DSS-02: Plots & Figures in `R`

Organizers:
Christina García, Ph.D., Spanish and Linguistics
Kelly G. Lovejoy, Ph.D., Spanish and Linguistics
Christopher G. Prener, Ph.D., Sociology

Fall, 2016

Contacts:
Christy: 1513 Morrissey; garciacm@slu.edu
Kelly: 1508 Morrissey; lovejoykg@slu.edu
Chris: 1918 Morrissey; prenercg@slu.edu

Seminar Meetings
Select Wednesdays
2:30pm to 3:30pm
Morrisey 3600 (GIS Lab)

Seminar Websites
GitHub Repositories: https://github.com/slu-data-science-seminar

---

Seminar Description

This is the second edition of the Data Science Seminar, a series we started to foster an open-source data analysis community in the social sciences and humanities here at SLU. This semester is focused on using `R` to produce plots and figures of your data. We'll pay particular attention to a well-known graphics package named `ggplot2`. Along the way, we'll also provide you with an overview of the nuts and bolts of `R`.

Using `R` for data analysis is becoming increasingly common because it is free, open-source, exceptionally flexible, and highly extensible - all features that set it apart from commercial statistical software. Learning `R` is something that we are excited about because it is where cutting edge analyses are happening in our respective disciplines.

In order to get the most from this seminar series, we recommend selecting an additional dataset (preferably one that you are already familiar with) to practice these techniques. Programming is a 'use it or lose it' skill-set - the more you practice and work with these techniques, the more comfortable you will feel using `R`.

Software

We will focus on using the software application `RStudio` (link), which is an integrated development environment for `R`. It offers a number of advantages over using `R`'s 'stock' programming environment.

It can be downloaded for free from its developers.[1] Using RStudio requires that R is already installed on your computer. If you are a Mac user, you will also need to install X11, a.k.a. XQuartz (link).

Resources

If you want a companion text for this seminar, we will be basing the seminar approach off of the following book:

1. Wickham, Hadley. 2015. *ggplot2: Elegant Graphics for Data Analysis.* 2nd edition. New York, NY: Springer.

Since we are focused on using R for a small number of specific tasks, there is much about R and RStudio that we will gloss over or skip entirely. SLU's library has a wide range of books available on using R. Of those, there are three in particular that provide introductory-level material and are available as e-books:

2. Adler, Joseph. 2012. *R in a Nutshell.* Sebastopol, CA: O'Reilly. (link)
3. Cotton, Richard. 2013. *Learning R.* Sebastopol, CA: O'Reilly. (link)
4. Van der Loo, Mark P. J. 2012. *Learning RStudio for R Statistical Computing.* Birmingham, UK: Packt Publishing. (link)

We will also being posting a number resources on Github, a website that is designed for sharing and collaborating on the development of computer code. GitHub will be used to post examples of R scripts, a log of each seminar session, and other replication files. You do not need a GitHub account to download seminar resources. If you want to contribute to the development of these resources, however, you will need a (free!) GitHub account.

One of the great things about R is the wide variety of resources that are available online. Among them are R Bootcamp (link), *Introduction to Probability and Statistics Using R* (link), Quick-R (link), and UCLA's Institute for Digital Research and Education (link). If you have specific questions, a quick Google search or a search of the forums at Stackoverflow (link) will often identify helpful materials. Finally, RStudio maintains a curated list of web resources for learning R (link). Working with R and RStudio typically requires a significant amount of trial and error, more so than with other statistical packages. The resources online are often helpful starting places but will require additional adaptation to be applied to the specific problem you are trying to solve.

Schedule

1. September 7th - Intro to R and Scripting
2. September 28th - Basic Graphics in R: plot and qplot
3. October 19th - ggplot2's Grammar
4. November 9th - Building a Plot with ggplot2
5. November 30th - Polishing a Plot with ggplot2

---

[1]Be aware that there is also a paid version of RStudio - you do not need to pay for this software.