

Exploring Opioid Prescriptions

11.09.2019

Austin Williams - A02170170 Ben Ovard - A01635364

> austin.michaelwill@gmail.com benovard@gmail.com

https://github.com/austin-michael/data_vis_final_project

Background and Motivation

While there may be different motivations for pursuing this project amongst our team, for me, Austin, this hits home. Many years ago a cousin of mine was arrested for stealing opiods to fulfill his addiction. Opioids are very addictive, often users will become dependent on them. We are interested in visualizing how many people are prescribed opioids and the impact they have on the United States. We also would like to see if weather has an impact on opioid prescriptions/overdoses. It will be fascinating to see how opioid dependency has either increased or decreased by state and overdoses amongst those users. Additionally, this visualization will allow users to dive into our visualizations and extrapolate where their interests lie.

Project Objectives

Primary Questions we are Trying to Answer:

- 1. How many people are prescribed opioids in the United States?
 - a. How many people are prescribed opioids in my area? In areas around me?
- 2. How many people die from opioids in the United States? In my area and areas around me?
- 3. Is there a correlation between weather, opioid prescriptions, and opioid related deaths?
 - a. I.e. Do cold months lead to more people 'needing' opioid prescriptions and ultimately more drug related deaths?

What we want to learn:

Our primary goal with this project is to explore visualization interactivity and learn how we can implement this successfully.. To do this, we are going to take a minimalistic approach: When the user initially loads the page the primary focus will be a map of the United States. It is then up to the user to select a point on the map to delve deeper into the visualization--allowing the user to only see the data they care to see and not be overwhelmed by data clutter.

What we want to Accomplish:

We wish to provide the user with an easy to understand, easy to use experience. Many data visualizations can be abstract, difficult to comprehend, and have a steep learning curve. Visualizations with these attributes may be pretty to look at, but ultimately serve no purpose to the average user. Our mission is to bypass these attributes and make our visualization usable by the average inexperienced user.

Data

Our main dataset has been acquired from kaggle. "Kaggle is an online community of data scientists and machine learners, owned by Google. Kaggle allows users to find and publish data sets..." Kaggle is well known for their reputable datasets. The dataset we have acquired is named "Pain Pills in the USA," which was originally posted by *The Washington Post*. This dataset consists of 380 million rows containing all records of opioid prescriptions in the years from 2006-2012. In addition to this large dataset, we have acquired two others from tableau public—"Tableau Public is a free service that lets anyone publish interactive data visualizations to the web." "Underlying Cause of Death, 2006-2016 - x40-44, Y10-Y14, Other drug - Drop Homicide, Suicide" and

"miami_fl_daily_weather_12_30_2007_-_7_31_2013#csv

(miami_fl_daily_weather_12_30_2007_-_7_31_2013." These two datasets include data on the number of deaths per county due to drugs from (2006 - 2016) and min, average, and max daily temperatures from (2007 - 2013), respectively.

Pain Pills in the USA: https://www.kaggle.com/paultimothymooney/pain-pills-in-the-usa

<u>Underlying Cause of Death, 2006-2016 - x40-44, Y10-Y14, Other drug - Drop Homicide,</u> Suicide:

https://public.tableau.com/profile/belliveaujd#!/vizhome/StateofAddiction-DrugDeaths200 6-2016/DrugDeathTrends

miami fl daily weather 12 30 2007 - 7 31 2013#csv (miami fl daily weather 12 30 2007 - 7 31 2013:

https://public.tableau.com/profile/jonboeckenstedt#!/vizhome/Temperatures/Temperature Dashboard

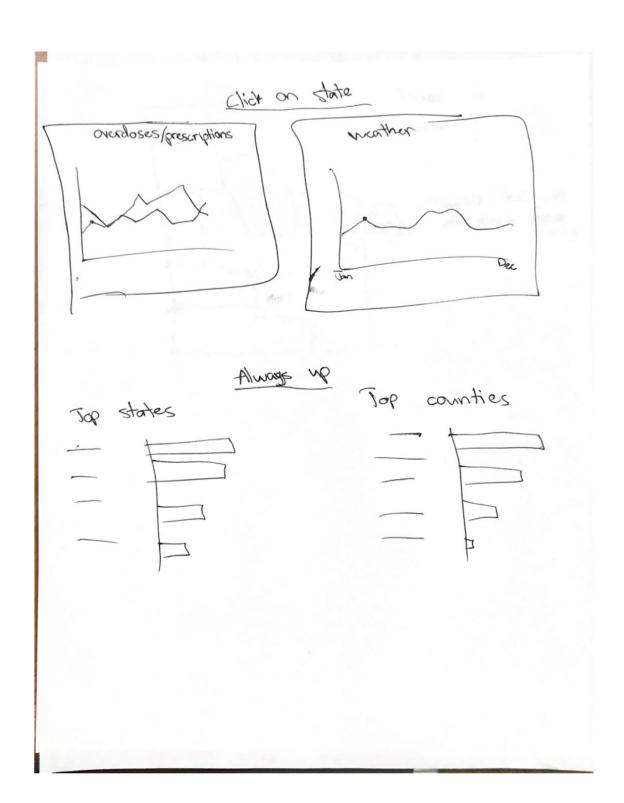
Data Processing

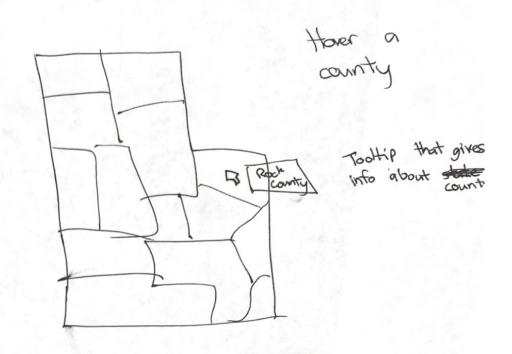
Most of the data sets we are using contain a lot more data than we need. The dataset of opioid purchases contains very detailed data about the purchaser, including their name, address, and other personal details. We simply want the number of purchases per county/city, so we will simply drop the columns that contain unnecessary data. Since each purchase it listed individually, we will also have to agglomerate the total number of purchases in each county into a new set, which will significantly reduce the size of our data. Our dataset of drug overdoses is pretty succinct. We will use almost all of the columns in that dataset, only dropping a couple. Our temperature dataset also contains a lot more data than we will need. It contains high and low temperatures of every day for over 6 years

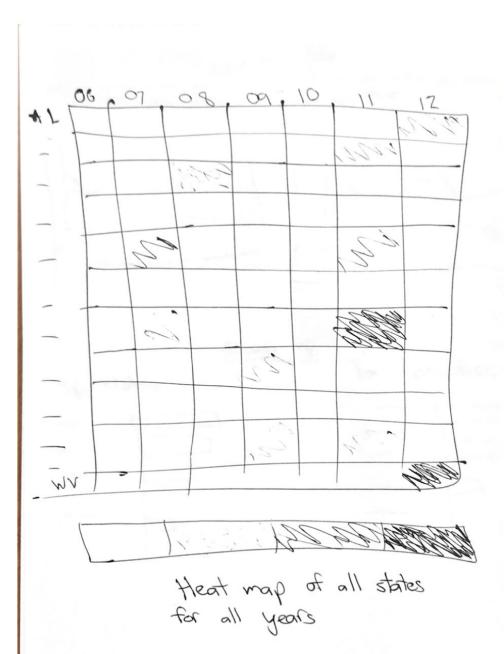
for a bunch of cities. We really only care about average high and low temperatures, so we will either get an average yearly high and low, or a total average high and low over all years. We will most likely do our data processing using the numpy and pandas packages, as they are suited for dealing with large amounts of data very quickly. They also make it easy to drop unneeded columns, and combine rows.

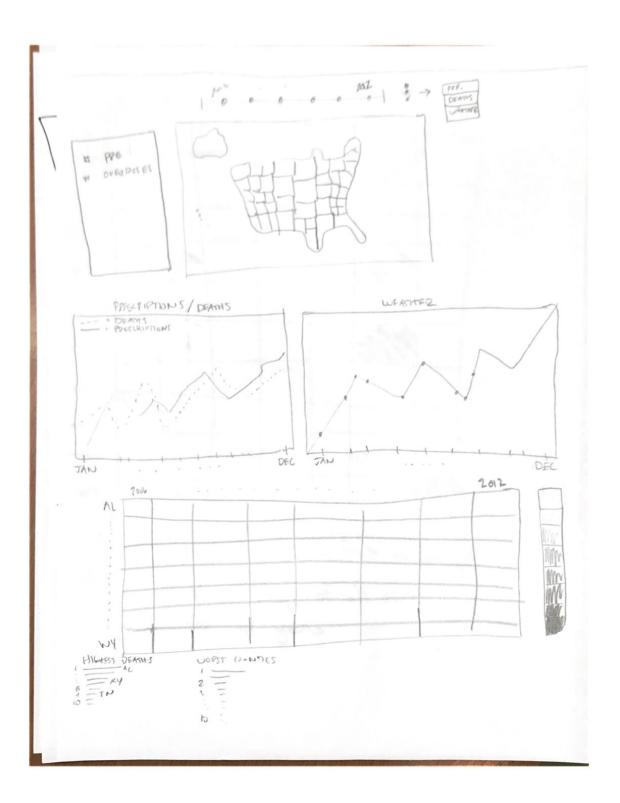
Visualization Design

The heart of our visualization is going to be a map of the United States. All interactivity will be derived here. We anticipate that we will divide the map by counties. The user will then click on a specific county and a dialog will display to the right of the map. The dialog will contain the total number of opiate prescriptions that were given in that year and the total number of deaths. We may also like to create small visualizations within this dialog. Under the map we would like to show two line charts. One line chart will consist of a comparison of prescription quantities and overdoses. The other will consist of weather data. The user will be able to click on a point in one chart and the corresponding point in the other chart will update as well. Underneath the two line charts we would like to include a heatmap showing the deaths by year for each state. Additionally, we would like to show an overall synopsis of the states and counties in two or more bar charts that represent the highest death rates by state, county, etc..









Must Have Features

Our project must have a map of some kind that shows opioid sales/use in the United States. This will be the core and center of the visualization. It also must be interactive. Interaction will be another core part of the visualization. We want users to be able to explore the map more fully by selecting and mousing over different parts.

Optional Features

One optional feature that we want to include are line charts that will appear when you click on a state/county. These will show more detailed about information over time for that region. We also want to combine various weather maps with the drug purchase maps in order to look for trends.

Project Schedule

Below is a tentative schedule for completing the project, dates may change as necessary.

Nov. 11th - Nov. 17th: Focus on data cleaning, processing, and building our data model for use with d3.js

Nov. 18th - Nov. 24th: Focus on getting the core features of our project underway (i.e. map functionality)

Nov. 25th - Dec. 1st: Focus on getting low priority visualizations implemented

<u>Dec. 1st - Due Date:</u> Test and cleanup

Works Cited:

"Academic Medicine's Response to the Opioid Crisis." AAMC, 2 Apr. 2018,

https://www.aamc.org/news-insights/opioids.

"Kaggle." *Wikipedia*, Wikimedia Foundation, 6 Nov. 2019, https://en.wikipedia.org/wiki/Kaggle.

Tableau Community Forums, https://community.tableau.com/docs/DOC-9135.