# Orenna Brand

Columbia University
Mentors: Emery Boose, Elizabeth Fong, Barbara Lerner
Group Project: Data Provenance in R

**Increasing the Use of Provenance Through User-Friendly Debugging in R**

For scientific results to be verified and expanded upon, they must be reproducible by other scientists. Precise reproducibility has historically been difficult to achieve; however, information about how data are analyzed can yield more transparent and repeatable results. This information is referred to as provenance—the record of all elements that contribute to a piece of data, including its intermediate values, operational dependencies, and computing environment. Essentially, it is documentation of how the data came to be in its current state. Our group previously developed tools that collect and visualize provenance in the R statistical language, such as RDataTracker and DDGExplorer, and that utilize the captured provenance, such as Rclean and Encapsulator However, more novel applications are afforded by this information, including tracing data lineage and debugging. We developed an R package, provDebugR, which uses provenance to facilitate postmortem debugging, or identifying and resolving issues in a script after it has completed execution. Through an iterative process of informally obtaining and integrating feedback from scientists, we implemented our application with intuitive functionality that lowers the barrier of entry and increases work efficiency. Features include an interactive graphical user interface and the ability to debug chunks of code at a time. Though the practice of collecting and archiving provenance is growing in popularity, it has yet to be fully integrated into existing workflows. The development of software applications that leverage the power of provenance, such as provDebugR, is important for shifting the focus from visualizing provenance to deriving meaningful insights.