INFO 5100

Project 3 Final Report

Kendall Hoffman (kah342), Tangwuyou (Christina) Su (ts789), Hanna Yen (hyy5), & Nimra Shakoor (ns924)

Data Description and Cleaning

The original meteorite landing dataset we found was from The Meteoritical Society posted on Data.gov. However, for the sake of time and simplicity, we used a cleaned version of the dataset which we found on Kaggle, and made additional changes to best suit our needs throughout the project. For example, we dropped all null values (most were present in the “mass” and “year” columns) to reduce noise or incorrect visualizations, and dropped years that were either negative or in the future (rows with year values greater than 2023). This allowed us to plot a histogram of the frequency of meteorites over time more accurately. We noticed that nearly all of the meteorites within the dataset had fallen or been found after the year 1800, most likely as a result of improved record keeping and the modernization taking place during the Industrial Revolution. As a result, we decided to filter the data to only include meteorite landings after the year 1800.

The dataset was still very large after this initial round of filtering so we decided to narrow our focus to include only meteorite landings in the United States (1,716 meteorites). We figured this would narrow the scope of the data, and could be more interesting to view on a map, as opposed to viewing dots on a global map. We did this using the longitude and latitude bounds of the United States' northernmost, southernmost, easternmost, and westernmost points. We then determined where each meteorite had landed using the “rec_lat” and ‘rec_long” columns and removed them if they fell outside of the “box” our filter had drawn. It is important to note that only 4% of the dataset was considered to have included meteorites that have fallen in this location. We thought it would be interesting to see how meteorite landings vary across different states and areas of the United States, and potentially consider how climate and ecosystems may contribute to more frequent meteorite landings. Due to time constraints, we were unable to incorporate weather or ecosystem data into our visualization but maintain this to be a goal of future iterations.

After filtering the dataset to include only United States meteorite landings starting in the year 1800, we further investigated the “rec_class” column of the dataset to determine how we could visualize different types of meteorites. We determined that there were 81 unique classes within the dataset so we sought to narrow this down significantly. We learned that meteorite classifications are based on composition and sub-classes (which were represented in our dataset) could be categorized into one of three categories: stony, stony-iron, and iron. We then determined which general class each of the sub-classes fit into and recategorized all the meteorites in our dataset to belong to only one of the three general classes.

Shown below is a table that details each column in our dataset, a description of the column, and any notes associated with it:

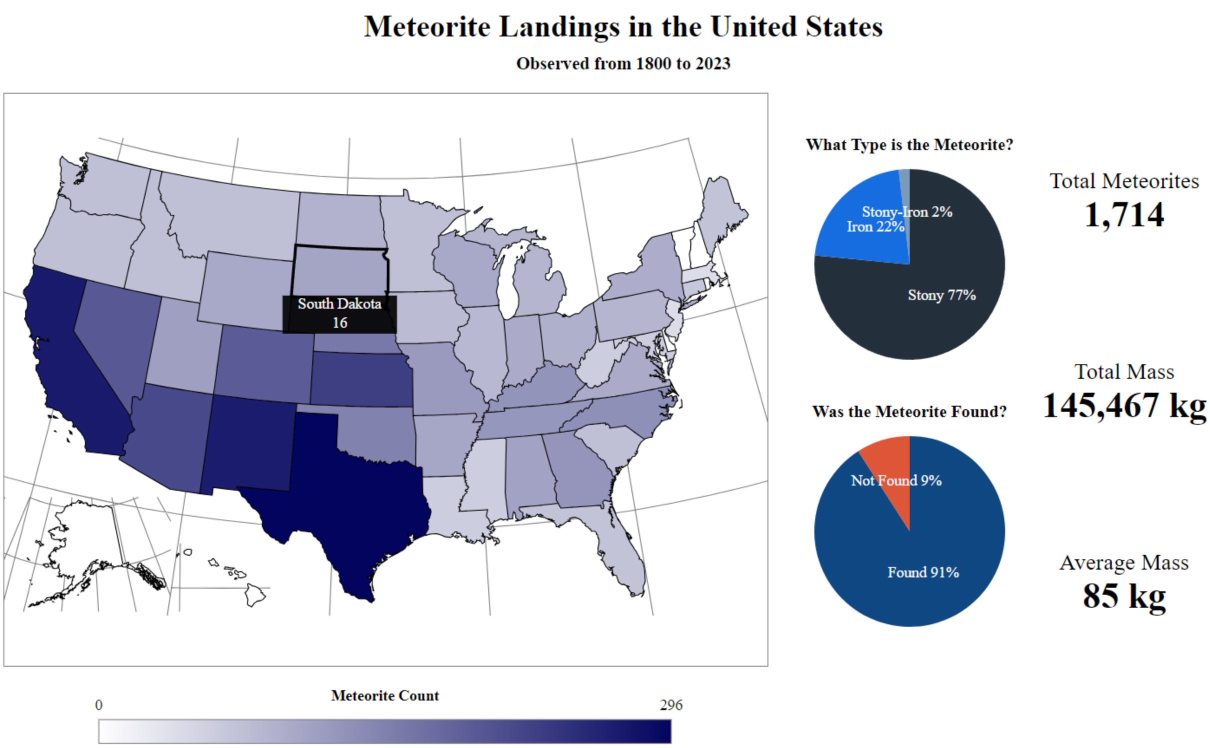
Data Column	Description	Notes
name	Name of the meteorite	Is sometimes the city of the meteorite’s location
id	The number of the meteorite	Not used in our analysis
name_type	Can be “valid” or “relict”. Valid is a normal meteorite while relict is a heavily weathered meteorite	Not used in our analysis
rec_class	The class of the meteorite	Classes are based on physical, chemical, and other characteristics. We changed this to be one of three more general classes: stony, stony-iron, or iron.
mass	The mass of the meteorite in grams	
fall	Can be “found” or “fell”. Fell is a meteorite that was observed falling to earth. Found is a meteorite that was found on the ground.	
year	The meteorite fell or the year it was found (depending on the value of fell)	
rec_lat	The latitude of the meteorite’s landing	
rec_long	The longitude of the meteorite’s landing	
geo_location	A tuple that has the latitude and longitude combined	Dropped, all NaN values
location_type	Can be either “Exact” or “Guessed”. Exact were carried over from the original dataset by the creator, or were Guessed based on information from the data by the creator.	

In addition, we needed a topoJSON dataset to plot each meteorite on a map of the United States. We used a United States JSON previously provided to us in class. Given the number of points in the dataset, we only wanted to display meteorites in their respective locations when a user clicks on a state of interest. To make this functionality possible we needed a way to determine which state each meteorite had fallen in using the given latitude and longitude. In a separate Python script, we used the Python package geopy.geocoders to identify the state of the meteorite landing. Some additional manual filtering was needed for a few meteorite landings that were from Canada or Mexico. Then the filtered data was grouped by state and aggregated by the row count to get the total number of meteorite landings in each state. Shown below is the aggregated dataset we used to create the initial top layer of our visualization:

Data Column	Description	Notes
state_name	Full name of the US state	
state_code	Numerical representation of US state	
state_short	US state abbreviation	Not used in our analysis
total	Total meteorite landings (from 1800 to 2023)	

Full details of our data cleaning process and final datasets can be found in our GitHub repository.

Visual Design Rationale



United States Map

The main part of our design is a map of the United States created using the aforementioned topoJSON file. Each of the 50 states is visible and outlined with a thin black line consistent with most map visualizations to ensure they do not interfere with the data being displayed. The latitude and longitude lines are also present to help show individual meteorite locations. When the entire country is in view, the color of the states corresponds to the total number of meteorites that fell there. A legend is provided to help convey the cutoff numbers for each color with 0 meteorite landings as the minimum and 296 as the maximum. Darker purple corresponds to more meteorites, whereas lighter purple represents fewer. A few states don't have any meteorites and are colored white to show this lack of data.

Individual Data Points (State Level)

The individual data points represent the meteorite landings at the exact location within each state. Meteorite landings in a selected state are revealed when a state is clicked on. The map also zooms into the center of the specified state view, and all state colors are proportionately lightened in the background to enhance the contrast between a state's color and the color of the data points that appear. Each data point is marked on the map with a dot. The dot's size correlates to its mass and the color of the dot corresponds to one of three general categories to show the composition of the meteorite. To make the representation of meteorite sizes more visually accessible, we employed a logarithmic scale for mapping the mass to the size of the dots. This approach ensures that even meteorites with smaller masses are visible on the map. However, using a logarithmic scale can lead to a misrepresentation of the actual size differences between meteorites. For instance, a meteorite that is twice as massive as another might not appear exactly twice as large due to the nature of the logarithmic scaling. Additionally, as we zoom in on each state of the USA, the zoom factor and the relative sizes of each state differ. The sizes of the dots representing the same mass meteorites may appear different when viewed in separate states. While this design choice aids in providing a detailed view of each state's meteorite landings, it's a tradeoff that can affect the perceived uniformity of meteorite sizes across different states. Stony meteorites are represented by dark gray dots, whereas Iron meteorites are represented as blue ones. When a state is selected, a legend is displayed to help show which classifications are represented by the data points. A "zoom-out" button is also presented in the bottom right of the map so users can navigate back to view the entire country again. This was to ensure easy navigation and interaction with the map.

Pie Charts

Two descriptive pie charts are statically displayed on the right-hand side of the screen to help show different information about all of the meteorites represented on the map. The top pie chart aims to show the distribution of each meteorite classification by showing how each type is represented throughout the data. The cool tone colors of the top pie chart are identical to the meteorite-type colors of the individual meteorite landing points for consistency. The bottom pie chart displays the proportion of meteorites that were observed as they fell to Earth, compared to the proportion of those that were found on the ground. For this bottom pie chart, we labeled the categories as "not found" and "found" in contrast with the original data's "fell" and "found" because we felt that "not found" would make more sense to the user than "fell" in the context of our graph; it could otherwise be confusing that meteors that were "found" were not also ones that "fell." The bottom pie chart colors are meant to blend well with all of the colors used previously, still provide an opposing distinction between the two categories of the pie chart, and accommodate color blindness.

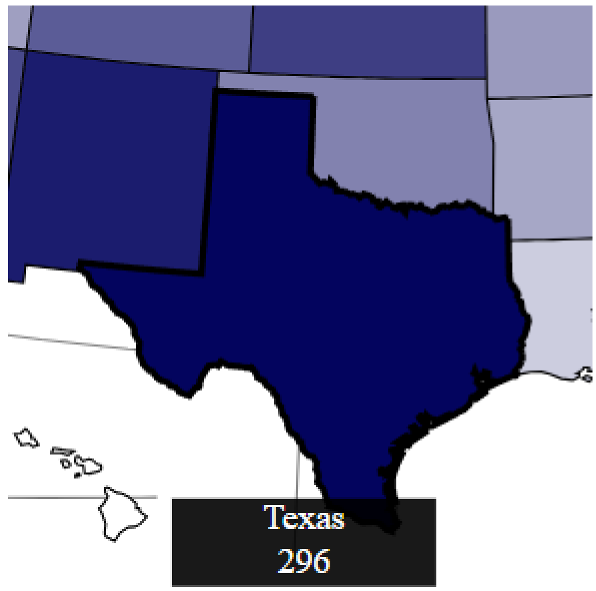
Additional Statistics

We also included other overall descriptive statistics such as the total number of meteors, total mass, and average mass. These numbers have commas on them and were rounded to whole numbers for readability. In this case, we felt that more precision was not as important since the number of significant digits was already quite high. Regarding the presentation of the numbers, we considered putting them in boxes with gray outlines to match the gray outline of the map. Ultimately, we decided on bolding the numbers and making them as big as possible in the space we had to call more attention to them. Putting outlines around the numbers would have limited the usable space.

Interactive Elements Design Rationale

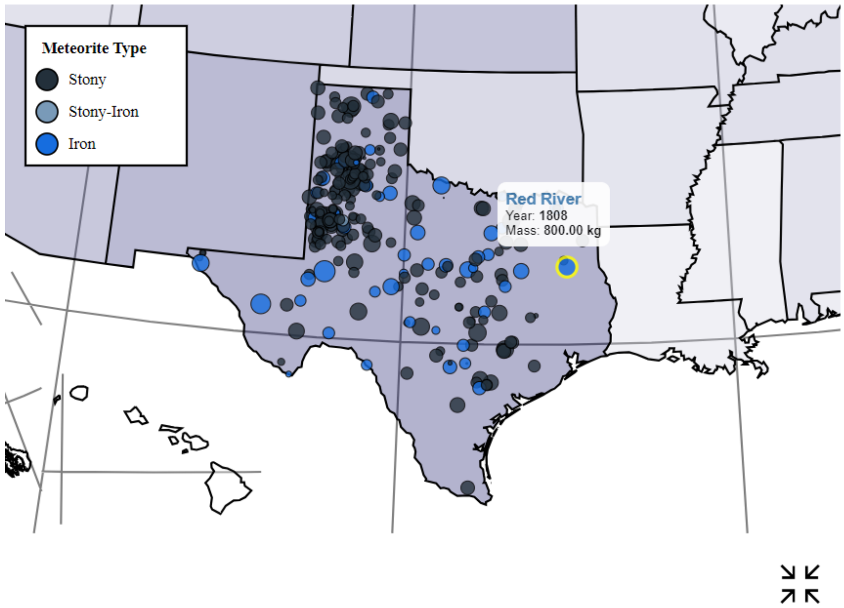
United States Map

All interactivity takes place within our map visualization to help control the amount of data presented at a time to viewers. At the country view, each state can be hovered over to reveal a descriptive label including the name of the state and the number of meteorites contained within the state. The cursor changes to a pointer to make the clickability of each state discoverable for viewers. The label is dark with white text to contrast with both the color of the states and the white background of the map. The border of the state becomes thicker as well to help viewers better visualize the state they are hovering over. Observe the image on the right for a reference.



Individual Data Points (State Level)

When a user clicks on a state, the map zooms in and centers the selected state in view. The data points (meteorite landings) that are revealed can be hovered over to reveal the name of the meteorite they represent, the year associated with it, and its mass. The name of the dot is represented in a light blue color to show its priority relative to the other information. The mass and year are represented in smaller, black text. As previously mentioned, the data points are distinguishable at first glance by color and size. In addition, the dot's border also becomes yellow which provides a stark contrast compared to the background and the other dots. This is to help viewers determine which dot they are hovering over at any given time which improves visibility when interacting with the data visualization. See below for a reference.



Additional Click and Zoom Functionality

While viewing a state, users can directly select another state that remains in view. Upon doing so, the dots from the original state appear to “move” to the new state. The map follows suit to ensure the new state is centered in view. Users can continue to navigate from state to state or return to the country view with the “zoom out” button in the bottom right corner.

<u>Visualization Objective and Insights</u>

Our main objective when creating this visualization was to determine where meteorites tend to land within the United States, and if there are any trends or patterns taking place. From the country view of our visualization, viewers can easily tell that meteorite landings are concentrated in the south and southwest of the country. Texas (296) appears to have the most meteorites, followed by California (227), and then New Mexico (221). While Texas and California are two of the largest states in the country, it is interesting to consider other factors that might cause meteorites to fall (or rather be located) in these locations disproportionately more than in states such as Vermont, New Hampshire, and Delaware which each have zero meteorites (and are also some of the smallest states in the country).

We also wanted to visualize the exact locations of meteorite landings within each state. Clicking on a state with a high frequency of meteorites such as Texas reveals nearly three hundred dots that seem to be clustered within the panhandle. In California, they are clustered in the south. Taking a closer look, we can determine that these areas are the High Planes and Mojave Desert ecoregions, respectively. While this may not influence the frequency at which meteorites land in these areas, their grassy and arid characteristics may make meteorites easier to find after they hit the ground. This can be compared to states with no meteorites like Vermont and New Hampshire which are Northeast Highlands – dominated by dense hardwood and coniferous forests. While we were unable to include ecoregion data directly in our visualization, knowledge of it provides valuable information and contexts regarding the data we were able to present.

It was surprising to us how few meteorites fell within the United States (only 4% of the original dataset were meteorites that fell within the United States). We assumed there would be more due to the prolific nature of The National Aeronautics and Space Administration (NASA) in terms of space exploration and discovery. We thought it was interesting that there weren't more meteorites detected within the United States, especially with modern technology and satellite systems. This is interesting when compared to the bottom pie chart which compares the proportion of meteorites that were detected as they were falling to Earth and those that were discovered on the ground at a later point. We expected that perhaps satellite imagery would assist in detecting when meteorites fall and that this number would outweigh the number of those found on the ground. This further spurs our curiosity regarding how meteorites are typically discovered when they are on the ground, and what landscape characteristics may contribute to the ease with which they can be found.

Team Contributions

- Hanna: Found JSON datasets, cleaned and reformatted JSON data, created US map with total meteorite landings, created meteorite count legend, created meteorite type legend, enabled different color and sizes for individual meteorite landings, assisted with report editing (6-7 hrs)
- Nimra: Calculated summary statistics and made the right-side non-interactive visualizations; chose color schemes for meteor classes; made edits to report. (6-7 hrs)
- Christina: Mapped different meteor classes to 3 classes in data cleaning; added meteor data display of each state, zoom in/out update function, mouseover/out (with info box) function, stylistic changes according to critiques; made edits to report. (6-7 hrs)
- Kendall: Found all of the meteorite datasets, imported them into our repository, and cleaned this dataset to make visualization easier. I was also responsible for writing the report, and critiquing the usability of our visualizations as I wrote about our project and provided my teammates with feedback so these edits could be incorporated into our visualization for better usability. (6-7hrs)

Appendix

Initial Sketches: