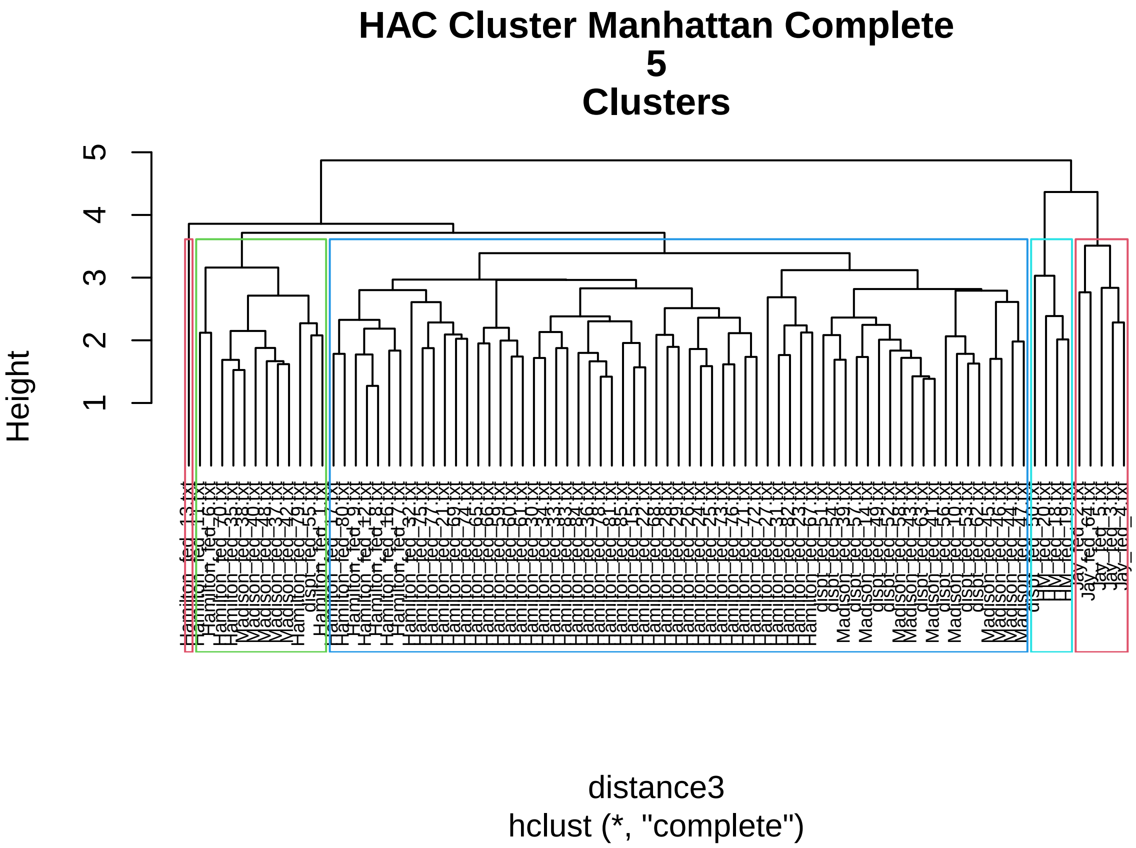

HAC Cluster Minkowski Complete
4
Clusters
Height
0.0
0.1
0.2
0.3
dispt_fed_50.txt
Hamilton_fed_74.txt
Hamilton_fed_1.txt
Hamilton_fed_13.txt
HM_fed_18.txt
Jay_fed_3.txt
Hamilton_fed_82.txt
Hamilton_fed_36.txt
Hamilton_fed_60.txt
Hamilton_fed_12.txt
Hamilton_fed_17.txt
Hamilton_fed_23.txt
Hamilton_fed_6.txt
Hamilton_fed_31.txt
Hamilton_fed_65.txt
Hamilton_fed_59.txt
Madison_fed_58.txt
Jay_fed_4.txt
Madison_fed_48.txt
dispt_fed_49.txt
Jay_fed_64.txt
Madison_fed_44.txt
Madison_fed_46.txt
Jay_fed_5.txt
dispt_fed_62.txt
Madison_fed_39.txt
Hamilton_fed_8.txt
Madison_fed_14.txt
Hamilton_fed_61.txt
HM_fed_20.txt
Hamilton_fed_7.txt
Jay_fed_2.txt
HM_fed_19.txt
Hamilton_fed_72.txt
Madison_fed_10.txt
dispt_fed_57.txt
dispt_fed_55.txt
Madison_fed_41.txt
Madison_fed_43.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_49.txt
Madison_fed_66.txt
Hamilton_fed_70.txt
Hamilton_fed_83.txt
Madison_fed_38.txt
Madison_fed_45.txt
dispt_fed_53.txt
Hamilton_fed_30.txt
dispt_fed_52.txt
dispt_fed_40.txt
Madison_fed_51.txt
dispt_fed_47.txt
Madison_fed_55.txt
Hamilton_fed_67.txt
Hamilton_fed_85.txt
Hamilton_fed_35.txt
Hamilton_fed_15.txt
Hamilton_fed_34.txt
Hamilton_fed_80.txt
dispt_fed_63.txt
Hamilton_fed_25.txt
Hamilton_fed_36.txt
Hamilton_fed_21.txt
Hamilton_fed_84.txt
Hamilton_fed_11.txt
Hamilton_fed_78.txt
Hamilton_fed_69.txt
Hamilton_fed_28.txt
Hamilton_fed_29.txt
Hamilton_fed_26.txt
Hamilton_fed_7.txt
Hamilton_fed_32.txt
Hamilton_fed_24.txt
Hamilton_fed_33.txt
Hamilton_fed_79.txt
distance5
hclust (*, "complete")

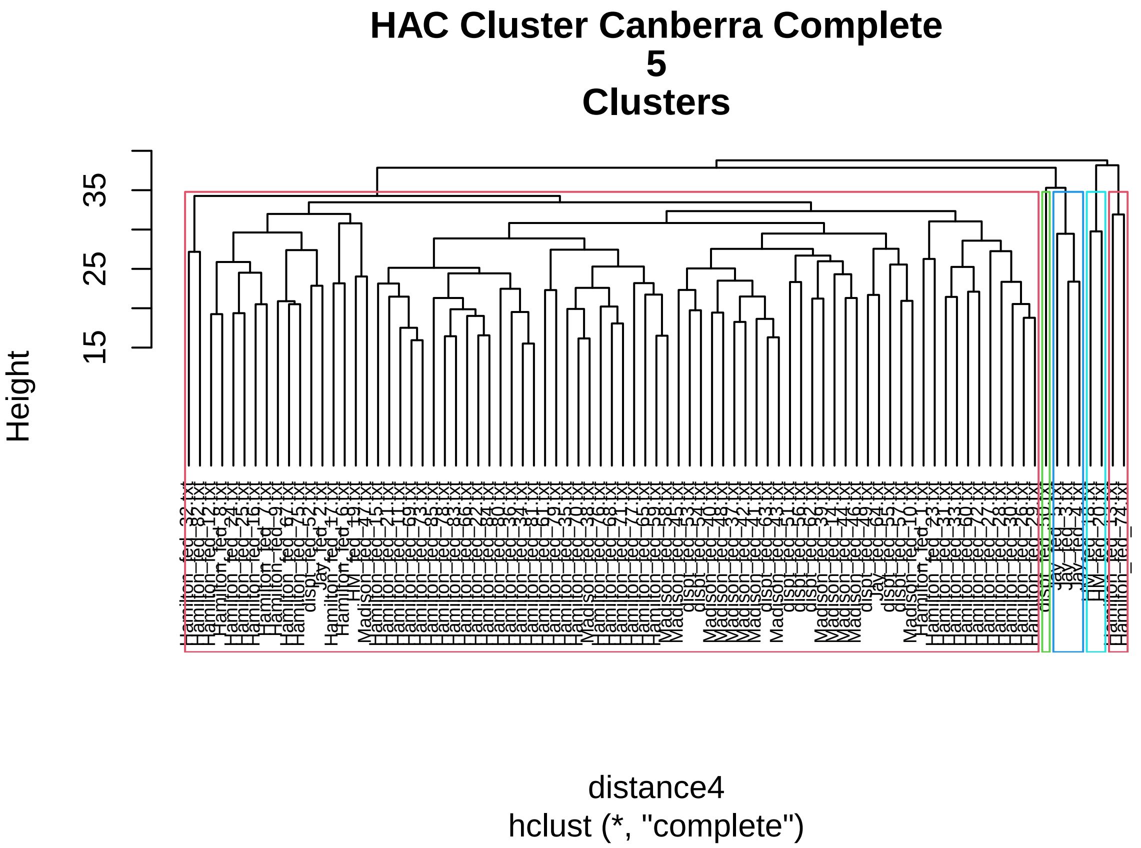
HAC Cluster Maximum Complete
4
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_23.txt
Hamilton_fed_82.txt
Madison_fed_45.txt
Madison_fed_46.txt
Hamilton_fed_78.txt
Hamilton_fed_38.txt
Hamilton_fed_60.txt
Hamilton_fed_61.txt
Hamilton_fed_59.txt
Hamilton_fed_81.txt
Hamilton_fed_65.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_66.txt
Hamilton_fed_69.txt
Hamilton_fed_74.txt
dispt_fed_51.txt
Madison_fed_58.txt
Madison_fed_44.txt
Madison_fed_47.txt
dispt_fed_50.txt
Madison_fed_40.txt
dispt_fed_63.txt
Madison_fed_42.txt
Hamilton_fed_84.txt
dispt_fed_55.txt
Madison_fed_14.txt
dispt_fed_43.txt
dispt_fed_52.txt
HM_fed_18.txt
Hamilton_fed_11.txt
Hamilton_fed_8.txt
dispt_fed_56.txt
Hamilton_fed_9.txt
dispt_fed_62.txt
Madison_fed_10.txt
Madison_fed_37.txt
Hamilton_fed_85.txt
Hamilton_fed_35.txt
Hamilton_fed_27.txt
Hamilton_fed_25.txt
Hamilton_fed_13.txt
Hamilton_fed_1.txt
Hamilton_fed_15.txt
Hamilton_fed_22.txt
Hamilton_fed_30.txt
Hamilton_fed_26.txt
Hamilton_fed_72.txt
Hamilton_fed_73.txt
Hamilton_fed_77.txt
distance6
hclust (*, "complete")

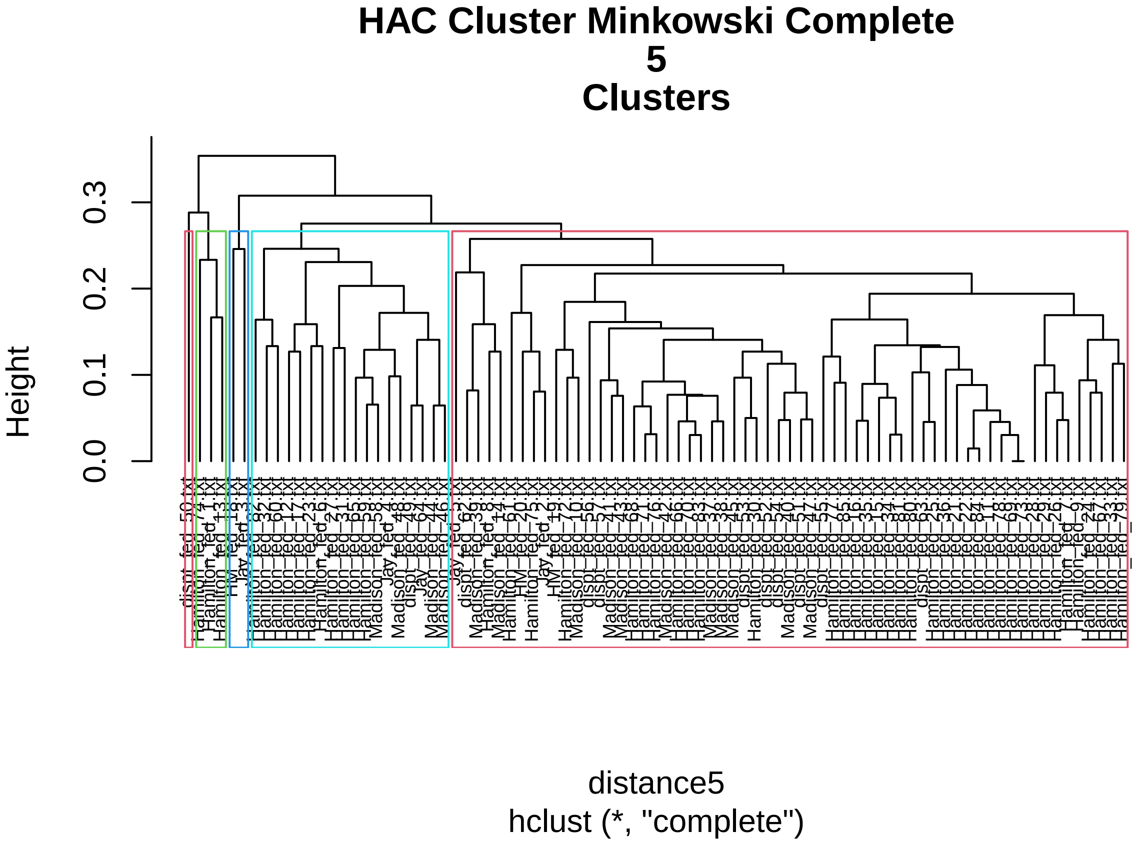

HAC Cluster Euclidean Complete
5
Clusters
Height
1.4
1.0
0.6
0.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_23.txt
Hamilton_fed_62.txt
Madison_fed_45.txt
Madison_fed_54.txt
dispt_fed_49.txt
Madison_fed_39.txt
Madison_fed_44.txt
Madison_fed_47.txt
dispt_fed_37.txt
Madison_fed_42.txt
Madison_fed_40.txt
Madison_fed_48.txt
HM_fed_18.txt
HM_fed_19.txt
dispt_fed_51.txt
Madison_fed_58.txt
Madison_fed_14.txt
dispt_fed_52.txt
Madison_fed_63.txt
dispt_fed_41.txt
Madison_fed_55.txt
dispt_fed_57.txt
dispt_fed_84.txt
Hamilton_fed_65.txt
Hamilton_fed_61.txt
Hamilton_fed_59.txt
Hamilton_fed_80.txt
Hamilton_fed_66.txt
Hamilton_fed_75.txt
Hamilton_fed_21.txt
Hamilton_fed_77.txt
Hamilton_fed_74.txt
dispt_fed_49.txt
Hamilton_fed_32.txt
Hamilton_fed_30.txt
Hamilton_fed_81.txt
Hamilton_fed_17.txt
Hamilton_fed_68.txt
Hamilton_fed_36.txt
Hamilton_fed_31.txt
Hamilton_fed_28.txt
Hamilton_fed_29.txt
Hamilton_fed_73.txt
Hamilton_fed_76.txt
Hamilton_fed_72.txt
Hamilton_fed_83.txt
Hamilton_fed_70.txt
Hamilton_fed_85.txt
Hamilton_fed_35.txt
Madison_fed_38.txt
Hamilton_fed_79.txt
Hamilton_fed_24.txt
Hamilton_fed_25.txt
Hamilton_fed_30.txt
Hamilton_fed_34.txt
Hamilton_fed_1.txt
Hamilton_fed_15.txt
Hamilton_fed_12.txt
dispt_fed_56.txt
Madison_fed_10.txt
dispt_fed_53.txt
dispt_fed_62.txt
dispt_fed_63.txt
Hamilton_fed_9.txt
Hamilton_fed_11.txt
Hamilton_fed_8.txt
distance
hclust (*, "complete")

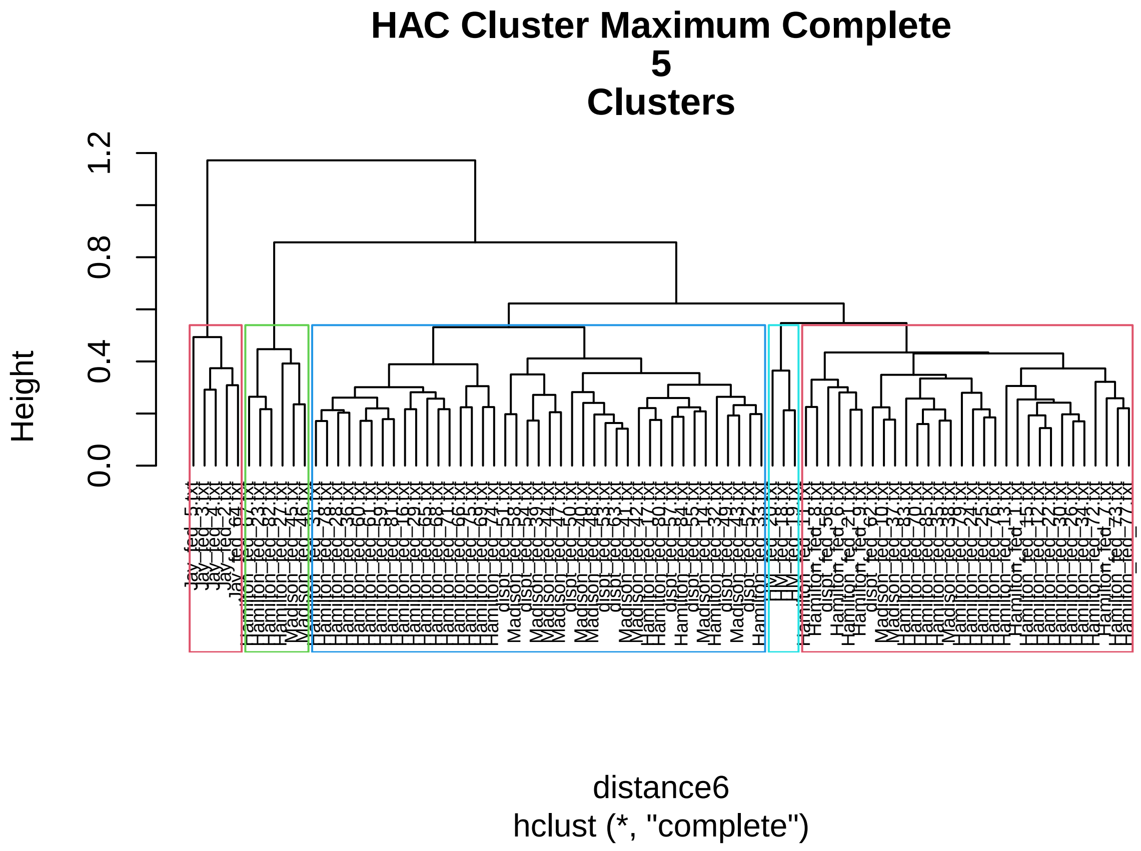

HAC Cluster Euclidean Single
5
Clusters
Height
0.20
0.35
Jay_fed_5.txt
Jay_fed_2.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance
hclust (*, "single")

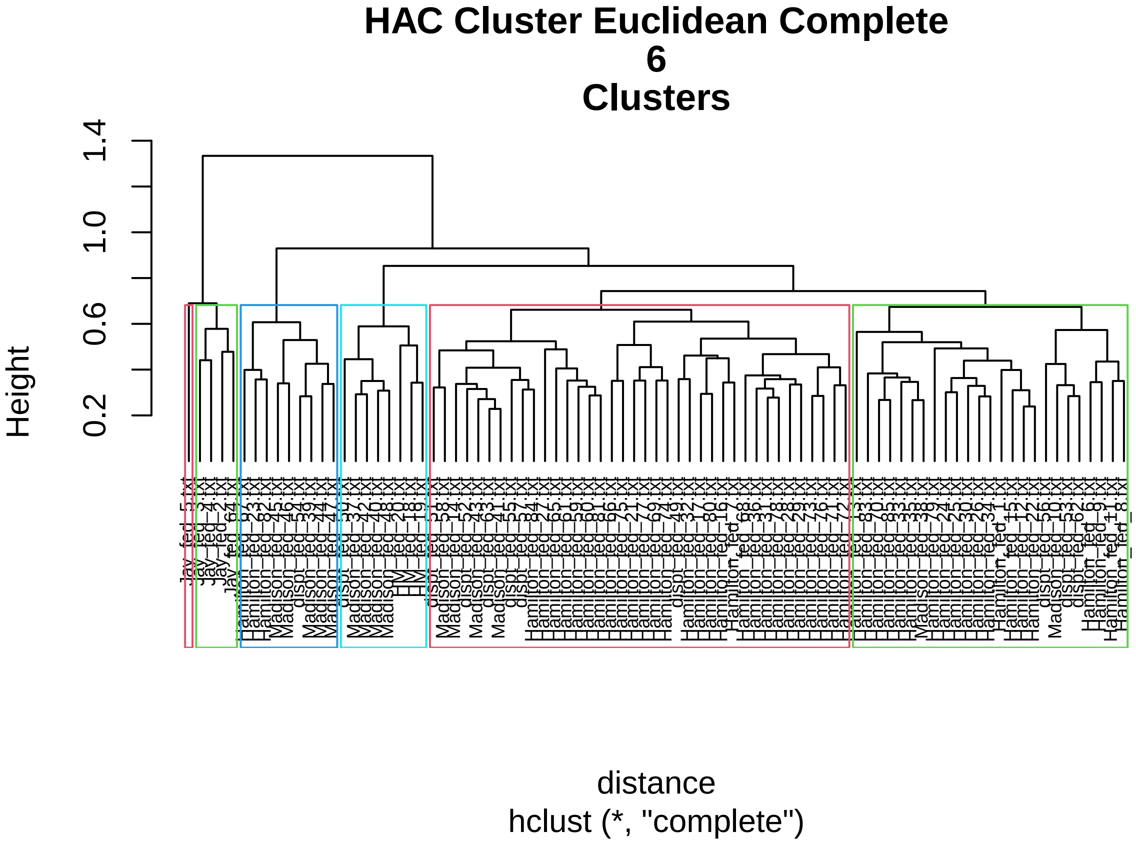

HAC Cluster Maximum Complete
5
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

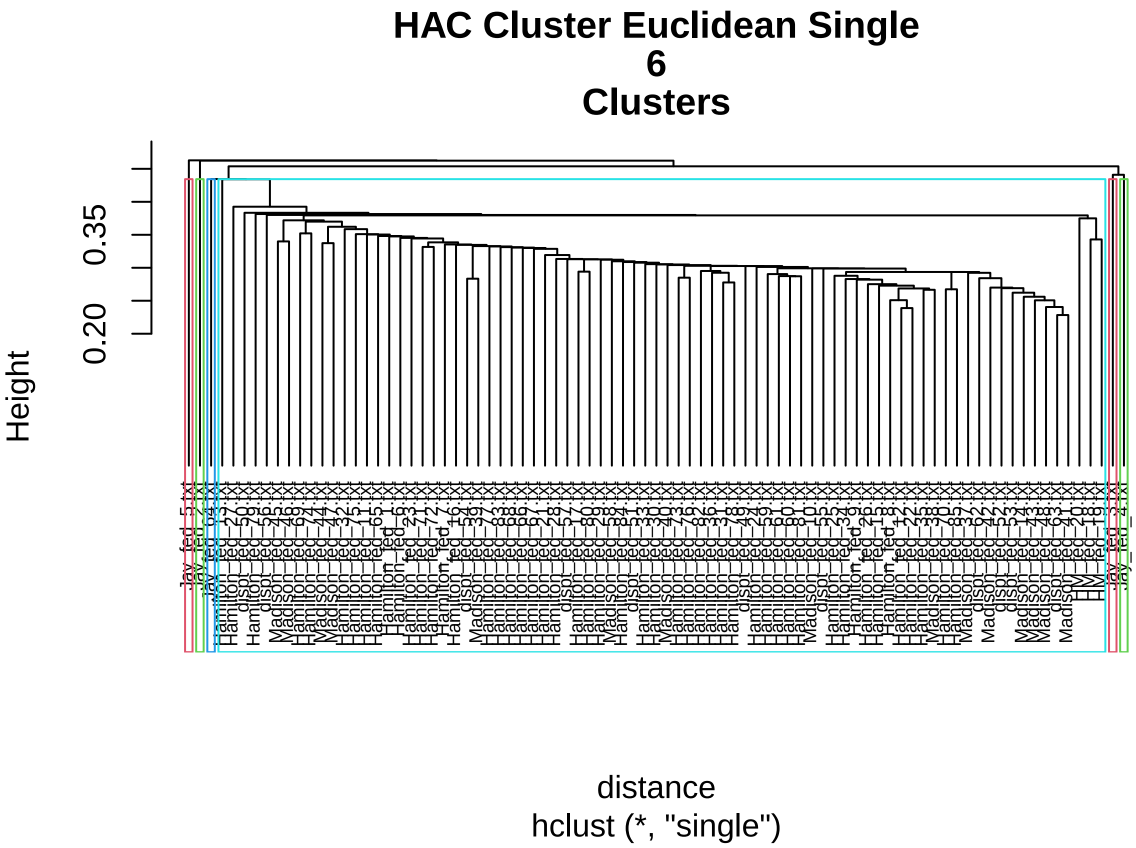

HAC Cluster Manhattan Complete
5
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")

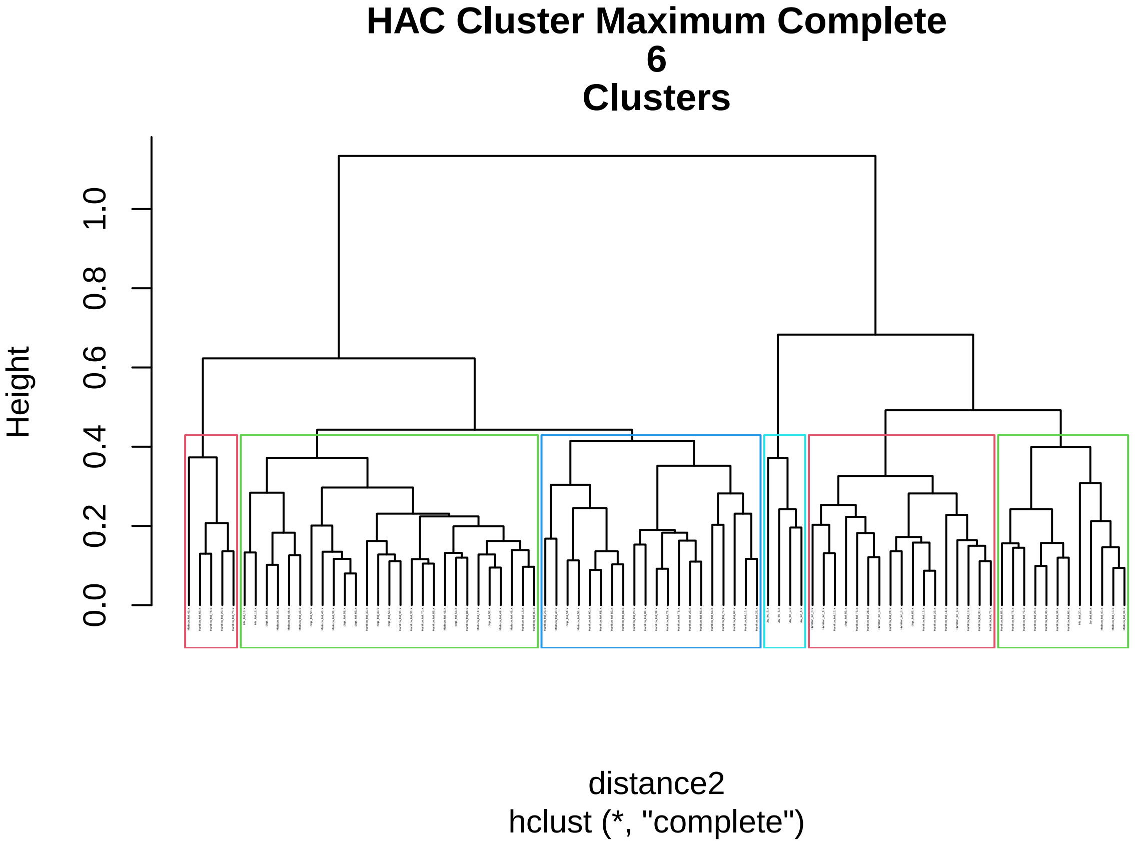

HAC Cluster Canberra Complete
5
Clusters
Height
15
25
35
Hamilton_fed_32.txt
Hamilton_fed_82.txt
Hamilton_fed_1.txt
Hamilton_fed_8.txt
Hamilton_fed_24.txt
Hamilton_fed_25.txt
Hamilton_fed_16.txt
Hamilton_fed_9.txt
Hamilton_fed_6.txt
Hamilton_fed_67.txt
Hamilton_fed_75.txt
dispt_fed_52.txt
Jay_fed_2.txt
Hamilton_fed_17.txt
Hamilton_fed_6.txt
HM_fed_18.txt
Madison_fed_47.txt
Hamilton_fed_15.txt
Hamilton_fed_21.txt
Hamilton_fed_69.txt
Hamilton_fed_73.txt
Hamilton_fed_85.txt
Hamilton_fed_78.txt
Hamilton_fed_83.txt
Hamilton_fed_59.txt
Hamilton_fed_84.txt
Hamilton_fed_80.txt
Hamilton_fed_36.txt
Hamilton_fed_34.txt
Hamilton_fed_81.txt
Hamilton_fed_79.txt
Hamilton_fed_35.txt
Madison_fed_70.txt
Hamilton_fed_38.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_77.txt
Hamilton_fed_65.txt
Madison_fed_58.txt
Madison_fed_45.txt
dispt_fed_53.txt
dispt_fed_54.txt
Madison_fed_40.txt
Madison_fed_37.txt
Madison_fed_42.txt
Madison_fed_41.txt
dispt_fed_63.txt
Madison_fed_51.txt
dispt_fed_56.txt
dispt_fed_62.txt
dispt_fed_39.txt
Madison_fed_14.txt
Madison_fed_44.txt
Madison_fed_46.txt
Madison_fed_49.txt
dispt_fed_64.txt
Jay_fed_64.txt
dispt_fed_55.txt
dispt_fed_57.txt
Madison_fed_10.txt
Hamilton_fed_1.txt
Hamilton_fed_23.txt
Hamilton_fed_31.txt
Hamilton_fed_33.txt
Hamilton_fed_60.txt
Hamilton_fed_72.txt
Hamilton_fed_27.txt
Hamilton_fed_28.txt
Hamilton_fed_30.txt
Hamilton_fed_26.txt
Hamilton_fed_29.txt
dispt_fed_50.txt
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
HM_fed_18.txt
HM_fed_20.txt
Hamilton_fed_13.txt
Hamilton_fed_74.txt
distance4
hclust (*, "complete")

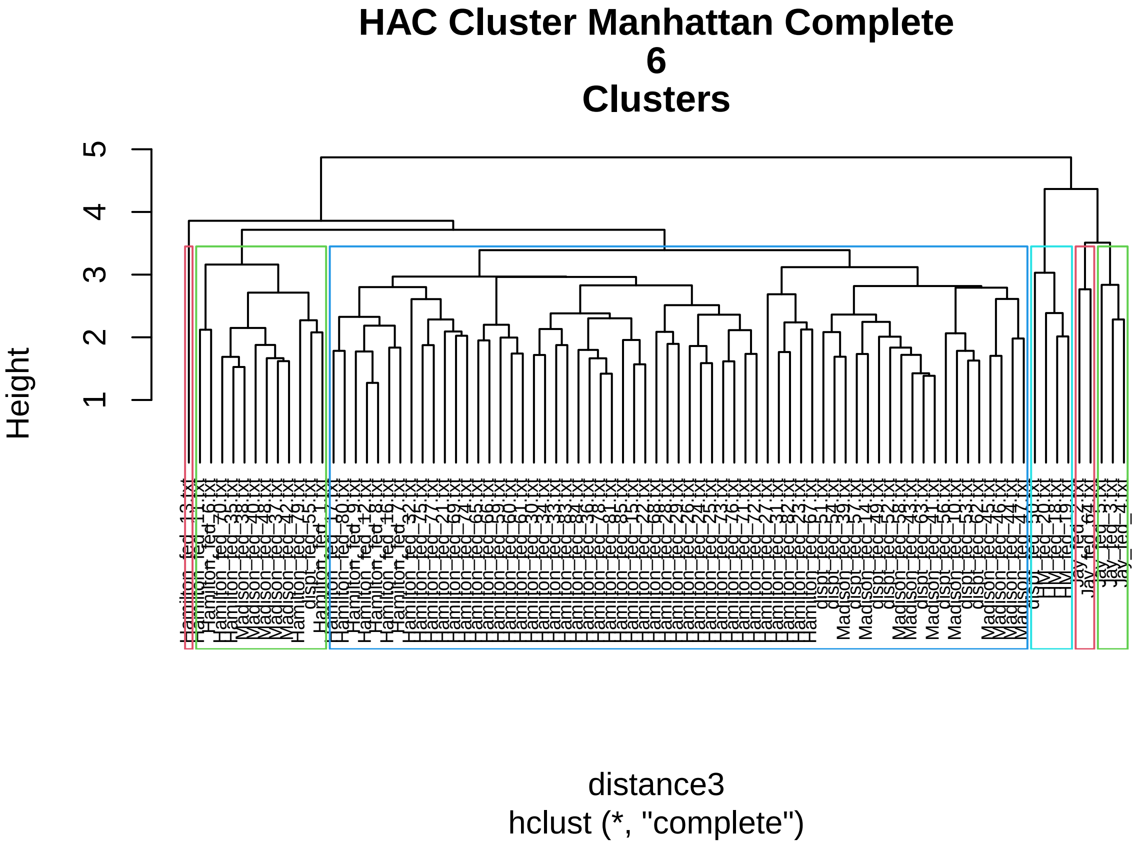

HAC Cluster Minkowski Complete
5
Clusters
Height
0.0
0.1
0.2
0.3
disp_fed_50.txt
Hamilton_fed_74.txt
Hamilton_fed_1.txt
Hamilton_fed_13.txt
HM_fed_18.txt
Jay_fed_3.txt
Hamilton_fed_82.txt
Hamilton_fed_30.txt
Hamilton_fed_60.txt
Hamilton_fed_12.txt
Hamilton_fed_17.txt
Hamilton_fed_23.txt
Hamilton_fed_6.txt
Hamilton_fed_21.txt
Hamilton_fed_31.txt
Hamilton_fed_65.txt
Hamilton_fed_59.txt
Madison_fed_58.txt
Jay_fed_4.txt
Madison_fed_48.txt
disp_fed_64.txt
Jay_fed_5.txt
Madison_fed_44.txt
Madison_fed_46.txt
Jay_fed_2.txt
disp_fed_63.txt
Madison_fed_36.txt
Hamilton_fed_8.txt
Madison_fed_14.txt
Hamilton_fed_61.txt
HM_fed_20.txt
Hamilton_fed_7.txt
Jay_fed_2.txt
HM_fed_19.txt
Hamilton_fed_72.txt
Madison_fed_10.txt
disp_fed_57.txt
disp_fed_59.txt
Madison_fed_41.txt
Madison_fed_43.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_49.txt
Madison_fed_66.txt
Hamilton_fed_70.txt
Hamilton_fed_83.txt
Madison_fed_38.txt
Madison_fed_45.txt
Madison_fed_53.txt
disp_fed_53.txt
Hamilton_fed_30.txt
disp_fed_52.txt
disp_fed_40.txt
Madison_fed_51.txt
disp_fed_47.txt
Madison_fed_55.txt
disp_fed_55.txt
Hamilton_fed_67.txt
Hamilton_fed_85.txt
Hamilton_fed_35.txt
Hamilton_fed_15.txt
Hamilton_fed_34.txt
Hamilton_fed_80.txt
Hamilton_fed_63.txt
disp_fed_25.txt
Hamilton_fed_36.txt
Hamilton_fed_21.txt
Hamilton_fed_84.txt
Hamilton_fed_81.txt
Hamilton_fed_78.txt
Hamilton_fed_69.txt
Hamilton_fed_78.txt
Hamilton_fed_29.txt
Hamilton_fed_26.txt
Hamilton_fed_7.txt
Hamilton_fed_9.txt
Hamilton_fed_24.txt
Hamilton_fed_67.txt
Hamilton_fed_33.txt
Hamilton_fed_79.txt
distance5
hclust (*, "complete")


HAC Cluster Maximum Complete
5
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_
Hamilton_fed_
Hamilton_fed_
Hamilton_fed_
Madison_fed_
Madison_fed_
dispt_fed_
HM_fed_18.txt
HM_fed_
distance6
hclust (*, "complete")

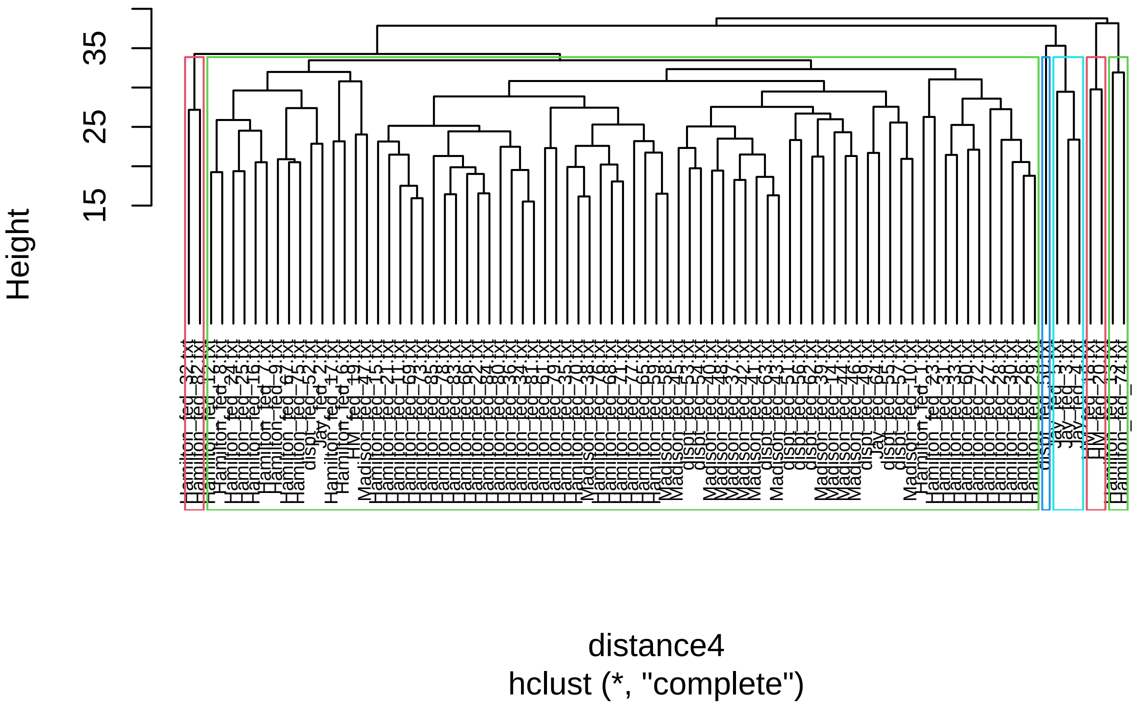

HAC Cluster Euclidean Complete
6
Clusters
Height
0.2
0.6
1.0
1.4
distance
hclust (*, "complete")

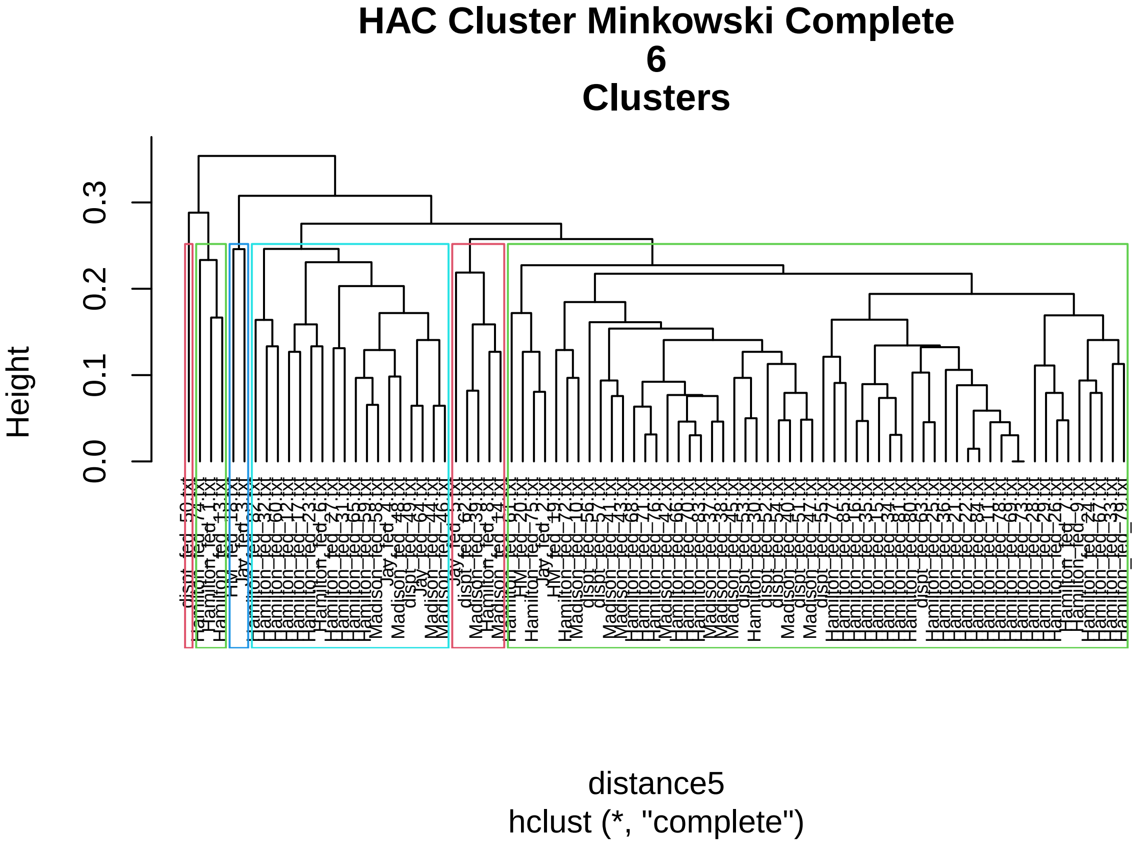

HAC Cluster Euclidean Single
6
Clusters
Height
0.20
0.35
Jay_fed_5.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_27.txt
Hamilton_fed_50.txt
dispt_fed_79.txt
Hamilton_fed_56.txt
dispt_fed_45.txt
Madison_fed_46.txt
Madison_fed_69.txt
Hamilton_fed_74.txt
Hamilton_fed_44.txt
Madison_fed_47.txt
Madison_fed_42.txt
Hamilton_fed_75.txt
Hamilton_fed_11.txt
Hamilton_fed_65.txt
Hamilton_fed_1.txt
Hamilton_fed_33.txt
Hamilton_fed_21.txt
Hamilton_fed_72.txt
Hamilton_fed_7.txt
Hamilton_fed_16.txt
dispt_fed_39.txt
Madison_fed_37.txt
Hamilton_fed_83.txt
Hamilton_fed_68.txt
Hamilton_fed_69.txt
Hamilton_fed_91.txt
Hamilton_fed_28.txt
dispt_fed_57.txt
Hamilton_fed_17.txt
Hamilton_fed_80.txt
Madison_fed_58.txt
Hamilton_fed_84.txt
dispt_fed_51.txt
Hamilton_fed_31.txt
Hamilton_fed_30.txt
Madison_fed_73.txt
Hamilton_fed_76.txt
Hamilton_fed_86.txt
Hamilton_fed_36.txt
Hamilton_fed_78.txt
dispt_fed_49.txt
Hamilton_fed_24.txt
Hamilton_fed_59.txt
Hamilton_fed_60.txt
Hamilton_fed_81.txt
Madison_fed_10.txt
dispt_fed_55.txt
Hamilton_fed_32.txt
Hamilton_fed_9.txt
Hamilton_fed_26.txt
Hamilton_fed_15.txt
Hamilton_fed_8.txt
Hamilton_fed_12.txt
Hamilton_fed_35.txt
Madison_fed_38.txt
Hamilton_fed_70.txt
Hamilton_fed_85.txt
Madison_fed_62.txt
dispt_fed_42.txt
Madison_fed_52.txt
dispt_fed_53.txt
Madison_fed_43.txt
Madison_fed_48.txt
Madison_fed_44.txt
dispt_fed_63.txt
Madison_fed_41.txt
HM_fed_20.txt
HM_fed_18.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance
hclust (*, "single")

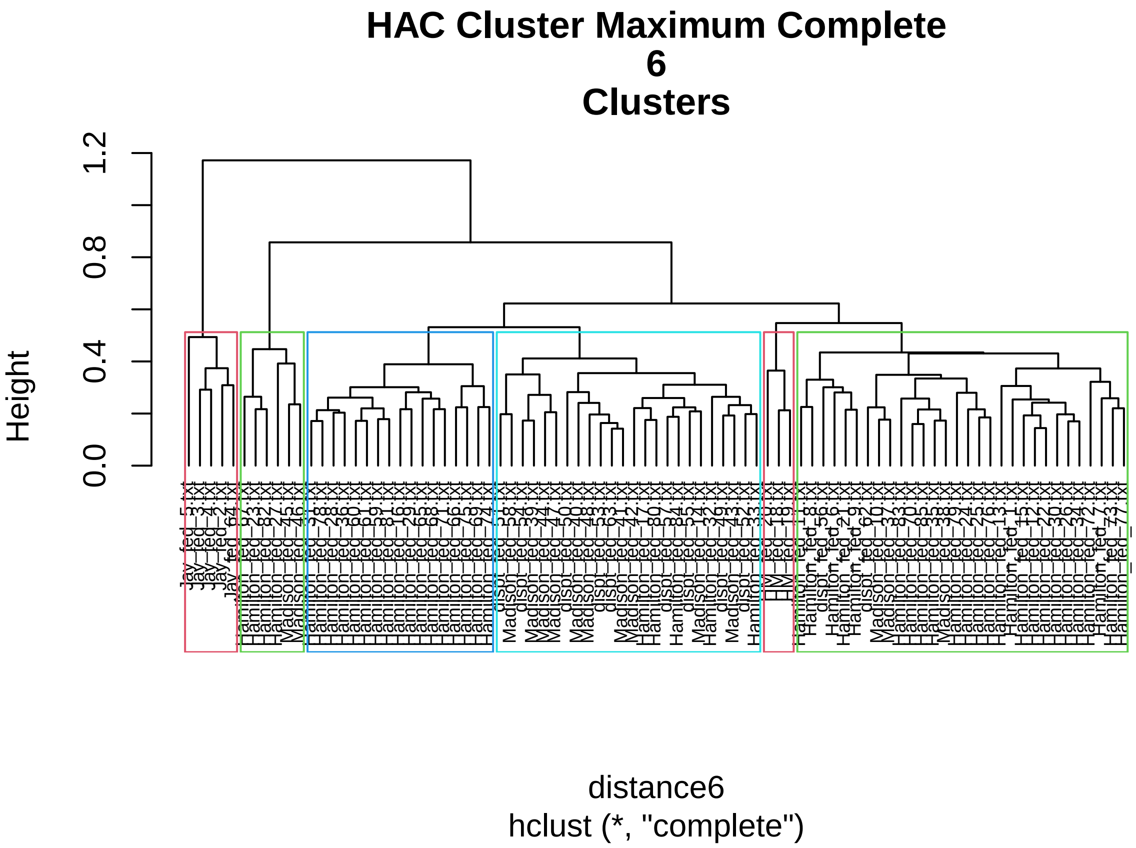

HAC Cluster Maximum Complete
6
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

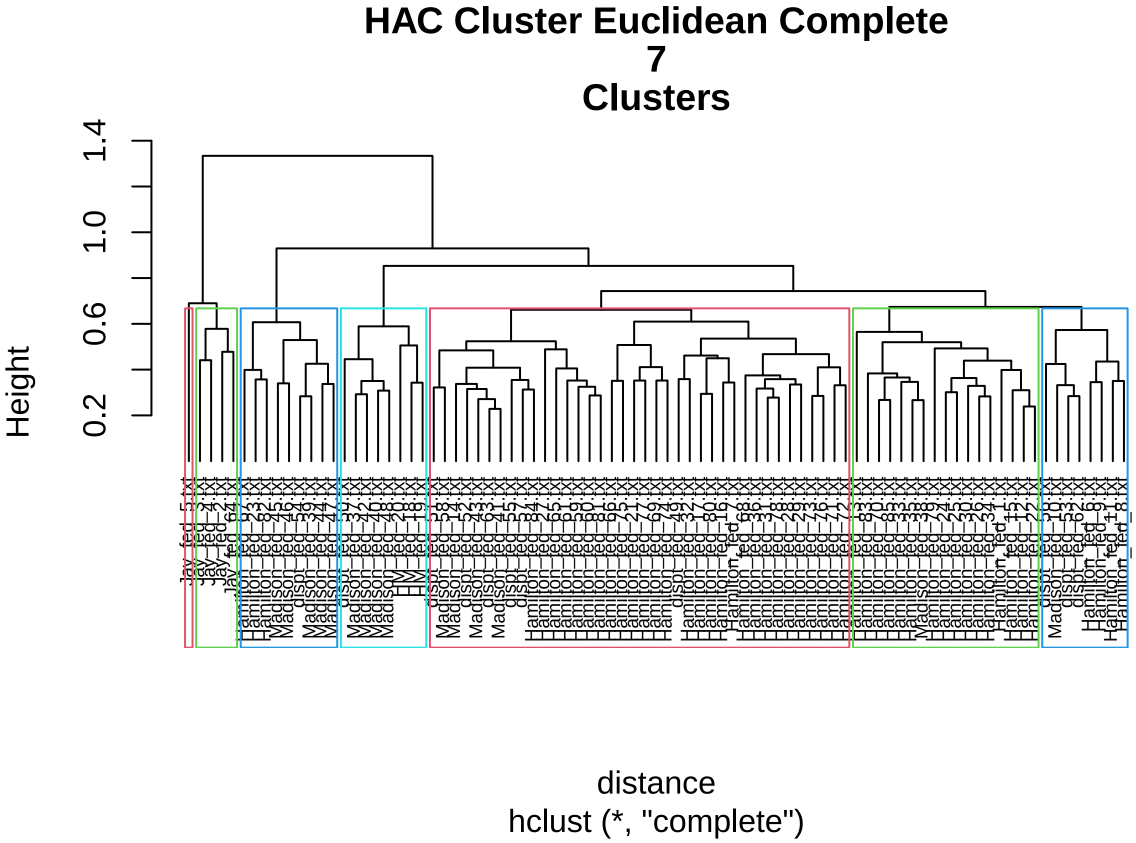

HAC Cluster Manhattan Complete
6
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")

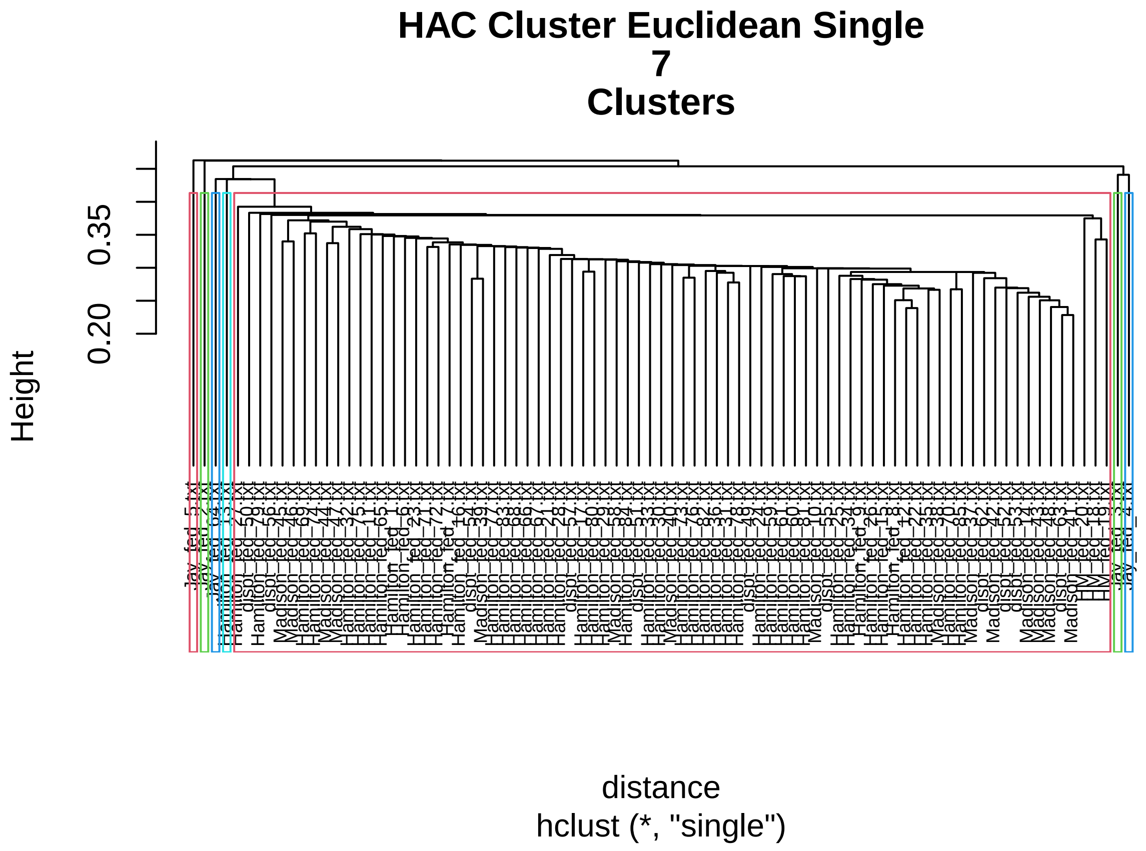
HAC Cluster Canberra Complete
6
Clusters

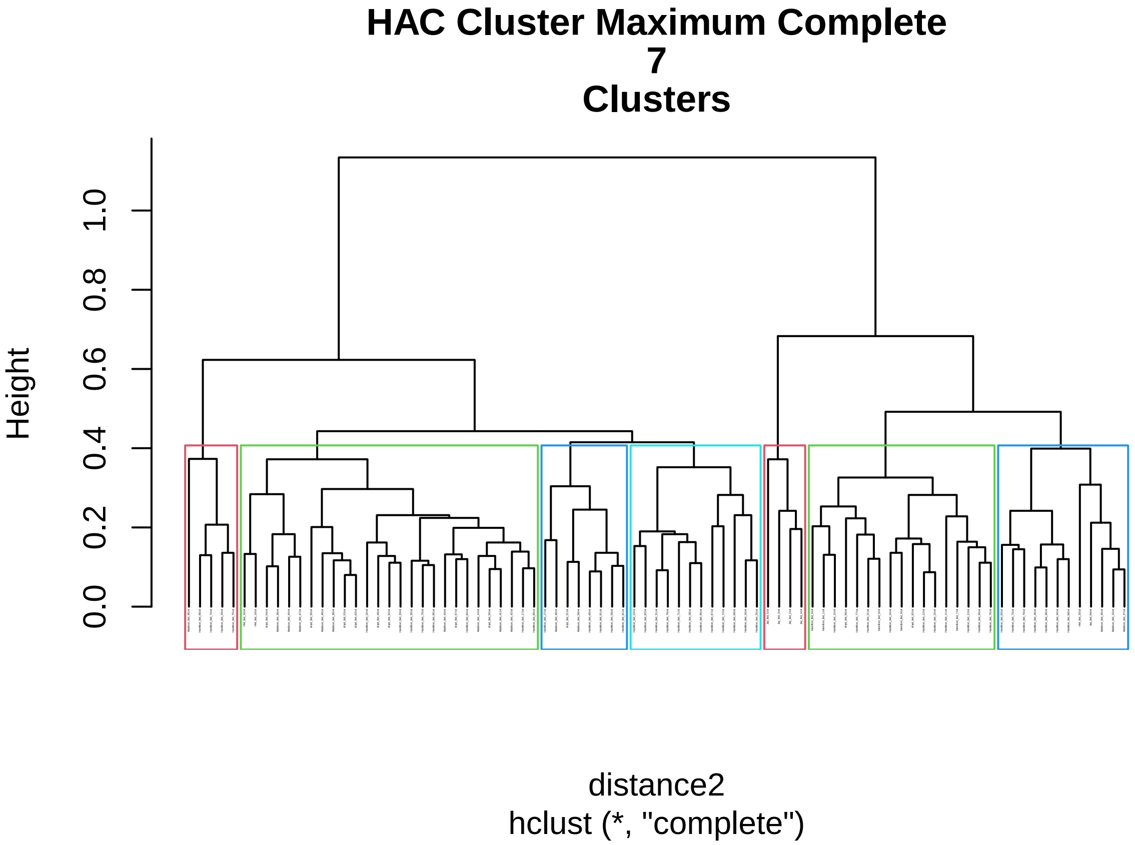



HAC Cluster Minkowski Complete
6
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")

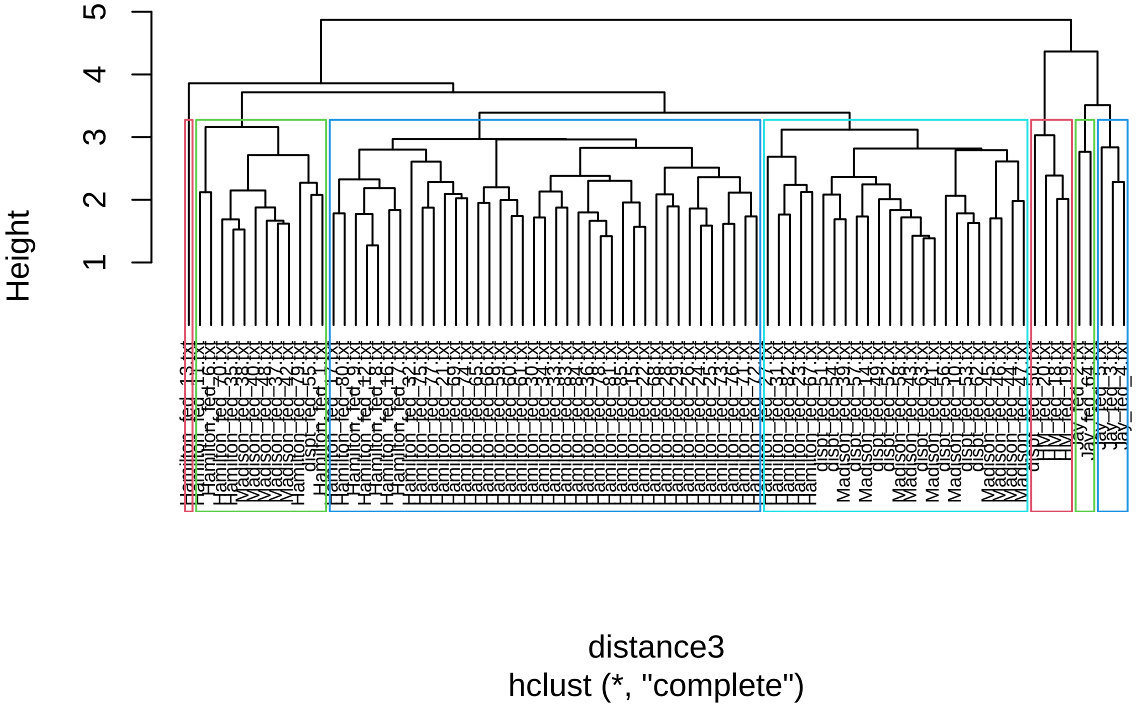

HAC Cluster Maximum Complete
6
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance6
hclust (*, "complete")

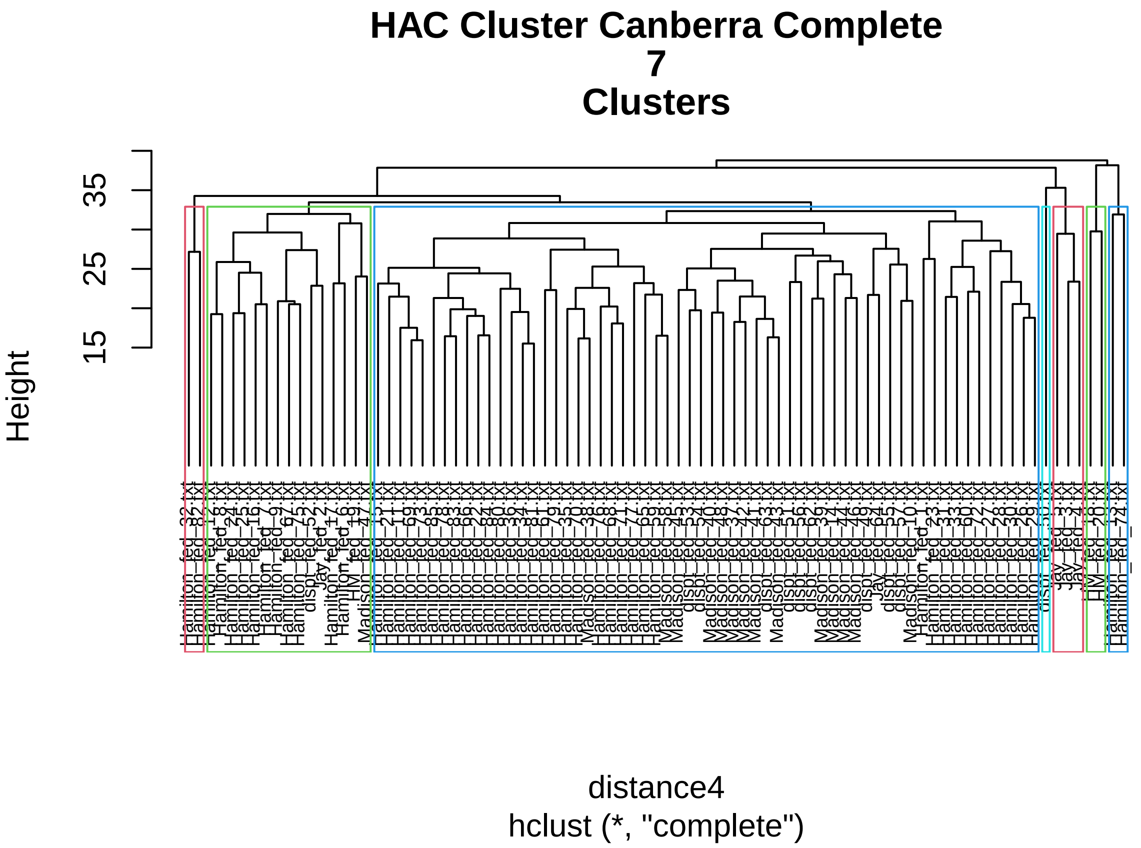

HAC Cluster Euclidean Complete
7
Clusters
Height
0.2
0.6
1.0
1.4
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_23.txt
Hamilton_fed_62.txt
Madison_fed_45.txt
dispt_fed_54.txt
Madison_fed_39.txt
Madison_fed_44.txt
dispt_fed_37.txt
Madison_fed_42.txt
Madison_fed_40.txt
Madison_fed_48.txt
HM_fed_18.txt
HM_fed_19.txt
dispt_fed_58.txt
Madison_fed_14.txt
dispt_fed_63.txt
Madison_fed_41.txt
dispt_fed_55.txt
Hamilton_fed_17.txt
Hamilton_fed_68.txt
Hamilton_fed_1.txt
Hamilton_fed_15.txt
Hamilton_fed_12.txt
Madison_fed_10.txt
dispt_fed_53.txt
dispt_fed_62.txt
Hamilton_fed_6.txt
Hamilton_fed_9.txt
Hamilton_fed_11.txt
Hamilton_fed_8.txt
distance
hclust (*, "complete")

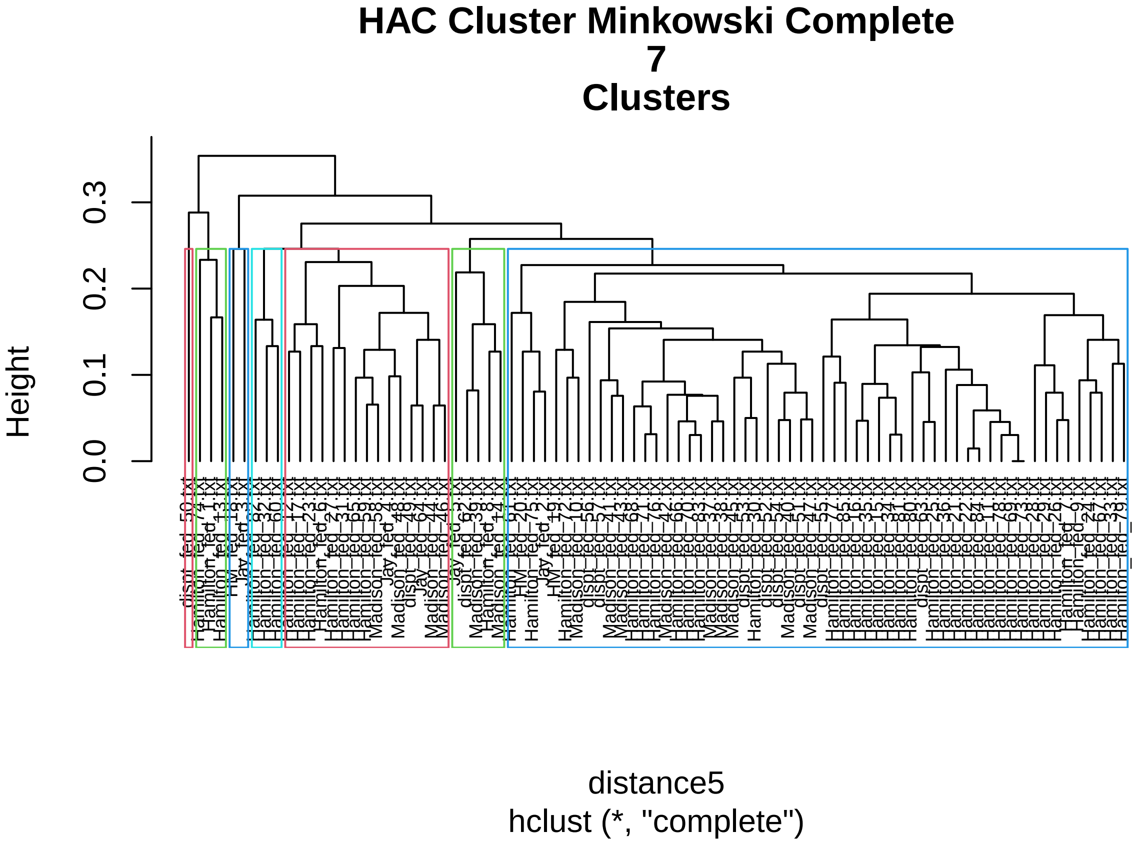

HAC Cluster Euclidean Single
7
Clusters
Height
0.20
0.35
distance
hclust (*, "single")

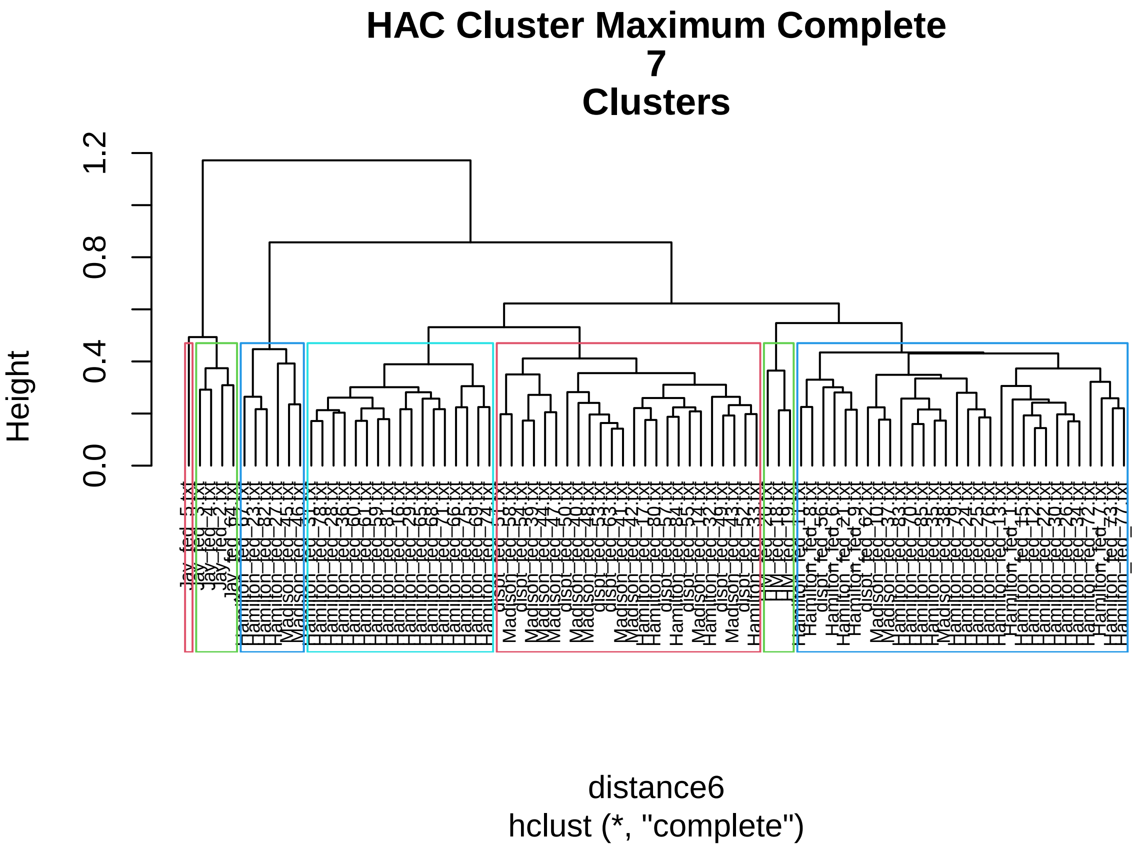

HAC Cluster Maximum Complete
7
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

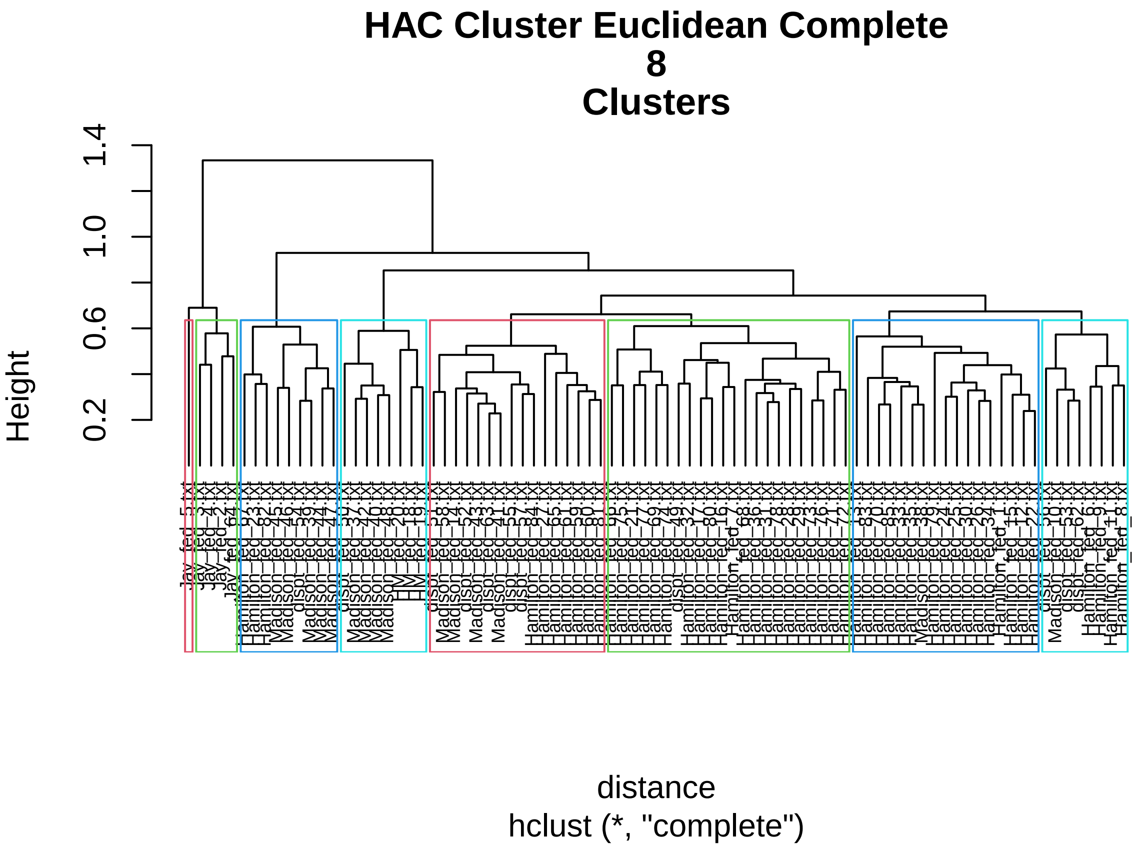
HAC Cluster Manhattan Complete 7 Clusters

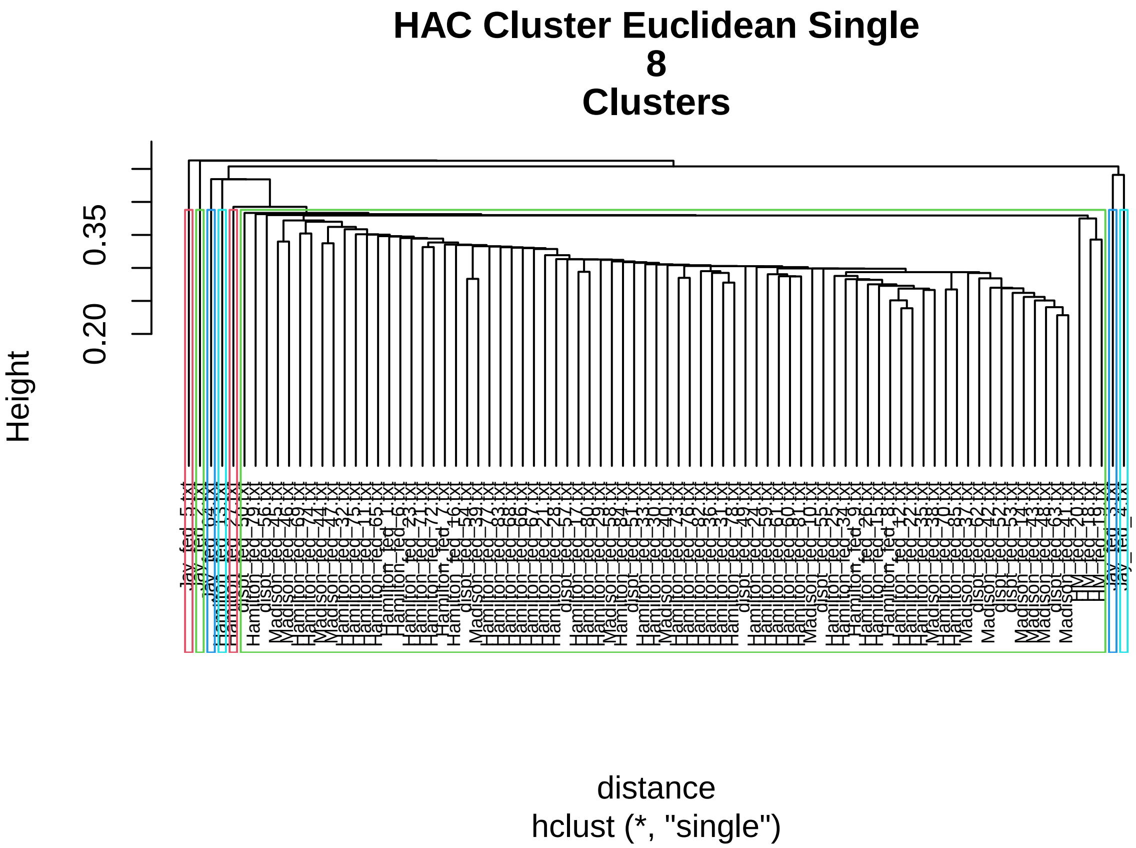


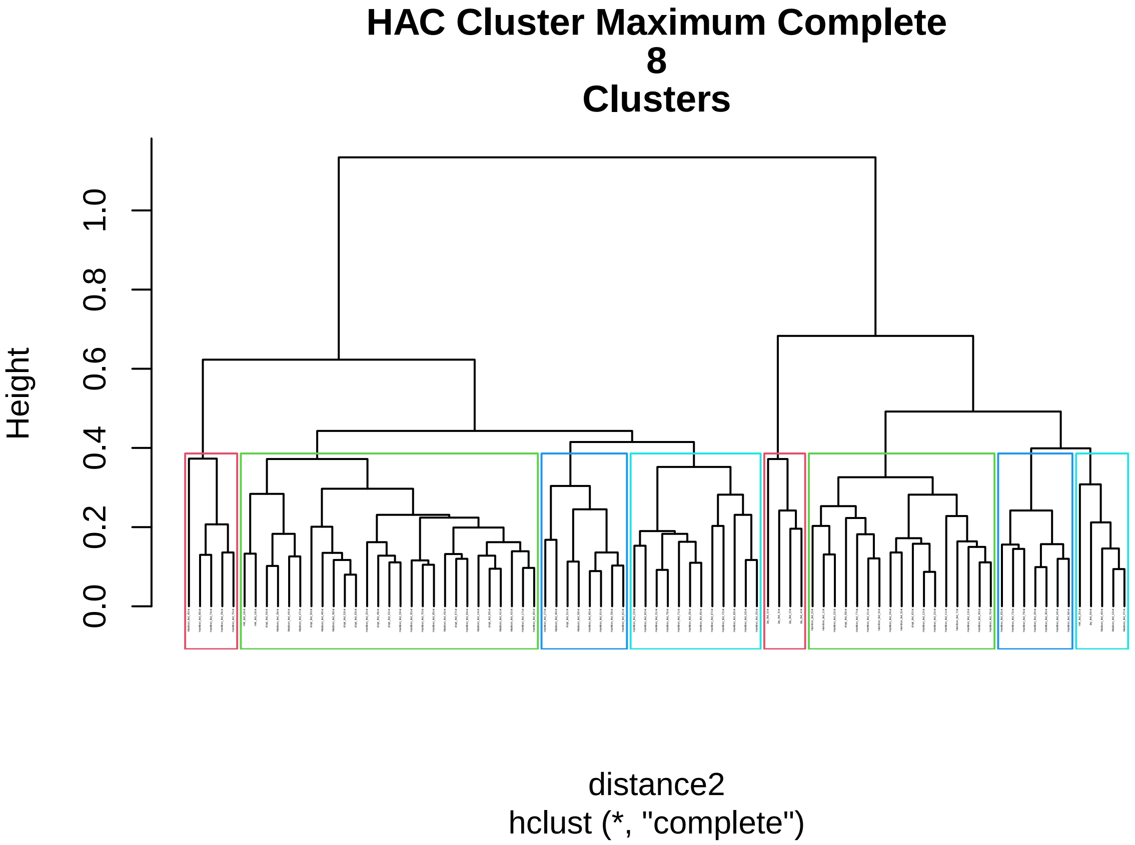
HAC Cluster Canberra Complete
7
Clusters
Height
15
25
35
Hamilton_fed_32.txt
Hamilton_fed_82.txt
Hamilton_fed_1.txt
Hamilton_fed_8.txt
Hamilton_fed_24.txt
Hamilton_fed_25.txt
Hamilton_fed_16.txt
Hamilton_fed_9.txt
Hamilton_fed_67.txt
Hamilton_fed_75.txt
dispt_fed_52.txt
Jay_fed_2.txt
Hamilton_fed_17.txt
Hamilton_fed_16.txt
HM_fed_19.txt
Madison_fed_47.txt
Hamilton_fed_21.txt
Hamilton_fed_11.txt
Hamilton_fed_69.txt
Hamilton_fed_73.txt
Hamilton_fed_85.txt
Hamilton_fed_78.txt
Hamilton_fed_83.txt
Hamilton_fed_59.txt
Hamilton_fed_64.txt
Hamilton_fed_80.txt
Hamilton_fed_36.txt
Hamilton_fed_84.txt
Hamilton_fed_81.txt
Hamilton_fed_61.txt
Hamilton_fed_79.txt
Hamilton_fed_35.txt
Madison_fed_70.txt
Hamilton_fed_38.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_77.txt
Hamilton_fed_65.txt
Hamilton_fed_58.txt
Madison_fed_45.txt
dispt_fed_53.txt
dispt_fed_54.txt
Madison_fed_48.txt
Madison_fed_37.txt
Madison_fed_42.txt
Madison_fed_41.txt
dispt_fed_63.txt
Madison_fed_51.txt
dispt_fed_56.txt
dispt_fed_62.txt
Madison_fed_39.txt
Madison_fed_14.txt
Madison_fed_46.txt
dispt_fed_49.txt
Jay_fed_64.txt
dispt_fed_55.txt
dispt_fed_57.txt
Madison_fed_10.txt
Hamilton_fed_1.txt
Hamilton_fed_23.txt
Hamilton_fed_31.txt
Hamilton_fed_33.txt
Hamilton_fed_30.txt
Hamilton_fed_72.txt
Hamilton_fed_27.txt
Hamilton_fed_28.txt
Hamilton_fed_30.txt
Hamilton_fed_26.txt
Hamilton_fed_29.txt
dispt_fed_50.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_5.txt
HM_fed_18.txt
Hamilton_fed_20.txt
Hamilton_fed_13.txt
Hamilton_fed_74.txt
distance4
hclust (*, "complete")

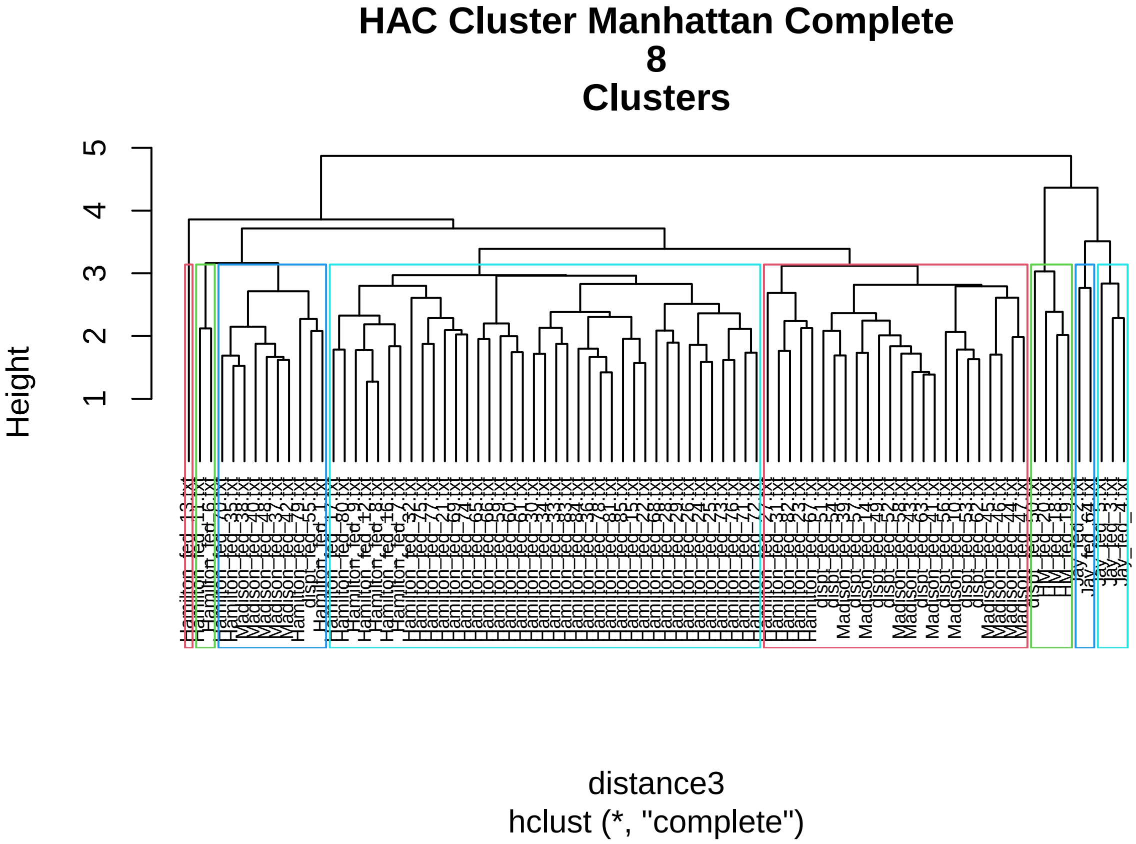

HAC Cluster Minkowski Complete
7
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")

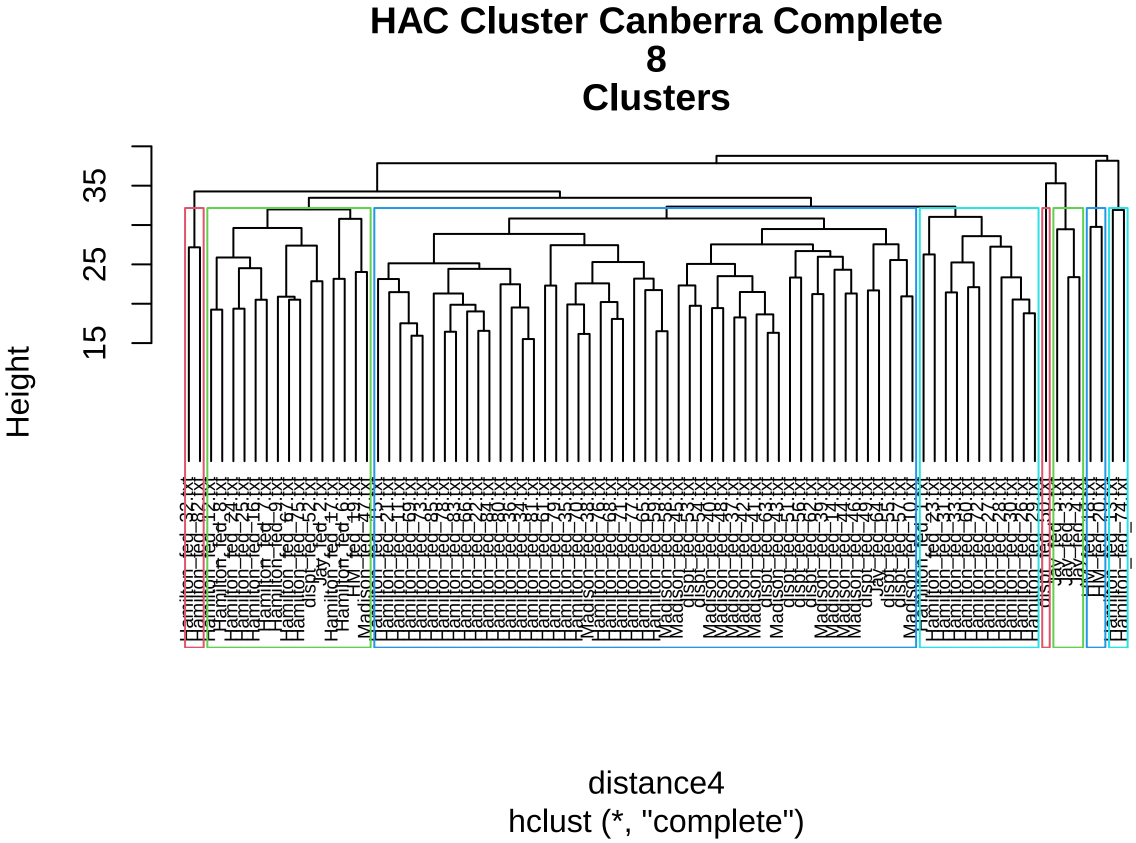

HAC Cluster Maximum Complete
7
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance6
hclust (*, "complete")

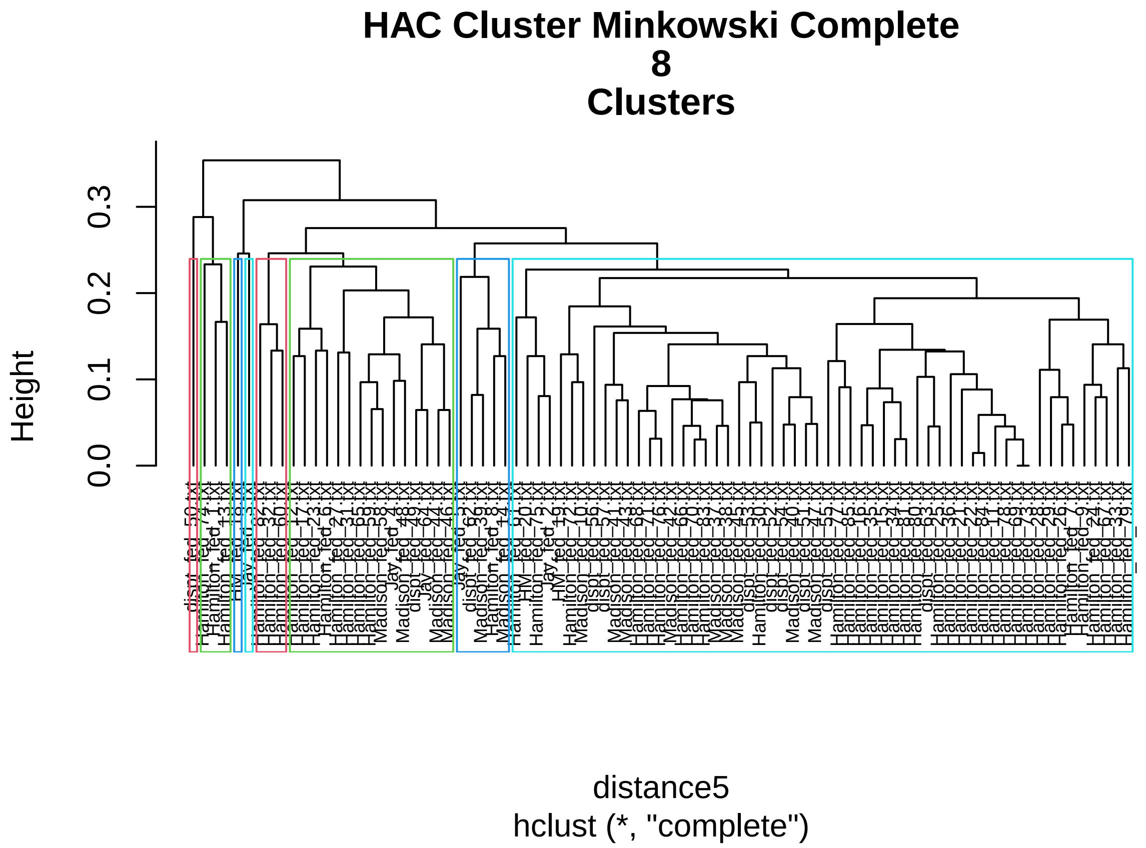

HAC Cluster Euclidean Complete
8
Clusters
Height
1.4
1.0
0.6
0.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
distance
hclust (*, "complete")

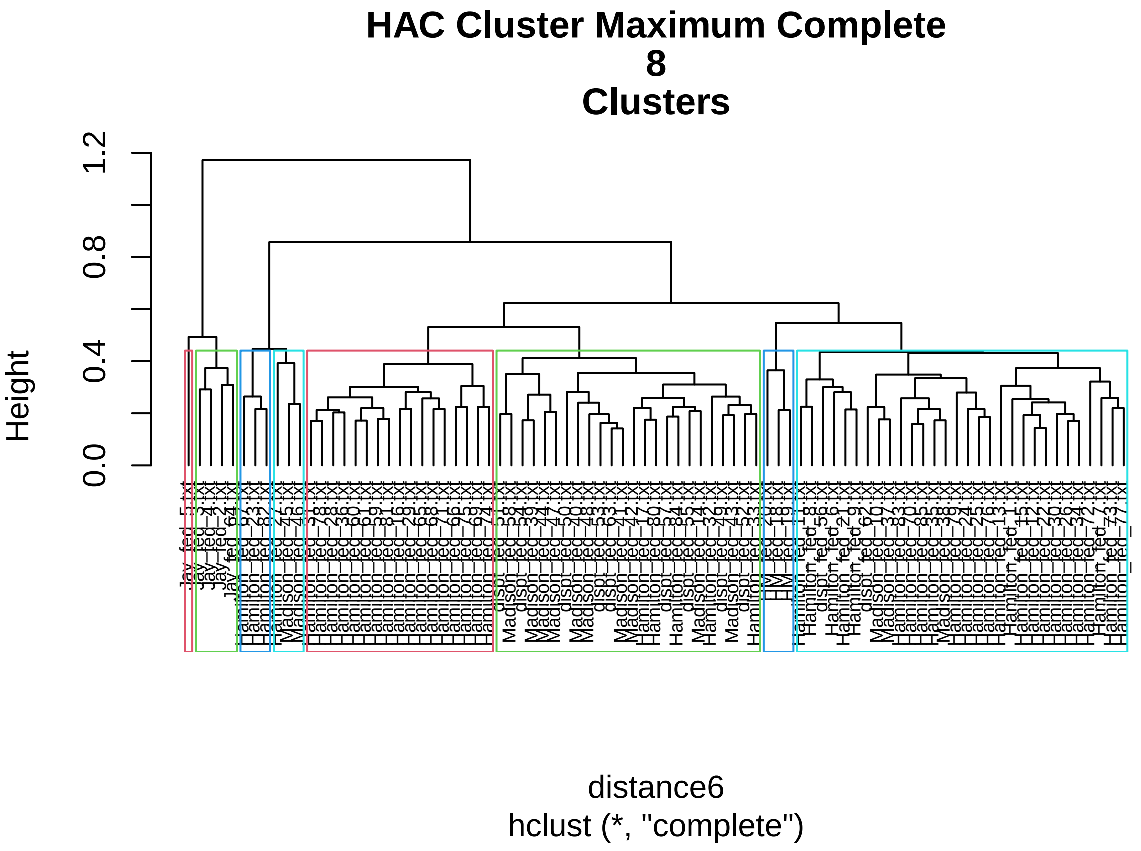

HAC Cluster Euclidean Single
8
Clusters
Height
0.20
0.35
Jay_fed_5.txt
Jay_fed_2.txt
Hamilton_fed_1.txt
Hamilton_fed_23.txt
Hamilton_fed_7.txt
Hamilton_fed_9.txt
Hamilton_fed_26.txt
Hamilton_fed_15.txt
Hamilton_fed_18.txt
HM_fed_20.txt
HM_fed_18.txt
Jay_fed_3.txt
Jay_fed_4.txt
distance
hclust (*, "single")

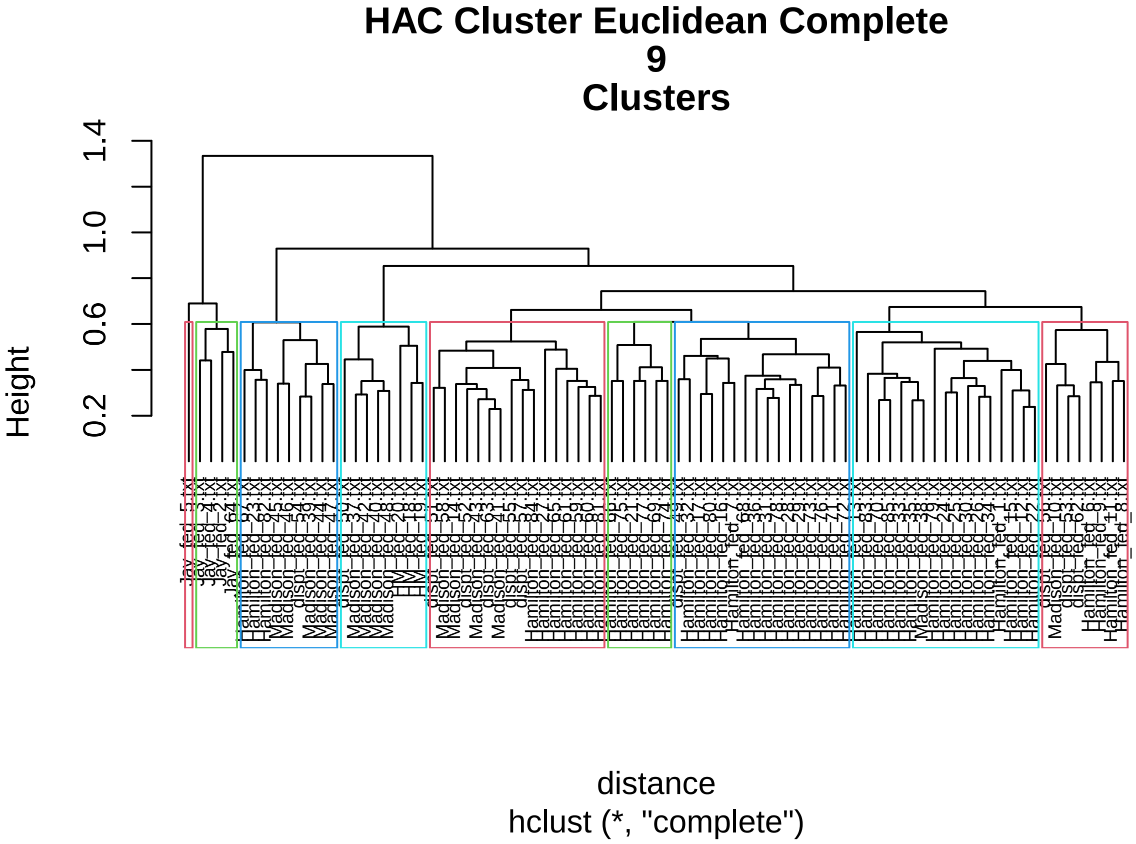

HAC Cluster Maximum Complete
8
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")

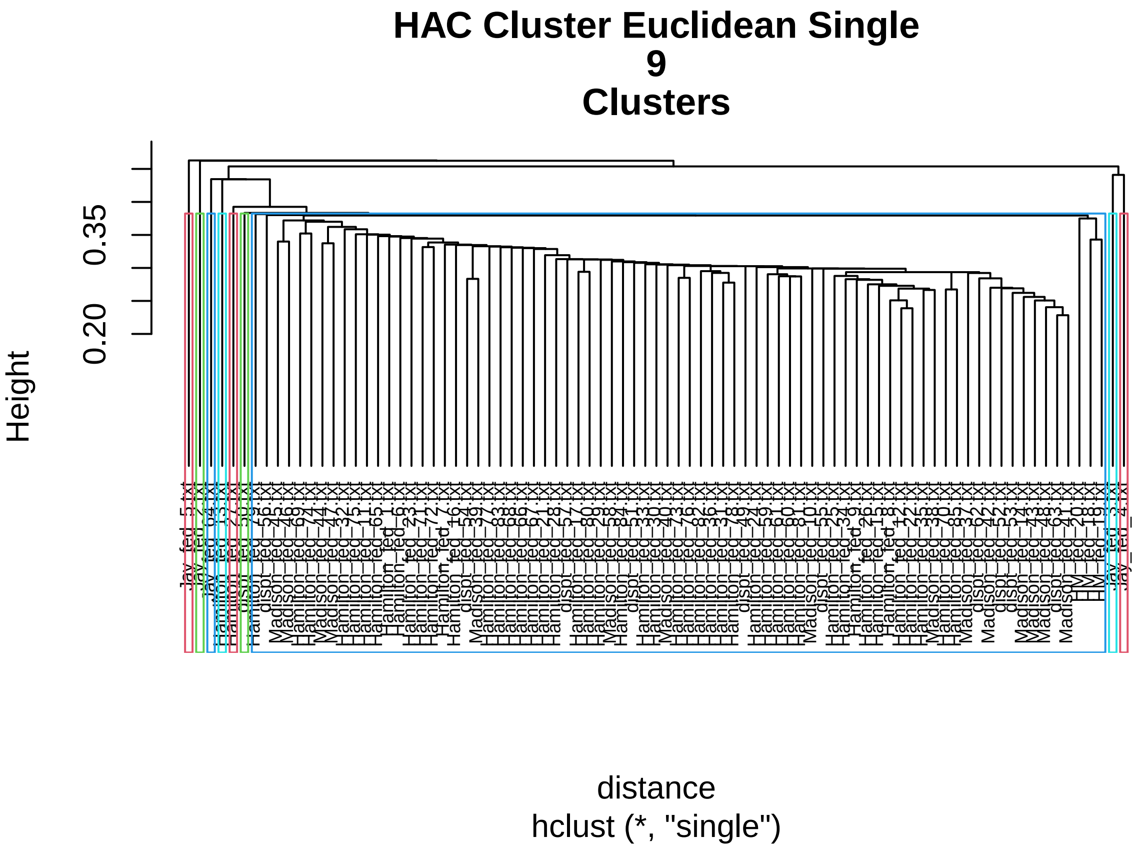

HAC Cluster Manhattan Complete
8
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")

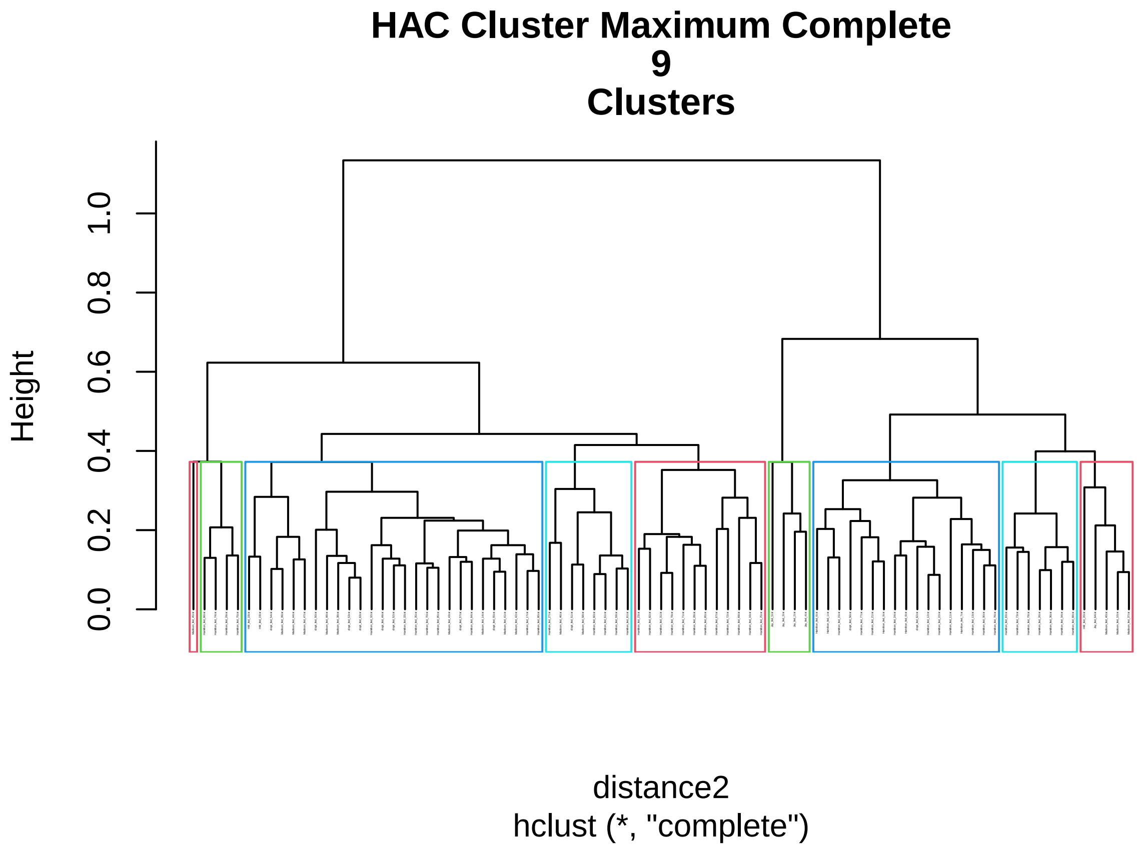

HAC Cluster Canberra Complete
8
Clusters
Height
15
25
35
distance4
hclust (*, "complete")

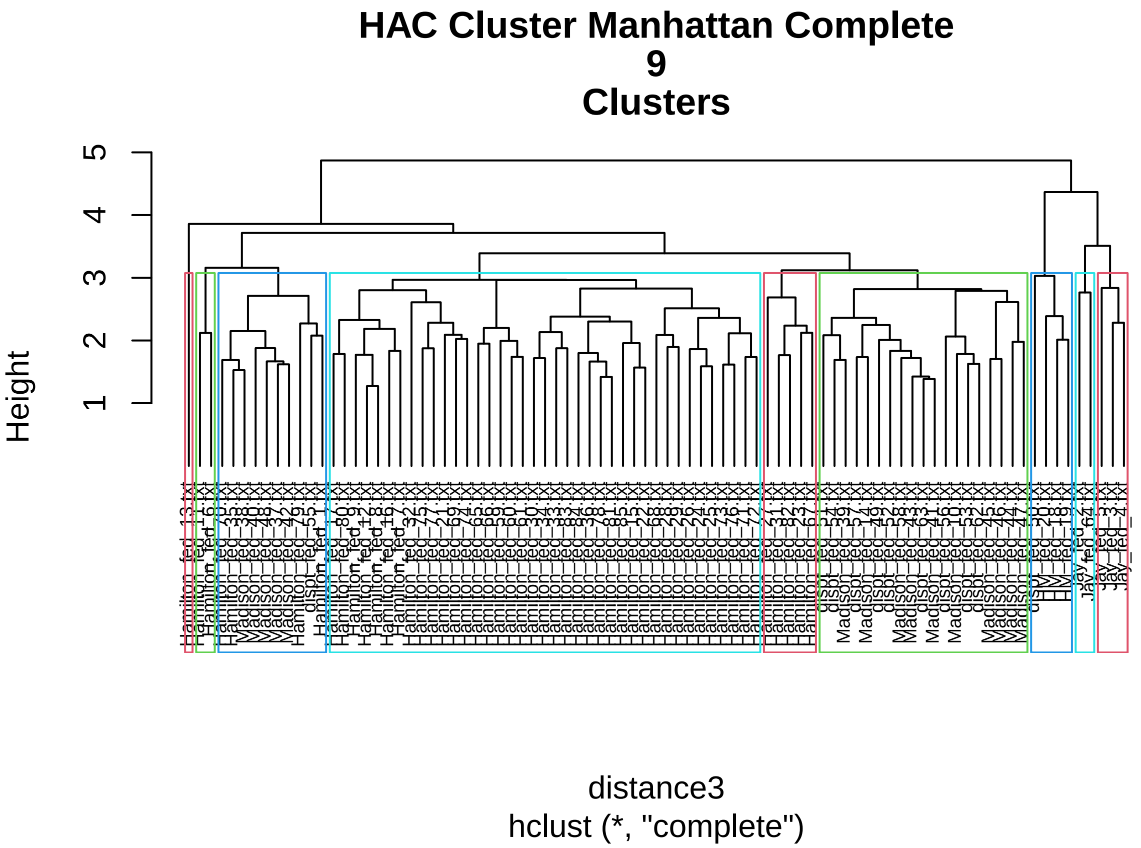

HAC Cluster Minkowski Complete
8
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")

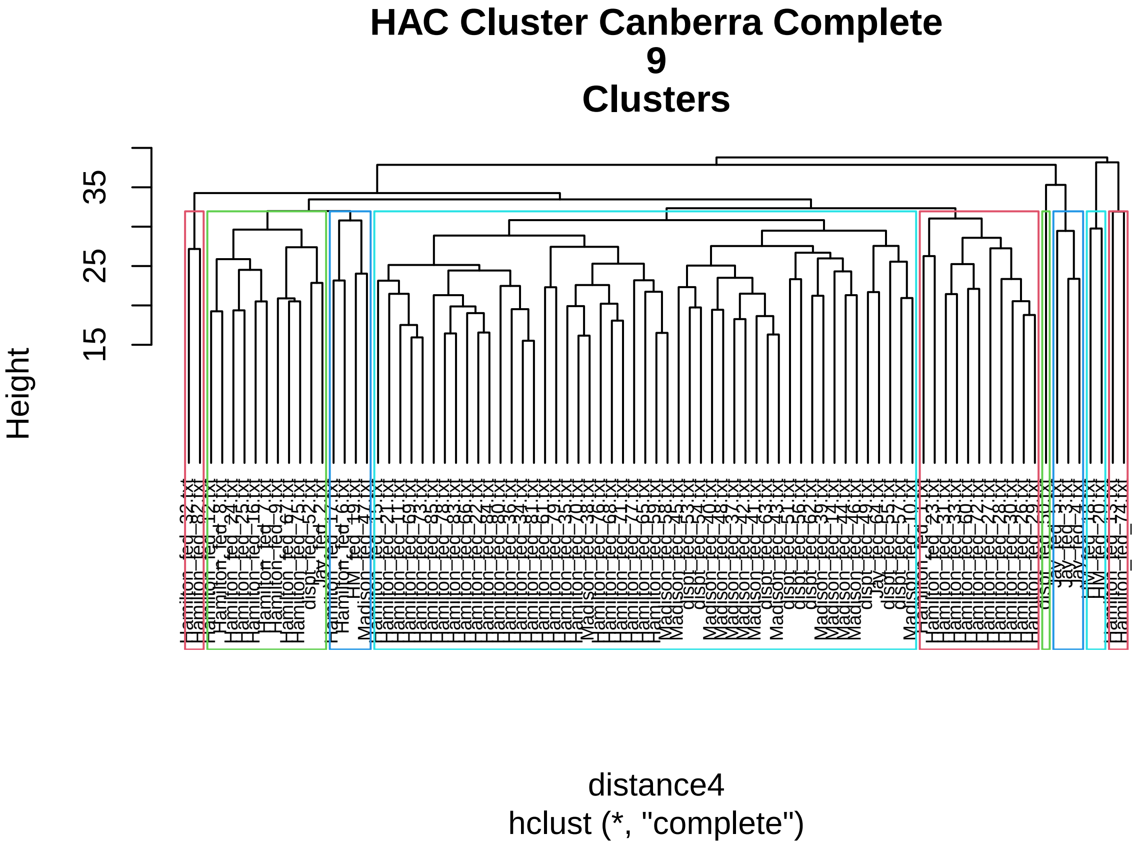

HAC Cluster Maximum Complete
8
Clusters
Height
0.0
0.4
0.8
1.2
Jay_fed_5.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_2.txt
Jay_fed_64.txt
Madison_fed_45.txt
Madison_fed_46.txt
distance6
hclust (*, "complete")

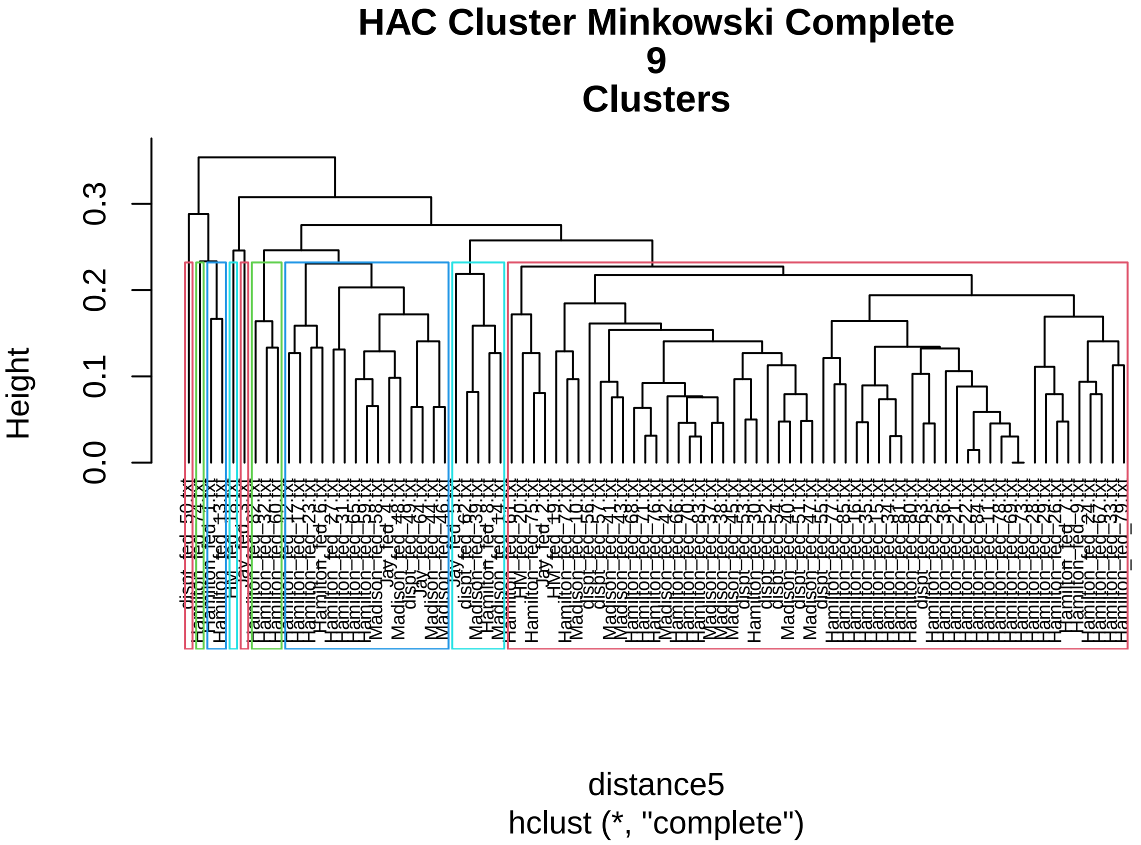

HAC Cluster Euclidean Complete
9
Clusters
Height
0.2
0.6
1.0
1.4
Jay_fed_5.txt
distance
hclust (*, "complete")

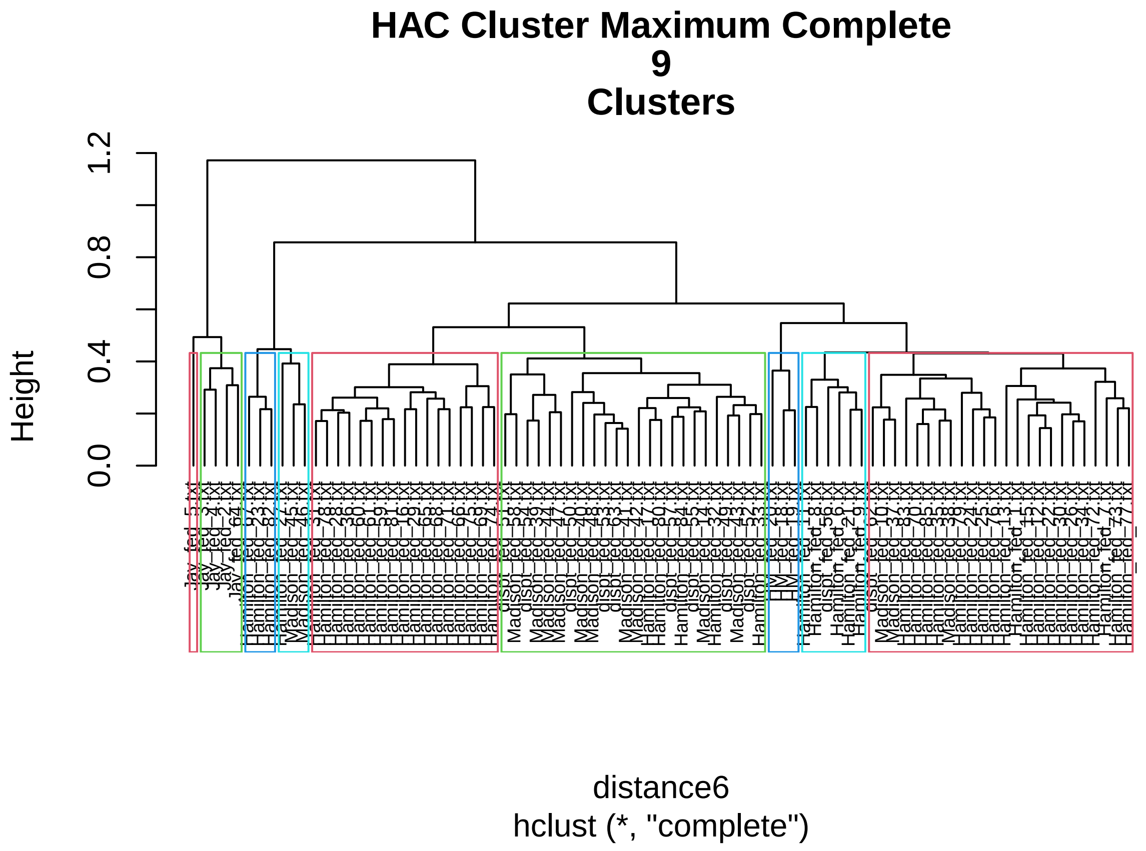

HAC Cluster Euclidean Single
9
Clusters
Height
0.20
0.35
distance
hclust (*, "single")
Jay_fed_5.txt
Jay_fed_2.txt
Jay_fed_64.txt
Hamilton_fed_17.txt
Hamilton_fed_27.txt
disp_fed_50.txt
Hamilton_fed_79.txt
disp_fed_56.txt
Madison_fed_45.txt
Madison_fed_46.txt
Hamilton_fed_69.txt
Hamilton_fed_74.txt
Madison_fed_44.txt
Madison_fed_42.txt
Hamilton_fed_75.txt
Hamilton_fed_11.txt
Hamilton_fed_65.txt
Hamilton_fed_1.txt
Hamilton_fed_23.txt
Hamilton_fed_21.txt
Hamilton_fed_72.txt
Hamilton_fed_7.txt
Hamilton_fed_16.txt
disp_fed_39.txt
Madison_fed_37.txt
Hamilton_fed_83.txt
Hamilton_fed_68.txt
Hamilton_fed_66.txt
Hamilton_fed_91.txt
Hamilton_fed_28.txt
disp_fed_57.txt
Hamilton_fed_17.txt
Hamilton_fed_80.txt
Madison_fed_58.txt
Hamilton_fed_84.txt
disp_fed_51.txt
Hamilton_fed_30.txt
Madison_fed_40.txt
Hamilton_fed_73.txt
Hamilton_fed_76.txt
Hamilton_fed_86.txt
Hamilton_fed_36.txt
Hamilton_fed_78.txt
disp_fed_49.txt
Hamilton_fed_24.txt
Hamilton_fed_59.txt
Hamilton_fed_60.txt
Hamilton_fed_81.txt
Madison_fed_10.txt
disp_fed_55.txt
Hamilton_fed_32.txt
Hamilton_fed_9.txt
Hamilton_fed_26.txt
Hamilton_fed_15.txt
Hamilton_fed_8.txt
Hamilton_fed_22.txt
Hamilton_fed_35.txt
Madison_fed_38.txt
Hamilton_fed_70.txt
Hamilton_fed_85.txt
Madison_fed_62.txt
disp_fed_42.txt
Madison_fed_52.txt
disp_fed_53.txt
Madison_fed_43.txt
Madison_fed_48.txt
disp_fed_63.txt
Madison_fed_41.txt
HM_fed_20.txt
HM_fed_18.txt
Jay_fed_3.txt
Jay_fed_4.txt


HAC Cluster Maximum Complete
9
Clusters
Height
0.0
0.2
0.4
0.6
0.8
1.0
distance2
hclust (*, "complete")


HAC Cluster Manhattan Complete
9
Clusters
Height
1
2
3
4
5
distance3
hclust (*, "complete")


HAC Cluster Canberra Complete
9
Clusters
Height
15
25
35
Hamilton_fed_32.txt
Hamilton_fed_82.txt
Hamilton_fed_6.txt
Hamilton_fed_24.txt
Hamilton_fed_25.txt
Hamilton_fed_16.txt
Hamilton_fed_9.txt
Hamilton_fed_67.txt
Hamilton_fed_75.txt
dispt_fed_52.txt
Jay_fed_2.txt
Hamilton_fed_17.txt
Hamilton_fed_18.txt
HM_fed_47.txt
Madison_fed_15.txt
Hamilton_fed_21.txt
Hamilton_fed_69.txt
Hamilton_fed_73.txt
Hamilton_fed_85.txt
Hamilton_fed_78.txt
Hamilton_fed_83.txt
Hamilton_fed_62.txt
Hamilton_fed_84.txt
Hamilton_fed_80.txt
Hamilton_fed_36.txt
Hamilton_fed_34.txt
Hamilton_fed_81.txt
Hamilton_fed_79.txt
Hamilton_fed_35.txt
Madison_fed_70.txt
Hamilton_fed_38.txt
Hamilton_fed_68.txt
Hamilton_fed_71.txt
Hamilton_fed_77.txt
Hamilton_fed_65.txt
Madison_fed_58.txt
Madison_fed_45.txt
dispt_fed_53.txt
dispt_fed_54.txt
Madison_fed_48.txt
Madison_fed_37.txt
Madison_fed_42.txt
Madison_fed_41.txt
dispt_fed_63.txt
Madison_fed_51.txt
dispt_fed_56.txt
dispt_fed_62.txt
dispt_fed_39.txt
Madison_fed_14.txt
Madison_fed_44.txt
Madison_fed_46.txt
dispt_fed_49.txt
Jay_fed_64.txt
dispt_fed_55.txt
dispt_fed_50.txt
Madison_fed_10.txt
Hamilton_fed_23.txt
Hamilton_fed_31.txt
Hamilton_fed_33.txt
Hamilton_fed_59.txt
Hamilton_fed_27.txt
Hamilton_fed_72.txt
Hamilton_fed_28.txt
Hamilton_fed_30.txt
Hamilton_fed_26.txt
Hamilton_fed_29.txt
dispt_fed_50.txt
Jay_fed_3.txt
Jay_fed_4.txt
Jay_fed_5.txt
HM_fed_20.txt
Hamilton_fed_13.txt
Hamilton_fed_74.txt
distance4
hclust (*, "complete")


HAC Cluster Minkowski Complete
9
Clusters
Height
0.0
0.1
0.2
0.3
distance5
hclust (*, "complete")



Other analysis

Load Data (as Corpus).

In this example, we will load the data in corpus form. We will need to do much of the data cleaning, text processing, ourselves.

###Load Fed Papers Corpus

```
FedPapersCorpus <- Corpus(DirSource("C:/Users/GeorgeSmith/Desktop/FedPapersCorpus"))
(numberFedPapers<-length(FedPapersCorpus))
```

```
## [1] 85
```

##The following will show you that you read in all the documents

```
(summary(FedPapersCorpus))
```

```
##                    Length Class             Mode
## dispt_fed_49.txt   2      PlainTextDocument list
## dispt_fed_50.txt   2      PlainTextDocument list
## dispt_fed_51.txt    2      PlainTextDocument list
## dispt_fed_52.txt    2      PlainTextDocument list
## dispt_fed_53.txt    2      PlainTextDocument list
## dispt_fed_54.txt    2      PlainTextDocument list
## dispt_fed_55.txt    2      PlainTextDocument list
## dispt_fed_56.txt    2      PlainTextDocument list
## dispt_fed_57.txt    2      PlainTextDocument list
## dispt_fed_62.txt    2      PlainTextDocument list
## dispt_fed_63.txt    2      PlainTextDocument list
## Hamilton_fed_1.txt  2      PlainTextDocument list
## Hamilton_fed_11.txt 2      PlainTextDocument list
## Hamilton_fed_12.txt 2      PlainTextDocument list
## Hamilton_fed_13.txt 2      PlainTextDocument list
## Hamilton_fed_15.txt 2      PlainTextDocument list
## Hamilton_fed_16.txt 2      PlainTextDocument list
## Hamilton_fed_17.txt 2      PlainTextDocument list
## Hamilton_fed_21.txt 2      PlainTextDocument list
## Hamilton_fed_22.txt 2      PlainTextDocument list
## Hamilton_fed_23.txt 2      PlainTextDocument list
## Hamilton_fed_24.txt 2      PlainTextDocument list
## Hamilton_fed_25.txt 2      PlainTextDocument list
## Hamilton_fed_26.txt 2      PlainTextDocument list
## Hamilton_fed_27.txt 2      PlainTextDocument list
## Hamilton_fed_28.txt 2      PlainTextDocument list
## Hamilton_fed_29.txt 2      PlainTextDocument list
## Hamilton_fed_30.txt 2      PlainTextDocument list
## Hamilton_fed_31.txt 2      PlainTextDocument list
## Hamilton_fed_32.txt 2      PlainTextDocument list
## Hamilton_fed_33.txt 2      PlainTextDocument list
## Hamilton_fed_34.txt 2      PlainTextDocument list
## Hamilton_fed_35.txt 2      PlainTextDocument list
## Hamilton_fed_36.txt 2      PlainTextDocument list
## Hamilton_fed_59.txt 2      PlainTextDocument list
## Hamilton_fed_6.txt  2      PlainTextDocument list
## Hamilton_fed_60.txt 2      PlainTextDocument list
## Hamilton_fed_61.txt 2      PlainTextDocument list
## Hamilton_fed_65.txt 2      PlainTextDocument list
## Hamilton_fed_66.txt 2      PlainTextDocument list
## Hamilton_fed_67.txt 2      PlainTextDocument list
## Hamilton_fed_68.txt 2      PlainTextDocument list
## Hamilton_fed_69.txt 2      PlainTextDocument list
## Hamilton_fed_7.txt  2      PlainTextDocument list
## Hamilton_fed_70.txt 2      PlainTextDocument list
## Hamilton_fed_71.txt 2      PlainTextDocument list
## Hamilton_fed_72.txt 2      PlainTextDocument list
## Hamilton_fed_73.txt 2      PlainTextDocument list
## Hamilton_fed_74.txt 2      PlainTextDocument list
## Hamilton_fed_75.txt 2      PlainTextDocument list
## Hamilton_fed_76.txt 2      PlainTextDocument list
## Hamilton_fed_77.txt 2      PlainTextDocument list
## Hamilton_fed_78.txt 2      PlainTextDocument list
## Hamilton_fed_79.txt 2      PlainTextDocument list
## Hamilton_fed_8.txt  2      PlainTextDocument list
## Hamilton_fed_80.txt 2      PlainTextDocument list
## Hamilton_fed_81.txt 2      PlainTextDocument list
## Hamilton_fed_82.txt 2      PlainTextDocument list
## Hamilton_fed_83.txt 2      PlainTextDocument list
## Hamilton_fed_84.txt 2      PlainTextDocument list
## Hamilton_fed_85.txt 2      PlainTextDocument list
## Hamilton_fed_9.txt  2      PlainTextDocument list
## HM_fed_18.txt       2      PlainTextDocument list
## HM_fed_19.txt       2      PlainTextDocument list
## HM_fed_20.txt       2      PlainTextDocument list
## Jay_fed_2.txt       2      PlainTextDocument list
## Jay_fed_3.txt       2      PlainTextDocument list
## Jay_fed_4.txt       2      PlainTextDocument list
## Jay_fed_5.txt       2      PlainTextDocument list
## Jay_fed_64.txt      2      PlainTextDocument list
## Madison_fed_10.txt  2      PlainTextDocument list
## Madison_fed_14.txt  2      PlainTextDocument list
## Madison_fed_37.txt  2      PlainTextDocument list
## Madison_fed_38.txt  2      PlainTextDocument list
## Madison_fed_39.txt  2      PlainTextDocument list
## Madison_fed_40.txt  2      PlainTextDocument list
## Madison_fed_41.txt  2      PlainTextDocument list
## Madison_fed_42.txt  2      PlainTextDocument list
## Madison_fed_43.txt  2      PlainTextDocument list
## Madison_fed_44.txt  2      PlainTextDocument list
## Madison_fed_45.txt  2      PlainTextDocument list
## Madison_fed_46.txt  2      PlainTextDocument list
## Madison_fed_47.txt  2      PlainTextDocument list
## Madison_fed_48.txt  2      PlainTextDocument list
## Madison_fed_58.txt  2      PlainTextDocument list
```

Data Cleaning

Here we investigate the data and vectorize it using DocumentTermMatrix.

We will ignore very infrequent words and very frequent words during the vectorization process.

Note: The DocumentTermMatrix method will perform much data cleaning for us.

Data Preparation and Transformation on Fed Papers

Remove punctuation,numbers, and space

```
(getTransformations())
```

```
## [1] "removeNumbers"     "removePunctuation" "removeWords"
## [4] "stemDocument"      "stripWhitespace"
```

```
(nFedPapersCorpus<-length(FedPapersCorpus))
```

```
## [1] 85
```

ignore extremely rare words i.e. terms that appear in less then 1% of the documents

```
(minTermFreq <- nFedPapersCorpus * 0.0001)
```

```
## [1] 0.0085
```

```
(minTermFreqNum <- 30)  #  min terms as a number
```

```
## [1] 30
```

###Ignore overly common words i.e. terms that appear in more than 50% of the documents

```
(maxTermFreq <- nFedPapersCorpus * 1)
```

```
## [1] 85
```

```
(maxTermFreqNum <- 1000)  # max terms as a number
```

```
## [1] 1000
```

```
MyStopwords <- c("will","one","two", "may","less", "well","might","withou","small", "single", "several"
                 "but", "very", "can", "must", "also", "very", "can", "any", "and", "are", "however",
                 "into", "almost", "can","for","add")

(STOPS <-stopwords('english'))
```

```
##   [1] "i"          "me"         "my"         "myself"     "we"
##   [6] "our"        "ours"       "ourselves"  "you"        "your"
##  [11] "yours"      "yourself"   "yourselves" "he"         "him"
##  [16] "his"        "himself"    "she"        "her"        "hers"
##  [21] "herself"    "it"         "its"        "itself"     "they"
##  [26] "them"       "their"      "theirs"     "themselves" "what"
##  [31] "which"      "who"        "whom"       "this"       "that"
##  [36] "these"      "those"      "am"         "is"         "are"
##  [41] "was"        "were"       "be"         "been"       "being"
##  [46] "have"       "has"        "had"        "having"     "do"
##  [51] "does"       "did"        "doing"      "would"      "should"
##  [56] "could"      "ought"      "i'm"        "you're"     "he's"      
##  [61] "she's"      "it's"       "we're"      "they're"    "i've"      
##  [66] "you've"     "we've"      "they've"    "i'd"        "you'd"     
##  [71] "he'd"       "she'd"      "we'd"       "they'd"     "i'll"      
##  [76] "you'll"     "he'll"      "she'll"     "we'll"      "they'll"   
##  [81] "isn't"      "aren't"     "wasn't"     "weren't"    "hasn't"    
##  [86] "haven't"    "hadn't"     "doesn't"    "don't"      "didn't"    
##  [91] "won't"      "wouldn't"   "shan't"     "shouldn't"  "can't"     
##  [96] "cannot"     "couldn't"   "mustn't"    "let's"      "that's"    
## [101] "who's"      "what's"     "here's"     "there's"    "when's"    
## [106] "where's"    "why's"      "how's"      "a"          "an"        
## [111] "the"        "and"        "but"        "if"         "or"        
## [116] "because"    "as"         "until"      "while"      "of"        
## [121] "at"         "by"         "for"        "with"       "about"     
## [126] "against"    "between"    "into"       "through"    "during"    
## [131] "before"     "after"      "above"      "below"      "to"        
## [136] "from"       "up"         "down"       "in"         "out"       
## [141] "on"         "off"        "over"       "under"      "again"     
## [146] "further"    "then"       "once"       "here"       "there"     
## [151] "when"       "where"      "why"        "how"        "all"       
## [156] "any"        "both"       "each"       "few"        "more"      
## [161] "most"       "other"      "some"       "such"       "no"        
## [166] "nor"        "not"        "only"       "own"        "same"      
## [171] "so"         "than"       "too"        "very"       "will"
```

```
Papers_DTM <- DocumentTermMatrix(FedPapersCorpus, control = list( stopwords = TRUE, wordLengths=c(3, 15
                                                                  removePunctuation = T, removeNumbers =
                                                                  stopwords = MyStopwords, bounds = list
```

use the "built-in" STOP words

#inspect FedPapers Document Term Matrix (DTM)

```
DTM <- as.matrix(Papers_DTM)
(DTM[1:11,1:10])
```

```
##                    Terms
## Docs                 abandon abat abb abet abhorr abil abject abl ablest abolish
##   dispt_fed_49.txt       0    0   0    0      0    0      0   2      0       0
##   dispt_fed_50.txt       0    0   0    0      0    0      0   0      0       0
##   dispt_fed_51.txt       0    0   0    0      0    0      0   1      0       0
##   dispt_fed_52.txt       0    0   0    0      0    1      0   1      0       0
##   dispt_fed_53.txt       0    1   0    0      0    0      0   0      0       0
##   dispt_fed_54.txt       0    0   0    0      0    0      0   0      0       0
##   dispt_fed_55.txt       0    0   0    0      0    0      0   0      0       0
##   dispt_fed_56.txt       0    0   0    0      0    0      0   0      0       0
##   dispt_fed_57.txt       0    0   0    0      1    0      0   0      0       0
##   dispt_fed_62.txt       0    0   0    0      0    0      0   1      0       0
##   dispt_fed_63.txt       0    0   0    0      0    0      0   4      0       0
```

Inspect Initial Cleaning Results

Look at word freuquncies

```
WordFreq <- colSums(as.matrix(Papers_DTM))
(head(WordFreq))
```

```
## abandon    abat     abb    abet  abhorr    abil
##       9       2       5       2       1      15
```

```
(length(WordFreq))
```

```
## [1] 4900
```

```
ord <- order(WordFreq)
(WordFreq[head(ord)])
```

```
##  abhorr  abject abraham   abreg  absenc  absolv
##       1       1       1       1       1       1
```

```
(WordFreq[tail(ord)])
```

```
## constitut       may     power    govern      will     state
##       686       811       937      1040      1263      1662
```

```
(Row_Sum_Per_doc <- rowSums((as.matrix(Papers_DTM))))
```

```
##    dispt_fed_49.txt    dispt_fed_50.txt    dispt_fed_51.txt    dispt_fed_52.txt
##                 758                 530                 923                 853
##    dispt_fed_53.txt    dispt_fed_54.txt    dispt_fed_55.txt    dispt_fed_56.txt
##                1035                 882                 968                 765
##    dispt_fed_57.txt    dispt_fed_62.txt    dispt_fed_63.txt  Hamilton_fed_1.txt
##                1023                1124                1432                 767
## Hamilton_fed_11.txt Hamilton_fed_12.txt Hamilton_fed_13.txt Hamilton_fed_15.txt
##                1164                1044                 479                1411
## Hamilton_fed_16.txt Hamilton_fed_17.txt Hamilton_fed_21.txt Hamilton_fed_22.txt
##                 918                 767                 937                1692
## Hamilton_fed_23.txt Hamilton_fed_24.txt Hamilton_fed_25.txt Hamilton_fed_26.txt
##                 828                 925                 927                1093
## Hamilton_fed_27.txt Hamilton_fed_28.txt Hamilton_fed_29.txt Hamilton_fed_30.txt
##                 690                 755                1010                 948
## Hamilton_fed_31.txt Hamilton_fed_32.txt Hamilton_fed_33.txt Hamilton_fed_34.txt
##                 797                 686                 773                1020
## Hamilton_fed_35.txt Hamilton_fed_36.txt Hamilton_fed_59.txt  Hamilton_fed_6.txt
##                1052                1272                 860                 984
## Hamilton_fed_60.txt Hamilton_fed_61.txt Hamilton_fed_65.txt Hamilton_fed_66.txt
##                1006                 681                 912                 997
## Hamilton_fed_67.txt Hamilton_fed_68.txt Hamilton_fed_69.txt  Hamilton_fed_7.txt
##                 781                 683                1359                1073
## Hamilton_fed_70.txt Hamilton_fed_71.txt Hamilton_fed_72.txt Hamilton_fed_73.txt
##                1436                 766                 925                1061
## Hamilton_fed_74.txt Hamilton_fed_75.txt Hamilton_fed_76.txt Hamilton_fed_77.txt
##                 478                 905                 883                 887
## Hamilton_fed_78.txt Hamilton_fed_79.txt  Hamilton_fed_8.txt Hamilton_fed_80.txt
##                1376                 478                 998                1132
## Hamilton_fed_81.txt Hamilton_fed_82.txt Hamilton_fed_83.txt Hamilton_fed_84.txt
##                1798                 749                2620                1907
## Hamilton_fed_85.txt  Hamilton_fed_9.txt       HM_fed_18.txt       HM_fed_19.txt
##                1264                 931                1029                1023
##       HM_fed_20.txt       Jay_fed_2.txt       Jay_fed_3.txt       Jay_fed_4.txt
##                 776                 804                 736                 780
##       Jay_fed_5.txt      Jay_fed_64.txt  Madison_fed_10.txt  Madison_fed_14.txt
##                 657                1072                1437                1016
##  Madison_fed_37.txt  Madison_fed_38.txt  Madison_fed_39.txt  Madison_fed_40.txt
##                1268                1529                1169                1340
##  Madison_fed_41.txt  Madison_fed_42.txt  Madison_fed_43.txt  Madison_fed_44.txt
##                1701                1330                1601                1382
##  Madison_fed_45.txt  Madison_fed_46.txt  Madison_fed_47.txt  Madison_fed_48.txt
##                1018                1233                1306                 846
##  Madison_fed_58.txt
##                 978
```

Normalization

Create a normalized version of Papers_DTM

```
Papers_M <- as.matrix(Papers_DTM)
Papers_M_N1 <- apply(Papers_M, 1, function(i) round(i/sum(i),3))
Papers_Matrix_Norm <- t(Papers_M_N1)
(Papers_M[c(1:11),c(1000:1010)])
```

```
##                    Terms
## Docs                crude cruel crush culpabl cultiv culumni cun cupid cure
##   dispt_fed_49.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_50.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_51.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_52.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_53.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_54.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_55.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_56.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_57.txt     0     0     0       0      0       0   0     0    0
##   dispt_fed_62.txt     0     0     0       0      1       0   0     0    0
##   dispt_fed_63.txt     0     0     1       0      0       0   0     0    0
##                    Terms
## Docs                curios curious
##   dispt_fed_49.txt      0       0
##   dispt_fed_50.txt      0       0
##   dispt_fed_51.txt      0       0
##   dispt_fed_52.txt      0       0
##   dispt_fed_53.txt      1       0
##   dispt_fed_54.txt      0       0
##   dispt_fed_55.txt      0       0
##   dispt_fed_56.txt      0       0
##   dispt_fed_57.txt      0       0
##   dispt_fed_62.txt      0       0
##   dispt_fed_63.txt      0       0
```

```
Terms
```

```
## function (x)
## UseMethod("Terms")
## <bytecode: 0x0000000018e869c0>
## <environment: namespace:tm>
```

```
(Papers_Matrix_Norm[c(1:11),c(1000:1010)])
```

```
##                    Terms
## Docs               crude cruel crush culpabl cultiv culumni cun cupid cure
##   dispt_fed_49.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_50.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_51.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_52.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_53.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_54.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_55.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_56.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_57.txt     0     0 0.000       0  0.000       0   0     0    0
##   dispt_fed_62.txt     0     0 0.000       0  0.001       0   0     0    0
##   dispt_fed_63.txt     0     0 0.001       0  0.000       0   0     0    0
##                    Terms
## Docs               curios curious
##   dispt_fed_49.txt  0.000       0
##   dispt_fed_50.txt  0.000       0
##   dispt_fed_51.txt  0.000       0
##   dispt_fed_52.txt  0.000       0
##   dispt_fed_53.txt  0.001       0
##   dispt_fed_54.txt  0.000       0
##   dispt_fed_55.txt  0.000       0
##   dispt_fed_56.txt  0.000       0
##   dispt_fed_57.txt  0.000       0
##   dispt_fed_62.txt  0.000       0
##   dispt_fed_63.txt  0.000       0
```

Data Structures

Convert to matrix and view

```
Papers_dtm_matrix = as.matrix(Papers_DTM)
str(Papers_dtm_matrix)
```

```
##  num [1:85, 1:4900] 0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ Docs : chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fed_52.txt"
##   ..$ Terms: chr [1:4900] "abandon" "abat" "abb" "abet" ...
```

```
(Papers_dtm_matrix[c(1:11),c(2:10)])
```

```
##                    Terms
## Docs               abat abb abet abhorr abil abject abl ablest abolish
##   dispt_fed_49.txt    0   0    0      0    0      0   2      0       0
##   dispt_fed_50.txt    0   0    0      0    0      0   0      0       0
##   dispt_fed_51.txt    0   0    0      0    0      0   1      0       0
##   dispt_fed_52.txt    0   0    0      0    1      0   1      0       0
##   dispt_fed_53.txt    1   0    0      0    0      0   0      0       0
##   dispt_fed_54.txt    0   0    0      0    0      0   0      0       0
##   dispt_fed_55.txt    0   0    0      0    0      0   0      0       0
##   dispt_fed_56.txt    0   0    0      0    0      0   0      0       0
##   dispt_fed_57.txt    0   0    0      1    0      0   0      0       0
##   dispt_fed_62.txt    0   0    0      0    0      0   1      0       0
##   dispt_fed_63.txt    0   0    0      0    0      0   4      0       0
```

#Also convert to DF

```
Papers_DF <- as.data.frame(as.matrix(Papers_DTM))
str(Papers_DF)
```

```
## ’data.frame':    85 obs. of  4900 variables:
##  $ abandon        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abat           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ abb            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abet           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abhorr         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ abil           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ abject         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abl            : num  2 0 1 1 0 0 0 0 0 1 ...
##  $ ablest         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abolish        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abolit         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abort          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abound         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abraham        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abreg          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abridg         : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ abroad         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abrog          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ absenc         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ absolut        : num  0 2 2 1 0 0 0 0 0 0 ...
##  $ absolv         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ absorb         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abstain        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abstract       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abstrus        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ absurd         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abund          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ abus           : num  1 1 2 1 1 0 0 0 0 0 ...
##  $ abyss          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acced          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acceler        : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ accept         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ access         : num  0 0 0 2 0 0 0 0 0 0 ...
##  $ accid          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accident       : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ accommod       : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ accomod        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accompani      : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ accomplic      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accomplish     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accord         : num  0 0 0 0 1 2 2 1 1 0 ...
##  $ account        : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ accret         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accru          : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ accumul        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accur          : num  1 0 0 0 1 0 0 0 0 1 ...
##  $ accuraci       : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ accus          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ accustom       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ achaean        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ achaeus        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ achaia         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ achiev         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acknowledg     : num  0 1 0 0 0 0 0 0 0 1 ...
##  $ acquaint       : num  1 0 0 0 2 0 0 2 0 1 ...
##  $ acquiesc       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acquir         : num  1 0 0 0 5 0 0 2 0 0 ...
##  $ acquisit       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acquit         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acr            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ act            : num  0 0 0 1 2 1 0 1 0 1 ...
##  $ action         : num  0 0 1 0 0 0 0 0 0 1 ...
##  $ activ          : num  0 4 0 0 0 0 0 0 0 0 ...
##  $ actor          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ actual         : num  1 2 0 0 4 0 0 0 1 0 ...
##  $ actuat         : num  0 0 0 0 0 0 1 0 1 0 ...
##  $ acut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adag           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adapt          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ add            : num  0 0 0 0 1 0 0 1 1 0 ...
##  $ addict         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ addit          : num  0 0 1 1 0 0 0 0 1 1 ...
##  $ address        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adduc          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adept          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adequ          : num  1 1 0 0 0 0 0 0 0 0 ...
##  $ adher          : num  0 0 1 0 0 1 0 0 0 0 ...
##  $ adjac          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adjoin         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adjourn        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adjud          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adjudg         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adjust         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ administ       : num  0 0 2 0 0 0 0 0 0 1 ...
##  $ administr      : num  1 2 1 0 0 0 0 0 1 0 ...
##  $ admir          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ admiralgener   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ admiralti      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ admiss         : num  0 0 0 0 0 1 0 0 1 1 ...
##  $ admit          : num  1 0 3 0 1 5 2 0 1 0 ...
##  $ admitt         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ admonish       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ admonit        : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ adopt          : num  0 0 0 1 0 1 0 0 0 1 ...
##  $ adroit         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ adul           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ advanc         : num  0 0 0 0 1 0 0 1 1 2 ...
##  $ advantag       : num  4 1 0 2 2 4 0 1 0 7 ...
##  $ adventiti      : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
(Papers_DF$abolit)
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 3 0 0 1 1 0 0 0 0 0 0 0 1 0 0 1 0 0
## [77] 0 0 0 0 0 0 0 0 0
```

```
(nrow(Papers_DF))   ## Each row is Paper
```

```
## [1] 85
```

Add row names

```
Papers_DF1<- Papers_DF%>%add_rownames()
```

```
## Warning: `add_rownames()` was deprecated in dplyr 1.0.0.
## Please use `tibble::rownames_to_column()` instead.
```

```
names(Papers_DF1)[1]<-"Author"
Papers_DF1[1:11,1]="dispt"
Papers_DF1[12:65,1]="hamil"
Papers_DF1[66:70,1]="jay"
Papers_DF1[71:85,1]="madis"
head(Papers_DF1[,1:2],20)
```

```
## # A tibble: 20 x 2
##    Author abandon
##    <chr>    <dbl>
##  1 dispt        0
##  2 dispt        0
##  3 dispt        0
##  4 dispt        0
##  5 dispt        0
##  6 dispt        0
##  7 dispt        0
##  8 dispt        0
##  9 dispt        0
## 10 dispt        0
## 11 dispt        0
## 12 hamil        0
## 13 hamil        0
## 14 hamil        0
## 15 hamil        0
## 16 hamil        2
## 17 hamil        0
## 18 hamil        0
## 19 hamil        0
## 20 hamil        0
```

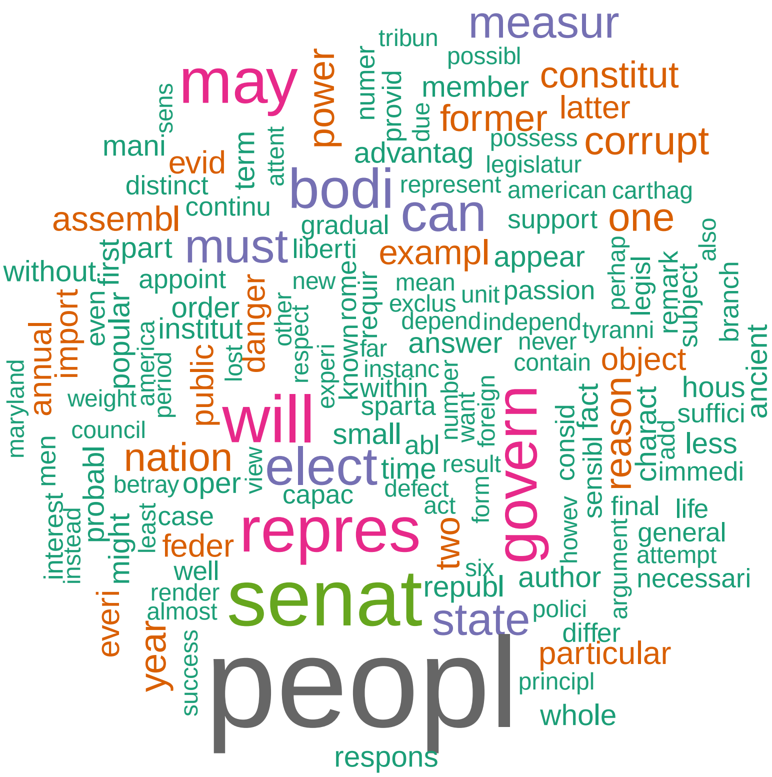
```
tail(Papers_DF1[,1:2],20)
```

```
## # A tibble: 20 x 2
##    Author abandon
##    <chr>    <dbl>
##  1 jay          0
##  2 jay          0
##  3 jay          0
##  4 jay          0
##  5 jay          0
##  6 madis        0
##  7 madis        0
##  8 madis        0
##  9 madis        1
## 10 madis        1
## 11 madis        0
## 12 madis        0
## 13 madis        0
## 14 madis        0
## 15 madis        0
## 16 madis        0
## 17 madis        0
## 18 madis        0
## 19 madis        0
## 20 madis        0
```

Example Word Cloud

Wordcloud Visualization Hamilton, Madison and Disputed Papers

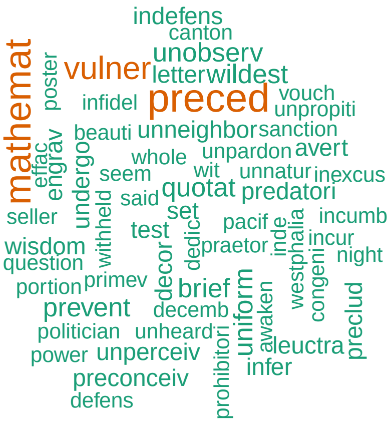
```
DisputedPapersWC<- wordcloud(colnames(Papers_dtm_matrix), Papers_dtm_matrix[11, ],
                             rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



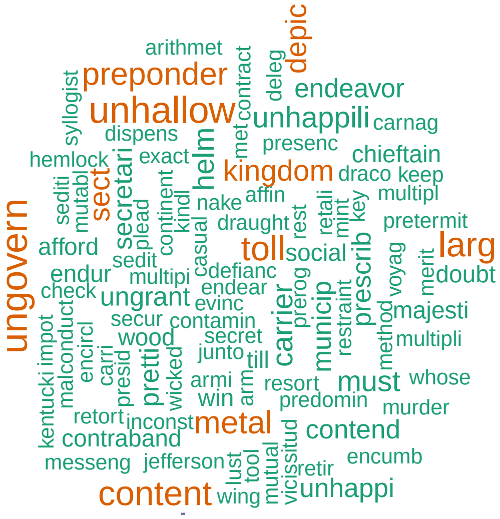
```
(head(sort(as.matrix(Papers_DTM)[11,], decreasing = TRUE), n=50))
```

```
##      peopl      senat       will        may     repres     govern       bodi
##         42         24         19         18         18         16         15
##        can      elect       must     measur      state    corrupt     nation
##         14         14         12         11         11          9          9
##        one  constitut     former      power     reason       year    assembl
##          9          8          8          8          8          8          7
##     exampl        two     annual     danger      everi       evid      feder
##          7          7          6          6          6          6          6
##     import     latter     object particular     public   advantag    ancient
##          6          6          6          6          6          5          5
##     answer     appear     author    charact       fact      first       hous
##          5          5          5          5          5          5          5
##   institut       less       mani     member      might       oper      order
##          5          5          5          5          5          5          5
##       part
##          5
```

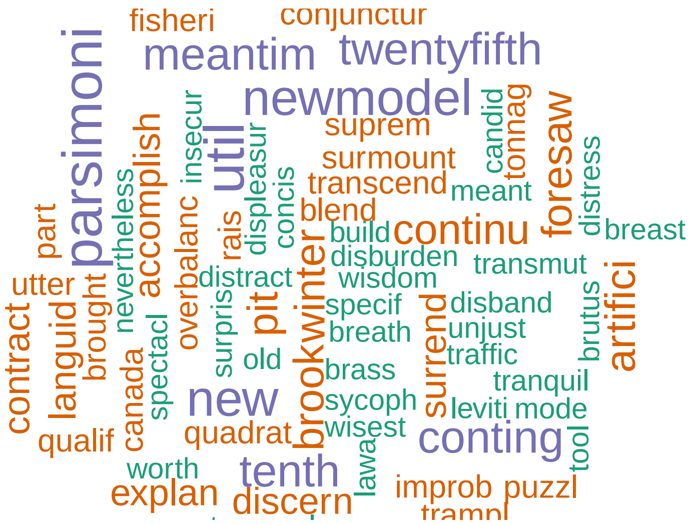
```r
HamiltonPapersWC <- wordcloud(colnames(Papers_dtm_matrix), Papers_dtm_matrix[12:65, ],
                              rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



```r
MadisonPapersHW <- wordcloud(colnames(Papers_dtm_matrix), Papers_dtm_matrix[71:85, ],
                             rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



```r
JayPapersHW <- wordcloud(colnames(Papers_dtm_matrix), Papers_dtm_matrix[66:70, ],
                         rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```



```r
DisputedWC <- wordcloud(colnames(Papers_dtm_matrix), Papers_dtm_matrix[1:11, ],
                        rot.per = .35, colors = brewer.pal(n = 8, name = "Dark2"))
```

stadthold
foresight
illinform
match
recognit
illhumor
mixtur
defeat
john
modif
diminut
matter
record
paus
squar
recours
cautionari
hee
held
creep
marlborough
deep
committe
defenseless
clash
materi
victim
dimens
illusori
heir
fox
somewhat
forego
modifi
inelig
signatur
marqu
stale
creditor
victori
iferior
dilemma
hirel
half
shut
forcibl
jar
illtim
jaw
illog
franc
stamp
mean
admiralti
recommendatori
impart
jealousi
shelter
purchas
mclean
judgement
volumin
stagnat
illicit
defer
stabil
cheaper
chord
irregular

################## # Analysis ################## # Distance Metrics ################## #Computing different distance matrices to determine which seems to work the best! ###Distance Measure

```
m <- Papers_dtm_matrix
m_norm <- Papers_Matrix_Norm
```

#m <- [1:2, 1:3]

```
distMatrix_E <- dist(m, method="euclidean")
```

```
#print(distMatrix_E)
```

```
distMatrix_M <- dist(m, method="manhattan")
```

#print(distMatrix_M)

```
distMatrix_C <- dist(m, method="cosine")
```

print(distMatrix_C)

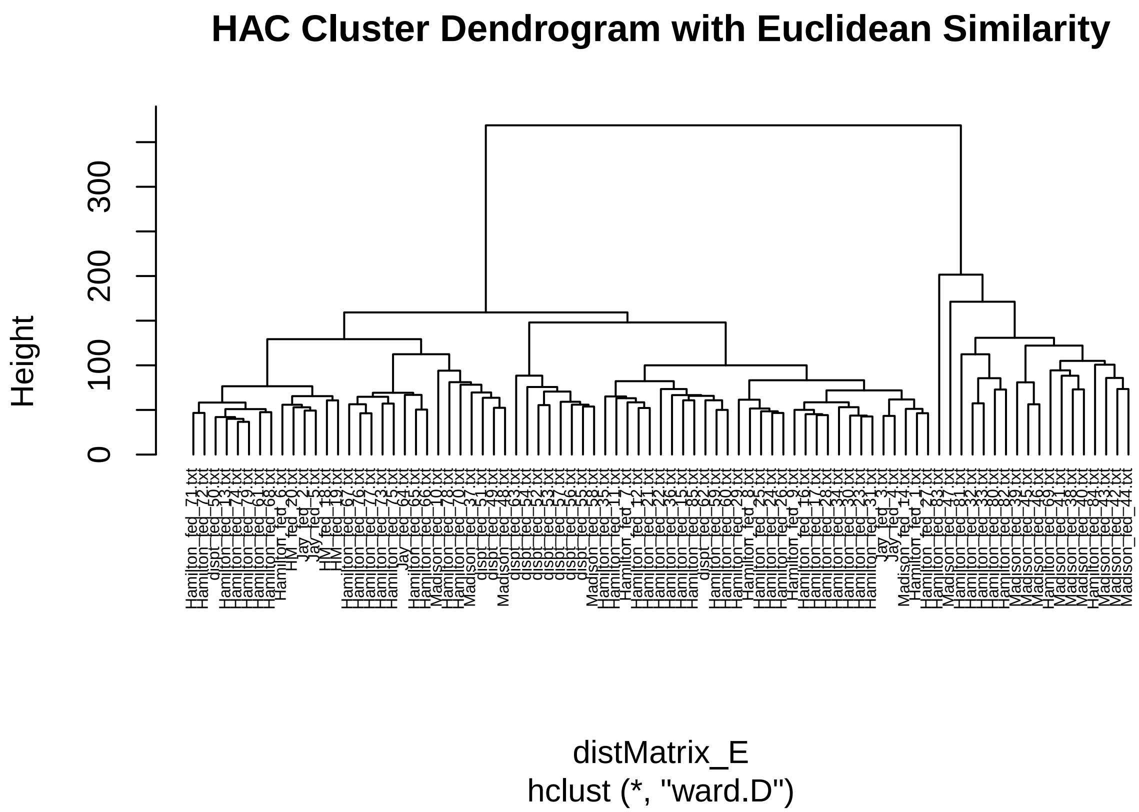
```r
distMatrix_C_norm <- dist(m_norm, method="cosine")
```

#print(distMatrix_C_norm)

## Clustering

###Clustering Methods: ## HAC: Hierarchical Algorithm Clustering Method ## Euclidean

```r
groups_E <- hclust(distMatrix_E,method="ward.D")
plot(groups_E, cex=0.5, font=22, hang=-1, main = "HAC Cluster Dendrogram with Euclidean Similarity")
```

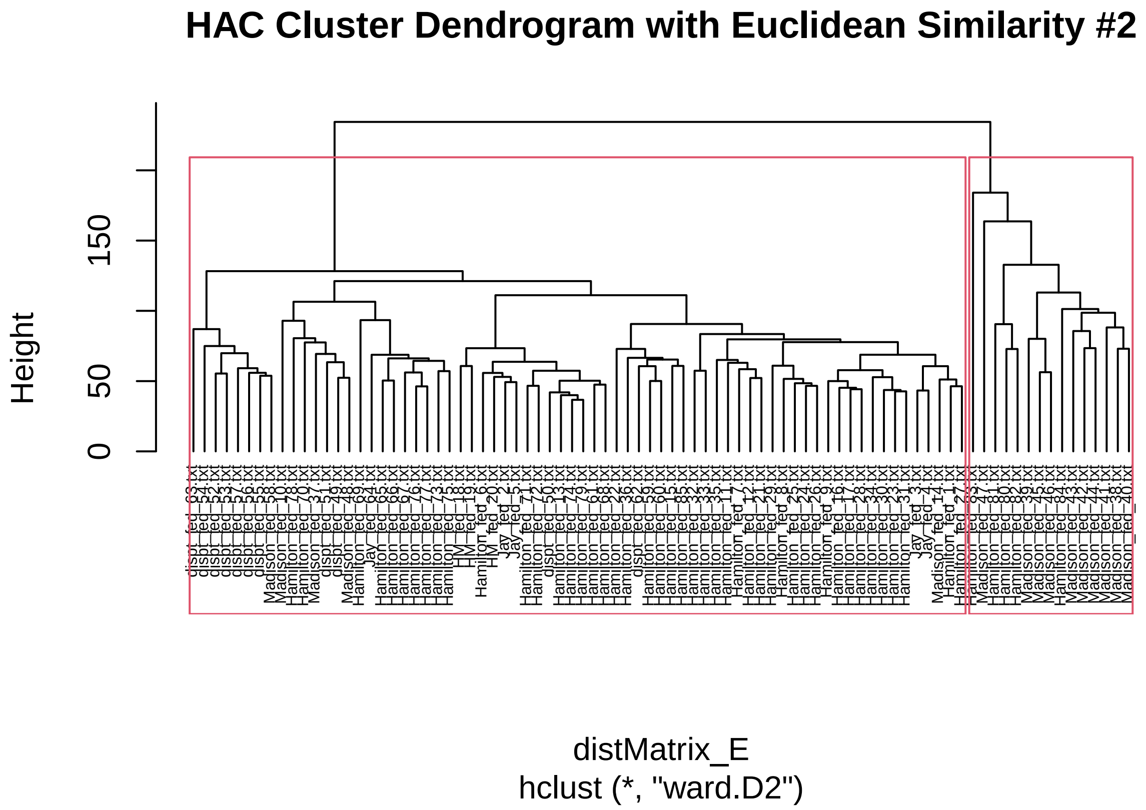


# Plots the separations

```r
#rect.hclust(groups_E, k=2)
```

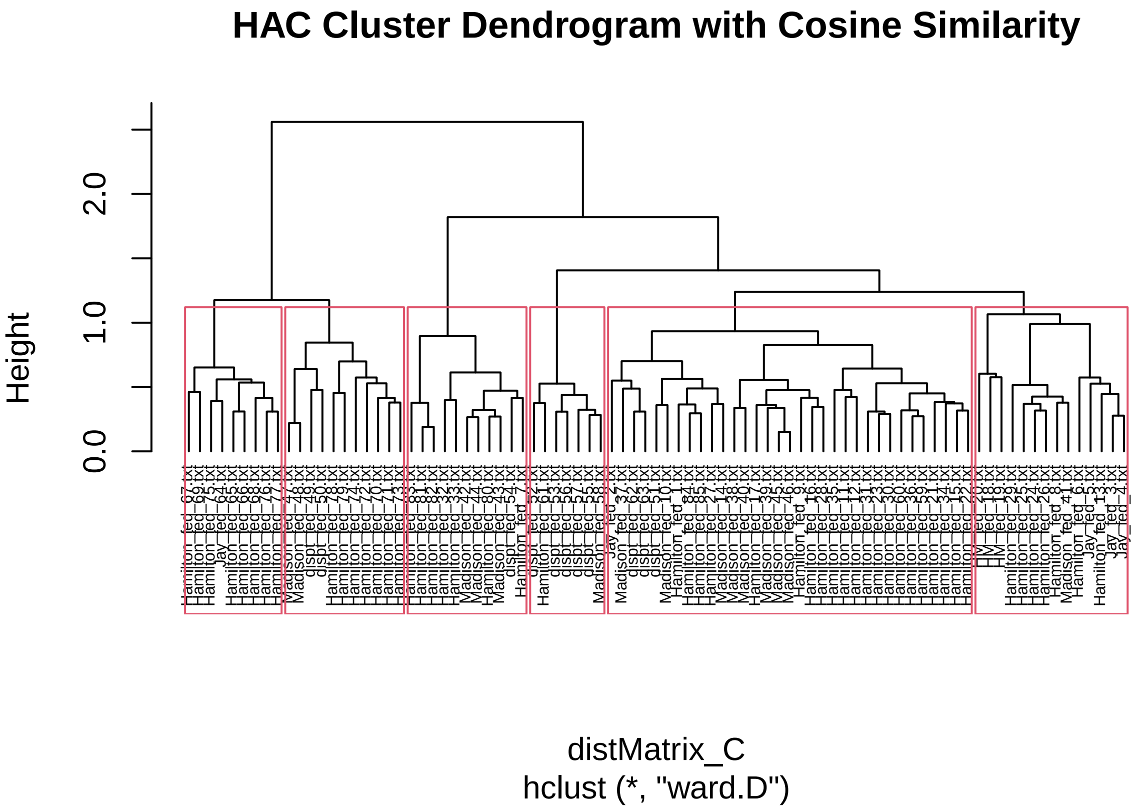
#HAC Cluster Dendrogram with Euclidean Similarity

```r
distMatrix_E1 <- hclust(distMatrix_E, "ward.D2")
plot(distMatrix_E1, cex=0.5, font=22, hang=-1, main = "HAC Cluster Dendrogram with Euclidean Similarity
rect.hclust(distMatrix_E1, k=2)
```



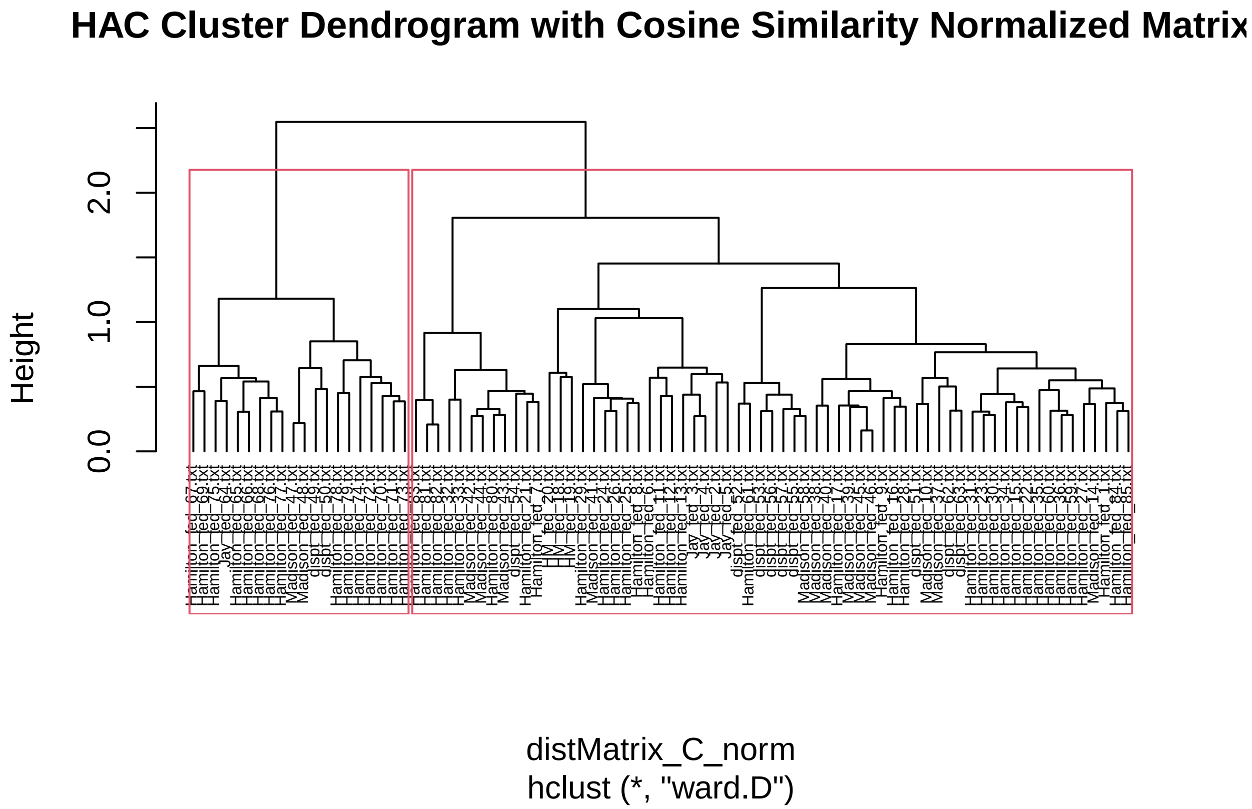
# HAC Cluster Dendrogram with Cosine Similarity

```
groups_C <- hclust(distMatrix_C,method="ward.D")
plot(groups_C, cex=0.5,font=22, hang=-1,main = "HAC Cluster Dendrogram with Cosine Similarity")
rect.hclust(groups_C, k=6)
```



## Cosine Similarity for Normalized Matrix

```
groups_C_n <- hclust(distMatrix_C_norm,method="ward.D")
plot(groups_C_n, cex=0.5, font=22, hang=-1, main = "HAC Cluster Dendrogram with Cosine Similarity Normal
rect.hclust(groups_C_n, k=2)
```



##################### # k means clustering Methods #####################

```
X <- m_norm
k2 <- kmeans(X, centers = 2, nstart = 100, iter.max = 50)
str(k2)
```

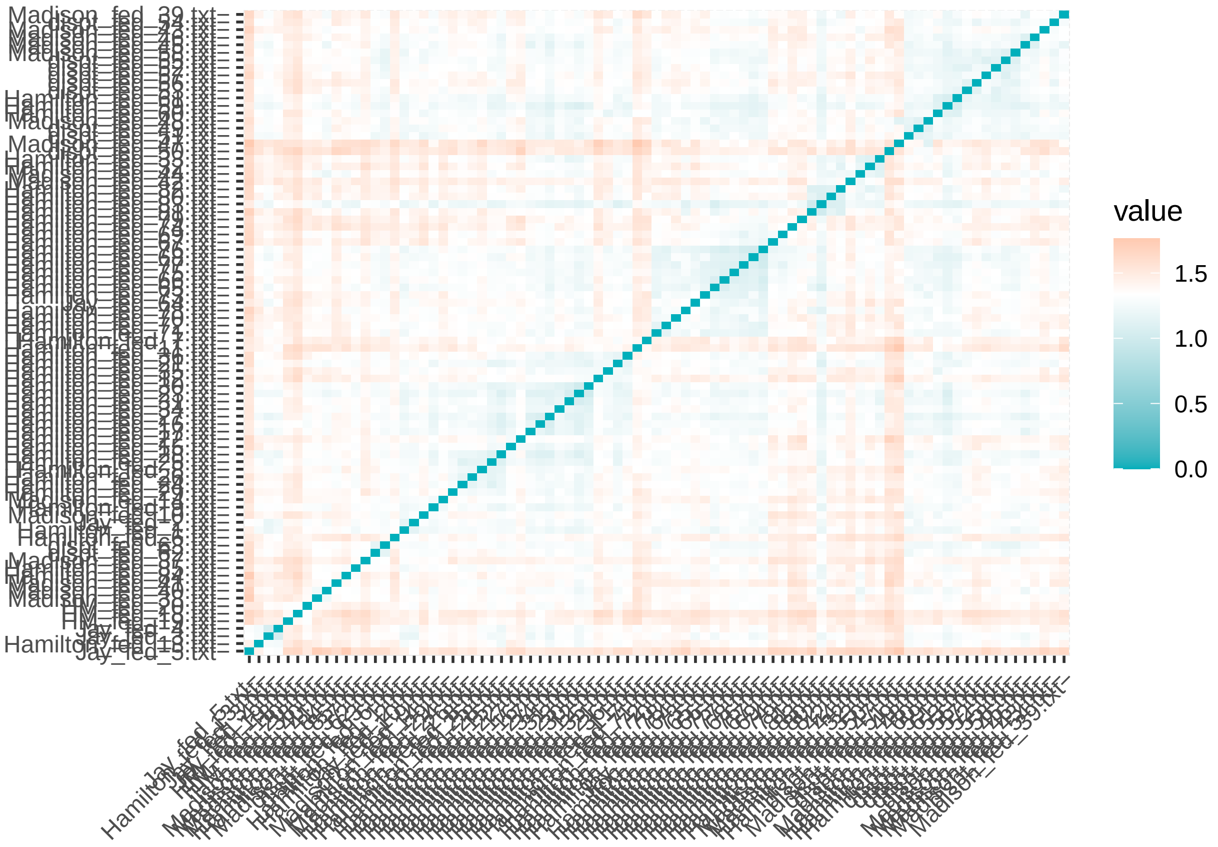
```
## List of 9
##  $ cluster     : Named int [1:85] 1 1 1 2 2 2 1 2 2 1 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:2, 1:4900] 1.09e-04 6.67e-05 1.82e-05 3.33e-05 9.09e-05 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "1" "2"
##   .. ..$ : chr [1:4900] "abandon" "abat" "abb" "abet" ...
##  $ totss       : num 0.216
##  $ withinss    : num [1:2] 0.1231 0.0794
##  $ tot.withinss: num 0.203
##  $ betweenss   : num 0.0137
##  $ size        : int [1:2] 55 30
##  $ iter        : int 1
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```
k3 <- kmeans(X, centers = 7, nstart = 50, iter.max= 50)
str(k3)
```

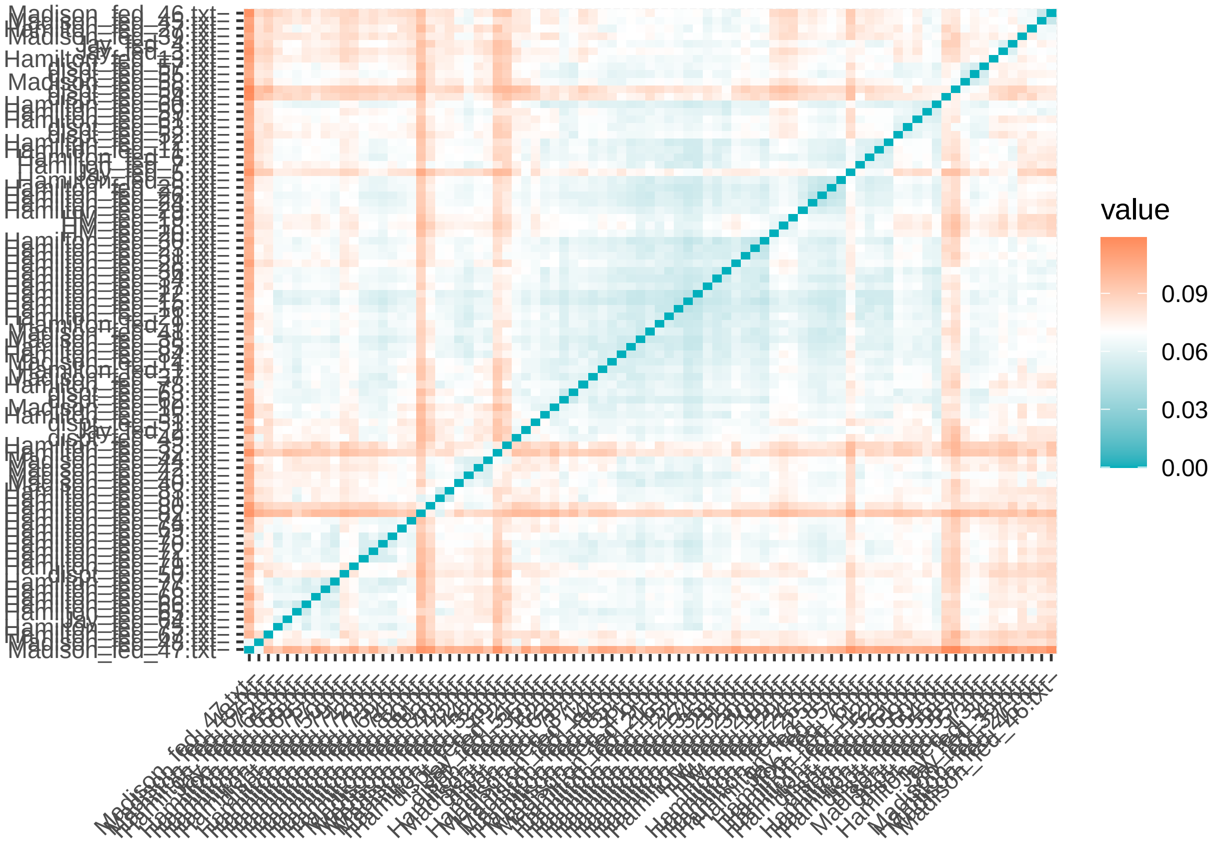
```
## List of 9
##  $ cluster     : Named int [1:85] 7 7 1 4 4 4 4 4 4 1 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_f
##  $ centers     : num [1:7, 1:4900] 0.000214 0 0.000125 0 0 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:7] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:4900] "abandon" "abat" "abb" "abet" ...
##  $ totss       : num 0.216
##  $ withinss    : num [1:7] 0.02827 0.01163 0.05749 0.01622 0.00201 ...
##  $ tot.withinss: num 0.163
##  $ betweenss   : num 0.0531
##  $ size        : int [1:7] 14 6 32 8 2 4 19
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

k means visualization results

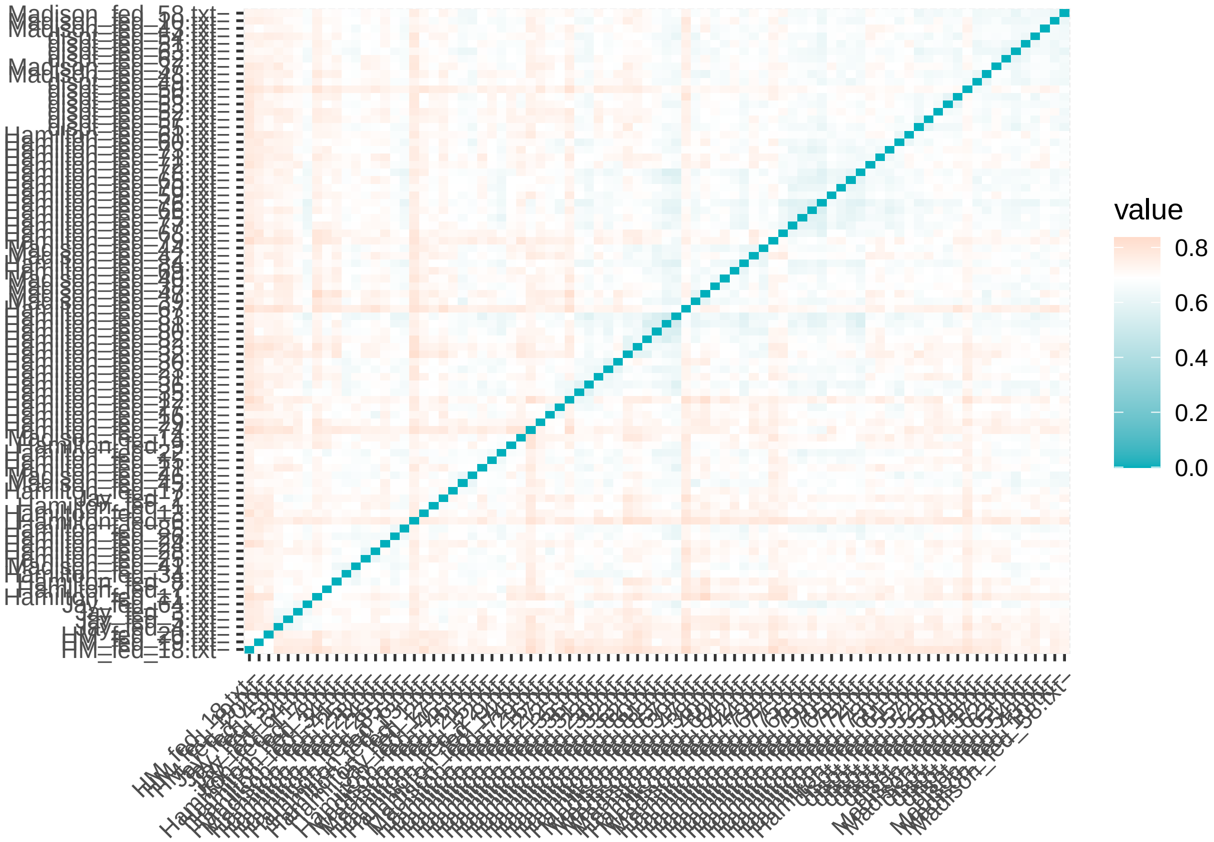
```
distance1 <- get_dist(X,method = "manhattan")
fviz_dist(distance1, gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07"))
```



```
distance2 <- get_dist(X,method = "euclidean")
fviz_dist(distance2, gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07"))
```



```
distance3 <- get_dist(X,method = "spearman")
fviz_dist(distance3, gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07", title= "Distance
```



# Visualize the k-means results

```r
str(X)
```

```
##  num [1:85, 1:4900] 0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ Docs : chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fed_52.txt"
##   ..$ Terms: chr [1:4900] "abandon" "abat" "abb" "abet" ...
```

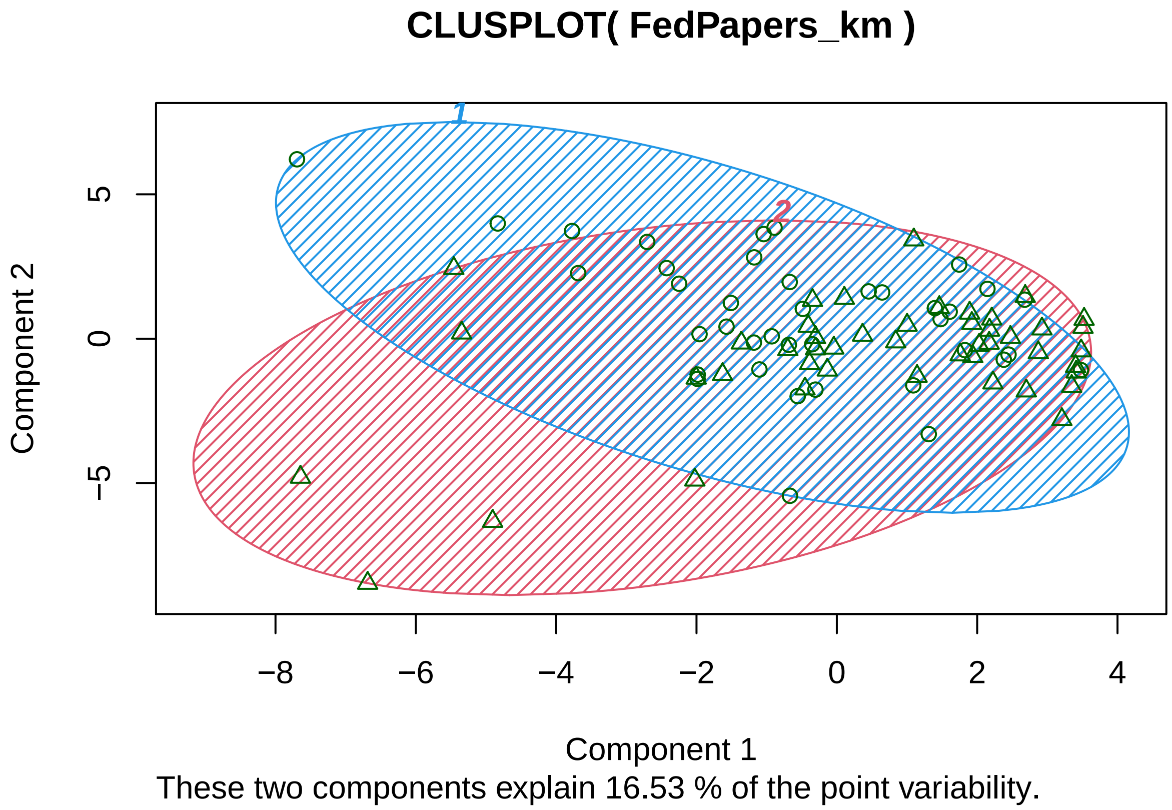
```r
kmeansFIT_1 <- kmeans(X, centers = 4)
summary(kmeansFIT_1)
```

```
##              Length Class  Mode
## cluster         85  -none- numeric
## centers      19600  -none- numeric
## totss            1  -none- numeric
## withinss         4  -none- numeric
## tot.withinss     1  -none- numeric
## betweenss        1  -none- numeric
## size             4  -none- numeric
## iter             1  -none- numeric
## ifault           1  -none- numeric
```

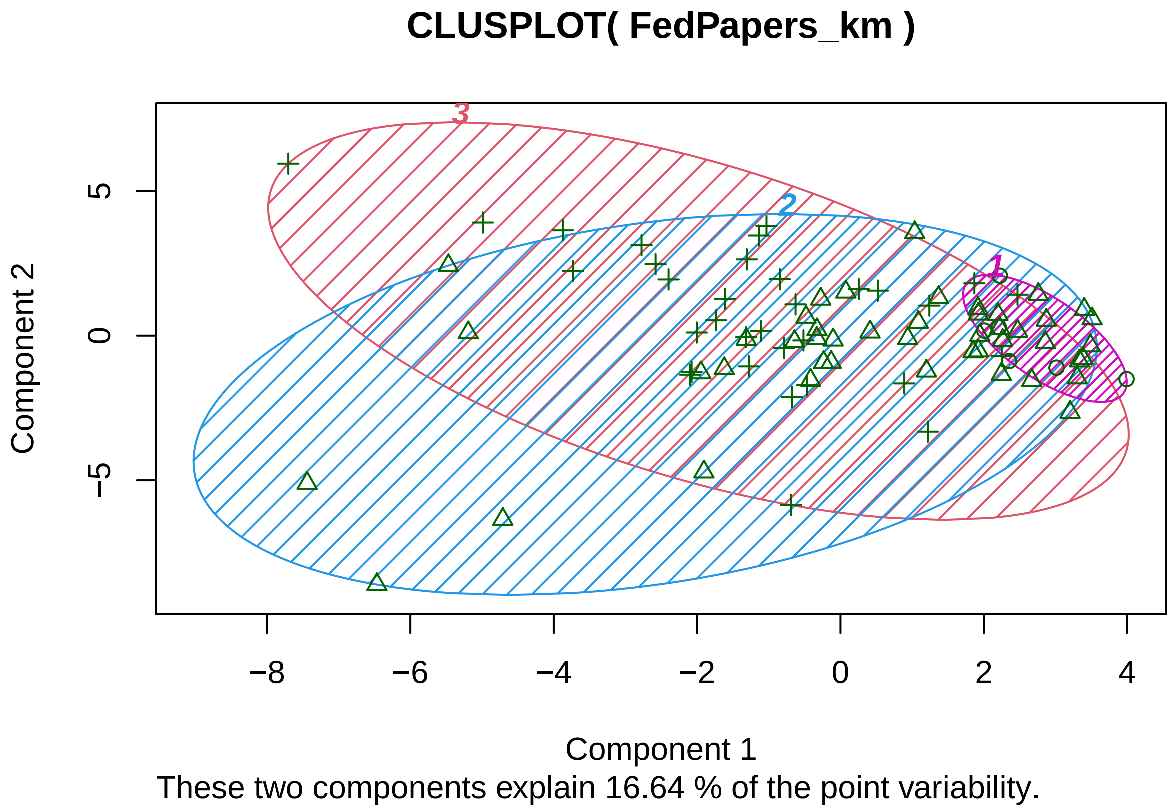
#Loop to be fancy

```
x <- c(2,3,4,5,6,7,8,9)
set.seed(20)
for (val in x){
  print(val)
  # run k-means
  Clusters <- kmeans(FedPapers_km, val)
  FedPapers_km$Clusters <- as.factor(Clusters$cluster)
  str(Clusters)
  Clusters$centers

  # Add clusters to dataframe original dataframe with author name
  FedPapers_km2 <- FederalistPapers
  FedPapers_km2$Clusters <- as.factor(Clusters$cluster)
  # Plot results
  #clusplot(FedPapers_km, FedPapers_km$Clusters, color=TRUE, shade=TRUE, labels=0, lines=0)

  clusplot(FedPapers_km, FedPapers_km$Clusters, color=T, shade=T, labels=4, lines=T)

  }
```

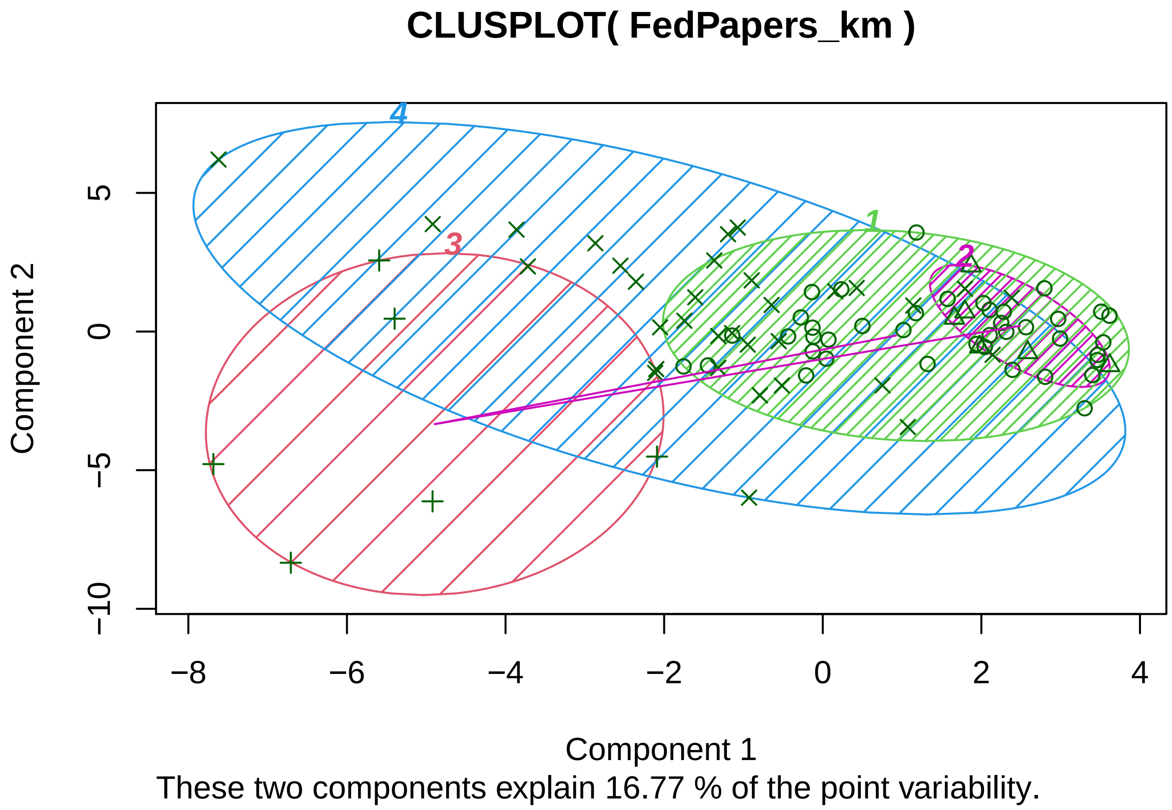
```
## [1] 2
## List of 9
##  $ cluster     : Named int [1:85] 2 1 2 1 1 2 1 1 2 1 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:2, 1:71] 0.3081 0.28063 0.05051 0.0548 0.00828 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "1" "2"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 515
##  $ withinss    : num [1:2] 33.1 79.5
##  $ tot.withinss: num 113
##  $ betweenss   : num 403
##  $ size        : int [1:2] 39 46
##  $ iter        : int 1
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



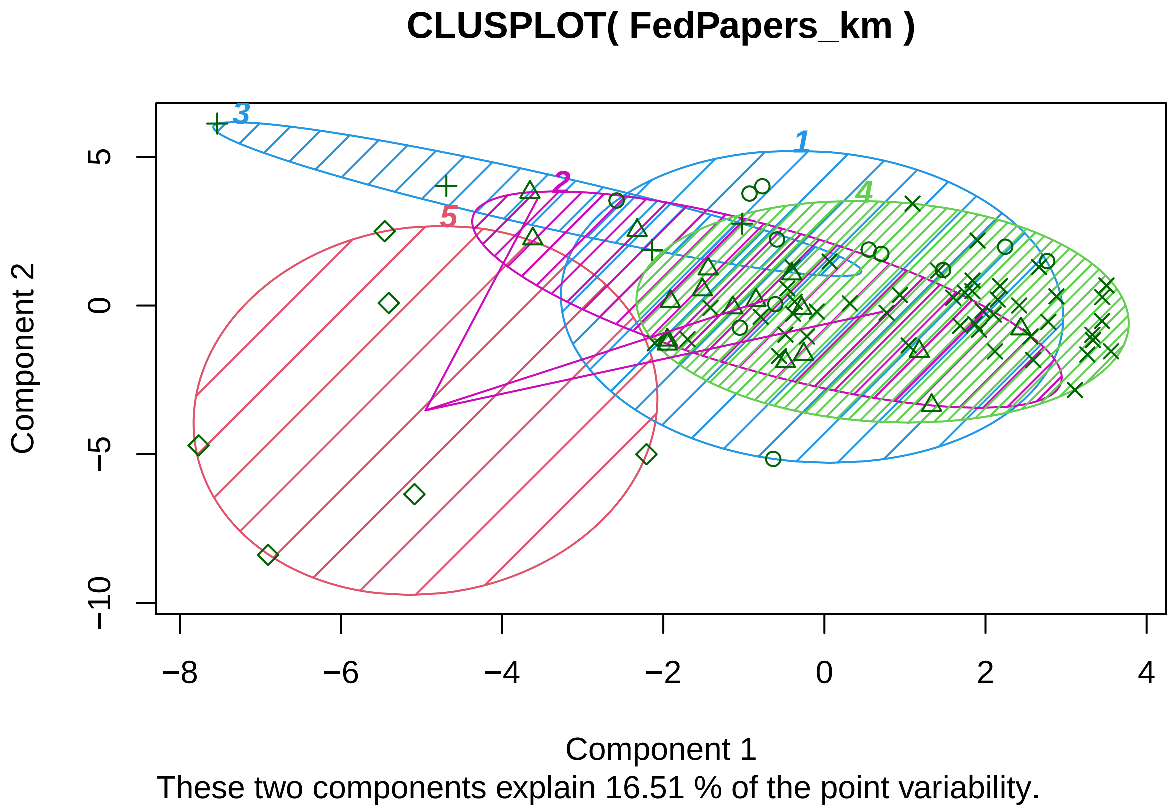
```
## [1] 3
## List of 9
##  $ cluster     : Named int [1:85] 2 3 2 3 3 2 3 3 2 3 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:3, 1:71] 0.3432 0.2806 0.3017 0.0372 0.0548 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3] "1" "2" "3"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 33.7
##  $ withinss    : num [1:3] 0.388 7.415 3.492
##  $ tot.withinss: num 11.3
##  $ betweenss   : num 22.4
##  $ size        : int [1:3] 6 46 33
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



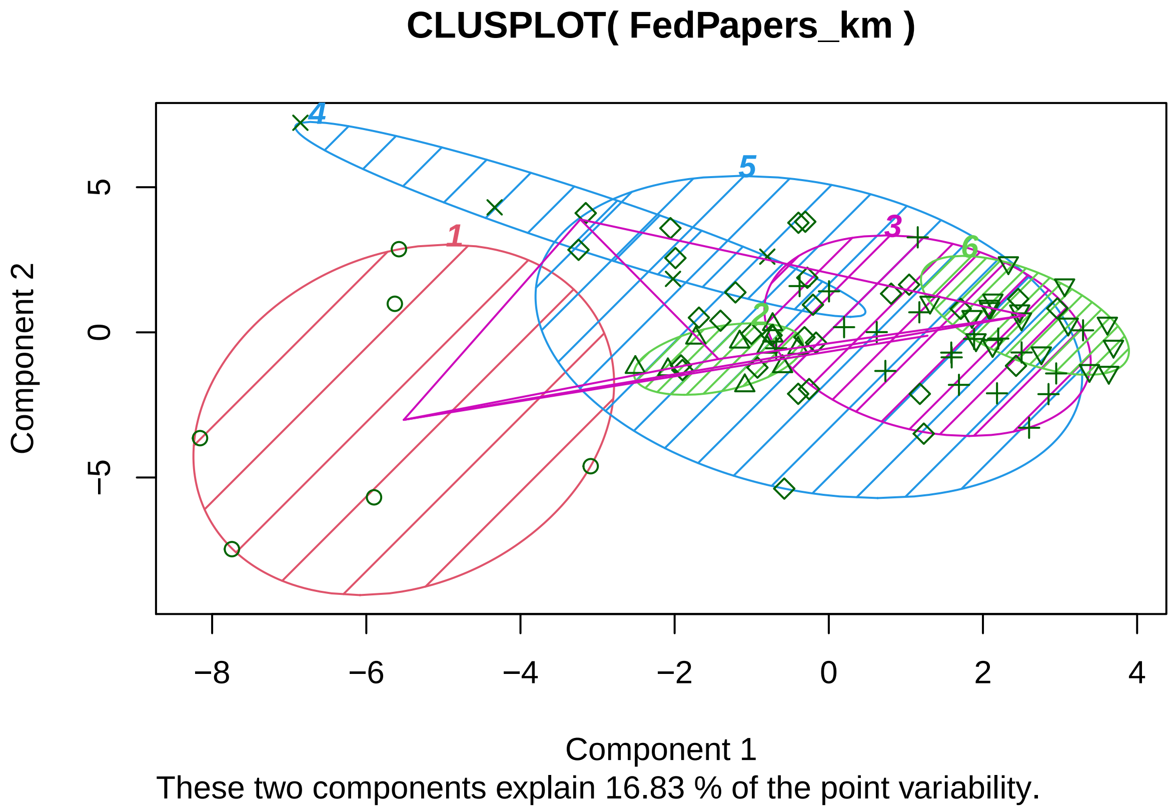
```
## [1] 4
## List of 9
##  $ cluster     : Named int [1:85] 1 4 1 4 4 1 4 4 1 4 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:4, 1:71] 0.2971 0.3432 0.1707 0.3017 0.0572 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "1" "2" "3" "4"
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 43
##  $ withinss    : num [1:4] 3.996 0.388 0.858 3.492
##  $ tot.withinss: num 8.73
##  $ betweenss   : num 34.3
##  $ size        : int [1:4] 40 6 6 33
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



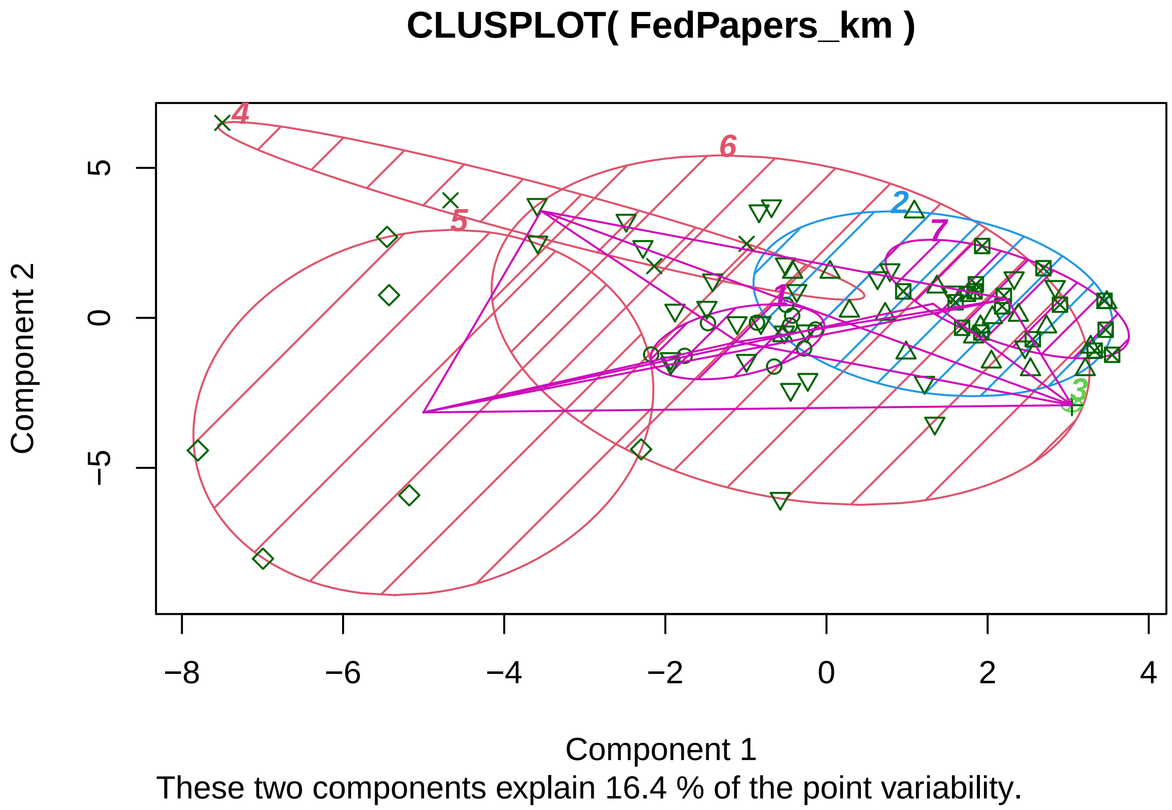
```
## [1] 5
## List of 9
##  $ cluster     : Named int [1:85] 4 2 4 2 2 4 2 1 4 2 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:5, 1:71] 0.355 0.285 0.213 0.303 0.171 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 179
##  $ withinss    : num [1:5] 1.008 1.195 0.264 9.921 0.858
##  $ tot.withinss: num 13.2
##  $ betweenss   : num 165
##  $ size        : int [1:5] 12 17 4 46 6
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



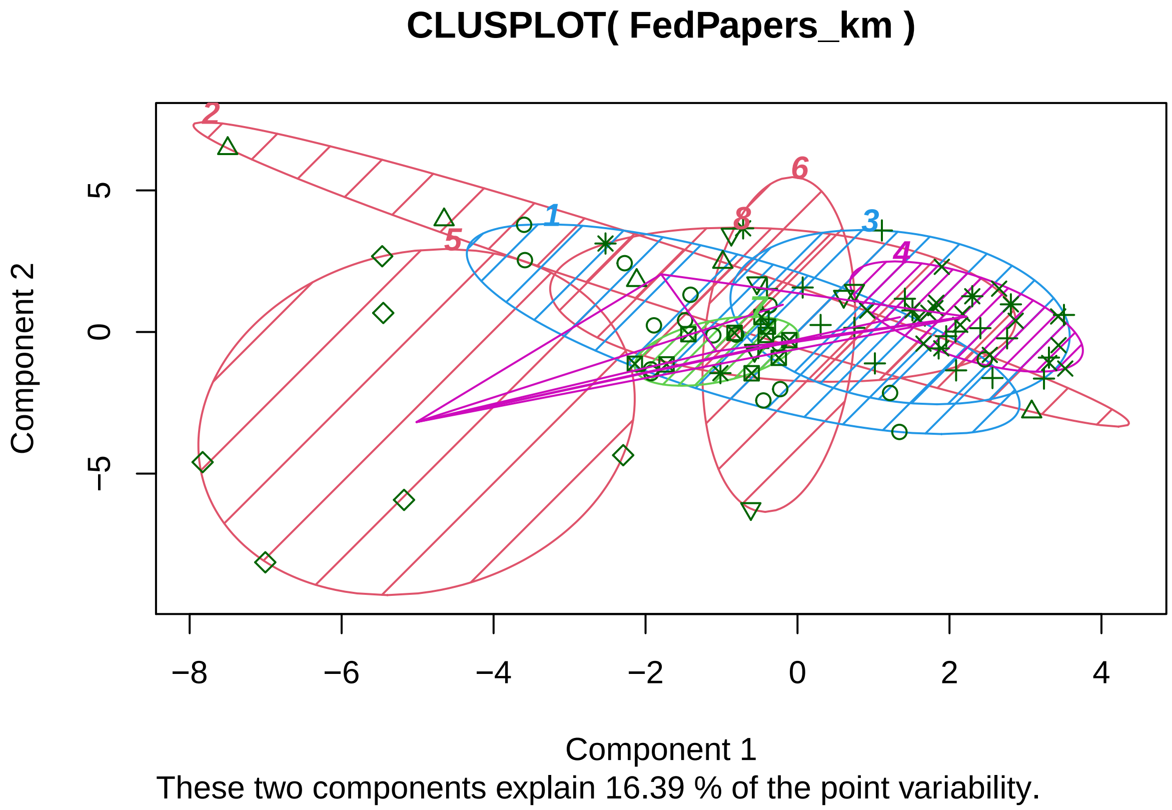
```
## [1] 6
## List of 9
##  $ cluster     : Named int [1:85] 2 5 2 5 5 2 5 5 2 5 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:6, 1:71] 0.171 0.27 0.286 0.213 0.314 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:6] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 144
##  $ withinss    : num [1:6] 0.858 0.755 1.416 0.264 9.734 ...
##  $ tot.withinss: num 14.5
##  $ betweenss   : num 130
##  $ size        : int [1:6] 6 10 19 4 29 17
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



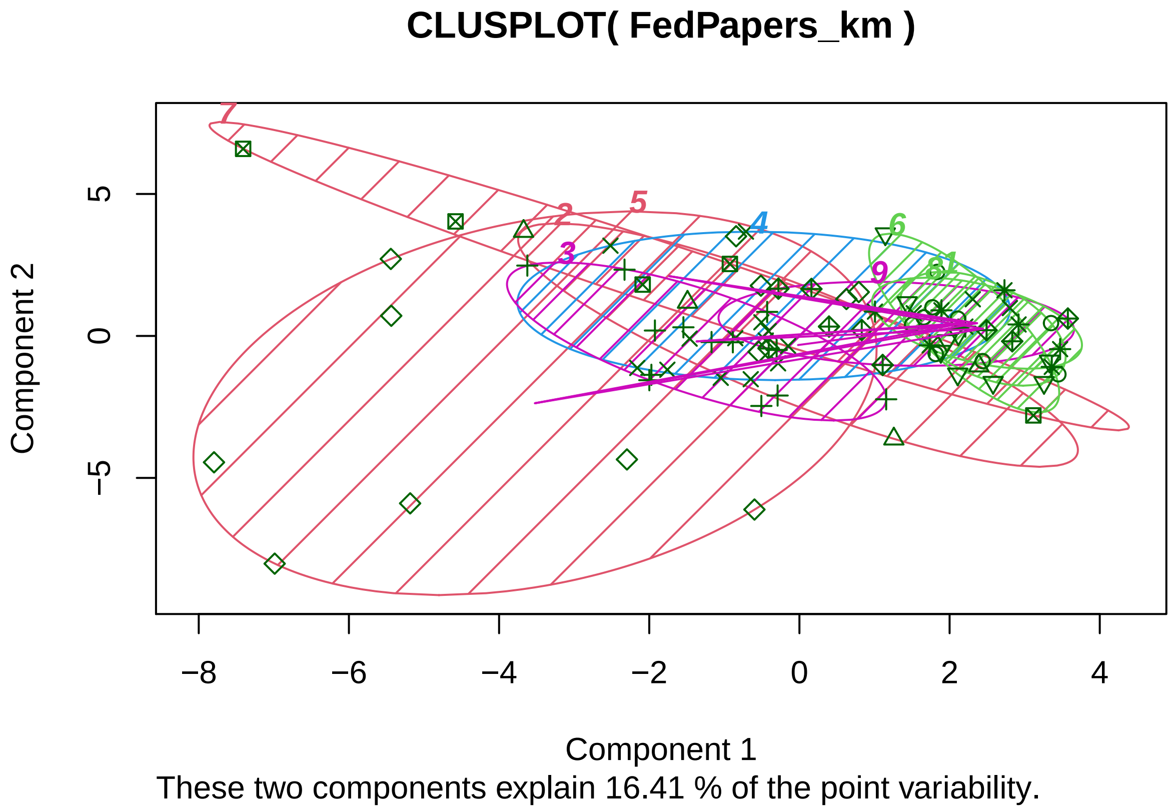
```
## [1] 7
## List of 9
##  $ cluster     : Named int [1:85] 1 6 1 6 6 1 6 6 1 6 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:7, 1:71] 0.27 0.286 0.27 0.213 0.171 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:7] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 222
##  $ withinss    : num [1:7] 0.755 1.26 0 0.264 0.858 ...
##  $ tot.withinss: num 7.3
##  $ betweenss   : num 215
##  $ size        : int [1:7] 10 18 1 4 6 29 17
##  $ iter        : int 2
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



```
## [1] 8
## List of 9
##  $ cluster     : Named int [1:85] 7 1 7 1 1 7 1 8 7 1 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:8, 1:71] 0.285 0.225 0.286 0.342 0.171 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:8] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 423
##  $ withinss    : num [1:8] 1.195 1.34 1.26 1.462 0.858 ...
##  $ tot.withinss: num 7.66
##  $ betweenss   : num 416
##  $ size        : int [1:8] 17 5 18 17 6 6 10 6
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

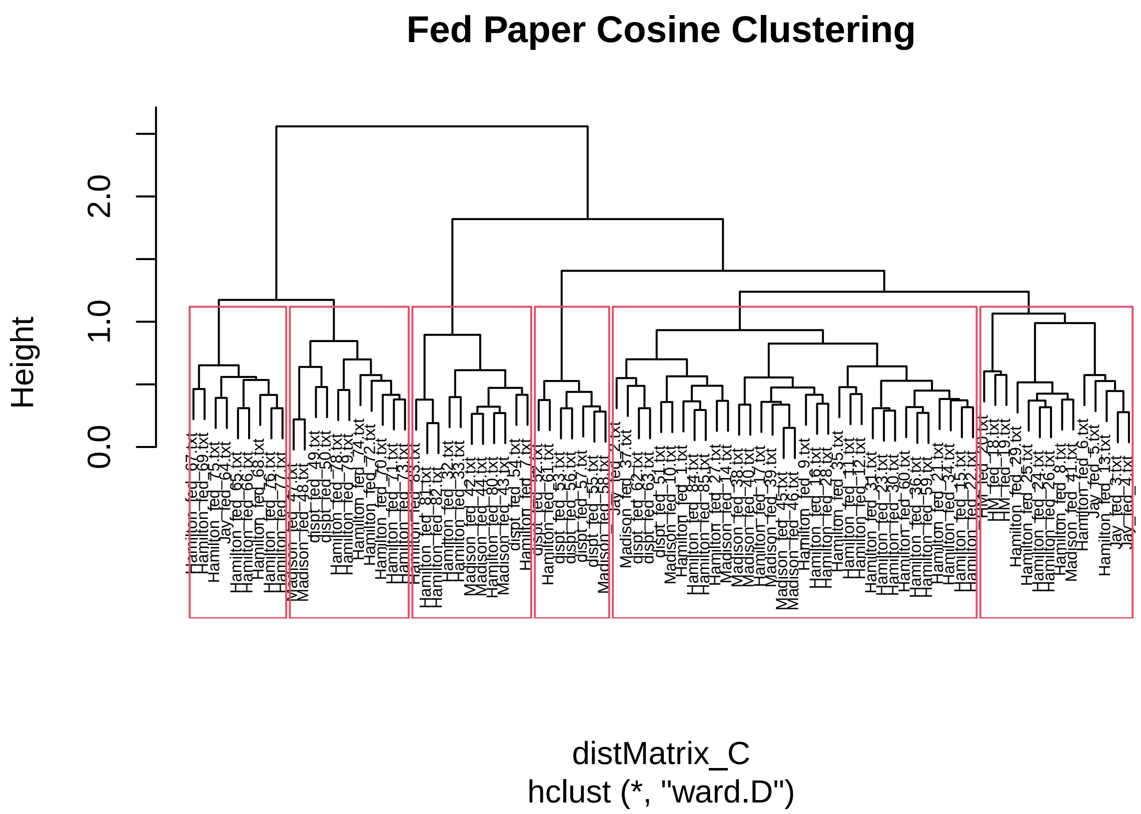


```
## [1] 9
## List of 9
##  $ cluster     : Named int [1:85] 4 2 4 2 3 4 3 4 4 3 ...
##   ..- attr(*, "names")= chr [1:85] "dispt_fed_49.txt" "dispt_fed_50.txt" "dispt_fed_51.txt" "dispt_fe
##  $ centers     : num [1:9, 1:71] 0.329 0.245 0.297 0.308 0.254 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:9] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:71] "a" "all" "also" "an" ...
##  $ totss       : num 419
##  $ withinss    : num [1:9] 0.675 0.289 0.744 5.623 5.194 ...
##  $ tot.withinss: num 14.7
##  $ betweenss   : num 404
##  $ size        : int [1:9] 9 4 13 16 12 9 5 8 9
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```



#Cosine Assignment of essays

```
plot(groups_C, main = "Fed Paper Cosine Clustering", cex = 0.5)
rect.hclust(groups_C, k=6)
```



```
authorCut <- cutree(groups_C, k = 6)

(Madison_cos <- FedPapers_km2[which((authorCut == "1") & FedPapers_km2$author == "dispt") ,c(1,2)])
```

```
##   author         filename
## 1  dispt dispt_fed_49.txt
## 2  dispt dispt_fed_50.txt
```

```
(Hamilton_cos <- FedPapers_km2[which((authorCut == "2") & FedPapers_km2$author == "dispt") ,c(1,2)])
```

```
##    author         filename
## 3   dispt dispt_fed_51.txt
## 10  dispt dispt_fed_62.txt
## 11  dispt dispt_fed_63.txt
```

conclusion

using using clustering algorithms k-Means, EM, and HAC techniques the authors of the federalist papers are no longer a mystery. I was able to generate multiple images that give clarity as to who wrote the disputed essays.