

Data Mining Techniques (Lab-3)

Bibek Sapkota

Install and Import packages

```
packages_to_install <- c("ggplot2", "factoextra")

for (package_name in packages_to_install) {
  if (!requireNamespace(package_name, quietly = TRUE)) {
    install.packages(package_name)
  }
}

library(ggplot2)
library(factoextra)
```

Welcome! Want to learn more? See two factoextra-related books at <https://goo.gl/ve3WBa>

task 2:set random seed

```
set.seed(7696)
```

task 3:Loading the data

```
data = read.csv("Mall_Customers.csv")
```

```
class(data)
```

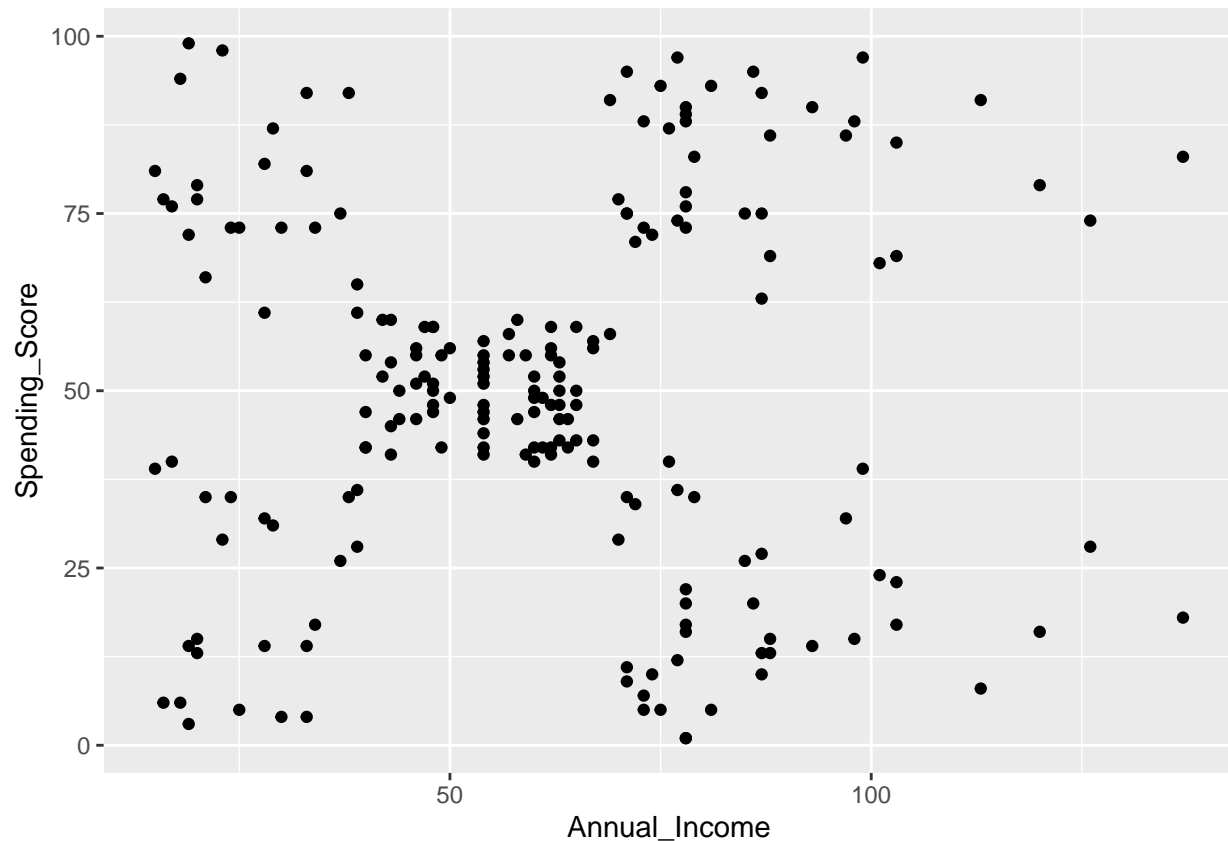
```
## [1] "data.frame"
```

task 4:filter data by column name

```
filtered_data1= data[, c("Annual_Income","Spending_Score")]
```

task5:Draw a scatter plot of data

```
ggplot(data, aes(x= Annual_Income, y= Spending_Score))+ geom_point()
```



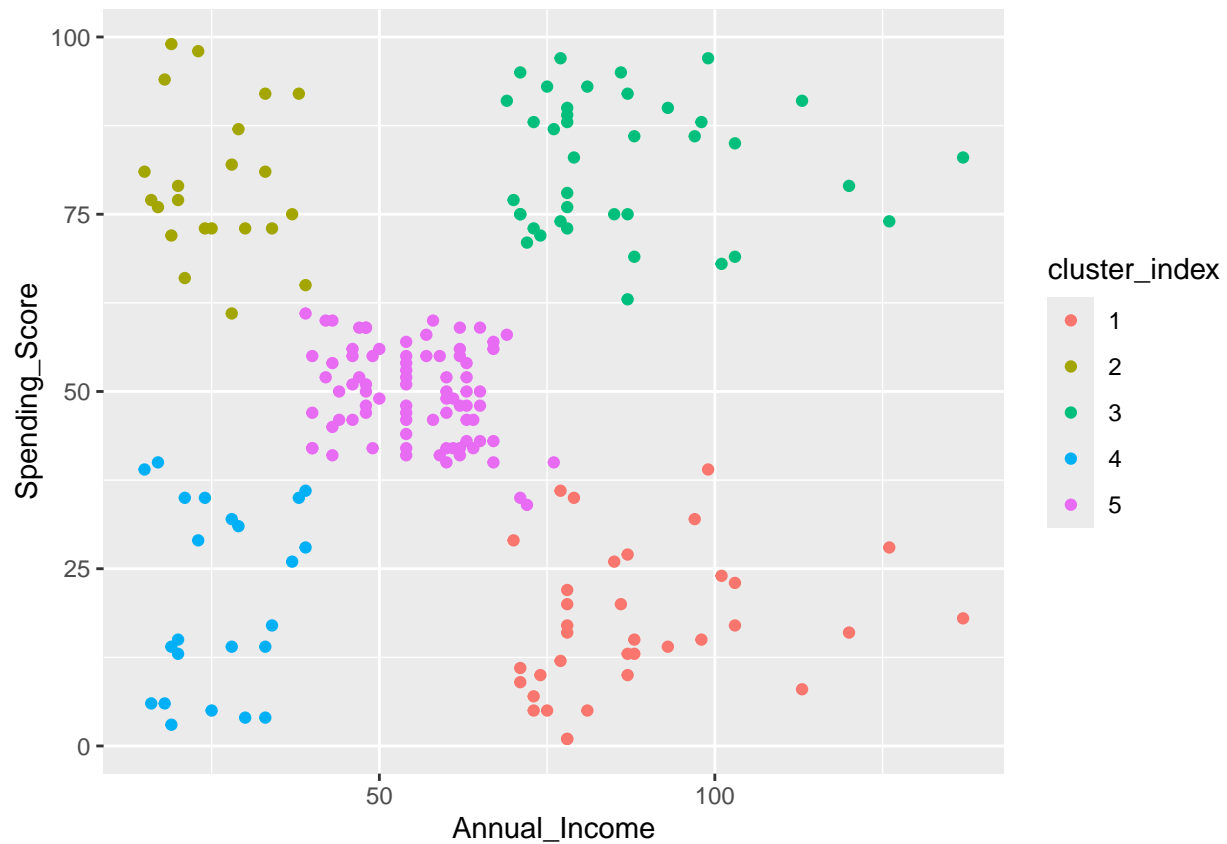
task 6: Finding 5 clusters using k-means clustering and viewing the details

[illegible]

```
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

task 7: Visualizing the cluster

```
cluster_index = as.factor(clusters$cluster)
ggplot(filtered_data1, aes(x=Annual_Income, y=Spending_Score, color= cluster_index )) +
  geom_point()
```



task 8: Filtering the data in filtered_data2

```
filtered_data2= data[, c("Age", "Annual_Income", "Spending_Score")]
```

task 9: Finding 5 clusters

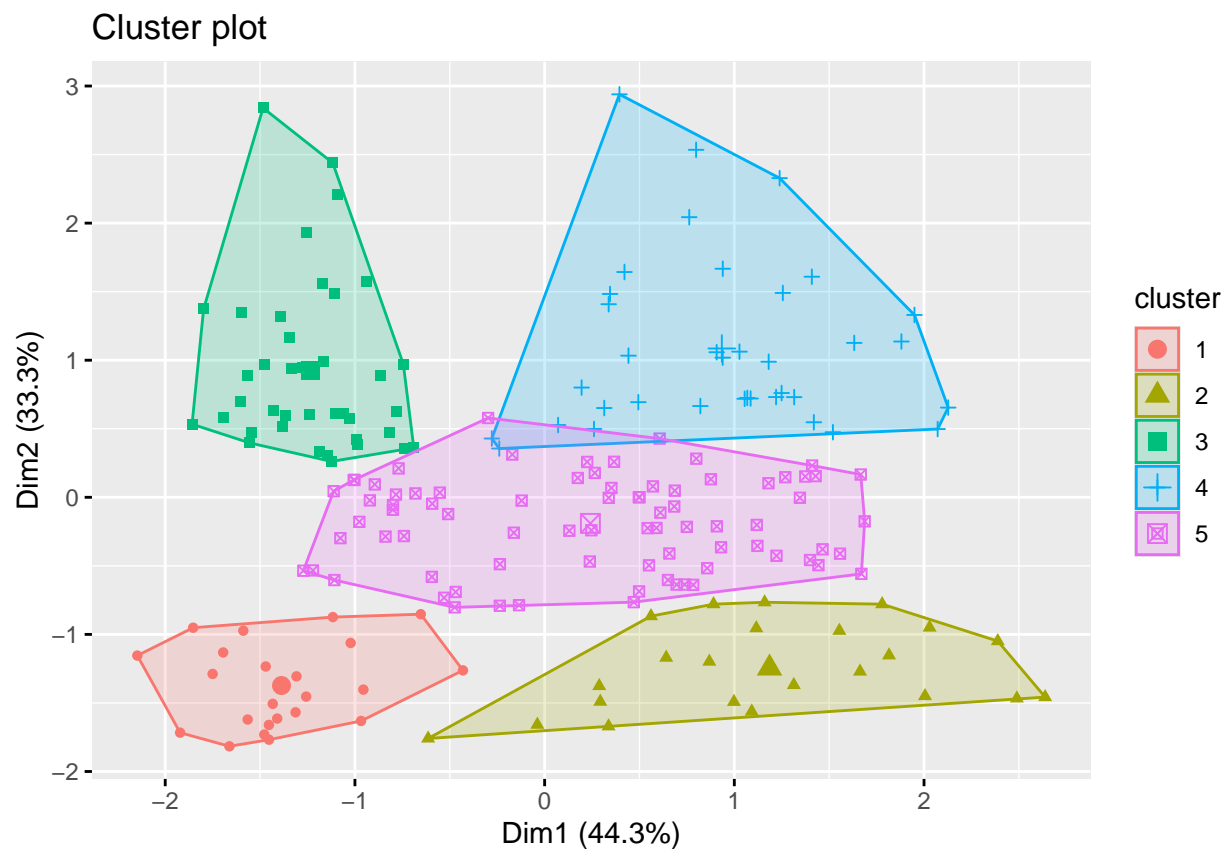
```
cluster2= kmeans(filtered_data2, 5)
cluster2
```

```
## K-means clustering with 5 clusters of sizes 23, 23, 39, 36, 79
##
## Cluster means:
##      Age Annual_Income Spending_Score
```

```
## 1 25.52174      26.30435      78.56522
## 2 45.21739      26.30435      20.91304
## 3 32.69231      86.53846      82.12821
## 4 40.66667      87.75000      17.58333
## 5 43.08861      55.29114      49.56962
##
## Clustering vector:
## [1] 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
## [38] 1 2 1 2 1 2 1 2 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
## [75] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
## [112] 5 5 5 5 5 5 5 5 5 5 5 5 3 4 3 5 3 4 3 4 3 4 3 4 3 4 3 4 3 5 3 4 3 4 3
## [149] 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4
## [186] 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3
##
## Within cluster sum of squares by cluster:
## [1] 4622.261 8948.609 13972.359 17669.500 30138.051
## (between_SS / total_SS = 75.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```

task 10:Ploting the cluster

```
fviz_cluster(cluster2, data= filtered_data2, geom= c("point"))
```



```
cluster3= kmeans(filtered_data2, 5 , nstart=25)
cluster3
```

[illegible]

```
fviz_cluster(cluster3, data= filtered_data2, geom= c("point"))
```

Cluster plot

