

Find me in a zone: the power of simple visualizations (VAST 2016 Mini Challenge 2)

Gosia Migut*

Bas Châtel

Leslie Dao

Wessel de Jong

University of Amsterdam, The Netherlands

ABSTRACT

Simple and common data visualizations, such as line graphs, bar charts, heatmaps are often overseen when dealing with complex spatio-temporal data. We show how those simple visualizations meet the challenge of analyzing complex data of Vast 2016 Mini-Challenge 2.

Keywords: visual analytics, small multiples, interactive visualization, intelligence analysis, human information interaction

1 INTRODUCTION

The GASTech company recently moved to a new building with highest energy efficiency standard, but faces several issues with sensory data and experiences some problems with new security procedures. We are provided with two weeks data and asked to use visual analytics to identify typical patterns and issues of concern.

Data consists of two types of variables: concentrations of chemical compounds (energy data) and detections of humans (prox data) in time and in space. The space is divided in three floors and every floor in several zones. There are two zones layout: prox and energy zones. The corresponding maps of the zones were provided.

The specific challenge questions include finding typical patterns in the prox card data, describing typical day for GASTech employees, finding the most interesting patterns and anomalies that appear in the energy and prox data and finding relationships between energy and prox data.

2 ANALYTIC AND VISUALIZATION PROCESS

In our analytic process we have followed the spatial and temporal dimensions of the provided dataset. We developed a set of web-based visualizations in D3 [1] to explore both spatial and temporal distribution of the variables. For both spatial and temporal dimensions we provide specialized tools to explore the energy and prox data. All measurements from energy data we call concentrations and are visualized both in: (1) space, using heatmap overlay on the provided floor maps, (2) and in time, using line plot. All measurements from the prox data we call detections and are visualized both in: (1) space, using heatmap overlay on the provided floor maps, (2) and in time using line plots, parallel coordinates plot and heatmap. The visualizations were not developed as an integrated tool but as prototypes that would give us more flexibility in the development cycle. We aim to show that data as complex as energy and prox data can be visualized effectively using simple visualizations.

2.1 Concentrations and detections in space

The two provided sets of maps form a basis for the visualizations of concentrations and detections in space. We visualize one floor at a time, the user can interactively select which floor to explore.

The concentrations' floor map is colored based on the value of measurements per zone for a specific hour. The darker the color the

higher the measured values. The user chooses a variable of interest with provided drop-down menu and timestamps with provided time slider. In this way we can directly compare the zones for each variable for each timestamps.

The detections' floor map is colored based on the amount of detections per zone for a specific hour. The darker the color the more detections. The timestamps can be chosen interactively with a time slider. Also, the mobile prox data from the robot is plotted. The dots are placed according to the coordinates from the provided dataset. Those do not correspond with the zones but always fall on the borders between the zones.



(a) Energy zones



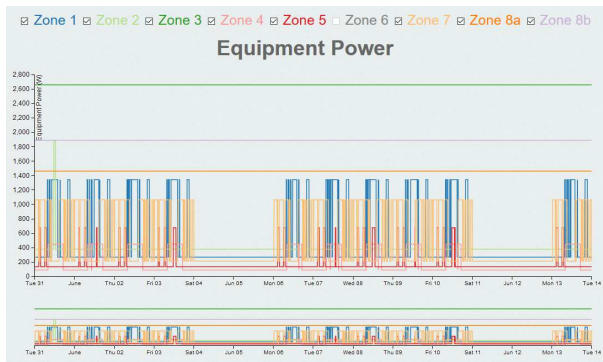
(b) Prox zones

Figure 1: (a) Heatmap of concentrations; (b) Heatmap of detections + robot data as points.

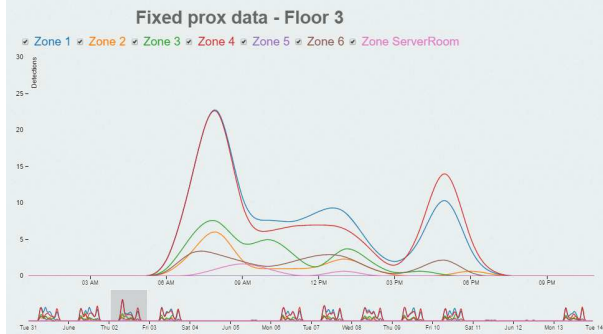
2.2 Concentrations and detections in time

To visualize the timeline we chose to consistently use a line plot for all provided variables. To visualize the spacial information we use the multi-series line graph, where each line represents a zone where the variable is measured. The zones can be interactively switched on and off using simple check boxes. The line graphs can be viewed per variable on demand or alternately could be implemented as small multiples [3]. The full timespan of two weeks does not allow exploration of data in detail for particular interest-

*e-mail: mmigut@gmail.com



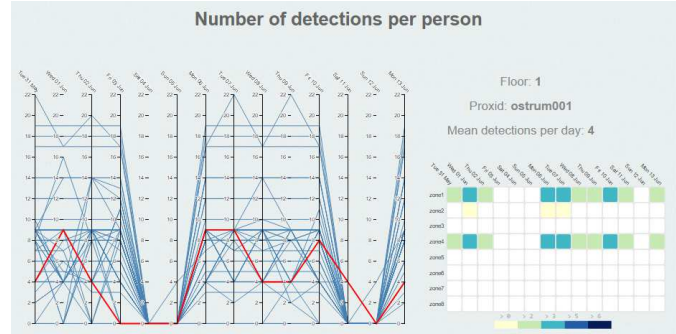
(a) Concentrations for all zones



(c) Detections aggregated



(b) Concentrations for 1 zone



(d) Detections per person

Figure 2: (a) Concentrations over time with focus and context via brushing. Context allows to see the overall patterns over the two weeks period while focus reveals routine and anomalies over selected narrowed time span; (b) Concentrations over time for one selected time zone; (c) Detections over time with focus and context via brushing. Context allows to see the overall patterns over the two weeks period while focus reveals routine and anomalies over selected narrowed time span; (d) Detections over time with parallel coordinates. The daily frequency of detections reveals the overall pattern over the two weeks period for each employee. Selection of one employee line shows frequencies of detections per zone per day on a heatmap.

ing moments in time, therefore a detail and overview windows are required. As it is essential to analyze both overview and detail at the same time we make use of focus+context technique [2], which is realized via brushing.

We use aggregated detections for all employees per timestamp on the line plot. To analyze the behavior of individual employees we propose to plot daily amount of detections per employee using parallel coordinates. The details-on-demand for parallel coordinates shows a heatmap of aggregated detections per day per zone.

3 FINDING ANOMALIES IN THE GASTECH DATA

The visualizations were designed to quickly find the patterns in temporal and spacial data. The simple visual representations of every variable in time with focus and context allows rapid identification of weekly and daily routines and at the same time all the outliers can be spotted. The dates can be directly identified where the measurements look suspicious and should be further investigated. The visualizations focusing on details per employee allow identification of their behavior in the building, including multiple prox card usage, being present in unexpected locations at unexpected time. Since our visualizations support the temporal variables well, it was particularly straightforward to identify the measurements' outliers in time.

4 DISCUSSION

A primary task of the VAST 2016 Mini challenge 2 is to find patterns and anomalies in the GASTech data. Our simple visualizations provide an overview of each variable in space and time. The interactive element allow zooming into details and elements of interest.

Those simple visualizations enable our team to explore the provided dataset and find regularities and abnormalities in the data. The valuable extensions of our visualizations would be a series of plots that allow to trace an employee in the space and time.

ACKNOWLEDGMENTS

The authors would like to thank Sascha van Essen for her creative inputs and feedback on the visualizations.

REFERENCES

- [1] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec. 2011.
- [2] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [3] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *Commun. ACM*, 53(6):59–67, 2010.