PROCESS BOOK

EXPLORING OPIOID USE IN AMERICA

OVERVIEW AND MOTIVATION

- Our goal was to create a visualization of opioid use in America that is pleasing to the eye, all information can be seen within one viewport, and there is reasonable amount of interactivity. We have found visualizations that require scrolling, panning, etc.. become cumbersome and key parts are hidden. To put a twist on things, we are introducing weather data to see if weather impacts opioid sales/deaths. Users will thus be able to extrapolate what they wish from our visualizations.

RELATED WORK

- A lot of our inspiration has come from Mike Bostock, he is widely known for some of his awesome map visualizations.
- Here is a link to go check out some of his work: https://github.com/mbostock?tab=repositories

QUESTIONS

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

We feel these are the most important questions to answer.

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
- Underlying Cause of Death, 2006-2016 - x40-44, Y10-Y14, Other drug - Drop Homicide, Suicide: https://public.tableau.com/profile/belliveaujd#!/vizhome/StateofAddiction-DrugDeaths2006-2016/DrugDeathTrends
- [TODO] Insert weather dataset name and link here

DATA PROCESSING

- Most of the data sets we used contained a lot more data than we needed. The dataset of opioid purchases contained very detailed data about the purchaser, including their name, address, and other personal details. We simply needed the number of purchases per county, so we dropped the columns that contained unnecessary data. Since each purchase is listed individually, we had to agglomerate the total number of purchases in each county into a new set, which significantly reduced the size of our data. Our dataset of drug overdoses is pretty succinct. We used almost all of the columns in that dataset, only dropping a couple. Our temperature dataset was pretty clean. It contained average high and low temperatures, so we got an average yearly temperature over the years 2006-2011.

DATA PROCESSING – IN DETAIL

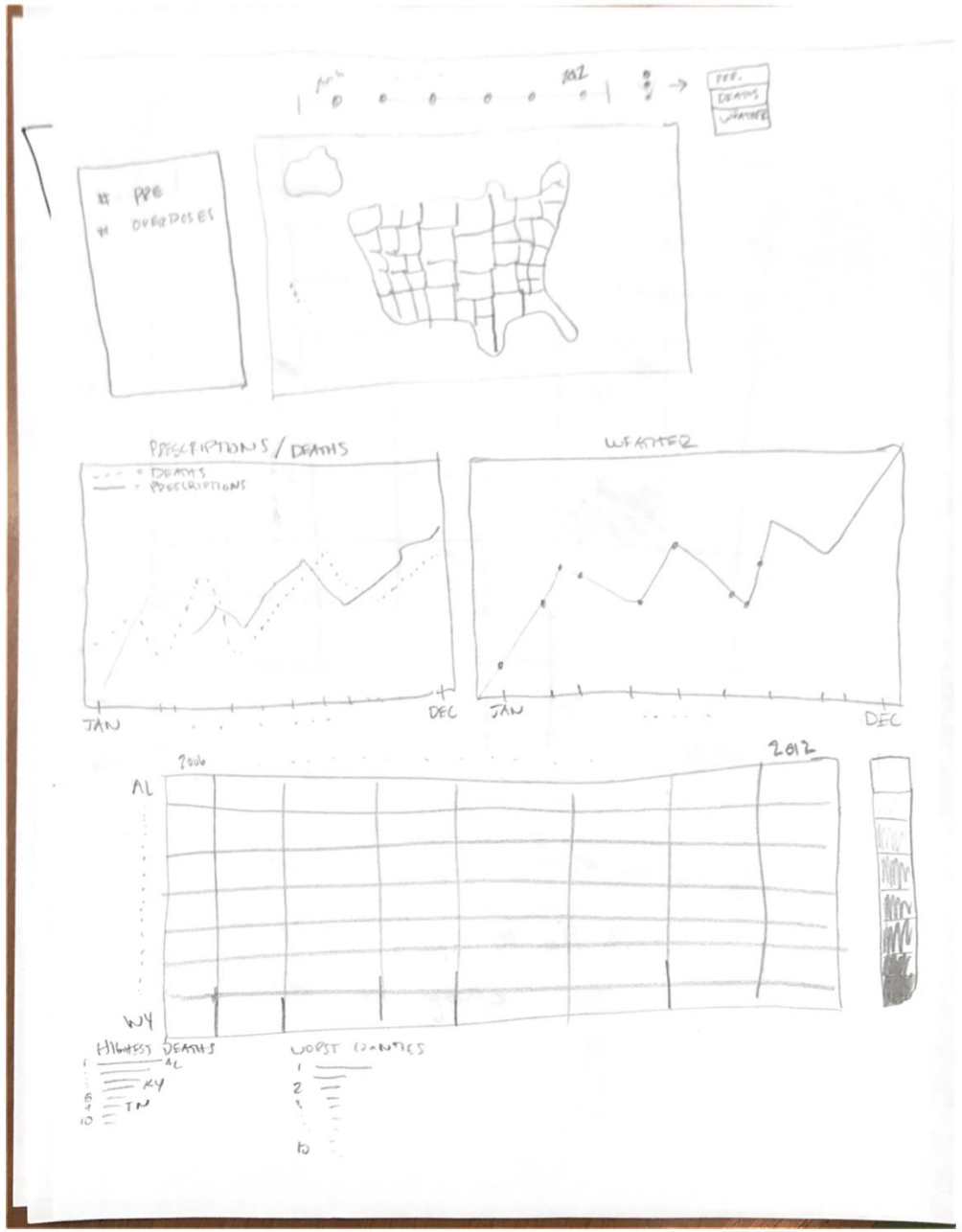
- [TODO]

EXPLORATORY DATA ANALYSIS

- We mainly used a map to visualize our data. One insight that we have gained, is that we did not have as much data as we thought we had. However, there was still sufficient data to proceed with the visualization and the user is still able select various years and get varying data.

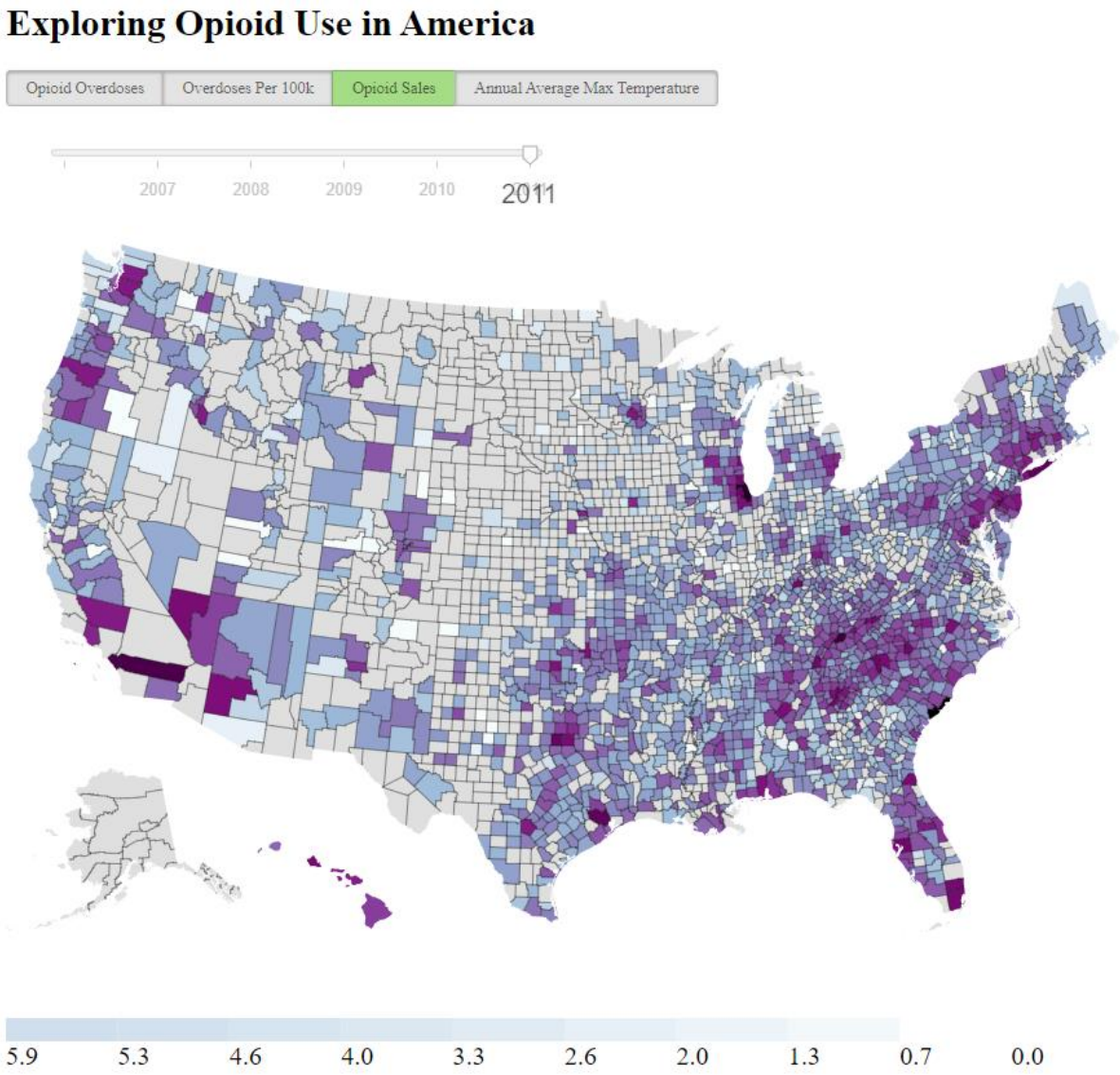
DESIGN EVOLUTION

- Our design remained mostly consistent. However, we messed around with the layout and considered using a linechart/areachart to compare total prescriptions, weather, and overdoses but time did not permit.

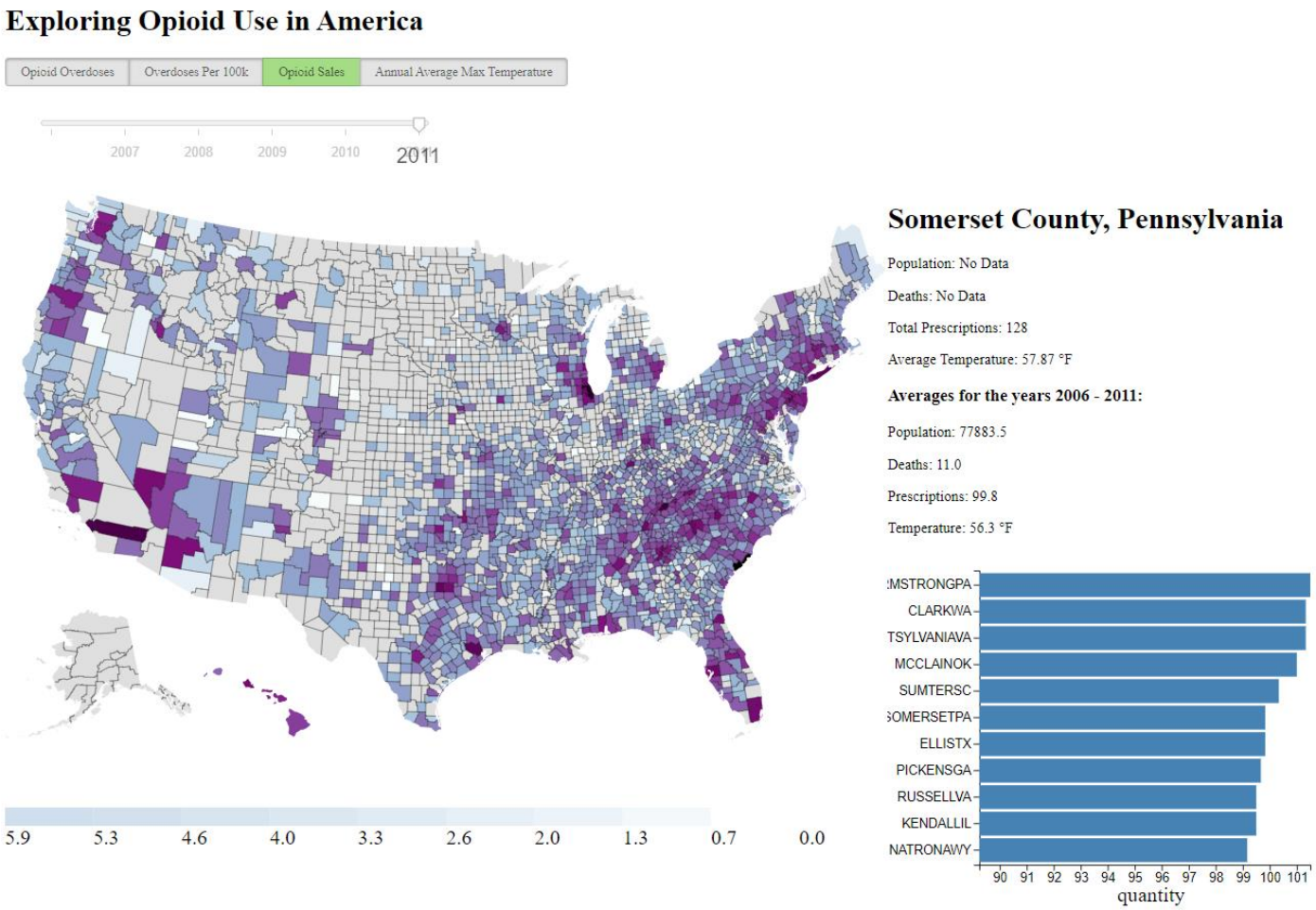


IMPLEMENTATION

- When a user first navigates to our site, they will be presented with a large map, year selections, and data selections:

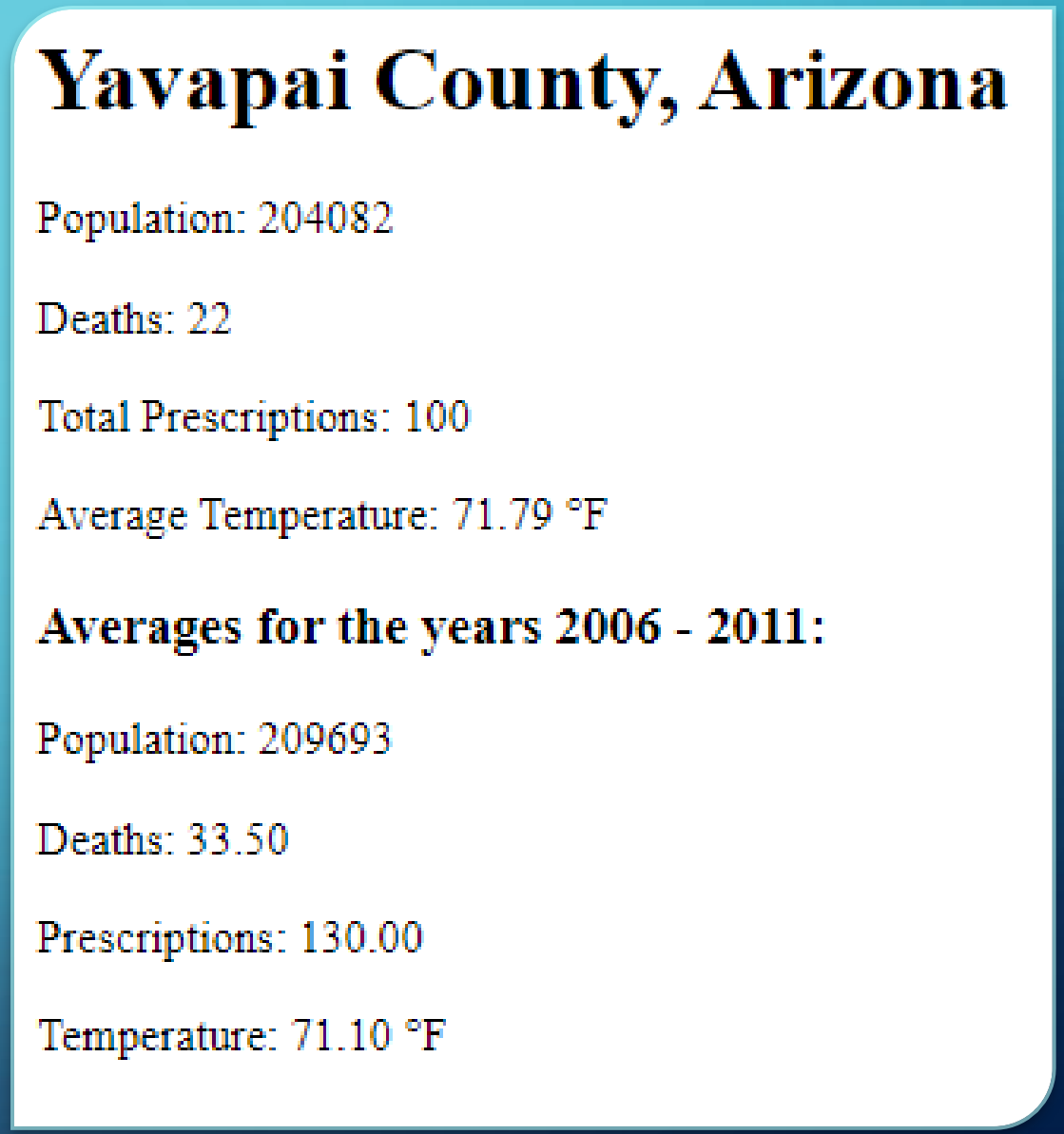


- The user will then select a dataset and year that they would like to examine. Upon hovering over a county a tooltip will display that provides the county name, state, and value of the current dataset they are viewing. If the user selects a county a menu to the right of the map will display, allowing the user to drilldown further into the data:

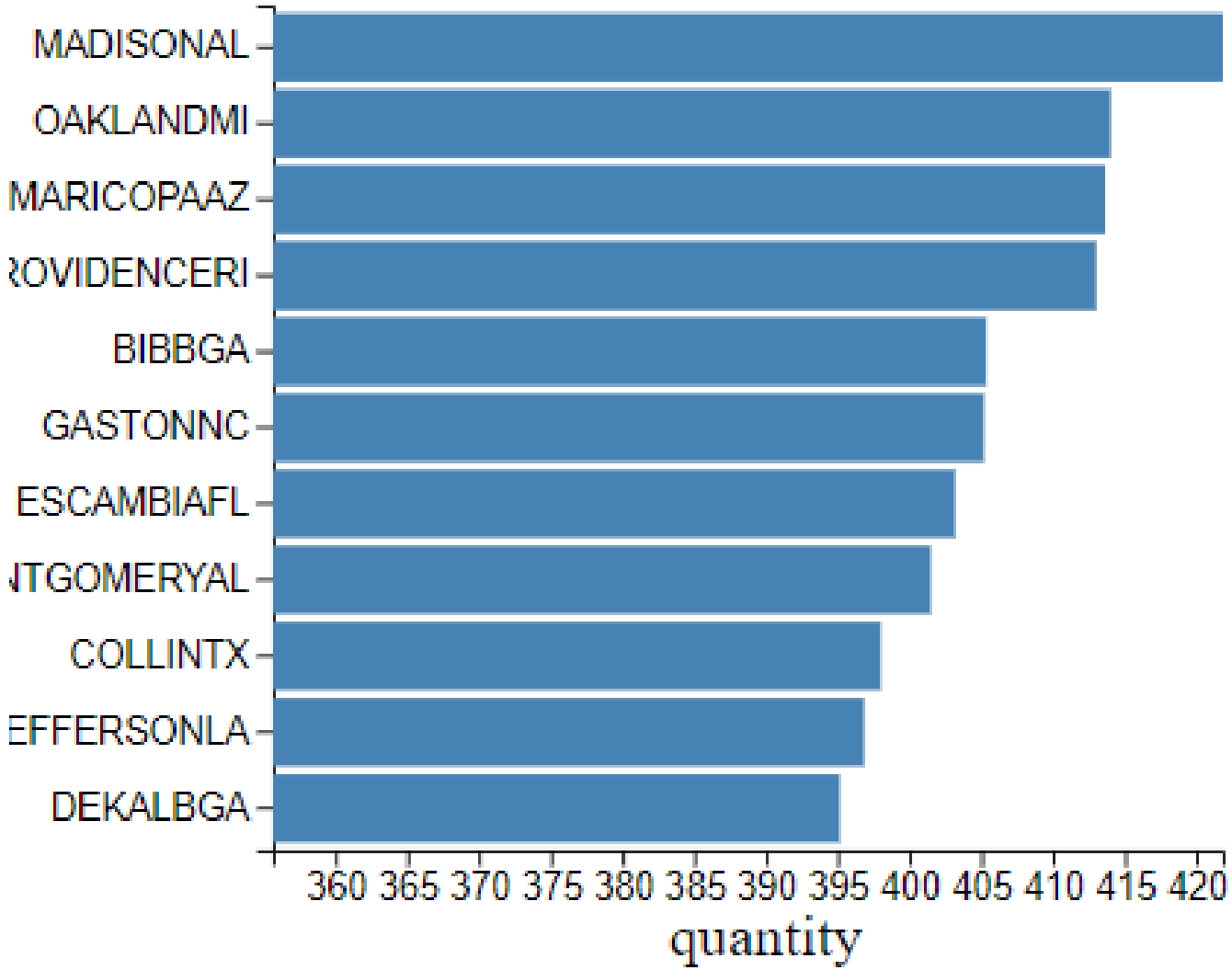


Here is a closer look of what the user can expect to see:

- Directly under the county name and state name, there is a list that contains the population, deaths, total prescriptions, and average temperature in degrees Fahrenheit for the currently selected year. Beneath this, the user can see an average of population, deaths, prescriptions, and temperature for the years 2006-2011.



- Additionally, the user will see a bar chart that appears next to the data drilldown. This bar chart shows the ranking of the currently selected county to the rank of similar counties within the U.S.. It displays 5 counties ranked higher than the selected county and 5 counties ranked lower.



EVALUATION

- We learned that there does not appear to be a correlation between Avg yearly temperature and opioid sales in America. There are quite a few things we could have done to improve our visualizations had time permitted. Namely, it would be more helpful to be able to see a direct comparison of the three datasets. Additionally, more interactivity could add to the over project.