**Blog Link:** <https://medium.com/@mkwils/privilege-and-poverty-the-wage-gap-affordable-housing-and-homelessness-in-america-a6ead139a04e>

**Title:** Privilege and Poverty: The Wage Gap, Affordable Housing, and Homelessness in America

**Introduction**

The dataset was created to tell a story, one where all the parts and pieces may be things that people know about, but by bringing the pieces together, a more nuanced picture might be revealed. One can look at the cost of shelter in a city, and separately look at the incomes for people in that city, but looking at them alongside estimates of the number of people in those locations experiencing housing insecurity may reveal whether the people can afford to live there.

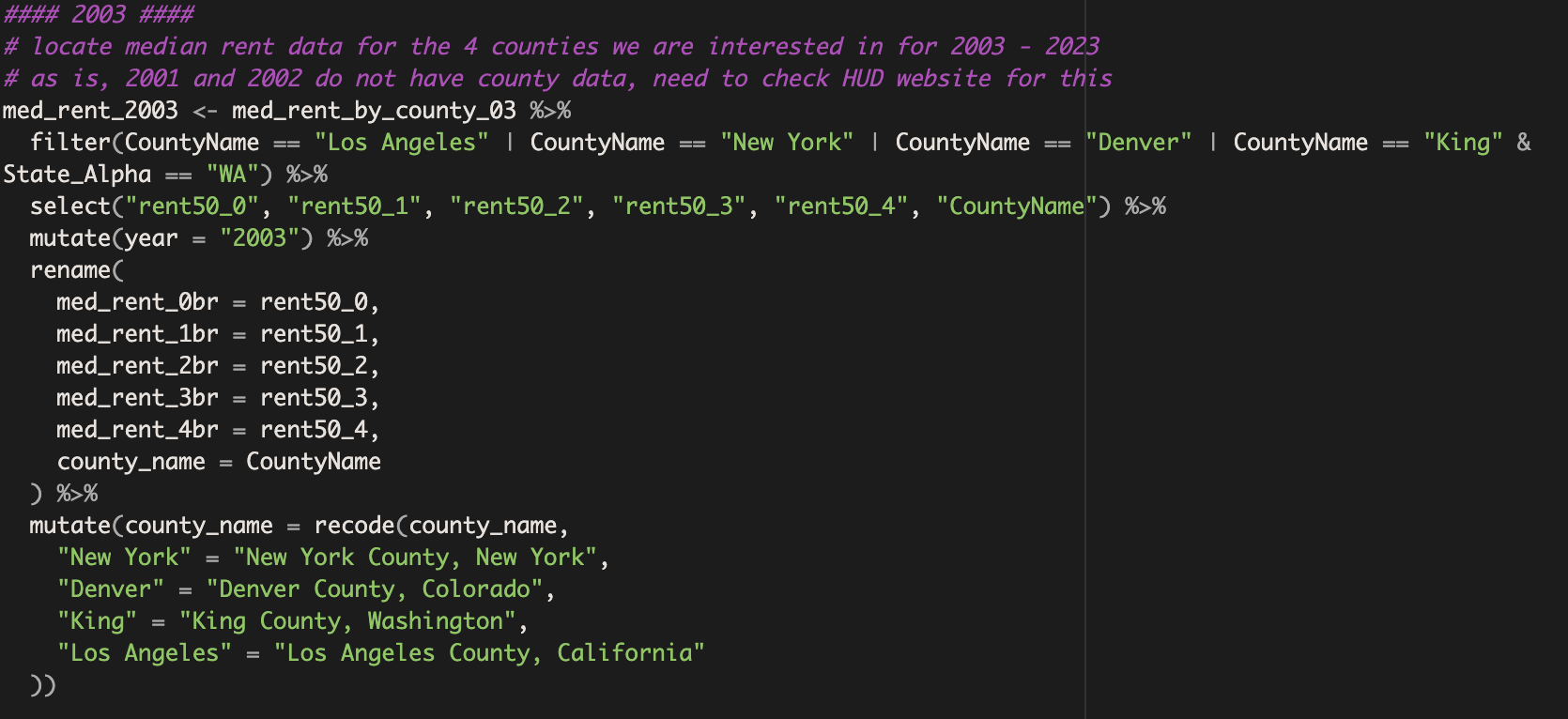
The original idea for this project was to look for a connection between the number of Airbnbs in an area, the cost of rent, and the number of unhoused people. The popularity of Airbnbs has caused some areas to have fewer places available for rent, and potentially contributed to increases in rent and fewer people being able to afford living there. After searching for the available data, the project took a slightly different direction. Airbnb historical data was not readily available without paying huge sums for access. Without this component, we collected data on wages, rent, and population counts of unhoused people to see if any connections existed.

The areas we chose to analyze encompass four of the five cities with the highest population of unhoused people in the United States and three were in the top ten most expensive cities to live in 2023. By analyzing data from multiple angles, the blame for homelessness can be shifted away from the individual as various factors are considered in this complicated story.

**Dataset**

The information to build our dataset came from three different sources. The data for **annual average wage** comes from the Quarterly Census of Employment and Wages performed by the U.S. Bureau of Labor Statistics [(QCEW Data Files, 2024)](https://www.zotero.org/google-docs/?I0Wt2r). The data was collected by tabulating all wage data collected from administrative data reported by employers. We utilized the files that are available by county for the years 2000 - 2022. According to the handbook of methods, this data represents 95% of all jobs in the United States. It is important to note that for this data, wage is inclusive of not just take-home pay, but total compensation [(Overview, 2023)](https://www.zotero.org/google-docs/?yruR2Y). The QCEW program has made their data available online in CSV format since 2014, although the Bureau has been publishing annual reports of employment and wages since 1975 [(History, 2023)](https://www.zotero.org/google-docs/?itZQiQ). This data was well organized and column titles were consistent throughout the various files, so no data cleaning was required. The files were loaded in using both the git2r and git2rdata libraries to make a temporary copy of the directory on our machine, using dir\_ls to search for the folder and files within we want, and then finally using read.csv to combine the files all into one data-frame.

The data for both the **Fair Market Rent** and the **median rent** comes from the U.S. Department of Housing and Urban Development’s Office of Policy Development and Research [(*Fair Market Rents*, 2023)](https://www.zotero.org/google-docs/?odM0bP). The Fair Market Rent represents the estimate of the 40th percentile of gross rent cost for “standard quality units” [(*24 CFR § 888.113*, 2023)](https://www.zotero.org/google-docs/?jFqKgN) and is calculated via statistical analysis performed on data collected by the American Community Survey. The median rent represents the 50th percentile and is also calculated using the data collected by the American Community Survey [(*50th Percentile Rent Estimates*, 2023)](https://www.zotero.org/google-docs/?uvturc). The Office of Policy Development and Research has made their data available online as downloadable Excel files dating back to 2001, but the format is inconsistent across the years. We created a script in R to clean up the data by first loading in all of the Excel files, then filtering the data for only the columns and geographical areas of interest. At this point, it was noticed that there were multiple entries for King County, so conditions were added to only retrieve data for King County, Washington. The script then standardized column names using the DPLYR functions “rename”, “mutate” and “recode”. These updated files were resaved under a new name to be used in later creation of a final dataset.



The data for the **number of unhoused people** comes from the Department of Housing & Urban Development (HUD) Exchange Continuum of Care (CoC) Homeless Assistance Programs Homeless Populations and Subpopulations Reports. According to the HUD Exchange website, the reports come from the Point in Time count performed in one 24 hour period in January [(*CoC Homeless Populations and Subpopulations Reports*, 2023)](https://www.zotero.org/google-docs/?hCtWAR). The annual reports for the years 2005-2023 were examined for our selected counties. Each year’s report was provided on the HUD Exchange website as a PDF. The total number of unhoused people for each county and each year was manually entered into a spreadsheet that was then merged into the final dataset. We chose to use these reports as the most consistent population data for the years and counties we were interested in. We felt it was important to have one source, with a single standard/set of rules for reporting data.

Once the data from all three sources was located and prepared, an R script was created to compile all of the various CSV files and Excel sheets into one dataframe. This script works by reading in all the files from the project [github](https://github.com/maurawilson/LIS-572-Final-Project) site and combining them by using the “merge” function.

**Ethical Concerns and Considerations and Limitations of Available Data**

Wage Data

One of the limitations of the wages portion of this dataset is that it only covers jobs eligible for unemployment. It does not include information on people who are working unreported jobs, which are more likely to be lower wage positions. Another limitation is that by using an average, the distribution of the wages is not shown, hiding the reality that many people make much less than the average wage. Ethically, it is important to ask who this data might misrepresent or exclude, and to be aware of that when analyzing the data.

Rent Data

Because the underlying motivation for our project is concern for people most susceptible to housing insecurity, Fair Market rent (FMR) is perhaps a better measure of rent costs for that population than average rent because HUD uses FMR data to set rates for low-income housing programs, whereas average rent amounts would include by definition rent for luxury housing accommodations that aren’t realistically attainable for most people.

FMRs are intended to estimate the cost of shelter plus major utilities (not including telephone, cable, satellite, television, or internet services)(H.U.D., nd). Calculation of FMRs is not an exact science, but rather a statistical exercise based on the outcome of survey data (the American Community Survey)[[1]](#footnote-1). “HUD calculates FMRs by using base year data from the American Community Survey (ACS) and then adjusting those numbers to the current year. For FY 2024, HUD uses 2021 5-year ACS estimates as its baseline, adjusts rents to 2022 levels using a mix of private sources and CPI rent of primary residence, and projects rents forward through the rest of 2023 and into 2024” (H.U.D., n.d.). Thus a limitation of this data is that the basis for estimates for a given year may be data several years old then adjusted mathematically, in addition to the limitations of the ACS survey data.

We have also included in our dataset information from H.U.D. denoting median rent data for each of our chosen locations [(*50th Percentile Rent Estimates*, 2023)](https://www.zotero.org/google-docs/?roqg8E). This data is based upon the same ACS information that informs the FMR data and is calculated by H.U.D. for every location for which FMR data is calculated. Median rent amounts are calculated by HUD for purposes of determining H.U.D. housing assistance amounts when the FMR alone is insufficient to achieve adequate housing choices in a particular metropolitan area (H.U.D., 2020, p. 8). Because our chosen locations are densely populated areas where adequate housing is likely to be more expensive, it is useful to consider this data in addition to the FMR rent data.

It is important to note that ACS data is not perfect in its sampling size, collection methods, survey content, resultant data processing, data analysis, and data dissemination (Division of Behavioral and Social Sciences and Education, 2015, pp. 19-22). Additionally, HUD’s data for our chosen locations is based on Metropolitan Areas, rather than city limits or counties.[[2]](#footnote-2) Thus the data may be overinclusive when compared to data based strictly on city limits, and underinclusive when compared to county-wide data. However, given the ubiquity of ACS-based data in governmentally compiled statistics as well as the other ACS-based data we are using in this project, the HUD FMR data is nonetheless informative and as near to apples-to-apples comparison with our other data points as we will likely be able to get. Because a two bedroom unit would likely be most useful for families with children, and because it is essentially the middle of the size range (0 to 4 bedroom) tracked by the ACS and H.U.D., we used two bedroom rents as a benchmark for comparison in many of our visuals. Most of our analysis settled on evaluating median rents because the H.U.D. median rent is employed to determine public housing benefit amounts in cities where fair market rent is insufficient to actually access quality housing in cities where rental unit availability is limited and/or cost is particularly high. (H.U.D., n.d.)

Data on Number of Unhoused People

A Continuum of Care (CoC) is a local or regional planning body that is responsible for coordinating the funding and delivery of housing and services for people experiencing homelessness in its service area. Each of the 381 CoCs in the US reports the number of unhoused people in their area as part of an application for funding. These numbers were collected and analyzed in our dataset.

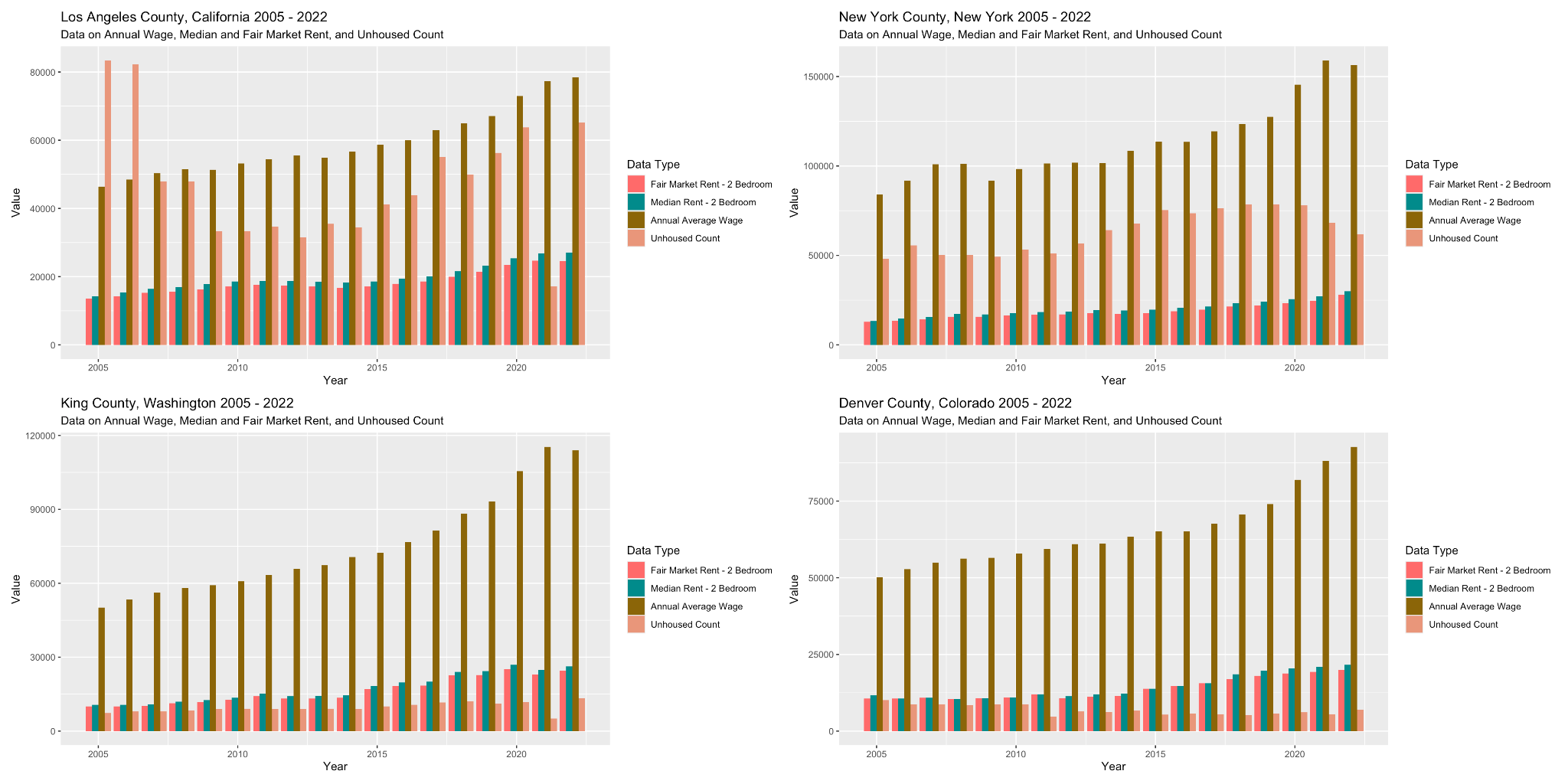
Each yearly report provides information about limitations of the data. One problem associated with using this data is that prior to 2011, the numbers submitted by the CoCs were not independently verified by HUD, and from 2012 on, only a limited review was performed on the data. (HUD, n.d.). Each report cautions that although standards exist for collecting counts of unhoused people, compliance with these standards, and therefore the accuracy of these numbers, may vary. Further, the accuracy of the count was further impacted during the COVID-19 pandemic when it was particularly difficult to obtain accurate numbers while comporting with social distancing measures. (HUD, 2021). Determining the number of unhoused people in a given area is challenging and this reporting system, even with its faults, is the best data we were able to find for all four of our chosen areas. (HUD, n.d.).

Another problem involves the fact that this is a point-in-time count and, of course, the numbers may fluctuate over the year between counts. For our particular chosen counties, the Denver data is actually for Metropolitan Denver and does not include the other cities included in Denver County. The New York data is for New York City, not New York County. This will cause the numbers of unhoused people for these to be lower than the entire county represented in the other data in our dataset (rent and income). However, these areas align with the metropolitan areas used in the HUD data for fair market rent discussed above.

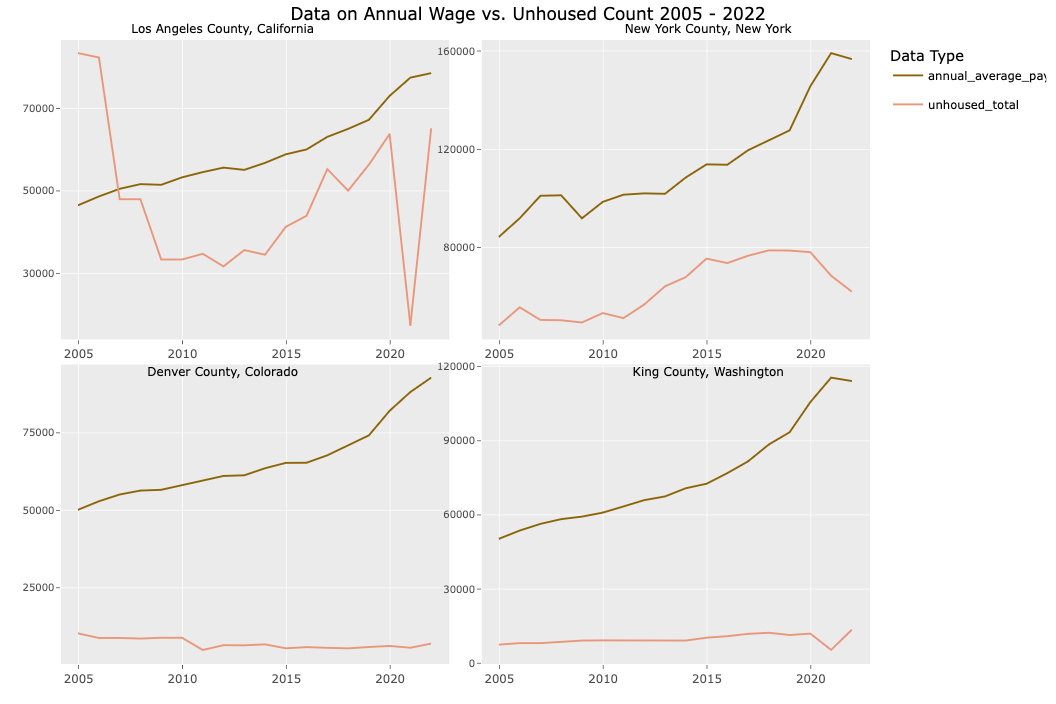
Finally, methods of taking a point-in-time count differ between communities in their collection practices. Some communities take actual counts of only a sample, and then use techniques to extrapolate that data to the larger community (HUD, 2023). Additionally, counts of unhoused youth often underestimate the number of youth who are actually unhoused. (National Alliance to End Homelessness, 2016).

**Findings and Data Visualizations**

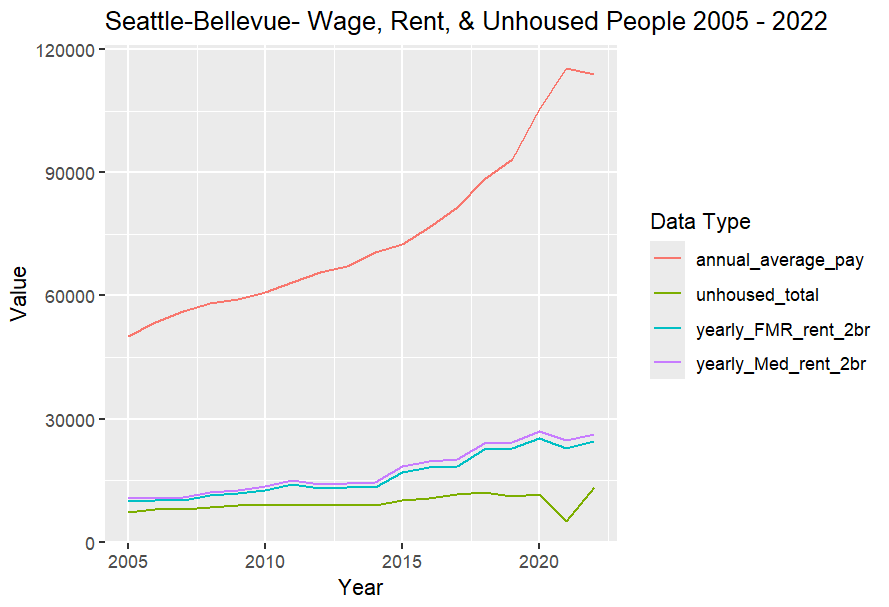
Initially, we thought plotting all the data together might immediately show a trend between the cost of rent, the average wage, and the number of unhoused people. This was not the case for our plot which showed the median and fair market rents for a 2 bedroom unit. The causes for this could be many. First, since average values tend to skew high and do not show distribution of wages, it is impossible from the data we have to show how many people are making average wage or above. In the four areas of the United States we chose, the average wage is going up, but since this data does not reflect actual take home pay, it is difficult to compare this value to the concrete numbers of median or Fair Market rent payment. Analyzing our data to see if the number of unhoused people is directly affected by people’s inability to afford rent payments also requires more information than just how much people make each year and how much rent costs in their city. Cost of living includes food, childcare, medical needs, car or transportation costs, and so much more. In the four areas we examined, the number of unhoused people is rising and falling from year to year in an undetectable pattern from simply plotting all the data in a bar plot. From this initial plot, we determined that in order to make sense of the data collected, we needed to look at multiple other views.



We chose to zoom in on the variables of total number of unhoused persons and average income in each city and compare those side by side in hopes that a simpler comparison could yield some useful insight.

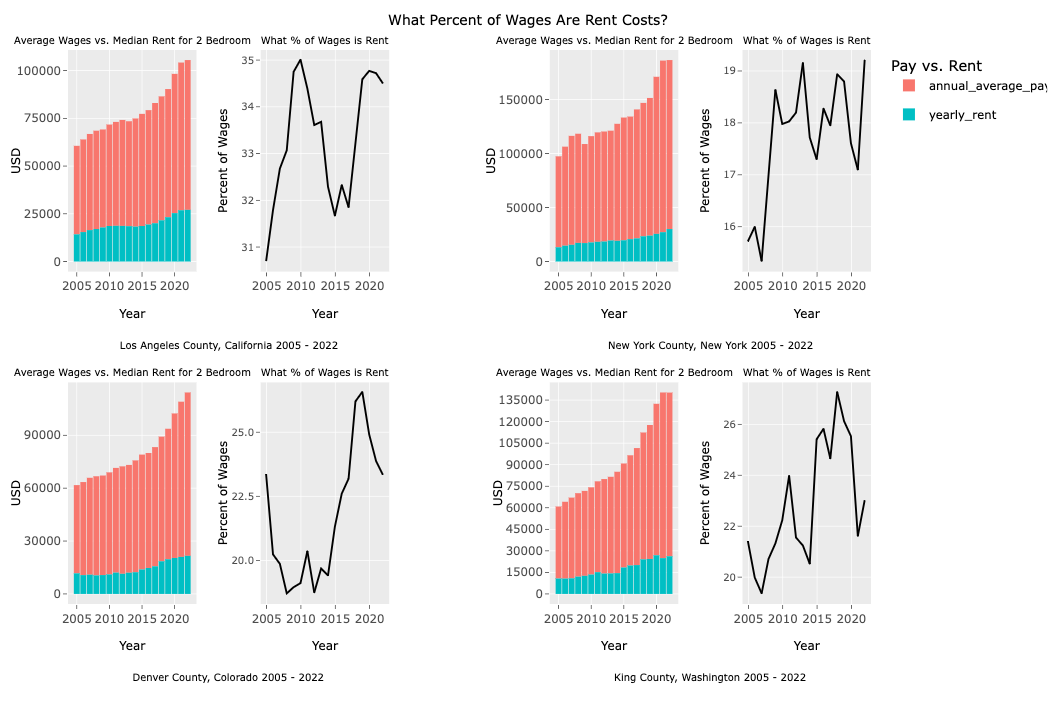


With wages generally increasing over time on average, why might the number of unhoused people not be decreasing over time? One might expect that with more income, more people would be able to afford housing and the number of unhoused persons would, thus, go down. However that does not seem to be the case. One explanation may be income inequality. For Seattle, for example, wages increased dramatically for a small part of the population but remained generally stagnant for most others, thus increasing income inequality over time (while inflating the average wage numbers) (Balk, 2017.)

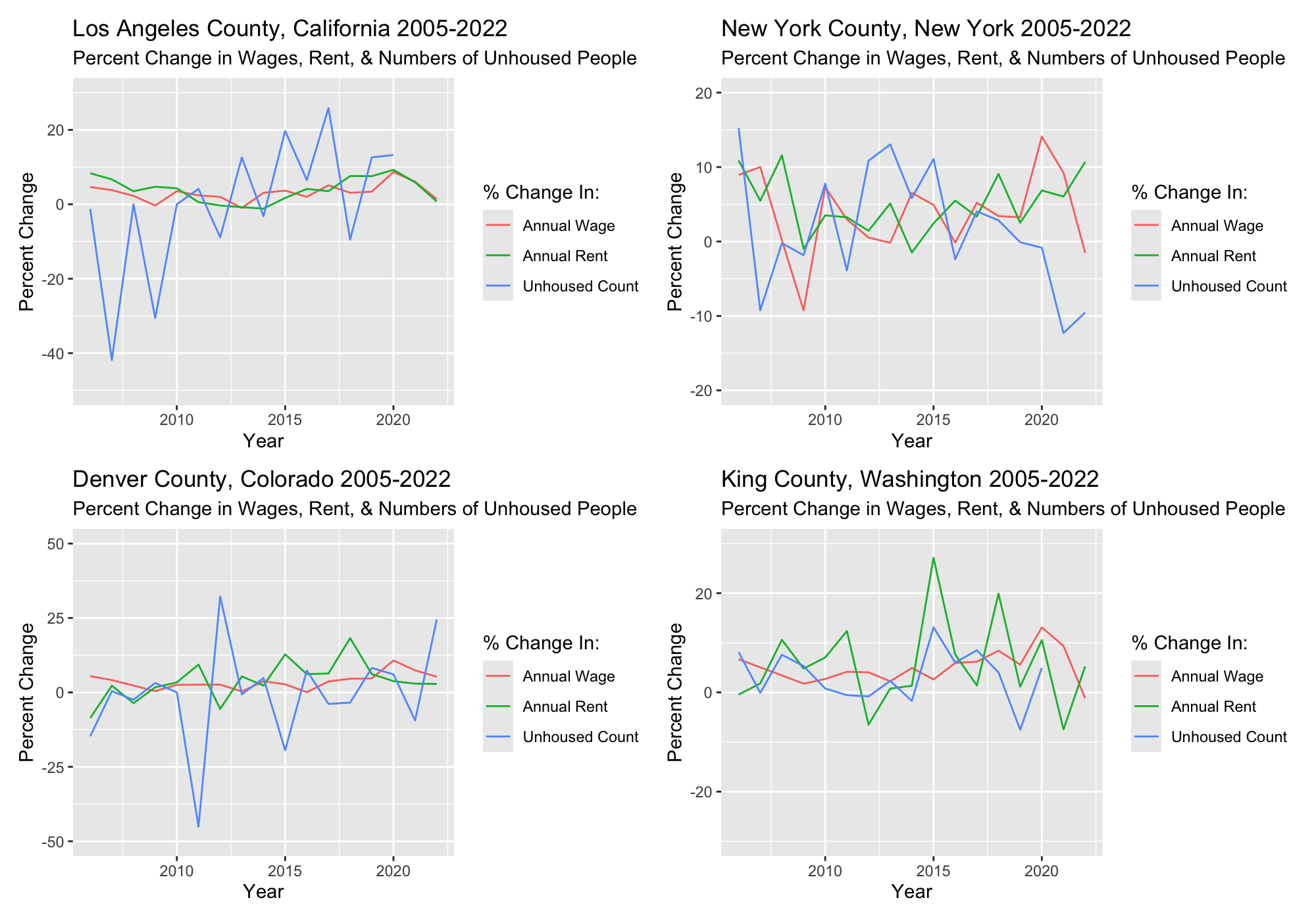


Comparing wages and rent across the years, it would appear that although rent numbers are increasing, wages are increasing as well, giving the illusion that affording rent should not be a problem. Looking at the data from a different angle, however, shows that increases in the percent of wages spent on rent could make it harder to afford the cost of living in the city.

King County, for example, shows wages increasing each year, but the percent of these wages spent on median rent for a 2 bedroom unit has also been steadily increasing.



A better way to compare changes in wages and rent, that allows us to also see changes in population numbers of unhoused people, is to view the numbers as percentages of change between the years. This gives us a clearer picture of what is actually happening. Looking at King County again we can see that the percent increase in rent is greater than the percent increase in wages multiple times over the span of analyzed years. Perhaps not coincidentally, there is also a percent increase in numbers of unhoused people at those times. This can be observed in all four of the areas we looked at.



Realizing that while this does not definitively prove that unaffordable rents are causing homelessness, it does give us something to think about when it comes to income and housing inequality.

**Future Work**

​​Given more resources, we would like to access data on the number of Airbnb units in each area since the inception of the company in 2008 to see if there are any identifiable trends between that data, the cost of living, and the number of people facing housing instability. While we found sources for some years of data, to access what we wanted, we would need a large amount of money or time to scrape the data ourselves. Our initial goal was to look at Airbnb as a potential contributor to housing instability but did not have the time or resources to accomplish this goal on the timeline we had.

Additionally, comparing the data we already have to populations in each of these cities over time could be useful to contextualize numbers of unhoused persons. The U.S. Census Bureau tracks income inequality each year in what is called the Gini Coefficient (Lamb, 2012), and adding this measure of income inequality to our dataset would add additional context and possibly allow additional conclusions to be drawn from the data.

Reference List

Balk, G. (2017, November 17). *Seattle hits record high for income inequality, now rivals San Francisco.* Seattle Times. <https://www.seattletimes.com/seattle-news/data/seattle-hits-record-high-for-income-inequality-now-rivals-san-francisco/>

[*CoC Homeless Populations and Subpopulations Reports*. (2023). U.S. Department of Housing and Urban Development. https://www.hudexchange.info/programs/coc/coc-homeless-populations-and-subpopulations-reports](https://www.zotero.org/google-docs/?M42DSV)

Colburn, Gregg; Aldern, Clayton Page. (2022). *Homelessness Is A Housing Problem*. University of California Press.

*Conducting the 2021 Unsheltered PIT Count.* (2021). <https://files.hudexchange.info/resources/documents/Conducting-the-2021-Unsheltered-PIT-Count.pdf>

[*Fair Market Rents*. (2023, April 19). U.S. Department of Housing and Urban Development’s Office of Policy Development and Research.](https://www.zotero.org/google-docs/?M42DSV) <https://www.huduser.gov/portal/datasets/fmr.html#data_2024>

[*50th Percentile Rent Estimates*. (2023, April 19). U.S. Department of Housing and Urban Development’s Office of Policy Development and Research. https://www.huduser.gov/portal/datasets/50per.html#year2016](https://www.zotero.org/google-docs/?M42DSV)

[*Handbook of Methods: U.S. Bureau of Labor Statistics*. (2023, November 20). Bureau of Labor Statistics. https://www.bls.gov/opub/hom/cew/home.htm](https://www.zotero.org/google-docs/?M42DSV)

[*History: Handbook of Methods*. (2023, November 20). Bureau of Labor Statistics.](https://www.zotero.org/google-docs/?M42DSV) <https://www.bls.gov/opub/hom/cew/history.htm#program-identification-and-data-publication>

*How to use sampling within a COC to conduct an unsheltered pit count - hud exchange.* (2023, April).   
https://www.hudexchange.info/resource/6940/how-to-use-sampling-within-a-coc-to-conduct-an-unsheltered-pit-count/

Lamb, Evelyn (2012, November 12). *Ask Gini: How to Measure Inequality*. Scientific American. <https://www.scientificamerican.com/article/ask-gini/>

[*QCEW Data Files*. (2024, March 6). Bureau of Labor Statistics.](https://www.zotero.org/google-docs/?M42DSV) <https://www.bls.gov/cew/downloadable-data-files.htm>

[*24 CFR § 888.113*. (2023, November 6). LII / Legal Information Institute.](https://www.zotero.org/google-docs/?M42DSV) <https://www.law.cornell.edu/cfr/text/24/888.113>

*What is a point-in-time count?* National Alliance to End Homelessness*.* (2016, October 19). https://endhomelessness.org/resource/what-is-a-point-in-time-count/

1. The American Community Survey (ACS) is a survey conducted by the U.S. Census Bureau that seeks to collect nationwide data on social, economic, housing, and demographic characteristics about the country’s population. Stakeholders such as government agencies use this information to decide how resources should be allocated, amongst other uses. The Census Bureau surveys only a small portion of the population; about 1 in 38 US households per year is invited to participate. The U.S. Census Bureau (n.d.). ACS Information Guide. Census.gov. Retrieved May 18, 2024, from https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS\_Information\_Guide.pdf [↑](#footnote-ref-1)
2. HUD FMR data calculates fair market rents for Los Angeles, California within the aggregate metropolitan area of “Los Angeles-Long Beach-Glendale”, Denver, CO as “Denver- Aurora- Lakewood”, New York City, Ny as “Ny, NY” and Seattle, Washington as “Seattle-Bellevue.” US. Housing and Urban Development (n.d.). FY 2024 Schedule of Metropolitan and Non-Metropolitan Fair Market Rents. HUD. Retrieved May 18, 2024, from https://www.huduser.gov/portal/datasets/fmr/fmr2024/FY2024\_FMR\_Schedule.pdf [↑](#footnote-ref-2)