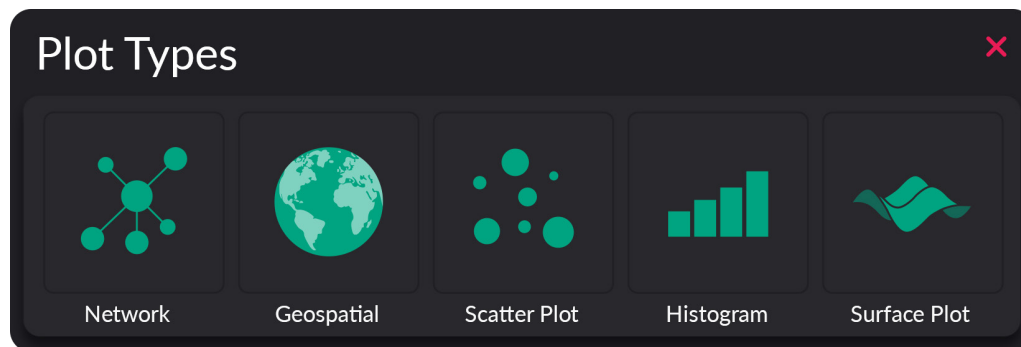# 3Data

Data

Formatting Guidelines

Data Formatting Guidelines

Purpose

The purpose of this guide is to explain how to format datasets in a way that will work best with the 3Data platform.

Plot Types

There are 5 primary plot types supported by the 3Data platform:

This guide will go through each of the graph types and detail some best practices for working efficiently and getting the most out of each of them.

Scatter Plots

Scatter plots are one of the most flexible plots that can be used on the platform. The X, Y, Z, color, size, shape, and rotation shelves can all be assigned to a different column of data within your file.

Recommended file types: CSV and Excel

The platform recognizes text, numbers, dates/time, and formatted number such as currency. The accepted formatted number and currency styles are U.S. style only.
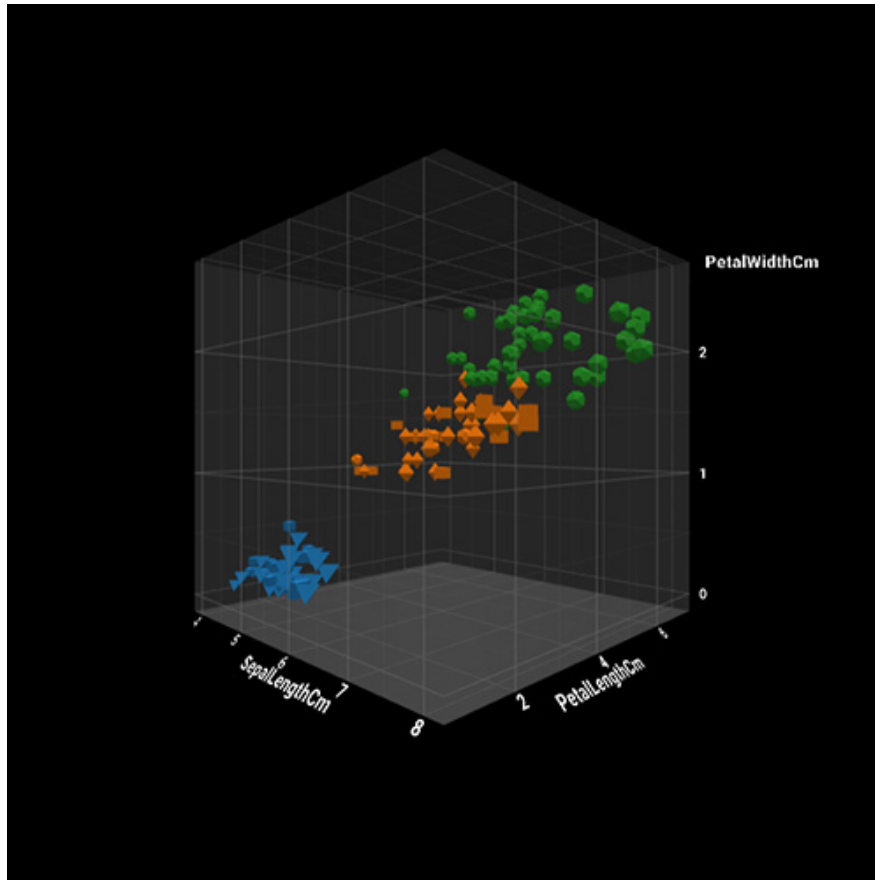
CSV

All data saved in a CSV file should be saved as a single table (denormalized). All data points in a single column should be of the same type (string, decimal, date, etc.) If both text and numbers are used, the program will default to treating the entire column as text. This is often times not desired.

The name used for the CSV file when it is uploaded will also be used as the name of the data set in the platform. Dates stored in CSV files should be stored in the "YYYY-MM-DD" format. This is also known as ISO 8601.

Excel Files (xls and xlsx)

The same rules generally apply for Excel files as they do for CSV files. The platform supports multiple data sets being loaded simultaneously when a single Excel file is loaded. Each tab of the Excel file will be loaded as a separate data set in the platform. The name of each data set will be based on the name of the tab that it came from. For this reason, it is recommended to name each of the tabs in the Excel document an appropriate name. (Please try to avoid naming your data "Sheet 1", for example.)

Dates stored in Excel format should be stored in the "YYYY-MM-DD" format. This is also known as ISO 8601. Excel's default date format is also supported.

Histogram Plots

Histograms are used to show the distribution and frequency of particular values of data. Once the user chooses which columns of data to uses on the X and Z axes, the platform will create hexagonal bins along that X/Z plane. It will then count the frequency of observations which fall into each bin and display the how many were located within each bin. By default, the height and color of each hexagon corresponds to the frequency of observations.

Recommended file types: CSV and Excel, generally follow the same rules as scatter plots. Using text is not recommended for the X or Z axes.

Text data is generally not recommended for this type of graph.
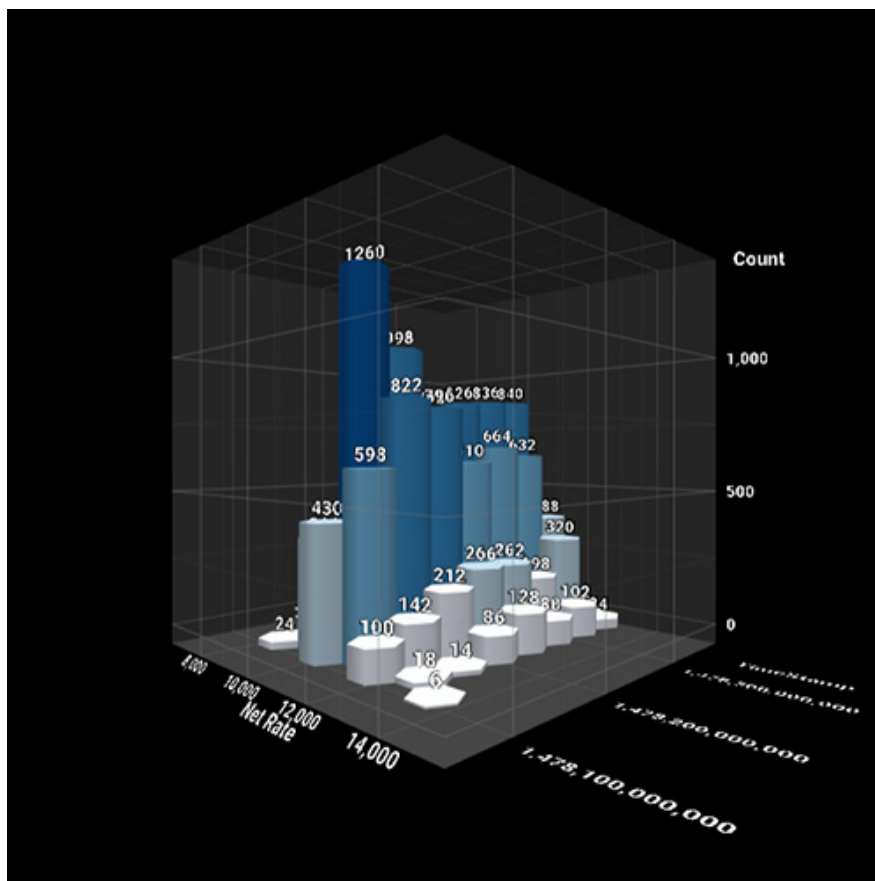
Surface Plots

Surface plots are used to show the contours of a 3D surface. Each combination of X and Z coordinates should have a unique Y value assigned to it in order to ensure a smooth surface plot.

Recommended file types: CSV and Excel, generally follow the same rules as scatter plots. Using text is not recommended for the X or Z axes.

Restrictions: Each data point in the column of data assigned to the vertical axis (Y) must be unique. Duplicate Y values in a data set for a single X/Z coordinate combination will result in errors.
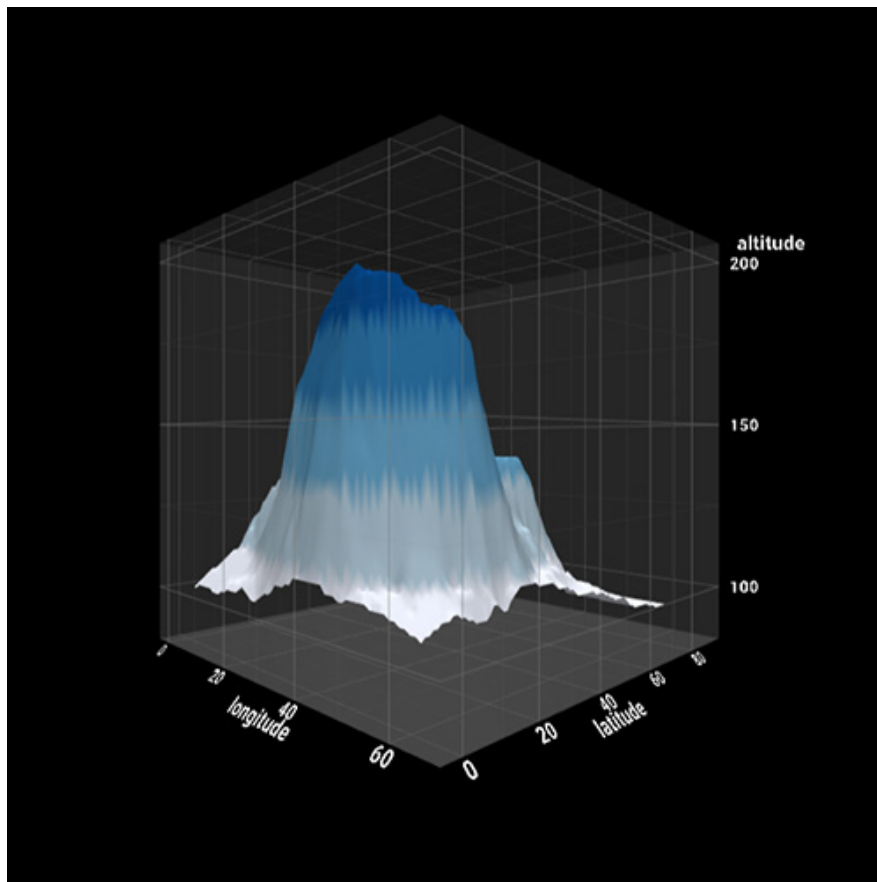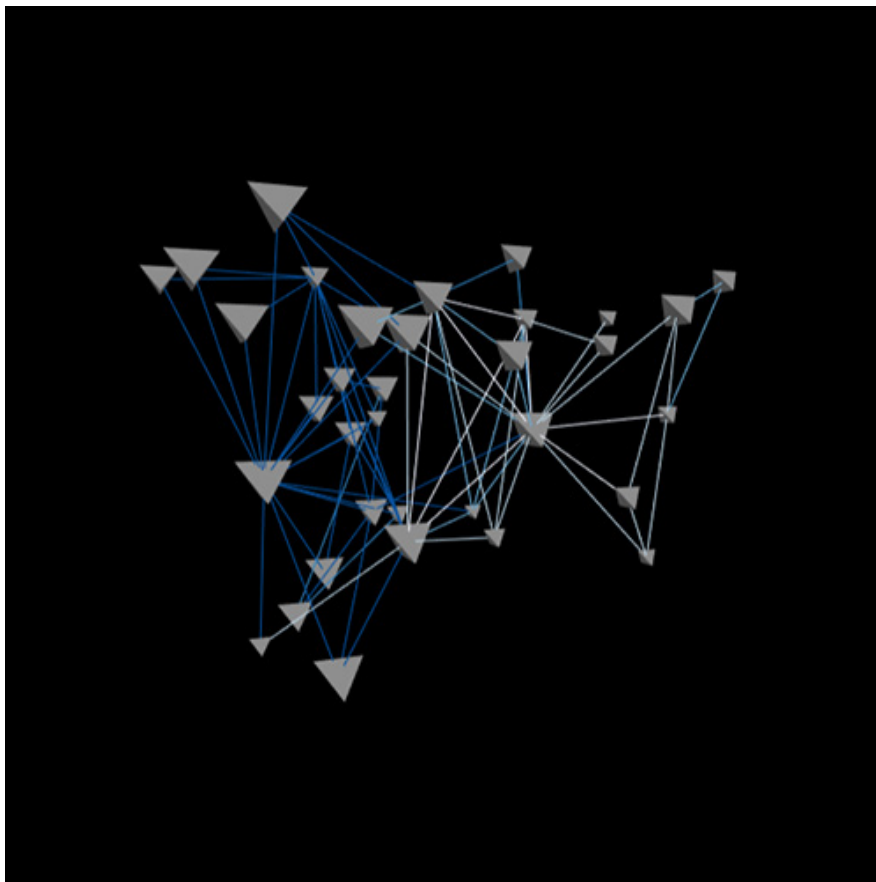
Network Graphs

Network graphs are used to show relationships between objects. Each object in a network graph can be linked to any number of other objects. The platform uses a force-directed graph drawing algorithm to maximize the distance between objects for greater clarity.

Network graphs are composed of nodes and edges. Nodes (sometimes called vertices or points) are the objects which can be assigned characteristics, such as color, shape, and size. Every ID used to create nodes must be unique for every object in the database. In addition, the number of nodes need to be the same as the number of attributes assigned to those nodes.

The nodes are connected to each other using edges (these are also sometimes known as links or lines). Edges are created when two different IDs are referenced in the source and target shelves. The first item in source is linked to the first item in target. The second item in source is linked to the second item in target, etc. For this reason, the source and target data columns must always have the same number of items.

Recommended file type: JSON

Network graphs work best when uploaded in the JSON file format. JSON files always start and end

with a curly brace. Generally, it is recommend to reformat the data so that each attribute has a header in quotes followed by a colon and a list in square brackets for every object in the data set. The source and target connection can be placed at the end of the file. Below is a sample file for a simple network graph visualization:

{"id":[1, 2, 3],
"person":["Joe", "Sam", "Anna"],
"type":["Boss", "Employee", "Employee"],
"source":[1, 1],
"target":[2, 3]}

Notice that "id", "person", and "type" all have a length of 3. "Source" and "target" then had a length of 2 objects, connecting id 1 to id 2 and then id 1 to id 3. Any source or target IDs which are referenced in the source or target columns of data must also be present in the "id" field.

The format for dates stored in JSON files does not necessarily matter because dates are treated as simple text in this example. The name used when the JSON file was uploaded will also be used as the name of the data set in the platform.

Geospatial Graphs

Geospatial graphs are used to show the latitude and longitude of a series of data points on a globe. The plot will automatically zoom in to the portion of the globe which corresponds to the underlying data.

Recommended file types: CSV and Excel. Using text is not recommended as a data type for this plot.

Restrictions: The values assigned to the column of data assigned to latitude must all be between -90 and 90. The values assigned to the column of data assigned to longitude must be between -180 and 180.
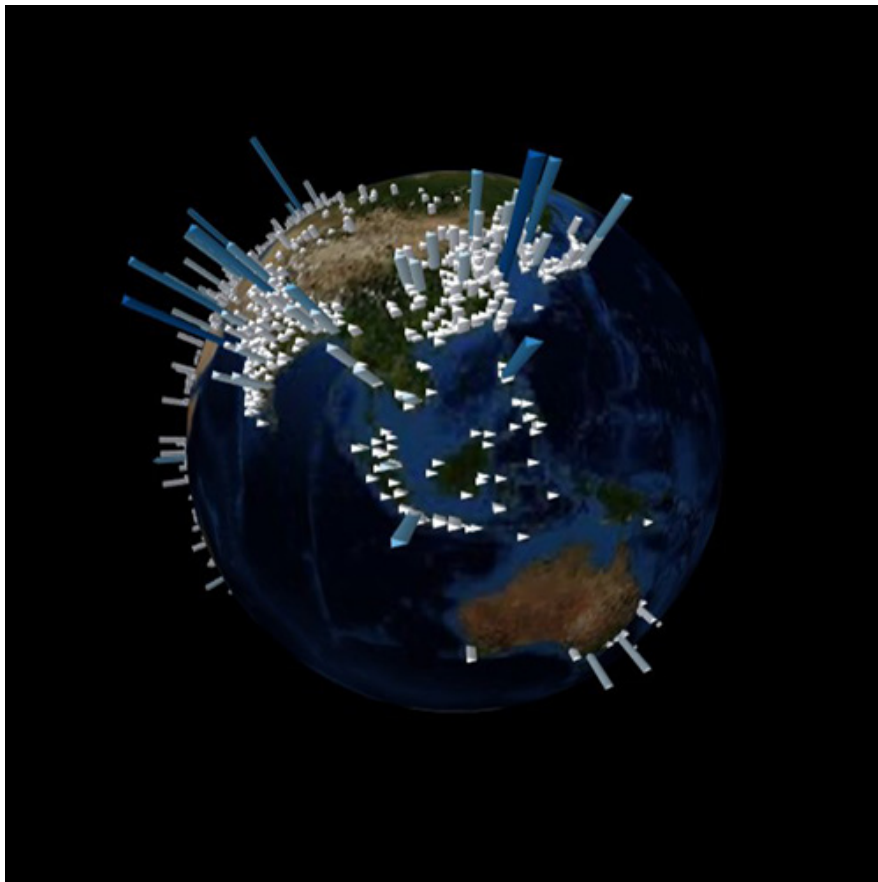
Null Values

Rows of data with null values are generally excluded from graphs. If you wish to include these rows of data, it is recommended to either remove the columns with the null values from the uploaded data set, or replace the null values with another value before the data is uploaded to the platform.

Performance Notes

To support smooth operation of the immersive VR mode, it is recommend to use a PC with the one of the following graphics cards:

- Nvidia GeForce GTX 1060 (or better)
- Radeon RX 570 (or better)

The platform will still run in 2D on much less powerful hardware, but the experience may be less than optimal in certain cases, especially when very large data sets are being displayed.