Response to Review

animbook: Visualizing Changes in Performance Measures and Demographic Affiliations using Animation

Krisanat Anukarnsakulchularp, Dianne Cook

Put in alt-text

Write short response

We thank the editor and reviewers for reading and offering improvements for our article. These are the changes we have made in response. *The reviewer comment is in italics*, and our response is regular text.

Reviewer 2

General

The article explores the use of animated ggplot2 plots with the animbook package. This package was inspired by an animated graphic from the New York Times and has the benefit of helping to visualize trends over time. The package takes a three stage approach, using a data cleaning step, then graphing step, and finally an animation step. The end result is an animated graphic that shows the relationship of categorical data over time.

Response: Nothing to change.

Detailed comments

Overall, the authors did a good job at guiding the narrative in a logical fashion. One major recurring theme that I noticed was that paragraphs tended to be short and had uneven transitions. My comments are below and follow the progression of the paper.

- *Page 1, second paragraph: this paragraph is really short and would benefit by describing the animation in detail.*

Response: We have added a short description of the New York Times animation.

- *Page 2, paragraph starting with Hvitfeldt: it might be helpful to include parts of the example from Hvitfeldt, or at least describe what parts of the example inspired animbook*

Response: We have emphasized that his example code that reproduces the New York Times visualization inspires this package.

- *Page 2, paragraph starting with "In the data structure": I suggest a little more pretext for the example of company name and where the company is registered. Or state the purpose of ID and grouping before giving a cohesive example.*

Response: Text has been modified to include an example using the US and Japan osiris data.

- *Page 3, paragraph starting with "One of the frequently": I was initially confused about how NA were handled, but it was clearer later in the paper that NA became its own category.*

Response: We have make it more clear in this paragraph.

- *Page 3, paragraph starting with "A principle important": switch to "An important principle"*

Response: Fixed

- *Page 3, paragraph starting with "There are multiple": avoid single sentence paragraphs. This could be included at the end of the previous paragraph to make a cleaner transition.*

Response: Fixed

- *Page 4, paragraphs 2 and 3: these paragraphs could be combined*

Response: Fixed and switching the order.

- *Page 5, paragraph 2: I don't think that redundancy needs to be capitalized here*

Response: Fixed

- *Page 5, last paragraph of section 5: "As illustrated in Figure 5." is left as its own sentence, but should be included with the surrounding sentence.*

Response: Fixed

- *Page 5, last sentence: change "Github" to "GitHub". Also, it would be helpful to include the GitHub link here as well.*

Response: Fixed

- *Page 6, Data pre-processing section: "ranking the sales" should be replaced with "ranking the observations".*

Response: Fixed

- *Page 6: the sentence with "When the observation", can be improved for clarity ("When an observation is moved up by a quantile, another one is bound to …")*

Response: Fixed

- *Page 6: Place tibble into table with a caption*

Response: Fixed

- *Page 7, first paragraph: replace "from the raw into" to "from the raw format into"*

Response: Fixed

- *Page 7, 3rd paragraph: include this information in another paragraph to help transitions.*

Response: Fixed

- *Page 7, 3rd and 4th paragraphs: the time_dependent argument only appears in the _plot() functions, so it should be mentioned in the later section.*

Response: Fixed

- *Page 8, table 2: "argumenst" should be "arguments"*

Response: Fixed

- *Page 8, paragraph 2/3: the proportional_shade() function can be discussed solely in one paragraph for better transition.*

Response: Fixed.

- *Page 8, paragraph 4: should "wallaby's plot" be "wallaby plot"?*

Response: Fixed.

- *Page 8, last paragraph: state that the data is included in the animbook package.*

Response: Fixed

- *Page 9: remove comma after "Figure 7"*

Response: Fixed

- *Page 10, applications: include sources at the first mention of each example.*

Response: ???

- *Page 13: use consistent naming for the plot objects (kan_p vs. p)*

Response: Fixed

- *Page 15, paragraph starting with "In Figure 11": There is a sentence prompt that should be filled in "(explain the difference in proportion …)"*

Response: We forgot to remove that.

- *Page 16, the two paragraphs before discussion: the second paragraph leads to explanation from the previous paragraph, so these ideas should stay together in one paragraph.*

Response: Fixed

- *Page 18: There is a reference starting with "———" that should be filled in.*

Response: Fixed

- *An example with the funnel_web_plot() would be a nice addition in the applications section.*

Response: ??? I think this is more like an experimental visualization.

Code

Usability:

- *Page 6: it states that users can specify breaks on a 0 to 1 scale. Is there a reason why users can't provide breaks on the raw data scale and have the function transforms the values to fit the 0 and 1 scale?*

Response: In the function, before separating the observation into groups, it will go through normalization (the data scale will be between 0 and 1). If we implement the raw data scale, it would require redesigning the function, as it would have to be able to distinguish between normalization and raw data scale. ??? is this too long?

- *Page 8: could the last_plot() function be used by default so that users don't have to save the plot into a ggplot2 object?*

Response: It should work by default.

- *In the wallaby plot, the applicable points move from all categories to a single category (or vice versa). Is there a way to choose which single category gets used?*

Response: Yes, we have made it clearer in the arguments description and on page 10.

Code:

- *When using custom breaks, the error displays "The breaks vector must have the same number of groups as ncat argument", but it is stated in the article that the length of breaks need to equal the number of categories plus 1.*

Response: The code has been fixed.

- *When running the code on an M1 MacBook Pro, some of the shaded sections tear apart and look like white circles. This happens particularly with the wallaby plots.*

Response: ??? It might be related to the default colour for shaded sections. Will need examples to replicate.

Other comments:

- *Are there plans to include additional visuals to help keep track of the points as they disappear? In the NYT article, the points are summed and proportions are calculated as the points reach the right side of the plot and I think this would be a nice addition in the future.*

Response: ??? Not at the moment since it can lead to a visual overload.

- *When there are many data points, it becomes difficult track the general movement of observations. Would there be any options to trace the paths to show where the points have moved (like a line with low alpha)?*

Response: The gganimate elements (e.g., shadow trail) work normally with this package. We have added an explanation on page 8.

- *Overall, I really like this package!*

Response: Thank you!

Reviewer 3

General

The paper presents a graphical tool for visualizing changes in some measurements over time. It also allows for multiple groups to appear in the Figure. At first glance, this sounds like a useful tool. Me personally, I was not amazed by this tool simply because of the lack of simplicity and straightforward interpretation. On the positive side, the paper is very professionally written, but it is not that easy to follow.

Response:

Detailed comments

1. Figure 1 for instance. What is the interpretation of this figure? I struggled to understand and did not manage to do so.

Response: ??? Already in the caption.

2. Figures 2 and 4. Since the data entered in R are numerical or categorical values, what is the use of these figures? Why should the reader be informed of the raw, handwritten data?

Response: Not change, this image is important base on discussion with user. (see how other use the package). (be more forceful)

3. I have spotted a few grammatical errors:

a) First paragraph after Figure 1. "This data…" > "These data…". The word data is plural, the singular form is datum.

Response: Fixed

b) Beginning of Section 5. Please delete the letters "R. E." from the reference of Mayer and Moreno (2002).

Response: All referencing are automated and the style is consistent with the R journal style so we have no control over this.

c) Page 13. First paragraph, 3rd and 4rth lines probably have a typo ("…and japan").

Response: "japan" is one of the variables in the osiris data. We have changed the text so that this is clearer.

4. The description you give for Figure 10 helps the reader understand the meaning of this plot. But, in the example in the R package with the aeles dataset, things are not so obvious. There are many observations that mask the visibility of the points.Also, what do the X and Y mean?

Response: XXX We have fixed the example on the package website and provide more documentation about the animation.

5. Why do the authors write "I acknowledge…" and not "We…" in the acknowledgements?

Response: Fixed

Code

I did not find the commands in the package very user-friendly. For instance, to produce a plot I must go through many functions, prior to typing the plot to appear. The example in the aeles dataset is such a case. I need to type 4 commands in order to see a plot, which I repeat, cannot understand, nor interpret easily. I would prefer a function that does all this preprocessing and plots the data. Legends are missing, and in my opinion, the help files are not very helpful for someone who wants to use this package easily, to visualize their data.

Response: This was a design choice for transparency so the user can see the steps. We have updated the web page to make the steps clearer, and also added a single function that does all the steps. XXX