

MichaelBasta_Assignment_5

Michael Basta

2022-11-28

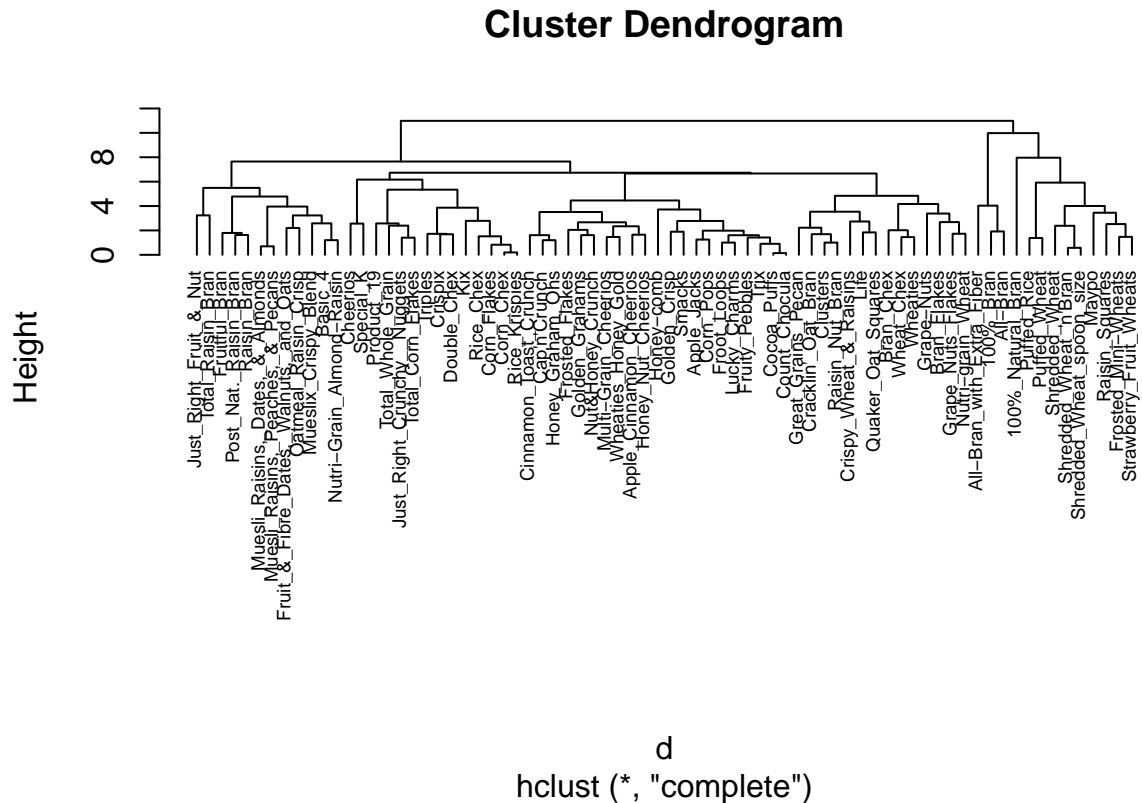
```
Cereals <- read.csv("/Users/michaelbasta/Documents/Fundamentals of Machine Learning /Module 8/Cereals.csv")
df <- Cereals
row.names(df) <- df$name
df <- na.omit(df)
# eliminating all the non
df <- df[,4:16]
df <- scale(df)
head(df)
```

```
##              calories    protein      fat      sodium
## 100%_Bran        -1.8659155  1.3817478  0.0000000 -0.3910227
## 100%_Natural_Bran  0.6537514  0.4522084  3.9728810 -1.7804186
## All-Bran        -1.8659155  1.3817478  0.0000000  1.1795987
## All-Bran_with_Extra_Fiber -2.8737823  1.3817478 -0.9932203 -0.2702057
## Apple_Cinnamon_Cheerios  0.1498180 -0.4773310  0.9932203  0.2130625
## Apple_Jacks      0.1498180 -0.4773310 -0.9932203 -0.4514312
##              fiber      carbo      sugars      potass
## 100%_Bran        3.22866747 -2.5001396 -0.2542051  2.5605229
## 100%_Natural_Bran -0.07249167 -1.7292632  0.2046041  0.5147738
## All-Bran        2.81602258 -1.9862220 -0.4836096  3.1248675
## All-Bran_with_Extra_Fiber 4.87924705 -1.7292632 -1.6306324  3.2659536
## Apple_Cinnamon_Cheerios -0.27881412 -1.0868662  0.6634132 -0.4022862
## Apple_Jacks      -0.48513656 -0.9583868  1.5810314 -0.9666308
##              vitamins      shelf      weight      cups
## 100%_Bran        -0.1818422  0.9419715 -0.2008324 -2.0856582
## 100%_Natural_Bran -1.3032024  0.9419715 -0.2008324  0.7567534
## All-Bran        -0.1818422  0.9419715 -0.2008324 -2.0856582
## All-Bran_with_Extra_Fiber -0.1818422  0.9419715 -0.2008324 -1.3644493
## Apple_Cinnamon_Cheerios -0.1818422 -1.4616799 -0.2008324 -0.3038480
## Apple_Jacks      -0.1818422 -0.2598542 -0.2008324  0.7567534
##              rating
## 100%_Bran        1.8549038
## 100%_Natural_Bran -0.5977113
## All-Bran        1.2151965
## All-Bran_with_Extra_Fiber 3.6578436
## Apple_Cinnamon_Cheerios -0.9165248
## Apple_Jacks      -0.6553998
```

```
d <- dist(df, method = "euclidean")
# Hierarchical clustering using Complete Linkage
```

```
hc1 <- hclust(d, method = "complete")
```

```
# Plot the obtained dendrogram
plot(hc1, cex = 0.6, hang = -1)
```



```
library(cluster)
```

```
# Compute with agnes and with different Linkage methods
```

```
hc_single <- agnes(df, method = "single")
```

```
hc_complete <- agnes(df, method = "complete")
```

```
hc_average <- agnes(df, method = "average")
```

```
# Compare Agglomerative coefficients
```

```
print(hc_single$ac)
```

```
## [1] 0.6067859
```

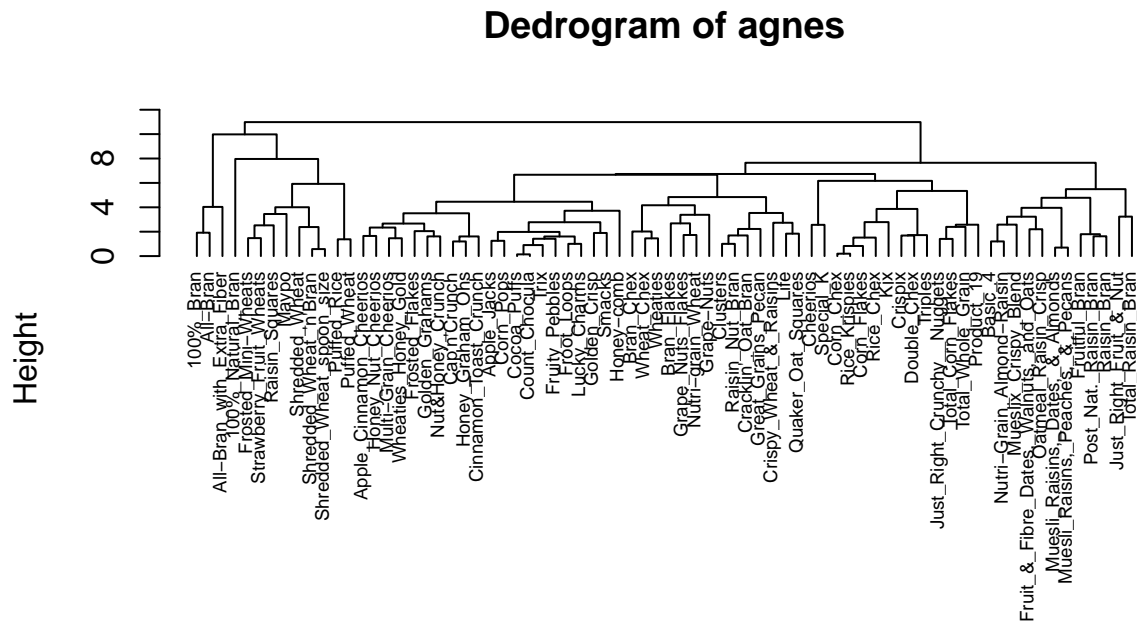
```
print(hc_complete$ac)
```

```
## [1] 0.8353712
```

```
print(hc_average$ac)
```

```
## [1] 0.7766075
```

```
pltree(hc_complete, cex = 0.6, hang = -1, main = "Dedrogram of agnes")
```



df
agnes (*, "complete")

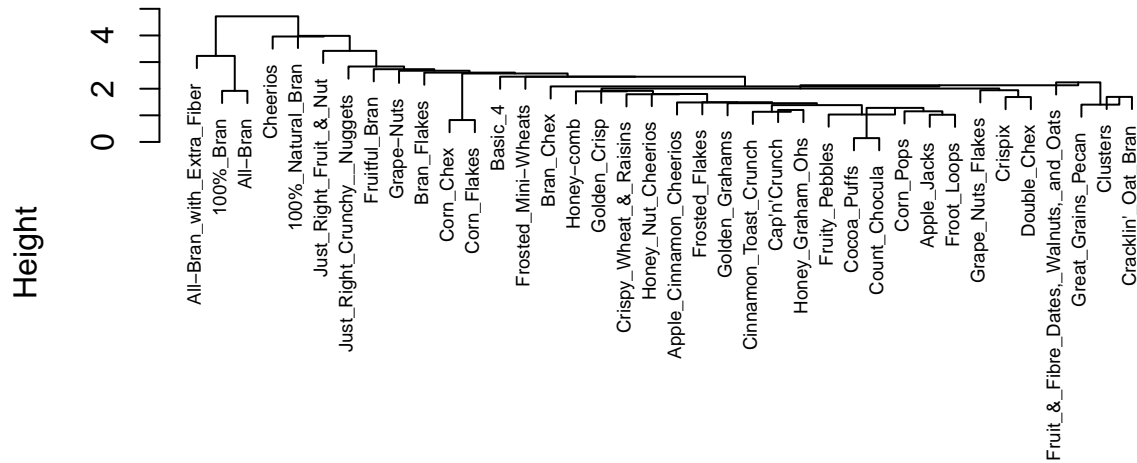
```
paste("Best Linkage method is the complete with the highest coefficient")
```

```
## [1] "Best Linkage method is the complete with the highest coefficient"
```

```
# Marking clusters to see the best number  
hc_complete <- hclust(d, method = "complete")
```

```
# plot dendrogram  
plot(hc_complete, cex = 0.6)  
rect.hclust(hc_complete, k = 7, border = 1:5)
```


Cluster Dendrogram



d
hclust (*, "centroid")

```
hc_single <- agnes(Train, method = "single")
hc_complete <- agnes(Train, method = "complete")
hc_average <- agnes(Train, method = "average")
```

```
# Compare Agglomerative coefficients
print(hc_single$ac)
```

```
## [1] 0.5853558
```

```
print(hc_complete$ac)
```

```
## [1] 0.7904243
```

```
print(hc_average$ac)
```

```
## [1] 0.7235901
```

```
## From what is seen on the partition dendrogram and
##the agglomerative coefficients we see the structure is less stable
```

```
# To be able to cluster "Healthy Cereals" We will
#consider only (Protien, Fiber, Vitamins, rating)
```

```
## Yes data should be normalized As we don't want the
##clustering algorithm to depend to an arbitrary variable unit
```

```
df <- Cereals
row.names(df) <- df$name
df <- na.omit(df)
# Only Selecting the columns contributing to healthy cereals
df <- df[,c(5,8,12,16)]
df <- scale(df)
head(df)
```

```
##               protein      fiber  vitamins      rating
## 100%_Bran      1.3817478  3.22866747 -0.1818422  1.8549038
## 100%_Natural_Bran 0.4522084 -0.07249167 -1.3032024 -0.5977113
## All-Bran      1.3817478  2.81602258 -0.1818422  1.2151965
## All-Bran_with_Extra_Fiber 1.3817478  4.87924705 -0.1818422  3.6578436
## Apple_Cinnamon_Cheerios -0.4773310 -0.27881412 -0.1818422 -0.9165248
## Apple_Jacks      -0.4773310 -0.48513656 -0.1818422 -0.6553998
```

```
d <- dist(df, method = "euclidean")
# Hierarchical clustering using Complete Linkage
hc1 <- hclust(d, method = "complete")

# Plot the obtained dendrogram
plot(hc1, cex = 0.6, hang = -1)
```

Cluster Dendrogram

