Textmining

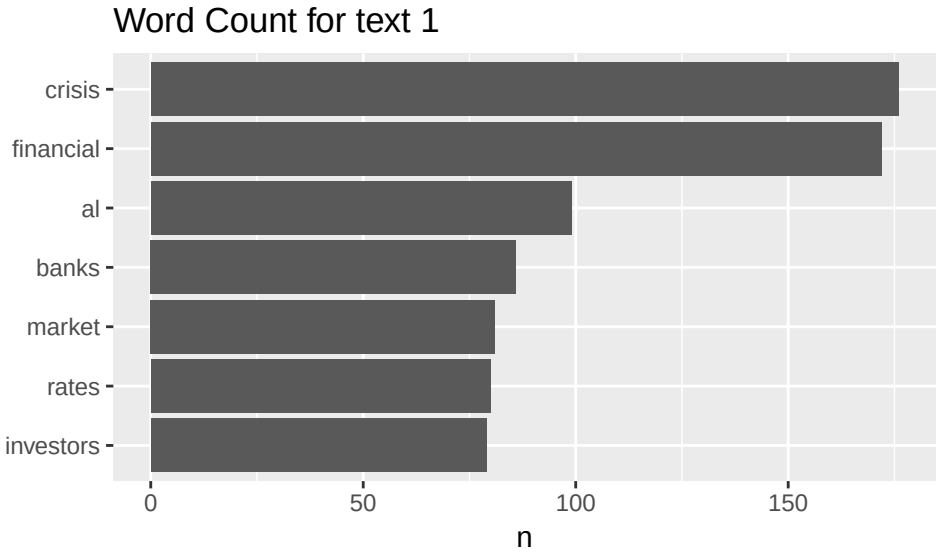
Sky Liu

November 4, 2018

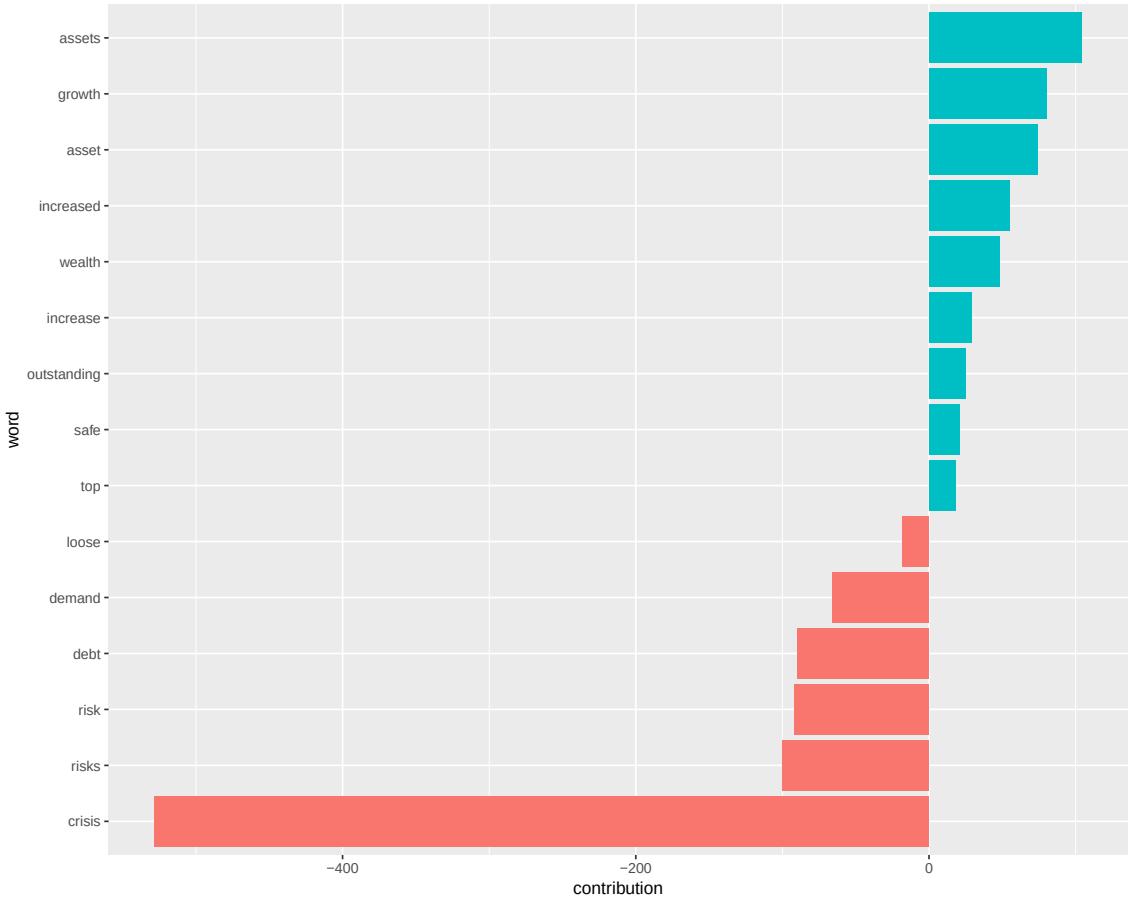
SCRAPING DATA FROM https://correlaid.org/blog/

GET A TIDY TEXT FORMAT & WORD COUNT

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



Sentiment Analysis With Tidy Data



Comparing the three sentiment dictionaries

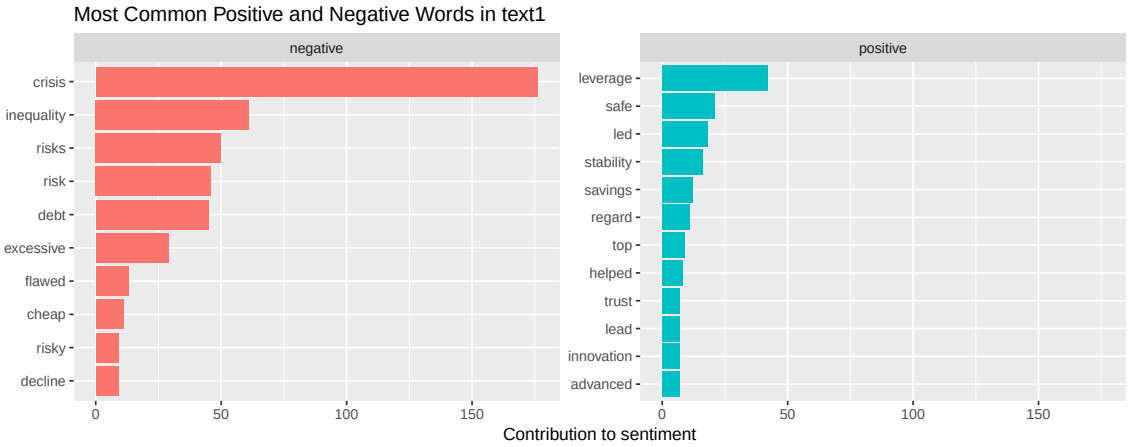
To compare Dictionary Bing with two other dictionaries AFINN AND NRC, we used blog 1 as an example.

From the plot, we could see that the analysis using AFINN dictionary gives most negative results, the analysis using NRC dictionary gives mostly positive results, while the analysis using Bing gives a result balanced from positive and negative.

Most common positive and negative words

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

```
## Selecting by n
```



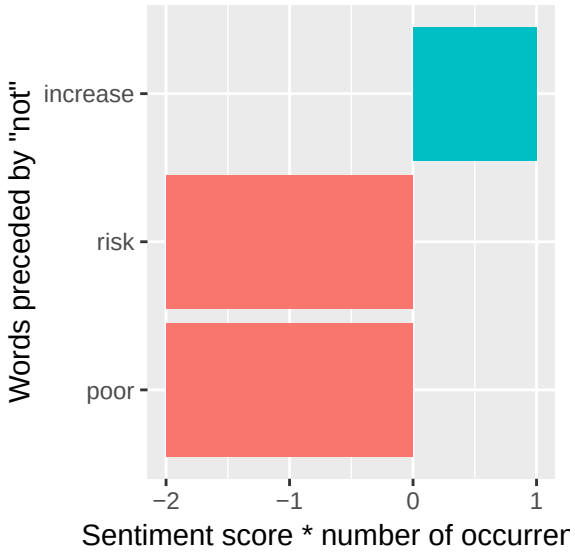
The result of the most common positive and negative words in blog 2 is obvious. Since blog 2 is about gaming AI, the common positive words would be "win" and the common negative words would be "lose" or "die" with no doubt.

Word Cloud

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

The word cloud analysis also gives us a sense of what the blog is mainly about. Take blog 2 as an example. We can clearly see that the blog is about a gaming AI model.

Bigram sentiment Analysis



To take a further look at sentiment analysis, we took blog 2 to look at bigrams. Since the blog isn't very long, we can see from the plot that the bigram begining with "not" is mostly not winning, which is of course a negative bigram.