Social Science Data and Model Visualization in R

ICPSR Summer Program
May 16-17, 2019
University of Houston
Houston, TX

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Contact Information</th>
</tr>
</thead>
</table>
| Boris Shor, PhD  
Assistant Professor  
Department of Political Science  
University of Houston | Office: Philip G. Hoffman Hall (PGH), room 389  
Email: boris@bshor.com  
Work Phone: 312-283-4599 |

Course Overview

One of R’s strengths is the flexibility, clarity, and sheer aesthetic beauty of the graphics it produces. That strength has been exponentially increased with Hadley Wickham’s ggplot2 package, part of the famous “tidyverse” ecosystem. Little wonder that sites like FiveThirtyEight have standardized on R and ggplot2.

This course will introduce social science students to modern methods of exploring and communicating data and models to a variety of audiences. The primary tools will be the R statistical programming language, RStudio as a development environment, and the tidyverse family of R packages (especially ggplot2) for visualization and exploratory data analysis. The object will be to enable students to feel confident in descriptively exploring their data, communicating that to scholarly and outside audiences, and communicating inferential results for presentation and paper writing purposes. Throughout, good workflow practices for scientific writing will be emphasized. A basic knowledge of R is required, which can be provided by the 3-day R course taught earlier in the week.

This course focuses on the “grammar of graphics” approach to visualization in R as exemplified in the ggplot2 package by Hadley Wickham, but does present some “classic” ways of plotting data in “base R.”

The goal of the course is to instill confidence in students’ own abilities to communicate about their data and models visually.

Schedule

Morning session: 9:00–12:00  
Lunch: 12:00–1:00  
Afternoon Session: 1:00–5:00
Textbook

We will be following Kieran Healy’s excellent *Data Visualization: A Practical Introduction*, which is available for free online from [http://socviz.co](http://socviz.co) or in book form from Amazon.

Outline

• Day 1 – We will cover the basics of visualization, including general principles and pitfalls. Then we will explore base R graphic briefly. Then, we will dive into the ggplot2 visualization package, includes its concepts of *geoms*, *facets*, and *aesthetics*. To explore our raw data, we will produce scatterplots, histograms, and density plots. We will discuss adjusting features of our plots, including labels and legends.

• Day 2 – We will show how to present map data. Then we will turn to plotting visual summaries of statistical models, including coefficient plots, interaction effects, and marginal effects plots. Finally, students can choose a special topic of their choice for us to work on together.