LIS 572 Introduction to Data Science

Jessica Fleming, Patty Titus, and Maura Wilson

5/5/24

Final Project Proposal

#### **Title:**

Our group is still considering a title, and the final choice will likely be influenced by what the data ultimately does or does not support. Here are a couple of current ideas:

#### Privilege and Poverty: The Inequality of Short-term Rental Riches and Affordable Housing

* Privilege and Poverty: An Examination of Airbnb’s Influence on Homelessness

# Introduction

Our project plan is to gather data on the numbers of unhoused people, average income, average rent, number of Airbnb properties and total number of properties in New York City, Denver, Los Angeles, and Seattle.

This information will be presented through a blog post and graphs plotting the data over time- ideally 2008-present if we are able to secure that data. We anticipate the data will show that the rise in the number of Airbnb properties may coincide with a rise in rent prices. We are then interested in seeing how the change in rent prices corresponds to changes in average income and number of unhoused people. We will be collecting data from a variety of datasets and combining them into a unique dataset. We think this is an interesting area to research because of the many false narratives that surround why people end up unhoused which mainly focus on drug use, mental illness or other “personal failings” instead of focusing on the systems that contribute to the housing crisis.

# Related Work

As far back as the 1980s, research has been done around the variety of causes for homelessness in American cities. Some studies focused on the individuals experiencing homelessness and what factors in their lives are the cause of their being unhoused. Other studies, however, take a more community-level approach, looking at factors such as decreases in available and affordable housing as a cause. *Homelessness Is a Housing Problem* posits that “Regional variation in rates of homelessness can be explained by the costs and availability of housing” [(Colburn & Page Aldern, 2022, p.10)](https://www.zotero.org/google-docs/?OgG8jQ). This stands in contrast to other explanations that emphasize individual characteristics and choices as the driver for rise in unhoused people such as mental health issues, drug addiction, or personal choice.

In her February, 2024, article, “The Role of Airbnbs in America’s Housing Crisis”, Tatum Joerndt explains the effect that property ownership solely for the purposes of Airbnb rentals is having on the availability of affordable housing to buy and rent.

People own more than one home and rent them out for profit, while others struggle to afford their first home at all. Being a renter through Airbnb allows the owner to make more money off these short-term stays than they would renting the property out as a home. [(Joerndt, 2024)](https://www.zotero.org/google-docs/?Ymy4IV)

A group of researchers from the U.S. Department of Veteran Affairs looked at the possible causes for homelessness beyond the focus on the individual and found “additional evidence that homelessness has its roots in housing market dynamics, and particularly in the difficulty in obtaining affordable housing.” [(Byrne et al., 2013)](https://www.zotero.org/google-docs/?bGekVK)

To further explore these issue we will gather data points over a period of time for four different cities and use R language to generate visualizations that help demonstrate any relationships between the number of unhoused people in those cities, average income, average rent, and the number of Airbnbs in each city over a number of years. This work will extend current research on this topic and hopefully create visualizations that will be useful beyond the limits of this course.

# The Dataset

The dataset will be created by combining datasets found for each of the selected cities. We are hoping to find information for the number of unhoused people, the number of Airbnb properties, the average cost of rent, the average income, and the number of residential properties in each city. As our project involves creating our own dataset out of pieces of the datasets of others, it will fall between the two established assignment tracks. We will be combining data from CSV files into one large dataset from which we can create multiple dataframes for each of our chosen cities. For example, Inside Airbnb is an open data project that has spent years gathering Airbnb information scraped from the Airbnb site. Current information on numbers of Airbnbs are provided for each of our cities, but the site notes it has historical archives available also [(*Inside Airbnb*, n.d.)](https://www.zotero.org/google-docs/?GxjMGD). We have written to their data request address and are hopeful they may be able to provide us with numbers of properties over time in our selected cities.

InsideAirbnb data scraped some identifying information from Airbnb into its dataset, like names of property reviewers. We neither need nor want to use identifying information nor reproduce into our own dataset, so we will take care not to pull any of that information into datasets we create. Also, it is worth considering whether it is unethical for InsideAirbnb to have scraped that information and, if so, is using any of their data tantamount to supporting those ethically questionable data gathering practices? To present our data in an ethical manner, we will provide information regarding the provenance or presentation of that data which the end-user or reader of our work may need to know in order to understand what our data says and what its limitations are.

While other issues will surely present themselves as we go along, we anticipate the following potential issues:

* Not being able to find the data across the same year span for all datasets. In this case we would have to get information about that variable from more than one source and combine them in our dataset. If differing data collection or presentation methods were used between sources that could mean that one year’s number is more or less accurate than another for the same variable.
* Point in Time (PIT) count data for unhoused people is famously imperfect, but it is widely available and fairly easy to ascertain. We will outline the potential issues with those data practices.
* Some data may be attainable in file formats other than .csv which presents a technical challenge of either file conversion or manual implementation.
* When combining data from multiple sources, there may be inconsistencies in data collection methods for data with the same name.
* Total number of residential housing in each city each year may also be difficult to access. We are looking but have not yet found reputable sources for this information. We may have to remove this variable from our project if it is too cumbersome or beyond our skill set.

# Analysis/Summary

We are planning to create visualizations for each of the four cities chosen. These visualizations will plot the five variables across, ideally, 2000-2023. Airbnb was launched in August of 2008 [(Schonfeld, 2008)](https://www.zotero.org/google-docs/?cvVUq6) and we would like to see data from before their launch through their growth in popularity to our current time. If this data is unavailable, we will determine what time frame we have available to us. It is possible this data will be available for some of our cities and not others, as the collection practices and agencies likely differ. We have chosen to look at the total number of Airbnb properties, total number of properties, average income, average rent cost, and number of unhoused people for each city. What type of plot we use to showcase these remains to be determined, but likely will be a line graph as it is generally good for showing changes in data over time.

# Why Is Your Work Useful?

Public libraries are, at their core, a public service benefitting their community. In order for a library to serve its community, the library and those people that make up the organization must understand how that community functions. It is a public tragedy that our neighbors may be unhoused in cities which simultaneously grant great financial privileges and opportunity to others. In order to best serve our community we must endeavor to understand this issue and its true root causes. This helps us avoid easy but ultimately destructive narratives about the drivers of economic inequality in our communities, and take responsibility for helping address structural issues where we can.

Libraries serve as a safe space for us and for our unhoused neighbors. In addition to furthering the current available research into this issue, our work will further the effort to discredit destructive narratives which lay the blame for being unhoused at the feet of unhoused people themselves- the very population least able to do anything about the structural issues that contributed to the situation in the first place.

# Challenges & Limitations

We currently envision the following challenges:

1. Data from reliable or usable sources may be difficult to find. In the event we are not confident about a data source but find no reasonable alternative, do we use it anyway?
2. How to achieve transparency regarding data source issues? It is important to us that the reader understands the limitations of our data.
3. It is possible that we may find our analysis doesn’t lead to any solid relationship between the number of unhoused people and the number of Airbnbs in a city. This might feel like failure but is not- this just means other issues are driving the problem. We will need to account for this when explaining our data in our written product.
4. We may find that we must aggregate data in various ways to use it. For example, some housing availability or rent information may be presented subdivided into categories based on the size of the property. Our project ultimately needs a single number for these variables for each city for each year. We will have to decide a principled method for doing this and apply it as equally as possible across cities and years.

Reference List:

[Byrne, T., Munley, E. A., Fargo, J. D., Montgomery, A. E., & Culhane, D. P. (2013). New Perspectives on Community-Level Determinants of Homelessness. *Journal of Urban Affairs*, *35*(5), 607–625. https://doi.org/10.1111/j.1467-9906.2012.00643.x](https://www.zotero.org/google-docs/?nfs1bQ)

[Colburn, G., & Page Aldern, C. (2022). *Homelessness Is a Housing Problem* (1st ed.). University of California Press.](https://www.zotero.org/google-docs/?nfs1bQ)

[*Inside Airbnb*. (n.d.). Retrieved May 4, 2024, from https://insideairbnb.com/get-the-data/](https://www.zotero.org/google-docs/?nfs1bQ)

[Joerndt, T. (2024). The Role of Airbnbs in America’s Housing Crisis. *LMU Law Review*. https://lmulawreview.scholasticahq.com/post/2332-the-role-of-airbnbs-in-america-s-housing-crisis.](https://www.zotero.org/google-docs/?nfs1bQ)

[Schonfeld, E. (2008, August 11). AirBed And Breakfast Takes Pad Crashing To A Whole New Level. *TechCrunch*. https://techcrunch.com/2008/08/11/airbed-and-breakfast-takes-pad-crashing-to-a-whole-new-level/](https://www.zotero.org/google-docs/?nfs1bQ)