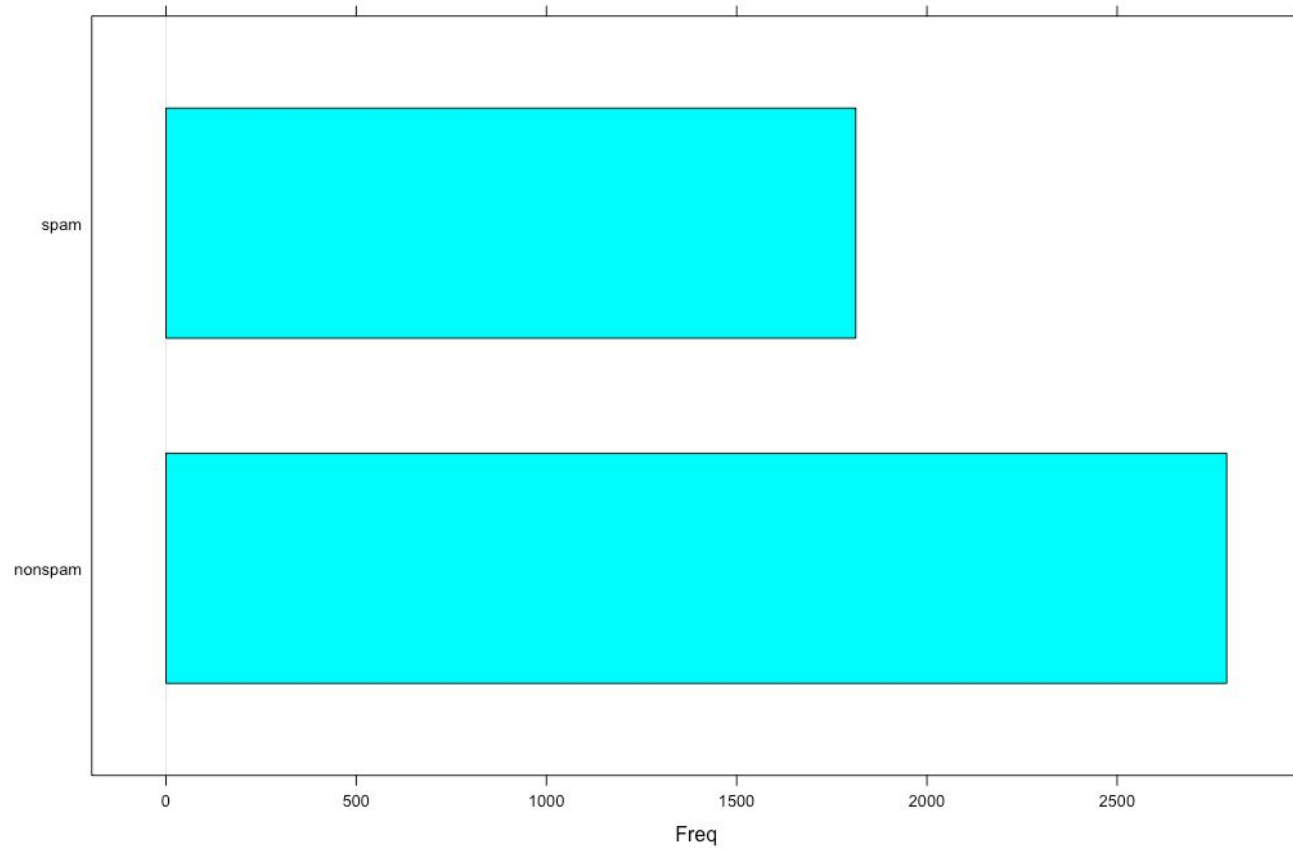
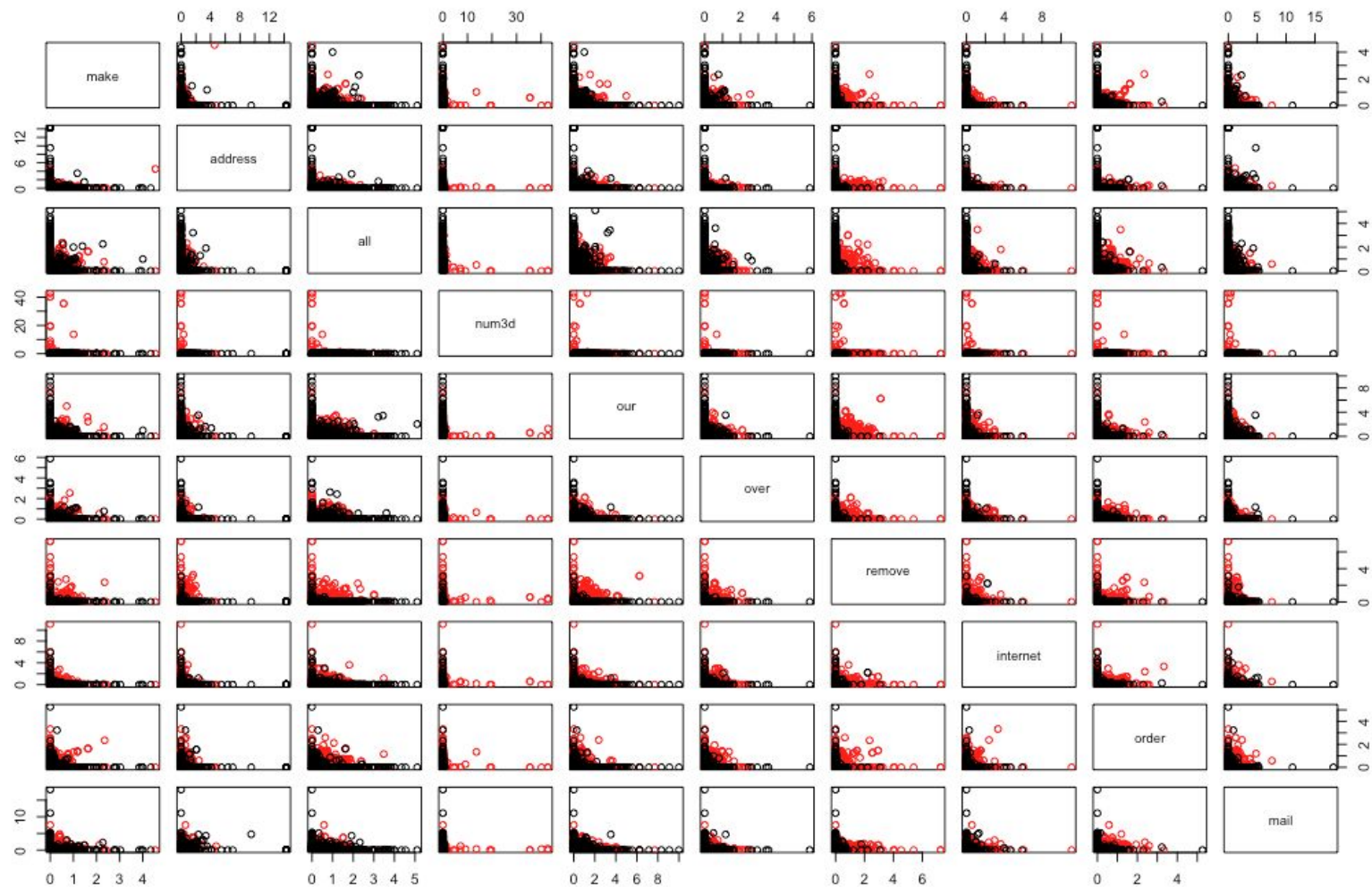


Spam Analysis

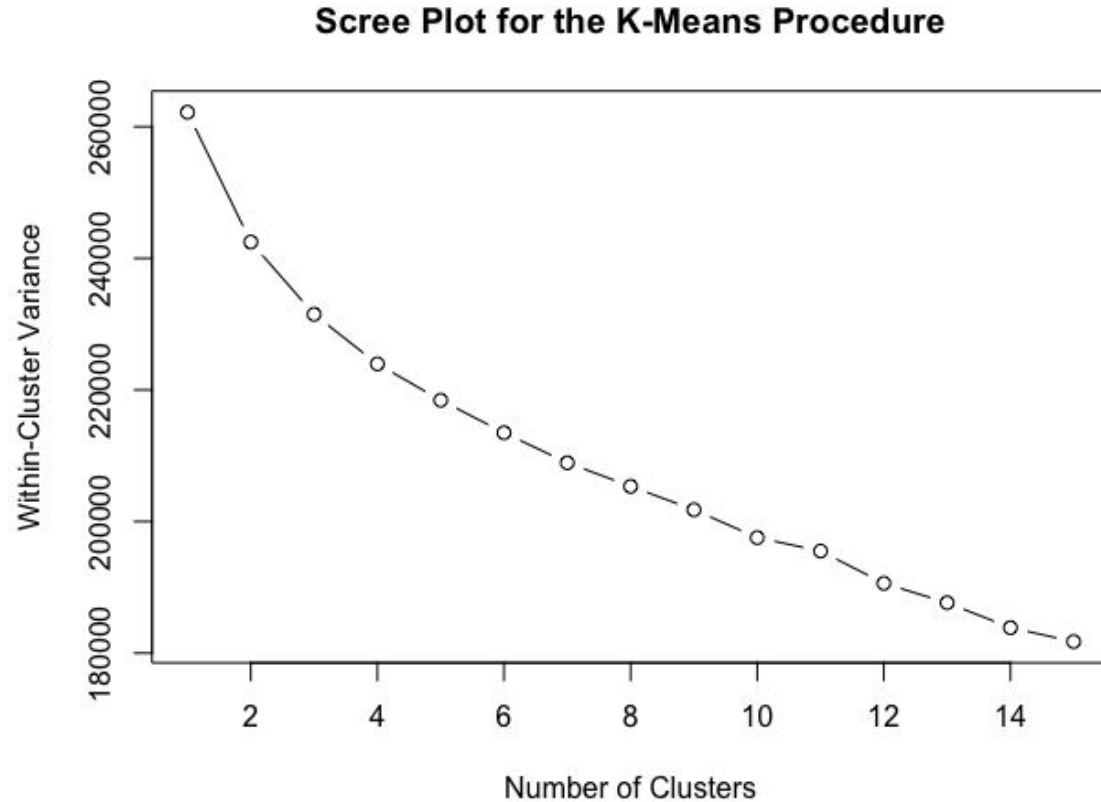
Group 5: Clustering

EDA



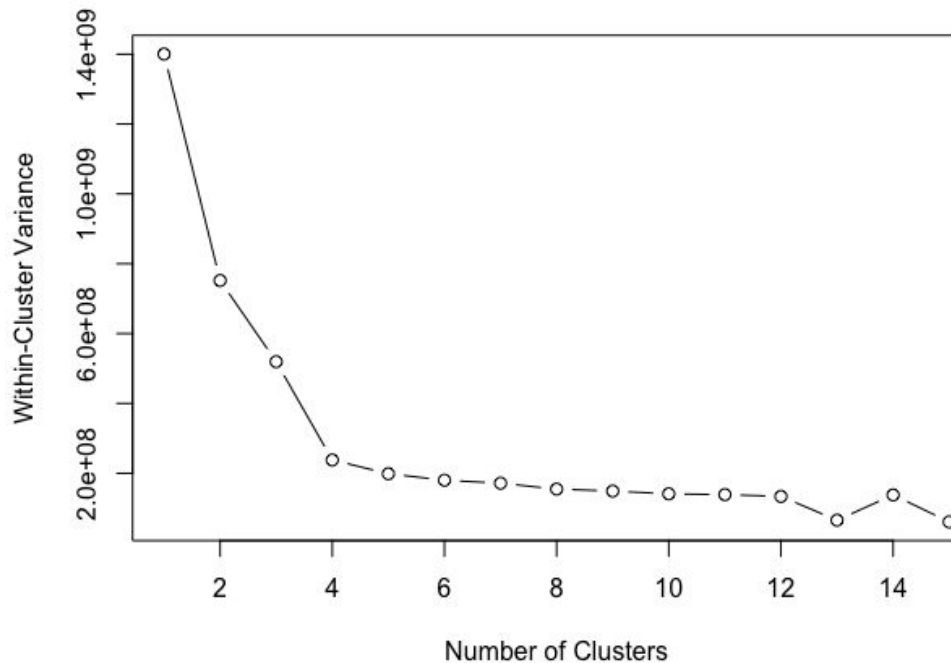


K Means Procedure

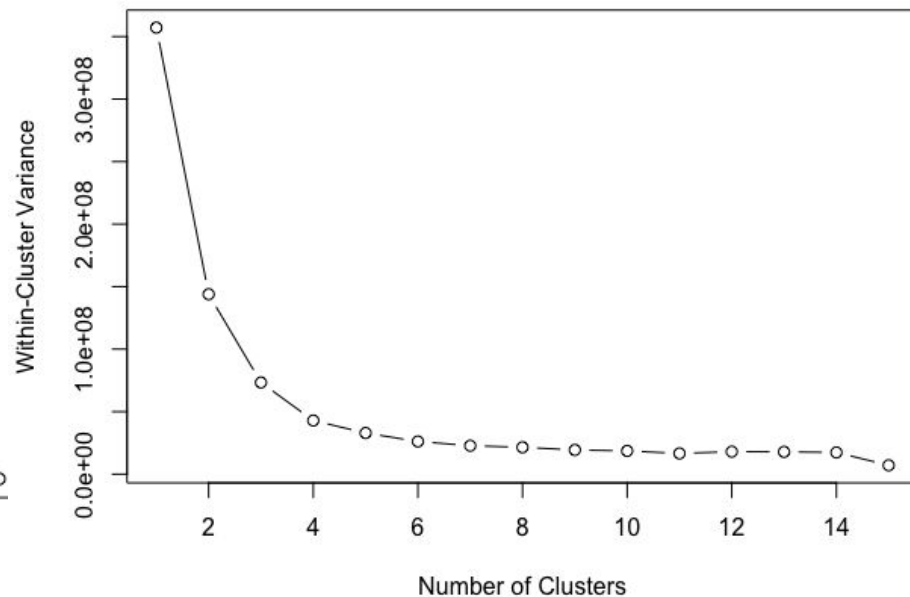


K-Means Procedure

Scree Plot for the K-Means Procedure: SPAM

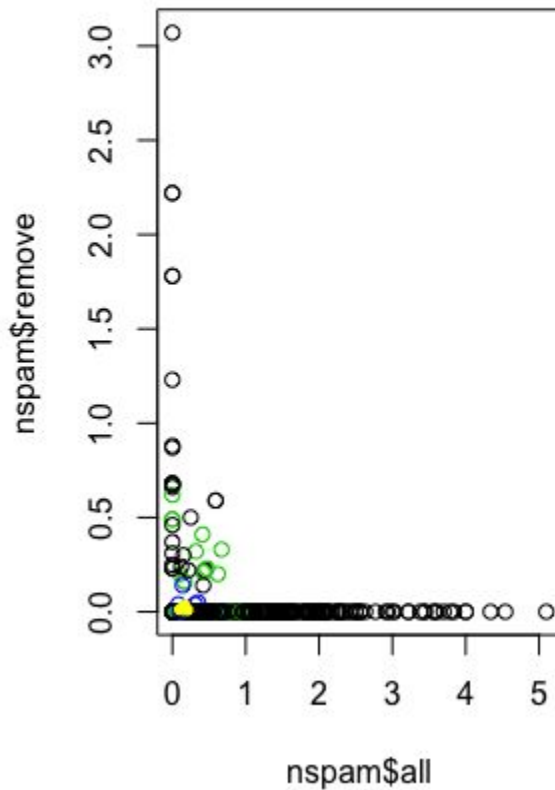
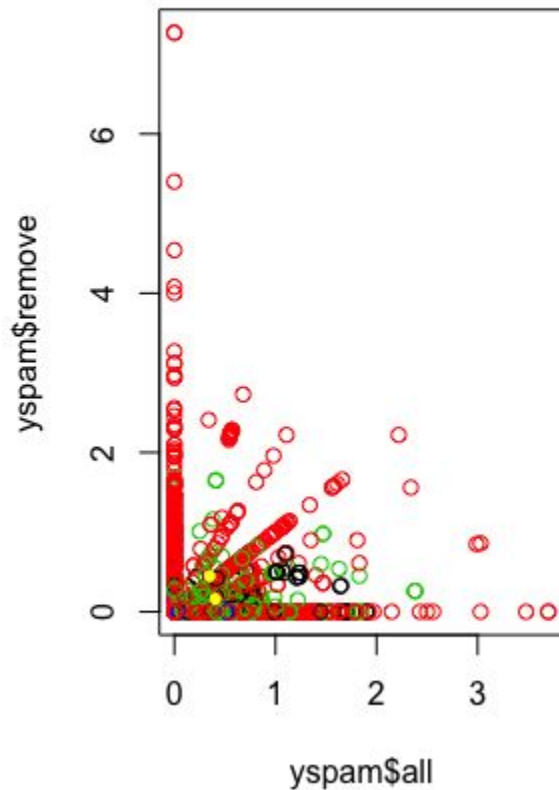


Scree Plot for the K-Means Procedure: NONSPAM



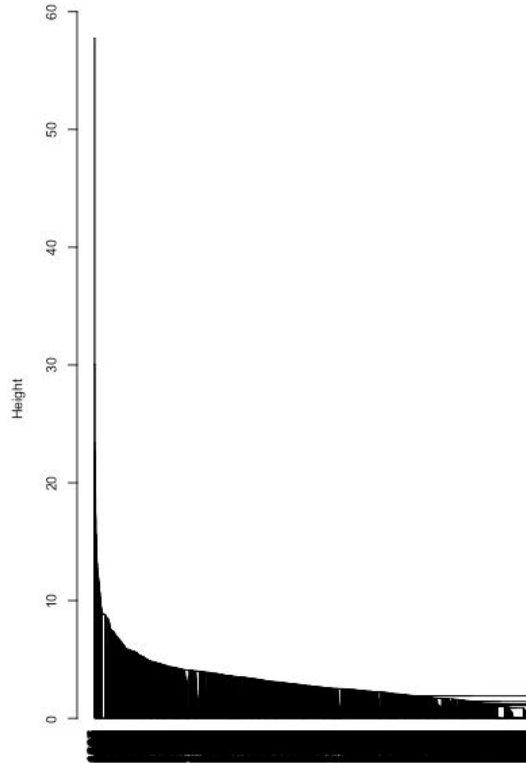
Cluster centers

SPAM: Single K-Means Attempt #NONSPAM: Single K-Means Attempt
WCV: 460517165.7302 WCV: 42886380.9881



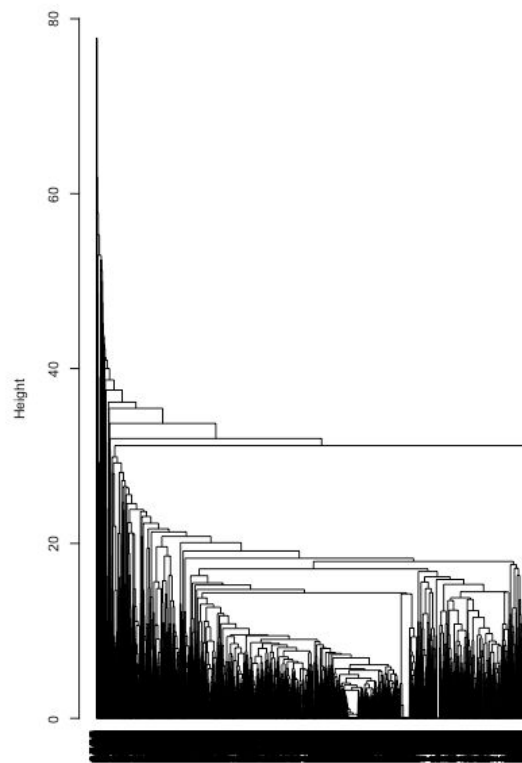
Dendograms- All data

Dendrogram of Single Linkage



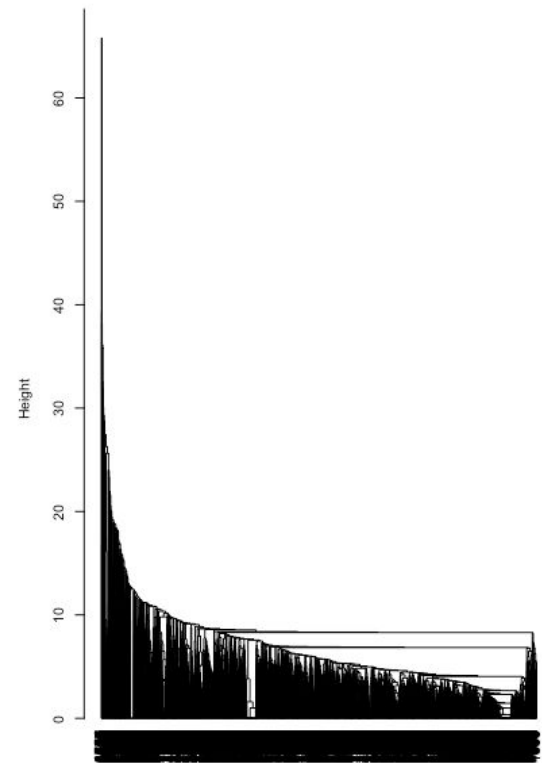
d
hclust("single")

Dendrogram of Complete Linkage



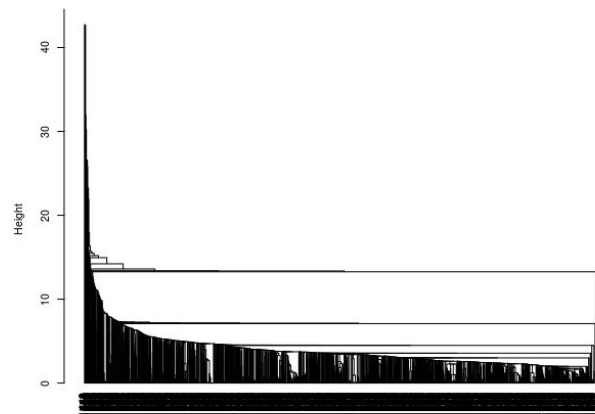
d
hclust("complete")

Dendrogram of Average Linkage

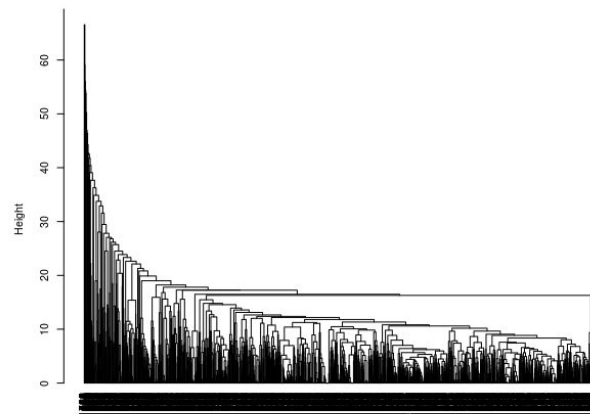


d
hclust("average")

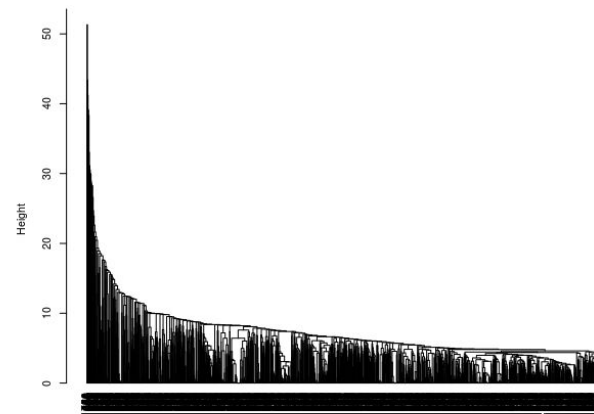
Spam Dendrogram of Single Linkage



Spam Dendrogram of Complete Linkage

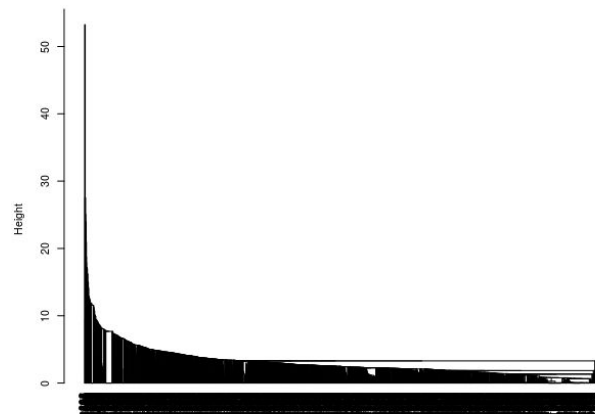


Spam Dendrogram of Average Linkage



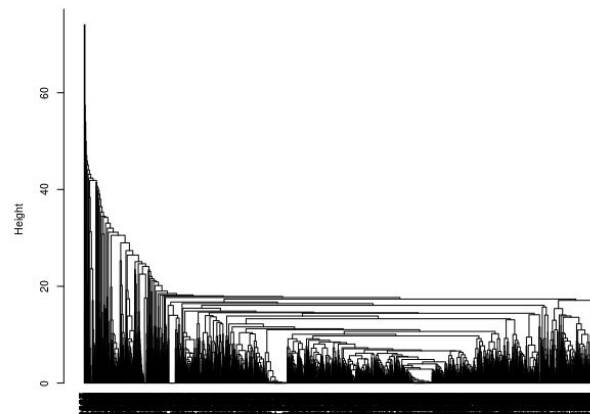
yd
holust ("single")

No Spam Dendrogram of Single Linkage



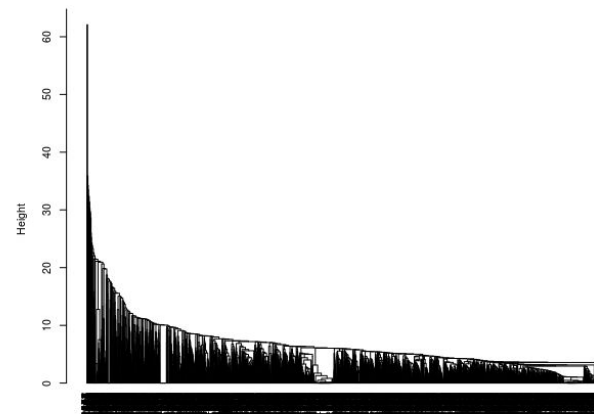
yd
holust ("complete")

No Spam Dendrogram of Complete Linkage



yd
holust ("average")

No Spam Dendrogram of Average Linkage



nd
holust ("single")

nd
holust ("complete")

nd
holust ("average")


```
yspam <- spam %>% filter(type == "spam")
```

```
nspam <- spam %>% filter(type == "nospam")
```

```
yspam.num <- yspam[, !sapply(spam, is.factor)]
```

```
nspam.num <- nspam[, !sapply(spam, is.factor)]
```

```
# #We should scale the data.
```

```
yspam.scaled = as.data.frame(scale(yspam.num))
```

```
nspam.scaled = as.data.frame(scale(nspam.num))
```

```
# #We need to calculate the pairwise distances between observations.
```

```
yd = dist(yspam.scaled)
```

```
nd = dist(nspam.scaled)
```

```
# #Using the hclust() function, we define the linkage manner by which we will
```

```
# #cluster our data.
```

```
yfit.single = hclust(yd, method = "single")
```

```
yfit.complete = hclust(yd, method = "complete")
```

```
#Splitting spam and nonspam
```

```
sspam = as.data.frame(scale(spam[, -58]))
```

```
sspam$type <- spam.raw$type
```

```
library(dplyr)
```

```
yspam <- spam.raw %>% filter(type == "spam")
```

```
nspam <- spam.raw %>% filter(type == "nonspam")
```

```
#Plotting scree plots for each group
```

```
wssplot(yspam[, -58])
```

```
wssplot(nspam[, -58])
```