```r
head(StudentsPerformance)
```

A tibble: 6 x 8

| gender <chr> | race/ethnicity <chr> | parental level of education <chr> | lunch <chr> | test preparation course <chr> | math score <dbl> |
|---|---|---|---|---|---|
| female | group B | bachelor's degree | standard | none | 72 |
| female | group C | some college | standard | completed | 69 |
| female | group B | master's degree | standard | none | 90 |
| male | group A | associate's degree | free/reduced | none | 47 |
| male | group C | some college | standard | none | 76 |
| female | group B | associate's degree | standard | none | 71 |

6 rows | 1-6 of 8 columns

## INTRODUCTION

For my project, I chose the "studentsPerformance" data set. I am hoping to find a correlation between each of the test scores. I also would like to see how a free/reduced lunch plays in with a students test scores. I am also interested in seeing how a parents degree can effect as students test score.

I will be focusing on all the variables above except "race/ethnicity." I do not have direction on what the groups mean. I contacted the owner of the data set but failed to get a response

```r
select(StudentsPerformance, 'test preparation course', 'math score')
```

A tibble: 1,000 x 2

| test preparation course <chr> | math score <dbl> |
|---|---|
| none | 72 |
| completed | 69 |
| none | 90 |
| none | 47 |
| none | 76 |
| none | 71 |
| completed | 88 |
| none | 40 |
| completed | 64 |
| none | 38 |

```{r}
select(StudentsPerformance, 'test preparation course', 'writing score')
```

A tibble: 1,000 x 2

| test preparation course <chr> | writing score <dbl> |
|---|---|
| none | 74 |
| completed | 88 |
| none | 93 |
| none | 44 |
| none | 75 |
| none | 78 |
| completed | 92 |
| none | 39 |
| completed | 67 |
| none | 50 |

1-10 of 1,000 rows

Previous 1 2 3 4 5 6 ... 100 Next

```{r}
select(StudentsPerformance, 'test preparation course', 'reading score')
```

A tibble: 1,000 x 2

| test preparation course <chr> | reading score <dbl> |
|---|---|
| none | 72 |
| completed | 90 |
| none | 95 |
| none | 57 |
| none | 78 |
| none | 83 |
| completed | 95 |
| none | 43 |
| completed | 64 |
| none | 60 |

1-10 of 1,000 rows

Previous 1 2 3 4 5 6 ... 100 Next

A closer look at the scores.

```r
mean_writing <- StudentsPerformance %>%
  group_by(`test preparation course`) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))

mean_writing
```

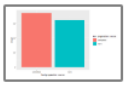
A tibble: 2 x 2

| test preparation course <chr> | mean <dbl> |
|---|---|
| completed | 74.41899 |
| none | 64.50467 |

2 rows

I used a group by to look at test prep and found the mean of the writing scores

```r
ggplot(data = mean_writing)+
  geom_col(mapping = aes(x = `test preparation course` ,y = mean, fill = `test preparation course`),stat = "identity")
```




```r
mean_reading <- StudentsPerformance %>%
  group_by(`test preparation course`) %>%
   summarise(mean = mean(`reading score`, na.rm = TRUE))

mean_reading
```

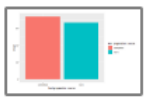
A tibble: 2 x 2

| test preparation course <chr> | mean <dbl> |
|---|---|
| completed | 73.89385 |
| none | 66.53427 |

2 rows

I used a group by to look at test prep and found the mean of the reading scores

```r
ggplot(data = mean_reading)+
  geom_col(mapping = aes(x = `test preparation course` ,y = mean, fill = `test preparation course`),stat = "identity")
```





We find that a test preparation course made a big difference between scores

```
mean_math <- StudentsPerformance %>%
  group_by(`test preparation course`) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))

mean_math
```

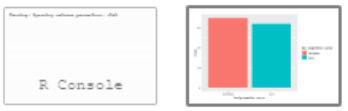
A tibble: 2 x 2

| test preparation course <chr> | mean <dbl> |
|---|---|
| completed | 69.69553 |
| none | 64.07788 |

2 rows

I used a group by to look at test prep and found the mean of the math scores

```{r}
ggplot(data = mean_math)+
  geom_col(mapping = aes(x = `test preparation course` ,y = mean, fill = `test preparation course`),stat = "i
```







Although the test preparation course did make a little over a 5% difference, the averages remain fairly low

```
mean_math <- StudentsPerformance %>%
  group_by(`lunch`) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))

mean_math
```

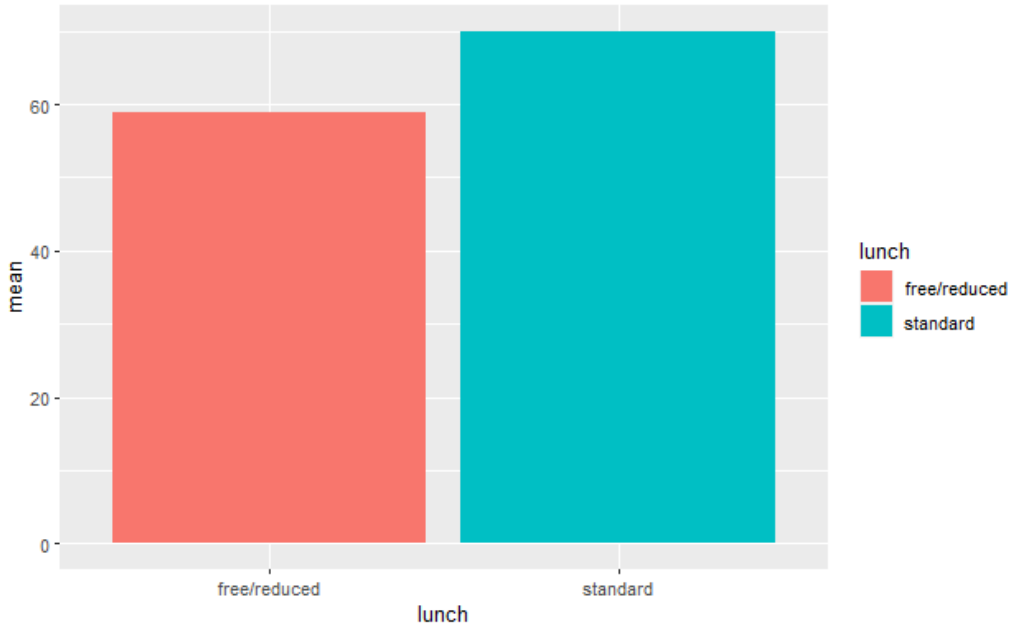
A tibble: 2 x 2

| lunch <chr> | mean <dbl> |
|---|---|
| free/reduced | 58.92113 |
| standard | 70.03411 |

2 rows

I used a group by to look at lunch and found the mean of the math scores

```{r}
ggplot(data = mean_math)+
  geom_col(mapping = aes(x = lunch ,y = mean, fill = lunch),stat = "identity")
```

```
mean_writing <- StudentsPerformance %>%
  group_by(`lunch`) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))

mean_writing
```

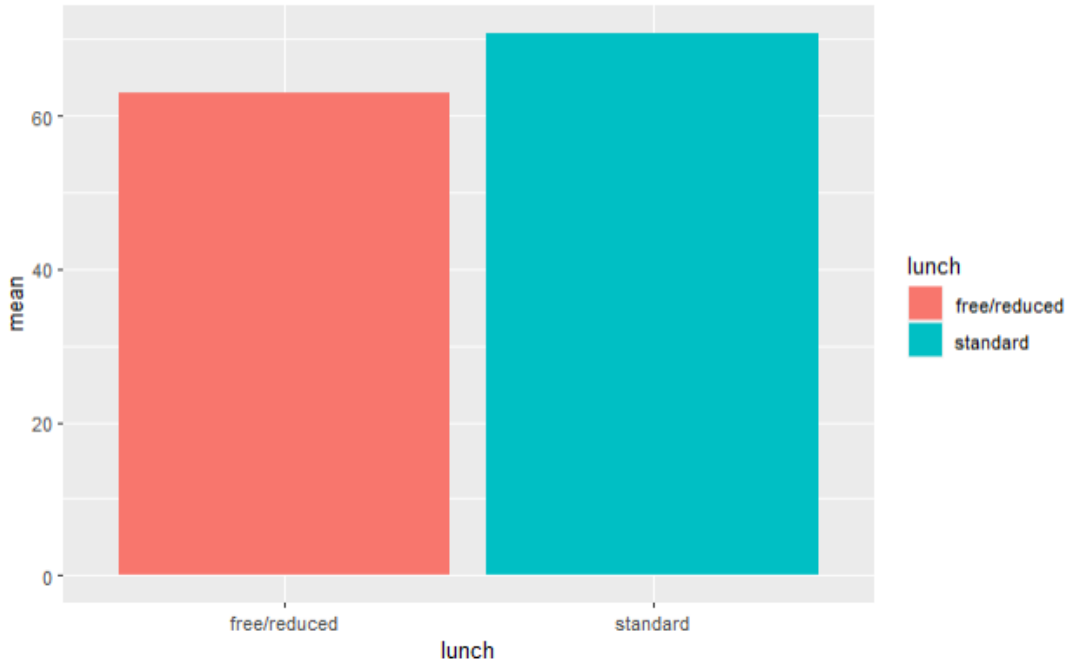
A tibble: 2 x 2

| lunch <chr> | mean <dbl> |
|---|---|
| free/reduced | 63.02254 |
| standard | 70.82326 |

2 rows

I used a group by to look at lunch and found the mean of the writing scores

```{r}
ggplot(data = mean_writing)+
  geom_col(mapping = aes(x = lunch ,y = mean, fill = lunch),stat = "identity")
```

```
mean_reading <- StudentsPerformance %>%
  group_by(`lunch`) %>%
    summarise(mean = mean(`reading score`, na.rm = TRUE))

mean_reading
```

A tibble: 2 x 2

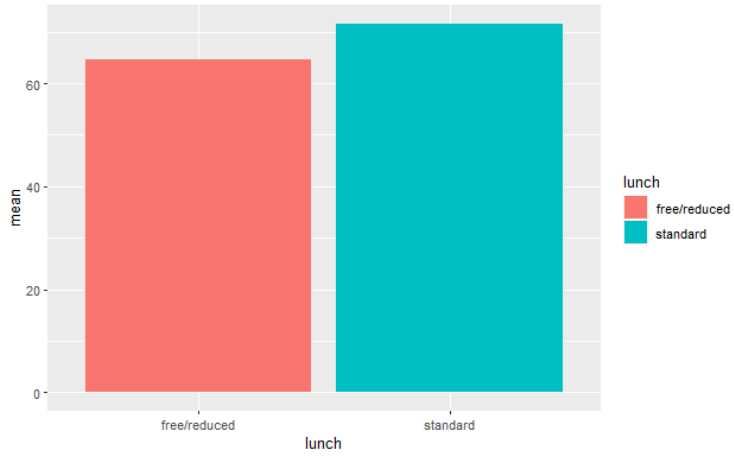
| lunch <chr> | mean <dbl> |
|---|---|
| free/reduced | 64.65352 |
| standard | 71.65426 |

2 rows

I used a group by to look at lunch and found the mean of the reading scores

```{r}
ggplot(data = mean_reading)+
  geom_col(mapping = aes(x = lunch ,y = mean, fill = lunch),stat = "identity")
```




Overall, a huge margin between those who have standard lunch and those who are on free/reduced lunch. It can be assume that the test preparation course was not free.

```{r}
mean_math <- StudentsPerformance %>%
  group_by(`parental level of education`) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))
mean_math
```

A tibble: 6 x 2

| parental level of education <chr> | mean <dbl> |
|---|---|
| associate's degree | 67.88288 |
| bachelor's degree | 69.38983 |
| high school | 62.13776 |
| master's degree | 69.74576 |
| some college | 67.12832 |
| some high school | 63.49721 |

6 rows

I used a group by to look at parental level of education and found the mean of the math scores

```{r}
ggplot(data = mean_math)+
  geom_col(mapping = aes(x = `parental level of education`,y = mean,fill = `parental level of education`),stat = "identity")

```

It appears that going to college made a big difference on whether or not a student performed well. I thought having a master's degree would create a bigger margin. Rather, whether or not the childs parents goes to college makes the difference.

```{r}
mean_writing <- StudentsPerformance %>%
  group_by(`parental level of education`) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))

mean_writing
```

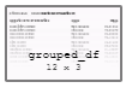
A tibble: 6 x 2

| parental level of education <chr> | mean <dbl> |
|---|---|
| associate's degree | 69.89640 |
| bachelor's degree | 73.38136 |
| high school | 62.44898 |
| master's degree | 75.67797 |
| some college | 68.84071 |
| some high school | 64.88827 |

6 rows

I used a group by to look at parental level of education and found the mean of the writing scores

```{r}
mean <- StudentsPerformance %>%
  group_by(`parental level of education`, lunch) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))

mean
```


A tibble: 12 x 3 Groups: parental level of education [6]

| parental level of education <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| associate's degree | free/reduced | 62.61039 |
| associate's degree | standard | 70.68276 |
| bachelor's degree | free/reduced | 63.00000 |
| bachelor's degree | standard | 73.18919 |
| high school | free/reduced | 54.51429 |
| high school | standard | 66.37302 |
| master's degree | free/reduced | 61.16667 |
| master's degree | standard | 75.62857 |
| some college | free/reduced | 60.12658 |
| some college | standard | 70.89116 |

1-10 of 12 rows Previous 1 2 Next

Here I am using a group by to see the parental level of education, the students lunch, and the mean of the math score

```
mean <- StudentsPerformance %>%
  group_by(`parental level of education`, lunch) %>%
    summarise(mean = mean(`reading score`, na.rm = TRUE))
mean
```


A tibble: 12 x 3 Groups: parental level of education [6]

| parental level of education <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| associate's degree | free/reduced | 67.64935 |
| associate's degree | standard | 72.66897 |
| bachelor's degree | free/reduced | 68.93182 |
| bachelor's degree | standard | 75.41892 |
| high school | free/reduced | 60.21429 |
| high school | standard | 67.19841 |
| master's degree | free/reduced | 70.29167 |
| master's degree | standard | 78.85714 |
| some college | free/reduced | 65.07595 |
| some college | standard | 71.81633 |

1-10 of 12 rows Previous 1 2 Next

Here I am using a group by to see the parental level of education, the students lunch, and the mean of the reading score

```{r}
mean <- StudentsPerformance %>%
  group_by(`parental level of education`, lunch) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```


A tibble: 12 x 3 Groups: parental level of education [6]

| parental level of education <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| master's degree | standard | 79.65714 |
| bachelor's degree | standard | 75.85135 |
| associate's degree | standard | 71.95172 |
| some college | standard | 71.59864 |
| master's degree | free/reduced | 69.87500 |
| bachelor's degree | free/reduced | 69.22727 |
| some high school | standard | 68.64407 |
| associate's degree | free/reduced | 66.02597 |
| high school | standard | 65.25397 |
| some college | free/reduced | 63.70886 |

```
mean <- StudentsPerformance %>%
  group_by(`parental level of education`, lunch) %>%
    summarise(mean = mean(`reading score`, na.rm = TRUE))
mean
```


A tibble: 12 x 3 Groups: parental level of education [6]

| parental level of education <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| associate's degree | free/reduced | 67.64935 |
| associate's degree | standard | 72.66897 |
| bachelor's degree | free/reduced | 68.93182 |
| bachelor's degree | standard | 75.41892 |
| high school | free/reduced | 60.21429 |
| high school | standard | 67.19841 |
| master's degree | free/reduced | 70.29167 |
| master's degree | standard | 78.85714 |
| some college | free/reduced | 65.07595 |
| some college | standard | 71.81633 |

1-10 of 12 rows Previous 1 2 Next

Here I am using a group by to see the parental level of education, the students lunch, and the mean of the reading score

```{r}
mean <- StudentsPerformance %>%
  group_by(`parental level of education`, lunch) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```


A tibble: 12 x 3 Groups: parental level of education [6]

| parental level of education <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| master's degree | standard | 79.65714 |
| bachelor's degree | standard | 75.85135 |
| associate's degree | standard | 71.95172 |
| some college | standard | 71.59864 |
| master's degree | free/reduced | 69.87500 |
| bachelor's degree | free/reduced | 69.22727 |
| some high school | standard | 68.64407 |
| associate's degree | free/reduced | 66.02597 |
| high school | standard | 65.25397 |
| some college | free/reduced | 63.70886 |

Despite children of high school drop out parents, student still performed worse despite their parents degree. having a free/reduced lunch has a huge impact regardless the degree.

Being able to afford a school lunch is highly significant to performing well on these exams. It is safe to say that those on free/reduced lunch most likely are apart of the lower class.

```
    summarise(mean = mean(`writing score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```

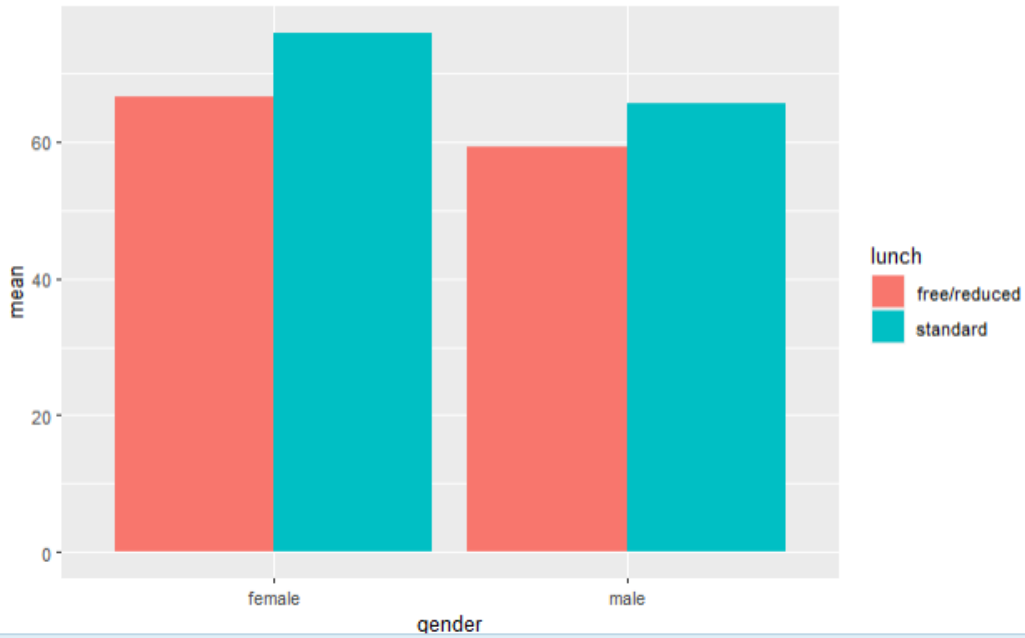


A tibble: 4 x 3 | Groups: gender [2]

| gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| female | standard | 75.92705 |
| female | free/reduced | 66.44444 |
| male | standard | 65.50949 |
| male | free/reduced | 59.12651 |

4 rows

```{r}
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = gender,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```



```
  group_by(`gender`, lunch) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```

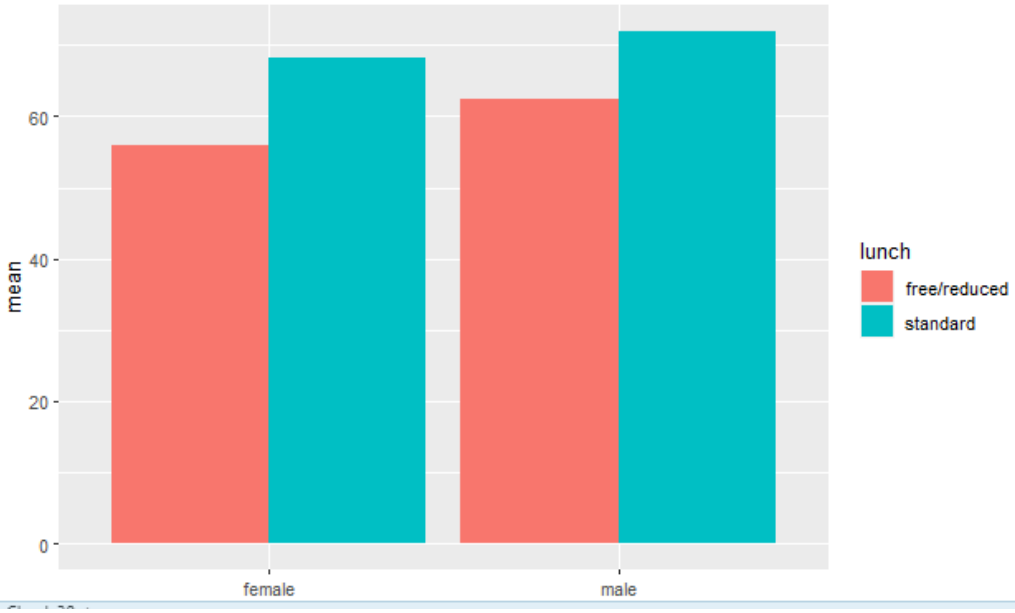


A tibble: 4 x 3 | Groups: gender [2]

| gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| male | standard | 72.02215 |
| female | standard | 68.12462 |
| male | free/reduced | 62.45783 |
| female | free/reduced | 55.81481 |

4 rows

```{r}
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = gender,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```



R Console

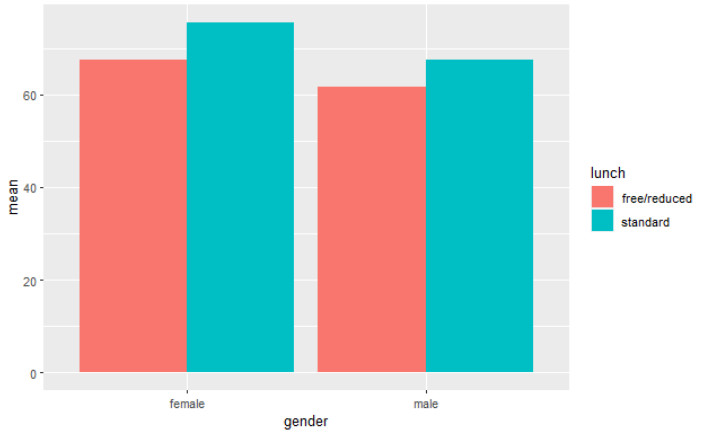
grouped_df
4 × 3

A tibble: 4 x 3 | Groups: gender [2]

| gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|
| female | standard | 75.60790 |
| male | standard | 67.53797 |
| female | free/reduced | 67.38624 |
| male | free/reduced | 61.54217 |

4 rows

```r
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = gender,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




Regardless of the lunch, the women dominate the reading and writing scores and men dominate the math scores. I am curious to wonder why this is occurring?

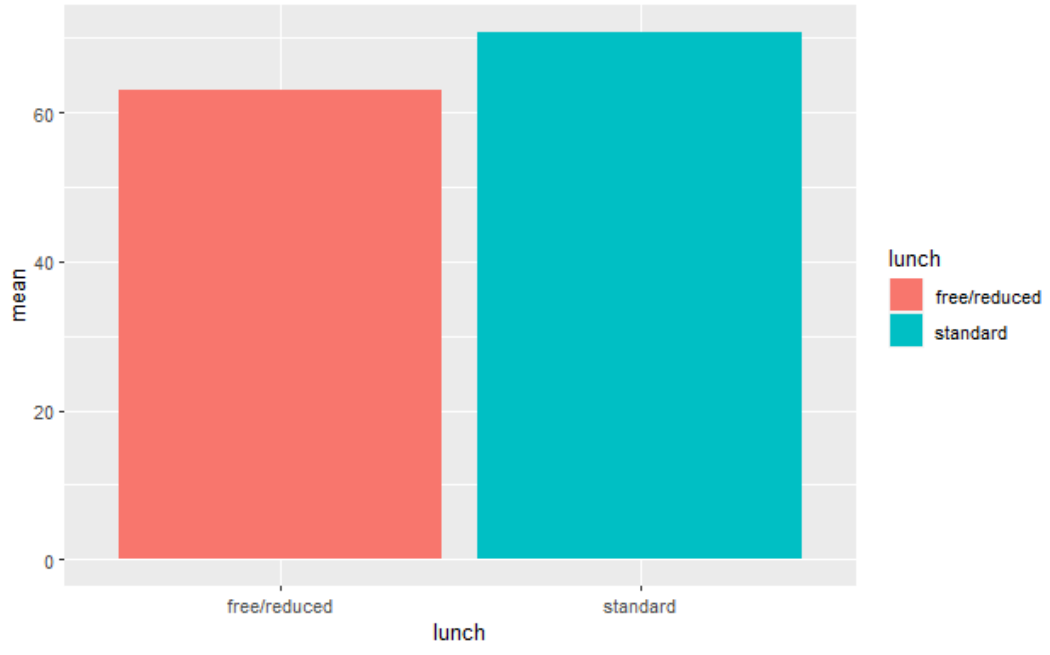
```r
mean <- StudentsPerformance %>%
  group_by(lunch) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))%>%
  arrange(desc(mean))

mean
```

A tibble: 2 x 2

| lunch <chr> | mean <dbl> |
|---|---|
| standard | 70.82326 |
| free/reduced | 63.02254 |

2 rows

```r
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = lunch,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




Chunk 33

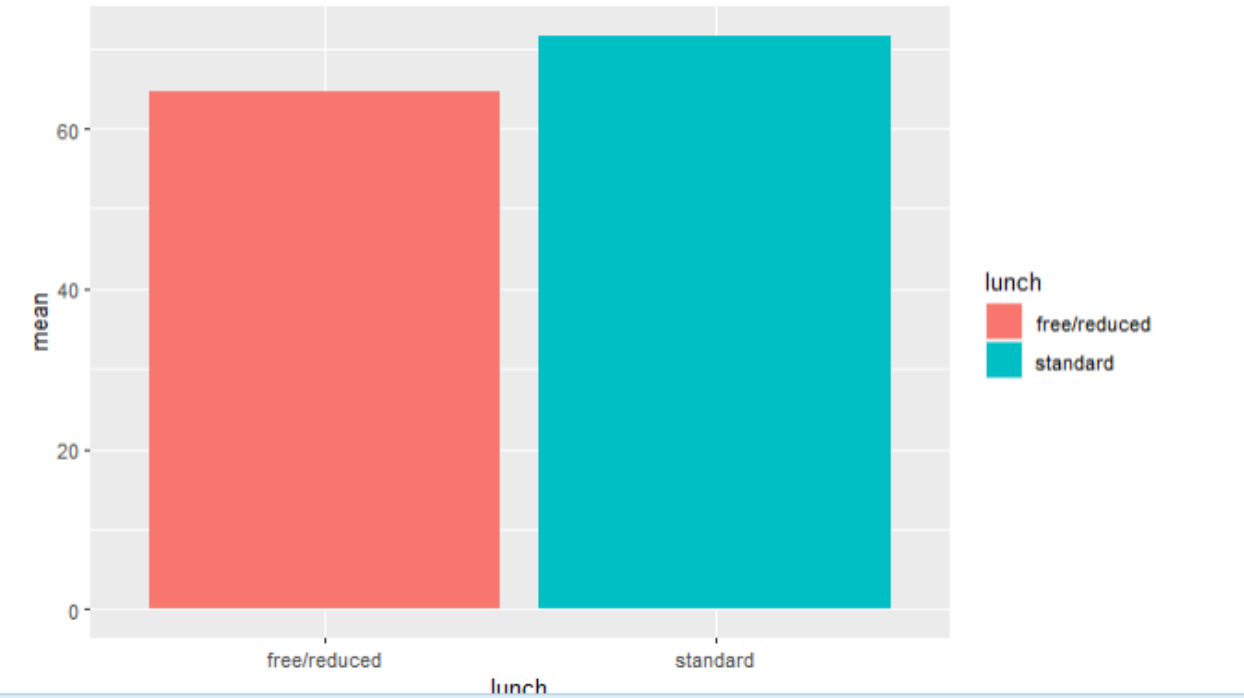
```r
mean <- StudentsPerformance %>%
  group_by(lunch) %>%
    summarise(mean = mean(`reading score`, na.rm = TRUE))%>%
  arrange(desc(mean))

mean
```

A tibble: 2 x 2

| lunch <chr> | mean <dbl> |
|---|---|
| standard | 71.65426 |
| free/reduced | 64.65352 |

2 rows

```r
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = lunch,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




Chunk 33

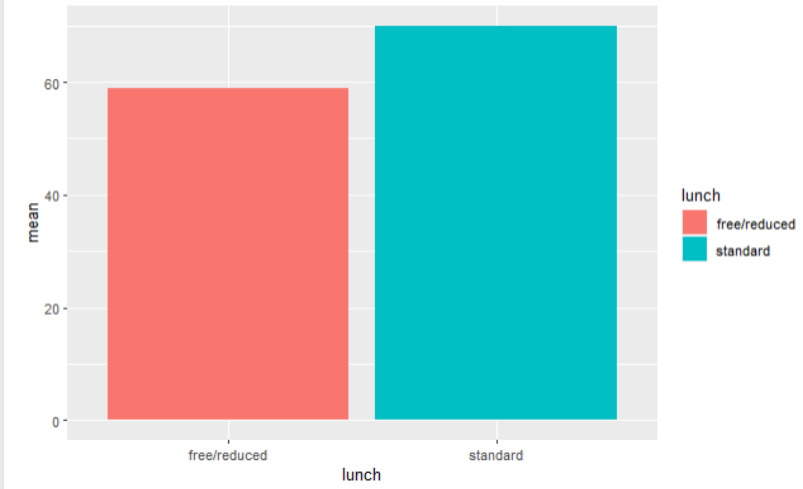
```r
  group_by(lunch) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```

A tibble: 2 x 2

| lunch <chr> | mean <dbl> |
|---|---|
| standard | 70.03411 |
| free/reduced | 58.92113 |

2 rows

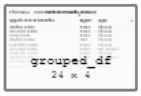
```r
ggplot(data = mean,                                    # Grouped barplot using ggplot2
       aes(x = lunch,
           y = mean,
           fill = lunch)) +
  geom_bar(stat = "identity",
           position = "dodge")
```




We see that students who are on a standard lunch are academically performing better  than those who are have free/reduced lunch

INTRODUCTION

```r
mean <- StudentsPerformance %>%
  group_by(`parental level of education`,`gender`, lunch) %>%
    summarise(mean = mean(`writing score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```




A tibble: 24 x 4 Groups: parental level of education, gender [12]

| parental level of education <chr> | gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|---|
| master's degree | female | standard | 81.22727 |
| bachelor's degree | female | standard | 79.75610 |
| some college | female | standard | 77.33784 |
| master's degree | male | standard | 77.00000 |
| associate's degree | female | standard | 76.02500 |
| bachelor's degree | female | free/reduced | 75.81818 |
| some high school | female | standard | 74.36364 |
| master's degree | female | free/reduced | 72.00000 |
| bachelor's degree | male | standard | 71.00000 |
| high school | female | standard | 70.66667 |

1-10 of 24 rows

```r
mean <- StudentsPerformance %>%
  group_by(`parental level of education`,`gender`, lunch) %>%
    summarise(mean = mean(`math score`, na.rm = TRUE))%>%
  arrange(desc(mean))
mean
```


A tibble: 24 x 4 Groups: parental level of education, gender [12]

| parental level of education <chr> | gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|---|
| master's degree | male | standard | 81.69231 |
| bachelor's degree | male | standard | 75.87879 |
| associate's degree | male | standard | 73.33846 |
| some college | male | standard | 72.26027 |
| master's degree | female | standard | 72.04545 |
| bachelor's degree | female | standard | 71.02439 |
| some high school | male | standard | 70.74603 |
| some college | female | standard | 69.54054 |

```r
mean <- StudentsPerformance %>%
  group_by(`parental level of education`,`gender`, lunch) %>%
    summarise(mean = mean(`reading score`, na.rm = TRUE))%>%
  arrange(desc(mean))

mean
```

R Console

grouped_df
24 x 4

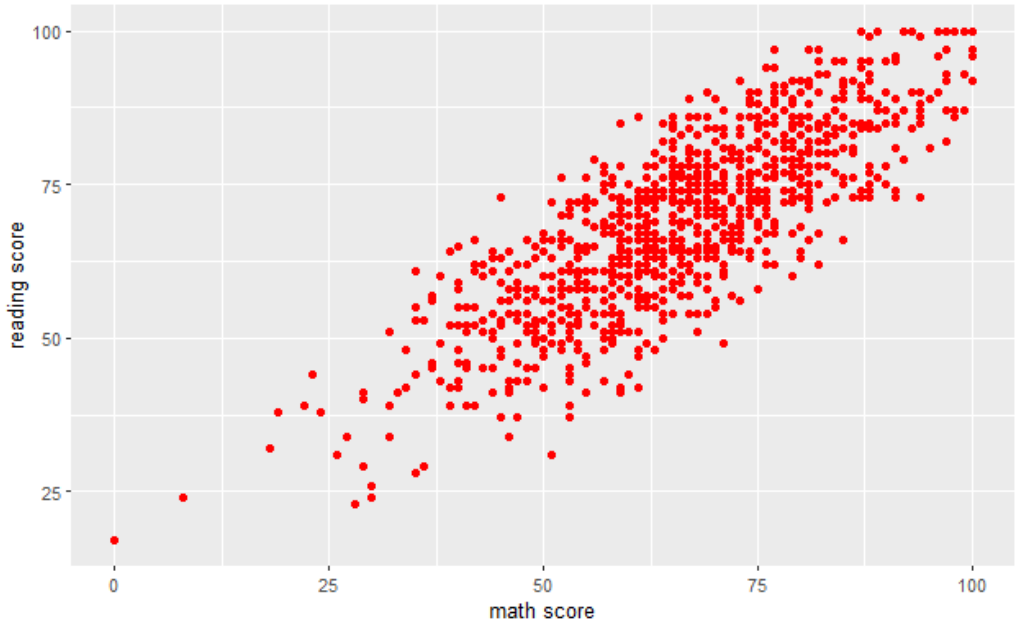
A tibble: 24 x 4 Groups: parental level of education, gender [12]

| parental level of education <chr> | gender <chr> | lunch <chr> | mean <dbl> |
|---|---|---|---|
| master's degree | female | standard | 79.72727 |
| bachelor's degree | female | standard | 79.04878 |
| master's degree | male | standard | 77.38462 |
| some college | female | standard | 76.16216 |
| associate's degree | female | standard | 75.86250 |
| some high school | female | standard | 74.47273 |
| bachelor's degree | female | free/reduced | 74.00000 |
| master's degree | female | free/reduced | 72.21429 |
| high school | female | standard | 71.56140 |
| bachelor's degree | male | standard | 70.90909 |

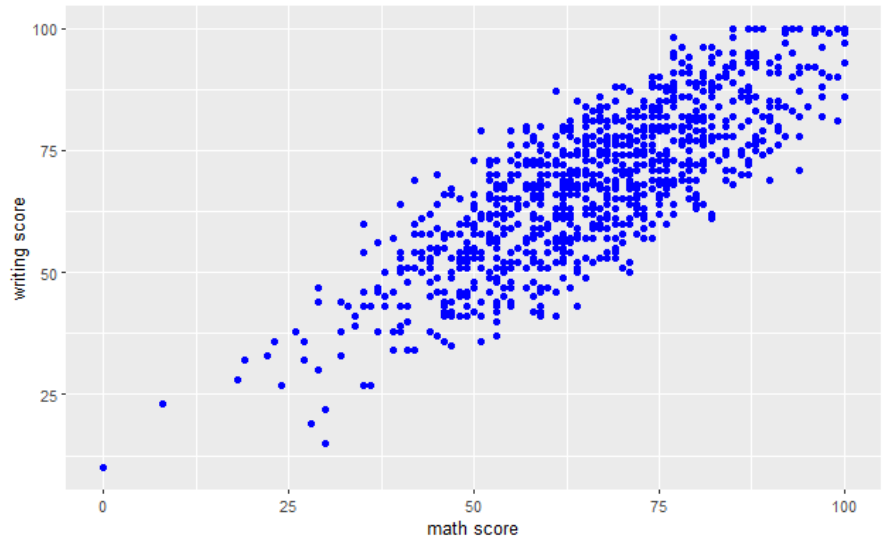
1-10 of 24 rows Previous 1 2 3 Next

As previously seen, women perform better on reading and writing exams and men preform better on math exams. What makes an exponential difference is the degree their parents have.

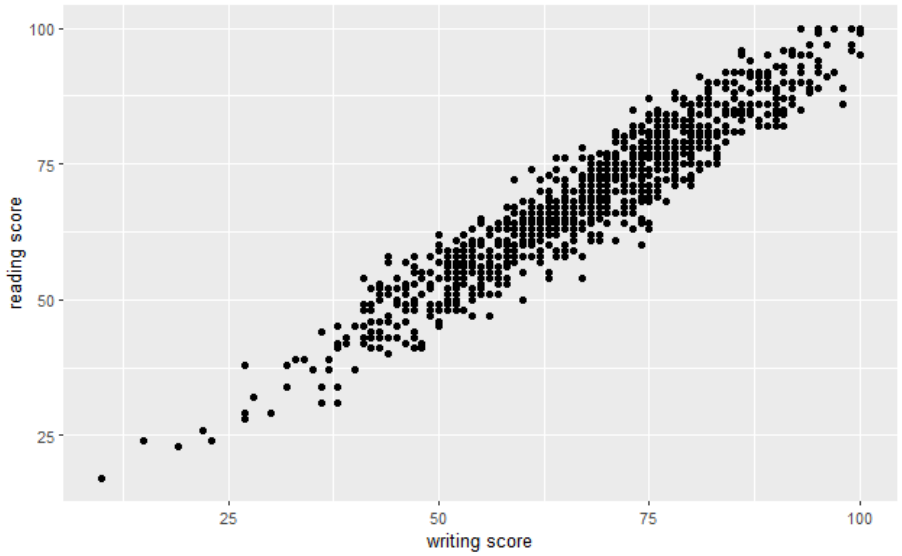
```r
ggplot(data = StudentsPerformance) +
  geom_point(mapping = aes(x = `math score` , y = `reading score`),  color = "red")
```



```r
ggplot(data = StudentsPerformance) +
  geom_point(mapping = aes(x = `math score` , y = `writing score`),  color = "blue")
```




```r
ggplot(data = StudentsPerformance) +
  geom_point(mapping = aes(x = `writing score` , y = `reading score`),  color = "black")
```



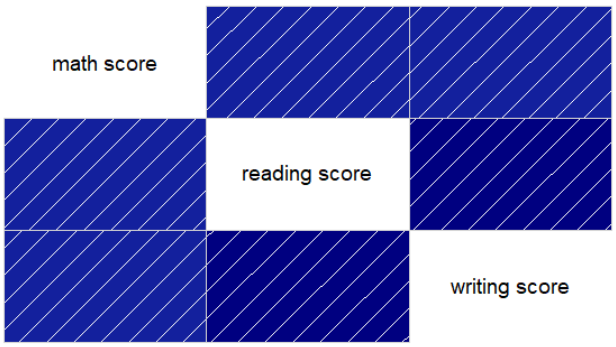

I created scatterplots to find a relationship between each of the scores. Below is where I found the correlation

```r
score <- select(StudentsPerformance, 'math score', 'reading score', 'writing score')
```

```r
cor(score)
```

```
              math score reading score writing score
math score     1.0000000     0.8175797     0.8026420
reading score  0.8175797     1.0000000     0.9545981
writing score  0.8026420     0.9545981     1.0000000
```

```r
#install.packages("corrgram")
library(corrgram)
corrgram(score)
```




We can see a positive correlation between all the scores. There is a deeper correlation between the reading score and the writing score. To me, this is no surprise knowing how closely related reading and writing are.

```r
score <- lm(`writing score` ~ `reading score`, data = StudentsPerformance)
```

```r
summary(score)
```

```
Call:
lm(formula = `writing score` ~ `reading score`, data = StudentsPerformance)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.9573  -2.9573   0.0363   3.1026  15.0557 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -0.667554   0.693792  -0.962    0.336    
`reading score`  0.993531   0.009814 101.233   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.529 on 998 degrees of freedom
Multiple R-squared:  0.9113,    Adjusted R-squared:  0.9112 
F-statistic: 1.025e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```

reading score
#Equation = y hat = -0.6676 + (-0.99)
#equation = y hat = 0.7872

writing score = -0.667 + 0.99 (reading score)

```r
#CONCLUSION
```

After thoroughly going through this data set, I have found many answers that I have been looking. I found that despite the student's parent's degree, or if they are on free/produced lunch, women perform better in reading and writing, and men perform better in math. In future projects, I would like to find more data that explains this situation. Due to running multiple tests on this, I do not believe this is a coincidence. I found students with free and reduced lunch perform exponentially worse than those on a standard lunch plan. If you are a student with a parent who went to college but is on free/reduced lunch, you remain a higher chance of outperforming your peers on a standard lunch plan with parents who did not go to college. Having parents who are well educated and can afford the school's lunch plays a crucial role in how well their child will do in school.

If I had to do further research, I would dive deeper into the parental background. I would like to see their annual income if they are divorced, and their racial/ethnic background. Once completed, I will have a more in-depth understanding of what students need from their parents to succeed.