```
# Cluster analyis
library(caret)
# load the data

airlines.df<-read.csv("EastWestAirlinesCluster.csv", header= T)
head(airlines.df)

# normalize
airlines.df.norm <- sapply(airlines.df[,-1], scale)

# kmeans
km <- kmeans(airlines.df.norm, 4)
km
table(km$cluster)

library(factoextra)
library(ggpubr)
fviz_cluster(km, data = airlines.df.norm)

# function to compute total within-cluster sum of square

fviz_nbclust(airlines.df.norm, kmeans, method = "wss")

# plot an empty scatter plot
plot(c(0), xaxt = 'n', ylab = "", type = "l", ylim =
       c(min(km$centers), max(km$centers)), xlim = c(0, 11))
# label x-axes
axis(1, at = c(1:11), labels = names(airlines.df[,-1]))

# plot centroids
for (i in c(1:4)){
  lines(km$centers[i,], lty = i, lwd = 2, col = ifelse(i %in% c(1, 3, 5),"black", "dark
grey"))
}
# name clusters
text(x = 0.5, y = km$centers[, 1], labels = paste("Cluster", c(1:4)))
```