MA615 Final Project Report

Xiang Zhao

12/18/2017

Introduction

2017 Cyberthreats Defense Report have done a survey and data analysis on successful cyberattacks a company or an organization experienced in 15 countries. It shows that 80.6% respondents in America have compromised by at least one successful cyberattack in the past 12 months. The beast situation happened in Brazil and the worst situation happpened in China, 64.7% and 95.9%.

It's not hard to see the severity of cyberattack and importance of cybersecurity. Thus, I choose cybersecurity and cyberattack(cybercrime) as my topics to see the reactions and comments people have on social media Twitter. I used text analysis and mapping to see where and what kind of reactions and comments peopel have gloabally. And also I selected first two best protection softwares against cyberattacks, using sentiment analysis to see people's attitude towards these two softwares.

Data & Method

Data

I collected data from Twitter using 'searchTwitter' in R through online api and access token. I searched 5 key words in total, cybersecurity, cybercrime, cyberattack, bitdefender and kaspersky. For "cybersecurity", there is 10000 rows and has text of twittes, created time, retwitted count, longitude and latitude and so on, only 9 of then have geolocation. The rest four datasets have same data format with different sample size.

Method

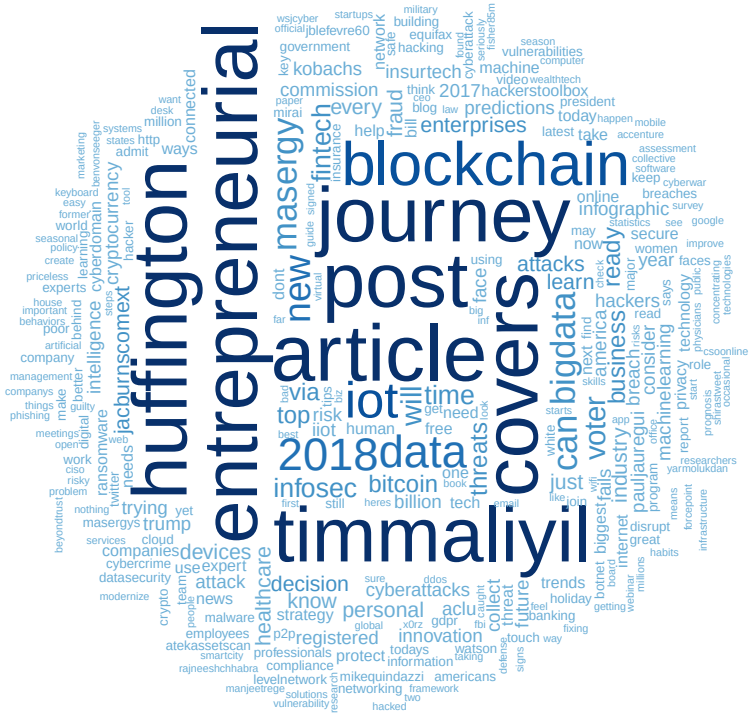
I used twitteR::searchTwitter to search for data, ggplot2::ggplot/ggplotly to draw map and other plots, wordcloud::wordcloud to do text analysis and draw word cloud and using sentiment analysis to analyze people' attitude toward protection softwares against cyberattack. And I use shiny app to show a interactive app of the result of my analysis.

Cybersecurit

Cybersecurity is the body of technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access. Cybersecurity is a hot topic especially at the data booming stage right now.

Word Cloud

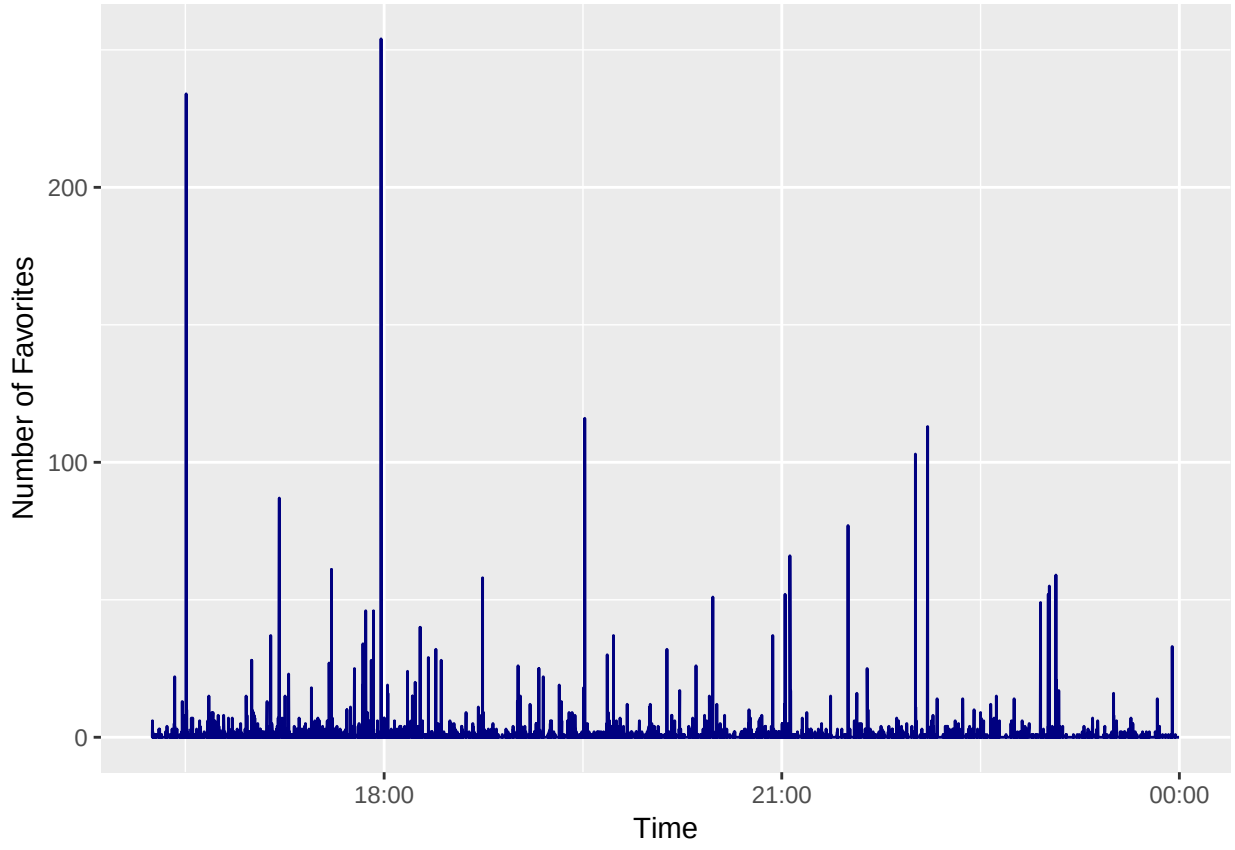
When talking about cybersecurity, we really want to see what are people's main reactions and comments on social media twitter? The word cloud below gives us answer.



In this word cloud, we can see several most frequent words like "entrepreneurial", "timmaliyil" and "blockchain" which are related to cybersecurity. They may reflect that cybersecurity is realy important for corporations and entrepreneurs because it's highly related to the network of companies or organizations, which may have an effect on trade secret which equals to profit. Tim Maliyil is a managed security services entrepreneur who Helps customers worldwide secure their data with encryption solutions. A blockchain, originally block chain, a continuously growing list of records, called blocks, which are linked and secured using cryptography is also a word relates to cybersecurity. Also, blockchain is highly correlated with bitcoin and fintech which have urgent needs of cybersecurity.

Timeline

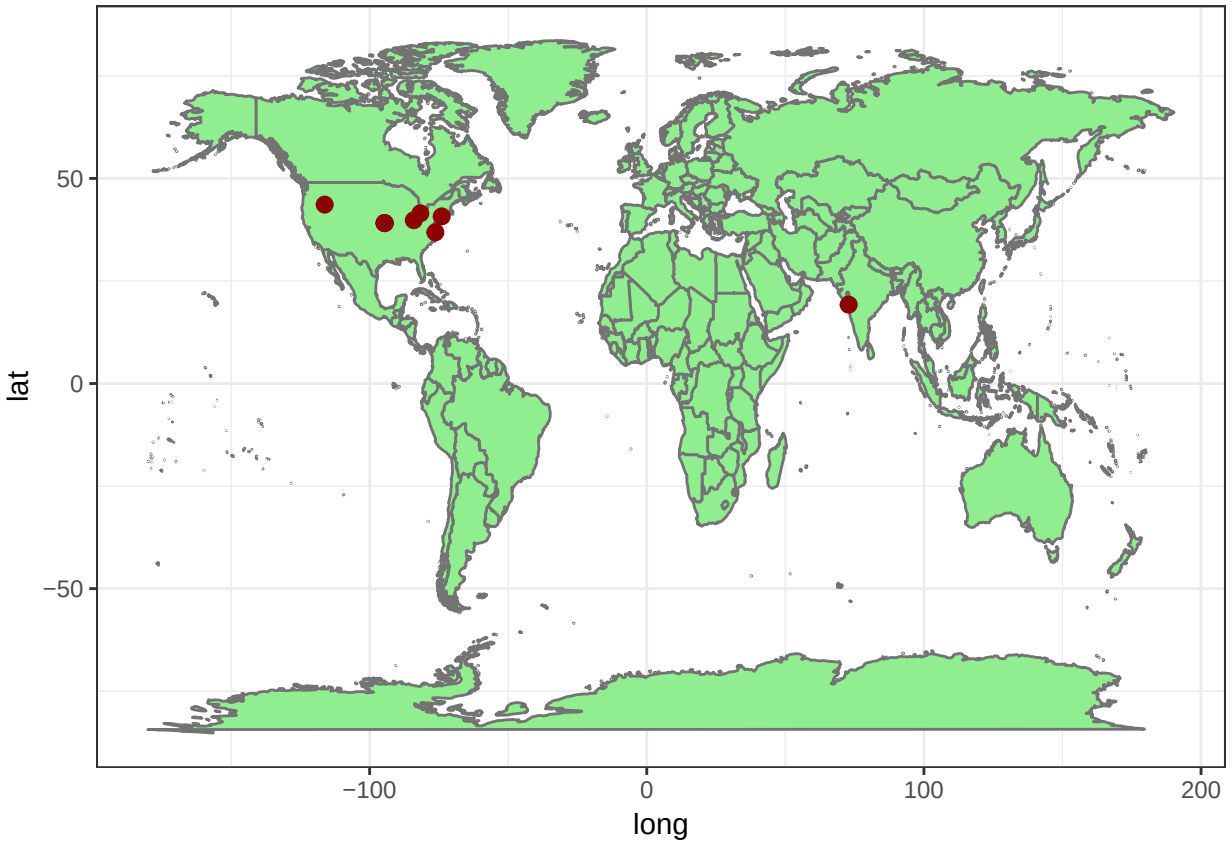
Also I'm interested in the count of favorites for the key word "cybersecurity", in order to see the number of favorites, I created a timeline to see when and the trend of number of favorites



This timeline is actually a 8-hour timeline which shows the count of favorites about cybersecurity, and the least unit of time is 1 second. The highest number of favorites is about more than 250 at around 18:00 December 13rd, 2017.

Map

Moreover, the geolocated distribution of tweets is also another interesting thing to explore since it reflects where has the biggest reaction to the topic.

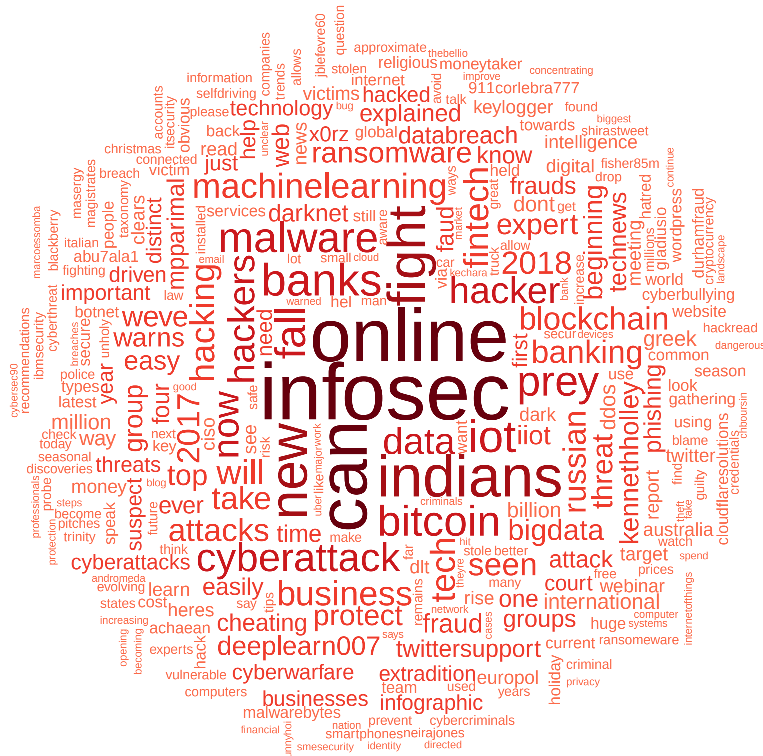


Through the map of geolocations of tweets on cybersecurity, we can see that almost all the spatial points drop in the USA,unifomally distributed from west to middle to east America. While there is one point dropping on India. Since there is only 9 geolocations, so it's too less to draw any conclusion through map.

Cyberthreat

The original definition of cyberthreat is the possibility of a malicious attempt to damage or disrupt a computer network or system, which is a bad thing for cybersecurity or the thing is fighted against. I searched "cyberattack" and "cybercrime" to see people's reactions toward these two negative word.

Word Cloud of Cybercrime



The word cloud of cybercrime shows several frequent words, "infosec", "bitcoin", "banking","hacking","frauds","big data" and "machine learning". We can see that cybercrime relates more to business and finance like banking system, bitcoin and may seek for information security because of the occurence of potential problem like frauding, and also big data and machine learning maybe signals for further developments or solutions for cybercrimes.

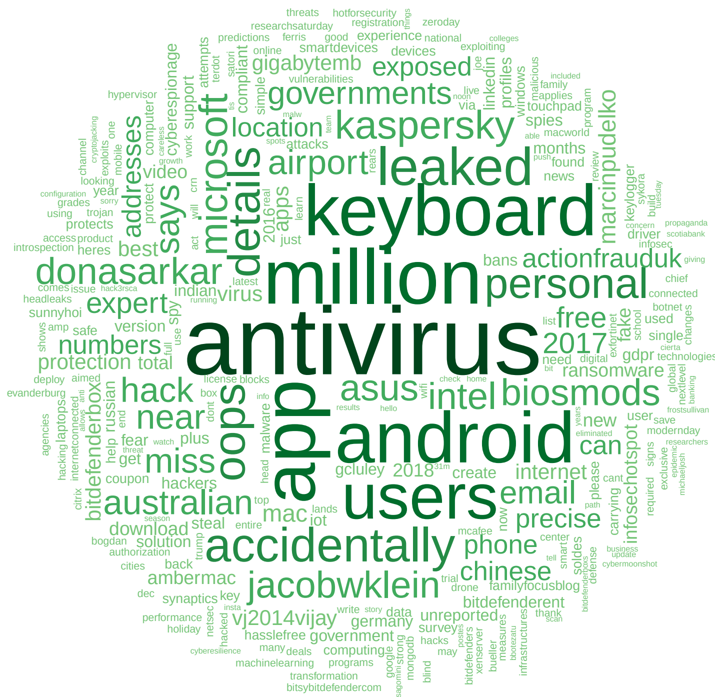
Word Cloud of Cyberattack

The word cloud of cyberattack is also something about "data", "hackers", "business" and "infosec" which indicates the potential risks hidden in business and the needs for information security.

Protection Software

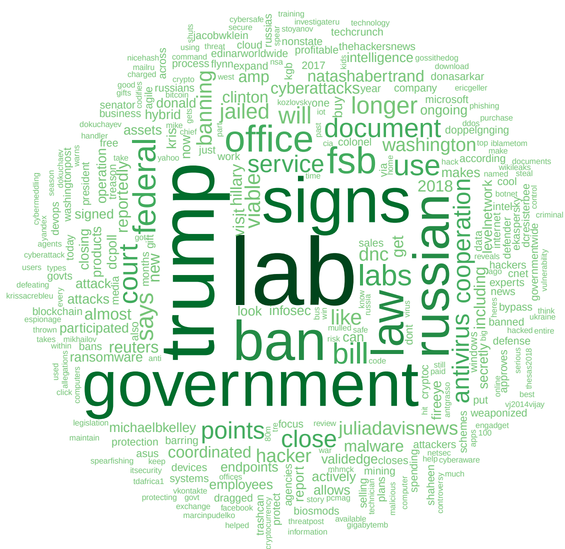
Pretection software can protect the computers and network of individuals or organizations. The two best protection softwares ranked in 2017 are "Bitdefender" and "Kaspersky". I'm eager to see people's attitude toward these two brands.

Word Cloud of Bitdefender



The word cloud of "Bitdefender" shows several main key words like "antivirus", "android", "microsoft" and "million". It may give a clue about bitdefender is an antivirus software having more than million users of android and microsoft. The word cloud nearly have no negative words, which may show that bitdefender has good effects and reputations.

Word Cloud of Kaspersky

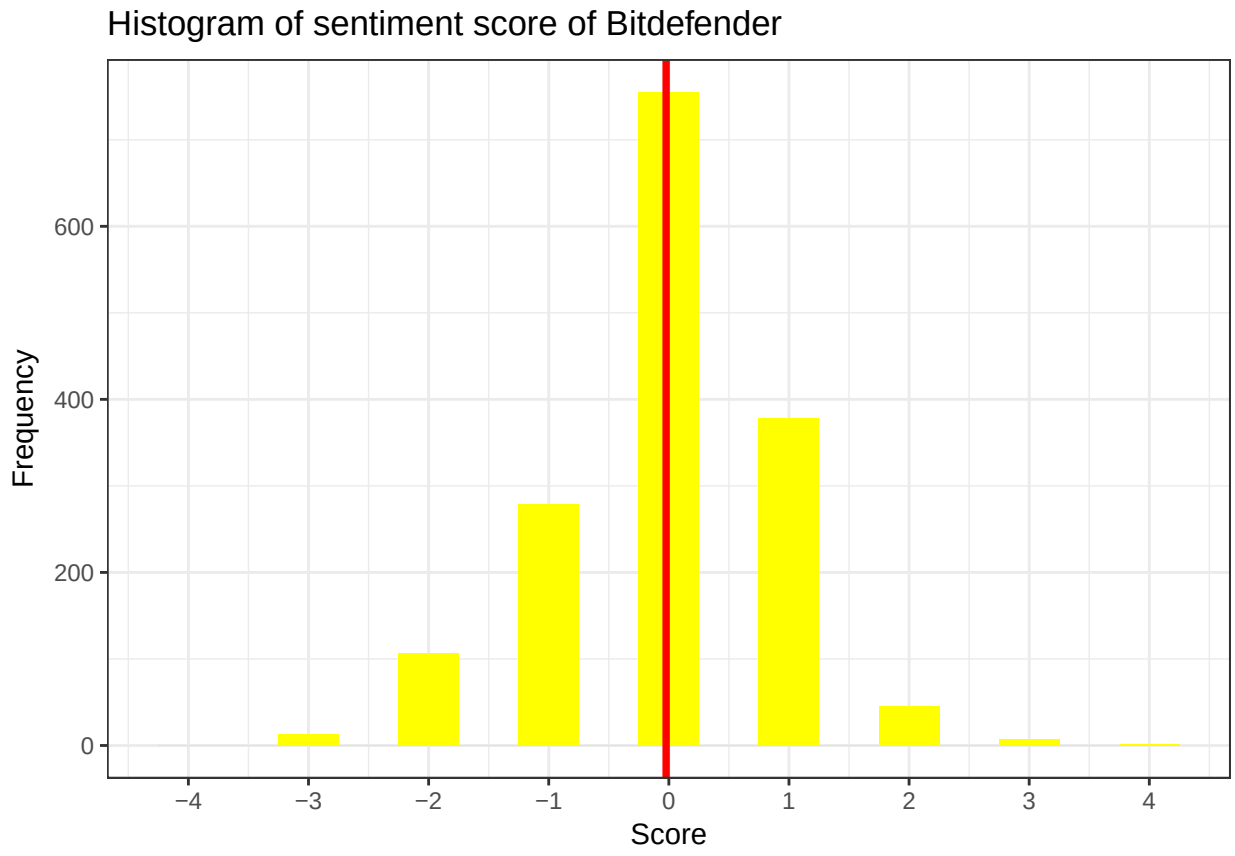


The word cloud of “Kaspersky” seems have no such detailed words for characteristics but something like “federal” and “government”, which may lead us guess that kaspersky has a better population and reputation among government departments.

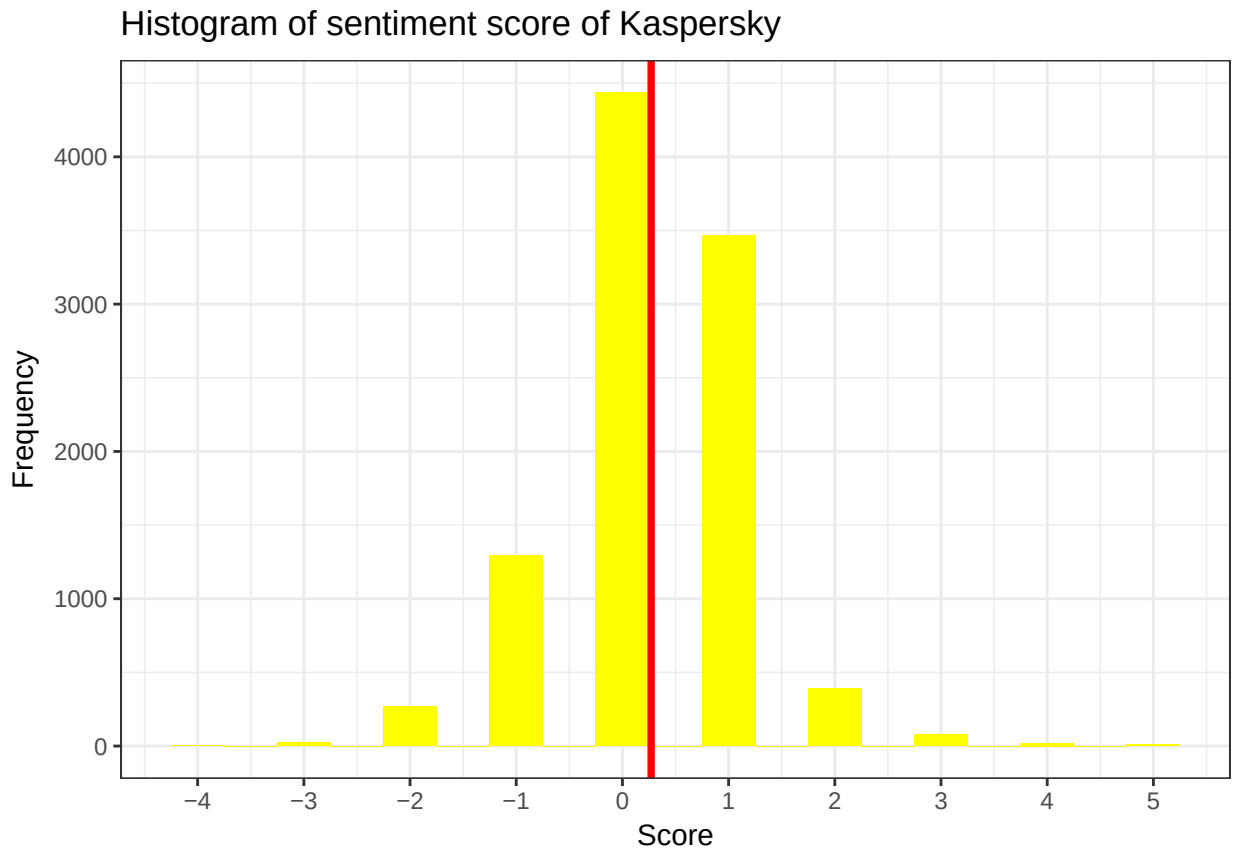
Sentiment Analysis

Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc., is positive, negative, or neutral. By doing sentiment analysis, we can figure out why the brand is having a positive or negative sentiment and how we can change it.

I broke down each sentence to single word of every tweet and compare the words to the positive/negative words sets to see every tweet's attitude for each brand. The bar plots shows the histogram of score of each brand's sentiment. The higher the score is, more positive is the attitude.



From histogram of sentiment score of Bitdefender, we see that the range is from -4 to 4, and a vast majority of sentences have 0 score which means most of them are neutral. The red line indicates the mean of score which is -0.22, so it's knid of lightly negative attitude of people towards Bitdefender.



From histogram of sentiment score of Kaspersky, we see that the range is from -4 to 5 which means a wider positive range than Bitdefender, and a vast majority of sentences have 0 or 1 score. The red line indicates the mean of score which is 0.27, so it's knid of lightly positive attitude of people towards Kaspersky. Thus, comparing Bitdefender and Kaspersky from tweets, we can conclude that Kaspersky has a relatively positive reputation than Bitdefender.