Lab 06 – MATH 240 – Computational Statistics

Harrison Wolfe
Colgate University
Math Department
hwolfe@colgate.edu

Abstract

The goal of this lab is to figure out how costs of living vary for home owners and renters in Connecticut. This data could be helpful to give people more infomred desicions when looking into housing in the Southern Connecticut area.

Keywords: graphs; cleaning data; us census

1 Introduction

This paper is intended to describe the process in which we analyze data regarding costs and incomes about housing in Southern CT. While there are only a few measurements on the topic like Median Income for Home Owners, Median Income for Renters, and Percentage of Income spent on Housing for each category, these categories are very important when analyzing housing in CT. The goal of this lab is to be able to summarize our findings into readable graphs which display conclusions regarding the two options.

2 Methods

2.1 Data Collection

In this lab we used data from the US Census in the southern CT area. We were able to obtain and organize this data through the `tigris` and `tidycensus` packages (Walker, 2024) (Walker and Herman, 2025). The US census contains numerous statistics about each area as they try and use this to make assumptions about areas as small as towns and as large as the United States. We were able to arragne the data in a way to be made into charts using the `tidyverse` package (Wickham et al., 2019).

2.2 Data Analysis

With this data we could find descriptive information about each "tract" (small areas within CT) and use that to summarize the region as a whole. We found information like the Percentage of Income used on Home Ownership or Gross Rent and Median Rent or Home Ownership costs. In 2022 the census changed the region name for Southern CT to Southern CT when before it was New Haven. We had to combine these data sets into one spreadsheet. We also had to rename data from 2013-2014 in this spreadsheet because they changed the name of some categoreis in 2015. We used these various statistics collected for the years 2013-2023 and then were able to graph the trend that was found within many of them.

3 Results

From this lab we were able to create three different types of graphs. One of the graphs displayed two maps of the areas in question and showed a color depending on the Median Gross Rent as a Percentage of Household of Income and the other map showed Median Owner Costs as a Percentage of Household Income. This was extremely helpful when determining where exactly people were spending a large portion of their income on housing and also gave a nice visual when looking at the difference between how it is affecting the people renting and the people who own their homes in general. Another graph we were able to create was a line plot showing the trend of various statistics collected. We were able to show the trends of Median Gross Rent ($), Median Owner Costs ($), Median Income of Owners ($), Median Income of Renters($), Median Rent as a Percentage of Household Income (%), and Median Owner Costs as a Percentage of Household Income (%) from 2013 to 2023. With these various line graphs we were able to see the trend of each of these statistics to determine not just the differences between owning and renting currently but to also compare how it has changed overtime. The final graph was a density plot of Median Rent($) for every year since 2013. This plot showed the distribution of the sample with every "tract" and showed how it moved overtime. All of these graphs can be found in the Appendix listed as Figures 1, 2, and 3. All of these graphs were made using `ggplot2`, organized with `gridExtra`, and the density plot was made using `ggridges` (Wickham, 2016) (Auguie, 2017) (Wilke, 2024).

4 Discussion

4.1 Discussion of Figure 1

Our initial task from this lab was to compare homeowners to renters in thier costs, incomes, and what percent of their income they use on their choice of housing. In Figure 1 we can see that homeowners all over CT are spending a much smaller percentage of their income on housing. Because we can see a lot more purple, blue and a little bit of red on this homeowner chart we know that the majority of people are spending 30 percent or less of their income on housing. This is completely different for the map displaying the renters in CT. We can see a lot more red and black showing a much higher percentage of poeple spending more than 30 percent of their income on gross rent. However, this makes sense when we look at Figure 2 when we see that people's income who own a home is higher on average than people who rent.

4.2 Discussion of Figure 2

Another important trend to analyze from Figure 2 is the spike in costs in 2020. While we do see a fluctuation for homeownership and static growth for rent costs over the years the costs went up a lot more in 2020 as we can see through the steeper line. This allows us to conclude that the COVID pandemic had a great impact on housing costs making them go up considerably. To sum this chart up we can see that every year homeowners are making more then renters, spending more on their cost of living while also spending a smaller percentage of their income. An interesting trend from the graph is when we look at costs as a percentage of income: we see that it is going down overtime. This means that people's incomes are growing more than their costs of housing which should make it more affordable for everyone.

4.3 Discussion of Figure 3

This figure is describing the prices of rent over time. As we can see it is increasing relatively steadily until 2020 where the variation starts to increase a lot year by year till 2023. The variation did not just increase but the mean increased as well showing that rent prices were increasing as well pretty dramatically when comparing it to the 7 years before that. As discussed before this is likely due to the pandemic showing that it increased rent prices and housing in general. However, this was combated as we saw before in Figure 2 with the increase in income as well. 2020 was an excellent year for the stock market which might have helped increase income for some of these people. That is pretty evident for homeowners as their income shot up quite a bit. Another thing to keep in mind for this time period was that the unemployment rate went up because of the pandemic which means that incomes for the employed may have been affected in the trend that is dispalyed in the chart. Overall, from this chart we can see that the costs of rent have grown and also the variation in the costs of rents have grown meaning that more people are paying more money for their rent.

References

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

Walker, K. (2024). *tigris: Load Census TIGER/Line Shapefiles*. R package version 2.1.

Walker, K. and Herman, M. (2025). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. R package version 1.7.1.

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

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wilke, C. O. (2024). *ggridges: Ridgeline Plots in 'ggplot2'*. R package version 0.5.6.

5 Appendix

These charts, and maps could not fit into the template above and for that reason I have placed them here. Each of them were made using `ggplot2` and Figure 2 was organized using `gridExtra` (Wickham, 2016) (Auguie, 2017).

	...1	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12	...13	...14	...15	...16	...17	...18	...19
1	Letters	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	Numbers	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

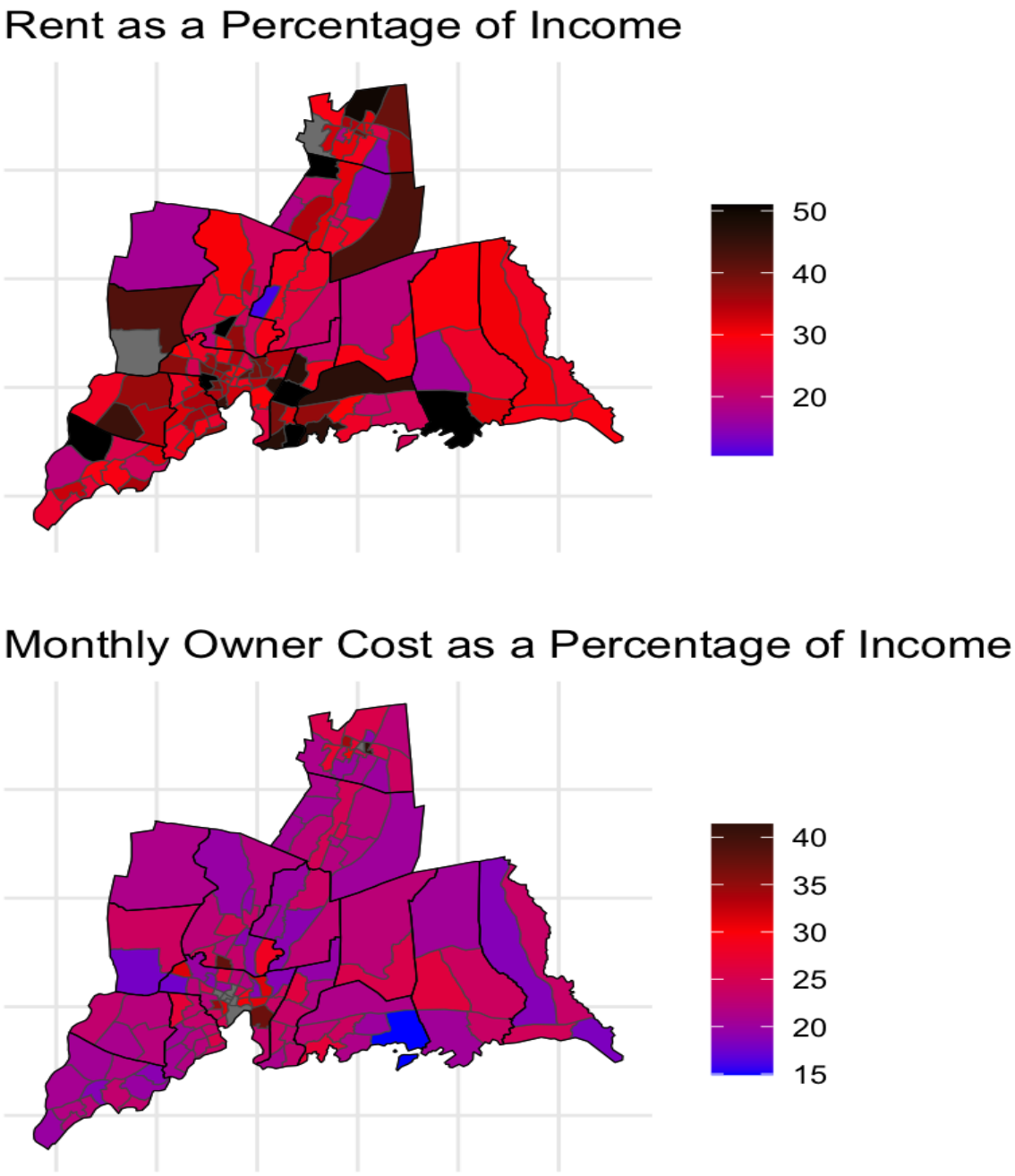

Figure 1: Median Rent and Owner Costs as a Percentage of Income on a Map

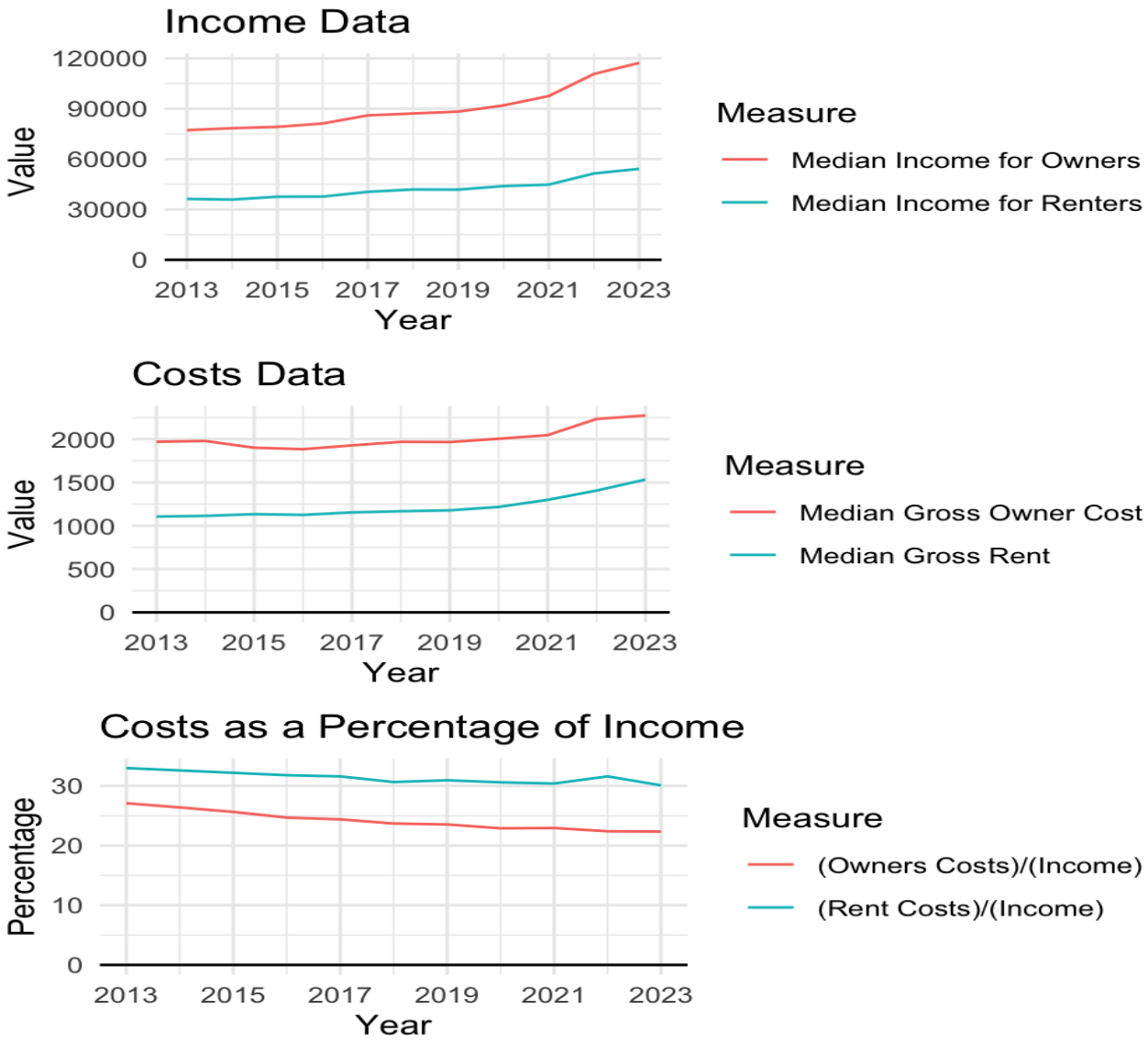

Figure 2: Income, Rent/Homeownership, and Rent/Costs of Homeownership as a Percentage of Income over Time

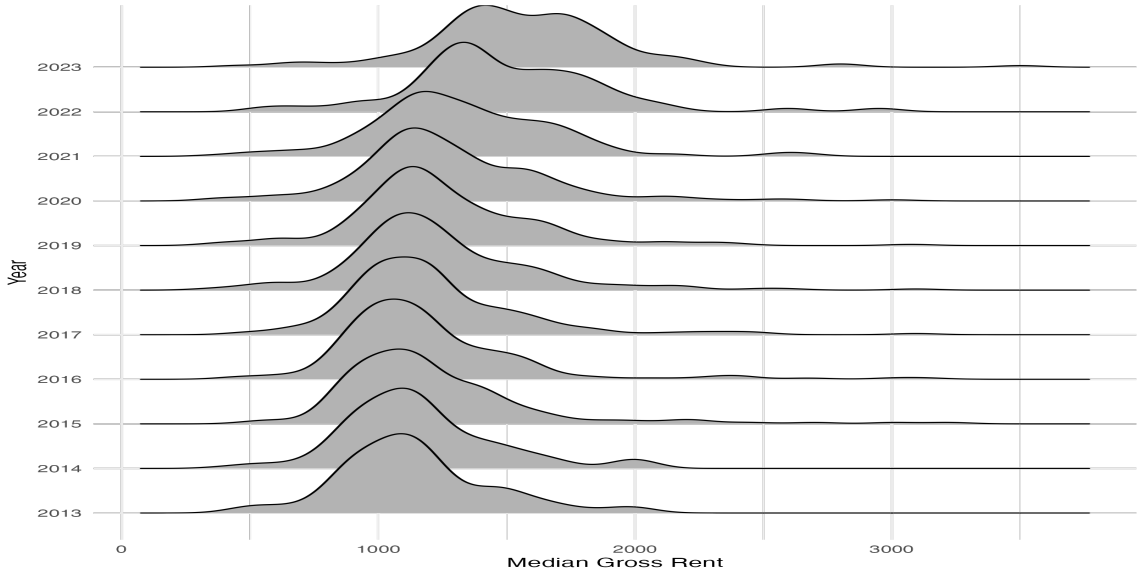

Figure 3: Density Plot of Median Rent over Time