How do Elon Musk’s tweets move Tesla’s stock price？

Group 9: Edmund Lam, Peishan Li, Xinrun Shan, Jingru Tang

How do Elon Musk's tweets move Tesla's stock price ?

Hypothesis:

A correlation exists between short-term stock performance and Twitter sentiment.

➢ **Behavioural finance evidence:** financial decisions can be driven by emotions and mood

➢ **Techniques used:** NLP, Sentiment analysis, Linear regression, Logistic regression

Data sources:

1. **% change in stock price of Tesla** *(Yahoo Finance)*

$$\text{Stock Returns}(R_{i,t}) = \frac{Price_t - Price_{t-1}}{Price_{t-1}}$$

2. **Tweets from Elon Musk throughout 2021 via Twitter API**

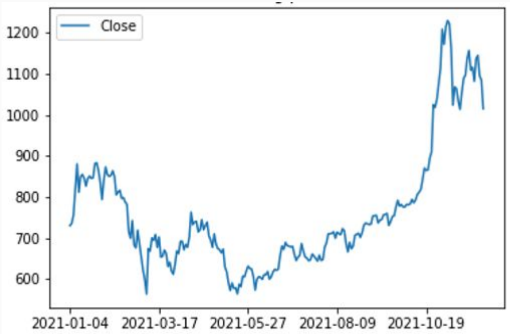
Data preprocessing

Clean
(lowercase; remove characters, stopwords, non-English words)
↓
Tokenization
↓
Lemmatization

	Date	
0	2021–01–28	[with, even, literally, great, game, the, dollar, short, indeed, great, use useful, product, provide, great, service, actually, matter, would, bette beholden, big, trading, house, get, pay, bill, somehow, even, fee, here,
1	2021–02–01	[it, fill, art, berlin, progress, work, super, hard, ensure, implant, safety, go, well, might, able,
2	2021–02–04	[yes, it, true, josh, yeah, by, default, engine, least, lever, arm, would throttle, point, risk, land, engine, mean, high, thrust, weight, away, foolish, u, start, engine,

Wordcloud visualization

Daily closing price of Tesla throughout 2021



MARKETS INSIDER

Subscribe

Elon Musk thinks investors should pay less attention to his tweets

- Elon Musk told Time that he doesn't think his tweets have much of an impact on the markets.
- "Markets move themselves all the time, based on nothing as far as I can tell," he said.

Polarity Categorization and Sentiment Scores

- TextBlob Model & Sentiment Dictionary of Hu & Liu (2004)

Numerical Sentiment Score Calculation:

➢ **TextBlob Model Lexicon Approach**
A sentiment or tone is defined by its **semantic orientation** and the intensity of each word in the sentence.
Assign each word in Elon Musk's tweet a sentiment score, final scores are calculated by taking the **average of all the sentiments in the bag of words**.

➢ **Sentiment Dictionary of Hu & Liu:**
Hu & Liu ('Mining and Summarizing Customer Reviews') provides **a list of 6800 words retrieved from Amazon categorized into Positive and Negative words**. The **rate of occurrence** of the pre-defined words provides information on the overall tone of each tweet.

➢ **Categorization:**
[0 - 1] : Positive Sentiment
0 : Neutral Sentiment
<0 : Negative Sentiment

```
#pip install textblob
import nltk
from textblob import TextBlob
```

```
def getPolarity(text):
    return TextBlob(text).sentiment.polarity
```

Out[27]:

	Unnamed: 0	Unnamed: 0.1	tweet	date	Same day stock closing price	Next day stock open price	Daily change	polarity
0	0	0	['gamespot', 'with', 'cyberpunk', 'even', 'hot...	2021-01-28	835.43	830.00	-0.65%	0.231818
1	1	1	['it', 'fill', 'graffiti', 'art', 'eiraum', 'g...	2021-02-01	839.81	844.68	0.58%	0.090278
2	2	2	['brendan2908', 'nasaspaceflight', 'yes', 'mik...	2021-02-04	849.99	845.00	-0.59%	0.119500
3	3	3	['the', 'entertaining', 'outcome', 'likely', '...	2021-02-10	804.82	812.44	0.95%	0.233333
4	4	4	['joerogan', 'spotify', 'great', 'interview', ...	2021-02-11	811.66	801.26	-1.28%	0.389394

Results: TextBlob

No statistical significant relationship discovered between characteristics of Elon Musk's tweets and price changes in Tesla's stock price.

- **Linear Regression Model**
  Independent Variable:
  Polarity
  Dependent Variable:
  Stock price daily percentage change
- **Logistic Regression Model**
  Independent Variable:
  Polarity
  Dependent Variable:
  Stock Price Change Direction

Sentiment Analysis: Linear Regression

	Dependent variable: DailyChange
polarity	0.0002
	(0.008)
Constant	0.002
	(0.002)
Observations	126
R2	0.00001
Adjusted R2	-0.008
Residual Std. Error	0.016 (df = 124)
F Statistic	0.001 (df = 1; 124)
Note:	*p<0.1; **p<0.05; ***p<0.01

Sentiment Analysis: Logistic Regression

	Dependent variable: Direction
polarity	-0.727
	(1.019)
Constant	0.658***
	(0.238)
Observations	126
Log Likelihood	-82.432
Akaike Inf. Crit.	168.865
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: Hu & Liu Dictionary (2004)

No statistical significant relationship discovered between characteristics of Elon Musk's tweets and price changes in Tesla's stock price.

- ➢ **Linear Regression Model**
  Independent Variable:
  Polarity
  Dependent Variable:
  Stock price daily percentage change
- ➢ **Logistic Regression Model**
  Independent Variable:
  Polarity
  Dependent Variable:
  Binary: Stock Price Change Direction

Sentiment Analysis: Linear Regression

	Dependent variable: DailyChange
Polarity	-0.002
	(0.002)
Constant	0.002
	(0.002)
Observations	113
R2	0.009
Adjusted R2	-0.0001
Residual Std. Error	0.015 (df = 111)
F Statistic	0.985 (df = 1; 111)
Note:	*p<0.1; **p<0.05; ***p<0.01

Sentiment Analysis: Logistic Regression

	Dependent variable: Direction
Polarity	-0.231
	(0.308)
Constant	0.555**
	(0.216)
Observations	113
Log Likelihood	-74.785
Akaike Inf. Crit.	153.570
Note:	*p<0.1; **p<0.05; ***p<0.01

Reference: Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews.* KDD 04.

Identifying Emotions in Tweets

Identify 8 emotions, including **anger, anticipation, disgust, fear, joy, sadness, surprise, trust**, as well as 2 sentiments, **negative and positive**, using NRC Word-Emotion Association Lexicon in the `tidytext` package in R.

Take angry words as an example:

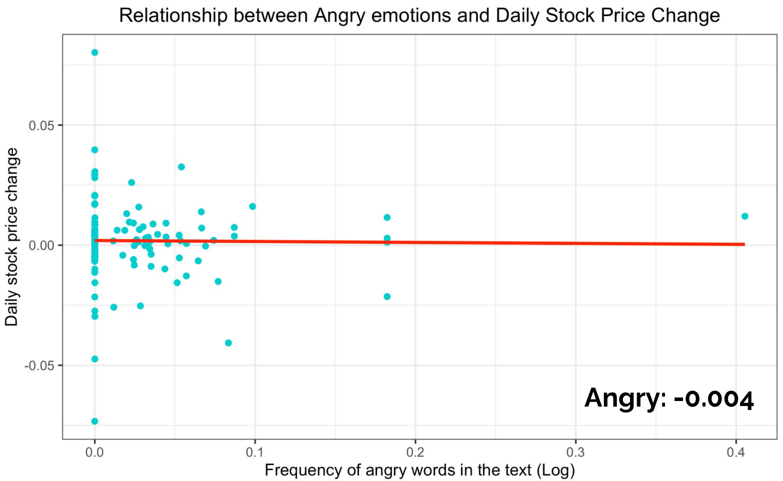
```
```{r NRC, echo=TRUE, eval=TRUE}
nrc_anger<-get_sentiments("nrc") %>%
  filter(sentiment=="anger")
```{r}
nrc_anger$word
```
```

```
 [1] "abandoned"      "abandonment"    "abhor"          "abhorrent"
 [5] "abolish"        "abomination"    "abuse"          "accursed"
 [9] "accusation"     "accused"        "accuser"        "accusing"
[13] "actionable"     "adder"          "adversary"      "adverse"
[17] "adversity"      "advocacy"       "affront"        "aftermath"
[21] "aggravated"     "aggravating"    "aggravation"    "aggression"
[25] "aggressive"     "aggressor"      "agitated"       "agitation"
[29] "agony"          "alcoholism"     "alienate"       "alienation"
[33] "allegation"     "altercation"    "ambush"         "anarchism"
[37] "anarchist"      "anarchy"        "anathema"       "anger"
[41] "angry"          "anguish"        "animosity"      "animus"
[45] "annihilate"     "annihilated"    "annihilation"   "annoy"
[49] "annoyance"      "annoying"       "antagonism"     "antagonist"
[53] "antagonistic"   "antichrist"     "antipathy"      "antisocial"
```

```
angryf<-function(words=c("Check out the frequency of words that represent emotions")){
  tok<-tokens(words)
  wordcount<-length(tok[[1]])
  angercount<-sum(tok[[1]] %in% nrc_anger$word)
  angerf<-angercount/wordcount
  return(angerf)
}
```

Calculate frequency of angry words for each tweet
```

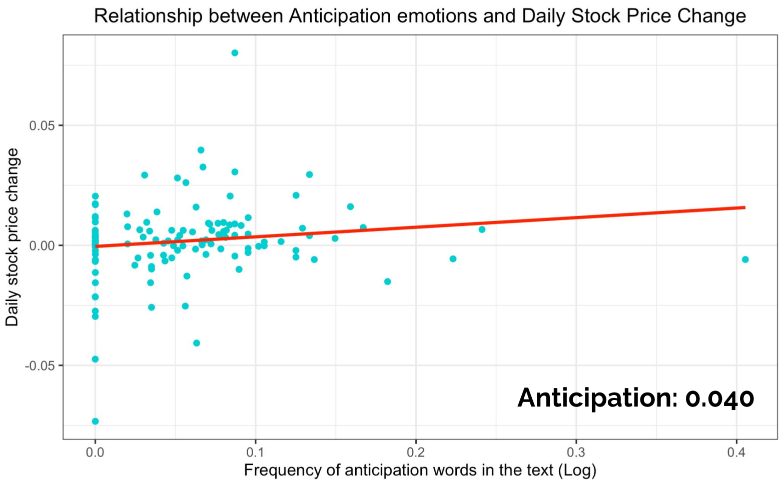
The dataset will look like this:

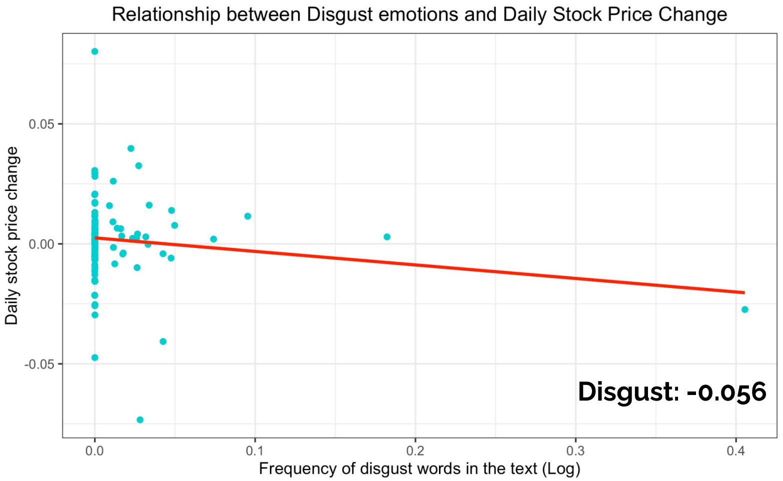
tweet	date	DailyChange	angryf	anticipationf	disgustf	fearf	joyf	sadnessf	surprisef	trustf	negativef	positivef
with even literally great game the dollar short indeed ...	2021-01-28	-0.0065	0.06666667	0.04444444	0.000000000	0.000000000	0.044444444	0.000000000	0.000000000	0.08888889	0.11111111	0.06666667
it fill art berlin progress work super hard ensure impl...	2021-02-01	0.0058	0.00000000	0.08333333	0.000000000	0.000000000	0.083333333	0.041666667	0.041666667	0.08333333	0.00000000	0.08333333
yes it true josh yeah by default engine least lever arm...	2021-02-04	-0.0059	0.02439024	0.14634146	0.048780488	0.097560976	0.073170732	0.048780488	0.048780488	0.07317073	0.17073171	0.14634146
the entertaining outcome likely just send just agree cl...	2021-02-10	0.0095	0.00000000	0.08333333	0.000000000	0.000000000	0.083333333	0.000000000	0.000000000	0.08333333	0.00000000	0.33333333
great interview doge underestimate congratulation ch...	2021-02-11	-0.0128	0.05882353	0.05882353	0.000000000	0.058823529	0.058823529	0.058823529	0.058823529	0.11764706	0.05882353	0.05882353
cover snow ice road mostly close power there definite...	2021-02-16	-0.0215	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.09090909	0.00000000	0.00000000
south central provide direct assistance feasible save c...	2021-02-18	0.0097	0.00000000	0.00000000	0.000000000	0.000000000	0.095238095	0.000000000	0.000000000	0.14285714	0.04761905	0.19047619
we much product complexity already tunnel hopefully...	2021-02-22	-0.0733	0.00000000	0.00000000	0.028571429	0.000000000	0.028571429	0.000000000	0.028571429	0.02857143	0.05714286	0.02857143
jeff rocket ridiculously hard sigh	2021-02-24	-0.0214	0.20000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
we use catapult air mattress land roof base jump nick...	2021-02-25	0.0261	0.02325581	0.05813953	0.011627907	0.011627907	0.034883721	0.023255814	0.023255814	0.05813953	0.03488372	0.16279070
note still well long term deal subscription sure green ...	2021-03-01	-0.0002	0.02500000	0.05000000	0.000000000	0.000000000	0.050000000	0.050000000	0.025000000	0.10000000	0.05000000	0.12500000
space from thence mar and hence star an area much l...	2021-03-02	0.0023	0.00000000	0.03846154	0.000000000	0.000000000	0.076923077	0.000000000	0.000000000	0.11538462	0.03846154	0.15384615
launch abort slightly conservative high thrust limit inc...	2021-03-03	0.0040	0.00000000	0.14285714	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.04761905	0.09523810
ford go bankrupt car easy production hard cash flow ...	2021-03-04	0.0074	0.09090909	0.18181818	0.000000000	0.272727273	0.090909091	0.000000000	0.090909091	0.09090909	0.18181818	0.27272727
not connect car terminal much big this aircraft ship la...	2021-03-08	0.0802	0.00000000	0.09090909	0.000000000	0.090909091	0.000000000	0.000000000	0.090909091	0.09090909	0.09090909	0.00000000
fair point if autogenous use bubble would likely rever...	2021-03-09	0.0397	0.00000000	0.06818182	0.022727273	0.045454545	0.045454545	0.022727273	0.045454545	0.04545455	0.04545455	0.06818182
doge i selling song	2021-03-15	-0.0065	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
actually feel quite right selling will pass	2021-03-16	-0.0296	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
we need launch tower hook height lift booster ship ne...	2021-03-18	-0.0100	0.00000000	0.09375000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.12500000
wow shame	2021-03-24	-0.0274	0.00000000	0.00000000	0.500000000	0.500000000	0.000000000	0.000000000	0.500000000	0.00000000	0.50000000	0.00000000
shame onion this people switch my father extend fam...	2021-03-25	0.0023	0.00000000	0.07142857	0.023809524	0.023809524	0.023809524	0.000000000	0.047619048	0.11904762	0.07142857	0.14285714
that would great possibly via video full access telemetry	2021-03-29	-0.0156	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.11111111
can little foggy sometimes barely scratch back stand ...	2021-03-30	0.0173	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
true over people need min airport min downtown righ...	2021-03-31	0.0306	0.00000000	0.09090909	0.000000000	0.000000000	0.090909091	0.000000000	0.000000000	0.09090909	0.00000000	0.09090909
tank lake	2021-04-06	-0.0067	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
a tidal wave vaccine produce probably j j good some ...	2021-04-07	0.0096	0.02173913	0.03260870	0.000000000	0.043478261	0.021739130	0.054347826	0.032608696	0.03260870	0.05434783	0.18478261
look pup soon shock absorption build tower arm sinc...	2021-04-08	-0.0088	0.03571429	0.03571429	0.000000000	0.035714286	0.000000000	0.071428571	0.000000000	0.03571429	0.07142857	0.14285714
thanks cast writer crew honor pleasure show	2021-05-10	-0.0474	0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.42857143	0.00000000	0.28571429

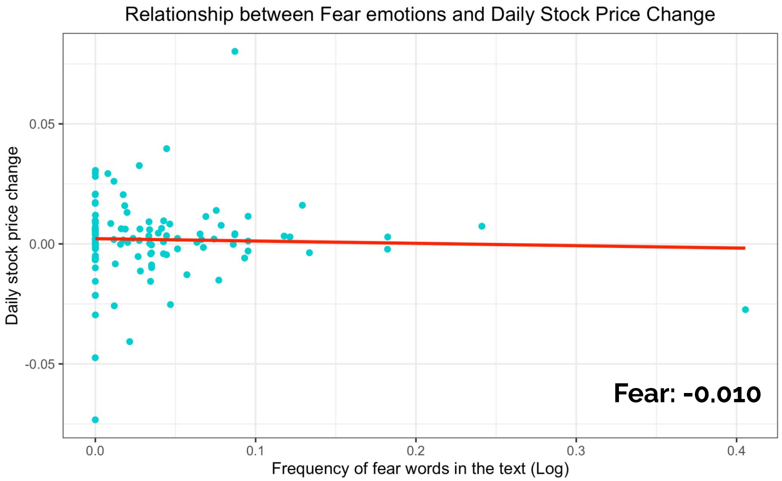
The dataset (Cont.) lemotionf=log(emotionf+1)

langryf	lanticipationf	ldisgustf	lfearf	ljoyf	lsadnessf	lsurprisef	ltrustf	lnegativef	lpositivef
0.06453852	0.04348511	0.000000000	0.000000000	0.043485112	0.000000000	0.000000000	0.08515781	0.10536052	0.06453852
0.00000000	0.08004271	0.000000000	0.000000000	0.080042708	0.040821995	0.040821995	0.08004271	0.00000000	0.08004271
0.02409755	0.13657554	0.047628049	0.093090423	0.070617567	0.047628049	0.047628049	0.07061757	0.15762894	0.13657554
0.00000000	0.08004271	0.000000000	0.000000000	0.080042708	0.000000000	0.000000000	0.08004271	0.00000000	0.28768207
0.05715841	0.05715841	0.000000000	0.057158414	0.057158414	0.057158414	0.057158414	0.11122564	0.05715841	0.05715841
0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.08701138	0.00000000	0.00000000
0.00000000	0.00000000	0.000000000	0.000000000	0.090971778	0.000000000	0.000000000	0.13353139	0.04652002	0.17435339
0.00000000	0.00000000	0.028170877	0.000000000	0.028170877	0.000000000	0.028170877	0.02817088	0.05556985	0.02817088
0.18232156	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
0.02298952	0.05651221	0.011560822	0.011560822	0.034289073	0.022989518	0.022989518	0.05651221	0.03428907	0.15082289
0.02469261	0.04879016	0.000000000	0.000000000	0.048790164	0.048790164	0.024692613	0.09531018	0.04879016	0.11778304
0.00000000	0.03774033	0.000000000	0.000000000	0.074107972	0.000000000	0.000000000	0.10919929	0.03774033	0.14310084
0.00000000	0.13353139	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.04652002	0.09097178
0.08701138	0.16705408	0.000000000	0.241162057	0.087011377	0.000000000	0.087011377	0.08701138	0.16705408	0.24116206
0.00000000	0.08701138	0.000000000	0.087011377	0.000000000	0.000000000	0.087011377	0.08701138	0.08701138	0.00000000
0.00000000	0.06595797	0.022472856	0.044451763	0.044451763	0.022472856	0.044451763	0.04445176	0.04445176	0.06595797
0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000
0.00000000	0.08961216	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.11778304
0.00000000	0.00000000	0.405465108	0.405465108	0.000000000	0.000000000	0.405465108	0.00000000	0.40546511	0.00000000
0.00000000	0.06899287	0.023530497	0.023530497	0.023530497	0.000000000	0.046520016	0.11247798	0.06899287	0.13353139
0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.10536052

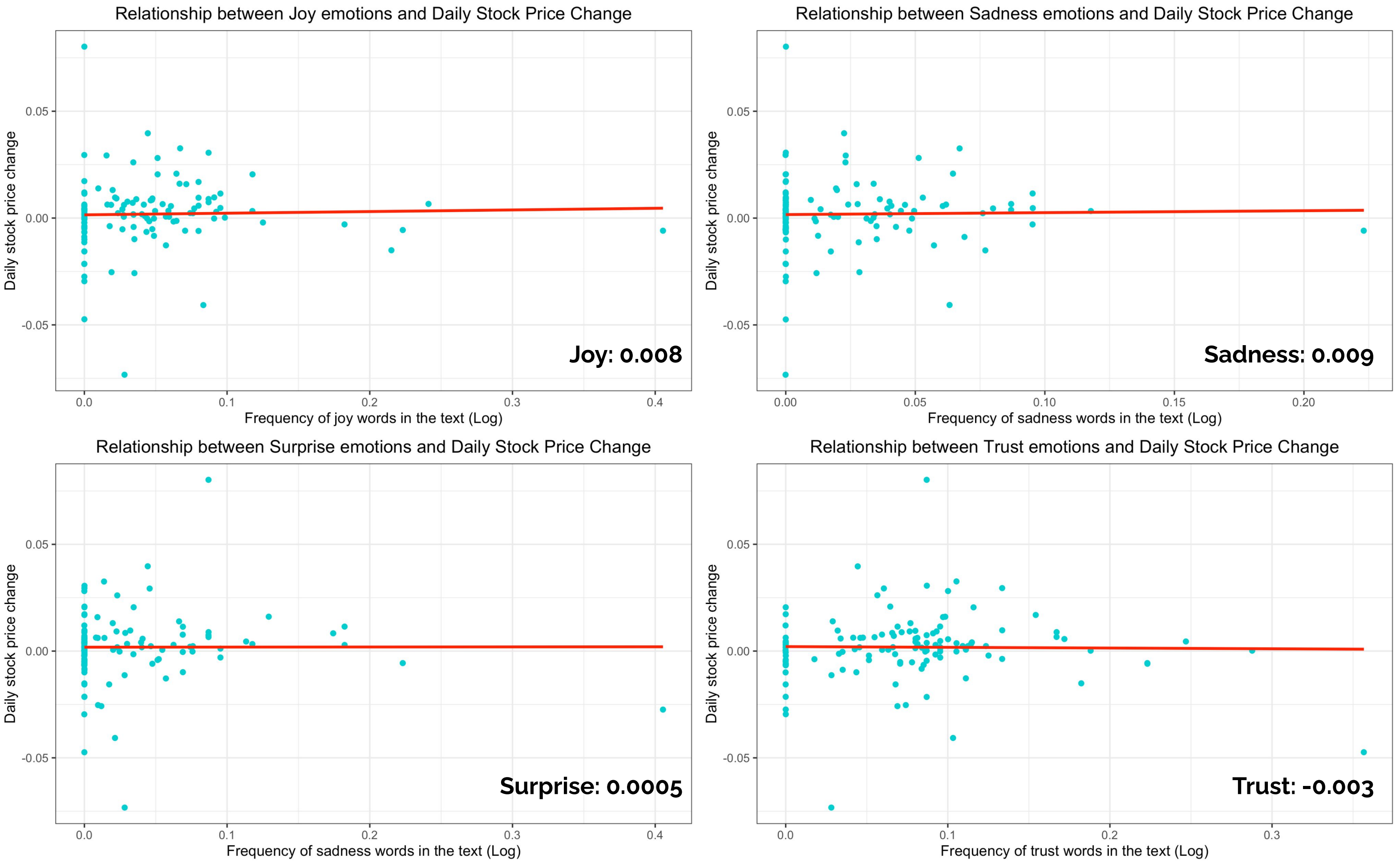




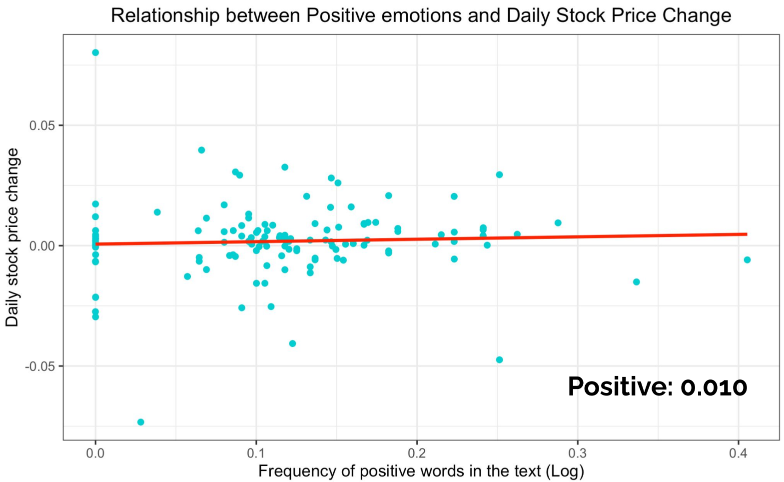


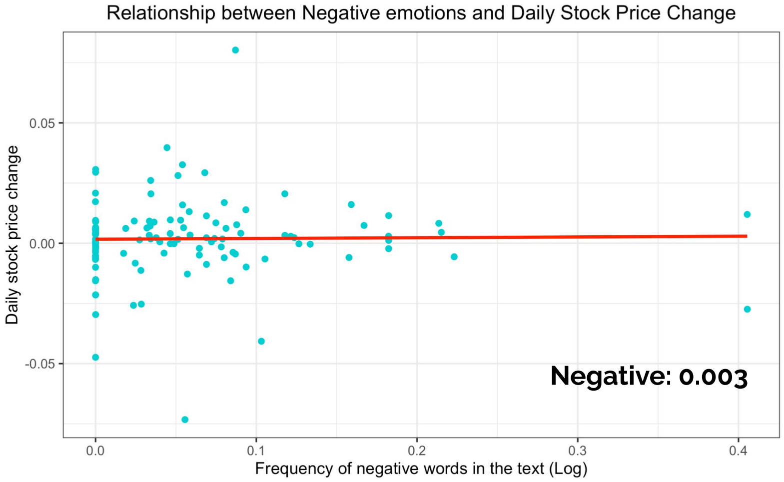


Linear Regression & Visualizations



Linear Regression & Visualizations (Cont)





Emotion Analysis: Multilinear Regression Result

	Dependent variable:
	DailyChange
langryf	0.003***
	(0.001)
lanticipationf	-0.004**
	(0.002)
ldisgustf	0.0001
	(0.001)
lfearf	0.0002
	(0.001)
ljoyf	-0.001
	(0.002)
lsadnessf	0.0002
	(0.001)
lsurprisef	0.0003
	(0.001)
ltrustf	-0.002
	(0.002)
lnegativef	-0.003*
	(0.002)
lpositivef	0.007***
	(0.002)
Constant	0.001
	(0.003)
Observations	122
R2	0.144
Adjusted R2	0.067
Residual Std. Error	0.016 (df = 111)
F Statistic	1.866* (df = 10; 111)
Note:	*p<0.1; **p<0.05; ***p<0.01

When performing multi-regression analysis, the log transformation of frequency of angry, anticipation, negative and angry words significantly decreases daily stock price percentage change:

1% increase of angry word frequency leads to 0.003 increase in stock price change, keeping other factors fixed;

1% increase of angry word frequency leads to 0.004 decrease in stock price change, keeping other factors fixed;

1% increase of positive word frequency leads to 0.007 increase in stock price change, keeping other factors fixed;

1% increase of negative word frequency leads to 0.003 decrease in stock price change, keeping other factors fixed.

Linear Regression & Visualizations (Cont), Robustness Check

Difference in word frequency plots & Logistic Regression

- We calculated word frequency for tweets associated with stock price increase & decrease separately
- We then filtered DataFrame with only words contained in the dictionary of each emotion category.
- Take sadness for example: in the graph below, i_tf_idf represents the tf_idf score for each sadness word in tweets associated with a stock price increase the day it was sent, and d_tf_idf decrease. Visualizations are based on the DataFrames.

term <chr>	i_tf_idf <dbl>	d_tf_idf <dbl>
mad	0.86876108	0.00000000
die	0.52152261	0.11535341
death	0.48264505	0.00000000
sue	0.48264505	0.31722187
bankrupt	0.43438054	0.00000000
sing	0.43438054	0.00000000
remove	0.43438054	0.00000000
tax	0.36457799	0.14364764
art	0.33187802	0.00000000
terminal	0.26886464	0.00000000

Logistic Regression

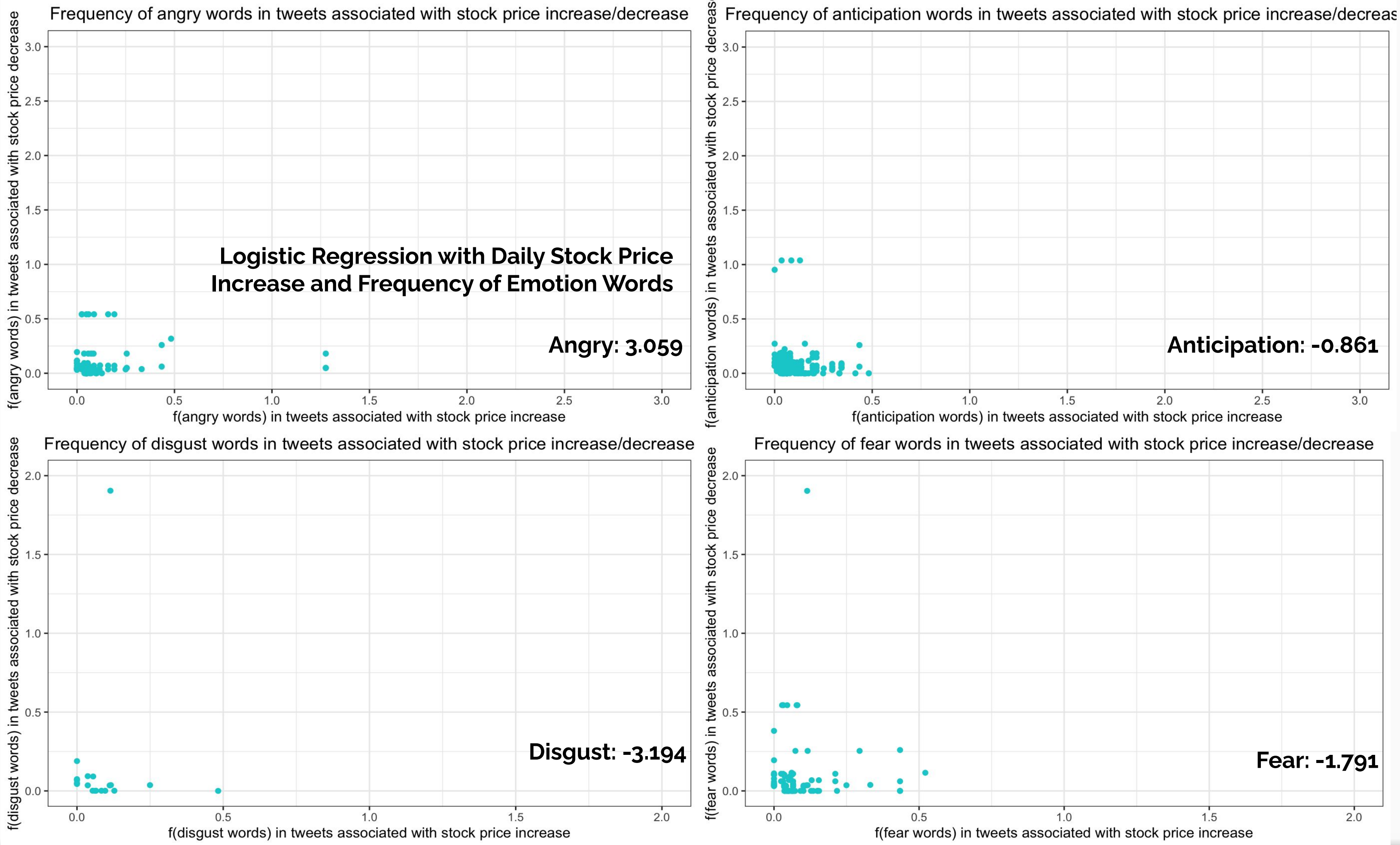
We've re-coded Daily Stock Price Change into three categories: 1 as Increase in stock price, -1 as Decrease in stock price, and 0 as No changes.

Frequency of words are taken logarithm to reflect the percentage change.

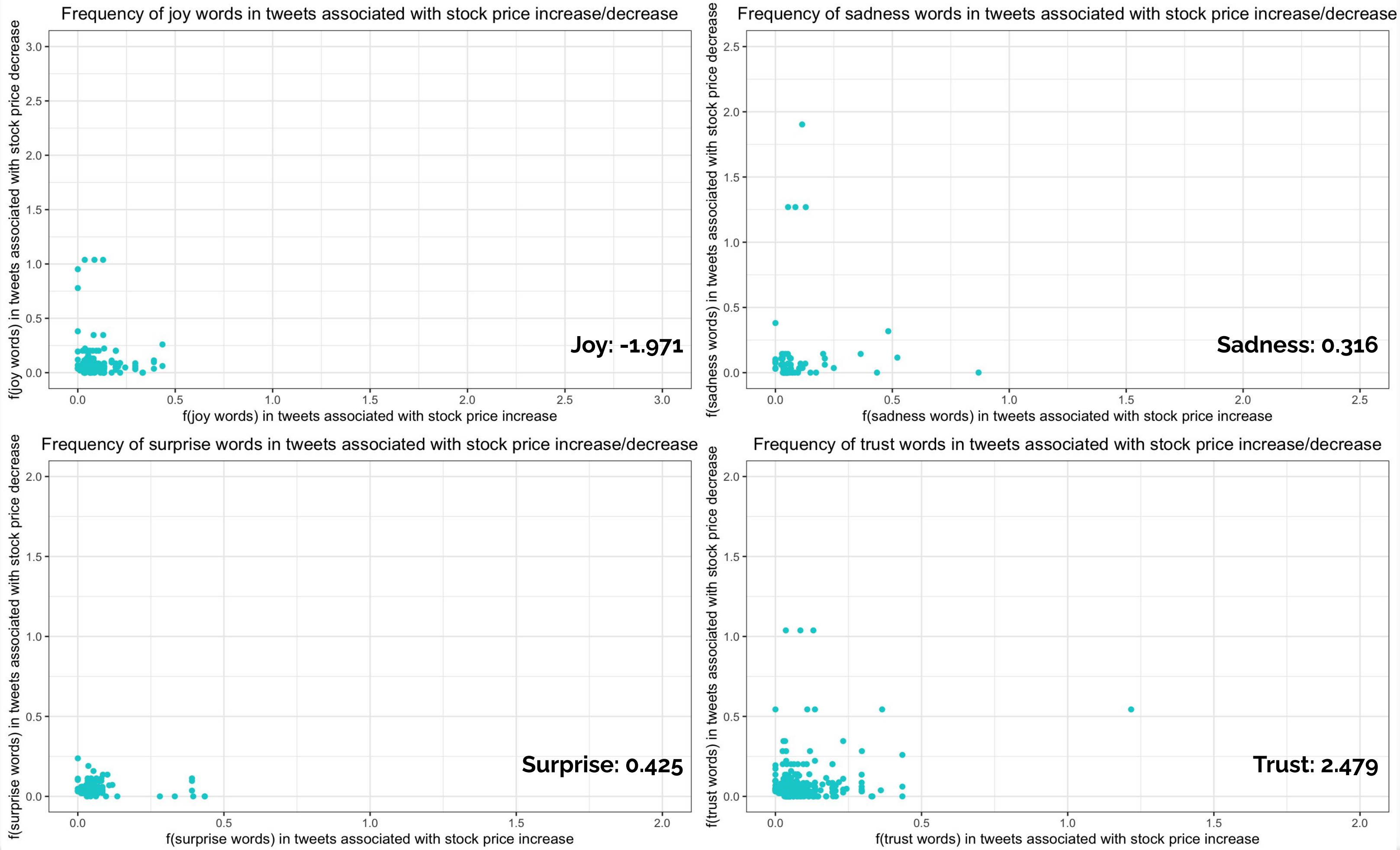
Again, take sadness as example:

```
logitsadness<-glm(Direction ~ lsadnessf,
emotionsfrequency, family=binomial(link ="logit"))
summary(logitsadness)
```

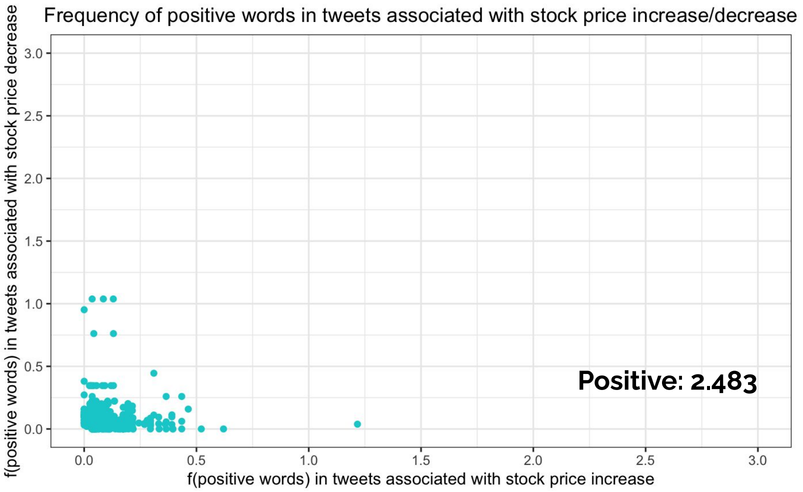
langryf	lanticipationf	ldisgustf	lfearf	ljoyf	lsadnessf	lsurprisef	ltrustf	lnegativef	lpositivef	Direction
0.06453852	0.04348511	0.000000000	0.000000000	0.043485112	0.000000000	0.000000000	0.08515781	0.10536052	0.06453852	−1
0.00000000	0.08004271	0.000000000	0.000000000	0.080042708	0.040821995	0.040821995	0.08004271	0.00000000	0.08004271	1
0.02409755	0.13657554	0.047628049	0.093090423	0.070617567	0.047628049	0.047628049	0.07061757	0.15762894	0.13657554	−1
0.00000000	0.08004271	0.000000000	0.000000000	0.080042708	0.000000000	0.000000000	0.08004271	0.00000000	0.28768207	1
0.05715841	0.05715841	0.000000000	0.057158414	0.057158414	0.057158414	0.057158414	0.11122564	0.05715841	0.05715841	−1
0.00000000	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.08701138	0.00000000	0.00000000	−1
0.00000000	0.00000000	0.000000000	0.000000000	0.090971778	0.000000000	0.000000000	0.13353139	0.04652002	0.17435339	1
0.00000000	0.00000000	0.028170877	0.000000000	0.028170877	0.000000000	0.028170877	0.02817088	0.05556985	0.02817088	−1
0.18232156	0.00000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.00000000	0.00000000	0.00000000	−1
0.02298952	0.05651221	0.011560822	0.011560822	0.034289073	0.022989518	0.022989518	0.05651221	0.03428907	0.15082289	1

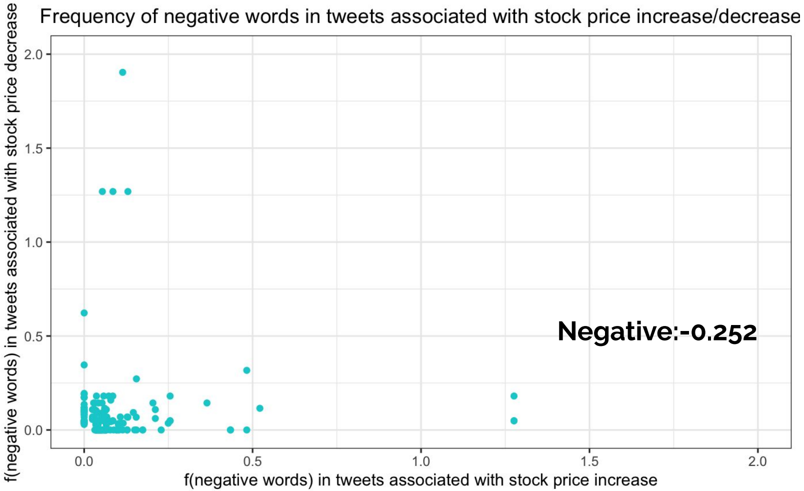


Logistic Regression & Visualizations on Increase/Decrease word frequency difference



Logistic Regression & Visualizations on Increase/Decrease word frequency difference (Cont)





Emotion Analysis: Multi-logistic Regression Result

	Dependent variable:
	Direction
langryf	5.647
	(5.067)
lanticipationf	-2.380
	(4.909)
ldisgustf	-7.209
	(8.111)
lfearf	-5.451
	(5.794)
ljoyf	-11.350*
	(6.498)
lsadnessf	4.862
	(7.987)
lsurprisef	12.813*
	(7.398)
ltrustf	2.327
	(4.224)
lpositivef	7.348*
	(3.848)
lnegativef	-2.879
	(5.006)
Constant	-0.043
	(0.446)
Observations	122
Log Likelihood	-75.710
Akaike Inf. Crit.	173.421
Note:	*p<0.1; **p<0.05; ***p<0.01

When performing multi-regression analysis, the percentage change of frequency of joy words significantly decreases logit of daily stock price change by 11.350, keeping other factors fixed.

When performing multi-regression analysis, the percentage change of frequency of surprise words significantly increases logit of daily stock price change by 12.813, keeping other factors fixed.

When performing multi-regression analysis, the percentage change of frequency of positive words significantly increases logit of daily stock price change by 7.348, keeping other factors fixed.

Conclusion and Limitations

- No significant relationship between sentiment and price change
- Some evidence of effects of certain emotions
- Large number of factors affect both stock price and Tweet content

- Data sampled only from 2021 for specific hours
- Other features of stock price may be relevant dependent variables
- Sentiment scores equally weights positive and negative words that can cancel out
- Did not make use of time component