Math Scores for Different Teaching Styles

Heather Shaw

04-28-2022

Three teachers at a junior high school have different opinions about which teaching method is most effective for 7th and 8th grade math students. Ms. Wesson uses a traditional approach, while Ms. Ruger and Ms. Smith use a standards-based method. The first task is to determine which approach is more effective, based on the math scores of their current students.

In addition, it has been suggested that each teacher has strengths that will make them more effective with students in certain ethnic groups. One teacher believes that students should be divided into classes based on ethnicity. The second task is to determine if there is a difference between student performance for each teacher, based on student demographics, that would justify such grouping.

Another proposal was that students should be grouped according to ability within each classroom. The teacher making this suggestion referenced an article, "Math and Reading Instruction in Tracked First-Grade Classes" (Stephen M Ross, et.al). The third task is to review this study to determine if ability grouping is supported by the findings.

The data includes math scores for students in grades 7-8 taught by the three teachers in question. It does not include scores for all 7th and 8th grade students in the school. The data includes limited demographic information regarding student ethnicity, gender, and free or paid lunch status.

The information regarding free or paid lunch is often used to give insight into student economic status. This can be helpful, but is not completely accurate. Families are not required to complete the application for financial assistance with meals, and there may be economically disadvantaged students not included in this statistic.

It should be noted that this data includes only one score for each student. A more effective method for measuring the quality of instruction would be a collection of multiple scores that could be used to demonstrate student growth. More information about this data project can be found at Kaggle.com.

```r
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
knitr::opts_chunk$set(echo = TRUE)
```

```r
library("tidyr")
library("dplyr")
library("foreign")
library("ggplot2")
#load data
math <- read.spss("1ResearchProjectData.sav", to.data.frame = TRUE)
```

Clean the Data

Preview Data Frame

```
head(math)
```

```
##   Student Teacher Gender           Ethnic   Freeredu Score      wesson
## 1       1   Ruger Female            Asian Free lunch    76 Ruger_Smith
## 2       2   Ruger Female         Hispanic Paid lunch    56 Ruger_Smith
## 3       3   Ruger Female African-American Free lunch    34 Ruger_Smith
## 4       4   Ruger Female            Asian Paid lunch    59 Ruger_Smith
## 5       5   Ruger   Male         Hispanic Free lunch    73 Ruger_Smith
## 6       6   Ruger   Male        Caucasian Paid lunch    58 Ruger_Smith
```

Tidy Columns

```
#rename columns to better describe data

math <- math %>%
  rename(Method="wesson")

math <- math %>%
  rename(Lunch="Freeredu")

math<- math %>%
  rename(Ethnicity="Ethnic")
colnames(math)
```

```
## [1] "Student"   "Teacher"   "Gender"    "Ethnicity" "Lunch"     "Score"
## [7] "Method"
```

```
#drop word "lunch" from lunch status descriptions
math <- math %>%
  mutate(Lunch=gsub(' lunch','',Lunch))

#change Method factors Ruger_Smith = standards, Wesson = traditional
math$Method <- sub("Ruger_Smith","Standards",math$Method)
math$Method <- sub("Wesson","Traditional",math$Method)

#check results
head(math)
```

```
##   Student Teacher Gender        Ethnicity Lunch Score    Method
## 1       1   Ruger Female            Asian  Free    76 Standards
## 2       2   Ruger Female         Hispanic  Paid    56 Standards
## 3       3   Ruger Female African-American  Free    34 Standards
## 4       4   Ruger Female            Asian  Paid    59 Standards
## 5       5   Ruger   Male         Hispanic  Free    73 Standards
## 6       6   Ruger   Male        Caucasian  Paid    58 Standards
```

Check for Duplicates

```
#check for duplicates
math %>%
  duplicated() %>%
```

```r
  table()
```

```
## .
## FALSE
##   217
```

```r
#none found
```

Omit Missing Values

```r
#remove missing values
math <- na.omit(math)
```

```r
#set plot theme for document
theme_update(plot.title = element_text(hjust = 0.5),
    panel.background = element_blank(),
    axis.line = element_line(color = "black"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 10))
```

Analysis

Inspect Population

```r
#total students
pop_total <- math %>%
  summarize(count=n())
pop_total
```

```
##   count
## 1   216
```

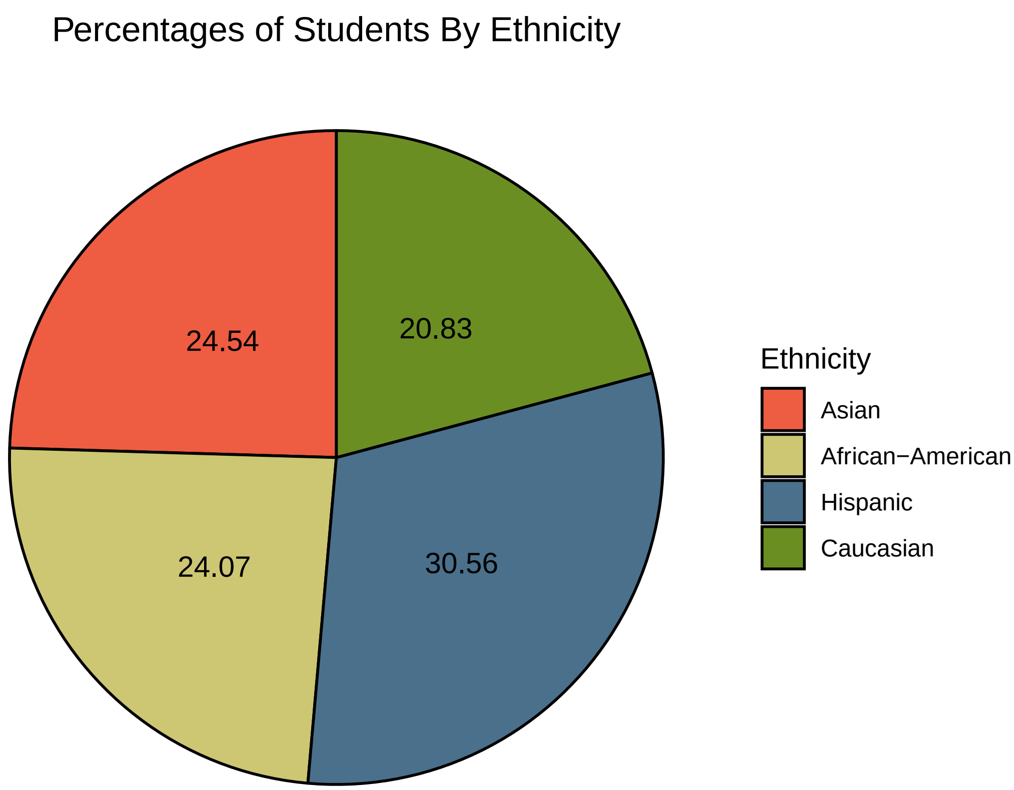
Student Demographics - Ethnicity

```r
#student ethnicity
pop_ethnicity <- math %>%
  group_by(Ethnicity) %>%
  summarize(count=n()) %>%
  mutate(Perc=(count/216)*100)
pop_ethnicity
```

```
## # A tibble: 4 x 3
##   Ethnicity        count  Perc
##   <fct>            <int> <dbl>
## 1 Asian               53  24.5
## 2 African-American    52  24.1
## 3 Hispanic            66  30.6
## 4 Caucasian           45  20.8
```

```r
#plot
pop_ethnicity_viz <- ggplot(pop_ethnicity,aes(x="",y = Perc, fill = Ethnicity,)) +
  geom_col(color="black")+
  scale_fill_manual(values =c("tomato2","khaki3","skyblue4","olivedrab"))+
  geom_text(aes(label = round(Perc,2)),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")+
  labs(title="Percentages of Students By Ethnicity")+
  theme_void()+
  theme(plot.title = element_text(hjust = 0.5))

pop_ethnicity_viz
```

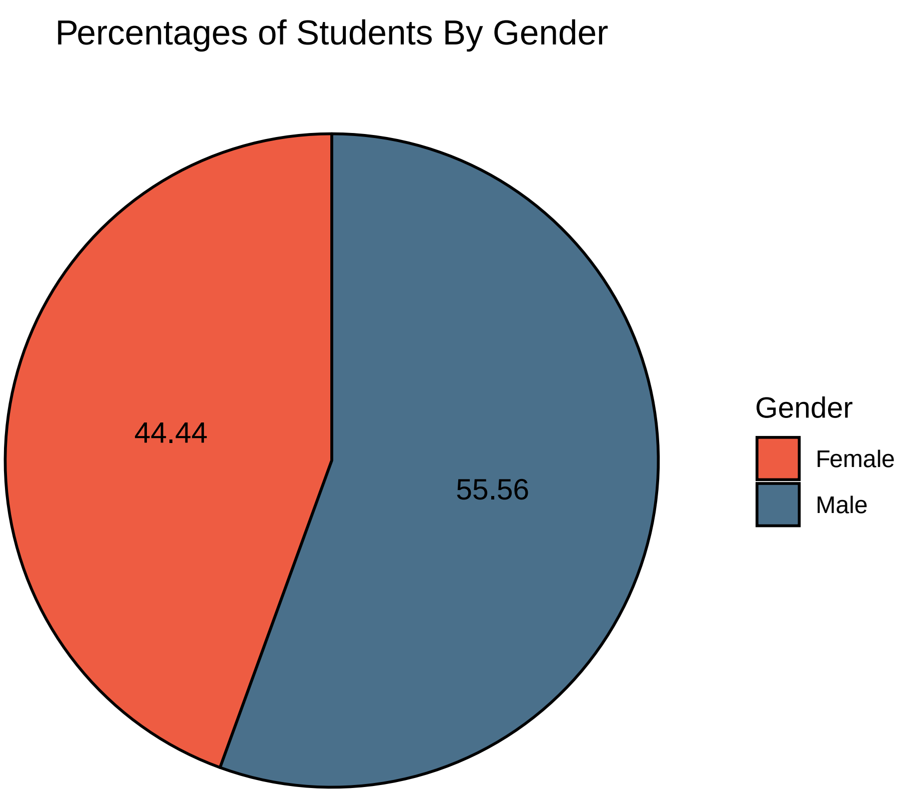

Percentages of Students By Ethnicity
24.54
20.83
24.07
30.56
Ethnicity
Asian
African−American
Hispanic
Caucasian

Student Demographics - Gender

```r
#Gender
pop_gender <- math %>%
  group_by(Gender)%>%
  summarize(count=n())%>%
  mutate(Perc=(count/216)*100)
pop_gender
```

```
## # A tibble: 2 x 3
##   Gender count  Perc
##   <fct>  <int> <dbl>
## 1 Female    96  44.4
## 2 Male     120  55.6
```

```r
#plot
pop_gender_viz <- ggplot(pop_gender,aes(x="",y = Perc, fill = Gender,)) +
  geom_col(color="black")+
  scale_fill_manual(values =c("tomato2","skyblue4"))+
  geom_text(aes(label = round(Perc,2)),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")+
  labs(title="Percentages of Students By Gender")+
  theme_void()+
  theme(plot.title = element_text(hjust = 0.5))
pop_gender_viz
```

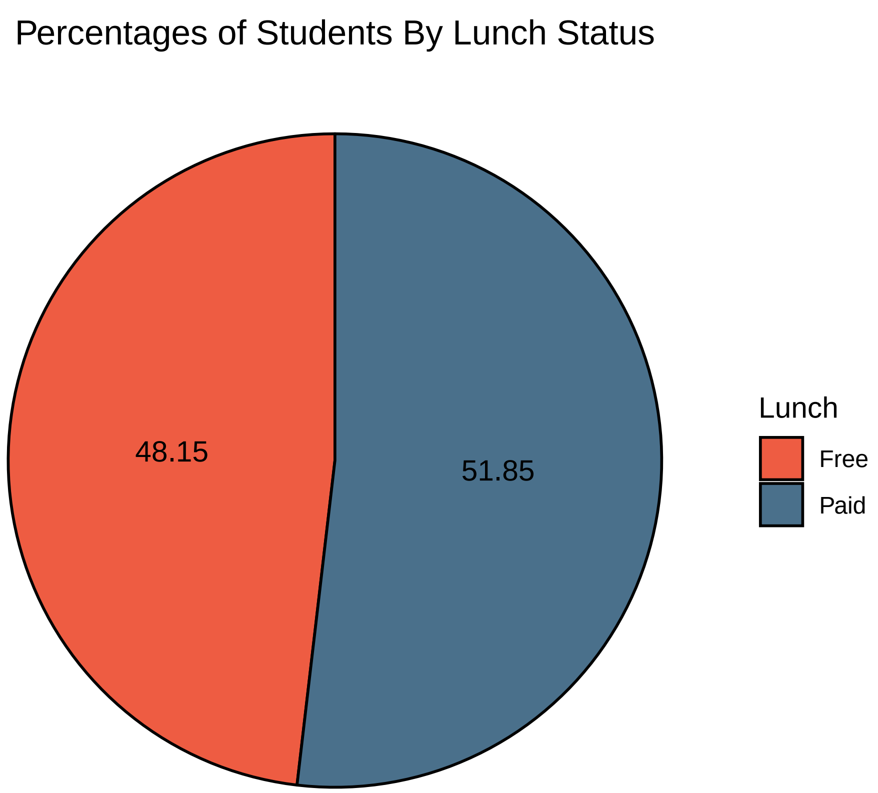


Student Demographics - Lunch Status

```
#Lunch Status
pop_lunch_status <- math %>%
  group_by(Lunch)%>%
  summarize(count=n())%>%
  mutate(Perc=(count/216)*100)
pop_lunch_status
```

```
## # A tibble: 2 x 3
##   Lunch count  Perc
##   <chr> <int> <dbl>
## 1 Free    104  48.1
## 2 Paid    112  51.9
```

```
#plot
pop_lunch_viz <- ggplot(pop_lunch_status,aes(x="",y = Perc, fill = Lunch,)) +
  geom_col(color="black")+
  scale_fill_manual(values =c("tomato2","skyblue4"))+
  geom_text(aes(label = round(Perc,2)),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")+
  labs(title="Percentages of Students By Lunch Status")+
  theme_void()+
  theme(plot.title = element_text(hjust = 0.5))
pop_lunch_viz
```



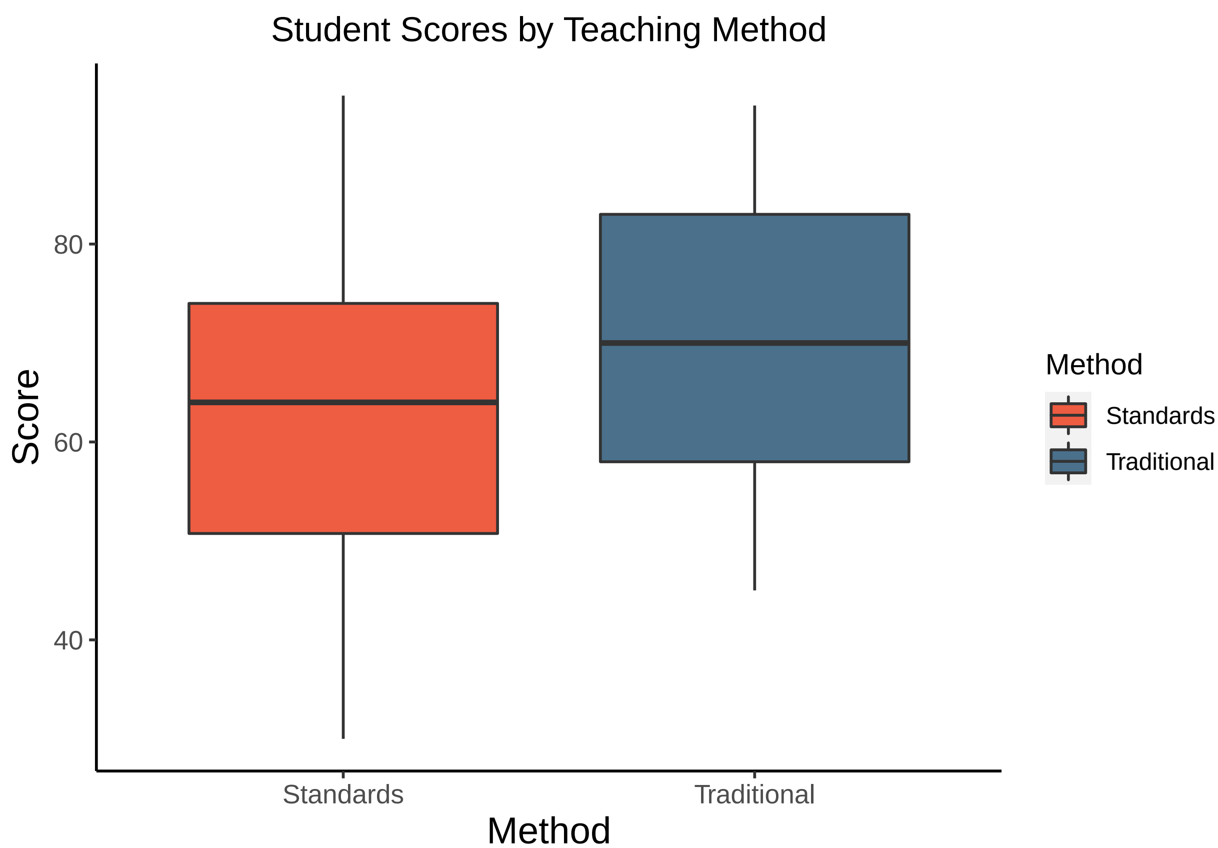
Compare Methods

Initial comparison of average test scores for each method shows higher scores for the traditional method (70.7), compared to the standards-based method (62.3). The difference between these two methods is statistically significant.

```
#avg score by Method
avg_score_method <- math %>%
  group_by(Method) %>%
  summarize(Avg_Score=mean(Score,na.rm=TRUE))
avg_score_method
```

```
## # A tibble: 2 x 2
##   Method      Avg_Score
##   <chr>           <dbl>
## 1 Standards        62.3
## 2 Traditional      70.7
```

```
qplot(data = math, x=Method, y=Score, fill = Method, geom = "boxplot")+
  labs(title="Student Scores by Teaching Method")+
  scale_fill_manual(values =c("tomato2","skyblue4"))
```




```
#H0: Mean Scores are the same for all methods
t.test(Score~Method,data=math)
```

```
##
##  Welch Two Sample t-test
##
## data:  Score by Method
## t = -3.8515, df = 174.6, p-value = 0.0001647
## alternative hypothesis: true difference in means between group Standards and group Traditional is no
## 95 percent confidence interval:
##  -12.626599  -4.070393
## sample estimates:
##   mean in group Standards mean in group Traditional
##                  62.33571                  70.68421
```

Compare Scores Within Demographic Groups

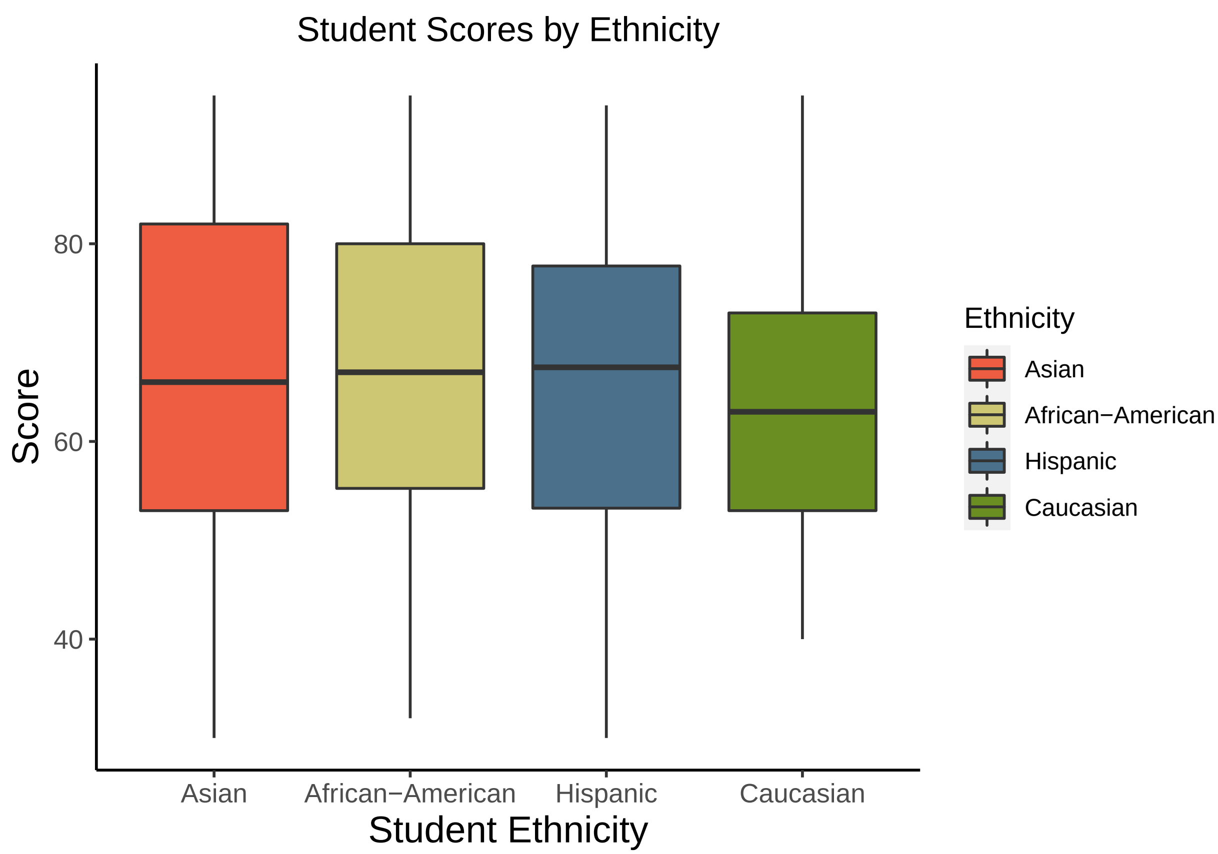
Analysis shows no significant difference in student scores when grouped by demographic criteria.

Ethnicity

```r
#avg score by ethnicity
avg_score_ethnicity <- math %>%
    group_by(Ethnicity) %>%
    summarize(Avg_Score=mean(Score,na.rm=TRUE))
avg_score_ethnicity
```

```
## # A tibble: 4 x 2
##   Ethnicity        Avg_Score
##   <fct>                <dbl>
## 1 Asian                 65.7
## 2 African-American      66.2
## 3 Hispanic              64.8
## 4 Caucasian             64.3
```

```r
#Comparison of scores for each ethnic group.
qplot(data = math, x=Ethnicity, y=Score, fill = Ethnicity, geom = "boxplot") +
  scale_fill_manual(values = c("tomato2","khaki3","skyblue4","olivedrab"))+
  labs(
    title = "Student Scores by Ethnicity", x = "Student Ethnicity", y = "Score")+
  theme(
    panel.background = element_blank(),
    axis.line = element_line(color = "black"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 10))
```



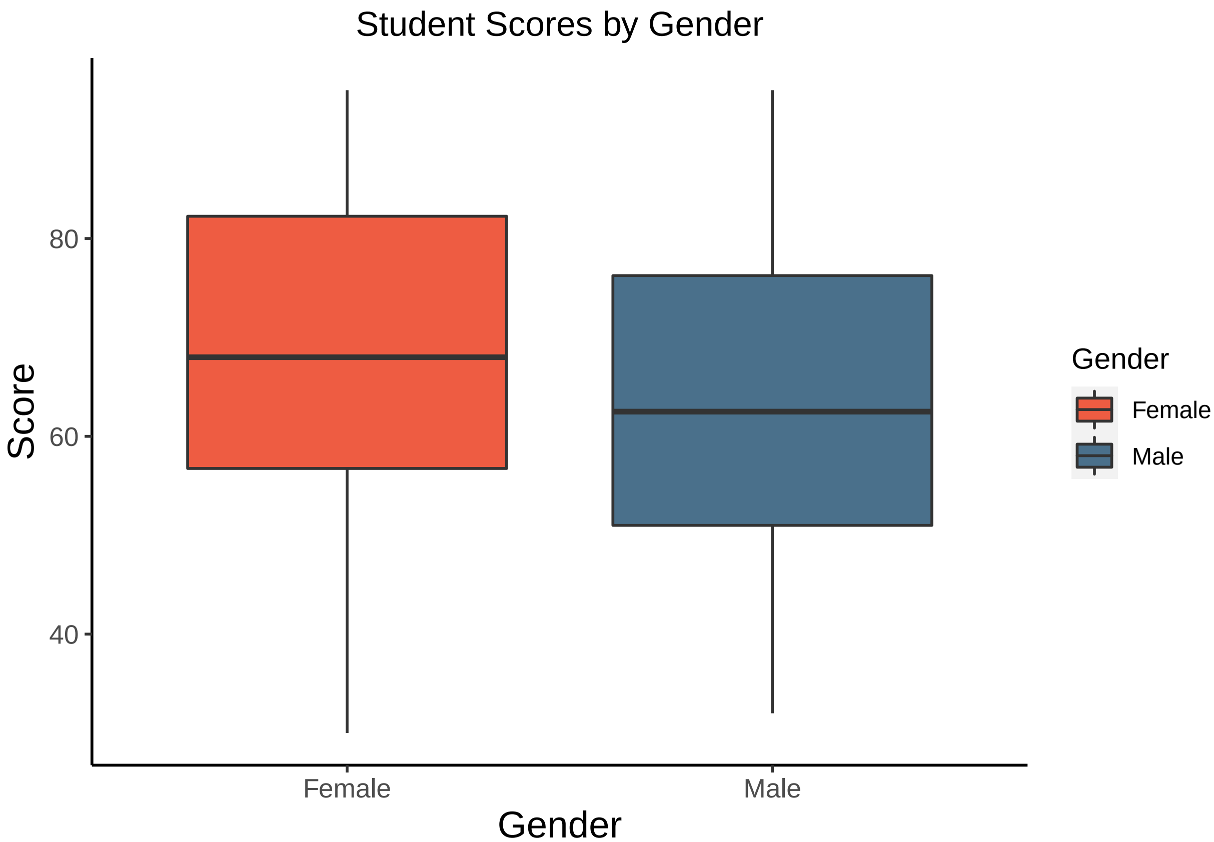

```
#H0: Mean Scores are the same for all ethnic groups.
anova1 <- aov(formula = Score ~ Ethnicity, data = math)
```

Gender

```
#avg score by gender
avg_score_gender <- math %>%
  group_by(Gender)%>%
  summarize(Avg_Score=mean(Score,na.rm=TRUE))
avg_score_gender
```

```
## # A tibble: 2 x 2
##   Gender Avg_Score
##   <fct>      <dbl>
## 1 Female      67.7
## 2 Male        63.4
```

```r
#Comparison of scores for each gender group.
qplot(data = math, x=Gender, y=Score, fill = Gender, geom = "boxplot")+
  labs(title="Student Scores by Gender")+
  scale_fill_manual(values =c("tomato2","skyblue4"))
```




```r
#H0: Mean Scores are the same for all gender groups.
 t.test(Score~Gender,data=math)
```

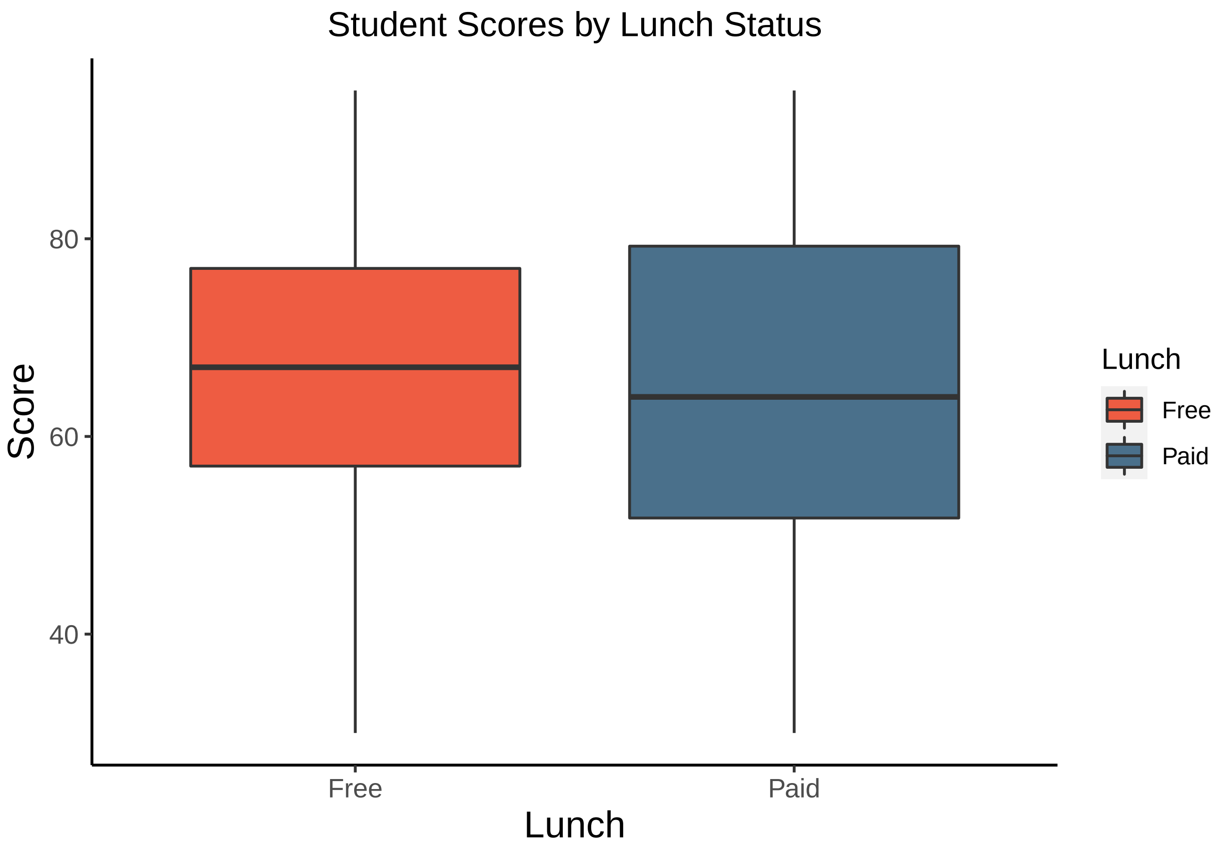
```
##
##  Welch Two Sample t-test
##
## data:  Score by Gender
## t = 1.9313, df = 205.5, p-value = 0.05482
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.0894566  8.6686233
## sample estimates:
## mean in group Female   mean in group Male
##             67.65625             63.36667
```

Lunch Status

```
#avg score by lunch status
avg_score_lunch <- math %>%
  group_by(Lunch)%>%
  summarize(Avg_Score=mean(Score,na.rm=TRUE))
avg_score_lunch
```

```
## # A tibble: 2 x 2
##   Lunch Avg_Score
##   <chr>     <dbl>
## 1 Free       66.0
## 2 Paid       64.6
```

```
#Comparison of scores for each lunch status group.
qplot(data = math, x=Lunch, y=Score, fill = Lunch, geom = "boxplot")+
  labs(title="Student Scores by Lunch Status")+
  scale_fill_manual(values =c("tomato2","skyblue4"))
```




```r
#H0: Mean Scores are the same for all lunch status groups.
t.test(Score~Lunch,data=math)
```

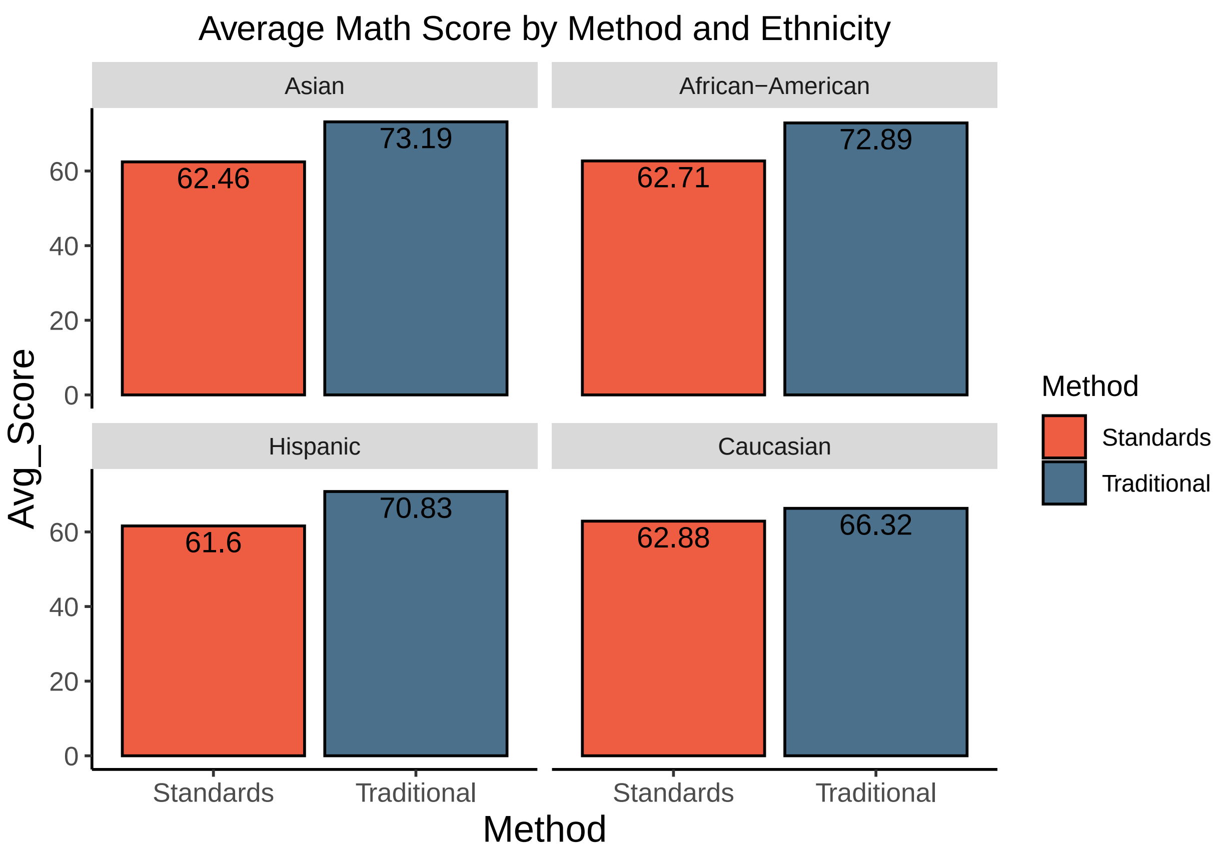
```
##
##  Welch Two Sample t-test
##
## data:  Score by Lunch
## t = 0.65536, df = 213.97, p-value = 0.5129
## alternative hypothesis: true difference in means between group Free and group Paid is not equal to 0
## 95 percent confidence interval:
##  -2.925998  5.840833
## sample estimates:
## mean in group Free mean in group Paid
##           66.02885           64.57143
```

Compare Student Scores Based on Method and Demographic Groups

When comparing student scores from different demographic groups, the traditional method continues to show a statistically significant higher score. Any observable difference between student groups accounting for demographic criteria is not statistically significant.

Method and Ethnicity

```
#plot
avg_method_ethnicity_viz <- ggplot(data=avg_score_method_ethnicity,aes(x=Method,y=Avg_Score,fill=Method)
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
  facet_wrap(~Ethnicity)+
  scale_fill_manual(values=c("tomato2","skyblue4"))+
  labs(title="Average Math Score by Method and Ethnicity")
avg_method_ethnicity_viz
```

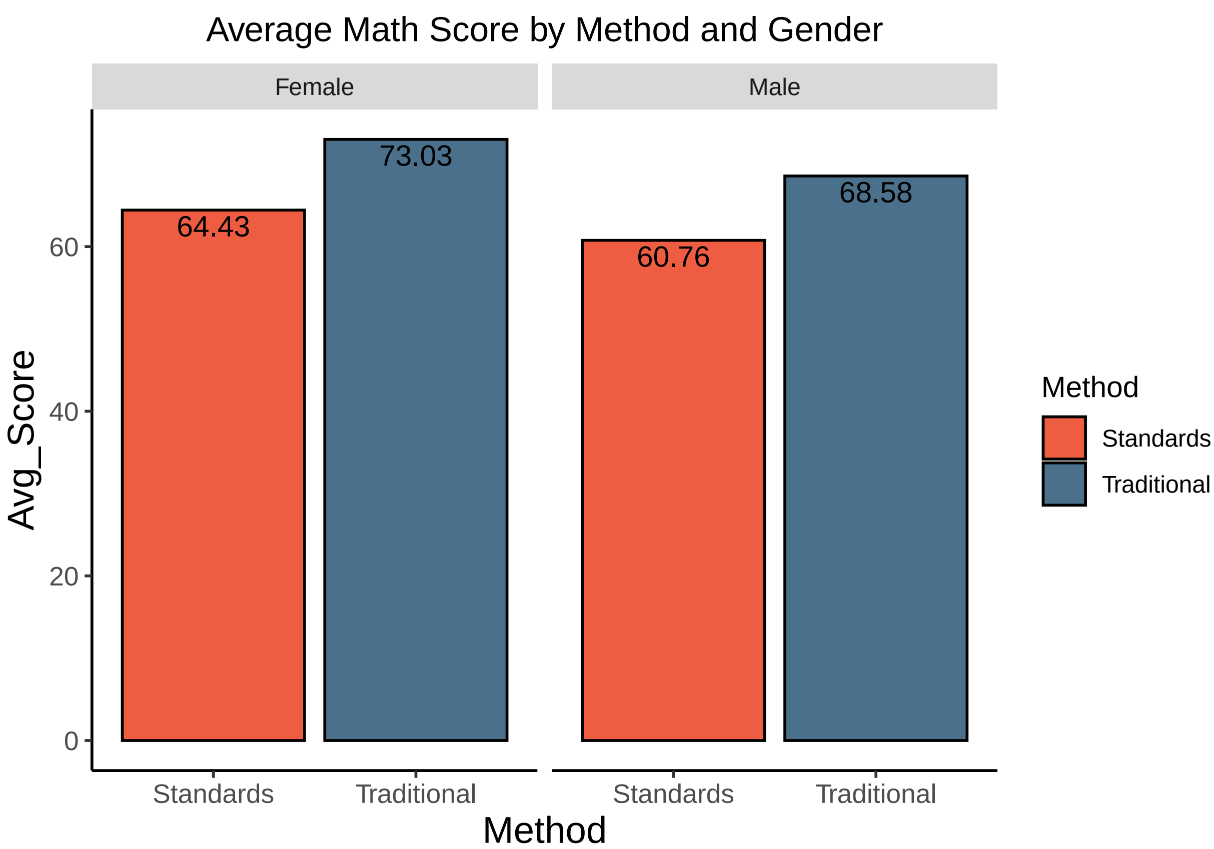


```r
anova4 <-aov(formula=Score~Method*Ethnicity, data = math)
summary(anova4)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)    
## Method             1   3433    3433   13.34 0.000329 ***
## Ethnicity          3    208      69    0.27 0.847243    
## Method:Ethnicity   3    379     126    0.49 0.689337    
## Residuals        208  53535     257                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Method and Gender

```r
#plot
avg_method_gender_viz <- ggplot(data=avg_score_method_gender,aes(x=Method,y=Avg_Score,fill=Method))+
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
  facet_wrap(~Gender)+
  scale_fill_manual(values=c("tomato2","skyblue4"))+
  labs(title="Average Math Score by Method and Gender")
avg_method_gender_viz
```



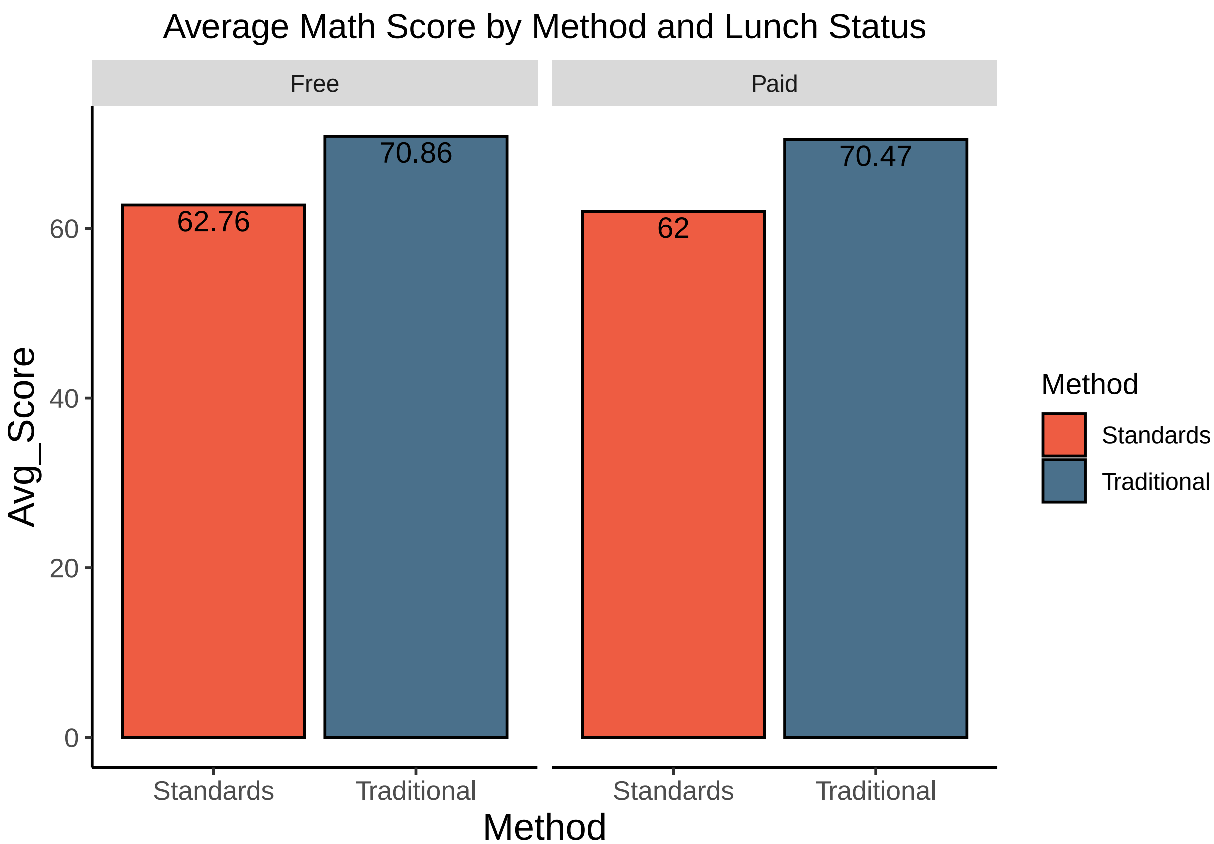

```
anova4 <-aov(formula=Score~Method*Gender, data = math)
summary(anova4)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Method          1   3433    3433  13.660 0.000279 ***
## Gender          1    830     830   3.303 0.070559 .
## Method:Gender   1      7       7   0.030 0.863366
## Residuals     212  53284     251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Method and Lunch Status

```
#plot
avg_method_lunch_viz <- ggplot(data=avg_score_method_lunch,aes(x=Method,y=Avg_Score,fill=Method))+
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
```

```
  facet_wrap(~Lunch)+
  scale_fill_manual(values=c("tomato2","skyblue4"))+
  labs(title="Average Math Score by Method and Lunch Status")
avg_method_lunch_viz
```




```
anova5 <-aov(formula=Score~Method*Lunch, data = math)
summary(anova5)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Method          1   3433    3433  13.454 0.000309 ***
## Lunch           1     21      21   0.082 0.774606
## Method:Lunch    1      2       2   0.007 0.935415
## Residuals     212  54099     255
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare Teachers

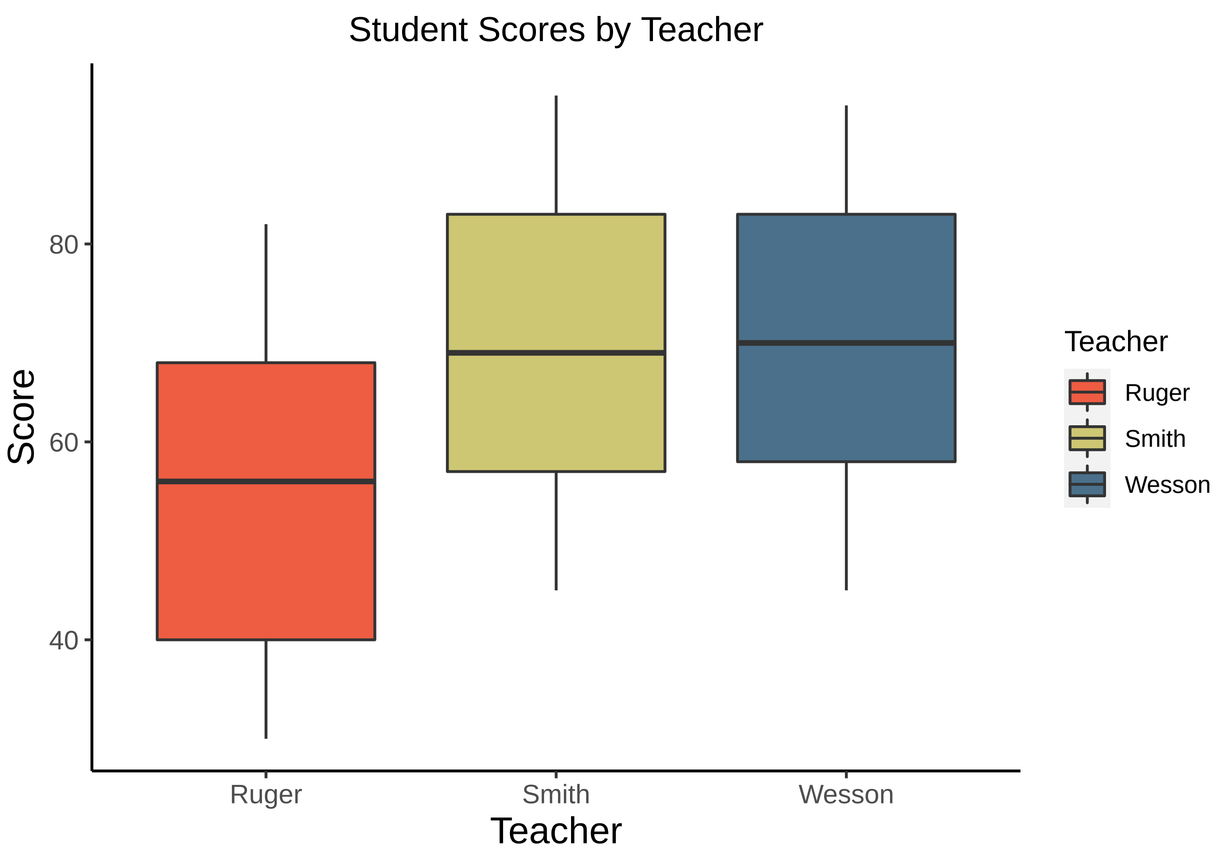
Compare Student Scores Based on Math Teacher.

Ms. Wesson and Ms. Smith have students with average scores of 70.7 and 69.6, respectively. The difference for these groups is not statistically significant. The students taught by Ms. Ruger have an average score of 55.26, which is lower than the other two classes by a statistically significant margin (p=3.59e-10).

```
#avg score by teacher
avg_teacher_score <- math %>%
    group_by(Teacher) %>%
    summarize(Avg_Score=mean(Score,na.rm=TRUE))
avg_teacher_score
```

```
## # A tibble: 3 x 2
##   Teacher Avg_Score
##   <fct>       <dbl>
## 1 Ruger        55.3
## 2 Smith        69.6
## 3 Wesson       70.7
```

```
qplot(data = math, x=Teacher, y=Score, fill = Teacher, geom = "boxplot")+
  labs(title="Student Scores by Teacher")+
  scale_fill_manual(values=c("tomato2","khaki3","skyblue4"))
```



```r
anova6 <- aov(formula = Score ~ Teacher, data = math)
summary(anova6)
```

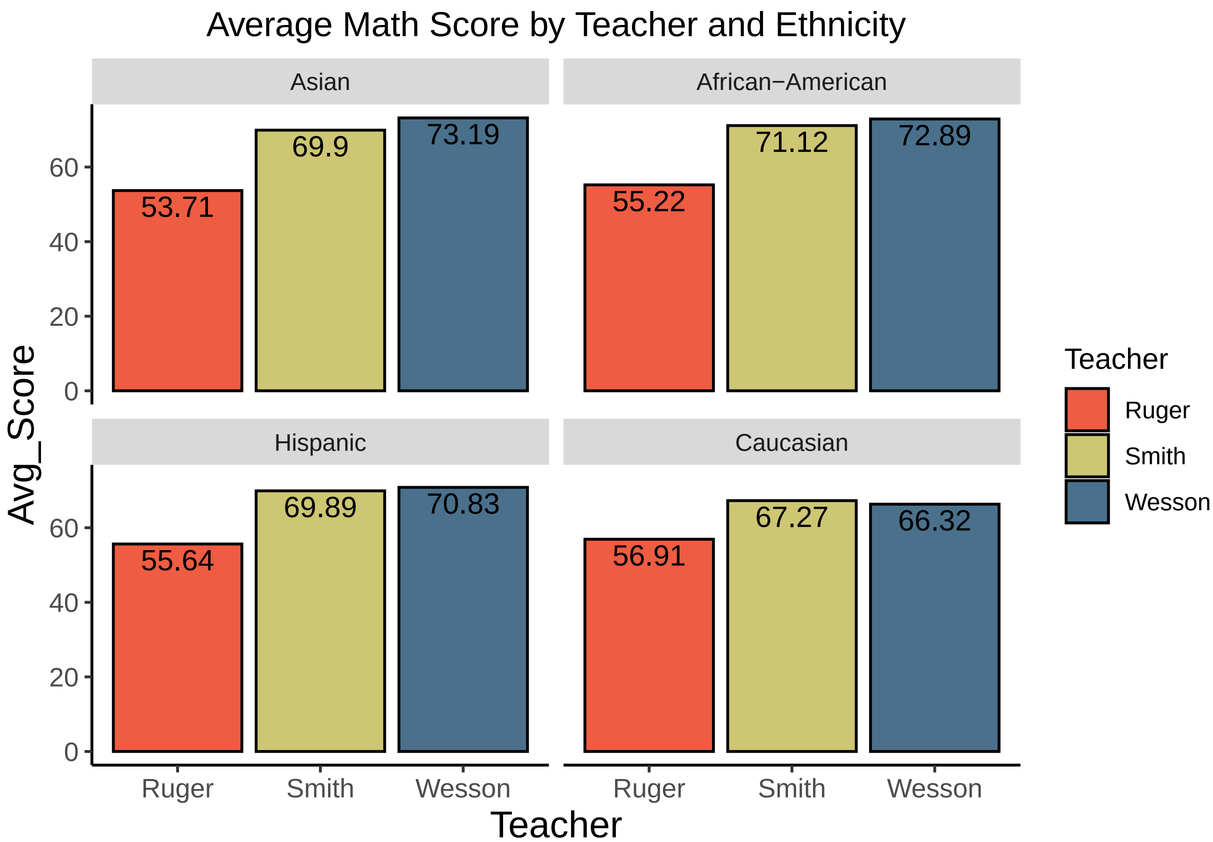
```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Teacher       2  10630    5315   24.13 3.59e-10 ***
## Residuals   213  46925     220                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare Student Scores Based on Math Teacher and Demographic Groups.

Next, we look at average scores for each teacher, accounting for demographic criteria. There is no significant difference in overall performance between demographic groups for any teacher. Ms. Ruger's class performs statistically lower in all comparisons.

Teacher and Ethnicity

```r
#plot
avg_teacher_ethnic_viz <- ggplot(data=avg_teacher_score_ethnicity,aes(x=Teacher,y=Avg_Score,fill=Teache
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
  facet_wrap(~Ethnicity)+
  scale_fill_manual(values=c("tomato2","khaki3","skyblue4"))+
  labs(title="Average Math Score by Teacher and Ethnicity")
avg_teacher_ethnic_viz
```



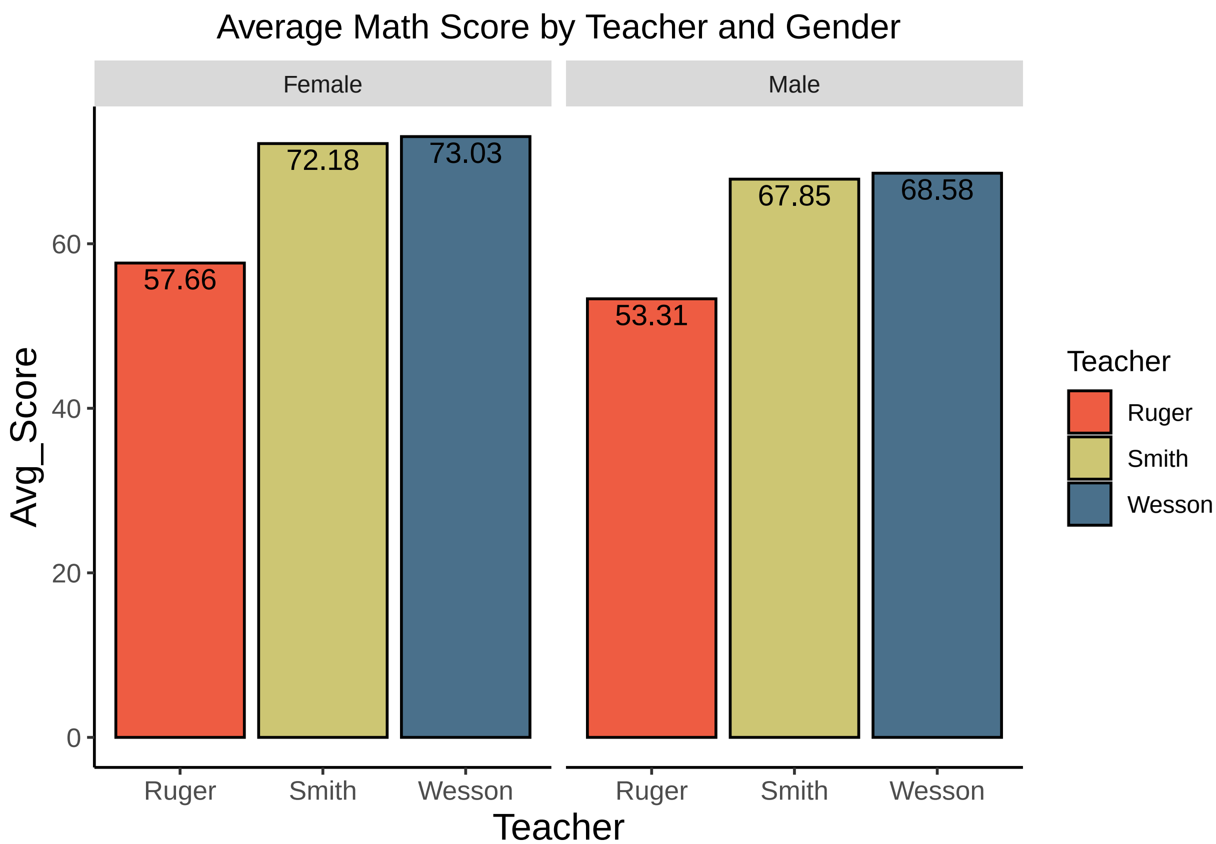

```r
anova7 <-aov(formula=Score~Teacher*Ethnicity, data = math)
summary(anova7)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## Teacher             2  10630    5315  23.481 6.64e-10 ***
## Ethnicity           3    313     104   0.462    0.709
## Teacher:Ethnicity   6    434      72   0.320    0.926
## Residuals         204  46177     226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teacher and Gender.

```r
#plot
avg_teacher_gender_viz <- ggplot(data=avg_teacher_score_gender,aes(x=Teacher,y=Avg_Score,fill=Teacher))+
```

```
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
  facet_wrap(~Gender)+
  scale_fill_manual(values=c("tomato2","khaki3","skyblue4"))+
  labs(title="Average Math Score by Teacher and Gender")
avg_teacher_gender_viz
```

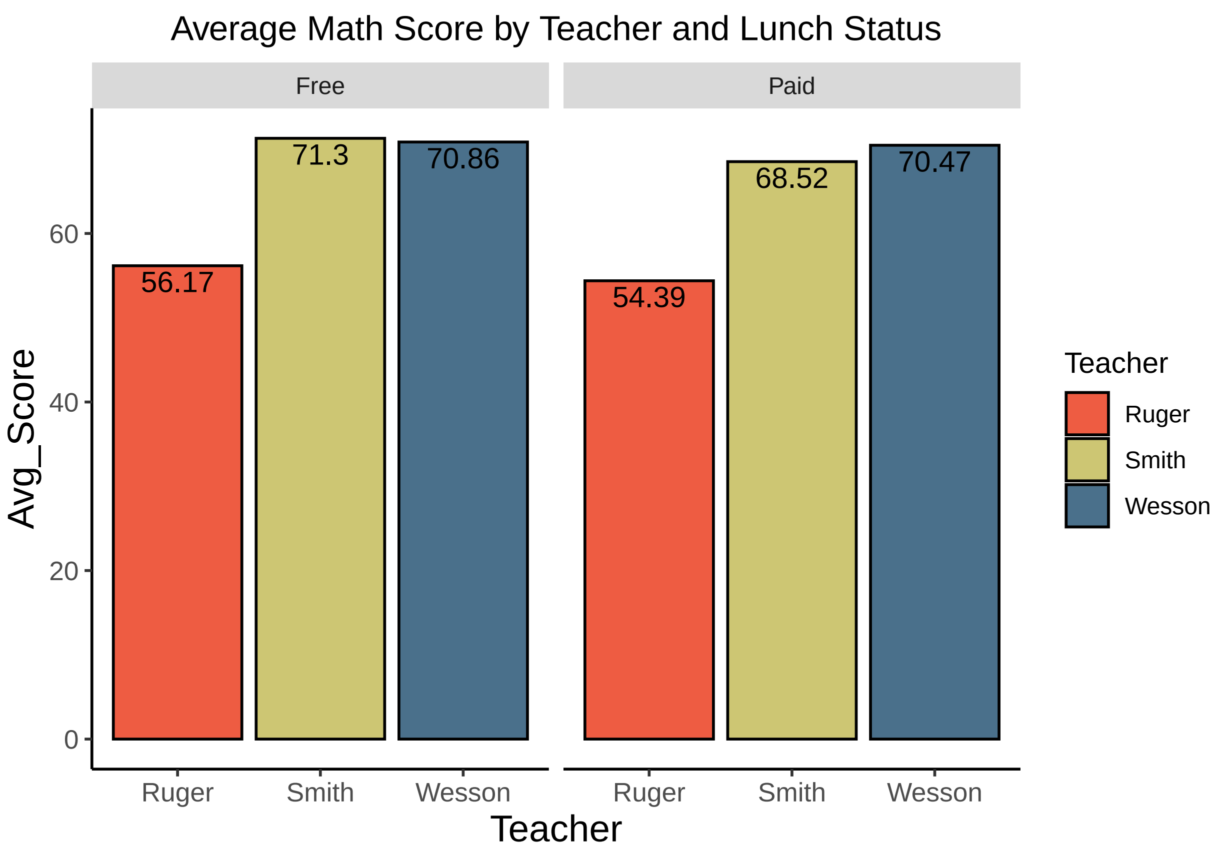



```
anova11 <-aov(formula=Score~Teacher*Gender, data = math)
summary(anova11)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## Teacher          2  10630    5315  24.314 3.18e-10 ***
## Gender           1   1019    1019   4.662    0.032 *
## Teacher:Gender   2      0       0   0.000    1.000
## Residuals      210  45906     219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teacher and Lunch Status.

```
#plot
avg_teacher_lunch_viz <- ggplot(data=avg_teacher_score_lunch,aes(x=Teacher,y=Avg_Score,fill=Teacher))+
  geom_bar(stat="identity")+
  geom_col(color="black")+
  geom_text(aes(label=round(Avg_Score,2)), vjust=1.25)+
  facet_wrap(~Lunch)+
  scale_fill_manual(values=c("tomato2","khaki3","skyblue4"))+
  labs(title="Average Math Score by Teacher and Lunch Status")
avg_teacher_lunch_viz
```




```
anova8 <-aov(formula=Score~Teacher*Lunch, data = math)
summary(anova8)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## Teacher        2  10630    5315  23.881 4.52e-10 ***
## Lunch          1    135     135   0.605    0.438
## Teacher:Lunch  2     51      25   0.114    0.892
```

```
## Residuals     210  46739     223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

When comparing traditional and standards-based methods among all students, the traditional method seems to result in higher scores. The difference is statistically significant, even when taking student demographic categories into account. Before we can conclude that method is the factor detrmining student scores, we must examine other possibilities.

The difference between the scores of students taught by Ms. Wesson (traditional method) and Ms. Smith (standards method) is not statistically significant. However, students taught by Ms. Ruger (standards method) performed significantly lower than Ms. Smith's students, even though these two teachers were using the same method. This remained true even when accounting for student demographic data. The data does not show any benefit for students being assigned to a specific teacher according to their demographic criteria that would outweigh the ethical issues that such groupings could create.

When ranking the performance of students in each class, the standards-based method ranks first and last, while the traditional method is a close second place. Considering the discrepancy between the two standards-based classrooms, we cannot definitively conclude that the teaching method is the determining factor affecting student performance. We can conclude, however, that Ms. Ruger's students did not perform as well as those taught by Ms. Smith or Ms. Wesson.

The final task was to evaluate the suggestion to group students by ability within the classroom. While the data set provided does not address this question, a careful read of the article cited by the teacher reveals that the data does not support the teacher's suggestion. The study observed the impact of grouping whole classes by ability, but did not address student grouping within a classroom. The paper showed no significant difference in teacher behavior or student performance between classes that were ability-tracked. However, there were negative differences in teacher attitude towards students in low-ability classrooms.

To find answers, it is necessary to look at other data. In his paper, "Ability Grouping in Mathematics Classrooms: A Bourdieuian Analysis," Robyn Zevenbergen found that ability grouping within the classroom can have a negative impact on how students perceive themselves and the subject of mathematics in general.

Further, "Within-Class Grouping: A Meta-Analysis" (Yiping Lou, et.al) found that, while small-group instruction within a classroom is preferable to whole-class instruction, the effect was largest if the teacher received training to adapt instructional delivery to each group. When comparing homogeneous or heterogeneous ability grouping within the classroom, the analysis found homogeneous groups benefit students with medium ability, but not low- or high-ability students. Additionally, a significant benefit of ability grouping was found in reading, but not in mathematics.

In summary, "Larger effects occurred when the group formation was based on mixed sources and involved more considerations than ability alone." These findings do support the use of small-group instruction with teacher training in adaptive methods for each group. They do not, however, support using ability as the exclusive criteria for forming small groups.

Recommendations

1. Allow teachers to continue using their preferred method. Collect more data that includes pre- and post-instruction scores to demonstrate growth, then reevaluate this issue based on future data.
2. Do not group classes according to ability or demographic criteria.
3. Encourage small-group instruction, and provide training to teachers on how to adapt instruction for each group. Small group criteria can include, but should not be limited to, mathematical ability.
4. The school administration should consider evaluating and offering professional development support for Ms. Ruger.