



Examining the Carnegie Classification Methodology for Research Universities

Robert Kosar & David W. Scott

To cite this article: Robert Kosar & David W. Scott (2018) Examining the Carnegie Classification Methodology for Research Universities, *Statistics and Public Policy*, 5:1, 1-12, DOI: [10.1080/2330443X.2018.1442271](https://doi.org/10.1080/2330443X.2018.1442271)

To link to this article: <https://doi.org/10.1080/2330443X.2018.1442271>



© 2018 The Author(s). Published with license by Taylor & Francis Group, LLC
Robert Kosar and David W. Scott



Accepted author version posted online: 26 Feb 2018.
Published online: 18 Apr 2018.



Submit your article to this journal [↗](#)



Article views: 2648



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

Examining the Carnegie Classification Methodology for Research Universities

Robert Kosar and David W. Scott

Department of Statistics, Rice University, Houston, TX

ABSTRACT

University ranking is a popular yet controversial endeavor. Most rankings are based on both public data, such as student test scores and retention rates, and proprietary data, such as school reputation as perceived by high school counselors and academic peers. The weights applied to these characteristics to compute the rankings are often determined in a subjective fashion. Of significant importance in the academic field, the Carnegie Classification was developed by the Carnegie Foundation for the Advancement of Teaching. It has been updated approximately every 5 years since 1973, most recently in February 2016. Based on bivariate scores, Carnegie assigns one of three classes (R1/R2/R3) to doctorate-granting universities according to their level of research activity. The Carnegie methodology uses only publicly available data and determines weights via principal component analysis. In this article, we review Carnegie's stated goals and the extent to which their methodology achieves those goals. In particular, we examine Carnegie's separation of aggregate and per capita (per tenured/tenure-track faculty member) variables and its use of two separate principal component analyses on each; the resulting bivariate scores are very highly correlated. We propose and evaluate two alternatives and provide a graphical tool for evaluating and comparing the three scenarios.

ARTICLE HISTORY

Received April 2017
Accepted January 2018

KEYWORDS

Data transformation;
Dimension reduction; Factor
rotation; Research university
rankings

1. Introduction

The Carnegie Classification (CC), now under the auspices of Indiana University Bloomington's Center for Postsecondary Research (Indiana University 2016), is one of the oldest regularly published rankings of university programs and reputations for doctorate-granting universities. It has been described as “the dominant classification system” for higher education research (Brewer, Gates, and Goldman 2002) and as “...the standard taxonomy for American higher education” (Graham and Diamond 1997). Published approximately every 5 years since 1973 and most recently in February 2016, the ranking currently assigns each of 335 US universities to one of three clusters: R1/R2/R3 or Highest/Higher/Moderate Research Activity, respectively. A stated goal of the classification is to provide a framework for researchers to compare programs among peer institutions (McCormick and Zhao 2005; Zhao 2011). Many universities designated as R2 have created institutional goals and research plans that include achieving an R1 rank in 5 or 10 years (The University of Nevada Las Vegas 2014; University of Houston 2015). Carnegie has often stated their classification is not a ranking and was not created for this purpose (Shulman 2001; McCormick 2005). However, the original Carnegie Classification described the R1 category as the “The 50 leading universities in terms of federal financial support of academic science...” (Carnegie Commission on Higher Education 1973) and the ordering of the current classes seems clear, so we shall use “ranking” and “classification” interchangeably.

The most prominent university ranking system in the United States (U.S.) for a wide variety of institutions and individual degree programs is published annually by U.S. News and World Report (USNWR). USNWR uses the institutional categories as defined by the Carnegie Foundation. Both USNWR and Carnegie create scores based upon a data matrix. While both solicit some data from the universities themselves, all of the Carnegie Rankings (CR) data are publicly available and published on their website. Conversely, USNWR solicits subjective data from high school counselors and academic peers that are unpublished. From a statistician's perspective, the major difference between the methodologies used to create the scores is the selection of weights applied to the data matrix. USNWR uses a subjective approach while CR uses weights derived from principal component analysis (PCA), which is a standard multivariate statistical technique. In the final analysis, USNWR provides a numerical ranking of the form 1, 2, . . . , N whereas CR only provides the three clusters mentioned above (Morse, Brooks, and Mason 2015).

The Carnegie Classification has evolved over time. The 1973, 1976, 1987, and 1994 iterations roughly evenly split the top 200 doctorate-producing universities into four categories based on federal spending and achieving a doctorate production above a certain threshold (50 doctorates per year to be classified as a research university) (Carnegie Commission on Higher Education 1973; Carnegie Council on Policy Studies in Higher Education 1976; The Carnegie Foundation for the

Advancement of Teaching 1987, 1994). The short-lived 2000 edition essentially combined the top two and bottom two categories and divided the resulting two categories based on how many doctorates schools granted without regard to research spending (The Carnegie Foundation for the Advancement of Teaching 2000). The latest scheme used in the 2005, 2010, and 2015 iterations of the Carnegie Classification is much more sophisticated than the earlier versions. It uses a more diverse set of variables and advanced statistical techniques to characterize schools. Despite the strengths of the modern classification, further refinement is still possible.

In this article, we examine in detail several aspects of Carnegie's methodology: Carnegie's data matrix, Carnegie's application of principal components, and Carnegie's formation of the three clusters in light of its stated goals. We also propose and examine two modifications that we believe are consistent with Carnegie's stated goals, more faithfully represent the structure of the data, and offer organizational advantages. We then evaluate their impact on the final clustering. The first modification involves the specific application of principal components to the data matrix. The second modification additionally involves the augmentation and redefinition of the data matrix that defines the per capita variables. We discuss the relative merits of these alternatives and Carnegie's technique.

2. Carnegie Methodology Synopsis

The Carnegie Classification for research universities is based on research expenditures, number of research staff, and number of doctorates granted. All values are annual. Research expenditures are divided into two categories: science, technology, engineering, and mathematics (STEM) spending and non-STEM spending. Doctorates are split into four categories: humanities, social science, STEM, and "professional." These seven variables are each transformed to ranks, which reduces the influence of extreme skewness and outliers. We shall assume that the rank-transformed data have been centered at zero and scaled to have unit variance. Ties are broken by taking the minimum assignable rank, rather than the average, which we believe is consistent with Carnegie's methodology. The first principal component score of these aggregate ranks (computed with loadings C_A) is combined with the first principal component score of per capita versions of the research expenditures and support staff (computed with loadings C_P) to produce a bivariate set of scores for each school. Note that there are two first principal components because there are two separate PCAs done on separate sets of variables. The per capita variables are the STEM and non-STEM research expenditures and the number of research staff each divided by the number of research faculty, transformed to ranks, and standardized. The final classification of each school is based on a standardized (Carnegie does not specify the exact formula) distance of the Carnegie principal component scores from a particular point (again unspecified, but approximately the minimum of each score). The very-high-research category is the set of schools farthest away from this point (The Carnegie Classification of Institutions of Higher Learning 2016).

The result of this transformation and subsequent classification as computed by the Carnegie Foundation is given in Figure 1 (The Carnegie Classification of Institutions of

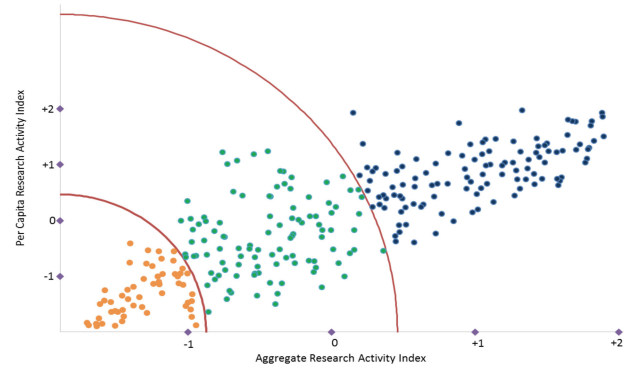


Figure 1. The Carnegie Foundation scores (figure taken from their website The Carnegie Classification of Institutions of Higher Learning 2016). The R3, R2, and R1 schools are shown in orange, green, and blue, respectively.

Higher Learning 2016). We provide a schematic of the Carnegie Foundation's approach in Figure 2, which illustrates how the aggregate and per capita variables are treated separately.

3. The Carnegie Data Matrix

The raw input for the Carnegie Classification comes from two sources: the Integrated Postsecondary Education Data System (IPEDS) and the National Science Foundation (NSF). IPEDS consists of a collection of surveys conducted under the auspices of the U.S. Department of Education. Participation in these surveys is required for schools that take part or apply to take part in federal student financial aid (IPEDS 2016). Reporting requirements are extensive, with some institutions employing multiple employees for IPEDS alone (Field 2008). The NSF conducts two surveys relevant to the doctoral university classification: the Higher Education Research and Development Survey (HERD), and the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS). Both have relatively low nonresponse rates of 3.2% and 0.3%, respectively (HERD 2016; GSS 2016).

The doctoral degree and faculty counts come from the 2013–2014 IPEDS data and fall 2013 IPEDS data, respectively. Here faculty is assistant, associate, and full professors. "Professional" doctorates consist of doctorates related to professional fields, but not so-called "professional practice" degrees such as M.D.'s and J.D.'s. Research expenditures are derived from the 2014 HERD. Institutional demarcation for research expenditures is somewhat unclear. For example, Johns Hopkins University more than doubles its reported expenditures by including its Applied Physics Laboratory (HERD 2017), which is an independent institution according to its website (APL 2016). This is not necessarily an error and Hopkins would still have a very high-ranked amount of research spending without this amalgamation; its class is not in doubt. However, this fact does illustrate that the inputs to the Carnegie Classification are partially affected by some subjective choices made by parties that may benefit from not choosing the conservative option. Finally, the number of doctorate-holding research staff is provided by the 2013 GSS. We provide estimates of the distributions of the variables used in Carnegie's system and our proposed systems in Figure 3 and summary statistics in Tables 1 and 2.

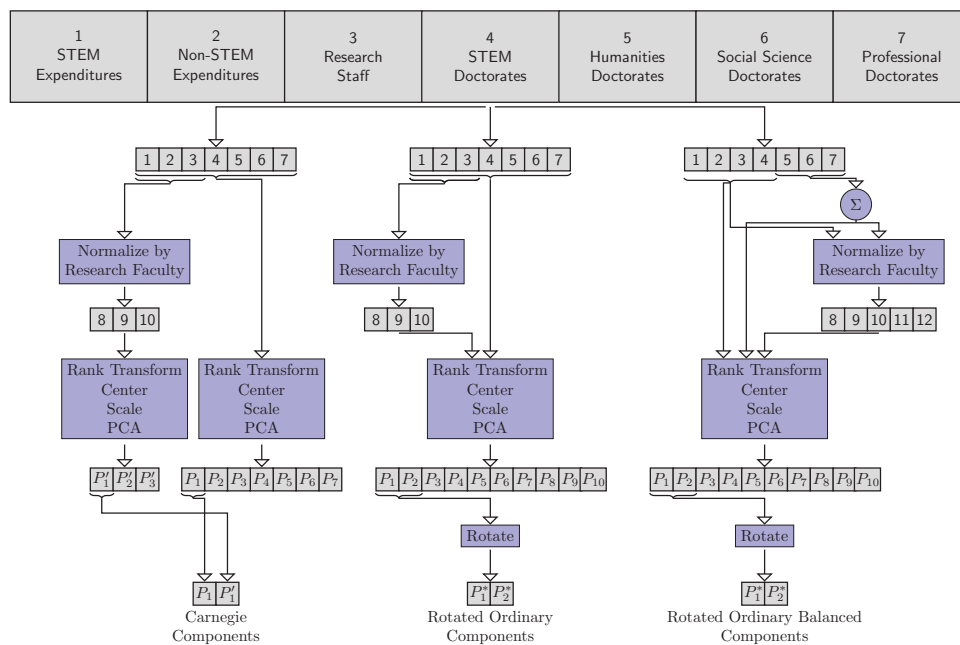


Figure 2. Schematic for showing how the variables are used in the three scoring systems described in this article. (1) The left scheme is currently used by the Carnegie Foundation. Two separate principal component analyses are performed on two disjoint, but highly correlated, sets of variables. The two first principal component scores form the basis of the Carnegie Foundation's classification scheme, pictured in Figure 1. (2) The middle scheme uses the same variables as the Carnegie method, but a single two-dimensional principal component analysis. We then perform a rotation to retain the aggregate and per capita interpretation of the Carnegie scores. (3) The right scheme is the same as the middle scheme, except that we balance the number of per capita and aggregate variables. Our proposed schemes are detailed in Section 7 but are included here for comparison.

Table 1. Summary statistics of the aggregate variables. Non-STEM doctorates is the sum of the last three categories of doctorates. Carnegie does not use the combined non-STEM doctorates.

	STEM expenditures USD (thousands)	Non-STEM expenditures USD (thousands)	Research staff	STEM doctorates	Non-STEM doctorates	Humanities doctorates	Social science doctorates	Professional doctorates
Min.	227	0	0	0	0	0	0	0
1st Q.	19,300	1260	14	15	31	0	2	21
Median	67,100	5190	56	51	70	8	13	39
Mean	198,000	11,200	273	100	105	25	22	58
3rd Q.	284,000	13,600	316	138	143	34	34	75
Max.	2,230,000	124,000	7300	589	535	179	129	495

4. Effect of the Rank Transform

Prior to being subjected to PCA, the input variables for the Carnegie Classification are transformed to ranks. This has a tendency to reduce differences between schools with large values on particular dimensions, which is advantageous because it mitigates the influence of outlying values on the subsequent PCA. However, it also has some negative effects. As Figures 1 and 4 demonstrate, transforming to ranks exaggerates differences between schools with small values along these dimensions. For example, there is a difference of only eight humanities

doctorates per year between the lowest ranked schools and the median ranked schools for that dimension. Conversely, the highest ranked school in that dimension produces more than 20 times that difference.

Additionally, transforming to ranks tends to mitigate even obvious clustering structure. As an example of this, Figure 5 demonstrates that even well-separated clusters may no longer be apparent after applying the rank transformation. In the top panels, clusters that are well separated in a dimension that is then rank transformed are difficult to distinguish. In the middle panels, clusters that are both correlated and separated along an axis

Table 2. Summary statistics of the per capita variables. Carnegie uses only the first three variables in their scheme.

	STEM expenditures USD (thousands) per capita	Non-STEM expenditures USD (thousands) per capita	Research staff per capita	STEM doctorates per capita	Non-STEM doctorates per capita
Min.	0.75	0.00	0.00	0.00	0.00
1st Qu.	42.10	2.59	0.03	0.03	0.06
Median	121.00	6.92	0.10	0.08	0.10
Mean	177.00	11.30	0.23	0.10	0.12
3rd Qu.	235.00	13.90	0.27	0.13	0.15
Max.	3950.00	93.00	6.12	0.69	1.73

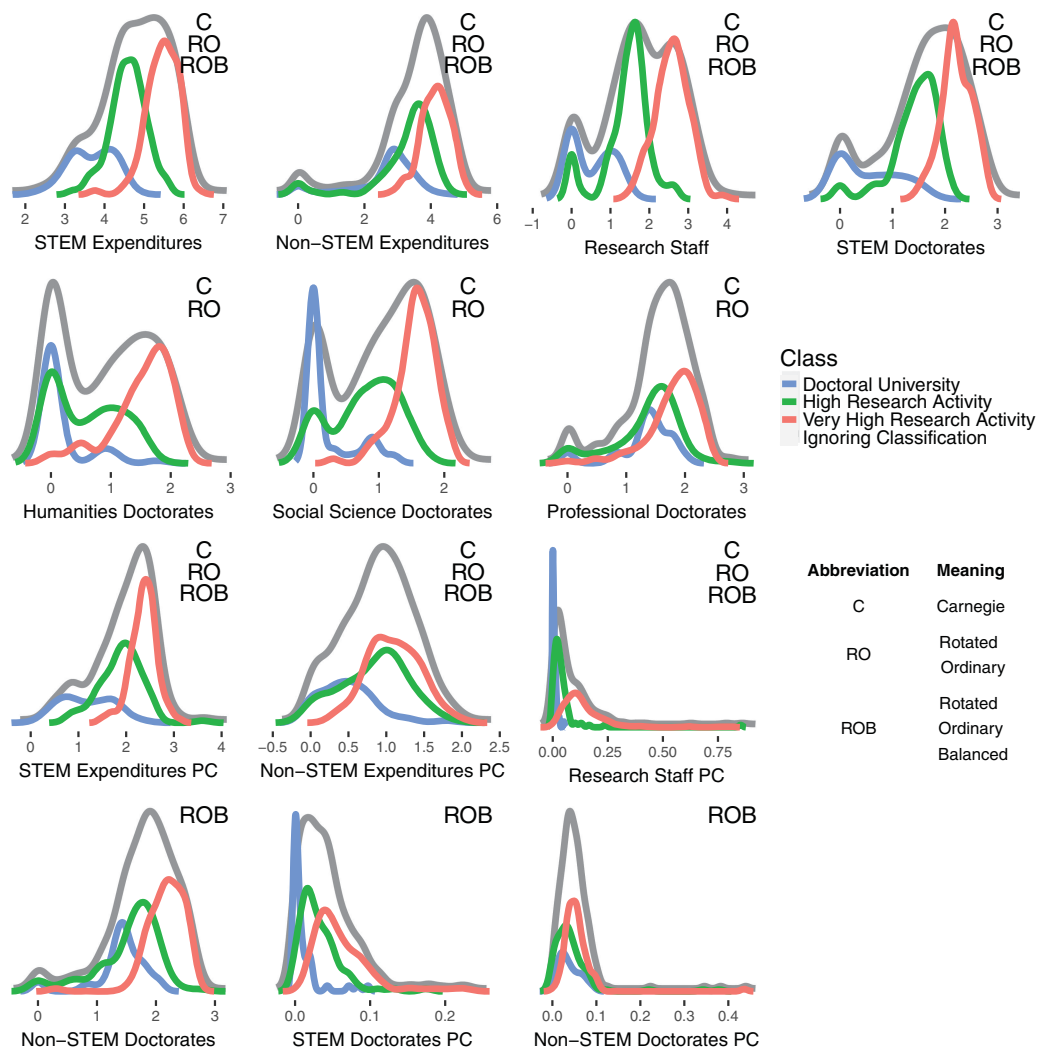


Figure 3. Density curves of the 13 metrics listed in Tables 1 and 2 used to classify schools. All values have been positively shifted by one and log-transformed. Expenditures are in thousands prior to this \log_{10} transformation. Colors correspond to the Carnegie categories. The gray curve is for the combined classes. The abbreviations indicate which variables are used in which schemes. Descriptions of the RO and ROB schemes can be found in Section 7.

are made difficult to distinguish by taking the rank transform of all dimensions. However, it is not always the case that the rank transform destroys the clustering structure. This is demonstrated in the bottom pair of panels where clusters are separated but not individually correlated with respect to an axis that is not itself rank transformed. In practice, the three common transforms of (1) simply standardizing, (2) log transforming, and (3) rank transforming often produce similar results when combined with PCA (Baxter 1995). We found that replacing the rank transform with a log transform offered no advantages for Carnegie's purposes and required arbitrary translation of the data away from zero. Despite this general consistency across transforms, the above caveats about the rank transform still apply since Carnegie is attempting to distinguish clusters along an axis with which the individual clusters are correlated.

5. Carnegie's Principal Component Analysis

Principal component analysis is a dimension reduction technique that seeks to capture maximal variance of the original data. One of the characteristics of a typical PCA is that the component scores are uncorrelated. However, since the

Carnegie Classification is based on the first principal components of two closely related sets of variables, the two component scores have correlation 0.84. If the numbers of doctorates in each category are not included in the aggregate measure, the correlation between the two component scores is 0.93. Evidently the per capita metrics are very similar to the aggregate metrics except they contain less information about non-STEM research. The fraction of total variation accounted for by the Carnegie components is 0.70, which is similar to the fraction accounted for by the single first component of all the variables 0.66 and not as much as is accounted for by the first two components of all variables 0.79. The fraction for each component is given in Table 4. Here, the fraction of total variation is taken to be the trace of the residual covariance divided by the trace of the covariance (correlation) matrix of all the variables. We define the residual to be the original 10-dimensional (7 + 3) point minus the specified components (Carnegie, or ordinary principal). The second ordinary component indicates a natural contrast between STEM research and other research that is not accounted for in the Carnegie components. The loadings for the schemes described in this article are given in Table 3. All loading signs have been chosen to be consistent with the Carnegie methodology.

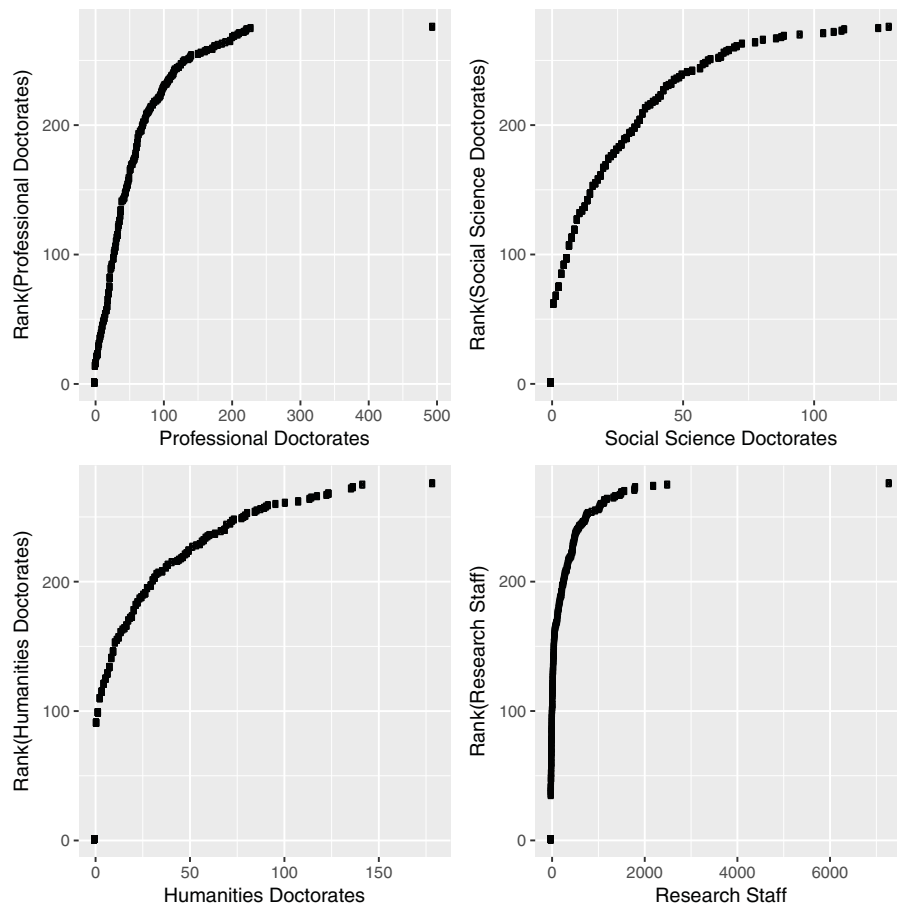


Figure 4. Rank versus nominal value for selected variables. The rank transform tends to exaggerate differences between low-ranked schools; however, it mitigates differences between high-ranked schools.

Table 3. PCA loadings of the Carnegie, ordinary, and our proposed (detailed in Section 7) schemes. Blanks indicate that a variable was not included in the formation of the loading. Per capita variables are highlighted in gray.

	C_A	C_P	First ordinary	Second ordinary	Rotated first ordinary	Rotated second ordinary	Rotated first balanced ordinary	Rotated second balanced ordinary
STEM expenditures	0.41		0.36	0.22	0.20	0.38	0.31	0.25
Non-STEM expenditures	0.36		0.31	−0.37	0.46	−0.16	0.38	−0.17
Research staff	0.41		0.37	0.22	0.20	0.38	0.32	0.22
STEM doctorates	0.41		0.36	0.15	0.23	0.31	0.33	0.20
Humanities doctorates	0.37		0.30	−0.24	0.38	−0.05		
Social Science doctorates	0.39		0.32	−0.22	0.39	−0.02		
Professional doctorates	0.28		0.20	−0.47	0.42	−0.30		
STEM expenditures PC		0.64	0.33	0.36	0.09	0.48	0.25	0.35
Non-STEM expenditures PC		0.42	0.22	−0.42	0.41	−0.25	0.31	−0.21
Research staff PC		0.64	0.34	0.33	0.12	0.46	0.28	0.29
Non-STEM doctorates							0.41	−0.37
STEM doctorates PC							0.23	0.31
Non-STEM doctorates PC							0.31	−0.58

6. Carnegie's Clusters

What clustering method is being used to determine how schools are categorized is not clear. The classification made by

Table 4. Percentage of variance explained by Carnegie's two 1-D PCAs compared to the ordinary 2-D PCA and our proposed rotation of the ordinary 2-D PCA discussed in Section 7.

	Carnegie	Ordinary	Rotated ordinary
1	48.94	65.57	51.51
2	21.06	13.30	27.36

the Carnegie Foundation is given in Figure 6. Ignoring the color cues, it is not visually obvious that there are any distinct or natural clusters. If there are indeed three true clusters, applying k -means clustering gives a different, albeit similar, classification. If a normal mixture distribution is fitted using the `mclust` (Fraley and Raftery 2002) package, having three clusters is determined to be optimal from an information criterion perspective; however, the three clusters are quite different from the three clusters given by the Carnegie Foundation and k -means clustering, as can be seen in the bottom left panel of Figure 6. Overall, there does not appear to be a reason to consider a given school more similar to schools in its assigned cluster than to other schools

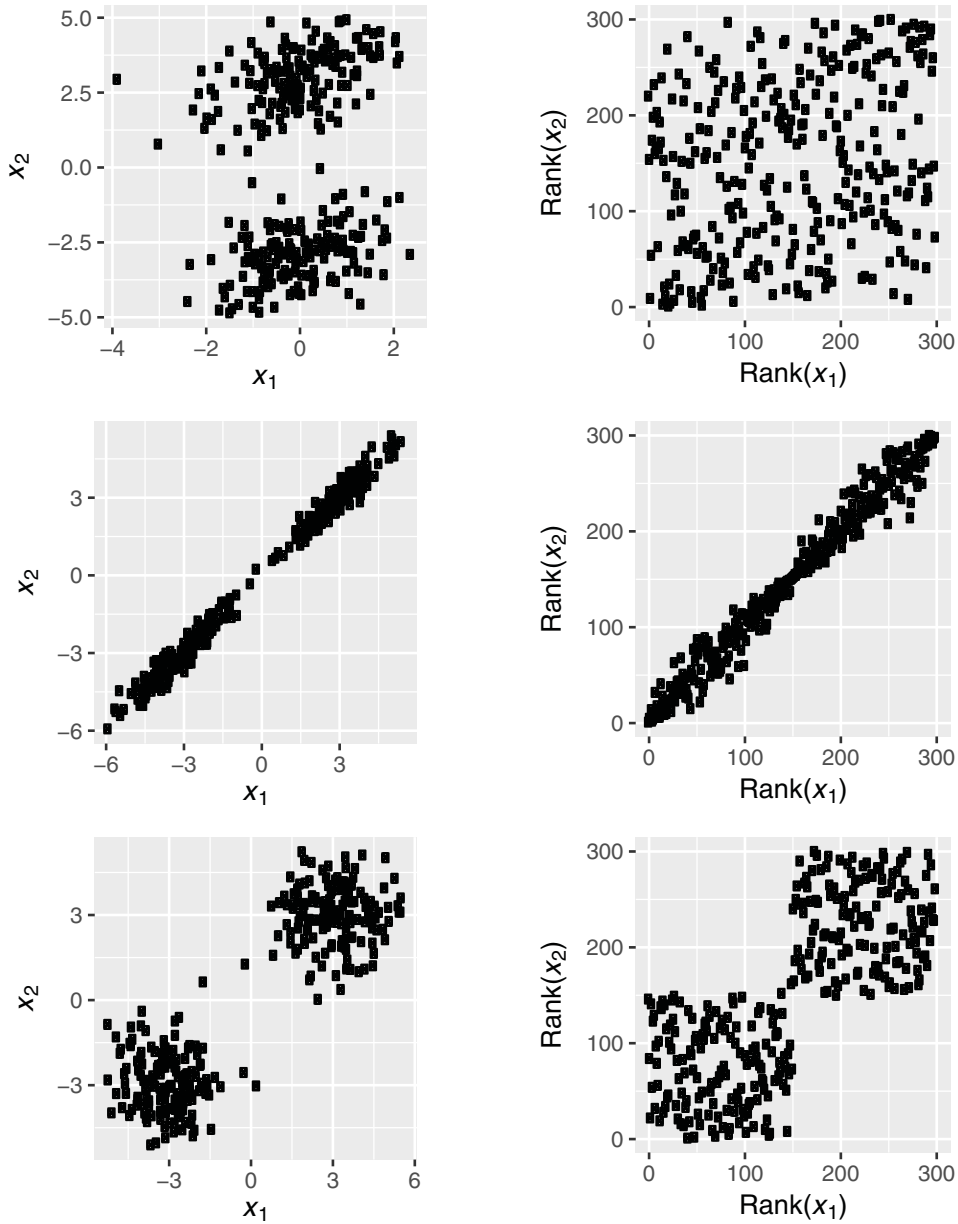


Figure 5. The effect of the rank transform on clustering structure with simulated data. The rank transform tends to reduce the separation between clusters (top panels), especially when variables are correlated and the separation is along the diagonal (middle panels). However, the clusters in the bottom panels are still easily visible after the rank transform. These data are simulated from bivariate normal distributions.

with more similar principal component scores that happen to lie across a cluster boundary determined by one of these methods.

7. Alternative Proposals

Rather than employing an ad hoc procedure to capture the aggregate and per capita components of the variation between schools, we propose using a rotation of the first two principal components such that the magnitude of the loadings is large for the aggregate measures on one component and large for the per capita measures on the other. To this end, define the diagonal matrices

$$D_A =: \text{diag}([1, 1, 1, 1, 1, 1, 1, 0, 0, 0])$$

$$D_{PC} =: \text{diag}([0, 0, 0, 0, 0, 0, 0, 1, 1, 1])$$

and let \mathbf{o}_1 and \mathbf{o}_2 be the first two ordinary principal component loadings for the combined Carnegie variables. By construction, \mathbf{o}_1 and \mathbf{o}_2 are orthogonal and of unit length. They may be rotated by introducing a constant $\alpha \in (-1, 1]$. Let

$$\mathbf{r}_1(\alpha) =: \alpha \mathbf{o}_1 + \sqrt{1 - \alpha^2} \mathbf{o}_2$$

$$\mathbf{r}_2(\alpha) =: -\sqrt{1 - \alpha^2} \mathbf{o}_1 + \alpha \mathbf{o}_2.$$

It is easy to verify that \mathbf{r}_1 and \mathbf{r}_2 are orthonormal and that the angle of rotation θ is such that

$$\cos(\theta) = \mathbf{r}_1^t \mathbf{o}_1 = \mathbf{r}_2^t \mathbf{o}_2. \quad (1)$$

Since principal component loadings with opposite signs are equivalent (we choose to have positive STEM signs for consistency with Carnegie), a rotation of θ is equivalent to a rotation of $180 + \theta$. We report the smaller in magnitude of the two as

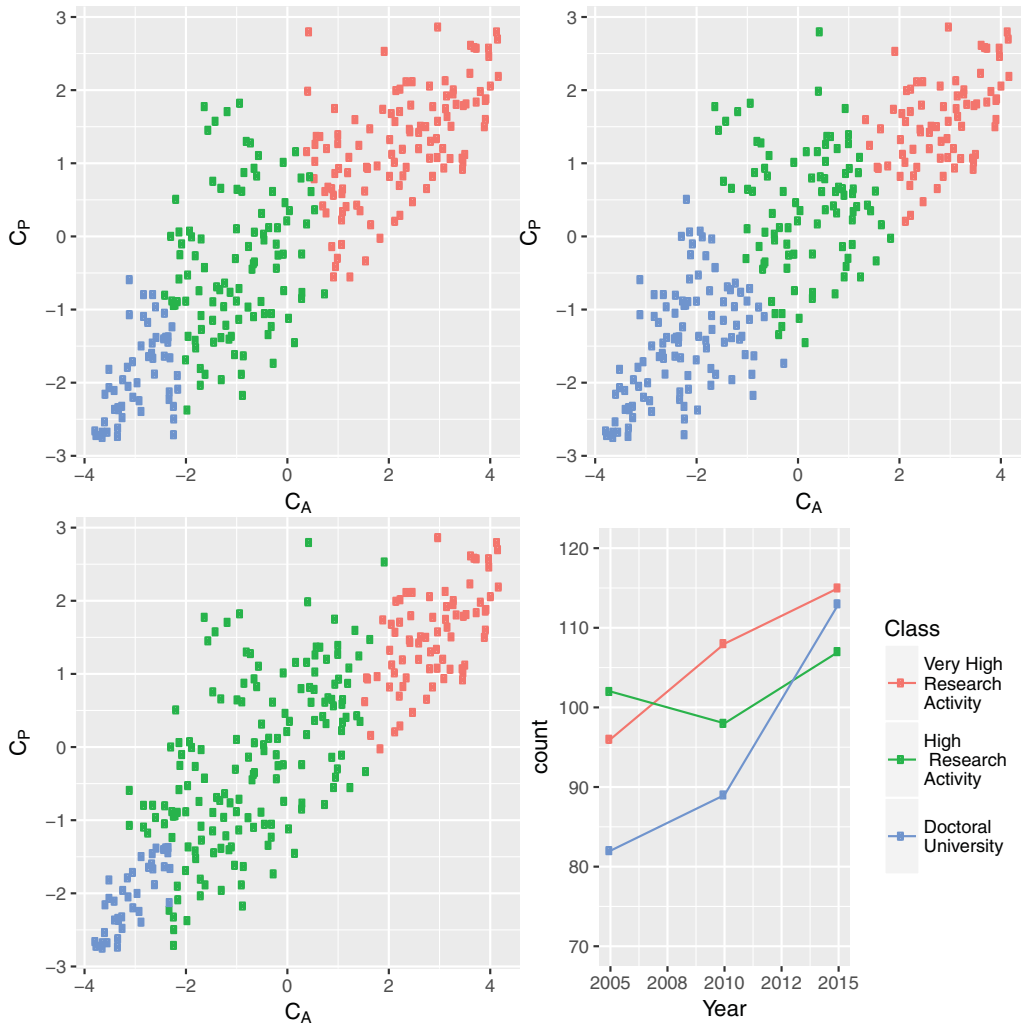


Figure 6. Carnegie clustering (top left), k -means ($k = 3$) clustering (top right), normal mixture clustering (bottom left), and Carnegie cluster counts for the past three iterations (bottom right). The total number of research universities in 2005, 2010, and 2015 was 280, 295, and 335, respectively.

the angle of rotation but always define θ as in Equation (1) for consistency with convention. We seek

$$\alpha^* = \operatorname{argmax}_{\alpha \in (-1, 1]} \|D_A \mathbf{r}_1(\alpha)\|^2 + \|D_{PC} \mathbf{r}_2(\alpha)\|^2. \quad (2)$$

Note that the norm is the Euclidean norm and that this rotation is very similar to the varimax rotation, which is one of the most popular (Jolliffe 2002) rotation criteria. This technique has a number of advantages. These two components account for the same amount of variance as the first two ordinary components. The parameterization of this rotation makes reference to the first two principal components easy. The transition from the Carnegie components to the standard principal components and these proposed components is given in Figure 8. This method is based on the same variables as those used by the Carnegie Foundation and we shall refer to it as the rotated and ordinary (RO) scheme. However, the Carnegie system may benefit from defining some new variables and combining some of the current variables.

Given the sensitivity of the location of some schools to small changes in the non-STEM doctoral variables, it may be desirable to sum them into a single category. It is unclear why the doctorates awarded by schools are not currently included among

the per capita variables. It does not seem implausible that the number of students graduated per research faculty member is important for characterizing schools. We also consider a version of the above proposed technique with the non-STEM doctorates summed into a single variable, and the number of STEM and non-STEM doctorates per capita as additional covariates. This balanced version of the data matrix has five aggregate variables and five per capita variables. The loadings are rotated in a way analogous to Equation (2). We shall refer to this second scheme as the rotated, ordinary, and balanced (ROB) scheme. Both schemes are visualized in Figure 2. The optimal angles of rotation for the RO and ROB schemes are 31 degrees and 19 degrees, respectively. The objective functions leading to these optima are pictured in Figure 7. As a result of these rotations, the principal component scores are somewhat correlated ($\hat{\rho}_{RO} = 0.62$, $\hat{\rho}_{ROB} = 0.41$) even though the loadings (given in Table 3) themselves are orthogonal.

To illustrate how much the new scoring systems would change the current classification system, we have added hypothetical boundaries to the bottom two panels of Figure 8. We attempted to preserve the current classification boundary as much as possible. Summary two-way tables are given in Table 5. These tables are sensitive to the precise choice of the boundary.

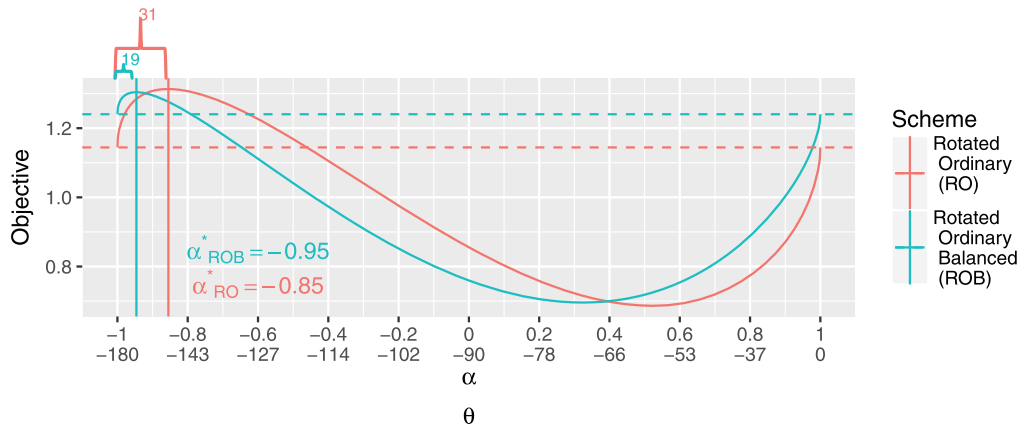


Figure 7. Objective function plotted over the domain of α for Equation (2). The dashed lines indicate the value of the objective function for no rotation. The vertical lines indicate the value of α^* for the RO and ROB schemes. Note that a 180 degree rotation is equivalent to a zero degree rotation since principal components with opposite signs are equivalent. Therefore, the positive values reported as the optimal rotation in the upper left corner are equivalent to the horizontal axis values of θ which is defined according to convention in Equation (1).

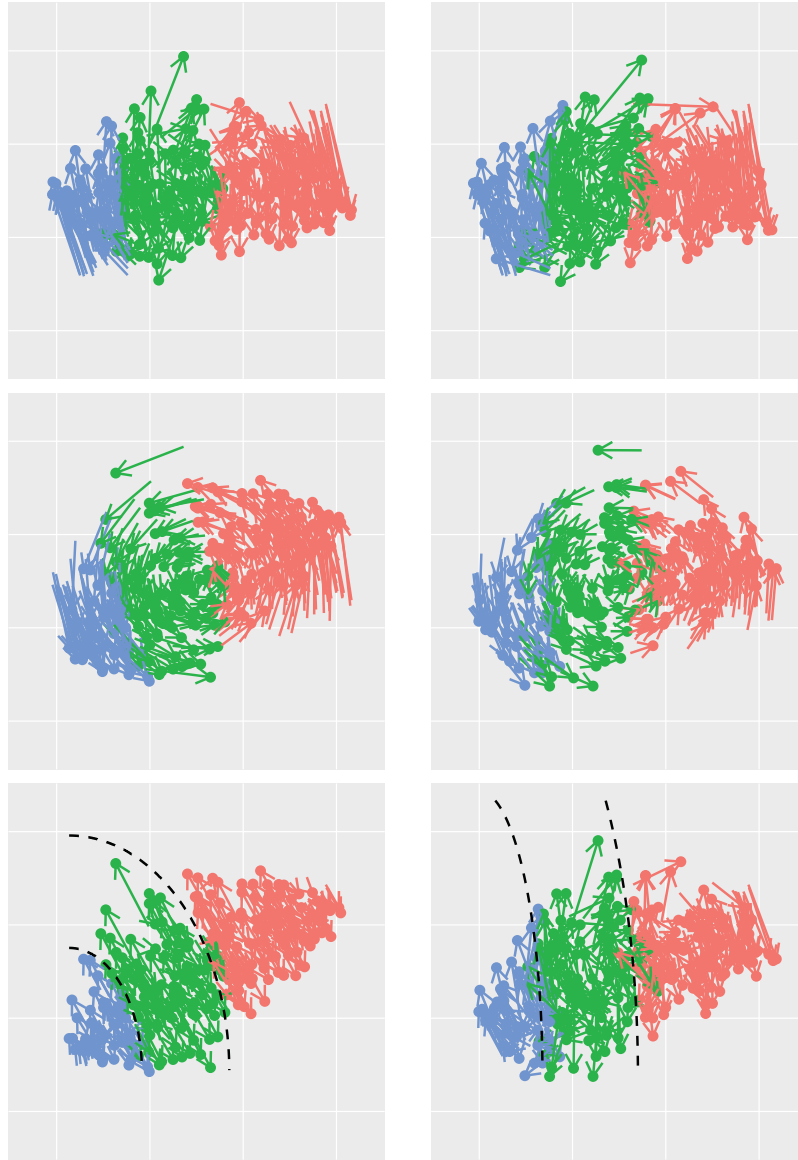


Figure 8. The panels in the left and right columns correspond to the RO and the ROB scores, respectively. The top row illustrates the transition from the Carnegie scores to the ordinary PCA scores; the middle row illustrates the effect of the rotation; the bottom row is the ultimate difference between the Carnegie scores and the proposed scores resulting from the top rows. Note that the combined effect for both schemes is largely to add an orthogonal component because we use a true second principal component. We have added hypothetical class boundaries to illustrate the minimal change in classes caused by the new scoring systems.

Table 5. Comparison of hypothetical classifications under the RO and ROB schemes and the current Carnegie scheme.

	Doctoral university (RO)	High research activity (RO)	Very high research activity (RO)	Doctoral university (ROB)	High research activity (ROB)	Very high research activity (ROB)
Doctoral university	51	3	0	40	14	0
High research activity	5	96	6	4	96	7
Very high research activity	0	0	115	0	1	114

8. Differential Effects of Investment

The sensitivity of a school's location within the Carnegie scheme as a result of the rank transform combined with the great difference in unit cost leads to an interesting calculus for schools that wish to move up in the Carnegie ranking. Figure 9 illustrates how much selected schools would move in the Carnegie scheme if they were to spend an additional \$10⁶ on a particular Carnegie variable assuming an annual unit cost given in the table at the top of the figure. Clearly spending money on some dimensions is much more effective than others for schools seeking to move up in the ranking. We provide an interactive graphical tool in the next section that allows users to manipulate the characteristics of individual schools and visualize the resulting movement in the Carnegie and our alternative schemes.

9. Viewing Application

We have created a Shiny (Chang et al. 2016) web application, pictured in Figure 10, to illustrate the three schemes described in this article.

The application displays the location of all the doctoral universities following the three transformations described in this article. Individual schools can be selected from the dropdown menu on the left. Once a school is selected, that school's marker in the plots is highlighted. Starting from the true values, users can adjust the input variables for a particular school by moving the sliders. To reset the sliders to the true values, users can press the "Reset to Truth Button" button. The extrema of the sliders are determined by the extrema of the variables for the doctoral universities so that an individual school can be changed to have the highest or lowest rank. Since the principal components are calculated based on the ranks of the schools, having even more extreme values would have no effect. If the slider values are changed, the new rank is computed by holding every other school constant. The transformation loadings are not changed by changing the sliders. To view the school corresponding to a particular point, users can click on that point. To view a larger version of a particular set of components, users should select the corresponding radio button. Some caution must be used when drawing conclusions from this exploratory application.

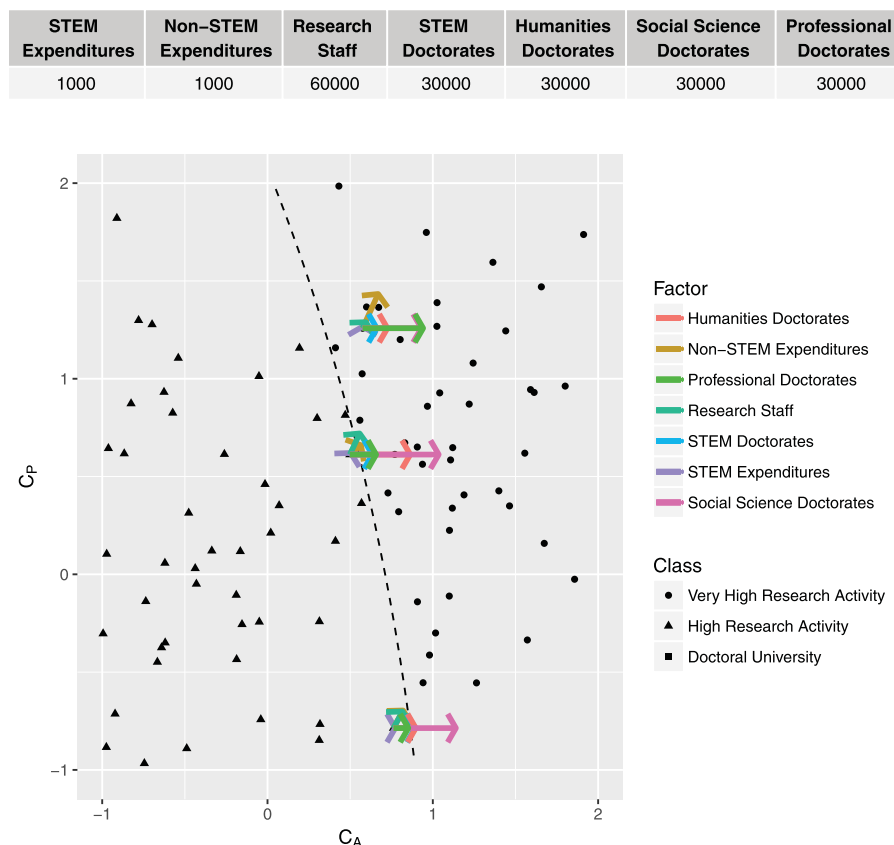


Figure 9. Hypothetical costs (top) and shifts (bottom) in selected schools' positions in the Carnegie scheme resulting from spending \$10⁶ extra on individual research dimensions. For example, all three schools shown here would benefit substantially in the Carnegie scheme by spending money to produce more social science doctorates, but the top school would similarly benefit from investing in professional doctorates. Note that the dashed boundary is approximate.



Certain changes to the attributes of a particular school may not be possible. For example, lowering the number of research faculty at a particular school may improve its location relative to the other schools since large per capita numbers are favored by the Carnegie methodology. However, even though this change is possible in the application, in reality it is probably not possible to lower this number without affecting the other dimensions.

10. Discussion

The Carnegie Classification of research universities is an effective and transparent research tool for many reasons. Carnegie publishes the data they use. Carnegie provides enough details about its methodology to (at least approximately) reproduce its scoring. Carnegie endeavors to use a well-established statistical method, PCA, to produce scores that represent the structure of the underlying data. It also employs transformations to limit the impact of outliers. Finally, Carnegie cautions against using its classification system as a ranking, which is appropriate given the uncertain nature of the divisions.

However, there are ways in which the Carnegie system may be refined. Carnegie uses the first components from two separate one-dimensional principal component analyses rather than one two-dimensional analysis and therefore does not optimally represent the underlying structure of the data from a variance perspective. We seek to improve this by computing the first two ordinary principal components, which accounts for more variance, and applying a rotation that attempts to preserve the interpretation of one loading representing aggregate research and the other representing per capita research. This largely succeeds. More variance is accounted for. However, the new second component as much contrasts how much STEM research versus non-STEM research is done at a particular school as it represents the per capita research output of schools. The second Carnegie component also measures faculty composition rather than per capita research output to some extent, as it heavily weights STEM over non-STEM research. It appears that there is more of a distinction between STEM and non-STEM research universities after accounting for aggregate research than there is wide variance in per capita output.

It is unclear why it is useful to separate non-STEM doctorates into three categories. Perhaps the fact that this favors non-STEM oriented schools helps balance the fact that research spending favors STEM (McCormick 2013), but it makes schools' positions very sensitive to small changes in doctorates granted for these separate categories. If it makes sense to separate these categories, why not decompose STEM? We propose combining the three non-STEM categories and also including per capita versions in our balanced alternative to the Carnegie system. As with the rotated ordinary principal components on the original data, the rotated ordinary principal components on this modified data mainly represent aggregate research in the first component and STEM versus non-STEM research in the second.

The class boundaries in the Carnegie scheme can be replaced with straight lines and approximately maintain the current classification. This is further evidence that the two components that make up the Carnegie scheme contain the same information. In our alternatives to the Carnegie scheme, roughly vertical lines

can be used to divide the Carnegie classes because our first component is similar to Carnegie's two components.

Given that all three systems produce useful bivariate scores (we assert that our proposed scores contain the most useful information), it seems unnecessary to insist upon discretization. Our interpretation of the first two ordinary principal components for Carnegie's data is that a majority of the variance in universities comes from aggregate research and STEM versus non-STEM composition. We consider neither concentration on STEM nor aggregate research output appropriate measures on which to rank schools because it is not clear that more of either is better. Researchers can select their own cohorts based on the scores rather than a classification that may not be well-suited for their purposes. For consistency with the historic Carnegie methodology and to emphasize per capita research, we suggest using our proposed rotation as well.

The focus of this article has been on how to summarize multivariate information about research universities given metrics that have already been chosen by Carnegie. The choice of what metrics to include is also of great importance. As outlined in the introduction, the Carnegie Classification (CC) has traditionally been based on research expenditures and doctorate production. The CC could be improved by including additional research metrics that have become available since its creation in the 1970s. In particular, we think that the inclusion of citation data would be useful for measuring research output. Citation quantity is not a perfect metric of the research quality of a University, but it does have some nice attributes:

- It is not based on a survey whose answers depend on a small number of people.
- It reflects the opinions of researchers rather than funding agencies.
- It seems, at least naively, to favor STEM less than research funding.
- It could fit neatly into Carnegie's aggregate versus per capita scheme.

One of the best qualities of the CC is that the "raw" data are made available by Carnegie. We would not desire the CC to lose that attribute to include more data. However, if Carnegie can find a source of citation data that they can publish, we think it would make a useful addition.

Funding

This work was partially supported by the sabbatical program of the Mathematical Sciences program of the National Security Agency, as well as a Rice University graduate fellowship administered by the Ken Kennedy Institute and funded by Shell Co.

References

- APL (2016), "About APL," available at <http://www.jhuapl.edu/aboutapl/default.asp>. [2]
- Baxter, M. J. (1995), "Standardization and Transformation in Principal Component Analysis, With Applications to Archaeometry," *Applied Statistics*, Series C, 44, 513–527. [4]
- Brewer, D., Gates, S., and Goldman, C. (2002), *In Pursuit of Prestige: Strategy and Competition in U.S. Higher Education*, New Brunswick, NJ: Transaction Publishers. [1]

- Carnegie Commission on Higher Education (1973), *A Classification of Institutions of Higher Education*, Berkeley, CA: The Carnegie Foundation for the Advancement of Teaching. [1]
- Carnegie Council on Policy Studies in Higher Education (1976), *A Classification of Institutions of Higher Education: Revised Edition*, Berkeley, CA: The Carnegie Foundation for the Advancement of Teaching. [1]
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2016), *shiny: Web Application Framework for R*, r Package Version 0.13.2. Available at <https://CRAN.R-project.org/package=shiny> [9]
- Field, K. (2008), "Pending Bill Would Double Colleges' Reporting Burden, Critics Say," *The Chronicle of Higher Education*. Available at <https://www.chronicle.com/article/Pending-Bill-Would-Double/13340> [2]
- Fraley, C., and Raftery, A. E. (2002), "Model-Based Clustering, Discriminant Analysis and Density Estimation," *Journal of the American Statistical Association*, 97, 611–631. [5]
- Graham, H. D., and Diamond, N. (1997), *The Rise of American Research Universities: Elites and Challengers in the Postwar Era*, Baltimore, MD: JHU Press. [1]
- GSS (2016), "Survey of Graduate Students and Postdoctorates in Science and Engineering (About the Survey)," available at <http://www.nsf.gov/statistics/srvygradpostdoc/#sd&q&tabs=1> [2]
- HERD (2016), "Higher Education Research and Development Survey (About the Survey)," available at <http://www.nsf.gov/statistics/srvyherd/#sd> [2]
- (2017), "HERD Table 17. Higher Education R&D Expenditures," Available at https://ncesdata.nsf.gov/herd/2014/html/HERD2014_DS_T_17.html [2]
- Indiana University (2016), "IU Research Center to House Carnegie Classification of Institutions of Higher Education," available at <http://news.indiana.edu/releases/iu/2014/10/carnegie-classifications-move-to-iu.shtml> [1]
- IPEDS (2016), "About IPEDS," available at <http://nces.ed.gov/ipeds/home/aboutipeds>. [2]
- Jolliffe, I. (2002), *Principal Component Analysis*, New York: Springer. [7]
- McCormick, A. C. (2005), "The Carnegie Classification of U.S. Institutions of Higher Education," Kraków, Poland, September 14, 2005. [1]
- (2013), *Classifying Higher Education Institutions: Lessons From the Carnegie Classification (Clasificación de las Instituciones de educación Superior: Lecciones de la Clasificación Carnegie)*, Pensamiento Educativo. *Revista de Investigación: Educacional Latinoamericana*, 50, 65–75. [11]
- McCormick, A. C., and Zhao, C.-M. (2005), "Rethinking and Reframing the Carnegie Classification," *Change: The Magazine of Higher Learning*, 37, 51–57. [1]
- Morse, R., Brooks, E., and Mason, M. (2015), "How U.S. News Calculated the 2016 Best Colleges Rankings," available at <http://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings>. [1]
- Shulman, L. S. (2001), *The Carnegie Classification of Institutions of Higher Education*, Menlo Park: Carnegie Publication. [1]
- The Carnegie Classification of Institutions of Higher Learning (2016), "Basic Classification Methodology," available at <http://carnegieclassifications.iu.edu/methodology/basic.php> [2]
- The Carnegie Foundation for the Advancement of Teaching (1987), *A Classification of Institutions of Higher Education 1987 Edition*, Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching. [2]
- (1994), *A Classification of Institutions of Higher Education: 1994 Edition*, Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching. [2]
- (2000), *A Classification of Institutions of Higher Education: 2000 Edition*, Menlo Park, CA: The Carnegie Foundation for the Advancement of Teaching. [2]
- The University of Nevada Las Vegas (2014), "The Path to Tier One," Technical Report, The University of Nevada Las Vegas. [1]
- University of Houston (2015), "University of Houston—Tier One FAQs," available at <http://www.uh.edu/about/tier-one/tier-one-faq/index.php> [1]
- Zhao, C.-M. (2011), "As Classificações Carnegie: Uma Breve Introdução," *Revista Ensino Superior Unicamp*, 4, 72–80 [English translation accessed online 23-January-2017]. [1]