Visualizing Graduate Admissions Data

Aaron Doubek-Kraft, adoubekk@ucsc.edu

March 6, 2017

Problem

- Given the unusually large applicant pool to the UCSC Computer Science graduate program, determine if any correlation exists between:
  - an applicant's quantitative records (GPA, test scores, etc.)
  - their potential to succeed in the program
  - their strength in a given research area
- Create a tool to visualize general multivariate datasets

Approach

- Parallel Coordinate Plots: Sets of parallel axes connected by polylines that represent an individual record
- Strengths
  - Standard and generic approach to multivariate visualization
  - Easy to identify trends in the data
  - Straightforward implementation
- Weakness
  - Quickly becomes visually cluttered for large datasets
- Overcoming Clutter
  - Map polyline color to group membership
  - Highlight group or subset by hiding or graying out other lines
  - Highlight data that falls within a particular range on an axis (brushing)

Implementation

- Operations on dataset: Python scripts
  - Normalize test scores
  - Strip rankings from research interests to improve grouping
- Visualization: Web Application
  - d3.js: Graphics
  - jQuery: DOM Manipulation
  - w3.css: Style Framework

Results

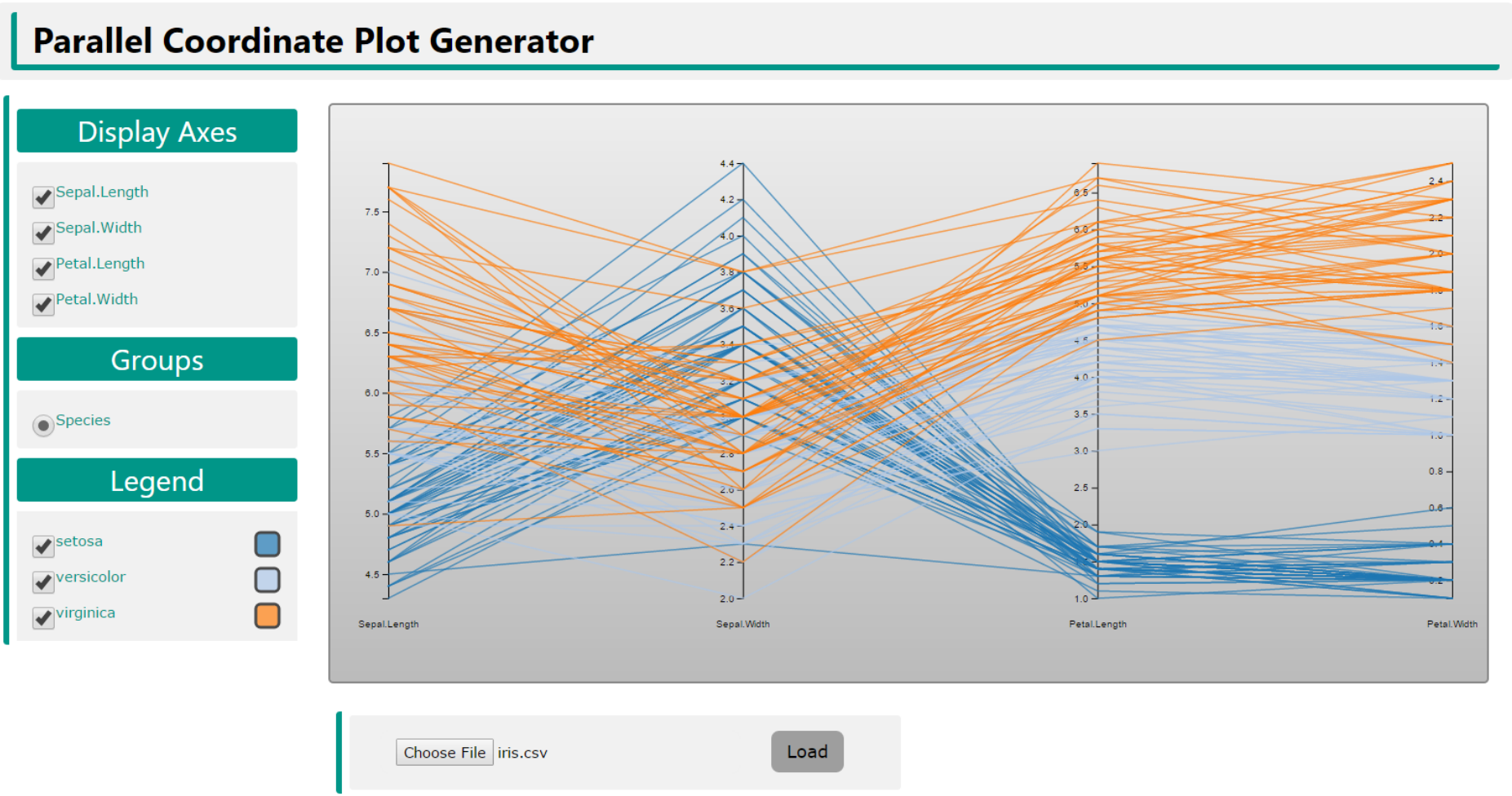

Figure: Visualizing Edgar Anderson's Iris dataset

To Complete

- Processing and analysis of graduate admissions dataset
- Additional user interactions, such as brushing
- User customization of appearance of visualization