

## Visual Analytic Design for Contextualising Sensor Data

VAST Award: Outstanding presentation of patterns in context

[gicentre.net/vastchallenge2016](http://gicentre.net/vastchallenge2016)

**Jo Wood**

giCentre, City, University of London

 [jwoLondon](https://twitter.com/jwoLondon)

This talk focuses on two aspects of visual analytic design to address the 2016 VAST Challenge MC2 – those driven by the nature of the data and those driven by the nature of the challenge tasks. After considering these two categories of design drivers, I will show a few examples of how the resulting design led to insights (and a couple that were missed).

**data → design**

Firstly, consider those aspects of the data we were working with and how that influenced design.

#timestamp,	type,	prox-id,	floor, zone
2016-05-31 00:05:00,	fixed-prox,	vawelon001,	1, 1
2016-05-31 00:20:00,	fixed-prox,	earpa001,	1, 1
2016-05-31 02:26:40,	fixed-prox,	earpa001,	1, 6
2016-05-31 02:31:41,	fixed-prox,	vawelon001,	1, 6
2016-05-31 03:12:00,	fixed-prox,	earpa001,	1, 1
2016-05-31 03:28:00,	fixed-prox,	vawelon001,	1, 1
2016-05-31 05:57:40,	fixed-prox,	earpa001,	1, 6
2016-05-31 05:57:41,	fixed-prox,	vawelon001,	1, 6
2016-05-31 06:47:00,	fixed-prox,	vawelon001,	1, 1
2016-05-31 06:59:00,	fixed-prox,	earpa001,	1, 1
2016-05-31 07:00:00,	fixed-prox,	ibarranco001,	1, 1
2016-05-31 07:00:00,	fixed-prox,	jsanjorge001,	1, 1

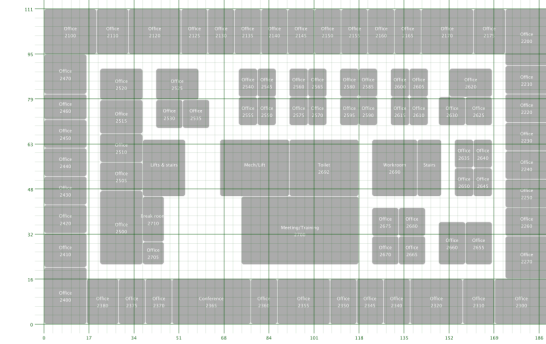
*Temporal referencing*  
(high precision)

*Spatial referencing*  
(low precision)

The fixed prox card records formed a large part of the data used in the challenge. Note the difference in spatial and temporal precision in these data.

The building sensor data (HVAC) also showed a relatively high temporal precision but much lower spatial precision.

In contrast, the mobile prox data showed higher spatial precision but represented very sparse trajectories through the space-time volume.

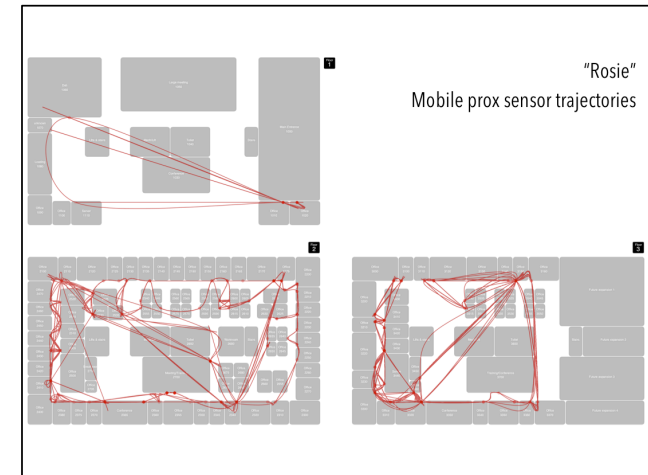


Although we have detailed spatial data about the building we were investigating, and the task would benefit from detailed spatio-temporal data, we were only provided with so-called 'checkpoint' data...



Transitions between a fixed set of zones at a much coarser resolution than the rooms of interest to our investigation.

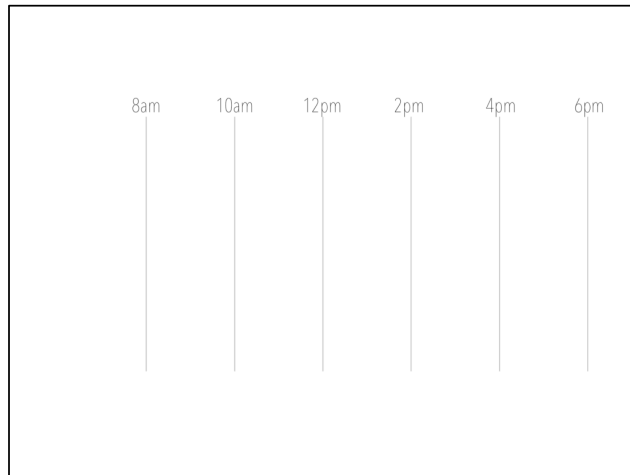
This is a common scenario for much spatio-temporal VA from sensor data (see two-page technical paper for more details).



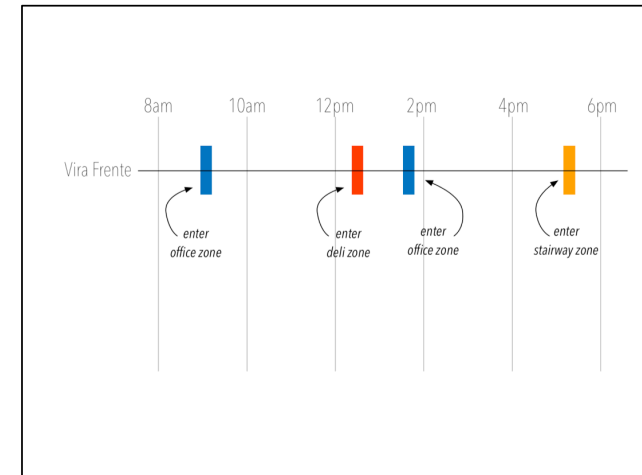
The mobile prox data, which had a much finer spatial resolution, had a (generally) poor temporal resolution and a very sparse sampling of people's movement.

This suggested that there is relatively little benefit in mapping as shown here.

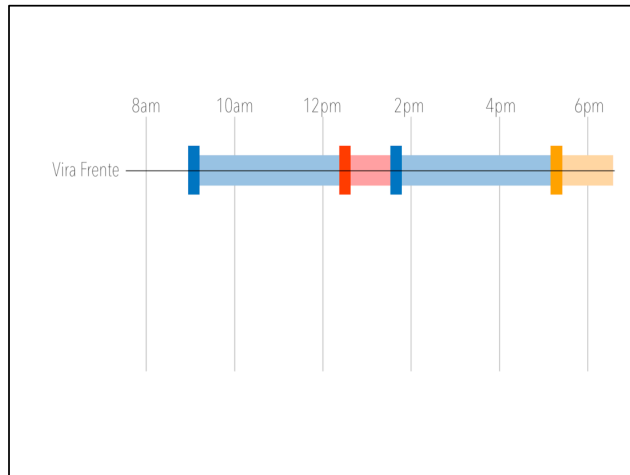
But note, one critical pattern was recoverable from these mobile data in relation to the theft of Patrick Young's card, but this was missed in the visualization design adopted - a lesson to be learned here.



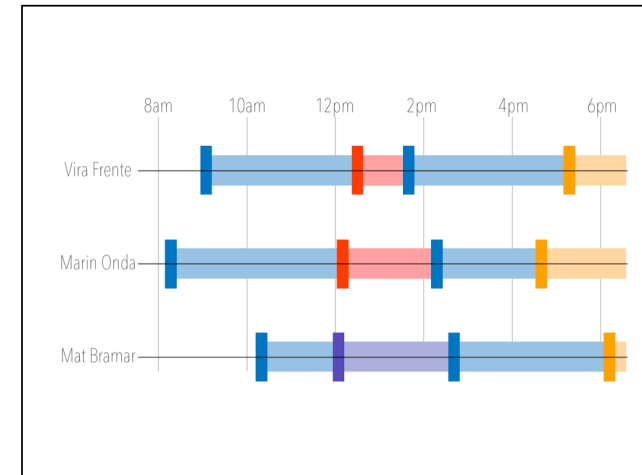
High temporal precision, low spatial precision reflected in the visual design.  
Position an expressive visual variable so map this to most precise data – time.  
Used the full width of the graphical space to record time (either over full period or over 24 hour cycles).



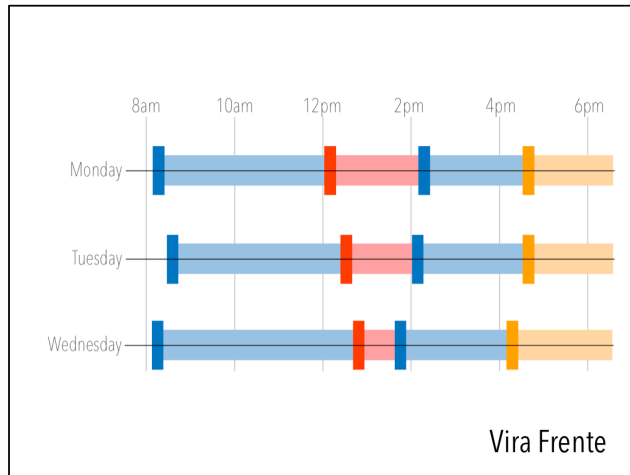
Individual prox card entries, both fixed and mobile can be symbolised along timeline  
Colour hue less expressive, so reflects the coarse spatial categorisation.



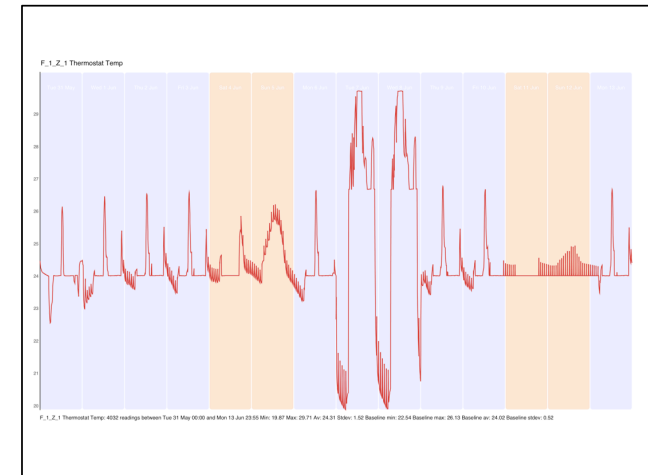
Assumed location can be interpolated between events  
 Reflected in colour encoding with reduced saturation for these more uncertain interpolated values.  
 But note uncertainty in real location of movement within zones or without prox card.  
 Can be particularly important for deliberate behaviour by individuals attempting to conduct unmonitored activity.



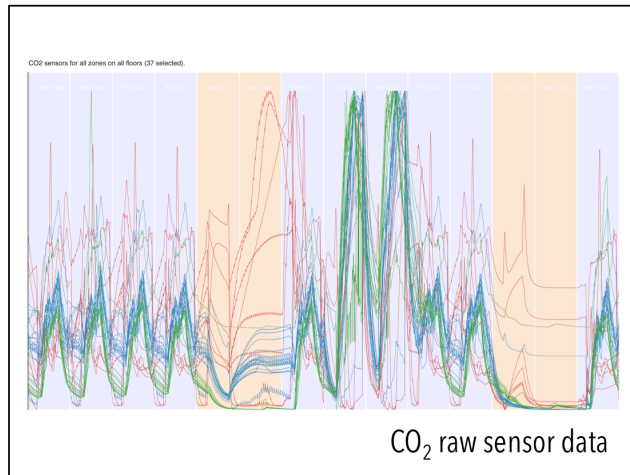
The same design allows comparison between individuals...



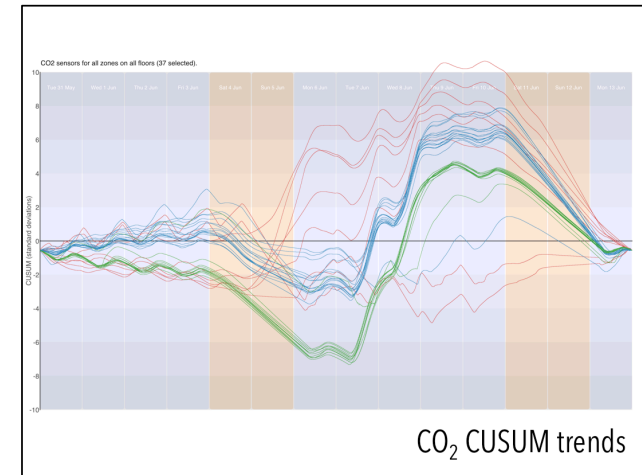
Or comparison over subsequent days for any individual (similar to 'calendar view')



Projecting building sensor data over the same temporal range allows comparison with prox zone movements.

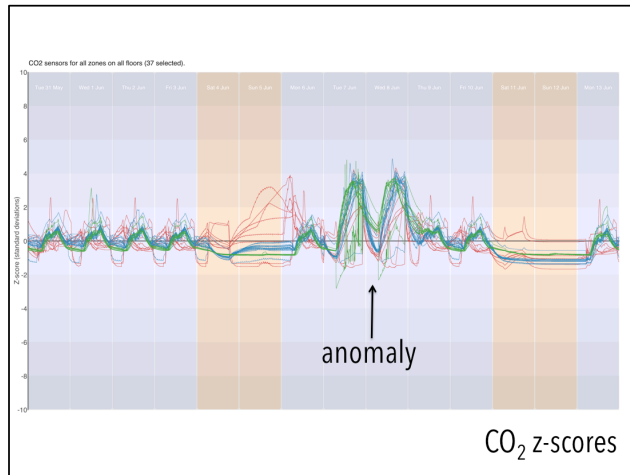


But raw sensor data over many different ranges makes comparison hard, so some simple processing of signals to make comparison easier...

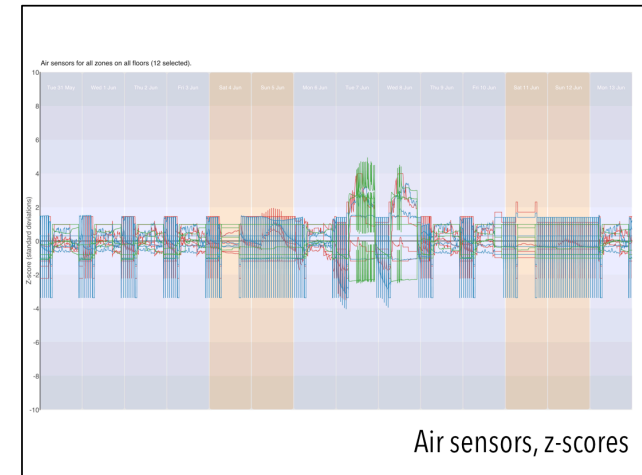


Cumulative Sum plots (CUSUM) allow early detection of deviations away from some expected cumulative average. This both smooths noise and shows systematic trends away from normal behaviour and so is suited to early detection of problems. Should be well suited to streamed visual analytics.

Red = floor 1  
Blue = floor 2  
Green = floor 3

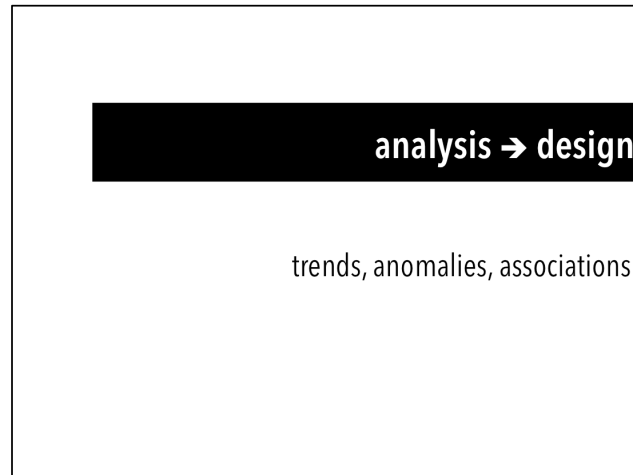


z-scores standardize ranges so allow both detail of signal to be shown but also comparison between data with different ranges.  
 Red = floor one  
 Blue = floor two  
 Green = floor three

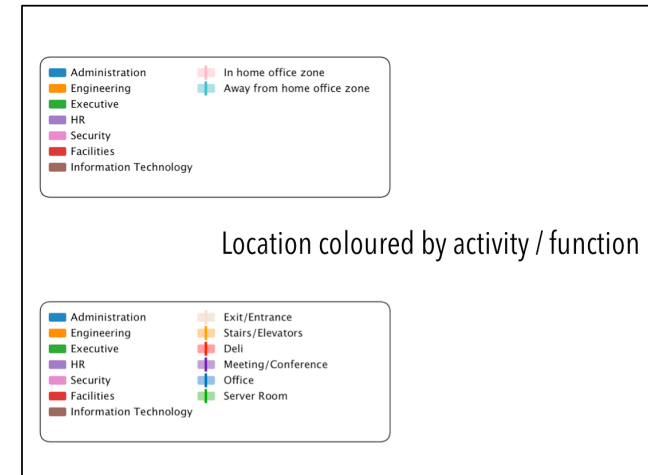


Can compare sensor data of different scales  
 Note here the anomaly on floor three (green) June 7-8th





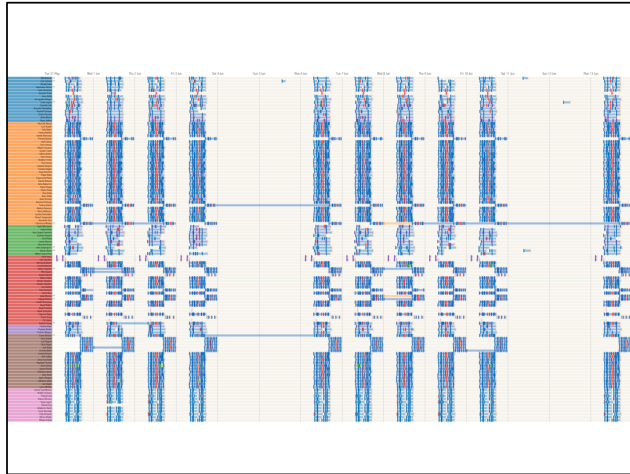
We move on to considering how the analysis required influenced design. As with many visual analytic tasks, three elements of the analysis dominant – detection of trends, which are required to spot anomalies, and the detection of associations between observations in order to hypothesise about explanations.



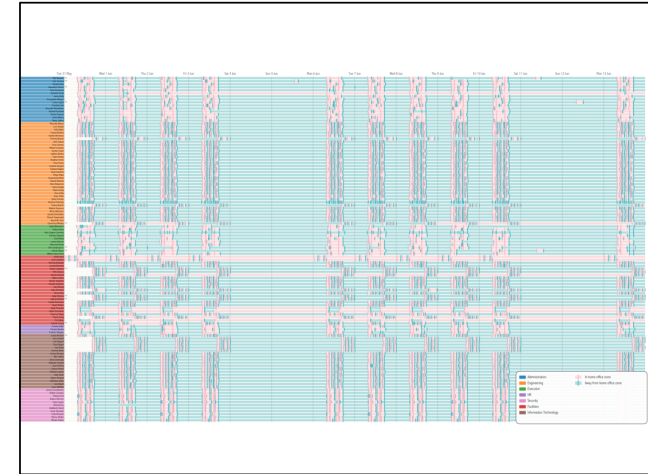
Rather than colour location by zone, each was re-classified into (a) home and away locations; (b) zone function.

Home-away classification a property both of the zones and of the individuals within them ('home' to one employee might be 'away' for another in the same zone).

Zone function is a simple reclassification of the original zones based on the function of the rooms within them. Interpolated colour between functions represents zones with multiple functions. Perceptual colour space was used so that interpolated colours for multi-functions had meaning.



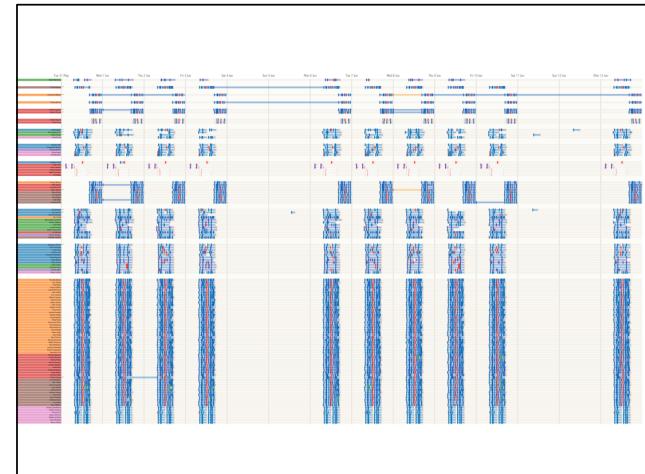
Some trends and anomalies are detectable with this layout, but it is easy to miss some patterns and would be harder still for larger datasets.



The binary home/away classification helps but still the same problem. Where should we direct our attention?

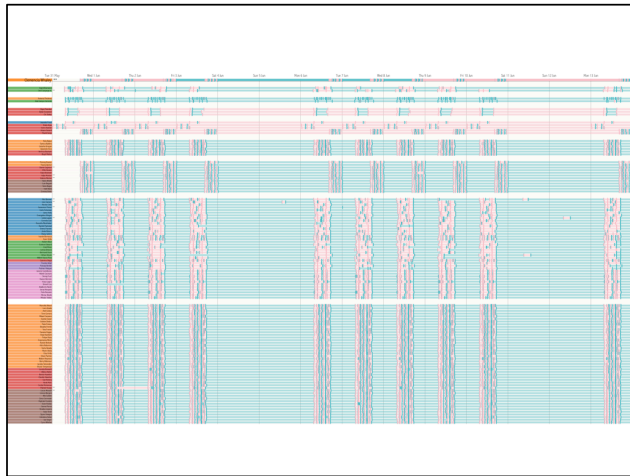
clustering to separate trends from anomalies

Clustering offers opportunity to prioritise some rows over others and separates trends from anomalies.



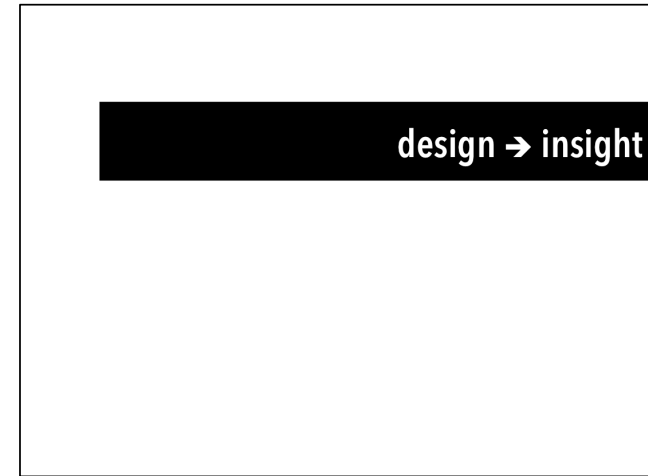
K-means++, vertical gap between clusters, ordered by cluster size.

Typical patterns towards bottom, anomalies towards top

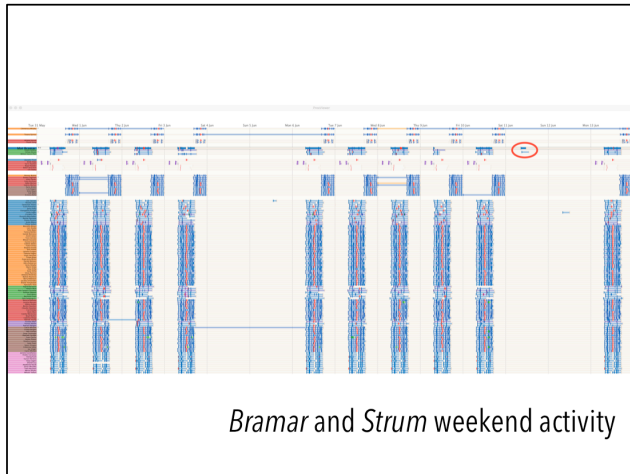


By sorting clusters by size, anomalous prox card patterns tend towards the top of the page.

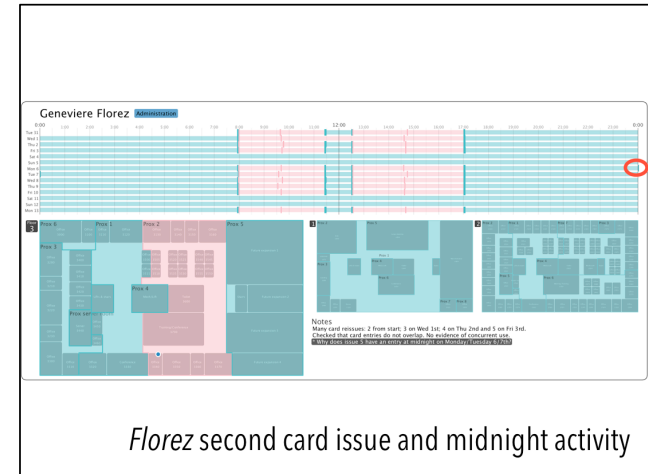
This solution would scale well to much larger datasets.



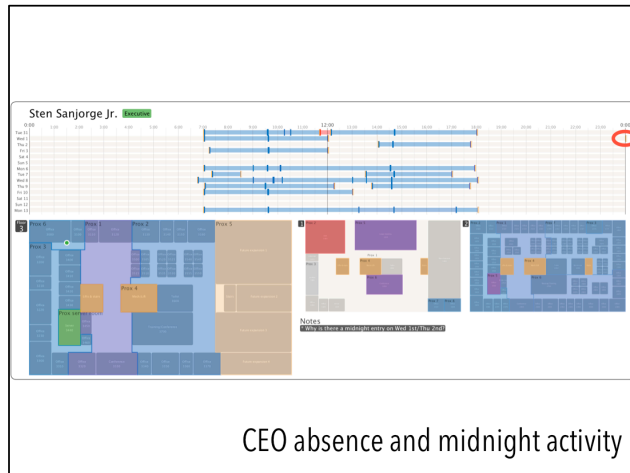
Finally, we look at some examples of insights generated following this design.



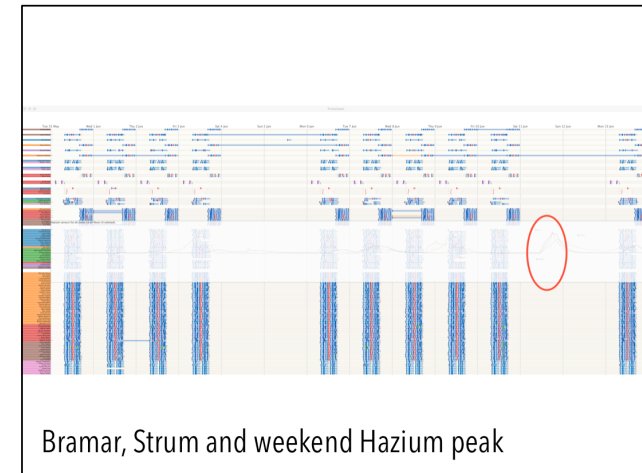
Comparatively rare weekend activity easily detectable towards top of layout in 'empty' columns.



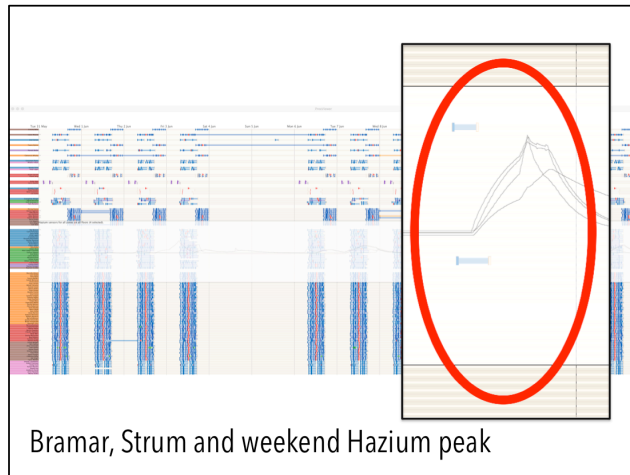
Midnight activity possible to detect, although short events do not occupy much space in layout.



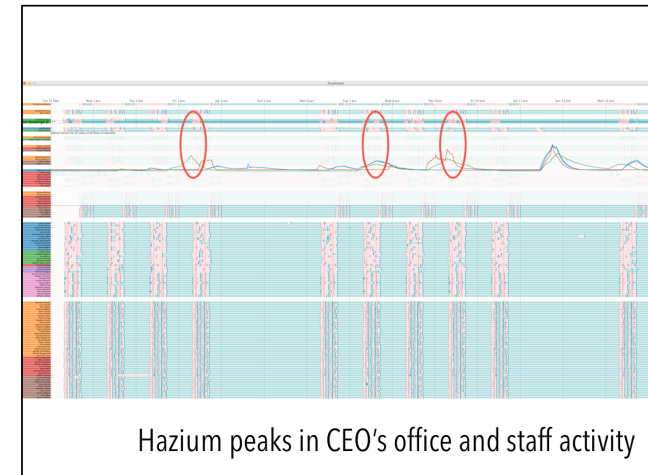
Sanjorge (the CEO) shows a more irregular daily pattern than most employees. Shows not only anomalous midnight activity, but no activity in the 12 hours before or 12 hours after.



Those anomalous weekend appearances by Bramar and Strum are associated with a large peak in Hazium just after they leave.



Here shown in more detail in the zoomed view.



Lines show Hazium peaks during normal office hours. Green line floor 3 around CEO Sanjorge's office at times when he is present.

## conclusions

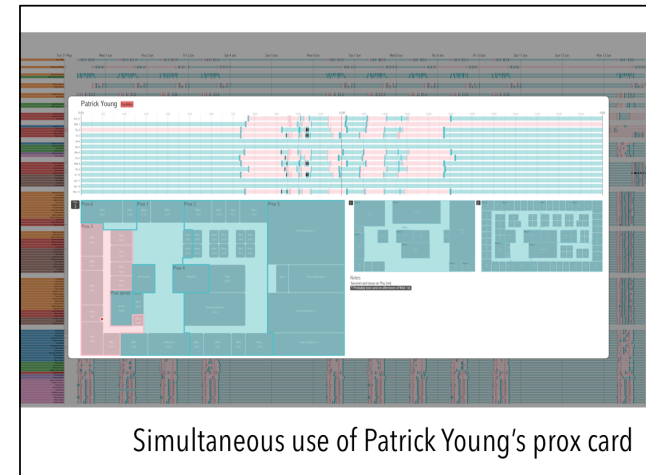
Transferrable observations...

Encoding reflects precision of data  
more expressive visual channels for higher precision data

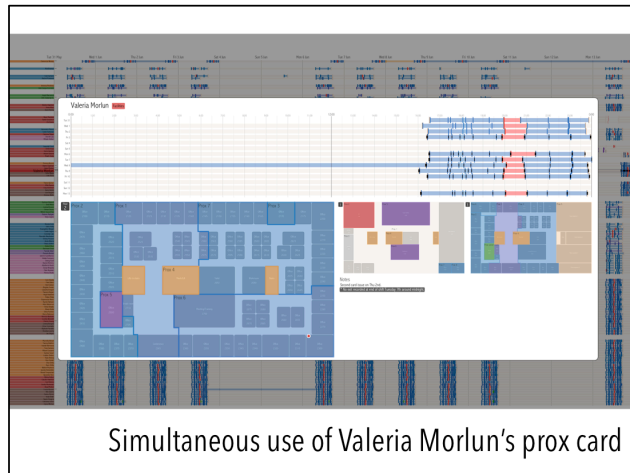
Common projection aids comparison  
timeline, CUSUM, z-scores

Detection of anomalies requires visualization of trends  
cluster sorting, common projection






Original analysis failed to spot that the first issue of Patrick Young's prox card was still active after his second card had been issued. These uses of the first issue card are highlighted in black here, but were not made salient during the original analysis. The framework allows such double entries to be symbolised, but a pertinent question is how do we design our Visual Analytics software to lead us to want to identify this in the first place?



Similar multiple use is seen in Valeria Morlun's prox cards. Next stage of analysis would be to trace the detailed movements of these double entries, with the possibility that Rosie the mobile prox robot would have revealed more detailed movement patterns (something designed out of the overviews above based on the general lack of spatial precision).

[gicentre.net/vastchallenge2016](http://gicentre.net/vastchallenge2016)

 jwoLondon

**Visual Analytic Design for Contextualising Sensor Data**

I will focus on two aspects of visual analytic design to address MC2 – those driven by the nature of the data and those driven by the nature of the MC2 tasks,