

Content Analysis of Scaling Up Nutrition (SUN) Movement Progress Reports from 2011-2017

Ernest Guevarra

17/05/2018

Re-structure dataset into a one-token-per-row format

```
tidy_reports <- progress_reports() %>%  
  unnest_tokens(word, text)  
  
tidy_reports  
  
## # A tibble: 354,209 x 5  
##   page linenumber chapter year word  
##   <int>      <int>    <int> <int> <chr>  
## 1     3          1      0  2011 preface  
## 2     3          2      0  2011 one  
## 3     3          2      0  2011 year  
## 4     3          2      0  2011 ago  
## 5     3          2      0  2011 i  
## 6     3          2      0  2011 joined  
## 7     3          2      0  2011 a  
## 8     3          2      0  2011 group  
## 9     3          2      0  2011 of  
## 10    3          2      0  2011 leaders  
## # ... with 354,199 more rows
```

Remove stop words - words not useful in analysis

```
data(stop_words)  
  
tidy_reports <- tidy_reports %>%  
  anti_join(stop_words)
```

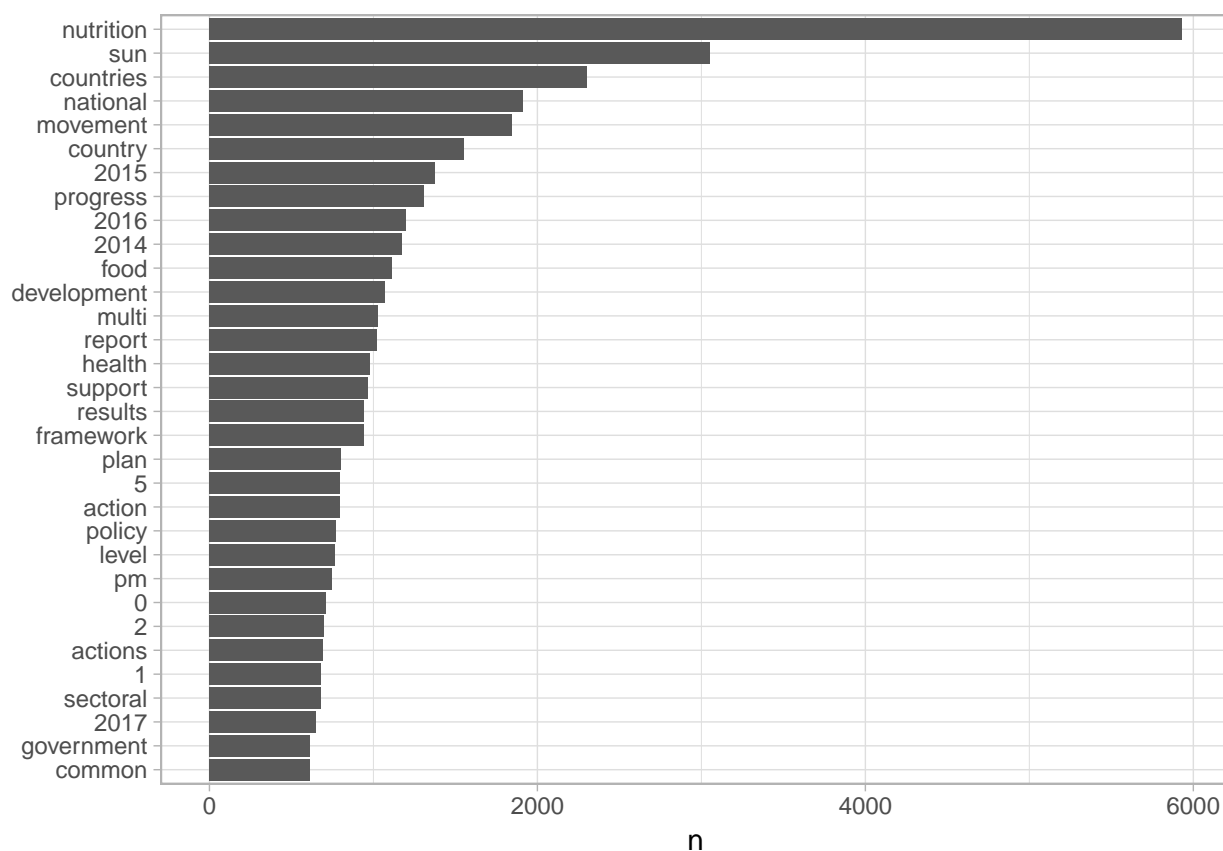
Find the most common words in all the reports as a whole

```
tidy_reports %>%  
  count(word, sort = TRUE)  
  
## # A tibble: 12,659 x 2  
##   word      n  
##   <chr>    <int>  
## 1 nutrition 5932
```

```
## 2 sun 3050
## 3 countries 2302
## 4 national 1914
## 5 movement 1843
## 6 country 1552
## 7 2015 1377
## 8 progress 1308
## 9 2016 1199
## 10 2014 1174
## # ... with 12,649 more rows
```

Visualise the most common words

```
tidy_reports %>%
  count(word, sort = TRUE) %>%
  filter(n > 600) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()
```



```
nrc_joy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")

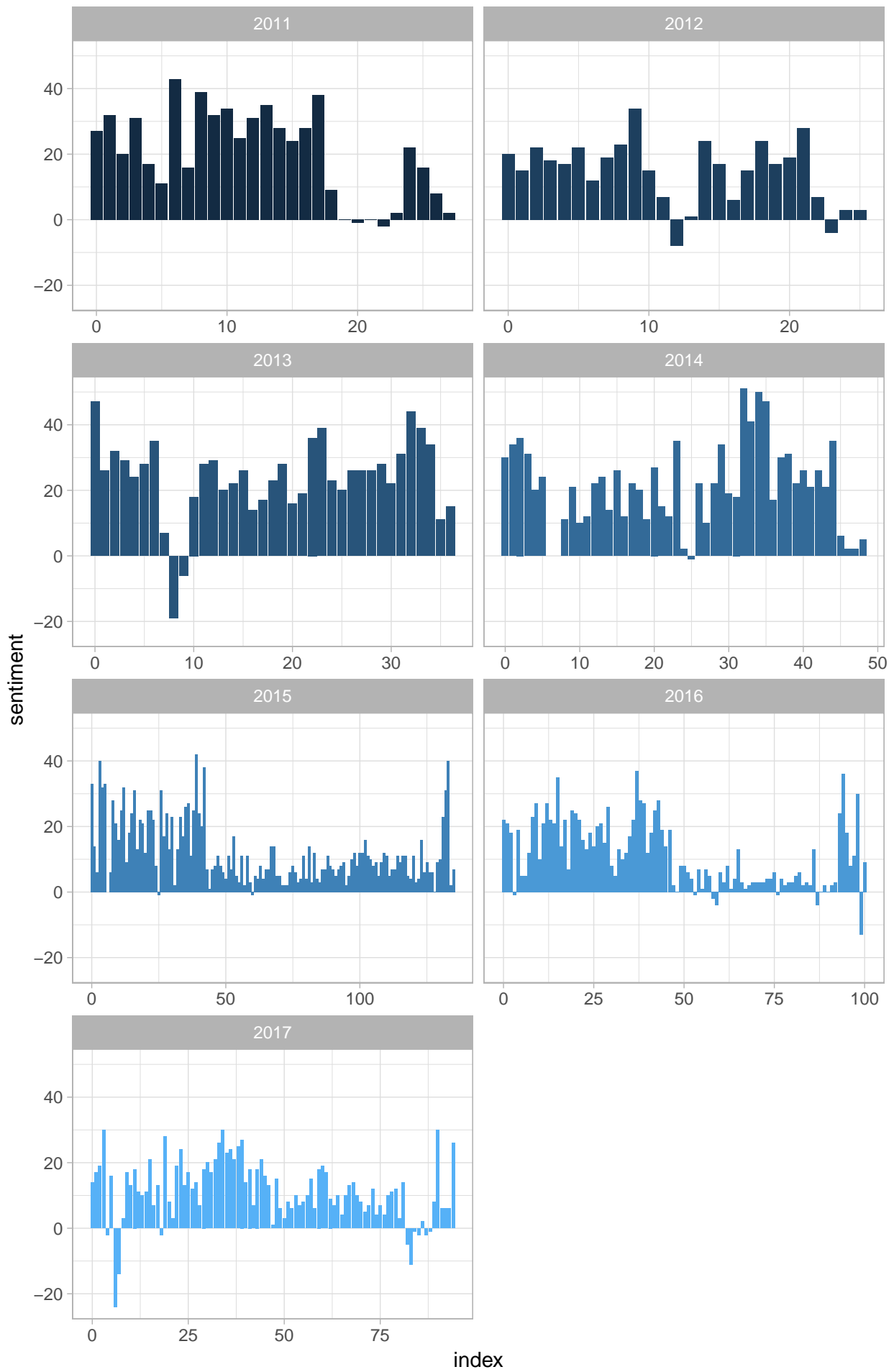
tidy_reports %>%
  filter(year == 2011) %>%
```

```
inner_join(nrc_joy) %>%  
count(word, sort = TRUE)
```

```
## # A tibble: 62 x 2  
##   word      n  
##   <chr>    <int>  
## 1 sun      236  
## 2 food     105  
## 3 progress  72  
## 4 child     35  
## 5 resources 30  
## 6 weight    19  
## 7 birth     18  
## 8 improve   17  
## 9 established 15  
## 10 engaged  13  
## # ... with 52 more rows
```

```
progress_sentiment <- tidy_reports %>%  
  inner_join(get_sentiments("bing")) %>%  
  #count(year, index = linenummer %/% 80, sentiment) %>%  
  count(year, index = page, sentiment) %>%  
  spread(sentiment, n, fill = 0) %>%  
  mutate(sentiment = positive - negative)
```

```
ggplot(progress_sentiment, aes(index, sentiment, fill = year)) +  
  geom_col(show.legend = FALSE) +  
  facet_wrap(~year, ncol = 2, scales = "free_x")
```



```
report_2011 <- tidy_reports %>%
  filter(year == 2011)
```

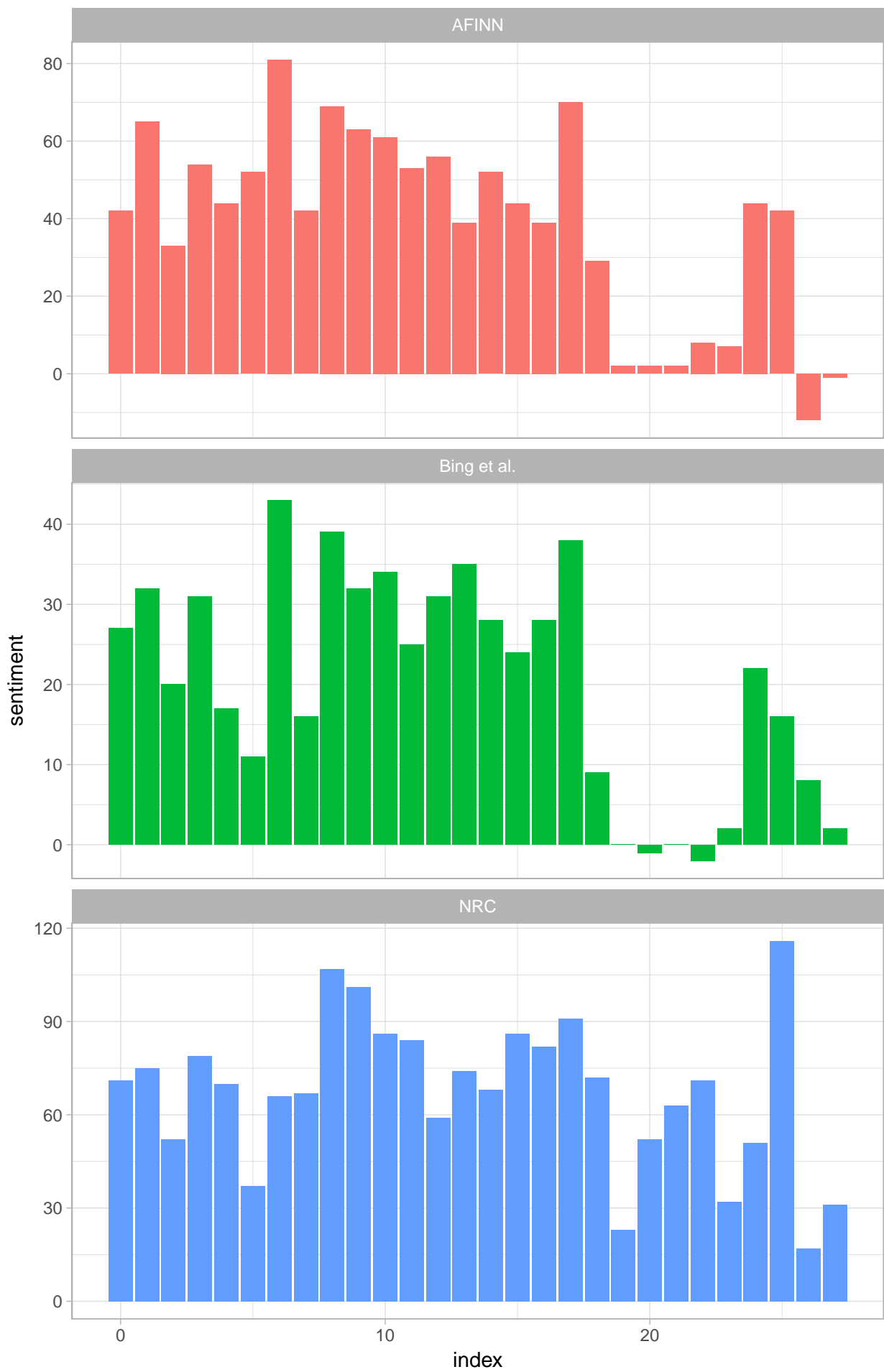
```
report_2011
```

```
## # A tibble: 14,590 x 5
##   page linenumber chapter year word
##   <int>      <int>    <int> <int> <chr>
## 1     3          1        0  2011 preface
## 2     3          2        0  2011 ago
## 3     3          2        0  2011 joined
## 4     3          2        0  2011 leaders
## 5     3          2        0  2011 pledging
## 6     3          2        0  2011 address
## 7     3          2        0  2011 global
## 8     3          2        0  2011 burden
## 9     3          2        0  2011 nutrition
## 10    3          3        0  2011 set
## # ... with 14,580 more rows
```

```
afinn <- report_2011 %>%
  inner_join(get_sentiments("afinn")) %>%
  #group_by(index = linenumber %/% 80) %>%
  group_by(index = page) %>%
  summarise(sentiment = sum(score)) %>%
  mutate(method = "AFINN")
```

```
bing_and_nrc <- bind_rows(report_2011 %>%
  inner_join(get_sentiments("bing")) %>%
  mutate(method = "Bing et al."),
  report_2011 %>%
  inner_join(get_sentiments("nrc")) %>%
  filter(sentiment %in% c("positive",
    "negative"))) %>%
  mutate(method = "NRC")) %>%
  #count(method, index = linenumber %/% 80, sentiment) %>%
  count(method, index = page, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)
```

```
bind_rows(afinn,
  bing_and_nrc) %>%
  ggplot(aes(index, sentiment, fill = method)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~method, ncol = 1, scales = "free_y")
```

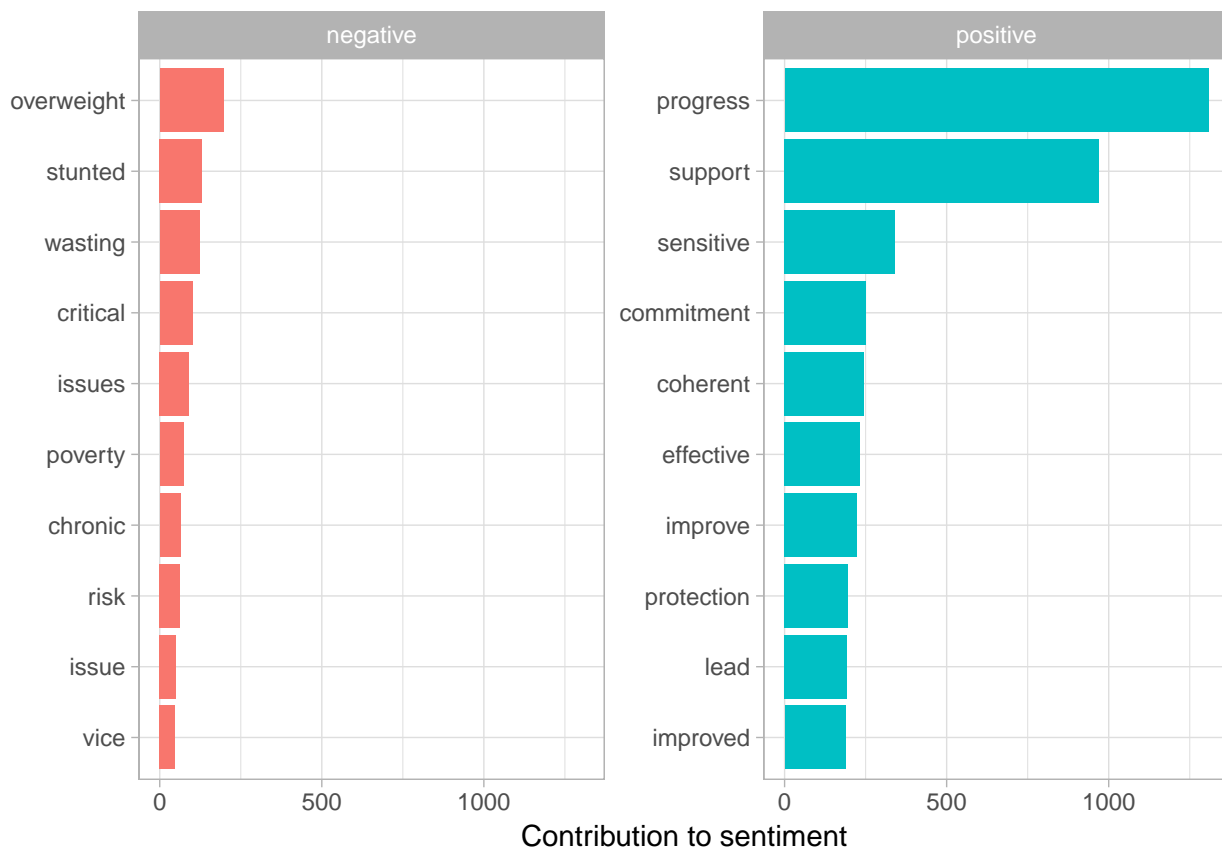


```
bing_word_counts <- tidy_reports %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
```

```
bing_word_counts
```

```
## # A tibble: 743 x 3
##   word      sentiment      n
##   <chr>      <chr>    <int>
## 1 progress  positive   1308
## 2 support   positive    969
## 3 sensitive positive    340
## 4 commitment positive    250
## 5 coherent  positive    244
## 6 effective positive    233
## 7 improve   positive    224
## 8 overweight negative    196
## 9 protection positive    195
## 10 lead      positive    192
## # ... with 733 more rows
```

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



```
library(wordcloud)

tidy_reports %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))
```

