Data and Model Visualization in R
ICPSR Summer Program
May 27-29, 2020
University of Houston

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

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.

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

Outline

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. 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.
Schedule

All times are in Eastern Time Zone.

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00–10:30am</td>
<td>Morning session 1</td>
</tr>
<tr>
<td>10:30–11:00am</td>
<td>Morning break</td>
</tr>
<tr>
<td>11:00–12:30am</td>
<td>Morning session 2</td>
</tr>
<tr>
<td>12:30–1:30pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:30–3:00pm</td>
<td>Afternoon Session 1</td>
</tr>
<tr>
<td>3:00–3:30pm</td>
<td>Afternoon break</td>
</tr>
<tr>
<td>3:30–5:00pm</td>
<td>Afternoon Session 2</td>
</tr>
</tbody>
</table>

Textbook

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

Details

Part 1 – Software Setup and Introduction

1. Software installation check
   a. R 4.0.0
   b. RStudio
   c. Required packages
2. R and RStudio introduction
3. R Markdown introduction

Part 2 – Visualization Introduction

1. gapminder example data
2. Base vs ggplot2 graphics

Part 3 – Introduction to creating plots

1. Mapping and setting aesthetics
2. Building layer by layer
3. Scatterplots
4. Smoothers
Part 4 – Making the right choices

1. Introducing small multiples
2. Combining techniques
3. Introducing geoms transformations
4. Histograms and density plots

Part 5 – Graph tables, add labels, make notes

1. Dplyr pipelines
2. Labeling points
3. Labeling axes, titles

Part 6 – Visualizing Models

1. Using predict
2. Using tidy methods
3. Marginal effects

Part 7 – Mapping geographical data

1. US State example
2. Choropleths

Bio

Boris Shor is an Assistant Professor at the Department of Political Science at the University of Houston. Previously, he held faculty positions in the Department of Government at Georgetown University and the Irving B. Harris Graduate School of Public Policy Studies at the University of Chicago. In 2011-2013, he was a Robert Wood Johnson Foundation Scholar in Health Policy, located at the University of California, Berkeley. He earned his Ph.D. in political science from Columbia University. His research interests include legislative institutions, ideology, political parties, and polarization often but not exclusively at the state level. He has been in published in the American Political Science Review, the American Journal of Political Science, the Quarterly Journal of Political Science, Political Science Research and Methods, and Political Analysis. He has specific policy interests in health care and policy reform. His data set on state legislative ideology is located at https://americanlegislatures.com and his regular research web site is located at https://research.bshor.com. His Twitter account is .