

# Data Mining HW9

---

Anil Varma B

November 28, 2019

## 1. R code: Use attached dataset and hierarchical clustering algorithm (single and complete) to find 5 or 6 clusters

```
library(cluster)
library(dendextend)
#Reading the file
a2<-read.csv(file.choose(),head = TRUE)
#Converting to a dataframe
a2_df<- as.data.frame(a2)
head(a2_df)
```

```
> a2_df<- as.data.frame(a2)
> head(a2_df)
  Company Fixed_charge RoR Cost Load_factor Demand_growth Sales Nuclear Fuel_Cost
1  Arizona      1.06   9.2  151      54.4          1.6  9077      0.0    0.628
2  Boston      0.89  10.3  202      57.9          2.2  5088     25.3    1.555
3  Central      1.43  15.4  113      53.0          3.4  9212      0.0    1.058
4 Commonwealth  1.02  11.2  168      56.0          0.3  6423     34.3    0.700
5      NY      1.49   8.8  192      51.2          1.0  3300     15.6    2.044
6  Florida      1.32  13.5  111      60.0         -2.2 11127     22.5    1.241
```

Figure 0.1: Dimensions of the data set

```
#Splitting the categorical column and numerical column
a2_col1<-a2_df[1]
head(a2_col1)
a2_df_num<-a2_df[2:9]
head(a2_df_num)
```

```

> a2_col1<-a2_df[1]
> a2_df_num<-a2_df[2:9]
> head(a2_col1)
  Company
1  Arizona
2  Boston
3  Central
4 Commonwealth
5      NY
6  Florida
> head(a2_df_num)
  Fixed_charge  RoR Cost Load_factor Demand_growth Sales Nuclear Fuel_Cost
1      1.06   9.2  151      54.4         1.6   9077      0.0    0.628
2      0.89  10.3  202      57.9         2.2   5088     25.3    1.555
3      1.43  15.4  113      53.0         3.4   9212      0.0    1.058
4      1.02  11.2  168      56.0         0.3   6423     34.3    0.700
5      1.49   8.8  192      51.2         1.0   3300     15.6    2.044
6      1.32  13.5  111      60.0        -2.2  11127     22.5    1.241

```

Figure 0.2: Splitting the columns

We are using the 'hclust' function to do the hierarchical clustering, we require the distance values which can be computed in R by using the 'dist' function. Default measure for dist function is 'Euclidean'

## 1 SINGLE LINKAGE

```

hc_s <- hclust(dist(a2_df_num), method = "single")
plot(hc_s, labels = a2_col1[hc_s$order,1])

```

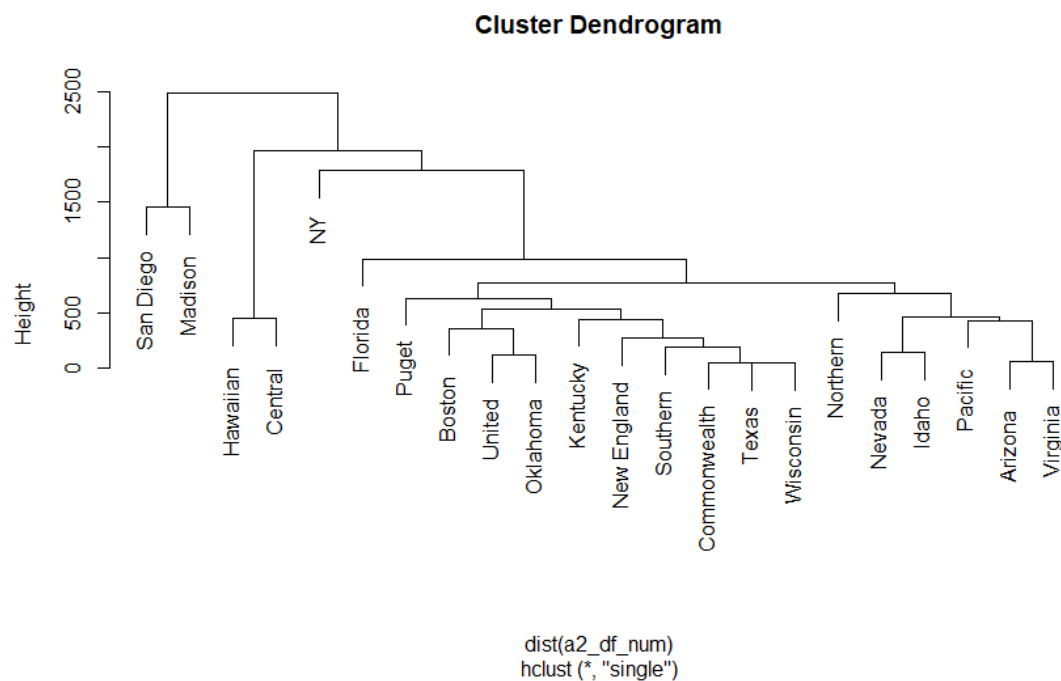


Figure 1.1: Single linkage with hclust

Finding 5 clusters from the single linkage dendrogram

```
rect.hclust(hc_s, k = 5, border = "red")
```

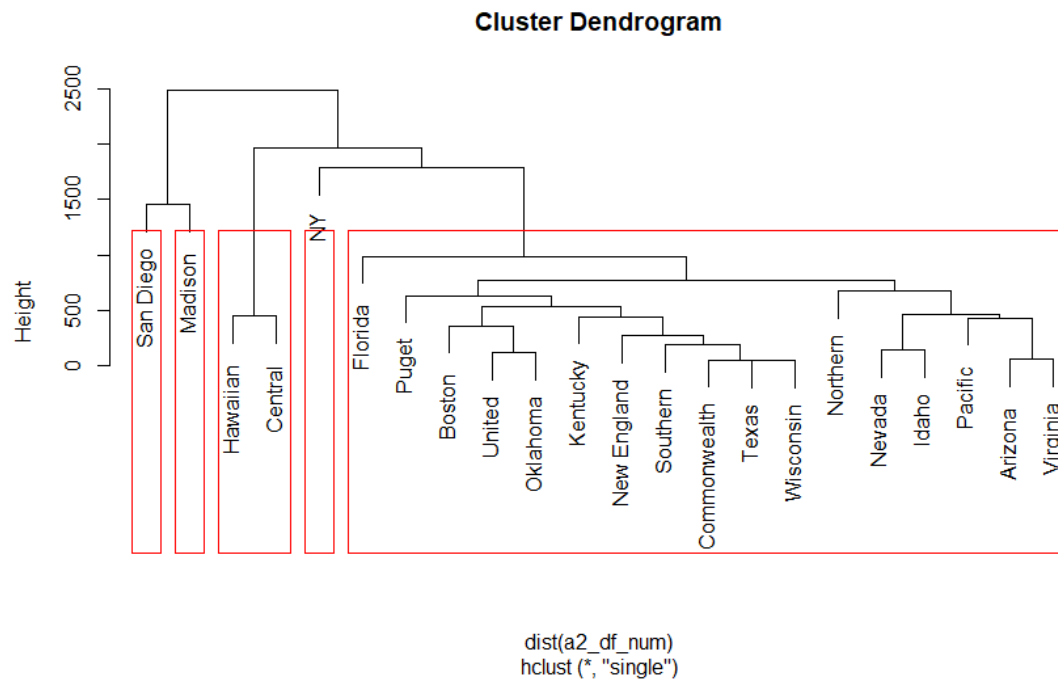


Figure 1.2: Highlighted clusters with red boxes

```
#selective coloring of branches AND labels
dend_s <- as.dendrogram(hc_s)
dend_s= color_labels(dend_s,k=5)
dend_s= color_branches(dend_s,k=5,groupLabels=TRUE)
plot(dend_s)
```

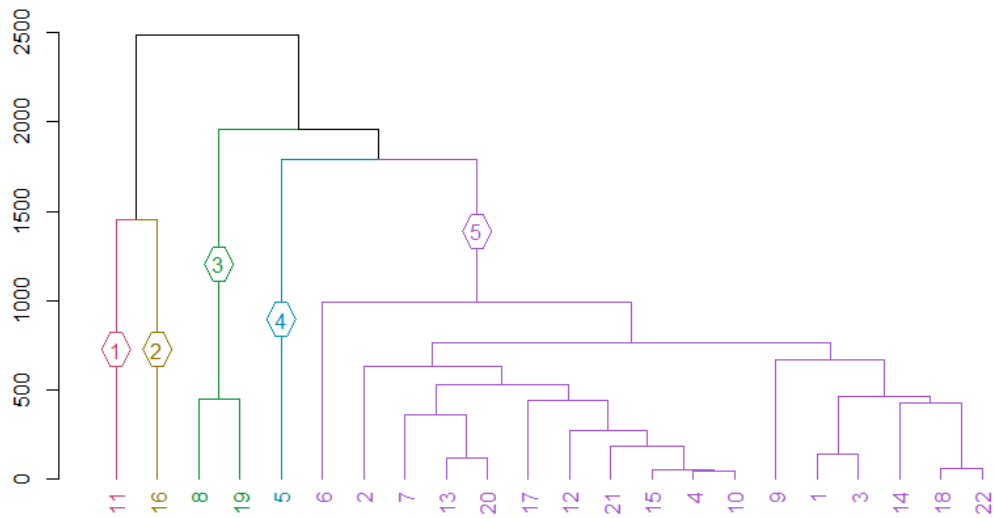


Figure 1.3: selective coloring of branches for single linkage clusters

## 2 COMPLETE LINKAGE

```
hc_c <- hclust(dist(a2_df_num), method = "complete")
plot(hc_c, labels = a2_col1[hc_c$order,1])
```

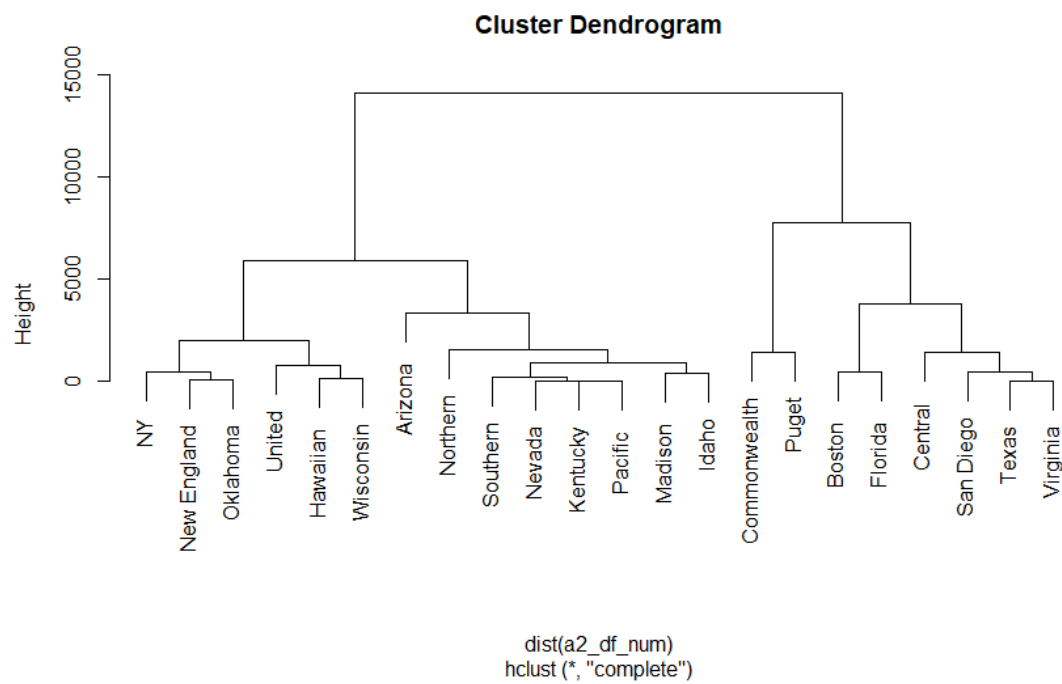


Figure 2.1: Complete linkage with hclust

Finding 5 clusters from the Complete linkage dendrogram

```
rect.hclust(hc_c, k = 5, border = "red")
```

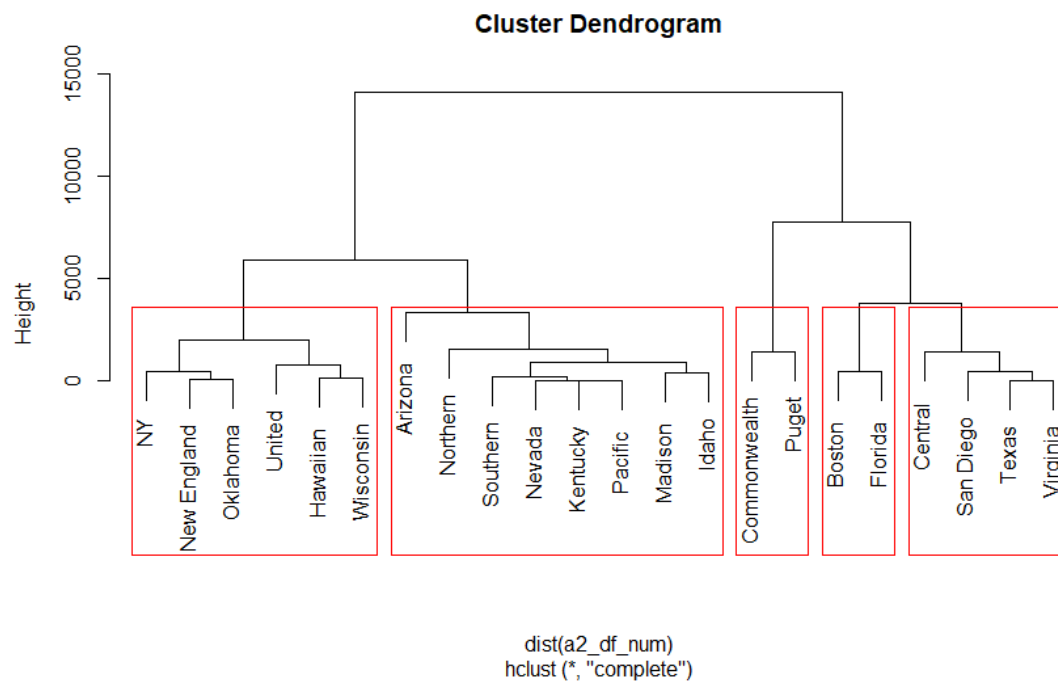


Figure 2.2: Highlighted clusters with red boxes

```
#selective coloring of branches AND labels
dend <- as.dendrogram(hc_c)
dend=color_labels(dend,k=5)
dend=color_branches(dend,k=5,groupLabels=TRUE)
plot(dend)
```

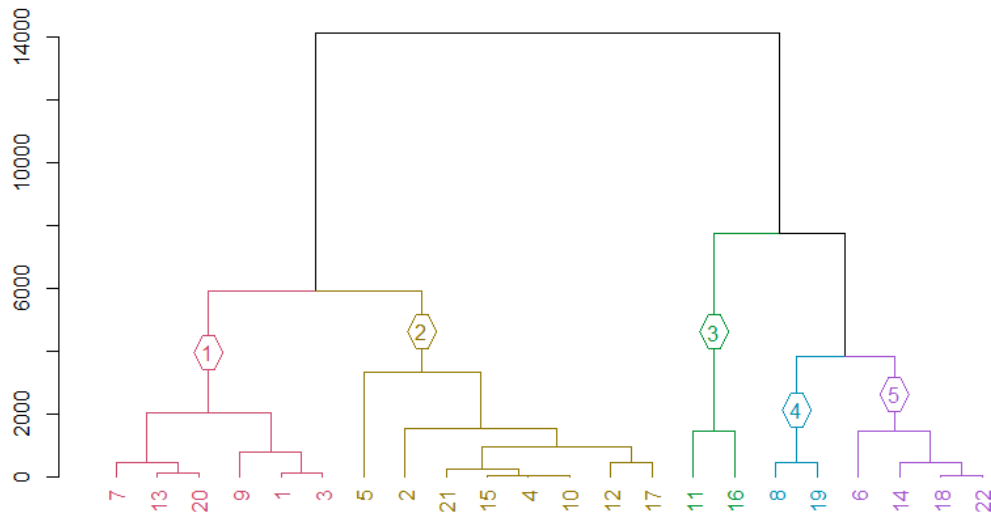


Figure 2.3: selective coloring of branches for Complete linkage clusters

Another additional method is to use the AGNES (acronym for Agglomerative Nesting) function. In this method, each observation is assigned to its own cluster. Then, the similarity (or distance) between each of the clusters is computed and the two most similar clusters are merged into one. Finally, steps 2 and 3 are repeated until there is only one cluster left

### 3 SINGLE LINKAGE

```
a2_agnes_single<-agnes(a2_df_num,diss=FALSE, stand=FALSE, method="single")
plot(a2_agnes_single,labels = a2_col1[a2_agnes_single$order,1])
```

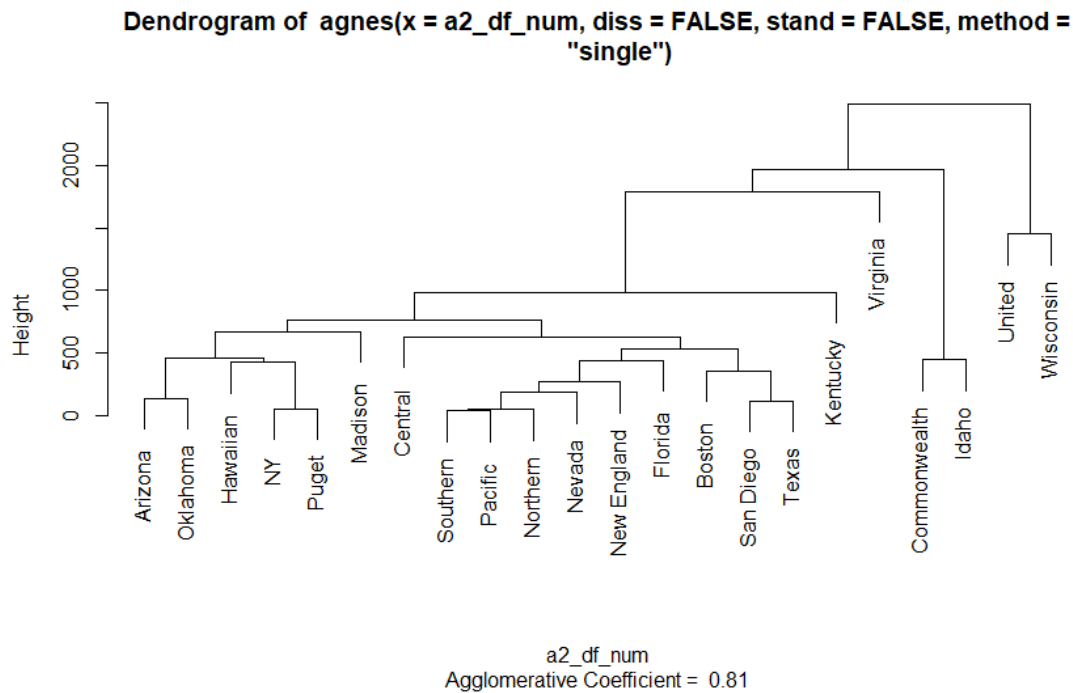


Figure 3.1: Single linkage with AGNES

#### 4 COMPLETE LINKAGE

```
a2_agnes_complete<-agnes(a2_df_num,diss=FALSE, stand=FALSE, method="complete")
plot(a2_agnes_complete,labels = a2_col1[a2_agnes_complete$order,1])
```

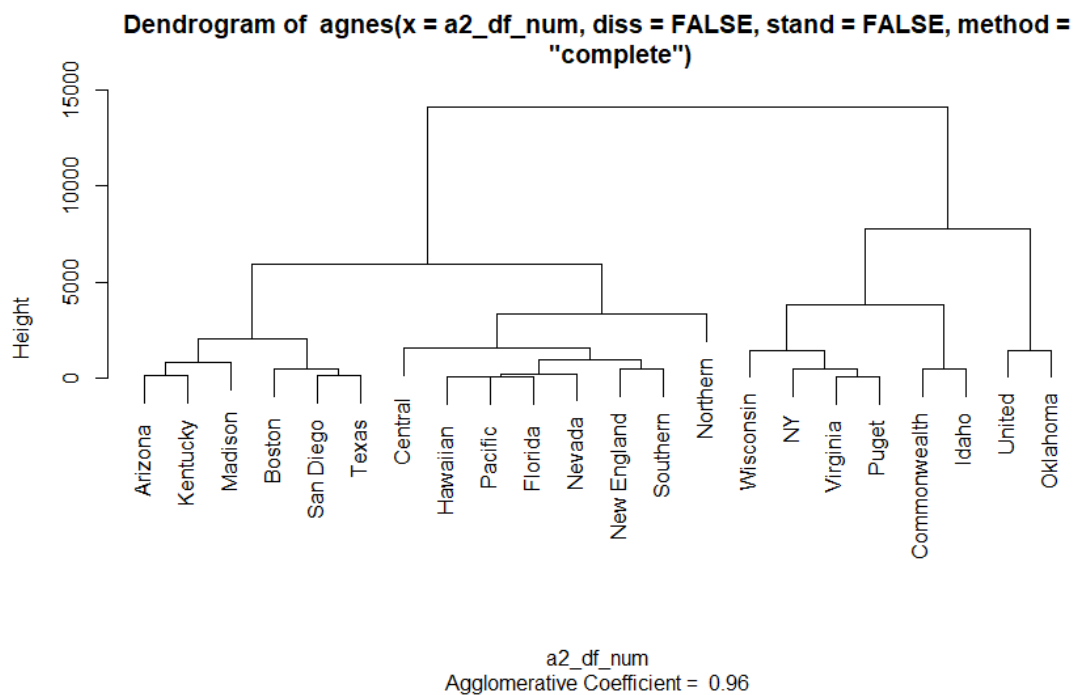


Figure 4.1: Complete linkage with AGNES