Sanaa Rafique

March 15, 2019

Data Management and Research Cycle – Spring 2019

**Final Project Proposal**

Research Question

For my final project, I will explore the relationship (if any) between ***total enrollment proportions of four main races*** (White, African American, Asian, and Hispanic) in 4-year public and private colleges in Texas and ***poverty rates of the zip codes of these colleges***.

A 2016 research study[[1]](#footnote-1) found that local colleges vary along lines of race and class. In particular, it found that communities with large Hispanic populations and low educational attainment had the fewest alternatives nearby, while White and Asian communities tended to have more. These can result in education deserts, or places where there are abundant opportunities for some communities and are rare in others.

This is an interesting topic from a policy perspective since it raises concerns about some populations, when compared to others, not having equal access to quality education because they live in a low-income and high poverty region. Poor access to educational opportunities nearby has implications for these communities in terms of income levels and standard of living. By expanding on this research topic, policymakers can better understand a need for injecting resources in a certain region and increasing educational access for certain communities and break the cycle of poverty and unequal opportunities for the same populations.

Research Design

For this project, I will harness the power of both, Python and ArcGIS to answer my research question. In ArcGIS, I will map the changes in enrollment proportions for each of the four main races from 2012-2017 for each college against the changes in poverty rates of the zip codes of these colleges from 2012-2017. In Python, I will build a correlation matrix and plot this matrix to better understand and visualize the correlations between different variables.

Data Sources

* U.S. Census, American Community Survey 2013-2017 American Community Survey 5-Year Estimates
* Integrated Postsecondary Education Data System (IPEDS) Survey 2012-2017

Validation Methods

There cannot be a validation technique in place to check for poverty rates of zip codes. The American Community Survey because of its large sample size is most preferred for subnational (including zip code-level) data on income and poverty by detailed demographic characteristics.

A validation check that might be put in place to authenticate enrollment proportions in Texas 4-year public and private colleges for each of the four main races between 2012 and 2017 is for an institution itself to create an account on National Student Clearinghouse’s website. The organization runs a program, StudentTracker which is the only nationwide source of college enrollment and degree data. More than 3,600 colleges and universities — enrolling over 99 percent of all students in public and private U.S. institutions — regularly provide enrollment and graduation data to the Clearinghouse. This information can be accessed to conduct different types of research analyses for educational purposes.

1. Hillman, Nicholas W. "Geography of college opportunity: The case of education deserts." *American Educational Research Journal* 53, no. 4 (2016): 987-1021. [↑](#footnote-ref-1)