Nicolas Aguirre

Spring 2016

## Summary

My project is an interactive data visualization, in the form of a website. The visualization focuses on revealing how many universities, especially for-profit institutions, exploit impoverished students. The project is well-substantiated by a massive volume of data. 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.

## Project Scope

The main goal of this web app 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 explore which schools are more affordable and accessible

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 primary goal of this project is to **inform** readers about the predatory practice of exploiting poor students by higher-education entities. This project does not aim to incite any action from readers.

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

* Show data in a way that is **intuitive** (needs little explanation) and **accessible** (understood by most people)
* Be **usable** – Able to be navigated and operated without error or frustration
* Be **compelling** – A story/narrative should be evident from the visualized data (although it may need to be augmented with written elements)
* Be **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*. Particularly:

* 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>)

I have met with both and discussed the subject, and the related data in depth. Others who stand to benefit from the app are:

* Potential education stakeholders – Students, 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
* Lawmakers and politicians – Those with the power to impart lasting change on universities or affect students’ access to higher 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 can negatively impact its success:

* Propaganda and advertising of for-profit institutions has the ability to obscure the app’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

Design objectives for this project:

* Intuitive and usable interface – The tool should be easy to use for almost everyone
* Aesthetic and minimal – Visual design should be practical, simple, and aesthetic elements should 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 should 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 should 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. I intend for the final product to 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.

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.