INST 447, Section: 0202
Project Report
11/23/2025
Group Members: Jincy Njenga, Josiah Altschuler, Kendall Hinz, Lily Oakes, Vincent Park

Related Variables

- Washington D.C. housing market 2024:
    - Complete address
    - Property type
    - Listing price
    - Amenities
- Inside Airbnb:
    - Amenities
    - Price
    - Latitude
    - Longitude
    - Availability_365
- DC Population by ZIP Code
    - Population

Potential Relationships

- Airbnb Density & Housing Price
    - We believed there to be a possible positive correlation between the concentration of Airbnbs and the price of housing. This turned out to be incorrect, which we discuss below.
- Airbnb Price & Population
    - We believed there to be a possible positive correlation between the price of Airbnbs and population count. In other words, the higher the population density, the higher the price of an Airbnb. This was also incorrect, and we will explain why below.
- Airbnb Density vs. Number of Zillow Listings Per ZIP Code
    - We believed there to be a possible negative correlation between Airbnb density and the number of home listings. This was also incorrect, as with the assumptions above. We will explore the reasons for this in the graphs below.

Data Cleanup

- Used an API to convert Airbnb listing coordinates addresses.
- Merged the Zillow dataset with the Inside Airbnb dataset
- Cleaned up the dataset by removing null listings (ones with zero bathrooms and/or zero rooms, etc.)
- Merged the DC Population By Zip Code Dataset

Key Characteristics

Zillow Dataset

- The average house price in DC is $1,303,691
- The median house price in DC is $759,000
- The most expensive zip code is 20007, 20015, 20016,20008, and 20036, with median housing costs being $2,000,000.0, $1,697,499.5, $1,400,000.0, $1,060,000.0, and $872,500.0respectively.
- Minimum house cost is $64,999 and maximum house price is $29,950,000
- Most homes have 2 beds and 1 bath

Airbnb Dataset

- The most expensive zip codes in DC on Airbnb are median house costs on Airbnb are 20420, 20006, 20004, 20712, and 20005 with median rental prices at $416, $351, $280, $273, and $260, respectively.
- The average Airbnb cost is $387.8174
- The median Airbnb cost is $134
- The minimum rental price is $10 and the maximum rental price is $5000

DC Population By Zip Code

- The population of each zip code in Washington DC, ranked from highest to lowest.

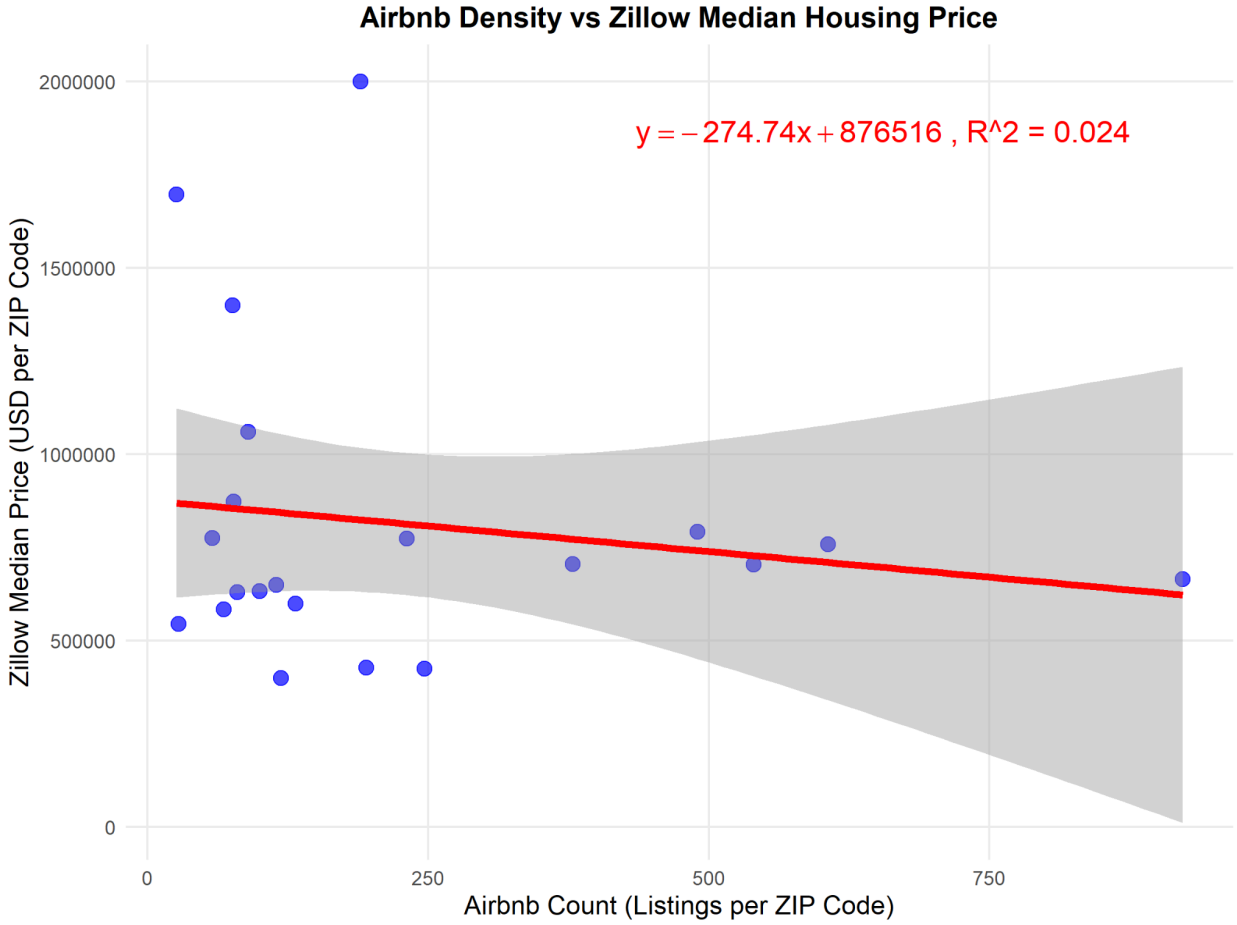
Graphs

Note from Vincent: Based on the 1st Deliverable research questions these are the 3 listings. Need 2 graphs for each research question. **Need to do database cleaning first before any graphing, Note: research question 3 been changed slightly**

1. Airbnb Concentration vs. Housing Price (What is the correlation between Airbnb concentration in a neighborhood and the housing costs in that neighborhood?)
2. Airbnb Price vs. Population (What is the relationship between Airbnb rental costs and the population in a given neighborhood?)
3. Airbnb Density vs. Availability of Housing (Do areas with more Airbnb rental properties have higher or lower housing availability, as measured by the number of listings?)

Question 1: Airbnb Density vs. Housing Price

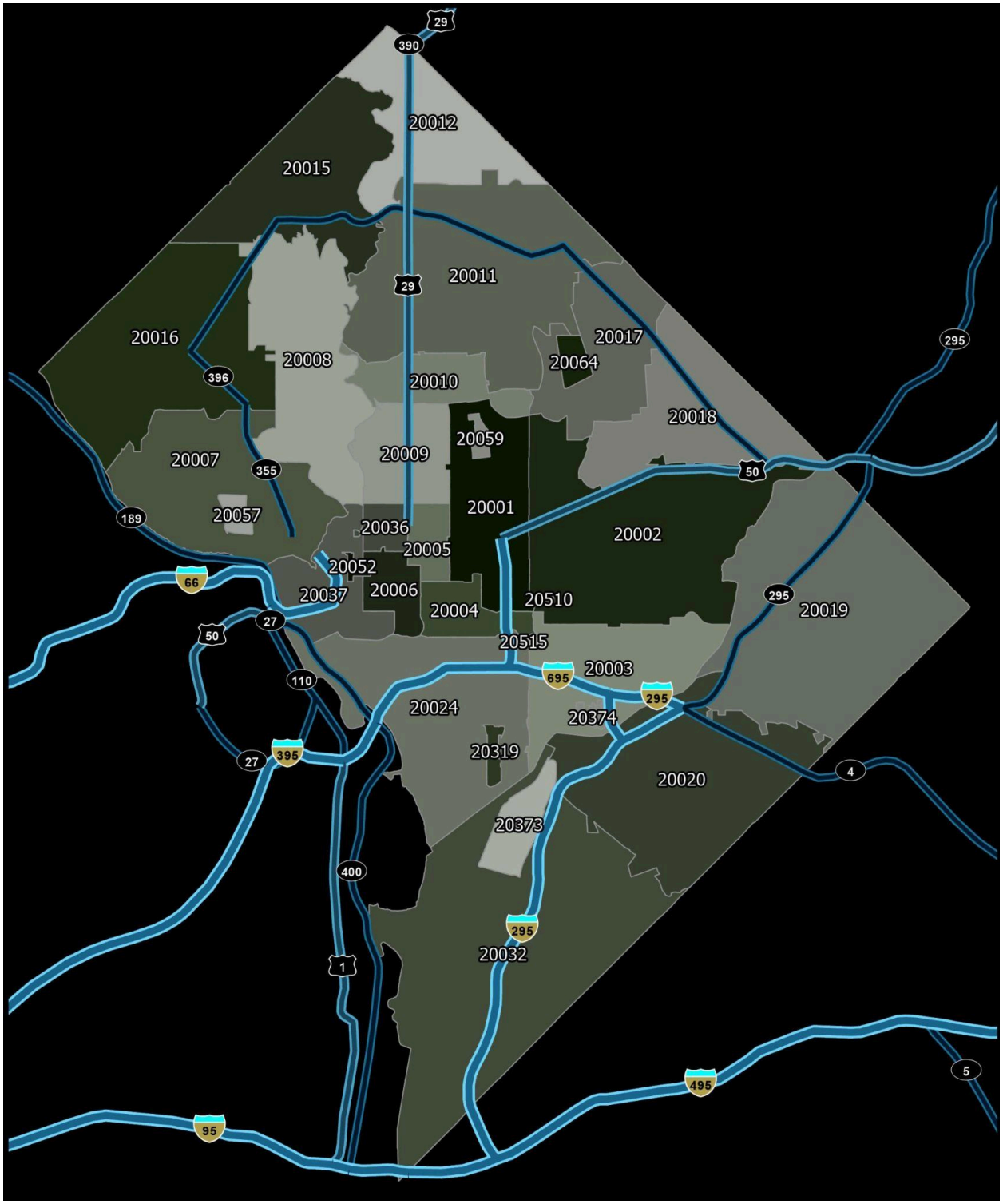
1. **Scatterplot Airbnb Density vs Zillow Median Housing Price**



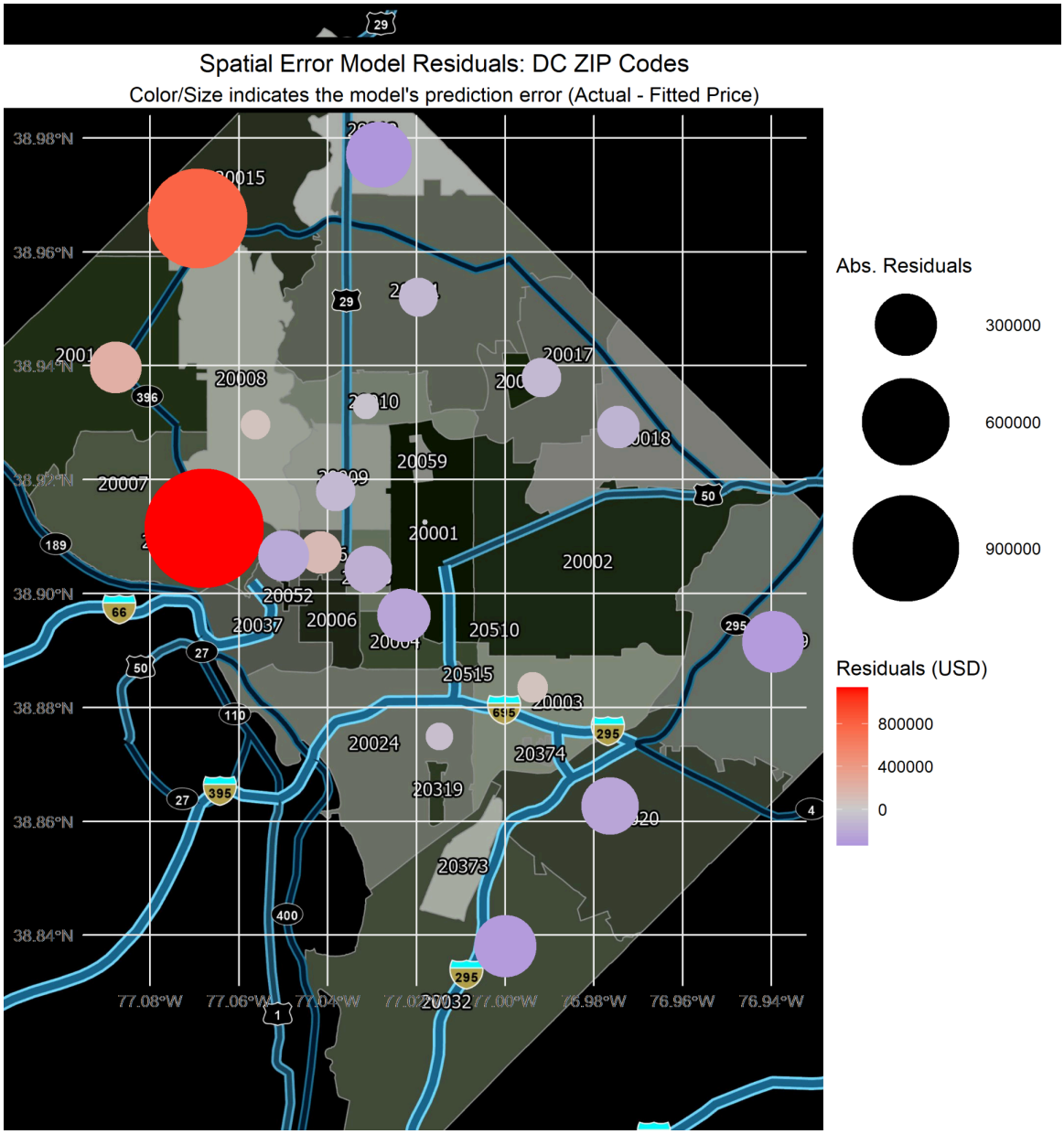
2. **Residuals Plot of Housing Price vs. Fitted Price**

Josiah: The two visualizations above were the first sign that something might be off with our data. Even after data cleanup, there are three significant outliers above 1.5 million dollars for the median Zillow price. These reduce our R^2 value below **0.1**. For this reason, we decided to do some digging. It turns out that one of the assumptions required for an R^2 test is that the points are not independent of each other, and for that reason we decided to employ a Spatial Econometrics Model for this analysis, which takes proximity into account, as well as utilize residuals, which is the amount of that each ZIP code's Zillow median is away from what is expected when using the line of best fit.

3. ZIP Code Map (Blank, for reference)



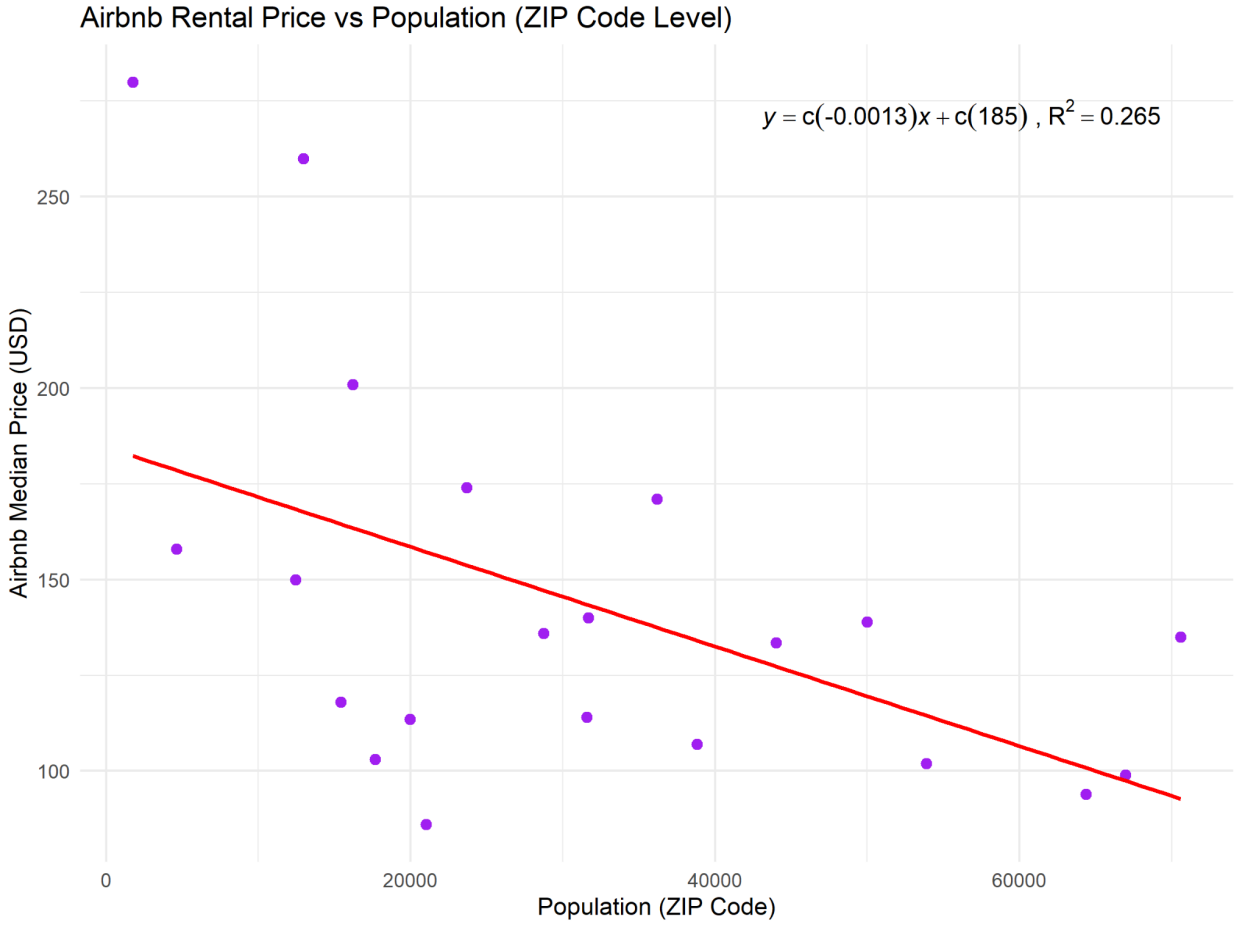
4. **Spatial Error Model using Residuals (How far off is a ZIP code from the predicted Zillow Price based on its number of Airbnbs?)**



Josiah: This visualization uses the geographical coordinates present in our dataset to map out the distribution of Zillow residuals. Each dot represents ZIP code. The larger and/or redder a circle is, the more expensive these ZIP codes are than expected, even after accounting for their Airbnb count and the influence of neighboring ZIP codes. The smaller and/or bluer a ZIP code is, the less expensive they are than expected, which indicates that something unmeasured is making housing cheaper in those areas. ZIP codes that are grey and small (including 20001, which is almost invisible), the closer it is to following the model.

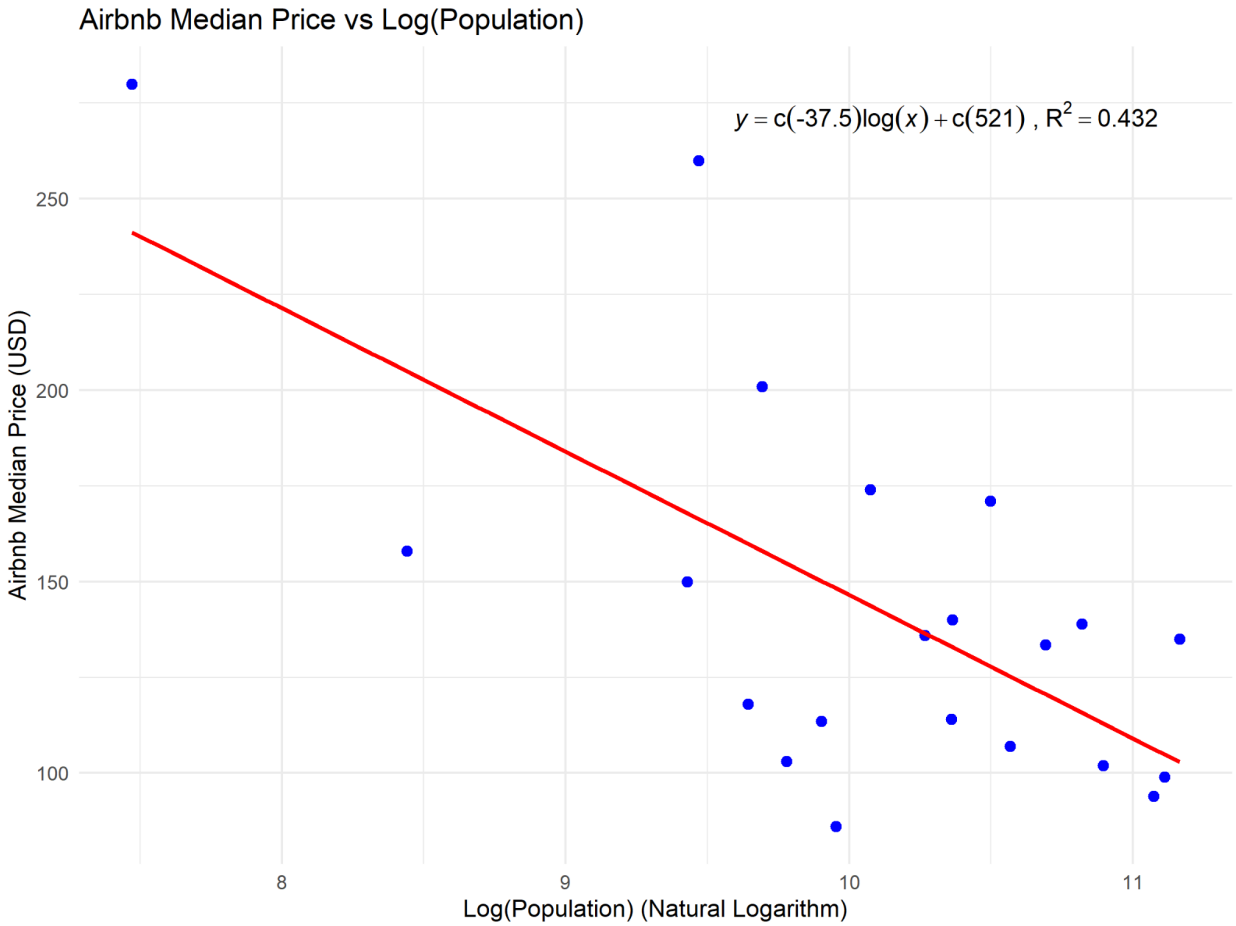
Question 2: Airbnb Price vs Population

1. **Scatterplot: Airbnb Price vs Population (Linear)**



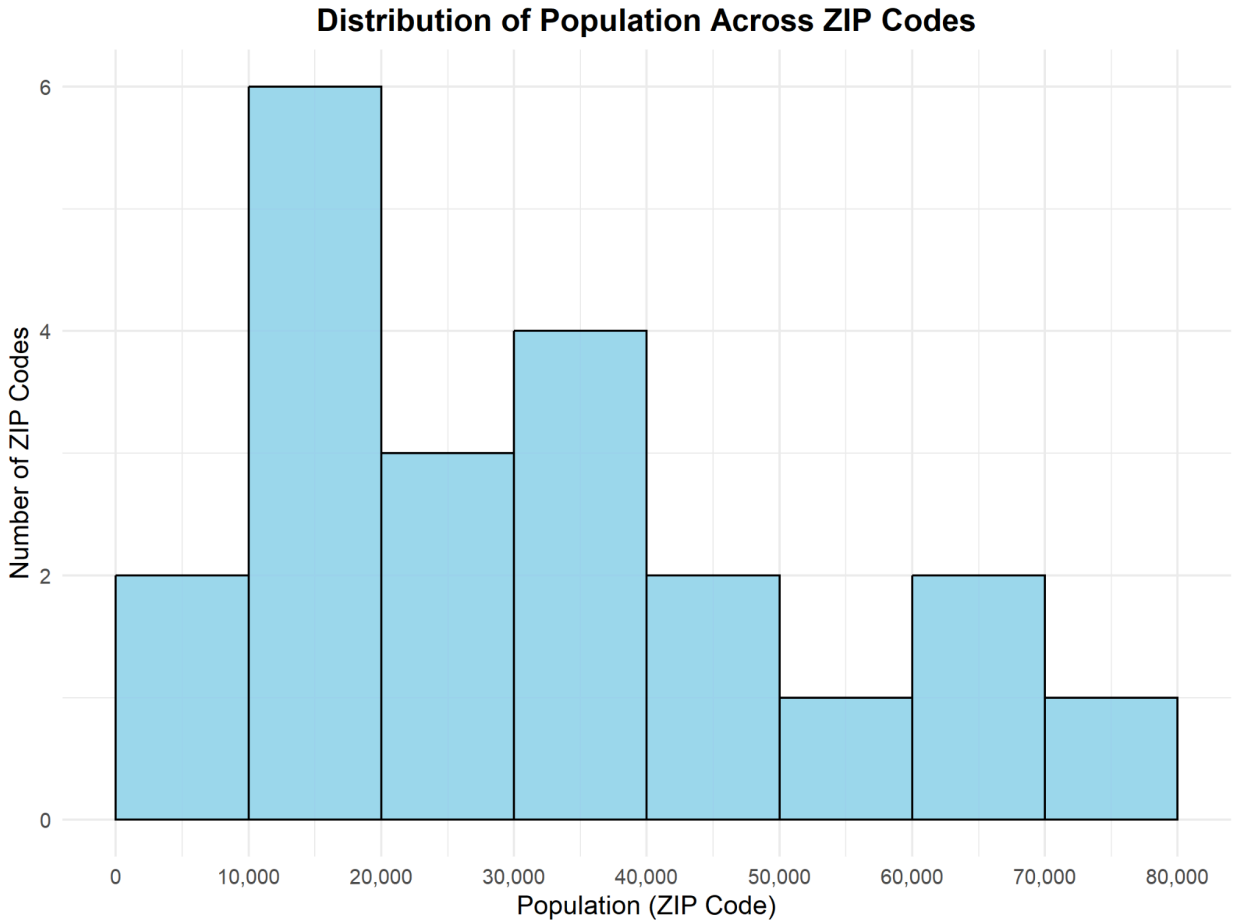
This scatterplot suggests that, indeed, there is a slight negative relationship between ZIP-code population and Airbnb median price, although the relationship isn't particularly strong. Altogether, more populated neighborhoods tend to have somewhat lower Airbnb prices, which fits the downward-sloping trend line. However, the points are spread out pretty widely, meaning ZIP codes with similar populations can still have very different prices. An R^2 value of about 0.26 also shows that population only explains a small part of what's going on-a quarter of the variation. So even though there was a small negative trend, population by itself doesn't really determine Airbnb pricing in DC. Other factors-just how desirable the neighborhood is, or what kinds of properties are available-probably matter a lot more.

2. **Scatterplot: Airbnb Price vs Population (Logarithmic – note higher R^2)**



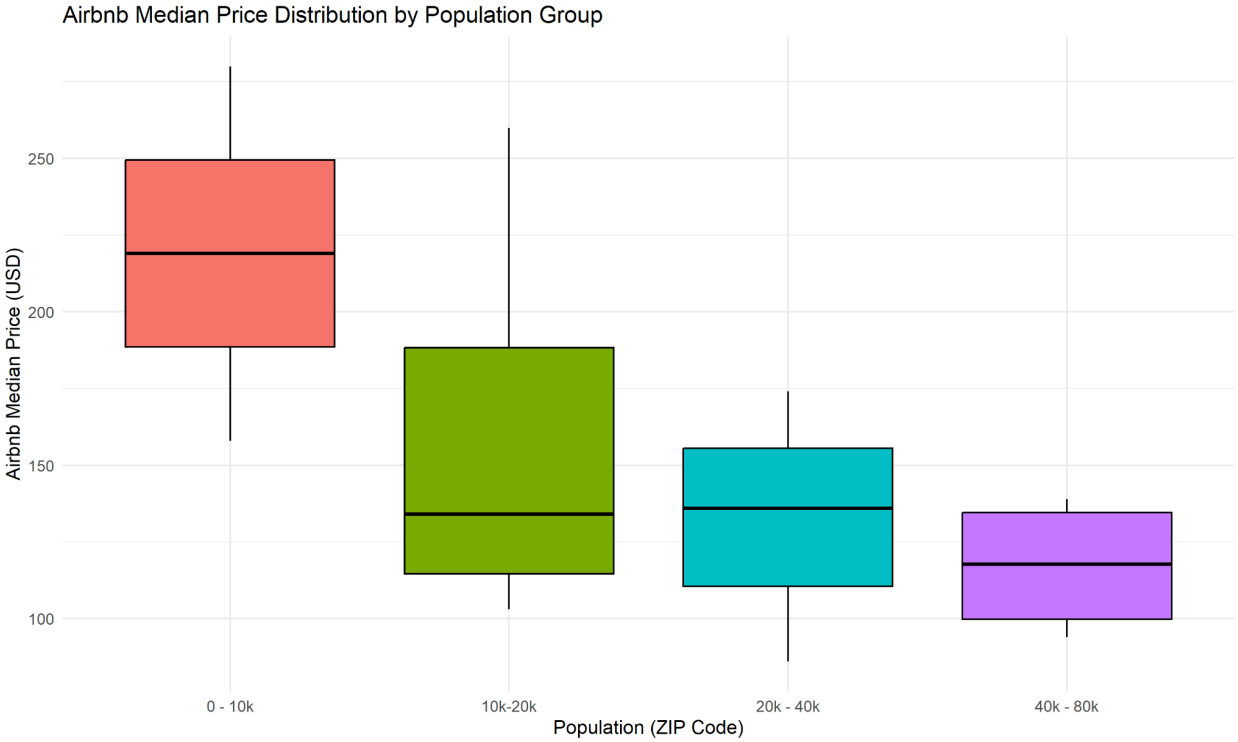
The higher R^2 value of our linear model (0.43 vs. 0.265) indicates that there is likely a negative correlation between median Airbnb price and population, but that negative correlation is more likely to be a logarithmic relationship than a linear one.

3. **Histogram: Distribution of Population Across ZIP Codes (How many ZIP codes have a given range of population?)**



This histogram displays the distribution of DC's population across zip codes, and informs us as to the distribution of the population across DC. Most ZIP codes have a low population, and only a few have a high population. Using this histogram, we initially grouped the bins into three groups: 0-20k (8 ZIP codes), 20k-40k (7 ZIP codes), and 40k+ (6 ZIP codes).

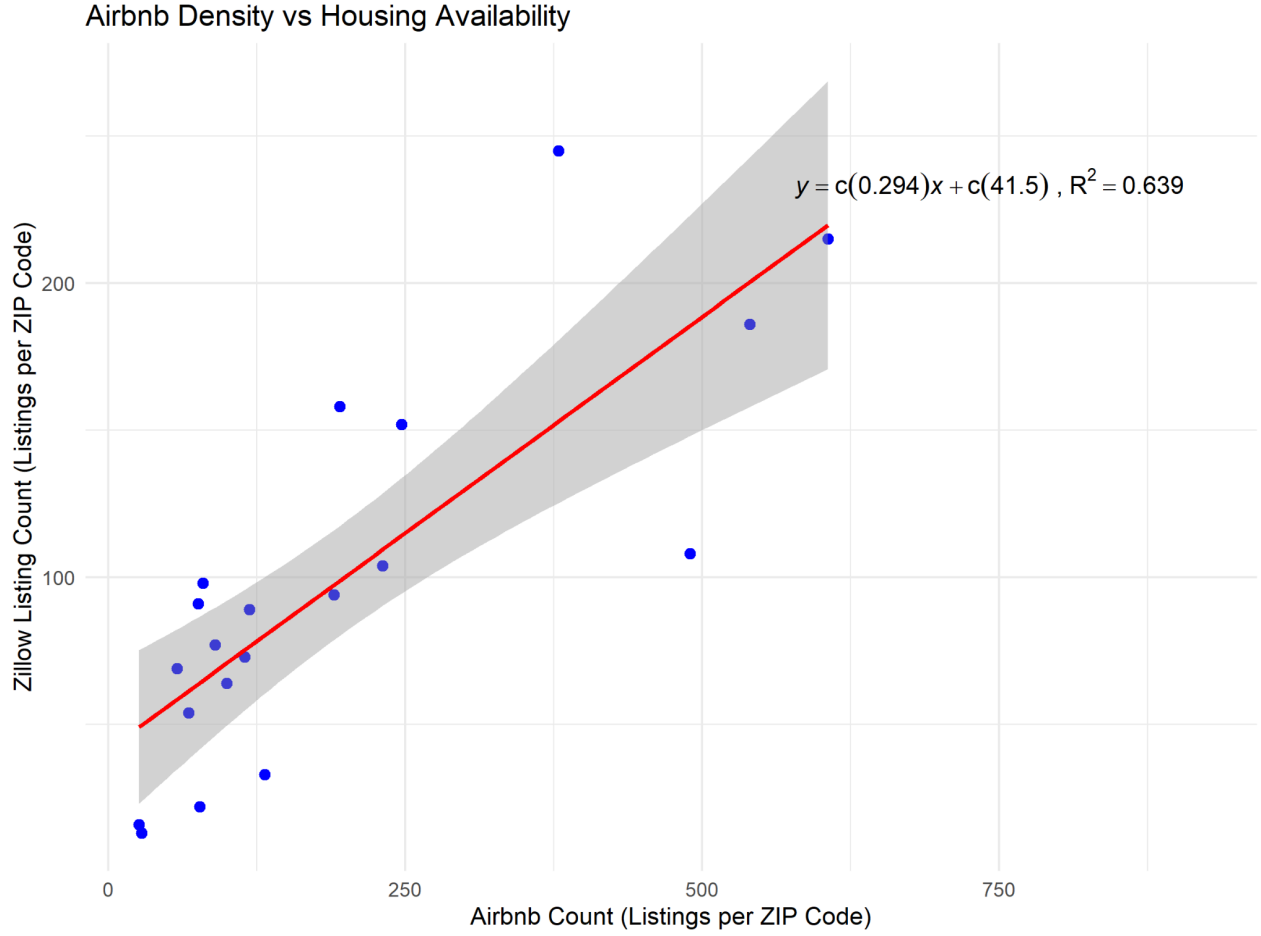
4. Boxplot of Airbnb Median Price Distribution by Population Group



This boxplot compares the population numbers of the ZIP code groupings we discussed above to the median price of Airbnbs in those same groupings. We decided to add another group – 0-10k, as we thought it might provide further information. This was correct. Through this graph, we have found that there is a negative correlation between ZIP code population and Airbnb median price, as the median Airbnb price lowers as the population increases in each ZIP code.

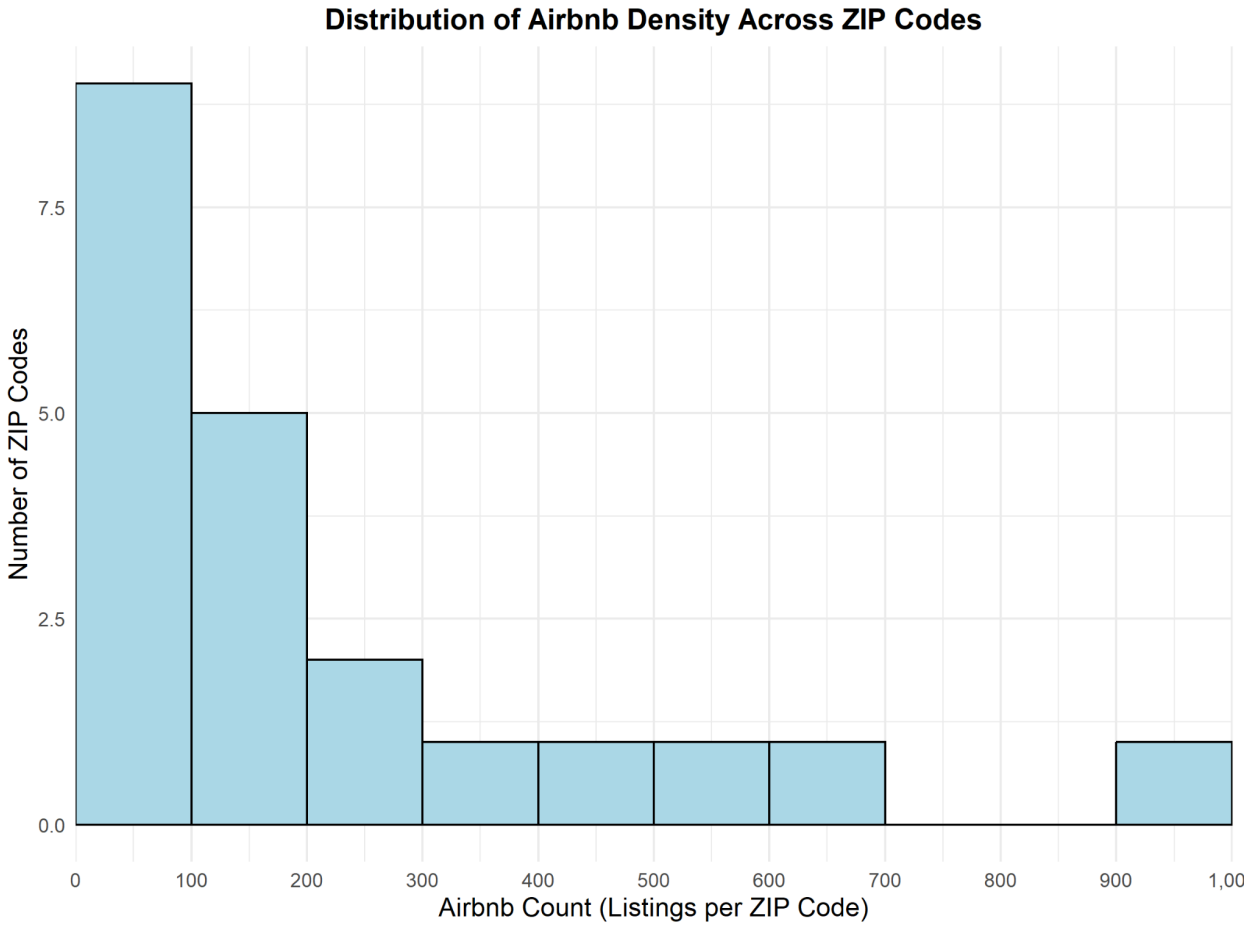
Question 3: Airbnb Density vs. Availability of Housing

1. **Scatterplot Airbnb Density vs Housing Availability**



This scatterplot displays the number of Airbnb listings in each ZIP code against the number of Zillow housing listings in that ZIP code. Most points cluster in the low Airbnb and Zillow count range, and the upward-sloping regression line indicates a positive association, and the R^2 of 0.64 indicates a decent degree of correlation. As such, we conclude that ZIP codes with more Airbnb properties also tend to have more housing listings overall. We think that high density for short-term rentals is also indicative of areas where more housing inventory is available, rather than Airbnbs showing up only in low-supply neighborhoods.

2. **Histogram for Distribution of Airbnb Density Across ZIP Codes**



This histogram provides the distribution of Airbnb listings across ZIP codes. Most of the ZIP codes have relatively low Airbnb density, generally below 200 listings, though there are only a few ZIP codes with extremely high Airbnb activity, ranging between 500 and over 900 listings. This reflects that the density of Airbnb is very unevenly distributed; there are only a few ZIP codes serving as the major hubs for short-term rental activity, while the counts for most others are moderate or low.