INFO 3300 / CS 3300
Project 2: Education and Expenditures

Adrian Lee, Faris Aziz, Reza Madhavan
awl68, fsa22, rm855

DATA

Our base dataset used was the Unification Project dataset from Kaggle. This dataset contains information about education expenditures, test scores, and enrollment data for all states spanning years from the 80's to present day. We further merged data for population counts from census data, and GDP data from the US Bureau of Economic Analysis. The geography file for rendering the map in the selection screen was obtained from the topojson repository here.

We then formed a .csv file where each row is a record for a state in a given year. We cleaned and merged data from our datasets in python with pandas and NumPy. Our criteria for acceptable records was that it must contain a valid year and state name, enrollment count, GDP and population count, and at least one standardized test score. Due to the large amounts of missing data, we filtered out unusable records and found that the range 2003-2015 was the only acceptable range with complete data points. We determined this to be suitable for the visualization.

Originally, we wanted to dive deeper into the race breakdowns for educational data to explore race and its relationship with education in the US. However, the data was not complete enough to properly conduct a visualization of this theme.

Having formed our final.csv file, we import the data into our app along with the shape file, and construct our StateDict object (defined in construct.js). This is a custom data structure that has getters and a constructor that makes loading in the .csv and getting data much cleaner.

OVERALL DESIGN RATIONALE

We decided to theme our entire visualization so that it fits a professional educational aesthetic. Since education is generally associated with intelligence and peacefulness, we decided to make all of our color schemes for our visualizations, and some of our buttons and popups fit a blue gradient that consists of light sky blue to dark navy blue. We decided to complement this color scheme with a black background and white outlines to further develop the contrast between our different visualizations.

For our chosen visualizations, we wanted to use our data about education and expenditures and learn more about the differences between states over time. Therefore we wanted to use a map to display the amount of educational expenditures in the most recent year in our dataset (2015) and to allow this map to give us the ability to learn more about the states with information regarding education and expenditure. Therefore, we have two analyses with analysis on a single state and a comparative analysis that allows us to see the differences in our data between the different states. For our analyses, we decided to share the same themes

within our various modules as well as our state selection module to provide a sense of continuity.

For our state selection visualization, we used data from the most recent year in our dataset, 2015, and produced a heatmap of the total expenditures per student across states. For our scale, we used a quantile scale with 4 bins to help see the different categories of states with low to high spending per student. We also used a white outline to help differentiate the different states for ease of state selection when moving towards the other analyses. We also included a label box that contains more information about the state when the state is hovered over. Because there are many significant elements on this screen (mode switch buttons, legend, label box, interactive map), we opted to limit the depth of the metrics shown on this screen, and not get too deep into the nitty gritty. The purpose of this screen is therefore to provide just enough information to the user for them to make an informed decision on where to explore next, at the cost of depth.

For our single state analysis, we wanted to create a screen that dives deeper into the state's management of education expenditures and achievement over time. We decided to use a multi-axis line graph to better show the relationships between metrics, with a slight cost to the immediate legibility of the graph. We use color to clearly mark the differences between metrics, and use a dashed red marker line to show the current year. On the right column, we use text metrics to give a quick high level description of key metrics for that year, and include a mosaic plot to show the education spending breakdown for that year.

For our comparative state analysis visualizations, we made mosaic plots for seven preselected categories to help compare how states weigh against each other for a given year. This analysis was chosen because it would be very clear to the users to see how much better a specific value is compared to another state, as we could relate it to 50%, which would be the comparative standard if both states were equal. The actual values are added to the sides of the plots for more specific information related to the comparison.

INTERACTIVE DESIGN RATIONALE

We have two main interactive elements: the map selection and the date slider.

Map Selection

We decided that a map selection is the most intuitive way for users to choose states, because it affords us the opportunity to simultaneously show the heatmap, which informs the selection process. We added hover styling and the label box to make the selection even more clear, and having multiple modes with clear active styling on the buttons, along with a dynamically changing prompt to action, makes it clear to the user what they are supposed to do with the map.

Date Slider

We decided to use a date slider as the primary method of exploring time varying changes in the single analysis and comparison analysis screens, because it is intuitive and does not conflict with the many other metrics we are trying to show. Understanding that time sliders can make it difficult to view trends over time, in the single analysis we made our line graph with the time on the x axis, such that temporal trends could still be viewed with the slider active. However, it is in the comparison screen that the slider becomes much more useful, as the focus is very much on the shifting of balances between the two states.

STORY

For our visualization, we wanted to see if there was a relationship between educational expenditures and academic achievement on a state level, and explore other discrepancies between states across the nation. Beginning at the heat map selection, we see that many states in the northeast spend much more on education per student. This can be for many reasons, but the GDP per capita metric tells us that it is simply access to greater wealth, and therefore greater state revenues, that drives more spending.

In the single state analysis, we see that the story varies very much by state. Some states see what seems to be a solid trend between education expenditures and academic performance in terms of standardized test scores. Some however, do not exhibit any such trend. In many states, however, a combination of metrics tells certain stories. For instance, there is an obvious drop in GDP per capita from 2007 to 2008 as per the financial crisis, and this in many states correlates to a drop in test scores. We also see that across most states, there is a general increasing trend in test scores over time, though of course it is subject to deviations. Furthermore, the increase in test scores over time is in the range of 1-2 percentage points, and does not exhibit any drastic change.

In the comparison analysis, we leave it to the user to explore states of interest to them. Here, again we see conflicting trends. In some states, a wealthier state with higher per student expenditures does indeed have higher test scores on average. However, when comparing other pairs, for example, New York and Kansas, we see that despite a large disparity in GDP per capita and expenditures per student, the average test scores are mostly equal.

In general, these findings tell us that our initial intuition of a positive relationship between expenditures and test scores isn't necessarily true - and that achievement is a much more nuanced metric that can't be easily predicted with expenditures, nor measured by one standardized test metric. And therefore the purpose of this visualization is therefore not to imply a broad, sweeping trend in a complex field, but to make the underlying data more accessible, such that the user can find smaller points of interest more easily.

WORK BREAKDOWN

Group / Pair Programming:

- Brainstorm and research (2 hours)
- Preprocessing .py scripts (3)
- Overall design iteration and visualization choices (1)

Adrian

- Color/ theme/ styling (2)
- Map selection heatmap and app structure / mode switching implementation (3)
- Client side data structure design and implementation (1)

Faris

- Mosaic plot infrastructure for the comparative analysis (3)
- Comparative analysis styling (2)
- Pandas data overview and splitting (2)

Reza

- Single state view styling (1)
- Mosaic svg on education expenditure breakdown (2)
- Line graph svg on gdp per capita vs. education expenditure per student vs. average student score (3)