Creating a Wordcloud

Ron Richardson

November 30, 2017

Outline

Install and Load Libraries

- `library(dplyr)`

Install and Load Libraries

- `library(dplyr)`

- `library(janeaustenr)`

Install and Load Libraries

- `library(dplyr)`

- `library(janeaustenr)`

- `library(tidytext)`

Install and Load Libraries

- `library(dplyr)`

- `library(janeaustenr)`

- `library(tidytext)`

- `library(stringr)`

Install and Load Libraries

- `library(dplyr)`

- `library(janeaustenr)`

- `library(tidytext)`

- `library(stringr)`

- `library(wordcloud)`

Getting Text to Parse

```
sns<-austen_books()
sns<-sns%>%
  filter(book == 'Sense & Sensibility')
```

Filter out Chapter Titles

```
sns<-sns%>%
  filter(!str_detect(sns$text, "^CHAPTER"))
```

Unpack the Words

```
words<-sns%>%
  unnest_tokens(word, text)
```

Filtering out Stop Words

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

Grouping and Summarizing the Data I

```
wordFreq<-words%>%
  group_by(word)%>%
  summarize(count=n())
```

Grouping and Summarizing the Data II

```
wordFreq

## # A tibble: 5,847 x 2
##            word count
##           <chr> <int>
## 1             1     1
## 2          1811     1
## 3           200     1
## 4         7000l     1
## 5     abandoned     1
## 6     abatement     1
## 7     abbeyland     1
## 8         abhor     1
## 9      abhorred     2
## 10   abhorrence     4
## # ... with 5,837 more rows
```

Creating the Wordcloud I

```
wordcloud(wordFreq$word, wordFreq$count, max.words=100)
```

Creating the Wordcloud II