February 2016

Understanding the Basic Carnegie Classification System

Recently, the Carnegie Classification of Institutions of Higher Education released an updated version of its classification system. Though some portions of Oakland University's classification were revised, there was no change to its 'basic' classification. (We are classified as an 'R3', or a Doctoral University – Moderate Research Activity).

But how close were we to being classified as an R2 (Higher Research Activity)? Based on our estimates, very close.

The [Carnegie Classification website](https://www.oakland.edu/OIRA) not only provides a copy of the dataset they use in their classification methodology, but it actually contains enough information to recreate their methodology. Though the Office of Institutional Research and Assessment (OIRA) is not able to reproduce the methodology exactly, we are able to recreate the classifications with an accuracy of about 95%. Thus, we can use the OIRA estimates to see how close Oakland University was to the demarcation line between R3 and R2 institutions.
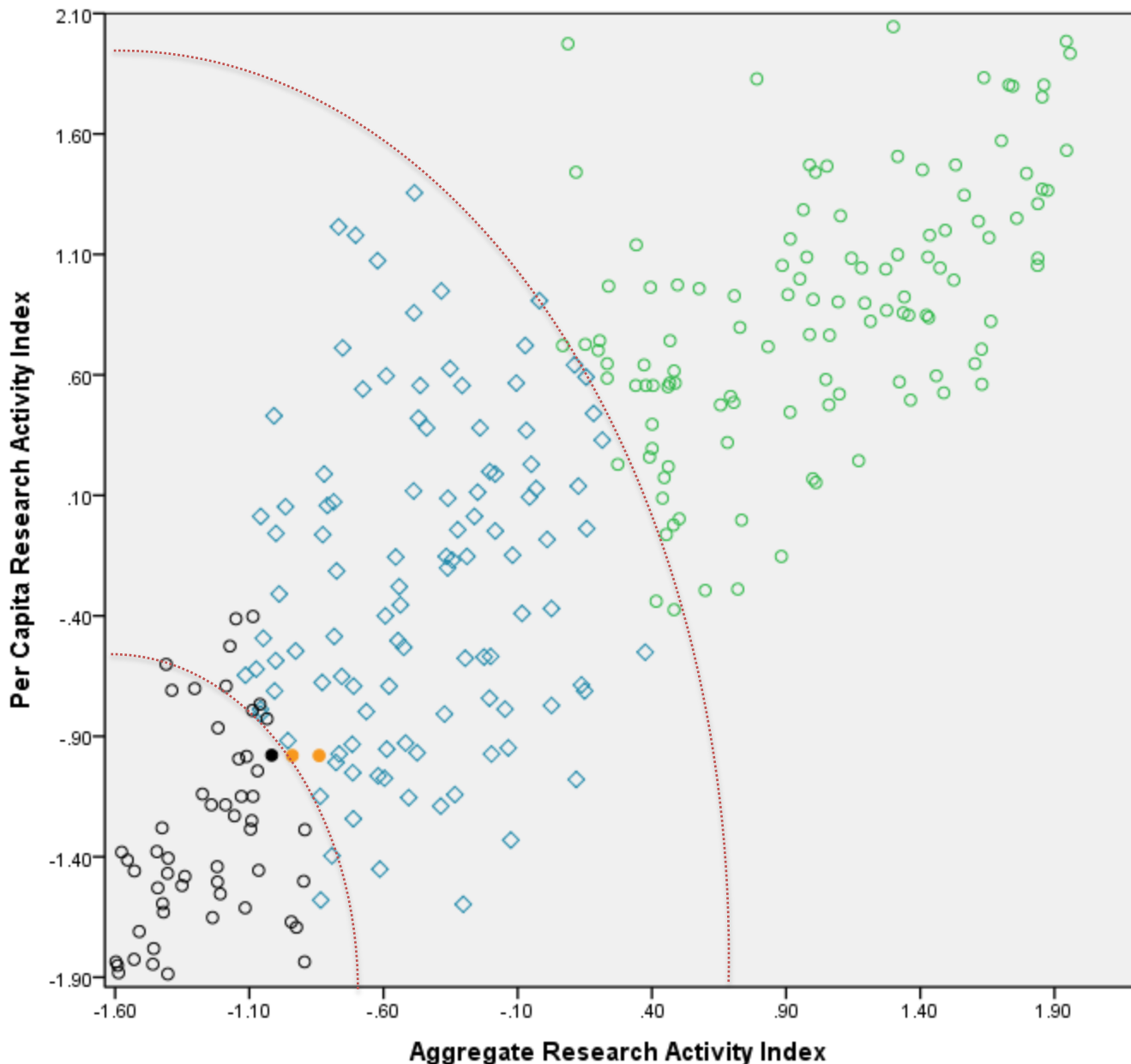
Carnegie Classification Data Points

1. STEM Research Expenditures

2. Non-Stem Research Expenditures

3. Number of Research Staff

4. Number of STEM Doctorates Awarded

5. Number of Humanities Doctorates Awarded

6. Number of Social Science Doctorates Awarded

7. Number of Other Doctorates (Excludes Professional Practice Degrees such as: DPT, DNP, MD)

The key to understanding the Carnegie Classification System is to understand how the system uses the base data to determine research output. The system uses only seven different data points (i.e. correlates of research output). From these data points, it creates two different metrics of research activity – An Aggregated Research Index and a Per Capita Research Index. These metrics are then plotted on a graph. The distance from the lowest institution determines the demarcation lines that separate R1, R2, and R3.

To create the Aggregated Research Index (ARI), each institution is first ranked on each of the 7 data points. Thus, each institution will get a number (1 through 276 for the 2015 ranking) for each of the 7 data points listed. These ranks are then summed. This sum is then standardized to produce a value between -2 and 2, where 0 represents the average summed rank. This means that for the ARI each of the 7 variables has exactly the same weight. Furthermore, the absolute value of the data point is irrelevant – only the rank is relevant.

The second index, the Per Capita Research Index (PCRI) uses only 3 of the 7 data points; STEM Research Expenditures, Non-STEM Research Expenditures, and the Number of non-Faculty Researchers & Post Docs. However, instead of ranking the total dollars amounts or numbers, each of these data points is divided by the total number of faculty at each institution – creating a per capita version of the existing data points.

Prepared by the Office of Institutional Research & Assessment. Visit us at: www.oakland.edu/OIRA

1

Once again, these three new data points are each ranked separately. The PCRI is then computed by summing the ranks of these three new data points. Just as with the ARI, the PCRI sum is then standardized to produce a number for each institution between -2 and 2, with zero representing the average sum of ranks for the three variables that comprise the PCRI.

Once each institution has both a standardized ARI and a PCRI score, they can be plotted on a simple graph. OIRA has recreated this graph using the method described above. The black dot below represents Oakland University. The colors of the circles represent the Carnegie Classification's designation between R1 (blue), R2 (Green), and R3 (Black).

Figure 1: Graphical Results of OIRA's Carnegie Classification Reproduction

Further Discussion

Figure 1 also includes demarcation lines that separate R3 from R2 and R2 from R1 (these are simply hand-drawn estimates). Oakland University (the black dot) is about as close to the line as possible without being over the line. It's worth noting that OIRA's reproduction is not exact, it's about 95% accurate, though the demarcation is better between R1/R2 than between R2/R3. (Additional methodological limitations are discussed later).

Figure 1 also contains two additional 'orange' data points. These data points represent hypothetical versions of Oakland University not contained within the original data. Before further discussion, it is important to note that in the data collection year used in Carnegie's classification system, Oakland University did not award a doctorate in either a social science field or a humanities field.

The leftmost orange data point represents where we would have scored if we had graduated just one PhD in social sciences and changed no other data points. The right orange data point represents where we would have scored if we had graduated one PhD in social sciences and one PhD in humanities. Both hypothetical data points rest above the demarcation line. The reason for the seemingly extreme changes with just one or two additional degrees is simple: the absolute number of doctorates an institution awards in a given category is not that important – it's all about the relative rank, and lots of institutions do not offer any degrees in various categories. Moving from 0 awarded doctorates to just 1 awarded doctorate improves the rank in that category by 30 or 40 places. In turn, this creates a small change when those ranks are aggregated to create the composite index located on the x-axis (the ARI).

Oakland University does not currently have a PhD program in a humanities discipline. (The PhD in Music Education is reported as an education degree). However, we do offer a PhD in psychology, which undisputedly lies within the social science PhD category. However, the program is new and had yet to award a doctorate within the data collection year used by the classification. Were the data collection gathered in a different year, our classification may very well have been R2 instead of R3.

Below is a breakdown of our rank for each of the 10 categories of data that the Carnegie Examines – the seven categories listed on page 1, plus the three per capita categories. Unlike many other ranking schemes, higher numbers are better, with a maximum rank of 276.

Carnegie Classification Data Points	Rank (Higher is Better)
1. STEM Research Expenditures	52 of 276
2. Non-Stem Research Expenditures	97 of 276
3. Number of Research Staff	66.5 of 276 (Tie; mean rank)
4. Number of STEM Doctorates Awarded	93.5 of 276 (Tie; mean rank)
5. Number of Humanities Doctorates Awarded	45.5 of 276 (Tied for lowest; mean rank)
6. Number of Social Science Doctorates Awarded	31 of 276 (Tied for lowest; mean rank)
7. Number of Other Doctorates	111 of 276
8. STEM Research Expenditures Per Faculty	48 of 276
9. Non-STEM Research Expenditures Per Faculty	111 of 276
10. Research Faculty & Post Docs Per Faculty	62 of 276

Caveats, Methodological Research, and Previous Research

The data presented here, particularly the graph on page 2, is simply an estimate. The methodological notes provided by the Carnegie Classification website do not quite provide enough detail to replicate the classification exactly. There are several methodological descriptions that are vague enough where multiple interpretations in data coding are possible.

In addition, it appears as if some of the methodology is purposely disguised. For example, the data download clearly identifies 54 R3 institutions and 107 R2 institutions. But when one examines the graph that comes with the data download, there are clearly 55 institutions plotted in the R3 category and only 104 institutions within the R2 category. Furthermore, while their methodological notes suggest the demarcation lines represent a circular region chosen from a central point, the graphical data strongly suggests that the region is highly elliptical, not circular. (The graph provided in the data download is warped - the distance on the x-axis is approximately twice the length as the y-axis. This makes the demarcation region looks as if it might be approximately circular, but it's not). Given all of this nuance, it becomes difficult to replicate the classification exactly.

Despite the noisy methodology provided in the data download, OIRA is still able to correctly classify about 95% of the given data using it's own process. This suggests that the foundation of the methodology is relatively accurate, even if some of the details are not exact.

This isn't the first time that OIRA has recreated the Carnegie Classification. It did so a few years ago in order to analyze the probability that Oakland University would be classified as an R2 during it's next iteration. In that report, we concluded that there would be a very low probability of moving into R2 designation by 2015, but a reasonable chance at moving into R2 by 2020. As it stands, Oakland University is very close to R2 designation right now, and we would revise OU's chance at reaching R2 designation in 2020 to be 'likely'.

However, R2 designation by 2020 would not be guaranteed. Most of OU's core metrics for the classification were essentially flat (or fell slightly) since 2010's classification. Yet, we are actually close to the demarcation line. The reason is that a sizeable portion of institutions have come into the bottom of the rankings, which has had the effect of essentially pushing everyone up. Several institutions with seemingly flat research activity were actually moved from R3 to R2. If this trend continues, Oakland University would likely be classified as an R2 institution by 2020 even if our research productivity were to fall slightly.