Results from Sentiment.sql:

1. Users:

1.1 Summary of the data set:

#users: the number of distinct users who posted at least one tweet
--code: SELECT COUNT(DISTINCT twitter_id) FROM volkswagen_15;

(1) theranos_1: theranos for 9th Oct. 2015 : one day

#users	#tweets	#mentions	#RTs	URL(%)
156	156		69	

(2) theranos_15: theranos for 16th Oct - 30th Oct 2015 : 15 days

#users	#tweets	#mentions	#RTs	URL(%)
62132	66449		19501	

(3) volkswagen_1: volkswagen for 11th Sept. 2015 : one day

#users	#tweets	#mentions	#RTs	URL(%)
2119	2119		553	

(4) volkswagen_15: volkswagen for 18th Sept - 2nd Oct 2015: 15 days

#users	#tweets	#mentions	#RTs	URL(%)
519042	529445		202203	

**QUESTION: mitigating error. Total tweets have both false positives (irrelevant tweets about the event containing the keyword) and true negatives (tweets about the event that do not include the keyword)

1.2 Number of all tweets, retweets, company responses according to the ymd

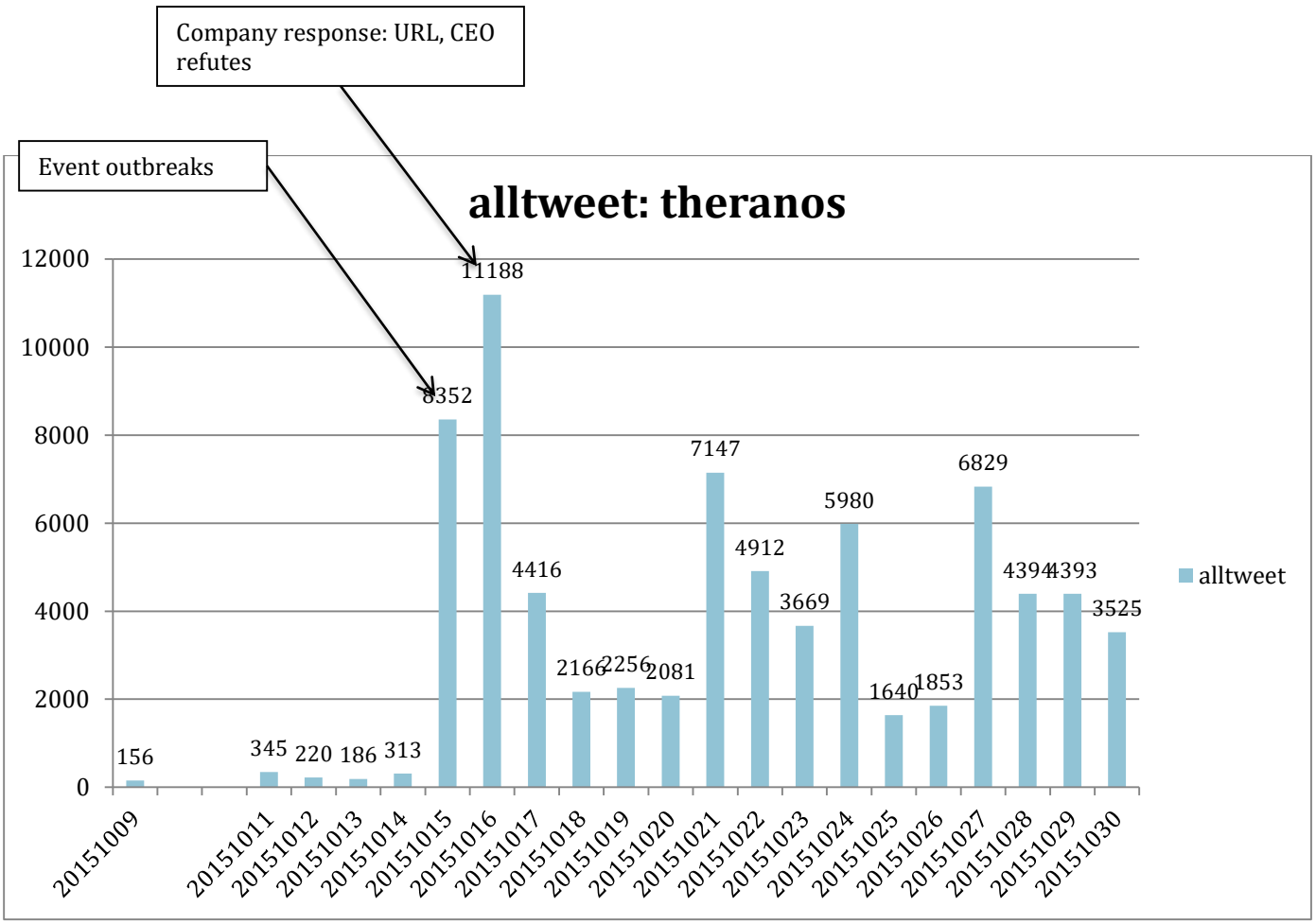
(1) Theranos

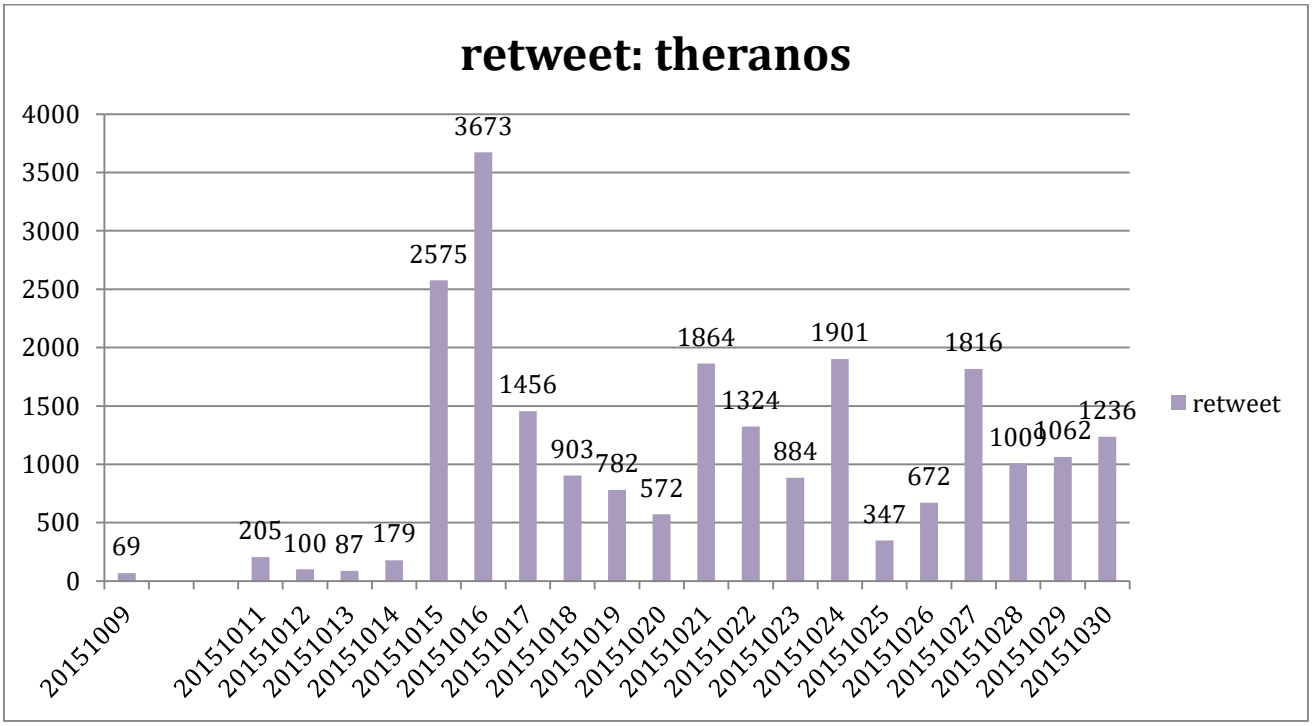
theranos_1: theranos for 9th Oct. 2015 : one day
theranos_5: theranos for 11-15th Oct. 2015 : five days
theranos_15: theranos for 16th Oct - 30th Oct 2015 : 15 days

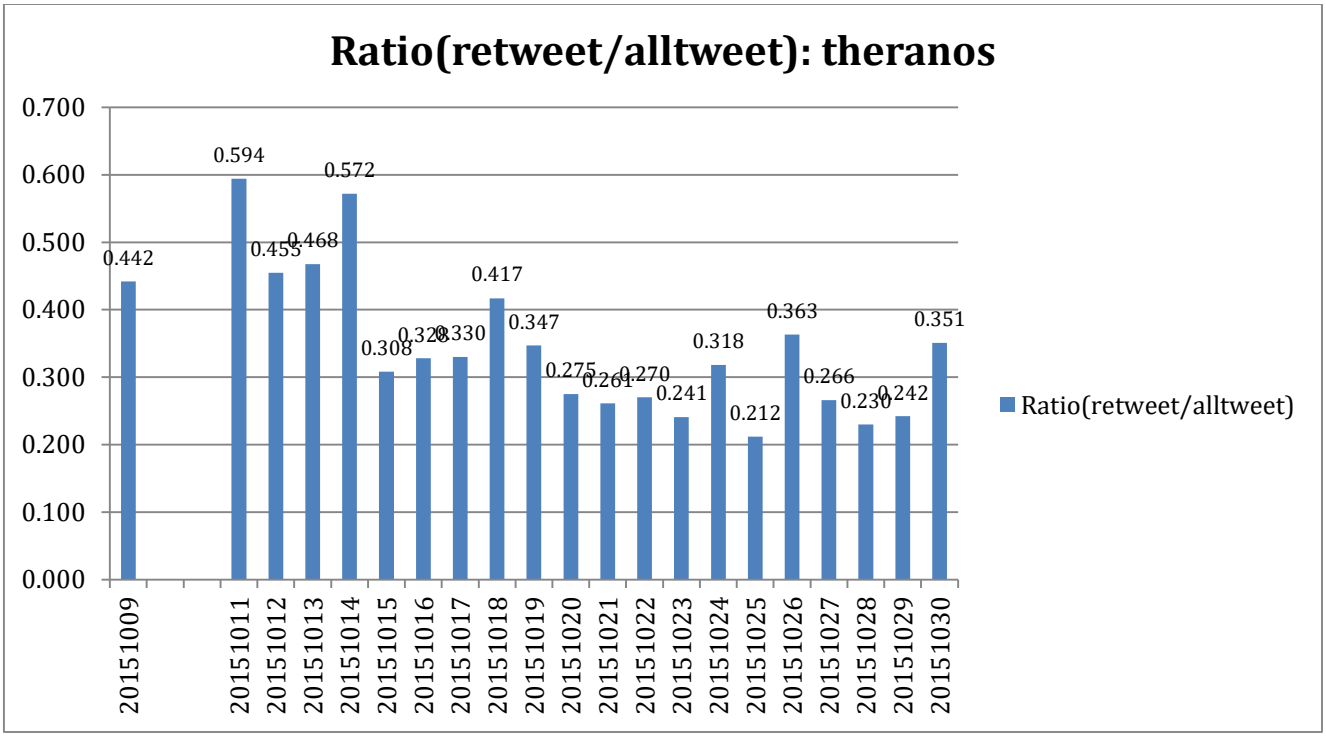
***15th Oct. break out

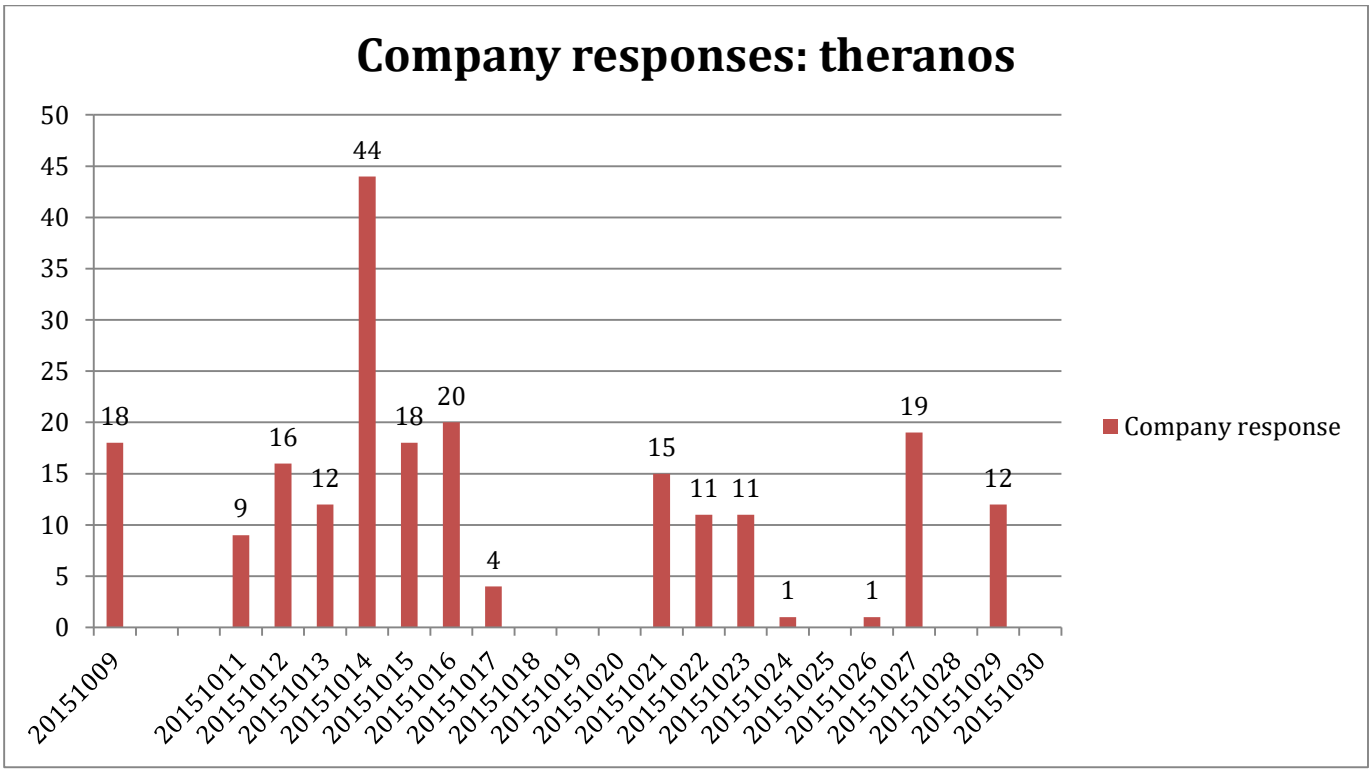
**Company responses: when displayname="theranos" or displayname="elizabeth holmes"

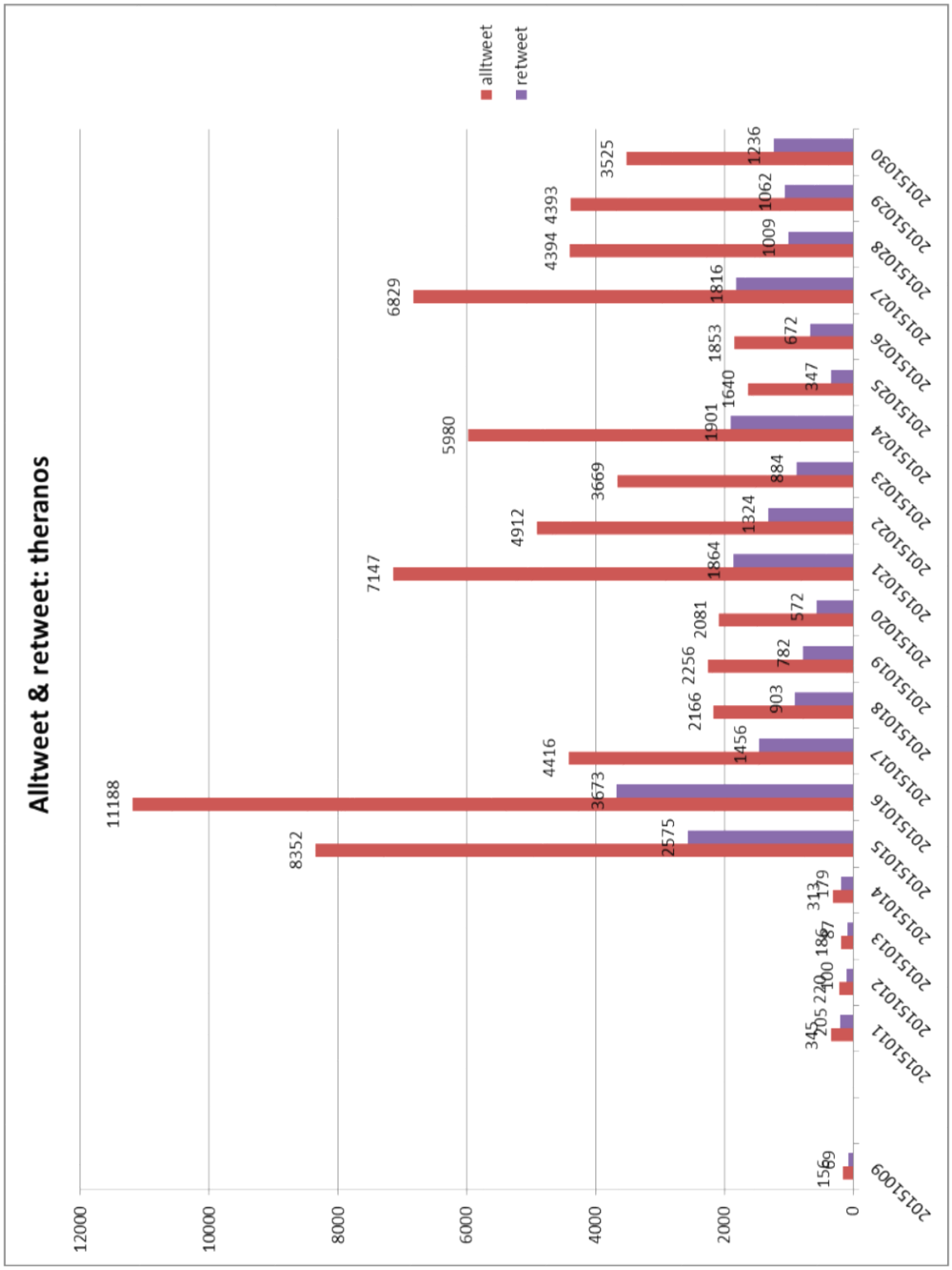
ymd	alltweet	retweet	Ratio(retweet/alltweet)	Company response
20151009	156	69	0.442	18
20151011	345	205	0.594	9
20151012	220	100	0.455	16
20151013	186	87	0.468	12
20151014	313	179	0.572	44
20151015	8352	2575	0.308	18
20151016	11188	3673	0.328	20
20151017	4416	1456	0.330	4
20151018	2166	903	0.417	
20151019	2256	782	0.347	
20151020	2081	572	0.275	
20151021	7147	1864	0.261	15
20151022	4912	1324	0.270	11
20151023	3669	884	0.241	11
20151024	5980	1901	0.318	1
20151025	1640	347	0.212	
20151026	1853	672	0.363	1
20151027	6829	1816	0.266	19
20151028	4394	1009	0.230	
20151029	4393	1062	0.242	12
20151030	3525	1236	0.351	


Company response: URL, CEO refutes
Event outbreaks
alltweet: theranos
12000
10000
8000
6000
4000
2000
0
156
345
220
186
313
8352
11188
4416
2166
2256
2081
7147
4912
3669
5980
1640
1853
6829
4394
4393
3525
20151009
20151011
20151012
20151013
20151014
20151015
20151016
20151017
20151018
20151019
20151020
20151021
20151022
20151023
20151024
20151025
20151026
20151027
20151028
20151029
20151030
alltweet


retweet: theranos
4000
3500
3000
2500
2000
1500
1000
500
0
69
205
100
87
179
2575
3673
1456
903
782
572
1864
1324
884
1901
347
672
1816
1009
1062
1236
20151009
20151011
20151012
20151013
20151014
20151015
20151016
20151017
20151018
20151019
20151020
20151021
20151022
20151023
20151024
20151025
20151026
20151027
20151028
20151029
20151030
retweet


Ratio(retweet/alltweet): theranos
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
0.442
0.594
0.455
0.468
0.572
0.308
0.328
0.330
0.417
0.347
0.275
0.261
0.270
0.241
0.318
0.212
0.363
0.266
0.230
0.242
0.351
20151009
20151011
20151012
20151013
20151014
20151015
20151016
20151017
20151018
20151019
20151020
20151021
20151022
20151023
20151024
20151025
20151026
20151027
20151028
20151029
20151030
Ratio(retweet/alltweet)


Company responses: theranos
50
45
40
35
30
25
20
15
10
5
0
18
9
16
12
44
18
20
4
15
11
11
1
1
19
12
20151009
20151011
20151012
20151013
20151014
20151015
20151016
20151017
20151018
20151019
20151020
20151021
20151022
20151023
20151024
20151025
20151026
20151027
20151028
20151029
20151030
Company response



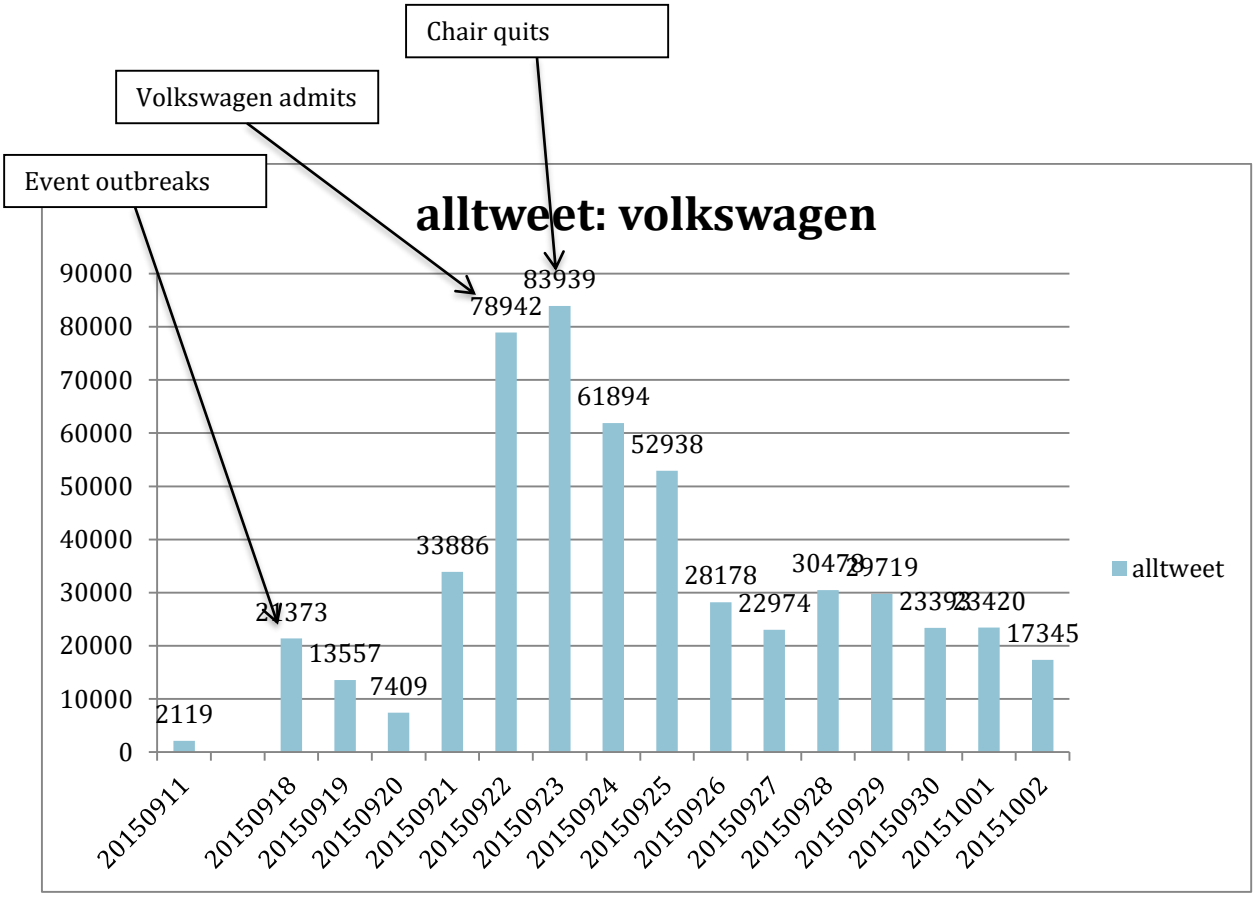
(2) Volkswagen

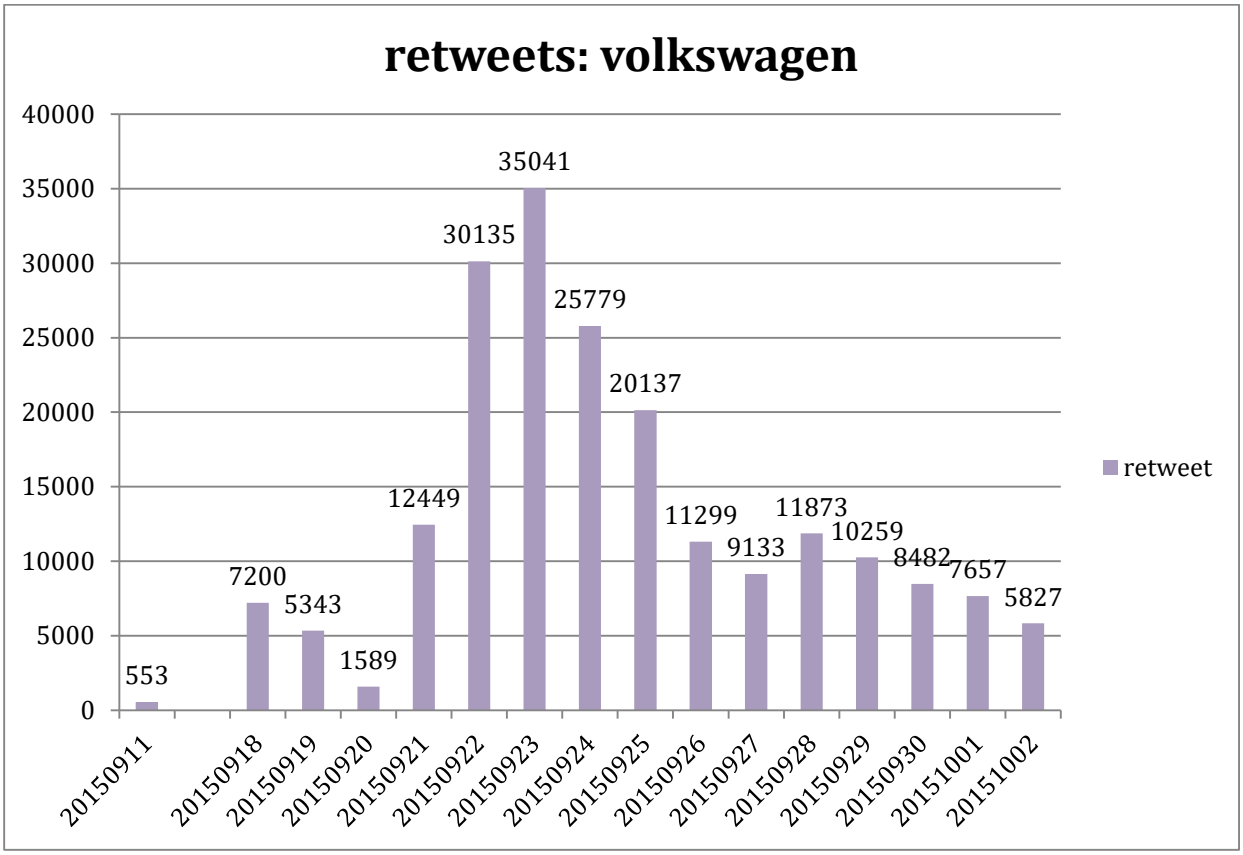
volkswagen_1: volkswagen for 11th Sept. 2015 : one day

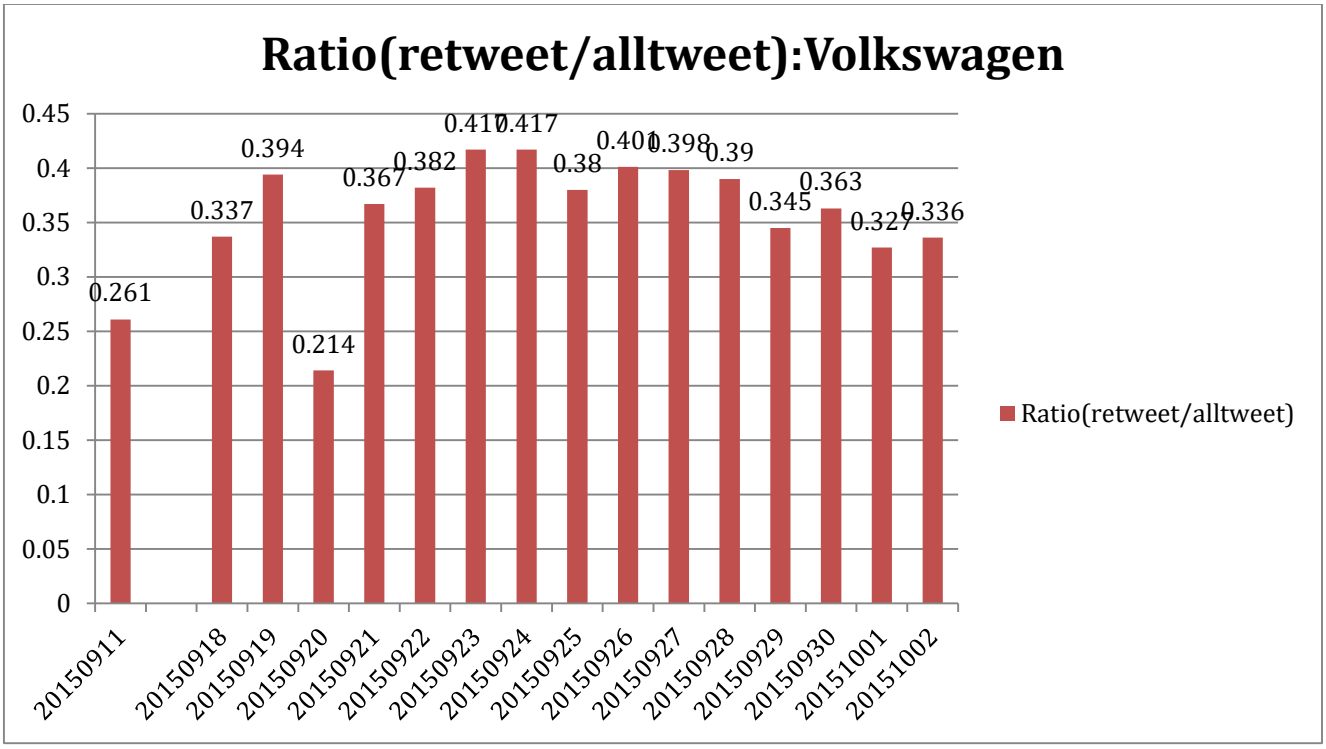
volkswagen_15: volkswagen for 18th Sept - 2nd Oct 2015: 15 days

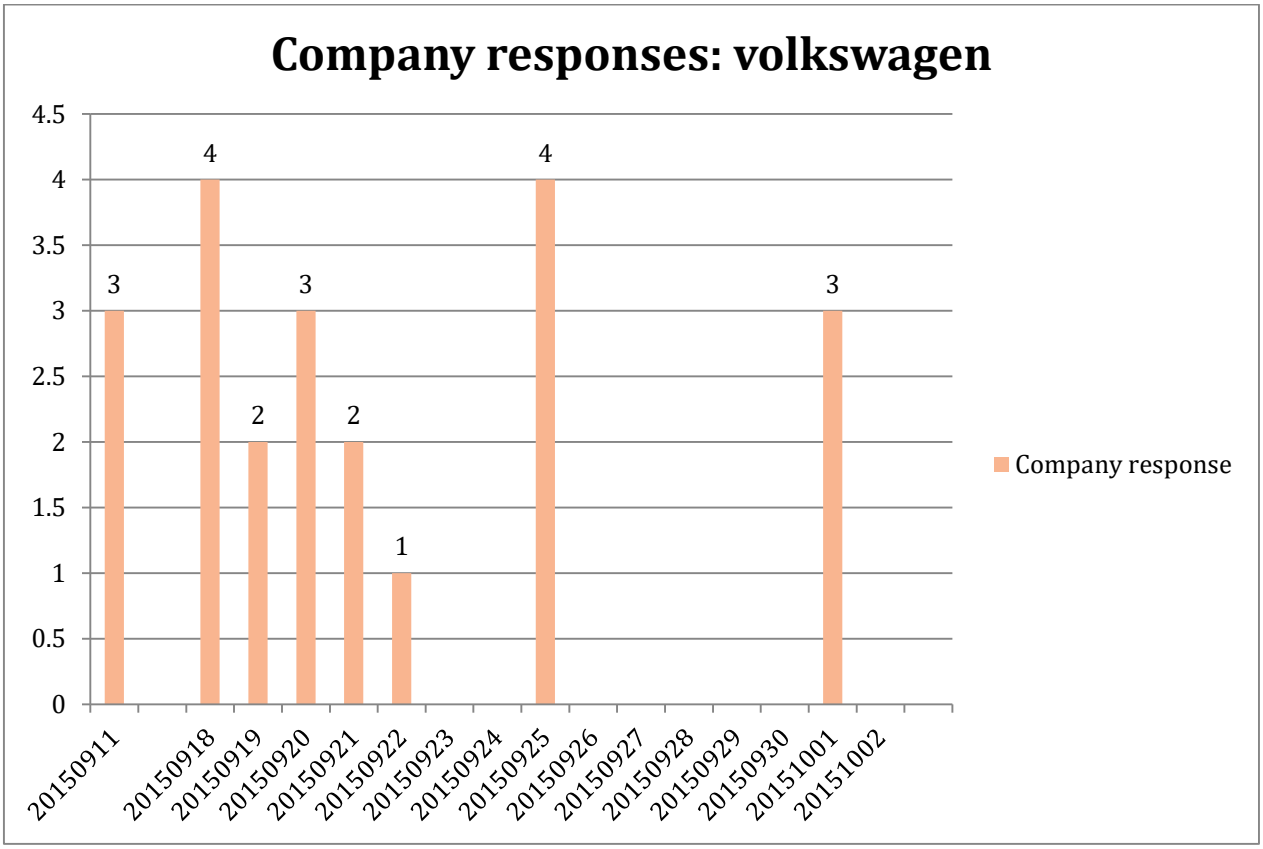
**Company responses: displayname="volkswagen" *and* preferredUsername="volkswagen"

ymd	alltweet	retweet	Ratio(retweet/alltweet)	Company response
20150911	2119	553	0.261	3
20150918	21373	7200	0.337	4
20150919	13557	5343	0.394	2
20150920	7409	1589	0.214	3
20150921	33886	12449	0.367	2
20150922	78942	30135	0.382	1
20150923	83939	35041	0.417	
20150924	61894	25779	0.417	
20150925	52938	20137	0.380	4
20150926	28178	11299	0.401	
20150927	22974	9133	0.398	
20150928	30478	11873	0.390	
20150929	29719	10259	0.345	
20150930	23393	8482	0.363	
20151001	23420	7657	0.327	3
20151002	17345	5827	0.336	

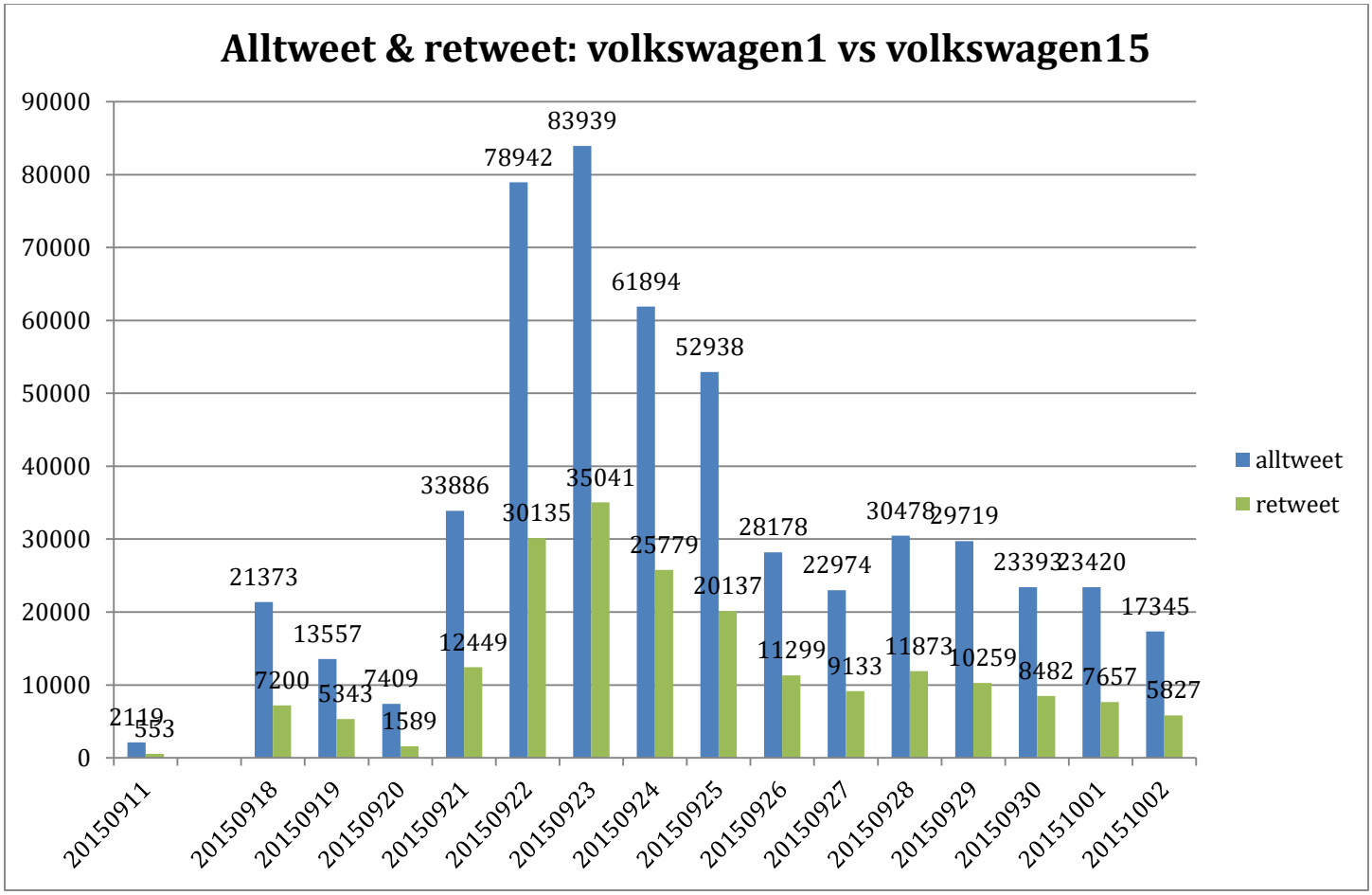

Chair quits
Volkswagen admits
Event outbreaks
alltweet: volkswagen
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
2119
21373
13557
7409
33886
78942
83939
61894
52938
28178
22974
30478
29719
23392
23420
17345
alltweet
20150911
20150918
20150919
20150920
20150921
20150922
20150923
20150924
20150925
20150926
20150927
20150928
20150929
20150930
20151001
20151002


retweets: volkswagen
40000
35000
30000
25000
20000
15000
10000
5000
0
553
7200
5343
1589
12449
30135
35041
25779
20137
11299
9133
11873
10259
8482
7657
5827
retweet
20150911
20150918
20150919
20150920
20150921
20150922
20150923
20150924
20150925
20150926
20150927
20150928
20150929
20150930
20151001
20151002





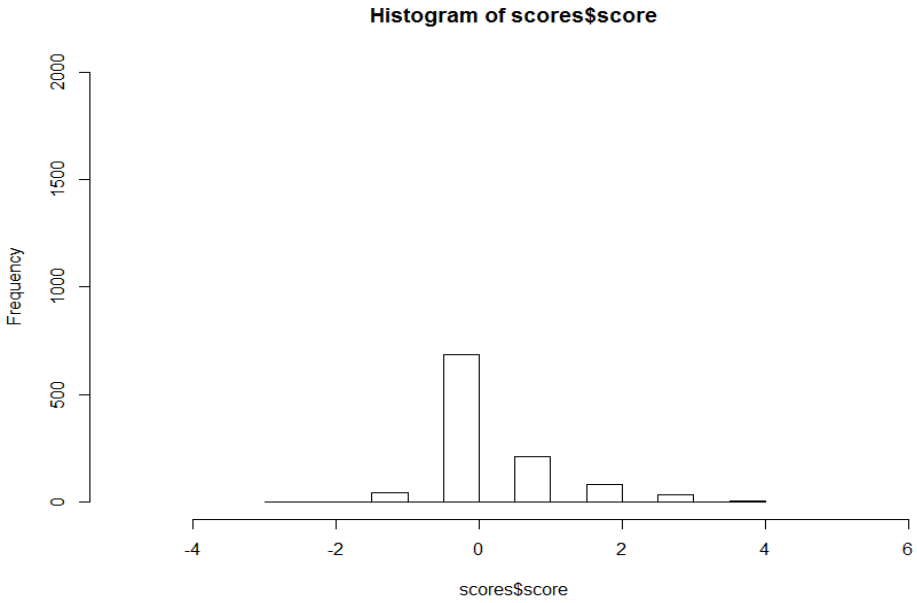
We can consider that Volkswagen hardly responses to the negative reactions on twitter.

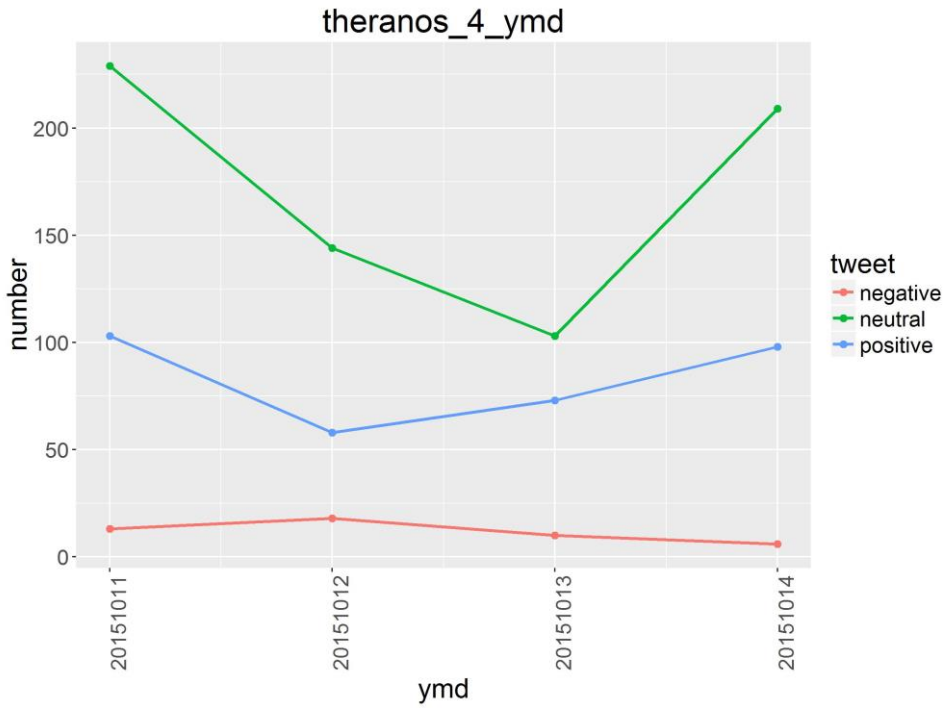

Alltweet & retweet: volkswagen1 vs volkswagen15
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
2119
553
21373
7200
13557
5343
7409
1589
33886
12449
78942
30135
83939
35041
61894
25779
52938
20137
28178
11299
22974
9133
30478
11873
29719
10259
23393
8482
23420
7657
17345
5827
20150911
20150918
20150919
20150920
20150921
20150922
20150923
20150924
20150925
20150926
20150927
20150928
20150929
20150930
20151001
20151002
alltweet
retweet

2. Sentiment Analysis

2.1 theranos_5: 11th – 14th October. control. Before the event

	min	1st Qu.	Median	Mean	3rd Qu.	Max.
	-3.0	0.0	0.0	0.4173	1.0	4.0

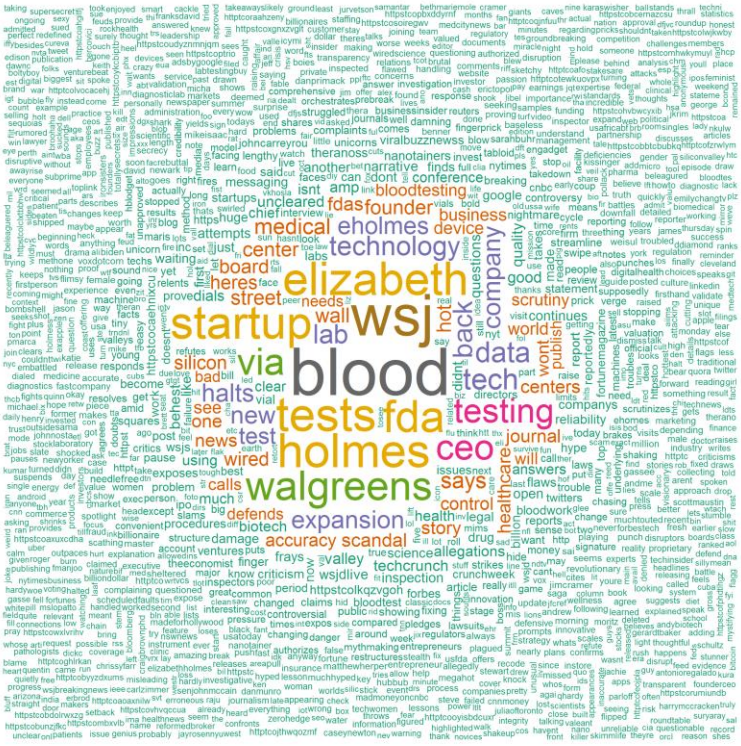




Mean is positive. Even thought there are high frequency of negative scores (-3), both mean and median >=0. According to the graph, tweets with negative sentiments are relatively lower than both neutral and positive sentiments before even outbreaks.

2.2 Theranos_15: 15th - 30th Oct 2015 : 16 days

(1) Wordcloud of thernaos_15: Removed word "Theranos"s

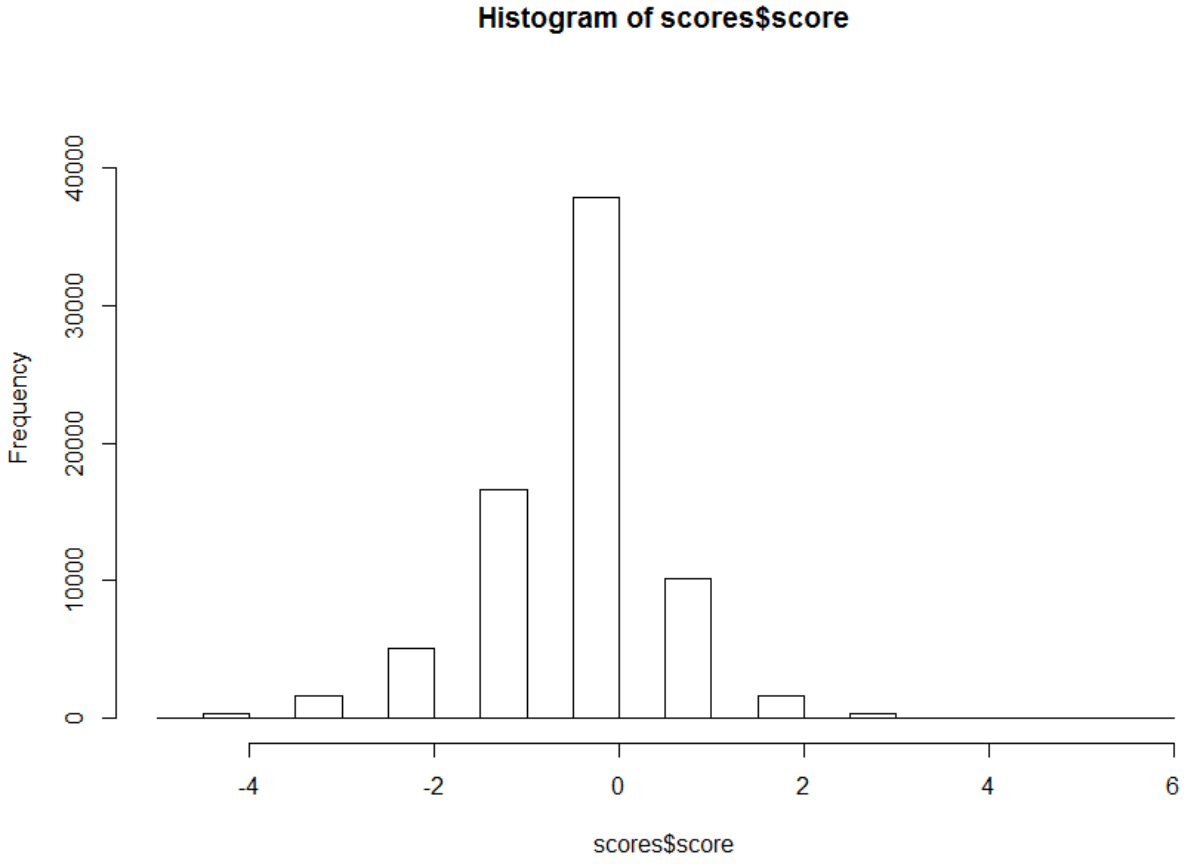


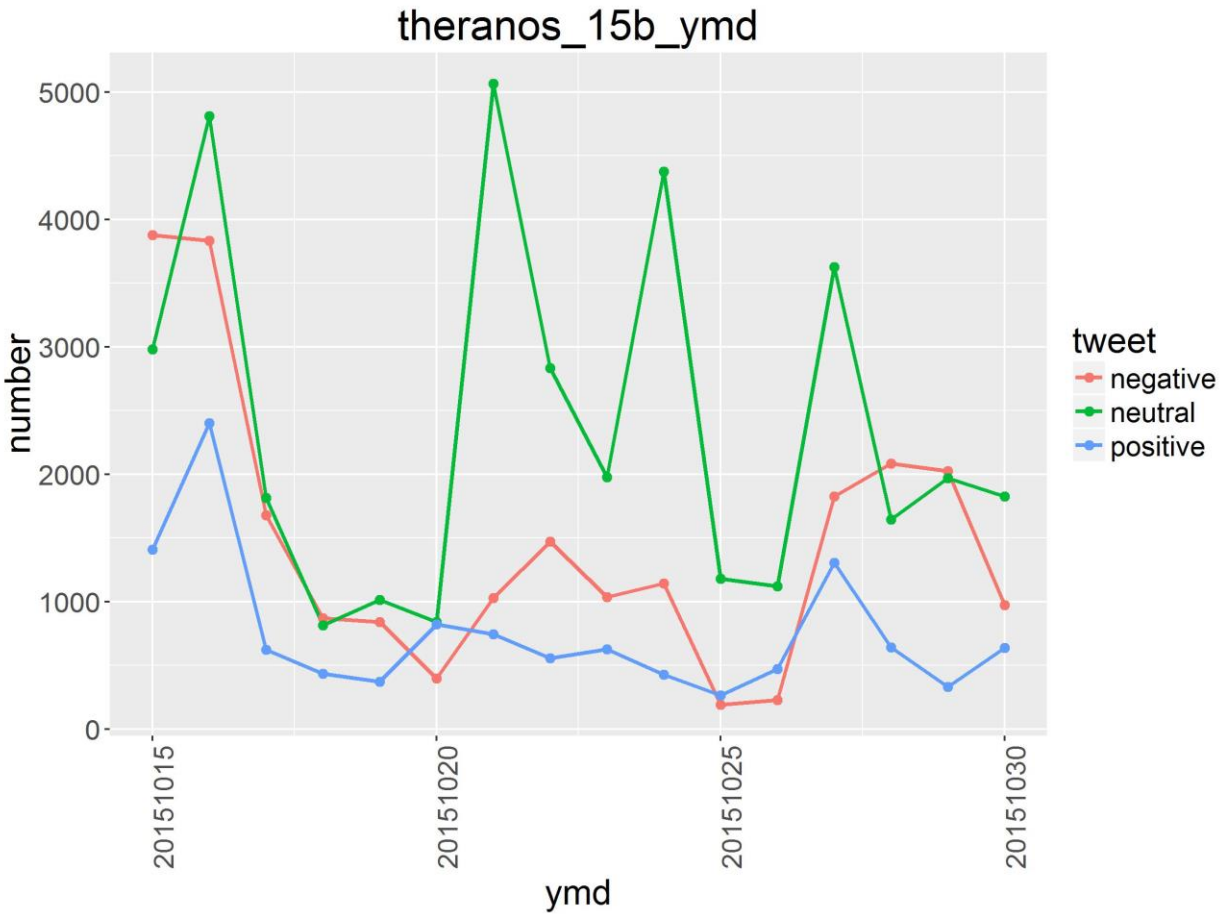
(2) Histogram of scores:

	min	1st Qu.	Median	Mean	3rd Qu.	Max.
	-5.0	-1.0	0.0	-0.2483	0.0	6.0

Histogram shows that sentiments are skewed to negative scores.
Also, according to the graph, frequency of tweets with negative sentiments is exclusively higher than neutral and positive sentiment tweets after the even outbreak.

**were there additional events after the 15th? Check the dates when negative sentiment tweets were higher than positive sentiment tweets. (2 points)


Histogram of scores$score
Frequency
0
10000
20000
30000
40000
-4
-2
0
2
4
6
scores$score

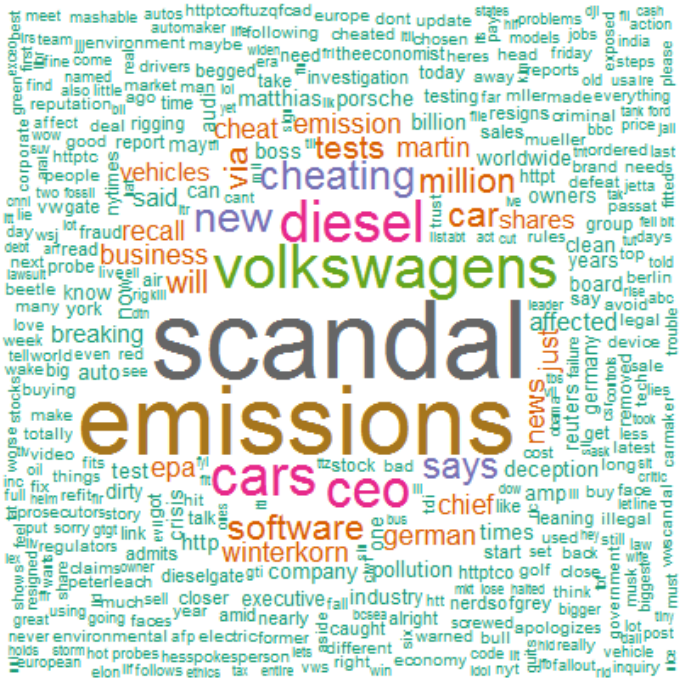

theranos_15b_ymd
number
0
1000
2000
3000
4000
5000
20151015
20151020
20151025
20151030
ymd
tweet
negative
neutral
positive

2.3 volkswagen_5

Data pending

2.4 Volkswagen_15

(1) Wordcloud of Volkswagen_15: Removed word “Volkswagen”

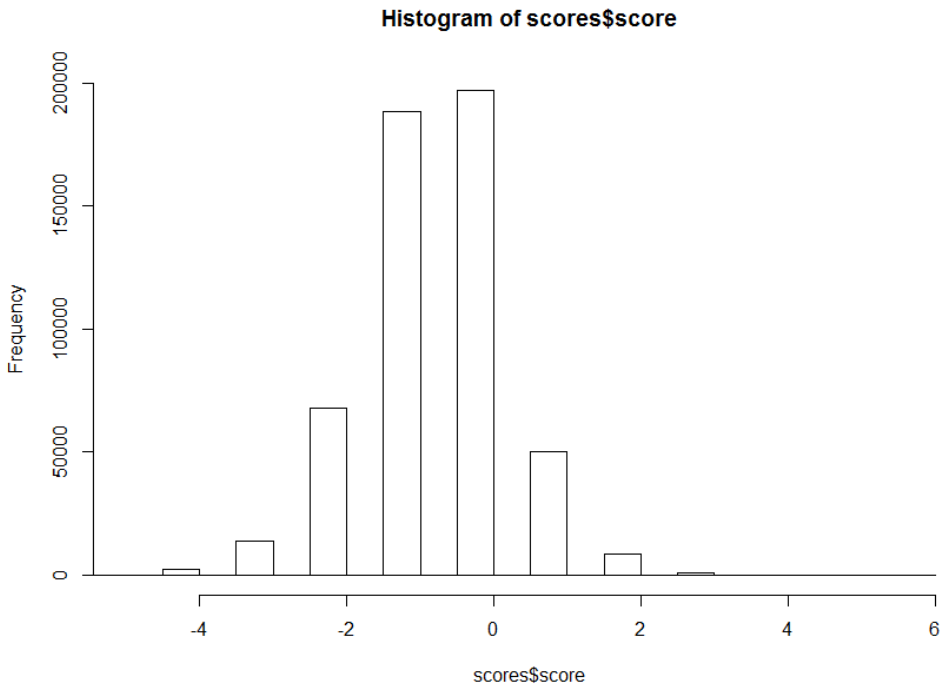


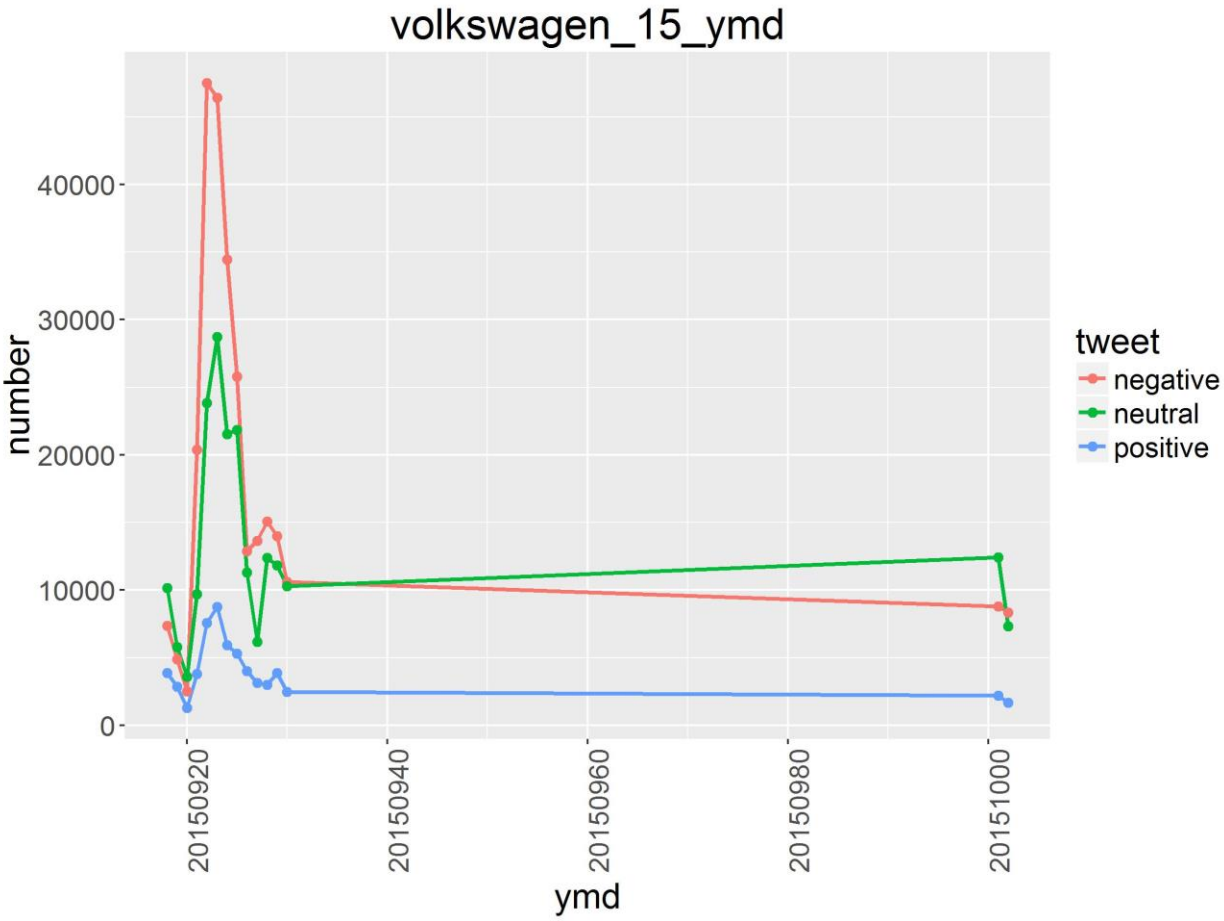
(2) Histogram of scores

skewed to negative scores.

```
#scores
```

	min	1st Qu.	Median	Mean	3rd Qu.	Max.
	-6.0	-1.0	-1.0	-0.5759	0.0	6.0





Negative sentiment surges after the chair has quit. I assume this is because Winterkorn quits but insists he is “not aware of any wrongdoing on [his] part

PLAN

1. Categorize firms that1) do response and 2) those that don't
   ➔ Game theory?

For (1), Company's responses:

- Among the companies that responses to the negative sentiments on twitter, find out what strategy is the best way to react. Come up with major strategies company takes.
- Example: Text, URL, media URL: **Use AWS**
- There might be consecutive effect from the first reaction to the negative sentiments if companies take several actions: *counterfactual* effect???

2. Machine Learning:
After causal evaluation, it might be better if we can show whether such model can predict the negative events beforehand.
i.e. Car industry: from the negative sentiments from small recall, we might be able to predict massive, major recall crisis

CONCERNS:

1. No significant data for companies' reaction.
i.e. no variety in a way companies responses

ISSUES TO THINK ABOUT:

1. Selecting companies:

(1) *Airline company*: sensitive to the twitter as sales highly depend on the firms' brand image
VS
(2) *Car industry*: Car companies tend to suffer from both minor and major recalls. As there are several negative events occurring within the company, we can examine stock prices in panel data. Also, we can control for the short and long term events as well as type or events.

(3) Some Criteria for the firms:

a. Companies that discloses **stock** exchange (check google finance: **public)**
b. Companies that are likely to be affected and uses **twitter/ SNS** a lot
c. Companies that have gone through **several events** i.e. recalls : Long term and short term event, small and large events

(4) Consideration: firms with good image vs firms with so-so image: impact of negative incident will be different-> need to specify what to control for this

2. Dependent variable:

1) Measure of *stock market performance* during the observation period: time it takes for the stock to become stable
   a. We can estimate how effective the certain strategy was by examining how quick the stock price stabilized

2) *Measure of google trend performance* during the observation period
    a. How much the keyword has been googled.?

3. Control Variable:
    - Variance between companies using twitter/SNS
    - Industry
    - Duration
    - Level of events: by estimating how much the stock falls we can control duration & size of events
    - Unobserved heterogeneity at the tweet level and the customer level:

Tweet level:	
Complaints within the last hour	Number of complaining tweets received by the airline during the last hour of receiving the current complaining tweet
Hashtag	Number of hashtags contained in the complaining tweet
@Order	The position of the airline twitter handle in the complaining tweet, relative to other username mentions if any
Customer level:	
Updates	Number of tweets ever posted by the user
Extrovert	Binary variable indicating whether the user's location, website or profile description (i.e., Twitter bio) is publicly available
Day of Week	Categorical variable indicating the day of week: fixed effect
Airline	Categorical variable indicating the company: fixed effect

4. Independent Variable:

Followers	Number of followers the user had, at the creation of the complaining tweet
Multilogue	Binary variable indicating whether the complaining tweet includes multiple
Retweet count	Number of retweet on the tweet
Response time	Time elapsed in seconds from the creation of the complaining tweet to the creation of the first reply tweet from the airline
Negative	Average fraction of negative words across all tweets about company in certain day
Positive	Average fraction of positive words across all tweets about company in certain day
Company Twitter $Account_{i,t}$	an indicator that equals one if company *i* has a company-managed Twitter account as of day *t* and zero otherwise
$\#CompanyTweets_{i,t}$	The number of tweets posted across all company-managed Twitter accounts on day *t*
Earnings $Announcement_{i,t}$	An indicator that equals one if company *i* announces earnings on day *t* and zero otherwise

Try regression
tweet ~ # of company tweet

23rd April, 2016: Accounting Scandals

- Twitter started form 2006, March
- Date range for accounting scandals: 2006 – 2016

Extract a sample of fraudulent firms from two sources:

1) SEC's (AAERs, from http://www.sec.gov/litifation)
   -Criteria uncertain, 50 companies that are likely to have fraud are selected and audited. Firms that went through accounting irregularities exist.
   -waiting for permission to access the data

2) Stanford Law School's Securities Class Action Clearinghouse (SCAC, http://securities.standford.edu)
   -Private security lawsuit. *Investors* select company with poor performance and suit it.
   -Twitter data + data from website "stocktwits".
   -Stocktwits: investors' tweets.

3) Stocktwits: can collect investors' messages about certain firm that goes through lawsuits. I am not sure whether they will include many negative sentiments, but this is well-known website and is used for sentiment analysis of investors in accounting research

Limitation: Companies' responses are in twitter not actual website.

A Communications Platform for the Investing Community

StockTwits® is a **financial communications platform for the financial and investing community**. The company was founded in **2008** by long-time investor Howard Lindzon. StockTwits created the **$TICKER(회사고유번호)** tag to enable and organize "streams" of information around stocks and markets across the web and social media. These streams provide new forms of insight, ideas and information that are used by investors, analysts, media and others as they research stocks and manage their investments.
Today, more than 300,000 investors, market professionals and public companies share information and ideas about the market and individual stocks using StockTwits, producing streams that are viewed by an audience of over 40 million across the financial web and social media platforms.

Share Ideas, news and information on StockTwits and across the web

StockTwits streams **consist of ideas, links, charts and other important financial data, summarized within 140 character messages**. Users, which include analysts, media and investors of all types, as well as the public companies themselves, contribute to the stream. Investors, and others interested in stocks and markets, can easily follow individual stocks, specific contributors, as well as view the StockTwits stream across dozens of financial sites that integrate the stream including Yahoo! Finance, CNNMoney, Reuters, TheStreet.com, Bing.com and The Globe and Mail.

The StockTwits platform **is also integrated with major social platforms, including Twitter**, Facebook and LinkedIn – making it easy to incorporate financial communications into your broader social interactions when, and only when, you choose to do so.

StockTwits content is focused solely on investing, and our technology and staff work to filter out unrelated messages and SPAM, ensuring the remaining content is only the most valuable and relevant discussions specifically about stocks and markets.

(4) Concerns for using accounting scandals: Companies that are listed in AAER have to publish. So companies cannot take different actions to the negative scandals.

Other ideas

1. Sports or entertainment industry, especially sports industry

When sports player are trying to change his/her agency, there tend to be many rumors which agency they will choose or how much guarantee will be.

Concerns:
- Negative sentiments on the agency from sports mania?
- Are there many events?

2. Companies around financial crisis
- Companies that AAER failed to audit beforehand
- select the companies that were not handled in AAER or IR etc. This is why such big scandal too place.