Task Three

Ingest book to tnum test2 space

```
tnum.authorize("mssp1.bu.edu")
```

```
## Available spaces: testspace, MEPED, alion-rf, shared-testspace, test2, alion, NCNM
, ED-900-Workshop, ED900_1, MSSP-TEXT-1, test3
```

```
## Numberspace set to: testspace
```

```
tnum.setSpace("test2")

uncletomref <- gutenberg_works(title == "Uncle Tom's Cabin")

### download the text
uncletom <- gutenberg_download(gutenberg_id = 203) ## download uncle tom's cabin
```

```
## Determining mirror for Project Gutenberg from http://www.gutenberg.org/robot/harve
st
```

```
## Using mirror http://aleph.gutenberg.org
```

```
#uncletomtxt <- readLines("pg203.txt")
source("Book2TN-v6A-1.R")
#tnBooksFromLines(uncletom$text, "stowe/uncletom")
```

Now, verify the book has been ingested properly

```
tnum.getDBPathList(taxonomy="subject", levels=2)
```

Sentimentr

With sentimentr, we produced a histogram to demonstrate the sentiment of the text. According to the histogram, most of the sentiment is positive.

From the sentiment histogram below, we can see that at sentence level, the sentiment is majorly neutral, and more skewed to positive sentiment.

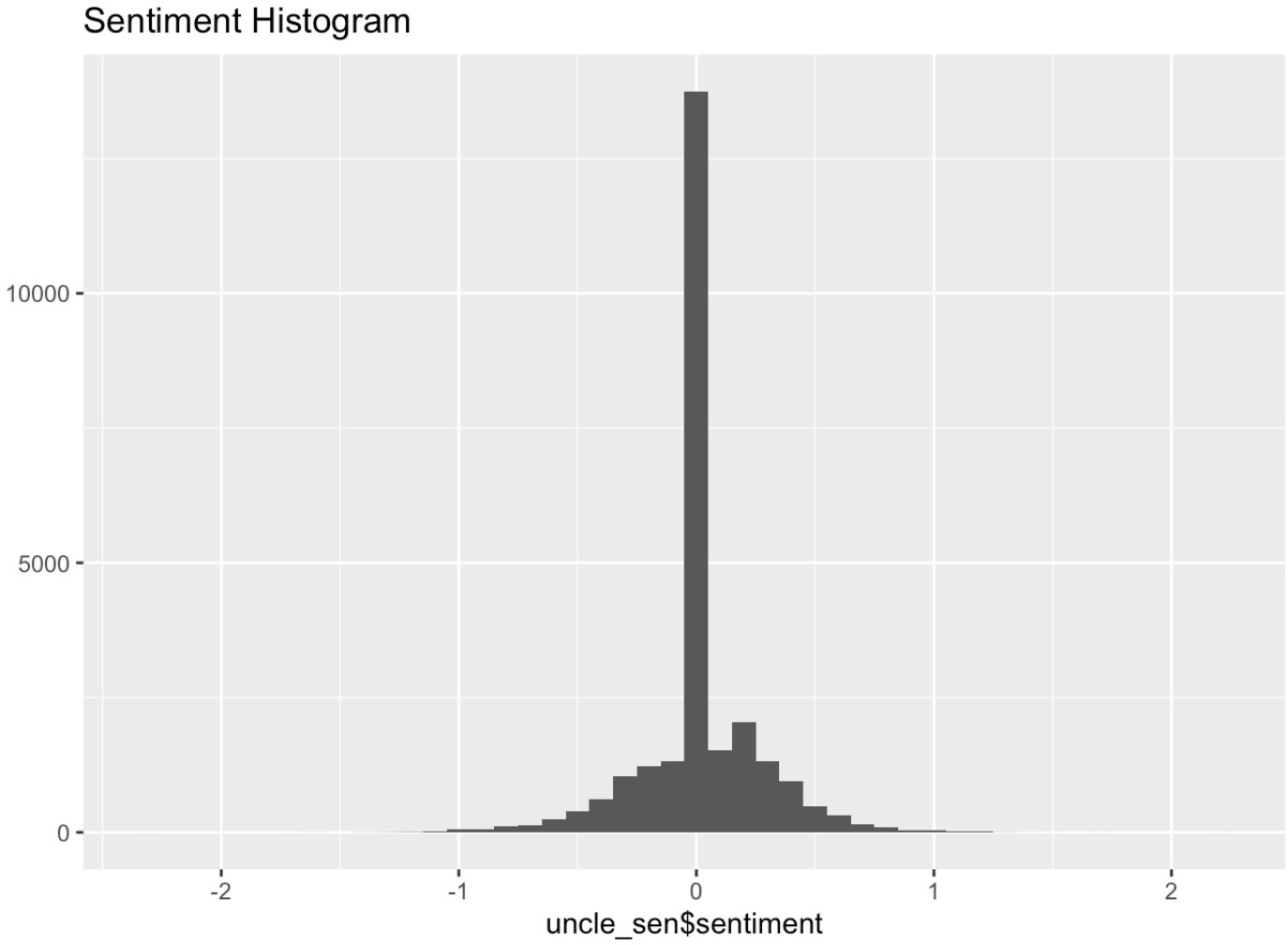
```
uncle_sen<-sentiment(uncletom$text)
```

```
## Warning: Each time `sentiment` is run it has to do sentence boundary disambiguatio
n when a
## raw `character` vector is passed to `text.var`. This may be costly of time and
## memory.  It is highly recommended that the user first runs the raw `character`
## vector through the `get_sentences` function.
```

```
summary(uncle_sen$ave_sentiment)
```

```
## Length  Class   Mode
##      0   NULL   NULL
```

```
qplot(uncle_sen$sentiment,
      geom = "histogram", binwidth = 0.1,
      main = "Sentiment Histogram")
```

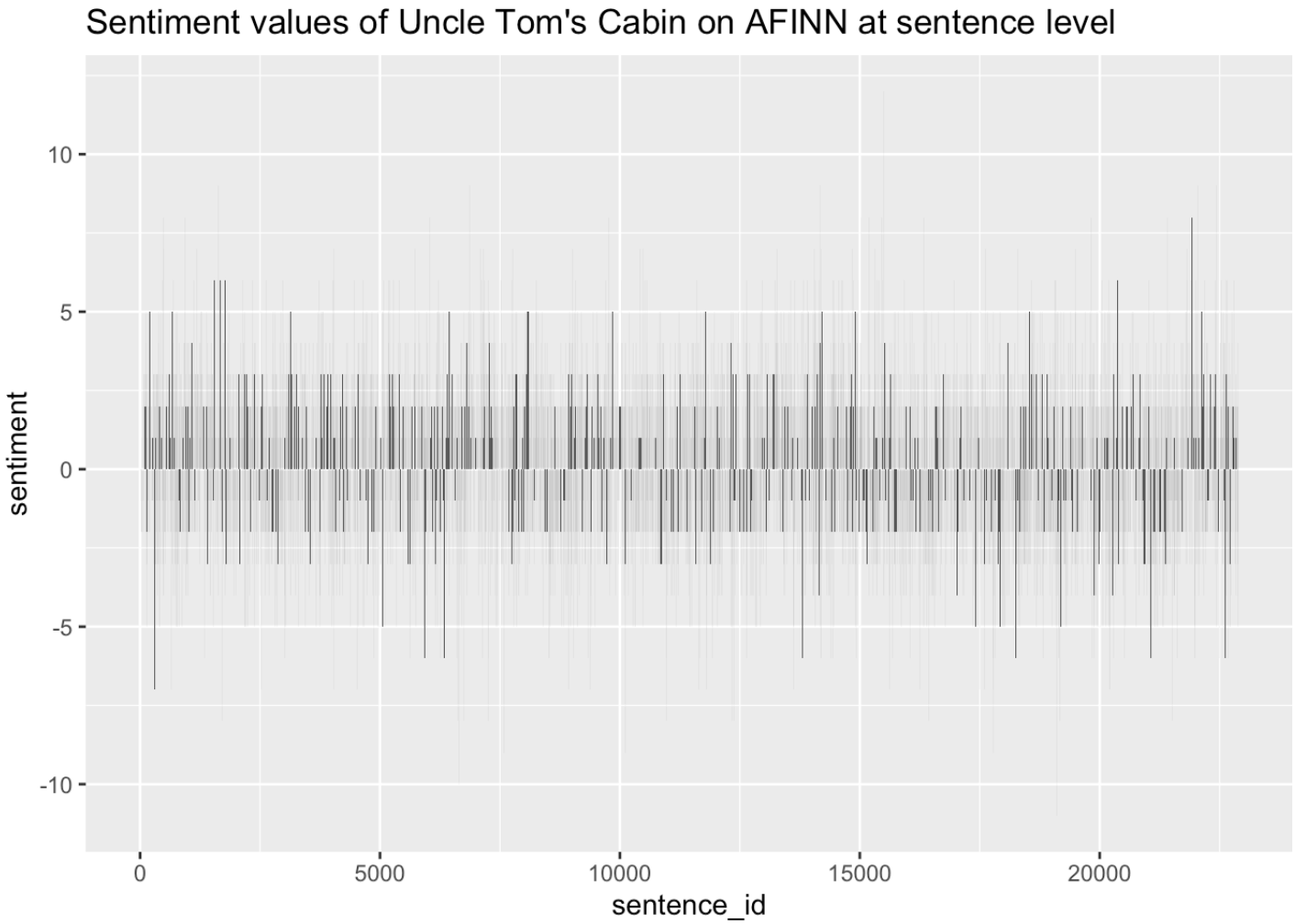


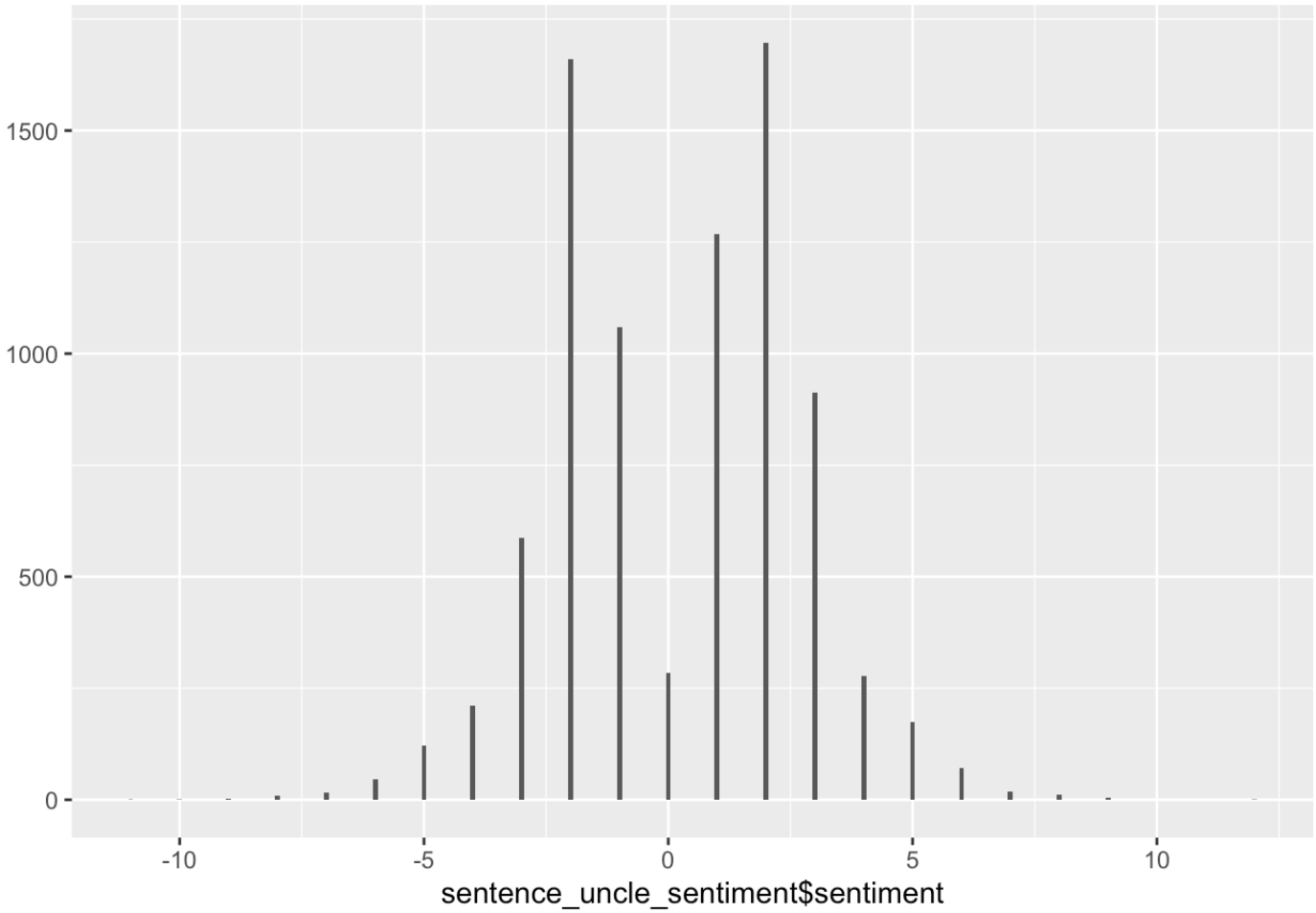
Bag of words comaprsion

In the task two, the analysis was conducted on a word basis, to compare the bag of words analysis and sentimentr, the bag of words need to adjusted to sentence level. Therefore, following we will conduct a sentence level analysis.

We used the sentence level analysis to produce over the sentence along with the plotline, what is the sentiment. We also plot the histogram based on sentiment. According to the histogram, the sentiment of sentences is almost symetric on both side of absolute neutral (0). With slightly skewed to the right,

representing the positive sentiment.




Sentiment Histogram at sentence level
1500
1000
500
0
-10
-5
0
5
10
sentence_uncle_sentiment$sentiment