Students Performance Analysis

Alexandria

2024-07-14

Exam Performance Scores

Here is a report following students performance scores and variables that share a realtionship. This dataset has been sorted and cleaned for the analysis. The visualizations provides insights on the data distribution and reflect the findings.

Firs, upload the csv file StudentsPerformance.csv

Then install necessary packages

```
install.packages(c("dplyr", "ggplot2", "readr"))
```

Load the packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readr)
```

Now the dataset can be loaded

```
students_data <- read_csv("StudentsPerformance.csv")
```

```
## Rows: 1000 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): gender, race/ethnicity, parental level of education, lunch, test pr...
## dbl (3): math score, reading score, writing score
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Here is the cleaning data process

Check For Missing Values

```
sum(is.na(students_data))
```

```
## [1] 0
```

Standardize the column names

```
colnames(students_data) <- tolower(gsub(" ", "_", colnames(students_data)))
```

Analyze the data

Summary Statistics

```
summary(students_data)
```

```
##     gender          race/ethnicity     parental_level_of_education
##  Length:1000        Length:1000        Length:1000
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##     lunch           test_preparation_course   math_score     reading_score
##  Length:1000        Length:1000             Min.   :  0.00   Min.   : 17.00
##  Class :character   Class :character        1st Qu.: 57.00   1st Qu.: 59.00
##  Mode  :character   Mode  :character        Median : 66.00   Median : 70.00
##                                             Mean   : 66.09   Mean   : 69.17
##                                             3rd Qu.: 77.00   3rd Qu.: 79.00
##                                             Max.   :100.00   Max.   :100.00
##  writing_score
##  Min.   : 10.00
##  1st Qu.: 57.75
##  Median : 69.00
##  Mean   : 68.05
##  3rd Qu.: 79.00
##  Max.   :100.00
```

Gender Disrtibution

```
gender_dist <- students_data %>% count(gender)
print(gender_dist)
```

```
## # A tibble: 2 x 2
##   gender     n
##   <chr>  <int>
## 1 female   518
## 2 male     482
```

Average Score For Each Gender

```
avg_scores_by_gender <- students_data %>%
  group_by(gender) %>%
  summarise(
```

```
    avg_math = mean(math_score),
    avg_reading = mean(reading_score),
    avg_writing = mean(writing_score)
  )
print(avg_scores_by_gender)
```

```
## # A tibble: 2 x 4
##   gender avg_math avg_reading avg_writing
##   <chr>     <dbl>       <dbl>       <dbl>
## 1 female     63.6        72.6        72.5
## 2 male       68.7        65.5        63.3
```

Parental Education Level Analysis

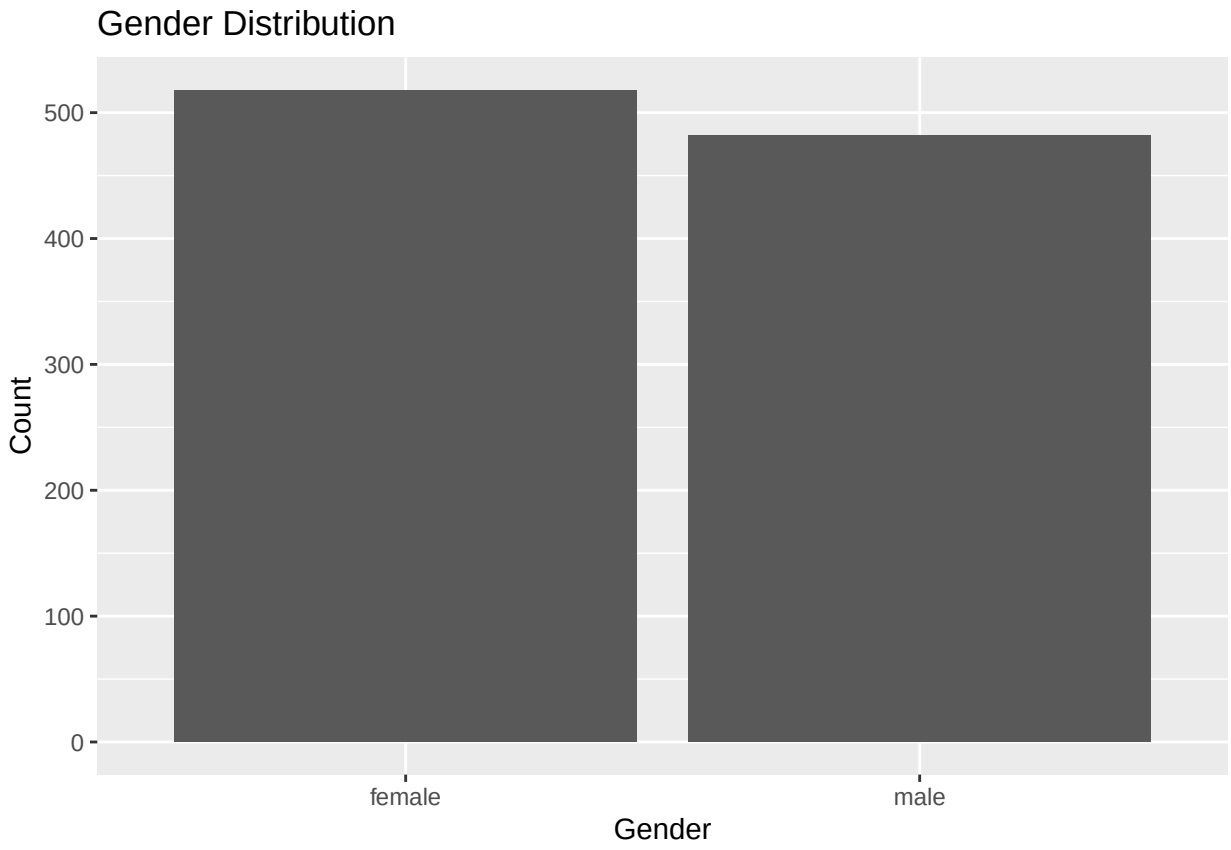
```
parental_education_analysis <- students_data %>%
  group_by(parental_level_of_education) %>%
  summarise(
    avg_math = mean(math_score),
    avg_reading = mean(reading_score),
    avg_writing = mean(writing_score)
  )
print(parental_education_analysis)
```

```
## # A tibble: 6 x 4
##   parental_level_of_education avg_math avg_reading avg_writing
##   <chr>                          <dbl>       <dbl>       <dbl>
## 1 associate's degree              67.9        70.9        69.9
## 2 bachelor's degree               69.4        73          73.4
## 3 high school                     62.1        64.7        62.4
## 4 master's degree                 69.7        75.4        75.7
## 5 some college                    67.1        69.5        68.8
## 6 some high school                63.5        66.9        64.9
```

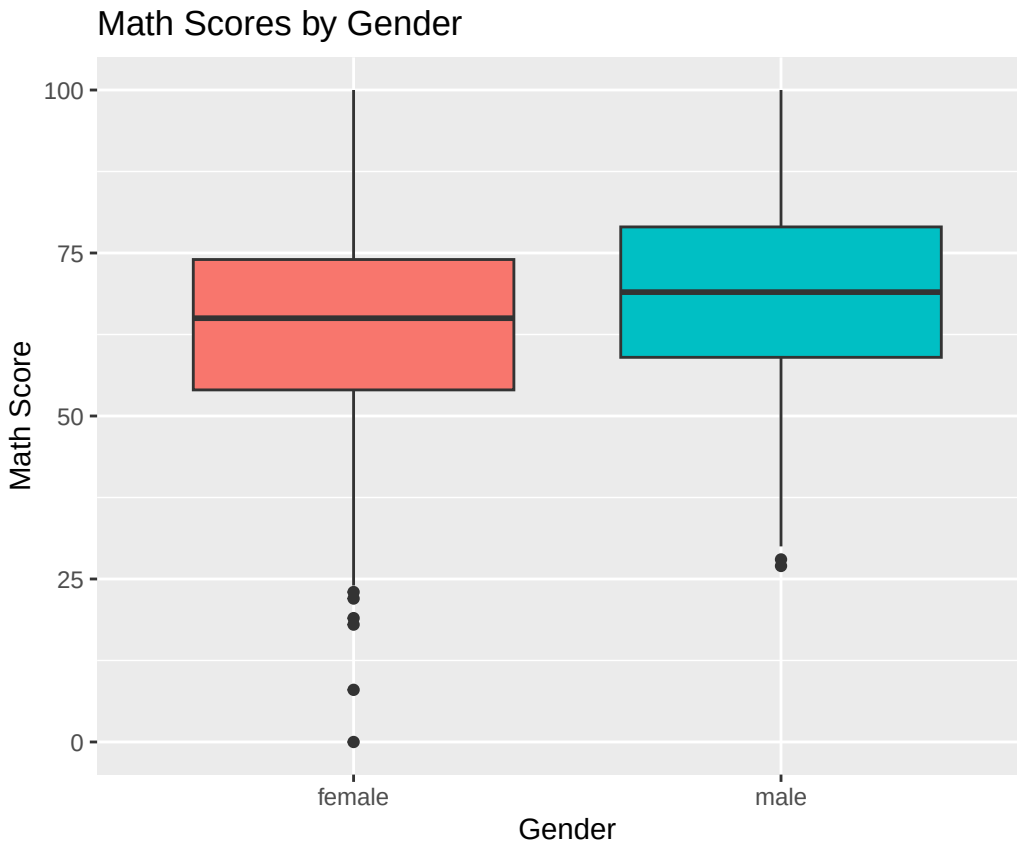
Data Visualization ### Bar Plots for gender disrtibution

```
library(ggplot2)

ggplot(gender_dist, aes(x = gender, y = n)) +
  geom_bar(stat = "identity") +
  labs(title = "Gender Distribution", x = "Gender", y = "Count")
```



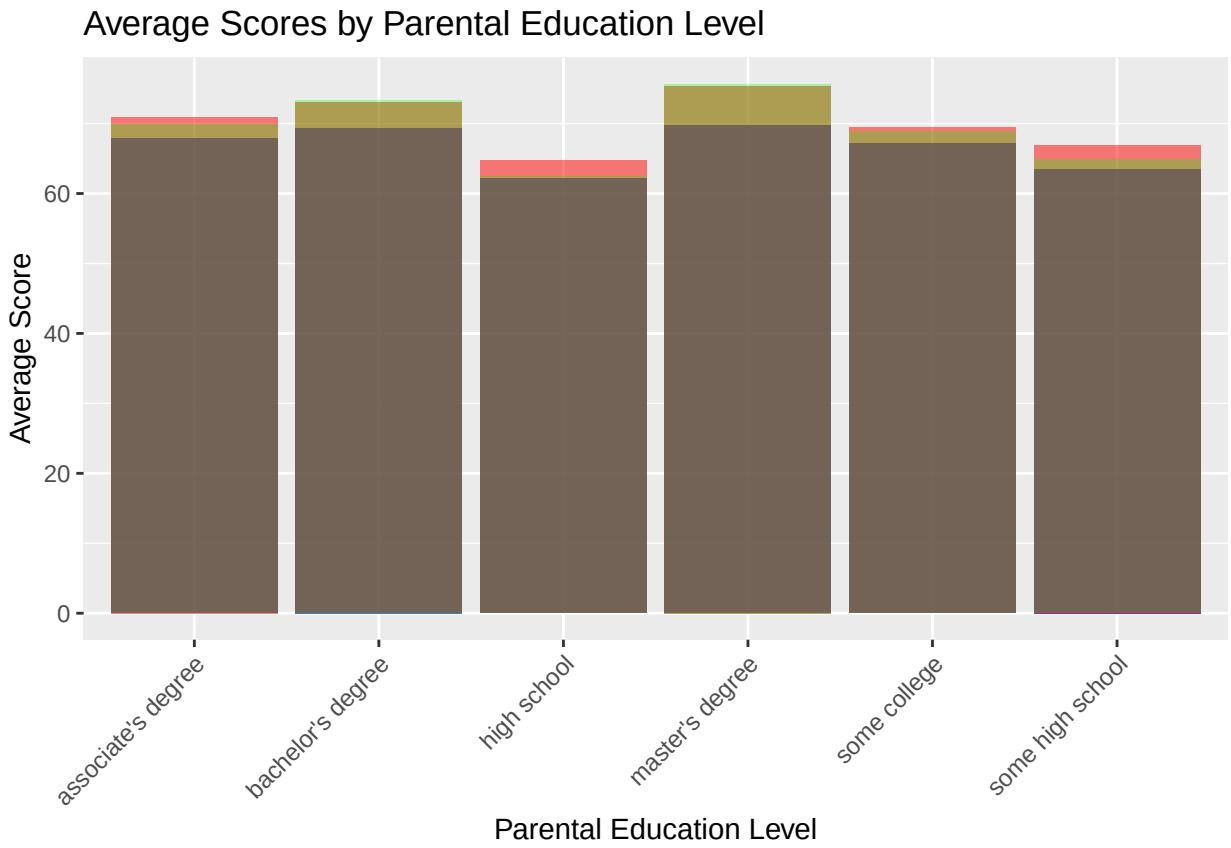

```
ggplot(students_data, aes(x = gender, y = math_score, fill = gender)) +
  geom_boxplot() +
  labs(title = "Math Scores by Gender", x = "Gender", y = "Math Score")
```



Box Plots for scores by gender

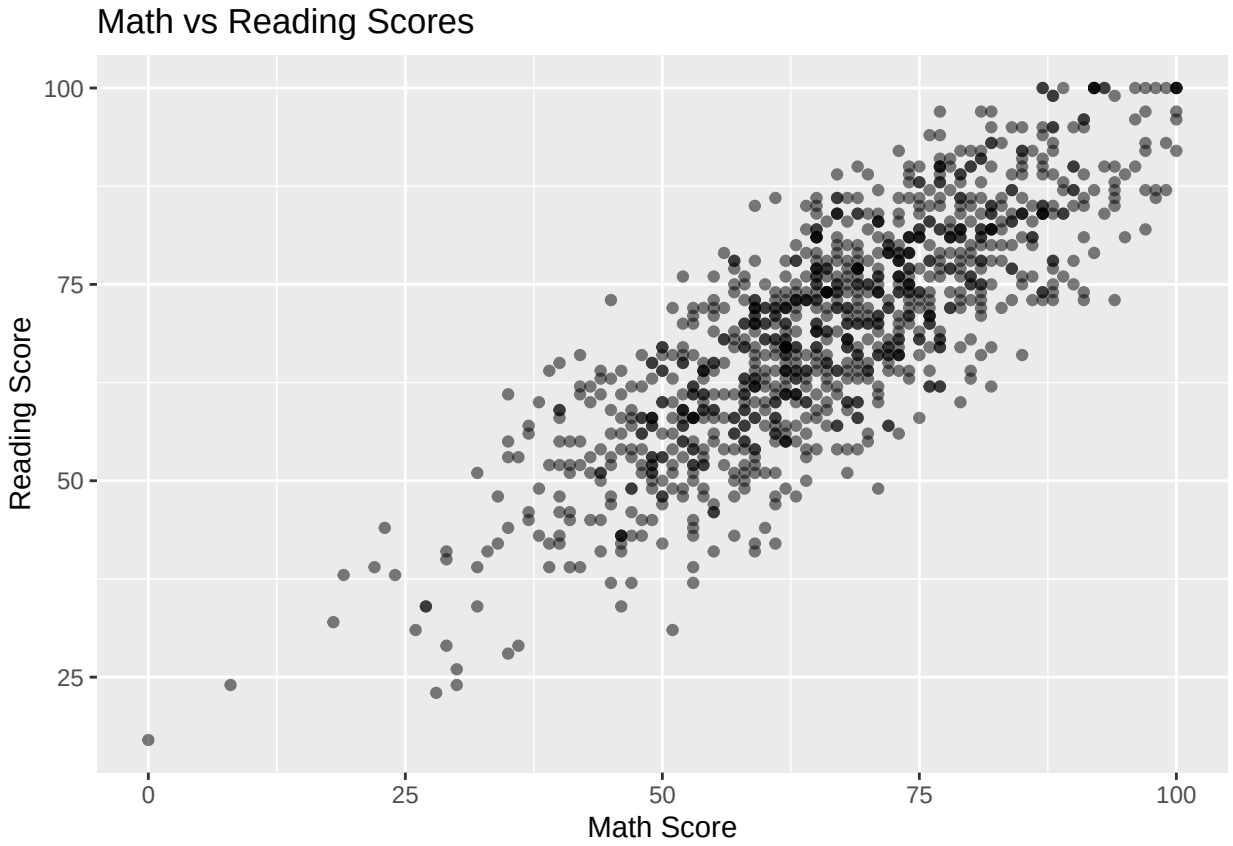
Bar Plot for Average Scores by Parental Education Level

```
ggplot(parental_education_analysis, aes(x = parental_level_of_education)) +
  geom_bar(aes(y = avg_math), stat = "identity", fill = "blue", alpha = 0.7) +
  geom_bar(aes(y = avg_reading), stat = "identity", fill = "red", alpha = 0.5) +
  geom_bar(aes(y = avg_writing), stat = "identity", fill = "green", alpha = 0.3) +
  labs(title = "Average Scores by Parental Education Level", x = "Parental Education Level", y = "Avera
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Scatter Plot for Math vs Reading Scores

```
ggplot(students_data, aes(x = math_score, y = reading_score)) +
  geom_point(alpha = 0.5) +
  labs(title = "Math vs Reading Scores", x = "Math Score", y = "Reading Score")
```



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.