

Visualization of Obesity Clinical Trials

Team members: Shawn Dai, Enming Zhang

Motivation

The obesity rate increases significantly each year in the US. Obesity has become a serious health issue facing the US. To understand the relation between the obesity rate and obesity-related clinical trials, we planned to create a visualization to display information including the obesity rate and related clinical trial information in previous years, from 2003 to 2017, including co-studied conditions, interventions, study type, enrollment amount, gender, study phase, and funding information. We are trying to explore the trends of obesity, trends of the obesity-related clinical trials, and the relationship between different variables.

Our initial research questions are:

- What is the distribution of obesity trials in the United States?
 - Which state in the United States has the most clinical trials of obesity and which is the least?
 - Is there any difference in clinical trials of obesity in different states? If there are differences, what are those differences?
 - Is there any difference in gender and age of clinical trials of obesity in different states? What's the specific feature?
 - What's the progress of the trials in different places?
 - What's the result of those trials if the studies have completed?
 - are the state demographic statistics such as population, obese rate, poverty rate, average income etc correlated with the attributes of clinical trials?
- What are the conditions that are frequently studied together with obesity?
- What are the interventions used?
- How did trials characters change over time? Are there any trends shown by the visualization?
- What are the most common sponsors?
- Are there any patterns on the participants, what are the common study design methods, how do these characters distribute over time and locations?

In order to address the above questions, we collect the following data in our visualization:

- Obesity rate by state
The data on obesity rate by states from 2003 to 2017 are retrieved from CDC.
- State demographic data
State demographic data include median household income, poverty rate, and state population from 2003 to 2017. The data is downloaded from US census bureau.
- Obesity Clinical Trials data
Obesity Clinical Trials data are obtained from clinicaltrials.gov. ClinicalTrials.gov is a database of privately and publicly funded clinical studies conducted around the world, managed by the U.S. National Library of Medicine. We will search the website by condition "obesity", "obese" or "overweight", the location is limited to the United States,

we will only focus on the completed trials, the active ones and withdrawn ones will be excluded. Trials that meet the search criteria can be downloaded directly from the website in XML format.

After getting the data, we used python to parse the XML files to extract the fields we want, then we cleaned those variables and aggregate the data to the structure we needed for the visualization.

Design process

To display the obesity rate and related clinical trial information in one page and also make that information easy to compare and analyze, we plan to use a timeline to separate the each year's information. To visually separate the obesity rate and clinical trial data, we designed an interactive timeline in the middle of the page. Above the timeline, we used a choropleth and a line chart to display the obesity rate in each state and in different years. Below the timeline, we used different individual charts to display clinical information from different types (Figure 1).

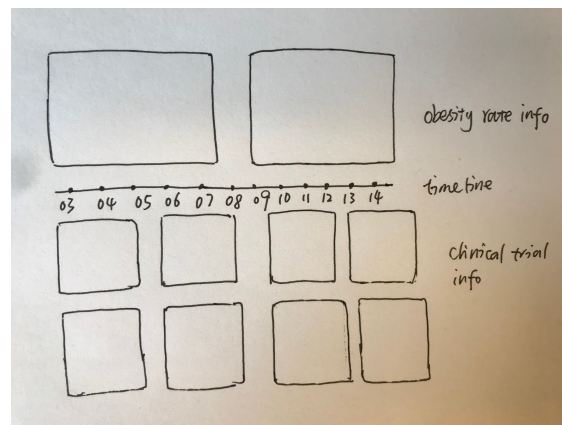


Figure 1. The layout of the visualization

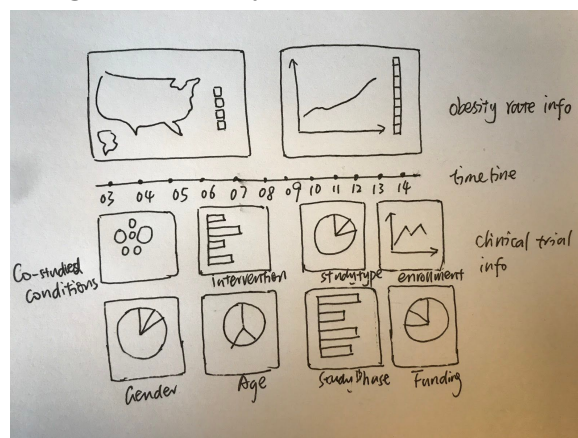


Figure 2. The details of the visualization

We plan to make the timeline as an interactive line. After hovering on the point indicating the different year, the obesity rate and the clinical trial information will change simultaneously,

showing corresponding information in that year. To explore the more details on the obesity rate, we used two different graphs to display the comparison of obesity rates among states and different years. On the left side, we used a choropleth to display the obesity rate in different states and different years. The severity of the obesity show different colors on the map: the darker the color, the much severe the obesity issue. Simultaneously, we tried to put the point data on the map representing the number of obesity-relevant clinical trials. We represent that information as circles on the states and the radius of the circle indicates the number of the clinical trials: the larger the circle, the more clinical trials in that state. Thus, it's easy to compare the obesity rates and the number of clinical trials among different states in each year by observing the choropleth. However, it's not easy to observe the trending of obesity rates. Thus, we designed the right side graph, the line chart, to display the trending of the obesity rate in each state. By clicking the legend of the line chart, it's easy to compare the trending information between or among different states (Figure 2). This line chart is also used to show how other State demographic statistics change over the year, the statistics include population, poor and poverty rate and median household income. We use a drop down menu to let user choose which State demographic they want to see.

For visualizing the clinical information, we chose different charts to display such information according to the number of the categories in each variable. For variables containing more than one hundred types, we chose to display them in bubble charts, for example, the co-studied conditions. For variables containing seven to ten categories, we chose to display them in histogram charts, for example, the intervention and the study phase. For variables containing less than six or five categories, we selected to display them in pie charts, for example, study type, gender, age, and funding source. All the clinical trial information change simultaneously in response to the hovering action on the timeline.

Design Justification

We designed a choropleth and a multi-series line chart to display obesity rate information. Then, we used individual charts to display clinical trial information.

1. Colorblind friendly

We used colorbrewer when choosing the color schema for the choropleth, to make it colorblind friendly.

2. Stability of graph

When hovering on the timeline, the choropleth and the clinical trial charts can change smoothly and simultaneously to show the corresponding information in the specific year, making it easier to explore and analyze the data.

3. Appropriate chart type selection

Most of the clinical trial information are categorical variables. According to the number of categories in each variable, we selected appropriate charts to display the information, making it easier to read and distinguish between different categories.

Thus, we believe our visualization is efficient and effective to conduct exploring and retrieving the required information and answering questions.

Findings

After building the visualization, we found that:

- The obesity rate is growing up fast nationally. There is also an upward trend for population and household income. There is no obvious trend for poverty rate.
- The amount of obesity-related clinical trial is increasing as time passing by, as well as the obesity rate. By exploring the choropleth, we found that Florida had the most obesity-related clinical trial among other states.
- We can also notice that there is a difference in the number of clinical trials in different states. It seems that more clinical trials are on the east side of the country relative to the west side of the country (Figure 3).

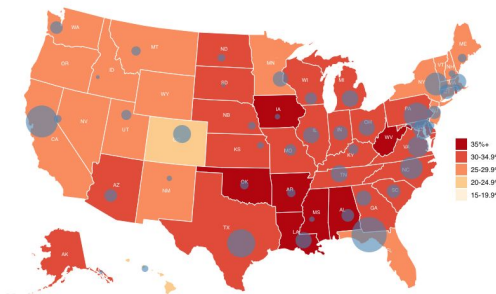


Figure 3. The obesity rate and the number of clinical trials on obesity in 2017

- To observe the Co-studied condition graph, we could see that diabetes has been studied most relevant to obesity in 2017, following by metabolism and nutrition disorder.
- The intervention used most is behavioral action from 2003 to 2017.
- For funding source, we found that NIH is the main funding source for all the clinical trial studies except others.

We also found that Colorado is the state with the lowest obesity rate in almost every year. However, the obesity rate and the number obesity-related clinical trials are still increasing.

Reflection

Through the project, we have experienced the whole process of building a visualization product to understand and explore data. By interacting with the visualized product, we could answer many questions more efficiently and effectively that we might not be able to do by only getting access to the data. Interactive visualization provides researchers and users a more direct way to explore and analyze the big deal of data we get.

Limitation and future work:

If we have more time, we can continue working on the choropleth map, adding more details about the obesity-relevant clinical trial information in each state. For example, enriching the clinical trial information in the label when hovering on the circle predicting the number of clinical trials in each state. We also consider connecting the choropleth and its right line chart to make them more holistically. For example, hovering on the state on the choropleth, it's better to show the state's trending information on the right.