Vasco A. Carinhas

University of Puerto Rico, Arecibo
Mentor: Emery Boose and Barbara Lerner

The Data's Story Made Accessible

The reproducibility of scientific results is a key element in any scientific experiment. In order to replicate the results, a storyline of the phases taken with the data should be delineated. With a Data Derivation Graph (DDG), a structured metadata that records the history of the data and its analysis, the scientist is able to establish the data's provenance, namely its derivation, by perusing the different steps the data took. The goal of this project is to make DDGs accessible to scientists from different computer backgrounds. To attain this objective, R was turned, with specific packages, into an interactive tool in which the user's understanding of R can be minimal to use R's features. *Quality Control Enforcer* is a Graphical User Interface tool that allows, using widgets, variables selection from a file and data input to perform modifications. The data can be subset by user input specifying a range. Also, quality control tests can be performed, with outliers and repeated values both depicted in different colors in the plot. Concurrently, the tool records the steps taken in a DDG. Such development in R allows an expansive experiment to test the DDG usefulness. The aims for the DDG development are to allow the scientists to explore the data analysis, and for the programmer to troubleshoot the R script and keep track of the program's process. A user-friendly platform is a crucial constituent to make the creation of DDGs easier in multiple scientific fields.