A True Lie about Reed College: U.S News Ranking

Reed College

Huaying Qiu, Wenxin Du, Bailee Cruger

April 28th, 2019

Acknowledgements

Our sincere gratitude goes to Mike Tamada, Ellen Kotler at the Office of Institutional Research at Reed College and Professor Kelly McConville at the Department of Mathematics at Reed College. Without them, this project would not have been possible.

Table of Contents

List of Tables

List of Figures

Abstract

The annual Best College Rankings published by U.S. News & Worlds Report are held by many as the prominent college rankings to consult with. However, the weight-and-sum model used has long been criticized for not reflecting the true educational quality of institutions. A few institutions, such as Reed College in 1995, announced refusal to continue participating in the USNWR college rankings. It's claimed that the ranks of the non-reporting institutions then penalized and deliberately under-ranked. This research used Principal Component Regression and Elastic Net Regularized Regression methods to build predictive models, aiming to reproduce the USNWR College Rankings published in 2009 and 2019 and assess if non-reporting schools are truly under-ranked. As a result, even thought no systematic under-rank of non-reporting institutions was found, Reed College was shown to be the only institution significantly under-ranked by USNWR both in 2009 and 2019.

Chapter 1

Introduction

U.S. News & World Report (USNWR) Best Colleges Ranking has long been held by many as the prominent ranking to consult regarding educational quality of universities and liberal arts colleges in the United States. In 2015 as survey on college freshmen by the UCLA's Higher Education Research Institute showed that 20% of college freshmen considered college ranking important to there choice in school (Eagan et al., 2016). On the other hand, given the determinant effect of rankings on students' application process, most of the administrators of universities consider rankings as an important, if not essential, marketing tool to attract applications. For them, rankings are so important that they have no scruple about spending hundreds of millions of dollars for an increase in ranking (Grewal, Dearden, & Llilien, 2008). While ranking has such prominent influence on the behavior of both students and schools, numerous concerns and criticism of the ranking process arise and question the validity of the ranking system. It is even suggested that the weight-sum model of the ranking system is fundamentally flawed since with a weight-sum model, the statistical significance of the difference in rankings can not be tested (Clarke, 2004). Therefore, it is unclear how big a difference makes a school significantly better than another school. Moreover, multiple analyses confirm severe multicollinearity within the criteria used by U.S. News (Bougnol & Dulá, 2015). This makes it difficult to tell how much of an effect individual varibles have on the final score of schools. Concerned by the credibility of U.S. News' ranking system and with the belief that simple quantification should not and can not serve as a measure of education quality, Reed College quit the ranking in 1995 by refusing to fill out the survey from U.S. News, and has continued to mantain this practice over time. Overtime, some other schools, such as St. John's College in New Mexico, also chose to quit the ranking system. Bizarrely, after those schools' exit, U.S. News still keeps them on the list while their rank droped remarkably. It's claimed that non reporting institutions are penalized and deliberately under-ranked, However, after extensive searching we were unable to find a study that examined if this claim was true. The current study attempts to reproduce the USNWR ranking process and explicate the true rankings for non reporting institutions.

1.1 Backgroud

1.1.1 U.S. News & World Report Best College Ranking

U.S. News & World Report Best Colleges Rankings are published annually ever since 1983, with an exception of year 1984. Schools are grouped into different categories based on the Carnegie classification's, including groups such as Masters School, Law School, and undergraduate colleges such as liberal arts and national, then ranked against schools in their class. Schools that offer a complete range of undergraduate majors, master's and doctoral programs and emphasize faculty research are classified as national universities. Schools offering at least 50% of degrees for majors in arts and sciences and focus mainly on undergraduate education are classified as national liberal colleges. The ranking methodology used for schools are almost the same between categories with subtle variation between.

The majority of data used by USNWR are directly reported from institutions through a questionnaire. Questions on the questionnaire include both questions incorporated from the Common Data Set initiative and proprietary questions from USNWR. The questionnaire is sent out in spring each year. The returned information are then evaluated by USNWR and the ranking results are published in the following year. The published ranking thus does not reflect the current information of institutions. In fact, the ranking of universities that is published in 2019 uses data collected from institutions in spring 2018, which means the data used are really from year 2017 or academic year 2016 - 2017.

Not all schools respond to USNWR surveys, and some schools do not answer every single question. For the 2019 rankings 92% of ranked institutions returned the survey during the spring 2018 data collection window.USNWR checks this data against previous years and third party sources. USNWR uses external data sources for information they fail to get directly from schools including using publicly available data from the Council for Aid to Education and the U.S. Department of Education's National Center for Education Statistics (Morse, Brooks, & Mason, 2018). For schools that chooose not to report at all aditional sources such as the schools' own websites and/or data collected by USNWR from previous years will be used (Sanoff, 2007).

The collected data are then grouped as indicators for different aspects of academic success. Each indicator is assigned a specific weight in the ranking formula used by USNWR, the weights of all indicators add up to 100%, and a score between 0 to 100 is calculated for each institution using the ranking formula and data collected. Final ranking results are generated based on this score.

Weightings change frequently. For example, USNWR surveys the presidents, provosts and deans of each institution to rate the academic quality of peer institutions, and also surveys about 24,400 high school counselors to do the same rating. The results are combined and grouped into the indicator Expert Opinion, which currently takes 20% weight in the ranking formula. Back in 2018 the indicator Expert Opinion takes 22.5% of the weights, and back in 2009 it takes 25% of the weights. The indicator Outcomes in 2019 adds a subfactor Social mobility which takes 5% of the total weights and was not considered in rankings from previous years. The frequent changes in weightings make it hard to do direct comparison of rankings year by year, since they are calculated based on different formula. Nonetheless, popular press, high schoolers and parents do so and tend to consider changes in rankings as important information that represents changes of institutions' academic quality.

1.1.2 Non-reporters and Under Rank

In 1995, believing that the methodology used by USNWR is "fundamentally flawed", then-president of Reed College Steven Koblik announced the refusal to responding to USNWR's surveys. Though Reed College refused to continue participating in rankings, USNWR kept assigning ranks to Reed College, without information provided by the school through the annual questionnaire. The impartiality of such ranking is questioned by the school and others, stating that USNWR purposely assign lowest possible score for Reed College in certain indicators and “relegated the college to the lowest tier” (Lydgate, 2018), which led the rank of the college to drop from top 10 to the bottom quartile from 1995 to 1996.

Reed College is not the only school protest against USNWR rankings. St. John College decided to not participate in college ranking surveys and refused to provide college information since 2005. Similar to Reed College, the school is still included in USNWR ranking and is now ranked in the third tier. President of the institution Christopher B. Nelson once stated, “Over the years, St. John’s College has been ranked everywhere from third, second, and first tier, to one of the”Top 25" liberal arts colleges. Yet, the curious thing is: We haven’t changed. … ” (Nelson, 2007) Less discussion are found on whether the current rank of the school is reliable.

Most of the evidence up to this point on non-reporting schools being ranked lower is anecdotal. For Instance, in 2001, a senior administrator from Hobart and William Smith Colleges failed to report their current year data to USNWR, followed by a decrease in the rank of the school from the second tier to the third tier. (Ehrenberg, 2002) It’s said by the USNWR that they used data of the school from previous year instead in 2001 for Hobart and William Smith Colleges, which lead to understating of many of the current performance of the school. (Ehrenberg, 2002) On the website of Reed College, Chris Lydgate stated that in May 2014, in a presentation to the Annual Forum for the Association for Institutional Research, the director of data research for U.S. News Robert Morse revealed that if a college doesn’t fill out the survey, the guidebook arbitrarily assigns certain key statistics at one standard deviation below the mean. (Lydgate, 2018) Though no further evidence can be found beyond the website of Reed College, this statement has increased our motivation to investigate if and how non-reporting schools appear to be under ranked by USNWR.

1.1.3 Modeling on U.S. News Ranking

Many studies have been done to find the important factors that affect schools ranking on USNWR and how meaningful the rankings are. In one previous study, researchers developed a model based on the weighting system and methodology provided by USNWR to reproduce USNWR rankings, trying to understand effects of subfactors and assess significance of changes in ranking. The predictive model generated in the study perfectly predicted 21.39% of the college ranking, with errors all be within ± 4 differences for the rest. Further, as a result they found that up to ± 4 changes in rank are simply noise and, thus, meaningless. (Gnolek, Falciano, & Kuncl, 2014) Due to the multicollinearity within the criteria used by U.S. News, it is hard to tell which criterion has the largest effect on a school’s rank. To tackle this problem, one research used principal component analysis to examine the relative contributions of the ranking criteria for those national universities in the top tier that had reported SAT scores and found that actual contribution of each criterion differed substantially from the weights assigned by U.S News because of correlation among the variables. (Webster, 2001) Another research was conducted on the 2003 U.S. News business and education rankings. Using a technique called jackknifing, the researcher was able to conduct hypothesis tests, which otherwise would be impossible, on the weigh-sum model. The result was appalling. The difference of rankings between most educational institutions were statistically insignificant. (Clarke, 2004)

In this study, we use principal component regression and elastic net regression to build predicative models aiming to reproduce the results of rankings from U.S. News. Then we apply these two models to data of non-reporting schools collected from a database called College Scorecards maintained by U.S Department of Education. With this method, we attempt to assess if non-reporting schools are under-ranked. If so, what results in their being ranked lower.

Chapter 2

Data, Method, Result

2.1 Data

The project started out with two datasets provided by the Office of Institutional Research at Reed College. Both of the datasets are supposed to come directly from USNWR. They will be referred to later as original 2009 dataset and original 2019 dataset. The 2009 dataset contains 124 liberal arts colleges ranked by USNWR with 36 variables. The 2019 dataset contains 172 liberal arts colleges ranked by USNWR with 27 variables. The list of variables in both datasets is presented in Table [2.1]. Given the intention to determine if Reed College is under-ranked by USNWR, the original datasets present several challenges. For example, comparing the varaiable available in the original 2019 dataset and the ranking system of USNWR, Table [2.2], one can see that: (1) **Social mobility** is completely absent. (2) For **faculty resources**, all sub-criteria are absent. Instead, an encapsulating variable, **faculty resource rank**, is given. (3) Similar to **faculty resources**, **financial resources rank** is given instead of the variables contributing to **financial resources per student**, which, according to USNWR, should be a logarithmic transformation of the quotient of the sum of expenditures on instruction, academic support, student services and institutional support, and the number of full-time-equivalent students.

Although USNWR has a detailed description of the criteria and weight of its ranking system, its methodology of standardizing the overall scores so that they are all within the range of 0 to 100 remains untold. Besides, when it comes to non-reporting schools, the data in the datasets are not consistent with those published by schools themselves in their Common DataSet (CDS). For Reed College specifically, it is found that, for 2019, the percent of classes under 20 students, percent of freshmen in top 10% of high school class, and SAT 25th-75th percentile are higher in the CDS than the variables given in the USNWR dataset.

In order to arrive at results as unbiased as possible, most missing variables are filled in with data from the Integrated Postsecondary Education Data System (IPEDS), a database maintained by National Center for Education Statistics (NCES). Data in IPEDS are collected through mandatory surveys authorized by law under the Section 153 of the Education Sciences Reform Act of 2002. All institution are *obligated* to complete all IPEDS surveys. With the additional data from IPEDS, the original datasets are expanded with the variables in Table [2.3]. However, class size related variables used to calculate **class size index** and **percent faculty with terminal degree in their field** are still missing since they are not required by NCES to be reported and therefore, not in any of the IPEDS datasets.

Table 2.1: A list of variables in both 2009 and 2019 datasets. The variables present are marked by ∘, the absent ×. 27 of the variables are shared across the two datasets while the remaining nine are only present in the 2009 dataset.

Variable name	2009	2019
Rank	∘	∘
School	∘	∘
Nonresponder	∘	×
State	∘	∘
Public/Private	∘	∘
New Category	∘	×
New School	∘	×
Overall Score	∘	∘
Peer Assessment Score	∘	∘
Graduation and Retention Rank	∘	∘
Average Freshman Retention Rate	∘	∘
Footnote	∘	∘
Predicted Graduation Rate	∘	∘
Actual Graduation Rate	∘	∘
Footnote_1	∘	∘
Over/Under Performance	∘	∘
Faculty Resource Rank	∘	∘
% of Classes under 20	∘	∘
Footnote_2	∘	×
% of Classes over 50 or more	∘	∘
Footnote_3	∘	×
Student/Faculty ratio	∘	∘
Footnote_4	∘	∘
% of Full-time Faculty	∘	×
Footnote_5	∘	×
Selectivity Rank	∘	∘
SAT/ACT 25th-75th percentile	∘	∘
Footnote_6	∘	∘
Freshmen in Top 10% of High School Class	∘	∘
Footnote_7	∘	∘
Acceptance Rate	∘	×
Footnote_8	∘	×
Financial Resources Rank	∘	∘
Alumni Giving Rank	∘	∘
Average Alumni Giving Rate	∘	∘
Footnote_9	∘	∘

Table 2.2: College ranking criteria and weights published by *USNWR* for 2019.

Ranking Indicator	National Schools	Regional Schools
Graduation and retention rates	22%	22%
Average six-year graduation rate	17.6%	17.6%
Average first-year student retention rate	4.4%	4.4
Social mobility	5%	5%
Pell Grant graduation rates	2.5%	2.5%
PG graduation rates compared with all other students	2.5%	2.5%
Graduation rate performance	8%	8%
Undergraduate academic reputation	20%	20%
Peer assessment survey	15%	20%
High school counselors' ratings	5%	0%
Faculty resources for 2017-2018 academic year	20%	20%
Class size index	8%	8%
Faculty compensation	7%	7%
Percent faculty with terminal degree in their field	3%	3%
Percent faculty that is full time	1%	1%
Student-faculty ratio	1%	1%
Student selectivity for the fall 2017 entering class	10%	10%
SAT and ACT score	7.75%	7.75%
High school class standing in top 10%	2.25%	0%
High school class standing in top 25%	0	2.25%
Acceptance rate	0%	0%
Financial resources per student	10%	10%
Average alumni giving rate	5%	5%
Total	100%	100%

Table 2.3: A table

Variable	Description
FTE faculty	Full-time-equivalent faculty is calculated by adding the number of full-time faculty and 1/3 of the number of part-time faculty
Total faculty	Total number of faculty including full-time and part-time
Faculty benefits	Cash contributions in the form of supplementary or deferred compensation other than salary, including retirement plans, social security taxes, medical/dental plans, guaranteed disability income protection plans, tuition plans, housing plans, unemployment compensation plans, group life insurance plans, worker's compensation plans, and other benefits in-kind with cash options.
Average faculty salaries	Average salaries equated to 9-months of full-time non-medical instructional staff
Pell Grant graduation rates	6-year graduation rate of students receiving Pell Grant
Instructional expenditure per FTE student	Instruction expenses per full-time-equivalent student includes all expenses of the colleges, schools, departments, and other instructional divisions of the institution and expenses for departmental research and public service that are not separately budgeted.
Research expenditure per FTE student	Expenses spent on research per full-time-equivalent student
Public service expenditure per FTE student	Expense spent on public service per full-time-equaivalent student
Academic support expenditure per FTE student	Expense spent on academic-support per full-time-equivalent student
Student service expenditure per FTE student	Expense spent on student service per full-time-equivalent student
Institutional support expenditure per FTE student	Expense spent on institutional support per full-time-equivalent student

2.2 Method

Some might ask why don't we just save all the hassles and test USNWR's model again using the expanded dataset just introduced. There are several good reasons. The first and most blunt one is that several points of their model are unclear so even with the expanded pool of variables we still don't know how they get to some of the numbers. For example, USNWR mentioned in the article about their methodology that one of the variables, *Class Size Index*, is calculated by the following method:proportions of undergraduate classes with fewer than 20 students contribute the *most* credit to this index, with classes with 20 to 29 students coming *second*, 30 to 39 students *third*, and 40 to 49 students *fourth*. Classes that are 50 or more student receive *no credit*. They told us the importance of each variables but didn't clearly say how they numerically contribute to *Class Size Index*. Similar problems will be discussed in detail in Chapter 3. Another problem with USNWR's model is that many of the variables are highly correlated with each other. The mullticollinearity problem can be immediately seen in the following correlation heatmaps of the variables.

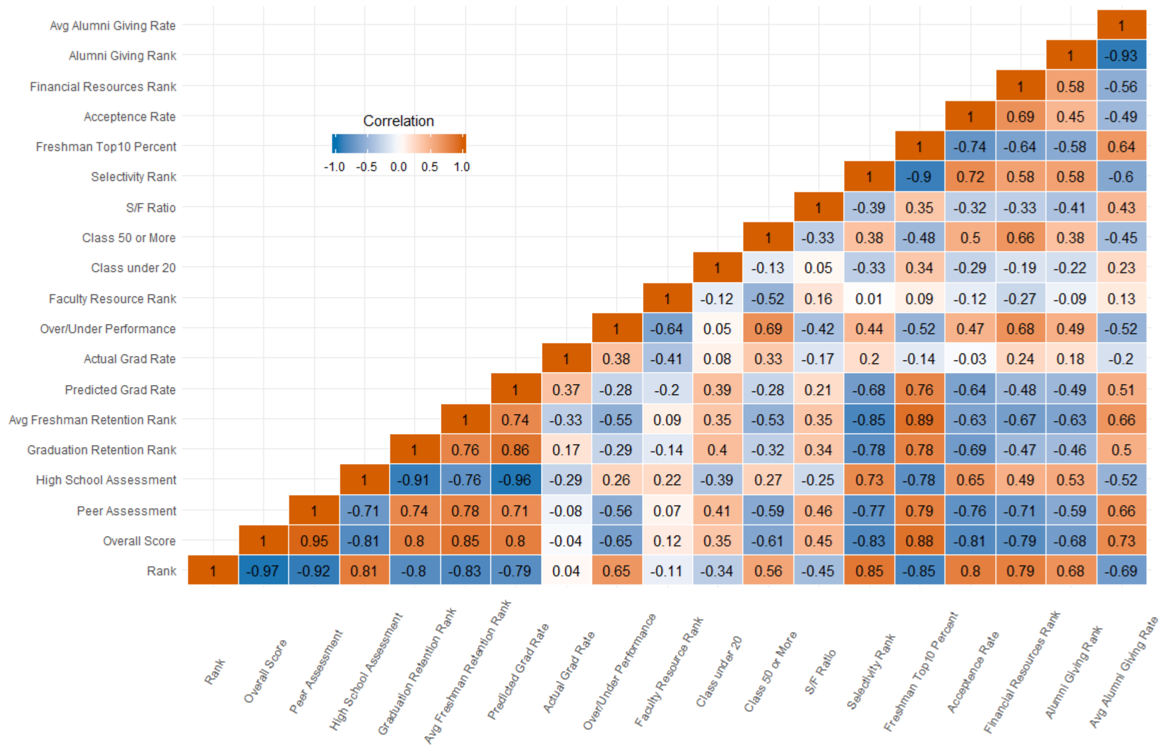

Figure 2.1: A correlation heatmap of all the variables in the original 2009 dataset. Many of the variables having major weight in the USNWR's weight-and-sum model are highly correlated with each other.

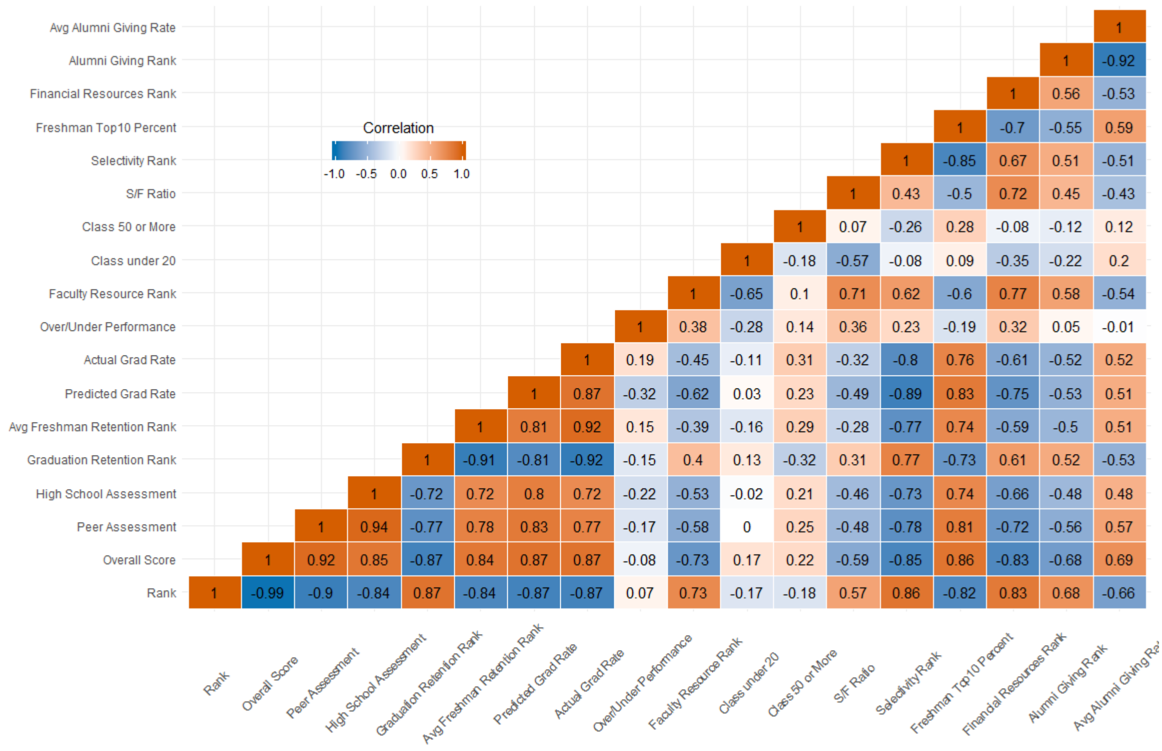

Figure 2.2: A correlation heatmap of all the variables in the original 2019 dataset. Like the original 2009 dataset, it also has severe mullticollinearity problem.

The final reason is that USNWR's weight-and-sum system does not generate standard error and thus uncertainty analysis is impossible. While in our case, if any difference in ranking is found, we certainly want to check whether the difference in the estimated ranks are statistically significant and USNWR's model would not allow us to achieve this goal.

2.2.1 Elastic Net

One of the approaches we took to replicate the USNWR National Liberal Arts Colleges ranking results for 2009 and 2019 is the Elastic Net, a regularized regression method that's usually used for linear regressions and logistic regressions. The Elastic Net method linearly combines the two most often used shrinkage methods, Ridge Regression and the LASSO, and thus overcomes the limitations of using either method while still producing well regularized regression results.

The ordinary least-squares regressions estimate the coefficients $\beta_0, \beta_1, ..., \beta_p$ by minimizing the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2$$

The Elastic Net method works in similar way as the least squares, except that instead of minimizing the RSS, the Elastic Net minimizes the linear combination of RSS and L1, L2, the shrinkage penalties of ridge and LASSO:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \alpha \lambda_1 \sum_{j=1}^{p} \beta_j^2 + (1-\alpha)\, \lambda_2 \sum_{j=1}^{p} |\beta_j| = RSS + L_1 + L_2$$

where $\lambda_1, \lambda_2 \geq 0$ and in which $L_1 = \lambda \sum_{j=1}^{p} \beta_j^2$ is the shrinkage penalty of Ridge, and $L_2 = \lambda \sum_{j=1}^{p} |\beta_j|$ is the shrinkage penalty of LASSO.

Ridge Regression

Simple ridge regression produces the model by minimizing the linear combination of RSS and a shrinkage penalty, L_1:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p} \beta_j^2$$

with $\lambda_1 \geq 0$.

The penalty term in ridge regressions, $L_1 = \lambda \sum_{j=1}^{p} \beta_j^2$, is smaller when $\beta_1, ...\beta_p$ are closer to zero. Thus minimizing this term causes the estimates of $\beta's$ to shrink toward zero. The tuning parameter λ_1 serves the role to control the effect of the shrinkage penalty. If λ_1 is 0, the penalty term goes away and the estimates are same as least squares estimates. As $\lambda_1 \to \infty$, the estimates approach zero as the shrinking effect increases. With different values chosen for λ, the penalty term affect the model differently, and thus produces different set of estimates. Usually cross-validation is performed to select the preferred value of λ_1.

The ridge regression works the best when the least squares produces estimates with high variance. The increase in λ would reduce the flexibility of the ridge regression estimate, increase the bias while decrease the variance. In cases when the least squares produce estimates with low bias but high variance, which can be caused by multicollinearity or over-fitting, this shrinkage penalty can reduce the variability, and thus avoid highly variable estimates. In this study, we have limited number of observations (institutions) in both datasets (maximum 172 schools), and relatively large number of variables available. (*need to be determined later*) This regularization could be used to reduce the variability in our estimates.

However, there exists limitations to ridge regression. The penalty L_1 shrinks coefficients toward zero but does not set any of them exactly to zero. Thus all variables are included in the final model produced by the ridge regression. Due to the limitations of data available in the original datasets provided by USNWR, we extracted additional variables from external sources (IPEDS, College Results) based on the description provided by USNWR. However, it is hard to be certain whether the variables selected match with the variables truly used by USNWR. Instead of assuming that all variables in our datasets were used by USNWR and have effect on the response variable, we use Elastic Net, which combines the ridge regression with LASSO, so that variable subsetting can be an option.

LASSO

The simple LASSO method estimates the model coefficients by minimizing the linear combination of RSS and a shrinkage penalty, L_2:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda_2 \sum_{j=1}^{p} |\beta_j|$$

with $\lambda_2 \geq 0$

When using LASSO and the tuning parameter λ is large enough, minimizing the shrinkage penalty $L_2 = \lambda \sum_{j=1}^{p} |\beta_j|$ can force some of the estimated coefficients to be 0, and thus allows LASSO to perform variable selection. Similar to ridge, when λ is zero, the penalty term goes to zero and the results produced is same as the one produced by the ordinary least squares. As λ increases, variables with sufficiently low estimated coefficient are thrown away, the flexibility of the estimates reduces, which brings more bias and less variance to the final results being produced. This allows LASSO to perform both shrinkage and variable selection.

In many cases, LASSO is sufficient to use as it performs both shrinkage and variable selection. This is not true, though, in the current study. Many of the variables in the datasets are highly correlated. With our prior knowledge from the description of method used by USNWR, the highly correlated variables can each have different effect on ranking results. LASSO, when dealing with highly correlated variables, would force some of the estimated coefficients of the correlated variables to be zero. Thus if we use simple LASSO model, many variables can be removed from the final model while are in fact influencial to the ranking result. Combining the ridge regression balances out this limitation.

Elastic Net Modeling

As mentioned earlier, the Elastic Net linearly combines the ridge and the LASSO model, which produce less variable estimates, and balance out the limitations of the two models.

USNWR assigns each school an overall score (0-100) based on their ranking formula, and the actual ranks are assigned based on the overall score. To best replicate the rankings, we chose the overall score to be the response variable. Cross Validation is used to choose values of λ_1 and λ_2 for the best model. The *train* function from R package *caret* is used to perform the cross validation. The variables are standardized during the modeling process, thus the estimated coefficients are not as directly interpretable as they would be in SLR. But the relative difference of the estimated coefficients shows the relative difference in the effect that the variables have on the response variables, where larger the estimated coefficient, larger effect it would have on the response variable. The selected model and estimated coefficients for 2009 and 2019 are listed in Table [2.4] and Table [2.5].

Table 2.4: Elastic Net Estimated Coefficients 2009. The variables with coefficients marked as - are discarded during the model selection process.

Variable	Coefficient
Peer Assessment Score	6.58
Average Freshmen Retention Rate	0.53
Predicted Graduation Rate	-
Average Graduation Rate	2.26
Graduation Performance	0.29
% classes size under 20	1.17
% classes size 50 or more	-0.71
Student Faculty Ratio	-0.27
% Full-time Faculty	0.07
% Freshmen High School top 10	1.75
Accept rate	-0.29
Average Alumni Giving Rate	1.35
Test Score (SAT, ACT)	0.90
Average Faculty Compensation	0.74
Expenditure per FTE	1.33

Table 2.5: Elastic Net Estimated Coefficients 2019. The variables with coefficients marked as - are discarded during the model selection process.

Variable	Coefficient
Peer Assessment Score	4.56
High School Counselor Assessment Score	0.64
Average Freshmen Retention Rate	1.55
Predicted Graduation Rate	-
Average Graduation Rate	1.23
Graduation Performance	0.87
% classes size under 20	1.64
% classes size 50 or more	-0.38
Student Faculty Ratio	-
% Freshmen High School top 10	0.36
Average Alumni Giving Rate	1.73
Test Score (SAT, ACT)	0.56
PG (Pell Grant Recipient) Graduation Rate	2.74
Ratio b/t PG and non-PG Graduation Rate	-0.55
% Full-time Faculty	0.05
Faculty Compensation	0.11
Expenditure per FTE	3.18

2.2.2 PCR

Due to the multicollinearity of the criteria used by U.S News, another method which can bypass this issue is *Principle Component Regression* (PCR). The basic idea of PCR is to use *Principle Components* generate through *Principle Component Analysis* as predictors in a regression model. To make sense of this method, we will start from the foundation, *Principle Component Analysis.*

Principal Component Analysis

Principle Component Analysis (PCA) is at its heart a method of dimentionality reduction. Suppose we have n observations and m variables, where $n-1 \geq m$[1]. We can represent the dataset by the matrix $\mathbf{X}$

$$\mathbf{X} = (X_1, X_2, \dots, X_m) = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & & x_{2m} \\ \vdots & & \ddots & \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

We wish to understand the relationship within the set of variables. One way to do so is to plot and examine pairwise scatterplots of $\mathbf{X}$. But since there would be $\binom{m}{2}$ plots in total, one can see that this task quickly becomes dreadful if not impossible. Now, we wish to have a low-dimensional representation of $\mathbf{X}$, which still encapsulates as much of the variance of $\mathbf{X}$ as possible and to generate a low-dimensional representation of high-dimensional data is exactly what PCA is for. Instead of looking at all m features,

[1]There are at most $\min\{n-1, m\}$ principal components.

PCA suggests we can instead examine a set of linear combinations of $X_1, X_2, \ldots, X_m$ called *Principle Components*. Each *Principle Component*, Z_i, can be calculated as

$$Z_i = \sum_{j=1}^{m} \phi_{ji} X_j, \quad \text{where } \sum_{j=1}^{m} \phi_{ji}^2 = 1$$

Let

$$\mathbf{Z} = (Z_1, Z_2, \ldots, Z_m) = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & & z_{2m} \\ \vdots & & \ddots & \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix}$$

$$\mathbf{\Phi} = (\Phi_1, \Phi_2, \ldots, \Phi_m) = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1m} \\ \phi_{21} & \phi_{22} & & \phi_{2m} \\ \vdots & & \ddots & \\ \phi_{n1} & \phi_{n2} & \cdots & \phi_{nm} \end{bmatrix}$$

Then

$$\mathbf{Z} = \mathbf{X\Phi}$$

One might wonder how ϕ is determined and the constraint $\sum_{j=1}^{m} \phi_{ji}^2 = 1$ might seem arbitrary at this moment, but it will make sense in the following steps. Since we want to capture as much of the variance of $\mathbf{X}$ as possible in one principle component, we want to maximize the sample variance by finding a set of ϕ such that

$$\text{Var}(Z_i) = \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{j=1}^{m} \phi_{ji} x_{kj} - \bar{Z}_i \right)^2$$

Since we are only interested in the variance of the dataset, we assume the $\bar{X}_i = 0$ for all i. Then $\bar{Z}_i = 0$ for all i as well. Then the problem becomes maximizing

$$\frac{1}{n} \sum_{k=1}^{n} \left(\sum_{j=1}^{m} \phi_{ji} x_{kj} \right)^2$$

Now, one can see that we need the constraint $\sum_{j=1}^{m} \phi_{ji}^2 = 1$ otherwise, we can make the sample variance arbitrarily large by making the absolute value of ϕ_{ij} arbitrarily large. After one principle component is found, we calculate another principle component that captures maximal variance out of all linear combinations of $X_1, X_2, \ldots, X_m$ that is *uncorrelated* with the previous one, i.e. find another set of ϕ that maximize the sample variance and all the sets of ϕ should ensure that Z_i's are *pairwise orthogonal*.

Why PCA in our case?

As mentioned above, the variables used by U.S News in their ranking system are highly correlated. If we use a regression model without a shrinkage method, the coefficients of the resulting model can be greatly affected by even a small change in the data or model. Therefore, the model fitted using such training dataset can perform very poorly in the test dataset. By applying PCA, we can create uncorrelated predictors that still capture the a large portion of the variance of the original predictors. Then using these principle components, we can build a linear regression model.

Final PCR Model

In our case, we have 14 variables for the 2009 model and 16 variables for the 2019 model. Therefore, 14 principal components were calculated for the 2009 model and 16 principal components were calculated for the 2019 model. Then, 14 linear regression models using $\{Z_1\}, \{Z_1, Z_2\}, \ldots, \{Z_1, Z_2, \ldots, Z_{14}\}$ as explanatory variables were built for 2009. Similarly, 16 linear regression models using $\{Z'_1\}, \{Z'_1, Z'_2\}, \ldots, \{Z'_1, Z'_2, \ldots, Z'_{16}\}$ as explanatory variables were built for 2019. With the intention to reduce the dimensionality of the data while capturing the majority of the variance in the data, we wanted to pick a model that has much less explanatory variables than the full-model while having a high explanatory power of the variance within the variables used to calculate the principal components. As a result, the model of 8 principal components were selected for both 2009 and 2019. For the 2009 dataset, the 8 principal components captures 94.73% of the variance within the 14 variables and explains 97.37% of the variance of the overall score given by USNWR. On the other hand, for 2019 dataset, the 8 principal components captures 93.22% of the variance within the 16 variables and explains 97.26% of the variance of the overall score given by USNWR.

2.3 Result

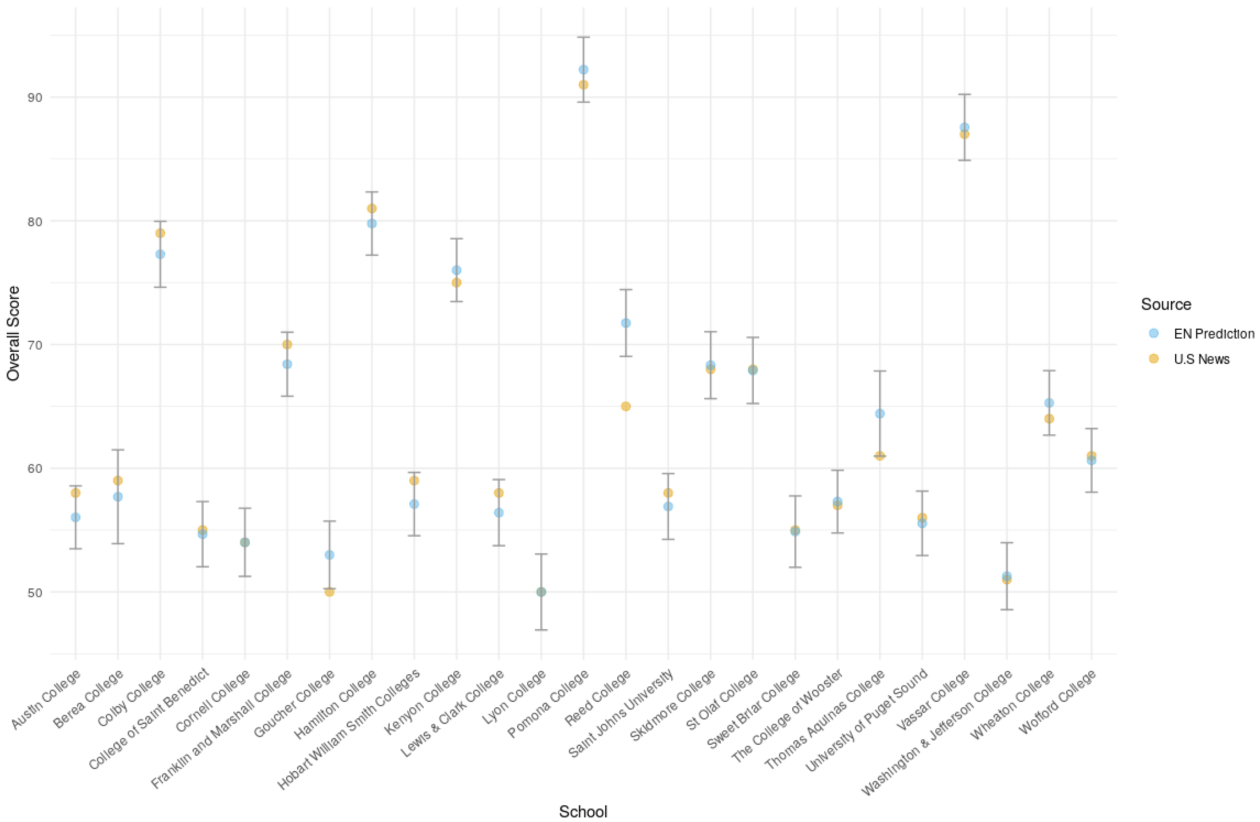

Figure 2.3: Elastic Net predictied overall score in the 2009 test dataset. The blue points represent scores predicted by the PCR model. On the other hand, the yellow points represent scores given by U.S News. The vertical bars represent the prediction intervals. For every school but Reed College, the U.S News score falls in the prediction interval.

Table 2.6: Elastic Net Prediction Results 2009.

School	USNews Overall Score	Predicted Score	Prediction Interval
Pomona College	91	92	[90, 95]
Vassar College	87	88	[85, 90]
Hamilton College	81	80	[77, 82]
Colby College	79	77	[75, 80]
Kenyon College	75	76	[73, 79]
Franklin and Marshall College	70	68	[66, 71]
Skidmore College	68	68	[66, 71]
St Olaf College	68	68	[65, 71]
Reed College	65	72	[69, 74]
Wheaton College	64	65	[63, 68]
Thomas Aquinas College	61	64	[61, 68]
Wofford College	61	61	[58, 63]
Berea College	59	58	[54, 61]
Hobart William Smith Colleges	59	57	[55, 60]
Austin College	58	56	[53, 59]
Lewis & Clark College	58	56	[54, 59]
Saint Johns University	58	57	[54, 60]
The College of Wooster	57	57	[55, 60]
University of Puget Sound	56	56	[53, 58]
College of Saint Benedict	55	55	[52, 57]
Sweet Briar College	55	55	[52, 58]
Cornell College	54	54	[51, 57]
Washington & Jefferson College	51	51	[49, 54]
Goucher College	50	53	[50, 56]
Lyon College	50	50	[47, 53]

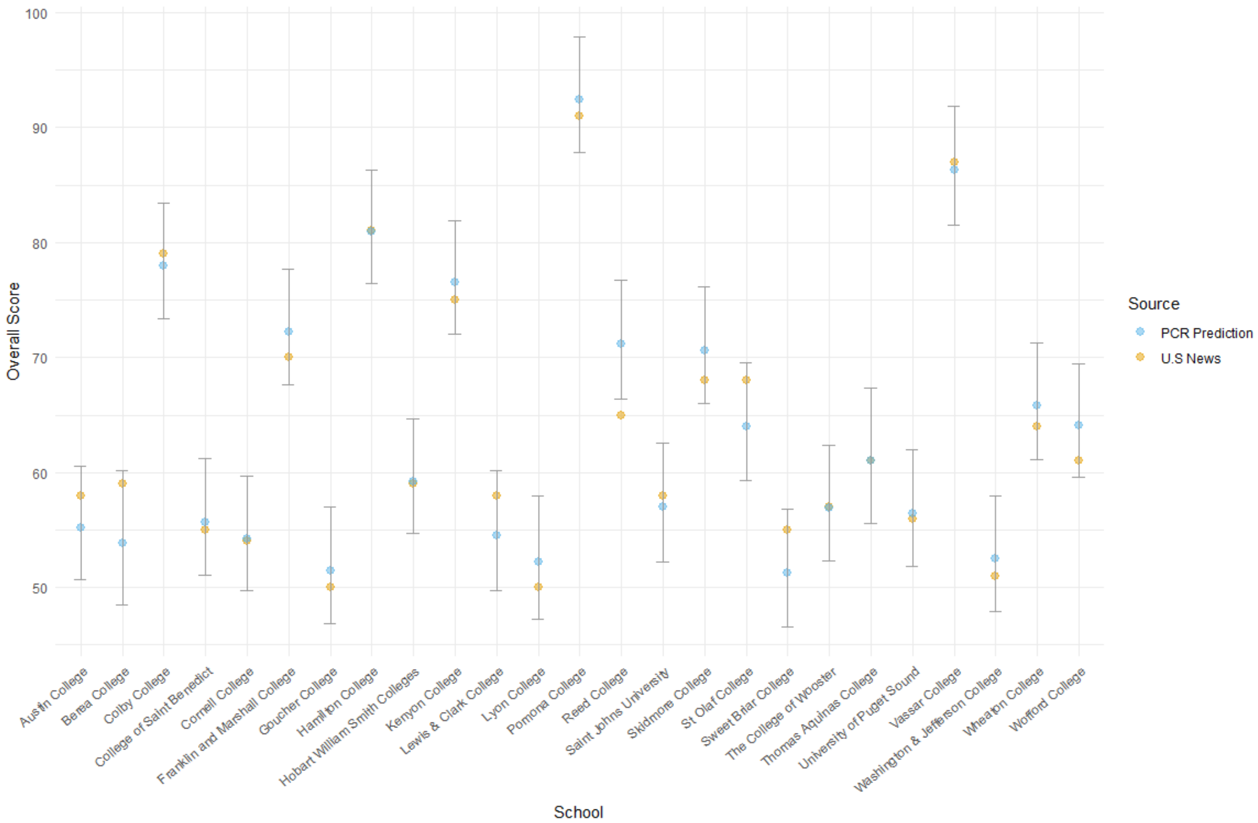

Figure 2.4: PCR predictied overall score in the 2009 test dataset. The blue points represent scores predicted by the PCR model. On the other hand, the yellow points represent scores given by U.S News. The vertical bars represent the prediction intervals. For every school but Reed College, the U.S News score falls in the prediction interval.

Table 2.7: PCR Prediction Results 2009.

School	USNews Overall Score	Predicted Score	Prediction Interval
Pomona College	91	93	[88, 98]
Vassar College	87	87	[82, 92]
Hamilton College	81	81	[76, 86]
Colby College	79	78	[73, 83]
Kenyon College	75	77	[72, 82]
Franklin and Marshall College	70	73	[68, 78]
Skidmore College	68	71	[66, 76]
St Olaf College	68	64	[59, 70]
Reed College	65	72	[66, 77]
Wheaton College	64	66	[61, 71]
Thomas Aquinas College	61	61	[56, 67]
Wofford College	61	65	[60, 69]
Berea College	59	54	[49, 60]
Hobart William Smith Colleges	59	60	[55, 65]
Austin College	58	56	[51, 61]
Lewis & Clark College	58	55	[50, 60]
Saint Johns University	58	57	[52, 63]
The College of Wooster	57	57	[52, 62]
University of Puget Sound	56	57	[52, 62]
College of Saint Benedict	55	56	[51, 61]
Sweet Briar College	55	52	[47, 57]
Cornell College	54	55	[50, 60]
Washington & Jefferson College	51	53	[48, 58]
Goucher College	50	52	[47, 57]
Lyon College	50	53	[47, 58]

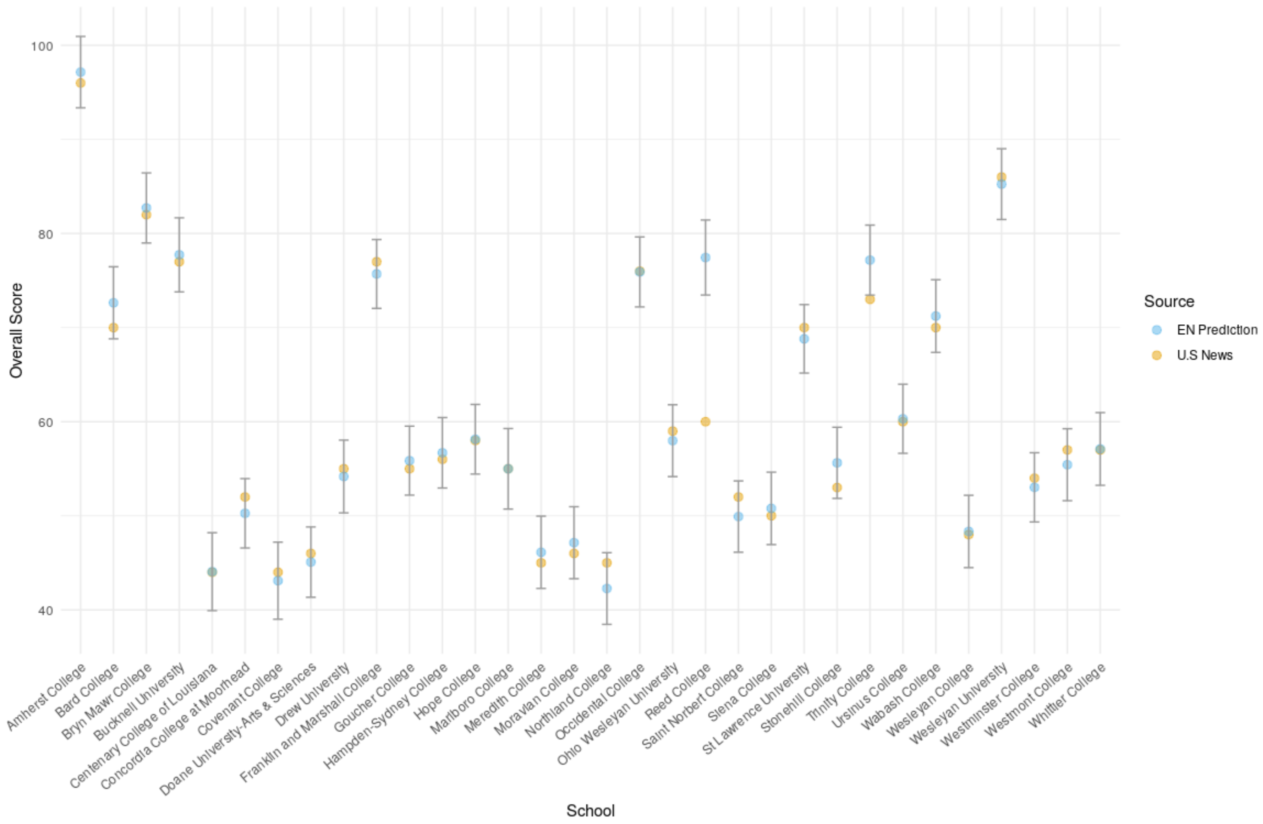

Figure 2.5: Elastic Net predictied overall score in the 2019 test dataset. The blue points represent scores predicted by the PCR model. On the other hand, the yellow points represent scores given by U.S News. The vertical bars represent the prediction intervals. For every school but Reed College, the U.S News score falls in the prediction interval.

Table 2.8: Elastic Net Prediction Results 2019.

School	USNews Overall Score	Predicted Score	Prediction Interval
Amherst College	96	97	[93, 101]
Wesleyan University	86	85	[81, 89]
Bryn Mawr College	82	83	[79, 86]
Bucknell University	77	78	[74, 82]
Franklin and Marshall College	77	76	[72, 79]
Occidental College	76	76	[72, 80]
Trinity College	73	77	[73, 81]
Bard College	70	73	[70, 76]
St Lawrence University	70	69	[65, 72]
Wabash College	70	71	[67, 75]
Reed College	60	77	[73, 81]
Ursinus College	60	60	[57, 64]
Ohio Wesleyan University	59	58	[54, 62]
Hope College	58	58	[54, 62]
Westmont College	57	55	[52, 59]
Whittier College	57	57	[53, 61]
Hampden-Sydney College	56	57	[53, 60]
Drew University	55	54	[50, 58]
Goucher College	55	56	[52, 60]
Marlboro College	55	55	[51, 59]
Westminster College	54	53	[49, 57]
Stonehill College	53	56	[52, 59]
Concordia College at Moorhead	52	50	[47, 54]
Saint Norbert College	52	50	[46, 54]
Siena College	50	51	[47, 55]
Wesleyan College	48	48	[44, 52]
Doane University - Arts & Sciences	46	45	[41, 49]
Moravian College	46	47	[43, 51]
Meredith College	45	46	[42, 50]
Northland College	45	42	[38, 46]
Centenary College of Louisiana	44	44	[40, 48]
Covenant College	44	43	[39, 47]

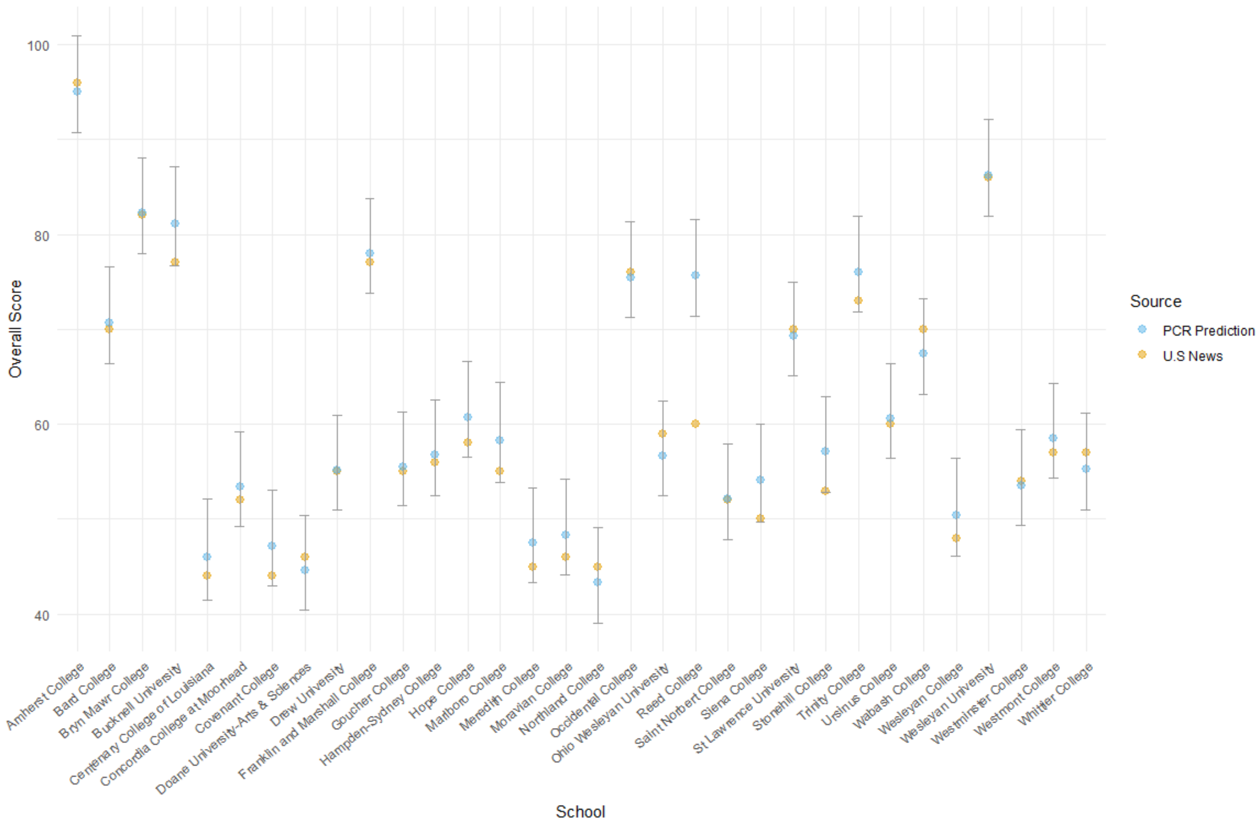

Figure 2.6: PCR predictied overall score in the 2019 test dataset. The blue points represent scores predicted by the PCR model. On the other hand, the yellow points represent scores given by U.S News. The vertical bars represent the prediction intervals. For every school but Reed College, the U.S News score falls in the prediction interval. Comparing with the result from 2009, the U.S News score is even lower than all the possible values in the prediction interval.

Table 2.9: PCR Prediction Results 2019.

School	USNews Overall Score	Predicted Score	Prediction Interval
Amherst College	96	96	[91, 101]
Wesleyan University	86	87	[82, 92]
Bryn Mawr College	82	83	[78, 88]
Bucknell University	77	82	[77, 87]
Franklin and Marshall College	77	79	[74, 84]
Occidental College	76	76	[71, 81]
Trinity College	73	77	[72, 82]
Bard College	70	71	[66, 77]
St Lawrence University	70	70	[65, 75]
Wabash College	70	68	[63, 73]
Reed College	60	77	[71, 82]
Ursinus College	60	61	[56, 66]
Ohio Wesleyan University	59	57	[52, 63]
Hope College	58	62	[57, 67]
Westmont College	57	59	[54, 64]
Whittier College	57	56	[51, 61]
Hampden-Sydney College	56	58	[52, 63]
Drew University	55	56	[51, 61]
Goucher College	55	56	[51, 61]
Marlboro College	55	59	[54, 64]
Westminster College	54	54	[49, 59]
Stonehill College	53	58	[53, 63]
Concordia College at Moorhead	52	54	[49, 59]
Saint Norbert College	52	53	[48, 58]
Siena College	50	55	[50, 60]
Wesleyan College	48	51	[46, 56]
Doane University - Arts & Sciences	46	45	[40, 50]
Moravian College	46	49	[44, 54]
Meredith College	45	48	[43, 53]
Northland College	45	44	[39, 49]
Centenary College of Louisiana	44	47	[41, 52]
Covenant College	44	48	[43, 53]

Chapter 3

Discussion

3.1 Is Reed College under-ranked?

As shown in the result section, for 2009 and 2019, the predicted overall scores of Reed College by both Elastic Net and PCR are much higher than the scores given by USNWR.

Table 3.1: Overall scores of Reed College for 2009 and 2019. The scores generated by both Elastic Net and PCR are higher than the scores given by USNWR.

Source	2009	2019
PCR	72	77
Elastic Net	72	77
USNWR	65	60

Further uncertainty analysis suggests that the difference in scores is significant for all years. The 95% prediction intervals for the overall score of Reed College predicted by Elastic Net are $[70, 74]$ for 2009 and $[73, 81]$ for 2019. On the other hand, the 95% prediction intervals for the overall score of Reed College predicted by PCR are $[66, 77]$ for 2009 and $[71, 82]$ for 2019. Clearly, the scores given by USNWR are in none of those intervals. Therefore, it is safe to say that Reed College is truely under-ranked as suspected. Referring back to overall scores given by USNWR and their corresponding ranks, by the predicted overall scores generated by PCR, Reed College should have been ranked at the 37th among the 124 liberal arts colleges rather than the 54th in 2009. On the other hand, it should have been ranked at the 36th among the 173 liberal arts colleges rather than the 90th in 2019.

A cautious reader might notice that in 2009, the overall score of Reed College given by USNWR is not as drastically to the left of the 95% prediction intervals as it is in 2019. Since the results of both models agree and such abnormality only applies to Reed College, the determining factor of the discrepancy is unlikely to be the predicative power of the models. Then, what is left to investigate is the data and indeed it is the cause of the abnormality. Comparing values of variables in the original 2019 dataset with the IPEDS data and Reed's Common DataSet, we found significant mismatch[1]. To give an extreme example, the original 2019 dataset has a variable called *Financial Resources*

[1]Refer to Appendix A for a detailed table comparing USNWR's data, IPEDS data, and Reed College's CDS data for 2019.

Rank[2] and Reed College is ranked at the 169th among 173 liberal arts colleges. However, a calculation based on USNWR's methodology[3] reveals that Reed College's expenditure per FTE student is higher than not only a school with the same financial resource rank but also a school with much higher financial resources rank. With the calculated expenditure per FTE student, Reed College's financial resource rank should be the 30th instead of the 169th among the 173 liberal arts colleges.

Table 3.2: 2019 *Financial Resource Rank* and expenditure per FTE student for Earlham College, Salem College, and Reed College. Although Reed College has the highest expenditure per FTE student among these schools, it has the lowest financial resource rank.

School	Financial Resource Rank	Expenditure per FTE Student
Earlham College	50	47956.31
Salem College	169	30004.51
Reed College	169	54566.76

As for the 2009, the data for Reed College from all three data sources are very close[4] except financial resource rank. Similar to the situation in 2019, Reed College's financial resource rank is also drastically under-ranked by 90.

Recall that another objective of this project is to investigate if schools are under-ranked because of their refusal to report statistics to USNWR. And our results show no systematic effect of non-reporting on ranks. In the 2009 dataset, Berea College is also marked as non-reporting. However, its overall scores given by USNWR for both 2009 and 2019 are close to the predicted overall score by Elastic Net and PCR and are within the prediction intervals. As for 2019, there is no variable in the dataset indicating whether a school is non-reporting or not so we assumed that Reed College is the only non-reporting school. And as it turns out, Reed College is the only school whose overall score given by USNWR is outside of the prediction interval.

At this point, it is clear that Reed College is under-ranked. However, it is not under-ranked because it is a non-reporting school. Although the true reason why Reed College is under-ranked can not be inferred by our research, how a lower rank is achieved is unveiled. Based on our results, the most credible conjecture is that the data of Reed College is somehow modified and thus resulting in a lower rank.

3.2 Potential Problems & Future Research

Although the results of our research seem to be promising, it is by no means perfect. The following section will introduce some major limitations and problems of our models and methodology. Then we will suggest some potential directions if one is intrigued to improve our results.

[2]Refer back to Table [2.1] for detailed list of variables and see Appendix A for a detailed variable description.

[3]USNWR states that *Financial Resource* is based on expenditure per FTE (full-time-equivalent) student. Therefore, to verify the data, we calculated expenditure per FTE student based on IPEDS data for every school in the 2019 dataset.

[4]Also refer to Appendix A for a detailed table comparing USNWR's data, IPEDS data, and Reed College's CDS data for 2009.

3.2.1 Unobtainable variables

The models generated by Elastic Net for both 2009 and 2019 assigned the largest effect on the overall score to *Peer Assessment Score*, with the coefficient of the 2009 model being 6.58 and the coefficient of 2019 model being 4.56. Having seen all the convoluted problems on data credibility in previous analysis, naturally, we wanted to verify the validity of the variable by comparing it with external credible data source. However, *Peer Assessment Score* is a variable specific to USNWR's college survey questionnaire and thus, we failed to find other credible sources containing this variable. This uncertainty can potentially rise or, less likely, lower Reed College's predicted overall score. Another similar variable that we can't verify the credibility of is *High School Counselor Assessment Score* but it has a relatively small effect. It is not used by USNWR in 2009 and in the 2019 Elastic Net model, it has a coefficient of 0.64.

The other two *unobtainable* variables are all related to faculty resource, the first one being *Percent of faculty with a doctoral degree* and the second being *regional cost of living*. *Regional cost of living* is related to faculty resource because USNWR uses it to scale faculty salaries, which we obtained from IPEDS. One might think such variable can be easily found in census data but the one used by USNWR, as mentioned in their own article from 2008, is an index from the consulting firm Runzheimer International. However, there is no further mention of such consulting firm in their most recent article on methodology. Therefore, it is unclear that what measure of cost of living are they currently using. Moreover, they only vaguely state their methodology of calculating faculty salaries as,"adjust for regional differences in cost of living". With too little information and limited time, we decided only to include unscaled faculty salaries in our model. Since regional cost of living is not account for by any variables in the model, its potential effect is inherited by the error term in our models, which can result in larger prediction intervals and less accurate prediction. Last but not least, *percent of faculty with a doctoral degree* is simply not included in any of the data sources we readily have. It is not even in the dataset from USNWR themselves. And such statistic is not one of the variables that needs to be report to NCES[5]

3.2.2 NA's

After we expanded our 2019 dataset with IPEDS data, there are two sources of NA's: (1) expenditure data (2) Pell Grant graduation rate data. Since there are 10 colleges lacking expenditure data, it wouldn't be ideal to replace NA's with mean or median. Therefore, we simply took the 10 colleges out of our dataset. To keep the methodology consistent, we also took out 2 colleges lacking Pell Grant graduation rate data. If one can come up with better ways to deal with these NA's or even find some source that does have the missing data, the two models will have stronger predicative power since the sample size is increased.

3.2.3 Future Research

The following are some directions to consider if one is interested in finding more accurate results:

[5]National Center for Education Statistics.

1. Either find data sources to replace the *unobtainable* variables with credible data or use some proxy to approximate their effects.

2. Find a better way to deal with NA's in the dataset.

3. To increase the size of test dataset, one might consider finding years that USNWR didn't change their weight system and use the whole dataset from one of the years as test dataset.

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(thesisdown))
  devtools::install_github("ismayc/thesisdown")
library(thesisdown)
if(!require(kableExtra))
  install.packages("kableExtra")
library(kableExtra)
```

In Chapter ??:

Appendix B

The Second Appendix, for Fun

References

Bougnol, M.-L., & Dulá, J. H. (2015). Technical pitfalls in university rankings. *Higher Education*, *69*(5), 859–866. http://doi.org/10.1007/s10734-014-9809-y

Clarke, M. (2004). Weighing Things Up: A Closer Look at U.S. News & World Report's Ranking Formulas. *College and University; Washington*, *79*(3), 3–9. Retrieved from https://search.proquest.com/docview/225608009/abstract/56FEE3878D134B69PQ/1

Eagan, K., Stolzenberg, E. B., Ramirez, J. J., Aragon, M. C., Suchard, M. R., & Rios-Aguilar, C. (2016). *The American Freshman: Fifty-Year Trends 1966-2015*. Los Angeles: Higher Education Research Institute, UCLA. Retrieved from https://heri.ucla.edu/publications/

Ehrenberg, R. G. (2002). Reaching for the Brass Ring: The U.S. News & World Report Rankings and Competition. *The Review of Higher Education*, *26*(2), 145–162. http://doi.org/10.1353/rhe.2002.0032

Gnolek, S. L., Falciano, V. T., & Kuncl, R. W. (2014). Modeling Change and Variation in U.S. News & WorldReport College Rankings: What would it really take to be in the Top 20? *Research in Higher Education*, *55*(8), 761–779. http://doi.org/10.1007/s11162-014-9336-9

Grewal, R., Dearden, J. A., & Llilien, G. L. (2008). The University Rankings Game: Modeling the Competition Among Universities for Ranking. *The American Statistician*, *62*(3), 232–237. http://doi.org/10.1198/000313008X332124

Lydgate, C. (2018, September). Reed College Admission College Rankings. Retrieved from https://www.reed.edu/apply/college-rankings.html

Morse, R., Brooks, E., & Mason, M. (2018, September). How U.S. News Calculated the 2019 Best Colleges Rankings. *US News & World Report*. Retrieved from https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings

Nelson, C. B. (2007, October). University Business: Controversy. Retrieved from https://web.archive.org/web/20071024014745/http://universitybusiness.ccsct.com/page.cfm?p=64

Sanoff, A. P. (2007). The "U.S. News" College Rankings: A View from the Inside. In *College and University Ranking Systems: Global Perspectives and American Challenges* (pp. 9–24). Washington, DC: Institute for Higher Education Policy. Retrieved from https://eric.ed.gov/?id=ED497028

Webster, T. J. (2001). A principal component analysis of the U.S. News & World Report tier rankings of colleges and universities. *Economics of Education Review*, *20*(3), 235–244. http://doi.org/10.1016/S0272-7757(99)00066-7