Hospital Performance Ratings

Edgar Cortes, Toni Trieu, Nitasha Gounder

Where is the data from?

- Hospital ratings dataset was posted to kaggle from Medicare.gov's hospital care-compare website collected by Centers for Medicare and Medicaid Services. The dataset contains information on the quality of care at over 4,000 Medicare-certified hospitals across the country.
- Income information was pulled from Census data sets at data.census.gov

What Complications did we run into?

- Funky Columns - Complications using CSV from certain sources
- .dtypes - learned a lesson about using dtypes early
- Data type conversions - proper sequence of data cleaning of string to numeric
    - Learned to go back to figure out why our charts were not looking correct
- gmaps learning curve - There is a lot of functionality and possibilities in gmaps

Funky Columns

- Census data has interactive tables that CSV are pulled from. This lead to some interesting initial dataframes.
- We had to play around to get column names usable (admittedly there are easier ways to do what needed to be done but it was fun to explore this process)

	Alabama						
	Households		Families		Married-couple families		Nonfamily households
Label	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate
▼ Total	1,855,184	±11,821	1,214,794	±12,486	879,227	±11,979	640,390
Less than $10,000	9.1%	±0.3	5.1%	±0.4	1.6%	±0.2	17.8%
$10,000 to $14,999	6.0%	±0.3	2.9%	±0.3	1.4%	±0.2	12.2%
$15,000 to $24,999	11.9%	±0.4	8.9%	±0.5	5.0%	±0.4	18.3%
$25,000 to $34,999	10.1%	±0.3	8.9%	±0.4	7.1%	±0.5	12.4%
$35,000 to $49,999	13.0%	±0.4	13.0%	±0.5	11.7%	±0.6	13.2%
$50,000 to $74,999	17.4%	±0.5	18.9%	±0.5	20.2%	±0.7	13.4%
$75,000 to $99,999	12.0%	±0.4	14.9%	±0.5	17.6%	±0.6	6.0%
$100,000 to $149,999	12.1%	±0.4	15.9%	±0.5	20.2%	±0.7	4.4%
$150,000 to $199,999	4.4%	±0.2	6.1%	±0.4	8.0%	±0.5	1.1%
$200,000 or more	4.0%	±0.2	5.4%	±0.3	7.2%	±0.4	1.2%
Median income (dollars)	49,861	±783	63,837	±1,085	78,805	±1,200	26,208
Mean income (dollars)	69,091	±894	82,884	±1,100	98,221	±1,551	40,905
▼ PERCENT ALLOCATED							
Household income in the past 12 months	36.5%	(X)	(X)	(X)	(X)	(X)	(X)
Family income in the past 12 months	(X)	(X)	37.6%	(X)	(X)	(X)	(X)
Nonfamily income in the past 12 months	(X)	(X)	(X)	(X)	(X)	(X)	33.3%

In [3]:
```
raw_income_df.head()
```

Out[3]:

	GEO_ID	NAME	S1901_C01_001E	S1901_C01_001M	S1901_C02_001E	S1901_C02_001M	S1901_C03_001E	S1901_C03_001M	S1901_(
0	id	Geographic Area Name	Estimate!!Households!!Total	Margin of Error!!Households MOE!!Total	Estimate!!Families!!Total	Margin of Error!!Families MOE!!Total	Estimate!!Married-couple families!!Total	Margin of Error!!Married-couple families MOE!!...	Estimate!! household
1	0400000US01	Alabama	1855184	11821	1214794	12486	879227	11979	
2	0400000US02	Alaska	254551	3266	167527	4191	127824	3685	
3	0400000US04	Arizona	2614298	11305	1713006	14051	1247270	13173	
4	0400000US05	Arkansas	1156347	7393	769134	11117	553461	9588	

5 rows × 130 columns

.dtypes

We learned a lesson about using dtypes early to understand what data we are working with. We all launched into our respective tasks after data cleanup and ended up with the same issue, we were trying to perform numeric operations on string data types. It gave us a good lesson in cleaning data and not making assumptions that lead to a team standstill.

```
TypeError: Could not convert Not Available to numeric
```

```
Hospital          object
Overall Rating    object
Address           object
City              object
State             object
Zip Code          object
dtype: object
```

Data type conversions of string to numeric

Once we identified our issue with .dtypes we had issues converting to numeric values. The issue was caused by certain values being “Not Available” as opposed to NAN. When trying to convert using .astype(float) we got the error below, prompting the removal of any rows with “Not Available”

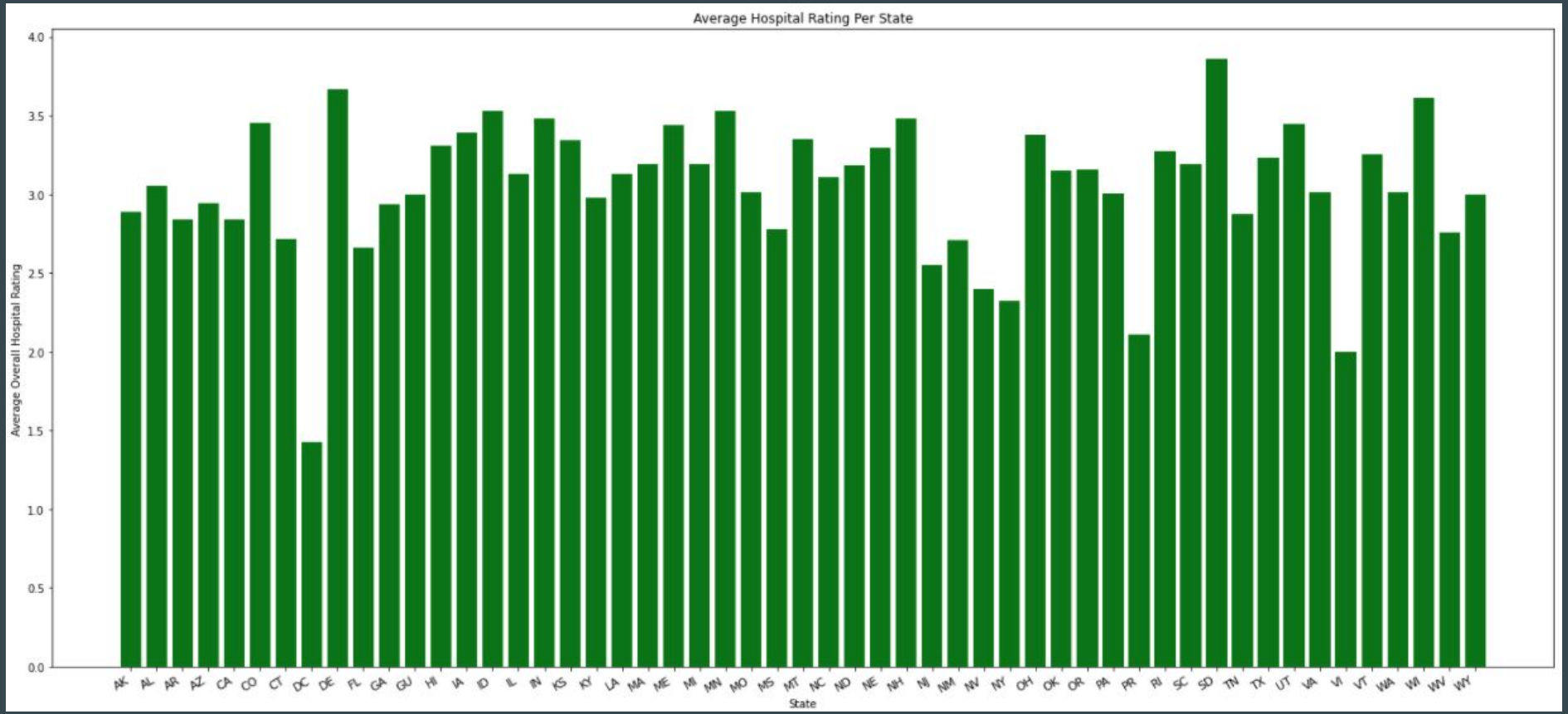
```
ValueError: invalid literal for int() with base 10: 'not'
```

gmaps

- We went over heatmaps in class but there is so much more there to do and learn. We played with infoboxes, markerlayers, different base map figures and more. We could keep going deeper and finding better ways to make the maps more useful and informative.
- One idea that came up after more research (but not enough time to execute) was to use a symbol layer with each rating category having a different colored symbol. This would make it so at a glance anyone can tell what each hospital is rated on the map.

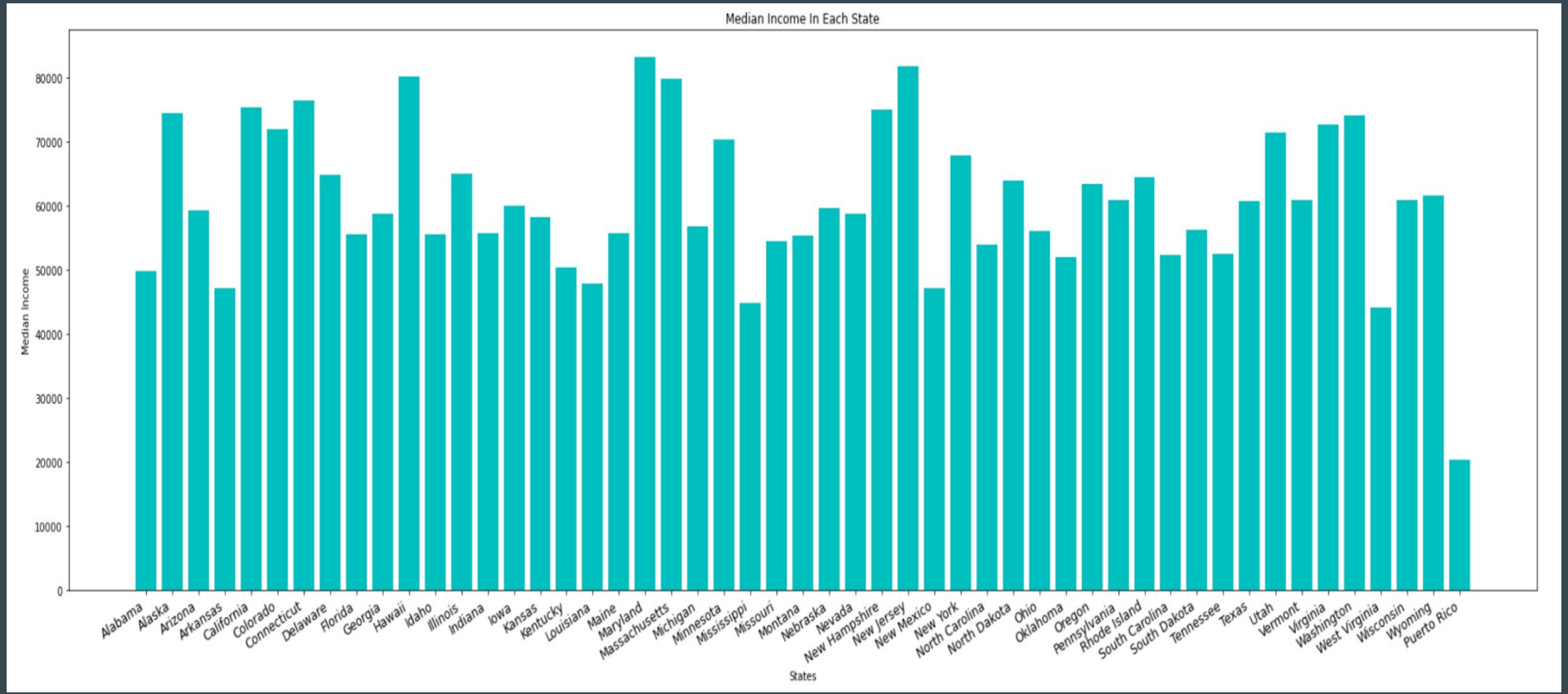
What were our findings?

We found how each States hospitals compare

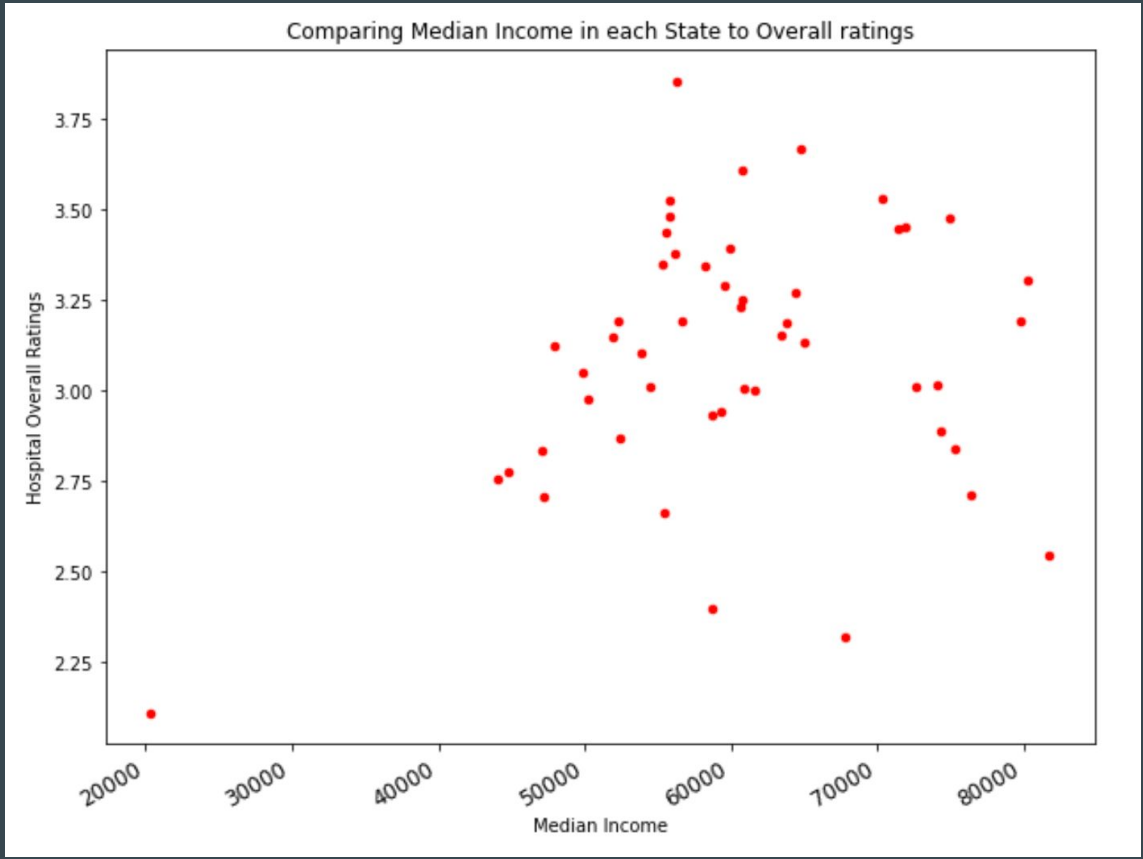


Do States with Higher Income Have better Hospital Ratings?

Median Income

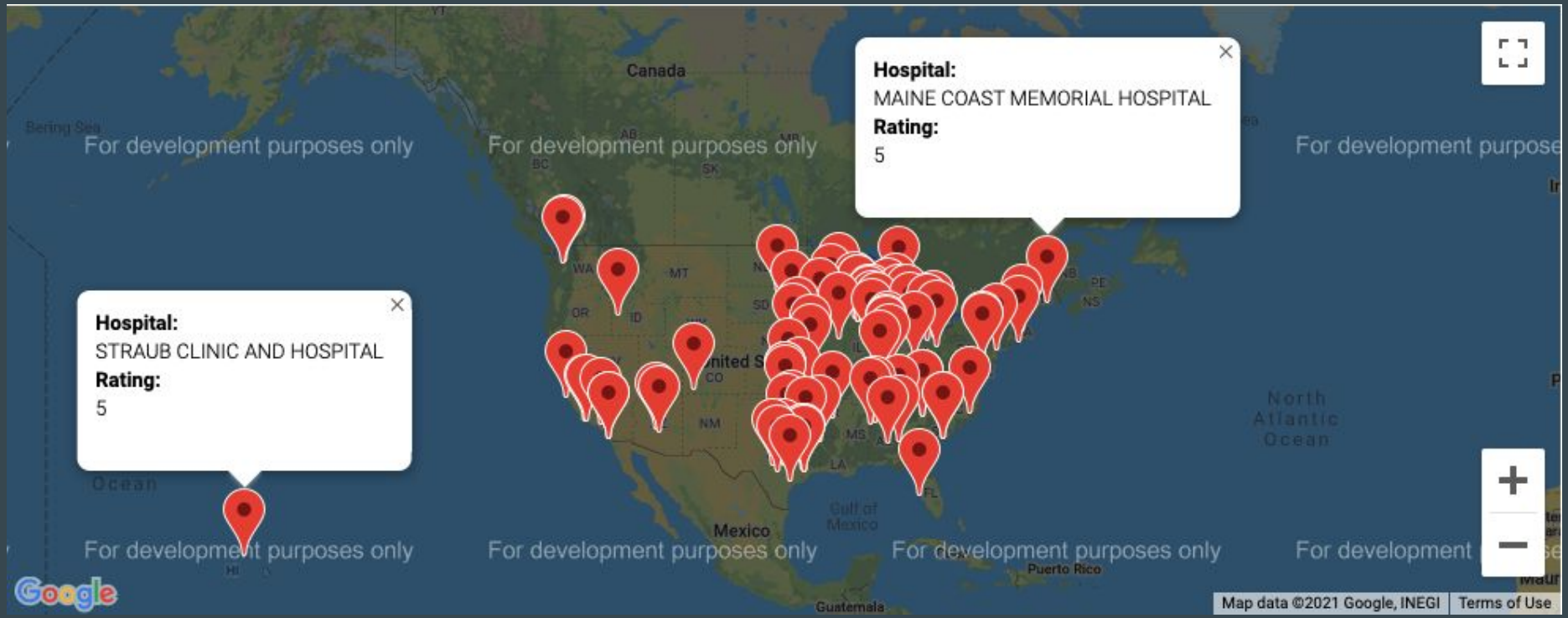


Comparing Median Income to Ratings



We found and visually represented where the best hospitals are located

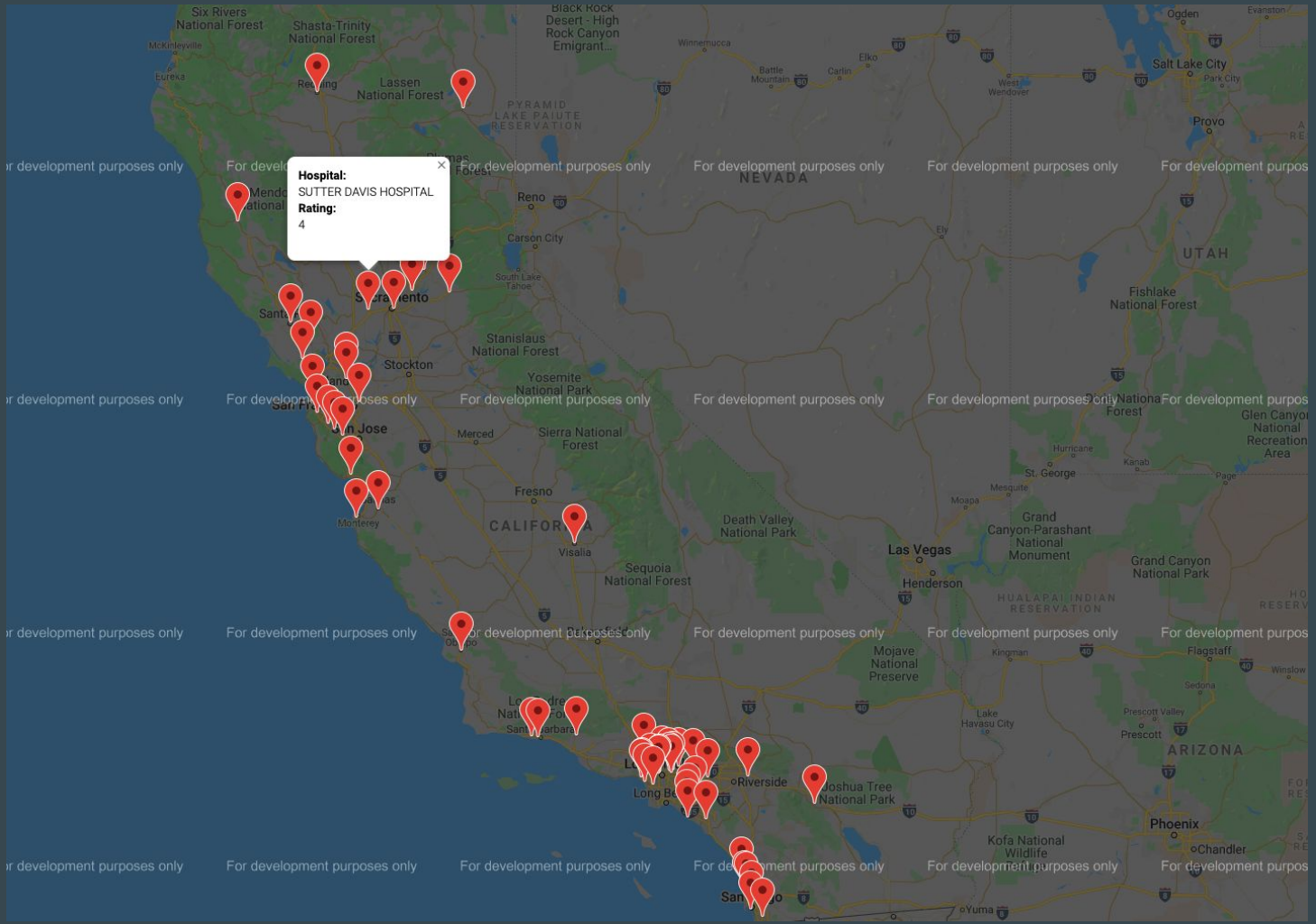
- We used the google maps geocode api to find the coordinates of the best rated hospitals
- Then a gmaps marker layer was used to plot these location on a map of the US with a helpful infobox of hospital information


Hospital:
MAINE COAST MEMORIAL HOSPITAL
Rating:
5
Hospital:
STRAUB CLINIC AND HOSPITAL
Rating:
5
Canada
Mexico
Guatemala
Puerto Rico
Gulf of Mexico
North Atlantic Ocean
For development purposes only
Google
Map data ©2021 Google, INEGI
Terms of Use

So What Factor?

Why does this class care about some of our findings?

1. If you want to move out of state and have health concerns then the graph showing the average hospital rating per state could help you decide where you might want to move
2. Most of us are in California currently, so we can make smart choices about what hospital should be our go to.
3. It will be useful to know a hospital's overall rating so we can make informed decisions about where we seek care. This can prove useful during a pandemic.


Hospital:
SUTTER DAVIS HOSPITAL
Rating:
4
For development purposes only


The End