CSE 163 Data Project

Analyzing Gender Wage Gaps

Sarah Schafer, Jessica Robinson, Vianne Bui Nguyen

March 15, 2022

Summary of Research Questions

1. **How does the wage gap differ between various occupations?**
   We identified and compared gender wage gaps for different occupations to determine which occupations have the highest gap and which ones have the lowest.

2. **How does the wage gap differ across education levels?**
   We computed the wage gap for our data set for various education levels to see what education levels have the greatest disparities.

3. **How does the gender wage gap differ across countries?**
   We visualized the gender wage gap for different countries in 2014 to identify which countries have the highest and lowest gender wage gaps.

4. **How has the gender wage gap changed over time?**
   We plotted the gender wage gap for different countries over time to determine the trends in gender wage gaps to see if they have decreased or increased over time.

5. **Does a country's proportion of women in management positions correspond to its gender wage gap?**
   We plotted the relationship between countries' percentage of women in management positions and gender wage gap to test our assumption that countries with the highest proportion of women in management positions have the lowest gender wage gaps.

Motivation

We want to uncover information about the gender wage gap in countries around the world in order to help reduce the issue. Our analysis of the data could help policymakers or industries see where the gender wage gap is occurring to the highest degree and show them where to invest their efforts. Without information about the current situation, it is difficult to make any progress in reducing the wage gap. Our goal is to make this information more accessible so that both policymakers and the general public can understand gender disparities and are motivated to act to reduce them.

Dataset

First dataset - Glassdoor data showing income for various job titles based on gender:
https://www.kaggle.com/nilimajauhari/glassdoor-analyze-gender-pay-gap
This dataset allows us to identify pay gaps between women and men for different occupations. The key features of this dataset include job title, gender, age, education, seniority, base pay, and bonus.

Second dataset - Organization for Economic Cooperation and Development (OECD) Gender Wage Gap Data:
Link to the website where we found the data:
https://data.oecd.org/earnwage/gender-wage-gap.htm
Link to the CSV file:
https://drive.google.com/file/d/1h9nsk_YfA-Mv6hCrcDMwy2gA2fcabFo5/view?usp=sharing
This dataset has relevant data about countries and their relative wage gaps for various years. The relevant columns are the location as a three-letter abbreviation of the country's name, the year (called time), and the wage gap of that country in that year in the value column.

Third dataset - Our World in Data Economic Inequality by Gender: Women in Senior and Middle Management Positions:
Link to the article: https://ourworldindata.org/economic-inequality-by-gender
Link to the CSV file:
https://drive.google.com/file/d/1IxCgZdMGOkQZ0tJ-ijcxGFokZGdMOMG6/view?usp=sharing
There are a number of interesting datasets in this article that may be useful for our analysis that are linked as CSVs for each chart in the download tab. We will use the dataset on women in managerial positions from the Our World in Data "Economic Inequality by Gender" article. This dataset, included in one of the charts in the middle of the article, has rows representing a country in a particular year. The columns are the entity (country), its code (three-letter abbreviation of the country), the year, and the proportion of the women in senior or middle management positions as a percentage. We may use some of the other datasets in this article to answer questions about women's ability to participate in economic household decisions, or the share of women in lower-paying or higher-paying jobs.

Challenge Goals

1. **Multiple datasets:**
   In analyzing the gender wage gap, it was difficult to find a single data set that could show how the gap has changed over time, how it varies between countries, and how it varies between occupations. To understand the topic, all of these aspects are crucial. Thus, we need multiple datasets to answer these questions.

2. **New library:**
   We decided to try to use the Plotly library because we wanted to learn how to make interactive visualizations with Plotly. We want to visualize our data to communicate it and allow people to view the information easily. Interactive visualizations help with this goal.

Method

1. **How does the wage gap differ between various occupations? - Vianne**
   Using our Glassdoor dataset, we applied the Pandas groupby method to group our data by occupation then computed the average base pay for each gender. We created a bar chart with the various occupations on the x-axis and the base pay in dollars on the y-axis to visualize the average pay of men and women respectively in each occupation.

2. **How does the wage gap differ across education levels? - Jessica**
   In order to apply the Pandas groupby method to compute the average base pay for both men and women by grouping the data by level of education we used our Glassdoor dataset. We then created a bar chart to visualize the average pay of men and women with the x-axis displaying the levels of education and the y-axis displaying the base pay in US dollars.

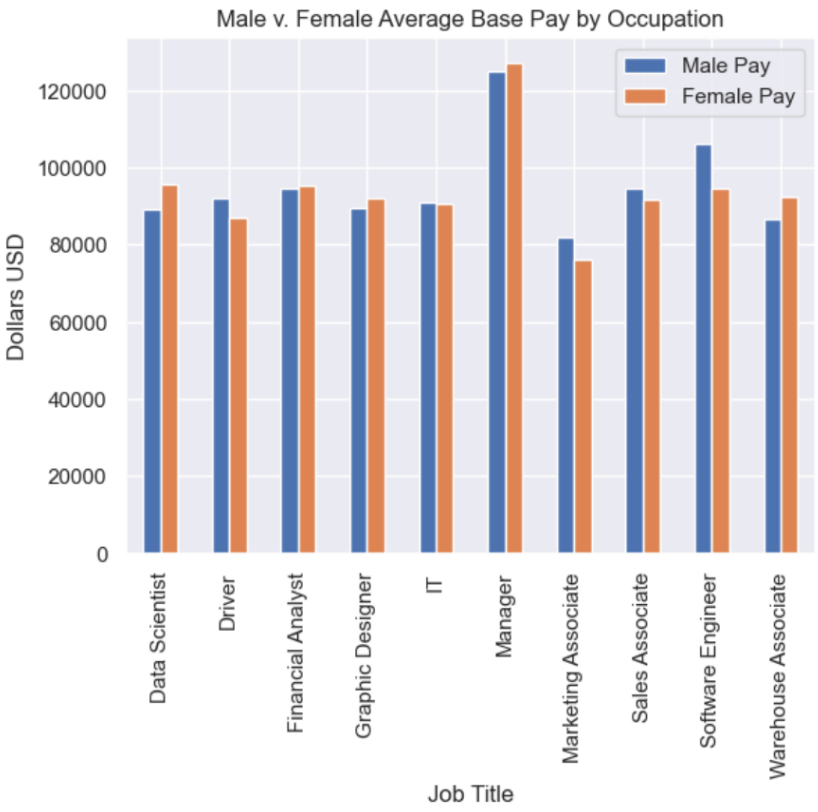
3. **How does the gender wage gap differ across countries? - Together**
   We merged the OECD gender wage gap dataset with a dataset that contains geospatial data for the world. We filtered the merged dataset to the year 2014 because it contained data for the most countries. First, we plotted a map of the world, then we plotted the gender wage gaps for all the available countries with a color scale where the darkest color indicates the lowest gender wage gap and the lightest color indicates the highest gender wage gap. We created a map visualization to show how the gender wage gap differs across countries.

4. **How has the gender wage gap changed over time? - Sally**
   For this question, we used the second data set (the same one we had used with the countries) We used a Plotly line plot, where each line represented a different country. Using Plotly, the viewer is able to see what line pertains to each country and information about specific years in hovering over the data. In order to make the graph readable, we could not use all of the countries. We wanted to be able to see the trends over the longest period of time, so we decided to only keep countries in the data set that had data before or starting in 1990; only ten countries fit this category. Because of the way we filtered the data, some of these countries appear as though their data starts after 1990. However, this is only because these countries did not have data specifically for the year 1990, but rather were included because they had data before 1990. Thus their trendlines start at their earliest year after 1990.

5. **Does a country's proportion of women in management positions correspond to its gender wage gap? - Together**
   We merged the OECD gender wage gap dataset and the women in management positions dataset and filtered to year 2014 because it contained the most data. We created a scatter plot for 2014 with the percentage of women in management positions on the x-axis and the gender wage gap in dollars on the y-axis. Each point on our plot represents a single country. We created this plot using Plotly, so that when the user hovers over a data point they can see the country name, its gender wage gap percentage, and its proportion of women in management positions.

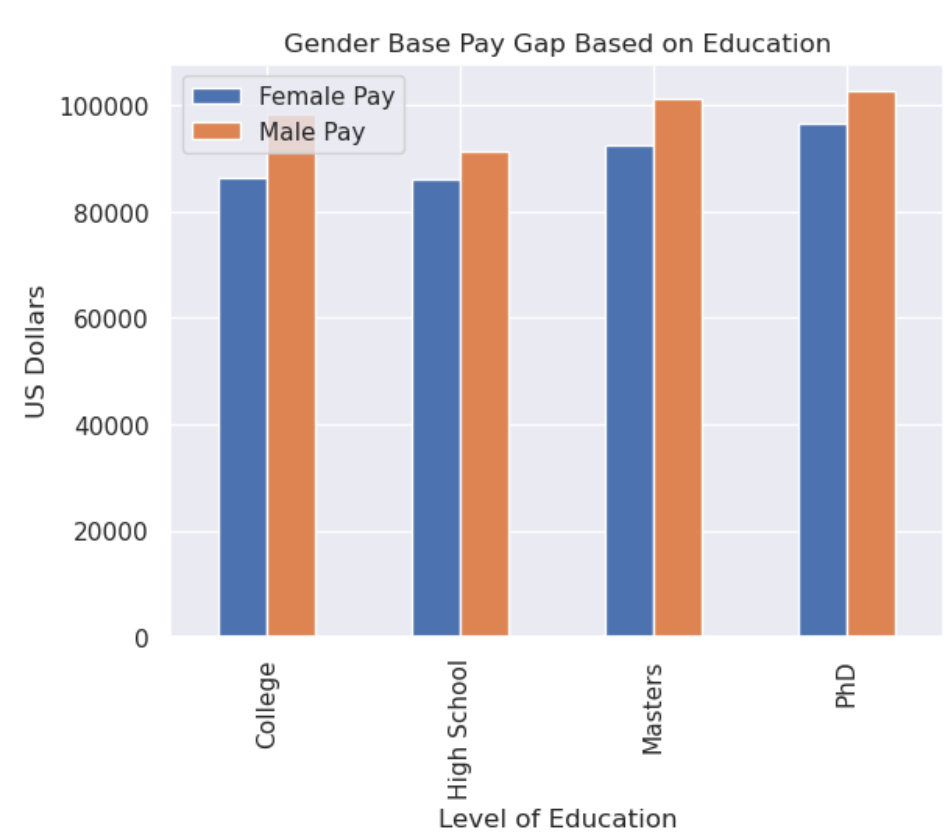
Results

1. **How does the wage gap differ between various occupations?**



This bar chart depicts the male and female base pay for 10 occupations represented in the Glassdoor dataset. Five of the 10 occupations show men earning more than women: Driver, IT, Marketing Associate, Sales Associate, and Software Engineer. This might be explained by the fact that these occupations tend to be male-dominated. For instance, 80% of delivery drivers in the US are known to be male while only 17% are known to be female (Zippia, 2021). Similarly, a 2021 global software developer survey shows the vast majority of developers are male as 91.7% of the respondents identified as male while only 5% of respondents identified as female (Statista, 2021). In male-dominated occupations, men are believed to be "more skilled", meaning women who work in these occupations are likely to be paid less than their male counterparts, even if they are similarly educated (Economic Policy Institute, 2016). To our surprise, 5 of the 10 occupations show women earning more than men: Data Scientist, Financial Analyst, Graphic Designer, Manager, and Warehouse Associate. We expected men to earn more than women in all 10 occupations since "women earn less than men in nearly all occupations", making "82 cents for every dollar a man earns" (US Department of Labor, 2021). One possible explanation for the large number of occupations with higher women's pay is that the average seniority of women in the dataset is 3.01 while the average seniority of men in the dataset is 2.93. Higher seniority levels would point to higher pay. Furthermore, since there are more men than women in the dataset, this could mean that women feel less inclined to post their pay if they feel like they are not compensated fairly. Even though the bar chart suggests male and female pay are similar and gender wage gaps trends are minimal, we understand that our dataset is not equally distributed and will interpret the results lightly.

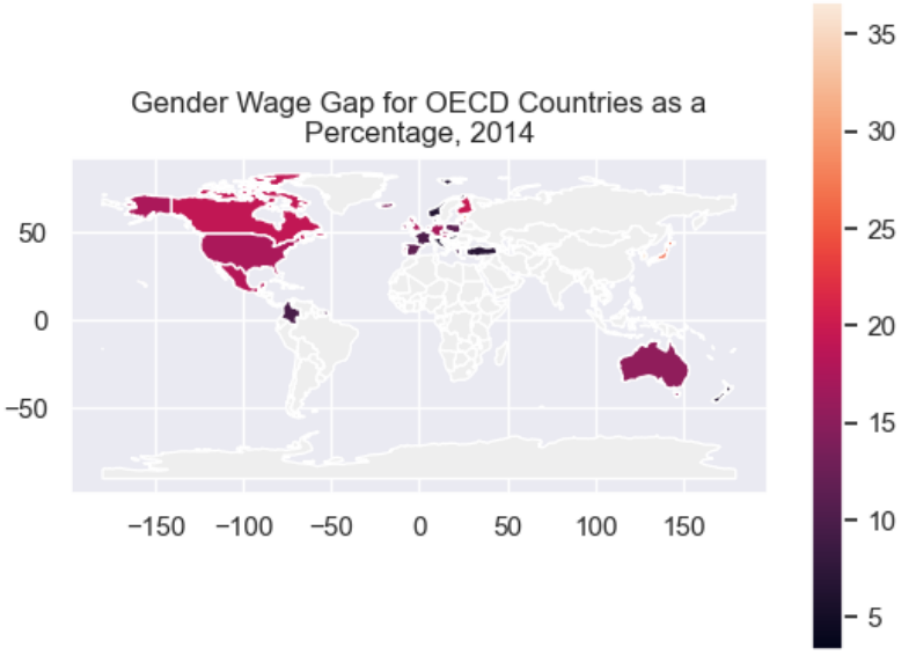
2. **How does the wage gap differ across education levels?**



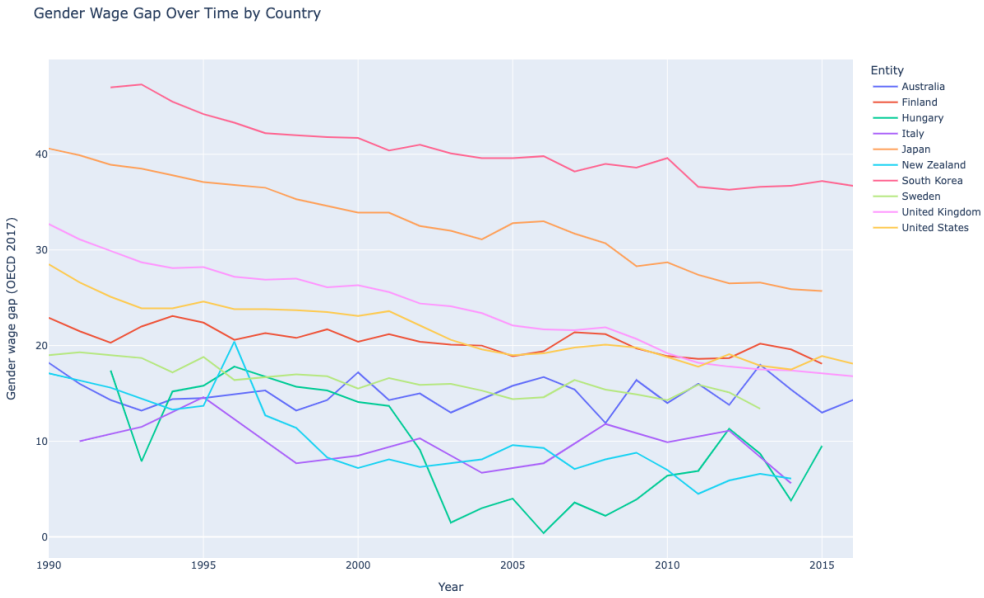
This bar chart illustrates the base pay for both men and women across four different levels of education from the Glassdoor dataset: high school, college, masters, and PhD. For each level of education, men are shown to have a higher base pay than women. It wasn't too shocking to see that there was a gap for each level of education, that was what we expected to find. However, it is a bit shocking that even when both men and women are matched with their level of education, there is still a disparity in their wages. It was interesting to see that the largest wage gap that can be seen for the different levels on the chart is college. According to the United States Census Bureau, "[a]mong workers with a bachelor's degree, women earn 74 cents for every dollar men make, which is less than the 78 cents for workers without [a] college degree" (United States Census Bureau, 2018). This provides some clarity that the results of our bar chart are not completely farfetched. But, even though the results are not far off, our dataset still is not perfect and the results should be taken into consideration with a grain of salt.

3. How does the gender wage gap differ across countries?

This map depicts the gender wage gap percentage for OECD (Organization for Economic Cooperation and Development) member countries in 2014. We chose to plot 2014 data because 2014 contained the highest number of member countries. The results show that European countries tend to have lower gender wage gaps than North American countries like the US. This may indicate that European countries are making greater efforts to bridge the gender wage gap than the US. According to the American Bar Association, "the U.S has not kept pace with the EU in the pay transparency context. The United States does not mandate pay transparency on a national level, while a growing number of European countries require employers to publish or provide their employees access to gender pay data" (2018). In 2014, the European Commission recommended that its member countries adopt the following policies: an employee's rights to request information on gender pay levels for the same work, an employer's duty to report gender pay levels, and an employer's duty to conduct pay audits on grounds of gender. Meanwhile, in the US, employers have objected to reporting gender pay data due on the grounds of burden and confidentiality, meaning they cannot be held accountable for pay disparities. Since there is a lack of pay transparency in the US, it would be easier for employers to get away with paying women less than men. However, since OECD countries are economically similar to the US, the laws of European member countries can offer insight as to how the US can adopt similar policies to address its own gender wage gap.



4. How has the gender wage gap changed over time?



This line chart, created with Plotly and downloaded as a png, represents how the gender wage gap for certain countries has changed since 1990 using the OECD data. For this data set, the wage gap is measured as the difference between average male earnings and average female earnings as a percentage of male earnings. Lower percentages indicate smaller gender differences. In the html file, hovering over each line displays details on specific years and countries.

The countries plotted are the countries in the dataset that had data before or starting in 1990, meaning they have tracked this data for the longest. A few of the countries (South Korea, Hungary, and Italy) have trend lines that start after 1990. These countries, though they started collecting data before 1990, did not have data specifically for 1990. We chose to select the countries this way because plotting all countries made the visualization unreadable, and we wanted to see data over the longest period of time.

From the graph you can see that overall in most countries, the wage gap has been decreasing slightly; however, there are large variations between years for each individual country. Further, we can see that all of the countries included in the plot do have a wage gap; the difference in male and female earnings is not zero for any country. Of these 10 countries, the country with the highest wage gap is South Korea, while the United States ranks fourth for the highest wage gap in most years in this subset. It is clear from the high wage gaps for all of the countries that the wage gap exists and needs to be addressed. Additionally, not all countries have equally high pay gaps, so measures that need to be put in place may differ between different countries. Possible reasons for the differences between countries include differences in culture and existing policy protections for women. Differences in accuracy and representation of women in data collection likely also plays a role, as some countries have different requirements for companies to report their wage gaps.

5. **Does a country's proportion of women in management positions correspond to its gender wage gap?**

This scatter plot, created with Plotly and downloaded as a png, shows how the wage gap for a country relates to the proportion of women in senior and middle management positions in 2014. Each point on this chart represents a single country, and using Plotly with the html file, you can hover over each data point to see which country it pertains to. To create this chart, we merged the OECD gender wage gap data with the Our World in Data senior and middle management data by country. This means that the countries included are just a small subset of all of the countries in the world; they are only the countries that were included in both of the data sets. In this case, senior and middle management positions mean "legislators, senior officials, and managers" (see Our World in Data).

From the chart, we can see that there is very little correlation between a country's wage gap and the percentage of women in senior and middle management for that country. There are no visible trends. This is interesting as our group expected that as the wage gap decreased, there would be a higher percentage of women in senior and middle management positions. However, it is notable that in this subset of the data, all of the countries have low proportions of women in senior and middle management positions; the country with the highest percentage of women in these positions was Latvia, which still had only 42.05% women. This lack of women in roles that could make a difference in the wage gap could explain why the wage gap does not correlate with the management data, as without a significant amount of women in these roles, change may not be able to be made. Further, the number of women in these roles does not necessarily indicate that they are paid equally. This means that even if there are women in high-ranking and high-earning positions, there could still be a significant wage gap between men and women, in high-paid or low-paid positions.

Impact and Limitations

Our results are meant to spark interest in, and begin to analyze, the gender wage gap. Our results are *not* meant to be the only source of information guiding policy decisions, and should not be viewed as such. They are not meant to excuse certain industries or countries that are doing better than others from dealing with the gender wage gap. Rather, our analysis is meant to show simply where we are, and show how everyone, to varying degrees, needs to improve.

Here are a few specific examples of limitations and possible harmful impacts of our data analysis, which should be kept in mind when processing our analysis:

- The data sets we used had limited information on **data collection.** This could lead to inaccuracy in the data.
    - For example, companies in some European countries have to report data on the wage gap, while America does not require this. This could lead to differences in data. (American Bar Association, 2018)
    - Furthermore, the Kaggle data set from Glassdoor had far more men in the data set than women, likely due to disparate representation of men in those industries, but also possibly due to the data collection process.
    - It is unclear whether people were allowed to self-identify their gender for the data sets. A lack of autonomy for this area could lead to inaccuracies and discrimination.
- The data sets we used had a strictly **binary** interpretation of gender. This excludes nonbinary people and people of other genders. These results should not be used to say that these groups are not significant or do not face challenges, but rather only that the datasets did not account for this. This lack of data and representation could harm these groups.
- The data did not allow us to analyze the gender gap for various **important demographics**, including race, ethnicity, religion, disability, sexual orientation, and more. It has been seen that the intersection of these demographics does also lead to pay gaps. According to an article from Pew Research Center in 2016, race can impact wage gaps (Pew Research Center, 2016). This article mentions that in 2015, black women in the US had average earnings of 65 cents to every white man's dollar, while white women earned 82 cents comparatively (Pew Research Center, 2016). A lack of attention to these compounding factors could lead to ineffective or harmful policy decisions for discriminated-against groups in these areas.

Work Plan Evaluation

Cleaning/preparing the data, visualizing the data, and testing code took about the same amount of time that we set out for. However, we estimated to spend only 6 hours in team meetings when in reality, we spent much more. This is because we met multiple times a week and used our meetings as synchronous work time to complete problems together instead of just updating each other about the progress we made individually.

We also divided the work differently. We each completed one research question individually and then came together to work on two research questions together. Each team member tested the research question they were responsible for while we collaboratively tested the two research questions we had worked on together.

Testing

We tested our code by creating and using small subsets of our data as well as using print statements. Through the use of these methods, we went in and created test versions of our functions from the analysis file in order to see the integrity of our code. The usage of small subsets of our data allowed us to check that our code would still work on a smaller, modified level. Also, we were able to make sure our code was functioning the way it is meant to through the visualization from the print statements. Testing our code was an important step and provided us with the peace of mind that our code was functioning correctly so we could continue on with our project. Because we were able to filter our data in the way we expected with these tests, our visualizations accurately represent the datasets we have.

Collaboration / Sources

Plotly resources:

- https://Plotly.com/python/line-and-scatter/
- https://Plotly.com/python/line-charts/
- https://Plotly.com/python/interactive-html-export/
- https://Plotly.com/python/figure-labels/

Other programming resources:

- https://www.w3schools.com/python/ref_func_round.asp
- https://stackoverflow.com/questions/11346283/renaming-column-names-in-pandas
- https://stackoverflow.com/questions/12096252/use-a-list-of-values-to-select-rows-from-a-pandas-dataframe

Wage gap Resources:

- https://www.statista.com/statistics/1126823/worldwide-developer-gender/#:~:text=According%20to%20a%20global%20software,reality%20of%20software%20development%20jobs.
- https://www.zippia.com/delivery-driver-jobs/demographics/
- https://www.epi.org/publication/what-is-the-gender-pay-gap-and-is-it-real/
- https://blog.dol.gov/2021/03/19/5-facts-about-the-state-of-the-gender-pay-gap
- https://www.americanbar.org/groups/labor_law/publications/ilelc_newsletters/issue-june-2018/eu-us-gender-pay-data/
- https://www.census.gov/library/stories/2019/05/college-degree-widens-gender-earnings-gap.html
- https://www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-in-u-s-despite-some-progress/