Tom Curran

Data Visualization

University of Chicago

Professor Benjamin Soltoff

Final Project Analysis and Write up

Over the past several years Oklahoma has been consistently ranked as one of the lowest achieving states in terms of education. The lack of achievement is often recognized by Oklahoma residents, and even more so by education professionals. As such, there have been a groundswell of reform efforts throughout the state, specifically within Oklahoma City, the state’s largest school district. Efforts such as charter school expansion to curriculum implementation have been attempted over the past several years. Though these efforts are enacted with genuine desire to improve educational outcomes in the district, they operate on minimal information. Critical data points such as a school’s A-F Grade (the State’s evaluation of a school’s performance), student achievement on tested subject, student body demographics, and socio-economic factors are often hard to find or disorganized (if available at all).

The Oklahoma City School Data Explorer seeks to centralize critical pieces of information so that groups ranging from parents of students, to non-profit groups and even policy makers can have a more precise understanding of the different variables that makeup Oklahoma City Public Schools. It should be noted that the data included in this dashboard are only for Oklahoma Public Schools and the district’s application schools, charter school and private schools have been left out due to lack of reportable data.

The Oklahoma City School Data Explorer targets policy makers and non-profit organizations for their audience. Both non-profit organizations and policy makers need information about schools and their students in order for non-profits to provide targeted services and for policy makers to make data driven policy decisions. As such, the dashboard is constructed in a very intentional manner by starting at a general overview of the district and with each additional page users can dig deeper into information about the district. The different levels that each part of the dashboard provide are intentionally designed for the user to be able to explore their own questions instead of having a research question asked by the author. Furthermore, the dashboard aims to use Alberto Cairo’s principles of data visualization in order to attract the user and have them enjoy their experience while simultaneously providing a high level of practicality. I built the dashboard(s) using the flexdashboard framework, leaflet API, ggplot, and Plotly (and several other libraries).

The data explorer is divided into three different sections; “District Map”. “Oklahoma City”, and “School Explorer”. The first section, “District Map”, uses the leaflet api with shiny to map the address of each school in the district. Oklahoma City is very much divided along geographic landmarks that act as neighborhood borders. As such, mapping the districts location and giving each school a color corresponding to the most recent (2016) letter grade, gives the user a sense of where OKCPS schools are struggling. A resident of Oklahoma City would understand the geography of the area and be able to draw some assumption or infer questions about why schools may be succeeding or failing. Furthermore, the map has an overlay of the district boundary.

The second tab, “Oklahoma City”, provides a summary view of the district by showing the number of schools by letter grade (A through F) since the evaluation system began in 2013, as well as a bar chart that displays the frequency of each letter grade in the specified year. Furthermore, the dashboard provides a list of schools based on the filters on the sidebar. This was done to allow the user to compile lists for their own needs as well as be given a general sense of how the district has been performing over time

The third tab, “School Explorer”, is given to the user as a means for digging deeper into individual school’s performance. The page provides seven critical fields. The first four are value boxes: student teacher ratio, number of enrolled students (for specified year), A-F grade (for the year specified by the user), and the grades served. These value boxes are meant to provide the user with a quick summary of information. The A-F grade is color coded to provide a sense of urgency for the school (i.e. if the school as an “F” letter grade than the box turns red). Below the value boxes are 4 graphs that show student demographics for a given year, over time, Free and Reduced Lunch overtime and test performance in a given year. These items are critical figures when creating policy as well as for non-profit groups. These graphs also intentionally highlight the demographic differences between schools. Also, it dispels some popular myths about classroom size in Oklahoma City Public Schools. It is widely believed that most schools have thirty or more kids in each class, that classrooms are often overcrowded. While this is certainly the case in some schools, the data shows that the most schools fall into the 12 to 16 students per teacher range. I believe that it is important to dispel popular notions like this in order to effectively allocate funds where school overcrowding is an issue instead of attempting to create a panacea or “one size fits all” funding policy. Particularly for the “failing” schools, having accurate information allows for more efficient and effective policies that can provide a greater impact. .

Given that Oklahoma City is very much a “neighborhood” city, it is necessary to recognize the demographics for each school, as well as how those demographics are changing. Furthermore, Free and Reduced lunch (expressed as percent of students at the school) is often used in policy making decisions when determining school funding, as well as an indicator of poverty in the school’s neighborhood. The last panel is test scores. The test scores are filtered by grade and by subject, and depict the percentage of students scoring proficient, and of those scoring proficient, the percent of students that scored advanced. Like the Free and Reduced Lunch metric, test scores often are used to determine education policies such as funding but are also an indicator of school performance as to why a school might have received the letter grade in the value box at the top of the page. These graphs were chosen to intentionally allow the reader to explore the relationship between these variables. Several of the plots were made interactive using the Plotly package to allow users to focus on specific areas and timeframes. For the test scores, I used a drop down as to reduce the noise on the graph as well as allow the user to focus on a specific subject and grade to prevent confusion and misinterpretation.

The last two tabs of the dashboard are the District’s revenues and expenditures. These two tabs are perhaps, at least in my opinion, the most important for this dashboard. As mentioned earlier, data on districts can be very hard to wrangle and use, the best of example of this is a district’s financial data. On the last two tabs of the dashboard I have included five components that attempt to circumvent the arduous process of data wrangling. On the first tab I create four value boxes for the total expenditure, and the top three most hotly debated expenditure codes; Instruction, Curriculum and Salary. The value boxes render show the total of all expenditure in those fields. The value boxes correspond with the given year chosen on the side bar. This choice was made in order for individuals to research historical financial records as well as simultaneously providing an “at a glance” look through the value boxes. Below the value boxes is a data table (taken from the DT library) showing a detailed breakdown of line item expenditures. I believed that this was important for the user to see in order to easily collect information. I furthered the utility of the table by making it functional with orderable columns and several options for downloading the aggregated information. The rendered data table contain buttons for copying or downloading the data in several formats (pdf, csv, excel). Given the volume of the financial records, it proved challenging to have a user-friendly way to allow for end-user to easily and quickly manipulate data in a way for exploratory purposes. Instead, I opted for the download buttons so that users could download and explore the data through a familiar tool (i.e. excel) to help drive further understand and increase the utility of the data explorer.

I used the same tools and approaches for the revenue tab. However, I aggregated several key pieces of revenue data that, like the value boxes of the expenditure page, are hotly debated in the education landscape. The four value boxes aggregate the total revenue of the district, the Ad Valorem tax revenue, Gross Production and motor vehicle tax revenues. The Ad Valorem revenue is the property tax that the district receives from the county. I decided to include the ad valorem tax as it has long been at the center of many education policy debates. The growing divide between schools in low income neighborhoods and high income neighborhoods, particularly when it comes to funding, has been attributed to this tax. The gross production and motor vehicle taxes are unique in Oklahoma and the state because so much of Oklahoma’s economy relies of oil and gas production. The Gross Production and Motor Vehicle Taxes are generated in part by the amount of oil and natural gas produced within the boundaries of the district. In many cases throughout the state school districts have been able to increase educational spending due to the oil boom of late, but with the recent decline in prices has presented a significant policy issue.

Ultimately, the Expenditures and Revenues tabs are meant to inform and provide those debating and researching the topic a vernacular. Often, different groups use different sources. With the Oklahoma City Schools Data Explorer, groups can use a common data resource while still being able to develop intendent models. Ideally, allowing for independent research using a common data set will not only enlightened new policy but more collaboration between groups, parties and organizations.

Based on the first tab, “District Map”, there are very few “A” level schools in Oklahoma City and those that are successful schools are not relegated to a geographic area within the district boundaries. That sentiment also extends to “B” level schools. Of the 80 plus schools in Oklahoma City Public Schools, a majority are receiving an “F” for their performance, with a higher concentration of failing schools in the north east and southern section of the district.

The “Oklahoma City” summary of the district performance affirms the intuition made on the map. Most schools in Oklahoma City are receiving a failing grade. What’s more is that the amount of F schools appears to be increasing since the initial implementation of the A-F evaluation scheme was implemented in 2013, the first graph shows an increase from 35 to 40 F schools between 2013 and 2016. Finally, on the third tab, “School Explorer”, when comparing an “F” school, for instance Capitol Hill High School, we see that a very large percentage of the student population qualify for “Free and Reduced Lunch”, an indicator that the areas that Capitol Hill serves are high poverty areas. Comparing Capitol Hill to a school such as Classen High school of Advanced Studies, which scored an A+ in 2016, shows a (relatively) low number of students that qualify for Free and Reduced Lunch status, suggesting that Classen does not serve as many economically disadvantaged students. This comparison is further complicated by the demographic makeup of both schools. Capitol Hill High School has a majority of Hispanic students, representing over 1000 individual students, while Classen High School of Advanced Studies is a majority of white students. Comparing these attributes are important because the imply a relationship between demographics, economic status, and the School’s Performance beyond as a simple measurement of test scores. The visualization reveal that socio-economic makeup of a school can be a heavy influencer of the school’s performance outside of tradition metrics such as amount of funding or as a function of test scores alone.

My design approach borrowed heavily from the Tufte school of thought in that I embraced the idea of minimalism to maximize the impact of the visualizations. The “Oklahoma City” and “School Explorer” tabs are the sections that embrace Tufte’s design philosophy the most. In the top two and bottom right graphs, I intentionally opted for the minimalist background on the graphs with also leaving the major gridlines to guide the user when observing the data. The minimalist components of the graph were also intentionally chosen for the color lines and points to stand out more. Particularly within the top two graphs of the Oklahoma City tab each letter grade is assigned a common color. The colors were chosen in order to make transitioning from graph to graph easier and establish a theme throughout the dashboard pages. Furthermore, the colors were chosen as to be associated with urgency. An “F” letter grade is red and “A” letter grade is green, while the letters between C, D and B were chosen to represent colors departing from an urgently red “F’ to an A letter grade, while still being able to stand apart from one another. This same design principal applies to the “School Explorer” tab where different demographic groups are consistently represented by the same colors, allowing the viewer to go from graph to graph with ease and circumventing the need for readjusting their frame of reference thereby making the ingestion of data easier.

In terms of appearance, the last two tabs, Expenditures and Revenues, are intended to embrace Tufte’s minimalism, but do so to less of an extent that the school explorer or the district overview. In my first draft of the visualization I sought to include more graphs in addition to the table and the value box, but found that with that much information on one page there was not enough space to accurately convey information to the user. As a result, I removed the graphs and kept the data table and value boxes. Ultimately, this is the better design decision despite it not fully embracing minimalistic aesthetics. I used color in the value boxes in order to attract the user’s eye to critical pieces of information. Unlike the “Oklahoma City” tab, I did not want to imply a sense of urgency but rather simply gain the reader’s attention. As such, I chose a bright yet toned down color pallet.

One of the most important tenants of Cairo’s design philosophy that I wanted the visualization to embrace was the notion of utility. In working in education, I found that tools that were not easily useable were quickly abandoned. I faced the challenge of making the visualization detailed while being useable and not overloaded with information. As one goes through the dashboard tabs you can see an increasing level of usability and customizability of the data. The map of Oklahoma City Public Schools has a low level of customizability as it is a simple map. The “Oklahoma City” table deepens the tool’s usability as it allows the user to drill down into the year and school level. Next, the “School Explorer” tab allows for the user to drill down into a specific school, year, grade and tested subject. Both tabs also employ the use of Plotly to enable the user to delve deeper into the graph through interaction. The use of Plotly comes in especially useful when attempting to employ Tufte’s minimalistic approach as hovering over data points can express the data dynamically, where as traditional static images may have created a lot of noise and skewed the interpretation of the graph. Finally, the Revenues and Expenditures tab give those that are seeking finite control of data the opportunity to explore the data by being able to export data sets. Ultimately, I believe the usability of this visualization is its best asset as few other data resource are able to cater to such a wide range of users.

Though the flexdashboard package provides a easy to manage framework and easily incorporates ggplot, leaflet, shiny and more, I found there are still several features missing. Particularly in the customizability of the layouts, I found it less advantageous than a traditional shiny application. Furthermore, there were some features of shiny that I found were difficult to implement with flexdashboard’s framework. Particularly with the “updateSelectInput” and “validate” functions, the implementation were straightforward in a traditional shiny runtime but proved confusing and difficult to do within a flexdashboard. This proved to be an important challenge as many of the selectInput options in my dashboard would benefit from being dynamically updated. Additionally, using the validate function that comes with shiny avoids the glaring red text when errors are thrown inside of the dashboard. I found that having such errors thrown not only looked unappealing but found that they left users confused.

The dashboard not only centralizes reporting information for Oklahoma City, but also allows for identifying possible relationships between variables. Visualizing this data can lead to asking important questions about how poverty, race and other socio-economic factors influence student performance. Furthermore, the data explorer is not targeted to solely identify “failing” schools but also to identify those schools that are being successful. Identifying the successes in the district can help influence the outcomes of different schools be recognizing and adapting best practices. From a policy lens, the dashboard can help summarize important findings that may simply be recognized but not quantified. For instance, it seems to be common knowledge that schools on Oklahoma City’s South and North East side are failing schools, but it is rarely quantified as the dashboard does. Furthermore, the dashboard proves that just because a school is “failing” does not mean it is relegated to the South or North East side of the city, when in fact there are several failing schools in the north and western parts of the district. Ultimately, the dashboards hopes to dispel misgivings or bias by quantifying and allowing users to explore the data for themselves. The functionality of the visualization aims to allow for the user to conduct some exploration, but ultimately aims to generate meaningful questions that drive important discussion in the education community.

Ultimately, I hope that users find the data explorer dashboards informative about the current state of education in Oklahoma. Given the highly sensitive social and political nature of education, it is easy for information to be lost or mistranslated. As such, creating the data explorer can circumvent any information loss by employing the utility of flexdashboard’s and embracing Tufte’s minimalism to drive visualization impact. The impact of these visualization emphasizes Cairo’s notion of truthfulness in both purpose and design principles, choosing aesthetics that embrace the spirit of improving education over favoring one group of students or a school. I believe that the data explorer to be very enlightening. It allows for people to see that issues facing the education community in Oklahoma are not happening in one neighborhood or in one part of the city, but help policy makers and non-profits alike realize that everyone in the greater Oklahoma City community is effected by issues in education.