Response to reviewer comments for paper: 'brolgar: An R package to BRowse Over Longitudinal Data Graphically and Analytically in R'

Nicholas Tierney & Dianne Cook

We thank the editor, associate editor and reviewers for their feedback on our paper. We have made all the requested changes, and we have also migrated the paper into the new `rjtools` format, and have taken advantage of the html output to make Figure 10 interactive. Please see below for our response to each comment from reviewers.

Article:

R1. The paper is rather clear although I think that certain statements are not so precise. I am referring, in particular, to the first two sentence of Section "Longitudinal data structures" where it is said that longitudinal data, time series, and so correspond more or less to the same data structures. I do not completely agree because, for instance, a time-series is typically a long series of observations referred to the same entity, whereas longitudinal data typically consist of short series of observations for many units. There are also differences between longitudinal and panel data. Please expand this part clarifying the several data structures.

We agree with the reviewer that there are differences between these data structures, which are not always clear. We have made changes to the first paragraph of the paper and to the first paragraph of this section that we hope clarifies the terms. It now reads:

> Longitudinal data is a sibling of many other temporal data forms, including panel data, repeated measures, and time series. The differences are many, and can be in data collection, context and even the field of research. Time series are usually long and regularly spaced in time. Panel data may measure different units at each time point and aggregate these values by a categorical or key variable. Repeated measures typically measure before and after treatment effects. We like to think of longitudinal as measuring the same individual (e.g. wage earner) over time, but this definition is not universally agreed on. Despite the differences, they all share a fundamental similarity: they are measurements over a time period.

R2. I think that also package pml can be mentioned together with panelr and other packages in Section "Background" because it provides formats for longitudinal data.

I believe the reviewer is referring to the `plm` package? https://github.com/ycroissant/plm. We have added information about this package.

R3. Overall this extension should be expanded.

We are not sure what extension the reviewer is referring to in this case. We think it might be that we referred to extension functions in `tsibble`. We have changed the wording to be "modified"" because this better reflects the type of functions.

R4. The type of observed variables is not clarified. I mean that it is given for granted that the variables represented in the plots are quantitative (typically continuous). Are there special tools for discrete/categorical and in particular binary variables (that may be difficult to be represented graphically).

The data that we are using is continuous, and the same methods work to a degree with binary or categorical data. However, there are some issues with binary and categorical data. We have added a paragraph in the final section of the paper discussing the issues around this and what some approaches could do to help resolve this. The section reads as:

> Visualizing categorical or binary data over a time period can be difficult as the limited number of values on the y axis leads to overplotting. This can conceal the number of values present at a given value. The tools discussed in `brolgar` facilitate this in the form of `facet_sample`, and `facet_strata`. Some special methods could be developed to add jitter or noise around these values on the y axis, while still maintaining the graphical axis and tick marks.

R5. Among the packages that may benefit from the proposed one, others can be mentioned such as LMest.

I am not sure where the reviewer is suggesting we mention this R package, but the LMest package, while relevant to fitting longitudinal data, does not have an emphasis on data structures like the other examples (panelr, pmdplyr) do. While this is a great package for fitting longitudinal models, in this paper we do not think it is feasible to have an extensive list of all R packages that work with longitudinal data.