Data Analysis of Meteorite Impact Locations

Matthew W. Younce

PSDS 4900 Capstone
Program of Study in Data Science

University of Missouri-Columbia
Data Science and Analytics

August 12, 2021

Abstract

This study hoped to determine if a comprehensive dataset of meteorite landings could be used to make predictions about the location of future events. Additionally, the study allowed the author to demonstrate his skills in coding both R and Python, data analysis, and data visualization.

The goals of the study included five key points. First, do meteorites fall at random? Second, are meteorite impacts consistent over history, or are there any changes? Third, does human population play any role in the locating of meteorites? Fourth, can trends be identified between meteorites that are seen falling and those that were found later? Fifth, can predictions about future events be made by studying this data?

Methods of analysis included the development of code, use of Jupyter Notebook and Github for processing and storing the results, the creation of a web app using R Shiny for visualization, and various techniques applied from the skills learned in the Program of Study in Data Science Program (PSDS), a collaboration between the University of Missouri-Columbia and the National Geospatial-Intelligence Agency.

While no clear predictions were made as a result of the study, the author was able to hone his skills in coding, analysis and visualization and brought some interesting results out of this dataset.

Introduction

I am the lead Geospatial Analyst in the Time-Dominant Operations Center (TDOC) at NGA, where the focus is on first-phase analysis. The TDOC's mission is to provide timely warnings to our military service members and national decision-makers by monitoring, analyzing and reporting imminent threats around the globe. My intent was to study a topic that could benefit my work in the TDOC. However, in this time of COVID and teleworking, it would not be wise to research a classified dataset, especially if I might not be able to go into a secure environment to perform the analysis.

Looking for unclassified data, I found a dataset of meteorite events. This included every meteorite "landing," a euphemism for impact events, recorded in history. I felt it was a suitable substitute, since it is a listing of events all over the globe that are classified and broken down into various categories, much like other natural disasters or man-made events of interest.

Additionally, this study served to demonstrate my skills in coding in R and Python, data analysis, and data visualization.

Data Discussion and Methodology

Meteorites are objects from outer space that survived their trip through the atmosphere to come to rest on the Earth's surface. I located a dataset from the Meteoritical Society and hosted by NASA[1]. This dataset contained information on all known meteorite landings on Earth, over 34,000 individual landings. The dataset includes basic information on the meteorite and the location it was found.

After a review of the data, and thinking about how my analytic methods could benefit my work in the TDOC, I planned to answer five questions in this study:

1. Do meteorites fall at random or are there patterns to their landing locations?
2. Are the numbers of meteorite falls consistent over history or are they changing?
3. Are there more falls recorded because more people are able to see them and are there more finds because more people are in the area?
4. Are there trends between meteorites that have been observed falling to Earth and meteorites that are found later?
5. Can predictions be made about future meteorite landings by studying these past events?

[1] https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh

I first wanted to scrub my data, and began reviewing the dataset. I created a GIT repository to host my data and analysis.[2] I found some discrepancies, such as over 6,000 meteorites at a coordinate of (000000N 0000000E) and almost 5,000 meteorites with the geolocation of (713000S 354012E), a location on the ice shield in Antarctica. I reasoned that perhaps the coordinates were the location of a lab where the scientists recorded their findings. I knew that meteoriticists have found it easy to locate meteorites in Antarctica because, "it's the only black thing when everything else is white as far as the eye can see."[3] However, that many meteorites in the same place seems unlikely. Entering those coordinates in Google Maps returned a photo of pancakes, as shown in Figure 1, so the mystery may never be solved.

FIGURE 1. A STACK OF PANCAKES AND 4,700 METEORITES IN ANTARCTICA. A RIDDLE, WRAPPED IN A MYSTERY, INSIDE AN ENIGMA, COVERED IN MAPLE SYRUP.

Other entries were missing some attributes, including geolocation, mass, or recorded class of meteorite. After completion of the scrubbing, I was left with 9,781 valid meteorites for my analysis.

After I reviewed my goals, I determined additional data would be needed to complete this research. First, I obtained was a meteorite classification guide, which is depicted below in Figure 2. This system, developed by Krot et. al. was immensely helpful, as the initial dataset only recorded the specific group, with a numeric subgroup. For instance, it described a meteorite as having a "recclass" of H5-6. There were a total of 311 different groups recorded.

[2] https://github.com/mwy912/capstone/

[3] Dr. Bruce D. Dod, Professor *emeritus* of Physics and Earth Sciences, Mercer University. Personal Communication, 1999.

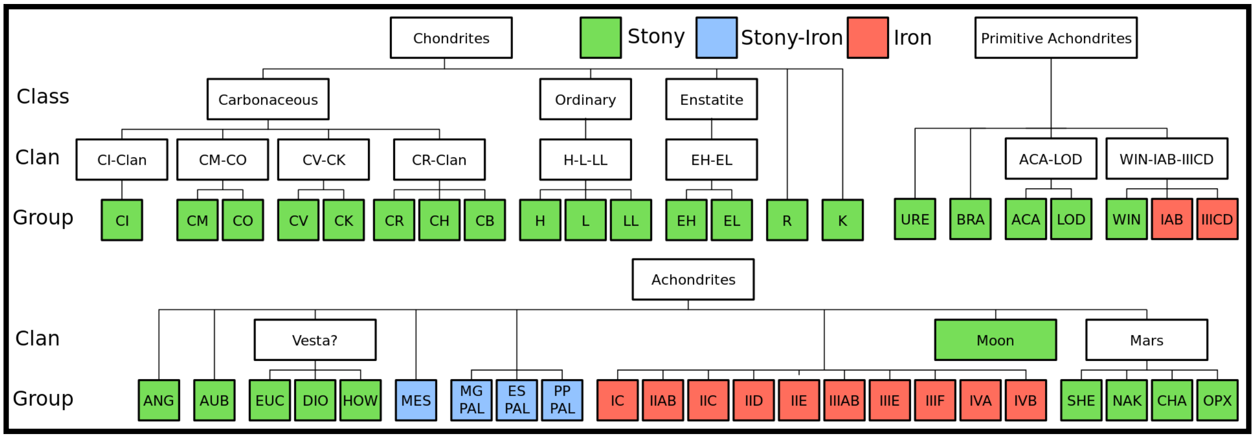

FIGURE 2. KROT SCHEME OF METEORITE CLASSIFICATION[4]

This classification first broke down all meteorites into chondrites, primitive achondrites, and achondrites. Chondrites are meteorites that contain chondrules, spherical grains of minerals that form very slowly in extremely cold temperatures, indicating they were formed early in the solar system. Achondrites do not contain chondrules. These three categories are further divided into classes, clans, and groups, all of which may relate to their origin (e.g Lunar or Martian meteorites), or their chemical composition. Additionally, meteorites can generally be identified as one of three types — iron, stony, or stony-iron. I recreated this classification chart as a data frame and joined it to the meteorite data. With Krot's classification guide, I was able to more thoroughly understand the hierarchy, and consolidate the data into 45 more manageable groups, instead of extremely specific ones. For instance, the H5-6 could also be described as all of the following: a stony meteorite; a chondrite; an ordinary class; an H-L-LL clan; and an H group, in addition to the H5-6.

The next dataset I acquired was a historical urban population dataset[5]. This contained locations of urban centers and their populations, from 3700 BC to AD 2000. I intended to use this to determine if a given meteorite was in an urbanized area at the time of their discovery. While the location given may today be a city, it may not have been 200 years ago. My process

[4] Michael K. Weisberg; Timothy J. McCoy; Alexander N. Krot (2006). "Systematics and Evaluation of Meteorite Classification" (PDF). In Lauretta, Dante S.; McSween, Jr., Harold Y. (eds.). Meteorites and the early solar system II. Foreword by Richard P. Binzel. Tucson: University of Arizona Press. pp. 19–52. ISBN 978-0816525621. Recreated at https://en.wikipedia.org/wiki/Meteorite_classification.

[5] Reba, M. L., F. Reitsma, and K. C. Seto. 2018. Historical Urban Population: 3700 BC - AD 2000. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4ZG6QBX.

was to buffer each meteorite location with a radius of 0.25°. Next, I located any urban areas in the buffer, and determined if they were present in the years immediately prior to the discovery of the meteorite. After lengthy processing, my analysis found that less than 250 of the meteorites were near an urban area at the time of their discovery.

Value	Description	Value	Description
0	Water Bodies	11	Temperate deciduous forest
1	Cultivated Land	12	Warm mixed forest
2	Pasture/Land used for grazing	13	Grassland/Steppe
5	Ice	14	Hot desert
6	Tundra	15	Scrubland
7	Wooded Tundra	16	Savanna
8	Boreal Forest	17	Tropical woodland
9	Cool conifer forest	18	Tropical forest
10	Temperate mixed forest	19	No data

TABLE 1. LAND USE/LAND COVER RASTER VALUES.[6]

The final dataset I acquired was historical land use and land cover rasters. These files were at a spatial resolution of 0.5° and temporal resolution of every 50 years from 1700 to 1950, as well as 1970 and 1990. The land use categories are shown in table 1. This information was used to process each location to determine what kind of land was present where the meteorite was found, when it was found. As the dataset began with 1700, I was able to assign a land cover to all but 16 early meteorites.

[6] Goldewijk, K.K. 2010. ISLSCP II Historical Land Cover and Land Use, 1700-1990. In Hall, Forest G., G. Collatz, B. Meeson, S. Los, E. Brown de Colstoun, and D. Landis (eds.). ISLSCP Initiative II Collection. Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/967

Results

I created a web app in Shiny for R to visualize the data and my analysis. Figure 3 shows the main screen of that app. The app is much more detailed and customizable than the results I will include here, so it is a complement to this research paper.

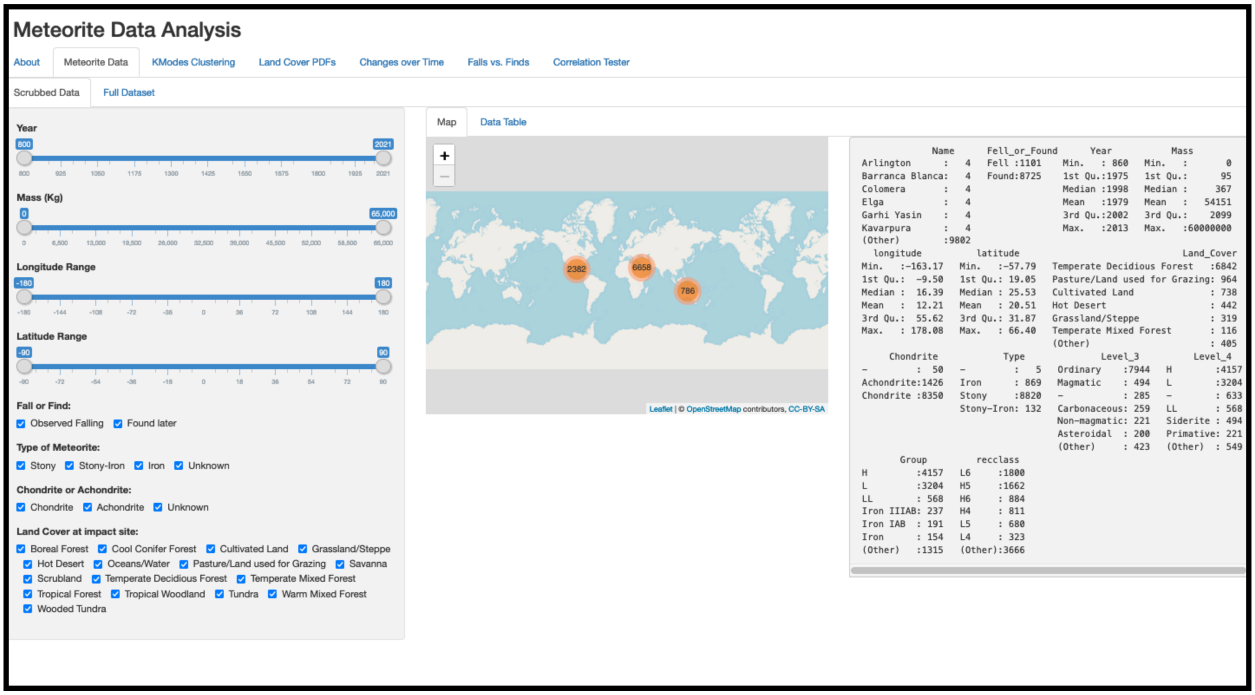

FIGURE 3. R SHINY WEB APP FOR METEORITE DATA ANALYSIS[7]

The first objective was to determine if meteorites landing locations were truly random or if there was any pattern to the locations. One technique to determine this was through an unsupervised machine learning method called *k*-modes. This is similar to a *k*-means method, except that it is run on categorical data. The *k*-modes technique determined that the data was best divided into seven clusters. Table 4 is a summary of those 7 clusters.

[7] https://shiny.sgn.missouri.edu/students/ymw.0780/Capstone/meteoriteanalysis/

Cluster	number	Mass		Likely dominant factor for cluster
		Mean	Median	
A	5023	5.50×10^3 g	2.38×10^2 g	98.9% Stony meteorites 93.9% Finds
B	1023	4.67×10^5 g	1.26×10^4 g	Median mass an order of 2 above others. 90.8% Iron meteorites (But 100% of the Iron meteorites in the sample)
C	238	1.49×10^3 g	2.14×10^2 g	100% Temp. deciduous forest 100% Finds 100% from 2005.
D	784	1.75×10^4 g	2.00×10^3 g	99.5% Stony meteorites 87.7% Falls
E	526	1.82×10^4 g	9.51×10^2 g	99.4% Finds
F	2009	2.85×10^3 g	2.40×10^2 g	99.9% Stony Meteorites 99.9% Finds
G	252	3.16×10^3 g	3.17×10^2 g	96.8% Hot Desert 99.2% Stony meteorites

TABLE 2. SEVEN CLUSTERS AS DEFINED BY UNSUPERVISED *K*-MODES CLUSTERING.

I have listed the dominant factor noted in each cluster. By far, the largest number of meteorites were in cluster A, with over half of all the meteorites. Each cluster had at least one category that was very prevalent, listed in the table. Even though these categories were prevalent, this doesn't preclude other clusters from having meteorites with the same categories. For instance, 99.9% of cluster F was stony meteorites, but the meteorites in cluster F represent only 23% of all the stony meteorites in the data set. So, for future predictions, another stony meteorite would not necessarily fit into cluster F.

The "most dangerous" meteorites in this list are cluster B, with a median mass 100 times any other cluster. Interestingly, this was a case of one cluster holding all the meteorites of a specific type. All iron meteorites in the dataset fell into Cluster B. Looking at Cluster B geospatially, however, does not show any particular pattern for location, as demonstrated in Figure 4.

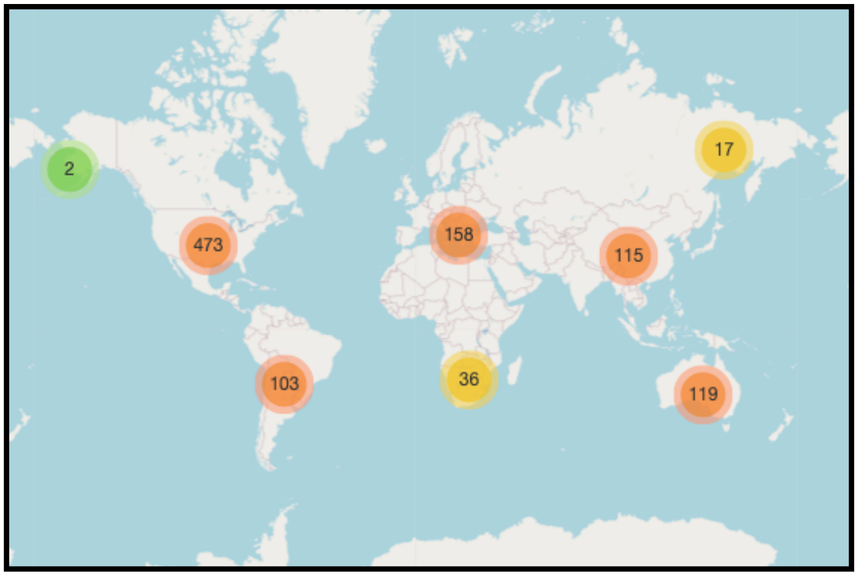

FIGURE 4. LOCATION OF METEORITES FOUND IN CLUSTER B.

Additionally, I performed a Probability Density Function, which showed the likelihood that a meteorite of a given mass was found in a given land cover type. The results, shown in Figure 5, demonstrate that wooded Tundra, Tropical Woodland and Tropical Forest have the lowest masses of any land cover. These peaks are quite dramatic compared to other land covers.

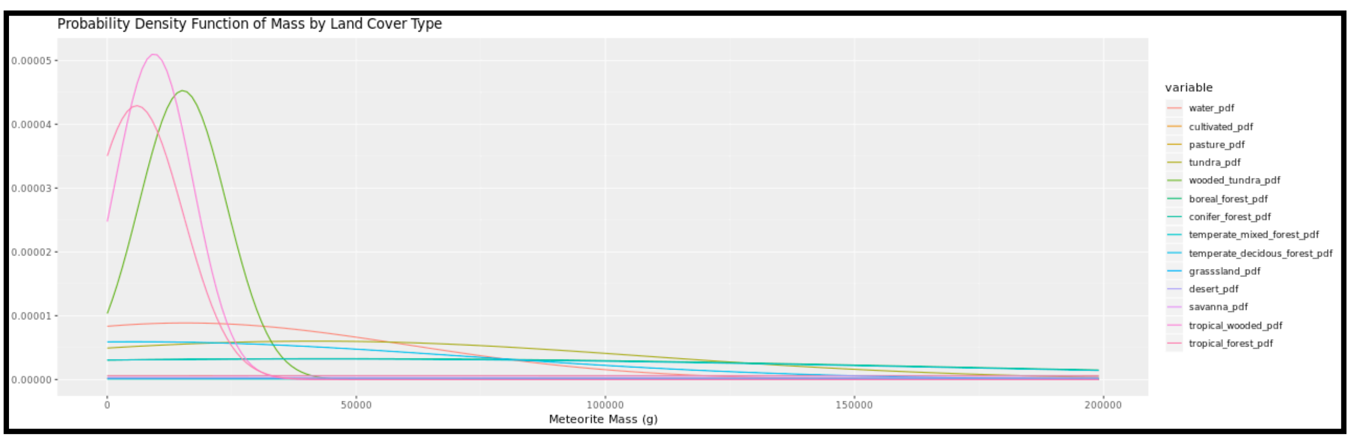

FIGURE 5. PROBABILITY DENSITY FUNCTION OF MASS BY LAND COVER TYPE.

The second and third goals can be combined together for this discussion, as they both considered changes in numbers of meteorites over time, and if there is any connection to the Earth's population. I found that the numbers of falls recorded is increasing, with a dramatic rise since the 1700s. However, this may be a spurious correlation, as there is likely much better record keeping in the recent past than the distant past. Figure 6 shows this increase.

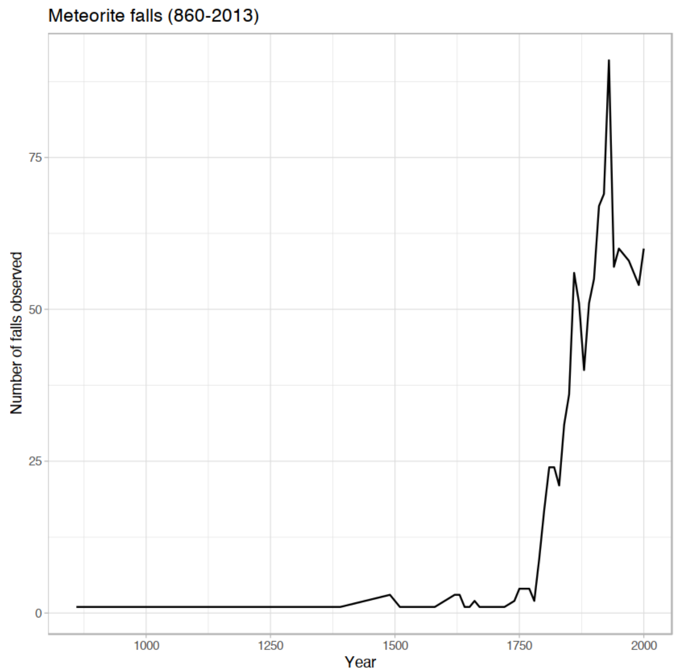

FIGURE 6. METEORITE FALLS OVER TIME.

Goal four was to determine if a relationship could be found between where meteorites are found and where they fall. This analysis was done by performing a floor function on latitude and longitude, which drops all decimals and leaves the whole integer. I then determined if a fall, a find, or both were recorded for each 1° x 1° area. There were 1063 grid squares where only finds were recorded, 712 where only falls were recorded, and181 with both. There did not seem to be a trend.

Goal five was to determine if past events can help predict future events with meteorite landings or impacts. I found that *if* it is possible, it would requiremuch more data than I obtained. While some trends were noted above, there was no “predicting” performed in my analysis. At least not yet.

Additional Research

Moving forward, it is clear that much more data would be needed before any sort of predictions could be made. Bringing this back to my work at NGA: A list of past terrorist events would suggest that there could be a future terrorist attack in the Middle East. However, It would not be possible from that list to determine an exact time and location. Additional intelligence could make a difference.

One avenue of future research could be to look at how latitude of meteorite landings changes over the course of the year. The Earth is at a 23.5° axial tilt relative to the plane of its orbit. This means that in December, the North Pole points away from the Sun, while, six months later, the North Pole points towards the Sun. This is diagrammed in Figure 7.

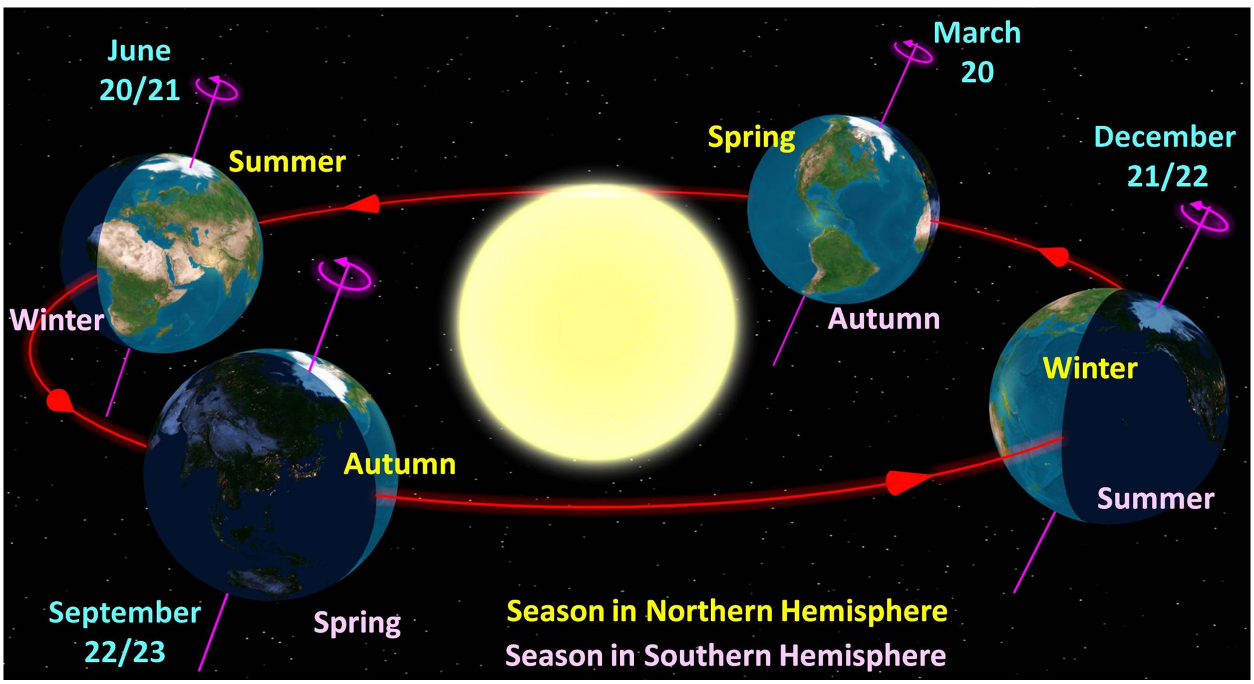

FIGURE 7. EARTH'S AXIAL TILT THROUGH THE YEAR.[8]

Effectively, the target that is the Earth is leaning towards or away from any potential impact. To pursue this analysis, I would need much more precise data, however, with times down to the second, if not at least the day, rather than only the year as was found in this dataset. Perhaps

[8] "The Earth at the start of the four seasons", at https://commons.wikimedia.org/wiki/File:North_season.jpg

NASA's data of bolides and fireballs (extremely bright meteors) recorded on government sensors[9], a study could be performed to compare the location to the day of the year.

Conclusion

This study did not have any dramatic findings. It did not give a prediction for when a single meteorite will impact the Earth. However, there were still many benefits from this study.

Through this research, I received a much better understanding of the data science tools than I did in any of my classes in the PSDS program. When it was my job to figure out what method would solve the questions I had, I learned how these techniques are applied. That was perhaps the most important outcome of this study. Learning not only how to code, analyze and visualize, but also which tools were right for which jobs will be immensely useful going forward. The classes have rekindled my excitement for education, and I plan to follow up this degree with a Ph.D. in the future.

It is appropriate that this paper is submitted on 12 August, the date that marks the peak of the Perseid Meteor Shower, when the Earth passes through debris left by Comet Swift-Tuttle. So, who knows? Perhaps the next entry in the NASA database will be written tonight.

[9] Fireballs and Bolides, https://cneos.jpl.nasa.gov/fireballs/