# CSE 512 Project Progress Report

Nicole Atherly, Mahir Kothary & Grace Telford

May 20, 2015

## 1 Problem Statement

We aim to create a tool that facilitates interactive exploration of correlations between parameters in a high-dimensional astronomy dataset containing measurements of various properties of 140,000 galaxies. This tool will allow the user to change filtering of the data, choose which dimensions are shown, and will allow for brushing/linking between plots.

## 2 Literature Review

The problem of effectively visualizing high-dimensional data is an active area of research, especially as large and complex datasets are becoming more common. Visual representations are limited to two or three dimensions, and as the number of dimensions increases, the time required to manually search for an informative low-dimensional projection of the data increases rapidly.

Various techniques have been devised to speed up the process of finding a useful projection of the data. The technique of parallel coordinates is used to display many dimensions of a dataset in two dimensions and allow the user to see structure in the data [2]. `Ggobi` [4] is an interactive visualization tool that "tours" the dataset by rotating through many low-dimensional projections of the data so that the user can rapidly see many different views and quickly identify interesting projections. It also allows the user to use brushing/linking between multiple graphs to explore structure in the data.

Several authors have devised methods of automating the choice of useful projections, using orthogonal [5] and non-orthogonal [6] projections. Studies that analyze the relative effectiveness of different methods of choosing low-dimensional projections have found that the optimal dimensionality reduction technique is highly dependent on the nature and size of the dataset [3, 1]. At present, there exist many methods of finding potentially useful low-dimensional projections of multidimensional datasets, but the success of each technique in identifying interesting structure in the data quickly varies by dataset and by task.

# 3 Project Plan

We have identified three major tasks that comprise this project. We will each be accountable for ensuring that one of these tasks is progressing appropriately, but we will all contribute to all tasks. In particular, all of us would like practice coding with D3.

- Mahir: design of user interface & layout of figures
- Grace: efficient, dynamic querying of dataset
- Nicole: Animation & brushing/linking

Since there are three weeks between now and the poster session, we also identified three milestones; we will achieve one by the end of each week.

- Week 1: static visualizations (2 prototype layouts) with real data displayed on graphs
- Week 2: interactive querying, filtering, and changing of displayed variables
- Week 3: brushing & linking between figures, final aesthetic touches

# References


[1] Etemandpour, R., et al., 2014, *Perception-Based Evaluation of Projection Methods for Multidimensional Data Visualization*, IEEE Transactions on Visualization and Computer Graphics, 21, 81 - 94

[2] Inselberg, A., 1997, *Multidimensional Detective*, IEEE Symposium on Information Visualization, 100 - 107

[3] Keim, D. & Krigel, H., 1996, *Visualization Techniques for Mining Large Databases: A Comparison*, IEEE Transactions on Knowledge and Data Engineering, 8, 923 - 938

[4] Swayne, D., Lang, D. T., Buja, A., and Cook, D., 2003, *GGobi: evolving from XGobi into an extensible framework for interactive data visualization*, Computational Statistics & Data Analysis, 43, 423 - 444

[5] Tatu, A., et al., 2009, *Combining Automated Analysis and Visualization Techniques for Effective Exploration of High-Dimensional Data*, IEEE Symposium on Visual Analytics Science and Technology, 59 - 66

[6] Yang, J., Peng, W., Ward, M. O., and Rundensteiner, E. A., 2003, *Interactive Hierarchical Dimension Ordering, Spacing, and Filtering for Exploration of High-Dimensional Datasets*, IEEE Symposium on Information Visualization, 105 - 112