Problemset 3

Kongpob Lee

2023-01-28

Part III - RMarkdown report and citations

About the data

The data contains information from the 1990 California census in which it indicates basic information about housing in California, including location of the houses, median value of the houses, households income, population, median income, etc (Pace & Barry, 1997).

Question 6

```
## [1] "housing.RDS"
```

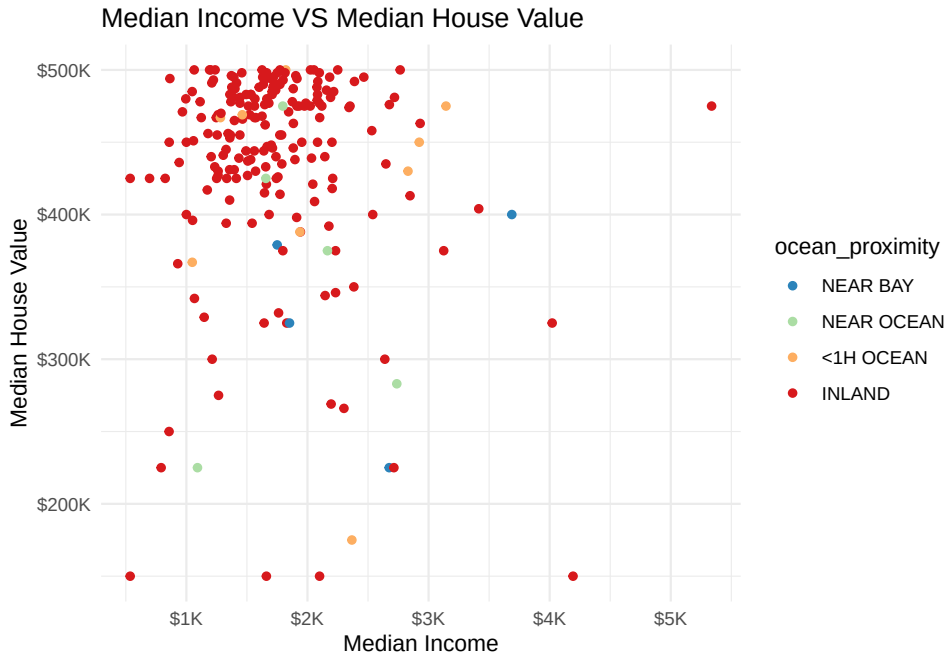
longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
-121.57	39.48	15	202	54	145	40	0.8252	42500	INLAND
-121.55	39.50	26	3215	827	2041	737	1.0585	45100	INLAND
-121.54	39.50	38	1438	310	779	275	1.3289	39400	INLAND
-121.55	39.48	41	461	107	284	90	2.2045	41800	INLAND
-121.55	39.44	31	1434	283	811	289	1.7727	49000	INLAND
-122.37	37.93	45	3150	756	1798	749	1.7500	37900	NEAR BAY
-122.32	37.93	33	296	73	216	63	2.6750	22500	NEAR BAY
-119.81	36.73	47	1314	416	1155	326	1.3720	49600	INLAND
-119.81	36.74	36	607	155	483	146	1.5625	47500	INLAND
-119.79	36.73	52	112	28	193	40	1.9750	47500	INLAND

Part IV - Creating and customizing plots

Data analysis

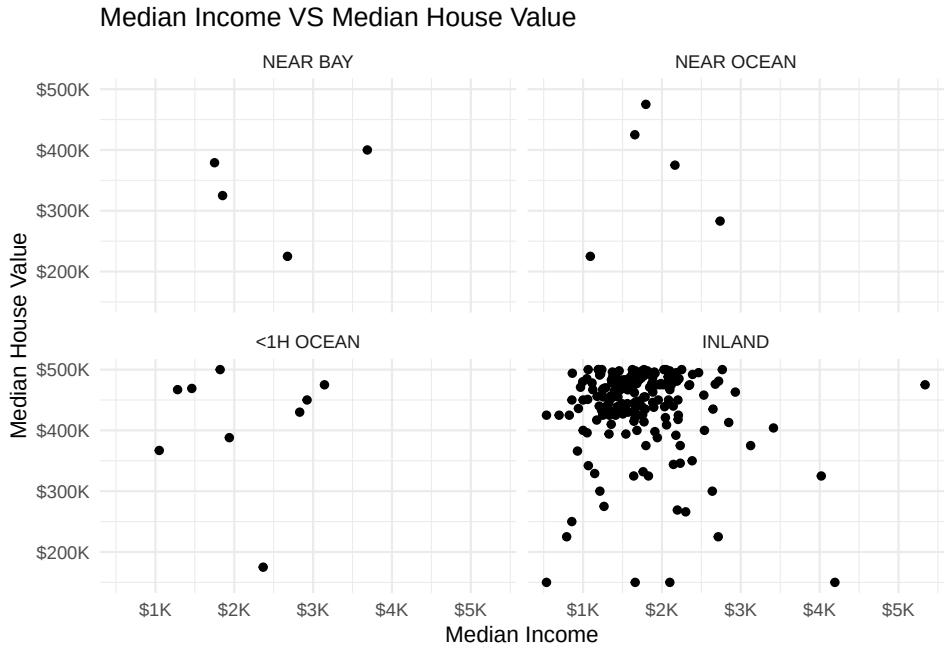
Question 1

Observation: the data indicates that most of the houses with high media house value were located in land (not close the ocean), this is a little bit different from today because houses near ocean are (probably) more expensive than houses located in land.



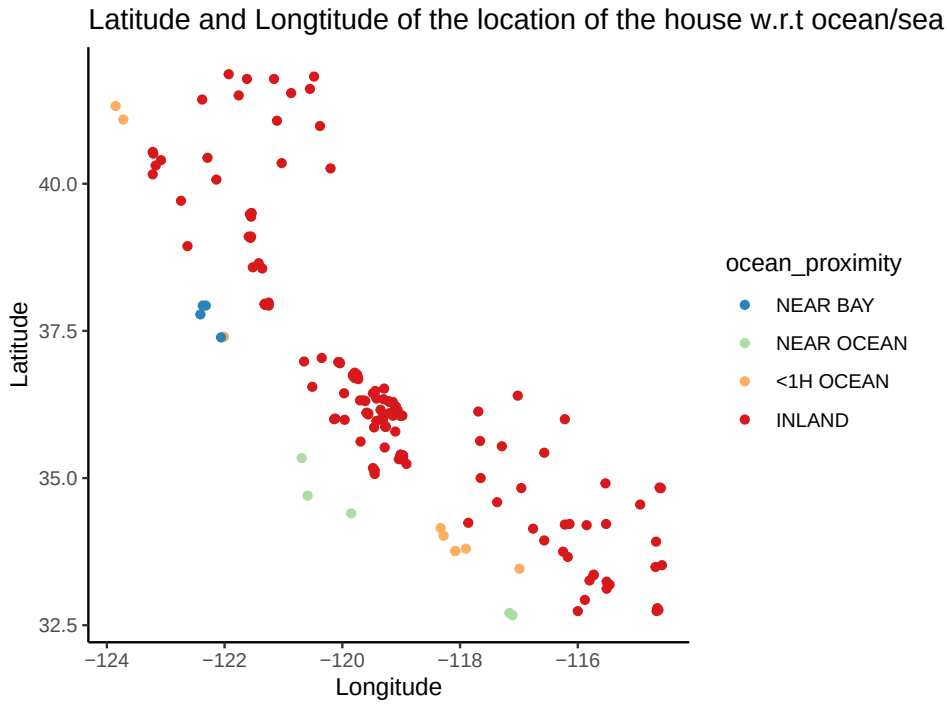
Question 2

Observation: When plotting data separately based on ocean_proximity – we can see that houses located near Bay had even less media house value than houses located near ocean. I assume that living near the ocean or the beach in the past might not be as convenience as today, which explains why the prices were lower in the past.



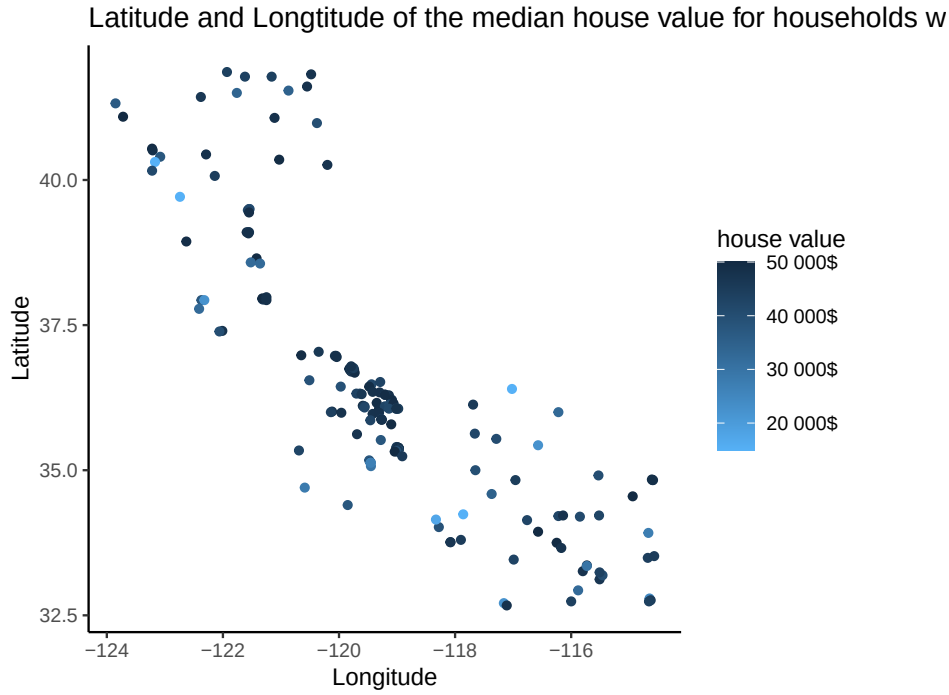
Question 3

Observation: Looking specifically at location of the houses (longitude and latitude), houses located in land are more likely to locate in higher latitude and longitude.



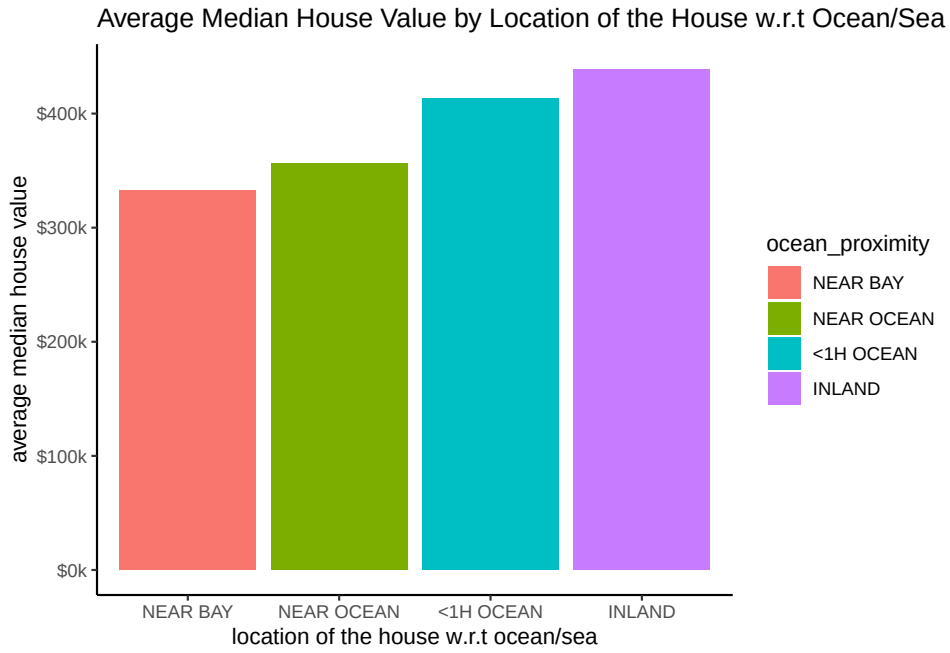
Question 4

Observation: As expected, houses located in land (high longitude and high latitude) had more house value than houses located near the beach or ocean in the past.



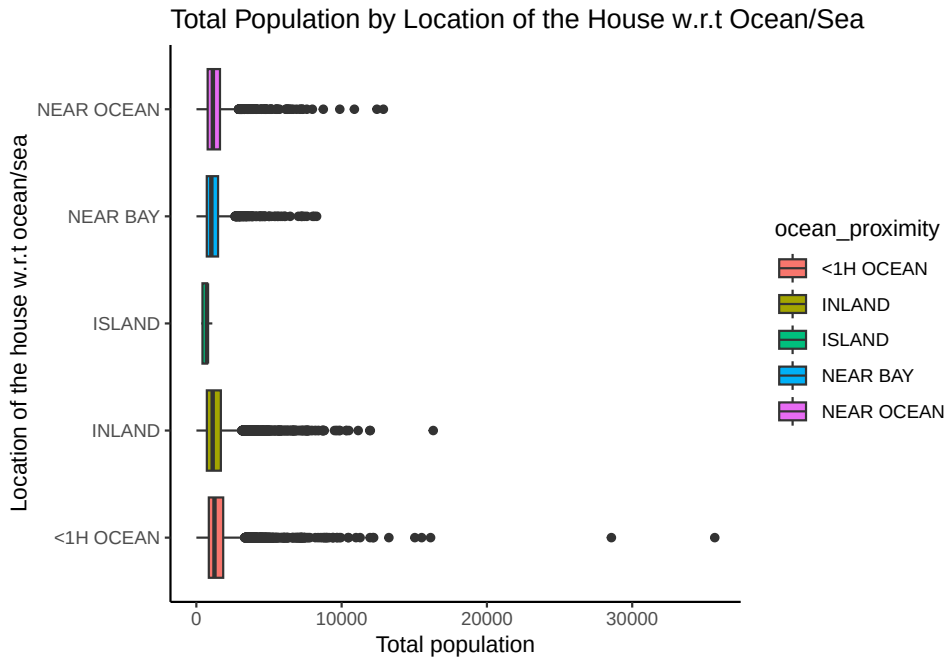
Question 5

Observation: As expected, the average median house value for the houses located in land (far from the ocean) is much higher than the houses located near bay or ocean! It seems like people like to stay away from the ocean (maybe afraid of the natural disaster?)



Part IV - Creating and customizing plots

Question 1 - Box plot



Observation: It seems like most people stayed in the house near the ocean and this is extremely contradicted to what we have assumed from the scatter plot earlier! So, houses with high median value (expensive houses)

were located in far from the ocean; however, most people in California cannot afford to those expensive houses, so most of the people were ended up living near the ocean instead. This might be the result of inequality in California area.

References

Pace, Kelley R., & Barry, R. (1997). Sparse spatial autoregressions. *Statistics and Probability Letters*, *33*(3), 291–297.