Text Mining - Middlemarch vs. Wuthering Heights

Mark Blackmore

January 30, 2018

Total Words

```
# Middlemarch
nrow(tidy_middlemarch)
```

```
## [1] 320374
```

```
# Wuthering Heights
nrow(tidy_wuthering)
```

```
## [1] 117111
```

Unique Words

```
# Middlemarch
(repo_count_middle <- tidy_middlemarch %>%
    summarise(keys = n_distinct(word)))
```

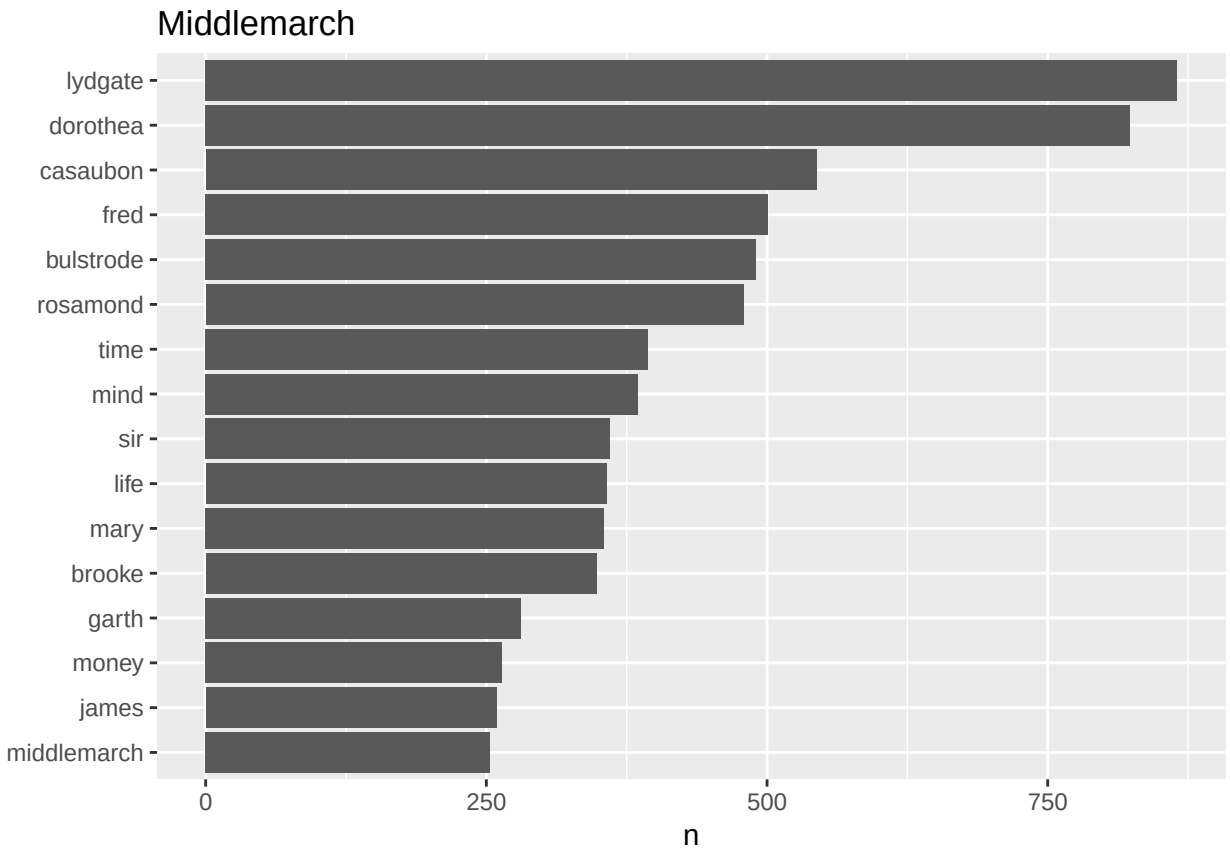
```
## # A tibble: 1 x 1
##    keys
##   <int>
## 1 15675
```

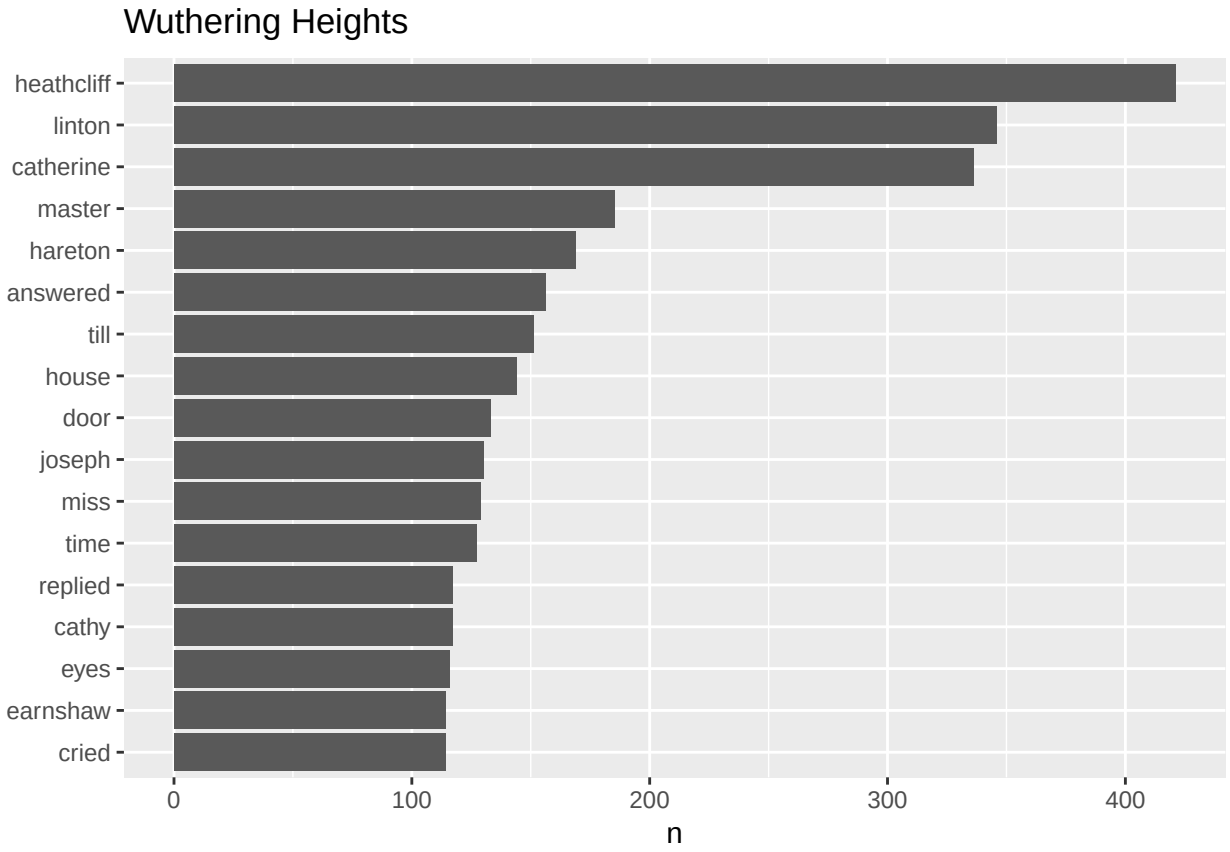
```
# Wuthering Heights
(repo_count_wuthering <- tidy_wuthering %>%
    summarise(keys = n_distinct(word)))
```

```
## # A tibble: 1 x 1
##    keys
##   <int>
## 1  9486
```

Most common words, excluding stopwords (e.g. is, the, are,. . .)

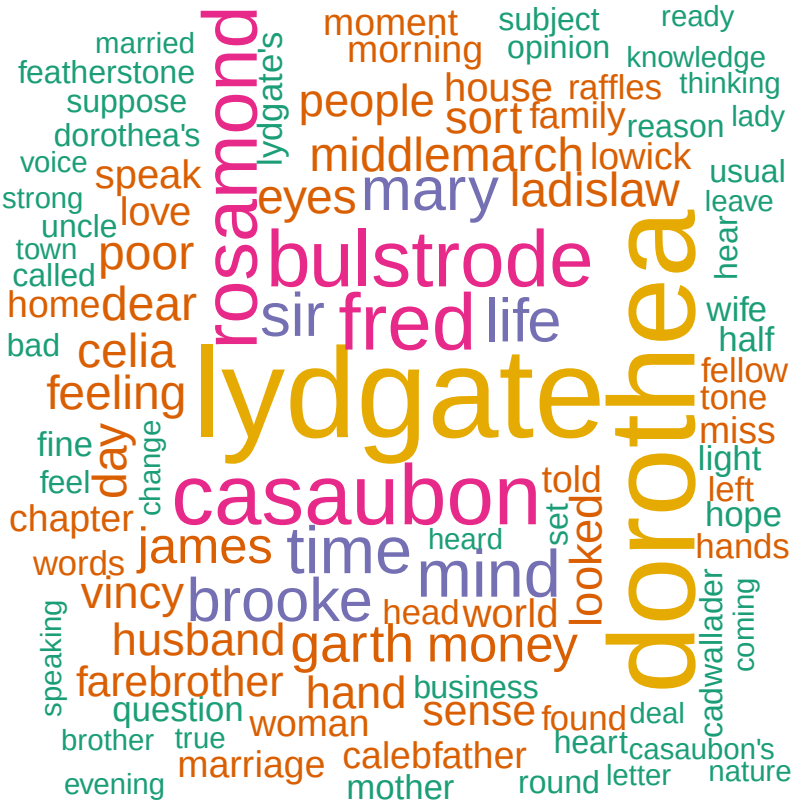
```
## Joining, by = "word"
## Joining, by = "word"
```


Middlemarch
lydgate
dorothea
casaubon
fred
bulstrode
rosamond
time
mind
sir
life
mary
brooke
garth
money
james
middlemarch
0
250
500
750
n

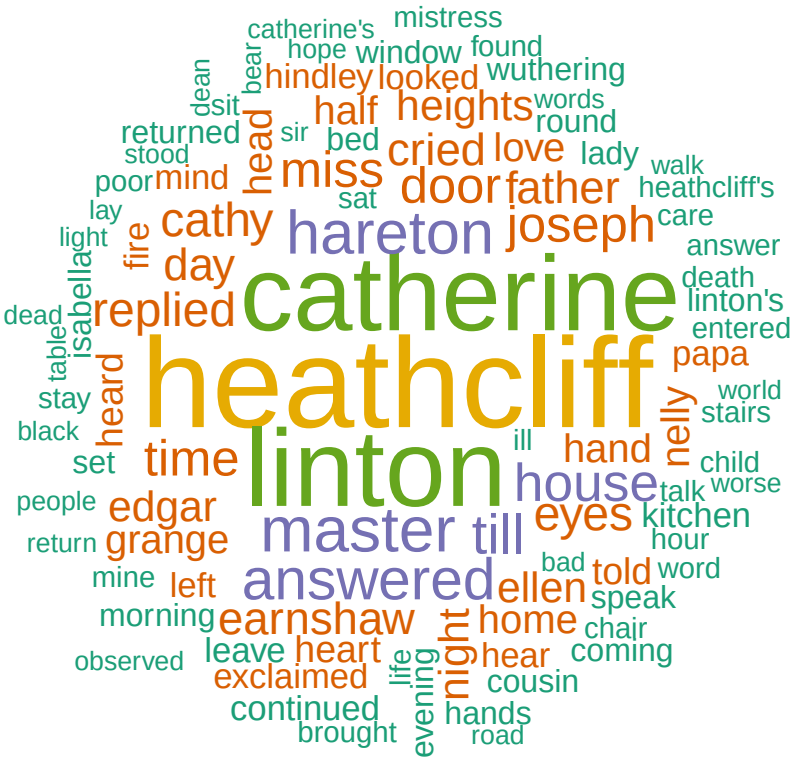

Wuthering Heights
heathcliff
linton
catherine
master
hareton
answered
till
house
door
joseph
miss
time
replied
cathy
eyes
earnshaw
cried
0
100
200
300
400
n

Wordclouds

Middlemarch

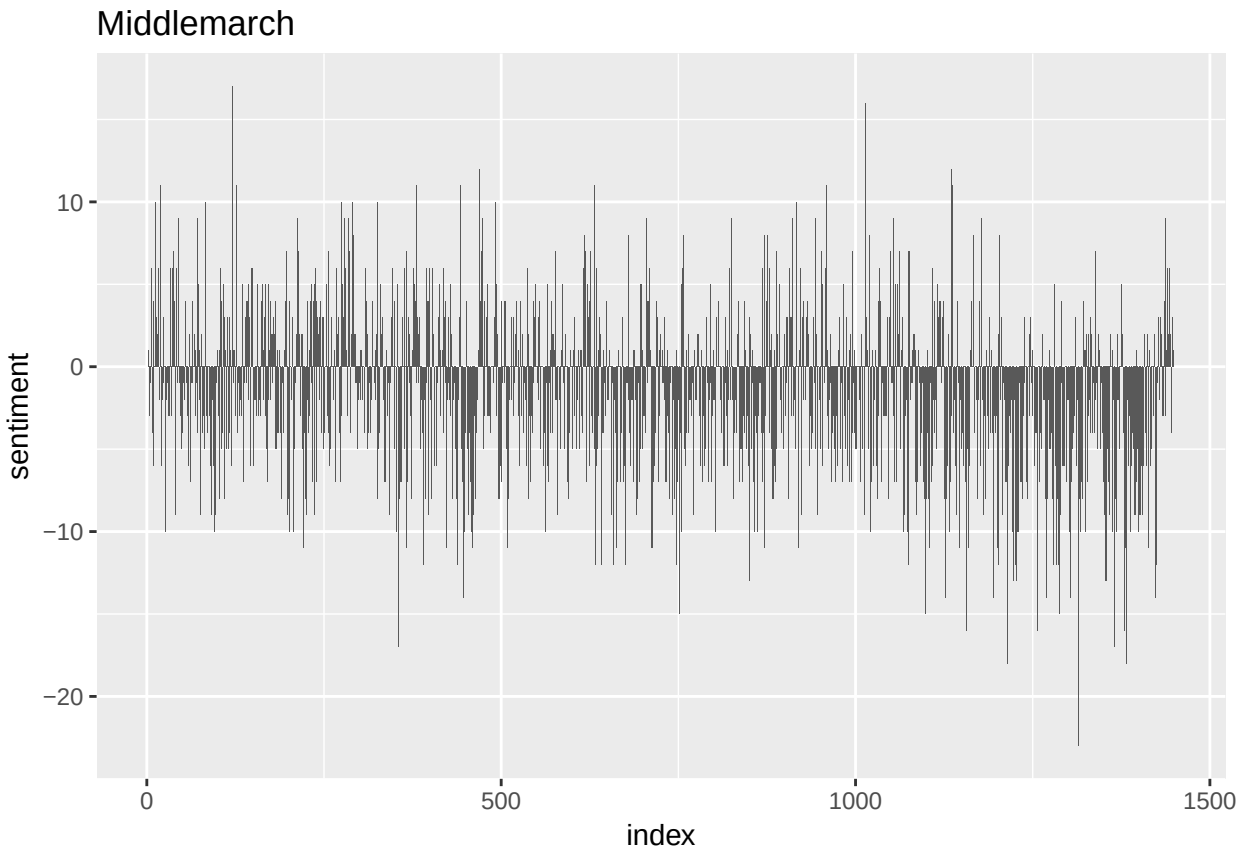


Wuthering Heights

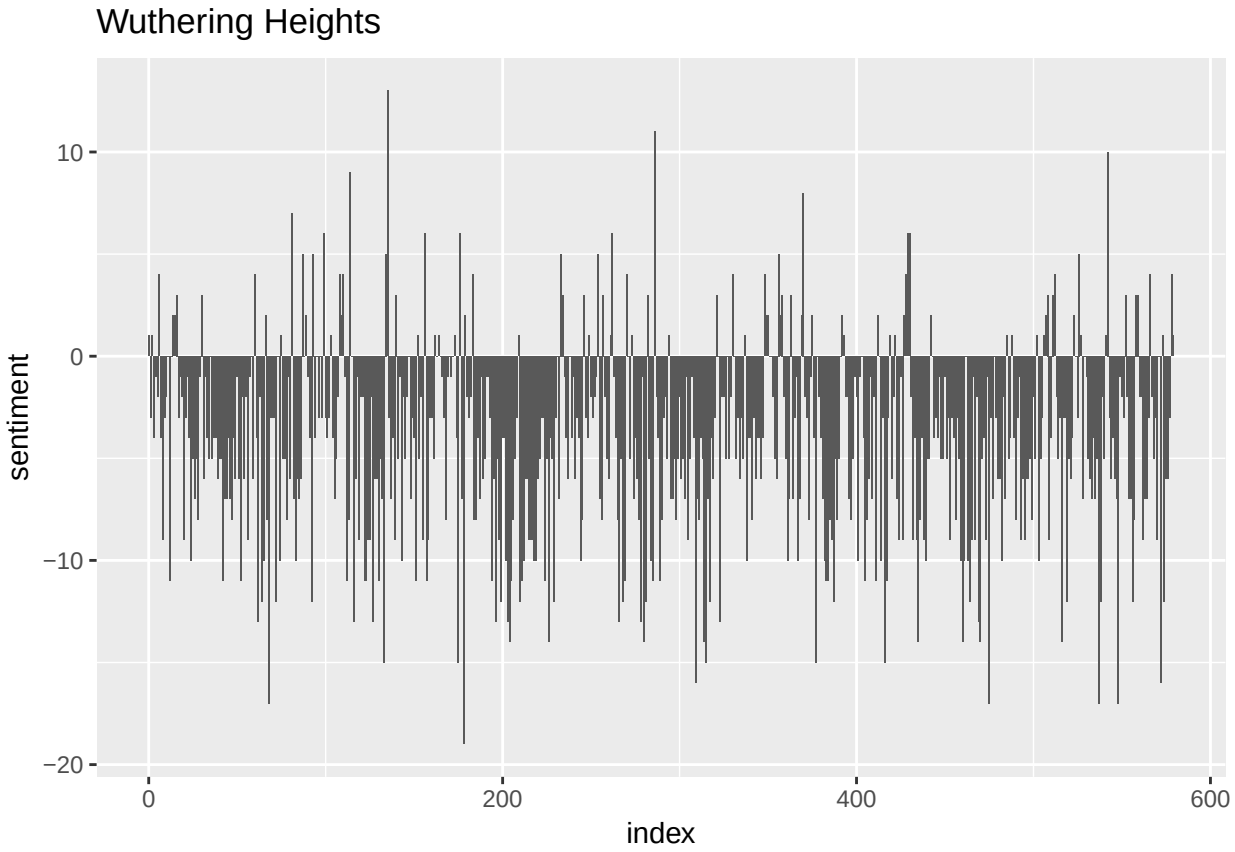


Sentimnt: Positive Words minus Negative Words by Paragraph

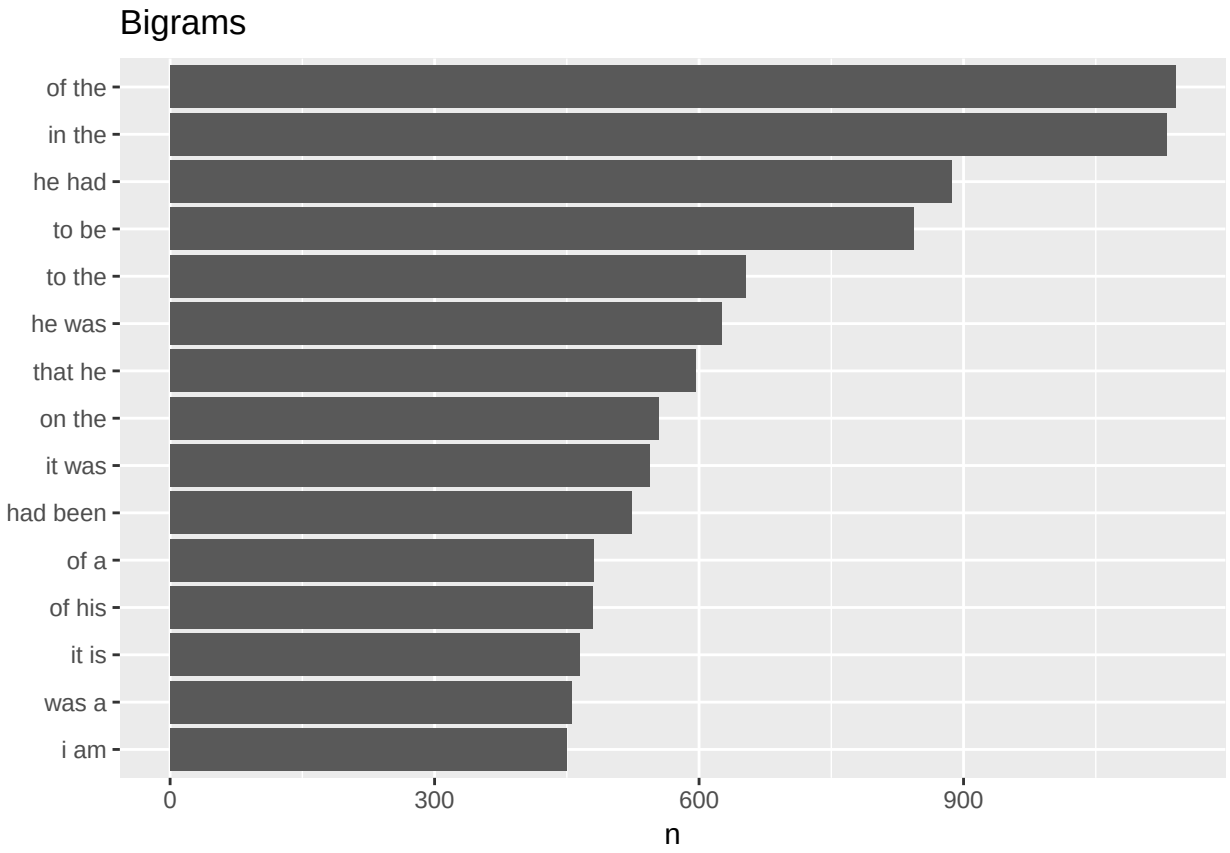
```
## Joining, by = "word"
```

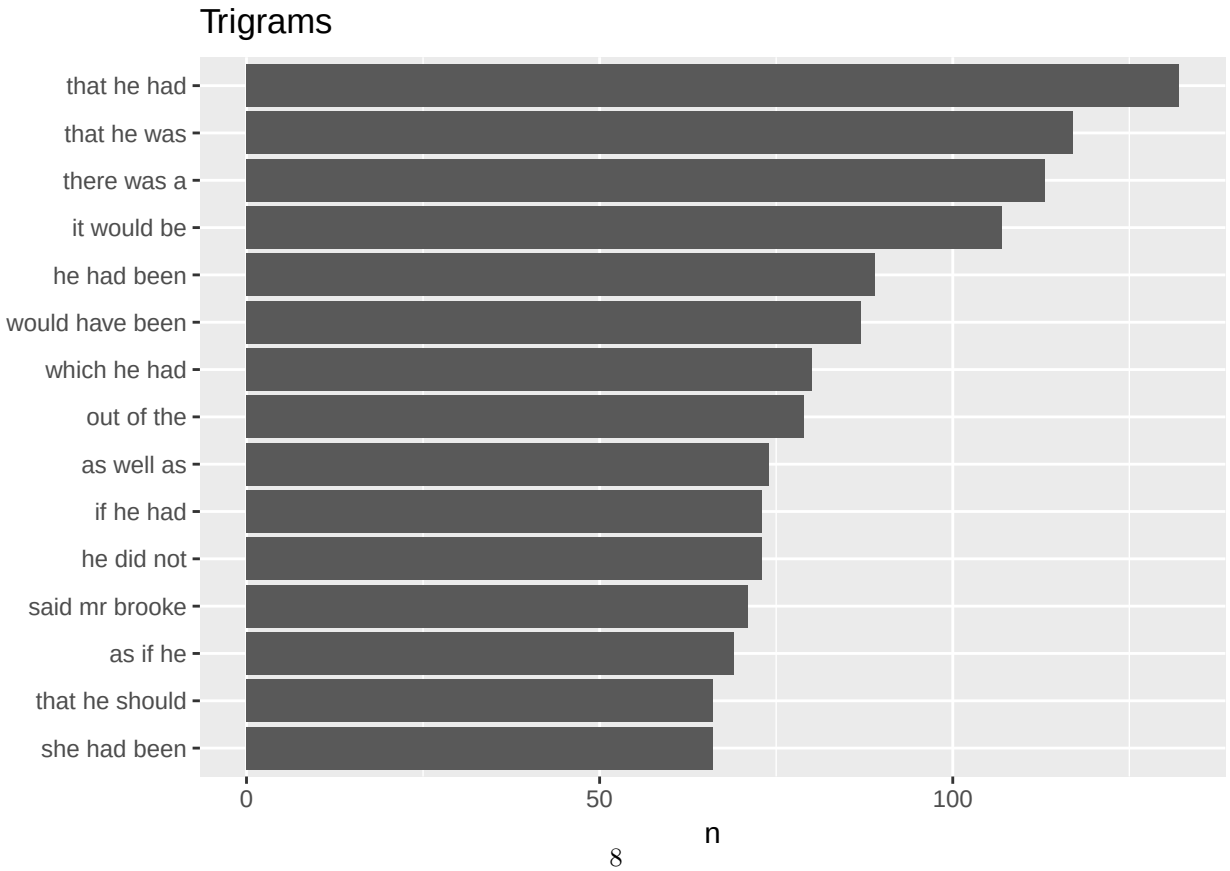


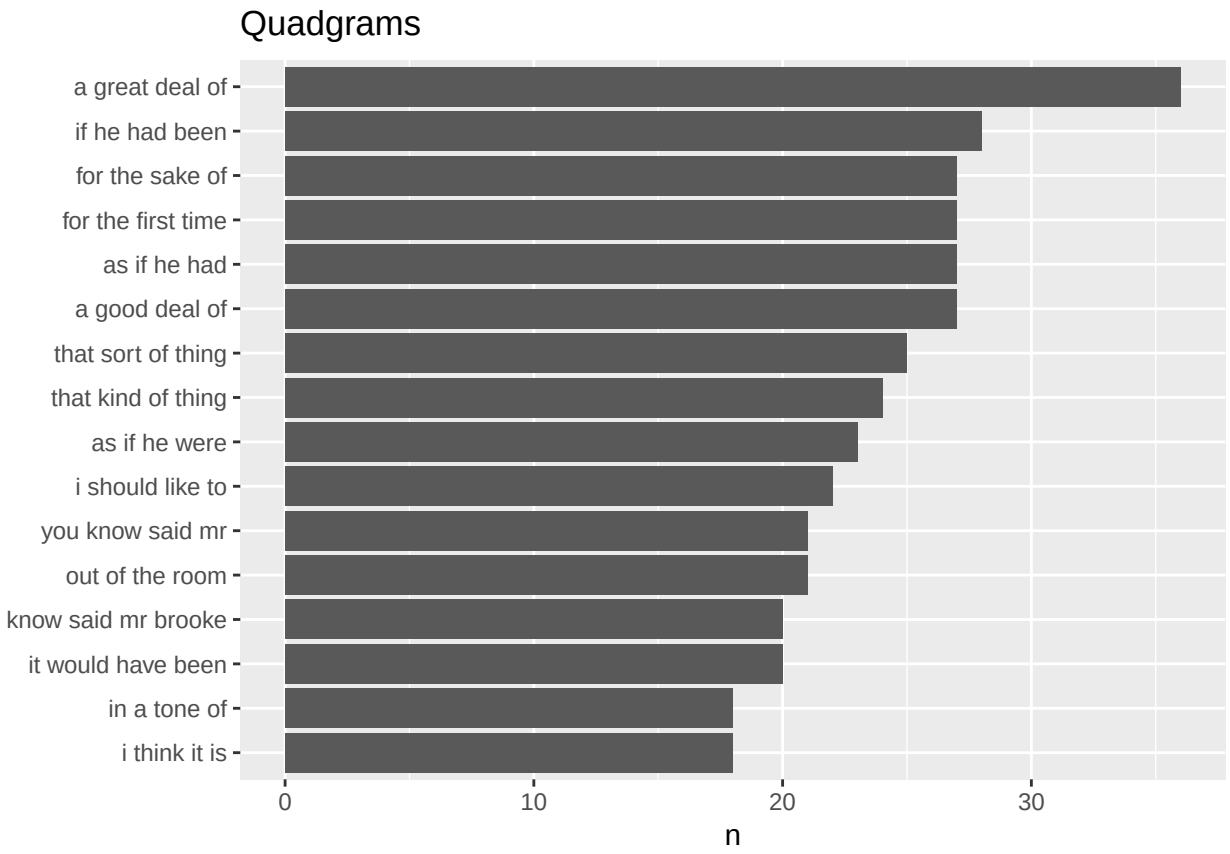

```
## Joining, by = "word"
```

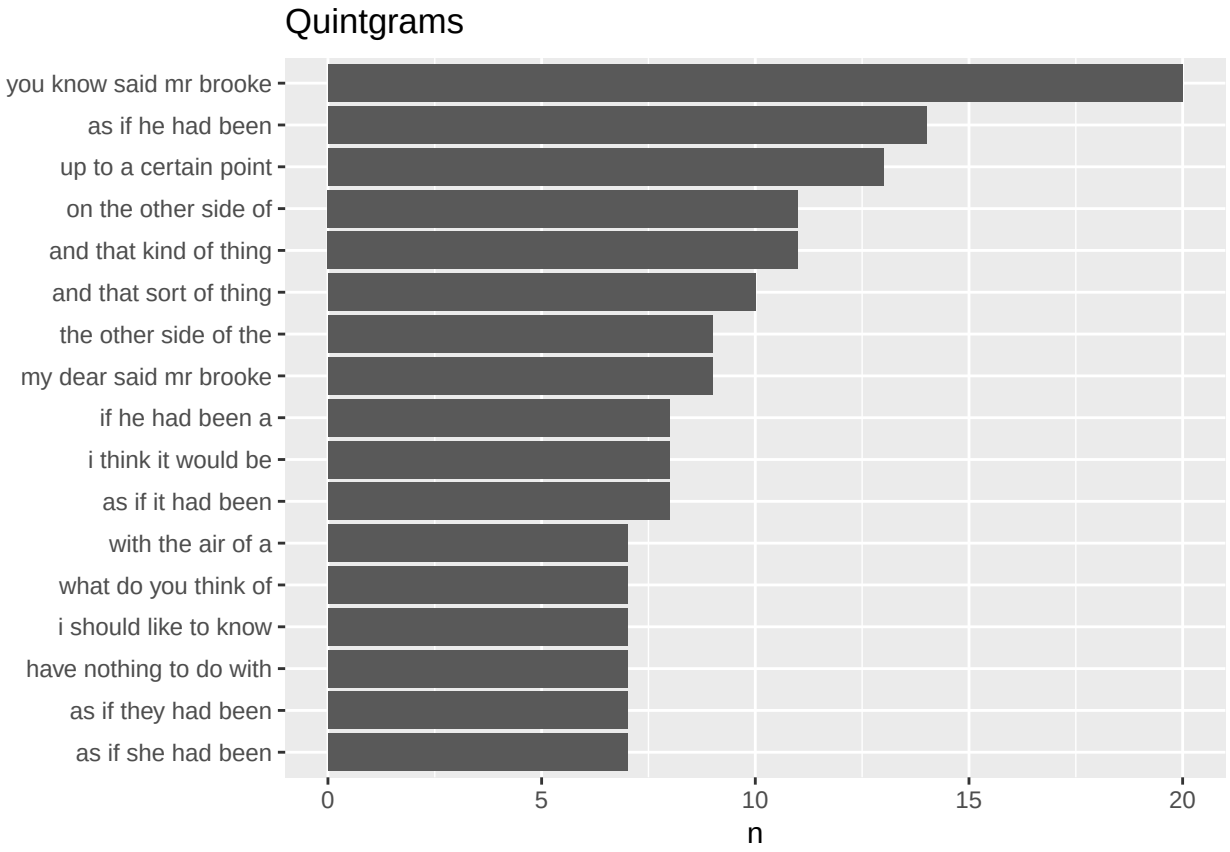

Wuthering Heights
10
0
−10
−20
sentiment
0
200
400
600
index

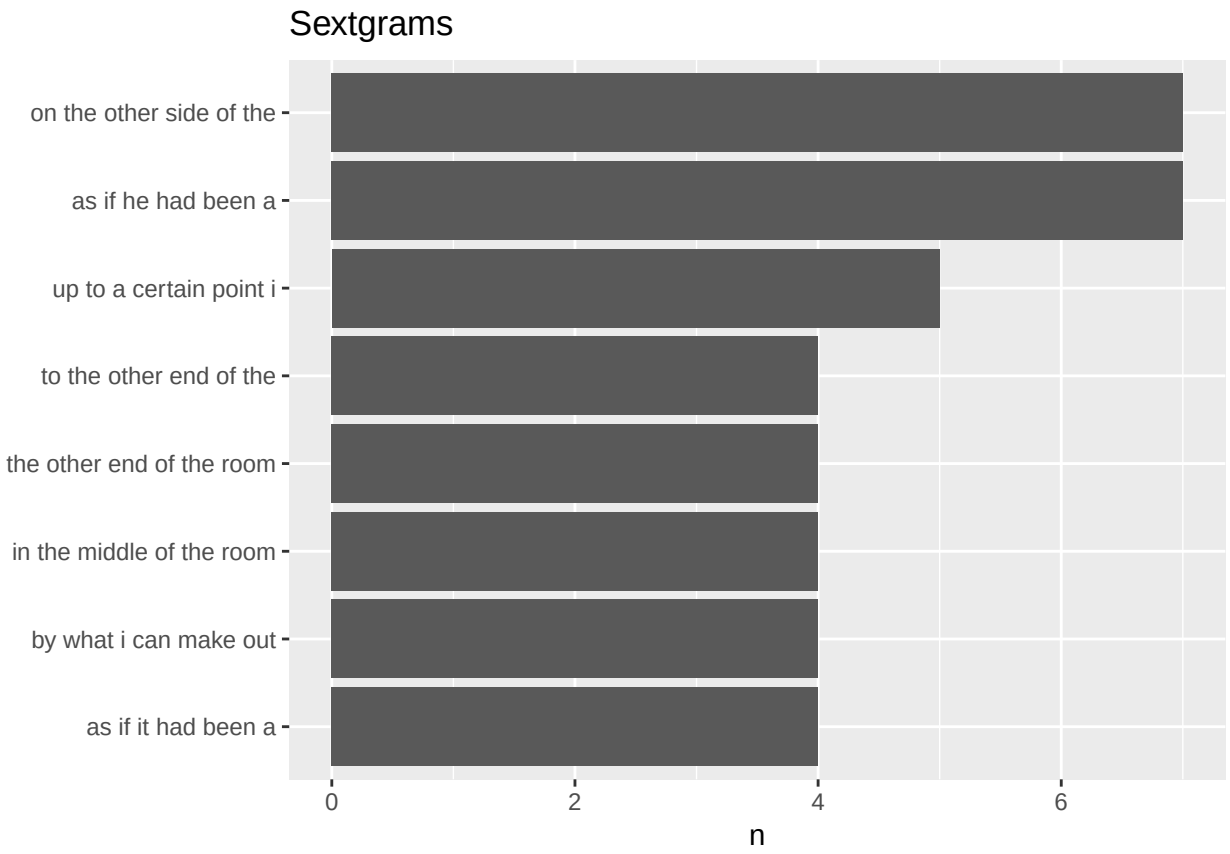
Middlemarch n-grams






Quadgrams
a great deal of
if he had been
for the sake of
for the first time
as if he had
a good deal of
that sort of thing
that kind of thing
as if he were
i should like to
you know said mr
out of the room
know said mr brooke
it would have been
in a tone of
i think it is
0
10
20
30
n


Quintgrams
you know said mr brooke
as if he had been
up to a certain point
on the other side of
and that kind of thing
and that sort of thing
the other side of the
my dear said mr brooke
if he had been a
i think it would be
as if it had been
with the air of a
what do you think of
i should like to know
have nothing to do with
as if they had been
as if she had been
0
5
10
15
20
n


Sextgrams
on the other side of the
as if he had been a
up to a certain point i
to the other end of the
the other end of the room
in the middle of the room
by what i can make out
as if it had been a
0
2
4
6
n