**V506 Fall 24 R Lab 6 In-Lab Exercise**

You will be using the [txhousing data](https://ggplot2.tidyverse.org/reference/txhousing.html) from the ggplot2 package. This is a data frame with 8602 observations and 9 variables:

* city: Name of multiple listing service (MLS) area
* year: Year
* month: Month
* date: Date of month
* sales: Number of sales
* volume: Total value of sales
* median: Median sale price
* listings: Total active listings
* inventory: "Months inventory": amount of time it would take to sell all current listings at current pace of sales.

1. Install default load packages.
2. Save the txhousing data as a data.frame or tibble in your R studio environment.
3. Explore the data using some of the common exploration functions we’ve used.
   1. This could be the glimpse, summary, head, View, etc. functions.
   2. You can also install the modelsummary package and using datasummary\_skim().
4. Replace the NAs in the sales, volume, median, listings, and inventory variables with the value 0. Save as *txhousing\_clean.*
5. Using the txhousing\_clean data, create a new dataframe called txhousing\_year. This is created by cleaning txhousing\_clean. Then, group by the city and year and find the average volume, average median home price, average sales, average listing, and average inventory.
6. Plot the *txhousing\_year* data using a scatterplot. Put the average sales on the x axis, the average median price on the y axis, and give this plot an appropriate title. Also, set the color of the points to be different for each year.
7. Create a histogram of the median price distribution. Use facet\_wrap to break out the plot by year. Again, use appropriate labeling.
   1. If you need to: you can change the x-axis text to be tilted. I used: “theme(axis.text.x=element\_text(angle = 45, hjust = 1))” but feel free to play around with it on your own.
8. Filter the *txhousing\_year* to be those for which the year is between 2007 and 2009 (that is, greater than or equal to 2007 and less than or equal to 2009).
9. Calculate the 95% confidence interval for the median price, using the *txhousing\_clean* data frame.

There is nothing to turn in for this exercise. When you are finished, you are free to leave.