For the following questions use the data in the file `cities1.xlsx`. It contains data on 325 metropolitan cities in the United States.

- Let column `Metropolitan_Area` be the row names of your dataframe.
- Remove the non-numeric variables, `Crime_Trend and Unemployment_Threat`).
- Use scale() function to scale all numeric columns.
- Use function dist to find the distance between cities (on the scaled data).

K-MEANS CLUSTERING

1. (10 pts) Use set.seed(123) and the user function twcv to find TWCV values for $k = 1 : 16$. Use nstart = 25. Display the elbow chart.
2. (10 pts) The best number of clusters is the smallest k such that the cluster plot shows the least amount of clusters overlap. Use `fviz_cluster( )` with argument geom = "point" to display cluster plots with no label names. Try `fviz_cluster( )` with different K. What is the best K? For this K find the number of cities in each cluster.
3. (10 pts) Find the median (or mean, if you prefer) of each numerical column (on the original un-scaled dataset). Write one sentence characterizing each cluster.

HIERARCHICAL CLUSTERING

4. (20 pts) Use function hclust with linkage ward.D to create object h1 and display the four clusters on the dendrogram. Use function `cuttree( )` to find the clusters. Find the number of cities in each cluster. Use `fviz_cluster( )` with argument geom = "point" to display the cluster plots of your choice with no label names. Find the CCPC for ward.D
5. (20 pts) Use function hclust with linkage complete to create object h2 and display the four clusters on the dendrogram. Use function `cuttree( )` to find the clusters. Find the number of cities in each cluster. Use `fviz_cluster( )` with argument geom = "point" to display the cluster plots of your choice with no label names. Find the CCPC for complete linkage.
6. (20 pts) Use function hclust with linkage average to create object h3 and display the four clusters on the dendrogram. Use function `cuttree( )` to find the clusters. Find the number of cities in each cluster. Use `fviz_cluster( )` with argument geom = "point" to display the cluster plots of your choice with no label names. Find the CCPC for average linkage.
7. (10 pts) What linkage do you prefer? For the clusters found for this linkage find the median (or mean, if you prefer) of each numerical column (on the original un-scaled dataset). Write one sentence characterizing each cluster for this linkage.

Submit your report (code and output) as a pdf file onto Blackboard (no screen captures). Read your pdf file before submitting.