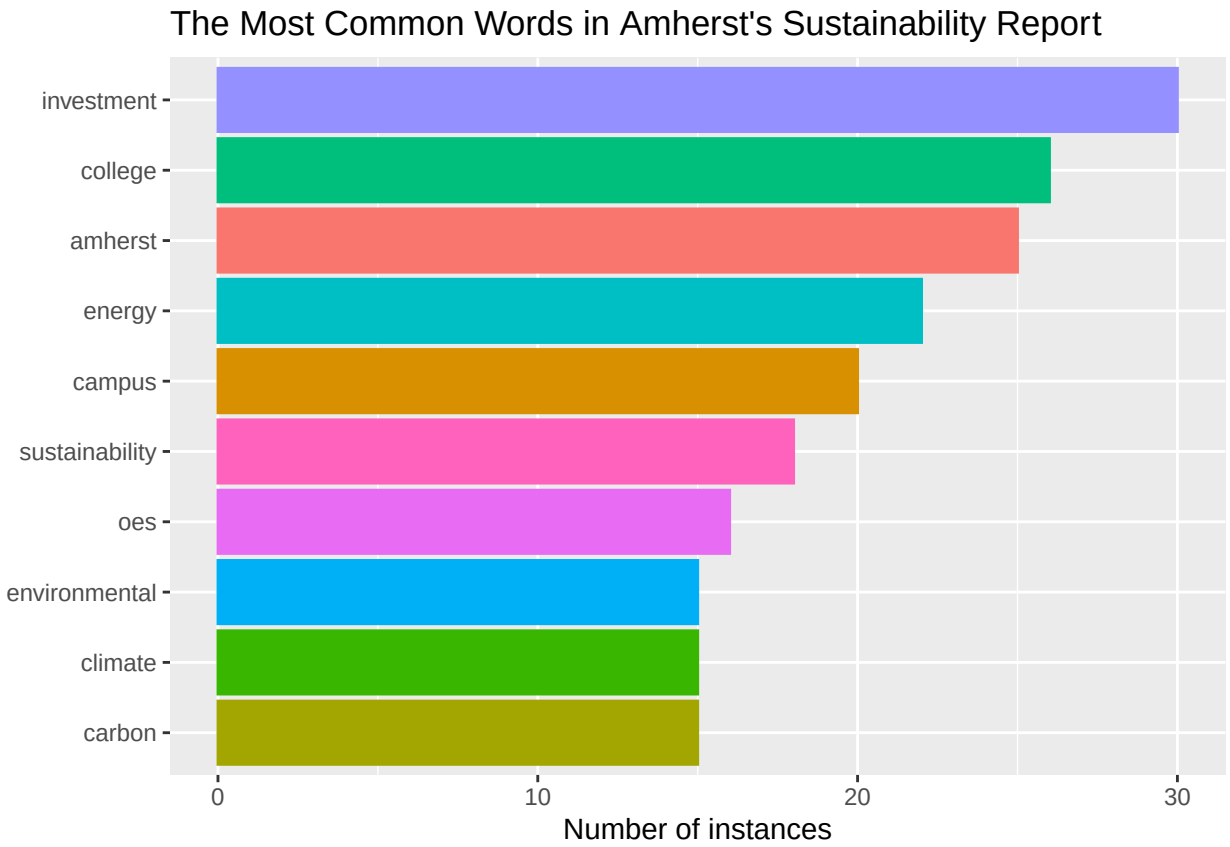
Sentiment Analysis

```
#Read in sustainability report
report <- read_csv("data/amherstsustainabilityreport.txt")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   x = col_character()
## )
```

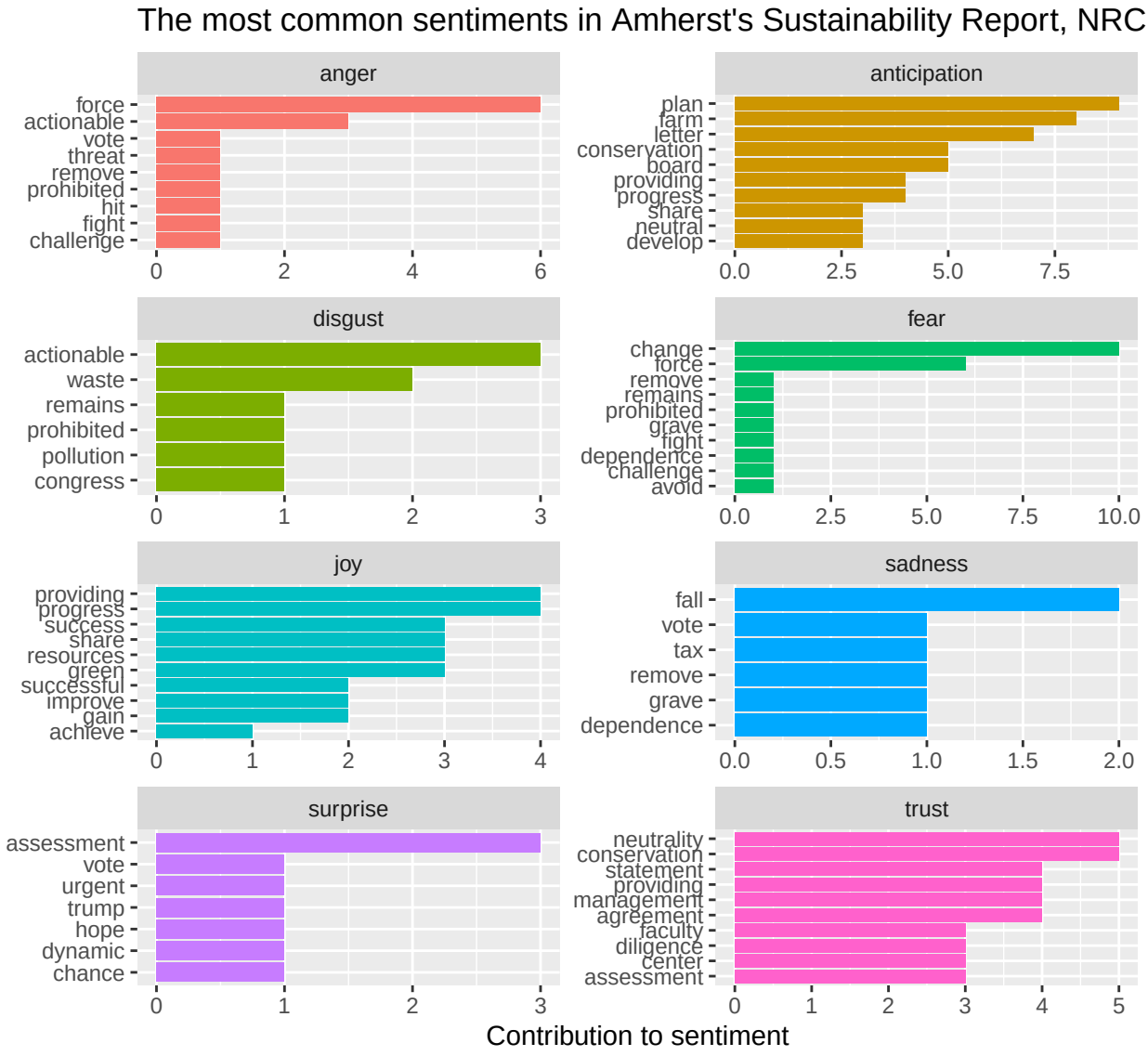
```
#Stop word data
data(stop_words)

#Wrangling
word_frequencies <- report %>%
  #Tokenize text into words
  unnest_tokens(output = word, input = x) %>%
  #Remove stop words
  anti_join(stop_words, by = "word") %>%
  #Word frequencies
  count(word, sort = TRUE)
```

```
#Common words plot
word_frequencies %>%
  slice(1:10) %>%
  ggplot(aes(x = reorder(word, n), y = n,
             color = word, fill = word)) +
  geom_col() +
  coord_flip() +
  guides(color = "none", fill = "none") +
  labs(x = NULL,
       y = "Number of instances",
       title = "The Most Common Words in Amherst's Sustainability Report")
```

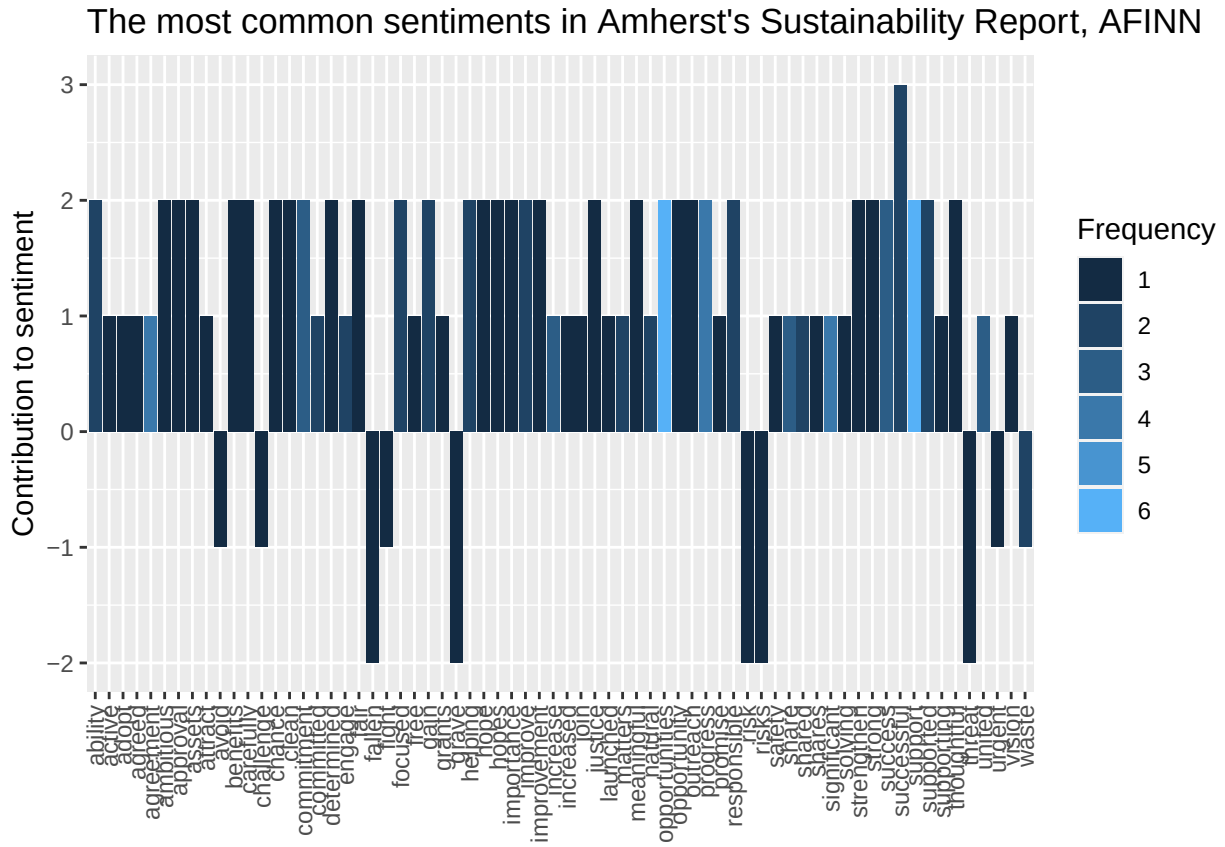



```
#Word Cloud
mypal <- brewer.pal(10, "Paired")

wordcloud(words = word_frequencies$word,
          freq = word_frequencies$n,
          min.freq = 5,
          max.words = 50,
          # plot the words in a random order
          random.order = TRUE,
          # specify the range of the size of the words
          scale = c(2, 0.3),
          # specify proportion of words with 90 degree rotation
          rot.per = 0.15,
          # colors words from least to most frequent
          colors = mypal,
          # font family
          family = "sans")
```


```
#NRC Lexicon
nrc_lexicon <- get_sentiments("nrc")

nrc_lexicon_formatted <- inner_join(nrc_lexicon, word_frequencies, by = c("word" = "word"))

nrc_graphic <- nrc_lexicon_formatted %>%
  filter(sentiment %in% c("anger", "anticipation", "disgust", "sadness", "fear", "joy", "surprise", "tr
  group_by(sentiment) %>%
  arrange(desc(n)) %>%
  slice(1:10) %>%
  ungroup()

nrc_graphic %>%
  ggplot(aes(x = reorder(word, n), y = n,
             fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  coord_flip() +
  facet_wrap(~ sentiment, ncol = 2, scales = "free") +
  labs(x = NULL,
       y = "Contribution to sentiment",
       title = "The most common sentiments in Amherst's Sustainability Report, NRC")
```



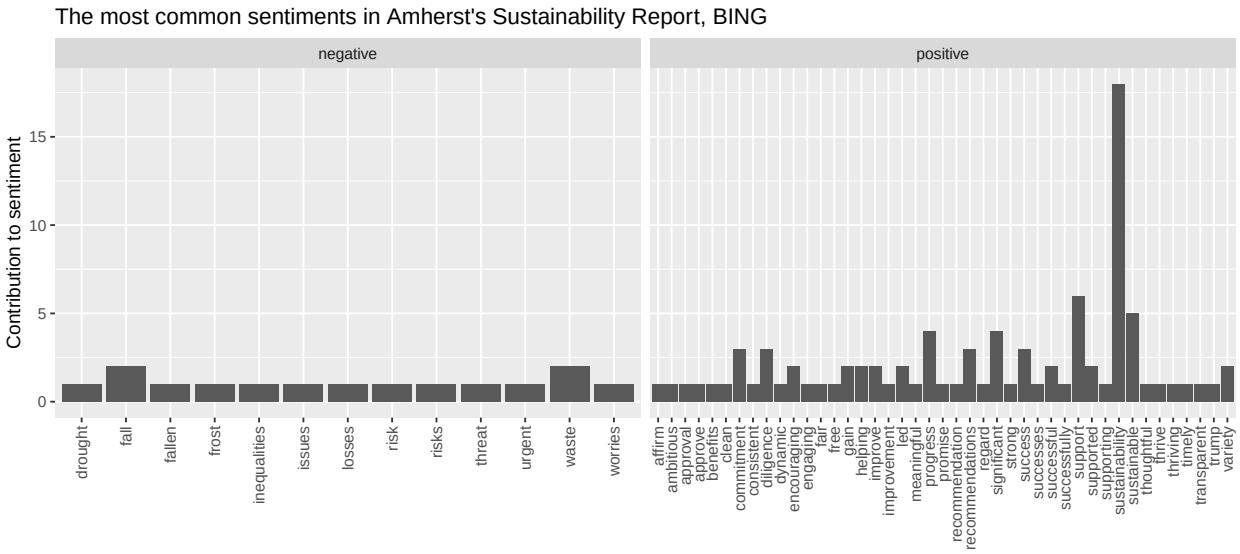

```
#AFINN Lexicon
afinn_lexicon <- get_sentiments("afinn")

afinn_lexicon_formatted <- inner_join(afinn_lexicon, word_frequencies, by = c("word" = "word"))

ggplot(afinn_lexicon_formatted, aes(x = word, y = value, fill = n)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  labs(x = NULL,
       y = "Contribution to sentiment",
       title = "The most common sentiments in Amherst's Sustainability Report, AFINN") +
        guides(fill = guide_legend(title = "Frequency"))
```




```
#BING Lexicon
bing_words <- word_frequencies %>%
  inner_join(get_sentiments("bing")) %>%
  ungroup()
```

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

```
ggplot(bing_words, aes(x = word, y = n)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  facet_wrap(vars(sentiment), scales = 'free_x') +
  labs(x = NULL,
       y = "Contribution to sentiment",
       title = "The most common sentiments in Amherst's Sustainability Report, BING")
```


The most common sentiments in Amherst's Sustainability Report, BING
negative
positive
Contribution to sentiment
0
5
10
15
drought
fall
fallen
frost
inequalities
issues
losses
risk
risks
threat
urgent
waste
worries
affirm
ambitious
approval
approve
benefits
clean
commitment
consistent
diligence
dynamic
encouraging
engaging
fair
free
gain
helping
improve
improvement
led
meaningful
progress
promise
recommendation
recommendations
regard
significant
strong
success
successes
successful
successfully
support
supported
supporting
sustainability
sustainable
thoughtful
thrive
thriving
timely
transparent
trump
variety