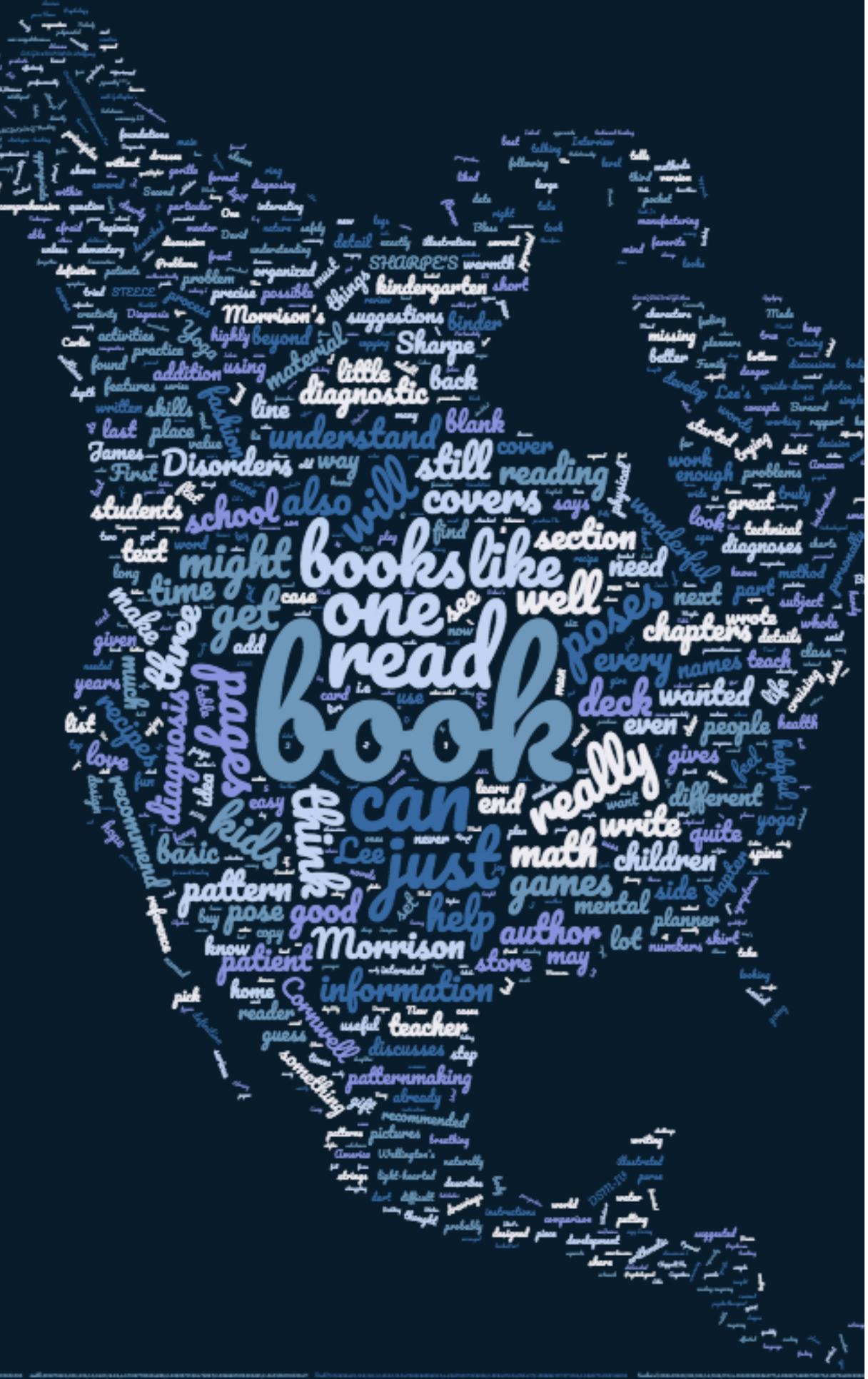
HELPFUL RATING METRIC

- Helpful Votes scaled between 0 and 1 and Helpful Vote Percentage can now be combined.
- Percentage of Helpful Reviews per Book added to balanced popularity.
- Multiplying the standard deviation of each column by 100 and rounding down gives 0.42 for Helpful Votes Scaled and 0.56 for Helpful Percentage.
- The remaining percentage is left for Percentage of Helpful Reviews per Book.

$$Helpful\_Rating = 0.42 * Helpful\_Votes\_Scaled + 0.56 * Helpful\_Percentage + 0.02 * Percentage\_Helpful\_Reviews\_Book$$



NATURAL LANGUAGE PROCESSING

- Before making predictions, book reviews must be converted into a corpus.
- Normalize corpus with lowercase letters; eliminate stop words and special characters.
- Use CountVectorizer and TfidfVectorizer to convert individual reviews into a sparse matrix.
- Iterate over n-grams. One word, is the default. Also try 2 word combinations and 3 word combinations.
- Best results consistently came from CountVectorizer(ngram_range=(1,2)) using, 1 and 2 word combinations.