UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

(FSKTM)

SEMESTER II 2024/2025

DATA MINING

BIT 33603

SECTION 03

LAB ASSIGNMENT 03

TITLE

DATA VISUALIZATION USING R

LECTURER'S NAME

DR. ROZITA BINTI ABDUL JALIL

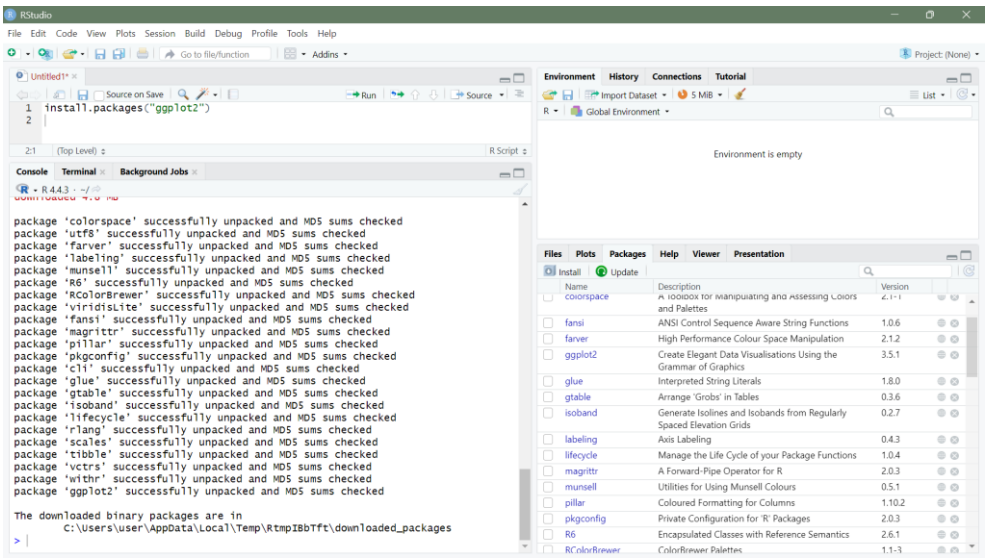
NAME	TUAN KHALIDAH SYAZWANA BINTI TUAN MOHD KASMAWI
MATRIC NUMBER	AI220118
DATE SUBMISSION	April 08, 2025

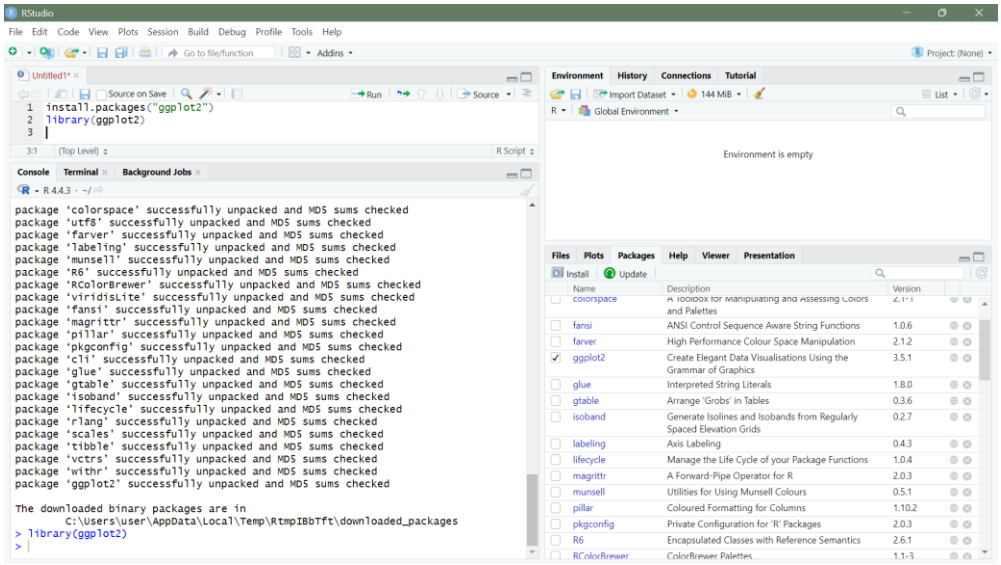
Assessment Question:

1. Run the provided code in R (Activity 1-7) and analyse the output.
2. Submit the visualizations as image snapshots along with a brief explanation of the insights gained.

Part 1: Introduction to ggplot2

1. Installation and Loading of Packages:
2. Understanding the Basic Structure of ggplot2:



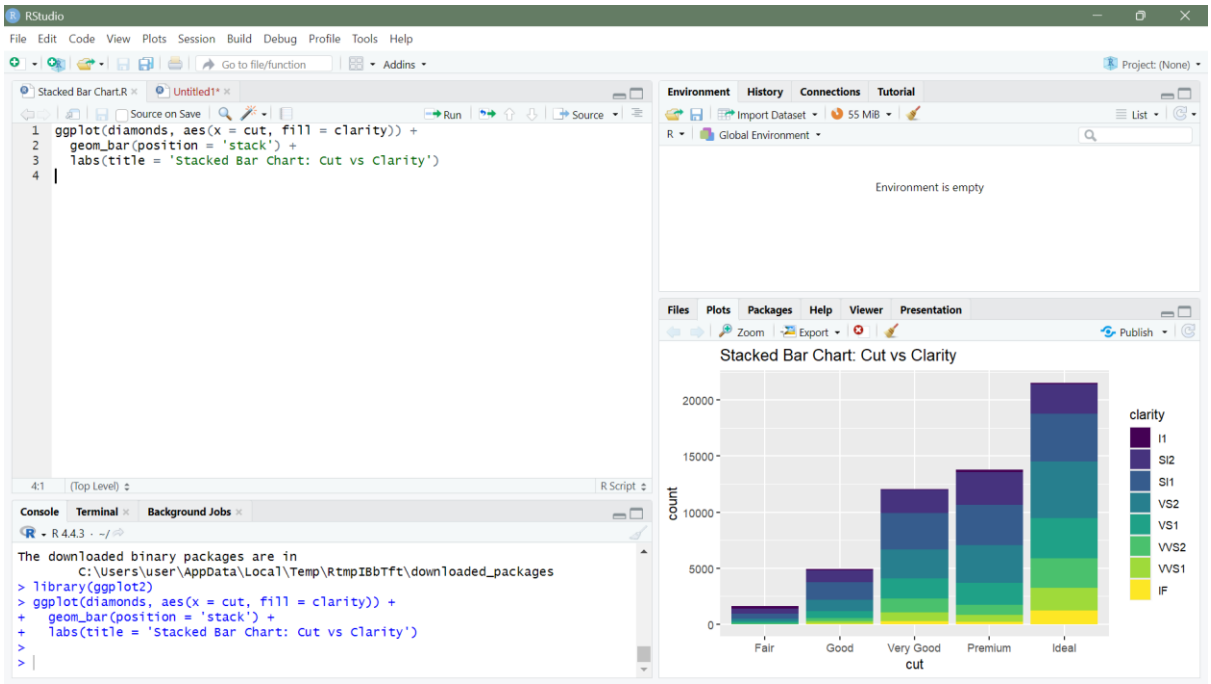


Part 2: Creating Various Types of Visualizations

Activity 1: Bar Chart (Stacked and Vertical Bar Chart)

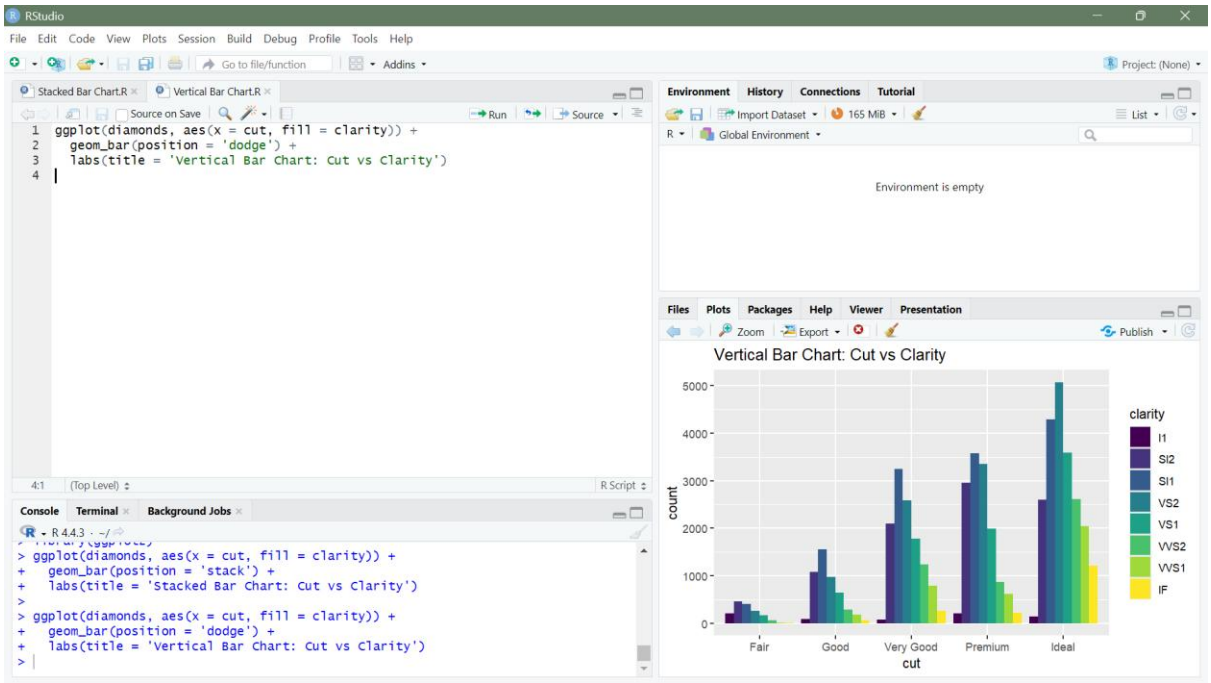
1.1 Stacked Bar Chart

Load the ggplot2 library and use the diamonds dataset to create a stacked bar chart. The x-axis should represent the cut of the diamond, and the bars should be filled based on clarity.



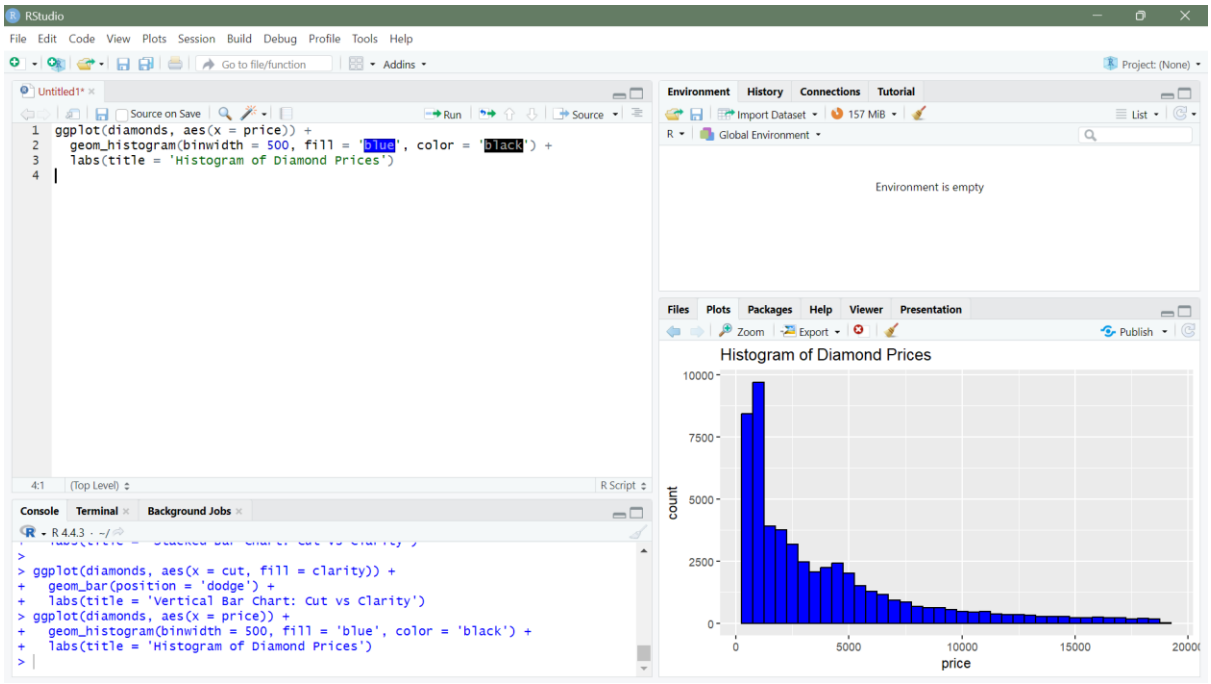
1.2 Vertical Bar Chart

Create a vertical bar chart using the same dataset. This time, use position='dodge' to separate the bars based on clarity.



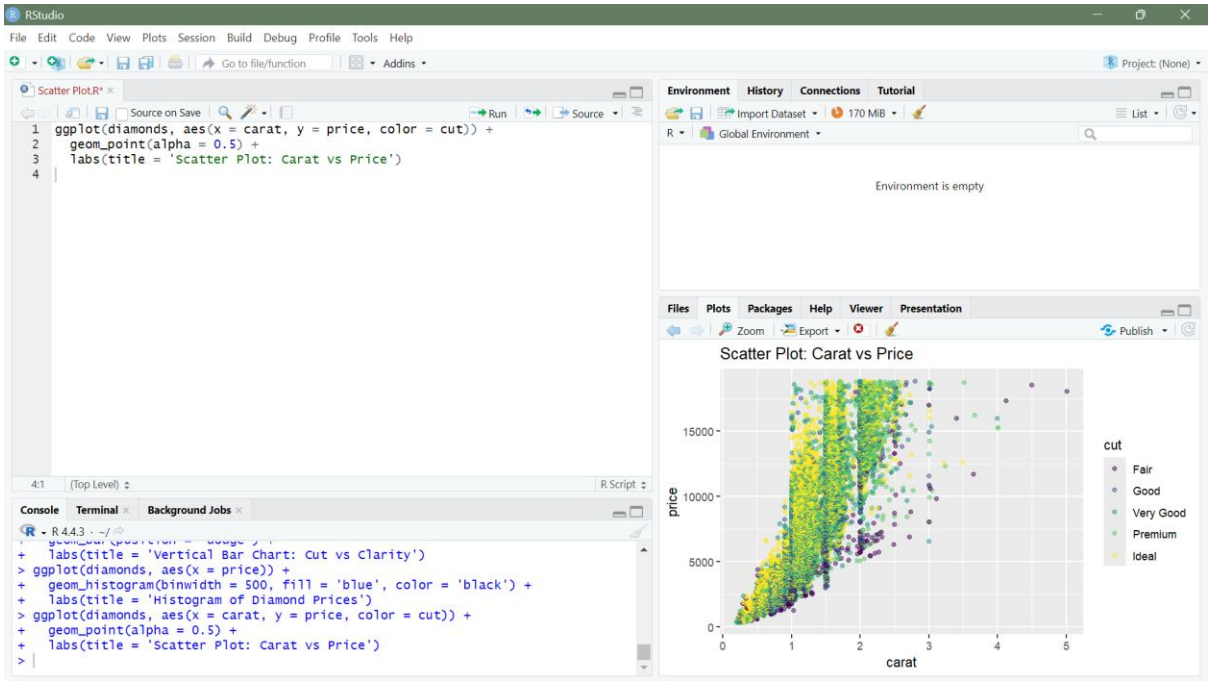
Activity 2: Histogram

Create a histogram to visualize the distribution of diamond prices using the diamonds dataset. Set the bin width to 500 and color the bars in blue.



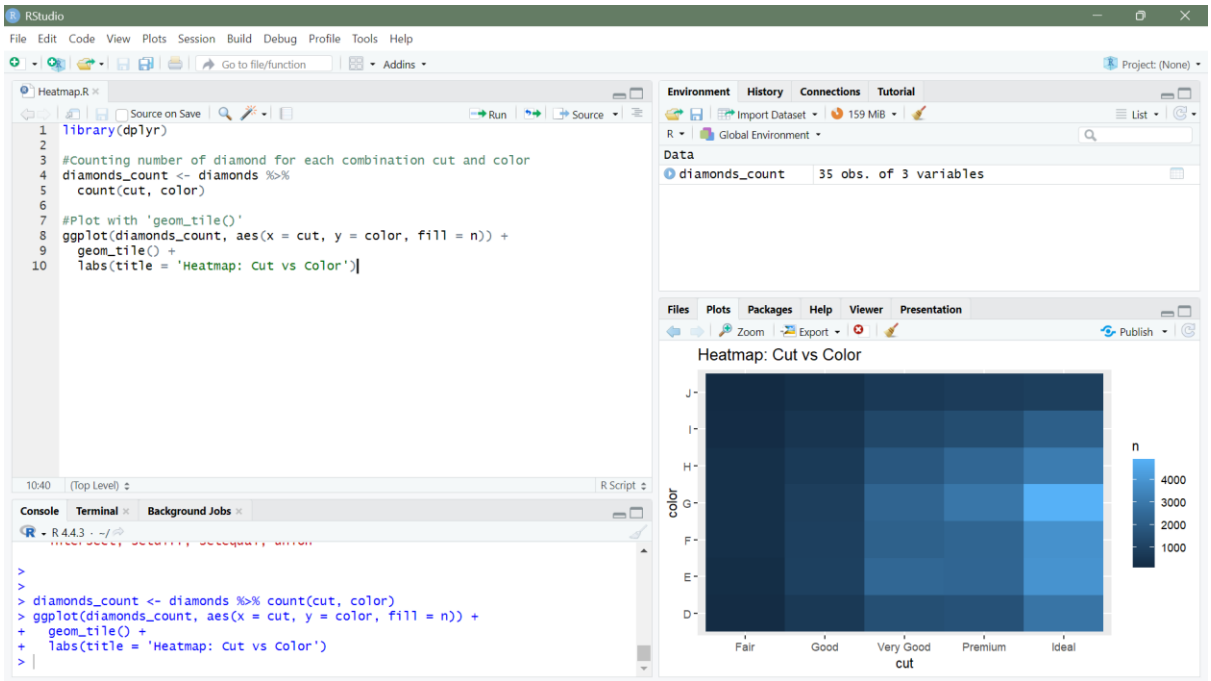
Activity 3: Scatter Plot

Create a scatter plot using the diamonds dataset. Plot the carat on the x-axis and the price on the y-axis. Color the points based on the cut of the diamond.



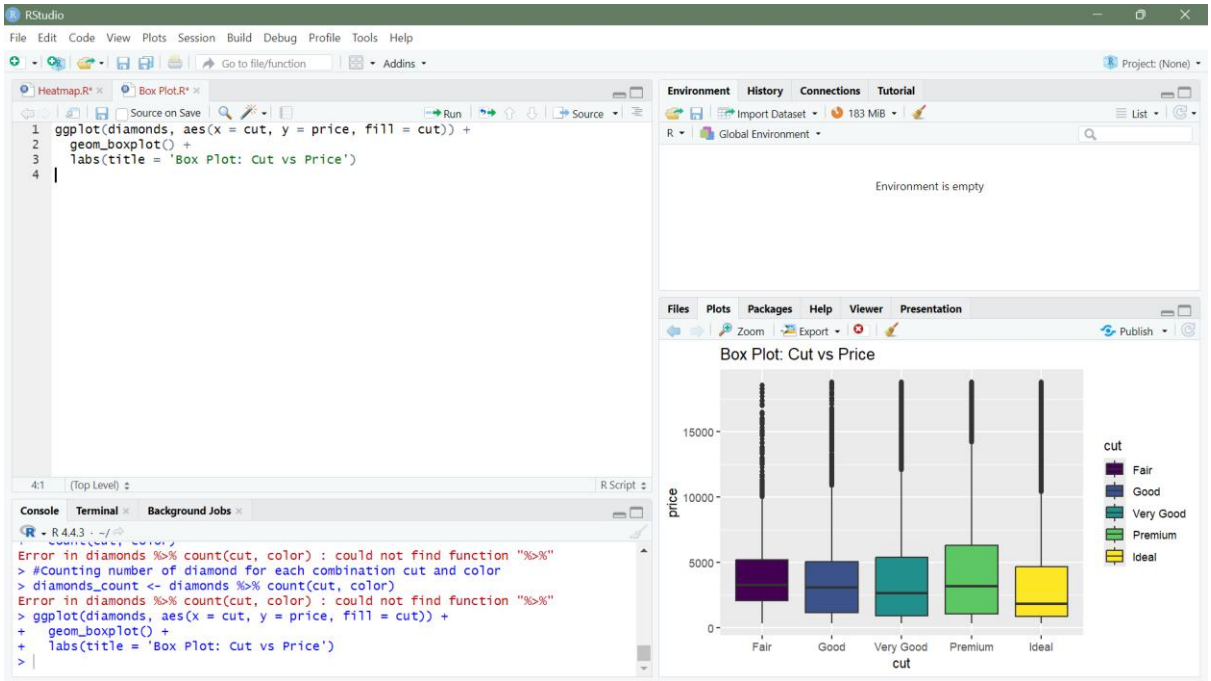
Activity 4: Heatmap

Create a heatmap to visualize the relationship between cut and color using the diamonds dataset. Fill the heatmap based on the count of diamonds for each combination of cut and color.



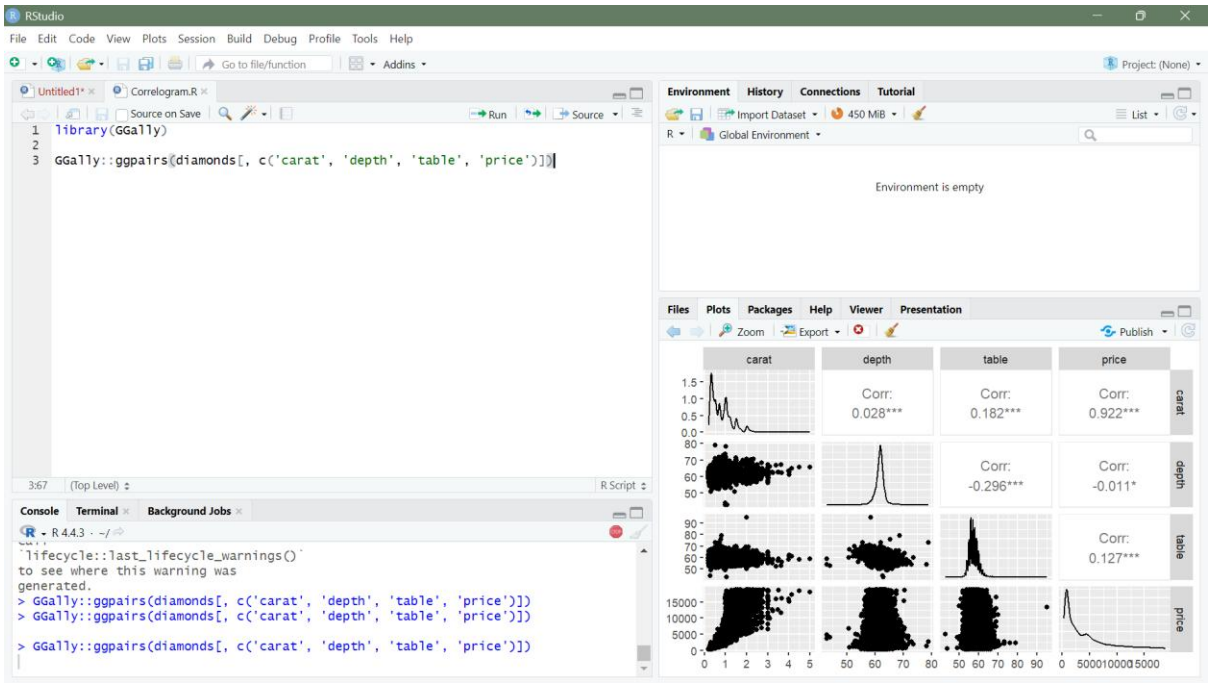
Activity 5: Box Plot

Create a box plot to compare the price distributions across different cut types in the diamonds dataset.



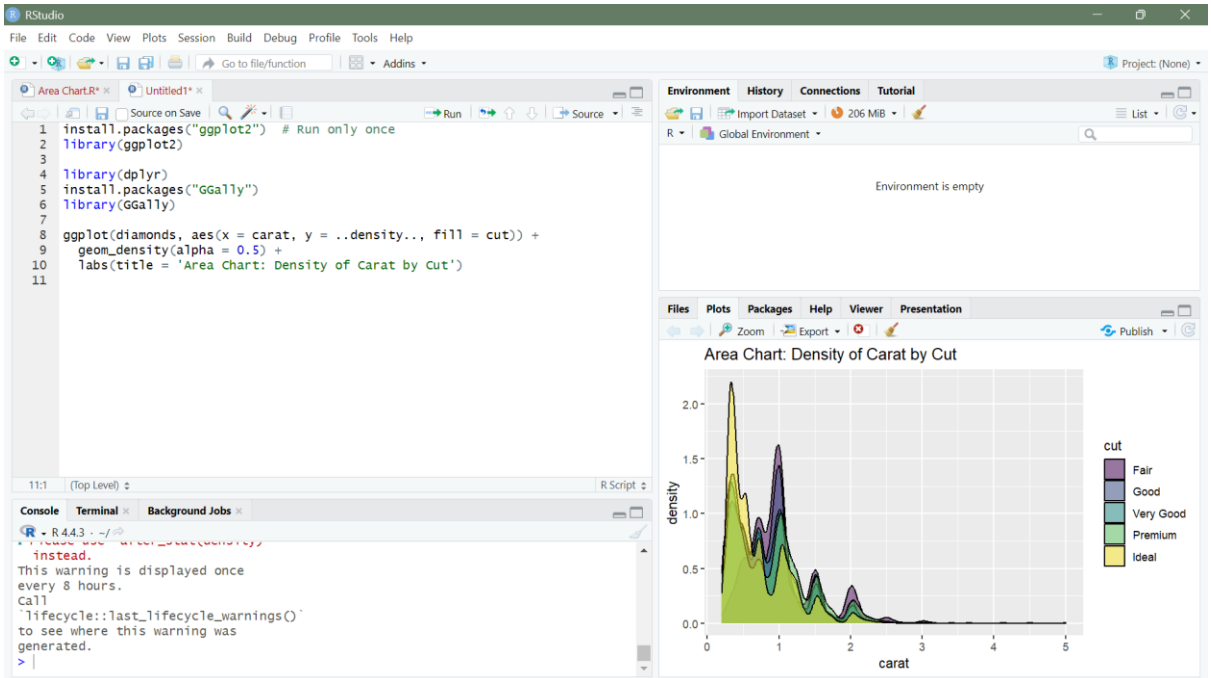
Activity 6: Correlogram

Use the GGally package to create a correlogram (pair plot) of several variables in the diamonds dataset, including carat, depth, table, and price.



Activity 7: Area Chart

Create an area chart to visualize the density of carat values by cut type in the diamonds dataset.



Questions to Test Understanding (Self-studies)

1. Based on the scatter plot of carat vs price, what trend can be observed? How does the cut of the diamond affect this trend?

As the carat (size) of the diamond increases, the price also increases. Diamonds with better cuts tend to have higher prices even if the carat is the same.

2. In the box plot visualization, what does the spread of price values indicate about the different cut types?

The spread (or height of the box) shows the price range. Some cut types like “Ideal” have a tighter (narrower) price range, while others like “Fair” show more variation in prices.