George Mason University

CS 504

Fall 2025

# Housing D.C.

# Analysis of Housing for Young Adults in the Washington, D.C. Metropolitan Area

Team: Join Genius

Leslie Rodriguez

Jennifer Reif

Juan Parada

Roger Rocha-Claros

Ryan Sollom

# Table of Contents

# Abstract

High housing costs in the Washington, D.C. metropolitan area pose major challenges for residents, especially young adults ages 20-29. Rising housing costs have outpaced wage growth since 2000, reducing the income available for other necessities. This study used data from the U.S. Census Bureau's American Community Survey to investigate the relationships between income levels, housing costs, and demographic information for young adults in the Washington, D.C. area. The analysis was performed in R using existing open-source packages and team-created mapping tool.

The study found that young adult housing type varies widely across the Washington, D.C. area. Young adults choose to live with family members, with roommates, or independently at significantly different rates across the subareas in the study. Within the D.C. city limits, nearly two-thirds of young adults live independently, compared with only 15% in some outer suburbs. The average monthly rent ranges from approximately $1,300 to $3,000 across subareas within the study region, with a regional average of approximately $2,000 per month. While many young adults in their twenties have found affordable housing, a significant number are cost-burdened where they spend more than 30% of their income on housing expenses. Cost pressure is most pronounced in southern Fairfax County, Virginia and eastern Prince William County, Virginia, where affordability challenges are more widespread.

Overall, housing remains a major expense for young adults in the D.C area and affordability varies widely across the region. Future research should improve housing cost estimates and incorporate newer ACS data to better understand the drivers in disparities.

# 1 Introduction

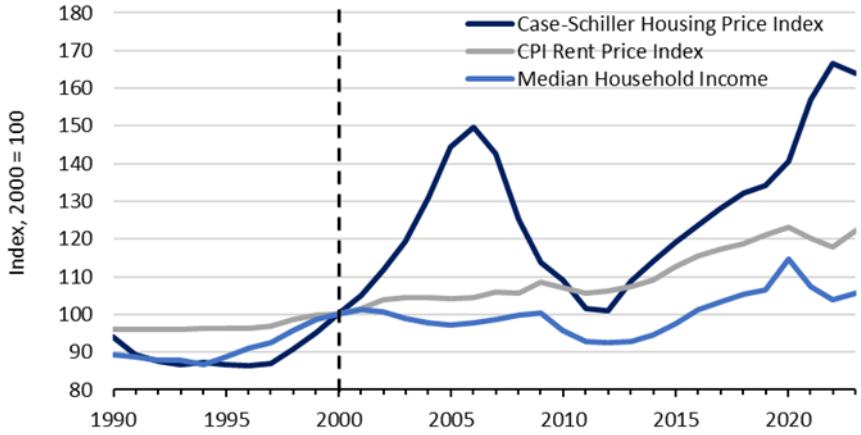
## 1.1 Background and Rationale

***The Rising Cost of Housing***

Rents and home prices have risen faster than incomes across many regions in the United States. In the time period between 2000 to 2020, more than 90 percent of Americans lived in areas where median rent and home prices grew faster than median incomes (Feiveson et al, 2024).

The U.S. Department of Treasury used data from the U.S. Bureau of Labor Statistics, S&P Dow Jones Indices; Current Population Survey; and Treasury staff calculations to develop a graph showing the relationship between median household incomes and housing costs (Figure 1). Rising housing costs have significantly outpaced wage growth since 2000. The growing gap between housing costs and incomes has resulted in an increasing number of households that struggle to find affordable housing.

**Figure 1: Housing Price, Rent and Wage Index Changes Over Time (Feiverson et al, 2024)**




Note: Inflation-adjusted by R-CPI-U-RS, not seasonally adjusted. CPI rent price index is the average rent of primary residences in U.S. cities.
Source: U.S. Bureau of Labor Statistics; S&P Dow Jones Indices; CPS; Treasury staff calculations.

***What is affordable housing?***

The United States Department of Housing and Urban Development defines affordable housing as "housing on which the occupant is paying no more than 30 percent gross income for housing costs, including utilities (U.S. Department of Housing and Urban Development, 2006). In 2022, a study by Virginia's Department of Housing and Community Development (DHCD) defines affordable housing as housing "within the financial reach" of residents across "the full spectrum of incomes and budgets" (Virginia Department of Housing and Community Development & Virginia Housing, 2022). The study further defines community affordability as a comparison between household incomes and the cost of homes to buy or rent.

***Why is it important?***

Stable, adequate housing is essential. Households that spend a higher percentage of their gross

income on housing will have less money available for food, medical care, clothing, and other necessities.

Many studies have shown that inadequate housing can lead to health problems due to faster spread of infectious diseases, inadequate plumbing, or the presence of lead-based paint (Gray, 2023), mold (Office of Disease Prevention and Health Promotion, n.d.), asbestos, or other hazardous materials (Braveman et al, 2011). High housing prices place a strain on budgets, that limit the amount of funds available for other necessities, like food, clothing, and medical care. Families with limited financial resources are more likely to have unhealthy living conditions and are generally least able to remedy the issues; widening disparities in health across socioeconomic groups (Braveman et al. 2011). Coley et al note that students living in housing with fewer deficiencies have better emotional and behavioral functioning and cognitive skills, on average (Coley et al, 2013).

The North Carolina Housing Finance Agency notes that "secure, stable, well-located housing in good condition is linked to better health and well-being". The lack of affordable housing is connected to a wide range of health risks and negative outcomes including postponing medical care, worsening hypertension and heart disease, difficulty leaving domestic violence situations, chronic illness, mental health issues, heart disease, asthma, lead poisoning, and more frequent hospitalizations (Flournoy et al, 2021).

The benefits of affordable housing extend beyond the individual or household to the overall community. The North Carolina Housing Finance Agency reports that the availability of affordable housing improves educational outcomes including increased test scores, school attendance and graduation rates; improved community health with a decrease in overall health expenditures, better local economic conditions including increased local revenue and jobs, and a decrease in crime rates (North Carolina Housing Finance Agency, n.d.).

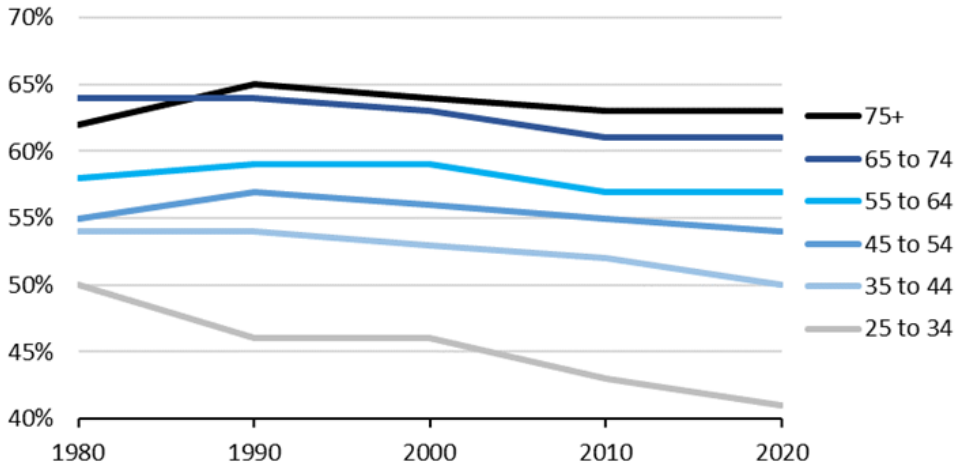
The Pew Research Center produced a comprehensive study of the impacts of housing on child development and found that children's development is affected by their environment in which they live and housing quality and neighborhood characteristics play a large part. The study notes that housing affordability has long-term consequences on children's development and success later in life (Roy et al, 2008). They find that multiple moves, living in overcrowded housing, in an unsafe or unhealthy environment, or in a neighborhood with few resources can adversely affect childhood development and future academic success. Both children and their parents are put under stress by adverse living conditions that strain relationships that are critical in children's early years. Low-earning families that spent more than 50% of their income on housing spent 30% less on food, 50% less on clothes, and nearly 70% less on healthcare than families with similar incomes but lower housing expenses. Reduced spending on nutrition and healthcare can have lifelong consequences.

In a 2019 report published by the Urban Institute, Veronica Gaitan states “Housing quality, instability, and unaffordability threaten the well-being of millions of children across the nation (Gaitan, 2019). Research shows that housing is the first rung on the ladder to economic opportunity and that a person’s access to opportunity is intrinsically linked with that of the community where they live.” She says housing has a central importance as a determinant of “wide-ranging outcomes for the country’s youngest generation.” Her report summarizes other studies that indicate poor living conditions contribute to increased levels of depression, anxiety, and aggression in children and young adults. Educational outcomes are also impacted by housing with poor quality housing associated with lower kindergarten readiness scores, lower graduation rates by age 19 and lower educational attainment by age 25.

***Living with family***

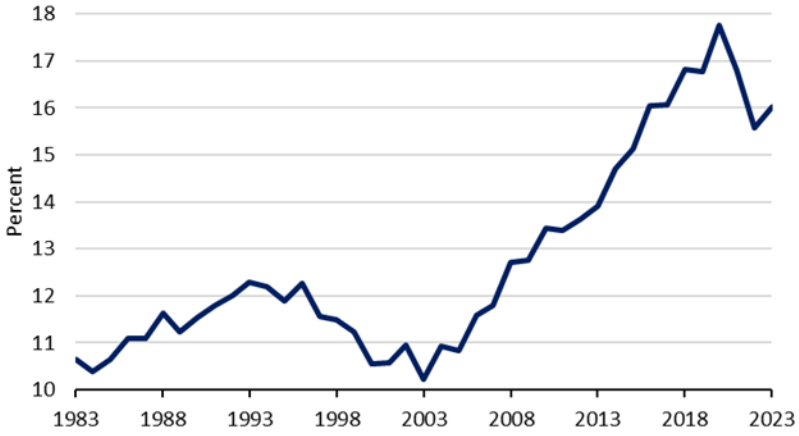
In order to adapt to the increasing disparity between income growth and rising costs, more young adults are choosing to live with family members or others than in previous years. Feiveson et al (2024) found that the percentage of young adults (aged 25 – 34) who were the head of their household declined steadily from approximately 50% in 1980 to approximately 40% in 2020 (Figure 2). This percentage is known as the “headship rate”. They found that the decrease in headship rates can also be seen in the increased number of young adults (ages 25 – 34) that lived with a parent over the same period (Figure 3). They attributed this rise in part to rising rents and house prices.

**Figure 2: Per-Age Rate of Headship (Feiveson et al, 2024).**




Source: Census Bureau; Treasury staff calculations.

**Figure 3: Percentage of Adults Aged 25 – 34 Living with a Parent (Feiverson et al, 2024)**




Source: Current Population Survey; Census Bureau.

In 2015, the Federal Reserve Bank of Boston found that young adults between the ages of 23 and 31 are more likely to live with their parents at that age than similarly-aged young adults in the 1990's. They found that increased housing costs had a meaningful impact on the proportion of young adults continuing to live with their parents. In areas with higher costs, they were 15 percent more likely to live with their parents than peers in lower-cost areas. These findings echo earlier work by Haurin, Hendershott, and Kim (1993) that note that higher housing costs slow the formation of new households. The rate of household formation has a wide range of impacts including the aggregate housing demand, population density, fertility, labor force mobility, and demand for public services (Garasky, Haurin, and Haurin, 2001).

As young adults continue to live with parents or other family members for longer periods of time, it may lead to delayed life milestones like marriage or parenthood (Mulder, 2006). It can also cause less economic mobility as young adults have fewer opportunities to develop rental histories and limited geographic mobility to relocate for jobs (Urban Institute, 2019). There may also social and psychological impacts of continuing to live as a family unit including additional stress or conflict with family members or fewer opportunities for social growth and relationships (Fry, 2025).

***The Local Experience***

It is expensive to live in the Washington, DC area. The Council for Community and Economic Research ranks the area as the 9$^{th}$ most expensive city in the country in 2024 (Council for Community and Economic Research, 2024). The average rental cost for a 1-bedroom apartment within the city limits was $2,308 in May 2025 (Apartments.com, 2025).

The Bureau of Labor Statistics provides age-related breakdowns of median incomes on a nationwide bases. They also provide an analysis of how incomes compare between major metropolitan areas; however they do not provide an age-related breakdown of median

incomes in the Washington, D.C. area. The bureau reported that the average hourly wage across the country was $32.66 in May 2024. It was also reported to be $43.47 (or approximately 33% above the national average) in Washington DC during the same time period (Bureau of Labor Statistics, 2025b). They also report a national average weekly income of $782 for young adults aged 20 – 24 and $1,139 for young adults ages 25 – 34 across the country (Bureau of Labor Statistics, 2025a). If differences in average income are consistent across all age groups, we can infer that the median incomes for young adults in the Washington, D.C. area would be approximately 33% higher than the national average for those age groups. Based on those assumptions, the average weekly income for 20 – 24 year olds in the Washington DC area would be $1,041 and the average weekly income for 25 – 34 year olds in the D.C. area would be $1,516.

**Table 1: Wage Comparison Between Washington, D.C. and National Averages**

| | Washington, DC area | National Average | Ratio |
|---|---|---|---|
| Average hourly wage (all ages) | $43.47 | $32.66 | 1.331 |
| Average weekly wage (Ages 20 – 24) | $1,041* | $782 | |
| Average weekly wage (Ages 25 – 34) | $1,516* | $1139 | |

*Calculated based on national average and ratio between local and national average hourly wages; assumes ratio is constant across all age groups

Based on the average wages calculated above and the median price of a 1-bedroom apartment in Washington, DC. The average 20 – 24-year-old would spend over half their gross income on housing costs. The average 25 – 34-year-old could expect to spend approximately 35% of their gross income on housing. These high percentages leave significantly less income available for other priorities like food, health care, education, or savings.

Similarly, the DC Fiscal Policy Institute found that more than one-third of renters (of all ages) who worked more than 40 hours a week spent more than 30% of their income on housing. They also noted that nearly 25% of workers who were on the job for at least 45 hours a week were also cost-burdened with housing spending in excess of 30% of their paycheck (DC Fiscal Policy Institute, 2025).

## 1.2 Research

The Pew Research Center published a study in 2024 that studies housing affordability across the United States (Desilver, 2024). In their study, they used data from the US Census Bureau including the American Community Survey, current Population Survey/Housing Vacancy Survey and Survey of Construction.

They found that the housing market across the United States dramatically decreased during the COVID-19 pandemic but has partially recovered in the last few years. They note that housing prices have continued to rise, with an increase of 57.8% in the price of single-family homes between July 2019 and July 2024. As a comparison, the Consumer Price Index rose by 22.8% between September 2019 and September 2024. They note that the following factors impact housing affordability: interest rates, new home construction, population growth, population shifts, rising home prices and rents, disposable incomes, and local economic conditions.

A March 2024 study in the "Econofact" newsletter published by The Fletcher School at Tufts University looked at how housing affordability varies across locations and income levels (Bhasin et al, 2024). They found that the ratio of median house price to median income has risen substantially over the last 20 years. In the past, median house prices were 4 to 5 times as high as the median income. In recent years, house prices are closer to 6 times the median income. In some high cost of living areas like San Francisco, this ratio can go up to 10 times or more. They note that renters tend to be considered "cost-burdened" at a higher rate than homeowners. In their study, approximately 40% of homeowners with incomes between $35,000 and $49,000 were cost-burdened, while over 60% of renters with similar incomes were cost-burdened by their housing expenses.

The authors note that an imbalance between housing supply and demand is contributing to the lack of affordable housing. A growing elderly population and an increasing number of older American aging in place in their homes has reduced available housing supply. In addition, the US is building new homes at a slower pace compared to other major countries. In addition, rising interest rates since 2022 have exacerbated housing supply issues. The increased rates have "locked in" to some homeowners who may be interested in "moving up" to a different home because the higher monthly payments from higher interest rates make it cost prohibitive in some cases. The higher interest rates and limited availability of "starter homes" make it more difficult for first-time buyers without home equity to enter the housing market.

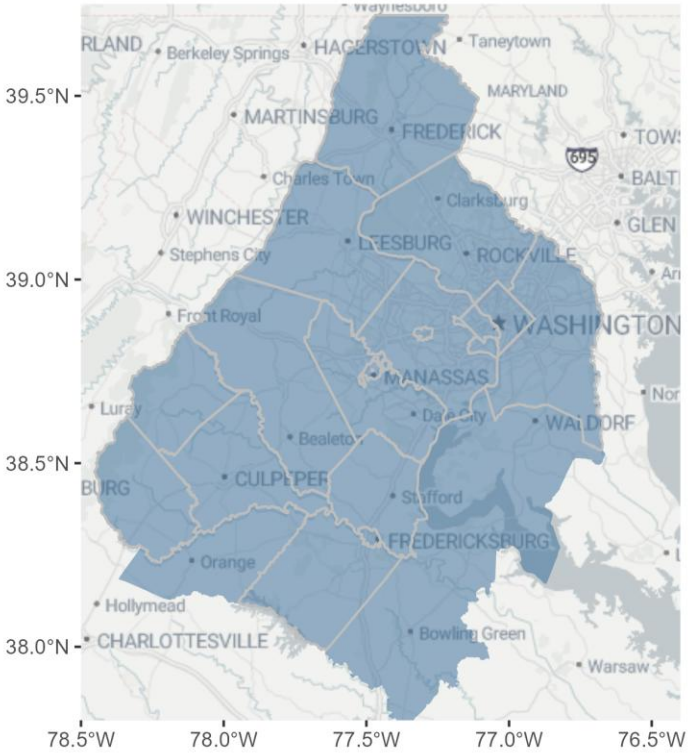
In 2024, the US Census Bureau published a study on housing affordability (U.S. Census Bureau, 2024b). They found that nearly half of renter households spent more than 30% of their income on housing costs in 2023. The study notes that a higher percentage of renters were cost-burdened than homeowners. They also note that there are differences across racial groups for the number of renters that are considered to be cost-burdened. A companion study released at

the same time (U.S. Census Bureau, 2024a) notes that the median cost of renting (rent plus average cost of utilities) grew faster than median home values in 2023.

## 1.3 Project Objectives

This analysis will aim to investigate the relationship between household income and housing affordability among individuals in their 20s living in the Washington, D.C. Metropolitan area from 2021 to 2023. Our study area is defined by the Washington Metropolitan Statistical Area (MSA) as defined by the U.S. Census Bureau (Figure 4). We will focus on how income levels influence housing costs by analyzing variables such as household income, mortgage status, gross rent, and the ratio of housing cost to income. Factors including household makeup, tenure type (owner vs. renter), and regional variations within the metropolitan area will also be considered to better pinpoint patterns and potential barriers.

**Figure 4: Washington, D.C. Metropolitan Statistical Area**




## 1.4 Problem Space

The problem space would be the relationship between income levels, housing costs, mortgage status, and rent among young adults between 2021 - 2023. This issue is relevant and continues to grow within the DC area. There is high demand but limited space and rising interest rates. This has put lots of pressure on the younger residents who are trying to weigh out their options of buying or renting.

## 1.5 Primary User Story (-ies):

As student researchers studying young adults in Washington, DC, we want reproducible estimates of housing cost burden. This is where spending on housing is above 30% for people who rent, people who have a mortgage, and people who own without a mortgage. This will be categorized by income bands, and wards/ neighborhoods from 2021 to 2023 so that policymakers can prioritize assistance and planning on areas where the burden is the highest.

## 1.6 Solution Space

Our project seeks to deliver insight into the relationship between income levels and affordability amount young adults living within the Washington, DC area between 2021 - 2023.

The goal of this analysis is to develop a data-driven model and dashboard that allows users to better understand how affordability challenges vary across income and tenure.

- Our solution:
    - Identify key patterns in affordability
    - Highlight disparities and outliers
    - Use visualization tools to identify neighborhoods with the highest costs
    - Support informed decision making for housing policy and community investment initiatives
- Project will include:
    - Data cleaning and preprocessing
    - Descriptive and inferential analysis using R to explore relationships between key variables.
    - Visualization outputs
    - A summary report or dashboard presenting findings in an accessible way for stakeholders.

## 1.7 Product Vision - Sample scenarios (why would someone want to use this)

- *For*: Individuals in their 20's
- *Who*: Live in the DC areas
- *The*: "Housing affordability study for young adults in the DC Metro area"
- *Is a*: study that focuses on how income levels influence housing affordability
- *That*: Helps show the financial challenges that young adults face when looking for affordable housing in a high-cost area.
- *Unlike*: General studies that look at multiple age groups
- *Our product*: Will focus on the unique barriers that are seen in the DC metro housing market by young adults.
- *Caveats:* Our findings may not be inclusive of other age groups or regions therefore; affordability patterns could be impacted by factors not captured in the available datasets.

### Scenario #1- Young Professionals Starting out

There is a 26-year-old that recently moved to the DC metro area for work and is starting an entry level job. Despite having a stable income, it has become increasingly more difficult to find affordable housing close to the city. They need insight into whether renting or buying may be better in the long term.

### Scenario #2- Graduate Student

A newly graduated student has just finished their master's program and is planning to stay in the DC area after graduation. They need to figure out what their income level needs to be in order to afford housing near the city.

## 1.8 Definition of Terms:

**Affordable Housing-** Affordable housing is generally defined as housing on which the occupant is paying no more than 30 percent of gross income for housing costs, including utilities. (U.S. Department of Housing and Urban Development, 2006)

**Cost burden** – Monthly housing costs (including utilities) exceeding 30% of monthly income. (U.S. Department of Housing and Urban Development, n.d.)

**Group Quarters** - place where people live or stay, in a group living arrangement, that is owned or managed by an entity or organization providing housing and/or services for the residents. Examples: college dormitory, healthcare facility, correctional facility (U.S. Census Bureau, n.d.)

**Family** – Related individuals living in the same household. The Census Bureau also tracks subfamilies (U.S. Department of Housing and Urban Development, n.d.)

**Household** – All people living in a housing unit. Members of a household can be related (see family) or unrelated. (U.S. Department of Housing and Urban Development, n.d.)

**Household Income** – Adjusted household income, which includes the income of all members of the household at the time of the survey, adjusted for inflation to reflect the most recent year of the data release (e.g. 2013 dollars in the 2009-2013 CHAS data). (U.S. Department of Housing and Urban Development, n.d.)

**Metropolitan Statistical Area** (MSA): having "at least one urban area of 50,000 or more population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties" (Office of Management and Budget, 2023)

**PUMA**- non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each. (U.S. Census Bureau, 2025)

**Severe cost burden** – Monthly housing costs (including utilities) exceeding 50% of monthly income (U.S. Department of Housing and Urban Development, n.d.)

**Young adult** – In this study, "young adults" are considered to be adults between the ages of 20 through 29.

# 2 Data Acquisition

## 2.1 Overview:

This study is based on data from IPUMS USA, a program that is hosted by the University of Minnesota and provides free access to harmonized microdata from the U.S. Census and the American Community Survey (ACS). The dataset includes information on demographics, housing, employment, income, and education. We have chosen to use an IPUMS USA dataset because it is designed to make large national survey data easier to explore and analyze. For this project, the dataset will be filtered to include individuals aged 20–29 living in the Washington, D.C. metropolitan area. This focus helps highlight housing and income patterns among young adults in urban regions.

The IPUMS data is in a 70 MB rectangular data file with 117 columns and over 218,000 rows. It was extracted from the larger IPUMS dataset at the University of Minnesota on November 7, 2025. We are authorized to use the data subject the conditions that we cite the data appropriately in our work and that we supply the IPUMS team with the title and full citation of any publications, research reports or educational materials that we develop using the data. We are also prohibited from charging any fees for the use and distribution of the data. We are also instructed to "Use it for GOOD – never for EVIL." in the licensing conditions.

## 2.2 Field Descriptions:

We obtained data from the IPUMS USA database. IPUMS is part of the Institute for Social Research and Data Innovation at the University of Minnesota. The organization compiles and hosts census and survey data from around the world. We extracted a custom data set from their repository, collecting information from American Community Surveys from 2021, 2022 and 2023 (Ruggles et al, 2025). The 2024 data has not been released to date.

After we obtained the data, we selected the variables that we will use in our analysis. Information about these fields is summarized in Table 1 below.

**Table 1: Summary of Variables Used in Analysis**

| ID | Label | Interval | Data Type | Concept |
|---|---|---|---|---|
| YEAR | Census year | discrete | Integer | Technical Variables -- HOUSEHOLD |
| HHTYPE | Household Type | discrete | Integer | Technical Variables -- HOUSEHOLD |
| PUMA | Public Use Microdata Area | continuous | Integer | Geographic Variables -- HOUSEHOLD |
| OWNERSHP | Ownership of dwelling (tenure) [general version] | discrete | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| OWNERSHPD | Ownership of dwelling (tenure) [detailed version] | discrete | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| MORTAMT1 | First mortgage monthly payment | continuous | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| MORTAMT2 | Second mortgage monthly payment | continuous | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| RENTGRS | Monthly gross rent | discrete | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| HHINCOME | Total household income | continuous | Integer | Economic Characteristic Variables -- HOUSEHOLD |
| RELATED | Relationship to household head [detailed version] | discrete | Integer | Demographic Variables -- PERSON |
| INCTOT | Total personal income | continuous | Integer | Income Variables -- PERSON |

## 2.3 Data Context:

Our dataset comes from IPUMS (Integrated Public Use Microdata Series), a website and database that compiles samples of data from 16 different federal censuses dating back to the year 2000. We narrowed down this large set of data to a subset relating to the greater Washington DC- Northern Virginia metropolitan area. The dataset consists of household information and their demographics to better understand how housing is impacting the population. The dataset contains data from 2021- 2023.

## 2.4 Data Conditioning:

The data used in this study is extremely large, consisting of 117 columns and over 218,000 rows. Because it is extremely comprehensive, and much of the data is not needed as it does not add any context or insight into our research, unused columns were dropped from the dataset. Of the columns that do pertain to our research, many of them consist of codes that correlate with places, demographics, or other identifiers whose values are labeled in a data dictionary provided by IPUMS. This dataset was conditioned using look up tables to translate these codes for ease of use in data analyses.

## 2.5 Data Quality Assessment:

To ensure the dataset is reliable and suitable for analysis, a structured data quality assessment was conducted using the 6 primary qualities.

- Completeness: The dataset is derived from ACS microdata, which achieves a high level of completeness through extensive sampling. There were some missing values observed in conditionally. Missing data in these fields is therefore expected and was addressed through filtering. Overall, completeness is considered high for core demographic and economic variables such as income ETC.
- Uniqueness: Each household and person record in IPUMS includes a unique identifier, ensuring no duplication within the dataset. These identifiers maintain one-to-one relationships between individuals and households. There are no personally identifiable details, but uniqueness at the household and individual levels remains intact.
- Accuracy: All data originates from the U.S. Census Bureau and undergoes rigorous validation and weighting before release. IPUMS enhances data accuracy by standardizing variable definitions across years. While survey-based data may contain self-reporting errors, the consistency of ACS methodology and large sample size minimize potential biases. To maintain accuracy, calculated metrics (e.g., housing cost burden) will be recomputed directly from the base fields to verify correctness.
- Atomicity: Each variable represents a single, well-defined attribute—such as age, income, or mortgage payment—that can be independently analyzed. Derived variables are clearly documented and calculated separately to preserve atomicity
- Conformity: All variables adhere to IPUMS naming and coding standards. Values for categorical variables. The dataset thus maintains conformity across both temporal and categorical dimensions.
- Overall Quality: The IPUMS USA data used for this project is of high overall quality and appropriate for quantitative analysis. It combines large sample size, robust documentation, and consistent variable structures. While the dataset is extensive, careful about preprocessing, such as code translation, filtering, and missing value handling. The primary limitations involve the geographic aggregation at the PUMA level and the exclusion of certain age groups, which may slightly limit generalizability. Despite these, the dataset's reliability and depth make it an excellent foundation for exploring housing affordability.

## 2.6 Other Data Sources:

The alternative data source that was initially considered for this project was going, directly to the U.S. Census Bureau's American Community Survey (ACS) Summary File. This source provides detailed pre-tabulated data on demographic, social, economic, and housing characteristics for all areas in the US. However, the team decided to not use this source because the datasets are extremely large and time-consuming to work with. In order to access specific variables, it requires that we extract and merge multiple separate tables making the process inefficient and would increase the chances of errors.

# 3 Analytics and Algorithms

## 3.1 Algorithm: Housing Cost Burden Calculation for Young Adult Renters and Owners

To quantify housing affordability for young adult renters and owners, we first implemented a housing cost-burden calculation using the 2021-2023 ACS data from IPUMS. Using the cleaned dataset created by team member Ryan Sollom, we generalized terms used to classify ownership in columns “OWNERSHP” and "OWNERSHPD” to either “Owner”, “Renter”, or "Other / N.A.". Using the new classifications, we created a new column “simple_tenure” to help create distributions for renters and owners. While looking at the summary statistics of the columns in the cleaned dataset I noticed that column “HHINCOME” had extreme values like 9,999,999 or 0 were treated as missing values. The column “RENTGRS” also had negative values which did not make sense for the monthly rent. The negative rent was treated as a missing value. To store the cleaned values, we created “HHINCOME_clean and “RENTGRS_clean".

One crucial step in the algorithm was determining who would be considered a renter or owner. Classifications in "OWNERSHPD” have labels such as “no cash rent”, “cash rent”, “Owned free and clear”, and "Owned with mortgage or loan". “No cash rent” and “cash rent” were classified as Renter. For “no cash rent” the housing cost was set to zero. In the “cash rent” classification we would just refer to “RENTGRS_clean”. “Owned free and clear” and "Owned with mortgage or loan" were classified as Owner. For “Owned free and clear” the housing cost was set to zero and for "Owned with mortgage or loan" we would refer to “MORTAMT1” and “MORTAMT2”. Using this information, we can create new columns “rent_for_burden” and “owner_for _burden”. These columns are used to help us with the calculation of rent burden and owner mortgage burden.

Using the cleaned monthly cost and income information, the algorithm computes a continuous rent-burden and owner-burden ratio for each rent and owner with valid data. This ratio is annualized rent or mortgage divided by household income.

$$\text{rent\_burden} = \frac{12 \cdot \text{rent\_for\_burden}}{\text{HHINCOME\_clean}} \tag{1}$$

$$\text{owner\_burden} = \frac{12 \cdot \text{owner\_for\_burden}}{\text{HHINCOME\_clean}} \tag{2}$$

This is only calculated when the household has a positive, non-missing income. The ratio is between 0 and 1 but occasionally produces values above 1. This is interpreted as the fraction of the household income that goes to rent or the mortgage, also known as the renter burden or

owner burden. For Classifications under 30 percent are considered not cost-burdened, 30 percent to 50 percent are cost-burdened, and larger than 50 percent is severely cost burdened.

**Algorithm Application 1:**

Measure housing affordability for young adult renters in the DC metropolitan area by year, using the 30% and 50% cost-burden thresholds.

**Algorithm Application 2:**

Compare housing cost burden across demographic and socioeconomic groups such as income bands, PUMA/ward, race/ethnicity, education, employment status by using the burden categories as an outcome variable in descriptive tables and plots.

**Implementation:**

The libraries implemented are ipumsr, dplyr, ggplot2, and scales. The housing burden calculation is coded directly. The code size to implement this will be about 100 lines. The code complexity is at about a medium level. To ensure the implementation is smooth, we will have the team look at the results and compare them with their analysis. I will also continuously check if the results are consistent with other distributions. The code will be placed under version control in a GitHub repository.

**Associated Risk 1:**

The misclassification of renters/owners due to top coded values. The probability of this happening is at a medium level. If this occurs, the impact on the results will be high. A mitigation strategy that can be used is the IPUMS metadata and documentation to re-code extremely high values in “HHINCOME” to missing value” and to enter a zero value for "no cash rent” classifications in “OWNERSHPD”. The risk has been mitigated.

**Associated Risk 2:**

The extreme outliers in rent burden and owner mortgage burden from reporting errors or very low income. The probability of this happening is high. The data needs to be checked for errors such as negative rent payments. The impact on the results will be high. A good mitigation strategy is to create a new column of rent payments with the reporting errors removed and zero values given to classifications of “no cash rent”. Another mitigation strategy can be to restrict the x-axis in ggplot from 0 to 100 percent. It is not necessary to see housing cost-burden exceeding 100 percent. The risk is currently mitigated.

## 3.2 Algorithm: Renter Housing Cost Analysis

This algorithm uses IPUMS ACS microdata to measure housing affordability for young adults ages 20–29 in the Washington, D.C. region. The process begins by classifying housing tenure using the OWNERSHP variable, placing each observation into one of three categories: renter, owner with a mortgage/loan, or not applicable. For renters, a monthly housing cost variable (rent_for_burden) is created using OWNERSHPD and RENTGRS_clean. Renters with cash rent use their gross rent, while "no cash rent" cases are assigned a value of zero. All other tenure types receive a missing value since rent burden does not apply.

Rent burden is then calculated for renters with valid income using:
**HHINCOME_clean**): rent = 12 * rent_for_burden / hhincome_clean

Each renter is assigned a HUD-standard burden category:

Not cost burdened: < 30%
Cost burdened: 30–49%
Severely cost burdened: ≥ 50%

The algorithm produces a summary table showing the count and proportion of renters in each burden category, with optional breakdowns by year, PUMA, and demographic characteristics. These outputs help quantify how affordable the region is for young adults.

**Application:**
This algorithm identifies the typical rent level and burden status for young adults who rent in the DC metro area. It provides insight into rental market pressures and overall affordability for this age group.

**Implementation Summary**
The algorithm is fully implemented in R using standard functions and packages. It consists of roughly 40–50 lines of moderately complex code that classify tenure, construct monthly housing costs, and calculate rent burden categories. Our group will validate the results by checking sample counts and confirming that median values align with the analysis output. All code is stored and maintained in the team's GitHub repository for version control.

**Risks Associated:**
1. Incomplete Housing Cost Data: Rent and mortgage values are only recorded for householders. Including non-householder observations, introduce zeros that distort results. The probability and impact of this issue are both medium, but it is mitigated by filtering the dataset

to include only cases where RELATED == "Head/Householder" before running any housing cost or affordability calculations. This risk is currently mitigated.

2. Few Mortgage Cases: Homeownership is rare among 20–29-year-olds, leading to many zero or missing mortgage values. The probability of this issue is high, with a medium level of impact. To address this, the analysis focuses primarily on renters and uses owner information only as a descriptive context, making this risk effectively mitigated.

3. Sampling Bias: Young adults who are householders do not fully represent the broader 20–29-year-old population, since many in this age group live with parents or roommates. The probability and impact of this bias are both medium-sized. To address it, the analysis clearly states that all findings apply specifically to renter householders rather than all young adults. This mitigation is still in progress as the team continues to refine how the results are presented.

### 3.3 Algorithm: Geographic Distribution and Mapping Tool

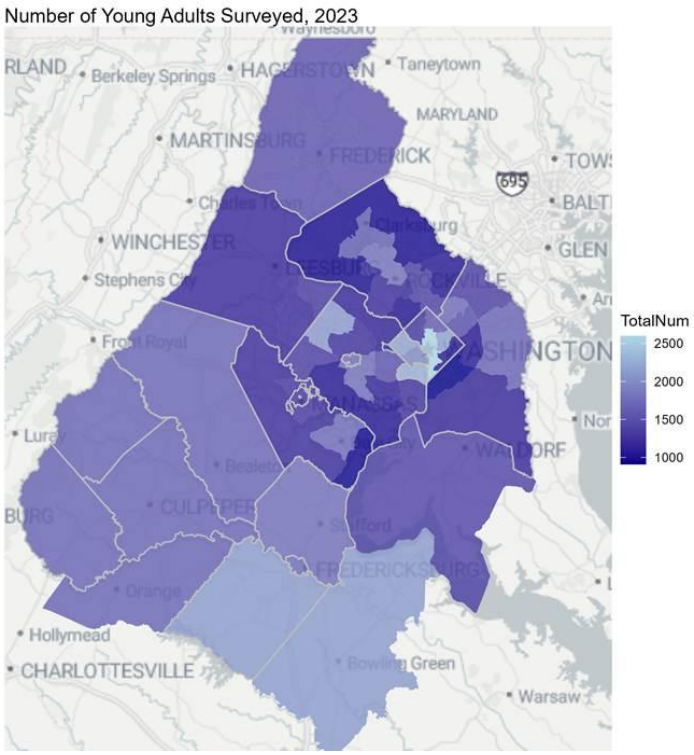
A mapping tool was created using the R programming language to provide visual representations of the geographic variation in the data points in this study. The tool joins our analytic results with data obtained from the U.S. Census Bureau and Stadia Maps. A raster "Alidade Smooth" basemap was obtained from Stadia Maps (Stadia Maps, 2024). The geographic limits were set to encompass the entire study area. Shapefiles containing the geographic boundaries of each Public Use Management Area (PUMA) and county in the study area were obtained from the U.S. Census Bureau (U.S. Census Bureau, 2024). Using the R programming language and the ***ggmap*** package, the analytic results were grouped by PUMA and combined with the shapefile information to create the maps.

The resulting maps were created by joining the PUMA boundaries with the analytic results and layering that on top of the Stadia Maps basemap. County boundaries at a reduced opacity were added on top of the analytic results to add additional context to the maps.

In order to collect and combine all required information, the code will be moderately complex. It requires two API keys and connections to both the US Census Bureau and the Stadia Maps websites. Information from these sources will be combined with our analytic results to produce the maps.

Version control will be maintained by using the team's shared GitHub repository. An example of the map showing the population of young adult survey respondents by area is below (Figure 5).

**Figure 5: Population of Young Adults Surveyed by PUMA Area**




The risks associated with the algorithm primarily focus on inaccuracies in joining multiple data sources together. Any layer misalignment could result in user confusion or misinformed decisions based on erroneous geographic information. These risks will be mitigated by careful tool development and an expectation that developers regularly check the join operations at each stage of the development process.

## 3.4 Algorithm: Spatial Distribution of Young Adult Living Situation

This algorithm explores the relationship between income level, housing cost, and living situation (living with friends, family, or independently). The results will be mapped using the mapping tool to evaluate any geographic differences in average housing cost, and the rates at which young adults live independently, with family members, or other adults.

**Algorithm Application:**

The data analysis will be performed using the R programming language with the assistance of the "tidyverse" package. The code is fairly straightforward with approximately 100 lines. One team member is developing the original code and the remainder of the team will verify the code and its execution.

The data grouped into categories based on whether the respondent lived independently, with family members, or with other peers like friends or roommates. The **RELATE** variable in the dataset identifies the relationship between the respondent and the "head of household". This information was used to categorize categories based on the **RELATE** variable in the dataset using the guidance in Table 2 below:

**Table 2: Relationship between RELATE variable and HoH Classification**

| Value | Label | HoH Category |
|---|---|---|
| 01 | Head/Householder | Head |
| 02 | Spouse | Head |
| 03 | Child | Family |
| 04 | Child-in-law | Family |
| 05 | Parent | Family |
| 06 | Parent-in-Law | Family |
| 07 | Sibling | Family |
| 08 | Sibling-in-Law | Family |
| 09 | Grandchild | Family |
| 10 | Other relatives | Family |
| 11 | Partner, friend, visitor | Friend/Peer |
| 12 | Other non-relatives | Friend/Peer |
| 13 | Institutional inmates | Friend/Peer |

Individuals that were identified as "Head/Householder" or "Spouse" are considered to be living independently. The remainder of the respondents are classified as living with family members or with peers if they live with roommates, visitors, or other non-relatives. One limitation of the

data and this approach is that it does not make a distinction between individuals living with roommates and individuals living with unmarried romantic partners.

The data was further cleaned by converting the PUMA values from integers to five-character text strings for use with the mapping tool. The **INCTOT** variable was cleaned by updating all the values of 9,999,999 to NA. In addition, **INCTOT** values less than zero (indicating a yearly loss) were changed to 0 for this section of the analysis.

The analysis includes a series of maps, including the percentage of young adults in the total population, as well as the percentage of young adults living independently, with family or with other adults. It also includes a comparison of the percentage of young adults in each living situation category based on the average housing cost in each PUMA area.

The risks associated with the algorithm are generally associated with anomalies in the dataset. While the extreme and NA values have been addressed in the development process, inaccurate values such at showing rent of $15,000 per month instead of $1,500 will cause errors in the resulting analysis. It is extremely difficult, if not impossible, to find all such anomalies since housing costs vary widely across the region and by housing type. The risks associated with these errors are moderate.

# 4 Visualization

## 4.1 Monthly Rent Distribution

This analysis examines the monthly gross rent paid by renters aged 20 to 29, using weighted microdata from IPUMS USA (Ruggles et al, 2025). After filtering out invalid and top-coded rent values, the dataset represents young adult renters across the United States.

**Table 3: Summary Rent Statistics**

| Metric | Value |
|---|---|
| Mean Rent | $2,089 |
| Median Rent | $2,067 |
| 25$^{th}$ Percentile (Lower Quartile) | $1,673 |
| 75$^{th}$ Percentile (Upper Quartile) | $2,594 |

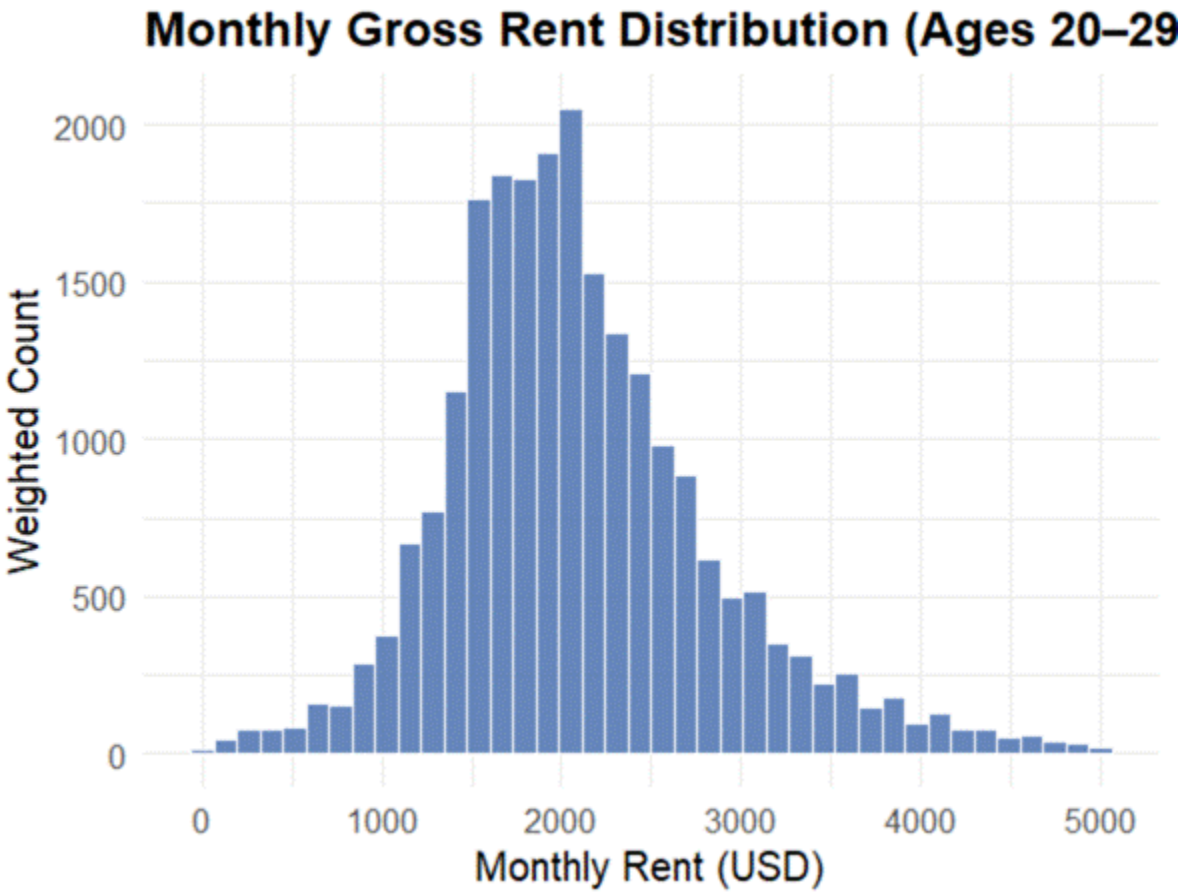
**Interpretation:**

- The **median** is slightly lower than the **mean**, suggesting a **right-skewed** distribution (a minority of high-rent units raising the average).
- Half of young adults pay between **$1,673 and $2,594** per month.
- Rents above **$3,000** are uncommon but present enough to influence the overall mean.

The histogram of monthly gross rent for young adults shows a **strong concentration of rent values** in the **$1,500–$2,500** range (Figure 6). This indicates that most renters in their twenties face mid-range rental costs nationwide. However, the distribution also displays a **long right tail**, reflecting a minority of young adults paying substantially higher rents—typically in high-cost metropolitan markets.

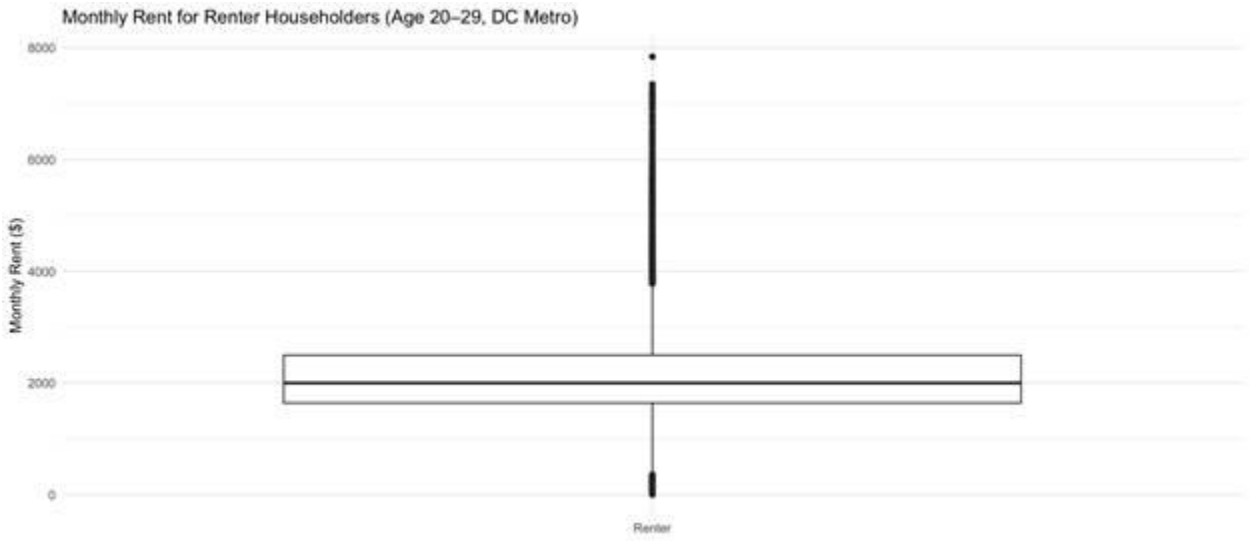
The shape of the distribution confirms a **classic right skew**, where most renters cluster around typical mid-range rents, and fewer renters pay premium prices. These high-rent outliers are responsible for the mean rent being slightly higher than the median.

**Figure 6: Monthly Gross Rent Distribution by Age**




The boxplot shown in Figure 7 illustrates the range of monthly rent paid by renter householders ages 20-29 in the Washington, D.C metro area. The box highlights the middle 50% of rent values; the lines inside mark the median. The whiskers indicate the typical lower and upper ends of the rent distribution, while the dots represent the higher priced outliers in the market.

**Figure 7: Monthly Rent Distribution for Renter Householders (20-29, DC Metro)**




**Application**

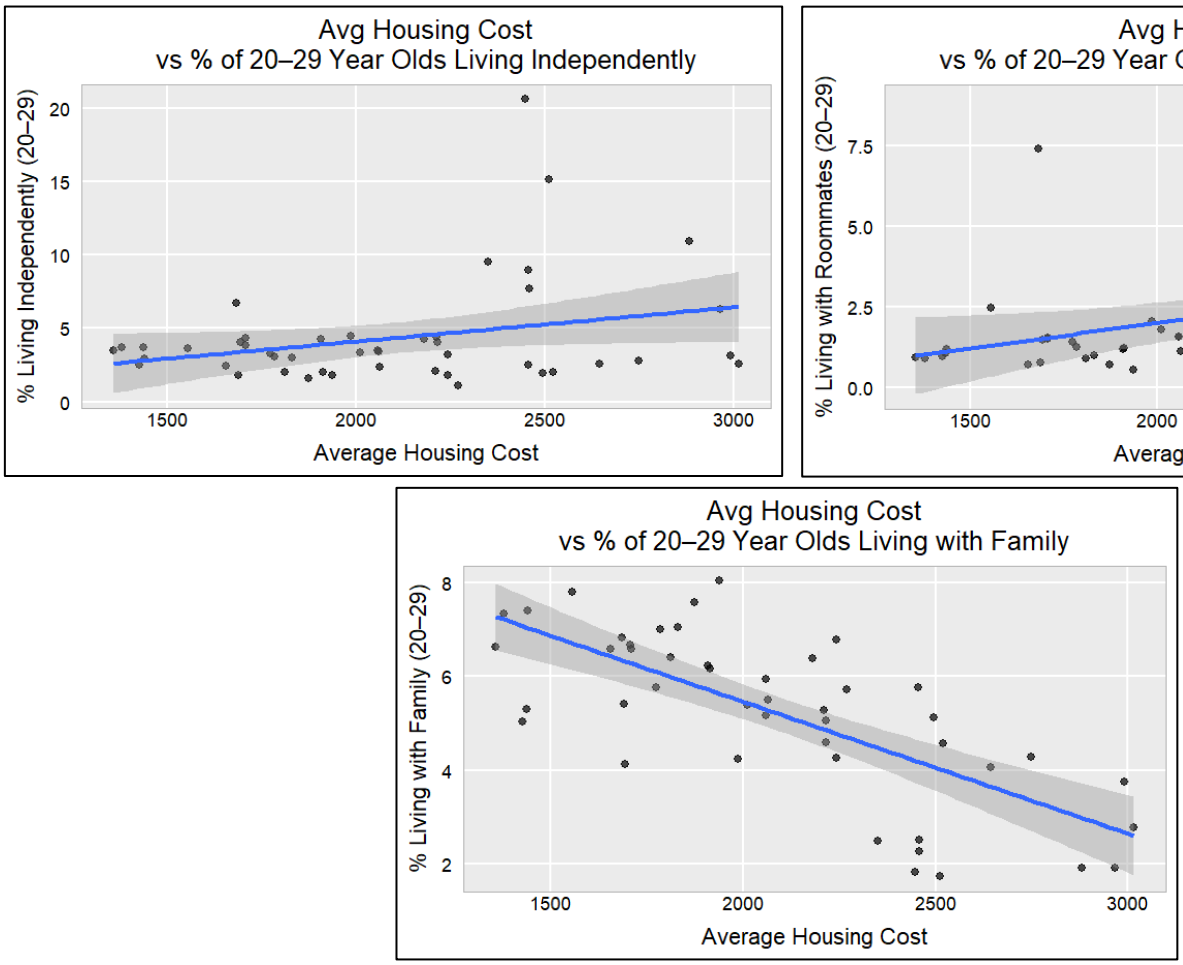
The boxplot is used to summarize and show how much young adult renter householders typically pay in rent and how rent levels vary in the region. This plot was created using R with the ggplot2 package and is already a part of the analysis script. The code complexity is roughly 5-10 lines and relies on basic filtering and plotting. Our group tested the implementation by confirming that the numbers of renter householders match the summary tables and checking that the median rent shown aligns with our calculated. The code was stored and version controlled in our GitHub repository.

**Risks Associated**

The main risks with this visualization relate to how viewers might interpret the boxplot. High rent outliers stretch the scale and make the typical rent range look a bit compressed, which may draw attention away from the median. The likelihood of this happening is medium, since outliers are visible and can draw attention away from the central trend. If this misinterpretation occurs, the impact is moderate. Another consideration that needs to be stated is that this is based on ACS data that may introduce bias. The probability of this affecting interpretation is low to medium. Ways to mitigate these risks are by having clear titles, captions, and consistent filtering to help ensure the visualization is interpreted correctly. As well as having additional notes in the report explaining the presence of outliers and the use of ACS microdata.

## 4.2 Distribution of Young Adult Living Arrangements

**Figure 8: Young Adult Living Arrangements**

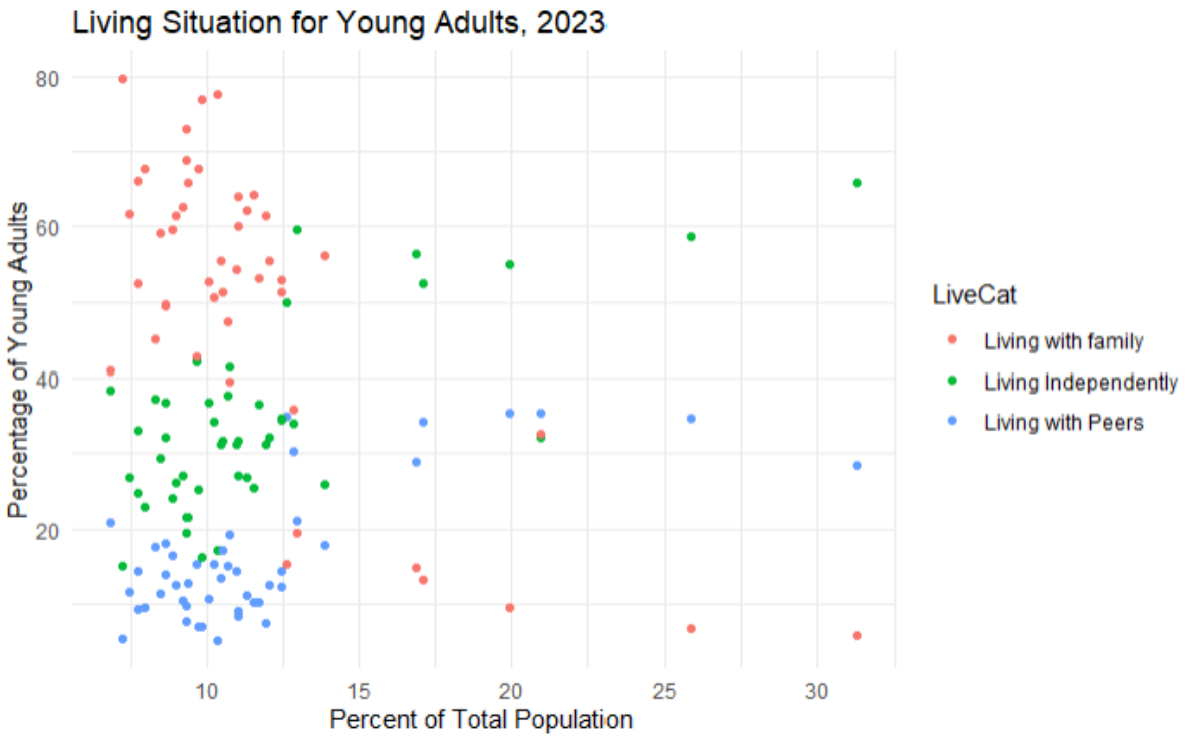



The three scatterplots above show the relationships between average housing costs and the living arrangements of the population of 20-29 year olds living in PUMAS around the DC metro area (Figure 8). The blue line shows the regression trend line of the dataset to better visualize the trend. The plots show that young adults are most likely to live with family members in areas where the average housing cost is lower. As the average cost of housing increases, young adults are more likely to live independently or with peers.

Figure 9 shows the relationship between the percentage of young adults in each living classification plotted against the proportion of young adults in the population for each PUMA. In many areas, most young adults in the 20 – 29 age band live with family members. It is interesting to note that when the concentration of people in their 20's exceeds about 12% of the total population these trends shift dramatically. Once the percentage of young adults crosses that threshold, far more of them live independently than live with roommates or family members. Based on the limited data shown here, it appears that the percentage of young

adults living independently is correlated with their relative representation in the overall community population. Because this range only covers six of the 47 PUMA areas studied, further investigation is needed about these areas. Due to the time constraints with this project, we do not have any additional information about those neighborhoods or any underlying reasons why they would have such a large percentage of independent young adult households.

**Figure 9: Percentage of Young Adult Living Situation vs Percentage of Population**




**Application**

The visualizations are used to show the cost burdens on young adults based on their living situations. These plots were created using the ggplot2 package on R. The code complexity consisted of 122 lines of code consisting of plotting, filtering and mutating the dataset. The team will test implementation by comparing the summary statistics with the outliers of the plots. The code will be stored and version controlled in our GitHub repository.

**Risks**

The risks associated with these plots are the axes and data interpretation. The axes of these plots differ due to the size of the different datasets and outliers stretching them. The risk is medium because viewers should be able to interpret the graphs accordingly, with the axes being relatively close in scale. Another risk is data interpretation due to the creation of variables. The team labeled young adults on their living arrangements based on information back assumptions of the original dataset. The risk is low because our team did the best based on the information given to us by IPUMS in regards to labeling of the original dataset.

## 4.3 Average Housing Costs in the Washington, D.C. Area

The areas in the southern end of the study area typically had lower average monthly housing costs than in areas closer to Washington, D. C. The lowest average monthly housing cost was in the District of Columbia (East) PUMA (#00104) where the average rent was $1,365 per month and the average mortgage cost was $1,346 per month. The most expensive area in the study area was the South Montgomery County- South Bethesda & Chevy Chase PUMA (#01008) where the average rent was $3,167 per month and the average mortgage was $2,977 per month. A complete list of all PUMA areas with the overall average housing cost, as well as the average mortgage rate and the average rent in the area is provided in Table 4.

The geographic distribution of the average monthly housing cost for renters is shown in Figure 10. Two PUMA areas have average rents that are significantly below the averages in the rest of the study area. In the Rappahannock-Rapidan Regional Commission PUMA (#06100), the average monthly rent is $1,377 per month.

It should be noted that this area is quite rural and covers the Counties of Fauquier, Rappahannock, Madison, Culpeper, and Orange in Virginia. Some towns in the PUMA are approximately 90 miles from Washington, D.C. Since it is considered part of the Washington, D.C. Metropolitan Statistical Area as defined by the U.S. Census Bureau, it was included in our study.

The other area with significantly lower rent is located within the boundaries of Washington, D.C. near the eastern border (District of Columbia (East)). The average rent in that area is $1,357 a month. This value may be incorrect for the area and warrants further study. It should be noted that a quick search of properties listed for rent on December 8, 2025 shows approximately 30 rental units in that area that are available for a monthly rate $1,300 or less; while there are nearly 200 units in the same area listed for rents in excess of $1,300 (Apartments.com, 2025). In addition, the calculated average monthly rent for this area is nearly $900 a month less than the next least expensive area in the District of Columbia. Due to project

time constraints, we are unable to investigate this issue further and flag it for follow up at a later date.

**Figure 10: Geographic Distribution of Average Monthly Rent**

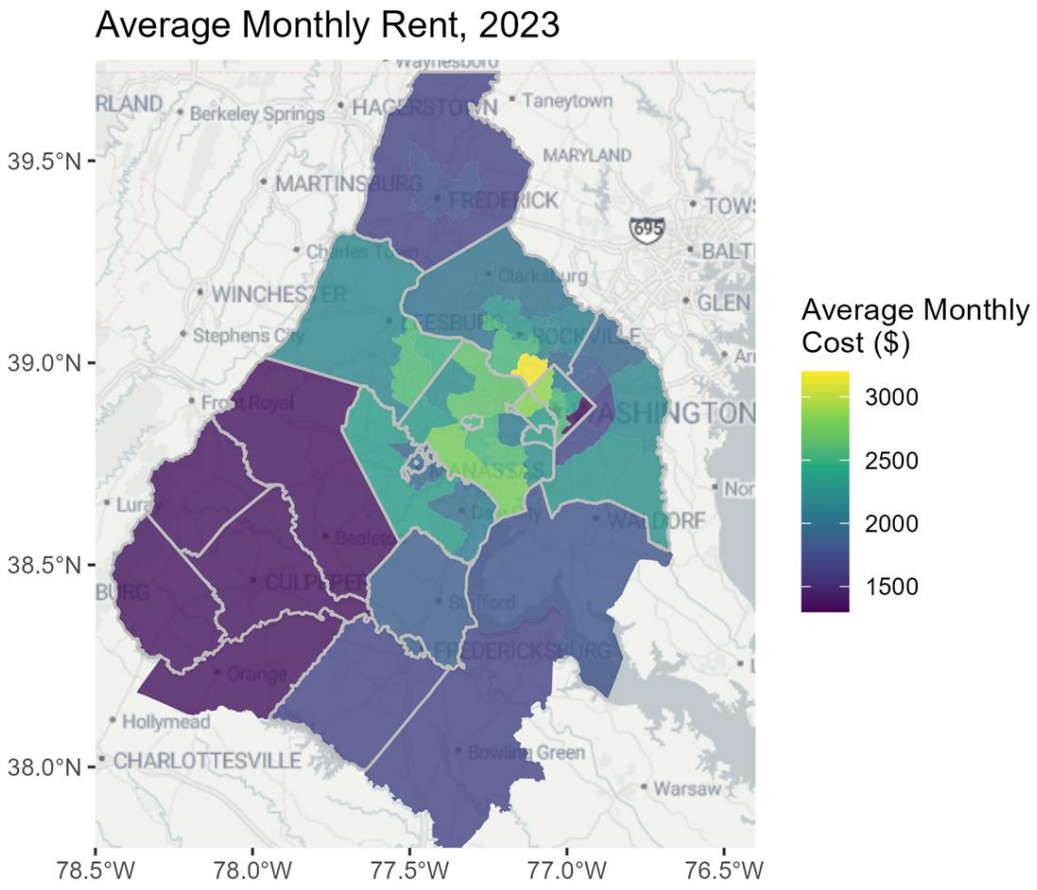


**Table 4: Mean and Median Monthly Housing Costs across the Washington, D. C. Area**

| PUMA | Description | Overall average housing cost | Average mortgage | Average rent |
|---|---|---|---|---|
| **District of Columbia** | | | | |
| 00101 | District of Columbia (West) PUMA | $2,966 | $3,010 | $2,871 |
| 00102 | District of Columbia (North) PUMA | $2,214 | $2,275 | $2,077 |
| 00103 | District of Columbia (Northeast) PUMA | $2,458 | $2,518 | $2,367 |
| 00104 | District of Columbia (East) PUMA | $1,357 | $1,346 | $1,365 |
| 00105 | District of Columbia (Central) PUMA | $2,512 | $2,468 | $2,551 |
| 00106 | District of Columbia (South Central) PUMA | $2,447 | $2,215 | $2,556 |
| **Maryland** | | | | |
| 00301 | Central Frederick County--Frederick City PUMA | $1,693 | $1,683 | $1,731 |
| 00302 | Outer Frederick County PUMA | $1,689 | $1,691 | $1,664 |
| 01001 | North Montgomery County--Olney, Damascus & Darnestown PUMA | $2,270 | $2,291 | $2,010 |
| 01002 | North Central Montgomery County--Germantown, Clarksburg & Montgomery Village PUMA | $1,784 | $1,719 | $2,008 |
| 01003 | Central Montgomery County--Gaithersburg & Rockville PUMA | $2,058 | $1,959 | $2,262 |
| 01004 | Southwest Montgomery County--North Bethesda & Potomac PUMA | $2,645 | $2,672 | $2,544 |
| 01005 | East Central Montgomery County--Wheaton, Aspen Hill & Glenmont PUMA | $1,912 | $1,845 | $2,157 |
| 01006 | East Montgomery County--Fairland, Colesville & White Oak PUMA | $1,810 | $1,720 | $2,105 |
| 01007 | Southeast Montgomery County--Takoma Park City & Silver Spring PUMA | $1,986 | $2,014 | $1,939 |
| 01008 | South Montgomery County--South Bethesda & Chevy Chase PUMA | $3,015 | $2,976 | $3,167 |
| 01101 | Inner Northwest Prince George's County--College Park City & Langley Park PUMA | $1,684 | $1,494 | $1,876 |
| 01102 | North Prince George's County--Laurel, Greenbelt & Beltsville PUMA | $1,707 | $1,596 | $1,937 |
| 01103 | Inner Northeast Prince George's County--New Carrollton & Hyattsville PUMA | $1,555 | $1,416 | $1,825 |
| 01104 | Central Prince George's County--Seat Pleasant City, Capitol Heights Town & Landover PUMA | $1,379 | $1,211 | $1,658 |
| 01105 | East Prince George's County--Bowie City, Kettering & Largo PUMA | $1,938 | $1,907 | $2,153 |
| 01106 | South Prince George's County--Clinton, Fort Washington & Rosaryville PUMA | $1,874 | $1,834 | $2,247 |
| 01107 | Southwest Prince George's County--Oxon Hill, Hillcrest Heights & Temple Hills PUMA | $1,439 | $1,304 | $1,704 |

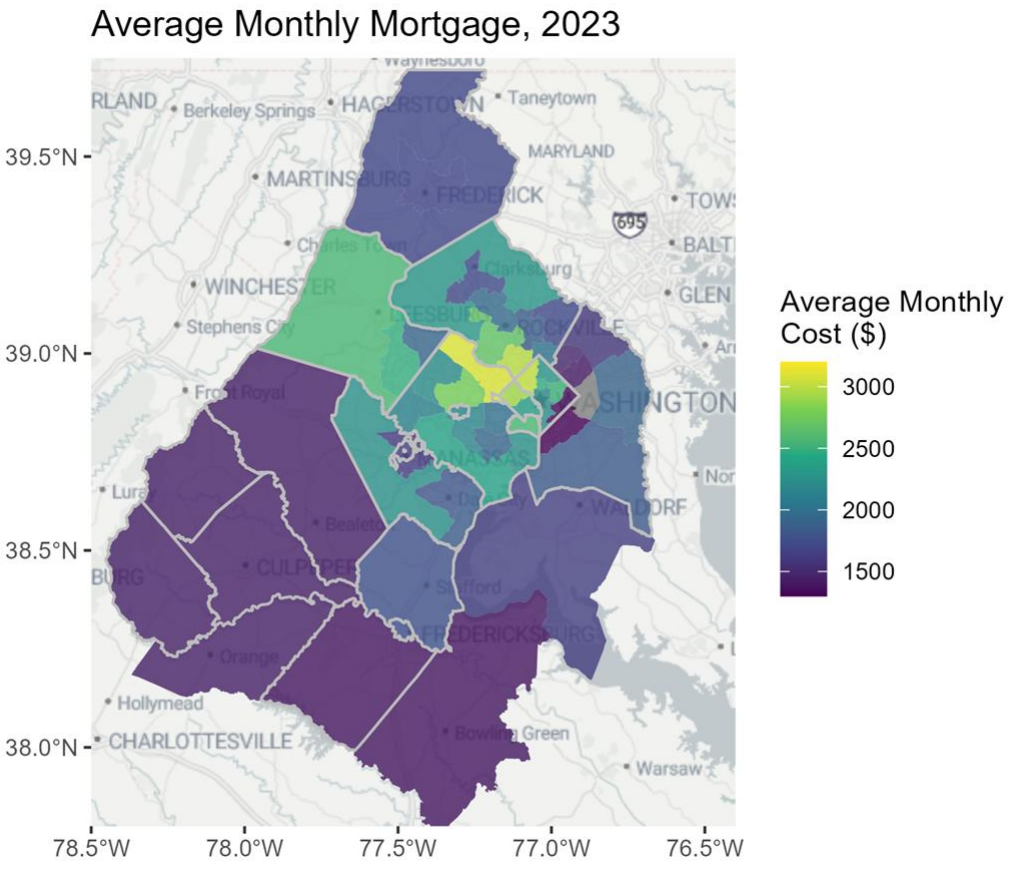
| | | | | |
|---|---|---|---|---|
| **Virginia** | | | | |
| 01301 | Arlington County (North) PUMA | $2,882 | $2,915 | $2,831 |
| 01302 | Arlington County (South) PUMA | $2,351 | $2,338 | $2,363 |
| 01700 | Charles County--La Plata Town & Waldorf City PUMA | $1,656 | $1,639 | $1,780 |
| 05901 | Fairfax County (West Central)--Centreville (North & West) & Chantilly (South & West) PUMA | $2,179 | $2,127 | $2,326 |
| 05902 | Fairfax County (Northwest)--Reston (North) & Franklin Farm PUMA | $2,216 | $2,215 | $2,220 |
| 05904 | Fairfax County (North Central)--Vienna Town, Oakton & Fair Oaks (East) PUMA | $2,748 | $2,771 | $2,654 |
| 05905 | Fairfax County (Southwest)--Centreville (Southeast) & Lorton PUMA | $2,455 | $2,376 | $2,767 |
| 05906 | Fairfax County (South Central)--Springfield (South), West Springfield & Franconia PUMA | $2,064 | $1,932 | $2,597 |
| 05907 | Fairfax County (East Central)--Annandale, West Falls Church & Bailey's Crossroads PUMA | $2,010 | $1,990 | $2,060 |
| 05908 | Fairfax County (Southeast)--Woodlawn, Rose Hill & Mount Vernon PUMA | $2,243 | $2,246 | $2,228 |
| 06100 | Rappahannock-Rapidan Regional Commission PUMA | $1,426 | $1,437 | $1,377 |
| 10701 | Loudoun County (Northeast)--Sterling & Ashburn (Northeast/Ashburn Village) PUMA | $2,059 | $1,922 | $2,509 |
| 10702 | Loudoun County (East Central)--Ashburn (Southwest/Belmont Ridge) & Dulles Airport PUMA | $2,495 | $2,482 | $2,573 |
| 10703 | Loudoun County (West & South)--Leesburg Town PUMA | $2,521 | $2,561 | $2,168 |
| 15301 | Prince William County (North & West)--Linton Hall & Gainesville PUMA | $2,242 | $2,234 | $2,331 |
| 15302 | Prince William County (Northeast) PUMA | $1,831 | $1,798 | $1,976 |
| 15303 | Prince William County (East) PUMA | $1,908 | $1,845 | $2,068 |
| 17700 | George Washington Regional Commission (South) PUMA | $1,436 | $1,393 | $1,666 |
| 17900 | George Washington Regional Commission (North) PUMA | $1,773 | $1,755 | $1,854 |
| 51000 | Alexandria City PUMA | $2,458 | $2,572 | $2,314 |
| 60001 | Fairfax County (Central)--Fairfax City & Burke PUMA | $2,210 | $2,153 | $2,540 |
| 61001 | Fairfax County (Northeast)--McLean & Idylwood PUMA | $2,992 | $3,093 | $2,650 |
| 68701 | Manassas, Manassas Park Cities & Prince William County (Central)--Buckhall & Sudley PUMA | $1,708 | $1,627 | $1,938 |

The area with the highest average rent was located close to the Washington, D.C. border in southern Montgomery County, Maryland. The average rent in that area was $3,167 per month. Other areas that tended to have higher average rents were in the suburbs close to Washington, D.C., including a large portion of Fairfax County, Virginia and the eastern portion of Loudoun County, Virginia.

## Average Monthly Mortgage

The average monthly mortgage payment was also mapped across the Washington, D.C. study area (Figure 11). Like the patterns seen in the average rental rates, the most expensive areas were close to the city, with less expensive areas in the rural, southwestern portion of the study area. The most expensive neighborhoods were in southern Montgomery County, Maryland, western Washington, D.C., and northern Fairfax County, Virginia. The range of monthly mortgage costs shows more variability across the region than the average monthly rent. In general, the eastern suburbs show lower average monthly mortgage costs than western suburbs at an equivalent distance from the city center. The reasons for this discrepancy go beyond the scope of this study and warrant further inquiry in the future. Like the rental average, the very eastern point of the Washington, D.C. city limits shows a relatively low average mortgage payment. Unlike the rental average, this area of lower cost extends into the western portion of Prince George's County, Maryland.

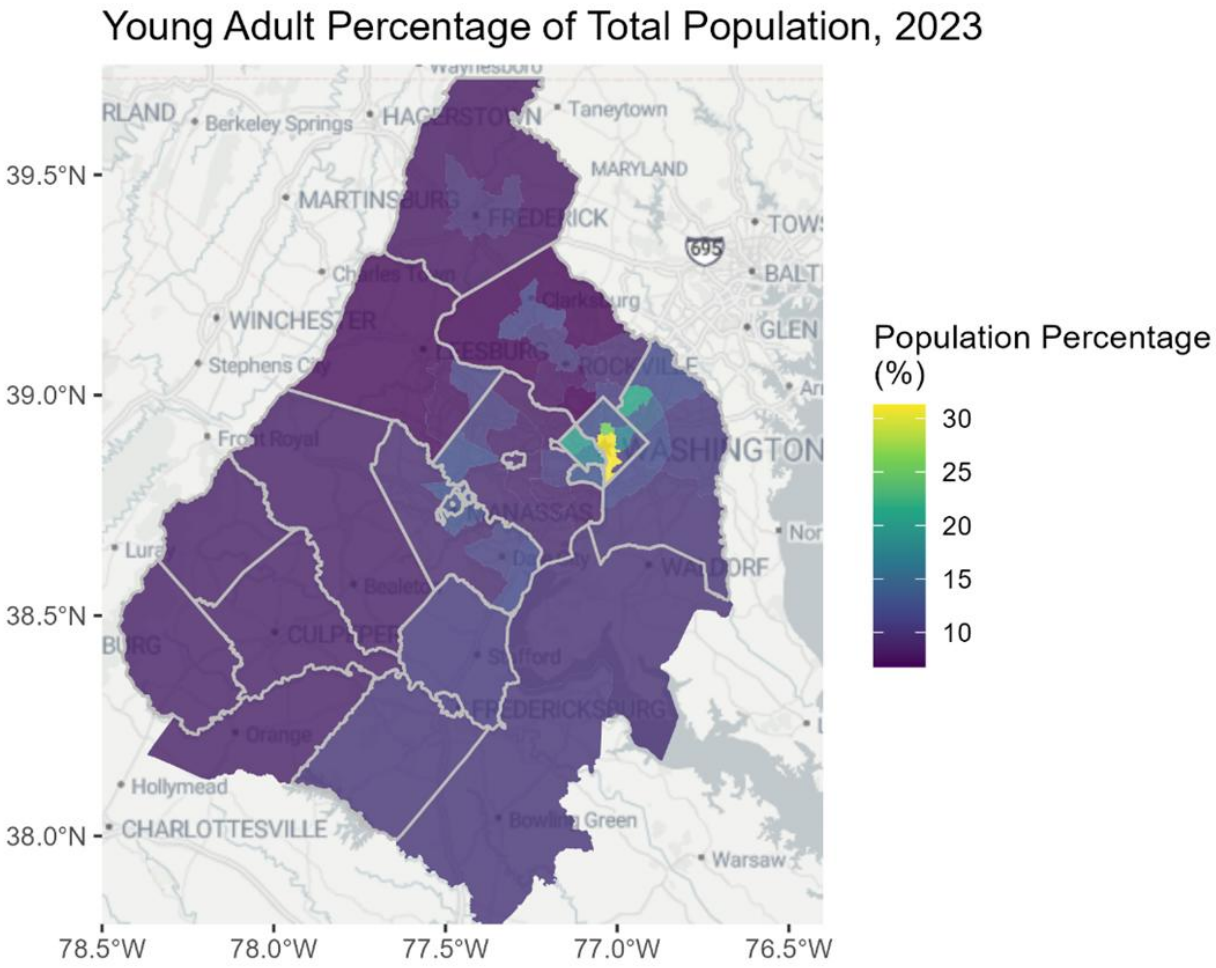
**Figure 11: Geographic Distribution of Average Monthly Mortgage Payment**



## 4.4 Where do young adults live?

Figure 12 shows the distribution of young adults in their 20's as a percentage of the overall population. The District of Columbia (South Central) PUMA has the highest percentage of young adult residents (31.3%). The District of Columbia (Central) PUMA and the Inner Northwest Prince George's County—College Park City & Langley Park PUMA follow closely behind with young adult populations reaching 25.8% and 20.9%, respectively.

**Figure 12: Percentage of Population Ages 20 – 29 by PUMA**



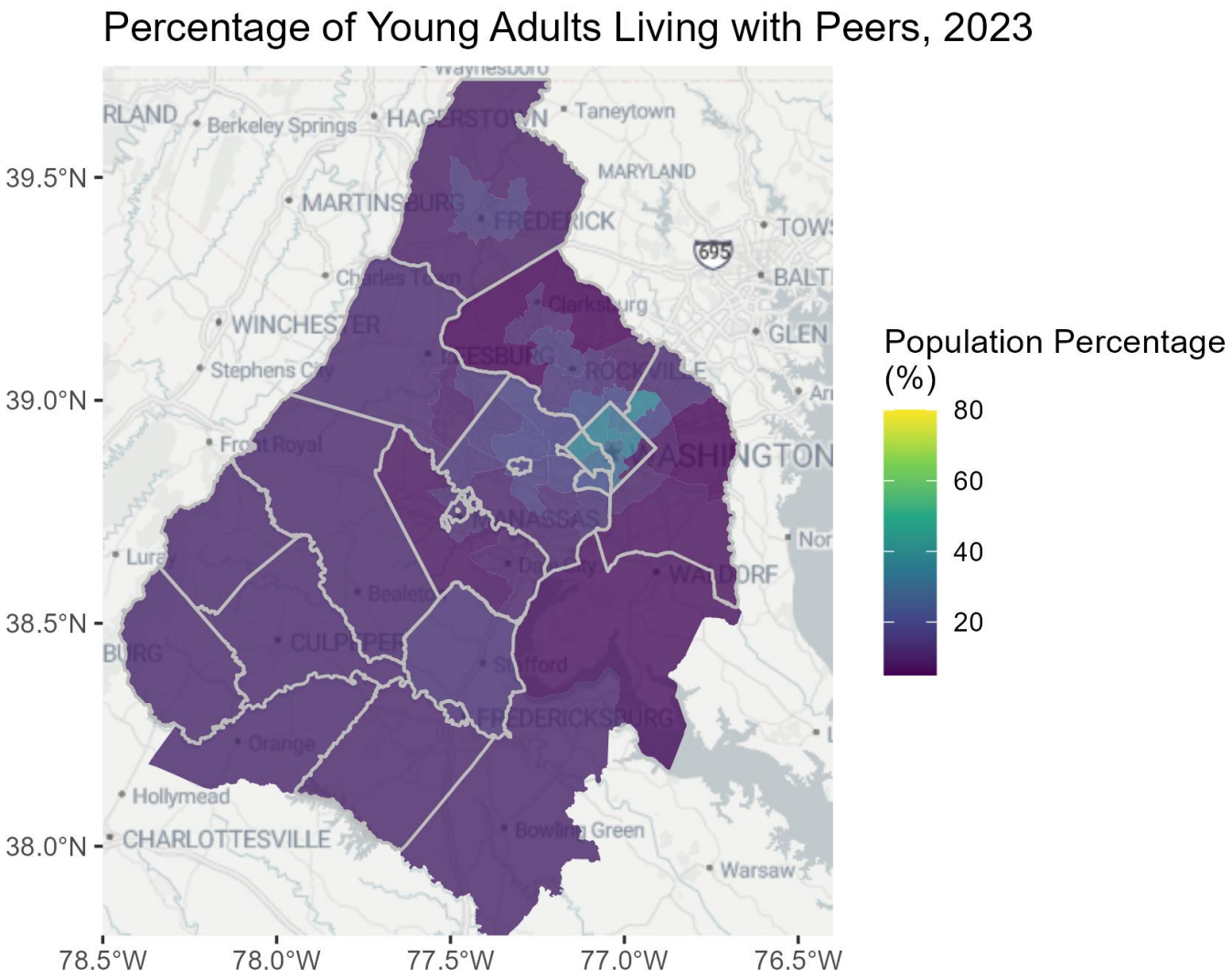

These numbers may be skewed because of the colleges and universities located in these areas. The University of Maryland-College Park, a major research university, is located within the "Inner Northwest" PUMA (#01101). The central part of Washington, D.C. also has a number of major universities including Howard University, George Washington University, and Galludet University. While not technically within the boundaries of the PUMAs listed here, Washington, D.C. is also home to Georgetown University, American University, Catholic University and the University of the District of Columbia. All of these institutions likely draw young adult residents to the nearby neighborhoods.

For the purpose of this study, the young adult respondents were grouped into three "living situation" categories based on their response in the RELATE variable in the dataset that indicates the relationship between the survey respondent and the "head of the household" using the procedure outlined on Page 22 of this report. Young adults were classified as "Living Independently" if their RELATE variable indicated that they were the head or spouse of the household. They were classified as "Living with Family" if they chose a family relationship for the RELATE variable. All other records were classified as "Living with Peers" (usually with roommates).

### Living with Peers

Respondents that were classified as "Living with Peers" showed less variation across the study area (Figure 13). There is a slightly higher concentration within the limits of Washington, D.C, where approximately 30% of young adults aged 20 – 29 live with peers. The lowest concentration of young adults living with roommates/peers was found in eastern Prince George's County, Maryland (PUMA #01105) and northern Montgomery County, Maryland (#01001) where approximately 5% of young adults lived with friends or roommates.

**Figure 13: Percentage of Young Adults Living with Peers, 2023**



### Living Independently

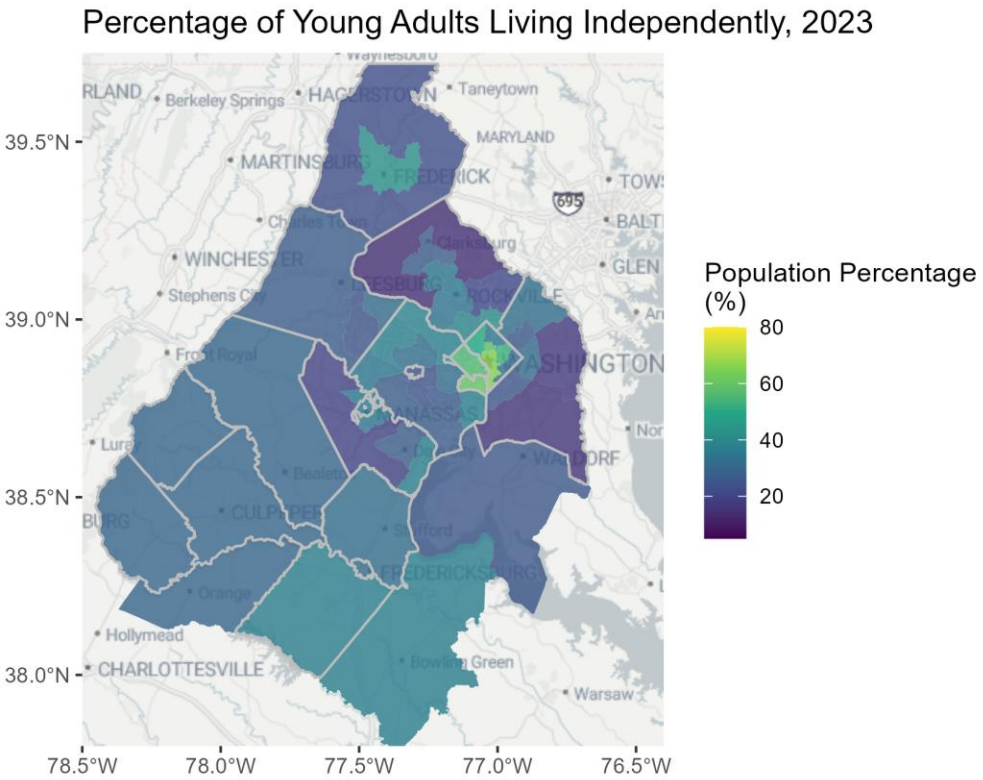
The percentage of young adults that were living independently in 2023 is mapped in Figure 14. The highest concentration of independent young adults can be found within close proximity to the Washington, D.C. city limits. Table 5 describes the regions where more than 50% of young adults live independently.

**Table 5: Areas with more than 50% of young adults living independently**

| PUMA | Area | % Young Adults Living Independently |
|---|---|---|
| 00106 | District of Columbia (South Central) PUMA | 65.8% |
| 51000 | Alexandria City PUMA | 59.5% |
| 00105 | District of Columbia (Central) PUMA | 58.6% |
| 01302 | Arlington County (South) PUMA | 56.3% |
| 01301 | Arlington County (North) PUMA | 55.0% |
| 00103 | District of Columbia (Northeast) PUMA | 52.5% |

There is a lower percentage of young adults living independently outside of the city limits. The areas of upper Montgomery County, Maryland (PUMA #01001, 15.0%) and southern Prince George's County, Maryland (PUMA #01106, 16.1%) have the lowest concentration of young adults living independently.

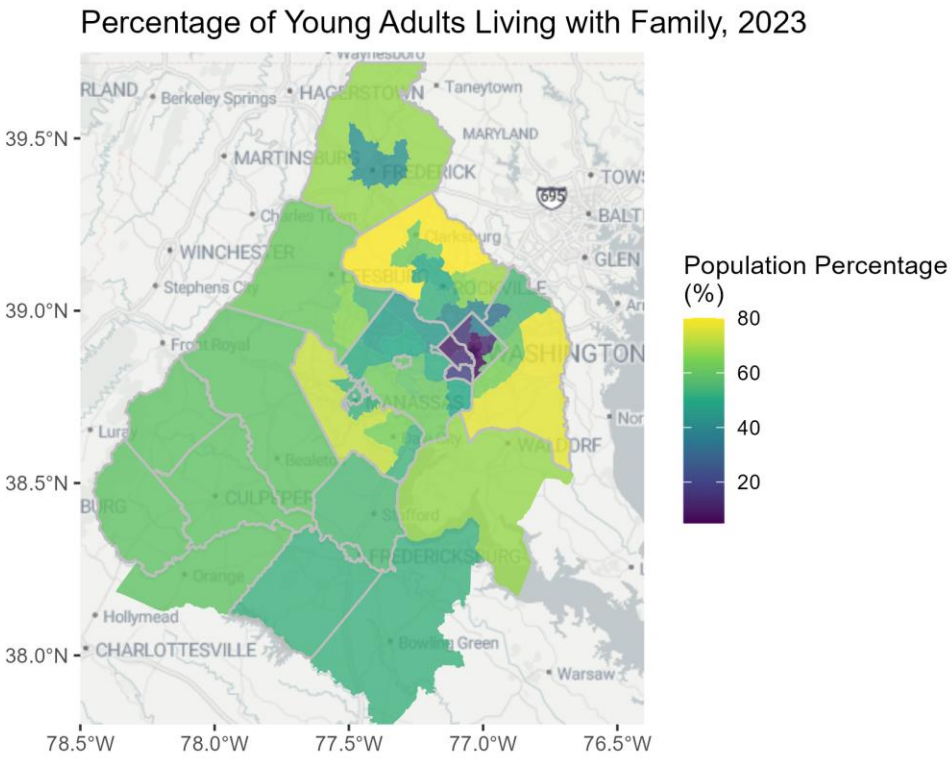
**Figure 14: Percentage of Young Adults Living Independently**



### Living with Family

The number of young adults living with family members has the widest variation of all three living situation classifications. In northern Montgomery County, Maryland (PUMA #01001), approximately 80% of young adults ages 20 through 29 live with other family members. In contrast, just 6% of young adults in the south central area in the District of Columbia (PUMA #00106) live with family. The geographic distribution of the percentage of young adults living with family members is shown in Figure 15. In some ways, Figure 15 appears to be an inverse of the data shown in Figure 14. The areas with the highest concentration of young adults living with family members correspond to the areas with the lowest concentration of young adults living independently.

**Figure 15: Percentage of Young Adults living with family, 2023**




## 4.5 Cost Burden Analysis

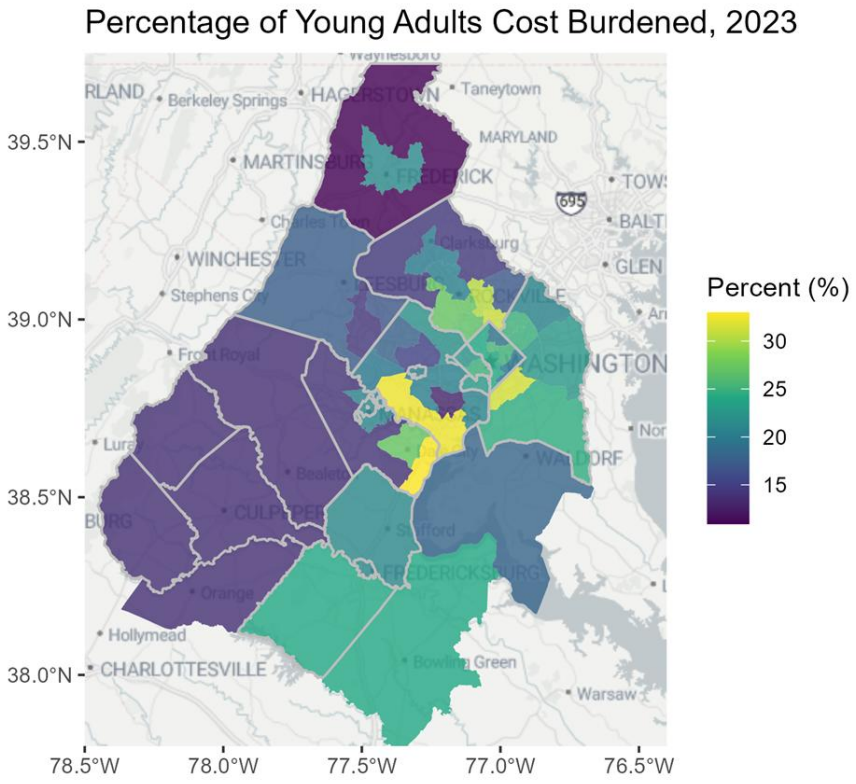
Young adults (ages 20 - 29) who were classified as living independently were categorized as "***Not Cost Burdened***", "***Cost Burdened***", or "***Severely Cost Burdened***" based on the percentage of income that went towards housing costs (Table 6).

**Table 6: Ratio between Household Income, Housing Cost, and Cost Burden Category**

| Ratio between Housing Cost and Income | Cost Burden Category |
|---|---|
| < 0.3 | Not Cost Burdened |
| ≥ 0.3 | Cost Burdened |
| ≥ 0.5 | Severely Cost Burdened |

The distribution of the percentage of young adults who are cost-burdened is mapped in Figure 16. Far fewer of the young adults living in the western suburbs are cost-burdened than their peers living further east. The areas with the highest percentage of independent young adults who were cost burdened are summarized in Table 7.

**Figure 16: Percentage of Cost Burdened Young Adults by PUMA**




The region immediately around the University of Maryland in College Park had the highest rate of cost-burden with over 63% of people in their twenties paying more than 30% of their income to cover their housing cost. The lowest percentage of cost burdened households were located

in Outer Frederick County, Maryland (PUMA #00302) where approximately 21% of young adults were considered to be const-burdened.

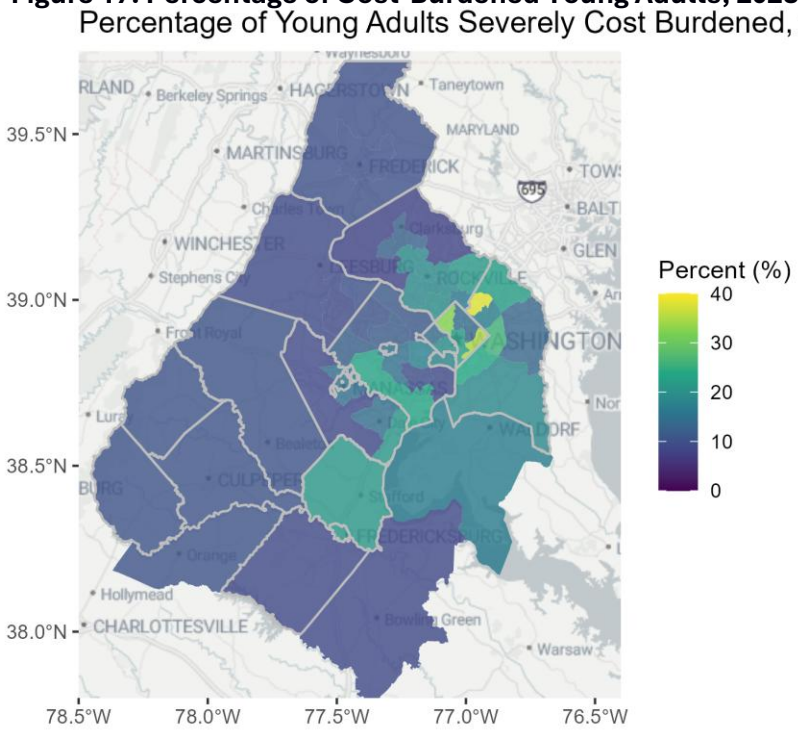
**Table 7: Areas with high concentrations of cost-burdened independent young adults**

| PUMA | Area | % Cost Burdened Young Adults |
|---|---|---|
| 15303 | Prince William County (East) PUMA | 53.8% |
| 00104 | District of Columbia (East) PUMA | 54.3% |
| 01107 | Southwest Prince George's County--Oxon Hill, Hillcrest Heights & Temple Hills PUMA | 55.1% |
| 05905 | Fairfax County (Southwest)--Centreville (Southeast) & Lorton PUMA | 55.4% |
| 00101 | District of Columbia (West) PUMA | 56.6% |
| 01101 | Inner Northwest Prince George's County--College Park City & Langley Park PUMA | 63.1% |

## Severe Cost Burden

When housing expenses consume more than 50% of a household's income, they are considered to be "severely cost burdened". While there are fewer households that fall into this category, there are still many pockets in the Washington, D.C. area where there are significant numbers of young adults spend more than half their paycheck on housing-related costs (Figure 17).

**Figure 17: Percentage of Cost-Burdened Young Adults, 2023**



The distribution of households that are severely cost burdened follows similar geographic patterns to the less extreme "cost burden" distribution. The area with the largest number of severely cost-burdened households is located in the vicinity of the University of Maryland (PUMA # 01101) where approximately 38% of independent household heads spend at least 50% of their income on housing. Following closely behind, approximately 30% of young adults in the District of Columbia (East) (PUMA #00104) and District of Columbia (West) (PUMA #00101) spend at least 50% of their income on housing expenses. By contrast, young adults in northern Montgomery County, Maryland (PUMA #01001) and western Prince William County, Virginia are the least likely to be severely cost burdened, with just approximately 8% of independent young adults paying more than 50% of their income towards housing.

## 4.6 Rent and Owner Burden Distribution by Year/Pooled

**Figure 18: Rent Burden Distribution for Young Adult Renters (2021-2023)**

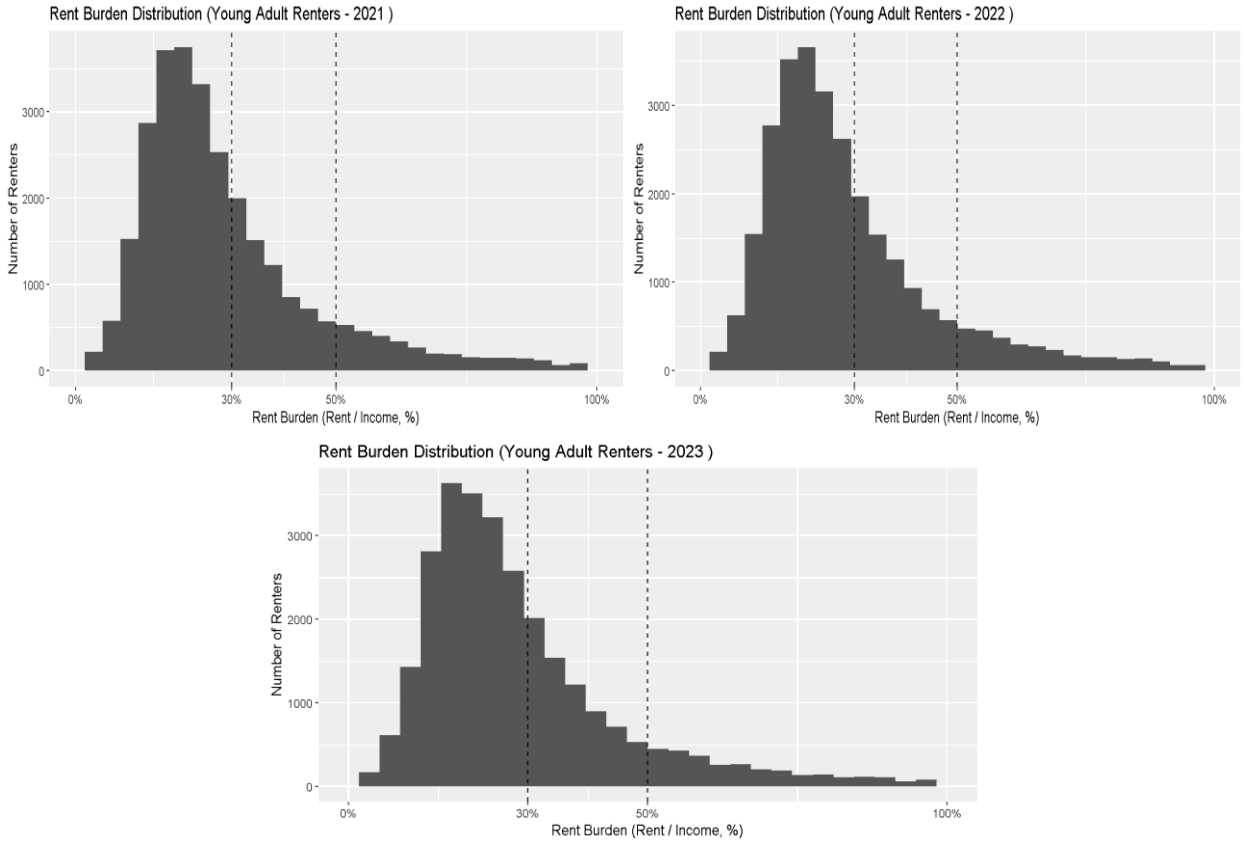



Figure 18 summarizes the rent burden of young adult renters across 2021-2023. In each histogram, the distribution has a peak centered right before the 30 percent cost-burden threshold. All three histograms show a right-skewed distribution with a long right tail. This

indicates most young renters are 20-30 percent cost-burdened, but a smaller group is facing severe cost-burden above 50 percent.

**Figure 19: Rent Burden Distribution for Young Adult Renters (Pooled Data)**

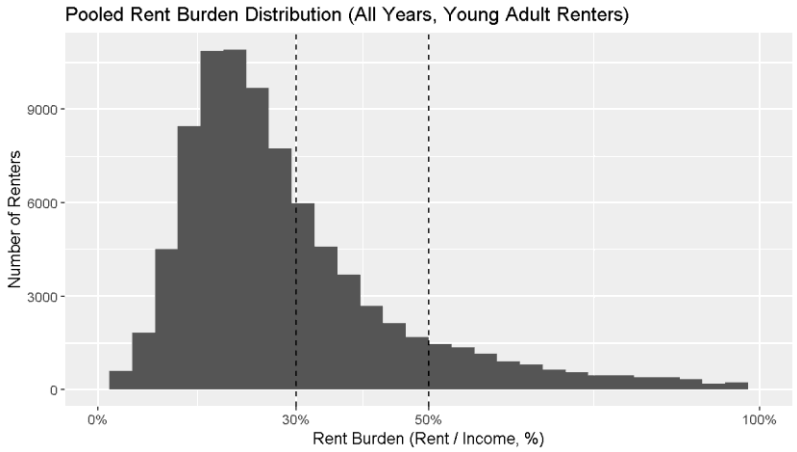



Figure 19 combines all young adult renters from 2021-2023 into a single pooled sample and plots the distribution of rent burden. The histogram shows clear indication of a right-skewed distribution that has a center peak right under 30 percent. There is a significant number of renters in the cost-burdened area (30%-50%) with a long right tail that extends beyond 50 percent. This means that while we have a huge group of renters in the 20-30 percent cost-burdened range, there exists a small important group that faces a severe rent burden that is above 50 percent.

**Figure 20: Owner Mortgage Burden Distribution for Young Adult Renters (2021-2023)**

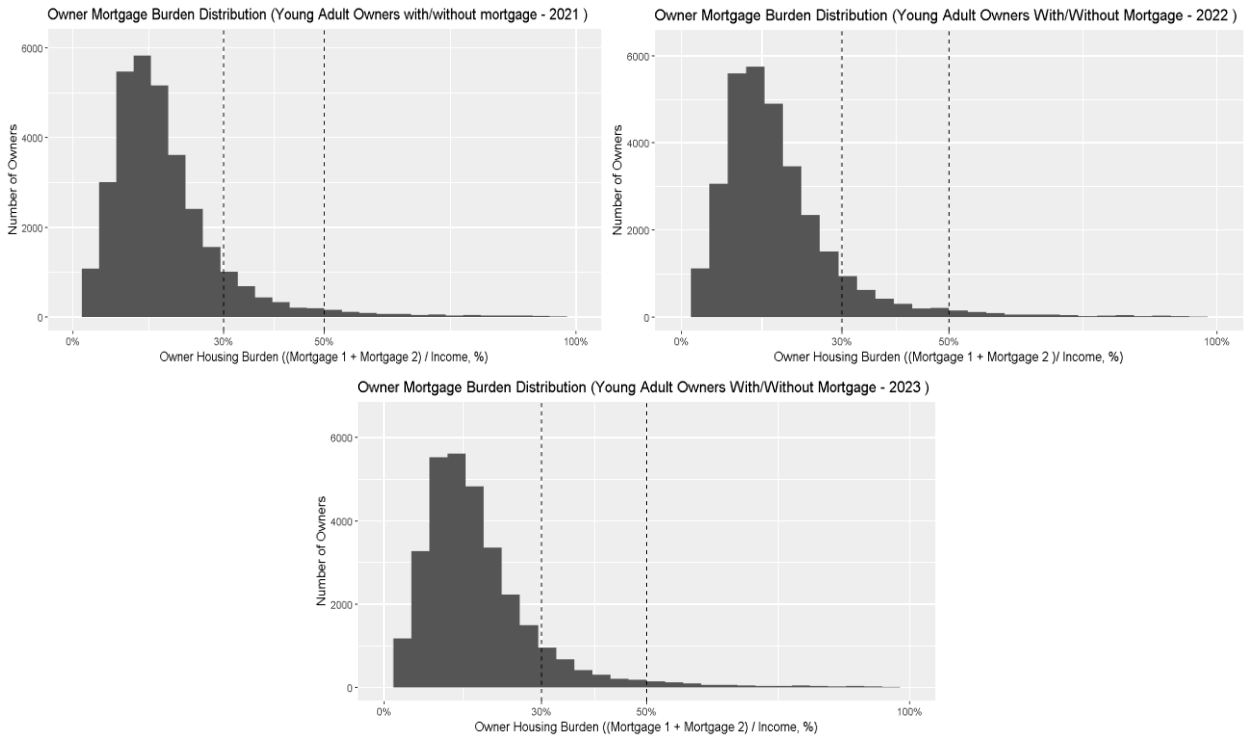



Figure 20 shows the owner's mortgage burden distributions from 2021-2023. Owners include people with and without mortgages. The owner's mortgage burden is calculated as the sum of the first and second mortgage payments divided by the household income. All yearly histograms are right-skewed with a peak centered below the 30 percent cost-burden threshold. Slightly lower for owners than renters. There are still a significant number of owners in the cost-burden range of 30 – 50 percent. There is also a long tail that goes right. Cost-burdened and severely cost-burdened categories are significantly less than the renter categories for cost-burdened.

**Figure 21: Owner Mortgage Burden Distribution for Young Adult Renters (Pooled Data)**

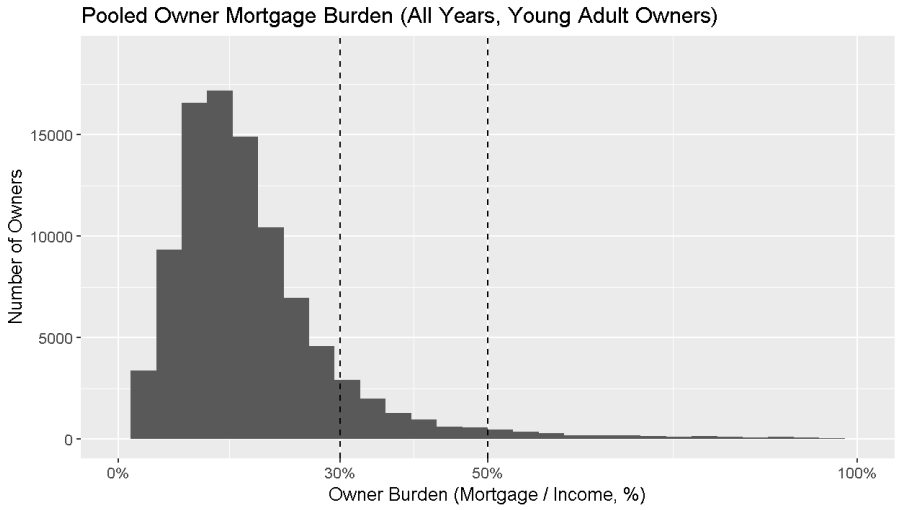



Figure 21 shows the pooled owner's mortgage distribution for young adults. The distribution is right skewed with a peak centered at a 15 percent owner mortgage burden. This is lower than the peak of the pooled renter's cost-burden distribution. There exists a long right tail in the severely cost-burdened category, but it is significantly lower than the severely cost-burdened category for the pooled renter's cost burden distribution.

**Application**

The histograms in figures 17 – 20 are being used to describe the housing cost burden that is distributed among young adult renters and owners. We are comparing the distributions of owners and renters. We compare each year from 2021 to 2023 and include a pooled distribution of all three years. We are looking to see if the distributions are shifting towards a higher burden over time. These comparisons help us support the main findings section by showing many households are centered below the 30 percent threshold and have a long right tail above 50 percent. This is coded in R (open source). The visualization tool is from the ggplot2 library. The histogram and vertical line thresholds are created using geom_histogram, geom_vline, scale_x_continuous, labs. This will use about 40 lines of code to implement per distribution. The code complexity is medium level. The only care is when setting values for certain classifications, because ggplot will be expecting a number for the axis. To test the implementation, we will examine the summary statistics of the burden variables and verify the counts of the histograms to match the data in the tables. The code used to plot the histograms will be in our GitHub repository.

**Risks**

Audience could misinterpret the bars as percentages or not understand that the thresholds for housing cost burden are what is considered for severe and not severe cases. This can lead to incorrect conclusions about how serious the burden is for renters and owners. The probability of this occurring is medium. Many people use histograms, but it's possible someone not in the field of analytics could see it confusing. The impact can be medium to high. If the visualizations are misunderstood, the audience might underestimate or overestimate young adults who are cost burdened. The avoid any confusion we need to use clear axis labels, include descriptive notes, complement histograms with summary tables. The risk is currently mostly mitigated. I will ask the team if they think a summary table or descriptive notes would be helpful in my visualizations.

# 5 Findings

In this research, the visualizations and analysis reveal that housing affordability for young adults in the Washington, D.C. metropolitan area is highly uneven across space, tenure type, and living arrangements. Using the 2021-2023 ACS microdata from IPUMS, we observe that there is substantial variation in both housing costs and cost-burden outcomes across PUMAS, and even within the same metro area. Average monthly rents range from roughly $1,300 in lower-cost PUMAs to more than $3,100 in high-cost areas such as South Bethesda and Chevy Chase. Mortgage costs follow the same general pattern, though far fewer young adults are homeowners.

**Average Housing Cost**

The lowest average monthly housing cost was in District of Columbia (East) PUMA #00104, with an average rent of $1365 and an average monthly mortgage payment of $1,346 per month. These values appear anomalous and need further investigation. The monthly rent is significantly below the surrounding areas and an online search of available rental properties in December 2025 showed a small number of available properties in that price range. Most available rentals were priced significantly higher (Apartments.com, 2025).

In contrast there are high-cost PUMAS such as South Montogomery County PUMA #01008. They have reported the highest average rent of $3,167 and the highest average mortgage $2,976. Also, Fairfax County Northeast PUMA #61001 has reported a high average mortgage of $3,093 and a high average rent of $2,650. Both regions are considered suburbs.

**Living Arrangements**

The scatterplots relating average housing cost to living arrangements show that in lower cost areas, young adults often live with family, but in higher-cost PUMA areas they have a larger share of independent living or living with peers. Figure 8 compares the shares of young adults. In PUMAs where young adults make up more than about 12% of the population, independent living becomes the dominant arrangement, a pattern observed consistently across the subset of areas that meet this threshold.

The heat maps show the differences in the distribution of varying living situations across the study area during 2023. District of Columbia (South Central) has 65.8 percent. Alexandria City shows 59.5 percent. District of Columbia (Central) shows 58.6 percent. North and South Arlington Country show 55 percent, and 56.3 percent respectively. Outside the city, the share of indent young adults drops significantly. Upper Montgomery Country PUMA #01001 has about 15 percent of young adults living independently and Southern Prince Goerge's PUMA #01106 has about 16.1 percent. Meanwhile, in Northern Montogomery Country 80 percent of young adults live with family compared to 6 percent in District of Columbia (South Central).

**Cost Burden**

Rent-burden distributions consistently show a strong right skew, with many young adults spending close to the 30% HUD affordability threshold and a smaller but meaningful share exceeding the 50% severe burden threshold. The highest cost-burden rates appear in eastern D.C., eastern Prince William County, southwestern Prince George's County, and the University of Maryland corridor, where more than half of independent young adults spend over 30% of their income on housing and up to 38% spend more than half.

In the cost-burden analysis we find that our map shows us that cost-burden among young adults is clustered by location. Fewer young adults are cost-burdened in western suburbs and in eastern, campus adjacent locations show a higher percentage of young adults cost-burdened. The severely cost burden heatmap shows it is less common but still concentrated. Near the University of Maryland PUMA #01101 shows 38 percent of independent young adults are severely cost-burdened. In D.C (East and West) 30 percent of young adults spend at least 50 percent of their income on housing. Northern Montogomery Country and Western Prince William County are less likely to be severely cost burdened with approximately 8 percent of independent young adults paying more than 50 percent of their income on housing.

In the analysis of Renters versus Owner distributions, we find peaks in the distribution right below the 30 percent threshold to be not cost-burdened for each year and in the pooled distribution. Most renters and owners are in the 20-30 percent burden range but show groups in the cost-burdened range and in the severely cost-burdened range. In the severely cost-burdened range there is a long right tail indicating they have a wide variety of severely cost-burdened groups. Overall, young adult renters are more likely to be cost burdened than owners based on the size of comparison of yearly distributions

The summary statistics of the monthly rent distribution in 2021 show the average rent is $2089, and the median is $2,067. The $25^{th}$ percentile shows $1,673. The $75^{th}$ percentile shows $2,584. Since the median is lower than the mean, this suggests a right-skewed distribution. 50 percent of young adults pay between $1673 and $2594 per month. Rent costs above $3,000 aren't as common but present a strong enough influence on the average.

Overall, the findings show that housing experiences for young adults in the D.C. region vary widely by location, with major differences in independence rates, average housing costs, and the prevalence of cost burden.

# 6 Summary

Overall the study found that young adults in the Washington D.C. metro area experience highly uneven housing arrangements shaped by where they live. The sharp geographic differences in independence rates, housing costs, and affordability demonstrate that no single experience represents the region as a whole. Young adults in and near D.C. are far more likely to live independently, even in the highest-cost areas, while those in outer suburbs often remain with family despite lower average housing costs. This suggests that other factors not examined in this study may influence living decisions as strongly as affordability itself.

Affordability pressures are widespread but concentrated unevenly. Although many young adults remain near the 30% HUD affordability threshold, specific areas show significantly higher rates of both cost burden and severe cost burden. These patterns indicate that a substantial share of young adults are struggling to meet housing costs despite regional economic growth.

The findings highlight a complex housing landscape in which location strongly shapes both living arrangements and financial burden. The study provides a data-driven foundation for understanding these disparities and underscores the need for targeted approaches to improving housing affordability for young adults.

# 7 Future Work

This study was an exploration of the relationships between income, geography, living status, and cost burden for young adults in the Washington, D.C. region. Our initial inquiry has revealed multiple aspects that warrant additional study.

Due to the time constraints with this project, we did not examine the impacts of race, gender, age, marital status or education level on the geographic distribution of young adults or their likelihood to experience a housing cost burden. Expanding the parameters included in the study may provide additional insights into varying rents and cost burdens across the region. This deeper understanding of the underlying phenomena would support more informed decision making to guide housing policy decisions.

Future work should investigate possible correlations identified in this study. For example, our results indicate that young adults are most likely to live independently in areas with the highest average housing cost. This unexpected finding suggests that there may be additional underlying area factors that drive decisions around living situations. Neighborhood characteristics and composition, proximity to major employment centers, housing supply, and income distribution may all play a part and warrant further inquiry.

Revising the age range included in the study, for example from 25 - 34 instead of 20 - 29, may also provide additional insights into the housing cost burden for younger residents. This change would reduce the influence of undergraduate students, who tend to have lower reported incomes and atypical housing arrangements. Their inclusion may skew the burden metrics even when their situations do not actually reflect a long-term financial instability.

Another avenue for future improvement is to implement a more accurate method for predicting housing costs. The methodology used in this study produced a number of records where the rent or mortgage cost was recorded as zero or a negative number. While it is possible that some young adults may be fortunate enough to live independently without a rent or mortgage payment, it is unlikely to occur at the rates seen in our data cleaning.

In summary, the areas highlighted show that this analysis has taken the initial steps towards understanding housing pressures on young adults in the D.C. region. Expanding the demographic variables and the age groups, exploring the unexpected correlations we identified as well as improving cost-estimation methods would help create a more complete picture. Future work in these areas would strengthen the insight that has been gained and help make more informed housing policies and planning.

# Appendix A: Links to Code

The analysis and visualizations created in this study were created in R (R Core Team, 2024). The GitHub linked below contains the relevant code used in this project.

**GitHub Repository:** https://github.com/Juan-GMU/CS504

The team used a number of open-source packages in the development of this study. A select list of packages used is provided below:

| R Package Name | Reference |
|---|---|
| dplyr | (Wickham et al, 2023) |
| ggmap | (Kahle & Wickham, 2013) |
| ggplot2 | (Wickham, 2016) |
| ipumsr | (Freedman Ellis, Burk & Roberts, 2025) |
| scales | (Wickham, Lin Pedersen, & Seidel, 2025) |
| sf | (Pebesma & Bivand, 2023) |
| tidycensus | (Walker & Herman, 2025) |
| tidyverse | (Wickham et al, 2019) |
| tigris | (Walker, 2025) |

# Appendix B: Risk

- Risk: Our data is too broad
    - Probability: medium
    - Impact: medium
    - Mitigations:
        - Take subset of data
    - Focus on a smaller population and scale the system
- Risk: Failure to obtain appropriate data
    - Probability: low
    - Impact: high
    - Mitigation:
        - Census.gov
        - Find other data sources

- Risk: Analytic methods fail
    - Probability: high for any one method-data combination
    - Impact: medium
    - Mitigations:
        - Prototype early
        - Apply multiple techniques
        - Communicate with Professor and Teammates as problems arise

# Appendix C: Agile Development

## Join Genius Team

Leslie Rodriguez: Project Owner
Jennifer Reif: Scrum Master
Juan Parada: Developer
Roger Rocha-Claros: Developer
Ryan Sollom: Developer

## Sprint 0: Team Formation

**Tools Used:** Microsoft Teams (communications), Zoom (online meeting), YouTrack (scheduling)
**Issues:** Identify team member roles, begin problem identification and scope
**Deliverables:** Summary of team name, roles, and a brief problem description
**Challenges:** Scheduling a common meeting time

## Sprint 1: Problem Definition and Project Plans

**Tools Used:** Microsoft Teams (meeting and communications), YouTrack (scheduling)
**Issues:** Refine problem identification, begin identification of potential data sources, start literature review and background research, identify risks and a mitigation plan
**Deliverables:** Sections 1.1 - 1.8 in paper template, corresponding slides
**Challenges:** Scheduling a common meeting time, data retrieval from the Census Bureau website was extremely slow and requests frequently timed out, identified a “Plan B” and obtained data hosted by the University of Minnesota in the IPUMS data repository (Ruggles et al, 2025).

## Sprint 2: Data Sets

**Tools Used:** Microsoft Teams (meeting and communications), YouTrack (scheduling)
**Issues:** Finalize data selection, obtain data and begin data cleaning
**Deliverables:** Sections 2.1 - 2.6 of paper template, corresponding slides
**Challenges:** Scheduling a common meeting time

## Sprint 3: Analytics / Algorithms

**Tools Used:** Microsoft Teams (meeting and communications), YouTrack (scheduling)
**Issues:** Developing and implementing algorithms for analysis, assessing code complexity, identifying and mitigating risks
**Deliverables:** Section 3 of paper draft and corresponding slides
**Challenges:** Scheduling a common meeting time

## Sprint 4: Visualizations

**Tools Used:** Microsoft Teams (meeting and communications), YouTrack (scheduling)
**Issues:** Define and implement visualizations, asses code complexity, identify risks and mitigation strategies
**Deliverables:** Section 4 (Visualizations) of the paper template and corresponding slides
**Challenges**: Scheduling a common meeting time

## Sprint 5: Reporting / Final Presentation

**Tools Used:** Microsoft Teams (meeting and communications), YouTrack (scheduling)
**Issues:** Draw conclusions, make recommendations for future work, finalize project paper and presentation
**Deliverables:** Final paper and presentation slides
**Challenges:** Scheduling a common meeting time, consolidating the work into one cohesive, concise “story” for the presentation

# Appendix D: References


198 Apartment Rentals Found Nearby. (n.d.). *Apartments.com.* Retrieved December 8, 2025, from https://www.apartments.com/over-1300/?bb=1t93yx0k0HqsrzpD

Apartments.com. (2025, May 29). *What renters in Washington, D.C. need to know in 2025*. https://www.apartments.com/blog/what-renters-in-washington-d.c.-need-to-know-in-2025

Bhasin, K., Loungani, P., & Sunderji, A. (2024, March 14). *Hitting Home: Housing Affordability in the U.S.* Econofact. Available: https://econofact.org/hitting-home-housing-affordability-in-the-u-s [Accessed 1 Nov 2025].

Braveman, P., Dekker, M., Egerter, S., Sadegh-Nobari, T., & Pollack, C. (2011, May 1). *Housing and Health*. Robert Wood Johnson Foundation. Available: https://www.rwjf.org/en/insights/our-research/2011/05/housing-and-health.html [Accessed 1 Nov 2025]

Coley, R. L., Leventhal, T., Lynch, A. D., & Kull, M. (2013). Relations between housing characteristics and the well-being of low-income children and adolescents. *Developmental Psychology, 49*(9), 1775–1789. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3766502/ [Accessed 1 Nov 2025]

Council for Community and Economic Research. (2024, May 31). *PRESS RELEASE – For Immediate Release – Q1 2024 – Cost of Living Index*. https://www.coli.org/press-release-for-immediate-release-cost-of-living-index/

DC Fiscal Policy Institute. (2025, April 14). *Nearly Half of All Renters and More Than Half of Black Renters in DC Struggle to Afford Rent*. https://www.dcfpi.org/all/nearly-half-of-all-renters-and-more-than-half-of-black-renters-in-dc-struggle-to-afford-rent/

Desilver, Drew, "A look at the state of affordable housing in the U.S.," *Pew Research Center*, Oct. 25, 2024. Available: https://www.pewresearch.org/short-reads/2024/10/25/a-look-at-the-state-of-affordable-housing-in-the-us/ [Accessed 1 Nov 2025]

Feiveson, L., Levinson, A., & Wertz, S. S. (2024, June 24). *Rent, house prices, and demographics*. U.S. Department of the Treasury. https://home.treasury.gov/news/featured-stories/rent-house-prices-and-demographics [Accessed 2 Dec 2025]

Flournoy, R., Reddy, A., Stagg, K., Davis, V., Haggerty, R., Grossman, D., & American Public Health Association, AcademyHealth, and Kaiser Permanente. (2021). *Housing*

*Affordability and quality: a community driver of health*. https://www.apha.org/-/media/files/pdf/pubs/housing_health_community_driver.pdf [Accessed 1 Nov 2025]

Freedman Ellis, G., Burk, D., Roberts, F. (2025). _ipumsr: An R Interface for Downloading, Reading, and Handling IPUMS Data. R package version 0.9.0, <https://CRAN.R-project.org/package=ipumsr>.

Fry, R. (2025, April 17). *The shares of young adults living with parents vary widely across the U.S.* Pew Research Center. https://www.pewresearch.org/short-reads/2025/04/17/the-shares-of-young-adults-living-with-parents-vary-widely-across-the-us/ (pewresearch.org)

Gaitán, V. (2019, January 2). *How housing affects children's outcomes*. Housing Matters. Available: https://housingmatters.urban.org/articles/how-housing-affects-childrens-outcomes

Garasky, S., Stewart, S., & Acock, A. (2001). *Group living decisions as youths transition to adulthood*. *Demography*, *38*(3), 443–461

Gray, M. N. "Problem: overcrowding," *City Journal*, Mar. 23, 2023. Available: https://www.city-journal.org/article/problem-overcrowding [Accessed 1 Nov 2025].

Haurin, D. R., Hendershott, P. H., & Kim, D. (1993). *The impact of real rents and wages on household formation*. *Review of Economics and Statistics, 75*(2), 284–293.

Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL: http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Mahoney, K. (2024, October 4). *The cost of living in Washington, DC, for millennials in 2024*. Illumint. https://illumintfc.com/what-is-the-cost-of-living-in-washington-dc-for-millennials/

Mulder, C. H. (2006). Population and housing. *Demographic Research*, *15*, 401–412. https://doi.org/10.4054/demres.2006.15.13

North Carolina Housing Finance Agency. (n.d.). *Housing Research*. Available: https://www.nchfa.com/about-us/research-reporting-and-policy/housing-research [Accessed 1 Nov 2025].

Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Roy, J., Maynard, M., & Weiss, E. (2008, December 16). *The hidden costs of the housing crisis: The long-term impact of housing affordability and quality on young children's odds of*

*success*. The Pew Charitable Trusts. Available: https://www.pew.org/en/research-and-analysis/reports/2008/12/16/the-hidden-costs-of-the-housing-crisis-the-longterm-impact-of-housing-affordability-and-quality-on-young-childrens-odds-of-success [Accessed 1 Nov 2025]

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J., Richards, S., Rodgers, R., Schroeder, J., & Williams, K. C. W., *IPUMS USA: Version 16.0* [dataset]. Minneapolis, MN: IPUMS, 2025. https://doi.org/10.18128/D010.V16.0

Office of Disease Prevention and Health Promotion. (n.d.). *Quality of housing*. In Healthy People 2030. U.S. Department of Health and Human Services. Available: https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/quality-housing [Accessed 1 Nov 2025].

Office of Management and Budget. (2023, July 21). *OMB Bulletin No. 23-01: Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas* (OMB Bulletin No. 23-01). Available: https://www.whitehouse.gov/wp-content/uploads/2023/07/OMB-Bulletin-23-01.pdf

Paciorek, A. D. (2013). *The long and the short of household formation* (Finance and Economics Discussion Series No. 2013-26). Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/pubs/feds/2013/201326/201326pap.pdf

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

U.S. Bureau of Labor Statistics. (2025a). *Median usual weekly earnings of full-time wage and salary workers by age and sex* [Chart]. https://www.bls.gov/charts/usual-weekly-earnings/usual-weekly-earnings-current-quarter-by-age.htm

U.S. Bureau of Labor Statistics. (2025b, May 8). *Occupational employment and wages in Washington-Arlington-Alexandria — May 2024*. https://www.bls.gov/regions/mid-atlantic/news-release/occupationalemploymentandwages_washingtondc.htm

U.S. Census Bureau. (2024, September 12). *Cost of Rent and Utilities Rose Faster Than Home Values in 2023*. (A) Available: https://www.census.gov/library/stories/2024/09/acs-rent-burden.html

U.S. Census Bureau (n.d.) *2024 American Community Survey/Puerto Rico Community Survey Group Quarters Definitions*. Available: https://www2.census.gov/programs-surveys/acs/tech_docs/group_definitions/2024GQ_Definitions.pdf [Accessed 16 Nov 2025]

U.S. Census Bureau. (2024, September 12). *Nearly half of renter households are cost-burdened, proportions differ by race* (Press Release No. CB24-150). (B) Available: https://www.census.gov/newsroom/press-releases/2024/renter-households-cost-burdened-race.html [Accessed 1 Nov 2025]

U.S. Census Bureau. (2023, September 7). *Public Use Microdata Areas (PUMAs)*. Available: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html. [Accessed: Nov. 11 2025]

U.S. Department of Housing and Urban Development. (n.d.). *CHAS: Background* [Dataset background]. HUD USER. [Online]. Available: https://www.huduser.gov/portal/datasets/cp/CHAS/bg_chas.html. [Accessed: Nov. 10, 2025].

U.S. Department of Housing and Urban Development. (2006, April 6). *Glossary of terms to affordable housing*. https://archives.hud.gov/local/nv/goodstories/2006-04-06glos.cfm [Accessed 1 Nov 2025]

Virginia Department of Housing and Community Development & Virginia Housing. (2022, Jan 14). *HB854 Statewide Housing Study: Current efforts, future needs, new strategies — Part 1: Why affordability matters*. https://dmz1.dhcd.virginia.gov/HB854/part-1-affordability.html [Accessed 1 Nov 2025].

Walker K (2025). _tigris: Load Census TIGER/Line Shapefiles_. R package version 2.2.1, <https://CRAN.R-project.org/package=tigris>.

Walker K, Herman M (2025). _tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames_. R package version 1.7.3, <https://CRAN.R-project.org/package=tidycensus>.

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr

Wickham H, Pedersen T, Seidel D (2025). _scales: Scale Functions for Visualization_. R package version 1.4.0, <https://CRAN.R-project.org/package=scales>.