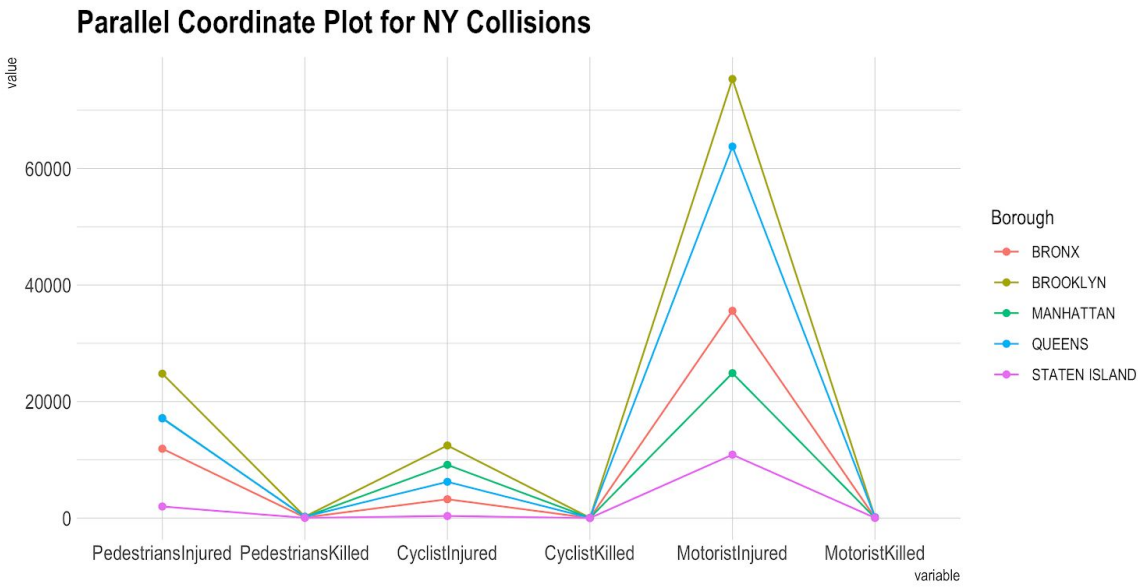
Assignment 4

Question 1 (20 points)

From the New York collision dataset create the following parallel coordinate plot



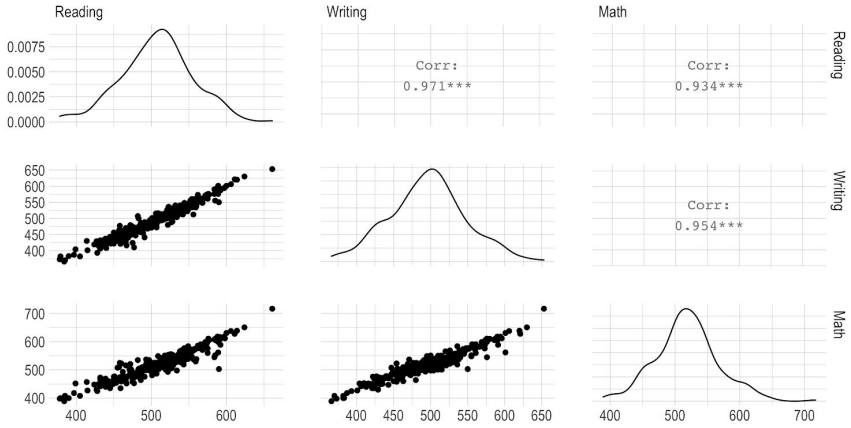
The font type, font case, color, and theme in your visualization can differ. Use the code below to generate the parallel coordinate plot.

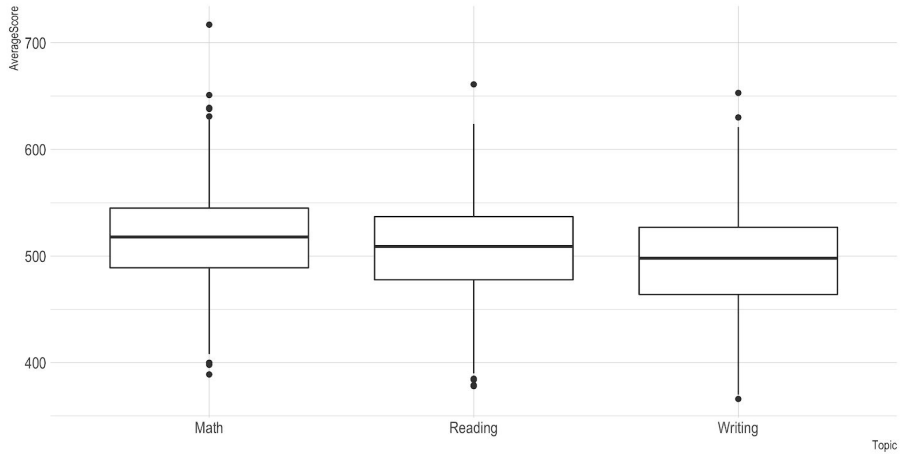
```
ggparcoord(df, columns =  , groupColumn = ,
           showPoints = TRUE,
           title = "Parallel Coordinate Plot for NY Collisions",
           scale="globalminmax")
```

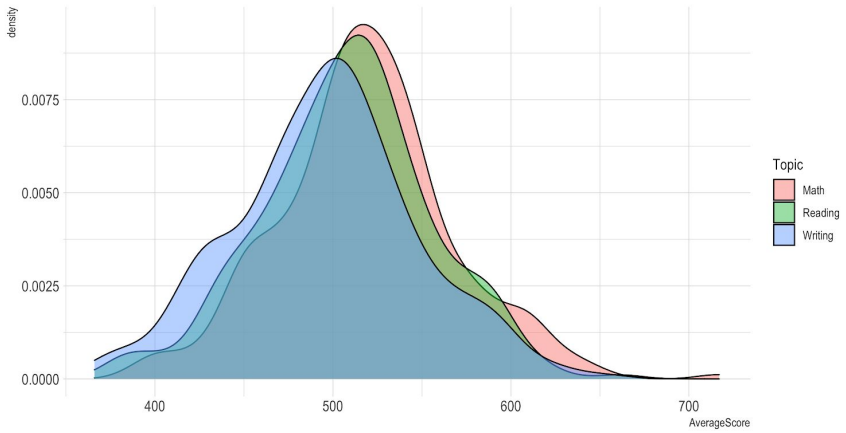
Critique the visualization and include your improved solution

Question 2 (60 points)

From the link (http://profiles.doe.mass.edu/statereport/sat.aspx) download the average SAT scores for the year 2013-14 and create the following plots







The font type, font case, color, and theme in your visualization can differ. Use the following codes to create the above three plots.

```
# For paired correlation
ggpairs(df)
# For boxplot
ggplot(df, aes(x=, y=)) + geom_boxplot()
# For density
ggplot(df, aes(x=, fill=)) + geom_density(alpha=0.5)
```

Question 3 (20 points)

Create a visualization that captures the relation between Avg. SAT scores (use data from question 2) and median household income in that school district. For the median household income of the school districts use http://www.usa.com/rank/massachusetts-state--median-household-income--school-district-rank.htm

Write your observations from the visualization

Submission Format

1. Submit your solution in .Rmd format
2. Some questions require you to write descriptive answers. Provide your answer to the question before the code chunk as follows

````
## The drawback of the parallel coordinate plot generated below....
```{r}
ggparcoord(df, columns =  , groupColumn = ,
           showPoints = TRUE,
           title = "Parallel Coordinate Plot for NY Collisions",
           scale="globalminmax")
```
````