

Process Book

CS 6630 | 13 November 2015

Holly Cordner | Craig Roddin

Background Information

Project title: A Look at Healthcare in the US: What Factors Actually Make a Difference?

Who we are:



Holly Cordner

holly.cordner@utah.edu

u0876675



Craig Roddin

u0380903@utah.edu

u0380903

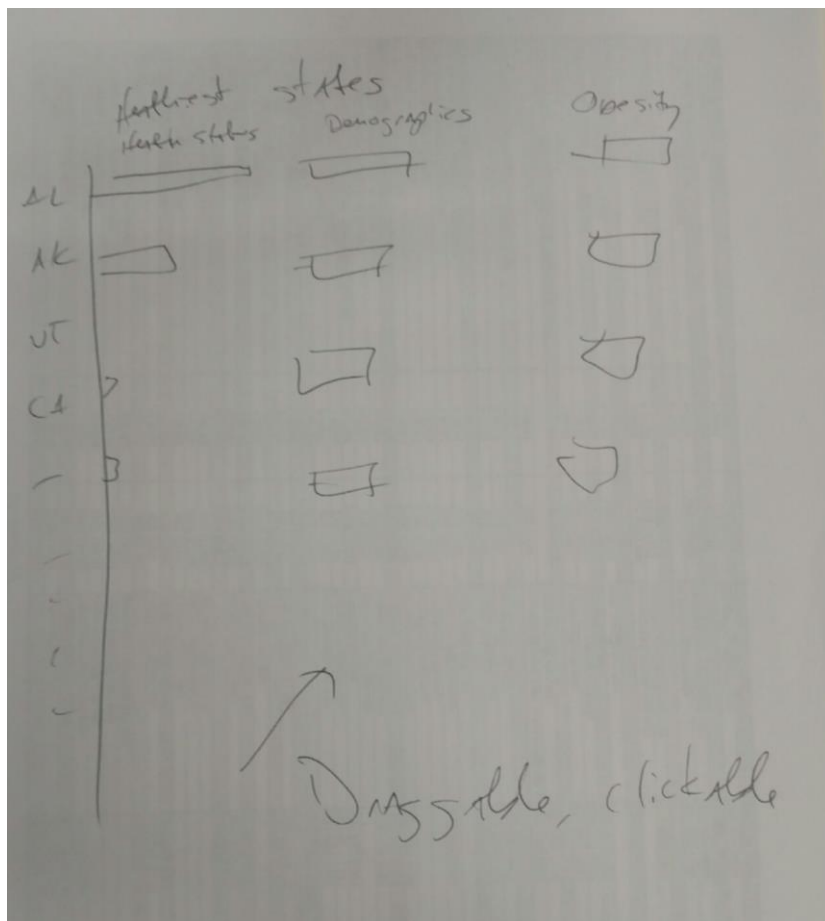
Project repository location: <https://github.com/croddin/dataviscourse-pr-a-look-at-healthcare-in-the-us>

Related Work

13 November 2015

We'd like to implement a stacked bar chart, similar to the one shown in class regarding college choice. This could be a top 10-20 counties or states that are the healthiest, and then a look at things like obesity, preventative services use, etc.

This chart would be interactive, in that users could drag bars around and reposition the order of the states/counties to look at the things of most interest to them.



Data

13 November 2015

The main source of our health data, related to outcomes will be the following dataset:

Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease and Cancer - <https://catalog.data.gov/dataset/community-health-status-indicators-chsi-to-combat-obesity-heart-disease-and-cancer>

We've decided to use this as our primary dataset. If time allows, we may implement some of the other data regarding spending.

We've eliminated some of the files that don't work for our purposes, and will work on eliminating some of the columns, too.

We still need to aggregate the data as state data, though it may be interesting to give users the option to switch between state data and county data.

Exploratory Data Analysis

13 November 2015

This early on, it's hard to see many patterns, but I'm encouraged by the visualizations that we are seeing. The scatterplot looks at various parameters against the self-reported health status of each state. To that end, you can see some patterns starting to emerge. It's interesting, that when comparing obesity rates against health status, for example, there seems to be a correlation between higher obesity and higher health status. What that means, or why, exactly that is, will take further exploration.

Design Evolution

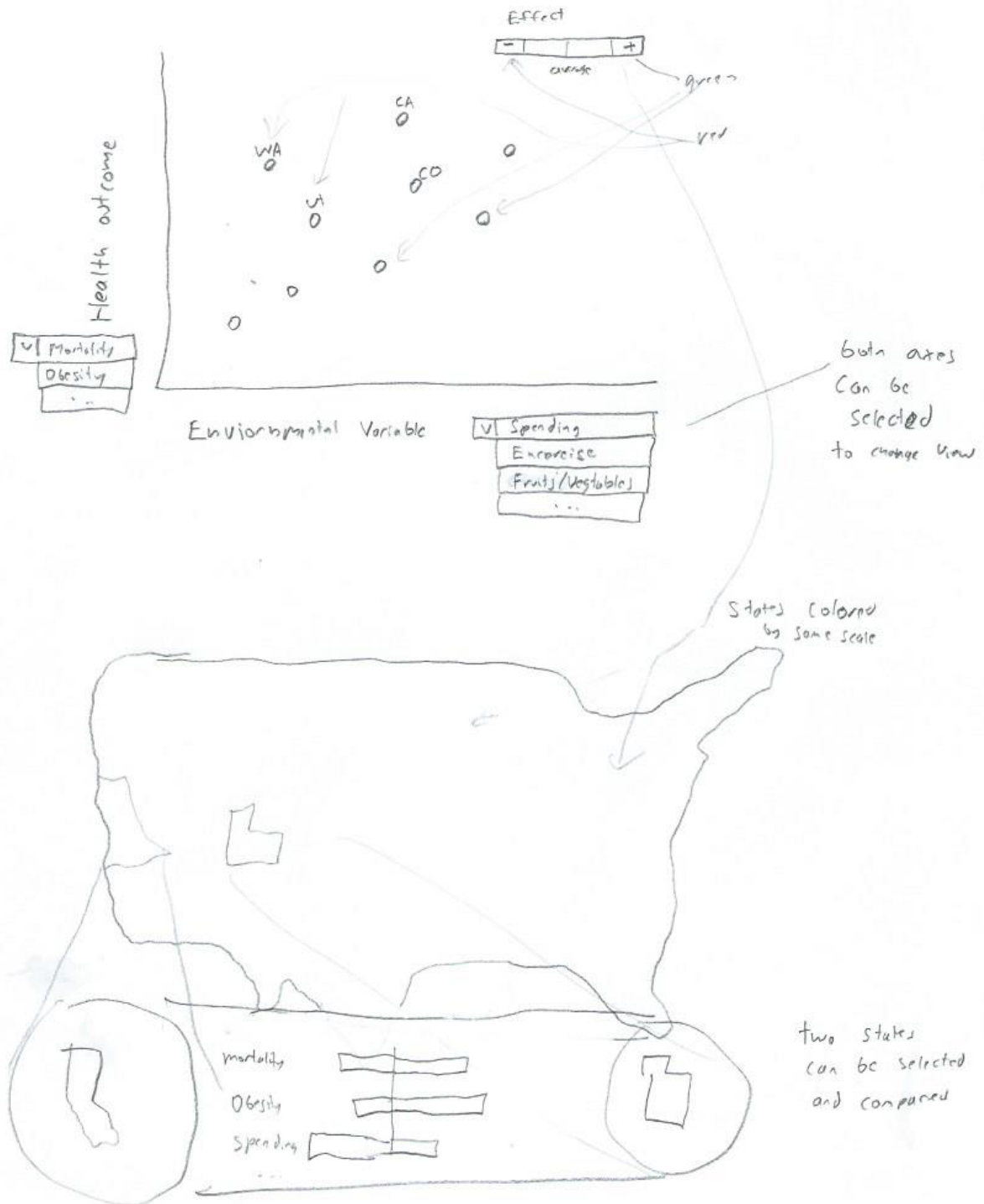
We both sketched out a number of options to determine the best way to represent the data. In the end, we decided that a scatter plot was the best way to observe a relationship between two variables. We'll show national data in a scatter plot where users can select the environment variables and health outcomes they'd like to examine.

Position would obviously be important in a scatter plot—the higher up on the “cancer” axis, for example, the worse the outcome, but we'd also like to redundantly encode that information using color. A low-spending state with good health outcomes, for example, would have a blue circle representation on the scatter plot, where a state with high spending and poor outcomes would be colored red.

We're also going to use a map as a state selector—people can click on the state they want to look at more closely in either the scatter plot or the map to see detailed numbers. We'd also like to implement a feature where people can compare states numbers side by side by selecting two states.

The states on the map will also be colored according to a metric such as health outcome or spending or population. This would also be selected via interaction.

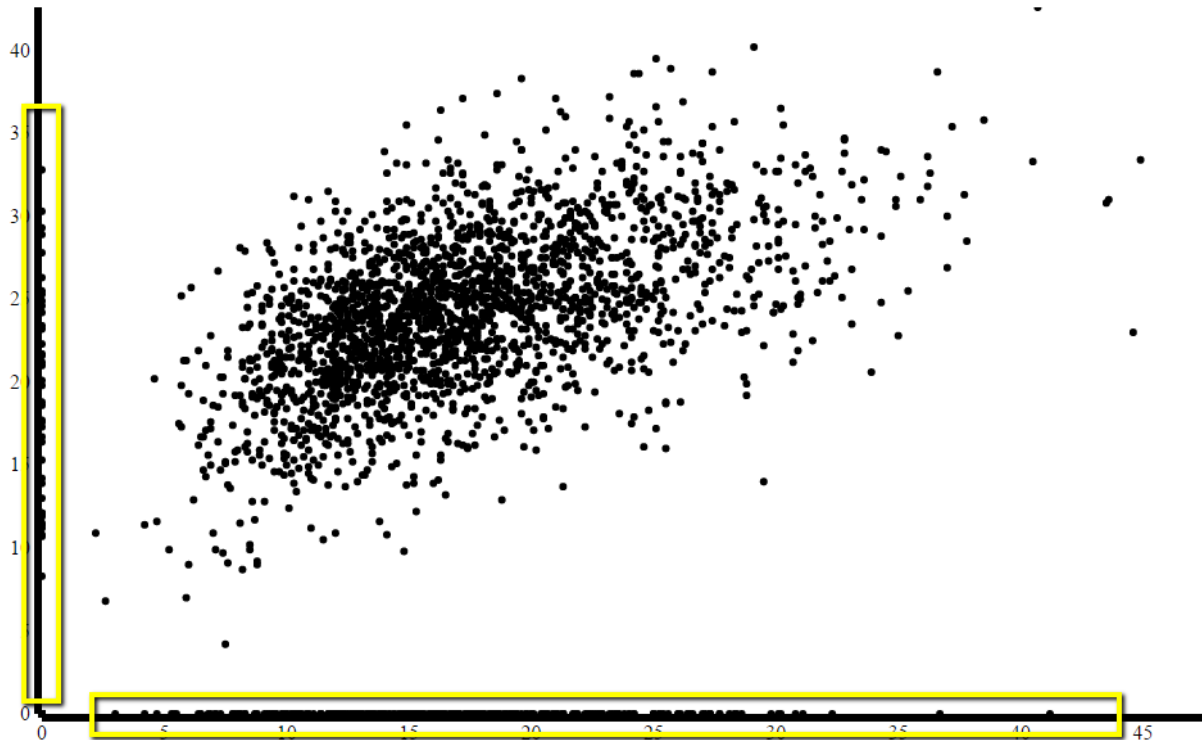
Here's a look at our final sketch for the visualization:



13 Novmeber 2015

This early on, we're already running into questions.

For example, in the scatterplot, how do we deal with zeros (as here):



We also need to come up with a good way to label the axes so that they're clear.

Another question is the process of selection. Right now, we're just using the column names straight from the data.

s:

This is obviously not ideal—we need to come up with clear names as well as an explanation for what each parameter means.

As mentioned earlier, originally we were just going to focus on aggregated state data, but after implementing the map and the scatterplot with the county data, it may be useful/interesting to give users the chance to decide which view they want to take.

We haven't linked the views yet, but we also need to come up with a good solution for that. We'll probably implement some sort of hover/click method, but it will take some experimentation to come up with a good solution.

We'd also like to highlight some interesting features of the data—for instance, if there is a correlation between obesity and health status, we'd like to point this out and suggest options for users to further explore the data.

When we implement the state view, we also need to think about how we're going to present that/allow users to select states or counties.

.