Question 1a

* Choice of package (dplyr): The use of the dplyr package is chosen because it provides efficient functions to help perform data manipulation task. It provides functions such as summarise(), which provides a way to combine values into single rows and group\_by() which groups data depending on the variables you give it. This allows for calculations to be done on big sets of data.
* Reading the sales data: To read in the CSV files I have made use of the read.csv() function is used to read the data from a CSV file named sales\_ug.csv. This is then stored in a data frame named sales\_data, this was named sales\_data as it contains all the sales data that was read from the sales\_ug.csv file.
* Computing the Daily Revenue: To compute the daily revenue for each store I have used the dplyr package and have taken the data frame of sales\_data, which was the data from the csv file and I applied the group\_by(store\_id, date) function. I used to function to group the data frame into groups of store id and date, in which I intend to combine as I am trying to calculate each store’s daily revenue. Then I apply the summarise (total\_revenue = sum(revenue)) function, with this I intend to combine the grouped variables from before and add the revenue values together into a new variable named total\_revenue. The result of this is a data frame in which I combine the dates and revenue for each store.
* Printing the daily revenue: To print out the data frame daily\_revenue, I utilised the function of print(daily\_revenue), This allows me to print out the data frame and display it for further analysis of each day.

Question 1b

* Choice of package (dplyr and ggplot2): I utilized the dplyr package again as it provides functions for data manipulation. In this code I used the group\_by() and the summarise() function again. The use of ggplot2 is due to the question asking for a plot of the weekly revenue for each store and I intend to use this package for functions that allow for a easier time to plot data.
* Reading sales data: To read in the CSV files I have made use of the read.csv() function is used to read the data from a CSV file named sales\_ug.csv. This is then stored in a data frame named sales\_data, this was named sales\_data as it contains all the sales data that was read from the sales\_ug.csv file.
* Computing Weekly Revenue: To calculate the weekly revenue of each store, I made use of the dplyr package, I have taken the sales\_data data frame, and grouped it using the group\_by(store\_id), this function groups the data frame by store id as I wanted to combine the revenue of each store over each date. Then I make use of the summarise(total\_revenue = sum(revenue)) function again, this allows me to combine the revenue of each store over the week into one.
* Plotting the data: This is where I utilise ggplot2 package to plot the data that I calculated from the code before. This is where I used the function ggplot(total\_revenue, aes(x=store\_id, y=total\_revenue, fill=store\_id)) I decided on a bar graph (geom\_bar(stat="identity")) as it provides the easiest way to compare the different stores’ revenue over the week. I made the x-axis the store Id and the y-axis the revenue as it allows for the bar graph to go up, which allows me to compare the stores’ revenue easier. I decided on removing the legend (theme(legend.position="none")) as it took up a lot of space from the plot and the x-axis is already showing each store.