

Josephs-Final

Moiya Josephs

2022-10-18

Data Preparation

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

Abstract

Research question

Does having a social life predict final grade?

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

```
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)
```

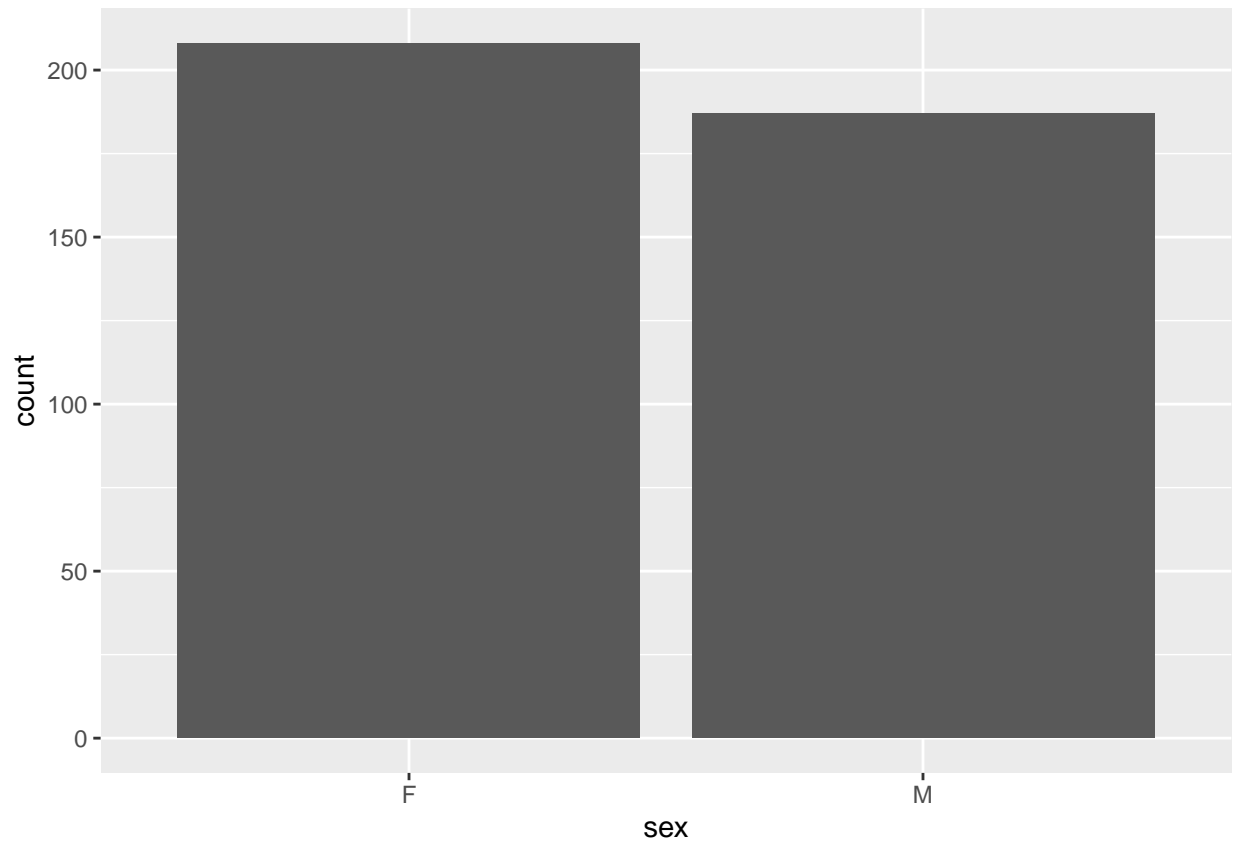
```
##      school      sex      age      address
## Length:395    Length:395    Min.   :15.0    Length:395
## Class :character    Class :character    1st Qu.:16.0    Class :character
## Mode  :character    Mode  :character    Median :17.0    Mode  :character
##                                     Mean   :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

```

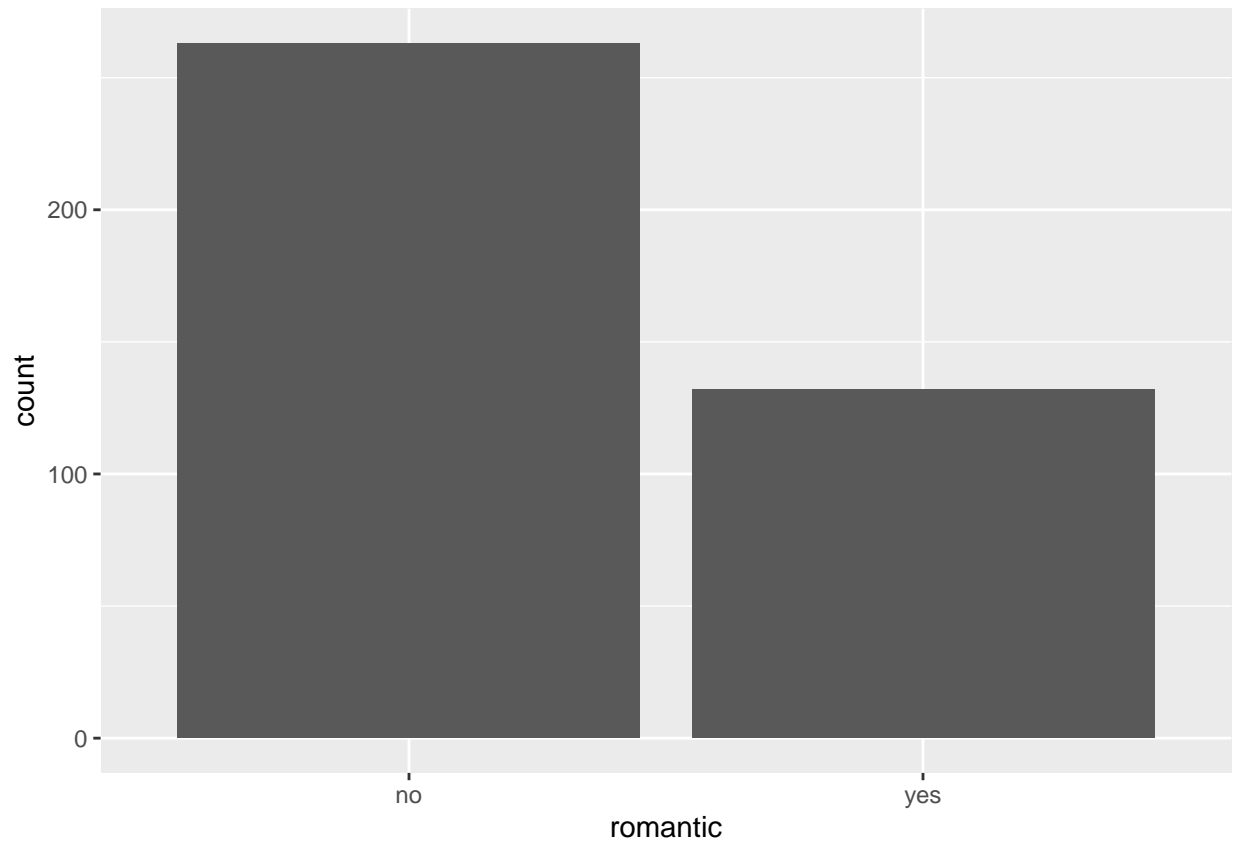
```
ggplot(student_mat, aes(sex)) + geom_bar()
```



Romantic

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

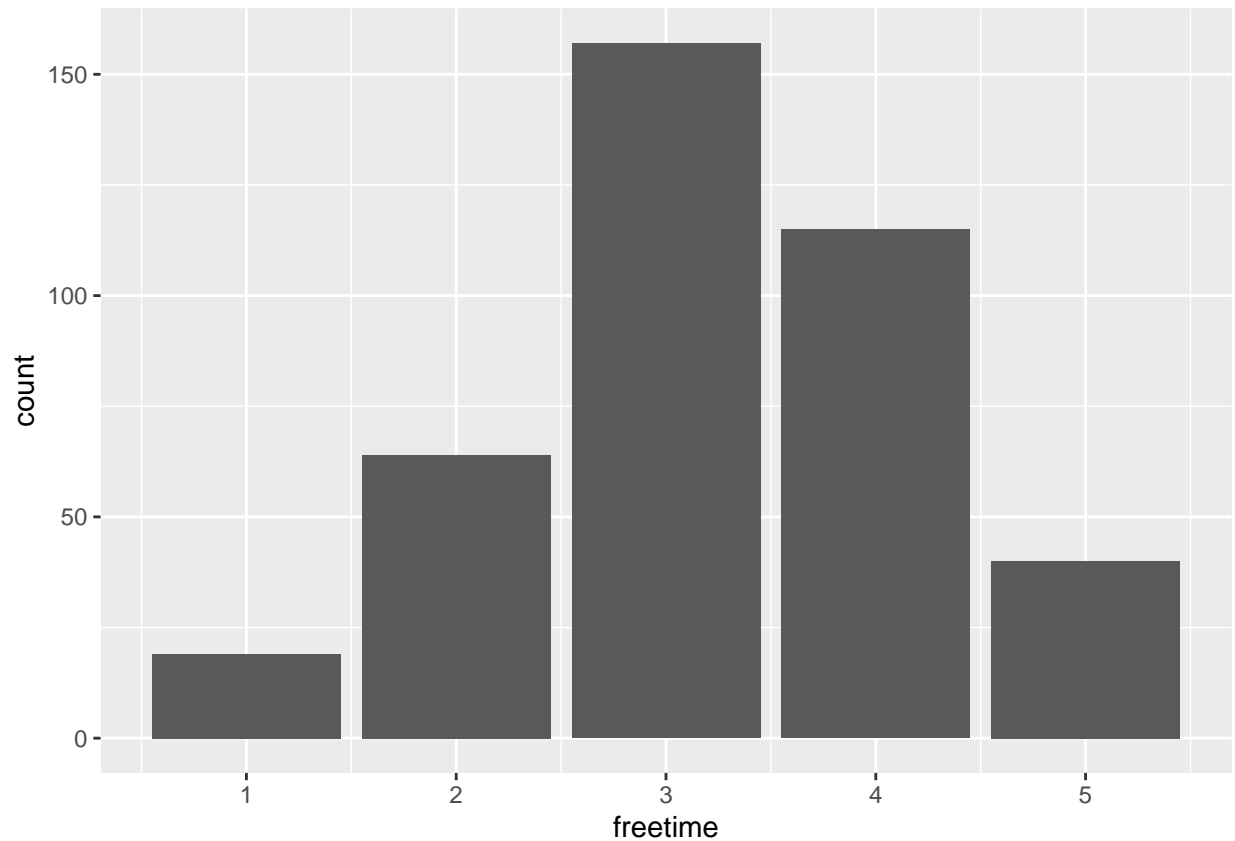
```
ggplot(student_mat, aes(romantic)) + geom_bar()
```



Freetime

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

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



Grades

```
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
```

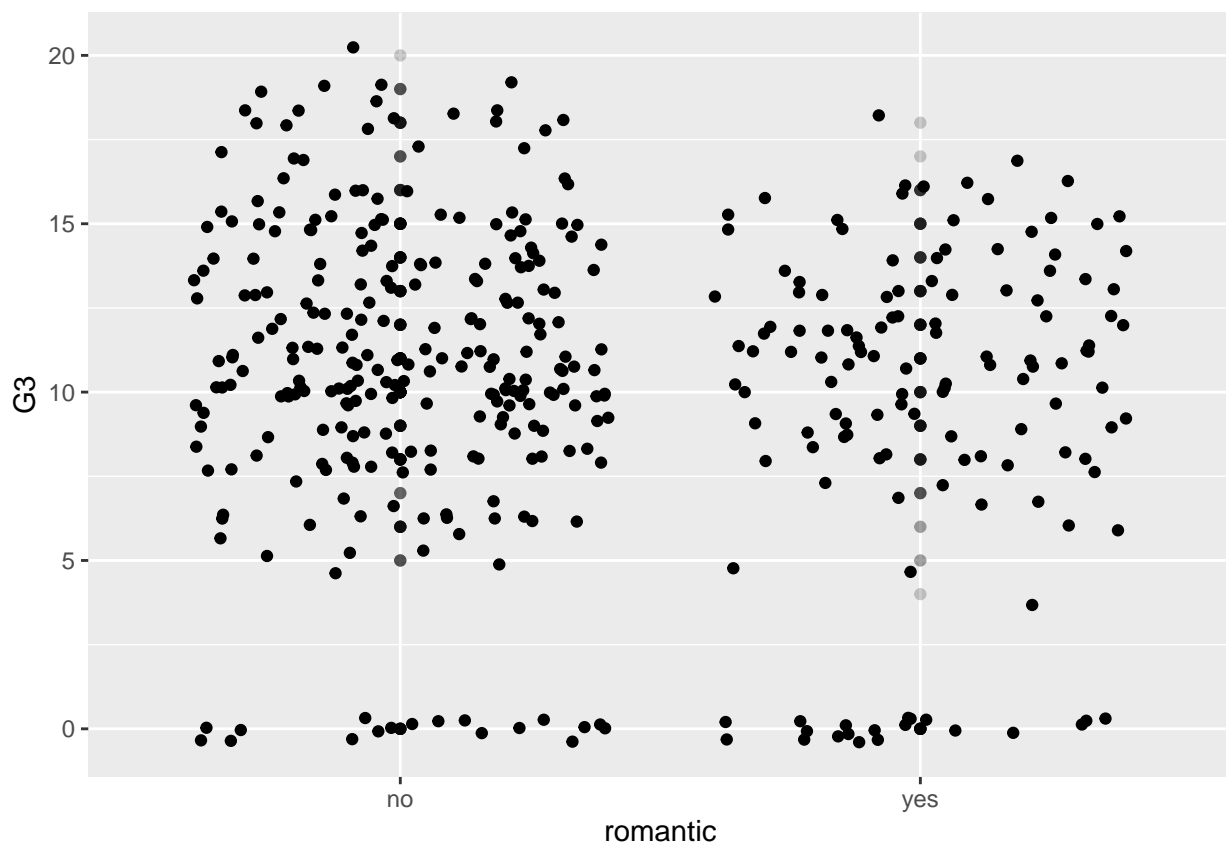
The Romance Model

```
romantic.model <- lm(G3 ~ romantic, data = student_mat)
summary(romantic.model)
```

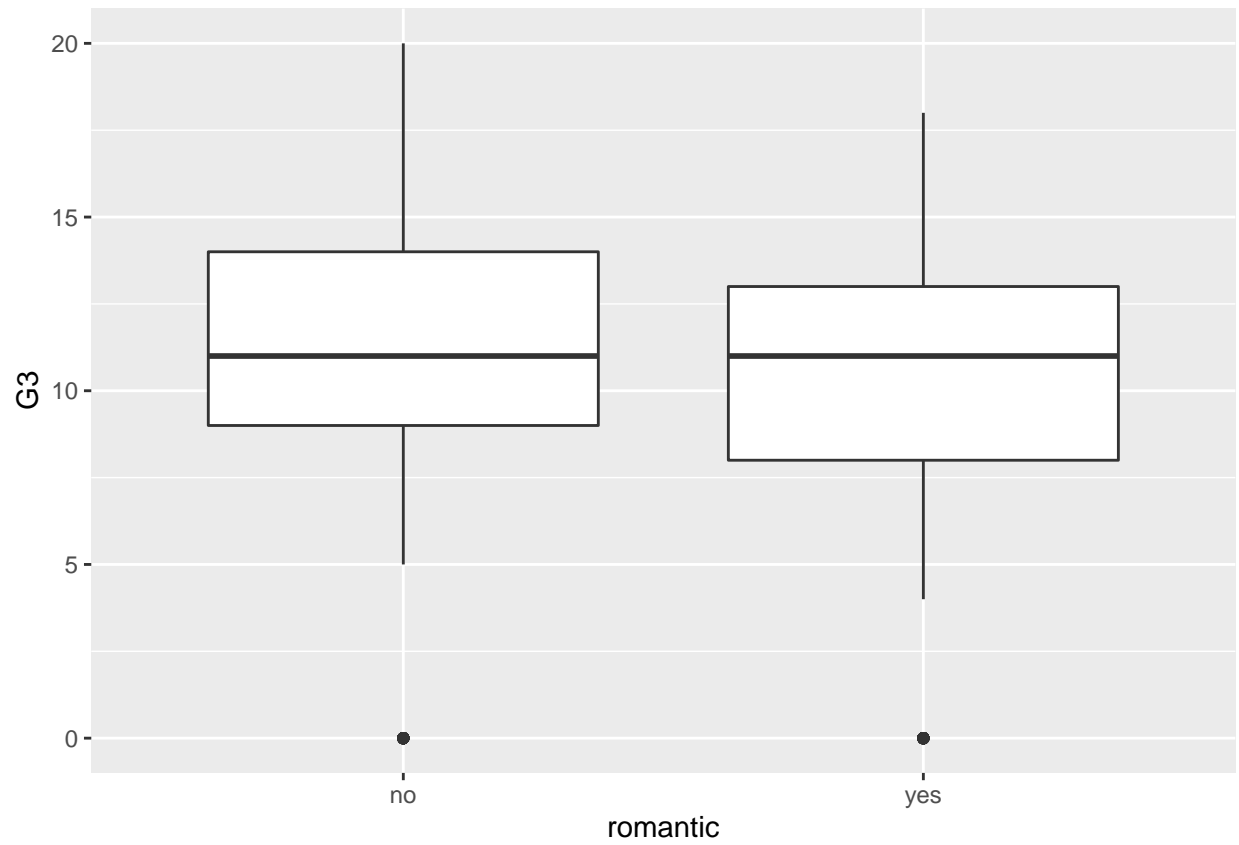
```
##
## Call:
## lm(formula = G3 ~ romantic, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8365  -1.8365   0.1635   3.1635   9.1635
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.8365     0.2805  38.638 < 2e-16 ***
## romanticyes  -1.2607     0.4852  -2.599  0.00971 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.548 on 393 degrees of freedom
## Multiple R-squared:  0.01689,    Adjusted R-squared:  0.01439
## F-statistic: 6.753 on 1 and 393 DF,  p-value: 0.009713
```

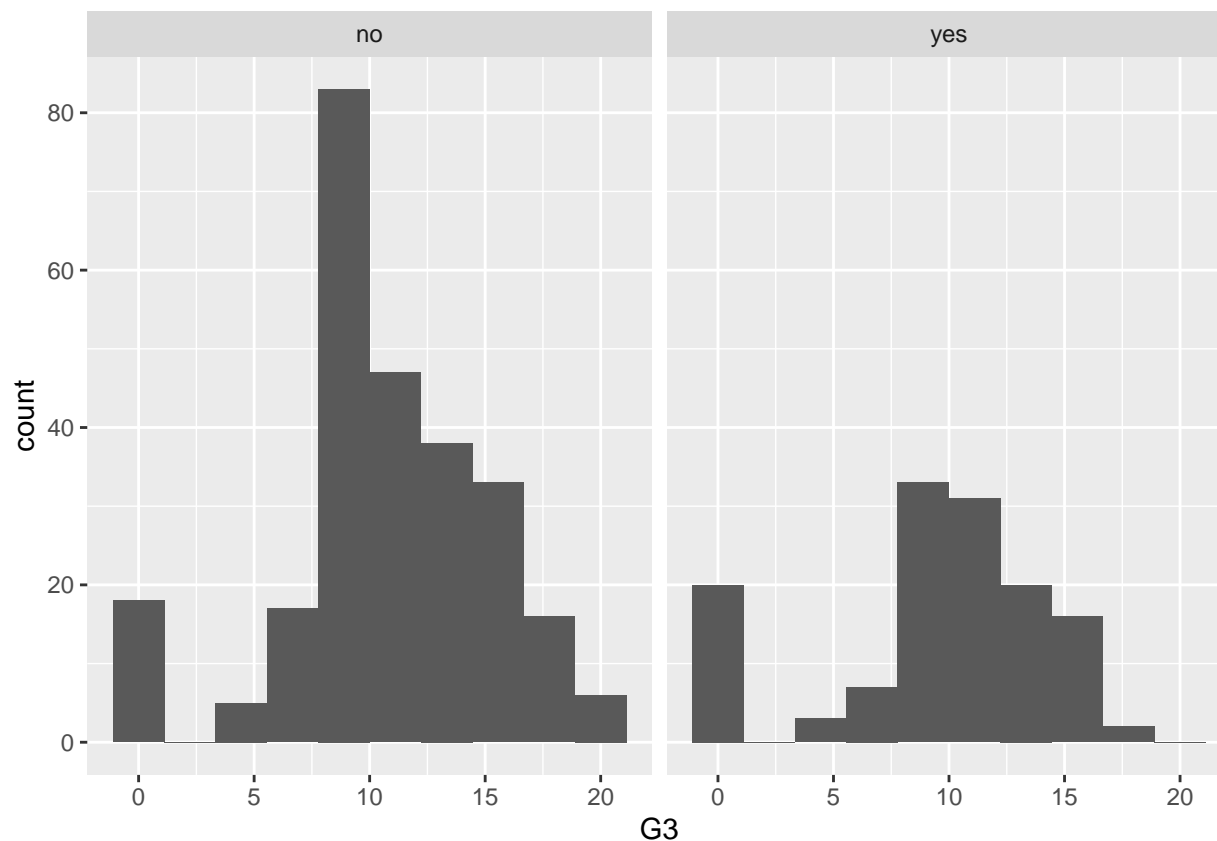
```
ggplot(student_mat, aes(x=romantic, y = G3)) +
  geom_jitter() +
  geom_point(alpha = 0.2)
```



```
ggplot(student_mat, aes(x=romantic, y = G3)) +
  geom_boxplot()
```



```
ggplot(student_mat, aes(x= G3)) +  
  geom_histogram(bins = 10) +  
  facet_wrap(vars(romantic))
```

```
summary_stats <- student_mat %>%
  group_by(romantic) %>%
  summarise(mean_by_group = mean(G3))
summary_stats
```

```
## # A tibble: 2 x 2
##   romantic mean_by_group
##   <chr>         <dbl>
## 1 no           10.8
## 2 yes           9.58
```

```
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
```

Making predictions:

```
explanatory_data <- tibble(romantic = c("yes","no"))
explanatory_data
```

```
## # A tibble: 2 x 1
##   romantic
##   <chr>
## 1 yes
## 2 no
```

```
explanatory_data %>%
  mutate(
    response_var = predict(romantic.model, explanatory_data)
  )
```

```
## # A tibble: 2 x 2
##   romantic response_var
##   <chr>           <dbl>
## 1 yes             9.58
## 2 no             10.8
```

Visualizing predictions

```
#Predict the grade for each student
predict(romantic.model)
```

```
##           1           2           3           4           5           6           7           8
## 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502 10.836502 10.836502
##           9          10          11          12          13          14          15          16
## 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502
##          17          18          19          20          21          22          23          24
## 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502
##          25          26          27          28          29          30          31          32
## 10.836502 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          33          34          35          36          37          38          39          40
##  9.575758 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          41          42          43          44          45          46          47          48
##  9.575758  9.575758 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          49          50          51          52          53          54          55          56
## 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502  9.575758
##          57          58          59          60          61          62          63          64
## 10.836502 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          65          66          67          68          69          70          71          72
##  9.575758 10.836502  9.575758 10.836502 10.836502 10.836502 10.836502 10.836502
##          73          74          75          76          77          78          79          80
##  9.575758 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          81          82          83          84          85          86          87          88
##  9.575758 10.836502 10.836502 10.836502 10.836502  9.575758 10.836502 10.836502
##          89          90          91          92          93          94          95          96
## 10.836502 10.836502  9.575758 10.836502 10.836502 10.836502 10.836502 10.836502
##          97          98          99         100         101         102         103         104
```

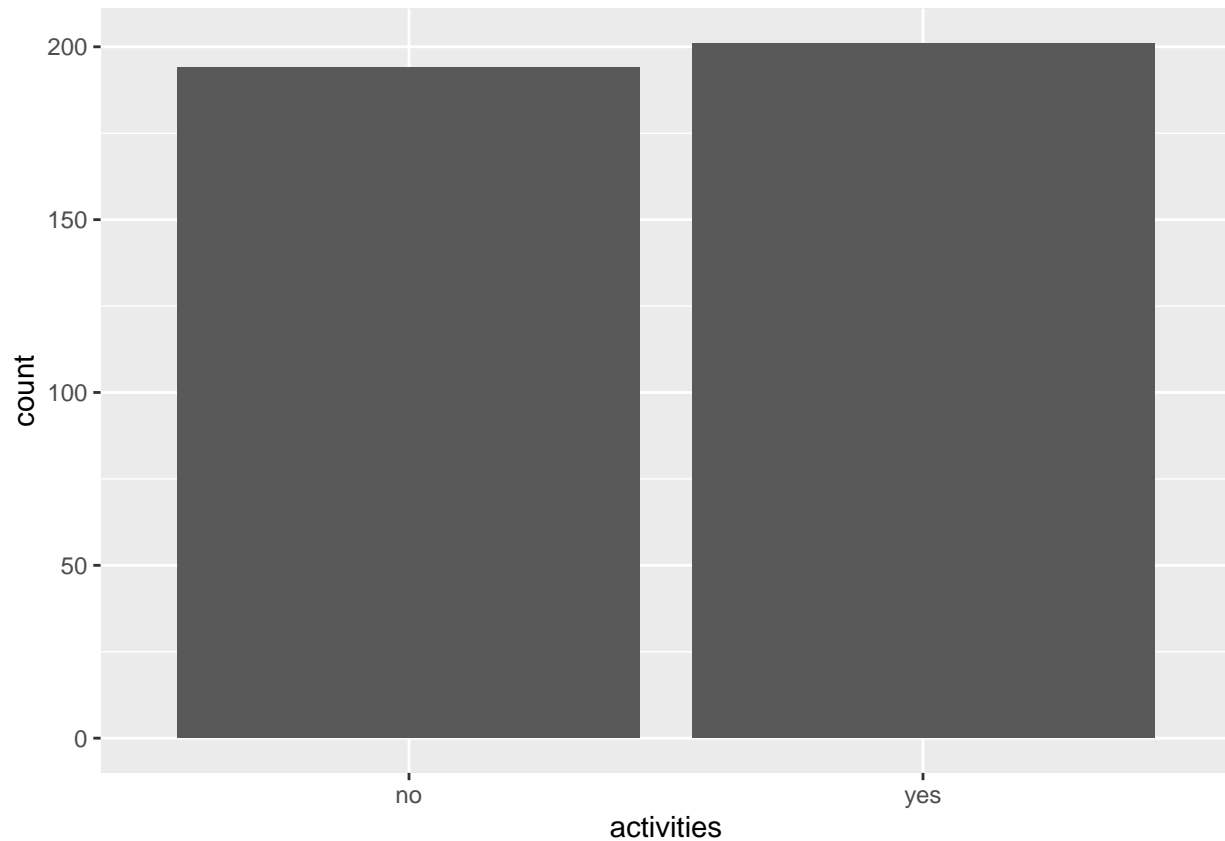
##	10.836502	9.575758	10.836502	10.836502	10.836502	9.575758	10.836502	10.836502
##	105	106	107	108	109	110	111	112
##	10.836502	10.836502	10.836502	10.836502	9.575758	9.575758	10.836502	10.836502
##	113	114	115	116	117	118	119	120
##	10.836502	10.836502	9.575758	10.836502	10.836502	10.836502	10.836502	10.836502
##	121	122	123	124	125	126	127	128
##	10.836502	10.836502	9.575758	10.836502	9.575758	10.836502	9.575758	10.836502
##	129	130	131	132	133	134	135	136
##	10.836502	10.836502	9.575758	9.575758	9.575758	10.836502	9.575758	9.575758
##	137	138	139	140	141	142	143	144
##	10.836502	9.575758	9.575758	10.836502	10.836502	10.836502	10.836502	10.836502
##	145	146	147	148	149	150	151	152
##	10.836502	10.836502	10.836502	10.836502	9.575758	10.836502	9.575758	9.575758
##	153	154	155	156	157	158	159	160
##	9.575758	9.575758	9.575758	10.836502	10.836502	10.836502	10.836502	9.575758
##	161	162	163	164	165	166	167	168
##	9.575758	9.575758	10.836502	10.836502	9.575758	10.836502	10.836502	9.575758
##	169	170	171	172	173	174	175	176
##	10.836502	9.575758	10.836502	9.575758	10.836502	9.575758	10.836502	10.836502
##	177	178	179	180	181	182	183	184
##	10.836502	10.836502	9.575758	9.575758	10.836502	9.575758	10.836502	9.575758
##	185	186	187	188	189	190	191	192
##	10.836502	9.575758	9.575758	9.575758	9.575758	10.836502	10.836502	10.836502
##	193	194	195	196	197	198	199	200
##	10.836502	10.836502	10.836502	9.575758	10.836502	10.836502	10.836502	10.836502
##	201	202	203	204	205	206	207	208
##	10.836502	10.836502	10.836502	10.836502	10.836502	9.575758	10.836502	9.575758
##	209	210	211	212	213	214	215	216
##	10.836502	9.575758	10.836502	9.575758	10.836502	10.836502	10.836502	10.836502
##	217	218	219	220	221	222	223	224
##	9.575758	10.836502	10.836502	10.836502	10.836502	9.575758	10.836502	10.836502
##	225	226	227	228	229	230	231	232
##	10.836502	9.575758	10.836502	10.836502	9.575758	9.575758	9.575758	10.836502
##	233	234	235	236	237	238	239	240
##	9.575758	10.836502	10.836502	9.575758	9.575758	9.575758	10.836502	10.836502
##	241	242	243	244	245	246	247	248
##	9.575758	10.836502	10.836502	10.836502	9.575758	10.836502	10.836502	9.575758
##	249	250	251	252	253	254	255	256
##	9.575758	10.836502	10.836502	10.836502	10.836502	10.836502	10.836502	10.836502
##	257	258	259	260	261	262	263	264
##	10.836502	10.836502	10.836502	9.575758	9.575758	10.836502	10.836502	10.836502
##	265	266	267	268	269	270	271	272
##	9.575758	10.836502	9.575758	10.836502	9.575758	9.575758	10.836502	9.575758
##	273	274	275	276	277	278	279	280
##	10.836502	9.575758	9.575758	9.575758	9.575758	10.836502	9.575758	10.836502
##	281	282	283	284	285	286	287	288
##	9.575758	10.836502	10.836502	10.836502	10.836502	10.836502	10.836502	10.836502
##	289	290	291	292	293	294	295	296
##	10.836502	10.836502	9.575758	10.836502	9.575758	10.836502	10.836502	10.836502
##	297	298	299	300	301	302	303	304
##	10.836502	9.575758	10.836502	9.575758	9.575758	10.836502	10.836502	10.836502
##	305	306	307	308	309	310	311	312
##	9.575758	10.836502	10.836502	9.575758	9.575758	10.836502	9.575758	9.575758
##	313	314	315	316	317	318	319	320

```
## 10.836502 9.575758 9.575758 9.575758 10.836502 9.575758 10.836502 10.836502
##      321      322      323      324      325      326      327      328
## 9.575758 9.575758 10.836502 10.836502 10.836502 10.836502 10.836502 10.836502
##      329      330      331      332      333      334      335      336
## 10.836502 9.575758 10.836502 9.575758 10.836502 9.575758 10.836502 10.836502
##      337      338      339      340      341      342      343      344
## 10.836502 9.575758 10.836502 10.836502 9.575758 10.836502 9.575758 9.575758
##      345      346      347      348      349      350      351      352
## 10.836502 9.575758 9.575758 9.575758 9.575758 10.836502 10.836502 10.836502
##      353      354      355      356      357      358      359      360
## 9.575758 10.836502 9.575758 9.575758 10.836502 9.575758 9.575758 10.836502
##      361      362      363      364      365      366      367      368
## 9.575758 10.836502 9.575758 9.575758 10.836502 10.836502 9.575758 9.575758
##      369      370      371      372      373      374      375      376
## 9.575758 9.575758 9.575758 9.575758 9.575758 10.836502 10.836502 10.836502
##      377      378      379      380      381      382      383      384
## 9.575758 10.836502 9.575758 10.836502 10.836502 9.575758 10.836502 10.836502
##      385      386      387      388      389      390      391      392
## 10.836502 10.836502 9.575758 10.836502 10.836502 10.836502 10.836502 10.836502
##      393      394      395
## 10.836502 10.836502 10.836502
```

The Activities Model

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

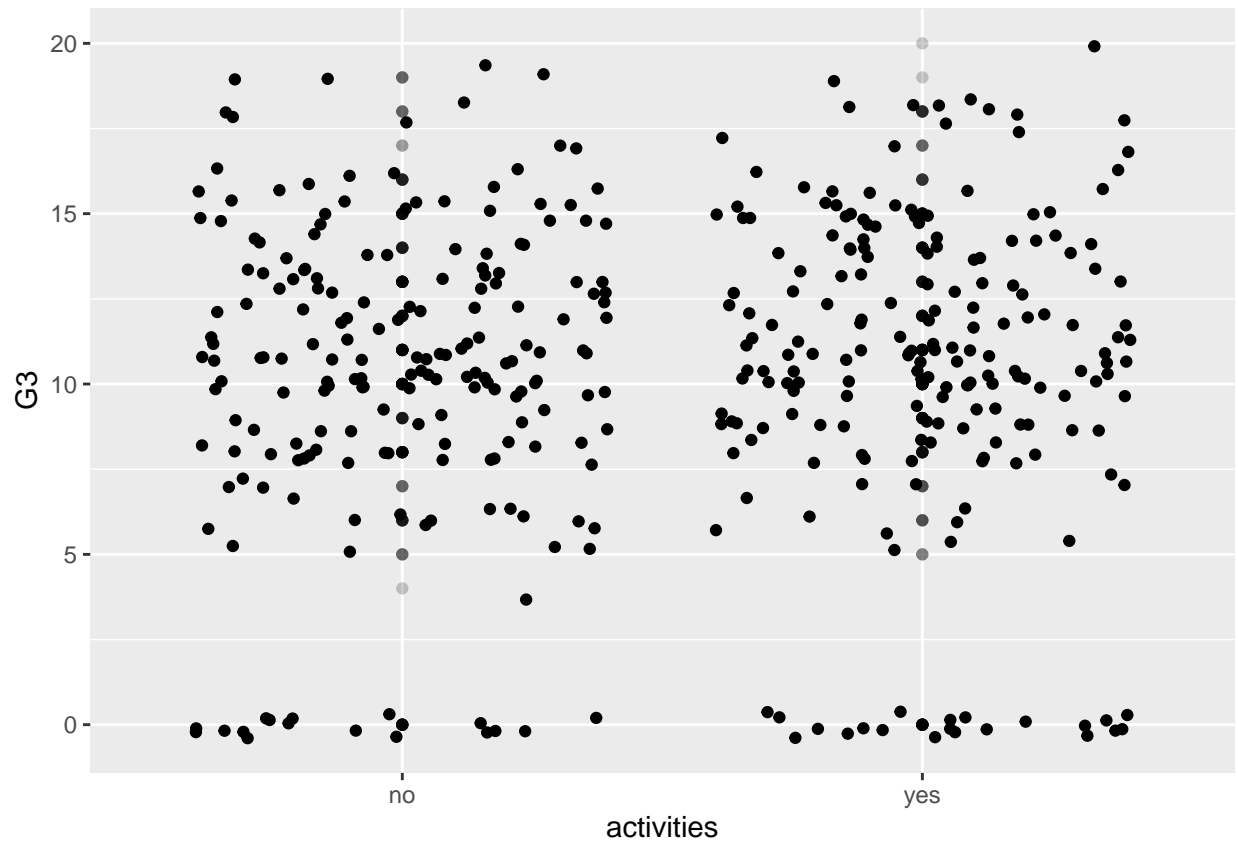
```
ggplot(student_mat, aes(activities)) + geom_bar()
```



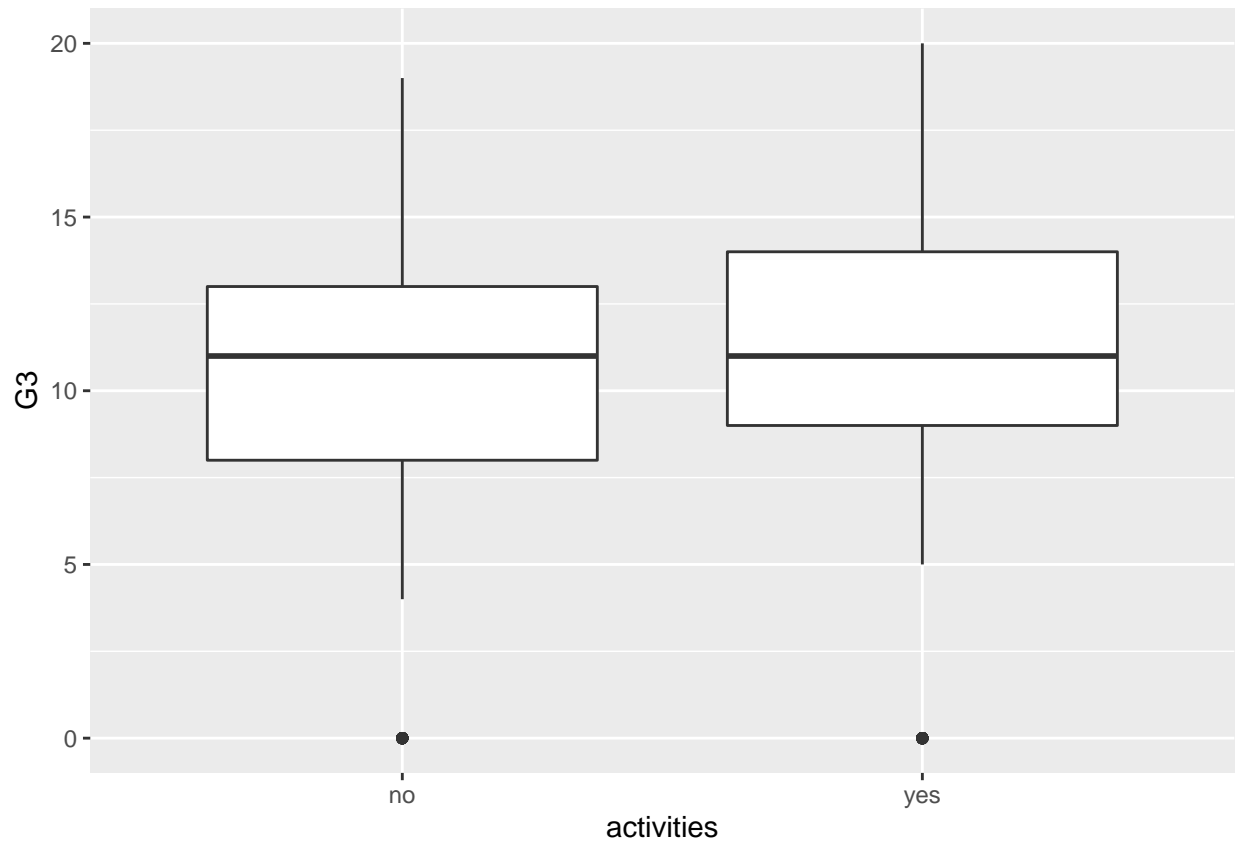
```
activities.model <- lm(G3 ~ activities, data = student_mat)
summary(activities.model)
```

```
##
## Call:
## lm(formula = G3 ~ activities, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4876  -2.3402   0.5124   3.5124   9.5124
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.3402     0.3293  31.400  <2e-16 ***
## activitiesyes     0.1474     0.4616   0.319    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.587 on 393 degrees of freedom
## Multiple R-squared:  0.0002592, Adjusted R-squared:  -0.002285
## F-statistic: 0.1019 on 1 and 393 DF,  p-value: 0.7497
```

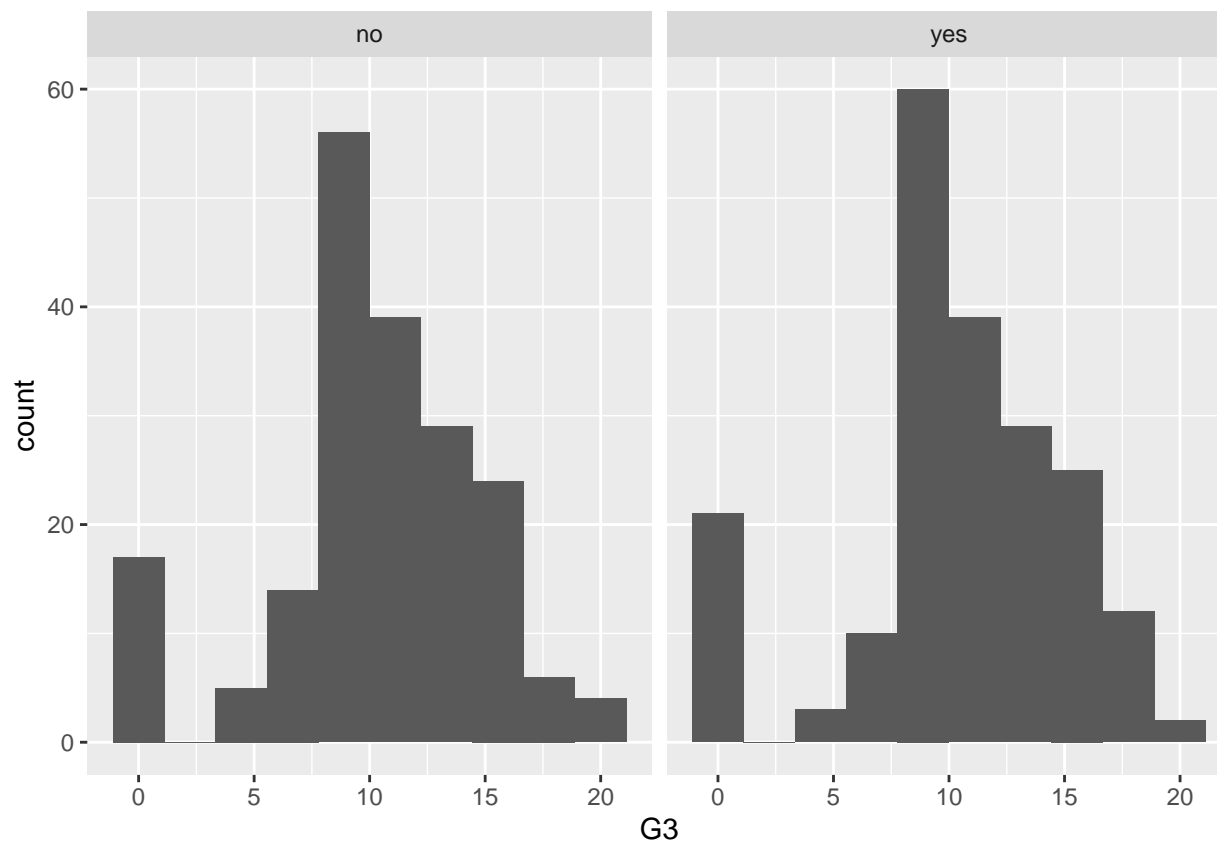
```
ggplot(student_mat, aes(x=activities, y = G3)) +
  geom_jitter() +
  geom_point(alpha = 0.2)
```



```
ggplot(student_mat, aes(x=activities, y = G3)) +  
  geom_boxplot()
```



```
ggplot(student_mat, aes(x= G3)) +  
  geom_histogram(bins = 10) +  
  facet_wrap(vars(activities))
```



```
summary_stats <- student_mat %>%
  group_by(activities) %>%
  summarise(mean_by_group = mean(G3))
summary_stats
```

```
## # A tibble: 2 x 2
##   activities mean_by_group
##   <chr>          <dbl>
## 1 no             10.3
## 2 yes            10.5
```

Making predictions:

```
explanatory_data <- tibble(activities = c("yes","no"))
explanatory_data
```

```
## # A tibble: 2 x 1
##   activities
##   <chr>
## 1 yes
## 2 no
```



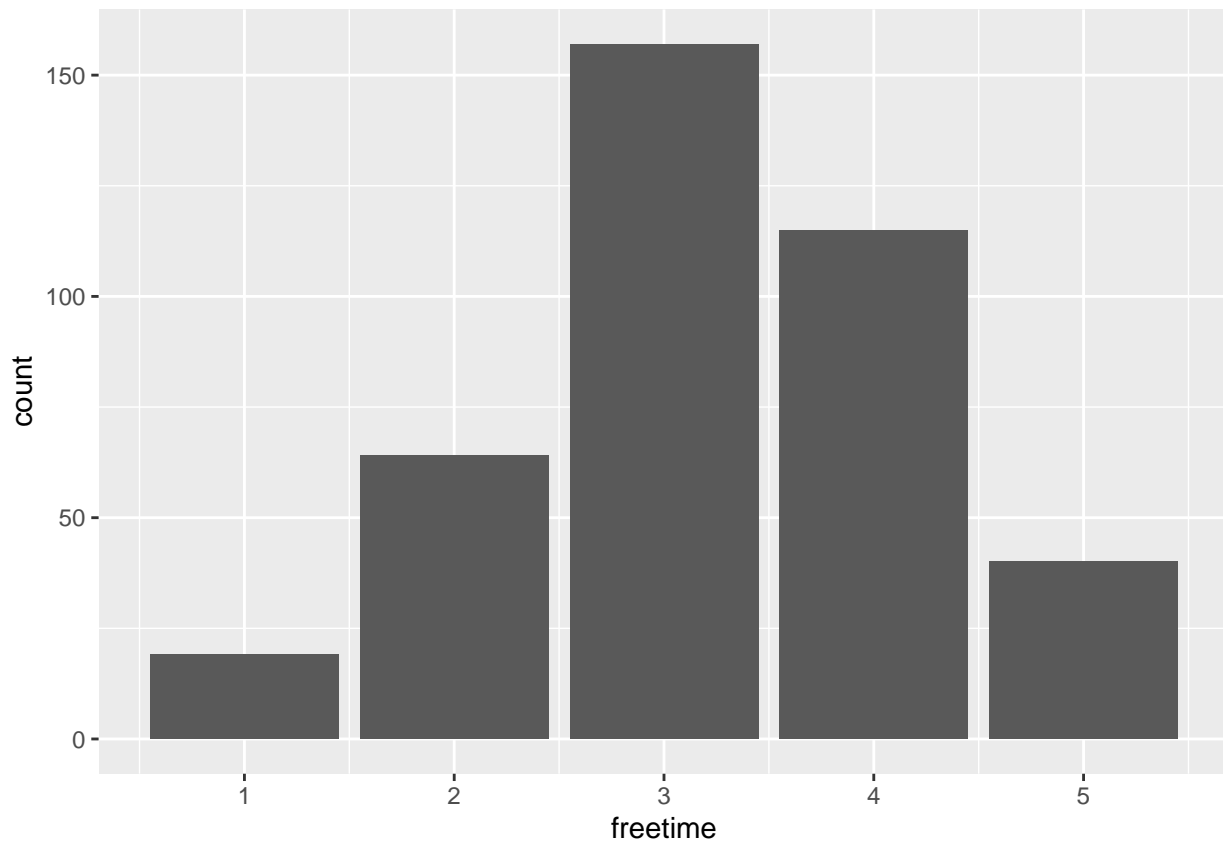
```
explanatory_data %>%
  mutate(
    response_var = predict(activities.model, explanatory_data)
  )
```

```
## # A tibble: 2 x 2
##   activities response_var
##   <chr>         <dbl>
## 1 yes           10.5
## 2 no            10.3
```

Freetime

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

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



```
freetime.model <- lm(G3 ~ freetime, data = student_mat)
summary(freetime.model)
```

```
##
## Call:
## lm(formula = G3 ~ freetime, data = student_mat)
```

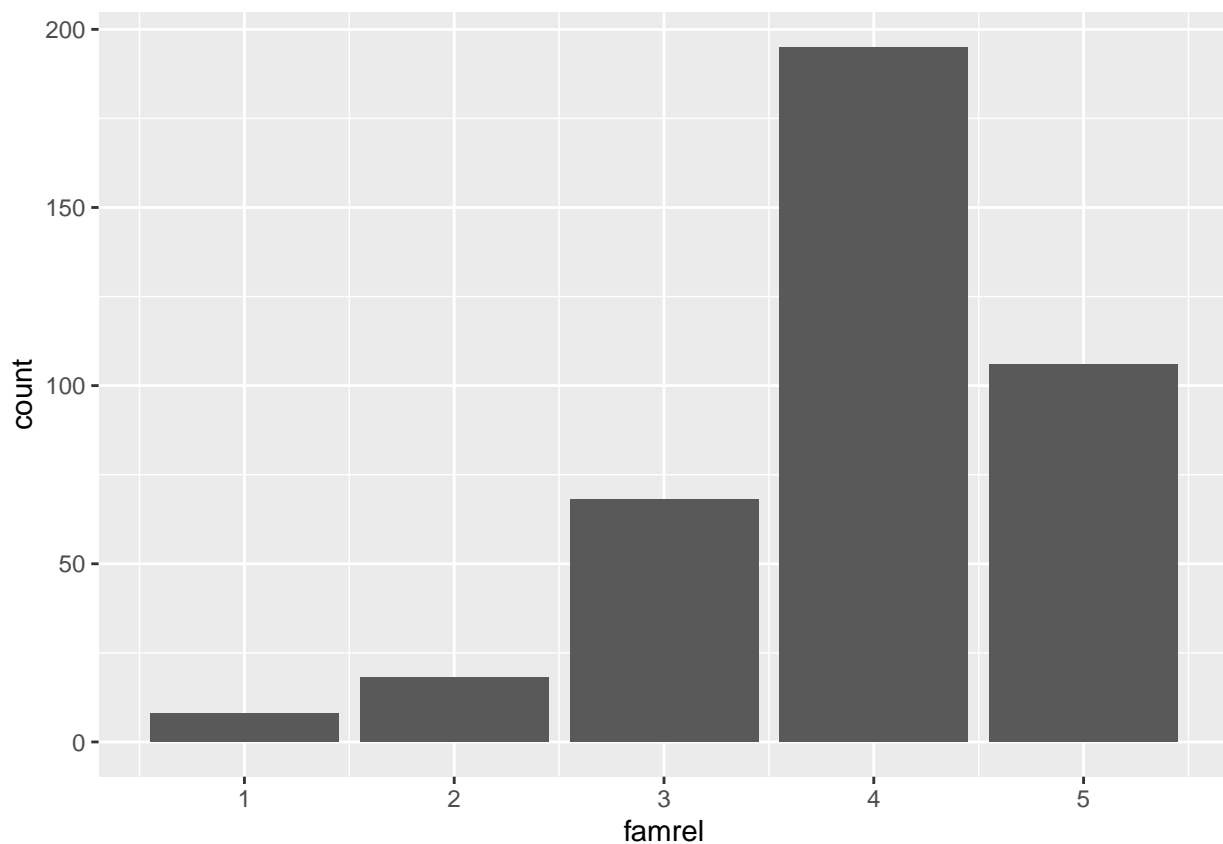
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5067  -2.3511   0.5452   3.4933   9.6489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.24739    0.78330  13.082  <2e-16 ***
## freetime     0.05186    0.23135   0.224    0.823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.587 on 393 degrees of freedom
## Multiple R-squared:  0.0001279, Adjusted R-squared:  -0.002416
## F-statistic: 0.05025 on 1 and 393 DF, p-value: 0.8227
```

Not significant!

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()
```



```
famrel.model <- lm(G3 ~ famrel, data = student_mat)
summary(famrel.model)
```

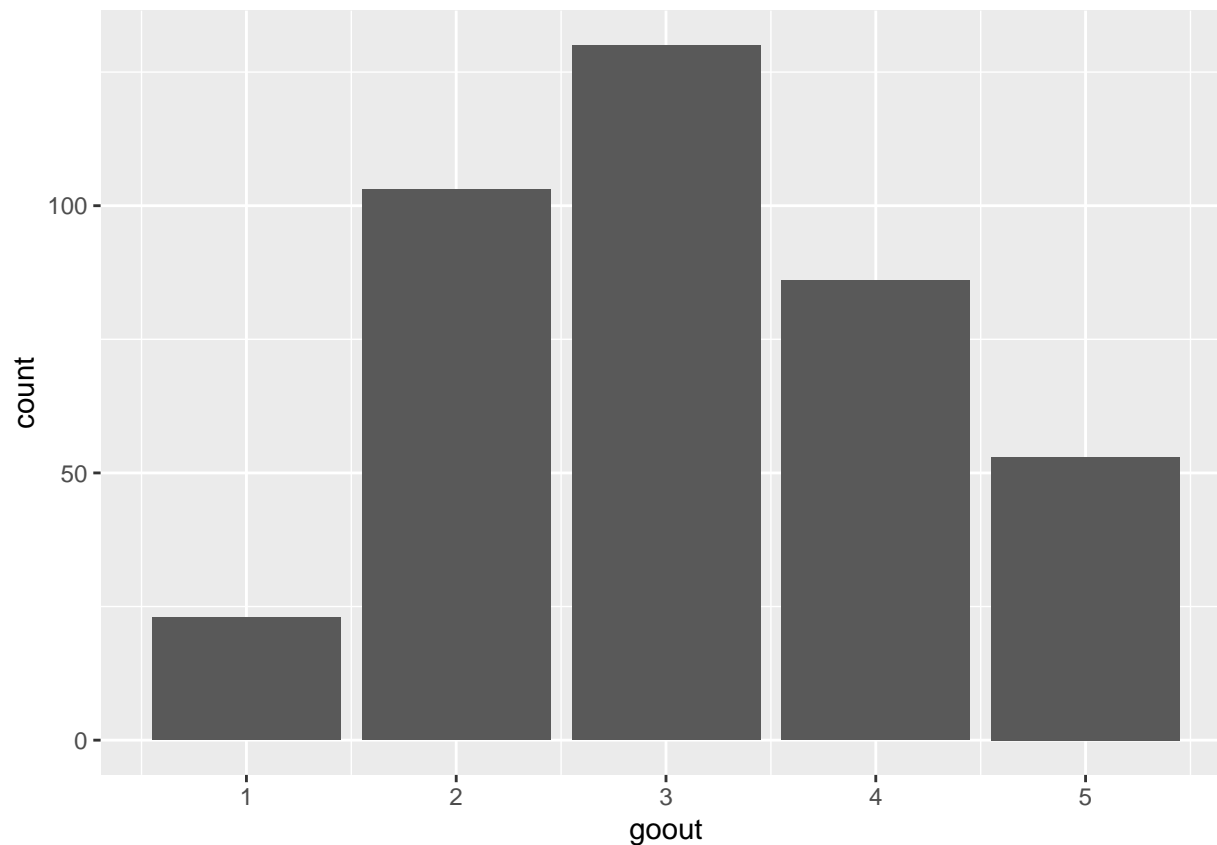
```
##
## Call:
## lm(formula = G3 ~ famrel, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6922  -2.1674   0.5702   3.3078   9.5702
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.3800     1.0411    9.01  <2e-16 ***
## famrel        0.2624     0.2574    1.02   0.309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.581 on 393 degrees of freedom
## Multiple R-squared:  0.002638,    Adjusted R-squared:  0.0001004
## F-statistic:  1.04 on 1 and 393 DF,  p-value: 0.3086
```

No significance

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()
```



```
goout.model <- lm(G3 ~ goout, data = student_mat)
summary(goout.model)
```

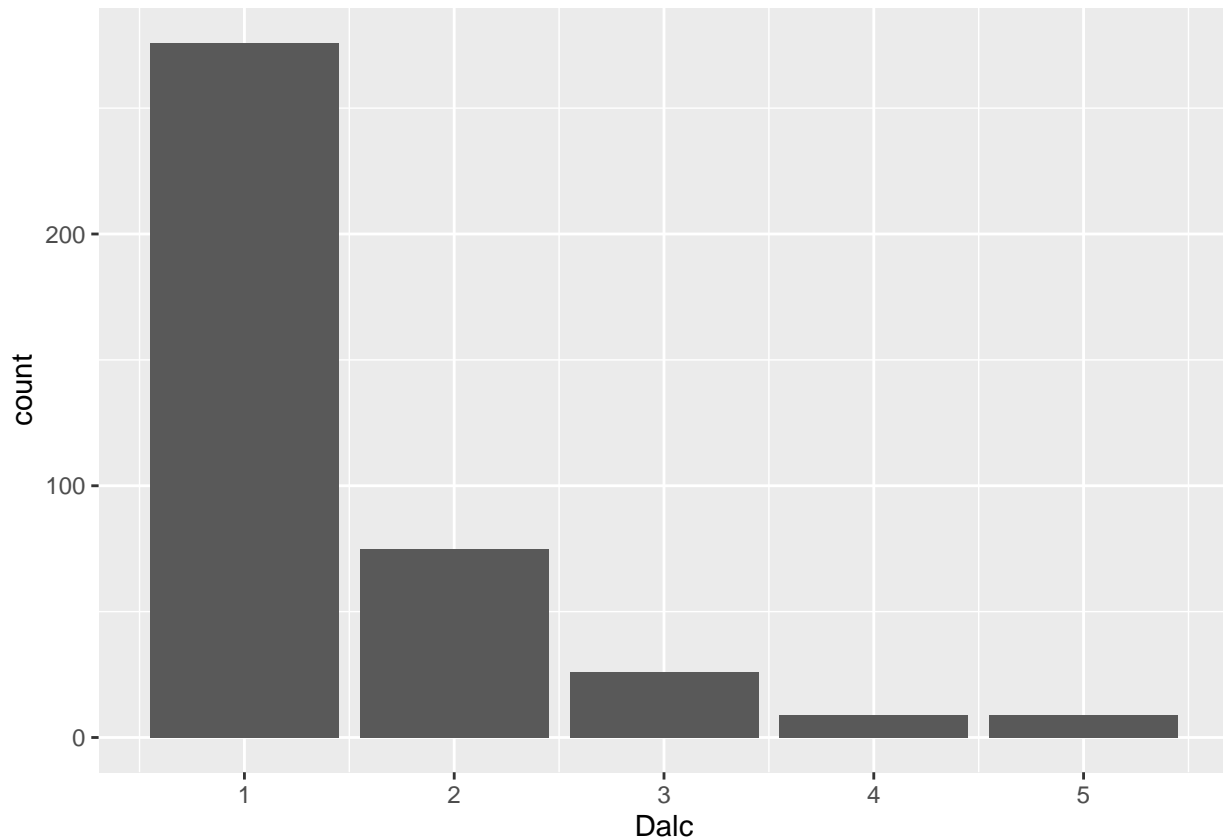
```
##
## Call:
## lm(formula = G3 ~ goout, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.5676  -1.9282   0.4324   3.0718   9.0718
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.1141     0.6793  17.833 < 2e-16 ***
## goout        -0.5465     0.2057  -2.656  0.00823 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.547 on 393 degrees of freedom
## Multiple R-squared:  0.01763,    Adjusted R-squared:  0.01513
## F-statistic: 7.054 on 1 and 393 DF,  p-value: 0.008229
```

Significance!!

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()
```



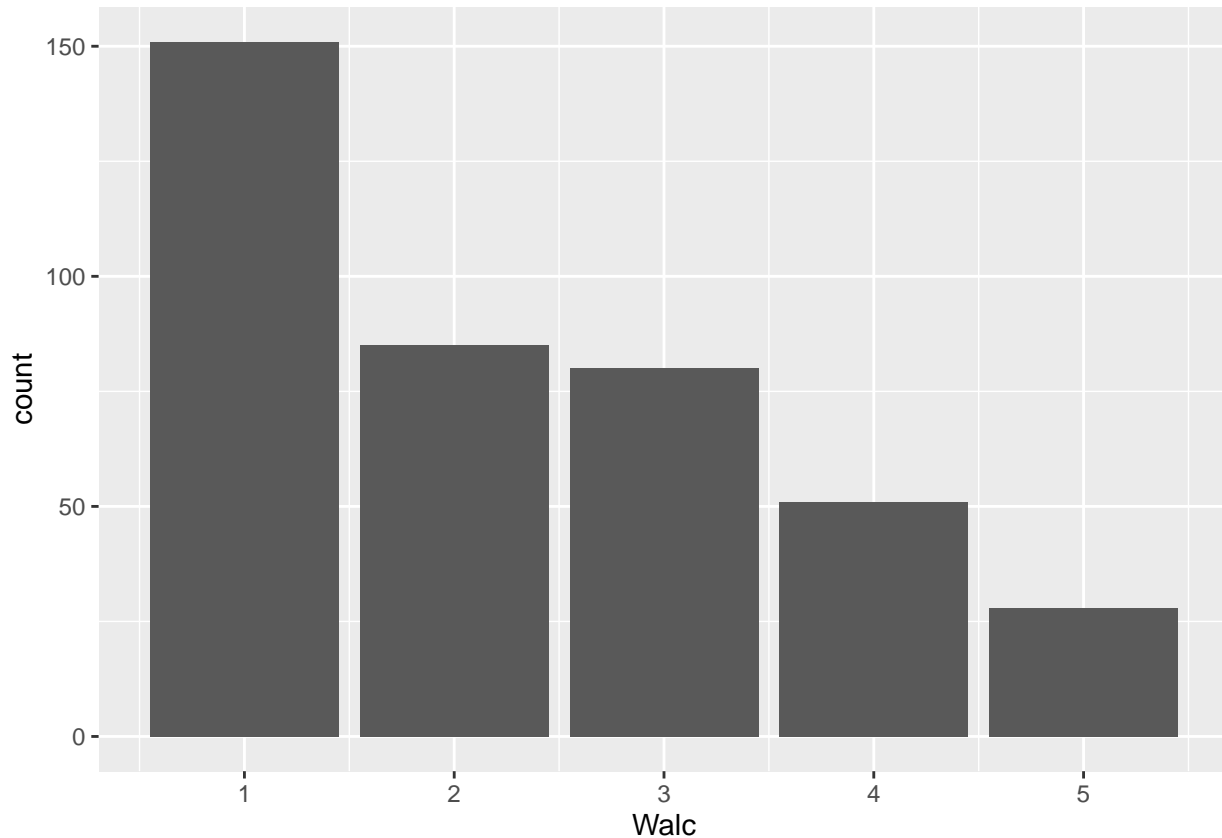
```
Dalc.model <- lm(G3 ~ Dalc, data = student_mat)
summary(Dalc.model)
```

```
##
## Call:
## lm(formula = G3 ~ Dalc, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5504  -1.9881   0.4496   3.4496   9.4496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.8316     0.4476  24.201  <2e-16 ***
## Dalc         -0.2811     0.2591  -1.085   0.278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 4.58 on 393 degrees of freedom
## Multiple R-squared:  0.002988,    Adjusted R-squared:  0.0004508
## F-statistic: 1.178 on 1 and 393 DF,  p-value: 0.2785
```

Not significant

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



```
Walc.model <- lm(G3 ~ Walc, data = student_mat)
summary(Walc.model)
```

```
##
## Call:
## lm(formula = G3 ~ Walc, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6537  -2.0071   0.3463   3.3463   9.3463
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.8385     0.4708  23.019  <2e-16 ***
## Walc         -0.1848     0.1792  -1.031   0.303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.581 on 393 degrees of freedom
## Multiple R-squared:  0.002698,    Adjusted R-squared:  0.00016
## F-statistic: 1.063 on 1 and 393 DF,  p-value: 0.3032
```

Not significant

Academic Help

```
model <- lm(G3 ~ paid, data = student_mat)
summary(model)
```

```
##
## Call:
## lm(formula = G3 ~ paid, data = student_mat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9227  -1.9860   0.0773   3.0773  10.0140
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.9860     0.3119  32.012  <2e-16 ***
## paidyes       0.9367     0.4608   2.033   0.0428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.563 on 393 degrees of freedom
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.007885
## F-statistic: 4.131 on 1 and 393 DF,  p-value: 0.04277
```

Domestic Life

Conclusion

Why is this analysis important? Limitations of the analysis?