MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Calendar-based graphics for visualising people's daily schedules

Earo Wang, Di Cook, Rob J Hyndman

June 2017

Working Paper no/yr

Calendar-based graphics for visualising people's daily schedules

Earo Wang
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: earo.wang@monash.edu
Corresponding author

Di Cook
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: dicook@monash.edu

Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: rob.hyndman@monash.edu

29 June 2017

JEL classification: C10,C14,C22

Calendar-based graphics for visualising people's daily schedules

Abstract

A brief summary of our ideas

Keywords: blah, blah

1 Introduction

Calendar-based visualisation proves useful in unfolding human-related activities in temporal context. For example, Van Wijk and Van Selow (1999) developed a calendar view of heatmap to represent the number of employees in the work place over a year, where colours indicate different clusters derived from the days. It effectively contrasts weekdays and weekends, highlights public holiday, and presents other known seasons like school vacations, all of which have influence over the turn-outs in the office. However, their technique is too constrained to colour-encoding graphics. We shall extend this calendar-based arrangement using linear algebra tools from a single type of heatmap to a broader range of glyphs, from univariate to multivariate cases, from one type of calendar to three types. The proposed algorithm has been implemented into the `sugrrants` package (**sugrrants**) using R (R Development Core Team 2017).

This work is partly motivated by studying hourly foot traffic in the city of Melbourne (City of Melbourne 2017). There have been 43 sensors installed across downtown Melbourne since 2009 in order to capture the pulse of city life. (FIGURE ?: weekday at multiple sensors) Figure (?) lends itself to addressing some of the issues: (1) to simultaneously exhibit multiple seasons including time of day, day of week, and day of year, (2) to visualise long historical time series data in an information-dense way at the overview level, (3) to compare and contrast multiple time series on a common scale, (4) to quickly look up the date when there are unexpected events happening without interactivity.

2 Construction

(NEED A BRIEF INTRO HERE)

2.1 Calendar grids

The algorithm for constructing a calendar plot uses linear algebra, similar to that used in the glyph map displays for spatio-temporal data (Wickham et al. 2012). To make a year long calendar, requires cells for days, embedded in blocks corresponding to months, organised into a grid layout for a year. Each month can be captured with 35 (5×7) cells, where the top left is Monday of week 1, and the bottom right is Sunday of week 5. These cells provide a micro canvas on which to plot the data. The first day of the month could be any of Monday-Sunday, which is determined by the year of the calendar. Months are of different length days, ranging form 28-31, and each month could extend over six weeks but the convention in these months is to wrap the last few days up to the top row of the block. The notation for creating these cells is as follows:

- $k = 1, \dots, 7$ is the day of the week that is the first day of the month
- $d = 28, 29, 30$ or 31 representing the number of days in any month
- (i, j) is the grid position where $1 \leq i \leq 5$ is week within the month, $1 \leq j \leq 7$, is day of the week.
- $g = k, \dots, (k + d)$ indexes the day in the month, inside the 35 possible cells

The grid position for any day in the month is given by

$$\begin{aligned} i &= \lceil (g \mod 35)/7 \rceil, \\ j &= g \mod 7. \end{aligned} \tag{1}$$

Figure 1 illustrates this (i, j) layout for a month where $k = 5$.

To create the layout for a full year, (m, n) denotes the position of the month arranged in the plot, where $1 \leq m \leq M$ and $1 \leq n \leq N$. Between each month requires some small amount of white space, label this b. Figure 2 illustrates this layout.

Each cell forms a canvas on which to draw the data. Consider the canvas to have limits $[0, 1]$ horizontally and vertically. For the pedestrian sensor data, within each cell hour is plotted horizontally and count is plotted vertically. Each variable is scaled to have values between $[0, 1]$, using the minimum and maximum of all the data values to be displayed. Let h be the scaled hour, and c the scaled count.

Then the final points for making the calendar line plots of the pedestrian sensor data is given by:

				k=5, g=5 *i=1, j=5*	*g=k+1* *i=1, j=6*	*g=k+2* *i=1, j=7*
g=k+3 *i=2, j=1*	*g=k+4* *i=2, j=2*	*g=k+5* *i=2, j=3*	*g=k+6* *i=2, j=4*	*g=k+7* *i=2, j=5*	*g=k+8* *i=2, j=6*	*g=k+9* *i=2, j=7*
g=k+10 *i=3, j=1*	*g=k+11* *i=3, j=2*	*g=k+12* *i=3, j=3*	*g=k+13* *i=3, j=4*	*g=k+14* *i=3, j=5*	*g=k+15* *i=3, j=6*	*g=k+16* *i=3, j=7*
g=k+17 *i=4, j=1*	*g=k+18* *i=4, j=2*	*g=k+19* *i=4, j=3*	*g=k+20* *i=4, j=4*	*g=k+21* *i=4, j=5*	*g=k+22* *i=4, j=6*	*g=k+23* *i=4, j=7*
g=k+24 *i=5, j=1*	*g=k+25* *i=5, j=2*	*g=k+26* *i=5, j=3*	*g=k+27* *i=5, j=4*	...	...	*g=k+d* *i=5, j=7*

Figure 1: *Illustration of the indexing layout for cells in a month.*

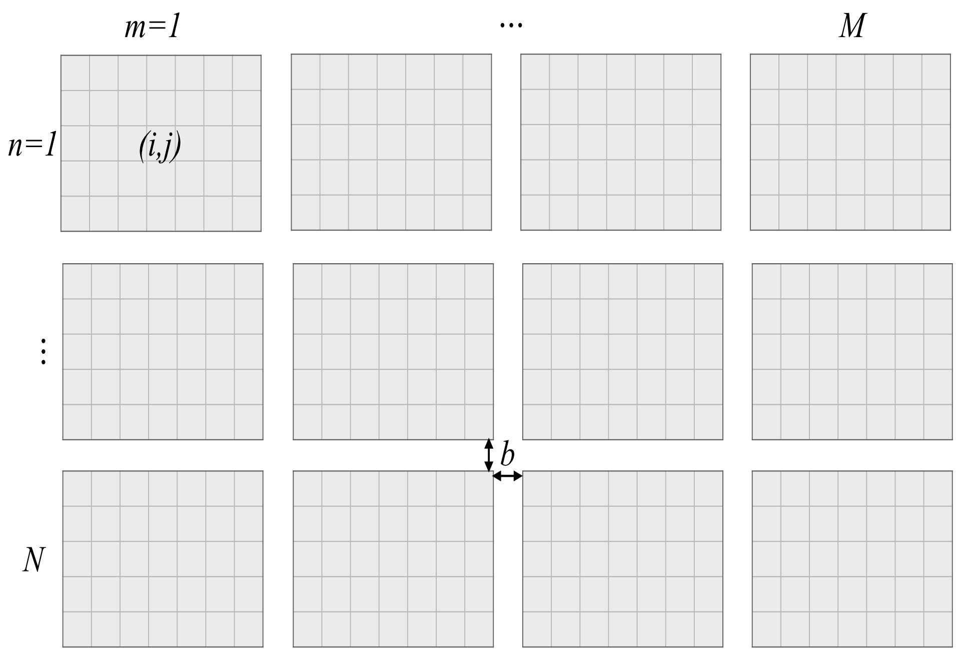

Figure 2: *Illustration of the indexing layout for months of one year.*

$$
\begin{aligned}
x &= i + (i-1) \times m + (m-1) \times b + h, \\
y &= -j - (j-1) \times n - (n-1) \times b + c.
\end{aligned} \tag{2}
$$

Note that for the vertical direction, the top left is the starting point of the grid, hence the subtractions, and resulting negative values to lay out the cells. Within each cell, the starting position is bottom left.

The algorithm can be extended relatively easily to layout multiple years, or to lay out different calendar formats like weeks of a year.

2.2 Scales

2.3 Reference lines and labels

The major reference lines are placed on the far left and the bottom for every month panel: for each m, the vertical separation is $\min(x)$; for each n, the horizontal separation is $\min(y)$. The month labels located on the top is obtaining $\max(y)$ for each n. The minor reference lines for every daily grids are placed on the starting positions: for each i, the vertical line is $\min(x)$; for each j, the horizontal line is $\min(y)$.

(INTERACTIVITY COULD BE DISCUSSED HERE) We shall enable interactivity in the calendar-based graphics for time series data. It will allow users to transform different temporal components and switch displays between overlaying and faceting through key strokes or mouse clicks.

3 Applications

The calendar-based visualisation provides data plots in the familiar (at least for the Western world) format of an everyday tool. Special events for the region, like Anzac Day in Australia, or Thanksgiving Day in the USA, more easily pop out to the viewer as public holidays, rather than a typical work day.

This sort of layout may be useful for studying consumer trends, or human behaviour, like the pedestrian patterns. It may not work so well for physical patterns like temperature, which are not typically affected by human activity.

4 Conclusions

The limitation is also evident: hard to perceive trend as not on the common scale.

References

City of Melbourne (2017). *Pedestrian Volumn in Melbourne*. Town Hall, 90-120 Swanston Street, Melbourne VIC 3000. http://www.pedestrian.melbourne.vic.gov.au.

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Van Wijk, JJ and ER Van Selow (1999). Cluster and Calendar Based Visualization of Time Series Data. In: *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*. IEEE, pp.4–9.

Wickham, H, H Hofmann, C Wickham, and D Cook (2012). Glyph-maps for Visually Exploring Temporal Patterns in Climate Data and Models. *Environmetrics* **23**(5), 382–393.