# Demystifying the Carnegie Classifications: A Sensitivity Analysis

Paul Harmon

Adviser: Dr. Mark Greenwood

Montana State University

Spring Semester 2017

# Contents

**Abstract**

The Carnegie Classifications of research activity are used to compare like institutions in higher education on a variety of research-related characteristics. In 2015, the newest update of the Carnegie Classifications were released, with Montana State University moving from the top-tier category of "Highest Research Activity" to second highest tier, "Higher Research Activity." The classification system is based on two separate indices calculated using Principal Component Analysis (PCA). The first index is based on a set of aggregate variables and the other on a set of per-capita metrics.

This analysis re-creates the calculation of the classifications and examines how sensitive they are to changes in the underlying characteristics of a given institution, focusing on Montana State University. Care should be taken when interpreting the results of this analysis; a static analysis of a dynamic process can illuminate the most sensitive aspects of the classifications but cannot forecast the way that other institutions will grow with respect to Montana State University. That being said, this analysis informs how difficult it would be to move from the R2 status to R1 in the future, but more importantly, it further illuminates the reasons for why Montana State was placed in the "Higher Research Activity" classification in 2015. Additionally, alternative methods for classification of the universities are discussed.

# 1 Introduction

The Carnegie Classifications of Institutions of Higher Education (CCIHE) are released every 5 years. They are intended to be used for institutions to identify other schools which are similar in size, research production, and research spending so that meaningful comparisons can be made between institutions. They are unfortunately often mistaken system ranking institutions based on research quality; however, the classifications of each institution are not meant to identify schools as being producing better or worse research than other institutions. Nevertheless, the classifications of doctoral-granting institutions (R1, R2, and R3) imply ordinal rankings, and thus administrators at various schools have interpreted them as such.

In 2014, the Center for Postsecondary Research at the Indiana University School of Education took over the formulation of the classifications from the Carnegie Foundation for the Advancement of Teaching. When the 2015 updates were released, Montana State University - among a cohort of several institutions - moved from the "Very High Research Activity" to the "Moderately High Research Activity" category. Institutions are scored on both an aggregate index of research productivity, which considers doctorates awarded along with expenditures and research staff, as well as a per-capita index of research activity.

This analysis seeks to recreate the classifications produced by the researchers at Indiana University. Further, I analyzed the sensitivity of the classifications to minor perturbations in the underlying indices used to calculate each school's score. This allowed me to determine which variables most strongly affect the score for a given institution, specifically focusing on Montana State. I also created an interactive web application that demonstrates where Montana State would end up relative to the other institutions in the dataset if it experienced these slight marginal changes. Administrators at Montana State University (and other institutions like it) have made obtaining R1 status an institutional goal; this sensitivity analysis shows that the path from the current classification to the higher one would be at least somewhat arduous, holding all other institutions at their 2015 levels. Finally, this paper covers several alternative methods for determining groups from the Carnegie Classifications using several clustering methods.

# 2 Methods and Data

## 2.1 The Data

The data used here were obtained from the Montana State University Office of Planning and Analysis but are available more generally from the Carnegie Classifications website at http://carnegieclassifications.iu.edu/. In either case, the data contain information pertaining to many levels of institutions; only those that grant doctoral degrees are of interest. The data are therefore processed in order to remove the non-doctoral granting institutions. The data that are reported come from a variety of sources, including the Integrated Postsecondary Education Data System (IPEDS), the CCIHE, and the National Science Foundation (NSF).

In the final dataset, 335 institutions granted doctoral degrees during the period of interest. However, some of the smaller institutions do not report expenditures for STEM-related fields. Therefore, I removed 59 institutions from the dataset, all of which were classified as "high research activity." Doing so leaves a count of n=276 schools on which the classifications were calculated. Table 1 shows the 8 variables used to calculate the Carnegie Classifications with the largest schools on each metric.

| Variable | Units | Mean | SD | School With Max Value |
|---:|---|---:|---:|---|
| Faculty Size | Count | 909.80 | 690.50 | U. of Michigan |
| HUM PhD | Count | 24.60 | 35.00 | Indiana U. Bloomington |
| OTHER PhD | Count | 58.30 | 58.40 | Nova Southeastern U |
| Social Science PhD | Count | 22.20 | 26.00 | CUNY Grad School |
| STEM PhD | Count | 100.20 | 122.50 | U. of California-Berkeley |
| Research Staff | Count | 272.90 | 588.50 | Harvard University |
| STEM exp | 1000’s of Dollars | 197783.90 | 285698.50 | Johns Hopkins U. |
| Non-STEM exp | 1000’s of Dollars | 11208.90 | 16513.90 | U. of Wisconsin-Madison |

Table 1: The variables used in the Carnegie Classifications. The schools with the largest values are also noted.

## 2.2 Principal Components Analysis

The Carnegie Classifications are built using a methodology known as Principal Components Analysis (PCA). The main goal of PCA is to do dimension reduction, meaning that this methodology is used to take a large set of variables and reduce it down to a more manageable number of variables.

Consider a set of p variables $x_1, x_2, ...x_p$. Methods for dimension reduction such as PCA seek to reduce the number of $p$ variables by creating a new set of variables $y_1, y_2, ...y_q$ that retain most of the variation in the original data. In PCA, this is done via an eigenvalue decomposition of either the correlation matrix or covariance matrix of the x's. These new $y$ variables are linear combinations of the x's; further, they are ordered so that the first new variable, $y_1$, accounts for the most variation in the original $X$ variables and $y_2$ accounts for the second most, orthogonal to $y_1$ (Everitt and Hothorn, 2011). Note that we have not yet achieved dimension reduction; PCA by itself returns the same number of variables, albeit re-parameterized, as what we started with. However, since the new variables contain information about much of the variation in the old variables, it is possible to use only a subset of the first few $y$ variables without losing much information about the variation in the $x$'s. In this way, often the first two or three $y$'s are used to describe the entire set of p $x$'s.

## 2.3 Calculating Scores and Loadings

Principal components analysis returns two key pieces of information: the **scores** and the **loadings**. The scores are the aforementioned $y$ values. The first principal component, $y_1$ can be represented as follows for the $x_1, ..., x_q$ original variables where $a_1 ... a_q$ are the (unstandardized) loadings from the PCA:

$$y_1 = a_{11}x_1 + a_{12}x_{12} + ... + a_q x_q.$$

The scores can be thought of as an "index" for a more complicated underlying process; each score can be interpreted as a weighted average of the original $x$ variables. In the Carnegie Classifications, the researchers explicitly define the scores used in the analysis as a "per-capita index" and an "aggregate index."

The loadings are the coefficients by which each of the original variables must be multiplied to obtain the PC scores. If PCA is done on the correlation matrix, then take the Z-scores and multiply them by the eigenvector coefficients. The scores can be interpreted as weighted averages of the original variables; the loadings are the weights on which those averages are calculated. The "standardized loadings" of the PCA are the eigenvalue coefficients multiplied by the square root of the eigenvalues.

PCA can be done on either the correlation matrix or the covariance matrix (Everitt and Hothorn, 2011). In the Carnegie Classifications, the correlation matrices were used. In R, calculation of the Principal Components Analysis can be done using the function `prcomp` or princomp (R Core Team, 2017). The two functions are similar; however, the former uses the singular value decomposition whereas the latter uses an eigenvalue decomposition.

## 2.4 Dimension Reduction: How Many Scores Should Be Used

The question of how many scores should be used has no definitive answer; it is often left up to the researcher to define the optimal degree of dimension reduction. Often this is achieved by examining the percentage of variation explained or screeplots of eigenvalues. Certainly, the number of scores used should be less than the original number of variables; otherwise, the researcher might as well work with the original data.

In the Carnegie Classifications, the researchers used only the first score for each index created. This allowed for researchers to directly plot two separately-calculated indices against each other; however, using only a single score for the index means that they explained only part of the variance in the original variables. Indeed, in the 2015 update, the aggregate index score only explained 70 percent of the variation in the original data and the per-capita index explained only 71 percent (Borden 2017).

## 2.5 Regression and Classification

PCA is a powerful tool for dimension reduction; it can be used both in a classification setting and a regression setting. For unsupervised classification, using the singular value decomposition to create a subset of k principal components scores allows for identification of groups on a reduced set of variables. In regression, the linear predictor can be rewritten as $y = \mathbf{X}\beta + \epsilon = \mathbf{F}\theta + \epsilon$ where **F** is called the Factor Matrix; it contains only the set of $k$ scores (West 2003). Singular Value Regression, or Principal Components Regression, is useful for some regression problems because it allows for regression on the re-parameterized indices, or scores, that come from the PCA. This analysis focuses mainly on the unsupervised classification using PCA scores; however, it is worth noting that for a given institution, one could use the Aggregate and Per-Capita indices of performance as derived inputs in a related regression problem. Nevertheless, as pointed out by Everitt and Hothorn, PCA is "overwhelmingly an exploratory technique" (Everitt and Hothorn 2011) and in the case of the Carnegie Classifications, it is used primarily to graphically describe the differences between institutions rather than to predict them.

# 3 Recreating The Carnegie Classifications

## 3.1 The 2015 Update

The classifications are calculated from two indices, one based on the aggregate PhDs awarded, expenditures, and research staff for each institution, and the other on a per-capita basis that uses the raw headcount of tenurable and non-tenurable faculty at each school to scale the expenditures and research staff variables.

In general, institutions with relatively large values in one index are likely to be relatively large in the other index. In 2015, the Pearson correlation between the two indices was 0.84. This suggests that the relationship between the indices was a strong, positive linear association.

The researchers generated the classifications based on calculating the two indices are calculated in the following way:

1. Each variable is ranked from smallest to largest. Ties are assigned the minimum rank.
2. Two PCAs are estimated using the correlation matrices of the ranked data.
3. An Aggregate and Per-Capita Index are calculated from the first Principal Component of each PCA.
4. The indices are rescaled. Then, the rescaled Aggregate Index is plotted against the rescaled Per-Capita Index to create Figure 1.
5. Arcs are drawn with arbitrary radii to define breakpoints between categories.

## 3.2 Why Rank the Data?

Instead of using rank, the indices could be based on the raw counts of PhDs, research staff, and expenditures. However, large schools tend to produce students and spend money at much higher rates than smaller schools. Counts of PhDs awarded, expenditures, and research staff/faculty sizes are severely positively skewed. For instance, Johns Hopkins University spent over 2 billion dollars on STEM expenditures in

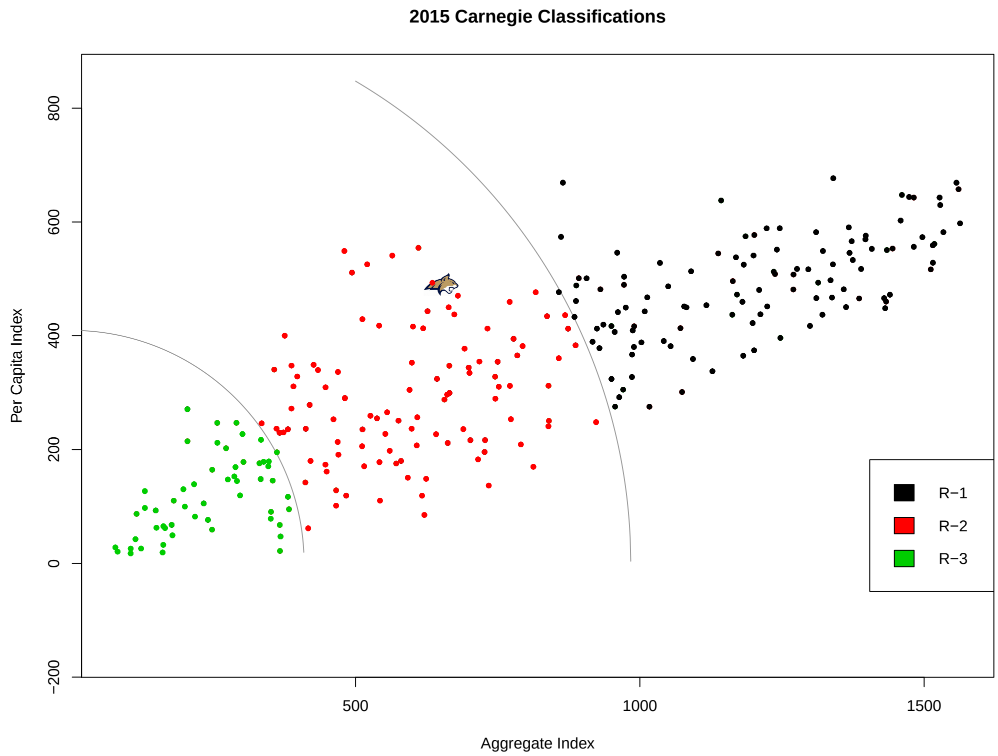


Figure 1: The 2015 Carnegie Classifications. Montana State was classified in the R2: Higher Research Activity group.

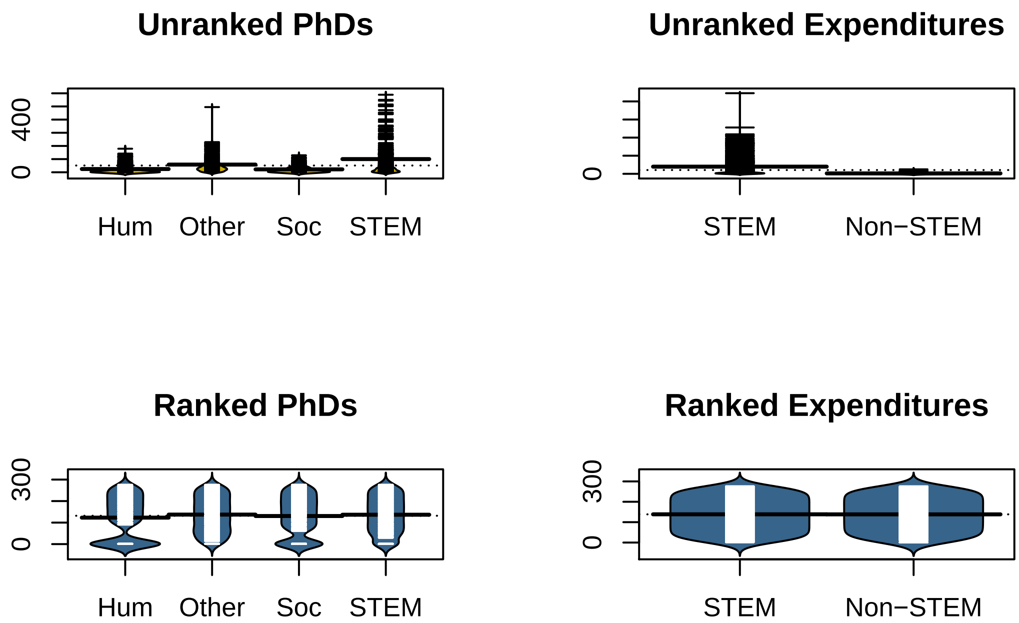


Figure 2: The raw data are positively skewed, but the ranking the data decreases the disparities between larger and smaller schools.

2015 compared to 104 million dollars for Montana State University, which was above the 50th percentile in the schools compared. Ranking the data prior to classification increases separation between institutions at the low end and decreases separation between institutions with large values. Figure 2 shows beanplots (Kampstra, 2008) that display the distributions of both the ranked and raw data. The narrow lines refer to each institution's value in the dataset; the wide lines refer to the mean values for each variable. For some of the PhD counts, the distributions of ranked counts appear bimodal because there were many schools that did not offer PhDs in that category in 2015. Those schools are all tied at 0 PhDs for those types of degrees.

## 3.3 Aggregate Index

The Carnegie Classifications are based on a plot of two indices, the aggregate index and the per-capita index. The x-variable in the display used to create the Carnegie Classifications is the aggregate index. It includes PhD degrees awarded in one of four categories: humanities, professional fields, social sciences, and STEM fields as well as the research staff, STEM expenditures, and non-STEM expenditures that are also used in the per-capita calculation.

The formula for the aggregate index for the $i^{th}$ institution is:

$$\begin{aligned} AggregateIndex_i = &.37HumD_i^* + .27STEMD_i^* + .39SocSciD_i^* \\ &+ .27OtherD_i^* + .40STEMExp_i^* + .38NonSTEM_i^* + .33ResStaff_i^* \end{aligned}$$

where $HumD_i^*$ is the standardized rank of number of humanities PhDs at the $i^{th}$ institution, *STEMD, SocSciD, OtherD* refer to STEM, Social Sciences, and Other PhD degree standardized ranks. $STEMexp_i^*$ and $NonSTEM_i^*$ refer to the standardized ranks of research expenditures related to STEM and non-STEM fields, respectively. $ResStaff_i^*$ refers to the standardized, ranked counts of research staff.

## 3.4 Per Capita Index

The per-capita index is the y-variable in the display used to create the Carnegie Classifications. It considers only three variables: non-faculty research staff, STEM expenditures, and non-STEM research expenditures all divided by the size of the faculty at the given institution. The weights on each variable are calculated from the loadings generated by the PCA of the correlation matrix of the three per-capita variables.

The formula for the per-capita index is given below for the $i^{th}$ institution:

$$PerCapitaIndex_i = \frac{.64ResStaff_i^* + .64STEMExp_i^* + .42NonSTEMExp_i^*}{FacultySize_i^*}$$

where $FacultySize_i^*$ is the standardized rank of number of Tenured/Tenure-able faculty at the $i^{th}$ institution. $STEMexp_i^*$ and $NonSTEM_i^*$ refer to the standardized ranks of research expenditures related to STEM and non-STEM fields, respectively. $ResStaff_i^*$ refers to the standardized ranked counts of research staff.

## 3.5 Combining the Indices

After calculating the individual per-capita and aggregate indices, the two are combined with a single plot. The per-capita index is plotted along the y-axis and the aggregate index is plotted along the x-axis, as previously noted. The Carnegie Classifications are then grouped into three groups regardless of how many distinct groups appear; in some years the three groups of institutions are fairly well-separated into three clusters and in other years they are not.

With higher values being "good" on each index, groups of universities were approximately formed along the major axis of the correlation line between the indexes, using arcs of circles with a common original to split the groups. The specific boundaries were subjectively chosen (roughly breaking the group into thirds) but did follow a specific curve. In previous years, the lines that divided groups were hand drawn. While three distinct clusters rarely form, in 2015 the data were decidedly not well-separated. The positions of the delineations between the groups were the most subjective aspect of the analysis.

## 3.6 Method For Ties

After choosing and obtaining the same variables they used in the Carnegie Classifications, a handful of decisions must be made. The most important choice comes in the ranking step. When ranking institutions that are tied, there are several methods for calculating ranks that have important consequences for an analysis, especially where an institution can move up or down in rank in subsequent years. In the statistical software package R (R Core Team 2017), the default setting of the rank function is to take the average of the ranks for any tied responses (Becker et al., 1988). For instance, if the first five institutions have the same value for STEM Expenditures, each institution would receive a rank value of 3. Alternative methods include taking the first, last, minimum, or maximum of the ranks. The first and last methods result in taking a permutation of each of the potential ranks with first involving increasing values and last involving decreasing values. The minimum rank method refers to the more commonly known ranking method used in sports; in the previous example, each of the five institutions would be ranked 1st. Similarly, the maximum rank gives the largest rank to all of the observations; in the previous example, each institution would be ranked fifth.

The literature on the Carnegie Classifications does not specify exactly the method that the institute used for tied ranks, but comparing each method to the final results indicates a clear picture. Table 2 is a table of standardized loadings for each method along with the loadings generated from the actual classifications. The standardized loadings did not exactly match for any method (largely due to rounding, software differences, and minor differences in the data used), but the minimum method had the closest results. The loadings for the average method did not match closely at all.

The loadings give some evidence that the minimum method was used to deal with ties in the rankings; however, the plots of the indices generated by each of the methods confirm that the minimum ranking was used. Other methods generate plots where institutions are classified that move across their provided boundaries. For the minimum ranking, most results respect the provided boundaries. For institutions that endeavor to improve their classification - or for institutions that seek to avoid dropping into the lower category - this is of particular importance. Breaking ties with one addi-

| **Aggregate Rankings** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ties Method | HUM | OTHER | SOSC | STEM | STAFF | STEM exp | NS exp |
| Average | 0.83 | 0.618 | 0.881 | 0.916 | 0.907 | 0.899 | 0.792 |
| First | 0.822 | 0.615 | 0.882 | 0.918 | 0.908 | 0.899 | 0.788 |
| Last | 0.824 | 0.62 | 0.876 | 0.913 | 0.907 | 0.899 | 0.795 |
| Min | 0.818 | 0.616 | 0.873 | 0.914 | 0.902 | 0.899 | 0.792 |
| Max | 0.837 | 0.619 | 0.886 | 0.917 | 0.912 | 0.899 | 0.792 |
| Actual | 0.82 | 0.617 | 0.874 | 0.915 | 0.902 | 0.9 | 0.79 |
| **Per Capita Rankings** | | | | | | | |
| Average | - | - | - | - | 0.93 | 0.932 | 0.615 |
| First | - | - | - | - | 0.93 | 0.933 | 0.61 |
| Last | - | - | - | - | 0.929 | 0.932 | 0.615 |
| Min | - | - | - | - | 0.928 | 0.93 | 0.616 |
| Max | - | - | - | - | 0.93 | 0.934 | 0.615 |
| Actual | - | - | - | - | 0.928 | 0.931 | 0.614 |

Table 2: Standardized Loadings: The standardized loadings generated with the minimum method for ties had the most exact matches and the closest overall matches to the actual Carnegie Classifications.

tional PhD in a category can lead to much larger changes in the ranked data if many institutions are given the minimum rank, compared to other methods of handling ties.

For Montana State, this implies that Social Science PhDs are important. Montana State University is tied with 61 institutions that produced 0 social science doctorates. Under the average ranking method, each of those institutions would be ranked as the average rank of 31; however, using the minimum rank, an institution that went from zero to one social science PhD would improve its position by 61 points. To be clear, the use of the minimum rank method for dealing with ties indicates that Montana State could gain substantial ground simply by going from 0 to 1 Social Science PhDs in the next iteration of the classifications. Moreover, increasing the number of STEM PhDs is unlikely to have the same impact as increasing Social Science PhDs because Montana State is tied with only 2 institutions for count of STEM PhDs.

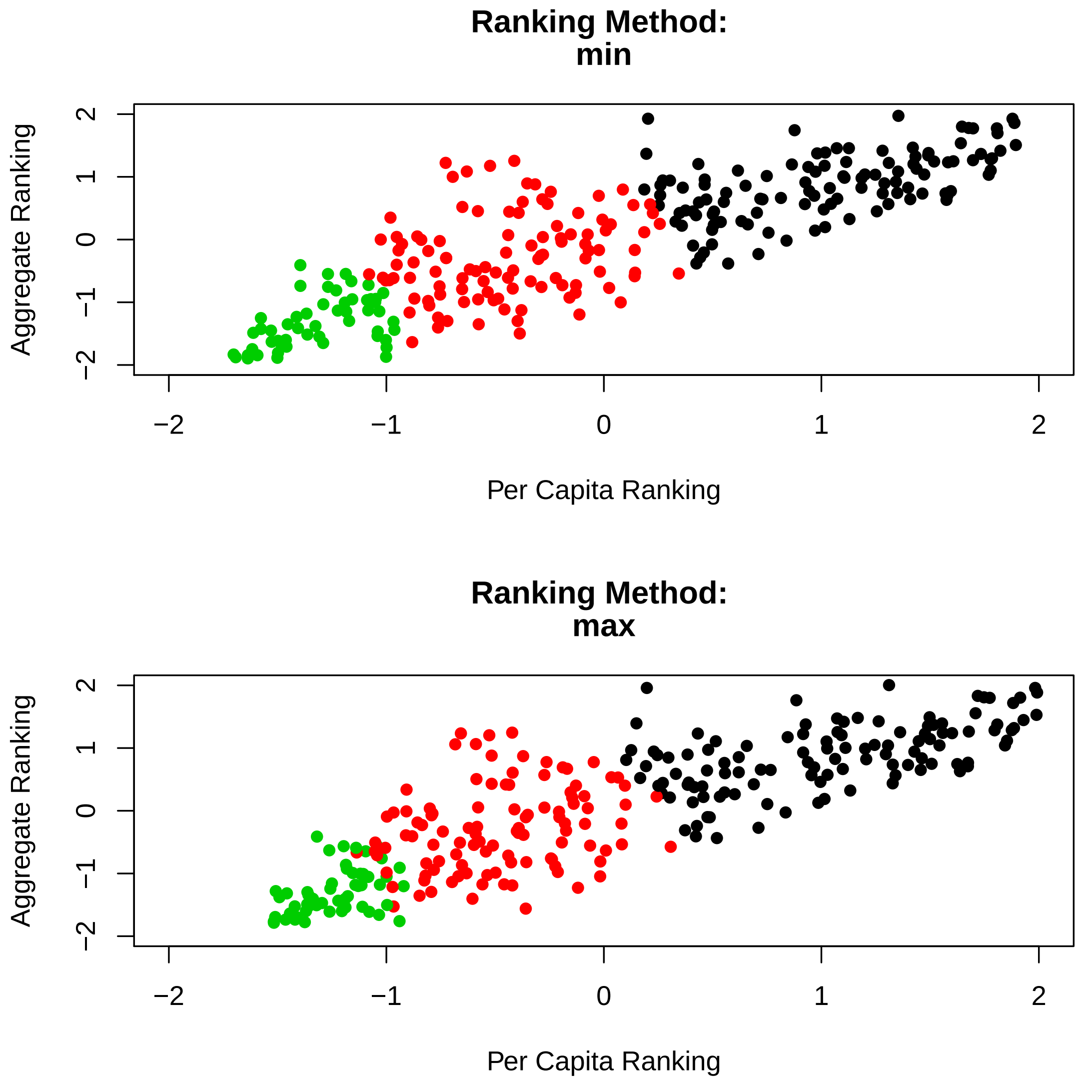


Figure 3: Classifications with different methods for ties. The overlapped groups in the lower panel and well-separated groups in the upper panel indicate that the researchers used the minimum method.

# 4 Exploring the 2015 Classifications

## 4.1 Similar Institutions

Recall that the purpose for which the classifications are designed is not to rank school quality but instead allow institutional researchers to make meaningful comparisons among schools with similar research qualities. Examining the classifications allows for administrators and other decision makers at Montana State to identify a cohort of similar institutions. By examining the plot of nearest neighbors to Montana State, it is possible to determine a list of like institutions. Figure 4 shows the plot of the most similar institutions to Montana State and Table 3 lists their names with their Euclidean distances to MSU.

It is evident that the University of Alaska-Fairbanks is the institution that most closely resembles Montana State. Interestingly enough, when comparing the raw numbers, they do not look all that similar; Montana State is clearly more productive in producing doctorates in every category save for Social Sciences. Further, while Alaska-Fairbanks spends more money on STEM expenditures, Montana State spends more on non-STEM and has a larger research staff cohort. The key similarities lie not in the raw data but the rankings across the variables. Table 4 gives the actual values of each variable for the two schools.

Institutional researchers and administrators ought to use the cohort of nearest neighbor schools in Table 3 to make meaningful comparisons. From a research perspective, for instance, Carnegie Mellon University, Oregon State University, and Rice University are more similar to Montana State in terms of research characteristics than they are to the large-scale R1 schools such as Stanford, Johns Hopkins or the University of Washington.

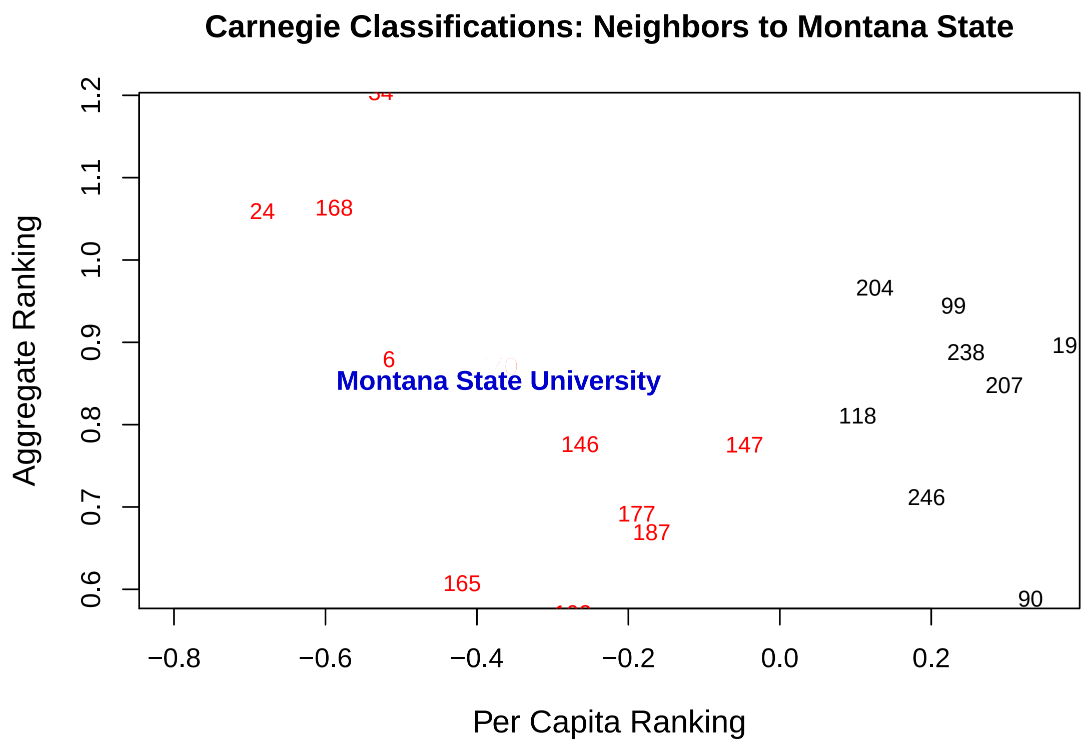


Figure 4: Neighbor institutions to Montana State University.

| ID Number | Name | Distance from MSU |
|---|---|---|
| 6 | University of Alaska Fairbanks | 0.02 |
| 146 | Dartmouth College | 0.12 |
| 177 | Yeshiva University | 0.24 |
| 165 | Rensselaer Polytechnic Institute | 0.28 |
| 34 | Colorado School of Mines | 0.36 |
| 187 | North Dakota State University-Main Campus | 0.36 |
| 24 | Naval Postgraduate School | 0.39 |
| 168 | Rockefeller University | 0.42 |
| 147 | University of New Hampshire-Main Campus | 0.44 |
| 118 | Tufts University | 0.56 |
| 204 | Oregon State University | 0.58 |
| 246 | The University of Texas at Dallas | 0.66 |
| 99 | Tulane University of Louisiana | 0.67 |
| 238 | Rice University | 0.67 |
| 207 | Carnegie Mellon University | 0.74 |

Table 3: Comparison of Montana State to nearest neighbors. Montana State's cohort of similar schools includes an impressive cohort of institutions, both R1 and R2.

| NAME | Faculty | HUM | OTHER | SOC | STEM | R.Staff | STEM exp | NS exp |
|---|---|---|---|---|---|---|---|---|
| Montana State | 456 | 2 | 9 | 0 | 45 | 75 | 104646 | 8702 |
| MSU Rank | 78 | 99 | 37 | 1 | 125 | 156 | 156 | 180 |
| Alaska-Fairbanks | 374 | 1 | 5 | 4 | 34 | 50 | 152352 | 3417 |
| UAF Rank | 47 | 91 | 23 | 85 | 112 | 131 | 178 | 113 |

Table 4: Comparison of Montana State to University of Alaska Fairbanks. Although not similar across all values on the raw data, both universities have similar scores on the Carnegie indices.

## 4.2 Montana State compared to R1 Institutions

In 2010, Montana State was classified as an R1 institution. The reason for this is not entirely clear; however, Montana State had a high per-capita index value and continued to have this in 2015. The per-capita score for Montana State was above the 75th percentile of all universities used in the classifications of doctoral-granting institutions.

Of all doctoral-granting institutions, Montana State's research expenditures on STEM-related fields were above the median value, as were its non-STEM expenditures. That, combined with the above-median research staff size, may explain why the Montana State scored highly on the per-capita scale. However, output in terms of PhDs awarded is less than the 50th percentile for all degree types awarded, even STEM doctorates. A naive examination of any single variable makes it appear that Montana State would be reasonably well-classified in the R1 group. As seen in Table 5, outside of Social Science PhDs, MSU produces PhDs, spends research money, and hires research staff at a rate that would be relatively similar to many of the smaller R1 schools.

| | HUM | OTHER | SOSC | STEM | Staff | STEM | NonSTEM |
|---|---|---|---|---|---|---|---|
| Montana State | 2.00 | 9.00 | 0.00 | 45.00 | 75.00 | 104646.00 | 8702.00 |
| Mean R1 | 51.00 | 87.00 | 44.00 | 202.00 | 604.00 | 411742.00 | 21672.00 |
| Median R1 | 45.00 | 76.00 | 37.00 | 152.00 | 387.00 | 319818.00 | 14914.00 |
| Min R1 | 0.00 | 0.00 | 1.00 | 27.00 | 32.00 | 5719.00 | 725.00 |

Table 5: Comparison of Montana State to R1 schools. MSU looks like it could fit in on any single metric, but is near the bottom when considering all variables.

Many R2 schools have instituted policy goals geared towards moving from the R2 category to achieving the R1 status. Indeed, both the University of Idaho (McClintick 2016) and the University of Montana (University of Montana 2017) are making R1 a policy goal. But is this a reasonable goal for Montana State? Does Montana State compare favorably with R1 schools on the metrics used to calculate the Carnegie Classifications?

However, comparing MSU on individual metrics to R1 schools is misleading since the Carnegie Classifications are based on weighted averages of these variables, not the individual variables themselves. A more fair comparison would be to look at Montana State across all metrics vs. R1 schools on all metrics. This can be achieved with a Parallel Coordinate Plot, as is given in Figure 5. The nearest R1 schools, Oregon State and Tufts Universities, are both highlighted as well. The difference between the majority of R1 schools and Montana State is striking; MSU produces fewer PhDs across the board and spends less money on research than nearly all of the schools in the top tier category. Even if Montana State were classified in the R1 group, it would be near the minimum in all categories. In 2015, Montana State lagged behind most of the

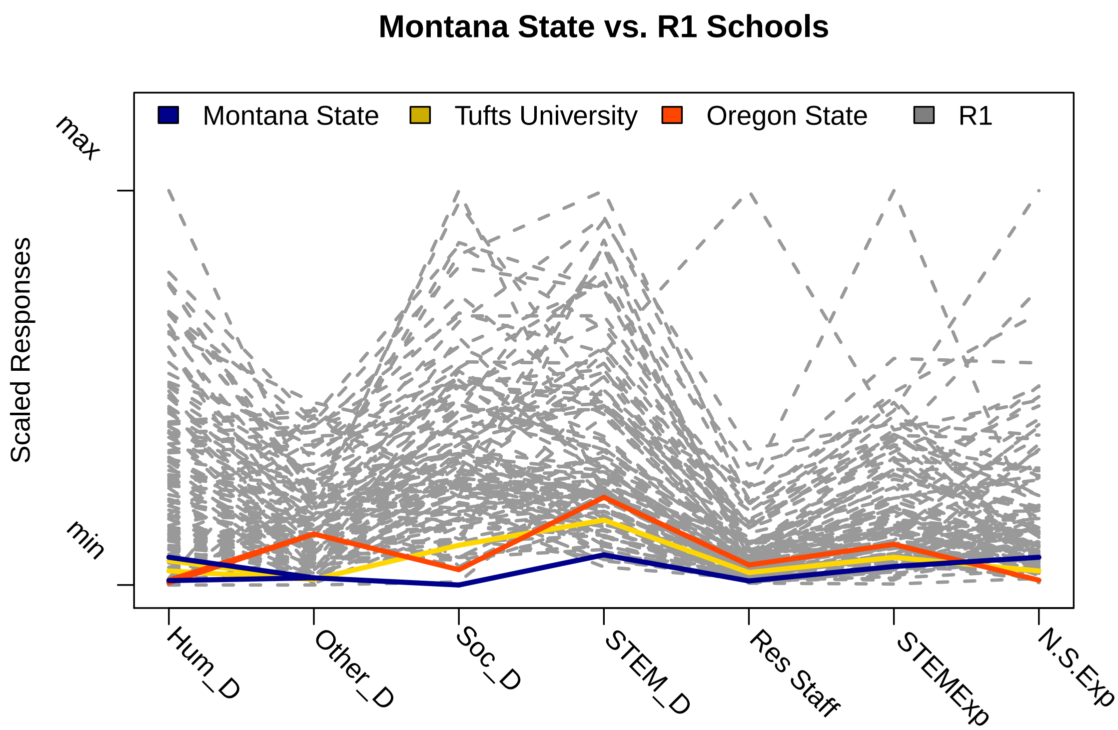


Figure 5: The above plot shows Montana State vs. the R1 Institutions on all of the variables used in the classifications scaled from minimum to maximum on each variable. Montana State would be near the bottom on all variables. If R2 and R3 schools were added, Montana State would be near the upper portion of schools.

R1 institutions in expenditures, staff sizes, and doctorates awarded. While Montana State had more Other, Humanities, and STEM doctorates awarded than the smallest R1 institutions, it lagged behind the mean and median R1 values by a large margin. Moreover, expenditures and research staff sizes were much smaller than the average R1 schools, even if they were slightly larger than the smallest R1 school.

# 5 Sensitivity of the Rankings

## 5.1 Single Variable Movement: Aggregate Index

One important question to examine is not only whether it is possible to attain the R1 status, but what might be the best way to get there? Certainly, there are many different ways to move up in classifications; in general, increases on both indices could be achieved by increasing any one of the variables that underlie the classifications. Could Montana State (or any institution, for that matter) focus solely on a single metric to move up to R1 status?

Recall that the aggregate scale is just a weighted average of four counts of Doctoral degrees awarded, STEM and Non-STEM Expenditures, and a headcount of research staff. The PhD counts only impact the aggregate index whereas the latter three variables also factor into the per-capita index. If Montana State were to focus solely on adding a single type of doctoral degree, could the institution move up from R2 to R1?

The plots in Figure 6 indicate the distance to the boundary associated with marginal increases in each variable. Starting with additional doctoral degrees in the Social Sciences, it is clear that a small gain from 0 to even a single PhD is associated with a jump towards the R1 boundary. Since Montana State University had no social science PhDs awarded in 2015, they were one of 61 tied institutions that did not grant any degrees. Moving up to a single PhD allows breaks that tie and gives Montana State a jump in 60 rank-units. Adding additional PhDs in the social sciences moves Montana State into and out of more ties, moving the institution closer to the boundary. After 58 Social Science Doctorates awarded, Montana State would hit the boundary and move across the classification border. At 130, the institution would cease to move as it would have attained the highest rank in this category; however, at that point it would already have reached the R1 status. While it is possible to move across the border based solely on

Social Science PhDs, it is neither efficient nor feasible to add 58 Social Science PhDs when none were awarded prior to 2015.

For the other doctorate types, it is not possible to move up to R1 status simply by increasing numbers of a single degree type. For STEM PhDs, Montana State already finds itself nearly halfway through the rankings - in 2015 it ranked 125th highest of the 276 ranks. Even if Montana State were to add 545 doctorates (in order to pass University of California-Berkley to attain top rank), it would still be roughly 60 units away from R1 status, holding all else equal. Admittedly, considering adding 545 additional PhDs is absurd. Humanities and Other doctorates experience the same problem; even though marginal gains in either degree type can yield differing increases in each rank, increasing either degree without changing anything else does not yield enough of a gain to move the institution across the R1 border.

This makes sense. The aggregate index is just a weighted average of seven different metrics. If one is changed, the effect it has may make a difference on the index itself; however, that effect is likely to be damped by the other six un-changed variables. The reason that Social Science PhDs can take Montana State all the way across the border is because the institution was at the lowest end-point already; on the other metrics, Montana State simply cannot grow enough to get across the threshold from R2 to R1.

## 5.2 Single Variable Movement: Both Indices

If increasing awarded doctoral degrees is not enough to move the institution to R1, could the variables that count for both indices work? Indeed, increasing PhDs awarded only moves a given institution to the right or left since those variables only influence the aggregate index; however, the STEM Expenditures, Non-STEM Expenditures, and Research Staff variables affect both the per-capita and aggregate indices. Thus, increases in either of those three variables could lead to a given institution moving both up and to the right. It would seem reasonable, then, that changes in any of

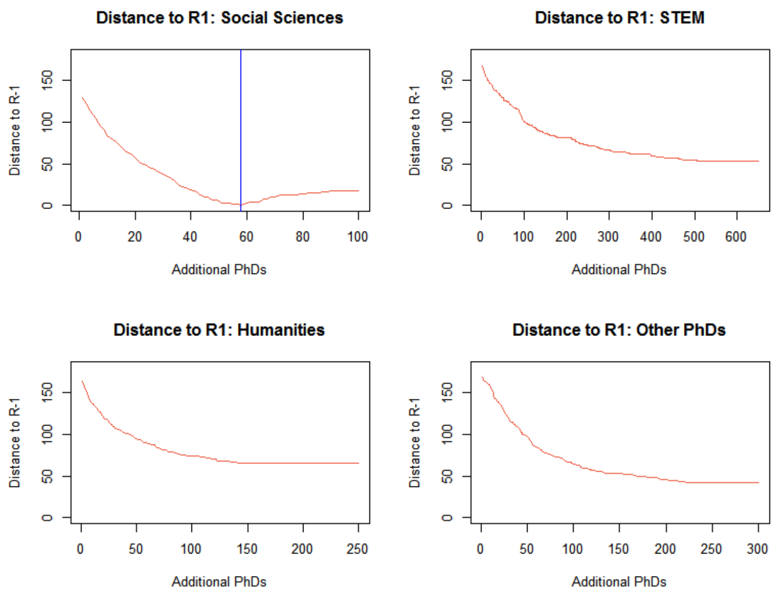


Figure 6: Single Variable Movement: Increasing counts of a single type of PhD lead to movement to the right on the plot. However, only a 58 PhD increase in Social Sciences would get Montana State across the R1 border.

the variables used in both indices could be enough to move Montana State across the R1 border.

Reasonable though the theory may sound, the plots in Figure 7 indicate that neither STEM expenditures, non-STEM expenditures, nor additional research staff can, in and of themselves, move Montana State into the R1 status. Montana State had 75 research staff in 2015; to move up to the top rank the institution would need to add 7223 additional researchers (to break a tie with Harvard, with 7297). Disregarding the absurdity of adding that many researchers, such an increase would still leave Montana State more than 80 units away from R1, a distance that could be obtained by adding fewer than 20 Social Science doctorates. Moreover, increasing STEM and non-STEM expenditures even by billions of dollars would only lead to modest decreases in the distance from the R1 boundary.

It is worth noting that some single metric changes are more reasonable than oth-

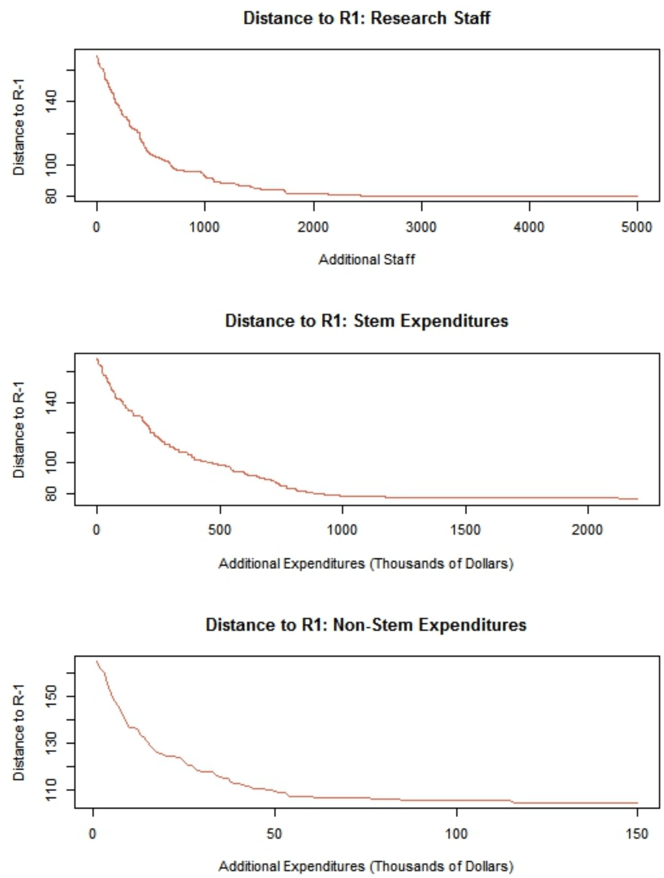


Figure 7

Single Variable Movement: Changes in expenditures and research staff lead to movement in all directions since they are used in both indices. However, increases in any single variable do not lead to large enough movement to get across the R1 threshold.

ers. For instance, while it may be reasonable to consider adding an additional STEM PhD without resulting in changes in the other variables. However, large increases in STEM expenditures would likely result in more research staff being hired and possibly more PhDs being produced. The above plots only consider movement on single variables without regard to the consequences of those movements. Ultimately, the narrative that this analysis informs is that in order to move up in the classifications from R2 to R1, decision makers must focus on a multi-dimensional approach. While increases in a single variable will help move the university up towards the boundary, changes in multiple variables simultaneously will prove much more effective.

## 5.3 Shiny App: Simultaneous Movements

Movement on multiple dimensions is hard to analyze for several reasons. First, there are eight variables to consider, all of which could increase or decrease in the next update. There are countless combinations of changes in each variable that could hypothetically occur. I created an interactive web application using the R package Shiny (Winston 2016) that allows for interactive modeling of changes in the classifications. The application is intended to be used by institutional researchers, administrators, and other stakeholders at Montana State for simulating the classifications under any of those circumstances.

The application can be found at https://paulharmon.shinyapps.io/Carnegie2/. The end user can adjust Montana State's counts of PhDs, expenditures, or research staff by moving the slide bars. The application then re-ranks the institutions, re-calculates the two PCAs, and plots the new indices. Small perturbations may change where Montana State is located in the plot, but they do not necessarily change the structure of the classifications; however, large changes in the values for Montana State can actually slightly change the locations of other schools in the ranked PCA, even though their values are held constant for all variables. This highlights an important feature of the Carnegie Classifications: changes in one school can actually impact the position of other schools.

The following plots illustrate a handful of changes that could be made. Figure 8 illustrates a STEM-heavy change:

- Increase non-STEM PhD counts by 1,
- Increase STEM PhDs by 15,

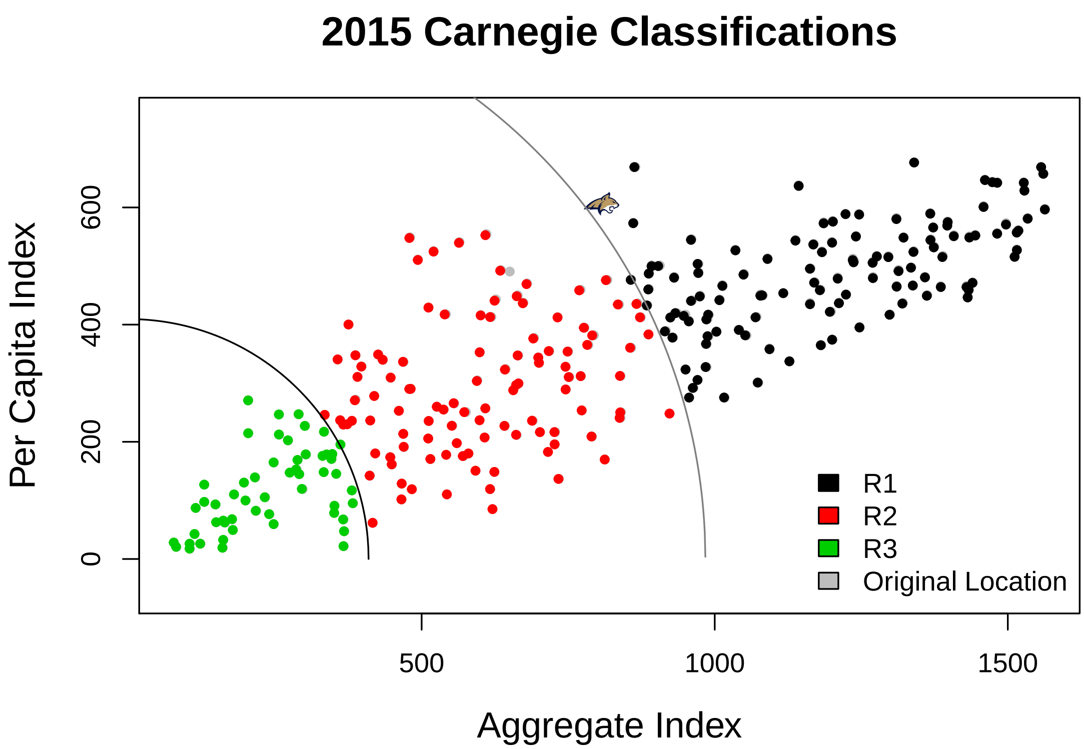


Figure 8: Large increases to STEM PhDs and expenditures could get Montana State across the R1 threshold, given small gains in non-STEM fields.

- Double STEM expenditures (increase by $104,646),
- Increase non-STEM expenditures by 5 million dollars, and
- Increase research staff by 75.

Another possible change that could be considered would be to focus solely on non-STEM degrees, as seen in Figure 9. Investing in humanities, social sciences, and other non-STEM fields may not be efficient for Montana State given the relative lack of non-STEM infrastructure, but could be considered. Given the dynamics of Montana State, I include small gains to the STEM fields as well. The below scenario involves the following changes:

- 10 additional PhDs in all non-STEM categories,
- 5 additional STEM PhDs,
- Increase research staff by 25,
- Increase STEM expenditures by 3 million dollars, and
- Increase non-STEM expenditures by 10 million dollars

Note that this change indicates another key element of the Carnegie Classifications. Well-rounded institutions generally have larger values on both indices than do institutions that specialize only in STEM or non-STEM fields. Compared to the STEM-heavy path, this path seems less arduous; however, there are only a few non-STEM doctoral programs offered at Montana State.

Reducing tenured/tenurable faculty could also help move Montana State towards the boundary; however, it would not be a particularly good long-term solution. Figure 10 shows where Montana State would be with a reduced tenurable teaching faculty but additional researchers. Reducing faculty may be a way to move towards R1, but it is likely not the best way to do it. The following scenario illustrated in Figure 10 considers a 25% reduction in the size of the tenurable/tenured faculty at Montana State along with modest increases in expenditures, PhDs, and a re-allocation of research faculty. The changes illustrated in Figure 10 are:

- Increase non-STEM PhDs by 1,
- Increase STEM PhDs by 5,
- Increase nontenurable research staff by 100,
- Increase STEM expenditures by 10 million dollars,
- Increase non-STEM expenditures by 5 million dollars, and
- **Reduce** tenured/tenurable faculty size by 25%

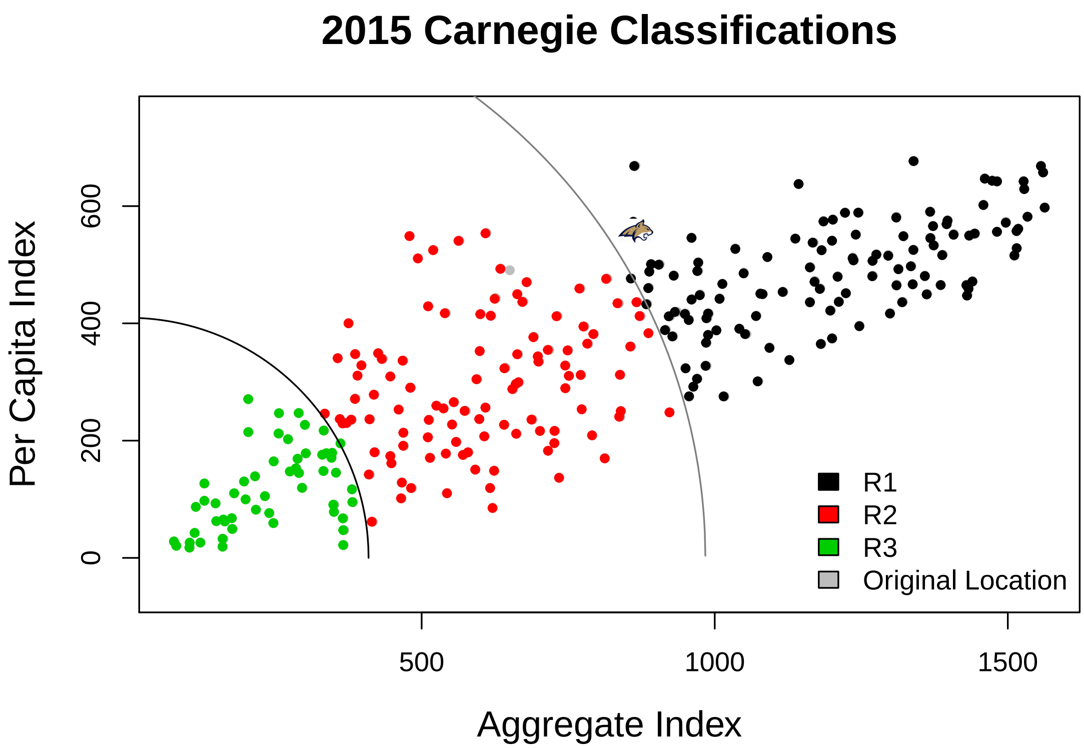


Figure 9: Increases in non-STEM fields could get Montana State across the border, but the university may not have the infrastructure to support large non-STEM growth.

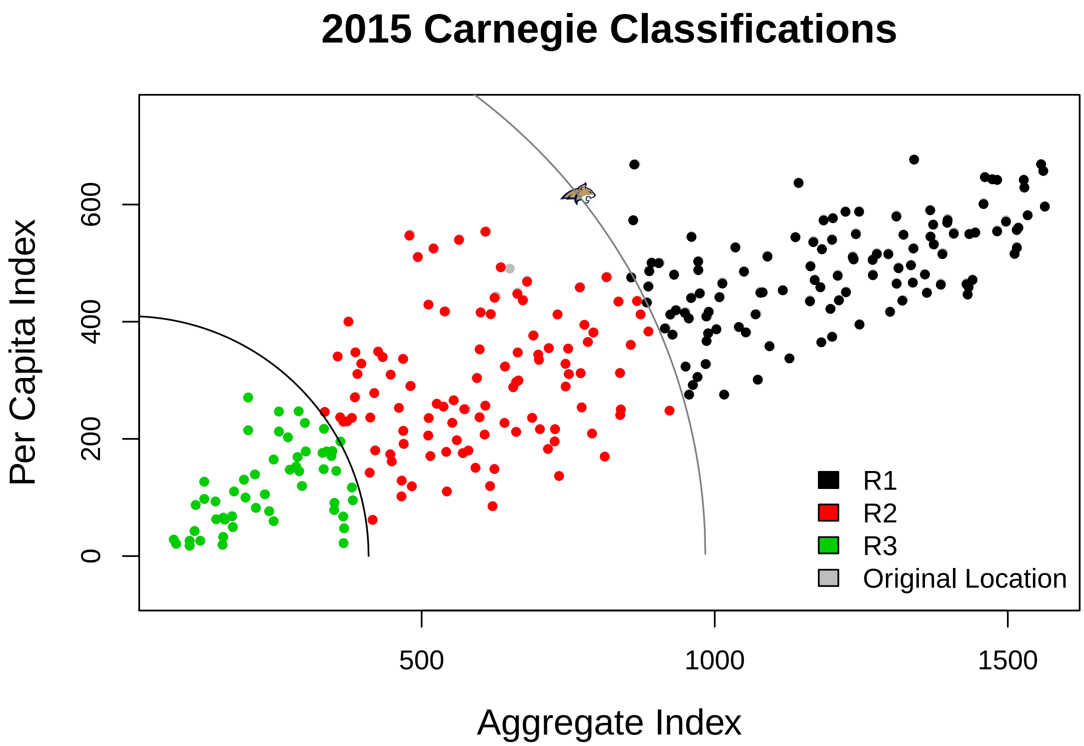


Figure 10: Reduction in faculty size by 25 percent, along with modest gains in other variables, could get Montana State across the border. However, long-term efficacy of cutting faculty is questionable at best.

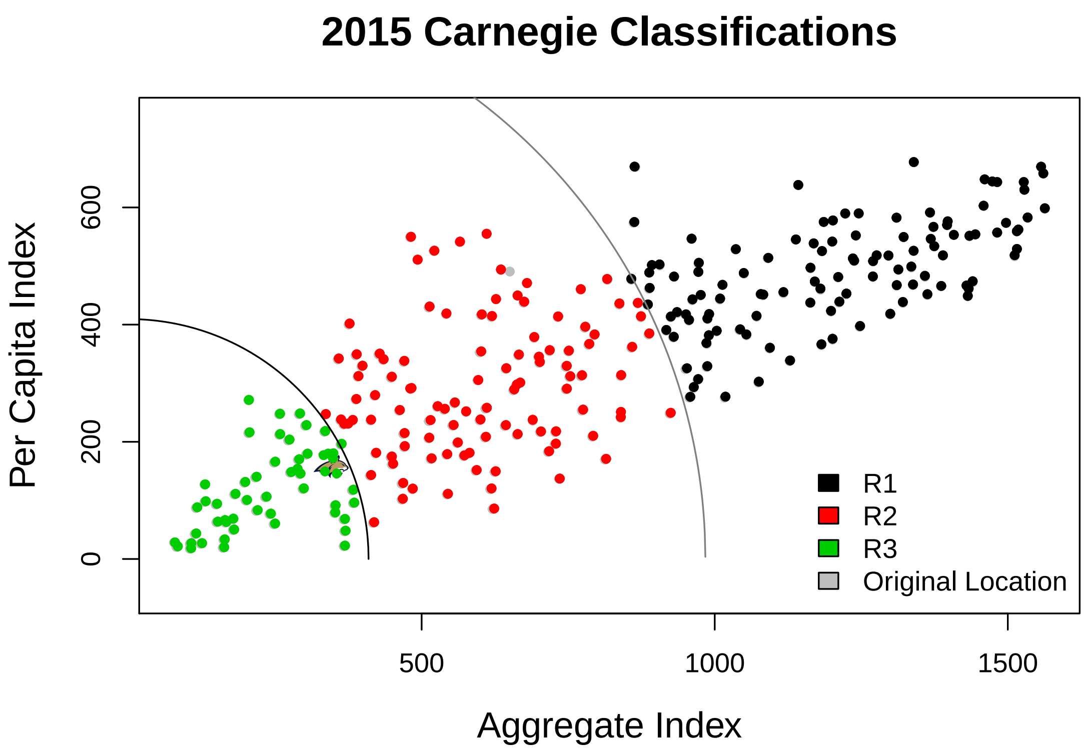


Figure 11: Removing STEM and non-STEM expenditures, along with small reductions in other variables would result in an R3 classification.

While moving to R1 is the policy goal envisioned by administrators, the specter of moving towards R3 is always a possibility. The following scenario illustrated in Figure 11 includes small reductions in the number of PhDs produced and research staff combined with wholesale cuts on research expenditures:

- Reduce all PhDs by 1,
- Reduce non-tenured research staff by 25,
- Reduce STEM expenditures entirely by 104 million dollars, and
- Reduce non-STEM expenditures by 8 million dollars.

The Carnegie Classifications App can be used to model these as well as many other hypothetical scenarios. As a tool for making decisions, it could be used as a starting point for doing economic Cost-Benefit Analysis. Decision-makers could plot out a potential change and then assess the accounting and economic costs of that change in order to determine the lowest-cost methods for potential movements towards R1. Care should be taken when considering these changes; although it may be more mathematically efficient to change one variable, it may not be economically efficient. For instance, adding 10 social science PhDs would likely cost the university more than adding 10 STEM PhDs simply because the academic infrastructure and programs may not exist. Finally, while moving towards R1 classification may not be a bad idea, it is worth remembering that moving towards R3 is just as feasible given the right negative circumstances.

# 6 Analyzing the Classifications

## 6.1 Model-Based Clustering

One of the most subjective aspects of the Carnegie Classifications is where the delineations between groups are drawn. Rather than hand-drawing lines and basing the three groups off of those arbitrary delineations, it might be worthwhile to let the data speak for themselves. This can be achieved via model-based clustering on the two indices. I tried several different clustering options:

- Model-Based Clustering with data-driven selection of the number of clusters
- Model-Based Clustering with 3 clusters
- K-Means Clustering with 3 clusters

### 6.1.1 MBC with Data-Driven Clusters

The model-based method of clustering (Fraley and Raftery, 2002) does not require prior specification of the number of clusters. Rather, it requires a three step process of initialization, maximum-likelihood estimation using the Expectation Maximization (EM) algorithm, and selects the model and number of clusters using BIC comparisons (Fraley and Raftery, 2003). Model-Based clustering can be implemented in R with the `mclust` package (Fraley, Raftery, Murphy, and Scrucca, 2012). I fit the model and made BIC comparisons for 1 to 9 clusters. The plot of those comparisons is shown in Figure 12. The "best" model is the one with the *largest* Bayes Information Criterion (BIC) value; the model with the largest BIC is the model that considers 4 clusters with ellipsoidal shapes with equal shape and orientation.

An advantage of model-based clustering is that it can generate either the best solution given a pre-specified number of clusters or it can generate the best solution for any number of clusters. The model-based clustering algorithm chose 4 clusters,

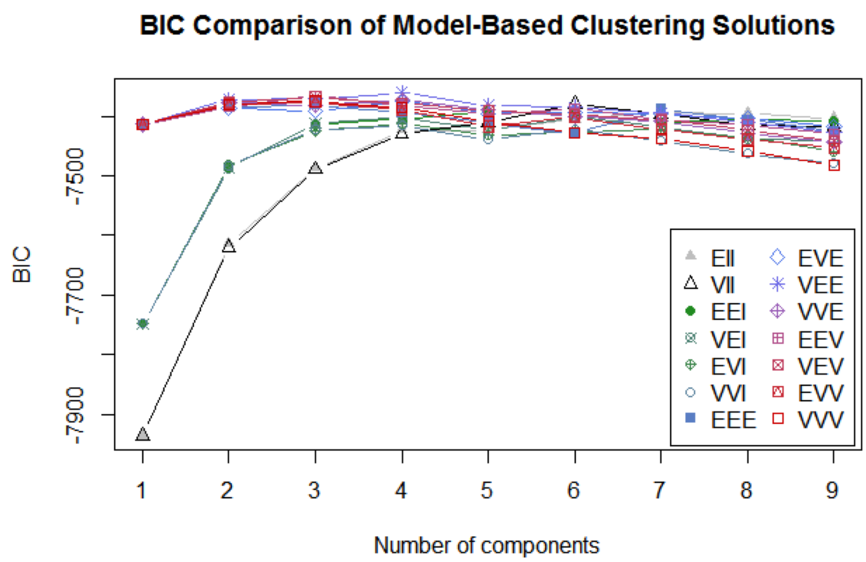


Figure 12: The model-based cluster solution that maximizes BIC is the ellipsoidal, equal shape and orientation (of the clusters) with 4 components.

which are shown in Figure 13. A contingency table showing the similarities between the actual classifications and the 4-group solution is given in Table 6.

| | R1 | R2 | R3 |
|---|---|---|---|
| MBC R1 (big) | 47 | 0 | 0 |
| MBC R1 (small) | 64 | 8 | 0 |
| MBC R2 | 4 | 99 | 31 |
| MBC R3 | 0 | 0 | 23 |

Table 6: The model-based cluster solution with 4 groups adds some institutions to the R2 category and splits the R1 category into sepearate large-institution and smaller-institution groups.

In order to keep things in line with the Carnegie Classifications, I named the two top groups R1 (big) and R1 (small) because the largest universities on both indices ended up in the "big" group and the smallest R1 schools were classified in the "small" category. I calculated a misclassification rate of about 16 percent (counting both R1 (big) and R (small) as classifications in the R1 category). This implies that only about

16 percent of institutions would be classified differently under this criterion for creating groups.

This solution places most of the smaller R3 institutions in a distinct category; it clusters the large R1 and most of the R2 schools together into a single group. While most of the R1 institutions would not drop be classified with the R2 schools, this algorithm breaks up the R1 group and classifies the largest few institutions separately. A contingency table showing the similarities between the actual classifications and the 4-group solution is given in Table 6. Even more fascinatingly, the schools with the largest scores on the Per-Capita ranking appear to get classified with their similar institutions on the Aggregate Index, even if they are outlying in the y-direction.

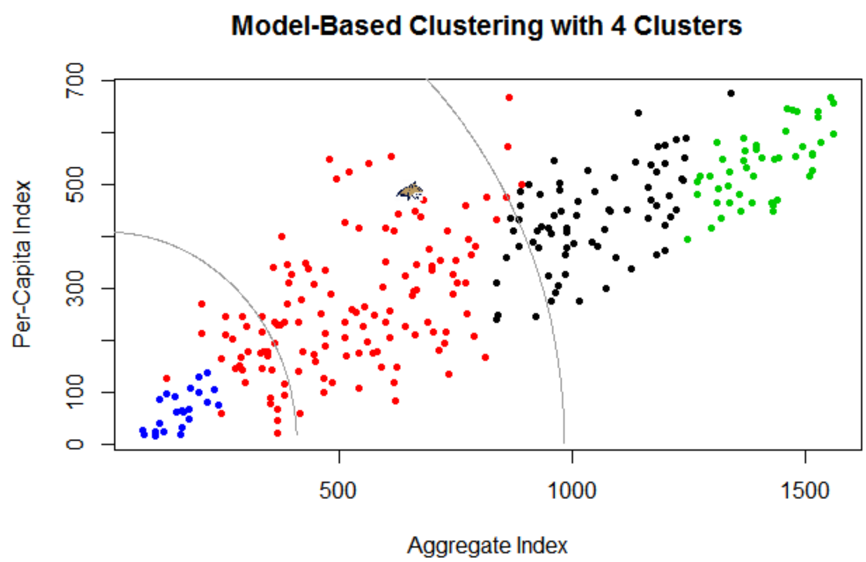


Figure 13: The model-based cluster solution is given below with the 4 number of clusters.

### 6.1.2 MBC with 3 Clusters

Since the Carnegie Classifications have only three groups, I tried fitting a model-based clustering solution with only thre clusters. Fixing the number of clusters has an interesting effect on the groups; Montana State would be an R3 school by this manner of delineation. Comparing group membership between the actual classifications and this method of creating groups allows for some interesting comparisons to be made. A contingency table comparing the groups in the classifications with the model-based clustering solution is given in Table 7.

| | R1 | R2 | R3 |
|---|---|---|---|
| MBC R1 | 93 | 1 | 0 |
| MBC R2 | 21 | 67 | 4 |
| MBC R3 | 1 | 39 | 50 |

Table 7: The contingency table comparing the optimal three group resultt to the actual Carnegie Classifications shows that using model-based clustering changes the groups fairly noticeably.

Whereas the actual classifications were based on two circles, these classifications look like they could be roughly replicated by drawing lines with positive slope. Although they look quite different from the actual Carnegie Classifications, only about 24 percent of schools would end up with a different status if the groups were calculated in this way. It does some nonsensical things in terms of classification as well; this classifies Massachusetts Institute of Technology as an R3 school. The 3-cluster model-based solution is given in Figure 14.

### 6.1.3 K-means

K-means(MacQueen, 1967) clustering is an unsupervised method that differs from model-based clustering in several ways. It can be implemented in R using the `kmeans` function (R Core Team, 2017). First, by specifying three groups, the K-means algorithm starts with three guesses at cluster means. Then, the algorithm iteratively moves the centers of each cluster in order to minimize within-cluster variance (Friedman, Hastie, and Tibshirani, 2009). The K-means clusters compared to the actual classifications are given in the Table 8. The misclassification rate is 30 percent, meaning that roughly a third of the schools would end up in different categories using this methodology for

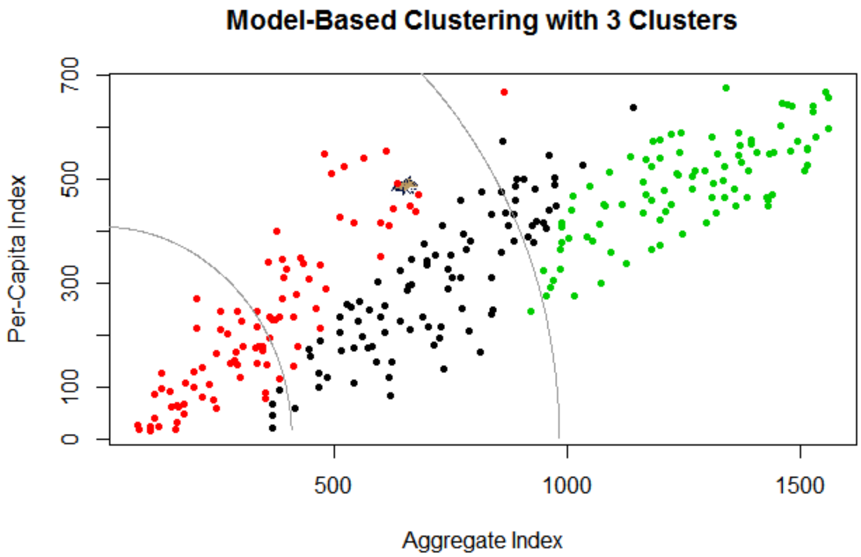


Figure 14: The model-based cluster solution is given below with the 3 number of clusters.

determining the boundaries.

| | R1 | R2 | R3 |
|---|---|---|---|
| K-means R1 | 79 | 0 | 0 |
| K-means R2 | 36 | 60 | 0 |
| K-means R3 | 0 | 47 | 54 |

Table 8: The contingency table comparing K-means with three groups to the actual classifications creates a larger cohort of R3 schools and a smaller R1 group.

Compared to the model-based clustering options, the K-means solution most resembles the groups given in the Carnegie Classifications; however, it also differs in several ways. The K-means solution increases the number of R3 institutions from 54 to 101 and the number of R1 institutions is reduced to 79 from 115. Montana State would remain classified in the R2 category; moreover, it would be farther away from the R1 group than in the Carnegie Classifications in their current form. This K-means solution is shown in Figure 15.

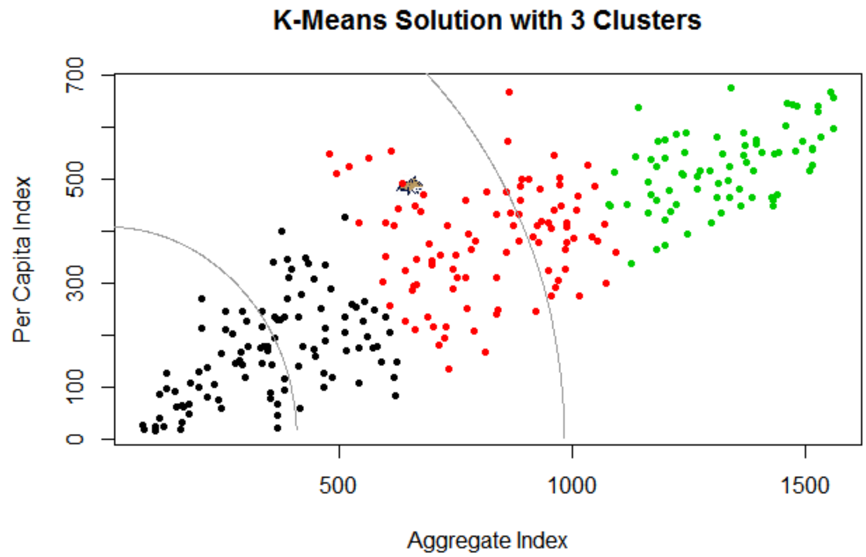


Figure 15: K-means solution with three clusters. This looks more like the Carnegie Classifications but it is more restrictive in terms of classifying schools as R1.

# 7 Discussion

## 7.1 Montana State as an R2 Institution

Montana State University had previously been classified in the highest tier of "Very High Research Activity" in the 2010 Carnegie Classification update. However, it may have been misclassified. Most R1 schools produce more doctoral degrees of each type than Montana State. The typical R1 institutions spend more money on research expenditures in STEM and non-STEM fields than did Montana State. It is worth recalling that the Carnegie Classifications are not intended to measure quality of educational experience; rather, they are used to identify similar institutions to a given school. In this case, Montana State's research output, size, and expenditures are more similar to North Dakota State University than to, for instance, the University of Michigan. The Carnegie Classifications reflect this.

## 7.2 Moving Up: How Hard Would It Be?

There is no single way to move from R2 to R1, nor is there a single way to move down to the R3 category. Both are possible for Montana State to achieve under the right set of circumstances. While breaking ties can help lead to large movements in some of the variables, this is most important for the Aggregate Index because the PhD counts are more likely to be tied. Montana State could improve from rank 1 to rank 61 in the Social Sciences PhD variable by adding just a single doctorate in the social sciences; this may actually occur with the addition of a Psychology PhD program in 2015.

Although Montana State is close to the R1 category, getting across the border would necessitate substantial gains in PhDs produced across all four categories and research staff size as well as large increases in STEM and non-STEM expenditures. The research staff and PhD counts are likely feasible; Montana State's growing infrastructure and faculty allow for increases in those variables that ought to be enough to at least move MSU towards the border. However, without nearly doubling non-STEM expenditures and adding STEM expenditures in the tens of millions of dollars, changes in the underlying variables simply are not enough to get MSU to R1.

## 7.3 Snapshots vs. Averages

The data are collected via a snapshot over the course of a year. To ensure that the data used are properly vetted, often data from the IPEDS several years prior to the update are used. For instance, in the 2020 update, the 2018 IPEDS is likely to be used. While expenditures and research staff may stay relatively constant over a period of a few years, counts of PhDs may be relatively variable. Given that uncertain nature of doctoral programs, it is difficult to predict exactly how many PhD students may graduate in a given year (at least compared to undergraduates or master's students). Year-to-year variation in the number of PhDs produced may drive some variability in the classifications. An interesting analysis - one that would require a great deal more data to be available in the Carnegie Classifications update - would be to examine a five-year average of all the variables at each institution and compare them to the snapshot-driven analysis performed by the Indiana University Center on Postsecondary Research.

## 7.4 Limitations of Static Analysis

This analysis considers changes at Montana State, holding the values at other institutions constant. Certainly, the assumption that other institutions will stay static over the next five years is unrealistic; however, we cannot predict with a high degree of certainty how other institutions will change before the next update.This analysis should be treated in a manner similar to that of a linear regression problem with multiple variables; in such a situation, parameters are estimated under the assumption that the others are held constant. In this instance, the same methodology can be applied. These small perturbations are made under the assumption that the data for other institutions are held constant.

One way for policy makers at Montana State to attempt to circumvent this problem would be to think less about the borders between the groups themselves and rather focus on moving to a point in the center of the R1 designation. The delineations between each group are the least predictable and most likely to change aspect of the classifications. Instead of trying to determine the most efficient way to get to the border, I would suggest that policy makers make getting to a central position in the R1 cluster a goal.

## 7.5 Final Thoughts

The Carnegie Classifications are useful for institutional research; however, they have little utility as a metric of institutional quality. R1 schools are not intended to be thought of as better than R2 schools; however, some campus administrators, faculty, and media have used them in this way. The delineations between each group are highly subjective; moreover, the nature of the dimension reduction used to create both indices leaves room for some debate. For instance, the Carnegie Classifications would contain more information about the underlying variables if more than a single Principal Component were used in each index. Finally, the nature of the self-reported snapshot data used in this analysis leads to the possibility of miscalculation due to either clerical errors, different definitions of variables at different schools, or other problems. The data, though mostly vetted, may not be entirely consistent across institutions. Given all this, the Carnegie Classifications should be taken with a grain of salt. They are useful tools for identifying similar institutions; however, prioritizing one status over another is not an appropriate policy goal for any institution of higher education, especially Montana State University.

# References

[1] Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth and Brooks/Cole.

[2] Borden, V. (March 2017). *Personal Communication*.

[3] Everitt, B., and Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Berlin: Springer.

[4] Friedman, J., Hastie, T., and Tibshirani, R. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

[5] Kampstra, P. (2008). *Beanplot A Boxplot Alternative for Visual Comparison of Distributions*. Journal of Statistical Software, Code Snippets 28(1). 1-9. `http://www.jstatsoft.org/v28/c01/`.

[6] Fraley, C., Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, 97(458), 611-631.

[7] MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: Statistics, 281–297, University of California Press, Berkeley, Calif., 1967. http://projecteuclid.org/euclid.bsmsp/1200512992.

[8] McClintick, W. (2016). *Carnegie Sensitivity Analysis: Moving from R2 to R1*. Poster Presentation at Rocky Mountain Association of Institutional Research Conference.

[9] University of Montana. (2017). *UM Strategic Mission 1.0 Creating Change Together*. Missoula, MT. *http://www.umt.edu/strategy/strategic-vision/UM_Strategic_Vision_1_0.pdf*

[10] West, M. (2003). Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. *Bayesian Statistics*, 723-732.

# 8 Appendix

## 8.1 Alternative Methods for Visualizing Schools

There are many ways to visualize differences between observations of multivariate data. The matplot discussed in Figure 5 does a nice job of comparing Montana State to the R1 insitutions; but schools that are more similar are harder to compare. One alternative method for comparing Montana State to its nearest neighbor institutions would be to use Chernoff Face Plots (Chernoff, 1973). These can be implemented in R using the `faces` function in the package `aplpack` (Wolf and Bielefeld, 2014). Such plots draw faces with different features to represent each variable in the dataset. For instance, eye color may represent one variable and nose height might represent another. Applied to the raw (unranked) data, I use these plots to get a sense for how similar institutions are on the raw counts of PhDs, expenditures, and faculty sizes. Figure 16 illustrates the differences betwween Montana State and its neighbors on the raw scale, with the legend included in Table 9.

Although Montana State is fairly close to several R1 schools such as Carnegie Mellon University and Rice University, these plots show that those institutions are pretty different from Montana State on the raw scales. Montana State appears to be relatively similar to Colorado School of Mines, the Naval Postgraduate School, Rensselaer Polytechnical Institute, and University of Alaska-Fairbanks; most of the differences between those schools tend to be driven by STEM expenditures and STEM PhDs produced.

## 8.2 More on Principal Components Analysis

One aspect of PCA omitted from this analysis is the directionality of the variables with respect to the new indices that are created. Biplots can be used to illustrate how related the original variables are to the new principal components. I include the biplots

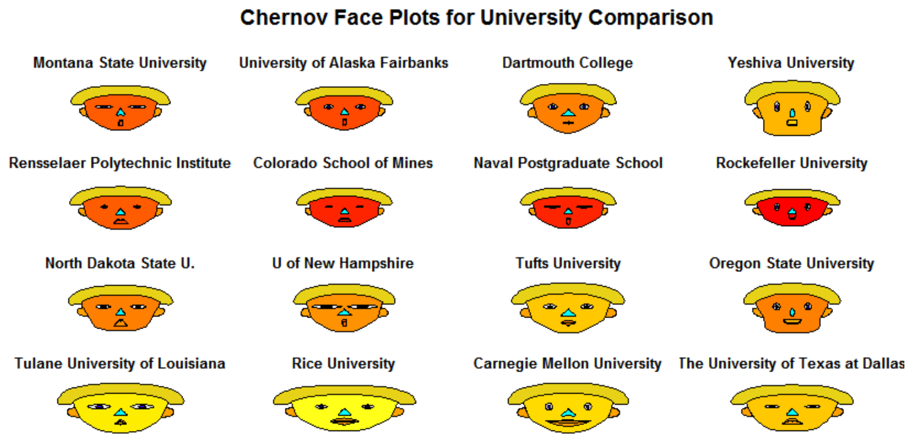


Figure 16: Face plots illustrate differences between schools. See Table 9 for interpretation of each facial trait on variable scales.

| Modified Item | Variable |
|---|---|
| height of face | Faculty Size |
| width of face | Humanities PhD |
| structure of face | Other PhD |
| height of mouth | Soc. Science PhD |
| width of mouth | STEM PhD |
| smiling | Research Staff |
| height of eyes | STEM exp |
| width of eyes | Non Stem exp |

Table 9: Legend for interpreting the Face Plots.

for both the Aggregate Index and Per-Capita Index in Figure 17.

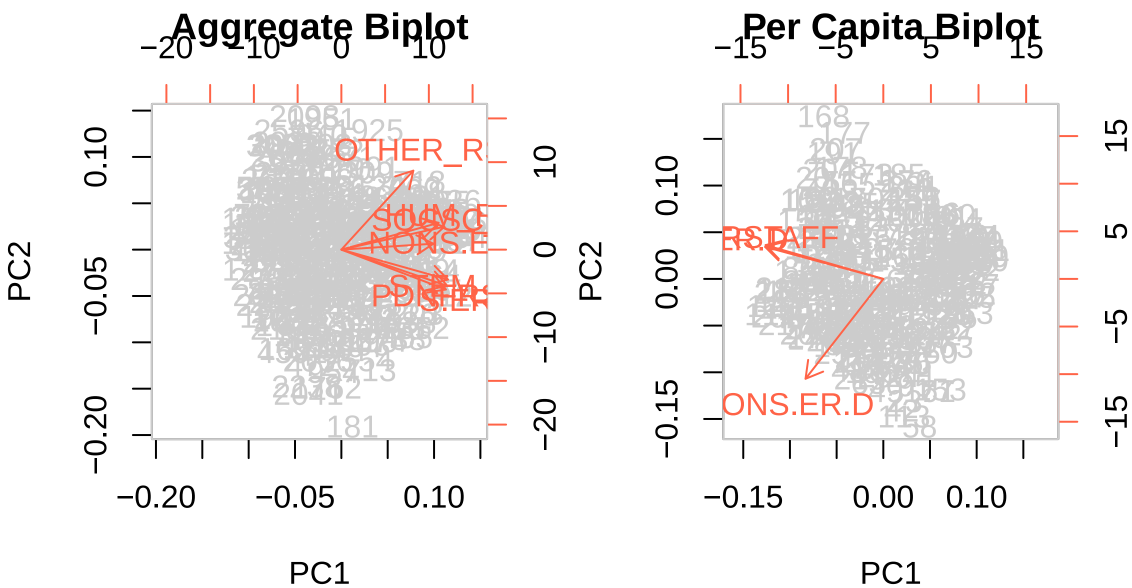


Figure 17: Biplots indicate that both indices were closely related to the first PC (partly why researchers used on the first PC). Note that for per-capita, the first PC scores were multiplied by -1 to make larger values on the first PC correspond with higher-ranked schools.