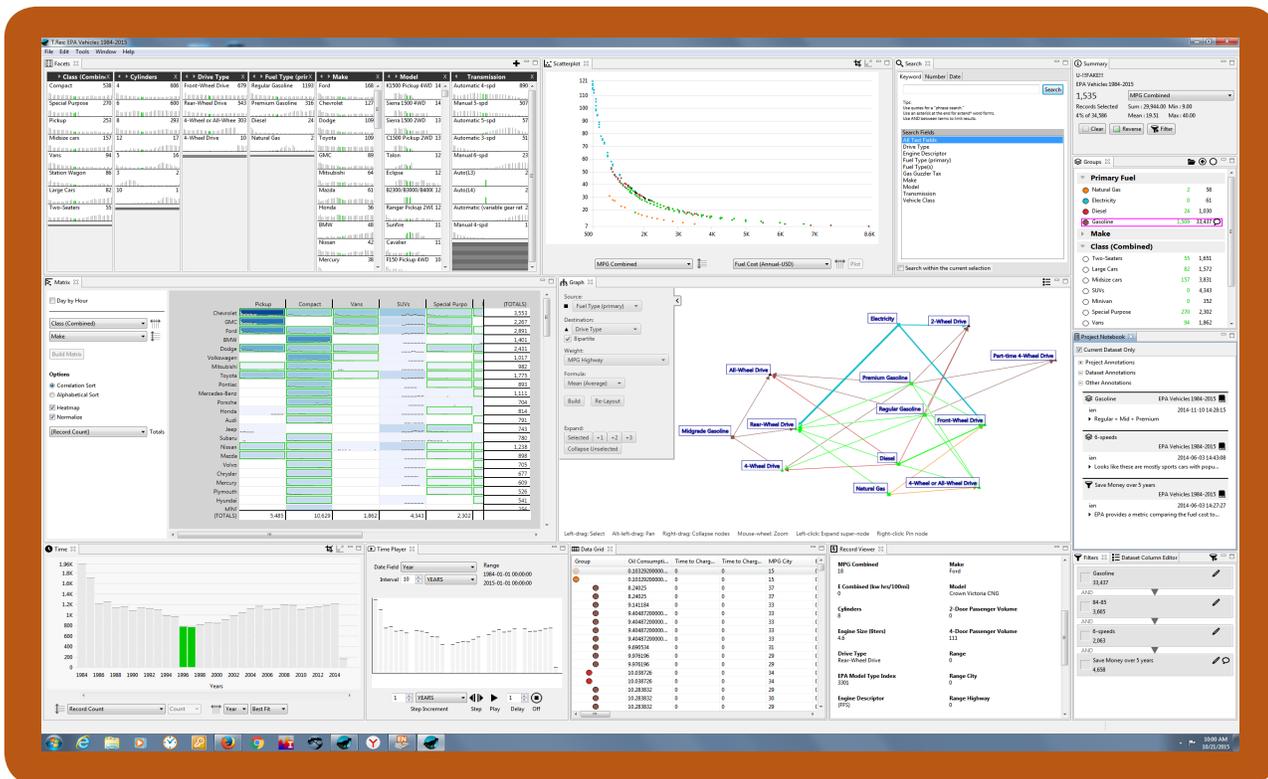




T.Rex Advancing Insight from Structured Data



T.Rex – Transaction Explorer – is a visual analytics tool that specializes in tabular structured data, like a spreadsheet or a CSV file. It’s a client-server application, allowing the server to do the heavy lifting, enabling the client to open spreadsheets with millions of rows. Developed at Pacific Northwest National Laboratory (PNNL), T.Rex provides multiple views allowing access to categorical, temporal, numerical, relational, and summary data. The interactivity allows users to look across data and see how things relate to each other.

CHALLENGE

Analysts who work with tabular data sets often struggle with characterizing and understanding patterns within their data, discovering anomalies,

and capturing and sharing these findings with other analysts. The size of many tabular data sets often exceeds the capabilities of standard desktop tools, and the tools that are able to open large datasets are often limited by their available views of the data. Analysts often have to work with data scientists or other data experts to begin to understand statistical qualities in their data, and to discover which columns of their data will be relevant to a given analytic task.

A need exists for a tool that can streamline this process for an analyst, enabling them to quickly assess a new data set, understand its relevance to an analytic problem, characterize its contents, perform a deep understanding of the relationships and patterns that exist, and share the results with other analysts.

This tool should be easy to load new data sets into, scale to dataset sizes of a few million rows, provide many complementary visualizations of the same data, then facilitate the user to capture new knowledge about the data set and export relevant results. It's simple enough to use for quick exploration of data sets, yet powerful enough for an analyst to perform deep analysis of a data set.

SOLUTION

T.Rex simplifies the translation and exploration of unknown tabular data sources, adding knowledge through discovery. It is intended as an assessment tool to help rapidly understand a previously unknown data set. Analysts can quickly identify patterns of interest in the records and annotate them to enable future collaboration.

The T.Rex interface is highly customizable to suit each analyst's preferences and analysis goals. Analysts can choose the visualizations and tools they would like to use for a particular session, ranging from tools that provide overviews of their data, to views that provide deep insights into their data. All views share the same selection model and update to show their membership with the current selection. When multiple views are used together, T.Rex dynamically reveals new insights that would normally be difficult to find with only a single, static view of a data set.

T.Rex aides in the discovery of patterns and relationships among categorical information, numeric information, temporal information, and free-form text. For initial exploration of a data set's categorical values, and how those values relate to each other, T.Rex's Facets view reveals values that co-occur, along with histograms to provide temporal awareness of these co-occurrences. The T.Rex Matrix view is used to dive deeper into a pair of categorical columns, to find highly correlated pairs, and the Graph view is used to discover connectivity patterns. Multiple other views aid in data visualizations to discover anomalies or changes in the data over time.

Further, the T.Rex system allows multiple users to access the same data sets. The annotations from project team members are gathered together into a Project Notebook, allowing discovered knowledge to pass between data sets and team members.

IMPACT

T.Rex is a powerful new tool for analysts to explore tabular datasets. It quickly ingests new data, visually explores the columns in these datasets, captures knowledge gained through this exploration, then exports subsets of the data. The core T.Rex system is designed to be generic enough to handle just about any kind of spreadsheet data, with scales up to a few million rows.

ABOUT PNNL

Interdisciplinary teams at PNNL address many of America's most pressing issues in energy, the environment and national security through advances in basic and applied science. Founded in 1965, PNNL employs 4,400 staff and has an annual budget of nearly \$1 billion. It is managed by Battelle for the U.S. Department of Energy's Office of Science.

For more information, contact:

David Gillen

Project Manager
National Security Directorate
david.gillen@pnnl.gov
(509) 375-5935

Dave Thurman

Manager, Computing Programs
National Security Directorate
dave@pnnl.gov
(206) 528-3221

Visit us online at vis.pnnl.gov



Proudly Operated by **Battelle** Since 1965

