

Process Book

Lisa Yao and Morgan Paull

CS171

Project 3

New Project Topic

For our project III we are interested in exploring what influences the quality of education outcomes in California public primary and secondary schools. We have gathered county-level data on educational outcome in California schools in the form of test scores and dropout rates, as well as a predefined aggregate identifier of educational outcome defined by the state (academic performance index or API). We will correlate these data with other variables such as average income, truancy and discipline rates, property tax values, and teaching staff education levels, and seek to reveal important patterns in the data on educational outcome in California.

Data + Sources

We have obtained data for each of the 58 counties in California for:

- STAR test scores
- Academic Performance Index
- Graduation and Drop-out rate
- Truancy rates
- Income per capita
- Property tax rates (directly funds education)
- Staff education level
- Number of students in each county

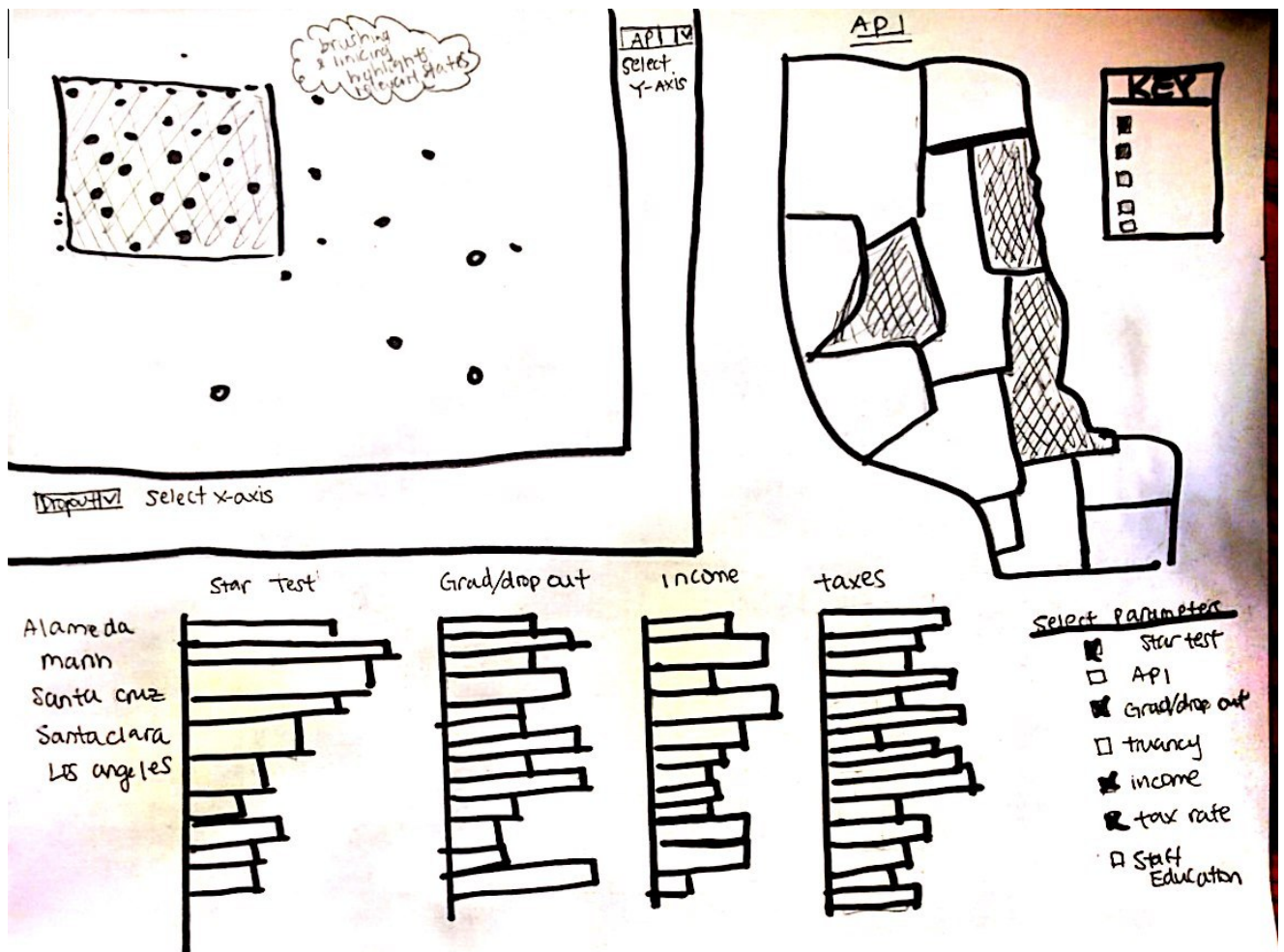
Some raw data is included in the submission (that we have saved in excel and not in a google doc)!!

Our scraper script is also included.

Visualization Plans

We plan to visualize this data on a variety of different dimensions (more details below the sketch).

- Scatterplot with brushing and linking to the map
 - Highlight relevant states
- Choropleth map showing API in gradient color
- Set of about 4 bar graphs that can be used to compare about 4 parameters across all counties
 - Checkbox menu to select parameter
- Chart to show ALL numbers for a particular county
 - Select county using a drop-down menu



County	Star-test	API	Grad rate	truancy	income	tax rate	Staff
Alameda	✓ 1500	304	92%	58%	58k	10%	146

Introduction

Purpose:

Defn of variables

Techniques/features and justifications

The county-level map of California will be a choropleth encoding API (academic performance index) by county. Counties will highlight when clicked and when linked objects are selected in other views.

The scatterplot will be an important element of data display for our project, allowing easy visual interpretation of the relationship between two variables. To allow the user control over the display the variables associated with the X and Y axes of the scatter plot will be mutable, and selectable by the user. The size of the circles will be associated to the number of active students in the associated county.

The bar charts allow the user to compare about 4 variables across all of the counties. The magnitude of the variable will be reflected in the bar. The 4 variables can be selected by the user by a checkbox on the right.

All elements of the display will be linked with highlighting and brushing, with county as the handle used for linking. So the user can brush a cluster of dots (representing counties) on the scatterplot, and the corresponding counties will be highlighted in the bar charts and on the map. Conversely, counties on the map may be selected, and those corresponding dots will be highlighted in the scatterplot and the bar charts.

The table will be displayed below the bar charts, and will act as a somewhat more secondary element of the visualization, to allow the user to see the actual numerical values of any information desired. When no elements of the display are selected, the entire dataset of county-level data is displayed. When some subset of the counties are selected, only those rows of the table devoted to those selected counties is displayed, to allow the user to more easily access and examine the specific quantitative values associated with that row.

4/17/13 - Scrape, sketch, brainstorm

We first decided on what data we would collect (listed above). Then we proceeded to collect all of our data from our online sources. We used a scraper to obtain data from the California Department of Education Website. We directly downloaded economic data such as income per capita and property tax rates from the Board of Equalization website and Wikipedia.

We also sketched our proposed visualization (detailed above) and searched for potential source codes, such as a d3 map of California.

4/16/13 - Meeting with Chelsea

We met with Chelsea to discuss our past project and plans for moving forward with our new project. We first explored the possibility of moving forward with our existing force-directed network project and adapting it to data scraped from PubMed. However, we decided that we could expand both the scope and storytelling aspects of our project if we pursued a new dataset and visualization techniques. Thus, we decided to start over with our project and visualize data on education in California by county.

Chelsea provided us with examples of good visualizations with brushing and linking, storytelling, and complexity.

Our main focus for Project 3 is to integrate *brushing and linking*, add *depth and complexity*, improve *storytelling*, and acquire *more data*.

Timeline:

Week of April 20th

- Solidify data set and clean
- Explore data in ManyEyes

Week of April 27th

- Implement basic features
- Add brushing and linking
- Refine formatting and visual aesthetics
 - Coloring, alignment

Week of May 3rd

- Resolve any bugs
- Add additional features
- Finish process book
- Ask for feedback from test users