

Analysis of Minimum Wage and Cost of Living by County

Todd Estep
DSC 530 Data Exploration and Analysis
Professor Metzger

03/02/2025

Introduction to Dataset

For this analysis I mostly used data from the Economic Policy Institute, with a focus on their minimum wage by county and cost of living by county.

I also used data from the Economic Research Service department of the US Department of Agriculture. I retrieved median household income data by county from here.

There are 3,144 counties in America, so my dataset has 3,144 data points for each variable.

For more information, please see the Works Cited page at the end.

Variables Analyzed

min_wage - The minimum hourly wage by county

a_min_income - The annual income of a full time minimum wage worker. Calculated by multiplying the min_wage by 2080 (40 hours per week * 52 weeks per year)

a_total_cost - The annual total cost of living for a single adult with no children, by county. The sum of housing, food, transportation, health, tax and miscellaneous other costs necessary for living..

a_min_ratio - The ratio of annual income of a minimum wage worker, and the cost of living of a single adult with no children. Calculated using $a_min_ratio = \frac{a_min_income}{a_total_cost}$

median_household - The median annual household income by county

median_ratio - The ratio of annual household income and cost of living. Note, household income may include multiple earners and additional dependents, while the cost of living only accounts for a single person, but this was necessary for the scope of this project. Calculated using $median_ratio = \frac{median_household}{a_total_cost}$

Calculated using the mean, std and mode functions in the Python pandas package.

Summary Statistics by variable

min_wage

Mean = \$9.71 per hour
Standard Deviation = 2.91

Mode = \$7.25

a_min_income

Mean = \$20,199.77 per year
Standard Deviation = \$ 6053.51

Mode = \$15080.00

a_total_cost

Mean = \$44989.54 per year
Standard Deviation = \$5894.57

Mode = \$48384.41

a_min_ratio

Mean = 0.4495
Standard Deviation = 0.1233

Mode = 0.3117

median_household_income

Mean = \$63,240.51 per year
Standard Deviation = \$16,248.92

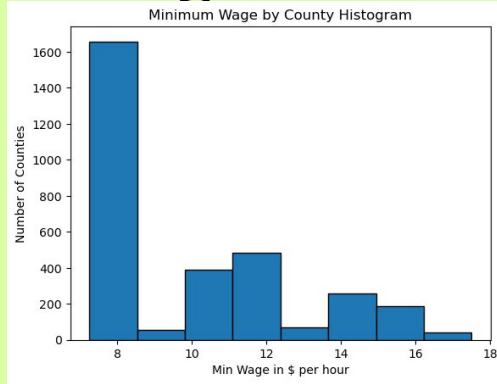
Mode = \$70,148

median_ratio

Mean = 1.3981
Standard Deviation = 0.2574

Mode = 1.4498

Histograms

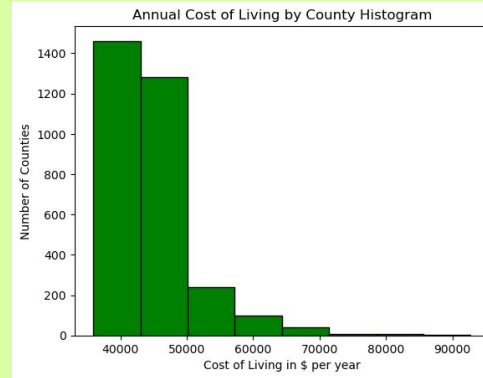


There are 4 counties with a minimum wage greater than 17, which are considered outliers. I think including them in the data is important.

Outlier Counties:

Los Angeles County, CA, San Mateo County, CA, Washington DC, and Montgomery County MD

This has a right tail.

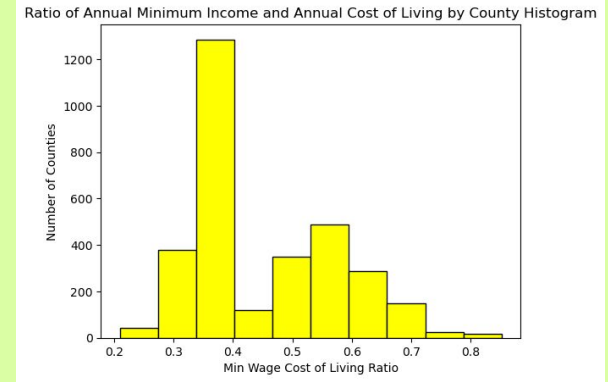


There are 72 counties with an annual cost of living greater than \$62,673.40. This is 3 standard deviations above the mean.

Many of these counties represent metropolitan areas with high populations, I will keep them.

No outliers on the minimum end.

This has a significant right tail



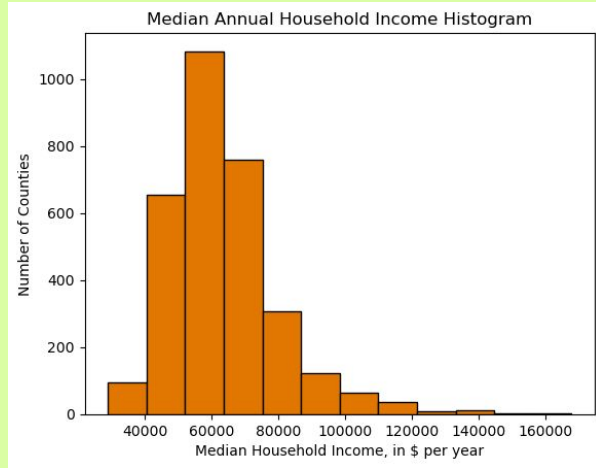
There are 20 counties with a Minimum Income and Annual Cost ratio above 0.758, which is 2.5 standard deviations above the mean. 19 of them are in Washington state, and 1 in Maryland.

I will keep these.

No outliers on the minimum end.

This has a right tail.

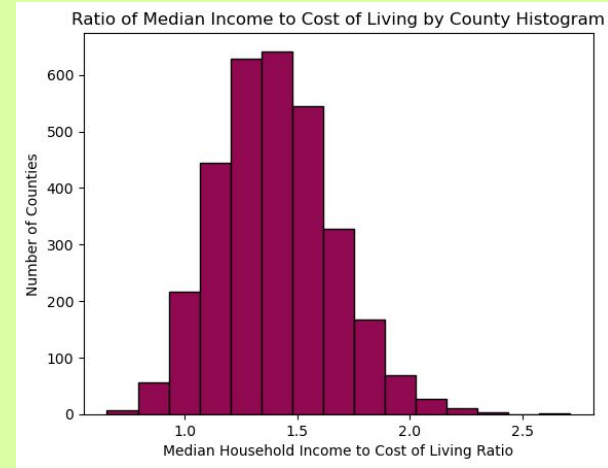
Histograms continued



There are 55 counties with a Median Annual Household Income over \$111,988. This is 3 standard deviations above the mean.

I will keep these datapoints, some represent counties with high populations.

This has a right tail.

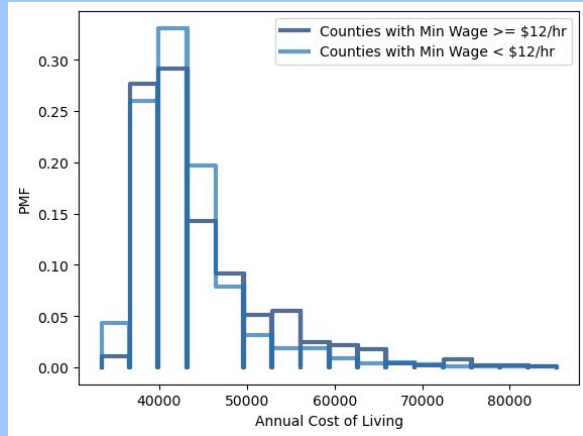


There are 12 counties with a Median Income to Cost of Living Ratio above 2.18. This is 3 standard deviations above the mean.

There are 6 counties with a Median Income to Cost of Living Ratio below 0.77, this is 2.5 standard deviations below the mean.

I will keep these data points.

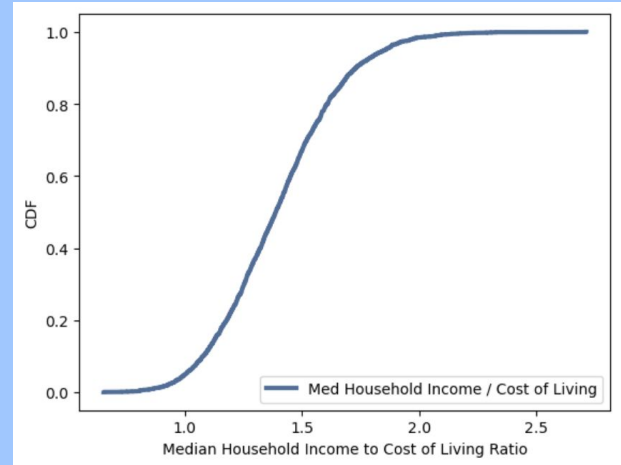
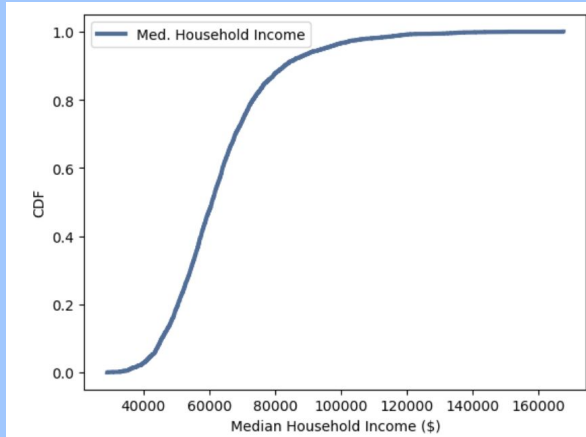
This has a slight right tail.



One common argument for keeping the minimum wage low is that it helps keep the cost of living low. However, this PMF shows no significant difference in cost of living for counties with higher minimum wages and those with lower minimum wages.

Annual Cost of Living PMF

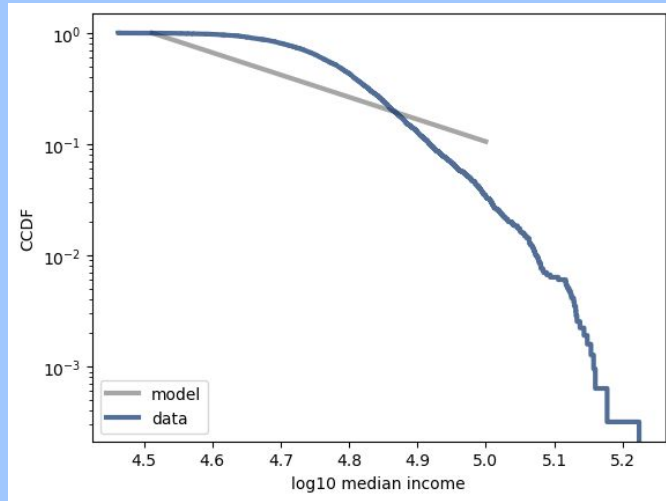
Median Income CDF



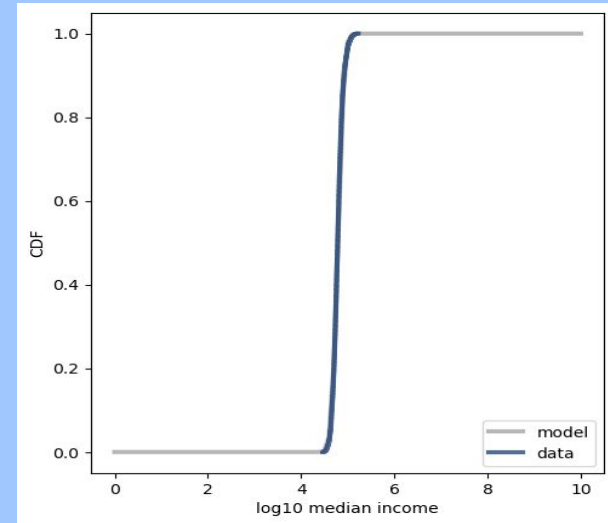
I wanted to compare the CDFs of the median household income and the median household income cost of living ratio. The ratio CDF has a slightly lower slope, but overall they take a similar shape. I think one of the conclusions this comparison hints at is that some counties with higher median household incomes have higher costs of living associated with them.

Analytical Distribution

Pareto Distribution Model

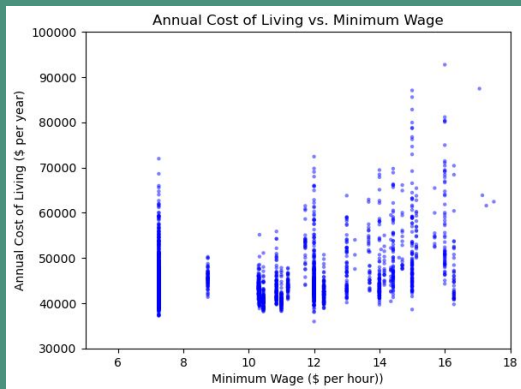


Lognormal Distribution Model



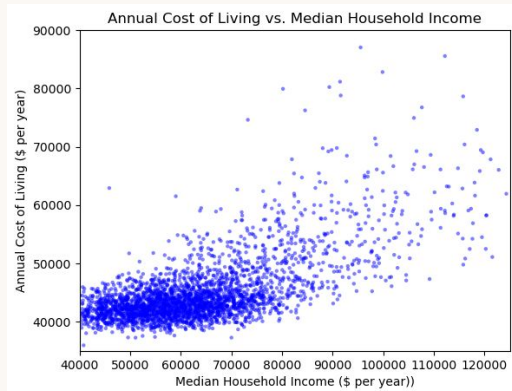
The Pareto Distribution Model does approach the data, but does a rather poor job of modelling it accurately. The Lognormal Distribution Model fits the data almost perfectly.

Cost of Living vs Minimum Wage



Pearson's Correlation: 0.3919
Spearman's Correlation: 0.3051

Cost of Living vs Median Income



Pearson's Correlation: 0.6967
Spearman's Correlation: 0.5506

The minimum wage has a weak positive correlation with the cost of living.

The median household income has a medium positive correlation with the cost of living.

In both cases, Pearson's Correlation was a stronger correlation, suggesting a better fit in a linear model. But a linear model would still not be a great fit.

Scatterplots and Correlation

Hypothesis Testing

I used the permutation test to test whether the observed correlation was statistically significant.

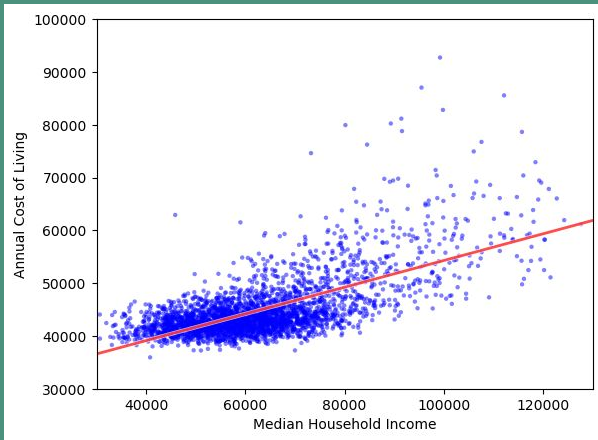
In both tests, annual cost of living vs. minimum wage and annual cost of living vs. median household income, the results returned a p-value of near 0, meaning in 1000 trials for each, we didn't see a correlation under the null hypothesis that exceeded the observed correlation.

This means our calculated correlation coefficients are statistically accurate and should be believed.

Regression Analysis

Annual Cost of Living vs. Median Household Income made more sense to do a regression analysis because it had a stronger correlation.

The red line of least squares follows the equation $y = 0.2527x + 29005$



Visually it looks like it goes through the center of the data and follows its flow.

Works Cited

Economic Research Service - County level datasets. (2025, January 31). Retrieved March 2, 2025, from

<https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data>

Family Budget map. (2025, January). Economic Policy Institute. Retrieved March 2, 2025, from

<https://www.epi.org/resources/budget/budget-map/>

Minimum Wage Tracker. (2025, January 1). Economic Policy Institute. Retrieved March 2, 2025, from

<https://www.epi.org/minimum-wage-tracker/>