CS5010 Project

Camille Leonard
Darren Frye
Matt Dakolios
Paul Hicks
Sudhanshu Luthra

Overview/Motivation.

Global deaths due to substance abuse have evolved considerably over the last two decades. The recent evolution has been driven in large part by the opioid crisis ultimately resulting in significant adverse public health impacts (Vereinte Nationen & Büro für Drogenkontrolle und Verbrechensbekämpfung, 2019). In some countries the resulting picture is brightening as creative public health policies have been developed to take on this challenge. Closer to home, America is experiencing its own acute opioid epidemic that represents over half of the drug deaths as of 2020 with the numbers cresting 40,000 deaths nationally(*Explore Drug Deaths in the United States | 2019 Annual Report*, n.d.). The opioid epidemic is having dramatic economic impacts on the country as well, with recent national estimates putting the costs on the order of 10's of billions of USD(Abuse, 2020).

In addition to drug abuse deaths, alcohol abuse and its associated consequences are a leading health concern. The World Health Organization estimates that currently over 5% (~3M) of all annual deaths worldwide can be attributed to alcohol abuse(*Alcohol*, n.d.). State and federal governments continue to implement policies ranging from legalization, decriminalization, and the status quo to slow down and reduce the effects of the drug epidemic. Time will tell which decisions drive a reduction in the drug abuse death rate more effectively. That said, in order to develop the necessary understanding of the size, scope, overall trends, and geographic bounds of the problem, we developed a software-based tool. Our solutions will help public health practitioners have a solid basis for discussion in data, key visualizations and flexible response tools to query available health data and inform future decisions.

To achieve this goal, we developed a data-querying tool that public health policy research teams working at both the state and national levels could use. Our code allows the user to gain an understanding of the broad trends and scope of the drug epidemic, for a specific time period. We think that there has been a mixed response(*State Marijuana Laws in 2019 Map*, n.d.) with considerable timing differences across the nation including early adopters of legalization of certain drugs for medicinal use(*Colorado, Washington Become First States to Legalize Marijuana | TIME.Com*, n.d.). Our intent was to create a flexible tool that was both national in scope and able to transition over a given time period to aid very specific research queries one may have.

Hence, after a lengthy data exploration phase we converged on a US national data set on the death rate and cumulative deaths attributed to drug use for an available time period of 2000-2017.

Turning back to our audience, we envision researchers having potential statistical, mathematical or computational backgrounds to be our primary audience and interested in our project. Through tool development, we want to put the following key research questions at their fingertips with only moderate coding required.

Key Research Queries.

Query 1. Individual researchers will be able to filter the top five states for drug abuse death rates (or cumulative deaths) by discrete drug types in the 15-49 age range for a variable time period within the years 2000-2017.

Query 2. Individual researchers will be able to filter the bottom five states for drug abuse death rates (or cumulative deaths) in the 15-49 age range within the 2000-2017 time period.

Query 3. Users will be able to graphically display the relative drug abuse death rates (or cumulative deaths) across the country.

Query 4. State level researchers will be able to filter and portray the death rates (or cumulative deaths) for their state for discrete drug types in variable time periods within the 2000-2017 range.

Data Set Description.

Our main data source came from the Global Health Data Exchange (GHDx(*GBD Results Tool | GHDx*, n.d.)). The GHDx tool allows one to scope specific global health concerns for a wide variety of countries and continents for defined time periods. The data is maintained by the Institute of Health Metrics and Evaluation (IHME) from the University of Washington. This data is aggregated through collaboration with the Centers for Disease Control and Prevention (CDC), the National Center for Health Statistics and the US Department of Defense(*Global Burden of Disease Study 2017 (GBD 2017) Data Input Sources Tool | GHDx*, n.d.).

We focused our data queries on the years 2000-2017 and within the US. While we could have conducted a global search, we felt it would be better to start with domestic data before launching into broader comparisons. As such, we used the GHDx tool to define our interest areas for drug use death rates and deaths, categorized by gender, and discreetly broken down by the available specific drugs, which included: methamphetamines, cannabis, opioids, cocaine and other.

Coding Module.

Each query, listed in the Key Research Queries section, has a one to one match with a function in our module. function 1 (“topState”) and 2 (“bottomState”) are written with default values for every argument so a user can execute them without additional input. On the other hand, functions 3 and 4 have default values for all but one argument to avoid being prescriptive to a potential user. The data used within each function comes from a single csv that is sourced at the beginning of the file.

All functions begin the same structurally. Once the functions read the arguments in, they are first validated by an internal function [to the module] to make sure that they are acceptable values to proceed. Once validated, functions 1 and 2 use the arguments to filter the main data set, aggregate when needed, sort the data, and extract the appropriate number of rows in a tabular output form. The output of functions 1 and 2 is a data frame and optionally a csv (or the same information) saved to the current working directory, shown later in the report. An individual researcher has the option to change the time period for the query, depending on their research interest.

Functions 3 and 4 follow the same general structure as functions one and two, in which they first manipulate the data based on input. That said, these functions diverge from the first two functions we discussed when it comes to class of output. The latter two functions produce interactive visuals that are launched in the user default browser. Function 3 returns a heatmap for the drug use death rates across the country, and function 4 returns a line chart or a bar chart depending on user input. The users are allowed to either define a single year of interest or a range of years (within the 2000-2017 timeframe).

Testing Approach.

In order to verify that our program achieves the intended goals and meets our requirements, we conducted validation through unit testing. Our testing focused on ensuring the exception handling worked properly on edge cases and testing that our query 1 and query 2 methods produced data frames of the correct sizes. Due to limited time and experience we did not test the graphical outputs produced by the query 3 and query 4 methods.

Tests were written in an iterative fashion. First, based on our user statements and requirements. Then, expanded to include exception handling after our methods were complete.

<u>Query 1: topState</u>

We validated that our query 1 method would create a new data frame of the top n states for drug abuse death rates (or cumulative deaths). This was tested for when no parameters were passed to the method and when the correct number of states, gender, drug type, start year, and end year were passed to the method.

A total of nine tests were written to validate the first query. Three of the tests checked that the function was executing correctly and returning the appropriate object. Six of the tests were testing the exception handling by using edge cases. For example, a test was written with the anticipation of a user choosing a start year that is chronologically greater than an end year. Our testing confirmed that the correct error message (that prompts the user to correct one of the values so that the end year is greater than the start year) was returned. We also validated an exception hierarchy in which a year selected outside of our acceptable frame would be addressed before an inappropriate chronological order.

For the majority of our unit tests, the assertEqual method was used to either check a return value or returned error message string. For example, we wrote a test using the assertEqual method that checks the appropriate error message is returned when a user selects a number of states outside of the 1-50 range. In addition to using assertEqual, we also used assertIsInstance to verify that the object returned from the topState method was indeed an object as expected.

<u>Query 2: bottomState</u>

The testing was nearly identical but executed on the bottomState method. We wrote such similar unit tests for Query 2, because we are passing the same parameters through both methods. However, it is imperative to verify that the method functionality and requirements are working properly as well as the exception handling written for bottomState.

<u>Query 3 stateHeatMap</u>

The testing for the stateHeatMap method focused on the exception handling of the measure (rate or deaths), gender (male, female, both) and drugType (all, opioid, cocaine, amphetamine) parameters. As with the previous methods, we checked edge cases to ensure the correct error message was returned. Three tests were written for this method.

<u>Query 4: stateBarChart</u>

Similar to the query 3 testing, we tested the exception handling of the parameters passed to the stateBarChart method. Five tests were written for this method. We validated that when the user passed a state name anything other than one of our 51 states (including Washington DC) the script "Please pass the correct value for State" is returned. Another test written for this method ensures that if an incorrect gender value is entered that the error message "Please pass the correct value for Gender. Options are : Male, Female or Both" is returned. Similar to Query 1 and 2, unit testing for this method also ensures that if years outside of 2000 to 2017 are entered or if they are entered in the incorrect chronological order, the appropriate error message is relayed.

User Experience.

To enhance the understanding of this tool and illustrate what uses are reasonable for its use, we provide some key examples below. In the first instance, assume you are a national level policy maker that wants to gain a sense of the drug abuse death rate trends in the most recent year (2017). Our tool enables you to display the following gradient heat map.The darker the color the worse the drug death rate. Clearly one can see that states such as West Virginia, Kentucky, and Ohio have high drug death rates, while states such as Nebraska and the Dakotas are less affected by the drug epidemic.

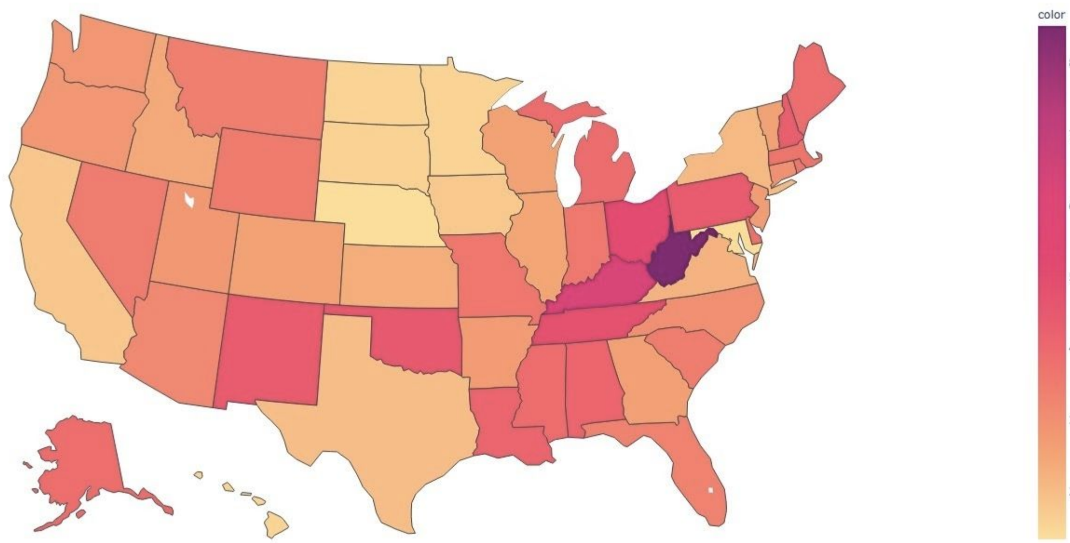

Fig. 1. 2017 National Level Drug Death Rates.

Turning to state-level research options, we developed a solution for follow-on queries to more closely examine defined time periods. We show in Figure 2 below a case where West Virginia's drug death rate from 2000-2017 is displayed. The graphic clearly demonstrates that the main driver behind the drug death rate epidemic over the set time period in West Virginia is opioids. The graphic also demonstrates there was steady drug death rate growth in the 2000's but a major uptick in about the year 2010. This uptick has persisted at least through 2017 suggesting a lack of public health measures in place to effectively blunt the growth of the epidemic in West Virginia until about 2016. We see many other state-level researchers potentially wanting to gain the same level of insight over certain time periods and the tool offers that opportunity.

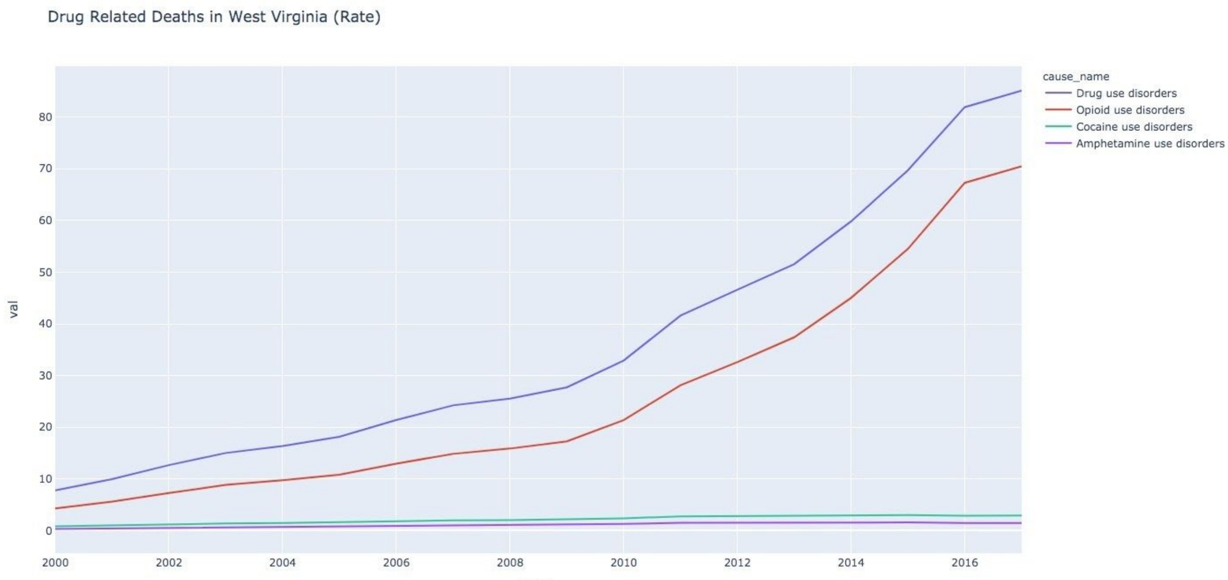

Figure 2. 2000-2017 Growth in Drug Death Rates.

Further, we chose to develop additional functionality that offers insight at multiple levels of policy development. To do so, we designed two queries that demonstrate groupings showing both **a)** (“topStates” query) highly affected areas and **b)** (“bottomStates” query) states with relatively less of a problem. In our view, this option would best serve research purposes if there was an ability to look at different time periods as well. As policies and legislation change over time, one would be able to see relative strengths and weaknesses of certain health plans and implementations against alternatives. Such comparisons would allow one to really dive into the data and see how long certain interventions take before one should expect to see a noticeable improvement. Similarly, one can envision a state level researcher being able to potentially transfer policies from areas that are having some success to ones that are not. These types of queries could drive outreach from governor’s to one another across the country. Of course, these types of results could drive national level funding as well as charitable foundation work to plan and put in place intervention centers in hard hit regions.

To take a look at the details of one is able to do with our tool, shown below are two tables that give examples to both of these potential queries over different time periods for the topState and bottomState functionality. In the query results shown in Figure 3, we can see that New Mexico, Pennsylvania, Nevada, Arizona and the District of Columbia in the years 2000-2005 were greatly plagued by drug deaths relative to other states.

Index	location_name	sex_name	age_name	cause_name	metric_name	val
31	New Mexico	Both	15–49 years	Drug use disorders	Rate	19.0137
38	Pennsylvania	Both	15–49 years	Drug use disorders	Rate	14.3379
28	Nevada	Both	15–49 years	Drug use disorders	Rate	14.0305
2	Arizona	Both	15–49 years	Drug use disorders	Rate	13.6539
8	District of Columbia	Both	15–49 years	Drug use disorders	Rate	13.5643

Figure 3. Top Five States for Drug Death Rates, 2000-2005.

In Fig. 4, below, we can see that milder impacts were observed in Nebraska, North Dakota, Maryland, South Dakota, and Minnesota suggesting potential positive interventions are in place in those states. Of course there are other contributing factors but the data provides a solid foundation for further queries, study and research.

Index	location_name	sex_name	age_name	cause_name	metric_name	val
27	Nebraska	Both	15–49 years	Drug use disorders	Rate	10.5662
34	North Dakota	Both	15–49 years	Drug use disorders	Rate	10.8081
20	Maryland	Both	15–49 years	Drug use disorders	Rate	11
41	South Dakota	Both	15–49 years	Drug use disorders	Rate	11.2617
23	Minnesota	Both	15–49 years	Drug use disorders	Rate	12.7593

Figure 4. Bottom Five States for Drug Death Rates, 2012-2017.

Extra Credit.

The main area where we developed advanced features beyond the original project requirements includes user experience and engagement. We have delivered a very flexible tool, driven by the acknowledgement that our target audience is potentially working at different levels of government support or within the government itself. We also know that research interests will widely vary over potentially different time periods as each state in the country was impacted differently and at different times. As such, we feel we have provided a very dynamic querying tool for advanced queries that allows for user input and interactive visuals. We have demonstrated many of these outputs above in the report.

Conclusions.

Our main goal was to gain insight into the drug death epidemic in our country and see what role the opioid crisis was playing in the overall picture at both the state- and national-level. The scope and trends have been demonstrably unfavorable for public health with major economic, social, and political costs accumulating. It is abundantly clear that in fact the opioid crisis is the main driver. One could see major research moving forward to further dive into the opioid crisis with a main goal of slowing the overall drug death rates across the country. Given the economic challenges with the drug crisis, it is clear that the benefits are there and that data is available to design effective outreach and interventions in the public health sphere.

We have definitely been able to gain insights into both the scale of the problem but also the state- and national-level picture. Researchers working across a number of areas in public health can access these insights through the tool we developed. The tool is also very flexible but admittedly it starts with having quality, structured and clean data. IHME and its sponsors have provided this opportunity and we acknowledge the benefits of what they have achieved.

Outlook.

We feel like we have just reached an early milestone and have improved the accessibility of the drug use death data available from IHME. Clearly more could be done. Some of the more interesting opportunities we see include researching the following:

- What insights can we glean from the “bottom five” states that may aid interventions in the “top five” states?
    - What did these “bottom five” states do well in terms of public health interventions?
    - How long did the interventions take before improvements were seen?
    - What were the associated costs of the interventions?

- What parts of the globe have the lowest per capita death rates over the last two decades?
    - What potential interventions did these states introduce and when?
    - Were there adverse impacts of any of these interventions?
    - What would the relative costs be if the US were to adopt these policies and by way of extrapolation, what would be the different challenges in domestic implementation?

- What are the underlying causes for the most impacted states?
    - Are legal prescription drug use rates correlated with drug use deaths?
    - What are the socioeconomic factors contributing to the drug death rates?
    - Of these factors, which areas should policies focus on to achieve the most dramatic reductions in drug abuse deaths?

Reference List

Abuse, N. I. on D. (2020, May 27). *Opioid Overdose Crisis*. National Institute on Drug Abuse. https://www.drugabuse.gov/drug-topics/opioids/opioid-overdose-crisis

Alcohol. (n.d.). Retrieved July 24, 2020, from https://www.who.int/news-room/fact-sheets/detail/alcohol

Colorado, Washington Become First States to Legalize Marijuana | TIME.com. (n.d.). Retrieved July 24, 2020, from https://healthland.time.com/2012/11/07/two-u-s-states-become-first-to-legalize-marijuana/

Explore Drug Deaths in the United States | 2019 Annual Report. (n.d.). America's Health Rankings. Retrieved July 24, 2020, from https://www.americashealthrankings.org/explore/annual/measure/Drugdeaths/state/ALL

GBD Results Tool | GHDx. (n.d.). Retrieved July 24, 2020, from http://ghdx.healthdata.org/gbd-results-tool

Global Burden of Disease Study 2017 (GBD 2017) Data Input Sources Tool | GHDx. (n.d.). Retrieved July 24, 2020, from http://ghdx.healthdata.org/gbd-2017/data-input-sources?locations=102&components=4&causes=561

State Marijuana Laws in 2019 Map. (n.d.). Retrieved July 24, 2020, from https://www.governing.com/gov-data/safety-justice/state-marijuana-laws-map-medical-recreational.html

Vereinte Nationen, & Büro für Drogenkontrolle und Verbrechensbekämpfung. (2019). *World* *drug report 2019*.