Practicum in Reproducible Research Methods

Professor Miriam Golden
office: Badia Fiesolana, BF263
email: `miriam.golden@eui.eu`
contact: Monika Rzemieniecka (`Monika.Rzemieniecka@eui.eu`)

<div align="center">

Second term seminar 2022–23
Tuesday 15:00-17:00
Seminar Room 2

Course site: https://mycourses.eui.eu/d2l/home/12844
Zoom information available on Brightspace

</div>

Course Description: This course walks you through all the steps involved in a complex collaborative reproducible research project, from thinking up a problem worth studying, to pre-registering your design, to collecting and analyzing your data, and writing up and posting your results. You will practice these steps using real data collected by the instructor. To master the skills involved, you will work in teams and hand in bi-weekly homework assignments.

Course Prerequisites: You will only be able to do the work in this course if you are familiar with statistical methods to analyze quantitative data. Thus, students are welcome in the course if they have taken at least one prior course in statistics (covering material through multiple regression) that used either Stata or R. All other skills required this term will be taught by the instructor.

Course Objectives: At the completion of this course, you will:

1. Have experience working with a complex multilevel dataset.
2. Have practiced many activities required to complete a large-scale reproducible research project.
3. Have experience collaborating with others on a research project.
4. Have worked with tools such as LaTeX, RMarkdown, and GitHub, and have developed familiarity with sites such as BITSS, OSF, EGAP, Dataverse, and others.

5. Have improved research practices and skills.
6. Be knowledgeable about the highest standards and practices associated with many aspects of reproducible research.
7. Have gained an overview of the entire research process that will give you a more realistic and complete idea of the timeframe, intellectual commitment, and skills involved in real research.

Course Format: This course is hands-on and requires active engagement with data, the internet, and various applications during class time. Please make sure your setup is appropriate. You will need a stable and robust internet connection to do the work during class meetings. Class sessions will ask you to work on your laptop and potentially to access the internet.

Readings: The readings are typically short and skills-oriented. Thus, they require a very specific kind of reading. You should first skim an assigned reading for the skills it covers and then carefully study the sections that are relevant to what you want to accomplish. At that point, you should **implement the techniques** that the reading discusses. If you are reading about how to structure your directories for a research project, take a look at your own directory structure for your most active research project and consider how the reading would suggest you modify it. Then go ahead and restructure your directories for that project. The only way to learn the skills that the practicum readings cover is to implement them, thereby developing "muscle memory." Over the course of the term, incorporate as many of the skills we study into your own workflow as possible. Your goal is to make them part of your standard research routines.

Most entries on this syllabus are hyperlinked to the publication site.

We read almost all of two books this term, which you may wish to purchase:

Gandrud, C. (2020). *Reproducible Research with R and RStudio*. CRC Press, Boca Raton, FL, 3rd edition.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science*. University of California Press, Oakland.

I will be teaching from the 3rd edition of the Gandrud book. It contains extensive new material on integrating RMarkdown into your workflow that is not available in the 2nd edition.

I also recommend reading much of Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded*. Oxford University Press, Oxford. Although the library has purchased access to the electronic version, only three readers can access it simultaneously. For this reason, you may wish to rent the book for the term or purchase personal access.

Finally, readings are front-loaded in the term. The first four weeks feature a lot of assigned reading. The quantity drops off very considerably after that. Your goal is to use the first month of the course to study and start to implement the new research skills we focus on. In the remaining six weeks, your goal is to extend the activities and to apply them and make them routine in how you work.

Data: You may select to work with one of two datasets this term:

- A dataset of all Italian legislators who served in the X and XI and the XV and XVI Legislatures, which includes information on charges of criminal activity and corruption, press coverage of the individual, and standard demographic and political characteristics. This study uses observational data.

- A dataset containing information on how a random sample of 14,400 respondents in 20 electoral constituencies in one Pakistani province reacted in 2018 after receiving questions delivered by Interactive Voice Response technology from 20 randomly chosen Members of the Legislative Assembly (MPAs) from all political parties. The data in this study was generated experimentally.

Each dataset is accompanied by an article that presents and analyzes the data. The datasets are available on Brightspace.

Your team will work with the same dataset all term. This allows you to gain familiarity with it so you focus on mastering new skills rather than figuring out a new dataset.

If you wish to work with your own data instead of instructor-provided data, please discuss this with me by email before the first class meeting.

Software(s): You may use either Stata or RStudio to do the statistical work for the course. For all assignments, you will write in RMarkdown. I will demonstrate how to integrate R and Stata into a Markdown document. You will produce graphics using ggplot2.

You are expected to present descriptive statistics and regression results in your work. You are discouraged from using more advanced statistical techniques in this course. This is a course in which you build research — not statistical — skills. All work must be accurate, complete, and thoughtful.

Homework Assignments: All homework should be submitted in .pdf format unless otherwise specified. Some assignments require you integrate your statistical code into your Markdown document. Code must be a clean run and must be properly annotated. Your code may be written in Stata or in R. **If your code does not run, your assignment will be considered incomplete and you may not receive credit for it**. Please make sure to write relative directory paths so that your code will run on my

computer with a single line change in the first section of the file indicating the name and location of the working directory.

Course Website: I plan to use the Brightspace Learning course site. Any materials not already in the public domain will be posted on Brightspace. You will also submit your homework assignments there. The site is also useful for conversations with other students, as well as to ask questions of me.

If there is student interest in switching to GitHub instead of Brightspace for submission of homework assignments, please let me know. Similarly, I am willing to adjust assignments and readings on the syllabus in response to student needs. Just let me know as soon as possible.

Collaboration: Students will work in teams for the term. Teams will be set up during the first class meeting. Team size will depend on the number of students enrolled, but you will probably be paired with one or two others. You will remain with the same team-mate(s) for the entire term. All homework assignments will be jointly submitted and jointly graded.

Requirement(s): To complete the course for a grade, each team will submit an assignment every other week. The assignments will be available on the Brightspace course site. They will be blind-graded.

Course Policies:

- Students should come to class meetings each week **already having read** the assigned material.

- Students should come to class meetings each week **already having practiced** whatever specific skills are assigned that week. That way, you will be well positioned to bring questions and problems to the attention of the instructor.

- Because you cannot learn skills without practicing them, this course is not open to students to sit in without fully participating. If you are not taking the course for credit, please contact Monika to be granted access to course materials.

- All students are expected to be active class participants.

- Materials are to be submitted on time to be given full credit. Please ensure that the timestamp for every submission is within ten minutes of the time due.

Other Resources:

In addition to readings and sites listed on the syllabus, you will find it useful to familiarize yourself with the following sites and resources:

Empirical Studies of Conflict (ESOC) resource links

Berkeley Initiative for Transparency in the Social Sciences (BITSS) resource library

World Bank Reproducibility Training

NBER Summer Institute 2019 Methods Lecture: Research Transparency and Reproducibility

2019 EASST/BITSS Transparency Training Workshop

Open Science Framework (OSF) Guides for Best Practices

Declare Design

Harvard Dataverse (data repository)

ICPSR (data repository)

Evidence in Governance and Politics (EGAP) Methods Guides

Evidence in Governance and Politics (EGAP) Policy Briefs

Center for Effective Global Action(CEGA) Policy Briefs

International Growth Centre(ICG) Impact Case Studies

CITI Program for Social-Behavioral-Education (SBE) Basic training in the protection of human subjects

Social Science Prediction Platform

Institute for Replication

Week One, January 10: Getting started with reproducible research

- Overview: why reproducible research matters
- Exploring your dataset
- Picking a research question
- Forming a research team
- Setting up SOP and lab arrangements
- Writing your team MOU
- Working in RMarkdown
- Setting up your project on GitHub

Gandrud, C. (2020). *Reproducible Research with R and RStudio.* CRC Press, Boca Raton, FL, 3rd edition, chs. 1–3 and 12.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science.* University of California Press, Oakland, ch. 11.

Lin, W., Green, D., and Coppock, A. (2016). Standard operating procedures for Don Green's lab at Columbia.

Shapiro, J. N. (2019). ESOC research production guide, v1.0.

Ram, K. (2013). Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine*, 8(7).

First homework assignment available. It will require each team to use RMarkdown to write a collaborative agreement, set up a GitHub repo for team use all term, and establish an initial structure for your project directories.

Week Two, January 17: Establishing your workflow

- Overview of reproducible workflow
- Organizing your repo
- Using readme documents
- Commenting your code

Gandrud, C. (2020). *Reproducible Research with R and RStudio.* CRC Press, Boca Raton, FL, 3rd edition, ch. 4.

Gentzhow, M. and Shapiro, J. M. (2014). Code and data for the social sciences: A practitioner's guide.

Wilson, G., Aruliah, D., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and WIlson, P. (2014). Best practices for scientific computing. *PLos Biology,* 12(1):1001745.

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and Teal, T. K. (2017). Good enough practices in scientific computing. *Plos Computational Biology.*

Bowers, J. and Voors, M. (2016). How to improve your relationship with your future self. *Revista De Ciencia Politíca,* 36(3):829–48

and/or

Evidence in Governance and Politics (EGAP). 10 things to know about project workflow.

Familiarize yourself with:

Open Science Framework (OFS). OSF Guides (Best Practices), especially the sections on File naming and Organizing files.

Datasets

Sometime before week four, when you will be required to work with data, please read the article that is relevant for the dataset your team will work with this term:

Asquer, R., Golden, M. A., and Hamel, B. T. (2020). Corruption, party leaders, and candidate selection: Evidence from Italy. *Legislative Studies Quarterly,* 45(2):291–325.

or

Golden, M. A. and Gulzar, S. (2022). "Press 1 for roads:" descriptive and experimental evidence on political communication. Unpublished paper.

First homework assignment due

Week Three, January 24: Planning your study

- Designing your (pilot) study
- Securing funding
- Setting up a project pipeline
- Thinking through your research design
- Establishing a partnership in the field

Gandrud, C. (2020). *Reproducible Research with R and RStudio.* CRC Press, Boca Raton, FL, 3rd edition, ch. 5.

Przeworski, A. and Soloman, F. (1995; rev. 1988). On the art of writing proposals.

Sohn, E. (2020). Secrets to writing a winning grant. *Nature*, 555:133–35.

Recommended: Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded.* Oxford University Press, Oxford, ch. 4.

Familiarize yourself with:

Blair, G., Cooper, J. J., Coppock, A., and Humphreys, M. (2019). Declaring and diagnosing research designs. *American Political Science Review*, 113(3):838–59.

and

Declare Design website

Week Four, January 31: Pre-analysis plans, data management plans, and pre-registration

- Writing your pre-analysis plan
- Writing your DMP
- Registering your pre-analysis plan at EGAP

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science.* University of California Press, Oakland, chs. 3–6.

Humphreys, M., Sanchez de la Sierra, R., and van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21(1):1–20.

This version: 2023/01/06 at 11:40:43

Evidence in Governance and Politics (EGAP). 10 things to know about pre-analysis plans.

McKenzie, D. (2012). A pre-analysis plan checklist. World Bank Blogs.

Ofosu, G. and Posner, D. N. (2022). Pre-analysis plans: An early stocktaking. *Perspectives on Politics*.

EUI Library (2022). Research data guide. 10th ed.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018.

Van den Akker, O. R., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P. E., Hall, A. N., Kosie, J. E., Kruse, E., Olsen, J., Ritchie, S. J., Valentine, K., van 't Veer, A. E., and Bakker, M. (2020). Preregistration of secondary data analysis: A template and tutorial.

Review carefully:

EUI Research Data Services, especially 3. Data Management Plans.

Second homework assignment available. It will require each team to examine the dataset provided by the instructor and write a PAP designed to answer a specified research question. Your PAP must include a Data Management Plan, which you will prepare via DMPonline.

Week Five, February 7: Research ethics and human subjects protection

- Ethical obligations to subjects
- Ethical obligations to staff
- Obtaining approval from the Ethics Committee or IRB
- CITI certification
- Informed consent

National Commission for the Protection of Human Subjects of Biomedical and Behavioral

Research (1978). *The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research.* U.S. Government Printing Office, Washington, D.C.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science.* University of California Press, Oakland, ch. 2.

APSA Ad-Hoc Committee on Human Subjects Research (2019). Report of the APSA ad-hoc committee on human subjects research.

World Health Organization, Research Ethics Review Committee. Informed consent template for qualitative studies. Geneva.

Procedure to request an Ethics Review by the EUI Ethics Committee. Read all linked documents and forms on the site.

Familiarize yourself with:

CITI Program for Social-Behavioral-Education (SBE) Basic training in the protection of human subjects. (*Note that the EUI is now making CITI training available to faculty and graduate students and that the Ethics Committee expects to make such training mandatory in the near future.*)

Second homework assignment due.

Week Six, February 14: Forecasting results

- Why forecast your results
- Programming your forecasting instrument in Qualtrics
- Using the Social Science Prediction Platform for forecasting

DellaVigna, S. and Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy,* 126(6):2410–56.

Della Vigna, S., Otis, N., and Vivalt, E. (2020). Forecasting the results of experiments: Piloting an elicitation strategy. *AEA Papers and Proceedings,* 110:75–79.

Green, K. C. and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research,* 68:1678–85.

Recommended: Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded.* Oxford University Press, Oxford, chs. 3, 16, and 20.

Sign up for an account and make at least one prediction on:

Social Science Prediction Platform.

Week Seven, February 21: Field work and data collection

- Staying safe in the field
- Managing staff and enumerators
- Ethics in the field
- Collecting your data
- Storing and organizing your data
- Assembling an operational dataset

Gandrud, C. (2020). *Reproducible Research with R and RStudio*. CRC Press, Boca Raton, FL, 3rd edition, chs. 6–7.

Davis, J. and Michelitch, K. (2022). Introduction to field experiments: Thinking through identity and positionality. *PS: Political Science & Politics*, 55(4):735–40.

Herman, Biz; Panin, A. W. E. I. . B. G. . P. L. D. O. K. O. . A. H. M. . G. A. N. . T. Y. D. A. P. . and Owsley, N. (2022). Field experiments in the global south: Assessing risks, localizing benefits, and addressing positionality. *PS: Political Science & Politics*, 55(4):769–72.

European University Institute (2017). EUI policy on risk missions.

Third homework assignment available. This will require each team to write a first draft reporting its analyis, based on the PAP submitted earlier. The draft should be as complete as necessary to submit to a journal for review as a Registered Report; i.e. it should contain everything except the actual data and results. It may be based on simulated data generated through Declare Design or another data simulator.

Week Eight, February 28: Reporting your study

- Writing (and submitting) a Registered Report prior to seeing your data
- Receiving and processing your data
- Analyzing your data
- Writing internal reports
- Using graphics effectively

Gandrud, C. (2020). *Reproducible Research with R and RStudio*. CRC Press, Boca Raton, FL, 3rd edition, chs. 9–11.

Schwabish, J. (2014). An economist's guide to visualizing data. *Journal of Economic Perspectives*.

Editors (2019). What science looks like. *Nature Human Behavior*, 3:763.

Rudis, B., Ross, N., and Garnier, S. The viridis color palatte.

Recommended: Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded.* Oxford University Press, Oxford, chs. 6–9.

Week Nine, March 7: Writing up final results

- Writing up results
- Writing an effective abstract
- Adhering to scientific reporting standards

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science.* University of California Press, Oakland, chs. 8–9.

Aczel, B. e. (2019). A consensus-based transparency checklist. *Nature Human Behavior*, Published online 2 Dec 2019.

DeScioli, P. and Pinker, S. (2021). Piled modifiers, buried verbs, and other turgid prose in the *American Political Science Review. PS: Political Science & Politics*, 55(1):123–28.

Asiedu, E., Karlan, D., Lambon-Quayefio, M., and Udry, C. (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences*, 118(29):e2024570118.

Recommended: Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded.* Oxford University Press, Oxford, chs. 11–16.

Third homework assignment due

Week Ten, March 14: Publishing, publicizing, and documenting your work

- Writing your policy brief
- Preparing replication code
- Preparing a publicly accessible dataset
- Preparing a codebook

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How To Do Open Science.* University of California Press, Oakland, ch. 10.

Kopper, S., Sautmann, A., and Turitto, J. (2020). J-PAL guide to publishing research data.

National Academies of Sciences, Engineering, and Medicine (2017). *Communicating Science Effectively: A Research Agenda.* The National Academies Press, Washington, D.C, chs. 1–2 and 5.

Lupia, A. (2019). Communicating science in politicized environments. *Proceedings of the National Academy of Sciences*, 110(Supplement 3):14048–54.

Read at least five policy briefs from Evidence in Governance and Politics (EGAP) Policy Briefs and/or Center for Effective Global Action(CEGA) Policy Briefs

Recommended: Schimel, J. (2012). *Writing Science: How to Write Papers That Get Cited and Proposals That Get Funded.* Oxford University Press, Oxford, chs. 18–20.

Fourth homework assignment available. This will require each team to write up final results in a professional format, prepare your replication code, prepare a codebook, and write a policy brief to publicize your results to a lawy audience. The assignment is due at 15:00 on March 21.

This version: 2023/01/06 at 11:40:43