The Adcentures of Sherlock Holmes

Akhilesh Atluri

December 13, 2017

Abstract

In this Article we are building wordcloud of the words from sherlock Holmes using sherlock Holmes clip art

The Adventures of Sherlock Holmes is a novel by SIR ARTHUR CONAN DOYLE

1 Getting the data and getting required packages

```r
library(dplyr)
library(tidytext)
library(stringr)
library(wordcloud)
library(tm)
library(gutenbergr)
library(knitr)
library(wordcloud2)

#After installing and bringing in all the packages let's bring in text for 'The Adventures

gutenberg_works(str_detect(title,'The Adventure'))

## # A tibble: 171 x 8
##    gutenberg_id
##           <int>
## 1            74
## 2           500
## 3           909
## 4          1194
## 5          1218
## 6          1372
## 7          1644
```

```
##  8         1661
##  9         1825
## 10         2343
## # ... with 161 more rows, and 7 more variables: title <chr>, author <chr>,
## #   gutenberg_author_id <int>, language <chr>, gutenberg_bookshelf <chr>,
## #   rights <chr>, has_text <lgl>

sherlock<-gutenberg_download(1661)
```

2 Getting the words

Now using dplyr lets get words without blank lines

```
sherlock_words<-sherlock%>%
  unnest_tokens(word,text)
```

Let's get rid of all the unwanted words such as 'the' ,'and' etc..

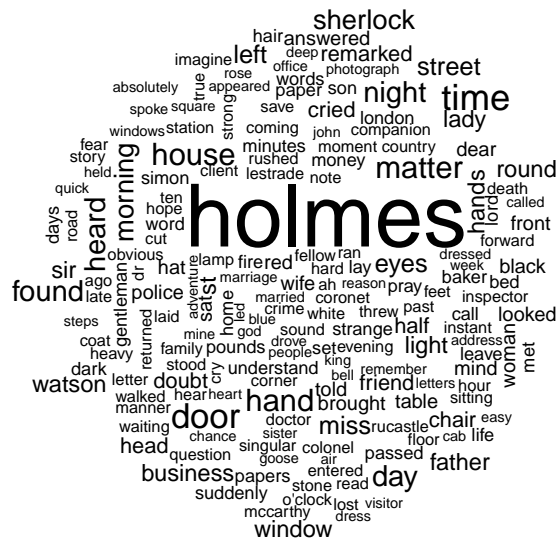```
sherlock_words<-sherlock_words%>%
  filter(!(word %in% stop_words$word))
```

Now let's group each word and get the count of each word

```
sherlock_freq<-sherlock_words%>%
  group_by(word)%>%
  summarize(count=n())
```

3 wordcloud2

```
wordcloud(sherlock_freq$word,sherlock_freq$count,min.freq = 25)
```

References

DOYLE, S. A. C. (2017). The adventures of sherlock holmes.