Answer of the authors to the review of
RJournal 2021-173

May 13, 2022

We would like to warmly thank the reviewers for their careful reading and their insightful comments that helped us to improve the paper. We have corrected and clarified the manuscript. Our specific answers to the reviewers' comments are below.

1 Reviewer # 1

1. The following suggestions can be considered to make the content more fluid:

    - *There are three section with the title is "Bibliography". First one may be differed as highlighting that it covers the related work about the missing data in literature. The second and third ones are duplicated.*
      Answer: Thank you for pointing out these duplicates. We have followed your recommendation and changed the title of the first "Bibliography" subsection in section *Structure and Content of the Platform* to "References on missing values" and removed duplicate at the end (before the appendix).

    - *At the last paragraph of the "Context and Motivation" part, there are some sentences highlight the remainders, but the numbers used for identifying the section numbers are not used in the section title. This is not enough for the reader to find the relevant chapters. It needs to be pointed out more descriptively.*
      Thank you for pointing out that we haven't adapted the overview to the R Journal article format without section number. We have now reformulated this paragraph at the end of this section to correct this:

        In the section entitled "Structure and content of the platform" we describe the different components of the platform, the structure that has been chosen, and the targeted audience. The section is organized as the platform itself, starting by describing materials for less advanced users then materials for researchers and finally resources for practical implementation. We then detail the implementation and use-cases

1

of the provided R and Python workflows in the following section entitled "Details of the missing values workflows". Finally, in the conclusion, we outline an overview of planed future developments for the platform and interesting areas in missing values research that we would like to bring to a wider audience.

- *Section 3 may be given under Section 2.1 to make more compact the structure of the manuscript.*
  Answer: We understand the argument for this suggestion but since we think that the workflows form a much larger contribution on this platform compared to the other sections, we would prefer to keep the current structure. We have however changed the title of Section 3 ("Details of missing values workflows" to highlight the difference with the content of Section 2.1).

- *A comparison table may be given to understand the difference of the scope between R and Python packages on workflows to make easier to understand for readers.*
  Answer: Thank you for your remark. Indeed, such a table can remove ambiguity between R and Python packages. We have added this table in the appendix and we reference it in the core text (page 10).

- *The graphs which are created by using ggplot may be improved in a data visualization expert's perspective. For example, adjusting the axis and choosing a more suitable chart type for Fig 6. Also the variable "imputation error" used in Fig 6 is not clear, so the more descriptive information about the variables used in the graphs can be given.*
  Answer: Thank you for your comment. The graph in Figure 5 has been adjusted and Figure 6 has been replaced by the (numerical) output of the R function, which is more readable.

- *As it stands, it is very difficult to separate the appendix from the rest of the manuscript, thus the appendix title must be added and the parts in there must be seperated and given under the subsections.*
  Answer: Thank you for pointing out this unclear separation that we oversaw when converting our manuscript into the R Journal format. We have now added a section title "Appendix" to clearly mark the beginning of the appendix.

2. The second point about the manuscript is that it has many typos, grammar mistakes and citation mistakes. For these, I suggest the following items:

   - There are many grammatical errors such as "... or researchers in a different fields using ..." need to be checked in the manuscript. I strongly recommend a grammar check by a native English speaker.

- There are also many typos need to be corrected such as "hanlde" in Lectures part.
- To the reference is "Tierney et al.", year information should be added.
- Extra parentheses were used in some references such as Breiman, 2001 in page 3.
- Some of the references are not seen in the text in page 5 and 8.
- Acknowledgement must be given as a section.
- Figure 8 and 9 must be referenced in the text.
- The double references are given in the superscripts must be seperated by commas.

Answer: Thank you for your detailed and careful reading of our manuscript and for taking the time to point out these various mistakes. We have taken into account all of your suggestions and corrected all mistakes noted above. Additionally, we have also asked a native English speaker to proof-read our manuscript to remove grammatical errors and incorrect expressions.

3. The third point is about the examples and R codes in the manuscript. The example codes should be given. The proposed R functions such as score_pred, plot_score_realdatasets and etc. must be introduced and the arguments must be given in Section "How to predict in the presence of missing values?". The R codes used in Sec 3 and Appendix should be shared to reproduce the results.

Answer: Thank you for your suggestion. In the main text, we have detailed the code examples. We have introduced more clearly the functions `produce_NA` (page 9), `how_to_impute` (page 10), `how_to_impute_real` (page 11), `score_pred` (page 11) and `plot_score_realdatasets` (page 19) and given their arguments. Additional details on the code are given in `https://github.com/R-miss-tastic/website`.

2 Reviewer # 2

1. *Some references are missing (only ? are presented), and some typos are in the text.*
   Answer: Thank you for your detailed and careful reading of our manuscript. We have corrected the citing mistakes and typos. We have also asked a native English speaker to proof-read our manuscript to remove grammatical errors and incorrect expressions.

2. *However, I think, for newcomers to the problem, the examples might be too brief.*
   Answer: Thank you for your remark. In the paper, we have added more details, examples code and better introduced the proposed functions. Different workflows, very detailed for a newcomer, are available on the

website: *how to impute missing values in R*, *how to impute missing values in Python*, *how to generate missing values in R*, *how to generate missing values in Python*, *how to estimate some parameters with missing values in R* and *how to predict with missing values in Python*. However, putting all the details of the workflows in the article would be too long. Besides, the notebooks presented on our platform Rmisstastic are frequently reworked according to the feedback that we get from users.

3. *The general quality of the paper is good, but after reading, I do have the impression of "and what". Everything in the paper is in the portal. Thus it is hard to see the point of repeating yourself and describing the portal, which is clear to use and read. Since it uses blogdown it is at it is :) The whole paper is the portal presentation, and that is the point - I think for promoting the topic, I would recommend doing the presentation at UseR! (or similar) conference than to write "presentation" paper. I would rather see the companion paper, which will provide more details on methods discussed in portal examples and workflows with more theoretical aspects. So that the reader will get more into the topic by providing him with additional information. To conclude, the idea is excellent, the topic is needed, and the paper is nicely written, but I am not sure that this is the proper way to promote the portal. I would suggest doing the conference presentations and workshops with the structure depicted in the paper and reworking the paper as companion material to the portal's content.*

   Answer: We can understand where your concerns are coming from. We also plan to present this platform in a conference. However, we remain convinced that this type of presentation in the form of an article is the best way to showcase and value the work that has been put into this platform and the results of this project that aims to offer a perennial hub and entry point for missing data problems. For example, describing the structure of the workflows reflects how this contribution was created and can help in the development of new workflows. In addition, we have already received feedback from several sources (after conference presentations, or through our github repository) that this platform has been well used for research and teaching purposes. Finally, this type of presentation in the form of an article also makes it easy to cite the work (currently, people quote the website, whereas quoting an article is simpler and more conventional). In other research fields such biology, it is common to describe the development and provision of a resource website in an article, such as `https://www.nature.com/articles/ng.2653`, `https://dx.doi.org/10.1093%2Fnar%2Fgkx1081`. Therefore, in order to enable a uniform way of referencing and promoting this project and sub-projects such as the workflows (instead of citing the platform through its current url), we hope that our manuscript will be accepted for publication in the R Journal.