Response to reviewers

JCGS-18-293-R1 18 October 2019

We thank the reviewers for their helpful feedback which has improved the paper. The point by point description of changes are below: the reviewers' comments are in red and our response is in black.

Associate editor

What if the data's collected over different time zones?

We added a short paragraph in Section 2.1 after 'Note that for the vertical direction' to address the time zone problem.

'The R package, **lubridate**, is used to extract the components of time, such as days of the week and the number of days in a month, that create the layout. These time variables are converted to integers for the modular arithmetic. Note that for any date-time information is associated with time zone. If your data is collected over multiple time zones, you will need to convert them to the same time zone before conducting any temporal analysis.'

Abstract: 'modular arithmetic' is more accurate. Or omit the sentence altogether. I agree with comments of reviewer 1 on this topic.

Thanks. We've revised to 'modular arithmetic'.

Page 2 par 2 first sentence. There are other broader applications than this, eg, pollution monitoring. Last sentence 'but conventional displays' (omit 'the')

We've rephrased 'people's daily schedules' to 'human activities' that would include pollution monitoring.

Page 3 Figure 1 legend. Say 'grey dots indicate sensor locations'.

Done.

Page 6 line 4. omit 'as' in 'as in R packages'.

Done.

Par 2. rephrase 'more easily pop-up to viewers'.

We've rephrased to 'to accentuate unusual patterns, such as those related to special events, for viewers'.

Page 7. Second line. You might want to add that fifth row could be blank. Wrapping around the last few days could be confusing. Could this be made optional?

Thanks, we added the note 'The fifth row could be blank for February if the month starts on Monday.' in the paragraph. We have mentioned the option of 6-week format as part of the future work.

Should be g= k, ... (k+d-1) I think.

Thanks for pointing it out. Corrected.

Equation 1. Modular arithmetic will give j in 0..6, not in 1..7.

Thanks. Added the missing $+1$ back.

Line 36, what are M and N? You explain them later, but should be at first usage.

Done.

Page 8. I agree with reviewer 1. These figures are not needed. Move to supplementary material.

We dropped figure 4 and 5 (diagrams for laying out the calendar).

Figure 6. Could the figure have some indicator of the wrapping of few days to first row? It would help the reader.

We noted this in the figure caption. If adding some indicators, these indicators would possibly receive more attentions from the viewers, thus deemphasizing the main message.

Page 11 line 1. Write 'For example in Figure 6. . . '

Done.

Page 12. Say why polar transformation is useful.

We added 'It is most useful in exhibiting cyclical patterns in the data.' at the end of paragraph.

In your figures is global scaling used, unless otherwise stated?

Yes.

Page 17, figure 10. I don't see negative correlation. Just two slopes, both positive. Am I missing something?

We deleted the sentence, 'This is due to the morning and afternoon commuter patterns in Figure 6: the adjacent hours are positively correlated when approaching to the peak hour but negatively correlated when moving away from the peak.', to mitigate the confusion.

Page 19, first sentence. The meaning of this is not clear. Rephrase.

Done. We rephrased to 'The previous calendar plots are static, made with **ggplot2**.'

Page 21, figure 12. I agree with reviewer 1 that this needs some tweaking. Too much vertical space given to panel labels. Time labels are confusing. Two of the line colours are very similar here and in other plots.

We've reduced the size of panel labels to give more room to the data display. Time labels are now made to show 6 and 18 to avoid confusion.

The colour scheme is suggested by reviewer 3 before, which is colour-blind and grey-scale print friendly.

Page 23. Suggest dropping figure 13. Or move to supplementary material.

We dropped the boxplots as in figure 13, and the corresponding paragraph.

Page 24 Line 15: 'calendar plots tell the stories'

Done.

Reviewer 1

'Would it be possible to have a plot for each day in a calendar plot with thick lines separating the months? That kind of plot is common for calendar heatmaps.' I think that is important. Perhaps they can build it in. If not, why not?

The white space between months already emphasises the separations between months. The lines separating the months are also thicker than the ones separating the days in the calendar plot. These sit in the background to make the data plot more visible, which follows the Tufte's rules.

The authors mention 'modulus arithmetic' three times and linear algebra once, so I expected complicated mathematics. The mathematics is simple. These remarks are not needed. Most people say 'modular arithmetic'.

Done.

p6 'Jones (2016), Wong (2013), Kothari & Ather (2016), and Jacobs (2017) implemented some variants of calendar-based heatmaps as in R packages: TimeProjection, ggTimeSeries, and ggcal respectively.' Four articles, three packages, one must be missing. Better when references are next to package names.

We removed the Jones (2016), since it's an online post. We also rephrased the sentence so that references are next to package names.

p7 'Each month conforms to a layout of 5 rows and 7 columns, where the top left is Monday of week 1, and the bottom right is Sunday of week 5 by default.' This is not true for six-week months.

Thanks. We revised this sentence to 'Each month conforms to a layout of 5 rows and 7 columns, where rows and columns refer to the week of month and the day of week respectively.'

p7 'Each variable is scaled to have values in [0,1], using the minimum and maximum of all the data values to be displayed, assuming fixed scales.' Can the scale be set by the user? Starting at 0, not minimum, can be useful.

Yes, it is possible with `facet_calendar()`.

p8 Figures 4 and 5 are unnecessary. Figure 4 is misleading, it does not apply to six-week months and the bottom right cell may be empty.

We dropped figure 4 and 5 (diagrams for laying out the calendar).

p10 Figure 6 Could the vertical scale be given in the caption?

We added the range of vertical scale in the caption.

p14 and 15-16 Figure 9 is better than Figure 8. Are both necessary?

We kept both figures in, to show what users can do with the calendar plots, for example both overlaying and faceting.

p14 'Almost any kind of plot can be shown in a calendar pane.' This is excellent. An impressive example would be useful. Figure 10 is disappointing. It is difficult to read and the information it provides could be seen in earlier figures.

We have rephrased this sentence to 'Many types of plot can be shown in a calendar pane, by taking advantage of the existing **ggplot2** plotting capabilities.'. You have inspired us to try to make more exciting calendar plots, and they are possible but not useful. The lag scatterplots (now figure 8) is the sensible data plot for the pedestrian dataset, that fits into the Section 2.3.3. To make the information easier to read, we plot a subset from week 1 to 17 instead of the whole year before.

p17 Figure 10 Why not align by days of the week?

We have changed figure 10 (now figure 8) to align by days of the week instead of by days of the month. Since the best display for the pedestrian dataset is monthly layout, we wanted to showcase other possible layouts.

p19 Text implies that more than one date can be selected in the Shiny app. The app was kludgy and did not work smoothly. It is a nice idea, but interaction has to be fast and work well. Zooming did not work properly and was not able to link from the weather plots. Fix or do not mention in article.

We deleted the sentence 'Conversely, selecting on weather data plots, linked to the calendar can help to assess if very hot/cold days and heavy rain, affect the number of people walking in downtown Melbourne.' from Section 2.3.4.

p21 Figure 12 is an interesting possibility, but not well implemented. The repetition of the scales in unhelpful. 0 12 24 looks like 240 12 240. There is too much labelling and too little information.

We've reduced the size of panel labels to give more room to the data display. Time labels are now made to show 6 and 18 to avoid confusion.

The case study is interesting. Why are the data for household 2 different (fewer possible values)?

Data is of the same time lengths (and the same amount of data points) across four households. Different households have different lifestyles.

p22 'with noticeable variations on Thursdays' based on only 26 data points is over-interpretation.

It's half-hourly data, meaning 26 Thursdays * 48 per day = 1248 data points. But we dropped the boxplot as in figure 13 including this sentence.

p22 'The other two households (1 and 4) tend to consume more energy with more variation on the weekends relative to the week days, reflecting work and leisure patterns.' The vertical scaling makes it very difficult to see this. Inspecting the data with individual scales offers some support but not a lot.

We dropped the boxplot as in figure 13, and the corresponding paragraph including this sentence.

p24 'All households have different week days versus weekends daily routines.' This does not look so true for households 3 or 4.

We removed this from the figure caption.

Reviewer 3

general: please check (multiple times): Shouldn't it be 'weekday[s]' (without a space) instead of 'week day[s]' (or speak of 'day of the week' if that is intended). Major online dictionaries only list the 'weekday' spelling.

Thanks. Done.

p2, l9: 'restructuring of the data, that can' - remove the comma

Done.

p7: In your (i, j) grid position notation, make clear immediately that i is the row and j is the column, with (1, 1) being in the upper left corner. This becomes apparent from the figures, but not from the text itself. Same for (m, n). Your first paragraph on p9 comes too late and only provides part of the explanation.

Done.

p10, l48: 'Note that, the algorithm' - remove the comma

Done.

p10, l50: 'which occurs to May and October.' -> 'which occurs in May and October.'

Done.

p11, l5: 'The date argument specifies the date variable used to construct the calendar layout.' Not very informative. Either omit or provide an example, e.g., date in which format (numerical, text, ...), from-to dates, etc.

We elaborated a bit on this. The `date` variable tells the range of dates plotted in the calendar.

Figs 6-9 captions: You speak of 'line glyphs', 'line graphs', and 'line charts'. These mean the same - so be consistent and use just one of the three terms throughout the figure captions (and main text).

Done. We used 'line graphs' consistently throughout the manuscript.

p22, l23: You wrote 'boxes for household 2 indicate greater variability in daily energy consumption with noticeable variations on Thursdays,' – isn't this on Fridays?

Yes, fixed.

p22 vs. Fig 14: You speak of 6am, 6pm, 3pm, etc. in the text, but use 0, 6, 12, 18, 24 hour labels in the graph. Be consistent and use the international 0-24 hour notation in the text as well.

Done.

Fig 15: It appears that April 30 is wrapped to the start of the monthly display? If so, add a comment similar to the one from Fig 6 to the figure caption.

Done.

p26, l20: 'and the another one starting' - delete 'the'

Done.

Kothari, A. & Ather (2016). I visited https://github.com/AtherEnergy/ggTimeSeries. Ather seems to be a company, so perhaps this needs to be listed as 'Ather Energy Pvt Ltd' as on the parent github page?!?

Done.