US University & College Ancestry Visualizations

Tracking the Evolution of Program Categorization Labels

Thomas Koutsidis
Indiana University
tkoutsid@iu.edu

Julia Nguyen
Indiana University
julinguy@iu.edu

Olusina Ojewola
Indiana University
oojewola@iu.edu

Mackenzie Ross
Indiana University
mackross@iu.edu

Keerthana Varakala
Indiana University
kvarakal@iu.edu

Abstract— Starting in 1973 the Carnegie Classification of Institutions of Higher Education (CCIHE) grouped universities and colleges in the United States into categories and subcategories based on the types of degrees awarded by the institutions. In total, the CCIHE has been updated ten times, most recently in 2021. The goal of the CCIHE is to provide a tool that students, researchers, and falculty can use to identify similar institutions. This projects aims to gain insights into how the classifications have changed over time through the use of static and interactive visualizations. The visualizations will provide researchers with an idea of how subcategory labels are connected and when major changes to the institutional classifications occurred.

Index Terms—Carnegie Classification of Institutions of Higher Education, visual analytics, education data

INTRODUCTION

The Carnegie Classification of Institutions of Higher Education (CCIHE) is the premier organization for cataloguing and describing diversity in institutions of higher education in the United States. In total the CCIHE has conducted ten surveys on US universities and colleges from 1973 to 2021. The classifications assigned to institutions by the CCIHE are objective and degree-based to identify similar institutions. These classifications can then be used in "research study design to ensure adequate representation of sampled institutions, students, or faculty" [1]. The surveys were not conducted at regular intervals, but at least one survey has been completed in each decade since the 1970s. The aim of the project is to use visualizations to gain insights about the data and track the changes to program sub-categorizations throughout the existence of the Carnegie Classifications.

1 INSIGHT NEEDS

The clients for this project are looking to (1) understand and identify the changes of institutional statuses to create a more comprehensive list of institutions, (2) visualize the evolution of institutional classifications to accurately provide insights for students, policymakers, and researchers, and (3) display program offerings and research expenditures. Thus, the insight needs for the project are: (a) trends, (b) categorization, and (c) comparisons. Trends in the number of sub-category labels over time will indicate which groups have changed the most/least and any years where major change occurred. Using the group labels and sub-category labels to categorize the visualizations will allow the client to investigate changes in their desired topic without having to look at all of the groups or sub-category labels. Finally, comparisons between the different groups can reveal patterns that have developed when the CCIHE performs updates. These insight needs will allow the clients to find the answers to their questions.

1.1 Stakeholder Analysis

The stakeholders for this project are the researchers at the CCIHE, specifically Dr. Victor Borden and PhD Candidate Daria Ivleva. Dr. Borden has directed the Carnegie Classifications since 2009. This project will be beneficial for Dr. Borden and PhD Candidate Ivleva because it will provide them with outside observations about the previous classification surveys and could help guide future iterations of the classifications.

2 DATA

The data for this project covers all ten updates to the CCIHE's classifications from 1973 to 2021. The data was provided by Daria Ivleva, a project research assistant at the CCIHE.

2.1 Description of Data

The data is contained in an Excel file with six sheets. The primary sheet for data acquisition was "Comb_1973-2021" which contain 15 columns and 38,811 rows of data. In total, the dataset contains information about 13,923 institutions. There are seven groups in the dataset: Doctoral, Master's, Bachelor's, Associates, Special Focus: 4 Year, Special Focus: 2 Year, and Bachelor's/Associate. In 1973, the earliest survey, a total of 2825 institutions were surveyed. In 2021, the most recent survey, a total of 3938 institutions were surveyed. Table 1 contains the column titles in the "Comb_1973-2021" sheet. The other sheet used for the visualizations and analysis was "YearCatsSlicer" which contained pivot tables with the count of institutions by subcategory labels for each year.

Table 1: Description of columns in "Comb_1973-2021" sheet

Column Title	Description
INSTNAME	Institution Name
UNITID	Institution's Unique Identifier
AllUpdates	Binary value indicating if an institution appears in all 10 updates
STATE	State where institution is located
CITY	City where institution is located
YEAR	Classification/update year
VALUE	Numeric classification value
YrValKey	Combination of YEAR and VALUE representing the classification label
CATEGORY	Label of classification for the year
IDYrKey	Combination of UNITID and YEAR to link information on state to "INSTNOT…" files
SubCatLbl	Sub-Category label for institution classification
GrpLbl	Group label for institution classification
Status	Institution status

LinkUnit	UNITID of an institution that was merged into or came out of
Notes	Any notes from the individual who tracked status

3 Analysis Methods

For this project, statistical, temporal, and relational analysis methods were used. Statistical analysis was used to gain insights into the overall dataset. Temporal and relational analysis was used to track changed to group and sub-category labels over time.

3.1 Analysis Methods

First, statistical analysis of the data was conducted using Python to get an idea of how many institutions were classified in each group for each update. Table 2 shows the breakdown of institutions by group label for each update.

Table 2: Number of institutions in each group for each update

	1973	1976	1987	1994	2000	2005	2010	2015	2018	2021
Doctoral	173	184	213	236	261	283	282	334	422	468
Master's	453	594	595	529	610	663	652	763	672	667
Bachelor's	718	583	572	637	552	647	634	572	575	531
Associates	1060	1146	1366	1471	1669	1705	1587	1113	1000	948
Special Focus: 4 Year	421	565	641	722	793	838	775	1036	965	782
Special Focus: 2 Year	-	-	-	-	-	-	-	444	435	340
Bacc/Assoc	-	-	-	-	57	229	220	403	269	202
#N/A	-	-	-	25	-	26	483	-	-	-
Total	2825	3072	3387	3620	3942	4391	4633	4665	4338	3938

Python was also used to calculate the number of entries in each category overall and number of entries in each status.

Table 3: Number of entries in each category

Category	Number of Entries
0 – Error, Not in Universe	534
1 – Doctoral	2856
2 – Master's	6198
3 – Bachelor's	6021
4 – Bacc/Assoc	1380
5 – Associates	13065
6 – Special Focus: 2 Year	1219
7 – Special Focus: 4 Year	7538

Table 4: Number of entries in each status

Label	Number of Entries
-8: Born out of	0
-2: Not yet born	0
-4: Dead no survivors	1064
-5: Stopout	76
-6: Merged into/related to	448
-7: Return	1
-9: Name change	17

These table reveal patterns in the data. For example, the Special Focus: 2 Year and Baccalaureate/Associate's categories are relatively new, starting in 2015 and 2000 respectively. Additionally, the category with the most institutions is the Associates.

4 Visualizations

For the visualizations, the handling of missing and null values was determined by the needs of the visualization. For example, when visualizing data about group and sub-category labels, rows that contained null values in the columns of interest were dropped. This solution was used because of the large amount of data used for the project.

Table 3 shows a summary of the number of institutions by group label for the combined data set. We can see that there are about seven categories of different levels, each category having its own preference among the stakeholders. For example, if an individual is searching for an institution of their liking to pursue master's degree, they prefer to filter out the institutions that do not provide that degree..

Figure 1: Bar chart of the number of institutions per category in 2021

Figure 1 shows a snapshot from an animated visualization of number of institutions by category for each update year. In the animation, there is a frame for each update year. Figure 1 shows the frame of the plot for the 2021 update. The Python library matplotlib was used to create the animation. The update function is called for each frame of the animation. The function clears the axes, selects the data for the current year, groups the data by category, and plots a bar chart. The animation can be saved each time the code is run. The categories, values and colors are extracted from the data which is grouped by category to get the count of the institutions. As the animation moves quickly through the plots for each year, it is not optimal for comparing number of institutions by category for each update.

4.1.2 Visualization 2

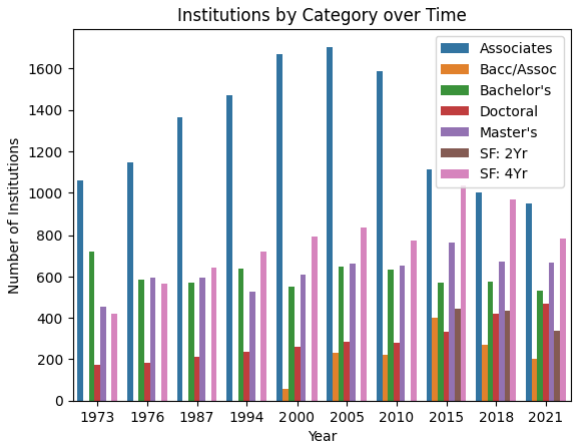

Figure 2: Bar chart of the number of institutions in each category for each update

Figure 2 shows a bar plot of the number of institutions in each category for each year. This visualization allows for comparisons between the different classification updates and is more effective than the animation described in the previous section. Many new institutions were added over time and, as a result, the number of institutions in each category increased. Until the year 2000, there were only five categories for institutions (Doctoral, Master's, Bachelor's and Special Focus: 4 Year). In the 2000 update, the Bachelor/Associate category was added. In the 2015 update, the Special Focus: 2 Year category was added.

4.1.3 Visualization 3

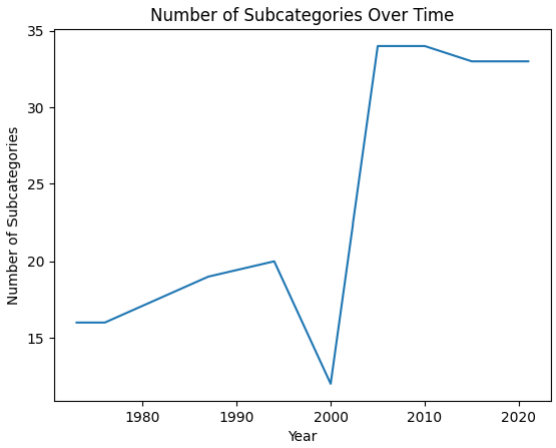

Figure 3: Line chart of the total number of subcategory labels over time

Initially, there were just over 15 total subcategory labels. The number of subcategories first peaked in the 1994 update, then dropped to an all-time low in the 2000 update. There is a sharp increase in the 2005 update. Since 2005, the number of subcategories has held steady.

4.1.4 Visualization 4

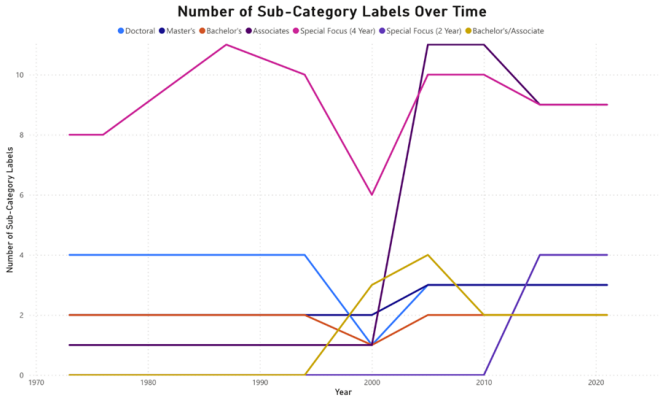

Figure 4: Line chart of the total number of subcategory labels by category over time

This line chart expands upon the line chart in Figure 3. Each degree category is represented by a different colored line, so we can see how the number of subcategories for each degree type changed with each classification update. As noted in the previous section, there was a large increase in the number of subcategories in the 2005 update. Figure 4 reveals that the number of subcategories in the Associates category increased from one to eleven. All categories experienced an increase in subcategories in the 2005 update, but the Associates category was by far the largest.

4.1.5 Visualization 5

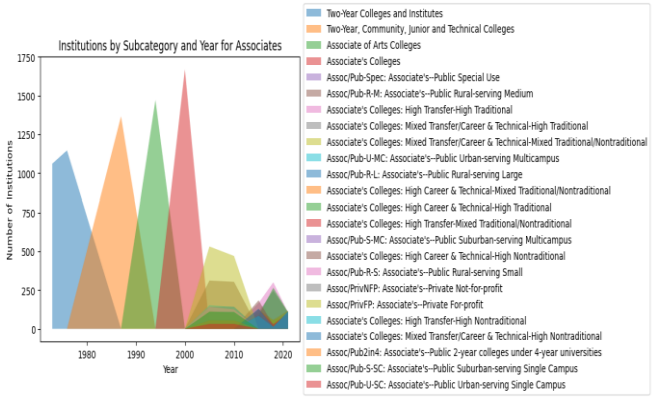

Figure 4: Distribution of institutions in the Associate's category by subcategory for all updates

The visualization and the following visualizations show the distribution of subcategories for each of the degree categories. Stacked area graphs are used with the area representing the number of institutions in each subcategory every year. Figure 5 shows the subcategory distribution for the Associates category. The plot shows that the Associates category has been active for each update to the CCIHE. After 2000, there is an increase in the number of subcategories, as described in Section 4.1.4. The graph is more clustered towards the end because of the increase in subcategories.

4.1.6 Visualization 6

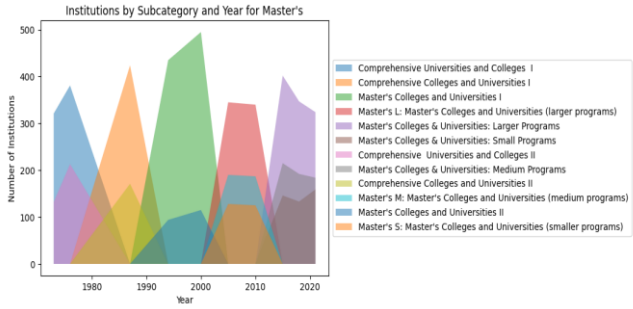

Figure 5: Distribution of institutions in the Master's category by subcategory for all updates

Figure 6 shows the distribution of the subcategories for the Master's category. This category has been active since 1973. This distribution of subcategories is relatively constant throughout the time period. There are some fluctuations in the number of institutions in each subcategory, but this graph looks more evenly distributed than the Associates graph. There are fewer subcategories in the Master's category than the Associates category.

4.1.7 Visualization 7

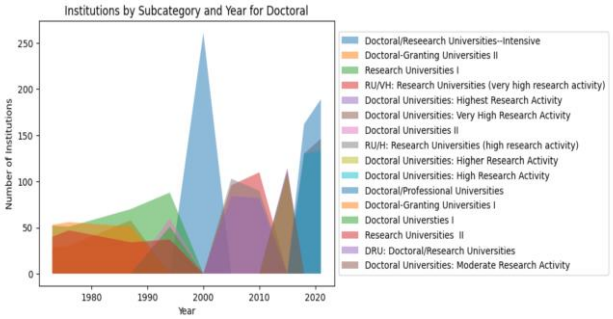

Figure 6: Distribution of institutions in the Doctoral category by subcategory for all updates

Figure 7 shows the subcategory distribution for the Doctoral category. Like the Associates and Master's categories, the Doctoral category has been present in the CCIHE for all updates. The distribution institutions per subcategory is mostly constant, but there is peak in 2000 when only one subcategory (Doctoral/Research Universities – Intensive) existed. The number of subcategories in this category is greater than the Master's category but less than the Associates category.

4.1.8 Visualization 8

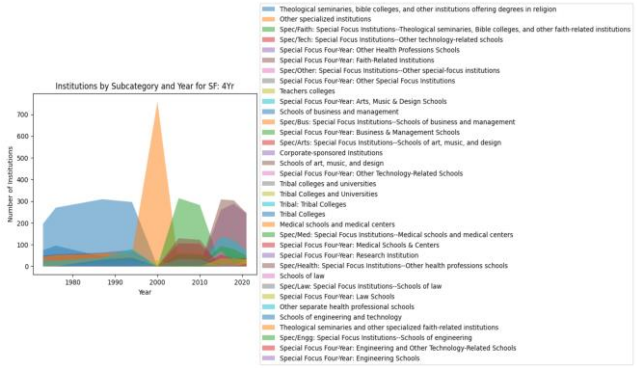

Figure 7: Distribution of Institutions in the Special Focus: 4 Year category by subcategory for all updates

Figure 8 shows the distribution of institutions by subcategory for the Special Focus: 4 Year category. This category has been since 1973 and is still active. Similar to the Doctoral category, this category also has a peak during the year 2000 when only one subcategory existed. Institutions do not appear to be equally distributed. Some subcategories have a greater number of institutions than others. The Special Focus: 4 Year category has the highest number of subcategories among all the seven categories.

4.1.9 Visualization 9

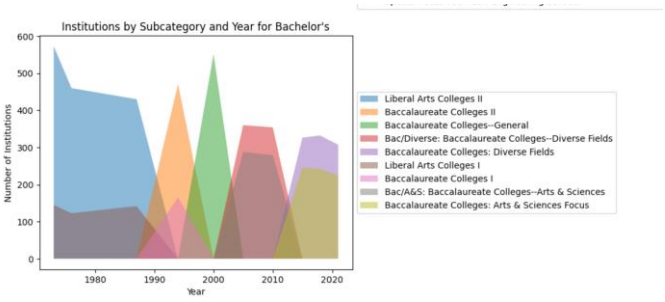

Figure 8: Distribution of institutions in the Bachelor's category by subcategory for all updates

Figure 9 shows the distribution for the Bachelor's category. This is the final category that has existed for all updates of the CCIHE. There are a greater number of peaks in this graph than the graphs for the other categories. This may indicate that more institutions are grouped into the same subcategory in this category than in other categories. There are fewer subcategories for the Bachelor's category in each update than in the previous categories.

4.1.10 Visualization 10

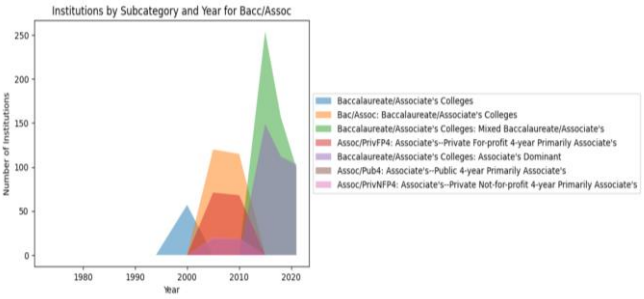

Figure 9: Distribution of all institutions in the Baccalaureate/Associate's category by subcategory for all updates

Figure 10 shows the distribution for the Baccalaureate/Associate's category. This is a newer category that first appeared in 2000. Initially, there was only one subcategory, but more subcategories were added in subsequent updates. In 2018, the number of institutions in the Mixed Baccalaureate/Associate's subcategory reached a high of almost 250 institutions. As this is a newer category, it will be interesting to see how the number of subcategories will change in future.

4.1.11 Visualization 11

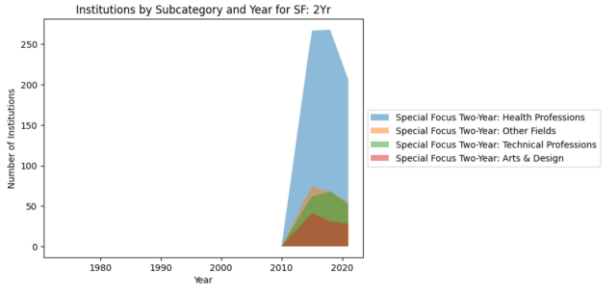

Figure 10: Distribution of all institutions in the Special Focus: 2 Year category by subcategory for all updates

Figure 11 shows the distribution of institutions by subcategory for the Special Focus: 2 Year category. This category is the newest category. It was added in the 2015 update. There are 4 subcategories present in the graph, and they have existed for the 2015, 2018, and

2021 updates. A majority of institutions have been assigned to the health professions subcategory in each update.

5 Interpretation of Results

Using multiple analytical methods on the CCIHE dataset, we discovered that the greatest change in the classifications occurred in the year 2000. The 2000 update was an inflection for the number of subcategories for each degree type. From 1994 to 2000, we saw the greatest drop in the number of subcategories. In addition to this, we saw the greatest increase in categories from 2000 to 2005. Since 2018, the number of subcategories has remained the same. The changes in subcategories are necessary to accurately describe and define universities.

The Carnegie Classification of Institutions of Higher Education has tracked Doctoral, Masters, Associates and Special Focus Four-Year programs since it began tracking in 1973. Special Focus Two-Year programs were introduced in 2015 and have been tracked ever since. The number of total institutions tracked has increased from 1973 to 2015 but saw a decline in 2018 and then again in 2021. Despite the decline in the number of institutions tracked, the number of degree programs that are tracked by the CCIHE has increased throughout its entirety. Almost every update has a change in the number of subcategories within each category, within each program.

Despite these results, there were 534 universities that did not exist in the universe of Carneige's classifications. These institutions could not be categorized into degree types because they were non-accredited or non-degree granting institutions in their respective years. For example, Selma University had lost its accreditation by the time the CCIHE dropped their 2010 update. This is reflected in the data as a category number of 0 (or ERROR: NOT IN UNIVERSE). When Selma University received accreditation in 2014, the 2015 update reflected this and had it categorized as a 7, meaning it was a Special Focus Four-Year program.

Acknowledgments

The authors wish to thank Dr. Victor Borden and Daria Ivleva. This work was supported in part by Indiana University and Dr. Andreas Bueckle.

References

[1] The Carnegie Classification of Institutions of Higher Education (n.d.). About Carnegie Classification. Retrieved April 20, 2023 from https://carnegieclassifications.acenet.edu/.

[2] J. Allebach. Binary display of images when spot size exceeds step size. Applied Optics, 15:2513–2519, August 1980.

[3] E. Catmull. A tutorial on compensation tables. In Computer Graphics, volume 13, pages 1–7. ACM SIGGRAPH, 1979.

[4] Peter Litwinowicz and Lance Williams. Animating images with drawings. In Andrew Glassner, editor, Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994),Com- puter Graphics Proceedings, Annual Conference Series, pages 409–412. ACM SIGGRAPH, ACM Press, July 1994.

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron Spectroscopy Studies on Magneto-Optical Media and Plastic Substrate Interface," *IEEE Trans. Magnetics,* vol. 2, pp. 740-741, Aug. 1987. (IEEE Transactions)

[6] S.P. Bingulac, "On the Compatibility of Adaptive Controllers," *Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory*, pp. 8-16, 1994. (Conference proceedings)

[7] J. MacQueen, "Some Methods for Classification Analysis of Multivariate Observations," *Proc. Fifth Berkeley Symp. Math. Statistics and Probability,* pp. 281-297, 1967. (Conference proceedings)