Given the data in the attached files, determine which of Jose Altuve or Mike Trout was the faster baserunner in 2016. The attached files include player tracking information for balls hit into play during 2016 when the player was either the batter or the runner at first or second base. The definition of faster speed in this baseball context is up to you to determine and justify. Construct a chart or other visualization to communicate your findings. Please include in the form of a Jupyter, RMarkdown, or similar notebook your R or Python code used to analyze the data, produce the visualization.

The x, y coordinates of both Altuve and Trout with their respective times gives rise to the following definition of velocity, aka speed, being measured in feet per second:

$$Speed = \frac{\Delta distance}{\Delta time} = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{(t_2 - t_1)}$$

The following is an overview of the results found in R:

José Altuve   Mike Trout

José Altuve (feet/sec)	Statistic Type	Mike Trout (feet/sec)
43.38	Max Speed	45.33
13.35	Average Speed	13.17
13.68	Median	12.54
8.44	Standard Deviation of Speeds	8.75
13.31	Lower Confidence Interval	13.13
13.39	Upper Confidence Interval	13.21
20.62	Mode Speed > 7 ft/sec	29.23
379	Frequency of Mode Value	399

Histogram of Altuve's Speed   Histogram of Trout's Speed
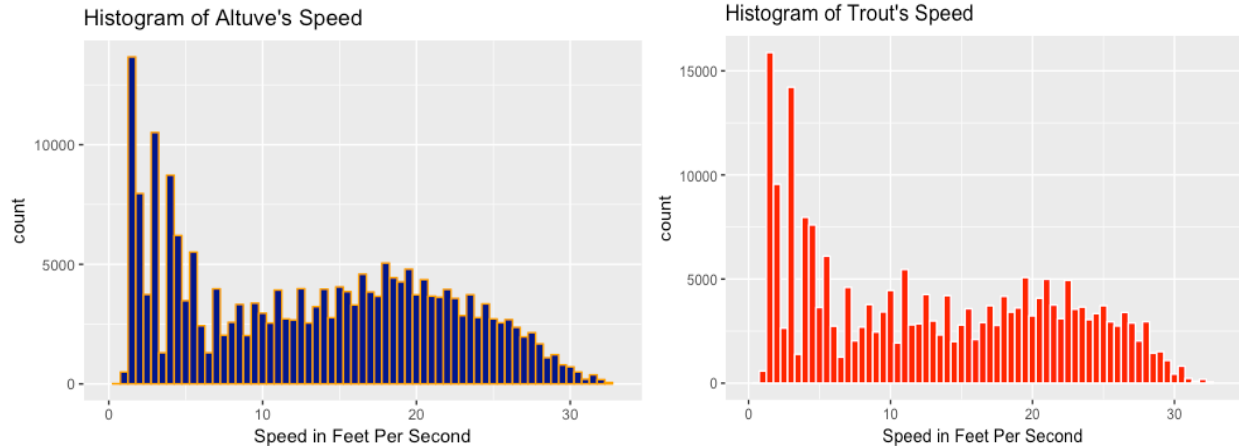
A two-sample t-test on the resulting calculated vectors of Altuve and Trout's speeds was run to ensure the means can be considered statistically different. The null hypothesis was that the mean speeds of both players are equal, and the alternative hypothesis was that the mean speeds of the players are not equal. The following results were obtained:

```
        Welch Two Sample t-test

data:  speed.altuve and speed.trout
t = 6.9606, df = 437650, p-value = 3.393e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1298809 0.2316928
sample estimates:
mean of x mean of y
 13.35050  13.16971
```

The p-value is much smaller than the alpha level of 0.05 at p-value = 3.39e-12. Therefore, we can reject the null hypothesis and state that the means of Altuve and Trout's speeds are indeed different. Therefore, we can continue with our comparison of their point estimates.

Based on these statistics and histograms, it is clear that Mike Trout's maximum and mode speeds are greater than that of José Altuve. However, the average speed, median speed and confidence intervals of Altuve's speed are higher. Additionally, the standard deviation of Altuve's speed is lower. These statistics show that although, Mike Trout was able to achieve a higher max and mode speed in 2016, José Altuve demonstrated a narrower spread of speeds, indicating that his speed is more consistent. You can also see this conclusion when observing Trout's histogram which shows a less normal distribution than that of Altuve. The spread of Trout's speeds more closely resembles the uniform distribution indicating that his probability of attaining varying speeds, such as an approximate 12 feet/second and 22 feet/second, were

roughly equal in 2016. Therefore, I would state that we can expect José Altuve to be on average, faster than Trout.

However, other confounding variables that were not controlled and could possibly affect speed should also be considered when reviewing these results. For example, the number of home runs a player hits can affect how likely they are to sprint to bases as it is unnecessary to run quickly if a clear home run is hit. This and other factors should be more closely examined in the final conclusions of speed analyses.