Gender Study: Gender Pay Gap with Real Estate Analysis

CSE 6242 Team 126 Project Report

Jia Wan, Wei Peng, Lucy Xinying Lu, Xuanming Lin, Hanna Choi
12-1-2022

1. Introduction

The gender pay gap (GPG) has been studied across different regions and occupations, but only a few studies have dived into the multi-faceted impacts of GPG on people's daily lives. Our project helps describe GPG in association with the real estate market and their changes in time in the United States. We believe everyone in our society should care about the study of GPG for the decrease in GPG confirmatively increases productivity growth in all business sectors (Wolszczak-Derlacz, 2013). Female should be aware of the existence and change of GPG as female wages are typically lower than male across all sectors (Miler 2008).

2. Problem Definition

What kind of living are we earning for ourselves with our salaries? How many years of one's salary does that take to buy a house to live in? Will answers to those questions differ for male and female? If there is a difference, will such difference change from the east coast to the central to the west coast of the country? Instead of trying to answer all the above questions ourselves, our team aims to provide an easy navigation tool for people to find answers on their own with a reliable data source.

GPG may have an impact on the housing market. We studied the relationship between income by gender and the variation in real estate costs by developing a user interactive dashboard to help users to visualize and to make prediction on the gender income, real estate cost and pay to house prize ratio in the next five years.

3. Literature Survey

Our project ultimately aims to develop an open-source website which provides an insight into the impact of the GPG on people's daily life, especially on their housing costs. To be more specific, our current goal is to develop an interactive and user-friendly dashboard (Murray, 2017) with functionalities including geographic data visualization and analytical graphs with user-defined filtering functions. On the backend, our database combines the census GPG data and the national real estate data (Captain, 2013). Algorithms implemented in our project include but are not limited to time-series prediction models such as ARIMA (Taieb, 2018). We hope to gain new insights from the visualizations we're planning to build with the historical census data, and to confirm or reach an alternative conclusion based on the pre-existing studies that the team researched on. (Lehmann, 2022)

When it comes to the study of GPG in the US, most research to date is conducted by survey and research paper (US Census Bureau, 2022). A lot of preexisting research has been done for a relatively fixed time frame and so far, it has been discovered the causes of the GPG are systematic, even controlling factors including age, location, education background and job level (Payscale, 2022) (Alkadry and Tower, 2014) (Bishu, 2016) (Sterling, Thompso, Wang, Kusimo, Gilmartin and Sheppard, 2020). There has also been study about the historical trend of the GPG. For instance, Blau and Kahn presented that the GPG has been narrowing from about 36-38% in 1970 to between 18% and 21% in 2010 as education and work experience explained about 27% of the gap in 2018 while only about 8% in 2010. (Blau and Kahn, 2017)

Current research on gender gap is narrowly focused on the absolute value of GPG with limited studies on its impacts on other aspects of life. Interactive and user-friendly visualizations are scarce in this research area. The research conducted in the U.S. health sector only concerns

real wage growth and GPG, suggesting that the earning growth rate is higher in female health workers versus male. (Barry, 2021) Research also shows that people with higher education are involved in more cognitive intensive and people-skill intensive jobs. (Bacolod, Marigee, Bernardo and Blum, 2010). One study suggests that commuting preference by gender affects gender wage gap. The gender wage gap is salient in metropolitan centers (Liu and Su, 2022). Another study supports the same idea and states that the GPG difference is about 10% between densely populated cities and rural areas (Hirsch, König, Möller, 2013). The report written by Gould, Schieder, and Geier in 2016 explored different ways to measure gender gap and how it is still a persistent issue in society. According to the report, the shrinking wage gap may not necessarily suggest that women are catching up to men. Since 1979, the median male worker's wage has declined and contributed to the 30% reduction in the gender wage gap (Gould, Schieder and Geier, 2016). Gender gap also affects the housing market and women tend to have worse experiences (higher purchase price and lower selling price) when compared to men (Goldsmith-Pinkham and Shue, 2020). Our team wants to expand on the current studies and deep dive into how exactly gender wage difference can affect their quality of life in housing and cost of living.

4. Proposed Method

In this project, our team successfully developed an interactive and user-friendly dashboard that visually demonstrates the impacts of GPG across various regions of the states. The dashboard allows users to not only explore such differences from region to region in the United States, but also be able to line up such gap with our dwelling availability. Our database combines the U.S. census data of individual earnings and national real estate data. Furthermore, to provide a trending insight for the near future, we implemented a collection of time-series models such as ARIMA and linear regression on the combined data sets.

4.1 Intuition

Our project serves as a user-interaction tool of the GPG from the view against the regional real estate market based on data in the past 10 years. Our research allows users to explore if there are correlations between resident real estate price and GPG. Furthermore, our project guides users to view resident real estate change side by side of the change in buying ability. To do that, we plan to develop models to forecast the GPG together with resident real estate prices for the next three years. To be more specific, we have learned that, with potential further evaluation and adjustments, ARIMA(1,0,1) is suggested for forecasting wages in a company (Toledo, 2022). We will present our research result and predictive data in an interactive geographic visualization dashboard with options to filter, which has not been configured in the current GPG or resident real estate data visualization websites.

4.2 Real Estate: Time Series Model for Housing Market Forecast

For real estate data, we obtained our zip code level dataset, 2.46 Gb size, from Redfin.com and used ARIMA to forecast future US housing median price by city. Data cleaning is performed by Python using the Pyspark and Pandas packages to fill in the blank data in the time series with the previous value. We noted there is a sizable portion of regions that lacked sold housing price data history for an extended period, which makes our method of filling in missing values less reliable. To address that issue, we introduced the US county and city table and grouped the data by US county. We calculated weighted average median price on numbers of sold properties per zip code in every county. Since our data is preprocessed, we introduced a python package called

pmdarima, which contains a useful auto_arima function. This auto_arima function can generate p,d,q parameters automatically based on time series data and seasonal parameter input, which is a powerful tool for our application as we would have to manually tune hundreds of US counties without auto_arima. The auto_arima function generates the future 3-year US housing median price by county. A sample Atlanta housing market forecast is shown in Figure 1 below. While blue line represents the historical data, the orange line is forecast, and blue shade indicates the 95% confidence interval of the forecast.

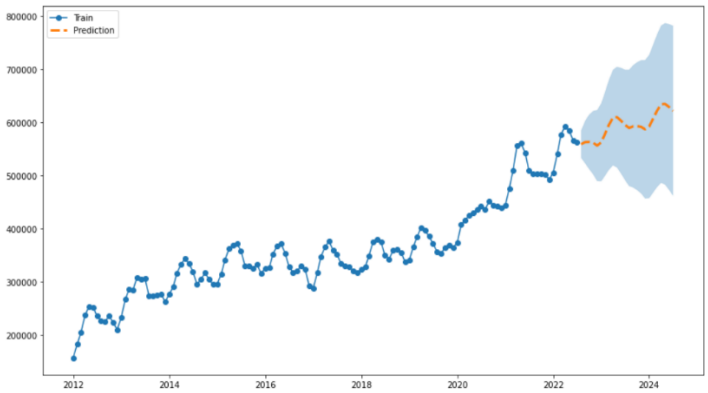

Figure 1. Housing market forecast in Atlanta

4.3 Gender Pay: Linear Regression Model for Individual Earning Prediction

We obtained both the zip-code level and county level median earning by sex for the US population from U.S. Census Bureau from 2011 to 2020. We conducted basic data cleaning by removing irrelevant columns and features, filling in missing data with the most recent prior year's available data, and regulated the geo data. We chose to apply the linear regression model on the time series data over the ten years, which helped us predict the near future income trend by sex for each county. Our linear regression model for income has a median R-squared value of 0.54, 0.61 and 0.64 among all 3220 data points for male, female and total, respectively. Hence, it shows a strong fit of the data in all three categories. On the other hand, given the limitation of ten years of public data we obtained from the U.S. Census Bureau; our model serves in prior as an aid for general suggestions rather than any accurate predictions of the economics. Figure 2 shows an example of the graph for the GPG for Mouheur County, OR, in 2012.

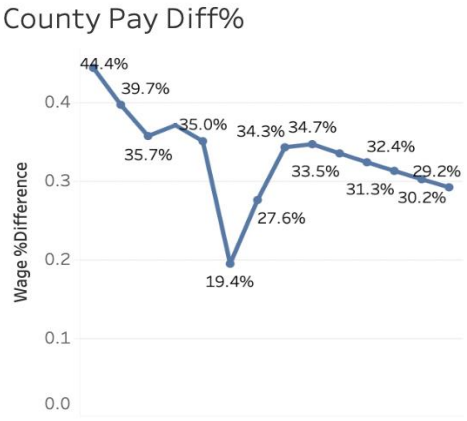

Figure 2. Gender Pay Gap for Mouheur County, OR, in 2012

4.4 User Interface Design and Implementation

Final visualization combines the real estate data and wage data with their comparisons on a map dashboard of Tableau. Using Alteryx, the two data sets were joined with respect to their geo-id. Alteryx also conducted a final round of data cleansing upon joining real estate data and wage data, upon with duplicates were removed and less than 0.5% of unmatched data were discarded. Data timestamps were matched between wage and real estate, from which an annual representation of house price would be extracted. The final visualization contains a map with county gender pay and house data on the left. House price-to-pay ratio is then color scaled from minimum to 50, which means at maximum color it would take a person 50 years or more to buy a house with cash in that geographic area with our model and predictions. Color scale on map can be selected among total, female, and male. Our visualization also includes six line charts, each representing selected county pay over years, selected states pay over years, national pay over year, and pay differences in percentage points between female and male by county, state, and nation. Whenever a user clicks on a county, all 6 charts would simultaneously update with corresponding county and state. A bar chart comparing house prices to wage between gender over years is also available when hovering over on available county. Users can also select which years to show on the map and charts.

Figure 3 shows the final dashboard that the team delivered. Users can select years and genders to observe changes in wage difference at county, state, and national level. Figure 4 shows the list of available years to visualize on map. Figure 5 allows users to hover over different regions on map to observe house price to wage ratios from 2012 to 2025.

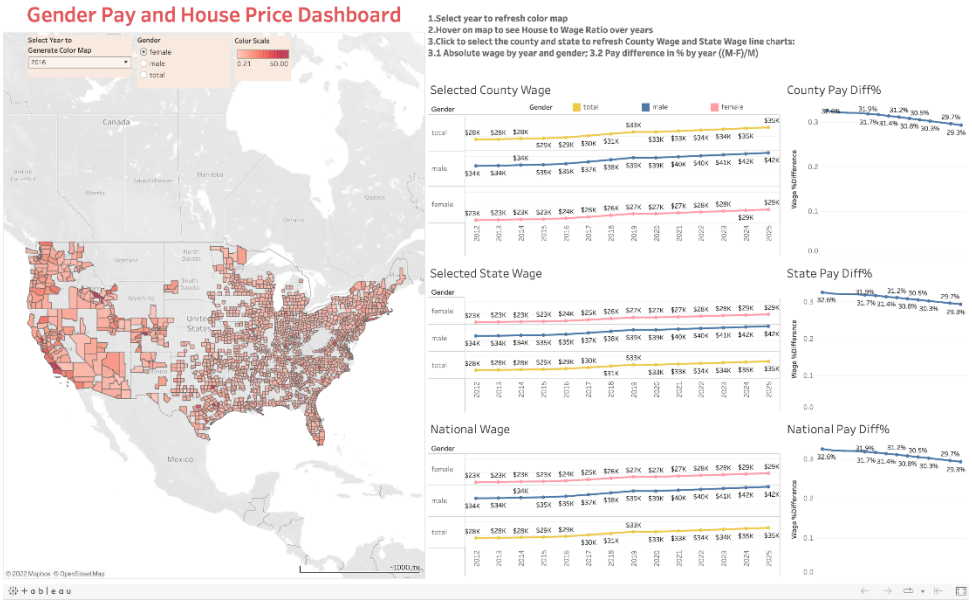

Figure 3. Gender Pay and House Price Dashboard

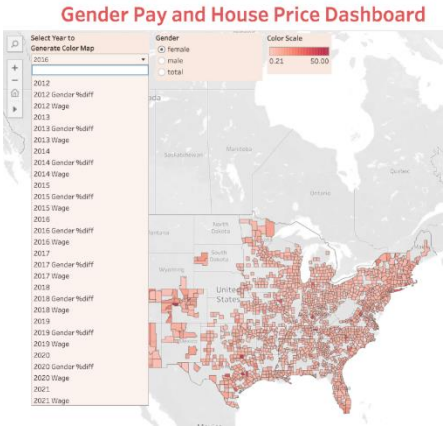

Figure 4. Navigation Drop Down

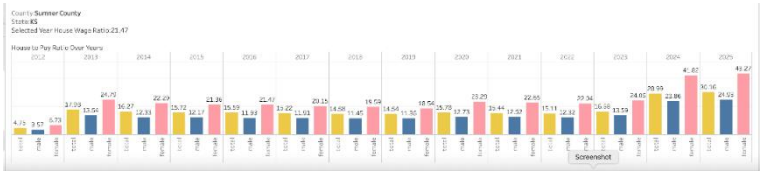

Figure 5. Map Hover Feature: house price to wage ratio

5. Experiments and Evaluation

Our product is accessible to the public and shared with researchers in economics and gender studies. We have a list of questions that our experiments are designed to answer, and we have concluded several observations throughout the experiment.

5.1 Questions and Observations:

Q1: How is the future GPG compared to historical GPG?

On a positive side, our data indicates that the national GPG is steadily decreasing. On the other side, the national GPG has been decreasing at the rate of about 0.23% per year. Starting from 32.6%, it will take over 130 years to eliminate it all, even which we all know as an extrapolation conclusion. The data suggests some urgent systematic change.

The trend for GPG change is noted to be reversed in certain regions. For example, in some fast-growing counties of CA, the GPG bumps up from the 20% range to the 30% or even 50% range.

Q2: How is the future housing price as compared to historical price and wage.

Our projections of future house prices on the US county level show that the prices will continue increasing in the next three years. Overall, the housing costs show an extensive range of change based on regions.

Q3: What is the year over year change of the GPG and how does it correlate with housing market changes?

As mentioned above, GPG has been slightly shrinking with a steady rate in next 3 years. While comparing to housing market, the affordability is expected to slightly worsen or stay flat in next 3 years for both genders depending on county. This increasing house to wage ratio in many areas, while it does not consider future impact of high interest rate and slow economy, indicates wage for both genders are not catching up with real estate based on historical data.

Q4: Any observations based on the past data that show a pattern in GPG or house price?

Though both male and female pay are on the rise over the past decade, the GPG is still present and prevalent, showing little signs of decrease (mostly less than 0.5% year over year). As for house prices, we have seen a steady increase over the years.

Our approach generated wage by-gender forecasts, and real estate price forecasts to calculate the ratio between the GPG and real estate price to reflect if females are bridging the gap with male in terms of pay and quality of living from a housing affordability perspective.

6. Conclusion and Discussion

For our project, we observed patterns and made predictions in GPG, real estate costs, and house price to income ratio. We created a dashboard through Tableau to allow users to easily interact with data by county, state and country and provided a direct visual comparison of house buying abilities between male and female.

To evaluate the dashboard, we conducted an anonymous survey of 24 researchers and students in gender study and economics. Among all survey respondents, 45.8% of responders were male and 50% of them were female. Overall, we received 91.7% positive feedback on our dashboard. Improvements such as rearranging the chart layout were made based on survey feedback.

GPG is an essential research topic because the decrease in GPG will increase productivity growth in all business sectors (Wolszczak-Derlacz, 2013). In addition, Females should pay more attention to GPG as female wages are lower than male across all sectors (Miler 2008). Learning more about the relationship between GPG and real estate costs will help us better understand how GPG impacted people in fields such as the housing market.

Overall, all team members have contributed a similar amount of effort during the project.

6.1 Future Research Recommendations

Ideally, we want to obtain as much data as possible for the GPG data, however, due to limitations in resources, we are only able to obtain the 10 years yearly summary data from U.S. Census Bureau. We recommend the future researchers to obtain a fuller dataset to make a more accurate prediction mode. We also suggest the future researchers to continue explore the GPG by occupation and invest additional time to build a more creative website to add more interactive tooltips and charts to improve user experience.

Reference

Press Release: Women could afford 18% more of the housing market if they made as much money as men. (2022). Dow Jones Institutional News.

Blau, Francine D., and Lawrence M. Kahn. (2017). "The Gender Wage Gap: Extent, Trends, and Explanations." Journal of Economic Literature, 55 (3): 789-865. DOI: 10.1257/jel.2016099

Murray. (2017). Interactive Data Visualization for the Web: An Introduction to Designing with D3. O'Reilly Media, Incorporated.

Taieb. (2018). Data Analysis with Python: A Modern Approach. Packt Publishing, Limited.

Lehmann, E.L. (2022). Testing Statistical Hypotheses, Springer Cham https://doi.org/10.1007/978-3-030-70578-7

US Census Bureau. (2022). Gender Pay Gap. Census.gov. Retrieved October 9, 2022, from https://www.census.gov/library/visualizations/interactive/gender-pay-gap.html

Payscale. (2022). 2022 State of the Gender Pay Gap Report. Retrieved March 15, 2022 from https://www.payscale.com/research-and-insights/gender-pay-gap/

Alkadry, M., & Tower, L. (2014.). Women in Public Service: Barriers, Challenges and Opportunities. Equality, Diversity and Inclusion an International Journal, 33(8), 806–808. https://doi.org/10.1108/EDI-07-2014-0059

Bishu, & Alkadry, M. G. (2017). A Systematic Review of the Gender Pay Gap and Factors That Predict It. Administration & Society, 49(1), 65–104. https://doi.org/10.1177/0095399716636928

Bacolod, & Blum, B. S. (2010). Two sides of the same coin: U.S. "residual" inequality and the gender gap. The Journal of Human Resources, 45(1), 197–242. https://doi.org/10.1353/jhr.2010.0006

Barry, J. (2021, 8). Real wage growth in the U.S. health workforce and the narrowing of the gender pay gap. Human Resources for Health, 19(105). https://doi.org/10.1186/s12960-021-00647-3

Captain, F. A. (2013). Six-Step Relational Database Design: A Step by Step Approach to Relational Database Design and Development. Createspace Independent Pub.

Goldsmith-Pinkham, & Shue, K. (2020). The Gender Gap in Housing Returns. https://doi.org/10.3386/w26914

Gould, E., Schieder, J., & Geier, K. (2016, October 20). What is the gender pay gap and is it real?: The complete guide to how women are paid less than men and why it can't be explained away. Economic Policy Institute. Retrieved October 11, 2022, from https://www.epi.org/publication/what-is-the-gender-pay-gap-and-is-it-real/

Liu, S., & Su, Y. (2022, 3). The Geography of Jobs and the Gender Wage Gap. The Review of Economics and Statistics, 1(27). https://doi-org.eu1.proxy.openathens.net/10.1162/rest_a_01188

Miler, P. W. (2008, 3). The Gender Pay Gap in the US: Does Sector Make a Difference? Journal of Labor Reserach, 30(1), 52-74. https://doi.org/10.1007/s12122-008-9050-5

Toledo, T. L. (2022). Forecasting Salaries and Wages using Auto Regressive Integrated Moving Average (ARIMA) Algorithm. International Journal of Computer Engineering and Sciences Research, 04(02), 18-24. https://www.ijcesr.com/Openaccess/v4i2/IJC856125481.pdf

Wolszczak-Derlacz, J. (2013). THE IMPACT OF GENDER WAGE GAP ON SECTORAL ECONOMIC GROWTH – CROSS-COUNTRY APPROACH. Equilibrium (Toruń), 8(3), 103-122. https://doi.org/10.12775/EQUIL.2013.022

A.Sterling, M. Thompso, S. Wang , A. Kusimo S. Gilmartin and S. Sheppard. (2020). The confidence gap predicts the gender pay gap among STEM graduates - PNAS. (n.d.). Retrieved October 14, 2022, from https://www.pnas.org/doi/10.1073/pnas.2010269117

Hirsch, B., König, M., & Möller, J. (2013). Is There a Gap in the Gap? Regional Differences in the Gender Pay Gap. Scottish Journal of Political Economy, 60(4), 412–439. https://doi.org/10.1111/sjpe.12017