Text, logo

Description automatically generated

Exploration of Student Alcohol Consumption Dataset

Data Processing Using R- Spring 2022

Submitted By- Kushal Ahuja

Table of Contents

[1. Introduction 3](#_Toc119966847)

[2. Dataset 3](#_Toc119966848)

[3. Body of the report 4](#_Toc119966849)

[3.1 Business Questions- 4](#_Toc119966850)

[3.2 User Interface (UI) 5](#_Toc119966851)

[3.3 Main Panel 6](#_Toc119966852)

[3.4 Server 6](#_Toc119966853)

[3.5 Interactive Tool 7](#_Toc119966854)

[4. Results and Discussions 8](#_Toc119966855)

[5. Conclusions 12](#_Toc119966856)

[6. References 13](#_Toc119966857)

# **Introduction**

This report is to analyse the student’s alcohol consumption; people are worried about the current student's consumption of alcohol but is backed without proof of distribution of consumption among the students. The dataset is cleaned and then deployed to R shiny for making the visualisation and interpreting the consumption in an efficient way.

The data was gathered through a survey of secondary school students taking mathematics and Portuguese language classes.

# **Dataset**

The dataset contains the data from the survey given by students, the attributes in the dataset are as below-

1. school - Student's school (binary)
2. sex - Student's gender (binary: 'F' - female or 'M' - male)
3. age - Student's age (numeric: 15 to 22)
4. address - Student's home address (binary: 'U' - urban or 'R' - rural)
5. famsize – Family’s size (binary)
6. Pstatus - Parent's cohabitation status (binary)
7. guardian – Guardian of the student (nominal: 'mother', 'father' or 'other')
8. studytime - weekly study time (numeric)
9. failures - number of past class failures (numeric: n if 1<=n<3, else 4)
10. schoolsup - extra educational support (binary)
11. famsup - family educational support (binary: yes or no)
12. activities - extra-curricular activities (binary)
13. higher - wants to take higher education (binary)
14. romantic – whether a student is in a romantic relationship (binary)
15. famrel – the quality of relationships with the family (numeric: from 1 - very bad to 5 - excellent)
16. freetime - free time after the school (numeric)
17. goout - going out with the friends (numeric)
18. Dalc - alcohol consumption on workdays(numeric: from 1 - very low to 5 - very high)
19. Walc - alcohol consumption on weekend (numeric)
20. health - health status of student (numeric)
21. absences - number of school absences (numeric)

These grades are related to the course subject:

1. G1 - First-period grade (numeric: from 0 to 20)
2. G2 - Second period grade (numeric: from 0 to 20)
3. G3 - Final grade (numeric: from 0 to 20, output target)

(Student Alcohol Consumption, 2016)

To solve the business question, I have chosen the following attributes.

* Famsup
* Dalc
* Sex
* Famrel
* Grade

# **Body of the report**

By analysing the dataset, it was not clear how the chosen attribute affects the consumption of alcohol, so as to investigate the attribute, the Rshiny is used and provide the relevant data-driven insights.

## **3.1 Business Questions-**

To know the attribute's dependence on consumption, many questions can be answered**,**

* Consumption of Alcohol by Students who have educational support vs those who don’t have educational support.
* Consumption of alcohol by students when in grades 1 and 2 and also in the final grade.
* Consumption of alcohol on the basis of Student’s gender.
* Consumption of alcohol by students is based on the quality of their relationship with their families.

## **3.2 User Interface (UI)**

User Interface in R Shiny acts like an interactive tool to choose the values from the data and interpret the changes in visualisation at the same time. This enables the user to get insights from the multiple attributes of the data in the same visualisation graph. It eliminates the process of creating another code or the visualisation of other attributes.

The user interface has three parts in this r shiny application.

1. sliderInput
2. radioButtons
3. selectInput

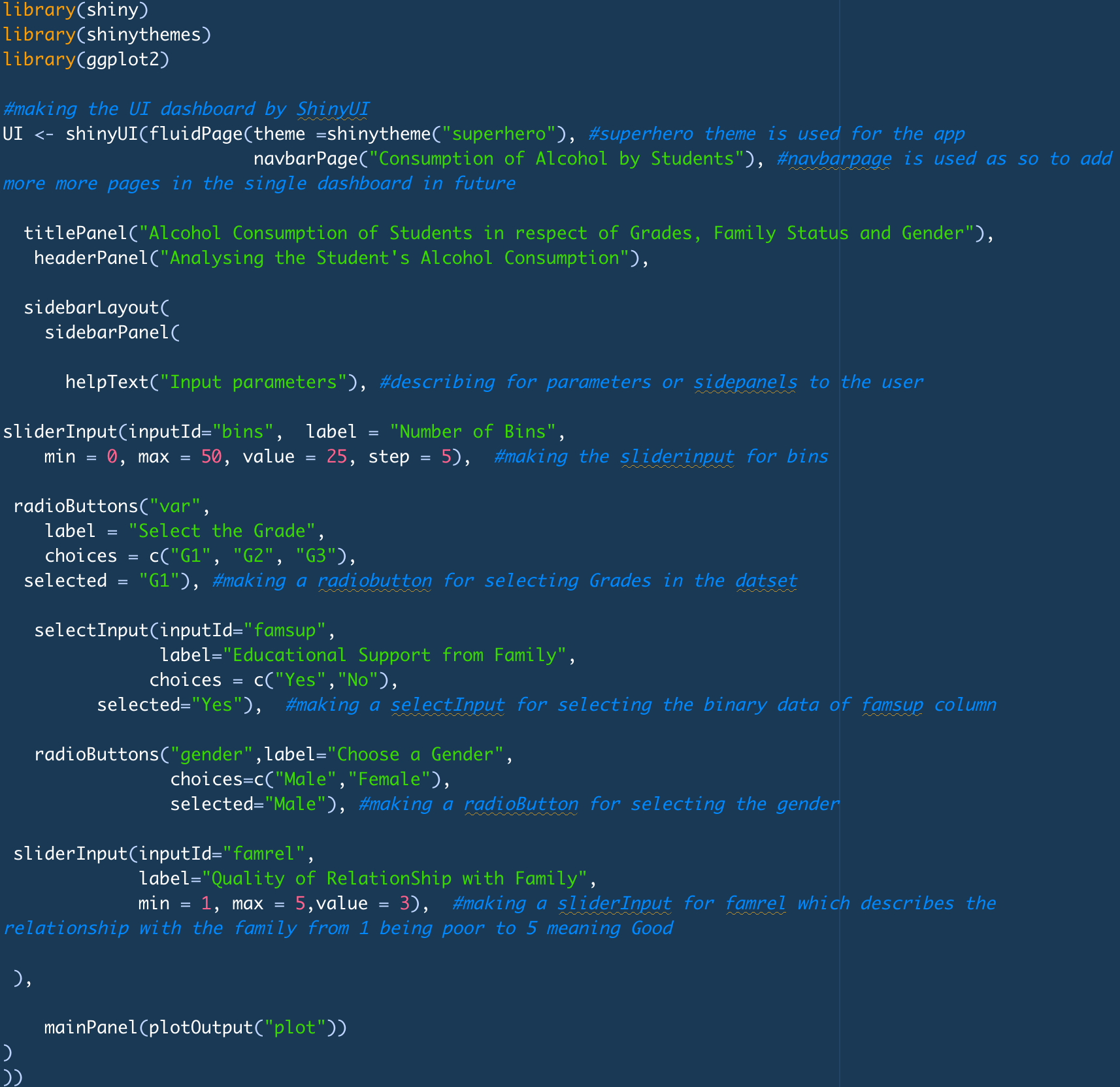
* sliderInput is used for defining the bin size of the plot and making the famrel attribute data i.e. numerical data to be selected with the slider panel.
* radioButtons are used to define the attribute named sex and the grade.
* selectInput is used to define famsup attribute data, i.e. binary data to be selected with the two choices.

Figure 1- User Interface

## **3.3 Main Panel**

The main panel here is used to create the output using the plotOutput function.

## **3.4 Server**

Server in R Shiny is a backend program that builds the web application according to the data and connects to the UI with the use of Shiny. The shiny server is also made to make the application available over the internet. The plot is defined with the data in the server panel, thus making the visualisation interactive and collaborative.

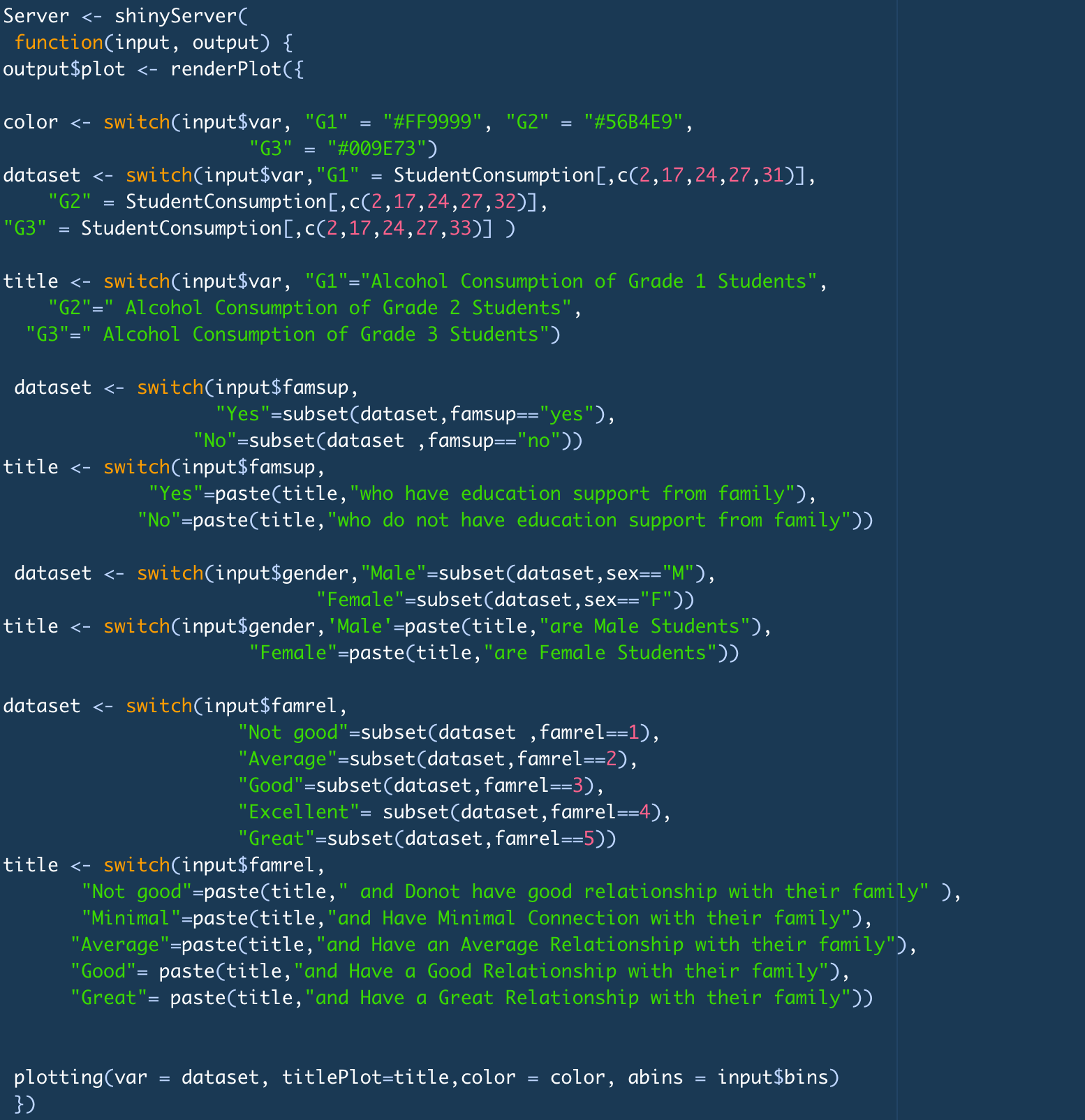
The server in this application is given below-

Figure 2- Server

## **3.5 Interactive Tool**

The interactive visualisation tool plays a significant role in the R Shiny and can be made by the User interface panel.

1. Various types of input are being added to this interactive tool, like sliderInput, radioButton and selectInput.

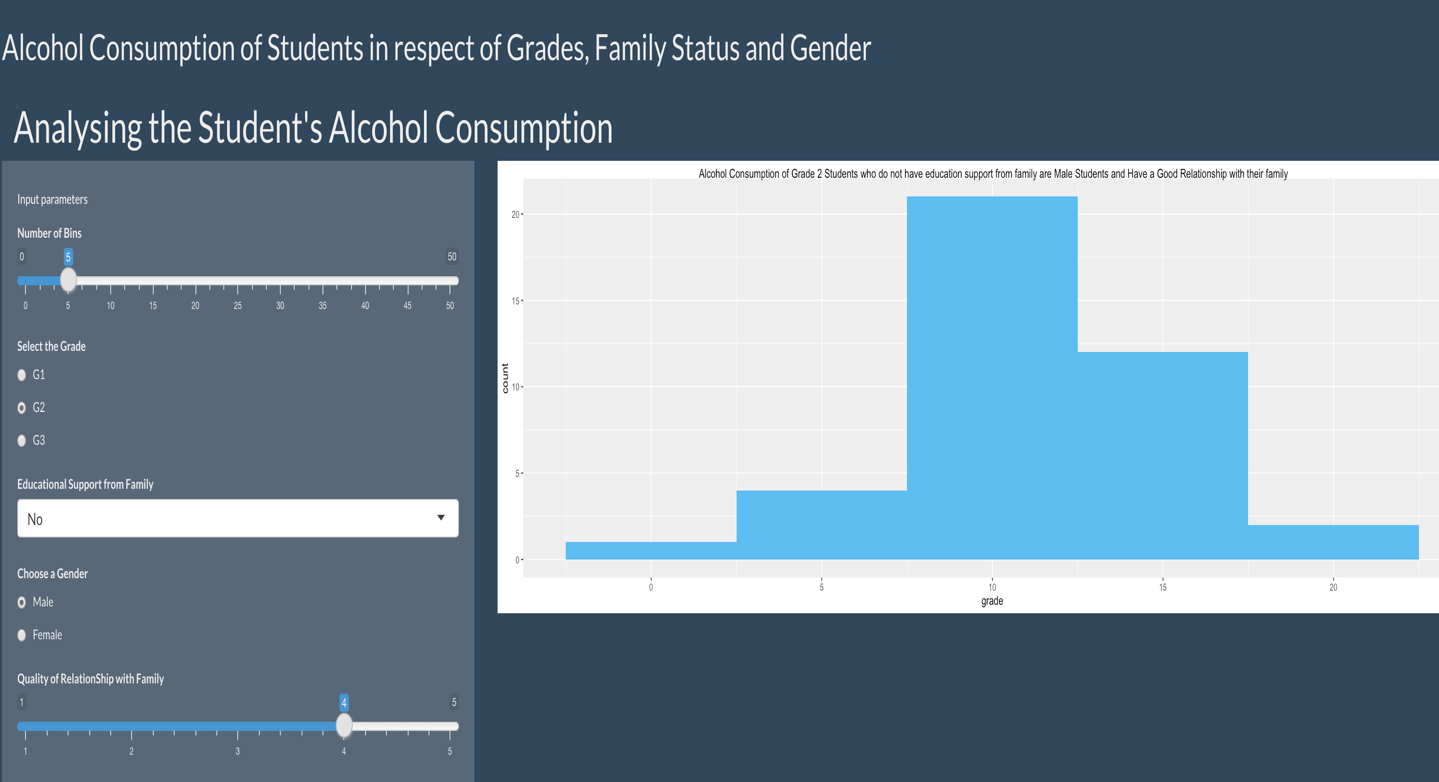
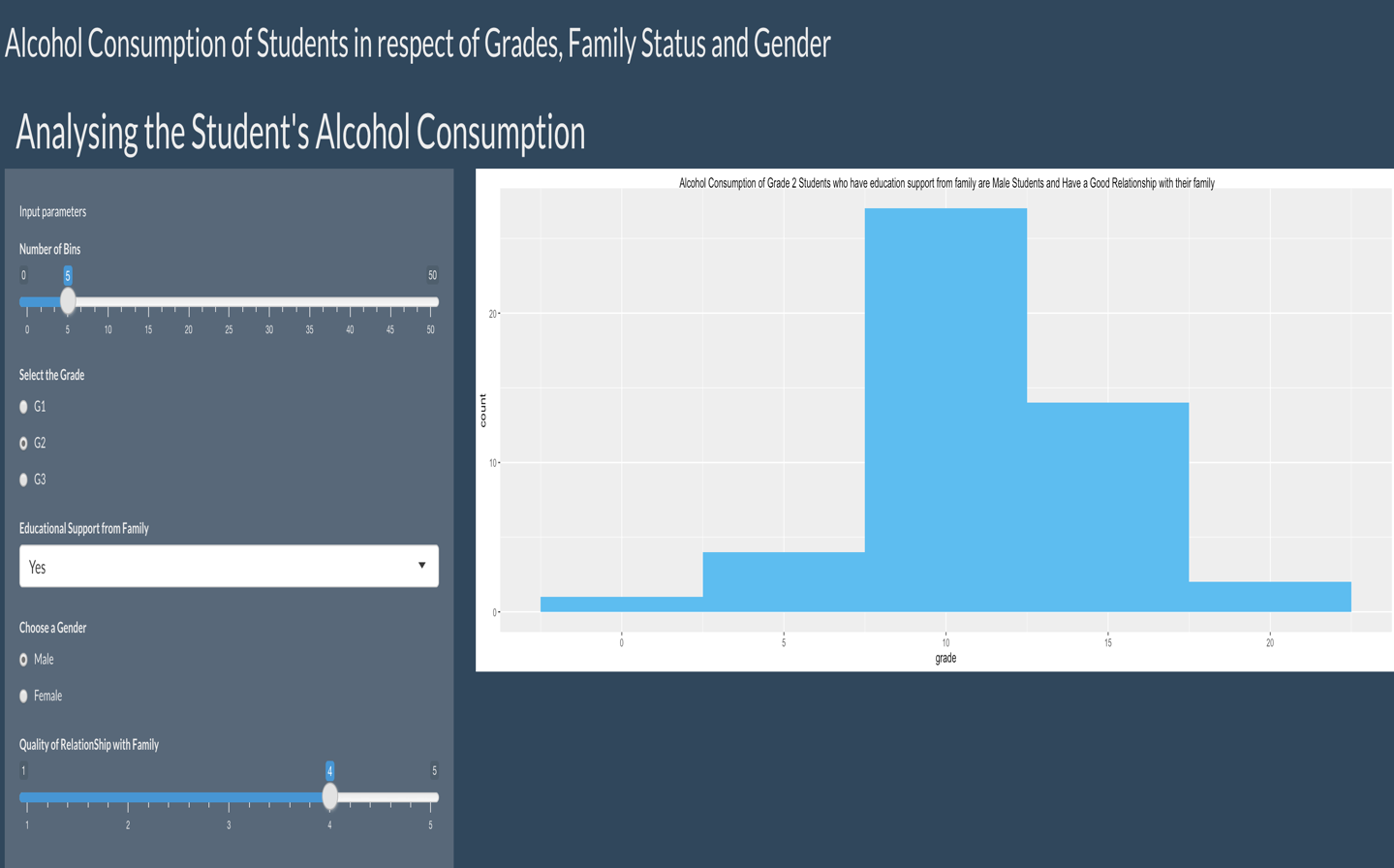
* Grade is the RadioButton
* Famsup is a selectInput
* Gender is a radioButton
* Famrel is a sliderInput

**Timeline

Description automatically generated**

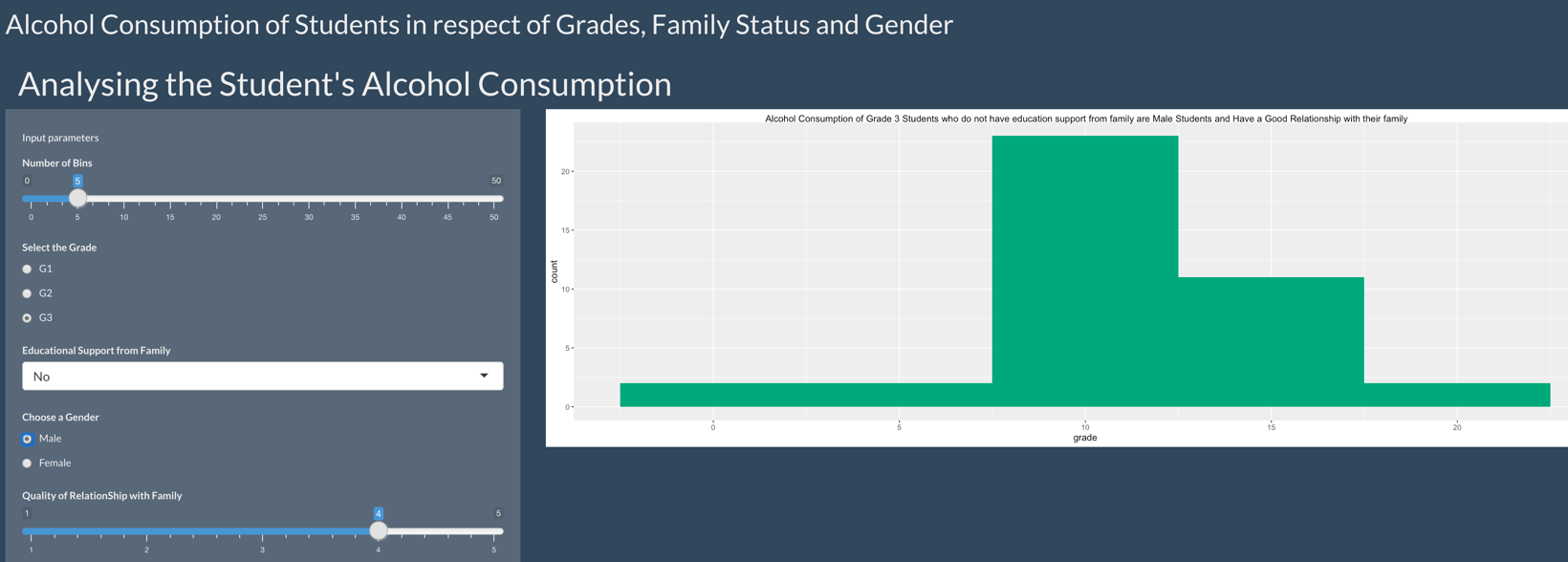
Figure 3- Interactive Tool

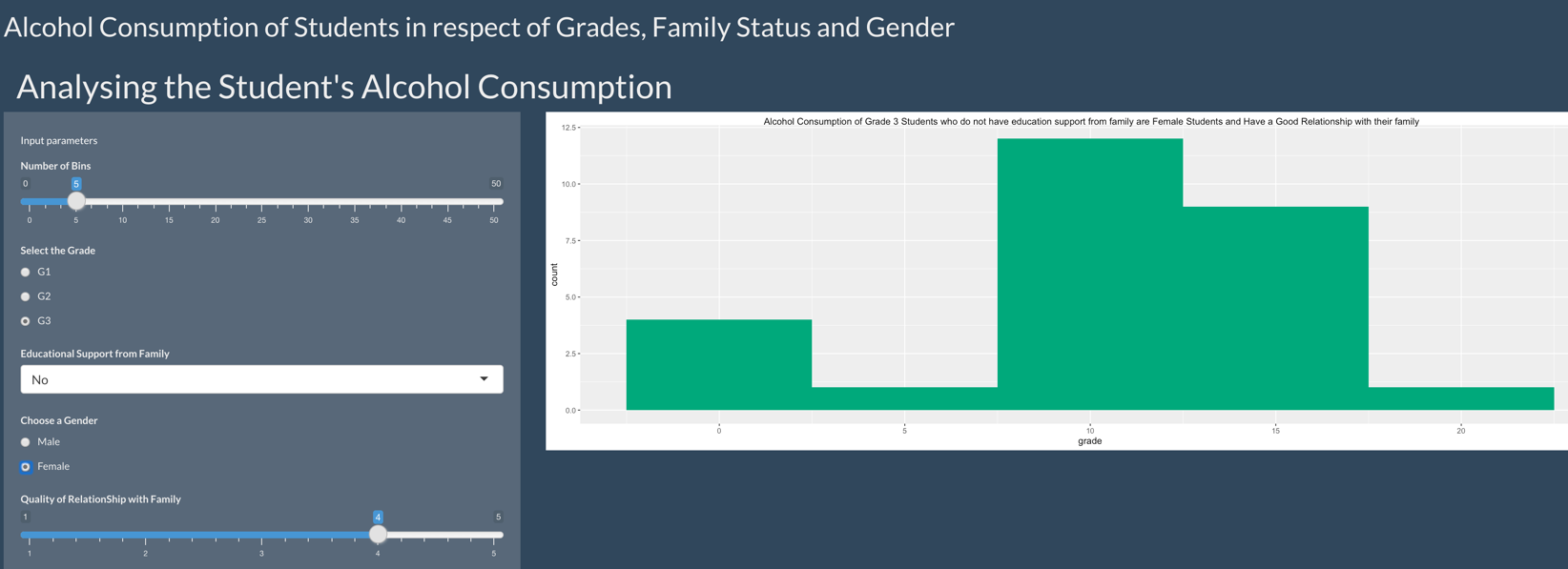
# **Results and Discussions**

**Consumption of Alcohol by Students who have educational support vs those who don’t have educational support.**

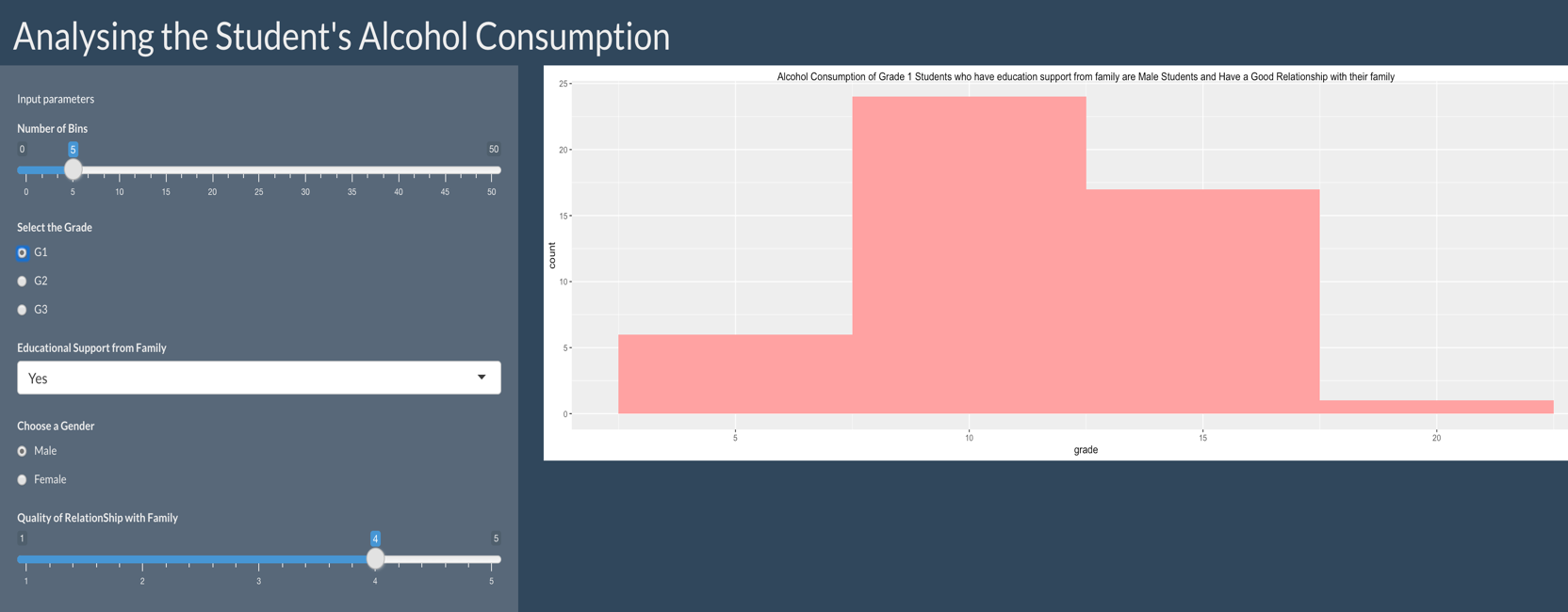
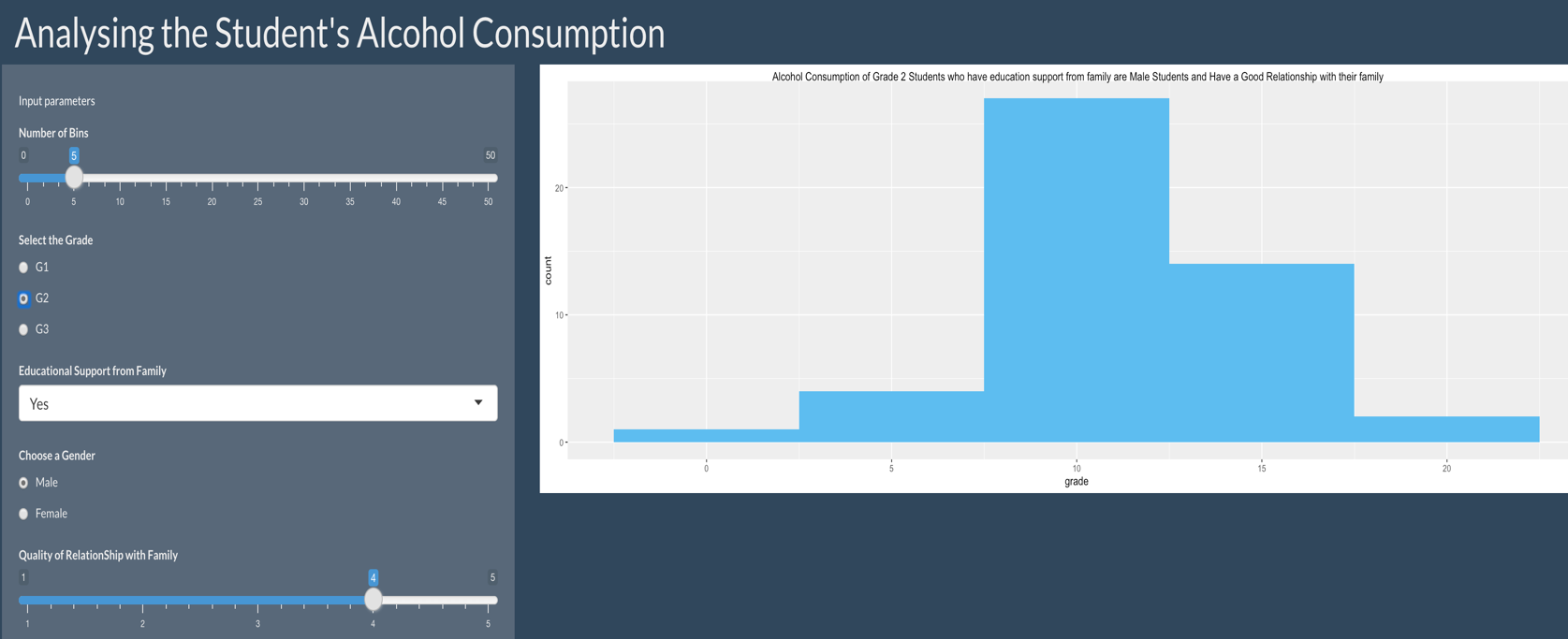
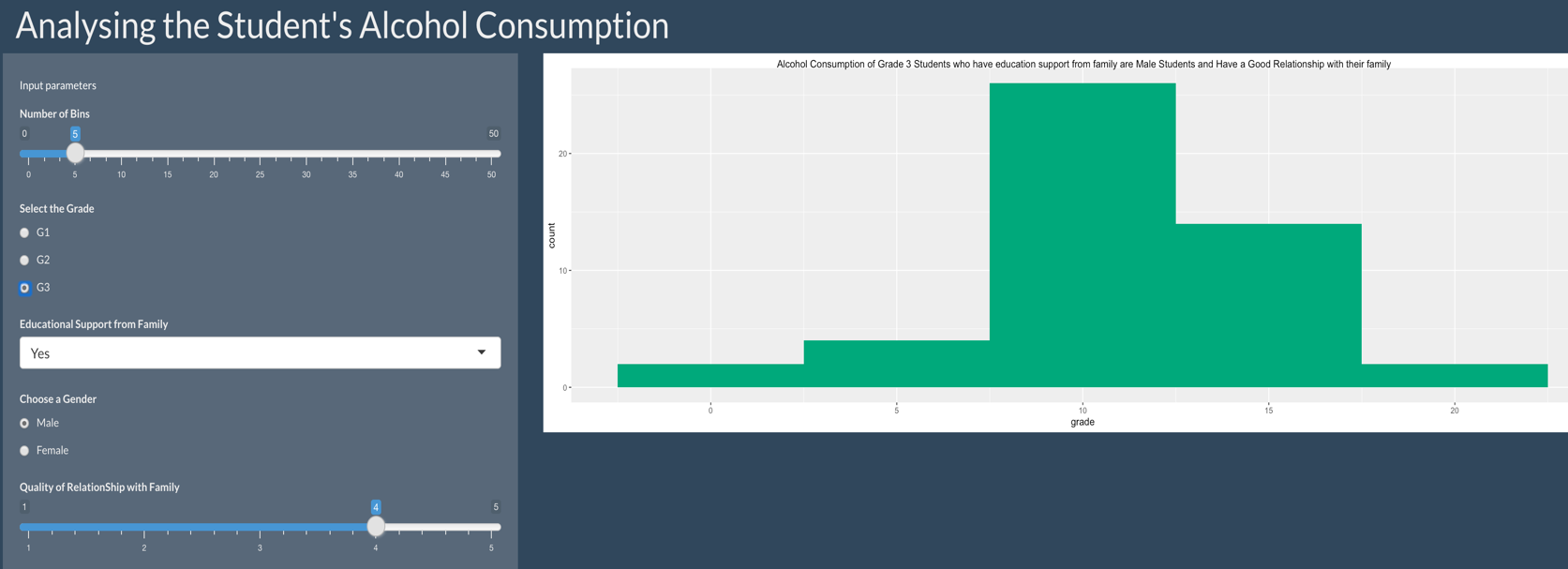
The above figure shows the difference of the consumption of alcohol by students who have educational support vs who don’t have. It is evident from the graph that there is minimal difference , leading us the insight that this attribute is not significant for deciding the alcohol consumption.

**Consumption of alcohol on the basis of Student’s gender**





The above figure shows the difference in the consumption of alcohol by students on the basis of their gender; it is evident from the graph that male students tend to consume more alcohol than female students, as demonstrated by the y-axis (count). This results in the valuable insight.

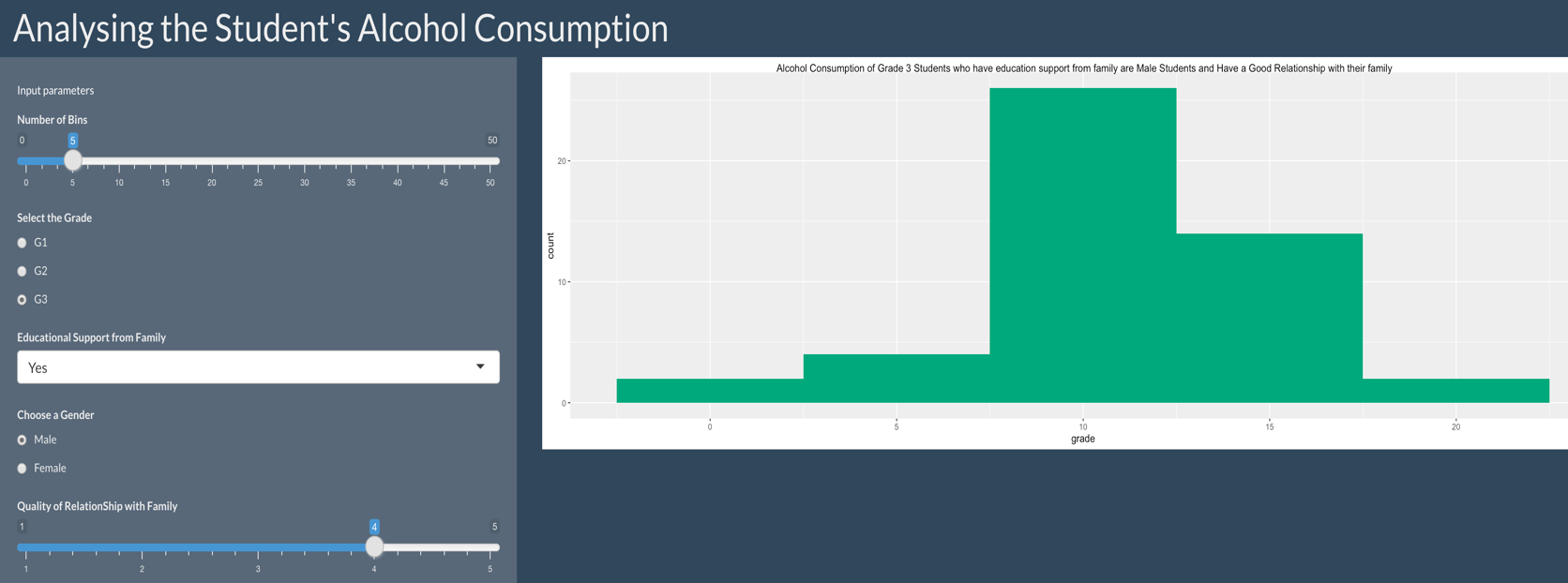
**Consumption of alcohol by students when in grades 1 and 2 and also in the final grade.**

The above figure shows the difference in the consumption of alcohol by students based on the grade of the first and second periods and the final period.

* G1 denoting the first grade
* G2 denoting the second grade
* G3 denoting the final grade

It is evident from the visualisation chart that alcohol consumption is little higher in grade 1 compared to the other grades.

Graphical user interface

Description automatically generated with low confidence**Consumption of alcohol by students is based on the quality of their relationship with their families.**

The above figure shows the difference in the consumption of alcohol by students based on the quality of their relationship with their family; It is evident from the graphs that the students who have a good relationship with their family tend to consume more alcohol than the students who have an average or great relationship with their family.

# **Conclusions**

After using the R Shiny and Data Visualisation techniques, the business questions were answered, and relevant insights were derived, which is helpful to the stakeholder and public while deciding an important attribute related to the student’s alcohol consumption.

The questions answered through this web application includes but not limited to the following

* To know the consumption of alcohol by students in respect of the reason being the educational support by family or not.
* To know the consumption of alcohol by students in respect of the reason having the type of quality of relationship with the family.
* To know the consumption of alcohol by students in respect of their grades in the various session.
* To know the consumption of alcohol by students in respect of the gender status.

Many more reasons can be derived by the use of RShiny application. Thus it is a very important and valuable package in R, which is used by the major organisation when deciding the operations of their company.

# **References**

*Student Alcohol Consumption*. (2016, October 19). Kaggle. https://www.kaggle.com/datasets/uciml/student-alcohol-consumption