College Cost-Benefit Analysis Using Linear Regression Analysis, Pandas, and Seaborn

Mario Sumali*

La Cueva High School, Albuquerque, NM

ABSTRACT

College is so expensive that student loans make up the second-biggest consumer debt in the nation. Some of the most expensive colleges produce alumni with low earnings. To slow down the growth of the problem, in 2015 the US Department of Education made available to the public the College Scorecard database which lists the costs and average alumni salaries, along with many other variables, of thousands of colleges. This paper features the Python data science tool Pandas and visualization tool Seaborn to reveal correlations among several important variables related to tuition and earnings. For a start, simple sorting of tuition shows that the few most expensive colleges do not produce alumni with the highest earnings. This paper then shows that the highest-earners-producing colleges achieve this to a large extent because those colleges give a large proportion of their degrees in graduate or professional degrees (MD, JD, Ph.D., MS, etc). That conclusion is enabled by a method presented here, based on linear regression analysis, to estimate Bachelor alumni's earnings from the Integrated Postsecondary Education Data System (IPEDS-reported) combined-Bachelor-and-Graduate alumni earnings. Furthermore, with or without adjustment for graduate earnings, highest-earners-producing colleges give a large proportion of their degrees principally in just a few fields of study: 1. Health professions (Bachelor's level, not including MD or Masters); 2. Engineering, including Computer Science; 3. Business. A few colleges produce high-earners in Social Sciences and Maritime Transportation (which were lumped into Engineering). The discovery of those highest income-producing fields from the College Scorecard is largely consistent with another study by Georgetown University which shows that STEM majors give significantly higher earnings than other majors. The study reports average earnings of college majors, which is important for students applying to colleges and may well determine the future of the students.

Major is a finer categorization of profession than fields of study. However, the Georgetown study does not attempt to correlate tuition of colleges to the alumni's earnings. Further analysis of the College Scorecard shows some correlation between the two. Colleges that tend to produce alumni with high earnings also tend to charge high tuition. Furthermore, public colleges tend to charge significantly less tuition than private colleges. However, colleges with high tuition do not necessarily produce alumni with high earnings. Thus, in terms of earnings, the major is far more important than the price tag or the name of the college. Nevertheless, for value comparison, the names of the top 32 colleges in terms of their average Bachelor's earnings (from College Scorecard adjusted for graduate degrees) and their tuition are given here, along with highest-earners-producing colleges in Health Professions and Engineering. College Scorecard reports the size of student debts at each college. The analysis shows that debt does not correlate with alumni earnings.

KEYWORDS: Choosing Colleges, Graduate Salaries, Graduate Boost Factor, University Tuition, Paying for College, College Debt

INTRODUCTION

A college education is very important for obtaining a well-paying job. On average, a Bachelor's degree results in roughly \$1 million in higher lifetime earnings compared to a high school degree (Carnevale et al., 2015). However, college is so expensive that student loans are the number two consumer debt in the nation—\$1.6 trillion in 2020, which is second only to home mortgages (Friedman, 2020). For comparison, far more consumers owe credit card debt than owe college debt, but credit card debt totals just above \$1 trillion for the whole nation. The average student loan debt is \$32,731, and 8.1 million Americans 50 years or older still owe around \$25,000 in student loan debt. (Friedman, 2020). Why do students incur so much debt? The simple answer is that college is expensive. A college that we shall call "College 1647" charged \$51,665 per year in tuition and fees in 2015. In addition to being expensive, that college does not produce alumni with high earnings. Six years after enrollment, the average graduate of College 1647 earns only \$21,093/yr. With that little earning, the alumni cannot pay down their college loan debt quickly. Several colleges that charge less than \$1,000/yr tuition have alumni with higher earnings than College 1647 alumni. Table 1 shows the wide range of average earnings of alumni of just a few colleges. Colleges that charge more in tuition do not necessarily produce high-earning alumni. Thus, today's applicants to colleges must be very careful in choosing which college to attend. The old idea of a "dream college" needs serious adjustment (Seeger, 2018).

College Name	Tuition, \$/yr	Earning, \$/yr
"College 1647"	51,665	21,093
"College 3874"	42,962	41,900
"College 1533"	37,489	18,500
Carnegie Mellon University	36,119	69,800
Worcester Polytechnic Institute	32,881	67,500
Bucknell University	31,788	57,700
Harvard University	43,938	70,300

Table 1. The high costs of colleges do not necessarily result in high earning.

DATA AND METHODS

The author wrote a Python code to read and process College Scorecard data and to present actionable results with lists and graphs for visualization. As a result of the US Department of Education's College Scorecard project, the Integrated Postsecondary Education Data System (*College Scorecard Data*, 2020) website reports hundreds of variables for each of over six thousand colleges in the US. The author downloaded all available databases from that website, dating back to 1996. The latest database, from 2016-2017, is too new to have data for earnings at 10 years after enrollment. Thus, the author uses the database from 2014-2015, the latest year for which the database contains earnings at 10 years after enrollment.

The Python code used the following libraries: **Pandas** is a module for data science. It implements numerous common statistics formulas. Additionally, **Scipy.stats** contains lower-level statistics functions, a few of which is used in the code. **Numpy** is necessary for nearly all significantly mathematical analyses. **Matplotlib** enables the creation of graphs for visualization. Additionally, **Seaborn** is a high-level library of common visualization routines, which runs on Matplotlib under the hood.

The variables used are listed in Appendix A. Most of the variable names follow the standard definitions given by the IPEDS database; they are written in ALL_CAPS. In contrast, the variables created by the author in the Python code are not in ALL_CAPS. Algebraic formulas derived and implemented by the author will be explained in the next section, which describes the research and coding in a logical sequence of data processing steps or conclusions. A few key lines in the code can be found in Appendix B as well, and some will be referred to in the discussions below.

Among over 7,700 colleges in the IPEDS database, the author is interested in applying to colleges that offer Bachelor's degrees. (In the database, HIGHDEG = 'Bachelor'.) This excludes two-year community colleges but includes all four-year undergraduate colleges and colleges that grant graduate (e.g., MS, Ph.D.) or professional (e.g., MD, JD) degrees in addition to Bachelor's degrees. There are almost 2,000 predominantly undergraduate colleges in the US. Most of them also give graduate degrees. In this analysis the author assumes that the most important variable for each college is the median earning of its alumni at six years after they enrolled, representing the early career stage after graduating with a bachelor's degree. In the IPEDS database, this variable is called MD_EARN_WNE_P6. Many colleges eliminated or suppressed the reporting of this earning, showing it as 'PrivacySuppressed'. The author excluded those colleges from the analysis. A few colleges are exclusively graduate/ professional schools in practice (Medical, Law, Business) even though they reported their 'predominant degrees' (PREDDEG) as "Bachelor's". Among 1684 colleges that are listed to give bachelor's degrees, only 1636 list the number of undergraduate students. The rest show no undergraduate population and are listed in Appendix C. Those colleges are actually graduate/professional schools and are therefore excluded from the analysis. The cleaned list from the database (just over 1600 colleges) is placed in the Pandas dataframe **df clean.** The author uses a Pandas command

$$df_clean = df_clean[df_clean['UGDS'].notnull()]$$
(1)

A few other similar steps are taken to exclude colleges that do not grant bachelor's degrees.

ALUMNI EARNING IS A FUNCTION OF GRADUATE STUDENT POPULATION

The most important variable in this analysis is the median earning at six years after graduation (MD_EARN_WNE_P6). Appendix D shows that this variable is distributed almost in a normal Gaussian fashion. A few colleges are on the low-end 'secondary' normal distribution. Those are mostly religious seminaries and colleges in US Territories outside the states. More interesting is the tail at the high-end of earnings, which are colleges that produce very high earners who are outside the main, approximate normal distribution. Colleges that are exclusively graduate or professional schools, e.g. Medical or Law schools, produce alumni with high earnings. Excluding those colleges, Table 2 lists the top few high-est-earners-producing colleges. Numbers in turquoise denote the undergraduate population that is smaller than the graduate population. Examination of Table. 2 and a few dozen more colleges on the top of the list reveals that most of these top highest-earners-producing colleges do. We also observe that Thunderbird School of Global Management produced alumni with earnings at six years after enrollment (MD_EARN_WNE_P6) of \$87,400/yr, which would rank it as the 4th highest-earners-producing "Bachelor's" college in the US. However, that college had 797 graduate or professional students and one undergraduate student. Therefore, we excluded that college from Table 2 although it passed the graduate are-and-professional-schools filter in code line (1) above.

From the observations made with Table 2, we statistically hypothesize that the highest-earners-producing colleges have high graduate student proportions in their populations. To examine this hypothesis, we perform a linear regression analysis of earning as a function of graduate student proportion. Figure 1 shows alumni earnings at six years after enrollment



Figure 1. Colleges whose graduate student ratio is near 100% (i.e. "near-totally graduate or professional" colleges) produce alumni with roughly 60% higher earnings than colleges whose graduate student proportion is near 0 (i.e. "near-totally undergraduate" colleges).

(MD_EARN_WNE_P6) as a function of *graduate student population fraction* at each college, defined in Pandas syntax as

df_clean['Grad_population'] = df_	_clean.GRADS / (df_clean.	.UGDS + df_clean.GRADS) (2)
-----------------------------------	---------------------------	-----------------------------

	College Name	Earning, \$/yr	Undergrads	Grads
1	St. Louis College of Pharmacy	120400	693	668
2	Albany Coll of Pharm&Health Sci	112100	1076	483
3	Samuel Merritt University	100100	523	979
4	Massachusetts Inst of Technology	82200	4476	6807
5	Oregon Health & Sci University	80000	847	2014
6	Louisiana State U Health Sci Ctr	78200	35	835
7	Duke University	76300	6480	9230
8	Mass Coll of Pharma & Health Sci	75700	3944	2908
9	SUNY Downstate Medical Ctr	75300	338	1521
10	Maine Maritime Academy	75200	1031	29
11	Roseman U of Health Sciences	74900	269	1135
12	West Coast U - Los Angeles	74600	1395	264
13	Calif State U Maritime Academy	73100	1047	NaN
14	Loma Linda University	72000	1039	3411
15	University of Pennsylvania	71600	10678	13258
16	Rose-Hulman Inst. of Technology	70700	2208	108
17	Kettering [GM] University	70700	1689	347
18	Babson College	70400	2107	942
19	Stanford University	70400	7018	9944
20	Harvard University	70300	7236	18453
21	U Maryland Baltimore	70000	786	5484
22	Carnegie Mellon University	69800	5819	6699
23	Colorado School of Mines	69200	4383	1506
24	Thomas Jefferson University	69000	705	2715
25	The U of Texas Medical Branch	68800	708	2503

Table 2. Highest-earners-producing colleges give high proportions of graduate degrees. Colors indicate the colleges' prime field of study. Blue = Health Professions; Red = Engineering or Transportation; Green = Business. Turquoise numbers denote the undergraduate population that is smaller than the graduate population.

Table 3 shows the statistical summary from the linear regression analysis. Earning is positively correlated with the graduate student population fraction. The regression line intercepts the MD_EARN_WNE_P6 axis on the left (i.e., Grad_population = 0) at \$31,305/yr. We call that intercept E6_ugds_mean. At the high extreme (Grad_population = 100%), the regression line intercepts the MD_EARN_WNE_P6 axis at \$49,998/yr. So the slope of the line is (\$49,998/yr - \$31,305/yr) / 1 = \$18,693/yr. The above numbers are obtained from the straight regression line among widely scattered data points. The correlation goodness of fit is somewhat low (*R* = 0.33). However, Table 3 is the best linear estimate that correlates alumni earning (MD_EARN_WNE_P6) with the Grads population.

Regression variable	Interpretation	Value	Unit
y at ($x = 0$)	Average earning of Bachelor's	31305	\$/year
<i>y</i> at ($x = 1$)	Average earning of Masters and above	49998	\$/year
Diff. between above	Graduate Earning Advantage	18693	\$/year
y(x=1) / y(x=0)	Graduate Boost Factor (GBF)	1.615	None
R	Goodness of fit	0.33	None

Table 3. Regression line shows that Earning is positively correlated with Graduate Population fraction.

The estimate means that, on average, colleges with 100% graduate students produce alumni with roughly \$19,000 (61%) higher earnings than colleges with 100% undergraduate students. To normalize the impact of graduate degree proportion to the alumni earning analysis, we introduce a concept called *Graduate Boost Factor (GBF)* which is the ratio of a college alumni's earning if the college had 100% graduate students to the college's alumni earning if the college had 0% graduate students:

In this analysis, GBF = 1.615. (For example, if Thunderbird School of Global Management (Grad_population = 100%) had only undergraduate students (Grad_population = 0%), then its MD_EARN_WNE_P6 would be \$87400/1.615 = \$54158.)

Colleges that award a large proportion of graduate degrees report higher alumni earnings, which is misleading to undergraduate program applicants. This article is about choosing colleges that give Bachelor's degrees. Thus, "Bachelor-Equivalent earning" is more important in their decisions than the average earning MD_EARN_WNE_P6, which is reported in the IPEDS database. "Bachelor-Equivalent earnings" are not listed in the database. Thus, we need to estimate "Bachelor-Equivalent Earning" from MD_EARN_WNE_P6, using the following algebra.

Assume that earning MD_EARN_WNE_P6 is an average of Bachelors_Earning and "100% Graduate" earning (E6_GRADS), each weighted by the proportions of Bachelor population fraction and Graduate population fraction:

Also, assume that Graduate earning is equal to Bachelor's earning times graduate boost factor GBF:

From (4) and (5), we calculate the Bachelor's earning of each college by implementing:

Bachelors_Earning = MD_EARN_WNE_P6 / (GBF * Grad_population + 1 - Grad_population)

Thus, we now have a method to estimate Bachelors' earnings from College Scorecard's MD_ EARN_WNE_P6 and Grad population. After applying the above code lines, the list of the few very top colleges with the highest Bachelor's earnings is shown in Table 4 (Colors indicate the colleges' prime field of study, mentioned in the note in Table 2). After considering graduate degree earnings, the top highest-earners-producing colleges seem to be colleges specializing in Health professions (Rank numbers shown in blue), Engineering/Computer Science/Technology (Rank numbers shown in red), or Business. The few colleges specializing in Marine Transportation are lumped into the Engineering category because of the small number.

	College Name	BS Earn, \$/yr	Median_P6 \$/yr	Grad Population
1	Albany Coll of Pharmacy& HealthSci	94170	112100	31%
2	St. Louis College of Pharmacy	92499	120400	49%
3	Maine Maritime Academy	73957	75200	3%
4	Samuel Merritt University	71471	100100	65%
5	Rose-Hulman Institute of Technology	68730	70700	5%
6	West Coast University-Los Angeles	67954	74600	16%
7	United States Merchant Marine Academy	64295	65200	2%
8	Kettering University	63997	70700	17%
9	Massachusetts Maritime Academy	63776	66300	6%
10	MCPHS University	60040	75700	42%
11	Massachusetts Institute of Technology	59967	82200	60%
12	Colorado School of Mines	59802	69200	26%
13	Rensselaer Polytechnic Institute	59528	66100	18%
14	Babson College	59166	70400	31%
15	Milwaukee School of Engineering	58269	61000	8%
16	Resurrection University	58127	65300	20%
17	SUNY Maritime College	58038	61200	9%
18	Bentley University	57437	65800	24%
19	Bucknell University	57123	57700	2%
20	Chamberlain University-Illinois	56989	67100	29%
21	Lehigh University	56184	66200	29%
22	Bryant University	56177	57600	4%
23	Duke University	56059	76300	59%
24	Oregon Health & Science University	55842	80000	70%
25	Worcester Polytechnic Institute	55768	67500	34%
26	Claremont McKenna College	55306	55900	2%

(6)

27	University of the Sciences	54691	68400	41%
28	Oregon Institute of Technology	53881	54400	2%
29	Cornell University	53530	64800	34%
30	University of Pennsylvania	53417	71600	55%
31	Georgia Inst Tech - Main Campus	53210	65500	38%
32	Carnegie Mellon University	52525	69800	54%

Table 4. Health professions (Blue) and Engineering/Computer Science/Technology (Red) give highest earnings.

THE MAJOR IS MORE IMPORTANT THAN THE COLLEGE NAME

Even after adjusting for graduate degree earnings, the top highest-earners-producing colleges seem to be colleges with high proportions of students in Health professions, Engineering (including Computer Science and Information Technology), Transportation of Materials, and Business. To test this conjecture, the author uses Data Visualization with Python Seaborn library. Figure 2 shows colleges with the top 25% (about 350) highest Bachelors' earnings (descending order from top to bottom). On the horizontal axis are 31 Fields of Study in the IPEDS database. Figure 2 suggests that the highest-earners-producing colleges give many of their degrees in just a few fields of study below:

- 1. Health professions (e.g., nursing, dental assistance, medical technicians). Note that the earnings considered here are Bachelor's earnings, thus excluding MD and Ph.D. earnings. Still, these top colleges produce alumni with very high earnings.
- 2. Engineering, which includes Computer Science and Information Technology since many Computer Science departments are part of a college of Engineering.
- 3. Business
- 4. Social Science
- 5. Transportation of Materials. Only a handful of colleges have this field of study as their largest. But most of those colleges produce alumni with remarkably high earnings.

The brightness of the rectangular symbols denotes the percentage of students at the college who are in the field of study. White/yellow is the brightest color, denoting that nearly 100% of the students at the college are in one major. Orange denotes a lower proportion than yellow, but still very high (around 80%). Red is a little less bright (meaning a little lower) than orange. Purple is lower than red, followed by dark purple, blue, etc. Black means near-zero.



Figure 2. Highest-Earners-Producing colleges give degrees in a few fields of study in common.

Some fields of study are most common in the highest-earners-producing colleges because they are the most popular fields for all other schools too. According to the National Center for Education Statistics (*Most Popular Majors*, 2020), "Of the 1,956,000 Bachelor's degrees conferred in 2016–17, the greatest numbers of degrees were conferred in the fields of business (381,000), health professions and related programs (238,000), social sciences and history (159,000), psychology (117,000), biological and biomedical sciences (117,000), engineering (116,000), communication, journalism, and related programs (94,000), and visual and performing arts (91,000)". Among the 400 highest-earners-producing colleges, Business appears in most places, especially in ranks between 100 and 300. However, the most dominant fields in the top 100 appear to be Health Professions and Engineering. Note that *Most Popular Majors* (2020) actually discusses fields of study.

To examine the above conjecture closer, Figure 3 shows the top 80 highest-earners-producing colleges, with only nine of the fields of study that these colleges teach. Lighter colors denote higher proportions of degrees awarded in the field of study. The colleges are shown from top to bottom in descending order of Bachelors' earnings. Figure 3 reveals that Biological/Biomedical Sciences is also a common field of study at highest-earners-producing colleges.



Figure 3. Top-80 Highest-Earners-Producing colleges teach mainly 1)Medical Profession; 2)Engineering including Computer Science/Technology; 3)Business; 4) Social Science; 5)Maritime Transportation; 6. Biological/Biomedical Sciences, and 7) Computer technology (IT, Cyber). Going back to the top quartile (roughly 400) colleges in terms of earnings, Fig. 2 suggests that:

- 1. Health professions are the most widely taught, from the very top highest-earnersproducing colleges all the way through rank 400 (and probably beyond).
- 2. Engineering is taught mainly at the very top 80 highest-earners-producing colleges, getting less common from there down.
- 3. Business and social science are widely taught everywhere throughout the 400 colleges.

Among the top-quartile-earnings colleges, Figure 4 shows that the very few top earners are colleges that have 100% of their students in Health Professions. In general, colleges with a high proportion of Health Professions students produce high Bachelor's earnings. Also among the top-quartile-earnings colleges, Figure 5 shows that colleges with a high proportion of Engineering students produce high Bachelor's earnings. Appendix E shows that proportion of Business students does not affect alumni earning significantly.



Figure 4. In general, colleges with a high proportion of Health Professions students produce high Bachelor's earnings. In particular, colleges that produce top earners have 100% of their students in Health Professions.

The hypothesis from Figure 5 is: The field of study is what makes college graduate alumni with high earnings. To investigate this hypothesis, the author refers to a report from George-town University (Carnevale et al, 2015), henceforth called "The Georgetown Report". Among other conclusions, the Georgetown Report says that a Bachelor's degree results in lifetime earnings of \$1 million higher than a high school diploma. However, a Bachelor's degree in a high-paying major results in lifetime earnings of \$3.4 million higher than a Bachelor's degree in a low-paying major. Thus, the popular question "How do we choose a college?" is far less



Figure 5. Colleges with a high proportion of Engineering/Computer Science students generally produce high Bachelors' earnings.

important than the better question: "How do we choose a college major?" Torpey (2003) can help answer the latter question, and the analysis below will answer both.

The Georgetown Report analyzes hundreds of college majors, which have an advantage over the College Scorecard for choosing a college major. The College Scorecard databases have only 38 fields of study because they "are easier for prospective students to understand and because combining six-digit CIP codes together leads to larger cell sizes, which in turn leads to fewer data points that need to be privacy-suppressed" (College Scorecard Data by Field of Study, 2020). In gathering and reporting information, the National Center for Education Statistics has used most of the Fields of Study since 1970 or earlier (Digest of Education Statistics (Fields of Study), 2019). Some of those 38 fields have all but disappeared or evolved far beyond their names. For example, in 1970-1971, Library Studies had 1013 new college graduates. In 2017-2018, it had 81, while Business had 386,201 new college graduates. (Health Professions had 244,909. Engineering, 121,956.) Other Fields of Study are also so small that they should be merged. In 2017-2018, Communication Studies had 4231 new graduates; Legal Professions and Studies, 4239; Military Technologies and Applied Sciences, 655; Precision Production, 45; and Transportation and Materials Moving, 4924. Therefore, today the college major is more relevant than the field of study discussed above. US colleges award degrees in hundreds of college majors; College majors have much finer granularity than the 38 Fields of Study. Most importantly, every college applicant needs to plan for, or often decide, which major-not field of study-she or he will pursue a degree in. The Georgetown report shows average earnings by major in Table 5 below. Outside Pharmacy, the highest paying majors are in Engineering (including Computer Science/Technology). Other Health Professions majors besides Pharmacy also result in high earnings (Carnevale et al., 2015).

2014-2015 Ranking	Median, \$/yr	25%ile, \$/yr	75%ile, \$/yr
Petroleum Engineering	120000	82000	189000
Pharmacy/Pharm Sci/Pharm Admin	105000	83000	120000
Math and Computer Science	98000	75000	134000
Aerospace Engineering	87000	60000	115000
Chemical Engineering	86000	60000	120000
Electrical Engineering	85000	60000	110000
Naval Architecture and Marine Eng	82000	44000	120000
Mechanical Engineering	80000	59000	105000
Metallurgical Engineering	80000	50000	106000
Mining and Mineral Engineering	80000	52000	125000

Table 5. Outside of Pharmacy, the top ten highest-paying majors are all in Engineering (including Computer Science and IT)(Carnevale et al., 2015).

College major is more important than the college name. Legendary elite colleges charge high tuition regardless of major or alumni earning. To get a good return on investment, a college applicant should first know (or at least make an educated guess of) which major he or she should pursue. Only then should he or she choose the colleges to apply to. To choose colleges to study the predetermined major, tuition must be considered very seriously. As exemplified in Figure 1, colleges with the highest tuition do not necessarily produce alumni with the highest earnings. The converse is also true. Table 6 shows a handful of rather extreme examples to drive the above point home. At the Bachelor's level, many other colleges produce alumni with equal or higher earnings than legendary elite colleges. (However, at the graduate school level, the elite colleges do produce alumni with very high earnings. See Appendix K for examples.)

College Name	Tuition, \$/yr	Major	BA/BS Start Earning \$/yr
Columbia U	61850	Architecture	47900
U Penn	57770	Psychology	52900
Yale U	55500	Sociology	47900
Texas A&M U ¹⁾	11232	Petroleum Engineering	111000
New Mexico Tech ²⁾	8156	Computer & Info Sci	73300

Table 6. The major determines earning far more than the tuition – a few examples. ¹⁾ Out-of-state \$37,726/yr; ²⁾ Out-of-state \$23,524/yr; All values are from 2019.

As discussed above, the handful of colleges that produce top earners are colleges that have almost 100% of their students in Health Professions. Financially, it would be wise to choose a major in Health Professions even at the undergraduate level for students. If they do so, Table 7 gives them an idea of which "Bachelor's colleges" to apply to. Engineering is another top-earning profession, including Computer Science. In further analysis in this paper, Computer Information, Engineering Technology, and Transportation are lumped into Engineering. If a student is interested in Engineering, Table 8 gives an idea of the colleges that

produce the highest Bachelor's earners. A little caveat: a few colleges in that table have fewer than 20% of their students in Engineering. Therefore, the high earners they produce might not be their engineers but other professionals including health professionals.

Rank	College Name	BS Earn, \$/yr	MD_EARN _WNE_P6, \$/yr	HealthPro Fraction %
1	Albany College of Pharmacy and Health Sciences	94170	112100	100
2	St. Louis College of Pharmacy	92499	120400	100
3	Samuel Merritt University	71471	100100	100
4	West Coast University-Los Angeles	67954	74600	100
5	MCPHS University	60040	75700	97
6	Resurrection University	58127	65300	100
7	Chamberlain University-Illinois	56989	67100	100
8	Oregon Health & Science University	55841	80000	100
9	University of the Sciences	54691	68400	81
10	Oregon Institute of Technology	53881	54400	51
11	Excelsior College	52299	54700	32
12	AdventHealth University	52071	54300	96
13	Bellin College	51455	55600	100
14	Mount Carmel College of Nursing	50959	55500	100
15	SUNY Downstate Medical Center	50105	75300	100
16	Roseman University of Health Sciences	50039	74900	100
17	Research College of Nursing	49564	59400	100
18	Louisiana State Uni Health Sci Center-Shreveport	49187	78200	100
19	Loma Linda University	48944	72000	99
20	Barnes-Jewish College Goldfarb School of Nursing	48746	53900	100
21	Texas Tech University Health Sciences Center	47704	68300	100
22	American University of Health Sciences	47275	48100	100
23	Saint Anthony College of Nursing	47037	54100	100
24	Saint Luke's College of Health Sciences	47031	52200	100
25	Blessing Rieman Coll of Nursing & Health Sci	46838	49000	100
26	The University of Texas Medical Branch	46516	68800	100
27	Mount Saint Mary's University	46471	51900	41
28	Thomas Jefferson University	46374	69000	96
29	Kettering College	45853	48800	100
30	Molloy College	45768	53100	54
31	University of Maryland Baltimore	45527	70000	100
32	MGH Institute of Health Professions	44887	67900	100

Table 7. Bachelor's degree highest-earner producing colleges in Health Professions. "BS" denotes all types of Bachelor's degrees. Earnings are the college's alumni earning six years after enrollment. "BS Earning" is estimated using Equation (6). Average Earning is from the College Scorecard database. Health Pro Fraction denotes how many percent of the college's students are in Health Professional majors.

Rank	College Name	BS Earn, \$/yr	MD_EARN _WNE_P6, \$/yr	Eng or CS Fraction %
1	Maine Maritime Academy	73956	75200	69
2	Rose-Hulman Institute of Technology	68730	70700	93
3	United States Merchant Marine Academy	64295	65200	100
4	Kettering University	63996	70700	90
5	Massachusetts Maritime Academy	63776	66300	57
6	Massachusetts Institute of Technology	59966	82200	62
7	Colorado School of Mines	59801	69200	91
8	Rensselaer Polytechnic Institute	59527	66100	63
9	Milwaukee School of Engineering	58268	61000	69
10	SUNY Maritime College	58037	61200	41
11	Bucknell University	57122	57700	20
12	Lehigh University	56183	66200	38
13	Duke University	56058	76300	17
14	Worcester Polytechnic Institute	55767	67500	78
15	Oregon Institute of Technology	53881	54400	17
16	Cornell University	53530	64800	22
17	Georgia Inst of Tech - Main Campus	53210	65500	68
18	Carnegie Mellon University	52525	69800	38
19	Clarkson University	52339	57700	57
20	University of Notre Dame	51994	61800	15
21	Stanford University	51753	70400	29
22	Stevens Institute of Technology	51693	68600	74
23	Wentworth Institute of Technology	51691	53300	17
24	Missouri University of Science and Technology	50928	58700	75
25	Cal Poly University-San Luis Obispo	50645	52100	27
26	Princeton University	50316	60800	20
27	Manhattan College	50313	54200	29
28	University of Portland	50279	53800	15
29	South Dakota School of Mines and Technology	49467	53200	83
30	Santa Clara University	49209	61100	19
31	Michigan Technological University	49014	55200	65
32	DigiPen Institute of Technology	47912	50400	_60

Table 8. Bachelor's degree highest-earners-producing colleges in Engineering/Computer Science. "BS" denotes all types of Bachelor's degrees. Earnings are the college's alumni earning six years after enrollment. "BS Earning" is estimated using Equation (6). Average Earning is from the College Scorecard database. Eng or CS Fraction denotes how many percent of the college's students are in Engineering/Computer Science majors.

HIGH-EARNERS-PRODUCING COLLEGES TEND TO CHARGE HIGH TUITION

Besides earnings, tuition is an important consideration in choosing colleges to apply to, since college tuitions today are almost the same magnitude as whole full-time salaries. Figure 6 shows the distribution of tuition based on IPEDS Scorecard data (Appendix F) after adjusting for undergraduate programs using Eq. (6). Linear regression analysis estimates a "Gradu-

ate Tuition Factor" (GTF) of 2.1. The coefficient of correlation R = 0.32. After that adjustment, Bachelor's tuitions are shown as distributions in Figure 6 based on whether the colleges are public, private nonprofit, or private for-profit. The horizontal axis is the bins (tuition ranges from \$0 to \$999, \$1,000 to \$1,999, \$2,000 to \$2,999, etc). The vertical axis is the number of colleges corresponding to each bin. Figure 6 shows that, generally, public colleges charge far lower tuition than private colleges. Private for-profit colleges do not seem to charge more than private nonprofit colleges. In fact, the most expensive colleges for Bachelor's degrees are private nonprofit. The following may help in reading the overlapping distributions in Figure 6: Public tails off around \$20,000/yr; Private Nonprofit tails off at over \$40,000/yr.



Figure 6. Distribution of Undergraduate Tuition. Generally, public colleges charge far less than private colleges.

The tuitions (including fees) listed in the Scorecard database are average values. The variable TUITFTE in the College Scorecard database is an average tuition that does not apply to most students. Each college applicant's situation determines the real tuition they will have to pay. Attending a public college within their state will likely cost far less (in tuition alone) than attending a private college or an out-of-state public college. Most colleges give financial aid to students from low-income families, some even to families with income over \$100,000/year. Students from higher-income families may have to pay far more in tuition than average. A very expensive Ivy League college may offer the best deal to a highly qualified student from a low-income family with the potential for merit-based scholarships. Tuition is the most widely varying, complicated, and least certain variable in the analyses in this paper. Although very important, the discussion on tuition in this paper is at the "macro-economic" level and not for choosing a college to attend.

Figure 7 shows that more expensive colleges tend to produce alumni with higher earning on average. Zero tuition corresponds to an earning of 25,677/yr. Every 1,000 increment of tuition corresponds to roughly 540 increment in earnings. The linear fit has considerable scatter (R = 0.36). Tuitions of the very top high-earners-producing colleges vary



Figure 7. Bachelor's earning versus Bachelors' tuition: High-earners-producing colleges tend to charge high tuition in general. However, the very top highest-earners-producing colleges, listed in the inset, do not charge very high tuition.

widely. However, they tend to be not the highest. Caveat: the most expensive colleges are not the highest earners producers. In fact, the three rightmost dots in Fig. 7 pertain to the lowest-earners producing colleges in Table 1. Another caveat: Although private for-profit colleges generally charge more than public colleges, they produce alumni with lower earnings than public or private non-profit colleges (Fig. 8).

ALUMNI OF HIGHEST-EARNERS-PRODUCING-COLLEGES DO NOT TAKE HIGH STUDENT LOANS IN GENERAL

Most college students have student loan debt, with no college in the US reporting zero student debt. In the College Scorecard database, the variable GRAD_DEBT_MDN is the median student loan debt accumulated at the college by all student borrowers of federal loans who graduate in a given fiscal year, measured at graduation. Figure 9 shows the distribution of colleges' median student debts in 2018 for 1977 four-year colleges in the US. (We use 2018 debt data because they were College Scorecard's latest at the time this analysis was done.) The horizontal axis is the bins or debt ranges. The vertical axis is the number of colleges corresponding to each debt range. Only Bachelors' debts are included, and private debts are

excluded as the US Department of Education collects data on federal debts only. No debts are in the names of the parents of the undergraduates.

Our data show that students at private for-profit colleges are the highest debtors. At two such colleges, the median debt at graduation in 2018 was \$47,000. So 50% of their graduating students owed more than that. Ninety-two and 88 percent of undergraduates at those two colleges took federal loans. The College Scorecard database does not yet have their alumni earning information. Because 1453 colleges (outside of 1977 in this analysis) did not report



Figure 8. Distribution of Bachelor's earnings: Private for-profit colleges tend to produce alumni with low earnings.



Figure 9. Distribution of Alumni Debt. Generally, alumni of public colleges accumulate less debt than alumni of private colleges.

GRAD_DEBT_MDN, it is quite likely that many colleges produced higher median debts than \$47,000.

Figure 10 shows that there is no correlation between debt and earning. The data are from years 2014-2015, the latest year that earning data are available. The slope of the line is -0.07; and the correlation coefficient is -0.002. Therefore, the plot is pure random scatter for all practical purposes. The complete absence of correlation between debt and earning means that colleges whose students have a lot of debt do not produce alumni with high earnings. Alumni of these colleges often take on large debts without gaining the resources to pay them off in the future. Before enrolling or applying to a college, students should consider the average student debt at that college.



Figure 10. Bachelor's earning versus alumni debt: Alumni of colleges with higher debt do not earn a higher salary.

HIGHEST-EARNERS-PRODUCING COLLEGES REQUIRE HIGH SAT SCORES BUT HAVE REASONABLE ADMISSION RATES

Analyses above revealed that highest-earners-producing colleges are not always the most expensive. We will now analyze if they are the most selective. Table 9 suggests that many highest-earners-producing colleges have reasonable admission rates and SAT scores. Regression analysis shown in Figure 11 shows that Bachelor's earning is correlated with combined SAT scores. At the low extreme, a 75th percentile SAT score of 800 corresponds to a Bachelor's earning of \$21,949/yr. Every 100-point increase in the SAT score corresponds to an earnings increase of \$2,872/yr. At the high extreme, a 75th percentile SAT score of 1600 (perfect score) corresponds to an earning of \$45,723/yr. The top 10 highest-earners-producing colleges have 75th percentile combined SAT scores between 1130 and 1400. Among all reported SAT scores, the median is 1150, and 1246 is the upper quartile. In conjunction with SAT scores, Appendix J shows that most high-earners-producing colleges do not have low admission rates. The admission rates to the 10 highest-earners-producing colleges are between 25% and 89%, mostly above 58%. These admission rates are far more generous than the admission rates to legendary colleges like the Ivy League colleges, or Massachusetts Institute of Technology where the admission rate is 8%. (Its 75th percentile SAT is 1570.)

Rank	College Name	BS Earn, \$/yr	BS Tuit \$/yr	SAT 75%ile	Admit Rate %
1	Albany College of Pharmacy and Health Sci	94170	18254	1260	67
2	St. Louis College of Pharmacy	92499	15554	1330	
3	Maine Maritime Academy	73957	16350	1130	79
4	Samuel Merritt University	71471	16267		
5	Rose-Hulman Institute of Technology	68730	21965	1400	59
6	West Coast University-Los Angeles	67954	23241		25
7	United States Merchant Marine Academy	64295	587		
8	Kettering University	63997	17555	1310	72
9	Massachusetts Maritime Academy	63776	8224	1190	62
10	MCPHS University	60040	18952	1190	84
11	Massachusetts Institute of Technology	59967	17152	1570	8
12	Colorado School of Mines	59802	14513	1400	36
13	Rensselaer Polytechnic Institute	59528	22181	1490	38
14	Babson College	59166	25090	1370	26
15	Milwaukee School of Engineering	58269	13768	1260	69
16	Resurrection University	58127	15362		
17	SUNY Maritime College	58038	5810	1190	53
18	Bentley University	57437	20489	1330	46
19	Bucknell University	57123	31201	1400	31
20	Chamberlain University-Illinois	56989	12589		66
21	Lehigh University	56184	19598	1410	34
22	Bryant University	56177	22952		75
23	Duke University	56059	13488	1550	11
24	Oregon Health & Science University	55842	13304		
25	Worcester Polytechnic Institute	55768	23627		44
26	Claremont McKenna College	55306	29500	1520	11
27	University of the Sciences	54691	15543	1260	58
28	Thunderbird School of Global Management	54158	33374		
29	Oregon Institute of Technology	53881	7051	1150	60
30	Cornell University	53530	18007	1510	14
31	University of Pennsylvania	53417	21153	1550	10
32	Georgia Institute of Tech - Main Campus	53210	8973	1490	33
33	Carnegie Mellon University	52525	22402	1540	25

Table 9. Bachelor's highest-earners-producing colleges require high SAT scores, but are not necessarily very hard to get into. "BS" denotes all types of Bachelor's degrees. Earnings are the college's alumni earning six years after enrollment. "BS Earning" is estimated using Equation (6). SAT 75% ile is the 75th percentile SAT total score. All data are from the College Scorecard database.



Figure 11. Bachelor's earning plotted against 75th percentile SAT score. High-earners-producing colleges tend to require high SAT scores.

As the last step of this study, the author has created a list of the top-quartile (25% = about 400) highest-earners-producing colleges for Bachelor's degrees, with their 1) Median Bachelor's earning at six years after enrollment; 2) Field of Study with most students; 3) Total SAT Score at 75th percentile; 4) Admission Rate; 5) Bachelor's Tuition (including fees); and 6) Median Debt. A college applicant can narrow down the list further by specifying his/her choice of major. The *List of Colleges with Highest Bachelor Alumni Earnings* (2020) will be maintained on the College Analytica website. Appendices G, H, I, and J show more information admission competitiveness and student debts. Appendix J lists the colleges with the highest bachelor's earnings. Figure 12 illustrates the earnings versus tuition confirming the conclusion from early in this paper: The very top high-earners-producing colleges have large proportions of their students in Health Professions or Engineering. Many high-earners-producing colleges are big on business, which is the most popular major, with more alumni per year than Health Professions and Engineering combined.

The IPEDS database used in this analysis is from the years 2014-2015 because those are the years for which earning data are available. For relative comparisons among colleges, the earning data are very useful. However, the absolute values are old and need to be updated. For the latest earning information, refer to US News and World Report's College Compass (premium/paid subscription). From that website, the part that gives median starting salaries



Figure 12. Bachelor's earning plotted against tuition, for the top quartile (25% = 400) high-earnersproducing colleges. Green o symbol means Business > 10%; Blue x means Health Profession > 10%; Red + means Engineering/Computer Science > 10% of undergrad population at the college.

by major does not say whether the starting salaries are those of Bachelors or graduate degree holders. Therefore, the starting salary information may have to be adjusted to account for the proportion of graduate students at each college. Use Eq. (6) to estimate Bachelors' earnings. As this research shows, colleges with high graduate student population proportions produce alumni with higher earnings than Bachelor's earning. Also, College Compass earning data are based on self-reported survey entries, unlike the IPEDS data that are based on tax returns and information required by student loan providers.

CONCLUSIONS

Data science and the US Department's College Scorecard database enable students to choose colleges based on college alumni's earnings. The most expensive or famous colleges do not necessarily produce alumni with the highest earnings. The highest-earners-producing colleges produce alumni who get the highest salaries, to a large extent because those colleges give a large proportion of their degrees in graduate or professional degrees (MD, JD/Law, Ph.D., MS, etc.). Therefore, the earnings reported by each college are skewed rather heavily by the college's graduate population fraction. Colleges that award a large proportion of graduate

degrees report higher alumni earnings, which are misleading to applicants to undergraduate programs. To correct for the graduate degree effect, we introduced the concept of "Graduate Boost Factor" (GBF) which enables us to estimate the earnings of each college's Bachelor's alumni based on the percentage of graduate students at that college.

Even after accounting for GBF, highest-earners-producing colleges give many degrees in just a few fields of study (as defined by the US Department of Education):

- 1. Health professions (e.g., nursing, dental assistance, medical technicians). Note that the earnings considered here are Bachelor's earnings, thus excluding MD and Ph.D. earnings. Still, these top colleges produce alumni with very high earnings.
- 2. Engineering, including Computer Science and technology
- 3. Business
- 4. Maritime Transportation. (These include only a handful of colleges. But their earnings are remarkably high).

College majors are a finer granularity than the fields of study. A college major is more important than the college name. Notable elite colleges charge high tuition regardless of major or alumni earning.

To choose colleges to study the predetermined major, tuition must be considered very seriously. Public colleges charge significantly less tuition than private colleges. High-earners-producing colleges tend to charge higher tuition than other colleges. Colleges with the highest tuition do not necessarily produce alumni with the highest earnings. Colleges with the lowest tuition do not necessarily produce alumni with low earnings.

High-earners-producing colleges tend to require higher SAT scores. However, admission rates to most of those colleges seem to be reasonable. The few colleges with low admission rates (the most extreme is 1 in 19 applicants) that are high-earners-producing colleges are top-tier elite colleges. However, at the Bachelor's level, many other colleges produce alumni with equal or higher earnings than those elite colleges. At the graduate school level, the elite colleges do produce alumni with very high earnings (Appendix K).

All colleges report significant amounts of student loan debt. Bachelor's alumni with larger college debts do not earn more than those with smaller debts. Private for-profit colleges incur huge student loan debts and produce alumni with lower earnings in general. College loan debts are so large that they are one of the top economic problems in the U.S.

The US Department of Education's College Scorecard database is very useful in revealing important information for choosing colleges to apply to and to attend. Ten years ago, there was no simple way to perform college cost-benefit analysis and reach the above conclusions. Python data science library Pandas and visualization library Seaborn (which uses and supplements Matplotlib) facilitated the analysis tremendously. Future work will incorporate machine learning to help students choose a major and the colleges that will give them the best financial value.

AUTHOR INFORMATION

Corresponding Author

*Mario Sumali, swimsuper12@gmail.com

ACKNOWLEDGMENTS

The author is grateful for the guidance of Ms. Yolanda Lozano, Computer Science teacher at La Cueva High School. The author also thanks Mr. Duncan Lee of MIT and Dr. Ronghua Wang of Intel for their valuable reviews and guidance.

REFERENCES

- *College Scorecard Data by Field of Study*; Technical Documentation; United States Department of Education. Washington, DC, 2021; <u>https://collegescorecard.ed.gov/assets/FieldOfStudyDataDocumentation.pdf</u>
- *Digest of Education Statistics (Fields of Study)*; National Center for Education Statistics (NCES), 2019; <u>https://nces.ed.gov/programs/digest/d19/tables/dt19_322.10.asp</u>
- Friedman, Z. *Student Loan Debt Statistics In 2020: A Record \$1.6 Trillion*; Forbes, Feb. 3, 2020; <u>https://www.forbes.com/sites/zackfriedman/2020/02/03/student-loan-debt-statis-tics/#d38db5e281fe</u>
- *List Of Colleges With Highest Bachelor Alumni Earnings*; <u>College Analytica website</u>; collegeanalytica.com.
- *Most Popular Majors*; Fast Facts; National Center for Education Statistics (NCES), 2020; <u>https://nces.ed.gov/fastfacts/display.asp?id=37</u>
- Seeger, J. A Starter Guide to College for Clueless Students & Parents: For a State College or the *Ivy League, Here's What You Need to Know*; Starter Guides LLC, 2018; ISBN 0999586025, 344p.
- Torpey, E.M. *The Class of 1993: Earnings and occupations by college majors, 1 and 10 years after graduation;* Occupational Outlook Quarterly, Summer 2008; <u>https://www.bls.gov/careeroutlook/2008/summer/art02.pdf</u>
- US News And World Report College Compass Premium Subscription; <u>https://premium.</u> <u>usnews.com/best-colleges</u>

APPENDIX A: LIST OF VARIABLES

From IPEDS full documentation, the author used the following variables in the analysis:

- 1. INSTNM: The institution's name, as reported in IPEDS.
- 2. TUITFTE: The net tuition revenue per full-time equivalent (FTE) student uses tuition revenue minus discounts and allowances, and divides that by the number of FTE undergraduate and graduate students.
- 3. MD_EARN_WNE_P6: 6_yrs_after_entry.median [working_not_enrolled.]
- 4. MD_EARN_WNE_P10: 10_yrs_after_entry.median [working_not_enrolled.]
- 5. PCT75_EARN_WNE_P6: The variable in the IPEDS database is called 6_yrs_after_entry.working_not_enrolled.earnings_percentile.75
- 6. PCT25_EARN_WNE_P6: 6_yrs_after_entry.working_not_enrolled.earnings_ percentile.25
- 7. PCT75_EARN_WNE_P10: 10_yrs_after_entry.working_not_enrolled.earnings_ percentile.75
- 8. PCT25_EARN_WNE_P10: 10_yrs_after_entry.working_not_enrolled.earnings_ percentile.25
- 9. UGDS: Number of degree/certificate-seeking undergraduates enrolled in the fall, as reported in the IPEDS Fall Enrollment component.
- 10. GRADS: Number of graduate students.
- 11. CONTROL: Governance structure: public, private nonprofit, or private for-profit.
- 12. PCIP01: Percentage of degrees awarded in Agriculture, Agriculture Operations, And Related Sciences. (program_percentage.agriculture in IPEDS database)
- 13. PCIP03: Percentage in Natural Resources And Conservation
- 14. PCIP04: Percentage in Architecture And Related Services
- 15. PCIP05: Percentage Area, Ethnic, Cultural, Gender, and Group Studies
- 16. PCIP09: Percentage Communication, Journalism, and Related Programs
- 17. PCIP10: Communications Technologies/Technicians and Support Services
- 18. PCIP11: Percentage in Computer and Information Sciences and Support Services
- 19. PCIP12: Percentage Personal and Culinary Services
- 20. PCIP13: Percentage Education
- 21. PCIP14: Percentage of degrees awarded in Engineering
- 22. PCIP15: Percentage Engineering Technologies and Engineering-Related Fields
- 23. PCIP16: Percentage Foreign Languages, Literatures, and Linguistics
- 24. PCIP19: Percentage Family and Consumer Sciences/Human Sciences
- 25. PCIP22: Percentage Legal Professions and Studies

- 26. PCIP23: Percentage English Language and Literature/Letters
- 27. PCIP24: Percentage Liberal Arts and Sciences, General Studies and Humanities
- 28. PCIP25: Percentage Library Science
- 29. PCIP26: Percentage Biological and Biomedical Sciences
- 30. PCIP27: Percentage Mathematics and Statistics
- 31. PCIP29: Percentage Military Technologies and Applied Sciences
- 32. PCIP30: Percentage Multi/Interdisciplinary Studies
- 33. PCIP31: Percentage Parks, Recreation, Leisure, and Fitness Studies
- 34. PCIP38: Percentage Philosophy and Religious Studies
- 35. PCIP39: Percentage Theology and Religious Vocations
- 36. PCIP40: Percentage Physical Sciences
- 37. PCIP41: Percentage Science Technologies/Technicians
- 38. PCIP42: Percentage Psychology
- 39. PCIP43: Pct Homeland Security, Law Enforcement, Firefighting and Related Protective Services
- 40. PCIP44: Percentage Public Administration and Social Service Professions
- 41. PCIP45: Percentage Social Sciences
- 42. PCIP46: Percentage Construction Trades
- 43. PCIP47: Percentage Mechanic and Repair Technologies/Technicians
- 44. PCIP48: Percentage Precision Production
- 45. PCIP49: Percentage Transportation and Materials Moving
- 46. PCIP50: Percentage Visual and Performing Arts
- 47. PCIP51: Percentage of degrees awarded in Health Professions and Related Programs
- 48. PCIP52: Percentage Business, Management, Marketing, and Related Support Services
- 49. PCIP54: Percentage History
- 50. ADM_RATE_ALL: Number of admitted undergraduates divided by the number of undergraduates who applied.
- 51. SATVR25, SATVR75, SATMT25, SATMT75, ACTCM25, ACTCM75: The files include the 25th and 75th percentiles of SAT reading (SATVR* for _25 and _75), writing (SATWR* for _25 and _75), math (SATMT* for _25 and _75)
- 52. COSTT4_A: Average annual cost of attendance includes tuition and fees, books and supplies, and living expenses for all full-time undergraduate.
- 53. TUITIONFEE_OUT: Tuition + fees, for out-of-state undergraduates.
- 54. TUITIONFEE_IN:Tuition + fees, for in-state undergraduates.
- 55. DEBT_MDN: median loan debt accumulated at the institution by all student

borrowers of federal loans who separate (i.e., either graduate or withdraw) in a given fiscal year.

- 56. PCTFLOAN: The share of undergraduate students who received federal loans.
- 57. LOAN_COMP_ORIG_YR4_RT: Percent of students who received a federal loan at the institution and who completed in 4 years at original institution
- 58. HIGHDEG: Highest award identifies the highest award level conferred at the institution.
- 59. PREDDEG: Predominant undergraduate award identifies the type of award that the institution primarily confers.

APPENDIX B: KEY LINES IN THE CODE

```
df clean = df clean[df clean['UGDS'].notnull()] #Excludes col-
leges that give no Bachelor's degrees.
df clean = df clean[(df clean['HIGHDEG'] == 'Graduate')]
df clean = df clean[df clean['MD EARN WNE P6']!= 'PrivacySup-
pressed']
df clean = df clean[df clean['MD EARN WNE P6'].notnull()]
df clean['MD EARN WNE P6'] = df clean['MD EARN WNE P6'].as-
type(int)
df clean['MD EARN WNE P6'] = pd.to numeric(df clean['MD EARN WNE
P10'], errors='coerce')
# Do above to all variables that need to be converted to numeric
sns.regplot(data = df clean, x = 'Grad population', y = 'MD EARN
WNE P6', fit reg = True, scatter kws = {'alpha':1, 's': 5})
GBF = (E6 ugds mean + Grad slope*1) / E6 ugds mean
df clean['Bachelor MD EARN WNE P6'] = df clean.MD EARN WNE P6 / (
GBF * df clean.Grad population + df clean.Grad population)
GTF = (Tuitn ugds mean + Grad tslope) / Tuitn ugds mean
print('Graduate Tuition Factor GTF = ', GTF)
#Plot color chart
img = plt.pcolormesh(df PCIP, cmap=cmap, vmin=0.0002, vmax=0.6)
```

Rank	College Name	Earning, \$/yr	UGDS	GRADS
1	Philadelphia Coll of Osteopathic Medicine	127400	NaN	2806
2	Western University of Health Sciences	108100	NaN	3842
3	Salus University	100400	NaN	1124
4	Southern College of Optometry	99300	NaN	527
5	A T Still University of Health Sciences	99000	NaN	3226
6	Illinois College of Optometry	96000	NaN	638
7	University of California-San Francisco	95400	NaN	3170
8	Northeast Ohio Medical University	91700	NaN	893
9	Brooklyn Law School	89300	NaN	1141
10	Midwestern University-Glendale	87500	NaN	3146
11	Midwestern University-Downers Grove	87500	NaN	2917
12	U of California-Hastings College of Law	86100	NaN	1003
13	Marshall B Ketchum University	83300	NaN	427
14	Southwestern Law School	74100	NaN	1106
15	New York Law School	72400	NaN	1029
16	New York Medical College	70300	NaN	1482
17	Albany Medical College	68800	NaN	823
18	CUNY School of Law	68500	NaN	325
19	Icahn School of Medicine at Mount Sinai	68400	NaN	1074
20	Rosalind Franklin U of Medicine and S	68100	NaN	2191
21	California Western School of Law	67400	NaN	781
22	Rensselaer at Hartford	66100	NaN	191
23	WV School of Osteopathic Medicine	65800	NaN	815
24	Weill Cornell Medical College	64800	NaN	1023
25	New England Law-Boston	64600	NaN	871
26	Meharry Medical College	63500	NaN	802
27	U of Massachusetts Medical School Wor	63500	NaN	1103
28	Albany Law School	61100	NaN	475
29	Mayo Clinic Coll of Medicine and Science	60300	NaN	228
30	Mayo Clinic Grad School of Biomed Sci	60300	NaN	291
31	Teachers College at Columbia University	60000	NaN	5011
32	Brite Divinity School	46900	NaN	196
33	Widener U-Commonwealth Law School	45300	NaN	330
34	Roger Williams University School of Law	42600	NaN	370
35	U of New Hampshire-School of Law	42400	NaN	271
36	New York Chiropractic College	42300	NaN	965
37	Bethel Seminary-San Diego	41100	NaN	176
38	Bethel Seminary-St Paul	41100	NaN	459
39	Penn State University-College of Medicine	40100	NaN	842
40	Penn State University-Penn State Great	40100	NaN	469

APPENDIX C: EXCLUSIVELY GRADUATE OR PROFESSIONAL COLLEGES

41	Penn State University-Dickinson Law	40100	NaN	153	
42	Asbury Theological Seminary	39000	NaN	1470	
43	Saint Vincent Seminary	37700	NaN	45	
44	LIU Hudson at Rockland	33900	NaN	239	
45	LIU Hudson at Westchester	33900	NaN	158	
46	Antioch University-New England	33500	NaN	705	
47	Trinity Law School	32300	NaN	193	
48	Western State Coll of Law at Argosy U	30300	NaN	353	
49	Evangel U - Assemblies of God Theolog	30300	NaN	327	
50	University of the D.CDavid A	29100	NaN	312	
51	National American U-Harold D. Bucking	28200	NaN	358	
52	Southern University Law Center	27300	NaN	644	

APPENDIX D: DISTRIBUTION OF MEDIAN EARNINGS AT SIX YEARS AFTER ENROLLMENT, FOUR-YEAR COLLEGES ONLY



NMJS 2020 | Volume 54 No. 2

APPENDIX E: BACHELOR'S EARNINGS NOT CORRELATED WITH PROPORTION OF BUSINESS MAJORS.



APPENDIX F. DISTRIBUTION OF TUITIONS, NOT ADJUSTED FOR GRADUATE SCHOOL TUITIONS.



Rank	College Name	Bach Earn, \$/yr	SAT Math 75%ile	SAT Verbl 75%ile	Admit rate 1/n	Bach Tuit \$/yr	Median Debt
1	Albany Coll of Pharmacy and Health Sciences	94170	610	650	1.48	18254	23222
2	St. Louis Coll of Pharmacy	92499	620	710		15554	17500
3	Maine Maritime Academy	73957	550	580	1.27	16350	27000
4	Samuel Merritt University	71471				16267	18750
5	Rose-Hulman Inst of Tech	68730	660	740	1.70	21965	25500
6	West Coast University-L.A Angeles	67954			3.92	23241	25000
7	US Merchant Marine Academy	64295				587	5500
8	Kettering University	63997	650	660	1.40	17555	18976
9	Massachusetts Maritime Academy	63776	570	620	1.62	8224	22050
10	MCPHS University	60040	570	620	1.19	18952	25000
11	MIT	59967	770	800	12.69	17152	13000
12	Colorado School of Mines	59802	680	720	2.74	14513	22000
13	Rensselaer Polytechnic Institute	59528	720	770	2.67	22181	26001
14	Babson College	59166	660	710	3.80	25090	26000
15	Milwaukee School of Eng	58269	580	680	1.46	13768	25000
16	Resurrection University	58127				15362	20000
17	SUNY Maritime College	58038	580	610	1.90	5810	20157
18	Bentley University	57437	640	690	2.17	20489	25821
19	Bucknell University	57123	680	720	3.26	31201	26467
20	Chamberlain University- Illinois	56989			1.51	12589	17294
21	Lehigh University	56184	670	740	2.92	19598	22250
22	Bryant University	56177			1.33	22952	25279
23	Duke University	56059	760	790	8.76	13488	6500
24	Oregon Health & Science University	55842				13304	10939
25	Worcester Polytechnic Institute	55768			2.28	23627	27000
26	Claremont McKenna College	55306	750	770	9.29	29500	11500
27	University of the Sciences	54691	610	650	1.73	15543	26935
28	Thunderbird School of Global Management	54158				33374	8000
29	Oregon Inst of Technology	53881	560	590	1.67	7051	15917

APPENDIX G: ADMISSION COMPETITIVENESS AND DEBTS

30	Cornell University	53530	740	770	7.05	18007	10912
31	University of Pennsylvania	53417	770	780	9.64	21153	12018
51	Georgia Tech -	55417		100	5.04	21133	12010
32	Main Campus	53210	720	770	3.00	8973	20500
33	Carnegie Mellon University	52525	740	800	4.07	22402	24827

APPENDIX H: CORRELATION BETWEEN EARNING AND COLLEGES' REJECTION RATE



- Lighter dots correspond to higher SAT scores
- About a thousand colleges admit almost 100% of applicants.
- A little over 100 colleges admit 1 in every 2 applicants.
- If you are top 1/10 applicant, then at least 99% of US colleges would admit you.
- Nine schools have admission rates lower than 1/11.
- The lowest admission rate, Stanford admits only 1 student per 19 applicants.

APPENDIX I: GRADUATE POPULATION PROPORTION HAS NO EFFECT ON DEBT



Grad_Dslope = 283.2
Debt_ugds_mean = 14821
R = 0.0169
Graduate Debt Factor GDF = 1.02

APPENDIX J: HIGHEST-EARNERS-PRODUCING COLLEGES

#	College Name	Bach Earn, \$/yr	Largest Field	SAT Total 75%ile	Admit rate 1/n	Bach Tuit \$/yr	Median Debt
1	Albany College of Pharmacy and Health Sciences	94170	Health Pro	1260	1.48	18254	23222
2	St. Louis Coll of Pharmacy	92499	Health Pro	1330		15554	17500
3	Maine Maritime Academy	73957	Engineer- ing/CS	1130	1.27	16350	27000
4	Samuel Merritt University	71471	Health Pro			16267	18750
5	Rose-Hulman Inst of Tech	68730	Engineer- ing/CS	1400	1.70	21965	25500
6	West Coast University-L.A Angeles	67954	Health Pro		3.92	23241	25000
7	US Merchant Marine Academy	64295	Engineer- ing/CS			587	5500
8	Kettering University	63997	Engineer- ing/CS	1310	1.40	17555	18976
9	Massachusetts Maritime Academy	63776	Engineer- ing/CS	1190	1.62	8224	22050
10	MCPHS University	60040	Health Pro	1190	1.19	18952	25000
11	MIT	59967	Engineer- ing/CS	1570	12.69	17152	13000
12	Colorado School of Mines	59802	Engineer- ing/CS	1400	2.74	14513	22000

13	Rensselaer Polytechnic Institute	59528	Engineer- ing/CS	1490	2.67	22181	26001
14	Babson College	59166	Business	1370	3.80	25090	26000
15	Milwaukee School of Engineering	58269	Engineer- ing/CS	1260	1.46	13768	25000
16	Resurrection University	58127	Health Pro			15362	20000
17	SUNY Maritime College	58038	Business	1190	1.90	5810	20157
18	Bentley University	57437	Business	1330	2.17	20489	25821
19	Bucknell University	57123	Social Sci	1400	3.26	31201	26467
20	Chamberlain University-Illinois	56989	Health Pro		1.51	12589	17294
21	Lehigh University	56184	Engineer- ing/CS	1410	2.92	19598	22250
22	Bryant University	56177	Business		1.33	22952	25279
23	Duke University	56059	Social Sci	1550	8.76	13488	6500
24	Oregon Health & Science University	55842	Health Pro			13304	10939
25	Worcester Polytechnic Institute	55768	Engineer- ing/CS		2.28	23627	27000
26	Claremont McKenna College	55306	Social Sci	1520	9.29	29500	11500
27	University of the Sciences	54691	Health Pro	1260	1.73	15543	26935
28	Thunderbird School of Global Management	54158	Business			33374	8000
29	Oregon Inst of Technology	53881	Health Pro	1150	1.67	7051	15917
30	Cornell University	53530	Engineer- ing/CS	1510	7.05	18007	10912
31	University of Pennsylvania	53417	Business	1550	9.64	21153	12018
32	Georgia Tech - Main Campus	53210	Engineer- ing/CS	1490	3.00	8973	20500
33	Carnegie Mellon University	52525	Engineer- ing/CS	1540	4.07	22402	24827

APPENDIX K: HIGHEST-EARNERS-PRODUCING GRADUATE SCHOOLS FOR 75TH PERCENTILE EARNERS 10 YEARS AFTER UNDERGRADUATE ENTRY

#	INSTNM	EARN 75_P10	TUITFTE	Adm rate 1/n	EngCS Tech Prop
1	Harvard University	166800	30239	16.8	0.09
2	MIT	164900	28992	12.7	0.62
3	Stanford University	164000	24668	19.6	0.29
4	Yale University	153300	15008	15.9	0.08
5	University of Pennsylvania	151600	34560	9.6	0.13
6	Maine Maritime Academy	150200	16862	1.3	0.69
7	California Institute of Technology	142400	16533	11.3	0.54
8	Columbia U in the City of New York	137600	33371	14.4	0.21
9	Dartmouth College	136800	29358	8.7	0.13
10	Duke University	134500	22556	8.8	0.17
11	Princeton University	132100	13255	13.4	0.20
12	Carnegie Mellon University	125400	36119	4.1	0.38
13	Massachusetts Maritime Academy	124300	8830	1.6	0.57
14	Cornell University	120500	25066	7.0	0.22
15	Tufts University	119000	29681	5.8	0.14
16	Colorado School of Mines	118300	18760	2.7	0.91
17	University of Notre Dame	118300	22629	4.7	0.15
18	Cali State Univ Maritime Academy	117400	7649	1.5	0.55
19	University of the Pacific	116500	30355	1.8	0.15
20	DigiPen Institute of Technology	115000	20838	1.5	0.60
21	University of Southern California	113200	30675	5.5	0.12
22	Lehigh University	112600	26103	2.9	0.38
23	SUNY Maritime College	110300	6399	1.9	0.41
24	Northwestern University	109800	28395	7.6	0.16
25	Case Western Reserve University	109600	22634	2.6	0.36