Josephs-Proposal

Moiya Josephs

2022-10-18

Data Preparation

```
# load data
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(ggplot2)
library(curl)
library(psych)
```

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Does having a social life predict final grade? I would use t-test to answer this question, comparing the groups that have a social life vs those who do not.

Cases

What are the cases, and how many are there?

There are 649 cases that represent students and their achievements in secondary education in two Portuguese schools.

Data collection

Describe the method of data collection.

The data was collected using school reports and questionnaires.

Type of study

What type of study is this (observational/experiment)?

This is observational study.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

The data was collected by University of Minho and the dataset can be found here: Source: https://archive.ics.uci.edu/ml/datasets/Student+Performance

Dependent Variable

What is the response variable? Is it quantitative or qualitative?

The response variable is the final grade of each student. It is quantitative.

Independent Variable(s)

The independent variable is the variables that describe the students social life. To answer this questions the variables I classify as describing a students social life is their activities, romantic, family relationship, free time, going out and alcohol consumption. They are all qualitative.

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

```
student_mat_csv <- "https://raw.githubusercontent.com/moiyajosephs/Data606-Final/main/student-mat.csv"
student_mat <- read_delim(curl(student_mat_csv),delim = ";")
```

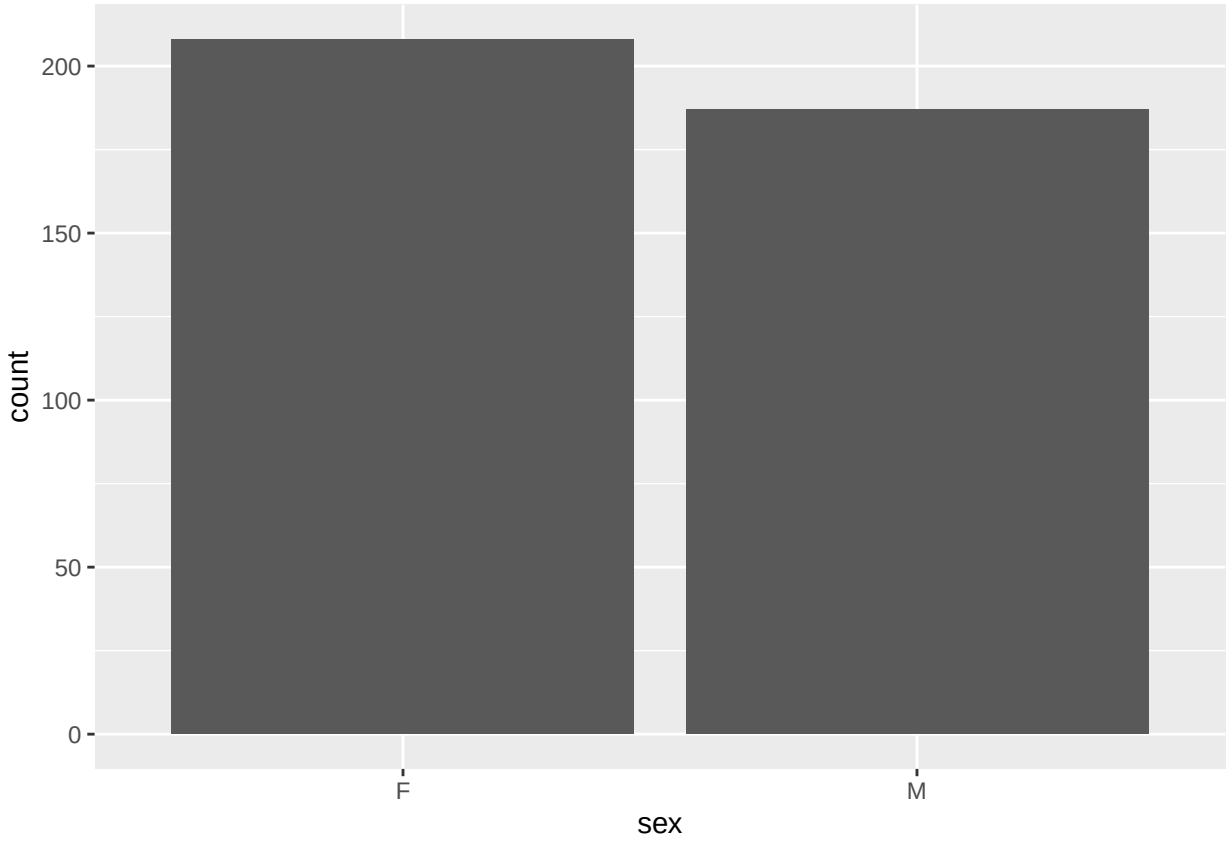
```
## Rows: 395 Columns: 33
## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (17): school, sex, address, famsize, Pstatus, Mjob, Fjob, reason, guardi...
## dbl (16): age, Medu, Fedu, traveltime, studytime, failures, famrel, freetime...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
summary(student_mat)
```

```
##     school              sex                 age         address
##  Length:395         Length:395         Min.   :15.0   Length:395
##  Class :character   Class :character   1st Qu.:16.0   Class :character
##  Mode  :character   Mode  :character   Median :17.0   Mode  :character
##                                        Mean   :16.7
##                                        3rd Qu.:18.0
##                                        Max.   :22.0
##    famsize            Pstatus              Medu            Fedu
##  Length:395         Length:395         Min.   :0.000   Min.   :0.000
##  Class :character   Class :character   1st Qu.:2.000   1st Qu.:2.000
##  Mode  :character   Mode  :character   Median :3.000   Median :2.000
##                                        Mean   :2.749   Mean   :2.522
##                                        3rd Qu.:4.000   3rd Qu.:3.000
##                                        Max.   :4.000   Max.   :4.000
##      Mjob               Fjob              reason            guardian
##  Length:395         Length:395         Length:395         Length:395
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
```

```
##
##    traveltime      studytime        failures       schoolsup
##  Min.   :1.000   Min.   :1.000   Min.   :0.0000   Length:395
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.0000   Class :character
##  Median :1.000   Median :2.000   Median :0.0000   Mode  :character
##  Mean   :1.448   Mean   :2.035   Mean   :0.3342
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :4.000   Max.   :4.000   Max.   :3.0000
##     famsup              paid            activities          nursery
##  Length:395         Length:395         Length:395         Length:395
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##     higher            internet           romantic             famrel
##  Length:395         Length:395         Length:395         Min.   :1.000
##  Class :character   Class :character   Class :character   1st Qu.:4.000
##  Mode  :character   Mode  :character   Mode  :character   Median :4.000
##                                                           Mean   :3.944
##                                                           3rd Qu.:5.000
##                                                           Max.   :5.000
##     freetime         goout            Dalc            Walc
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :3.000   Median :3.000   Median :1.000   Median :2.000
##  Mean   :3.235   Mean   :3.109   Mean   :1.481   Mean   :2.291
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##      health         absences            G1              G2
##  Min.   :1.000   Min.   : 0.000   Min.   : 3.00   Min.   : 0.00
##  1st Qu.:3.000   1st Qu.: 0.000   1st Qu.: 8.00   1st Qu.: 9.00
##  Median :4.000   Median : 4.000   Median :11.00   Median :11.00
##  Mean   :3.554   Mean   : 5.709   Mean   :10.91   Mean   :10.71
##  3rd Qu.:5.000   3rd Qu.: 8.000   3rd Qu.:13.00   3rd Qu.:13.00
##  Max.   :5.000   Max.   :75.000   Max.   :19.00   Max.   :19.00
##        G3
##  Min.   : 0.00
##  1st Qu.: 8.00
##  Median :11.00
##  Mean   :10.42
##  3rd Qu.:14.00
##  Max.   :20.00
```

```
ggplot(student_mat, aes(sex)) + geom_bar()
```

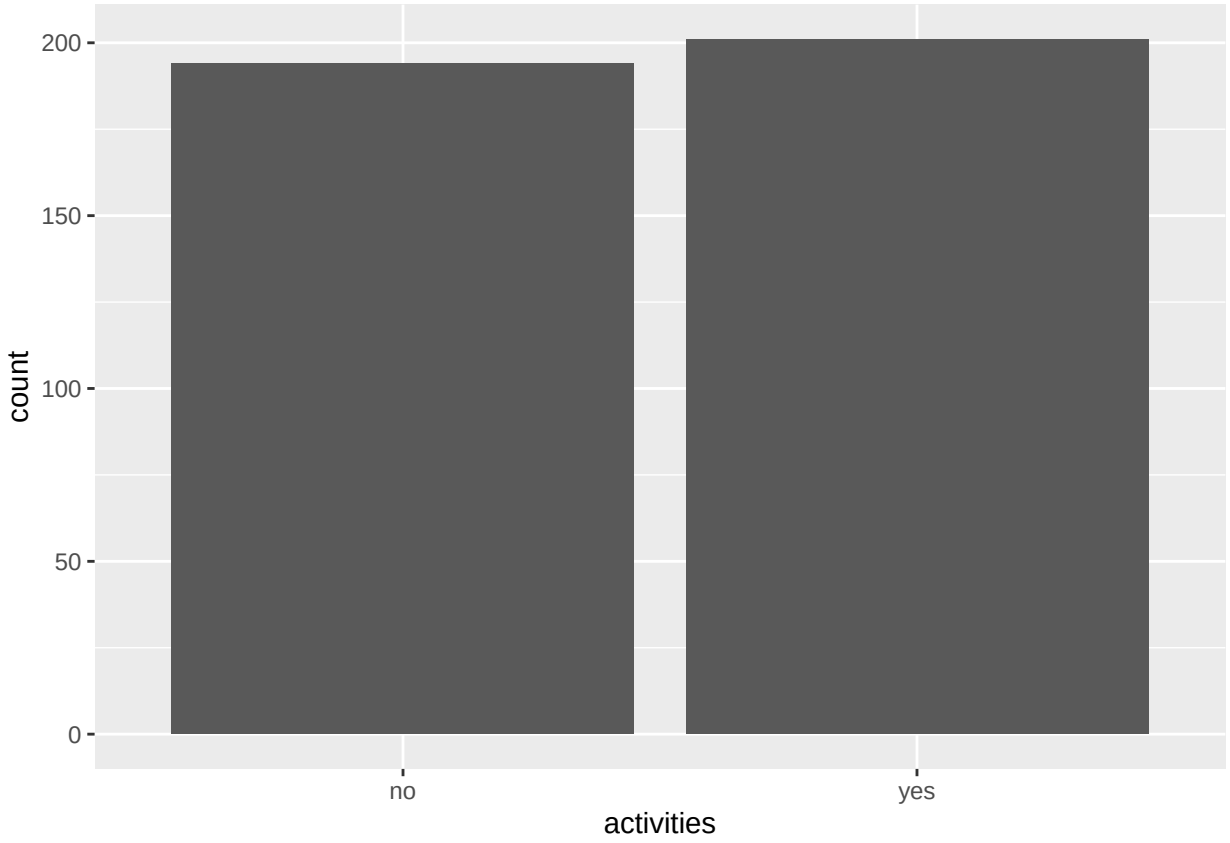


Activities Students reply yes or no if they have extra curricular activities, indicated by the `activities` variables.

```
table(student_mat$activities)
```

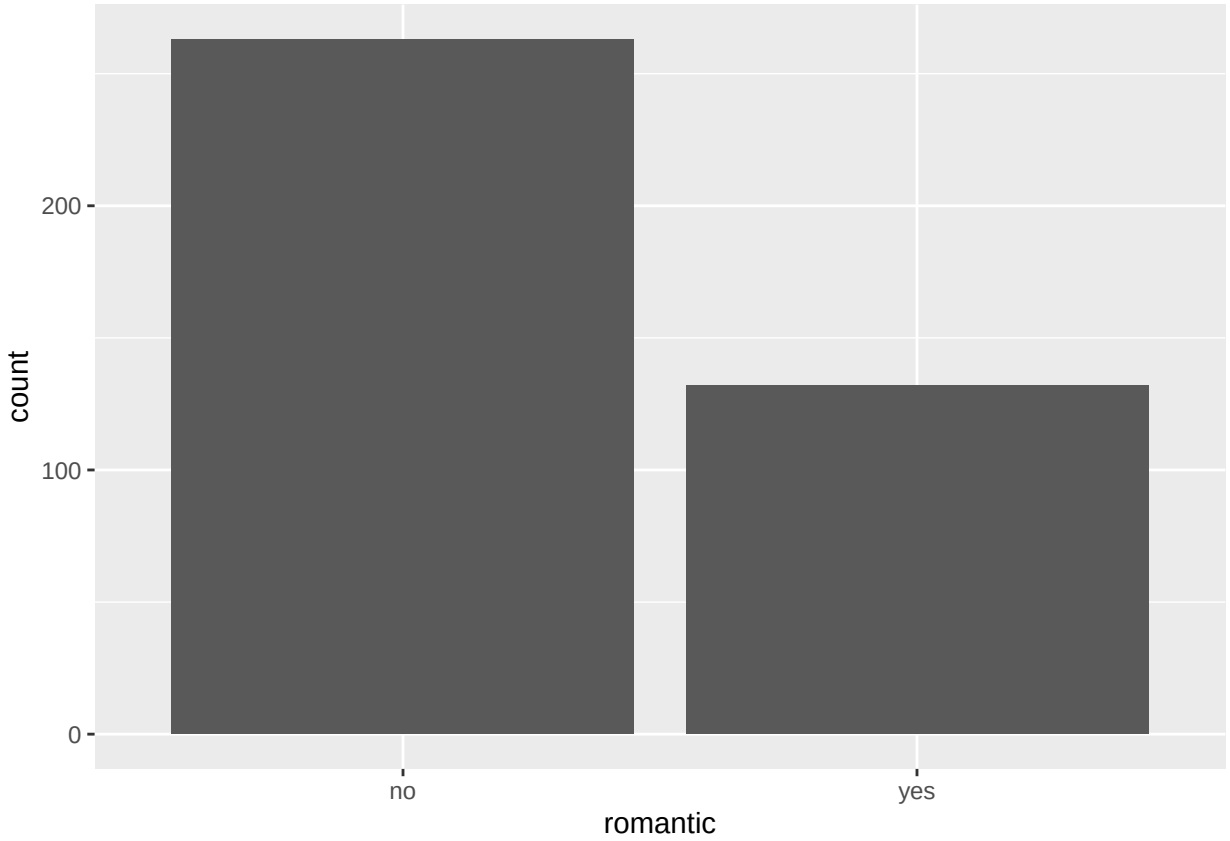
```
##
##  no yes
## 194 201
```

```
ggplot(student_mat, aes(activities)) + geom_bar()
```



Romantic Students reply either yes or no if they are in a romantic relationship, indicated by the `romantic` variable.

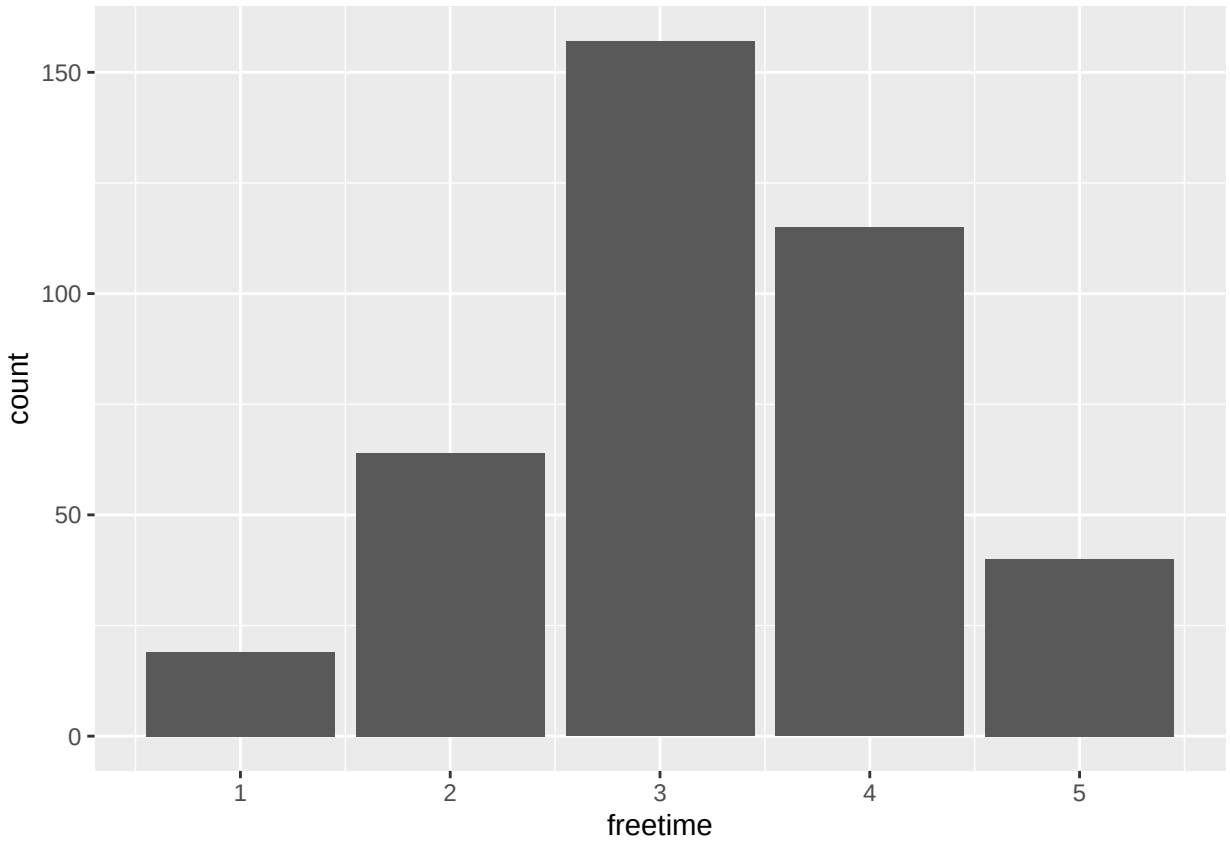
```
ggplot(student_mat, aes(romantic)) + geom_bar()
```



Freetime

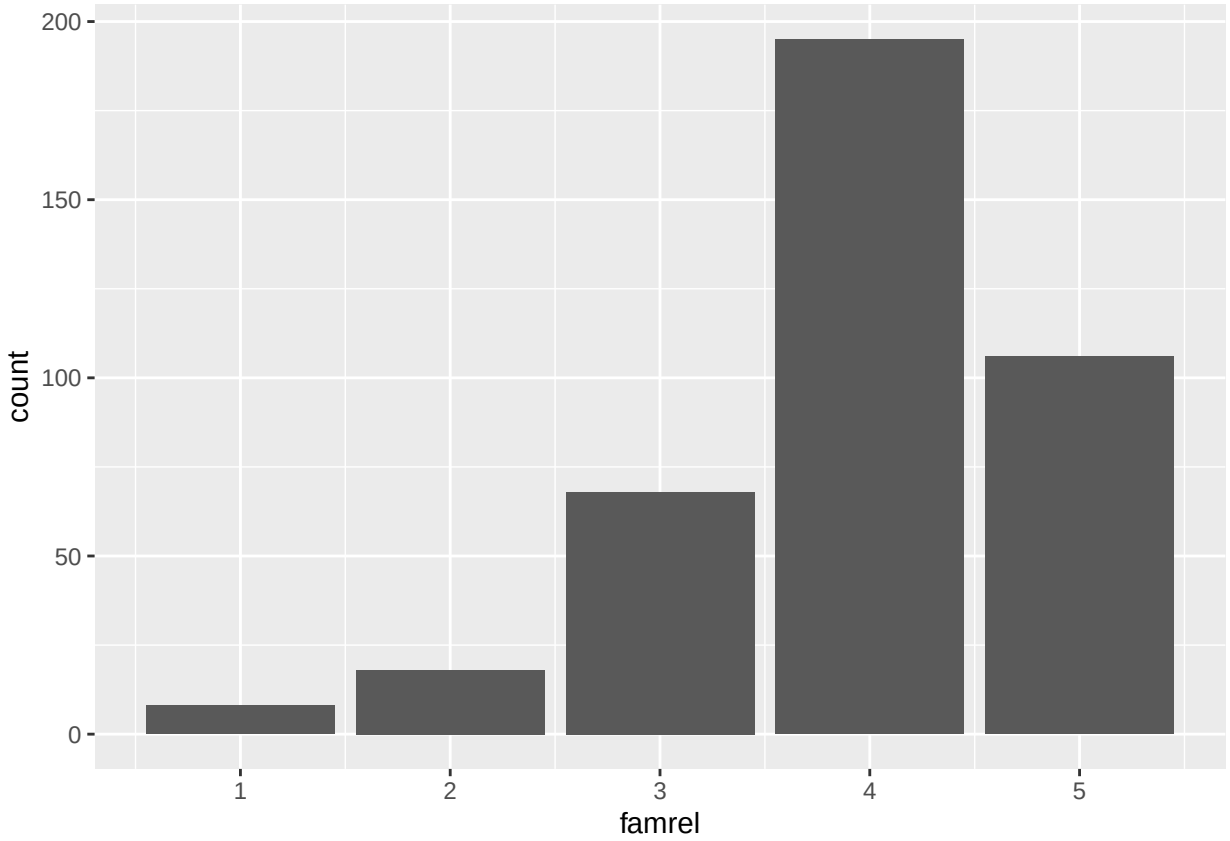
Freetime after school rated from 1 very low to 5 very high.

```
ggplot(student_mat, aes(freetime)) + geom_bar()
```



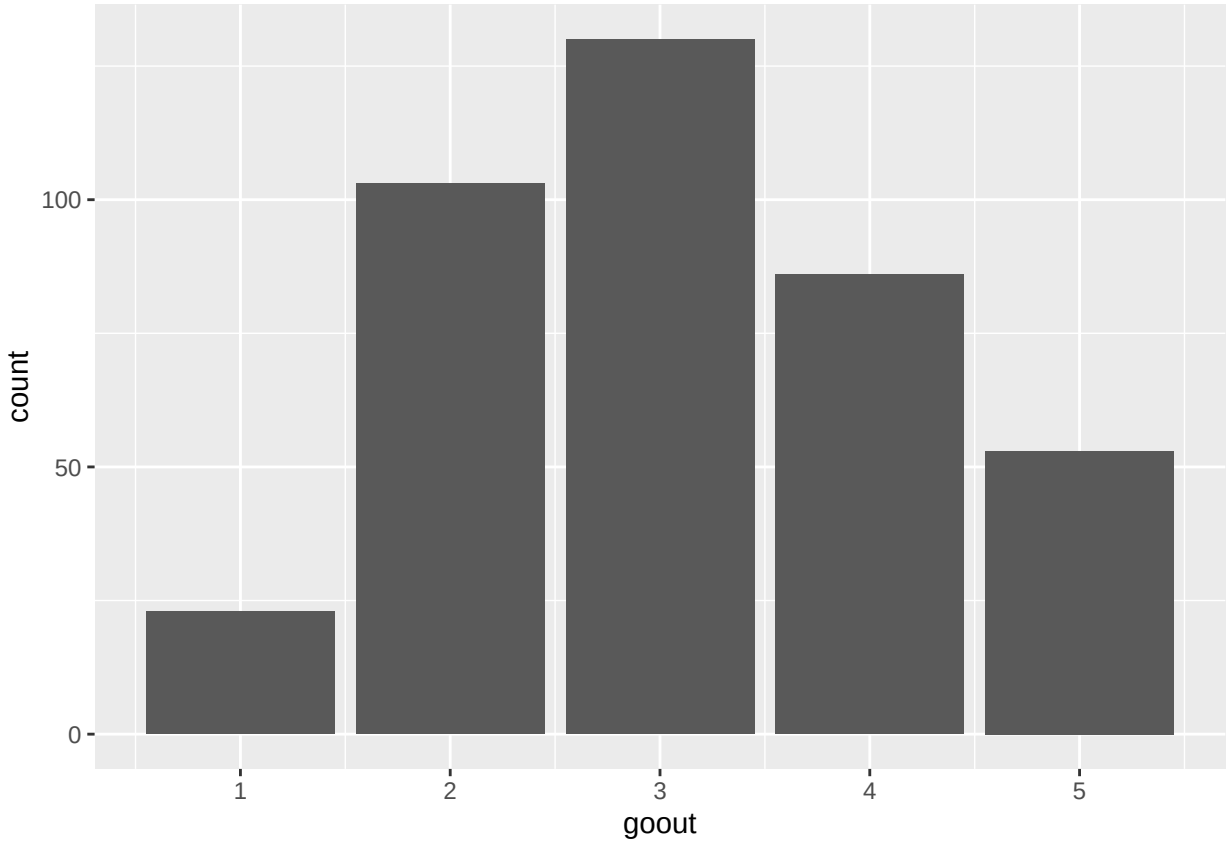
Family Relationship `famrel` indicated the quality of family relationships on a scale from (numeric: from 1 - very bad to 5 - excellent)

```
ggplot(student_mat, aes(famrel)) + geom_bar()
```



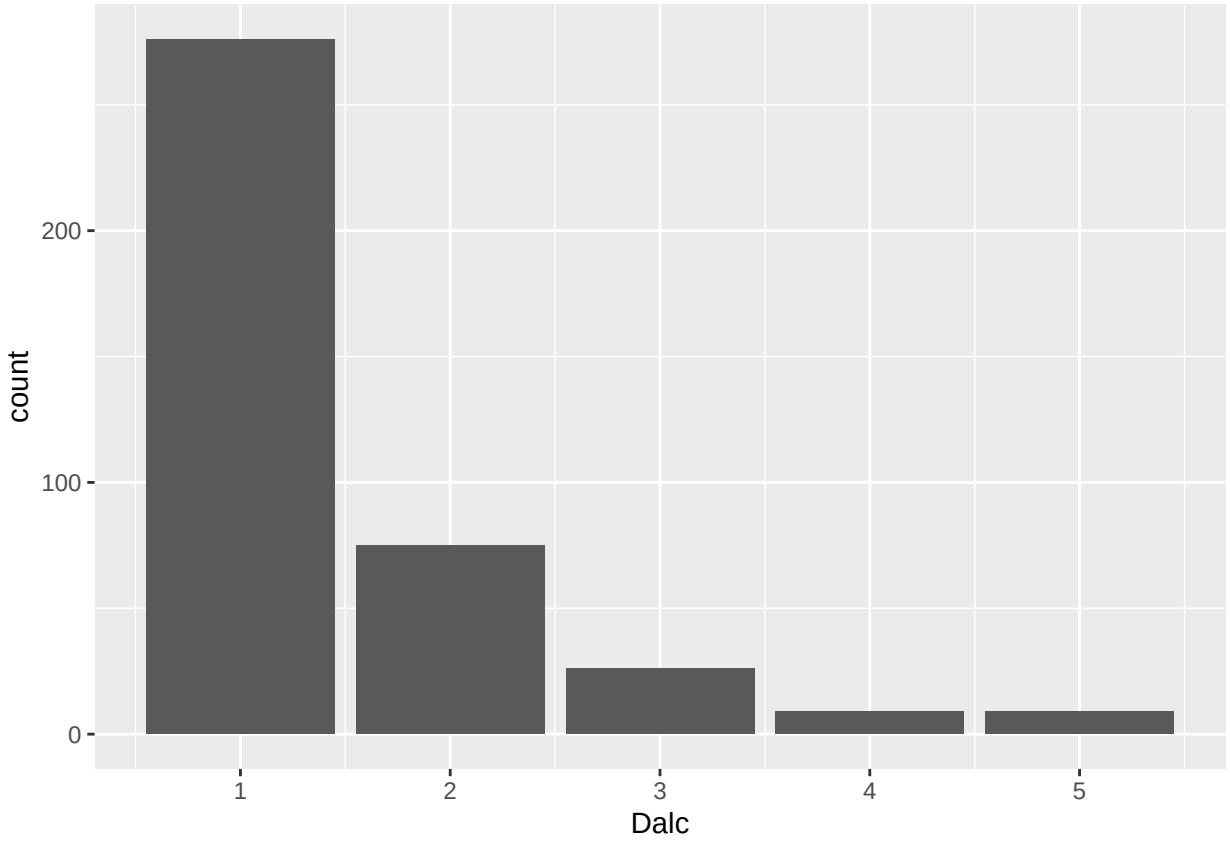
Going Out `goout` indicates if students go out with friends rated from (numeric: from 1 - very low to 5 - very high)

```
ggplot(student_mat, aes(goout)) + geom_bar()
```

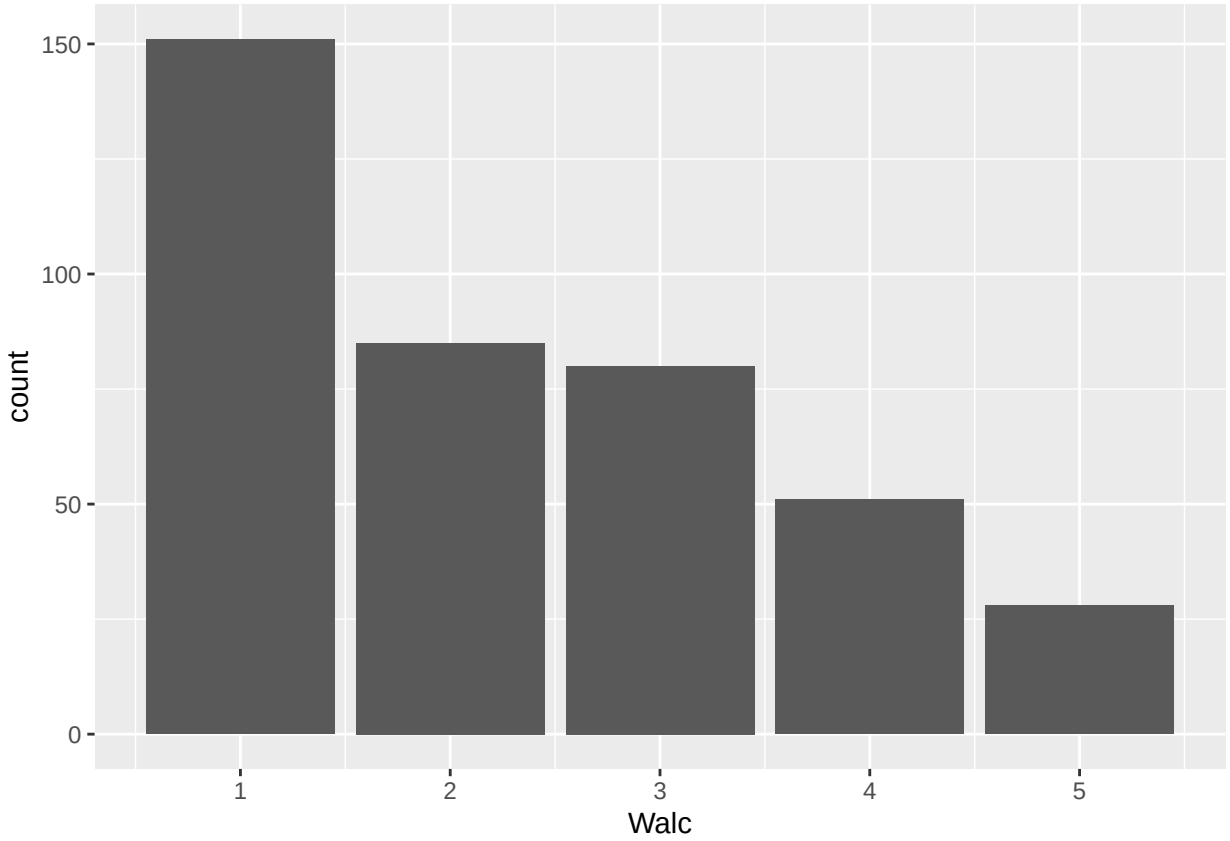


Alcohol Consumption Dalc indicated the workday alcohol consumption from very low to very high and Walc indicates weekend alcohol consumption, also from very low to very high.

```
ggplot(student_mat, aes(Dalc)) + geom_bar()
```



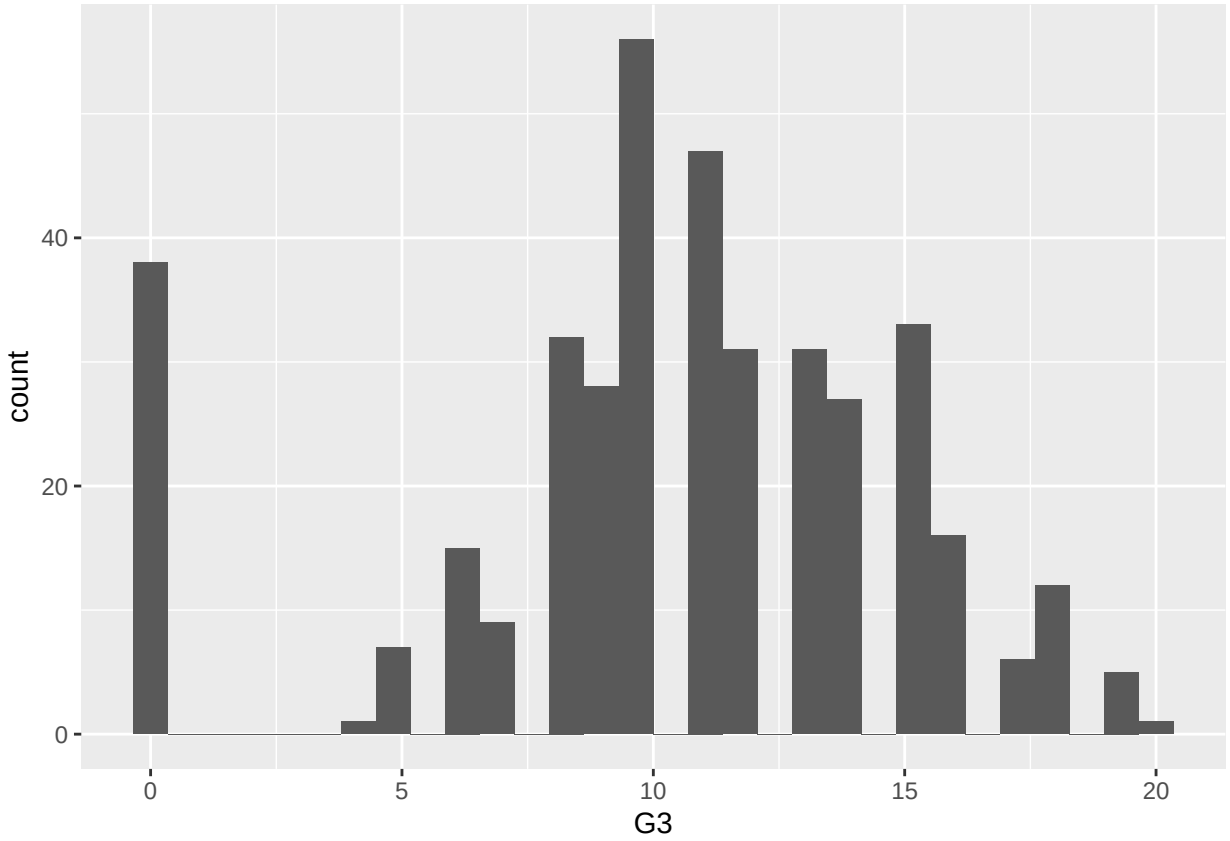

```
ggplot(student_mat, aes(Walc)) + geom_bar()
```




```
ggplot(student_mat, aes(x=G3)) + geom_histogram()
```

Grades

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Sample comparison For the final project I will compare the final grades with the other social factors. Here I compare with the student romantic social life.

```
describe(student_mat$G3)
```

```
##    vars   n  mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 395 10.42 4.58     11   10.84 4.45   0  20    20 -0.73     0.37 0.23
```

```
describeBy(student_mat$G3,
           group = student_mat$romantic, mat=TRUE)
```

```
##     item group1 vars   n      mean       sd median   trimmed    mad min max
## X11    1     no    1 263 10.836502 4.385946     11 11.113744 4.4478   0  20
## X12    2    yes    1 132  9.575758 4.856916     11  9.971698 3.7065   0  18
##     range       skew   kurtosis        se
## X11    20 -0.6186634  0.5452026 0.2704490
## X12    18 -0.8212904 -0.2272406 0.4227403
```