Oromocto | March 21, 2018
Word Count: 600

Clusters and Calendars Revisited: A Vision Science Review of Classic Visualization Work

Nearly 20 years ago, van Wijk & van Selow (1999) published a clever representation of time series data over multiple time scales. In this paper, the authors juxtapose a view of multi-colored line charts (depicting clusters of similar daily patterns of number of employees in the office, during a single day), with a calendar view showing the same information (Figure 1). The authors conclude that this interface helps users see average patterns of behavior over time, as well as identify anomalous events on specific days. The findings from this paper are frequently cited as classic examples of successful visualization work (Munzner, 2014; Google Scholar*). Despite its popularity, the paper lacks empirical analysis, offering no rigorous explanation for why the interface works. Here, we go beyond the original conclusion's appeal to reader intuition. Rather, we provide the missing perceptual account of this seminal work, in order to promote awareness about cognitive psychology, and improve evaluation and reporting in visualization research.

We first explain the success of the multi-colored cluster line charts. In this view, the authors claim two major tasks can be accomplished: a) recognizing that office hours are followed strictly, and b) identifying 7 unique patterns of behavior. Disappointingly, no reason is provided in the original text. Fortunately, Szafir, et al., (2016) recently showed that ensemble processes are used to understand line charts. Based on this work, Task a) can be achieved quickly via trend detection. Spatial segmenting and structure identification of ensembles helps users see the tight clusters of lines starting between 8:30-9:00 and at 17:00, representing times when employees arrive to and leave the office. Furthermore, the use of ensemble color feature mapping readily facilitates Task b).

Ensemble processes also account for pattern detection tasks in the calendar view, but van Wijk & van Selow's primary task for this encoding is the identification of anomalous work hours on December 5th. This is achieved through visual search (Wolfe, 1994). The unique shade of red guides viewer attention to that date. Interestingly, because of hue similarity, shape congruency with the other warm colored days, and the presence of another unique color for December 31st, December 5th does not fully pop-out of the display. Feature integration theory (Treisman & Gelade, 1980) suggests that the interface could be improved with increased hue saliency for the 5th and 31st. This could be accomplished with larger steps along a diverging color palette.

Finally, we explain the advantages of the cluster and calendar view juxtaposition. The authors note that each of their tasks can be understood and compared between both the calendar and line chart; but again, there is no justification. Side-by-side views provide a useful extension of the visual system (Munzner, 2009). Viewers do not need to hold a

*A quick search reveals that this paper has been cited over 400 times in visualization literature.

scene of hourly clusters in memory while they navigate to a new part of the interface for a weekly and daily view. Instead, they can make quick saccades between the visual encodings, examining less than 4 clusters, or days, at a time. This allows for pattern comparisons to be made fully within the bounds of viewer working memory capacity (Luck & Vogel, 1997).

To summarize our review, the designs and task realizations from van Wijk & van Selow (1999) can be elucidated with evidence from perceptual literature. In fact, we show that earlier consideration of vision science by the authors could have led to more successful design choices. In conclusion, vision science is useful, and visualization experts should reflect on this to publish more rigorous task design validation.

Figures

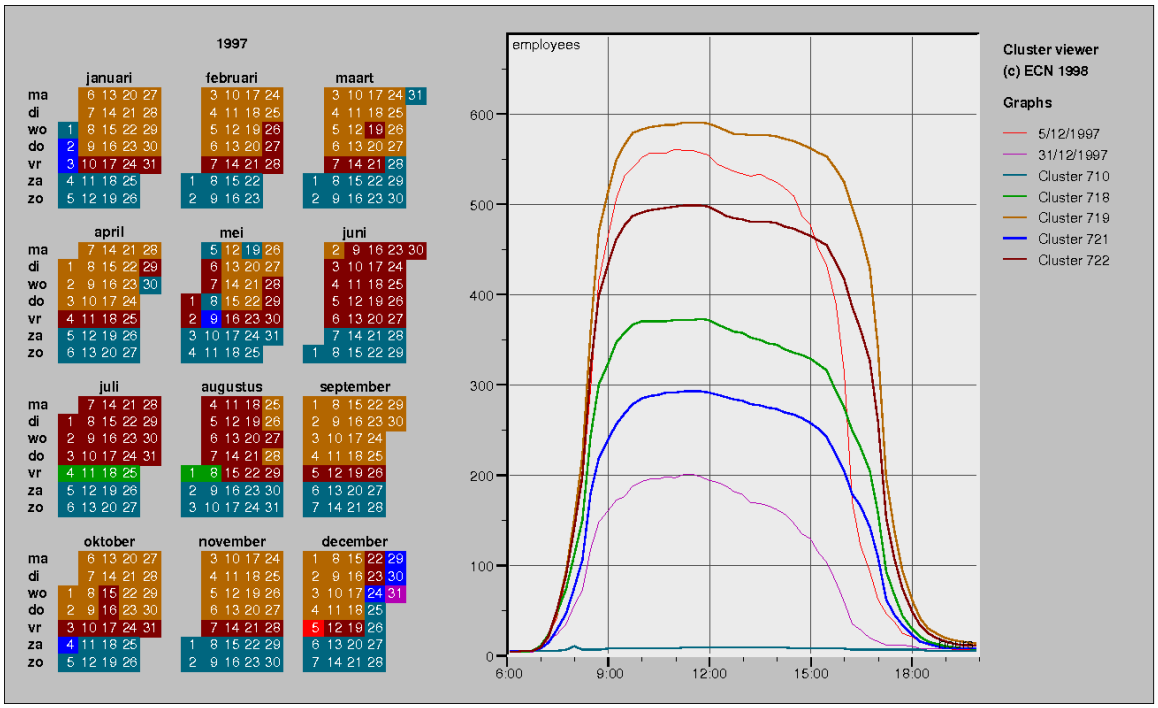

Figure 1. Calendar + Line Views from van Wijk & van Selow (1999): Cluster Visualization of Number of Employees in the Office based on Hourly Measured Observations.

References

Google Scholar Search: https://scholar.google.ca/scholar?hl=en&as_sdt=0%2C5&q=Cluster+and+Calendar+based+Visualization+of+Time+Series+Data&btnG=

Munzner, T. (2009). Visualization. In *Fundamentals of Computer Graphics* (pp. 675–707).

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of vision*, *16*(5), 11-11.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Van Wijk, J. J., & Van Selow, E. R. (1999). Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on* (pp. 4-9). IEEE.