Wordcloud

Tiffany Gonzalez

December 8, 2017

Outline

Install and Load Libraries

- `library(dplyr)`

Install and Load Libraries

- `library(dplyr)`

- `library(tidytext)`

Install and Load Libraries

- `library(dplyr)`

- `library(tidytext)`

- `library(gutenbergr)`

Install and Load Libraries

- `library(dplyr)`

- `library(tidytext)`

- `library(gutenbergr)`

- `library(stringr)`

Install and Load Libraries

- `library(dplyr)`

- `library(tidytext)`

- `library(gutenbergr)`

- `library(stringr)`

- `library(wordcloud)`

Access Gutenberg database

```
df<-gutenberg_works(str_detect(title,'Peter Pan'))

   df$gutenberg_id

## [1]  1332 24012 39755

   df$title

## [1] "Peter Pan in Kensington Gardens"
## [2] "The Peter Pan Alphabet"
## [3] "The Story of Peter Pan, Retold from the fairy play
```

Download Peter Pan

```
          peter_pan<-gutenberg_download(39755)
colnames(peter_pan)

## [1] "gutenberg_id" "text"
```

Unpack the words

```
                 words_df<-peter_pan%>%
  unnest_tokens(word,text)

  words_df[1:8,2]

## # A tibble: 8 x 1
##           word
##          <chr>
## 1 illustration
## 2         with
## 3          the
## 4       spring
## 5        comes
## 6        wendy
## 7          the
## 8        story
```

Remove Common Words

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

words_df[1:8,]

## # A tibble: 8 x 2
##    gutenberg_id        word
##           <int>       <chr>
## 1        39755 illustration
## 2        39755       spring
## 3        39755        wendy
## 4        39755        story
## 5        39755        peter
## 6        39755          pan
## 7        39755       retold
## 8        39755        fairy
```

Frequency Count

```
                word_freq<-words_df%>%
group_by(word)%>%
summarize(count=n())

  word_freq[1:8,]

## # A tibble: 8 x 2
##      word count
##     <chr> <int>
## 1 _colour     8
## 2 _first_     1
## 3  _hear_     1
## 4   _his_     1
## 5     _i_     1
## 6  _kiss_     1
## 7  _like_     1
## 8    _me_     1
```

The Wordcloud

```
wordcloud(word_freq$word,word_freq$count,mi
```

wendy

land crocodile
home boys
wicked
tinker shadow nana window
live mother house plate_
told poor
children lived girl ship return
heard left story round
nursery lost captain bell night
pirate cried fairy kiss
pirates bed boy
smee
time illustration sitting fairies
found michael called father
colour kennel
darling john pan hook cabin

peter