# Exploratory Data Analysis Project

*Kelly Scott*

*4/2/2019*

```r
HS2beyond <- read.table("/Users/kellyscott/Desktop/math130/data/hsb2.txt", header=TRUE, 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
```

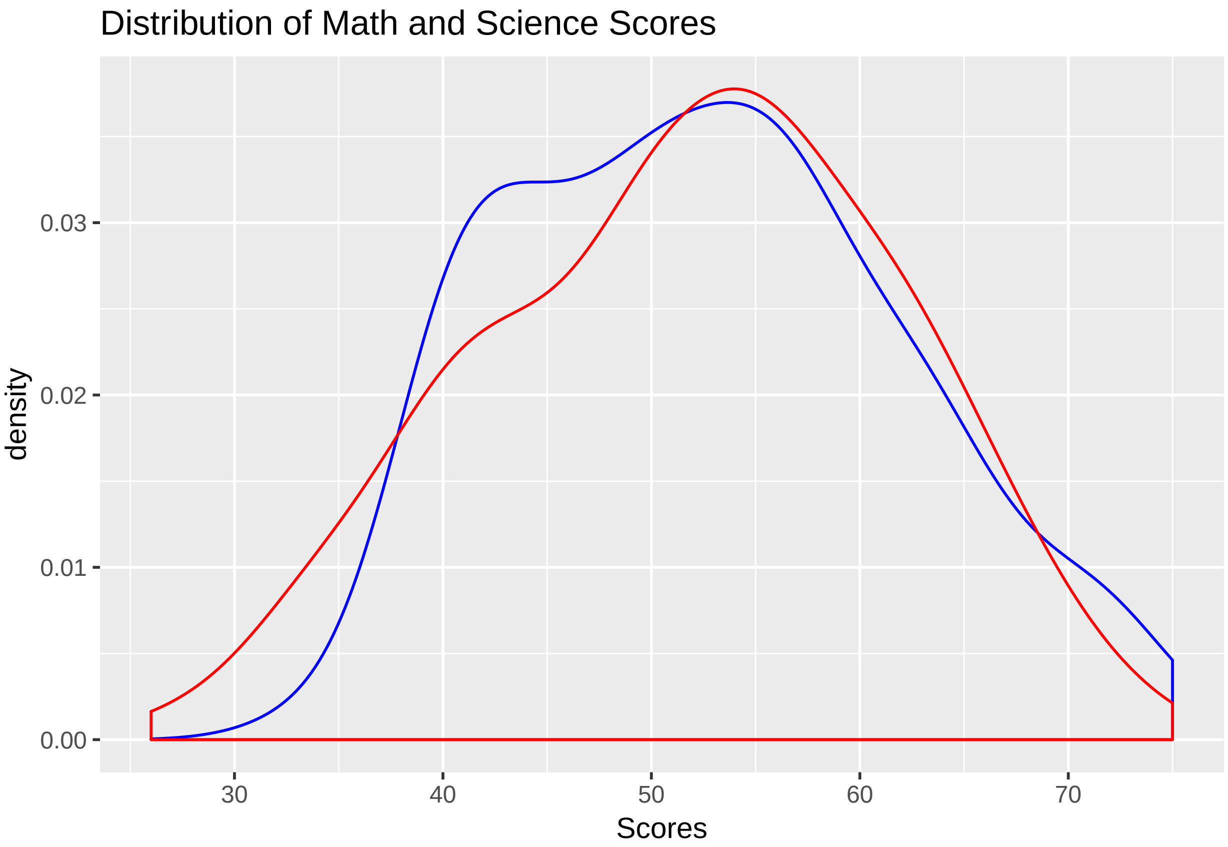
```r
library(ggplot2)
library(knitr)
```

## Introduction

The data I will be analyzing is the High School and Beyond data set. This is a longitudinal study that looked at a group of high school students and how their lives progressed after they graduated. The study looks at a number of variables relating to the students backgrounds, how they preform in school, and what type of education they pursued after high school. I am interested in how socioeconomic status affects math and science scores.

## Visualization of math and science scores

```r
HS2beyond %>%
  select(math, science, ses) %>%
  ggplot() + geom_density(aes(x=math), color="blue") + geom_density(aes(x=science), color='red') + xlab
```



In this graph we can see the distribution of math (blue line) and science (red line) scores for high school students. We can see they both have a bell shaped curve with most students scoring in the 50 to 60 range.

## Average math and science scores

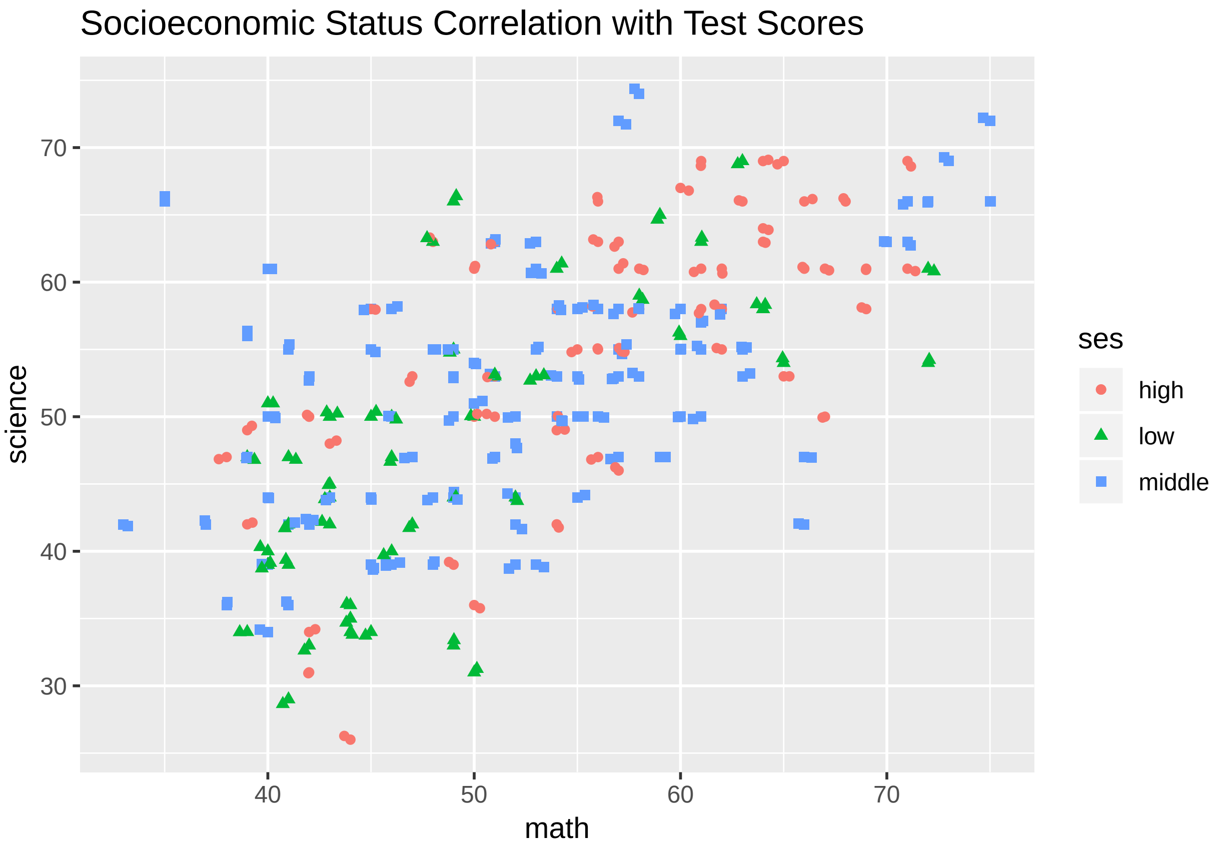
```
summarise(HS2beyond, avg_science=mean(science), avg_math=mean(math))
```

```
##   avg_science avg_math
## 1       51.85   52.645
```

We can see that the average math and science scores reflect what we saw in the density plot. Most students score in the low 50's with slightly better math scores.

## Socioeconomic Status Correlation with Scores

```
HS2beyond %>%  ggplot(aes(x=math, y=science, shape=ses, color=ses)) + geom_point() + geom_jitter() + gg
```



```
HS2beyond %>% select(ses, math, science) %>%
  group_by(ses) %>%
  summarise(avg_math=mean(math), avg_science=mean(science))
```

```
## # A tibble: 3 x 3
##   ses    avg_math avg_science
##   <fct>     <dbl>       <dbl>
## 1 high       56.2        55.4
## 2 low        49.2        47.7
## 3 middle     52.2        51.7
```

There is clearly a difference between class and average test scores. Students from lower socioeconomic class preformed on average 7 points lower than students from a higher socioeconomic class in math and 8 points lower in science. More efforts should be made to support students from lower class house holds to ensure they have the same opprotunities as students from higher class house holds.