Analyzing the Historic Maryland Newspapers Project Collection

Pamela A. McClanahan | University of Maryland Libraries

SAA Research Forum | August 5, 2020

Project Background

The Historic Maryland Newspapers Project (HMNP) at University of Maryland Libraries is the Maryland state awardee of the National Digital Newspaper Program (NDNP), a partnership between Library of Congress (LC) and National Endowment for the Humanities (NEH). This program has created the Chronicling America online database with historic newspapers from nearly every state, which is freely available on the LC website. Awardees digitize about 100,000 newspaper pages per grant ground that continue to be added to Chronicling America (https://chroniclingamerica.loc.gov/), which has over 16 million pages so far.

HMNP has been participating since 2012 and has digitized 316,168 newspaper pages. HMNP is currently in the midst of its 4th round of NDNP, which is set to end in Summer 2020 resulting in the addition of approximately 110,000 more pages. If awarded a 5th grant round, HMNP would be making final title decisions in Fall 2020.

The HMNP advisory board has prioritized titles from communities traditionally underrepresented in archival records including immigrant communities, women, political minority groups, and labor groups. Newspapers are often one of the only primary sources for these communities' activities as they are not always otherwise well-documented in the historical record.

Research Statement

Goal - To better understand the Maryland newspaper collection in Chronicling America to make future digitization selection decisions prioritizing diversity and to share this information with researchers and archives, libraries, and other cultural heritage institutions throughout the state.

Objective - To gather and evaluate data from the Chronicling America Maryland newspaper collection comparing the number of titles, issues, and pages with such factors as geographic and temporal distribution, and distribution of newspapers from underrepresented communities.

Methodology

Data Collection - Title data was gathered from Chronicling America, US Newspaper Directory, and HMNP's own metadata collation of titles and input into an Excel spreadsheet. Fields collected include Title, Date Range, City of Publication, County, Pages, Issues, and Language. Additional fields are Region and Underrepresented groups. These fields were manually determined and added. Underrepresented titles for the purposes of this project include non-english language, immigrant communities, political minorities, labor groups, and women.

Analysis - Excel pivot tables were used to sum data including number of titles, number of issues, and number of pages by county and by region. Sorting and filters in Excel allowed for quick display of this information. Tableau was used to create map based data visualizations. Using Federal FIPS county codes and ESRI US county shape data along with the Excel pivot tables in Tableau, a map was created of the different sum factors by county with a density measure (lighter color = lower number and darker color = higher number). Four maps were created including Sum of Titles by County, Sum of Issues by County, Sum of Pages by County, and Sum of Underrepresented Group Titles by County. The maps are interactive and allow you to hover over a polygon in the map and it will display the county name and the sum factor. Regional sums were not used at this time, but are easily visualized when seeing where a county is on the map and the density colors for the surrounding area.

Preliminary Results
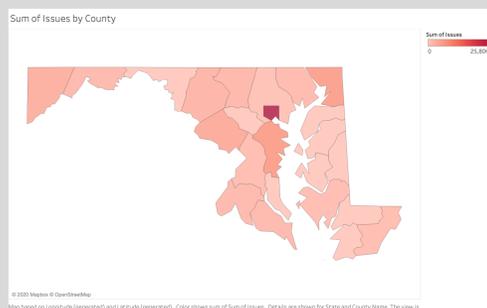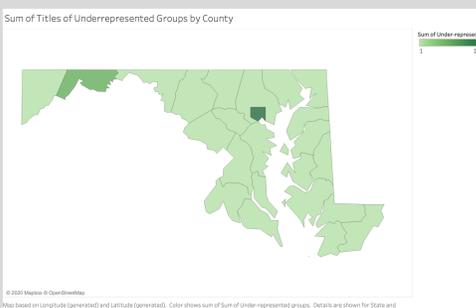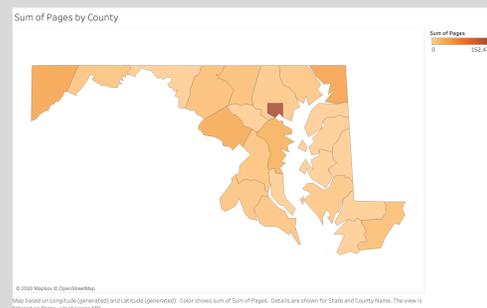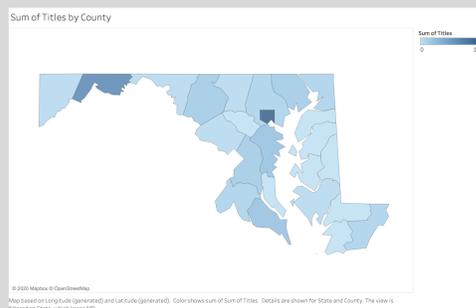
Next Steps

Moving forward with analyzing the collection, items to consider include:

- additional data collection, comparisons, and/or computations;
- other ways of displaying and visualizing the data;
- adding a process to view the data as families of titles;
- most effective way to include temporal data and visualizations;
- continued research in best practices for diversity and inclusion in digitization selection;
- utilizing data from the digitized newspaper survey that is conducted as a part of the project; and
- ways to make this process more automated and efficient for the future.

This research will be shared with the HMNP advisory board in making future title selection decisions and the program will continue to evaluate policies and procedures to incorporate diversity and inclusion data in selection and decision making. The HMNP staff and advisory board will also determine the best way to share this information with archives, libraries, and museums throughout the state and make it available on the HMNP website.

Contact:
Pamela A. McClanahan
Digital Projects Librarian
pmcclana@umd.edu
301-405-1798

Historic Maryland Newspapers Project
lib-newspapers@umd.edu
www.lib.umd.edu/digital/newspapers/
@HistoricMDNews

UNIVERSITY OF MARYLAND 1856

UNIVERSITY LIBRARIES