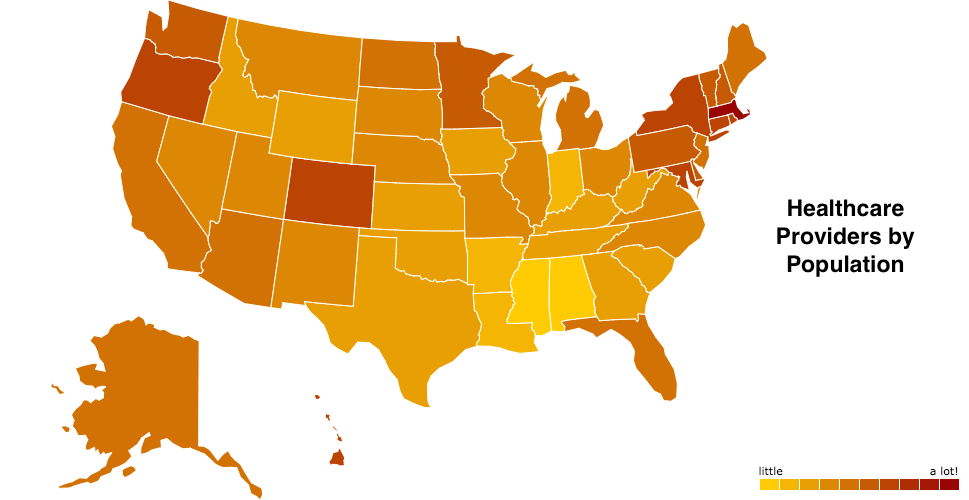
BEM 150: Check-in 1

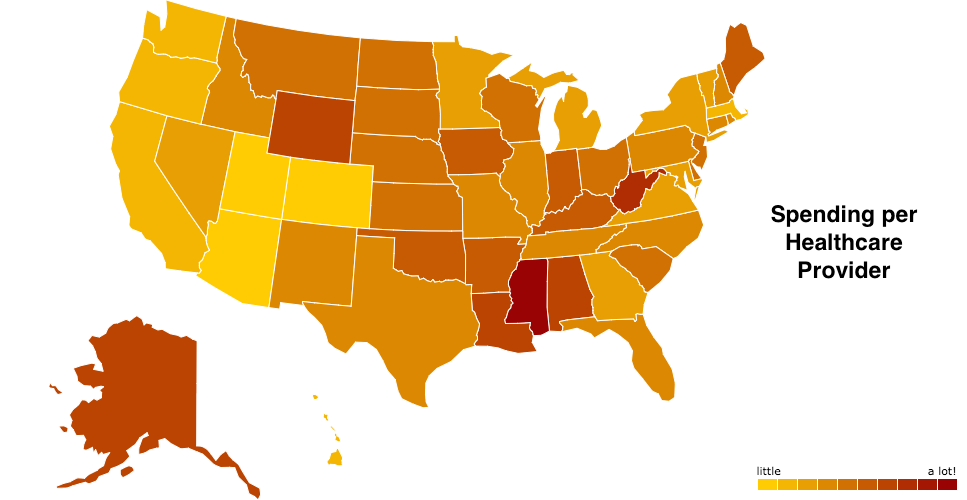
Advith Chelikani, Nick Haliday, Logan Cross

We’ve done some preliminary exploration of our data, but have run into many issues with Factual data (which will be discussed later). For our first visualizations we contructed the following:

Below is a heat map of the United States based on number of health care providers per person available in the state. We see the New England area is on the higher end of the spectrum while many Southern states are lacking. To construct this visualization, we used population date from the census (2015) as well as the number of providers in Factual’s healthcare database for each state. The visualization is actually an interactive one made with JavaScript’s AMCharts library.



The second visualization (below) we performed involved using regarding healthcare spending per capita for each state. We took each the product of spending per capita and population (total spending) and divided by the number of healthcare providers to create our second heat map. We notice that a lot of the states with fewer health care providers also spend the most per healthcare provider.



We ran into a lot of problems with the Factual data. Firstly, we cannot download the data so we can only access the data through specific calls to the API. What makes this even worse is that each call is limited to having only 50 results. For example, if I queried Factual to find how many health care providers are located in Texas, it would only give me the first 50 results. This is very problematic as it is very difficult to run analytics on the data when retrieval is structured in this way. Factual has a dashboard where you can make queries to the entire data set with unlimited results, but it’s a graphical interface so there isn’t anyway to do it programmatically. As a result, to make the heat map above, I had to run a query on the dashboard and manually copy each of the 50 results over to a text file. I emailed the Factual team but they said the only way to download the data is to pay for a yearly license. We’d love some assistance on figuring out how to best work with the Factual data in this respect.

Also, we noticed that while almost none of the location data was missing, a huge portion (>50%) of the education, years experience, and affiliation data is missing. Again, this is a pretty major roadblock and we aren’t sure what to do about it.