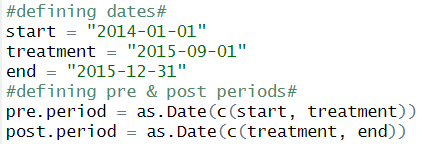
*The Great Diesel Dupe Scandal – Using Data Analysis Techniques to Determine the Causal Impact of Volkswagen’s Nefarious Behaviors*

In this case study, I will review the now infamous *Diesel Dupe* scandal. In September 2015, the Environmental Protection Agency (EPA) found that many Volkswagen vehicles being sold in America essentially had a defeat device, which would detect when they were being tested. When tested, the Volkswagen vehicles would change the emissions performance of the vehicle to improve results. Eventually, the EPA published their findings, and the German car giant has since admitted to cheating emissions test in the United States.

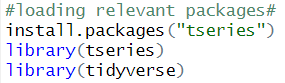
For some background information, Volkswagen had a major push to sell diesel cars in the United States in the mid-2010s. In fact, Volkswagen labored a huge marketing campaign, trumpeting its cars’ low emissions. However, the EPA ultimately found out about the so-called defeat device implemented in their vehicles through a study of over 482,000 cars in the United States. As aforementioned, Volkswagen later admitted that about 11,000,000 cars worldwide were produced with this defeat device.

Naturally, a scandal of this magnitude had significant impact in Volkswagen stock price and reputation. But, to what extent? In this case study, I will answer that question by analyzing the scandal’s impact through historical time-series data of Volkswagen stock prices from the beginning of 2014 to the end of 2015. To explore this relationship, I will be using R.

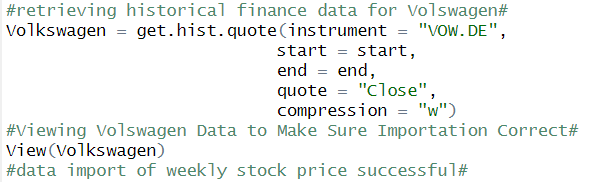
The first thing I want to do is define the dates I would like to study. To do so, I will define a start date, end date, and, most important, a treatment date. In this instance, the treatment date is defined as September 2015 as this is when the scandal was first discovered and being published.



Next, I want to shift my focus by loading the relevant packages I will need to perform this analysis in R.

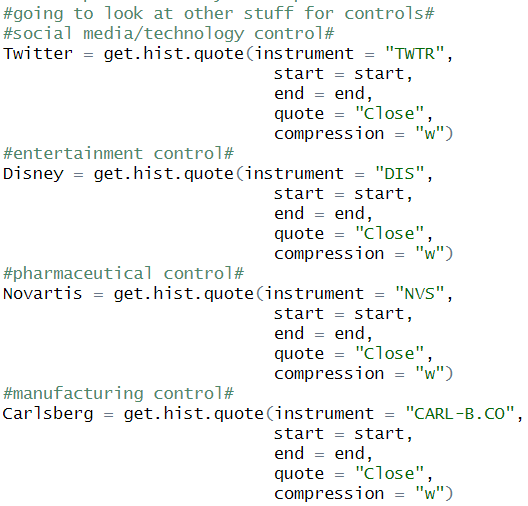


From there, I want to shift my focus to loading & retrieving historical stock price information directly related to Volkswagen.

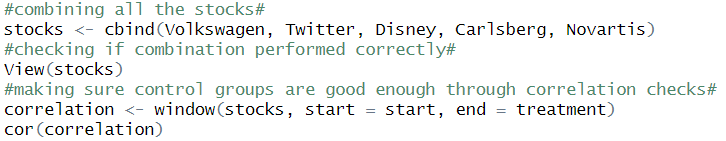


Now, this data is not useful to me by itself. My ultimate goal is to determine the causal impact of this scandal. To help me determine this, I want to include control groups. Stock markets are complex interconnected systems where various local factors can cause global changes in the behavior of the entire market. In particular, I want these control groups to be correlated with the stock price of Volkswagen to represent contagion. Contagion in financial markets essentially implies the propagation of impact between different components of the market.

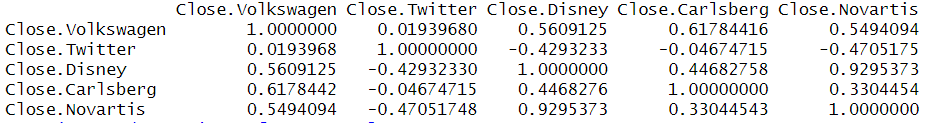
Ultimately, I decide on using four control groups, consisting of social media/technology, entertainment, pharmaceutical, and food & beverage industries respectively. I determine that I will use Twitter, Disney, Novartis, and Carlsberg as representative companies.



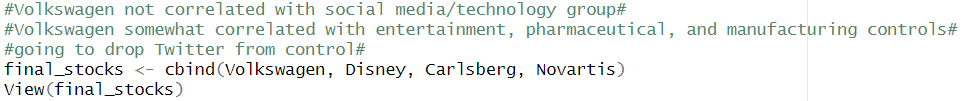
Now, I want to combine all the stocks into a single data frame. From there, I want to check if this combination was performed correctly and run correlation checks to make sure contagion is represented in my model to determine causal impact.



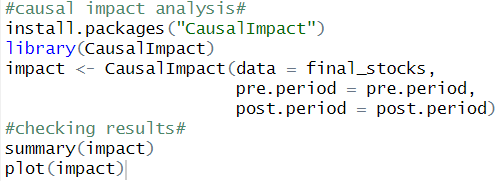
Upon running the code displayed above, I can see the correlation in the stock price changes over time with respect to Volkswagen, Twitter, Disney, Carlsberg, and Novartis.



From the table above, I can clearly see that I captured that aspect of contagion with respect to Disney, Carlsberg, and Novartis. In fact, it is clear that the market price is somewhat or moderately correlated with Volkswagen, being above a value of .5. However, Twitter has a very low correlation value at ~.02. Moving forward, I decide to drop Twitter from my model. To do so, I create a new data frame titled final\_stocks that notably does not include Twitter.

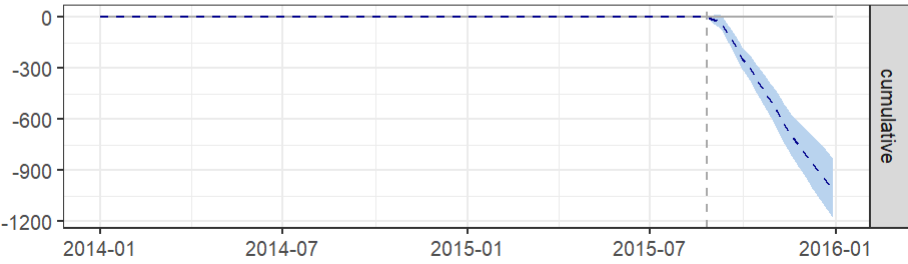


Now that I have done so, I can finally prepare for my causal impact analysis. First, I install the CausalImpact package as I have not worked with this package in the past, and I load it into my library in R. I will then create a new data frame titled impact and check results from there.



From the results, there are a few things I want to highlight in particular – the absolute effect & a visualization that will help me communicate my results through the plot function.

First, I want to highlight the visualization I produced as I think that tells the *real* story.

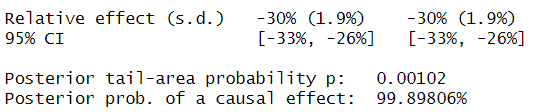


Here, I want you to notice a few things. Firstly, you can clearly identify the vertical dotted line. The vertical dotted line represents when the EPA first published their findings and when the scandal first started picking up media attraction. Notice how this line directly lines up to what I referenced at the beginning of this case study – September 2015. This line represents the origin of the now infamous *Diesel Dupe* scandal.

Second, notice how the horizontal dotted line remains at zero until this point. This dotted line essentially represents the stock price fluctuation in Volkswagen over time up until the treatment takes place. In this case, the treatment is scandal I have been trying to explore throughout this case study. The dotted line remains at zero up until the scandal took place, and then the line clearly displays a rapid negative decline.

To put it simply, the visualization tells the story of scandal’s impact with respect to Volkswagen stock price. The scandal’s effect on the market capitalization price of Volkswagen was practically immediate. As you would expect, a scandal of this magnitude contributed to a rapid negative impact of Volkswagen stock price. And, the stock price & Volkswagen as a whole seemingly never recovered from the great *Diesel Dupe* of our times.

Likewise, I would like to bring your attention to the summary statistics of this impact data frame. Throughout this case study, we have focused on stock price, which is a relative key performance indicator. The magnitude of this impact can be defined in the relative effect summary statistics.



The table above tells us a number of things. One, we are almost certain that the causal impact on Volkswagen’s stock price is from the publication and popularization of this scandal. And, secondly, the -30% can be interpreted as Volkswagen’s stock price losing 30% of its value over the course of four months *after* this scandal came to light.

For more background info: <https://www.bbc.com/news/business-34324772>