Bradley Higdon

9/8/20

Matplotlib Data Trends

The first major data trend that I can see when looking at the various charts and graphs would be a somewhat positive correlation between timepoint and tumor volume. The relationship is very consistent, as the shape of the line chart is fairly flat while rising steadily. Overall, the line chart shows that there is a high likelihood that tumor volume increases as the time point increases.

Looking at the bar and pie charts, the dataset appears to be fairly well rounded. The pie charts show that there is almost perfect gender parity in the treatments, minimizing the probability of an error in the data occurring should there be differences between how the drug regimens would affect different genders. In addition, the bar charts show that every drug regimen was treated thoroughly in sheer quantity. Propriva was noticeably a bit lower than the others in the amount of treatments conducted, but the variance was not high enough to be of much concern.

The summary statistics of a dataset can give a lot of quick information on both correlations and relationships in the data before looking at more targeted graphical information that isolates smaller sections of the data. Capomulin and Ramicane stood out for their low averages in particular, but no one stood out for having a particularly high average. The median category for tumor volume shows a similar story, with both Capomulin and Ramicane having noticeably lower numbers than the rest of the treatments. Both Capomulin and Ramicane show low variance as well, in a sharp contrast to others such as Ketapril and Naftisol, both of which had variance more than double those of Capomulin and Ramicane.