Rating Colleges: Using the College Scoreboard Data in Introductory Statistics

Christina Wang and Fei Wang

December 15, 2015

1. Abstract

We use the publicly-available datasets from College Scorecard (compiled by the Department of Education) to allow instructors to use large dataset to bring interesting questions into their introductory statistics courses. We outline useful strategies to introduce this rich dataset with descriptive statistics and univariate visualizations. We set forth several statistical questions that shed light on relationships of interest, then describe how students could answer them using simple multivariate data analysis techniques. The goal of this case study is to familiarize students with statistical concepts and language that are crucial to describing trends and relationships between variables. In this case study, we explore how costly higher education institutions are by analyzing their completion rates, median debt level of graduates, and their early-career earnings.

Keywords: student debt; college admissions; College Scorecard; data wrangling; visualization

2. Introduction

A college degree has two prices – declares a recent National Public Radio (NPR) news article (Bui 2015). According to the report, each college has a "sticker" price, or the price that is listed on brochures and websites, and a "net" price, which is the sticker price less any financial aid and scholarships. The net price is the price a college student actually pays to attend the college. The actual price of attending college may differ substantially for families in different income brackets. Amherst College, one of the elite liberal arts insitutions with a high price tag, is chosen as an example of an expensive college that is affordable for lower-income families (Bui 2015).

We can view the *value* of a college as its benefits less its costs. From a purely monetary perspective, the most straight-forward and visible measure of the benefits of a college education is a graduate's expected earnings, and the main cost of going to college is the net price. Another way to measure the cost of college is to calculate how much debt a family has to take out in order to support a student. We believe that debt level is a good proxy for the cost of college because it represents a student and their family's willingness to forgo future earnings in exchange for an education. We are interested in exploring the relationship between college debt and early career earnings as the main factors in the cost-benefit analysis for college education.

In addition to investigating the relationships between debt and earnings, we also want to understand how these relationships differ for different types of institutions. Specifically, we analyze whether "selectivity" (determined by both admission rate and completion rate) and "governance" (whether a college is public, private non-profit, or private for-profit) have an impact on how students choose to take on debt and how they perform in terms of income after completing college. In order to do so, we first select one year of data to explore at the possible interactions between these variables. We chose data from 2011 because it is the most recent year that income and earnings data is available. Please refer to Appendix for instructions to access data of other years.

2.1. Data

Analyses like NPR's were made possible because of ongoing effort at the U.S. Department of Education (DOE) to provide transparency and comprehensive information on U.S. higher education institutions to students and their families. The College Scorecard is an interactive site the Department of Education unveiled in

2013 that aims to "provide students and families the critical information they need to make smart decisions about where to enroll for higher education. (DOE 2013)" It is designed to help students and families evaluate colleges, or as President Obama put in his 2013 State of the Union address, to get the most "bang for your educational buck."

In September 2015, the Department of Education released the data behind the interactive tools on College Scorecard, including supporting data on student completion, debt, repayment, employment and earnings, etc. These datasets are freely downloadable through the U.S. Department of Education's College Scoreboard website (https://collegescorecard.ed.gov/data/). Since its release, many blogs and news articles have provided guides on how to use and interpret the massive dataset. For instance, Vox Education has an easy-to-read introduction navigating College Scoreboard site (Nelson 2015). For readers unfamiliar with the data, College Scoreboard's interactive search is a good visualization tool to compare colleges.

2.2. Goals

In the sections below, we pose and answer statistical questions using univariate, bivariate, and multivariate analyses that can be addressed using these data. We hope to achieve the following goals:

- Shed light on whether the amount of loans taken during college years has any association with future earnings, and provide an innovative perspective for students considering taking federal loans;
- Provide basic computing framework and data visualization tools for students, so that they can apply to future explorations of data on their own; and
- Encourage introductory statistics course instructors to incorporate more relevant real-life data and examples that pertain to students' interests.

3. Univariate Analysis

We first look at some descriptive statistics for variables of interest. In this section, we examine national data first. Then, based on students' interests, instructors can narrow the scope of the analysis down to one region or state.

3.1. Median debt

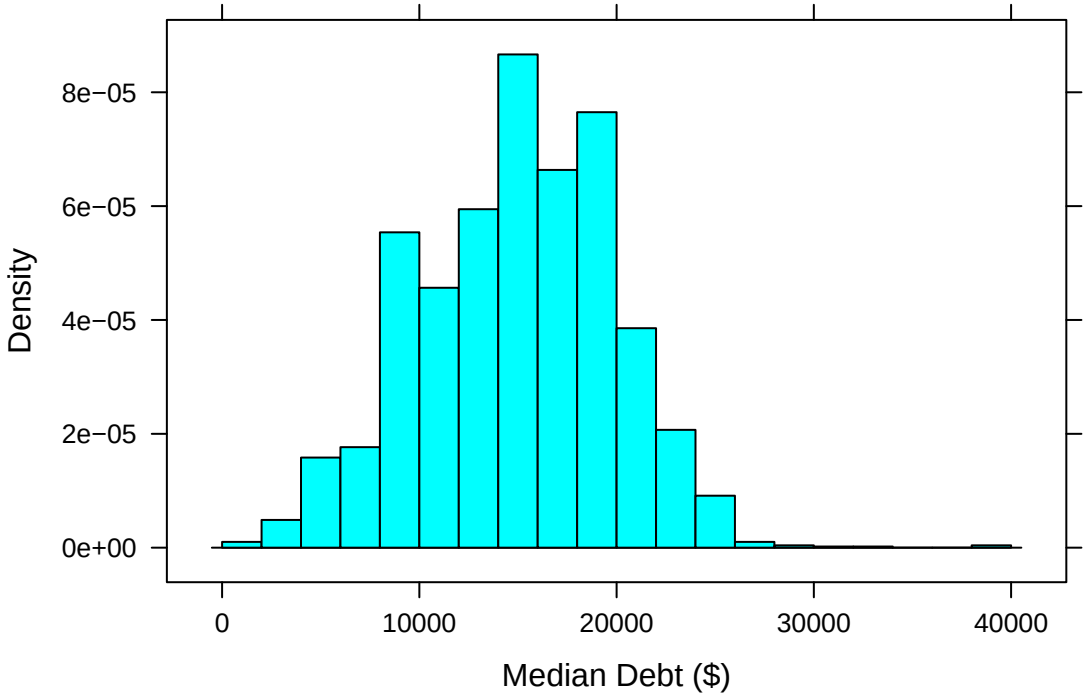
We investigate the distribution of median debt graduates take on at different colleges.

Teaching Tips: Data related to debt is only based on students who are on financial aid, and is not representative of the entire student population.

min	Q1	median	Q3	max	mean	sd	n	missing
1000	11748	15000	18750	39912	14968	4936	2464	381

Table 1: Summary Statistics for Median Debt

Figure. 1: Median Debt Level for College Students with Financial Aid



Summary statistics for the median debt (variable `DEBT_MDN_SUPP`) level for individual colleges across the United States are displayed in Table 1. Figure 1 shows that very few colleges have students who graduate with an extraordinarily high debt level (median debt $>$ 25,000). So what are those schools with very high debt level?

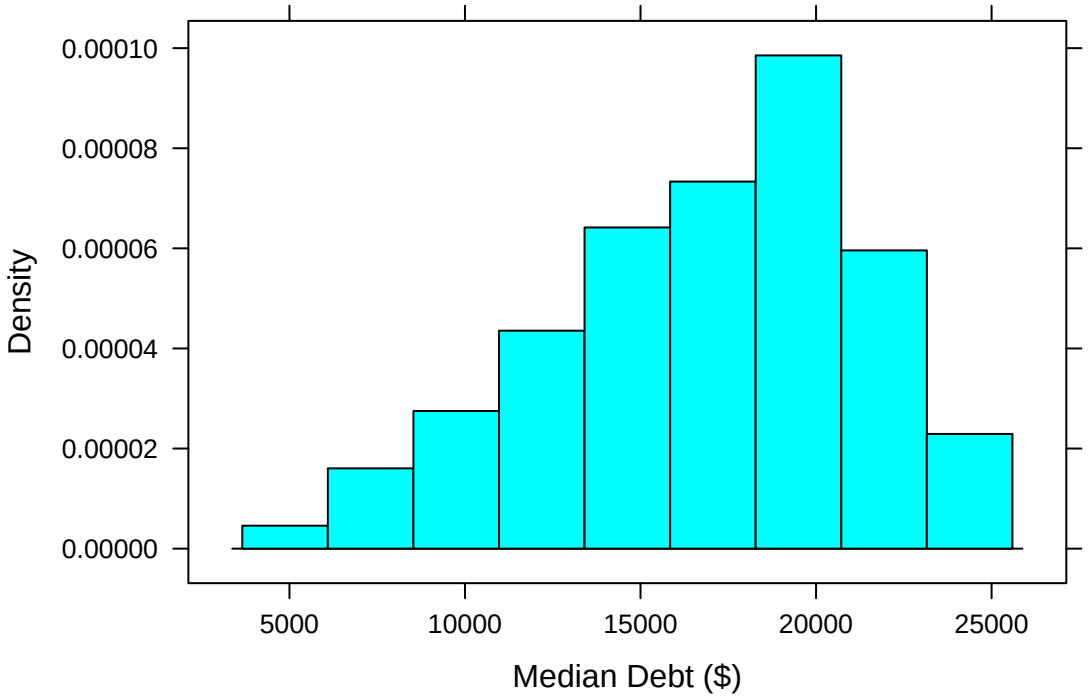
	INSTNM	DEBT_MDN_SUPP
1	Platt College-Aurora	39912
2	Northwest College of Art	39581
3	Southern California Institute of Architecture	32500
4	Careers Unlimited	31660
5	Art Center College of Design	29031
6	Design Institute of San Diego	28750
7	Beulah Heights University	27796
8	Newschool of Architecture and Design	27500
9	Missouri Tech	27500
10	Brooks Institute	26246
11	Digital Media Arts College	26062
12	Woodbury University	25750
13	Fontbonne University	25750
14	California Institute of the Arts	25312
15	Grambling State University	25286
16	Charles R Drew University of Medicine and Science	25250
17	Expression College for Digital Arts	25188

Table 2: Outlier Colleges for Median Debt

Table 2 shows that a large number of these colleges are art, architecture, and design related institutions.

Teaching Tips: Does this pattern still hold when we shift our focus to a specific region in the U.S.? Instructors should feel free to use a subset of the full dataset by choosing a region according to students' interest. As such, the analysis will be more relevant to students. Here, we look at the pattern in New England as an example.

Figure 2: Median Debt Level for College Students in New England



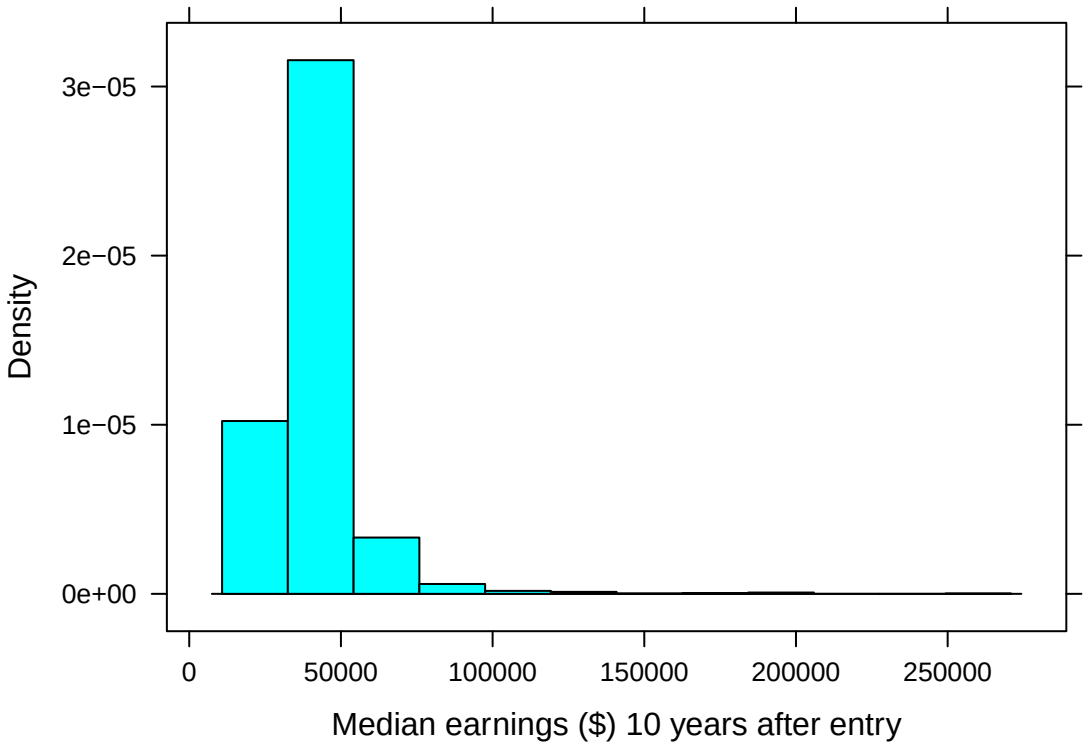
The median value of the Median Student Debt is $17,500 (which is higher than national median), the minimum is $5,500 and the maximum is $25,000. Figure 2 shows that in New England, the difference between the median debt in different schools can be as much as around $20,000.

3.2. Early career earnings (10 years after college entry)

The summary statistics and histogram below indicate that the median of early-career earnings is $39,000, but the maximum is $250,000, more than 6 times higher than the median. We also noticed that the graph has a very long right tail, indicating that there are quite a number of outliers (i.e., schools with students who earn exceptionally high earnings 10 years after school).

min	Q1	median	Q3	max	mean	sd	n	missing
11600	33700	39000	46500	250000	41518	15550	2370	475

Table 3: Summary Statistics for Median Earnings



Teaching Tips: Before displaying the results for outliers, ask students to identify if any interesting patterns in the graph merit further investigation (i.e., outliers). Ask students which types of institutions are more likely to have higher median earnings 10 years after entry, and then guide them to test their hypothesis.

A closer look at the schools with the highest earnings shows that all of them are medical or healthcare-related institutions.

	INSTNM	md_earn_wne_p10
1	Medical College of Wisconsin	250000
2	Albany Medical College	201200
3	A T Still University of Health Sciences	199600
4	West Virginia School of Osteopathic Medicine	198300
5	University of Massachusetts Medical School Worcester	184900
6	New York Medical College	169600
7	Rosalind Franklin University of Medicine and Science	166200
8	Philadelphia College of Osteopathic Medicine	148400
9	University of North Texas Health Science Center	132400
10	Baylor College of Medicine	129000

Table 4: Outliers: Colleges with High Early-Career Earnings in the U.S.

Does this trend apply to New England? We are interested in knowing if the highest-earning institutions in New England also mostly consist of medical schools.

Not surprisingly, we find that the college with the highest early-career median earnings is the Univeristy of Massachusetts Medical School, Worcester. Within the top five high-earnings colleges, three are medical-related schools and the other two are MIT and Harvard. It is understandable because besides doctors, graduates from the Ivy's or those who come from a technical background are also popular job candidates, and therefore more

	INSTNM	md_earn_wne_p10
1	University of Massachusetts Medical School Worcester	184900
2	Massachusetts College of Pharmacy and Health Sciences	116400
3	Massachusetts Institute of Technology	91600
4	New England College of Optometry	88500
5	Harvard University	87200
6	Babson College	85500
7	Rensselaer Hartford Graduate Center Inc	81700

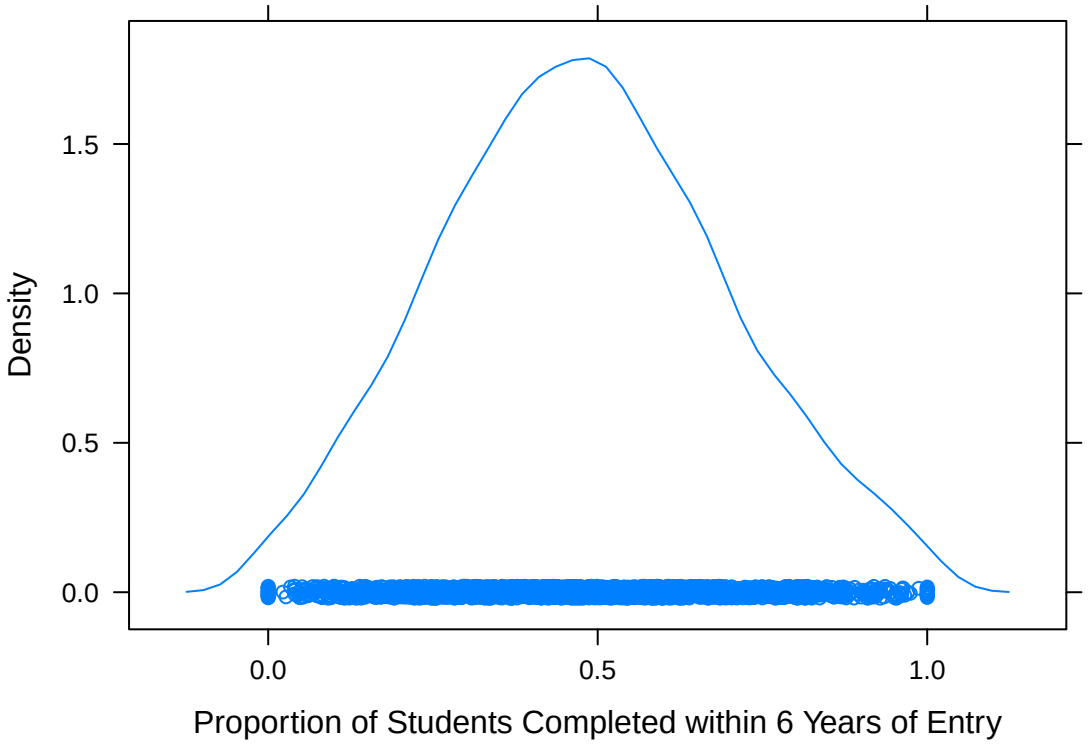
Table 5: Outliers: Colleges with High Early-Career Earnings in New England

likely to get a higher-paying job. Although Babson College is a liberal arts college, it focuses on business and entrepreneurship; hence, it's reasonable that its graduates tend to earn more.

3.3. Completion rate

As Figure 4 demonstrates, the nationwide college completion rate is approximately normally distributed, with a mean of approximately 0.5.

Figure 4: Density Plot of Completion Rate in the U.S.



What are the institutions that have extremely low (0%) or extremely high completion rate (100%)? Below we filtered institutions with either 0% or 100% completion rate, and listed 10 examples from each category. Due to limited space, we do not discuss the possible reasons behind these outliers. Instructors can encourage students to explore the characteristics that made these colleges so "successful" (or "unsuccessful") in producing graduates by analyzing other variables such as `selectivity`.

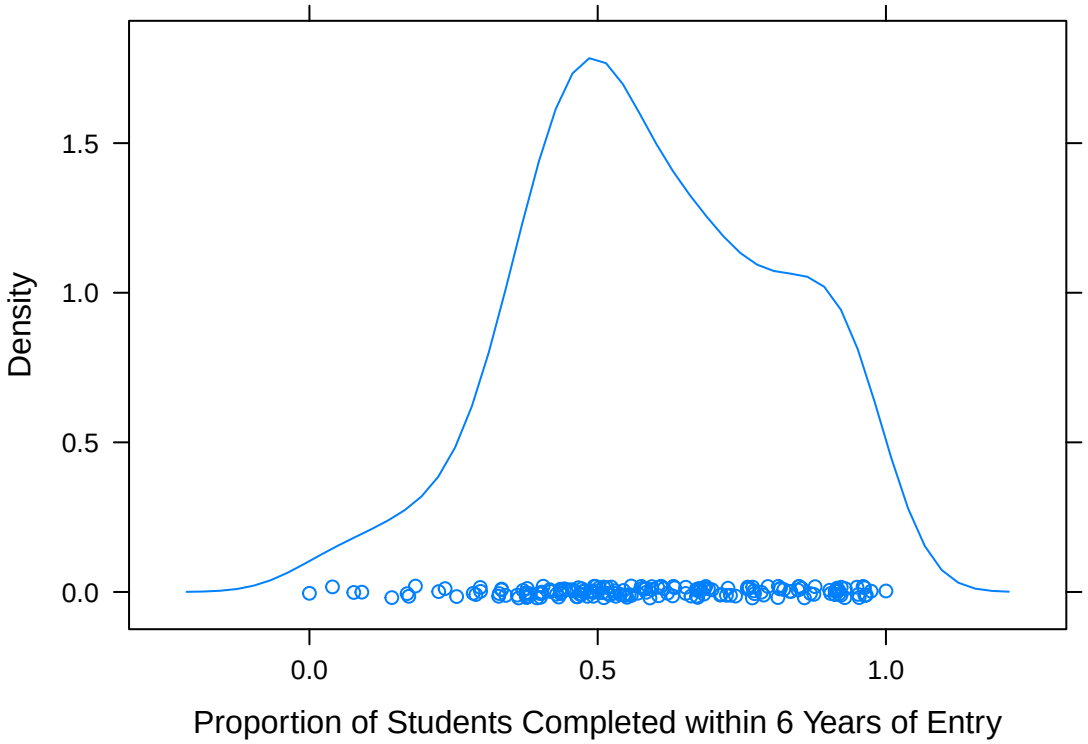
Teaching Tips: Ask if colleges in the region where students come from perform better or worse than the national average in terms of undergraduate completion rate.

	INSTNM	C150_4
1	Heritage Christian University	0.00
2	Charles R Drew University of Medicine and Science	0.00
3	National American University-Colorado Springs	0.00
4	Bangor Theological Seminary	0.00
5	Sacred Heart Major Seminary	0.00
6	Cleveland Chiropractic College	0.00
7	Kol Yaakov Torah Center	0.00
8	Mesivta of Eastern Parkway-Yeshiva Zichron Meilech	0.00
9	Talmudical Institute of Upstate New York	0.00
10	Yeshivath Zichron Moshe	0.00
11	Lincoln University	1.00
12	Central Baptist Theological Seminary	1.00
13	Laboure College	1.00
14	Barnes-Jewish College Goldfarb School of Nursing	1.00
15	Carolina Christian College	1.00
16	Pontifical College Josephinum	1.00
17	Tri-State Bible College	1.00
18	American Baptist College	1.00
19	Virginia University of Lynchburg	1.00
20	Universidad Teologica del Caribe	1.00

Table 6: Completion Rate Outliers

Again, we use New England to illustrate. The distribution of the completion rate (shown in Figure 5 below) in New England area is bi-modal, peaking at 0.5 and 0.8. The distribution skews towards right and hence has a long tail on the left hand side. This shows that a majority of first-time, full-time students at four-year institutions in New England will complete the study within six years. This is a good example to demonstrate that national trend or distribution may not necessarily represent that on a more regional level.

Figure 5: Density Plot of Completion Rate in New England



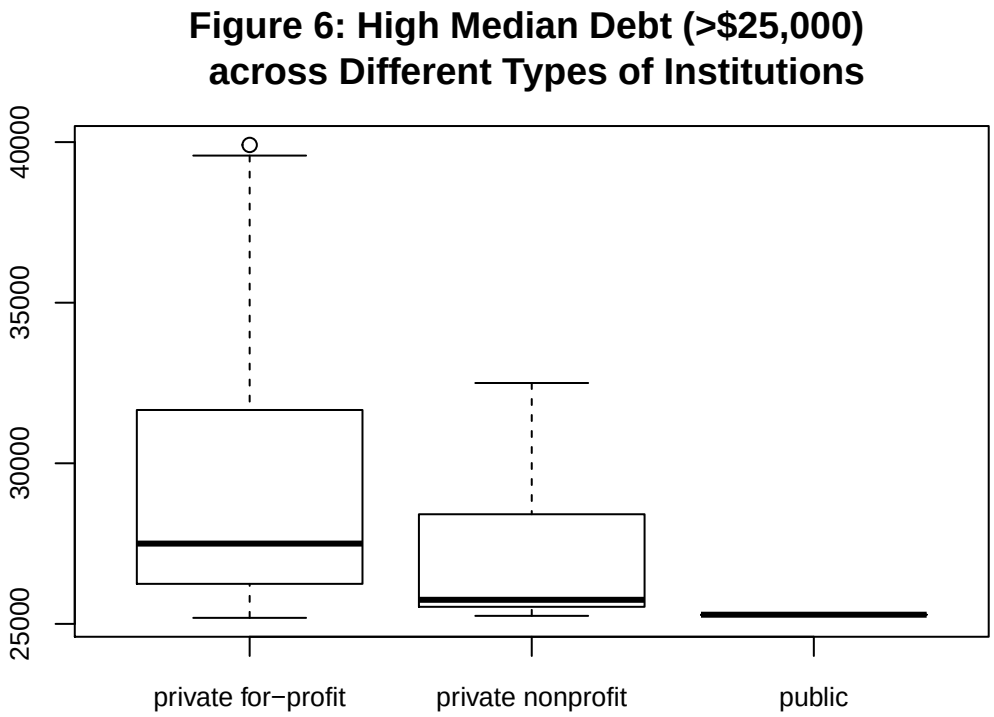
4. Bivariate Analysis

After understanding the distribution of each of the variables of interest, we move on to explore the relationships between different variables. First, we compare debt levels aross different types of institutions.

Teaching Tips: Ask students to identify the colleges with highest debt levels; encourage them to pinpoint the characteristics (attributes of variables) those colleges are likely to be associated with.

	INSTNM	DEBT_MDN_SUPP	CONTROL
1	Platt College-Aurora	39912	private for-profit
2	Northwest College of Art	39581	private for-profit
3	Southern California Institute of Architecture	32500	private nonprofit
4	Careers Unlimited	31660	private for-profit
5	Art Center College of Design	29031	private nonprofit
6	Design Institute of San Diego	28750	private for-profit
7	Beulah Heights University	27796	private nonprofit
8	Newschool of Architecture and Design	27500	private for-profit
9	Missouri Tech	27500	private for-profit
10	Brooks Institute	26246	private for-profit
11	Digital Media Arts College	26062	private for-profit
12	Woodbury University	25750	private nonprofit
13	Fontbonne University	25750	private nonprofit
14	California Institute of the Arts	25312	private nonprofit
15	Grambling State University	25286	public
16	Charles R Drew University of Medicine and Science	25250	private nonprofit
17	Expression College for Digital Arts	25188	private for-profit

Table 7: Outliers: Colleges with High Median Debt



We find that, among the colleges with extremely high median debt levels, a large number of them are for-profit colleges.

4.1. Median earnings 10 years after entry vs. Median debt

By introducing bivariate analysis, we will give students tools to answer more interesting and complex statistical questions, such as how current debt level might affect future earnings? To begin, we look at some descriptive statistics of earnings and student debt.

The range of Median earnings 10 years after entry is (11,600, 250,000), with a median of $39,000. The range

min	Q1	median	Q3	max	mean	sd	n	missing
1000.00	18593.00	22750.00	25000.00	47186.50	21866.90	6848.45	2417	428

Table 8: Summary Statistics for Median Debt of Graduates

min	Q1	median	Q3	max	mean	sd	n	missing
11600.00	33700.00	39000.00	46500.00	250000.00	41518.35	15549.89	2370	475

Table 9: Summary Statistics for Median Earnings

of median debt is (1,000, 47,186), with a median of $22,750.

Figure 7: Median earnings 10 years after entry vs. Median Debt

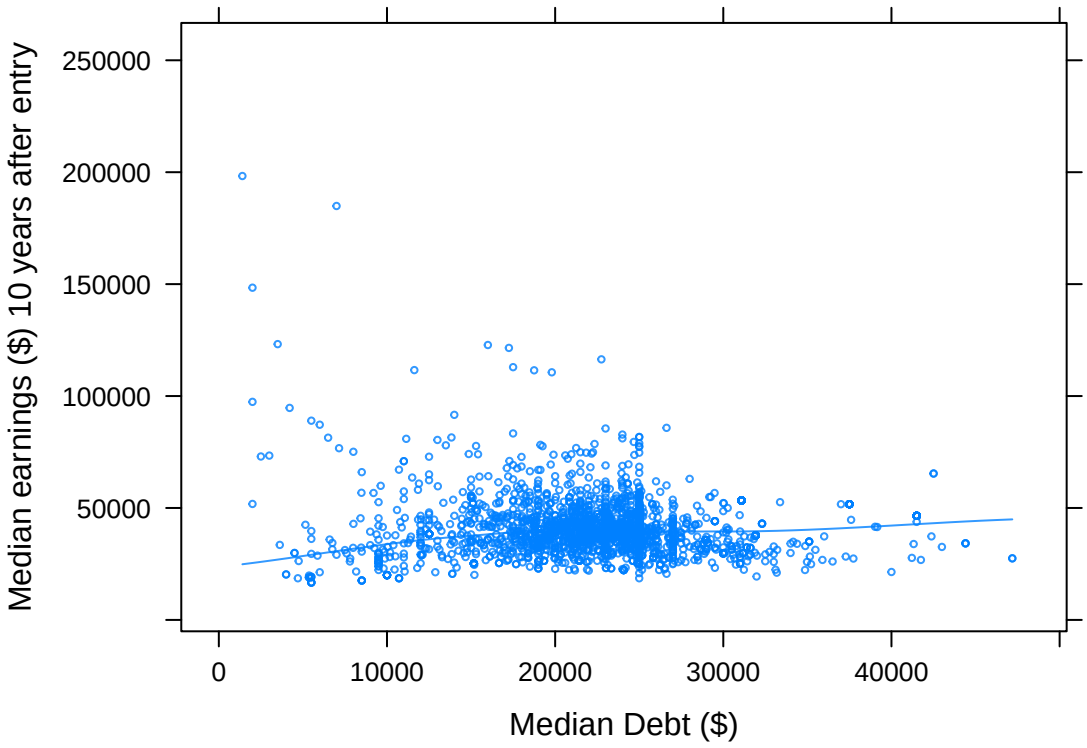
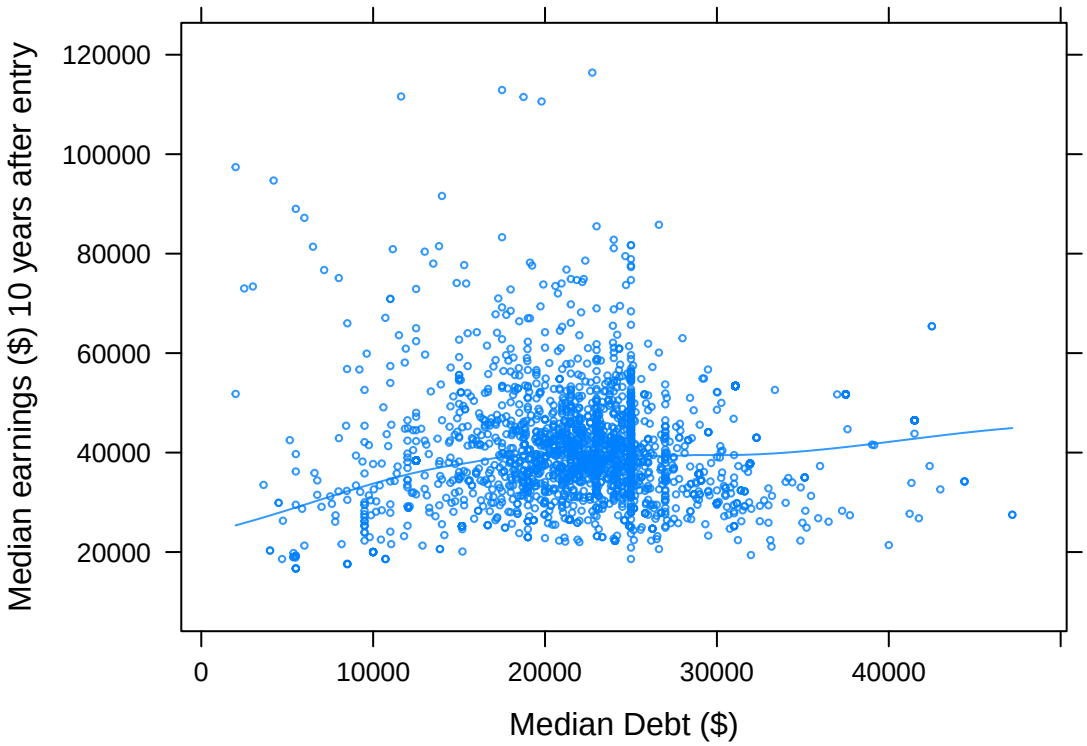


Figure 8: Median earnings 10 years after entry vs. Median Debt



Both graphs show a general trend that debt is not strongly associated with earnings 10 years after college entry. However, after removing outliers (institutions with extremely high median earnings), it seems that earnings and debt are weakly but positively correlated.

Teaching Tips: Ask students what other patterns they find interesting (hint: a vertical line connected by a series of data points all with median debt = $25,000). Ask them to identify the institutions involved by filtering through the dataset.

From the list, we notice that different schools have the same `GRAD_DEBT_MDN`. Ask students to suggest some explanations.

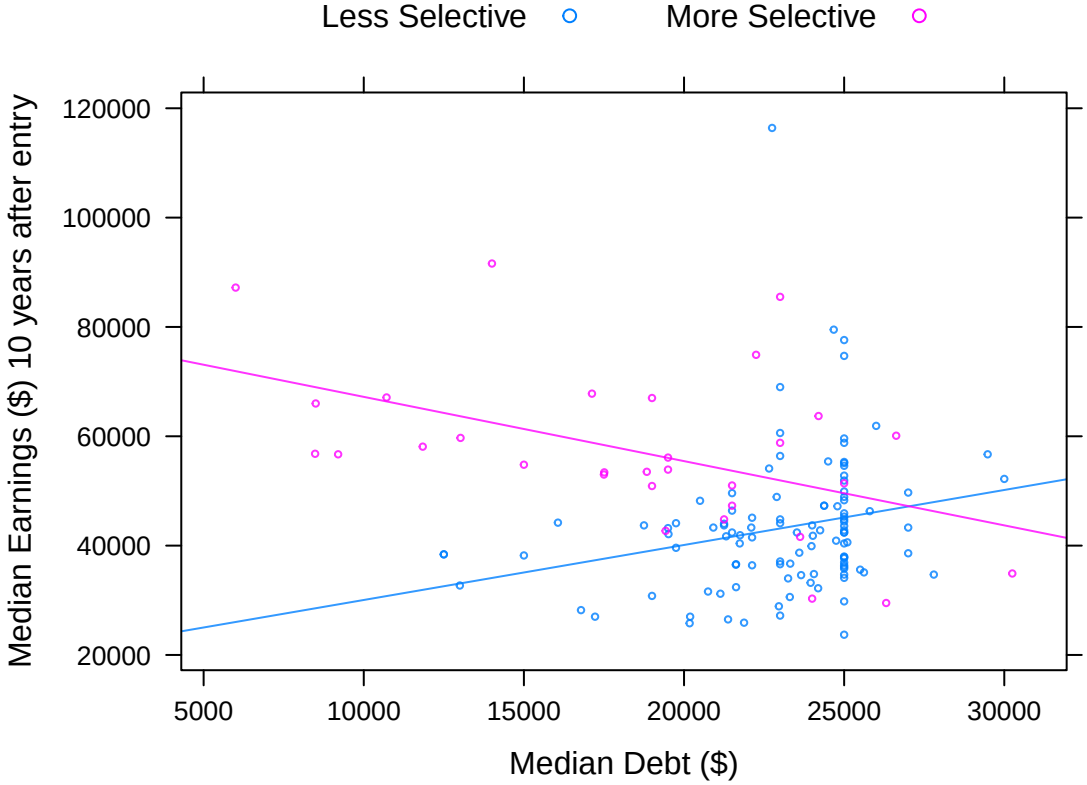
5. Multivariate Analysis

Next, we consider if the relationships revealed in the previous analysis differ for different types of institutions.

5.1. Median Earnings 10 years after entry vs. Median debt across colleges with different admissions rate

How would the admission selectivity of the schools affect future earnings, given that peers in another type of school have the same amount of student debt? In this case, since the underlying relationship is quite complicated and may vary across different regions, we only examine the trends in New England. We group colleges into two categories based on admission selectivity; colleges whose admission rate is less than or equal to 50% are defined as "more selective", and the rest as "less selective".

Figure 9: Median Earnings vs. Median Debt by Selectivity



Interestingly, the "return on investment" seems to go different directions for colleges with high and low admission rate. For the more selective colleges, the median earnings decrease as median debt level increases. For less selective colleges, however, the opposite trend seems to prevail (i.e., there is a positive correlation between median earnings and median debt).

Overall, college graduates who attended more selective institutions tend to have higher median earnings. For the same level of median debt, people from more competitive institutions tend to earn more 10 years after they first enter college, especially when the median debt is on the lower end. However, when median debt level increases, the difference in early-career earnings narrows.

Teaching Tips: After showing sample analyses to students, ask them what kinds of variables they may be interested to explore, and how they want to present the data visualization. Show them the work from other student projects as inspiration. A good source to consult: https://www.kaggle.com/c/us-dept-of-education-college-scorecard/scripts

6. Conclusion

The College Scorecard data allows the exploration of median debt, early career earnings, and completion rate of college students on financial aid. Whether on a national or regional level, the median of the median debt is around $15,000. However, students (on financial aid) from for-profit as well as arts and design institutions usually bear much higher debt. Healthcare-related professionals and graduates from top research universities top the early career earnings rankings.

We have also explored the relationship between median earnings and median debt among graduates. Overall, we did not observe a significant relationship on a national level. However, given the same level of median debt, public and private non-profit institutions students usually earn more than their peers in private for-profit institutions, although earnings and debt level seem to be positively correlated among private for-profit institutions. As for admission rate, we found that, in New England area, more selective colleges' median

earnings are negatively correlated with the median debt level of its graduates; a positive correlation between median earnings and median debt is observed for less selective colleges.

In addition, a student's choice of major also matters in determining the return on educational investment. Even though both medical schools and arts institutes are expensive and their students usually incur a high level of debt, the future earnings prospects for graduates from med school vis-à-vis those from art school differ quite substantially. Therefore, we suggest that high school seniors and parents factor the above findings into consideration when deciding where to spend the next four years.

7. Acknowledgement

We would like to thank the U.S. Department of Education for making the dataset of CollegeScorecard available. We would also like to express our appreciation for Nicholas Horton for his help and guidance on earlier drafts of this manuscript.

8. Codebook

We extracted the following variables from the master dataset and stored them in a new dataset called df2011:

1) INSTNM: Institution name (string)
2) ZIP: ZIP code (integer)
3) st_fips: FIPS code (integer); five-digit Federal Information Processing Standard (FIPS) code which uniquely identifies counties and county equivalents in the United States, certain U.S. possessions, and certain freely associated states.
4) region: Region (IPEDS); 1 = New England (CT, ME, MA, NH, RI, VT), 2 = Mid East (DE, DC, MD, NJ, NY, PA), 3 = Great Lakes (IL, IN, MI, OH, WI), 4 = Plains (IA, KS, MN, MO, NE, ND, SD), 5 = Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV), 6 = Southwest (AZ, NM, OK, TX), 7 = Rocky Mountains (CO, ID, MT, UT, WY), 8 = Far West (AK, CA, HI, NV, OR, WA), and 9 = Outlying Areas (AS, FM, GU, MH, MP, PR, PW, VI) [Will omit 9 in our final analysis]
5) CONTROL: Control of institution; 1 = public, 2 = private non-profit, and 3 = private for-profit
6) DEBT_MDN_SUPP: Median debt, data suppressed for institutions where sample size n is less than 30
7) GRAD_DEBT_MDN: Median debt for students who have completed; range (1000, 47190)
8) WDRAW_DEBT_MDN: Median debt for students who have not completed; range (1678, 29640)
9) GRAD_DEBT_N: The number of students in the median debt completers cohort (integer); range (0, 80140)
10) WDRAW_DEBT_N: The number of students in the median debt withdrawn cohort (integer); range (0, 291000)
11) mn_earn_wne_p10: Mean earnings of students working and not enrolled 10 years after entry (all earnings integer)
12) md_earn_wne_p10: Median earnings of students working and not enrolled 10 years after entry
13) pct10_earn_wne_p10: 10th percentile of earnings of students working and not enrolled 10 years after entry
14) pct25_earn_wne_p10: 25th percentile of earnings of students working and not enrolled 10 years after entry
15) pct75_earn_wne_p10: 75th percentile of earnings of students working and not enrolled 10 years after entry
16) pct90_earn_wne_p10: 90th percentile of earnings of students working and not enrolled 10 years after entry

Some control variables to consider:

- ADM_RATE: Admission rate; range (0.0626, 1.0000)
- PCTPELL: Percentage of undergraduates who receive a Pell Grant; range (0, 1)
- C150_4: Completion rate for first-time, full-time students at four-year institutions (150% of expected completion time); range (0, 1)

** Note (for GRAD_DEBT_N)

- For some public universities (in this case, University of Phoenix), the number of students with debt is calculated for all campuses and then this value is assigned to each of the individual campuses. Hence, the value of their GRAD_DEBT_N is particularly high compared to others. If this variable is chosen, instructors may advise students to exclude these values.

Appendix A: Data sources

The complete dataset is available as CollegeScorecard_Raw_Data.zip at https://collegescorecard.ed.gov/data/, including comma-separated value Excel files ranging from 1996 to2013 (e.g. MERGED1996_PP.csv), data dictionary and so on. Several other more specific data sets are also available for downloading. For our case study, we use the raw data from MERGED2011_PP.csv and consult CollegeScorecardDataDictionary-09-12-2015.pdf.

The filtered dataset based on the raw dataset of MERGED2011_PP.csv. is stored in files `df2011.rda` and `df2011.csv`.

Appendix B: R Markdown

```
## Load data from derived
load("df2011.rda")
```

```
## Set up xtable
options(xtable.comment = FALSE)

## Descriptive statistics of median debt
debtX <- xtable(favstats(~ DEBT_MDN_SUPP, data = df_full),
                caption = "Summary Statistics for Median Debt",
                digits=0)
print(debtX)

histogram( ~ DEBT_MDN_SUPP, data = df_full, xlab = "Median Debt ($)",
           main = "Figure. 1: Median Debt Level for College Students \nwith Financial Aid",
           width = 2000, center = 1000)
```

```
# Outlier colleges with very high debt level
debtOl <- df_full %>%
  select(INSTNM, DEBT_MDN_SUPP) %>%
  filter(DEBT_MDN_SUPP > 25000) %>%
  arrange(desc(DEBT_MDN_SUPP))

## Debt outlier output
options(xtable.comment = FALSE)
debtOlX <- xtable(debtOl, caption = "Outlier Colleges for Median Debt", digits=0)
print(debtOlX)
```

```
## Filter & analyze colleges in New England region
df_ne <- df_full %>%
  filter(region == 1)
histogram( ~ DEBT_MDN_SUPP, data = df_ne, xlab = "Median Debt ($)",
           main = "Figure 2: Median Debt Level for College Students \n in New England")
```

```
## Descriptive statistics of median earnings
options(xtable.comment = FALSE)
earningsX <- xtable(favstats(~ md_earn_wne_p10, data = df_full), digits=0,
                    caption = "Summary Statistics for Median Earnings")
print(earningsX)

## Histogram of median earnings
histogram( ~ md_earn_wne_p10, data = df_full,
           xlab = "Median earnings ($) 10 years after entry",
           main = "Figure 3: Median Early-Career Earnings Students")
```

```
## Top earnings - national
earningsHi <- df_full %>%
  select(INSTNM, md_earn_wne_p10) %>%
  filter(md_earn_wne_p10 >13000) %>%
  arrange(desc(md_earn_wne_p10))

earningsHiX <- xtable(head(earningsHi, n = 10),
                      caption = "Outliers: Colleges with High Early-Career Earnings in the U.S.",
                      digits=0)
print(earningsHiX)
```

```
## Top earnings - New England
earningsNe <- df_ne %>%
  select(INSTNM, md_earn_wne_p10) %>%
  filter(md_earn_wne_p10 > 80000) %>%
  arrange(desc(md_earn_wne_p10))

earningsNeX <- xtable(head(earningsNe, n=10),
                      caption = "Outliers: Colleges with High Early-Career Earnings in New England",
                      digits=0)
print(earningsNeX)
```

```
densityplot(df_full$C150_4, na.rm = TRUE,
            xlab="Proportion of Students Completed within 6 Years of Entry",
            main="Figure 4: Density Plot of Completion Rate in the U.S.")
# 150% of expected completion time
```

```
## Outliers for national completion rate
completionLo <- df_full %>%
  select(INSTNM, C150_4) %>%
  filter(C150_4 == 0)

completionHi <- df_full %>%
  select(INSTNM, C150_4) %>%
  filter(C150_4 == 1)
```

```
completionOl <- rbind(completionLo[1:10,],completionHi[1:10,])
xtable(completionOl, caption = "Completion Rate Outliers")
```

```
densityplot(df_ne$C150_4, na.rm = TRUE,
            xlab="Proportion of Students Completed within 6 Years of Entry",
            main="Figure 5: Density Plot of Completion Rate in New England")
```

```
## Outliers in Median Debt (nationwide)
outliers <- filter(df_full, DEBT_MDN_SUPP > 25000) %>%
  select(INSTNM, DEBT_MDN_SUPP, CONTROL) %>%
  mutate(CONTROL=ifelse(CONTROL==1, "public",
                        ifelse(CONTROL==2, "private nonprofit", "private for-profit"))) %>%
  arrange(desc(DEBT_MDN_SUPP))

xtable(outliers, digits = 0, caption = "Outliers: Colleges with High Median Debt")

boxplot(DEBT_MDN_SUPP ~ CONTROL, data=outliers, par(cex.axis=0.8),
        main="Figure 6: High Median Debt (>$25,000) \n across Different Types of Institutions")
```

```
## Descriptive statistics on median debt & early-career median earnings
options(xtable.comment = FALSE)
debtX <- xtable(favstats(GRAD_DEBT_MDN, data=df_full),
                caption = "Summary Statistics for Median Debt of Graduates")
earnX <- xtable(favstats(md_earn_wne_p10, data=df_full),
                caption = "Summary Statistics for Median Earnings")
print(debtX)
print(earnX)
```

```
## Median earning vs. median debt
xyplot(md_earn_wne_p10 ~ GRAD_DEBT_MDN, auto.key=list(columns=3),
       data=df_full, type = c("p", "smooth"), lwd=1, alpha=0.8, cex=0.4,
       main="Figure 7: Median earnings 10 years after entry vs. Median Debt",
       xlab="Median Debt ($)",
       ylab="Median earnings ($) 10 years after entry"
       )
```

```
## Remove outliers and re-analyze
df_full1 <- df_full %>%
  filter(md_earn_wne_p10 < 120000)

xyplot(md_earn_wne_p10 ~ GRAD_DEBT_MDN, auto.key=list(columns=3),
       data=df_full1, type = c("p", "smooth"), lwd=1, alpha=0.8, cex=0.4,
       main="Figure 8: Median earnings 10 years after entry vs. Median Debt",
       xlab="Median Debt ($)",
       ylab="Median earnings ($) 10 years after entry"
       )
```

```
outliers <- filter(df_full, GRAD_DEBT_MDN == 25000) %>%
  select(INSTNM, GRAD_DEBT_MDN, CONTROL) %>%
  arrange(INSTNM)

# xtable(outliers, caption = "Colleges with Median Debt 25000", digits=0)
```

```
df_admit <- df_ne %>%
  filter(ADM_RATE > 0, na.rm=TRUE) %>%
  filter(GRAD_DEBT_MDN <= 32000) %>%
  mutate(selectivity = ifelse(ADM_RATE > 0.5, "Less Selective", "More Selective"))


xyplot(md_earn_wne_p10 ~ GRAD_DEBT_MDN, groups = selectivity, auto.key=list(columns=2),
       data=df_admit, type = c("p", "r"), lwd=1, alpha=0.8, cex=0.4,
       main="Figure 9: Median Earnings vs. Median Debt by Selectivity",
       xlab="Median Debt ($)",
       ylab="Median Earnings ($) 10 years after entry"
       )
```

References

Bui, O. (2015), "What You'll Actually Pay At 1,550 Colleges," *NPR Planet Money*, Available at http://www.npr.org/sections/money/2015/09/30/444446022/what-youll-actually-pay-at-1-550-colleges.

Department of Education (DOE) (2013), "Education Department Releases College Scorecard to Help Students Choose Best College for Them," *U.S. Department of Education Website*, Available at http://www.ed.gov/news/press-releases/education-department-releases-college-scorecard-help-students-choose-best-college-them.

Nelson, L. (2015), "The government's new College Scorecard helps you pick a college. Here's how to use it," *Vox Education*, Available at http://www.vox.com/2015/10/26/9617010/college-scorecard-how-to.

Rothwell, J. (2015), "Understanding the College Scorecard," *Brookings Opinion*, Available at http://www.brookings.edu/research/opinions/2015/09/28-understanding-the-college-scorecard-rothwell.

Stone, A. (2015), "Weekly Address: A New College Scorecard," *The White House Home Blog*, Available at https://www.whitehouse.gov/blog/2015/09/12/weekly-address-new-college-scorecard.