Performance Task For HISD

Ndubuisi Chibuogwu

2024-04-13

Objective

To independently explore the provided data files, which include students' performance on STAAR EOC for Algebra I and student demographic information. Merge the data sets, perform an exploratory analysis, and communicate my findings via PowerPoint slides. This assignment aims to evaluate my proficiency in data cleaning, analysis, and visualization techniques.

```
# Load necessary packages
library(readxl)
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(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3

## Warning: package 'tibble' was built under R version 4.1.3

## Warning: package 'readr' was built under R version 4.1.3

## Warning: package 'forcats' was built under R version 4.1.3

## Warning: package 'lubridate' was built under R version 4.1.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.5.0     v stringr   1.5.1
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidyr)
library(purrr)
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(latexpdf)
```

```
##
## Attaching package: 'latexpdf'
##
## The following object is masked from 'package:GGally':
##
##     wrap
```

Importing of the data file

Exploratory Data Analysis

```
## [1] 236  20
```

```
## tibble [236 x 20] (S3: tbl_df/tbl/data.frame)
##  $ D_StudentID                    : num [1:236] 1415085 1578121 1424549 1549105 1557558 ...
##  $ D_CampusName                   : chr [1:236] "Campus A" "Campus A" "Campus A" "Campus A" ...
##  $ D_CampusID                     : num [1:236] 100 100 100 100 100 100 100 100 100 27 ...
##  $ Sex_Code                       : chr [1:236] "M" "F" "F" "F" ...
##  $ Date_of_Birth (MMDDYYYY)       : num [1:236] 5022004 6022004 3192004 3062003 1162003 ...
##  $ Ethnicity_Race_Category        : chr [1:236] "H" "H" "H" "H" ...
##  $ Economically_Disadvantaged_Code: num [1:236] 1 0 1 1 1 0 1 1 1 1 ...
```

```
##  $ S_StudentID                    : num [1:236] 1415085 1578121 1424549 1549105 1557558 ...
##  $ S_CampusName                   : chr [1:236] "Campus A" "Campus A" "Campus A" "Campus A" ...
##  $ S_CampusID                     : num [1:236] 100 100 100 100 100 100 100 100 100 27 ...
##  $ End_Of_Course_Code             : chr [1:236] "A1" "A1" "A1" "A1" ...
##  $ Score_Code                     : chr [1:236] "S" "S" "S" "S" ...
##  $ Test_Version                   : chr [1:236] "A" "S" "S" "S" ...
##  $ First_Time_Tester_Retester_info: chr [1:236] "F" "F" "F" "F" ...
##  $ BLANK16                        : logi [1:236] NA NA NA NA NA NA ...
##  $ Raw_Score                      : num [1:236] 37 51 30 27 29 27 35 39 12 34 ...
##  $ Scale_Score                    : num [1:236] 4156 5127 3888 3780 3851 ...
##  $ Approach_Grade_Level           : num [1:236] 1 1 1 1 1 1 1 1 0 1 ...
##  $ Meets_Grade_Level              : num [1:236] 1 1 0 0 0 0 1 1 0 1 ...
##  $ Masters_Grade_Level            : num [1:236] 0 1 0 0 0 0 0 0 0 0 ...
```

```
##   D_StudentID      D_CampusName         D_CampusID       Sex_Code
##  Min.   :1159459   Length:236         Min.   :  6.00   Length:236
##  1st Qu.:1373464   Class :character   1st Qu.: 27.00   Class :character
##  Median :1513117   Mode  :character   Median : 27.00   Mode  :character
##  Mean   :1458300                      Mean   : 45.75
##  3rd Qu.:1557621                      3rd Qu.: 27.00
##  Max.   :1629992                      Max.   :250.00
##
##  Date_of_Birth (MMDDYYYY) Ethnicity_Race_Category
##  Min.   : 1012003         Length:236
##  1st Qu.: 3184504         Class :character
##  Median : 7042004         Mode  :character
##  Mean   : 6680012
##  3rd Qu.:10044503
##  Max.   :12302003
##
##  Economically_Disadvantaged_Code  S_StudentID      S_CampusName
##  Min.   :0.0000                  Min.   :1159459   Length:236
##  1st Qu.:1.0000                  1st Qu.:1373464   Class :character
##  Median :1.0000                  Median :1513117   Mode  :character
##  Mean   :0.8814                  Mean   :1458300
##  3rd Qu.:1.0000                  3rd Qu.:1557621
##  Max.   :1.0000                  Max.   :1629992
##
##    S_CampusID     End_Of_Course_Code  Score_Code         Test_Version
##  Min.   :  6.00   Length:236         Length:236         Length:236
##  1st Qu.: 27.00   Class :character   Class :character   Class :character
##  Median : 27.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 45.75
##  3rd Qu.: 27.00
##  Max.   :250.00
##
##  First_Time_Tester_Retester_info BLANK16          Raw_Score      Scale_Score
##  Length:236                      Mode:logical   Min.   : 7.00   Min.   :2886
##  Class :character                NA's:236       1st Qu.:19.00   1st Qu.:3500
##  Mode  :character                               Median :30.00   Median :3888
##                                                 Mean   :29.69   Mean   :3917
##                                                 3rd Qu.:39.00   3rd Qu.:4242
##                                                 Max.   :53.00   Max.   :5640
##                                                 NA's   :1       NA's   :1
```

```
##  Approach_Grade_Level Meets_Grade_Level Masters_Grade_Level
##  Min.   :0.0000       Min.   :0.000     Min.   :0.0000
##  1st Qu.:0.0000       1st Qu.:0.000     1st Qu.:0.0000
##  Median :1.0000       Median :0.000     Median :0.0000
##  Mean   :0.7034       Mean   :0.428     Mean   :0.2161
##  3rd Qu.:1.0000       3rd Qu.:1.000     3rd Qu.:0.0000
##  Max.   :1.0000       Max.   :1.000     Max.   :1.0000
##
```

Data Cleaning & Preprocessing

```
##  [1] "D_StudentID"                    "D_CampusName"
##  [3] "D_CampusID"                     "Sex_Code"
##  [5] "Date_of_Birth (MMDDYYYY)"       "Ethnicity_Race_Category"
##  [7] "Economically_Disadvantaged_Code" "S_StudentID"
##  [9] "S_CampusName"                   "S_CampusID"
## [11] "End_Of_Course_Code"             "Score_Code"
## [13] "Test_Version"                   "First_Time_Tester_Retester_info"
## [15] "BLANK16"                        "Raw_Score"
## [17] "Scale_Score"                    "Approach_Grade_Level"
## [19] "Meets_Grade_Level"              "Masters_Grade_Level"
```

```
## [1] 238
```

```
##                     D_StudentID                     D_CampusName
##                               0                                0
##                      D_CampusID                        Sex_Code
##                               0                                0
##        Date_of_Birth (MMDDYYYY)         Ethnicity_Race_Category
##                               0                                0
## Economically_Disadvantaged_Code                      S_StudentID
##                               0                                0
##                    S_CampusName                       S_CampusID
##                               0                                0
##              End_Of_Course_Code                       Score_Code
##                               0                                0
##                    Test_Version First_Time_Tester_Retester_info
##                               0                                0
##                         BLANK16                        Raw_Score
##                             236                                1
##                     Scale_Score             Approach_Grade_Level
##                               1                                0
##               Meets_Grade_Level              Masters_Grade_Level
##                               0                                0
```

```
## [1] "BLANK16"     "Raw_Score"   "Scale_Score"
```

```
## [1] 236
```

```
## [1] 235
```

```
##                     D_StudentID                     D_CampusName
##                               0                                0
##                      D_CampusID                        Sex_Code
##                               0                                0
##        Date_of_Birth (MMDDYYYY)         Ethnicity_Race_Category
##                               0                                0
## Economically_Disadvantaged_Code                      S_StudentID
##                               0                                0
##                    S_CampusName                       S_CampusID
##                               0                                0
##              End_Of_Course_Code                       Score_Code
##                               0                                0
##                    Test_Version First_Time_Tester_Retester_info
##                               0                                0
##                          BLANK16                        Raw_Score
##                             235                                0
##                     Scale_Score             Approach_Grade_Level
##                               0                                0
##               Meets_Grade_Level              Masters_Grade_Level
##                               0                                0
```

```
## [1] "BLANK16"
```

```
## [1] 16
```

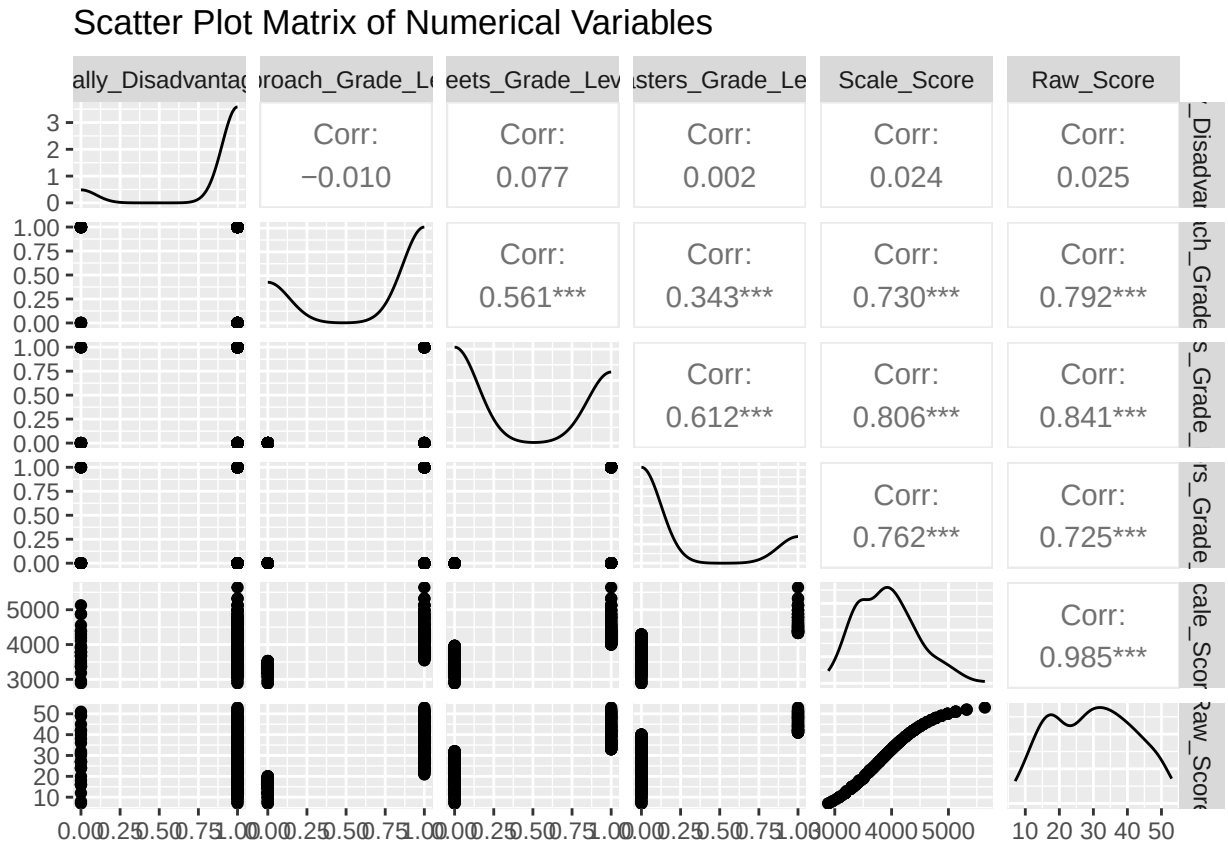
```
##  [1] "D_StudentID"                     "D_CampusName"
##  [3] "D_CampusID"                      "Sex_Code"
##  [5] "Date_of_Birth (MMDDYYYY)"        "Ethnicity_Race_Category"
##  [7] "Economically_Disadvantaged_Code" "End_Of_Course_Code"
##  [9] "Score_Code"                      "Test_Version"
## [11] "First_Time_Tester_Retester_info" "Raw_Score"
## [13] "Scale_Score"                     "Approach_Grade_Level"
## [15] "Meets_Grade_Level"               "Masters_Grade_Level"
```

```
##  [1] "D_StudentID"                     "D_CampusName"
##  [3] "D_CampusID"                      "Sex_Code"
##  [5] "Birthdate"                       "Ethnicity_Race_Category"
##  [7] "Economically_Disadvantaged_Code" "End_Of_Course_Code"
##  [9] "Score_Code"                      "Test_Version"
## [11] "First_Time_Tester_Retester_info" "Raw_Score"
## [13] "Scale_Score"                     "Approach_Grade_Level"
## [15] "Meets_Grade_Level"               "Masters_Grade_Level"
```

```
##   [1] "2004-05-02" "2004-06-02" "2004-03-19" "2003-03-06" "2003-01-16"
##   [6] "2003-10-29" "2004-03-04" "2004-05-25" "2003-08-10" "2003-08-22"
##  [11] "2003-09-09" "2004-03-09" "2003-06-16" "2004-03-10" "2003-11-11"
##  [16] "2004-03-04" "2004-07-22" "2004-08-04" "2004-05-11" "2003-10-17"
##  [21] "2002-09-17" "2004-07-14" "2003-12-20" "2004-04-26" "2002-12-11"
##  [26] "2001-05-17" "2004-08-27" "2003-02-04" "2004-03-22" "2004-07-22"
##  [31] "2003-12-19" "2003-08-03" "2004-02-22" "2003-11-11" "2002-11-28"
##  [36] "2003-04-25" "2002-11-11" "2003-09-17" "2004-04-20" "2004-06-25"
##  [41] "2004-06-25" "2003-12-02" "2004-08-24" "2002-08-18" "2002-02-06"
##  [46] "2004-06-21" "2001-12-11" "2004-08-20" "2003-09-26" "2004-07-19"
```

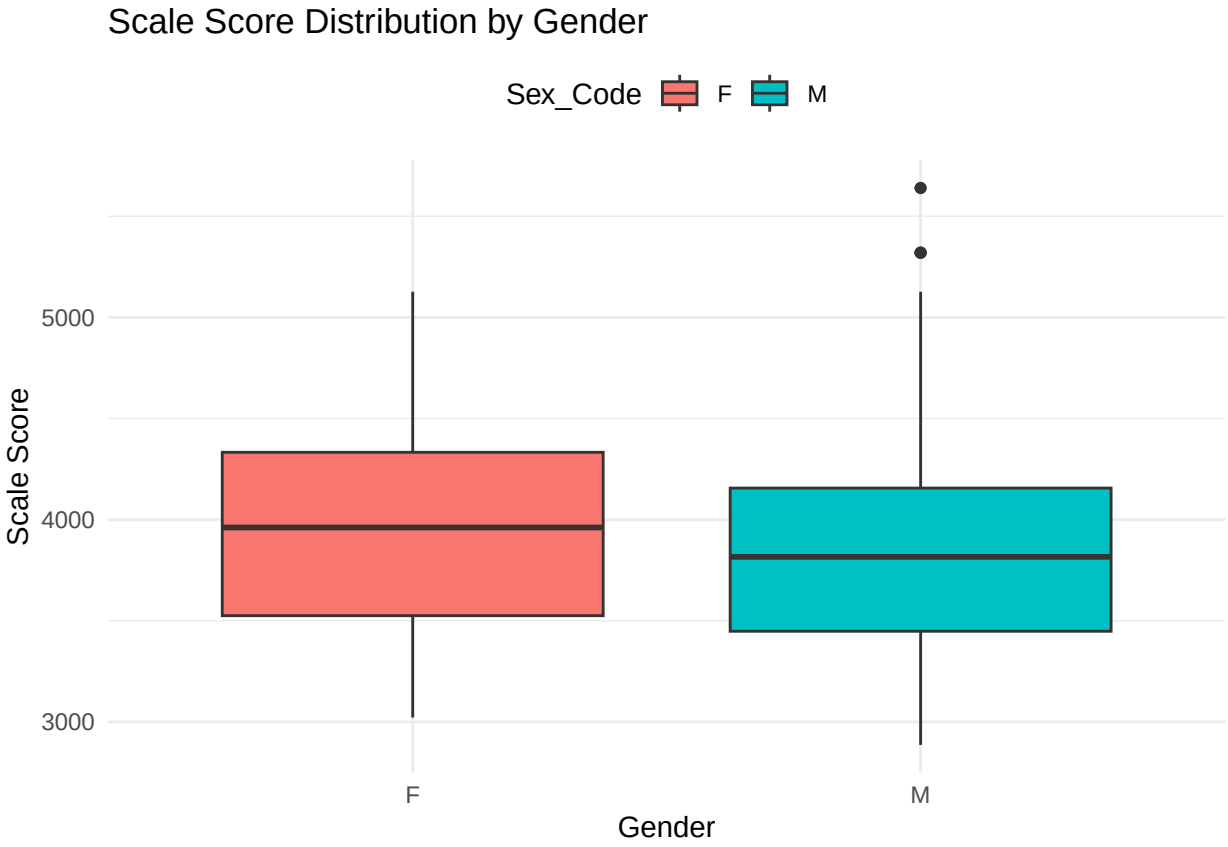
```
##  [51] "2003-11-26" "2003-10-11" "2003-11-30" "2003-09-24" "2004-05-22"
##  [56] "2004-02-13" "2004-05-02" "2003-01-07" "2004-08-13" "2004-07-20"
##  [61] "2003-02-21" "2003-09-04" "2003-09-13" "2001-05-12" "2002-08-26"
##  [66] "2003-01-08" "2004-01-30" "2003-08-12" "2004-06-27" "2002-03-13"
##  [71] "2003-08-23" "2004-01-13" "2004-01-05" "2004-08-27" "2004-05-10"
##  [76] "2004-08-03" "2004-07-15" "2004-01-21" "2004-01-29" "2002-11-01"
##  [81] "2003-06-30" "2002-03-13" "2002-06-28" "2004-02-18" "2003-12-30"
##  [86] "2003-10-09" "2004-06-02" "2001-01-23" "2003-11-04" "2004-06-16"
##  [91] "2003-08-04" "2004-04-09" "2003-10-03" "2003-07-04" "2002-10-21"
##  [96] "2002-10-09" "2002-10-11" "2002-09-25" "2004-05-04" "2003-11-07"
## [101] "2003-05-20" "2003-12-23" "2004-08-11" "2003-09-03" "2003-11-13"
## [106] "2003-11-10" "2003-02-11" "2003-07-16" "2003-09-04" "2003-02-09"
## [111] "2003-12-17" "2004-01-22" "2004-05-31" "2003-10-09" "2004-05-03"
## [116] "2004-04-15" "2004-03-03" "2004-06-03" "2003-12-12" "2004-05-13"
## [121] "2004-04-18" "2003-12-08" "2003-12-08" "2003-06-04" "2004-02-20"
## [126] "2003-12-01" "2003-09-03" "2004-02-06" "2003-10-03" "2004-04-13"
## [131] "2004-01-23" "2003-11-13" "2003-02-20" "2004-06-20" "2003-10-08"
## [136] "2003-07-20" "2004-02-17" "2003-04-12" "2003-12-13" "2003-12-13"
## [141] "2003-09-07" "2004-08-17" "2003-10-11" "2003-02-22" "2003-12-04"
## [146] "2003-01-01" "2004-01-27" "2004-02-09" "2004-06-26" "2004-02-10"
## [151] "2003-03-24" "2004-08-24" "2002-07-23" "2004-02-19" "2003-01-21"
## [156] "2004-07-21" "2004-03-21" "2004-07-13" "2004-04-14" "2004-01-30"
## [161] "2003-01-09" "2004-07-04" "2003-12-16" "2001-12-16" "2004-03-16"
## [166] "2003-04-24" "2004-02-20" "2001-12-30" "2004-03-25" "2002-01-21"
## [171] "2003-10-03" "2004-08-18" "2004-06-13" "2004-01-30" "2004-01-08"
## [176] "2002-09-27" "2003-11-21" "2004-01-28" "2003-09-04" "2003-08-09"
## [181] "2003-04-16" "2002-10-09" "2002-10-11" "2002-09-25" "2004-05-04"
## [186] "2003-11-07" "2003-05-20" "2003-12-23" "2004-08-11" "2003-09-03"
## [191] "2003-11-13" "2003-11-10" "2003-02-11" "2003-07-16" "2003-09-04"
## [196] "2003-02-09" "2003-12-17" "2004-01-22" "2004-05-31" "2003-10-09"
## [201] "2004-05-03" "2004-04-15" "2004-03-03" "2004-06-03" "2003-12-12"
## [206] "2004-05-13" "2004-04-18" "2003-12-08" "2003-12-08" "2003-06-04"
## [211] "2004-02-20" "2003-12-01" "2003-09-03" "2004-02-06" "2003-10-03"
## [216] "2004-04-13" "2004-01-23" "2003-11-13" "2003-02-20" "2004-06-20"
## [221] "2003-10-08" "2003-07-20" "2004-02-17" "2003-04-12" "2003-12-13"
## [226] "2003-12-13" "2003-09-07" "2004-08-17" "2003-10-11" "2003-02-22"
## [231] "2003-12-04" "2003-01-01" "2004-01-27" "2004-02-09" "2004-06-26"
```

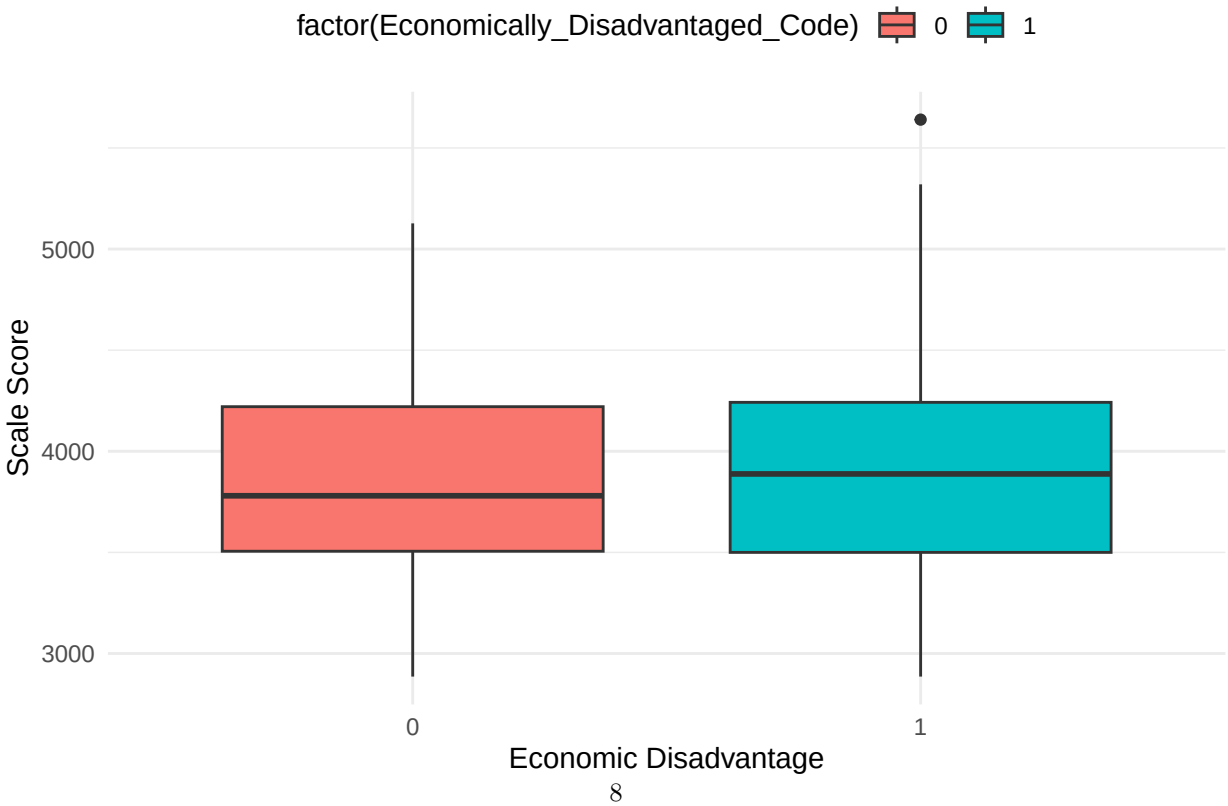
Correlation Analysis




```
##                                 Economically_Disadvantaged_Code
## Economically_Disadvantaged_Code                     1.000000000
## Approach_Grade_Level                               -0.009777657
## Meets_Grade_Level                                   0.077438958
## Masters_Grade_Level                                 0.002440700
## Scale_Score                                         0.023553779
## Raw_Score                                           0.024591255
##                                 Approach_Grade_Level Meets_Grade_Level
## Economically_Disadvantaged_Code         -0.009777657        0.07743896
## Approach_Grade_Level                     1.000000000        0.56058331
## Meets_Grade_Level                        0.560583310        1.00000000
## Masters_Grade_Level                      0.342912427        0.61170645
## Scale_Score                              0.730180396        0.80634135
## Raw_Score                                0.792331412        0.84076600
##                                 Masters_Grade_Level Scale_Score  Raw_Score
## Economically_Disadvantaged_Code           0.0024407  0.02355378 0.02459125
## Approach_Grade_Level                      0.3429124  0.73018040 0.79233141
## Meets_Grade_Level                         0.6117065  0.80634135 0.84076600
## Masters_Grade_Level                       1.0000000  0.76229364 0.72491225
## Scale_Score                               0.7622936  1.00000000 0.98512316
## Raw_Score                                 0.7249123  0.98512316 1.00000000
```

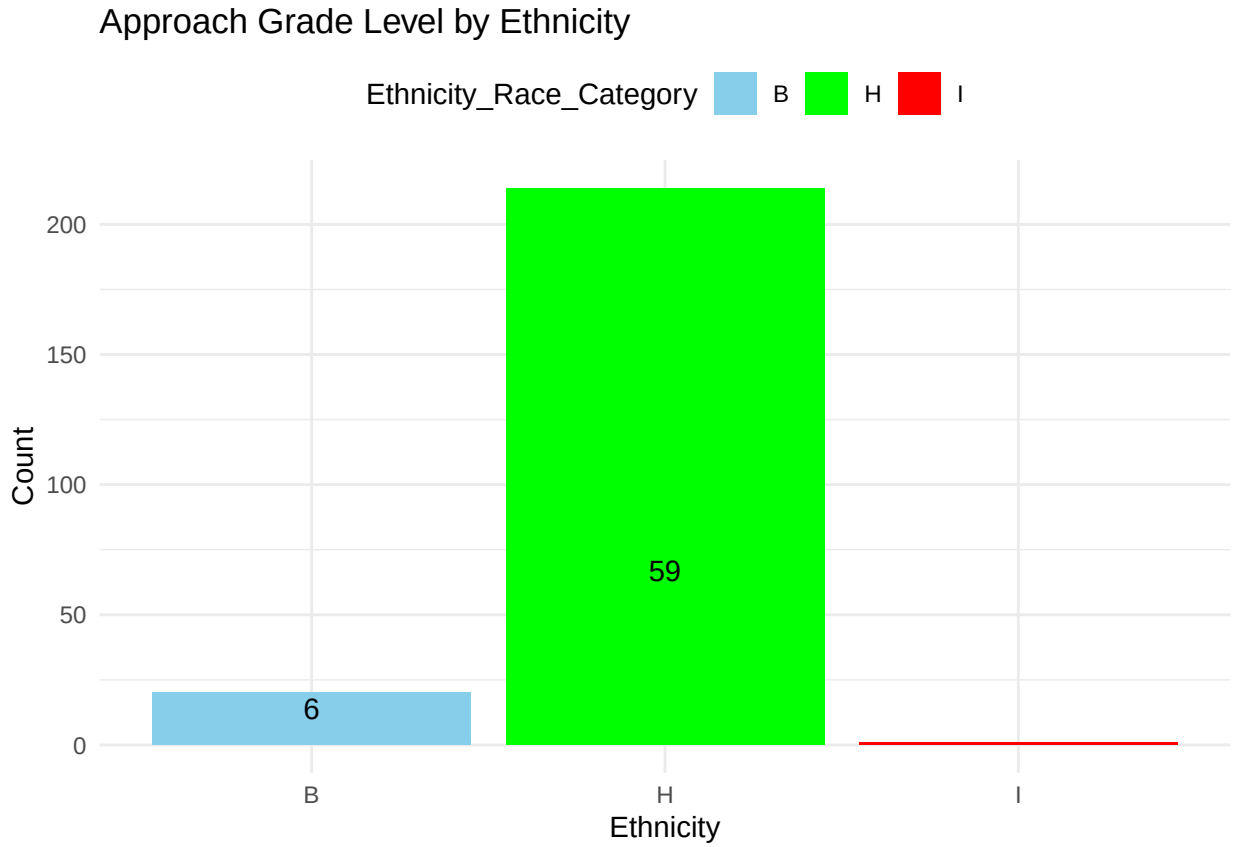
Data Analysis & Visualization

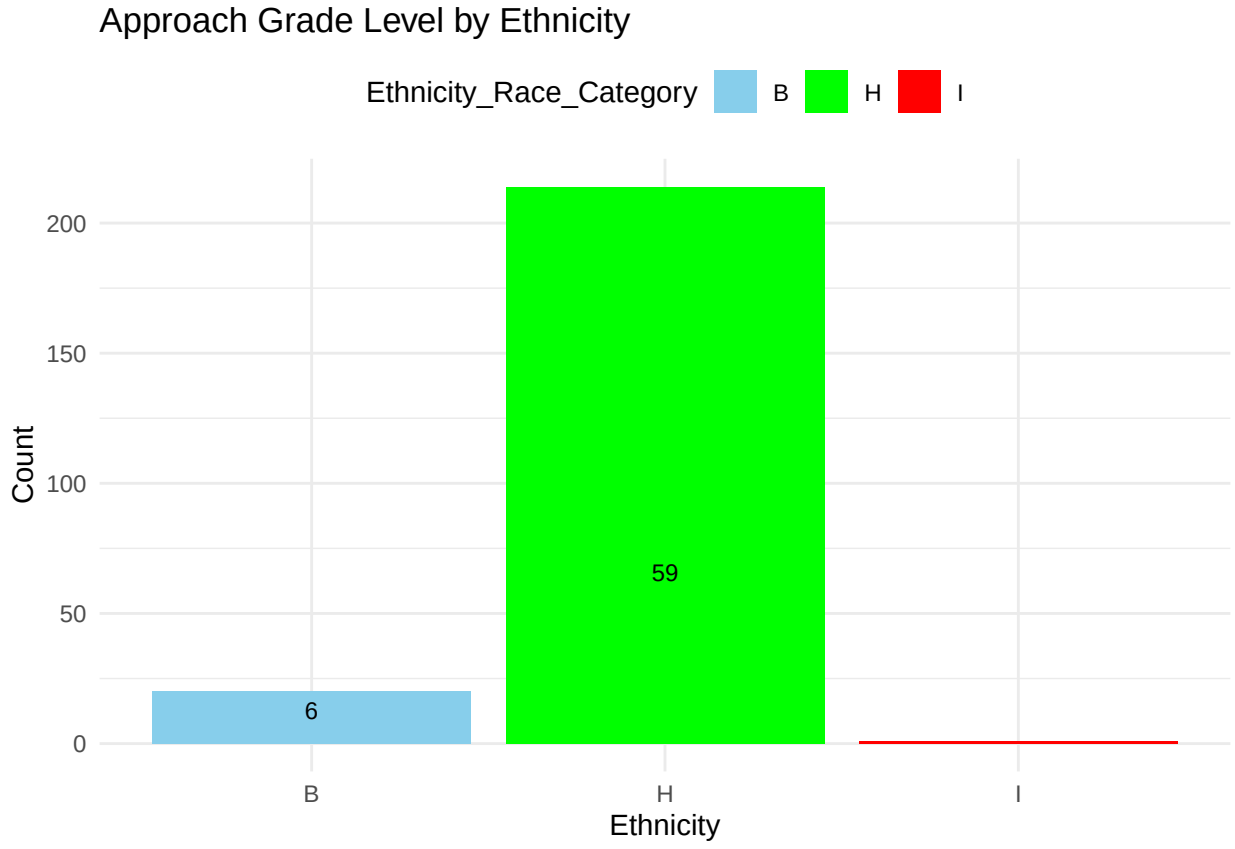




Male and female students have similar median scale scores. However, females show a slightly wider interquartile range, indicating greater variability in their scores. Poor students tend to have lower median scores compared to their non-disadvantaged peers, highlighting a potential achievement gap

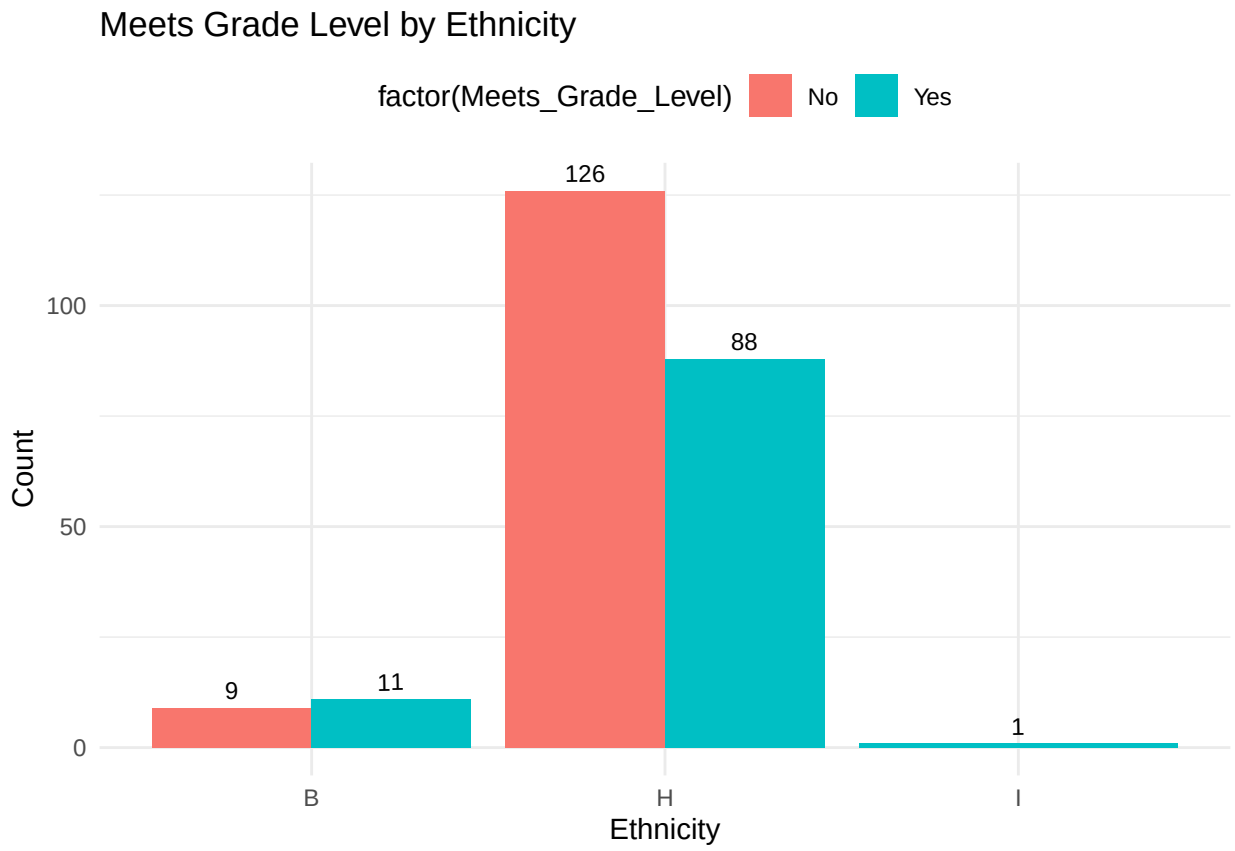
Demographic Effects on Academic Performance



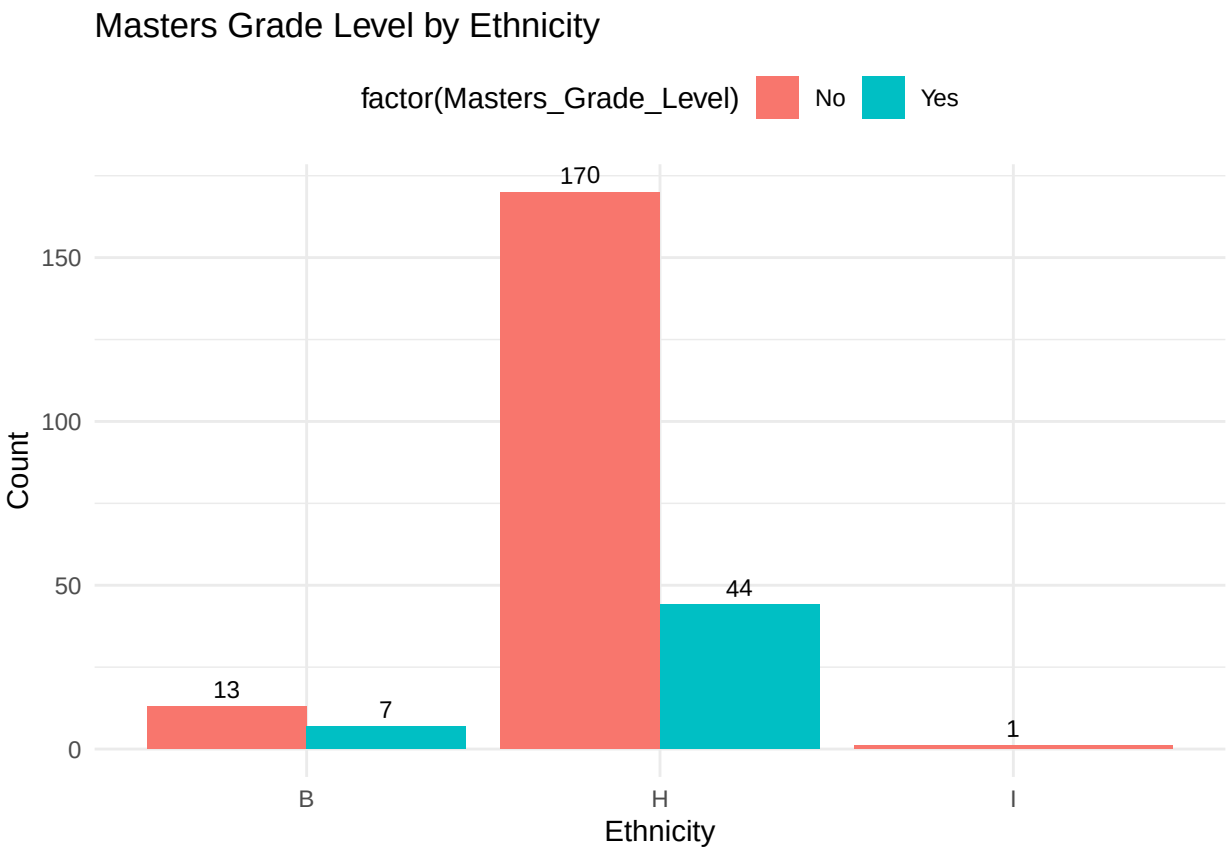


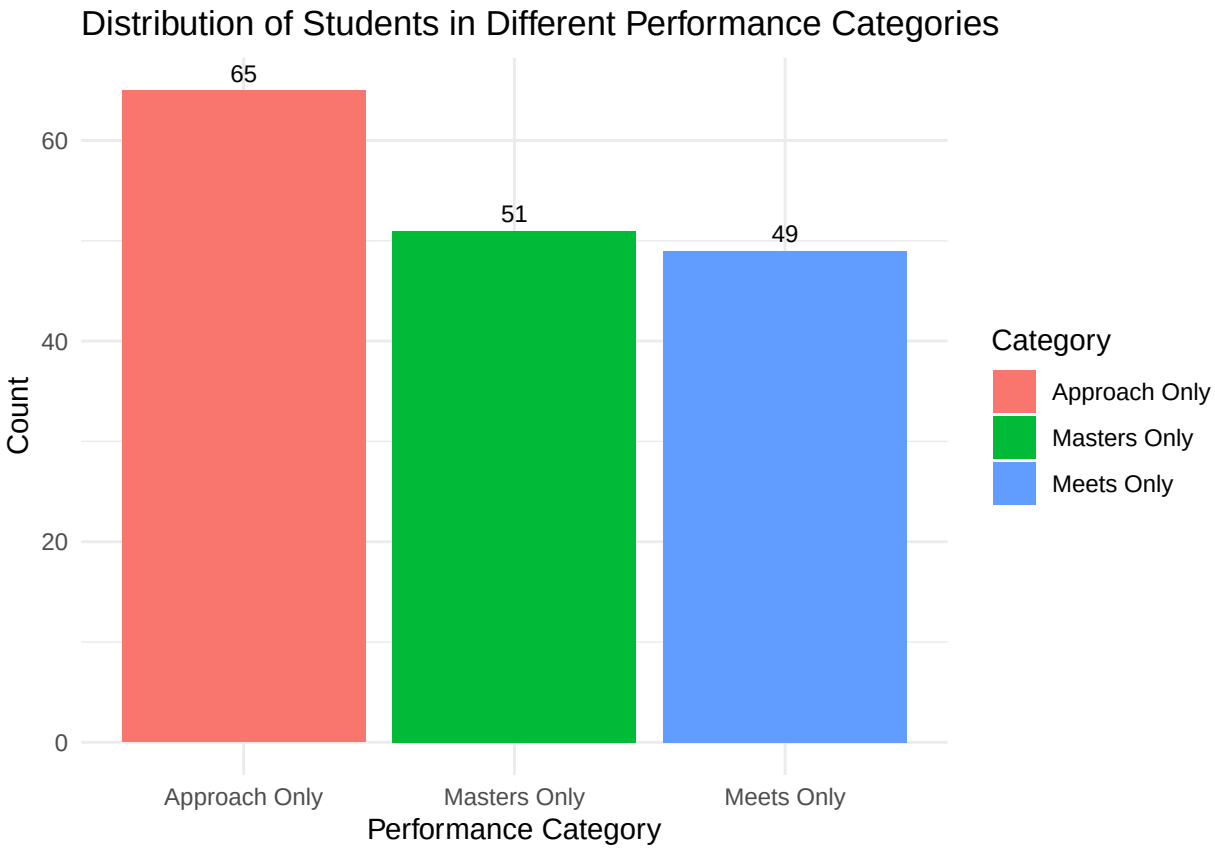
```
## [1] 49
```

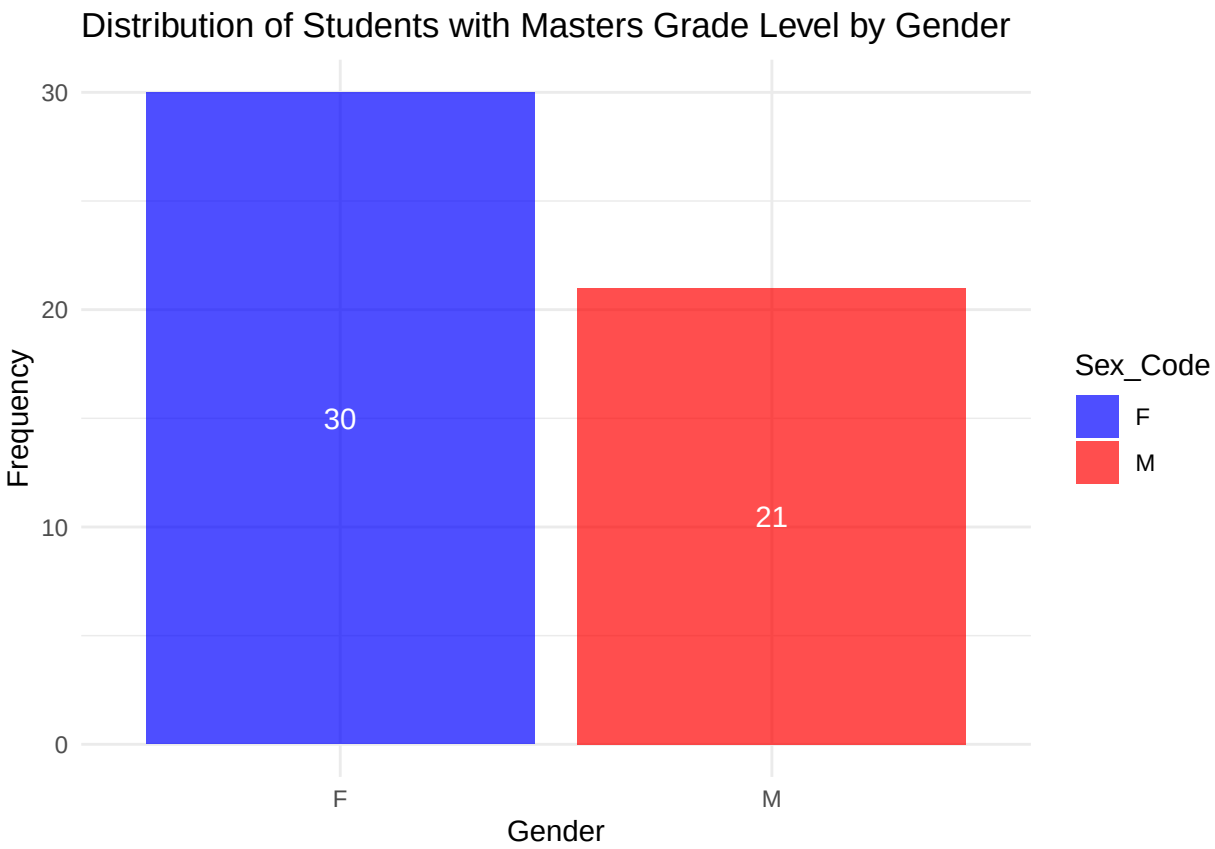
```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




```
## [1] 51
```


Masters Grade Level by Ethnicity
factor(Masters_Grade_Level)
No
Yes
Count
150
100
50
0
170
44
13
7
1
B
H
I
Ethnicity

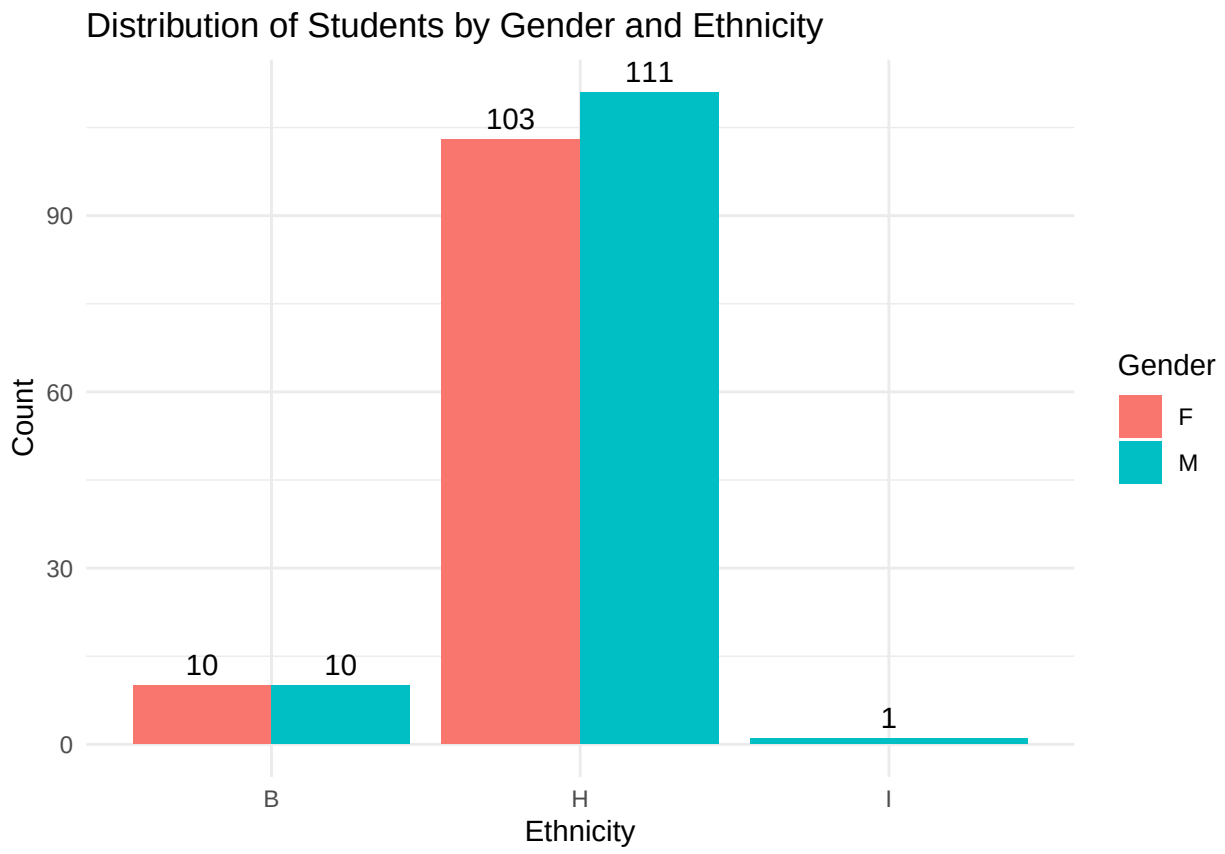

Distribution of Students in Different Performance Categories
65
51
49
60
40
20
0
Count
Approach Only
Masters Only
Meets Only
Performance Category
Category
Approach Only
Masters Only
Meets Only



Meets Grade Level by Ethnicity: Most students in each ethnic category did not meet the grade level, with Hispanic students having the highest number of students both meeting and not meeting the grade level. This is reflective of the predominant demographic in the data set.

Masters Grade Level by Gender: A higher proportion of female students appear to master the grade level compared to male students, suggesting better performance at the highest achievement level.

```
## `summarise()` has grouped output by 'Sex_Code'. You can override using the
## `.groups` argument.
```




```
##
##       B   H   I
##   F  10 103   0
##   M  10 111   1
```

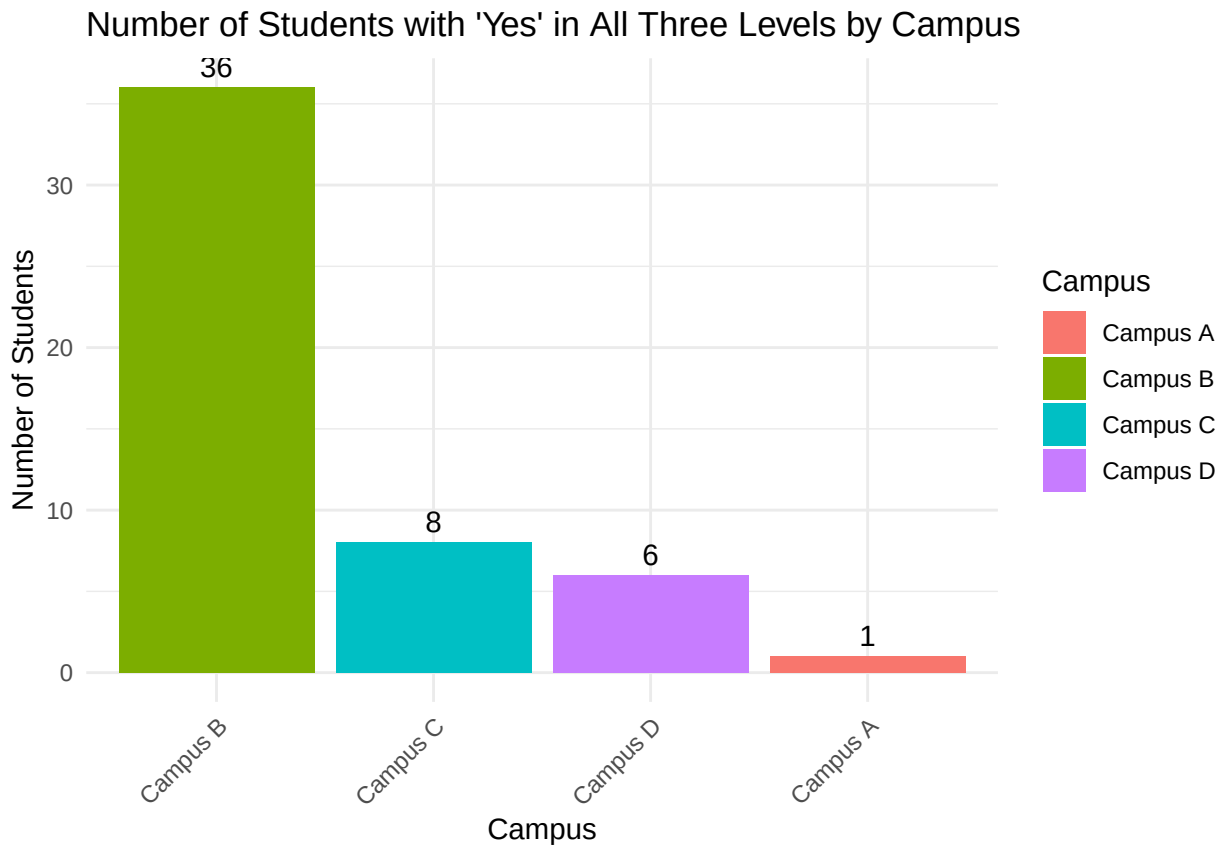
Hispanic make up over 90% of students in all the campuses in the Houston ISD, with male Hispanic students accounting for about 47% of the total population

```
# Filter data for students who have "Yes" (1) in all three levels
  students_all_yes <- Combined[Combined$Approach_Grade_Level == 1 &
                               Combined$Meets_Grade_Level == 1 &
                               Combined$Masters_Grade_Level == 1, ]

# Count the number of students from each campus
  campus_counts <- table(students_all_yes$D_CampusName)

# Convert campus counts to data frame
  campus_counts_df <- data.frame(Campus = names(campus_counts),
                                 Students_with_All_3_Levels = as.numeric(campus_counts))

  print(view(campus_counts_df))
```

```
##     Campus Students_with_All_3_Levels
## 1 Campus A                          1
## 2 Campus B                         36
```

```
## 3 Campus C                          8
## 4 Campus D                          6
```

```r
# Create bar plot
  print(ggplot(campus_counts_df,
         aes(x = reorder(Campus, -Students_with_All_3_Levels),
         y = Students_with_All_3_Levels, fill = Campus)) +
         geom_bar(stat = "identity") +
         geom_text(aes(label = Students_with_All_3_Levels), vjust = -0.5, color = "black") +
         labs(title = "Number of Students with 'Yes' in All Three Levels by Campus",
         x = "Campus", y = "Number of Students") +
         theme_minimal() +
         theme(axis.text.x = element_text(angle = 45, hjust = 1)))
```




```r
# Calculate average Raw_Score by campus
  average_scores <- aggregate(Raw_Score ~ D_CampusName, data = Combined, FUN = mean)
  average_scores <- average_scores %>%
  mutate(Raw_Score = round(Raw_Score, 2)) %>%
  select(D_CampusName, Raw_Score)  # Keep only the rounded numeric column and the categorical column

# View the modified data frame
  print(view(average_scores))
```

```
##   D_CampusName Raw_Score
## 1     Campus A     31.89
```

```
## 2     Campus B     29.65
## 3     Campus C     31.03
## 4     Campus D     26.75
```

```r
# Create bar plot
# Define custom colors for each campus
  campus_colors <- c("Campus A" = "blue", "Campus B" = "red", "Campus C" = "green", "Campus D" = "purple

# Bar plot with distinct coloring for different campuses
  print(ggplot(average_scores,
         aes(x = reorder(D_CampusName, -Raw_Score),
         y = Raw_Score, fill = D_CampusName)) +
         geom_bar(stat = "identity") +
         geom_text(aes(label = round(Raw_Score)), vjust = -0.5, color = "black") +
         scale_fill_manual(values = campus_colors) +
         labs(title = "Average Raw Score by Campus",
         x = "Campus", y = "Average Raw Score") +
         theme_minimal() +
         theme(axis.text.x = element_text(angle = 45, hjust = 1)))
```