

# Unsupervised Learning - 1

Abhirup Sen

11/06/2021

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

## K-Means Clustering

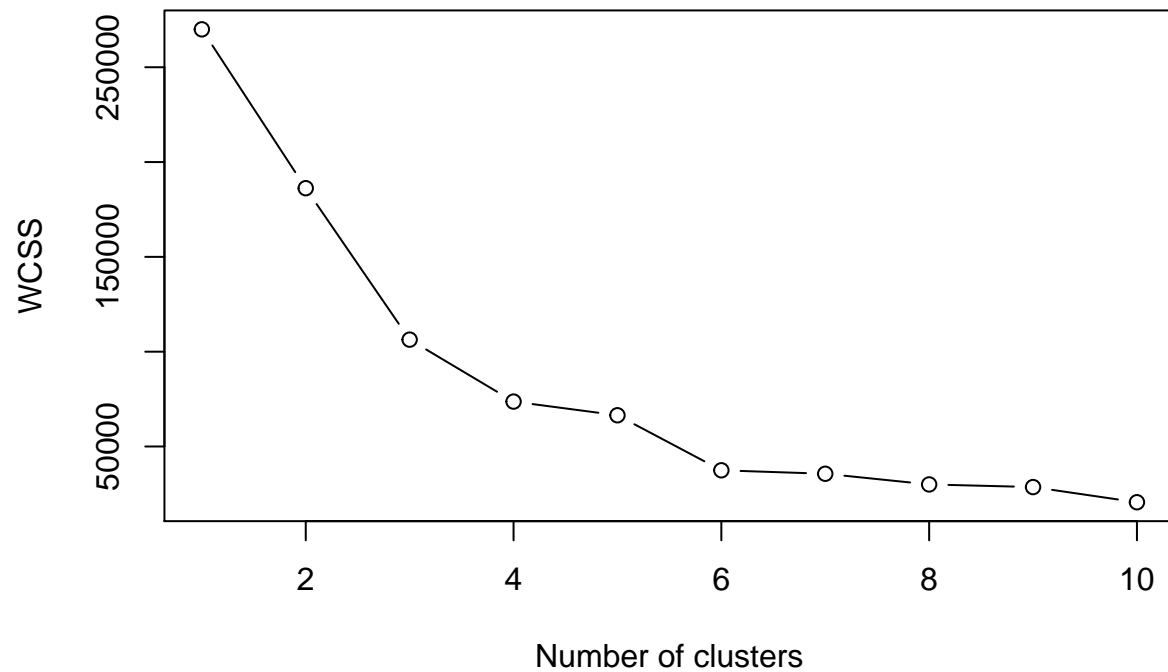
```
# Importing the dataset
```

```
dataset = read.csv('Mall_Customers.csv')  
dataset = dataset[4:5]
```

```
# Using the elbow method to find the optimal number of clusters
```

```
set.seed(6)  
wcss = vector()  
for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss)  
plot(1:10,  
     wcss,  
     type = 'b',  
     main = paste('The Elbow Method'),  
     xlab = 'Number of clusters',  
     ylab = 'WCSS')
```

## The Elbow Method



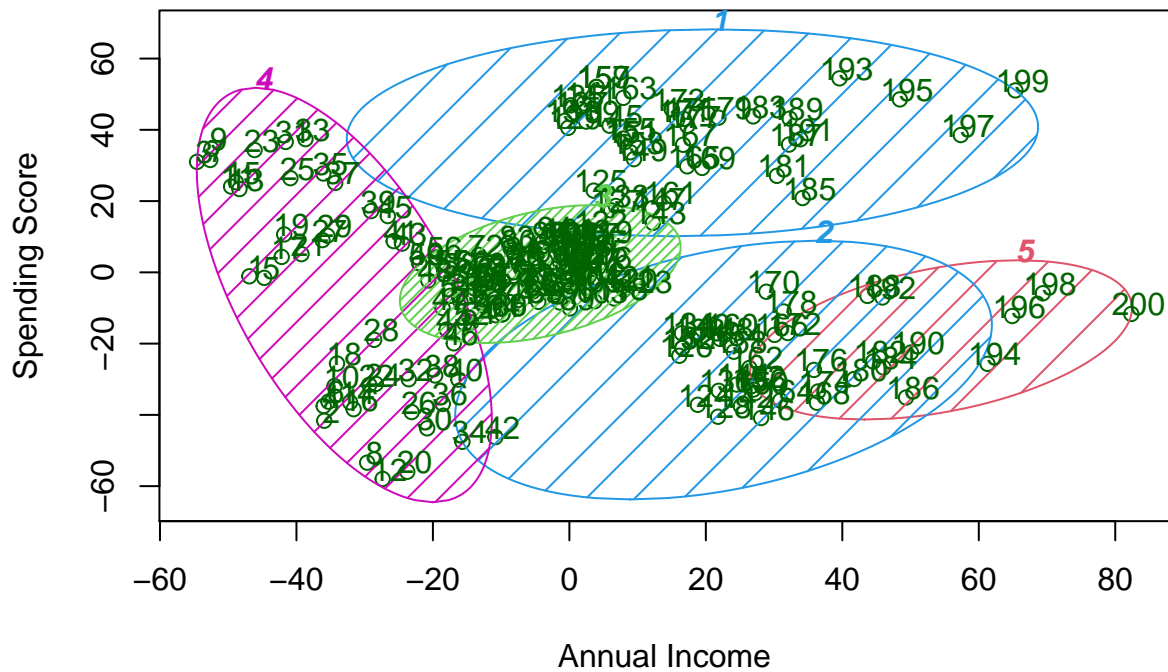
```
# Fitting K-Means to the dataset
set.seed(29)
kmeans = kmeans(x = dataset, centers = 5)
y_kmeans = kmeans$cluster
```

```
# Visualising the clusters
library(cluster)
?clusplot()
```

```
## starting httpd help server ... done
```

```
clusplot(dataset,
  y_kmeans,
  lines = 0,
  shade = TRUE,
  color = TRUE,
  labels = 2,
  plotchar = FALSE,
  span = FALSE,
  main = paste('Clusters of customers'),
  xlab = 'Annual Income',
  ylab = 'Spending Score')
```

## Clusters of customers



These two components explain 100 % of the point variability.

```
# Visualising the cluster Center
```

```
plot(dataset, col = kmeans$cluster, main = "K-Means with 5 clusters")
kmeans$centers
```

```
##   Annual.Income..k.. Spending.Score..1.100.
## 1      88.20000      17.11429
## 2      75.20000      82.56667
## 3      56.31579      49.52632
## 4      27.06250      47.70833
## 5     108.18182      82.72727
```

```
points(kmeans$centers, col = 1:3, pch = 8, cex=3) #cex = font size, pch = symbol
```

### K-Means with 5 clusters

