

ABSTRACT OF THE DISSERTATION

Bridging the Gap Between Tools for Learning and for Doing Statistics

by

Amelia Ahlers McNamara

Doctor of Philosophy in Statistics

University of California, Los Angeles, 2015

Professor Robert L Gould, Co-chair

Professor Fredrick R Paik Schoenberg, Co-chair

Computers have changed the way we think about data and data analysis. While statistical programming tools have attempted to keep pace with these developments, there is room for improvement in interactivity, randomization methods, and reproducible research.

In addition, as in many disciplines, statistical novices tend to have a reduced experience of the true practice. This dissertation discusses the gap between tools for teaching statistics (like TinkerPlots, Fathom, and web applets) and tools for doing statistics (like SAS, SPSS, or R). We posit this gap should not exist, and provide some suggestions for bridging it. Methods for this bridge can be curricular or technological, although this work focuses primarily on the technological.

We provide a list of attributes for a modern data analysis tool to support novices through the entire learning-to-doing trajectory. These attributes include easy entry for novice users, data as a first-order persistent object, support for a cycle of exploratory and confirmatory analysis, flexible plot creation, support for randomization throughout, interactivity at every level, inherent visual documentation, simple support for narrative, publishing, and reproducibility, and flexibility

to build extensions.

While this ideal data analysis tool is still a few years in the future, we describe several projects attempting to close the gap between tools for learning and tools for doing statistics. One is curricular, a high school level Introduction to Data Science curriculum. Others are technological, like the experimental LivelyR interface for interactive analysis, the **MobilizeSimple** R package, and **Shiny** microworlds.

Much of this work was inspired by Biehler (1997), which describes the attributes necessary for a software package for learning statistics. Biehler’s vision has largely come to light, but advancements in computing and ‘data science’ have moved the goalposts, and there is more to learning statistics than before. This work envisions a tool not only encompassing these advancements, but providing an extensible system capable of growing with the times.