

## Workshop Descriptions and Instructions

### Computational Social Science Summit 2015

Northwestern University

Friday, May 15 2015

Workshops Organizer: Laura K Nelson, Northwestern University

If you have questions about the workshops email Laura at [laura.nelson@kellogg.northwestern.edu](mailto:laura.nelson@kellogg.northwestern.edu).

We find that many social science departments do not have faculty with the skills or resources to teach computational methods. To begin to fill this gap we are offering four different workshops that cover core tools and skills for computational research: the bash shell and Git, collecting and cleaning social media data using DiscoverText, network analysis using R, and computer-assisted text analysis using Python.

Below you'll find descriptions of the four workshops. Before you arrive at the conference you should read these descriptions and decide which two you will attend—one during the morning session and one during the afternoon session. **To get the most out of these workshops you should bring your own laptop with the appropriate software installed and datasets downloaded for your chosen workshops.** We will not provide laptops to workshop attendees. Specific instructions about necessary software and datasets for the different workshops follow the workshop descriptions below.

#### Morning Session: 10:30AM-1:30PM

*Choose one*

#### Workshop 1a: Computational Research Fundamentals: the bash shell and Git

Instructor: Jackie Milhans, Lead Computational Specialist, Northwestern University

Location: Jacobs, 166

#### Workshop Description:

In this hands-on workshop using materials from [Software Carpentry](#), we will cover basic concepts and tools, including using the command line, traversing through file directories, creating and manipulating files, task automation, and version control. Participants will be encouraged to help one another and to apply what they have learned to their own research problems. Knowing bash commands and version control is the foundation of efficient and effective computational research; all other tools build on this foundation. If you aren't yet familiar with these tools, this workshop is for you.

#### Pre-Workshop Instructions:

In preparation for the workshop on the Unix Shell and Version Control with Git, please be sure to have the following installed:

## Windows Users:

The bash shell – You can install Git Bash, which will install both Git and Bash for you.

<http://msysgit.github.io/>

Text editor – Nano <http://www.nano-editor.org/download.php> or Notepad++

<http://notepad-plus-plus.org/>

Git will be installed with Git Bash

## Mac OS X Users:

**For OS X 10.8 and higher**, install Git for Mac by downloading and [running the installer](#). After installing Git, there will not be anything in your /Applications folder, as Git is a command line program. **For older versions of OS X (10.5-10.7)** use the most recent available installer for your OS [available here](#). Use the Leopard installer for 10.5 and the Snow Leopard installer for 10.6-10.7.

## Linux Users:

If Git is not already available on your machine you can try to install it via your distributions's package manager. For Debian/Ubuntu run **sudo apt-get install git** and for Fedora run **sudo yum install git**.

-----

So that you can follow along, here are the links for the dataset we will be manipulating and lesson plans for the day:

### The Unix Shell

- [Click Here to Download Data](#)
- [Introducing the Shell](#)
- [Files and Directories](#)
- [Creating things](#)
- [Pipes and Filters](#)
- [Loops](#)
- [Reference...](#)

### Version Control with Git

- [Setting up Git](#)
- [Creating a repository](#)
- [Tracking changes](#)
- [Exploring history](#)
- [Ignoring things](#)
- [Collaborating](#)
- [Conflicts](#)
- [Reference...](#)

**Instructor Bio:**

Jackie Milhans, PhD is Lead Computational Specialist in NUIT Research Computing at Northwestern University. Her role involves consulting and training researchers at Northwestern in high-performance computing, programming, operating systems, and research data management. Her postdoctoral work at Los Alamos National Laboratory, Los Alamos, NM, concentrated on examining the effects of friction on microstructural evolution in polycrystalline ductile metals in extreme environments. Her research background also includes multi-scale modeling of mechanical and thermal material properties, homogenization of multi-phase materials, inverse materials design, and statistical continuum mechanics. She received her Ph.D. in Materials Science and Engineering at the Georgia Institute of Technology in Atlanta, GA and her B.S. in Materials Science and Engineering at Carnegie Mellon University in Pittsburgh, PA.

**Workshop 1b: Text Analytics for Social Data Using DiscoverText & Sifter**

Instructor: Dr. Stuart Shulman, Founder and CEO, Texifter

Location: Jacobs, G40

**Workshop Description:**

Attend this workshop to learn how to build custom machine classifiers for sifting social media data. All attendees can sign up for a free trial of [DiscoverText](#) (30-days) with immediate access to Facebook, Twitter, Tumblr, Disqus, WordPress and more. Bring your laptop to tap into day-forward full fire hose social data streams. Join the live beta test of [Sifter](#), a free tool for estimating the cost to license historical Twitter datasets. Learn how to construct social data fetch queries, search resulting archives, filter on metadata or other project attributes, count and set aside duplicates, cluster near-duplicates, crowd source the human coding (also called labeling or annotation), and build high quality word sense disambiguation engines for cleaning up messy social data streams. Use basic research measurement tools, developed over ten years of National Science Foundation funding, to improve human and machine performance over time. Learn to reach and substantiate inferences on a solid basic research model informed by 14 years of interdisciplinary scientific research into the text classification problem.

**Instructor Bio:**

Dr. Stuart W. Shulman is founder & CEO of Texifter, and a Research Associate Professor of Political Science at the University of Massachusetts Amherst. He was the founding Director of the Qualitative Data Analysis Program (QDAP) at the University of Pittsburgh and at UMass Amherst, and Associate Director of the National Center for Digital Government. Dr. Shulman is Editor Emeritus of the Journal of Information Technology & Politics, the official journal of Information Technology & Politics section of the American Political Science Association.

**Afternoon Session: 2:30PM - 5PM***Choose one***Workshop 2a: Social Network Analysis Using R**

Instructor: Lorien Jasny - National Socio-Environmental Synthesis Center, University of Maryland

Location: Jacobs, 166

**Workshop Description:**

This workshop session will serve as a basic introduction to the importation, manipulation, and descriptive analysis of social network data within the R/statnet platform. Topics covered will include: basics of R syntax; an overview of basic R functions and data types; importation of network data into R; network data manipulation; management of metadata for complex networks; visualization of network data; calculation of network descriptives (e.g., centrality scores, graph-level indices); and use of classical network analytic techniques (e.g., blockmodeling). No prior experience with R or statnet is assumed, but attendees should have familiarity with the basic concepts of descriptive network analysis.

**Pre-Workshop Instructions:**

In preparation for this workshop you should have [R](#), [RStudio](#), and the R packages statnet, sna, MASS, numderiv, and coda installed on your laptops. To install packages, from the R console run the command:

```
install.packages(c("statnet", "sna", "MASS", "numderiv", "coda"))
```

In RStudio click on the tab "Packages" and then "Install," which brings up a dialogue box that will allow you to search for, and install, packages.

**Instructor Bio:**

I am a computational social scientist focusing on questions of public involvement in environmental decision making. My research agenda focuses on two related themes – how the structure and dynamics of inter-organizational networks affect policy change, and how the structure and dynamics of belief networks affect behavioral change. Substantively, I study how people try to bring about societal change in response to political and environmental concerns. Methodologically, the need to grapple with these often complex phenomena requires the use and development of techniques for handling large, dynamic, and relational datasets.

**Workshop 2b: Natural Language Processing and Text Analysis in Python**

Instructor: Brian Lange - Data Scientist, Datascope Analytics

Location: Jacobs, G40

**Workshop Description:**

Dive into the world of Natural Language Processing with the power of Python and its many analysis libraries. We'll use NLTK, TextBlob, and more on real text data, learning about stemming, word counts, sentiment analysis, and ways to summarize bodies of text. We'll also talk a little about some of the more advanced techniques available today, like text extraction, topic modeling, and machine learning techniques.

**Pre-Workshop Instructions:**

Workshop attendees should have the following installed on their laptops:

- Python 2.7 – instructions [here](#)
- the ability to run IPython 3.0 notebooks (aka Jupyter notebooks) — instructions [here](#)
- the Python libraries NLTK and TextBlob installed

If anyone is brand new to Python or having trouble with installation, [Anaconda](#) is a distribution of Python that includes IPython Notebook and NLTK out of the box, among many other commonly used libraries. Some things will work slightly differently than they do with the instructor's setup, but he has found it to be an easy way to get most things you need installed in one shot.

**Instructor Bio:**

Brian Lange is a Data Scientist at Datascope Analytics, where he leads design process exercises and creates algorithms, web interfaces and data visualizations. He has contributed to projects for P&G, Thomson Reuters, Motorola and other well known companies and his work has been featured on FlowingData. While he's not nerding out about typography and machine learning techniques, he enjoys science and comedy podcasts, brewing beer, and listening to weird music.