

Affordable Housing in Southern California

September 6, 2019

1 Where is affordable housing urgently needed in San Diego?

Our final project will deal with location where more affordable (Section 8) housing can be developed. To do this, we will first look at existing Section 8 housing within San Diego county. We feel this is an important issue because of a shortage of affordable housing in San Diego. For example, in a building that was supposed to have 33 units built, only 4 were built.

For our current existing affordable housing dataset, we scraped <https://www.sdhc.org/housing-opportunities/affordable-rentals/rent-from-sdhc/sdhc-owned-properties/> for their backend dataset of housing units at <https://www.sdhc.org/wp-content/themes/sdhc/properties.php?type=sdhc-owned-properties%2Chousing-development-partners&query=>, which returns a JSON file with all the current SDHC affordable housing units in San Diego. We then downloaded this for use as the file “properties.json”.

2 Background and Literature

2.0.1 A senior on California's streets with little chance of a home

This identifies the purpose of the research question, homelessness is a major issue in San Diego and neighbouring Southern California Counties. With the creation of greater public transportation links comes the ability to expand the affordable housing sector in areas not previously accessible before. With one of the 22 ‘promise zones’ in San Diego, in addition to a large number of work-force coming in from the Mexico-US border, housing supply needs to keep up with an increased demand; and the need to match the country wide inflation.

2.0.2 The Housing Crisis Is Obvious – but the Data Detailing Its Scope Is Not

“A city report says just 33 units for middle-income units were built in seven years. The real number is actually four”

Expanding on the point above, the previous problem these housing units are not adequately budgeted and if they are the proposals end up getting rejected by the lobbyists pushing for welfare and safety of its current citizens. With the public transport expansion(planned and current) it is possible to have these units away from the city. This creates neighbourhoods and communities, increasing the trickle down by adding jobs and infrastructure to different parts of the city.

2.0.3 SDHC - Affordable Housing Crisis Action Plan

Action plan from the San Diego Housing Commission about the current housing crisis in San Diego and what can be done to address it. Points include having annual goals of housing production, introducing tax rebates for housing, opening more vacant land for development, and allocating more resources towards low-income housing.

2.0.4 Chicago Affordable Rental Housing

This project was inspired from the Chicago Affordable Rental Housing project, the idea is to create a similar dataset for San Diego County by merging already available housing data and the San Diego County socio-economic and transportation patterns to add efficiency to this pre-existing model.

3 Python Libraries and ArcGIS modules used

Standard Python modules used

1. geopandas
2. pandas
3. Shapely
4. numpy
5. arcgis

GIS operations used

1. dissolve_boundaries
2. overlay_layers
3. geoenrichment
4. SpatialDataFrame & related functions
5. create_buffers
6. gis.map
7. spatial.join
8. Other functions on ArcGIS Online

This list evolved as we realized the limitations of the Zillow dataset. We had to use alternative datasets that were more detailed than the Zillow dataset. The inclusion of geopandas, pandas, Shapely, and numpy were needed in order to preprocess data into shape files and feature layers for ease of use as well as adding/dropping columns as part of data cleaning.

4 Data Sources

1. [SDHC existing properties](#): JSON file traced from SDHC homepage AJAX requests that loaded pins on Google Maps for existing SDHC properties
2. [San Diego County ZIP codes](#): Map of San Diego County ZIP codes
3. [ZIP Code mean and median incomes](#): University of Michigan dataset derived from the ACS survey that details mean and median incomes by ZIP code
4. [Transit Stops GTFS](#): Dataset of stop IDs and location

5. [Transit Stop Routes](#): Stop IDs and routes that serve every stop
6. [Transit Routes GTFS](#): Route geometry for SDMTS and NCTD
7. [City-owned Land](#): City-owned land for all parcels either owned by the city or leased to city.

In our project proposal, we were going to use homeless spot counts for analysis. However, we realized that a much better indicator of the need for affordable housing is to look at rent as a percentage of income as well as area median incomes, as the homeless population is very small when compared to the former. We also removed the Zillow dataset as a principal dataset because it did not give us fine enough data on the ZIP code level, which we had to use the UMich dataset as a substitute.

Many of our data sources used were from SANDAG or the open-data warehouse from the city for city resources (public transit, land). We could not obtain the existing SDHC properties through conventional means, so we had to resort to digging around on a web browser console and listening to web requests in order to gather our data.

5 Data Cleaning

The first point of data cleaning we had to do was to move our existing SDHC properties data to a GIS-compatible format from just a simple JSON file. To do this, we used pandas, geopandas, and Shapely to convert the data and then save it with geopandas to a .shp file type, which we could then upload onto ArcGIS Online. We thought that most of our data would be available already in some form of a GIS warehouse, so when we were confronted with this issue we had to think of a solution we used in previous MPs for this project.

The next point in cleaning we had to do was joining the 3 transit datasets in order to get meaningful information (routes that serve stops and geometry all together). We had to fill NULL values for stop_ids as well as route_ids in order to both merge and drop them.

Finally, a last part of cleaning we had to do was using numpy in order to do calculations on an enriched dataframe to go from total in an area to a percentage.

In our project proposal, we did not expect this amount of data cleaning, which soon became apparent as our first step was to preprocess the SDHC existing properties data into a usable format.

6 Descriptive Statistics

In the development of the project analysis, we first saw clusters of affordable housing scattered around maps. As we discussed in class spatial autocorrelation can be used to find clusters of similar data (+ve Morans I). The goal in relation to the current housing locations was to have to the new housing locations further away from the current sites. This in some sense could be considered to be a variation of negative spatial autocorrelation while we look for sites in inverse value of affordable housing availability.

For the most part, with the housing sites and the developable and vacant land plot we see random point patterns dispersed around the city based on land vacancy. Additionally, the results we can up with can be considered random point patterns, as the are not spatially correlated and depend heavily on city owned sites.

The description of the random point patterns in the data also explain how there is a shortage of spatial statistical descriptions. The diversity of datasets and criteria developed in the analysis do not allow for extensive statistical significance.