Visualizing the Opioid Epidemic in New York State from 2003 to 2019

Adelaide Bsharah, Marley Ferguson, Robbie Hoyler, Aashvi Shah

Northeastern University

ABSTRACT

With over 20 million people in the United States suffering from substance use disorders, drug addiction is a disease that deserves far less stigma, and far more research. Since 2015, the number of drug-involved overdose deaths has been on a sharp incline, doubling from about 52,000 deaths to more than 102,000 in 2021 [4]. In this paper we decided to look at New York state and support two domain tasks when examining opioid related deaths. We will first analyze which year since 2003 had the greatest number of opioid deaths per county. Each county's drug use trends change due to residents' socioeconomic status, access to healthcare, and public policy, so creating a visualization will assist policy makers and public health officials to identify patterns over time. We will then determine which counties in New York currently have the greatest total number of opioid-poisoning deaths. This visualization will allow those same officials to target their efforts towards the areas of the state that need the most assistance, and possibly determine characteristics that high risk counties share with one-another. Our first task is to display the number of opioid poisoning deaths per year for each county in an interactive bar graph. Our second task will be to compare the number of opioid deaths per year and overall totals between 2003-2019 between counties. It is important to support both domain tasks because they provide insight on which counties in New York state have been most affected by the opioid crisis, as well as trends/changes in the deaths in both the state and individual counties overtime.

1 INTRODUCTION

Within public health, visualizing data is an essential component to interpreting the severity of public health crises. This project aims to support public health officials from New York State in better understanding the progression of the opioid epidemic from 2003- 2019 per county. This can be done by analyzing the trends in the number of opioid poisoning deaths that have occurred. To do this, two domain tasks and subsequent task abstractions have been identified. The first domain task is: We want to see which year had the greatest number of opioid deaths per county in New York State. It is important to support this task because it will give public health officials insight on exactly when the opioid crisis initially took off within the state. We are looking purely at the number of deaths, but we know that the opioid deaths are just a fraction of the number of opioid users, so officials can use these numbers as indicators for increased drug use as well. Officials could use this to better understand why and how the opioid epidemic occurred. Additionally, it can provide insight on how effective certain public health policies (that were passed before 2019) were in addressing/limiting opioid related deaths over time. The task abstraction for this domain task is as follows: Display changes in the number of opioid poisoning deaths over time per county in New York State (Mid-level, Search, Look-up). The second domain task is: We want to see which counties in New York State have the total greatest number of opioid poisoning deaths. It is important to support this task because it will give public health officials insight on which New York counties have been hit the hardest by the opioid epidemic, and who requires the most aid. This visualization tool will be utilized by public health officials who want to better understand the progression of the opioid crisis in both the state overall and in specific counties. This visualization tool can also be made public for educational purposes to inform individuals on how the opioid epidemic is affecting their communities.

2 RELATED WORK

Related Work 1:
The first related work was found in a report from the Massachusetts Department of Health, An Assessment of Fatal and Nonfatal Opioid Overdoses in Massachusetts (2011 – 2015). In this report, data on opioid overdose is examined and then it is further analyzed based on the source of opioid. The first two figures in the report use a map to indicate the density of opioid deaths in each town in the state. The respective number of deaths are represented by the hue that was shaded in, with light pink representing 0 deaths per 100,000 citizens and the dark red hue representing greater than or equal to 15.4 deaths per 100,000 citizens. We adopted this map model for one of our visuals. In our visual, we have the map of New

York state which is then broken down by counties. The hue of the county's shading represents the total number of opioid deaths from 2003 to 2022. The number of deaths increases as the shade of blue gets darker. The report also further observed that the most deaths across all the counties were the highest amongst young males. This led us to adding another dataset that could give us information about the subgroups of the population of opioid related deaths. The new dataset includes subgroups of age and gender. The drop-down feature will allow us to zoom in specific age groups for each gender to confirm if this pattern is accurate in New York and if it holds in New York City and out of the city[1].

Related Work 2:
In Christopher R. Herrmann's, Street-Level Spatiotemporal Crime Analysis: Examples from Bronx County, NY (2006–2010), it is discussed how geographic clusters of information, in his case crime-scale analysis, can be confusing to the reader. Herrmann goes through the various ways to address and understand these visuals. He includes graphics that show the various "crime hotspots" using heat maps to highlight patterns and trends between geographic locations and types of crime [3]. We decided to incorporate a similar type of map to our visualization, however we expanded from just the Bronx County to the entire state, with each county being represented by a color signifying the number of deaths that have occurred due to opioid-related causes. Something that stuck out beyond the heat map however was the way Herrmann used a line chart that connected to the data [3]. He used it to show overarching trends over time and give an overview of the swath of data in a simplified way. We thought this was a great way to help users get a better overall grasp on the data, so we decided to incorporate a similar idea, but instead of using a line chart, we decided to use a bar chart. For the map of New York counties, we want users to be able to click on a county and be shown a bar graph representing the number of deaths per each individual year from 2003 to 2022, instead of just the overall total.

3 USE CASE

In order to highlight the severity of the opioid crisis that has ravaged New York State, a visualization tool will be designed for public health officials who want to track the number of opioid related deaths that have occurred over the last two decades per county and by human characteristics. The tool will utilize data from New York State's public health database titled, "Vital Statistics: Opioid-Related Deaths by County: Beginning 2003." The visualization will have multiple components, starting with a drop down menu including all the counties within New York. Once a user selects a county, an interactive map of the New York counties appears, with the selected county being highlighted in pink. Hovering over any county will highlight its area in pale yellow on the map and display a tooltip with its name and total number of opioid deaths that occurred within that county. Simultaneously, a bar chart will display the total number of opioid related deaths (per year) that have occurred in the county selected from the dropdown menu from 2003 to 2019. This visualization tool will allow public health officials to visualize trends in the frequency of opioid-related mortality in a particular county as well as identify areas in dire need of drug/overdose prevention assistance, and push for the necessary policies to be passed by local and state governments. Additionally, the tool could be made publicly available for educational purposes, thereby informing the public of the current status of the opioid crisis and how it is affecting their community.

4 Data

Initially we were looking at two datasets, both owned and published by the New York State Department of Health. Data from both data sets were collected starting in the beginning of 2003. In terms of ethical considerations, both datasets are free of any personally identifiable information, making them compliant with privacy and HIPAA regulations. The New York State Bureau of Vital works in cooperation with the New York State Department of Health to provide information (including deaths), which was then used to create these datasets. Both datasets are monitored by the Office of Quality and Patient Safety. The final dataset we ended up using, Opioid-Related Deaths by County, seems to have no missing data values or outliers/unexpected values [2]. This data set is very limited and only includes three attributes: year, county and number of opioid related deaths. We would have liked to visualize other demographic information associated with this dataset including race, sex, and age, but this may have caused privacy concerns. The manner in which the data was input remains consistent throughout the dataset, making it very cohesive. Overall, this dataset is very clean and does not require additional cleaning on our end. In terms of biases in the dataset, this information must have been collected at the county level, and later conglomerated by New York State. It is possible that inconsistencies in data collection among numerous local departments may skew numbers for certain counties. The other dataset we were looking at but didn't end up using was the

Opioid-Related Deaths by Age Group dataset. Instead of grouping the number of deaths by county, the opioid deaths were categorized by attributes. There seems to be some missing data values, but they are minimal. For the race and ethnicity attribute, there's "not stated". For the age group attribute, there's "unknown". There are no outliers/unexpected values (total opioid poisoning deaths increased dramatically, but that is expected). This data has slightly more attributes including: year, region, race/ethnicity, sex, age group and opioid poisoning deaths. However, the data is somewhat limited, particularly the race/ethnicity category, which only includes Black Non Hispanic, White Non Hispanic, Hispanic and Not Stated. Additionally, there seems to be some spacing issues within the age group category. The manner in which the data was input remains consistent throughout the dataset, making it easy to interpret. In terms of biases in the data, it is possible that the limited range for representing race/ethnicity may skew the numbers inaccurately to one category. By not having proper representation, the data does not provide an accurate picture of the most afflicted groups. In terms of data cleaning for the Opioid-Related Deaths by County dataset, the data within the County column was converted to its proper data type (string). In terms of the Opioid-Related Deaths by Age Group dataset, the data within the Region, Race or Ethnicity, Sex and Age Group columns were converted to proper data type (string). All spaces around the string data types were removed using the strip() function. No additional attributes were added to either data set. Missing values were not replaced seeing as they still represent relevant data value (would skew data if imputation was used) In terms of data cleaning for the Opioid-Related Deaths by County dataset, the data within the County column was converted to its proper data type (string). In terms of the Opioid-Related Deaths by Age Group dataset, the data within the Region, Race or Ethnicity, Sex and Age Group columns were converted to proper data type (string). All spaces around the string data types were removed using the strip() function. No additional attributes were added to either data set. Missing values were not replaced seeing as they still represent relevant data value (would skew data if imputation was used).

5 DESIGN PROCESS

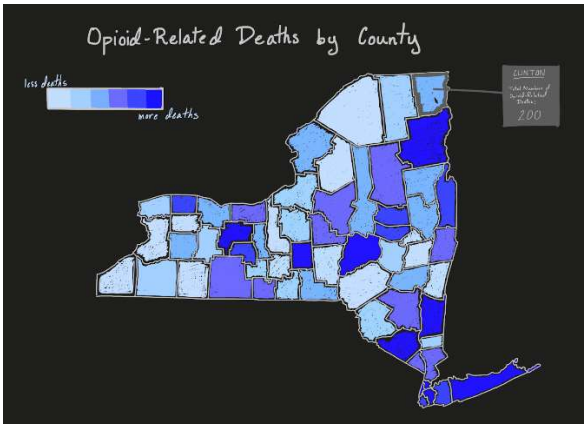

Figure 1: The various counties in New York and the number of total opioid-related deaths that have occurred between 2003 and 2022 using our first dataset. The color of each county correlates to how many deaths have been recorded, with the darkest color signifying more deaths, and the lighter colors signifying fewer. By hovering over a county, it will highlight and a pop-up will appear showing the county name and total number of deaths.

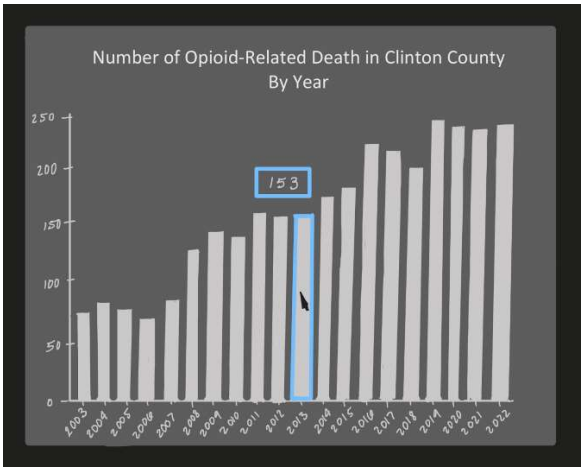

Figure 2: Relates to Figure 1 showing the counties of New York because when a county is clicked on, it will change this bar graph to represent the data for that specific county. Instead of just showing the total number of deaths, this bar graph will show the number of deaths that occurred each year, and if a bar is highlighted, that specific number appears.

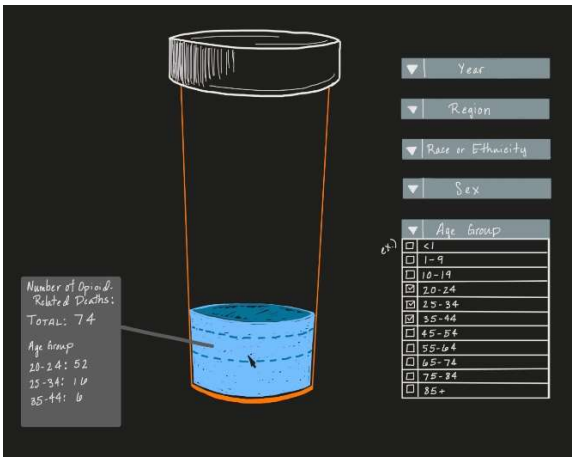

Figure 3: Uses the second dataset about opioid-related deaths in New York. It contains information such as the year, region, race, sex, and age, as well as number of deaths. Each drop down menu will contain various characteristics that can be selected, such as specific age- groups. When selected, the pill bottle will either increase or decrease in volume proportional to the number of deaths that fit the selected characteristic. If the volume of the bottle is hovered on, a pop-up will show the total number of deaths, as well as the number of deaths associated with each selected characteristic.

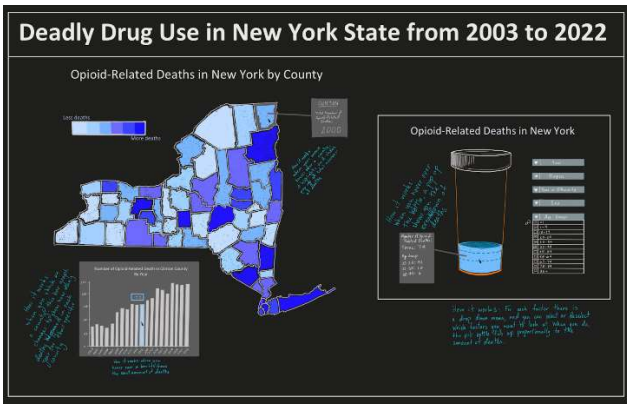

Figure 4: A combination of all the previous figures, as well as some added notes about interaction. This is a representation of what the final visualization would appear similar too, with both datasets being represented and user interaction. It should be noted however that any "data" being shown isn't accurate, and can be considered "placeholder" data. Each part of the visualization relates to the other as they are representing the same topic, and the map of New York and the bar graph are intrinsically connected, because when a county on the map is selected, the graph adjusts accordingly. All 3 visuals use area as a mark, as well as vertical and horizontal position as a channel. The New York counties visual and the characteristic visual both use color as a channel, with the colors representing different groups. They also use size and area as a channel, and the bar graph also uses size, but as length. The length of the bars is important as longer/taller bars represent more deaths. The map of New York allows users to get a better geographical idea of where these events are located, the bar graph shows trends over time, and the pill bottle is a creative visual to engage the user

After finishing our sketches and beginning the actual implementation, we decided against using the second dataset including the attributes, meaning Visual 3 would no longer be included in the design. Due to time constraints and issues with linking the two datasets, we removed the stacked bar chart in the shape of a pill bottle and focused instead on the count map.

6 FINAL DESIGN

Our final design shifts the focus entirely on the map of New York counties and the number of deaths over time and in total. We decided that to make it easier for users to locate the county they wanted to look at, we would have users use a drop down menu to select a county instead of needing to know its location geographically and click it. Otherwise the design is quite similar, with the counties being colored a shade of blue associated with the number of total deaths. A small color legend beneath the map shows the darker blue counties have a higher number of opioid deaths. The bar graph automatically changes to show the data for whichever county the user selects from the drop down menu. The largest modification from the initial idea we had was the decision to leave out the stacked bar chart of all the attributes. This visualization, while helpful in potentially identifying personality or demographic attributes that make a person more at risk for opioid overdose, wasn't linked to our domain tasks. Its potential inclusion was meant to be engaging for the user and to allow them to explore the data, so removing it from the final visualization didn't impact the task abstractions we set out to complete.

7 DISCUSSION

Using some preliminary user testing, we wanted to look at how effective our visualizations and webpage actually were at completing the domain tasks. We wanted to test the bar chart specifically to see if it was an effective means at conveying the change over time for total deaths per year. By seeing how long it took a user to identify the year with the highest number of deaths, we were able to interpret this as whether or not the chart is easily understood, while also showing focus on the yearly change. We wanted to test the map visualization and the corresponding

color legend to understand how easily a user was able to identify that darker colored counties had more opioid deaths while lighter colored ones had less. We can assume that the shorter the time, the more intuitive the color scale is.
Immediately our test user knew where to look to find the answers to the questions we asked them. We also wanted to test whether or not a user would figure out that the map of New York counties was interactive with minimal prompting from us. We want the user to understand everything they can do with our visualization, so this was essentially testing whether or not we needed to add instructions explicitly stating you can hover (and later click) on a county. Our final part of user testing was testing how easily a user could find a county in New York State, even if they aren't previously familiar with the state. We need our visualization to be accessible to everyone, so we wanted to ensure that by using the hover feature a user could identify a specific county. From this test we actually decided to implement the drop down menu instead of requiring the user to hover over each county until they find the correct one they want to look at.

When finding the year with the highest number of deaths they knew to look at the bar chart and go to the tallest bar. From there they were able to identify the specific year. When asked which county had the highest number of total opioid deaths, they immediately scrolled down to look at the map, and within a second knew that the darker colored counties meant a higher concentration of deaths, so they went to Suffolk county first, which also shows that the user intuitively knew that they could hover over the map to get more information. Of course this may be affected by both the age of the user, and the fact that they are also in our class so they were expecting to have to interact with the webpage. The longest part of our test was having the user find a specific county and the number of total opioid deaths because the user had to hover over multiple parts of the map. They were successful with finding the county however, and they knew exactly how to get the information we asked them to find. From this test we actually decided to implement the drop down menu instead of requiring the user to hover over each county until they find the correct one they want to look at.

8 CONCLUSION

Tackling the opioid epidemic in New York state is a daunting task due to the magnitude and severity of the crisis. Over time, each county has experienced fluctuations in the death toll based on a variety of outside factors, such as public health policy, income, access to healthcare, and a variety of other influences. Thus, by creating this interactive visualization, policy makers can determine which areas of New York need the most assistance, along with accessing historical data from 2003 to 2019 to examine how the crisis has evolved with time. The officials can then adjust how they relocate their resources to most efficiently tackle the crisis. Additionally, policy makers can examine laws in the afflicted areas alongside the time series death toll data to better understand which were successful and which failed to alleviate the crisis. By creating this visualization, the information about the impact of the opioid crisis can be navigated in a convenient and accessible manner from people of all backgrounds and skill levels.

9 REFERENCES

[1] C.Baker, M.Bharel, M.Sudders, and K.Polito. An Assessment of Fatal and Nonfatal Opioid Overdoses in Massachusetts (2011 – 2015). Massachusetts Department of Public Health, pp.7-9, Aug.2017

[2] Department of Health Agency. (2022, May 11). Vital statistics: Opioid-related deaths by county: Beginning 2003: State of New York. Vital Statistics: Opioid-Related Deaths by County: Beginning 2003 | State of New York. Retrieved April 18, 2023, from https://health.data.ny.gov/Health/Vital-Statistics-Opioid-Related-Deaths-by-County-B/sn5m-dv52

[3] Herrmann, C. R. (2012). Street-level Spatiotemporal Crime Analysis: Examples from Bronx County, NY (2006–2010). Crime Modeling and Mapping Using Geospatial Technologies, 73–104. https://doi.org/10.1007/978-94-007-4997-9_4

[4] U.S. Department of Health and Human Services. (2023, March 31). *Drug overdose death rates*. National Institutes of Health. Retrieved April 18, 2023, from https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates#:~:text=More%20than%20106%2C000%20persons%20in,drugs%20from%201999%20to%202021

Appendix A. Data Abstraction:

Opioid-Related Deaths by County:

- Year: Attribute - Ordered, Ordinal, Sequential
- County: Attribute - Categorical
- Opioid Poisoning Deaths: Attribute - Ordered, Quantitative, Sequential

Opioid-Related Deaths by Age Group:

• Year: Attribute - Ordered, Ordinal, Sequential
• Region: Attribute - Categorical
• Race or Ethnicity: Attribute - Categorical
• Sex: Attribute - Categorical
• Age-Group: Attribute - Categorical
• Opioid Poisoning Deaths: Attribute - Ordered, Quantitative, Sequential

Appendix B. Task Abstraction: For our first domain task, we want to observe which year had the greatest number of opioid deaths per county in New York state. The following task abstraction would be to display the changes in the number of opioid poisoning deaths over time per county in New York State (Mid-level, Search, Look-up). For the second domain task, the goal is to see which counties in the New York state have the greater number of opioid overdose deaths. The following task abstraction would be to compare the number of opioid deaths per year and overall totals between 2013-2019 between counties (Low-level, Query, Summarize, Compare).