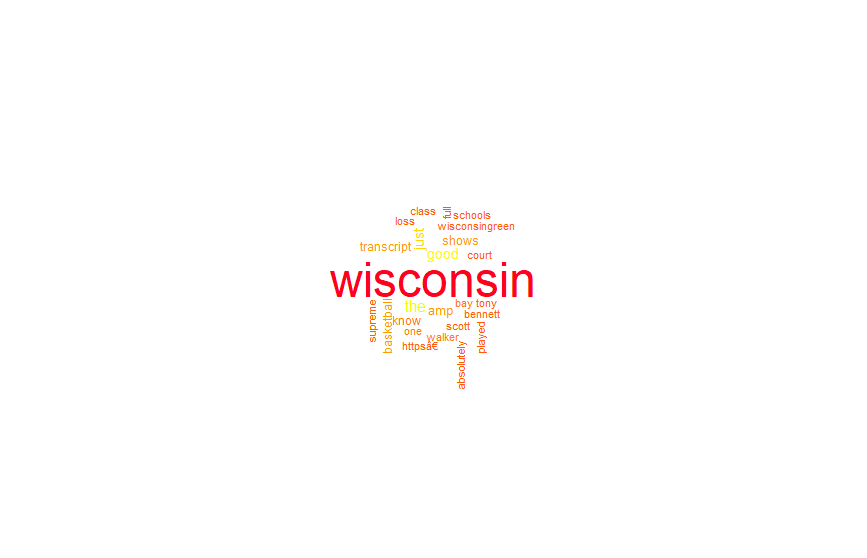
Matt Allen

Ds 700

Assignment 2

**Exercise 1: Wisconsin word cloud**

**In R Studio, create a word cloud for the term “Wisconsin” using at least 1,000 of the most recent tweets from Twitter. Show the 25 most frequent words in your word cloud. Take a screenshot of the word cloud in R Studio and paste it into a Word document for submission. Also save your R syntax as an R file.**

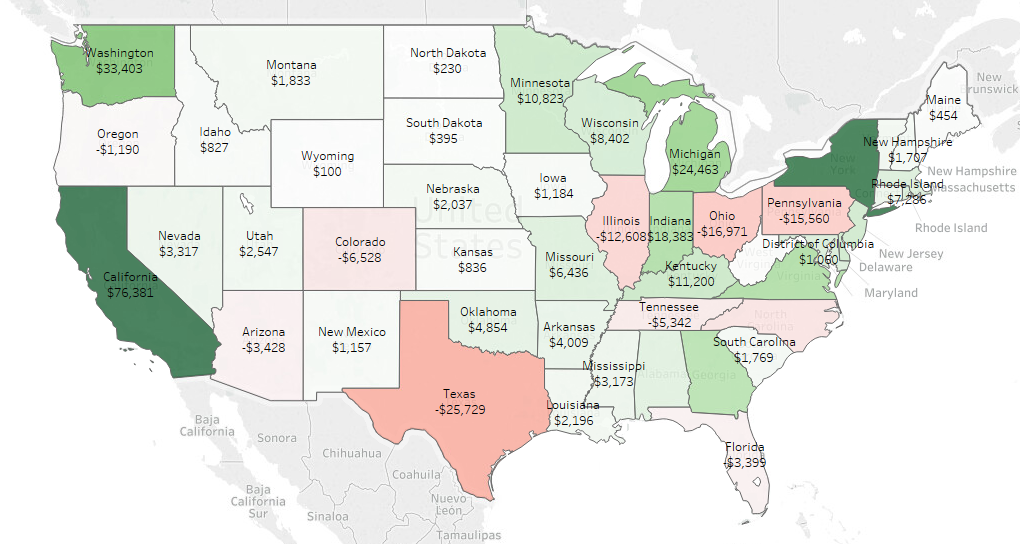


**Exercise 2: Tableau Map Visualizations**

**Locate the Sample - Superstore Subset dataset, which comes built in to Tableau. Using this dataset, create two map visualizations:**

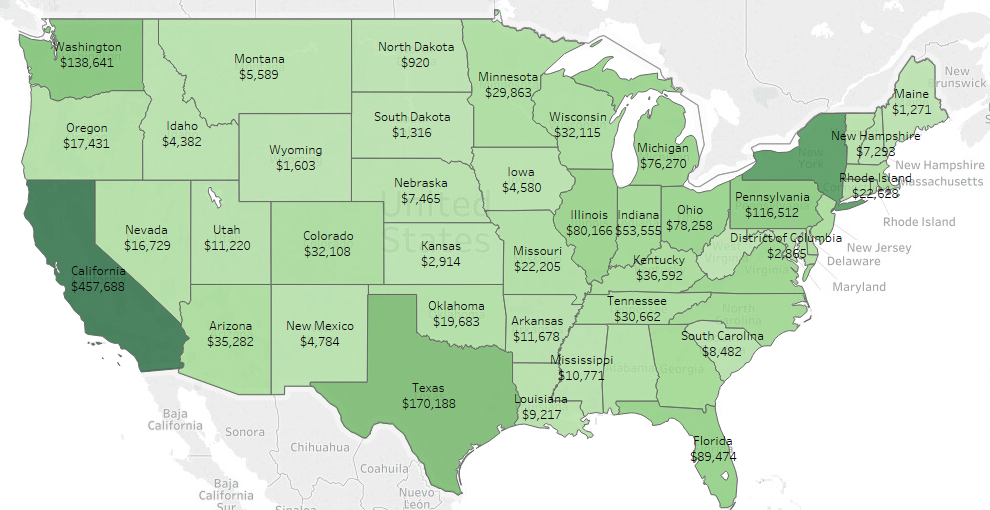
**Are there any U.S. states in the dataset that have lost money? Show the profitability of each state using a color scale. For example, the most profitable states should be a darker shade of green than less profitable states. Similarly, use a red color scale (darker to lighter) to show non-profitable states. (Worth 15% of the total assignment grade.)**

Profitability of States



Yes, there are states that have lost money. They are represented as a shade of red. Texas lost the most money at -$25,729. California was the most profitable at $76,381.

**Which state has the highest total sales in the dataset? Using a color scale, show the total sales for each U.S. state. Label each state with the state name. (Worth 15% of the total assignment grade.)**

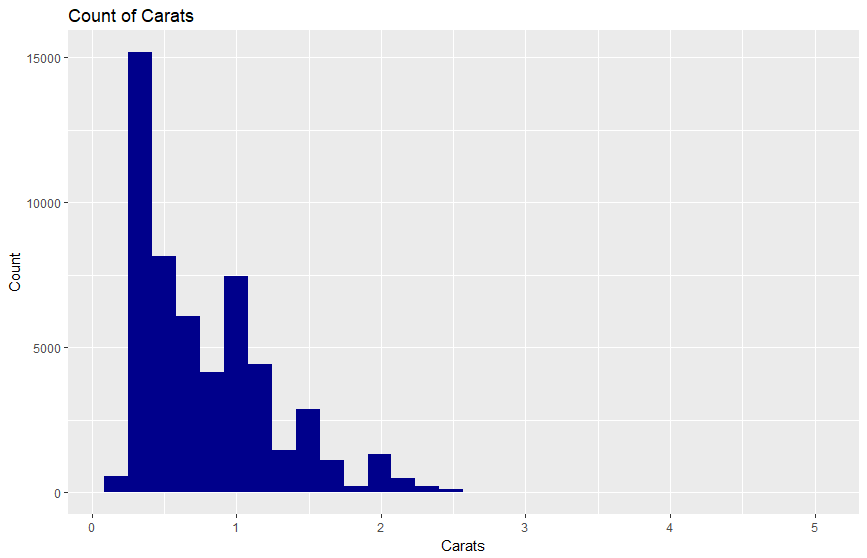


California has the highest total sales at $457,688.

**Exercise 3: Visualizations in R Studio**

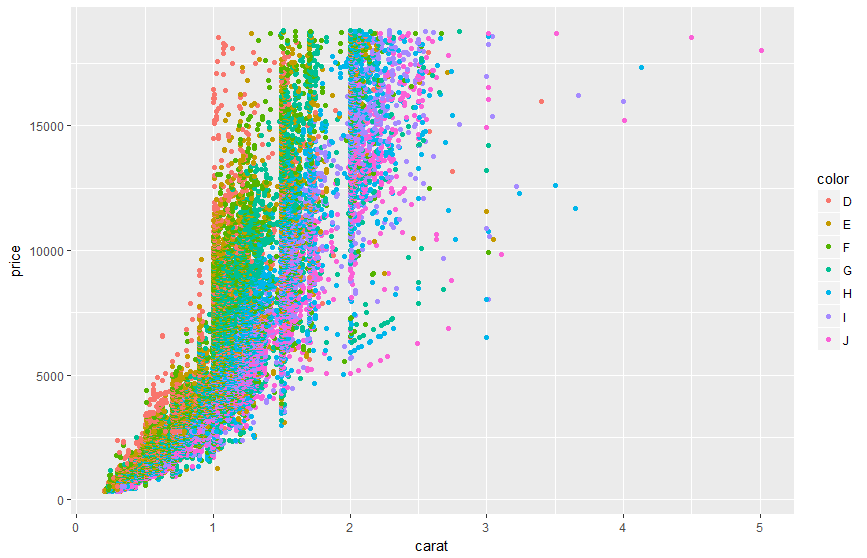
**Using the sample Diamonds dataset in R Studio, perform the following operations:**

**A histogram of carats using ggplot2. Explain your findings.**



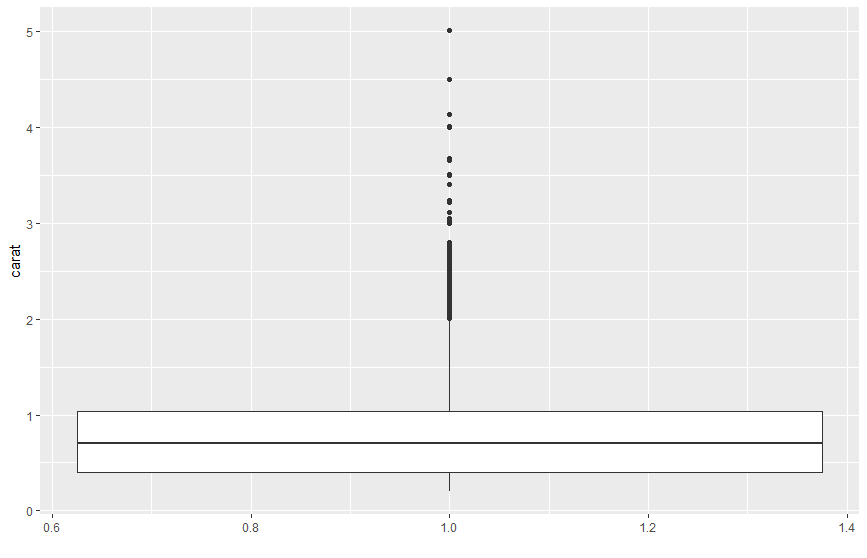
The distribution of carats is right skewed. The mean carat is 0.7979397, the median carat is .7, and standard deviation is 0.4740112.

**A scatterplot mapping carats (x) and price (y) using ggplot2. Use the color aesthetic argument. Explain your findings.**



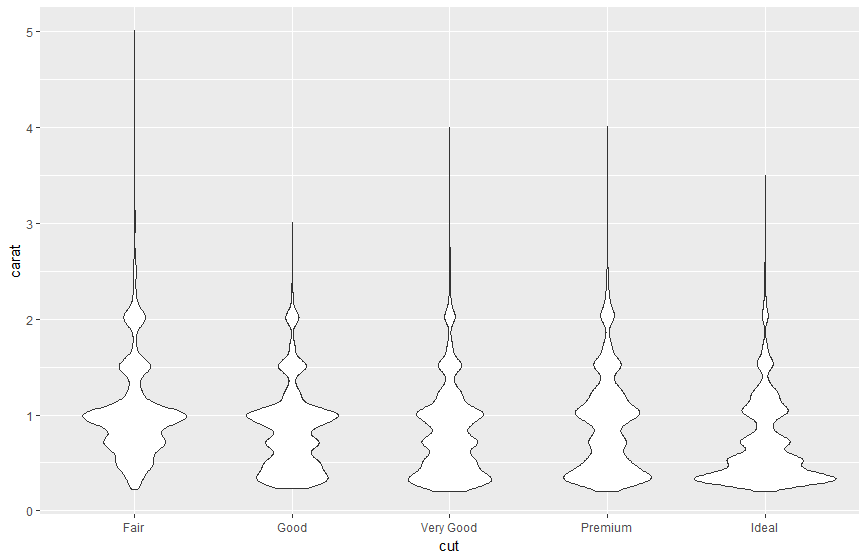
There appears to be a positive relationship between carats and price. As the number of carats increases the price rapidly increases. Also, the color of the diamond seems to have an impact on number of carats. The higher in the alphabet which is represented by cooler colors, then the higher the number of carats. In general, the warmer colors which represent lower letter in the alphabet appear to have fewer carats. The color is spread across the price. The spread may be variation in clarity.

**A boxplot of diamond carats using ggplot2. Explain your findings.**



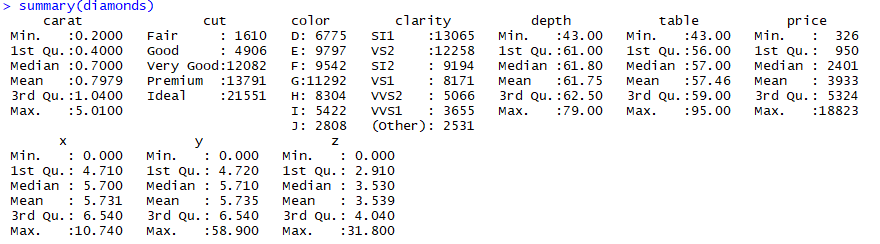
There are a lot of outliers in the number of carats. This is more evidence of the right skewness of the distribution of carats. The median carat is 0.7.

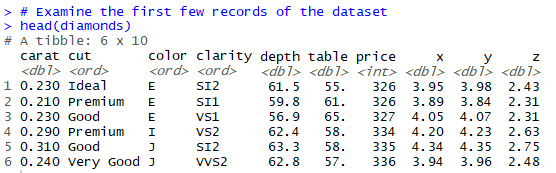
**A violin plot of diamond carats by cut using ggplot2. Explain your findings.**



On the far-left side at the fair cut, it looks like the majority of the carats tend be around 1. While on the far-right side at ideal cut, the carats are mostly less than .5. This may be because more a diamond is cut the lower its carats or mass.

**Using summary() and head() functions, provide an overview of the dataset.**

****

****