

Introduction

In recent years, the higher education industry has been going through some challenges including: a stagnating 18-year-old population, increased competition over enrollment, geographical pressures, and expensive price tag of liberal arts education compared to public institutions [1]. Although most reasons highlighted play a factor in the well-being of higher education institutions, the stagnation of 18-year-old population affects the core business model of it. In extreme cases where enrollment falls, there is an increase in debt due to accumulation of deficits as revenue is less than the ever-increasing costs, leading to a decrease in endowments, bankruptcy, and closure of schools [2]. Thus, our research aims to discern by comparison the resilience of small colleges to big colleges and looks at which factors contribute significantly to the closure of schools.

Definition of Small Colleges

Small colleges, in terms of revenue, are defined as private schools with revenues less than \$100 million or public schools with revenues less than \$200 million per year. In terms of institutional size, private small colleges can be defined as schools with fewer than 5000 students [3].

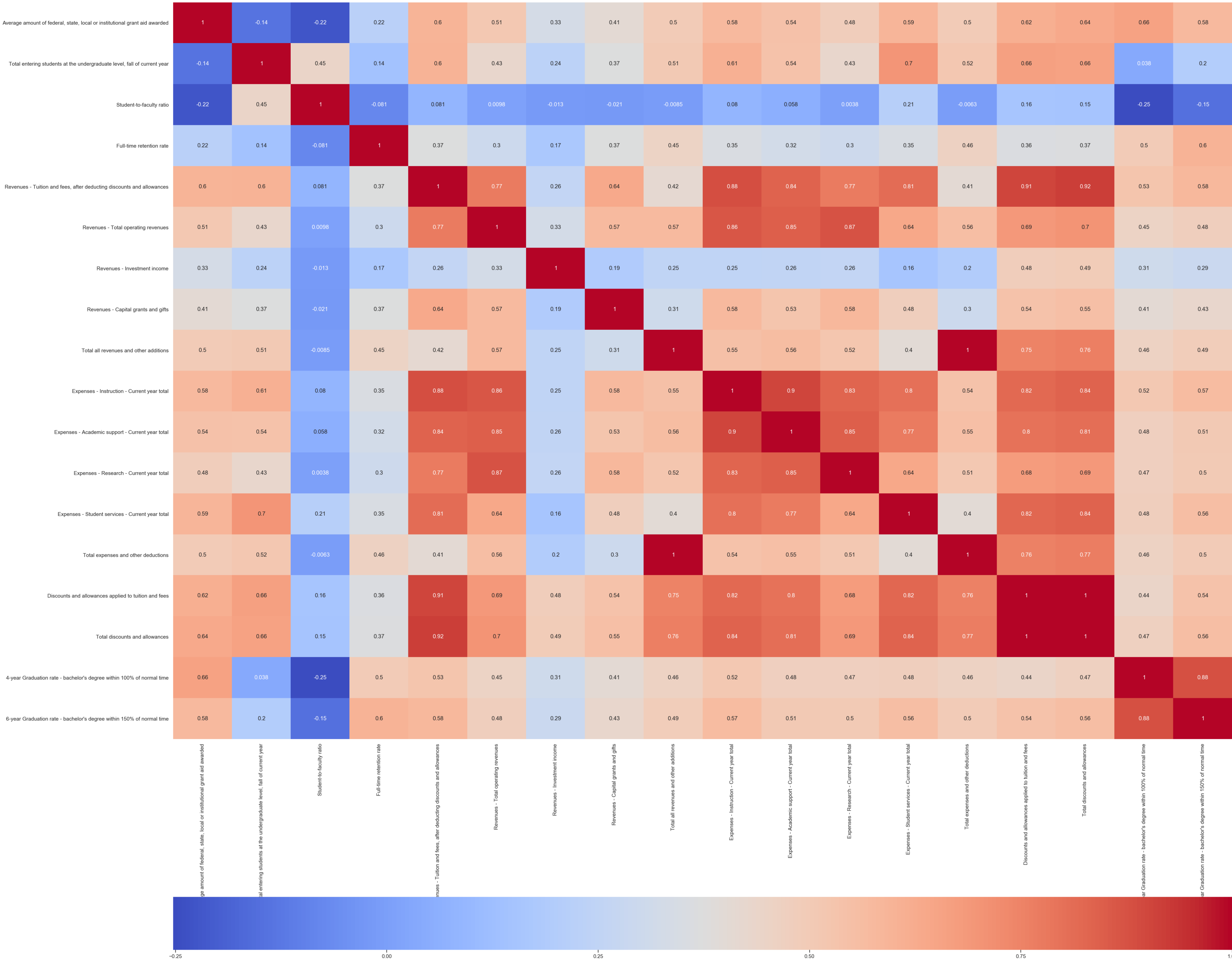
Statistical Methods and Model Design

We collected our data from the Integrated Post-secondary Education Data System (IPEDS) and used data sets from 2010 to 2017 for our statistical analyses. The data sets include information on enrollments, annual revenues and costs, federal financial aid, etc. from higher education institutions [4]. We performed data cleaning, and correlation matrix to determine explanatory variables using Python. We first divided our samples into each year by size: large schools= $\{\text{schools} \geq 5000 \text{ students}\}$ and small schools= $\{\text{schools} < 5000 \text{ students}\}$.

We did this to compare the performance of several indicators between the two independent samples of small and big schools, such as full-time retention rate, and average financial aid awarded using one-tailed 2-sample t-test. Using Python (Numpy, Pandas) and R libraries (dplyr, ggplot2), we discovered some general trends on closed colleges and universities, analyzed statistically our data set with hypothesis testing and formulated regression models that predict a certain feature based on various other features.

To build the multiple linear regression model, we set y = full-time retention rate as our target and selected 6 other variables that were highly correlated (correlation coefficient > 0.4) with y , including the average amount of grant aid awarded, total expenses per total entering students, total revenues per total entering students, total current asset per total entering students, 4-year graduation rate, and 6-year graduation rate. We labeled them x_1, x_2, \dots, x_6 and transformed x_1, x_2, x_3 and x_4 into their logarithmic scale to derive the equation. We then calculated the correlation coefficients of the equation from year 2010 to 2017.

Correlation Matrix



Hypothesis testing

SCI= Small college indicator

LCI= Large College indicator

Variables	2010	2011	2012	2013	2014	2015	2016	2017
Published in-state tuition and fees	8.89e-20	1.81e-20	4.80e-25	2.87e-39	2.82e-31	5.86e-12	1.26e-19	3.48e-11
Average net price-students awarded grant and scholarship aid	2.61e-55	1.97e-60	1.33e-60	1.14e-62	2.09e-56	3.34e-54	1.88e-72	5.35e-61
Percent of full-time first-time undergraduates awarded student loans	2.96e-15	9.05e-28	3.58e-27	4.64e-32	6.24e-31	3.96e-11	7.98e-12	8.50e-6
6-year Graduation rate - bachelor's degree within 150% of normal time	0.001079	1.72e-6	3.42e-10	1.95e-14	1.28e-14	5.72e-16	8.50e-15	2.66e-16
Full-time retention rate	8.41e-7	1.13e-6	5.36e-7	1.8669e-15	2.17e-16	6.30e-16	9.05e-11	2.05e-8
Total amount of grant and scholarship aid awarded, all income levels per entering students	0.0084	0.0874	0.3531	0.4540	0.2402	0.4376	0.3487	0.1682
Total all revenues and other additions per entering students	0.080	0.1740	6.94e-6	4.73e-6	3.2e-5	0.000314	0.000502	0.00026
Total discounts and allowances per entering students	0.0012	0.0050	3.91e-5	2.68e-6	4.05e-6	3.25e-5	9.73e-5	6.19e-6

Table 1. p-values for 2-sample t-test with the alternative hypothesis being the means of each feature is smaller in the small college sample than the big college sample

Regression Models

The general multiple linear regression equation is $y = a_0 + a_1 \log(x_1) + a_2 \log(x_2) + a_3 \log(x_3) + a_4 \log(x_4) + a_5 x_5 + a_6 x_6$ in which a_0, a_1, \dots, a_6 are constant coefficients. In 2017, $(a_0, a_1, \dots, a_6) = (5.6315, 2.8657, -1.8462, 3.8542, -0.2844, -0.2294, 0.6205)$

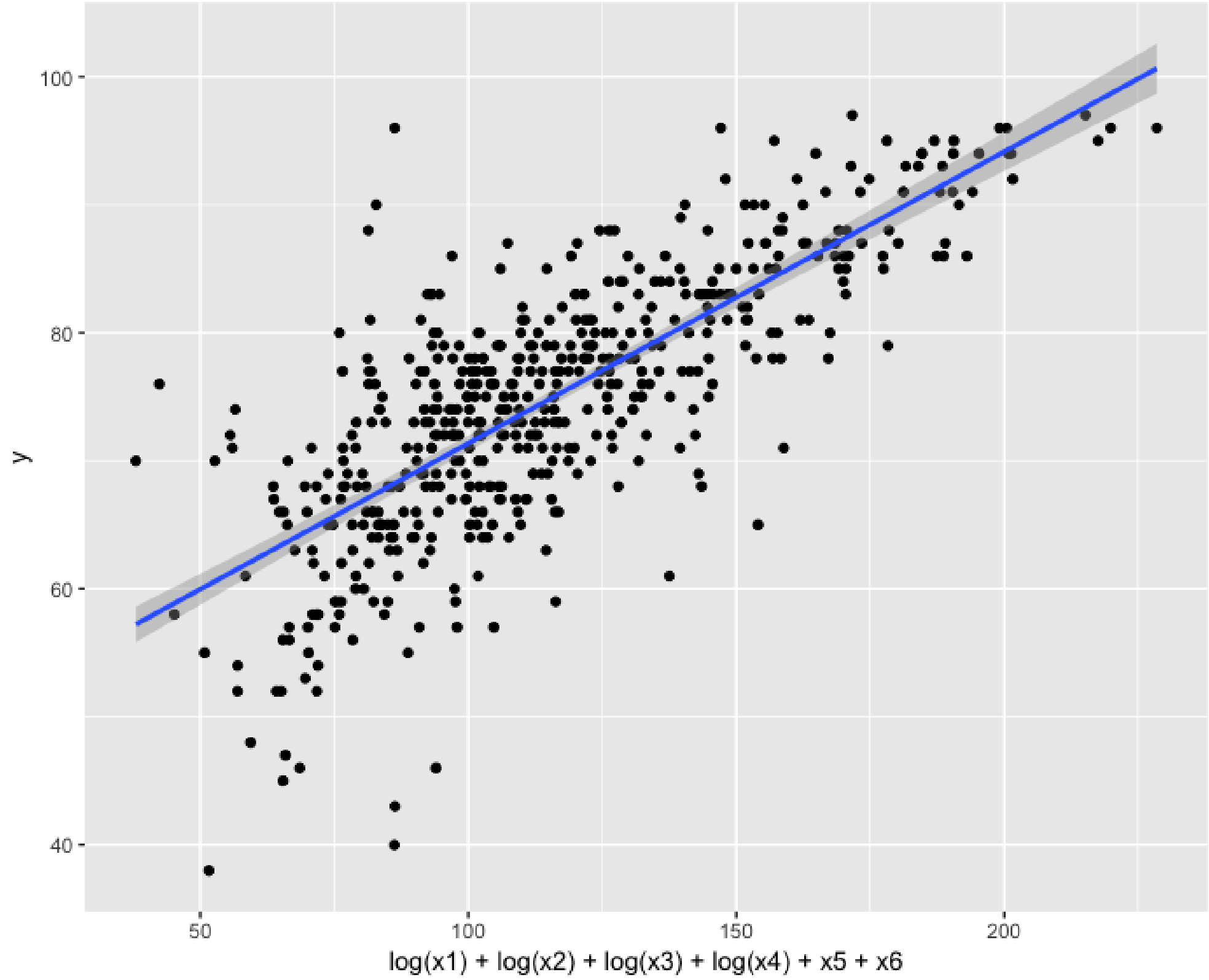


Figure 1. Multiple Linear Regression Model for predicting full-time retention in 2017

Results and Discussion

From the analysis that we carried out we discovered that there is strong relationship existing between dependent variable (retention rate) and our other independent variables grouped together (graduation rate, total revenues, total assets, average federal, state or local grant aid, and total expenses). Over the years 2010 – 2017, the correlation coefficient $0.6431 \leq r^2 \leq 0.6996$. In addition, from the 2-sample t-test, we discovered that the mean of all features except "Total amount of grant and scholarship aid awarded,all income levels per entering students" in the small college sample was significantly smaller than that in the big college sample.

Future Directions

Our paper was limited because of missing data, and we believe with better quality of data there is much more that can be done. The possibilities include: developing a prediction model that determines which schools are in danger of closure, benchmarks that illustrate whether a school is in distress or not.

References

- [1] M. Horn, Will Half Of All Colleges Really Close In The Next Decade?, Forbes, December 13 2018, <https://www.forbes.com/sites/michaelhorn/2018/12/13/will-half-of-all-colleges-really-close-in-the-next-decade/33e5dc3552e5>
- [2] A. Harris, Here's How Higher Education Dies, The Atlantic, June 5, 2018. <https://www.theatlantic.com/education/archive/2018/06/heres-how-higher-education-dies/561995/>
- [3] Small US colleges feeling the squeeze from "soft" tuition revenues, ICEF Monitor, December 18 2015, <http://monitor.icef.com/2015/12/small-us-colleges-feeling-the-squeeze-from-soft-tuition-revenues/>
- [4] About IPEDS, National Center for Education Statistics, <https://nces.ed.gov/ipeds/about-ipeds>