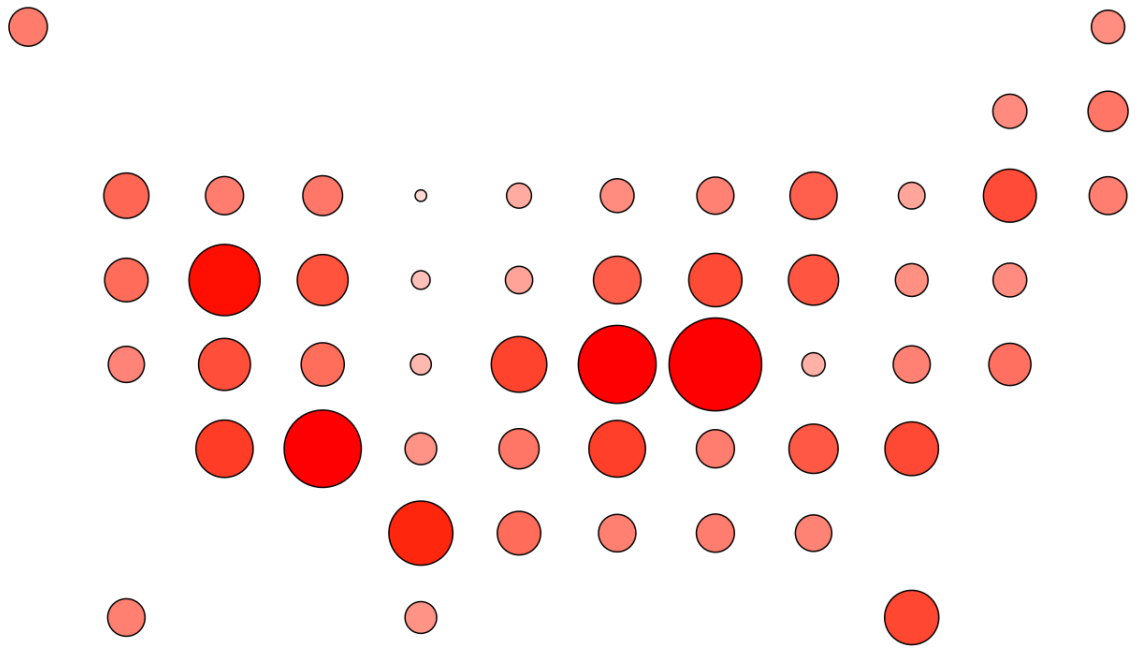
DRUG OVERDOSE DEATHS IN THE UNITED STATES

PROCESS BOOK



Jacob Crump
Jean Frédo Louis
Max Taggart

OVERVIEW AND GOALS

The visualization is designed to show how deaths due to drug overdose have changed over time in the United States. Our goal is provide an interface for users to explore national trends, including geographical patterns and distributions of drug overdose deaths across genders, races, and days of the week. While the simplicity of this visualization is perhaps best suited for the general public, more sophisticated views could easily be incorporated into this basic framework for use in public health and epidemiology.

RELATED WORK

The Utah department of health produces a report anually that details opioid overdose deaths in the state. The report is available as a static PDF at https://health.utah.gov-/vipp/pdf/RxDrugs/PDODeaths2015.pdf.

It includes figures like the one on the following page showing overdose deaths as a heatmap organized by county. This approach suffers from the problems that plague other geographical heatmaps, namely that large geographical regions often appear more severe due solely to their size. We consider geographical awareness of opioid deaths to be a key factor in understanding national trends, however we sought an approach that does not give undue importance to states based on the amount of land they posses or the sheer size of their population.

The New York Times also produced a story about the opioid epidemic in the united states, including figures like Figure 2, similar to the UT DOH report. This county-based heat map provides high-resolution information about the problem but also gives undue weight to larger counties.

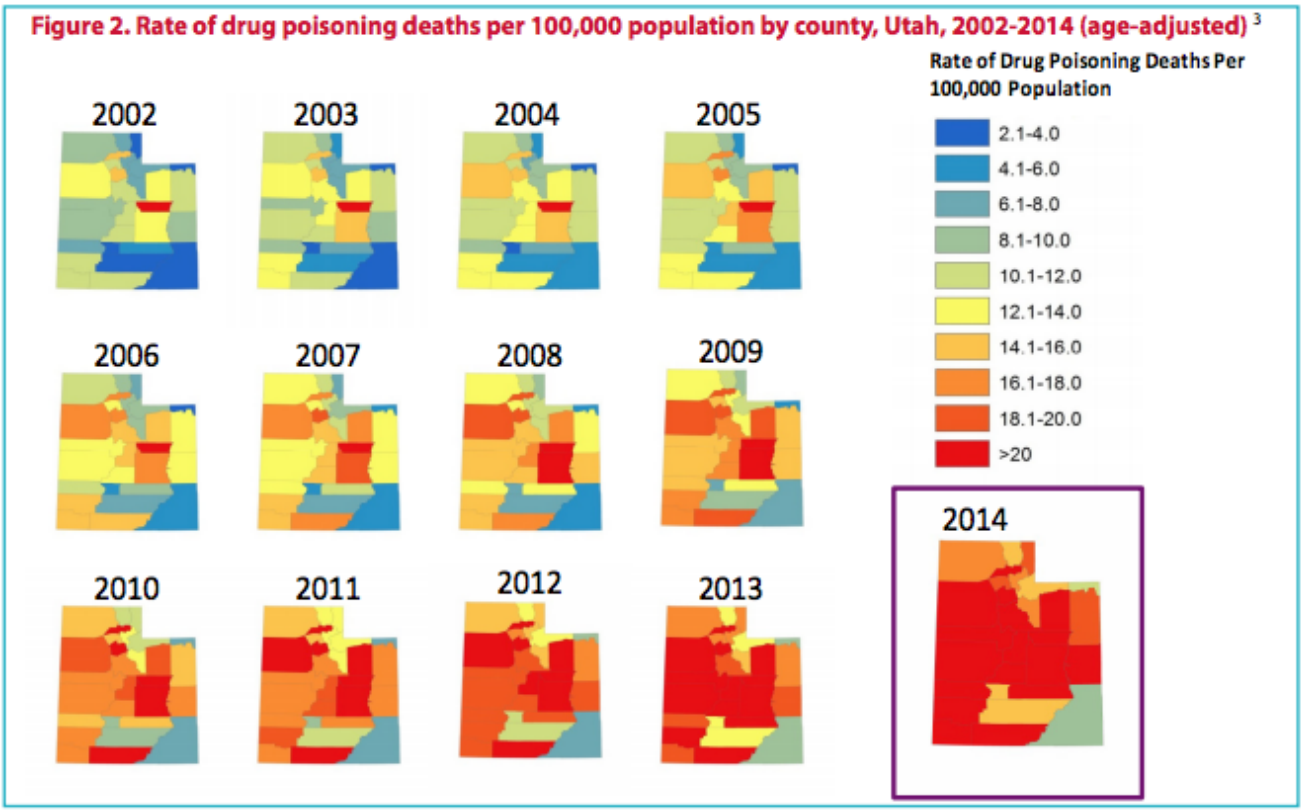


Figure 1. The Utah Department of Health annual report showing drug overdose deaths in utah.

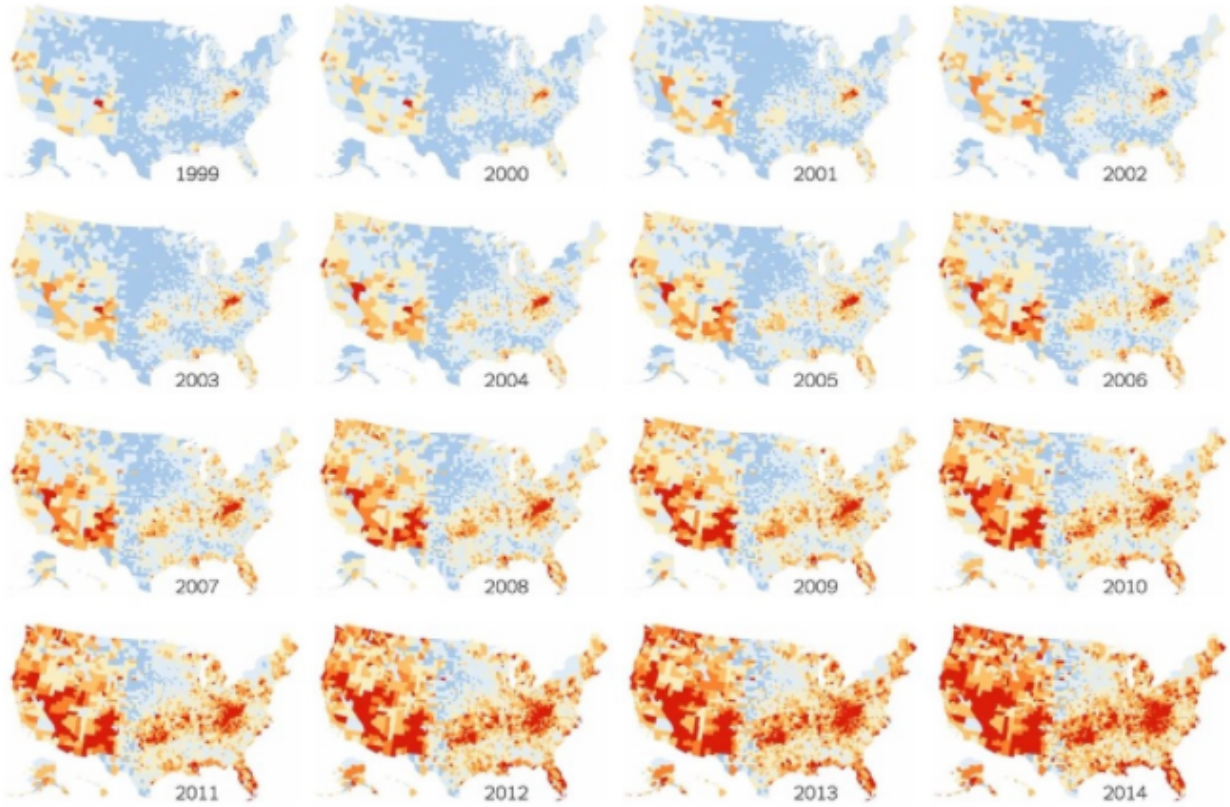


Figure 2. The New York Times heatmap of drug overdose deaths nationally.

INITIAL OBJECTIVES

Initially we were in contact with the Utah Department Of Health who had communicated that they might have a dataset available in time for our project. The dataset contained detailed information about prescriber practices and county-level opioid abuse. We were hoping to visualize patterns in prescribing practices and overdose deaths over time. However, the dataset from the Utah DOH was not available in time and we defaulted to a dataset from the Centers for Disease Control. Since this dataset contains national data our design questions shifted accordingly to an analysis of national trends including a comparison of deaths between genders, races, and geographical regions. We also wished to be able to visualize geographical trends over time to know if opioid use has spread from one part of the country to another or if the regions of high opioid abuse are geographically isolated.

DATA WRANGLING

We obtained our data from the CDC Wonder database, found at wonder.cdc.gov. The data are made available by the CDC in tabular format, however, for ease of use and to reduce application load times we condensed the data into a single JSON object using Python. We also wrote a JavaScript wrapper around the JSON data to provider a simple API for accessing the data from the different views of the visualization. The python scripts for data-wrangling and the JavaScript data wrapper are both available on the Github repository in DataPrepScripts/ and app/public/js/DataModel.js respectively. Python's pandas and matplotlib packages were used to perform initial data exploration.

DESIGN EVOLUTION

The first question we faced in the design of the visualization was what level of granularity we wished to display in terms of time and geography. The data from the CDC included county-level information that came in time increments as fine-grained as one month. Thus our initial sketches included a view that would show individual states so the user could view county-level information in a tooltip or detail view (Figure 3).

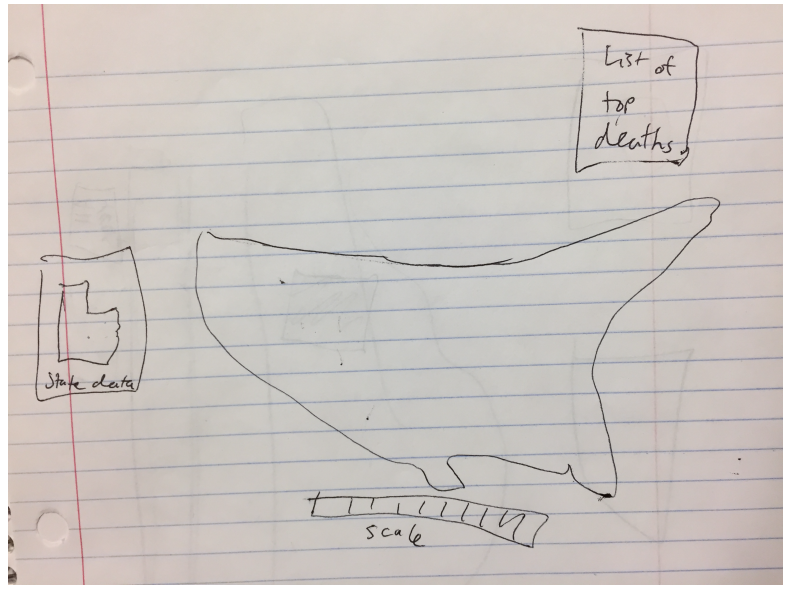


Figure 3. Initial sketches including a state-specific view.

It quickly became clear from examining the data however that month-level and county-level views were infeasible due to the fact that the CDC suppresses data values when the number of samples is low, which is the case with large portions of county and month-level data. From there we settled on showing state and year-level information. The next question was then how to allow the user to select their desired level of granularity and specify a specific state and/or year to examine.

Our next rudimentary sketches include a simple timeline above the map to allow users to select a specific year. Interacting with the map would allow the user to select a state. At this point we also included the possibility of detail views wherein the user could see a distribution of other variables, such as gender, day of the week, and race (Figure 4).

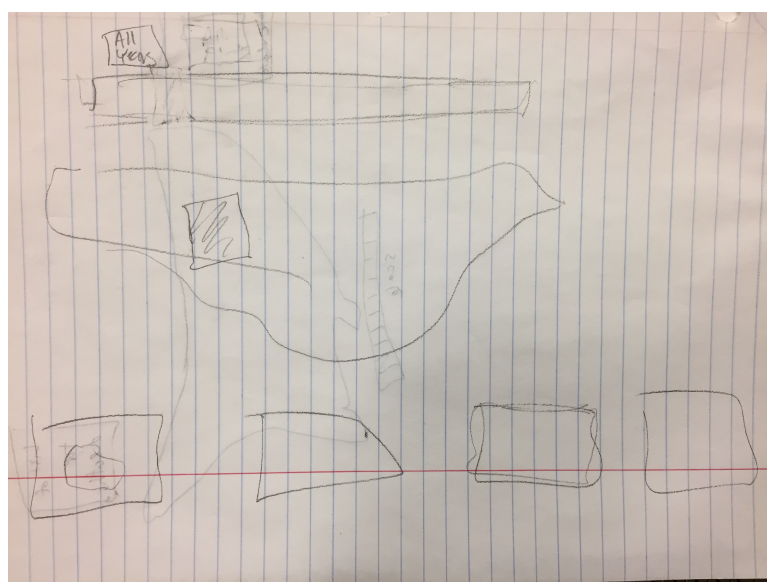


Figure 4. Sketches showing detail views and a simple timeline.

We also wished to help the user perceive changes over time in the all the views. So rather than producing a separate line graph or other representation of each view over time we decided to include a play bar similar to a movie, in which the user would be able to select a single year or to allow the visualization to advance upon activation of the “play” feature (Figure 5).

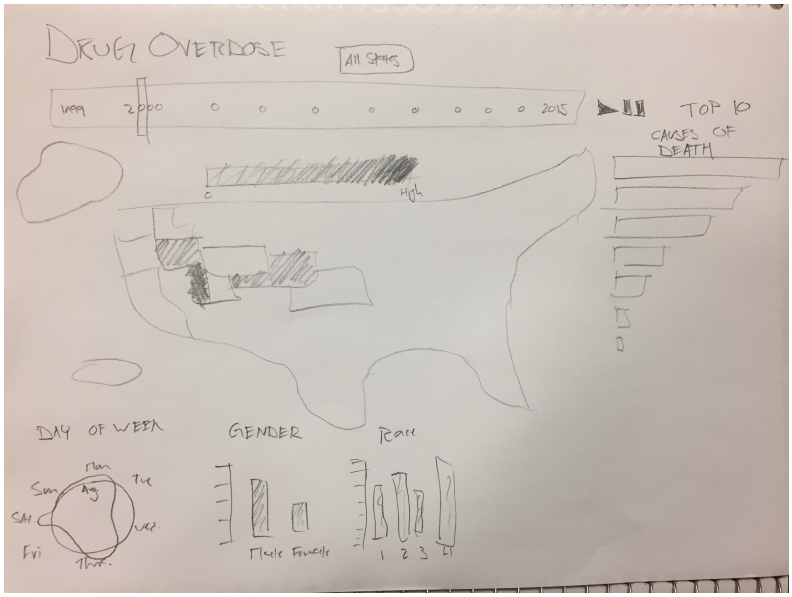


Figure 5. Inclusion of a play bar and specific detail views.

After talking to Carolina our conception of the map began to change and we began to consider abstract representations of U.S. geography. We explored possible representations of states with varying degrees of geographical fidelity (Figure 6).

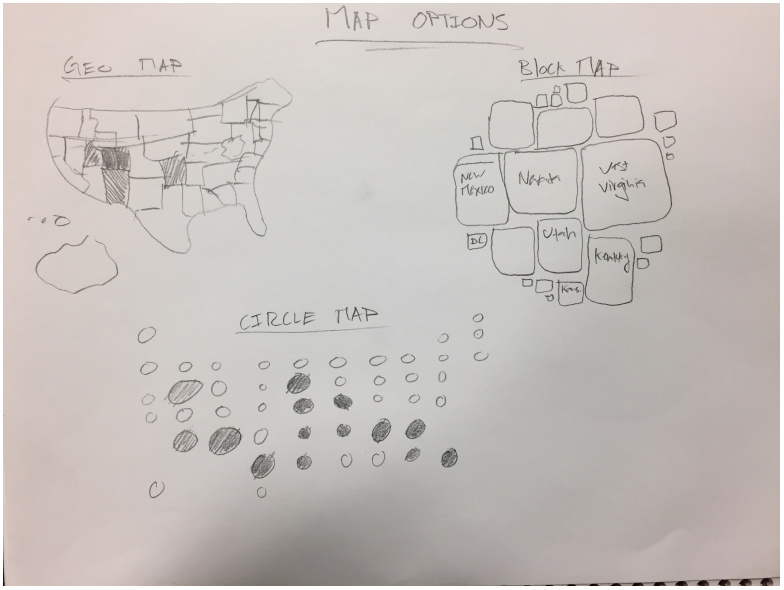


Figure 6. Exploration of different map representations.

We settled on the circle map representation for states since it preserved a general idea of national regions. The circle map representation also avoided the problem of giving undue weight to states with large areas or populations, given that rate of death was used instead of number of deaths. We also decided to include the detail views next to the map view to make all the views visible to users with smaller screens (Figure 7).

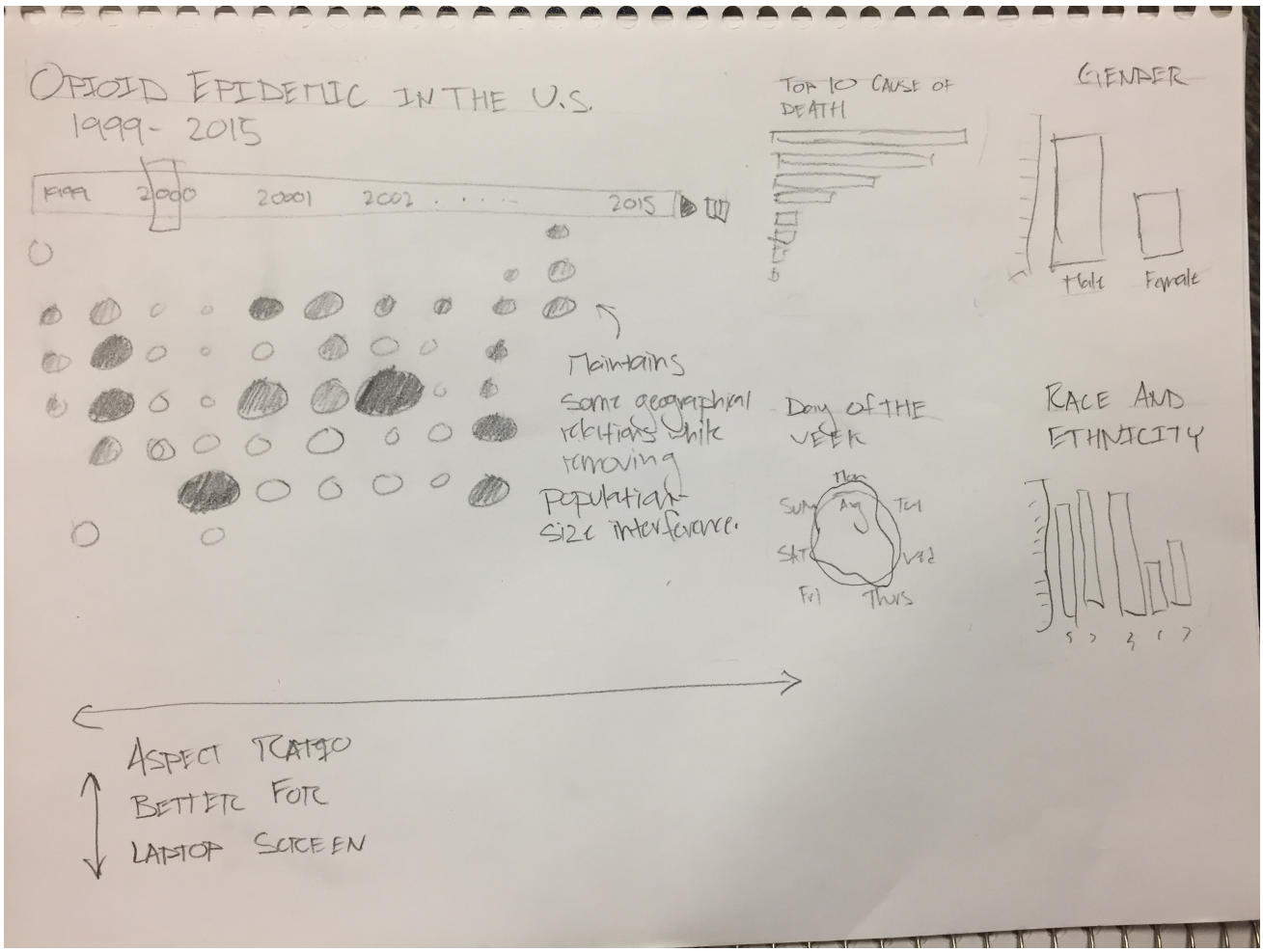


Figure 7. Final sketch incorporating the circle map and repositioning the detail views next to the map.

IMPLEMENTATION

One of the challenges of the implementation was how to deal with data whose value came from the CDC as “suppressed” or “unreliable” rather than as a numerical value. In the interest of time we chose to simply substitute 0 for these values. However, if we were to continue improving the design of the visualization it would be advantageous to make the user aware of the fact that data is not available for all variables, for all states, for all years. It would nice to have a visual queue that data is missing for a given view rather than to leave the user wondering why the view is blank or there is no bar on the bar chart.

A static view of the final implementation is shown in Figure 8.

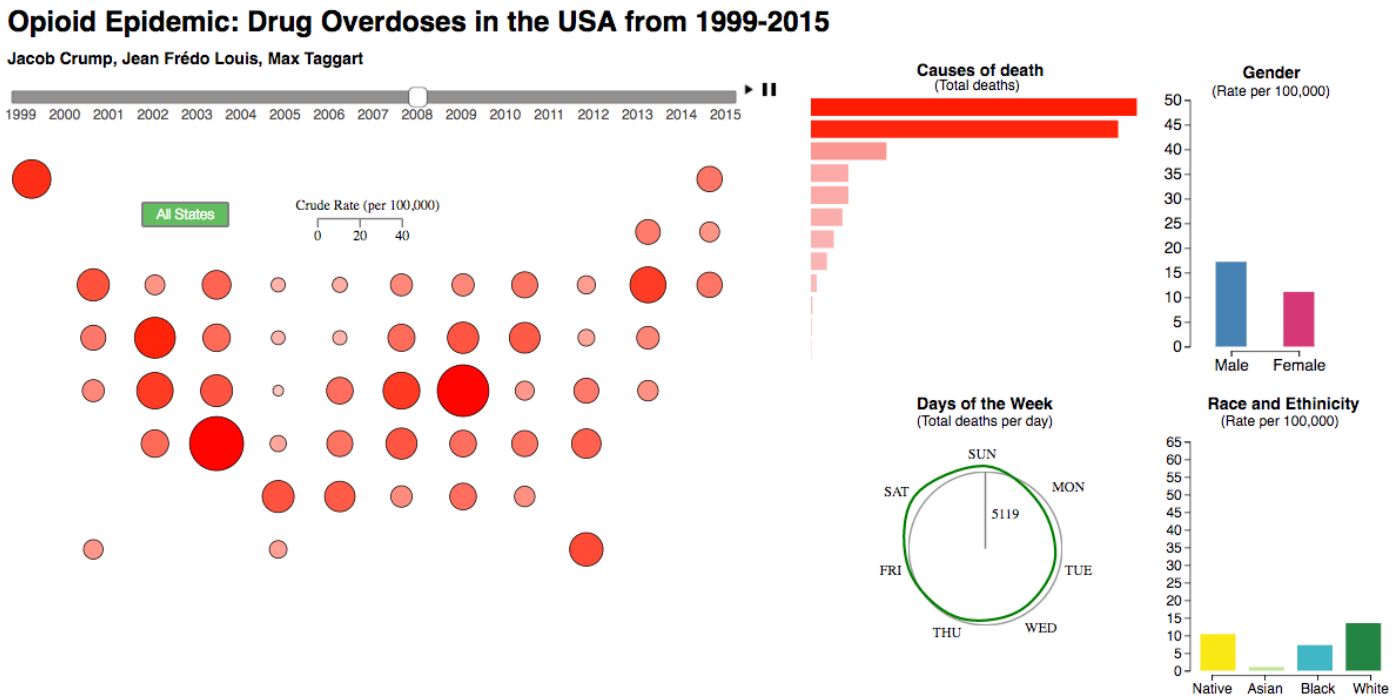


Figure 8. A screenshot of the final implementation showing the circle map and the detail views.

EVALUATION

Overall the visualization displays a clear representation of deaths due to opioid overdose by state and by year. It allows the user to examine multiple variables over time by stepping through years. It also allows the user to compare variables across geographic locations by selecting different states.

Admittedly one of the weaknesses of the visualization is that it does not incorporate any explicit storytelling component, leaving it up to the viewer to find salient points on their own. Another more subjective weak point of the visualization is the absence of a single view that displays trends. For example, a line chart of deaths over time. This is due to the fact that we chose to focus on displaying information for one year at a time.

Arguably it may be more informative to display states as a sorted bar chart and to keep track of states' changes from year to year than to display them as circles on a map. These additional views were not included in the visualization due to time constraints.