

Our project team worked with Anouk Lang, a researcher of 20th and 21st century literature, to analyse data from [LibraryThing.com](https://www.librarything.com/). The website provides a platform to catalog one's books and find others with similar book collections. Some users have catalogued book collections on behalf of famous authors, such as Ernest Hemingway and Virginia Woolf. Lang downloaded data about the book collections of 13 American and British writers from the 20th and 21st centuries. As a digital humanist, Lang was interested using technology to understand the network of relationships between book owners, book authors, and the books themselves.

Anouk Lang gave us a CSV file of text and numeric data with 23,351 rows and 11 columns. There are 13 book owners in the dataset, and information about book titles, their authors, their publication dates and locations, and some book owners' libraries' locations. Before visualization, we needed to do clean the data because there were missing values and inconsistent formatting of the values in each column of the dataset. For example, many book publication dates are missing or were given as a range of years instead of a single year. Our researcher gave us advice that we could ignore the book collection column because that column was empty for many book owners. As for the missing dates, our researcher recommended we consult a website called JSTOR to find the valid book publication dates.

Our audience for our data visualizations can be divided into two groups: literary researchers and fans of the authors. To communicate the breadth of our dataset, we decided to create a data comic about one of the book owners, Ernest Hemingway. To explore relationships between Hemingway's life and book collection information, we searched online for information beyond what was in the dataset, such as the length of staying in different places, friendship with other authors and his own writing experience. To communicate the large size of the dataset, we created visualizations in Tableau that show a broader view of the dataset.

To interest literary researchers (following our researcher's recommendation), we created networks to show relationships between authors based on the books in their libraries, which could vividly show connections between authors. This can be a good starting point for the literary researchers if they are looking to better understand relationships between authors or patterns of influence among authors. Both of our audience groups can benefit from the data comics we created. The comic zoomed in from the whole dataset to a specific author to show the stories that can be discovered using the visualizations as a starting point. This can also attract fans of the authors because they may want to know the details and stories related to the author instead of just seeing the summary of statistics.

The main purpose of the data comic was to use a typical example to show what we can know about people from their books, so we chose Earnest Hemingway's book collection information to make a visualization. The first part is an autobiography of one Hemingway's book as a comic to communicate how far books can travel; the rest of the comic focuses on Hemingway's life experiences. We used a timeline combined with his book collection information to illustrate the how his role as an author and book collector developed over time. At the third part, we put a world map on our comic canvas and use different nodes in varying shades to demonstrate the location links.

After showing the relationships between books and the owner, the comic explores relationships between people. We show the most popular authors in Hemingway's library and put several important relationships he had with other people in the dataset (which we discovered through online research) to our data comic. For the final part of the data comic, we chose a relationship network made with Vistorian.com to visualize the relationships between the libraries of Ernest Hemingway, Gertrude Stein, Ezra Pound and Georges Simenon.

The challenges we faced creating the data comic was having an adequate amount of time to draw most of the story elements. We also spent a lot of time thinking about how to sequence the different story elements logically.

Our project team also faced a challenge in finding a visualisation that communicated the scope of the dataset while still providing discernible patterns for a viewer. To create node-link diagrams that did not have an excess number of links, we created subsets of the data based on a book's decade of publication. These data subsets could be visualised as networks between book owners and the authors in their libraries. We also found that a node-link diagram of books written by the 13 owners created a network with visually discernible relationships. We used Python to generate CSV files of the data subsets and Vistorian.com to visualise the networks.

For a broader view of the dataset, we used Tableau's tree maps and area charts. The tree maps communicate the size and diversity of each book owner's library. Upon visualising the number of books in each owner's library (figure 1), and visualising the number of authors in each owner's library (figure 2), one can see that in our data, the size of an owner's library indicates the diversity of authors in the library.

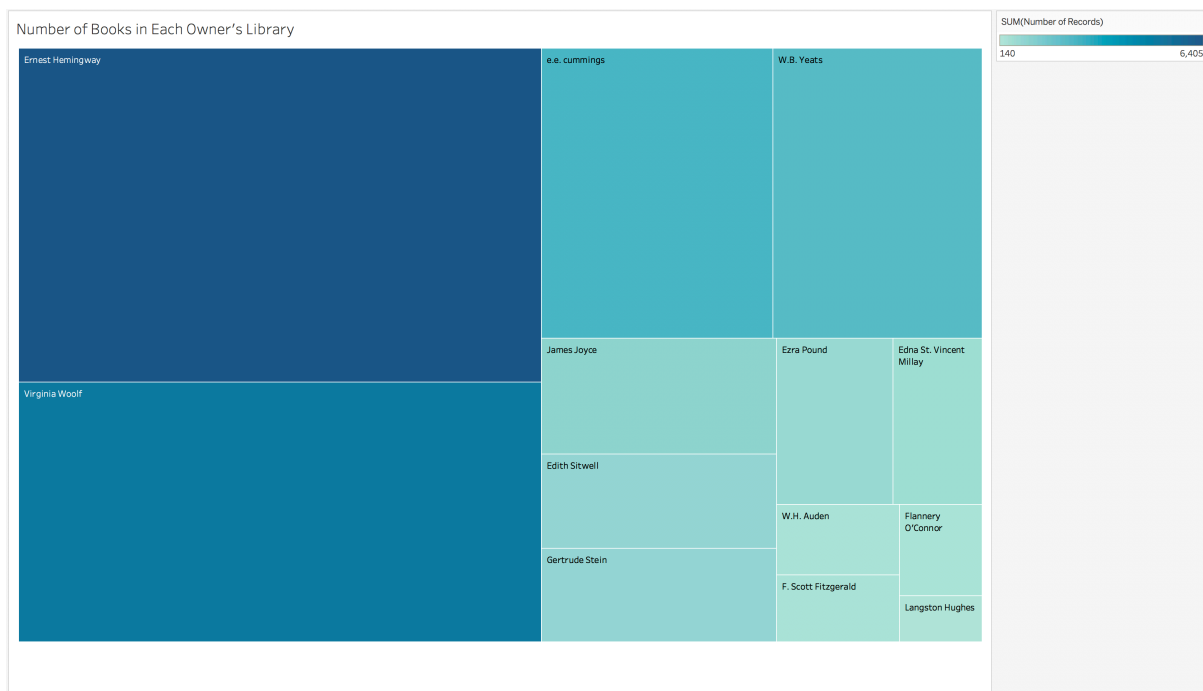


Figure 1 The number of physical books present in each owner's library, with larger numbers indicated by a darker colour and bigger rectangular area.

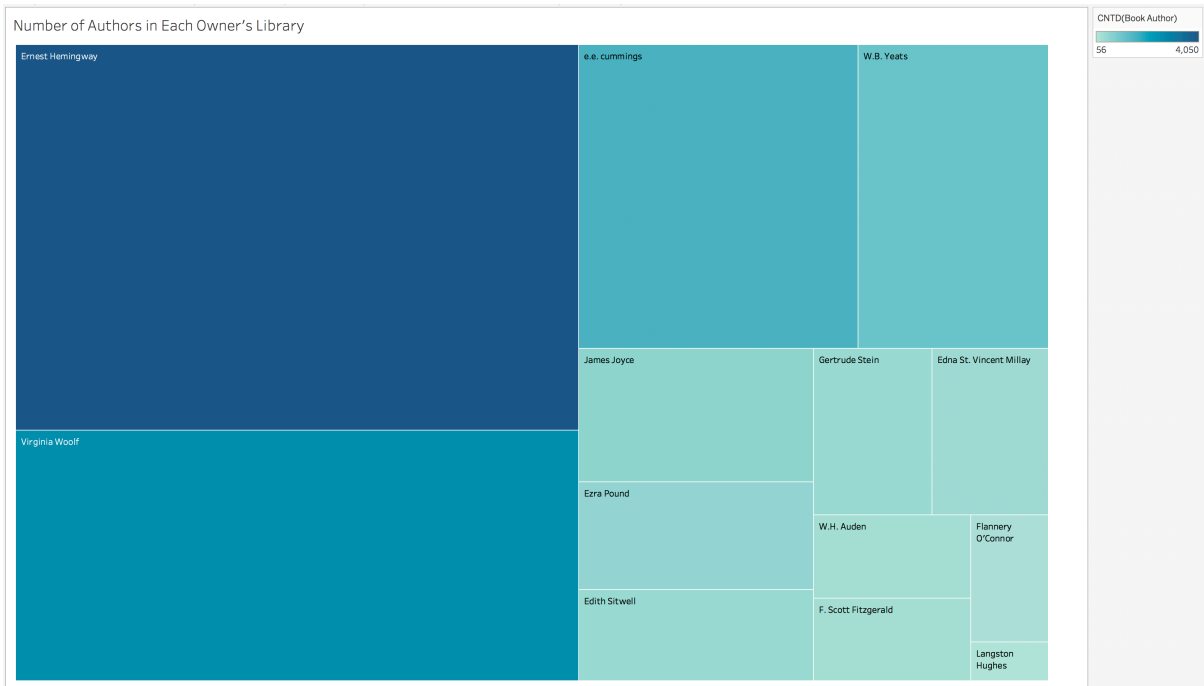


Figure 2 The number of unique author names present in each owner's library, with larger numbers indicated by a darker colour and bigger rectangular area.

The area charts communicate the range of publication years in each owners' library, and allows for comparison across authors.

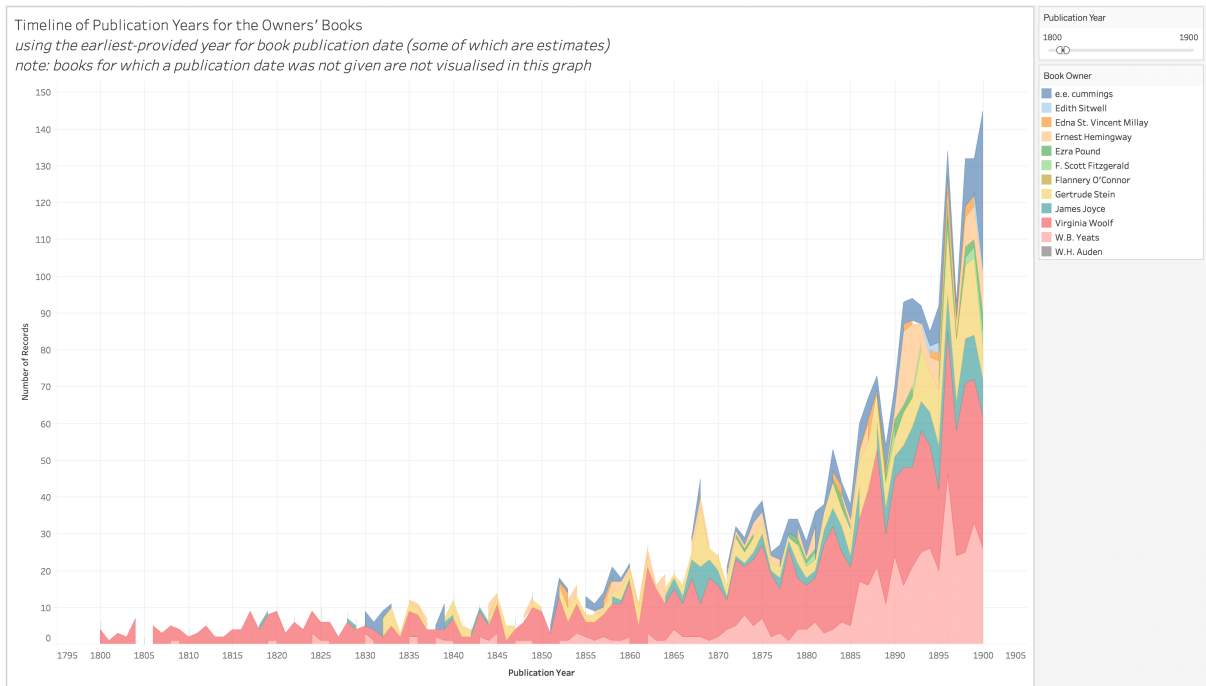
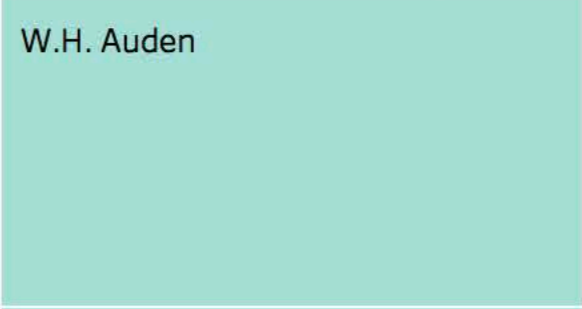
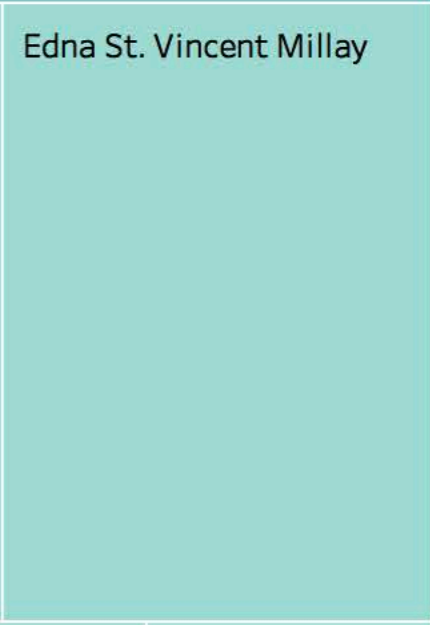
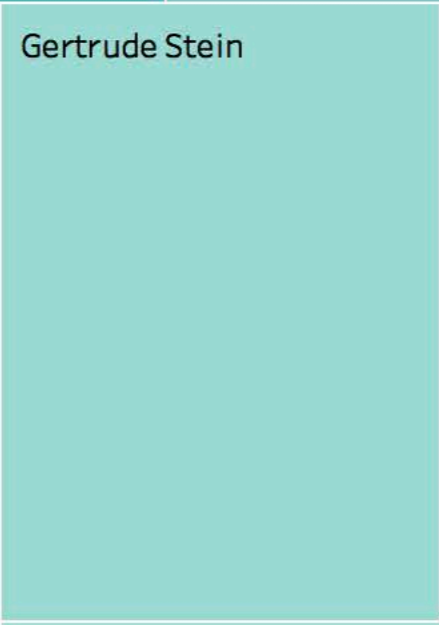
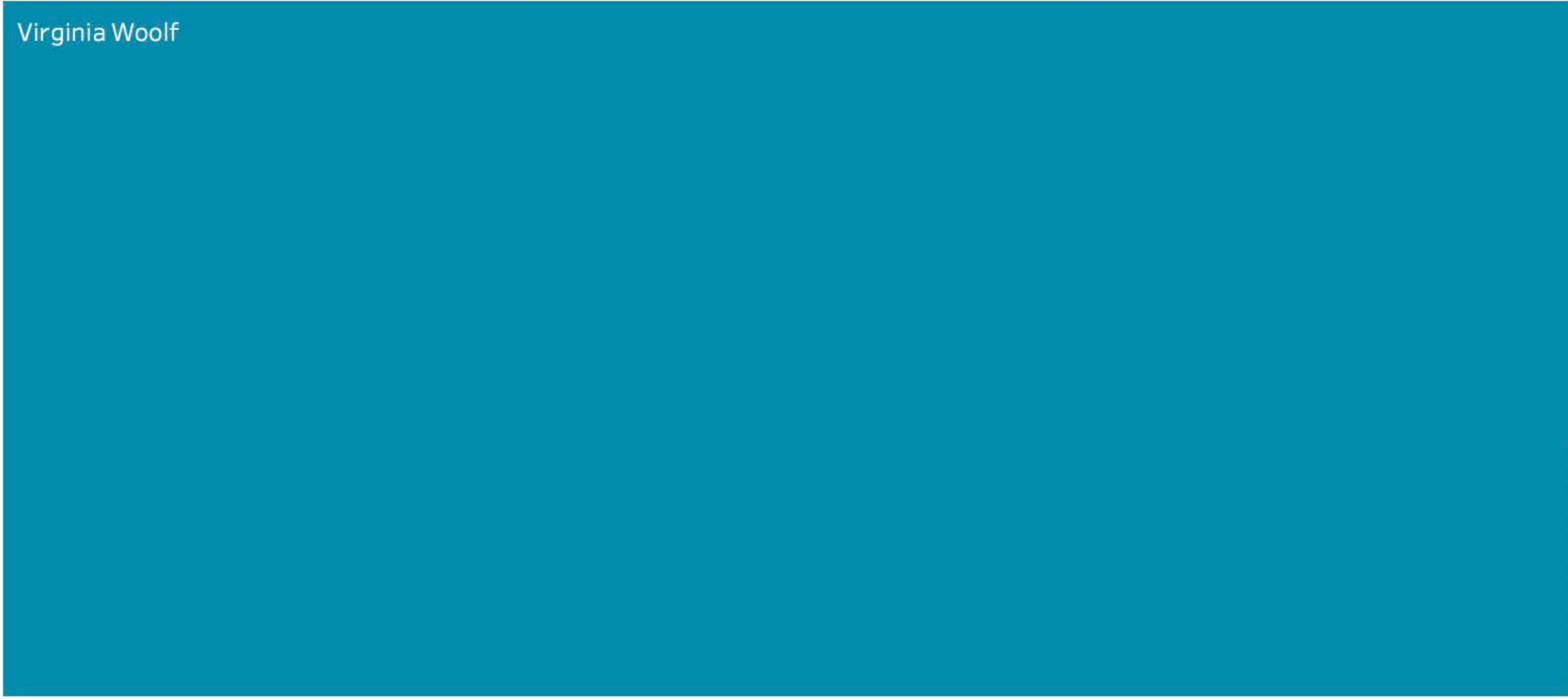
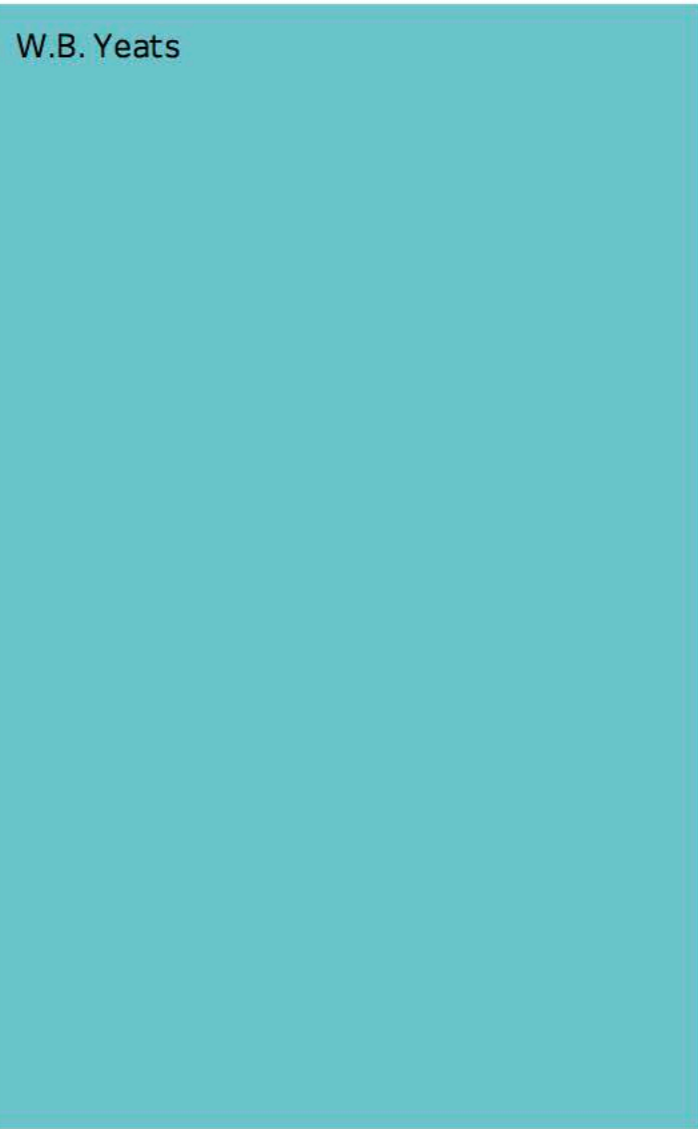
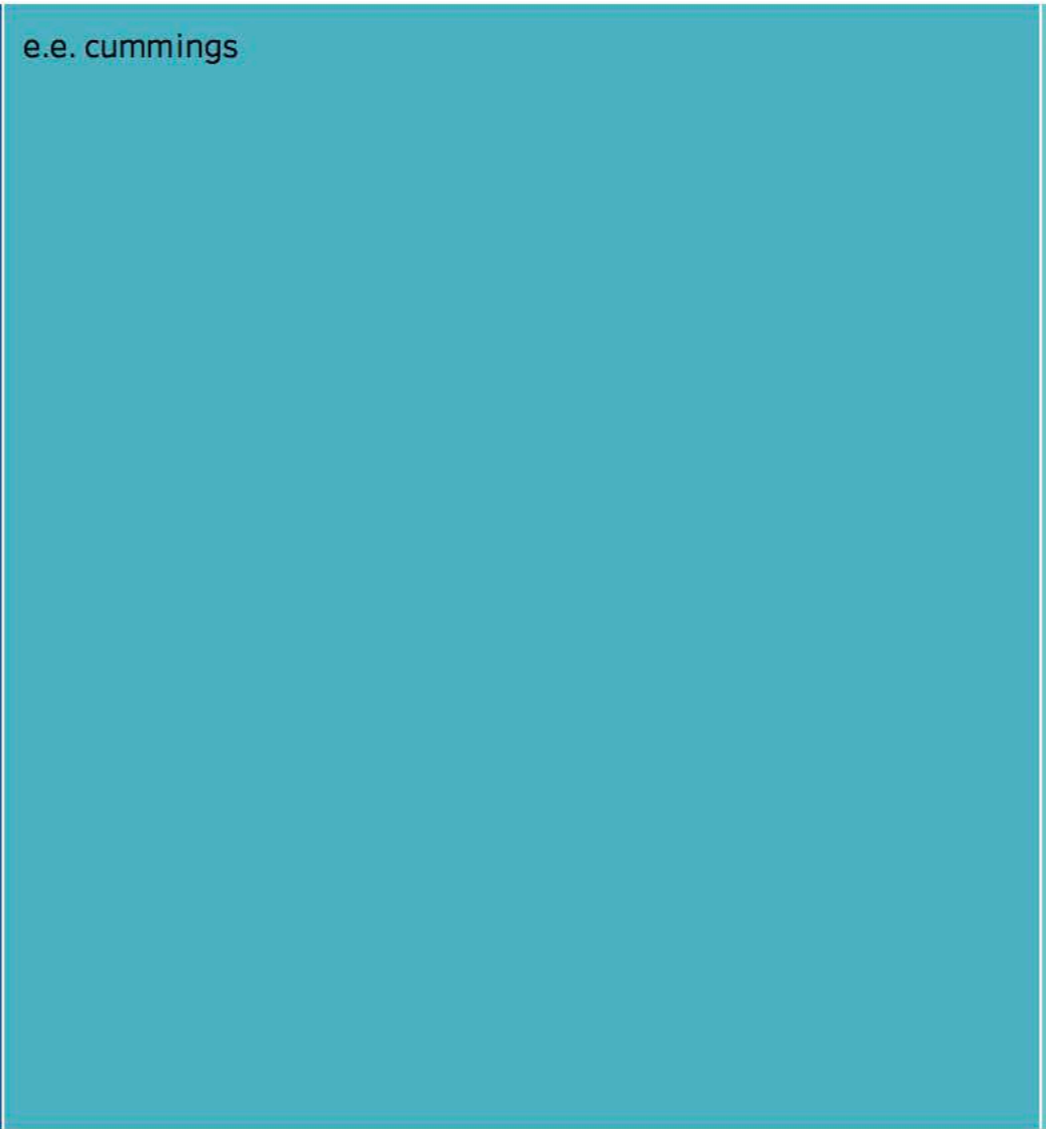
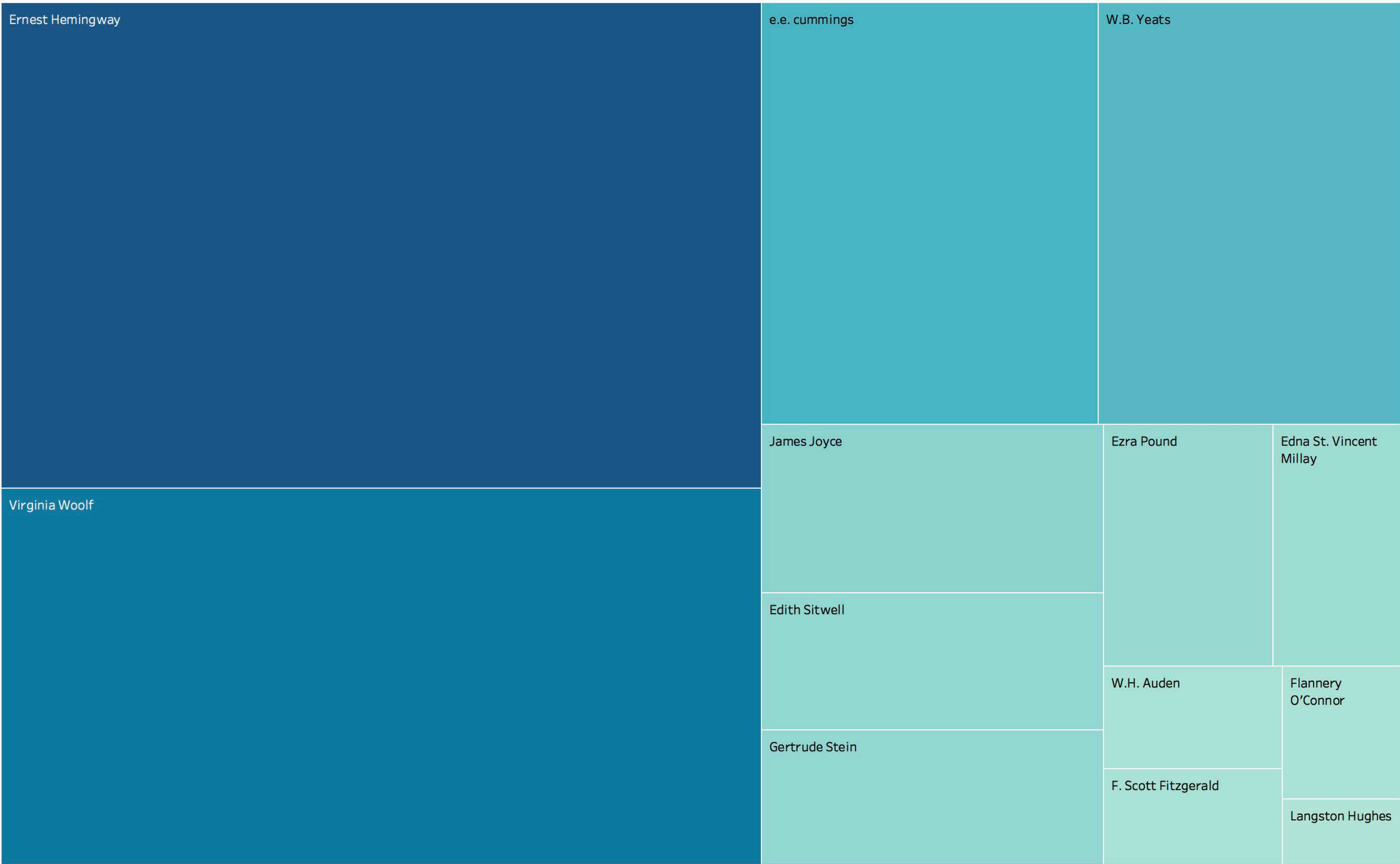


Figure 3 The number books published in each year that each owner had in their library, with the number of books on the y-axis, the year on the x-axis, and the owner indicated by the colour of the graphed area.

Number of Authors in Each Owner's Library



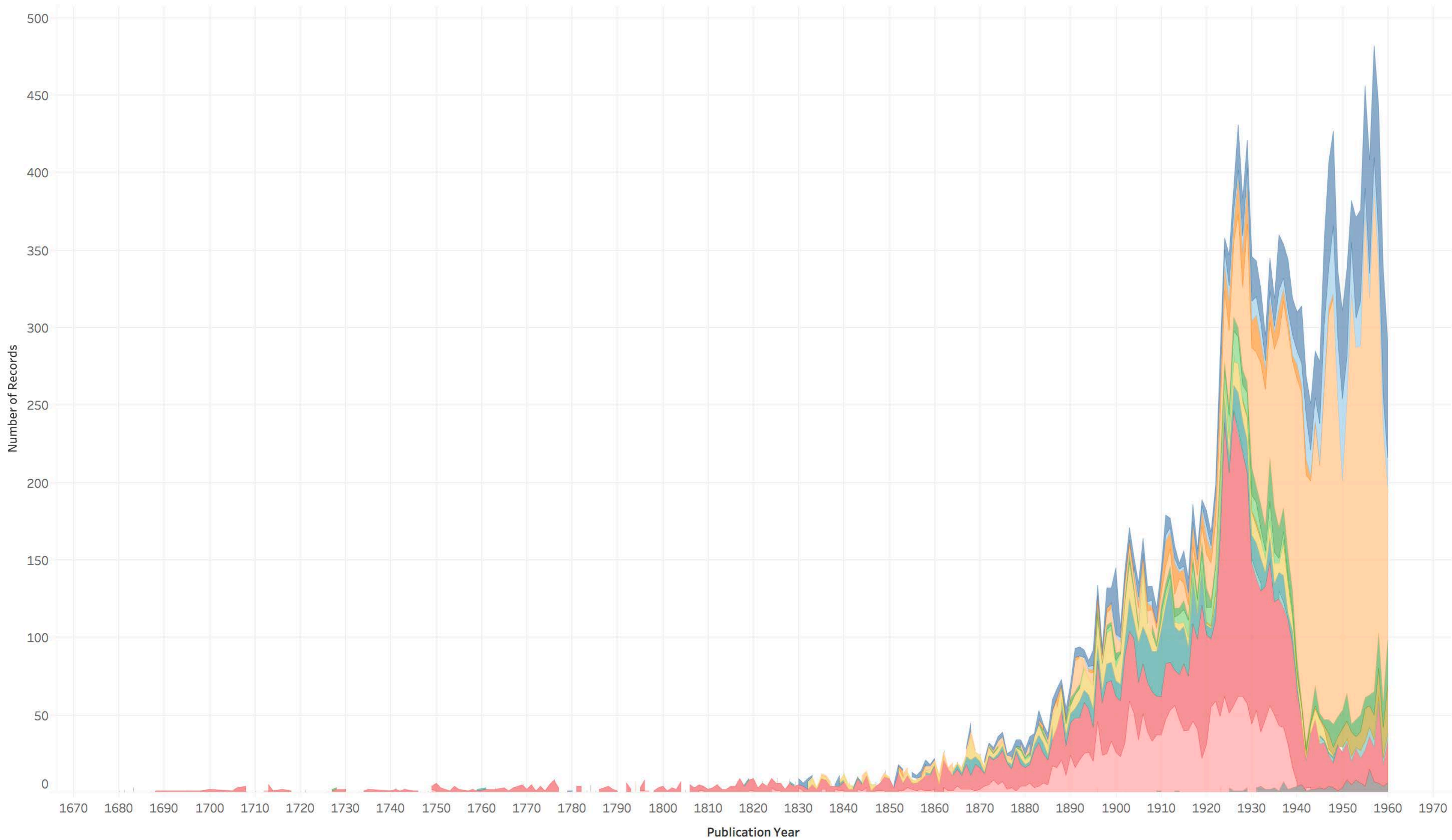
Number of Books in Each Owner's Library



Timeline of Publication Years for the Owners' Books

using the earliest-provided year for book publication date (some of which are estimates)

note: books for which a publication date was not given are not visualised in this graph



Publication Year

1680 1960

Book Owner

- e.e. cummings
- Edith Sitwell
- Edna St. Vincent Millay
- Ernest Hemingway
- Ezra Pound
- F. Scott Fitzgerald
- Flannery O'Connor
- Gertrude Stein
- James Joyce
- Langston Hughes
- Virginia Woolf
- W.B. Yeats
- W.H. Auden


Timeline of Publication Years for the Owners' Books

using the earliest-provided year for book publication date (some of which are estimates)

note: books for which a publication date was not given are not visualised in this graph

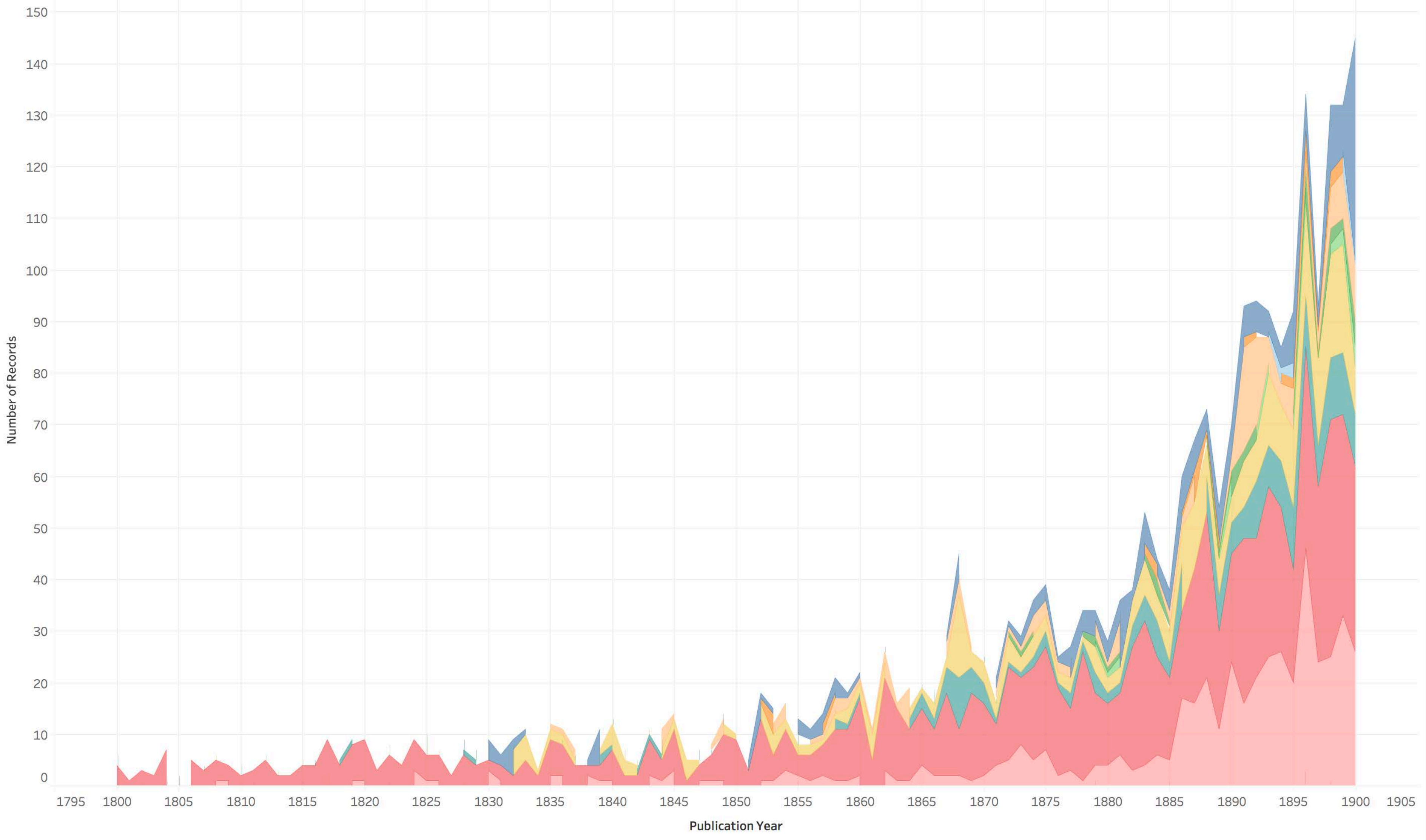
Publication Year

1800 1900



Book Owner

- e.e. cummings
- Edith Sitwell
- Edna St. Vincent Millay
- Ernest Hemingway
- Ezra Pound
- F. Scott Fitzgerald
- Flannery O'Connor
- Gertrude Stein
- James Joyce
- Virginia Woolf
- W.B. Yeats
- W.H. Auden



Timeline of Publication Years for the Owners' Books

using the earliest-provided year for book publication date (some of which are estimates)

note: books for which a publication date was not given are not visualised in this graph

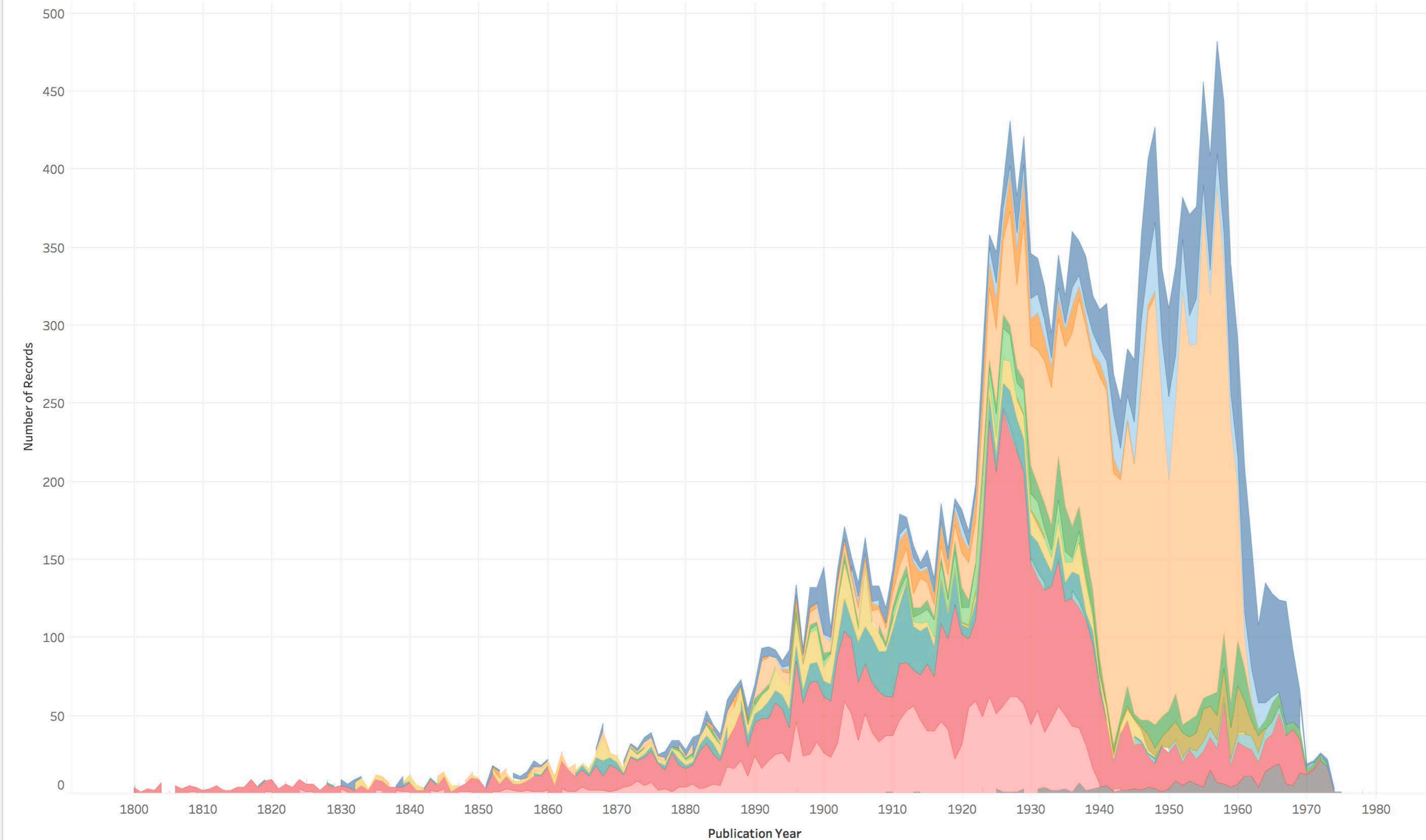
Publication Year

1800 1980

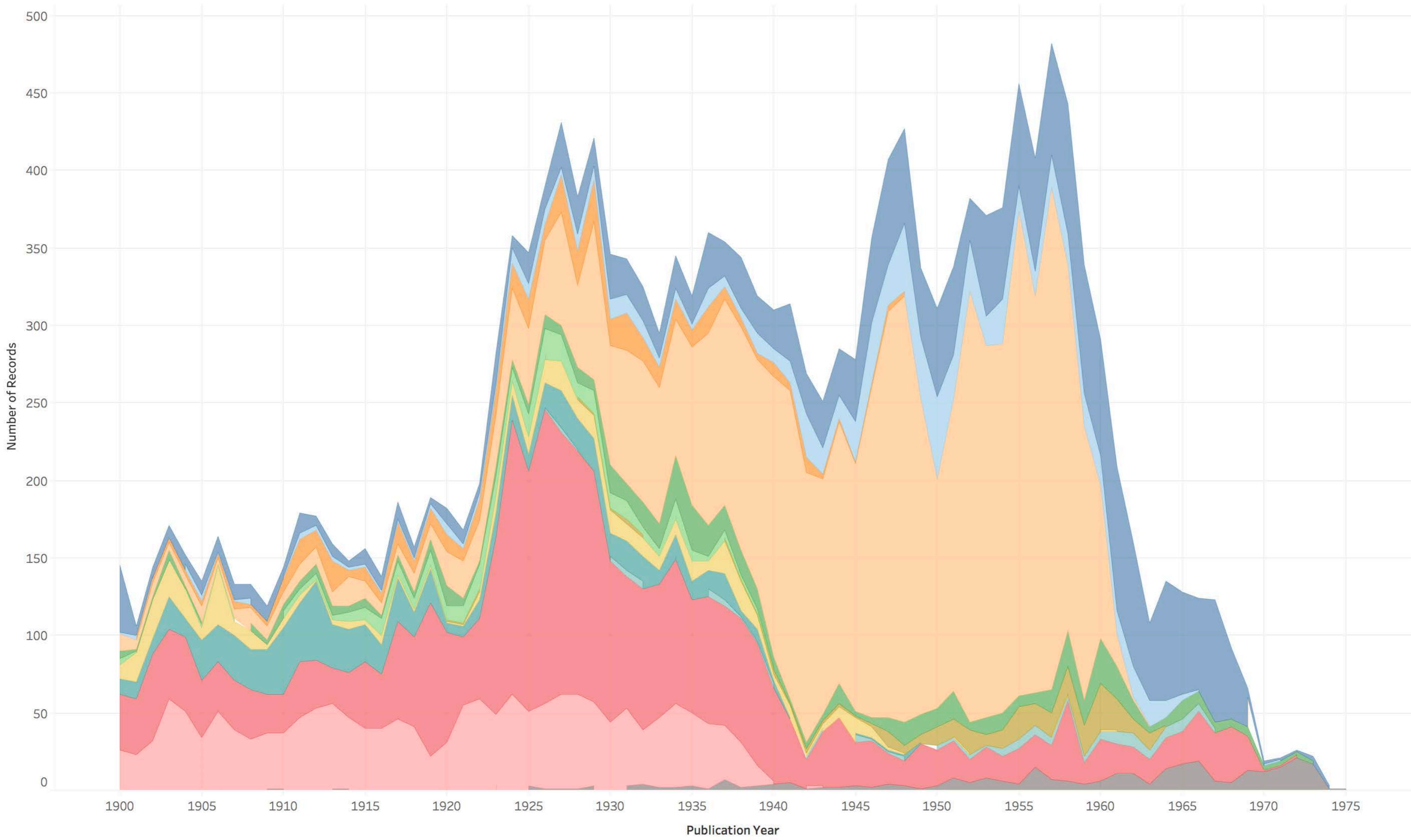


Book Owner

- e.e. cummings
- Edith Sitwell
- Edna St. Vincent Millay
- Ernest Hemingway
- Ezra Pound
- F. Scott Fitzgerald
- Flannery O'Connor
- Gertrude Stein
- James Joyce
- Langston Hughes
- Virginia Woolf
- W.B. Yeats
- W.H. Auden



Timeline of Publication Years for the Owners' Books
using the earliest-provided year for book publication date (some of which are estimates)
note: books for which a publication date was not given are not visualised in this graph



Publication Year

1900 1975

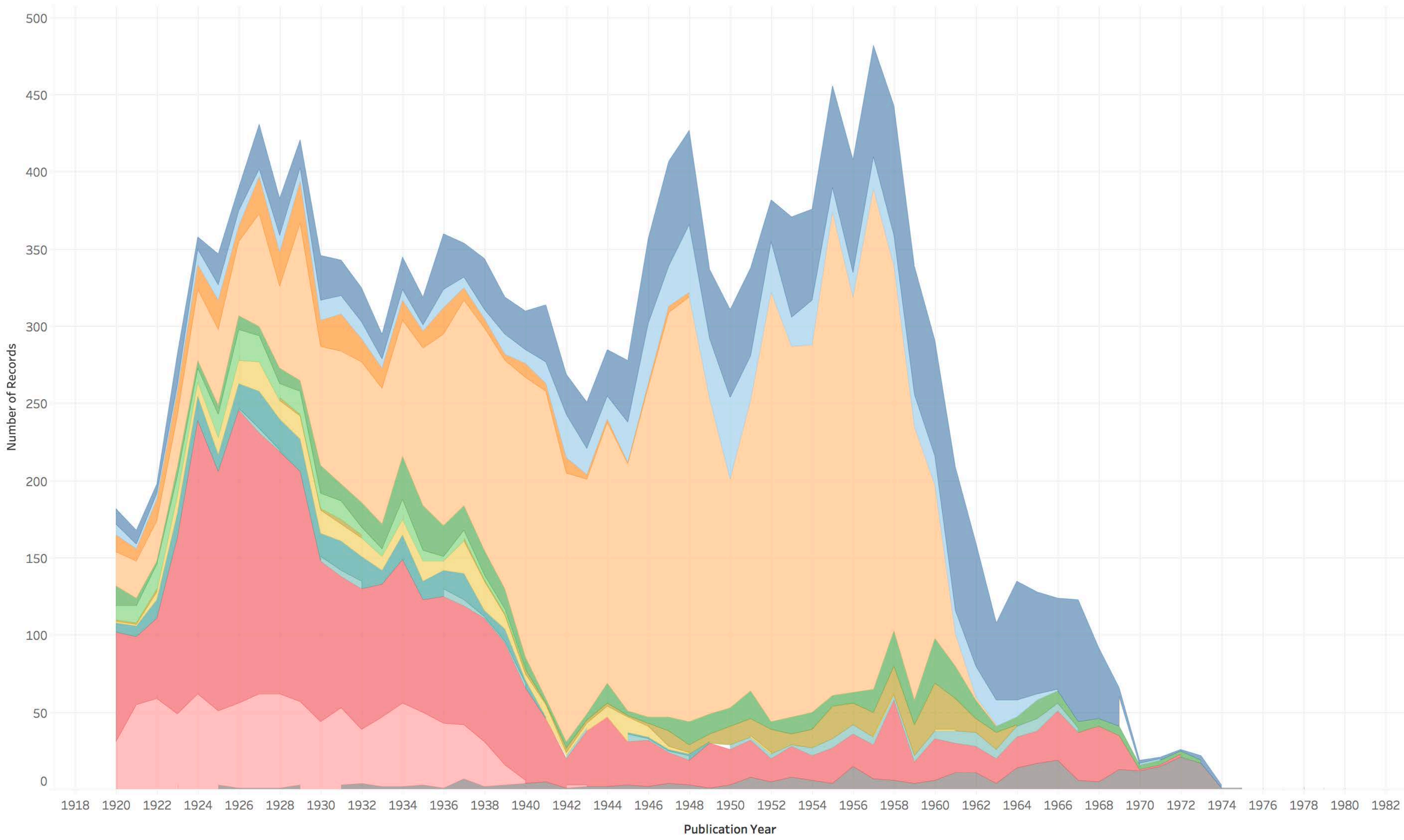
Book Owner

- e.e. cummings
- Edith Sitwell
- Edna St. Vincent Millay
- Ernest Hemingway
- Ezra Pound
- F. Scott Fitzgerald
- Flannery O'Connor
- Gertrude Stein
- James Joyce
- Langston Hughes
- Virginia Woolf
- W.B. Yeats
- W.H. Auden

Timeline of Publication Years for the Owners' Books

using the earliest-provided year for book publication date (some of which are estimates)

note: books for which a publication date was not given are not visualised in this graph



Publication Year

1920 1980

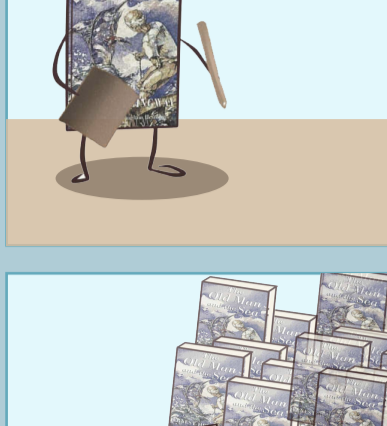
Book Owner

- e.e. cummings
- Edith Sitwell
- Edna St. Vincent Millay
- Ernest Hemingway
- Ezra Pound
- F. Scott Fitzgerald
- Flannery O'Connor
- Gertrude Stein
- James Joyce
- Langston Hughes
- Virginia Woolf
- W.B. Yeats
- W.H. Auden

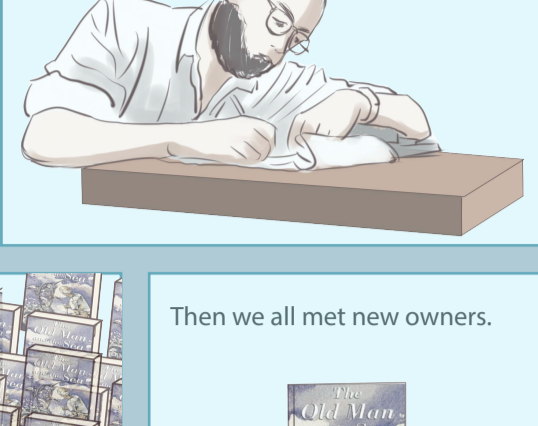
Ernest Hemingway and his library

-- What can we learn about people from their books?

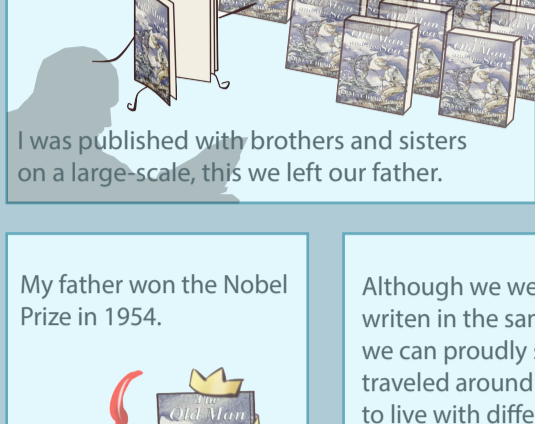
Qinru Shi, Yuchen Ye, Lucy Havens



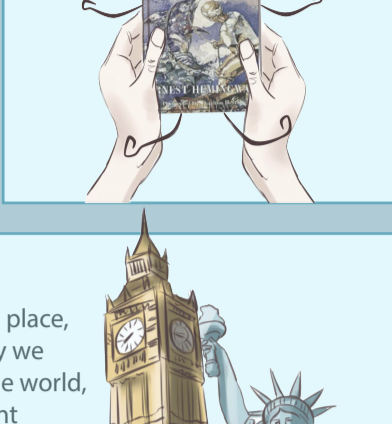
I am a book.



It took my father 8 weeks to let me see the world.



I was published with brothers and sisters on a large scale, this we left our father.



Then we all met new owners.



My father won the Nobel Prize in 1954.



Although we were written in the same place, we can proudly say we traveled around the world, to live with different owners.

The book is the old man and the sea, and Ernest Hemingway wrote the book in Key West.

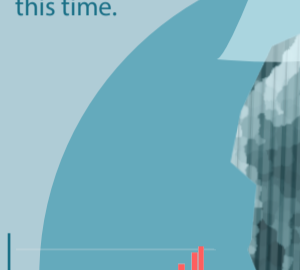
From high school in Chicago, Ernest Hemingway showed an interest and talent in writing.



His experience in World War I in Italy in 1918 lay the foundation for his writing style.



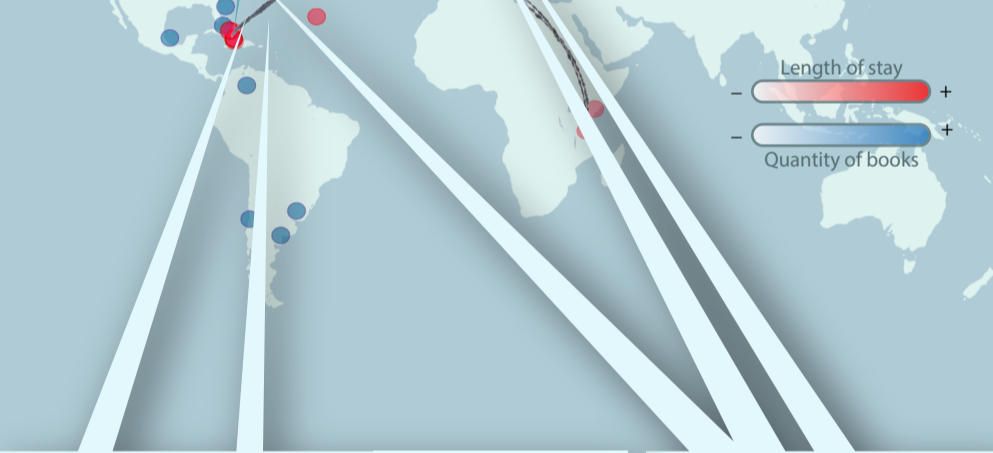
He published his first book Three Stories and Ten Poems in Paris in 1923. His book collection started to grow faster around this time.



L I F E
S P A N

Ernest Hemingway spent almost 20 years in Cuba, and during this period of time between 1940 and 1960, he wrote most of his work.

61 percent of his book collection was published during this time.

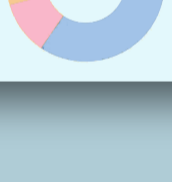


Finca Vigia is the largest bookcollection source and over 6000 thousand books in Ernest Hemingway's library were collected from there.



Key west is the second largest bookcollection source. This place was Ernest Hemingway's home in America. He lived and wrote here for more than 10 years. And the book the old man and the sea was written here.

Almost 80 percent of the books collected by Ernest Hemingway were published in New York (60%), London (11%), and Paris (8%).



There are many book authors in Hemingway's library.



The 2 most popular authors in Hemingway's library were Georges Simenon and Ezra Pound.



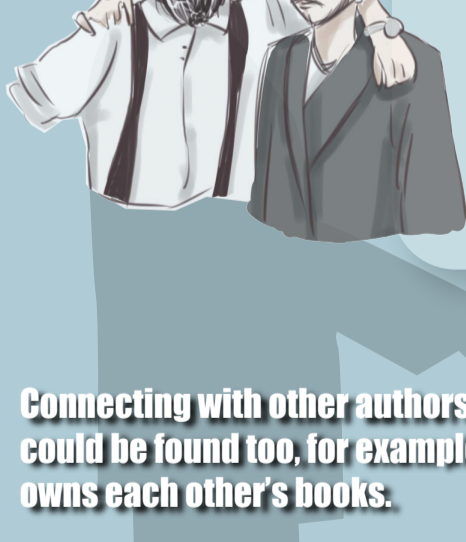
What relationships did Hemingway have with the authors in his library?

Georges Simenon(1903-1989)

Ernest Hemingway said that he admired and enjoyed Simenon's work, so it makes sense that Simenon's work is the most popular in Ernest Hemingway's library.



Hemingway owned **43** books written by Georges Simenon. Simenon is the **most popular** author in Hemingway's library.



Ezra Pound(1885-1972)

The **second most popular** author in Hemingway's library is Ezra Pound. He was a good friend of Hemingway's.

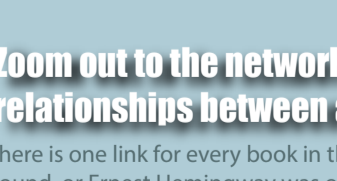


Hemingway owned **30** books written by Ezra Pound.

Connecting with other authors' libraries, some relationships could be found too, for example, Gertrude Stein and Hemingway owns each other's books.

Gertrude Stein (1874-1946)

Thus, Hemingway's writing style was affected by her.



She owned **3** books owned by Hemingway, and Hemingway owned **8** books written by her.



Zoom out to the network to see the relationships between authors.

There is one link for every book in the dataset for which Georges Simenon, Gertrude Stein, Ezra Pound, or Ernest Hemingway was owner or author.

