High School and Beyond

Anna Sutton

2/24/24

```
hsb2 <- read.delim("hsb2.txt", header=TRUE, sep="\t")
dim(hsb2)
```

```
## [1] 200  11
```

Introduction

The data set I will be analyzing is the High School and Beyond data set. This data was conducted by the National Longitudinal Studies Program regarding the educational development of students between the years of 1980 and 1992. The data set has 200 observations and 11 variables. With this data set, I want to explore the relationship between gender and test scores, but more specifically within the subject of mathematics. Does gender influence test scores? I hypothesize that there could be a difference in scores between female and male students, due to the fact that math could potentially be viewed as male dominate class.

Univariate Exploration

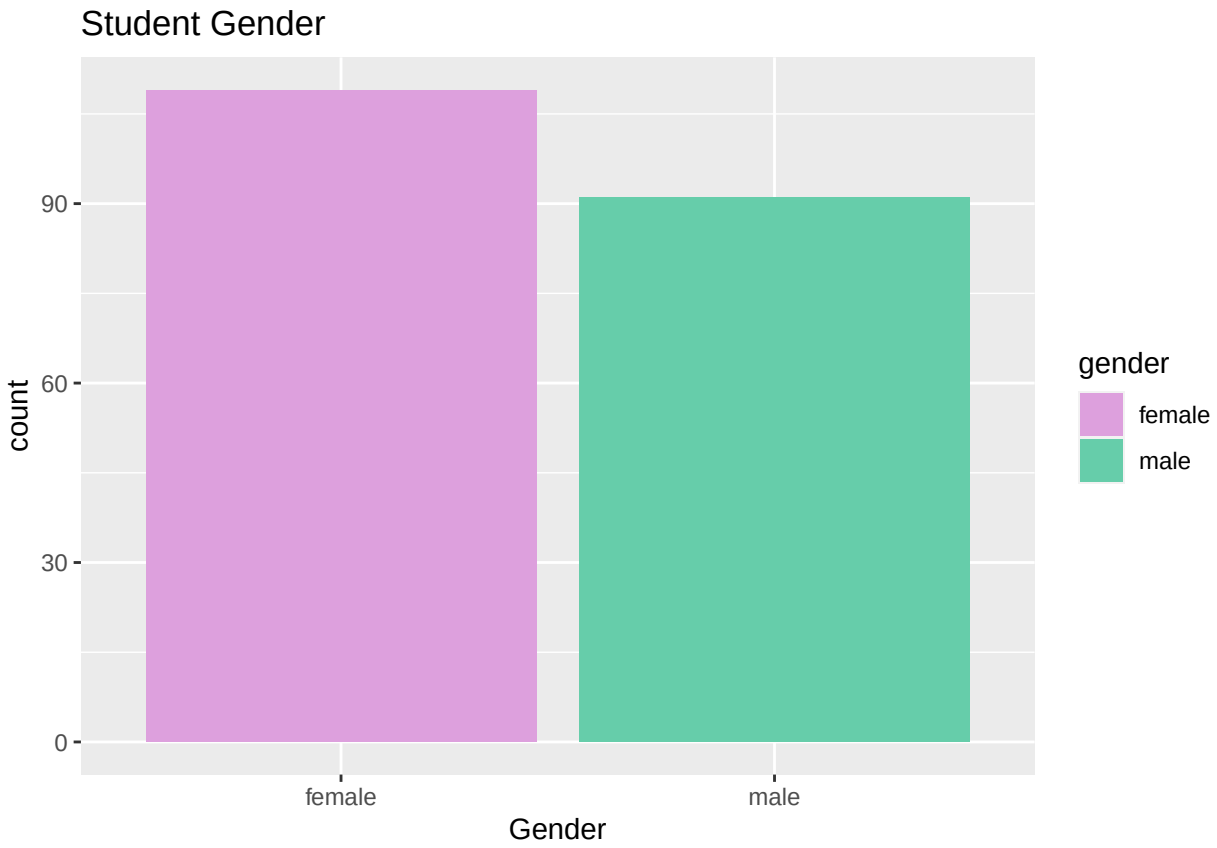
```
table(hsb2$gender)
```

```
##
## female   male
##    109     91
```

To start off, I frist wanted to see the number of male and female students by using a table. There were 109 females and 91 males that this program collected data from.

```
library(ggplot2)
```

```
ggplot(hsb2, aes(x=forcats::fct_infreq(gender), fill=gender)) + geom_bar() + xlab("Gender") + scale_fil
```



This bar graph represent both genders from the data. From the graph itself, and also the numbers from the table, it seems that the number of male and female students are close in numbers.

```
table(hsb2$math)
```

```
##
## 33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##  1  1  1  2  6 10  7  7  7  4  8  8  3  5 10  7  8  6  7 10  5  7 13  6  2  5
## 61 62 63 64 65 66 67 68 69 70 71 72 73 75
##  7  4  5  5  3  4  2  1  2  1  4  3  1  2
```

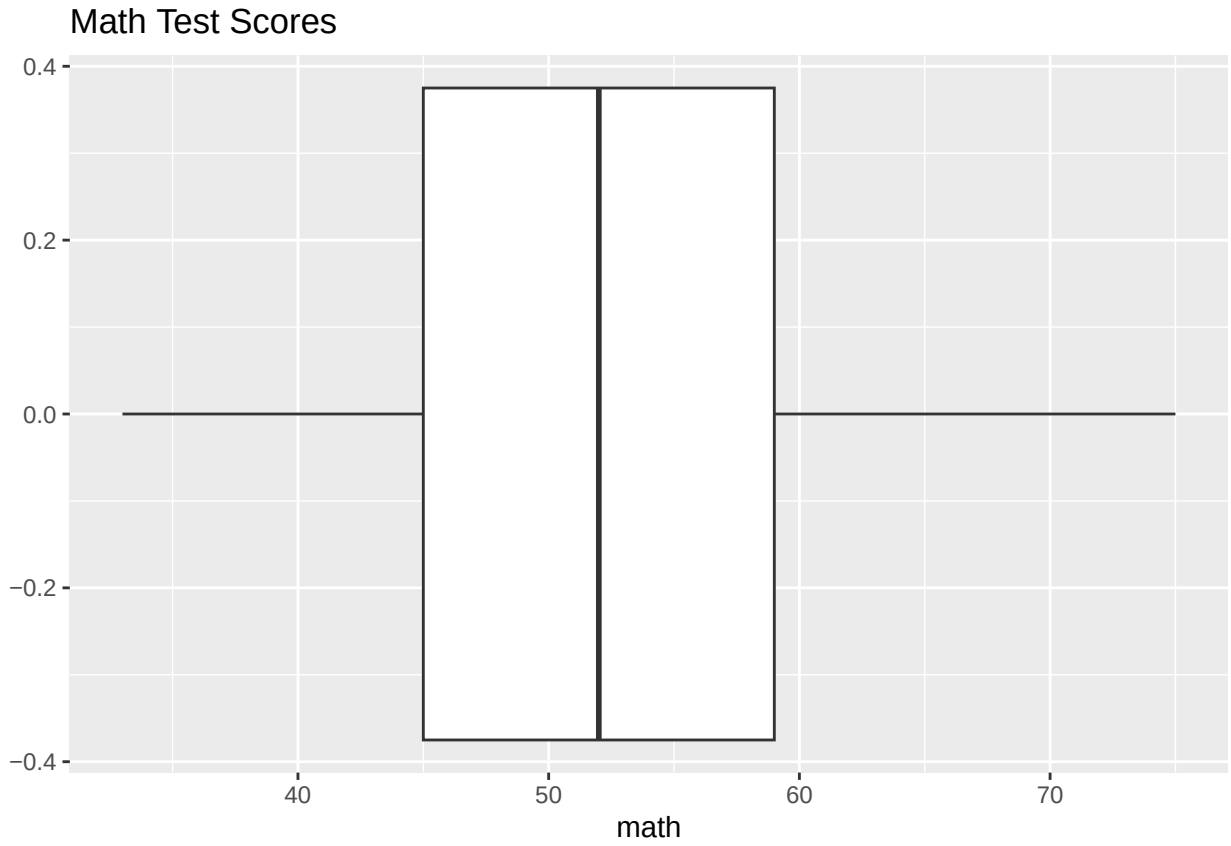
I then went on to examine the math scores from all the students within the math class by using another table.

```
summary(hsb2$math)
```

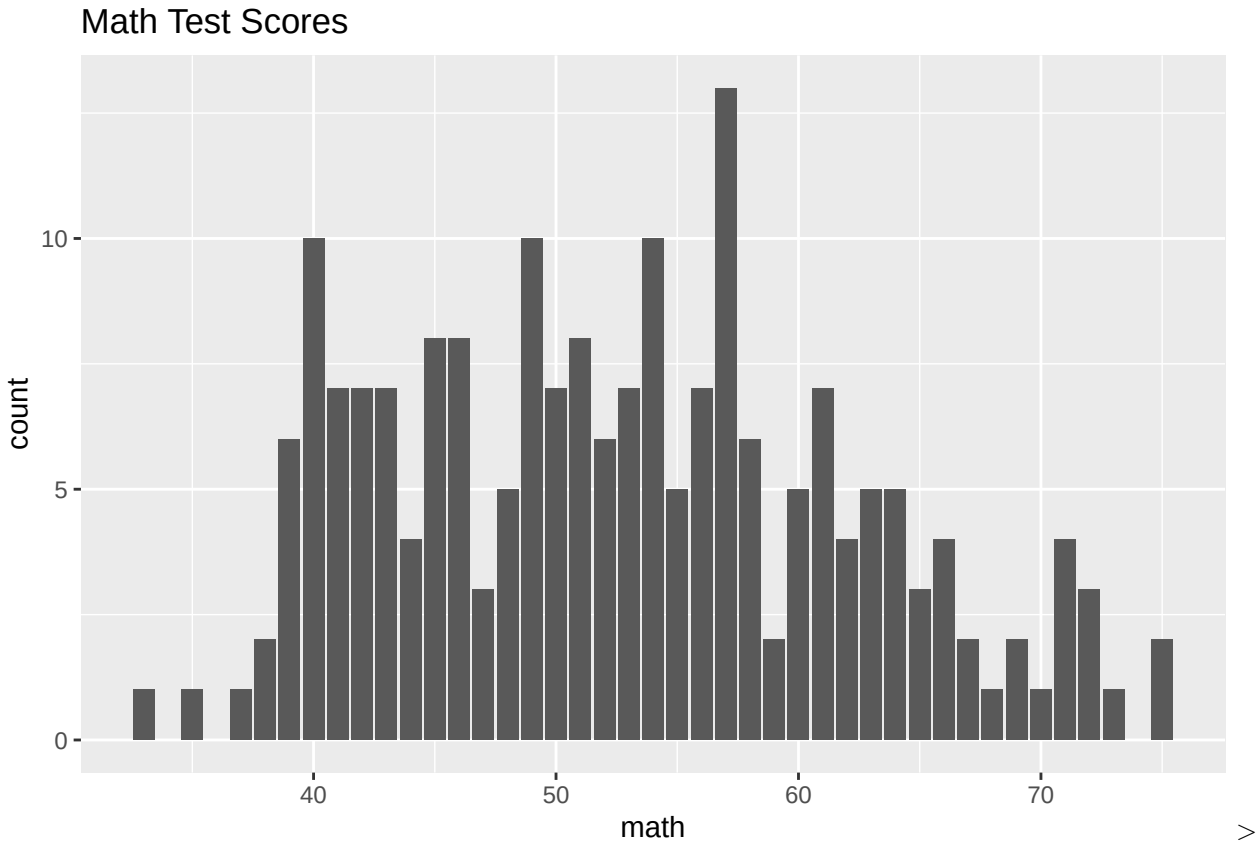
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   33.00   45.00   52.00   52.65   59.00   75.00
```

Next, I summarized the data of the math test scores. This data is telling us how the students scored on their math tests. Therefore the average of the test scores was 52.65, and the heigest score was a 75 and the lowest was a 33.

```
ggplot(hsb2, aes(x=math)) + geom_boxplot() + ggtitle("Math Test Scores")
```




```
ggplot(hsb2, aes(x=math)) + geom_bar() + ggtitle("Math Test Scores")
```



The boxplot and the Bar graph are both representations of the math test scores from both male and female students. From this data we are able see that most of the test scores lie in the middle, somwhat between 40 and 60.

Bivariate Exploration

```
table(hsb2$gender, hsb2$math)
```

```
##
##          33 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
##   female  1  0  1  1  2  6  4  5  4  2  4  4  1  3  5  4  3  4  7  5  4  6  4
##   male    0  1  0  1  4  4  3  2  3  2  4  4  2  2  5  3  5  2  0  5  1  1  9
##
##          58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 75
##   female  3  0  3  3  2  2  3  3  2  2  0  2  0  1  3  0  0
##   male    3  2  2  4  2  3  2  0  2  0  1  0  1  3  0  1  2
```

?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
```

```
##     intersect, setdiff, setequal, union
```

```
hsb2_male <- filter(hsb2, gender=="male")

hsb2_female <- filter(hsb2, gender=="female")
```

```
dim(hsb2_male)
```

```
## [1] 91 11
```

```
dim(hsb2_female)
```

```
## [1] 109  11
```

```
dim(hsb2)
```

```
## [1] 200  11
```

```
mean(hsb2_male$math)
```

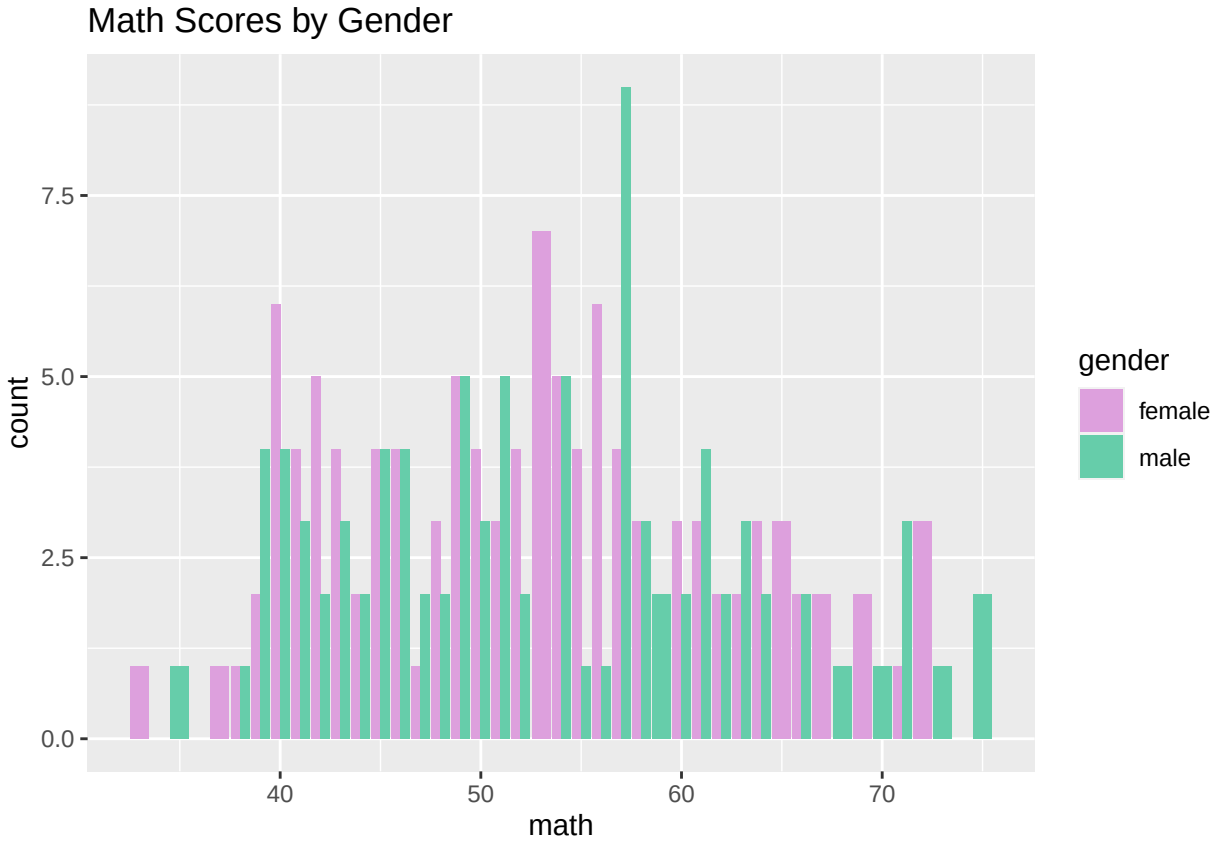
```
## [1] 52.94505
```

```
mean(hsb2_female$math)
```

```
## [1] 52.3945
```

I calculated the mean of the test scores of the male and female students. From this calculation we are able to see that the averages between both geneders are quite similar.

```
ggplot(hsb2, aes(x=math, fill=gender)) + geom_bar(position ="dodge") + scale_fill_manual(values=c("plum
```



This graph represents the math test scores by gender and shows how similar the scores were between genders. From this graph, we could also see which students had higher and lower scores.

Conclusion

Overall, from the observed data I noticed that gender does not influence the math test scores. Even thought we did not know the total amount of points of the test, we were still able to see the relasthip between the two variables. I was not too surprised that the test scores were similar between males and females because the test scores can also depend on other factors like comprehension or study time. I would also say that this data did not support my prior hypothesis, because the data displayed that gender does not impact test scores.