California Housing

Andrew Pang

anypang@ucsc.edu

University of California, Santa Cruz

CMPS 161/L - Introduction to Data Visualization

Professor Alex Pang

Winter 2019

1. Abstract

Housing storage is a serious issue in California. Many are driven out of their neighborhoods due to a variety of reasons. My project aims to help inform and visualize how different California counties compare to one another. Built with the Google Maps API, this project will take data and compile them into a neat google map format. With HTML, CSS, and JavaScript, the user can obtain of visualization heatmaps and color-coded stats with interactive data selection

2. Introduction & Approach

To start, my program will be powered by the Google Maps API. The choice to use Google Maps API is because of the large documentation and support that already exist within the community. Next is the datasets. The housing dataset I used is from Kaggle and the county dataset is from 2017 US Census(see Reference Section for link, more info about datasets can be found in the Technical Detail Section). The program will be mainly comprised of three files: HTML, CSS, and JavaScript. The HTML file will set up the structure of the whole program. CSS will stylized it. The JavaScript file will parse the datasets and make the page interactive. Since the two dataset are entirely separate, a huge challenge was to find a way to merge or bridge the two sets. Displaying stats for a county required: (1) which county shape from the county dataset did the user want and (2) which entry in the housing dataset is contained in that county shape bounds. This required a parsing of two list, or O(n^2) computation time. This was especially slow within a web browser. To solve this, I ran through the two file once and created an additional property for each housing entry to add the county name of which it belonged to. Due to time, I wasn't able to make it as smooth and responsive as I wanted but no interaction should take more than 2 seconds.

3. Technical Details

3.1. Datasets

- Housing Data
    - The data I used for housing is from Kaggle, an online community of data scientists and machine learners, owned by Google LLC. The raw filetype was in csv format but was converted to JavaScript Object Notation(JSON) format for JSON consistency. The dataset is based on the 1990 census in California. The data columns are:
        i. longitude
        ii. latitude
        iii. housing_median_age
        iv. total_rooms
        v. total_bedrooms
        vi. population
        vii. households
        viii. median_income

ix. median_house_value
x. Ocean_proximity

- 2017 California County Boundary Shapefile Data
  - The data I used to display California Counties is from the US Census Bureau. The file is in GeoJSON format, a geospatial data interchange format based on JSON. The GeoJSON type is a “FeatureCollection” and some main feature/properties are:
    i. COUNTYFP (Current county FIPS code)
    ii. COUSUBFP (Current county subdivision FIPS code)
    iii. COUSUBNS (Current county subdivision GNIS code)
    iv. NAME (Current county subdivision name)

  Note that many more properties are available and will be linked in the references section.

3.2. Programming Languages, Libraries, and Scripts

The Project consist of several files and is meant to run in the browser.

3.2.1. HTML

The HTML is pretty standard. It consist of the head and body. In the body tag, there are several different divs. One is the main map div for Google Maps. Next a small box div, meant for county name display whenever the user hovers over a county. Next, there is the option sidebar. In it, there are three tabs. The first tab is the State-Level data visualization options. User can select which data values to consider then the program will create a heatmap overlay. The second tab is the County-Level data visualization options. Here the user can click on the county shapes on the map. A list of selected counties will be displayed. From there, the user can select one of the button choices available such as displaying the county stats, displaying the markers that make up the individual data entries, along with clearing the selection. The third tab is the county index. Here, a list of counties is available. Each county on the list can be clicked and selected. Multi select is supported with the user holding command or the control key on Mac/Windows keyboard respectively.

3.2.2. CSS

This file styles with HTML file. Most of the CSS file is deciding how the tab will look.

3.2.3. JavaScript & Google Maps API

The Map - Maps Class

The google.maps.Map class is the main project. It is the Google Map itself and its map bounds are restricted to California.

Coordinates - LatLng Class

The google.maps.LatLng class is used to store the latitude and longitude of a data point. The location values are pulled from the housing dataset.

The Event system - click, mouseover, etc.

The google.maps.event namespace was used to make the map more interactive.

Data Points - Markers, Info Window & MarkerClusterer.js

Google Markers were used to display individual data point locations. Each data point corresponds to a row in the housing dataset. Info Window was used to display individual data point values.

To condense the huge amount of data point markers so that Google Maps won't be flooded with clutter, I used a library called MarkerClusterer. MarkerClusterer is a JavaScript library that creates and manages per-zoom-level clusters for large amounts of markers. The library is community supported. The library is well supported, "We're comfortable enough with the stability and features of the library that we want you to build real production applications on it."

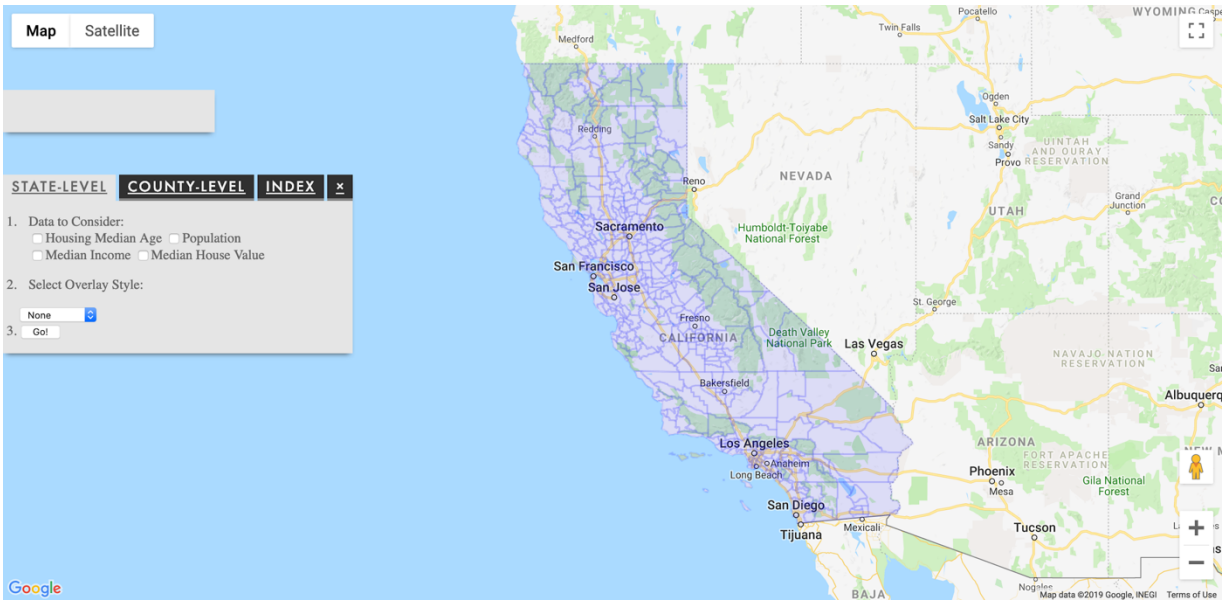
County Shapes - Polygon Class

The google.maps.Polygon class was used to create the shape of each county. Each individual data point is checked via this polygon class if that coordinates are contained in this polygon.

Heatmaps - Heatmap Layer Class

The google.maps.visualization.HeatmapLayer class is used to display the data values set by the user. Each data point entry for a heatmap is weighted by its relevant data value.

4. Results



The result is a web application that runs on the web browser. It was tested and developed with Firefox.

Please see User Documentation for instructions.
Please see Appendix for more result pictures.

5. Conclusion

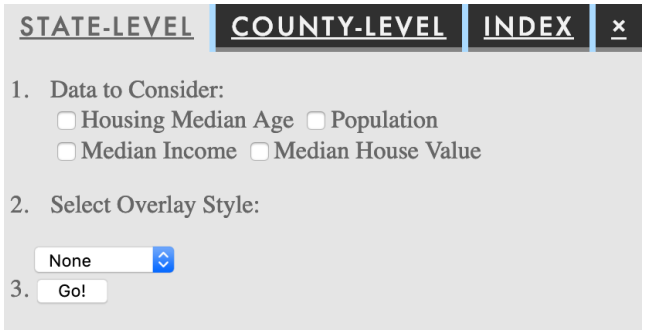
Top major cities such as LA or SF in 1990 are still major cities today so although the housing dataset is from the 1990 California census, it gives a rough idea of what it still can look like today. Major cities continue to have to have the most population, highest house values, and continue to struggle with average income for the housing market. One huge challenge was trying to work with huge sets of data. Simple quick edits to data files were not realistic. For example, to assume the dataset was ready to work immediately was a risk as I later found some data values were missing thus causing unexpected errors. Due to huge data files, I was not able to look through the file manually as it were there were 20,000+ entries so a post-processing approach was need. Also as an unexpected end, thanks to HTML and JS running entirely in the web browser, this program has no real issues running on mobile web browser and therefore can be used as a mobile web application as well. To improve the mobile experience, I added an open/close button for the sidebar as screen size is limited.

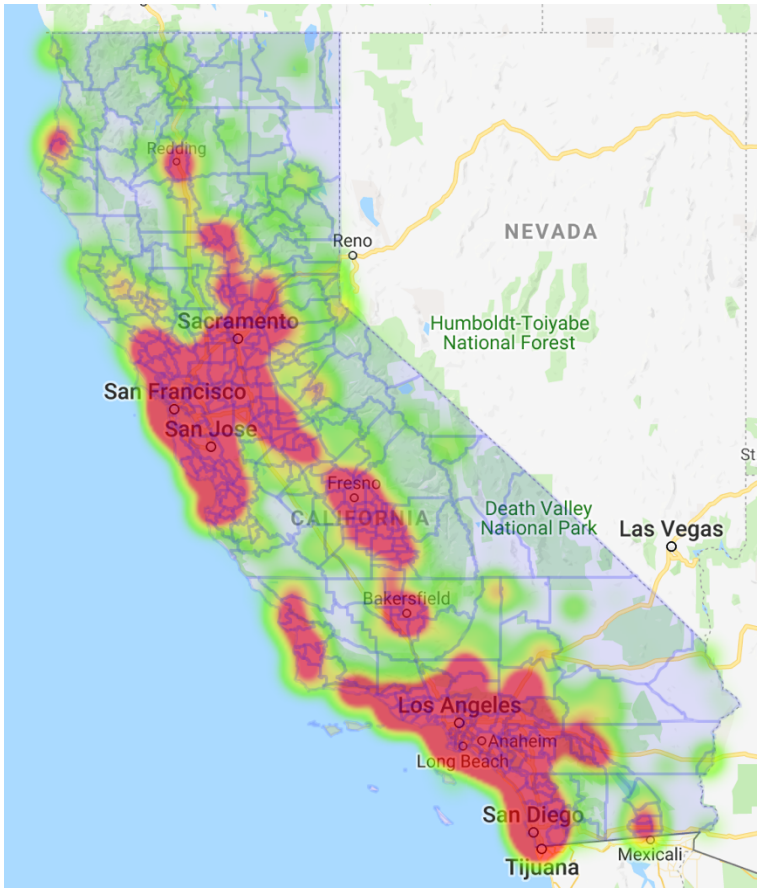
6. References

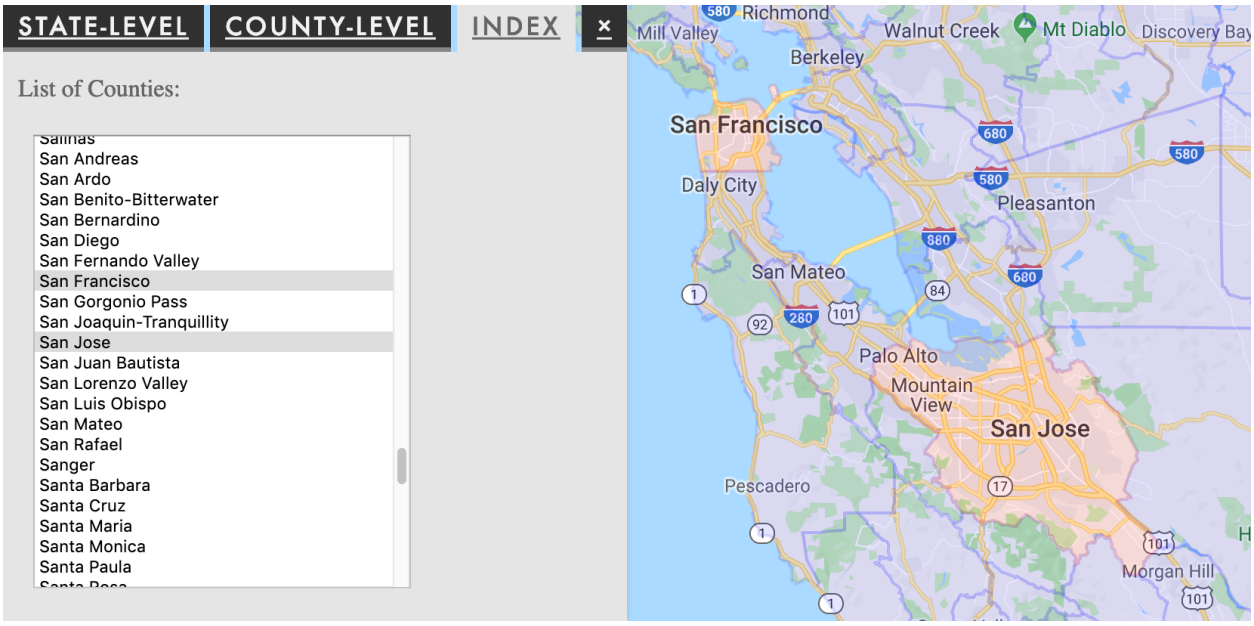
- 1990 California Census Housing Data
  - https://www.kaggle.com/camnugent/california-housing-prices
- 2017 US Census Bureau Shapefile Data for California County Boundary
  - Data
    - https://github.com/uscensusbureau/citysdk/blob/master/v2/GeoJSON/500k/2017/06/county-subdivision.json
  - Documentation
    - https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2017/TGRSHP2017_TechDoc.pdf
- Google Maps API
  - Documentation
    - https://developers.google.com/maps/documentation/javascript/reference/
  - MarkerClusterer for Google Maps
    - https://github.com/googlemaps/v3-utility-library/tree/master/markerclusterer

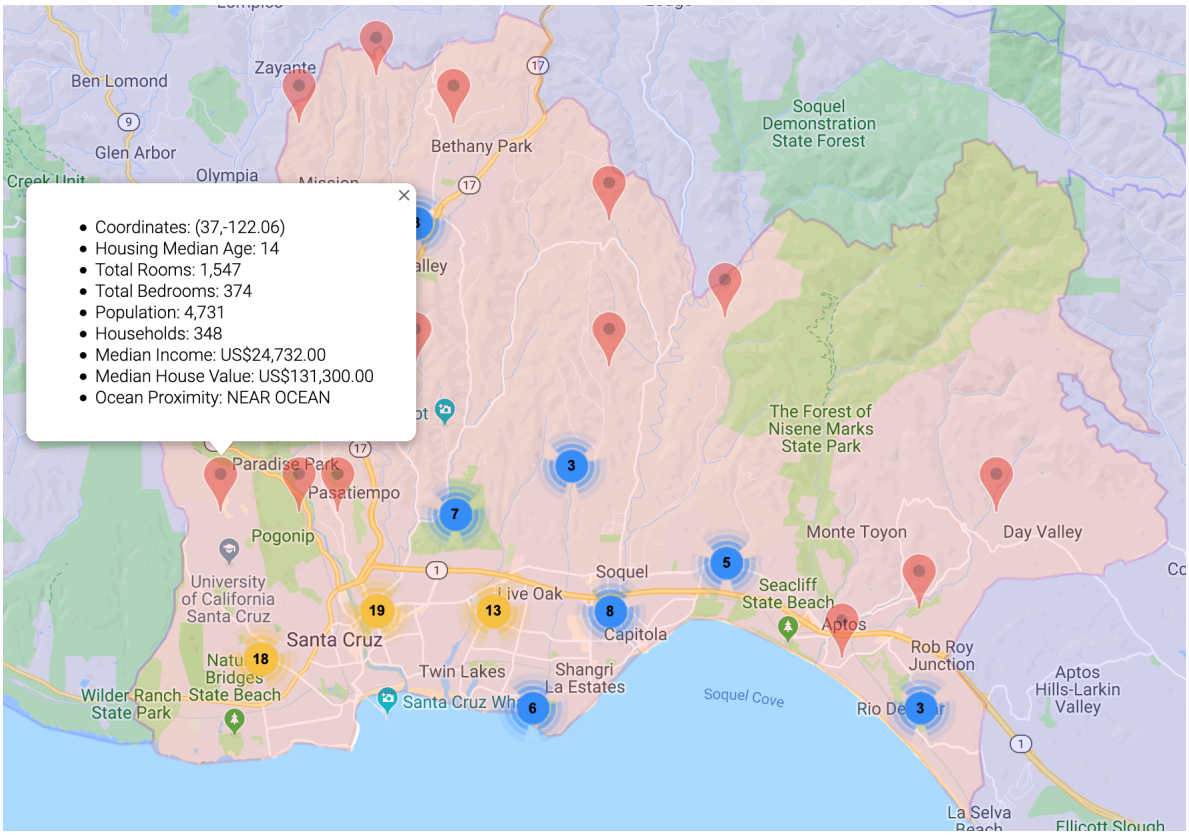
7. Appendix

Result Pictures:









University of California, Santa Cruz