

The Trial

Chris Kozak

December 15, 2017

Outline

Load Packages

The Trial

Unnest the Tokens

Bing Sentiment Lexicon

Data Transformation

Positive Words

Negative Words

Load Packages

- `library(dplyr)` *#for querying data*
- `library(ggplot2)` *#for visualizing data*
- `library(gutenbergr)` *#for open-source docs*
- `library(stringr)` *#for working with text data*
- `library(easyGgplot2)` *#easily renders plots together*
- `library(tidytext)` *#for quantifying sentiment in text*

About The Trial

German author Franz Kafka wrote *The Trial* around 1915, but it wasn't published until 1925, after his death. It's the bleak, authoritarian story of a man on a trial for a crime he's not even aware of. This version was translated by David Wyllie. Naturally, other translations may feature different word choices, as the translator's goal here is to adapt a book written in an older style of German to a much newer style of English.



Download The Trial

```
trial <- gutenbergl_download(7849) #download The Trial  
head(trial$text) #preview the text
```

```
## [1] "The Trial"  
## [2] "Franz Kafka"  
## [3] ""  
## [4] "Translation Copyright (C) by David Wyllie"  
## [5] "Translator contact email: dandelion@post.cz"  
## [6] ""
```



Unnest the Tokens

```
trial_words <- trial%>%  
  unnest_tokens(word, text) #separates each word  
trial_words$guttenberg_id <- NULL
```

Bing Sentiment Lexicon

```
#for positive/negative sentiments  
bing <- get_sentiments('bing')
```

```
head(bing) #preview bing
```

```
## # A tibble: 6 x 2  
##           word sentiment  
##      <chr>    <chr>  
## 1    2-faced  negative  
## 2    2-faces  negative  
## 3         a+   positive  
## 4  abnormal  negative  
## 5   abolish  negative  
## 6 abominable negative
```

Merging the Lexicon

```
#merge bing with The Trial  
trial_words <- inner_join(trial_words,bing)  
trial_words$gutenberg_id<-NULL #delete extra column  
head(trial_words) #preview the dataframe/tibble
```

```
## # A tibble: 6 x 2  
##           word sentiment  
##       <chr>    <chr>  
## 1      miss  negative  
## 2      lies  negative  
## 3    wrong  negative  
## 4  unusual  negative  
## 5 disconcerted negative  
## 6    knock  negative
```

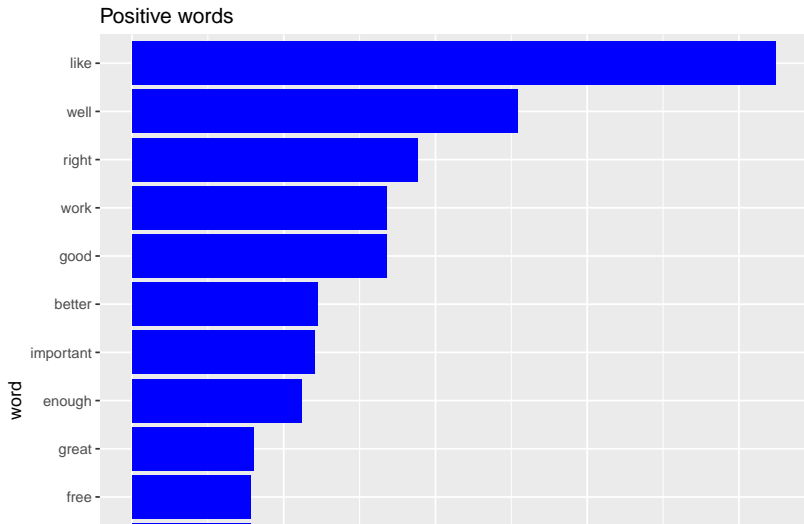



Positive Words Transformation

```
trial_pos <- trial_words%>%  
  filter(sentiment=='positive')%>%  
  group_by(word)%>%  
  summarize(count=n(), sentiment=first(sentiment))%>%  
  arrange(count)%>%  
  top_n(15, wt=count)  
  
trial_pos$word <- factor(trial_pos$word,  
                        levels=trial_pos$word)  
  
pos_plot <- ggplot()+  
  geom_bar(data = trial_pos,  
          aes(x=word, y=count), stat='identity', fill='blue')+  
  coord_flip()+  
  ggtitle('Positive words')
```

Positive Plot

```
pos_plot
```





Negative Words Transformation

```
trial_neg <- trial_words%>%  
  filter(sentiment=='negative')%>%  
  group_by(word)%>%  
  summarize(count=n(),  
             sentiment=first(sentiment))%>%  
  arrange(count)%>%  
  top_n(15, wt=count)  
  
trial_neg$word <- factor(trial_neg$word,  
                        levels=trial_neg$word)  
  
neg_plot <- ggplot()+  
  geom_bar(data = trial_neg,  
           aes(x=word, y=count), stat='identity', fill='red')+  
  coord_flip()+  
  ggtitle('Negative words')
```

Negative Plot

```
neg_plot
```

