Student Performance

Aidan Blassage

6/3/2021

In this project I analyzed Student performance from the Kaggle Dataset “Students Performance in Exams” made availabe by Royce Kimmons at <http://roycekimmons.com/tools/generated_data/exams>

colnames(StudentsPerformance)

## [1] "gender" "race.ethnicity"   
## [3] "parental.level.of.education" "lunch"   
## [5] "test.preparation.course" "math.score"   
## [7] "reading.score" "writing.score"   
## [9] "avg\_score"

# Ask

* Do male and female students score the same?
* Do low income families whose children receive sponsored lunches do better or worse?
* Does test preparation increase scores?

# Process

* I took the data to R to begin the analysis. I split the gender column into male and female so I could begin.

df <- (StudentsPerformance)  
df1 <- subset(df, df$gender == "male")  
df2 <- subset(df, df$gender == "female")

* I then averaged the scores between math, reading, and writing for both male and female test takers.

df1$math.avg = mean(df1$"math.score")  
df1$reading.avg = mean(df1$`reading.score`)  
df1$writing.avg = mean(df1$`writing.score`)  
  
df2$math.avg = mean(df2$'math.score')  
df2$reading.avg = mean(df2$`reading.score`)  
df2$writing.avg = mean(df2$`writing.score`)

\*After finding all of the averages I recombined the two genders into a new table for further analysis.

male\_fem\_comb <- rbind(  
 within(df1, {DS <- 'df1'}),  
 within(df2, {DS <- 'df2'})  
)

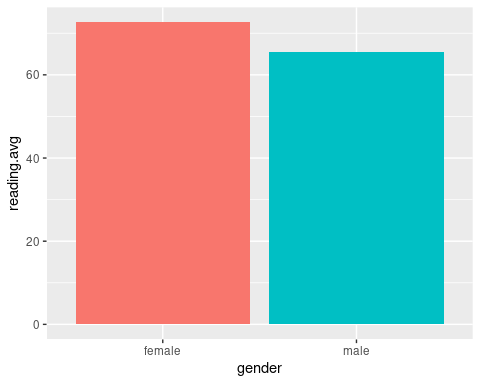
* After this I ran a check to ensure that the data combined correctly.

colnames(male\_fem\_comb)

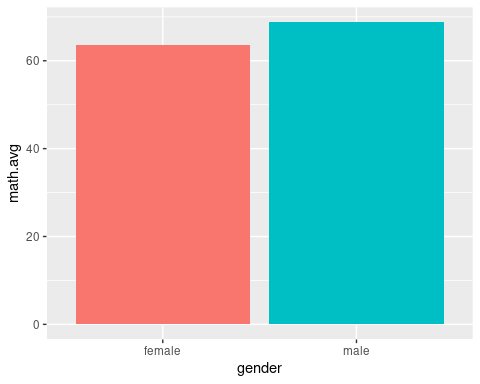
## [1] "gender" "race.ethnicity"   
## [3] "parental.level.of.education" "lunch"   
## [5] "test.preparation.course" "math.score"   
## [7] "reading.score" "writing.score"   
## [9] "avg\_score" "math.avg"   
## [11] "reading.avg" "writing.avg"   
## [13] "DS"

\*I then began the analysis within R

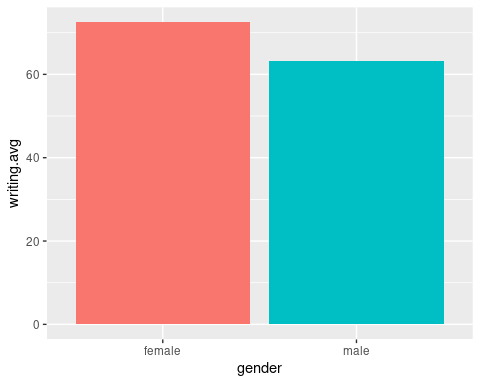
ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=gender, y = reading.avg,fill=gender), position = 'dodge')+  
 theme(legend.position = "none")



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=gender,y=math.avg,fill=gender), position = 'dodge')+  
 theme(legend.position = 'none')



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x =gender,y=writing.avg,fill=gender), position = 'dodge')+  
 theme(legend.position = 'none')



* We can see that there is a difference in performance between genders on different subjects, males lead in math, while females lead in writing and reading.
* My next step is to see if there is any effect on test scores based on gender and lunch assistance or no lunch assistance.

df3 <- subset(df, df$lunch == 'free/reduced' )  
df4 <- subset(df, df$lunch == 'standard')

df5 <- subset(df3, df3$gender == 'male')  
df6 <- subset(df3, df3$gender == 'female')  
df7 <- subset(df4, df4$gender == 'male')  
df8 <- subset(df4, df4$gender == 'female')

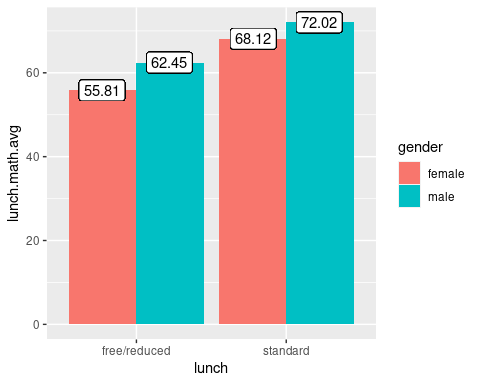
* I split the data into 4 different groups, based on lunch style and gender.

df5$lunch.reading.avg = mean(df5$"reading.score")  
df6$lunch.reading.avg = mean(df6$"reading.score")  
df7$lunch.reading.avg = mean(df7$"reading.score")  
df8$lunch.reading.avg = mean(df8$"reading.score")  
  
df5$lunch.math.avg = mean(df5$"math.score")  
df6$lunch.math.avg = mean(df6$"math.score")  
df7$lunch.math.avg = mean(df7$"math.score")  
df8$lunch.math.avg = mean(df8$"math.score")  
  
df5$lunch.writing.avg = mean(df5$"writing.score")  
df6$lunch.writing.avg = mean(df6$"writing.score")  
df7$lunch.writing.avg = mean(df7$"writing.score")  
df8$lunch.writing.avg = mean(df8$"writing.score")

* Across the 4 different groups I now have the averages of all the scores within each subcategory and can begin my analysis

male\_fem\_comb <- rbind(  
 within(df5, {DS <- 'df5'}),  
 within(df6, {DS <- 'df6'}),  
 within(df7, {DS <- 'df7'}),  
 within(df8, {DS <- 'df8'})  
)

ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=lunch,y=lunch.math.avg,fill=gender), position = 'dodge')+  
 geom\_label(label ="62.45", x = 1.22, y = 62.45)+  
 geom\_label(label = "55.81", x = .77, y = 55.81)+  
 geom\_label(label ="72.02", x = 2.22, y = 72.02)+  
 geom\_label(label ="68.12", x = 1.78, y = 68.12)



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=lunch,y=lunch.reading.avg,fill=gender), position = 'dodge')+  
 geom\_label(label = "61.54", x = 1.22, y = 61.54)+  
 geom\_label(label ="67.38", x = .77, y = 67.38)+  
 geom\_label(label = "67.53", x = 2.22, y = 67.53)+  
 geom\_label(label = "75.61", x = 1.78, y = 75.61)



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=lunch,y=lunch.writing.avg,fill=gender), position = 'dodge')+  
 geom\_label(label ="59.12", x = 1.22 , y =59.12)+  
 geom\_label(label ="66.44", x = .77 , y =66.44)+  
 geom\_label(label ="65.51", x = 2.22, y =65.51)+  
 geom\_label(label ="75.92", x = 1.78, y =75.92)



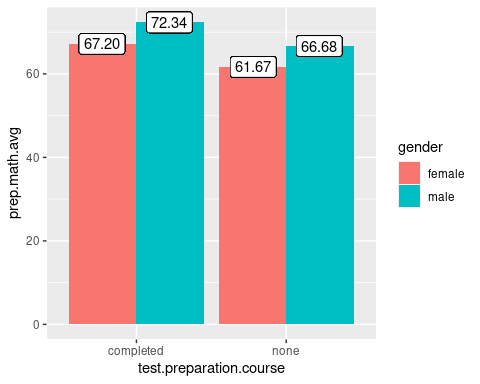
* These charts show that free/reduced lunch has a negative correlation with test scores within this dataset.
* Females have a stronger negative correlation when on free/reduced lunch concerning test scores. For the writing portion, female scores lowered by 12.5% compared to 9.7% for males. For reading, female scores lowered by 10.9% compared to 9.0% for males. Finally, for math female scores were lowered an astounding 18.1% compared to males with a 13.3% lowering compared to their counterparts who had standard lunch.

df9 <- subset(df, df$test.preparation.course == 'completed' )  
df10 <- subset(df, df$test.preparation.course == 'none')  
  
df11 <- subset(df9, df9$gender == 'male')  
df12 <- subset(df9, df9$gender == 'female')  
df13 <- subset(df10, df10$gender == 'male')  
df14 <- subset(df10, df10$gender == 'female')

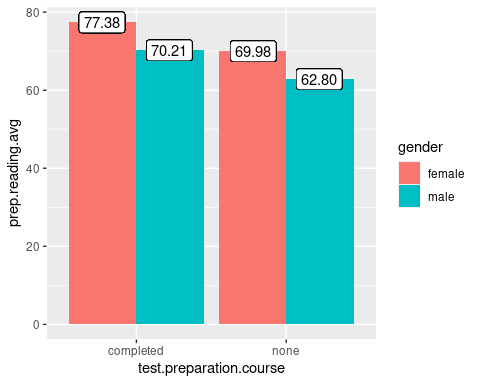
df11$prep.math.avg = mean(df11$'math.score')  
df12$prep.math.avg = mean(df12$'math.score')  
df13$prep.math.avg = mean(df13$'math.score')  
df14$prep.math.avg = mean(df14$'math.score')  
  
df11$prep.reading.avg = mean(df11$'reading.score')  
df12$prep.reading.avg = mean(df12$'reading.score')  
df13$prep.reading.avg = mean(df13$'reading.score')  
df14$prep.reading.avg = mean(df14$'reading.score')  
  
  
df11$prep.writing.avg = mean(df11$'writing.score')  
df12$prep.writing.avg = mean(df12$'writing.score')  
df13$prep.writing.avg = mean(df13$'writing.score')  
df14$prep.writing.avg = mean(df14$'writing.score')

male\_fem\_comb <- rbind(  
 within(df11, {DS2 <- 'df11'}),  
 within(df12, {DS2 <- 'df12'}),  
 within(df13, {DS2 <- 'df13'}),  
 within(df14, {DS2 <- 'df14'})  
)

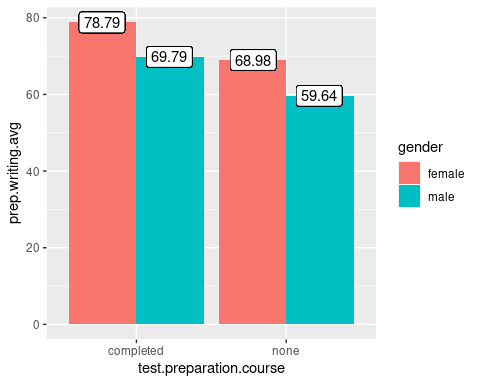
ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=test.preparation.course,y=prep.math.avg,fill=gender), position = 'dodge')+  
 geom\_label(label ="72.34", x = 1.22, y = 72.34 )+  
 geom\_label(label = "67.20", x = .77, y = 67.20)+  
 geom\_label(label ="66.68", x = 2.22, y = 66.68 )+  
 geom\_label(label ="61.67", x = 1.78, y = 61.67 )



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x= test.preparation.course,y=prep.reading.avg,fill=gender), position = 'dodge')+  
 geom\_label(label = "70.21", x = 1.22, y = 70.21)+  
 geom\_label(label ="77.38", x = .77, y = 77.38)+  
 geom\_label(label = "62.80", x = 2.22, y = 62.80)+  
 geom\_label(label = "69.98", x = 1.78, y = 69.98)



ggplot(data = male\_fem\_comb)+  
 geom\_col(mapping = aes(x=test.preparation.course,y=prep.writing.avg,fill=gender), position = 'dodge')+  
 geom\_label(label ="69.79", x = 1.22 , y =69.79)+  
 geom\_label(label ="78.79", x = .77 , y =78.79)+  
 geom\_label(label ="59.64", x = 2.22, y =59.64)+  
 geom\_label(label ="68.98", x = 1.78, y =68.98)



* Those who completed the test preparation course scored higher in all areas. A similar proportion of females and males took the test preparation course, though potential bias may occur as better students that already would have scored higher may have taken the test preparation course and the gap could be explained by that.
* The charts show that the test prep had a slightly larger impact on males than females. For the writing, females saw a 12.5% increase if they did the test prep, while males saw a 14.5% increase. For the reading, females saw a 9.5% increase while males saw a raise in score of 10.5%. Lastly, females saw a larger increase in math with the test prep with 8.2% increase, while males saw a 7.8% increase, though this is the smallest gap between increases and males scored higher overall in math.

**Conclusion**

Through this data set we found that female students score higher than males except on the math portion. We also found that students who receive free or reduced lunches score lower than their counterparts that pay standard lunch. Lastly, we found that test preparation raises the scores of everyone that participated but this finding may have a bias due to the students that would choose to do the test preparation.