Visualizing the Opioid Epidemic in New York State from 2003 to 2022

Adelaide Bsharah, Marley Ferguson, Robbie Hoyler, Aashvi Shah

Northeastern University

ABSTRACT

We seek to support two domain tasks when examining opioid related deaths in New York. First, we will determine which years per county had the greatest number of opioid deaths. Over the years, trends in drug use change based on 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. Second, we will determine which counties in New York have the overall greatest number of opioid poisoning deaths. Policy makers and public health officials will be able 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, using a visual representing the number of deaths that involve certain characteristics such as age. Our first task is to display the changes in the number of opioid poisoning deaths over time per county in New York State (Mid-level, Search, Look-up). Second, we will compare the number of opioid deaths per year and overall totals between 2003-2022 between counties (Low-level, Query, Summarize, Compare). It is important to support both domain tasks because they provide insight on which counties in New York state were most affected by the opioid crisis, as well as trends/changes in death in both the state and 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 2003-2022 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. 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 2022) 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. We hope to visualize the data in both the "Opioid-Related Deaths by County" dataset and the "Opioid-Related Deaths" by Age Group dataset. We will have visualization to break down the number of opioid poisoning deaths in New York state and all counties, as well as a tool that visualizes certain demographic data.

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 hot-spots" using heat maps to highlight patterns and trends between geographic locations and types of crime [2]. 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 [2]. 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." We will also use the "Opioid-Related Deaths by Age Group" dataset from New York state to provide additional insights using various demographic data. The visualization will have multiple components, starting with an interactive map of New York State separated by county. Hovering over a particular county will highlight its area on the map and display its name. Simultaneously, a bar chart will display the total number of opioid related deaths (per year) that have occurred in that specific county from 2003 to 2022. This will allow public health officials to visualize trends in the frequency of opioid-related mortality in a particular county. Additionally, there will be a bar chart that presents the sum-total (all years added together) of opioid-related deaths for each New York county. The data of the particular county being viewed will also be highlighted and enlarged on the bar chart in conjunction with the other components of the visualization. This visualization tool will allow public health officials to 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

Both datasets are 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.

In regard to the Opioid-Related Deaths by County dataset, there seems to be no missing data values or outliers/unexpected values. 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.

In the Opioid-Related Deaths by Age Group dataset, 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. There will be some additional data cleaning required on our end. 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.

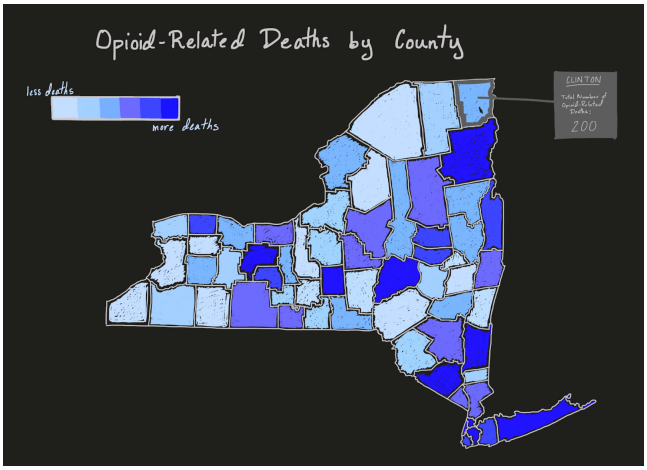

Figure 1

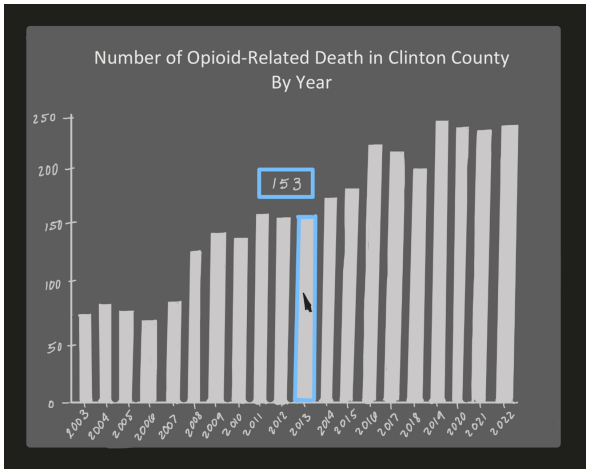

Figure 2

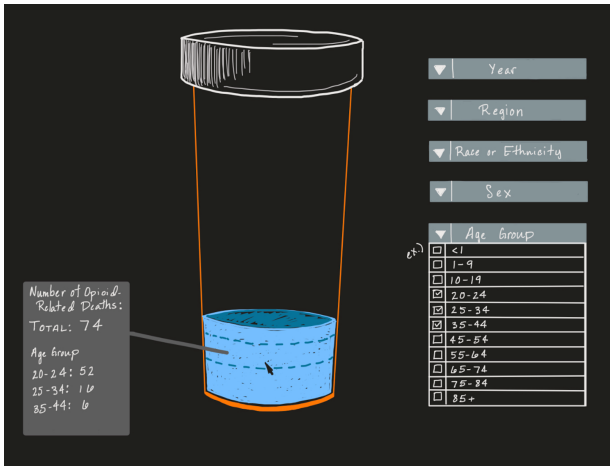

Figure 3

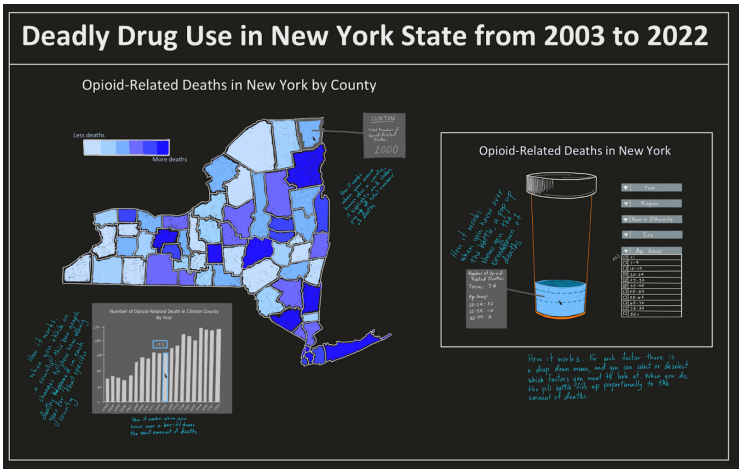

Figure 4

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.

6 Final Design

7 Discussion

8 Conclusion

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] 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

10 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

11 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).