# Response to reviewers

JCGS-18-293 23 April 2019

Thanks to the three reviewers for some very helpful comments which have led to many improvements of the paper. My response are in black, the reviewers' comments are in red.

## Reviewer 1 comments

Figure 2 and Figure 10 show much the same information. You can see the Australian Open period in both. There is more detail in Figure 10 for the State Library. The gaps in the Birrarung Marr data are easier to see in Figure 2.

We add a new subsetion '2.4 Reasons to use calendar-based graphics' to explain why to use calendar plots, and what information revealed by calendar plots is but not revealed by other plots. The calendar display doesn't replace traditional displays, but serves to complement to further tease out interesting structures in the data.

Figure 11 is not described well. What does the V-pattern mean? How can you see the strong autocorrelation at weekends? There are only dots.

We have switched to use the local scaling for the lag scatterplot (as Figure 10 now) to accurately represent the autocorrelation. The strong autocorrelation is clearly revealed at weekends.

We have elaborated more about the V-shape on weed days.

The colours are not well chosen in Figure 15. It is hard to see the averages for households 1 and 3.

Done.

There are too many large plots in the paper. Figure 6 is not needed and Figure 12 is not needed. How do you know that the New Year's Eve outliers in Figure 12 are from the same sensor? That is not in the plot.

Thanks for correcting it. We have pointed out the variability instead of the outliers on New Year's Eve.

Do Figures 16-19 show much? Drawing all four series as rows in facets one above the other would show the holidays better. Comparisons of the households would be easier. The six-weeks months confuse looking at holidays. What are the one-day peaks in the middle of holidays?

ToDo.

There is a shiny app that presents some of the data. It is not well described. The plot at lower left is cumulative precipitation for each month, but it does not say so. Why is cumulative precipitation by month interesting? Months are of different lengths. You can select single days in the time series. Can you select more than one day?

ToDo.

## Reviewer 2 comments

p.10, l.1-15 It wasn't clear to me why the subtraction was performed in the equation (2) because I couldn't find any description for the global coordinate system of the canvas. The value of y could be negative. Please add the description how to deal with the coordinate system, especially the y-axis.

The reason why the subtraction is performed was originally given in the following paragraph of Equation (2).

## Reviewer 3 comments

p1, l32: 'information, and events.' - remove the comma

We have removed the comma and elsewhere, and adjusted the language based on the comments.

p2, l9: 'grammar of graphics' - provide reference immediately here and not only later in the article and explain what the grammar of graphics is; also 'piped into' may not be understood by outsiders

Done.

Fig.1 (& related figs): I suppose you use a '3-class Dark2' color scheme from Rcolorbrewer? This does not work well when printed in gray scale: The colors become almost indistinguishable. ...

We have changed the colour scheme from 'Dark2' to '4-class PuOr' (without faint orange) in Figure 1 and all consecutive figures.

Fig.4: Colors orange & purple are hard to distinguish in grayscale. Moreover, you should use different colors here as these 2 colors are already related to the 3 stations.

We have changed the colour to red and blue picked from a '4-class RdYlBu' colour scheme, which is different from the colours used for the 3 stations.

Fig.2: Instead of 'Jan 2016' ... 'Jan 2017', be specific and list 'Jan 1, 2016' ... 'Jan 1, 2017'

Done.

Fig.2: Can you match the minor grid lines with the start of a month, rather than the middle of a month? This will make it much easier to identify the approximate dates for some of the other spikes. No need to label these minor grid lines.

Done.

Fig.2: You mention 'small multiples' only in a figure caption. This concept is a central part of your calendar graphics and should be summarized in more details (including references) in the main text. Also mention Unwin & Valero-Mora's 'Ensemble Graphics', JCGS, 27(1), as a major concept that applies to both sets of related figures in your article.

We have explained the idea of faceting and also referred to small multiples, trellis charts, and ensemble graphics at the end of the second paragraph in Introduction section.

Fig.3: Add tick marks at 0 and 24. See whether tick mark labels fit. If not, OK to omit those for 0 and 24.

Done.

p6, l8: You mention ggTimeSeries & ggcal. For completeness, also cite Jones (2016) Calendar Heatmaps, https://rpubs.com/haj3/calheatmap, and possibly Wong's TimeProjection R package, https://cran.r-project.org/web/packages/TimeProjection/index.html

Done.

p6, l49-. . . : Starting a section called 'Data transformation' with an example/figure seems to be strange. Can you first start with the formal steps and then place this example/figure after the formal part, i.e., around p.8, l38.

Done.

Fig.4: Colors orange & purple are hard to distinguish in grayscale. Moreover, you should use different colors here as these 2 colors are already related to the 3 stations.

Done.

Fig.4: Prior to reading the text on p8, I was really confused and even assumed there was a major bug in your R code. Having 2 days in early May and 1 day in early Oct, then 3 days missing, and then the remainder of the month is really confusing for someone who only looks at the figure without reading the text in detail. At least mention this layout anomaly in the figure caption.

Thanks. I have pointed out the layout anomalies for May and October in the Figure 6 caption (previously Figure 4).

p8, l19: 'wrap the last few days up to the top row of the block': This answers my comment for Fig.4 now, but this layout still remains misleading. We expect to see some similar temporal pattern in nearby graphs. But there could be considerable differences over a 30-day period, e.g., in your example from Section 3: What if these 1 or 2 days are after some summer vacation with lots of air conditioning use, but the vacation already started in the previous month and continued through the middle of the current month. Suddenly, there will be a few huge spikes that interrupt the overall low-energy pattern. I could think of 2 possible solutions: (i) Add a 6th week for each month; or (ii) Add the extra days to the start of the next month. This is sometimes called a 'Calendar Heatmap Tetris Chart', see for example https://stackoverflow.com/questions/27000131/calendar-heat-map-tetris-chart [both of these features could become additional user options for your R function]

It is intentional to conform a calendar layout of 5 rows and 7 columns for the purpose of compactness. Only one or two months in a year will span over 6 weeks, which would result in more unused space for a year-long calendar plot. We don't shift the extra days to the next month because it is difficult for readers to distinguish which month contains these additional days without proper labelling. But we agree that it is disruptive since time is usually perceived as linear.

That said, the algorithm that lays out the calendar is flexible to incorporate these two options. We have included them as the options for future work. Thanks for the suggestions.

p8, l44: 'Between each month requires some small amount of white space, denoted by b.' - strange sentence; rephrase

Done.

p11, l44: 'star plots': briefly explain what these are and cite a basic reference for these.

Done.

Fig.7: Where is 'noon' and 'midnight' here - on top/bottom or on the 0/180 degree position on the right/left side? And what is the direction? I suppose clockwise, but this needs to be mentioned somewhere.

The star plots have been removed from the paper, since it doesn't add more information to other existing plots.

p13, l39: 'Figure 13 shows the same plot as Figure 12 labelled': You can't jump forward to a figure before all intermediate figures have been introduced. The previous fig was Fig.8. Figs 9-11 have not been mentioned yet. Either rearrange your figures so that 12-13 become 9-10, or use Chinese characters for one of the previously introduced figs (1-8).

Sure. 'Figure 13 shows the same plot as Figure 12 labelled' is deleted.

p13, possible subsection 2.2.6: You seem to have a 'sunday = FALSE' default. If TRUE, does this make Sunday the start of the week (as used in the US & Canada)? If so, mention this as an option. No need for a figure.

Done. We deprecate argument `sunday` in favor of `week_start` that appears more general. We added a new subsection on 'Start of the week' to briefly explain the option. We have reorganized the order of the subsections to be consistent with the function argument's ordering.

p13, possible subsection 2.2.7: How can you enter specific holidays? These are different from country to country, e.g., based on national holidays such as Thanksgiving, Labor Day, independence days, religious holidays, etc.

We used `tsibble::holiday_aus()` for generating Australian public holidays, prior to creating calendar displays. Whether to use holidays or not is a data-specific problem, which is beyond the scope of the calendar plot.

Fig.8: 'impossible to compare the size of peaks between days.' - Correct; since this is based on ggplot, could you use colors from a sequential color scheme that maps the counts, e.g., from faint yellow (0) to dark red (for the overall max)? This might be another useful argument for a future extension of your function.

ToDo.

p15, l12: 'idea of faceting': explain what 'faceting' means in general (and cite 1 or 2 main references)

We added the explanation of 'faceting' in the Introduction section.

p15, l16: 'In particular, it can be immediately learned that when Birrarung Marr was busy and packed, for example Australian Open in the last two weeks of January.' -> 'In particular, it can be immediately learned that Birrarung Marr was busy and packed, for example during the Australian Open, a major international tennis tournament, in the last two weeks of January.'

Done.

p15, l45: 'the day before Christmas, go shopping on the Boxing day, and stay out for the fireworks on New Year's Eve.': Some of these names may not be known to non-Christian / non-British readers. List the date in parentheses, e.g., December 24, December 26, etc.

Done.

Fig.12: 'loess smooth': cite main reference

Done.

p21, l36: 'these four households are the data of colleagues of the authors.' - strange sentence; rephrase

Done.

p22, l17-35: inconsistent use of times, e.g., 'before 6', 'around 18', '3pm', etc. - adjust to times used in the figures and be consistent. Also adjust caption of Fig.15.

Done.

p22, l36: 'Figure 16, 17, 18 and 19' -> 'Figures 16-19'

Done.

'individually for each household': This seems to suggest that a scale from 0 to max(household_i) [but not to max(all households)] has been used. But what is this max in each of the 4 figures? The reader has no way to easily infer this. Also mention that these 4 figs are not true small multiples as the vertical scales for the 4 households differ from figure to figure.

ToDo.

Fig.15: Use labels 6, 12 & 18 and extra tick marks at 0 & 24, matching Fig 3.

Done.

p25, l3: What about April? If added here, also add to caption of Fig.19.

Done.

p25, l14: 'Anzac Day in Australia, or Thanksgiving Day in the USA,': Never mentioned before in this article. Refer back to the special events/holidays from earlier in the article.

ToDo.

p25, after l35: List my suggestions for software additions from above that you do not want to implement now as options for future work.

ToDo.

Fig.16 caption: 'Calendar display for household 1, indicates higher weekend usage, and in the summer months, November-February. It seems that they took a vacation in June.' -> 'The calendar display for household 1 indicates higher weekend usage and higher usage in the summer months (November-February). It seems that this household took a vacation in June.'

Done.

Fig.17 caption: 'Calendar display for household 2, reveals their tendency to use air conditioning and heating continuously. Not many vacation were taken.' -> 'The calendar display for household 2 reveals their tendency to use air conditioning and heating continuously. Some vacation days were taken in late December and in the second week of June.'

Done.

Fig.17: I am concerned: Is this a 'fair' comparison when you drop the months August-October (& part of November) 2017 for household 2, while these 3+ months are part of the graphics for households 1, 3 & 4? Wouldn't it be better to restrict all households to Dec 2017 - July 2018? This would drop the winter/spring months from the aggregated displays in Figs.14-15. But household 2 still would have most variation, but now, this is based on matching time periods.

ToDo.

Fig.18 caption: 'Calendar display for household 3. Their energy use reveals higher energy use in the winter months, with multiple peaks daily on both week days and weekends. There are some high peaks in summer, perhaps indicating occasional air conditioner use. There have been several long vacations in the past year.' -> 'The calendar display for household 3 reveals higher energy use in the winter months (May-August), with multiple peaks daily on both week days and weekends. There are some high peaks in summer (November-February), perhaps indicating occasional air conditioner use. There have been several long vacations ... [match the main text].'

Done.

Fig.19 caption: 'Calendar display for household 4, shows energy use mostly in the evenings and on weekends. Three short trips were taken in October, December, and June.' -> 'The calendar display for household 4 shows energy use mostly in the evenings and on weekends. Three short trips were taken in October, December, and June.' [also April? - match with p25, l13]

Done.

p.30: City of Melbourne (2017) [& other URLs]: Does JCGS require the 'last accessed' information for web-based references? If so, add where needed.

Done.