# Apply machine learning

It is an unsupervised as a machine learning problem.

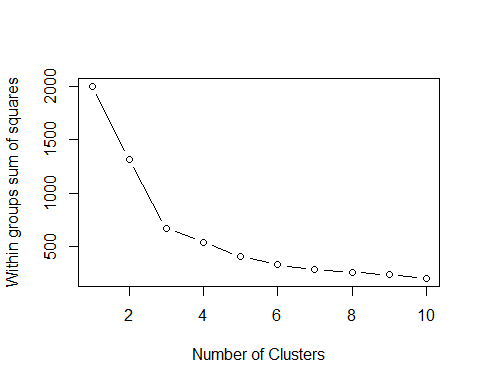
## Speed\_kms Diam\_min Moondist\_LD\_nom  
## 1 8.69 250 3.75  
## 2 10.65 37 15.55  
## 3 5.55 610 8.09  
## 4 3.13 40 12.63  
## 5 4.27 25 6.86  
## 6 10.97 92 19.09

## num [1:1000, 1:2] 0.811 -0.538 3.092 -0.519 -0.614 ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : chr [1:1000] "1" "2" "3" "4" ...  
## ..$ : chr [1:2] "Diam\_min" "Moondist\_LD\_nom"  
## - attr(\*, "scaled:center")= Named num [1:2] 121.9 12.5  
## ..- attr(\*, "names")= chr [1:2] "Diam\_min" "Moondist\_LD\_nom"  
## - attr(\*, "scaled:scale")= Named num [1:2] 157.86 4.73  
## ..- attr(\*, "names")= chr [1:2] "Diam\_min" "Moondist\_LD\_nom"

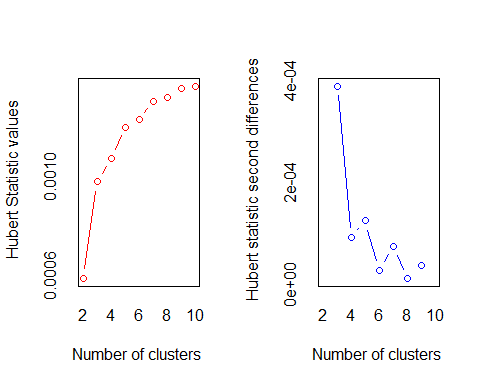
### Method 1: A plot of the total within-groups sums of squares against the

### number of clusters in a K-means solution can be helpful. A bend in the

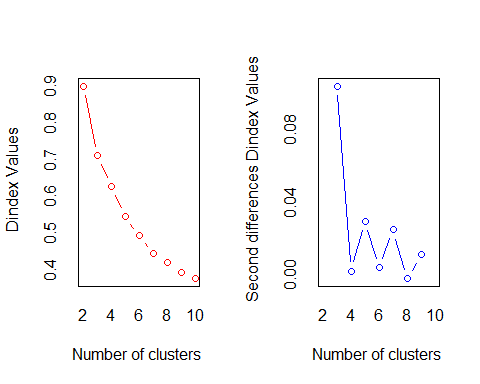
### graph can suggest the appropriate number of cluster



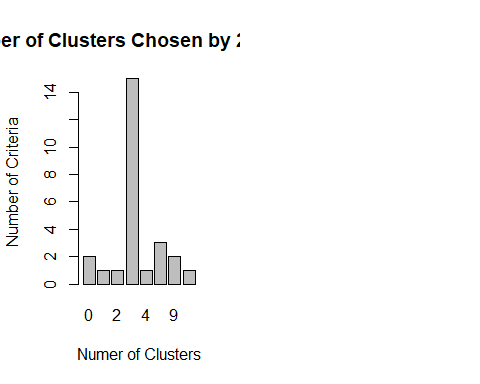
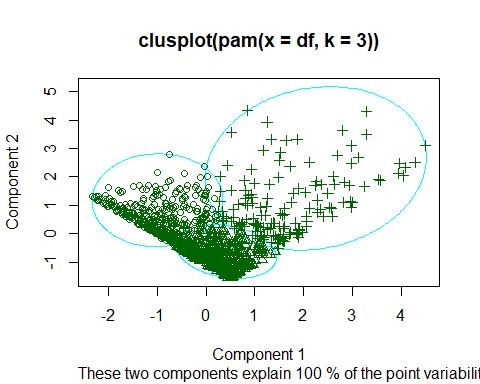
### Barplot sugested 10 clusters and number 3 is the most favorable.



## \*\*\* : The Hubert index is a graphical method of determining the number of clusters.  
## In the plot of Hubert index, we seek a significant knee that corresponds to a   
## significant increase of the value of the measure i.e the significant peak in Hubert  
## index second differences plot.   
##



## \*\*\* : The D index is a graphical method of determining the number of clusters.   
## In the plot of D index, we seek a significant knee (the significant peak in Dindex  
## second differences plot) that corresponds to a significant increase of the value of  
## the measure.   
##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*   
## \* Among all indices:   
## \* 1 proposed 2 as the best number of clusters   
## \* 15 proposed 3 as the best number of clusters   
## \* 1 proposed 4 as the best number of clusters   
## \* 3 proposed 6 as the best number of clusters   
## \* 2 proposed 9 as the best number of clusters   
## \* 1 proposed 10 as the best number of clusters   
##   
## \*\*\*\*\* Conclusion \*\*\*\*\*   
##   
## \* According to the majority rule, the best number of clusters is 3   
##   
##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

### \* Visualize these clusters using function clusplot() from the cluster library

### \* Would you consider this a good clustering?