# Roadmap to Carnegie SEM Paper

1. Carnegie Classifications
   1. Context – Why do we care about CC?
   2. Methodology - Paul
   3. Critique – Paul/Mark?
      1. Sensitivity of loadings of variables - Paul
2. SEM
   1. Context - Paul
   2. **Methodology (Sarah will fill it in)**
      1. **Carnegie Method on SEM**
   3. Diagnostics (Laura/Sarah/Paul)
   4. Variability of the variables – look at which variables are impacting things the most or the least (analogue of looking at loadings on a PCA) (Paul and Sarah)
3. Model-Based Clustering (Mark)
   1. Proposed Classifications
4. **Shiny App and Sensitivity**
   1. **Shiny App for Carnegie Classifications**
   2. **Shiny App for SEM-based classifications**
5. Justification
   1. Discussion of why the model is better

Roadmap:

Model

Interpret

Compare

Shiny App

MBC

Fit Values

For meeting with Mark: (let’s try to schedule something in the next week or so)

* Cleaned up version of the code, (bare minimum)
* Give him the scores
* Have him look at clustering over the break
* Give him the shiny apps if we have them

# Introduction: Institutional Classifiers

## The Rising Importance of the Carnegie Classifications

Systems for institutional classification, ranking, and comparison are increasingly important to administrators, faculty, and students at institutions of higher education. The Carnegie Classifications for Institutions of Higher Education, the US News Ranking, and other metrics of institutional quality are utilized not only by students to make decisions as to where to go to pursue their education, but by prospective faculty in deciding where to pursue employment and research opportunities. The Carnegie Classifications are intended to be used for identification of peer institutions, not to create a ranking system; however, they are often misconstrued as metrics of institutional quality.

As such, these metrics are used to direct institutional policy. Several universities in the United States have explicitly set out policy goals directed towards improving their standing in the Carnegie Classifications. For schools in the top group, maintenance of status is the priority – schools that are in the “Highest Research Category” implement policies intended to keep them in the top group and often market themselves as such. The effect of institutional classifications may be more evident in the “High Research Activity” and “Higher-Research Activity” categories – several of these schools have implemented policy goals and timelines explicitly oriented towards achieving so-called R1 status.

While the Carnegie Classifications, US News World Rankings, and other metrics of institutional characteristics provide interesting and useful data for decision makers at institutions, they lack reproducibility and transparency. This paper attempts to illustrate the concerns associated with the current available arsenal of institutional classifications, and presents an alternative classifier based on a single score from a Structural Equation Model.

## Methodology of the Carnegie Classifications

Although the exact methodology is not documented, the Carnegie Classifications can be roughly replicated via the following high-level algorithm. The data are separated into two distinct datasets, an aggregate and a per-capita dataset. The aggregate dataset contains the following 7 variables:

* Stem PhDs
* Humanities PhDs
* Social Science PhDs
* Other PhDs
* Stem research expenditures
* Non-Stem research expenditures
* Research Staff size

The per-capita dataset contains the last three variables mentioned above, but divides them by the size of the tenured/tenure-able faculty.

Because the size of some of the largest universities dwarfs the size of many of the smallest institutions in the dataset, the data are highly skewed. Rather than using the raw data, the schools are ranked and the ranked datasets are put into two Principal Component Analyses (PCAs). The first scores from each of the PCAs are taken to form an Aggregate Index and a Per-Capita index. In previous iterations of the Carnegie Classifications, these indices only explained between 68 and 72 percent of the variation in the underlying data. In the 2015 update, these were 70 and 72 percent for the Aggregate and Per-Capita scales, respectively.

The Aggregate scores (x-axis) for each school were plotted against the Per-Capita scores (y-axis). The plot of scores is then partitioned into thirds by hand – this is the most subjective part of the Carnegie Classifications – by drawing concentric circles that separate points in the data cloud. Schools in the bottom left corner of the plot are in the “High Research” category, and those in the top-right corner are in the “Highest Research” category, with the “Higher Research” category in the middle.

## Problems with the Carnegie Classifications

The Carnegie Classifications are intended to be tools for institutional comparisons, not for ordinal rankings of schools. Based on the metrics used in their calculation, they have little to say about undergraduate education or outcomes. However, they are routinely used by policymakers on campuses to drive institutional goals and academic development. Some universities in the second and third tier of institutions have even gone so far as to set explicit policy goals directed at moving into the Highest Research Category.

Universities that choose to move towards the next highest goal, however, must shoot at a moving target. Because the data used to calculate the Carnegie Classifications in any given year are based on ranked snapshots at a single time point, the weighting of a single factor can change from year to year. This is not always drastic; however, if large changes were to occur in the characteristics of many of the universities in the calculation, it is possible that the loading for a given variable, say STEM PhDs, could be different in the next release of the classifications.

At some schools, this has driven the narrative that the classifications substantively changed, and that these changes resulted in shifts in the classifications for certain borderline schools. A school may determine, based on the weights used in the 2015 calculation of the classifications, that it needs to gain a certain number of PhDs in several fields and increase expenditures by some amount. That school might implement those changes, only to find out that because of the changes in the underlying PCA that generates each school’s score, those policy goals were not enough in the updated 2018 classifications.

This illustrates an important shortcoming of the Carnegie Classifications, as their loadings can change from one release to another simply due to changes in the underlying data. While this is not likely to be extreme, this can affect institutions trying to move up a category because policy goals tied to sensitive metrics in the current update may not have the desired effect.

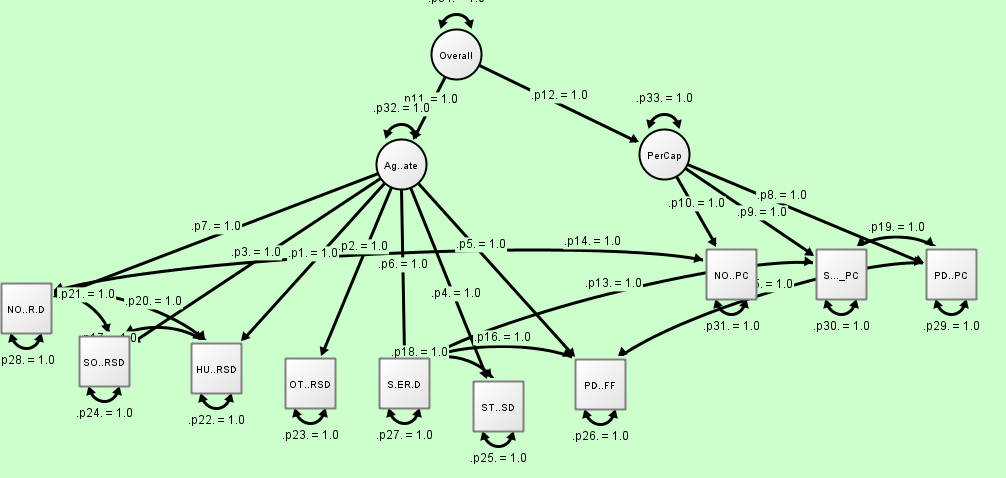
The SEM will have this same issue but we can go back to the model and easily explain why things change from year to year.

Does SEM model address this issue? Need to write a section about how SEM would do better with prior choices of loadings.

Arbitrary cutoffs are a problem! Write a section about this, and how model-based clustering should be

# An Alternate Approach: Structural Equation Models

Using a PCA-based approach is not the only way to develop an index for institutional characteristics. Alternatively, we propose a classification system built on Structural Equation Models, or SEMs (Bollen \_\_\_). Rather than creating two individual indices of institutional characteristics, a SEM allows for modeling of latent traits that can be built from a set of manifest variables that are measured. Using the same dataset, we proposed fitting an SEM-based analogue to the Carnegie Classifications.

The manifest variables would be mapped to two latent traits in the same way as before, with the Aggregate latent trait comprising the counts of PhDs awarded, Research Staff, STEM-based expenditures, and non-STEM research expenditures. In the same vein, the per-capita latent trait would similarly be based on the per-capita versions of the research expenditures and Research Staff. Structural Equation Modeling allows for modeling of covariances between the manifest variables loaded onto different latent traits, so we proposed a model that allowed for correlation between the per-capita and raw versions of each variable. Finally, the two latent traits would be combined to form a single index, called a factor of factors. This single score would be used to cluster schools based on a more objective method than splitting into three groups; rather, we proposed to use univariate model-based clustering methods to determine the optimal cluster solution. 

Additional Section on Sensitivity of the Carnegie Classifications vs. the Sensitivity of the SEM Classifications based on the two Shiny Apps

# Model Process

Structural Equation Modeling (SEM) was used to replicate the Carnegie Classification Method. Two latent factors were constructed: one with the 7 aggregate variables and another with the 3 per capita variables. These two latent factors were loaded onto a single factor of factors (**Figure of SEM Paths of this model**). However, this model failed to converge, which suggested a model misspecification, namely that the two latent factors were too similar to be separated. Although the ranked per capita manifest variables are not exactly correlated with their ranked aggregate counterparts, the correlations are very close to 1 (“SERD”: 0.92, “Non-SERD”: 0.88, Research Staff: 0.96). This is a serious issue in SEM because we are unable to identify how much variability in the data is from each manifest variable (**citation**).

A correlation matrix plot (**Figure of correlation matrix**) showed that the aggregate variables are naturally divided into two groups. The first includes **[STEM variable names]** and the second includes **[Non-STEM variable names]**. A new model was constructed which loaded these variables to two latent factors, cross-loading number of faculty onto both latent factors to emulate the per-capita variables without using them directly. The two latent factors were then loaded onto a factor of factors. This model was able to converge with Huber-White robust standard errors.

# Model Fit Assessment

Model fit was fair (RMSEA: 0.141, CFI: 0.958, Chi-Square: 6.5), with fit characteristics suggestive of good relative fit, but an absolute fit that may need improvement. That is, the model is comparatively better than an alternative with only an intercept, but does a middling job recreating the variability of the data. The proportion of the variability in the manifest variables explained by the latent traits was above 0.7 for all variables with the exceptions of “Other RSD,” “Non-SERD,” and number of faculty (**Table(s) of variability proportions??**).

This model gives a different interpretation than the Carnegie Classifications. Scores are computed as averages of the ranks and weights associated with each manifest variable. Each university receives three scores, a humanities score and a STEM score, and a weighted average aggregate score of the two. **[Need some help with estimate interpretation.]**

# Determining Scores: Univariate Clustering

In contrast with the Carnegie Classifications, which used two indices to determine cluster membership for each university, the SEM-based rankings could use the latent factor of factors, or overall score. This gives a single number for each school that could be partitioned into groups in several ways. Determination of the cluster partitions may be the most subjective aspect of the Carnegie Classifications because the lines dividing groups are hand drawn each year. In years when the data are poorly separated, determining an optimal place for the lines to be drawn is ambiguous at best. Rather, we seek an objective method for defining both the number of optimal groups and cluster membership.

# Model Based Clustering

[fill in section with some discussion of MBC method and talk about the results that we had]

[include plots from Mark’s output]

## Sensitivity of the Carnegie Classifications

It is of interest to determine how sensitive both metrics are to changes in the underlying data, especially since institutions are driving policy intended to move up in Carnegie rank. We developed a Shiny application designed to allow the user to select a school and assess the sensitivity of that school’s classification to changes in the underlying variables. The user could select a school and then use a slide bar to either increase or decrease the number of PhDs awarded in each category, research staff size, or research expenditures. Changes can be made to either a single variable, all of them, or just a select group. The application takes the user input and re-calculates the PCA-based indices based on the new dataset, and shows where the university would be relative to other schools in that update.

[insert sections from writing project]

## Sensitivity of the SEM Classifications

#Conditional Panel for Shiny App

* Might work better for having different things pop up when you select a school