Willamette University

Final Project:
Meteorite Landings Data Analysis

Anujan Tennathur, Matteo Alemay, Lane Mikkelsen
Dr. Kristen Gore
Data Science 151
December 14, 2023

Intro/Background:

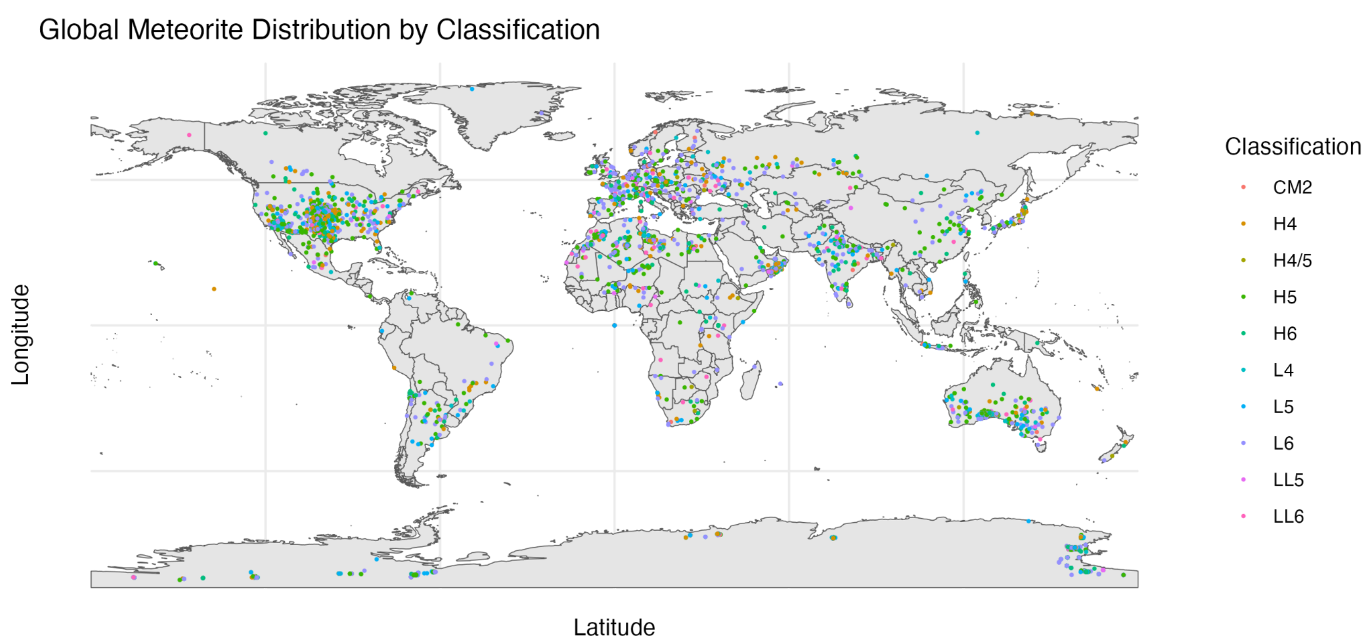
For our exploratory data analysis, our group sought an interesting and complex dataset from which we could learn about the natural world. We found just that on Data.gov: an extensive dataset with information about every meteorite known to have fallen to Earth. “Meteorite Landings” was collected by Javier de la Torre of the Meteoritical Society and published by NASA Public Data. Though we first encountered this dataset on Data.gov, it is also available on Kaggle, which offers some helpful suggestions for analysis that will be discussed later. We were drawn to this dataset because it stood apart from many others we looked at. The dataset is enormous, contains several variables that caught our attention, and pertains to a topic that inspires awe in children and scientists alike. Meteorites are fragments of interplanetary material that enter Earth’s Atmosphere, heating up due to friction and creating glowing trails. Most are far older than Earth's rocks; “Some meteorites even contain tiny particles that formed around other stars that existed before our Sun,” (“What Is a Meteorite?”). Scientists use meteorites to investigate the history of our solar system and how they have impacted history on Earth. In analyzing “Meteorite Landings” we got to be involved, even for a small amount of time, in this process.

Perhaps the most compelling reason for selecting this dataset was the opportunity for exploratory data analysis. As soon as we looked over the variables, we saw great potential in the Geolocation variable to create a map depicting all meteorite landings on Earth. We wondered what the distribution would look like. Would meteorites be evenly dispersed on Earth's surface or clustered in certain areas? We were also intrigued by the Recclass, or classification, variable. Though we know little about how scientists classify meteorites, we wondered if classification had any relationship with a meteorite’s mass and with the frequency at which it fell to Earth. Do different classifications of meteorites fall at different frequencies? In addition, we wanted to analyze if the total landings of meteorites by year (regardless of classification) changed over time. We had some difficulty with our linear regression because we didn’t expect there to be causal or correlational relationships between the variables. Nonetheless, we used linear regression to investigate the relationship between certain classifications of meteorites’ mass and the years they fell or were found. These are the most compelling of the guiding questions we identified from the “Meteorite Landings” dataset that inspired the data analysis and visualizations presented in the upcoming sections of this paper.

Data and Methodology:

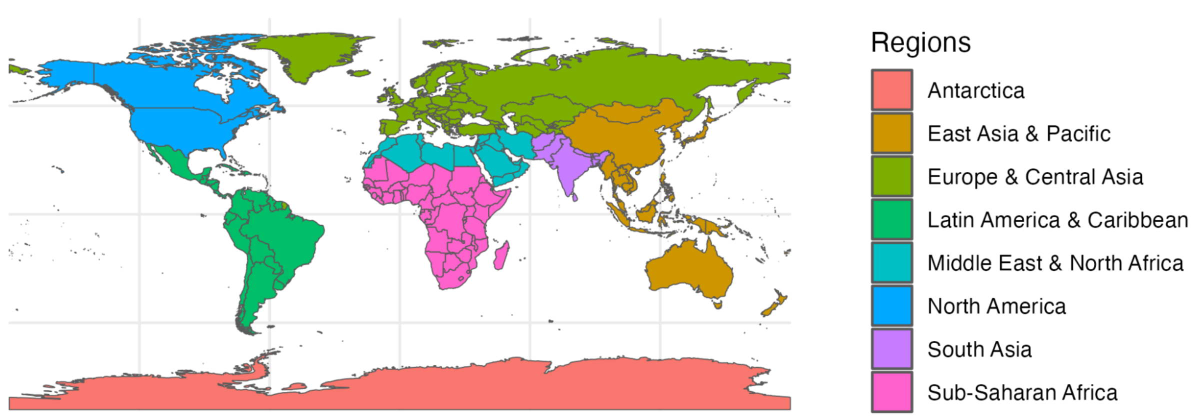
Our dataset was gathered via an observational study and the observational units are individual meteorites. The sampling scheme is a census in order to make the dataset comprehensive. As such, this dataset is very large. The sample size is 45,716 and we filtered the data to remove inaccuracies and make our visualizations more comprehensible. As recommended by Kaggle, we filtered out any years before 860 and after 2015 because the data before and after these times is less accurate. Additionally, for our analysis about Classification, we chose to select only the top ten classifications (CM2, H4, H4/5, H5, H6, L4, L5, L6, LL5, and LL6) to make the visualization easier to understand. Any other filtering will be discussed later in the description of our visualizations. The dataset contains ten variables. Name refers to the name of the meteorite, which is typically a location and is often modified with a number, year, or composition. ID is a unique identifier for the meteorite. Nametype is either Valid, meaning a typical meteorite, or Relict, meaning a meteorite that has been highly degraded by the weather on Earth. Recclass, which we renamed Classification for clarity, is the class of the meteorite. Meteorites are classified according to a variety of characteristics, especially mineralogical, petrological, chemical, and isotopic properties. Mass is the mass of a meteorite, measured in grams. Fall is whether a meteorite was seen falling to Earth or was discovered after its impact: Fell means the meteorite's fall was observed and Found means the meteorite's fall was not observed. Year is the year the meteorite fell or the year it was found. Reclat is the latitude of the meteorite's landing and Reclong is the longitude of the meteorite's landing. Geolocation combines Reclat and Reclong into a parentheses-enclosed, comma-separated tuple. We decided to focus our visualizations on Classification, Mass, Year, and Geolocations because we believed these variables offered the most interesting insights for our questions of interest. Our response variables were not innate to the raw dataset but we looked at several covariates for our analysis, including Reclat, Reclong, Mass, and Reclass.

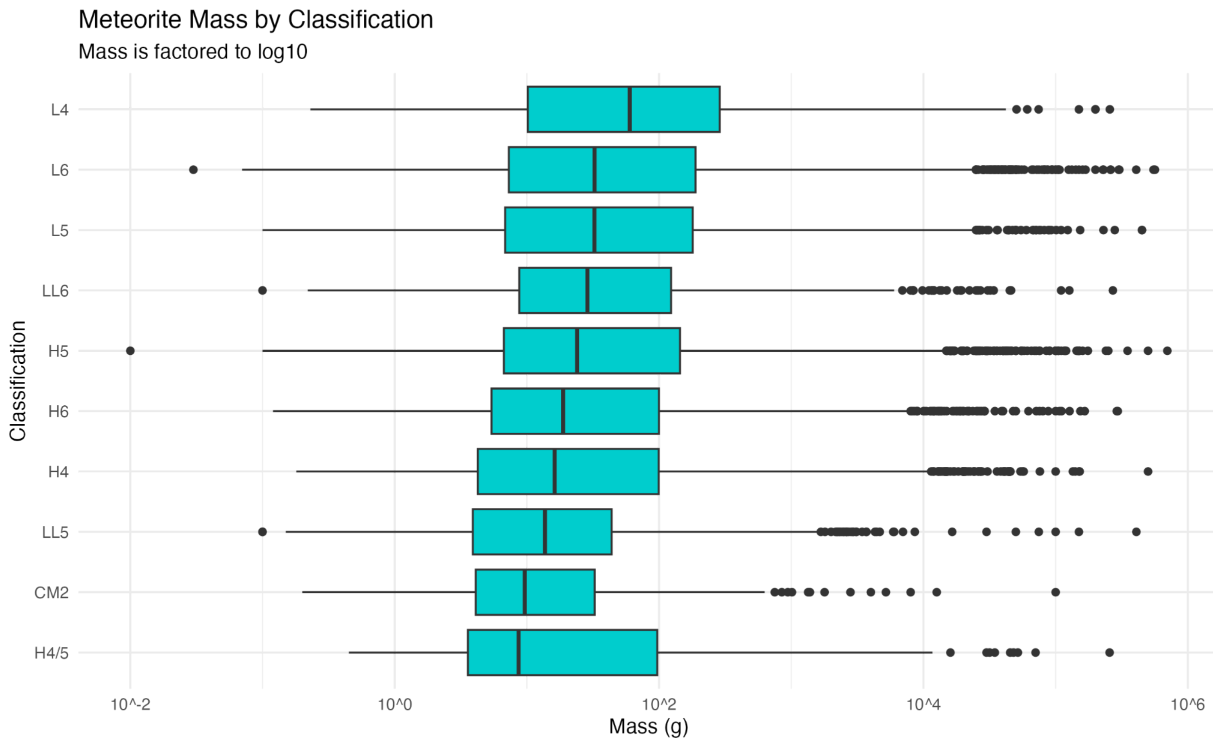
Results:



Region	Meteorite Count
Antarctica	17381
East Asia and the Pacific	496
Europe and Central Asia	321
Latin America and the Caribbean	402
Middle East and North Africa	3934
North America	1101
South Asia	111
Sub-Saharan Africa	335
NA	6371

Our first question of interest was "What is the global distribution of the top ten classifications of meteorites?" To answer this question, we overlaid our dataset on RStudio SF global mapping tool with the help of Dr. Gore. Because there are so many classifications, we limited this visualization to the ten most common. This visualization depicts the meteorites distributed unevenly across the globe with clusters in certain regions. There are dense clusters of meteorites in the United States and Western Europe and less dense clusters in South America, North Africa, India, and the Pacific Islands. In order to provide compelling descriptive statistics we separated the globe into the regions specified above and generated a meteorite count in each region with the help of Spencer the tutor. However, there is a mismatch when comparing the meteorite count by region to the world map. Most notably, the meteorite count in Antarctica is supposedly 17,381 but our world map depicts maybe a few dozen. Spencer hypothesized that there is an inconsistency between the SF mapping tool and our data points causing many of the meteorites in Antarctica to not be depicted. In addition to this inconsistency, we believe that the uneven distribution of meteorites may be due in part to a data collection bias. As previously stated there is a much higher distribution of meteorites in heavily populated areas. This is to be expected because meteorites are more likely to be seen falling or found in areas that are inhabited. For our group, this begs the question of what a more accurate and complete meteorite distribution would look like.

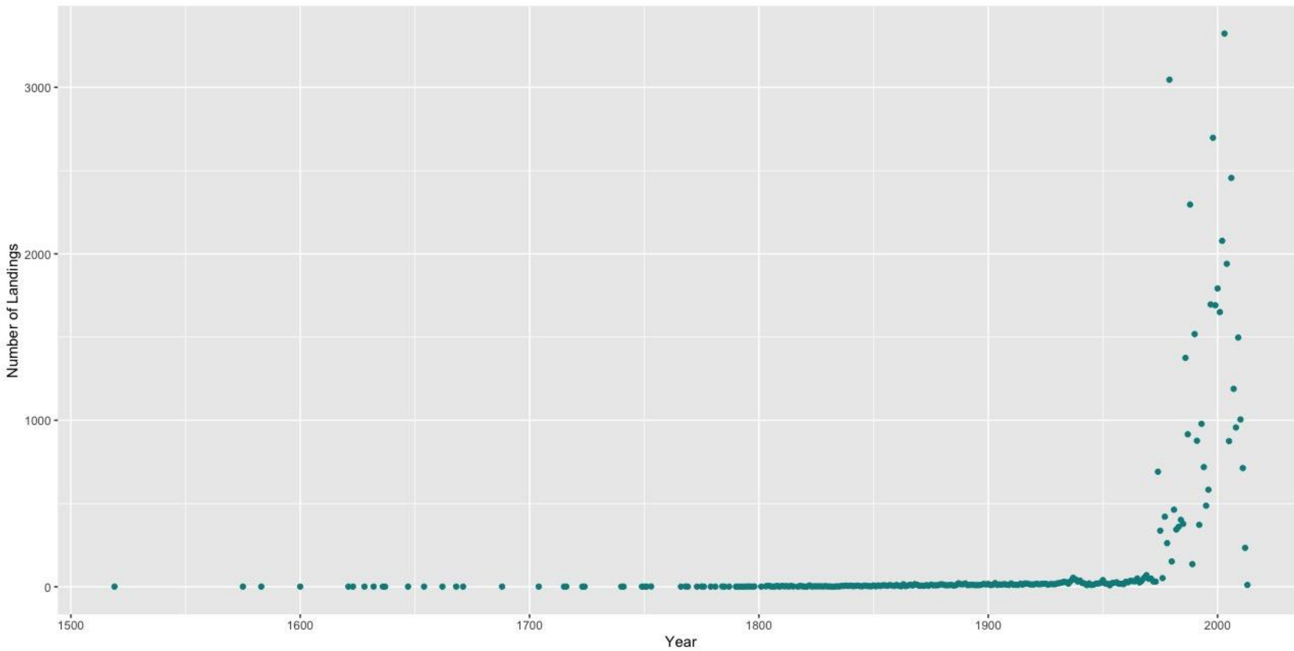




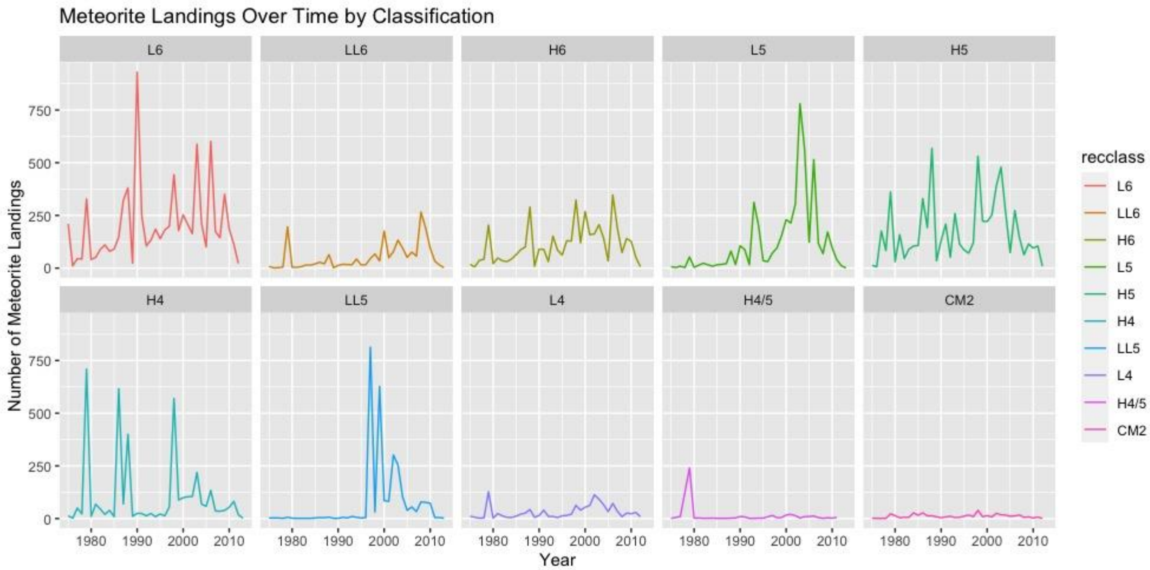
Classification	Minimum	Median	Maximum
L4	0.23	60.00	257000
L6	0.03	32.50	564000
L5	0.10	32.50	1750000
LL6	0.10	28.70	271000
H5	0.01	24.00	4000000
H6	0.12	18.79	295000
H4	0.18	16.21	500000
LL5	0.10	13.70	408000
CM2	0.20	9.64	100000
H4/5	0.45	8.65	256000

Our second question of interest is "How does mass vary between the top ten classifications of meteorites?" To answer this question, we created boxplots to visualize the mass (in grams) of the top ten classifications. Our boxplots depict the lower-bound outliers, the first quantile, median, third quantile, and upper-bound outliers. To represent all of our data points, which ranged from .01 grams to 4,000,000 grams, we factored our mass scale to log10. We found relatively little variation in median mass between classifications and the median masses were surprisingly low; they ranged from 8.65 grams (H4/5) to 60.00 grams (L4). Though there are far more upper-bound outliers, the median range contradicted our assumption that meteorites tend to be very massive. Upon doing more research it makes sense because "large meteoroids fragment as they come through the atmosphere and land" which suggests that some of the small meteorites likely originated from larger ones ("How big are meteorites?").

Number of Meteorite Landings by Year

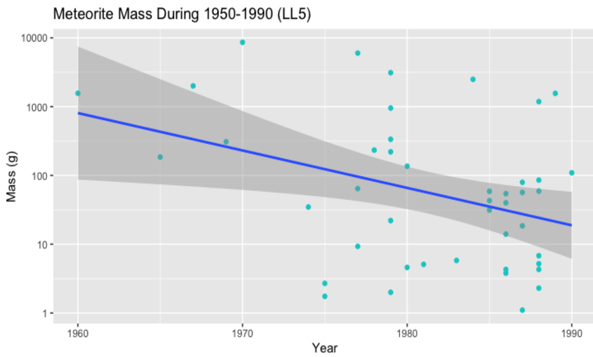


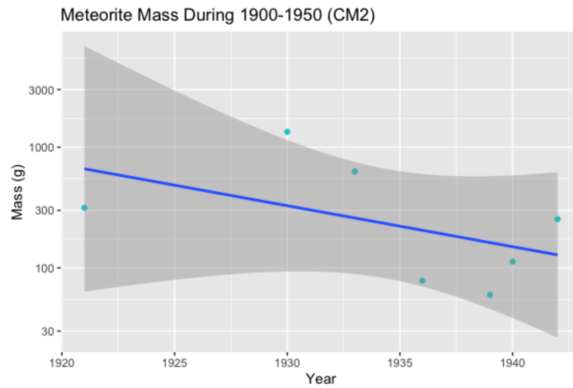
Another of our questions of interest was how the number of recorded meteorite landings changed over time. To answer this question, we created a scatter plot of the number of landings recorded for each year from 1519 to 2013. To improve the readability of our visualization and exclude leverage points, we filtered out the few meteorite landings recorded before 1519. We also filtered any recorded past 2013 based on the advice of Kaggle, to exclude data that was potentially inaccurate. As we expected, the number of recorded landings generally increased over time; however, the number of recorded landings began to trend much higher after the year. Because of this, the range of recorded landings per year is very wide: many years had as few as one or zero recorded landings, while the greatest number of recorded landings in a year was over 3,000. This is also why the mean and median recorded landings are drastically different: the median year saw around 10.5 meteorite landings, but the average number of landings per year was 176. It should be noted that this visualization incorporates data on every classification of meteorite, not just the top ten most common.

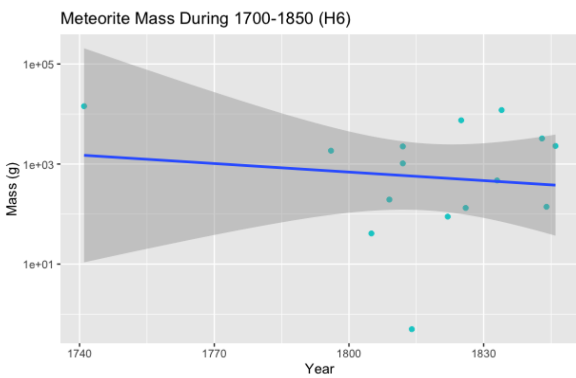


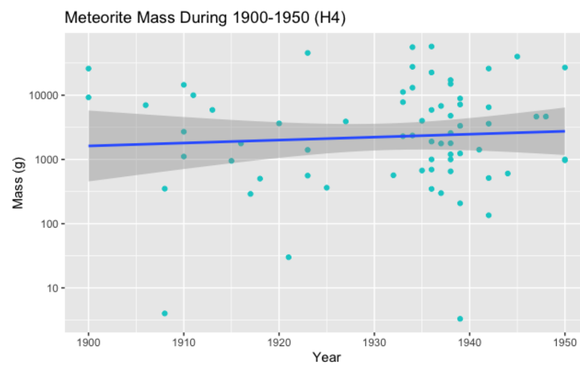
Classific ation	Minimum	Median	Maximum
L4	2.0	21.0	127.0
L6	10.0	168.5	930.0
L5	1.0	53.0	780.0
LL6	1.0	20.0	266.0
H5	6.0	118.0	569.0
H6	6.0	88.0	347.0
H4	2.0	39.0	709.0
LL5	1.0	6.0	813.0
CM2	1.0	9.0	39.0
H4/5	1.0	3.0	240.0

Our fourth question of interest was how the number of landings varied between meteorites of different classifications. We chose to restrict these graphs to the period after 1975, to make the data more reflective of the actual frequency of each type of meteorite and avoid bias created by technological limitations and incomplete information. As with the other visualizations, we limited ourselves to the ten most common classifications: L6, LL6, H6, L5, H5, H4, LL5, L4, H4/5, and CM2. The table below depicts the minimum, median, and maximum number of landings per year of each classification. All of these classifications are chondrites, meteorites formed from accumulated dust, and small grains of rock. Almost all are what are known as ordinary chondrites. CM2 is unique because it is a carbonaceous chondrite, which contains much higher carbon and organic compounds than ordinary chondrites; CM2 is also the least common classification within the top ten. The letters ‘LL’, ‘L’, and ‘H’ refer to the iron content of the meteorite, with H chondrites having higher amounts of iron, L chondrites having lower amounts, and LL chondrites having the least iron and low metal content in general. The numbers refer to different geological characteristics. The frequency of each group is relatively similar, though LL meteorites are less common than L or H meteorites. There is an interesting variation year-to-year for each classification: for example, there is a dramatic spike in the number of LL6 landings after 1995.









For our linear regression model, our question of interest is "What is the relationship between the year a meteorite fell or was found and the meteorite mass?" In other words, is there a correlation between the Mass and Year variables? Similar to our other visualizations, we filtered down our data by classification. By filtering for a certain classification, we can narrow down the number of meteorite landings, and determine if there are trends based on the classification and not a general dataset with meteorites of all classifications In the example regressions above, we chose the classifications LL5, H4, CM2, and H6. In addition, limiting the time period for our analysis allows us to examine the data more closely without a large dataset. In the example regressions above, we selected the time periods 1700-1850, 1950-1990, and 1900-1950. We also filtered out strong outliers in each graph to generate linear models that are less skewed.

To accurately assess if a correlation exists between the year a meteorite fell or was found and its mass, we use the *R-value* of the linear regression model. The R-value is a variable used to assess *linear trends*, whether they exist, if they are positive or negative, and how strong or weak the trends are. The range of R-values is $-1<R<1$. If R is 0, there is no linear relationship observable. The greater the absolute value of R, the stronger the linear trend is. All of the classifications visualized above have very small correlations between Year and Mass, and as such, their R-values are small as well.

During 1950-1990, LL5 meteorites have an R-value of -0.12524. This means that for every year during this time period, the mass of the meteorites found decreased by approximately 12%. During 1900-1950, CM2 meteorites had an R-value of -0.07818, meaning that for every year during this time period, the mass of the meteorites found decreased by approximately 7%. During 1900-1950, H4 meteorites had an R-value of 0.01053, meaning that for every year during this time period, the mass of the meteorites found increased by approximately 1%. Finally, during 1700-1850, H6 meteorites had an R-value of -0.01308, meaning that for every year during this time period, the mass of the meteorites found decreased by approximately 1%.

Conclusion:

Our group chose the "Meteorite Landings" dataset with the hopes of generating compelling conclusions about the distribution of meteorites over time and space and the relationship between other variables of interest, like mass, year, and classification. We were able to generate a visualization of the global distribution of meteorite landings and concluded that meteorites are more likely to be found or seen falling in inhabited areas, leading to interesting clusters of meteorites. Additionally, when investigating the relationship between classification and mass, we found the median mass to be relatively similar regardless of classification. The most interesting conclusion drawn from this visualization is the huge range of mass and surprisingly low median values. We also concluded that any change in the number of recorded meteorite landings over time is likely due to technological advancements that allow data to be more accurately recorded and that the large majority of meteorite landings are chondrites.

We were able to analyze our data with a primary focus on meteorite classification. The classification, or recclass information, within our dataset allowed us to break down our dataset into smaller sizes, and test out our hypotheses on more manageable data. Furthermore, by narrowing our focus to specific classifications, we were able to observe real trends, such as clusters of meteorite landings depending on classification, the median mass depending on classification, and other analyses. The classification of meteorites was an extremely helpful tool that provided us with the chance to closely analyze our data and come to more specific and informative conclusions.

Limitations:

The first limitation of analysis that we encountered was the lack of information available about the data collection methodology. We would have liked to know more about the methodology so that we could account for any potential bias in our analysis. Without information about how this data was collected, it is difficult to determine the accuracy and completeness of our dataset. There is also little available information about the criteria used to classify meteorites, and whether those criteria have changed over time. This makes it difficult to draw concrete conclusions about the significance of variations in mass and frequency between different classifications of meteorites.

We also encountered issues mapping our data using RStudio's global mapping tools. Due to an inconsistency between our data and the tool, many data points did not appear on the map, making it difficult to draw informed conclusions from this visualization. Furthermore, we had to omit many data points from the map because they had incomplete or missing geolocations. From this map, we were however able to infer that there was a data collection bias causing there to be much higher meteorite distributions in densely populated areas.

Another limitation is the accuracy of our dataset. Though we filtered before 860 and after 2015 as recommended by Kaggle, we still utilized very old data points that may lack accuracy in our analysis. Additionally, while the number of recorded meteorites drastically increased in the last few decades of the 20th century, this is not because meteorite landings are becoming more frequent over time. This is very likely because recent technological advancements have allowed us to more accurately record the number of meteorites that enter Earth's atmosphere. This means that the data we have on meteorite landings for most of human history should be regarded as incomplete. We should consider this when trying to draw conclusions about our dataset and meteorites in general.

Future Directions for Research:

To draw more compelling conclusions from this dataset, it would be helpful to reduce the limitations discussed above. In particular, we would like to know more about the data collection methodology to inform our analysis. Additionally, with a more comprehensive and accurate dataset, we would generate a more insightful map depicting the distribution of meteorites. We would like to know if meteorites are truly more likely to land in certain regions or if this clustered distribution is due to the biases that have been discussed. Finally, we would be interested in learning about how meteorites are classified and including more of the classifications in our analysis.

Works Cited

"How Big Are Meteorites? | Some Meteorite Information." *Washington University in St. Louis*, sites.wustl.edu/meteoritesite/items/how-big-are-meteorites.

Meteorite Landings - Catalog. 6 Dec. 2023, catalog.data.gov/dataset/meteorite-landings.

"Meteorite Landings." *Kaggle*, 5 Nov. 2016, www.kaggle.com/datasets/nasa/meteorite-landings.

"What Are Meteorites? Ancient Clues to Our Solar System | AMNH." *American Museum of Natural History*, www.amnh.org/exhibitions/permanent/meteorites/meteorites/what-is-a-meteorite#:~:text=In%20simplest%20terms%2C%20a%20meteorite,comets%20%E2%80%93%20in%20our%20solar%20system.