Project #3: An Evaluation of Exam Scores

Justin Long

2023-11-28

Data Introduction

The purpose of this data is to examine the affects that different factors have on exam scores. These factors include test preparation, lunch, and parental level of education.

The data can be found at this link: Exam Scores (http://roycekimmons.com/tools/generated_data/exams)

Because the original compiler of this data mentioned it, it will be mentioned here as well, this data set is completed generated for the purpose of practice and does not reflect any real life students.

Preparing the Data

```
library(readr)
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(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
setwd("/Users/jakelong/Documents/Data\ Exploration/Project3")
SP <- read.csv("StudentsPerformance.csv",
               header = T)
PSP <- filter(SP, test.preparation.course == "completed")
NPSP <- filter(SP, test.preparation.course == "none")
```

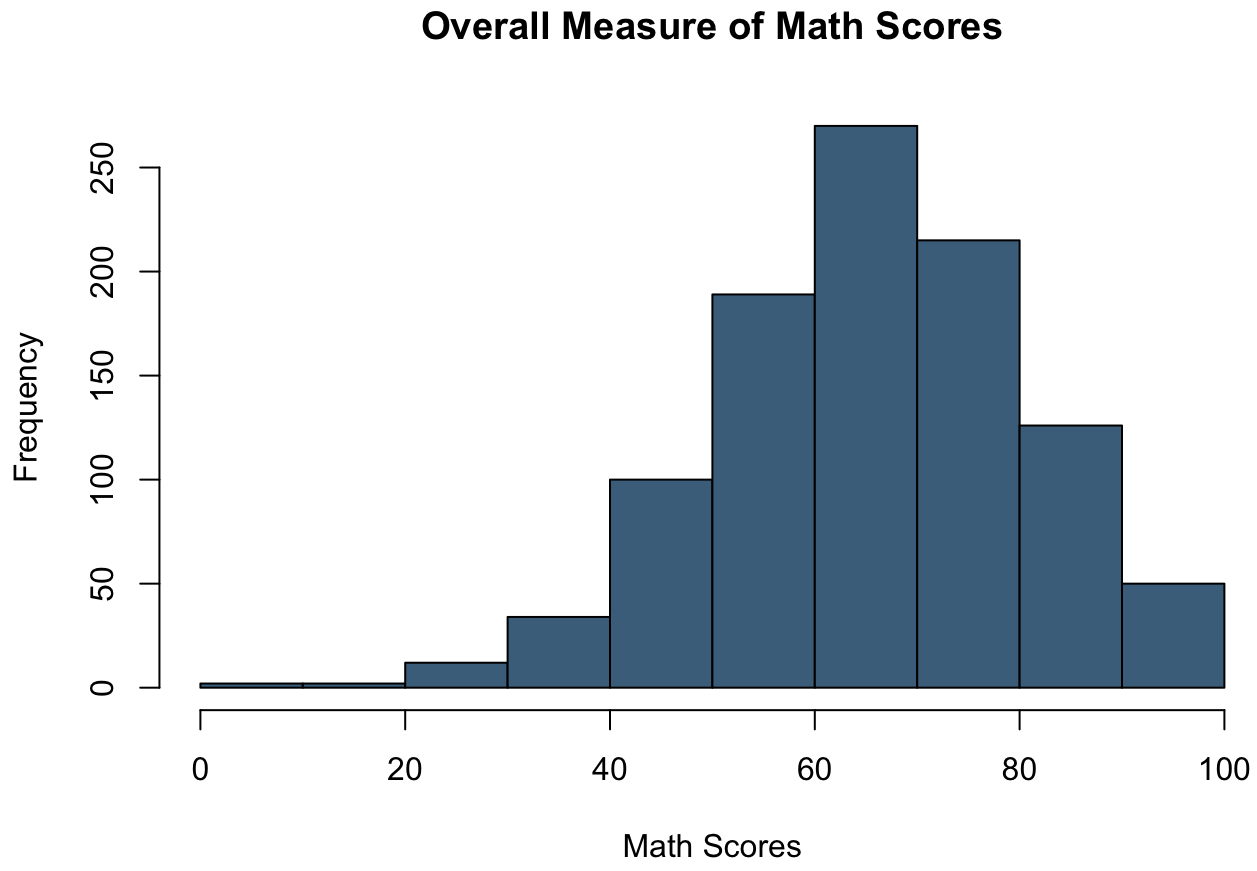
The first data set, *SP*, represents the complete data set in its entirety. The second and third data set, *PSP* and *NPSP*, are filtered sets of the first data set. They are split between students who prepared for the test and those who elected not to prepare. These data sets will be useful shortly.

Understanding the Data

Before any tests are run on the data, it is important to create a reference point to compare to. The average exam scores for each subject will be set to variables in order to make future calculations easier.

Math Scores

```
hist(SP$math.score, main = "Overall Measure of Math Scores",
     xlab = "Math Scores", ylab = "Frequency", col = "skyblue4")
```




```
describe(SP$math.score)
```

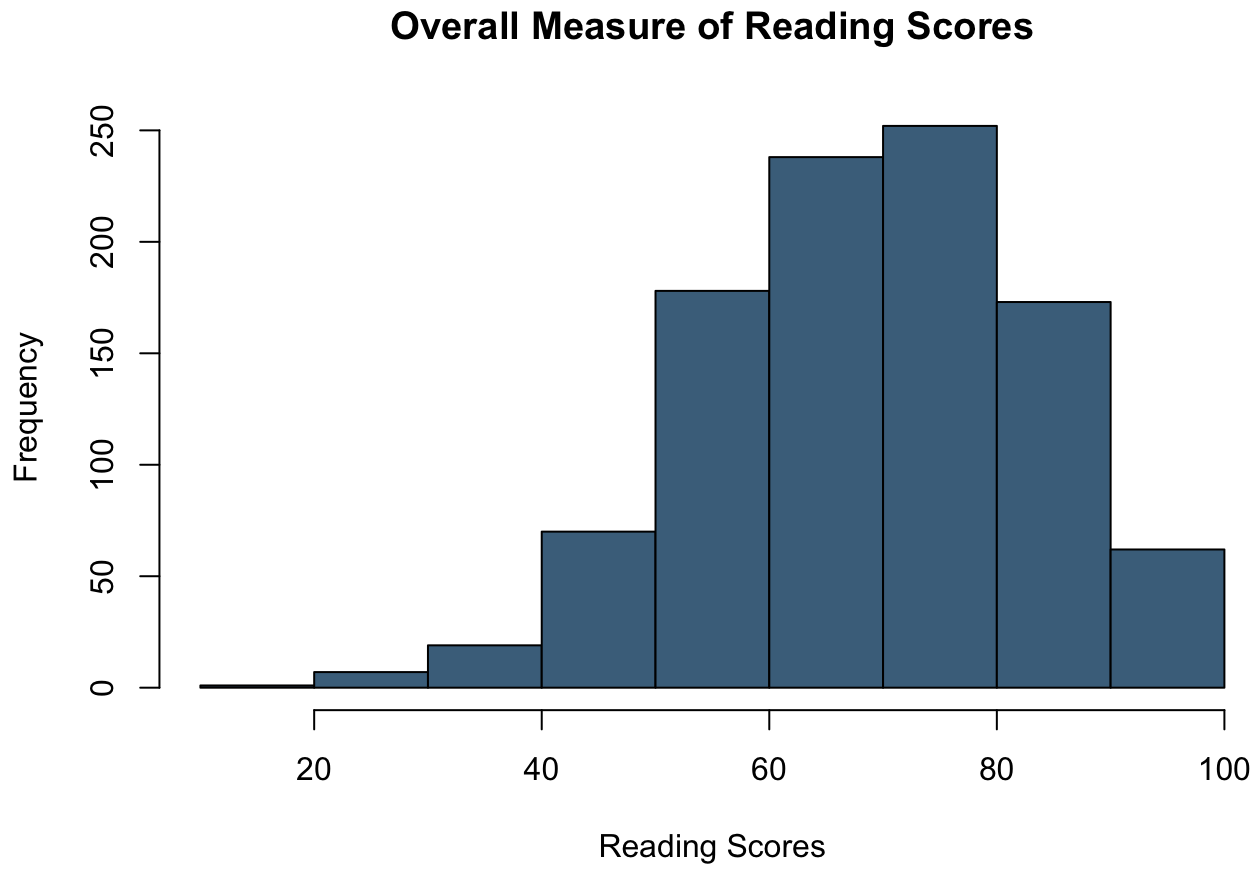
```
##    vars    n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 1000 66.09 15.16     66   66.38 14.83   0 100   100 -0.28     0.26 0.48
```

```
MathAVG <- mean(SP$math.score)
```

The math scores appear unimodal and skewed to the left. The average exam score is 66.

Reading Scores

```
hist(SP$reading.score, main = "Overall Measure of Reading Scores",
     xlab = "Reading Scores", ylab = "Frequency", col = "skyblue4")
```




```
describe(SP$reading.score)
```

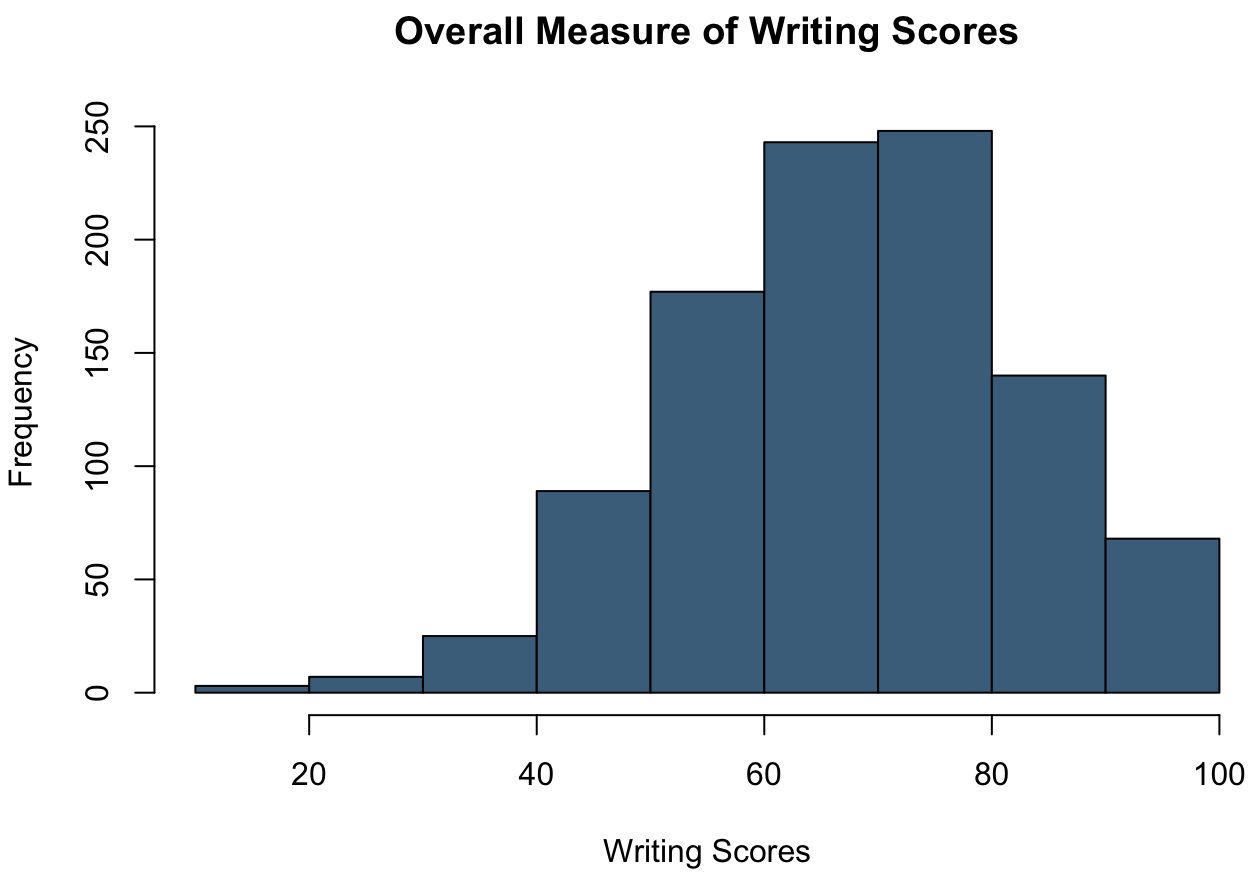
```
##    vars    n  mean   sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 1000 69.17 14.6     70    69.5 14.83  17 100    83 -0.26    -0.08 0.46
```

```
ReadingAVG <- mean(SP$reading.score)
```

The reading scores appear unimodal and are skewed to the left. The average exam score is 69.

Writing Scores

```
hist(SP$writing.score, main = "Overall Measure of Writing Scores",
     xlab = "Writing Scores", ylab = "Frequency", col = "skyblue4")
```




```
describe(SP$writing.score)
```

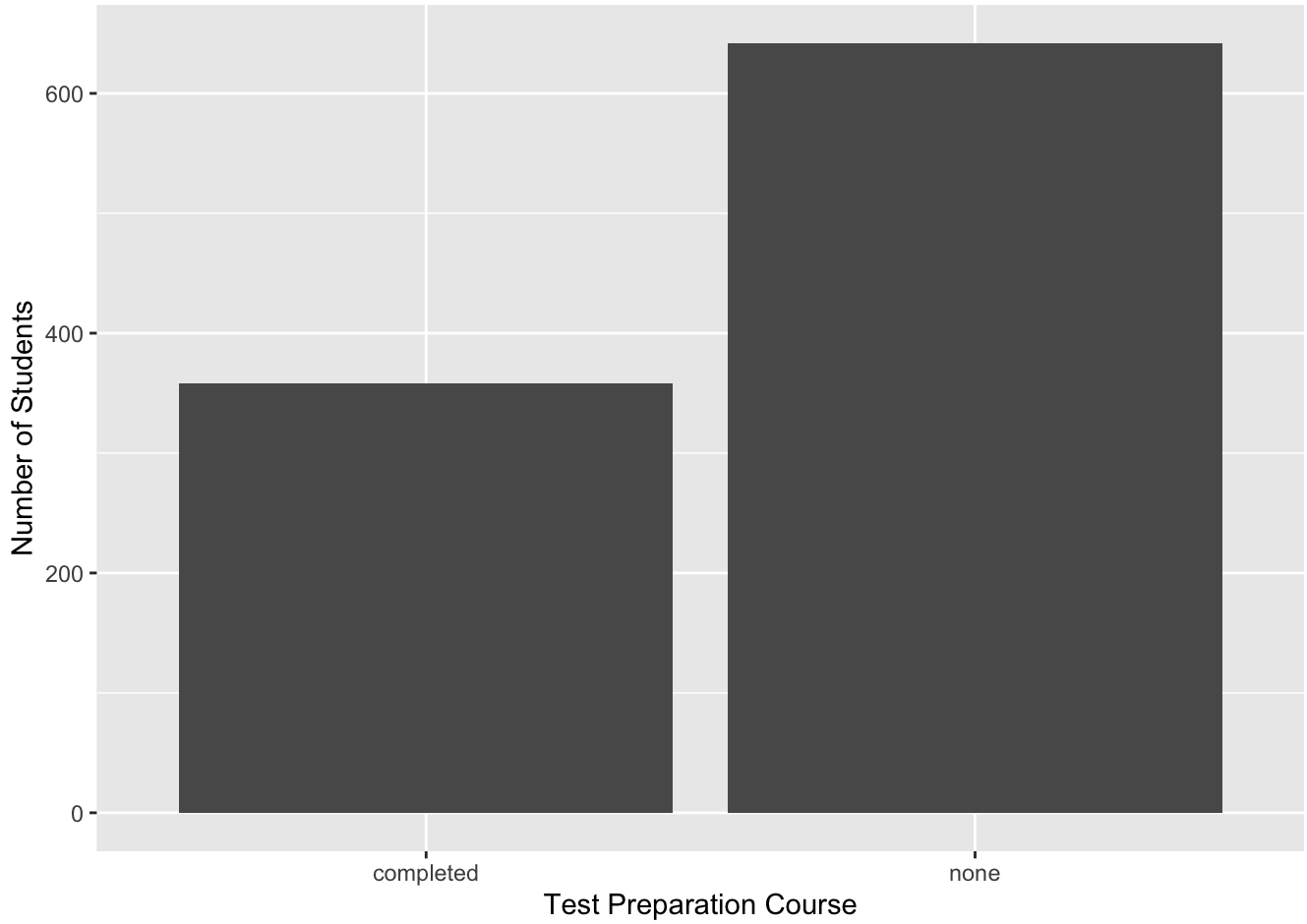
```
##    vars    n  mean   sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 1000 68.05 15.2     69   68.41 16.31  10 100    90 -0.29    -0.05 0.48
```

```
WritingAVG <- mean(SP$writing.score)
```

The writing scores appear unimodal and are skewed to the left. The average exam score is 68.

Student Test Preparation

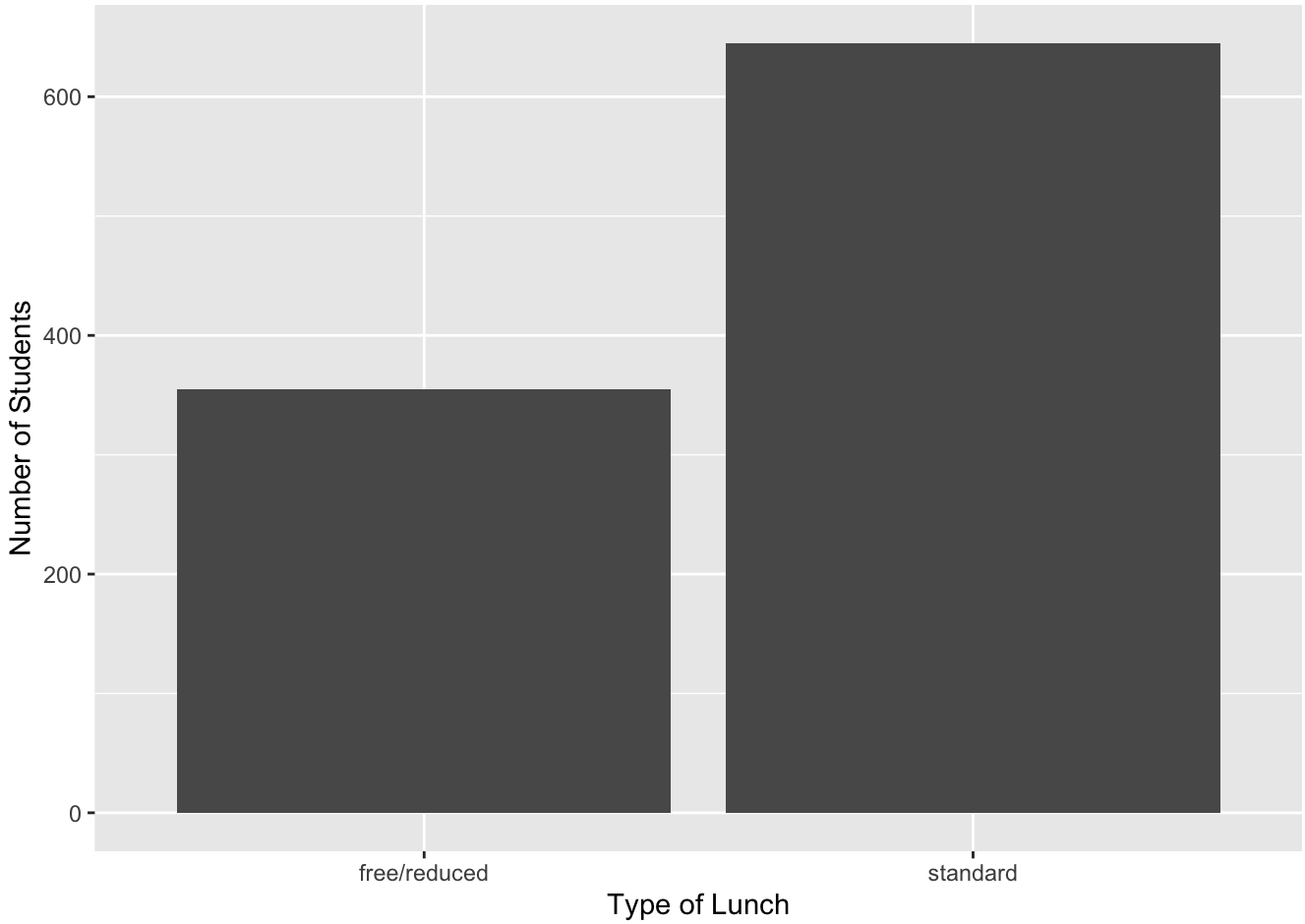
```
ggplot(SP, aes(test.preparation.course)) +
    geom_bar() + labs(x = 'Test Preparation Course', y = 'Number of Students')
```



It is clear that the majority of students choose not to prepare for the exam.

Student Lunch

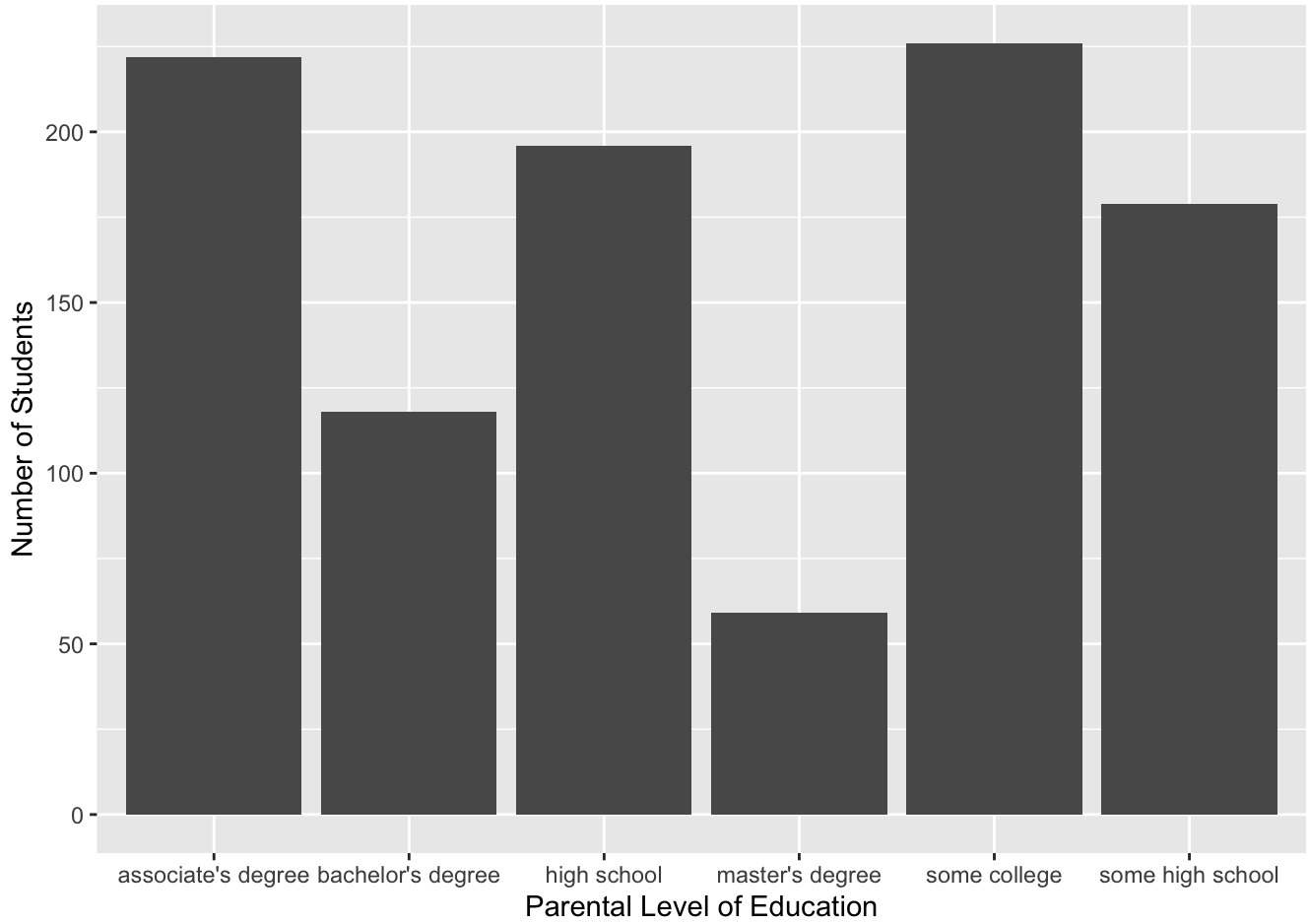
```
ggplot(SP, aes(lunch)) +
    geom_bar() + labs(x = 'Type of Lunch', y = 'Number of Students')
```



The majority of students receive a standard lunch as opposed to a free/reduced lunch.

Parental Level of Education

```
ggplot(SP, aes(parental.level.of.education)) +
    geom_bar() + labs(x = 'Parental Level of Education', y = 'Number of Students')
```



Parents with an Associate's degree with or some college education are the most common. Parents with a high school diploma or some amount high school education are the third and fourth most common respectively. The numbers decrease a lot here as a parent with a Bachelor's degree is the fifth most common. As one may expect, a parent with a Master's degree appears the least.

Task 1: Examining How Test Prep Affects Scores

This section is going to examine the grades of the students on the basis of exam preparation. The data will be split between students who chose to prepare and those who did not. The average scores from each of the two groups will then be compared.

Students who Prepared

Math

```
describe(PSP$math.score)
```

```
##    vars   n mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 358 69.7 14.44     69   69.85 14.83  23 100    77 -0.15    -0.19 0.76
```

```
PrepMath <- mean(PSP$math.score)
```

Reading

```
describe(PSP$reading.score)
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 358 73.89 13.64     75   74.29 13.34  37 100    63 -0.29    -0.38 0.72
```

```
PrepReading <- mean(PSP$reading.score)
```

Writing

```
describe(PSP$writing.score)
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 358 74.42 13.38     76   74.91 13.34  36 100    64 -0.35    -0.21 0.71
```

```
PrepWriting <- mean(PSP$writing.score)
```

Students Who Did Not Prepare

Math

```
describe(NPSP$math.score)
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis  se
## X1    1 642 64.08 15.19     64   64.44 14.83   0 100   100 -0.33     0.38 0.6
```

```
NPMath <- mean(NPSP$math.score)
```

Reading

```
describe(NPSP$reading.score)
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 642 66.53 14.46     67   66.83 13.34  17 100    83 -0.23     0.05 0.57
```

```
NPReading <- mean(NPSP$reading.score)
```

Writing

```
describe(NPSP$writing.score)
```

```
##    vars   n mean sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 642 64.5 15     65   64.74 14.83  10 100    90 -0.21     0.06 0.59
```

```
NPWriting <- mean(NPSP$writing.score)
```

Difference

Math

```
PrepMathDiff <- PrepMath - NPMath; PrepMathDiff
```

```
## [1] 5.617649
```

Reading

```
PrepReadingDiff <- PrepReading - NPReading; PrepReadingDiff
```

```
## [1] 7.359587
```

Writing

```
PrepWritingDiff <- PrepWriting - NPWriting; PrepWritingDiff
```

```
## [1] 9.914322
```

Students who prepared for the exam performed better across the board. Writing exam scores benefited the most from the preparation. Reading exam scores benefited the second most, and math scores benefited the least.

Task 2: Comparing Lunch Type With Exam Scores

Free/Reduced Lunch

Math

```
describe(SP$math.score[SP$lunch == "free/reduced"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis  se
## X1    1 355 58.92 15.16     60   59.31 16.31   0 100   100 -0.31     0.36 0.8
```

```
FLM <- mean(SP$math.score[SP$lunch == "free/reduced"])
```

Reading

```
describe(SP$reading.score[SP$lunch == "free/reduced"])
```

```
##    vars   n  mean   sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 355 64.65 14.9     65      65 14.83  17 100    83 -0.23    -0.04 0.79
```

```
FLR <- mean(SP$reading.score[SP$lunch == "free/reduced"])
```

Writing

```
describe(SP$writing.score[SP$lunch == "free/reduced"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 355 63.02 15.43     64   63.37 16.31  10 100    90 -0.25     0.08 0.82
```

```
FLW <- mean(SP$writing.score[SP$lunch == "free/reduced"])
```

Standard Lunch

Math

```
describe(SP$math.score[SP$lunch == "standard"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 645 70.03 13.65     69    70.1 14.83  19 100    81 -0.1    -0.16 0.54
```

```
SLM <- mean(SP$math.score[SP$lunch == "standard"])
```

Reading

```
describe(SP$reading.score[SP$lunch == "standard"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 645 71.65 13.83     72    71.9 14.83  26 100    74 -0.21    -0.24 0.54
```

```
SLR <- mean(SP$reading.score[SP$lunch == "standard"])
```

Writing

```
describe(SP$reading.score[SP$lunch == "standard"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 645 71.65 13.83     72    71.9 14.83  26 100    74 -0.21    -0.24 0.54
```

```
SLW <- mean(SP$writing.score[SP$lunch == "standard"])
```

Difference

Math

```
LMDiff <- SLM - FLM; LMDiff
```

```
## [1] 11.11298
```

Reading

```
LRDiff <- SLR - FLR; LRDiff
```

```
## [1] 7.000742
```

Writing

```
LWDiff <- SLW - FLW; LWDiff
```

```
## [1] 7.800721
```

The data shows that students who receive free/reduced lunch perform worse on the exams.

Task 3: The Impact of Parental Education Level on Student Exam Scores.

Some High School

Math

```
describe(SP$math.score[SP$parental.level.of.education == "some high school"])
```

```
##    vars   n mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 179 63.5 15.93     65   64.23 14.83   0  97    97 -0.62     0.79 1.19
```

```
SHMath <- mean(SP$math.score[SP$parental.level.of.education == "some high school"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "some high school"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 179 66.94 15.48     67   67.66 16.31  17 100    83 -0.4    -0.07 1.16
```

```
SHReading <- mean(SP$reading.score[SP$parental.level.of.education == "some high schoo
l"])
```

Writing

```
describe(SP$writing.score[SP$parental.level.of.education == "some high school"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 179 64.89 15.74     66   65.54 17.79  10 100    90 -0.46     0.07 1.18
```

```
SHWriting <- mean(SP$writing.score[SP$parental.level.of.education == "some high schoo
l"])
```

High School

Math

```
describe(SP$math.score[SP$parental.level.of.education == "high school"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 196 62.14 14.54     63   62.65 13.34   8  99    91 -0.44     0.52 1.04
```

```
HMath <- mean(SP$math.score[SP$parental.level.of.education == "high school"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "high school"])
```

```
##    vars   n mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 196 64.7 14.13     66   65.02 14.83  24  99    75 -0.27     -0.1 1.01
```

```
HReading <- mean(SP$reading.score[SP$parental.level.of.education == "high school"])
```

Writing

```
describe(SP$writing.score[SP$parental.level.of.education == "high school"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 196 62.45 14.09     64   62.86 14.83  15 100    85 -0.33     0.02 1.01
```

```
HWriting <- mean(SP$writing.score[SP$parental.level.of.education == "high school"])
```

Some College

Math

```
describe(SP$math.score[SP$parental.level.of.education == "some college"])
```

```
##    vars   n  mean    sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 226 67.13 14.31   67.5   67.32 12.6  19 100    81 -0.22     0.47 0.95
```

```
SCMath <- mean(SP$math.score[SP$parental.level.of.education == "some college"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "some college"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 226 69.46 14.06   70.5   70.04 14.08  23 100    77 -0.37    -0.03 0.94
```

```
SCReading <- mean(SP$reading.score[SP$parental.level.of.education == "some college"])
```

Writing

```
describe(SP$writing.score[SP$parental.level.of.education == "some college"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis se
## X1    1 226 68.84 15.01     70   69.47 14.83  19  99    80 -0.42     0.07  1
```

```
SCWriting <- mean(SP$writing.score[SP$parental.level.of.education == "some college"])
```

Associate’s Degree

Math

```
describe(SP$math.score[SP$parental.level.of.education == "associate's degree"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 222 67.88 15.11     67   67.84 17.79  26 100    74 0.01    -0.71 1.01
```

```
ADMath <- mean(SP$math.score[SP$parental.level.of.education == "associate's degree"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "associate's degree"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 222 70.93 13.87   72.5   71.12 15.57  31 100    69 -0.17    -0.47 0.93
```

```
ADReading <- mean(SP$reading.score[SP$parental.level.of.education == "associate's degre
e"])
```

Writing

```
describe(SP$writing.score[SP$parental.level.of.education == "associate's degree"])
```

```
##    vars   n mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 222 69.9 14.31   70.5   70.08 15.57  35 100    65 -0.14    -0.64 0.96
```

```
ADWriting <- mean(SP$writing.score[SP$parental.level.of.education == "associate's degre
e"])
```

Bachelor's Degree

Math

```
describe(SP$math.score[SP$parental.level.of.education == "bachelor's degree"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 118 69.39 14.94     68   69.55 13.34  29 100    71 -0.11    -0.29 1.38
```

```
BDMath <- mean(SP$math.score[SP$parental.level.of.education == "bachelor's degree"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "bachelor's degree"])
```

```
##    vars   n mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 118   73 14.29     73   72.94 14.83  41 100    59 0.03    -0.49 1.32
```

```
BDReading <- mean(SP$reading.score[SP$parental.level.of.education == "bachelor's degre
e"])
```

Writing

```
describe(SP$writing.score[SP$parental.level.of.education == "bachelor's degree"])
```

```
##    vars   n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 118 73.38 14.73     74   73.64 14.08  38 100    62 -0.16    -0.48 1.36
```

```
BDWriting <- mean(SP$writing.score[SP$parental.level.of.education == "bachelor's degre
e"])
```

Master's Degree

Math

```
describe(SP$math.score[SP$parental.level.of.education == "master's degree"])
```

```
##    vars  n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 59 69.75 15.15     73    70.1 17.79  40  95    55 -0.22    -1.15 1.97
```

```
MDMath <- mean(SP$math.score[SP$parental.level.of.education == "master's degree"])
```

Reading

```
describe(SP$reading.score[SP$parental.level.of.education == "master's degree"])
```

```
##    vars  n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 59 75.37 13.78     76   75.33 13.34  42 100    58 -0.02    -0.78 1.79
```

```
MDReading <- mean(SP$reading.score[SP$parental.level.of.education == "master's degree"])
```

Writing

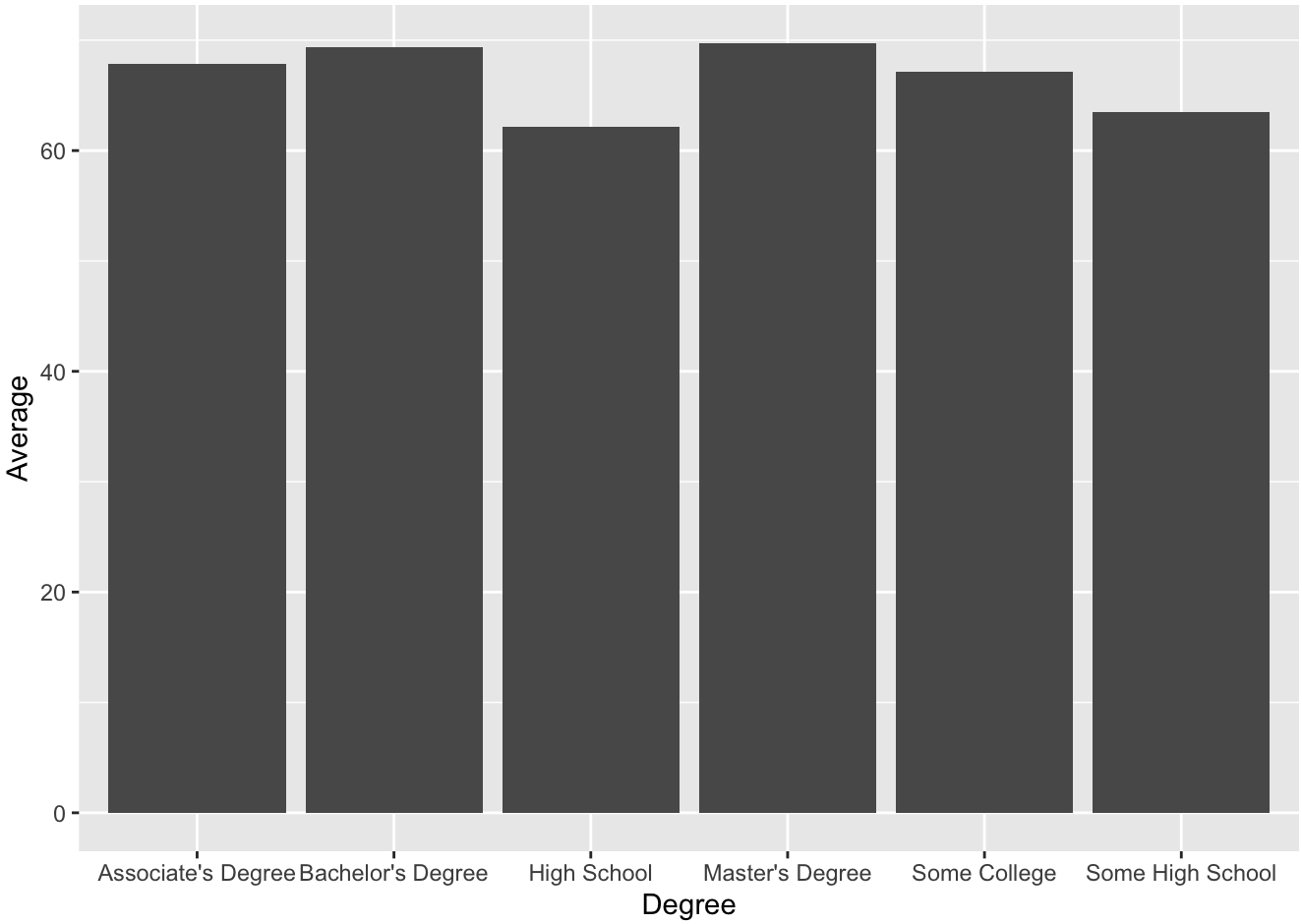
```
describe(SP$writing.score[SP$parental.level.of.education == "master's degree"])
```

```
##    vars  n  mean    sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 59 75.68 13.73     75   75.78 13.34  46 100    54 -0.05    -0.76 1.79
```

```
MDWriting <- mean(SP$writing.score[SP$parental.level.of.education == "master's degree"])
```

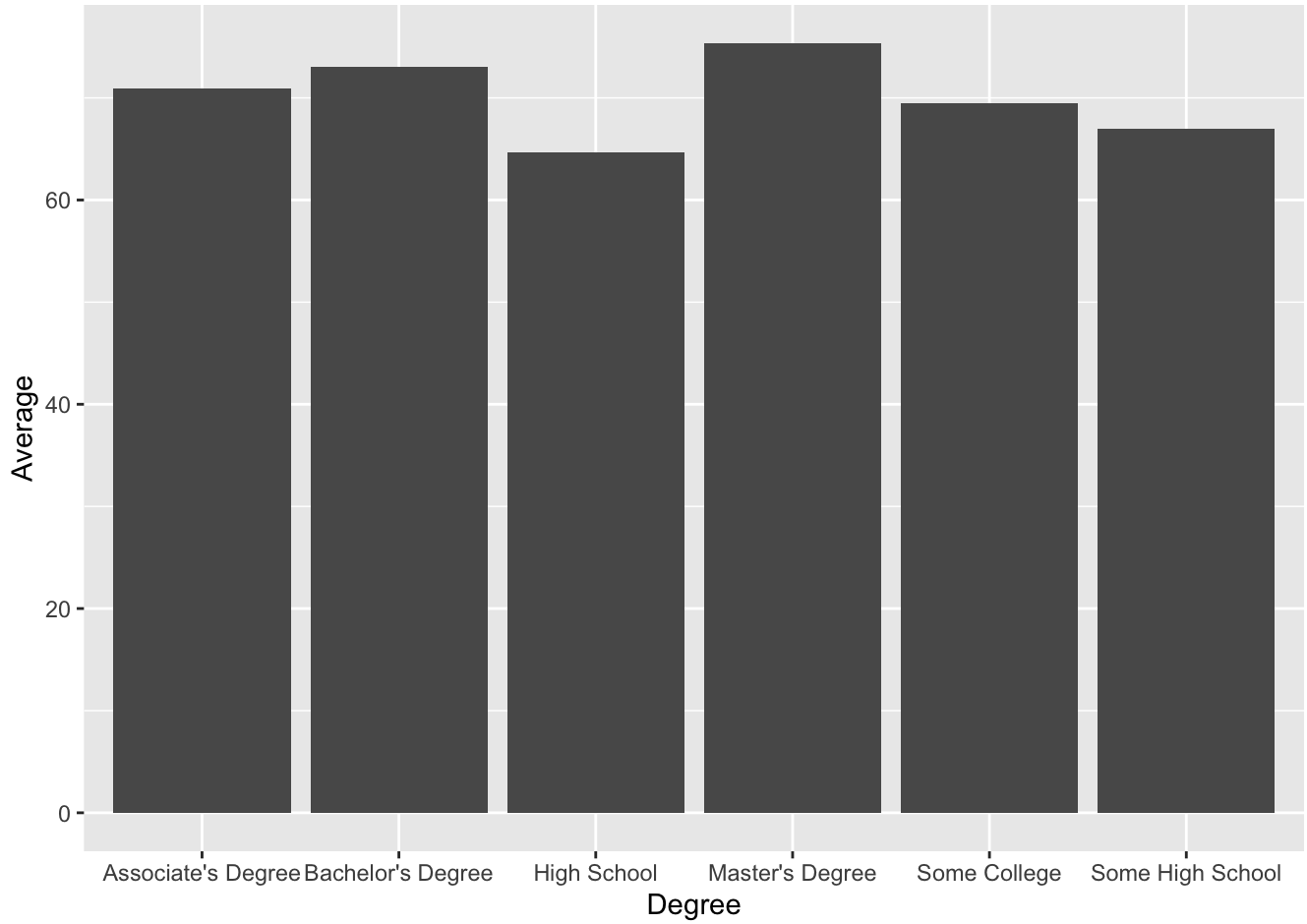
Math Comparison

```
cf <- data.frame(
  Degree=c("Some High School", "High School", "Some College", "Associate's Degree", "Bac
helor's Degree",
         "Master's Degree"),
  Average=c(SHMath, HMath, SCMath, ADMath, BDMath, MDMath)
)
ggplot(cf, aes(x=Degree, y=Average)) +
  geom_bar(stat = "identity")
```



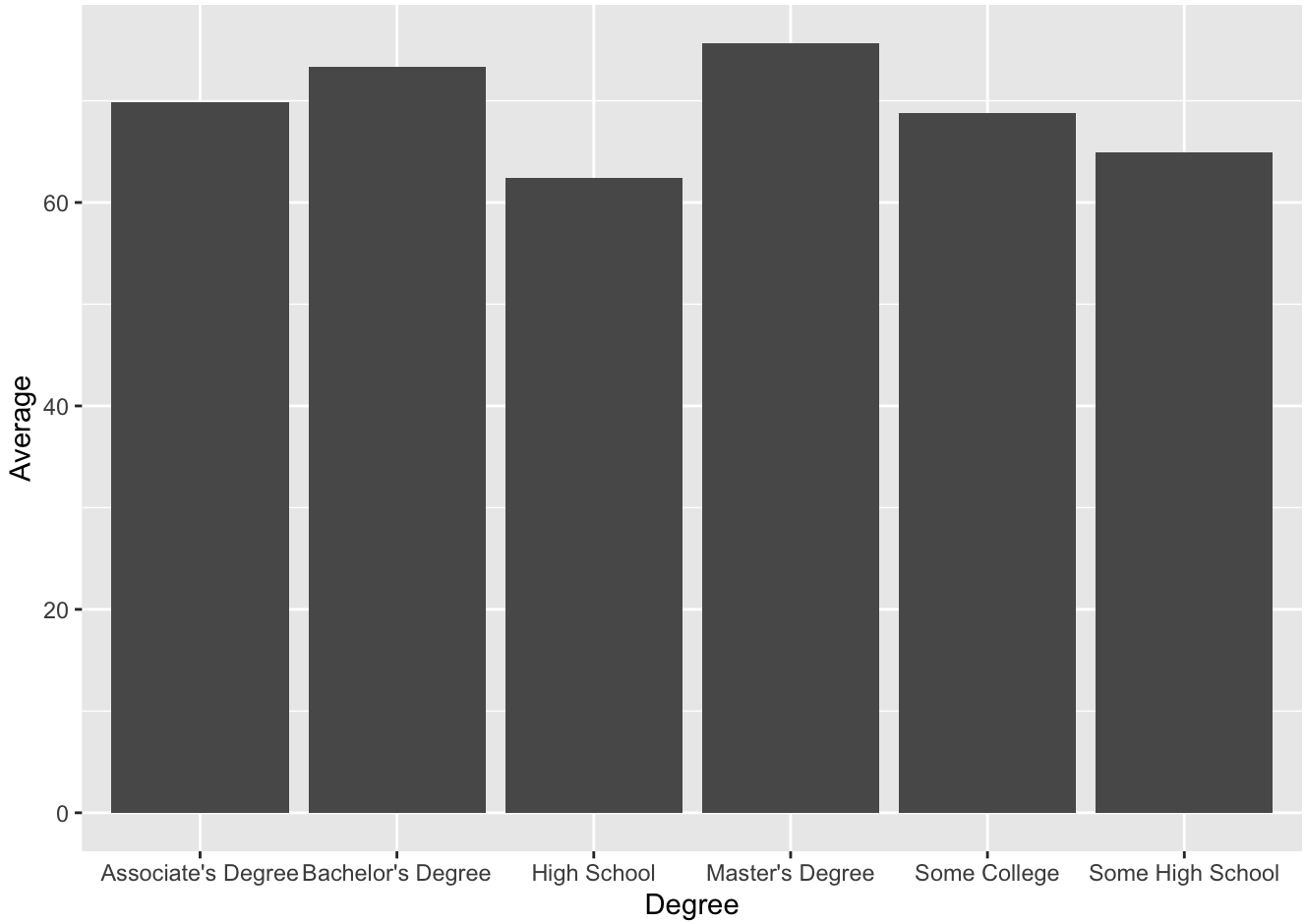
Reading Comparison

```
rdf <- data.frame(
  Degree=c("Some High School", "High School", "Some College", "Associate's Degree", "Bac
helor's Degree",
           "Master's Degree"),
  Average=c(SHReading, HReading, SCReading, ADReading, BDReading, MDReading)
)
ggplot(rdf, aes(x=Degree, y=Average)) +
  geom_bar(stat = "identity")
```



Writing Comparison

```
wdf <- data.frame(
  Degree=c("Some High School", "High School", "Some College", "Associate's Degree", "Bac
helor's Degree",
           "Master's Degree"),
  Average=c(SHWriting, HWriting, SCWriting, ADWriting, BDWriting, MDWriting)
)
ggplot(wdf, aes(x=Degree, y=Average)) +
  geom_bar(stat = "identity")
```



Looking at the data, the appears to be a general trend among all three subjects, the higher level of parental education, the higher the average exam scores are. The exception to this trend is that students whose parents only completed some high school, on average, perform better than those students whose parents have a high school diploma.

Putting It All Together

Through out this study, three factors and their effects on exam scores have been examined: test preparation, free/standard lunch, and parental level of education. In order to examine the impact of these factors, the students were broken up into groups and the exam averages were compared to overall averages for each subject.

Test Prepartion

Looking at the effects of test preparation first, the average exam scores of students who prepared for the test were higher than their counter parts who elected not to prepare. This outcome is expected as preparing for an exam is a logical way to improve the outcome. One helpful extension of this study would be comparing the amount of time studying and the corresponding average grade.

Free/Reduced Lunch vs. Standard Lunch

The data tests concluded that the students who received free or reduced lunch performed worse than the students that received the standard lunch. Unlike the preparation tests, the explanations are not immediately obvious. It is important to take these tests and look beyond. For example, students who receive a free or reduced lunch typically come from a worse economical situation. In this situation, the lunches are not as impactful as general economic situations of the students. Another reason could be that the lunches are physically different.

The standard lunch could be comprised of a meat and vegetables and the free/reduced lunch could be a simple PB&J, for example. In this situation, examining the nutritional components of the lunches and how nutrition affects exam scores could be beneficial.

Parental Level of Education

Much like the lunch tests, the correlation between student exam scores and the level of parental education is less obvious. One possible explanation is that parents who completed a higher level of education can better assist their children in learning themselves. Another possibility is that more educated parents place more importance on the value of education and thus create a home environment that encourages educational success. There is likely no one correct explanation. It is more likely that the correct explanation is a mixture of multiple factors.