Yishan Wang F20DV Lab 4

8 April 2022 demonstrated to

Shuangjiang Xue

All the exercise can be viewed on this link.

https://jamesw99.github.io/F20DV/lab4/index.html

In lab4, we analysed house values in California.

The dataset was obtained from the following repo: https://github.com/ageron/handson-ml

Additionally, I am utilising maps and data from the following repo:
https://github.com/ropensci/geojsonio

This dataset provides the following feature.

Longitude, latitude, housing_median_age, total_rooms, total_bedrooms, population, households, median_income, median_house_value, ocean_proximity

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY

The entire dataset is separated into five regions based on their proximity to the sea: 1H OCEAN, INLAND, NEAR OCEAN, NEAR BAY, and ISLAND. This means that the dataset does not include the names of the cities in which the communities are located; instead, only the towns' approximate locations are given. This is ambiguous.

We began by analysing the complete dataset.

The third row's pie chart (population with ocean proximity) depicts the population distribution for the entire state of California.

Additionally, when you mouse over any region of the pie chart, the chart displays the population in that area.
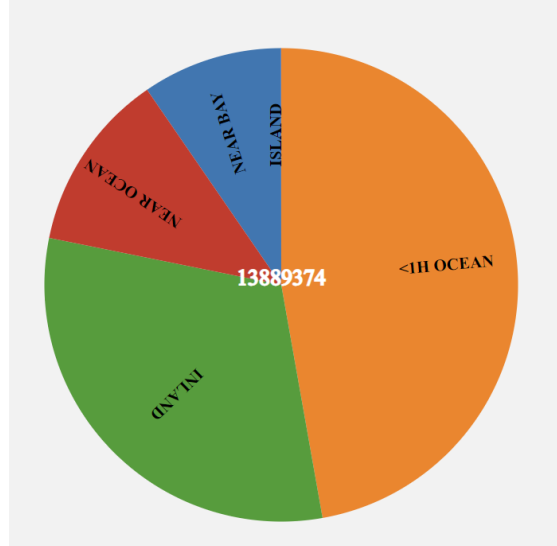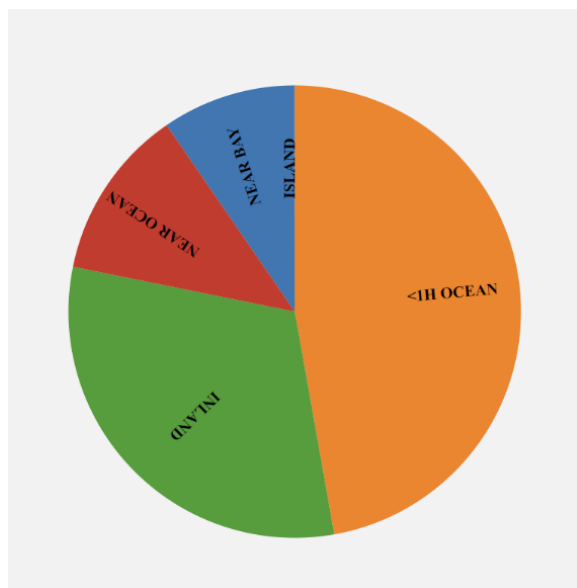
The association between average income and latitude was then visualised using line graphs (latitude, median income).
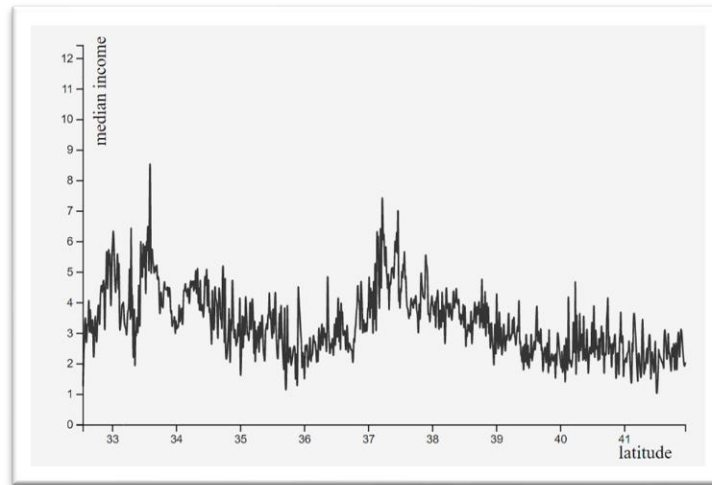
Two significant bumps can be seen in the diagram, as California is long and all of the main cities are concentrated along the coast. They were identified as San Francisco and Los Angeles after being compared to the map.

Although the dataset does not include information on housing prices in individual cities. However, the dataset includes latitude and longitude, which we can use in conjunction with the map to determine the position of the dots. We used california.geojson to produce a map of California, identifying neighbourhoods based on the latitude and longitude of each data entry in housing.csv.
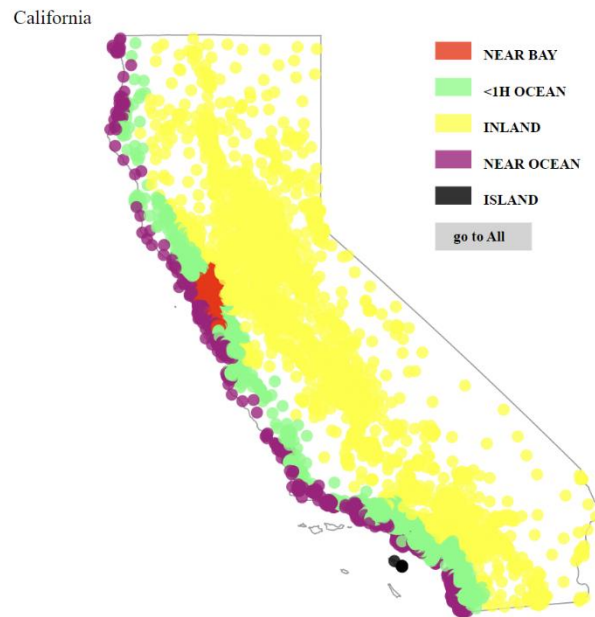
We constructed a circle for each entry, whose position is determined by the longitude and latitude, and whose colour is determined by the house's value, which ranges from "rgb(255, 151, 119)" to "rgb(126, 10, 16).
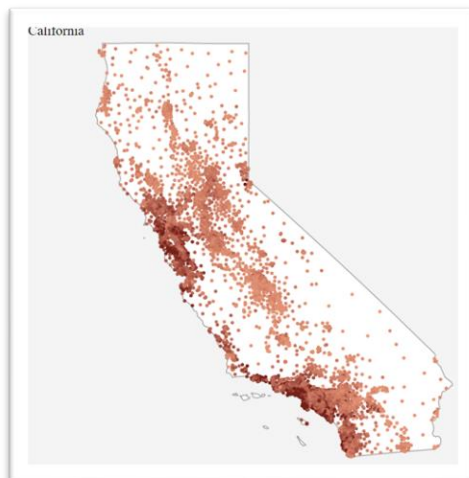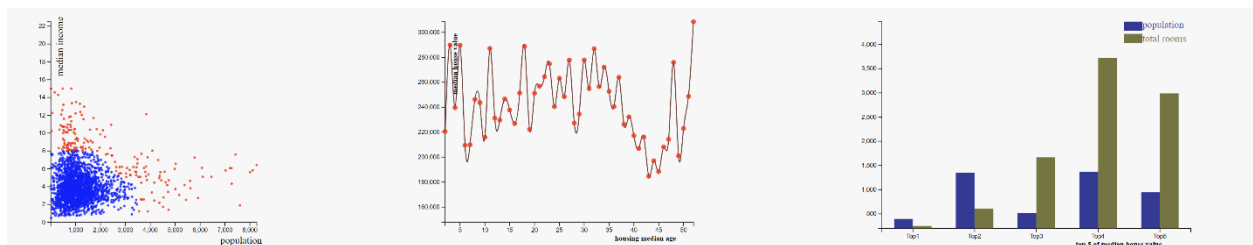
There is no doubt about which locations have higher housing values.

Here is a graph with all of the data.

The following step is to color-code the various areas on the map.

Click on a different area and the chart below the map will be refreshed.

A scatter plot is used in the first graph. In terms of population and average income, we can observe that the vast majority of people live below the 3000 mark and earn less than $8.

The second graph is a line graph that illustrates how the growth in the area's housing median age affects the median house value. We can observe that there is virtually no association between median house value and median housing age in the NEAR BAY area. This is not true in all locations.

The third graph depicts the region's current five most costly neighbourhoods, together with their population and room count. As can be seen, the most expensive areas are sparsely populated.

To the right of the image is another map that depicts the house price trend in the specified area. When the price is increased, the colour becomes darker.
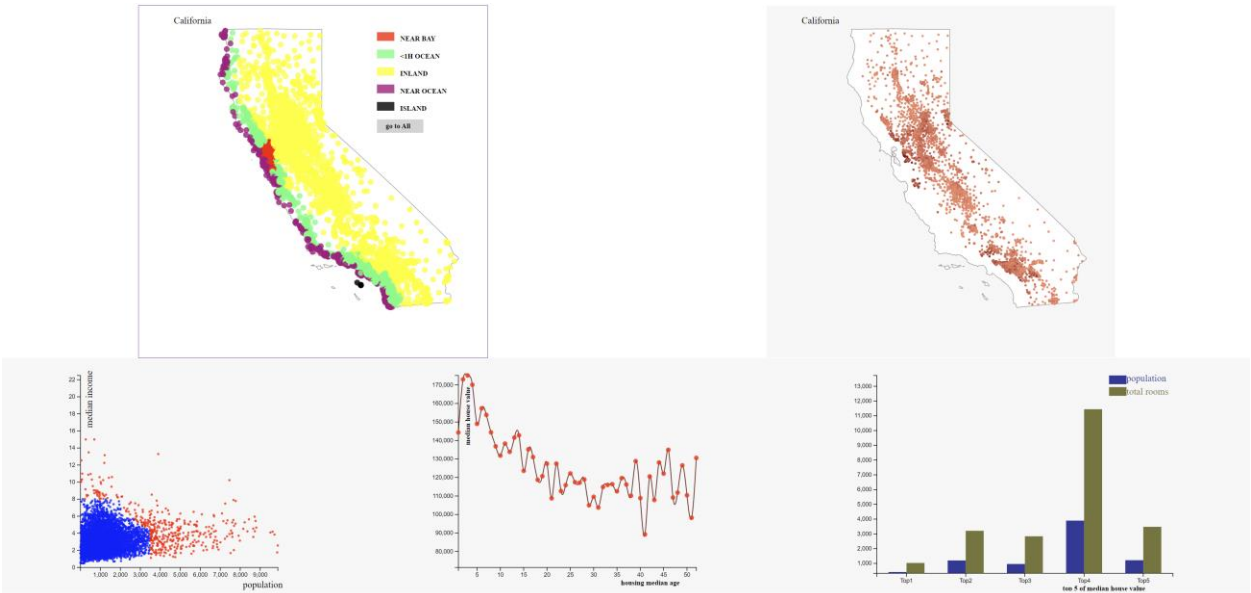

When the page is first opened. The map will display information for the NEAR BAY neighbourhood, but you can click on different regions of the map to display information for other neighbourhoods.
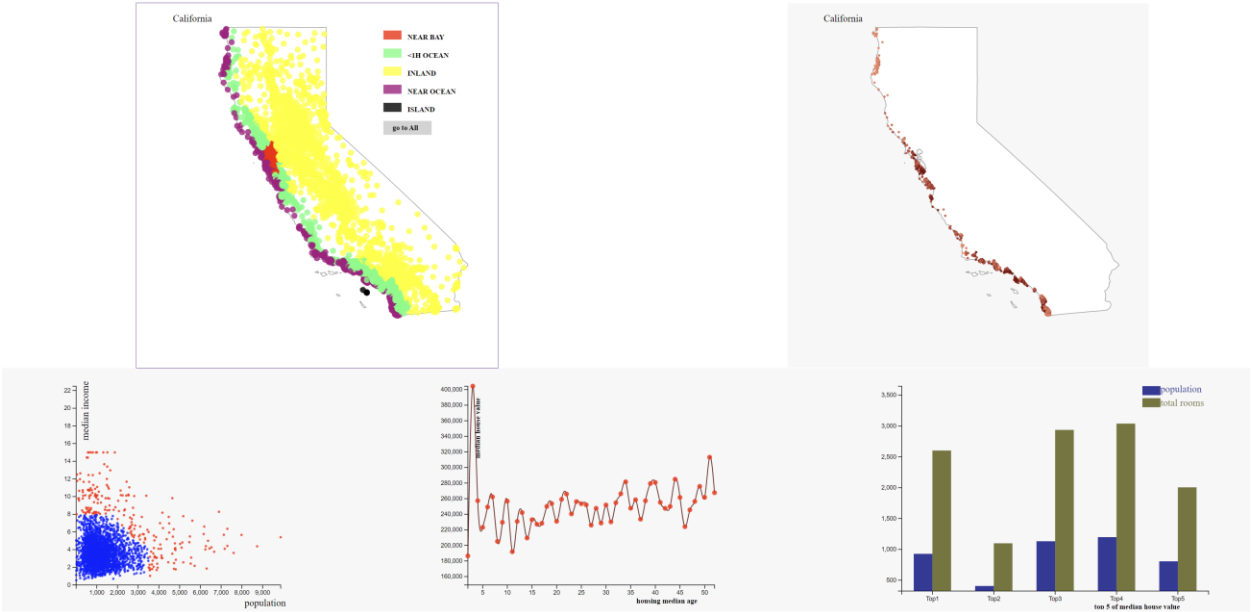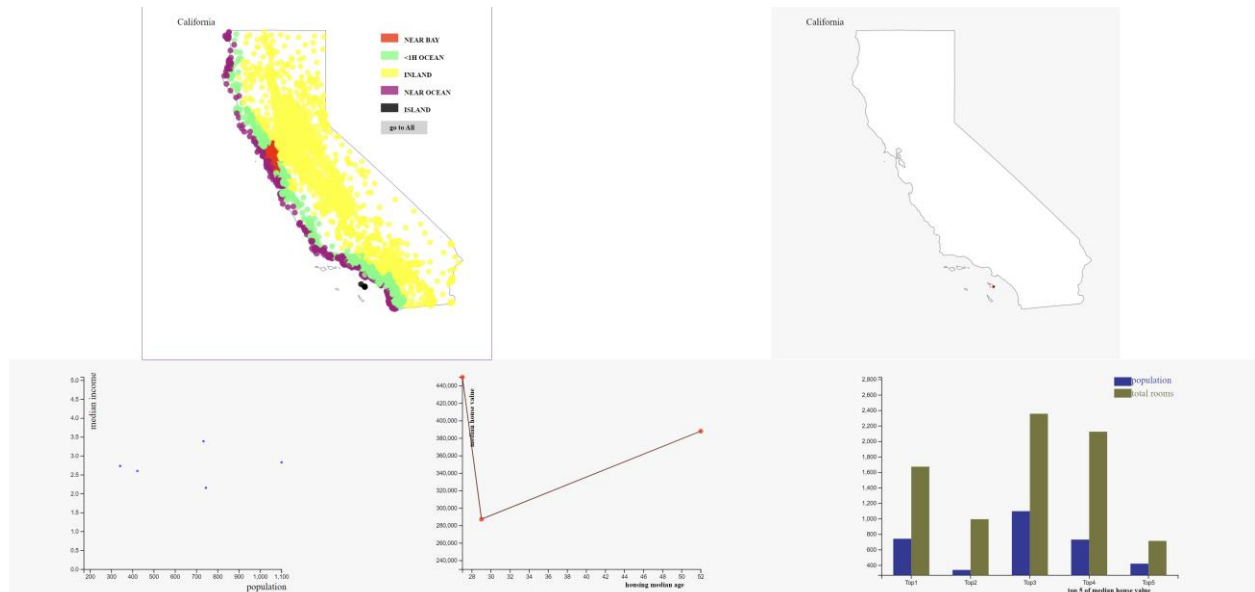
For instance:

1H OCEAN
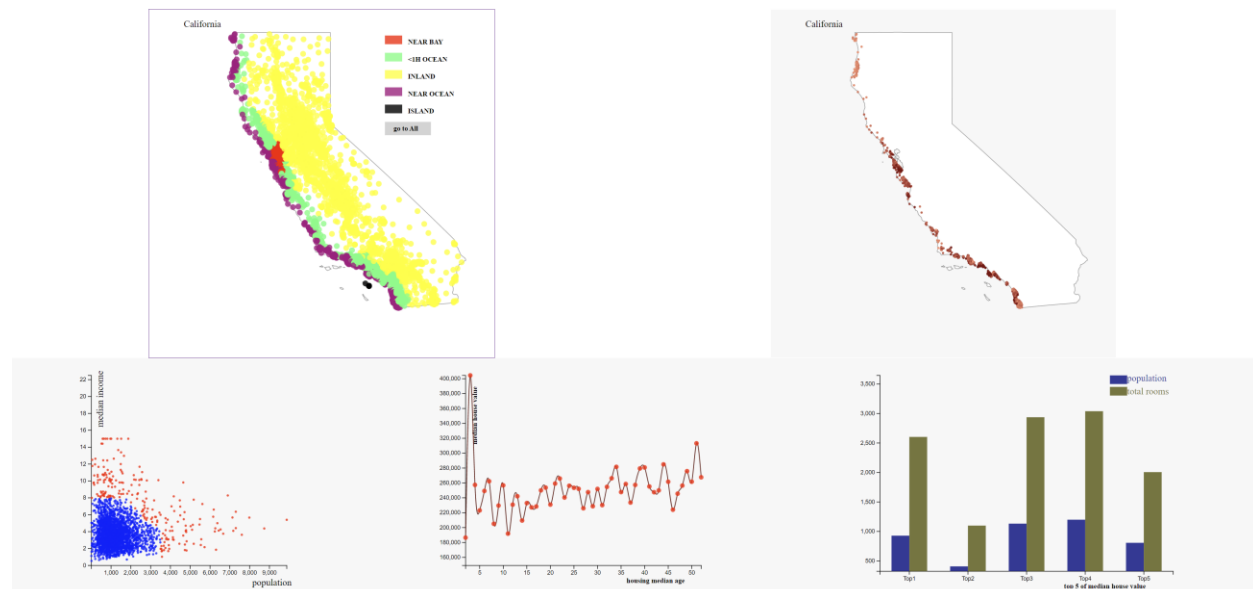
Inland

Near ocean

Island

Unlike the previous tabs, this one has all of the content on a single page. Thus, we can compare the separate data visually.