Nora Mendez
UA591 Applied Analytical Methods
Luis Santiago
6 December 2023

Rent burden in the city of Boston, 2022

In this project I explore the phenomenon of rent burden in the city of Boston with 2022 data. Rent burden, which is the housing cost burden specific to renters (as opposed to owners/buyers), has become a pressing issue in most US big metros. It is a popular rule of thumb that spending up to 30% of your gross income on rent is a financially sound measure. This budgeting standard dates back to 1981 when the US government stipulated that those spending more than 30% on housing were "cost-burdened." To be severely rent burdened, which happens increasingly to more people, is to be spending 50% or more of your earnings on rent.

There is much literature and research done on the drivers, impacts, and characteristics of cost-burdened populations in different cities. Boston, as one of the most expensive cities in the US (up with San Francisco, Los Angeles, and and New York) has some of the highest cost of rent rates in the country, and with that, a high proportion of cost-burdened and severely cost-burdened renters. Why is this? And how can these cities work toward making housing more affordable for its residents? The goal of this project is to look at the data that relates rent burden with household income in the city of Boston and to test their correlation. I attempt to use a regression model to quantify the influence that factors such as income have on housing affordability. Though obvious, this analysis provides evidential rigor to the understanding that the trend of increasing rent burden affects those with lower incomes most and who are already disadvantaged with limited housing options. Measuring how rent burden affects low income populations can prove useful in any effort to decrease the burden of housing.

Literature Review

I used the following research and articles to inform this project:

- Najjar, Amelia. *Renting in Boston*, boston.gov
- *American Families Face a Growing Rent Burden*, Pew Research Center

- *When the Rent Eats From Incomes Large and Small, Is the Traditional Measure of Cost Burden Still Useful*, HUD User
- *If Renters Weren't Paying Too Much Rent, They Could Spend More on Family Needs And in The Community*, National Equity Atlas
- *A Typology of U.S. Metropolises by Rent Burden And Its Major Drivers*, Link Springer

Rent burden is determined from the calculation of a household's expense on rent (including utilities) over its gross income:

$$\frac{monthly\ rent}{household\ income}$$

From this equation we can see that if you decrease what you pay for rent you will be less rent burdened, and likewise if you increase your yearly gross income. It makes sense to consider average incomes in an area as an indicator of a higher probability of rent burden, especially if the rental market is already tight. I will test this relationship in the city of Boston.

However, there are many potential determinants that factor into rent burden, such as demographic characteristics including race/ethnicity, immigration histories, age, gender, marital status, presence of children, single parenthood and household size. Here is a list provided by researchers Mikhail Samarin and Madhuri Sharma about how different research has shed light on how these interact, often in complex ways:

- Immigrants from developing countries and undocumented immigrants both suffer from higher housing cost and rent burden compared to U.S.-born minority counterparts.
- Similarly, disadvantaged minorities, particularly Blacks and Latinx, are more affected by rent burden.
- Some immigrant groups may experience lower rent burden (compared to other low-income groups) because of living in multigenerational or multifamily households, thereby increasing their aggregate household income relative to rent.

- Higher educational attainments in a metropolis and the greater presence of more educated cohorts both increase housing prices, thus making rentals less affordable for low- and moderate-income groups.
- At the same time, higher educational attainments at the individual level reduce the incidence of rent burden among more educated people due to their higher income potentials.
- Single female-headed households with children are more susceptible to living in poverty because of the double penalty of being able to work part-time only.
- Higher income inequality aggravates unaffordability among low-income families because the presence of high-income groups inflates housing prices and simultaneously reduces low-cost options.
- The extent and intensity of rent burden are also attributed to lower housing vacancy rates and an inadequate supply of low-income housing.

Their research paper actually seeks to push for a theory that considers the occupational specializations of metropolises through a geographical lens. What this means is that increasing rent burden should be considered as a problem at the regional economic level and local job market level as well.

Data sources and methodology

I will be using US Census Bureau microdata collected from the American Community Survey (ACS) 1-Year Estimates Public Use Microdata Sample (2022). With the ACS, data is collected on the individual level, but it is a sample of the overall population it represents. Though having inherent sampling error, the Census Bureau uses this data to provide estimates on a range of demographic characteristics by person and by household. The PUMS (Public Use Microdata Sample) is a set of sample data that the public is given access to, that is detailed enough for analysts to perform their own estimates.

Specifically, the Census Bureau MDAT tool allows users to select the desired variables and tailor a table with the data needed of the sample population, in the geographies selected (Public Use Microdata Areas). For my analysis, I selected five geographic regions, listed below:

- Boston city–Allston, Brighton & Fenway (PUMA 801)
- Boston city–Back Bay, Beacon Hill, Charlestown, East Boston, Central & South End (PUMA 802)
- Boston city–Dorchester & South Boston (PUMA 803)
- Boston city–Mattapan & Roxbury (PUMA 804)
- Boston city–Hyde Park, Jamaica Plain, Roslindale & West Roxbury (PUMA 805)

All of these together comprise the city of Boston (excluding Winthrop, part of Suffolk county).

Choosing the variables was a little more tricky, as not all of them have the same units of measurement. As mentioned above, the ACS collects data at the levels of the individual person, or the household. For example, the variable I am most interested in this project, labeled GRPIP, measures gross rent as a percentage of household income in the past 12 months. I am using this variable as a measure of rent burden *by household*. In any cross-tabulation with this variable I should use variables with household units of measurement. That means that with this data it becomes more difficult to analyze household rent burden against other interesting variables such as age, race, gender, natality, etc, that are usually measured on the individual person level. Thankfully, ACS has recoded variables for some of these that are adjusted to the household unit of measurement. I wanted to look at race and was able to use the variable HHLDRRAC1P which is the ‘recoded detailed race code of the householder’ and assigns the householder’s race to a category from 1 to 9. The other variables I chose to extract the data with are included in the following:

- Household income (HINCP), interval-ratio, 0 to 99999999
- Geographies (the PUMA neighborhood clusters listed above), categorical 801-805
- Monthly rent (RNTP), interval-ratio, 1-99999, excluded 0

- Gross rent as a percentage of household income in the past 12 months (GRPIP), interval-ratio, 1-100 percent, excluded N/A coded as 0 (Group quarters, owned or being bought/occupied without rent payment/no household income)

  Once I had uploaded the data to SPSS I was able to start running analyses.

Findings

- **Descriptive analysis:**

I ran some descriptives for the interval-ratio variables GRPIP, RNTP and HINCP, caring about their mean and median for the 5 geographies combined

Statistics

		GRPIP	HINCP	RNTP
N	Valid	174749	174749	174749
	Missing	0	0	0
Mean		41.29	93958.51	1867.92
Median		31.00	61200.00	1800.00
Mode		101	100000	2000
Std. Deviation		29.059	115071.405	1204.219
Percentiles	25	20.00	25000.00	1000.00
	50	31.00	61200.00	1800.00
	75	53.00	125000.00	2500.00

The average renter in the city of Boston has a household income of almost 94K, pays 1867.92 in monthly rent, and has a rent to income ratio of 41.29. This means the average household renter is rent burdened. The median household income is lower at $61,200, with rent of $1800, and rent to income ratio of 31%, meaning at least half of the households in the sample are rent burdened. They are also considered slightly rent burdened. Since the median is lower than the mean, that means the majority (more than half) is below the average, probably because a smaller number of households are making over-proportionally more (for HINCP at least), skewing the curve to the right. We can also see with the 75% percentile that over 25% of households are severely rent burdened.

- **Creation of the BUSTA variable:**

I created a variable first in Excel and then passed it over to SPSS, that categorized the rent to income variable GRPIP into severely rent burdened, rent burdened, or not rent burdened, as such:

BUSTA: If GRPIP ≤ 30 → 0 (not rent burdened)
If 30 < GRPIP ≤ 50 → 1 (rent burdened)
If GRPIP > 50 → 2 (severely rent burdened)

Then, I ran a frequency table. According to the results, the percent of rent burdened households in the sample to be 100 – 48.6 = 51.4, and the percent of severely rent burdened households to be 100 – 72.8 = 27.2. This makes sense with the quartile values we calculated earlier for GRPIP.

BUSTA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	84847	48.6	48.6	48.6
	1	42457	24.3	24.3	72.8
	2	47445	27.2	27.2	100.0
	Total	174749	100.0	100.0	

Frequency table for the new variable BUSTA that categorizes GRPIP into three groups, 2022 data

- **GRPIP, HINCP and RNTP comparison of median values across the different PUMA:**

When looking at these three variables by specific neighborhood clusters (PUMA), we can observe patterns as well. GRPIP seems to increase with lower HINCP across these 5 geographical groups, and decrease with higher median HINCP. RNTP doesn't seem to correlate strongly with either.

Statistics

		RNTP	GRPIP	HINCP
N	Valid	35726	35726	35726
	Missing	0	0	0
Mean		2124.25	43.86	91178.67
Median		2000.00	34.00	66860.00
Mode		1900	101	60000

PUMA 801

Statistics

		RNTP	GRPIP	HINCP
N	Valid	48648	48648	48648
	Missing	0	0	0
Mean		2056.29	40.93	102541.58
Median		2000.00	32.00	70000.00
Mode		2200	101	100000

PUMA 802

Statistics

		RNTP	GRPIP	HINCP
N	Valid	31832	31832	31832
	Missing	0	0	0
Mean		2068.84	37.81	117117.13
Median		2000.00	27.00	86000.00
Mode		5800	101	145000

PUMA 803

Statistics

		RNTP	GRPIP	HINCP
N	Valid	36366	36366	36366
	Missing	0	0	0
Mean		1266.76	44.17	62725.84
Median		1200.00	35.00	38000.00
Mode		1200	101	35000

PUMA 804

Statistics

		GRPIP	HINCP	RNTP
N	Valid	22177	22177	22177
	Missing	0	0	0
Mean		38.19	97583.21	1739.13
Median		28.00	71450.00	1800.00
Mode		101	160000	1400

PUMA 805

PUMA	RNTP	HINCP	GRPIP
Allston, Brighton & Fenway (PUMA 801)	$2000	$66860	34%
Back Bay, Beacon Hill, Charlestown, East Boston, Central & South End (PUMA 802)	$2000	$70000	32%
Dorchester & South Boston (PUMA 803)	$2000	$86000	27%
Mattapan & Roxbury (PUMA 804)	$1200	$38000	35%
Hyde Park, Jamaica Plain, Roslindale & West Roxbury (PUMA 805)	$1800	$71450	28%

Median values for each PUMA for monthly rent, household income, and gross rent as a percentage of household income

If you put the median values in a table, and then in a table ordered by HINCP:

PUMA	RNTP	HINCP	GRPIP
Mattapan & Roxbury (PUMA 804)	$1200	$38000	35%
Allston, Brighton & Fenway (PUMA 801)	$2000	$66860	34%
Back Bay, Beacon Hill, Charlestown, East Boston, Central & South End (PUMA 802)	$2000	$70000	32%
Hyde Park, Jamaica Plain, Roslindale & West Roxbury (PUMA 805)	$1800	$71450	28%
Dorchester & South Boston (PUMA 803)	$2000	$86000	27%

Median values for each PUMA for monthly rent, household income, and gross rent as a percentage of household income

- **<u>Correlation between GRPIP and HINCP:</u>**

To explore this correlation, I carried out a Pearson correlation test. Pearson correlation has a set of assumptions:

- Both variables must be continuous (interval-ratio). This is true for HINCP and GRPIP.

- Both variables have approx normal distribution. Though both are probably skewed, I went ahead and made this assumption.
- Relationships are linear. I wasn't sure about this one, so tested it later with a scatter plot.

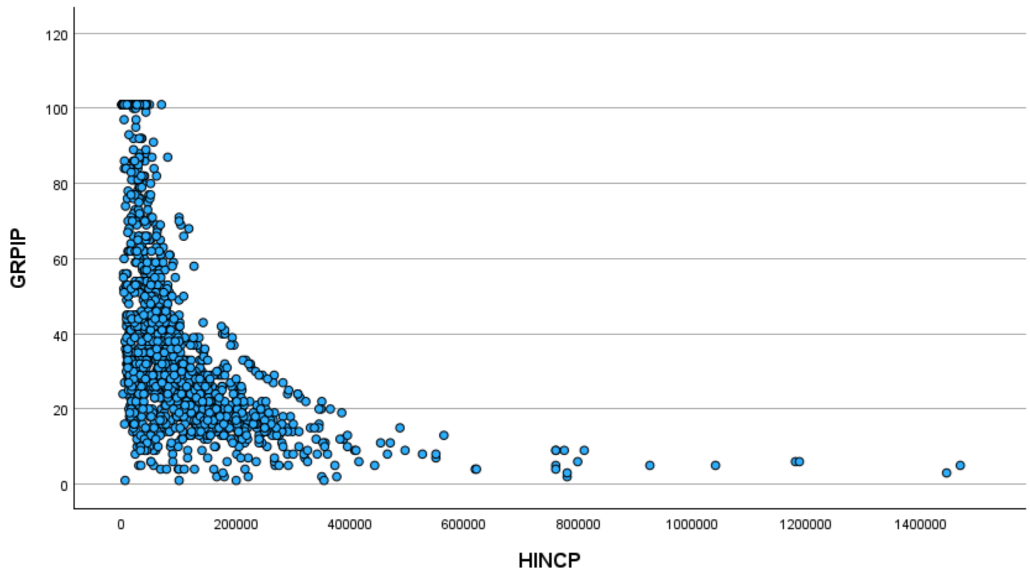
The Pearson correlation coefficient obtained with this data was -.489. The correlation is significant at the 0.01 confidence level. This correlation is moderately strong, and is negative, which I expected.

Correlations

		GRPIP	HINCP
GRPIP	Pearson Correlation	1	-.489**
	Sig. (2-tailed)		<.001
	N	174749	174749
HINCP	Pearson Correlation	-.489**	1
	Sig. (2-tailed)	<.001	
	N	174749	174749

**. Correlation is significant at the 0.01 level (2-tailed).

Given that the correlation coefficient wasn't the strongest (strong correlation strength is considered at the 0.5 mark or more) I wanted to see why and ran a scatter plot of the two variables. It definitely showed a negative relationship, but 1) the curve that seems visible is more of a logarithmic one instead of linear and 2) points populated most of the space below the curve rather than having them all clustered closer to the line. Elaborating on this second point, there seemed to be a point y at each point x past which there were no more values and these formed the contour of a logarithmic curve. It also hints there is some predicting power: if x is a certain income, we can say with some confidence according to this sample that their rent to income ratio will be capped at a certain percentage. For example, if a household is making $20,000 a year, we can say that according to this sample, it's unlikely that their rent burden will be past 40%.



Cases weighted by WGTP

Because what seems to be the best curve follows a logarithmic rather than linear relationship, we can also test the correlation with the Spearman's correlation test, since it does not require linearity. Here are the assumptions for this test:

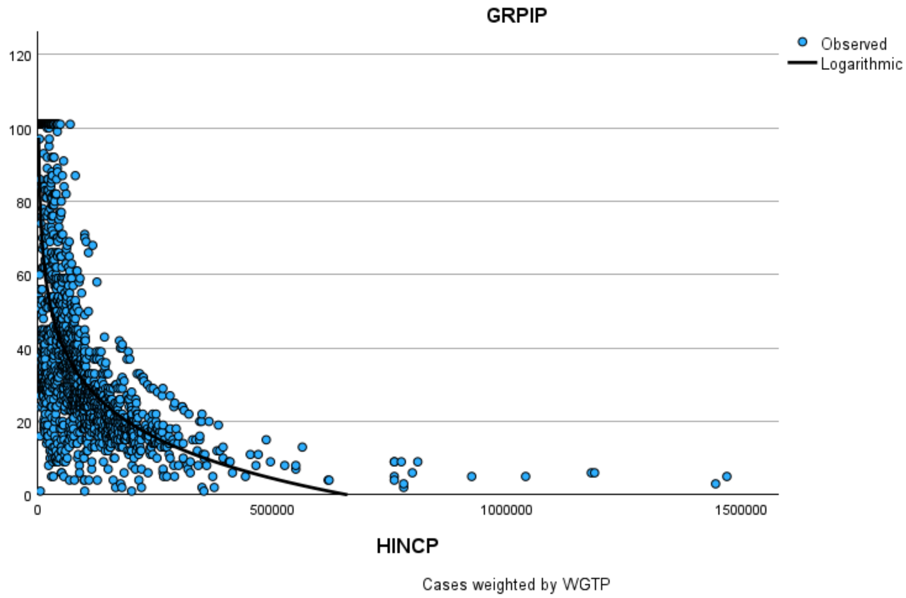
- Variables are ordered (can be continuous)
- Relationship is monotone

A monotone relationship is one in which as x increases, y either increases for every change in x, or decreases for every change in x. Logarithmic curves are monotone so we can assume this.

As may be expected when using a better fit correlation test, the correlation coefficient turned out stronger. The test measured a correlation coefficient of -.701 significant at the 0.01 level. This further shows a relationship between HINCP and GRPIP that is statistically significant.

- **Regression:**

Given the strong correlation, regression can tell us if there is some predictive power in the relationship and if it is significant and valid. I ran a curve estimation test selecting the option of 'logarithmic' for the best-fit curve.



Model Summary and Parameter Estimates

Dependent Variable: GRPIP

	Model Summary					Parameter Estimates	
Equation	R Square	F	df1	df2	Sig.	Constant	b1
Logarithmic	.494	170538.158	1	174747	<.001	216.226	-16.136

The independent variable is HINCP.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
ln(HINCP)	-16.136	.039	-.703	-412.963	<.001
(Constant)	216.226	.426		506.986	<.001

The Model Summary table shows an F statistic of 179538.158 at a significance level of p < .001, meaning that the model as a whole is significant and that there is less than a one chance in a thousand that our results are due to chance. The R square value represents 49.4%. We can interpret this as HINCP explaining 49.4% of the value for GRPIP. The prediction lies in being able to say that, statistically, according to this sample, close to half of a household's rent to income ratio variation in Boston city is determined by the household income. Additionally, the t-statistic of the *lm* variable is

-412.963 so our significance level is beyond 0.001 and we can be 99.9% certain that our coefficient estimate is not due to chance.

Limitations of the Analysis:

In the GRPIP variable there were households with observation values of 0 and 101. Both of these were a catch category for:

- 0: N/A observations, group quarters, owned or being bought/occupied without rent payment/no household income.
- 101: no household income, any ratio where more rent was to be paid than income earned.

Because of the broad options these two variable values for GRPIP had very high frequencies (the 0 value made up 38% of cases) and skewed the distribution of values 1 to 100. However, they tell important cases of rent burden and rent free living. Whether to include or exclude them can be considered for each analysis depending on its purpose. I chose to exclude the 0 values since I was not interested in no answer values or rent situations with no rent payment or group quarter situations. I wanted to include the 101 values because I thought they were still relevant to the discussion of rent burden. Understanding what cases exactly fall into the 101 value would help make a decision as to what to include or omit.

GRPIP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	109312	38.5	38.5	38.5
	1	334	.1	.1	38.6
	2	547	.2	.2	38.8
	3	252	.1	.1	38.9
	97	147	.1	.1	91.8
	99	62	.0	.0	91.9
	100	234	.1	.1	91.9
	101	22904	8.1	8.1	100.0
	Total	284061	100.0	100.0	

high frequencies

The ACS also makes available PUMS 1-year or 5-year estimates. I used the 2022 1-year estimate, but the 5-year estimate could have captured different information useful to a time-series analysis. Lastly, the nature of the AMS as a survey with weighted cases has sampling bias and other errors that can cause data to not represent the populations fully or to hide a bias that we may not realize. Regardless, the data in this sample is statistically significant and hopefully accurate and if these are true we can still draw some useful conclusions.

Conclusions:

To many Bostonian renters and households, rent burden is evident. And different citywide and statewide measures and legislations could drastically help those most affected by it (e.g. rent control, subsidized housing, construction of affordable housing, support for multifamily housing). Data shows that rent burden affects low income groups the most. According to this sample, household income explains nearly half (.494) of rent-to-income ratio variation in the city of Boston. Housing cost burden is a phenomenon that comes about through a variety of realities ranging from the demographic, systemic, geographical, and economical, and this analysis comes with a series of sampling bias, data choice, and survey error. On the demographic side at least, this sample can point to, through correlative and predictive statistics, what many people already can tell you: lower income households have a much harder time escaping rent burden. In conclusion, it is important to understand the drivers of housing cost burden in conjunction with effective policies catered to the realities of renters most straddling with its effects.