# INFO 4310 Final Project Report

Jiaxin Li (jl3579) & Yvette Hung (yh387)
Site link: https://pride-and-prejudice-sentiment.onrender.com/

## Goals & Motivation

1. Analyze the sentiment and emotional trajectory of Pride and Prejudice. By visualizing the sentiment scores for each chapter, we aim to identify patterns and trends in emotions as the story progresses, allowing readers to better understand the journey of the characters and the overall tone of the book.

2. Visualize character relationships quantitatively. By quantifying and displaying the relationships between characters visually, we offer readers an overview of the connections and interactions among characters. This could help readers identify key relationships and understand their significance in the story.

3. Facilitate further exploration and discussion of the novel. An interactive platform for examining the book might encourage readers to delve deeper into its themes and characters, stimulating further discussion and analysis.

## Audience & Use Cases

1. Analyze sentiment trends for a closer understanding: an instructor teaching Pride and Prejudice in class could use the sentiment analysis visualization to highlight the emotional fluctuations in the novel. By discussing these trends with students, teachers can help them better understand the significance of events and character development throughout the story.

2. Help new readers familiarize themselves with the novel: someone who hasn't read the book but plans to do so can explore the visualizations to get a sense of the mood and character network. This can provide them with a foundation to appreciate the story as they read, making the experience more engaging.

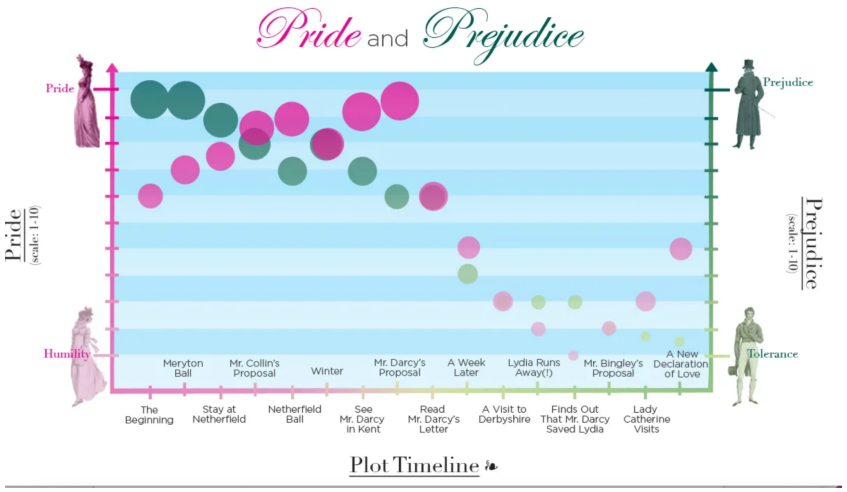
3. Compare and contrast different novels: a student conducting an analysis of multiple novels could use the visualizations as a baseline to understand the emotional trajectory of the story. If this sort of representation is applied to other books as well, it can help them identify similarities and differences across novels.

## Related Materials

- https://p-mckenzie.github.io/2018/01/11/Jane-Austen/: a previous analysis of three Jane Austen novels, including Pride and Prejudice, was an important inspiration for us

in terms of the possibilities available for visualizing the contents of a large body of text.

- https://www.vox.com/2015/1/28/7922617/pride-and-prejudice-charts: an example of a bubble chart helped us decide to create a similar visualization - labeling important events in the book seemed like a particularly intuitive and useful thing to do.




# Data Source & Processing

The raw text file of Pride and Prejudice was downloaded from Project Gutenberg at https://www.gutenberg.org/ebooks/1342. Irrelevant parts of the file, such as the introduction and metadata, were trimmed, and the text separated into 61 chapters in total.

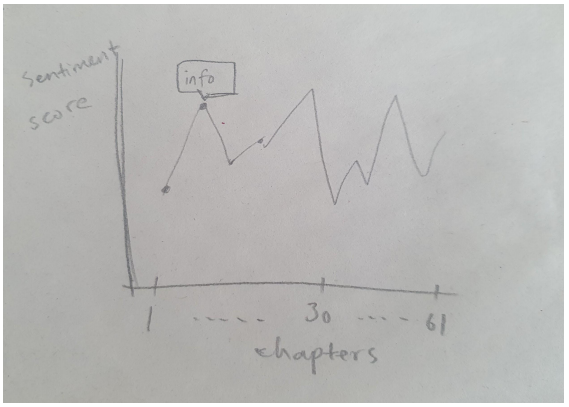
For the sentiment portion, we used NLTK's model to calculate the sentiment score for each sentence and take the average by chapter, resulting in a score for each chapter.

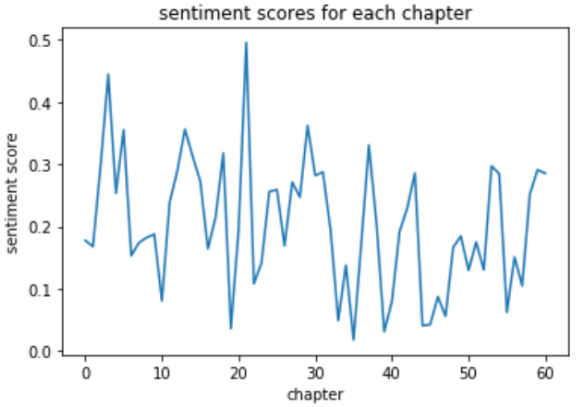
To quantify character relationships, we decided to take the eight most important characters in the book and measure co-occurrences for each pair of characters. If two characters are named within three sentences of each other, it was counted as an instance of co-occurrence. We recognize that this approach may result in inaccuracies such as some characters not being counted properly or ambiguous context, but for the scope of this project, we think this is a reasonable method.

Finally, to get the emotional intensity for each of the four main characters at each important event, we averaged the sentiment score for all lines mentioning the character throughout the corresponding chapter. This approach has similar tradeoffs to our method of measuring character co-occurrences.

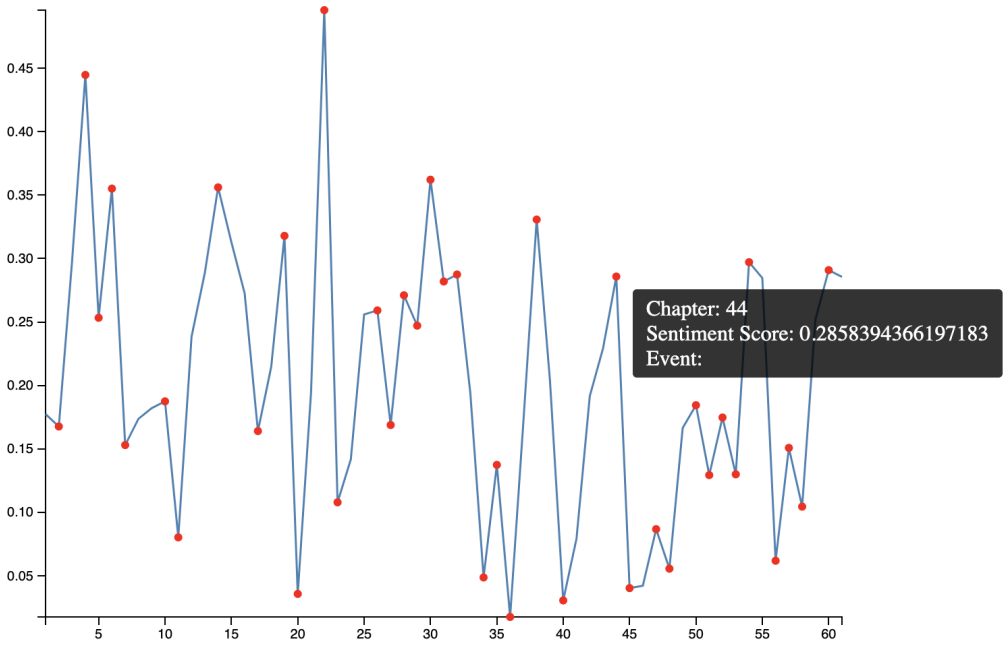
# Design Process

Our idea for the project started with the sentiment line chart and chord diagram of character relationships. For the line chart, we wanted a simpler overview of emotions throughout the entire novel, with hover interactivity showing the sentiment scores. Shown below is a rough sketch and a simple chart generated during data processing with pyplot.

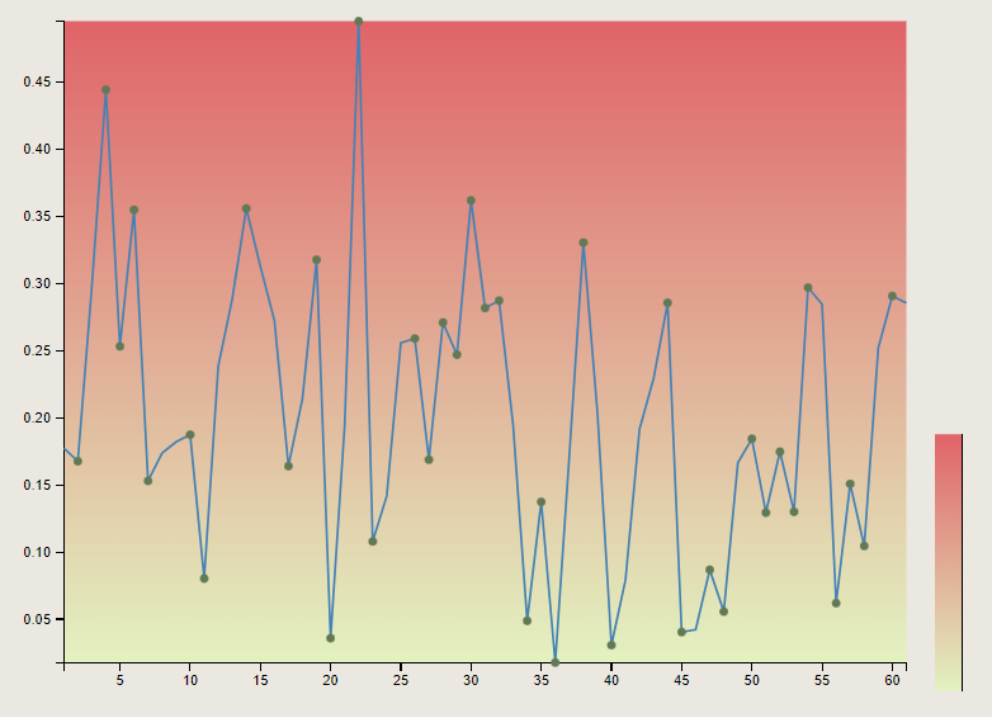


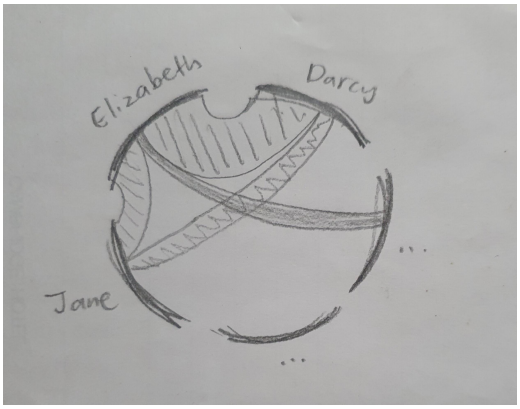

Below is a preliminary version of the line chart in d3.



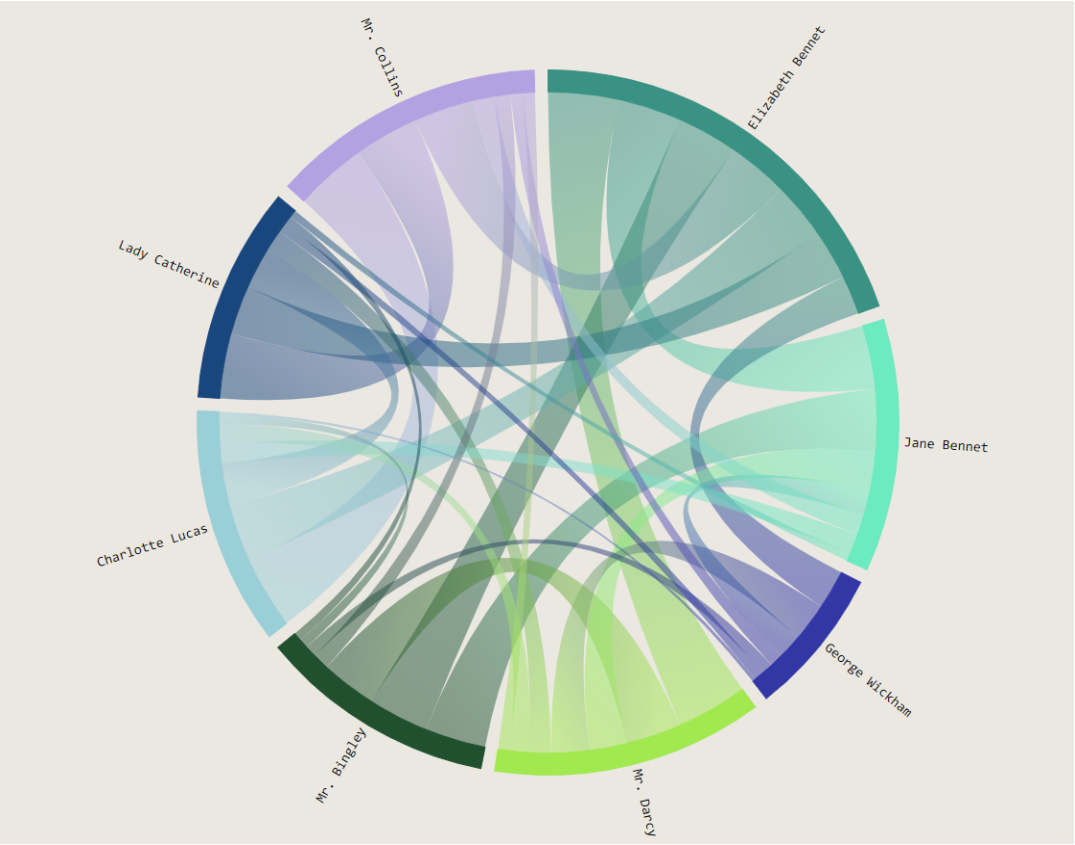

Our line chart of sentiment scores for each chapter features a color scale transitioning from yellow to red. The yellow background indicates low or sad emotions, while the red background represents strong, intense emotions. A legend is provided to accompany the chart, clarifying the color meanings for easy interpretation. This visual presentation allows users to quickly understand the emotional trajectory of the characters as the story unfolds.



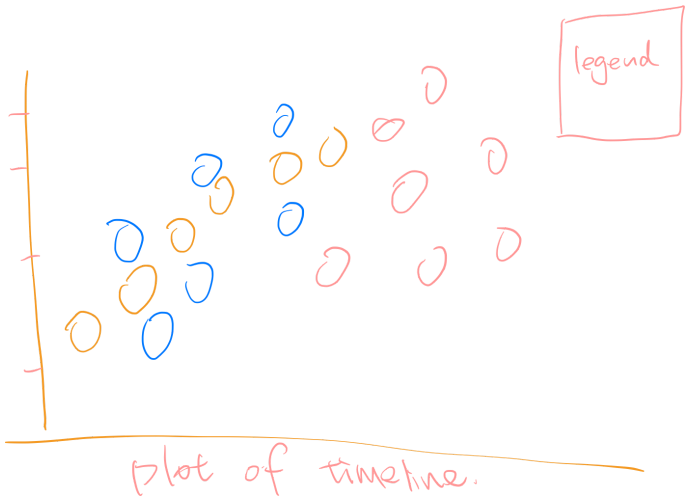

Next, we imagined a chord diagram as an intuitive way to visualize the strength of relationships between characters. We also briefly considered a Sankey diagram, as it also shows a sense of flow, but decided that the circular shape of a chord diagram would better show the interconnectedness of all eight characters that we chose to focus on.

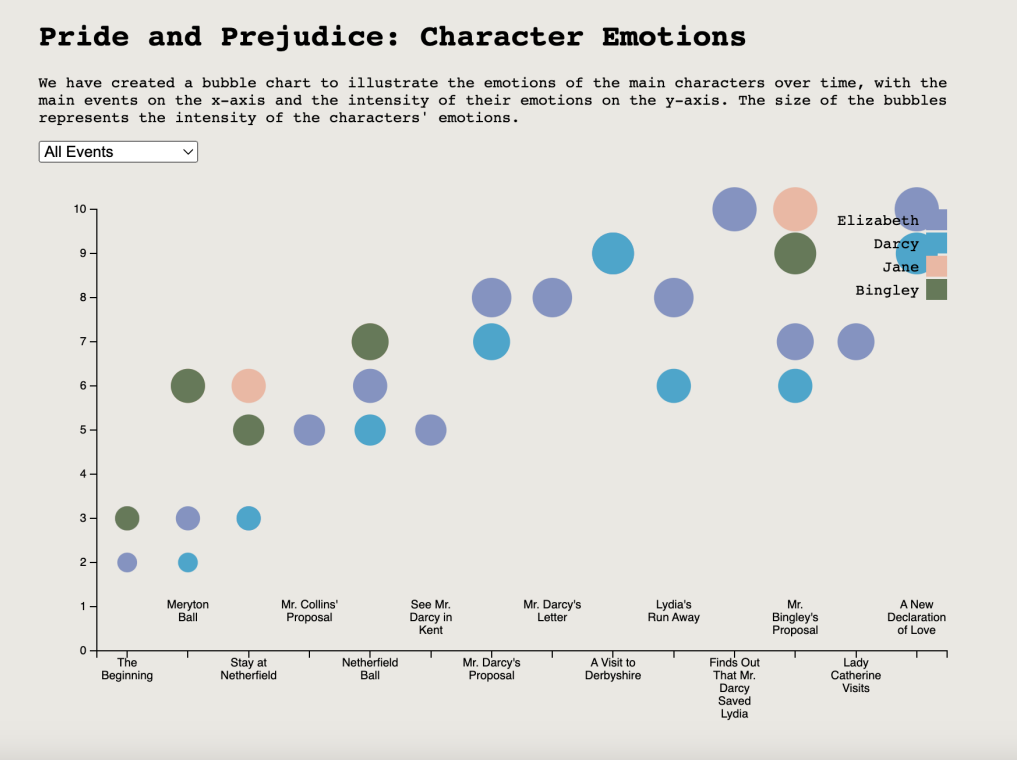



Since there are eight characters, we had to consider the problem of color choice for each character. Initially, we used generated colors as a placeholder, but at this stage we also thought about unification with our other charts, i.e., using the same color for the same character across visualizations.



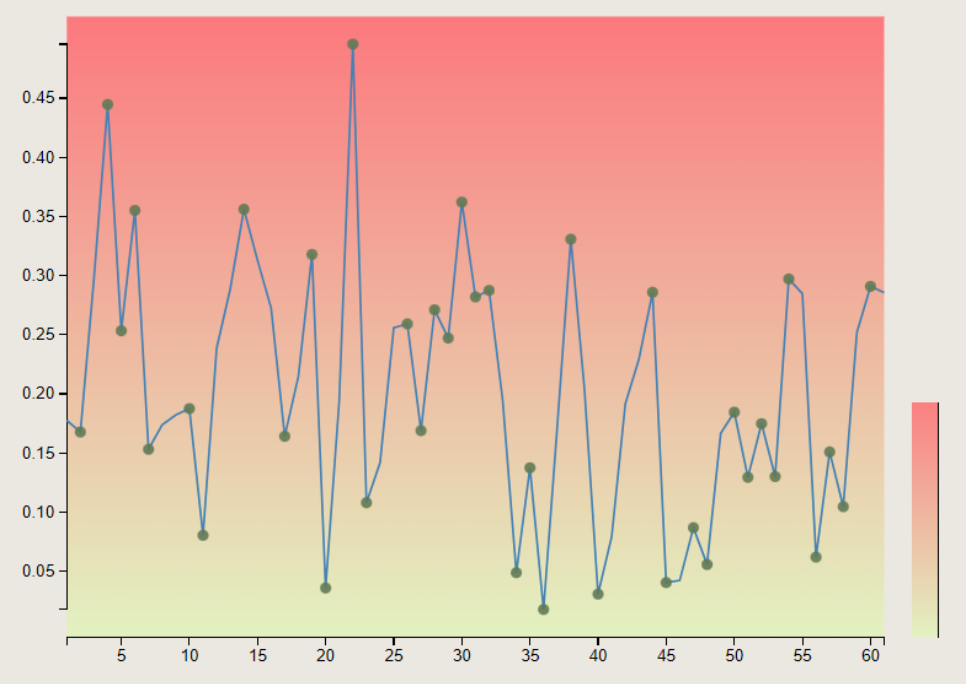

Finally, we considered creating a bubble chart based on the previous two charts that showcases the emotions of the four main characters in Pride and Prejudice over time. The goal is to understand whether they are happy, sad, or angry throughout the book and capture the overall tone of their thoughts. However, as we thought it was redundant for the bubble size and the y-axis to convey the same information, we decided on a regular scatter plot instead.



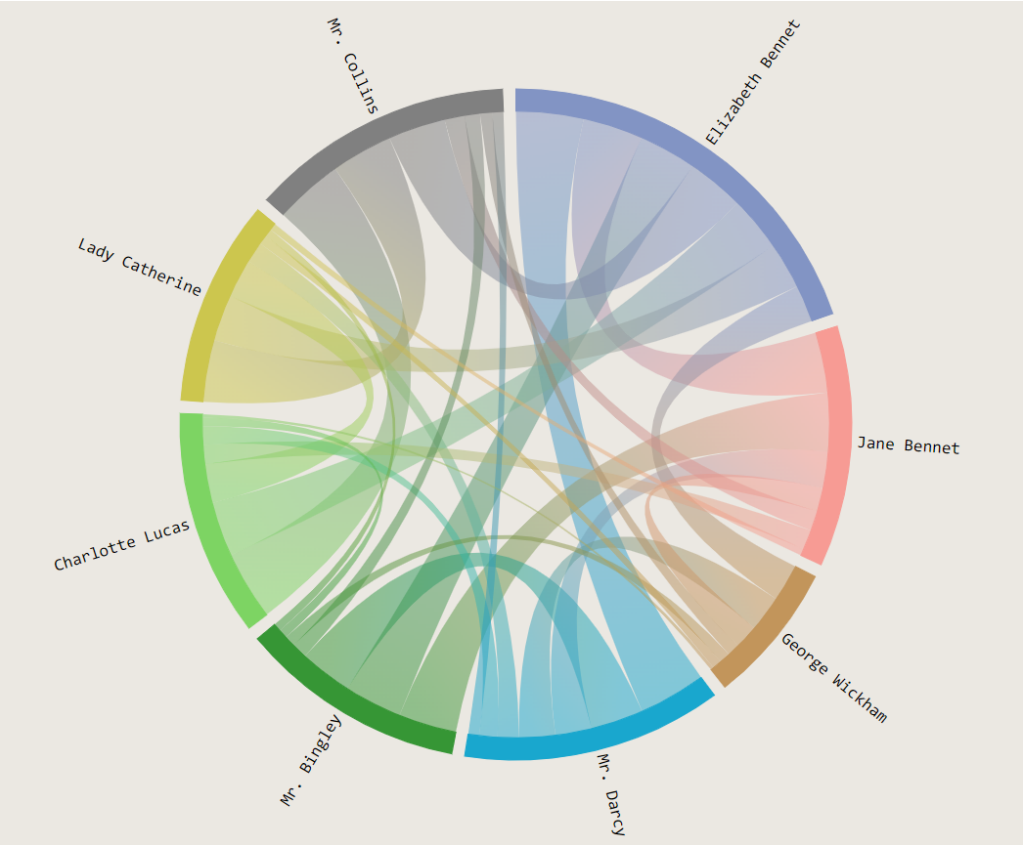



# Final Design & Implementation

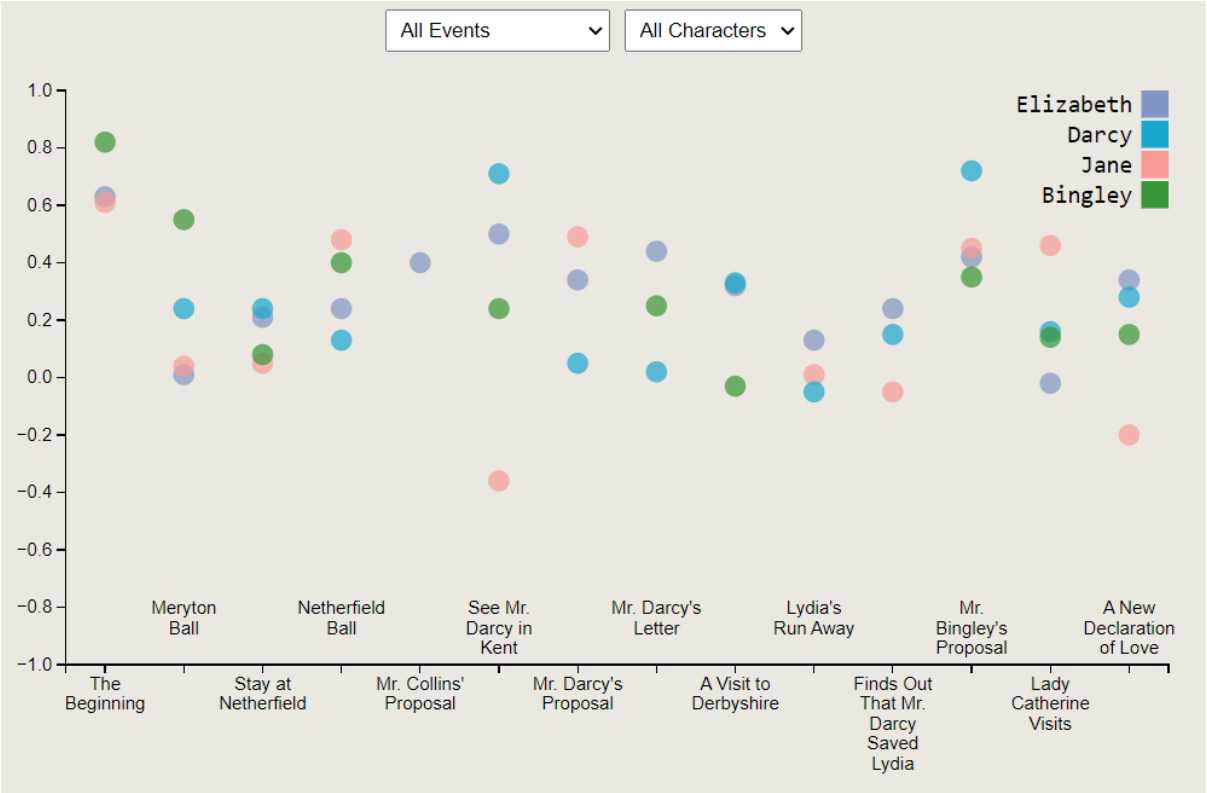
In our line chart, the color gradient signifies change in emotional intensity. This acts as a visual helper to the y-axis, which shows the sentiment scores. The top gradient color is adjusted so that the lines and circles contrast more with the background. Each circle represents a local high or low point, which avoids overwhelming users with information from all 61 chapters. Interactivity is implemented by displaying a tooltip on hover.



In our chord diagram, we use an intuitive flow to represent the relationships between characters, with the width of each ribbon showing how strongly two characters are related in comparison to other pairs. The color for each character was chosen manually for distinctiveness and loosely corresponds to the character's traits. For instance, Mr. Darcy is somewhat aloof, while Mr. Bingley is more peaceful or stable with regard to emotions. Interactivity is implemented by highlighting a character's relationships on hover.




Our final visualization was modified from the initial bubble chart into a scatter plot, with circles representing characters and the y-axis showing the intensity of their emotions, while the x-axis is a timeline of important events throughout the novel. Each color corresponds to a character, and as part of unifying our color scheme, we used the same colors for the four main characters as in the chord diagram. In addition to hover functionality, we also incorporated two filter dropdowns. Filtering by event type (e.g., social and confrontation) allows users to observe general emotional intensity, while filtering by character shows the trajectory of that character as the story progresses.




The suggestions we received were useful in helping us improve readability and choose a more suitable color scheme. For example, label sizes are increased throughout our visualizations.

# Team Member Contributions

- Jiaxin: brainstorming, project & milestone reports, presentation, line chart, scatter plot
- Yvette: brainstorming, project & milestone reports, presentation, chord diagram, data processing