In [ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

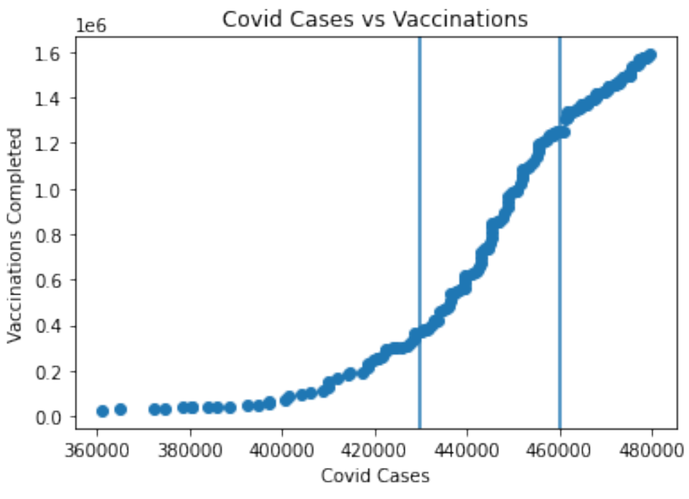
In [ ]:
```
!pip install pandas
!pip install matplotlib
!pip install seaborn
```

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.dates as mdates

path = '/content/drive/My Drive/Schmidt Futures/Data/LA_singlestate_timeseries.csv'
df = pd.read_csv(path)
```

B.1 Data Analysis

1) Graph the relationship between actual covid cases and vaccinations completed.

In [ ]:
```
x_axis = df['actuals.cases']
y_axis = df['actuals.vaccinationsCompleted']

plt.scatter(x_axis, y_axis)
plt.title('Covid Cases vs Vaccinations')
plt.xlabel('Covid Cases')
plt.ylabel('Vaccinations Completed')
plt.axvline(x=430000)
plt.axvline(x=460000)
plt.show()
```



1 cont.) Are there any salient trends or strong relationships? For which time periods do these trends exist?

I've divided the graph into three periods.

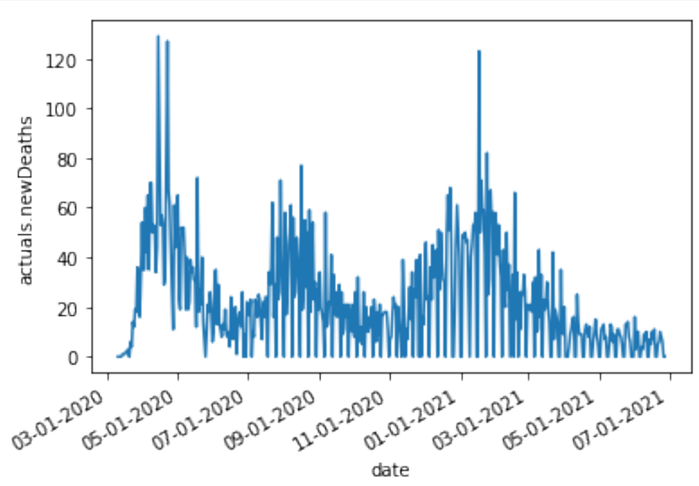
The first period occurs from 1/14/2021 - 2/27/2021, up to 430,000 cases. There is relatively little rise in vaccinations as Covid cases progress.

The second period occurs from 2/28/2021 - 5/3/2021, corresponding to a range of 430,000 cases to 460,000 cases. There's a strong positive correlation, as Covid cases increase so do vaccinations completed.

The third period occurs from 5/4/2021 - 6/25/2021, from 460,000 cases and beyond. Here, the relationship between Covid cases and vaccinations is not quite as strongly correlated.

2) Graph a time series of new covid deaths.

In [ ]:
```
df['date'] = pd.to_datetime(df['date'])

y_col_name = 'actuals.newDeaths'
x_col_name = 'date'

# build the figure
fig, ax = plt.subplots()
sns.lineplot(data=df, x='date', y='actuals.newDeaths', ax=ax)

# assign locator and formatter for the xaxis ticks.
ax.xaxis.set_major_locator(mdates.AutoDateLocator())
ax.xaxis.set_major_formatter(mdates.DateFormatter('%m-%d-%Y'))

# put the labels at 45deg since they tend to be too long
fig.autofmt_xdate()
plt.show()
```



2 cont.) Are there any salient trends you can see? For which time periods do these trends exist?

There are three major spikes.

Spike #1: March 2020 - June 2020

Spike #2: August 202 - September 2020

Spike #3: December 2020 - March 2021

B.2 Open-Ended Questions

1. **Why do you think there is very limited data on hospital and ICU bed usage in Pennsylvania before mid-May 2020? Limit your answer to 1-2 sentences.**

   COVID Act Now ingests data from the HHS. The article below states that as of July 22, 2020 just 28% of Louisiana hospitals were reporting through the new HHS system.

   https://www.cnbc.com/2020/07/22/us-hospitals-scramble-to-adopt-new-hhs-coronavirus-data-system-some-states-see-data-blackout.html

1. **Given what you know about covid, refer to your time-series graph to explain which societal factors, if any, may have impacted the number of new covid deaths. Please hypothesize why you believe so. State any assumptions you have made. Limit your answer to no more than 2 paragraphs.**

   The spikes in COVID deaths in Louisiana closely mirrors national trends, although Lousiana has been hit particularly hard.

   Spike #1 (March 2020 - June 2020) reflects a time period when cases were soaring nationwide, and little could be done besides social distancing. Key events include the LA stay at home order expiring on May 15th 2020, school closures on March 13 2020, and the absence of a mask mandate in this period. (As far as I could tell) Additionally, LA has issued no travel restrictions to date and the popularity of New Orleans as a spring break destination may have spurred outbreaks and deaths throughout the state.

   Spike #2 (August 2020 - September 2020) could be influenced by the opening of higher education institutions, although I struggled to find aggregate date on this. Spike #3 (December 2020 - March 2021) is likely to have been spurred by Thanksgiving travel, again with LA having no travel restrictions. The end of this spike can likely be attributed to the availability of vaccines, with all adults having access to vaccination as of March 28 2021.

1. **What are some possible limitations of this dataset? Limit your answer to 2-3 bullet points.**

- relationships between vaccinations and outbreaks / deaths may be difficult to interpret without regional granularity
- Lousiana was late in achieving widespread reporting to the HHS system, with resulting gaps in this state's dataset

1. **Briefly explain 1-2 recommendations you would relay to the team that maintains this dataset to overcome any of the limitations you mentioned above. Please keep your recommendations related to improving data quality, and not improving Louisiana's overall covid response. We are looking for how well you can explain and persuade through writing. Limit your answer to no more than 2-3 sentences.**

   To understand relationships between vaccinations and new cases / deaths, the available county level data from COVID act now should be utilized. This would allow for a more targeted re-opening plan, and could improve the response of residents at the county level, since the data basis for the decision would be specific to their needs and attitudes.