Who wrote the scariest novels/poems

Contents

Introduction

In this exploratory analysis, the work of three masters of horror novel/poem will be examined. The analysis tries to find out who wrote the scariest novels/poems. Three authors, namely Edgar Allan Poe, Mary Wollstonecraft Shelley and Howard Phillips Lovecraft???are popular horror authors in their time and in their countries.

This analysis is mainly based on sentiment analysis and is divided into 2 major sections: single unit (word) analysis and full unit (paragraph) analysis. Packages used in this analysis include ggplot2, tibble, tidyr, stringr, tidytext, wordcloud, ggridges, dplyr, stopwords, sentimentr and tidyverse.

Single Unit Analysis

Seperating dataset

The dataset (spooky.csv) contains texts from works written by spooky authors: Edgar Allan Poe, Howard Phillips Lovecraft and Mary Wollstonecraft Shelley. The dataset has three columns: id, text and author. First, let's explore the data.

```
##    Length     Class      Mode
##     19579 character character
```

We could see that, among 19579 rows, 7900 texts belong to EAP (Edgar Allan Poe), 5635 texts belong to HPL (Howard Phillips Lovecraft) and 6044 texts belong to MWS (Mary Wollstonecraft Shelley).

To further analyse each word in given texts, I separated the dataset by authors' names and built up three datasets for each author.

```
##       id                text           author
##  Length:7900        Length:7900        EAP:7900
##  Class :character   Class :character   HPL:   0
##  Mode  :character   Mode  :character   MWS:   0
```

```
##       id                text           author
##  Length:5635        Length:5635        EAP:   0
##  Class :character   Class :character   HPL:5635
##  Mode  :character   Mode  :character   MWS:   0
```

```
##       id                text           author
##  Length:6044        Length:6044        EAP:   0
##  Class :character   Class :character   HPL:   0
##  Mode  :character   Mode  :character   MWS:6044
```

Figure 1: Let's get started

	id	author	word
1	id26305	EAP	process
2	id26305	EAP	afforded
3	id26305	EAP	means
4	id26305	EAP	ascertaining
5	id26305	EAP	dimensions
6	id26305	EAP	dungeon
7	id26305	EAP	circuit
8	id26305	EAP	return

Figure 2: Tidy Data Example

Cleaning data

Then, I cleaned the data by removing all punctuation and transform all words into lower case using unnest_tokens and removing stopwords using anti_join.

Next

```
## [1] 729 563 559 559 540 516
```

```
## [1] 260 230 178 174 168 164
```


```
## [1] 254 238 188 186 174 169
```


```
## [1] 329 273 262 248 243 231
```

The above word clouds for three au-

thors show the most frequent 100 words in their works. We could see that in EAP used words such as time and length most frequently in his work, HPL used words such as night, strange and heard most frequently in his work while MWS used words such as raymond, love and life most frequently in her work (which do not seem scary to me).