

*To submit this assignment, upload the full document on Canvas, including your R Markdown file and the output as a **knitted.docx file (for problem 1)** and **.html file (for problem 2)**.*

Problem 1 (40 Marks)

Plot the Boston house price dataset:

- a) Using the ggplot function to plot a basic scatterplot with x-axis as dis (weighted distances to five Boston employment centers) and y-axis as medv (Median value of owner-occupied homes in \$1000). (6 marks)
- b) Modifying the scatterplot from the last step in terms of color, size, and transparency. (5 marks)
- c) Adding a line of best fit, and then briefly discuss what is the relationship between dis and medv. (5 marks)
- d) Grouping the points with different colors in the scatterplot based on chas (Charles River dummy variable (1 if tract bounds river; 0 otherwise)). (5 marks) note: before running this code, you need to convert chas variable into a categorical variable. You could achieve this by using this code:
Boston\$chas <- as.factor(Boston\$chas)
- e) Using the facet_wrap function to generate two scatterplots based on chas. (5 marks)
- f) Including the appropriate labels in the plot. (6 marks)
- g) Picking a theme for the plot. (2 marks)
- h) Producing a correlation plot for all numeric variables in the dataset. (6 marks)

Problem 2 (10 Marks):

Using the leaflet package to generate an interactive graph that includes a place in Google Map. Please pick a place, which is meaningful to you. For example, it could be your hometown or the first place that you want to travel to after the end of the pandemic. Also include short explanations using the popup argument.

To find a location's Latitude & Longitude in Google Maps, please refer to the instructions here:

<https://support.google.com/maps/answer/18539?co=GENIE.Platform%3DDesktop&hl=en&oco=0>

Also, feel free to share your place with your classmates on the discussion board of Week 2.

Note: There are 13 variables in the dataset. They are:

1. **crim** - per capita crime rate by town
2. **zn** - proportion of residential land zoned for lots over 25,000 sq.ft.
3. **indus** - proportion of non-retail business acres per town.
4. **chas** - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5. **nox** - nitric oxides concentration (parts per 10 million)
6. **rm** - average number of rooms per dwelling
7. **age** - proportion of owner-occupied units built prior to 1940
8. **dis** - weighted distances to five Boston employment centers
9. **rad** - index of accessibility to radial highways

10. **tax** - full-value property-tax rate per \$10,000
11. **ptratio** - pupil-teacher ratio by town
12. **lstat** - % lower status of the population
13. **medv** - Median value of owner-occupied homes in \$1000

