

Time Series Clustering

Henrique Lispector *lispe001@umn.edu*

```
library(Quandl)

## Loading required package: xts
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

#https://github.com/aayushmit/Data-science-presentation/blob/master/Deep\_dive\_in\_hierarchical\_clustering

Quandl.api_key("X8fn4LKgpiW3kRNoLAQn")

library(ggplot2)
library(gridExtra)
library(ggdendro)
library(zoo)
library(TSclust)

## Loading required package: wmtsa
## Warning: package 'wmtsa' was built under R version 3.4.3
## Loading required package: pdc
## Loading required package: cluster

pg <- Quandl('EOD/PG', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

## Warning in strptime(xx, f <- "%Y-%m-%d", tz = "GMT"): unknown timezone
## 'zone/tz/2017c.1.0/zoneinfo/America/Sao_Paulo'

apple <- Quandl('EOD/AAPL', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

visa <- Quandl('EOD/V', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

uhg <- Quandl('EOD/UNH', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

cocacola <- Quandl('EOD/KO', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

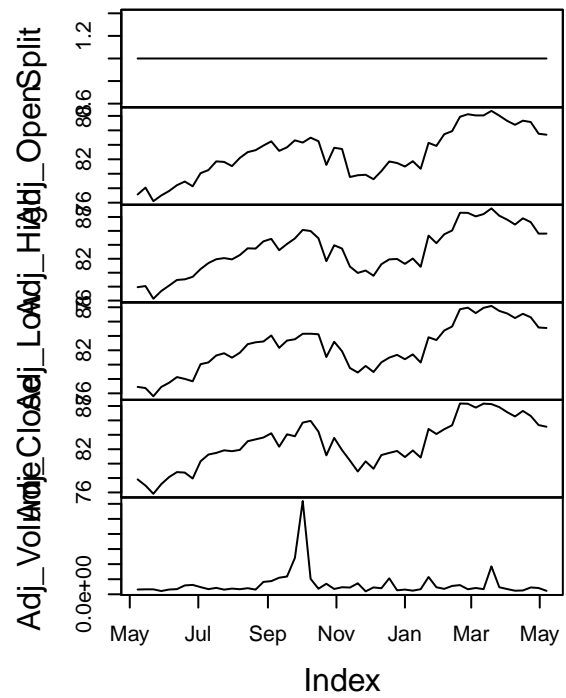
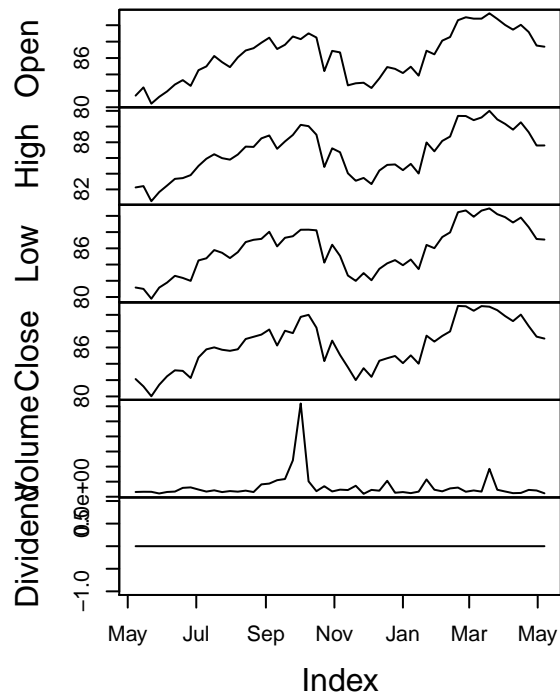
goldmansachs <- Quandl('EOD/GS', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

walmart <- Quandl('EOD/WMT', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

merk <- Quandl('EOD/MRK', start_date="2016-05-01", end_date='2017-05-01', collapse = 'weekly', type='zoo')

plot(pg)
```

pg

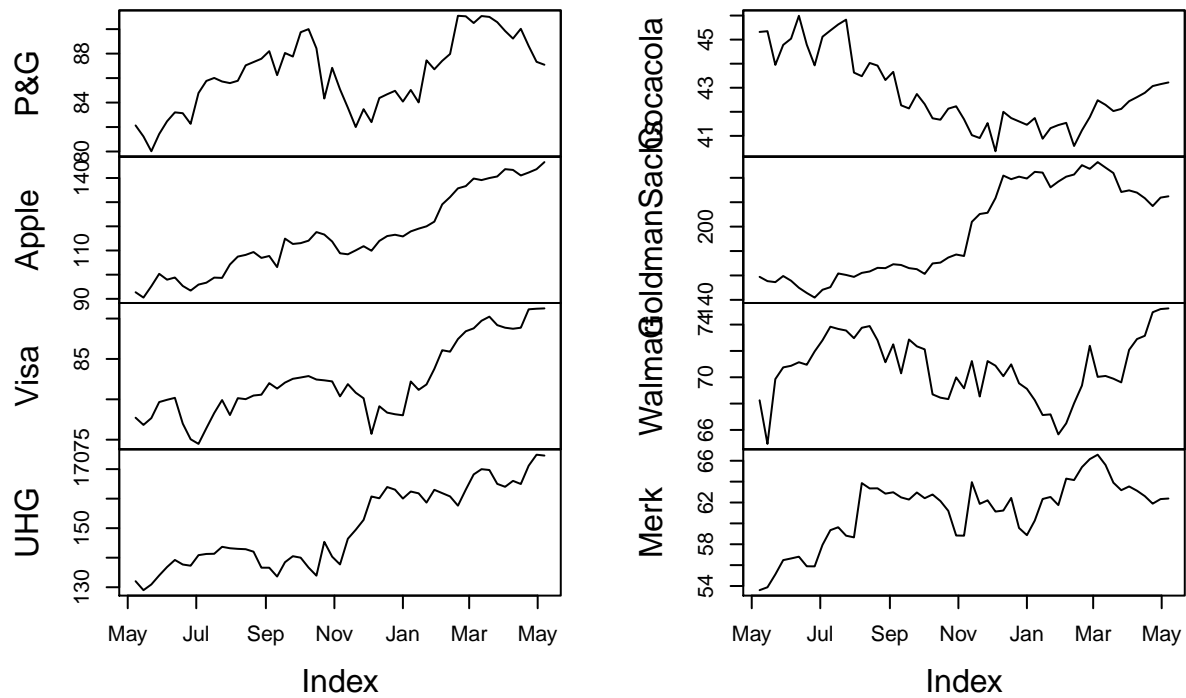


```
joined_ts <- cbind(pg[,4], apple[,4], visa[,4], uhg[,4], cocacola[,4], goldmansachs[,4],
                  walmart[,4], merk[,4])

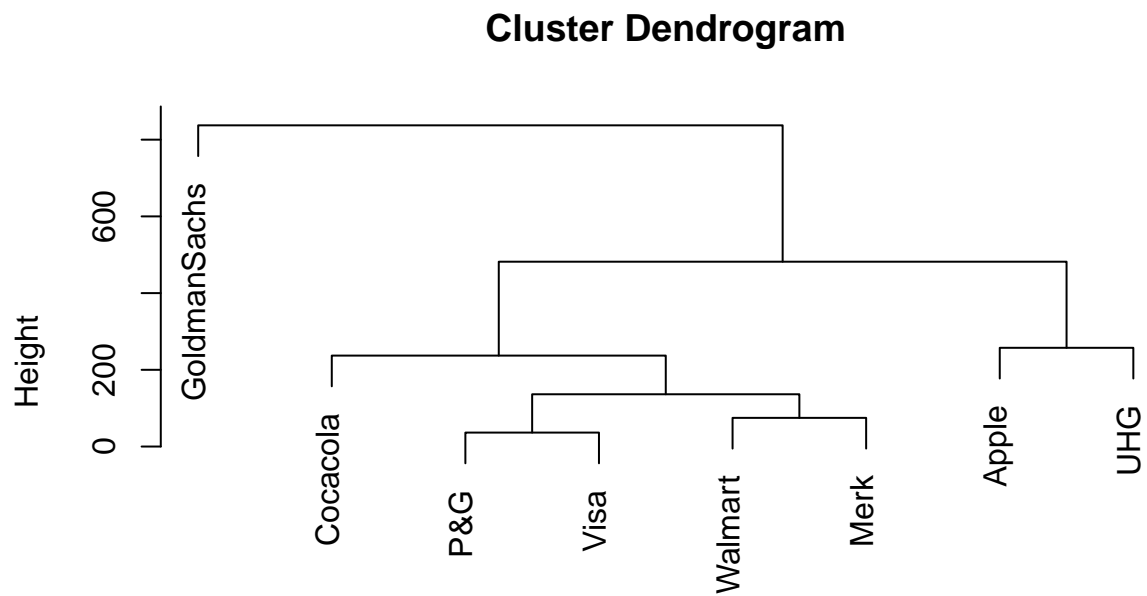
names(joined_ts) <- c('P&G', 'Apple', 'Visa', 'UHG', 'Cocacola', 'GoldmanSachs', 'Walmart',
                     'Merk')

plot(joined_ts)
```

joined_ts

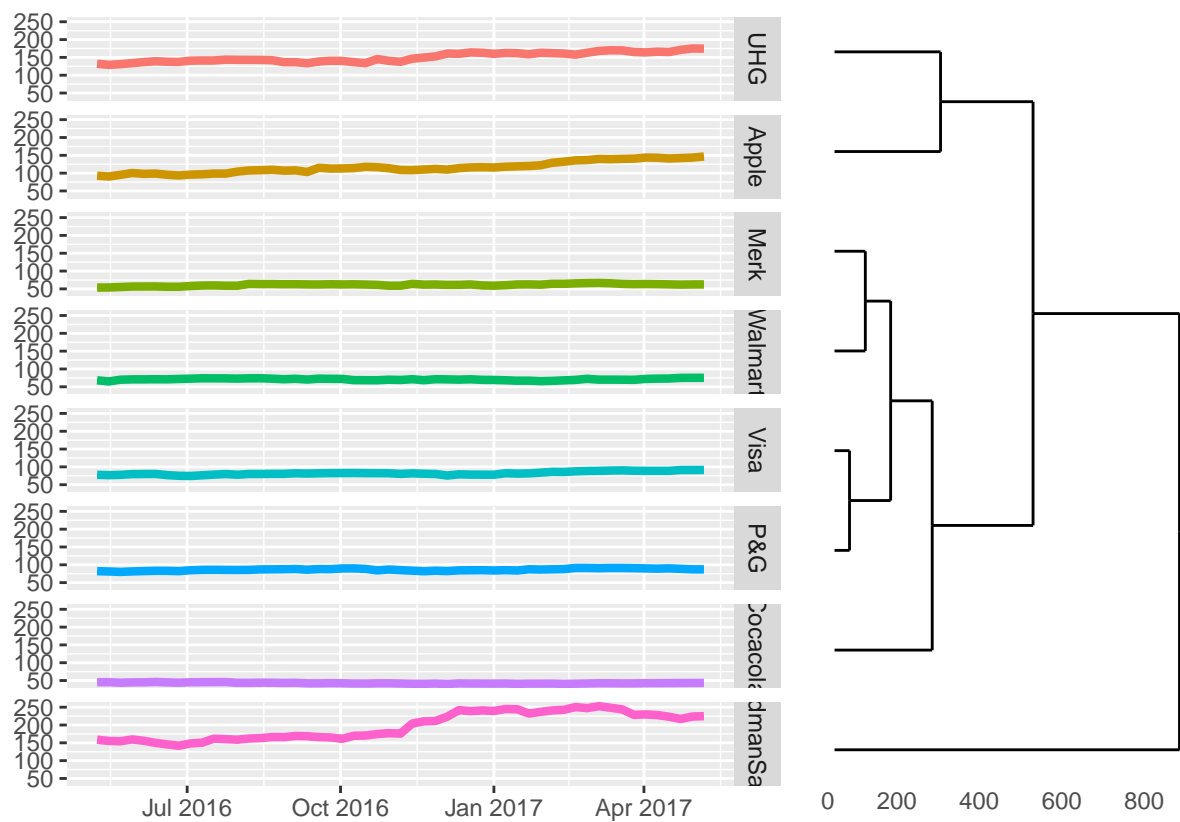


```
hc <- hclust(dist(t(joined_ts)), "ave")
plot(hc)
colours_hc <- cutree(hc, h=2)
hcdata <- dendro_data(hc)
names_order <- hcdata$labels$label
hcdata$labels$label <- ''
p1 <- ggdendrogram(hcdata, rotate = T, leaf_labels = F)
new_data <- joined_ts[,rev(as.character(names_order))]
p2 <- autoplot(new_data, facets = Series ~ . ) + aes(colour = as.character(rep(colours_hc, each=53)), l
gp1 <- ggplotGrob(p1)
```



```
dist(t(joined_ts))  
hclust (*, "average")
```

```
gp2 <- ggplotGrob(p2)  
grid.arrange(gp2, gp1, ncol = 2, widths = c(4,2))
```



```

maxs <- apply(joined_ts, 2, max)

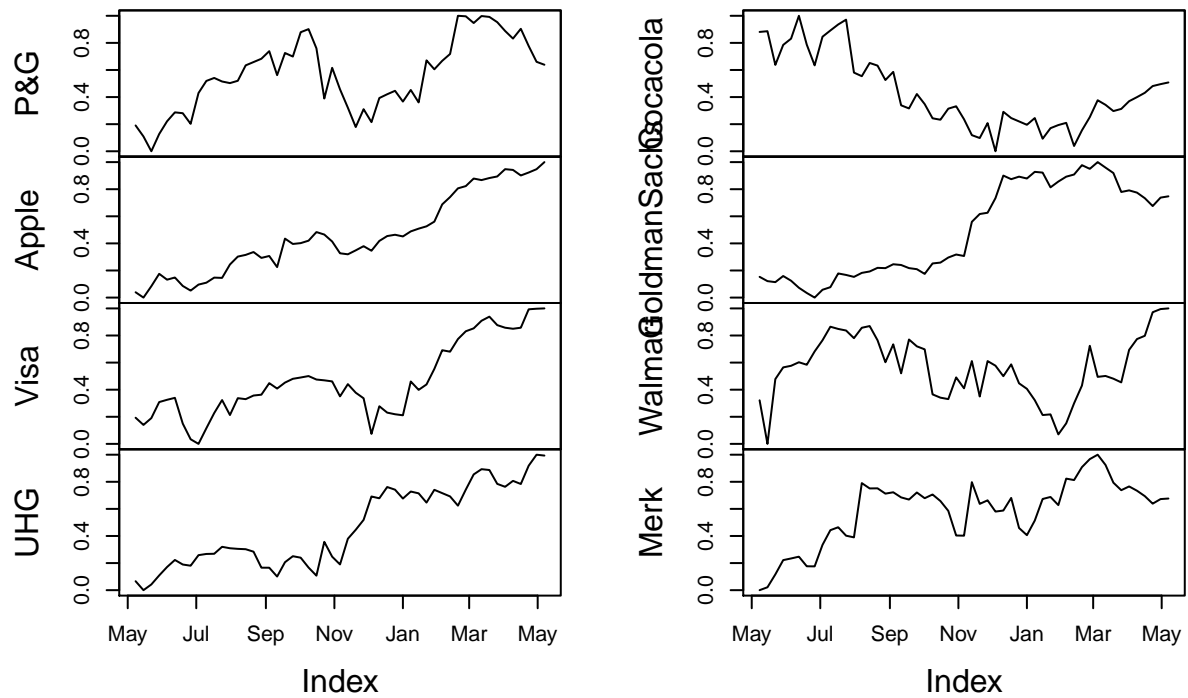
mins <- apply(joined_ts, 2, min)

joined_ts_scales <- scale(joined_ts, center = mins, scale = maxs - mins)

plot(joined_ts_scales)

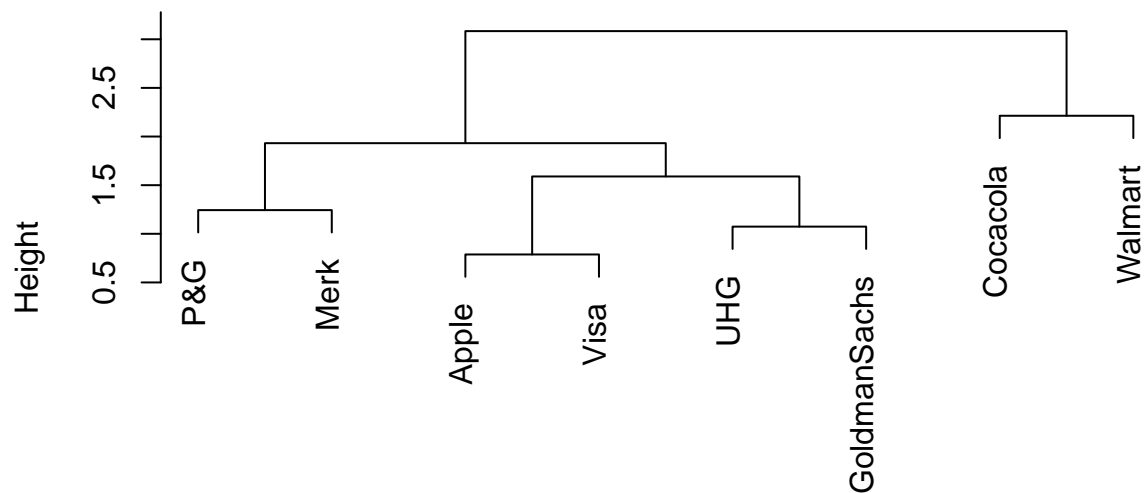
```

joined_ts_scales



```
hc <- hclust(dist(t(joined_ts_scales)), "ave")
plot(hc)
```

Cluster Dendrogram



```
dist(t(joined_ts_scales))
hclust (*, "average")
```

```

colours_hc <- cutree(hc, h=2)

hcdata <- dendro_data(hc)

names_order <- hcdata$labels$label

hcdata$labels$label <- ''

p1 <- gg dendrogram(hcdata, rotate = T, leaf_labels = FALSE)

new_data <- joined_ts_scales[, rev(as.character(names_order))]

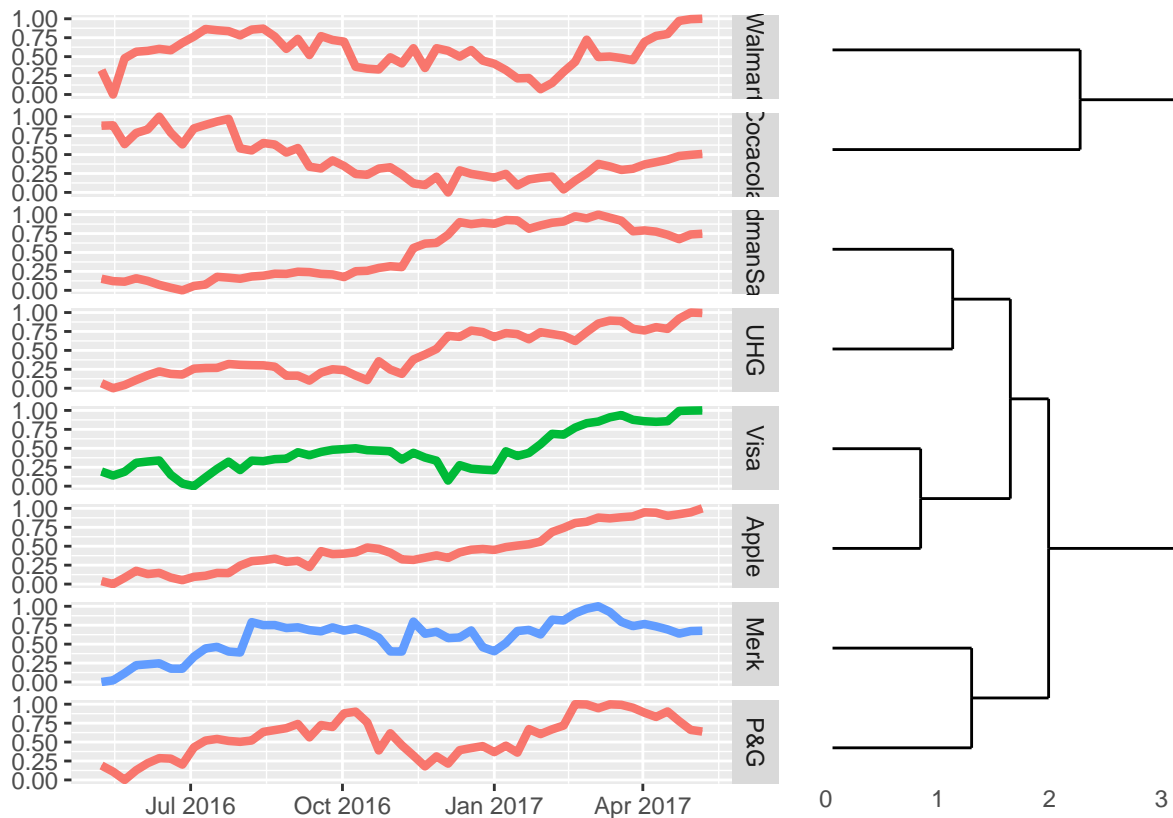
p2 <- autoplot(new_data, facets = Series ~ .) + aes(colour=as.character(rep(colours_hc, each = 53)), lin

gp1 <- ggplotGrob(p1)

gp2 <- ggplotGrob(p2)

grid.arrange(gp2, gp1, ncol=2, widths=c(4,2))

```



```

data <- data.frame(joined_ts)

data_modified <- data

rownames(data_modified) = 1:nrow(data_modified)

data_modified <- (data_modified[2:53,] - data_modified[1:52,])*100/(data_modified[1:52,])

```

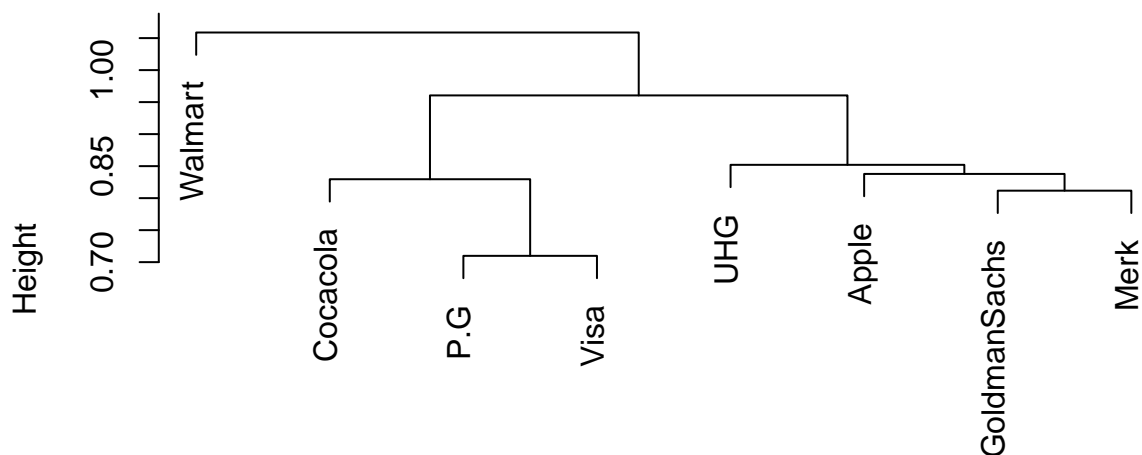
```
head(data_modified)
```

```
##          P.G          Apple          Visa          UHG          Cocacola GoldmanSachs
## 2 -1.09582369 -2.3727351 -1.1451364 -2.302333  0.06619594  -2.2096317
## 3 -1.48959744  5.1922227  1.0933229  1.503876 -3.08710033  -0.5343118
## 4  1.76205949  5.3875236  2.5621218  2.336948  1.88850967   3.2489806
## 5  1.27717058 -2.4215247  0.3514938  2.119403  0.58061635  -2.4196076
## 6  0.88517036  0.9293301  0.3002252  1.753873  2.10923623  -3.7129826
## 7 -0.08413462 -3.5414348 -3.9785483 -1.113186 -2.60926288  -2.8354126
##          Walmart          Merk
## 2 -4.8498168  0.5223881
## 3  7.5762242  2.2828508
## 4  1.2739765  2.4859372
## 5  0.1696113  0.2832861
## 6  0.3809793  0.3001412
## 7 -0.2670790 -1.6194332
```

```
hc <- hclust(diss(t(data_modified), "ACF"), "ave")
```

```
plot(hc)
```

Cluster Dendrogram



```
diss(t(data_modified), "ACF")
hclust (*, "average")
```

colours_hc

```
<- cutree(hc, h=2)
```

```
rownames(data_modified) <- rownames(data)[1:52]
```

```
data_modified <- as.matrix(data_modified)
```

```
class(data_modified[1:53,1])
```

```
data_modified <- xts(data_modified, as.POSIXct(rownames(data_modified)))
```