INTRODUCTION

In the data deluge of the modern era, an efficient method for identification, extraction and visualization of pertinent data is an absolute necessity. To this end interdisciplinary and international collaborations are a definite necessity; however there are many obstacles that prevent effective communication between engineers and researchers in such an environment. Communicative devices such as mathematical diagrams and computer models serve to mitigate the problems arising from disciplinary and language barriers.

The goal of this project is to create discipline-specialized tools that take advantage of an individual’s knowledge base and expertise in an interdisciplinary collaborative environment and allowing them to present the result of their analytical work using expressive models. These tools are intended to significantly improve the speed and efficiency of data analysis.

This project takes advantage of an interdisciplinary undergraduate research group led by two faculty advisors, also known as an Affinity Research Group (ARG), to investigate and implement a toolset for visual analysis of statistical data in various disciplines and areas. As depicted in figure 1, the end result of the project will be tools that allow for effective extraction of information from raw data.

METHOD

The research is conducted in two phases. The first consisted of researchers familiarizing themselves with the intended development environment while investigating its potential applications and limitations. The second phase is focused on the development, testing and eventual distribution of the toolset. This phase was executed by subdividing the research team into task specific subgroups, with each individual contributing to three separate subgroups. This allowed researcher to remain task oriented while remaining aware of the interdependencies of separate components. Results from each subgroup were periodically distributed to other members to allow for deeper collaboration and constructive feedback.

TOOLS

The primary tools used for the project are the R statistical environment and its Rcmdr and RGL packages. This project takes advantage of existing R features and plug-ins to build a better overall suite of statistical visualization and analysis tools. The project takes advantage of the Graphical User Interface (GUI) provided by the Rcmdr package and the RGL interactive 3-dimensional (3D) rendering tools in the RGL package.

Utilizing existing R packages as prerequisites takes advantage of future improvement in these components. Also the dissemination mechanism for R packages allows users to download only updates for the packages/prerequisites necessary for the task at hand, thereby sidestepping the difficulties associated with updating monolithic solutions. Future implementations may be updated to take advantage of technologies such as cloud computing.

PROGRESS AND RESULTS

The first phase of the project has been complete and the second phase is well under way. Results are categorized as technical requirements, algorithms and software.
development. Current results include an outline of technical requirements of the final product, partial result in algorithm development and some completed visualization tools and GUI components.

The software suite or toolsets is designed with the memory limitation of 32 bit operating systems in mind. Since the general expectation is that datasets used for input may be larger than 2GB two modes of operation are desirable. The first mode operation, which is being implemented currently, relies on pre-processing of data to create a new aggregated data for input. The second mode of operation which being investigated and will be implemented at a later date involves aggregation and analysis of data as a stream of information during input. Both methods strive to reduce data to its smallest yet most meaningful form.

Algorithms currently in development are adaptation of statistical aggregation techniques with built in parameters allowing for customization of processes with respect to disciplinary context of the input data. The objective is to minimize noise and maximize retention of pertinent data and patterns. Also being investigated is possible context specific customization to sorting algorithms. Sorting algorithm are important due to the potential savings in computing time, however it should be noted that in many cases, such as time-series, data should not be reordered.

The project supplements the GUI provided by Rcmdr through the addition of discipline specific menus. Each menu item is a drop-down revealing more options with discipline specific terminology. Select options provide pop-up windows allowing for a more refined control over the analytical process. The customized vocabulary creates a user friendly environment and also allows for a degree of internationalization, since a majority of professional engineers and scientists are familiar with the English nomenclature of concepts particular to their field of expertise.

Researchers are also researching discipline-customized visualizations that are clear while communicating a more complete and informative overview of data they represent. Current works are predominantly focused on interactive 3D models such as the one depicted in figure 2.

In this example one set of data is represented as clustered bars while another set pertinent to the context of research is presented as mesh overlay allowing users to compare and contrast information very quickly. In general the toolset is designed so that diagrams such as this can be generated with only a few clicks rather that the copious amounts of scripting normally necessary for customized graphics.

Limited testing of tools is currently underway, though upon completion tools will be distributed to collaborators in different institutions and departments for thorough applied testing.

ACKNOWLEDGEMENTS

Material generated by this project is based upon work supported by the U.S. Department of Homeland Security's International Center for Command Control and Interoperability - Visual Analytics for Command, Control and Interoperability Environments (VACCINE) Center under Award Number 2009-ST-061-CI0001.

Figure 1: 3D model comparing to data sets