R- Assignment 5

Text-

1. What are two different ways to detect outliers in your data?

When plotting your data with large y values, you can use coord\_cartesian() to zoom in and see data point outliers with small y values. You can also use xlim() and ylim() functions to find and throw away data outside the limit.

R Script-

1. Using the nycflights13::flights data, describe the distribution of arr\_delay:
   1. What are the min and max values of arr\_delay?
   2. Create a histogram graph of the values in arr\_delay.
   3. Create a histogram graph of the values in arr\_delay that are greater than 300.
   4. What are the cutoff values that split the values of arr\_delay into deciles (10 bins with an equal number of observations in each bin)? *Hint: The quantile() function is one possible way to do this.*
   5. Describe in words what these things teach you about the distribution of arr\_delay.
2. Use the forcats::gss\_cat data:
   1. Collapse the categories of *marital* so that all the values are either Currently Married (for people marked as “Married”), Not Married (for those who did answer, but are not married) or No Answer (for respondents who submitted “No Answer”).
   2. Create a summary table that calculates the number of people identifying as Currently Married, Not Married, or No Answer for each year.
   3. Create line plots to show how the proportions of people identifying as Currently Married, Not Married, or No Answer have changed over time.