II. Data Visualization

1. Intro

The initial dataset from the NASA had 422 meteorite classes in it. A high number that did not necessarily reflect the distribution of meteorites. Indeed, as shown in the Phase-1 (Data Wrangling), the top ten meteorite classes encompassed approximately 80% of the total number of meteorites. For that reason, I decided to neglect – for now, the remaining 20% of the data – for a more concrete analysis.

2. Visualization

2.1. Scatter Plot

After extracting the needed segments, I used the Matplotlib Library to generate a Scatter Plot (meteorite classes on the x-axis and their mass on the y-axis) to see the distribution of the mass per class. Surprisingly, with the notable exception of a few outliers, the distribution was quite homogenous among the ten classes, with the majority of the meteorites having a mass range 0-500kg. Here are some conclusions:

- The heaviest meteorite belongs to the H5 class;
- The class L6 is more homogenous;
- The class CM2 has the lightest meteorites.

2.2. Seaborn lmplot

I used the Seaborn Library to visualize a sort of a time-series plot showing the distribution of the landings throughout the entire timescale of the dataset. Not surprisingly, most of the landings occurred after the 19th century, which can be explained by the restricted technological advances prior to that time, where only landings that were actually observed by humans were recorded.
The third plot shows the mass of the meteorites that landed throughout the timescale. Here are some conclusions:

- Accurate landing records really started in the 19th century;
- Recent meteorites seem to be heavier – which is probably due to the drastic increase of the landing records recently. More records => more meteorite landings => potentially heavier meteorites.

2.3. Ploty Library

2.3.1. Scatter Plot

I opted for the Ploty Library for more esthetic plots. It's a bit slower and requires an API-Key to run adequately but it provides a variety of options that cannot be found elsewhere. The scatter plot shows the distribution of the top ten meteorite classes. We can learn, for example, that the L6 class is the oldest class observed and that the H4/5 class is contemporary.

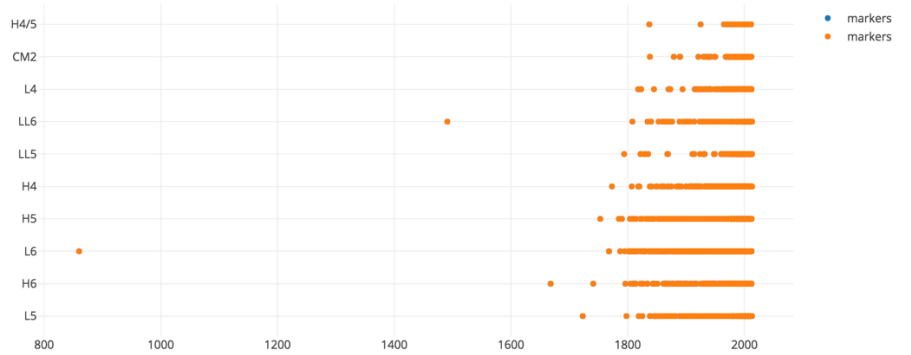
2.3.2. Scatter Map Box

This is, I believe, the most interesting plot provided by the Ploty Library. I tried to plot the geo-location of about 35.000 the meteorite on the dataset, based on the latitude and the longitude of the places where they landed. Surprisingly, only a minority of the meteorites landed in the seas/oceans, despite them constituting around 70% of the planet.

Note

Because the library required an API-Key to run, I couldn't share the plots publicly on the Jupyter Notebook. Screenshots of the Scatter Plots and the Scatter Map Box can be found below:

Out[112]:



EDIT CHART

In [125]:
```
#Same scatter plot but with the Plotly Library - it's, I believe, more esthetic.
data = [go.Scatter(x=df.year,
           y=df.mass, mode = 'markers'

             )]
py.iplot(data, filename='jupyter-basic_bar')
```