



# Visual Analysis of IEEE VAST Challenge 2016 Data



Ian Turk, Zhou Xinan, Jun Tao, Lei Shi, and Matthew Sinda  
Advisors Chaoli Wang and Qi Liao

## Abstract

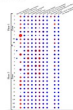
The IEEE Vast Challenge 2016 is a competition intended to make researchers break new ground in visual analytics. The challenge provides competitors with fourteen days of simulated data from a hypothetical laboratory building, including variables such as temperature, gas concentration, electrical power usage, and employee positions in various zones and floors.

The researchers must use visual analytics to identify patterns in the data, connect related events, and predict possibly harmful occurrences. Our team decided to achieve this by using a dynamic, interactive visualization. The visualization has two parts. The first part is a view centered on the scalar data, showing a snapshot of each variable in each zone at a particular time as a dot with color representing value and radius representing distance from the average, as well as a traditional line graph to see trends around the snapshot. The second part is a view focused on the location-based data, and can be configured to show a heat map of a particular variable, the relative time spent by specific employees in a particular zone, or both.

Both views feature automatic anomaly detection using cross-correlation, integrals, and Kulback-Leibler divergence. These tools allowed us to find the patterns and anomalies required by the VAST challenge.

## Evolution of Design

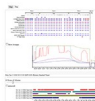
The first view we built was a matrix view, representing the value of the variables as size and their rate of change as color. We would later change this to represent value as color and distance from average as size.



Next, we created a line graph view, a simple but effective visualization for seeing the changes in data over time. We added the ability to select variables to display by clicking on the matrix view, and added simple anomaly detection that only used correlation. We also added a framework for the addition of multiple tabs.



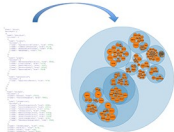
Finally, we combined all of our views into one program, and updated the anomaly detection to calculate the integral of the difference of variables with their average values and use Kulback-Leibler Divergence on histograms made from the proximity data. We also added a heat map function to the proximity data view.



## What is Data Visualization?

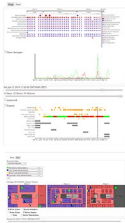
Data Visualization is the process of taking raw data and turning it into an image that can effectively portray the important information and patterns it contains. A good visualization will:

- Show the data without distorting it
- Make important trends obvious
- Present many numbers in a small space
- Allow datasets to be easily compared



## Final Visualization

To allow our program easily analyze the data, we first organized it into a series of javascript objects stored in arrays and maps.



## Automatic Anomaly Detection

Our visualization emphasizes abnormal events in the data based on:

- Cross Correlation
- Integral of Difference
- Kulback-Leibler Divergence

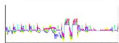
## Results

Shown below are four of the anomalies we found in the building sensors data: Several spikes in the concentration of Hazium (a hypothetical gas that the laboratory is studying), an anomalous fluctuation in thermostat temperature on Tuesday the 7th and Wednesday the 8th, Higher-than-average spikes in CO2 concentration in several different zones, and an unusual fluctuation in VAV Damper Position on Tuesday the 7th that may be related to a large Hazium spike.

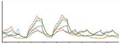
Hazium Concentration over Two Weeks



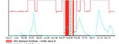
Anomalies in Thermostat Temperature



Return Outlet CO2 Concentration for Various Floors and Zones



Floor 3 - Building Sensors



## Minard's Map

A famous example of data visualization is Charles Joseph Minard's graph of Napoleon's Russian campaign of 1812. The diagram effectively shows six variables in a small space: Number of troops, distance traveled, position in latitude and longitude, direction of travel, and time.

