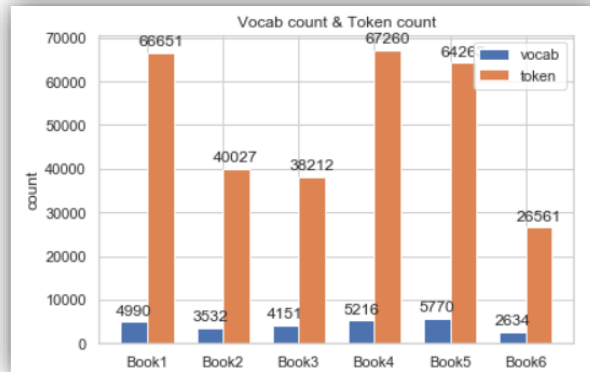
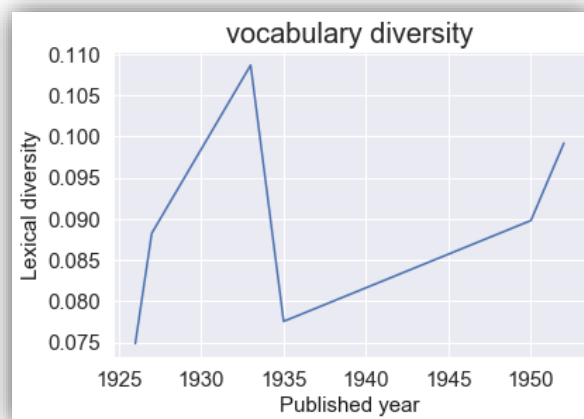


Data Visualization-Page1

Vocab count & Token count



Vocabulary diversity



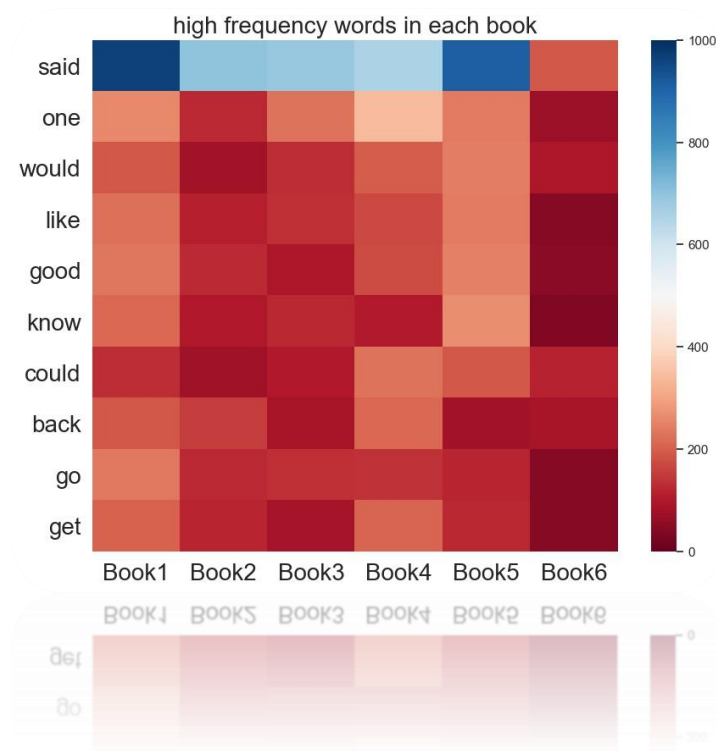
Wordclouds of Hemingway's book



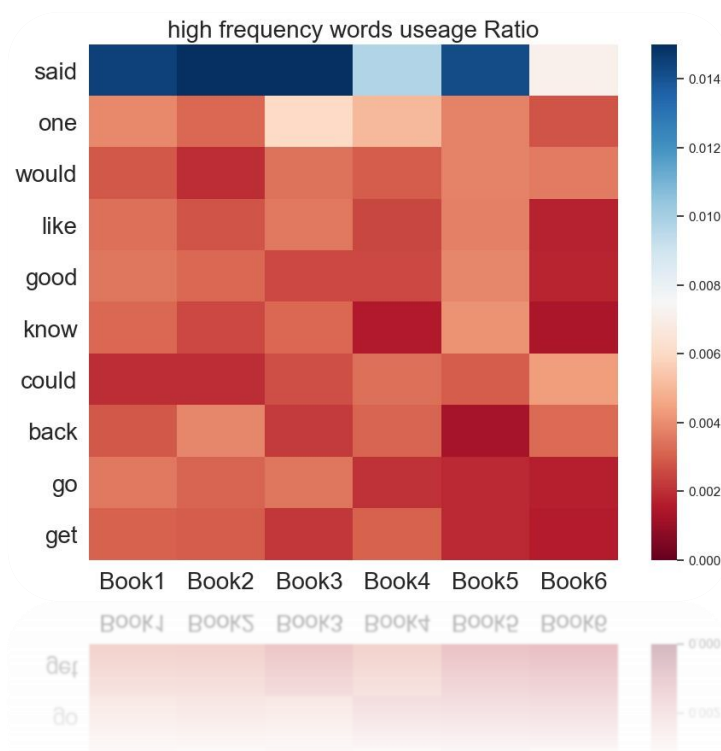
Louis - Kaiyang Liu

Data Visualization-Page2

High-frequency words in each book



High-frequency words usage ratio



A Summary of Hemingway's six novels

Louis, Kaiyang Liu

Ernest Miller Hemingway, who is one of the most famous novelists of the 20th century, created many works throughout his whole life. In this report, we mainly analyze six main articles of Hemingway using data visualization and give a summary of the insights of these data.

To begin with, we count the number of the words and vocabulary precisely of each book, which is shown in image 1, and we calculate their lexical diversity, which is shown in image 2. We can conclude that Book3: *Winner Take Nothing* is the highest diversity vocabulary usage book and Book1: *The Sun Also Rises* owns the lowest lexical diversity among those six novels. And the lexical diversity gradually increases since 1935.

Furthermore, analyzing high-frequency vocabulary that Hemingway used is also interesting. Given us six books of Hemingway, we first generate the texts into a single text and create a word cloud image to show the most commonly used vocabulary. We can see that "Said", "Know", "Like", "Get", "Back" and others verb are the bigger words in this image, those vocabulary shows that Hemingway's expression is usually declarative and active. What's more, adjective words like "Old", "Good" are also high-frequency word, those word are very basic vocabulary and easy to understand. To get deeper insights of Hemingway's word choice, we get top 10 high-frequency vocabularies that Hemingway used in the whole six novels text, which are shown in the axis of image 4&5, and we would like to see that whether those 10 vocabularies are also the top 10 words in each novel or not, or maybe one of the high-frequency vocabulary is extremely commonly used in one single writing and some words are less used in some writing. Combining with image 4 & 5, the heatmap shows that Book6: *The Old Man and Sea* are very special, its high-frequency words usage ratio is quite different from the others, the top 10-word ratio is relatively lower, which means the vocabulary Hemingway choose in this book are not as he used in the past, his vocabulary structure has undergone a slight change. The interesting thing is that it is precisely because of the book *The Old Man and the Sea* helped Hemingway won the Nobel Prize for Literature in 1954.

In conclusion, this report gives a summary of the data visualization based on Hemingway's six famous books. We analyze the word count, vocabulary count of each book, and shows their lexical diversity. Besides, we also get the point that Hemingway's expression are always statements and active, and the word choice are very simple. We also find that the high-frequency words usage ratio of *The Old Man and the Sea* is relatively lower and the word choice of this book is different from the others in some degree, which might be a factor that helps him won the Nobel Prize in Literature.

Louis - Kaiyang Liu

Appendix

<i>Book1</i>	<i>1926</i>	<i>Sun Also Rises</i>
<i>Book2</i>	<i>1927</i>	<i>Men Without Women</i>
<i>Book3</i>	<i>1933</i>	<i>Winner Take Nothing</i>
<i>Book4</i>	<i>1935</i>	<i>Green Hills of Africa</i>
<i>Book5</i>	<i>1950</i>	<i>Across the River and Into the Trees</i>
<i>Book6</i>	<i>1952</i>	<i>The old Man and the Sea</i>