AN INTRODUCTION TO R

PART 1

Joey Stanley
University of Georgia
joeystan@uga.edu @joey_stan
joeystanley.com/r

PREPARATION
install R at r-project.org
install R Studio at rstudio.com
"The bad news is whenever you’re learning a new tool, for a long time you’re going to suck. It's going to be very frustrating.

But, the good news is that that is typical, it’s something that happens to everyone, and it’s only temporary.

Unfortunately, there is no way to go from knowing nothing about a subject to knowing something about a subject and being an expert in it without going through a period of great frustration and suckiness."

- Hadley Wickham
GOALS FOR THE SERIES

1. Show what’s possible.
2. Point you in the right direction.
3. Equip you with some new skills.
4. Shorten that period of suckiness.
Iris k-means clustering

X Variable
Sepal.Length

Y Variable
Sepal.Width

Cluster count
3

Shiny
(FEBRUARY 2 & 9)
VISUALIZATIONS USING GGPILOT2
(FEBRUARY 16, 23 & MARCH 2)
Great NYT Interactive -- Now Reusable with rCharts

Disclaimer and Attribution
I claim absolutely no credit for this visualization, which I consider one of the most beautiful I have ever seen. All credit belongs to the original source. If anybody believes this to be not true, I will take it down immediately. I am explicitly assuming approval for this task due to the DataJournalism interview.

Another Favorite from NYT
I think we all know the data visualization team at NYT is truly amazing. Earlier this year in my post "Is Barack Obama's Climate Plan Adrenaline or Adrenaline?" I watched and measured the 2004 White House to work with a data through Twitter. Unfortunately I was not as creative enough to think of other data sets to plug into the visualization. When Scott Lee my friend sent

A Pandoc Markdown Article Starter and Template*
Steven V. Miller  Clemson University

This document provides an introduction to R Markdown, argues for its benefits, and presents a sample manuscript template intended for an academic audience. I include basic syntax to R Markdown and a minimal working example of how the analysis itself can be conducted within R with the knitr package.

Keywords: pandoc, r markdown, knitr

Introduction
Academic workflow, certainly in political science, is at a crossroads. The American Journal of Political Science (AJPS) announced a new initiative which authors who are tentatively accepted for publication at the journal must hand over the raw code and data that produced the results shown in the manuscript. The editorial team at AJPS then reproduces the code from the manuscript. Pending successful replication, the manuscript moves toward publication. The AJPS might be at the fore of this movement, and it could be the most aggressive among political science journals, but other journals in our field have signed the joint Data Access & Research Transparency (DART) initiative. This, at a bare minimum, requires uploading code from quantitatively-oriented published articles to in-house directories hosted by the journal or to services like Dataverse.

There are workflow implications to the LaCour controversy as well. Political science, for the foreseeable future, will struggle with the extent of the data fraud perpetuated by Michael LaCour in an article co-authored with Donald P. Green in Science, the general scientific journal of record in the United States. A failure to reproduce LaCour's results with different samples uncovered a comprehensive effort by LaCour to "fake" data that provided results to what we felt or believed to be true (i.e. "truthfulness"). However, fake data can have real consequences for both the researcher and those who want to learn from it and use it for various purposes. Even research done honestly may suffer the same fate if researchers are not diligent in their workflow.
THE TIDYVERSE
(MARCH 23 & 30)

- **Tidy**
  - `tidyr`
  - `dplyr`
  - `forcats`
  - `purrr`
  - `readr`
  - `lubridate`

- **Get data into R**
- **Rearrange**
  - Work with categorical variables
- **Work with dates**
- **Advanced loops and functions**
- **Make code easier to read and write**
SPECIAL TOPICS

(APRIL 6, 13, & 20)

Regression and mixed-effects modeling

Network Analysis

Working with Text

Mixed model: random intercept and slope

stringr
1. Learn about R.

2. Learn about RStudio.

INTRO TO R
R is an open source programming language for statistical computing.

- open source: people can contribute to it
  - User-submitted contributions usually called “packages” or “libraries”

- programming language
  - Python, C, Java, C++, C#, Javascript, PHP, Go, Swift, Perl, Ruby…

- statistical computing
  - not really designed for building software
ALTERNATIVES TO R

SPSS
- Common in the humanities.
- Proprietary software. Only on Windows.

SAS
- Common in the sciences.
- Proprietary software.

Stata
- Economics and epidemiology
- Proprietary software

MATLAB
- Common in Mathematics and engineering
- Proprietary software

JMP
- No coding required. Point-and-click.
- Proprietary software.

Python
- Actually, Python’s not bad…

Why R?
- It’s free
- all major operating systems
- widely used
- extensive documentation/help online
**R vs RStudio**

R is the programming language.  
- Comes standard on many computers.  
- Stand-alone  
- Not pretty.

RStudio is a pretty wrap-around environment  
- Not standard, but it is free.  
- Requires R to run.  
- Less not pretty.
RS**TUDIO**

- Scripts
- Environment and History
- Console
- Files, plots, packages, help, etc.
RStudio also has other capabilities that (as far as I know) don’t come with R.

Shiny (shiny.rstudio.org)
  Builds interactive webpages in R.
  Workshops February 2 and 9.

R Markdown (rmarkdown.rstudio.org)
  Build HTML, Word files, or PDFs directly in R.
  Workshop March 9.

Easier to write and maintain packages.
Installing R: https://www.r-project.org

Installing RStudio: https://www.rstudio.com
An Introduction to R by Venables, Smith, and the R Core Team (2017).

This R Cookbook site is great and has helped me a lot.

The tidyverse website is the launchpad for learning to use the tidyverse package.

The Springer’s Use R! series (50+ volumes for free)

Lynda.com, which is free for UGA students, has some great help for learning R.
TODAY’S GOALS

1. Learn about R.
2. Learn about RStudio.
Thanks!

<table>
<thead>
<tr>
<th>Date</th>
<th>Workshop Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/26</td>
<td>Intro to R Part 2</td>
</tr>
<tr>
<td>2/2</td>
<td>Building Interactive Webpages in R: Introduction to Shiny Part 1</td>
</tr>
<tr>
<td>2/9</td>
<td>Building Interactive Webpages in R: Introduction to Shiny Part 2</td>
</tr>
<tr>
<td>2/16</td>
<td>Visualizations I: Introduction to ggplot2</td>
</tr>
<tr>
<td>2/23</td>
<td>Visualizations II: Customizing plots in ggplot2</td>
</tr>
<tr>
<td>3/2</td>
<td>Visualizations III: Advanced topics in ggplot2</td>
</tr>
<tr>
<td>3/9</td>
<td>Communicating to your audience with R Markdown</td>
</tr>
<tr>
<td>3/23</td>
<td>Clean and tidy data: Tidyverse Part 1</td>
</tr>
<tr>
<td>3/30</td>
<td>Transform, reshape, and modify your data: Tidyverse Part 2</td>
</tr>
<tr>
<td>4/6</td>
<td>Special topics: Regression and mixed-effects modeling</td>
</tr>
<tr>
<td>4/13</td>
<td>Special topics: Network Analysis</td>
</tr>
<tr>
<td>4/20</td>
<td>Special topics: Working with text</td>
</tr>
</tbody>
</table>