Response to review: New and simplified manual controls for projection and slice tours, with application to exploring classification boundaries in high dimensions

Ursula Laa, Alex Aumann, Dianne Cook, German Valencia

2023-01-05

Editor

A referee, an Associate Editor (AE), and I have reviewed your paper. There's enthusiasm for your paper but a number of issues that came up in the reviews require additional attention. As such, I'd like to offer the opportunity to submit a revised version that addresses the issues brought up by the AE and referee.

Thank you for the careful review of our paper. We have done our best to make changes to address each comment, as detailed in blue below.

1- Your revision should not increase the length of the manuscript.

2- I will download and run your code as part of the next round of reviews. I suggest, but don't require, the tidyverse style for your code and, in any case, ask that you submit code that is documented and as user-friendly as possible. Before resubmitting please run your code in a "clean" environment as issues such as required packages that aren't loaded or calls to files that don't exist are common and can delay publication (I'm not saying these exist in your case, just asking you to be aware!).

3- Your paper will be printed in black and white. The online version will be in color but to accommodate printing without color please be sure you don't refer to color in either the figure captions or text and that all figures are still interpretable in black and white.

4- Please be sure all figures have font size of labels and any text in the figure of at least 10pt.

Articles published in JCGS are ordinarily accompanied by online publication of the data and computer code used to illustrate the proposed methods. The goal is to enable readers to replicate much of the analysis and simulations described in a paper and to apply proposed methods in their own data without having to write code from scratch. When it is not practical to include a whole dataset in the manuscript, the manuscript should state how the complete dataset can be obtained. Please let me know if you would not be able to publish your data for reasons of security or confidentiality. Please provide the computer code and/or data that you hope to publish along with your revision. I have attached the instructions for including online materials here.

Associate Editor

Summary

This paper describes a method for interacting with multidimensional data using slice tours in Mathematica (as well as a reduced-feature implementation in R).

Major Issues

- Reproducibility:
  - Code should be reproducible and not depend on user paths.

  - Data files should be uploaded along with code files (they don't seem to be part of supplemental material?)
  - Not having the ability to actually run the code (because the data wasn't included) makes it somewhat difficult to assess the software side of this paper.

- Grammar: There are a lot of minor grammatical issues with this paper. I suggest utilizing a tool such as Grammarly to spot and fix them. The ones that I noticed as I read through the first section are detailed in minor issues below, but I am trying to focus more on the content than the style, so I will try to not nit-pick through the rest of the paper and trust that the authors can use the online tools.

- Cite your software packages appropriately - they're tools you depend on to do your work and should be cited in the same way papers are cited.

  – Page 3, line 28: Capitalize Mathematica and cite it appropriately.

  – Page 3, line 32: Cite R

  – Page 7, line 28: Cite Shiny

- Notation

  – Define your notation better: V_m is used without being defined, and it would probably be helpful to have a clearer mathematical definition of the relationship between your basis A and V_m.

- What developments are specific to Mathematica and are not yet implemented in R? You allude several times to R's deficiencies in this space but don't ever explicitly spell out what is missing, beyond a nod to mouse tracking capabilities in Shiny (which would likely depend on the user's screen resolution, but should be fairly precise once you condition on that - why would Mathematica have better resolution for mouse controls?) Below is a set of comments relating to this issue, written as I read through the paper (so hopefully the chronology is preserved and you can see why this isn't as direct as it should be).

  - Page 7, line 46: "Mathematica provides much utility and versatility, such as inbuilt data visualization, data manipulation and analysis, dynamic functionality, and sym- bolic and numeric computation. Most of this inbuilt functionality is user-friendly and described in the Mathematica documentation; typically, numerous examples are provided within the documentation, and some possible issues are outlined there." The same could be said of R, so I'm not sure why this is relevant. It seems mostly like Mathematica propaganda.

    – Page 7-8: "Importantly for our work, it is relatively simple to create dynamic objects via the inbuilt commands Manipulate or DynamicModule. Control objects, such as sliders, locator panes, and input fields, can then be used on dynamic variables, and when there is a change in the dynamic variable the dynamic objects which contain that variable will be updated. This is the essential ingredient for this implementation of the manual tour." Again, the same dynamic control creation exists in Shiny, along with the basic reactivity to user inputs. It's hard to see why Mathematica is different here.

    – Page 8, line 11: "The most relevant inbuilt Mathematica function for our purpose is the LocatorPane. This creates a region on the screen where the position of the mouse is captured and then converted to input that updates the graphics functions, enabling the manual navigation of the tour."

  - Finally, after all of this time, this is the one input that might be a bit harder to implement in R/Shiny, though I think via WebGL it should be workable. This is where the focus should be, rather than the previous two paragraphs, which are a distraction.

Minor issues and suggestions

- Pg 2, line 8: "the more useful methods for data analysis need to work for arbitrary dimension." Perhaps instead: methods which work for arbitrary dimensions are more useful/have more general utility? Either way, dimension should be plural.

- Pg 2, line 17: "could theoretically show the viewer the data from all sides." It might be helpful to have a small figure here to demonstrate what this might mean, for those who aren't familiar with tours and/or who can't visualize n-dimensional space projections into 3D out of thin air. Alternately, link to e.g. a YouTube video showing what you mean by "from all sides"
- Pg 2, line 23: "A video illustrating the capabilities is available through video library of ASA Statistical Graphics Section (2022)." Add "the" before video

Reviewer 1

In this paper a new manual controls for projections and slice tours are presented. The controls are designed to explore high-dimensional data based on linear projections and slices. In the paper a Mathematica package is described to explain the application.

Some comments:

In the Abstract you have mentioned "Some limited implementations has also been made available in the R package tourr" but in Sub-Section 3.2 you can mention in more detail the limitations in tourr implementation

Section 1 line 29 "The software section describes mathematica package" mathematica should be capitalized as you did through the paper in most cases. Also since is the first time you have mentioned Mathematica you should include the reference here (Wolfram reference appears in page 4 line 26 for first time)

Section 2: "manual tour allows the user to alter the coefficients of one (or more) variables" but in the explanation it is not clear what is V_m and how to alter the coefficients for more than one variable.

In page 4 line 2.2 "The components corresponding to V_m are directly controlled by cursor movement which updates row m of A". Maybe a video will be illustrative in this sub-section.

Section 3: will be useful to include a schematic diagram illustrating the interactivity

In the manual tour implementation. This part is difficult to follow, it is more clear after reading the application Section because you have included some videos but I think you need to include something here to help the reader.

In the introduction "The software section describes a mathematica package that is used for the application, and describes the interactive environment that would be desirable within R as new technology becomes available" I think you need to be more specific about the R limitations. Iin Section 3. Page 7, line 20 "One might implement such an interactive interface via R Shiny, but in particular the tracking of the mouse position in small increments might pose a challenge" This is the only possible R limitation you have mentioned.

Section 4: Page 10 line 36: Figure 3 summarizes (space)

Section 5: page 15 line 26 "If new technology for interactive graphics became available" be more specific about the needed technology in R to implement the manual controls as you did with Mathematica