Fortune Finders: The living wage expedition

Most Economical Places to Live: where workers make the most and pay the least, focusing on computer science positions

Phuc Nguyen, Nathan Kitts, Chris Struck
Winter 2024

Introduction: Finding the Treasure Map

The Background

An important goal for many people is being able to find a job that provides good compensation and a comfortable living standard. However, most job seekers who are entering the workforce these days face significant challenges to achieve this goal. This problem is even more relevant for new computer science graduates with limited working experience trying to navigate the current complex job market. While high figures income might be attractive at a glance, the real value of the compensation is tied to the cost of living within the job location. **Failing to take into consideration these geographic differences in living expenses can mislead individuals to take on positions that seem lucrative initially but end up failing to provide competitive compensation**. A high paying job in a high cost area like Manhattan, New York might not be as financially rewarding as a medium paying one in a much lower cost of living area.

With the **lack of accessible resources that integrate salary and cost of living metrics** for job seekers, it would be **difficult for one to successfully assess the actual financial comfort job opportunities** would provide. This knowledge gap can result in individuals taking on positions that do not financially align with their interests and needs, creating stress and dissatisfaction, leading to a lower career potential.

Why It Matters

The project goal is to **address these knowledge gaps and potential challenges** by conducting a comprehensive analysis of the real financial value of employment in different geographic regions around the US. We will integrate salary data, especially in the field of data science, with the associated cost of living. The goal is to **create tools that consider these important metrics and provide job seekers with the knowledge** they need to make informed decisions about their career choices and financial expectations.

The analysis will not only be an academic exercise but also serve as a practical resource for promoting wage equity. Our objective is to **enable job seekers with the ability to negotiate for fair compensation** that aligns with their financial goals and standard of living. Through the examination and presentation of data on the relationship between salaries and cost of living, the prospector should gain the confidence and leverage needed to negotiate a fair salary for their employment.

What is Needed

For this project, we will primarily utilize and examine the data tables from the **MIT Living Wage, Census, and Bureau of Economic Analysis (BEA)** to examine the relationship between typical salaries and cost of living across various roles, states, and counties. They would also the our guide to provide needed context on the meaning of different variables and their functions.

Note: All data used for the project are obtained from 2022.

Questions to answer

1. Which regions within the US would be the most ideal place to live in terms of salary-to-living wage ratio for a typical data science career?

2. How does the relative standard of living and financial comfort vary across different regions, both within states and across the country?

3. Are there any regional patterns or relationships that can help determine the most ideal places financially and comfortably for data science occupations?

Data Sources

	Bureau of Economic Analysis			
Dataset	**Regional Price Parities (RPP)**	**Personal Income by Industry (SAINC5N)**	**MIT Living Wage**	**Census Data**
Description	Presents regional cost comparisons across states and metro areas against the national average, covering housing and services for 2022.	Provides detailed income data across industries by NAICS, including GDP and employment figures.	Outlines earnings needed for a living wage in the US, detailing income and expenses for various family types and occupations.	Offers yearly demographic data for US citizens, with information on counties, metropolitan areas, and states.
Records	364	480	370,860	3,143
Year	2022	2022	2023	2023
Source	BEA RPP Dataset	BEA Personal Income	MIT Living Wage	US Census Data
Format	CSV	CSV	CSV, JSON	xlsx
Access	API	API	Scraping, Potential API	Web Tables

Methodology

Data Acquisition (BEA)

- Raw dataset is pulled from BEA using their API containing price parities between states and per capita personal income of each states. The API returns the data in JSON format.
- The JSON data is converted into Pandas dataframes.

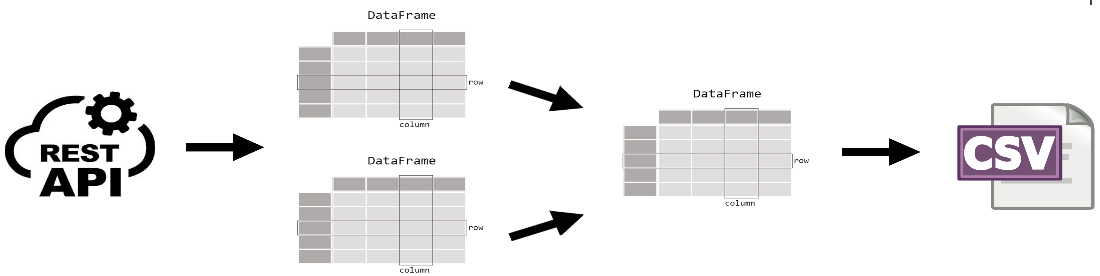
Data Cleaning (BEA)

- Merge both Dataframe into one based on the the state.
- Extract only the needed columns into a new dataframe. Just the columns for State, Year, Income, and Price Parities are kept.
- Convert the Year column from text to datetime format, then extract just the year as integers. This makes the year data easier to work with if needed.
- Convert the Income and Price Parities columns from text to floats (decimal numbers). This allows mathematical operations on them.
- Drop the row for District of Columbia. It's not needed for state comparisons.

Data Manipulation (BEA)

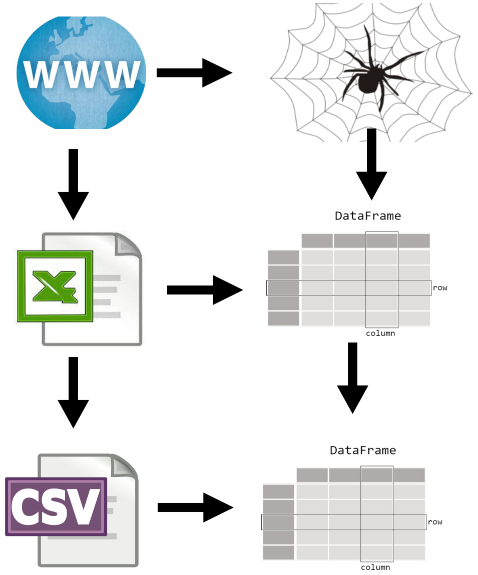
- Standardize the Income column by dividing it by the Price Parities column. This adjusts income for cost of living based on the BEA's dataset.
- Round the standardized income to 2 decimal places.
- Sort the data frame by standardized income in descending order. Puts highest at top.
- Export all three Dataframe to CSV.

Notes: All map was sourced from the US Census Bureau.[8]



Data Acquisition (MIT and Census)

- The data used for analysis was scraped from the MIT living wage website. utilizing a web crawler. The crawler accessed ~ 3100 pages and extracted the web hosted tables into CSV for each page. These CSVs were then compiled into a compound data set and subsequently into dataframes.
- Raw dataset is pulled from census using their table downloads option. The information gathered was the county population, which we used to consolidate the MIT data at a state level. The data was downloaded as an excel spreadsheet. The spreadsheet was converted into Pandas dataframes.



Data Cleaning (MIT and Census)

- MIT data had info about poverty wage, and state level taxes(earnings before and after we kept %income taxed) that we dropped. Occupations were filtered for computer science only. Computer science data was provided at a state level and did not need to be weighted.
- Census data had state county naming conventions that had to be reconciled. Header, footer, and DC rows were dropped.

Data Manipulation (MIT and Census)

- County population was merged into the MIT data frames from the census on county and state name. MIT crawler had some different county names, so state level projections were created utilizing population to weight values. price parities were then merged to the state level MIT data. Parity ratio was then utilized to recalculate wage information for state to state comparisons.

Finding the Treasure:
In which US states does your $ go furthest?

Regional Price Parities(RPP) are defined by the BEA as **weighted price indexes** used to account for **geographic cost of living differences** when comparing economic data across areas.[1] They show how prices in one area compare to the overall national average. An RPP of **100** represents the **national average**. An RPP of **115** for a state means prices there are **15% above average**.

To **compare per capita personal income amounts across states**, the nominal income can be **adjusted using RPPs**. For our analysis, RPPs were converted to a ratio of 1 instead of an index of 100 prior to income adjustment. We **adjusted the nominal incomes by dividing them by the RPPs ratios** to deflate based on each state's price deviation from the average. States with higher price levels will see incomes adjusted downward more. This results in "real" incomes adjusted for cost of living differences. Comparing these adjusted incomes **allows an apples-to-apples comparison** of the equivalent purchasing power of an income amount across states.

After adjusting for prices, the 5 states with the highest per capita personal incomes are:

1. **Wyoming:** $79,701
2. **North Dakota:** $79,363
3. **Connecticut:** $77,940
4. **South Dakota:** $77,481
5. **Massachusetts:** $77,300

Wyoming and North Dakota, which were among the top 10 in nominal incomes, moved to the top after adjustment, suggesting the lower cost of living plays a large factor in how high the "real" incomes are.[6]
To **focus specifically on incomes for computer science occupations**, additional data on average salaries by state from the US census was used. Adjusting these computer science salaries for price parities using the same method stated above with RPP, the top 5 states are:

1. **Washington:** $144,033
2. **Virginia:** $125,363
3. **New York:** $111,716
4. **California:** $109,579
5. **Connecticut:** $109,137

States like Washington and Virginia, which have high computer science salaries after adjustment, may offer better spending power for professionals in this field. This result stayed in line with the expectation that **states with a high concentration of tech firms would be ideal for computer science occupations** despite higher than average price parities.

Accounting for purchasing power/cost of living differences by adjusting incomes using the BEA's Regional Price Parities **provides a more accurate comparison of purchasing power and living standards across states.** The adjusted incomes for computer science occupations show different top states than just looking at the national average salaries. This methodology reveals different rankings than just looking at unadjusted per capita incomes.

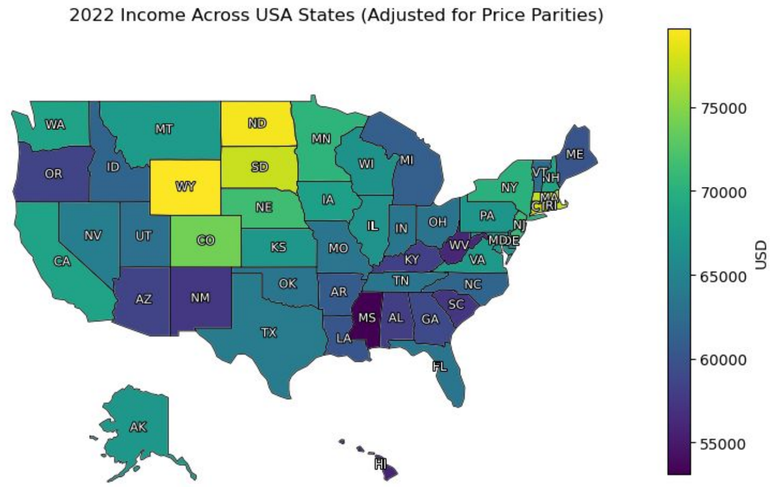


Figure 1: Map of average income across USA after adjusting for price parities.

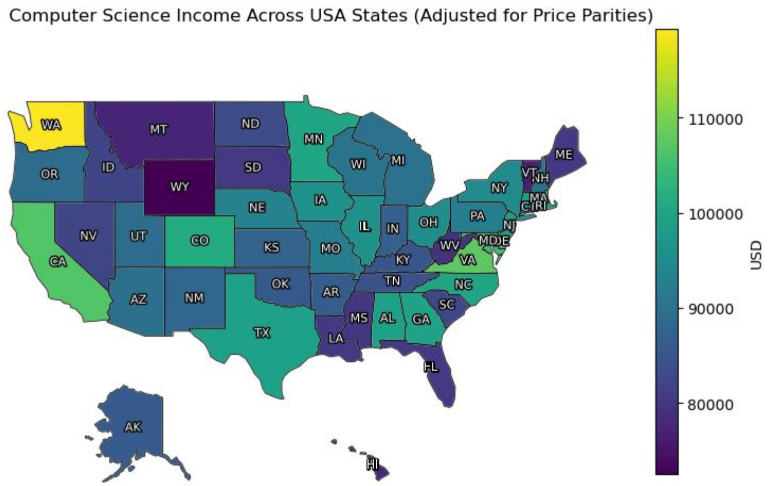


Figure 2: Map of average computer science income across USA after adjusting for price parities.

Comparing BEA Salary to MIT's Wage and Computer Science Salary

There is unfortunately almost **no correlation between the estimated living wage data for each state and state income**, shown in the table below. The difference between states depending on Living Wage and Cost of Living is also almost zero. The range for Average Wage divided by Average Cost of Living shown to the right is between 1.001 and 1.002. However, this does show that for three states in the general top 5 adjusted salary categories, these are also near the best from a living wage to cost of living standpoint: North Dakota, South Dakota, and Wyoming.

Because MIT's living wage is calculated based off of the cost of living, these results show perhaps in a more telling way that **average salary in general does not correlate with cost of living across states**. On the other hand, the average computer science salary by state does positively correlate at medium strength with cost of living. While this was also provided by MIT, it is in the same measurement as the general salary, provided by BEA, as average statewide salary.

Correlation Matrix

	Income v Parity	Wage Average	Cost Average	CompSci Salary
Income v Parity	1	.271	.271	.259
Wage Average	.271	1	1	.669
Cost Average	.271	1	1	.669
CompSci Salary	.259	.669	.669	1

Figure 1 *represents the correlation between living wage, cost of living, income adjusted by price parity, and computer science salary*

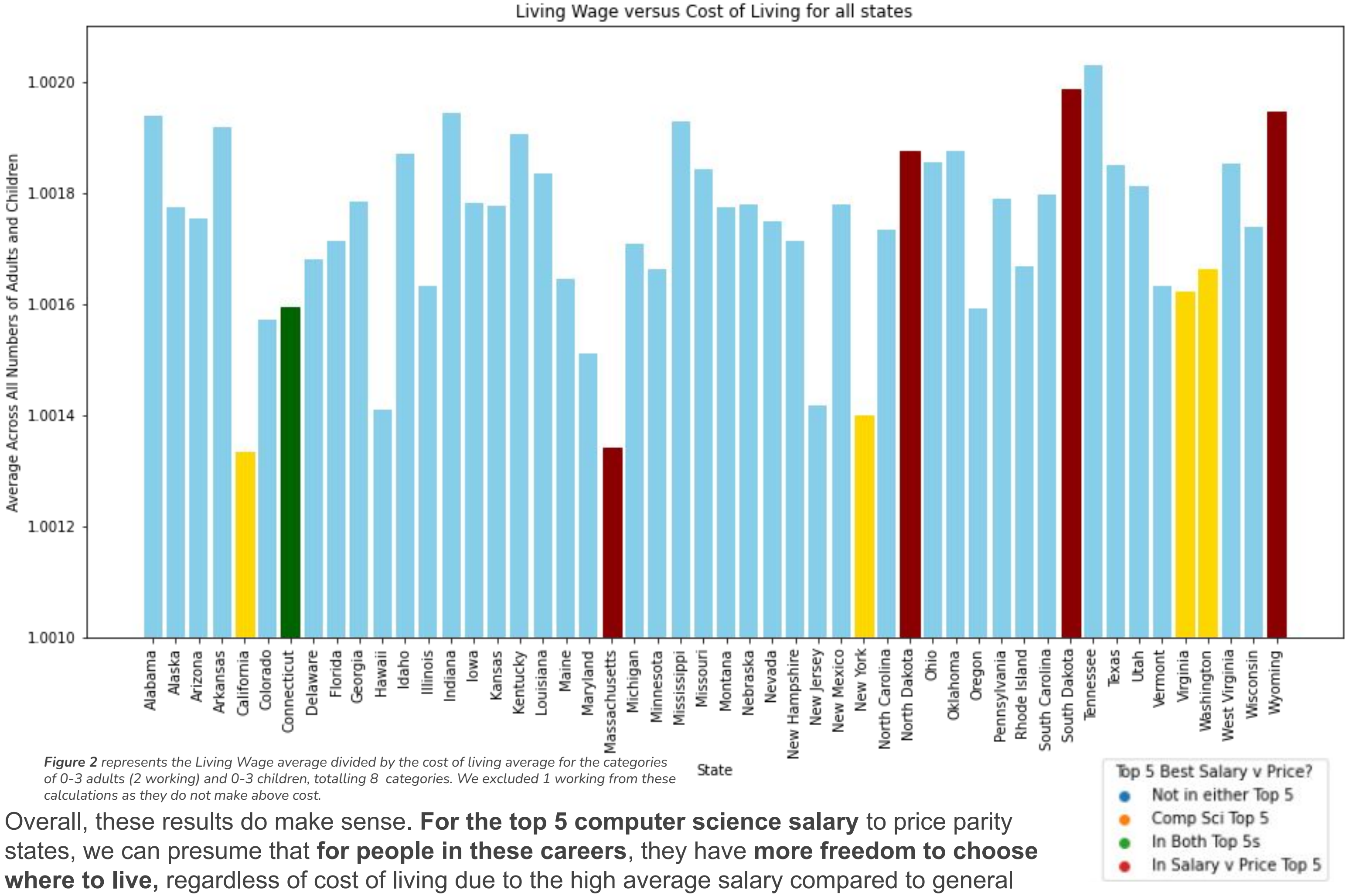


Figure 2 *represents the Living Wage average divided by the cost of living average for the categories of 0-3 adults (2 working) and 0-3 children, totalling 8 categories. We excluded 1 working from these calculations as they do not make above cost.*

Overall, these results do make sense. **For the top 5 computer science salary** to price parity states, we can presume that **for people in these careers**, they have **more freedom to choose where to live,** regardless of cost of living due to the high average salary compared to general salaries. Based off of this analysis, **we decided to focus more specifically on Computer Science Salaries** (and not living wage) **versus Cost of Living** for the states with the highest Computer Science Salary versus Price Parity as shown on slide 5. Living wage turned out to be a less reliable income metric as it is directly connected to cost of living.

Investigating Computer Science Salaries versus Cost of Living

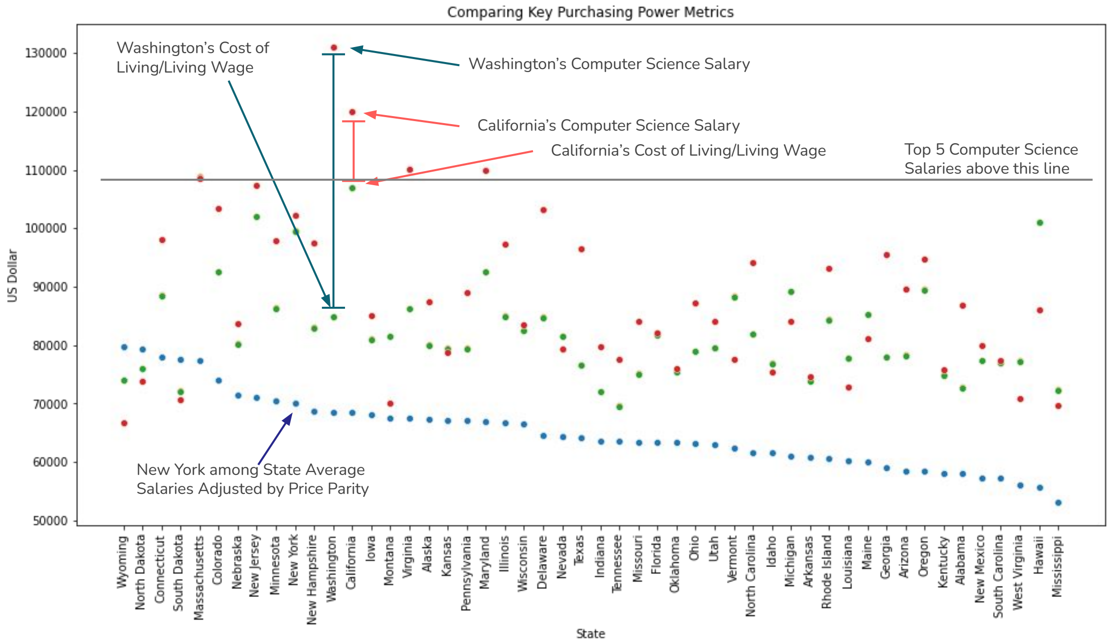


Figure 1 shows the salary adjusted by price parity (blue) plotted against average wage (yellow), cost of living (green), which is just below wage data for each state, and computer science salary (red)

Looking in more detail between states on computer science salary versus adjusted salary (also shown in the heat maps on slide 5) and living wage/cost of living (shown in direct comparison in the bar graph on slide 6), we can see more s**ignificant differences between states due to how much impact the higher computer science salary has**. For example, at the difference between Washington's Computer Science Salary versus its Cost of Living and the same metrics for California, it is clear that **Washington is a much better place for unlocking the dollar value of one of the highest salaries**. For most states, it should be noted that the average adjusted salary for one person is below the average cost of living. Perhaps, paradoxically, it can be seen to the left that some of the states that would be expected to be the highest cost of living like New York, are in fact middle of the road and also among the highest adjusted salary. This leads us to the next step of our investigation where we look in more detail at key states at a county level to decipher where the true best place to live is within these states for seeing how far one can stretch their dollar. For our analysis, the top 5 states by computer science salary will be our focus.

Investigating these states further, we can see that while computer science salary, adjusted average salary (shown by state on slide 5), and adjusted wage follow a similar pattern for these five states, the significant variability (shown on slide 6) between cost of living from state to state, demonstrates **the importance of where a person lives to the value of their higher salary**. While these positions tend to pay the best by a significant margin in these states, the cost of living is also significantly higher for California, New York, and Connecticut than Washington.

With the rise in remote work and from-home hybrid models, there is now more than ever the flexibility to move and decide on places to live. The New York Times wrote an article on the topic stating that: "Employees who only need to be in the office two or three days a week can tolerate longer commutes in return for a home office and more outside space."[2] **A deeper look into the geography and cost of living for the top five states** with the highest salary for computer science occupations can provide more insights on which counties would be the most ideal place within each state. We investigate this further on the following slide.

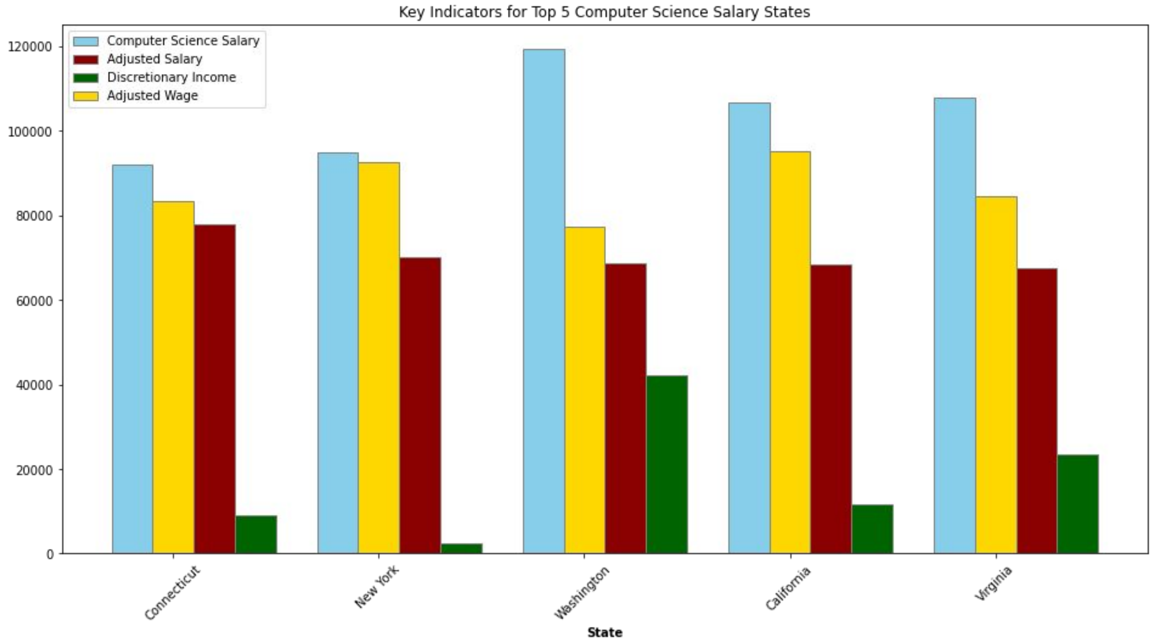


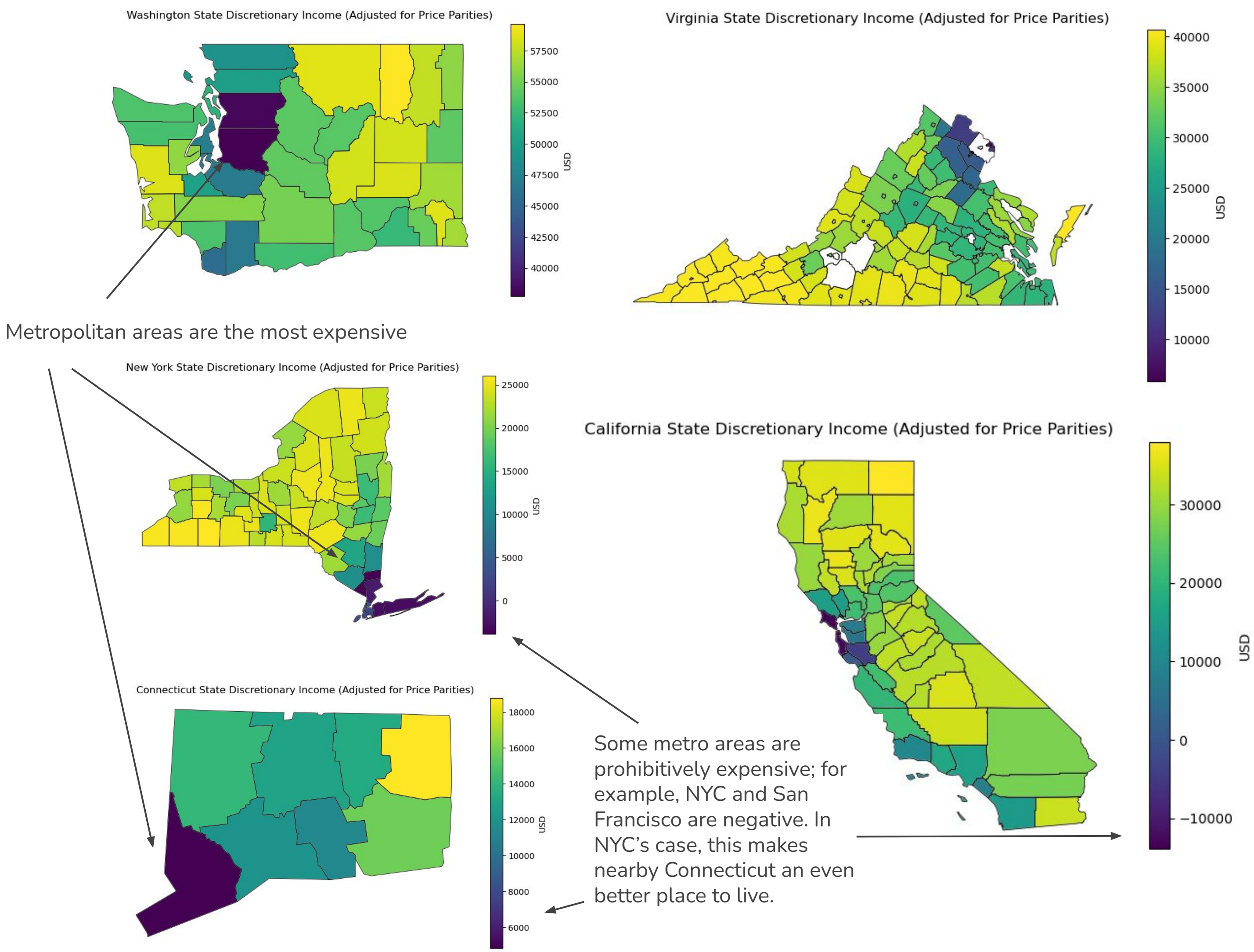
Figure 2 shows the key indicators against each other; computer science salary. average salary, and living wage, adjusted by price parity. Discretionary Income is calculated as Computer Science Salary minus Cost of Living, then adjusted for price parity

County Level Detail using MIT’s Cost of Living Data

To assess the relative standard of living and comfort level across different counties across the US, an **adjusted discretionary income metric** was calculated. This metric calculates **how much income is left over after typical living costs are paid** in an area, which allows us to **compare purchasing power and standard of living** across different geographic locations on an equal footing. Similar to prior analyses, regional price parities (RPPs) were converted to ratios before adjusting incomes. The income per state data came from the U.S. Census, while county-level cost of living estimates utilized the MIT living wage dataset. By **subtracting the local cost of living from income**, the discretionary income for each county was derived. This was then **divided by the corresponding RPP** ratios to adjust for cost of living disparities between counties. The resulting adjusted discretionary income figure allows for a **direct comparison** of purchasing power and standard of living across counties across all states. Counties with a **higher adjusted discretionary income** indicate a **greater level of financial comfort for residents** when considering both earnings and expenses in that specific location.

Shown to the right are the top five states for computer science adjusted discretionary income compared between counties. While the values for each county are adjusted to allow for comparison across the US, the **choropleth color scale is set to be relative within each state** to analyze regional relationships. Within states, there's a **clear pattern** if you meet typical state salary expectations you achieve **more net take-home pay the further you are from metropolitan areas**. At the state level, it's clear that **Washington has the tightest and highest range** for Computer Science adjusted discretionary income, so it stands out as a **good place to work**, especially for those with computer science positions.

When reviewing the data, we found not only that **metropolitan areas regardless of state are the most expensive**, but that they can be prohibitively expensive in some states. This is primarily **due to the cost of raising kids** as shown by MIT Living Wage data[3]. As the **number of children increases from one to three**, the required salary to support families of this size begins to exceed the average living wage across the same categories in places where the cost of living is abnormally high.



Figures 1-5 *represent the top 5 Computer Science Salary states broken down by computer science adjusted discretionary income across counties*

Summary:
The map leads to Washington for Computer Science

As we investigated the best states in general, we found that the center of the country, centering around the Dakotas was generally the best place to live from the perspective of average state salary compared to price parity. However, **we were focused more specifically on Computer Science salaries** from the start, given this is our intended field of work, post-graduation. We found that when adjusting for these higher salaries, **a new top 5 emerged** with more states along the coast, **including New York, California, and Washington**. Connecticut was in both top 5s.

When diving further into the county data available from MIT, **we found that the average statewide computer science salaries had a distribution more similar to the average cost of living** from MIT with a 0.67 positive correlation. **This shows somewhat that these higher salaries are more in line with the expected regional costs**. However, when adjusting these by price parity, **we found that the discretionary income varied wildly among the top 5 computer science salary states**.

At the county level, the differences were even more dramatic. Not only did we find the expected result that **metropolitan areas in these states were the most expensive**, but we also found that **in California and New York, especially, these areas were cost-prohibitive** and even negative for discretionary income, even for these higher average salaries. Contrary to this trend, as one can see to the right, Washington metropolitan area around Seattle remains near positive USD 40,000. This is even more significant, given the fact that these are for single salaries marked against the estimated costs averaged from 1 to 5-person households, across MIT's categories, which include the expected costs of up to 2 adults and 3 children.

In conclusion, when weighing the pros and cons of a particular place to live, the higher salary makes a meaningful difference to living quality, even in places with high expected cost of living like California. **Choosing a high-salary position is therefore the key first step to making a decision on where to live**. From there, it is then important to consider how best to unlock the cost of living and price parity of your local region. Ideally, **for a computer scientist, a rural area near Washington State's metro area around Seattle would be best**. If you have to take a job in New York though, for example, it may make sense to instead move to Connecticut and commute, depending on other factors such as whether you want to raise kids as this is a key driver of higher costs.

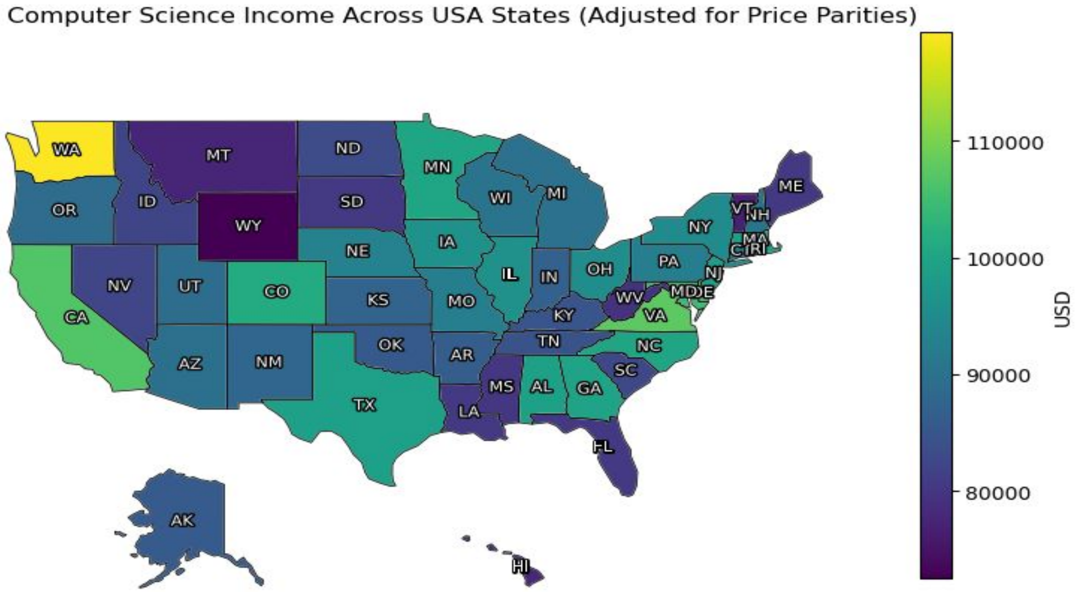


Figure 1 maps of average computer science income across USA after adjusting for price parities.

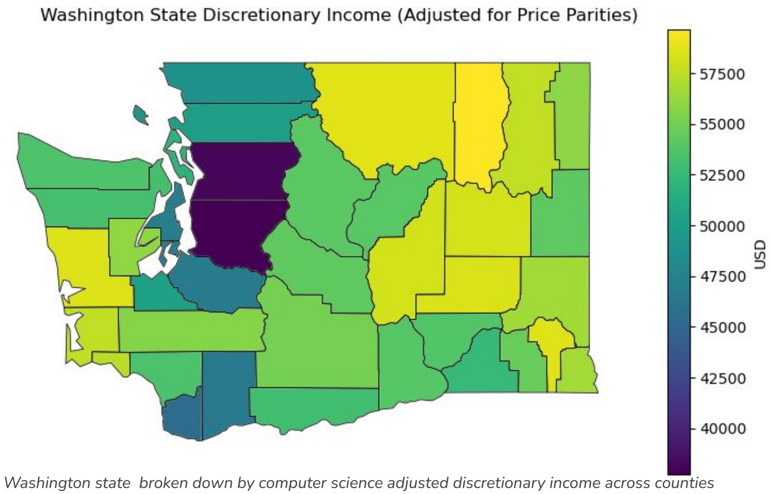


Figure 2 represents Washington state broken down by computer science adjusted discretionary income across counties

Limitations

Price Parities

The U.S. Bureau of Economic Analysis describes limitations in their document, stating that the **Regional Price Parities (RPPs) come from surveys by U.S. federal agencies**, such as the Consumer Price Index (CPI) from the Bureau of Labor Statistics and the American Community Survey (PUMS) from the U.S. Census Bureau. These surveys look at prices and consumer spending across different places. One problem is that the **surveys don't always have enough respondents from every area**. Some places have many people answering the survey, but other places only have a few. This makes the data for those with **lower sample areas less accurate**. Furthermore, the **original CPI surveys were not designed for place-to-place comparisons**, which leads to volatile results for areas with smaller sample sizes or uneven sampling distributions. To make the datasets comparable, they **must be aligned to a common unit by allocating them to the county level using housing or income shares**. Previous RPP estimates averaged CPI results over 5 years, but the current methodology uses annual data.[7] These **limitations can skew the calculations and comparisons** between adjusted per capita personal incomes and adjusted discretionary incomes. Such inaccuracies could then lead to flawed rankings or comparisons of adjusted incomes across states or counties, **potentially producing unreliable conclusions** about which areas offer the best purchasing power or standard of living for different occupations. To mitigate some of these limitations, the RPP data quality was closely examined and cross-referenced with MIT living wage data to filter out potential outliers. However, given these constraints, it **would be more beneficial to further investigate the accuracy of RPPs for specific regions** before drawing concrete conclusions.

MIT Living Wage

Like the BEA's RPP, The basis for **expenditures calculated by MIT stem from U.S. federal surveys**, such as the consumer expenditure survey from the Bureau of Labor Statistics (BLS) BLS, and the Cost of Foods Report from the U.S. Department of Agriculture (USDA). As such it also likely suffers from **Response bias in survey data**.

Our team initially **utilized a custom-built scraper** that functioned as intended**,** allowing us to capture significant data. However, after an initial data retrieval, we discovered **MIT's request not to scrape** on their website. Although **MIT offers data sharing** upon request, our request coincided with their annual survey update in January, leading to a **delayed response** on their part. Consequently, **we relied on the collected data**, aligning with our expectation of accessing the API before project submission, and **chose not to scrape further data to adhere to MIT's request.** Efforts therefore to refine the scraper were halted which led to **missing data**. **Naming inconsistencies** in states like Virginia to miss or duplicate certain counties. **25 counties out of the total 3143** across states were missed this impacted 12 states, notably Alaska Boroughs and Virginia "City Counties", **necessitating adjustments at state level,** via the Census, to **account for these gaps.**

We observed a **lack of detailed salary information** for various occupations at the **county level**. This omission likely reflects the sparse distribution of some jobs across regions, protecting against individual identification. Therefore, **we decided against collecting this information** as well.

Final Considerations

The **geographical insights** derived from our analysis, while intriguing, are **not yet suitable for career making decisions**. Our findings are based on **state-level salary averages, which may not accurately reflect the job market in every county.** Moreover, our analysis does not account for the variability in job availability, laws, and cultural factors affecting salaries across states and counties. Lastly, considerations such as commuting across counties or states, the practicality of living in underpopulated areas, and the impact of local laws and culture on salaries were outside the scope of our current analysis.

Future studies might include scraping job boards for available roles, allowing for more accurate salary estimates for counties, and analyzing traffic patterns to better understand commute implications, offering a more comprehensive view of the job landscape.

Statement of Work and References

Work was highly collaborative. Team met frequently and kept consistent communication through slack. Each member provided substantial contributions to all parts the project. Team members had their own notebooks and collaborated via GIT for code. Code was compiled into one Notebook at the end. Below is a further breakdown of individual responsibilities.

Phuc:
Provided domain research and literature review. brought domain knowledge of Regional Price Parity (RPP) to group. Collected and cleaned RPP data. Researched tools specifically geopandas. Generated choropleth visualizations and provided authorship and feedback of choropleth slides.

Nathan:
Initiated project and provided domain knowledge from prior familiarity with the MIT Data. Built the web scraper that pulled the MIT data. Assisted with cleaning and manipulation. Assisted with code blocks for visuals. Reviewed, validated and corrected code across notebooks. Final Compilation of Code. Provided feedback and editing of final report.

Chris:
Project Management, lead communication and organization. Data preparation of MIT Data, Cleaning: Combining tables and handling missing values, Manipulation: Merging tables, creating visualizations. incorporated census data and utilized to handle missing values in MIT Data. Primary writer and editor of final presentation

[1] U.S. Bureau of Economic Analysis. (2022). Regional price parities by state and metro area. Retrieved January 19, 2024, from https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area

[2] Barrero, J. M., Bloom, N., & Davis, S. J. (2022, December 22). Working from home is the trend of the year—and next year too. *Time*. Retrieved from https://time.com/6243148/working-from-home-is-the-trend-of-the-year-and-next-year-too/

[3] Massachusetts Institute of Technology. (2023). Living wage calculator methodology. Retrieved January 19, 2024, from https://livingwage.mit.edu/pages/methodology

[4] Nadeau, C. A., Glasmeier, A. K., & Swalina, C. (2023). Living wage calculator: Living wage calculator: User's guide / Technical notes (2022-2023 Update) . Massachusetts Institute of Technology, Department of Urban Studies and Planning.

[5] U.S. Census Bureau. (2023). County population totals and components of change: 2020-2022. Retrieved from https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html

[6] U.S. Bureau of Economic Analysis. (2022). Personal Income by Major Component and Earnings by NAICS Industry (SAINC5N). Retrieved January 19, 2024, from BEA's Interactive Data Tables.

[7] U.S. Bureau of Economic Analysis. (2022). Methodology for Regional Price Parities, Real Personal Consumption Expenditures, and Real Personal Income. Retrieved February 1, 2024, from https://www.bea.gov/system/files/methodologies/Methodology-for-Regional-Price-Parities_0.pdf

[8] U.S. Census Bureau. (2018). Cartographic Boundary Files - Shapefile. Retrieved February 1, 2024, from https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html