

BUS 240 Information Visualization

Final Report

Group 5: Zhao, Victoria; Tang, Grace; Ge, Elva; Chen, Yushang

Table of Contents

Context / Domain / Market of Consultancy	3
Focus of the Consultancy	4
Data Quality Assurance	5
Correctness	5
Completeness	7
Data Visualizations and Discussion from Different Aspects	9
Argument 1: Does fertility affect female labor force participation rate?	9
Argument 2: Does public spending on family benefits encourage women's employment?	10
Argument 3: Do Women's weekly working hours affect female labor force participation rates in European countries?	12
Argument 4: Do different working fields affect the female labor force participation rate?	14
Argument 5: Does the female employment to population ratio affect the female labor force participation rate?	17
Interpretations, Insights, and Reflections	19
Insights	19
Reflection of the journey as a team	20
Additional Data Sources	21
References	22

Context / Domain / Market of Consultancy

Women's status in the area of employment and earnings has significantly improved since the last century. However, a gender gap in earnings persists across races, regions, and sectors of employment, and greatly expands with age. Gender equality is not a zero-sum game. Many workers will benefit from greater diversity in the workplace. According to a McKinsey Global Institute (MGI) report, \$12 trillion could be added to global GDP by 2025 if all countries match the progress toward gender parity of the fastest-improving country in their region (McKinsey Reports, 2015).

Companies with greater gender flexibility can add more competitive advantage in the industry, drive higher profitability, and better withstand unanticipated challenges. However, although many corporate managers now understand the importance of gender equality and inclusion if they want to improve business performances, only a few know the circumstances and what to do, even fewer have been successful in closing the gender gap. Our consulting firm focuses on gender equality in workforces and strives to provide a range of services to help our potential clients advance gender equality.

Our services will mainly focus on companies in European countries. In this paper, we will present some key facts and drivers of the level of gender equality in the European labor market in the past twenty years to provide some insights for our clients.

Focus of the Consultancy

Building gender-inclusive workplaces in the global labor force can improve outcomes for all levels of society and boost socio-economic progress. We help our clients to achieve these critical goals by working alongside clients to evaluate the organization's current level of gender equality at the first stage. After creating a roadmap to get started, we develop effective approaches to gender equality to help our clients to embrace a more inclusive future.

We are adept at both short-term consultancies (for instance, recruiting evaluation, anti-discrimination guidance, flexible working models) and longer-term support (for instance, providing an evaluation framework to guide the design and implementation of the gender equality monitoring and evaluation system) to help our clients.

Specifically, our approach will mainly focus on four stages to develop and accelerate gender equality and inclusion services (still will vary from client to client):

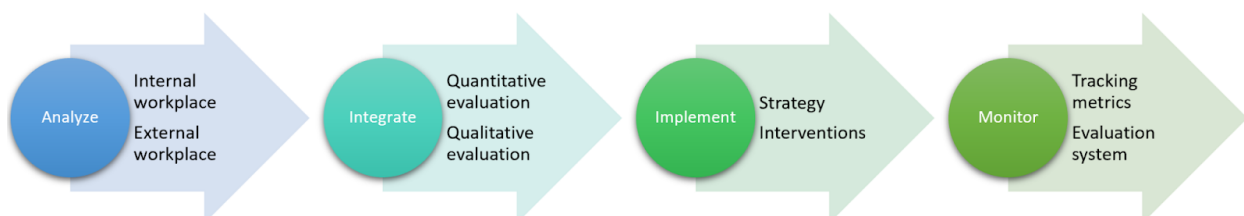
Stage 1: analyze gender equality and inclusion performance in the internal workplace and external environment (same region/same industry).

Stage 2: integrate both quantitative evaluation (collect and analyze internal data) and qualitative evaluation (create an internal survey: how employees assess company's diversity and inclusion performance) to understand the starting point.

Stage 3: identify strategic priorities and implement applicable interventions to help clients learn which shifts they need to make and where to focus investments.

Stage 4: build clear tracking metrics and evaluation system to monitor the progress toward gender equality.

By achieving these stages our advisory services can help clients to strike for gender equality



Data Quality Assurance

Correctness

In terms of correctness, we used 6 data sets from 4 credible resources to maintain the quality of our analysis. To generate the insights in this report, we researched, cleaned, and manipulated data from the following sources:

- [OECD Statistic:](#)
 - Public Spending on family Benefits
 - Female labor force participation OECD
 - Weekly Average Working hours

OECD Statistics was started with the name OEEC(Organization for European Economic Cooperation) in 1948, after World War II. Its goal was to help European governments recognize their economic interdependence. Canada and the United States joined the OEEC and created the OECD on December 14, 1960. Since OECD was one of the roots of the European Union, we cited their data and trust the quality of it.

- [International Labor Organization \(ILO\)](#)
 - Unemployment Male v.s Female
 - Female Labor Force by Age

ILO was growing out of labor and social movements which culminated in the demands for social justice and higher living standards for the working people over the world in the nineteenth century. As the first specialized agency associated with the United Nations from 1946, it has over 121 member countries. Since ILO is one of the most used resources of publications, we believe it is worthy to cite data from ILO.

- [The World Bank](#)
 - The Sectoral Composition of Female Employment (called “World Development Indicators” in WorldBank datasets)
 - Fertility (added since report 2)

The Work Bank Group was created in 1944. With the mission “To end extreme poverty and sustainably promote shared prosperity”, the World Bank collects and processes large amounts of data and gradually publicizes them. Since The World Bank database is widely considered to be the most consistent and reliable database, we believe the insights and analysis generated from their database are trustworthy.

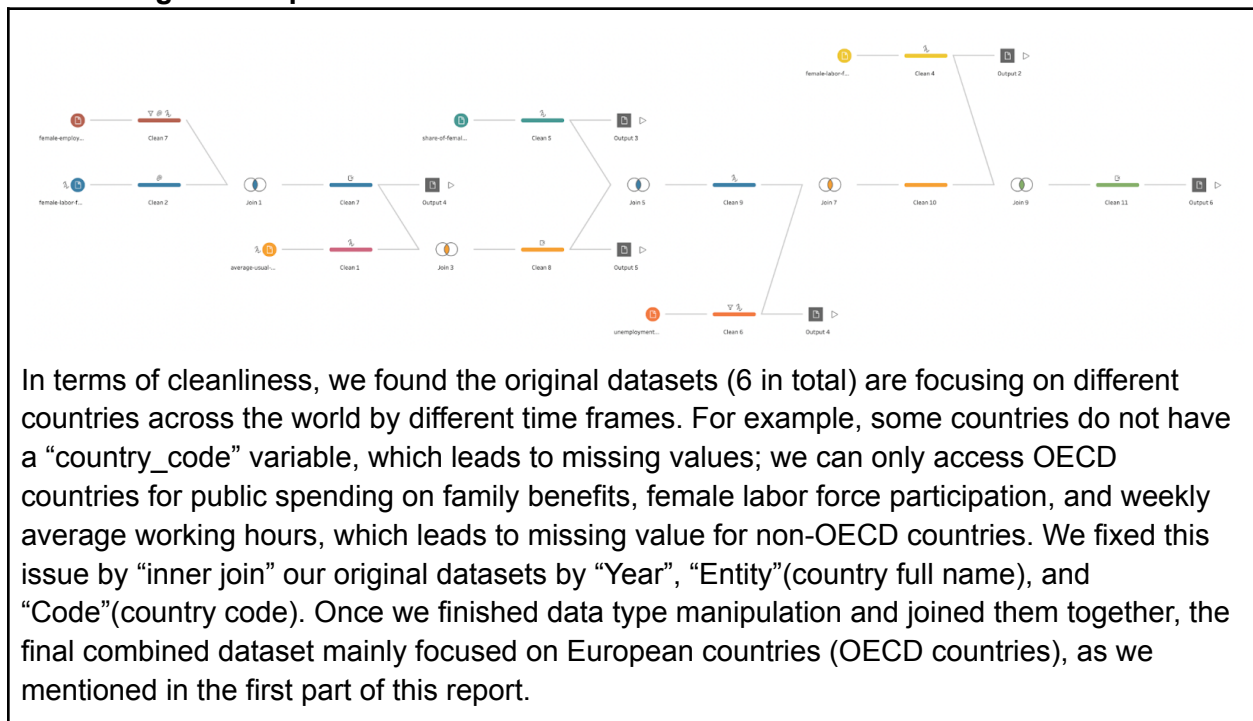
In report 2, we planned to test whether the fertility rate negatively affects female labor force participation “statistically significant” for our client company. Then we can future investigate how companies could help female employments overcome the negative effects brought by high fertility. Now, we included the fertility rate for each country by each year as the baseline.

- [Kaggle](#) :
 - Country to Continent

In report 2 we fixed the continent variable by adding the “country to continent” dataset in R studio. This data was well-formatted and has been cited many times by users of Kaggle. Since we double-checked the accuracy of this data and it was completed, we included this data to provide needed information of continents of country.

Cleanliness

Precleaning and Prepare in Tableau:



Precleaning and Prepare in R:

The data added in report 2, Fertility rate data has a similar structure as the previous data, which means it has records for each country and each year. For the country to continent data, we manipulated the mismatched country names for the countries. To be more specific, two

countries with mismatch names are “The United Kingdom” and “Czechia” in the previous dataset, and they were named "United Kingdom of Great Britain and Northern Ireland" and “Czech Republic”. We fixed this by changing the names of these two countries to merge them.

To create reader-friendly plots, we need to make sure our data will be explained in the right plot in the right format. We used the `pivot_longer()` function in `r` to achieve it. (See `r` script to take a closer look)

For outliers, we used a `boxplot.stats()` to identify and remove the outliers. This line of code can remove the existing outliers without identifying new outliers. This function comes from “[Box Plot Statistics](#)” in the `R` library.

Another assumption we need to make before analyzing and visualizing this dataset is “No measurement error in the original 6 datasets”. Since our original data all came from reliable and most often used sources, we are confident to make this assumption.

Overall, we believe our combined dataset has good quality in terms of cleanliness.

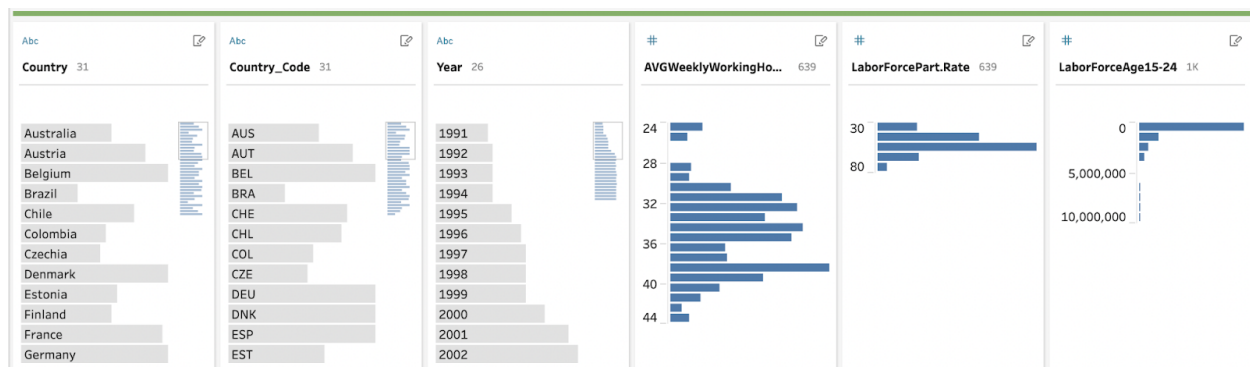
Completeness

In terms of completeness, we have 31 countries in total once we completed the manipulation and cleaning. Some of the records in Eastern Europe are not available. To avoid “extrapolation” in generating insights, the countries mentioned in the rest of the reports mean 31 countries in the datasets.

The time frame expressed by our data and this report was between 1991 and 2016. Overall, we got a dataset with good quality in terms of completeness.

Completeness of additional data:

No missing value found in the combined dataset with continent and fertility rate variables.

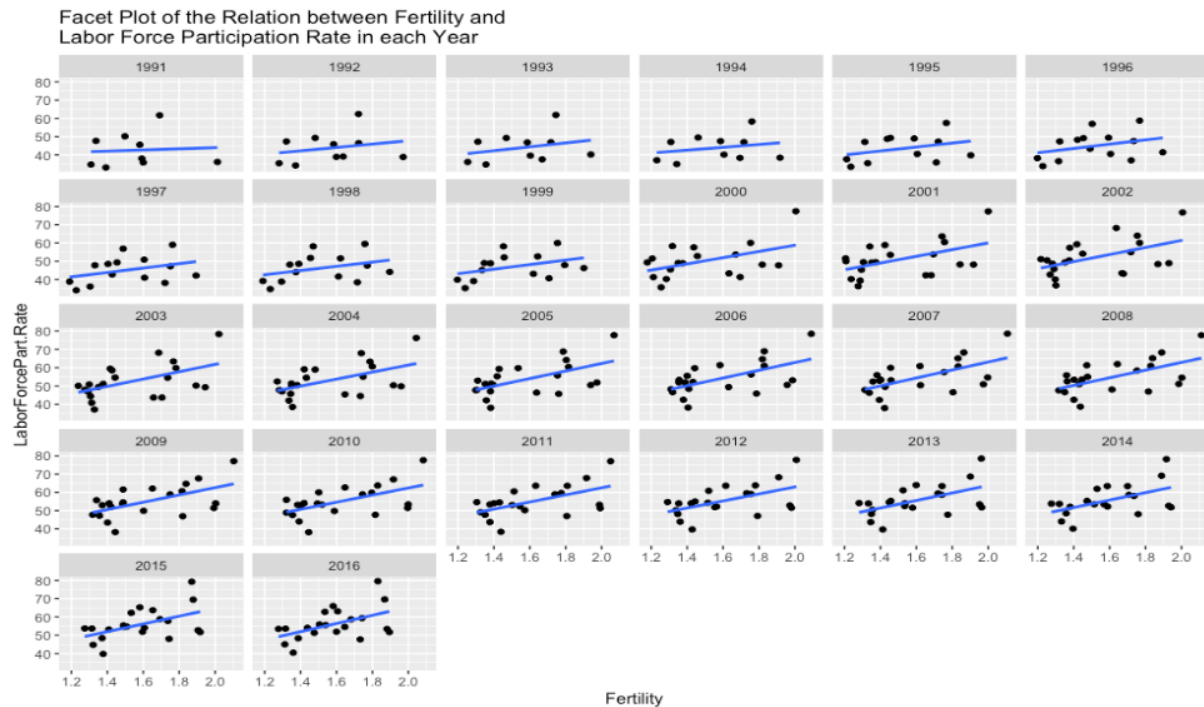




(Continent and Fertility has been added to the dataset)

Data Visualizations and Discussion from Different Aspects

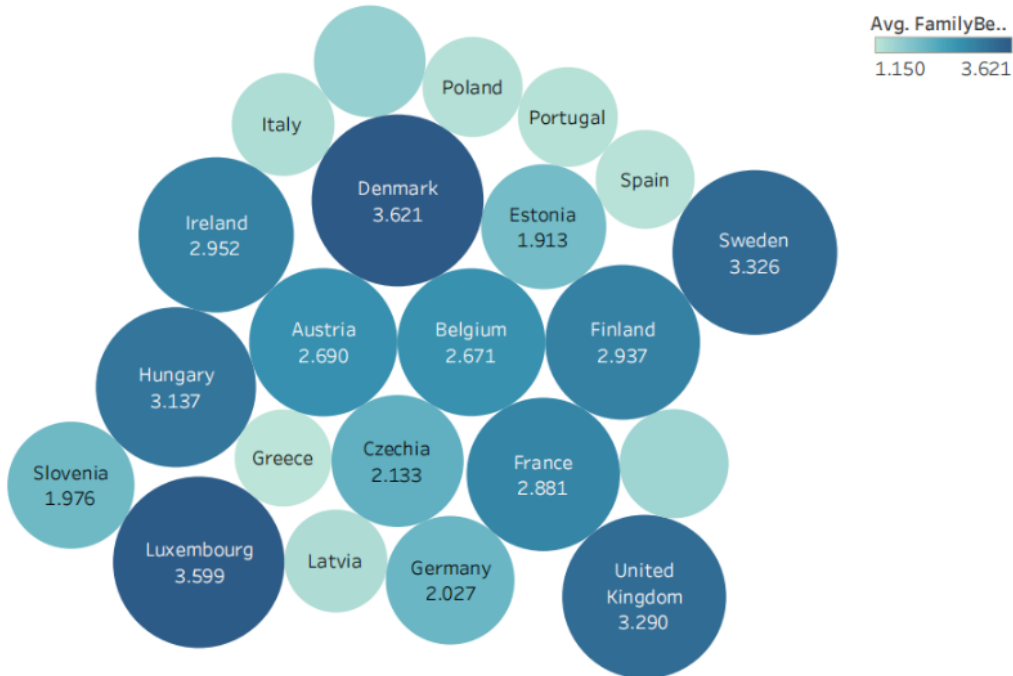
Argument 1: Does fertility affect female labor force participation rate?



The facet plot showed the correlation between Europe's female fertility and labor force participation rate in each year. From the visualization, we can see that fertility rate and labor force participation has a positive relationship and it's more obvious as the year increases. The graph also tells us that overall European females slightly tend to have lower fertility since 2000. On the other hand, the female labor force participation rate is slightly increasing as the year goes by. Before we generated the graph, we thought that fertility should have a negative effect on the female labor force participation rate since with more children, mothers need to spend more time taking care of them and have no time to participate in work. However, from the visualization result, we surprisingly discover a positive correlation between fertility and female labor force participation rate. We think that there should be some deep reasons behind this result.

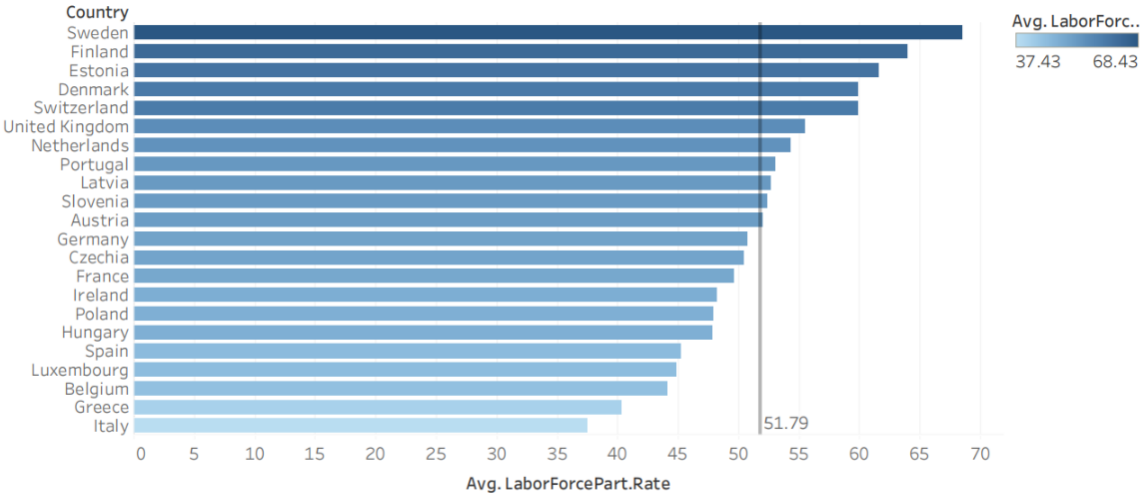
Argument 2: Does public spending on family benefits encourage women employment?

Packed Bubbles_Public Spending on Family Benefits



Country and average of FamilyBenefitsPubSpending (% GDP). Color shows average of FamilyBenefitsPubSpending (% GDP). Size shows average of FamilyBenefitsPubSpending (% GDP). The marks are labeled by Country and average of FamilyBenefitsPubSpending (% GDP). The data is filtered on European Countries, which keeps European Countries.

Bar Chart_Female Labor Force Participation Rates in European Countries



Average of LaborForcePart.Rate for each Country. Color shows average of LaborForcePart.Rate. The data is filtered on European Countries, which keeps European Countries.

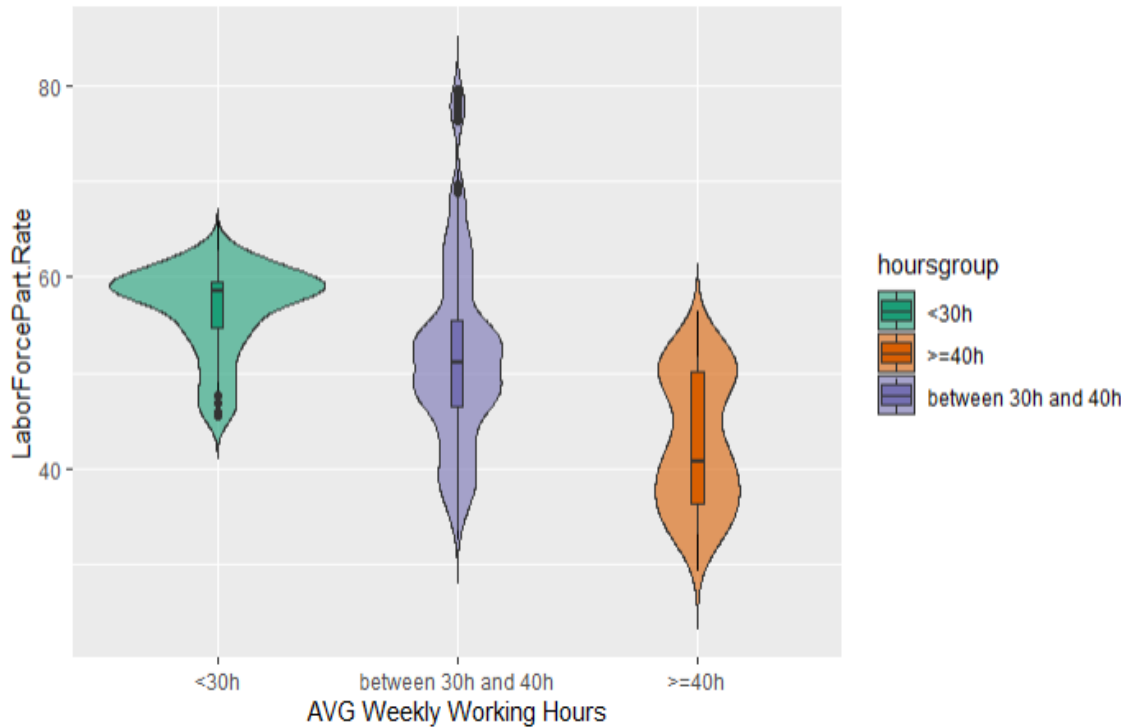
In the first packed bubble graph, we plot the government's public spending on family benefits for all European countries. We mark the measure of the percentage of GDP that the government spends on providing family benefits by color and size. As a result, the top 5 European countries that have the highest percentage of GDP in public spending on family benefits are Denmark (3.621%), Luxembourg (3.599%), Sweden (3.326%), the United Kingdom (3.29%), and Hungary (3.137%). One disadvantage of this packed bubble graph is that for those countries with the least government spending on family benefits, their names and details are omitted on the graph due to the limited size of their bubbles.

In the second bar chart, we graph average female labor force participation rates for all European countries from 1990 to 2016. We filter the magnitude of participation rate by color and add a reference line to see the average of the distribution. To make the graph more clear, we sort all countries' female labor force participation rates in descending order. As a result, we can see that Sweden has the highest female labor force participation rate, while Italy has the lowest. The average female labor force participation rate across all countries is 51.79%.

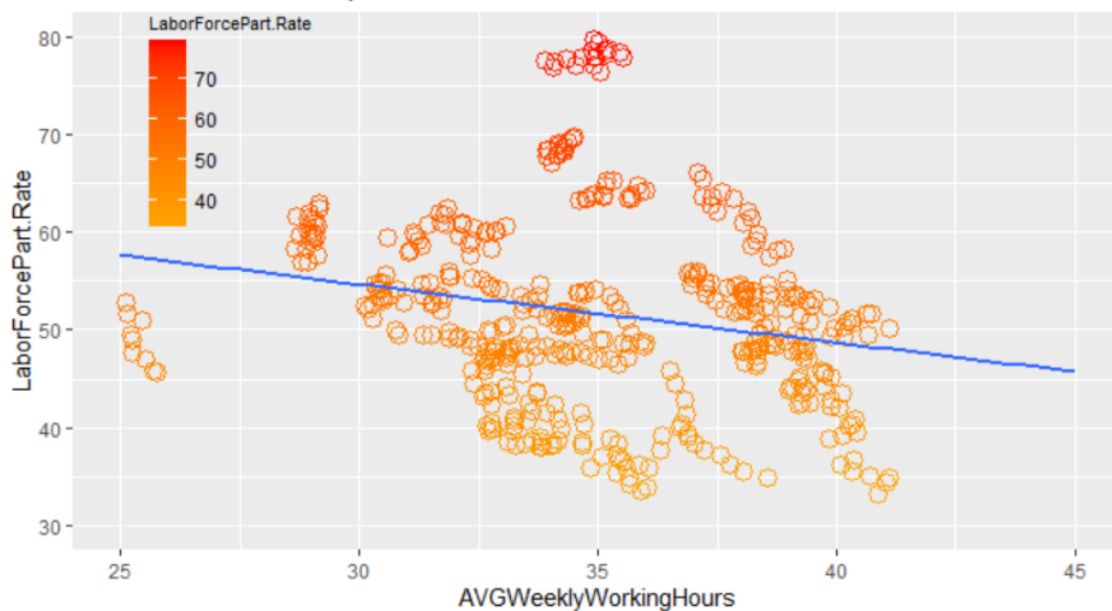
Combining two graphs, we found that countries with higher government public spending on family benefits encourage more women to participate in the labor force. As the country with the highest female labor force participation rate, the Swedish government spends 3.3% of its GDP on public spending devoted to family benefits. Finland spends 2.9% of its GDP. Denmark spends 3.6% of its GDP. The United Kingdom spends 3.3% of its GDP. All the above-mentioned countries have female labor force participation rates much higher than average, and their governments spend a significant amount of their GDP on public spending on family benefits.

Argument 3: Do Women's weekly working hours affect female labor force participation rates in European countries?

Violin Plot of Labor Force Participation Rate by Avg Weekly Working Hours



Regression with Outliers_Long Working Hours Discourages Female Labor Force Participation Rate

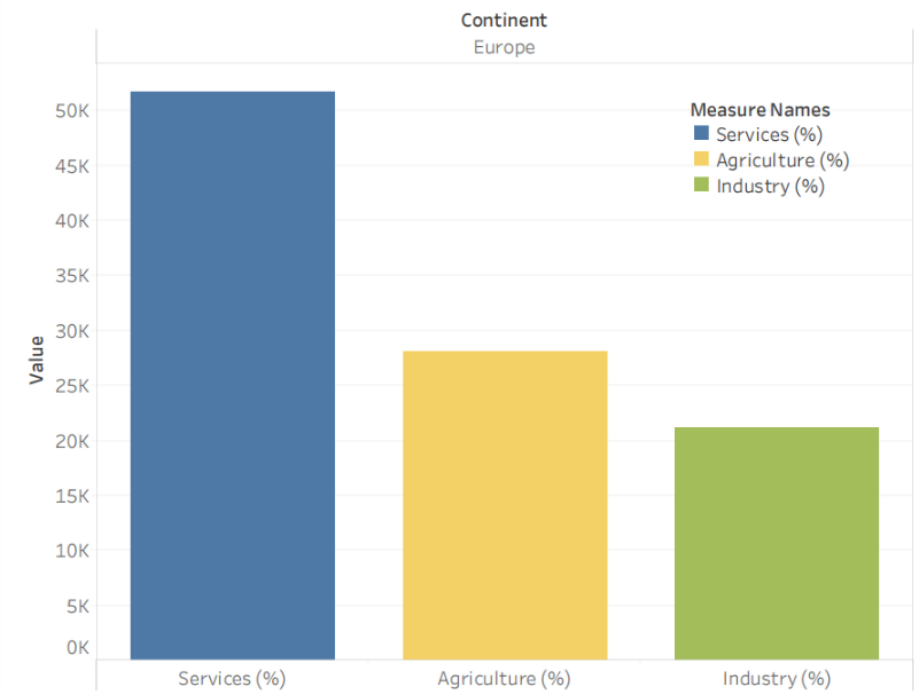


In the first violin plot, we plot the labor force participation rate by average weekly working hours. We grouped weekly working hours into three categories (average weekly working hours under 30 hours, between 30 and 40 hours, and above 40 hours). The visualization told us that overall females have a higher labor force participation rate if they have lower average weekly working hours. From the boxplot, we can see the medians of each group are around 60%, 50%, and 40%. “<30h” and “between 30h and 40h” two categories have a few outliers. One’s outliers are low extreme values and the other’s are high extreme values. The majority of weekly working hours are between 30 and 40 hours and the labor force participation rate has the largest variation. From the graph, we can see a negative correlation between weekly working hours and female labor force participation. Medians of the three groups, which represent the female labor force participation rates of each group, decrease, as the average weekly working hours increase from one group to another.

To further prove our finding, we regress the female labor force participation rate on average weekly working hours in the second scatterplot. This regression line shows a negative correlation between female labor force participation rate and weekly working hours, which proves our hypothesis that longer working hours would discourage women’s participation rate in the labor force.

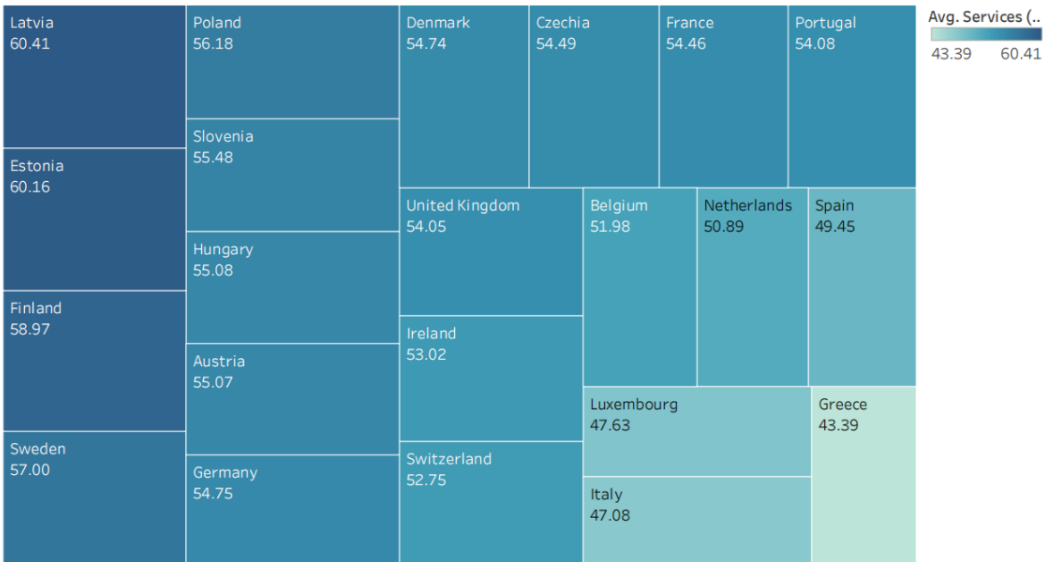
Argument 4: Do different working fields affect the female labor force participation rate?

Bar Chart_Female Share of Employment in Economic Sectors

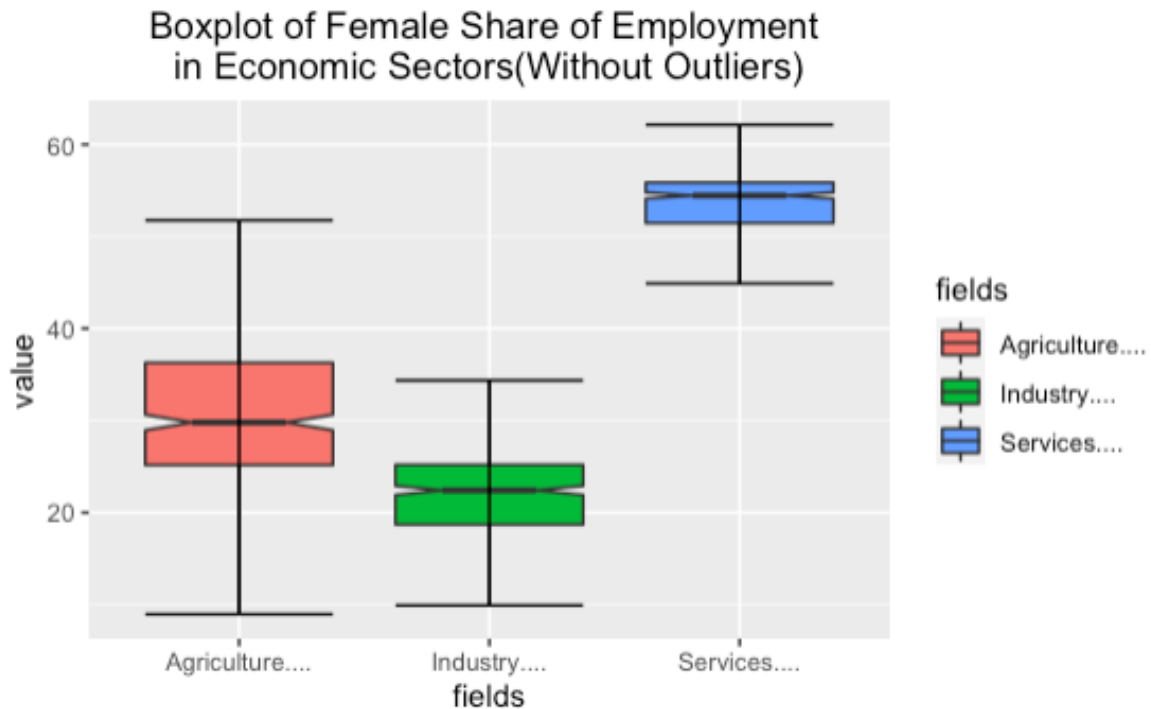


Agriculture (%), Industry (%) and Services (%) for each Continent. Color shows details about Agriculture (%), Industry (%) and Services (%). The view is filtered on Continent, which keeps Europe.

Treemap_Women's Share of Employment in the Service Sector



Country and average of Services (%). Color shows average of Services (%). Size shows average of Services (%). The marks are labeled by Country and average of Services (%). The data is filtered on European Countries, which keeps European Countries.



In the first bar chart, we plot the female share of employment in three economic sectors: services, agriculture, and industry. We found that most female employment occurs in the service sector. The Percentage of women working in service is about 60% in Europe, while agriculture is 45% and industry is only 30%. Thus, companies and countries that focus on service areas are more likely to hire women. Or females are much more likely to work in the service field than others. Either way might influence the total female labor force participation rate.

In the second treemap, we look into women's share of employment in the service sector specifically in all European countries. We found the top 5 countries that have the highest percentage of women working in the service sector are Latvia (60.41%), Estonia (60.16%), Finland (58.97%), Sweden (57%), and Poland (56.18%).

In the third boxplot, we further analyzed women's labor force participation rate in different working fields. The boxplots above show the female share of employment in three economic sectors: industry, agriculture, and services. These sectors are highlighted in different colors. The horizontal lines in each boxplot's middle are the mean of each sector. While the triangle cut off from the mean line is the 95% confidence interval of the mean. The upper and lower edge of the colored box is the 75th and 25th percentiles. While the error bar at the top and bottom of each box plot is the largest value within 1.5 times the interquartile range above the 75th percentile, and the smallest value within the 1.5 times interquartile range below the 25th percentile. As we can see, the industry field (shown in green boxplot) has the lowest percentage of women employment while the services field (shown in blue boxplot) has the highest.

Combining three graphs, we found that women tend to be disproportionately concentrated in service occupations. Moreover, countries with high female labor force participation rates such as Sweden, Finland, and Estonia, also have a high share of women's employment in the service sector. If the government could encourage companies in the service sector to hire more women, it would lead more women to participate in the labor force and create more value for society.

Argument 5: Does female employment to population ratio affect female labor force participation rate?



Call:

```
lm(formula = LaborForcePart.Rate ~ EmploymentPopulationRatio...,  
    data = Emp_Lab)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.923	-2.231	-1.104	1.103	9.752

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.98710	0.95599	8.355	8.6e-16 ***
EmploymentPopulationRatio...	0.94992	0.02084	45.572	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.417 on 441 degrees of freedom

Multiple R-squared: 0.8248, Adjusted R-squared: 0.8244

F-statistic: 2077 on 1 and 441 DF, p-value: < 2.2e-16

The scatterplots above show the relationship between the female labor force participation rate and women's employment to population ratio. Each circle represents one specific country's data in a specific year. The color of the circles represents the magnitude of the women's employment to population ratio. High women employment ratios are represented by darker orange circles on the upper right corner of the graph, and low employment ratios would have yellow circles. The black line is the linear regression line between these two factors. We ran an 'lm test' in R to obtain the best fit.

Interpretations, Insights, and Reflections

Insights

Based on our analysis, we summarized our five conclusions as follows:

- **Improving the welfare and social security of women giving birth to children might encourage females to participate in working without any concerns.**

One possible reason is that countries with higher fertility rates in Europe might have better welfare and social security for pregnant women and mothers. For example, some countries would give money rewards or lower the tax rate for families with children. These countries might also have better babysitting and childcare institutions. With these benefits, women do not need to worry about having a new child that would prevent them from joining the workforce. In conclusion, we don't think increasing fertility is a good idea to encourage women to participate in the labor force. Instead, we should increase welfare and social security for pregnant women and mothers.

- **Providing more financial support for families and children can encourage more women to participate in the labor force**

The government should provide more financial support for families with children to encourage more women to participate in the labor force. For example, the government can provide generous subsidies for childcare services, child-related cash transfers to families with children, tax exemptions for families with children, etc.

- **Long working hours discourage women's participation rate in the labor force**

we found that most European countries with longer weekly working hours on average tend to have a lower women labor force participation rate. More interestingly, those countries with shorter weekly working hours and a higher women labor force participation rate are overall richer countries in Europe (e.g. Sweden, Finland, Estonia, Denmark, Switzerland, the U.K.). Some potential reasons causing this phenomenon could be working flexibility or social pressure. Richer countries may enact more policies, for example, maternal health and childcare policies, to ensure gender equality, which provided lots of support for women's rights and greatly encouraged women's participation rate in the labor force.

- **Companies and countries that focus on service areas are more likely to hire women**

We concluded that countries with higher women labor force participation rates also have a stronger service sector. From our perspective, we would strongly recommend companies in the service sector, such as health care, food service, and nursing industries, build gender-inclusive workplaces and increase female leadership proportion as women are the main population in this area. On the other hand, this disproportionate distribution also reflected that women only hold a limited portion in the high-tech industry. For our potential clients in the tech industry, some of the ways to shorten the gender gap in the workplace and encourage more female candidates to apply for this type of job

could be holding women in STEM workshops, creating internship programs only for female applicants in college

- **If companies or the society employs more women, it would encourage women to participate in the labor force more actively**

From our graph, we can see that there is a positive linear relationship between the female labor force participation rate and women's employment to population ratio. The coefficient of the Employment ratio is nearly 0.95 and the P-value is about 45.5 which means this result is significant. As women's employment to population ratio increases, the female labor force participation rate also increases. We speculate the possible reason behind this phenomenon is that if companies hire more female workers, it would encourage more women to step out of the home and participate in the labor force because of the openness of society.

Reflection of the journey as a team

- **Data Cleaning**

As a group, we all agree on the importance of data quality. We tried our best to have data be cleaned and ready to use. Another important thing that we found essential is Version control. Whenever we make changes or updates, we'll keep others posted so no one would repeat the thing that has been done before.

- **Information Visualization**

The R library is one of the most helpful resources. We learned many comments from here like "pivot" from their site. Visualizations make sense only if people can understand them. So, we make our visualization as simple and reader-friendly as we can.

- **Team Work**

It was a pressure to work as a group. Everyone in our group adapts to competitive advantages. We are good at time management, actively participated in weekly meetings, and are happy to help others. Communication is important in teamwork and we did a good job on that.

References

Skadsen, Chaney. (2017). *Chaney Skadsen Fertility and Female Labor Force Participation: The Role of Legal Access to Contraceptives*. Illinois State University. Retrieved from <https://ir.library.illinoisstate.edu/cgi/viewcontent.cgi?article=1029&context=scced#:~:text=When%20the%20total%20fertility%20rate,level%20of%20significance%2C%20economically%20significant.>