

Given the vast amount of data contained in the National Register of Historic Places dataset, there were a number of questions that could be explored: What state has the most buildings listed? What category of building is most represented? What areas of significance are most commonly referenced? What terms appear most frequently in the properties' given names? Quantitative methods help answer these questions, providing insight into the types of places that the National Register deems worthy of preservation.

### **Quantitative Analysis with Python**

The raw dataset was first cleaned in OpenRefine. This involved standardizing the values in the "Category of Property" field to make them uniformly capitalized, adding a column to show only the year of listing (derived from the "Listed Date" field), removing empty columns, and deleting a row that contained only a property name with no other relevant information. The clean dataset was saved as a CSV file. A Python script was used to count the number of properties in each state, category, and area of significance. The script wrote the counts to three CSV files.

### **Data Visualization in Tableau Public**

The CSV files produced by the Python script were imported into Tableau Public to visualize the data. The count by state was illustrated through a map visualization where a state's color corresponds to its count. The count by category was visualized as a treemap, and count by

area of significance as a packed bubble chart where the size and color of the bubbles indicate the count. The visualizations were made freely available on Tableau Public.

### **Text Analysis in Voyant**

Returning to the cleaned CSV file from OpenRefine, the data from the “Property Name” column was copied into Voyant Tools. The tool analyzed the frequency with which words appeared in the text, producing a count for each word (excluding automatically-detected stopwords). It also produced a word cloud visualization where each word’s size corresponds to its number of appearances in the corpus. The word cloud was saved as a PNG file.

### **Next Steps**

The Python script will be augmented to retrieve the count of listings by year in which they were listed, to explore potential temporal trends in the National Register. The resulting CSV file will be visualized in Tableau Public. All Tableau visualizations will be further refined to effectively communicate the findings. The word cloud from Voyant Tools may also be refined to remove more stopwords, such as “la” and “san.”