Lab Exercises

1. Consider the dataset *USArrests* which is in the package **base** in R. Your task is to group the 50 states in United States on the attributes of crime namely, Murder, Assault, UrbanPop, Rape. Use k-means clustering and also determine the optimal number of clusters. Try number of clusters from 2 to 15.

2. Consider the dataset wine which is in the package *gclus* in R. Your task is to group the wines on the following attributes (Ignore the variable `Class`):

```
1) Alcohol
2) Malic acid
3) Ash
4) Alcalinity of ash
5) Magnesium
6) Total phenols
7) Flavanoids
8) Nonflavanoid phenols
9) Proanthocyanins
10)Color intensity
11)Hue
12)OD280/OD315 of diluted wines
13)Proline
```

Use hierarchical clustering to group the wines. Use “ward.D” method with distance as “euclidean”. Try number of clusters from 2 to 15. Find the optimal number of clusters. Assign the clusterIDs to the respective observations.