**V506 Fall 24 R Lab 3 In-Lab Exercise**

Use Rmarkdown to document your response to each question.

1. Load the “dplyr”, and “ggplot2” packages using “pacman”.
2. Ggplot includes the dataset *diamonds* which has 53940 rows and 10 variables. The variables are:
   * **price**: price in US dollars (\$326–\$18,823)
   * **carat**: weight of the diamond (0.2–5.01)
   * **cut**: quality of the cut (Fair, Good, Very Good, Premium, Ideal)
   * **color**: diamond colour, from D (best) to J (worst)
   * **clarity**: a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
   * **x**: length in mm (0–10.74)
   * **y**: width in mm (0–58.9)
   * **z**: depth in mm (0–31.8)
   * **depth**: total depth percentage = z / mean(x, y) = 2 \* z / (x + y) (43–79)
   * **table**: width of top of diamond relative to widest point (43–95)

**Load this data set and, using pipes, print out the summary statistics of this dataset.**

1. Find the median diamond price and the mean diamond price. Are they different?
2. Use the t.test() function to test the hypothesis that the price of diamonds is $1,500. Now use the t.test() function to test the hypothesis that the price of diamonds is $3,950. What can you say about each of these t-tests?
3. Find the 95% confidence interval of the average diamond price. Hint: use the t.test function to test function to test that the price of diamonds is 0. Store the test as an object and explore the output it provides.
4. Using ggplot(), create a histogram of the price distribution of the diamonds. Does the shape of the distribution make sense given the mean and median price?
   1. This chart should follow chart best practices—labels on the axes, a title, an appropriate binwidth, etc.
5. Save a new dataset called *diamonds2* where you have added a variable that is equal to 1 if the price of the diamond greater than or equal to the median price. This variable will be equal to 0 if the price of the diamond is below the median price. Call this new variable **high\_price.** (You should use ifelse(), mutate, and pipes to create this variable).
6. Using ggplot(), create a scatterplot with carat on the x-axis, price on the y-axis. Use the facet\_wrap() to create 5 separate plots for each “cut”. Again, include the axis titles and a main title.
7. Using ggplot(), create a bar chart where the x-axis is the “color” of the diamond. Again, include the axis titles and a main title.

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