|  |
| --- |
|  |
| Capstone Spring 2016 Project |
| Nicolas Aguirre |
| EduViz |

Nicolas Aguirre

Spring 2016

## Summary

**eduViz** is a visualization project that aims to make sense of the recently-released [College Scorecard Data](https://collegescorecard.ed.gov/data/). The dataset, provided by the Department of Education, contains information on higher-education institutions. Some of the metrics provided are admission rates, student completion rates, debt and repayment, earnings, demographics and more. The dataset contains nearly twenty years of data, with about 1700 variables tracked. Most of this visualization will focus on data from 2013.

## Project Scope

The primary goal of this project is to **inform** readers about the predatory practice of exploiting poor students by higher-education entities.

Last semester, I visited *ProPublica* in New York to meet with some of their reporters. In the process, I became more acquainted with the dataset and some of the stories that could emerge from a compelling visualization.

While there are extensive datasets with relevant variables, this project will be unique because:

* It highlights variables of importance – With over 1700 columns of variables, this is an absolute necessity (particularly for those interested who have no statistics acumen.)
* It visualizes, in an intuitive way, the magnitude of select variables. This is intended to help readers understand how a given school compares to others

## Objectives

The main goal of this visualization story is to serve as a tool that:

* Shows a general trend of schools exploiting impoverished students
* Reveal specific institutions as culprits of this practice
* Allows students, parents, and professors to compare and contrast which schools are more affordable and accessible

To convey a successful idea of this visualization, the website should be:

* **Intuitive** (needs little explanation) and **accessible** (understood by most people)
* **Usable** – Able to be navigated and operated without error or frustration
* **Compelling** – A story/narrative should be evident from the visualized data (although it may need to be augmented with written elements)
* **Flexible** – Versatile in its potential uses and ability to relay information

The main goal is to take an immense and somewhat cryptic dataset and transform it into a visualization, or series of visualizations that is sensible and more readily available to the general public; in its current state, the data caters mostly to statisticians and researchers, as well as though who can manipulate data.

## Audience

The intended client, to whom the final deliverable will be given, is the reporting staff at *ProPublica*. For distribution:

* Annie Waldman – A reporter covering education (<http://www.propublica.org/site/author/annie_waldman>)
* Scott Klein – Team director, whose team focuses on data-driven interactive stories

(<https://www.propublica.org/site/author/scott_klein>)

* Potential education stakeholders – Students seeking to further education, as well as professors and others affected by performance of universities
* Members of education-driven entities such as *Kaplan, CollegeBoard*, and *ACT*
* Persons of power in universities – Those who direct financial resources or influence the school’s approach in providing affordable education
* General Public – This app will exist firstly as a story-telling mechanism, with the primary purpose to inform.

## Marketplace and Barriers

While this project doesn’t strive to make a profit, there are several forces that may negatively impact the project’s success:

* Propaganda and advertising of for-profit institutions has the ability to obscure the story’s message.
* Skepticism of readers – this can be mitigated with the availability of source data and a comprehensive description of methodology.
* Lack of comprehension – If the visualization is too robust, complicated, or requires special knowledge; it will limit the number of people who understand it. The project aims to be highly intuitive and accessible.

## Design

* Intuitive and usable interface – The tool should be easy to use for almost everyone
* Aesthetic and minimal – Visual design will be practical, simple, and aesthetic elements will serve a functional purpose (e.g., color of data points as a representation of ordinal category)
* User’s experience with the site should be unobtrusive, satisfying and memorable. One’s reading and interaction with the app will make for a lasting impression

## Deliverable

The website will be hosted on my domain (<http://cyni.co>), and its code will exist as a private repository on GitHub. The deliverable will be completely transparent, making all source code and data visible to recipient. A current mockup view can be seen here at <http://cyni.co/edu-viz/>

While the website will exist as a **single-page application**, “views” can be enabled by the user. The final product will have these features:

* A scatterplot, that encodes an X variable, a Y variable, and up to three additional variables (radius, fill, opacity). The scatterplot is intended to show interestingly correlated variables. E.g., comparing median first-generation student debt to median Pell grantee debt.
* A detail pane, which will feature a bar chart of relevant metrics for a given school. The idea is to give an overall picture with the scatterplot, and then provide a more granular view with the detail pane. For example, a user might want to know what a “Pell grant” is, how one qualifies for it, and what it means if a given school has high median Pell grantee debt.

If time permits, a geospatial component is desirable. Being able to visualize a region and the access to schooling that it permits could prove interesting. This is a non-critical, but still useful feature.