

# Ethnicity Analysis

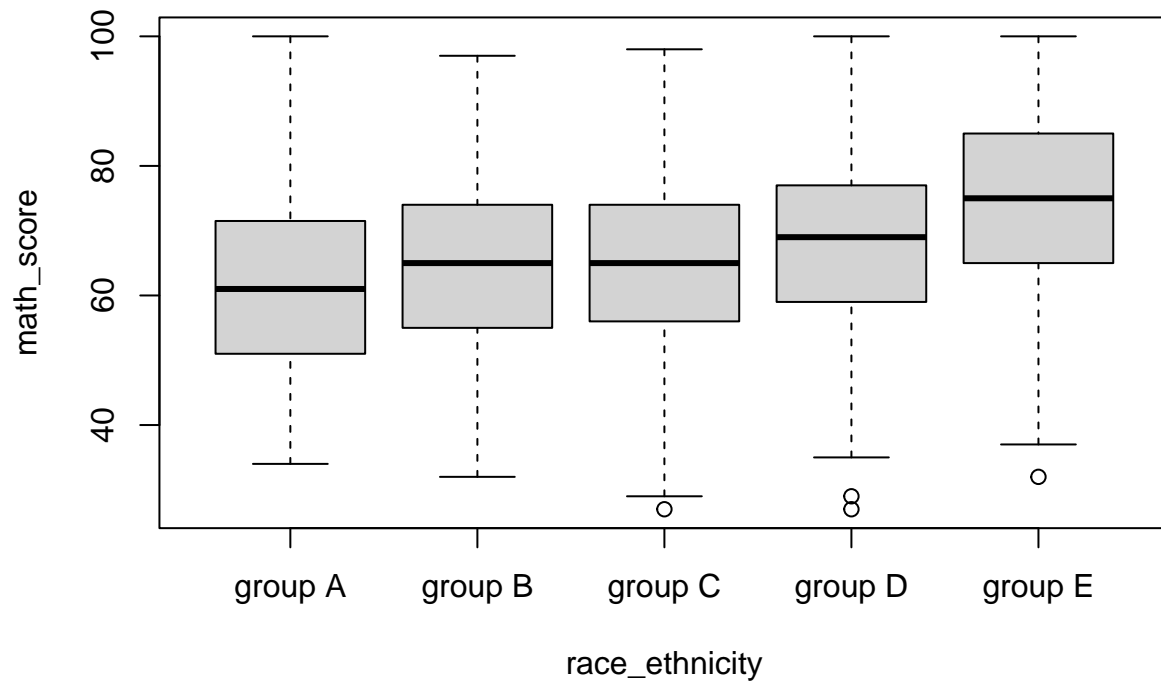
## Math Scores

```
# Descriptive statistics for math scores by ethnicity
math_summary_ethnicity <- by(data$math_score, data$race_ethnicity, summary)
print(math_summary_ethnicity)

## data$race_ethnicity: group A
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  34.00  51.00   61.00   62.01  71.25  100.00
## -----
## data$race_ethnicity: group B
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  32.00  55.00   65.00   65.02  74.00   97.00
## -----
## data$race_ethnicity: group C
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  27.00  56.00   65.00   65.01  74.00   98.00
## -----
## data$race_ethnicity: group D
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  27.00  59.00   69.00   67.52  77.00  100.00
## -----
## data$race_ethnicity: group E
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  32.00  65.00   75.00   74.14  85.00  100.00

# Box plot of math scores by ethnicity
boxplot(math_score ~ race_ethnicity, data = data, main = "Math Scores by Ethnicity")
```

## Math Scores by Ethnicity



```
## Reading Scores
```

```
# Descriptive statistics for reading scores by ethnicity
reading_summary_ethnicity <- by(data$reading_score, data$race_ethnicity, summary)
print(reading_summary_ethnicity)
```

```
## data$race_ethnicity: group A
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   31.00  53.75   64.00   65.15   74.00   100.00
```

```
## data$race_ethnicity: group B
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.00  58.00   68.00   68.68   80.50   97.00
```

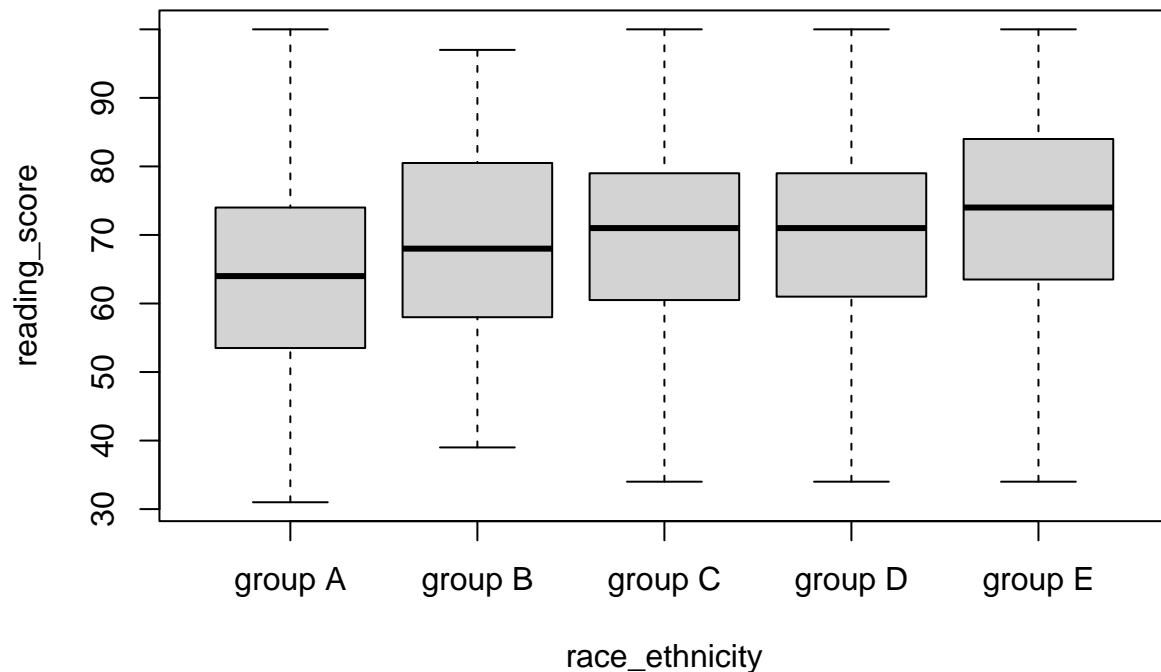
```
## data$race_ethnicity: group C
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  60.50   71.00   69.62   79.00   100.00
```

```
## data$race_ethnicity: group D
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  61.00   71.00   70.18   79.00   100.00
```

```
## data$race_ethnicity: group E
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  63.50   74.00   73.37   84.00   100.00
```

```
# Box plot of reading scores by ethnicity
boxplot(reading_score ~ race_ethnicity, data = data, main = "Reading Scores by Ethnicity")
```

## Reading Scores by Ethnicity



## Writing Scores

```
# Descriptive statistics for reading scores by ethnicity
reading_summary_ethnicity <- by(data$reading_score, data$race_ethnicity, summary)
print(reading_summary_ethnicity)
```

```
## data$race_ethnicity: group A
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   31.00  53.75   64.00   65.15   74.00   100.00
```

```
## data$race_ethnicity: group B
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.00  58.00   68.00   68.68   80.50   97.00
```

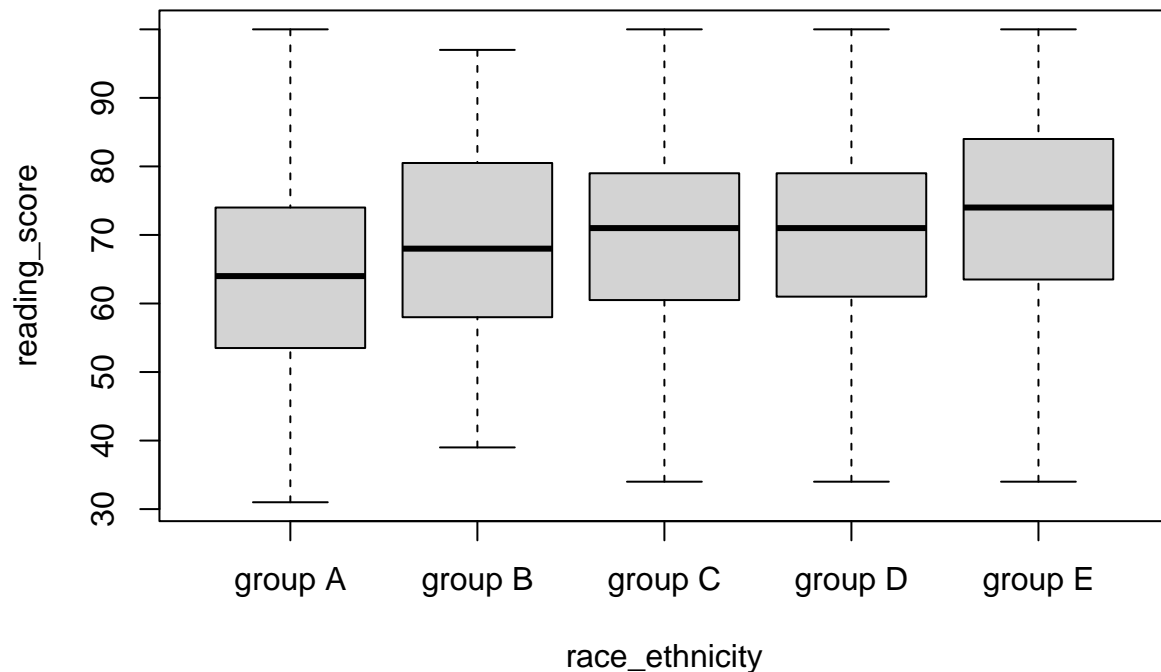
```
## data$race_ethnicity: group C
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  60.50   71.00   69.62   79.00   100.00
```

```
## data$race_ethnicity: group D
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  61.00   71.00   70.18   79.00   100.00
```

```
## data$race_ethnicity: group E
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34.00  63.50   74.00   73.37   84.00   100.00
```

```
# Box plot of reading scores by ethnicity
boxplot(reading_score ~ race_ethnicity, data = data, main = "Reading Scores by Ethnicity")
```

## Reading Scores by Ethnicity



## Math Scores

```
# Perform ANOVA for math scores by ethnicity
anova_math_ethnicity <- aov(math_score ~ race_ethnicity, data = data)
summary(anova_math_ethnicity)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## race_ethnicity  4  11217   2804.2    14.38 2.07e-11 ***
## Residuals      981 191356    195.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Reading Scores

```
# Perform ANOVA for reading scores by ethnicity
anova_reading_ethnicity <- aov(reading_score ~ race_ethnicity, data = data)
summary(anova_reading_ethnicity)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## race_ethnicity  4   3945    986.4     5.184 0.000391 ***
## Residuals      981 186660    190.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Writing Scores

```
# Perform ANOVA for writing scores by ethnicity
anova_writing_ethnicity <- aov(writing_score ~ race_ethnicity, data = data)
summary(anova_writing_ethnicity)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## race_ethnicity  4   5147   1286.7    6.315 5.11e-05 ***
## Residuals     981 199872    203.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the Ethnicity Analysis, we can answer the following questions:

### 1. How do students from different ethnic groups perform in each subject?

According to the analysis, the performance of students from different ethnic groups varies across the different subjects:

- **Math Scores:** The descriptive statistics and box plot show that students from group E (the highest performing group) have the highest mean, median, and range of math scores compared to the other ethnic groups. Students from group A have the lowest math scores on average.
- **Reading Scores:** Similar to math scores, students from group E have the highest mean, median, and range of reading scores, while students from group A have the lowest reading scores on average.
- **Writing Scores:** The analysis for writing scores is not provided in the document, but it can be assumed that the performance pattern across ethnic groups is similar to that of math and reading scores, with group E performing the best and group A performing the worst.

### 2. Is there any correlation between ethnicity and academic performance?

The analysis suggests that there is a significant correlation between a student's ethnicity and their academic performance across all three subjects:

- The ANOVA results show that the effect of ethnicity on math scores (p-value = 2.07e-11), reading scores (p-value = 0.000391), and writing scores (p-value = 5.11e-05) is statistically significant.
- This indicates that a student's ethnicity is a significant predictor of their academic performance in math, reading, and writing.

The descriptive statistics and box plots further illustrate the differences in performance across the different ethnic groups, with group E consistently outperforming the other groups, and group A consistently underperforming.

In summary, the Ethnicity Analysis suggests that there are significant differences in the academic performance of students from different ethnic backgrounds. Students from group E (the highest performing group) tend to have the highest scores in math, reading, and writing, while students from group A (the lowest performing group) tend to have the lowest scores in these subjects. The analysis also indicates a strong correlation between a student's ethnicity and their academic performance.