I examined the FatRats dataset, which measures the gain of what I assume to be a rat’s weight over a given time period by measuring against two variables. The two variables within the study include a high/low protein intake, and different sources of food (ex: beef). I first performed univariate discovery by grouping dataset by the Source variable. Overall, beef produced the highest mean gain. However, pork produced the higher median gain, along with the highest maximum value. This leads me to believe that although pork may have had some higher outliers, beef produced the highest consistent Gain since the mean was the highest. I also performed univariate discovery by grouping the dataset by the high/low variable. Overall, a high protein intake produced a larger gain.

Additional data visualization was performed to affirm my conclusions. Sure enough, when visualizing the Source (x-axis) and the Gain (y-axis), the pork source had longer tails than the beef source. This supports the finding of a larger spread and therefore a higher median. An additional finding showed that cereal appeared to be the most negatively skewed in comparison to the Beef and Pork sources. When comparing the Protein (x-axis) variable in comparison to Gain (y-axis), there was an outlier on the high Protein variable. Additional findings may be present if this individual outlier were to be examined in further detail.

Finally, the measures of central tendency, skewness, and kurtosis were examined. Measurements such as the trimmed mean, the winsorized mean, and the interquartile mean of the Gain were all very similar measurements at around 88. Although I was not given units of measurement for this study, I assume this to be calculated in grams. In comparison to the mean relative to the various sources, this is consistent with beef and pork, and about 4 grams higher than the mean Gain relative to cereal. The skewness of the gain was calculated to be -0.282497163714679, meaning the data is close to symmetrical relative to Pearson’s Second Coefficient. This also means the horizontal pull on the data is minimal. Kurtosis of the Gain was also measured and calculated to be a 2.55883965202515. This leaves the Excess Kurtosis with a value of 3-2.55883965202515=0.4411603. Overall, given that the kurtosis is close to 3, the vertical pull is minimal in the dataset, with a fairly normal peak in the data.