Creating a Wordcloud

Ron Richardson

November 30, 2017

Outline

Install and Load Libraries

Getting Text to Parse

Filter out Chapter Titles

Unpack the Words

Filtering out Stop Words

Grouping and Summarizing the Data

Creating the Wordcloud

► library(dplyr)

- ► library(dplyr)
- ► library(janeaustenr)

- ► library(dplyr)
- ► library(janeaustenr)
- ► library(tidytext)

- ► library(dplyr)
- ► library(janeaustenr)
- ► library(tidytext)
- ► library(stringr)

- ► 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
## 2 1811
## 3 200
## 4
         70001 1
##
   5 abandoned 1
##
   6 abatement
## 7
     abbeyland
         abhor
##
   9 abhorred
##
## 10 abhorrence
## # ... with 5,837 more rows
```

Creating the Wordcloud I

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

Creating the Wordcloud II

behaviour to poor willoughby sister colonel morning town acquaintance park e edward farther sisters to la speak street hopeparty engagement voice immediately pspiritsmindcried happiness day found brought subject manner left comfort brandon pleasure