Review of JCGS-18-293

"Calendar-based graphics for visualizing people's daily schedules"

by Earo Wang, Dianne Cook & Rob J Hyndman

Nice manuscript. Well written overall. I have 3 main types of minor comments regarding general readability, information for non-statisticians, and possible (future) extensions of the software. If you don't like my language adjustments, make further adjustments yourself.

In my opinion, this article is also of considerable interest for readers from outside of the Statistical Graphics community. For these readers, some detailed background information (including references) is often missing. See details below.

Specific comments:

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

-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

-p2, l14: "new package" -> "new R package" [also add R elsewhere, e.g., p2, l44 "geofacet R package"; p25, l43: "sugrrants R package"]

-p2, l23: "till" -> "until" [same later on, e.g., p22, l22]

-p2, l30: "The faceted plots in Figure 2, give" - remove the comma

-p2, l50: "to digest multiple seasonalities, and special events." - remove the comma

-p3,l45: "Birrarung Marr:" - use "--" (same as for the 2 other stations)

-Fig.1 (& related figs): I suppose you use a "3-class Dark2" color scheme from Rcolorbrewer? This does not work well when printed in grayscale: The colors become almost indistinguishable. How does a "4-class PuOr" work in grayscale, leaving out the faint orange (#FDB863)? The 3-class version of PuOr has a faint gray that may be hard to see on a white/gray background. If this proposed scheme (or any other scheme you can think of) works on grayscale, then also adjust the colors in all consecutive figs.

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

-Fig.2: Can you match the minor gridlines 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 gridlines.

-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.

-Fig.3: Add ticmarks at 0 and 24. See whether ticmark labels fit. If not, OK to omit those for 0 and 24.

-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

-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.

-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.

-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.

-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]

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

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

-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.

-p13, l28: "swapped in the Equation 1." -> "swapped in Equation (1)."

-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).

-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.

-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.

-p13, l49: "The comparison of sensors can be done by overlaying plot the values for each" - strange sentence; rephrase

-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.

-p15, l5: "distinct temporal trend" -> "distinct temporal pattern"

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

-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."

-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.

-p15, l51: "calendar-based display" -> "calendar-based displays"

-Fig.12: "loess smooth": cite main reference

-p21, l7: "Options in function frame calendar" -> "Options in the frame_calendar function "

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

-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.

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

-p22, l36: "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.

-p22, l38: "you can see" -> "we can see"

-p22, l43: "summer months" -> "summer months (November-February)" ?!? Many readers
likely will be from the Northern hemisphere.

-p22, l45: "evident in the mid July; while household 1 uses" ->
"evident in mid July. In contrast, household 1 uses"

-p22, l52: "holidays during Christmas and the second week of June." ->
"vacation in late December and in the second week of June."

-p22, l54: "September till early October and in June/July," ->
"September until early October and in June/July."

-Fig.15: Use a different color scheme here as this is no longer related
to the 3 stations from Sections 1 & 2.

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

-p25, l3: "and household 4" -> "Household 4"

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

-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.

-p25, l31: "summaries and deviations, in order" - remove the comma

-p25, l34: "faceting method, with formal labels and axes." - remove the comma

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

-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."

-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."

- 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.

-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]."

-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]

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