Jupyter Notebook: Vaccin Status in Each State as of 10/27/2021

By: Ashley Krause

1. Introduction:

For this project, our group explored Vaccine data in the US since the Covid-19 vaccines came out. Our main question we wanted to answer was: **"How is the United States doing as a whole with vaccination percentages, as well as how is each individual state doing?"**

We decided to ask the following questions: Which companies have administered more doses of the vaccine? Which states have the highest and lowest percentage of their populations vaccinated? Is there any age group that is slacking in percentages vaccinated overall? Which states rely more on different vaccines than other states? And lastly, how many vaccines are being wasted (distributed vs administered)?

> Distributed being the vaccines that were given to the states and Administered being the vaccines that were given to the people in each state

We found that the vaccines were distributed to states based on population more so than politics as the top distributed state was CA and followed by TX who are both very different politically leaning states but two of the biggest in the country. This is not to say that population is a sole factor though as New York was fourth on the list for distributed behind Florida in third.

When looking at the line graphs that show the distributed over time per vaccine type that J&J did not really have a chance to take off as it was not distributing anything for the majority of the year till about midway through April. This explains why in the vaccine distribution bar graph for J&J is so low as it had hardly been distributed throughout the year.

2. Data

A direct link to the data can be found at the following website.

- https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc/data

This is the CDC's official vaccine database. It has data from each state for each day since the vaccinations have been introduced. The dataset updates daily, but was downloaded on 10/27, so the data ends then. The dataset was also run through Tableau Prep Builder to remove unnecessary fields and rows.

```
[ ] ... #load data into a data frame
    import pandas as pd
    url = 'https://raw.githubusercontent.com/tjjoe/CS260FinalProject/main/VaccineStatistics.csv'
    vaccines = pd.read_csv(url, parse_dates = ["Date"])
    vaccines
```

	Date	MMWR_week	Location	Distributed	Distributed_Janssen	Distributed_Moderna	Distributed_Pfizer	Dist_Per_100K	Distributed_Per_100k_12Plus	Distributed_Per_100k_18Plus	Di
0	2021-10-27	43	UT	4328350	195700	1592040	2540610	135010	166972	190276	
1	2021-10-27	43	KY	6240815	310500	2665520	3264795	139688	163966	180120	
2	2021-10-27	43	NV	4267880	211600	1531060	2525220	138561	162784	178758	
3	2021-10-27	43	ME	2327320	127900	961380	1238040	173136	196409	212469	
4	2021-	43	TN	9213340	419100	3585220	5209020	134911	157882	173212	

2.1 Variables Explained

Below is a list of each variable in the dataset and its explanation.

- **Date** - Date
- **MMWR_week** - Morbidity and Mortality Weekly Report (MMWR) (Early data in the set has a later MMWR week because vaccine came out late in the year last year)
- **Location** - State/Territory/Federal Entity
- **Distributed** - Total number of distributed doses
- **Distributed_Janssen/Moderna/Pfizer** - Total number of doses delivered for each respective company
- **Distributed_per_100k** - Delivered doses per 100,000 census population
- **Distributed_per_100k_12Plus/18Plus/65Plus** - Total number of delivered doses per 100,000 in each age group
- **Administered** - Total number of administered vaccines based on state where administered
- **Administered_12Plus/18Plus/65Plus** - Total number of doses administered to people in each age group based on the jurisdiction where recipient lives
- **Administered_Janssen/Moderna/Pfizer** - Total number of doses administered for each company
- **Admin_per_100k** - Total number of doses administered per 100,000 census population based on the jurisdiction where recipient lives
- **Admin_per_100k_12Plus/18Plus/65Plus** Total number of doses administered to people in each age group per 100,000 each age group population
- **Recip_Administered** - Total number of administered vaccines based on the jurisdiction of residence
- **Administered_Dose1_Recip** - Total number of people with at least one dose based on the jurisdiction where recipient lives
- **Administered_Dose1_Pop_Pct** - Percent of population with at lease one dose based on the jurisdiction where recipient lives
- **Administered_Dose1_Recip_12Plus/18Plus/65Plus** - Total number of people in each age group with at least one dose based on the jurisdiction where recipient lives

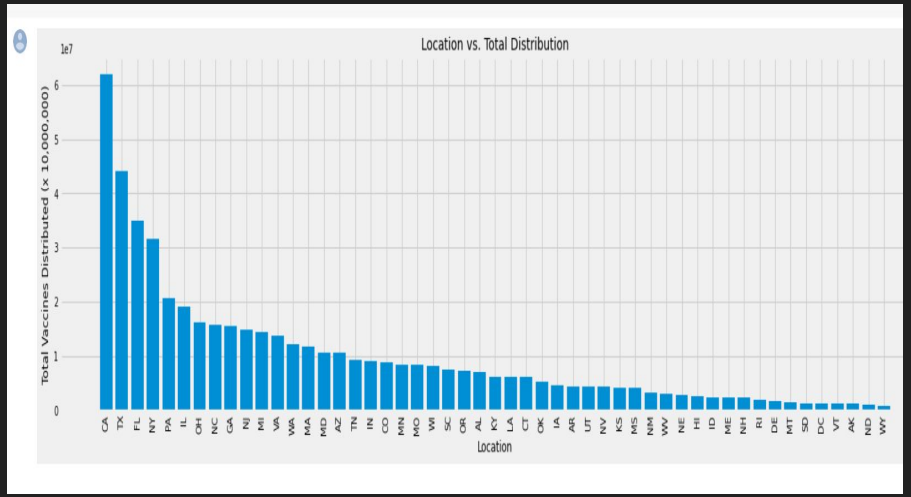
- **Series_Complete_Yes** - Total number of people who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction where recipient lives
- **Series_Complete_Pop_Pct** - Percent of people who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction where recipient lives
- **Series_Complete_12Plus/18Plus/65Plus** - Total number of people in each age group who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction where recipient lives
- **Series_Complete_12Plus/18Plus/65PlusPop_Pct** - Percent of people in each age group who are fully vaccinated (have second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction where recipient lives
- **Series_Complete_Janssen/Moderna/Pfizer** - Total number of people who are fully vaccinated with each vaccine based on the jurisdiction where recipient lives
- **Series_Complete_Janssen/Moderna/Pfizer_12Plus/18Plus/65Plus** - Total number of people in each age group who are fully vaccinated with the each vaccine based on the jurisdiction where recipient lives
- **Additional_Doses** - Total number of people who are fully vaccinated and have received a booster (or additional) dose.
- **Additional_Doses_Vax_Pct** - Percent of people who are fully vaccinated and have received a booster (or additional) dose.
- **Additional_Doses_18Plus/50Plus/65Plus** - Total number of people in each age group that are fully vaccinated and have received a booster (or additional) dose.
- **Additional_Doses_18Plus/50Plus/65Plus_Vax_Pct** - Percent of people in each age group who are fully vaccinated and have received a booster (or additional) dose.
- **Additional_Doses_Moderna/Pfizer/Janssen** - Total number of fully vaccinated people who have received a booster from any company (or

3. Results

```
[ ]  import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     plt.style.use('fivethirtyeight')
```

```
[ ]  vaccines = vaccines.sort_values(by='Distributed', ascending = False
```

```
[ ]  plt.figure(figsize =(20,5))
     plt.rcParams['font.size'] = '10'
     plt.xticks(rotation = 90)
     plt.bar(vaccines['Location'],vaccines['Distributed'])
     plt.title("Location vs. Total Distribution")
     plt.ylabel("Total Vaccines Distributed (x 10,000,000)")
     plt.xlabel("Location");
```

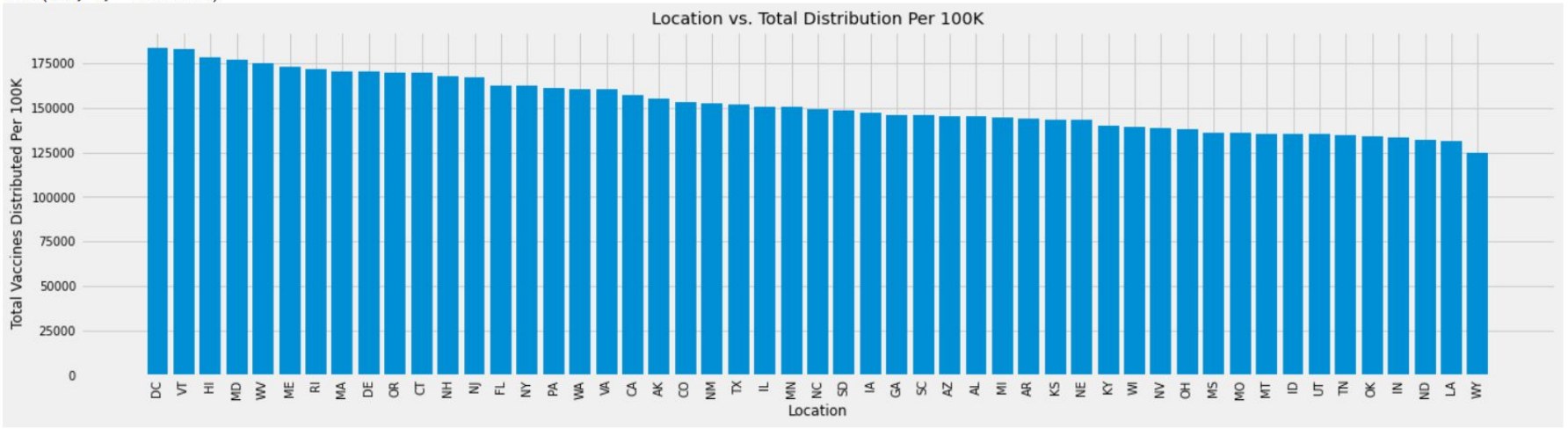


Graph 1: When looking at this graph we can take a look at the different locations and see which areas have recieved distributions of the vaccine more than others. Here we can see that California is the has had more vaccines distributed. While Wyoming has the the least amount of vaccines distributed. This can most likely be attributed to population, as the more people are in a state, the more vaccines will be needed.

```
[ ] vaccines = vaccines.sort_values(by='Dist_Per_100K', ascending = False)
```

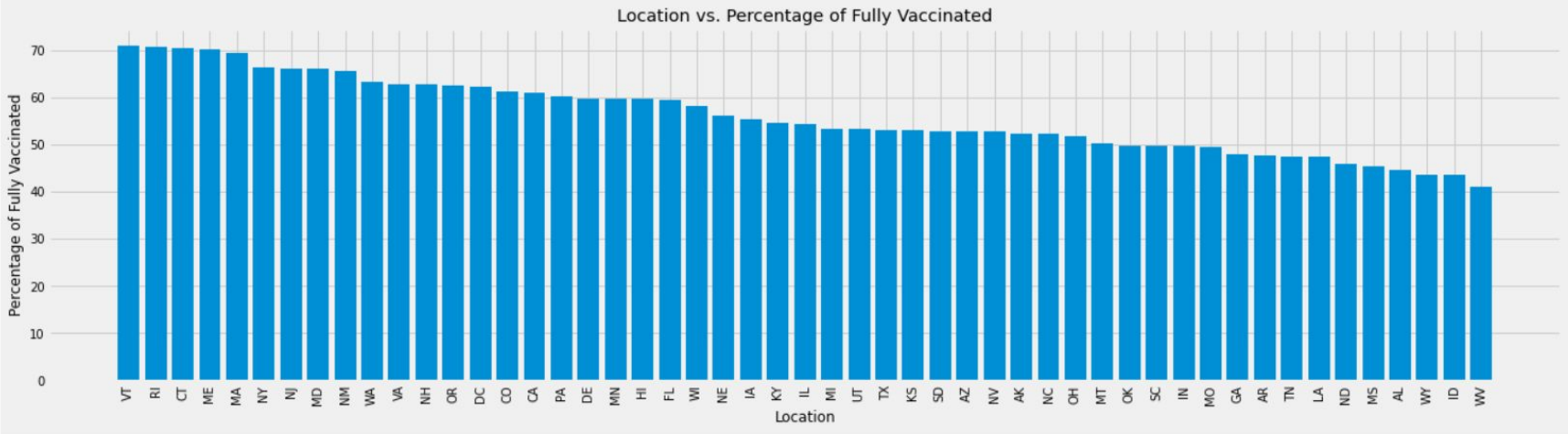
```
[ ] plt.figure(figsize =(20,5))
    plt.rcParams['font.size'] = '10'
    plt.xticks(rotation = 90)
    plt.bar(vaccines['Location'],vaccines['Dist_Per_100K'])
    plt.title("Location vs. Total Distribution Per 100K")
    plt.ylabel("Total Vaccines Distributed Per 100K")
    plt.xlabel("Location")
```

```
Text(0.5, 0, 'Location')
```

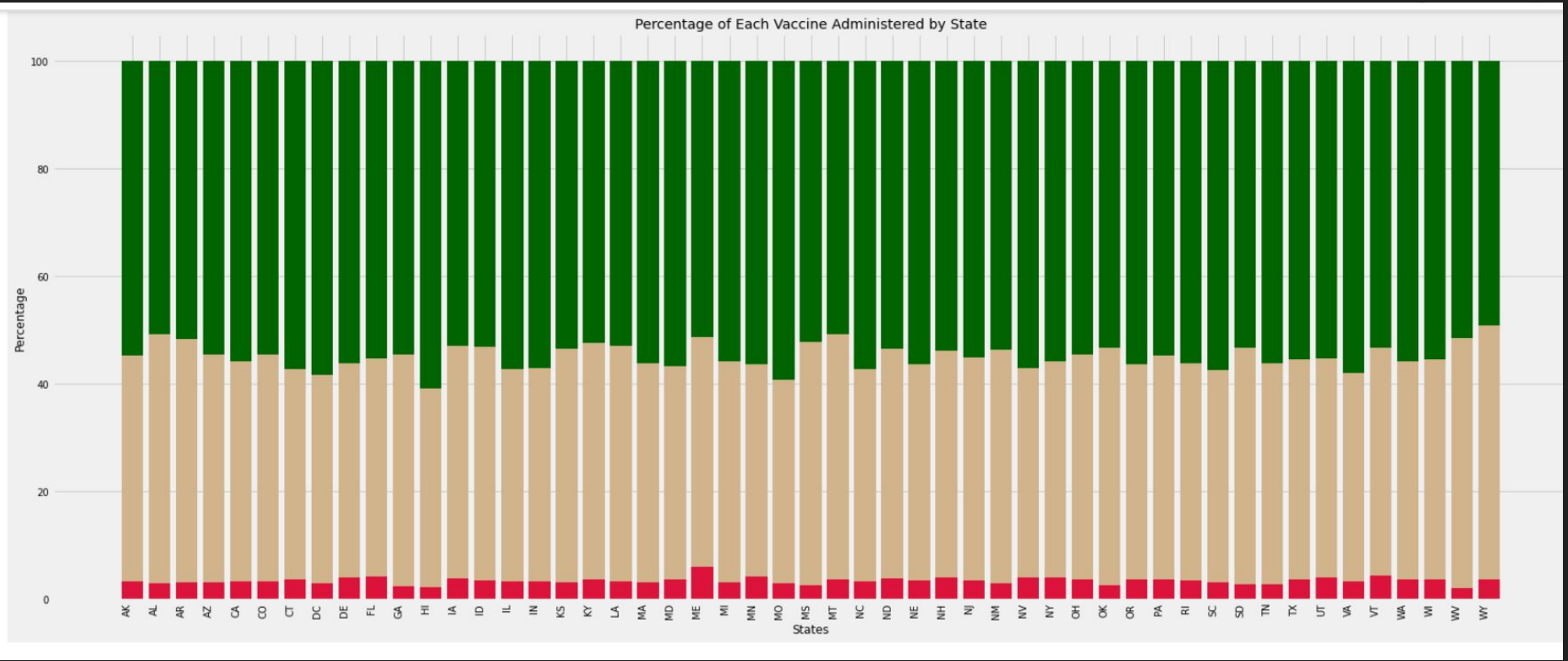


```
[ ] vaccines = vaccines.sort_values(by='Series_Complete_Pop_Pct', ascending = False)
```

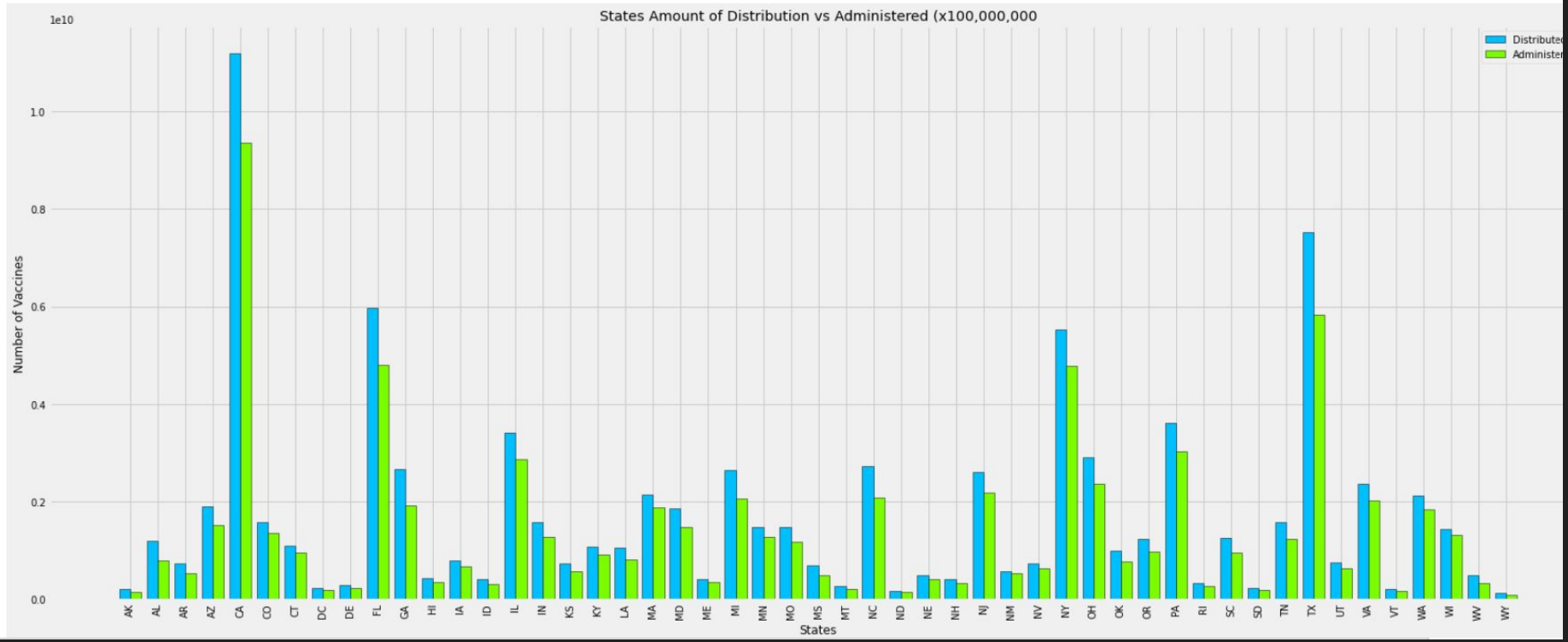
```
[ ] plt.figure(figsize =(20,5))
    plt.rcParams['font.size'] = '10'
    plt.xticks(rotation = 90)
    plt.bar(vaccines['Location'],vaccines['Series_Complete_Pop_Pct'])
    plt.title("Location vs. Percentage of Fully Vaccinated")
    plt.ylabel("Percentage of Fully Vaccinated")
    plt.xlabel("Location");
```



```
groupedvstates = groupedvstates.sort_values(by='Location', ascending = True)

total = groupedvstates['Administered_Pfizer'] + groupedvstates['Administered_Moderna'] +groupedvstates['Administered_Janssen']
proportion_pfizer = np.true_divide(groupedvstates['Administered_Pfizer'], total) * 100
proportion_moderna = np.true_divide(groupedvstates['Administered_Moderna'], total) * 100
proportion_jj = np.true_divide(groupedvstates['Administered_Janssen'], total) * 100

plt.figure(figsize =(25,10))
ind = np.arange(51)
plt.bar(groupedvstates.index, proportion_pfizer, width=0.8, label='Pfizer', color='darkgreen', bottom=proportion_jj+proportion_moderna)
plt.bar(groupedvstates.index, proportion_moderna, width=0.8, label='Moderna', color='tan', bottom=proportion_jj)
plt.bar(groupedvstates.index, proportion_jj, width=0.8, label='Janssen', color='crimson')


plt.ylabel("Percentage")
plt.xlabel("States")
plt.title("Percentage of Each Vaccine Administered by State")
plt.ylim=1.0

# rotate axis labels
plt.setp(plt.gca().get_xticklabels(), rotation=90, horizontalalignment='right')

plt.show()
```

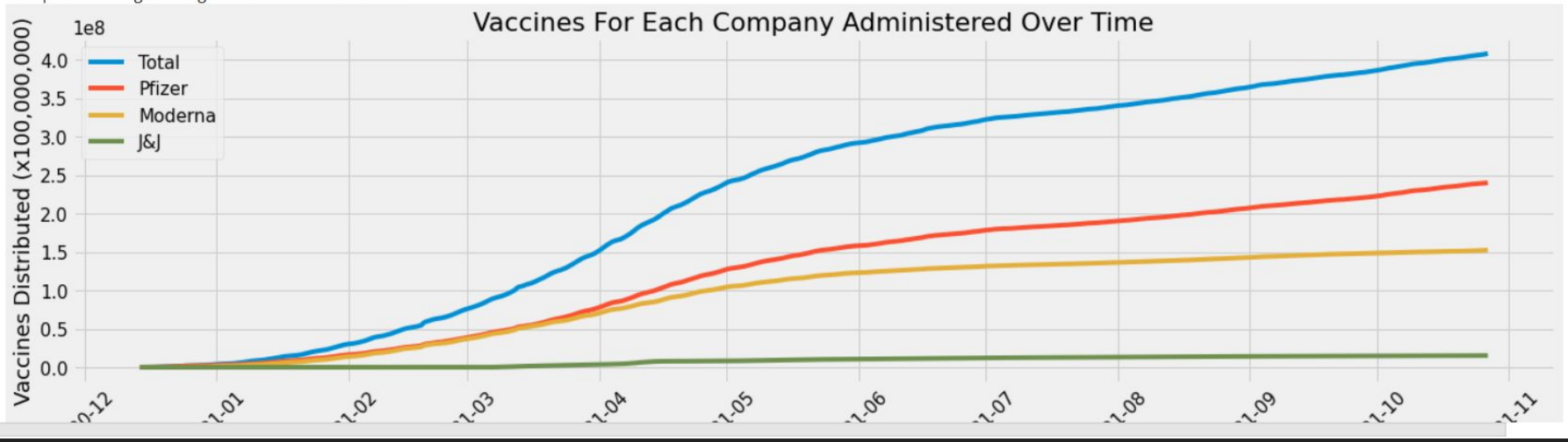

Percentage of Each Vaccine Administered by State
Percentage
100
80
60
40
20
0
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
States

```
[ ]  plt.figure(figsize =(25,10))
     bystate = groupedvstates.index
     dis = groupedvstates['Distributed']
     adm = groupedvstates['Administered']

     X_axis = np.arange(len(bystate))

     plt.bar(X_axis - 0.2, dis, 0.4,color='deepskyblue',edgecolor = 'black',label = 'Distributed')
     plt.bar(X_axis + 0.2, adm, 0.4,color = 'lawngreen',edgecolor = 'black',label = 'Administered')

     plt.xticks(X_axis, bystate)
     plt.xlabel("States")
     plt.ylabel("Number of Vaccines")
     plt.title("States Amount of Distribution vs Administered (x100,000,000")
     plt.legend()
     plt.xticks(rotation = 90)
     plt.rcParams['font.size'] = '15'
     plt.show()
```


States Amount of Distribution vs Administered (x100,000,000
1e10
1.0
0.8
0.6
0.4
0.2
0.0
Number of Vaccines
Distributed
Administered
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
States

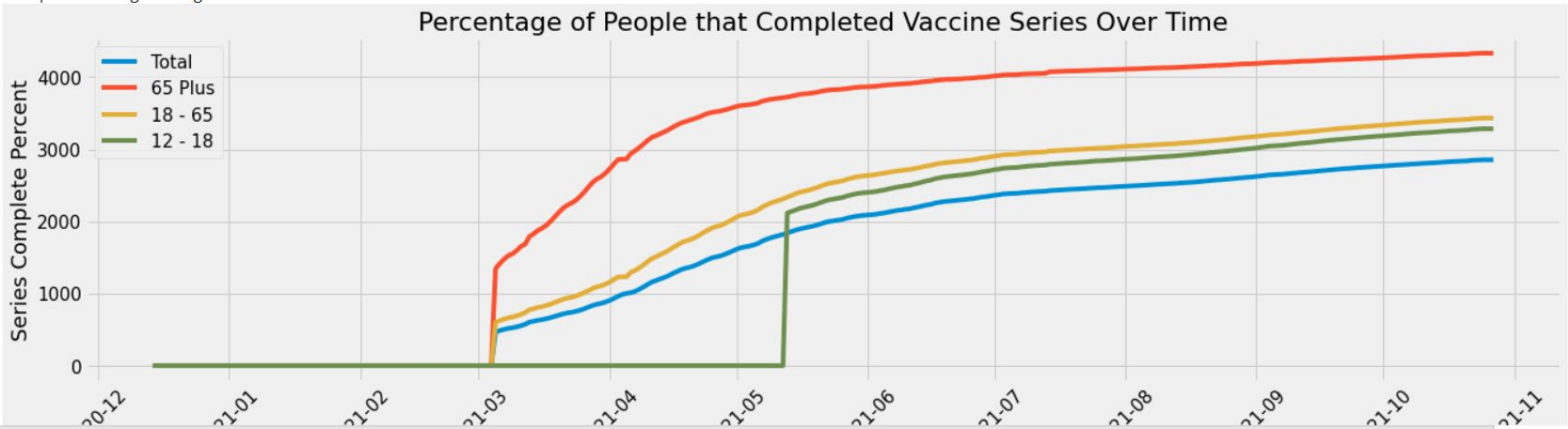
```
[ ]  plt.figure(figsize =(20,5))
     plt.plot(vaccineDates.index, vaccineDates['Administered'])
     plt.plot(vaccineDates.index, vaccineDates['Administered_Pfizer'])
     plt.plot(vaccineDates.index, vaccineDates['Administered_Moderna'])
     plt.plot(vaccineDates.index, vaccineDates['Administered_Janssen'])
     plt.xticks(rotation = 45)
     plt.title('Vaccines For Each Company Administered Over Time')
     plt.xlabel('Date')
     plt.ylabel('Vaccines Distributed (x100,000,000)')
     plt.legend(['Total', 'Pfizer', 'Moderna', 'J&J'])
```

```
<matplotlib.legend.Legend at 0x7fed0416ab90>
```



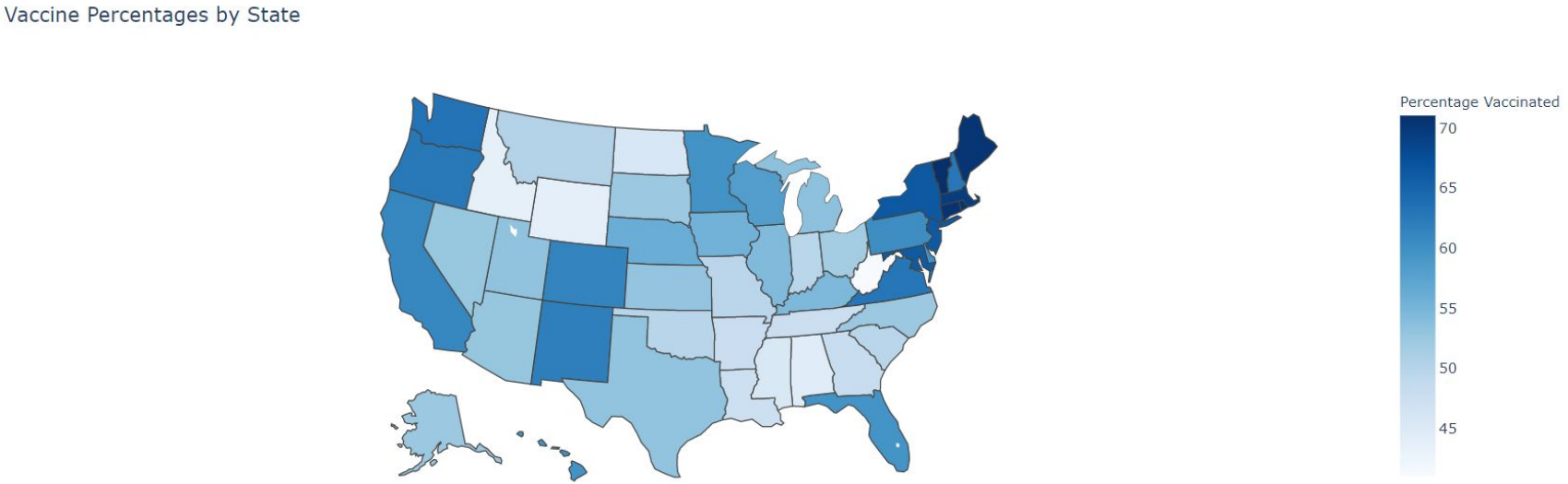
```
[ ] plt.figure(figsize =(20,5))
    plt.plot(vaccineDates.index, vaccineDates['Series_Complete_Pop_Pct'])
    plt.plot(vaccineDates.index, vaccineDates['Series_Complete_65PlusPop_Pct'])
    plt.plot(vaccineDates.index, vaccineDates['Series_Complete_18PlusPop_Pct'])
    plt.plot(vaccineDates.index, vaccineDates['Series_Complete_12PlusPop_Pct'])

    plt.xticks(rotation = 45)
    plt.title('Percentage of People that Completed Vaccine Series Over Time')
    plt.xlabel('Date')
    plt.ylabel('Series Complete Percent')
    plt.legend(['Total','65 Plus', '18 - 65', '12 - 18'])
```

```
<matplotlib.legend.Legend at 0x7fed04c4bdd0>
```



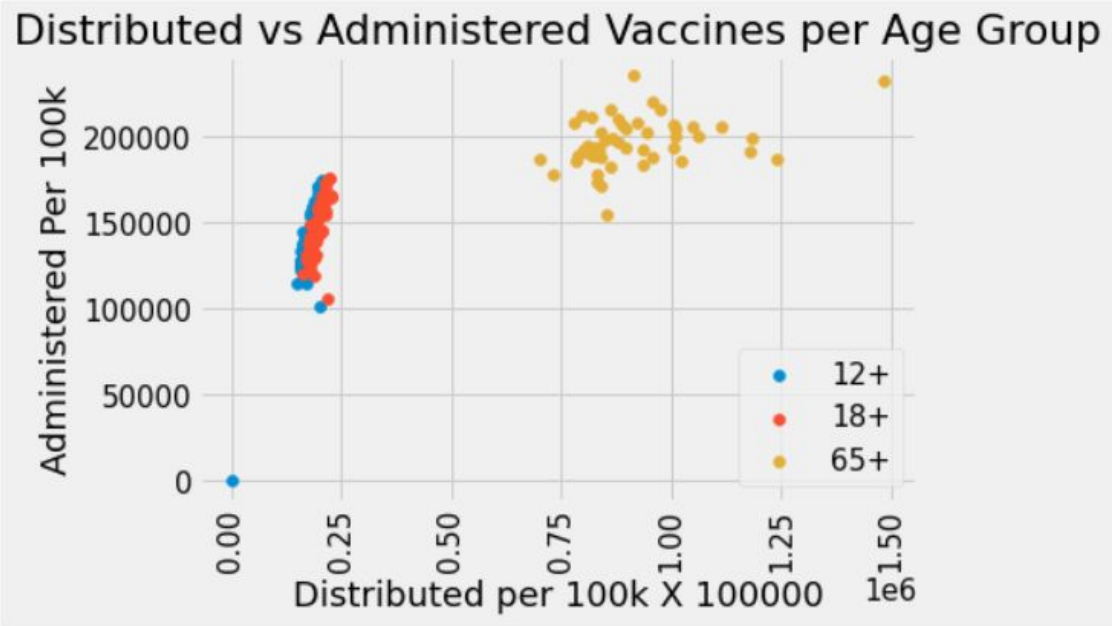
```
[ ] import plotly.graph_objects as go
    vaccines = vaccines.sort_values(by='Date', ascending = False)
    vaccines_lastDay = vaccines.head(51)
    vaccines_lastDay


    fig = go.Figure(data=go.Choropleth(
        locations=vaccines_lastDay['Location'], # Spatial coordinates
        z = vaccines_lastDay['Series_Complete_Pop_Pct'].astype(float), # Data to be color-coded
        locationmode = 'USA-states', # set of locations match entries in `locations`
        colorscale = 'Blues',
        showscale = True,
        colorbar_title = "Percentage Vaccinated",
    ))

    fig.update_layout(
        title_text = 'Vaccine Percentages by State',
        geo_scope='usa',
    )

    fig.show()
```

[ ]

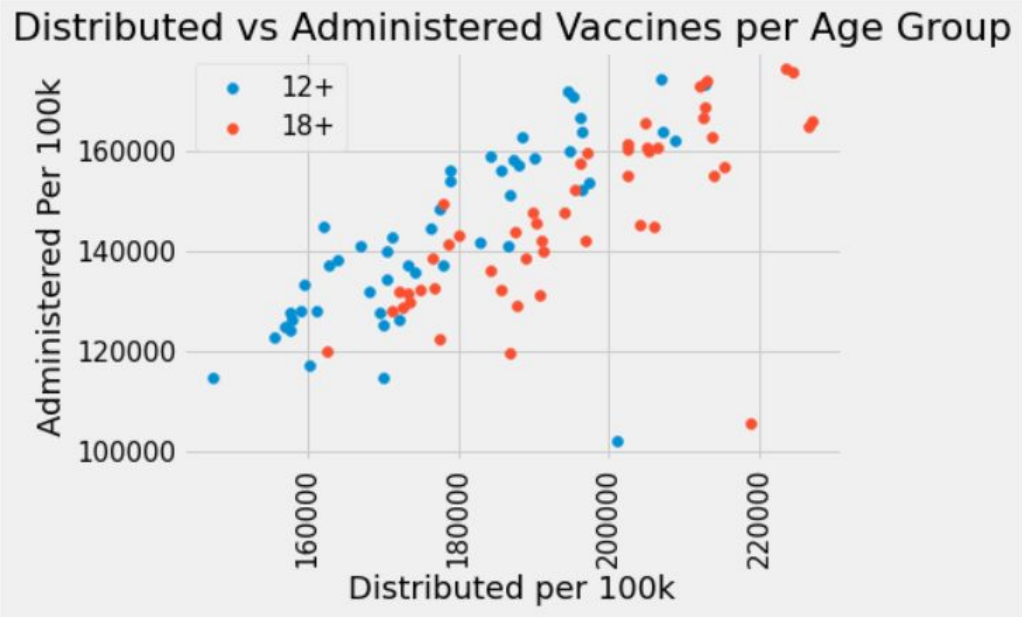


Graph 7: This map gives a nice color coded view of each state where it can be seen what states are doing well. The darker blue means more people are vaccinated, while the lighter means less are vaccinated.

```
plt.scatter(vaccines_lastDay['Distributed_Per_100k_12Plus'],vaccines_lastDay["Admin_Per_100k_12Plus"])
plt.scatter(vaccines_lastDay["Distributed_Per_100k_18Plus"],vaccines_lastDay["Admin_Per_100k_18Plus"])
plt.scatter(vaccines_lastDay["Distributed_Per_100k_65Plus"],vaccines_lastDay["Admin_Per_100k_65Plus"])
plt.xticks(rotation = 90)
plt.legend(['12+', '18+', '65+'], loc = 0)
plt.xlabel("Distributed per 100k X 100000")
plt.ylabel("Administered Per 100k")
plt.title("Distributed vs Administered Vaccines per Age Group");
```



```
[ ]  where_not0 = (vaccines_lastDay['Distributed_Per_100k_12Plus'] > 0)
     where_not0 =vaccines_lastDay[where_not0]
     plt.scatter(where_not0['Distributed_Per_100k_12Plus'],where_not0["Admin_Per_100k_12Plus"])
     plt.scatter(where_not0["Distributed_Per_100k_18Plus"],where_not0["Admin_Per_100k_18Plus"])
     plt.xticks(rotation = 90)
     plt.legend(['12+', '18+', '65+'], loc = 0)
     plt.xlabel("Distributed per 100k")
     plt.ylabel("Administered Per 100k")
     plt.title("Distributed vs Administered Vaccines per Age Group");
```



4. Conclusion

Throughout this analysis we have learned much about the vaccine and the different rates of vaccintion versus location and vaccination versus age range. We began the analysis by looking into the ditribution rates versus location. The orginal story started by lookinig at the total distribution based on location. Here we saw California as having the most distributed and Wyoming being the least amount distributed. This caused the question of if this was due to population, because California has a population of about 40 million and Wyoming only has a population of 600,000. SO then we loked into total ditribution by 100K. This created a better analysis due to the fact that we could compare apples to apples. We then saw that California was near the middle and DC had the most distributed per 100k with Wyoming still having least distributed. This led us to question the percentage of each vaccine used per each state. We created a bar chart that showed the amounts of each vaccine give out. Here we saw Pfizer was the most taken, then Moderna, and then J&J. After looking at the breakdown of each vaccine that was given out we questioned the percentage of fully vaccinated within each state. We first started by creating a bar graph to show this. This analysis had shown us that Vermont has the highest percentage of fully vaccinated and West Virgina has the lowest perentage. Once looking at the bar graph we decided to create a map graph to create more of a visal in vaccinated percentage by state. This ultimately was a great visual for the eye and backed up what we had found within the bar chart. This then made us question the timleine of the vaccine. Here we just created a simply timeline with a breakdown of each vaccine. Here we can see similar results with Pfizer taking off and J&J remaining low, most likely due to the problems that had occured during the pandemic with J&J. We then strted to question age range and distribution. We had created a scatter plot that showed the distribution per 100K based on age group versus administered by 100k. Here we can see as distribution increases, so does distribution. Going along with age group we decided to look at the timeline of fully vaccinated by age group. here we can see 12+ have a very rapid spike around May 2021 and 65+ seems to have the highest percentage over time. Overall we have learned a great amount about the vaccine based on location, age range, and percentage of fully vaccinated. We would be interested in learning more. Such as does political party have an affect of vaccine percentages? Does gender have an affect of percentage of fully vaccinated? What is the correlation with fully vaccinated that have recieved the booster shot? All this and more.