In Fall 2017, a group of four librarians embarked on a project to learn more about GW library’s rare book holdings using MARC data. The goals for this project were to gain a more complete understanding of the strengths, coverage, and gaps of the rare book collection and to strengthen our nascent data analysis skills. We plan to use the results from our data analysis to guide collection development and instruction.

How many rare books do we have?
What is the range of publication dates?
Where are the places of publication?
Who were the printers and/or publishers?
What languages are represented?
What are the subjects covered?

Using a dataset that captured specific MARC fields within the bibliographic records that held the potential answers to our research questions we used various methods of data clean-up and organization to analyze and visualize the data.

73,000+ records in initial data extractions

Issues encountered:
• Difficult to capture relationships between fields and subfields
• Inconsistent cataloging practices
• Duplicate MARC fields

We were able to locate titles that remain from the original library collection of the Columbian College. We also located University Archives that were cataloged as books.

A page from the original inventory of the Columbian College library’s collection, written in 1825.

The note in the 500 field indicating books listed in the above 1825 inventory.

Next Steps
• Explore data in 5xx and 7xx fields
• Use data to support strategic collection development
• Disseminate data to support discovery and collections as data research