

# UVM ScholarWorks

## A Temporal Analysis of Humanities Representation in Higher Education Course Catalogs

Item Type	undergraduate thesis
Authors	Brim, Ryan;Chawla, Aviral;Lovato, Juniper
Download date	2026-06-09 21:38:02
Item License	<a href="http://creativecommons.org/licenses/by-nc-nd/3.0/">http://creativecommons.org/licenses/by-nc-nd/3.0/</a>
Link to Item	<a href="https://hdl.handle.net/20.500.14849/5645">https://hdl.handle.net/20.500.14849/5645</a>

# A Temporal Analysis of Humanities Representation in Higher Education Course Catalogs

Ryan Brim<sup>a</sup>, Aviral Chawla<sup>a,b</sup>, and Dr. Juniper Lovato<sup>a,b</sup>

<sup>a</sup>Department of Computer Science, University of Vermont, Burlington, 05405, U.S.A.

<sup>b</sup>Vermont Complex Systems Institute, University of Vermont, Burlington, 05405, U.S.A.

## ABSTRACT

The "death of the humanities" has been a subject of much discussion, with some lamenting the death of these subjects so important to our experience as humans, and some celebrating the greater focus on the sciences and "more important" fields. But, what are the actual trends in the prevalence of the humanities: are they even really in decline? In this study, we analyzed course catalogs for 25 American institutions between academic years 2003-04 and 2021-22, along with statistics on enrollment demographics and academic library spending from the National Center for Education Statistics as part of the Integrated Postsecondary Education Data System (IPEDS), with the goal of examining the trends in the humanities over time as well as the relationships between the prevalence of the humanities, student demographics, and library spending. Our findings suggest that the humanities are not in decline and that the difference between academic library spending on one-time purchases and subscriptions can be a good predictor of the prevalence of the humanities. Our results also suggest that institutions with more students of Asian or Pacific Islander descent and non-United States resident students are more likely to have a greater focus on the humanities. In contrast, institutions with greater enrollment of American Indian or Native Alaskan, white, black, and Hispanic students are more likely to have less focus on the humanities. These results point to possible preferences of students of different demographics. We hope these insights can inform discussions on popular views of humanities subjects in higher education.

## 1 Introduction

The humanities encompass fields central to the human experience, and human expression has been a core element of higher education from the very beginning. However, in recent years, there has been much discussion about the "death of the humanities" and a noted decline in the prevalence of the humanities fields in higher education. With modern technical advances, it is feared that non-humanities fields, such as those in the sciences, medicine, and engineering, have superseded the humanities in perceived importance and priority. Is this fear justified? In an effort to examine this question, this thesis will investigate course catalogs from academic years 2003-04 to 2021-22 for 25 American universities.

Course catalogs are a largely untapped source of data on the evolution of higher education. Through their information on course availability, they provide insight into the development of fields over time, including how courses spread across institutions and the impact of historical biases and prejudices. Through the analysis of these historical course catalogs, this thesis examines the trends in the prevalence of the humanities over time at these institutions and will seek to understand if the humanities are, in fact, dying and what factors might be related to their decline.

### Research questions:

- RQ1: How has the relative prevalence of the humanities changed over time?
- RQ2: How do undergraduate student demographics such as gender, race, and United States residency status relate to the trends in the humanities?
- RQ3: How does university library spending on one-time purchases and subscriptions relate to the trends in the humanities?

This thesis utilizes course information from university course catalogs collected from university websites alongside demographic and spending information collected by the National Center for Education Statistics as part of the Integrated Postsecondary Education Data System (IPEDS). The analysis of these data was done in Python.

Section 2 will review relevant literature, section 3 will discuss the data and methods used in analysis, section 4 will discuss the results of the analysis, and section 5 will discuss the findings, implications, and limitations of this work.

## 2 Background

The "death of the humanities" is often talked about as a given, but what are the actual trends in institutional investment and student interest in humanities fields? In his paper "Ending the budget wars: Funding the humanities during a crisis in higher education," Newfield breaks down data from the National Center for Education Statistics and shows that the percent change from 1970 to 2006 in the number of bachelor's degrees conferred in humanities fields is comparable to other fields, with only English and foreign languages having shown a decrease.<sup>1</sup> In another article by Zuckerman and Ehrenburg, they discuss how funding for the humanities has increasingly gone to places like museums and historical projects instead of the academic humanities.<sup>2</sup> They also discuss how lower rates in the growth of expenditure on monographs, longer and more specialized books which are preferred by the humanities, as compared to subscriptions to peer-review journals may show a shift in focus away from the humanities and towards the sciences. Going against the view that the humanities are certainly declining, Pippins, Belfield, and Bailey argue that, even though the number of bachelor's degrees awarded has fallen, when associate's degrees awarded by community colleges are factored in, the number of humanities degrees has risen overall.<sup>3</sup> They also point out that humanities courses continue to be a core part of college curricula, even for STEM students, so the humanities will likely continue to have a presence even as STEM programs grow. The focus on the humanities in higher education does seem to be declining, but perhaps not at such a drastic rate. This thesis offers further insight into these trends in the humanities and examines them over a longer time span than the previous literature.

Why is the prevalence of the humanities declining? Perhaps it is because the sciences, engineering, and medicine, which require much more funding than the humanities to conduct research, are growing and eating into the budgets for humanities fields.<sup>1</sup> Taylor, Cantwell, and Slaughter conceptualize universities as being in a quasi-market controlled by government policy makers and hypothesize that policy makers that control federal grants may favor schools with less emphasis on the humanities.<sup>4</sup> Hearn and Belasco examine a variety of factors that might affect the humanities.<sup>5</sup> They show a positive association between undergraduate humanities degree production and smaller student enrollment, greater proportions of full-time students, older institutions, locations with higher per capita income, and more competitive colleges. They also show a negative association between humanities degree production and women's enrollment. These trends imply that as higher education continues to become more accessible and diverse, the humanities will continue to decline in prevalence. Moran discusses the attitudes of prominent British academics and policy makers about higher education and displays common views that may have led to the current position of the humanities.<sup>6</sup> Some of these views are that competition is essential to education, and humanities fields have too lax standards, the sciences are more important when facing modern problems such as climate change, and the humanities in schools are the root of progressive ideas and actions which are often found distasteful by more conservative people. Many of these views are again displayed by students and academics in Heller's article.<sup>7</sup> Students at top American universities are shown to view the humanities as passion projects instead of serious career options, and many students, especially first-generation students and immigrants, seem to be afraid of being viewed as less intelligent for pursuing the humanities. Heller also shows the impact that large tech companies have in drawing students towards business and science, technology, engineering, and mathematics (STEM) fields by being present on campus. This thesis extends our understanding of how many of these variables are correlated with the prevalence of the humanities by tracking their relationship across the past two decades.

The use of course catalogs to examine different factors in higher education has some precedent. Course catalogs have been used to investigate the structure of sociology departments,<sup>8</sup> as well as the content of introductory sociology courses.<sup>9</sup> They have also been used to study the presence of environmental journalism courses,<sup>10</sup> and to see trends in the pedagogical development of journalism courses.<sup>11</sup> Beyond specific fields and courses, course catalogs have also been used to track the academic mobility of faculty in undergraduate programs.<sup>12</sup> Many of these studies were very limited in scope, with only Yuret examining historical catalogs for only six universities.<sup>12</sup> This thesis expands the use of course catalogs for studying higher education, utilizing a modern optical character recognition (OCR) model to greatly speed up the analysis of the catalogs.

## 3 Methods

### 3.1 Data

#### 3.1.1 Course Catalogs:

Course catalogs were collected for the 25 institutions distributed across the United States. As shown in Figure 1 there was solid state coverage, so the results of our analysis do not just pertain to a single geographical area within the country. These catalogs were scraped from institutional websites as PDFs. An optical character recognition model was used to read the contents of each PDF into parsable text and decompose the catalogs into different sections. A data frame of courses was created, with each row having the title, description, number, number of credits, and department of the course, the start year, end year, and catalog type (undergraduate, graduate, or both) of the catalog the course came from, and the IPEDS ID and name of the institution the catalog belongs to.

This data was cleaned by removing duplicates, ensuring that each course had a title, ensuring a valid number of credits



**Figure 1.** Geographical distribution of institutions of interest.

was supplied for each course, and condensing department names by matching shortenings and slight differences in naming conventions to one common set of department names. The number of courses for each academic year after cleaning is shown in Figure 2.

A column for quinquennium was added using the start year in each row. A column called TitleDescription was created by combining the title and description columns, and then a column named TitleDescriptionCleaned was created by converting the contents of TitleDescription to lowercase and removing stop words using the nltk library.

Humanities courses were classified using the lexicon of humanities words shown in Appendix A.3. The number of lexicon words in TitleDescriptionCleaned was counted for each course, and the column is\_humanities was added, having the value true if a course had at least one humanities word and false otherwise. A column, is\_humanities\_dept, was also added classifying whether a course was in a humanities department; this was done using the list of humanities departments shown in Appendix A.4.

The percentage of humanities courses was then calculated per institution per year.

### 3.1.2 IPEDS:

IPEDS data collected during academic years 2004-05 to 2022-23 were acquired from the IPEDS website as Microsoft Access databases. These databases contained tables that provided information from the academic year before the collection year on institution characteristics, as well as statistics on student enrollment, admissions, degrees awarded, financial aid, hiring, and spending. In these tables, institutions are identified by unique six-digit numeric UnitIDs. In this analysis, two tables were used from each database: EFFY20XX, which contained statistics on student enrollment and demographic information for the entire academic year, and DRVAL20XX, which contained derived values on academic library spending and collections. The variables of interest in the EFFY20XX tables were total enrollment, total female enrollment, total enrollment by race with five categories of White, Black, Hispanic, American Indian or Native Alaskan, and Asian or Pacific Islander, and total non-United States resident enrollment. These values were examined for undergraduate students at each institution. The EFFYLEV variable defines the level of study that is described by a row. For tables with data collected during academic years 2004-05 to 2010-11, a value of 1 identifies undergraduate, and for all subsequent academic years, a value of 2 identifies undergraduate. The EFFY20XX table was present in all 19 databases. From academic years 2008-09 through 2010-11, there was a transition in how the variables in EFFY20XX were identified; the column names in each table are as follows.

Variable Names by Academic Year Collected			
Variable	2004-05 to 2007-08	2008-09 to 2010-2011	2011-12 to 2022-23
Total enrollment	FYRACE24	EFYTOTLT	EFYTOTLT
Total female enrollment	FYRACE16	EFYTOTLW	EFYTOTLW
Total white enrollment	FYRACE22	FYRACE22 and EFYWHITT	EFYWHITT
Total black enrollment	FYRACE18	FYRACE18 and EFYBKAAT	EFYBKAAT
Total Hispanic enrollment	FYRACE21	FYRACE21 and EFYHISPT	EFYHISPT
Total American Indian or native Alaskan enrollment	FYRACE19	FYRACE19 and EFYAIANT	EFYAIANT
Total Asian or pacific islander enrollment	FYRACE20	FYRACE20	
Total Asian enrollment		FYASIAT	FYASIAT
Total native Hawaiian or pacific islander enrollment		FYNHPIT	FYNHPIT
Total non resident enrollment	FYRACE17	FYRACE17 and EFYNRALT	EFYNRALT

During the transition period between academic years 2008-09 to 2010-11 institutions were given the option of which form to report data in. For the institutions examined in this research, there were no instances of data being reported in a combination of formats during this time period; this allowed for the data in these related columns to be combined into a single column for each variable. In the case of the transition from reporting Asian or Pacific islander enrollment together to reporting them separately as Asian enrollment and native Hawaiian or pacific islander enrollment, the values given for Asian enrollment and native Hawaiian or Pacific Islander enrollment were added together to find the total combined enrollment for comparison with the earlier collected data.

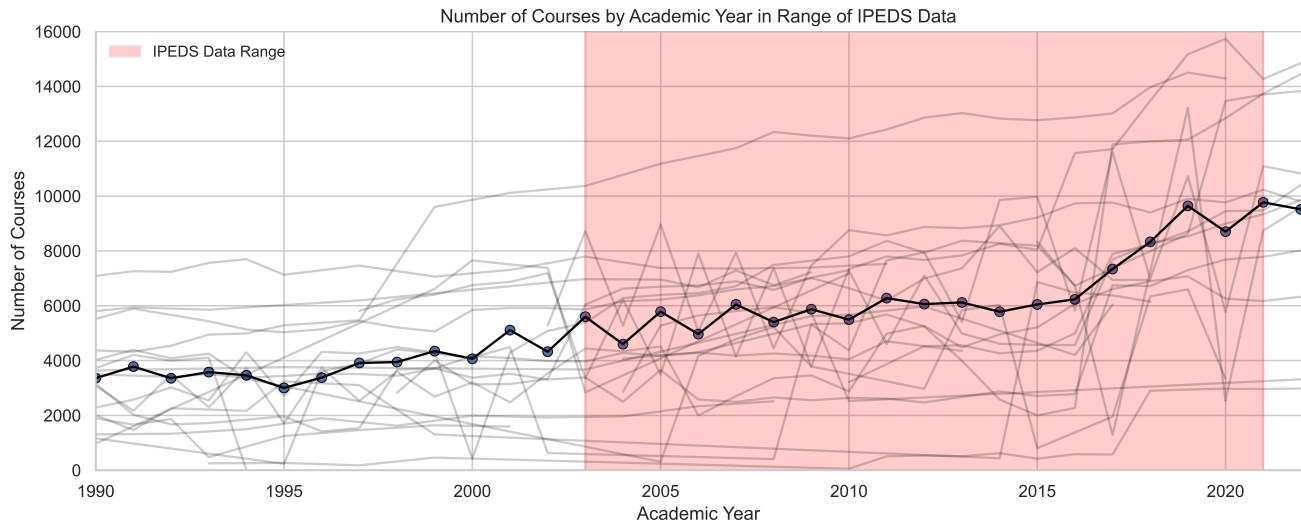
25 rows from each EFFY20XX table were selected, one for each institution of interest, combined into a data frame with 475 rows, each uniquely identified by a combination of IPEDS ID and the beginning year of the academic year the data describe. To allow for comparisons between institutions of different sizes, columns were created for the percentage of total enrollment for each demographic by dividing the enrollment of each demographic by the total enrollment. The more recent databases include a table DRVEF1220XX, which contains these derived values, but since not all years have this table, the values were calculated to ensure uniformity across years. A column nonwhite\_enrollment was created by adding the total enrollments for the demographics Black, Hispanic, American Indian or Native Alaskan, and Asian or Pacific Islander. Then, the column percent\_nonwhite\_enrollment was calculated by dividing nonwhite\_enrollment by total\_enrollment. The column inst\_name was also added by matching each the IPEDS ID of each row to its institution.

This data was merged with the course catalog data using an inner join on ipeds\_id, inst\_name, and start\_yr to create a data frame with 258 rows, each uniquely identifiable by ipeds\_id and start\_yr. The 17 columns in this data frame are ipeds\_id, inst\_name, start\_yr, end\_yr, quinquennium, total\_enrollment, percent\_humanities\_courses, and percent enrollment for each demographic.

The collection of data on academic library spending for IPEDS began in the academic year 2014-15, so the DRVAL20XX tables were only available for data collected in academic years 2014-15 to 2022-23. The variables of interest from the DRVAL20XX tables were the total academic library spending by full-time equivalent enrollment, the percent of academic library spending on one-time purchases, and the percent of academic library spending on subscriptions. These variables are identified by the names LEXPTOTF, LEXMSBBP, and LEXMSCSP, respectively.

25 rows from each EFFY20XX table were selected, one for each institution of interest, and combined into a data frame with 225 rows, each uniquely identified by a combination of IPEDS ID and the beginning year of the academic year the data describes. The column inst\_name was also added by matching each of the IPEDS IDs of each row to its institution.

This data was merged with the course catalog data using an inner join on ipeds\_id, inst\_name, and start\_yr to create a data frame with 121 rows, each uniquely identifiable by ipeds\_id and start\_yr. The 9 columns in this data frame are ipeds\_id, inst\_name, start\_yr, end\_yr, quinquennium, total\_lib\_spending\_by\_FTE, percent\_humanities\_courses, percent\_one\_time\_spending, and percent\_subscription\_spending.



**Figure 2.** Change in the number of courses in examined catalogs over time for all catalogs (undergraduate and graduate) for each institution. The IPEDS data used alongside these catalogs is available in the range of academic years shaded in red: 2003-04 to 2021-22.

### 3.2 Analytical Methods

The relationships between demographic variables and library spending with the percentage of humanities courses were analyzed using linear regression. This model was chosen to show simple trends between the variables of interest. The values used to evaluate the linear regressions were slope, correlation,  $r^2$ , and p-value. These values show the trends and the relationship between the variables, as well as the significance of these findings.

## 4 Results

### 4.1 RQ1: How has the relative prevalence of the humanities changed over time?

As shown in Figure 2 the number of courses offered in the course catalogs has risen over time: In 1990 the mean number of courses was about 3,357 and in 2021 the mean number of courses was about 9,773. This was an increase of about 191%. The Mann–Whitney U test gives a p-value of 0.00064 for this increase in courses and shows that this increase is statistically significant.

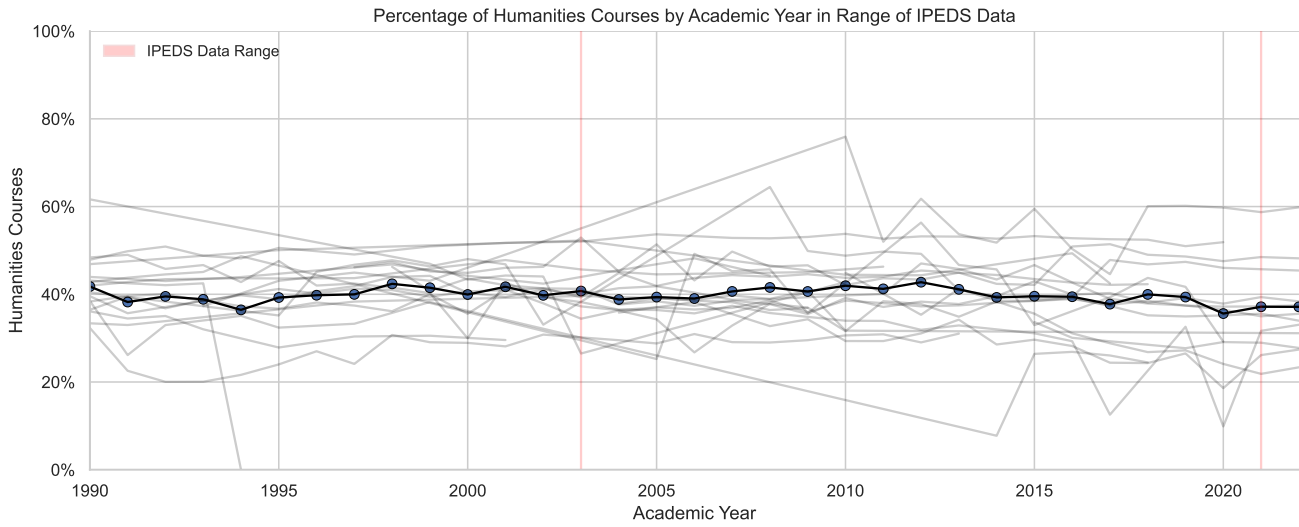
Despite this rise in courses over time, the proportion of humanities courses does not appear to change. Figure 3 shows the percentage of courses classified as humanities courses remaining stable. In 1990 the mean percentage of humanities courses was about 41.77% and in 2021 the mean percentage of humanities courses was about 37.15%. The Mann–Whitney U test gives a p-value of 0.1679 for this decrease and shows that this is not a statistically significant change.

### 4.2 RQ2: How do undergraduate student demographics relate to the trends in the humanities?

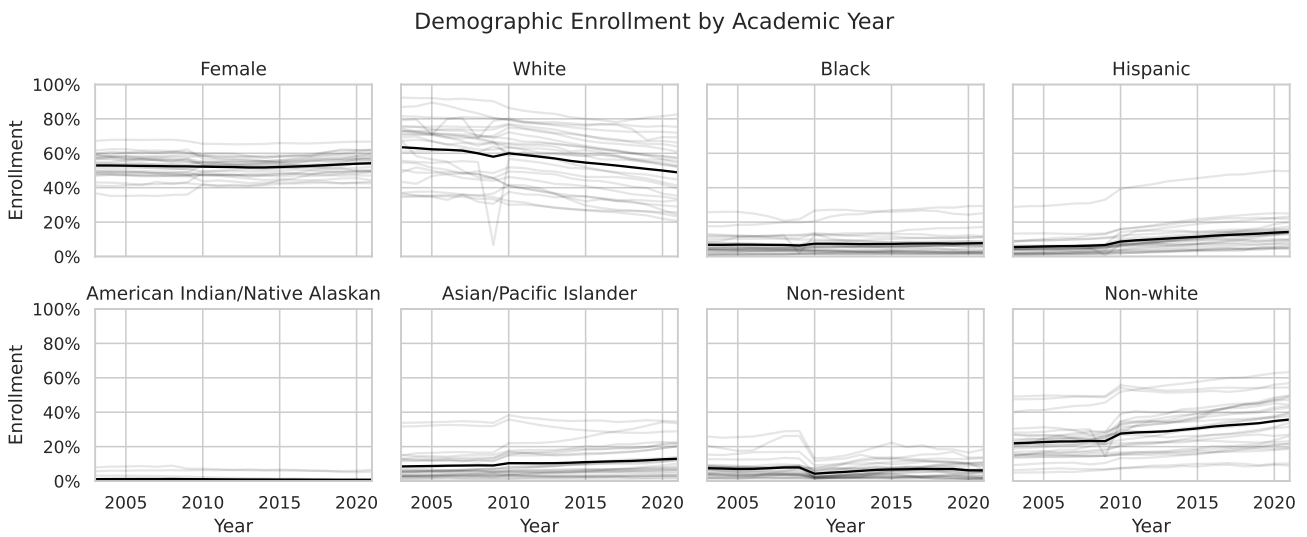
The changes in undergraduate student demographics are shown in Figure 4. Our analysis of the IPEDS data between academic years 2003-04 and 2021-22 for these demographics using the Mann–Whitney U test showed that there were statistically significant changes in the enrollment percentages for the demographics of white, non-white, Hispanic, and American Indian or Native Alaskan. The percentage of white students and the percentage of American Indian or Native Alaskan students decreased with changes of -23.05% and -35.99% respectively, and the percentage of Hispanic students and the percentage of all Non-white students increased with changes of 160.27% and 63.28% respectively.

Our analysis also showed that these demographic changes occurred while there was no statistically significant change in total enrollment within this time period. The Mann–Whitney U test produced a p-value of only 0.14561 for the change of -14.07% (28364 undergraduate students in 2003 to 24372 undergraduate students in 2021).

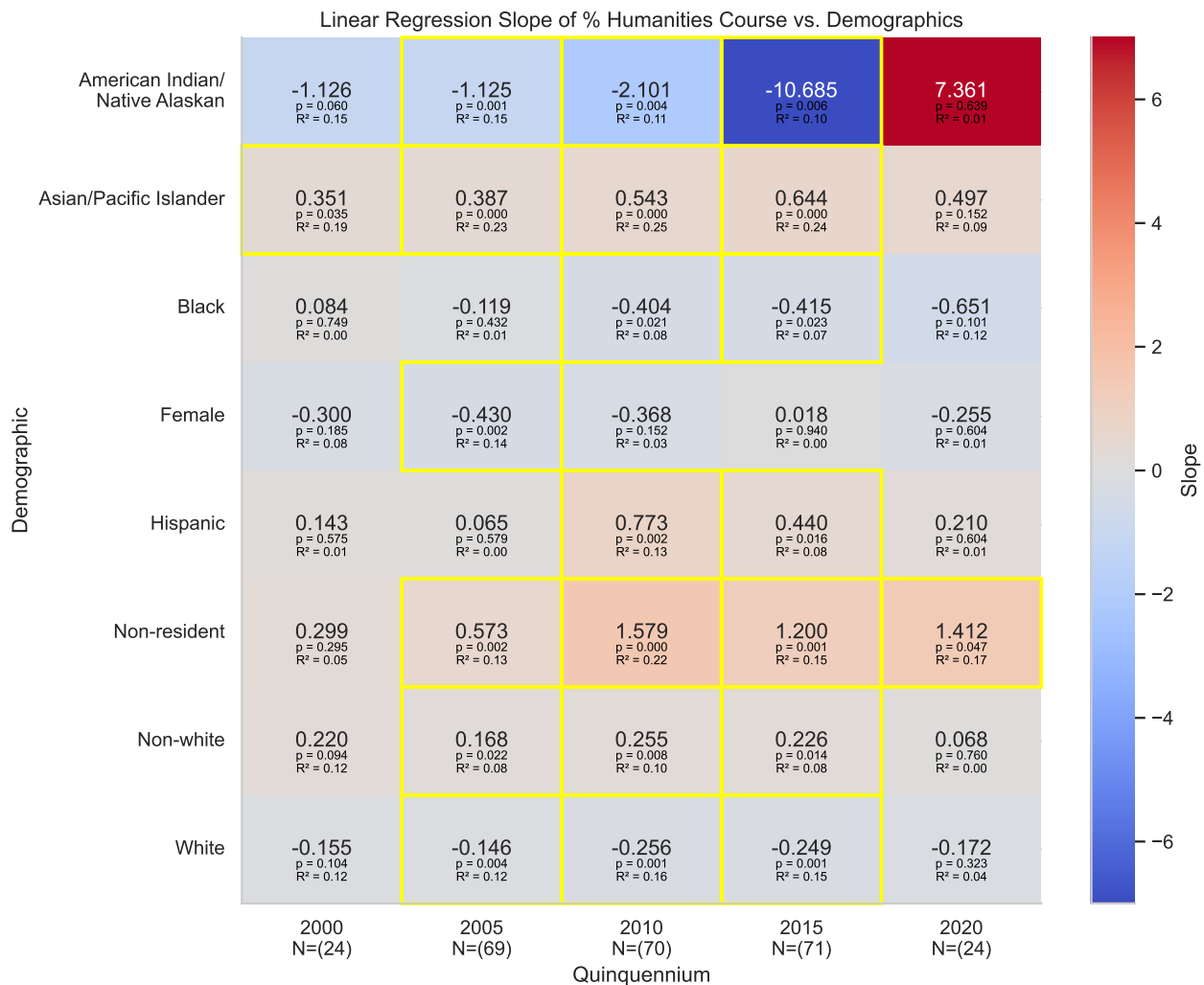
The association between these undergraduate student demographics and the percentage of humanities courses is shown in Figure 5. The demographics of Asian or Pacific Islander, Hispanic, non-resident, and non-white were found to have positive associations with the percentage of humanities courses, while the demographics of American Indian or Native Alaskan, Black, Female, and White were found to have negative associations with the percentage of humanities courses. Generally we found that these associations have become more pronounced over time for all of the demographics.



**Figure 3.** Change in the percentage of courses classified as humanities courses over time for all catalogs (undergraduate and graduate) for each institution. The IPEDS data used alongside these catalogs is available in the range of academic years shaded in red: 2003-04 to 2021-22.

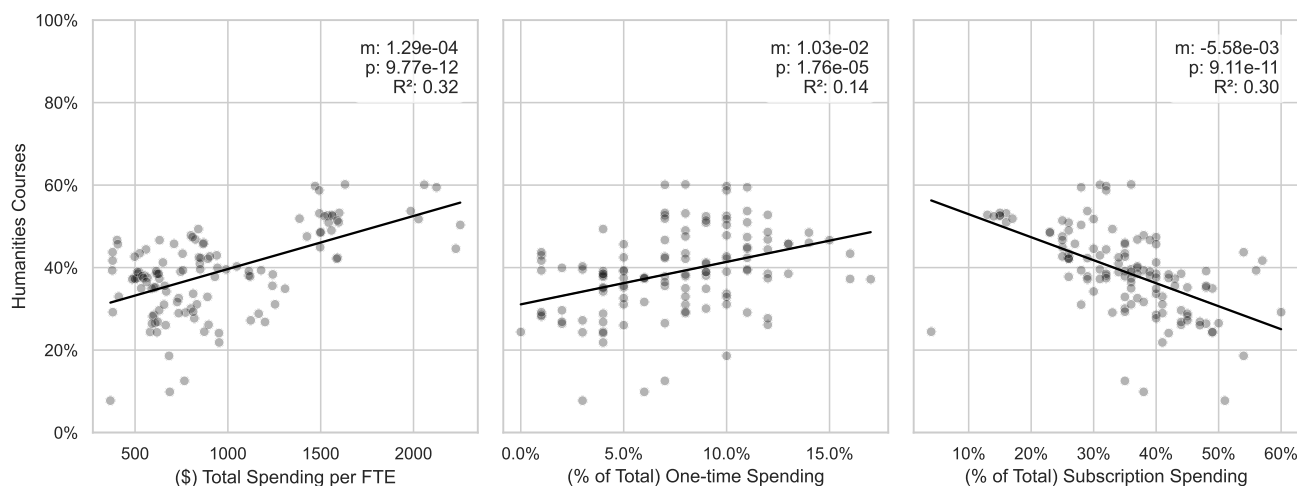


**Figure 4.** Changes in the percent of students of different demographics enrolled over time according to IPEDS data.



**Figure 5.** Heatmap showing the slopes of the linear regression of the percentage enrollment of various demographics and the percentage of humanities courses for quinquenniums 2000-2020. Positive slopes are shaded red and negative slopes are shaded blue. Cells highlighted with yellow outlines are statistically significant with p-value < 0.05. Cells are annotated with the slope, p-value, and  $r^2$  for the associated linear regression.

### Library Spending vs. Humanities Courses



**Figure 6.** Associations between percent of humanities courses classified as humanities courses and total spending per full time enrollment equivalent, percent of spending spent on one-time purchases, and percent of spending spent on subscriptions. The slope (m), p-value, and  $r^2$  for each linear regression are shown.

#### 4.3 RQ3: How does university library spending relate to the trends in the humanities?

The association between library spending statistics (total spending per full time enrollment equivalent, spending on one-time purchases, and spending on subscriptions) and the percentage of humanities courses are shown in Figure 6. We found a positive association between the percentage of humanities courses and both total spending per FTE, and spending on one-time purchases. We also found a negative association between the percentage of humanities courses and spending on subscriptions. These findings are made very credible due to their very low p-values and because of the high values of  $r^2$  for the associations with total spending per FTE and spending on subscriptions.

## 5 Discussion

### 5.1 RQ1: How has the relative prevalence of the humanities changed over time?

We did not find a decline in the prevalence of the humanities over time and instead found that humanities courses have continued to be offered at roughly the same proportion. This suggests that the humanities are not dying. This result goes against common sentiment, but is in line with the findings of Newfield that the amount of bachelors degrees in the humanities have not clearly declined.<sup>1</sup> The humanities have been central to higher education since its beginning, and according to our analysis, still continue to be today.

However, even though the percentage of courses offered in the humanities may not be declining, the perception that the humanities are dying may be coming from other sources of tension within institutions. Students and faculty working in the humanities may have this perception due to lack of institutional support, salary and funding imbalances, or any other perceived inequities in treatment or attention. These factors are not explored in this thesis, but they are important components of the discourse around the state of the humanities, and should be considered in future work.

### 5.2 RQ2: How do undergraduate student demographics relate to the trends in the humanities?

Our analysis of the trends in the demographics of undergraduate students found that the percentage of undergraduate students that are white has decreased while the percentage of non-white students has increased. This indicates that higher education is becoming more racially diverse. This trend demonstrates the importance of investigating and understanding the preferences and goals that students from these historically underrepresented groups have for their time in school. As more students from these demographics enter into higher education institutions must be able to adapt their priorities to align more with those of these students.

Our finding for the associations between undergraduate student demographics and the percentage of humanities courses contradicts the opinions portrayed in Heller's article that students of color and immigrants are more likely to feel the need to avoid the humanities because they fear being perceived as less intelligent.<sup>7</sup> We found instead that increases in the percentage of

Asian or Pacific Islander, Hispanic, non-resident, and non-white students was associated with an increase in the percentage of humanities courses offered in course catalogs.

Our analysis also showed that the associations between the demographics and the percentage of humanities courses have become more pronounced over time. This may be caused by a feedback effect where perhaps historically students of specific demographics have been drawn to specific institutions because of their preferences for subjects that those institutions prioritized, and over time more students of those demographics have been drawn to those same schools because of the existing populations of students of their demographic, in turn causing the institutions to prioritize those subjects even more greatly to cater towards the growing segment of their student bodies. This could indicate a self selection bias which may hinder institutional efforts to encourage diversity by creating disparities in the demographics of students interested in attending an institution, not only because of the existing demographic makeup of a school but also because of the subject prioritization due to that existing makeup.

### **5.3 RQ3: How does university library spending relate to the trends in the humanities?**

We found a strong positive association between total spending on academic libraries per FTE and humanities prevalence. This result is not surprising; many humanities fields are heavily focused on materials that are provided by academic libraries, so greater prevalence of humanities should certainly entail greater library spending.

Breaking down academic library spending, we found a positive association between humanities prevalence and one-time purchases and a negative association between humanities prevalence and spending on subscriptions. This result is in line with the idea that less spending on monographs shows priority shifting away from the humanities.<sup>2</sup> Humanities subjects are much more focused on works that are one-time purchases: work done in subjects such as history, literature, and philosophy are often compiled into longer monographs. Subjects such as the sciences and medicine are much more dependent on journal articles purchased through subscriptions to journals. The strong negative association between humanities prevalence and spending on subscriptions shows this: spending going towards subscriptions is much more likely to be for non-humanities subjects than humanities subjects. These results support the idea of tracking academic library spending as a way to examine institution priorities in humanities subjects.

### **5.4 Limitations**

The selection of 25 institutions used in this thesis is only a small portion of American institutions; future work should expand on the collection of course catalogs to allow for more broad examination of trends and for exploration of the relationships of catalog data to institutional characteristics such as geographic location, control (private or public), and institution age. Future work should also consider more complex methods for the classification of courses. Machine learning models could be trained to classify humanities courses with a greater accuracy than the lexicon approach used in this thesis.

Our analysis is also limited because we do not know which courses in the catalogs are being offered or how many sections of courses there are. Often course catalogs are not completely up to date with the courses that are currently being offered, and courses that have not been offered for a few years are still in the catalog. We also do not know how many students are taking each course or the demographics of the students taking each course.

Our understanding of why the humanities are perceived of as declining is limited because we do not have data on institutional support of the humanities. Our data only pertains to the amount of humanities courses at each institution, and does not have any information on the amount of funding going towards humanities subjects, or the salaries of professors working in humanities subjects compared to those working on STEM. We also do not have data on the non-monetary support that professors working in humanities subjects receive from institution administration.

## **Acknowledgments**

Thank you to Sam Zhang, Laurent Hebert-Dufresne, and Jonathan St-Onge for their previous collaboration on the course catalogs project.

## **References**

1. Newfield, C. Ending the budget wars: Funding the humanities during a crisis in higher education. *Profession* 270–284 (2009).
2. Zuckerman, H. & Ehrenberg, R. G. Recent trends in funding for the academic humanities & their implications. *Daedalus* 138, 124–146 (2009).
3. Pippins, T., Belfield, C. R. & Bailey, T. Humanities and liberal arts education across america's colleges: How much is there?. *Community Coll. Res. Center, Teach. Coll. Columbia Univ.* (2019).

4. Taylor, B. J., Cantwell, B. & Slaughter, S. Quasi-markets in us higher education: The humanities and institutional revenues. *The J. High. Educ.* **84**, 675–707 (2013).
5. Hearn, J. C. & Belasco, A. S. Commitment to the core: A longitudinal analysis of humanities degree production in four-year colleges. *The J. High. Educ.* **86**, 387–416 (2015).
6. Moran, J. The humanities and the university: a brief history of the present crisis. *Critical Q.* **64**, 5–28 (2022).
7. Heller, N. The end of the english major: Enrollment in the humanities is in free fall at colleges around the country. what happened? *The New Yorker* **27** (2023).
8. Kain, E. L. The sociology major at institutions of higher education in the united states. *Teach. Sociol.* **35**, 31–47 (2007).
9. Rowell, K. R. & This, C. Exploring the sociology curriculum at community colleges in the united states. *The Am. Sociol.* **44**, 329–340 (2013).
10. Schmidt, H. C. Training tomorrow's environmental journalists: Assessing the extent of environmental-themed training in college-level journalism programs. *Appl. Environ. Educ. & Commun.* **16**, 17–28 (2017).
11. Spillman, M., Kuban, A. J. & Smith, S. J. Words du jour: An analysis of traditional and transitional course descriptors at select j-schools. *Journalism & Mass Commun. Educ.* **72**, 198–211 (2017).
12. Yuret, T. Tenure and turnover of academics in six undergraduate programs in the united states. *Scientometrics* **116**, 101–124 (2018).
13. 89th Congress (1965-1966). National foundation on the arts and the humanities act of 1965 (1965).
14. Wikipedia. Humanities (2025).

## A Supplementary Materials

### A.1 What are the Humanities?

There is much debate about what subjects are humanities and especially about whether certain subjects should be considered as part of the humanities or as social sciences. For the purposes of this analysis, the definition given in the "National Foundation on the Arts and the Humanities Act of 1965" was followed. The act defines humanities courses as the study of "language, both modern and classic; linguistics; literature; history; jurisprudence; philosophy; archeology; the history, criticism, theory, and practice of the arts; and those aspects of the social sciences which have humanistic content and employ humanistic methods."<sup>13</sup> This definition was used to guide the selection of words for the humanities lexicon and the selection departments for this list of humanities departments. Words and departments pertaining to the fields listed were included, and words and fields pertaining to the social sciences were excluded.

### A.2 Creation of Humanities Lexicon

To get a good sampling of the many fields in the humanities and the words associated with them, the creation of the humanities lexicon began with an analysis of the Wikipedia page on humanities.<sup>14</sup> Python scripts were written to create lists of the section headers, links, and all nouns on the page. The wikipedia library was used to parse the page, and the nltk library was used to filter out only the nouns. These lists were then manually scanned, starting with the section headers, and all humanities-related words were added to the lexicon. The lexicon was then expanded by adding words from closely related subjects already in the lexicon. For example, coming from language and english which were already in the lexicon, the commonly taught foreign languages spanish, french, german, and italian were added. To ensure that the lexicon was a fair representation of the language used in university course catalogs, the lexicon was then validated through comparison with the language used in humanities course descriptions in University of Vermont course catalogs from throughout the university's history. Next, slight variations and plurals of words in the lexicon were added to minimize the effect of phrasing and specific word choice within the catalogs. Finally, an investigation of words common in courses within humanities departments was performed and the most common humanities related words were added to the lexicon. During selection of words, attention was paid as to not include words that were too broadly common to courses of all fields not just those in the humanities.

### A.3 Humanities Lexicon

acting	african	american	ancient	anthropological	anthropology	antiquity
arabic	archaeological	archaeology	art	artistic	arts	asian
asl	bible	british	buddhism	chinese	choral	christian
christianity	classical	classics	contemporary	cultural	culture	cultures
dance	design	drama	drawing	english	ethical	ethics
europa	european	feminism	feminist	fiction	film	folk
french	german	grammar	greece	greek	hebrew	hinduism
historical	historical	history	humanities	islam	italian	japanese
jazz	jewish	judaism	language	languages	latin	linguistic
linguistics	literacy	literature	literatures	medieval	moral	music
musical	myth	mythological	mythology	opera	painting	performance
philosophical	philosophy	phonetics	phonology	photography	piano	poem
poetry	poets	prose	quran	reading	readings	religion
religious	renaissance	revolution	rhetoric	roman	russian	sculpture
shakespeare	spanish	studio	theater	theatre	theatrical	themes
theology	thought	torah	translation	western	write	writers
writing	writings					

### A.4 Humanities Departments

african american studies	afro-american studies	american studies
anthropology	arab	arabic
art	art history	asian american studies
asian studies	chinese	classics
comparative literature	dance	drama
english	ethnomusicology	film
film and television	fine arts	french
german	germanic languages	greek
grmn	hebrew	history
history art	humanities	italian
japanese	japn	language
languages	latin	latn
linguistics	modern and classical languages	modern languages
music	near eastern languages	near eastern languages and cultures
near eastern studies	oriental languages	philosophy
portuguese	religion	religious studies
rhetoric	romance languages	russian
scandinavian	slavic	slavic languages and literatures
spanish	thea	theater
theatre arts	women's studies	