In the past, much of the study of Chinese narrative literature has focused on a very small sample of novels – usually those that were best known. This project’s aim is to broaden the scope of fiction studied. In order to accomplish this, Digital Humanities approaches were explored. Making data abstract allows for easier visualization and facilitates the study of patterns, rather than of details alone.

Ōtsuka Hidetaka’s *A Bibliography of Chinese Popular Fiction* was used to create a database which serves as the basis of all other work in this project, along with the other references cited below. In order to fully test some possible digital tools, six collections of Chinese vernacular short stories were chosen from this database for analysis. One in particular, *Strange Views Old and New* (Jīngǔ qíguān 今古奇觀) was of particular interest, as it seems to be one of the most popular works of vernacular fiction in Ming-Qing China, having 55 printed editions in total. For the purpose of analysis, information from the database was used to create graphs, and the text of these collections were run through “The Late Imperial Primer Literacy Sieve”. This tool allows for Chinese texts to be run against chosen primers, showing the percentage of unique characters that the subject would be able to read had they studied those primers. Supposing that the popularity of *Jingu qiguan* could be due to accessibility by wider audiences of readers, we chose the most basic primers- The Three-character classic (Sān zì jīng 三字經), Hundred family names (Bǎi jiā xìng 百家姓), and Thousand-character essay (Qiān zì wén 千字文) - in order to analyze of the relative accessibility of each collection of vernacular fiction.

*A comparison of aggregate collections against chosen primers, the cross section representing the accessibility of the texts (on the left) if the subject had studied the primers (on the top).*
A complete view of output generated by the literacy sieve. On the bottom left is a visual representation of the stats on the right, the reddened and blacked out characters being the ones unknown to a subject who had studied the selected primers.

Through the use of these digital tools, it was found that the accessibility of the texts tended towards an association with the time-period in which the story was written, the most accessible stories being those that were written in the middle period (ca. 1450-1550). However, the stories that were anthologized in the *Jingu qiguan* story collection were almost entirely late-period stories – those which our analysis showed were the least accessible of all the periods. The popularity of *Jingu qiguan* does not seem to be due to greater accessibility by a wider audience, so other explanations must be found. This brings to light some interesting questions and possible patterns relating to the process of anthologizing or transmitting of fiction, which may aid understanding of the history of the Chinese vernacular story.

References


Acknowledgement

This work was supported by the University of Utah Office of Undergraduate Research.