Text analytics of the TED talks by Gever Tulley and Greg Gage

Introduction

In this report, We will analyze and compare the **word count** and **sentiment** of two TED speakers. For this report, I am using the Data Set of ted_talks which is present inside dsEssex package. For this analysis the two speakers are **Greg Gage** and **Gever Tulley**.
Greg Gage is a Neuroscientist and co-founder of Backyard Brains. In April 2017, In one of his TED talks, he talk about how plants use electrical signals to convey information and plant-to-plant communication. He also shows how plants translate information between plants.
In one of his TED talks in March 2015, he also talks about the importance of learning Neuroscience. Tools to study neuroscience are so complicated and costly that most universities are not able to buy them. He also predicted that 20 percent of the world population will have a neurological disorder. So he and his lab mate decided to build a simple and cheap tool so anyone can learn about Neuroscience in their school and university. In this TED talk, he also gave demonstrated his equipment. He shows us how to control someone else arm with our brain.
Gever Tulley is a Computer scientist and also the founder of Tinkering School. In his school, there is no fixed curriculum and they do not follow any schedule and do not have any assessment. He believes that children have more ability than they know. Learning happens through projects and practical implementation of thoughts.
In one of his TED talks in March 2007, he talks about a summer program that is organized by his school. This program helps kids to acquire knowledge of constructing new things that come to their minds. He mostly talks about child's imaginations and creative problem-solving techniques. There is a total of four TED talks in the dataset. Each speaker has two talks.

Methods

In this project, we have to compare word frequencies and sentiment analysis.
For comparing word frequencies we are using the below steps
Step 1
First, we will import libraries and load our data set(ted_talks) and filter our two speakers(Greg Gage, Gever Tulley).
Step 2
After filtering the data we want to count which words occur most frequently so we have to convert our text data(transcript) into words. To do this, we use tidytext's **unnest_tokens()** function.
Now that the data is in one-word-per-row format, we can handle it with tidy tools like dplyr. In-text analysis, we wanted to remove stop words. Stop words are words that are not useful for analysis. We can remove stop words by using **stop_words()** function with an **anti_join()** function.
Step 3
After converting the transcript into words and removing stopwords we have to count the word frequency of each speaker. Here we use **count()** and **slice_max()** to find top ten words of each speakers.To Visualize word frequency count we plot a bar graph using the ggplot package.

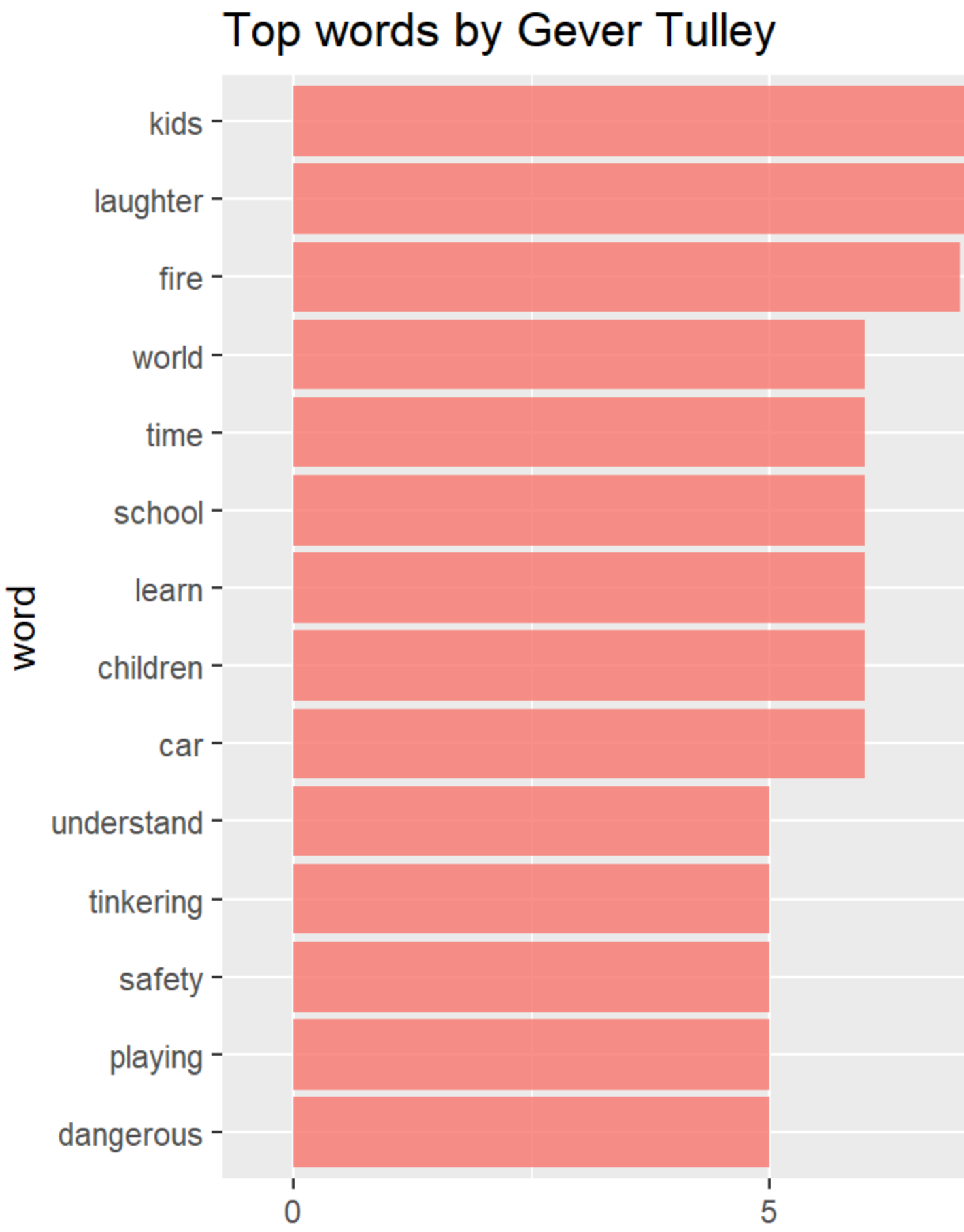
Also, we want to know what are the most common words of both speakers. For a better understanding of common words used by speakers, we are using the **geom_text_repel()** function.

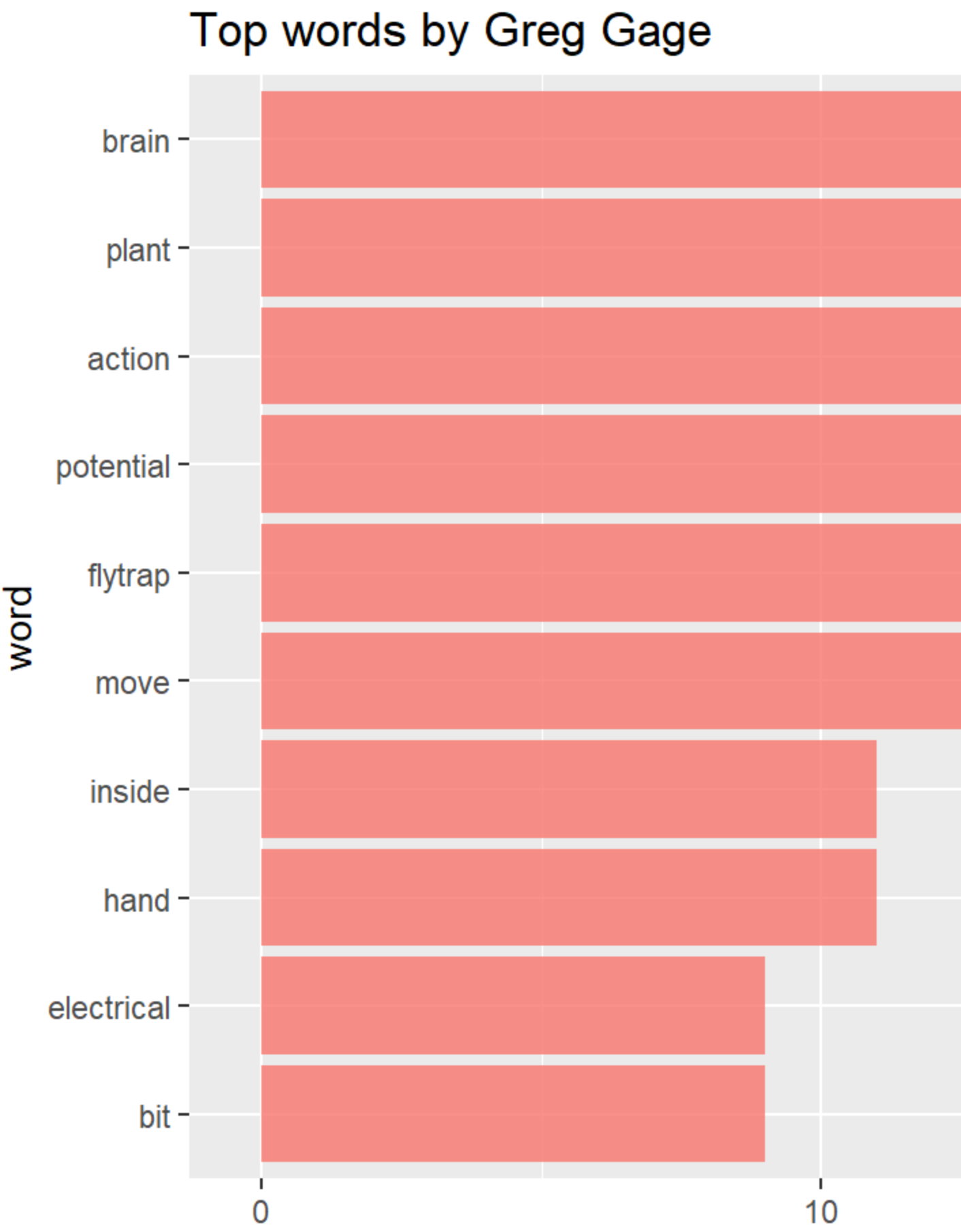
Sentiment Analysis

There are many types of sentiment lexicons but in this project, we are using only two types of *bing* lexicons which gives categorizes words in a binary form(positive and negative), and *NRC* lexicons which link with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). To use particular sentiment lexicons we are using **get_sentiments()** function. It is done with the **inner_join** function
With the help of bing lexicons, I am able to compare who talks more positive words between two speakers, and also I came to know how many positive and negative words each speaker used in their talks.
Using NRC lexicons we come to know what are the different type of emotions each speaker use in their talks.

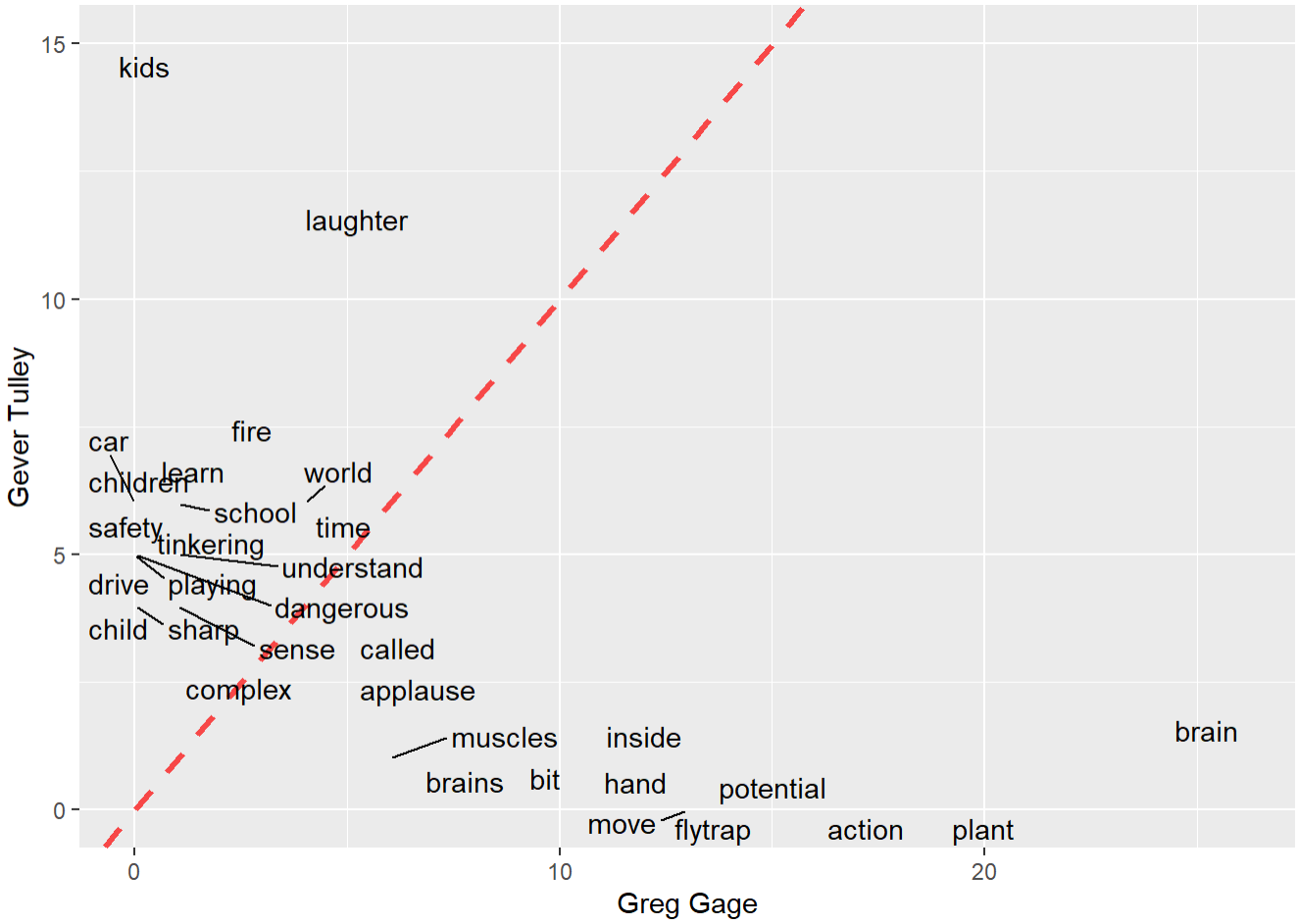
Results

In the below figure, we are anticipating what are the most common words used by each speaker.

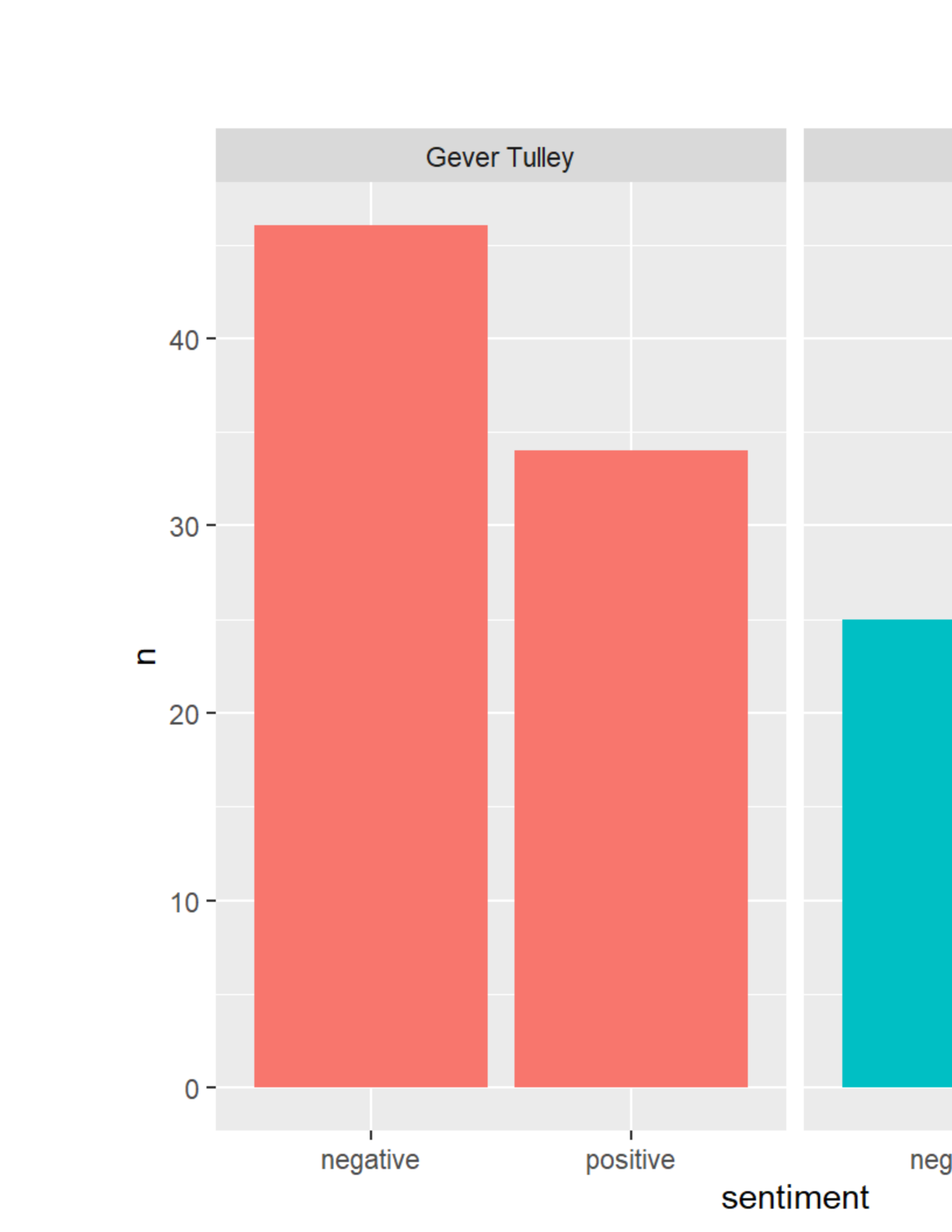

Top words by Gever Tulley
word
kids
laughter
fire
world
time
school
learn
children
car
understand
tinkering
safety
playing
dangerous
0
5

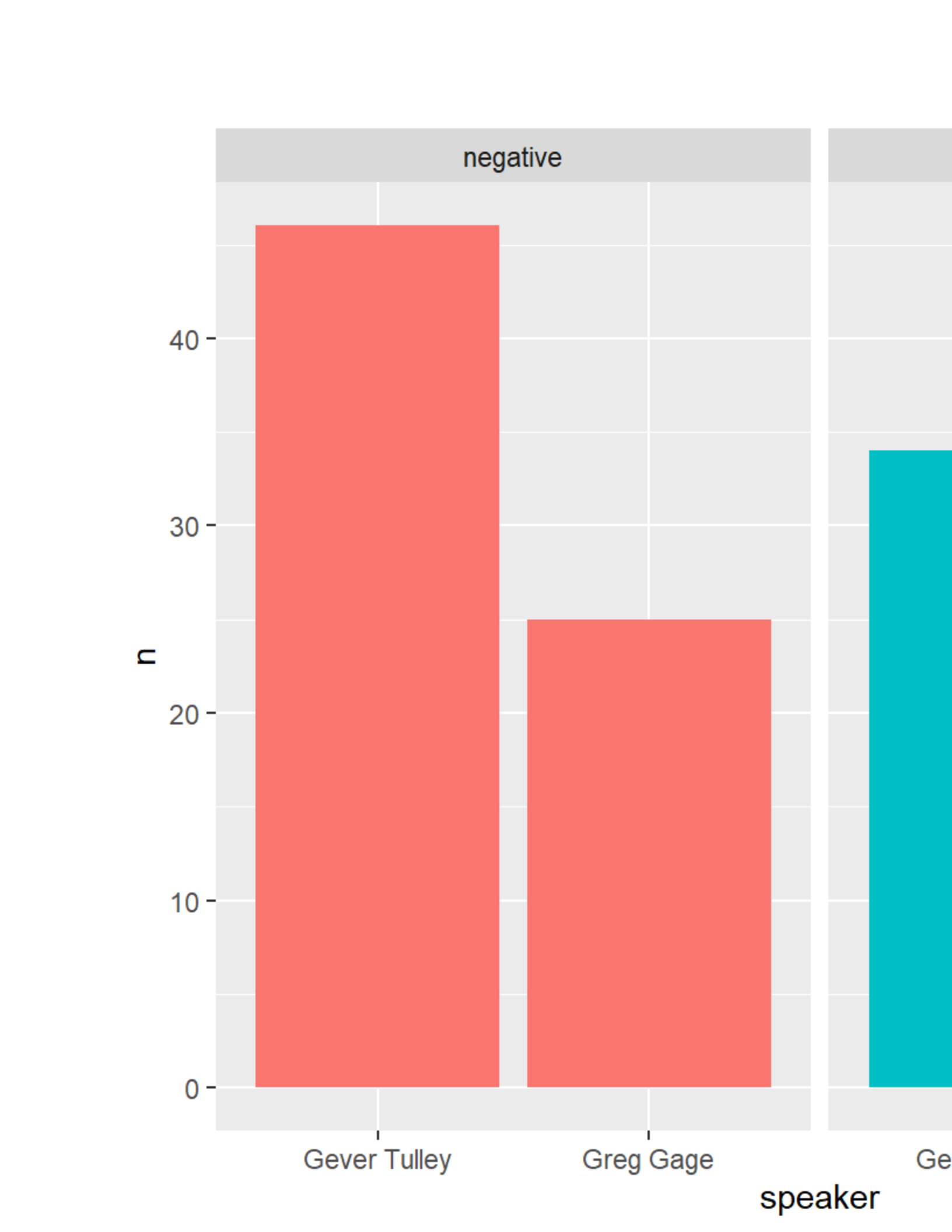

Top words by Greg Gage
word
brain
plant
action
potential
flytrap
move
inside
hand
electrical
bit
0
10

From the above bar graph, we see that In **Gever Tulley** talks he use kids, children, learn, and school frequently whereas In **Greg Gage** talk he often use words like the brain, action, potential, and many more.
In the below plot, we show what are the common words between Greg Gage and Gever Tulley.

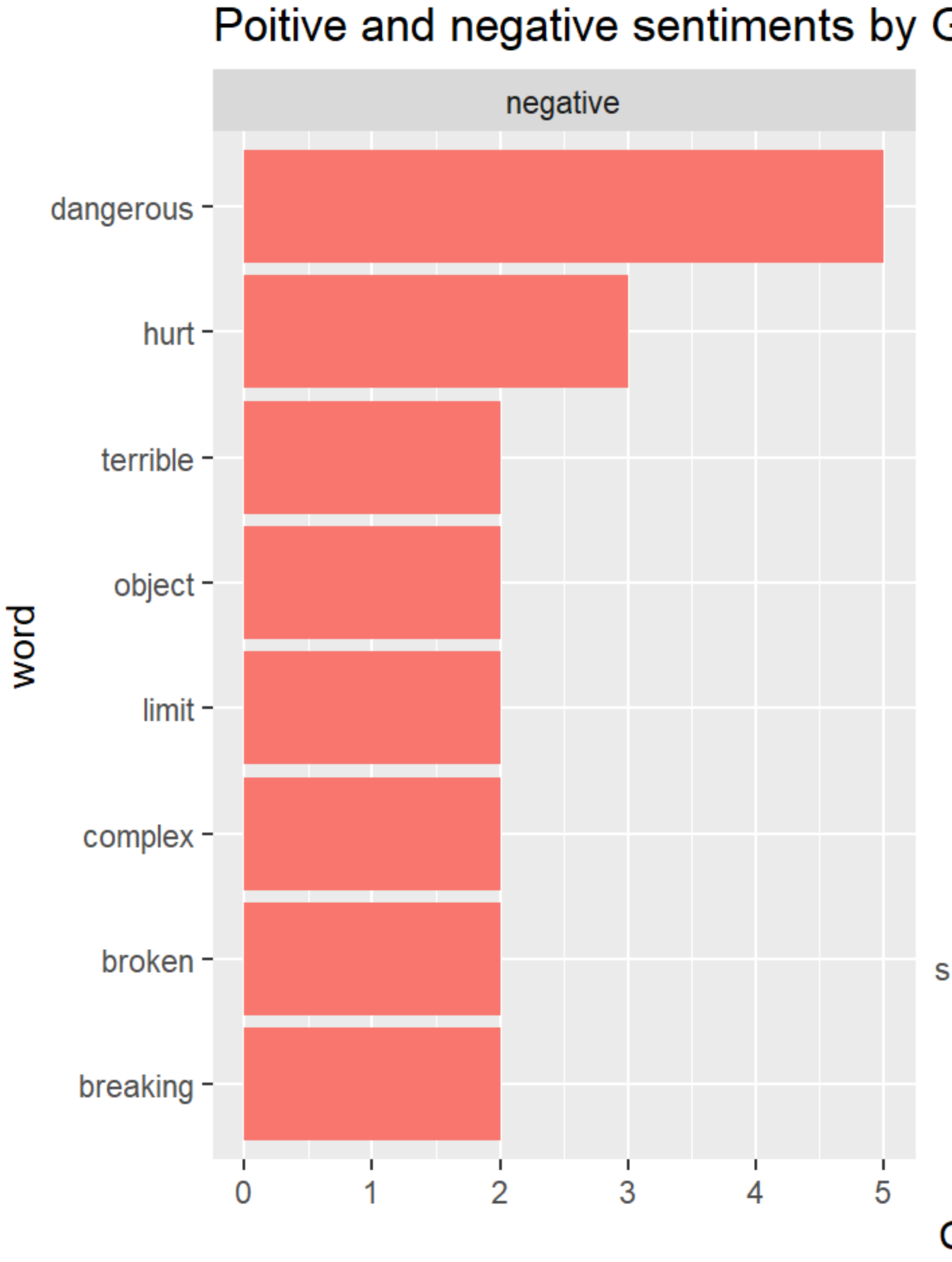


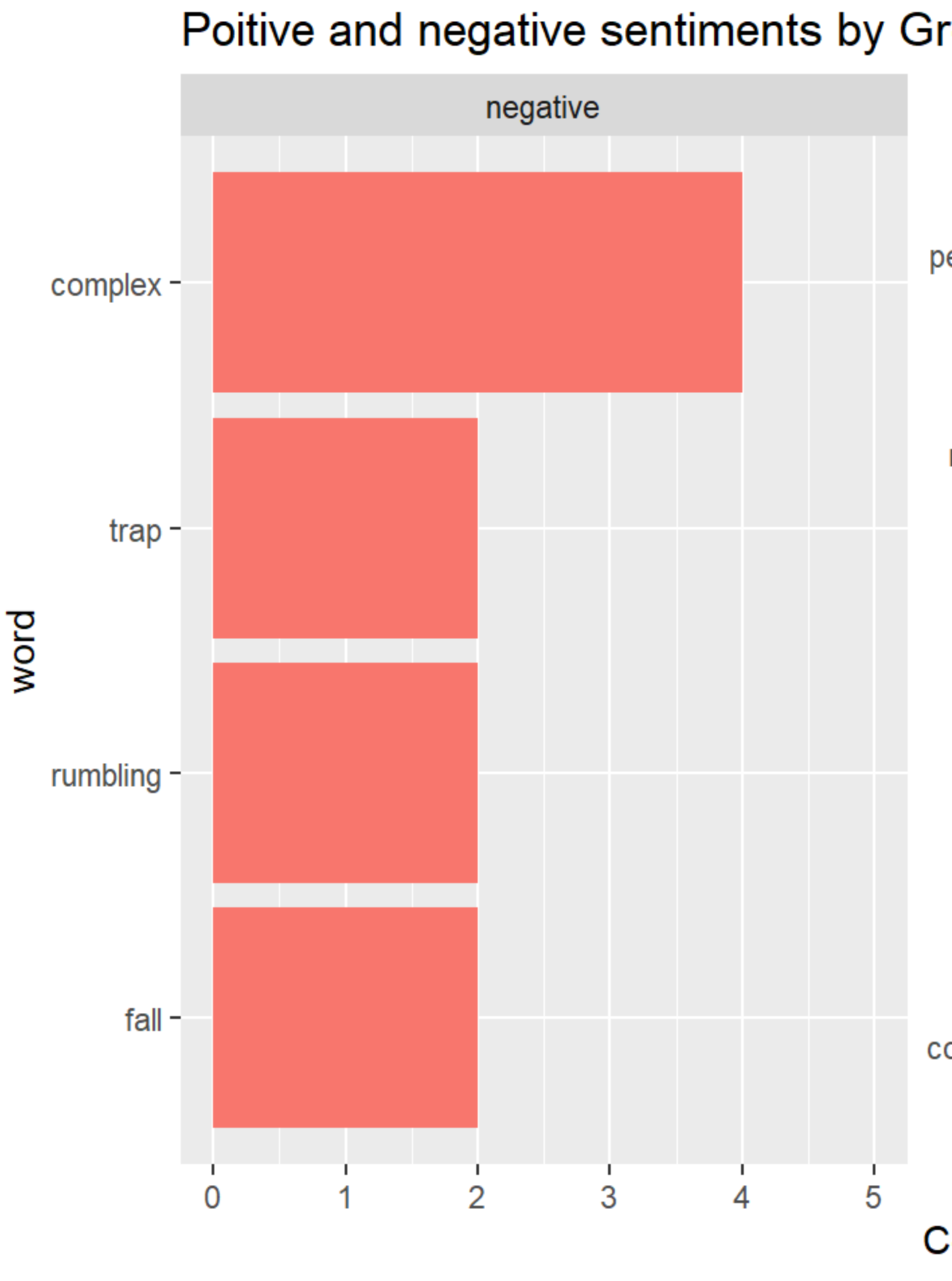
Sentiment Analysis between two speakers
In the below plot, we compare positive and negative words by each speaker (Left diagram). we also compare positive and negative words *between* each speaker (Right diagram).


Gever Tulley
40
30
20
10
0
n
negative
positive
neg
sentiment

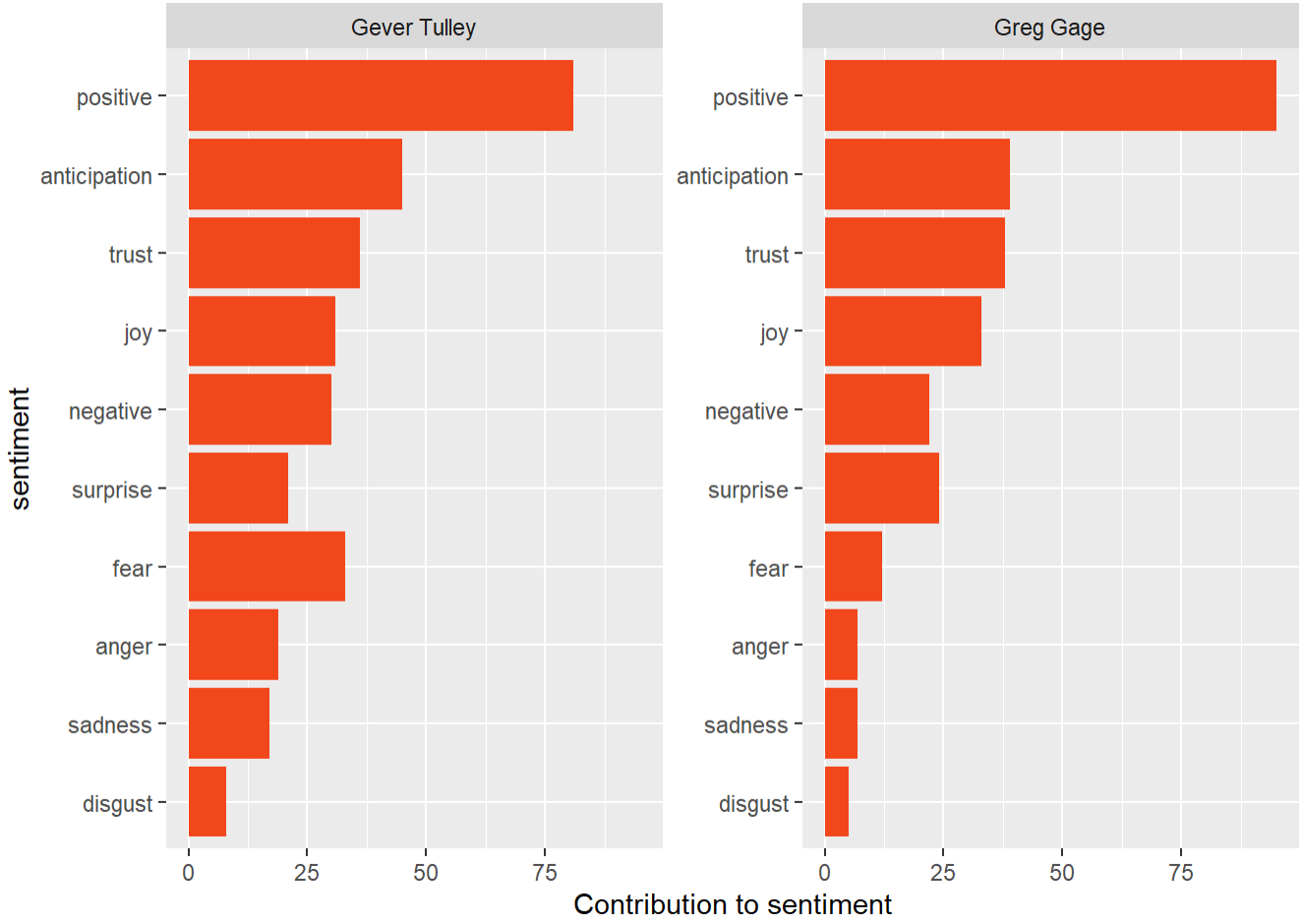

negative
40
30
20
10
0
n
Gever Tulley
Greg Gage
Ge
speaker

From the above bar charts, we can say that Greg Gage use positive sentiment and negative sentiment words nearly the same number of times but Gever Tulley use more negative sentiment words as compared to positive words.
Also, we analyze that Gever Tulley uses more negative sentiment words than Greg Gage. In the above plots, we use the bing lexicon.


Poitive and negative sentiments by G
negative
dangerous
hurt
terrible
object
limit
complex
broken
breaking
word
0
1
2
3
4
5
s


Poitive and negative sentiments by Gr
negative
complex
trap
rumbling
fall
word
0
1
2
3
4
5
C

From the above sentiment analysis we can say that commonly use negative words by Gever are *dangerous, hurt, terrible* and positive words are *sharp, trusted, success*. Whereas most negative words used by Gerg are *complex, fall, trap* and positive words are *perfect, ready, fast*.



From the above chart, we can say that both the speakers have used a lot of positive words throughout their talk. We can see that count of positive words for Greg is more which is obvious as the total word count for him is more.
We can observe that *anticipation*, *trust*, *joy* are present in both the talks.
We can see that the *fear*, *anger* and *sadness* sentiment is high for Gever.
From the above plot, we can conclude that positive sentiments are high in Greg's talk.

Discussion

As both the speakers come from different backgrounds Gerg is a neuroscientist and Gever is a computer scientist and teacher. Both of them are working on different domains and have a different set of audiences. Gever would be addressed to parents of school-going kids and talks of Greg is mostly viewed by students at high school or university. So there are very less common words used by both the speakers and it's difficult to determine the common words used are referring to the same context.
Though if we look on a broad level both are trying to modify or evolve something in the education field. Both of them are working on some equipment or doing some experiments. Greg

is trying to optimize the equipment related to Neuroscience so that they can be easily available and cost-friendly. Whereas Gever is helping kids to design things for fun and utility and he wants kids to think out of the box and then discover and design.

Challanges and Limitations

Data sets have very less things in common so it's a bit difficult to draw the conclusion for sentiments.

Word counts and number of talks are less to determine more about sentiments and personality of speakers

Improvements

The common things which I mentioned above are noted by the observations after watching the videos. If we use some libraries which look at synonyms of words and group them under a section, for example, words like students, university, learning can be related to education. We can also determine the speed by which the presenter is talking like slow, medium, or fast by calculating the number of words per minute so that a recommendation for play speed like 0.75x or 1.25x can be recommended.