Kyle Dunn

Text mining midterm

**(1)**

**A screenshot of a computer code

AI-generated content may be incorrect.**

Harry potter without stop words removed – from previous assignments ‘the’ being on top of the list makes sense. The most interesting word in the list is harry as I am pretty sure he is the sole focus of the books having his name 16 thousand times makes sense.

A screenshot of a computer code

AI-generated content may be incorrect.

Harry potter with stop words removed – ‘Harry’ having over double the count as the second entry makes sense. ‘Ron’ and ‘Hermione’ are his friends, the series takes place in a school so ‘professor’ having a high count makes sense. A book series about magic having ‘wand’ on the top ten list makes sense.

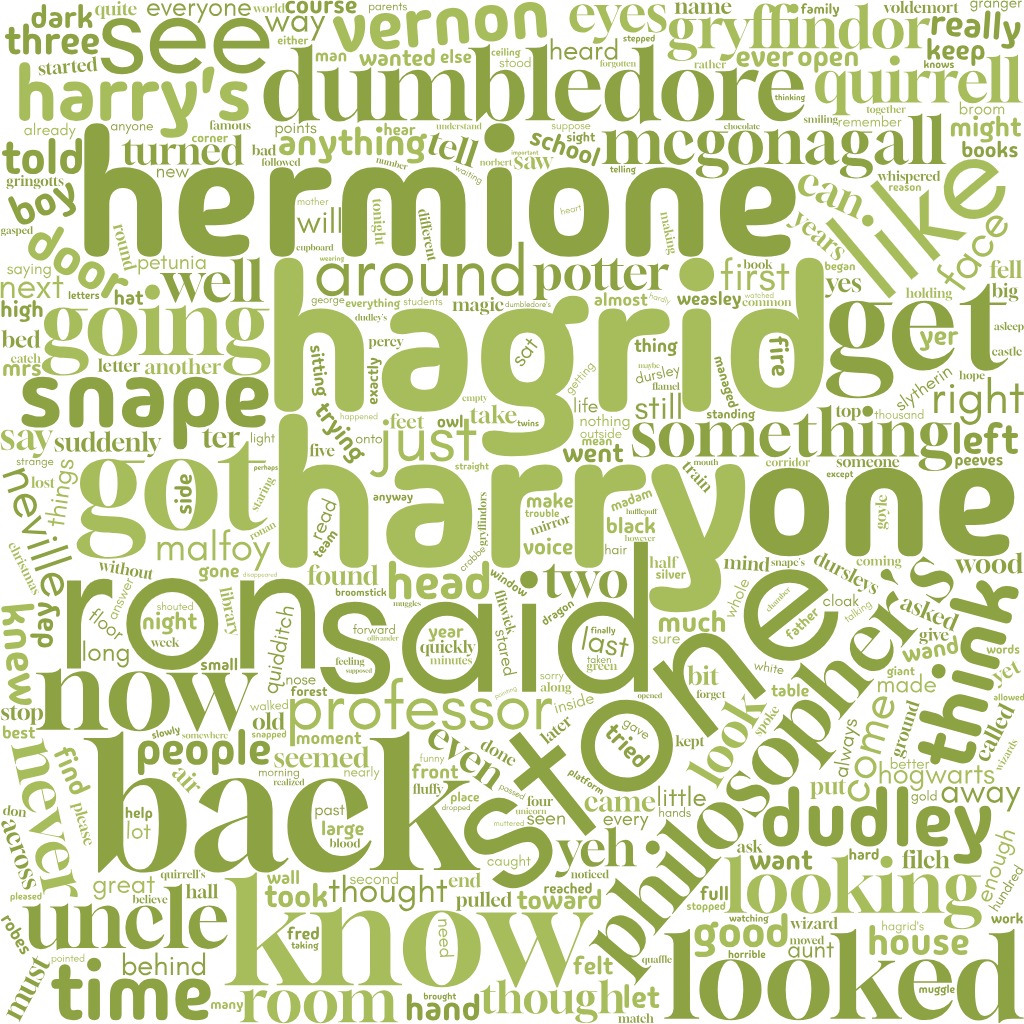
**(2)**

A screenshot of a graph

AI-generated content may be incorrect.

The 7 bar graphs harry being the most frequently used word in each book makes sense. Then for the other 9 is it commonly characters names, most commonly Ron and Hermione. The word professor is in 6 of the 7 graphs which for a book series set in a school makes sense. Sirus, Dumbledore, Hagrid Snape and Malfoy are important characters in the series it is understandable there appear in a majority of the bar graphs.

**(3)**

****

word cloud for philosopher’s stone – the way a person could tell this is a word cloud based off of that book is the word stone and philosophers are prominently displayed in the word cloud.

**(4)**

**Chosen Author Rick Riordan**

**A screenshot of a computer

AI-generated content may be incorrect.**

**(5)**

**A graph showing a line of data

AI-generated content may be incorrect.**

**(6)**

**Riordan**

**TTR - 0.100723833985173**

**HAPAX - 5361**

**JK Rowling**

**TTR - 0.0581304447669163**

**HAPAX - 9093**

**For Riordan the TTR is higher than JK Rowling which means Riordan uses a wider pool of unique words in their work. JK Rowling has a greater hapax than Riordan that means JK Rowling has a higher count of unique words like names or other harry potter terms.**

**(7)**

**A white background with text and colorful text

AI-generated content may be incorrect.**

**A diagram of a cluster analysis

AI-generated content may be incorrect.**

From the two stylo charts I believe Galbraith\_Cuckoo.text – Excerpt of “A Cuckoo’s Calling”, is written by JK Rowling. I believe “Harry Potter and Leopard-Walks-Up-To-Dragon” is also written by JK Rowling