Investigating the 16th Century Book Trade with Data Science and Machine Learning

By Jamie Locke-Jones, University of St Andrews
Supervised by Graeme Kemp, School of History

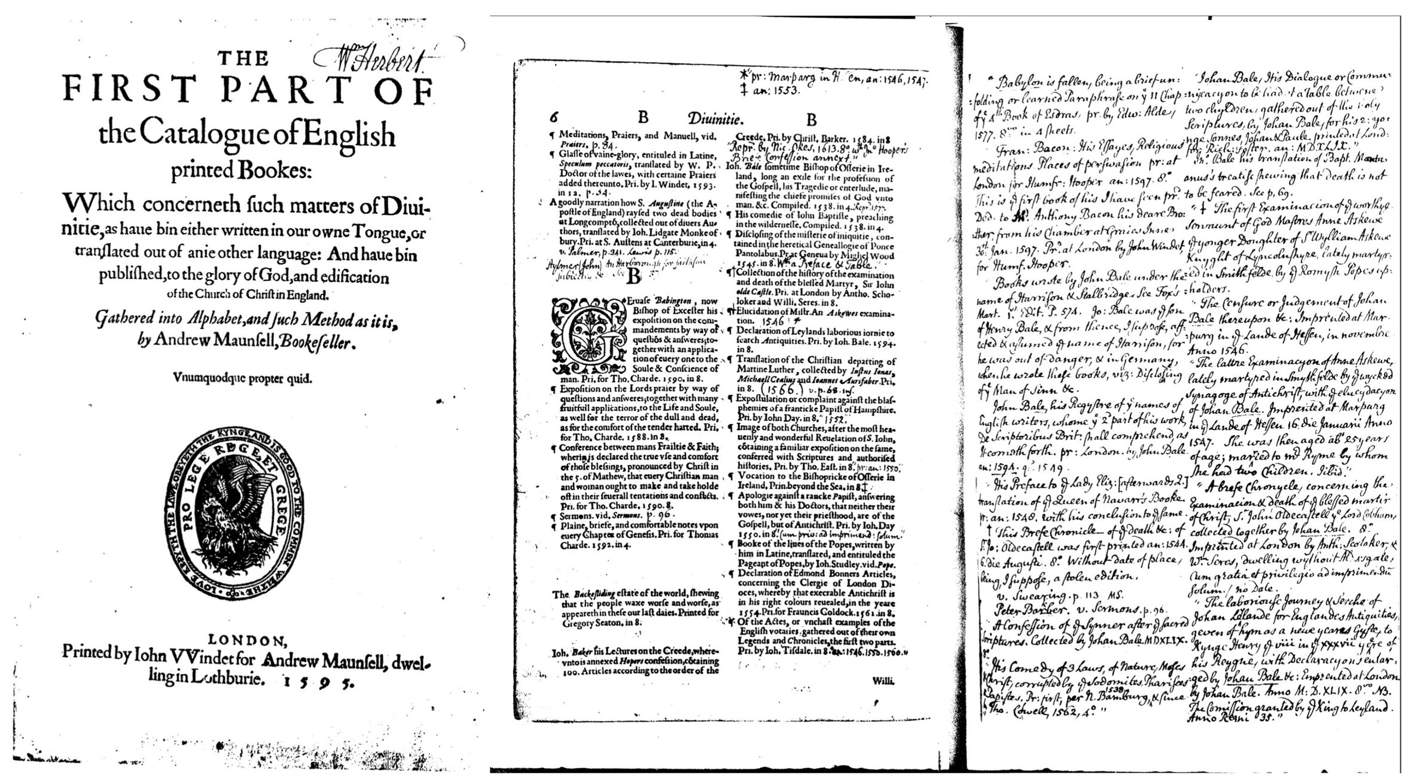
Introduction

By the 1500s, books were the primary means of communicating ideas in Europe. In a time of huge artistic and scientific development, the books being printed illustrate the priorities of contemporary figures, their opinions, and what they thought the reading public should know.

Unfortunately, many of these books have since been destroyed. They now exist only in records such as national bibliographies, of which Andrew Maunsell's Catalogue of English Printed Books is the first in the UK. Maunsell printed two parts of his catalogue, the first focusing on religion, the second on science.

Many copies were heavily annotated by collectors. William Henry Miller, a Scottish parliamentarian and book collector, wrote extensive notes in his copy, almost doubling the number of books recorded in the catalogue.

Until now, no-one has systematically studied Maunsell's book or its annotations. My project has produced a freely available transcription of the entire book, including Miller's annotations, produced using machine learning character recognition and encoded with valuable bibliographic data. This enables some key insights into the 16th century book trade.



From left to right: The title page of Maunsell's catalogue; a printed page of the catalogue, and a blank page interleaved after publication and annotated by William Henry Miller.

Methodology

The transcription was completed by training a character recognition model through the software suite Transkribus. I then used Python and R to convert the Transkribus output to XML, a standard language for encoding text with additional data. My XML links entries to the English Short Title Catalogue, an online repository which contains more information about each book. Scan the QR code to view my transcription on Github.

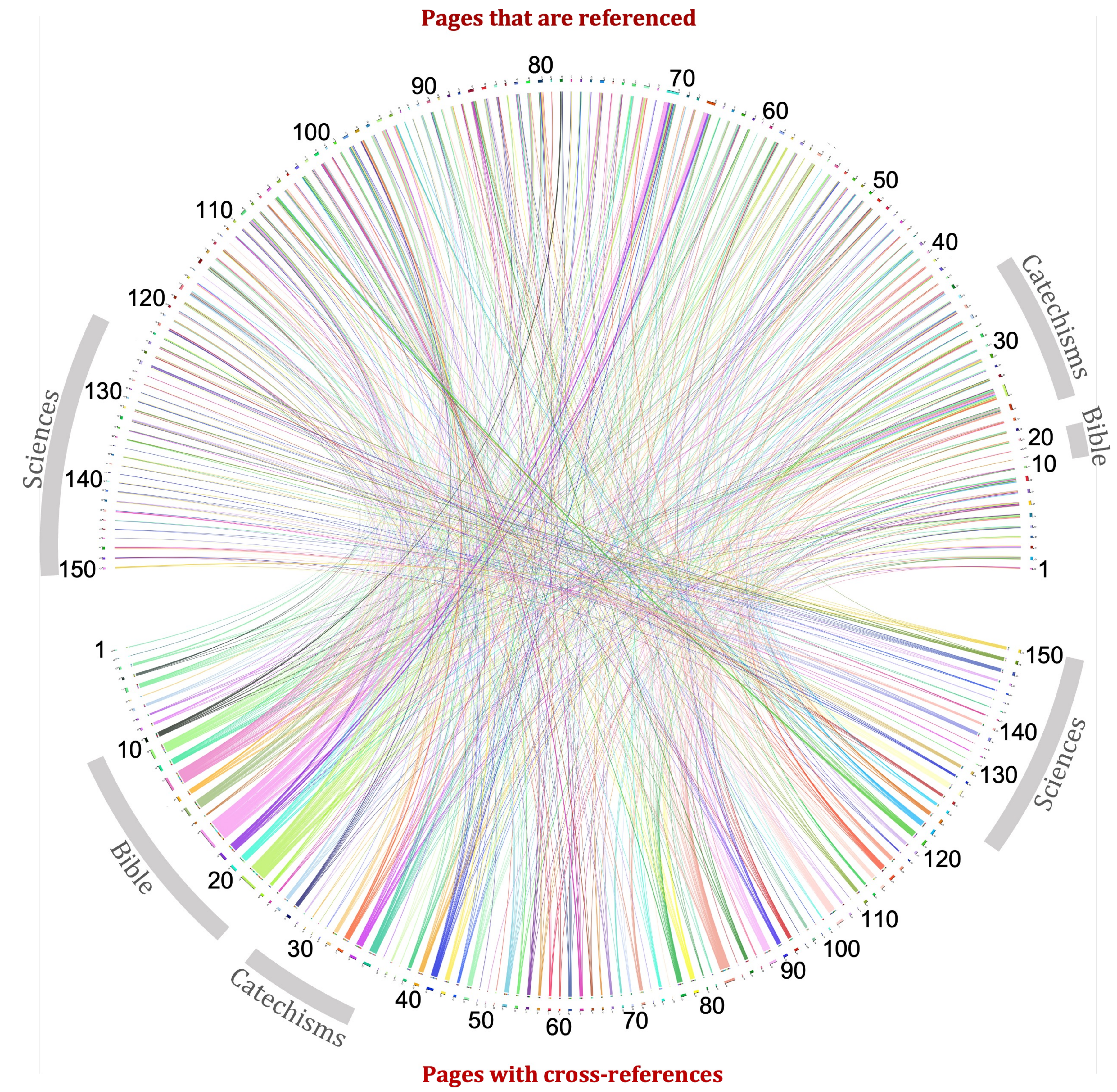
SCAN ME

There is a lack of interaction between religion (pages 1-123) and the sciences (pages 124-150). Religious teachings played a huge role in day-to-day life, but even so they were kept separate from the more down-to-earth teachings of disciplines such as cookery and navigation.

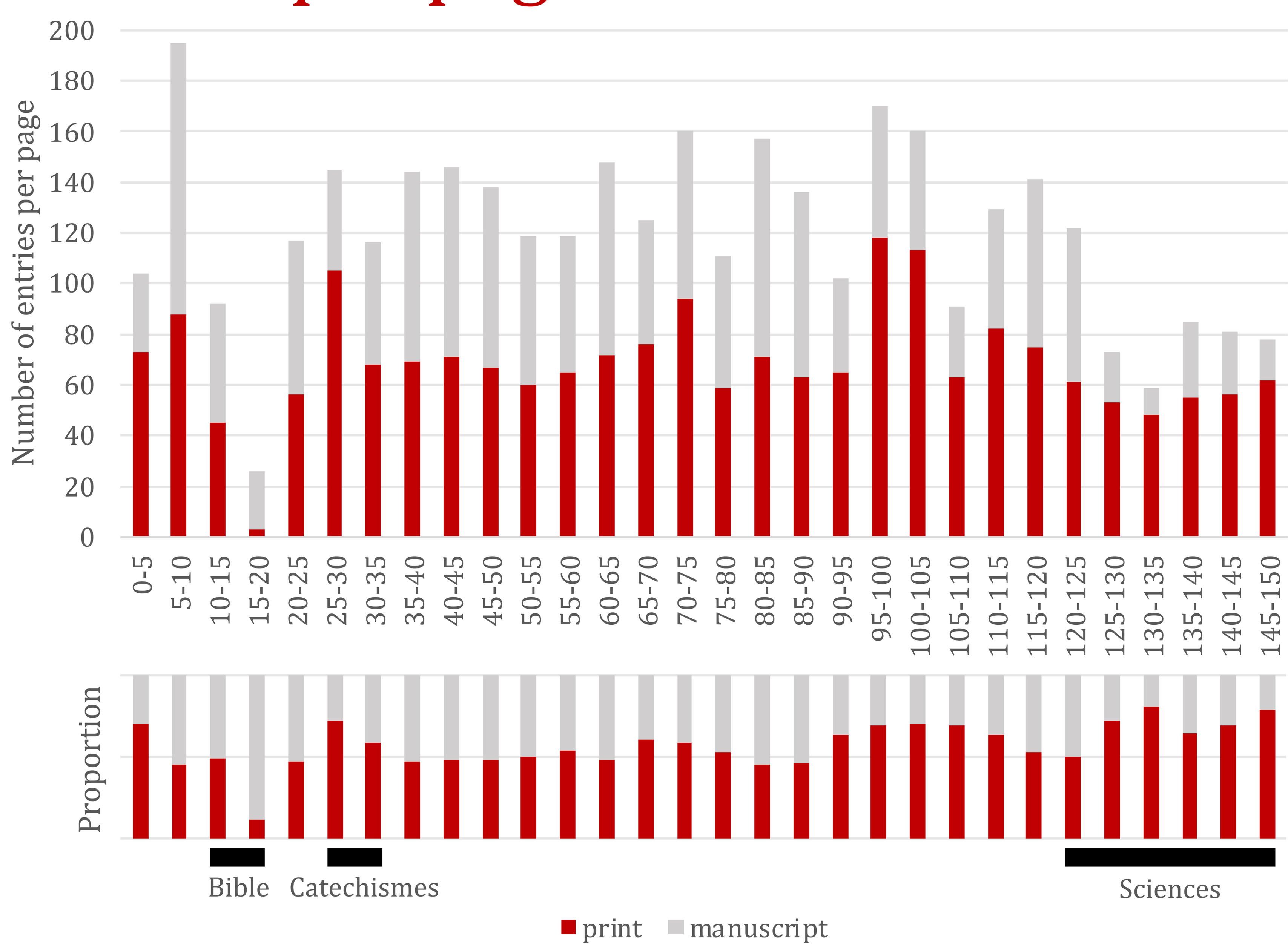
The section with the most cross-references per page is the Bible, which shows that, far from sticking to one accepted interpretation, the 16th century saw many different authors tackle the centrepiece of their religion.

Cross-references

Maunsell's catalogue contains over 3,000 cross-references to other sections of the book, i.e. from the Bible section to the analyses of religious scholars. Miller's main contribution to many of the printed pages is to number the cross-references, showing that he used this book as a reference for his collecting. The references paint a picture of how interconnected the book trade was:



Entries per page



The number of manuscript entries demonstrates how incomplete Maunsell's initial catalogue was. That Maunsell, an experienced bookseller, managed to catalogue only a fraction of the industry is a testament to its size and complexity even at this early stage.

Religion had a much larger place in the industry relative to science. This is notable considering Maunsell's outsized definition of 'science' – he considers navigation, cookery, and even music part of the sciences.

It seems Maunsell's experience as a member of the book industry gave him a more well-rounded understanding of the trade than Miller, who is more knowledgeable about religion than science. While collectors often gain specialised knowledge, tradespeople in the book industry were not particularly specialised in terms of genre.

University of St Andrews | FOUNDED 1413
The Carnegie Trust for the Universities of Scotland