**e. Analysis of what this project helped you learn about data science, e.g., the exploration of data which is what you have been doing – 2 points. You must argue persuasively. Review the documentation for the packages and functions that you use.**

There are three major areas this project taught our team about data science to include the benefits of data visualization, principal component analysis, and importance of researching and leveraging well documented resources and packages. First, the project taught our team a lot about the effectiveness of data visualization throughout the project. We were able to use a variety of plotting and graphing libraries such as ggplot2 and GGAlly to explore our dataset, identify relationships amongst variables, and quickly characterize and interpret the distribution, frequency or trends within the dataset and across variables. We leveraged the ggpairs function to examine and compare our multivariate dataset in different dimensions and across subsets (such as edible and poisonous). We were also able to validate PCA results when using different approaches (and libraries such as prcomp and FactoMineR) to our PCA analyses by plotting and comparing our results. Leveraging data visualization enabled our team to effectively explore, analyze, and understand our dataset and project results.

Secondly, this project also taught us a lot about the use and benefits of the variable reduction technique, principal component analysis (PCA), in the data science realm of processes, methods, and approaches. After exploring and conducting our PCA analysis, we incorporated our resulting components into our various clustering methods and found improving results to include less parameters and computation incorporated in our analysis enabling us to analyze and understand our results more simply through visualization.

Lastly, this project taught us about the importance of researching and leveraging well documented resources and packages. We found that there is a wealth of resources, packages, and methods that can be done in data science especially in the open source world. What we found that despite the many roads and approaches to finding an answer or performing a particular analysis, what’s (also) important is ensuring each package and method used from a particular package is clear, well documented, and easy to understand. We found a variety of packages that could support us in this project effort but we wasted a lot of time researching packages that had missing information in their documentation or guides (such as FactoMineR). This is not only important for data science consumers of particular packages across academia and business but also important for the developers of libraries and packages.