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"
```

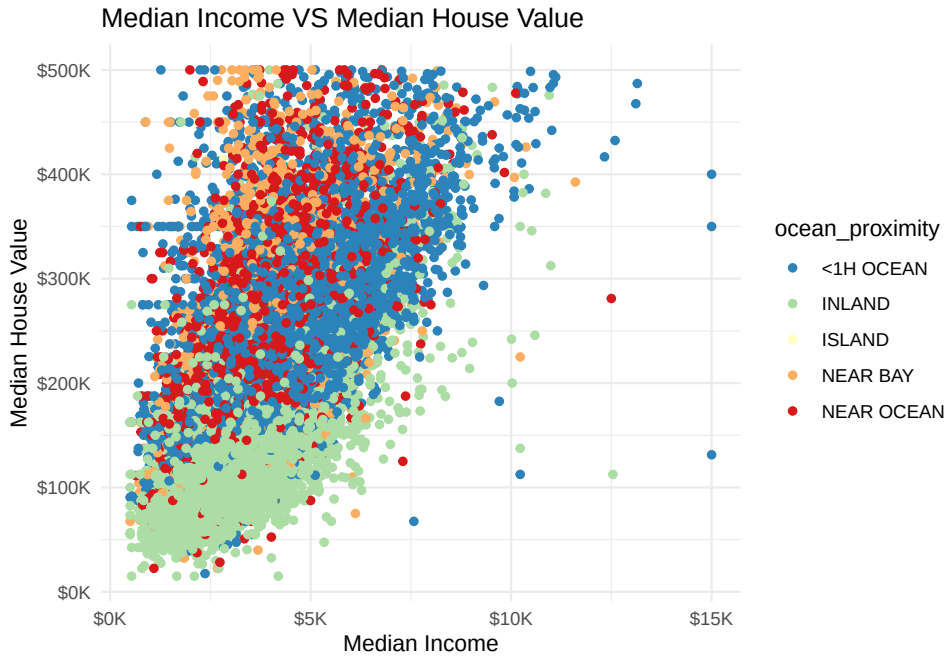
longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
-122.23	37.88	41	880	129	322	126	8.3252	452600	NEAR BAY
-122.22	37.86	21	7099	1106	2401	1138	8.3014	358500	NEAR BAY
-122.24	37.85	52	1467	190	496	177	7.2574	352100	NEAR BAY
-122.25	37.85	52	1274	235	558	219	5.6431	341300	NEAR BAY
-122.25	37.85	52	1627	280	565	259	3.8462	342200	NEAR BAY
-122.25	37.85	52	919	213	413	193	4.0368	269700	NEAR BAY
-122.25	37.84	52	2535	489	1094	514	3.6591	299200	NEAR BAY
-122.25	37.84	52	3104	687	1157	647	3.1200	241400	NEAR BAY
-122.26	37.84	42	2555	665	1206	595	2.0804	226700	NEAR BAY
-122.25	37.84	52	3549	707	1551	714	3.6912	261100	NEAR BAY

Part IV - Creating and customizing plots

Data analysis

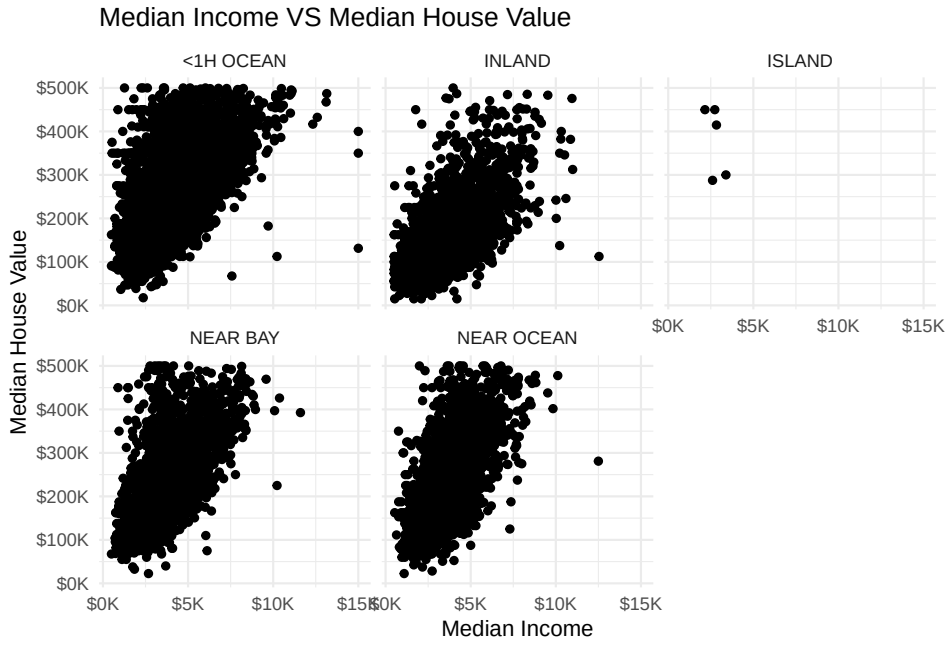
Question 1

Observation: The data suggests that houses located near the ocean are more expensive than houses located far away from the ocean. At you can see from the graph, most of the dots in the bottom of the graph are categorized as “in land”. This is not so surprising to me as I expecte the houses near the ocean to be more expensive!



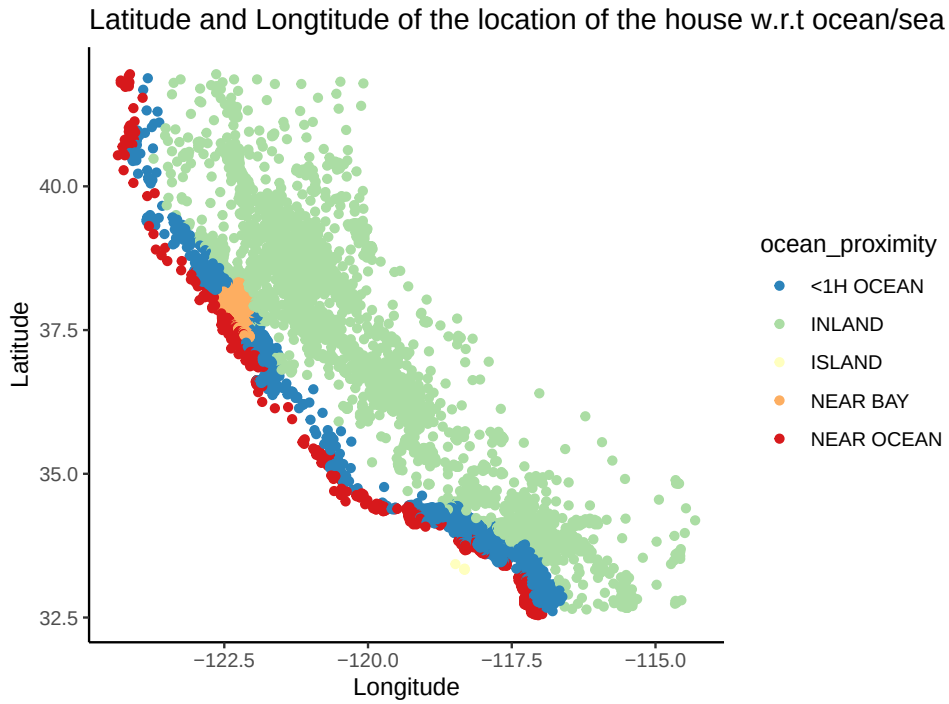
Question 2

Observation: When plotting the data separately based on the ocean proximity, we can clearly see that the houses located near the ocean have more value than houses located far from the ocean. However, it is interesting to me that these expensive houses (located near the ocean) are owned by people with average income (I thought it would be only own by millioniare)



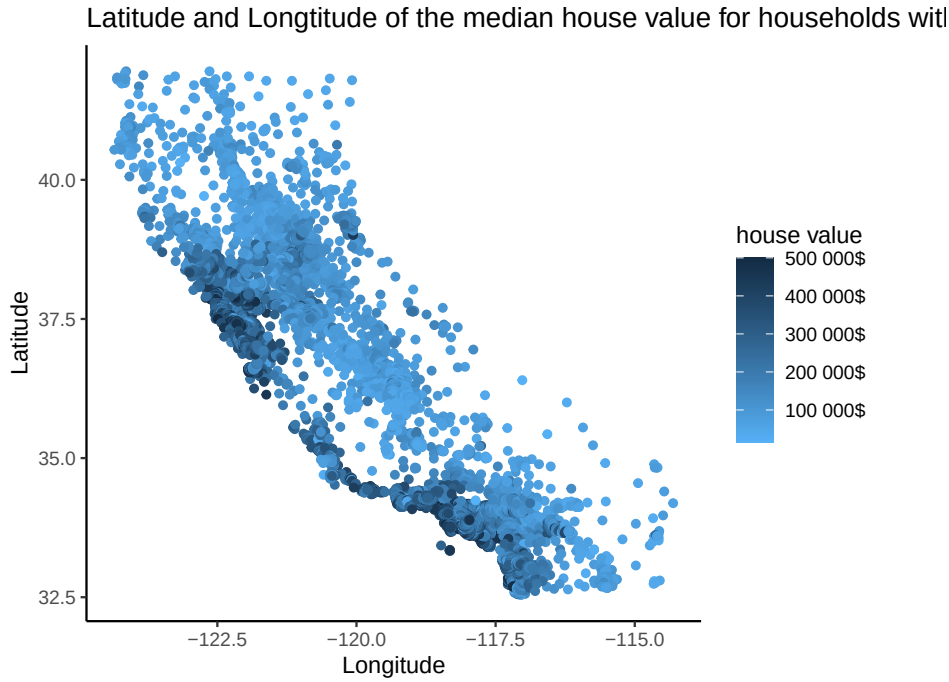
Question 3

Observation: As expected for the longitude and latitude, houses located near the ocean have lower latitude and longitude as its location is lower than houses located far away from the ocean.



Question 4

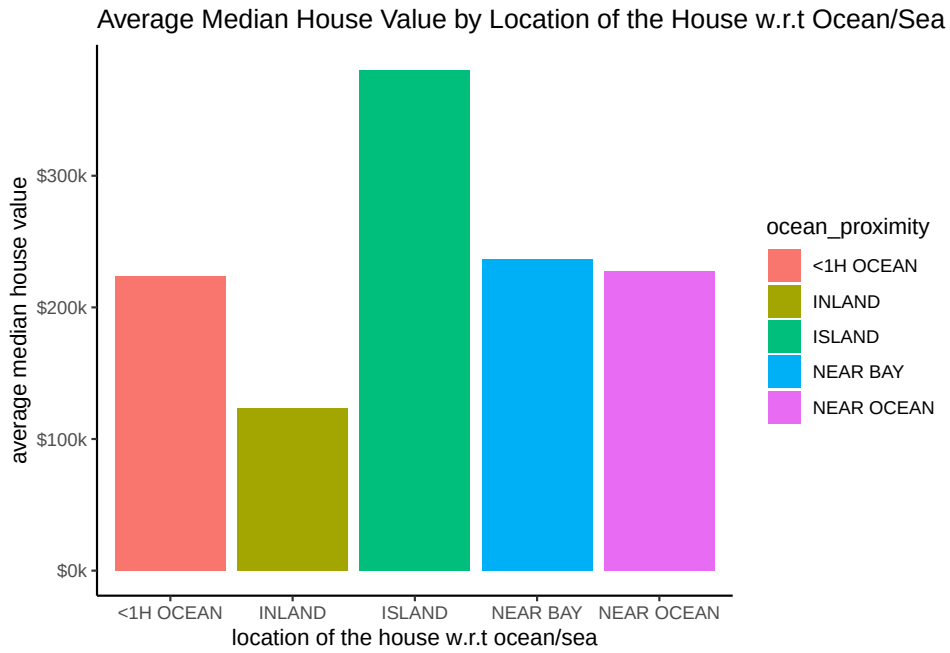
Observation: As expected, houses located near the ocean (with lower latitude and longitude) have median house value than the houses located far away from the ocean.



Question 5

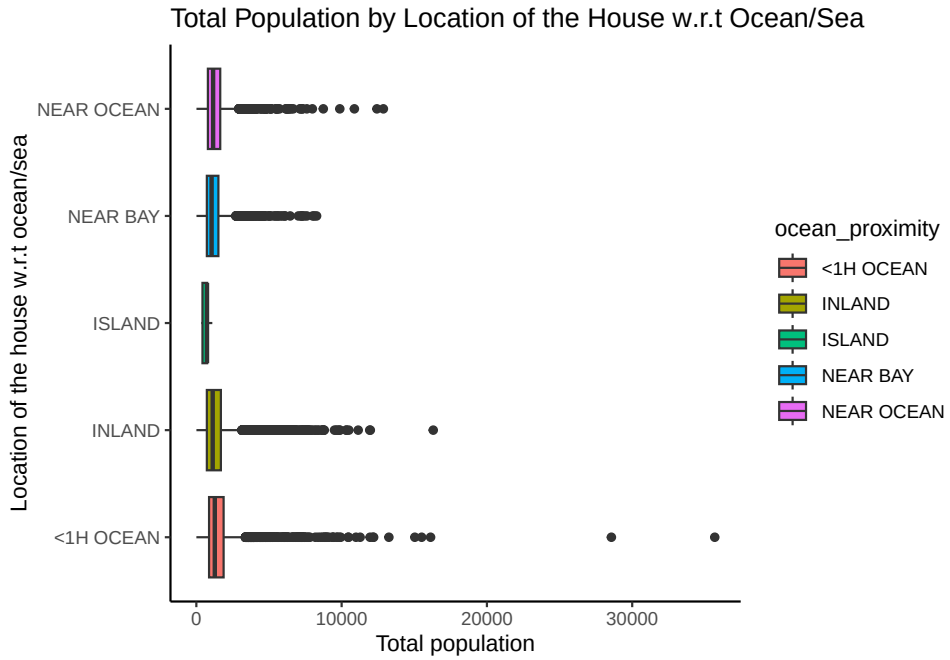
Observation: Surprisingly, the houses located in island are the type of property with highest median house value - these houses must be owned by the millionaire and billionaire (the top 1%). When you look at these

graphs, you can see that houses located near the ocean have more median house value. However, this is the data in the past, which means that these houses' values might have been increased a lot by now and only people with money can afford to live at these houses!



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 I have assumed from the scatter plot earlier! I assumed that most population will be living in the houses located far from the ocean, but the data indicates that total population are higher in the houses located near the ocean. This might mean that people who cannot afford the houses near the ocean in California might move out to other states. However, we might need to do more investigation to find out what is going on with the housing in California!

References

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