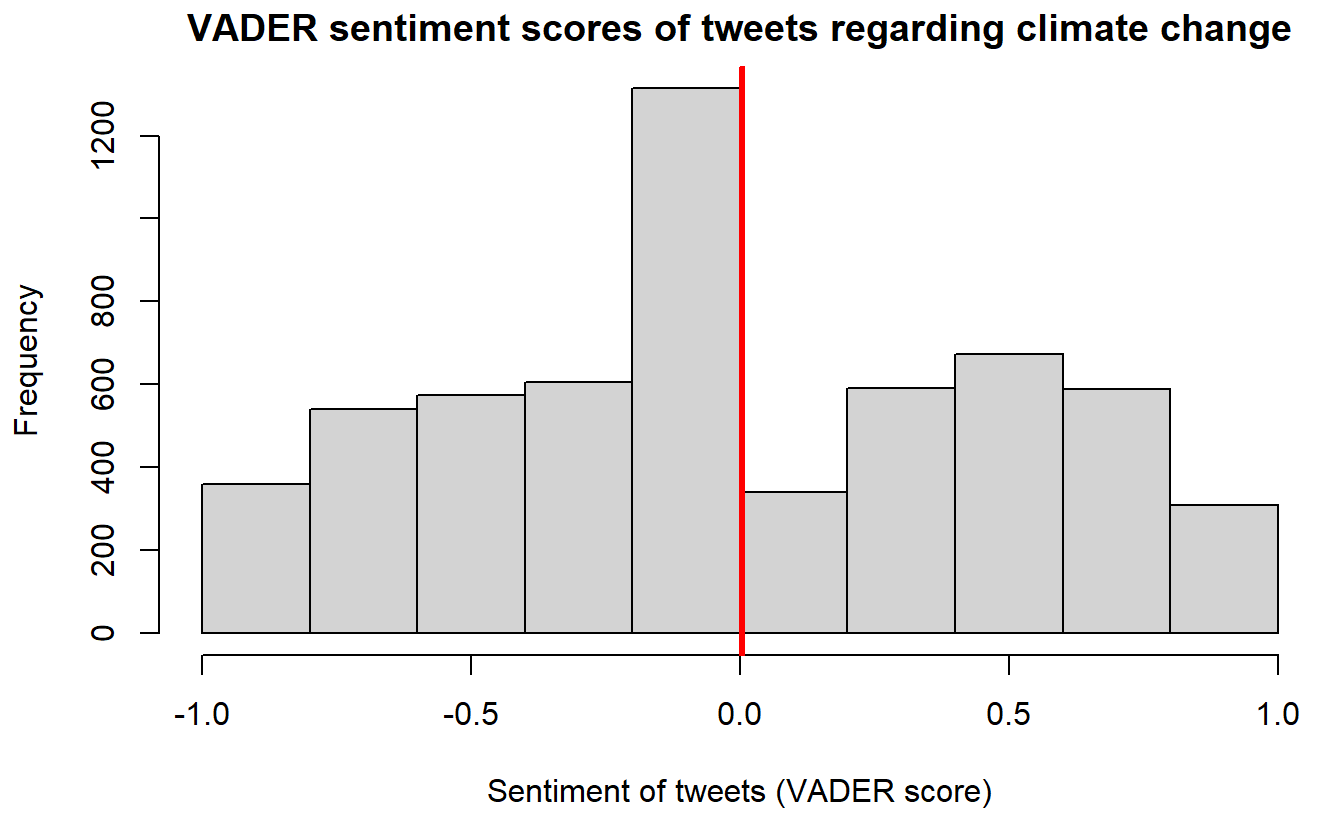
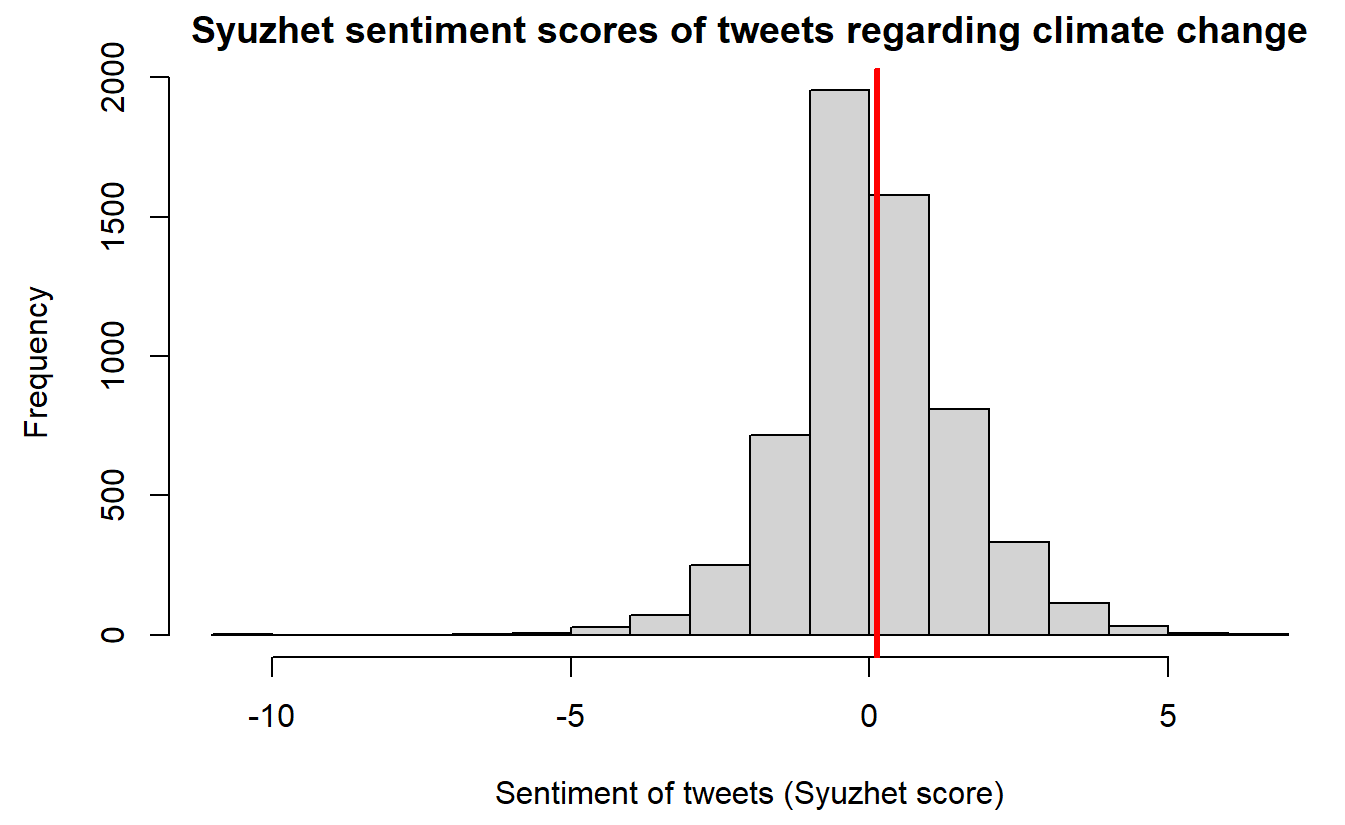
**CLIMATE CHANGE**

**Do tweets that use more emotional language in respect of climate change have a higher engagement rate?**

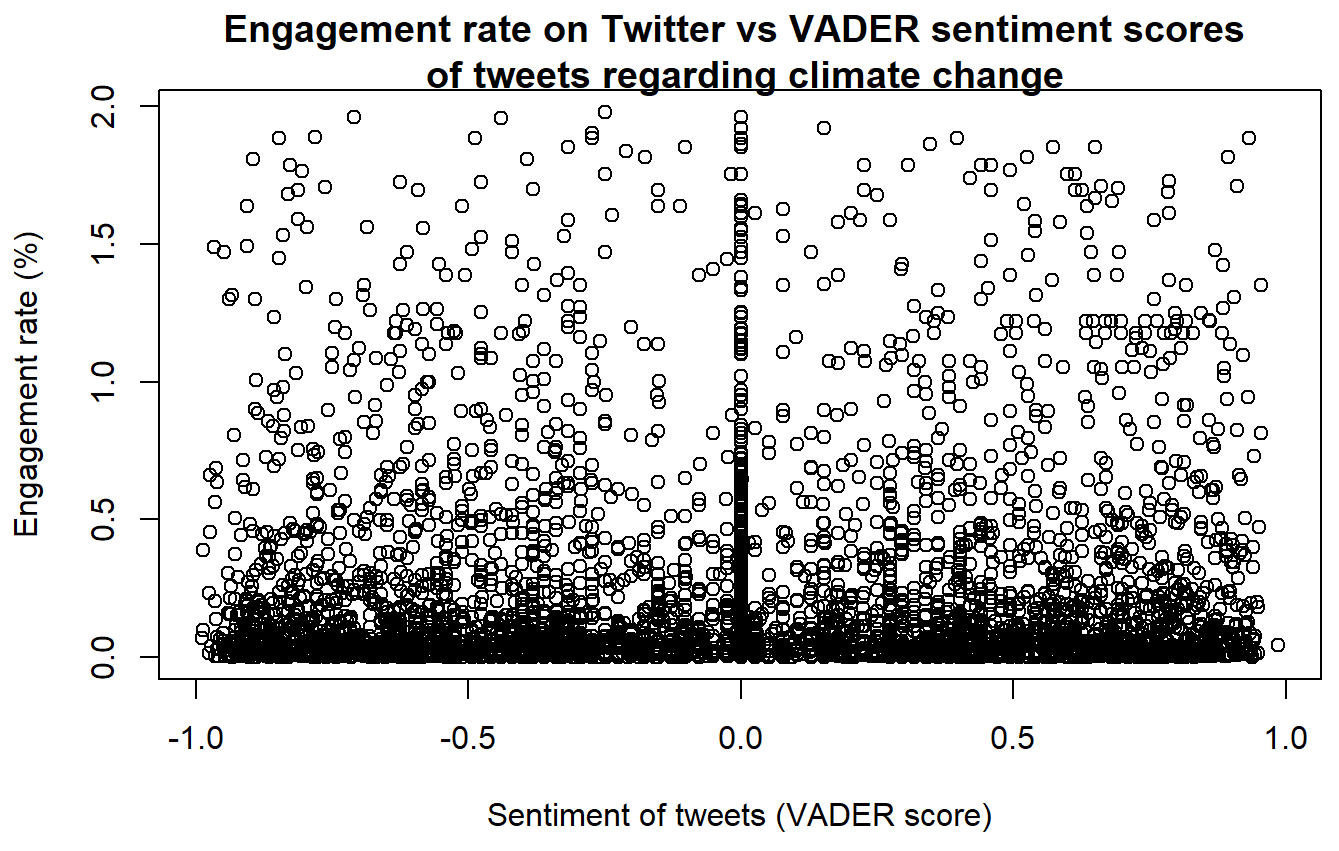
Average engagement rate (for info, in case helpful in discussion): 1.64%, 95% CI [1.29, 1.98]

Frequency of sentiment scores with average plotted as a red vertical line:

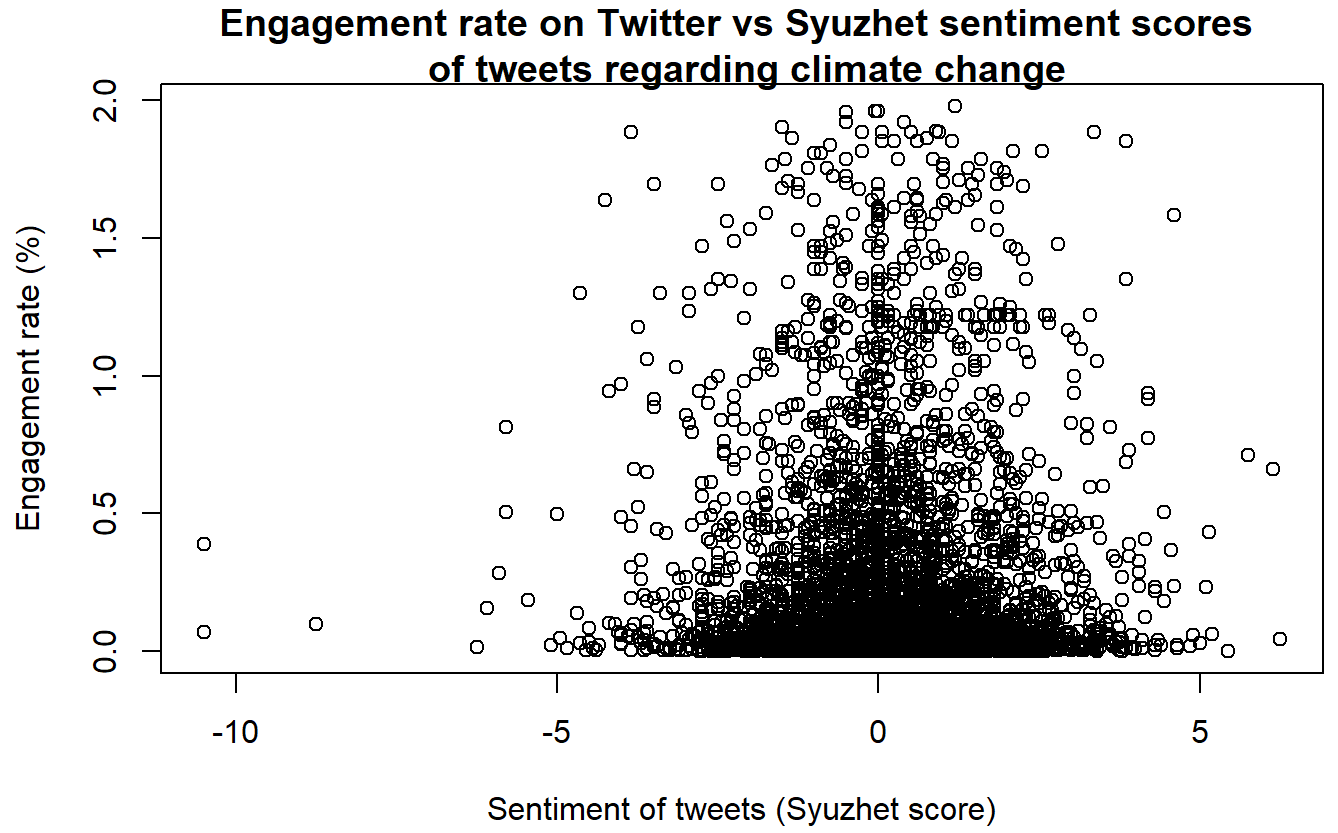




Scatter plots showing engagement rates vs sentiment scores to look for any observable relationships (NOTE: Just for the plots, but not the statistical analysis, data was filtered so that the upper limit of the 95% confidence interval was used as the max engagement rate. This is because there are some large outliers that otherwise make it difficult to see what is going on.):



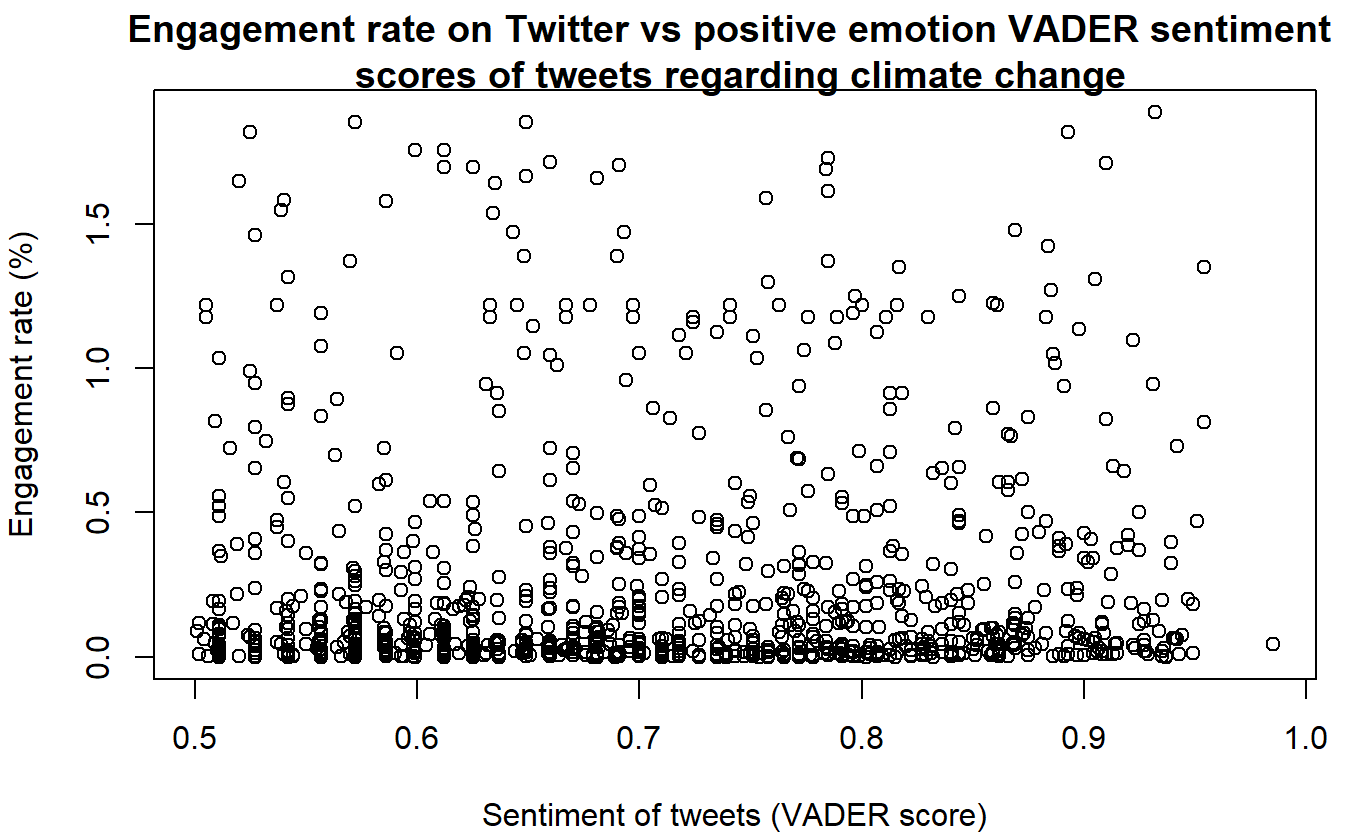
Simple linear regression was used to test if VADER sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = 0.00038, 95% CI [-0.00057, 0.00134], p = 0.432, adjusted R2 = -6.491e-05). Assumptions of the linear regression model were checked.



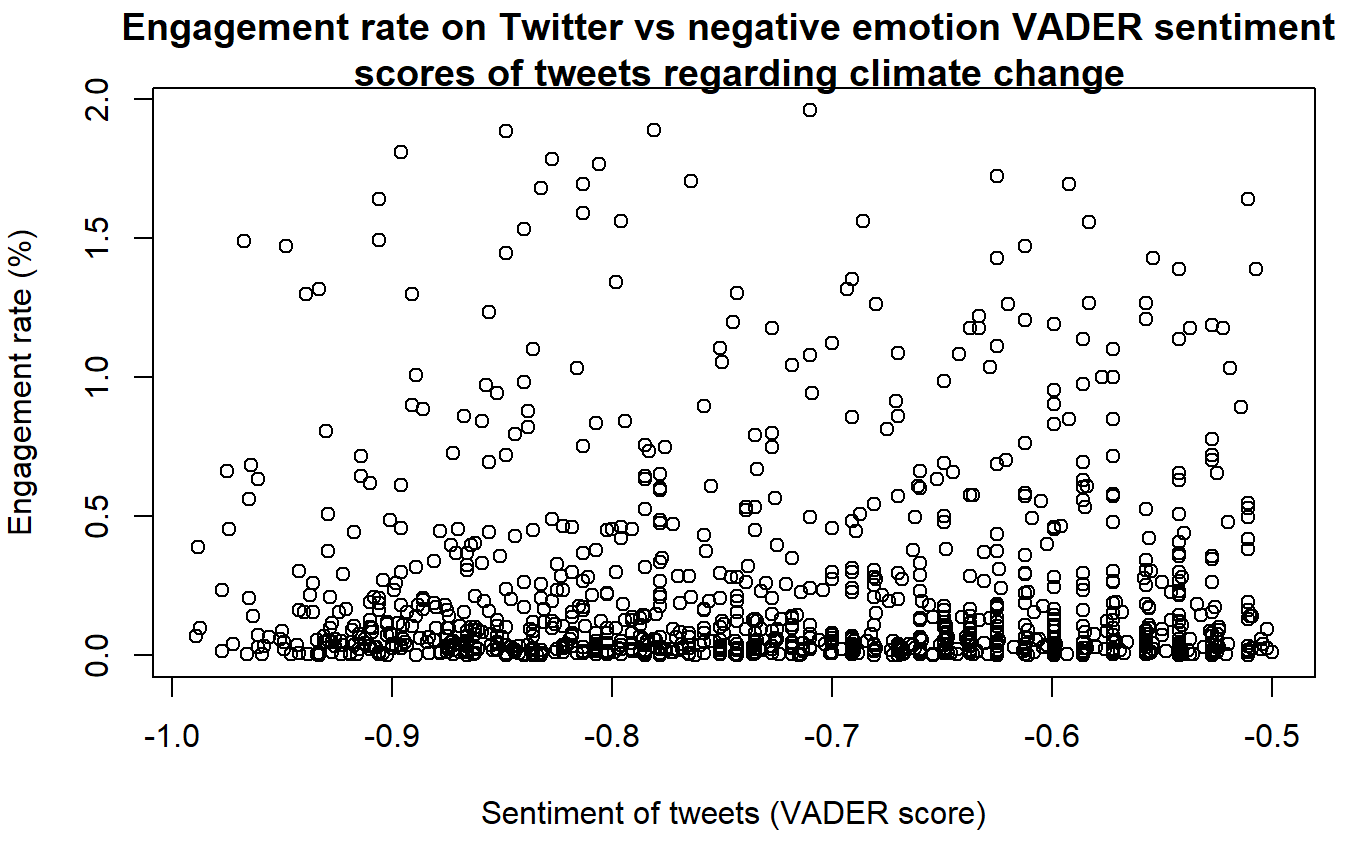
Simple linear regression was used to test if Syuzhet sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = 0.00259, 95% CI [-6.370279e-05, 0.00524], p = 0.0558, adjusted R2 = 0.00045). Assumptions of the linear regression model were checked.

We also separately looked at the relationship between engagement rates and positive and negative sentiment scores to check if there is a relationship that was not observable when the entire dataset was used (NOTE: Just for the plots, but not the statistical analysis, data was filtered so that the upper limit of the 95% confidence interval was used as the max engagement rate. This is because there are some large outliers that otherwise make it difficult to see what is going on.):

* VADER

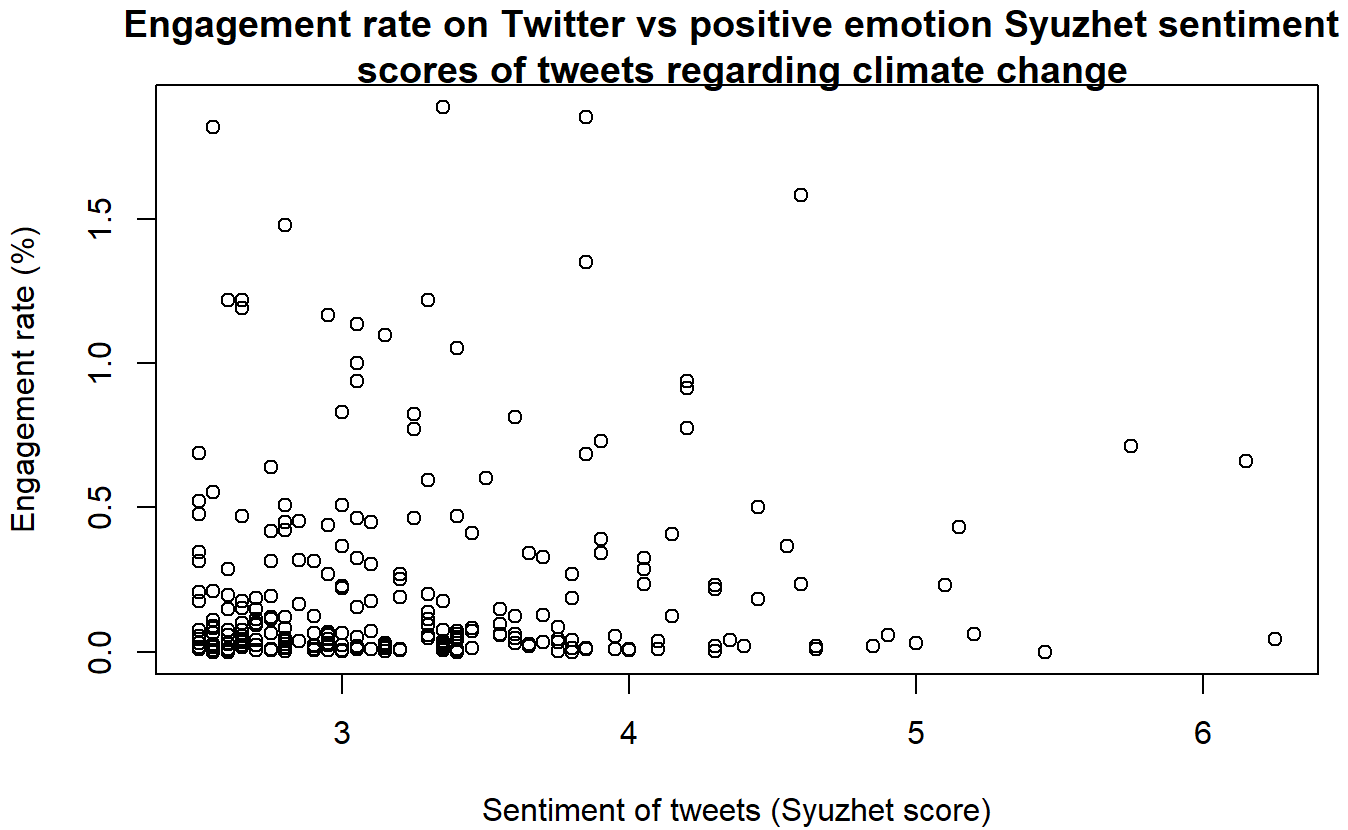


Simple linear regression was used to test if positive VADER sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = 0.00038, 95% CI [-0.00029, 0.00106], p = 0.262, adjusted R2 = 0.00022). Assumptions of the linear regression model were checked.

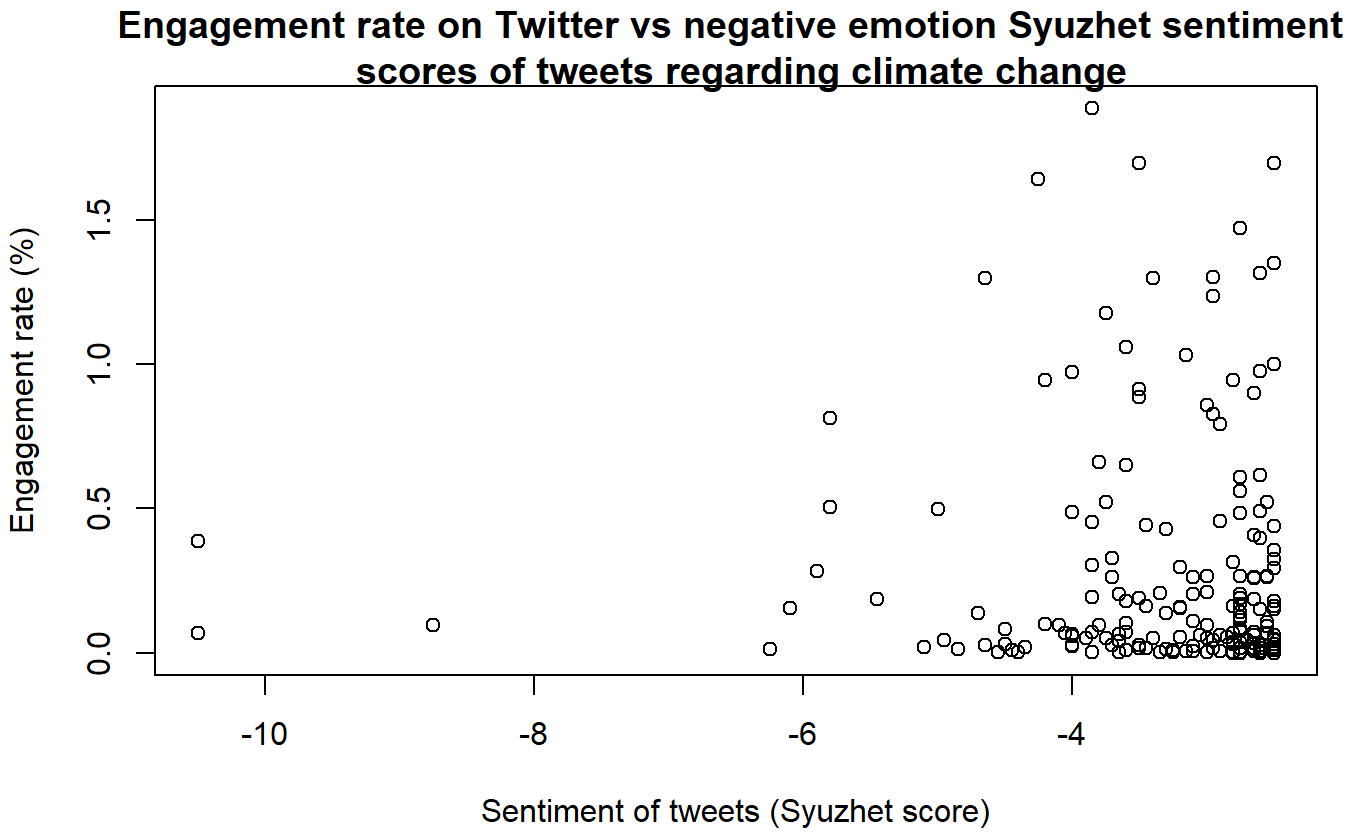


Simple linear regression was used to test if negative VADER sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = -7.059e-05, 95% CI [-0.00124, 0.00110], p = 0.906, adjusted R2 = -0.00084). Assumptions of the linear regression model were checked.

* Syuzhet



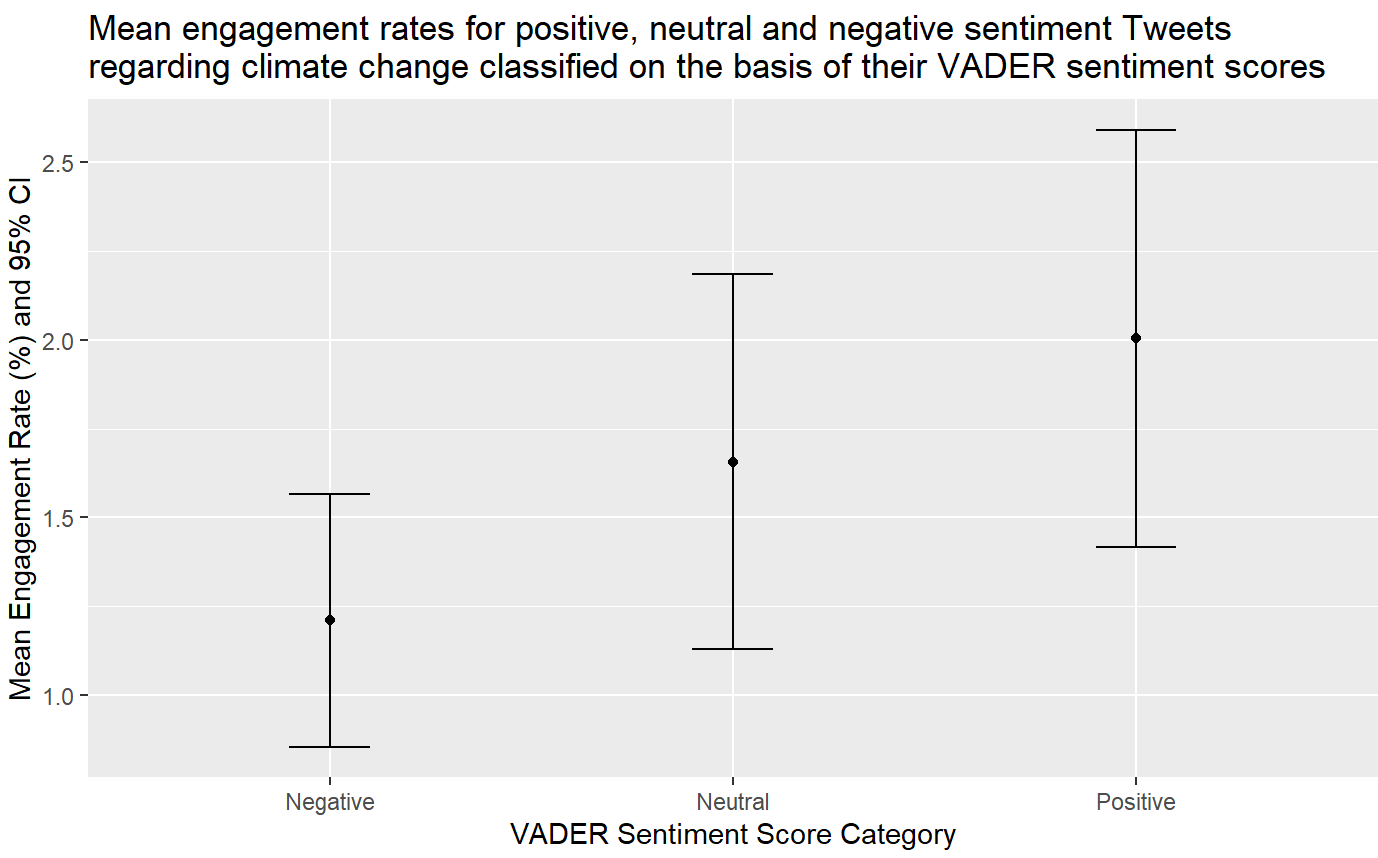
Simple linear regression was used to test if positive Syuzhet sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = 0.00037, 95% CI [-0.00518, 0.00592], p = 0.896, adjusted R2 = -0.00338). Assumptions of the linear regression model were checked.



Simple linear regression was used to test if negative Syuzhet sentiment scores of Tweets significantly predicted the engagement rate and no statistical significance was found (β = -0.04387, 95% CI [-0.10117, 0.01343], p = 0.133, adjusted R2 = 0.00608). Assumptions of the linear regression model were checked.

**Is there a difference between engagement rates of tweets containing positive and negative emotional language?**

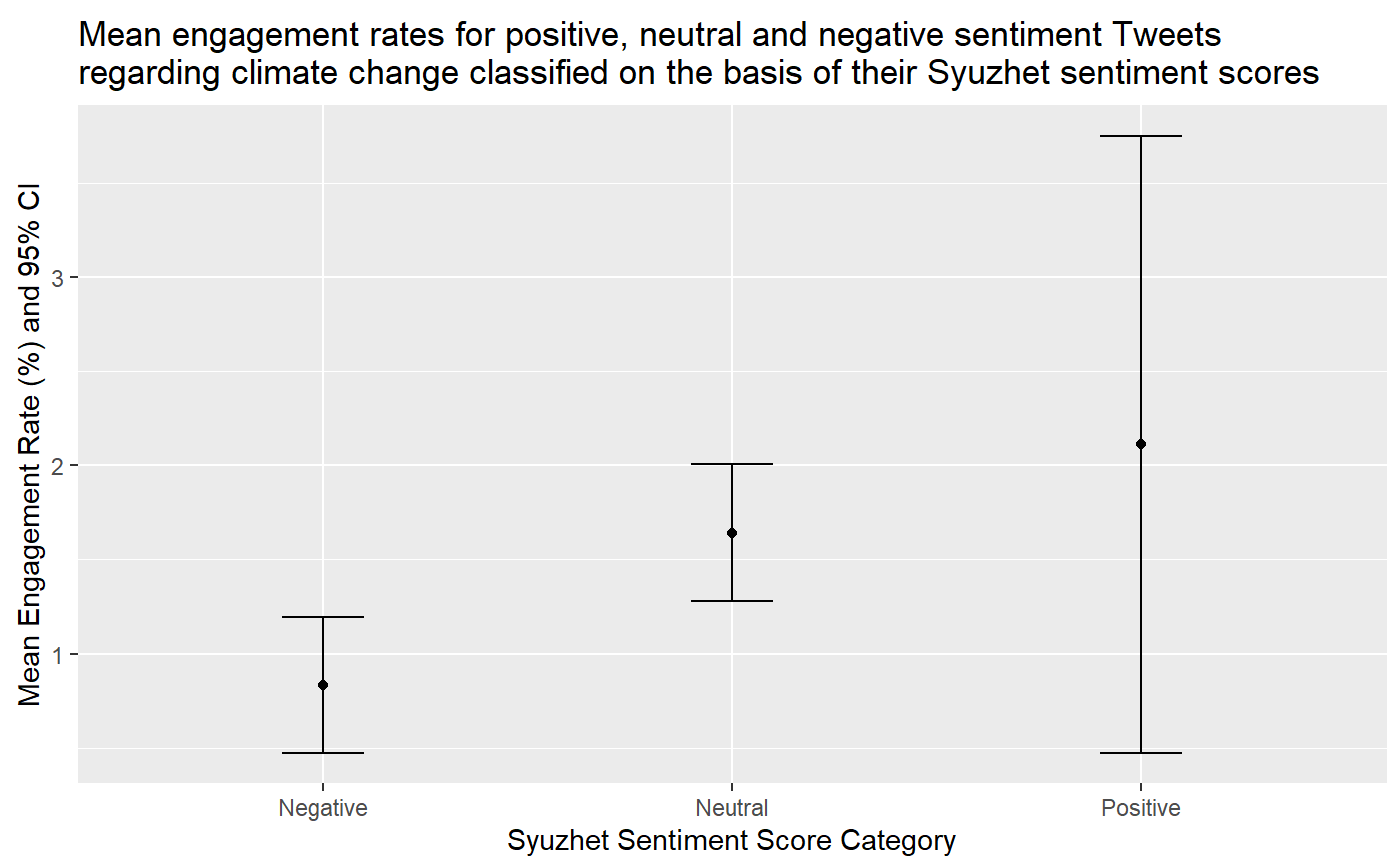
Plots of mean engagement rate showing 95% CI for three categories of sentiment (positive, neutral and negative):



Negative = scores from -1 to -0.5; neutral = scores from -0.4 to 0.4; positive = scores from 0.5 to 1.

To determine which statistical test to perform, data was tested for normality. Engagement rates of Tweets with negative and positive VADER sentiment scores were found to have non-normal distribution using Shapiro-Wilk's test (W = 0.17003, p-value < 2.2e-16; W = 0.17414, p-value < 2.2e-16; respectively).

Wilcoxon rank-sum test was carried out to test if the mean engagement rates of Tweets with negative and positive VADER sentiment scores are significantly different from each other and no significant difference was found at 5% significance level (W = 725524, p-value = 0.08948).



Negative = scores from -5 to -2.5; neutral = scores from -2.4 to 2.4; positive = scores from 2.5 to 5.

To determine which statistical test to perform, data was tested for normality. Engagement rates of Tweets with negative and positive Syuzhet sentiment scores were found to have non-normal distribution using Shapiro-Wilk's test (W = 0.31275, p-value < 2.2e-16; W = 0.12002, p-value < 2.2e-16; respectively).

Wilcoxon rank-sum test was carried out to test if the mean engagement rates of Tweets with negative and positive Syuzhet sentiment scores are significantly different from each other and no significant difference was found at 5% significance level (W = 28874, p-value = 0.2396).

—-----------------------------------------------------------------

**Exploratory Analysis**

Note: Retrieved 22/01/09 17:09, 14936 entries before filtering

Top 5 tweets according to engagement rate:

1. *Take note: those pics of Seawall, Kits, Spanish Banks, just a hint of what climate change will do to city.* [*https://t.co/k2vO5T0lZU*](https://t.co/k2vO5T0lZU)18 retweets, 3 followers, 600% engagement rate   
   VADER score 0, Syuzhet score 0  
   Not a famous person, not a lot of followers, but the tweet quotes a famous Canadian meteorologist who has over 19k followers on twitter and that might explain the popularity of this tweet.
2. *@GretaThunberg is a joke. The elite propped up a kid to scream their "tax and control" agenda. She doesn't have real climate change solutions. She doesn't talk about biochar, urban farming, mycelium. She's absolutely clueless about practical climate mitigation efforts #byegreta*  
   2 retweets, 1 follower, 200% engagement   
   VADER score -0.380, Syuzhet score 1.50  
   Not informative - only two retweets and engagement rate is high due to low number of followers.
3. *@indyfromspace @doritmi Geologists : this is what an extinction level event looks like. Politicians : Climate change isn’t real.*58 retweets, 37 followers, 156.75% engagement rate   
   VADER score 0.361, Syuzhet score 1.15  
   Not a famous account. This tweet is a reply to another tweet from a Professor who has 80.6k followers which might explain its success.
4. *The recent blizzards &amp; subsequent loss of life in Pakistan's Punjab Province's hill station, #Murree are a direct consequence of man-made Climate Change. We need to learn the consequences of our actions, &amp; to own up to our mistakes. Please check the forecast before travelling.* [*https://t.co/dsHk4CcOEo*](https://t.co/dsHk4CcOEo)  
   9 retweets, 7 followers, 128.57% engagement rate   
   VADER score -0.361. Syuzhet score 0.15  
   Not informative. Engagement rate is high due to low number of followers.
5. *We really are seeing a clear distinction between Labour and LNP even if it's not readily apparent. Labour is about universal healthcare, fair wages for essential workers and progress on climate change. The LNP is about getting re elected. #auspol*28 retweets, 22 followers, 127.27% engagement rate   
   VADER score 0.772, Syuzhet score 4.40

Interestingly, the top 4 tweets with lowest engagement rates are all from the same user - The Economist. All have positive sentiment but low engagement rates (low rates are not just due to time of posting and data retrieval).