

Reproducible research

An introduction for the R novice

Richard Layton

Department of Mechanical Engineering
Rose-Hulman Institute of Technology
Fall 2016



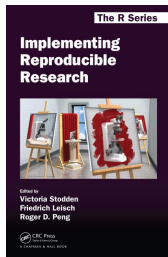
Getting started

- ▶ Introductions
- ▶ Handouts

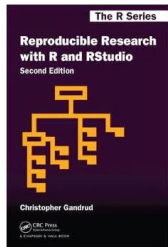
Getting started

- ▶ Introductions
- ▶ Handouts
- ▶ Mystery question: *What is reproducible research?*

Practitioners tell us:



Research is reproducible when the data and the code used to obtain a finding are available and sufficient for an independent researcher to recreate the finding.



- ▶ computational, data-intensive
- ▶ spans the full data, analysis, & publication workflow
- ▶ most of us have received only perfunctory training (if any)

Events tell us:

More accountability is needed because of

- ▶ data falsification
- ▶ erroneous analysis
- ▶ misleading presentation of results

Karen EC Levy & David Merritt Johns, *When open data is a Trojan Horse: The weaponization of transparency in science and governance*, *Big Data and Society*, 2016.



Reproduction showed their primary findings were false.

Results were used to justify austerity policies, but the major effect disappeared after correcting for

- ▶ coding errors
- ▶ selective exclusion of available data
- ▶ unconventional weighting of summary statistics



Kenneth Rogoff & Carmen Reinhart

Reproduction showed he falsified data.

To obtain the results he wanted, he altered data in several ways. Consequences included:

- ▶ terminated clinical trials (real patients)
- ▶ civil suit by patients
- ▶ cancelled research funding
- ▶ retracted journal articles (11 to date)



Anil Potti, formerly a cancer researcher at Duke University

“Weaponized” reproduction is used to discredit scientists.

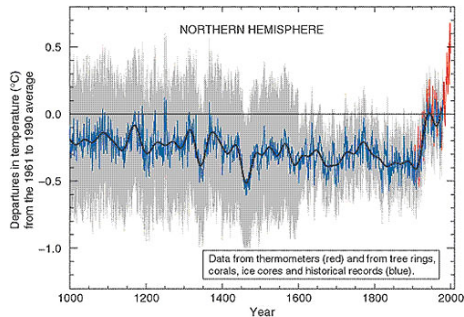
Mann’s work has withstood 15 years of scrutiny—and still holds up. But he refused to share.

Scientists and “skeptics” are in a knife fight, and you don’t bring data to a knife fight.

— Paul Erlich

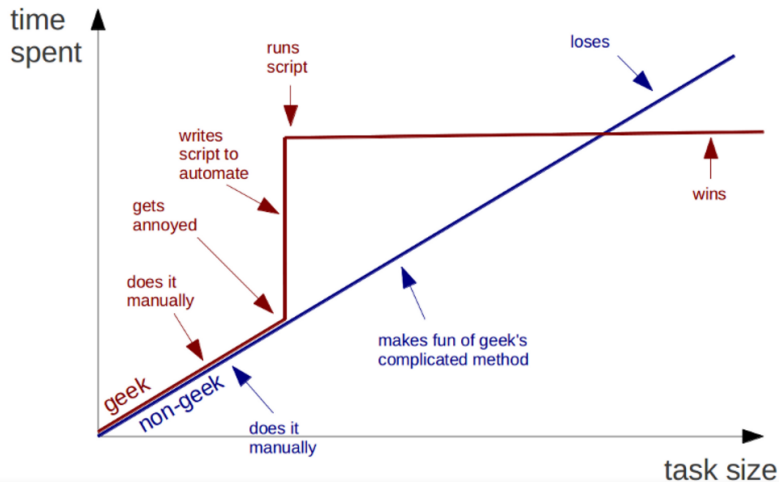
Why should I make the data available to you, when your aim is to try and find something wrong with it?

— Phil Jones



1000 years of temperature variation
with uncertainties, Michael Mann

The benefits first accrue to you.



- ▶ reproducible for your future self
- ▶ faster updating of results
- ▶ faster resumption of work after hiatus

Steps towards reproducibility

- ▶ Write scripts (avoid manual copy, paste, mouse-clicks)
- ▶ Plan the organization and naming scheme for files
- ▶ Strive for simplicity & readability
- ▶ Write for reusability & testability
- ▶ Agree on a workflow for collaborating before starting a manuscript
- ▶ DRY (don't repeat yourself)
- ▶ Link files explicitly
- ▶ Use version control
- ▶ Plan data management
- ▶ License your software
- ▶ Manage package dependencies

Steps towards reproducibility: in this course.

- ▶ Write scripts (avoid manual copy, paste, mouse-clicks)
- ▶ Plan the organization and naming scheme for files
- ▶ Strive for simplicity & readability
- ▶ Write for reusability & testability
- ▶ Agree on a workflow for collaborating before starting a manuscript
- ▶ DRY (don't repeat yourself)
- ▶ Link files explicitly
- ▶ Use version control
- ▶ Plan data management
- ▶ License your software
- ▶ Manage package dependencies

How the course work is organized.

- ▶ Course materials are reproducible.

<https://dsr-rhit.github.io/me497-reproducible-research/>

- ▶ Syllabus

- ▶ Week-0 assignment

Consider a sample report.

- Imagine that you were the author of the “Load cell calibration report”

Load-cell calibration report

Richard Layton

2016-08-24

Introduction

The goal of this analysis is to determine the calibration equation and sensor accuracy for an Omega LCL-005 (0--5 lb) load cell.

The test setup is illustrated in Figure 1. Precision weights (0.1% accuracy) are used to apply the reference force (lb) to the load cell and the resulting voltage readings (mV) from the sensor are recorded. The test procedure follows the ANSI/ISA standard.

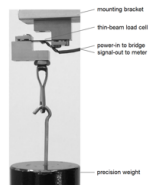


Figure 1. Load cell calibration test setup

Data

The calibration data are shown in Table 1. The maximum force (4.5 lb) is 90% of the 5 lb sensor limit, per the ANSI/ISA standard. The NA entries in the first and last columns are artifacts of the ANSI/ISA test procedure (the test starts and stops at a mid-range test point in the same direction).

Consider a sample report.

- Imagine that you were the author of the “Load cell calibration report”
- Mystery question: Identify as many “manual operations” as possible.

Load-cell calibration report

Richard Layton

2016-08-24

Introduction

The goal of this analysis is to determine the calibration equation and sensor accuracy for an Omega LCL-005 (0–5 lb) load cell.

The test setup is illustrated in Figure 1. Precision weights (0.1% accuracy) are used to apply the reference force (lb) to the load cell and the resulting voltage readings (mV) from the sensor are recorded. The test procedure follows the ANSI/ISA standard.

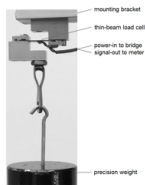


Figure 1. Load cell calibration test setup

Data

The calibration data are shown in Table 1. The maximum force (4.5 lb) is 90% of the 5 lb sensor limit, per the ANSI/ISA standard. The NA entries in the first and last columns are artifacts of the ANSI/ISA test procedure (the test starts and stops at a mid-range test point in the same direction).