Julie Carlson
INFO 628: Data Librarianship & Management
Final Project Report
December 20, 2021

Inspired by my encounters with landmarks across the nation, I chose to analyze listings from the National Register of Historic Places in order to explore what kinds of properties are listed, and what these properties tell us about our country's history. In my initial research, I found numerous articles about the criteria for listing properties on the National Register (see Maskey et al., 2009; Heath, 2019; Sprinkle, 2007), but nothing that applied data analysis methods to the entirety of the listings. Aiming to fill this gap, I downloaded the complete dataset of listed properties through June 17, 2021, and used quantitative methodology to explore them. Through my analysis, I have seen that the National Register disproportionately represents properties from a particular geographic region and properties associated with themes that are perhaps not representative of the country as a whole.

The National Register is especially well-suited for quantitative analysis because of the sheer volume of properties listed. The National Park Service has maintained the National Register since 1966, and in these 55 years, more than 96,000 properties have been listed. As Lemercier & Zalc note, quantitative methods aid in "identifying broad patterns" from large datasets; with 96,644 rows and 21 columns, the National Register is certainly vast enough to lend itself to such analysis (2019). In order to identify patterns, I first cleaned the raw dataset in OpenRefine. As detailed in my analysis statement, this involved standardizing values from the "Category of Property" field to make them uniformly capitalized, adding a column to show only the year listed (derived from the "Listed Date" field), removing seven empty columns, and deleting a row that contained only a property name with little other useful information. From there, I saved the processed dataset as a CSV file, which I ran through a Python script that I

wrote. Using the script, I was able to count the number of properties in each state, category, area of significance, and year listed. The script wrote the counts to four CSV files, which I used to create visualizations that aided in my analysis.

I used the Python-derived CSV files and Tableau Public to create four separate visualizations that show the count of properties by state, by category, by area of significance, and by year listed. A map was a natural fit for the count by state information; in my Tableau visualization, a state's color corresponds to its count of listings, with darker states having more listed properties. Through this visualization, the most-represented states on the National Register become apparent: New York has 6,276 listed properties, Massachusetts has 4,419, Ohio has 4,122, and Pennsylvania has 3,512. All told, these four states have 18,329 properties, 18.97% of the 96,642 total properties listed. Given that these states are contiguous, the visualization reveals how properties on the National Register are heavily clustered in the same geographic region.

To illustrate the count of properties by category of property, I created a tree map in Tableau Public. The categories defined by the National Register are quite broad. Properties can be categorized as a building, district, site, structure, or object. Perhaps unsurprisingly, the vast majority are buildings; 66,712 listings (69% of the total listings) are categorized as such.

In order to obtain more insight into these buildings, I analyzed their associated areas of significance from the dataset. For the areas of significance, I created a packed bubble chart where the size, color, and location of the bubbles in the visualization indicate their respective counts. From this graphic, it becomes clear that most properties are designated as significant because of their architectural value. A single property can be listed with multiple areas of significance, but the "Architecture" designation has been applied to 71,758 properties, 74.25% of the total dataset. Other highly-represented areas include "Commerce" (12,892, 13.34%),

"Politics/Government" (8,385, 8.67%), "Industry" (7,034, 7.28%) and "Exploration/Settlement" (6,093, 6.3%). These figures and the accompanying visualization seem to reflect the National Register's focus on locations associated with capitalism, commercialism, and colonization. It is striking to compare these numbers to those of listings deemed significant for their association with racial and ethnic groups that have historically been considered minorities in the United States. There are 2,163 properties associated with the area of significance simply deemed "Black," 1,201 marked as "Historic - Aboriginal," 1,047 as "Native American," 210 as "Ethnic Heritage," 151 as "Hispanic," 100 as "Asian," 88 as "Ethnic Heritage-Black," 55 as "Pacific-Islander" 52 as "Other-Ethnic," 14 as "Ethnic Heritage-Native American," 7 as "Archaeology-Historic Aboriginal," 4 as "Ethnic Heritage-Asian," 4 as "Ethnic Heritage-Other-Ethnic," 2 as "American Indian," 1 as "Ethnic Heritage-Hispanic," and 1 as "Ethnic Heritage-Pacific Islander." In sum, these listings cover 5,100 properties or 5.28% of the total listings, a smaller percentage put together than "Exploration/Settlement" alone.

To reveal more descriptive information about the listed properties, I processed the data from the dataset's "Property Name" column in Voyant Tools. The tool analyzed the frequency with which words appeared in the text, producing a count for each word (excluding stopwords, which were both automatically detected by Voyant and input by me). Voyant also produced an accompanying word count visualization where a word's size corresponds to its number of appearances in the corpus. As anticipated given the category information, most of the large words are types of buildings; house, school, and church are particularly prominent. It is worth noting that, like "Church," many of the terms in the word cloud are associated with Christianity (Presbyterian, Baptist, Episcopal, Methodist, Catholic, Chapel, Congregational), while only one term (Temple) is commonly associated with non-Christian religions. Additionally, it is interesting

to see that the word cloud includes a number of typically male first names (Robert, Joseph, Charles, Thomas, Samuel, James, George, John, Henry, William) and only one typically female name (Mary). This imbalance likely reflects the fact that the many historic houses in the National Register are more commonly associated with the men who lived there rather than the women.

Given the listing characteristics that emerged from this analysis, I was curious to see trends about when properties were added to the National Register. Returning to Tableau Public, I created a line chart visualization showing that—with the notable exception of 1981—years in the 1980s had a relatively high number of listings added. From 1980-1989, there were 32,222 properties listed, 33.34% of the total dataset. In contrast, from 2010-2019, only 10,352 properties were listed, 10.71% of the total dataset. It could be argued that the number of historical properties remaining to be listed has simply diminished over time if many are already on the National Register. However, the evaluation criteria for potential listings stipulates that a property must generally be at least 50 years old, so theoretically new properties would become eligible every year (National Park Service, 2021). With more time and resources, further research should explore potential legislative and cultural reasons why the 1980s were a particularly active time for the National Register. This may shed light on the trends in the listings themselves, and give insight into the relatively low number of listings in recent years.

In fact, my project leaves much room for further analysis and research. Through quantitative analysis, we have seen that the National Register's listings emphasize commerce, religion, and prominent male figures. My hope is that identifying these trends will lead to more critical inquiries about the National Register and its representation of American history.

Bibliography

Heath, K.W. (2019). Toward a Humanist Approach to Historic Preservation. *Journal of American Folklore 132*(526), 390-411. https://www.muse.jhu.edu/article/734364.

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

Maskey, V., Brown, C., & Lin, G. (2009). Assessing Factors Associated With Listing a Historic Resource in the National Register of Historic Places. Economic Development Quarterly, 23(4), 342–350. https://doi.org/10.1177/0891242409342571.

National Park Service (2021). *How to List a Property.* https://www.nps.gov/subjects/nationalregister/how-to-list-a-property.htm.

National Park Service (2021). *National Register of Historic Places Listed properties (listings up to 06/17/2021)* [Data set]. https://www.nps.gov/subjects/nationalregister/data-downloads.htm.

Sprinkle, J. H. (2007). "Of Exceptional Importance": The Origins of the "Fifty-Year Rule" in Historic Preservation. *The Public Historian*, *29*(2), 81–103. https://doi.org/10.1525/tph.2007.29.2.81.