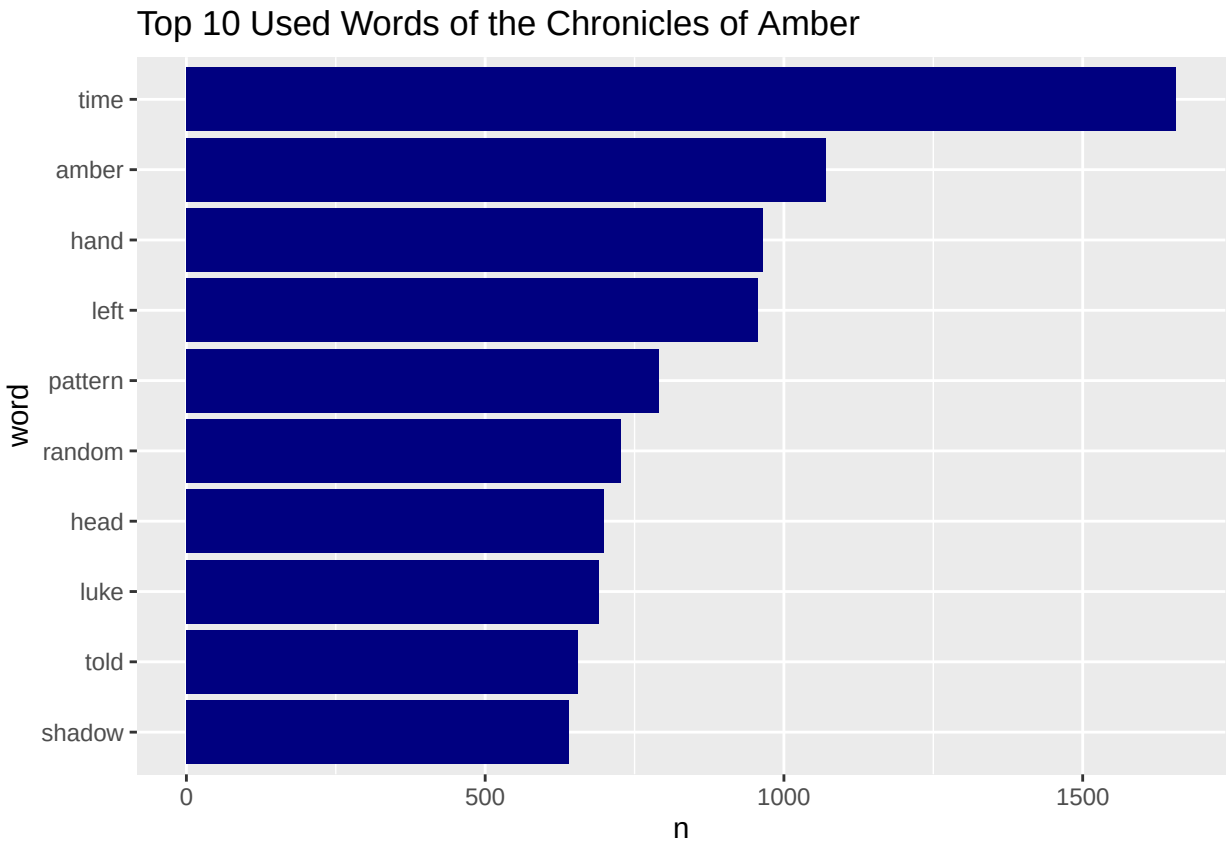
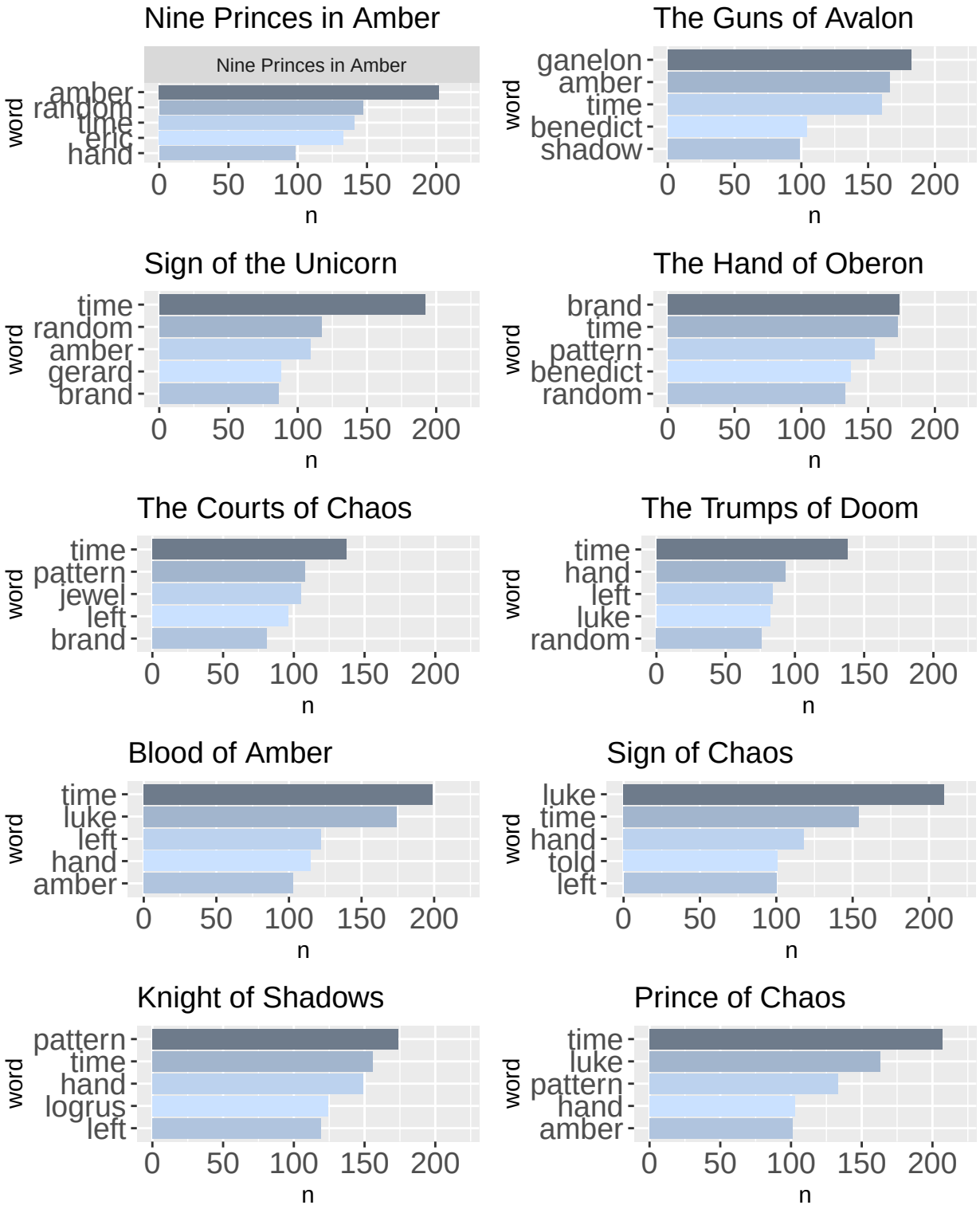
Chronicles_of_Amber

Crystal Sawtelle

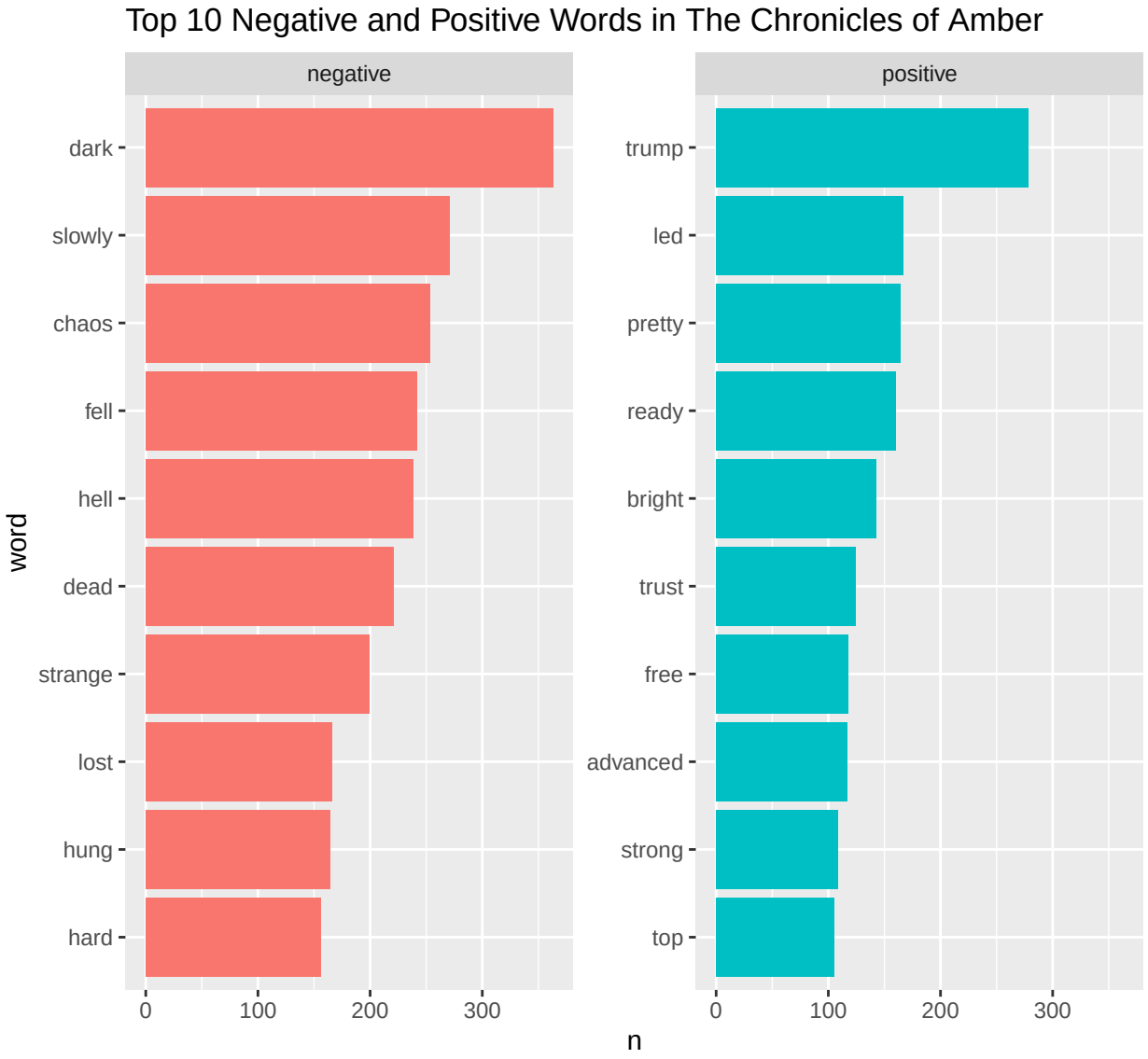
STA-486C



Graph 1: Displaying the top 10 used words for the Chronicles of Amber

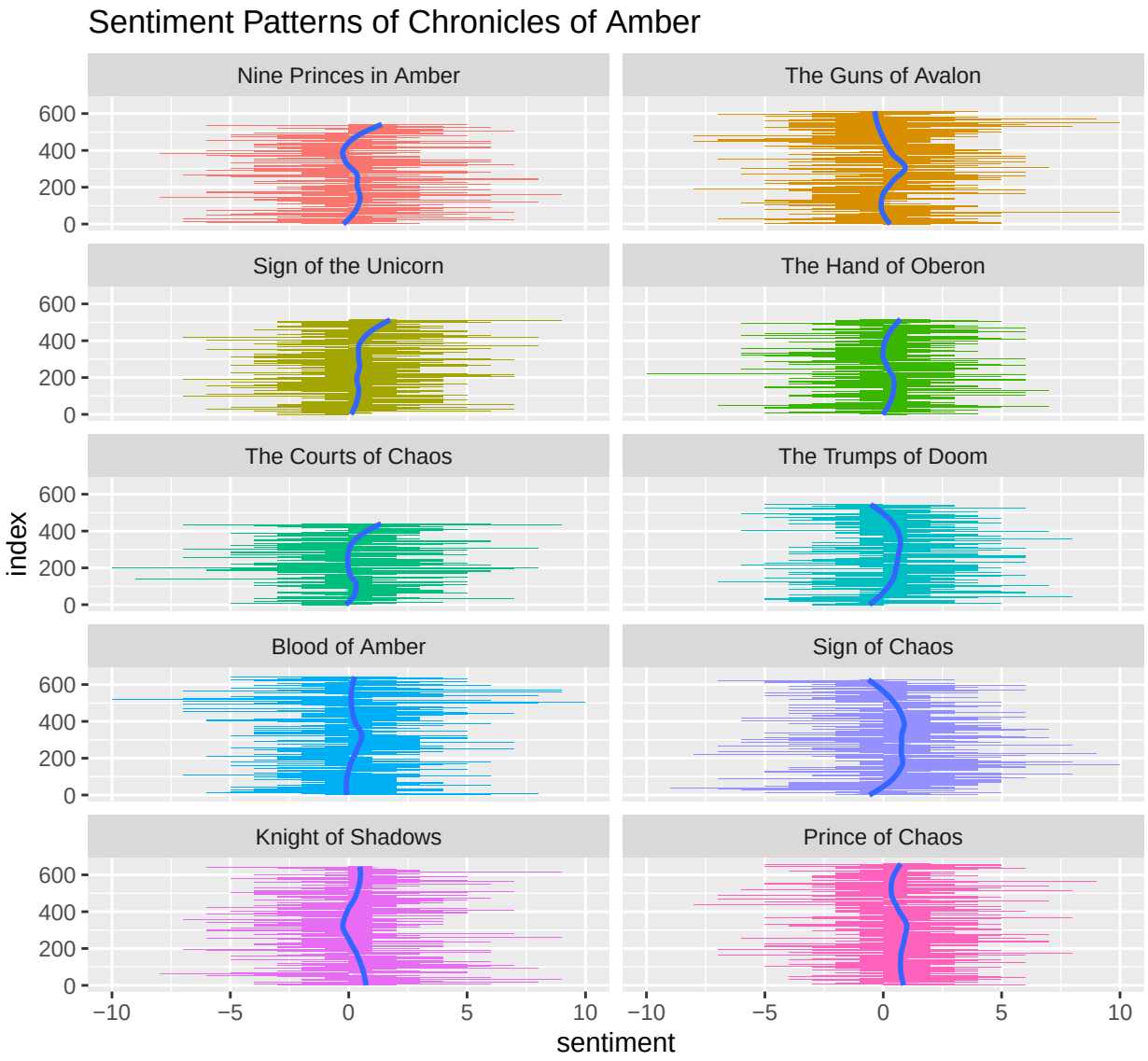


Graph 2: Displaying top 5 used word in each book.

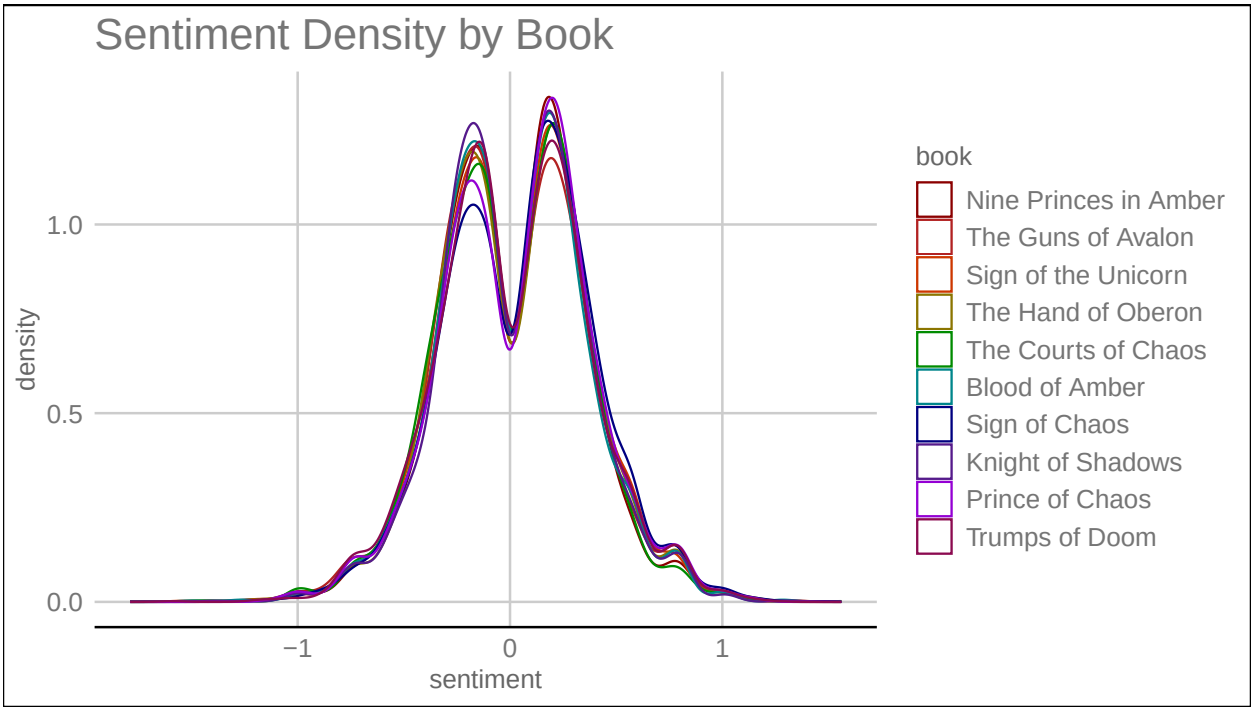


Graph 3: Identifying the top 10 positive and top 10 negative words in the book series. Will remove words that are proper nouns because they do not have a sentiment (i.e. corwin, random, amber, and luke). The **negative** graph includes the word *"chaos"*, which is misclassified because it is a place, *The Courts of Chaos*. The **positive** graph has also misclassified the word *"trump"*, it is in reference to a *trump card*, which they shorten in the series to just trump.

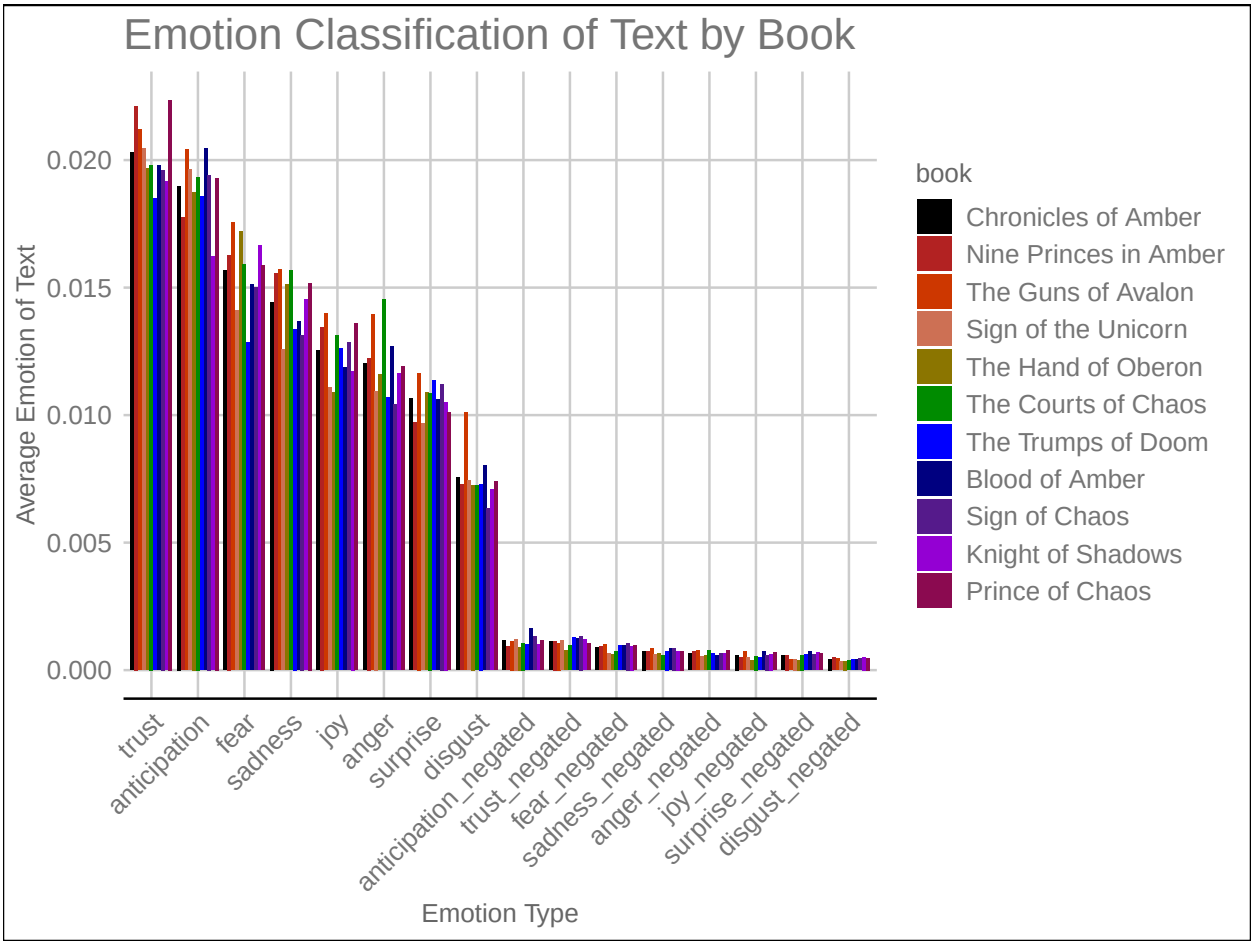
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



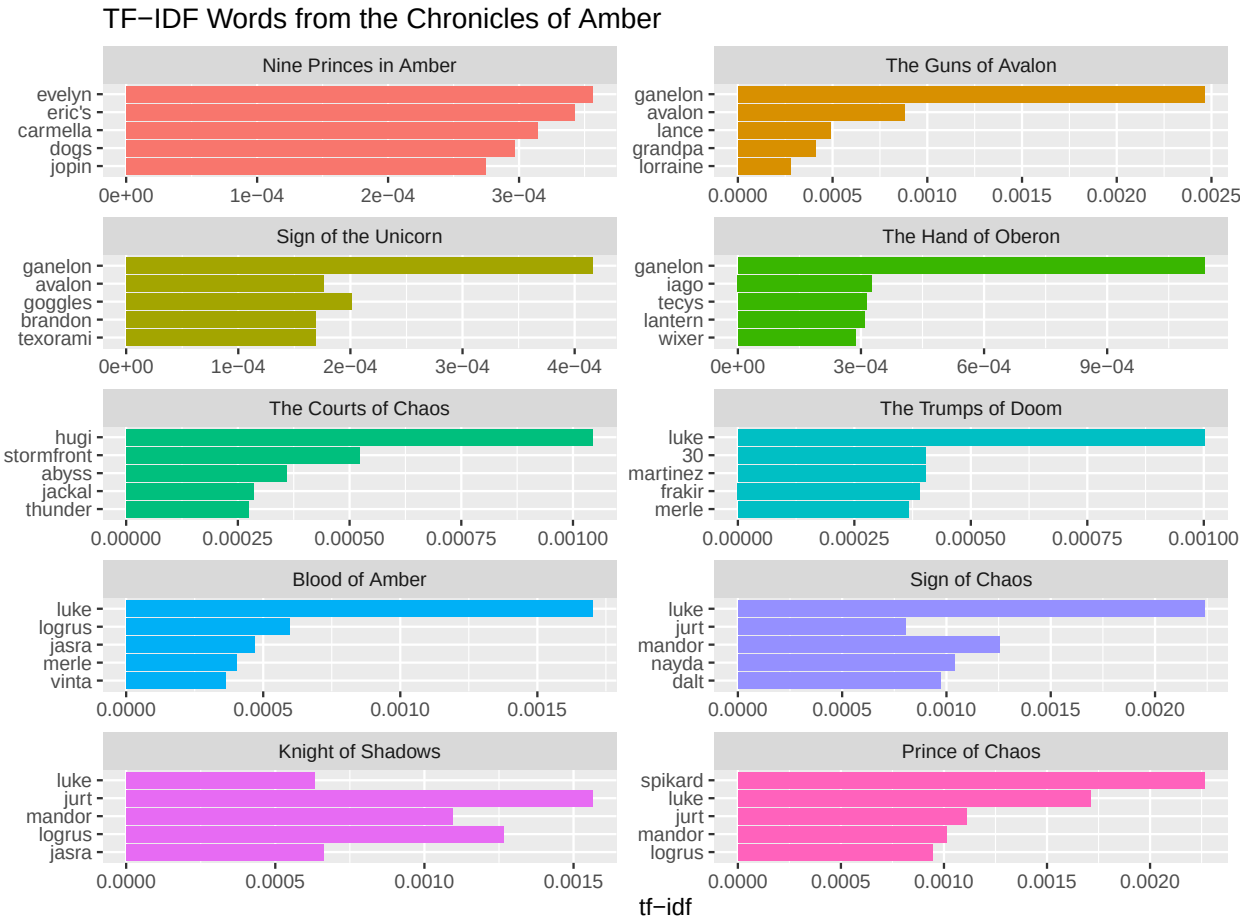
Graph 4: Graphing using "nrc" lexicon, filtering only words associated to the sentiment of positive and negative. Displays the difference between the number of positive and negative words found in every 10 lines of text. Similar patterns of sentiment can be seen between **Nine Princes of Amber**, **The Hand of Oberon**, *and* **The Courts of Chaos**, starting more negative increasing to positive, then decreasing negative and finishing the book positive. Another similar pattern can be found between **The Guns of Avalon**, **Blood of Amber**, *and* **Prince of Chaos**, increasing from more negative to positive, then decreasing again more negative. **Trumps of Doom** and **Sign of Chaos** also have a similar pattern going from negative to positive ending negative. Both **Sign of the Unicorn** and **Knight of Shadow** have unique patterns in sentiment.



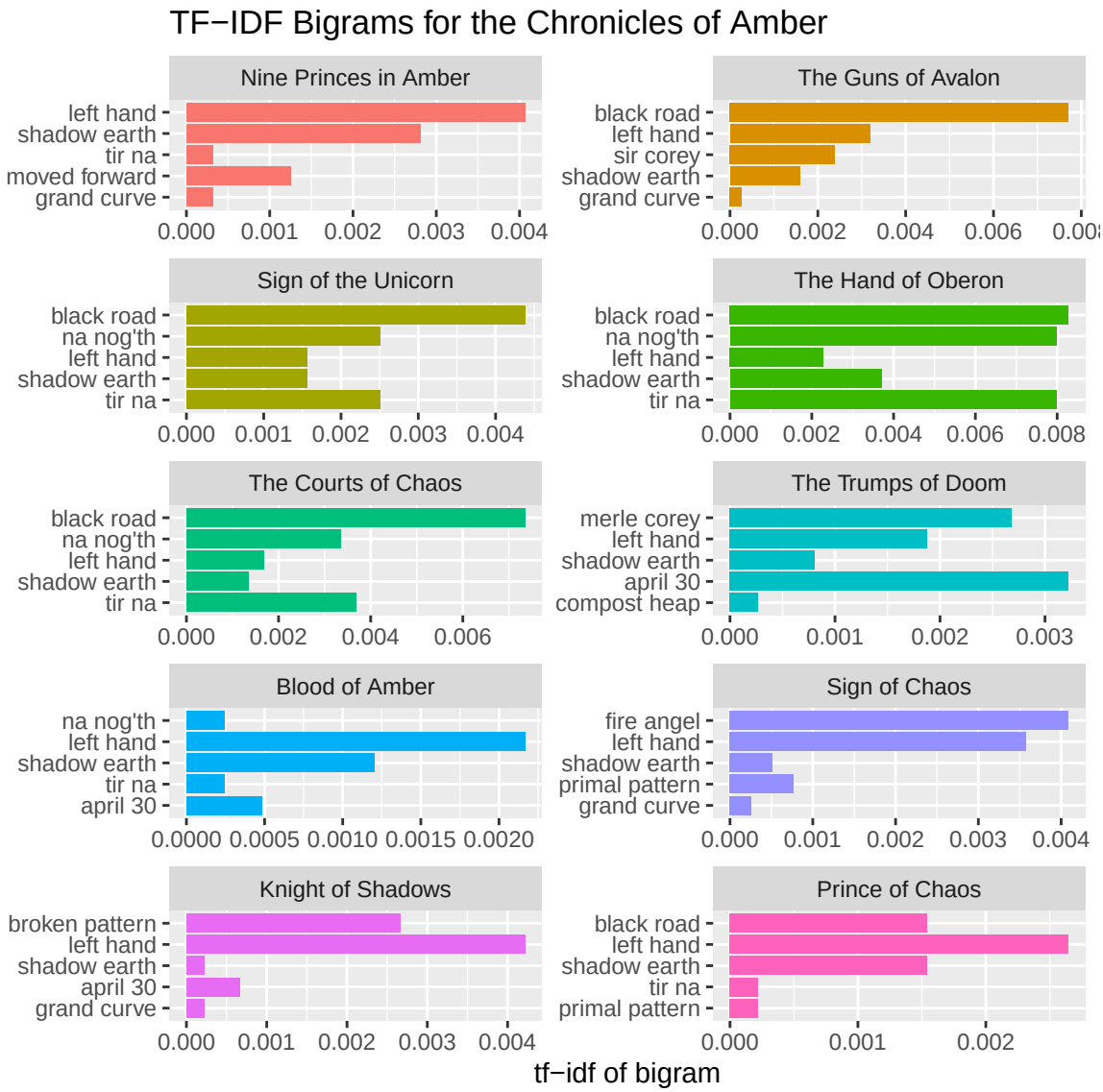
Graph 5: Displays the density of the sentiment of each book excluding when sentiment equals zero. Words with a zero sentiment cannot be categorized into a positive or negative sentiment. Each book has a very similar density.



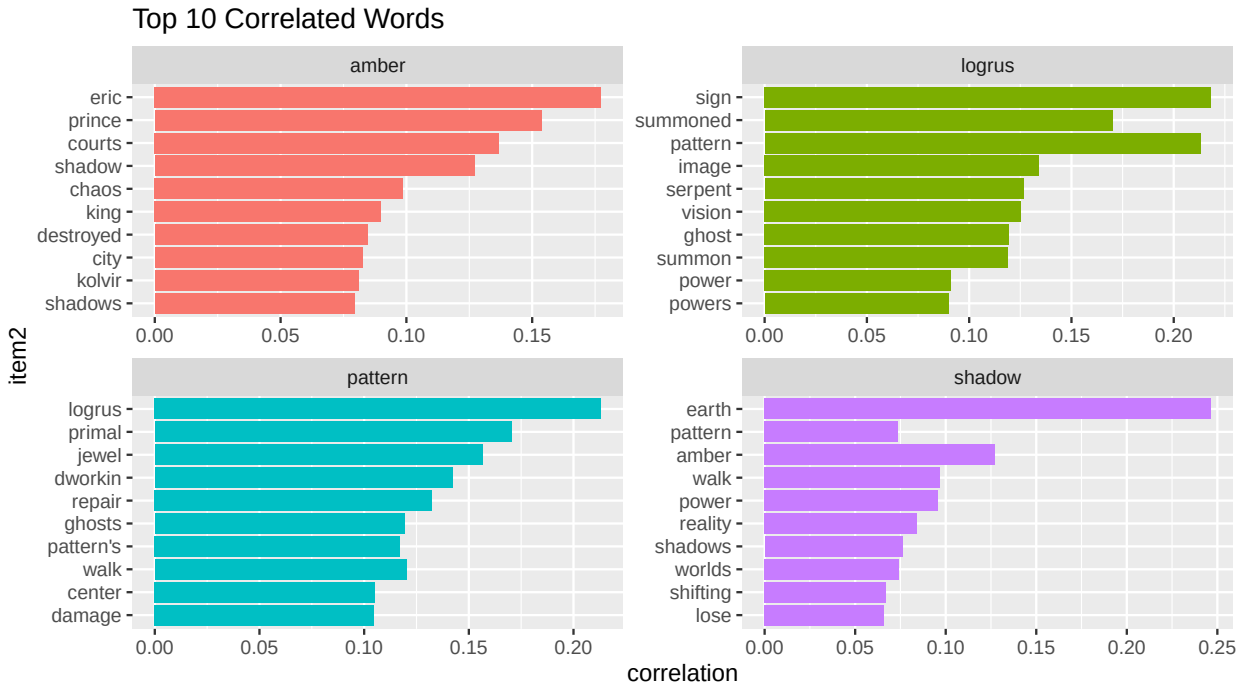
Graph 6: Displaying emotion count by each line of text, returning a count for each emotion it detects (anger, anticipation, disgust, fear, joy, sadness, surprise, trust, and any negation to those emotions.). Trust is the most used emotion in all of the books with an average of approximate 0.21.



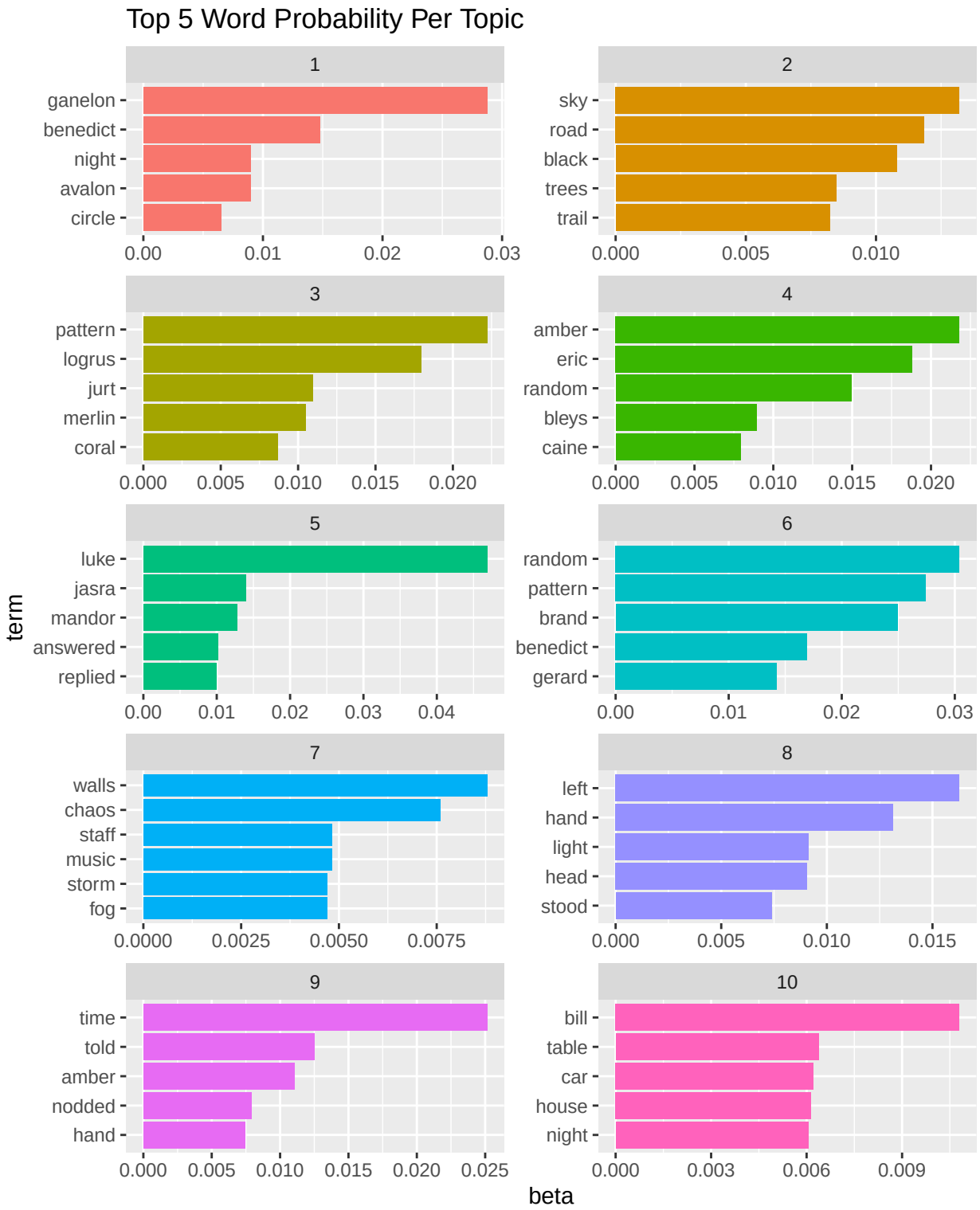
Graph 7: Displaying the tf-idf words for each book. The term frequency - inverse document frequency (tf-idf) is the term frequency (tf) multiplied by the inverse document frequency (idf). The higher the term frequency, the lower the inverse document frequency. The tf-idf attempts to find the words that are important/common in the text, but not too common. When idf and tf-idf are zero, these are extremely common words and thus not as important. This approach decreases the weight for those common words and allows us to find words that are characteristic for one book within all the books; i.e., words that are more common in one book than another.



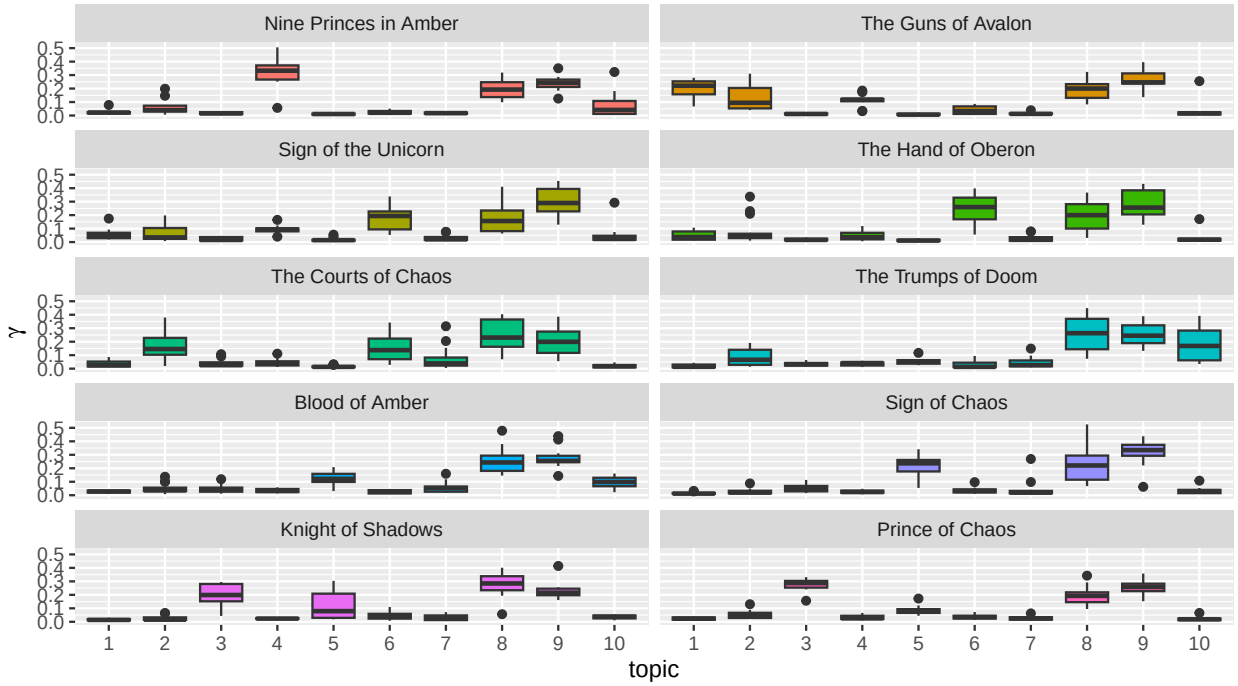
Graph 8: Displaying the top five tf-idf bigrams for each book. Bigrams identify how often one word is followed by another word. Using bigrams in conjunction with tf-idf we can find the common, but not too common bigrams for each book.



Graph 9: Displaying the top 10 words correlated with **Amber**, **Logrus**, **Pattern**, and **Shadow**. This uses the Phi coefficient for measuring binary correlation based on how often a word appears in the same 10 line section.



Graph 10: Visual representation of top five word with the highest probability of showing up in each topic.



Graph 11: Displaying Topic modeling - per-document-per-topic probabilities, γ. Where γ is an estimated proportion of words from that document that are generated from that topic.