

LIS 572 Introduction to Data Science

Jessica Fleming, Patty Titus, Maura Wilson

May 19th, 2024

Dataset Bibliography

This 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 is 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 as to the number of people in those locations experiencing housing insecurity reveals whether the people can afford to live there, and may also reveal an underlying cause of homelessness in the city. Each group member has contributed to collecting data sources. The areas we chose to analyze encompass four cities that are in the Top 5 cities with the highest homeless populations. As noted below, because our data comes from a few different sources, the boundaries of each location represented by each data's source is not always perfectly matched. HUD data for these locations, for example, maps to areas that are larger than one city's limits but smaller than an entire county. This will be addressed more in detail in our final deliverable. Our dataset combines data from three different sources into one dataset.

Wage Data

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). 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 - 2023. 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). 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).

The amount of background information needed to make sense of the data is minimal. In the county level files, the information is presented in a concise manner, in plain English. If one was looking to break down annual wage by industry, they would need to understand the North American Industry Classification System (NAICS) to ensure understanding of the professions each “Industry” includes. As stated above, the definition of wage is important to know to properly understand the data. It is important to understand who exactly the data covers, and how it was collected. The data is collected from unemployment insurance accounting systems in each state, which means there are some jobs it does not cover and it is important to be aware of that when analyzing the data.

Coverage only of jobs eligible for unemployment is one of the limitations of this dataset. It does not include information on people who are working unreported jobs, which are more likely to be lower wage positions. Another limitation is in using an average, which doesn’t show the distribution of the wages and therefore hides the reality that many people make much less than the average. Ethically, it is also important to ask questions of who this data might misrepresent or exclude, and include that in any analysis performed using the data.

Rent Data

Fair Market Rent (FMR) data is compiled by the US Department of Housing and Urban Development and can be found here in Excel formats here:

https://www.huduser.gov/portal/datasets/fmr.html#data_2024. Because the underlying motivation for our project is concern for people most susceptible to housing insecurity, it 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. Fair market rent data is easily available on the HUD website.

“Fair Market Rents (FMRs) are used to determine payment standard amounts for the Housing Choice Voucher program, initial renewal rents for some expiring project-based Section 8 contracts, initial rents for housing assistance payment (HAP) contracts in the Moderate Rehabilitation Single Room Occupancy program (Mod Rehab), rent ceilings for rental units in both the HOME Investment Partnerships program and the Emergency Solutions Grants program, maximum award amounts for Continuum of Care recipients and the maximum amount of rent a recipient may pay for property leased with Continuum of Care funds, and flat rents in Public Housing units. The U.S. Department of Housing and Urban Development (HUD) annually estimates FMRs for Office of Management and Budget (OMB) defined metropolitan areas, some HUD defined subdivisions of OMB metropolitan areas, and each nonmetropolitan county. 42 USC 1437f requires FMRs be posted at least 30 days before they are effective and that they are effective at the start of the federal fiscal year (generally October 1). Fair Market Rents, as defined in [24 CFR 888.113](#) are estimates of 40th percentile gross rents for standard quality units within a metropolitan area or nonmetropolitan county.” (H.U.D., n.d.)

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. “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.). Ethical use of this data must be transparent about its provenance, that the basis for estimates for a given year may be data several years old that is then adjusted mathematically, as well as the purposes for which it was collected.

We have also included in our dataset information from H.U.D. denoting median rent data for each of our chosen locations, located here:

<https://www.huduser.gov/portal/datasets/50per.html#year2016>. 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,

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

HUD's FMR data for our chosen locations is based on Metropolitan Areas, rather than city limits or counties.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. Ethical use of this data must be transparent about the geographic areas to which the data applies.

Data on Unhoused Populations

The data for homeless populations 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 are based on Point-in-Time (PIT) information provided to HUD by CoCs in the application for CoC Homeless Assistance Programs. The PIT Count provides a count of sheltered and unsheltered homeless persons on a single night during the last ten days in January.

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 Homeless Persons number 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 homeless 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.

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

The social context of this data lies in the fact that the reported numbers determine CoC funding from HUD. Each person represents a dollar amount. In their *Guide to Counting Homeless People*, HUD presents the reason for the count as one of raising awareness in the community as well as planning and advocating for resources. The effort that goes into the count (especially of the unsheltered) would suggest that the numbers are as accurate as possible, but are more than likely undercounted.

The background information needed to understand this data is the concept of the Community of Care. 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. CoCs are tasked to track and manage the homeless community in their area. HUD divides the US into 381 Continuums of Care (CoC), which are responsible for coordinating homelessness services in their area, and report the number of homeless persons as part of an application for funding.

A problem associated with using this data is that from 2005-2011, HUD did not independently verify the data and from 2012-2023, HUD “conducted a limited data quality review but has not independently verified all information submitted”. Each report states “The reader is therefore cautioned that since compliance with these standards may vary, the reliability and consistency of the homeless counts may also vary among CoCs”. Determining the number of homeless persons in a given area is challenging and this reporting system, even with its faults, may be the most reliable for counts.

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 homeless numbers 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 H.U.D. data for fair market rent discussed above.

Reference List

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>

History: Handbook of Methods. (2023, November 20). Bureau of Labor Statistics. <https://www.bls.gov/opub/hom/cew/history.htm#program-identification-and-data-publication>

QCEW Data Files. (2024, March 6). Bureau of Labor Statistics. <https://www.bls.gov/cew/downloadable-data-files.htm>

National Alliance to End Homelessness Community of Care Fact Sheet
<https://endhomelessness.org/wp-content/uploads/2016/08/Fact-Sheet-CoC-2-1-2010.pdf>

HUD Exchange
<https://www.hudexchange.info/programs/coc/coc-homeless-populations-and-subpopulations-reports>

HUD (n.d.). *Fair Market Rents (40th Percentile Rents)*. HUD Office of Policy Development And Research. Retrieved May 18, 2024, from <https://www.huduser.gov/portal/datasets/fmr.html#null>.

Program Parameters and Research Division, Office of Economic Affairs (n.d.). *Fair Market Rents Introductory Overview*. US Housing and Urban Development. Retrieved May 18, 2024, from https://www.hud.gov/sites/dfiles/Phttps://www.huduser.gov/portal/sites/default/files/pdf/fmr-overviewFY24.pdfA/documents/FMR_FAQs.pdf

HUD (n.d.). *Frequently Asked Questions: FY 2024 Fair Market Rents*. US Housing and Urban Development. Retrieved May 18, 2024, from https://www.hud.gov/sites/dfiles/PA/documents/FMR_FAQs.pdf

Panel on Addressing Priority Technical Issues for the Next Decade of the American Community

Survey, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. (2015). *Realizing the Potential of the American Community Survey*.

National Academies Press.

H.U.D. (2020). *Housing Choice Voucher Program Guidebook*.

https://www.hud.gov/sites/dfiles/PIH/documents/HCV_Guidebook_Payment_Standards_November_2020.pdf