

EDA_CalebCamara

2024-03-05

Introduction

The High school and Beyond data presents vast array of highschool students along with their scores in a variety of subjects, race, socioeconomic status, and school type. The two main data points that I will be analyzing will be the school type and the math test scores. The reasoning I chose these two data points is to see if the education system at a private school proves to be a better educational choice for parents that want the best education for their kids. School type and math

```
hsb2 <- read.table("C:/Users/jeffr/Downloads/Math130/hsb2.txt", header=TRUE, sep="\t")
library(ggplot2)
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(RColorBrewer)
```

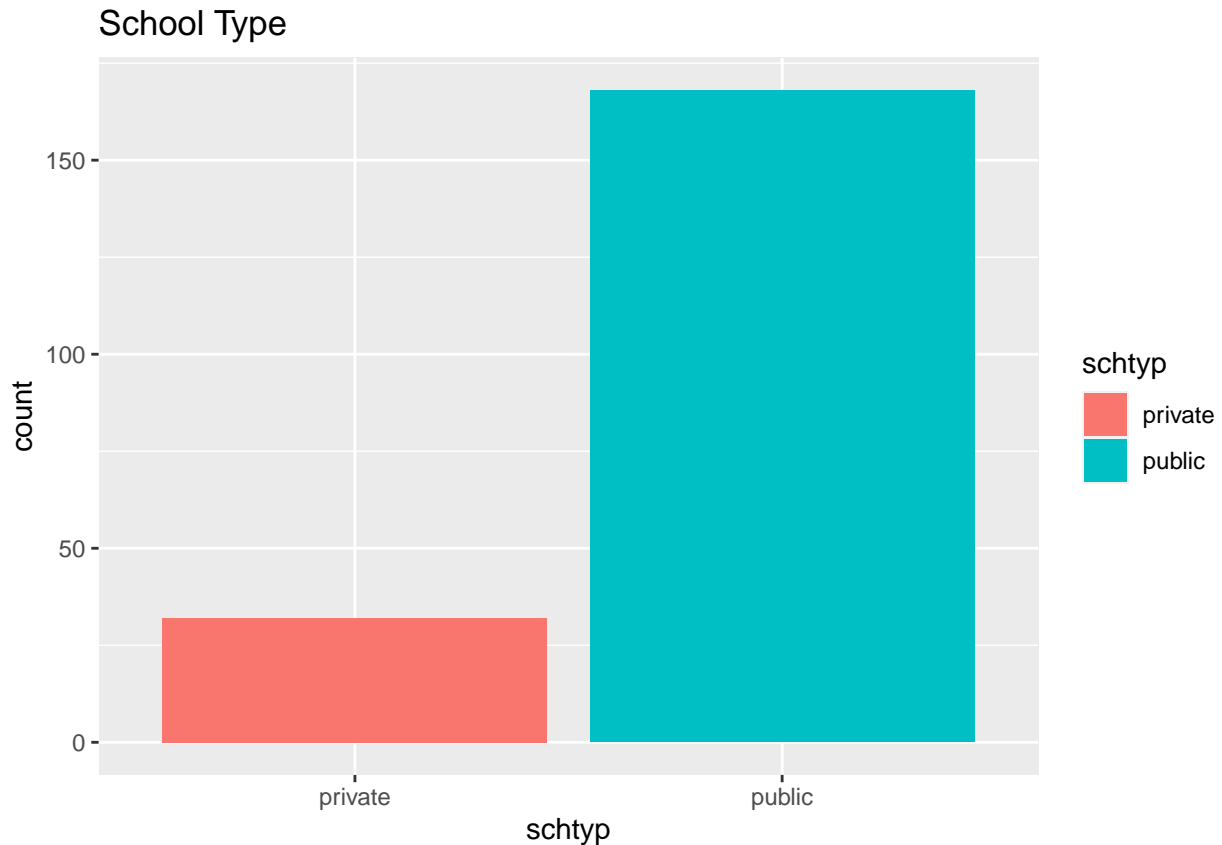
Univariate Data

School Type

```
table(hsb2$schtyp)
```

```
##
## private  public
##      32      168
```

```
ggplot(hsb2, aes(x=schtyp, fill=schtyp)) + geom_bar() + ggtitle("School Type")
```



This bar chart represents the highschool students that are either in a private school (32) or a public school (168)

Math Scores

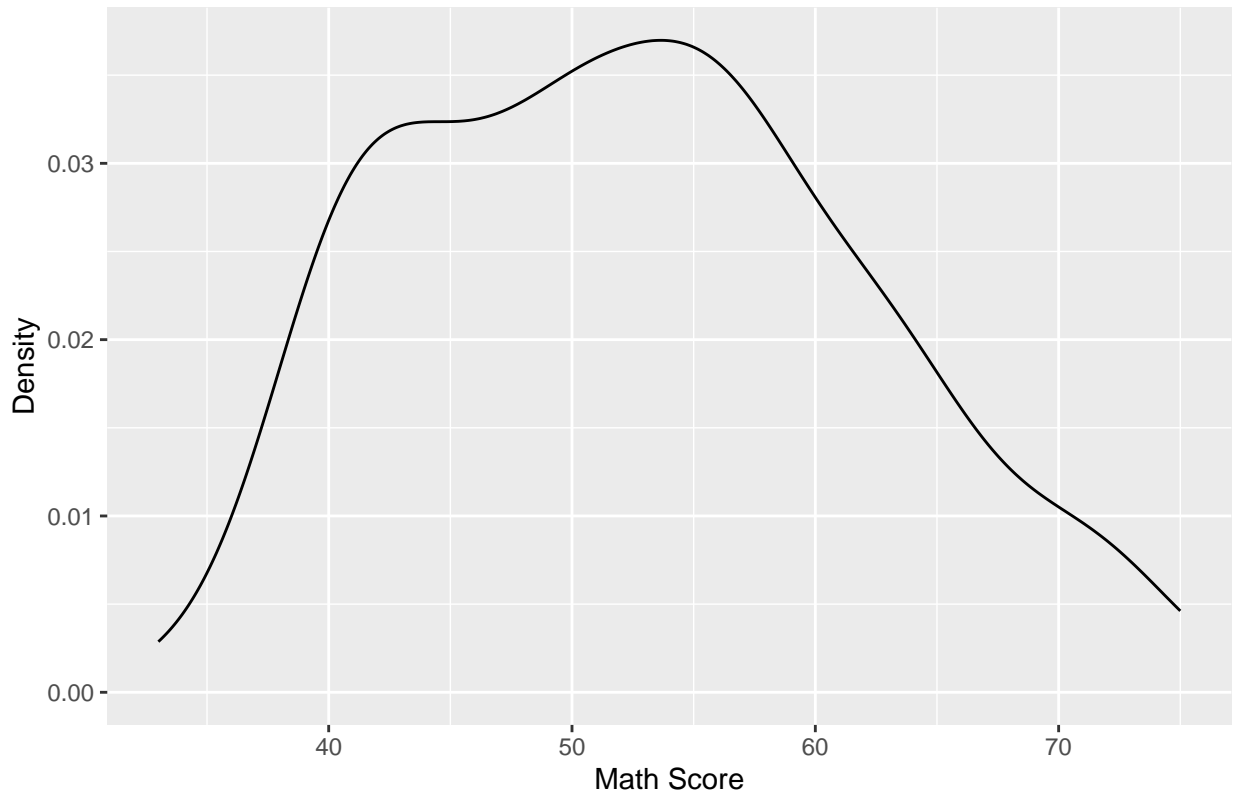
```
table(hsb2$math)
```

```
##
## 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
```

```
ggplot(hsb2, aes(x=math, fill=math)) + geom_density() + ggtitle("Math Scores") + xlab("Math Score") + ylab("Density")
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
##   variable into a factor?
```

Math Scores

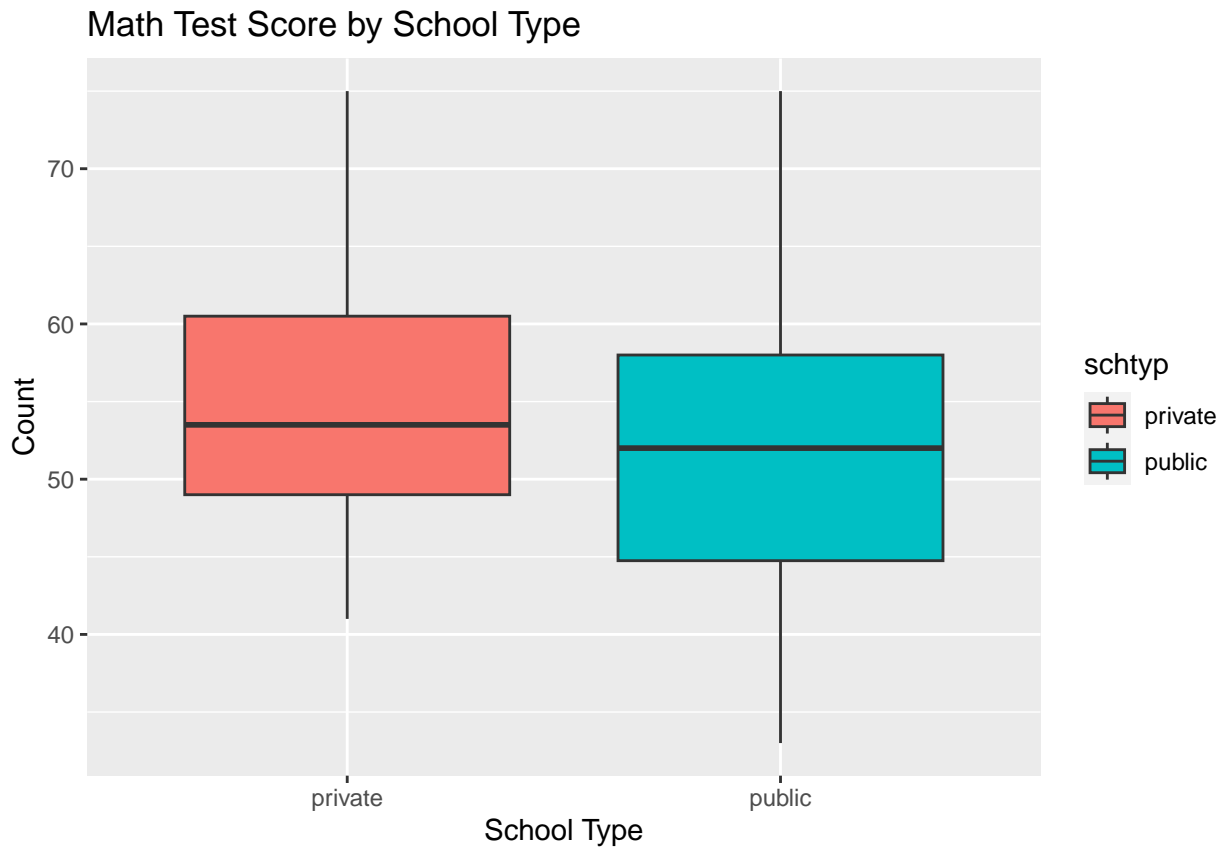


Bivariate Data

```
table(hsb2$schtyp, hsb2$math) %>% prop.table(margin=1) %>% round (5)
```

```
##
##           33      35      37      38      39      40      41      42
## private 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.03125 0.03125
## public  0.00595 0.00595 0.00595 0.01190 0.03571 0.05952 0.03571 0.03571
##
##           43      44      45      46      47      48      49      50
## private 0.06250 0.00000 0.03125 0.03125 0.00000 0.03125 0.09375 0.09375
## public  0.02976 0.02381 0.04167 0.04167 0.01786 0.02381 0.04167 0.02381
##
##           51      52      53      54      55      56      57      58
## private 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.09375 0.00000
## public  0.04167 0.02976 0.03571 0.05357 0.02381 0.03571 0.05952 0.03571
##
##           59      60      61      62      63      64      65      66
## private 0.00000 0.06250 0.00000 0.03125 0.09375 0.00000 0.00000 0.00000
## public  0.01190 0.01786 0.04167 0.01786 0.01190 0.02976 0.01786 0.02381
##
##           67      68      69      70      71      72      73      75
## private 0.00000 0.00000 0.06250 0.00000 0.03125 0.00000 0.00000 0.03125
## public  0.01190 0.00595 0.00000 0.00595 0.01786 0.01786 0.00595 0.00595
```

```
ggplot(hsb2, aes(x=schtyp, y=math, fill=schtyp)) + geom_boxplot() + ggtitle("Math Test Score by School Type")
```



This table shows the density with the averages included

Conclusion

According to the data the private high school students did end up having a higher overall average, but in terms of highest score, there was no difference. The public school students did have lower scores than the private students did. That being said, the scores represented in the data does not prove that kids who attend a private school will end up having a better test scores.