Visualizing Social Mobility in the United States

Zhixin Lin, Christabel Sitienei, and Yen-Ting Wang

ABSTRACT

We present an interactive web tool for visualizing social mobility data in relation to college education, parental income, race, and neighborhood in the United States. This tool explores the technique of "scrollytelling" - storytelling through animation on scroll along with a variety of graph interactivity. It also takes the opportunity to engage the audience to speculate a data point before revealing it, which helps them to better internalize the insights.

Keywords: interactive data visualization, social mobility, college education

1. INTRODUCTION

Almost everyone works hard for a chance at succeeding in life. For some, it is for a chance to move out of poverty. We sought answers to questions on matters of social mobility. The data is published by the Opportunity Insights Project, which contains intergenerational income mobility data from 2202 colleges and over 30 million college students from 1999-2013 (Chetty et al., 2020).

Questions our visualization aims to answer include: How does college education impact social mobility in the US? What are your chances of moving out of poverty? What about your parent's income, neighborhood, and race?

2. RELATED WORK

There have been several past visualizations on this dataset by online media with different levels of interactivity. For example, American Public Media Reports published an article on the topic (Zhou, 2018) that uses bar charts and line charts visualization. The interactivity is limited to tooltips and readers primarily consume text-based content. Another example is the New York Times' interactive project "Economic Diversity and Student Outcomes at America's Colleges and Universities: Find Your College" (2017). This project provides much more diverse and engaging interactivity for the audience to explore the work. It allows users to search for a specific college they care about and benchmark its mobility stats with a variety of visualization. We took inspiration from some of NYT's work, such as the tier-based clustering in the "Family income vs. student income at age 34" scatterplot. However, the extensive use of tables in the project is not very intuitive for the understanding of the complex concepts and massive data that is available. Building upon past works, our visualization aims to a) increase the overall intuitiveness of visualization through interactive storytelling and b) further enhance audience engagement by inviting users to draw on a graph.

3. METHODS

The first step is to understand the related concepts in the data. We have investigated the following aspects - family income, college attendance between age 19-22 income to evaluate social mobility. Social mobility is defined as the fraction of its students who come from the bottom percentile and end up moving to the top percentile, which can be calculated by multiplying access with success rate. The access rate is defined as the bottom quintile of distribution of parents' income by average annual household total income before tax which is adjusted for inflation to 2015. As for the success rate, it is done by characterizing students' earnings at ages 32-34 for 1980-1982 birth cohorts ranked conditions on their parents' income level. Both lists are ranked and computed by percentiles. Such social mobility metric is evaluated for each college. In addition, the dataset includes geographical information such as states and neighborhoods, which we turn into map-based visualization.

The overall visualization storytelling is thus broken into four steps from broad to narrow:

1. Background introduction
2. Social mobility characterized by State
3. Social mobility characterized by neighborhood
4. Social mobility characterized by college

4. RESULTS

We develop a "scrollytelling" website to tell the story about social mobility in the US step by step through animation on scroll and interaction with graphs.

In the beginning, there is a brief introduction about our visualization background and the data source. The audience can easily reference it for more detailed background information. The US map colored by average social mobility rate is shown first to help the audience create a broad picture among all states. The state with darker blue represents a higher average social mobility rate. On the

right-hand side, the map is paired with a table listing top 10 colleges with the highest mobility rate. The audience can select a specific state on the map to check out the information and switch around to compare with other regions.

In the next section, we zoom into the neighborhood scale. The US map is shown again but this time the regions are divided in a more granular scale. There is a color bar showing the ranking of individual income. The more redness means lower individual income whereas the more blueness represents higher income in the neighborhood. On the right, there are two lists of radio buttons to filter - family income and the child race. The parent income level has been categorized into high/medium/low corresponding to 75th/50th/25th percentile ranking. The audience can select a different parent income level and child race to see the geological distribution changes of individual income level.

The last section is visualizing social mobility across colleges. The primary visual encoding includes a scatter plot mapping family income and student income of all colleges on the left and a square matrix representing the mobility rate on the right. The two graphs are dynamically linked so that users could easily observe any patterns among family income, college education, and mobility rate.

For the scatter plot, the default view is showing all colleges with each dot color-coded based on its tier so that users get an overview of the distribution first. When users click on each dot, the same college tier cluster will be highlighted, and the annotative information of a specific college is updated on the right. Users could also choose to search for a particular college that they care about. Instead of showing the result upfront, we invite users to draw their guesses on the plot to encourage active thinking. They could then click on the “show actual” button to reveal the actual position and compare it with theirs. Through this approach, users could develop a deeper understanding of the problem space and better internalize the take-aways.

For the square matrix, the intention is to visualize the multi-layer aspects of the mobility rate metric - access rate and success rate, which were discussed in detail in the method section. The data source computes mobility rate by multiplying access rate and success rate. We found it unintuitive for people without any background knowledge. Thus we chose to visualize it on a scale of 100 people abstracted by square shapes. We color-coded squares medium-blue for the number of students who come from a poor family (access rate), and deep-blue for the number of students who become a rich adult among the previous group (success rate). In this way, we managed to reveal both layers without getting into the jargon. For example, a school might have a high access rate but a low success rate, leading to an overall low mobility rate, and vice versa.

5. DISCUSSION

The maps can help the audience get a more concrete insight into the geological distribution of social mobility and the linkage between parent income and race. It is not just a list of colleges and the corresponding values of social mobility. It turns out to be very vivid and clear for the audience to understand the regional impact as well as the racial distribution. From the color scale on regional distribution, it is found that the states with high social mobility are located near the coast. As for middle states, the average social mobility is lower. In addition, the audience can see the top-ranking colleges in that state as well as the value of social mobility in that specific school. If the audience is interested in seeing the information of a specific college but not found in the top 10 college list, there is a search box in the next section for the audience to engage with.

Combining the first map showing social mobility with the second one showing the parents’ income, there is a positive relationship between these two factors. With higher average social mobility, the parental income also displays more blueness in the neighborhood map. As for states with lower average social mobility, the regions also mapped into the more redness area in the neighborhood map.

For the last section of the scatter plot and square matrix, we think the linking, drawing, and tooltips interaction is effective in communicating correlations and encouraging exploratory acts. Higher family income is generally positively correlated with higher student income. The college tier shows a clustering effect on the plot. Schools that are more selective cluster around the upper right corner (high family income, high student income). At the same time, it is easy to identify outliers in the pattern, such as medical schools.

6. FUTURE WORK

One area for further refinement is to extend the application of actively engaging the audience to speculate a data point before revealing it in the maps. In addition, it is possible to enrich the story by including more linked graphs on information such as the student gender ratio and race among colleges. There are a number of issues to further investigate, such as whether there are demographic patterns in different college tiers associated with student income and mobility rate.

REFERENCES

"Economic Diversity and Student Outcomes at America's Colleges and Universities: Find Your College," *The New York Times*, 18-Jan-2017. [Online]. Available: https://www.nytimes.com/interactive/projects/college-mobility. [Accessed: 19-May-2021].

K. Zhou, "Changing Class: Are Colleges Helping Americans Move Up?," *APR Reports*, Minnesota Public Radio, 20-Aug-2018.

R. Chetty, J. N. Friedman, E. Saez, N. Turner, and D. Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States*," *The Quarterly Journal of Economics*, vol. 135, no. 3, pp. 1567–1633, 2020.