Data Visualization of West Virginia's Drug Epidemic

Voke Brume, Amy Farson-Ciccone, and Moriah Frazier
Lane Department of Computer Science and Electrical Engineering
West Virginia University
Morgantown, WV

Introduction

In the United States, drugs have been a problem for decades and have gradually evolved into an epidemic. Many states currently have high opioid-related overdose deaths, but over the past several years, no more than West Virginia (WV). This has resulted in WV investing considerable resources into various approaches to tackle the drug epidemic. One such approach is massive statewide data collection of numerous factors influencing the rate of increase in opioid-related overdose deaths. The widespread adoption of technology and the ever-increasing ability to store large amounts of data have caused researchers to recognize the vital role of Big Data as an opportunity to access empirical evidence in the health sector. The current challenge is linking these multivariate, discrete datasets from different sources so that new insights into the state's drug epidemic root cause, growth, and impacts can be investigated. This report addresses this problem.

Related Work

Over the past two decades, data visualization has become an increasingly pertinent tool for gaining insights into healthcare data sets. The importance of sequential color schemes to represent geographically distributed datasets (choropleth maps) in data-driven knowledge discovery was recognized by Koua and Kraak, 2004. Gil-Garcia and Pardo, 2005, identified usability and ease of use as vital technical challenges in healthcare data visualization. Tominski, Schulze-Wollgast, and Schumann, 2008, impressed on the importance of temporal and spatial factors in multivariate data representation. This work was furthered by the Community Health Map proposal, enabling users to visualize healthcare data geospatially and in multivariate space (Awalin Sopan et al., 2012). Benjamin Bearnot MD et al., 2018, utilized publicly available, crowdsourced data to understand the opioid overdose epidemic by analyzing geospatial trends of discarded needles. More recently, C. K. Leung et al., 2020, presented a Big Data visualization and visual analytics tool for visualizing and analyzing COVID-19. Rui Li, 2021, visualized county-level data in the state of New York using a thematic map utilizing a choropleth design to display confirmed COVID-19 cases. This report builds on these works by utilizing techniques and visualization types to create a user-friendly dashboard that explores multivariate factors influencing the drug epidemic in West Virginia.

Data

For this project, the need to utilize an easy-to-use platform that works well with big datasets was crucial. After much research, Tableau was deemed the best platform to do the job. Tableau is a visual analytics platform that helps transform and translate data to solve problems, view and predict trends, and for compelling storytelling. Multiple data sets containing factors that impact the drug epidemic's root cause, growth, and impact were sourced and explored. Ruhm C.J 2019 cites economic factors as a significant indicator of drug use and mortality along with other factors such as the cost of drugs, gender, and demographic distribution. This report uses poverty and unemployment rates as economic indicators, visualizing the growth over time. Each factor is visualized individually for each county, after which their relationships with factors like life expectancy, drug arrests, total arrests, drug mortality, and marijuana and non-marijuana use are then explored using appropriate visualization techniques. The trends for drug arrests and mortality were then forecasted using Tableau's inbuilt exponential smoothing/weighted moving average model for the next three years.

Fig. 1. Structure of report's relational database

One of the primary goals of this report is to make the dashboard user-friendly for not just administrators but also the general public. Since most individuals have access to a mobile device, it is easy to access the dashboard and explore the visualizations anywhere and at any time because Tableau's dashboards have the option of a mobile device view.

Method

For this report, the need to utilize an easy-to-use platform that works well with big datasets was crucial. After much research, Tableau was deemed the best platform to do the job. Tableau is a visual analytics platform that helps transform and translate data to solve problems, view and predict trends, and for compelling storytelling. Multiple data sets containing factors that impact the drug epidemic's root cause, growth, and impact were sourced and explored. Ruhm C.J 2019 cites economic factors as a significant indicator of drug use and mortality along with other factors such as the cost of drugs, gender, and demographic distribution. This report uses poverty and unemployment rates as economic indicators, visualizing the growth over time. Each factor is visualized individually for each county, after which their relationships with factors like life expectancy, drug arrests, total arrests, drug mortality, and marijuana and non-marijuana use are then explored using appropriate visualization techniques. The trends for drug arrests and drug mortality were then forecasted for the next three years using Tableau's inbuilt exponential smoothing/weighted moving average model.

$$E_i = WY_i + (1-W)E_{i-1}$$

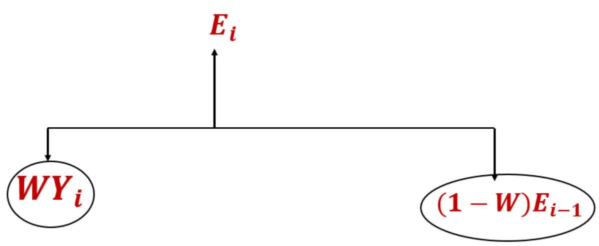

Fig. 2. Equation for calculating exponential moving average

This quantitative forecasting model forecasts data trends by multiplying the average values of a contiguous data set by a weight value that ranges from zero (0) to one (1). The bigger the weight value, the more the value is given to the time series current values compared to previous values.

$$\hat{Y}_{i+1} = W \cdot Y_i + W \cdot (1-W) \cdot Y_{i-1} + W \cdot (1-W)^2 \cdot Y_{i-2} + \ldots$$

W is...	Weight		
	Prior Period	2 Periods Ago	3 Periods Ago
	W	*W*(1-*W*)	*W*(1-*W*)^2
0.10	10%	9%	8.1%
0.90	90%	9%	0.9%

Fig. 3. Table showing how weight value assigns more importance to current values as compared to previous values

Finally, a four-page intuitive Tableau dashboard containing the visualizations is designed for users to explore these factors efficiently.

Achievements and Key Innovations

Completing this report required, first and foremost, a good understanding of Tableau and a relational database of all data sets used, meaning the first achievement was accomplishing these feats. The second achievement required using Tableau to generate multiple visualizations for the different data sets. The main accomplishment was creating the interactive dashboard with toggles, filters, and sliders so users can explore the data, draw correlations between data sets, and forecast future trends.

A key innovation was the "Explore relationships" tab of the dashboard, where the relationships between the different data sets were visualized. This enabled new insights to be derived and opens up the possibility for users to derive even more than documented in this report. Also a key innovation is the visualization to these data sets in the county level

Results

Visualizations were generated on an educational licensed Tableau Desktop software. The first step was to generate visualizations for individual worksheets, then a choropleth map of West Virginia with the color gradient representing the population of each county.

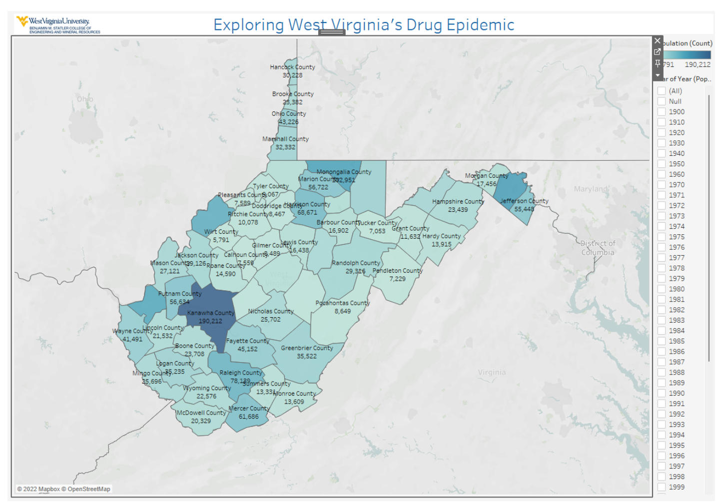

Fig. 4. Choropleth map with year filter representing a selected year's population on as a color gradient.

This map contains a filter toggle that switches between years or a sum of selected years. Hovering over individual counties shows the values of the data sets used for that particular year. Clicking on any county leads to multiple time series visualizations of the datasets for that county. Also, a forecasted trend of drug arrests and drug mortality data sets is visualized.

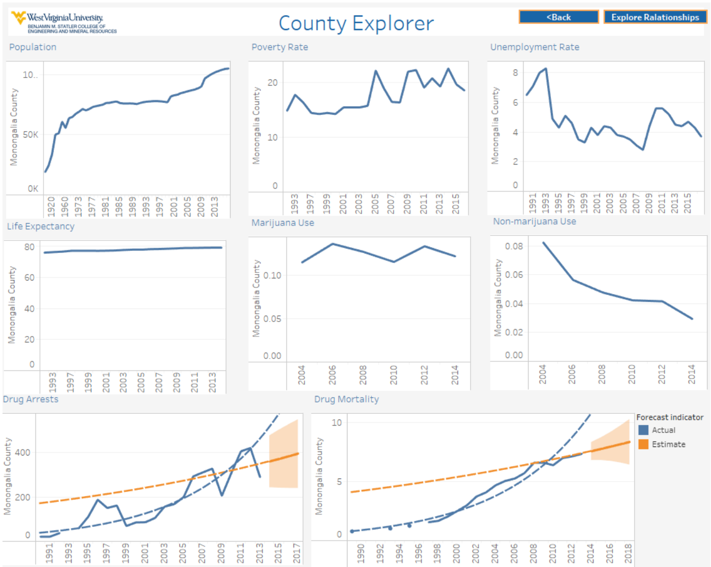

Fig. 5. County explorer on the dashboard showing time series data of all data sets and forecasting of drug arrests and mortality

The dashboard also contains multiple relationship pages. The first visualizes population growth and its relationship with poverty and unemployment rates. Users have the option of visualizing this in a video-like mode.

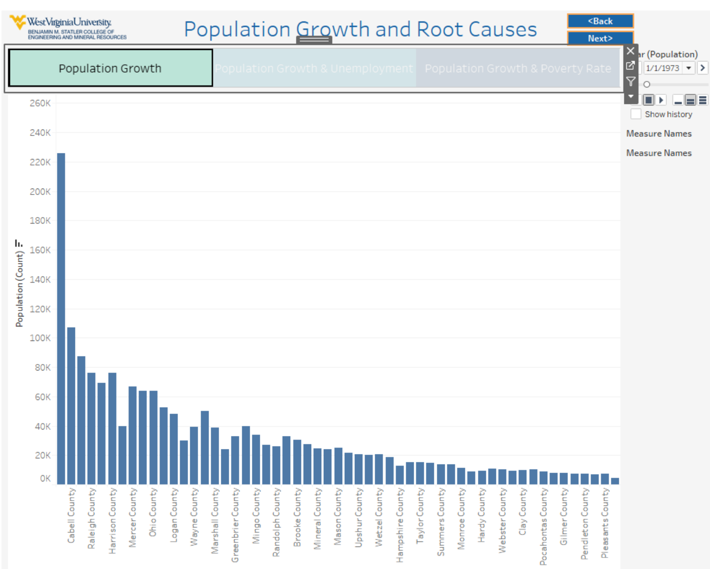

Fig. 6. Dashboard page showing population growth, and its relationship with poverty rate and unemployment rate

The last page of the dashboard contains multiple visualizations showing the interrelationships between the data sets. Relationships such as drug mortality and root causes, marijuana and non-marijuana use, drug arrests and total arrests, life expectancy, and drug mortality are visualized.

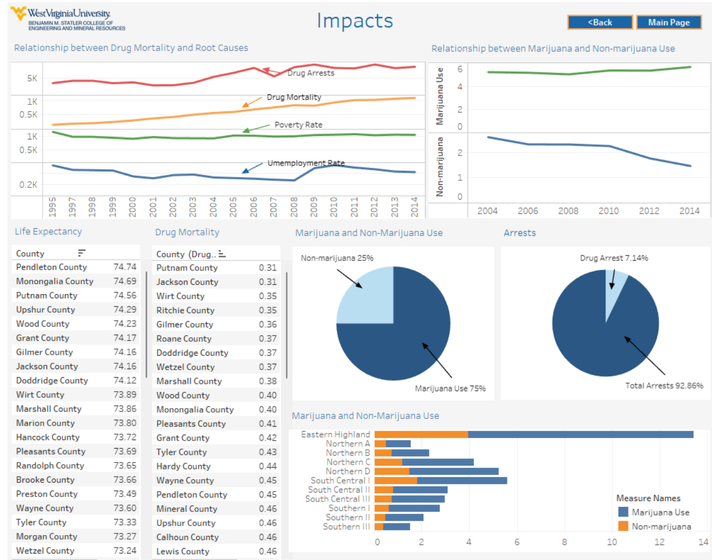

Fig. 7. Dashboard page showing other interelationships between factors

Key Findings

A significant finding in this report is the distribution of drug use, represented as marijuana and non-marijuana use. The result shows that counties in the Eastern Highland region of West Virginia, such as Tucker County, Barbour County, Berkeley County, and Randolph County, have a highly disproportionate drug use compared to other regions of the state. These are also small, rural counties with lower population density compared to counties such as Brooke County, Hancock County, Logan County, and Mingo county, in regions of the state characterized by lower drug use.

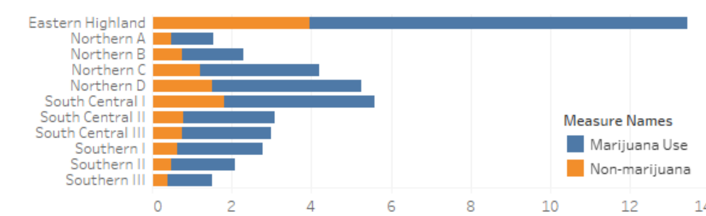

Fig. 8. Bar chart showing the distribution of drug use in different regions of West Virginia

While population growth was minimal across West Virginia, poverty and unemployment rates dropped. This indicates that, individuallly, population growth can not account for trends in the observed root causes of the drug epidemic in West Virginia.

Another key finding is that drug arrests account for about 7.14% of total arrests in West Virginia. According to the Department of Justice, this is considerably lower than the regional average of 17.1%. Also, although delegated as an illegal drug in West Virginia, marijuana accounts for about 75% of the total drugs used.

Fig. 9. Relationship between population growth and Unemployment rate in West Virginia

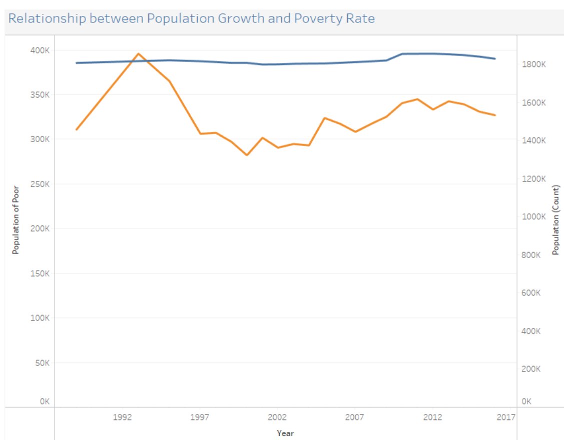

Fig. 10. Relationship between population growth and poverty rate in West Virginia

This report (Fig. 12) shows that while observed economic factors stayed steady, drug arrests and mortality increased, suggesting that the observed root causes can not individually account for increased drug arrests and mortality statewide.

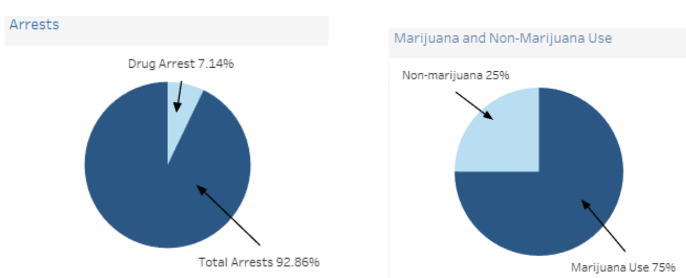

Fig. 11. Relationship between Drug Arrests and Total Arrests, and Marijuana and Non-marijuana Use

.

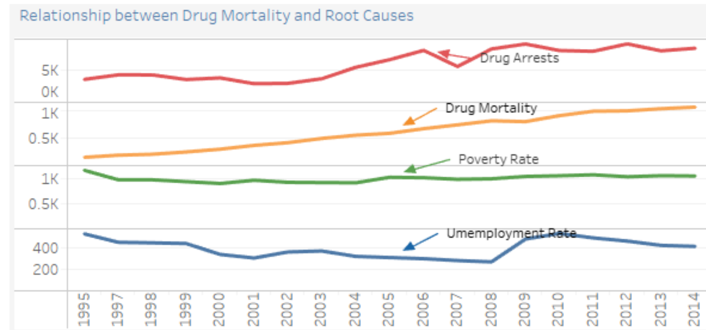

Fig. 12. Relationship between trends of Drug Arrests, Drug Mortality, Poverty Rate, and Unemployment Rate

Finally, Fig. 13 shows a non-linear relationship between the trends of marijuana and non-marijuana use. Recent trends show that while marijuana use increases slightly, non-marijuana use decreases. Comparing this to the trend of drug arrests and drug mortality, an inverse relationship can be observed. The National Institute of Drug Abuse (NIDA) suggests this is a result of an increase in the use of synthetic opioids, such as Fentanyl, that are more addictive and kill more users.

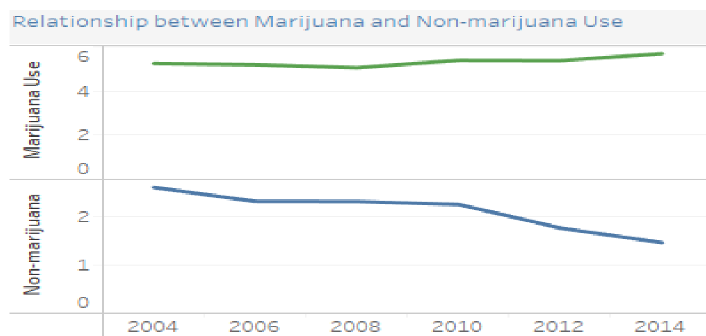

Fig. 13. Relationship between trends of Marijuana and Non-marijuana Use

Challenges and Limitations

The first challenge encountered with this project was finding the ideal platform that would best serve the needs of our project. Some compromises had to be taken into account since some platforms were better at specific types of visualizations than others, while most platforms do not work well with others.

Our greatest challenge was understanding how to work with Tableau to get our desired results. Fortunately, there is a plethora of instructional videos and web pages to help.

Another challenge we faced while working on the project was creating a central location for all datasets accessible to all group members. This made it a considerable challenge for individual group members to work simultaneously on the same datasets. Using google drive seemed like the best approach, but it was soon discovered that only the creator of the drive, not the shared participants, could access the files through Tableau. Moreover, having the files in the Tableau cloud required each group member to pay for the service.

A fundamental limitation in this report is the number of data sets used. Furthermore, there is the assumption that these economic factors and resulting factors were as key to the drug epidemic in West Virginia as they are in other states.

Conclusion

After completing our visualizations, we found some correlations between the datasets; some correlations were expected, and some were not.

It was expected that rural counties suffer more from drug abuse compared to more urbanized counties. However, the data showing a non-linear relationship between economic factors such as unemployment and poverty rate and drug mortalities statewide was entirely unexpected. Other findings include marijuana being the most used illicit drug in West Virginia, and a decrease in non-marijuana use (opioids) did not lead to a decrease in drug mortality. Finally, drug Arrests in the State of West Virginia account for a small fraction of total arrests.

Future Work

The learning curve of working on Tableau and time constraints did not allow for implementing some ideas. One idea was to expand the current dataset to include other root cause and impact data sets. Furthermore, some of the data sets stopped after a specific year. It would be helpful to obtain values for those missing years so the prediction of trends could become more accurate, and it could allow the comparison of the trend predicted now to see how accurate they were. Expanding the number of data sets could also yield new insights into current and future trends.

Another idea was to incorporate a "What If" section into the dashboard. The "What If" section would allow users to manipulate the data to see how specific changes affect the data trends. For example, what if the unemployment rate in West Virginia or a specific county goes up/down? How would that affect drug mortality or drug use in the state/county?

Data Sources

Population:
1900-1960:
https://www.census.gov/population/cencounts/wv190090.txt
1970-2009:
http://www.nber.org/data/census-intercensal-county-population.html
2010-2017:
https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2017/PEPANNRES/0400000US54.05000
Unemployment Rate:
https://data.bls.gov/cgi-bin/dsrv?la
Poverty Rate:
https://www.census.gov/data-tools/demo/saipe/saipe.html?s_appName=saipe&map_yearSelector=2016&map_geoSelector=aa_c&menu=grid_proxy&s_year=2016,2015,2014,2013,2012,2011,2010,2009,2008,2007,2006,2005,2004,2003,2002,2001,2000,1999,1998,1997,1996,1995,1993,1989&s_state=54
Life Expectancy:
http://ghdx.healthdata.org/record/united-states-life-expectancy-and-age-specific-mortality-risk-county-1980-2014
Marijuana Use:
2002-2004:
https://web.archive.org/web/20060816072445/https://oas.samhsa.gov/substate2k6/substate.pdf
2004-2006:
https://web.archive.org/web/20090324081746/https://oas.samhsa.gov/substate2k8/substate.pdf
2006-2008:
https://www.samhsa.gov/data/sites/default/files/Substate2k10-ChangeTables/ChangeTables/NSDUHsubstateChangeTabs2010.pdf
2008-2010:
https://www.samhsa.gov/data/sites/default/files/NSDUHsubstateChangeTabs2012/NSDUHsubstateChangeTabs2012.pdf
2010-2012:
https://www.samhsa.gov/data/sites/default/files/NSDUHsubstateTrendTabs/NSDUHsubstateTrendTabs2012.pdf
2012-2014:
https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHsubstateChangeTabs2014/NSDUHsubstateChangeTabs2014.pdf
2014-2016:
https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHsubstateChangeTabs2016/NSDUHsubstateChangeTabs2016.pdf
Non-marijuana Use:
2002-2004:
https://web.archive.org/web/20060816072445/https://oas.samhsa.gov/substate2k6/substate.pdf
2004-2006:
https://web.archive.org/web/20090324081746/https://oas.samhsa.gov/substate2k8/substate.pdf
2006-2008:
https://www.samhsa.gov/data/sites/default/files/Substate2k10-ChangeTables/ChangeTables/NSDUHsubstateChangeTabs2010.pdf

2008-2010:
https://www.samhsa.gov/data/sites/default/files/NSDUHsubstateChangeTabs2012/NSDUHsubstateChangeTabs2012.pdf
2010-2012:
https://www.samhsa.gov/data/sites/default/files/NSDUHsubstateTrendTabs/NSDUHsubstateTrendTabs2012.pdf
2012-2014:
https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHsubstateChangeTabs2014/NSDUHsubstateChangeTabs2014.pdf
Drug Arrests:
Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data
Annual series reports from https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/57?q=county
Total Arrests:
Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data
Annual series reports from https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/57?q=county
Drug Mortality
http://ghdx.healthdata.org/record/united-states-substance-use-disorders-and-intentional-injuries-mortality-rates-county-1980

References

[1] Awalin Sopan, Angela Song-Ie Noh, Sohit Karol, Paul Rosenfeld, Ginnah Lee, Ben Shneiderman, Community Health Map: A geospatial and multivariate data visualization tool for public health datasets, Government Information Quarterly, Volume 29, Issue 2, 2012, Pages 223-234, ISSN 0740-624X, https://doi.org/10.1016/j.giq.2011.10.002.

[2] Benjamin Bearnot, John F. Pearson, and Jorge A. Rodriguez: Using Publicly Available Data to Understand the Opioid Overdose Epidemic: Geospatial Distribution of Discarded Needles in Boston, Massachusetts
American Journal of Public Health 108, 1355_1357, 2018. https://doi.org/10.2105/AJPH.2018.304583

[3] “Big Data Visualization: What is it, Tips & Tools (beginner's guide),” *Visme Blog*, 27-Apr-2022. [Online]. Available:
https://visme.co/blog/big-data-visualization/#:~:text=Big%20Data%20Visualization-,What%20is%20Big%20Data%20Visualization%3F,chart%20types%20or%20visual%20representation. [Accessed: 24-Oct-2022].

[4] C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang, Y. Wen and A. Cuzzocrea, "Big Data Visualization and Visual Analytics of COVID-19 Data," 2020 24th International Conference Information Visualisation (IV), 2020, pp. 415-420, doi: 10.1109/IV51561.2020.00073.

[5] Gil-Garcia, J. R., & Pardo, T. A. e-Government success factors: Mapping practical tools to theoretical foundations. Government Information Quarterly, 22(2),
187–216. 2005.

[6] Koua, E. L., & Kraak, M. J. Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geography, 3(1),
12. 2004

[7] Li, R. Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps. Hum Behav & Emerg Tech. 2021; 3: 97 – 106. https://doi.org/10.1002/hbe2.248

[8] M. Bowers, “An overview of common data visualization mistakes” Toptal Design Blog, 20-Oct-2020. [Online]. Available:
https://www.toptal.com/designers/ux/data-visualization-mistakes. [Accessed: 24-Oct-2022].

[9] “Office of Drug Control Policy - Data Dashboard” *Data Dashboard*. [Online]. Available: https://dhhr.wv.gov/office-of-drug-control-policy/datadashboard/Pages/default.aspx. [Accessed: 25-Oct-2022].

[10] “Opioid data analysis and resources” *Centers for Disease Control and Prevention*, 01-Jun-2022. [Online]. Available:
https://www.cdc.gov/opioids/data/analysis-resources.html. [Accessed: 24-Oct-2022].

[11] Ruhm CJ. Drivers of the fatal drug epidemic. J Health Econ. 2019 Mar;64:25-42. doi: 10.1016/j.jhealeco.2019.01.001. Epub 2019 Jan 14. PMID: 30784811.

[12] Tominski, C., Schulze-Wollgast, P., & Schumann, H. Visual methods for analyzing human health data. Encyclopedia of Healthcare Information Systems, 1357–1364. 2008.