

Student Demographics, Attendance, and Course Outcomes

2025-08-02

```
list <- read_csv("../Data/TUMO Yerevan Center Report_Students List_Table - Sheet1.csv")
perf <- read_csv("../Data/TUMO Yerevan_Students Performance_Table - Sheet1.csv")
work <- read_csv("../Data/TUMO Armenia Center Report_Workshops Statistics_Table.csv")

list <- list %>% mutate(TumoID = as.character(TumoID))
perf <- perf %>% mutate(TumoID = as.character(TumoID))
work <- work %>% mutate(TumoID = as.character(TumoID))

gender_info <- work %>%
  select(TumoID, Gender) %>%
  distinct(TumoID, .keep_all = TRUE)

main_df <- list %>%
  select(TumoID, Age, PresenceRatio) %>%
  left_join(gender_info, by = "TumoID") %>%
  left_join(perf %>% select(TumoID, Completed, Incomplete, Withdrawn, Participated), by = "TumoID") %>%
  mutate(
    PresenceRatio = str_replace_all(PresenceRatio, ",", "."),
    PresenceRatio = as.numeric(str_extract(PresenceRatio, "\\d+(\\.\\d+)?")) / 100
  ) %>%
  select(TumoID, Gender, Age, PresenceRatio, Completed, Incomplete, Withdrawn, Participated)

main_df
```

```
## # A tibble: 10,428 x 8
##   TumoID Gender   Age PresenceRatio Completed Incomplete Withdrawn Participated
##   <chr>   <chr>   <dbl>         <dbl>     <dbl>      <dbl>      <dbl>      <dbl>
## 1 2.303~ Male     15         0.85         2          1          1          1
## 2 2.301~ <NA>     14         0.9          NA         NA         NA         NA
## 3 2.404~ <NA>     13         0.86         NA         NA         NA         NA
## 4 2.504~ <NA>     12         0.81         0          0          0          0
## 5 2.211~ <NA>     14         0.84         NA         NA         NA         NA
## 6 2.306~ Male     15         0.81         NA         NA         NA         NA
## 7 2.310~ <NA>     14         0.83         NA         NA         NA         NA
## 8 2.412~ Female   14         0.85         1          0          0          1
## 9 2.412~ <NA>     12         0.75         NA         NA         NA         NA
## 10 2.408~ Female   12         0.95         1          1          1          0
## # i 10,418 more rows
```

Attendance by Gender

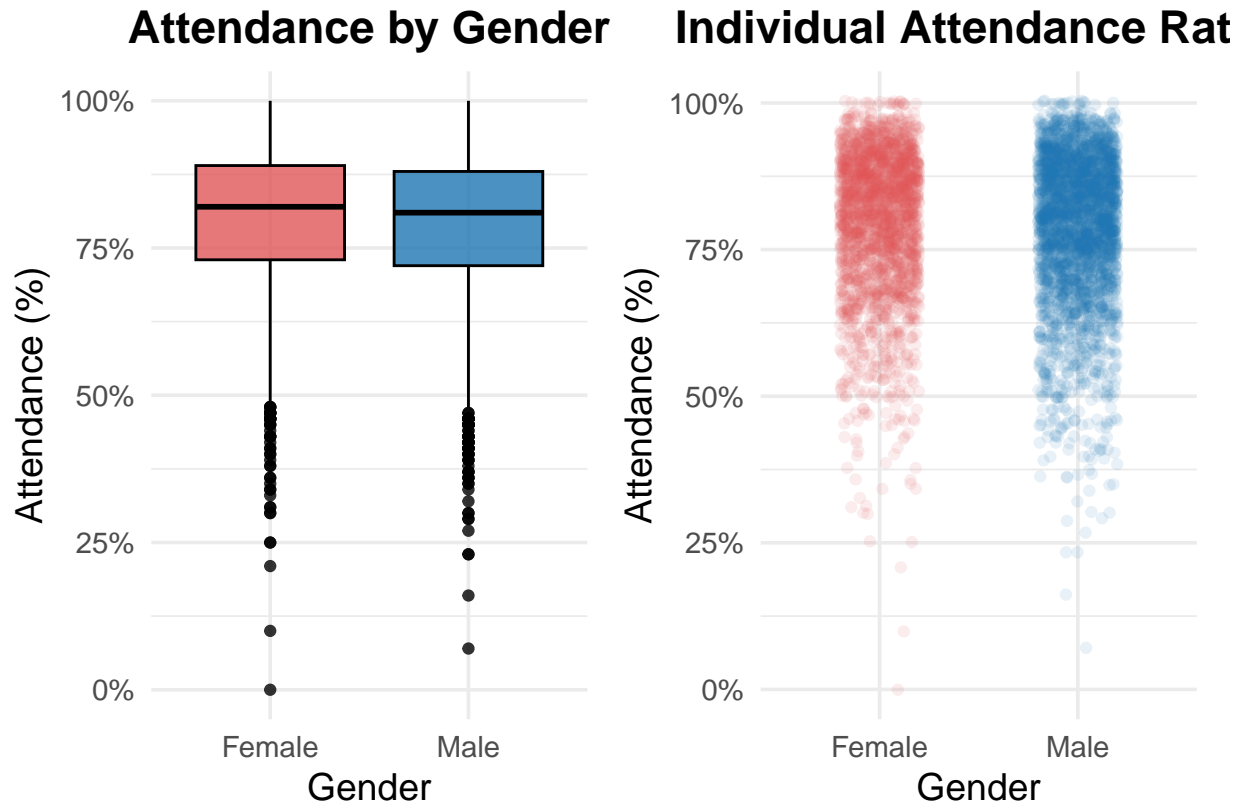
```

box_plot <- main_df %>%
  filter(!is.na(Gender), !is.na(PresenceRatio), PresenceRatio <= 1) %>%
  mutate(Gender = case_when(
    str_to_lower(str_trim(Gender)) %in% c("male", "m") ~ "Male",
    str_to_lower(str_trim(Gender)) %in% c("female", "f") ~ "Female"
  )) %>%
  ggplot(aes(x = Gender, y = PresenceRatio, fill = Gender)) +
  geom_boxplot(alpha = 0.8, color = "black") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  scale_fill_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(title = "Attendance by Gender", x = "Gender", y = "Attendance (%)") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", hjust = 0.5), legend.position = "none")

dot_plot <- main_df %>%
  filter(!is.na(Gender), !is.na(PresenceRatio), PresenceRatio <= 1) %>%
  mutate(Gender = case_when(
    str_to_lower(str_trim(Gender)) %in% c("male", "m") ~ "Male",
    str_to_lower(str_trim(Gender)) %in% c("female", "f") ~ "Female"
  )) %>%
  ggplot(aes(x = Gender, y = PresenceRatio, color = Gender)) +
  geom_jitter(width = 0.2, size = 1.5, alpha = 0.1) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  scale_color_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(title = "Individual Attendance Rates", x = "Gender", y = "Attendance (%)") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", hjust = 0.5), legend.position = "none")

box_plot + dot_plot

```



Boxplot Interpretation:

The left-hand boxplot gives us a quick summary of how consistently male and female students attend their sessions. Both groups show very similar attendance patterns — the median is slightly higher for female students, but overall both boxes are close together and sit around the 75–90% range. This tells us that, on average, **students of all genders are attending regularly**, with only a few outliers missing a lot of sessions.

Individual Student Patterns:

The right-hand scatter plot shows **each student** as a point. We can clearly see that most dots are around the 80–90% zone. That means the **majority of students attend often and consistently**, no matter their gender. The spread below 50% is very light, so there are only a small number of students who have a poor attendance rate.

There's no obvious gap between male and female groups — **both show very similar behavior**, confirming that gender doesn't play a big role in who shows up more often.

Gender-Based Course Enrollment Analysis

```
bar_plot <- main_df %>%
  filter(Gender %in% c("Male", "Female")) %>%
  count(Gender) %>%
  ggplot(aes(x = Gender, y = n, fill = Gender)) +
  geom_col(width = 0.6) +
  scale_fill_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(title = "Total Course Enrollments", x = "Gender", y = "Enrollments") +
  theme_minimal(base_size = 14) +
```

```

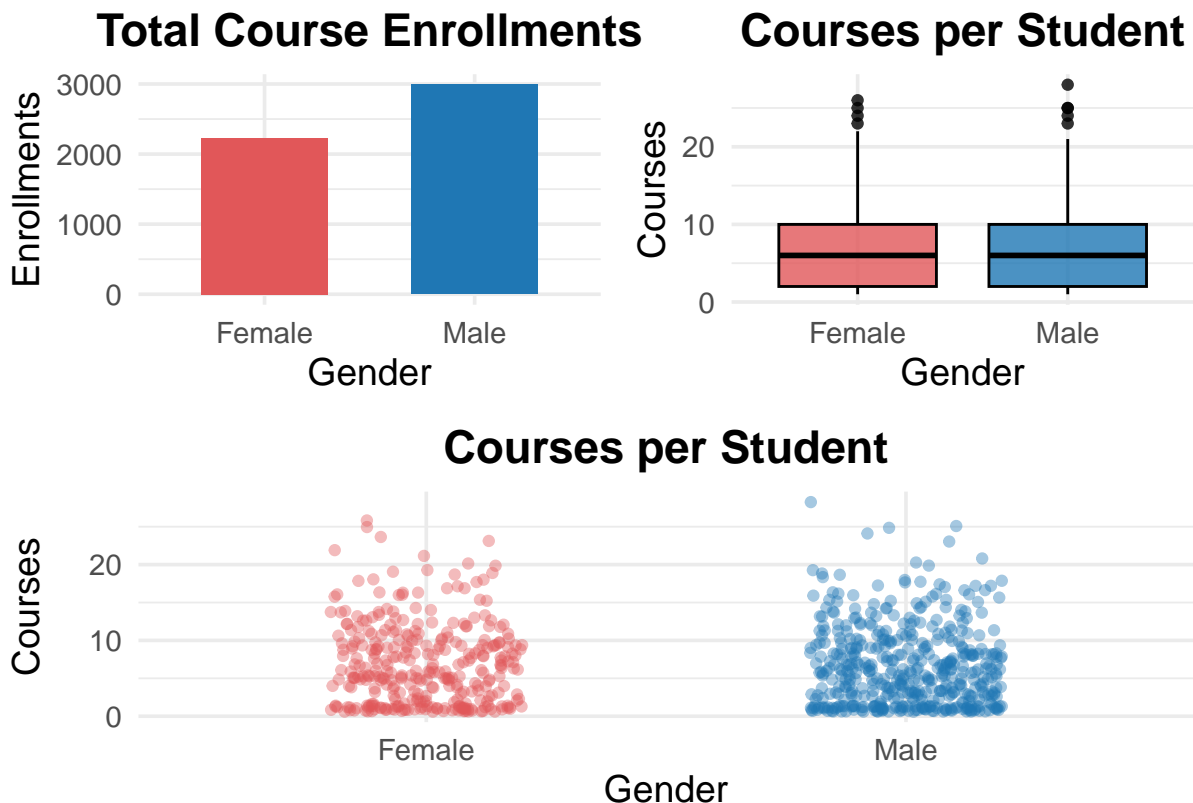
theme(plot.title = element_text(face = "bold", hjust = 0.5), legend.position = "none")

box_plot <- main_df %>%
  filter(!is.na(TumoID), Gender %in% c("Male", "Female")) %>%
  count(TumoID, Gender) %>%
  ggplot(aes(x = Gender, y = n, fill = Gender)) +
  geom_boxplot(alpha = 0.8, color = "black") +
  scale_fill_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(title = "Courses per Student", x = "Gender", y = "Courses") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", hjust = 0.5), legend.position = "none")

jitter_plot <- main_df %>%
  filter(!is.na(TumoID), Gender %in% c("Male", "Female")) %>%
  count(TumoID, Gender) %>%
  ggplot(aes(x = Gender, y = n, color = Gender)) +
  geom_jitter(width = 0.2, alpha = 0.4, size = 1.5) +
  scale_color_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(title = "Courses per Student", x = "Gender", y = "Courses") +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", hjust = 0.5), legend.position = "none")

(bar_plot | box_plot) / jitter_plot

```



Total Enrollments by Gender The bar chart in the top-left shows that **male students have slightly more total course enrollments than female students**. While the gap isn't massive, it suggests that males are enrolling in TUMO courses a bit more often.

Courses Taken per Student (Boxplot) The top-right boxplot summarizes how many courses each student takes, by gender. The **median** number of courses is quite similar between males and females, but males show a **wider spread** — meaning some male students take many more courses than the average.

Individual Course Counts (Jitter Plot)

The scatter plot represents each individual student as a point, showing their total course counts.

We can see that:

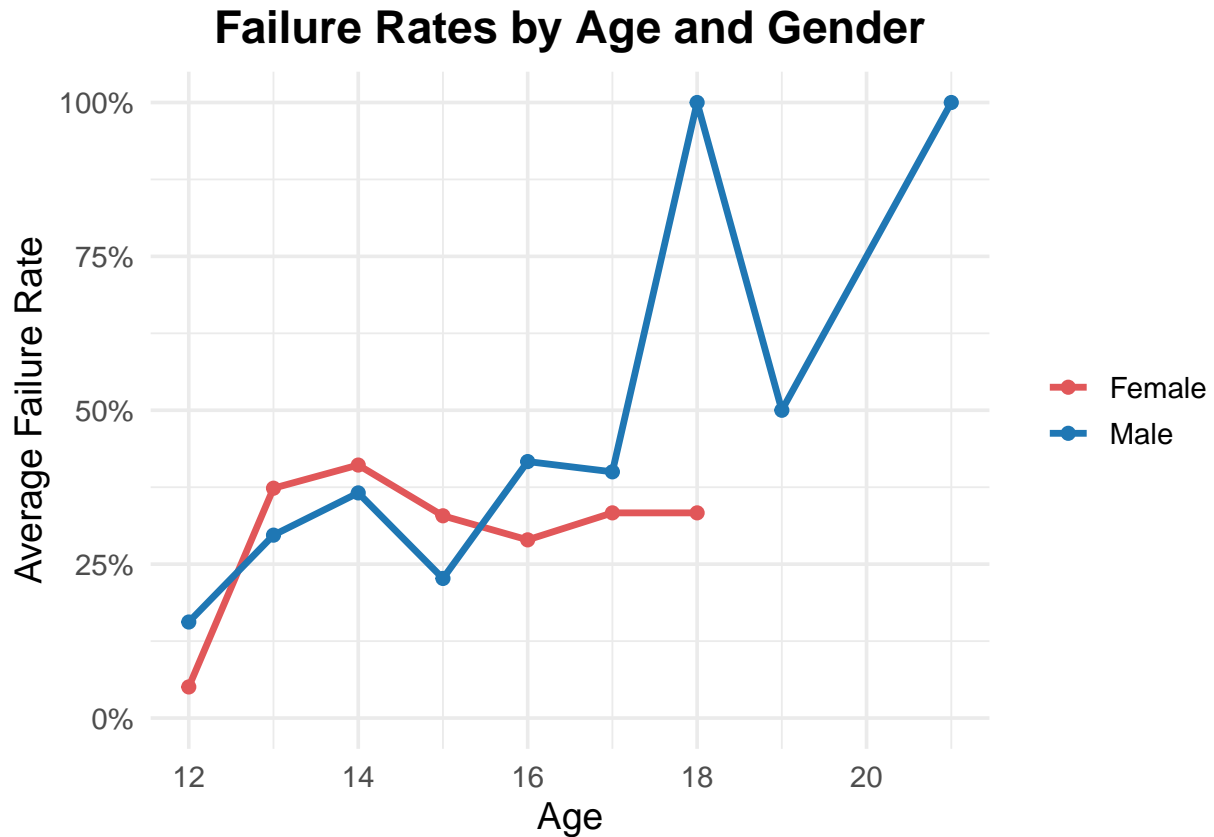
Most students take between 1 and 2 courses. Both genders have a similar concentration of students in the mid-range. Male students have slightly more variation, including more students who take over 20 courses. Female students have fewer extreme outliers.

Overall, the patterns are comparable, with no huge gender differences.

Failure Rates by Age & Gender

```
failure_df <- main_df %>%
  filter(!is.na(Gender), !is.na(Age), !is.na(Participated), Participated > 0) %>%
  filter((Incomplete + Withdrawn) <= Participated) %>%
  mutate(failure_rate = (Incomplete + Withdrawn) / Participated) %>%
  group_by(Gender, Age) %>%
  summarise(mean_failure = mean(failure_rate, na.rm = TRUE), .groups = "drop")

ggplot(failure_df, aes(x = Age, y = mean_failure, color = Gender)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1), limits = c(0, 1)) +
  scale_color_manual(values = c("Male" = "#1f77b4", "Female" = "#e15759")) +
  labs(
    title = "Failure Rates by Age and Gender",
    x = "Age",
    y = "Average Failure Rate"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    legend.title = element_blank()
  )
```



Line plot shows how average failure rates (Students who either withdrew or did not complete a course) vary by age and gender. By graph failure rates generally increase with age, especially among male students. Both genders start with relatively low failure rates around ages 12–13, a clear gap appears by age 17–18.

Female students show more stable failure rates, mostly between 25% and 40% across ages. Male students show greater fluctuations, with sharp spikes at ages 18 and 21 where the failure rate reaches 100%.

Overall, the data suggests that older male students tend to have more difficulties completing courses, while female students show more consistent outcomes across age groups. However, these trends could also be influenced by sample size or participation differences at each age level.