Project Report

Housing Cost and Availability in Boston

Authors:

Muhammad Mehdi Ali

Bharath Raam Vaduvoor Srinivasan

Alexander Thrush

Summary:

This project investigates housing cost, availability, and related factors in Boston, Springfield and Worcester. The questions addressed by this report are: What factors contribute to a lack of housing and rising housing costs in Boston[i]? Is there a shortage of housing in Boston, and if so, are residents changing their behavior as a result of it? Additionally, are trends in Boston mirrored in Springfield and Worcester?

In recent decades the cost of housing has increased dramatically in many cities, especially in already densely populated areas such as Boston, and other coastal cities[ii]. To investigate the problems of housing availability and cost, we used datasets from the American Community Survey[iii] (ACS), which is conducted by the U.S. Census Bureau. Specifically, we analyzed data for Boston, Springfield, and Worcester, for the years 2005 to 2021 (excluding 2020 due to lack of reliable data). The most relevant datasets for answering our questions are called "Selected Housing Characteristics"[iv] and "Age and Sex"[v]. Relevant data within these datasets are population, number of housing units, vacancy rates for rental units, median rents, and rents in relation to household income. While the ACS does also track data about costs to homeowners, we focused on rent prices for this project as renters make up a significant portion of the population in cities, and renting is typically more accessible as a downpayment is not required, nor is approval for financing.

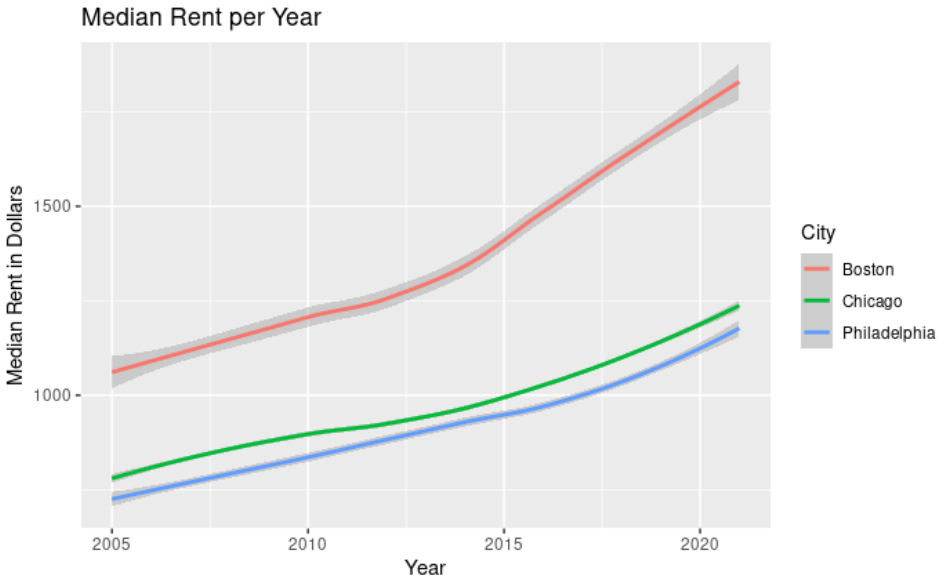
Before we were able to gather meaningful insights from the data, we identified the cleanest source for it, which is called tidycensus. We then pulled data for Massachusetts cities, cleaned and tidied it, and combined it together. With the data processed, we created several visualizations that show changes in the number of occupants in each housing unit, median rent, median rent as a percentage of household income, and number of occupants per room in housing units. Some key results that we gathered from our visualizations are that median rents in all three cities have increased significantly during the time period in which data is available. In contrast, median rent as a percentage of income fluctuated noticeably, but ended up around the same or lower in 2021 than in 2005. Interestingly, the number of occupants per housing unit in Worcester increased, in Springfield increased less, and in Boston increased greatly before returning roughly to where it started.

Methods:

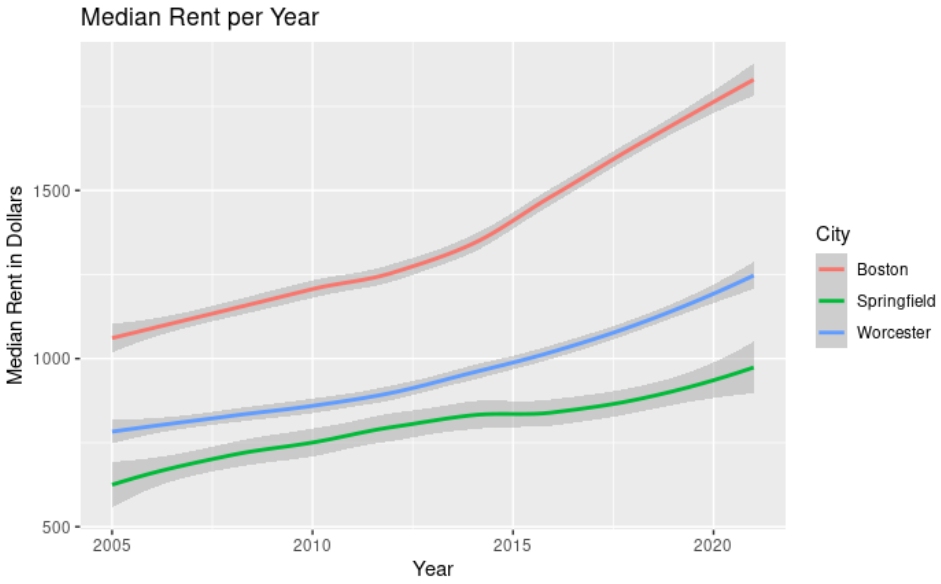
Methods for our project can be found in the relevant Project Methods document.

Results:

Before focusing in on the three Massachusetts cities that are the subject of this project, we thought it would be useful to compare Boston to some other cities in the US, Chicago Illinois and Philadelphia Pennsylvania. We chose these cities as they all share similar qualities of having been well established by the early 20th century and having had populations that peaked in the mid-twentieth century. As can be seen in the below visual, in 2005 median rents are already about $400 higher than in Philadelphia and $300 higher than in Chicago, by 2021 this gap widens by to about $600 and $550 respectively.

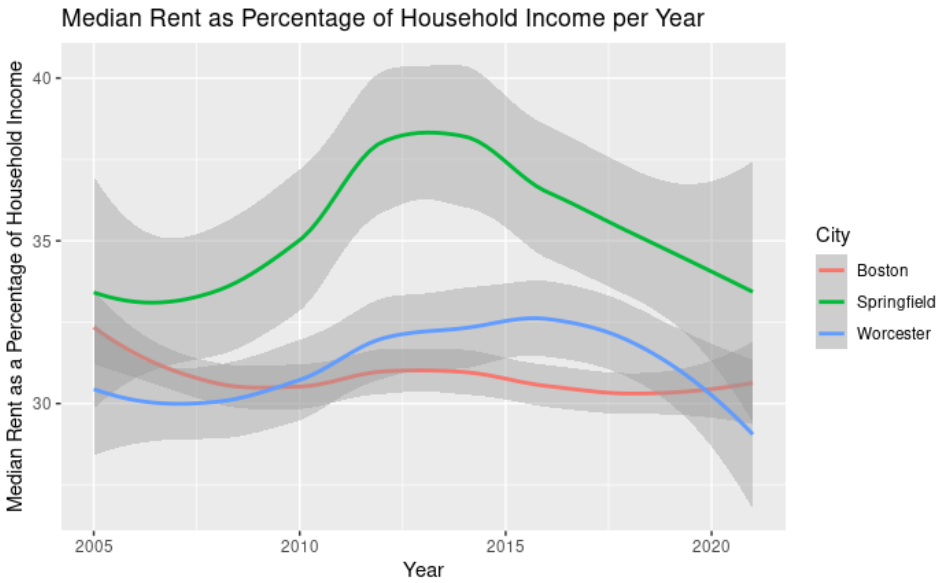


Turning our eye to the cities that are the focus of this project, we see that median rent has increased year over year for the time period for which data is available. The rate of increase in median rent increased for Boston and Worcester grew noticeably in 2013-2014. Increases in median rents in Springfield plateaued in 2015, and then began increasing again the following year.

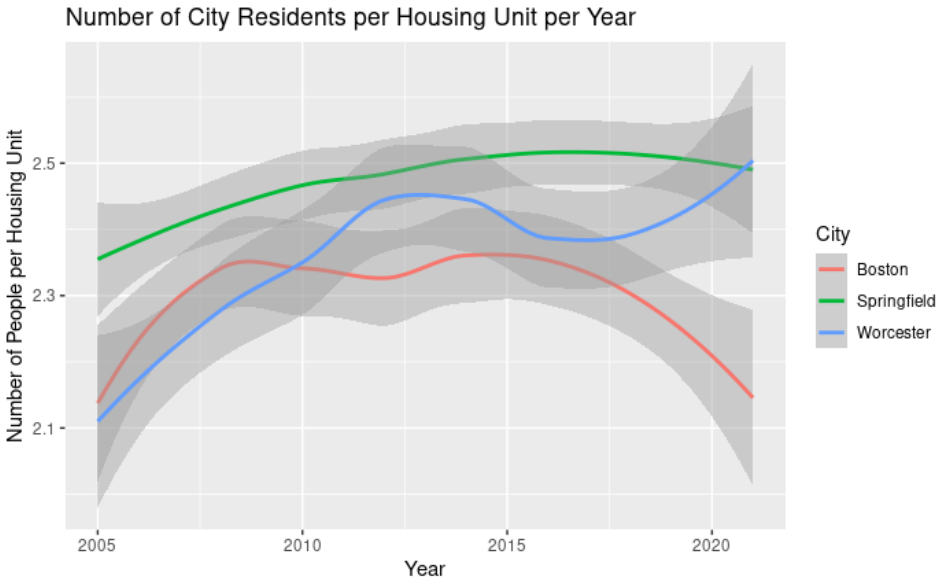


While median rents in the three cities have increased consistently, median rents as a percentage of household income (MRPHI) changed much more erratically over the same time period. In Boston we see a MRPHI of 32.5% in 2005 with some movement up and down over the next 16 years, ending up 31% in 2021. An overall decrease in MRPHI is also seen in Worcester, starting at 30.5% and ending at 29%. Springfield had the greatest fluctuation in MRPHI of the three, starting at 33.5%, peaking at 38%, and ending at 33.5%. Based off the changes in MRPHI and median rent, it appears that while medians were reliably increasing from 2005 to 2021, household income was changing significantly over that same time period. Additionally, we can see that except for Worcester in 2021, MRPHI is at or above 30% for each

city. The significance of this is that it means housing costs break a common rule of thumb[vi] (that they shouldn't exceed 30% of one's gross income) for many residents of these cities.

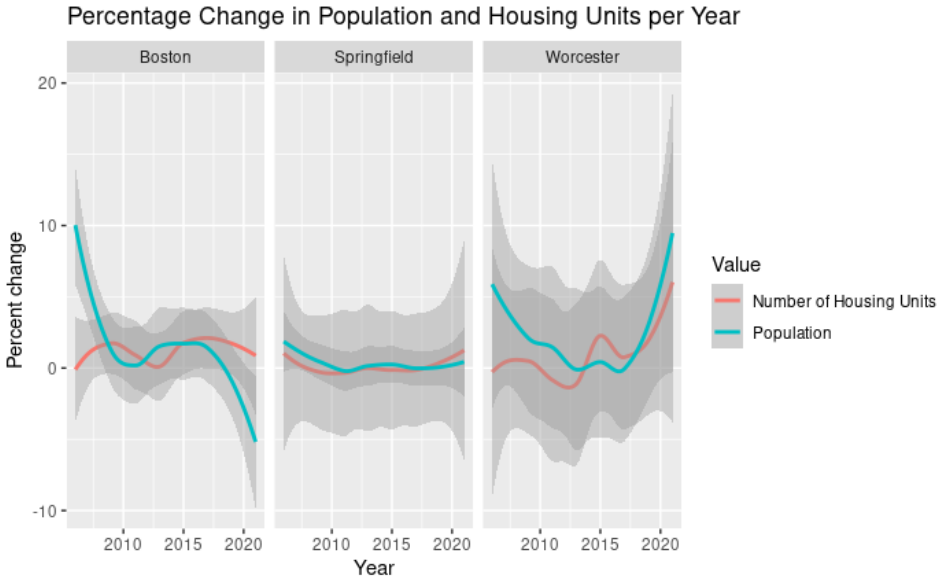


Another possible reason for changes in MRPHI (other than fluctuations in the individual incomes of members of households) could be changes in the number of people per housing unit (NPPHU). This does indeed appear to be a factor for Boston, as the number of people per housing unit changes in the opposite direction to changes in MRPHI. This may indicate that decreased MRPHI may be due simply to there being more occupants in a given housing unit (and therefore additional household income) and vice versa. A similar trend can be seen for Worcester, but not for Springfield. The change in Springfield's NPPHU is positive for the most part, but the rate of increase decreases until change in NPPHU becomes negative in 2015-2016. This is in contrast to the dramatic changes in Springfield's MRPHI.

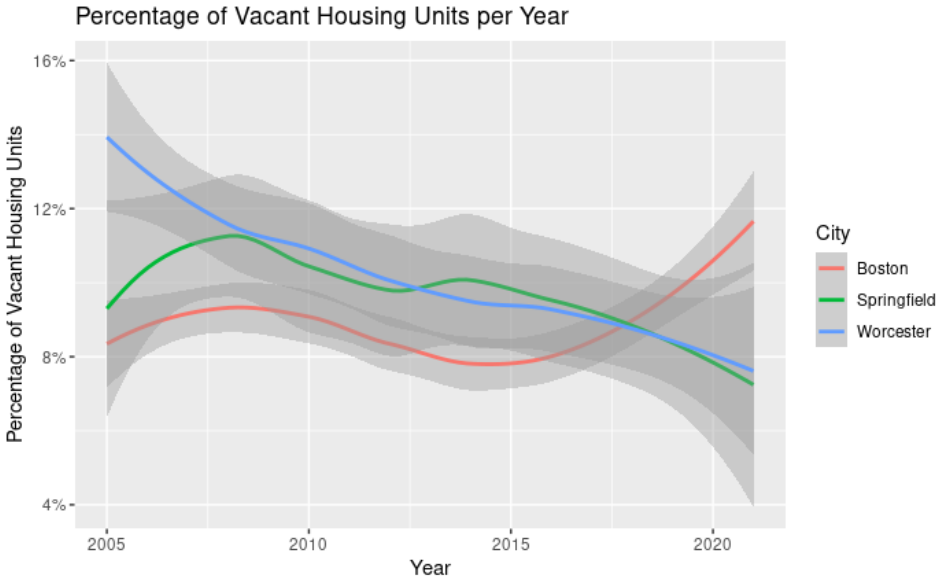


The below visualization illustrates changes in both population and number of housing units. We can see that trends in Boston are relatively stable, with percentage change in number of housing units and number

of residents not passing 2.5%. While changes in Springfield and Worcester are more extreme we see an overall positive change in population and number of housing units.



Finally we wanted to see vacancy rates for housing generally in our focus cities. Below we see that vacancy rates have been decreasing in Springfield and Worcester, though they have increased significantly in Boston in recent years. Given the increasing population of the city of Boston, and the number of city residents per housing unit increasing only very slightly over the 16-year period for which data is available, this is a surprising trend.



Discussion of Results:

The principle, and to many unsurprising, result of this project is that for these three major Massachusetts cities, median rent has been increasing from 2005 to 2021 and it appears this trend is likely to continue. In

relation to this result, we do also see that renters appear to be willing to cohabitate/live with roommates in greater numbers in order to mitigate the impact of ever-increasing rents.

Another important result of this report is that housing is unaffordable to many residents of these cities, as median rents generally exceed 30% of household income. This has many negative impacts for renters, a very significant one being that it makes saving for a home more difficult, solidifying class/social structure in these cities.

Regarding whether or not there is a shortage of housing, there are some mixed signals in our results. While we do see that rents have been increasing for all three cities, in Boston we see that vacancy rates have also been increasing in the past several years, though they've been decreasing in both Worcester and Springfield. Being somewhat surprised by this combination of trends, we looked for an alternate source of information about housing vacancy rates, the Saint Louis Federal Reserve. Their data[vii] shows that rental vacancy rates in Massachusetts were 5.7% in 2005 and now stand at 2.8%. The Saint Louis Fed cites the Census Bureau in providing this information, so it's unclear what the cause of this discrepancy is.

Several groups of people may benefit from the information found in this report, such as current and future residents of Boston, Springfield and Worcester, as well as policy makers. People who will live in or already do live in one of these cities may be able to more easily choose to stay where they are or move somewhere else based on projected changes in housing cost and availability. By looking at this report city government officials may better understand the high cost of housing and how that impacts residents. With this information they would be better equipped to make policy changes that allow housing to be built more quickly by both private and public organizations.

This project could be improved by also investigating the number of residential building permits approved or denied. While we have looked at changes in housing supply in this report, analyzing the rate of approval of residential building permits would provide an insight into whether the local reviews process is hindering growth in the number of housing units.

We have used this analysis to predict the median_rent_allstate_cities_estimate using Random Forest Regressor, and with the columns that affect the rent prices. This would give us the future scope to predict how the rent may increase or decrease given the details.

Contributions:

Muhammad Mehdi Ali

- Data visualizations

Bharath Raam Vaduvoor Srinivasan

- Data pre-processing and tidying
- Data modeling

Alexander Thrush

- Initial exploration of census data to gauge suitability for project
- Investigation of alternative methods to access census data (tidycensus)

References can be found in the endnote section of this document.

Appendix:

Predictive Modeling Codes and Result:

```
findRMSE <- function(formula, data) {
  set.seed(123)
  index <- createDataPartition(data$median_rent_allstate_cities_estimate, p = 0.5, list = FALSE)
  train_data <- data[index, ]
  test_data <- data[-index, ]

  data_cv <- tibble(train = list(train_data), test = list(test_data))

  data_cv %>%
    mutate(fit = map(train,
                     ~ randomForest(formula, data = .)),
         rmse = map2_dbl(fit, test, ~ sqrt(mean((.y$median_rent_allstate_cities_estimate - predict(.x,
    summarize(cv_rmse = mean(rmse)) %>%
    pull(cv_rmse)
}


findRMSE(median_rent_allstate_cities_estimate ~ GEOID, data)
```

```
data<-select(data, c('GEOID','city_names','year','owner_occupied_units_allstate_cities_estimate','rent

set.seed(123)
train_index <- sample(1:nrow(data), round(0.7 * nrow(data)), replace = FALSE)
train_data <- data[train_index, ]
test_data <- data[-train_index, ]

rf_model <- randomForest(median_rent_allstate_cities_estimate ~ ., data = train_data, ntree = 100, mtr

rf_predictions <- predict(rf_model, newdata = test_data)

rf_mse <- sqrt(mean((rf_predictions - test_data$median_rent_allstate_cities_estimate)^2))
print(paste0("Random Forest MSE: ", rf_mse))
```

```
## [1] "Random Forest MSE: 130.435460524528"
```

actualRent <dbl>	predictedRent <dbl>
1075	1169.3512
794	916.3215
1265	1342.3543
659	720.6252
890	895.0738
850	850.7457
883	886.8297
635	669.1023
1258	1240.4323
1035	1063.3207
1223	1208.9065
603	657.1503
780	769.4112

[i] Zillow Home Value Index for Boston MA - https://www.zillow.com/home-values/44269/boston-ma/

[ii] https://thehill.com/changing-america/sustainability/infrastructure/3712231-housing-costs-are-highest-in-these-us-cities-report/

[iii] About the American Community Survey - https://www.census.gov/programs-surveys/acs/about.html

[iv] SELECTED HOUSING CHARACTERISTICS
-
https://data.census.gov/table?q=rent+in+Boston+city,+Massachusetts&tid=ACSDP5Y2021.DP04

[v] AGE AND SEX - https://data.census.gov/table?q=rent+in+Boston+city,+Massachusetts&tid=ACSST5Y2021.S0101

[vi] https://www.nerdwallet.com/article/finance/money/how-much-should-i-spend-on-rent

[vii] Massachusetts Rental Vacancy Rate - https://fred.stlouisfed.org/series/MARVAC