

ADAM RULE'S DEFENSE DESIGN AND USE OF COMPUTATIONAL NOTEBOOKS

ABSTRACT:

Individuals and organizations increasingly rely on data analysis to generate insights and make decisions. Yet, small changes to how data are collected, cleaned, or modeled can lead to vastly different results. If data-driven insights are to be reviewed, reused, or trusted the process used to generate them must be tracked and communicated in detail. But data analysis is typically an iterative and exploratory process that is hard to articulate, especially when it involves programming. Computational notebooks aim to ease tracking and sharing of complex analyses by enabling analysts to write rich *computational narratives* combining executable code, interactive visualizations, and explanatory text in a single document. While millions of people use computational notebooks, we know little about how they use them, or how well they help people track and share complex analyses.

In this dissertation I present three studies of how people currently use computational notebooks, demonstrating that few notebooks, even those published alongside academic papers, have much in the way of narrative. Instead, most notebooks are loose collections of notes and scripts that even the original analyst struggles to understand. I then present two systems demonstrating how computational notebooks might be designed to support clearer communication of complex analyses. The first system, Janus, shows how current notebooks might be modified to aid both ongoing analysis and later communication by adding interactive hierarchy for selectively showing and hiding portions of the notebook. The second system, ActiveNotes, a prototype clinical note editor, demonstrates how computational notebooks might support data-driven work even when programming is not the primary means of interacting with data.

Together, these studies demonstrate that tracking and sharing complex analyses is hindered by a tension between exploration and explanation, but that computational notebooks and other media can reduce this tension by supporting not only the combination of, but also flexible organization and navigation of analytical steps, explanatory text, and results.

ADAM RULE TO
DEFEND

WEDNESDAY,
JUNE 6TH, 2018 AT
3:00PM

COGNITIVE SCIENCE
BUILDING,
ROOM 180

COMMITTEE MEMBERS:

James Hollan, Chair, Cognitive Science

Robert El-Kareh, Medicine

Gloria Mark, Informatics at UC Irvine

William Griswold, Computer Science &
Engineering

Philip Guo, Cognitive Science

Bradley Voytek, Cognitive Science