TED Talks by Speaker Laurie Garrett and Speaker Ted Halstead

Ibitayo, Ademola

Introduction

The aim of this project is to extract specific speakers' Ted talk data from the provided Ted talk data and analyse the data to see the words used by the speakers, the words' sentiments and observe patterns for useful insights.

The Ted talk topics and speakers considered in this report include Lessons from the 1918 flu by Laurie Garrett (TED2007) and A climate solution where all sides can win by Ted Halstead (TED2017).

Laurie Garrett's talk was on pandemic threats and the level of preparations that should be made towards them. She focused on the Avian Influenza H5N1 which first emerged in Guangdong, China, 1997. She discussed the level of severity and leveraged this to recommend ways to prepare which include: 1. Research on masks, hygiene, Tamiflu and home isolation. 2. Pressure local political leaders to provide better plans 3. Pressure national leaders.

Ted Halstead's talk was on a climate solution through a killer app to climate policies. He spoke about how the climate killer app helps to overcome psychological, geopolitical and partisan barriers to climate progress. He co-authored The Conservative Case For Carbon dividends. He mentioned the four pillars of a carbon dividends plan which include: 1. Gradually rising carbon tax. 2. Carbon dividend for all. 3. Regulatory rollback. 4. The climate domino effect.

Methods

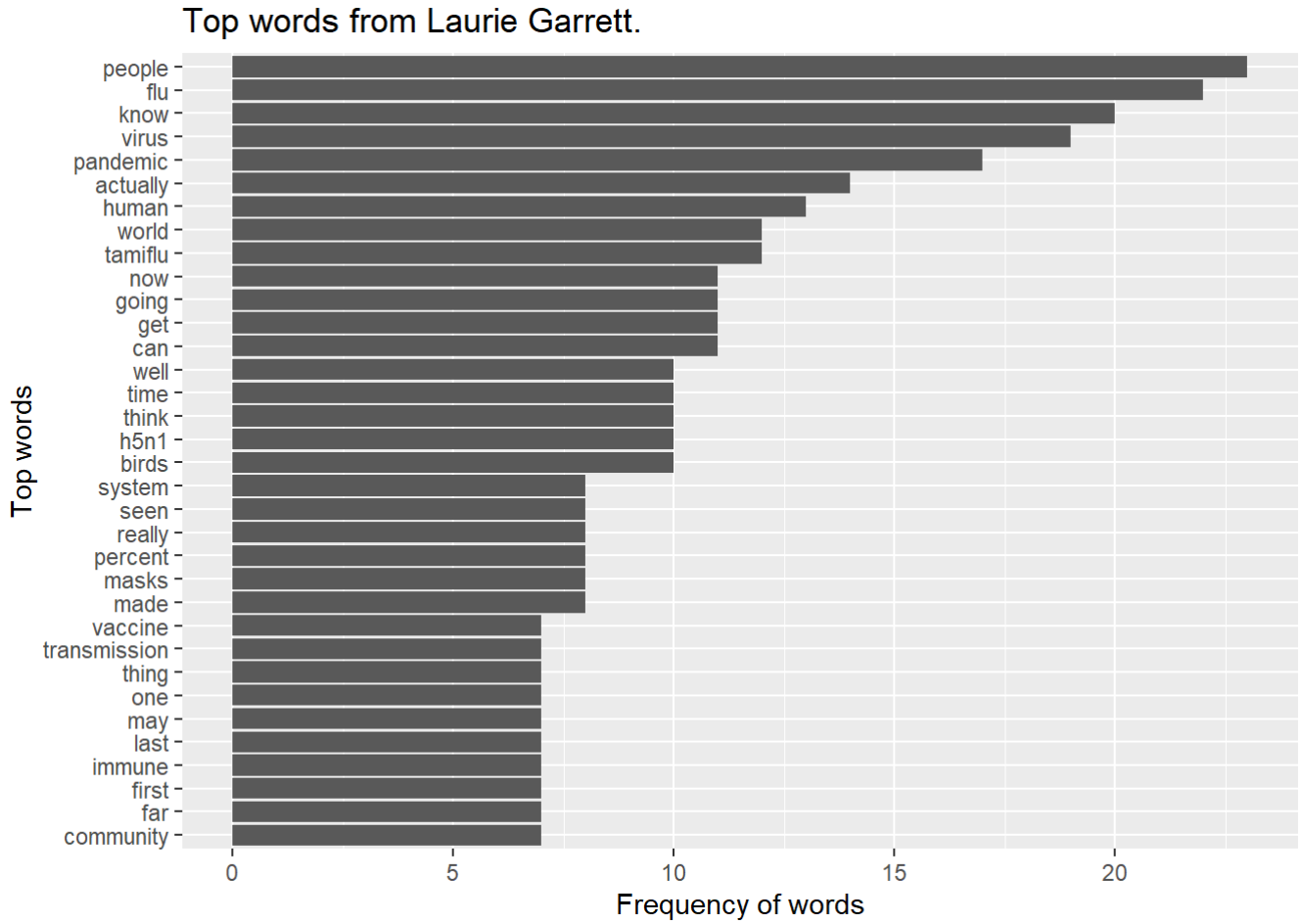
1. The packages needed for the analysis were loaded and they include (`dsEssex` , `tidyverse` , `tidytext` , `ggplot` , `ggrepel`).
2. The Ted talk data was loaded from the given MA331-Report template. An overview of the ted_talk data frame structure was gotten using `glimpse()` .
3. The data for the two speakers to be considered were extracted from the ted_talk data using `filter()` and the data frame summary was viewed.
4. The text data for the two speakers were tidied and tokenized using `unnest_tokens()` . Stopwords were removed using `get_stopwords()` passed into `anti_join()` and the new data frame summary was viewed.
5. The words for each speakers were extracted into separate data objects for analysis using `filter()` and the frequency of each speaker's words was computed and sorted in descending order using `count()` .
6. The top 25 words from each speaker data were visualized separately using `slice_max()` to get top words and `ggplot()` + `geom_col()` to view them based on the categorical variable m_word defined by `mutate()` .
7. A new date frame was created by binding the different speakers data using `bind_row()` and the words from both speakers were visualized and compared on a plot.
8. The sentiment data of the speakers' words were obtained using `get_sentiment()` with the `nrc` lexicon and stored in a data frame.

9. The odds ratio (`OR`) of the sentiments were computed using `compute_OR()` . The logarithm of the odd ratios (`log_OR`) were computed using `log()` and the sentiments were ordered according to their corresponding log_OR values using `reorder()` .

10. A bar chart was plotted for the `log_OR` of the sentiments with color-filled bars to differentiate between negative and positive log_OR values using `coord_flip()` and `scale_fill_manual()` .

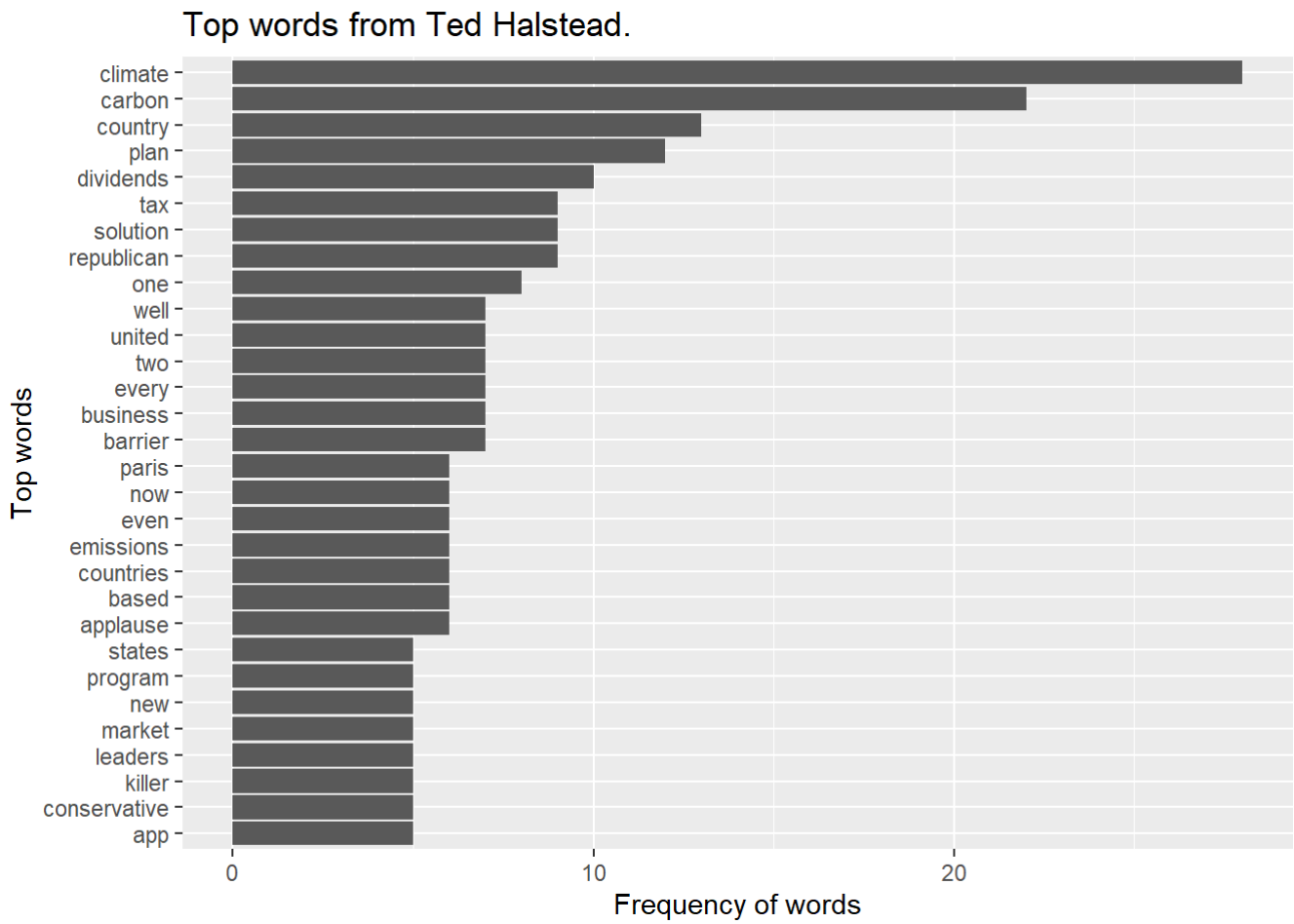
Results

`Plot 1` Top words from Laurie Garrett



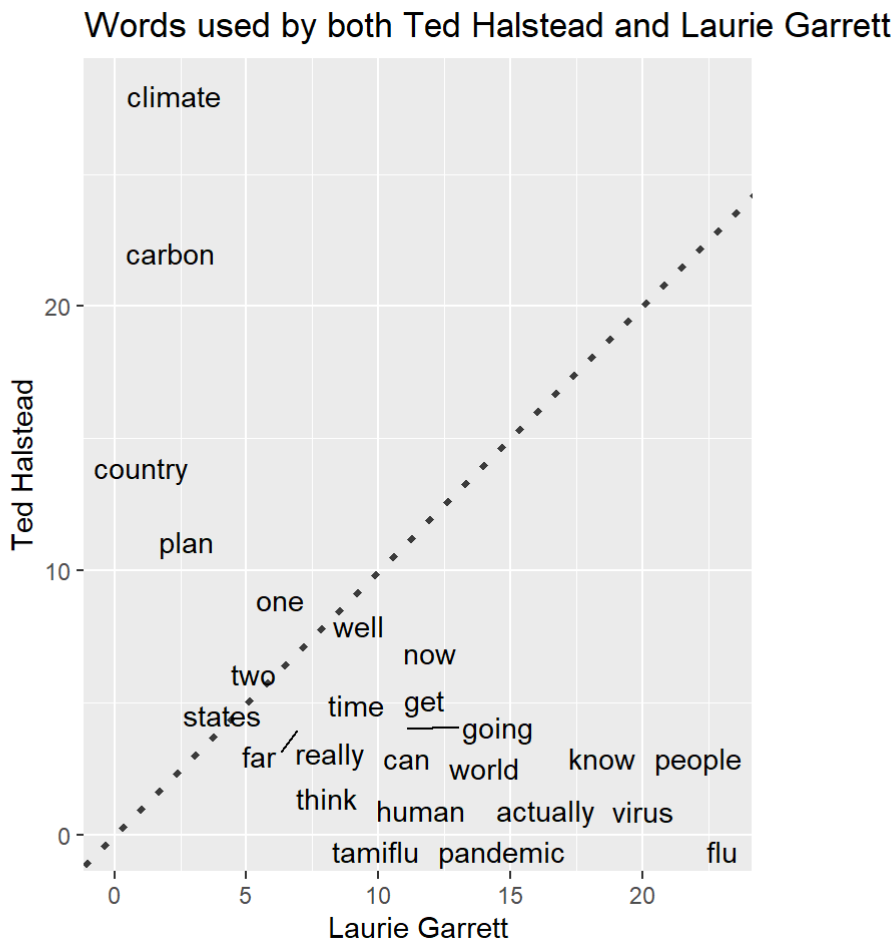
The `plot 1` below shows the top words used by Laurie Garrett. The most commonly used words based on their frequencies are flu, people and virus.

`Plot 2` Top words from Ted Halstead



The `plot 2` above shows the top words used by Ted Halstead. The most commonly used words based on their frequencies are climate, carbon, and country

`Plot 3` The top words from both Ted Halstead and Laurie Garrett



The `plot 3` above shows some of the top words used by both speakers. There is a wide difference with the main words used by each speakers given that they spoke on different topics. The keywords are peculiar to the topics under which they are used. The climate It is also observed that few words slightly shared based on the distance to the line. They include time, plan, people, human which are more generic words across industries.

Sentiment Analysis

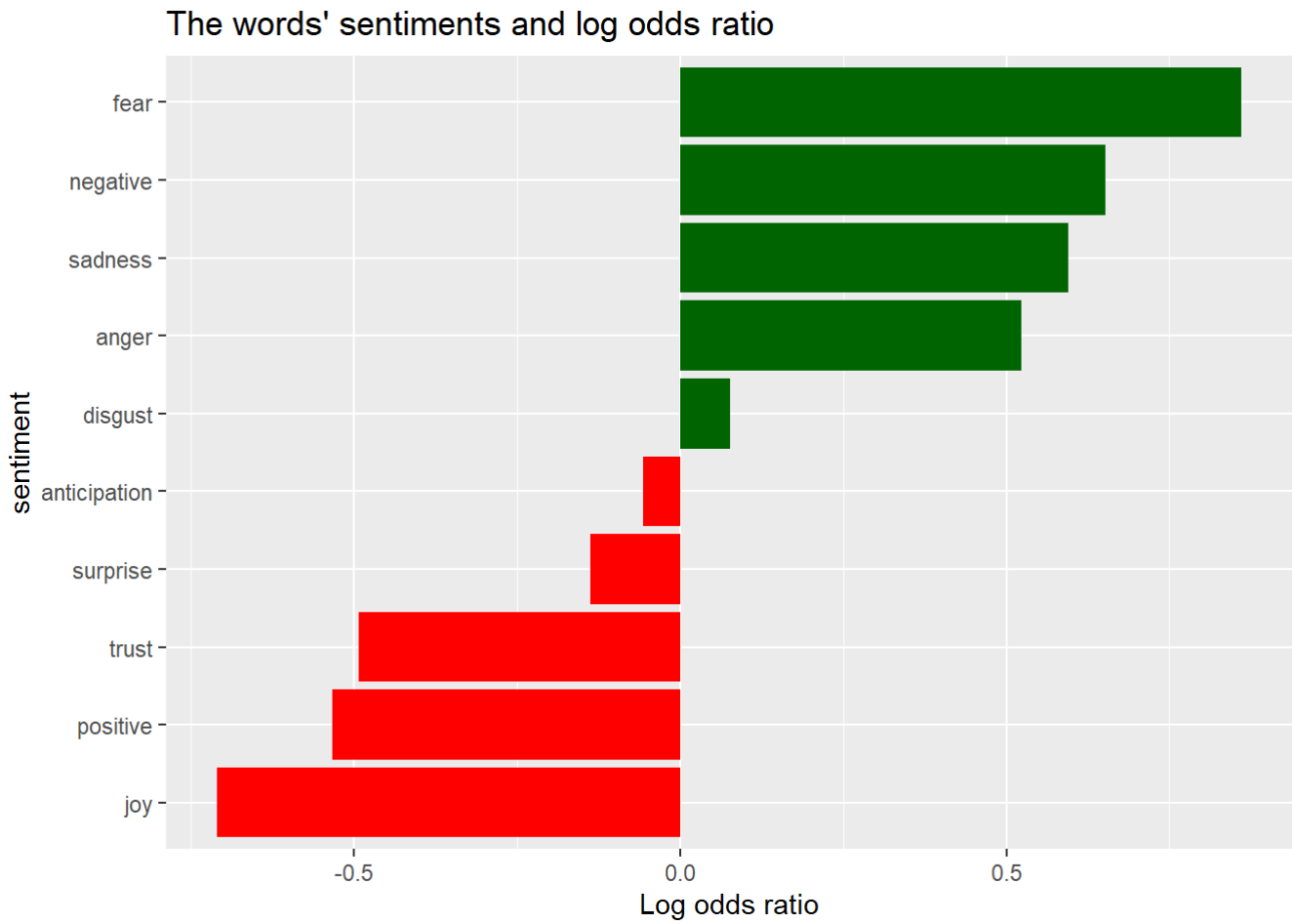
Sentiment values are assigned to speakers' words based on their emotional content. The National Research Council (NRC) lexicon are used to generate the sentiment values for the speakers' words.

`Table 1` Sentiments occurrence for each speakers.

```
# A tibble: 10 x 5
   sentiment    `Laurie Garrett` `Ted Halstead`    OR  log_OR
   <fct>                   <int>          <int> <dbl>   <dbl>
 1 anger                      25             10 1.69   0.523
 2 anticipation               39             27 0.944 -0.0573
 3 disgust                    18             11 1.08   0.0765
 4 fear                       50             15 2.36   0.859
 5 joy                        18             23 0.492 -0.710
 6 negative                   72             27 1.92   0.652
 7 positive                   77             76 0.587 -0.533
 8 sadness                    37             14 1.81   0.595
 9 surprise                   16             12 0.872 -0.138
10 trust                      48             48 0.611 -0.493
```

The `Table 1` above shows the sentiments occurrence for each speakers and the corresponding odds ratio and logarithm of odds ratio. Given that the total number of words vary for each speaker, the log OR is computed to have a basis for comparing the sentiments for the speakers.

`Plot 4` The words' sentiments and log odds ratio



The `Plot 4` shows the words' sentiments and log odds ratio for the speakers. The joy, positive, trust, surprise and anticipation sentiments are found more in second speaker's words (Ted Halstead's transcript) than in the first speaker's words (Laurie Garrett).

The disgust, anger, sadness, negative, and fear sentiments are found more in the first speaker's words (Laurie Garrett's transcript) than in the second speaker's words (Ted Halstead).

The findings on the sentiments above can be traced back to the two speakers' talks. Ted Halstead's talk was about a solution where everyone wins and this makes it more on the positive side than Laurie Garrett who spoke about a pandemic that could have stirred up urgency and some level of fear for actions to be taken.