Effect of Socio-Economic Factors on Student's Performance

Jishant Acharya

22/03/2022

Introduction :

School has become a place to compete against the world to prove yourself and set yourself apart from the crowd to excel. This is done to measure everyone's grasping power with the same yardstick marks. This is in no way the right manner to judge a person's intelligence. Each individual is different, everyone comes from a varied set of backgrounds. There can never be a single yardstick that can measure each and everyone. This analysis is trying to prove that.

Dataset Description

The following is an extensive dataset that has many social factors also included in the records. There are 3 types of marks. G1, G2 and G3 are three marks that correspond to the first period, second period and the final marks. There are other factors present in the dataset that range from being related socially to being able to show economical situation of the student.

```
data_scheme <- read.csv("./Book1.csv")
kable(data_scheme,caption="Data Description")
```

Table 1: Data Description

ï..Variable	Type	Description
school	categorical	One of the two schools
sex	categorical	Male or Female
age	continuous	Age from 18 -22
address	categorical	Urban or Rural
famsize	categorical	LT3 - Less Than 3 ; GT3 - Greater than 3
Pstatus	categorical	Living 'T' (Together) or 'A' (Apart)
Medu	categorical	Level of Education (5 levels - From 0 to 4)
Fedu	categorical	Level of Education (5 levels - From 0 to 4)
Mjob	categorical	Types of Jobs
Fjob	categorical	Types of Jobs
traveltime	categorical	Level of Travel (4 levels - From 1 - 4)
studytime	categorical	Level of Study Time (4 levels - From 1-4)
failure	categorical	Past failures (n if 1<=n<3 or 4
schoolsup	categorical	Yes / No for support from school
famsup	categorical	Yes / No for support from family
paid	categorical	Yes / No for paid classes
nursery	categorical	Yes / No for nursery attendance

ï..Variable	Type	Description
internet	categorical	Yes / No for availability
goout	categorical	Level from 1-5
romantic	categorical	Yes / No from involvement in romantic activities
freetime	categorical	Level of Free time from 1-5
health	categorical	Quality of health from 1-5
G1	continuous	Marks for first period
G2	continuous	Marks for second period
G3	continuous	Marks for third period

```
head(student.mat,3)
```

```
##   school sex age address famsize Pstatus Medu Fedu    Mjob    Fjob traveltime
## 1     GP   F  18       U     GT3       A    4    4 at_home teacher          2
## 2     GP   F  17       U     GT3       T    1    1 at_home   other          1
## 3     GP   F  15       U     LE3       T    1    1 at_home   other          1
##   studytime failures schoolsup famsup paid nursery internet romantic famrel
## 1         2        0       yes     no   no     yes       no       no      4
## 2         2        0        no    yes   no      no      yes       no      5
## 3         2        3       yes     no  yes     yes      yes       no      4
##   freetime goout health absences G1 G2 G3
## 1        3     4      3        6  5  6  6
## 2        3     3      3        4  5  5  6
## 3        3     2      3       10  7  8 10
```

Analysis

Gender Distribution

```
ggplot(student.mat, aes(x=sex)) + geom_bar(stat = "count",width = 0.5,fill=primary)
```

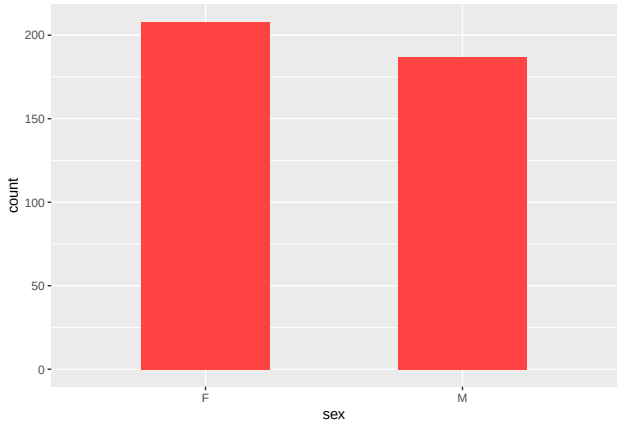


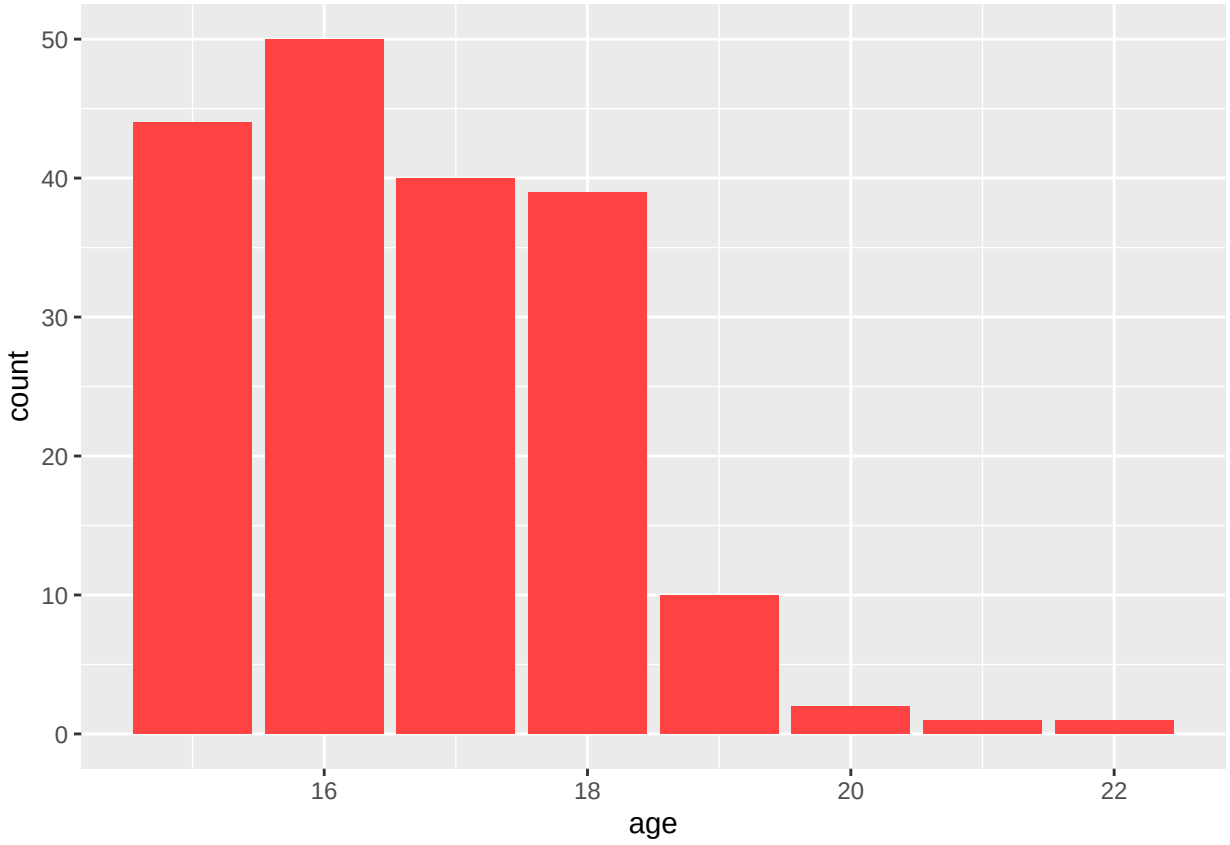
Figure 1: Population distribution between genders

This graph shows that there is no class bias in the dataset, the number of male is almost similar to the number of females.

** Frequency Distribution of Male Population **

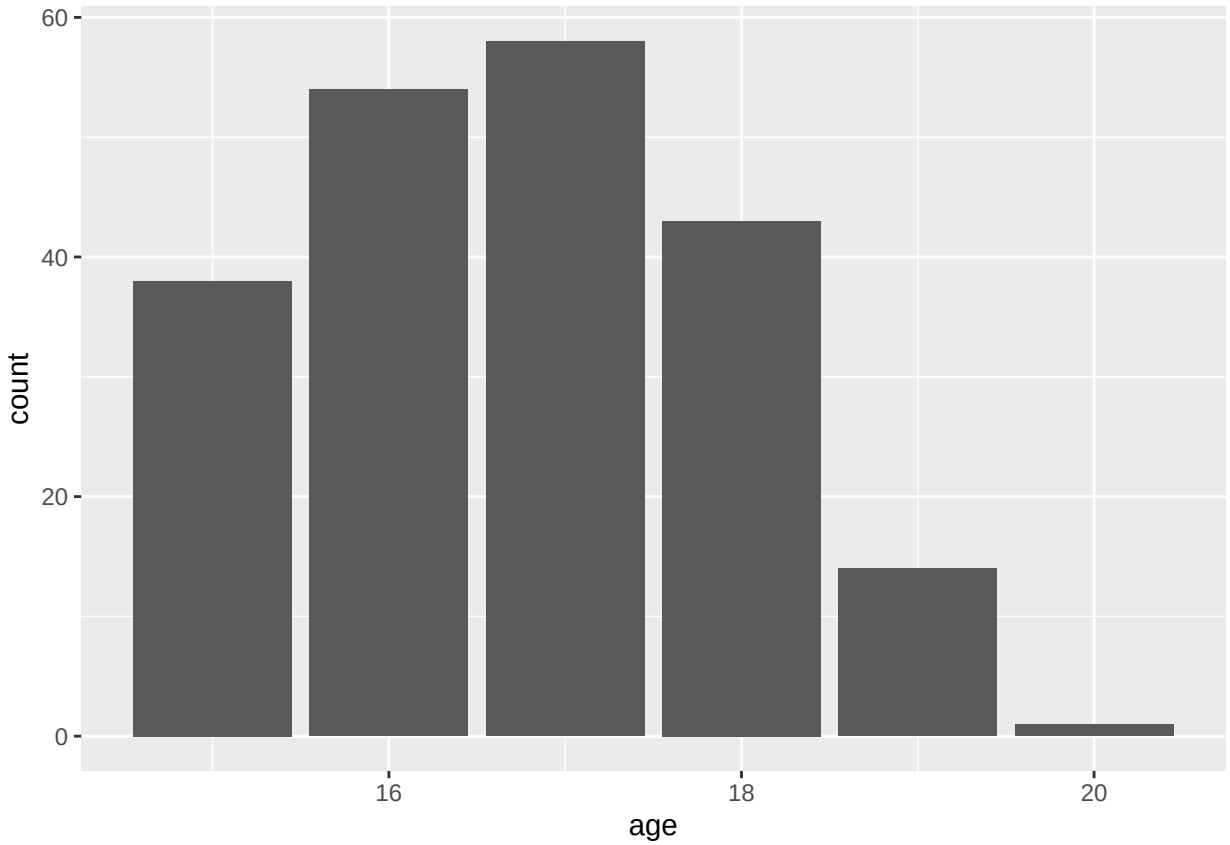
```
ggplot(student.mat[student.mat$sex == "M",], aes(age)) + geom_histogram(fill=primary,stat = "count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```
ggplot(student.mat[student.mat$sex == "F",], aes(age)) + geom_histogram(stat="count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```


60
40
20
0
count
16
18
20
age