PROJECT TIER

Teaching Integrity in Empirical Research

Spring 2017 Faculty Development Workshop

Haverford College

March 31-April 1, 2017

@Project_TIER

www.projecttier.org

Project TIER is supported by the Alfred P. Sloan Foundation.

What is Project TIER About?

Broadly, our goal is to collaborate with instructors and researchers from across the social sciences to develop and disseminate curriculum and tools for teaching principles and methods of transparent quantitative research to students in all disciplines of the social sciences and related fields.

But let's start by discussing some narrow goals in a concrete way.

From that basis, some broader and more conceptual aspects of what we are doing will emerge.

To date, the focus of Project TIER has been on teaching students to assemble replication documentation for empirical research papers they write.

And the main purpose of the documentation is to make it easy for someone else (like the course instructor or thesis advisor) to replicate all the data processing and analysis conducted for the project that is necessary to reproduce the results reported in the paper.

Why did we choose this focus for Project TIER?

Because of a concern about a crisis of credibility in social science research?

Because of our devotion to principles of the scientific method?

Out of zeal for combatting flawed, misleading and even fraudulent us of statistics in the social sciences?

Although all of those issues have turned out to be relevant, the answer is “none of the above.”

In fact, it was a trick question, because we didn’t start out by choosing a focus for a project.

What started all this was a prosaic problem:

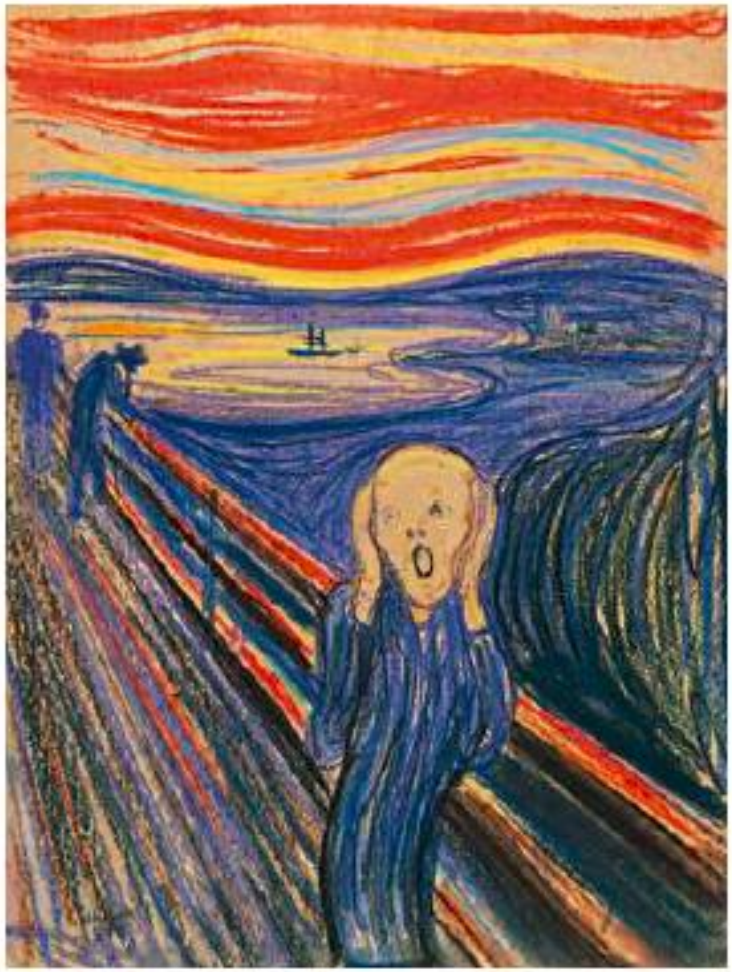
We couldn't understand what students were talking about in their research papers.

In particular, we couldn't understand

--their descriptions of the data they had started out with
--how they had processed the data, constructed new variables, chosen which cases to exclude, etc.
--how the variables in the final data set were defined
--exactly what tests or procedures they had conducted to generate the results

And it gets worse:

> Conversations in which we asked students to clarify these questions for us revealed that they themselves did not understand well the data they had started with, how they had manipulated them, and the analyses they had conducted.



But fixing that is easy, right?

Just get them to turn in their data and command files, as illustrated in this excerpt from the instructions to our Fall 2005 class:

--At the same time you turn in your final papers, you should create an electronic folder whose name contains the names of the members of your group, and in it you should place electronic copies of your three do-files (import.do, cleaning.do, results.do) and of the raw data set you downloaded from ICPSR. Then copy that folder into the folder called "Econ 203 Final Projects (Do-files and data sets)" that is in the public read_write folder in my personal folder on the storage server. The path to get there is:

\\storage\users\r\rball\public_readwrite\Econ 203 Final Projects (Do-files and data sets)

It took a few iterations, but by about 2010 we had developed these instructions into what we eventually started calling the TIER Documentation Protocol (version 1).

Since about 2010, students in our intro statistics classes have been following this protocol and producing replication documentation for their papers, and achieving a high standard of success at a high rate.

Also around 2010, some of our econ majors started preparing similar documentation for their senior theses, again with good success. In cases where the replication documentation students produced was really excellent, we began posting the replication files along with the pdf of the thesis on Haverford's public archive of senior theses.

A few years ago the Haverford economics department agreed on the policy that all seniors who write empirical theses must prepare complete replication documentation. In practice…we make small steps toward achieving that goal every year.

Once we had formulated a protocol that was working tolerably well, it occurred to us that this approach to reproducible research might be of interest to other faculty who teach statistics and social science research methods:

> --When we talked about this work with friends and colleagues, many said they thought this is important and that they had been vaguely considering doing something along these lines, but just had not gotten around to figuring out exactly how to integrate it into their classes or to type up a handout.

--And it occurred to us that this was related to increased concerns being expressed about transparency/reproducibilitly/credibility/robustness of professional research in the social sciences.

So we started Project TIER to learn what others are doing in this area and to share our experiences with interested instructors.

That's how we arrived at the broad goals we stated at the outset.

The main “products” Project TIER has promoted to date are:

The TIER Documentation Protocol

Using the Open Science Framework (OSF) to manage a reproducible workflow

The focus of those products is computational reproducibility.

However:

--Teaching students to take measures to ensure their research is computationally reproducible leads to a number of other pedagogical benefits, some of which we believe are central to the entire enterprise of education.

--We are working with colleagues to develop curriculum that addresses other dimensions of research transparency (e.g., p-hacking, pre-analysis plans, experimental replicability, protection of the welfare of human research subjects)

The main principles underlying the TIER Protocol:

Replication documentation should be

--Complete (soup-to-nuts)

--Independent

--Portable

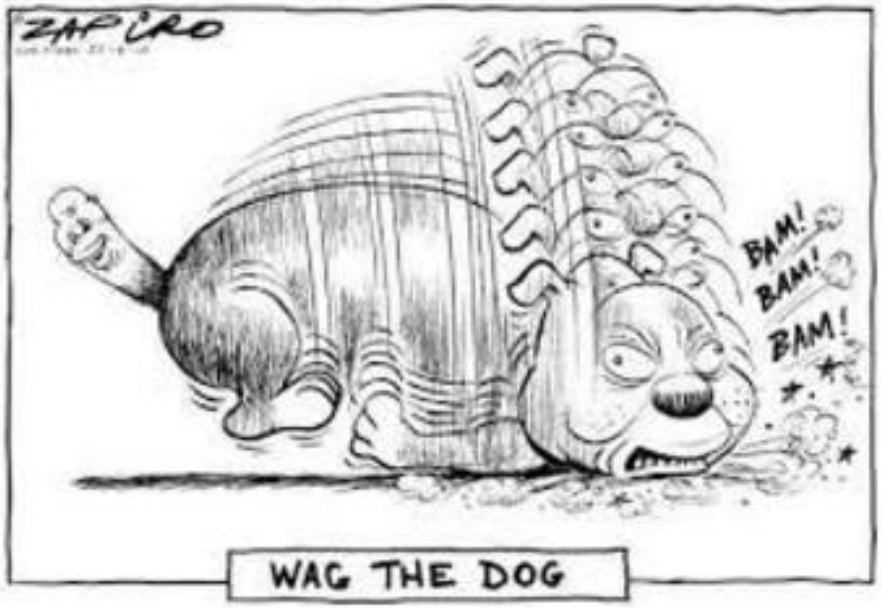
The TIER Protocol is written in a way that makes it sound didactic.

--That is because we want it to provide concrete guidance that is easy to follow. (We are not by nature bossy.)

--Unusual situations will arise in which the particular specifications of the Protocol are poorly suited or infeasible; creative adaptation is necessary and encouraged.

One last bit of perspective:

The purpose of making good replication documentation is to improve the quality and usefulness of the research that is being conducted; it is not an end in itself.



Incorporating reproducibility and transparency in classes and research advising

I. In the context of a semester-long course with a research paper or a thesis or dissertation

A. The TIER Protocol "Process"

"Pain points"

--establishing a workflow and file organization system—OSF can help a lot

--"take it from the top," hands-on advising

--no data processing by hand

--data appendix before analysis

--relative directory paths and portability??

B. Preparing written instructions for your students

--TIER website vs. customized handouts (or a hybrid solution)

--The installment plan: scaffolded assignments

C. Pre-project warm-ups

Provocative readings about high-profile failures of reproducibility

Broader readings about research transparency and open science movement

Work through demo project or previous student work

Try an “ex-post replication” before doing your own “ex-ante documentation”

D. Evaluation

E. How to make room for reproducibility in your already packed curriculum?

II. Other contexts, such as large classes in which individual or small group research projects may not be feasible

"Soup to nuts" reproducibility exercises