Yugine Lama- Yugineruit2012@gmail.com

Melanie Li- [limelanie01@gmail.com](mailto:limelanie01@gmail.com)

Salma Elhassa [salmae309@gmail.com](mailto:salmae309@gmail.com)

Falande Alexandre [Alex.falande@gmail.com](mailto:Alex.falande@gmail.com)

**Question set: Submit only the questions, not the rubric above.**

**1. If you used your modified SQL data, what did your modifications do? If you didn’t, how did you manipulate your graphics to show what you wanted to show?**

Salma - so yes, I had to make a couple of modifications – some of the data in the .csv file was unnecessary to what I wanted to show in my graphs. For example, each state was given a certain amount of doses of the vaccine – I only wanted to analyze the data of the administered doses. I ended up using only these charts in my data tables – State, Total Doses Administered, Republican, No Lean, Democratic, and Total Doses Administered for those 18+. Also, the data I found online came in two separate .csv files so I merged into one to create the graphs. (<https://covid.cdc.gov/d2ed157e-2e63-4e4f-acde-05786d08e760>) (<https://covid.cdc.gov/94384369-ac73-405e-b447-9d712becbc62>) (<https://www.bea.gov/data/income-saving/personal-income-by-state>) (<https://www.pewforum.org/religious-landscape-study/compare/party-affiliation/by/state/>)

Melanie- I used my modified SQL data. The modifications specified what I wanted to be shown for the specific question. For example, in question 5, it asks for the state that has the most vaccinations, so my SQL code found the sum of vaccinations from each entity and ordered them in a descending order. Then, I copied that data along with the entity’s name. With that information, I simply created a graph using data wrapper and got the illustration I wanted. A similar method was also used for question 7. In addition, I have added a new line to the codes for questions 5 and 7 in order to copy that data and put it into a csv file that helped generate the graphs.

Falande: For specific graphics that I did not have to modify the SQL data. I obtained the data via the various sites and collected the raw data in a csv file. From there I used the data in excel and data wrapper to make visual representations of the data. I had to manipulate the data in multiple ways, first I had to remove multiple items of data were did not the concern the questions at hand and I went through the process of picking the best graphic to represent our data.

Yugine:

**2. For each of your graphics answer the following: If you had multiple authors, kindly indicate which author answered about which graphic. What does your graphic show? What can be learned from the graphic/from the data? What sort of graph did you use (histogram, barplot, etc.)**

**Question 1-** which areas have lower/higher access to vaccines based on income?

Salma - For question one both graphs show vaccination rates based on income. The first one is a horizontal bar graphs that represents the multiple pieces of data in once graph. It represents COVID-19 vaccinations in the U.S. for those 18 and above, total doses admitted, and income per capita for personal income for 2020 – where the pandemic started. The second representation is a line graph. The states are numbers from 1-52 (other territories were considered). Instead of per state on the y-axis I decided to use vaccination doses administered in total. Then specifying via the legend.

**Question 2: Falande**

In the graph for this question, it shows the amount of people vaccinated per hundred people in each state on May 12th 2021. From this graph we can see the states and their total amount of vaccinations. We can total the values of all states to determine the total vaccines given in the US on that date. Additionally, we can use the data on a federal level to see the rates in which people get vaccinated in each state and whichever states need the most federal aid will receive it to aid in vaccinating a larger number of people. For this data I used a bar graph to illustrate how we can answer the question.

**Question 3 -** which areas have been/not been vaccinated based on political affinity?

Salma - For question three the graphs are a bit unique. See it was a bit difficult to combine percentages with vaccination numbers (rates were difficult to find). So I separated the graphs a bit. The first one shows the vaccination rates beginning with the lowest – Alabama all the way to Wyoming with the highest rate. The percentages on the x-axis represent the population and the legend represents the political affiliation. The second graph is cool in my opinion. Okay so I made a map via excel representing states v.s. total doses administered. It’s by color at the top and when hovered over the exact number and state will appear (it’s hard to show in a picture). The second part of the graph is a line plot showing only political affiliation and states to compare it to that map above it.

**Question 5- Melanie**

Which state has the most vaccinations up to this day (May 12, 2021) according to the dataset?

**a. What does your graphic show?**

My graphic shows the name of the entities (on the left) from the dataset and the sum of vaccinations (on the right) for each entity. This data is based on the sum of vaccinations, which are ordered in a descending order from the highest amount to the lowest amount.

P.S.- Due to the large amount of entities from the dataset, I have decreased the size of the image, so if needed, it is possible to increase the size without the quality of the graphic declining.

**b. What can be learned from the graphic/from the data?**

From the graphic, it is clear that California has the highest amount of vaccinations due to its bar being the longest out of the other entities, specifically California has a sum of 64,896,030 vaccinations.

**c. What sort of graph did you use (histogram, barplot, etc.)**

I used a bar chart in order to compare many values. I chose not to use a column chart in order to avoid the clutter from the many values I want to illustrate.

**Question 7- Melanie**

Which day had the highest vaccination record in New York?

**a. What does your graphic show?**

My graphic shows the dates (on the left) available from the dataset and the daily vaccinations (on the right) for each date. This data is based on the daily vaccinations, which are ordered in a descending order from the highest amount to the lowest amount.

P.S.- Due to the large amount of entities from the dataset, I have decreased the size of the image, so if needed, it is possible to increase the size without the quality of the graphic declining.

**b. What can be learned from the graphic/from the data?**

From the graphic, April 4, 2021 has the highest amount of daily vaccinations due to its bar being the longest out of the other entities, specifically on that day, it has an amount of 266,430 vaccinations.

**c. What sort of graph did you use (histogram, barplot, etc.)**

I used a bar chart in order to compare many values. I chose not to use a column chart in order to avoid the clutter from the many values I want to illustrate. Initially, I wanted to use a line graph to depict the amount of vaccination per day over time, but when I saw the result, it was very difficult to see the dates, so I decided to use a bar chart since it clearly shows the information I want.

**Question 12: Falande**

This data shows the percentage of people in various age groups that have been vaccinated in the US. The purpose of this data is to show which age group is most likely to be vaccinated in the US. We can use this data on a state and federal level to see how they can place vaccine sites and how one can possibly create incentives for certain age groups to be more comfortable getting vaccinated. The graph that I used to represent this data is a bar graph.

**3. What if anything, did the project teach you? Do you have any suggestions to improve this project? What issues did you face when trying to answer your questions/copying the data (or any other issues)?**

Salma - I think the data was a bit overwhelming, some things were more important than others. There was a lot of data of federal entities that I did not want to include because it would be an outlier in my data, and they were not any states or territories. It was really useful because I now understand how the CDC and The World Bank uses their sql database to make graphs that the news and really the common person can read easily. I had issues with combining the data of percentages to that of totals – as in numbers and not percentages of the whole. Maybe that would have been helpful but then the population of each state would have to be calculated which is again not necessary for what I was trying to analyze. Having .csv files of data as recommendations might be helpful – I did do the research but again I ran into too much data I did not believe I needed.

Melanie- This project taught me how to export the data from my SQL code into a csv file, I personally thought it was cool and useful to know, that way I don’t have to manually input the information myself. I have no suggestions, this project wasn’t as hard as it seemed. A minor issue I faced when copying the data was the line spacing. Initially, I wasn’t aware that if there was a new line in the middle of the /copy statement it wouldn’t work; luckily, I solved that problem soon after I noticed it.

Falande: Overall I feel like this project taught me that SQL can be applicable in an actual real life setting. It was very interesting to create the code and graphics from raw data we found on the internet. One way I would improve this project would be focusing of csv file data more during class, however other than that I feel like our lecture were clear and concise enough to understand this project as a whole.

**4. If you had unlimited time and resources, what would you different with the graphics?**

Salma - I would have searched for actual vaccination rates instead of total doses given. I also do not think looking at states by income year of 2020 would not be useful. A lot of people were given unemployment benefits and it increased the income rate by so much – which is why many states have such high rates. I think next time I would pick just New York and focus on that – it would have been more manageable data to work with. I also would do some pie graphs if I had only percentages to work with.

Melanie- Personally, I think my graphics are straightforward and clear. So, if I had unlimited time and resources, I would ask the opinions of other people on their thoughts of the graphics I created to see if there was anything they disliked or would like to be fixed. With their opinions, I can hopefully implement their ideas and improve on the graphics.

Yugine: I believe there is always a better product output if we get unlimited resources and time for any kind of project. Similarly for this project, we would have definitely done much better by collecting data from one state and focusing on that as well as getting the opinion of other people would have helped better. However, we would have done good graphings of all if we had more time.

Falande: If I had an unlimited amount of time and resources I would probably make the graphics look nice. It is more of an aesthetic thing rather than an issue with the data. But I feel like when images are aesthetically pleasing it draws people attention and allows them to empathize with whatever issues we are talking about. I would also most likely add my graphics within an infographic that would give more background information and importance to the data.

**5. Did you enjoy this class? What did you like? What did you dislike? Do you have any advice for improvements or other suggestions? If you have multiple authors, please separate by author)**

Salma - I really enjoyed this class as I learned how many companies and businesses use this to make so much money. And a bioinformatics major – and this sounds a little nerdy – but I am really glad I learned how to use sql databases to make meaningful graphs that even the science industry can use. I think it might have been a bit fast paced but eventually I did catch up, so it wasn’t too bad. I really hope my future computer science classes are taught like this and not with exams.

Melanie- I really enjoyed this class as well. I really like the idea of not having exams, it definitely takes off the pressure. To add on, I believe it’s more important to familiarize ourselves with the reason why we chose to code that certain line and not just remembering the surface of coding and theories. I also like the real time captions you had because my internet isn’t always that great so at times it’s useful to read what you had said in case I missed it. A suggestion I would make is to make a list of the types of joins we learned and write a basic description of each of them in that same list because I’m still forgetting the join types and the difference amongst them.

Yugine: I had fun taking this class honestly. I took this class as I wanted to learn SQL. I have taken several CS classes ranging from 100 level to 400 levels but this was one class I thought was worth taking. Many of my friends who are working in different companies told me to be familiar with sql language and taking their suggestion i took this class. Honestly, it was also because of professor Ariel that things went much better and I am also planning to take more advanced sql classes in future.

Falande: I really enjoyed the class, I have had issues learning coding and it was very difficult for me to understand certain languages at one point. But, SQL is really easy and digestible for a beginner like myself. Additionally, the real world application made it easier to conceptualize when I would write code. I liked the coding language itself. An improvement that could be done for this class may be the usage of many more open-ended assignments similar to our project. Just allowing us to pick an issue or industry and creating tables and writing code for said table. I feel like something like that would have made me a bit more engaged.

**Graphics we created:**

1 - which areas have lower/higher access to vaccines based on income?

create table vaccination\_rates(

age integer,

name varchar,

vaccination\_status varchar,

income integer,

state varchar

);

\copy vaccination\_rates(age, name, vaccination\_status, income,state) FROM 'C:\Users\Salma\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\our-world-in-data-vaccination-coverage-by-income.csv' DELIMITER ',' CSV HEADER

SELECT

AVG(ALL vaccination\_status),

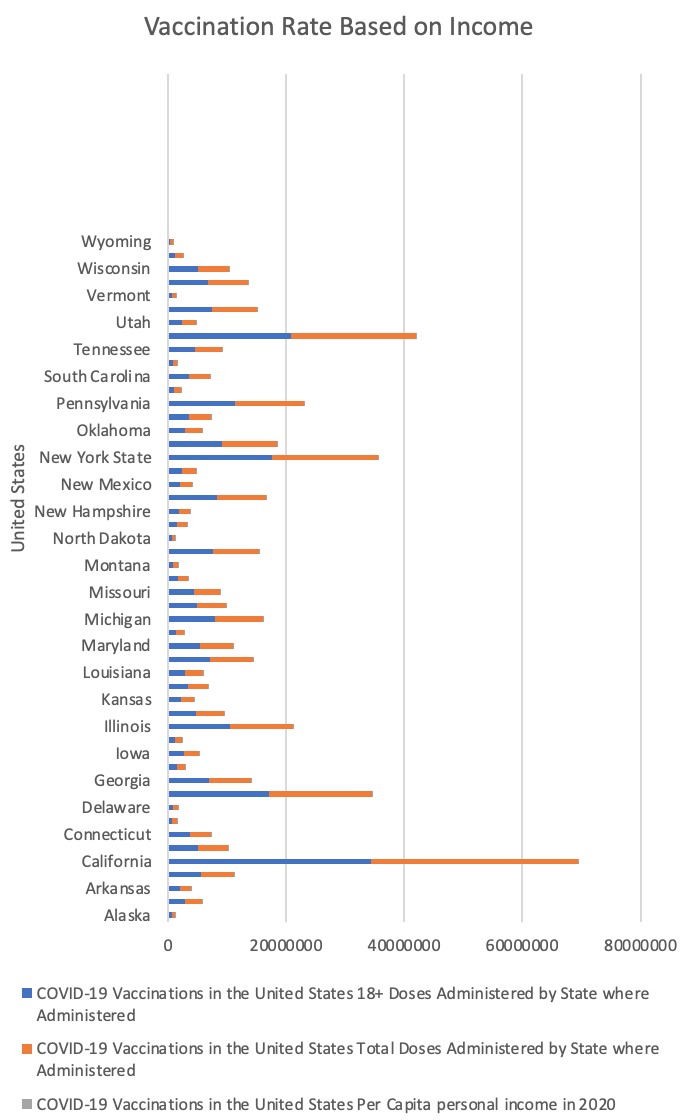
avg\_vaccination\_rate

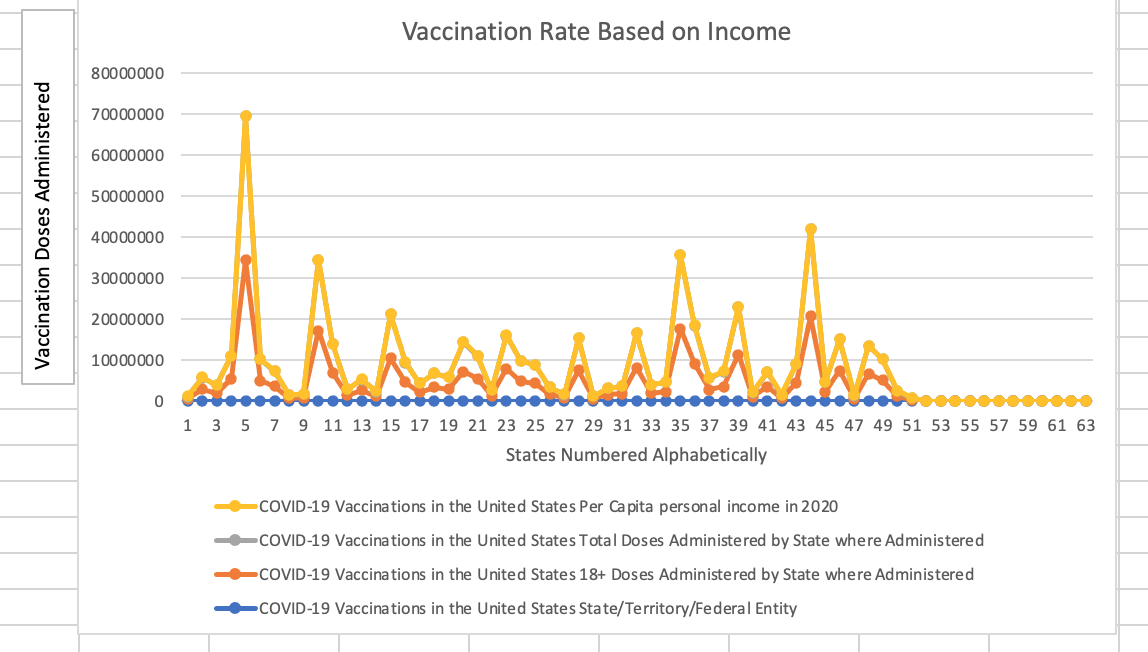
FROM

vaccination\_rates

GROUP BY

income, state;





2 - Which day had the lowest vaccination record and in which state?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

--FIRST SOLUTION

select entity, daily\_vaccinations, day

from us\_vaccination

group by entity, day, daily\_vaccinations

order by daily\_vaccinations, entity asc limit 1;

--SECOND SOLUTION

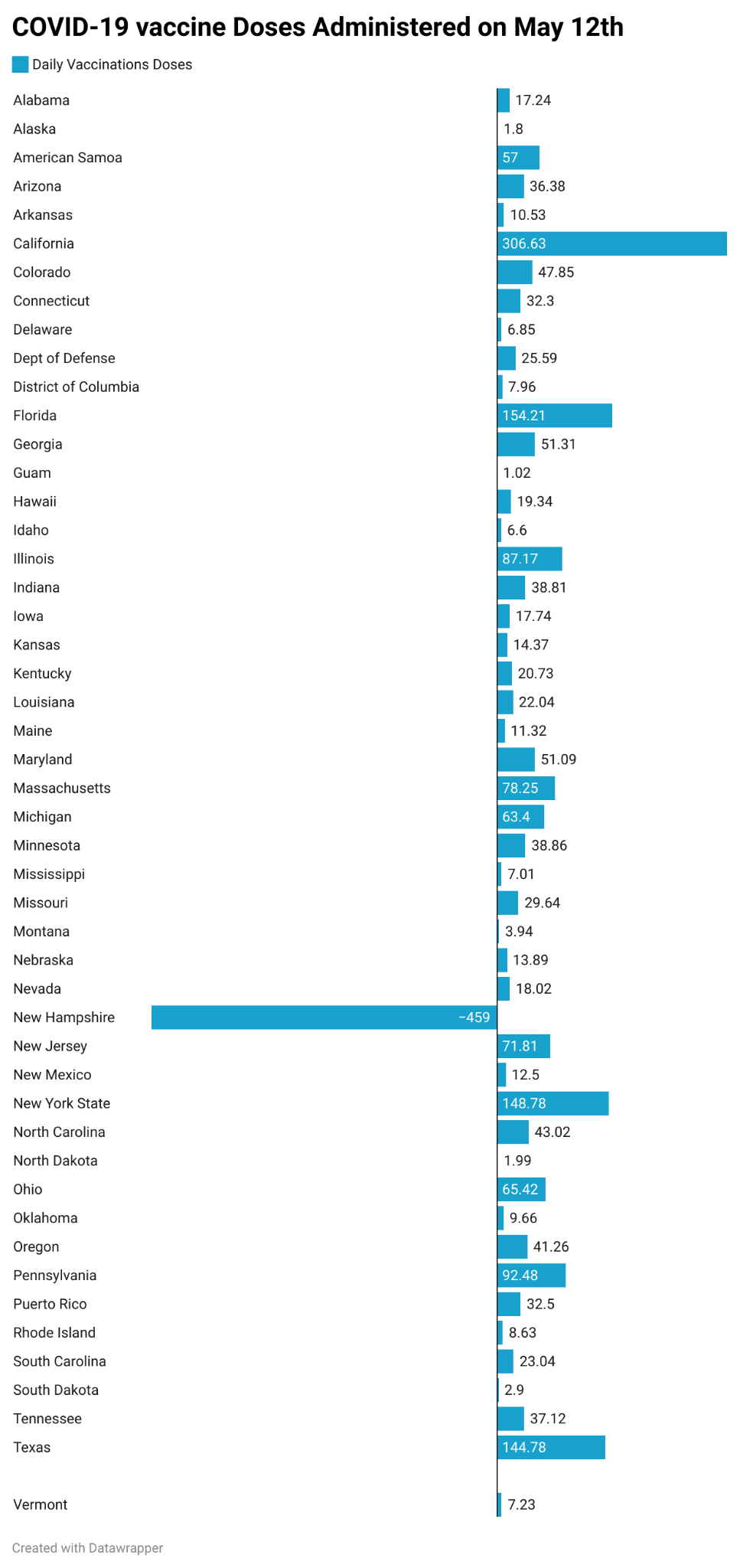
select entity, daily\_vaccinations, day

from us\_vaccination

where daily\_vaccinations >= 0

group by entity, day, daily\_vaccinations

order by daily\_vaccinations, entity asc limit 1;



3 - which areas have been/not been vaccinated based on political affinity?

ALTER TABLE vaccination\_rates

ADD political\_affiliation varchar:

\copy vaccination\_rates(age, name, vaccination\_status, income,state,political\_affiliation) FROM 'C:\Users\Salma\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\covid-cdc-gov-covid-data-tracker.csv' DELIMITER ',' CSV HEADER

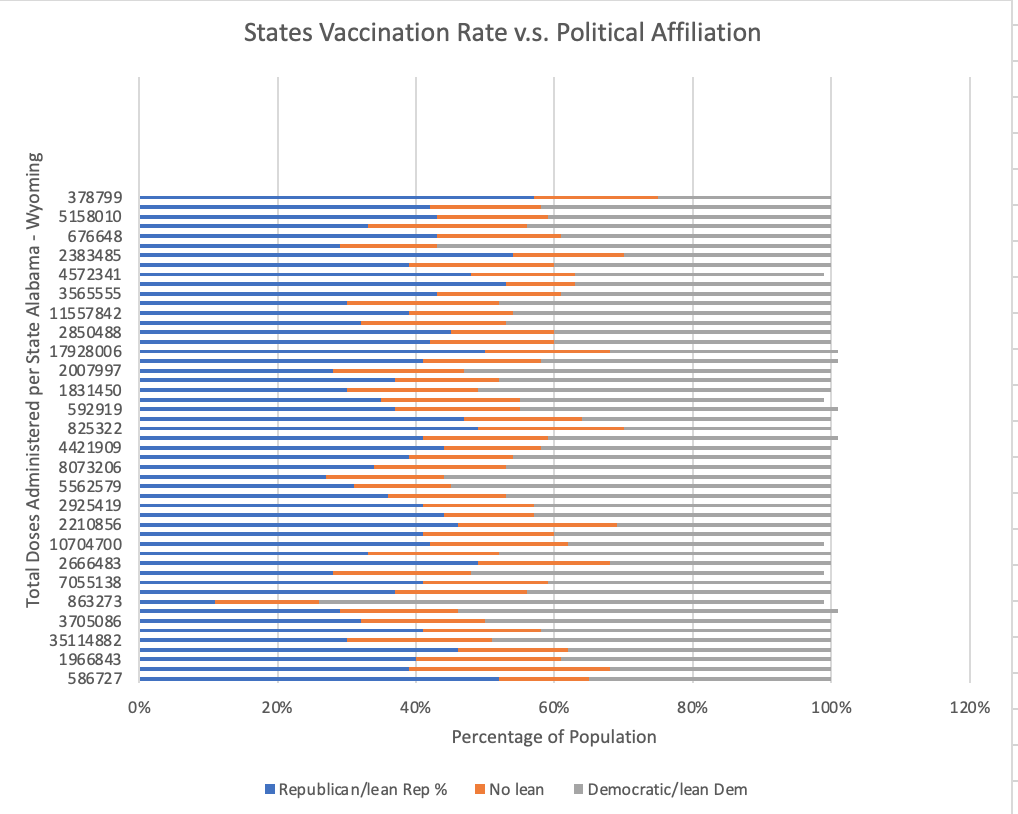
SELECT vaccination\_rates, political\_affiliation

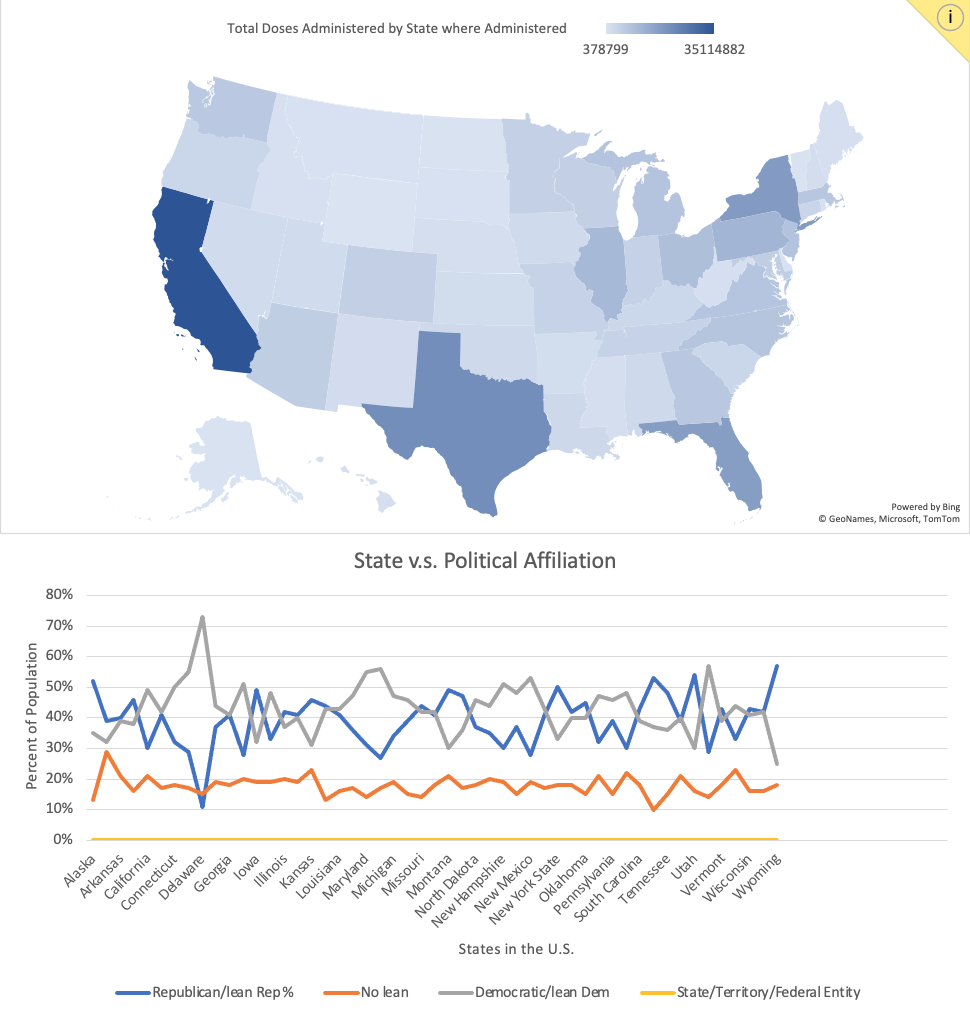
FROM table\_name

WHERE democratic OR republican OR third\_party OR none\_of\_the\_above

GROUP BY

political\_affiliation, state;





4- how many vaccine hub per area to find out if everyone has access to vaccines

5 - Which state has the most vaccinations up to this day (May 12, 2021) according to the dataset?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

select entity,

round(sum(daily\_vaccinations), 2) as sum\_vaccine

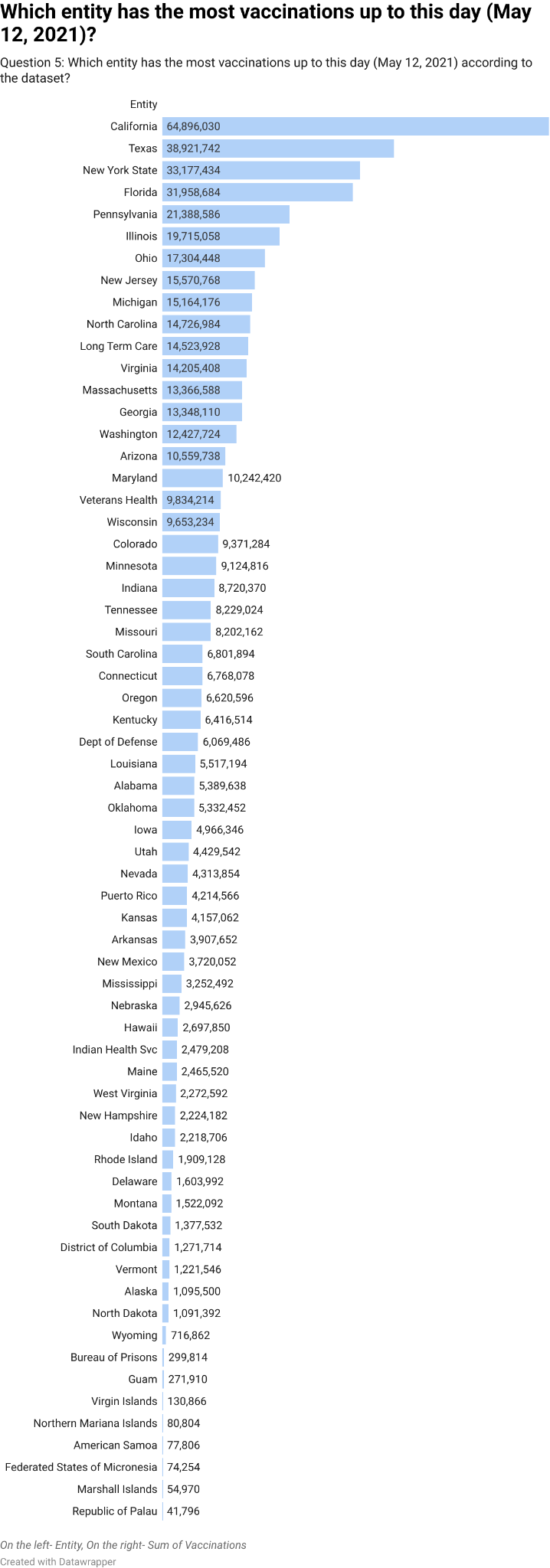
from us\_vaccination

where entity != 'United States'

group by entity

order by sum\_vaccine desc;

\copy (select entity, round(sum(daily\_vaccinations), 2) as sum\_vaccine from us\_vaccination where entity != 'United States' group by entity order by sum\_vaccine desc) to 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\Project 3- Q5.csv' with csv



6- how can we encourage more people to get the vaccine?

We can use our overall analysis of data. As you can see from our queries of the data set we can put faith within the American people that the usage of vaccines has substantially aided in the decrease of covid cases. This is in addition to the fact that we do not see any adverse effect of the vaccine as more Americans continue to receive it. Overall, the fact that we have covid deaths, but not covid vaccine deaths should be reason enough for people to get vaccinated.

7- Which day had the highest vaccination record in New York?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

select entity, daily\_vaccinations, day

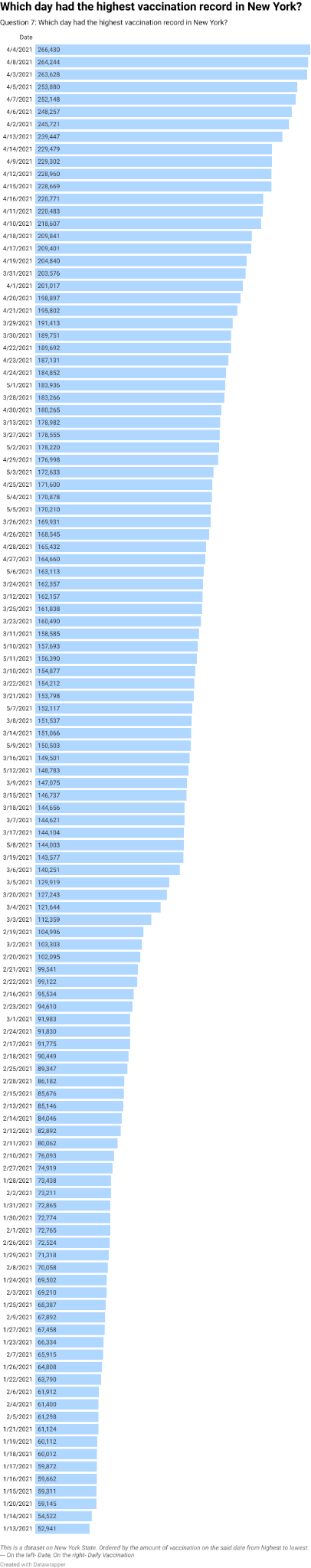
from us\_vaccination

where entity = 'New York State'

group by entity, day, daily\_vaccinations

order by daily\_vaccinations desc;

\copy (select entity, daily\_vaccinations, day from us\_vaccination where entity = 'New York State' group by entity, day, daily\_vaccinations order by daily\_vaccinations desc) to 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\Project 3- Q7.csv' with csv



8- Which entity in the dataset should have an increase of access to vaccinations?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

select entity,

round(avg(daily\_vaccinations), 2) as avg\_vaccine

from us\_vaccination group by entity order by avg\_vaccine asc;

9. How many Americans are currently fully vaccinated as of May 12th in the United States?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM ‘'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

SELECT entity, sum(daily\_vaccinations)

FROM us\_vaccination

where entity = 'United States'

group by entity;

10. How many people under 18 have been vaccinated ?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int,

age int

);

SELECT age, COUNT(daily\_vaccinations)

FROM us\_vaccination

WHERE age < 18

group by age;

11. What is the average rate of vaccination per day?

create table us\_vaccination(

entity varchar,

code varchar,

day date,

daily\_vaccinations int

);

\copy us\_vaccination(entity, code, day, daily\_vaccinations) FROM 'C:\Users\Melanie\OneDrive\Desktop\Comp Sci- 232 (Spring Term 2021)\us-daily-covid-vaccine-doses-administered.csv' DELIMITER ',' CSV HEADER

select day,

round(avg(daily\_vaccinations), 2) as avg\_vaccine

from us\_vaccination

group by day

order by day asc, avg\_vaccine;

12. Which age group is more likely to get vaccinated?

create table us\_vaccination(

entity varchar,

code varchar,

age int,

day date,

daily\_vaccinations int

);

select entity, age,

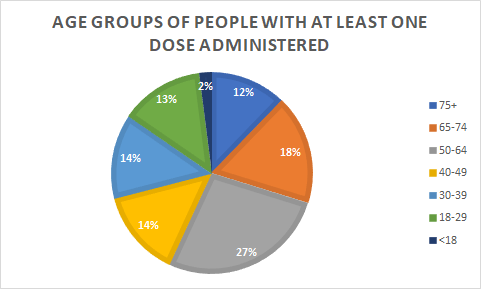
round(avg(daily\_vaccinations), 2) as avg\_vaccine

from us\_vaccination

group by age

order by avg\_vaccine desc, age;

https://covid.cdc.gov/covid-data-tracker/#vaccination-demographic

****