

CENTER FOR INTERNATIONAL SOCIAL SCIENCE RESEARCH

# PlainText Working Group

Timothy Elder

[timothyelder@uchicago.edu](mailto:timothyelder@uchicago.edu)

PhD Candidate, Sociology

Jenny Trinitapoli

Associate Professor, Sociology

Faculty Sponsor

[jennytrini@uchicago.edu](mailto:jennytrini@uchicago.edu)

WINTER AND SPRING 2023

Pick Hall RM118

## WORKSHOP DESCRIPTION AND OBJECTIVES

This working group is meant to give a short introduction to a set of plain text and open source tools for doing and writing up research. Many people have been exposed to these tools before, but the startup costs are non-negligible and we tend to fall back on the tools that are most ubiquitous rather than those that best suit our work. Unlike a traditional workshop at the University of Chicago, or a one time tutorial, we intend for this working group to allow interested students, post-docs and faculty to overcome the initial time investment and troubleshoot together to adopt a different and more appropriate set of tools for our research.

Some of these tools are applicable to anyone doing research and writing, in the humanities, social sciences or natural sciences, and some of the tools will be most helpful to scholars doing quantitative research. The tools that we are going to be working with represent a fundamentally different workflow and ethos than the Office toolkit and is distinguished in three ways: portability, reproducibility and elegance. More than just typesetting documents and using tools that are free, we will be thinking about how we do our research in a more orderly and, most importantly, *reproducible* way. We will be covering the following topics:

1. The Command Line and Project Organization
2. Document composition and typesetting with  $\text{\LaTeX}$ , pandoc, Markdown and Make
3. Creating Beautiful and Convincing figures in R with ggplot
4. Creating Cogent and Potent Tables
5. Bibliographic management with Zotero and Bibtex
6. Version Control and Collaboration with Git and GitHub

No previous programming or computing experience is expected, just literacy with a keyboard and mouse is required.

## THE COMMAND LINE AND PROJECT ORGANIZATION

In this meeting we will begin by discussing our experiences with collecting data and writing up our research, with a particular emphasis on the means by which we get things done. Important to note that we are not lingering on this process for its own sake but with a particular goal in mind: communicating insight in a transparent and reproducible way. We will then cover the basics of the Command Line and then how to organize a directory with all the material for your research. Though we will not directly address how you should organize the process of completing your project, the workaday techniques of interacting with your files and folders related to your project will help to keep both mind and research ordered and clear.

Meeting Date: February 3rd, 2023 — Noon to 2PM

## DOCUMENT COMPOSITION: $\text{\LaTeX}$ , pandoc, MARKDOWN AND MAKE

This is the heart of our working group; at a certain point after completing your data collection and doing your analyses, you are going to want to communicate what you have learned. At this point, the document you are composing becomes the heart and focus of your research. Academics, researchers and scientists of all stripes use documents to both get clear on how they are thinking about the things they are learning, and to communicate those things to their colleagues and general audiences. Taking very seriously how we both make an argument and share that argument will improve the fruits of our research projects. In this meeting we focus on the latter and will learn how to compose documents in Markdown and use pandoc and  $\text{\LaTeX}$  to typeset them into beautiful formats. What more, we will touch on using Make to compile your projects. This last feature will be particularly useful to quantitative scholars, who need to update tables and figures in their documents when the underlying data changes.

Meeting Date: February 17th, 2023 — Noon to 2PM

## CREATING BEAUTIFUL AND CONVINCING FIGURES IN R WITH ggplot

Pictures may not be able to say a thousand words but they are certainly an excellent way for quickly communicating complex relationships to your audience, particularly when you are using quantitative data. What more, when you have to present your research at a conference or a job talk, visualizations are an important tool for maintaining your audience's interest. Unfortunately, many scientific visualizations are clumsily made, and do a better job of obscuring an argument or relationships than explicating them.

In this meeting we will go through some of the most important features of visualizations, what to do and what not to do, and then how to actually create figures in R with the ggplot library. We will also cover diagramming, a helpful tool for everyone and discuss the Matcha Notebook software and the LaTeX library TikZ.

Meeting Date: March 3rd, 2023 — Noon to 2PM

## CREATING COGENT AND POTENT TABLES

Tables, like visualization, are an excellent tool for summarizing critical information from your data and emphasizes the power of your technically complicated data analyses. But just like scientific visualizations, tables are often times poorly constructed and do a better of job of obscuring your argument than clarifying it. Table construction abides by a different set of principles than visualizations and we will cover them in this workshop and share some good examples of what to do and not do.

Meeting Date: TBD, Spring Quarter

## BIBLIOGRAPHIC MANAGEMENT WITH ZOTERO AND BIBTEX

Mounting a new argument with the data you collected yourself and analyzed almost certainly will require that you reference past works, and keeping track of what you are referencing. Including references to the documents that have guided and informed your own research is critical to ensuring that you are not inadvertently committing fraud and for making your own process of argumentation clearer. This task is somewhat burdensome but fortunately there are several tools which help managing references and automatically generating bibliographies in your documents, with all the proper formatting, very very easy. That means you will not be endlessly googling what the appropriate formatting is for APA style citations, or ASA style, or MLA, or any other esoteric and obtuse academic convention no one ever taught you. We will be advocating the use of one bibliographic management tool, Zotero, but pretty much any of the major bibliography softwares will be compatible with the other tools we are using, as we will be relying upon the Bibtex format to store compiled lists of references.

Meeting Date: TBD, Spring Quarter

---

## VERSION CONTROL AND COLLABORATION WITH `git` AND GITHUB

One of the primary benefits of doing and writing your research in a plain text format is that whole collections of plain text files can be tracked with a software known as `Git`. This software allows you to track all the changes that occur to all the files in your project directory, as well as keeping backups in a remote location, and a complete log of all changes that have occurred to them. What more, it allows you to share your work with collaborators, track all the changes that they make to the documents and to arbitrate when there are conflicts between versions.

Meeting Date: TBD, Spring Quarter

## REQUIREMENTS

To participate in the working group it will be important that you have a laptop that is capable of accommodating the software we will be learning to use. The performance requirements for the software are not burdensome so pretty much any modern computer will be usable but you will need sufficient memory to install LaTeX and the other software. About 7GBs of memory is sufficient for everything. Guides for installing all the required software will be distributed in advance of the workshop and the expectation is that you will come to working group meetings ready to start using and troubleshooting the software.