SOC 4650/5650 Users Guide

Christopher Prener, Ph.D.

2016-10-20

Contents

Preface

This text is a companion document for **SOC 4650/5650 - Introduction to Geographic Information Sciences**. It is designed to help you be *successful* in this course. The idea behind a course Users Guide is to create a reference for many of the intangible, subtle or disparate skills and ideas that contribute to being a successful researcher. In creating a Users Guide, I draw inspiration from the work of Donald Knuth.[1] Knuth has discussed his experiences in designing new software languages, nothing that the developer of a new language

> …must not only be the implementer and the first large-scale user; the designer should also write the first user manual… If I had not participated fully in all these activities, literally hundreds of improvements would never have been made, because I would never have thought of them or perceived why they were important…

While there is nothing particularly new about what I am writing here, and I am certainly not developing a new language for computing, the goal of the **Users Guide** remains similar to Knuth's experience. However, by distilling some of key elements for making a successful transition to being a *professional developer* of knowledge rather than a *casual consumer*, I hope to both improve the course experience itself and also create an environment that fosters a successful learning experience for you.

If you read through the course objectives included in the syllabus, you will note that creating maps is only one of them. As much as this is a GIS course, it is a course in research methods. We are concerned not just with any research methods, but *high quality* research methods and the *process* of conducting research.

There is an important distinction to be made between research methods and methodology.[2] Research methods are the mental habits and technical practices that make you a successful researcher. Some of the skills and techniques that we will discuss this semester are not taught as often in graduate programs. Instead, they are often the products of "learning the hard way". These "habits of mind and habits of method" are broadly applicable across methodologies and disciplines.

[1] Donald Knuth is the developer of TeX, a computer typesetting system that is widely used today for scientific publishing in the form of LaTeX. He also established the concept of literatue programming, which forms the basis of some of the practices we will follow with Stata this semester.

[2] Methodology refers to the *reasons* for implementing a particular approach to research while method refers to the tools used in the research process (Clough and Nutbrown 2012:25).

Chapter 1

Getting Started

Before you begin the semester, there are a number of things that I recommend that you do to help set yourself up for success. Before you do *anything* else, you should read through the **Syllabus** and the **Reading List**. Make sure you have a good sense of what is *required* for the course. If you have questions, bring them to the first day of class!

1.1 Prep Your Computer

Before you do anything else for this course, make sure you get your computer ready for the work you are about to undertake:

1. Make sure your operating system is up-to-date. If you are able, I would also recommend upgrading your computer to the most recent release of its operating system that the computer can run.
2. We'll be sharing computer files throughout the semester, so you should ensure that you have functioning anti-virus software and that it is up-to-date.
3. You'll also need to download files, so you'll need to make sure you have some free space on your hard drive. If you have less than 10GB of free space, you should de-clutter!
4. Make sure you know how to access your computer's file management system.
    - On macOS, this means being comfortable with Finder.app.
    - On Windows, this means being comfortable with Windows Explorer.

This of course assumes that you own a computer. Owning a computer is not required for this course. All students who are enrolled in SOC 4650 or SOC 5650 will be given 24-hour swipe access (*just what you always wanted!*) to Morrissey Hall to facilitate access to lab computers.

1.2 Create Accounts

There are two major web services that we will use this semester, and you'll need to create accounts for both:

- **GitHub** - you can sign-up at GitHub.com. Once you've signed up, fill out your profile, set-up two-factor authentication, and let Chris know (via email) what your user name is. Once he has it, he can add you to the SOC 4650/5650 organization.
- **Slack** - you can ask Chris (via email) for an invitation to sign-up for our team. Once the sign-up process is complete, you can log-in by going to our team's Slack site. Fill out your profile, set-up two-factor authentication, and change your timezone.

1.3 Download and Install Software

There are a number of software applications that we will use this semester. Most of them are free, and I recommend downloading those free ones right away. All of these applications are available for macOS and Windows.

- **Atom** - Atom is a flexible, open-source text editor that is produced by GitHub. You can download it from Atom's website.
- **GitHub Desktop** - GitHub makes a desktop client that you can use to easily interact with repositories that are stored on the site. You can download it from GitHub's website after you sign-up for an account there. You'll need that account information to complete the desktop client's set-up process.
- **Slack** - Slack has a number of applications for desktop and mobile operating systems. I recommend downloading Slack on your personal computer, and optionally installing it on your mobile device as well. You can download their desktop applications from their website and the mobile applications from your App Store.

For Graduate Students *only*

If your computer meets the operating system requirements for ArcGIS and you think you'd benefit from having access to the software at home, let Chris know (via email).

If you are in the Public and Social Policy Ph.D. program and your computer meets the hardware and software requirements for Stata, you should consider purchasing it for yourself. I recommend purchasing a perpetual license for Stata/IC. This is the most cost-effective solution for typical students.

1.4 Buy Course Materials

Books

There are three required books for this course:

1. Brewer, Cynthia. 2015. *Designing Better Maps: A Guide for GIS Users.* Redlands, CA: ESRI Press. ISBN-13: 978-1589484405; List Price: $59.99; ebook versions available.
2. Gorr, Wilpen L. and Kristen S. Kurland. 2013. *GIS Tutorial 1: Basic Workbook.* 10.3.x edition. Redlands, CA: ESRI Press. ISBN-13: 978-1589484566; List Price: $79.99; ebook versions available.
3. Thomas, Christopher and Nancy Humenik-Sappington. 2009. *GIS for Decision Support and Public Policy Making.* Redlands, CA: ESRI Press. ISBN-13: 978-1589482319; List Price: $24.95.

There is one additional book that is optional:

- Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook.* College Station, TX: Stata Press. ISBN-13: 978-1597180764; List Price: $48.00.

Buying Mitchell (2010) is *highly* recommended for graduate students who will continue using Stata in the future and those who are concerned about the command-line interface. I recommend waiting for a week or two before purchasing this.

External Media

You will need a USB external storage device (either an external hard drive or a thumb-style drive) that has at least 20GB of storage capacity. This will be used for storing spatial data for this course.

1.5 Download Course Data

Mots of the course data is available for download via Dropbox in a single `.zip` file. If you want, you can let Chris know (via email) that you'd like to download these data before the beginning of the semester. Once you download them, extract the data from the `.zip` file and transfer them to your external storage device.

Chapter 2

Introduction to GitHub

Much of our interaction this semester outside of class will utilize GitHub.com (or just "GitHub"). GitHub is a web service that is a social network for programmers, developers, data scientists, researchers, and academics. It is also a tool for collaborating on projects, especially projects that involve writing code.

2.1 Git

GitHub is a web application that utilizes Git:

> Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Essentially, Git is a project-wide system for tracking changes to files. Think of it as Microsoft office's track changes feature on steroids - every change to every file in a directory (a "repository" or "repo" in Git-lingo) is tracked. You do no need to host files online to use Git. If you have a project saved locally (say, a doctoral thesis), you could utilize Git to version control that project without ever uploading it to the Internet.

For our purposes, this is just about all you need to know about Git. If you want to learn more, Git's 'About' page is a great place to start.

2.2 More Git-lingo

Beyond "repositories", there are a few additional terms that are specific to Git and that are helpful to know:

- **Clone**: Make an identical copy of a repository on your local hard drive.
- **Fork**:
- **Commit**: Approve any changes you have made to a repository.
- **Pull Request**:
- **Sync**: For cloned repositories, files that have been changed need to be pushed to GitHub.com after they are committed.

2.3 GitHub.com

GitHub is a web service that can host projects using Git's version tracking. It is widely used by programmers, software developers, data scientists, and academics to host and collaborate projects.

GitHub is an excellent way to backup files for a project since you can “sync” changes made to a repository up to GitHub’s servers. It is also an excellent way to collaborate on files with colleagues while also using Git’s version tracking. Repositories can be either public (like all of the repos for our seminar) or private, which means that only people who have been given access to can view the contents of the repo. Private repos require an upgraded account, which retails for $7/month.

Students can get access to GitHub’s paid services for free, however, by signing up for a free student account. This will give you access to private repositories for as long as you are a student.

2.4 GitHub Repositories

Users of GitHub.com adhere to a couple of norms with their repositories that are worth knowing about. Repositories cannot have spaces in their names (much like variables in Stata), so the naming conventions that we will discuss in relation to Stata this semester all apply to GitHub as well!

Public GitHub repositories also contain (typically) at least three core files:

1. A **license** file - since the data is out there for public consumption, it is important to think about how that data is licensed. The norm among GitHub users has been to use open source licenses, which let others edit and adapt your work. There are a range of licenses that are commonly used on GitHub.
2. A **README** file - this describes the purpose and content of the project.
3. A **.gitignore** file - this stops certain types of files from being swept up by GitHub when a user syncs their files with a server.

Another norm is to write using a markup language known as Markdown. Markup languages allow users to specify exactly how they want their text to appear when it is parsed and processed by special software. This is different from, say, Microsoft Word, which is known as a “what you see is what you get” or **WYSIWYG** editor, which uses a graphical interface for constructing documents.

2.5 Storing GitHub Repositories

When you clone your repositories, you will be prompted to save them on your computer. There are a number of ways in which this process can introduce sources for trouble down the road:

1. External media - storing data on devices like thumb drives or external hard drives can be a part of a backup workflow. However, I have seen issues where this has appeared to contribute to sync errors with GitHub Desktop, particularly on Windows.
2. Cloud storage services (Dropbox, Google Drive, etc.) - like external drives, these services can be a part of a backup workflow. However, like external drives, I have seen issues whee this has to contribute to sync errors with GitHub Desktop.

In order to avoid any issues, I suggest storing GitHub repositories on your computer’s hard drive and not a thumb drive or other external device. Make sure your are saving your files in a place not backed up to Dropbox or another cloud storage service.

2.6 GitHub Issues

GitHub has a powerful tool for interaction called Issues. These can be accessed by opening a repository and then clicking on the “Issues” tab. Issues can be “opened” by anyone with access to the repository. They allow for a conversation to occur in the form of messages posted within the Issue itself. Files can be attached to Issues, and the messages can contain Markdown formatting. Once the conversation is complete, issues can be marked as “closed”, which moves them into a secondary view on the website so that they are archived.

2.7 GitHub Desktop Application

GitHub Desktop is a tool that allows you to easily clone repositories hosted on GitHub, commit changes to them, and then sync those changes up to the website. You can also create new repositories, however this is not task you will have to do this semester. GitHub Desktop is not a fully functional desktop version of GitHub. For our purposes, it is important to note that the Desktop application will not let you easily identify when repositories have been updated by other users, view Wikis associated with repositories, or view Issues.

2.8 Learning More

GitHub has a resources page with links to websites that are great for helping you learn more about how Git and GitHub work!

Chapter 3

Methods

We describe our methods in this chapter.

Chapter 4

Applications

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

Chapter 5

Final Words

We have finished a nice book.

Bibliography