Jacobs University Bremen Apr, 23, 2018

Frenci Rakipi, f.rakipi@jacobs-university.de Fanlin Wang, f.wang@jacobs-university.de

Project Milestone 1 Report

Link to the webpage prototype: `https://frenci8.github.io/vis-project/`

It has been a common understanding that open national budget data is important for effective governance. Publishing budget data alone, however, is not enough, since few people would be interested in tedious raw data. In order to open up for increased citizen attention and participation, it would be meaningful to design a visualization method that provides tools for a wider audience. Our web page, which takes US budget on education as an example, will be based on dataset mainly from `www.usaspending.gov` and `data.gov`. Our visualization will be showing how public funds develop over time and what is the geological distribution of the money. Such visualization, for example, will give information about the emphasis of government education policy by analyzing the distribution of federal spending on different disciplines, and furthermore serves as an indication to the underlying trends in job markets and future active field of research. Another user case could be allowing for comparison of the geological distribution of educational input with other economic indicators. Current available tools for demonstrating federal budget in this way include IPEDS trend generator[1], which provides an excellent tool for interactively exploring educational budget distribution and displaying the trend. Based on that, we want to further improve such tool to make a clearer user-interface, and add additional features such as displaying geological distribution.

So far we have implemented a treemap of displaying distribution of federal spending on different kinds of education. We plan to add several visual encodings such as scatterplots, choropleth maps and line charts to further illustrate the different ways in which budget and spending data can be visualized to bring out correlations between variables, to show trends in spending over the years, or to enable users to easily see where money has been spent. Interaction techniques like filtering, zoom, and linking will also be implemented to meet the need of users. We believe that the overall design will make it easier to understand the meaning of financial budget reports.

1. Scatterplot:

We need scatterplots because they are effective for showing the correlation between two attributes. Here they can be used to for users to judge the correlation between the education budget and the economic indicators chosen by the user. In our case, each point mark represents a region of the US, while the horizontal and vertical spatial position encodes the amount for educational budget and the economic indicator of respective region correspondingly. The reason that scatterplot can be applied here is that since we only have less than 100 regions, the number of points is not so huge and the scalability of a scatterplot is well suited.

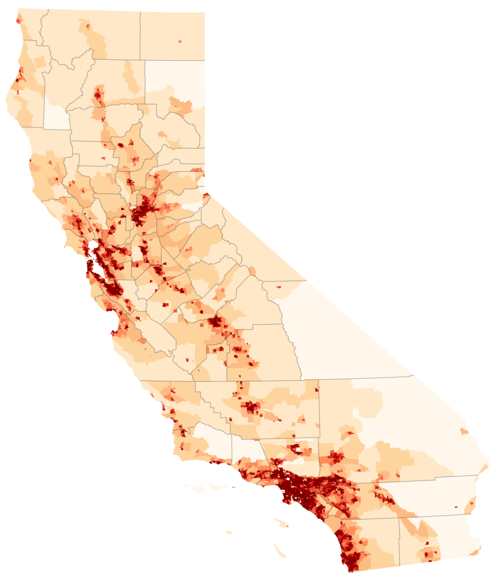
An alternative way of plotting correlations could be using parallel coordinates, which allows for visualizing many quantitative attributes at one using spatial position(Munzner, 2014). They can be used for judging correlation over multiple variables and they are scalable for hundreds of points. However, it takes more than intuition for our users, who by assumption should not be required to taught on design technique, to understand the pattern of parallel coordinates.

Scatterplot Matrix (SPLOM) is a matrix where each cell contains an entire scatterplot chart (Munzner, 2014), and is heavily used for finding correlations; the reason it is not optimal for our study is that so many scatterplots at one time can be a little overwhelmed to the users.

2. Choropleth Maps

   Regarding geographic data, we chose choropleth map which shows the quantitative attribute(the education budgets) encoded as color over regions. [2]

   Figure 1: An example of choropleth maps

3. Scented widgets

It is a common design choice that is employed in almost every interactive visualization design. The usage is straightforward and effective in reducing elements of no interests to the users. The effect of filter can be immediately shown through dynamic queries, according to Munzner, "a display showing a visual encoding of the dataset is used in conjunction with controls that support direct interaction, so that the display updates immediately when the user changes a setting". In our design, users are allowed to choose certain range of time, different disciplines, etc. In terms of filtering control, we choose scented widgets to display high information density by adding visual encoding information directly to graphical widgets.

An alternative approach introduced in the textbook is DOSFA, which stands for dimensional ordering, spacing, and filtering approach. The advantage of it is that it regards many value attributes, however, the visual pattern is difficult to interpret for our targeted users and our dataset does not have such high dimensions.

Figure 2: Examples of scented widgets from the textbook

In order to implement the visual representation of data, we want to use a tool such that it can effectively help us to organize and display data in a easy yet flexible way. D3 is introduced in class and it provides us with a well-written API. We choose D3 over another option, Tableau, because it gives more flexibility of representing data. Since D3 is based on JavaScript, we plan to use other libraries of JavaScript. For example, JavaScript InfoVis Toolkit allows for building rich interactive visualization; extended d3 API such as d3-geo-projection can be useful in creating a choropleth map.

Reference:

https://medium.com/@mbostock/command-line-cartography-part-1-897aa8f8ca2c

https://nces.ed.gov/ipeds/trendgenerator/default.aspx

Visualization Analysis and Design, Tamara Munzner (A K Peters Visualization Series, CRC Press, 2014).

Task division:

Fanlin wrote the report and Frenci worked on the webpage prototype.