

Astrological Analysis with Mutual Information

W

Aditi Chegu, Olivia Fang, Tejan Patel

Mentored by: Dan Shumow, Junaid Hasan

Astrology

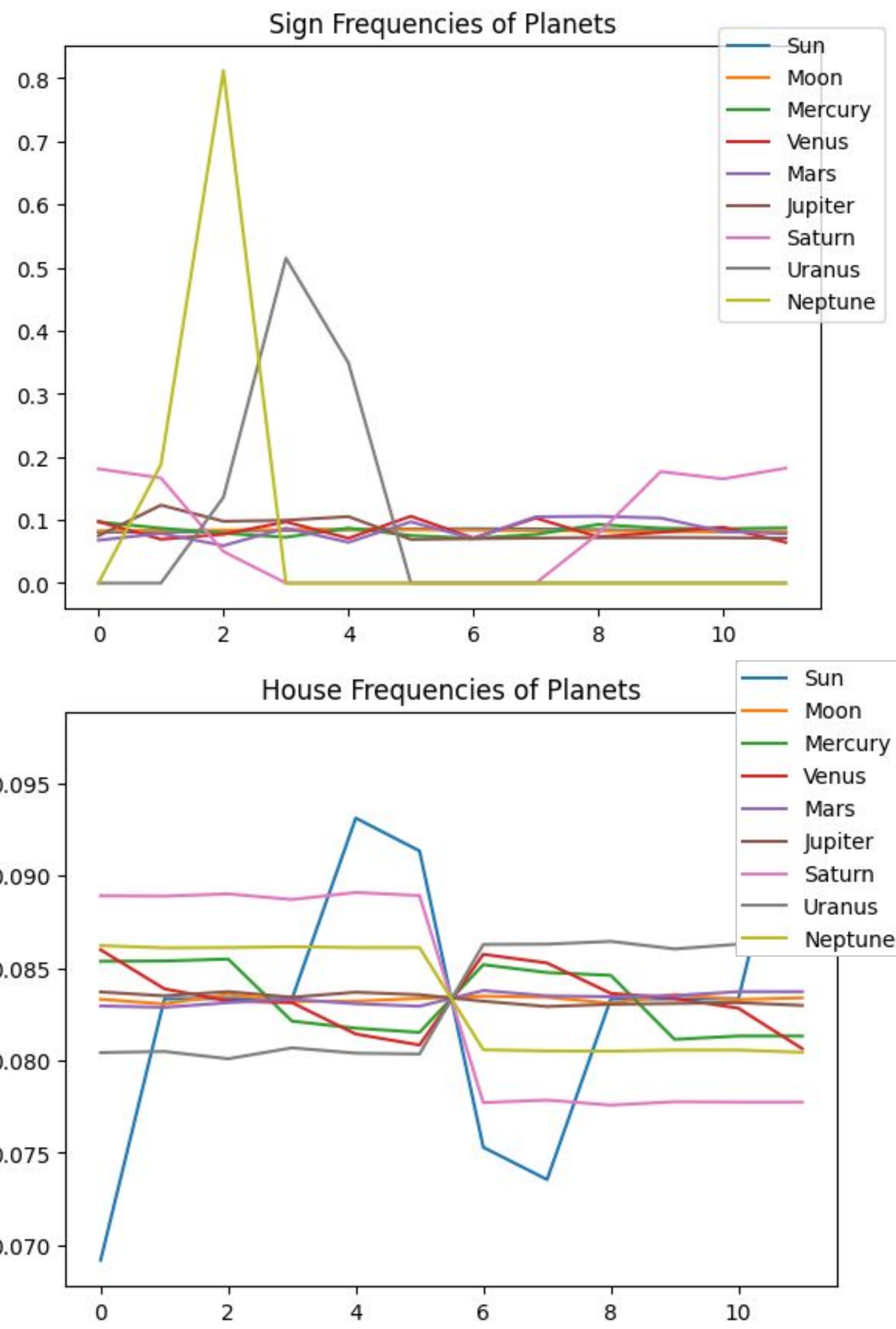
Astrology is a divination system in which people attempt to analyze an individual's personality and make predictions based on the location of celestial objects in the sky at the place and time of an individual's birth. Astrology classifies people based on their natal charts; this is the principal type of data we are analyzing. We are looking at astrological data, as it would be held by astrology apps, such as 'Co-Star,' or 'The Pattern.' The goal is to analyze whether or not individuals can expect to be uniquely identified by their astrological data.

Natal Charts

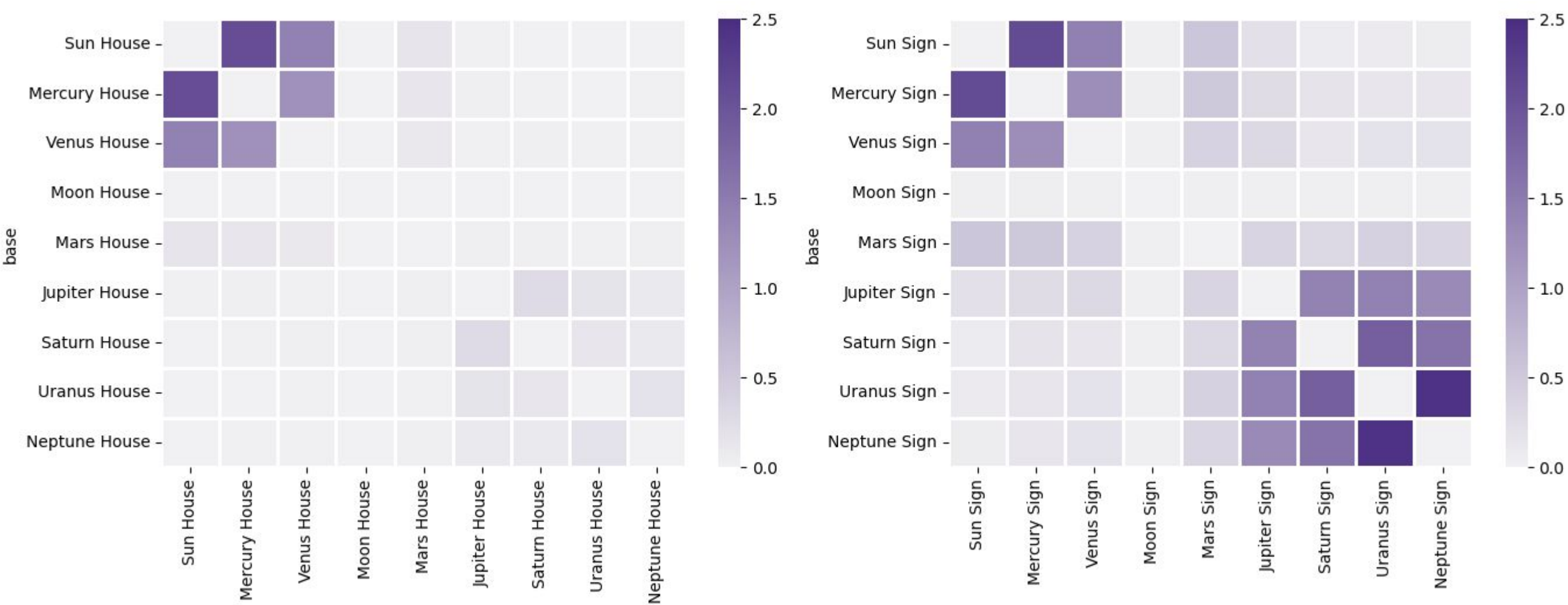
A natal chart (or birth charts) contains information about the location of all the planets on the exact time someone is born. The sky is divided into 12 regions that are assigned a house and sign. Each planet will be found in one of these regions and will then be assigned with the region's house and sign. Given the location of each of the planets we can construct a specific natal chart.

Sign vs. House

House and sign both specify the location of planets in a person's birth chart in different ways. The main difference between the two is that a planet's astrological sign at a given time is independent of location, while the house of a planet varies depending on location (on Earth). In addition to the planets, an object called the "ascendant" is represented by the eastern horizon and is also given a sign and house.



Observed Mutual Information



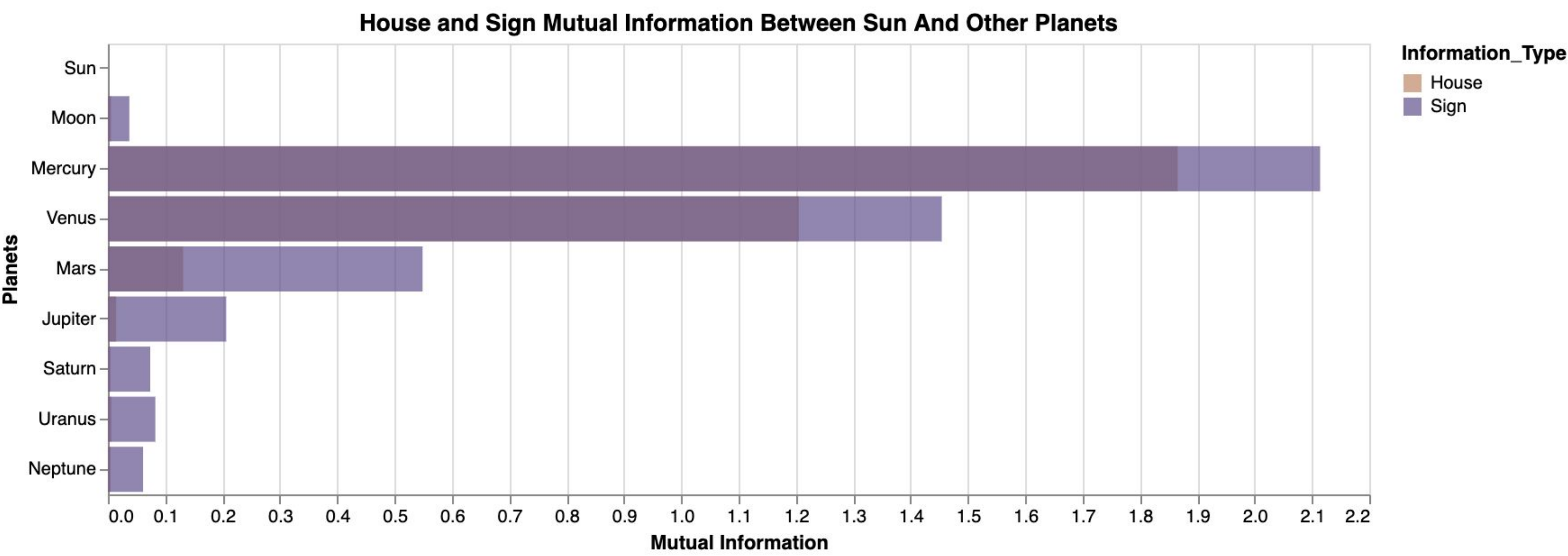
The procedure for generating this data is as follows: Over the last 200 years, we selected 25 different years uniformly at random. Within those 25 years, we selected 100 days and 12 sample times within each day, again both uniformly at random. From this, we had 30,000 uniform samples. We then generated a data frame containing the house and sign data for each celestial object.

From this, for each combination of two celestial objects, the data frame was filtered to generate a 12x12 probability grid and their respective signs or houses. Each probability within the matrix represented the first object was in a certain house and the second object was in a certain house. From each probability grid of two celestial objects, mutual information can be calculated. This led us to have two 9x9 data frames, one for house and one for sign, that contained the mutual information for every combination of planets.

Mutual Information and its Role

Mutual information is the measure of distance between two random variables. Informally, it measures how much information ones gets about a random variable after knowing the result of another random variable. For our project, we used mutual information to gauge the variability of one celestial object's house or sign when given a different object's house or sign.

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{(X,Y)}(x,y) \cdot \log \left(\frac{P_{(X,Y)}(x,y)}{p_X(x)p_Y(y)} \right)$$



Incorporating ML Models

We wanted to test the vulnerability of a database that was organised without differential privacy. That is to say, oblivious to any privacy preserving techniques. To do this we attempted to create a model that could predict the position of a planet given the positions of all the other planets.

Since we know the mutual information of planets in relation to one another, we can use this information to help us create a model. A high mutual information between planets (such as the sun and mercury) means that they are always within a certain, predictable distance of one another. We thought this would lend itself nicely to a linear regression model.

Hence, the model we currently have is acceptable for planets that have a high mutual information with other planets (the sun, mercury).



Takeaways

High mutual information between planetary signs/houses indicates that we need to consider more than just the entropy of each of these individual properties to preserve privacy. In our calculations, we see that the signs have more mutual information than the houses. Hence, the signs are more correlated. We can use this information to model a database that incorporates aspects of differential privacy to make it more secure. This could involve either limiting the number of queries a user can make or ordering the information while keeping the relative mutual information of each planet in mind.

Future Work

Now that we have a solid foundation of astrology and trends within astrological data, we would like to take an experimental approach to see how astrology apps such as Co-Star work. We hypothesize that astrology apps purchase user information, utilize machine learning algorithms, and output that information to the users. Essentially, we think astrology apps force feed users their own data using astrology. Our approach would be centered around generating fake accounts and attempting to bait the apps to output specific data that we gave them. From this, we will be able to make conclusions centered around how astrology apps work and if they actually use real astrological data.