STAT 388 Homework 8

Charles Hwang

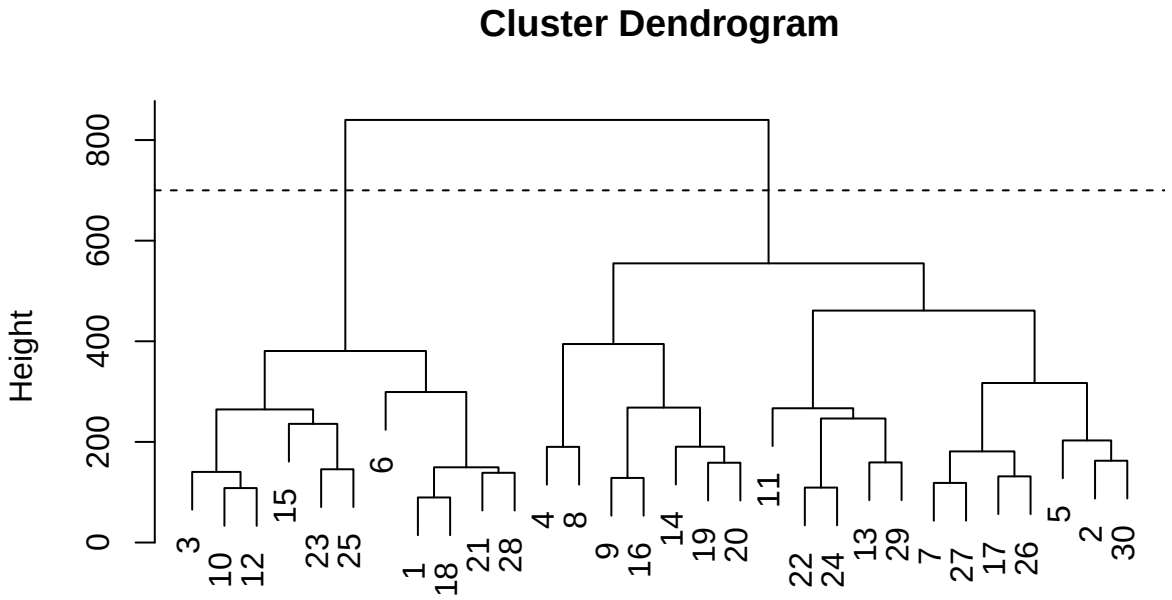
Professor Matthews

STAT 388-001

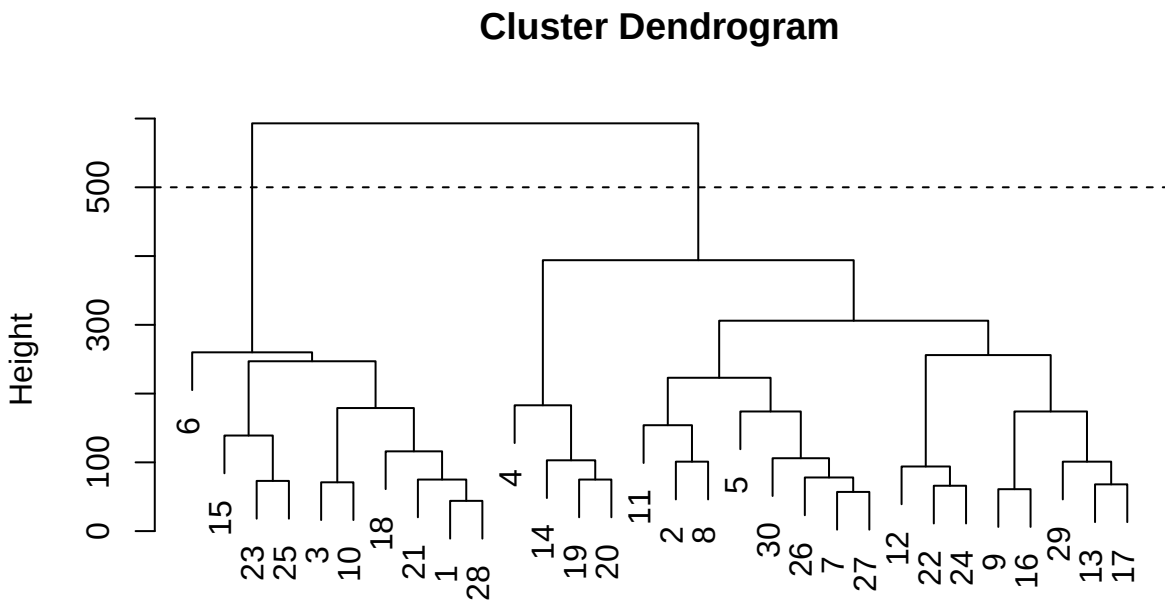
7 December 2019

Problem 1

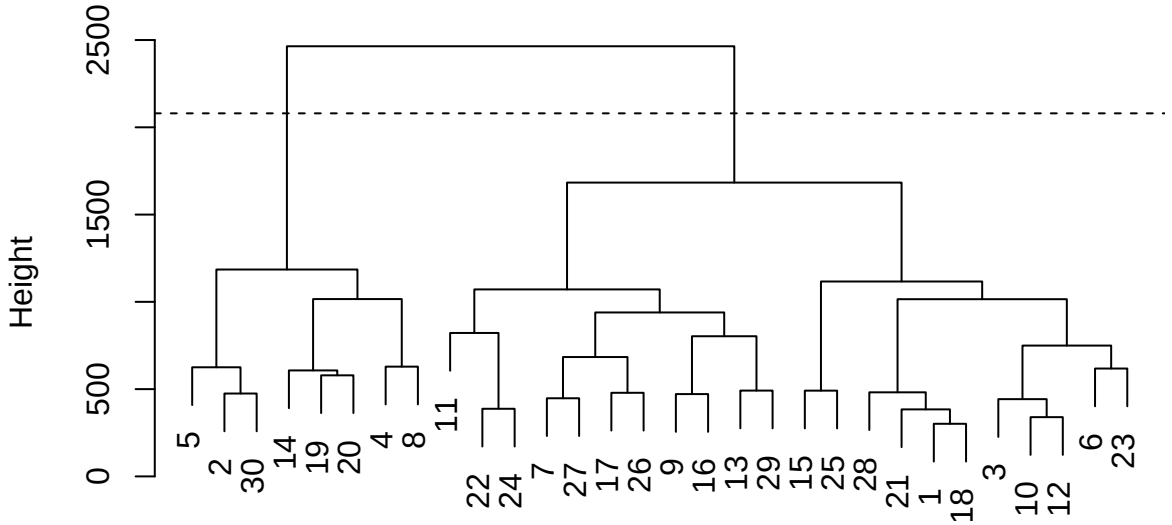
```
rm(list=ls())
library(BBmisc)
library(stats)
data <- read.csv(file="/Users/newuser/Desktop/Notes/Undergraduate/STAT 338 - Predictive Analytics/MLB20
names(data)[1] <- "Team"
names(data)[4] <- "RpG"
names(data)[17] <- "K"
names(data)[22] <- "OPS+"
data$Team[30] <- "WSH"
data$Team[32] <- "Total"
data$X.Bat[31] <- mean(data$X.Bat[1:30])
data$X.Bat[32] <- sum(data$X.Bat[1:30])
plot(hclust(dist(data[-(31:32),],method="euclidean")),xlab="",sub="") # Hierarchical clustering on orig
abline(700,0,lty=2)
```



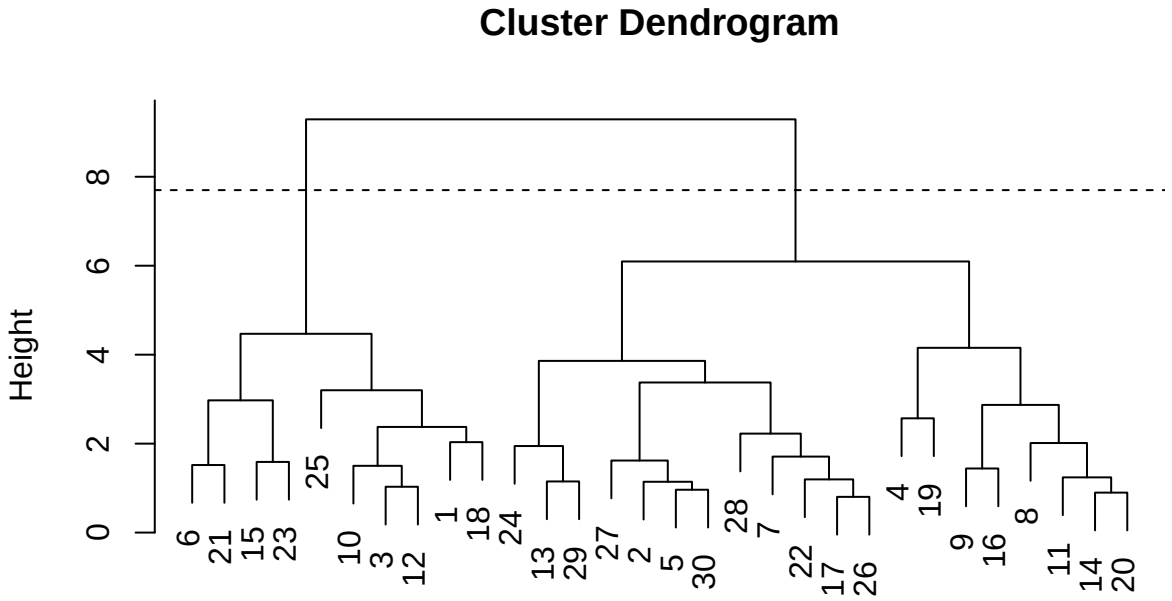

```
e <- table(cutree(hclust(dist(data[-(31:32),],method="euclidian")),h=700))
plot(hclust(dist(data[-(31:32),],method="maximum")),xlab="",sub="")
abline(500,0,lty=2)
```




```
max <- table(cutree(hclust(dist(data[-(31:32),],method="maximum")),h=500))
plot(hclust(dist(data[-(31:32),],method="manhattan")),xlab="",sub="")
abline(2080,0,lty=2)
```



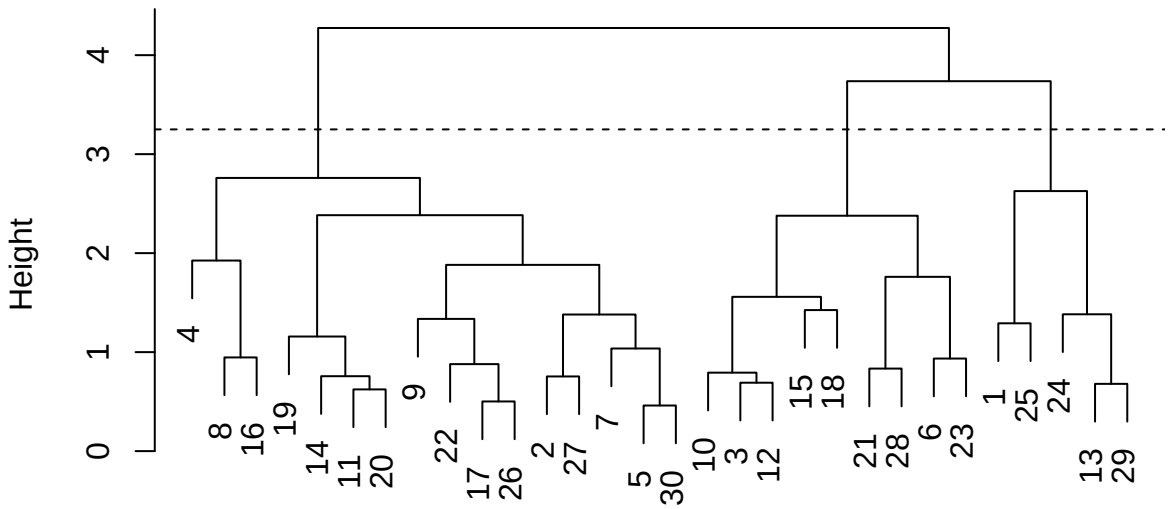

```
man <- table(cutree(hclust(dist(data[-(31:32),],method="manhattan")),h=2080))
data <- normalize(data)
plot(hclust(dist(data[-(31:32),],method="euclidean")),xlab="",sub="") # Hierarchical clustering on stan
abline(7.7,0,lty=2)
```




```
table(cutree(hclust(dist(data[-(31:32),],method="euclidian")),h=7.7))
```

```
##
##  1  2
## 10 20
```

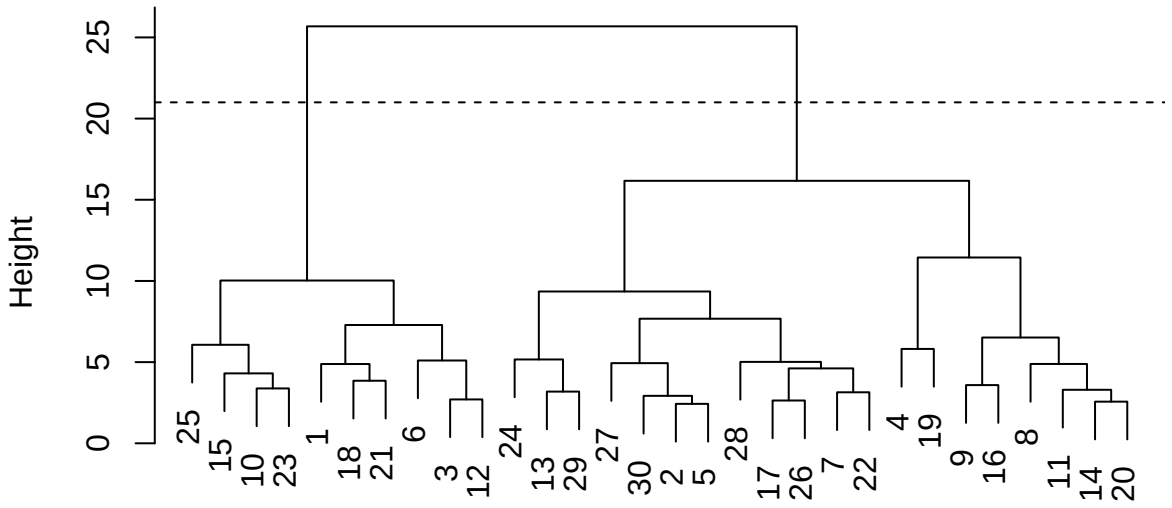
```
plot(hclust(dist(data[-(31:32),],method="maximum")),xlab="",sub="")
abline(3.25,0,lty=2)
```




```
table(cutree(hclust(dist(data[-(31:32),],method="maximum")),h=3.25))
```

```
##
##  1  2  3
##  5 16  9
```

```
plot(hclust(dist(data[-(31:32),],method="manhattan")),xlab="",sub="")
abline(21,0,lty=2)
```




```
table(cutree(hclust(dist(data[-(31:32),],method="manhattan")),h=21))
```

```
##
##  1  2
## 10 20
```

```
cat("              Euclid  Maximum  Manhattan \n Regular Data",e," ",max,"    ",man,"\n Standardized",ta
```

```
##               Euclid  Maximum  Manhattan
##  Regular Data 11 19   10 20      22 8
##  Standardized 10 20   5 16 9     10 20
```

```
# The processes using euclidean and Manhattan distances are similar because they both yield similar den
```