Gender Pay Inequality in the EU: Sectoral and Regional Analysis

Jana Fľaková 2856763, Aliya Ismayilova 2845411, Vasiliki Tyrogianni 2819431, Leyan Shen 2801691, Elo

2025-06-23

Names: Jana Fľaková, Leyan Shen, Aliya Ismayilova, Vasiliki Tyrogianni, Eloy Puga, Sofia Christodoulou

Tutorial Group: 2 tutorial 3 group

Lecturers: Chantal Schouwenaar, Jack Frederick Fitzgerald

Part 1 - Identify a Social Problem

1.1 Describe the Social Problem

This study report analyzes the gender pay inequality in Europe, with focus on a sectoral structure of Germany as a subgroup. To do so, we estimate average male and female incomes per sector that adjust for female employment share, male employment share, and wage gap.

Relevance of the issue

The gender pay gap remains a significant social issue in Europe today. The gender pay refers to the difference in male and female earnings. Although there has been progress in promoting gender equality, women in the European Union still earn less than men. This inequality has an impact on women's economic independence as lower earnings reduce women's financial stability and independence. Moreover, it also increases the risk of poverty, particularly for elderly women and single mothers. Hence, addressing the gender pay gap issue is essential because equal pay will benefit the whole economy as it increases consumer spending power and promotes social justice. Therefore, our purpose is to analyze the gender pay gap in Europe for the years 2021, 2022, and 2023 to analyze the way it changed over time.

Part 2 - Data Sourcing

2.1 Load in the Data

```
# Read CSVs from project GitHub repo structure
TableCountries = read_csv("data/TableCountries.csv")
Germany_subgroup = read_csv("data/Germany_subgroup/Germany_subgroup.csv")
```

2.2 Summary of the Dataset

```r
head(TableCountries)
```

```
## # A tibble: 6 x 7
##    ...1 country  year gender_pay_gap monthly_income p_female p_male
##   <dbl> <chr>   <dbl>          <dbl>          <dbl>    <dbl>  <dbl>
## 1     1 Austria  2021           19.1          4014.    0.410  0.590
## 2     2 Austria  2022           18.7          4228.    0.414  0.586
## 3     3 Austria  2023           18.3          4542.    0.412  0.588
## 4     4 Belgium  2021            1.8          4141.    0.429  0.571
## 5     5 Belgium  2022            0.7          4452.    0.434  0.566
## 6     6 Belgium  2023            0.7          4832.    0.432  0.568
```

```r
head(Germany_subgroup)
```

```
## # A tibble: 6 x 6
##   sector                 income_eur gpg_2023 occupation_category p_female p_male
##   <chr>                       <dbl>    <dbl> <chr>                  <dbl>  <dbl>
## 1 Financial and insuran~       5841       26 Professionals          0.501  0.499
## 2 Information and commu~       5769       21 Professionals          0.501  0.499
## 3 Professional, scienti~       5436       26 Professionals          0.501  0.499
## 4 Electricity, gas, ste~       5352       14 Technicians and as~    0.537  0.463
## 5 Education                    4733        9 Professionals          0.501  0.499
## 6 Mining and quarrying         4544        2 Craft and related ~    0.102  0.898
```

The EU-wide data set includes country-level variables such as the gender pay gap, average monthly income, and male and female share in workforce for 2021–2023. The Germany subgroup includes sector-level earnings, gender pay gaps, and gender share in workforce.

2.3 Describe the Type of Variables

The datasets are compiled from administrative sources such as Eurostat and Destatis. Variables include:

- `gender_pay_gap`: Percentage difference in male vs female income
- `monthly_income`: Average monthly gross pay (€)
- `p_female` / `p_male`: Employment shares by gender
- `income_eur`: Average monthly income by sector in Germany
- `gpg_2023`: Sector-specific gender pay gap

Part 3 - Quantifying

3.1 Data Cleaning

```r
TableCountries$year = as.factor(TableCountries$year)
Germany_subgroup$income_quartile = cut(Germany_subgroup$income_eur,
  breaks = quantile(Germany_subgroup$income_eur, probs = seq(0, 1, 0.25), na.rm = TRUE),
  include.lowest = TRUE,
  labels = c("Q1 (lowest)", "Q2", "Q3", "Q4 (highest)"))
```

3.2 Generate Necessary Variables

To estimate male and female income levels based on the reported average income and gender pay gap, we apply the following logic. The overall average income in each country or sector is calculated as a weighted sum:

$$\text{Average Income} = p_{\text{male}} \times \text{Male Income} + p_{\text{female}} \times \text{Female Income}$$

The unadjusted gender pay gap (GPG) expresses the percentage by which female income falls short of male income:

$$\text{Female Income} = (1 - \text{GPG}) \times \text{Male Income}$$

Substituting this into the first equation allows solving for male income:

$$\text{Average Income} = p_{\text{male}} \times \text{Male Income} + p_{\text{female}} \times (1 - \text{GPG}) \times \text{Male Income}$$

$$\text{Average Income} = \text{Male Income} \times (p_{\text{male}} + p_{\text{female}} \times (1 - \text{GPG}))$$

Thus, male income can be isolated as:

$$\text{Male Income} = \frac{\text{Average Income}}{p_{\text{male}} + p_{\text{female}} \times (1 - \text{GPG})}$$

Once male income is known, female income follows directly as:

$$\text{Female Income} = (1 - \text{GPG}) \times \text{Male Income}$$

```r
# EU country-level income estimation
male_income = numeric(nrow(TableCountries))
female_income = numeric(nrow(TableCountries))
for (i in 1:nrow(TableCountries)) {
  income = TableCountries$monthly_income[i]
  gpg = TableCountries$gender_pay_gap[i] / 100
  p_f = TableCountries$p_female[i]
  p_m = TableCountries$p_male[i]

  if (any(is.na(c(income, gpg, p_f, p_m))) || (p_m + p_f * (1 - gpg) == 0)) {
    male_income[i] = NA
    female_income[i] = NA
  } else {
    male_income[i] = income / (p_m + p_f * (1 - gpg))
    female_income[i] = male_income[i] * (1 - gpg)
  }
}
TableCountries$male_income = round(male_income, 2)
TableCountries$female_income = round(female_income, 2)
```

```r
# Germany sectoral level
male_income = numeric(nrow(Germany_subgroup))
female_income = numeric(nrow(Germany_subgroup))
for (i in 1:nrow(Germany_subgroup)) {
  income = Germany_subgroup$income_eur[i]
  gpg = Germany_subgroup$gpg_2023[i] / 100
  p_f = Germany_subgroup$p_female[i]
  p_m = Germany_subgroup$p_male[i]
  male_income[i] = income / (p_m + p_f * (1 - gpg))
  female_income[i] = male_income[i] * (1 - gpg)
}
Germany_subgroup$male_income = round(male_income, 2)
Germany_subgroup$female_income = round(female_income, 2)
```

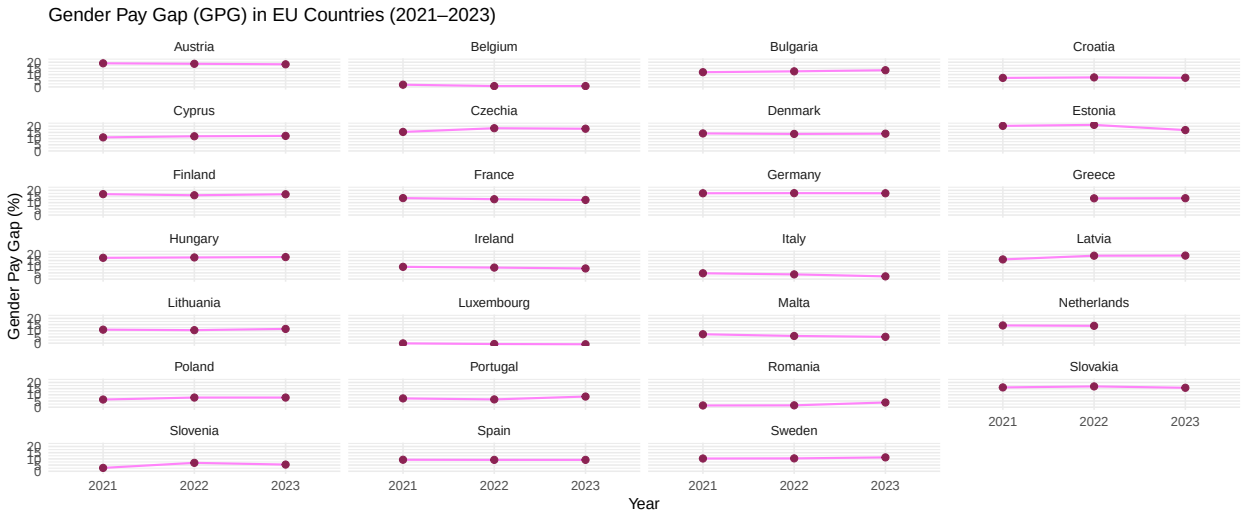
3.3 Visualize Temporal Variation

```r
eu_countries = c("Austria", "Belgium", "Bulgaria", "Croatia", "Cyprus", "Czechia", "Denmark", "Estonia"

TableCountries_filtered = TableCountries %>%
  filter(country %in% eu_countries) %>%
  arrange(factor(country, levels = eu_countries))

ggplot(TableCountries_filtered, aes(x = year, y = gender_pay_gap, group = 1)) +
  geom_line(color = "orchid1", linewidth = 0.7) +
  geom_point(color = "violetred4", size = 2) +
  facet_wrap(~ country, ncol = 4, nrow = 7) +
  scale_y_continuous(limits = c(-1, 22)) +
  labs(
    title = "Gender Pay Gap (GPG) in EU Countries (2021-2023)",
    x = "Year", y = "Gender Pay Gap (%)"
  ) +
  theme_minimal(base_size = 11)
```

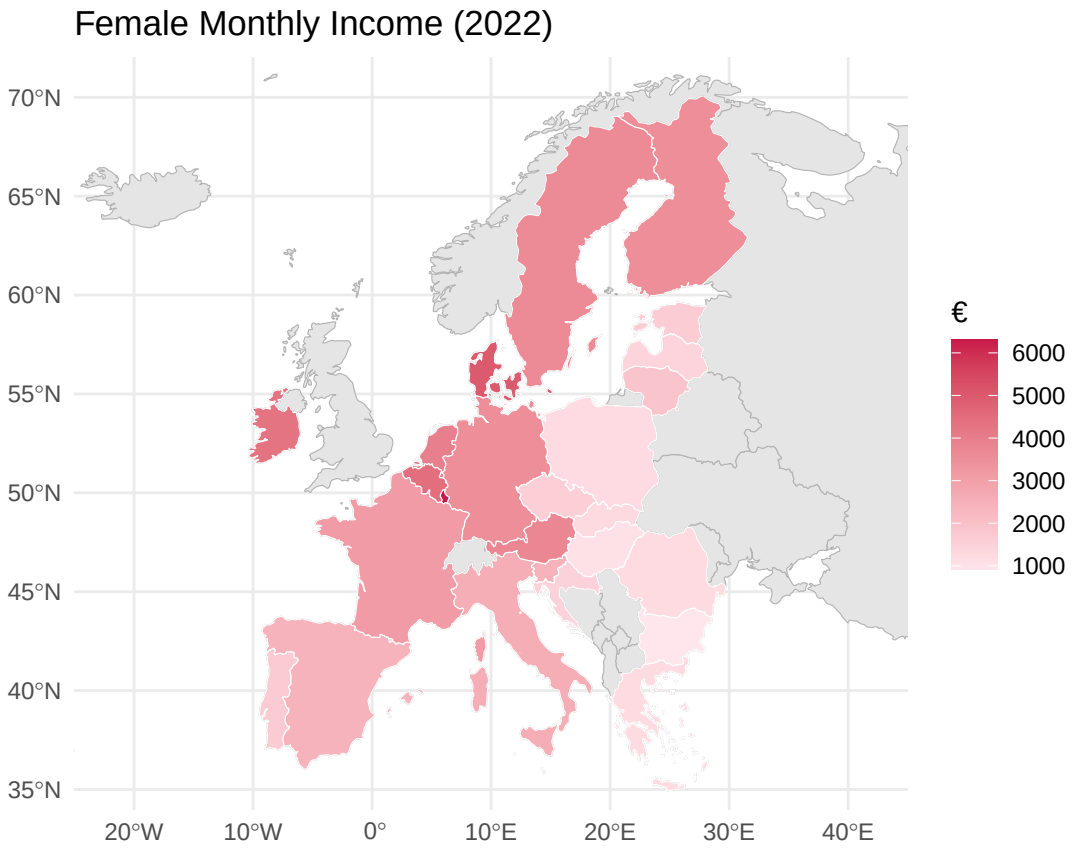
```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



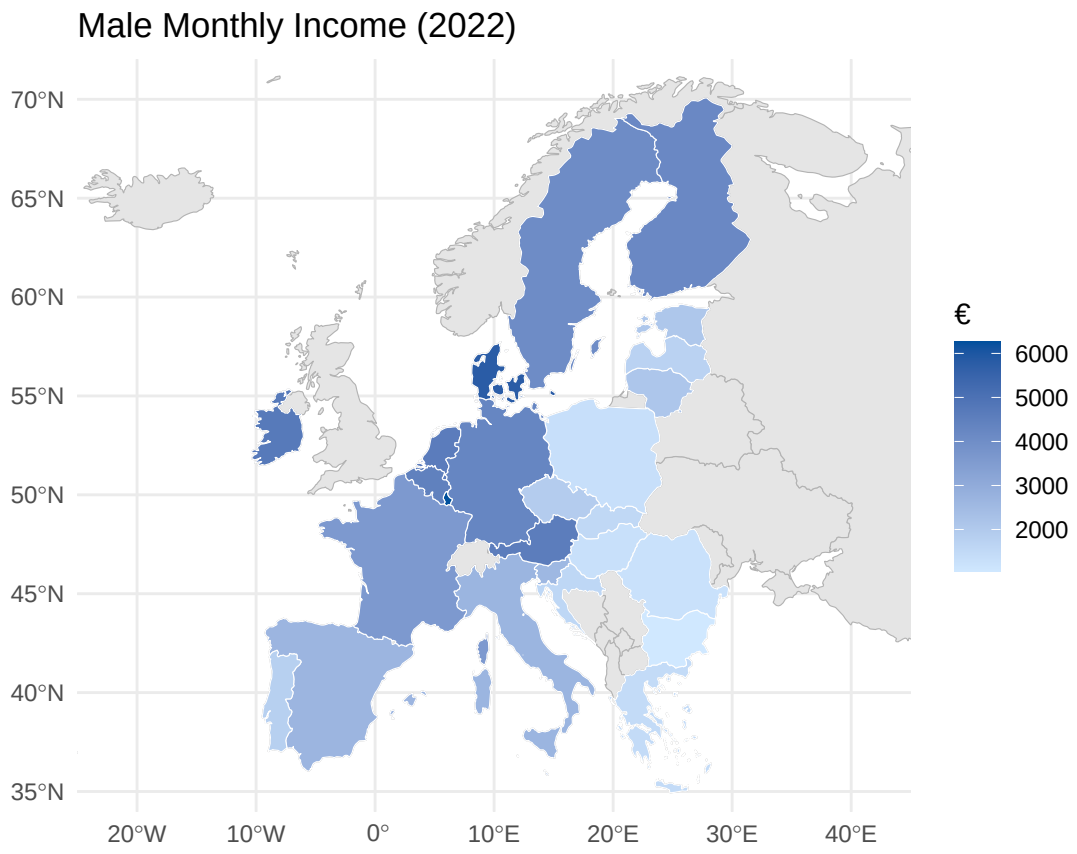
3.4 Visualize Spatial Variation

```
all_europe = rnaturalearth::ne_countries(scale = "medium", returnclass = "sf") %>%
  filter(region_un == "Europe")
eu_data_2022 = TableCountries %>% filter(year == 2022)
europe_map = all_europe %>%
  left_join(eu_data_2022, by = c("name" = "country")) %>%
  filter(!is.na(monthly_income))

ggplot() +
  geom_sf(data = all_europe, fill = "grey90", color = "grey70", size = 0.2) +
  geom_sf(data = europe_map, aes(fill = female_income), color = "white", size = 0.3) +
  scale_fill_gradient(low = "#FFE5EC", high = "#C9184A", name = "€") +
  labs(title = "Female Monthly Income (2022)") +
  coord_sf(xlim = c(-25, 45), ylim = c(34, 72), expand = FALSE) +
  theme_minimal()
```




```
ggplot() +
  geom_sf(data = all_europe, fill = "grey90", color = "grey70", size = 0.2) +
  geom_sf(data = europe_map, aes(fill = male_income), color = "white", size = 0.3) +
  scale_fill_gradient(low = "#D0E8FF", high = "#00509D", name = "€") +
  labs(title = "Male Monthly Income (2022)") +
  coord_sf(xlim = c(-25, 45), ylim = c(34, 72), expand = FALSE) +
  theme_minimal()
```



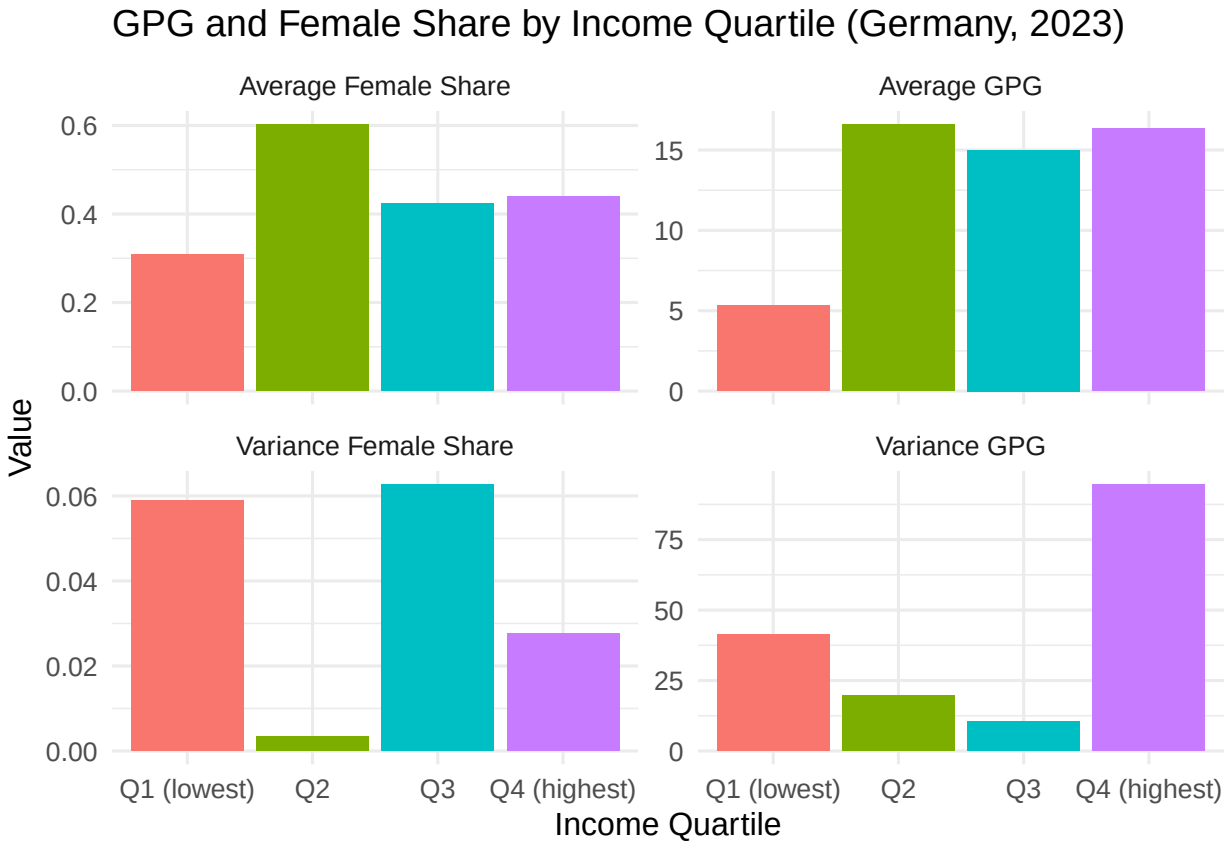
3.5 Visualize Sub-Population Variation

```
quartile_summary = Germany_subgroup %>%
  group_by(income_quartile) %>%
  summarise(
    avg_income = mean(income_eur),
    avg_gpg = mean(gpg_2023),
    var_gpg = var(gpg_2023),
    avg_p_female = mean(p_female),
    var_p_female = var(p_female),
    n_sectors = n()
  )

quartile_long = quartile_summary %>%
  select(income_quartile, avg_gpg, var_gpg, avg_p_female, var_p_female) %>%
  pivot_longer(cols = -income_quartile, names_to = "metric", values_to = "value") %>%
  mutate(metric = recode(metric,
    avg_gpg = "Average GPG",
    var_gpg = "Variance GPG",
    avg_p_female = "Average Female Share",
    var_p_female = "Variance Female Share"
  ))

ggplot(quartile_long, aes(x = income_quartile, y = value, fill = income_quartile)) +
```

```
  geom_col(show.legend = FALSE) +
  facet_wrap(~ metric, ncol = 2, scales = "free_y") +
  labs(
    title = "GPG and Female Share by Income Quartile (Germany, 2023)",
    x = "Income Quartile", y = "Value"
  ) +
  theme_minimal(base_size = 12)
```



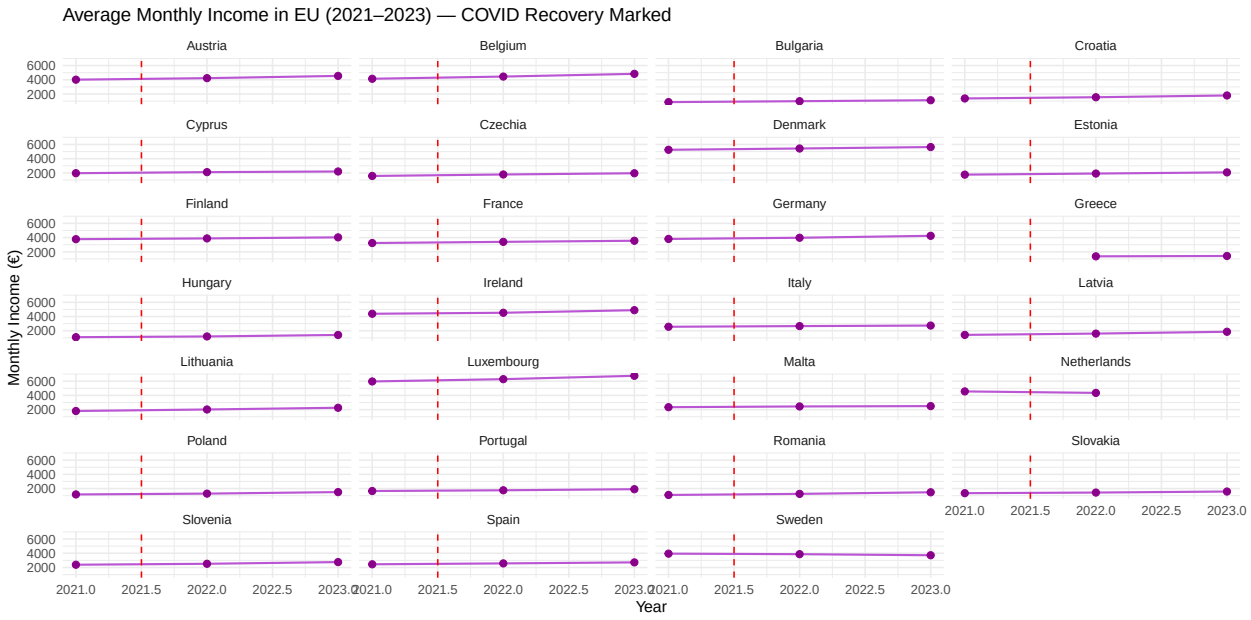
3.6 Event Analysis

```
filtered_data = TableCountries %>%
  filter(country %in% eu_countries) %>%
  mutate(year_numeric = as.numeric(as.character(year)))

ggplot(filtered_data, aes(x = year_numeric, y = monthly_income, group = 1)) +
  geom_line(color = "mediumorchid", linewidth = 0.7) +
  geom_point(color = "darkmagenta", size = 2) +
  geom_vline(xintercept = 2021.5, linetype = "dashed", color = "red", linewidth = 0.5) +
  facet_wrap(~ country, ncol = 4, nrow = 7) +
  labs(
    title = "Average Monthly Income in EU (2021-2023) - COVID Recovery Marked",
    x = "Year", y = "Monthly Income (€)"
```

```
  ) +
  theme_minimal(base_size = 11)
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Part 4 - Discussion

4.1 Discuss Your Findings

Our analysis reveals persistent income and gender inequalities across the European Union, both regionally and within sectors. Despite ongoing policy efforts, the Gender Pay Gap (GPG) has remained largely stable between 2021 and 2023, with minimal progress in closing the earnings gap between men and women. Countries such as Germany, Austria, and Finland continue to exhibit some of the highest GPG levels, particularly within high-income economies.

The COVID-19 pandemic acted as a clear economic disruptor, especially in sectors like retail, hospitality, and education (European Parliament, 2021). Following widespread lockdowns and business closures in 2020, most countries experienced stagnant wages in 2021. By 2022 and 2023, as economies reopened, average monthly incomes recovered. However, this recovery was uneven: Western and Northern countries maintained the highest incomes but saw slower relative growth, while Eastern and Southern European countries recorded lower incomes but occasionally larger percentage gains.

Spatial analysis of 2022 income data further highlights regional disparities. Countries like Luxembourg, Germany, and the Netherlands report significantly higher average wages than Bulgaria, Romania, or Portugal. Strong public sectors and wage regulation in some Western countries contribute to more stable income patterns, but high absolute wages do not guarantee gender equality.

A closer look at Germany's 2023 sectoral data exposes structural gender inequalities within the labor market. High-paying industries like finance, IT, and professional services have both low female representation and significant GPGs, whereas sectors like education and public administration show narrower gaps, partly due to standardized pay structures.

The relationship between income level and inequality is evident in the income quartile breakdown for Germany. Female representation is highest in mid-to-lower income sectors but significantly lower at both the top and bottom of the wage distribution. Meanwhile, the GPG is greatest in the highest income quartile, suggesting that even when women access high-paying roles, they still face considerable earnings disadvantages. Variance in GPG is also high at the top, pointing to structural barriers that persist even in elite sectors.

Overall, our findings indicate that simply increasing average wages is not sufficient to reduce income inequality or close the gender pay gap. Targeted policy interventions addressing occupational segregation, wage transparency, and sector-specific barriers are necessary to create more equitable labor markets across Europe.

Part 5 - Reproducibility

5.1 Github Repository Link

https://github.com/Janusenka/Programming-group-3

5.2 Reference List

- Average gross monthly earnings by sector Germany 2023. (n.d.). Statista. https://www.statista.com/statistics/1284149/average-gross-monthly-earnings-by-sector-germany/
- Boll, C., & Lagemann, A. (2019). The Gender Pay Gap in EU Countries – New Evidence Based on EU-SES 2014 Data. *Intereconomics*, 2019(2), 101–105. https://www.intereconomics.eu/contents/year/2019/number/2/article/the-gender-pay-gap-in-eu-countries-new-evidence-based-on-eu-ses-2014-data.html
- European Parliament. (2020, January 13). Understanding the gender pay gap: definition, facts and causes. Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20200109STO69925/understanding-the-gender-pay-gap-definition-facts-and-causes
- European Parliament. (2021). Impacts of the COVID-19 pandemic on EU Industries. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662903/IPOL_STU(2021)662903_EN.pdf
- Eurostat. (2019). Gender Pay Gap in Unadjusted Form. Europa.eu. https://ec.europa.eu/eurostat/databrowser/view/sdg_05_20/default/table?lang=en
- Gender Equality Index | 2021 | Work | BE | European Institute for Gender Equality. (2021). https://eige.europa.eu/gender-equality-index/2021/domain/work/BE
- Gender Equality Index | 2022 | Work | BE | European Institute for Gender Equality. (2022). https://eige.europa.eu/gender-equality-index/2022/domain/work/BE
- Gender Equality Index | 2023 | Work | BE | European Institute for Gender Equality. (2023). https://eige.europa.eu/gender-equality-index/2023/domain/work/BE
- Klein, F. B., Hill, A. D., Hammond, R., & Stice-Lusvardi, R. (2020). The gender equity gap: A multistudy investigation of within-job inequality in equity-based awards. *Journal of Applied Psychology*, 106(5). https://doi.org/10.1037/apl0000809
- Participation of women in economic activity by occupations. (2025, May 19). Federal Statistical Office. https://www.destatis.de/EN/Themes/Labour/Labour-Market/Quality-Employment/Dimension1/1_4_ParticipationWomenOccupation.html
- Unadjusted Gender Pay Gap (GPG) by economic sector. (2025, February 13). Federal Statistical Office. https://www.destatis.de/EN/Themes/Labour/Earnings/GenderPayGap/Tables/ugpg-03-by-economic-sector-at2014.html
- Understanding the gender pay gap: definition and causes - EU monitor. (2023, April 12). Www.eumonitor.eu. https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vl59mpupm0vq?ctx=vk4jic6t1dxz

- Eurostat. (2023). Employment by sector and full-time equivalents (custom dataset). https://ec.europa.eu/eurostat/databrowser/view/NAMA_10_FTE___custom_4232263/bookmark/table?lang=en&bookmarkId=fafb4e3b-f3aa-4907-9102-16be8df6f775
- Eurostat. (2024). Employment by sector and full-time equivalents (custom dataset). https://ec.europa.eu/eurostat/databrowser/view/nama_10_fte___custom_13597179/bookmark/table?lang=en&bookmarkId=c29eed24-377e-4763-aaf0-0419906d2ecd