Poetixment

Sentiment Analysis for Multi-stanza Poetry

### Author: Nathan Parker

Author Contributions:

This framework was developed by the author, with blueprint starter code provided by Dr. John Rachlin of Northeastern University.

Overview:

The Poextiment framework functions by loading files that contain poems. The framework has built-in support for loading .txt and .json files, but a user may input their own parser if they so wish; it just needs to result in a string that contains the poem. Each poem is then processed, determining word counts, the sentiment score of each line, and the sentiment score of every stanza. The framework then has built-in visualizations for the previously mentioned data.

Ultimately, a case study of the sentiment in some of Edgar Allan Poe's most famous poems was decided as the focus of this framework, as Poe has a reputation for bleak and depressing pathos. His poetry was transcribed from https://poestories.com/poetry.php and www.poetryfoundation.org/poets/edgar-allan-poe into .txt files.
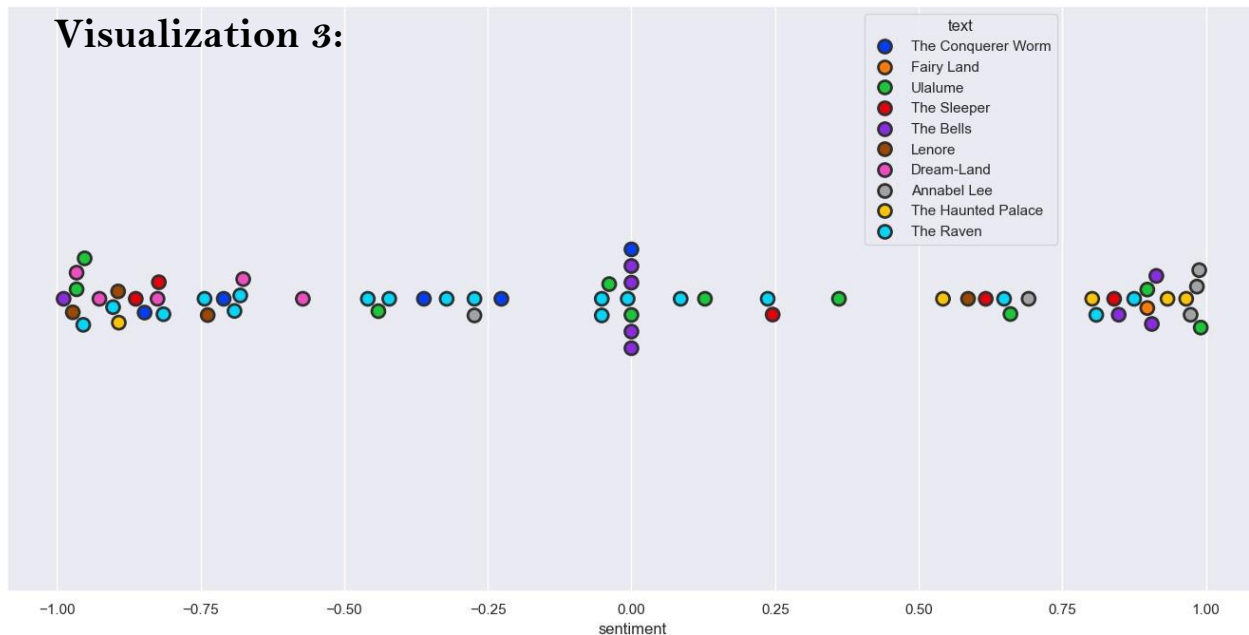
Visualization 1:
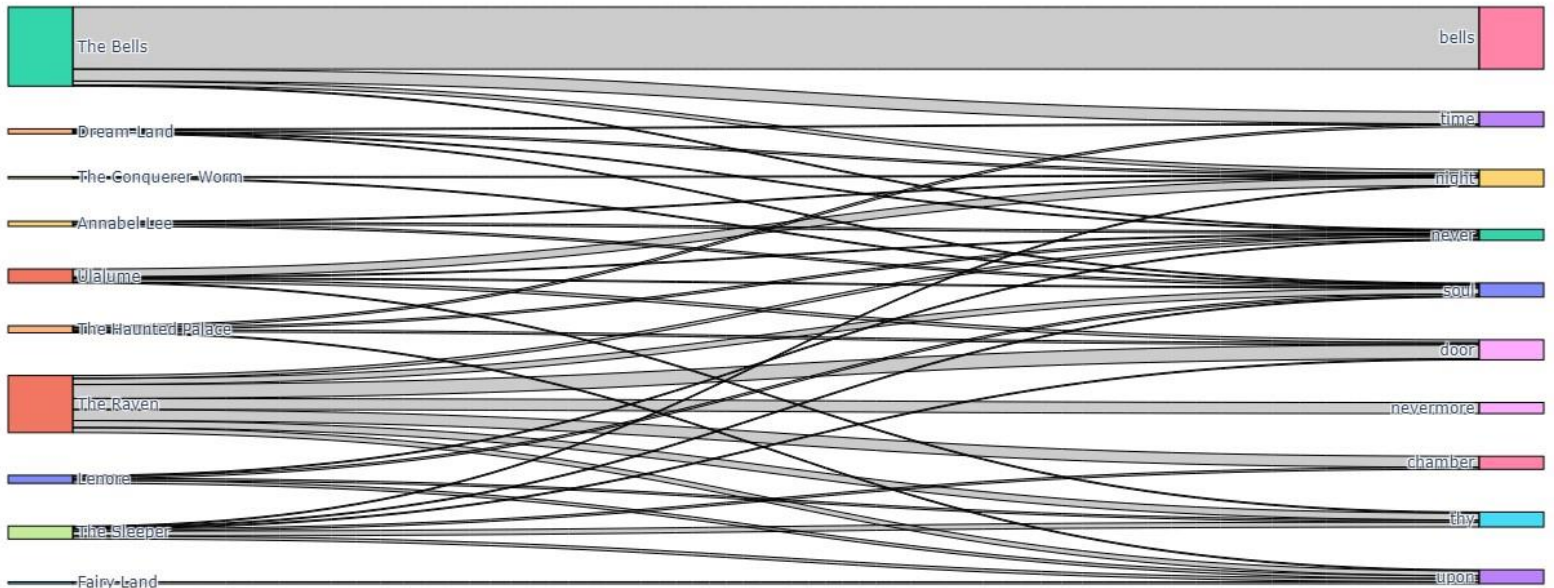
The first visualization is a Sankey diagram from the loaded poems to different words, where the width of each link represents the count of how often a word appears in that poem. A user may either input a list of words that the poems may map to, or they can select the k most common words between all poems. In the above visualization, the poems map to the ten most common words (those being, in no particular order, "bells", "time", "night", "never", "soul", "door", "nevermore", "chamber", "thy", and "upon"). These common words invoke dark imagery.

Visualization 2:

The second visualization is a subplot of histograms, where each subplot contains the distribution of line sentiment values. Every line was assigned an emotional value between -1 and 1, where -1 is overall negative and 1 is overall positive. This visualization then displays the count of lines that fall in 0.1 increments of

sentiment between -1 and 1. A key detail about this visualization is many lines tend to fall under neutral values (as the diction choices of certain lines hold no sentiment without context), so neutral values were removed for ease of viewing. The option to display neutral values is included in the framework. Despite Poe's reputation as a depressing figure, many of his famous poems have very positive lines, resulting in an overall even spread across his different works.

Visualization 3:

The third visualization is a swarm plot, where each mark represents a stanza from one of Poe's poems. Once again, the values are arranged from -1 to 1 to represent the sentiment value of the stanza. This visualization allows users to compare the stanzas from various poetry works, while also having distinct colors to see the sentiment of all the stanzas in a single work. Also once again, there is an even distribution of sentiment across Poe's various stanzas, demonstrating he also wrote hopeful and positive poems in addition to his infamous dark poems.

Conclusion:

Despite his reputation as a gloomy figure, the Poextiment framework demonstrates that Edgar Allan Poe's works often contained positive prose as well. His work often centered around themes of loss and grief, which makes it easy to forget that something must be highly valued before it can be grieved. It is reductive to view Poe's works as sinister and gloomy, without appreciating the love he often demonstrates through his descriptions.