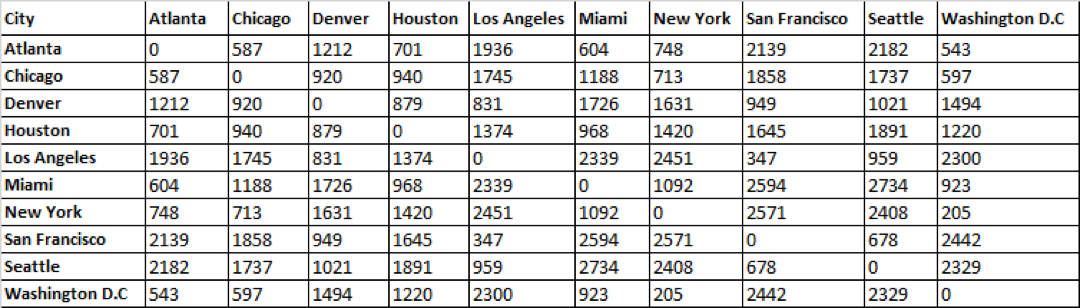
**Project 5: Unsupervised Learning**

Part I: Use NCI60 data of ISLR2 package and page 540 of ISLR2 book to do as follows in R Studio to knit PDF output:

1. Define nci labels (NCI$labs) as nci.labs and nci data (NCI$data) and nic.data
2. Check dimension of nci.data object and interpret it carefully
3. Check first four cancer types using nci.labs object
4. Fit principal component analysis (PCA) on nci.data with scale = TRUE argument as pr.out object
5. Create a plot showing first three PCA components with three different colors
6. Get summary of pr.out object and interpret it carefully
7. Plot pro.out object and interpret it carefully
8. Create custom scatterplots with principal components in x-axis and proportion variance explained (PVE) in y-axis for the first plot and cumulative PVE in the y-axis for the second plot and interpret them carefully
9. Perform PCA with varimax rotation and compare it with the PCA result obtained above
10. Write summary of the results and conclusion based on your findings

Part II: Use the distance between 10 US cities provided below in R Studio to knot PDF output:



1. Get dissimilarity distance as city.dissimilarity object
2. Fit a classical multidimensional model using the city.dissimilarity object
3. Get the summary of the model and interpret it carefully
4. Get the bi-plot of the model and interpret it carefully

Part III: Part I: Use NCI60 data of ISLR2 package and page 543 of ISLR2 book to do as follows in R Studio to knit PDF output:

1. Scale the nci.data as sd.data object
2. Fit hierarchical cluster analysis on the sd.data using complete, average and single linkage methods, show the results with dendrogram and interpret them carefully
3. Find the best number for clusters using “cutree” function with best distance value
4. Use your roll number as set.seed and perform k-means clustering on sd.data with the best number of clusters/distance value with nstart=20
5. Get summary of the k-means clustering and interpret them carefully
6. Plot this k-means results using base r plot and cluster package and interpret them carefully

Part IV: Use “Groceries” data available in the “datasets” package to do as follows in R Studio to knit PDF output

1. Load “arules” and “arulesViz” libraries
2. Load “Groceries” data, check its structure and interpret it carefully
3. Get Frequent Item frequencies using itemFrequencyPlot function and interpret it carefully
4. Set a priori rule with support = 0.001 and confidence = 0.8 and interpret the output carefully
5. Show the top five rules using inspect and round the results to two digits
6. Sort the rule by confidence in decreasing order
7. Use “whole milk” as target item and show the items in “lhs” with decreasing order of confidence and show the top five rules
8. Use “whole milk” as target item and show the items in “rhs” with decreasing order of confidence and show the top five rules
9. Write summary and conclusion based on your findings above