

# Mark Riley – Visualization Project

DS 745 – Fall 2019

## Introduction

The purpose of these visualizations is to determine if there is a relationship between tweets done by Elon Musk, CEO of Tesla, on the following day's Tesla stock (TSLA) trade volume between 2013 and 2016.

Elon Musk is one of the most controversial Twitter users among CEOs of publicly traded companies. In August of 2018, Musk tweeted his intent to take Tesla private at a substantial premium per share, and that funding had been secured.<sup>1</sup> This tweet landed him in crosshairs of the Securities and Exchange Commission (SEC), which charged Musk with fraud. Musk and the SEC reached a settlement in September 2018, requiring Elon to apologize, pay a \$40M penalty, step down as Tesla's Chairman, and for Tesla to appoint additional independent directors to the company.<sup>2</sup>

According to marketrealist.com, "Volume and pattern are combined to identify the trend direction. Some traders use this combination for entry and exit signals."<sup>3</sup> This information could be useful to investors if it is possible to show a correlation (positive or negative) between Elon's tweeting and Tesla stock activity.

## Data Source

The data for these visualizations came from two Kaggle datasets.

- <https://www.kaggle.com/kulgen/elon-musks-tweets>
- <https://www.kaggle.com/rpaguirre/tesla-stock-price>

I combined the datasets together by looking at the stock metrics for the day (open, high, low, close, etc.) against the count of tweet types (retweet, reply, original, total) that Elon did the day before, if any. Only tweets from the day prior to trading were included in this analysis.

In addition to being CEO of Tesla, Elon Musk also owns several other private companies including SpaceX, Hyperloop, OpenAI, The Boring Company, and several others. All of Elon's tweets were included in this analysis even if they were not specifically about Tesla because he is such a high-profile leader whose identity is closely tied to Tesla.

## Data Manipulation

I added a Category field to distinguish between three types of tweets. A tweet was deemed to be a Retweet if the "Retweet from" column had a value in it. Otherwise if the first character of the tweet text was the symbol "@" then the tweet was categorized as a Reply. If a tweet was neither a Retweet nor Reply, then it was considered an original Tweet. The Total Tweets is calculated by adding the Retweets, Replies, and Tweets together.

---

<sup>1</sup> <https://twitter.com/elonmusk/status/1026872652290379776?lang=en>

<sup>2</sup> <https://www.sec.gov/news/press-release/2018-226>

<sup>3</sup> <https://articles2.marketrealist.com/2014/12/interpret-volume-technical-analysis/>

# Mark Riley – Visualization Project

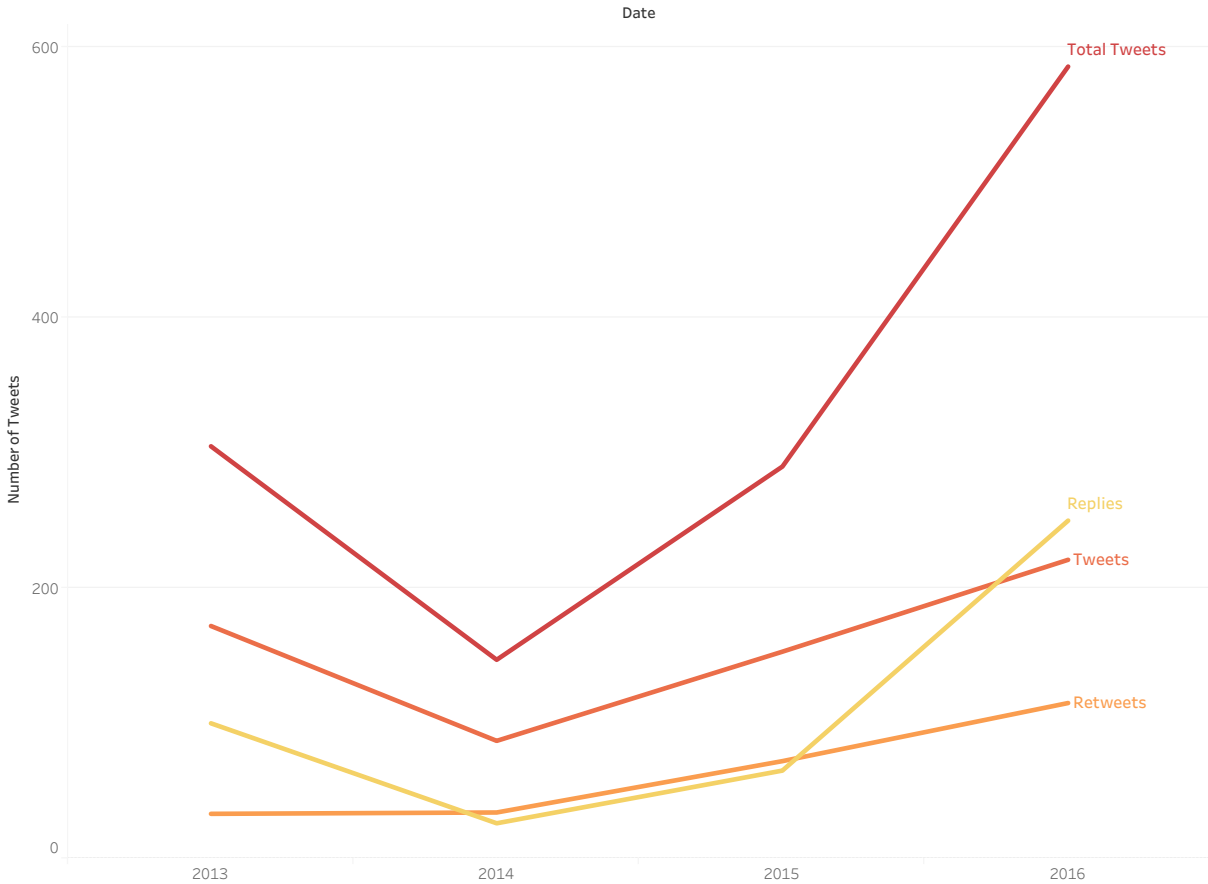
DS 745 – Fall 2019

The range of dates between the datasets did not overlap. Elon’s tweet data ranged been mid-November 2012 to the end of September 2017. The Tesla stock data ranged between the end of June 2010 to mid-March 2017. To ensure the comparisons would be equal across time periods, I deleted all data prior to 2013 and after 2016. This leaves only full-year comparisons on this dataset.

## Elon Musk’s Twitter Use

This graphic is not officially part of the visualization process. I want to give an overview of how Elon tweets before performing my analysis. Between the beginning of 2014 and the end of 2016, Elon more than tripled his Twitter volume and has his engagement with other Twitter users, as evidenced by the increasing number of Replies.

Elon Tweeting Habits  
How Often Elon Tweets by Type Over Time

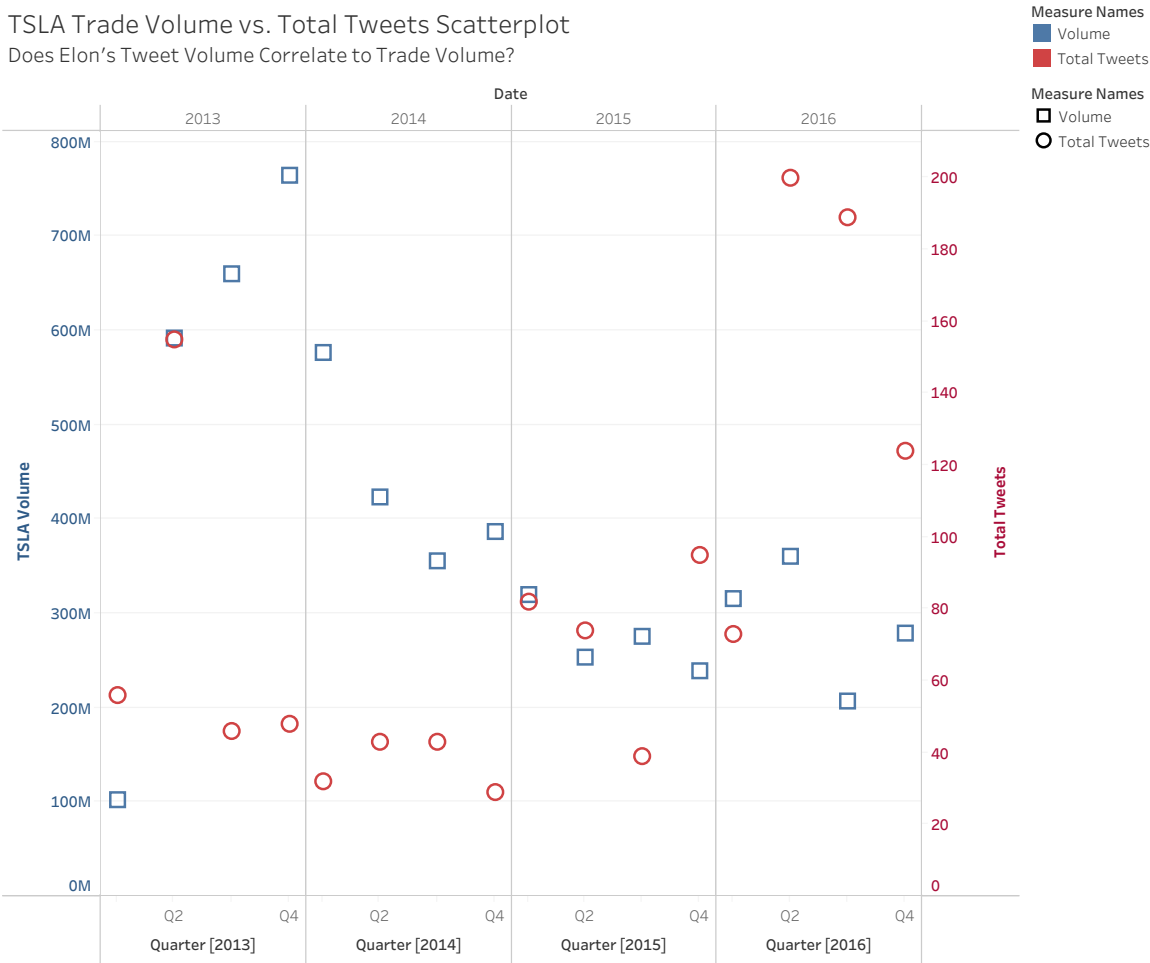


# Mark Riley – Visualization Project

DS 745 – Fall 2019

## Visualization 1

TSLA Trade Volume vs. Total Tweets Scatterplot  
Does Elon’s Tweet Volume Correlate to Trade Volume?



### Pros

- Uses a combination of color and shape to distinguish between trade volume and tweet volume data points (Gestalt Law of Similarity).
- The colors on the axis match up with colors of the points on the graph for additional clarity.
- Minimal chart junk and good data-ink ratio.

### Cons

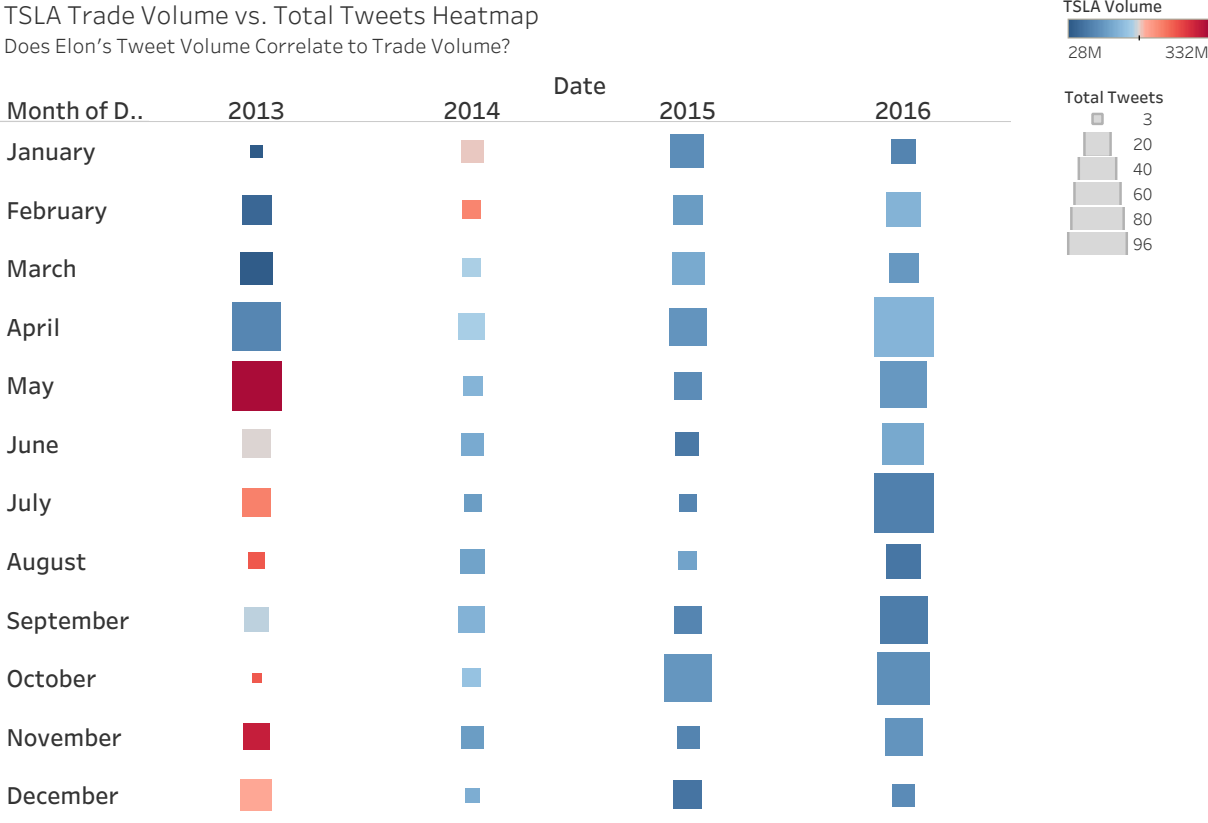
- With different scales between the stock volume (left) and tweet volume (right) it can be easy to incorrectly think that if the points overlap then there is correlation. For example, in Q1 2015, the points overlap but trade volume decreased over the prior month, but Elon’s tweet volume increased.
- A scatter plot is not a good format for this type of data because you need to understand the trends over time as well as the values in each period.
- Difficult to scan horizontally for pre-attentive processing.
- Unable to combine the legends for color and shape in Tableau, which could be confusing.

# Mark Riley – Visualization Project

DS 745 – Fall 2019

## Visualization 2

TSLA Trade Volume vs. Total Tweets Heatmap  
Does Elon’s Tweet Volume Correlate to Trade Volume?



### Pros

- Unique way of looking at the relationship of two variables over time rather than the standard line graphs or bar graphs.
- I like being able to compare trends across each month across the four years and then the year from top to bottom – grouping by time horizontally and vertically.
- After some time processing the picture you can get a sense that there does not appear to be correlation between TSLA share volume and Elon’s tweet volume.
- Good data-ink ratio.
- Good spacing between the columns (Gestalt Law of Proximity) to tie the data together in each year.

### Cons

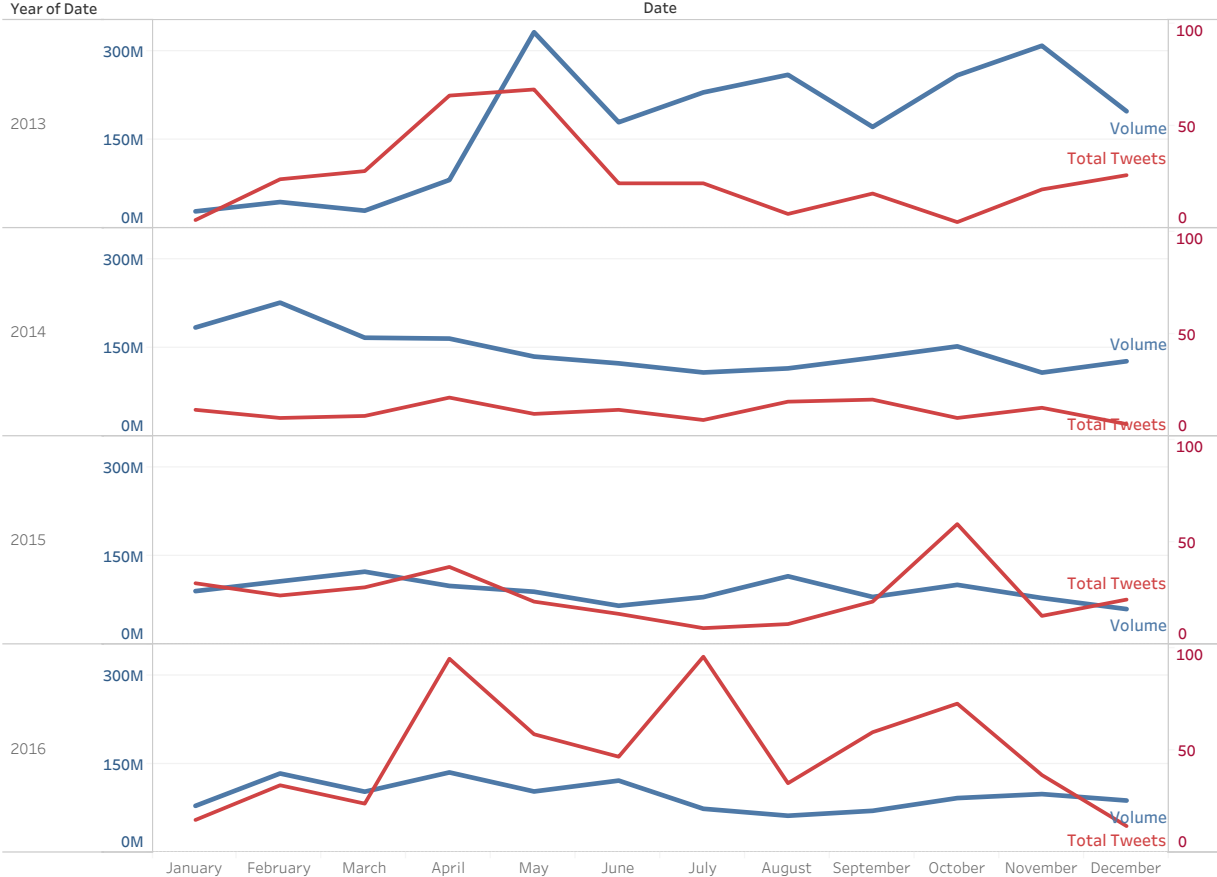
- It is not as intuitive as I would like. It takes several references back and forth between the legend and the graphic to understand that color is the volume (blue lower, red higher) and the size is the total tweets for that month. Since the data types are similar (volumes of tweets and trades) it would be easy to mix up the size vs. color association.
- Need to scan vertically rather than horizontally to see the trend over each year.

# Mark Riley – Visualization Project

DS 745 – Fall 2019

## Visualization 3

TSLA Trade Volume vs. Total Tweets Line Graph  
Does Elon’s Tweet Volume Correlate to Trade Volume?



### Pros

- Clearly shows the trends over time for both trade and tweet volume.
- Easy to do a comparison of trends between the two variables to assess the correlation (or lack thereof).
- Lower chance of mistaking overlap of the lines for correlation.
- Axis labels, lines, and line labels colors correspond to their variables for clarity.
- Easily read left to right.
- Minimal chart junk.
- Good data-ink ratio.

### Cons

- Difficult to compare each month across years, e.g. January 2013, 2014, 2015, 2016.
- Horizontal lines for trade volume do not line up with tweet volume labels.