Partnership between RCC and SSRC (Social Science Research Center)
Social Science Computational Resources Series
Brooke Luetgert

Abstract:
This five session workshop series is designed to introduce students of the Social Sciences to computational tools and resources for data analysis and visualization. Using relevant macro economic and socio-political data, this introductory series will focus on the potential and limitations of various software packages. We will take a detailed tour through readily available data sources, organizing and cleaning data in OpenRefine, generating descriptive statistics in Stata, basic regression analysis in R, visualization alternatives in R as well as more advanced modeling resources in Python. Participants are warmly encouraged to practice what they have learned in a mini-Hackathon, designed to have researchers work with their own data, while receiving the technical support of devoted RCC staff specialized in data science. Workshops will be held bi-weekly from 2-4 PM and meet in Room 344 at 1155 60th Street.

Objective:
The purpose of this series is to expose participants to working with data in various software environments. While many tools will often get the job done, knowing the tips and tricks of multiple software alternatives helps the user to work more efficiently and confidently through data analysis and presentation. Our focus is on good data practice, documentation and reproducibility across software platforms.

Prerequisites:
Participants must bring a laptop on which they have administrative rights in a Windows, Linux or Mac OS. These workshops do not assume any previous knowledge of the software or scripting languages.

Date	Workshop Topic
February 4, 2-4 PM	**Find, Clean, Assemble**: Where can we find social science data? How can we clean and organize our indicators for easy input into our analysis with OpenRefine? How do we document our cleaning and assembly process in Markdown? Learn those first steps in good data management that will help take your analysis, documentation and dissemination to the next level.
February 18, 2-4 PM	**Deep Dive into Stata:** Starting with our clean and robust data, we begin with descriptive statistics, creating a codebook, entering basic regressions and creating plots of our data and resulting analysis. We explore the efficiency gains of canned regression analysis in a comfortable graphical user interface. Finally, we will look at exporting our results, tables and plots for inclusion in our written results.
March 3, 2-4 PM	**The R alternative:** As a language written for statistical analysis, the computational power of R is unquestionable. The flexibility of R comes with a steeper learning curve, but RStudio makes data analysis both convenient and transparent. The easy user interface and popular packages such as ggplot2 allow for the rapid production of beautiful graphics and quickly modifiable, publication quality plots and tables. Numerous packages and a supportive user community enhance the applicability of R to include textual analysis, network analysis, Bayesian statistics and so much more.
March 17, 2-4 PM	**Python for Data Science:** Python is a widely-used, general purpose scripting language with extensive data science libraries. Because of Python's wide application base, it may be the better choice for those interested in machine learning and data analysis tasks that need to be integrated with web applications. In this introduction, you will learn Python fundamentals such as reading and manipulating csv files, as well as how to handle data structures such as lists and Pandas Data Frames. You will learn about creating Python functions and customizing your own plots based on real data.
March 31, 9AM- 4 PM	**Hackathon:** This mini-hackathon is your opportunity to bite the bullet and join other data scientists for a full day of data assembly, evaluation and initial visualization. You will have access to expert support while you work with your data and advance your skill set. Participants of earlier series events are welcome to register and enjoy the opportunity to play all day and work through any data struggles together with RCC staff.