

# Stat159 Project 3: Add a title

Liang Hao, Bret Hart, Andrew Shibata, Gary Nguyen

December 5, 2016

## 1 Introduction

We are assuming the role of consultants to an NGO which primarily desires to provide at-risk, low-income, minority students a portfolio of colleges which best serve their needs: a high-quality, low-cost education. The interactive final product which allows student input is the shiny tool. Students enter in various criteria (their GPA, location, desired field of study, income bracket, first-generation status, etc) and it outputs a list of colleges ranked according to a score which we arbitrarily, but intelligently create as an ad-hoc indicator of "quality" of school for varying types of students. The responses are sorted upon this composite metric of "quality," and are readjusted in real time if the student changes their inputs. This "college quality score" is our response variable that students see and interact with, and the method the output list of schools are sorted upon.

The actual disparate pieces of the response score are included as a result of a variety of exploratory analyses. At first, we underwent the daunting task of reading through the Data Dictionary given with the College Scorecard data - a detailed account of the content of around 2000 columns. After we became familiar with the general structure and material within the dataset, we moved on to research and exploratory data analysis. While there is some interesting philosophical merit to deliberating upon what makes a school "good," the definition of "good" we adopted allowed for a much more supervised process of creating a response variable - the majority of emphasis was placed on cost, community, and projections of success after college. Thus, we carried out research on relevant statistical literature. By looking through the data dictionary and reading some analytical publications that used the College Scorecard dataset, we began to put together a set of salient quality indicators, such as average salary after college, completion rate, median debt, or net price relative to a student's income bracket. This piece, from page 7 on, was of huge help to creating this performance metric.

We then combine these responses into one "college quality score" for the school, with weights for different variables in the composite metric based upon experimentation and what the literature suggests. Additionally, We use combined data from the last 5 years for our score to reduce variance, with each

year prior given less weight. The interesting specificity of our model is that it is geared not toward creating a score for general school quality and a prediction for the average student, but specifically for helping at-risk students who need assistance in deciding which schools are worth the application in their specific circumstances. These students need far more help than most and we truly want to accomodate them.