Text analysis on Uncle Tom's Cabin

TASK ONE: Select the book

The book selected is Uncle Tom's Cabin by Harriet Beecher Stowe, finished in 1852. This book is an anti-slavory novel about the evil and immorality of slavery.

TASK TWO: Bag of words analysis

The text is downloaded from Gutenberg.org, and convert the text of the novels to the tidy format using `unnest_tokens()`.
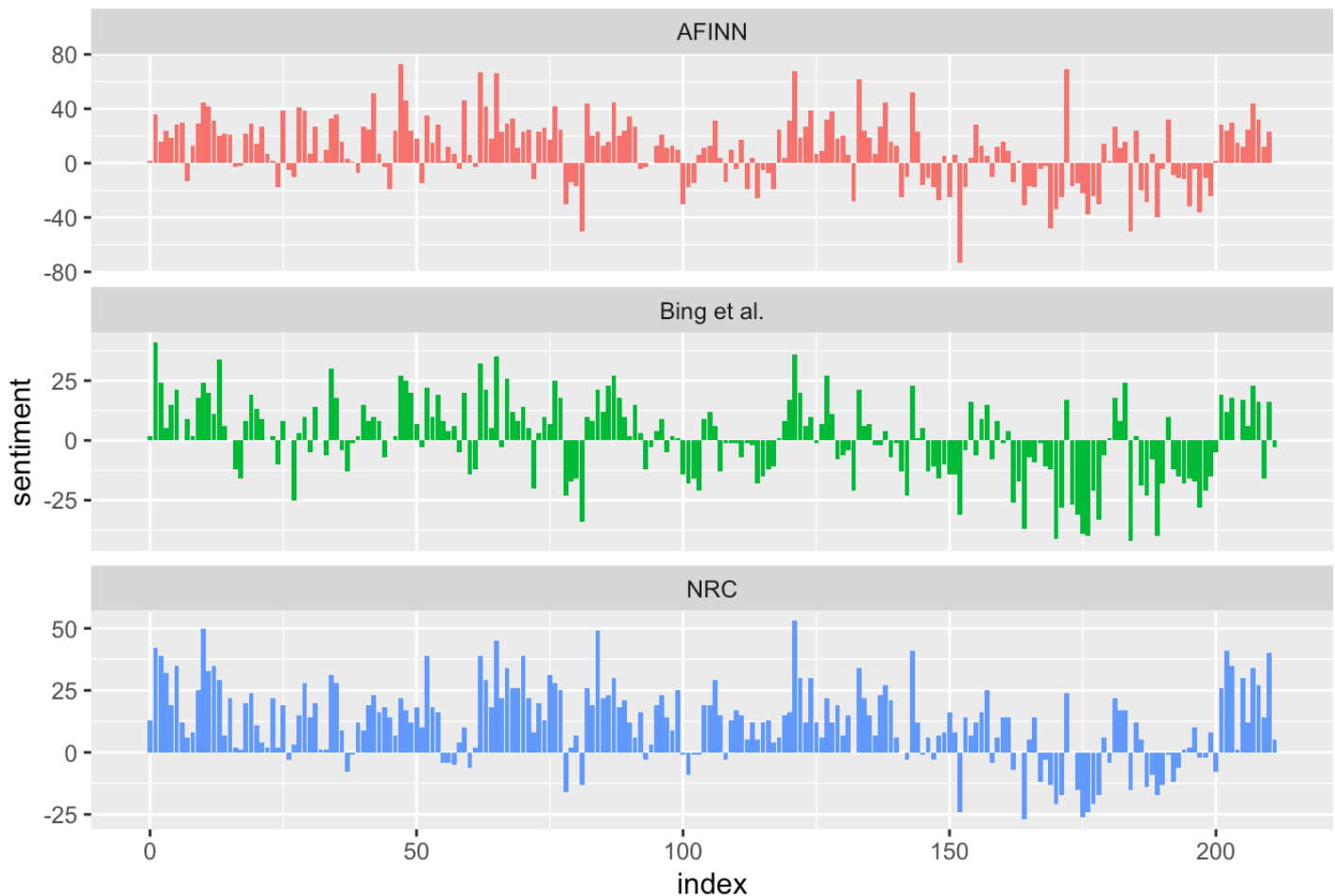
To text out which method for sentiment lexicons fits the plotline of Uncle Tom's Cabin better, we use all three sentiment lexicos to track and compare the sentiment changes across the narrative arc of Uncle Tom's Cabin.

We define an index of 100, which counts up sections of 100 lines of text and count how many positive and negative words.

AFINN, Bing et al., and NRC has similar relative trajectories through the novel. There are similar ips and peaks in sentiment at about the same places in the novel, but the absolute differences are different. The overall trends in the sentiment are simimlar among three methods but the absolute values and variances are different. The AFINN lexicon has the largest absolute values, with mostly positive values. The NRC sentiment is high, but Bing et al. has more variance with more negative values. Bing et al. catches more sentiment variance with more extreme sentiment flowing.

According to Uncle Cabin plotline, Bing et. al would be more suitable, as more negative sentiment are more frequent overall.

Sentiment values of Uncle Tom's Cabin over three methods

Summary Statistics of three sentiment lexicons

	Absolute Value	Variance
AFINN	73	42.0
Bing et al.	53	597.4
NRC	272	237.5

Next, we find out the most frequent positive and negative words. We noticed that in positive words, master ranked one of the highest positive words, however, according to the plotline of Uncle Tom's Cabin, master is a prefix for the characters in the book. Therefore, we will add master to our customized stopword list.
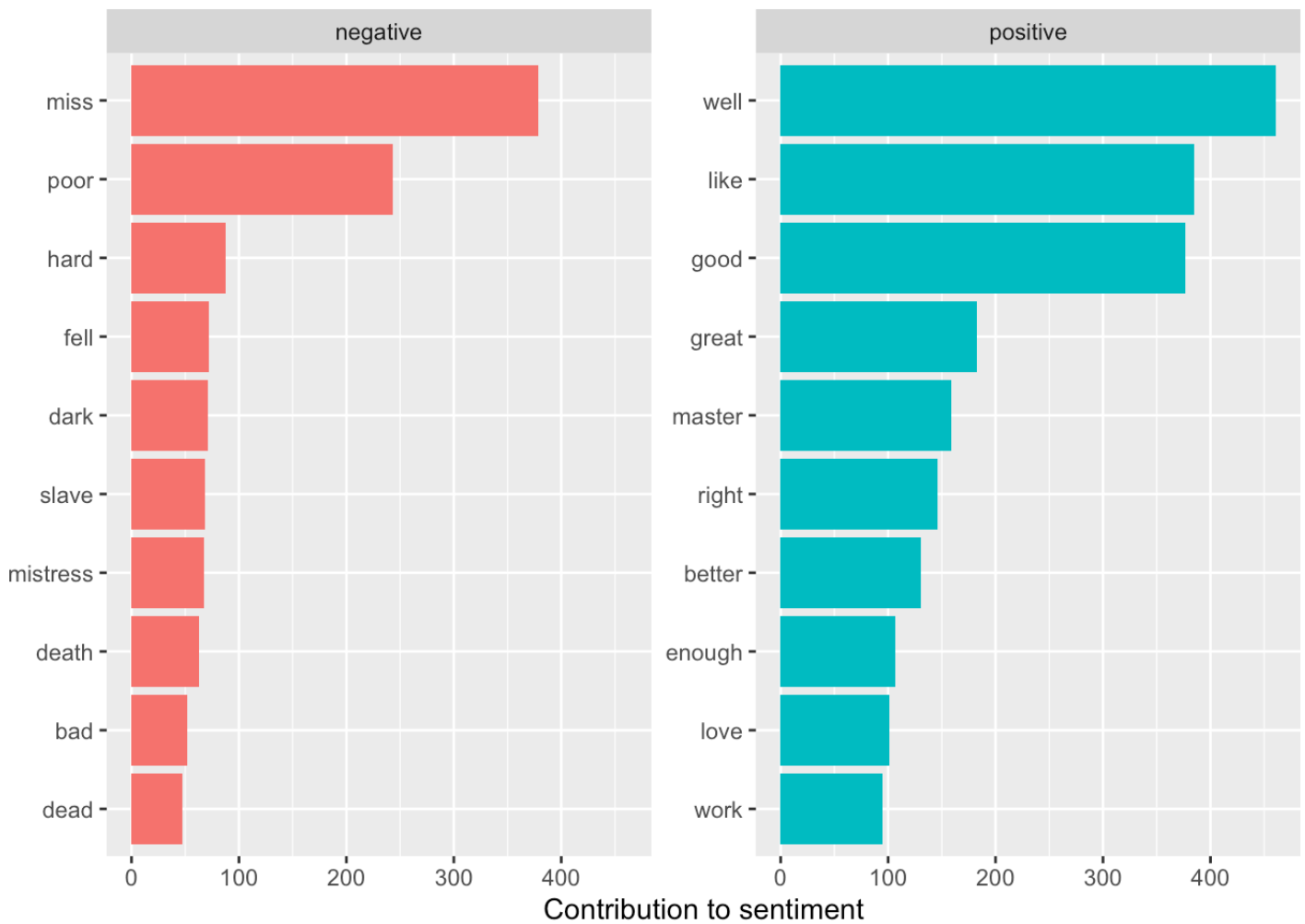
```
bing_word_counts <- tidy_book %>%
   inner_join(get_sentiments("bing")) %>%
   count(word, sentiment, sort = TRUE) %>%
   ungroup()
```

```
## Joining, by = "word"
```

```
bing_word_counts
```

```
## # A tibble: 2,025 × 3
##    word    sentiment      n
##    <chr>   <chr>      <int>
##  1 well    positive     461
##  2 like    positive     385
##  3 miss    negative     379
##  4 good    positive     377
##  5 poor    negative     243
##  6 great   positive     183
##  7 master  positive     159
##  8 right   positive     146
##  9 better  positive     130
## 10 enough  positive     107
## # … with 2,015 more rows
```

```
bing_word_counts %>%
  group_by(sentiment) %>%
  slice_max(n, n = 10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(x = "Contribution to sentiment",
       y = NULL)
```

```
custom_stop_words <- bind_rows(tibble(word = c("master"),
                                      lexicon = c("custom")),
                              stop_words)
```
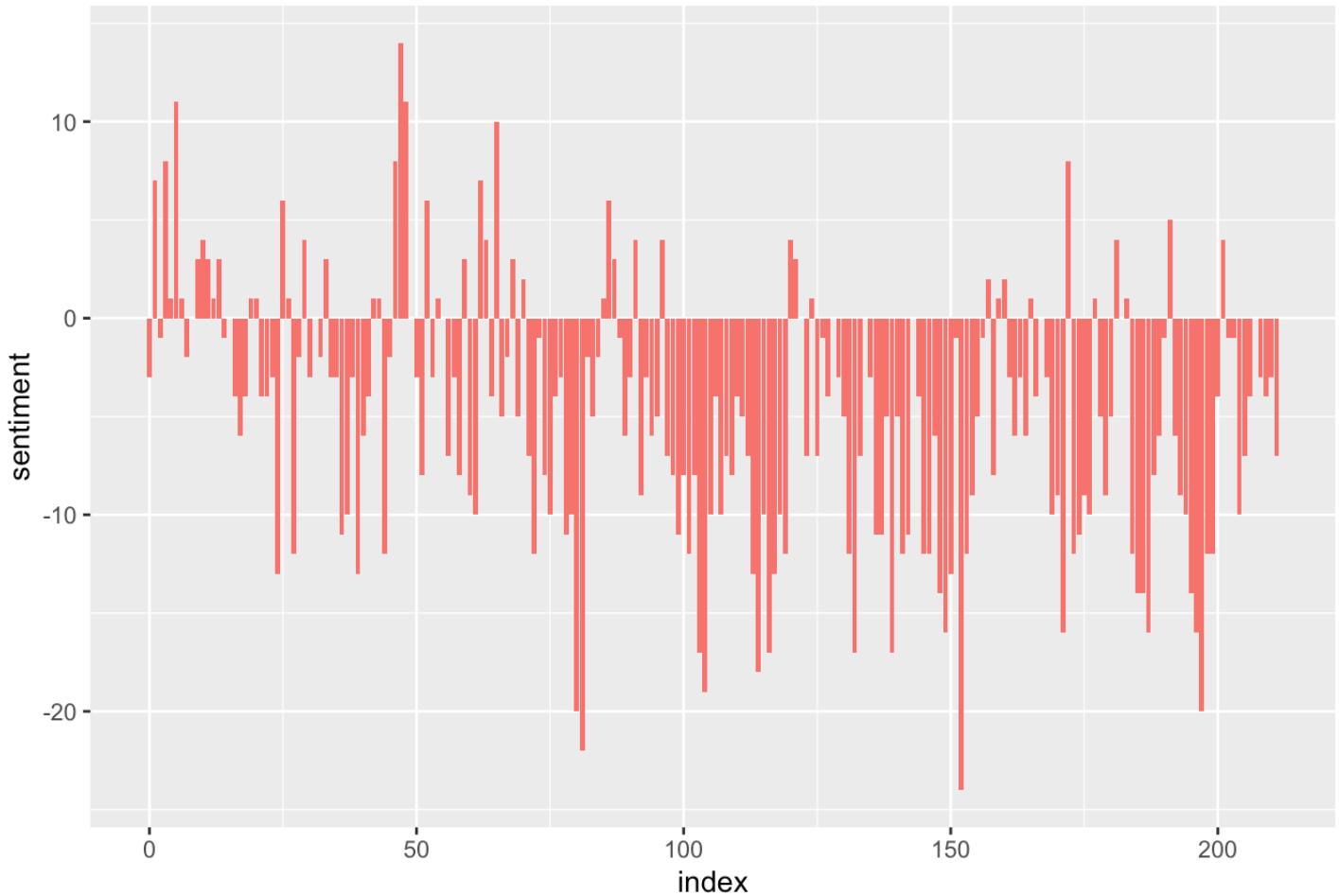
Extra creidt - Loughran

Next we explore one extra lexicon, Loughran, which is another available lexicon from tidytext package. However, this lexicon has a large disadvantage as it collected the lexicon from financial documents, which might no be as accurate as Bing et al. would be for novel literatures, such as the book we are using.

From the graph below, we found that from Loughran, the loughran lexicon has the smallest absolute values compared to other three lexicons, with mostly negative values. And the trend is less similar as other three methods to each other. The sentiment from loughran lexicon starts with small positive sentiment then goes

down to negative sentiment, with larger negative proportion and absolute values. This indicates that the book from Loughran is mostly negative.

Sentiment values of Uncle Tom's Cabin with Loughran Lexicon

Plotline

The major plotline of Uncle Tom's Cabin is about the life of Uncle Tom and other characters, as slaves. At the beginning, the story starts with a Kentucky farmer named Arthur Shelby needs to sell two of his slaves to slave trader to raise money to avoid losing everything. Uncle Tom and Harry are two slaves that are on the selling list.

Overhearding the decision on selling Harry, Aunt Chloe, Harry's mother, flew to the North, hoping to find freedom in Canada. The salve trade hires a slave hunter named Loker to bring Harry back to Kentucky while Harry makes his way to a Quaker settlement and receieved help transport them to safety.

This first two parts of the story are where the sentiment starts with positive sentiment as shown in Bing et al.'s graph, going down quickly then go back up a little bit to positive as Harry settled to Quaker settlement.

Meanwhile, Uncle Tom sadly leaves his family being transported to a slave market. On the boat, Tom meets an angelic little white girl named Eva, who quickly befriends him. Tom saveds Eva's life when she falls into the river and her father St. Clare gratefully agrees to buy Tom from the slave trader. Tom travels with the St. Clares to New Orleans, where he grows increasingly invaluable to the St. Clare household and increasingly close to Eva, with whom he shares a devout Christianity.

Up North, when the slave hunter Loker attempts to capture Eliza, he was shot in the side, and the other slave hunters retreat. However, Eliza persuaded Quakers to bring Loker to the next settlement to heal.

This second part concords with the sentiment plot that the there are up and down, but mostly positive sentiment, as Tom was able to be friend with the girl, being bought by grateful father and become invaluable to the new household. While up north, the conflict and still the nice of Eliza shows the positive sentiment.

After Tom has lived with the St. Clares for two years, Eva becomes weaken then dies, with a vision of heaven before her. Her death has a profound effect on everyone who knew her: Ophelia resolves to love the slaves, Topsy learns to trust and feel attached to others, and St. Clare decides to set Tom free. However, before he can act on his decision, St. Clare is stabbed to death while trying to settle a brawl. As he dies, he at last finds God and goes to be reunited with his mother in heaven.

St. Clare's cruel wife, Marie, sells Tom to a vicious plantation owner named Simon Legree. Tom is taken to rural Louisiana with a group of new slaves, including Emmeline, whom the demonic Legree has purchased to use as a sex slave, replacing his previous sex slave Cassy. Legree takes a strong dislike to Tom when Tom refuses to whip a fellow slave as ordered. Tom receives a severe beating, and Legree resolves to crush his faith in God. Tom meets Cassy, and hears her story. Separated from her daughter by slavery, she became pregnant again but killed the child because she could not stand to have another child taken from her.

This later part, the plotline also matches the sentiment plot. As the most story line is about the death of both Eva St. Clare, and resell of Tom to a much worse situation. The small positive part reflects the image of God before Eva dies and St. Clare finds God.

Around this time, with the help of Tom Loker—now a changed man after being healed by the Quakers—George, Eliza, and Harry at last cross over into Canada from Lake Erie and obtain their freedom. In Louisiana, Tom's faith is sorely tested by his hardships, and he nearly ceases to believe. He has two visions, however—one of Christ and one of Eva—which renew his spiritual strength and give him the courage to withstand Legree's torments. He encourages Cassy to escape. She does so, taking Emmeline with her, after she devises a ruse in which she and Emmeline pretend to be ghosts. When Tom refuses to tell Legree where Cassy and Emmeline have gone, Legree orders his overseers to beat him. When Tom is near death, he forgives Legree and the overseers. George Shelby arrives with money in hand to buy Tom's freedom, but he is too late. He can only watch as Tom dies a martyr's death.

Taking a boat toward freedom, Cassy and Emmeline meet George Harris's sister and travel with her to Canada, where Cassy realizes that Eliza is her long-lost daughter. The newly reunited family travels to France and decides to move to Liberia, the African nation created for former American slaves. George Shelby returns to the Kentucky farm, where, after his father's death, he sets all the slaves free in honor of Tom's memory. He urges them to think on Tom's sacrifice every time they look at his cabin and to lead a pious Christian life, just as Tom did.

This last part also matches the sentiment plot as Tom experiences extreme harships and his death, but as the story goes forward, other slaves start to have new free life in Canada and the Christian life.