

Empirical Social Research - Final Assignment

Raphael Reichel

2024-07-02

This essay is aimed at **Question Q4** (Data Visualisation) of the final assignment.

The source files of this essay are available at <https://github.com/dir-rafa/WU-Assignment-ESR-2024-Academic-Gender-Gap>

1 Academic Gender Gap

The main variable of focus in this essay is the **academic gender gap** in degree programs in Austria. This is understood as an uneven distribution of places in degree programs at universities between female and male students (no more genders are available within the available data set). The gap emerges if either more female or male students are enrolled in a specific program. The assumption that equal participation of genders would mean 50% female and 50% male students is made since roughly 50% of the overall Austrian population is female (Statistik Austria).

1.1 Methodology

For investigating the academic gender gap we will focus on comparative statistics. The used data set was created by Statistik Austria (Statistik Austria, 2024) and downloaded via their database access [STATCube](#).

In a first step the overall gender distribution will be explored across all studies and fields. Afterwards, the composition of the total population of students will be assessed by splitting them by their degree program. In the final section those dimensions will then be combined and the gender gap will be determined for each degree program.

1.2 The Data Set

The aforementioned data set holds information about the amount of students enrolled in types of academic degree programs (bachelor studies, masters studies, doctoral studies or diploma studies) and the academic field the degree is associated with (science or arts). The data is further divided into sex (female or male) and sorted by semester. The overall available variables therefore are as follows:

- **Semester:** Represents time periods - e.g. “summer semester 2021”
- **Sex:** Shared gender of a subgroup of students - “female” or “male”
- **Field:** A broad differentiation between categories of academic disciplines - “University of the sciences” or “University of the arts”
- **Studies:** The level of education - “Bachelor studies”, “Master studies”, “Doctoral studies”, or “Diploma studies”

1.3 Visualisation

To understand the development of the academic gender gap over time, line graphs will be used for analysing the different specified dimensions of academic education. Moreover, mostly weighted regression lines will be used for displaying the trend instead of raw data lines, since academic years are split into two semester and typically more new students register for the winter semester, which leads to a fluctuating levels of students

in between two academic years. To avoid overplotting this simplification will be employed for the first two sections of this essay. In the last section a different simplification method will be used.

1.4 Data Preperation

This page holds the code that is used for preparing the data set.

```
# Function to convert semester strings to startdate of semester
semester_to_startdate <- function(semester) {
  if (grepl("winter semester", semester)) {
    year <- as.numeric(sub(".*([0-9]{4})/([0-9]{2}).*", "\\1", semester))
    return(as.Date(paste(year, "-10-01", sep="")))
  } else if (grepl("summer semester", semester)) {
    year <- as.numeric(sub(".*([0-9]{4}).*", "\\1", semester))
    return(as.Date(paste(year, "-03-01", sep="")))
  } else {
    stop("Unknown semester format")
  }
}

# Function to convert semester strings to academic year
semester_to_academic_year <- function(semester) {
  if (grepl("winter semester", semester)) {
    year_start <- as.character(sub(".*([0-9]{4})/([0-9]{2}).*", "\\1", semester))
    year_end <- as.character(sub(".*([0-9]{4})/([0-9]{2}).*", "\\2", semester))
    return(paste(year_start, "/", year_end, sep=""))
  } else if (grepl("summer semester", semester)) {
    year <- as.numeric(sub(".*([0-9]{4}).*", "\\1", semester))
    return(paste(year - 1, "/", sub("[0-9]{2}([0-9]{2})", "\\1", year), sep=""))
  } else {
    stop("Unknown semester format")
  }
}

read_csv( # read the csv file
  "data/table_austria_public_university_studies.csv",
  skip = 5,
  show_col_types = FALSE
) %>%
slice_head(n = -10) %>% # remove trailing comments/information
rename(
  Field = "University (Ebene +1)",
  Study = "Kind of study"
) %>%
select(Semester, Sex, Field, Study, Number) %>% # filter collumms
mutate( # convert strings to r data formats
  Sex = factor(Sex),
  Field = factor(Field),
  Study = factor(Study),
  Number = as.integer(round(Number))
) %>%
rowwise() %>%
mutate(
  AcademicYear = semester_to_academic_year(Semester),
  SemesterStart = semester_to_startdate(Semester)
) %>%
ungroup() %>%
relocate(AcademicYear, SemesterStart, .before = Semester) -> df #save data to df
```

1.5 Sex of the general Student Population

In Figure 1 the general population of students across all studies and fields is separated by their sex.

```
df %>%  
  summarise(Number = sum(Number), .by = c(SemesterStart, Sex)) %>%  
  ggplot(aes(SemesterStart, Number)) +  
  geom_point(aes(color = Sex)) +  
  geom_smooth(aes(color = Sex), method = "loess", formula = y ~ x) +  
  scale_y_continuous(labels = unit_format(unit = "k", scale = 1e-3)) +  
  labs(  
    title = "Student sex across all Studies and Fields",  
    x = "Semesters", y = "Students (in thousand)"  
  ) +  
  theme_R()
```

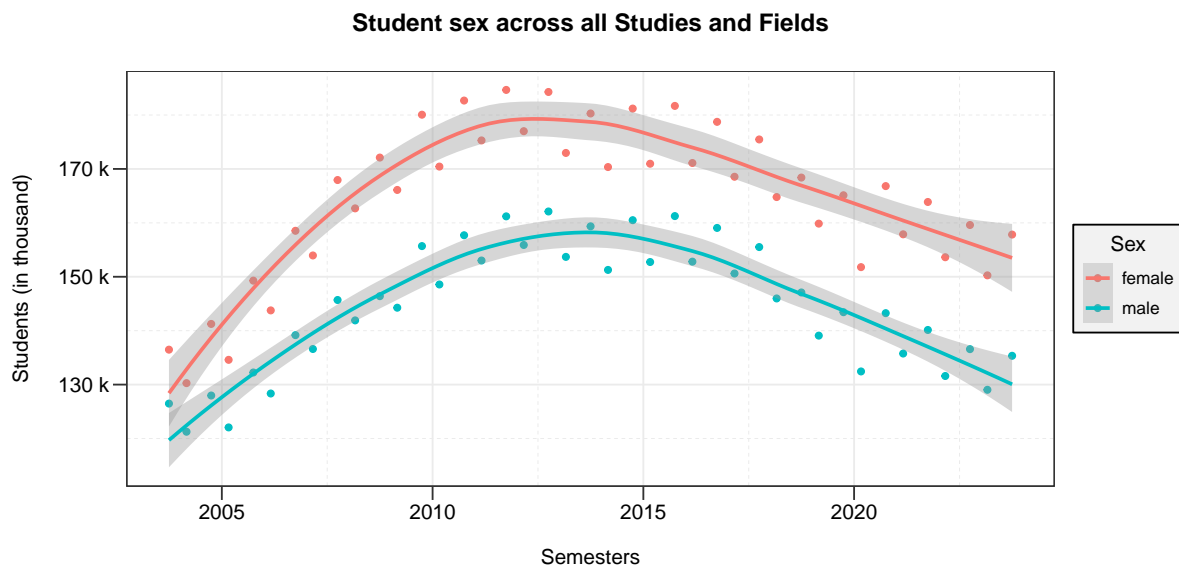


Figure 1: Students split by their sex with regression line.

The chart shows that overall more female students are enrolled in academic degree programs than male students.

In Figure 2 this general student population is further split into the field of study they are enrolled in.

```
df %>%
  summarise(Number = sum(Number), .by = c(SemesterStart, Sex, Field)) %>%
  ggplot(aes(SemesterStart, Number)) +
  geom_point(aes(color = Sex)) +
  geom_smooth(aes(color = Sex), method = "loess", formula = y ~ x) +
  scale_y_continuous(labels = unit_format(unit = "k", scale = 1e-3)) +
  labs(
    title = "Student split by Sex and Fields",
    x = "Semesters", y = "Students (in thousand)"
  ) +
  facet_wrap(vars(Field), dir = "h", scales = "free_y") +
  theme_R()
```

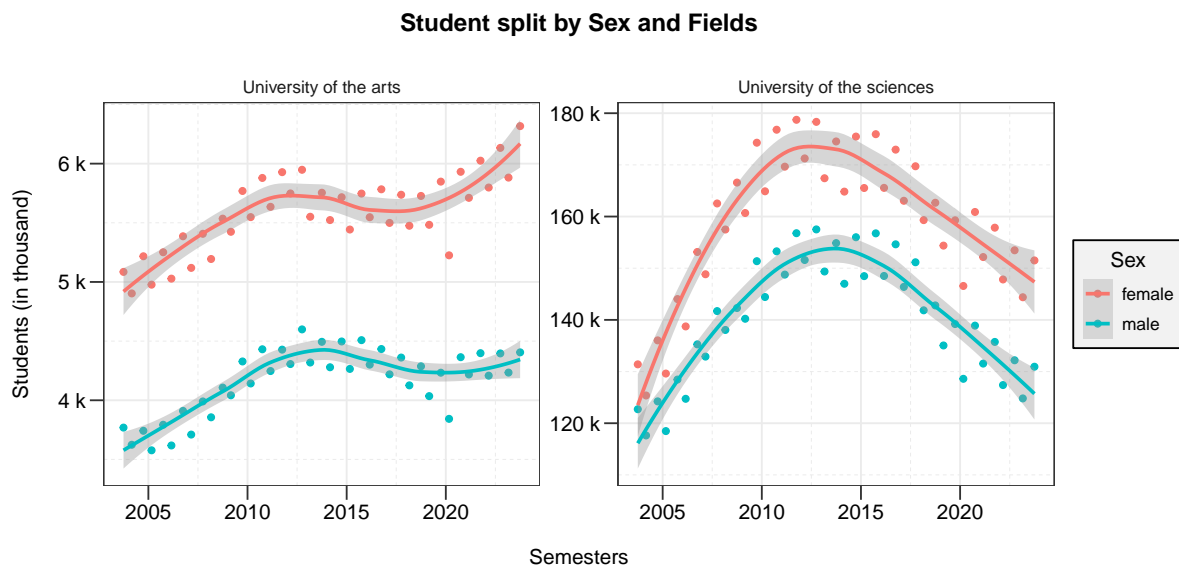


Figure 2: Students split by their sex and field with regression line.

Although, female students are dominant in both arts and sciences, the gap between the genders is even wider in arts. Those two graphs are important for viewing the later graphs about gender.

1.6 Students split by their Studies

To further specify the gap shown in the last two charts, we now will take a look at the study area of students. In Figure 3 the overall student population is split by their type of study.

```
df %>%  
  summarise(Number = sum(Number), .by = c(SemesterStart, Study)) %>%  
ggplot(aes(SemesterStart, Number)) +  
  geom_point(aes(color = Study)) +  
  geom_smooth(aes(color = Study), method = "loess", formula = y ~ x) +  
  scale_y_continuous(labels = unit_format(unit = "k", scale = 1e-3)) +  
  labs(  
    title = "Students across all Fields split by Studies",  
    x = "Semesters", y = "Students (in thousand)"  
  ) +  
  theme_R()
```

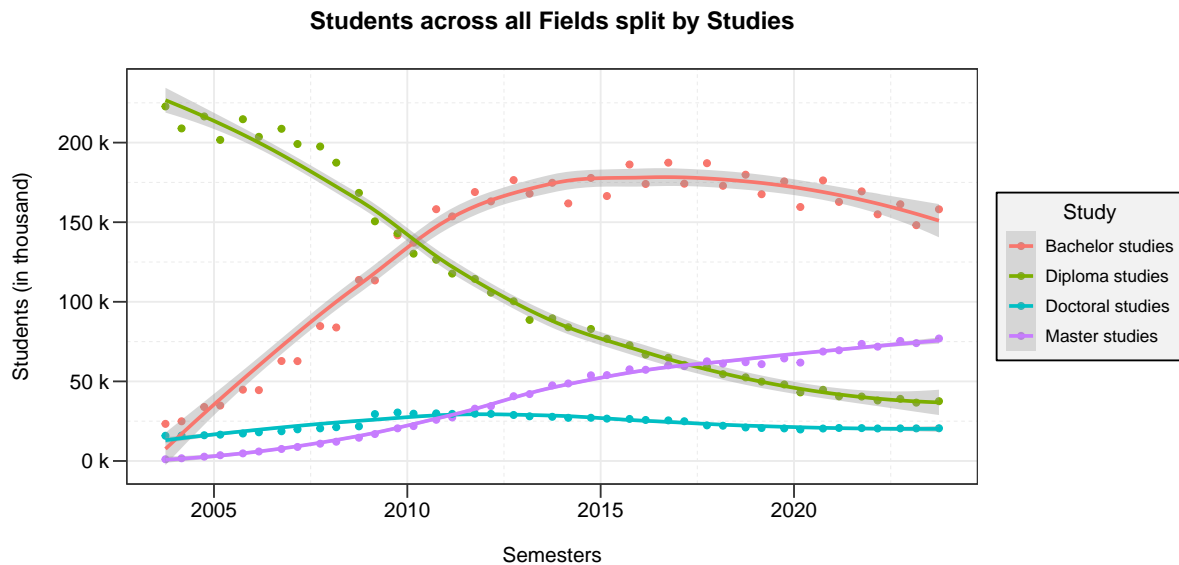


Figure 3: Total amount of students split by their studie.

This shows that most students are enrolled in a bachelor program, which drastically gained popularity over the course of the last 20 years, while diploma studies decreased simultaneously. Master degrees steadily gained popularity too, while doctoral studies remained even in comparison to the others.

This can be again drilled down by the field of studies:

```
df %>%
  summarise(Number = sum(Number), .by = c(SemesterStart, Study, Field)) %>%
  ggplot(aes(SemesterStart, Number)) +
  geom_point(aes(color = Study)) +
  geom_smooth(aes(color = Study), method = "loess", formula = y ~ x) +
  facet_wrap(vars(Field), dir = "h", scales = "free_y") +
  scale_y_continuous(labels = unit_format(unit = "k", scale = 1e-3)) +
  theme_R() +
  labs(
    title = "Students split by Studies and Fields",
    x = "Semesters", y = "Students (in thousand)"
  ) +
  theme(legend.position = "bottom")
```

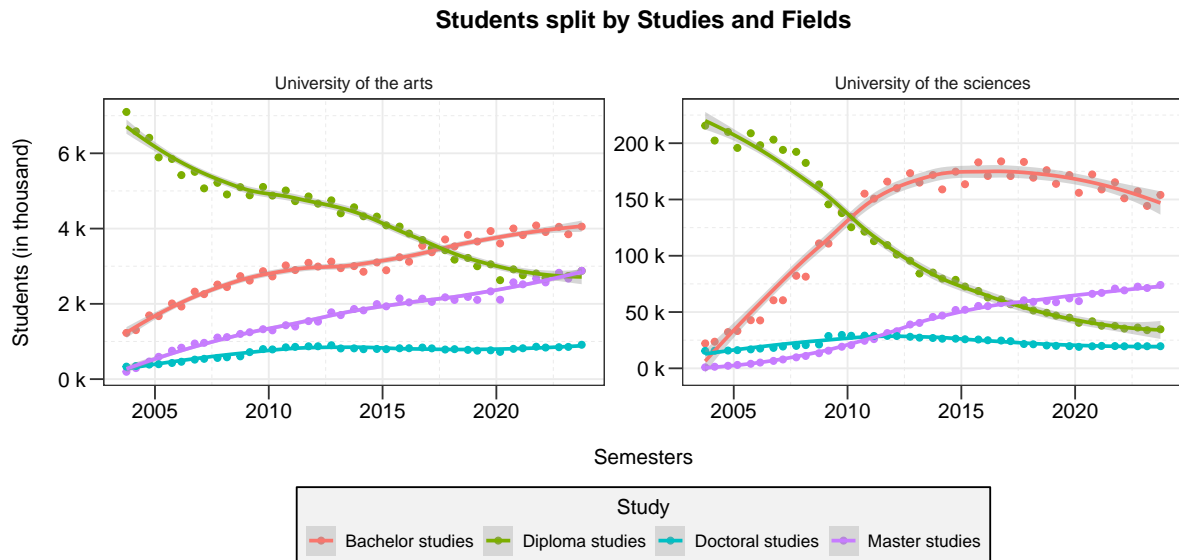


Figure 4: Students split by their studie and field.

While the aforementioned development holds true for the study programs of universities of science, the university of arts observed slightly different trends in the last 20 years. Bachelor and master studies develop parallel to another. However, the overall trends follow a similar direction

1.7 Investigating the Gender Gap

After gaining insight on the overall gender distribution in the population of students and the distribution of students across their study programs, those dimensions can be combined to investigate the distinct gender gap in studies. In Figure 5 the gender gap can be viewed in isolation and absolute numbers for each study and field:

```
ggplot(df, aes(SemesterStart, Number)) +
  geom_smooth(aes(linetype = Sex), method = "loess", formula = y ~ x) +
  facet_wrap(Study ~ Field, ncol = 2, scales = "free_y") +
  scale_y_continuous(labels = unit_format(unit = "k", scale = 1e-3)) +
  labs(
    title = "Students split by Sex, Studies, and Fields",
    x = "Semesters", y = "Students (in thousand)"
  ) +
  theme_R() +
  theme(strip.background = element_rect(fill = "gray90"))
```

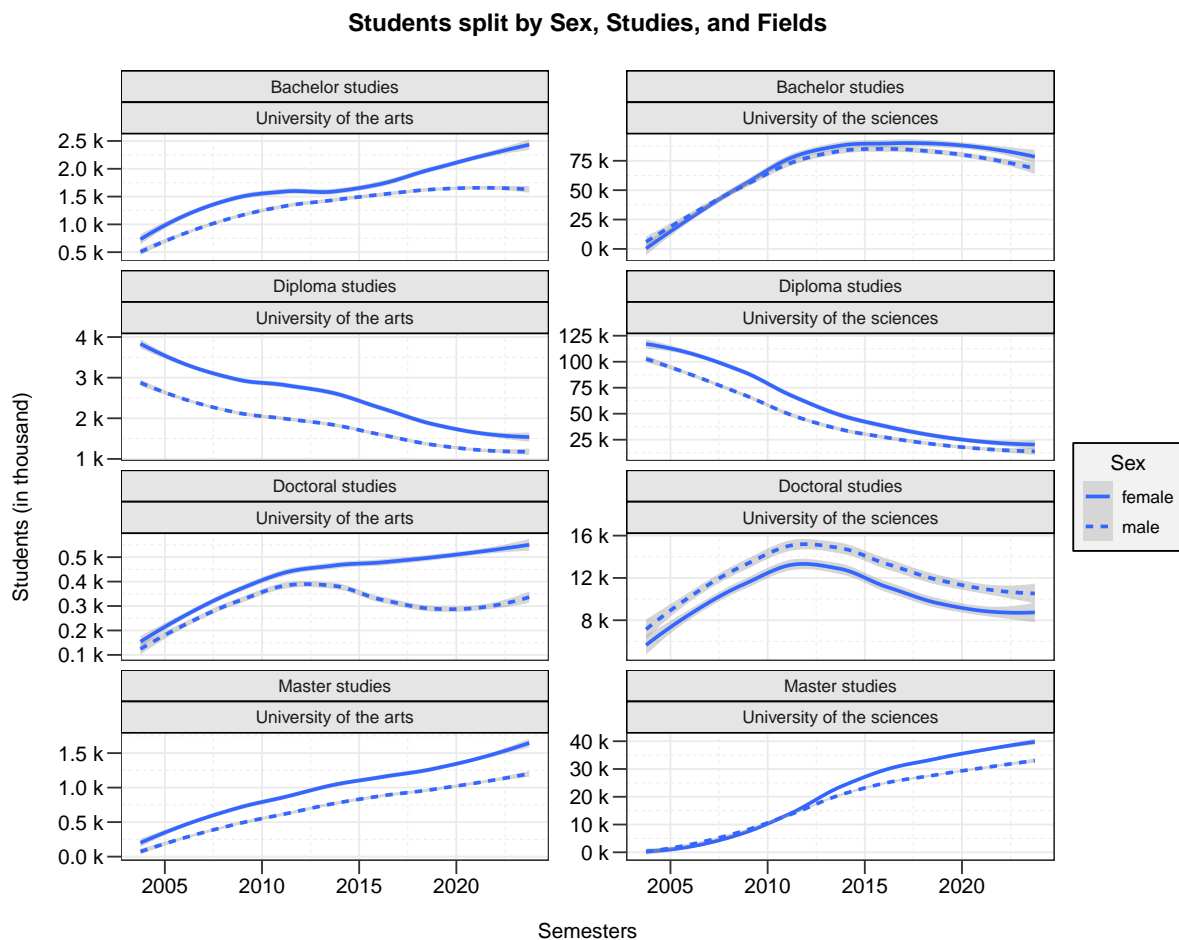


Figure 5: Isolated gender gap of individual studies split by field.

Although this representation gives insight on the gender gap, it is hard to visualise the difference in relation to the individual studies. Therefore, Figure 6 uses the ratio of female and male students instead of absolute numbers. The ratio R is calculated as follows:

$$R = S_f / S_m$$

S_f being the amount of female students in a study

S_m being the amount of male students in a study

Lines above the 1.0 mark sign that female students are more commonly enrolled in the specific study than male students. Lines below the 1.0 mark sign the opposite.

```
df %>%
  select(-Field) %>%
  summarise(Number = sum(Number), .by = c(SemesterStart, Sex, Study)) %>%
  pivot_wider(id_cols = c(SemesterStart, Study), names_from = Sex, values_from = Number) %>%
  mutate(FM_Ratio = female / male) %>%
ggplot(aes(SemesterStart, FM_Ratio)) +
  geom_line(aes(color = Study)) +
  geom_point(aes(shape = Study, color = Study)) +
  geom_hline(aes(yintercept = 1), linetype = "longdash") +
  scale_y_continuous(breaks = seq(from = 0, to = 1.5, by = 0.2)) +
  labs(
    title = "Students female-male Ratio by Studies",
    x = "Semesters", y = "Ratio female-male students"
  ) +
  theme_R()
```

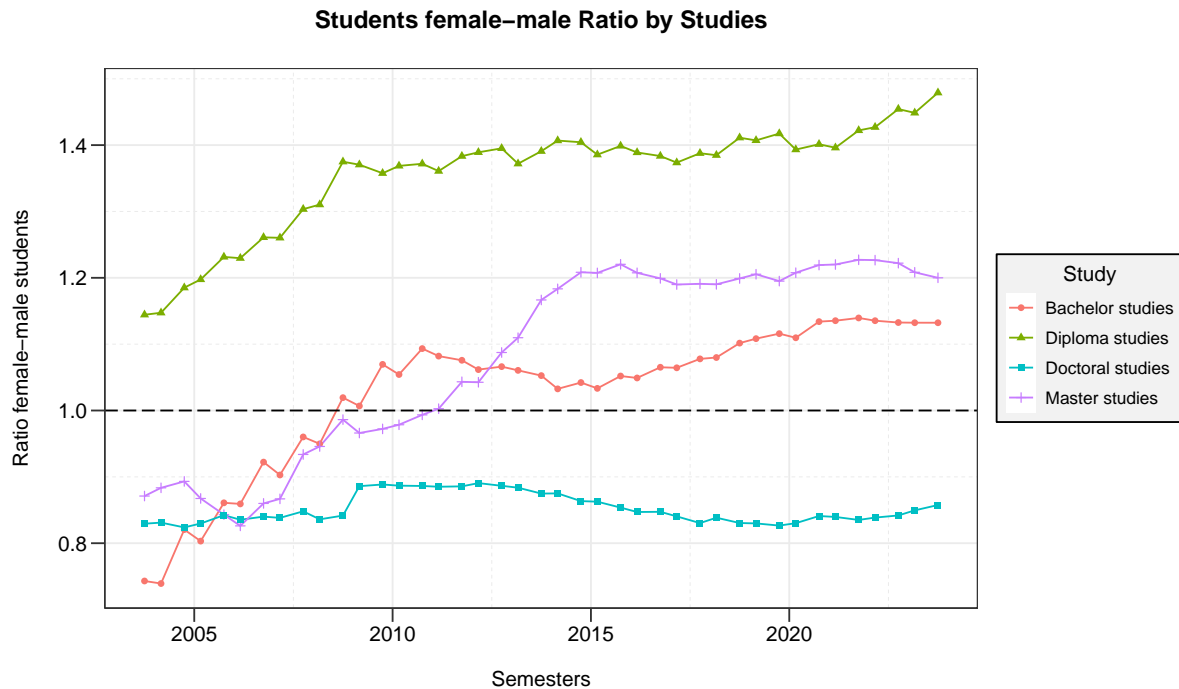


Figure 6: Gender gap in studies, visualised by ratio of female to male students.

This shows that in bachelor, diploma, and master studies, female students are overrepresented in comparison

to male students. This fits in line with the distribution of gender in the overall population of students that was shown in Figure 1. However, in doctoral studies men are overrepresented, although they are the minority group in the overall population of students.

This representation is once again split by study field in Figure 7:

```
df %>%
  summarise(Number = sum(Number), .by = c(SemesterStart, Sex, Study, Field)) %>%
  pivot_wider(id_cols = c(SemesterStart, Study, Field), names_from = Sex, values_from = Number) %>%
  mutate(FM_Ratio = female / male) %>%
  ggplot(aes(SemesterStart, FM_Ratio)) +
  geom_line(aes(color = Study)) +
  geom_point(aes(shape = Study, color = Study)) +
  geom_hline(aes(yintercept = 1), linetype = "longdash") +
  facet_wrap(vars(Field), dir = "h", scales = "free_y") +
  scale_y_continuous(breaks = seq(from = 0, to = 4, by = 0.25)) +
  labs(
    title = "Students female-male Ratio by Studies and Fields",
    x = "Semesters", y = "Ratio female-male students"
  ) +
  theme_R() +
  theme(legend.position = "bottom")
```

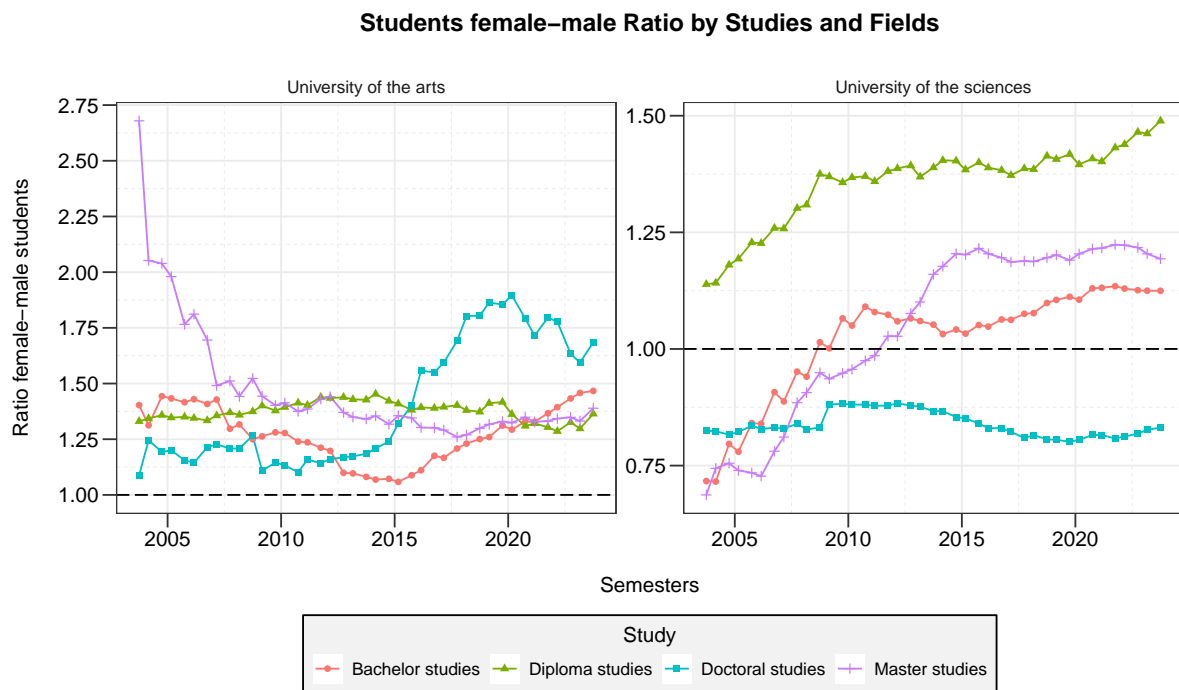


Figure 7: Gender gap in studies split by field, visualised by the ratio of female to male students.

For the studies of the science field nothing dramatically changes, but studies of arts show that females are overrepresented in the doctoral studies too. Moreover, doctoral studies appear to be more popular in studies of arts.

1.8 Related Work

Francesconi & Parey (2018) used a similar visualisation technique in their article “Early gender gaps among university graduates” for universities in Germany. While they mainly focused on graduates and processed data about the labor market too, their analysis follows a similar path to this essay.

They started by plotting the gender gap of the overall population of university students and then drilled down the gap by splitting their data into subjects (e.g. Art, Engineering, Medicine, etc.) and proceeded by comparing those gender gaps by calculating the female-male ratio of students.

In their evaluation they found that more female students graduate university, while enrollment count is roughly equal between genders (Francesconi & Parey, 2018). This appears not to be the case for Austrian universities, as shown in this essay.

2 References

- Francesconi, M., & Parey, M. (2018). Early gender gaps among university graduates. *European Economic Review*, 109, 63–82. <https://doi.org/https://doi.org/10.1016/j.euroecorev.2018.02.004>
- Statistik Austria. *University students, studies*. https://www.statistik.at/fileadmin/pages/358/Infotext_GeunderStatistik_DemographischeMerkmale.pdf
- Statistik Austria. (2024). *University students, studies*. <https://www.statistik.at/en/statistics/population-and-society/education/university-students-studies>