Homework 3

Cory Costello

October 17, 2017

I'll start by reading in the data, loading the tidyverse, and setting the global theme

```
library(rio)
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

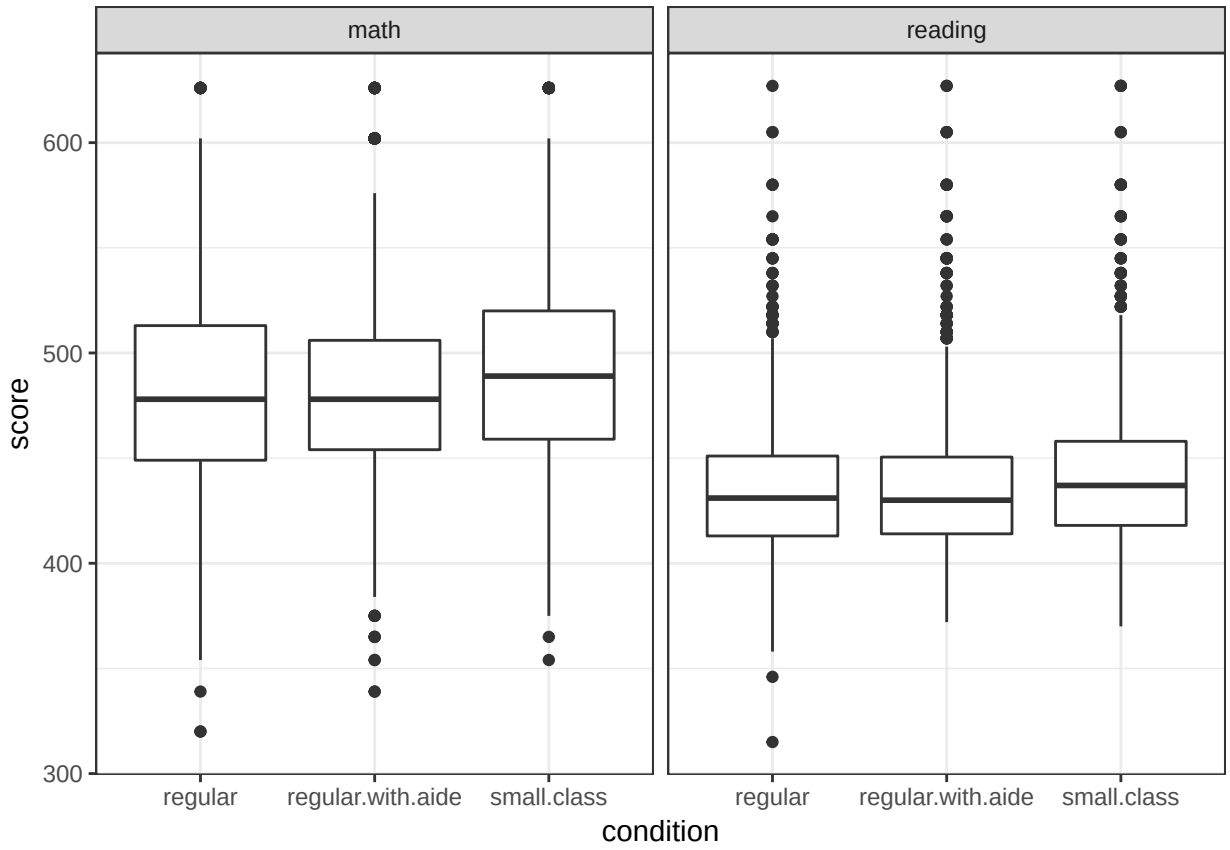
```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
star_data<- import("STAR_tidy2.csv")

theme_set(theme_bw())
```

Plot 1 - Box-plot by condition and content area

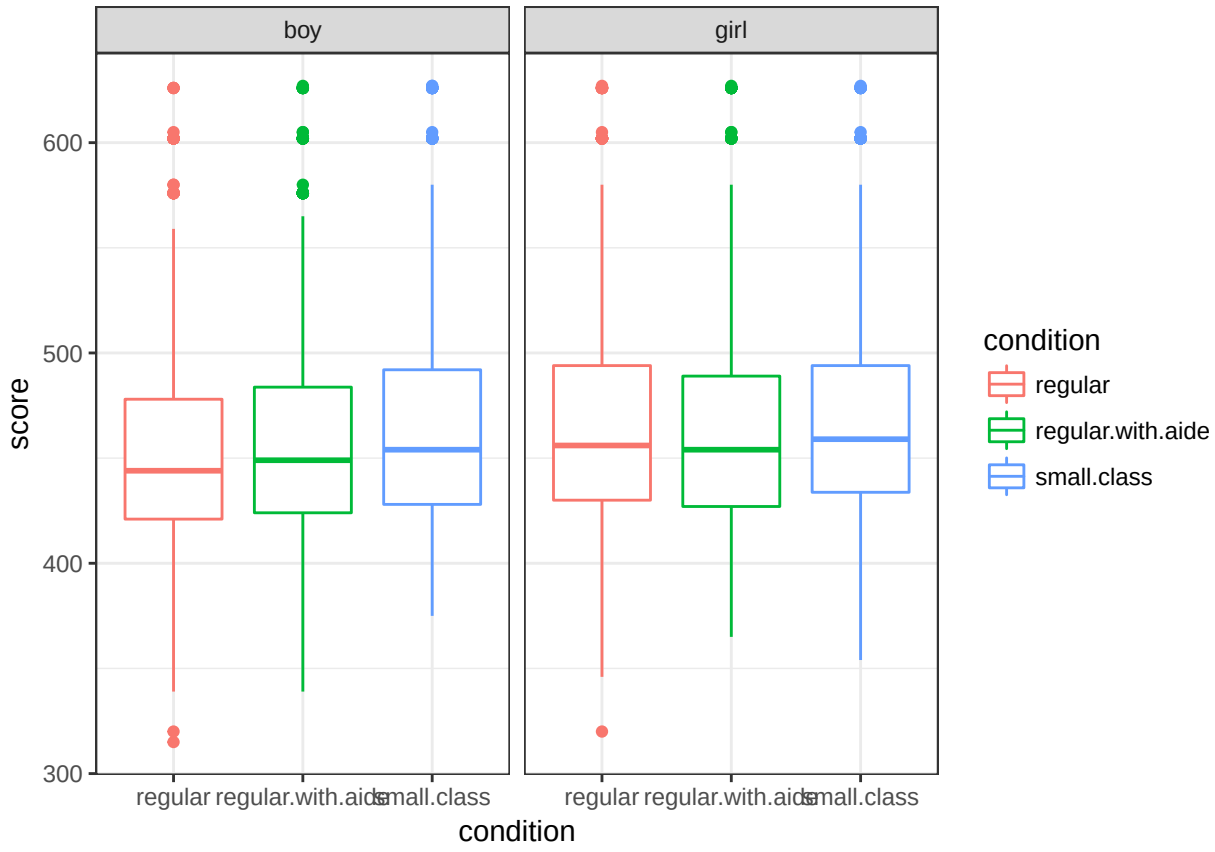
```
ggplot(star_data, aes(x = condition, y = score)) +
  geom_boxplot() +
  facet_wrap(~content_area)
```



The above plot demonstrates a number of things. First, median scores tend to be higher for math than reading (though it's not clear if that is a significant difference). Second, there appears to be very little effect of condition, at least in terms of the order-statistics (quartiles, medians, etc.). Third, it appers there are some outliers in both content areas, but there are far more outliers for reading than math. Additionally, there are both positive and negative outliers in math, whereas the bulk of the outliers are positive outliers for reading (especially in the conditions regular with aide and small class).

Plot 2 box-plots by condition and gender

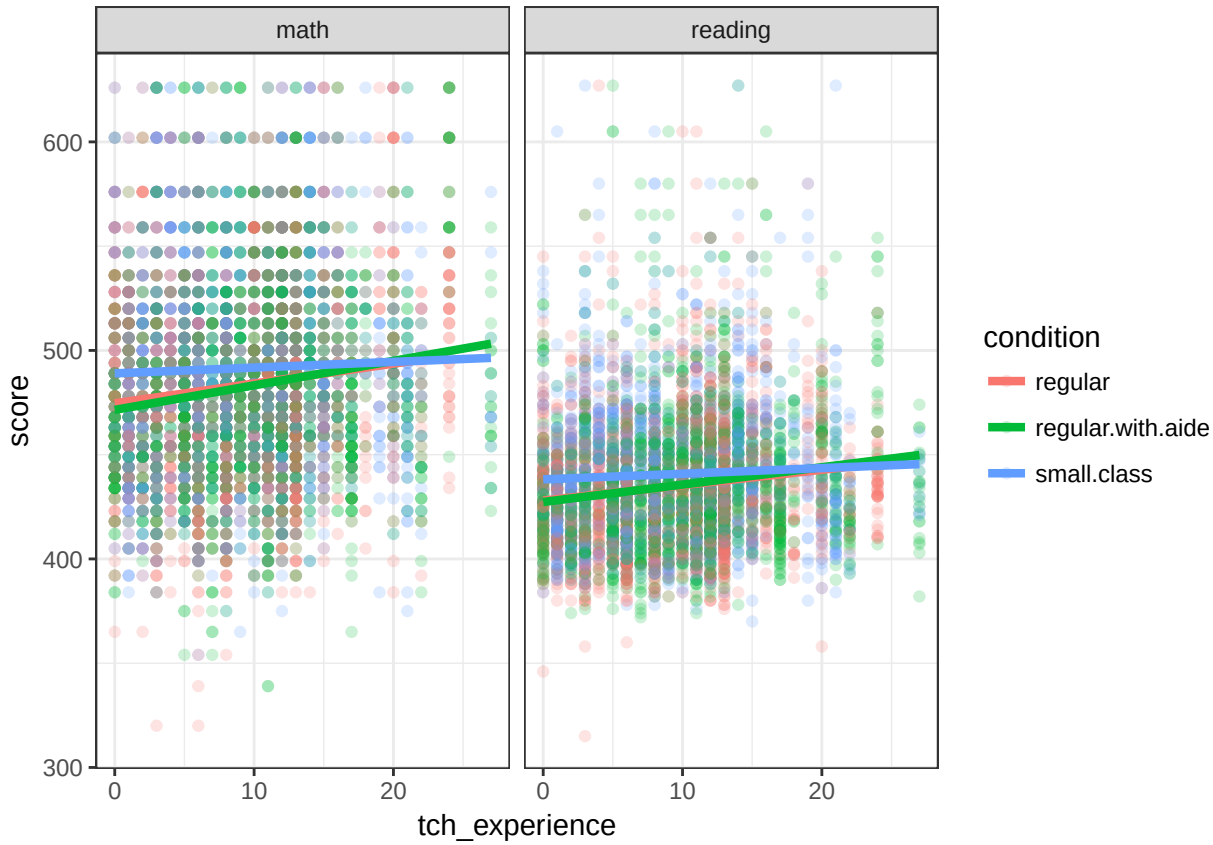
```
ggplot(star_data, aes(x = condition, y = score, color = condition)) +
  geom_boxplot() +
  facet_wrap(~sex)
```



Looks like male and female students get similar scores (collapsed across content area). Also looks like there's no condition by gender effect (at least for the order-stats). It's somewhat hard to tell, but it looks like there may be more outliers among male than female students (I think turning alpha down could help make this clearer).

Plot 3 - Teaching experience and total score by condition

```
ggplot(star_data, aes(x = tch_experience, y = score, color = condition)) +
  geom_point (alpha = .2) +
  geom_smooth (method = "lm", size = 1.5, alpha = 0) +
  facet_wrap (~content_area)
```

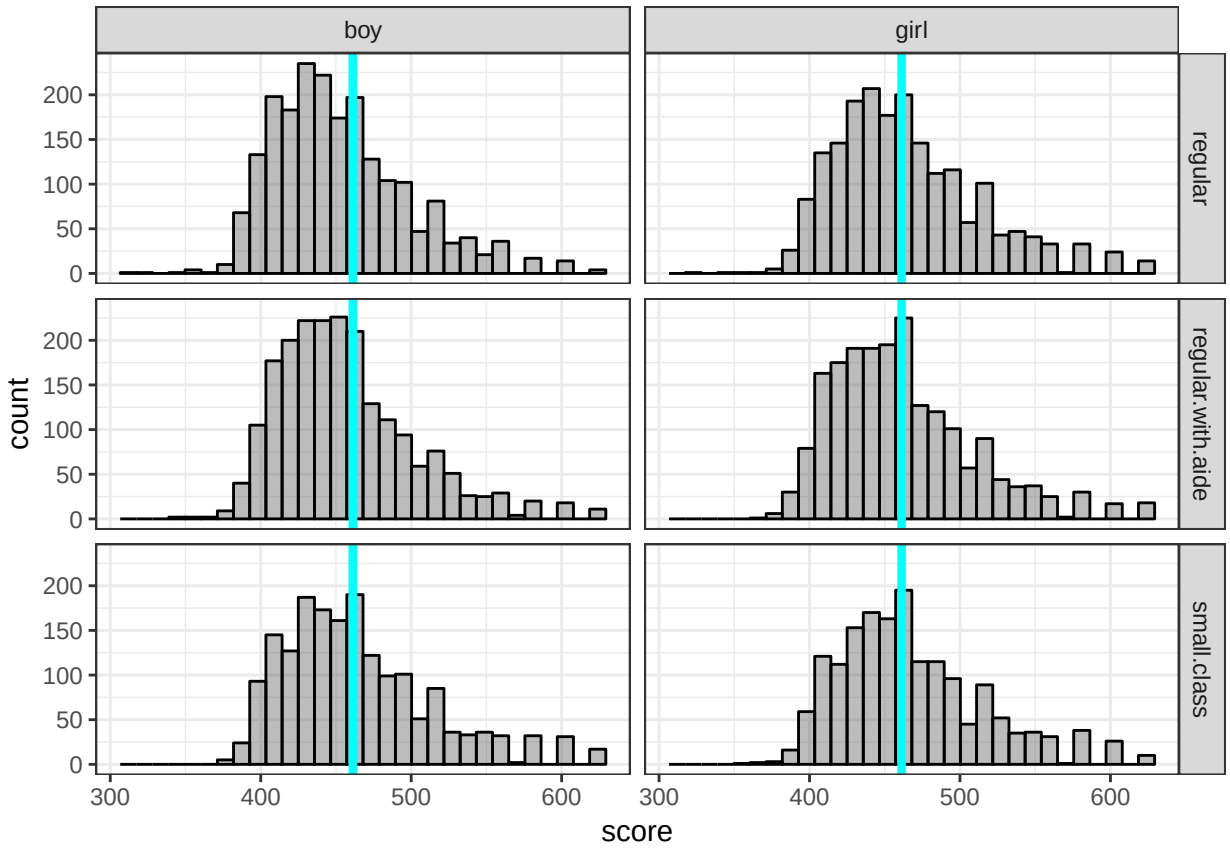


It looks like there may be a differential effect of teaching experience on test scores based on condition. Specifically, it appears that teaching experience has a larger effect for teachers in the regular and regular with aide condition relative to the small class condition. Put differently, it looks like teaching experience matters less in the small class condition than it does in the other two conditions. It looks like this differential effect is stronger in math than reading (probably not significantly so).

Plot 4 - Histograms by condition and sex

```
ggplot(star_data, aes(x = score)) +
  geom_histogram(alpha = .4, color="black")+
  # I'm sure there is a more elegant solution for
  # putting a line at the mean,
  # but this is what occurred to me first and I stuck with it.
  geom_vline(xintercept = mean(star_data$score), color = "cyan", size = 1.5)+
  facet_grid (condition~sex)
```

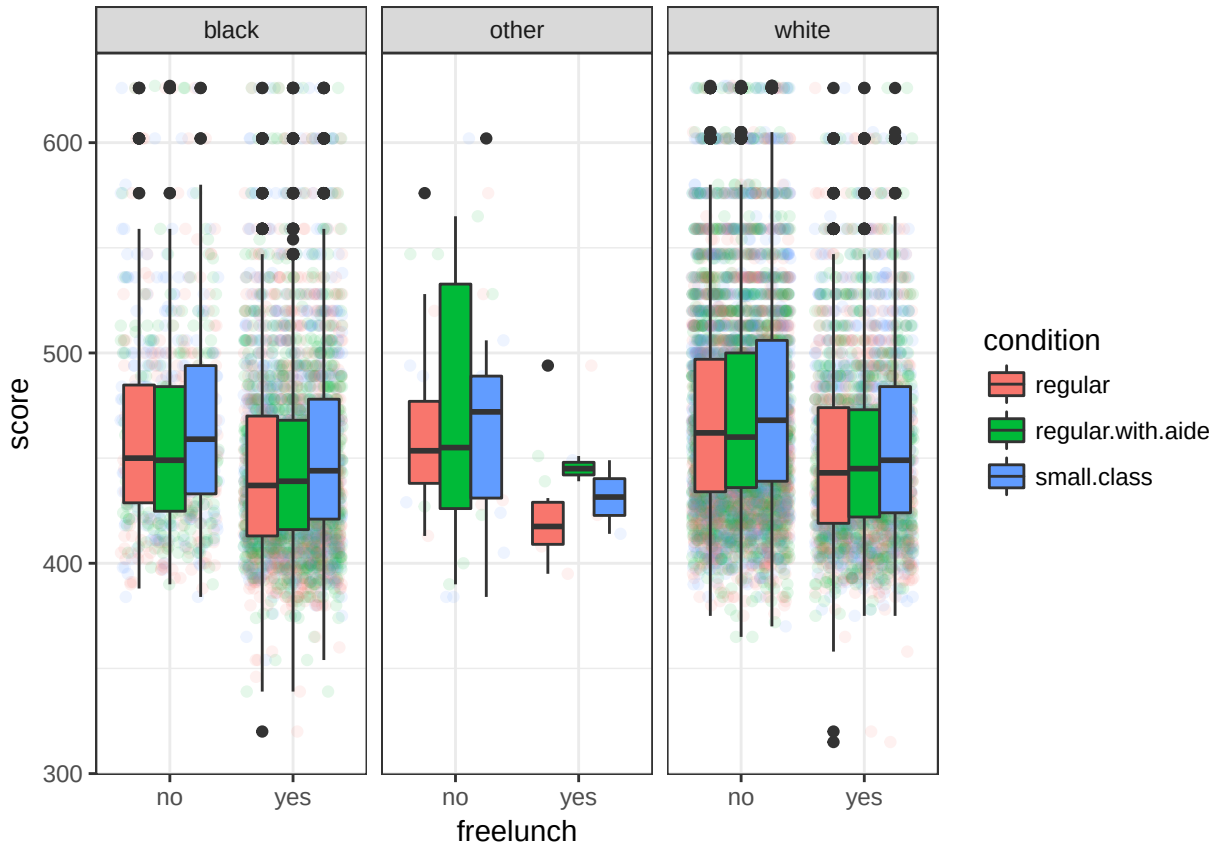
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Damn if those aren't similar distributions for male and female students. The vertical line at the mean seems to further suggest no main effect of condition.

Plot 5 - condition effect by free lunch status and race

```
ggplot(star_data, aes(x = freelunch, y = score, fill = condition)) +
  geom_jitter(aes(color = condition), height = 0, alpha = .1)+
  geom_boxplot()+
  facet_wrap (~race)
```



It looks like there isn't a differential effect of condition by race and free lunch status. You can see this in the above graph in that the order of conditions appears to be relatively preserved across race and free lunch status combinations.