Ikhlas Attarwala
MSAI 339 – Data Science Seminar
11-15-18 – Dr. Jennie Rogers

Summarizing "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases" by Chris Stolte, Diane Tang and Pat Hanrahan

Prior to 2002, the use of multidimensional relational databases was limited by form of interface demand, lack of dense, graphical representations, and a need for rapidly changing visualizations. Stolte, Tang and Hanrahan sought to find a solution here and created Polaris at Stanford University. This project is no longer active, however commercial projects were translated into Tableau Software. Then, a major challenge when using databases was to discover structure, pattern and relationships when presented ever-increasing sizes of data sets. Stolte et al. walk us through their project by discussing the various graphical representations they offer, options for better navigating the database, the need for more interactivity through transformation, and example scenarios to help readers in similar fields see a potential affinity between their work and Polaris.

One of the stronger points of this paper is that as it's related to a product the team is advertising, they do a great job at detailing all of its features through a well-detailed "manual". Throughout the paper they go into great depth in the proper ways to display different representations of tables/graphics, how to extract more information from original input data through Polaris, and most importantly the various tools they offer and what to know about these tools. Another point is that in section 7, they provide scenarios, which is great for consumers on the chance they can relate to the problem and are looking for a solution. They do well to provide different situations too, where the first scenario concerns a use of RDs in day to day work, and the second recreating an analysis of a research project at Stanford through Polaris. Finally, it's always great for a researcher to discuss what direction they aim to take their project in. One of the most important I found was the need to explore performance issues. With the ever-growing complexity of information obtainable, it's important to patch for unexpected developments in larger data sets (or else your product may lose traction in the long run).

This paper lacked a few points I thought could have helped. Firstly, while a Discussion section exists that compares Polaris to graphical efforts by other researchers, and there are many graphs and screenshots of Polaris' interface in the paper, it doesn't show much comparison with any older existing programs. I'm not sure if there weren't many others around 2002, but I'd assume that if the algebra existed, some other system for representing databases should have existed. One such example could be where we cannot compare the older versions of Pivot Tables to their new implementation and extension of in Polaris' interface. Furthermore, while future works is a great section to highlight the potential of Polaris, this paper doesn't discuss limitations to what the Polaris team has had difficulty with other than maybe the core limitation of the CUBE operator computing aggregates.

I think Polaris has created a fantastic project for representing quick multidimensional models in relational databases. These visuals (possibly in various graphical representations) help analysts quickly explore and sift through potential details that they may have missed just staring at numbers all day. My biggest suggestion for future improvement is understanding how to better perform in the presence of large dynamic data sets. I don't know if you can update input data once it's already implemented into Polaris/Tableau, but as I mentioned before, even just data size will continue to grow exponentially, and the need to be able to handle such will be key for Polaris'/Tableau's success. I haven't yet looked into the types of data Polaris can read, but I'd be interested to see if a histogram of elements classified in video footage can also be graphically represented.