

Motivation and Audience

We intended for our visualization to inform the user of how different demographics and regions within the US might be inclined to die from different causes. We hypothesized that these differences might be would be negligible for universal killers such as cancer, but would be much more pronounced for less common causes of death, such as digestive diseases.

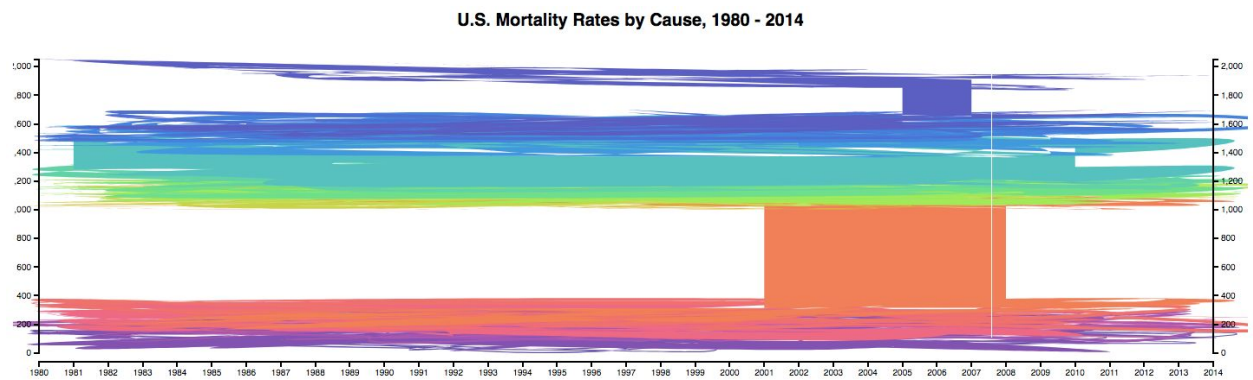
Visualization Design Evolution

Brendan had two different visualization ideas that were eventually combined into one. The first idea was to create a parallel coordinates graph which plotted how each county voted in each presidential election for the Republican and Democratic candidates. The user would be able to select counties from a map to alter the dataset. The second idea was a streamgraph showing how causes of death have changed over time. Andrew preferred the second idea, but we concluded that the streamgraph alone would be too small of a project. So we combined the two by allowing the user to choose the dataset in the same way they would be able to in the first idea.

To begin making the streamgraph, Andrew modified Will Turman's d3 [streamgraph example](#), which Prof. Cutler provided to us for Homework 6. First we tried modifying it to build a streamgraph with death_data.csv (79.8 MB) as the dataset, stitched together from each of the data source files. But the graph took far too long to render to debug comfortably.

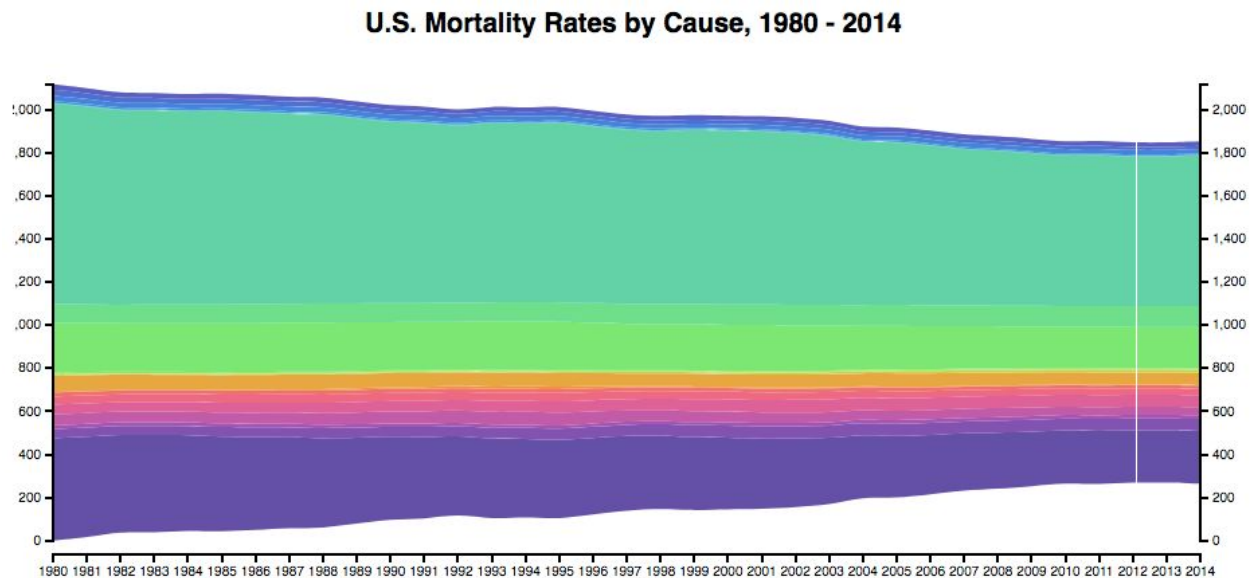
To simplify debugging, we decided to display the mortality rates of the entire US—not just individual counties—as a proof of concept. So Andrew wrote another Python script to average together the mortality rates between every state for each cause of death. The result was death_data_annual.csv (data-initial.png), 48 KB, which allowed for much more rapid

iterations in visualization code. However, the initial graph (streamgraph-initial.png) was a tangled mess of SVG <path>s that certainly did not reflect the real data.



The initial graph.

To make checking the data easier, we reformatted `death_data_annual.csv` to have its rows ordered by year then alphabetical cause of death. This had the unexpected effect of correcting the streamgraph's appearance.



The corrected graph.

Features/Challenges

One of our biggest initial challenges was trimming the mortality data to a more manageable size. The dataset for our streamgraph from healthdata.org is over 1 GB. It contains mortality rates, separated by cause of death, for every US county from 1980-2014. On healthdata.org, the yearly (non-quinquennial) data is contained within a couple dozen separate CSV files, separated alphabetically by state name. Andrew wrote a Python script to condense all of these CSVs into a single CSV. It reduced our data to 222.4 MB from 1.09 GB, removing information such as mortality rates split by sex, and upper and lower bounds on mortality for a 95% uncertainty interval. However, this was too big for GitHub, which (without the Large File Storage extension) only stores files < 100 MB. So I reduced the data even further to 79.8 MB, by removing repeated information between mortality statistics for the same state/county and cause of death.

Even after bringing the filesize down to 79.8 MB, GitHub wouldn't let us push our changes, because the 222.4 MB version of the data still existed within previous commits. [This article](#) within GitHub's help pages contains a helpful "git filter-branch" command with appropriate flags to remove a specified file from the entire history of every branch and tag. Our compressed mortality dataset, at 66.1 MB in size, is able to load in 674 ms.

Andrew thought that restricting the streamgraph's colors to a single hue, as in the six predefined colors from the initial code, would make it too difficult to distinguish each of the 21 causes of death. So we assigned each cause of death to a different uniformly spaced color from d3's `interpolateRainbow(t)` function, which produces a color scale whose varying lightness is a "less-angry rainbow". [This blog post](#) explaining the perceptual rainbow, though short, may serve as interesting reading material for a future Interactive Visualization semester.

Feedback

Unfortunately, most of the feedback we received from Interactive Visualization classmates was irrelevant, because it suggested we implement features we already included. However, there were a few suggestions that we have since implemented. One classmate suggested that we give the user the ability to apply more than one demographic to our selection of the map data. For example, a user might be interested in counties that are predominantly white, and are also urban. This suggestion led to us reconsidering how we allow the user to manipulate the data. We have since updated the design so that a user may ask for a range for a variety of demographics, including race, education, urbanization, age, and income. The user may also select which mortality causes they want to be included in the streamgraph. Another classmate recommended that, instead of continuously updating the streamgraph with each changed selection, we use an “apply” button that the user would press after selecting the counties and causes of death they are interested in. After implementing this change, our visualization ran much more smoothly.

One feature that we would like to add in a future iteration of this visualization is smooth transitions between streamgraphs with different parameters (selected counties / causes of death), instead of erasing and redrawing the streamgraph every time parameters are changed, in the style of [this example](#). Andrew attempted to integrate these transitions into the visualization about $\frac{2}{3}$ of the way through the project, but it requires a deeper understanding of d3 and JavaScript callbacks to work correctly.

Division of Responsibilities

The project conveniently divided itself into two separate subprojects: the map and the streamgraph. Brendan collected census data on the demographics of different counties and

used it to create the selection map. Andrew focused his efforts on the streamgraph, as well as finding and manipulating the mortality data.

Background

Koeze, Ella. "35 Years Of American Death." FiveThirtyEight, 20 Apr. 2017, projects.fivethirtyeight.com/mortality-rates-united-states/.

United States, Congress, "American FactFinder." American FactFinder. factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. US Census - American Fact Finder

Bostock, Mike. "Paint by Numbers." Bl.ocks.org, 27 Nov. 2017, bl.ocks.org/mbostock/9907392.

"United States Mortality Rates by County 1980-2014." *United States Mortality Rates by County 1980-2014* | GHDx, Global Health Data Exchange, 28 Sept. 2017, ghdx.healthdata.org/record/united-states-mortality-rates-county-1980-2014.