Process Book

CS 6630 | 4 December 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.
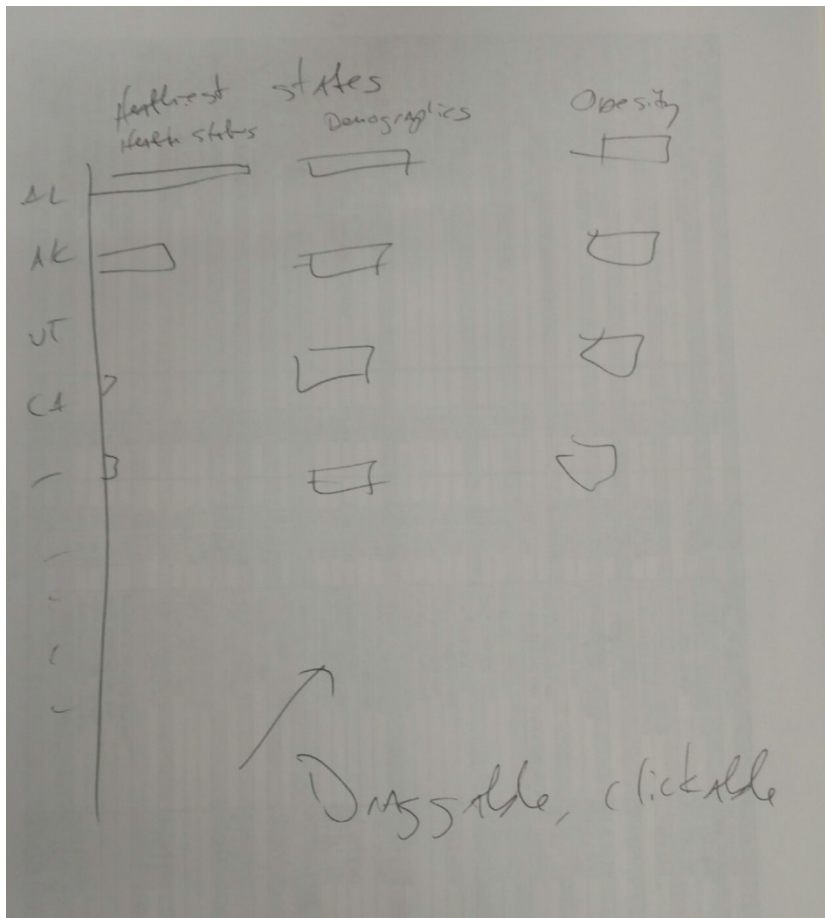
2 December 2015

The final bar chart ended up being a variation on the one described above. We decided that it would be best to have two charts showing the healthiest and unhealthiest counties in the United States. The reason for this was clarity—we wanted to depict clear metrics showing health outcomes that would support the scatter plot and the map that were already implemented.
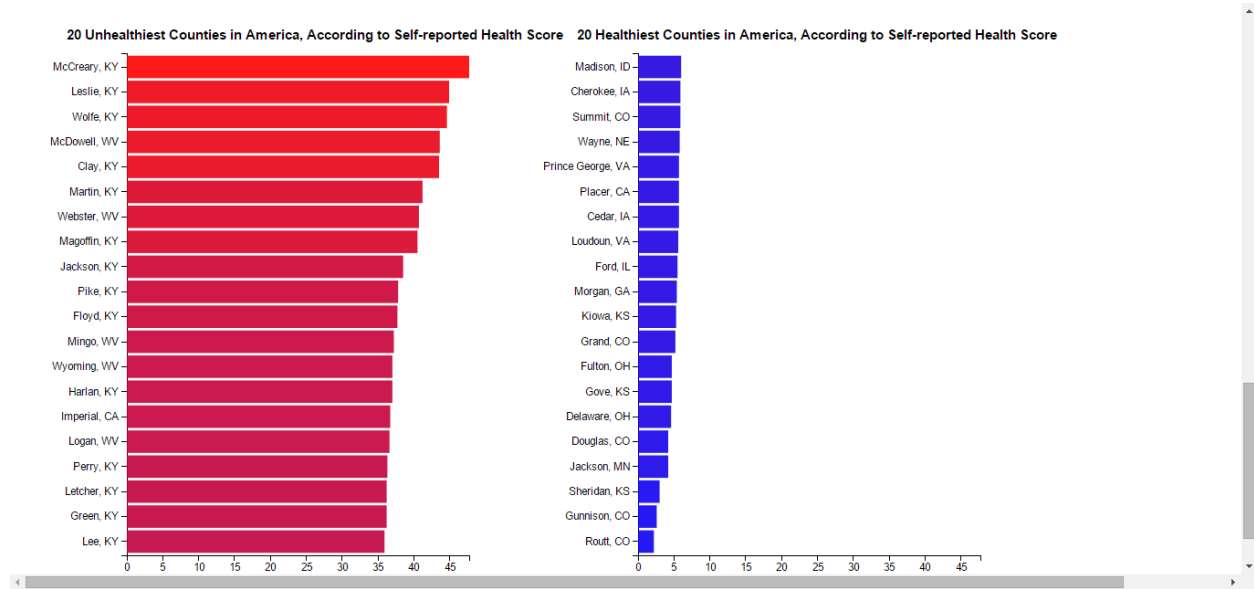
At first, this was our solution:
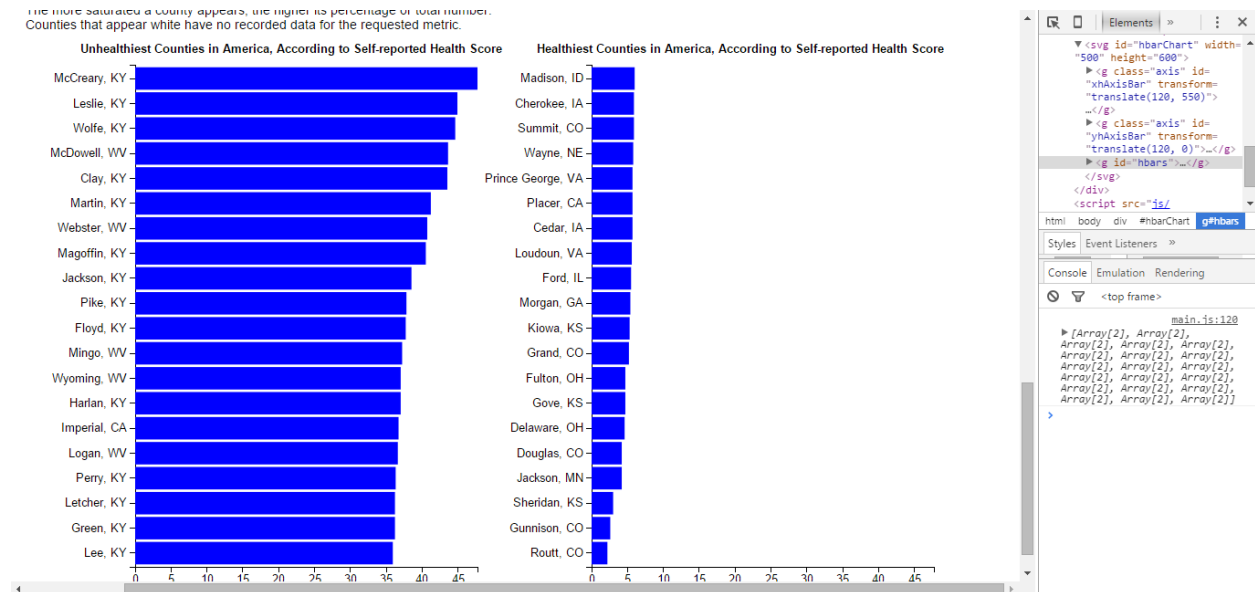
This wasn't very clear—it looked like the counties on top of the chart were actually the ones that were the healthiest, when the opposite is true.

To correct this, we added a color scale that uses a red/blue mapping where red represented counties that had a "worse" health score, and blue represented counties with a "better" health score. We hoped that by using this well-recognized color scheme (red = worse, blue = better) that it would be more clear to users, at a glance, which counties reported better numbers.

We also added some additional explanation at the bottom of the chart. You can see the final result here:

20 Unhealthiest Counties in America, According to Self-reported Health Score

County	Score
McCreary, KY	
Leslie, KY	
Wolfe, KY	
McDowell, WV	
Clay, KY	
Martin, KY	
Webster, WV	
Magoffin, KY	
Jackson, KY	
Pike, KY	
Floyd, KY	
Mingo, WV	
Wyoming, WV	
Harlan, KY	
Imperial, CA	
Logan, WV	
Perry, KY	
Letcher, KY	
Green, KY	
Lee, KY	

20 Healthiest Counties in America, According to Self-reported Health Score

County	Score
Madison, ID	
Cherokee, IA	
Summit, CO	
Wayne, NE	
Prince George, VA	
Placer, CA	
Cedar, IA	
Loudoun, VA	
Ford, IL	
Morgan, GA	
Kiowa, KS	
Grand, CO	
Fulton, OH	
Gove, KS	
Delaware, OH	
Douglas, CO	
Jackson, MN	
Sheridan, KS	
Gunnison, CO	
Routt, CO	

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.

4 December 2015

This was a massive dataset, and we're using a lot of data from it. We did pare down the data to be more manageable, however. First, we removed the quartile data that was available as well as data that didn't seem to be useful for our purposes.

Our primary goal for this visualization is to find patterns in the data that so people can see what things, generally, make a difference for better health outcomes. So, then, we focused on both positive (preventative services) and negative (risk factors) that may have an effect on health. We also left in demographic data to see if there was a correlation between certain populations and healthier outcomes.

As part of our data processing, we also attempted to make the data more human-readable by, for instance, adding longer descriptions and human-readable column names.

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 worse health status.
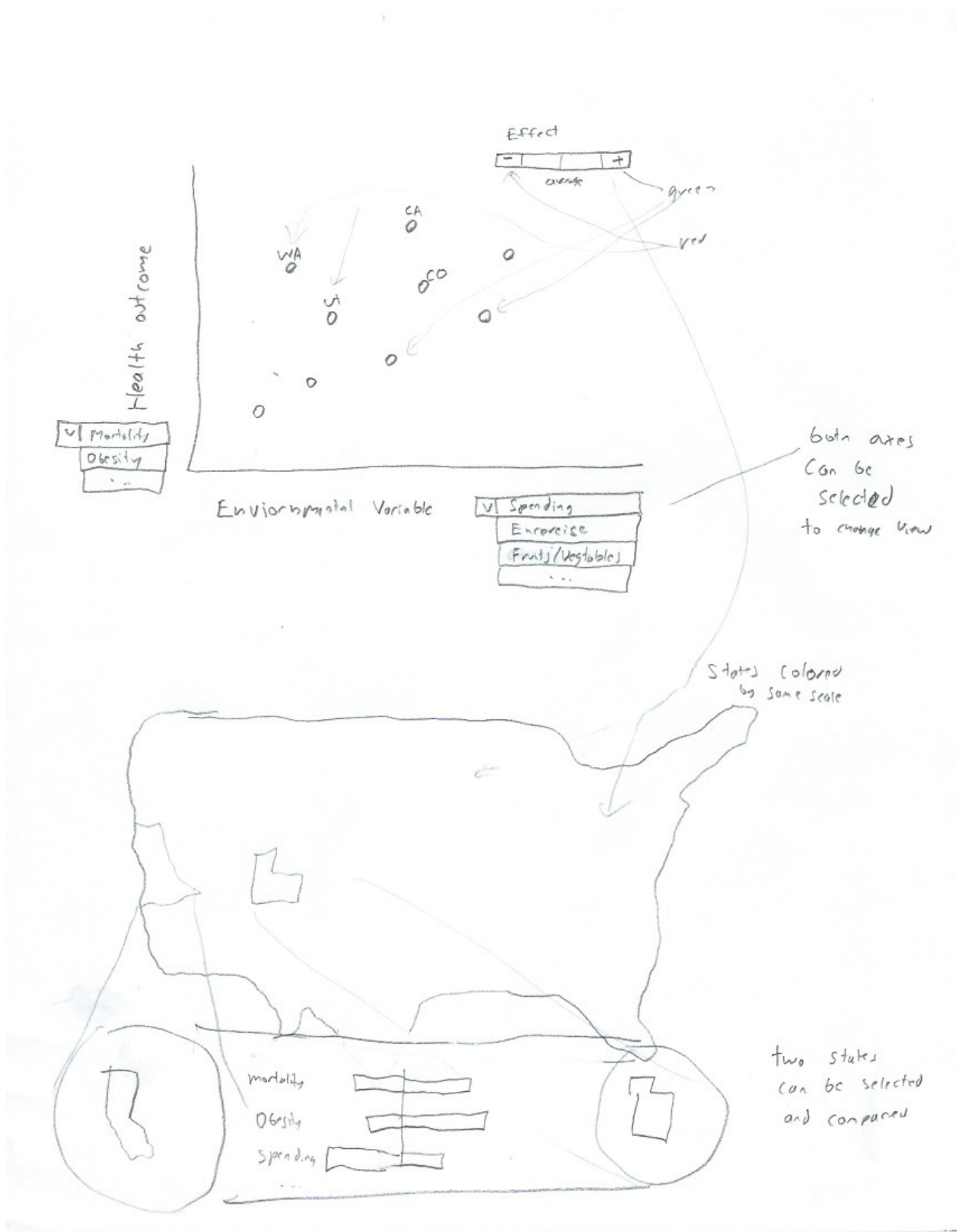
4 December 2015

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 be 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:

Effect

− average + green

red

Health outcome

CA

WA

50

CO

Mortality
Obesity
...

Enviornmental Variable

Spending
Exercise
Fruits/Vegtables
...

both axes
can be
selected
to change view

States colored
by same scale

mortality

Obesity

Spending

two states
can be selected
and compared

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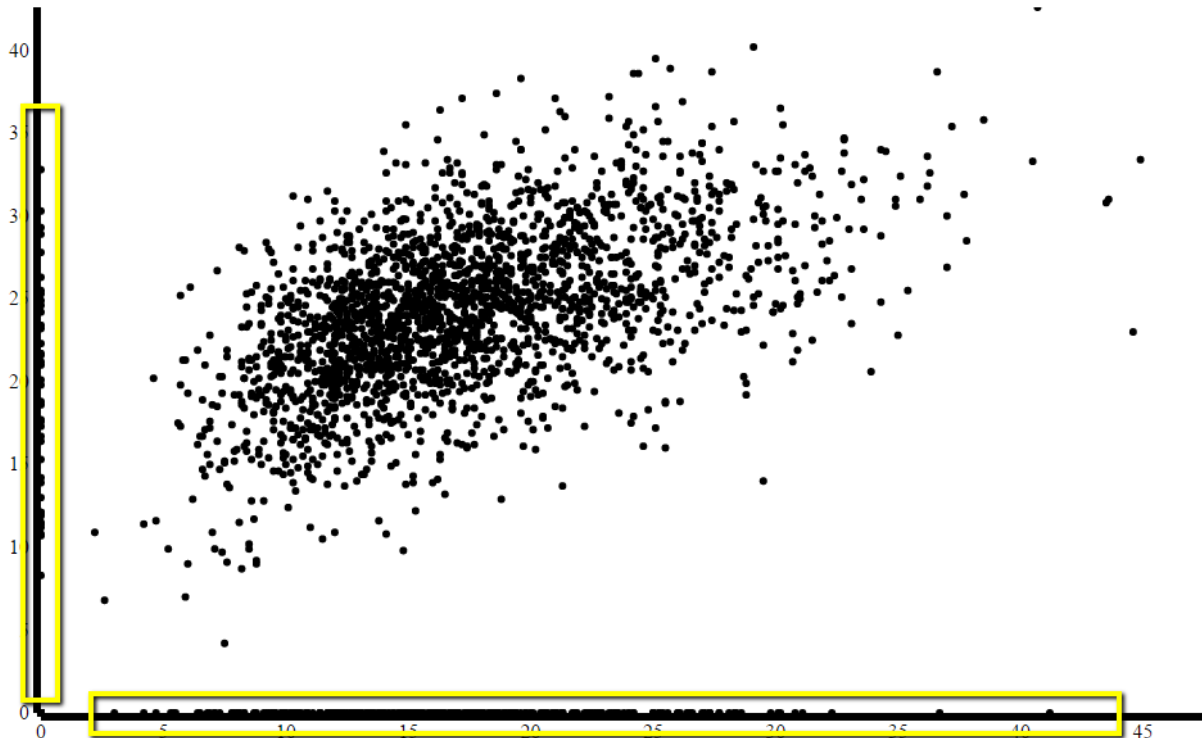
We also need to come up with a good way to label the axes so that they're clear.

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

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 intersting 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.

4 December 2015

Our final design solution has stayed pretty close to what we had envisioned with our project sketches.

As mentioned, we did have a problem with the "zero" data, and ultimately decided that it would simply be better and easier to filter that data out, like so:
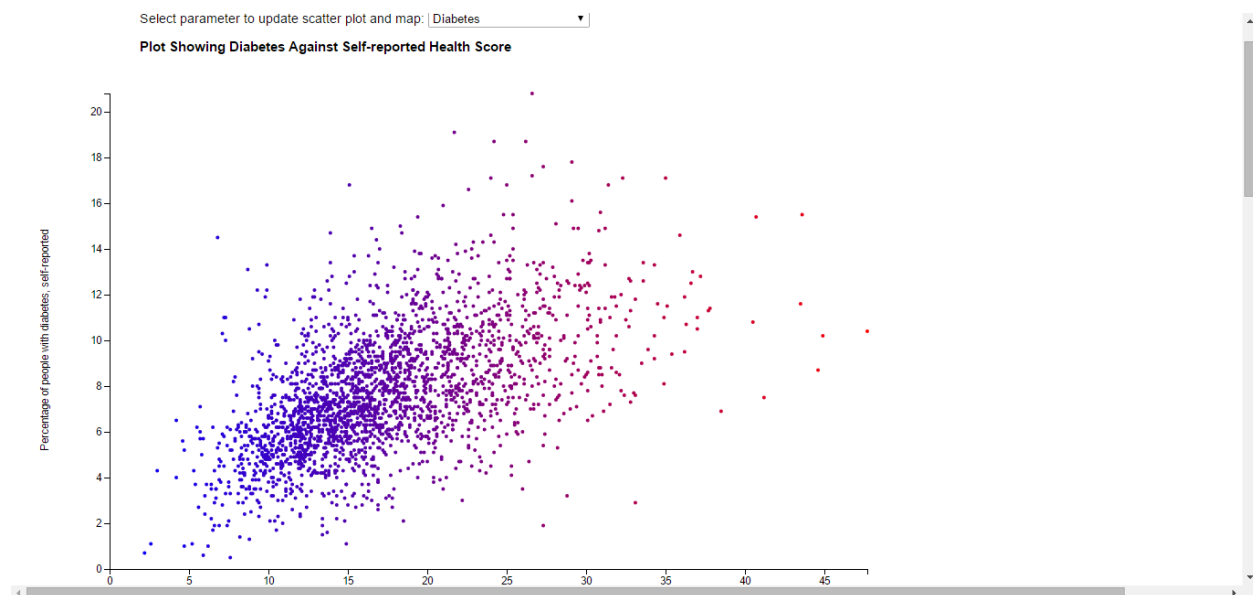
We also did come up with some human readable names and descriptions that appear when the user selects the parameter he or she wants. An example is below:

Percentage of adults reporting eating fewer than 5 fruits and vegetables a day

We used these descriptions to inform our tooltip design for hovering, too. We wanted it to be clear to the user at all times which county they were looking at and the exact values that were being depicted:
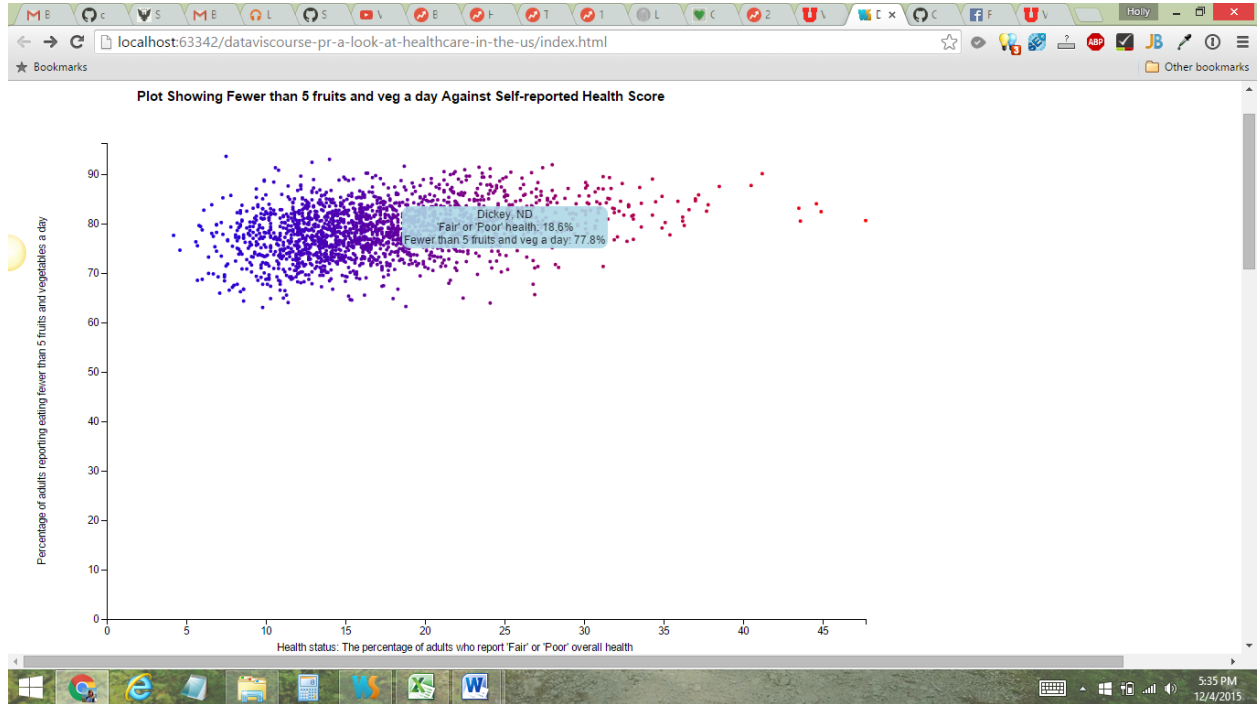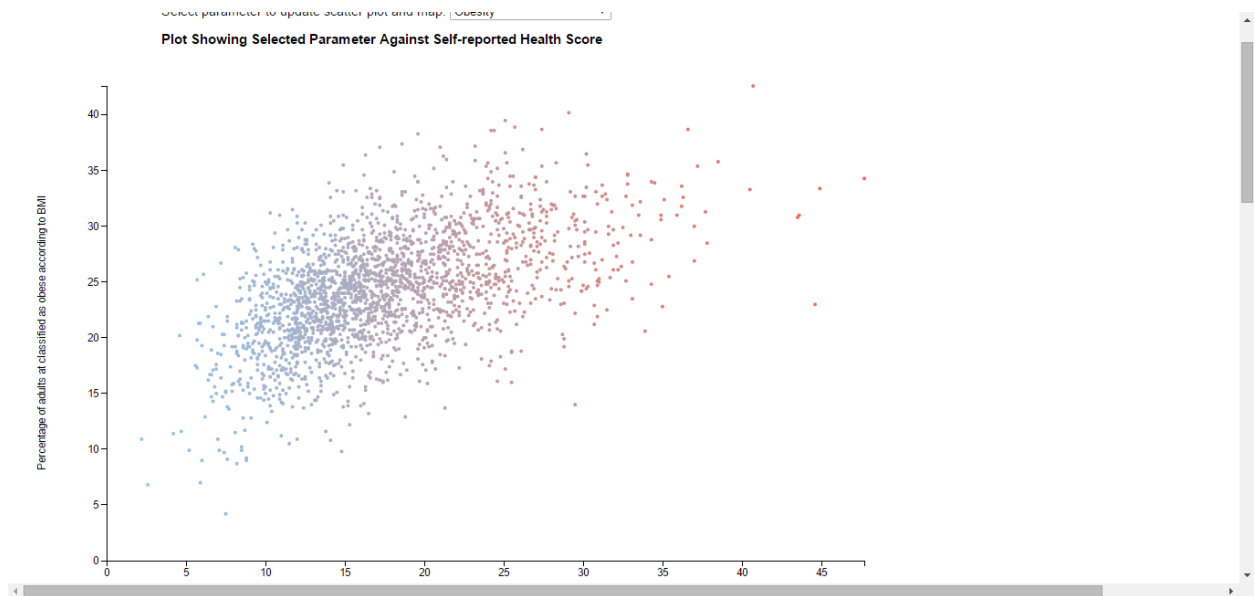
We also considered how to best represent the change in health status for the scatter plot and map, and for both we used a color/saturation representation.

Here's what the plot looked like before the color scale was added:
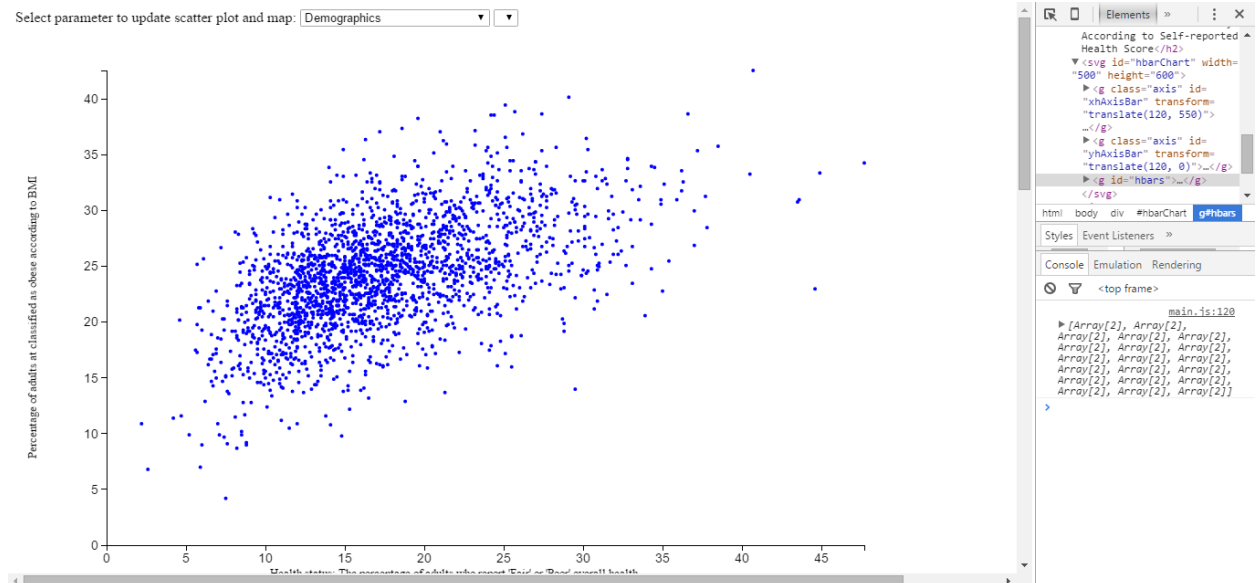
And after:

Implementation

As already discussed, our design features four main views: the scatter plot, the map, the "healthiest/unhealthiest" bar charts, and the comparison view for a more granular view of the stats for each county.

Each of these views provides a slightly different view of the data, but are linked with each other.

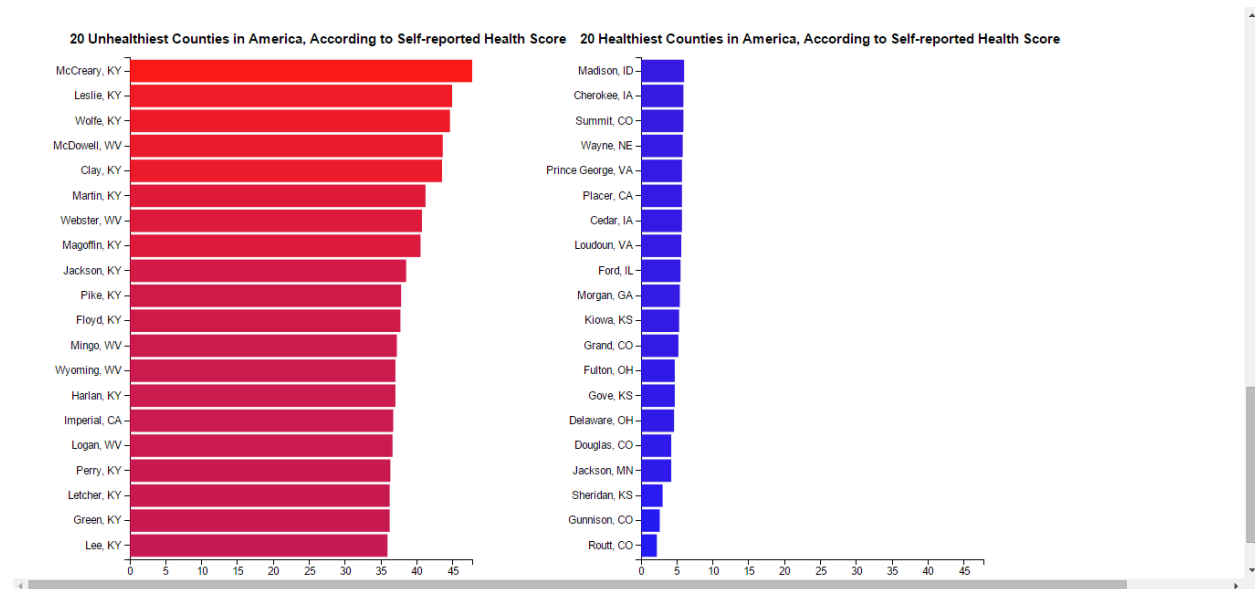The scatter plot shows how each metric compares to the overall health score of the counties. This is good for showing potential correlations and ways in which the data interacts:

The map view shows trends nationally and allows for deeper discussion of policies and changes that may be affecting health in certain geographical locations:

Map Depicting Health status: The percentage of adults who report 'Fair' or 'Poor' overall health by County
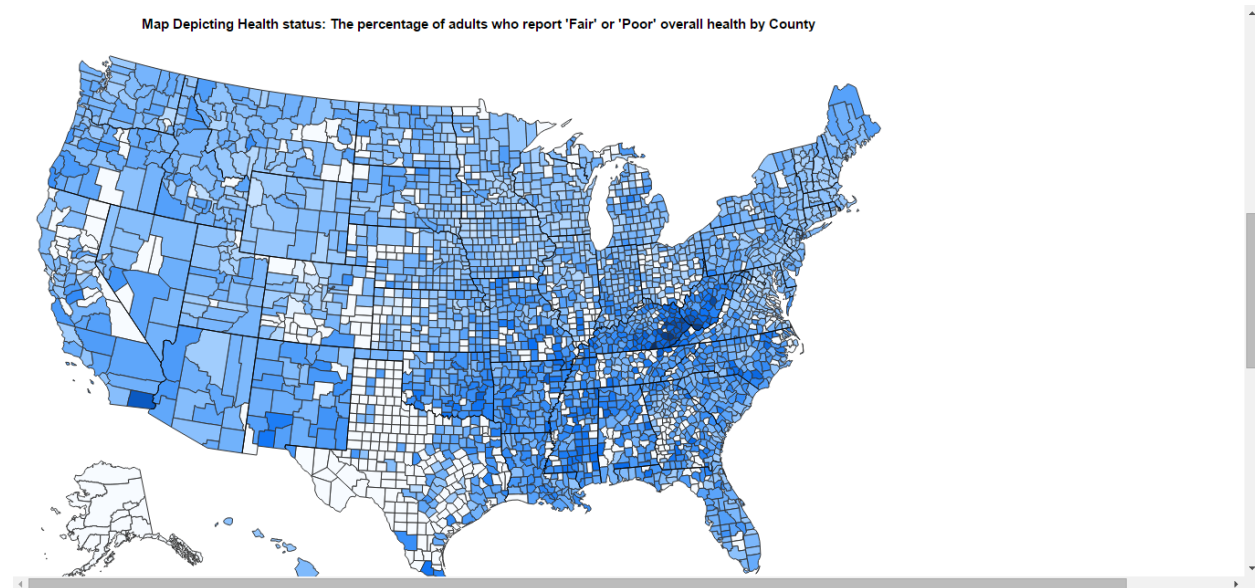
The bar chart view shows the top 20 healthiest/unhealthiest counties, and, like the map, allows for an instant view of which regions are doing well or poorly in the area of public health:

20 Unhealthiest Counties in America, According to Self-reported Health Score   20 Healthiest Counties in America, According to Self-reported Health Score

The comparison view allows for a deeper understanding of how two counties differ in terms of metrics. This makes it easier to compare counties side by side to look at how regions with similar demographics may differ in health outcomes:
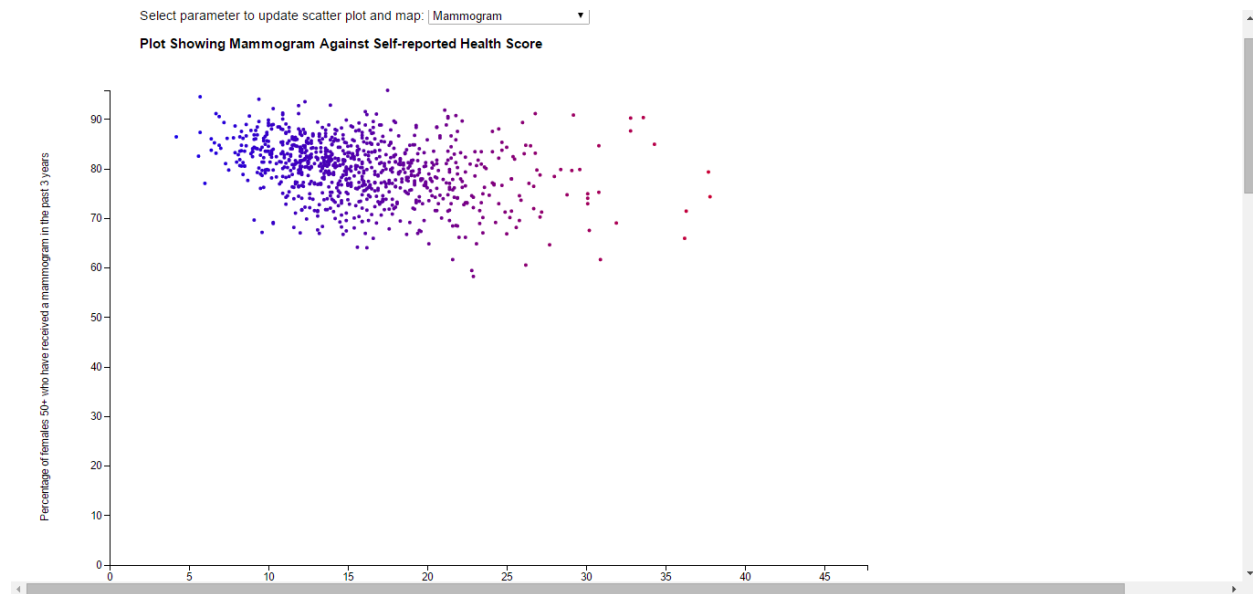
Evaluation

Our data shows some clear correlations between health status and some risk factors that you would expect like obesity:
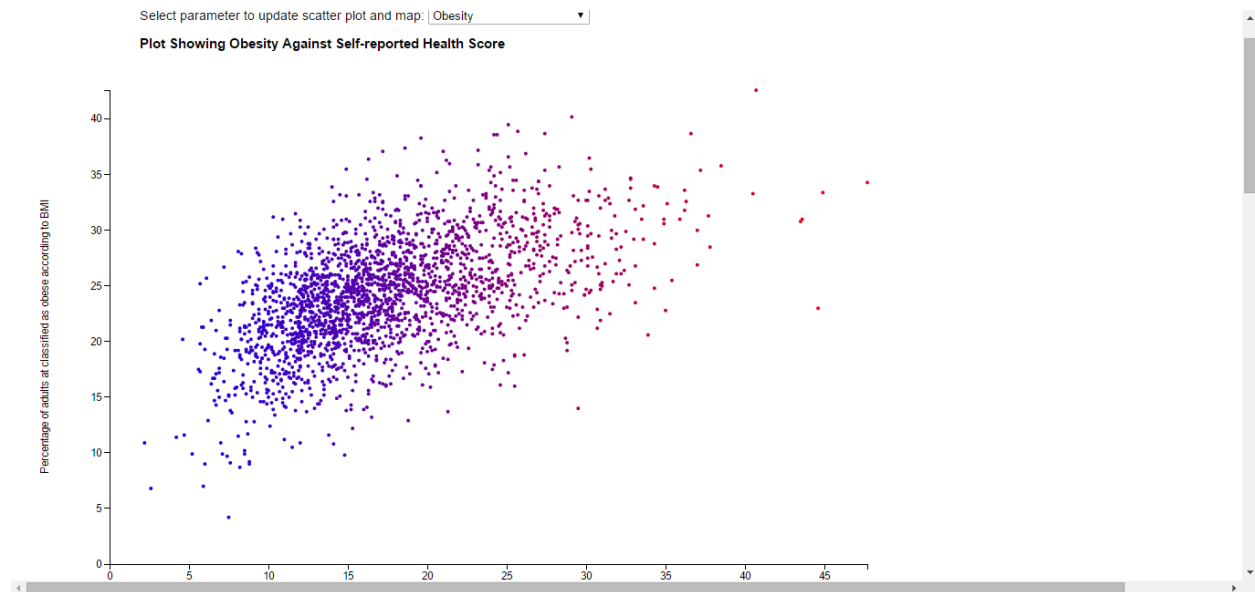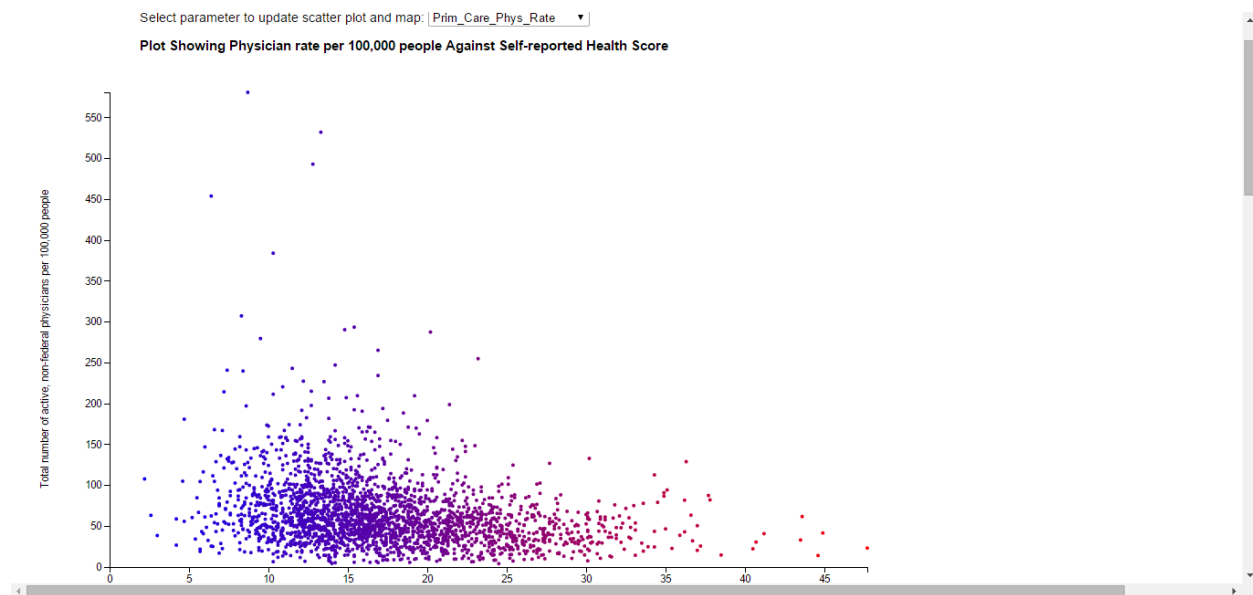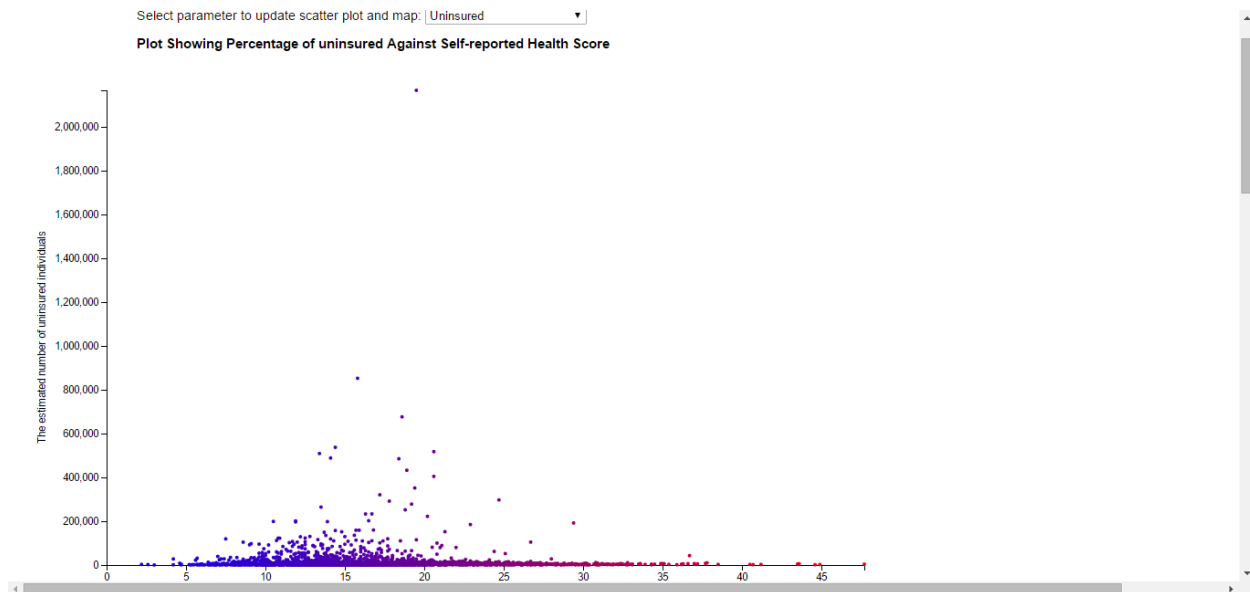

And diabetes:
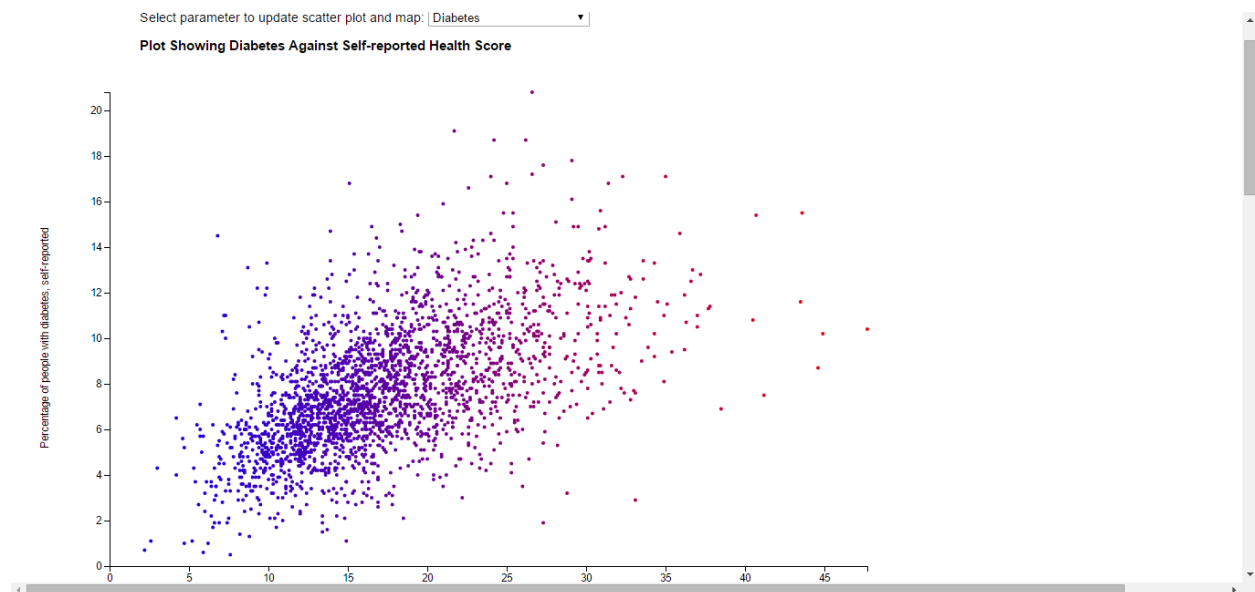
But just as interestingly, in our opinion, is the fact that the data do not show some correlations where you might expect them to be, such as the percentage of women who get mammograms:
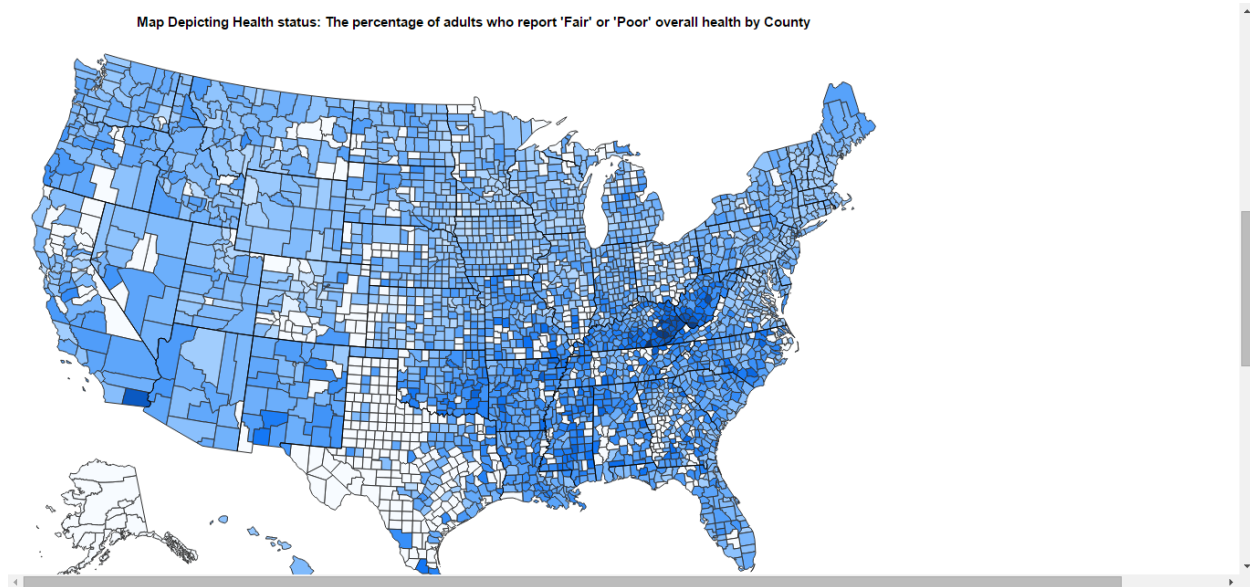
Or the rate of physicians per 100,000 people:

Or, perhaps most telling for healthcare policy, the number of uninsured:

It is also interesting to look at the health status of the country as a whole. You can identify pockets of health and unhealthiness and ask what kinds of policies or environments are creating these conditions:

While we cannot come to clear conclusions from these visualizations alone, of course, they may suggest areas for further study.

While we do feel like our visualization is effective, there are almost certainly ways we could improve it. One way to do so may be to use additional metrics to evaluate health outcomes—the self-reported health score is useful, but, practically, limited. It may be interesting to find a more objective and measurable metric for health.