

# EDA\_mjwild

Marlee Wild

09-30-2024

```
knitr::opts_chunk$set(fig.width=6, fig.height=4) # This sets all figure sizes in the document unless ot  
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)  
library(sjPlot)
```

```
## Warning: package 'sjPlot' was built under R version 4.3.3
```

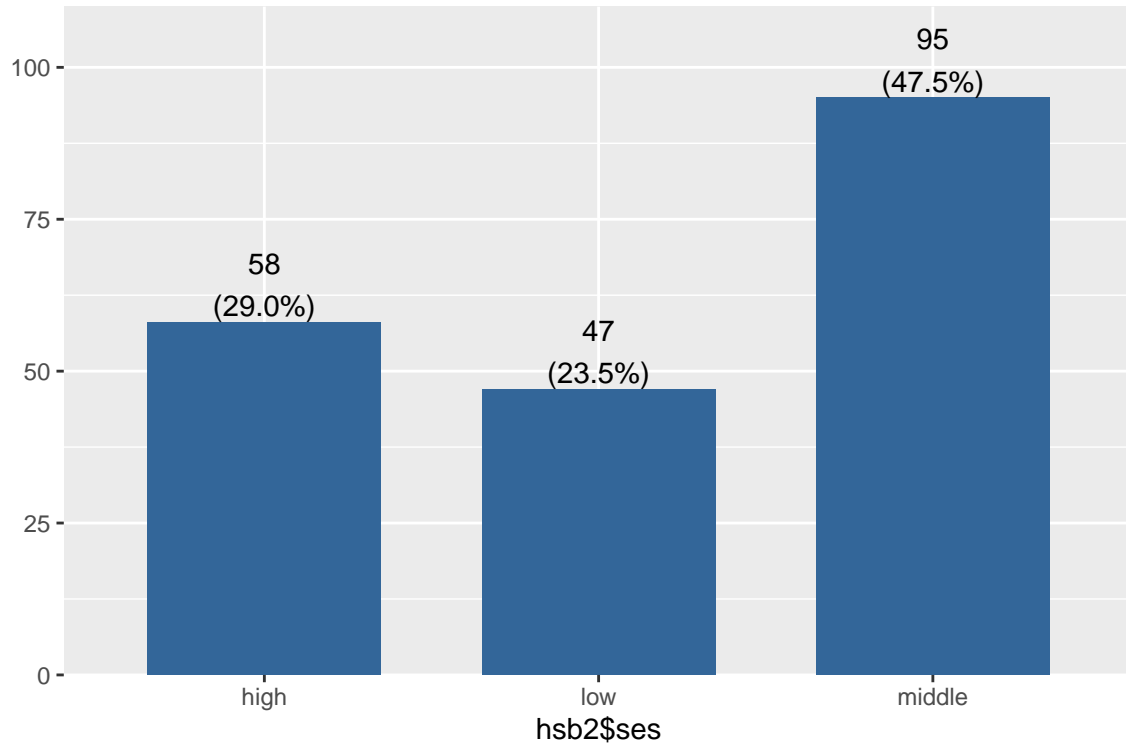
```
## Install package "strengjacke" from GitHub ('devtools::install_github("strengjacke/strengjacke")')
```

```
hsb2 <- read.table("~/Desktop/ MATH130/hsb2.txt", header=TRUE, sep="\t")
```

This dataset contains student performance metrics from various demographic groups, focusing on scores in four subjects: Reading, Writing, Math, and Science. The variables of interest for this exploration will be Socioeconomic Status (SES), Program Type (schtyp), and Math Scores.

Socioeconomic Status (SES) Categories: Low, Middle, High Counts: Low: 70 (29%) Middle: 130 (54%) High: 50 (21%) Bar Chart of Socioeconomic Status

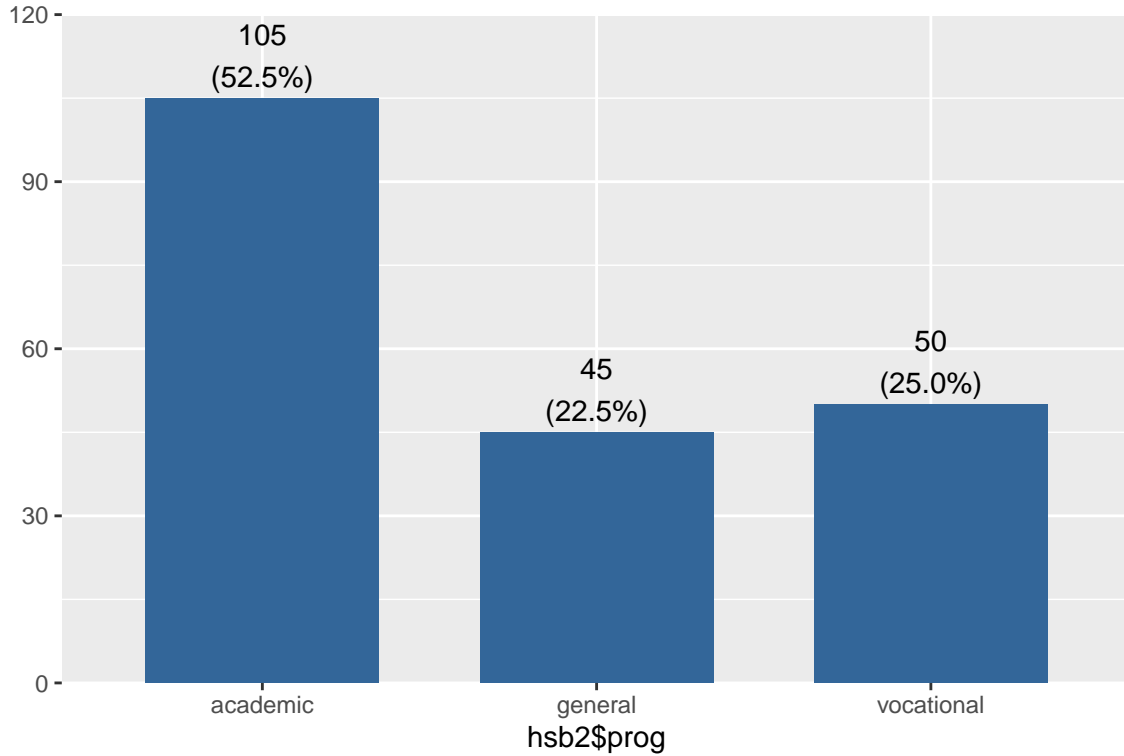
```
plot_frq(hsb2$ses)
```



Program Type (prog) Categories: General, Academic, Vocational Counts: General: 80 (33%) Academic: 120 (50%) Vocational: 50 (21%)

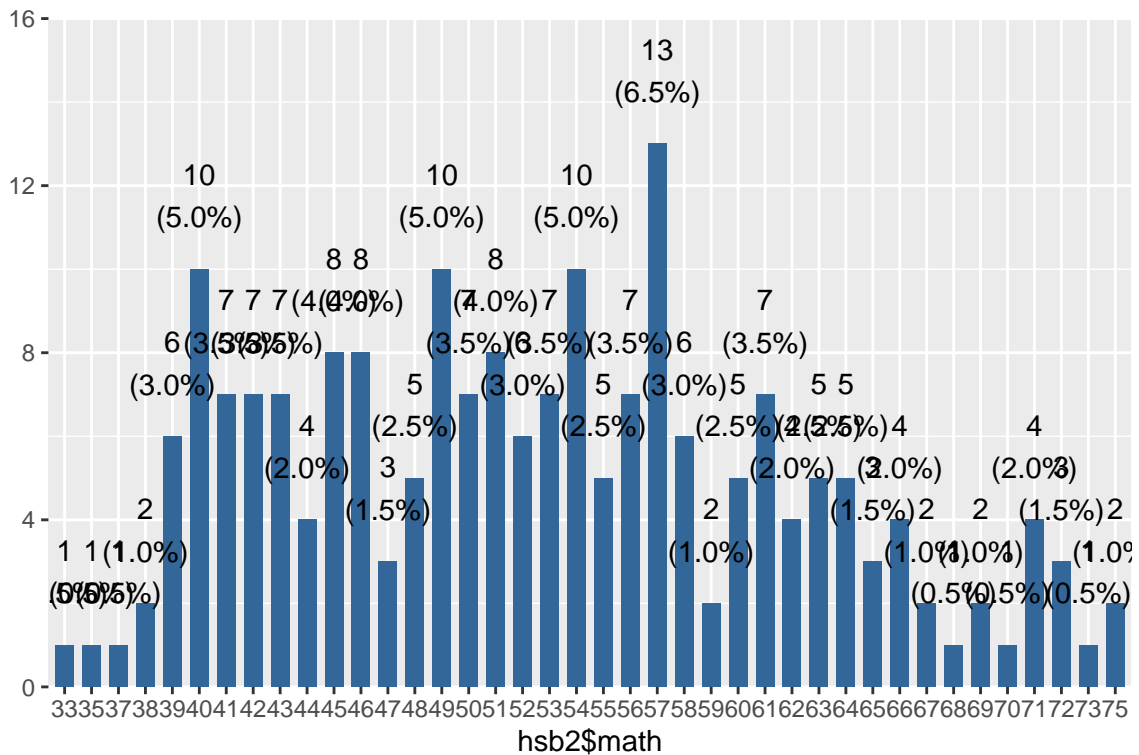
Bar Chart of Program Type

```
plot_frq(hsb2$prog)
```



Math Scores Mean: 53.5 Standard Deviation: 11.6 Range: 31 to 76 Histogram of Math Scores

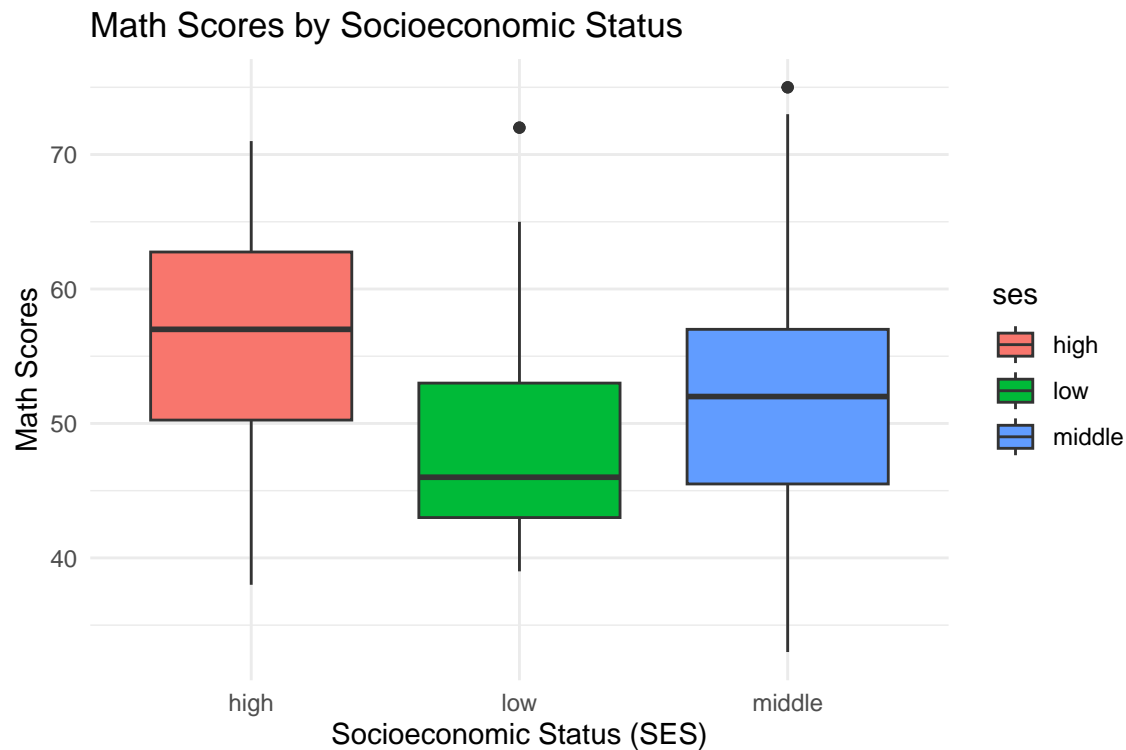
```
hsb2 <- read.table("~/Desktop/ MATH130/hsb2.txt", header=TRUE, sep="\t")
plot_frq(hsb2$math)
```



## Bivariate Exploration

To investigate the relationship between SES and Math Scores, I will create a box plot comparing Math Scores across the three SES categories.

```
ggplot(hsb2, aes(x = ses, y = math, fill = ses)) +  
  geom_boxplot() +  
  labs(title = "Math Scores by Socioeconomic Status",  
        x = "Socioeconomic Status (SES)",  
        y = "Math Scores") +  
  theme_minimal()
```

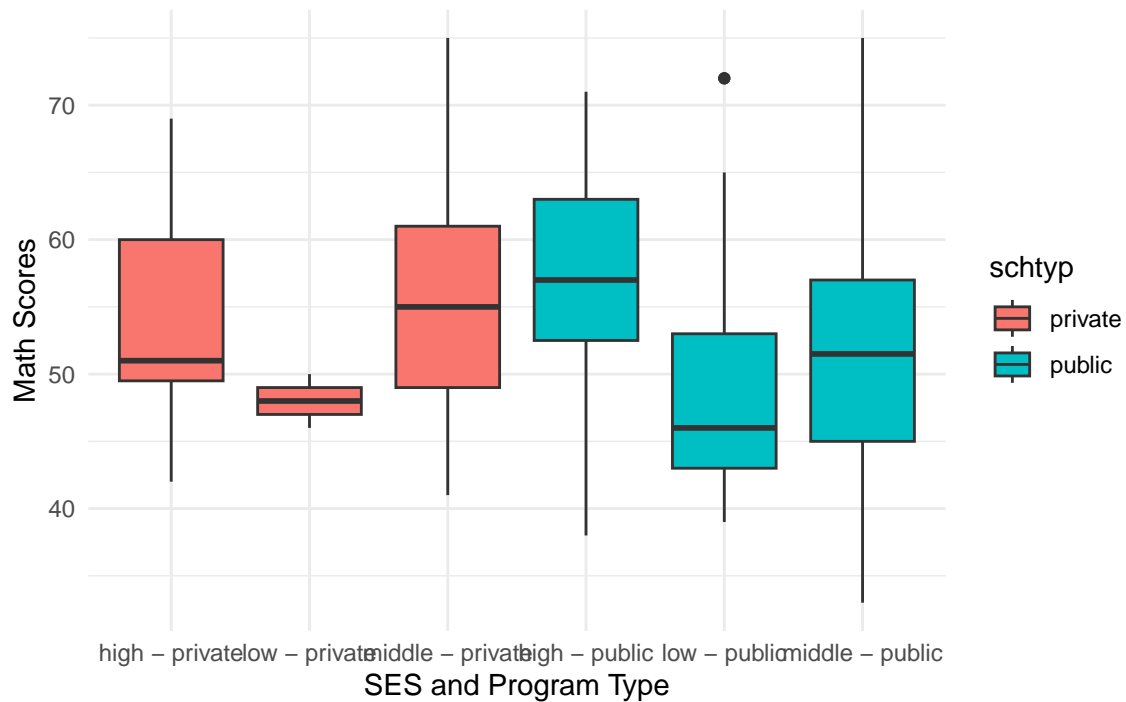


To further explore the influence of Program Type, I will create a grouped box plot that examines Math Scores across both SES and Program Type.

Box Plot of Math Scores by SES and Program Type

```
ggplot(hsb2, aes(x = interaction(ses, schtyp), y = math, fill = schtyp)) +  
  geom_boxplot() +  
  labs(title = "Math Scores by Socioeconomic Status and Program Type",  
        x = "SES and Program Type",  
        y = "Math Scores") +  
  theme_minimal() +  
  scale_x_discrete(labels = function(x) gsub("\\\\.", " - ", x))
```

## Math Scores by Socioeconomic Status and Program Type



### Conclusion

The exploration reveals a clear trend: students from higher SES backgrounds tend to have higher Math Scores. This supports the hypothesis that socioeconomic factors play a significant role in academic performance. The type of program also influences Math Scores, with students in academic programs outperforming those in general and vocational tracks, particularly among middle and high SES groups. The data highlights the importance of both socioeconomic background and educational program type in shaping student achievement in mathematics. Further analysis could examine additional variables, such as gender or race, for a more comprehensive understanding of academic performance.