



psyphr.read: wrangle and tidy psychophysiological data in R

Kathleen E. Wendt, Siqi Zhang, Mallory J. Feldman, G.Brooke Anderson, & J. Douglas Coatsworth

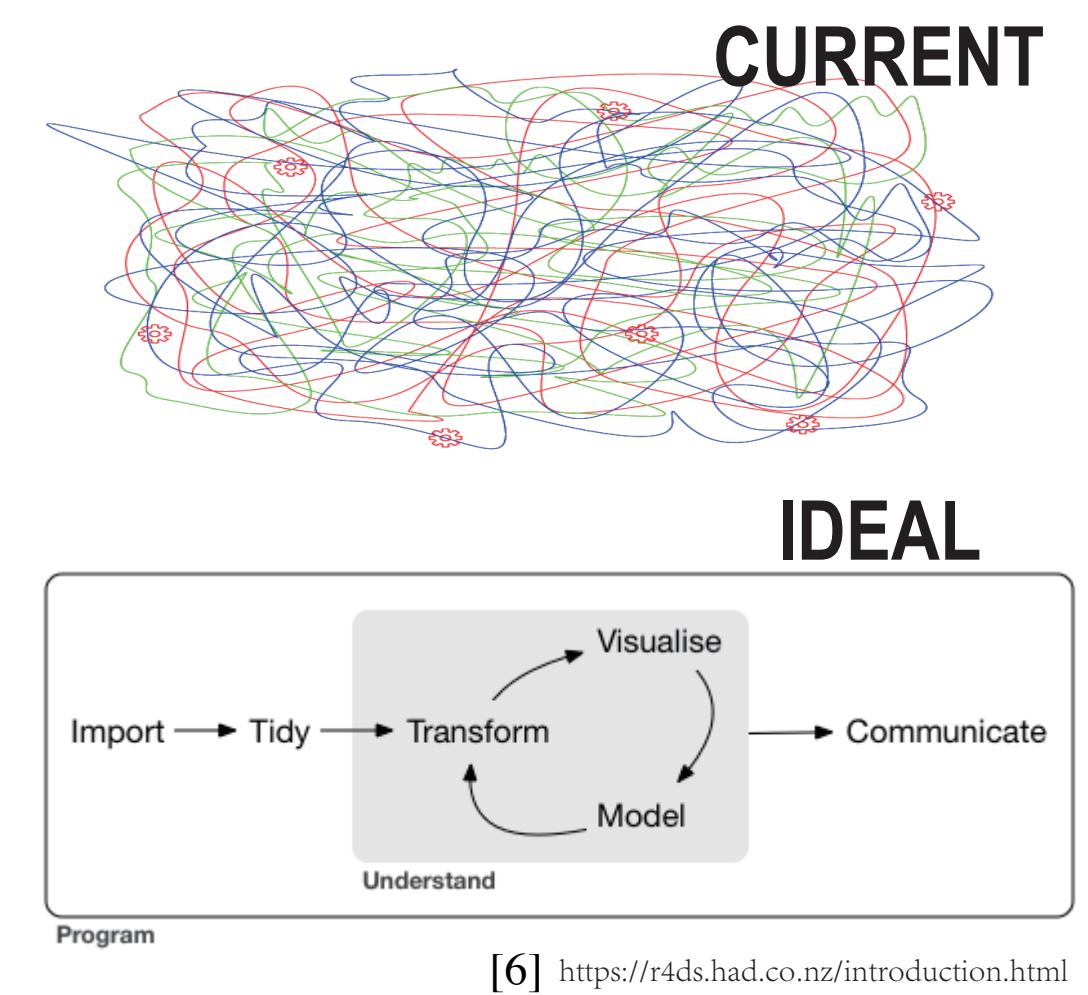


1 a burgeoning field

Physiological measures are increasingly popular in social, behavioral, and psychological research but yield unwieldy amounts of data and bring their own challenges [1]. To date, few tools have been developed [2] to support reproducibility, efficiency, and accuracy [3-5] in psychophysiological data processing and management.

2 workflow concerns

- Output files from proprietary software are often compiled/aligned by hand (e.g., via copy-paste)
- Manual processes introduce opportunities for bias and error
- Error-ridden data produces error-driven inferences
- Lack of standardization obstructs shareability and reproducibility

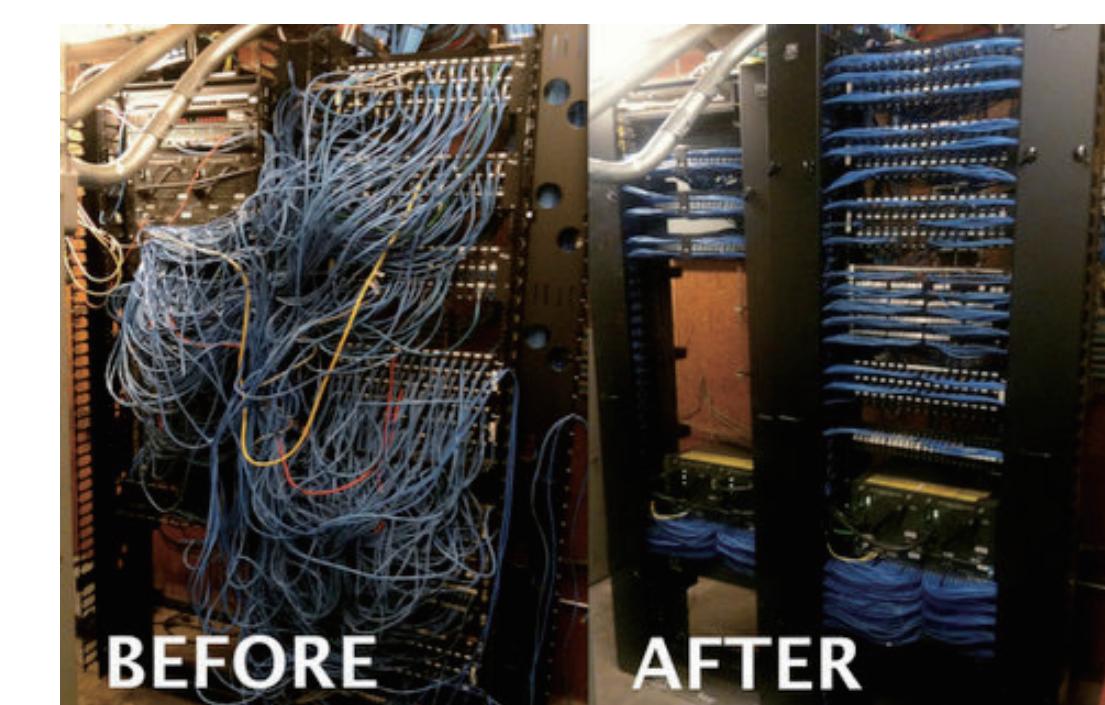


3 tidy data

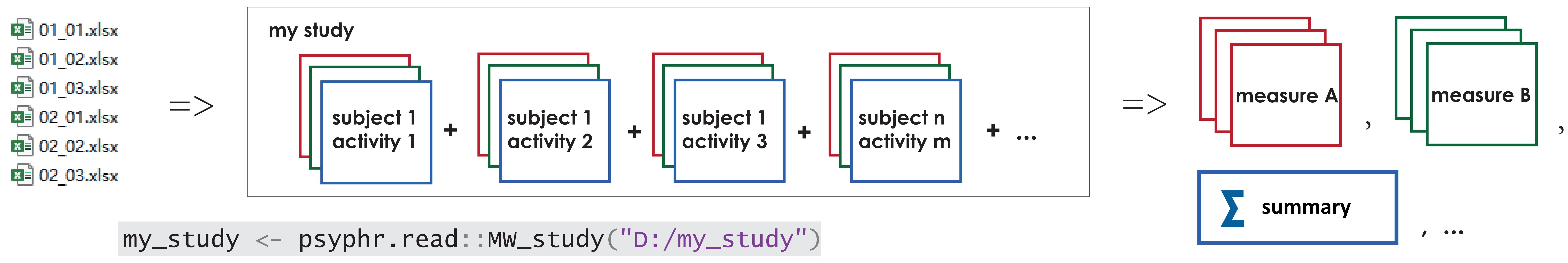
A main principle of computational reproducibility = tidy data [7].

- Each column as a variable
- Each row as an observation
- Each table as a measure

Our work extends these tidy principles.



<https://advwireonline.com/cable-management.html>



4 `psyphr.read` package

... is the first in a suite of R packages. With one line of code, the user combines and organizes the entire study (see above illustration). In this tidy form, the user can quickly and easily generate summary statistics and exploratory plots. `psyphr.read` also includes simple data validation techniques to catch errors or problems early.

5 `psyphr` suite; future directions

<code>psyphr.read</code>	<code>+</code>	<code>psyphr.check</code>	<code>+</code>	<code>psyphr.plot</code>	<code>+</code>	<code>psyphr.model</code>	<code>+</code>	<code>psyphr.bids</code>
<i>Ingesting and parsing data from proprietary systems</i>		<i>Sophisticated data validation schemes</i>		<i>Publication-grade plots</i>		<i>Statistical models for psychophysiological measures</i>		<i>Compatibility with Brain Imaging Data Structure (BIDS) [8]</i>

The full `psyphr` suite will help researchers spend less time on data management and more time “doing science.” This work will provide the foundation of open science practices for psychophysiological research and promote computational reproducibility in fields with increasingly large and complex datasets.

