



# 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](https://www.r-project.org)  
install R Studio at [rstudio.com](https://www.rstudio.com)



# WORDS FROM THE WISE...

---

"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

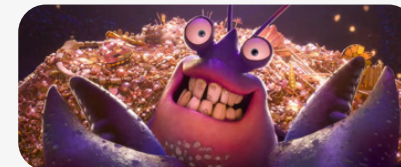
---

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.



# SHINY

(FEBRUARY 2 & 9)



## Iris k-means clustering

**X Variable**

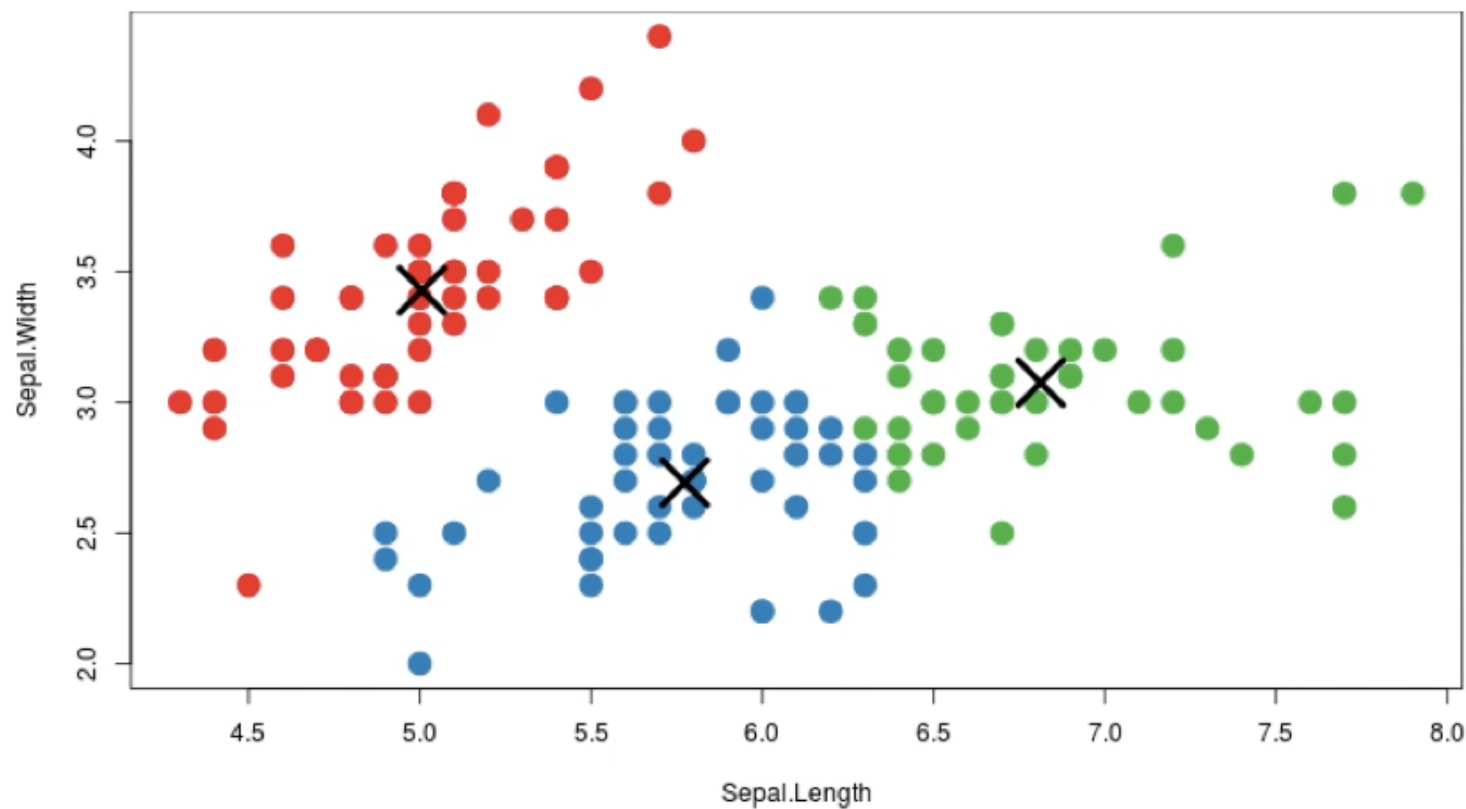
Sepal.Length

**Y Variable**

Sepal.Width

**Cluster count**

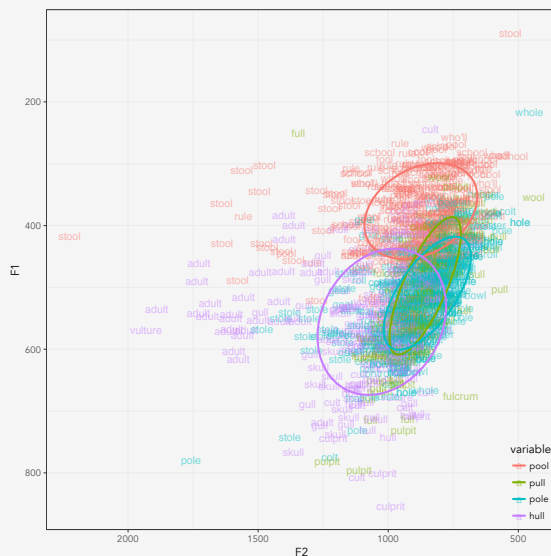
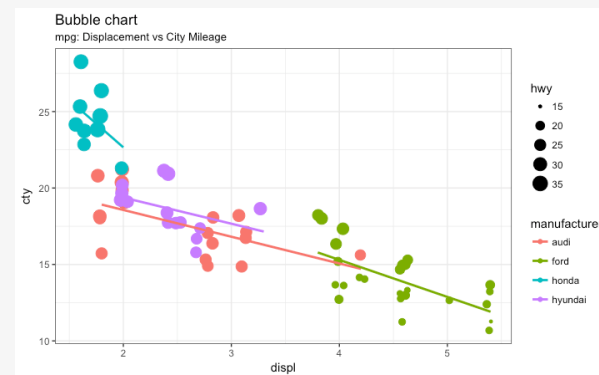
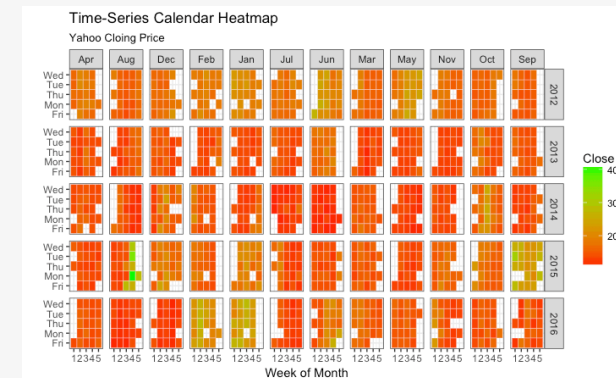
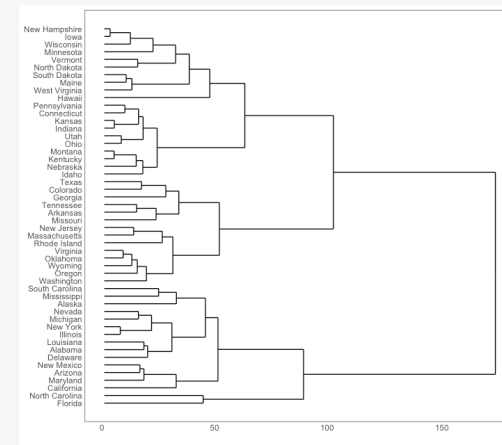
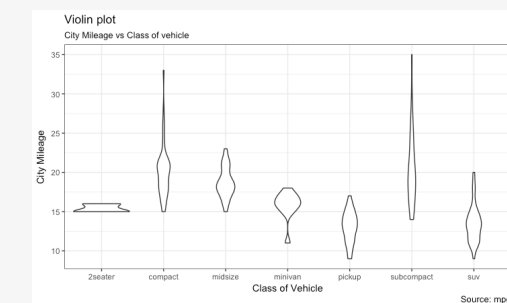
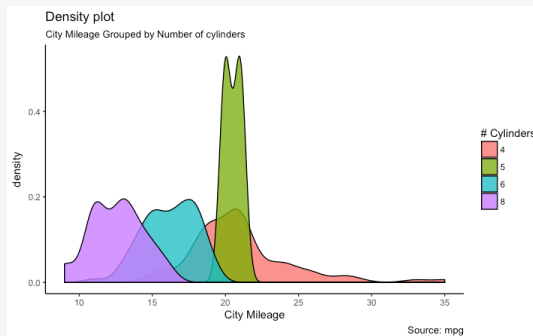
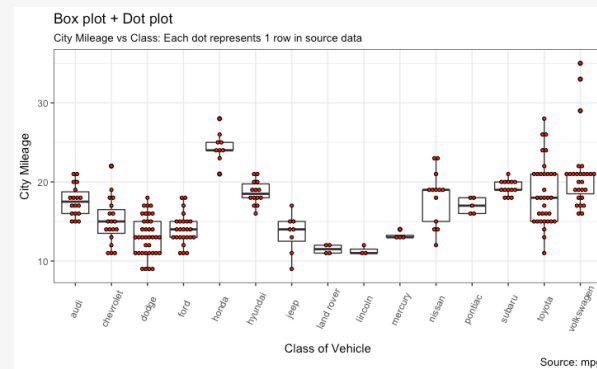
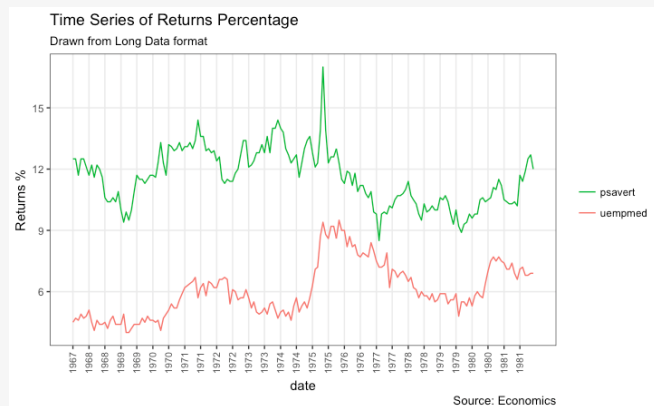
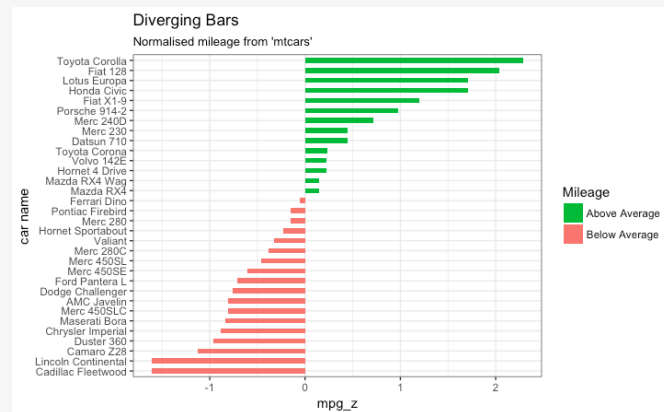
3





# VISUALIZATIONS USING GGPLOT2

(FEBRUARY 16, 23 & MARCH 2)





# R MARKDOWN

(MARCH 9)

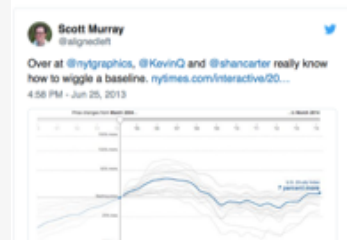
## Great NYT Interactive -- Now Reusable with rCharts

### Disclaimer and Attribution

I claim absolutely no credit for this visualization, which I consider one of the most best I have ever seen. All credit belongs to the [original source](#). If anybody believes this to be not fair use, I will take it down immediately. I am implicitly assuming approval for this fork due to the [data.stories interview](#).

### Another Favorite from NYT

I think we all know the data visualization team at NYT is simply amazing. Earlier this year in my post [d3.js - R with rCharts and slidify](#) I adapted and recreated the [512 Paths to the White House](#) to work with [rCharts](#). Unfortunately, I was not creative enough to think of other data sets to plug into the visualization. When Scott Murray tweeted,



## A Pandoc Markdown Article Starter and Template\*

Steven V. Miller *Clemson University*

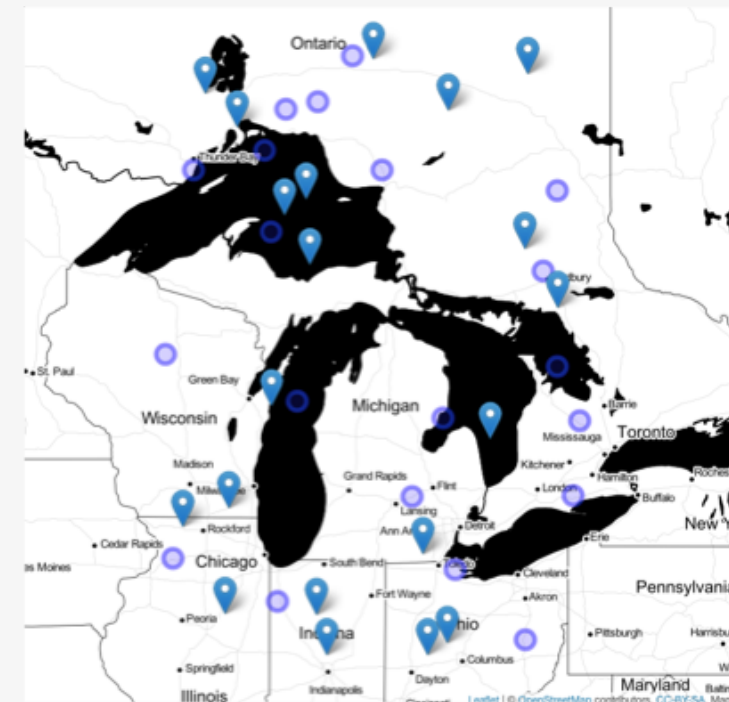
This document provides an introduction to R Markdown, argues for its benefits, and presents a sample manuscript template intended for an academic audience. I include basic syntax to R Markdown and a minimal working example of how the analysis itself can be conducted within R with the `knitr` package.

**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 (my words) “[show your work](#)” initiative in 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](#).

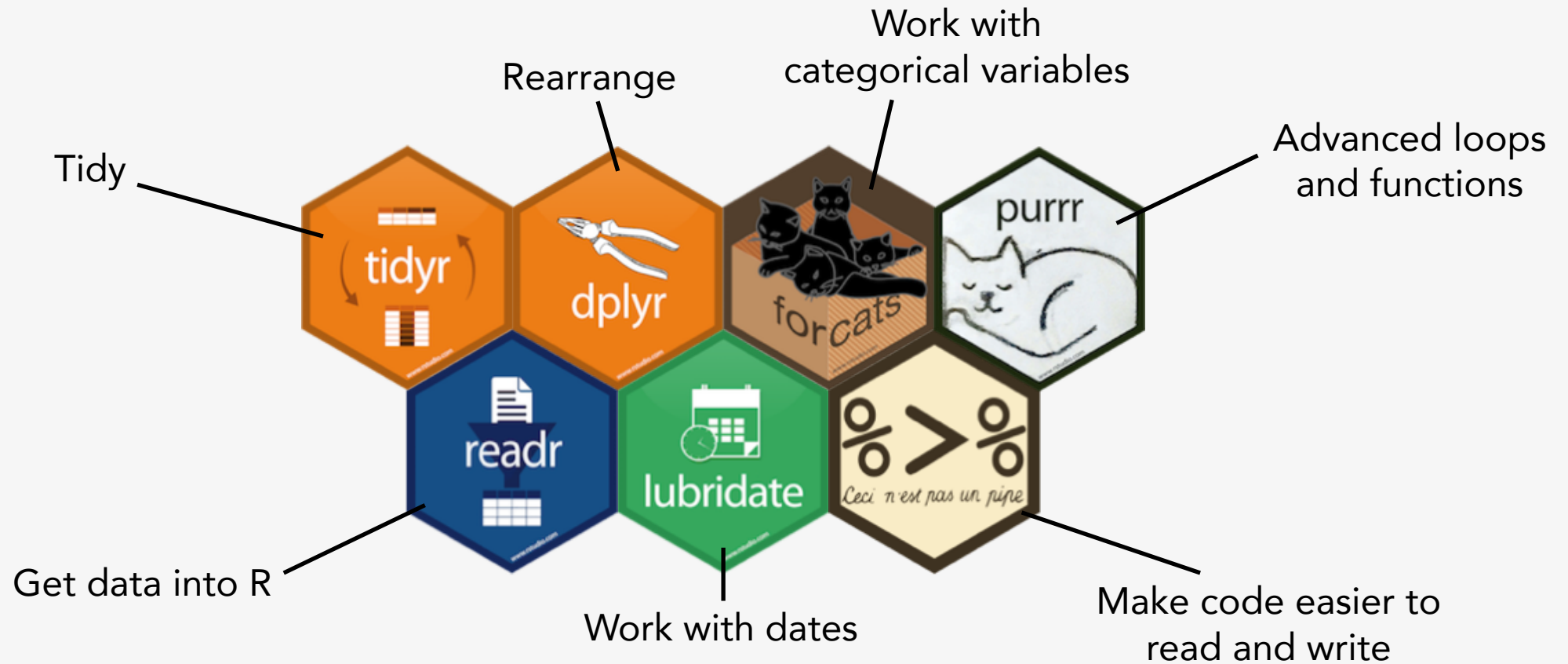
There are workflow implications to the LaCour controversy as well. Political science, for the foreseeable future, will struggle with the extent of [the data fraud perpetrated by Michael LaCour](#) in an article co-authored with Donald P. Green in *Science*, the general scientific journal of record in 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. “truthiness”). 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)



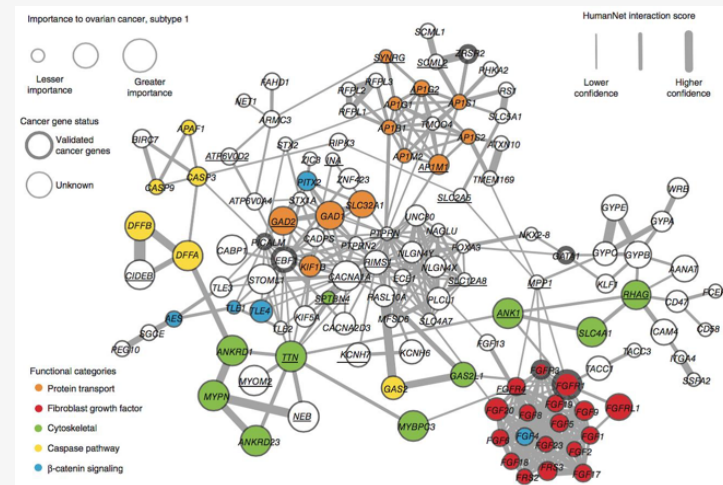
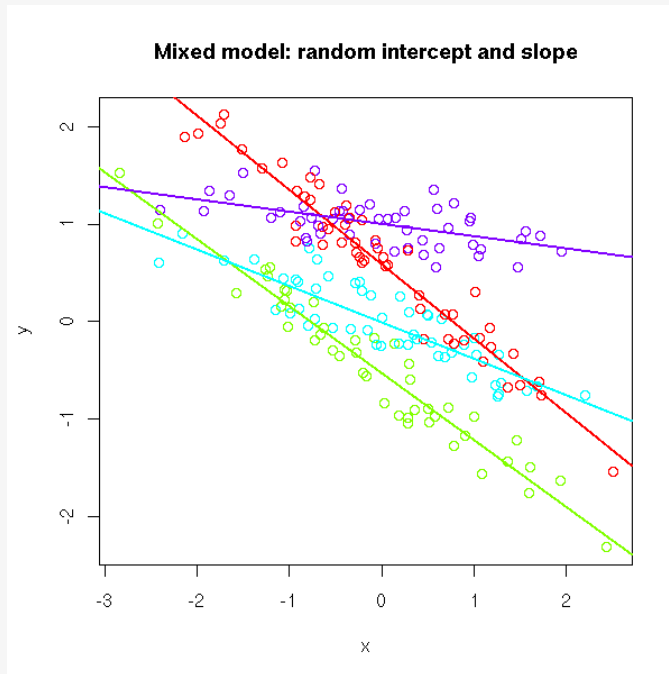
# SPECIAL TOPICS

(APRIL 6, 13, & 20)

## Regression and mixed-effects modeling

## Network Analysis

## Working with Text





# GOALS FOR TODAY

---

1. Learn about R.
2. Learn about RStudio.
3. Get R and RStudio running.

# INTRO TO R

---

# WHAT IS 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

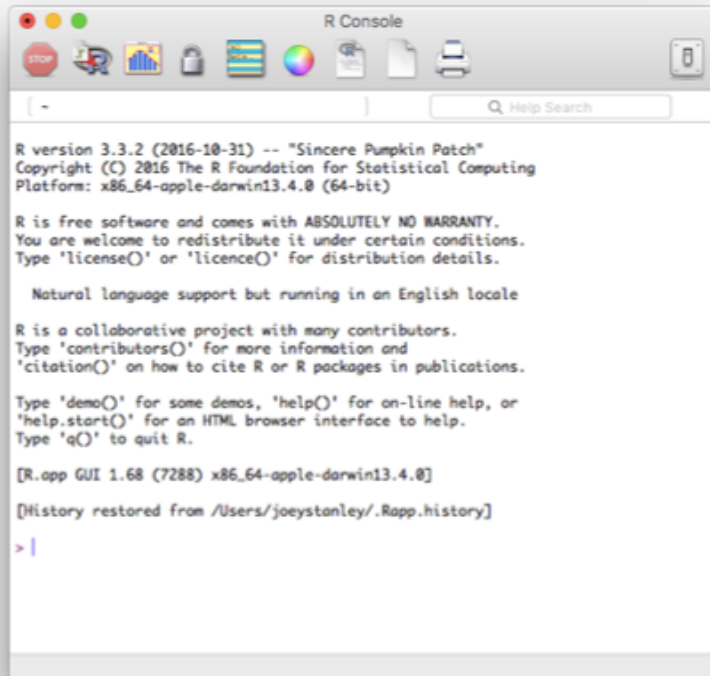
# R vs RStudio

R is the programming language.

Comes standard on many computers.

Stand-alone

Not pretty.



```
R Console
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

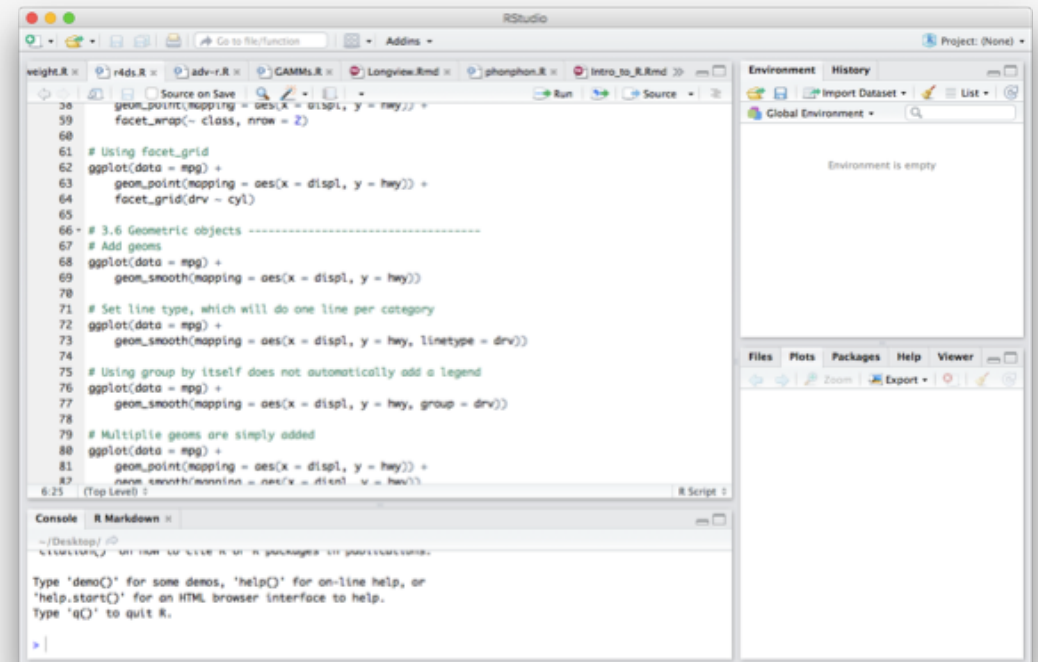
[R.app GUI 1.68 (7288) x86_64-apple-darwin13.4.0]
[History restored from /Users/joeystanley/.Rapp.history]
> |
```

RStudio is a pretty wrap-around environment

Not standard, but it is free.

Requires R to run.

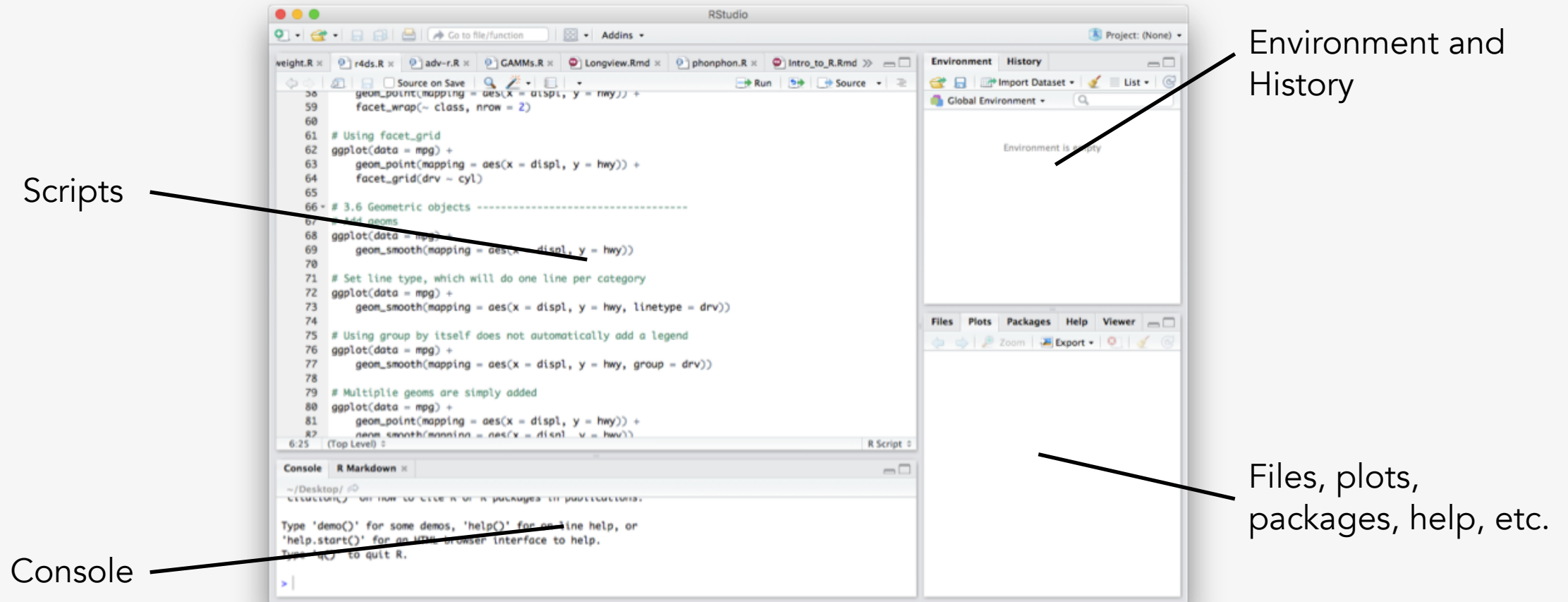
Less not pretty.



```
RStudio
veight.R x rds.R x adv-r.R x GAMMs.R x Longview.Rmd x phangon.R x Intro_to_R.Rmd
Source on Save | Go to File/Function | Adds +
60
61 # Using facet_grid
62 ggplot(data = mpg) +
63   geom_point(mapping = aes(x = displ, y = hwy)) +
64   facet_grid(drv ~ cyl)
65
66 # 3.6 Geometric objects -----
67 # Add geoms
68 ggplot(data = mpg) +
69   geom_smooth(mapping = aes(x = displ, y = hwy))
70
71 # Set line type, which will do one line per category
72 ggplot(data = mpg) +
73   geom_smooth(mapping = aes(x = displ, y = hwy, linetype = drv))
74
75 # Using group by itself does not automatically add a legend
76 ggplot(data = mpg) +
77   geom_smooth(mapping = aes(x = displ, y = hwy, group = drv))
78
79 # Multiple geoms are simply added
80 ggplot(data = mpg) +
81   geom_point(mapping = aes(x = displ, y = hwy)) +
82   geom_smooth(mapping = aes(x = displ, y = hwy, group = drv))
83
6-25 (Top Level) | R Script

Console | R Markdown
~/Desktop/...
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> |
```

# RSTUDIO





# RSTUDIO

---

RStudio also has other capabilities that (as far as I know) don't come with R.

Shiny ([shiny.rstudio.org](https://shiny.rstudio.org))

- Builds interactive webpages in R.

- Workshops February 2 and 9.

R Markdown ([rmarkdown.rstudio.org](https://rmarkdown.rstudio.org))

- Build HTML, Word files, or PDFs directly in R.

- Workshop March 9.

Easier to write and maintain packages.

# INSTALLATION

---

Installing R: <https://www.r-project.org>

Installing RStudio: <https://www.rstudio.com>

# HELP

---

[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.
3. Get R and RStudio running.

Thanks!

joeystanley.com/r  
@joey\_stan  
joeystan@uga.edu

## Don't miss our future workshops!

1/26 Intro to R Part 2

---

2/2 Building Interactive Webpages in R: Introduction to *Shiny* Part 1

---

2/9 Building Interactive Webpages in R: Introduction to *Shiny* Part 2

---

2/16 Visualizations I: Introduction to ggplot2

---

2/23 Visualizations II: Customizing plots in ggplot2

---

3/2 Visualizations III: Advanced topics in ggplot2

---

3/9 Communicating to your audience with R Markdown

---

3/23 Clean and tidy data: Tidyverse Part 1

---

3/30 Transform, reshape, and modify your data: Tidyverse Part 2

---

4/6 Special topics: Regression and mixed-effects modeling

---

4/13 Special topics: Network Analysis

---

4/20 Special topics: Working with text