Chloe Wohlgemuth ASTR 200 S19 Prof. Follette Final Project

Final Project: Orbital Properties of Comets in Our Solar System

BACKGROUND:

I was interested in learning about smaller-body objects. Comets or asteroids caught my interest, after some exploring some astronomy scientific papers, as I find the motion and interactions between astronomical objects to be fascinating. There is a lot of phenomena and behaviors of such bodies that we never really talked about in class. We mainly focused on planets and stars. However, for this project, I wanted to explore the smaller bodies.

Comets are celestial objects consisting of a nucleus formed from ice and dust. When near the sun, a "tail" of gas and dust particles pointing away from the sun are produced. Comets are primitive bodies left over from the formation of our--or others--solar system. They were among the first solid bodies to form in the solar nebula. The solar nebula was a cloud or collection of dust and gas that is believed to have been what the sun, planets, moons, etc. were formed from around 4.5 billion years ago. Comets are likely "the earliest record of material from the nebula." [1] Consequently, comets are of particular interest to astronomers and scientists researching fields or theories related to the formation of our solar system.

I arrived at the JPL Small-Body Database while exploring the JPL Solar System Dynamics Database, which was listed as a Suggested Database Link in our Prelab 9. Upon opening the site, the page "Physical Data & Dynamical Constants" popped up. Under the sub-heading of "Comets and Asteroids," the text included a link to their small-body database search engine.[2] This database stores orbital and physical parameters for small-bodies, specifically asteroids and comets. It contains over hundreds of thousands of data values. With the JPL Small-Body Database Search Engine, I was able to filter the data based on my desired constraints. Though the database lacked some the physical properties of comets, i.e. mass or density or material composition, that I was thinking of exploring, there was substantial data on orbital properties.

The data set that I exported from this database included orbital properties of comets. I constrained the included comets to those that have recent times of perihelion passages and that have orbital paths contained loosely within the outer bounds of the Kuiper Belt, corresponding to semi-major axes less than 60 au and greater than 30 au. The Kuiper Belt is a region of our Solar System outside Neptune's orbit that contains many small bodies, including comets, made largely of ice. An "au" is an astronomical unit, or the average distance between the Earth's and sun's centers (≈149.6 million kilometers).

At the perihelion, where the object is closest to the sun (or other central object it is orbiting around), the comets are closer to us. Consequently, they have a much larger proportion of their orbit within an area/range that we can detect, and thus acquire information more readily and reliably. Comets that have more of their orbital paths somewhat close to Earth and Sun are more easily detectable. Hence, I constrained the comets that were included in the dataset by their time of perihelion passage [**TP**], the most recent time that they have passed through their perihelion, and semi-major axis, half of the longest diameter of the orbit. This reduces mis-representation and selection biases from comets largely outside of our detection limits.

Filtering the data produced 50 rows. Each row of the data set corresponded to a comet in the data set. Each column contained the data entries of a given measurement. In addition to the

---

[1] https://stardust.jpl.nasa.gov/comets/comets.html
[2] https://ssd.jpl.nasa.gov/sbdb_query.cgi

designated name for each comet, the columns include the following measurements: eccentricity, semi-major axis (units: astronomical units [au]), inclination (units: degrees), time of perihelion passage (units: Barycentric Dynamical Time or TDB). TDB is a relativistic coordinate time scale, in which larger values correspond to more recent or current dates for the event. The uncertainties, estimated differences/range between the measured and true values, in each of these 4 properties were also exported. Each of these four properties are defined and explained below.

Orbital eccentricity is a Likert scale for how much an orbit deviated from a perfect circle, where 0 represents a perfectly circular orbit, 1 represents a parabolic orbit, and >1 represents a hyperbolic orbit. Comets can be categorized by such shapes of their orbits. The major classifications are depicted in the figure below.

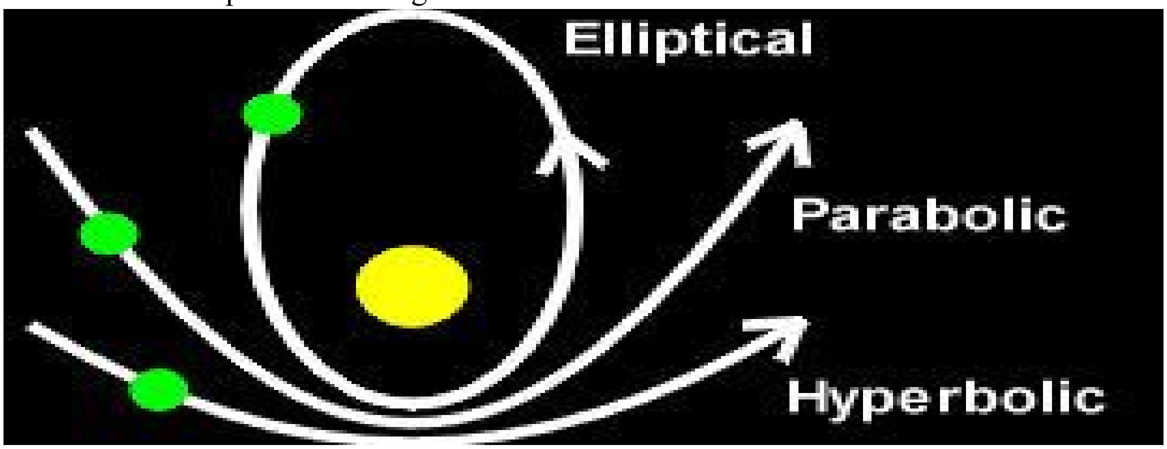

FIGURE 1: Orbit Shapes. This figure displays three general orbit shapes for comets. The white lines indicate the orbit path and shape. Arrows indicate direction along the path. Green circles represent an example comet. The yellow circle represents the object that the comet orbits about. (Source: www.cometwatch.co.uk/comet-info/types-of-comet-2/)

Parabolic comets have just enough-, Hyperbolic comets have more than enough-, and elliptical orbit comets have less than enough- escape velocity than is needed to escape the gravitational influence of the object they are orbiting. This is asserted in Kepler's first law that satellites, objects in orbit, have elliptical orbits with the object being orbited−at one focus. Comets within my chosen dataset are all periodic and elliptical.

Within their orbits, properties related to positions or locations of comets are often characterized by measures of time and distance. Explanations of such orbital properties are best accompanied by a visual diagram/illustration for reference.

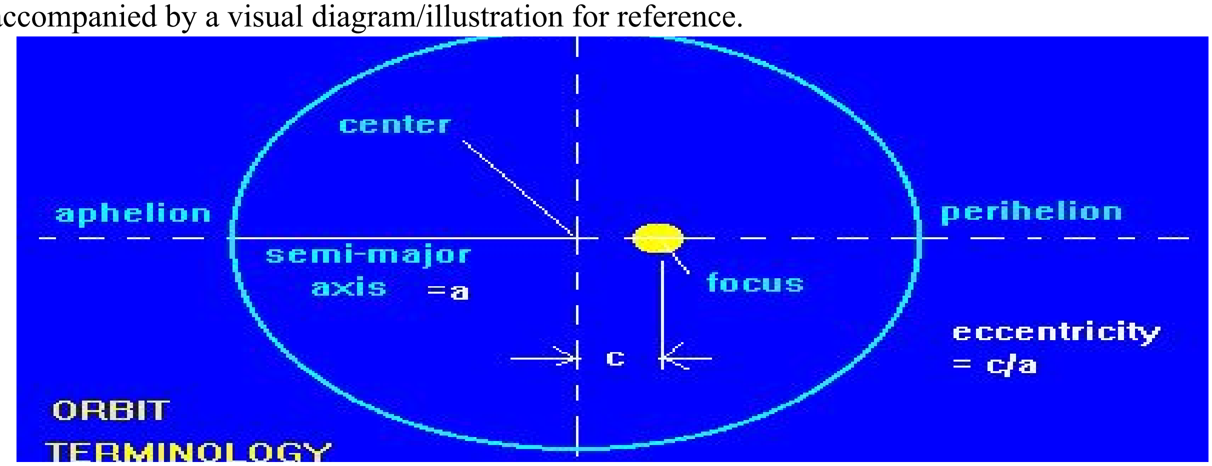

FIGURE 2: Orbit Terminology.This figure identifies and maps key locations and measurements for elliptical (and circular) orbits. The yellow object indicates the sun or object being orbited, and the light blue ellipse represents the orbital path of the object (in this investigation, a comet) orbiting the sun. (Source: http://www.bogan.ca/orbits/geometry.html)

At its perihelion, a comet is closest to the sun (or other orbited object). This is also around where the comet would be closest to us, since the relative distance between Earth and the sun is negligible in relation to the distance between us and the comet. They have a much larger proportion of their orbit within an area/range that we can detect and thus collect information more readily and reliably.

Comets with smaller semi-major axes, half of the longest diameter of the orbit, have shorter orbital periods. An object's orbital period is the time it takes to complete one orbit. For an elliptically orbiting object, Kepler's third law directly relates these two variables: $period^2 = a^2$. As a consequence of this phenomena, short period comets primarily have more circular orbits. This is expected, since a shorter period tends to mean a shorter path that was traveled; due to conservation of energy and momentum.

The orbital inclination of an object measures the tilt of its orbit around a celestial body, in our case the Sun. For an analogy, the orbit is on or part of a 2-dimensional plane. While all orbiting around the same central object, not all the orbiting objects need to have orbits on the same planes. They can be at an angle, or inclination, from the aforementioned "ecliptic" plane.

The periodic motion of orbits allows astronomers to study current phenomena to understand those of the past. In our solar system, all planets occupy the same orbital plane. This is not always the case for other planet-star systems, or for other orbiting bodies/systems. Explanations and theories about what happened during the formation of our solar system, as well as other systems, are proposed, formulated, and explored extensively by astronomers. Studying the orbits of comets within our solar system provides insight about the early conditions and formation of our solar system. For instance, our system's planets occupy the same orbital plane, while many other systems' planets do not. Comet's orbits could provide insight about the early conditions and formation of our solar system, and how they may differ from others.

<u>DATA EXPLORATION and PROCEDURE</u>:

Through using the JPL Small-Body Database Search Engine, I filtered the database entries for Comet objects with the previously explained constraints on time of perihelion passages and semi-major axes. The database allows the user to select and specify properties and constraints to include, and then has two button options for exporting the data. These two options, after the filtering options, included an HTML table or a csv file. I downloaded the data as a csv file named "jpl.csv". This format made it easy to read into a Jupyter notebook as a workable form for Python's Pandas library. I called its built-in function, pandas.read_csv('jpl.csv'), to convert csv files into parseable and manipulatable tables called dataframes.

In my initial exploration of the dataset, I used scatterplots, histograms, and boxplots to investigate the distributions. I first plotted the histograms, using plt.hist(). of each of the 4 variables of interest– semi-major axis, eccentricity, time of perihelion passage, and inclination.

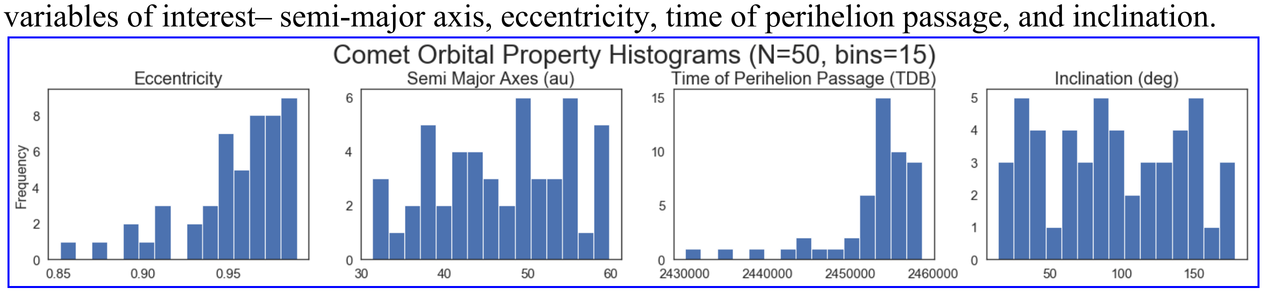

FIGURE 3: Orbital Property Distributions. This figure shows histograms of the 4 orbital properties. The x-axis represents the data values−for the variable given above each plot−and the y-axis plots the count frequencies.

The histograms helped illustrate the distributions of the data. The term skew reflects the extent to which a distribution differs from a normal distribution.[3] The eccentricity data ranged from 0.85- 0.99 and was moderately skewed left, meaning that more of the data has higher eccentricity values. Similarly, the perihelion passages times were strongly left-skewed shape. The large majority of the comets passed through their perihelion more recently; as was expected, since the data was constrained to recent TPs to reduce selection biases. Inclination appeared somewhat uniform. Semi-major axis was slightly uniform or bi-modal. Their shapes cannot be clearly determined, perhaps due to the small sample size.

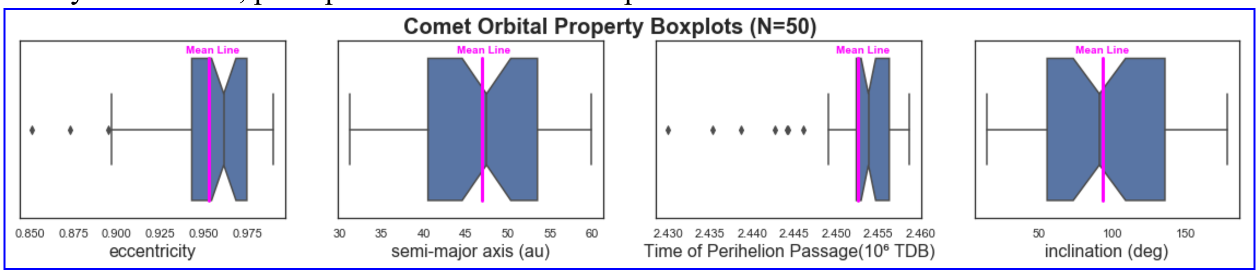

FIGURE 4: Orbital Property Boxplots. These boxplots were produced using seaborn.boxplot(). The x-axis represents the variable for which the values are plotted. The notches are confidence intervals on the medians, which was done instead of on means due to skewness and asymmetry in the distributions. For reference, fuchsia lines have been placed–and annotated using plt.text()–to indicate the mean values for each distribution.

I constructed boxplots for each of the variables in order to identify potential outliers in the dataset. Confidence intervals are ranges of values defined so there is a specified probability that the value of a parameter lies within it.Of the variables–e, a, and i–only eccentricity had outliers. However, all of these were lower-value outliers within ≈0.05 of the mean and median. Since eccentricity is a 0 to 1 likert scale, this difference can be considered negligible.

Throughout my exploration, I did not consider Time of Perihelion Passage [TP] when removing outliers or scatter plotting. TP was plotted *only* with semi-major axis. I plotted it with semimajor axis to verify if the data was spread uniformly across the data set, which was constrained to higher TP values.

I scatter plotted to see if there were interesting trends or relationships.

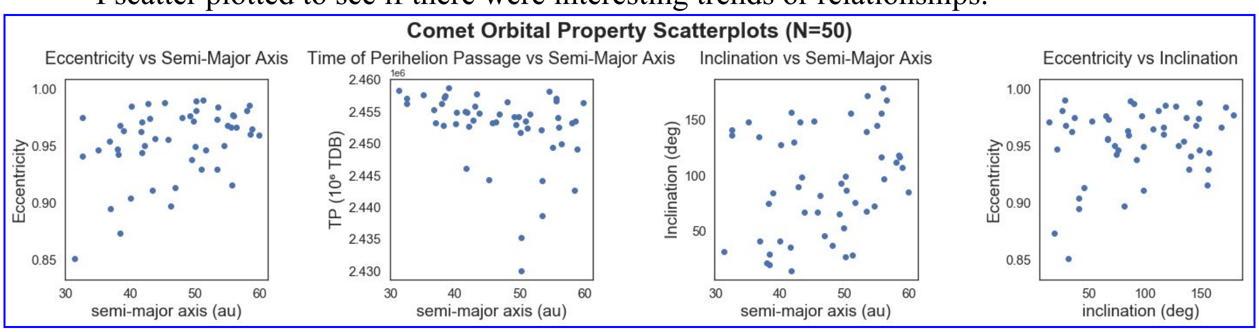

FIGURE 5: Orbital Property Scatterplots. Scatterplots of variables: e vs a, TP vs a, i vs a, and e vs i–respectively. Units are given for each axes–and have been defined previously. Produced using plt.scatter().

The data appeared to be evenly spread across the range of semi-major axes, and are all mostly located at the larger time of perihelion passage values. This was expected, since the data set was constrained to comets with recent observed perihelion passages–hence with larger perihelion passage times. Though not necessary for the investigation, this plot verified that the comets' semi-major axes data was spread across higher TP values.

[3] Normal distribution: the distribution of many random variables as a symmetrical bell-shaped graph

Scatter plotting each pairing of orbital properties revealed potential trends and correlations. A correlation is a -1 to 1 scale measuring how variables fluctuate in parallel if positive, inversely if negative, or not at all if zero. Eccentricity seems to have a moderately positive correlation with semi-major axis, though there is more data for the higher eccentricity values than lower values. The inclination vs semi-major axis plot reveals a somewhat random (thus un-correlated) scatter or very slight positive correlation between the two variables. The eccentricity vs inclination plot indicates no correlation between the two variables, with the randomly spread points; similar to TP vs a.

As mentioned in the Background Section, the database lacked sufficient data on physical properties of comet. Thus, I refined my topic of focus to orbital properties. Kepler III directly related $period^2$ to a^3. Consequently, I only included one of them in my data set. The comets that pass through our solar system more often have their semi-major axes within distances that we can detect and observe than their orbital periods. As shown in Figure 2, the semi-major axis is the center to farthest-radius distance. This will usually be about half the amount of distance needed to be detected and observed as orbital period. Hence, I chose semi-major axis. Outside of Kepler III and $P^2=a^3$, my research on eccentricity, inclination, and semi-major axis did not yield any definitively proven dependence.

I wanted to explore how comets'−orbiting in our solar system−orbital properties relate to one another as well as vary with one another. The two graphs with semi-major axis−with inclination and eccentricity−displayed possible correlations or relationships. Additionally, semi-major axis had a relatively uniform or slightly bimodal distribution. Consequently, after creating and analyzing my exploratory plots, I became interested how the other orbital properties possibly vary with and possibly relate to semi-major axis. Specifically, I wanted to see if short and longer period comets have different behaviors/properties. I created and used the function filter_periods() to store "short" or "long" under a new categorical variable of period length; based on which side of 45 (the dividing value) their semi-major axis was.

I chose the main question for this investigation: ***How do inclination and eccentricity relate and/or vary with semi-major axis (and thus orbital period)***?

<u>DISCUSSION and ANALYSIS</u>:

In attempts to exhaust what information I could obtain from the plots, I considered if log-scaling any of the axes would be insightful for scatter plotting. However, this was a dead end since none of the data points were clustered within small ranges, with the exception of TP. Consequently, log scaling provided no additional insight or improvements.

I first sought to establish the possible groupings present. I then wanted to compare them and see if they were significantly different. In addition, I also wanted to determine how well each could be modeled and predicted by regression

Based on the distributions from my initial exploration of the data, I tried dividing eccentricity and semi-major axis according to their respective breaks. These breaks were visually identified on the histograms, from which I then manipulated the bins and axes and checked by displaying the dataFrame in order to find the range or value that divided the sides of the data.

Eccentricity contained a break at e=0.916 to e=0.929 in the histogram. This is possibly supported by the (leftmost) scatter plot in Figure 5. The data above 0.92 have slight to no correlation and the data below 0.92 have a stronger positive correlation. When I split the dataset, however, it only yielded two roughly uniform/flat scatters both with semi-major axis and with inclination. Exploring orbital property relationships with eccentricity as the independent variable resulted in a dead end. After doing the same for semi-major axis and observing what each variable produced, I chose to use semi-major axis as the primary independent variable. It had more even and workable sub-sets. Whereas separating by eccentricity resulted in N=8 and N=42 sample sizes, dividing the semi-major axis based on the break at ≈46 resulted in N=21 for short period comets and N=29 for long period comets. Though still small, this was more sufficient for comparisons and testing. I graphed side-by-side boxplots in order to compare the categories of "short" vs "long" periods.

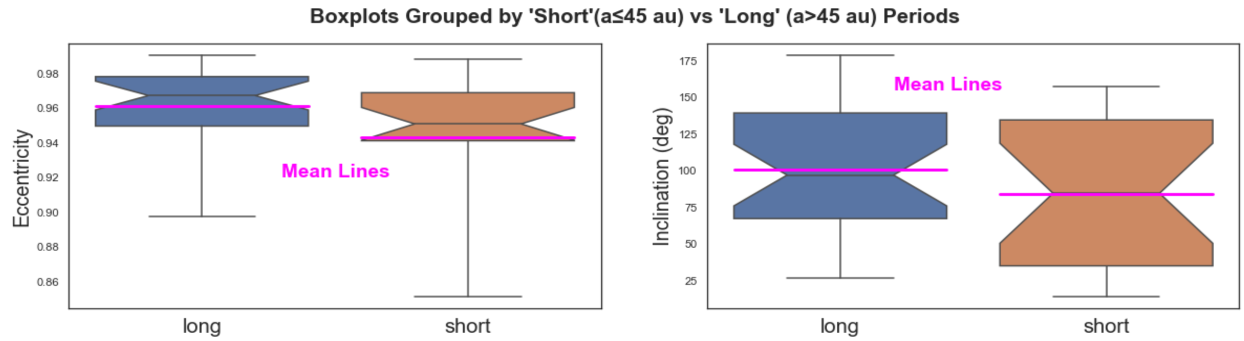

FIGURE 6: Boxplots grouped by semi-major axis. These boxplots were produced using seaborn.boxplot(). x-axes are the groups: "long" and "short semi-major axis, and hence the directly related orbital period. The notches are confidence intervals on the medians. Fuchsia lines were placed—and annotated using plt.text()—for means.

The boxplot for Inclination shows how the medians of long and short periods are very close, and within the confidence interval of one another. Their spreads and shapes are also similar, though the shorter period comets have a more symmetrical shape. The eccentricity boxplot does exhibit different centers and spreads between the short and long period groups. Though both are prominently left-skewed, the short period comets have a larger range and lower center or median (as well as mean); their means and medians are not contained within one's interval; though the ends of the intervals do appear to barely overlap.

I checked these visual estimations and inferences with np.nanmean() and np.nanmedian() for the measures of center (mean and median), and confidence intervals with stats.t.interval().

```
        SHORT period - ECCENTRICITY
Mean is:         0.94304
Median is:       0.95049
St. dev. is:     0.03558
95% conf interval:      (0.9264395260163022, 0.9596327596979836)
99% conf interval:      (0.9203976766811934, 0.9656746090330923)
        LONG period - ECCENTRICITY
Mean is:         0.96081
Median is:       0.96681
St. dev. is:     0.02429
95% conf interval:      (0.9514048027112321, 0.9702085765991124)
99% conf interval:      (0.9481237221568699, 0.9734896571534746)
```

```
        SHORT period - INCLINATION
Mean:   84.03192
Median is:      84.34110
St. dev. is:    49.65811
95% conf interval:      (60.86960750360399, 107.19423249639601)
99% conf interval:      (52.43757562036123, 115.62626437963877)
        LONG period - INCLINATION
The mean is:    100.36894
The median is:  96.75817
The st. dev. is:        43.40470
95% conf interval:      (83.56643457258436, 117.17144749638119)
99% conf interval:      (77.70267790911875, 123.0352041598468)
```

FIGURE 7 Confidence Intervals and Statistics. This figure displays the output for several relevant descriptive statistics: mean and median (measures of center) and standard deviation. The 95% and 99% confidence intervals were also calculated and displayed.

Eccentricity has overlapping confidence intervals at 95% confidence level: (0.9264, 0.9596) for Short and (0.9514, 0.9702) for Long. Inclination has overlapping confidence intervals at 95% confidence level: (60.8696, 107.1942) for Short and (83.5664, 117.1714) for Long.

Though both eccentricity and inclination have different means and medians between short and long period comets, the difference between them is within a standard deviation of each value; hence the statistical strength of the difference is questionable. Eccentricity has a median of 0.9505 with σ=0.0356 for short and 0.9668 with σ=0.0243 for long period comets. Neither of these values are outside of one standard deviation of the other. This is the same with inclination as well, in which the median is 84.3411 with σ=49.658 for short and 96.7582 with σ=43.405 for long period comets. The lack of difference could also be a consequence of the constrained range or spread of the data, due my restricting all semi-major axes within 30-60 au.

I realized that my data set was not giving very conclusive results with regards to how the orbital properties' distributions vary with one another. I planned to come back to such tests and comparisons with a Monte Carlo simulation approach; once I had completed the relational and model fitting processes with the original data.

I created scatterplots from the original data, with error bars and linear regressions.

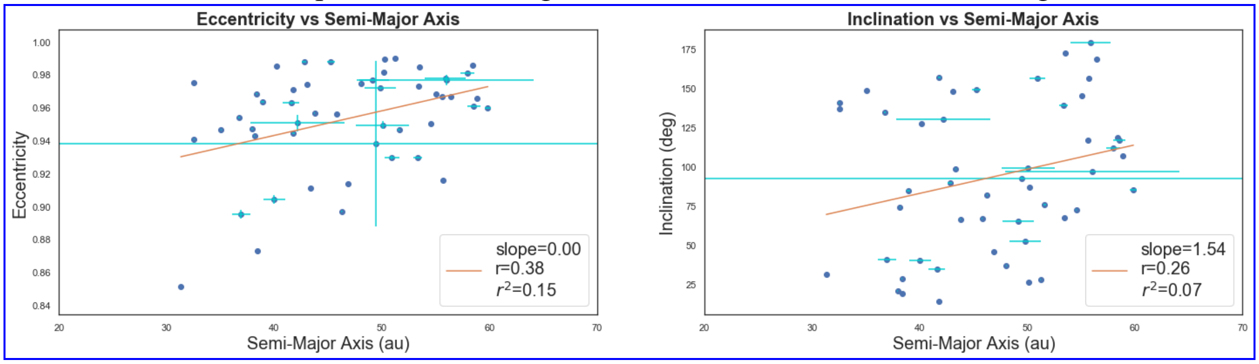

FIGURE 8: Original Data Scatter Plot Fits. This figure displays the plots for e vs a and i vs a, respectively. Linear regression was conducted for each using stats.linregress() and is graphed as an orange line. The turquoise error bars are the uncertainties (in standard deviations) of the values as provided in in the dataset. The "r" correlation coefficient and "r^2" quality of fit metric were both calculated and included in the legends as well.

The scatterplots related/plotted eccentricity and inclination each against semi-major axis. The correlation coefficients between the variables in both cases reflects slight positive correlations. R-squared, r^2, is a goodness-of-fit measure for how well the computed linear regression models fit the data. It indicates the percentage of variance in the dependent variable that the independent variables explain collectively. In both cases, the r^2 values were relatively low. They imply that semi-major axis explained only 38% of the variance in eccentricity and 7% of the variance in inclination.

I could not conduct a two sample t-test, as I originally had planned to, because the data set did not meet the requirements. The data set met the conditions of: continuous dependent variable without outliers (eccentricity and inclination both are continuous and do not have outliers), and independent observations (data consists of different and independent comets). However, the dependent variables, eccentricity and inclination, are not normally distributed and therefore could not meet the requirements for a 2-sample t test for difference in means.

Instead, I conducted a 2 sample Mood's Median test, a non-parametric test to determine whether the median of two independent samples are equal. I tested for the difference of eccentricity medians between 'Short'($a \leq 45$ au) and 'Long' ($a > 45$ au) period comets passing through our solar system. Using stats.median_test(), the following results were produced:

```
Mood's Median Test for difference in ECCENTRICITY medians
        test statistic =          1.3136288998357966
        p-value =                 0.2517386729294735
```

```
Mood's Median Test for difference in INCLINATION medians
        test statistic =          0.32840722495894914
        p-value =                 0.5665984802832496
```

FIGURE 9: Mood's Median Hypothesis Test. This figure provides the outputted results for the Mood's Median tests conducted for eccentricity and inclination of short vs long period comets. H_0 = No significant difference in eccentricity (left) and inclination (right) medians of 'short' ($a \leq 45$ au) and 'Long' ($a > 45$ au) period comets.

Since the p-values for both eccentricity, p=0.252, and inclination, p=0.567, are greater than the significance level (0.05), we CANNOT reject the null hypothesis of equal Eccentricity medians. Hence, we are unable to make a definitive conclusion about the difference in median inclination and median eccentricity between short and long semi-major axis comets.

Overall, my investigation of the single sample dataset was not very conclusive. It was not especially informative either, beyond the stated slight correlations and trends.

I used Monte Carlo simulation to repeatedly draw samples from the original data set. I did this for two main reasons. The first was so that the conditions related to sample size and shape for modeling and hypothesis tests on the data For instance, large enough sample size to satisfy the central limit theorem and have a normal distribution shape. The second reason was so that I could obtain distributions of Eccentricity vs Inclination; such as slopes, correlations, and the r^2 quality of fit metric. Taking the slope for instance: Monte carlo was necessary for this because I would not have the distribution of slopes without it. I would have only a single slope value, or insufficiently representative fractions of the original data. Simulating numerous samples that are sufficient in size for plotting allowed me to estimate the distributions.

I create a Monte Carlo simulation function called dataSim() that has the optional parameter of "points" that defaults to 30, as well as 3 required parameters. The function requires two specified variables of data and an "n" number of realizations or samples that is will simulate. It returns 3 arrays, which store the slopes, the correlation coefficients, and the simulated samples data. This simulation was used in order to produce multiple samples from the original data set "population." Therefore, the set of data values for which the simulated samples would be based off of were necessary. The covariance and mean of these values would be used to generate random samples based on the original data. The sample size and number of simulated realizations were necessary so that the correct desired number of samples could be produced, with the correct number of values included in each sample.

I simulated 1,000 realizations of 30 points each. I wanted a distribution with a large number of samples that could fulfill the Central Limit Theorem [CLT]. This theorem states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population. Thus, the previously failed condition of normality and/or large enough samples size would be fulfilled for testing and graphical analysis. I ran the simulation for each of the variables with semi-major axis. I then created and ran a function to plot the simulated samples, plotSamples(), which displayed only the first 8 of each simulation so as to avoid graphical clutter and redundancy. I provided part of an example row of plotSamples() display below.

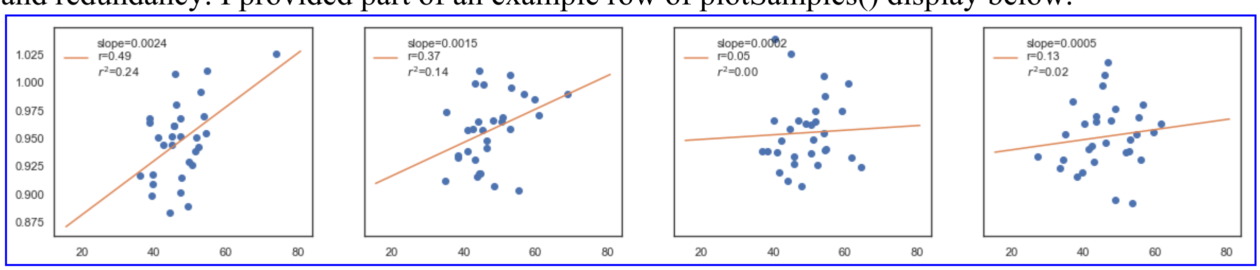

FIGURE 10: Plotted Realizations. This is one sub-plot row of the output for plotSamples(), of a simulation of 1000 realizations of 30 points each, for Eccentricity vs Semi-Major axis. The x and y axes are the independent and dependent variables, respectively; in this case, eccentricity and semi-major axis (au).

I printed the average r^2 for each simulation as well, using np.mean(np.square(r)); where r is the corresponding correlation array returned from each simulation.

```
                SIMULATING 1000 RANDOM 30 POINT REALIZAITIONS
e vs a realizations (N=30 each):        Average R-squared = 0.07609740293336141
TP vs a realizations (N=30 each):       Average R-squared = 0.000688941847750761
i vs a realizations: (N=30 each):       Average R-squared = 0.11565174384163066
```

FIGURE 11: Quality of Fit Metrics. The average R-squared value, a quality of fit metric, for the realizations.

From the plots and the R^2 values, it is evident that there is not a significant amount of the variance in plotted data. The semi-major axis only explains 7.61% of variation in the eccentricity data and only 11.57% of variation in the inclination data. The R^2 values imply that a linear regression model is not of good quality for modeling the data. Hence, it is unlikely that either of eccentricity or inclination would be in a linear relationship with semi-major axis.

The various tests and searches for relationships and variations between orbital properties was largely inconclusive. There is a visible and understandable positive trend between semi-major axis and eccentricity, but it was not statistically significant. The same went for the measures of centers between the shorter vs longer period comet groups; no significant difference could be determined. Even so, I learned a lot from this project. The ability to search for, parse through, and manipulate data was necessary to devise the research question and then conduct an exploration to try and answer it. All of this was largely self-driven. Especially when dealing with large, overwhelming and unfamiliar databases, all of these tasks/abilities were imperative. Additionally, being able to devise, execute, and connect different approaches to manipulating data for the investigation was something that I struggled a bit with. Even though there were so many hypothesis tests, grouping and plotting options, choosing the correct one and then figuring out where to go from there was not as straightforward as I originally thought. When one part of my plan was inconclusive, I had to adjust my exploration and think of a different approach to try. Even though the procedures and tests conducted in this project were inconclusive, I learned a lot.

If I were to continue or build on this project, I would be very interested in investigating more orbital properties and with larger ranges. I had constrained my data set a lot, bringing it from N=3000 to N=50, to reduce selection bias. However, if different properties were available and there was enough data to explore, I would like to investigate them. Another route would be how orbital shapes relate or vary with other orbital properties. For instance, there are comet classes for Hyperbolic orbits and Parabolic orbits. A study exploring physical, orbital, or some other properties of comets (or asteroid or others!) would be interesting.

<u>TUFTE APPENDIX</u>: Aspects of the graphics that

Aspects of the graphics that…

FIGURE	***Adherences***	***Violations***
2	• A lot of info in little space → visually maps and defines (w/o words) orbital locations and terminology in a single diagram. • Purpose: map out orit properties/terms • All items on graph are relevant.	• Not multivariate (only has position/locations) • Lacks detailed labels. Terms are placed without descriptions or any written supplement • Data Ink and aesthetic appeal. Same color with different blinding shades used for background and diagram. Background is distracting
3 & 4	• Large quantity of data in small space coherently with consistent dimensions • Purpose (illustrate/describe distribution) • Multivariate: frequency or counts and corresponding variable measurements • Small multiples allow for easy comparisons of distribution shapes • Data-ink ratio - all ink on plot is for data; no clutter or unnecessary empty space	• Tick-marks → not always known what location on bars/whiskers corresponds to what x-axis value • Forcing multiples layout makes the distributions with small spreads (ie. IQRs with the boxplots) less readable

5	• Large quantity of data in small space coherently with consistent dimensions • Multivariate: independent and dependent variables are plotted • Small multiples allow for easy comparisons • Recognition of trends and patterns	• In forcing larger tick labels for legibility, loses clarity for where exactly on the axis a value corresponds to • 4 separate plots might not be necessary ; could have the three semi-major axis graphs on same plot, different colors. Might improve trend and comparison recognitions, and optimize space more
6	• Large quantity of data in small space coherently with consistent dimensions • Clear purpose (illustrate distributions for comparison between groups) • Multivariate: frequency or counts and corresponding variable measurements • Small multiples allow for easy comparisons of distribution shapes • Data-ink ratio - all ink on plot is for data; no clutter or unnecessary empty space • Adjacent placement with grouping → ease of comparison between groups and sub plots	• Tick-marks → not always known what location on bars/whiskers corresponds to what x-axis value • Labeling of color coding (i.e. in legend) to make graphic more clear and quickly informative • Color palette - might not be the friendliest to color sensitive of blind people; the pink can be faded by the dark blue of the box and whiskers.
8	• Large quantity of data in small space coherently with consistent dimensions • Labeling and use of Color to distinguish elements on graph from one another • Data-ink ratio - all ink on plot is for data; no clutter or unnecessary empty space • multi-variate	• Color pallette choice - not friendly for those without access to such colors, blues should not be repeated when have other options • Clutter / clarity: the error bars sometimes lay over the other points, which draws focus away from data. Should have them be a less highlighted color • Distortion/representation: error bars force x-range to be larger, and thus squish the data points into small portion of x-axis. • Labeling: sizing of ticks, markers, and labels should be bigger for easier viewing
10	• Large quantity of data in small space coherently with consistent dimensions • Multivariate: independent and dependent variable • Data at levels of detail: regression and correlation calculations give details about correlation and possibly how well fit to linear relationship • Small multiples allow for easy comparisons of distribution shapes • Data-ink ratio - all ink on plot is for data; no clutter or unnecessary empty space • Shared axes scales for ease of comparison between adjacent plots	• Lack of labels for axes; did not specify because differences depending on which plot is being displayed/regressed, and I did not have a variable to determine it • Labeling: sizing of ticks, markers, and labels should be bigger for easier viewing