Introduction

The Spanish Artists Dictionary (SAD) is a reference source created by scholars at the Frick Art Reference Library. Originally a print publication, the dictionary was formatted as a Filemaker database in the early 1990s and made available through the Frick’s online research portal. The database consists of 5,186 records describing Spanish artists; these records include information about artist name, alternate names, dates (birth, death, and/or activity), field of artistic endeavor, bibliographic references, and Frick Photoarchive holdings.

The data from SAD’s three Filemaker tables has been made available for this project as CSV files.

Visualizations

Artists and Exhibition Dates:

These visualizations came from a script, which parsed the original CSV data and looked through the “activity date qualifier” column (labeled OQN). It got rid of a square bracket character using a regular expression as a way to normalize the data. Then, it searched for rows in which the “activity date qualifier” was “exh.” and wrote a new CSV containing the artist’s name and the data from the activity year(s) column (OPRN). Finally, new rows were manually created (each year of activity given its own row) so that Tableau would properly understand the data as date ranges.

In the original CSV data, there were five uncertain dates (qualified with a question mark) that we excluded from the data used for the visualizations.

Visualization #1: Displays all the artists whose activity dates were qualified with the term “exh." (exhibited). It shows the number of years an artist exhibited, organized in descending order. It is also interactive and can be searched by artist name and exhibition date.

Visualization #2: This visualization plots the exhibition years on a line graph. There is a noticeable spike in the number of artists exhibiting in 1944 and 1951. The bulk of the artists in the dataset exhibited around 1900.

Visualization #3: This visualization counts the number of years an artist exhibited. It clusters the artists by size and color. The darker the bubble, the more years an artist exhibited. Hover over the bubbles to find out how many years an artist exhibited.

Gender of Artists in SAD:

To create the visualization, we wrote a script which parsed through the original CSV file, pulled out the columns that included the artists name and gender, and then wrote this data out into a new CSV. Afterwards, the CSV with the artists name and gender was imported into Tableau.

Name Matching

We also wrote a script to address a problem with linking from SAD records to the library catalog. From a SAD record, when a user clicks on one of the three links to a library catalog—the Frick’s library catalog (FRESCO), Arcade (the NYARC catalog), or Worldcat)—the search box that appears in the catalog is populated from a field in the Filemaker database labeled “LC Name.” However, many entries in SAD have incorrect or outdated names in the “LC Name” field, or they do not include a name in the “LC Name” field, in which case the name is populated from the main name field. This produces faulty results (i.e. Goya, Francisco, 1746-1828 not Goya, Francisco de, 1746-1828) (1181 vs. 2 hits in FRESCO).

To address this problem, the SAD\_FrescoMatch.py script compares the main name entry in the SAD database with a list of Spanish names pulled from the library catalog’s subject heading list. A portion of the CSV version of the database (comprised of the 374 female artist names in SAD) was parsed using the CSV DictReader method. Each row of the CSV became a dictionary object. The token\_set\_ratio method was then used from FuzzyWuzzy, a string matching library for Python. The method breaks two strings (in this case artists’ names) into words and compares the degree of similarity between the two strings. A match was considered those results that achieved a .token\_set\_ratio above 90. 49 matches were found, including two false matches that probably resulted because of the presence of substrings (Romero López, José María, 1815-1880. matched falsely with López, María and Rodríguez de Losada, José María, 1826-1896 matched falsely with Rodríguez, María. These matches were then inserted back into the Python dictionary rows to populate a new field called FRESCOSH. The updated data was written as a CSV file, which can now be uploaded back to the FileMaker database.