

Publishing Patterns and Title Themes in British Drama

(1660–1900)

Xinran Liu (r 1069247)

Jan 2, 2025

1. Introduction

The 16th to 19th centuries saw significant shifts in British society, shown also in the changes in the drama publishing industry. Based on the dramatic metadata dataset of the British library, this report examines these developments using quantitative analysis.

The analysis is divided into two main parts. Section 2 focuses on general publishing patterns, while section 3 analyzes dramatic titles in greater detail.

Python was used to process the data, involving further data wrangling, text analysis on title keywords, and initial data visualization. (Codes and methodology shown in Jupyter Notebook and GitHub). Tableau was then used to visualize the resulting data, and results are displayed on this [dashboard](#).

2. Descriptive analysis of the drama datasets

The part of analysis focuses on general publishing patterns, including fluctuations in publication volume and genre, authorship trends, and the concentration of publishers.

2.1. Publications Over Time

There are a total of 1615 publications in the cleaned drama datasets, 4 of them without publication year recorded, leaving 1611 rows of data being analyzed.

Overview of publications over time

Looking at the timeline, there are two noticeable peaks, with a relatively quiet period in between. After a very slow beginning in the mid-17th century, publication **rises sharply** in the 1690s, reaching 92 entries. This first boom is followed by a **decline** across much of the 18th century. From the early 19th century onwards, however, the numbers **increase again** and continue to grow steadily, eventually reaching a **high point** in the 1890s with around 200 publications.

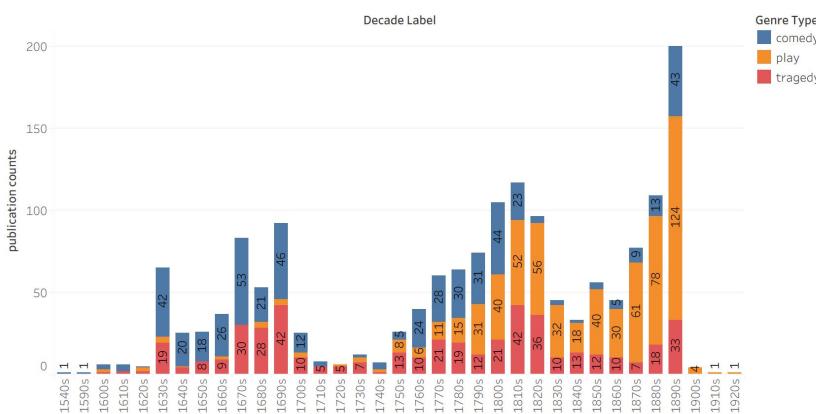


Figure 1. Numbers of Dramas Published by Decade (comedy, play, tragedy and all combined)

Difference among the three sub-types

A similar change can be seen when looking at the types of drama. In terms of overall numbers, the three main categories are relatively close: Play appears most frequently (636),

followed by Comedy (525) and Tragedy (450). What differs much more clearly is their distribution over time. In the late 17th century, **Comedy** dominates the market, often appearing at roughly twice the rate of the other genres. By the 19th century, however, the more general genre **Play** becomes increasingly common. From the early 1800s onward, its numbers rise steadily, and by the 1890s it has become the leading category with 124 entries. Comedy, in contrast, declines in number, while **Tragedy** retains a more stable middle position, never becoming dominant but not disappearing either.

2.2 Authorship over time

The dataset contains a total of 1,611 records, **85.72%** of the entries (1,381 works) list an identified author, while the remaining **14.28%** (230 works) are labeled as “Unknown.” This suggests that authorship information is generally well recorded in the dataset, although a small portion remains unclear.

With **850** unique individual authors in total, the “Top 10” bar chart shows that Richard Cumberland is the most productive author, with 24 published works, followed by John Dryden with 21 publications, and Thomas Shadwell with 19. William Shakespeare also appears among the top authors with 14 works.

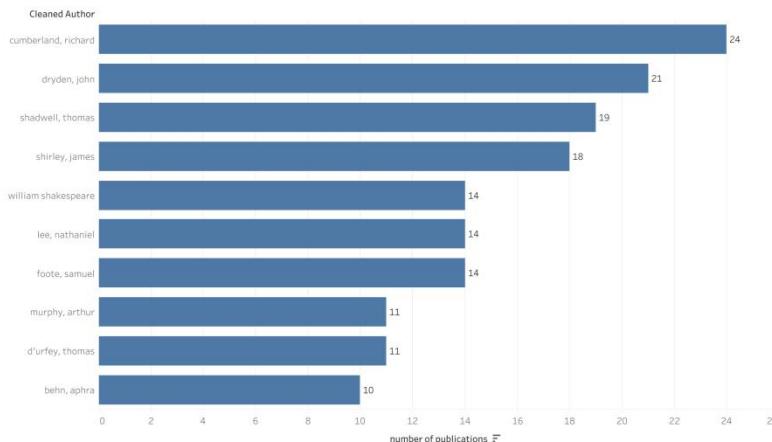


Figure 2. Top 10 authors and number of publications

Unique authors over time

First, there is a **steady increase** reaching a peak in the 1670s, with around 41 authors, followed by a **sharp decline** throughout the 18th century, when the number drops to 6 authors in the 1720s. In the final stage, the trend changes again, and the number of unique authors **rises rapidly** towards the end of the timeline, reaching its highest point in the 1890s with 171 authors.

These changes can be linked to broader historical contexts. The initial rise in the late 17th century corresponds to the Restoration period, when theatres reopened after 1660. In contrast, the decline during the mid-18th century may be related to the Licensing Act of 1737, which imposed strict controls on theatres and plays. Finally, the dramatic increase in the 1890s might reflect the impact of the Victorian era, driven more by developments in mass printing and publishing.

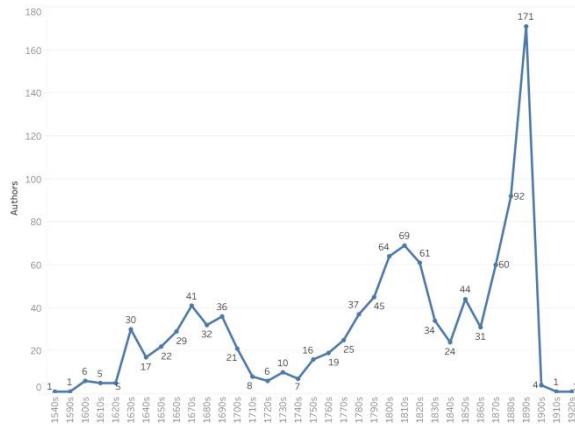


Figure 3. Unique authors over time

It is noteworthy that publication volumes and the number of unique authors follow a rather similar trajectory over time, which align with broader shifts in the literary marketplace. The initial **rise** in the late 17th century correlates with the **Restoration stage** following the reopening of theatres in 1660. The marked **decline** during the mid-18th century parallels the impact of the **Licensing Act of 1737ⁱ**, which imposed strict censorship and restricting patent theatres. Finally, the dramatic resurgence in the 1890s might reflects the industrialization of print in the **Victorian era**, featured by the expansion of mass publishing and a growing readershipⁱⁱ.

Unknown Authors over time

In the early decades (1540s–1590s), almost no works lack author attribution, but the ratio quickly rises after 1600 and then fluctuates between the early 17th and mid-19th centuries. Several peaks occur around the 1730s and 1760s, when about one quarter of works are unattributed, while sharp drops appear in the 1680s and 1700s with ratios falling below 0.05.

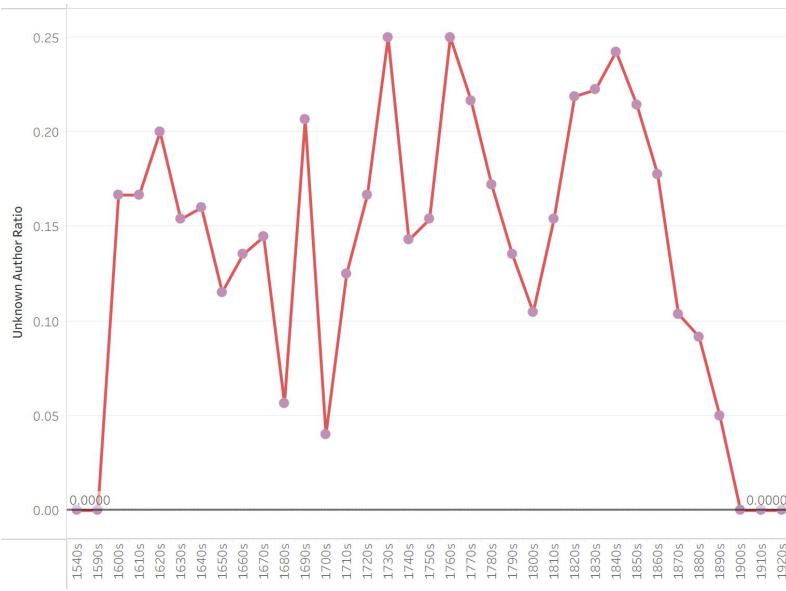


Figure 4. Unknown authors over time

2.3 Publishers over time

The total unique publishers over this period are 624, with 991 publications without a documented author, equaling a **Missing Data Rate of 42.33%**, significantly higher than that of the **authors (14.28%)**, which may suggest the publishing world and the audience generally prioritize authors over publishers over this period.

Here concentration analysis was conducted with **concentration ratio (CR3, top 3 publishers)**. It is shown the British drama publishing market was generally competitive, with no single group of publishers dominating the industry for most of the period. In most decades, the concentration ratio stays **below 25%**, indicating a fragmented market structure. An exception appears between the 1710s and 1750s, when the ratio rises to **around 50%**. This temporary increase suggests a looser form of oligopoly, which may be linked to changes in copyright regulation following the Statute of Anneⁱⁱⁱ in 1710, the first statute on copyrights.

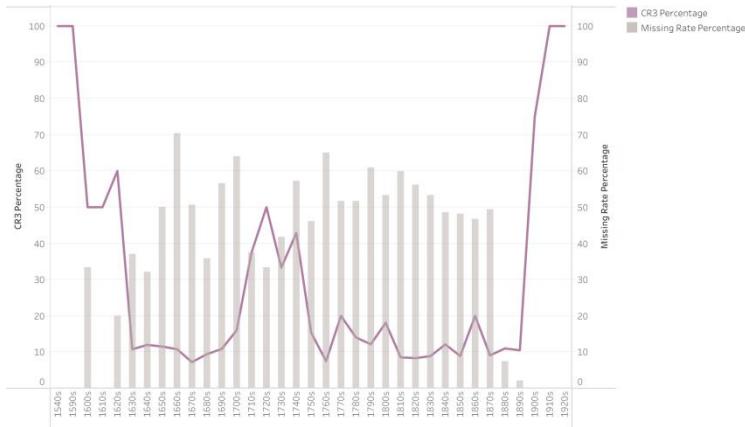


Figure 5. Publisher CR3 and missing rate over time

However, these results should be treated with caution. As shown by the grey bars, the proportion of **missing data** is often very high, frequently exceeding 50%. In addition, the sharp concentration levels at the very beginning and end of the timeline are more plausibly caused by the limited survival of archival records in those periods.

3. Analysis of Titles and Socio-cultural Reflections

This section analyzes dramatic titles in greater detail, using keyword trends to explore how plays reflected expanding global interests and changing ideas about private life.

3.1. Evolution of Geographic Discourse in British Drama Titles

In this part of the analysis, several methodological approaches were tested before a hybrid strategy was adopted. After several attempts, a hybrid approach combining historical knowledge and computational analysis was used to build a geographic keyword list. First, the **top 500 frequent words** from titles were extracted in Python and grouped into three categories: **Continental (Europe)**, **Colonial (New World)**, and **Oriental (Exotic East)**. Then, a full-corpus scan was utilized to identify the words **co-occurring** with these seeds to validate the list, with ambiguous terms manually filtered out. Relative frequencies by decade were calculated using Regex matches, and 100% stacked bar charts were used to visualize temporal changes in geographic themes. The analysis of title keywords indicates three distinct phases in terms of geographic space:

High Volatility (1660s–1760s). Over this period, several decades display a 100% share for a single category. For instance, the 1740s are entirely dominated by "Oriental" references, while the 1760s by "Colonial" terms. These absolute values are likely attributable to **small sample sizes** rather than sudden cultural shifts.

Diversification of Themes (1770s– 1860s). A more **mixed distribution** is observed during this period, during which no single category completely dominates the discourse. An exception is observed in the 1860s, where "Colonial" themes account for the entire share of geographic references. However, this sudden spike may again reflect sample limitations.

The Resurgence of Continental References (1870s–1890s). Starting in the 1870s, the "Continental" category began to increase significantly. By the 1880s, **Continental references regained a dominant position**, accounting for approximately 80% of the geographic titles. While the "Colonial" and "Oriental" became marginalized during this period, they remained present **as a stable minority** (roughly 20%), showing that while the focus of the British authors and readers turned towards Europe, the interest in other parts of the world did not vanish entirely.

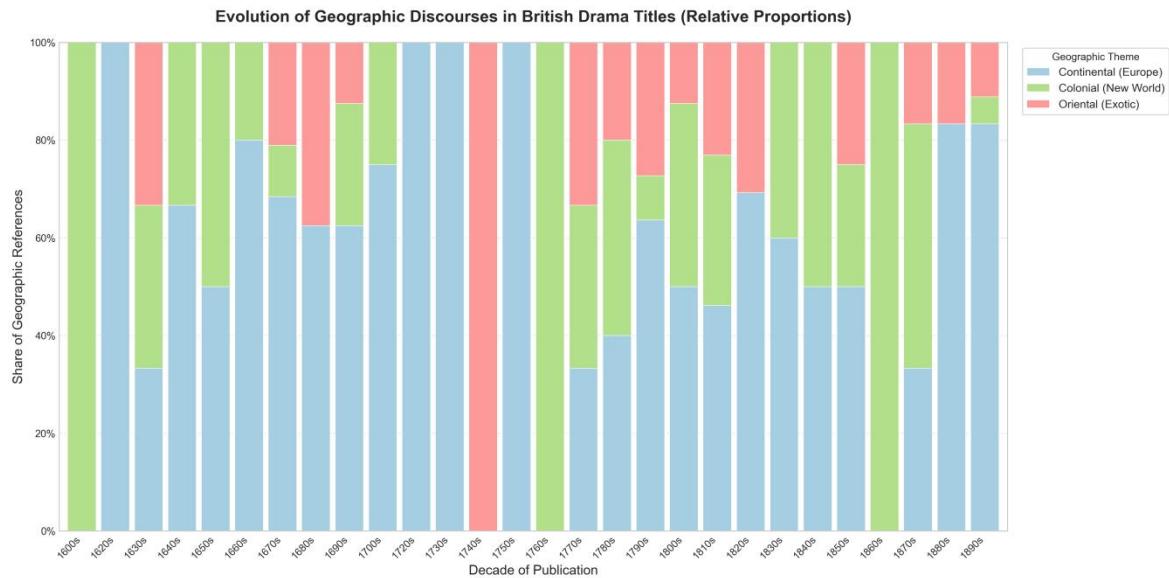


Figure 6. Evolution of Geographic Discourse in British Drama Titles (relative proportions)

3.2 The decline of the Aristocratic and the rise of the Domestic

The evolution of aristocratic / domestic themes in drama titles was examined through an **iterative semi-supervised co-occurrence analysis** with Python. Initial textual scans showed that the 'Aristocratic' category in the 19th century was frequently linked to specific historical names (e.g., Henry, Mary), whereas the 'Private' category was strongly associated with words of emotion and tension (e.g., Secret, Fatal), in addition to traditional kinship terms. The **keyword dictionaries** were **dynamically expanded** to incorporate these specific markers. Three distinct phases in the evolution of titling conventions are revealed by the data.

Mainly Aristocratic Dominance and Volatility (1660s–1730s). After the Restoration in 1660, titles related to aristocracy **surged** and then experienced a **decline around 1700**, before **rebounding** sharply. This resulted in a peak of around 24% in the 1720s, showing the aristocratic discourse remained dominant but contested. Meanwhile, the private sphere had a more stable increase during this period.

Domestic Dominance (1740s–1790s). A significant shift is observed after the Licensing Act of 1737. A **sharp decline** is recorded in the aristocratic domain—dropping from around 32% to under 5% within two decades. In contrast, the private lexicon maintained at a stable level of 12% to 15%. The private domain **surpassed** the aristocratic in this period.

Aristocratic Resurgence (1800s–1900). The 19th century was marked by a **resurgence in aristocratic terms**, forming a second peak in the 1830s (around 24%). Lexicon-wise, this 19th-century rise is driven by **specific monarch names and historical terms**, which were identified in the coding process. By the final period, a downward slope occurred in both groups.

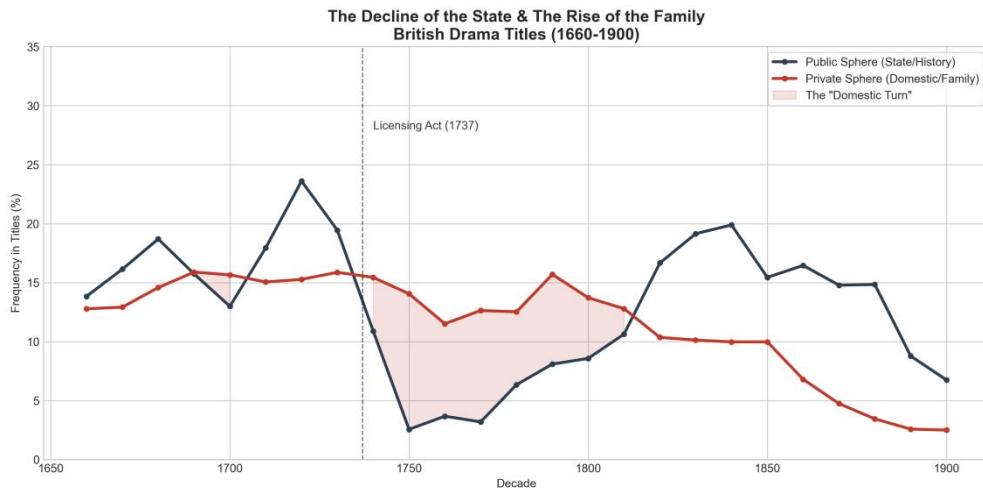


Figure 7. The decline of the state and the rise of the family in British drama titles (1660-1900)

4. Discussion and limitation

Several limitations regarding data quality, the initial state of the dataset, and methodological assumptions must be acknowledged.

4.1 Assessment of Dataset Suitability

The structure and cleaning status of the original dataset were found to be generally suitable for immediate quantitative analysis. However, additional manipulations were also required for achieving research objectives.

Structural Inconsistencies: The primary temporal field, p year, contained date ranges and non-numeric characters rather than single integers.

Need for Feature Engineering: Essential analytical categories, such as Decade_Label had to be generated to enable the aggregation required for the study.

4.2 Data Bias

It was observed during the data cleaning phase in OpenRefine that a significant portion of the dataset **lacked author attribution**. Although these entries were standardized into a distinct "Unknown" category (representing 14.28% of the corpus) or merged via the other author (n) column, the analysis of authorship trends is inherently biased towards established figures.

In Spatial analysis, explicit geographic keywords were found in only 6.07% of the total dataset N=98. Due to this small sample size, percentage-based visualizations (100% stacked bar charts) are used to reflect the changes in raw counts. Therefore, sudden spikes in specific regions may be attributed to the publication of a single popular play or statistical noise.

4.3 Reflection on Quantitative Methods

During the adjustments of codes, it is found that the interpretability of the results was found to be heavily dependent on the defined noise words. The removal of such words may have smoothed over subtle shifts in the charts.

For the classification of semantic themes, three approaches were tested, which shows the limitation of subjective categorization: Expert-Built Dictionaries, Frequency-Based Selection, and finally Semi-Supervised Co-occurrence (final method).

i The 1737 Licensing Act and Its Impact. <https://doi.org/10.1093/oxfordhb/9780199600304.013.015>

ii The Victorian Era in English Literature – A Complete Guide to Literature, Culture, and Life.

<https://engliteratureguru.com/the-victorian-era-literature-features-writers/>

iii The Statute of Anne; April 10, 1710. https://avalon.law.yale.edu/18th_century/anne_1710.asp