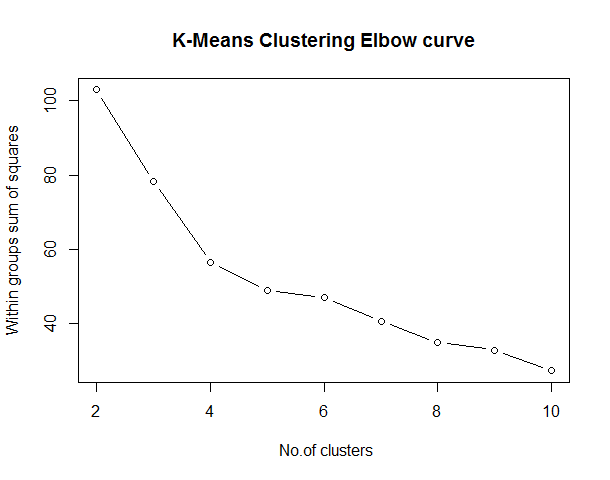
1. **Crime data set**

**Business objective:** Minimize public’s Risk

**INFERENCE:**  The given data set contains crime data in different states in USA. So ,this data set contains the no.of murders,rapes,pop culture areas, assaults. After importing the data and normalizing all the columns except ‘states’ and afternormalising the data ,I have experimented various iterations from (2:8) and I have plotted the elbow curve



I have decided to apply 3 clusters .

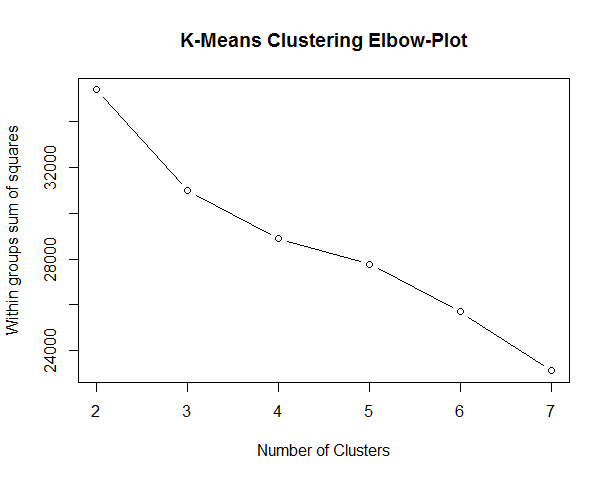
1. Safest cities
2. Risky cities
3. Dangerous cities

In the first cluster .The crime rate is very low and in second cluster crime rates are moderate and in cluster 3 the crime rate is too high and it is too risky to stay in that city.

1. **East west airlines**

**Business objective:** To identify customer segments via kmeans\_clustering**.**

**Inference**: The main objective is to segment the customers based on their flying patterns,frequently use of rewards and use of airline credit.This is a large data set, this contains 4000 rows and 12 columns. After importing the data and normalizing all the columns except ‘ID’ and after I have used iterations from 2 to 7 .Then when I have plotted the elbow curve



After watching the elbow curve ,I have decided to cluster the customers in 3 clusters

. **Group.1 Balance Qual\_miles cc1\_miles cc2\_miles**

**1 1 73134.42 143.9474 2.059339 1.014522**

**2 2 1704838.00 0.0000 1.000000 1.000000**

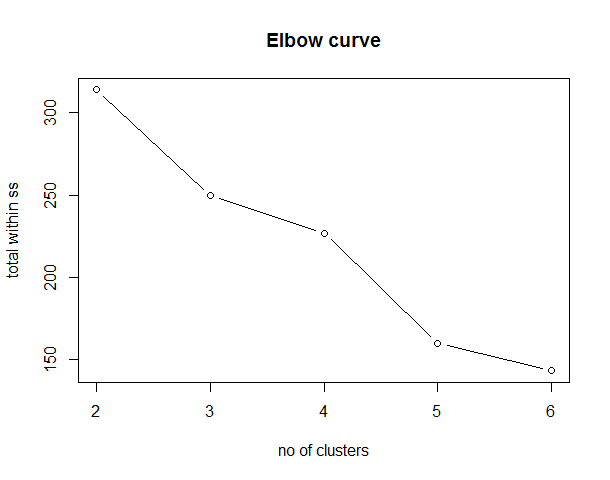
**3 3 131999.50 347.0000 2.500000 1.000000**

So,here the group 3 people are the people who travel most frequently Based on number of miles counted as qualifying for Topflight status,group 3 customers also earn moe cc1 miles .Then comes group 2 where the people are least travelled.They earn very less cc1 miles.Then comes group customers where are more frequentlu travelling customers , have a average earned cc1,2 miles.

1. **Insurance data set**

**Business objective:** To identify customer segments via kmeans\_clustering**.**

**Inference**: The main objective is to cluster the customers into clusters.First I have imported the data set and I have normalized the data set and then I have used 2:5 iterations and then I have plotted the elbow curve.

****

**Group.1 Premiums.Paid Age Days.to.Renew Claims.made Income**

**1 23164.286 61.09524 152.57143 26153.455 155500.00**

**2 9010.526 47.10526 244.52632 17312.250 77263.16**

**3 9942.917 38.55 69.833 6329.067 91525.00**

**Then I have clustered them into 3 clusters .When we observe the clusters the cluster 0 they are the old age people who are 60+, so they have high chance of getting hospitalized and claiming the insurance and they all are having very high incomes and they are also claiming the highest of all the 3 clusters customers.So we can increase the premium for these customers. In the second cluster there are middle aged men 40+ and we can see that they are claiming 17k and the premium paid is 9k and they are renewing after 244 days .In the last cluster where we have bunch of young people.They are the people who have less chanance to get hospitalized and they also maynot get that high income to pay high premiums .So ,If we can concentrate on the top 2 clusters add more plans and increase the premiums we can get profits as every one are earning high salaries.**