Julie Carlson
INFO 628: Data Librarianship & Management
Methododology Statement

Since 1966, the National Park Service has maintained the National Register of Historic Places, "the official list of the Nation's historic places worthy of preservation" (National Park Service, 2021). As of June 2021, this list contains more than 96,000 properties—data about these places can be found on the National Park Service's website. Although the data is freely available, there is a dearth of comprehensive analysis that examines the characteristics of listed properties. Where are they primarily located? What is the predominant category of the properties (e.g. building, district, site)? What "areas of significance" are the most represented? Using the National Park Services' dataset as my raw data, I will explore the National Register through quantitative analysis to consider what kind of places are deemed worthy of preservation.

The National Park Services' dataset is particularly suitable for quantitative analysis because of its large size. The dataset contains 96,644 rows and 21 fields; quantitative methods will assist me in "identifying broad patterns" from this vast collection of data (Lemercier & Zalc, 2019). After cleaning the dataset in OpenRefine, I will use Python scripts to perform statistical analyses. This work will result in .csv files, which I will use to visualize my findings with Tableau Public. Visualizing my results will help me explain my findings, but could also illuminate aspects of the data. As Lemercier and Zalc (2019) note, visualizations "give information on aggregate patterns […] and detect exceptions and 'outliers.'" Perhaps we'll learn that a certain state has far fewer properties than others have listed.

I also plan to perform quantitative text analysis using Voyant Tools to explore the National Register through a different lens. Certain fields in the dataset have a fixed set of values. For example, the "Category of Property" is only described as "Building," "District," "Object,"

"Site," or "Structure." While statistical analysis will allow me to count the instances of these particular property types, nuance about the properties is lost; a house, a school, and a post office are all reduced to a "building." The "Property Name" field contains more descriptive information that could be valuable to analyze, so I will use Voyant Tools and its built-in visualizations to see which terms appear most frequently there. Analyzing text in Voyant Tools has helped scholars "frame important terms through count, frequency, and relativity that ultimately gave way to new areas of desired inquiry" (Miller, 2018). Given that there has yet to be a thorough overview of the National Register of Historic Places, I hope that my methodology will both provide insight about the places and suggest future areas of inquiry.

References

Lemercier, C., & Zalc, C. (2019). *Quantitative Methods in the Humanities: An Introduction.* University of Virginia Press.

Miller, A. (2018). Text Mining Digital Humanities Projects: Assessing Content Analysis Capabilities of Voyant Tools. *Journal of Web Librarianship*, *12*(3), 169-197. https://doi-org.ezproxy.pratt.edu/10.1080/19322909.2018.1479673

National Park Service (2021, August 9). *National Register of Historic Places: Data Downloads.* https://www.nps.gov/subjects/nationalregister/data-downloads.htm