COVIDMINDER Bootcamp Notebook

Alex Chan

5/05/2020

Contents

Final Project: Submission Links

This should be the first section of your final project notebook. Fill out the following according to how you submitted your notebook!

- github repository: https://github.com/TheRensselaerIDEA/COVID-Notebooks (for reference)
- My github ID: *alexchan18*
- github issues addressed by this work: `#27`
- Github branch name of my submitted notebook: *feature-27*
- link to merged notebook (post these to LMS!:
  - https://github.com/TheRensselaerIDEA/COVID-Notebooks/blob/feature-27/MATP-4400-FINAL/covid-notebook-Chana6.Rmd
  - https://github.com/TheRensselaerIDEA/COVID-Notebooks/blob/feature-27/MATP-4400-FINAL/covid-notebook-Chana6.pdf

Overview & Problems Tackled

On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic. The number of cases reached a point where the entire world needed to take action. In the United States, state governors began closing down states to reduce the rate of infection. This meant that all citizens had to remain in their homes until further notice. Fast forward 6 weeks, the citizens of the United States are eager to leave their homes and return to normal life. The decision to do open up the country will be made based on the rate of infection and the increase in new cases. I looked at the data for this in New York State to see if Governor Cuomo should make the decision to open.

The first problem was visualizing the number of new deaths per day.

The second problem was visualizing the number of deaths per 100k.

Data Description

I used the data for every county in the United States that was updated daily

From this large dataset, I extracted the data for every county in New York State

Then, I had to extract the FIPS code, confirmed cases, and deaths for each county and store this information in a data frame because the other information was not of use.

Results

The graphs show that there is a significant decrease in new deaths per day in the counties that are located below Weschester. These counties were largely affected by COVID-19 and the number of cases increased significantly. This could be due to the high population density of these areas. As for the counties north of Weschester, the number of new deaths were very small and the changes between days were either positive or negative. The only county that showed a significant number of deaths each day was Buffalo, located in western New York on the NY/Canadian border.

Problem 1

We need to visualize the number of new deaths per days in New York, separated by counties. The graph will be represented 5 times, every 3 days since April 23rd.

Methods

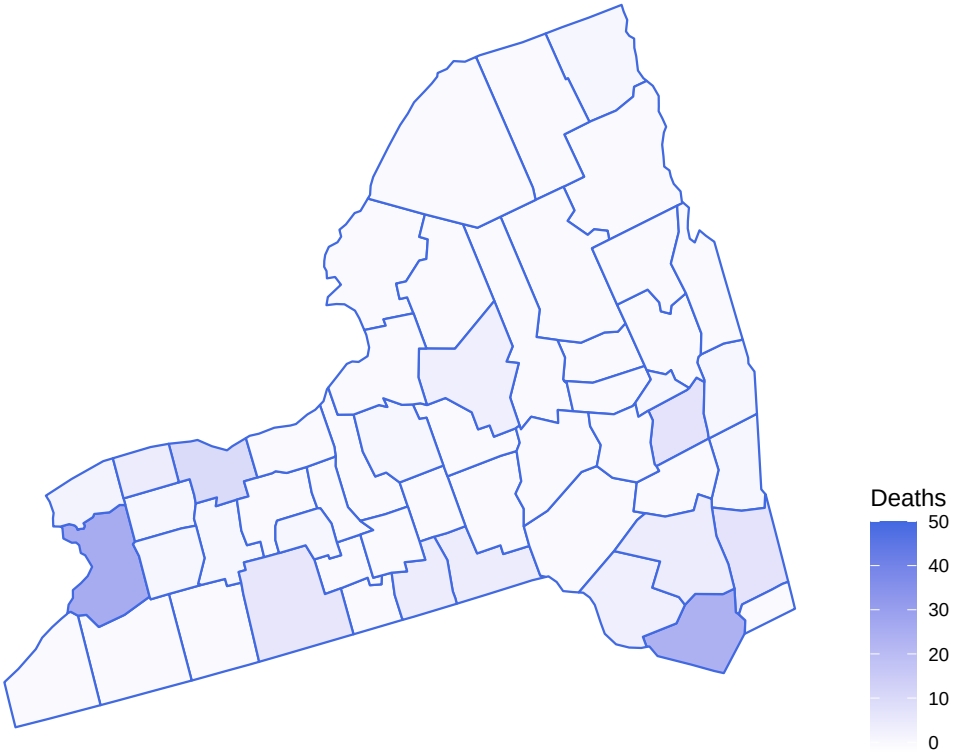
Take country data for April 23rd, April 26, April 29th, May 2nd, and May 5th. Extract data for only New York counties. Create graphs from data frames created through extraction. Then, find difference in number of deaths between days to determine number of new deaths per day.

For each day:

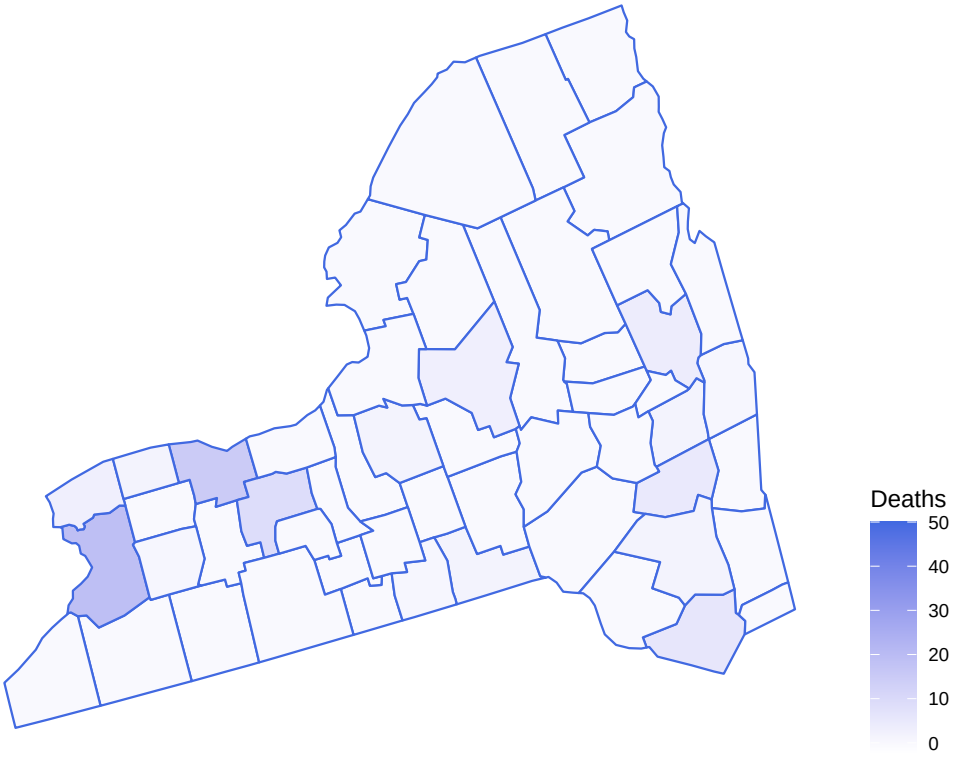
1. Read in dataset for desired date
2. Extract data for all rows with New York as state
3. Extract FIPS code, number of confirmed cases, and number of deaths for each county
4. Remove NULL rows from data frame
5. Determine number of new deaths by subtracting cumulative deaths of old day from new day

The Following 5 graphs show the number new deaths in each county north of Weschester

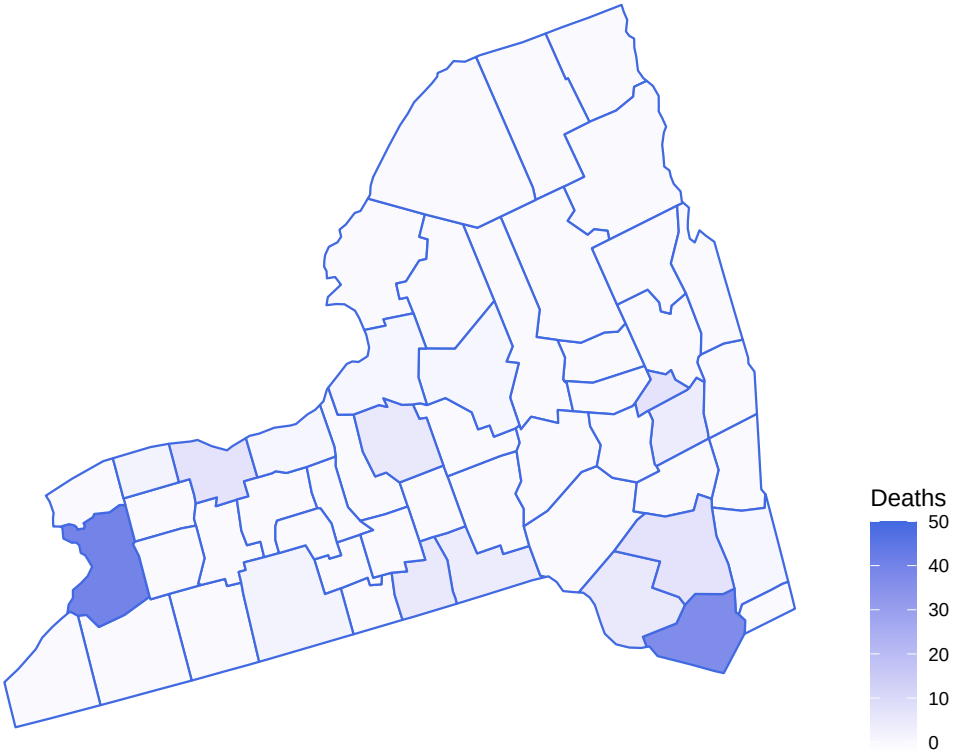
April 23rd



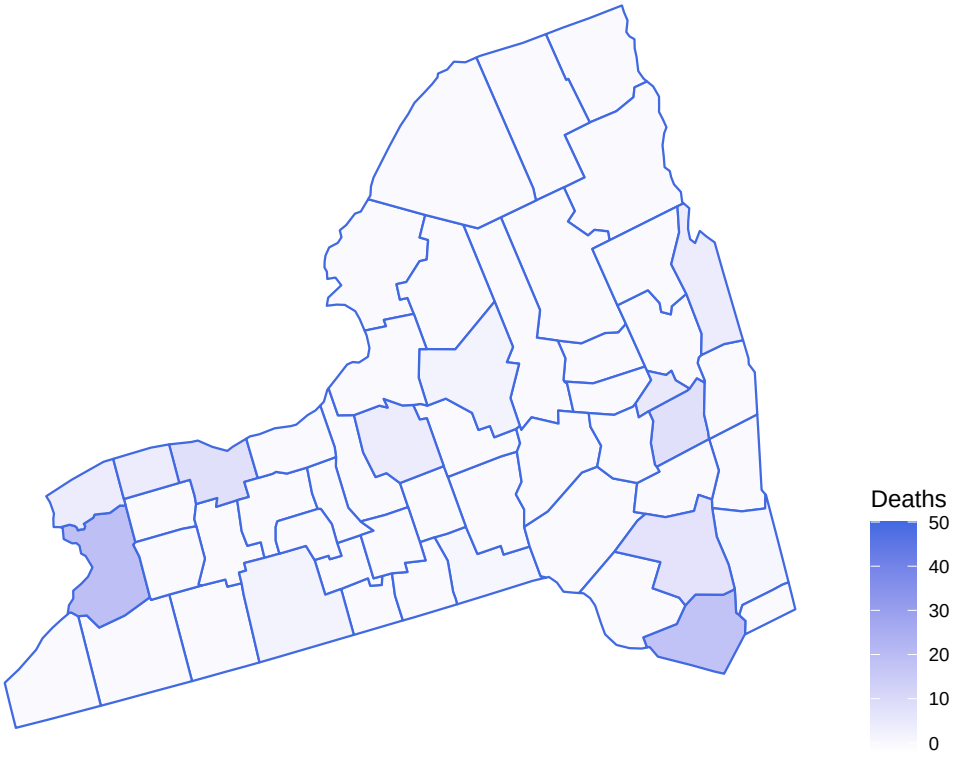
Aril 26th



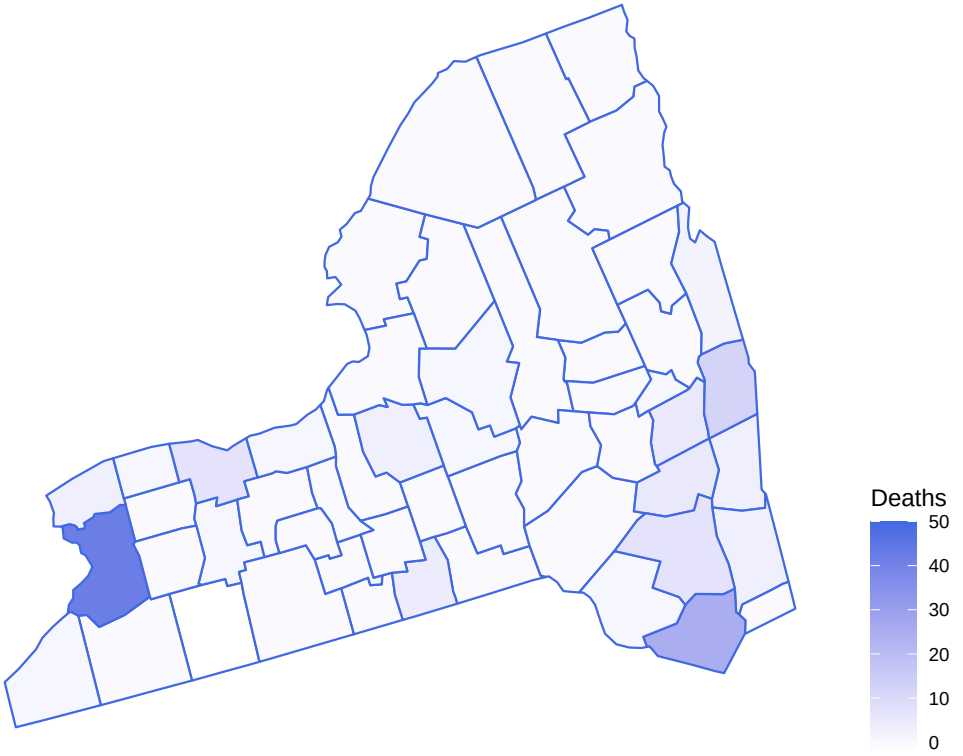
Aril 29th



May 2nd

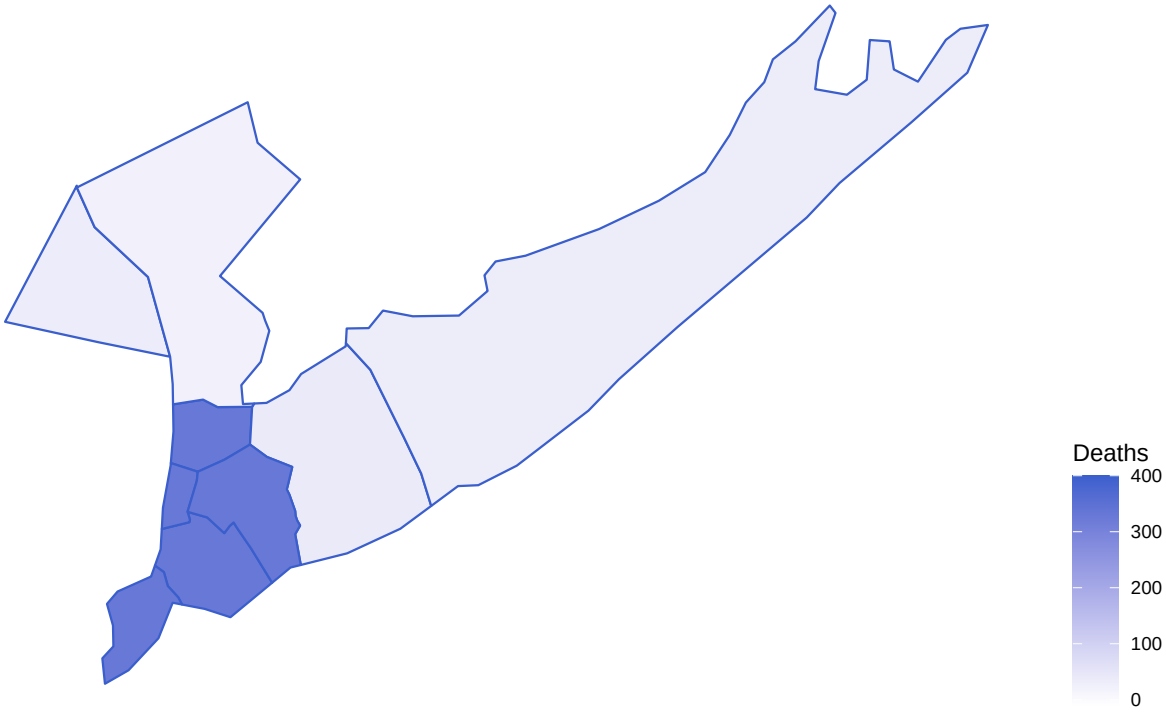


May 5th

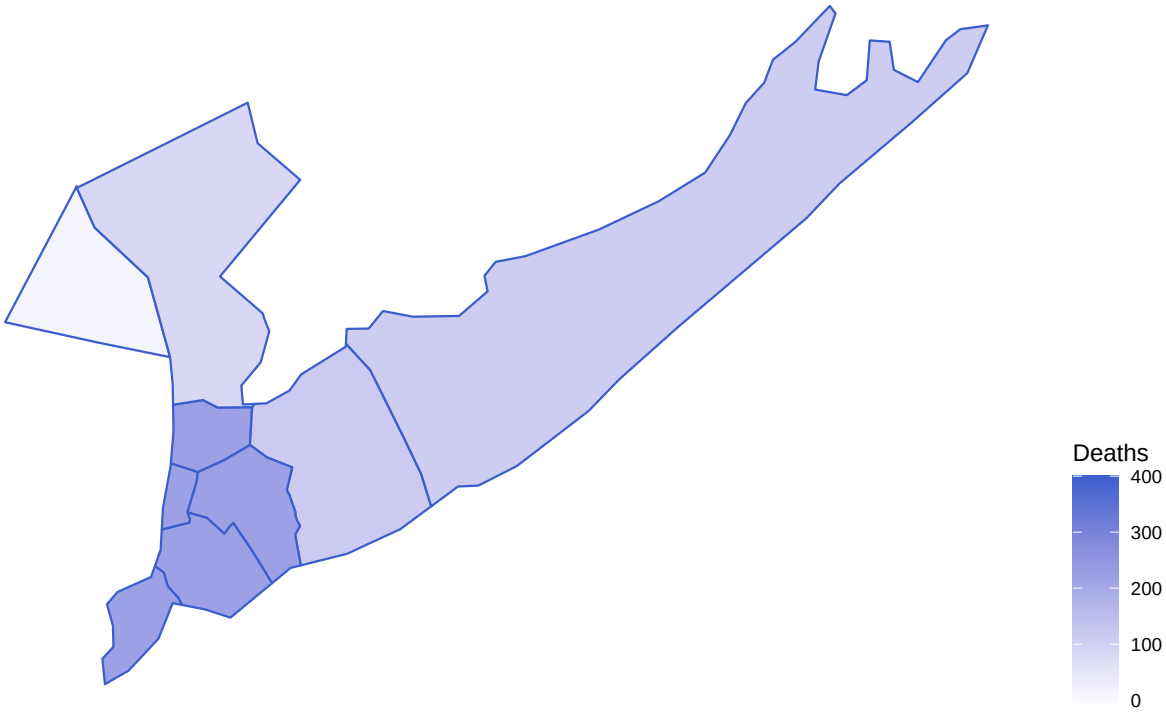


The following 5 graphs show the number of new deaths in the counties south of and including Weschester

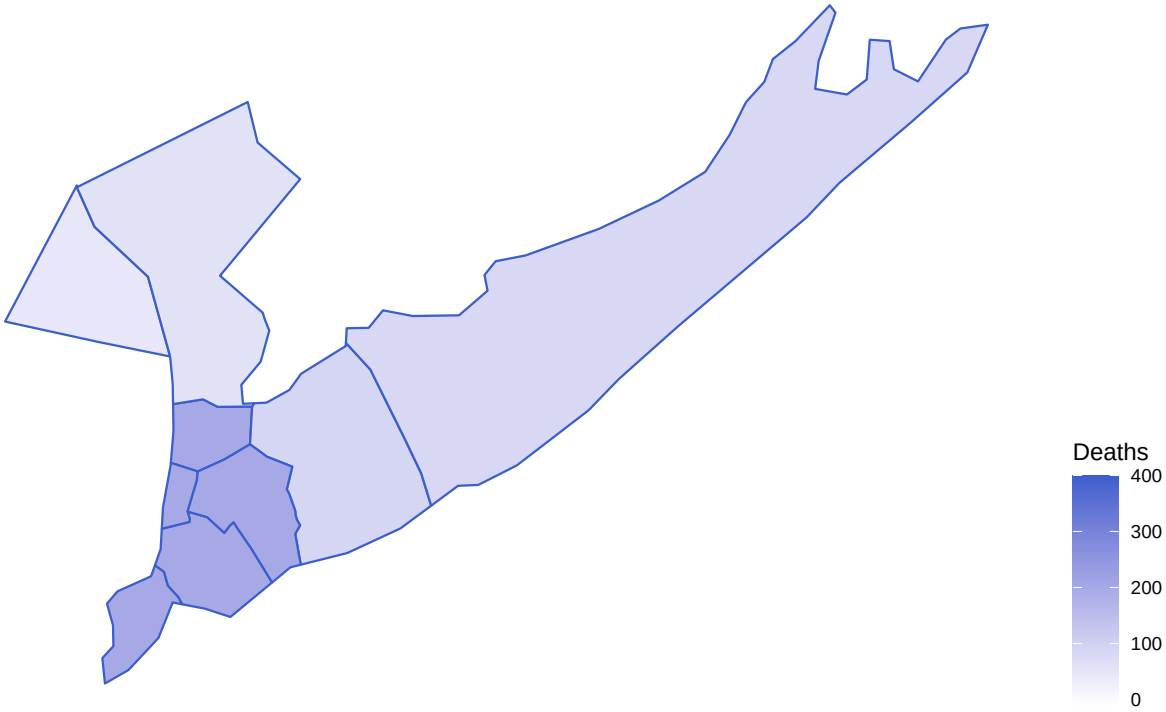
April 23rd



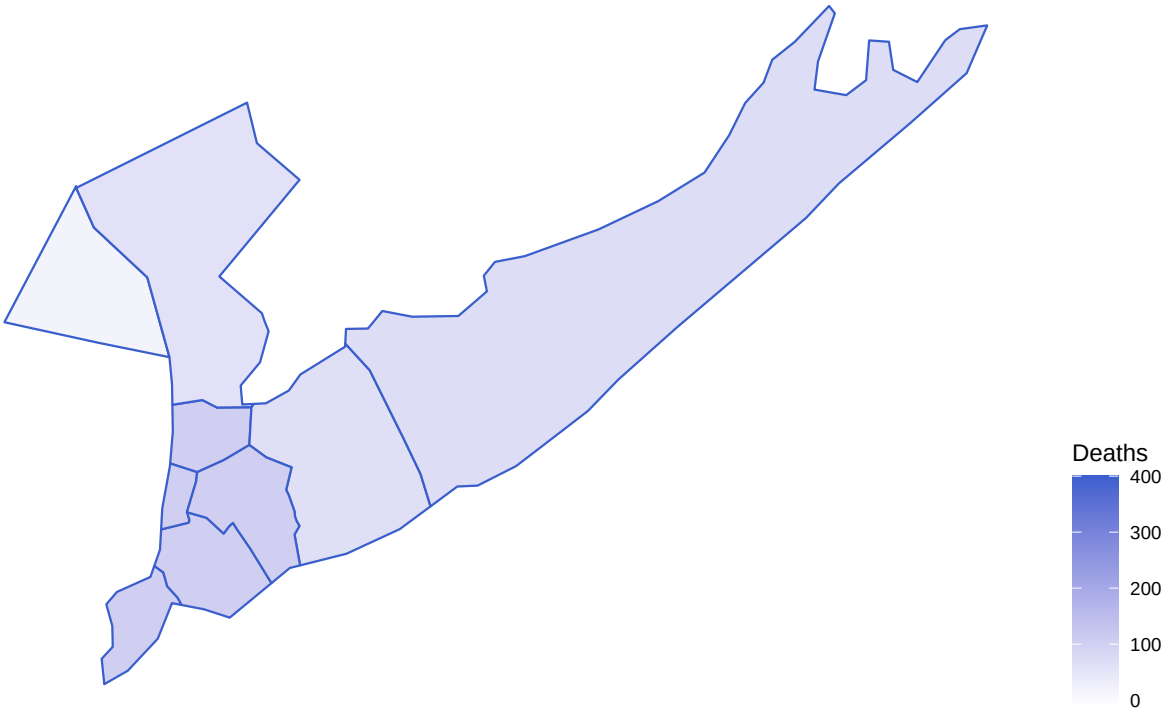
April 26th



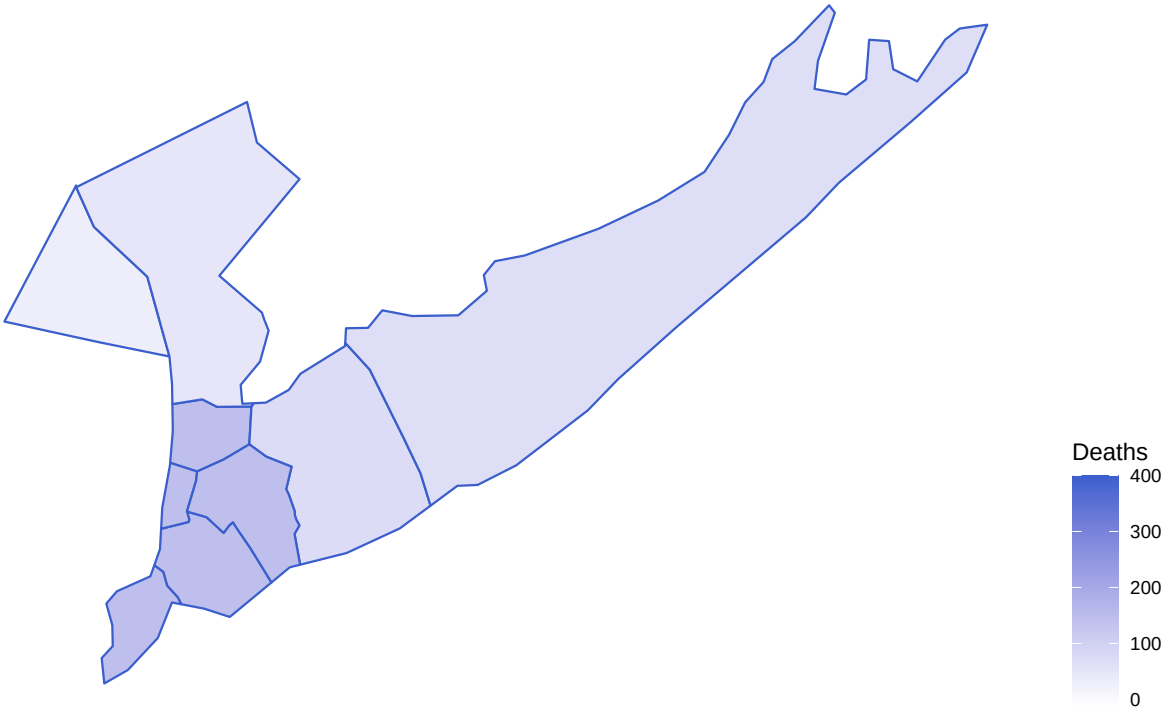
April 29th



May 2nd



May 5th



Note: When extracting the county data from the overall country data, the 5 bouroughs of New York City were grouped together under one county. I needed to divide the number of deaths in New York City by 5 to be able to map the counties in the plot_usmap function

Results

For the counties north of Weschester, all but 2 had less than 10 new deaths every 3 days, showing that these areas were not largely impacted by COVID-19. The 2 counties that had a relatively high number of deaths during these days were Buffalo and Rockland. However, the range for the number of new deaths in these 2 counties varied, from around 50 to as low as 20.

For the counties south of Weschester, the number of new deaths every 3 days started very high, but gradually decreased over time. The 5 boroughs in particular had very high number of deaths, starting at 400 on April 23. Over the course of the past 2 weeks, however, the number of new deaths decreased, especially in the Suffolk and Nassau counties.

Discussion

These graphs show that the number of new deaths has decreased since New York was declared a hot zone. In terms of reopening, the decision should be to remain closed until number of new cases is small. The number of new deaths just shows how fatal the virus is. Until a vaccine is discovered, the number of deaths will always be greater than 0 because people who have the virus are not safe from it.

There are some strengths to these results. By mapping out each county in New York we can determine which counties are vunerable to fatalities and therefore pay more attention to slowing down the rate of infection. Clearly, the areas with high population densities are suffering the most, while the counties north of Westchester have little to no fatalities.

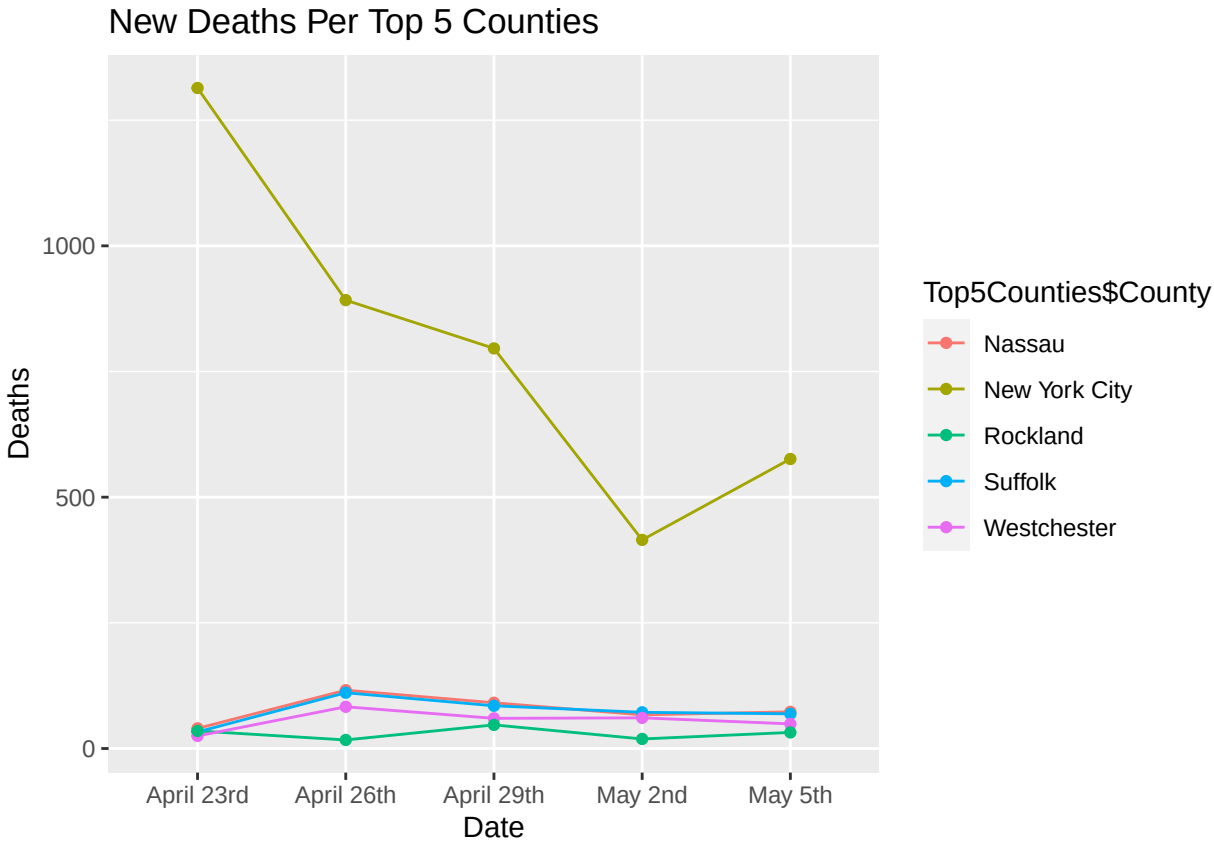
There are also some limitations to these results. The graphs do not account for population density, which could help explain the reason for such high numbers of deaths in the souther counties of New York. New York City has 20,000 people per square mile which would clearly lead to an increase in the number of cases and furthermore the number of deaths.

Problem 2

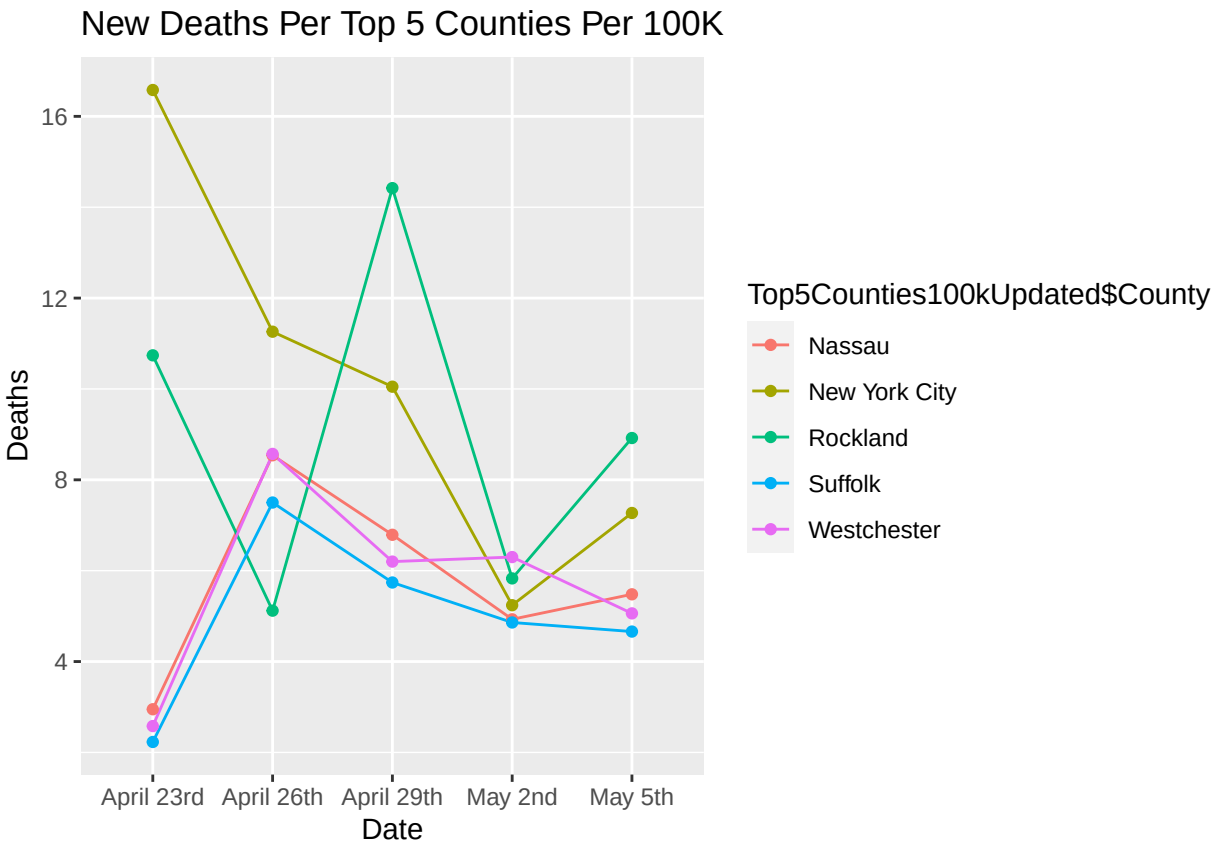
Similar to the first problem, except we are graphing deaths per 100k. Normalizing the data this way will allow us to see a better visualization of the data.

Methods

- I plotted a line graph of the new deaths per day in the top 5 worst counties. This included New York City, Nassau, Rockland, Suffolk, and Westchester.
- Then, I normalized the data so that the graph represented deaths per 100,000 people. This was based on the populations for each county



- I had to extract the population data for each county, then divide it by 100,000.
- I took the number of deaths for each county for each day and divided it by these new numbers.



Results

Based on the first graph, it is easy to see that New York City completely skews the data. This is because the number of deaths was significantly larger than the remaining counties in the top 5. Nonetheless, the line for NYC slopes downward, meaning the number of new deaths decreases.

For the remaining 4 counties, the lines are relatively flat, showing the number of new deaths does not change significantly. However, this is only how it appears. Again, NYC skews the data so the shape of the lines for the 4 other counties appear to be flat.

Discussion

We can see that normalizing the data based on population showed us a better representation of the change in the number of deaths for the top 5 worst counties in New York. For NYC, the number of new deaths per 100k drops significantly over time. For the remaining 4 counties, the number of new deaths per 100k varies, but is on average around 6.

Summary and COVIDMINDER Recommendations

- Overall, what insights did you find about the COVID-19 epidemic in your analysis?
  - I discovered that the number of deaths due to COVID is steadily decreasing in New York, which was considered a hot zone recently. This is a good sign and shows that the work Governor Cuomo is doing to slow the rate of infection/death is really working.
  - I also realized that graphing data can be useful but also misleading. It is important to understand the scale of what is being represented and to know when to normalize data. It's easy to look at a graph and determine that a certain county is doing well because the slope of their line is close to 0. But normalizing the data will give us a better picture of the fatalities of a certain area.

- What recommendations do you have for COVIDMINDER for Data utilization, Analytics, Visualizations, User interface design, etc.
  - I would recommend creating an animation with a slider of a map of New York or even the entire country that includes the changes in cases/deaths/active.
  - Would you recommend that your analysis be included in COVIDMINDER? Why or Why not?
    * Yes, I would recommend that my analysis be included in COVIDMINDER. It is important to have visualizations of deaths in a certain area in order to decide where to focus attention. Also, when deciding, on a larger scale, when to reopen the country, it is important to look at the number of new deaths over time. Obviously, we would want to reopen when the number of new deaths is 0. This would mean a vaccine is available and those who are infected would have the chance to get cured.

References

None

Appendix

None