

Response to AE and Reviewers

ggESDA: An R Package for Exploratory Symbolic Data Analysis using ggplot2

Bo-Syue Jiang^a and Han-Ming Wu^b

^a*Department of Statistics, National Taipei University*

^b*Department of Statistics, National Chengchi University*

April 15, 2023

Dear Editor of the R Journal,

Thank you for providing us with the opportunity to revise our manuscript. We greatly appreciate the comments from the Editor, AE, and both reviewers. In response to these comments, we have made revisions to the article, package, and codes. Below, you will find our point-by-point response to the reviewers' comments. The reviewers' suggestions and criticisms have significantly improved the manuscript.

Once again, we would like to express our gratitude for considering our paper for publication in the R Journal. We welcome any further constructive comments. We look forward to hearing from you soon. Sincerely Yours,

Wu, Han-Ming (Hank), Associate Professor

Department of Statistics, National Chengchi University

No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan

Tel.: +886-2-29393091

E-mail: wuhm@g.nccu.edu.tw

Website: <http://www.hmwu.idv.tw>

Manuscript ID: RJournal 2022-156

Title: ggESDA: An R Package for Exploratory Symbolic Data Analysis using ggplot2

Responses to Editor's Comments

- Please consider restructuring your paper into the new journal template described at <https://journal.r-project.org/submissions.html>. This will produce both pdf and html versions, and makes your work more accessible to visually impaired readers.

Authors: Thank you for your valuable feedback. Our paper has been restructured into the new journal template using `rjtools` as described on the website <https://journal.r-project.org/submissions.html>. We agree that having both pdf and html versions of our work will improve accessibility for visually impaired readers, which is an important aspect of making research more inclusive.

- Please ensure that the text of all of your figures is readable. Currently the plots are very small in the pdf, and the text is microscopic.

Authors:

- Captions need to follow the structure described in Figures section of <https://rjournal.github.io/rjtools/articles/format-details.html>: "Note that the structure of a thorough caption is three components: (1) what is the plot/table about, (2) specific details of plot/table, like what type of display and how variables are mapped, (3) the most important thing that the reader should learn."

Authors:

Manuscript ID: RJournal 2022-156

Title: ggESDA: An R Package for Exploratory Symbolic Data Analysis using ggplot2

Responses to AE's Comments

- The article is about 'for Exploratory Symbolic Data Analysis using ggplot2'. Both reviewers agreed that this work needed major revision. Both reviewers proposed a very large number of comments on the structure of the article (the first reviewer said it was confusing, the second said it was good, but I rather agree with the former, it needs some work), on the specific wording of the paper as well as on the software. There are a lot of changes, but I don't see any blockers, the authors should be able to make them all.

Authors: The tidyverse style guide: <https://style.tidyverse.org/>

Manuscript ID: RJournal 2022-156

Title: ggESDA: An R Package for Exploratory Symbolic Data Analysis using ggplot2

Responses to Reviewer 1's Comments

- The manuscript describes a new R package, ggESDA that extends ggplot2 to visualization of symbolic data, represented as intervals. The structure of the manuscript is very good, authors explain various design decisions and naming conventions. Authors give examples of existing R packages that include similar functionality and give a satisfactory explanation of differences between the other packages and ggESDA.

Authors:

- In general, the manuscript gives a good introduction to the topic of symbolic data analysis and visualization, as well as a great overview of the many functions implemented in the new package. However, both the manuscript, the code attached to the manuscript, and the code in the ggESDA source needs to be heavily revised. Below, I provide detailed comments.

Authors:

— ARTICLE: —

- very good structure of the article!
- when including the R code, try to include it so that it can be copied and pasted - try using the **reprex** package, then you wouldn't have to include the code as a separate R-script

Authors: Jiang/Wu

- English could be improved

Authors:

- some mistakes:

- on page 1, 7th line from the bottom - wrong name of the package and wrong link!

Authors:

- page 6, "Aggregation by categorical variable" - you mention that RSDA package has the function `classic.to.sym()`, while you've implemented a similar one, `classic2sym()` - why did you do that? Can't the user use the one from RSDA? What is the improvement of your implementation over the one from RSDA?

Authors: Jiang

- page 7, "Descriptive statistics for interval-valued data" - this paragraph needs to be revised, the sentences are not logically connected with one another, and there need to be some more references

Authors:

- page 13, 3rd line from the bottom - both dates are the same

Authors:

- page 14, "The min max plot" - a bit misleading description, because the blue points on the plot have $x=y$ (minimum, as I understand), while the red points have $x=\min$ and $y=\max$

Authors:

- page 20, last sentence before the code snippet for Figure 13(a) - there is a mistake in examples, both AH and POLITICS show the same type of distribution, while from the text it seems like the authors wanted to show various types of skewed distributions

Authors:

- code in the text is not nicely styled, much nicer code is in the vignette!

Authors: The tidyverse style guide: <https://style.tidyverse.org/>

- when running many functions from ggESDA, I get many warnings indicating that the implementation uses deprecated versions of various functions or arguments and mixing of `tidyverse` and base R - since you're extending `{ggplot2}`, you should stick to tidyverse style.

Authors: Jiang

— R SCRIPT: —

- don't use `"rm(list=ls())"`! (you can read more about this e.g., here: <http://www2.stat.duke.edu/~rsc46>)

Authors:

- naming of variables with `'.'` is not recommended

Authors: Jiang/Wu, The tidyverse style guide: <https://style.tidyverse.org/>

- give output of `sessionInfo()` at the end - this will show all the versions of the packages you were using, your OS, and R version

Authors:

- I get warning when using `classic2sym()`:

```
"Warning message:
```

```
'as.tibble()' was deprecated in tibble 2.0.0.
```

```
Please use 'as_tibble()' instead.
```

```
The signature and semantics have changed, see '?as_tibble'."
```

Authors: Jiang

- general: in the .R script, use "----" after the name of the section to create a nice 'table of contents', instead of "#####"

Authors: Jiang/Wu

- it's not recommended to load libraries in the middle of the code

Authors: Jiang/Wu

- don't use global variables! (e.g., code to Fig.6 and 7, and other) - you should instead create a variable inside the data table that you are visualizing

Authors: Jiang/Wu

- Fig.6: - title is not very informative; this is a paper on visualization, so the examples should also be well constructed

Authors:

- Fig.6: - code shows the use of `classic2sym()` function, but the argument 'groupby = "customize"' is not explained well neither in the manuscript, nor in the R help

Authors: Jiang/Wu

- Fig.7: - in the R-script, there is superassignment used, apparently to be able to use the function argument in the 'aes()' mapping - this is not a correct way; the correct code would be something like the following:

```
ggInterval_minmax(facedata, aes(!!sym(x), size = 2))
```

please, check the {rlang} documentation (e.g., here: <https://rlang.r-lib.org/reference/injection-operator.html#injecting-expressions>)

Authors: Jiang

- Fig.9 - warning:

```
"> ggInterval_scatter(facedata, aes(x = BC, y = AD, fill = Subjects),
+                       col = "black") +
+   scale_fill_brewer(palette = "Set1") +
+   labs(fill = "Subjects")
```

Warning messages:

- 1: Ignoring unknown parameters: x, y
- 2: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> = "none")' instead.
- 3: Use of 'd\$x1' is discouraged. Use 'x1' instead.
- 4: Use of 'd\$x2' is discouraged. Use 'x2' instead.
- 5: Use of 'd\$y1' is discouraged. Use 'y1' instead.
- 6: Use of 'd\$y2' is discouraged. Use 'y2' instead."

Authors: Jiang

- Fig.10(b) - what does the 'col = "white"' argument inside the 'aes()' mapping mean? Are there specific 'aes' arguments that ggESDA implements?

Authors: Jiang

- Fig.12(a) - lots of warnings and the figure looked differently than the one in the manuscript:

```
"> # Figure 12(a)
> library(ggthemes)
> p <- ggInterval_radar(Environment, plotPartial = c(4, 6),
+                       showLegend = F, base_circle = F, base_lty = 1,
+                       addText = F, addText_modal = F) +
+   scale_fill_manual(values = c("darkred", "darkblue")) +
+   scale_color_manual(values = c("darkred", "darkblue")) +
+   labs(title = "") +
+   theme_hc()
```

Scale for 'fill' is already present. Adding another scale for 'fill', which will replace t

Scale for 'colour' is already present. Adding another scale for 'colour', which will repla

Warning messages:


```

1: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> = "none")' instead.
2: Using alpha for a discrete variable is not advised.
> gridExtra::marrangeGrob(list(p), nrow=1, ncol=1)
There were 20 warnings (use warnings() to see them)
> warnings()
Warning messages:
1: Use of 'd$x' is discouraged. Use 'x' instead.
2: Use of 'd$y' is discouraged. Use 'y' instead.
3: Use of 'myPathData$x1' is discouraged. Use 'x1' instead.
4: Use of 'myPathData$y1' is discouraged. Use 'y1' instead.
5: Use of 'myPathData$x2' is discouraged. Use 'x2' instead.
6: Use of 'myPathData$y2' is discouraged. Use 'y2' instead.
7: Use of 'polyDf$cos' is discouraged. Use 'cos' instead.
8: Use of 'polyDf$sin' is discouraged. Use 'sin' instead.
9: Use of 'polyDf$group' is discouraged. Use 'group' instead.
10: Use of 'polyDf$obsGroup' is discouraged. Use 'obsGroup' instead.
11: Use of 'polyDf$obsGroup' is discouraged. Use 'obsGroup' instead.
12: Use of 'plotMin$cos' is discouraged. Use 'cos' instead.
13: Use of 'plotMin$sin' is discouraged. Use 'sin' instead.
14: Use of 'plotMax$cos' is discouraged. Use 'cos' instead.
15: Use of 'plotMax$sin' is discouraged. Use 'sin' instead.
16: Use of 'd$x' is discouraged. Use 'x' instead.
17: Use of 'd$y' is discouraged. Use 'y' instead.
18: Use of 'plotMax$Variables' is discouraged. Use 'Variables' instead.
19: Use of 'newD$x' is discouraged. Use 'x' instead.
20: Use of 'newD$y' is discouraged. Use 'y' instead."

```

(it seems like there is one plotting device not closed because when I ran 'dev.off()', I got output 'null device 1')

Authors: Jiang

- similar warnings for Figure 12(b)

Authors: Jiang

- the same warnings and problem with device graphics for Fig.13(a) and 13(b); additionally, I get a figure that's showing darkest inside and lightest outside, while in the manuscript, it's reversed

Authors: Jiang/Wu

- general comment: avoid abbreviating boolean constants (i.e., write out 'TRUE' and 'FALSE' instead of 'T' or 'F')

Authors: Jiang/Wu

— PACKAGE: —

- there are no tests in the source code - you should create those, you can find `{testthat}` useful for this purpose

Authors: Jiang, *testthat:* Unit Testing for R: <https://cran.r-project.org/web/packages/testthat/index.html>

- vignette is fine, but may be too long - you might wanna split it into smaller, more focused chunks

Authors: Jiang

- naming of the scripts in the R directory is both ".R" and ".r" - decide on one

Authors: Jiang/Wu

- loading the package also loads `{ggplot2}` and `{tidyverse}`
 - `ggplot2` is part of the `tidyverse`, so you don't have to load it separately
 - dependence on `{tidyverse}` is highly not recommended since `{tidyverse}` is a huge bundle of packages - you should check which functions do you need and import only those
 - you might find this info useful: <https://devguide.ropensci.org/building.html#pkgdependencies> (the entire book is very useful as well!)

Authors: Jiang

- in the code, the naming of the variables is not consistent: once it's with '.', once camelCase; and the styling should be improved: use spaces and new lines to make the code more readable (`{styler}` package might help here: <https://styler.r-lib.org/>)

Authors: Jiang/Wu, camelCase: pineApple, Snake Case: pine_apple_pineapple

- the exported functions have a nice documentation, but all the other ones are not documented at all; comments in the code are scarce

Authors: Jiang

- you might wanna consider submitting the package to rOpenSci Statistic Software Review: <https://ropensci.org/stat-software-review/>

Authors: Jiang/Wu

Manuscript ID: RJournal 2022-156

Title: ggESDA: An R Package for Exploratory Symbolic Data Analysis using ggplot2

Responses to Reviewer 2's Comments

Installing the package

Firstly, I have been unable to install the package, because I get the following error when loading the dependency “RSDA”:

```
Error: package or namespace load failed for ‘RSDA’ in dyn.load(file, DLLpath =  
DLLpath, ...):  
unable to load shared object ‘/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources
```

```
dlopen(/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/library/Matrix/libs/M  
/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libquadmath.0.dylib
```

```
Referenced from: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/library/Matr
```

```
Reason: tried: ‘/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libquadma  
(no such file)
```

I am running on an M1 Mac. Could the authors please advise on how to solve this?

Authors: Jiang/Wu

Review

- While there is value in the functionalities provided by authors, the paper is poorly written and needs substantial improvements before being considered for publication. Language usage throughout the manuscript needs to be greatly improved, since there are numerous ungrammatical sentences which make ideas difficult to understand. The structure of many sections is chaotic, leading to confusion when trying to understand the functionalities.

Authors:

- The arguments of the plotting functions are not properly described anywhere. Sometimes, possible values are given for some of the arguments but the outcome of providing all the possible different values is not explained (e.g., for arguments taking logical values, the outcome is explained sometimes for FALSE or TRUE value, but not for both). Default behaviors are not clearly stated either.

Authors:

- Below are some instances of the issues found, but these are not comprehensive. The authors need to identify all instances of problems similar to those listed below them and correct them:

Authors:

- Inconsistent verb tense usage (e.g., page 1, “These packages implement statistical and machine learning methods, primarily for interval-valued data, such as dimension reduction, clustering, and regression, among others. Furthermore, some of these packages provided basic and limited graphical tools such as scatterplots “).

Authors:

- Another example of inconsistent verb tense usage, in abstract: “Nowadays the collected data keeps getting bigger and complex. The description of data was no longer stored by a form of a single value but the intervals, histograms and/or distributions.”. First verb is “keeps” (present tense), second is “was” (past tense). In general, this two sentences in the abstract sound ungrammatical/confusing.

Authors:

- Page 2: “The first part is devoted to data manipulation, such as read and write from/to files, and aggregation.” It should be “... such as reading and writing

...”. Additionally, as mentioned further below, authors do not really provide any functionality in their package for reading/writing files.

Authors: Jiang/Wu

- Page 2: “the `ggplot2`”, change to just “`ggplot2`” or “the `ggplot2` package”. The usage of “the” needs to be fixed throughout the manuscript.

Authors:

- Page 4: “For instance, the packages of `ggtern` (Hamilton and Ferry, 2018), `ggnatogram` (Maag, 2018), `ggstatsplot` (Patil, 2021), and etc.”. The expression “the packages of XXXX”, sounds strange. Replace with “the XXX packages” or “the packages XXX, XXX and XXX”. Also, avoid expressions such as “and etc.”

Authors:

- Paragraph describing face recognition dataset “The six measurements designed to identify each face are expressed as the number of pixels in an image. Specifically, they are the distance spanned by the eyes (denoted by $X1 = AD$), the distance between the eyes ($X2 = BC$), the distance from the outer right eye to the upper middle lip between the nose and mouth ($X3 = AH$), the corresponding distance for the left eye ($X4 = DH$), the distance from the upper middle lip to the outside of the mouth on the right side ($X5 = EH$), as well as the distance to the left side of the mouth ($X6 = GH$). These measurements were taken from a sequence of more than 1,000 images and cover a wide range of values which make them interval variables. “It is confusing to describe the dataset by first saying that the six measurements are expressed as the number of pixels in an image, and then say what each measurement is. State first the different measurements, and then state that the unit for all of them is pixels.

Authors:

- Page 5: “When applied to an interval variable, summary function returns the summary statistics...”. It should be “the summary function ...”.

Authors:

- Missing short description of the 2nd included dataset (Environmental questionnaire). Similarly as what is done for the face recognition dataset, do not just point to an external reference, but include a brief summary.

Authors:

- Description of `classic2sym` function is confusing and needs major improvements.

Authors: Jiang/Wu

- Page 6: “One can also possible aggregate observations based on user-defined categories.” Ungrammatical sentence. Also, provide details about the usage of user-defined categories.

Authors:

- Authors state: “Alternatively, users can create interval-valued data tables using the function `classic.to.sym` in the RSDA package.”. How does it differ from the `classic2sym` function introduced by the authors?

Authors: Jiang/Wu `classic.to.sym`, `classic.to_sym?`, `classic2sym?`

- Authors say package provides functions for reading/writing from files, but these are provided from RSDA package. Therefore, remove the statement that the presented package provides such functionalities, or implement them.

Authors:

- Page 7: “For example, the five-number summary of numerical data (minimum, 25% quartiles, median, 70% quartiles, and maximum).” The upper quartile is the 75% percentile. It does not make sense to say “25% quartiles”. Amend this to refer to the 25th and 75th percentiles as lower and upper quartiles. Better say lower and upper quartiles.

Authors:

- Page 7: "...should be provided before (some) graphing techniques..." . Remove "(some)".

Authors:

- Page 7: "The SDA literature contains some algorithms for calculating descriptive statistics for interval-valued data. Denote the i th observation of a univariate interval-valued variable X by $[a_i, b_i], i = 1, \dots, n$ and the i th observation of a bivariate interval-valued variables (X_1, X_2) by $([a_{i1}, b_{i1}], [a_{i2}, b_{i2}]), i = 1, \dots, n$. We have implemented basic descriptive statistics for univariate and bivariate interval-valued variables according to Irpino and Verde (2015). Formulas can be found in Table 3. The following code illustrates how to obtain these descriptive statistics. It must go as superindex." This paragraph is confusing, both due to the language and structure of ideas. Firstly, a notation is used, then a citation is presented, and then the previously presented notation is not used, and no more details of the chosen methodology are given. This needs complete rewording.

Authors:

- The authors state they have implemented functionalities to calculate descriptive statistics. However, judging from the example by the end of page 7, 2 of these are simply applications of the standard "mean" and "sd" functions. For the other 2, i.e., cov and cor, these are also implemented in the RSDA package. How do they differ, if there is any difference? In general, it seems the authors have not provided much, if any, novel functionality for calculation of descriptive statistics. Either more work would be required in this aspect of the package, or the wording in the paper should be changed to reflect this. A possible idea might be a function that directly calculates and presents in a proper way all of the mentioned descriptive statistics.

Authors: Jiang/Wu

- Section Standardization of interval-valued data : the implemented functions to perform this operation are not described. I assume it is the scale function, but this needs proper explanation.

Authors: Jiang/Wu

- Page 9: “With it, you can quickly scan large amounts of data in order to seek out anomalies and unusual observations.” Avoid using sentences such as “you XXX” for consistency and style.

Authors:

- Page 9: wrong usage of “The” in “The figure 2(a)”, change to Figure 2(a).

Authors:

- With titles of sections of plots, avoid naming them as “The XXX”, instead refer to them directly as XXX.

Authors:

- Index plot: “The Figure 2(c) shows a variant of index plot with an image strip that displays the intervals’ range through sequential colors.” It should be noted the color does not add additional information, since it just emphasizes the X range. Change this description to be more clear.

Authors:

- Description of figure 2(d) also very unclear.

Authors:

- Page 9: “As we can see, the set of three faces within each subject is cohesive within that subject. Obviously, subjects LOT, KHA, INC, and FRA have lower AD values than average.” A few issues with this:
 - There is no LOT, KHA, INC and FRA subjects. These are categories, rephrase this. In general, the authors throughout the paper seem to refer to the different categories (nationalities) as individuals. This seems to be a wrong interpretation of the dataset. Please correct throughout:
 - Avoid wording such as “Obviously, XXXX”.

- Avoid usage of “As we can see”

Authors:

- Figure 2, titles of panels (e.g., Index plot-Column Condition), what are these? These titles are confusing. Change or remove.

Authors:

- Page 9: “we observed that two people, HUS and KHA, have different quantities of AH and DH measurements.” As mentioned before, these are 2 nationalities , i.e., 2 groups of people, not individual people. Also, change this sentence to be more descriptive, not just stating “X and Y differ in Z and W”. E.g., how to they differ? Both higher in one group? Both lower?

Authors:

- Page 10: “Note that with the argument `plotAll = T`”. what happens if `plotAll = F`? Describe.

Authors: Jiang/Wu

- Boxplot section: what Is explained here, and presented in the corresponding figure (Figure 4) differs from my standard understanding of a boxplot (a box representing the 5-number summary). Change section to explain this.

Authors:

- Page 12: ”For the above steps, the subinterval I_g may not be necessary to have the same width. All boundaries of the intervals can be used to compute the breaks ξ_g . To be more precise, we pool a_i and b_i into a new vector and sort it as $(x^{(1)}, x^{(2)}, \dots, x^{(2n)})$ where $x^{(j)} \leq x^{(j+1)}$, $j = 1, 2, \dots, 2n - 1$. The frequency for the subinterval $I'_g = [x^{(j)}, x^{(j+1)})$ can be obtained from the Equation (1). This histogram is known as the non-equidistant-bin histogram.” Multiple sentences of this paragraph are ungrammatical and make the paragraph very confusing. Fix.

Authors: Jiang/Wu

- In general, the histogram section is confusing. Maybe change focus to describe differences (if any) with standard histograms of non-interval variables.

Authors:

- Page 12: "The line plot In a line plot, data is displayed along an ordered sequence or number line (usually on x-axis), showing how the data changes over time and if the data points are random or exhibit any patterns." Line plot refers in general to a plot of Y vs X where the different values are connected by lines, instead of just dots as in a scatterplot. Not necessarily vs time, although frequent. Change this part to better explain this type of plot.

Authors:

- Page 13: "The interval time series plot shown in Figure 6 is accomplished by the functions `ggInterval_index` and `geom_line`, with an economist theme." Not necessary to explicitly state the ggplot theme in text.

Authors:

- Minmax plots:
 - Explicitly say that each pair of dots connected by line corresponds to an interval.
 - Explicitly state that the reference line would correspond to intervals where `min = max`.

Authors:

- Page 14: "As an alternative, the literature commonly uses center and range to represent intervals". Avoid sentences such as "the literature communly uses". Directly say "Another commonly used method to XXX is YYY ..." or similar.

Authors:

- Page 15: “The bivariate plot provides the means to determine how variables relate to each other when measured on the same sample of subjects.” Avoid referring to the bivariate plot as if it was a single type of plot, when as described next, it is a full ensemble of plots.

Authors:

- Page 15: “Two main groups of observations can be distinguished. The AD and BC measurements of one group are larger while those of the other group are smaller.” Avoid such long wording of simple statements, e.g., say “Two main groups of faces can be observed, differing on their AD and BC values”.

Authors:

- Page 17, “except in the special case of classical data”, is classical data “non-interval data”? If so, say explicitly - Multivariate plots section:
 - I assume these refer to cases of more than 2 variables. If so, say explicitly (as opposed to bivariate)
 - Avoid expressions such as “the multivariate data sets”. Change to “multivariate data sets”, since otherwise, it gives the false impression that you are referring only to a specific ensemble of multivariate data sets.

Authors:

- Page 18: “The scatterplot matrix is commonly used to display multivariate data on two dimensions. It provides the bivariate relationships between combinations of variables using a grid (or matrix) of scatterplots.” Avoid this language, plots don’t provide things, but a means to visualize things.

Authors:

- Page 20: “As a result, they can be used to identify which variables are scoring high, low, or similar, or whether any variables consist of outliers.” Scoring in what scale? Confusing sentence.

Authors:

- Page 20: “To represent modal multi-valued variables, the stacked bar is used”: change to “...stacked bars are used...”
- Image plot section: a “matrix display condition” is mentioned, how is this even turned on? Confusing section - Multiple instances of the issues described above are found in the rest of the sections, as well as before. Authors should identify them and fix them.

Authors:

As a general comment

- As a general comment, the structure of the paper and that of the different subsections is confusing. For example, it might be clearer to first introduced the different datasets that will be used during the examples, rather than introducing them ad-hoc as required. For the subsections for each plot, the authors need to state more clearly and in a much more structured way which function produces them, the main arguments and the effect of different values, how the plot differs from the equivalent for non-interval data and the purpose of each plot.

Authors:

- A comment about the code of the package itself: many functions use repetitive code patterns, such as:

```
if(any(unlist(lapply(as.data.frame(data[,1:p]),FUN=identical,x=eval(this.x)))){ attr<-whi
}else if(any(unlist(lapply(as.data.frame(data[,1:p]),FUN=identical,x=eval(this.y)))){ att
}else
```

Avoid such repetitions of code by placing this code into an auxiliary function and calling it when required.

Authors: Jiang