High School and Beyond Analysis

Aidan Delaney

2022-09-26

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

```
library(ggplot2)
```

The High School and Beyond Longitudinal Study was established to study the educational, vocational, and personal development of young people. This second study by the NCES' National Longitudinal Studies Program covered more than 30,000 high school seniors and 28,000 high school sophomores. With the information from 200 students provided in the High School and Beyond dataset, this analysis aims to compare student gender and math test performance.

VARIABLE ANALYSIS GENDER RATIO

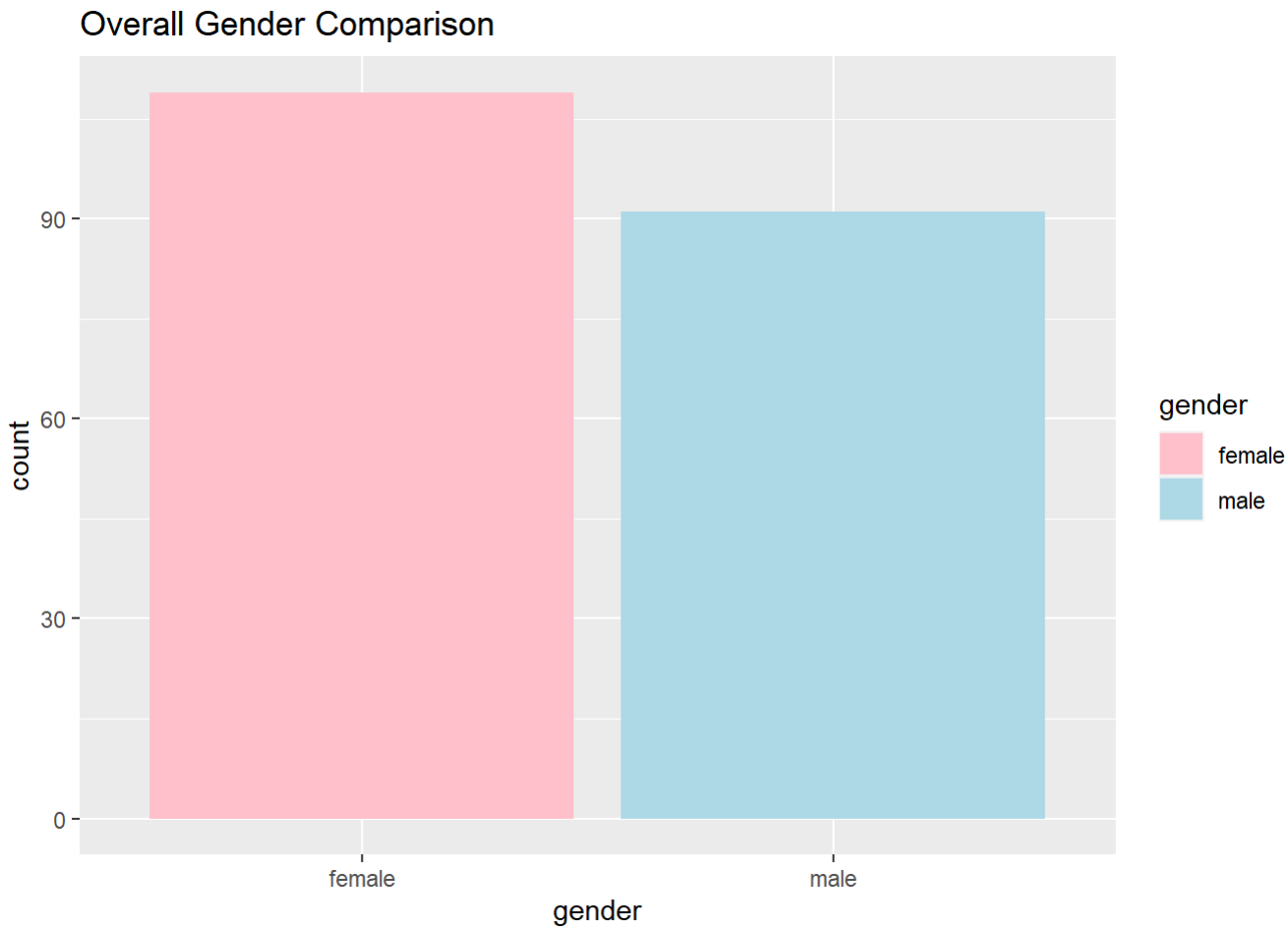
```
table(hsb2$gender)
```

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

```
prop.table(table(hsb2$gender))
```

```
##
## female   male
##  0.545  0.455
```

```
ggplot(hsb2, aes(x=gender, fill=gender)) + geom_bar() + scale_fill_manual(values=c("pink", "ligh
tblue")) + ggtitle("Overall Gender Comparison")
```



There are 109 female and 91 male students in our dataset of 200.This means that approximately 54% are female, and 46% are male, a roughly equal gender ratio.

Survey responses collected scores in standardized testing. Here, we will examine math test scores. MATH TEST SCORES

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

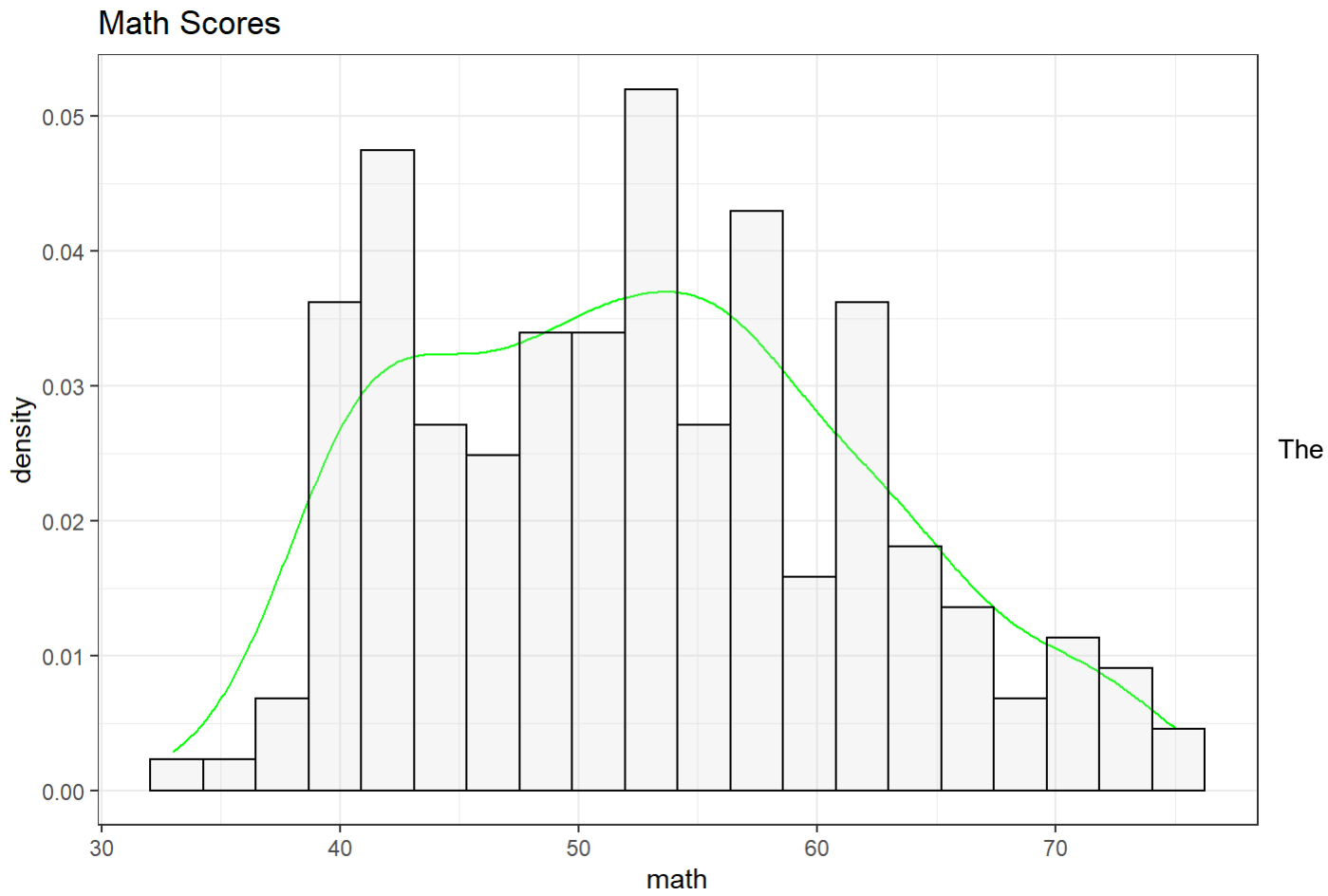
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   33.00   45.00   52.00   52.65   59.00   75.00
```

```
sd(hsb2$math)
```

```
## [1] 9.368448
```

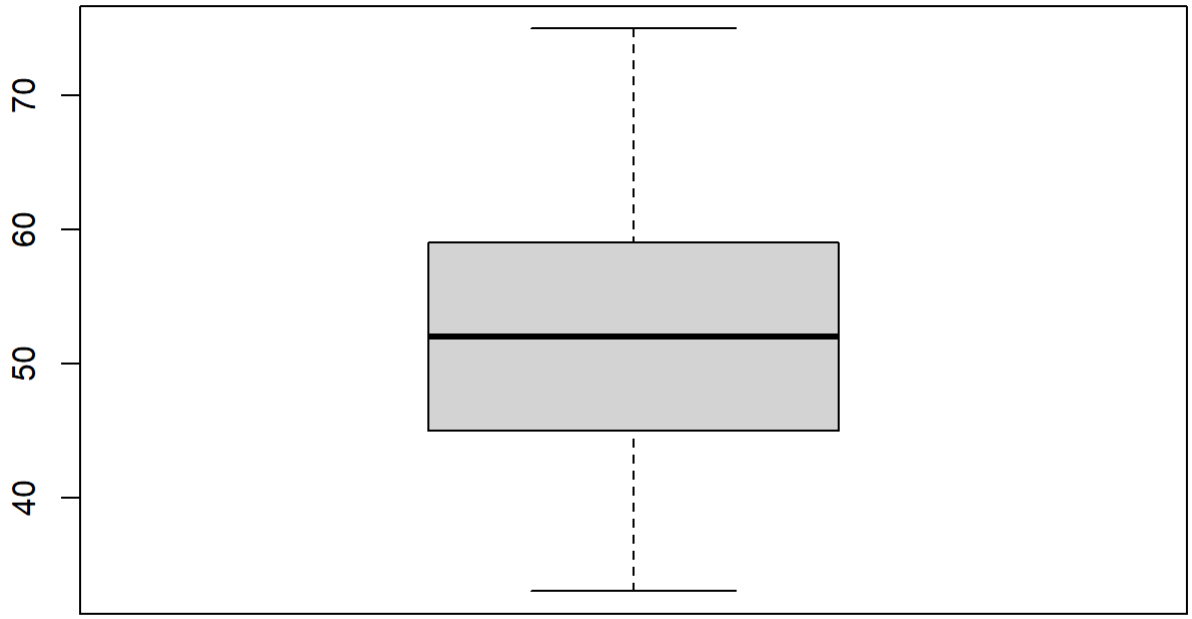
As seen by the tables data, math test scores range from thirty three to seventy five, with a median score of 52 and mean of 52.65.

```
ggplot(hsb2, aes(x=math)) + geom_density(col="green", alpha = 1) + geom_histogram(aes(y=..densit
y..), bins = 20, col="black", fill= "light grey", alpha = .2) +  ggtitle("Math Scores") + theme_
bw()
```



The histogram and density plot for Math Scores appears slightly skewed to the left, and is unimodal.

```
boxplot(hsb2$math)
```



The boxplot of the math scores show that there are no significant outliers, and that the scores are roughly equal about the median.

BIVARIATE COMPARISON

```
ggplot(hsb2, aes(x=math, col=gender)) + geom_density() + ggtitle("Density Plots of Math Scores")
+ theme_bw()
```

The density plot comparison between genders shows a similar relationship as compared to math test scores. The female density appears unimodal, while the male density is more bimodal about the median.

Perhaps a closer look is needed to determine which gender is doing better overall on Math Tests.

```
female_math <- filter(hsb2, gender == "female")
summary(female_math$math)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   33.00   45.00   53.00   52.39   58.00   72.00
```

```
sd(female_math$math)
```

```
## [1] 9.151015
```

```
male_math <- filter(hsb2, gender == "male")
summary(male_math$math)
```

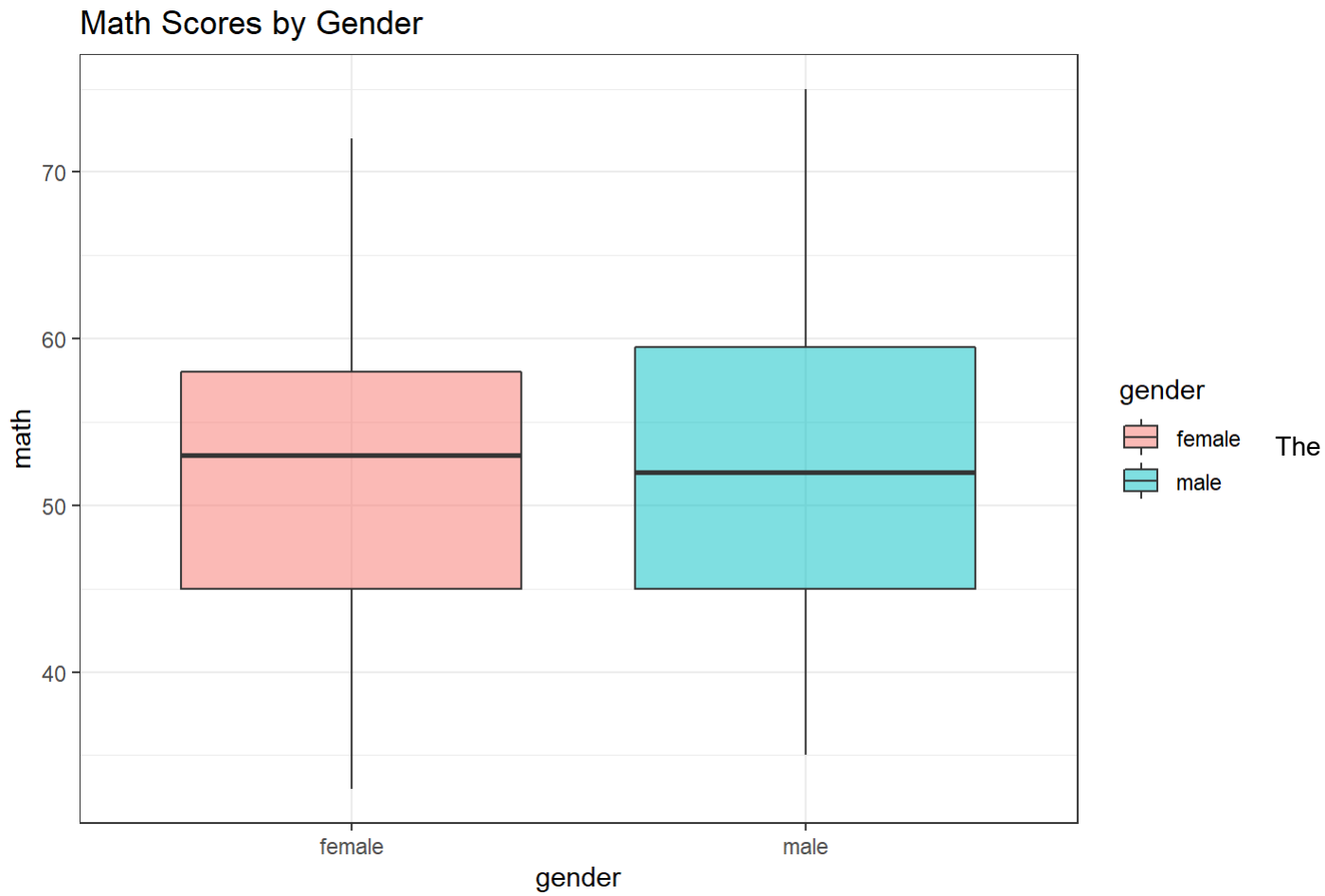
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.00   45.00   52.00   52.95   59.50   75.00
```

```
sd(male_math$math)
```

```
## [1] 9.664784
```

Female math scores obtained a median of 53.00 and a mean of 52.39, with a standard deviation of 9.15. Male math scores obtained a median of 52.00 and a mean of 52.95, with a standard deviation of 9.66.

```
ggplot(hsb2, aes(x= gender, y=math, fill = gender)) + geom_boxplot(alpha = .5) + ggtitle("Math Scores by Gender") + theme_bw()
```



The grouped box plot displays the similarity in math test scores by gender. The interqartile range is larger for males, with the encompassed scores being only slightly higher on average than female scores.

CONCLUSIONS

FEMALE MATH MEAN :52.39 FEMALE MATH MEDIAN : 53.00

MALE MATH MEAN : 52.95 MALE MATH MEDIAN : 52.00

After analyzing math test scores compared to gender, I can conclude that neither gender did significantly better compared to the other's math test scores.