

Five Rs

2023-03-20

	Replicable	Reproducible	Repeatable	Rerunable	Reusable
Pseudo Code	4	4	4	4	4
Python	2	2	2	3	2
C++	3	3	3	2	3
R package	1	1	1	1	1

Replicability

The replicability is the best for R package and worst for pseudo code since many things are open for interpretation. For example, the algorithm omits the pathological case when 3 or more points are collinear on the hull. In this case, the algorithm calculates the distance for each of the collinear points and chooses the closest one. However, opposite approach of choosing the furthest point would also satisfy the definition (there is more than one correct approach). The Python is ranked higher than C++ due to the accessibility of the code since Python is less wordy language and provides nice plots.

Repeatability

Since there is no randomness involved in the algorithm, the results are the same each time the algorithm is run. The algorithm works on distinct set of points and orders them, the output is identical each time the algorithm is run on a given hull of points (even if they are given in different order). However, the `<` operator applied in C++ and analogous dataframe arrangement are not described in pseudo code. R package uses R 4.1.2 and the other two languages use Python 2 and C++17. As of March 2023, minimum supported version of R for tidyverse is 3.5.

Reproducibility

The behaviour of the R package installed from GitHub is same on different machines with R version 4.1.2. Same holds for Python and C++ with their respective environments, though these lack detail of imported package versions. All 4 in the table above are publicly available, which is a condition of reproducibility.

Re-runnability

In R package, versions of imported and dependent packages are documented in `DESCRIPTION`. Wickham (2015) points out the ‘double-edged sword’ nature of dependencies, which means that if the major `tidyverse` dependency changes and breaks the code, maintenance work is required. However, this downside outweighs the benefits of using the library such as piping syntax to improve code readability (`margittr` package) and `ggplot2` package for plotting. Python and C++ code does not document the versions of the packages used, but Python is ranked higher since Python is more understandable and generally used language.

Reusability

The R package documents the code with comments and docstrings. The syntax is easy to read using pipes and documentation files can be navigated via `README.md`. Furthermore, the Python and C++ notebooks show examples of the algorithm and are also included in the package. Python is ranked slightly above C++ since the language is generally easier to write and use though C++ specifies the variable types beforehand and generally has more rigorous approach.

References

Wickham, H. (2015), *R packages: organize, test, document and share your code*, O'Reilly.