

Spatial and Temporal University Database*

Brendan Apfeld[†]

Department of Government, University of Texas at Austin

July 15, 2019

Abstract

This paper introduces a novel dataset of world universities. In addition to including all four-year institutions of higher education, it includes their founding dates and geographic coordinates. The introduction of this data may help to advance research agendas that consider institutions of higher education as both cause and effect.

1 Introduction

Higher education increasingly is a topic of great interest to political science. Recent research explores universities as both a dependent and an explanatory variable. See, for example, Huang (2012) on education system expansion, Rauh, Kirchner, and Kappe (2011) and Rizzo (2006) on state expenditures and higher education, or Dahlum and Knutsen (2017) on the role of higher education in human capital creation. Others have looked at how higher education can be the basis of party attempts to support and increase their base (Busemeyer 2009); how higher education has become a partisan issue (Garritzmann 2017); or the effect that university education has on democratic attitudes (Solis 2013). This list of questions only scratches the surface of a broad and deep field of research that looks at political, economic, and social causes and effects of higher education.

While there exists deep research and highly detailed data for a subset of universities, there is no extant database that provides both breadth of coverage and institution-level data. Existing research has relied instead on either micro level data for small-N studies, or aggregate measures of university education (such as average years of post-secondary education for a population) for large-N studies. I argue that a university database that is both wide and deep can help to advance agendas on these questions. At least two such databases exist already — 4ICU (Atuzzo 2019) and the World Higher Education Database (WHED) (International Association of Universities 2019) —

*This document is a DRAFT. Comments welcome. Please do not quote or cite without prior permission.

[†]brendan.apfeld@utexas.edu

but neither is available in a format that is usable for research. Ideally, this data would be available in a “tidy” tabular format with universities as the unit of observation and relevant quantities of interest as the variable columns.

This brief paper introduces such a dataset. I have collected a list of all universities in the world along with their geolocations and other variables of potential interest, including year of founding and estimated enrollment. The paper first defines the scope of the universities in this collection. The third section outlines the data sources used to create the database and the methods used to collect the data and the fourth provides an overview of some of the data provided. The final section concludes with some proposed research questions that might be answered with this data and recommendations for expanding the database in the future.

2 Universities

What is a university? In broadest terms, we think of universities as degree-granting institutions of higher education. But there are important practical questions to answer in order to measure universities: what types of degrees are necessary to qualify as a university? Are certain types of programs (such as military academies) included? I lay out here the definitional decisions I make in determining if institutions were included in this database.

First, I limit the scope of institutions of interest to those that grant at least one bachelor’s degree or its equivalent. It should be noted that I use the term “university” broadly here to mean any institute of higher education and not necessarily those that fit the American definition of “university” as opposed to “college”, where traditionally the former was defined as the subset of colleges that granted doctoral degrees. By this definition I exclude two-year universities, those that only grant associates’ degrees, as well as any technical schools that would otherwise be defined as post-secondary non-tertiary education. This translates to including only institutions that fit into the International Classification of Education (ISCED) levels 6–8 (UNESCO Institute for Statistics 2012).

Second, I remove any specialized institutions that fit any of the following categories: military, theological training centers,¹, special maritime universities, nursing schools, police academies, social

¹Institutions that include seminaries or other training for the clergy are included in the database, but those that exclusively provide this type of education are excluded.

work schools, and yoga schools. Many of these institutions are excluded because they do not grant the equivalent of a bachelor’s degree. Others do, but are so narrowly focused that they fall outside the scope of the nature of education universities provide.

3 Dataset Sources and Collection Methods

The primary list of universities and data about them comes from uniRank, a website hosted at 4icu.org that collects basic information about world universities and attempts to rank them (Atuzzo 2019). This database contains over 10,000 institutions of all types and locations², although it is created with an eye toward providing a proprietary ranking of world universities.

I collected data from the uniRank collection at 4icu.org using a scraping algorithm implemented in Python. After cleaning the data, I used the Google Maps API to convert university addresses into latitude and longitude coordinates. The coordinates are as precise as the university information available. Thus, for universities that provide a street address, the coordinates reflect this location. For those that have only a city the coordinates are at its geographic center.

The data is also compared against a second university collection, the World Higher Education Database (WHED), which is maintained by the International Association of Universities (IAU) (International Association of Universities 2019). At the time of data collection, the WHED contained 17,581 entries.³ Although the WHED provides high data quality, it employs a more generous definition of “university” without an easy or consistent mechanism for filtering out those that do not fit the definition employed above. This is particularly true in Brazil, Mexico, and the Philippines, where institutions that do not grant the equivalent of a bachelor’s are included but not easily identifiable. Outside of these countries, the two databases correlate at a very high rate, providing a basic form of validation for the uniRank enumeration.

The data included here is limited by the input received from uniRank, which in turn is limited by its data collection strategy. uniRank uses a combination of automatic and manual methods for collecting and updating data on universities. In many cases, it relies on members of the public submitting requests for updates or amendments to university entries. In general terms, the data is strongest for well-established institutions and weaker for newer ones. A preliminary validation

²At the time the data was gathered, the website listed 10,182 universities.

³The WHED now claims to contain information on over 19,000 institutions of higher education (International Association of Universities 2019)

suggests that the data of most interest — founding dates and physical locations — are also reliable, but that other quantities such as enrollment and programs offered are subject to greater variation in quality and higher levels of missingness. Finally, recently opened universities are unlikely to be included in his data.

4 World Universities

One of the contributions of this data is geocordinate information about the locations of universities. Patterns of university locations conform to expectations (wealthier and larger countries have more universities and universities tend to cluster in densely populated areas), but knowing the precise coordinates opens the possibility of leveraging this information to provide more precise causal effects. Figure 1 shows the locations of universities around the world by control type (public or private).

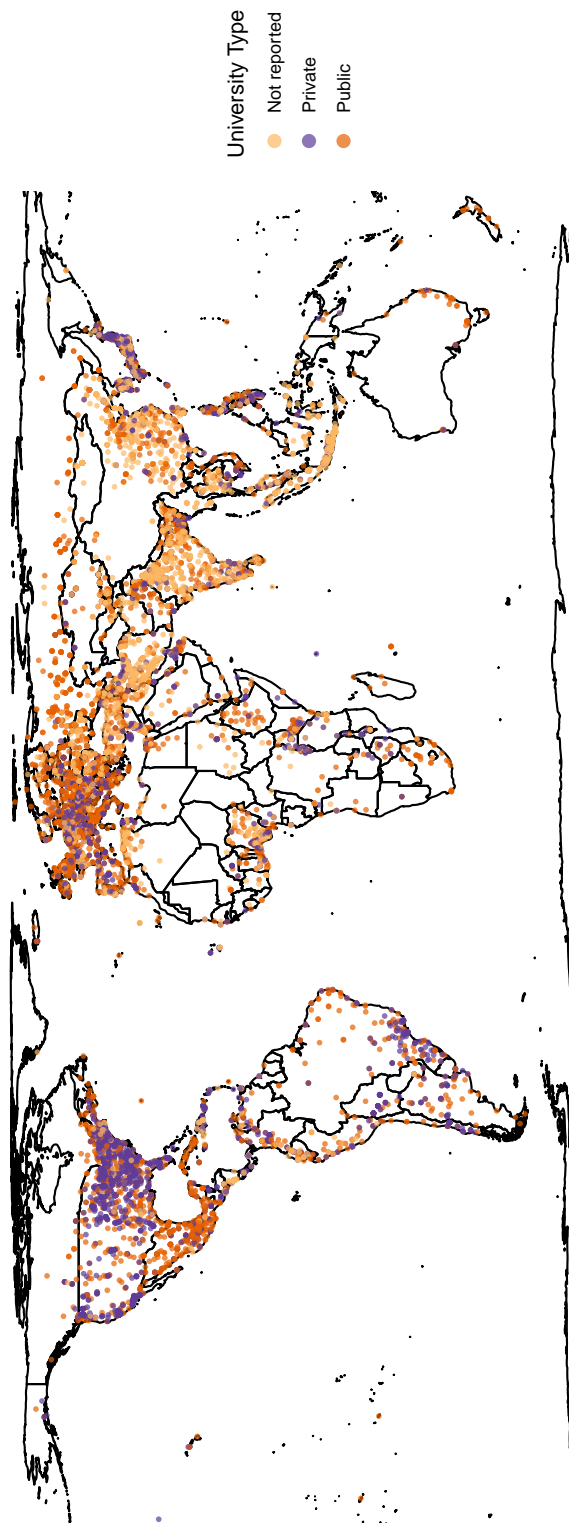


Figure 1: Current listing of universities around the world by control type.

The second contribution of this database is the inclusion of university founding dates. As with location, information on temporal development of institutions for higher education opens the possibility of leveraging this information for new or more precise explanations. Figure 2 shows universities in Europe by founding date. Although the graph shows only Europe with founding dates grouped by century, the underlying data includes year-level precision for the entire world.

Note that although Europe is home to most of the most ancient universities and experienced the first wave of university expansion, it has also experienced lots of recent growth. This mirrors the explosion of universities throughout the rest of the world, particularly the developing world.

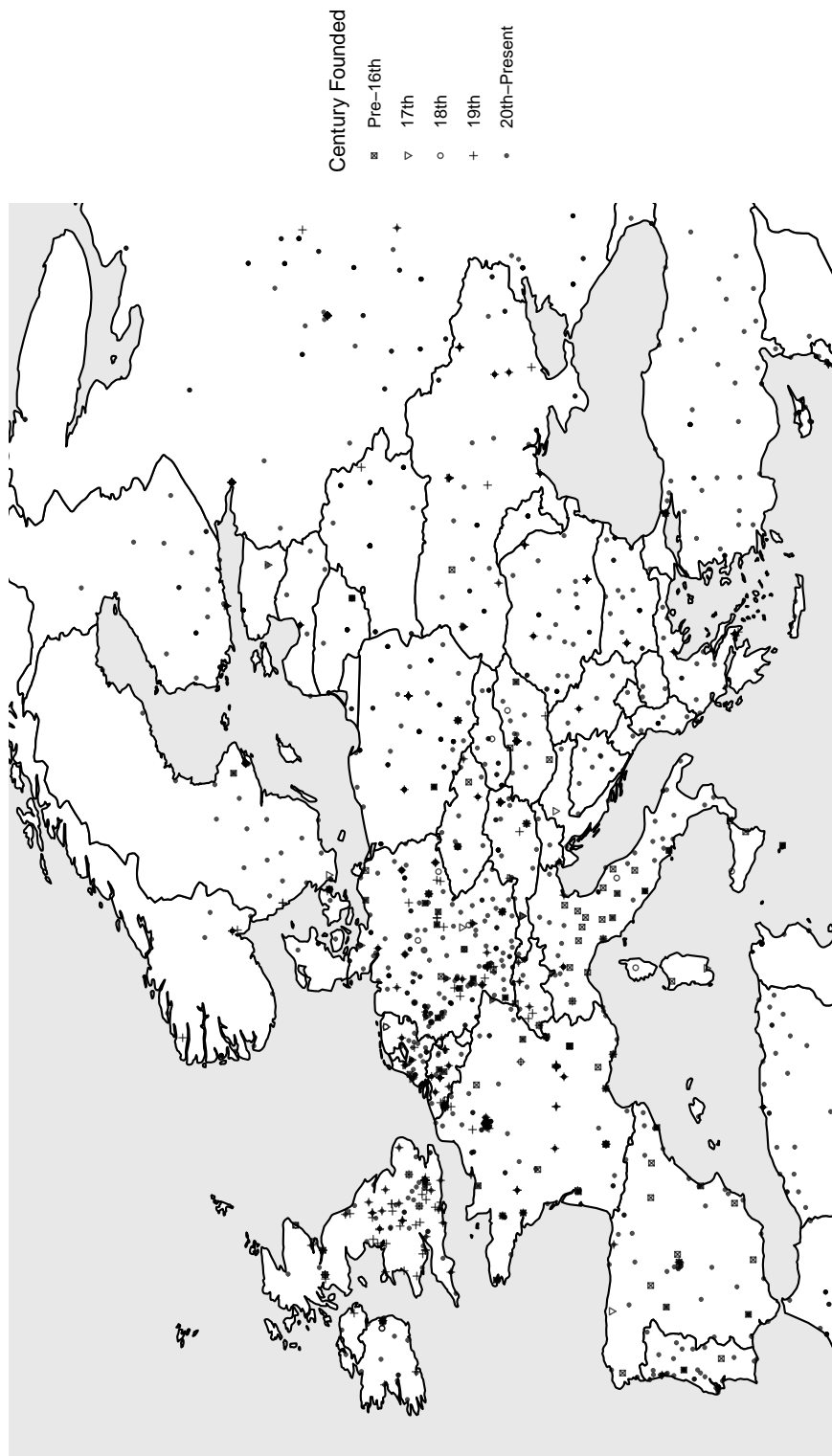


Figure 2: European universities by century founded.

Indeed, the growth of universities over time shows exponential growth. Figure 3 shows this growth since the founding of the University of Pisa in 1088. The number of new universities has slowed in the previous few years, though this may simply reflect a delay between university foundings and their appearance in this database.⁴

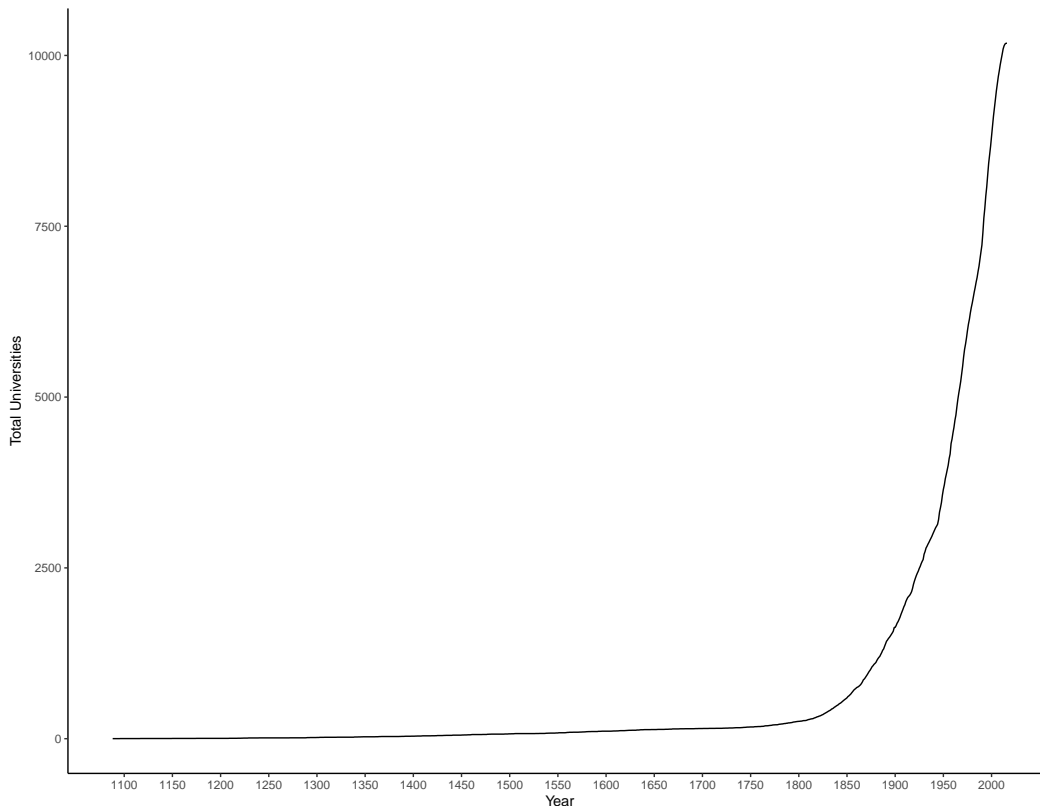


Figure 3: Total universities in the world, 1088–2016

5 Conclusion

In introducing this dataset, I hope to provide a resource to researchers looking into universities as both cause and effect. The inclusion of temporal data about universities (e.g. founding date) and spatial data (geocoordinates) may facilitate more sophisticated answers to important political, economic, and social questions. It holds the potential to contribute to both macro explanations of the relationship between universities and development, technology, and state power as well as causal explanations that rely on more advanced identification strategies.

One branch of questions this data may help to answer treats universities as a causal of some

⁴Growth has, in fact, nearly flat-lined according to the data, though it is difficult to see in this visualization.

political phenomenon. For example, what effect do universities have on political, economic, and social attitudes? Knowing the exact locations of universities may allow researchers to provide causal explanations for the diffusion of beliefs. Similarly, what effect have universities have on economic development or the expansion of the liberal state?

A second branch of questions this data may help answer treats universities as effect of some political process. For example, what are the driving forces that create universities? Previous research has traced the expansion of the university system around the world through several distinct periods. Sequencing of university expansion, economic growth, and population shifts across regions or the world is now possible with the addition of university founding dates and locations.

Although this database provides new access to important numbers, there are several ways in which it could be improved. First, university locations where only a city is currently available could be expanded to the street address (and thus latitude and longitude). Doing so would require a very high level of human input, however. Second, the WHED data could be used to supplement the uniRank data, particularly for enrollment figures. Obstacles to this improvement include seeking permissions from the International Association of Universities to use the data in this way, collecting the updated figures from a redesigned website, and finding a reliable way to merge data on university names. Finally, the data could be updated to reflect any new universities opened in the previous five years that are not currently included.

References

- Atuzzo, Fabio (2019). “uniRank”. URL: <https://www.4icu.org>.
- Busemeyer, Marius (Jan. 2009). “Social Democrats and the New Partisan Politics of Public Investment in Education”. *Journal of European Public Policy* 16.1, pp. 107–126.
- Dahlum, Sirianne and Carl Henrik Knutsen (2017). “Do Democracies Provide Better Education? Revisiting the Democracy-Human Capital Link”. *World Development*.
- Garritzmann, Julian L. (Apr. 2017). “The Partisan Politics of Higher Education”. *PS: Political Science and Politics* 50.2, pp. 413–417.
- Huang, Fali (2012). “Why Did Universities Precede Primary Schools? A Political Economy Model of Educational Change”. *Economic Inquiry* 50.2, pp. 418–434.
- International Association of Universities (2019). “World Higher Education Database”. URL: <https://www.whed.net>.
- Rauh, Christian, Antje Kirchner, and Roland Kappe (2011). “Political Parties and Higher Education Spending: Who Favours Redistribution?” *West European Politics* 34.6, pp. 1185–1206. URL: <https://doi.org/10.1080/01402382.2011.616659>.
- Rizzo, Michael J. (2006). “State Preferences for Higher Education Spending: A Panel Data Analysis, 1977-2001”. In: *What’s Happening to Public Higher Education?* Ed. by Ronald G. Ehrenberg. Series on Higher Education. Westport, CT: Praeger.
- Solis, Alex (2013). *Does Higher Education Cause Political Participation?: Evidence From a Regression Discontinuity Design*. 13. Uppsala Universitet.
- UNESCO Institute for Statistics (2012). *International standard classification of education: ISCED 2011*. Montreal: UNESCO Institute for Statistics.