

Introduction

Understanding collections as data is a new paradigm within cultural heritage institutions and presents both challenges and opportunities for professionals working in libraries, archives, and museums (Padilla and Higgins). As part of my work as the Humanist-In-the-Loop Graduate Research Fellow at the Harry Ransom Center, I was tasked with thinking about how to present the rich data of archival collections to diverse audiences while also teaching students about primary source literacies. Paying particular attention to the wealth of descriptive metadata in the Movie Poster Collection housed within the Ransom Center, I was able to generate data visualizations and online educational resources in the service of rethinking discoverability, sparking user engagement, and reimagining how to teach with the primary resources housed within cultural heritage institutions.

Research Goal

Use cultural heritage data to present archival collections to users and generate educational resources at the intersections of critical archival studies, digital humanities, and primary source literacies.

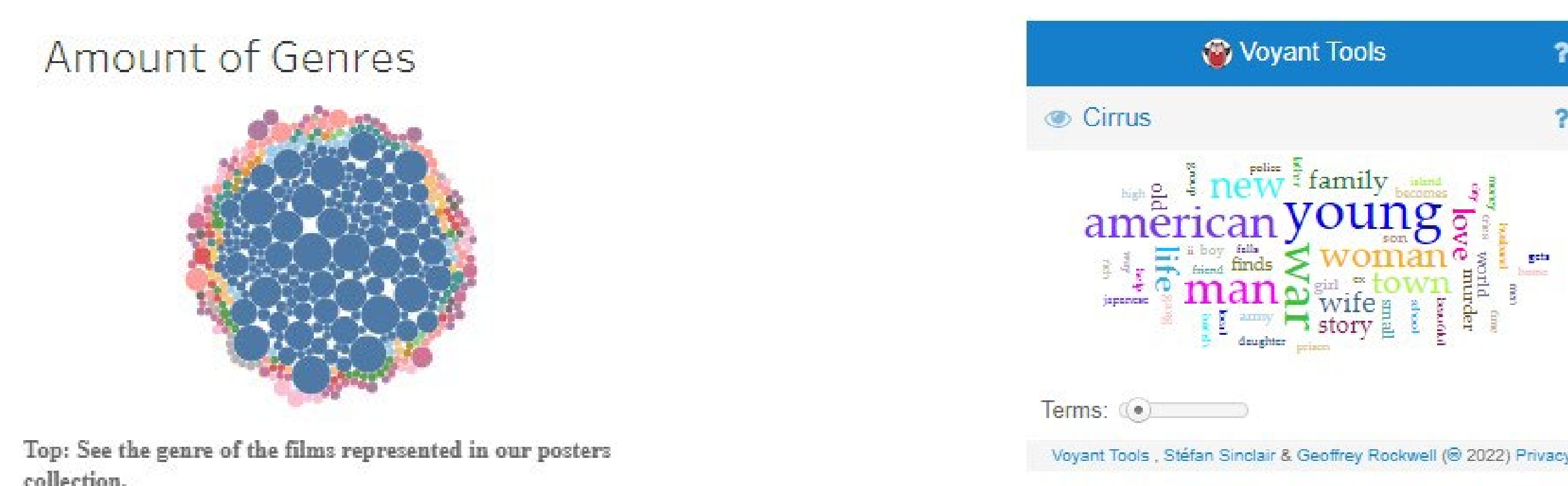
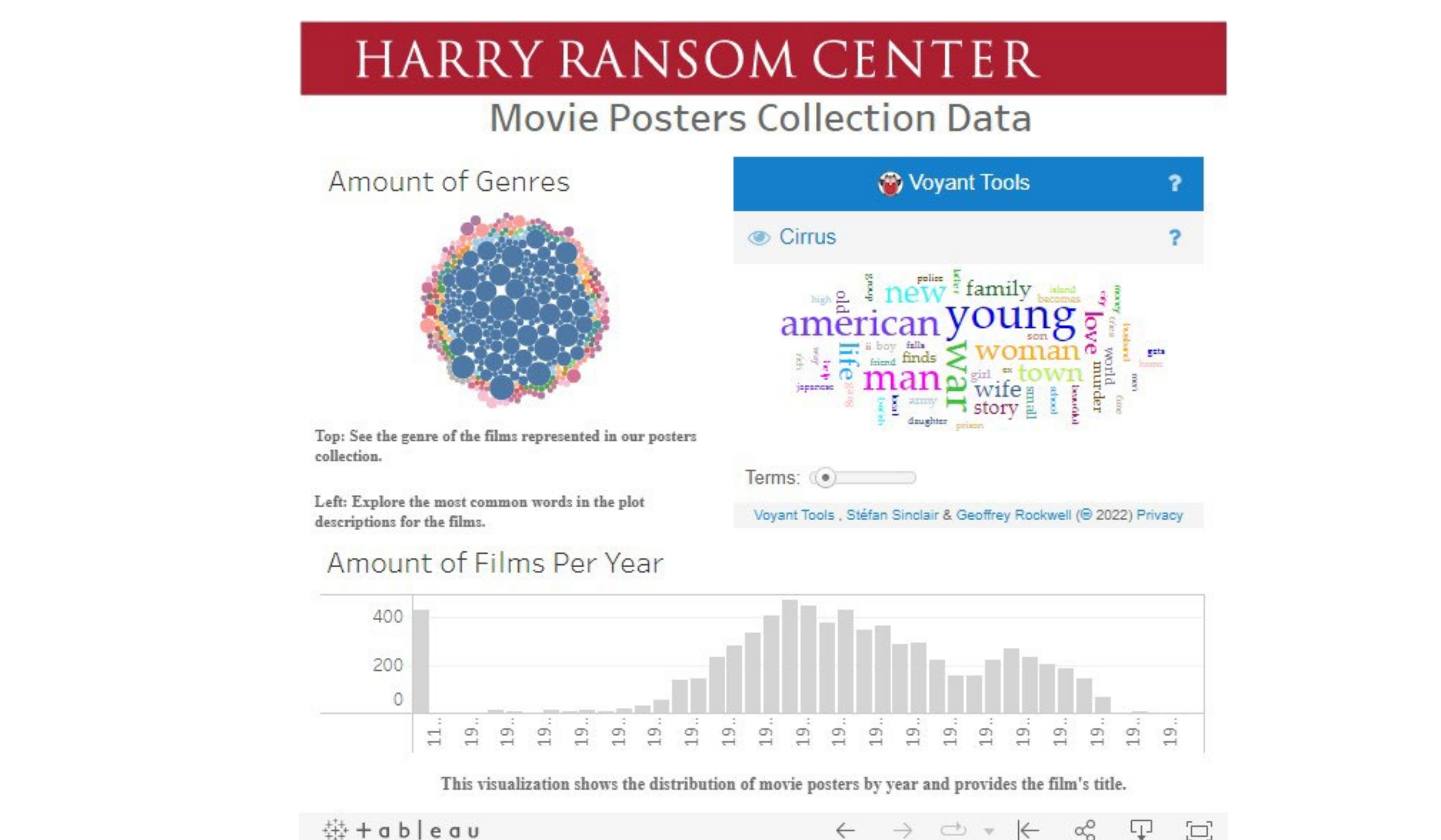
Methods

The first part of this process was learning more about the Movie Posters Collection and how these archival holdings came into being. This took the form of various interviews and conversations with Ransom Center staff members. Subsequently, it was decided that the most pertinent data to be mined from this collection consisted of descriptive metadata pertaining to year, genre, plot, and title.

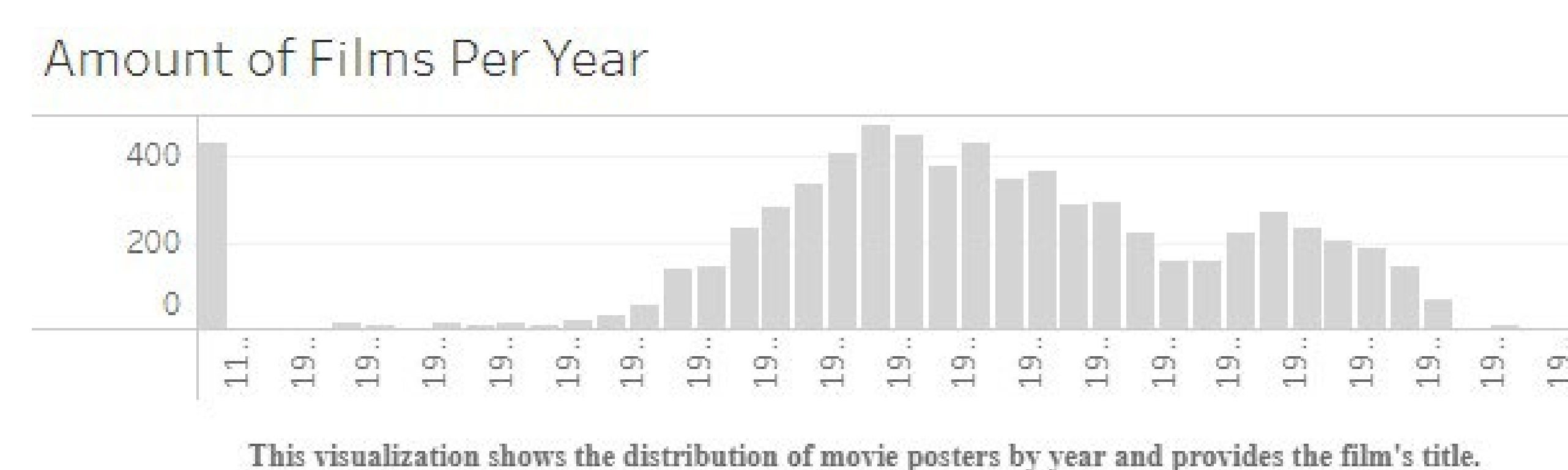
Metadata was mined from both the collection holdings itself and cross-referenced with data from the Open Media Database. Once the data was mined and cleaned up, it was uploaded to Tableau Public so that visualizations could be generated. After the data was uploaded and visualizations were created. A set of three lesson plans and corresponding Google jamboards were created.



Figures and Results



The visualization of the data was generated using Tableau Public. This visualization now lives online as an open-access data tool available to anyone on the web. The visualization focuses on three area of interest that were found significant throughout the planning and data mining process. These areas are as follows: Genre, Plot descriptions, & Year/Date. Both the graphs for the genre and year are interanimated, so that students can explore the genre designations for the poster while also seeing what year they were shown.



Visualization information metadata concerning the plot of the films was more challenging, as this description was not extant in the records of the Ransom Center itself. Therefore, data regarding the plot of the films was mined from the Open Media Database and new data visualization tool was employed alongside Tableau Public. Voyant is a text-based data visualization tool that reads uploaded texts and generates visualizations. The word cloud that Voyant created from the plot metadata shows the most common words used in describing the general plot of the films. Users can interact with this graph to think through key terms that recur when describing the films and their narrative content.

Conclusion

This project is just one way to think through collections as data and the opportunities this paradigm may yield for scholars and educators working with cultural heritage data. However, there are certain challenges that prospective researchers and educators should keep in mind as they move forward with work at the intersections of critical archival studies, digital humanities, and primary source literacies.

There was a significant learning curve as someone trained in the liberal arts/humanities to become conversant in both technologies and terminologies. If this kind of new and innovative work is to be sustainable, then researchers and funding entities must rethink how to collaborate, fund, and build capacities for these endeavors.

Most importantly, timelines for producing this work must be stretched to account for the ways in which humanities students/scholars do not generate findings in the same ways or within the same timeframes as researchers within the traditional scientific fields. With that being said, this project serves a singular entry point into the field of collections as data and the occasions for collaboration and scholarly inquiry are much vaster than this study has presented here.

Acknowledgments

I would like to thank Good Systems for sponsoring this project and my position. This experience has been invaluable to my growth as a scholar and educator. I would also like to share my deep appreciation for the members of the Humanist-In-the-Loop grant team for their insights, guidance and enthusiasm. Much thanks and boundless gratitude to Tanya Clement, Andi Gustavson, Lauren Walker, Allyssa Guzman, and Aaron Choate.

References

Padilla, Thomas G., and Devin Higgins. "Library collections as humanities data: the facet effect." *Public Services Quarterly* 10.4 (2014): 324-335.