

Text & Joins

MA 615

Problem 1

Install Bradley Hoehmke's Harry Potter R Package!

```
devtools::install_github("bradleyboehmke/harrypotter")
```

(If you install using the RStudio popup you will install the **WRONG** package)

In this assignment we will deal with the book *Harry Potter and the Chamber of Secrets*. You may remember that in this book, sword of Gryffindor plays an important role when Harry enters the Chamber of Secrets and destroys Tom Riddle's diary with a basilisk fang. If you don't remember the story from the book or the film (in which Richard Harris plays Dumbledore for the last time), you can read about it [Harry Potter and the Chamber of Secrets](#).

- a) Using the phi coefficient for binary correlation, what word in the book has the highest correlation with "sword?"
- b) What is the correlation between "sword" and "riddle?"

Problem 2

Before we leave *Harry Potter and the Chamber of Secrets*, let's do some word counting.

Begin by loading the tidytext package, which comes with the stop_words list we have used.

- a) Is the word "itself" in the stop_words list?
- b) Is "highest?"
- c) Does the word "highest" appear in the text you downloaded using Bradley Hoehmke's Harry Potter R Package?
- d) How many times?

- e) Now remove the stop_words from the book text file. How many times does “highest” appear in the book text file?