

Final Report

Spring 2020 CSCI 3390 Network Science

Group Members: Lizzy Hanley, Kieran Roth, Catriona Sullivan, Kevin Li

Monday, May 4th, 2020

Abstract	2
Data Sources	2
Our Motivation	2
How Our Work Helps Others	3
Publication	3
Listening to Feedback	3
Adjusting Our Final Milestone	3
Map Prototype Iterations	4
Early Stage Map Prototype	4
Figure 1: Airbnb price heatmap of NYC	4
Final Stage Map Prototype	5
Figure 3: ArcGis MapLayer of Airbnb locations in NYC	6
Figure 4: ArcGis NYC Airbnb prices map	6
Figure 5: ArcGis NYC Airbnb prices map, only 1st quintile is toggled on	7
Instructions for viewing our maps	8
Figure 6: ArcGis NYC Airbnb prices map legend	8
Website Design	9
Figure 7: Components Folder	9
Figure 8: index.js File	10
Github Link: https://github.com/kproth99/airbnb-wealthdistr	10
Github Pages Link (use this to see website): https://kproth99.github.io/airbnb-wealthdistr/	10
Meeting & Exceeding our Goal	10
Conclusion: Airbnb prices do not predict wealth distribution in a city	11
Group Report Questions	11
Did you meet your milestone stated in the initial report? If not, what made you fail to achieve it?	11
What did you learn from this project?	11
Do you agree to share your code/report with other students, say in other classes or future Algorithms class?	12

[Bonus: up to 5pt] Do you have any suggestions on the final project? E.g., more help? More resources? Format of poster session? Workload?...etc. Please be specific.

12

Abstract

Does the distribution of Airbnb prices in a city accurately represent the wealth distribution of that city? The three cities our group has selected to investigate are Austin, TX, New York, NY, and San Francisco, CA. For each city, we created a visualization of its wealth distribution where each real estate tract in the city is categorized into one of five income tiers. We then added a layered visualization of the pricing of Airbnb listings in that city as of February 12th. The Airbnb price layer categorizes each listing into one of five pricing tiers as well, following a quintile distribution.

Data Sources

SF, NY, and TX Airbnb data: <http://insideairbnb.com/get-the-data.html>

2019 USA Median Household Income:

<https://www.arcgis.com/home/item.html?id=a95db032417f40579fb68ff98ca14847>

Our Motivation

Airbnb is an online marketplace that allows travelers to find temporary housing from hosts who want to rent out their home. An individual or family household that takes in a high income is more able and likely to afford a high quality home. Therefore, it makes sense that homeowners who rent out these homes through Airbnb will charge a fee proportionate to the quality of housing. The decision for an individual to rent an Airbnb instead of booking a hotel is motivated by the belief that Airbnb homes give a customized and personalized experience. As a result, travelers can explore a new city in the most authentic way possible. Each member of our group has used Airbnb and is curious about how a price listing might reflect the actual wealth of the population in that area. Our motivations for using Airbnb were consistent: we were travelling to a new city and needed to find a place to stay for a short period of time.

After brainstorming various project ideas that could connect the topics in network science to services and platforms that people use frequently, we came up with the idea to base our project on Airbnb. We wanted to investigate how Airbnb prices in a city relate to the wealth distribution of that same city. Therefore, we sought to create a map that includes visualizations of both wealth distribution and Airbnb price distribution. To make this project feasible within the time constraint of one semester, we focused on three major cities across the USA: Austin, Texas, New York, New York, and San Francisco, California. These are cities our group members have traveled to and loved, and include a variety of wealth levels.

How Our Work Helps Others

We will be using Airbnb listing data from Inside Airbnb and [2019 Median Household Income in the United States](#) to measure wealth on a visual map.

For many folks, the decision to rent an Airbnb rather than a hotel is motivated by the belief that an Airbnb is a personalized experience, and thus an authentic way to explore a new city. We hope our visualization of these two datasets will create value for renters on the Airbnb platform by allowing them to understand if the listing they are looking to rent will be as genuine of an experience as Airbnb intends it to be. In addition, studies and articles that we have looked at (see initial report) allude to the point that Airbnb may be the cause of disproportionate wealth distribution. We kept this in mind when building our project and our map allows users to view Airbnb listings while also viewing the median household income of the surrounding area.

Publication

A great way to publicize our work would be to incorporate it into the Inside Airbnb tool. Because Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world, our analysis would enrich their platform. Our widget could be incorporated into the “see the data visually” page on InsideAirbnb’s website!

Listening to Feedback

In our initial report we stated that our goal is to test the hypothesis that the distribution of Airbnb prices in a city accurately represents the wealth distribution in that city. We defined ‘accurate’ as a 70% similarity or higher. Since then, we received feedback regarding how to better validate our criteria. Our group found a [Harvard study](#) that supports our choice to use median household income as a wealth distribution metric. We also decided to remove the 70% benchmark from our investigation as there was no data to back this number up.

Adjusting Our Final Milestone

With the interruption of COVID-19 into our semester, we decided that modifying our final report milestone was a prudent choice. After considering two options, one that focused on the technical analysis of our data and one that focused on the presentation of our project, we chose the latter. With only so much time in the semester left, we wanted to make the most out of our project by creating something interesting, while learning new web development skills.

Our new goal became creating a website for our project that would present the ArcGIS maps we had already made. We decided to use interactive maps, so a user could click on data points to view the price of an Airbnb and the median income in that area. We did not intend to do anymore work on our ArcGIS maps or make a heatmap of the Airbnb pricing data. Instead, we wanted our time to be spent

allowing individuals in our group to gain exposure to web technology and, for the group members who have done web development before, strengthen their skills!

Map Prototype Iterations

We knew we wanted to gain the most knowledge that we can from completing this project. Therefore, we researched several options on creating maps with data representation and through trial and error, were able to narrow down the best implementation method.

Early Stage Map Prototype

We wanted to use Python as our primary coding language because of its many powerful data analysis libraries and we were all comfortable coding in this language. In our early stage prototype, we focused on creating a heatmap of Airbnb listings across a city.

The data that we downloaded from InsideAirbnb had numerous fields for each Airbnb listing ranging from address, to reviews, to descriptions of the Airbnb home. Although our dataset had over 300 fields, we were only interested in three: nightly price, latitude, and longitude of the Airbnb. We were able to create a vivid heatmap using a python package called folium by iterating over the dataframe for each pair of coordinates and then adding a marker point to the map. We then plotted a heatmap of the prices of each listing, creating a complete visual representation (Figure 1) of the Inside Airbnb cleansed and filtered dataset.

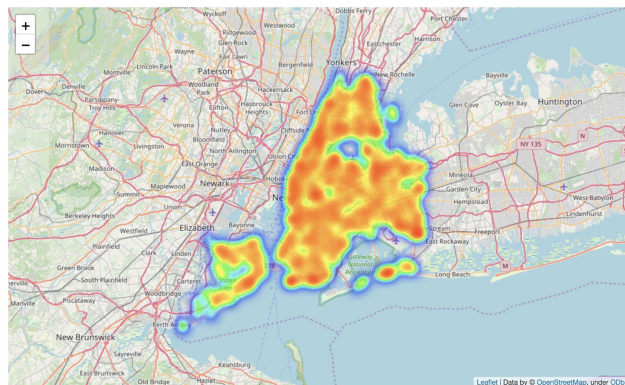


Figure 1: Airbnb price heatmap of NYC

Now, we knew ahead of time that our next objective was to create the same type of heatmap but using median household income in the United States. The challenge was finding usable data that would have the median incomes, latitude, and longitude of each area in the city. Datasets that we had looked into either were missing certain essential attributes like latitude and longitude, or required us to pay to access their full datasets.

We then found ArcGIS, which is a platform that allows organizations to create and share projects involving geographical data. ArcGIS consists of developer tools that allowed us to create various layers of

data points onto a map. We discovered a detailed ArcGIS map that illustrated the [2019 Median Household Income in the United States](#) (Figure 2) divided into different sections of a city.

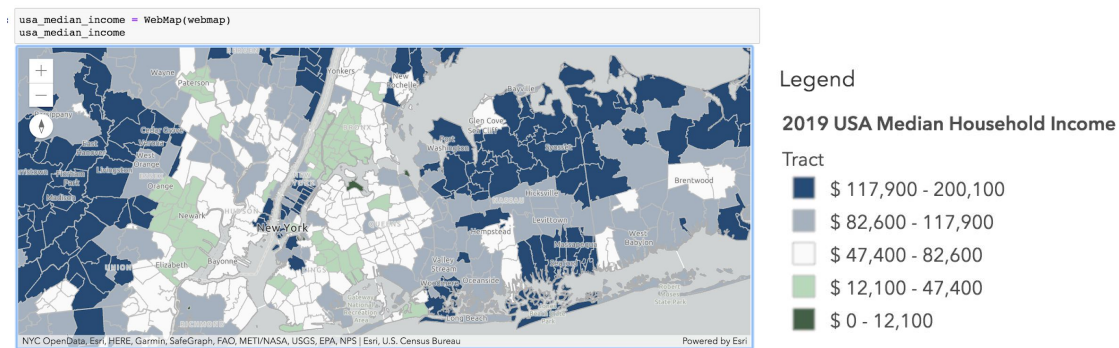


Figure 2: Wealth distribution in NYC and map legend

Our discovery of this platform marked the start of us pivoting our project milestone. Although we were able to import this web map into Python, we were unsuccessful in combining the two together to show a detailed visual comparison between the Airbnb price listings and the 2019 U.S. median incomes for different areas of a city. This propelled our decision to explore the ArcGIS platform more extensively. Our decision to delve deeper into exploring ArcGIS ultimately narrowed us down to our final stage map prototype. Unfortunately, technical problems on Lizzy's laptop resulted in the loss of our Jupyter notebook containing code for our early stage map prototype (she spilled water on her computer and broke her harddrive). This was the same Jupyter notebook that we presented during our midterm oral report presentation. Fortunately, our final stage map prototype no longer required code from the notebook, so this disaster ended up not affecting us at all.

Final Stage Map Prototype

Although we had not set goals for continuing to work on our maps, we were able to allocate time to learning how to use ArcGIS, thus we implemented readability improvements. Originally, our ArcGIS maps looked like this:

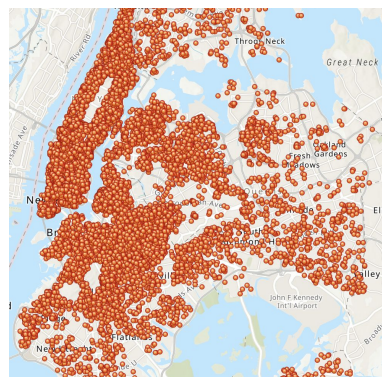


Figure 3: ArcGIS MapLayer of Airbnb locations in NYC

As you can see, we plotted our Airbnb listings using coordinates and this visual did not convey any information about pricing distribution unless the user clicked on a data point to see more details. This was okay, but a far reach from the heatmap of pricing distribution we had intended to create.

In order to learn how to use ArcGIS, we had to read many pages of documentation describing how to create custom layers from our own datasets, how to edit the display of these feature layers, and how to create a map legend. After we felt more confident using this new technology, we were better able to assess what would be feasible for us to build.

Even after this research, we were still struggling to create a heatmap following the same color scheme as the 2019 median household income layer. This meant we had to think of a different way to visualize pricing distribution. Since the 2019 median household income layer categorizes income into one of five quintiles, we decide to create a quintile distribution for our Airbnb data. We did this by ordering the data in our csv files in ascending order according to price, and then dividing the data five, roughly equally new files. Each file was roughly $\frac{1}{5}$ the size of our original file for that city because we wanted each file to have unique price boundaries. Each csv file now represented a single quintile of Airbnb pricing data. We then uploaded each file to ArcGIS and created a feature layer for each quintile. From here, we added each feature layer to our map, using the 2019 median household income layer as our base layer. Lastly, we edited the symbols for each of our Airbnb feature layers so they followed the same color scale as the 2019 median household income layer.

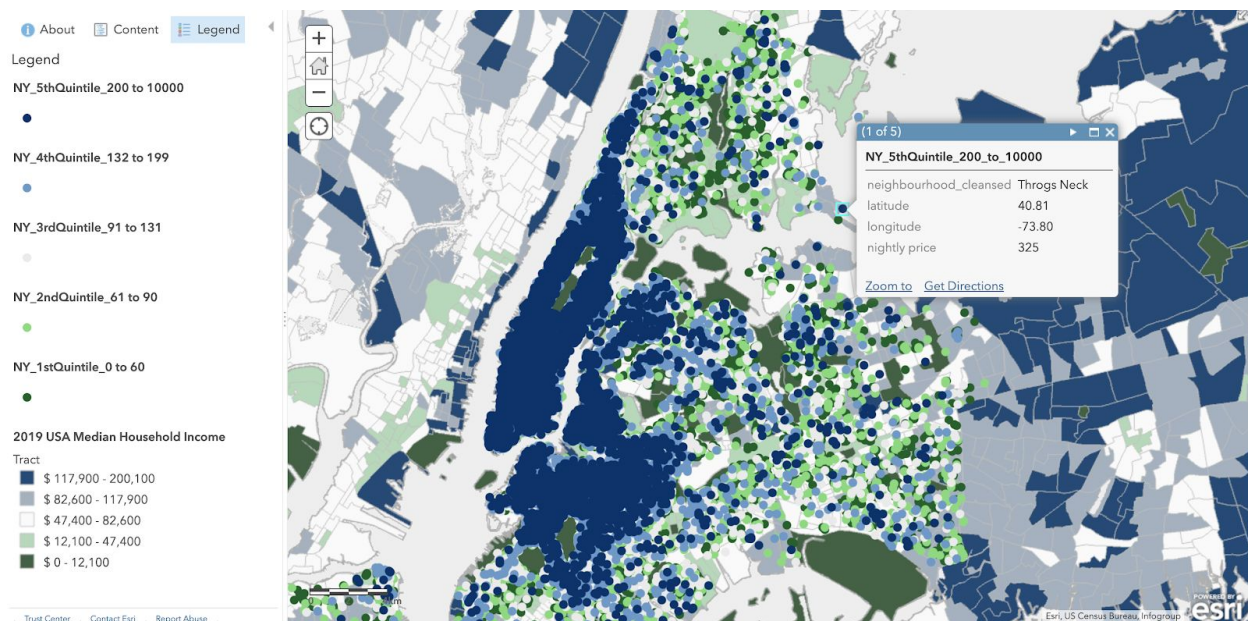


Figure 4: ArcGIS NYC Airbnb prices map

Figure 4 shows the final version of our NYC Airbnb prices map. When a datapoint is selected, a menu opens to show more details about that listing: its neighborhood, its coordinates, and its nightly

price. Each of our quintile layers can be toggled on and off. We added this feature so a user can select a single quintile at a time, and view where those listings are.

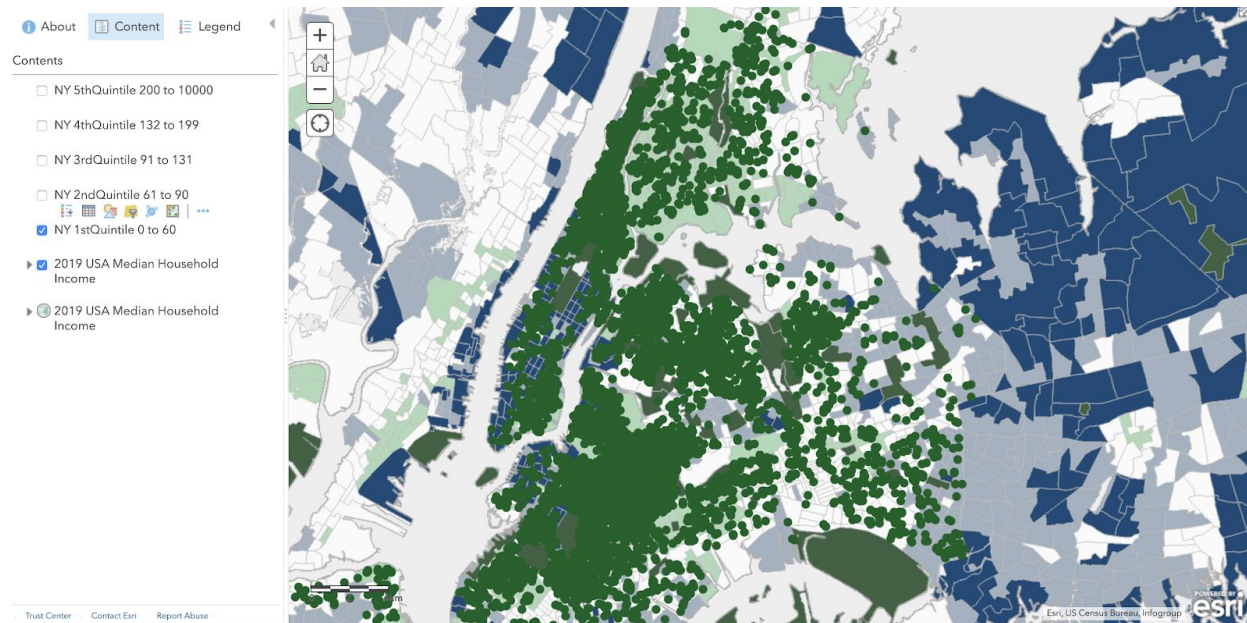


Figure 5: ArcGis NYC Airbnb prices map, only 1st quintile is toggled on

As you can see from Figure 5, for New York, listings in the first quintile of pricing (dark green dots) are not limited to areas of the city with the lowest income (dark green real estate tracts). In fact these cheapest listings can be found anywhere in New York, including the most expensive areas!

Links for the final Austin and San Francisco maps are included below, as well as the 15 feature layers we had to create.

[Austin Airbnb price map](#)

[1st quintile, listings range from \\$0 - \\$73](#)

[2nd quintile, listings range from \\$74 - \\$115](#)

[3rd quintile, listings range from \\$116 - \\$194](#)

[4th quintile, listings range from \\$195 - \\$399](#)

[5th quintile, listings range from \\$400 - \\$14298](#)

[New York City Airbnb price map](#)

[1st quintile, listings range from \\$0 - \\$60](#)

[2nd quintile, listings range from \\$61 - \\$90](#)

[3rd quintile, listings range from \\$91 - \\$131](#)

[4th quintile, listings range from \\$132 - \\$199](#)

[5th quintile, listings range from \\$200 - \\$10000](#)

[San Francisco Airbnb price map](#)

[1st quintile, listings range from \\$0 - \\$89](#)

[2nd quintile, listings range from \\$90 - \\$129](#)

[3rd quintile, listings range from \\$130 - \\$184](#)

[4th quintile, listings range from \\$185 - \\$270](#)
[5th quintile, listings range from \\$271 - \\$10000](#)

Instructions for viewing our maps

You may be prompted to log into ArcGIS and so if you already have an account feel free to use it. Otherwise, you can log in using the login credentials we emailed you and commented on our canvas submission.

To see areas of each city in more detail, zoom in and out using the “+” and “-” buttons. The more you zoom in, the clearer each individual listing becomes. In some areas of our maps, all five quintiles of Airbnb listings can be found in a single real estate block!

On the top left corner you have the option to click the arrow button to filter and view some of the map layer details. You will be able to view all of our layers in the “Content” section, and also select which layers to view at a time. The default view displays all the layers at once, but feel free to toggle on and off each quintile map layer of the Airbnb listings. To understand the color scheme of each map, click on “Legend” on the left side of the map. Here you will see an explanation of what the color of each Airbnb listing data point corresponds to in terms of pricing quintile and pricing range. You can also view the legend for the 2019 median household income layer and the income ranges each block color corresponds to.

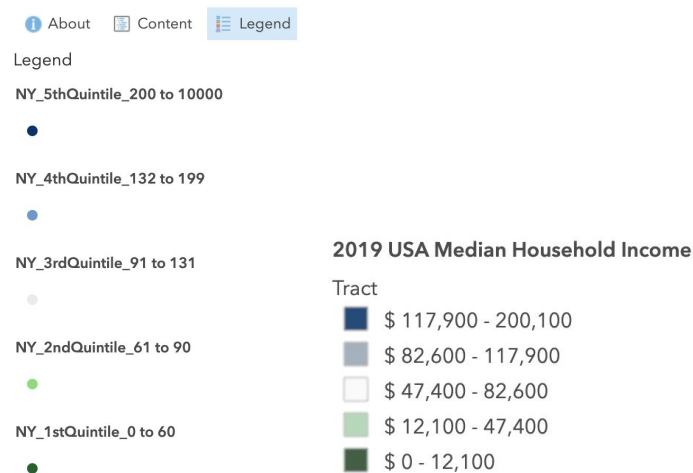


Figure 6: ArcGis NYC Airbnb prices map legend

Lastly, you can click a data point to reveal additional details, including the neighborhood, nightly price, latitude, and longitude of that Airbnb listing. Clicking on a block group will display the median household income of that area in comparison to the national median household income, and a bar graph categorizing median household income by age groups. You can hover over each bar of the graph to show the exact median household income for that specific age group.

Website Design

We constructed our website using mainly JavaScript and HTML rendered within JS files. The website is structured as a portfolio to present and display our project. We start with a gallery display of each group member, and under that is a brief description of what you will find on our page and about the project in general. Below that there is an icon for each of the three cities along with a link to each corresponding map. Once you click on the link you are taken to another page with the embedded interactive map where you can click and zoom in using the ArcGIS functionality. Then you are able to click the “Go Back” button to be taken back to the main page. The final section lists our Project Deliverables (Final Report, Midterm Report, and Initial Report), along with an icon in which the user can click to access our PDF files for the reports and learn even more about our project!

Now to discuss the actual code of the website. Our goal was to make the code as easy to read as possible for an outsider. In our Components folder you will see multiple .js files, for example About.js or MainPage.js. Each of these files contain export functions that specifically render a certain part of the website: the About file renders the about section, the Navbar file renders the navigation bar at the top of the website, the Project file renders the project section, and so on. Using these render functions, instead of just coding an incredibly long index.html file which would contain all the information in one place, allows not only for an easier follow-along for the user, but also allows for the use of database in which we can edit and store the information displayed on our site.

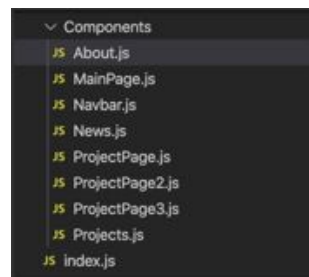


Figure 7: Components Folder

Using a .json file to store our information proved incredibly helpful when it came to creating the separate pages for each city map. In the above picture, you will see multiple ProjectPage files. Each file contains the same exact code except for the specific map, as each file uses different embedding code given to us by ArcGIS, while still all pulling from the json database in order to display the description. In order to get the website to display the different files for each separate project, we needed to edit the index.js file and include a conditional statement.

```

fetch("assets/data.json")
  .then(resp => resp.json())
  .then(data => {
    console.log(data);
    const params = new URLSearchParams(window.location.search);
    console.log(params);
    console.log('projects.param', params.get('project'));
    if (params.get("project") === null) {
      renderMainPage(data);
    }
    else if (params.get("project") === "austin"){
      let id = params.get("project");
      console.log(id);
      let project = data.projects.find(function(project){
        console.log(project.id)
        return project.id === id;
      });
      renderProjectPage(project);
    }
    else if (params.get("project") === "newyork"){
      let id = params.get("project");
      console.log(id);
      let project = data.projects.find(function(project){
        console.log(project.id)
        return project.id === id;
      });
      renderProjectPage2(project);
    }
    else if (params.get("project") === "sanfran"){
      let id = params.get("project");
      console.log(id);
      let project = data.projects.find(function(project){
        console.log(project.id)
        return project.id === id;
      });
      renderProjectPage3(project);
    }
  });

```

Figure 8: index.js File

In this code above, you can see that we are fetching the data from our json file and attempting to match up that data in order to render the proper page. In the json file, each map is assigned a different id (Austin receives “Austin, New York receives “newyork”, etc.) and thus we can utilize that id in order to render the correct page. So if the user clicks the link listed under the Project section and the parameter of “newyork” is received by the website, then renderProjectPage2 is called and the New York page is loaded. With this structure, we hoped to build a website which is not just information hardcoded into one HTML file, but instead a website which utilizes higher-level structures and rendering processes while still being easy to understand by the user.

Github Link: <https://github.com/kproth99/airbnb-wealthdistr>

Github Pages Link (use this to see website):
<https://kproth99.github.io/airbnb-wealthdistr/>

Meeting & Exceeding our Goal

We were able to meet the goals we set out for ourselves and even put more work into our maps than we were originally intending. The foundation of our goal was to have a price distribution map and wealth distribution map for each city, which we definitely met. Once we started learning more about ArcGIS, we realized that it would be very beneficial to gain these new skills and also add more functionality to our maps. ArcGIS gave us the ability to have more readability because we could color coordinate the data points and include a legend for both the price distribution and median income distribution on the side. Also, it was easier to create separate layers for the price data points rather than make a heat map, and this resulted in a final map that was equally informative. Additionally we didn't

have a strong idea of whether we wanted to have a website or display the maps only on ArcGIS, but when we decided to create a website we were able to come up with a product that exceeded the original vision. We not only embedded our maps in our website, but we also created an “About” section to introduce the team and project, and a “Deliverables” section with all of our reports and video! Ultimately, every member of our team developed new skills and got to use them to make a project we are very proud of.

Conclusion: Airbnb prices do not predict wealth distribution in a city

Although we decided not to go forward with a comparison metric for the Airbnb price distribution and wealth distribution, by exploring the maps we came up with some useful observations. If you zoom into a smaller area of the map that is part of the lowest income quintile, there are still data points representative of multiple Airbnb price ranges. For example, in one heat map block of San Francisco, the median household income is \$151,588, which is in the highest income range, but there is one data point for a neighborhood called Diamond Heights in that area that has an Airbnb nightly price of \$80, which is in the lowest price range. From this observation we can note that Airbnb prices do not predict wealth distribution for a city.

From a real life standpoint, if someone is looking to find a cheap Airbnb rental they do not necessarily have to look in lower income neighborhoods to find it. Many people are not planning travel in the current COVID-19 situation, however, it is still helpful to be well informed for future travel. A potential extension to the project we have created is to add maps for even more cities to see if the same pattern is present. That being said, the three cities we did choose show a good range of wealth distributions and are major travel destinations for people from all over the world.

Group Report Questions

Did you meet your milestone stated in the initial report? If not, what made you fail to achieve it?

In the initial report our milestone for the final product was to create a web app that would display a heat map of the wealth distribution of a city with a price heatmap of Airbnb listings layered on top. The web app would allow the user to select from three cities. Both the data points and the heat map would follow a color scale according to price and highest median income respectively. This is something we were definitely able to accomplish. We built a website with links to maps of Austin, New York City, and San Francisco. Using ArcGIS we were able to create interactive maps with Airbnb listings as data points and a wealth distribution layer. In our milestone we also added that we wanted to incorporate a comparison metric if we had extra time. We did not end up finding a comparison metric, but instead used our time to focus on enhancing the design aspect of our website.

What did you learn from this project?

From this project we learned how to use new technologies and set manageable goals to meet each milestone. ArcGIS was a new software for all of us but by experimenting and doing research we were able to get to the final product we were looking for. We learned how to create map layers and transform

our data into visualizations that convey meaningful conclusions. We collaborated on the web design part of the project with tools some of us had learned from previous classes. There are many different avenues to take when building a website, so it was interesting to learn how HTML is sometimes rendered within JavaScript versus having completely separate files for each.

Do you agree to share your code/report with other students, say in other classes or future Algorithms class?

We agree to share our code and report with other students. We didn't do something that was directly related to networks, but it provides a good example of using outside technologies and web design elements to make an engaging product. Students who are considering renting an Airbnb in Austin, San Francisco, or New York City might also find our work useful!

[Bonus: up to 5pt] Do you have any suggestions on the final project? E.g., more help? More resources? Format of poster session? Workload?...etc. Please be specific.

It would have been helpful to have more guidance in the beginning stages of the project. Since the project is more open-ended it was difficult to come up with a topic, question to investigate, and find a dataset. It was good to have the examples from previous Algorithms projects, but it would have been better to either have more network-related examples or links to resources giving tips on how to formulate good project ideas.

Also, it felt slightly repetitive to create a very detailed report and, on top of that, record an oral presentation. Getting presentation skills is very useful, but perhaps the report could have been shortened or converted to a slide deck. We really enjoyed the format of the midterm oral presentation: book a time and explain the project in 15 minutes. We would have much preferred doing it the same way for the final presentation.