* Choice of package (dplyr): Again The \*\*`dplyr`\*\* package is chosen again because it provides efficient functions to help perform data manipulation tasks. It also offers functions that allow the code to express data transformation steps in a clear and readable manner. In this \*\*`summarise()`\*\*, \*\*`count`\*\*, \*\*`arrange`\*\*, \*\*`group\_by`\*\* is used.
* Reading the data: I used the read.csv function again to read the sales\_ug.csv the sales\_data and store\_cities.csv into store\_cities data frames
* Merging the dataframes: Using the merge() function, I merge the ‘sales’ dataframe with the ‘store\_cities’ dataframe using the common column of ‘store\_id’ . This allowed me to join the data frame to be together in one data frame named merged\_data.
* Calculating most common storetype: Here I used the count() function to count the unique values of the store types and add them together to allow me to find the most common store type and used the arrange() function to place the counted values into a descending list allowing me to find the most common store types as the top 2 values will me the most common store type. This is all then put into the most\_common\_storetypes data frame for further analysis later.
* Calculating sales volume: Here I take the merged\_data data frame a apply the group\_by(storetype\_id) function as it groups the data by store type for further calculations. Then I use the summarise(total\_sales = sum(sales)) to collapse the all the store types into one and add the sales together into one, resulting in a data frame which contains each store type and the amount of sales they each have, This will allow me to compare the sales volume between the 2 most common store types as the question requires. Then this calculations have been stored within the sales\_volume data frame.
* Calculating total revenue: For calculating total revenue, again I take the merged data frame and apply the group\_by(storetype\_id) function, this will allow me to group the data by store type to calculate for each store type. Using the summarise(total\_revenue = sum(revenue)) function again but this time I have used the revenue column, which allows me to calculate the total revenue each store has made. This all goes into a data frame called revenue
* Printing calculations: The print() function is used again to print out the resulting data frames calculated from before. I have printed the most\_common\_storetypes, sales\_volume and the revenue data frames, This allows me to further analyse the data frames produced by the calculations.

For revenue for each store type

* Choice of package (dplyr and ggplot2): Again the Dplyr is used as it provides functions like summarise() and ggplot’s inclusion here is so I can plot the results in a bar graph as I believe it would be easier to visualize if there was a relationship between the store size and its revenue produced
* Reading the data: I used the read.csv() function again to read the sales\_ug.csv and store\_cities.csv into data frames.
* Merging the dataframes: Using the merge() function, I merge the ‘sales’ dataframe with the ‘store\_cities’ dataframe using the common column of ‘store\_id’ . This allowed for the necessary information to be joined together in one data frame named merged\_data.
* Calculating Revenue based on store size: I take the dataframe of the merged csv files and group the data frame with the function `group\_by(store\_size)’ this groups the dataframe by the store size numbers and allows for calculations to be done for each store size. The `summarise()` function is then applied to each group within the group, in which I then calculate the sum (`sum`) of the revenue for each store type. This all results in a dataframe in which all the store types are listed with the revenue they made on the side
* Plotting the data: Here I utilized the ggplot2 package and plot the data frame so it will be easier to analyse. I have made the x-axis the store ID and the y-axis the total revenue as it allows for the bars to go in an upwards direction, helps me determine if there was a relationship between the store size and the revenue it made. This plot would grant me the means of determining if a relationship has formed.