**Data Visualization**

**Why Visualize Data?**

Visualizing the data obtained to complete your machine learning is a vital step. Using graphs or other software such as facets can help determine trends or correlations in the data to assist in the creation of the model. It can also help determine any outliers that may be caused due to missing or incorrect data. Correcting or removing these missing or inaccurate values will create more accurate data during the normalization process to aid in a more successful model.

**Facets**

**What is Facets?**

Google open-sourced Facets: a data visualization tool to explore data for machine learning scientists.

Facets aim is to make [big data sets understandable and interpretable](https://pair-code.github.io/facets/). Facets consists of two visualizations: Facets Overview and Facets Dive. Overview takes input feature data and analyses it feature by feature. Its goal is to give developers a quick understanding of the features in their dataset, their distribution, and unexpected values. Facets Dive is an interactive tool in which developers can bucket items in multiple dimensions based on their feature value.[[1]](#endnote-1)

**Using Facets**

Facets may be accessed directly through <https://pair-code.github.io/facets/>. The page has a try it out feature that then allows the user to upload a CSV file of their data into either the overview or dive visualization. They may then manipulate the data within these two visualizations as they please. Additionally, the user may create code that displays their data in an html page in either the dive or overview option. For the purposes of this project the dive option was selected.

1. <https://www.infoq.com/news/2017/07/facets-big-data-visualisation> [↑](#endnote-ref-1)