

# Applications of Machine Learning

## Homework 1

Matthew Ferguson



- 1.) I successfully downloaded the TSV file from [healthdata.gov](https://healthdata.gov) containing COVID 19 patient impact data.
- 2.) I then read the file into excel and converted the TSV file to an XLSX file.
- 3.) Upon inspection of the data, I noticed that there were impossibly large values of “-999999” as well as values for the 7-day average occupied beds of “64459.1” and similarly high values. I filtered out -999999 from my data set and had determined that I could filter out the top remaining 1% of values. I hesitated and made a phone call to Mercy One Hospital in Iowa, the hospital with 64459.1 occupied beds. I talked to a nurse on the phone who said Mercy One was a chain of over 20 hospitals and could conceivably have 64459 beds (She did ask though how one could have 0.1 of a bed amusingly). I was doubly tempted to apply statistical methods and filter out the top 1% of values but after hearing this piece of information from the nurse I decided to keep all values besides -999999. About 2,400,000 instances of -999999 were replaced with empty cells.

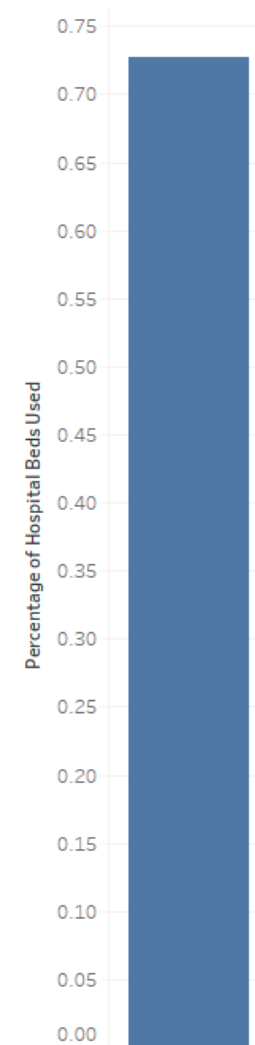
4.) For each column I calculated the set of descriptive statistics you requested keeping the left and rightmost six columns:

Statistics	hospital_pk	collection_week	state	ccn	hospital_name	addresses	previous_week_personnel_covid_vaccinated_doses_administered_7_day	total_personnel_covid_vaccinated_doses_none_7_day	total_personnel_covid_vaccinated_doses_one_7_day	total_personnel_covid_vaccinated_doses_all_7_day	previous_week_patients_covid_vaccinated_doses_one_7_day	previous_week_patients_covid_vaccinated_doses_all_7_day
Mean	267615.9	44316.6	#DIV/0!	267572.3	#DIV/0!	42071.0	82.0	378.9	498.0	795.8	169.4	164.5
Min	10001	44043	0	10001	0	42071	0	0	0	0	0	0
Max	677297	44589	0	677297	0	42071	539270	520037	520037	520037	520037	520037
Range	667296	546	0	667296	0	0	539270	520037	520037	520037	520037	520037
Median	260009	44316	#NUM!	260009	#NUM!	42071	0	83	32	163	0	0
Mode	160080	44190	#N/A	50515	#N/A	42071	0	0	0	0	0	0
Variance	24513913052.9	25375.8	#DIV/0!	24518455790.7	#DIV/0!	0	4002780.9	2817084.3	5128307.9	5956082.8	4366297.8	3655142.4
Std Dev	156569.0	159.3	#DIV/0!	156583.5	#DIV/0!	0	2000.7	1678.4	2264.6	2440.5	2089.6	1911.8
Q1	140172	44176	#NUM!	140167	#NUM!	42071	0	10	0	32	0	0
Q2	260009	44316	#NUM!	260009	#NUM!	42071	0	83	32	163	0	0
Q3	390265	44456	#NUM!	390265	#NUM!	42071	10	292	237	650	6	5

5.) I then proceeded to load the hospital data into tableau.

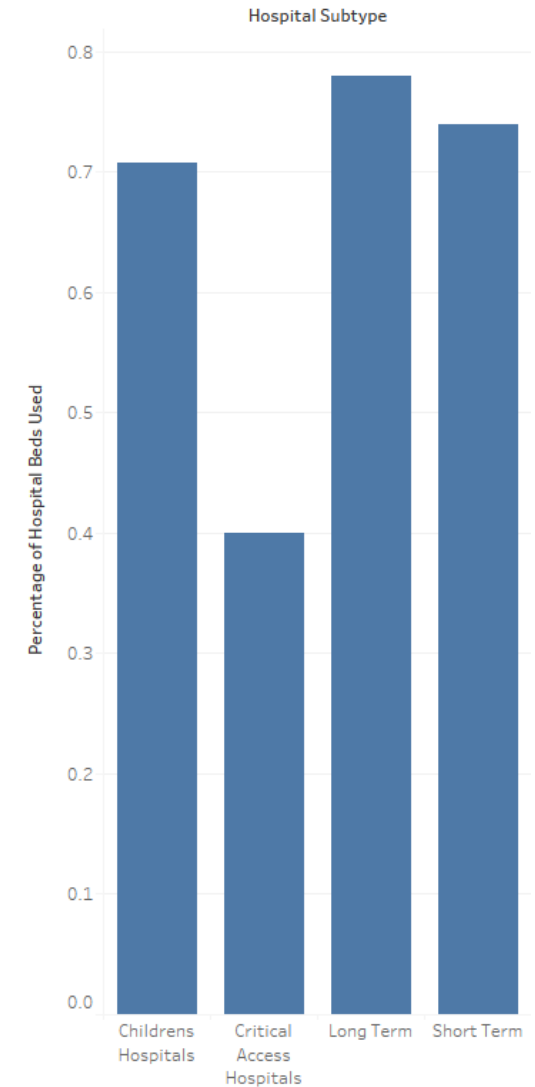
6a.) Here we can see the average percentage of inpatient beds that are occupied both for all hospitals and for hospital subtype. The USA healthcare system was dangerously close to running out of capacity for COVID patients. I think this provides some critical context regarding COVID relief bills passed by Congress. These bills provided enormous sums of money to the healthcare industry and this graphic makes the case that this money was critical in preventing an over capacity situation.

Average Bed Consumption



SUM([Inpatient Beds Used 7 Day Avg])/SUM([Inpatient Beds 7 Da....

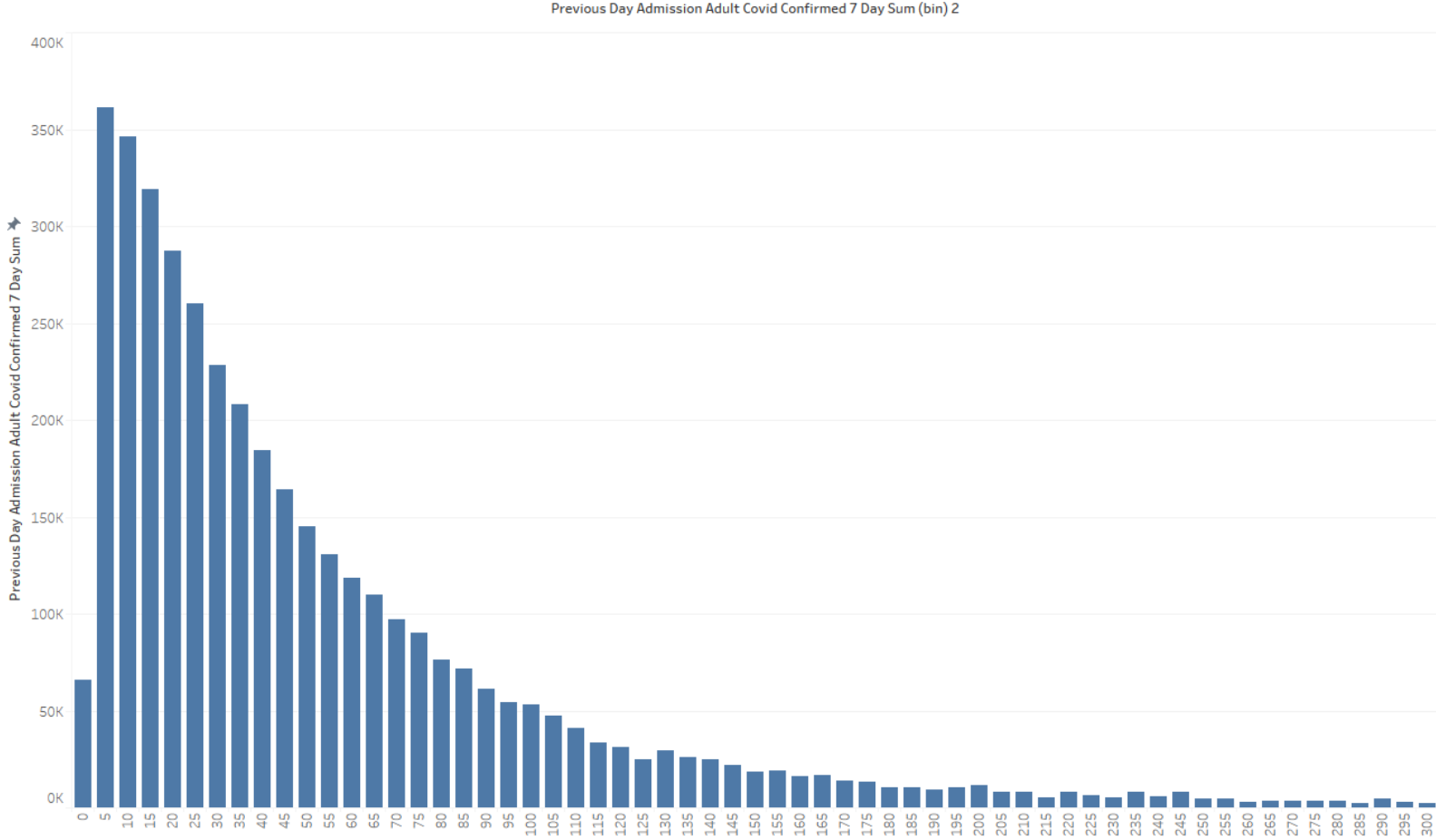
Bed Consumption by Hospital Type



SUM([Inpatient Beds Used 7 Day Avg])/SUM([Inpatient Beds 7 Da... for each Hospital Subtype.

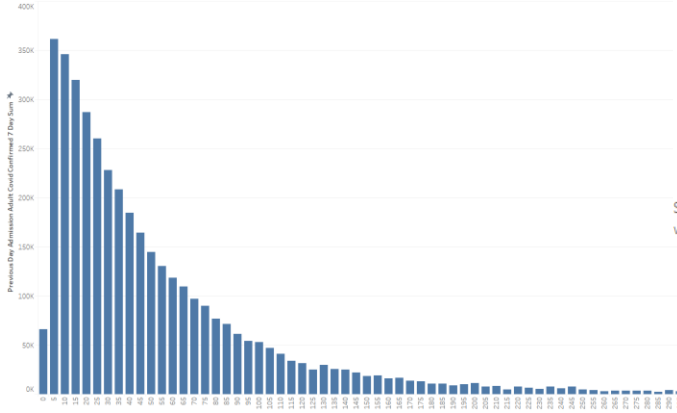
6b.) Using bin size 5 we display a histogram of “Previous Day Admission Adult Covid Confirmed 7 Day Sum”. Here we display the Histogram with (right) and without (bottom) outliers removed.

Prev Day Admission Hist



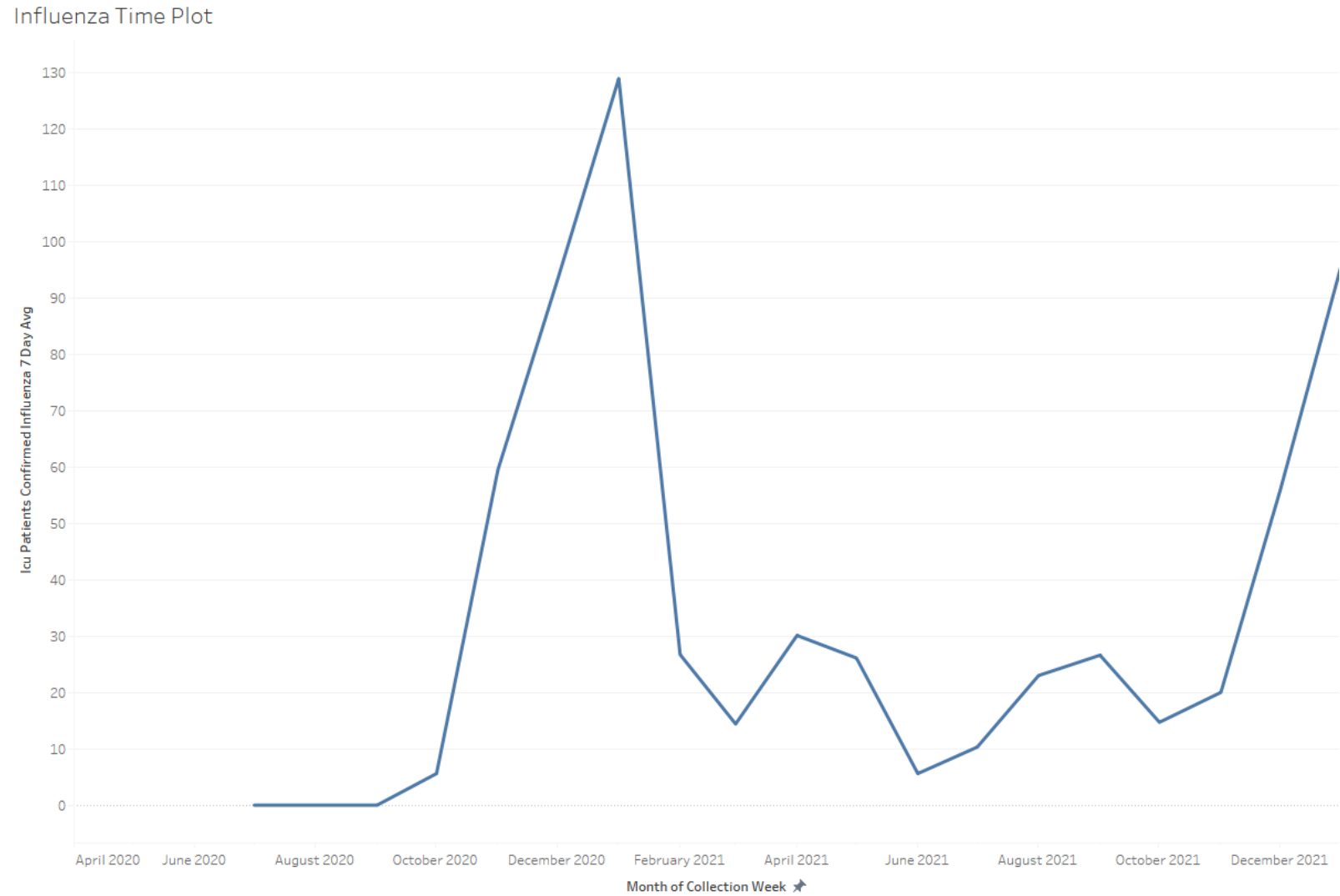
Sum of Previous Day Admission Adult Covid Confirmed 7 Day Sum for each Previous Day Admission Adult Covid Confirmed 7 Day Sum (bin) 2. The view is filtered on Previous Day Admission Adult Covid Confirmed 7 Day Sum (bin) 2, which keeps 61 of 143 members.

Prev Day Admission Hist



Sum of Previous Day Admission Adult Covid Confirmed 7 Day Sum for each Previous Day Admission Adult Covid Confirmed 7 Day Sum (bin) 2. The view is filtered on Previous Day Admission Adult Covid Confirmed 7 Day Sum (bin) 2, which excludes Null and -6.

6c.) Here we can see the time plot of ICU patients with confirmed cases of Influenza as a seven-day average. The period is from April 2020 through December 2021. Note the lack of data for April and June. We can see spikes of Influenza ICU patients in the Winter. This makes sense intuitively as Influenza transmits better in cooler temperatures. My understanding is the water droplets which carry viruses are less likely to evaporate in colder temperatures and thus more likely to infect someone.



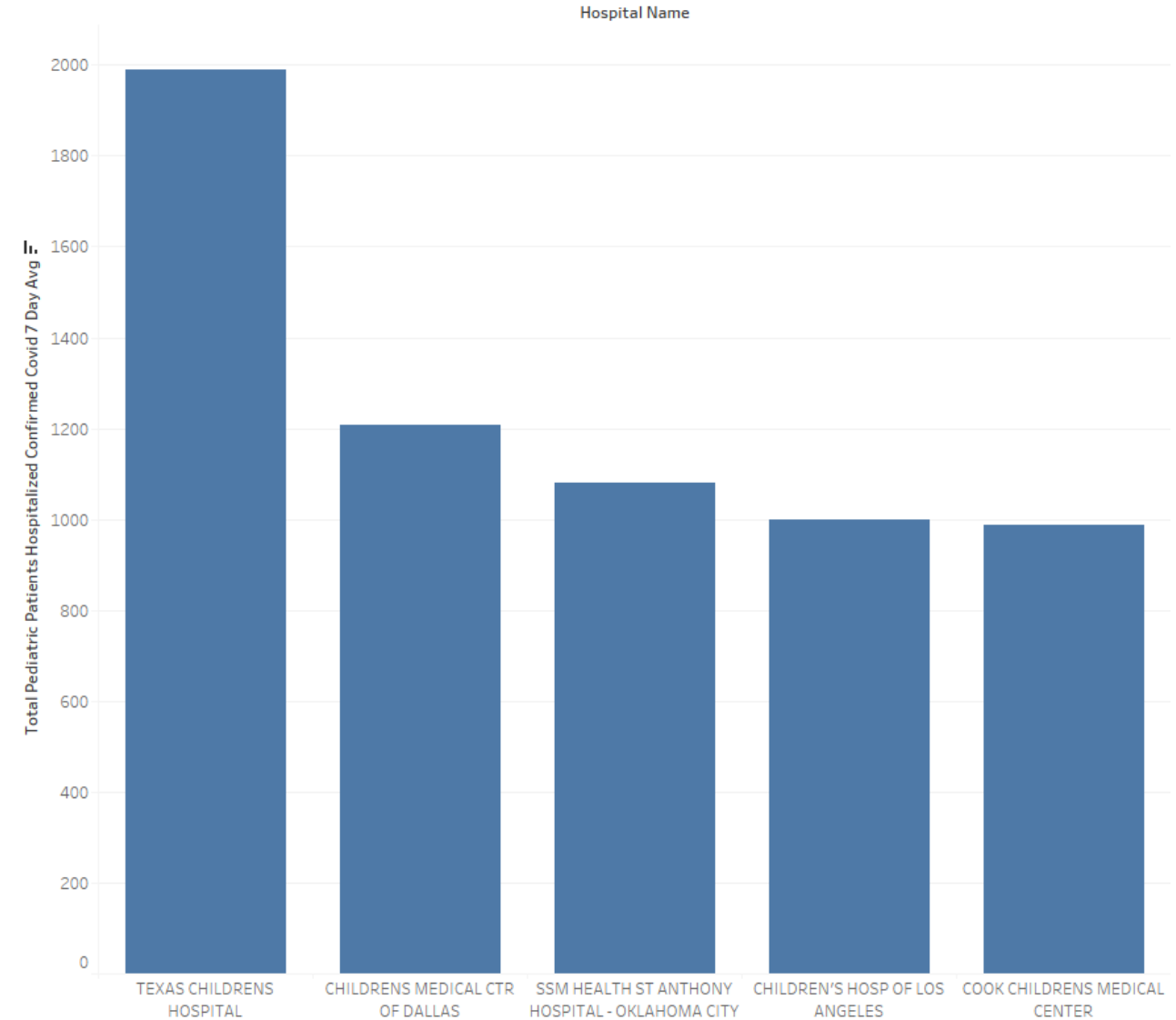
The trend of sum of Icu Patients Confirmed Influenza 7 Day Avg for Collection Week Month.

6d.) This graph identifies the top 5 hospitals by the seven-day average total pediatric patients with confirmed cases COVID 19. These Identified hospitals are:

1. The Texas Children's Hospital
2. Children's Medical Center of Dallas
3. SSM Health St Anthony Hospital of Oklahoma City
4. Children's Hospital of Los Angeles
5. Cook Children's Medical Center

As an Oklahoma native it is disheartening to see that Texas and Oklahoma hospitals lead the pack for high levels of pediatric covid cases.

Top 5 Hospitals



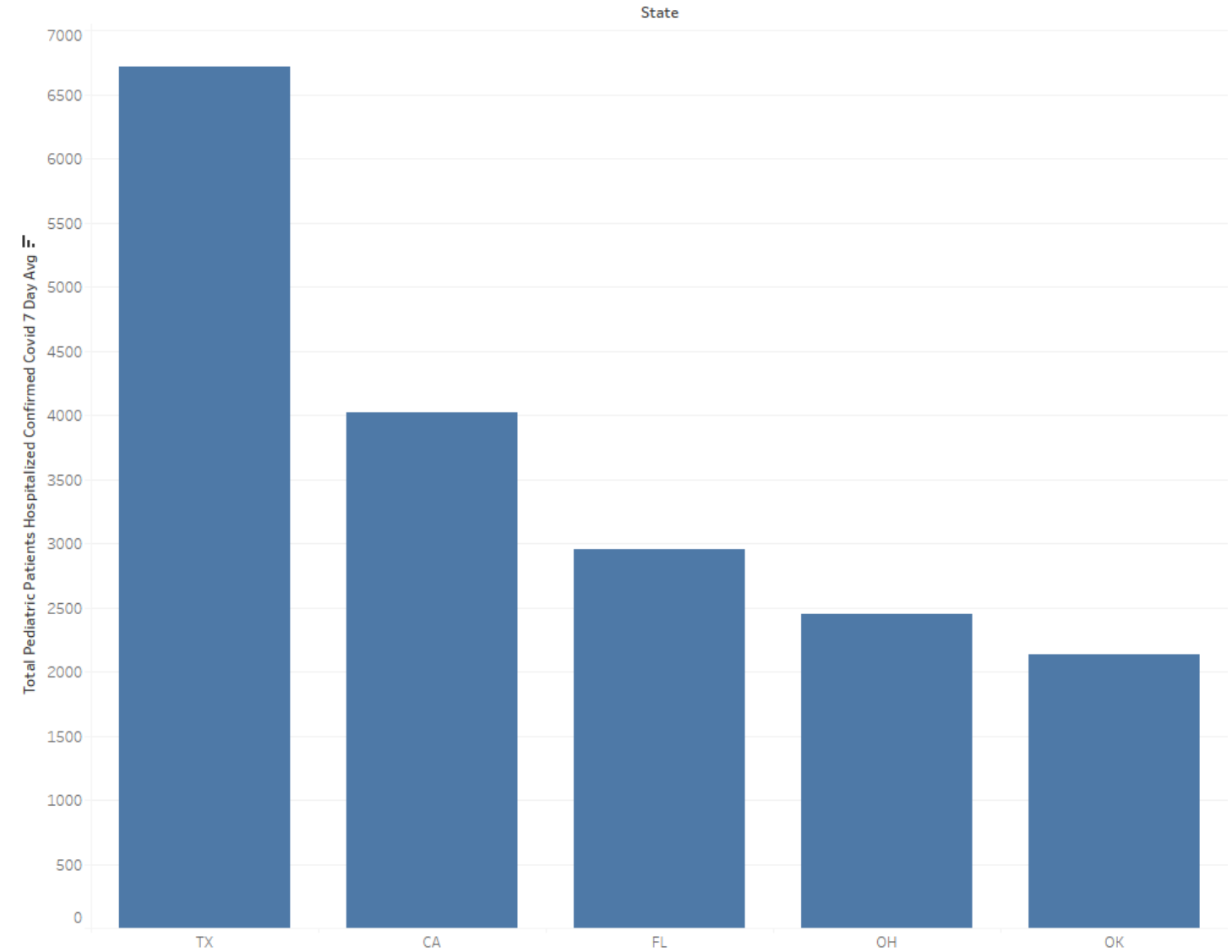
Sum of Total Pediatric Patients Hospitalized Confirmed Covid 7 Day Avg for each Hospital Name. The view is filtered on Hospital Name, which keeps CHILDREN'S HOSP OF LOS ANGELES, CHILDRENS MEDICAL CTR OF DALLAS, COOK CHILDRENS MEDICAL CENTER, SSM HEALTH ST ANTHONY HOSPITAL - OKLAHOMA CITY and TEXAS CHILDRENS HOSPITAL.

6e.) This visualization shows the top 5 states by seven-day average of total pediatric patients with confirmed cases of COVID 19. These identified states are:

1. Texas
2. California
3. Florida
4. Ohio
5. Oklahoma

Again, as an Oklahoma native it is disheartening to see Oklahoma have high levels of pediatric covid cases. A per capita basis would likely make these numbers look worse as Oklahoma has a small population relative to Texas and California.

Top 5 States

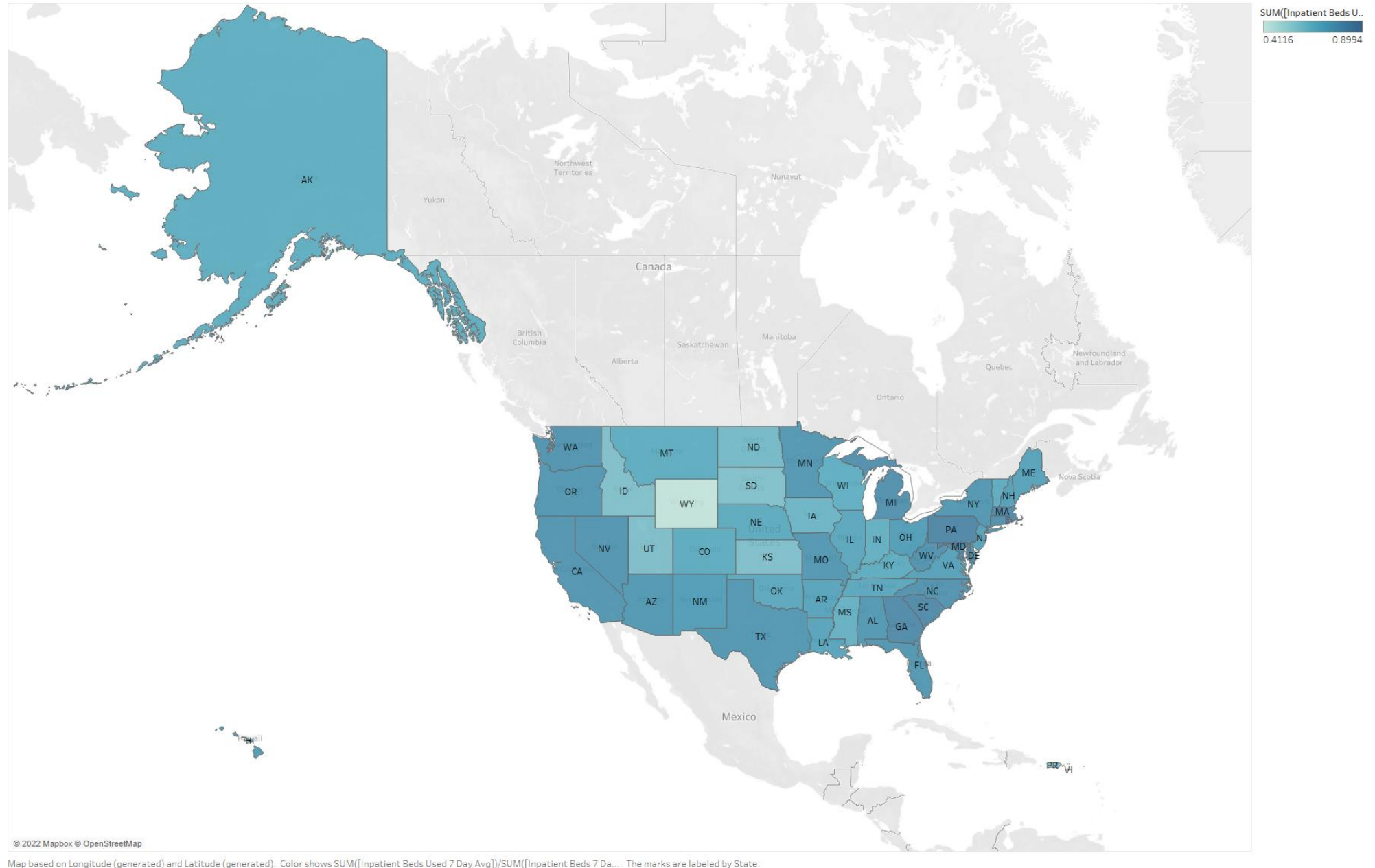


Sum of Total Pediatric Patients Hospitalized Confirmed Covid 7 Day Avg for each State. The view is filtered on State, which keeps CA, FL, OH, OK and TX.



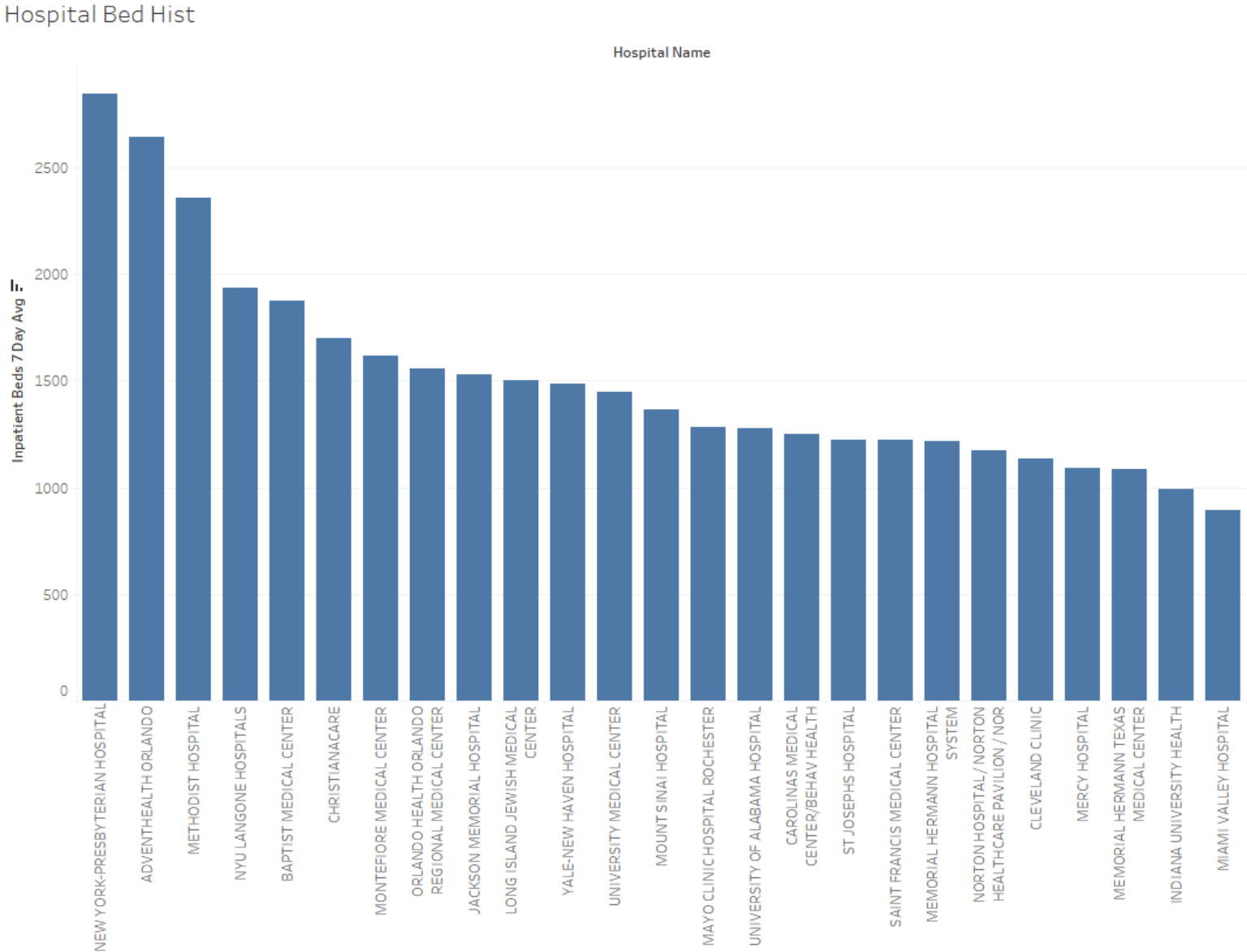
6f.) This visualization shows the percentage of seven-day average inpatient bed capacity used. We can see that parts of the south, the west, and northeast came close to going over capacity. The Midwest and Wyoming especially had more capacity comparatively.

Bed Consumption Map



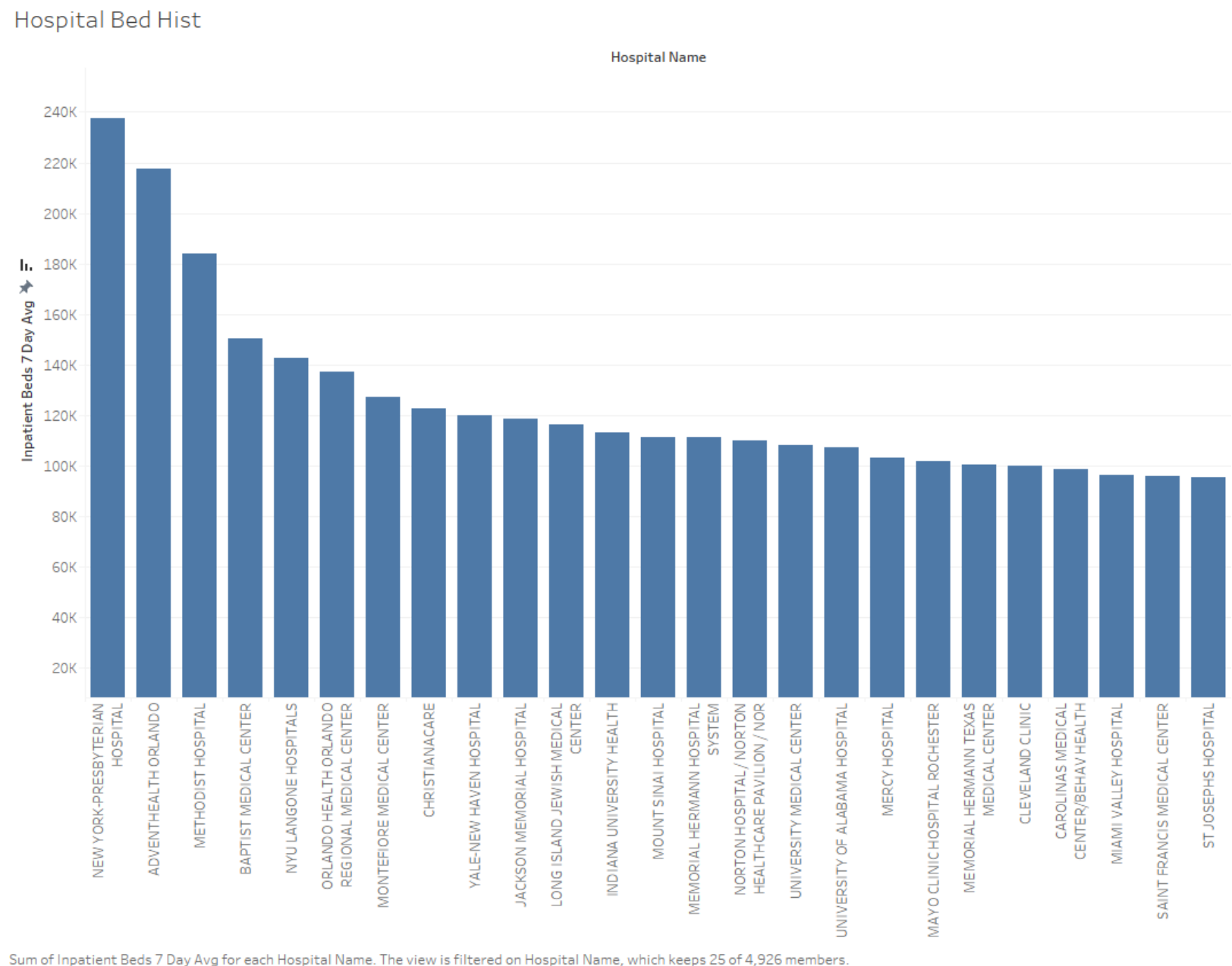
6g.) This visualization shows that New York Presbyterian Hospital had the most hospital beds. We filtered the seven-day average by date so only those hospital bed values on the final date in the data set show here.

Your question asks which hospital *HAS* the most beds. As this data set does not extend to the current date, we can only use the final day of the set if we seek current bed count.



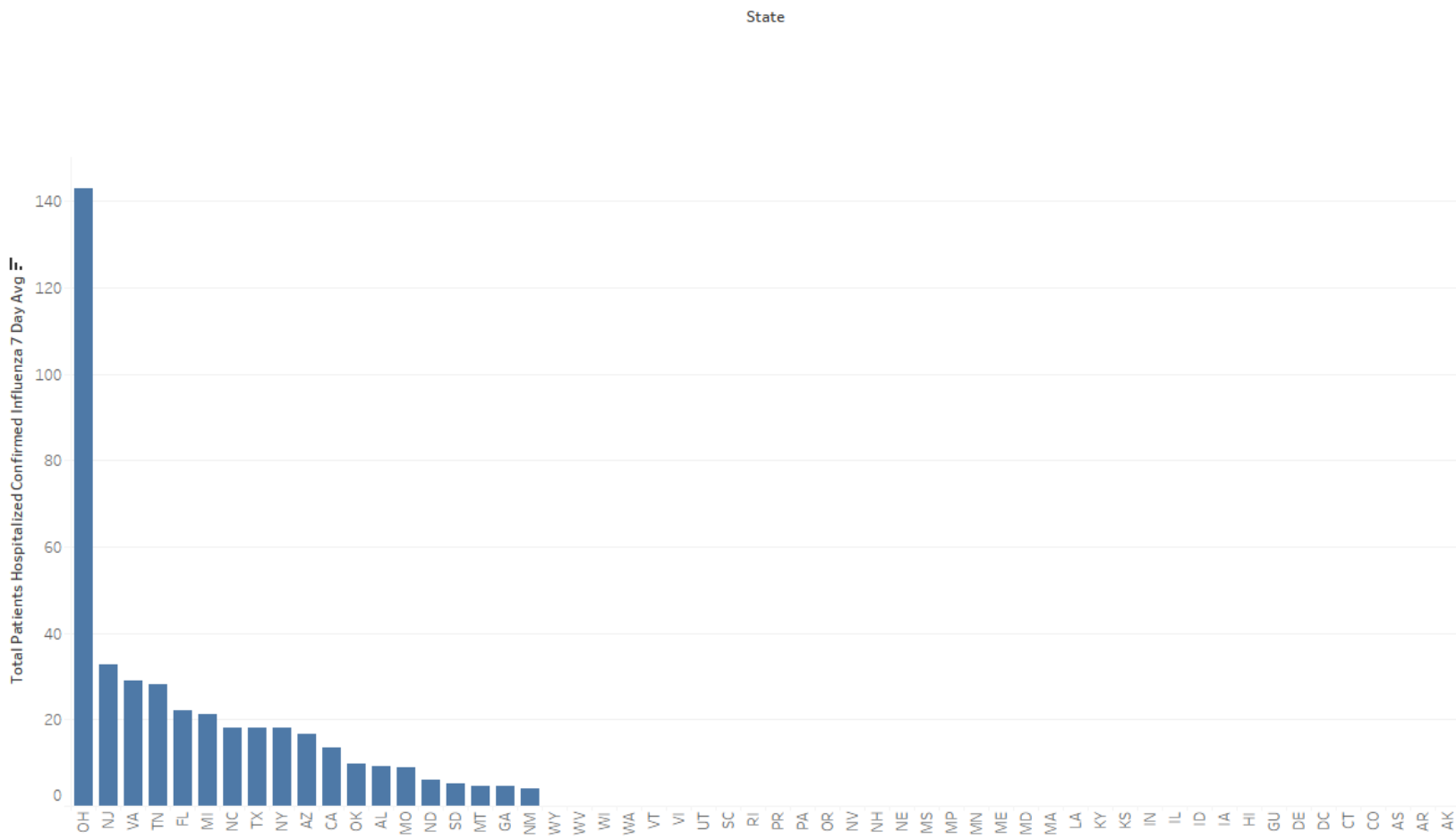
Sum of Inpatient Beds 7 Day Avg for each Hospital Name. The data is filtered on Collection Week, which ranges from 01/28/22 to 01/28/22. The view is filtered on Hospital Name, which keeps 25 of 4,926 members.

6g. continued) We can approach the question of “which hospital has the most inpatient beds” differently. Interpret the question as meaning “which hospital provided the most inpatient bed capacity during the pandemic”. We reach an answer for that question by taking the sum of hospital inpatient bed counts reported (seven-day average). This visualization shows the sum of total daily bed capacity provided (seven-day average). From this view New York Presbyterian Hospital provided the most daily capacity of inpatient beds throughout the pandemic.



6h.) Here we show the total influenza cases reported literally over the last 7 days of the data set (1/21/22 through 1/28/22). Ohio has the most reported cases. Note however we interpret the question differently on the next page. It is surprising to see how low reports of influenza cases are. Many people have stated to me that there is cross contamination between influenza cases and suspected COVID cases. This graph provides plausibility to that viewpoint.

Past 7 Days Literal



Sum of Total Patients Hospitalized Confirmed Influenza 7 Day Avg for each State. The data is filtered on Collection Week, which ranges from 01/21/22 to 01/28/22.

6h.) Here we interpret your question of “Which states have the most influenza hospitalizations for the past seven days” as which states for the total period have the highest hospitalizations due to influenza (as a sum by seven-day average).

Note that New Jersey has the highest reported hospitalizations due to influenza, followed by California. This is surprising as on a per capita basis California has very few cases in comparison to New Jersey and other small states. Large states like New York report zero Influenza hospitalizations which is also unusual.

Past 7 Days Global

