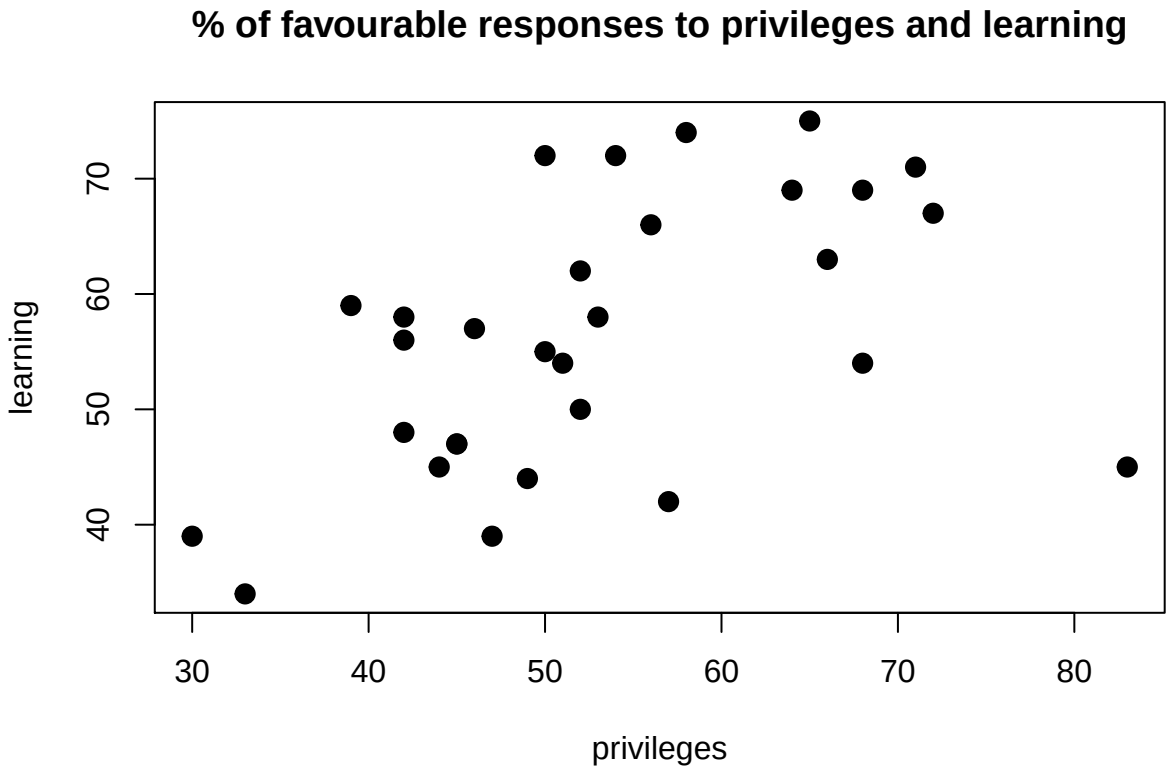
K-Means Clustering example.

Suchitra

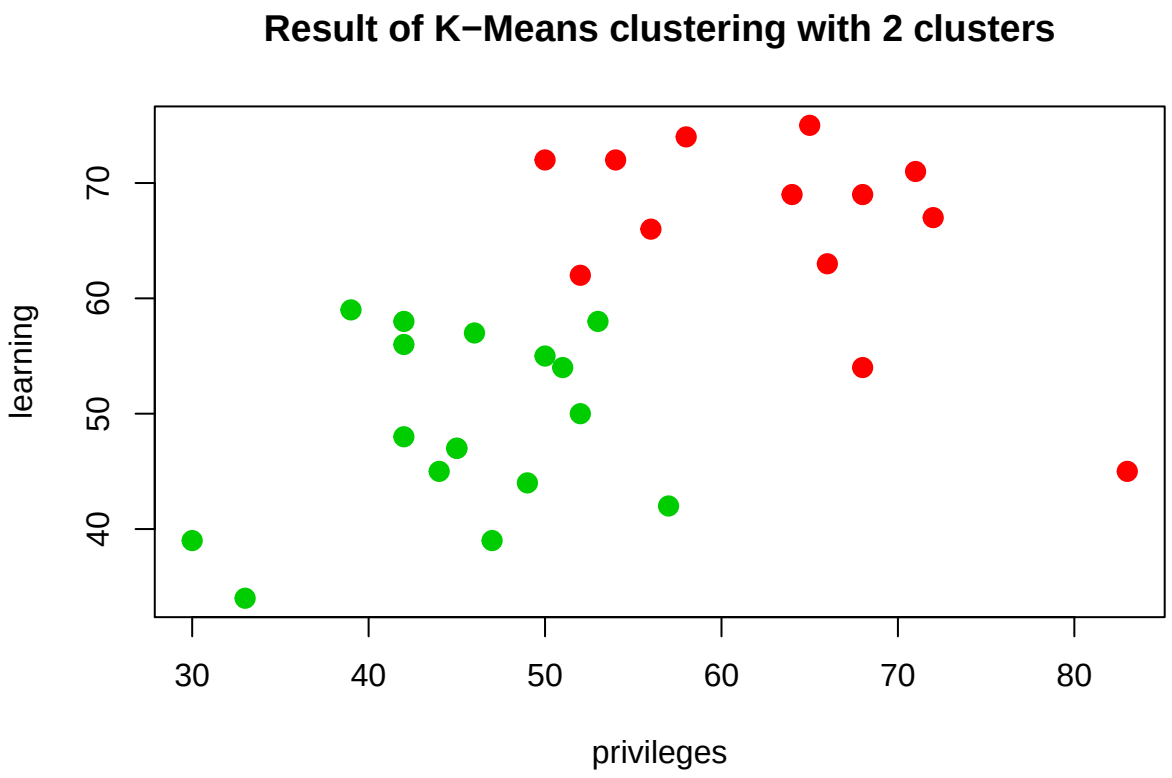
4/5/2017

```
data <- attitude[,c(3,4)]
plot(data, main="% of favourable responses to privileges and learning",pch=20, cex=2)
```



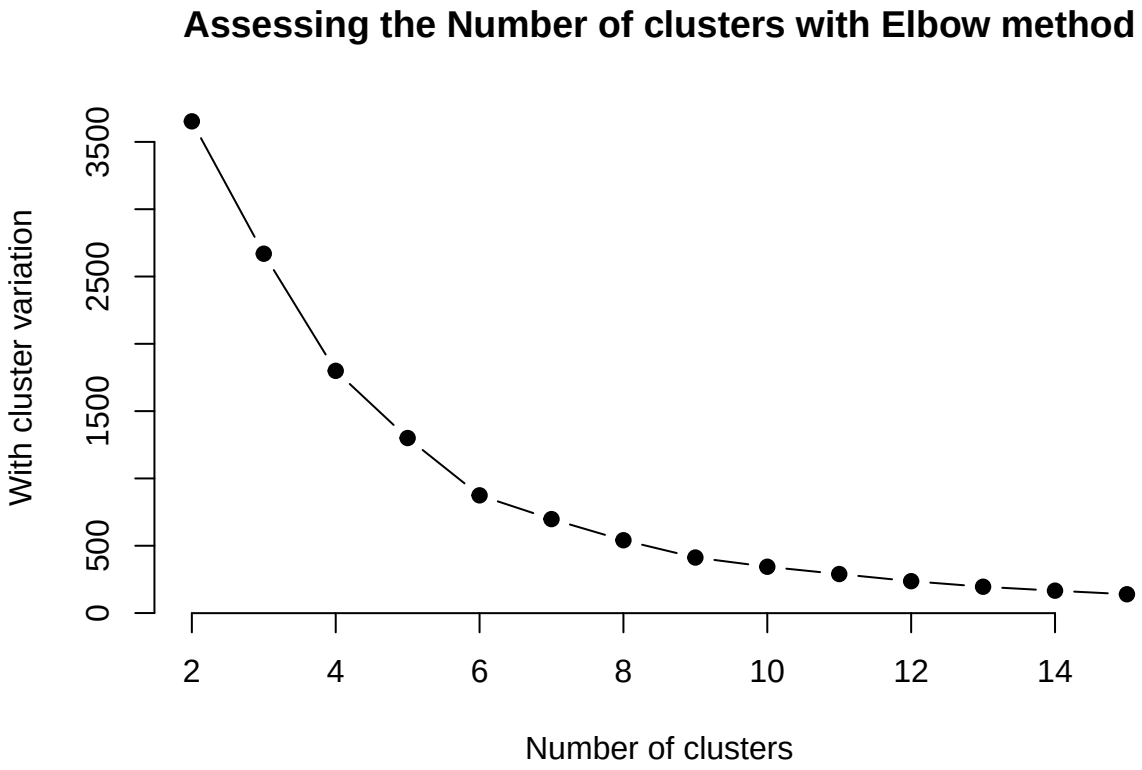

```
#Perform K-Means with 2 clusters
set.seed(7)
km1= kmeans(data, 2, nstart = 100)

plot(data, col= (km1$cluster + 1), main="Result of K-Means clustering with 2 clusters", pch=20, cex=2)
```



Elbow method to decide the number of clusters. Plotting the within cluster dissimilarity

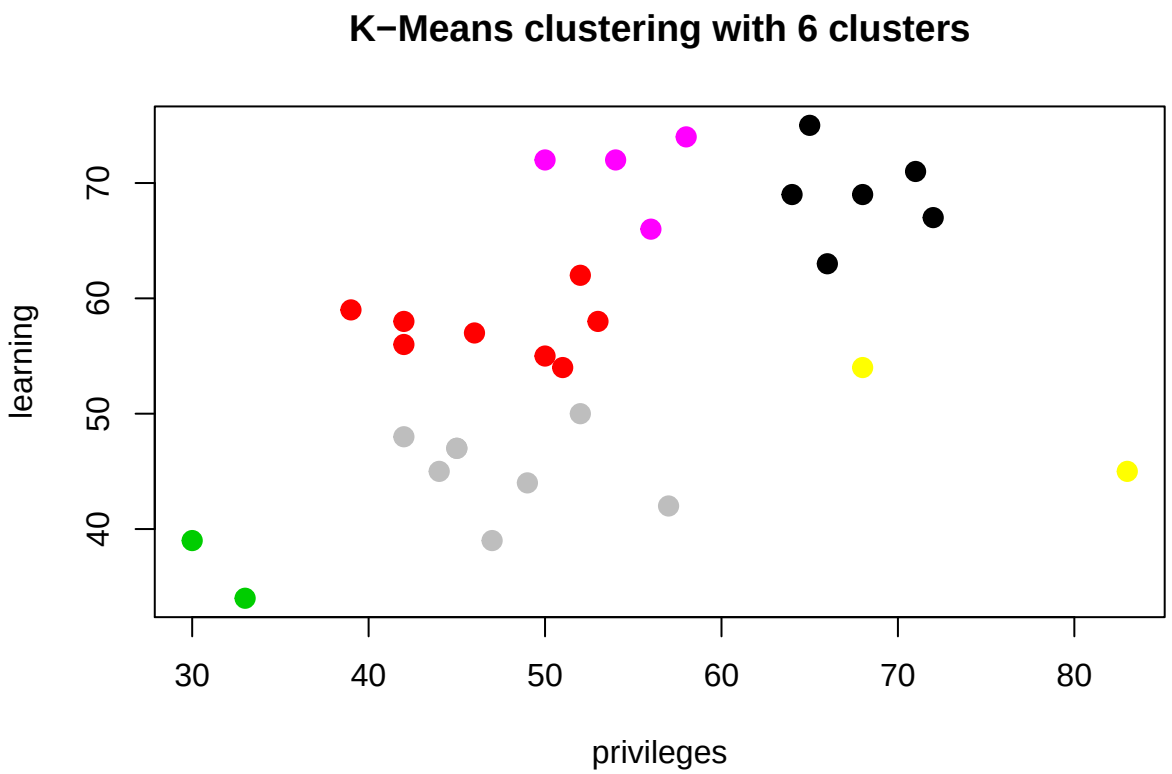
```
wss <- sapply(2:15, function(k){kmeans(data,k,nstart=100)$tot.withinss})
plot(2:15, wss, type='b', xlab = "Number of clusters", ylab = "With cluster variation", pch=19, frame=F
```



This plot is used to find the optimal number of clusters to perform K-Means clusterin on the given dataset. In this, we apply the k-Means function on a range of cluster values and find the within cluster variation(euclidean distance between the data points and the centroid of the cluster). The optimal cluster number is chosen when the within cluster variation stabilizes. From the above plot, we can see that the variation decreases negligibly after the cluster number reaches 6. Thus we can perform K-Means clustering using 6 as the cluster number.

```
set.seed(7)
km2=kmeans(data,6,nstart = 100)

plot(data, col=(km2$cluster + 5), main="K-Means clustering with 6 clusters", pch=20, cex=2)
```



The above plot shows 6 groups of departments having distinct responses about the prvileges and learning in the financial organization. This information could be of great use to understand if ceratin departments have more privileges and learning opportunities than the other and understanding the reasons behind them.