Question 1

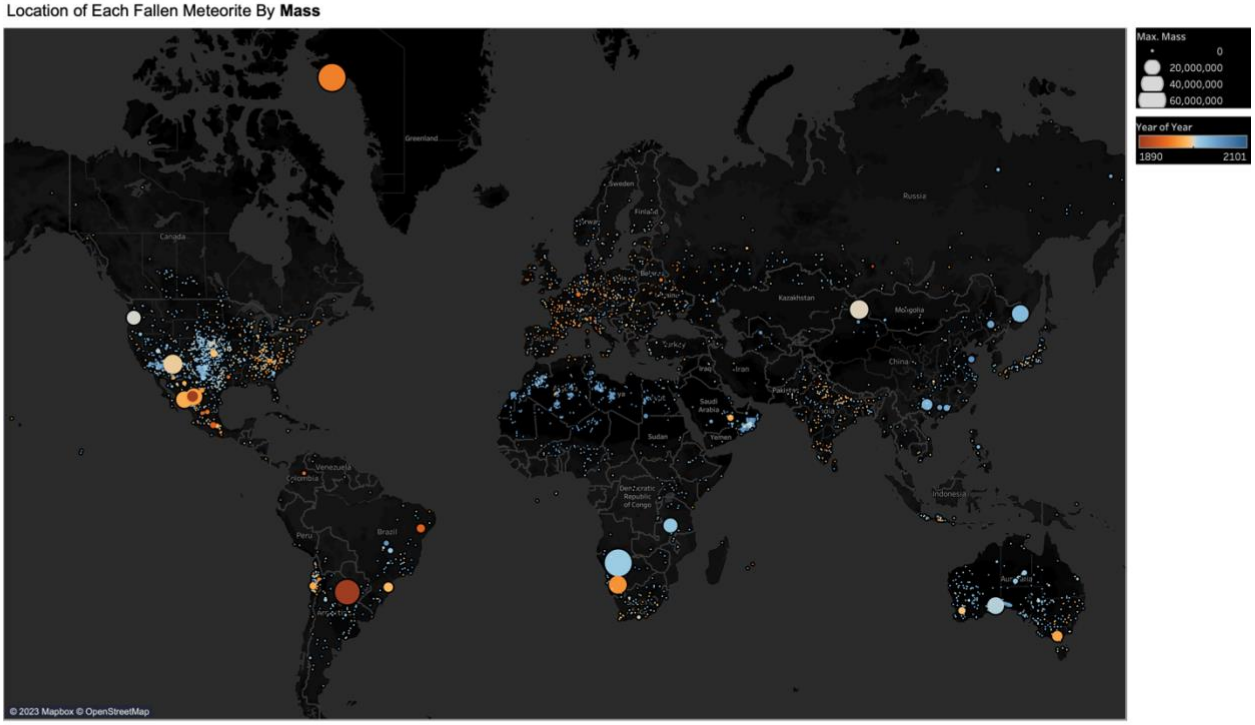
- **Looking at the data, what two variables do you find critical to the analysis?**

When analyzing the meteorite data provided, two critical variables that stand out are the "Mass" and "Location" columns.

Mass is an important variable to consider when analyzing meteorites as it can provide insights into the size and composition of the meteorites. Studying the distribution of mass across different locations and over time can help identify patterns and trends that can inform our understanding of the origin and nature of meteorites.

Location is another important variable to consider when analyzing meteorites as it can provide insights into the geographic distribution of meteorites and their potential impact on the environment. Analyzing the location of meteorites can help identify areas with a high concentration of meteorites and potential areas of impact, which can inform our understanding of the risk and impact of meteorite strikes on the Earth. Additionally, location can also be used to study the relationship between meteorites and other environmental factors, such as climate and geology.

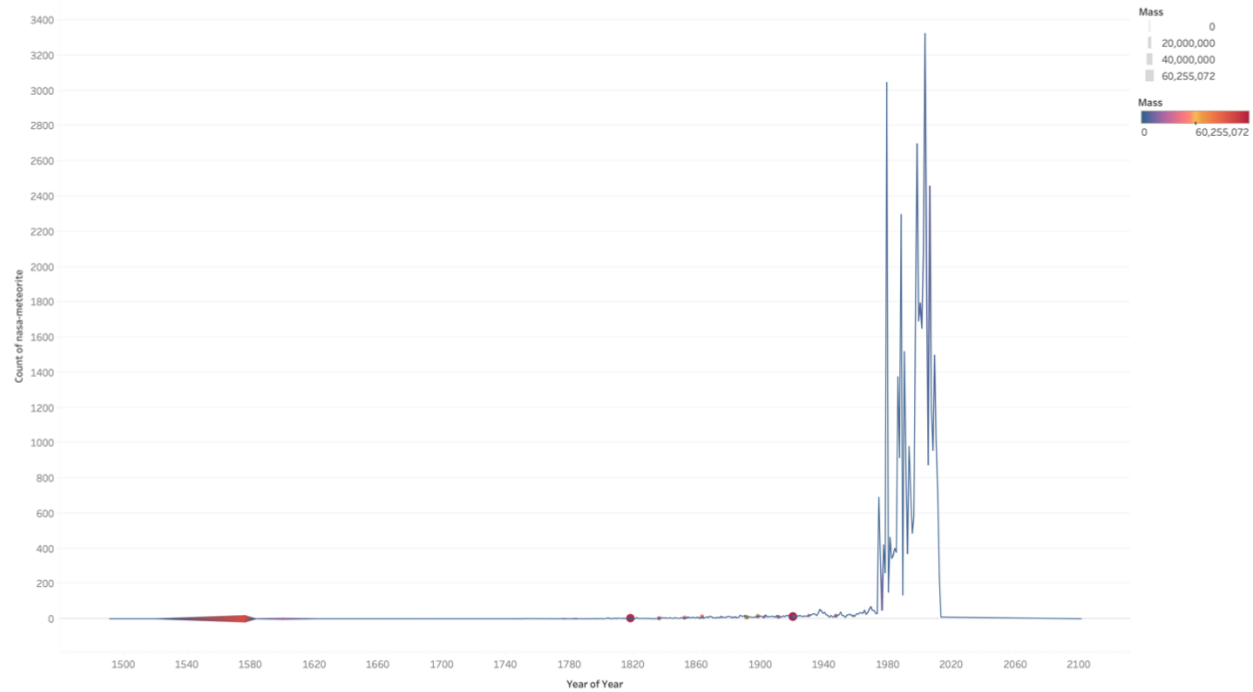
Location of Each Fallen Meteorite By **Mass**



Map based on Long and Lat. Color shows details about Year Year. Size shows maximum of Mass. The view is filtered on Long, which keeps non-Null values only.

Aside from the two variables mentioned above, the "Year" column is also essential because it allows us to see when the meteorite falls occurred. By analyzing the distribution of meteorite falls by year, we can identify any patterns or trends in the data and focus on recent falls. This can help us determine which locations to investigate further and potentially locate recent meteorites that have fallen to Earth.

Distribution of Meteorite Falls By **Year**



The trend of count of nasa-meteorite for Year Year. Color shows sum of Mass. Size shows sum of Mass. The view is filtered on Year Year, which ranges from 1473 to 2101.

- **Can we use the data to map where each meteorite has fallen?**
- **If this is possible, what data field is critical for this?**

Yes, we can use the "Location" variable in the data to map where each meteorite has fallen. The "Location" data field is critical for mapping where each meteorite has fallen. This data field provides the latitude and longitude coordinates of the location where the meteorite impact occurred. By using this information, we can plot the location of each meteorite on a map using Tableau or any other mapping tool. This can provide valuable insight into the geographic distribution of meteorite impacts and help us identify areas where recent meteorite falls may have occurred.

- **Are there any changes to the data that would make this possible?**

To map the meteorite falls accurately, the "Location" field needs to be in a format that can be recognized by mapping tools. Currently, the "Location" field in the given data set is in a format that combines latitude and longitude values into a single field separated by a space, which can be problematic when trying to plot the locations on a map.

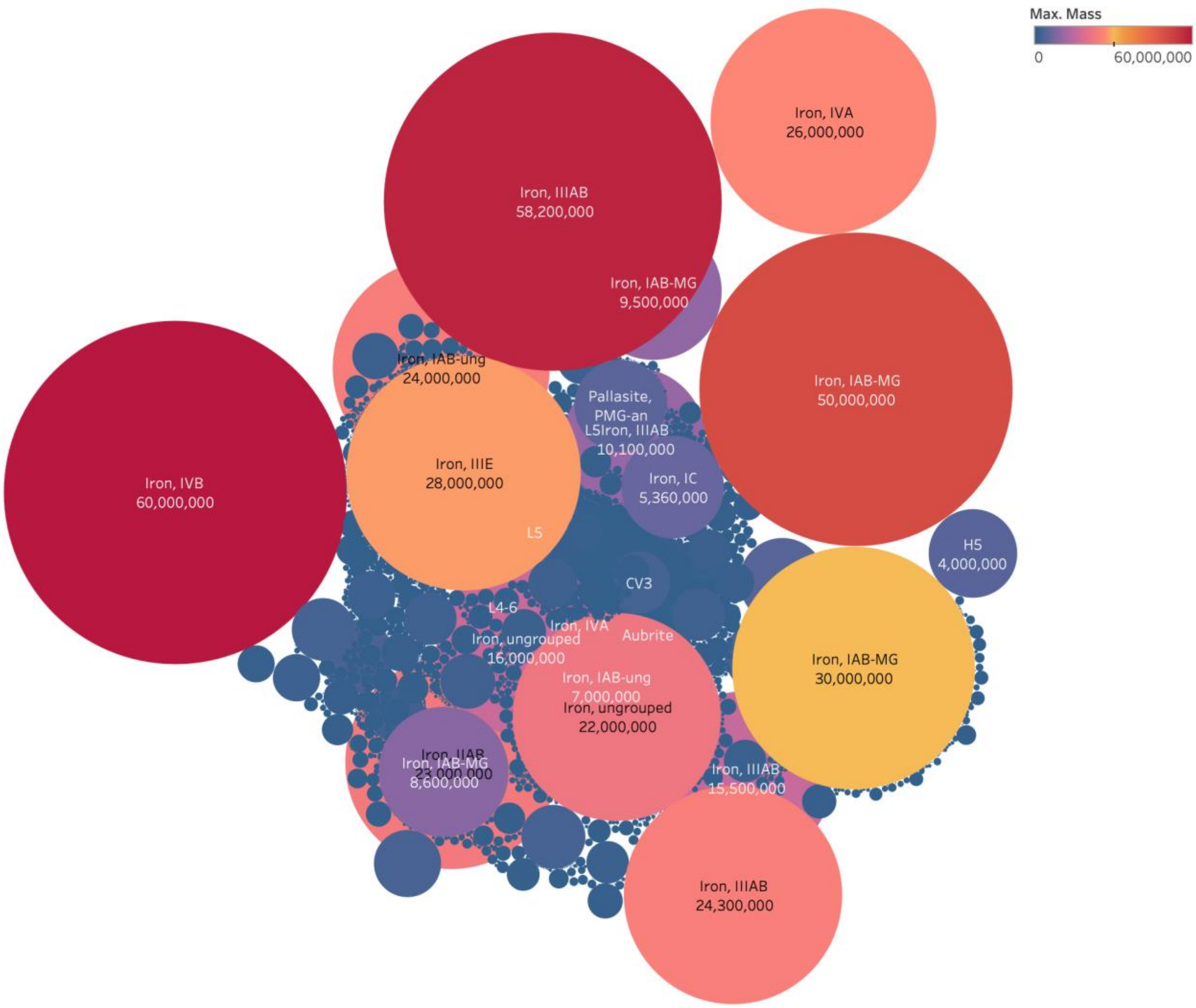
To make mapping possible, it would be better to separate the latitude and longitude values into separate fields. Alternatively, we can split the "Location" field into two separate fields, one for latitude and one for longitude, using Tableau's "Split" function. This will allow us to plot each meteorite's fall accurately on a map.

In addition, having more accurate or precise location data for some of the meteorites in the data set would be helpful. For example, some of the location values are only listed at the city or region level, which could make it difficult to pinpoint the exact location of the meteorite fall. Having more precise location data would increase the accuracy of the map visualization and make it easier to identify areas where recent meteorite falls may have occurred.

- **Can you determine the largest meteorites that have fallen to Earth**?

Yes, we can determine the largest meteorites that have fallen to Earth using the "Mass" field in the given data set.

Largest Meteorites By **Mass**



Recclass as an attribute and maximum of Mass. Color shows maximum of Mass. Size shows maximum of Mass. The marks are labeled by Recclass as an attribute and maximum of Mass. Details are shown for Id.

- **Each column in the data set was initially created by a computer program. Do these names look correct? Do you see any issues that should be fixed?**

At first glance, the given data set's column names appear correct and descriptive.

The "Fall" column indicates whether the meteorite fell to the Earth or was found on the Earth's surface. The "Id" column provides a unique identifier for each meteorite. The "Name" column provides the name of the meteorite. The "Nametype" column indicates whether the meteorite is a "Valid" meteorite or a "Relict" meteorite. The "Recclass" column provides the class of the meteorite. The "Location" column provides the latitude and longitude of the location where the meteorite was found or fell. The "Year" column provides the year in which the meteorite was found or fell. The "Mass" column provides the mass of the meteorite in grams.

However, it's always a good practice to verify the column names and their meanings with the data source documentation or the data provider to ensure that they are correct and consistent with the data source.

Based on the data set provided, there are a few issues that could be improved: Fixing these issues would improve the usability and readability of the data set.

1. There are two “Id” columns. While one of the “Id” is computer generated, the other seems to be in the right format.
2. The "Location" column combines latitude and longitude into a single field, separated by a space. It would be more helpful to separate these into two separate columns, one for latitude and one for longitude. This would allow for easier mapping and analysis.
3. The "Year" column is in a date-time format, which includes the time of day, but all of the values in the column are set to 12:00:00 AM (00:00:00). Since the time information is not needed for this data set, it would be better to remove the time component and keep only the date component.
4. The "Mass" column is in grams, which could be difficult to read and analyze for very large or very small meteorites. It may be more helpful to convert the mass into a more human-readable format, such as kilograms or metric tons.

- **Using the provided data, organize each data type into one of the following categories: Qualitative and Quantitative.**

- **For the Quantitative data, highlight discrete values in blue and continuous values in green.**

Based on the provided data set, the variables can be organized into the following categories:

Qualitative variables:
- Name (of the meteorite)
- ID (of the meteorite)
- Name Type
- Reclassified Type
- Fall (whether the meteorite was seen falling)

- Location (the location where the meteorite was found)

Quantitative variables:

- Mass (in grams)
- Year (when the meteorite was found/fell)

The qualitative variables describe the qualities or characteristics of the meteorites, such as their name, type, classification, and location. The quantitative variables are numerical measurements of the meteorites, such as their unique identifier, the year they were discovered, and their mass.

Question 2

Here are three weaknesses of using a pie chart to visualize total sales by category:

Pie Chart:

- Difficulty in comparing absolute values: Pie charts are better suited for showing proportions and percentages rather than absolute values. While the pie chart shows the relative sizes of different sales categories, it is challenging to determine the actual sales amounts for each category. This makes it difficult to compare the total sales between categories, especially when the difference in size is not visually apparent.

- Lack of Precision: The pie charts are not as precise as other chart types, such as bar charts or tables, in conveying exact values. Since our goal is to provide exact sales figures, a pie chart may not be the best option.

- Difficulty in Highlighting Differences: The pie chart does not allow for highlighting the differences in sales between categories, as the slices do not have a common baseline. In contrast, bar charts or stacked bar charts can show the differences more clearly, allowing for easier comparisons between categories.

Pie charts can be useful in some situations, but they may not be the most effective way to represent total sales data by category. Other chart types, such as bar charts or stacked bar charts, may be more appropriate for highlighting differences in sales between categories.

Packed Bubbles:

Here are three weaknesses of using the packed bubbles chart to visualize the average discount amount:

- Difficulty in Comparing Size and Proportions: The packed bubbles chart makes it difficult to compare the sizes and proportions of the bubbles accurately, especially when there are many bubbles in the chart. It can be challenging to determine the exact values of the average discount amounts for each bubble as well as compare them with one another.

- Limited Information: The packed bubble chart does not provide enough information to accurately represent the average discount amount by category. It is difficult to show the average discount amounts across the regions.

- Misleading Interpretations: The packed bubbles chart is misleading because the bubbles are not proportional to the actual average discount amounts, viewers may misinterpret the data presented. Additionally, the chart does not provide additional context, such as the names of the regions, viewers may draw incorrect conclusions about the data.

Packed bubble charts can be useful in certain situations, but they may not be the most effective way to represent average discount amount data. Alternative chart types, such as bar charts will be more effective for accurately representing the average discount amount by region.

Map:

Here are three weaknesses of visualizing profit by country using maps:

- Distortion of Geographic Area: Maps can distort the geographic area of different countries based on the projection used, making it difficult to accurately compare the profits between countries. For example, larger countries may appear larger on the map, which can create a false impression that they have higher profits than smaller countries.

- Lack of Context: Maps can provide limited context and make it difficult to compare profits between countries if other factors influencing profit are not considered. For example, if two countries have similar profits, it may be misleading to assume that the countries are equally successful, as there may be differences in population size, market size, or other factors that can influence profit.

- Uneven Data Distribution: Maps can be less effective if the data is not evenly distributed across countries. If most of the profits come from a few countries, these countries will dominate the map, and it can be difficult to discern differences in profits between other countries. Additionally, if some countries have no profits or low profits, they may appear as blank spaces on the map, which can create a false impression of their economic status.

We could use chart types, such as bar charts or tables for a more effective comparison of profit between countries. But we will use maps to provide a visual representation of profit by country.