

## CodeRbook: R package for Interactive Exploratory data analysis and for data checks to build metadata

ByungJu Kim<sup>1,2,3</sup>, and Hyung Wook Kwon<sup>1,2,\*</sup>

<sup>1</sup>*Convergence Research Center for Insect Vectors, Incheon National University.*

<sup>2</sup>*Human Genome Research Center, Incheon National University.*

<sup>3</sup>*NGENI Foundation, San Diego, CA 92127, USA*

\*Corresponding author: [hwkwon@inu.ac.kr](mailto:hwkwon@inu.ac.kr)

Data cleaning, manipulation, and Exploratory data analysis is the most time-consuming process during the entire data analysis process. As there is a quote "Garbage in, Garbage out", The processes are key steps that repeat multiple times to get the right feature selection to get better outcomes. Meanwhile, Metadata that explains the data briefly is also very important for understanding the data, but currently, metadata for clinical data that has been accumulated for more than several decades has many problems, such as conflicts between variables. To reduce the time-consuming burden, We developed *CodeRbook* package aiming to help people with exploratory data analysis and writing and reviewing the metadata. The package interactively visualizes the overall data structure, distribution of values, and missing data, and could make summary graphics between two or more variables. Indeed, every function in the package was modularized so it could be reusable in various Shiny applications and databases based on the Shiny platform. In conclusion, The development of *CodeRbook* will help exploratory data analysis and data checks and reduce the most time-consuming processes during the entire analysis.

**Acknowledgment:** This work was carried out with the support of the Cooperative Research Program for Agriculture Science & Technology Development (Project No. RS-2023-00228817), the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03041954), and computing resources from Human Genome Research Center at Incheon National University.