

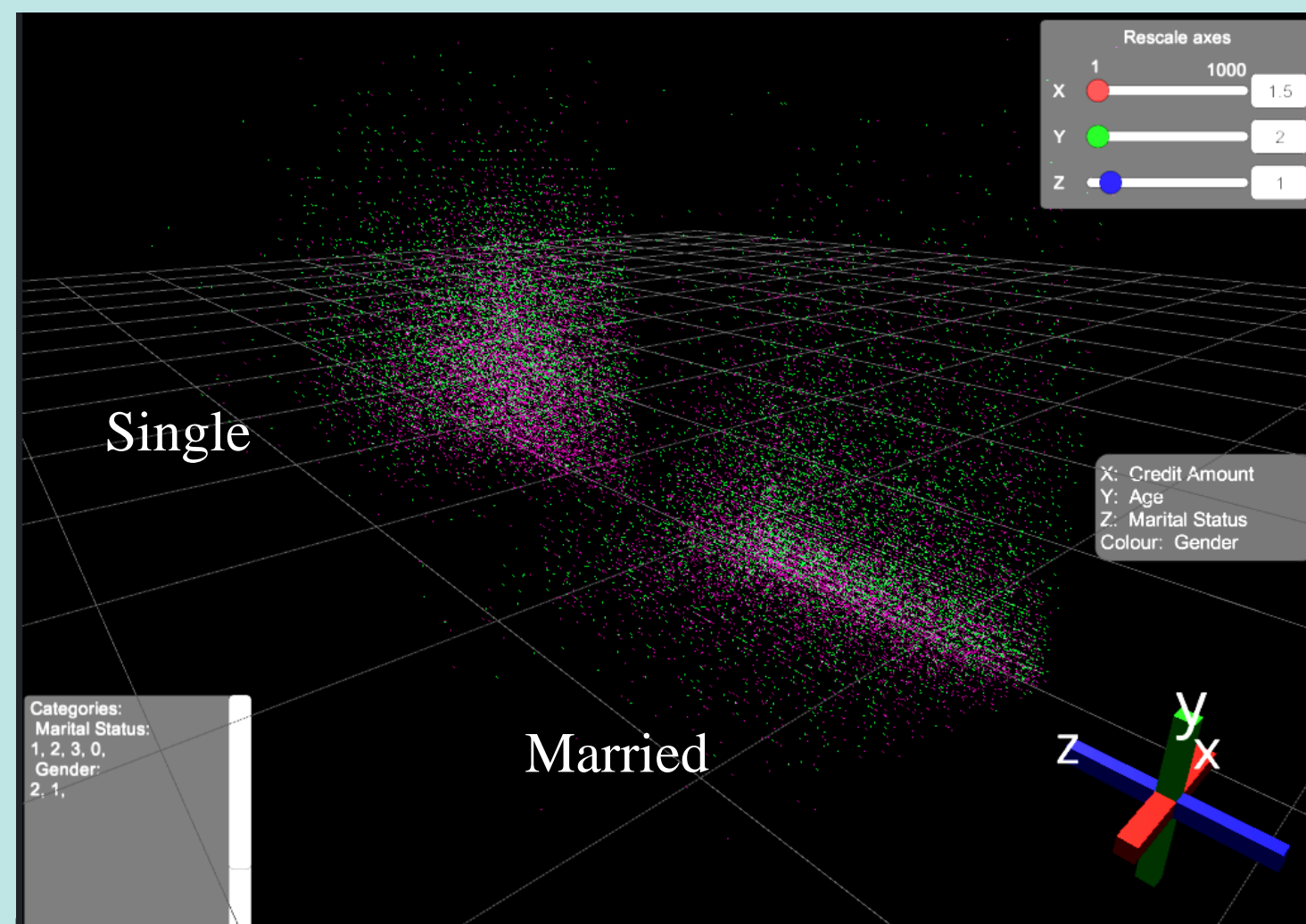
Dimensionality, Scale and Cognition in Large-Dataset Visualisation

Frederick Leighton

MSc in Big Data

Objective

This project aims to create a prototype data visualiser that can handle large, high dimensional datasets to facilitate understanding of complex datasets. The issues regarding this are the size and complexity of the data and how to generate a meaningful visualisation.



Visualisation of 30,000 credit card clients. By mapping credit amount, age and marital status on the x, y and z axes, we can see that people have the most credit at middle age

Credit data from Yeh, I. C. & Lien, C. H - UCI machine learning repository. Brain cell data provided by A. Leighton & C. Lohmann, Synapse and Network Development, Netherlands Institute for Neuroscience

Dimensionality

High dimensional datasets contain many columns (a.k.a. fields, categories). This visualiser allows a user to up to four columns to display in three dimensions with colour or use dimensional reduction techniques.

Optimisation

When rendering millions of data points performance issues must be considered. This visualiser is optimised to read data quickly, with minimal memory usage while maintaining a responsive interface.

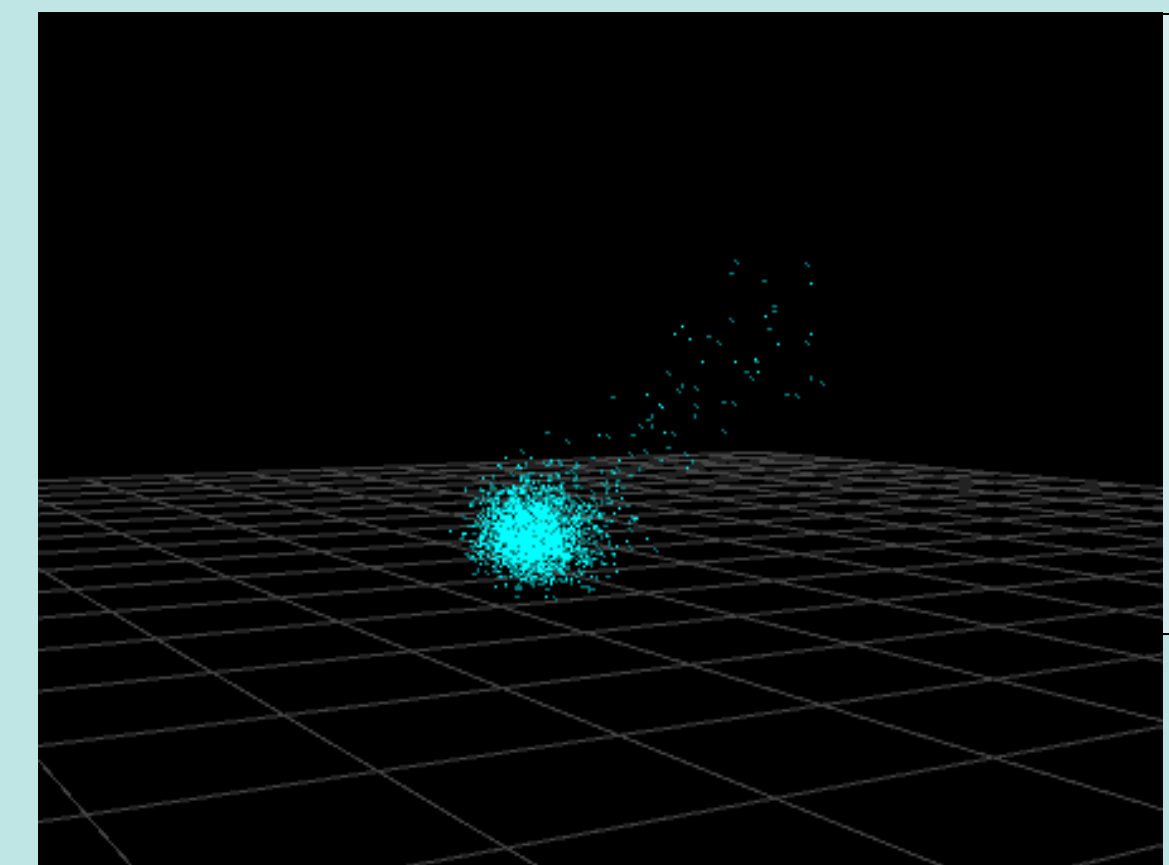
Implementation

A row of data is represented as a 3-dimensional vector, which will become coordinates in space. It supports combinations of categorical or numeric data.

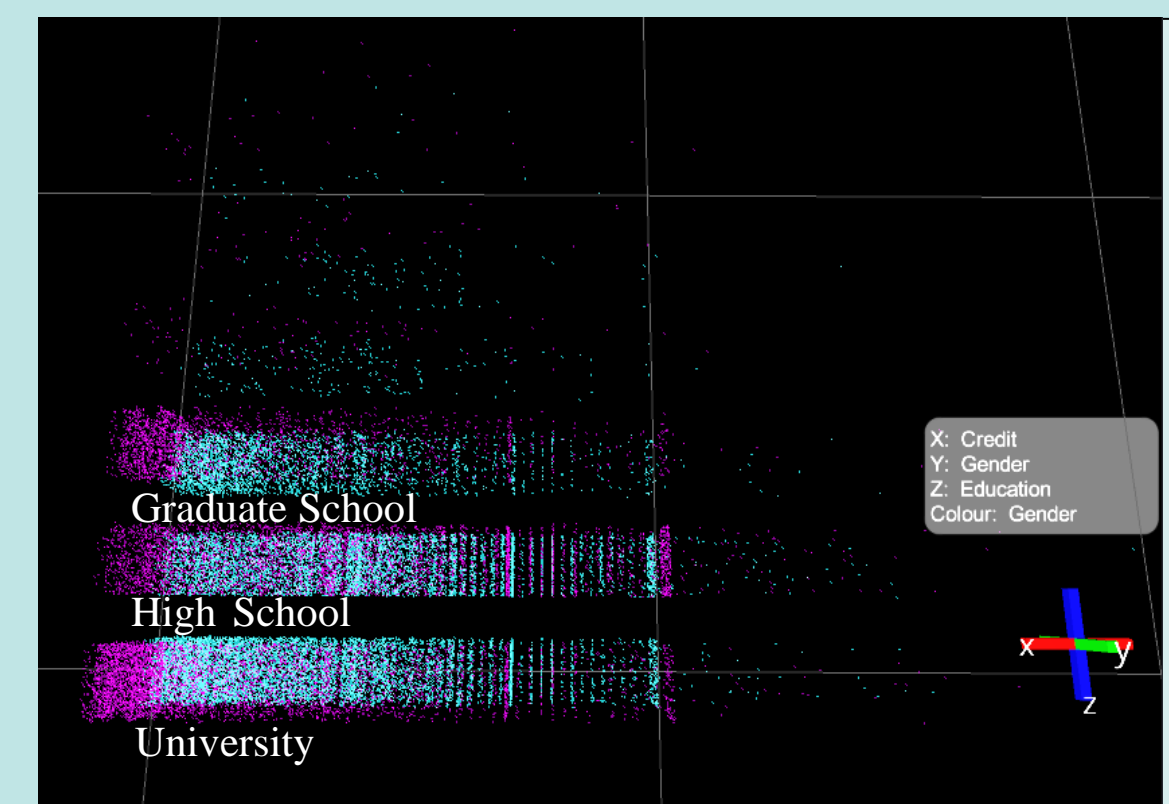


Understanding

Gaining a comprehensive understanding of data relies on using the appropriate visualisation technique. This tool provides interactivity within a 3-dimensional space to facilitate this



Dimensional reduction (PCA) performed on brain cell activity in the visual cortex of a mouse.



Credit card clients, visualising credit by education. High school and University have the highest credit.