**DATA 180 Review Questions: All questions require importing the “tidyverse” library. Remember, you can input ?variable\_name to receive a description of the pre-loaded data.**

1. In txhousing data, calculate the average sales prices for each city, in each month.
   1. Which cities are performing best over time? Visualize this information in some coherent way.
   2. Now aggregate this up to the year level. Report the average for each year. Use ggplot to compare cities in the year 2014. Which city has the highest average price home?
   3. In the year 2012, perform a hierarchical clustering process on the set of cities. Remove the year, month, and data columns. Which cities are deemed as similar? Why might that be?
   4. Now look at the listings, sales, volume, and average sales price columns you have created for each year. Conduct a kmeans clustering process to evaluate what years are similar to one another. Use the optimal number of clusters.
2. In mtcars, evaluate the similarity of these mixed data. Run the calculation in separate ways.
   1. With both the 0-0 matches containing information in vs and am:
   2. With neither vs and am 0-0 matches containing information:
   3. With one containing information and not the other.
   4. How does this change the distance between the objects?
3. In longley data, complete the following:
   1. Calculate the GNP per Capita for each year and plot the data. What is the trend?
   2. Create a scatter plot of the relationship between GNP per Capita and the unemployment rate.
   3. Create a linear regression model with a dependent variable of GNP per Capita and this unemployment rate. What is the interpretation?
   4. Now add the GNP deflator to the regression. How does this change your coefficients?