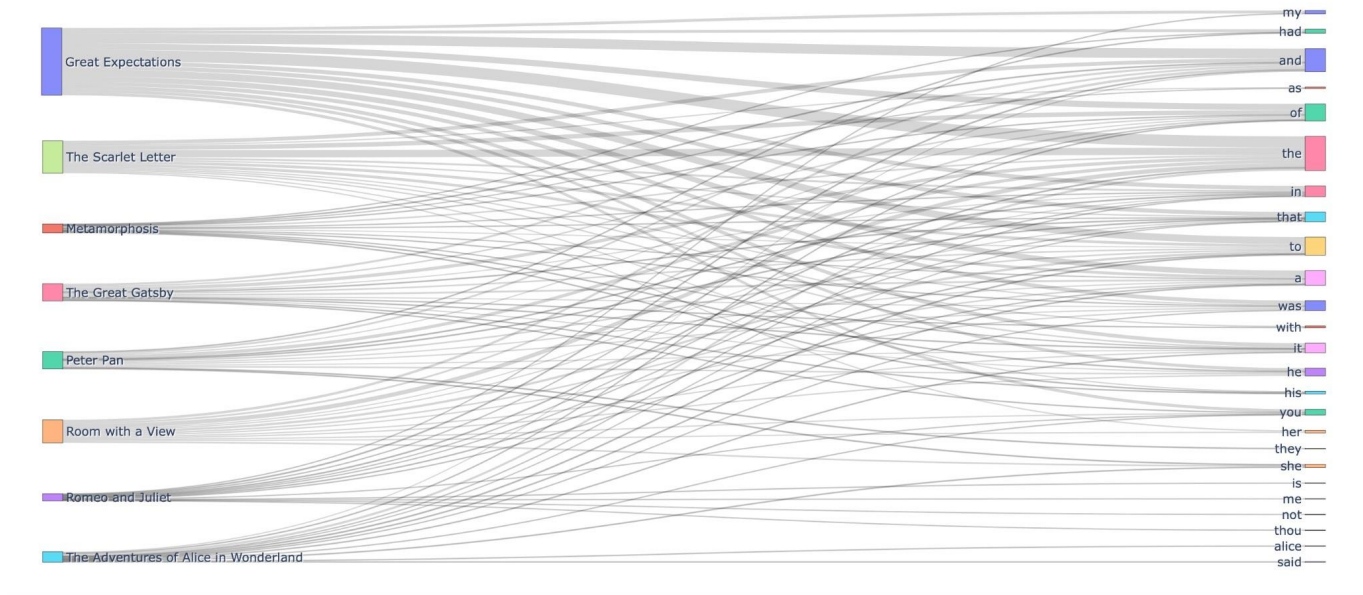# NLP Analysis of Popular Novels (Project Gutenberg)
Justin Guthrie, Ryan Huang, Julia Luo

Sankey Diagram (Left):

This Sankey diagram visualizes the novel on the left side and the word on the right side. Depending on the word count of that particular word in the novel determines how thick the connecting line is. The right side contains the 25 most common words out of all the novels, and the left side contains 8 of the 10 novels selected.

References:
"Alice's Adventures in Wonderland by Lewis Carroll." *Project Gutenberg*, 14 November 2022, https://www.gutenberg.org/ebooks/11.
"A Room With A View by E. M. Forster." *Project Gutenberg*, 1 May 2001, https://www.gutenberg.org/ebooks/2641.
"Great Expectations by Charles Dickens." *Project Gutenberg*, 1 July 1998, https://www.gutenberg.org/ebooks/1400.
"Metamorphosis by Franz Kafka." *Project Gutenberg*, 17 August 2005, https://www.gutenberg.org/ebooks/5200.
"Moby Dick; Or, The Whale by Herman Melville." *Project Gutenberg*, 1 July 2001, https://www.gutenberg.org/ebooks/2701.
"Peter Pan by J. M. Barrie." *Project Gutenberg*, 25 June 2008, https://www.gutenberg.org/ebooks/16
"Romeo and Juliet by William Shakespeare." *Project Gutenberg*, 11 May 2022, https://www.gutenberg.org/ebooks/1513.
"The Great Gatsby by F. Scott Fitzgerald." *Project Gutenberg*, 30 October 2021, https://www.gutenberg.org/ebooks/64317.
"The Scarlet Letter by Nathaniel Hawthorne." *Project Gutenberg*, 19 October 2022, https://www.gutenberg.org/ebooks/25344.

Polarity vs Subjectivity Scatter Plot (Right):

Plots polarity versus subjectivity for each of the novels, overlaying them onto the same plot. This measure of sentiment analysis examines how positive or negative the text is as the subjectivity of the text increases.

Polarity vs Subjectivity for all the Books