Reproducing and Replicating Spatial Data Science:

Tutorial with Breakout Discussion

Scientific research is increasingly expected to be reproducible as a matter of transparency and public trust in research, such that other researchers can use the same data and methods to produce the same results. Reproducibility and replicability are also integral to the mechanisms of self-correction and theory development in science. Reproduction studies are needed to evaluate the internal validity of prior research findings. Replication studies with different data and in different geographic contexts are needed to assess the external validity and generalizability of prior research claims. Researchers are increasingly motivated to adopt reproducible research practices to meet expectations of publishers and funding agencies and to expand the broader impacts of their work. Researchers and students are motivated to reproduce and replicate prior studies in order to learn from their methods, assess their validity, and extend from or build upon their prior work. There are concurrent and interdependent needs to develop:

1. infrastructure to facilitate reproducibility and open science,
2. exemplar cases of reproduction and replication studies in spatial data science, and
3. a reproducibility and replicability curriculum.

We will present working prototypes of infrastructure, exemplar cases, and curriculum developed over the first two years of the National Science Foundation award, *Transforming theory-building and STEM education through reproductions and replications in the geographical sciences*. Following presentations, we will form breakout groups to discuss applications to individual research programs and future steps for scaling up reproducible research practices in spatial data science.

We will present *infrastructure* in the form a Git repository template for reproducible research compendia and handbook. The template research compendium and handbook help guide the research process and organize research materials while maximizing reproducibility, and may be applied to both individual and collaborative research; and to original studies or reproduction/replication studies. The compendium includes space for project-level metadata and organizing documentation, intellectual property license, and preferred citation. The directory structure organizes space for protocols and code, proprietary and public data, metadata, results, documents, and manuscripts. We include templates for pre-analysis plans and post-analysis reports and guidance for registrations and integration with the Open Science Foundation (OSF). Finally, we include sample Rmarkdown and Python Jupyter Notebook files with useful code and structure for maximizing reproducibility. Our tutorial introduction to the *infrastructure* will be sequenced and paired with our *approach* to conducting and publishing reproduction and replication studies. This infrastructure has been developed and refined over the course of completing seven reproduction or replication studies with undergraduate students in three courses, graduate students in two courses, and numerous graduate and undergraduate research assistants and independent researchers.

We will present our *infrastructure* using an *exemplar case* of a spatial data science study we have reproduced with our students. The study models best practices for open and reproducible science while highlighting the contributions and advantages of reproduction studies. The case also highlights approaches and advantages to using reproduction studies as an integral component of spatial data science *curriculum*. The majority of the reproduction study has been implemented, written, and presented by students. The infrastructure (template and handbook), exemplar case, and course curriculum are all being made available to the public with open access licensing so that tutorial participants can review and reuse the materials in their own scholarship.

We will conclude the tutorial with *breakout groups* to discuss:

1. sharing questions or concerns about our approach to reproducibility and replicability (R&R),
2. developing action plans for adoption in current research and teaching programs, and
3. collectively discussing next steps to scale up R&R in spatial data science.

Evaluation

We are interested in conducting pre- and post-surveys based on The Unified Theory of Acceptance and Use of Technology (UTAUT) survey instrument with tutorial participants as part of our research on pedagogy and reproducibility.

Team members

- Dr. Joseph Holler, Middlebury College
- Dr. Peter Kedron, Arizona State University
- PhD Candidate Sarah Bardin, Arizona State University

Expected participation

We anticipate that this tutorial will be interesting and valuable for the following groups:

- Graduate students interested in learning methods from the literature and either publishing reproduction or replication reports or extending prior research in their own master's theses or PhD dissertations.
- Faculty and career researchers interested in adopting more open and reproducible research practices for their own original work, or simply crafting more competitive data management plans and working more efficiently and accurately.
- Professors interested in teaching reproduction or replication studies in their courses or as part of their advising and mentoring
- Graduate advisors and research mentors interested in training their research assistants / advisees on more reproducible practices
- Journal editors interested in publishing reproduction or replication studies, or more thoroughly incorporating reproducibility into their review of original research

Website: http://sdss2023.spatial-data-science.net/ Contact: Kitty Currier kcurrier@ucsb.edu