

Sentiment dictionaries

INTRODUCTION TO TEXT ANALYSIS IN R



Marc Dotson

Assistant Professor of Marketing

Bing dictionary

```
get_sentiments("bing")
```

```
# A tibble: 6,788 x 2
  word      sentiment
  <chr>     <chr>
1 2-faced  negative
2 2-faces  negative
3 a+       positive
4 abnormal negative
5 abolish negative
# ... with 6,783 more rows
```

Bing dictionary

```
get_sentiments("bing") %>%  
  count(sentiment)
```

```
# A tibble: 2 x 2  
  sentiment      n  
  <chr>      <int>  
1 negative    4782  
2 positive    2006
```

Afinn dictionary

```
get_sentiments("afinn")
```

```
# A tibble: 2,476 x 2
  word      value
  <chr>    <int>
1 abandon     -2
2 abandoned  -2
3 abandons   -2
4 abducted   -2
5 abduction  -2
# ... with 2,471 more rows
```

Afinn dictionary

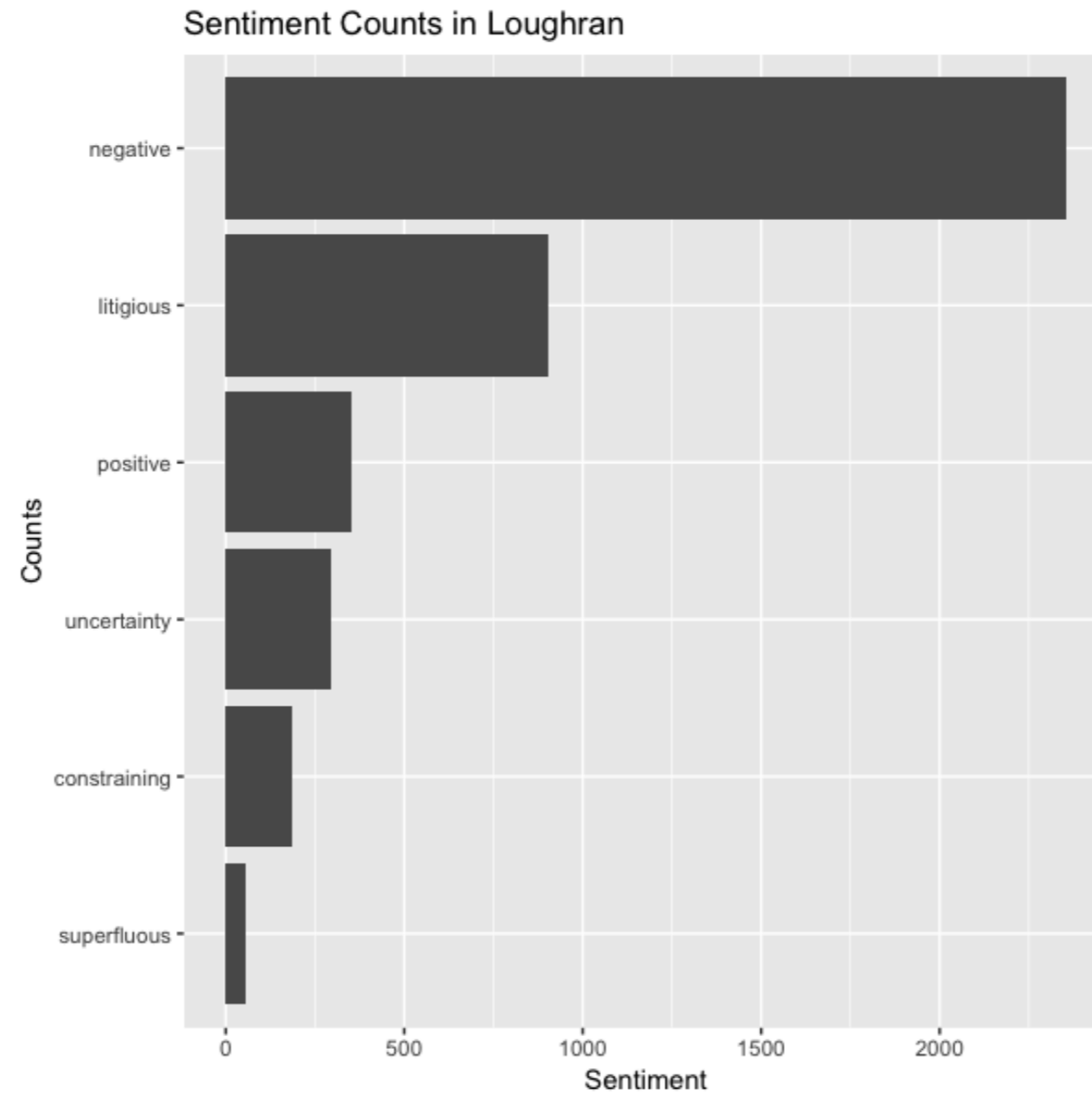
```
get_sentiments("afinn") %>%  
  summarize(  
    min = min(value),  
    max = max(value)  
  )
```

```
# A tibble: 1 x 2  
  min    max  
  <dbl> <dbl>  
1    -5     5
```

Loughran dictionary

```
sentiment_counts <- get_sentiments("Loughran") %>%  
  count(sentiment) %>%  
  mutate(sentiment2 = fct_reorder(sentiment, n))  
ggplot(sentiment_counts, aes(x = sentiment2, y = n)) +  
  geom_col() +  
  coord_flip() +  
  labs(  
    title = "Sentiment Counts in Loughran",  
    x = "Counts",  
    y = "Sentiment"  
  )
```

Loughran dictionary



Let's practice!

INTRODUCTION TO TEXT ANALYSIS IN R

Appending dictionaries

INTRODUCTION TO TEXT ANALYSIS IN R



Marc Dotson

Assistant Professor of Marketing

Using inner_join()

```
tidy_review %>%  
  inner_join(get_sentiments("Loughran"))
```

```
# A tibble: 3,960 x 6  
   id date      product      stars word      sentiment  
   <int> <chr>    <chr>      <dbl> <chr>    <chr>  
1     5 12/22/15 iRobot Roomba 650 for Pets     5 slow      negative  
2     5 12/22/15 iRobot Roomba 650 for Pets     5 easily    positive  
3     5 12/22/15 iRobot Roomba 650 for Pets     5 random    uncertainty  
4     5 12/22/15 iRobot Roomba 650 for Pets     5 easy      positive  
# ... with 3,956 more rows
```

Counting sentiment

```
sentiment_review <- tidy_review %>%  
  inner_join(get_sentiments("Loughran"))  
sentiment_review %>%  
  count(sentiment)
```

```
# A tibble: 6 x 2  
  sentiment      n  
  <chr>      <int>  
3 negative    1795  
4 positive    1568  
# ... with 4 more rows
```

Counting sentiment

```
sentiment_review %>%  
  count(word, sentiment) %>%  
  arrange(desc(n))
```

```
# A tibble: 598 x 3  
  word      sentiment      n  
  <chr>    <chr>    <int>  
1 easy     positive   297  
2 happy    positive   107  
3 trouble  negative    58  
# ... with 595 more rows
```

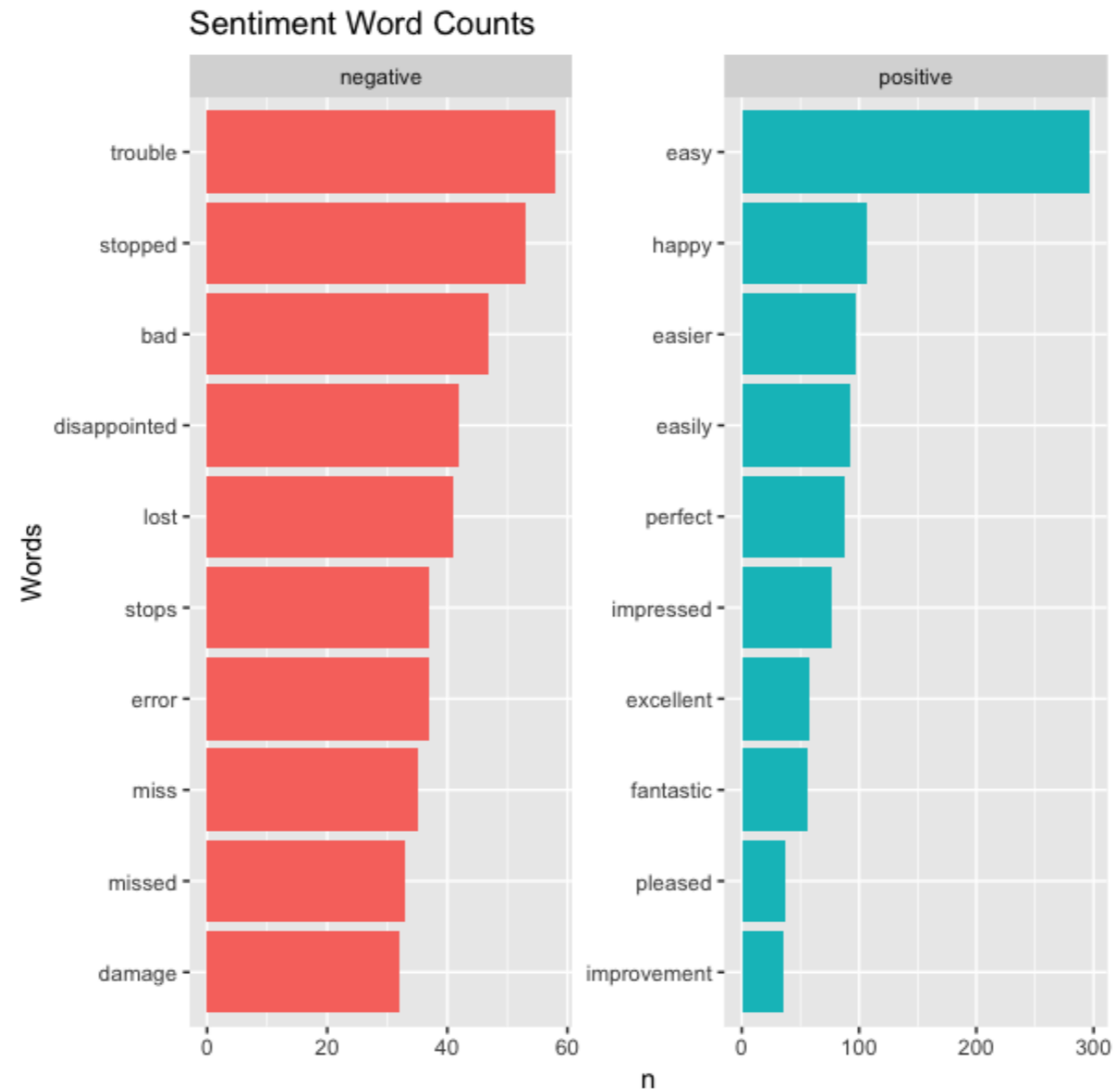
Visualizing sentiment

```
sentiment_review2 <- sentiment_review %>%  
  filter(sentiment %in% c("positive", "negative"))  
word_counts <- sentiment_review2 %>%  
  count(word, sentiment) %>%  
  group_by(sentiment) %>%  
  top_n(10, n) %>%  
  ungroup() %>%  
  mutate(  
    word2 = fct_reorder(word, n)  
  )
```

Visualizing sentiment

```
ggplot(word_counts, aes(x = word2, y = n, fill = sentiment)) +  
  geom_col(show.legend = FALSE) +  
  facet_wrap(~ sentiment, scales = "free") +  
  coord_flip() +  
  labs(  
    title = "Sentiment Word Counts",  
    x = "Words"  
  )
```

Visualizing sentiment



Let's practice!

INTRODUCTION TO TEXT ANALYSIS IN R

Improving sentiment analysis

INTRODUCTION TO TEXT ANALYSIS IN R



Marc Dotson

Assistant Professor of Marketing

Count sentiment by rating

```
tidy_review %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(stars, sentiment)
```

```
# A tibble: 10 x 3  
  stars sentiment     n  
  <dbl> <chr>    <int>  
1     1 negative   381  
2     1 positive  241  
3     2 negative  384  
# ... with 7 more rows
```

Using spread()

```
tidy_review %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(stars, sentiment) %>%  
  spread(sentiment, n)
```

```
# A tibble: 5 x 3  
  stars negative positive  
  <dbl>     <int>     <int>  
1     1     381     241  
2     2     384     247  
# ... with 3 more rows
```

Computing overall sentiment

```
tidy_review %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(stars, sentiment) %>%  
  spread(sentiment, n) %>%  
  mutate(overall_sentiment = positive - negative)
```

```
# A tibble: 5 x 4  
  stars negative positive overall_sentiment  
  <dbl>   <int>   <int>         <int>  
1     1     381     241         -140 ...  
5     5    3705    5083         1378
```

Visualize sentiment by rating

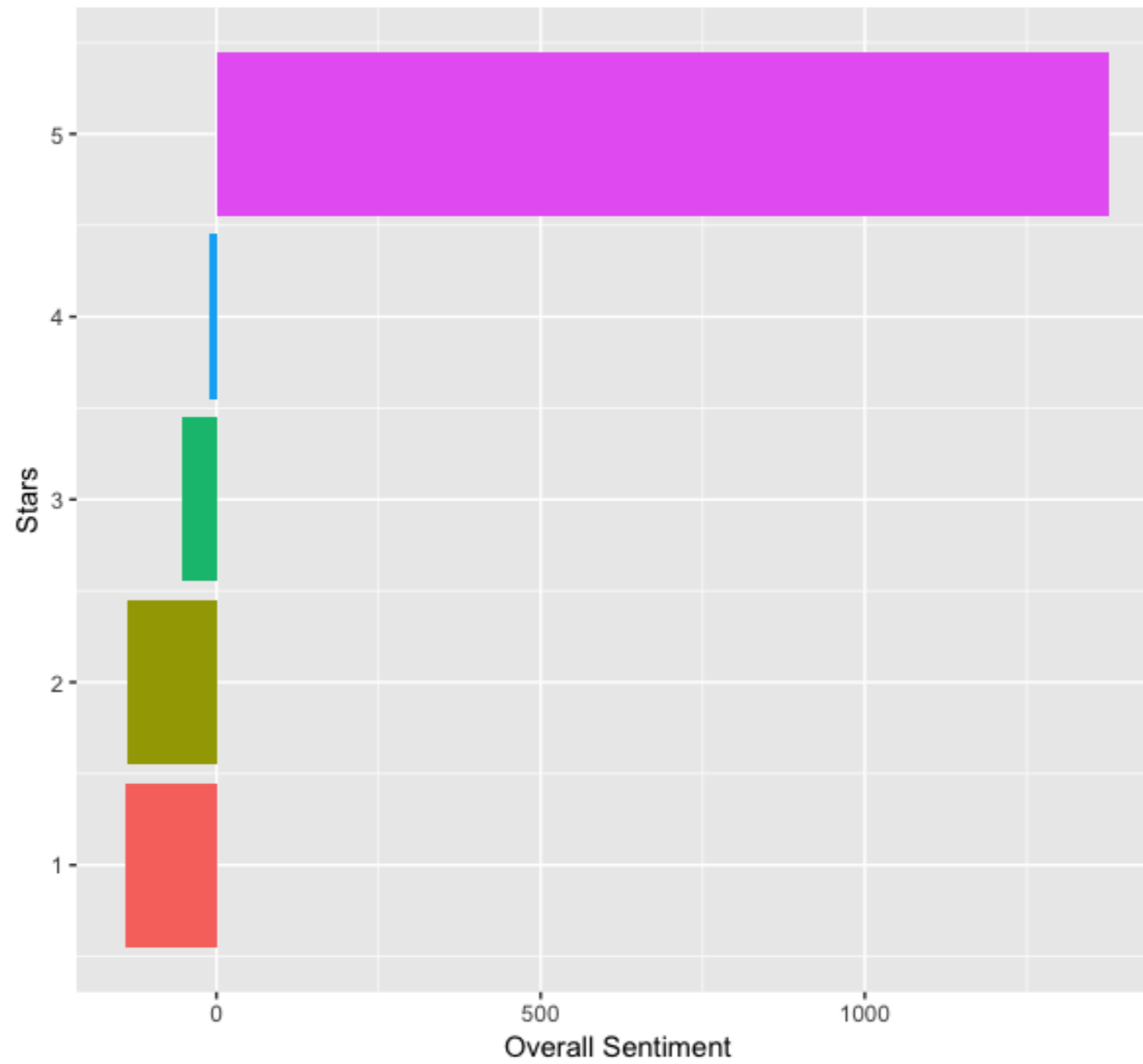
```
sentiment_stars <- tidy_review %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(stars, sentiment) %>%  
  spread(sentiment, n) %>%  
  mutate(  
    overall_sentiment = positive - negative,  
    stars = fct_reorder(stars, overall_sentiment)  
  )
```

Visualize sentiment by rating

```
ggplot(  
  sentiment_stars,  
  aes(x = stars, y = overall_sentiment, fill = as.factor(stars))  
) +  
  geom_col(show.legend = FALSE) +  
  coord_flip() +  
  labs(  
    title = "Overall Sentiment by Stars",  
    subtitle = "Reviews for Robotic Vacuums",  
    x = "Stars",  
    y = "Overall Sentiment"  
  )
```

Overall Sentiment by Stars

Reviews for Robotic Vacuums



Let's practice!

INTRODUCTION TO TEXT ANALYSIS IN R