

# GETTING STARTED

Jeff Goldsmith, PhD  
Department of Biostatistics



# Overall goals

- The short course is intended to introduce some common tools
- This lecture will get us started by focusing on:
  - RStudio
  - Some coding best practices
  - R Markdown
  - Project organization



# Why are we using RStudio?

- Makes life *much* easier for useRs (not a typo – people who use R are sometimes referred to as useRs...)
- The RStudio folks are also leading the development of a new analytic framework within R, and that work is integrated into RStudio





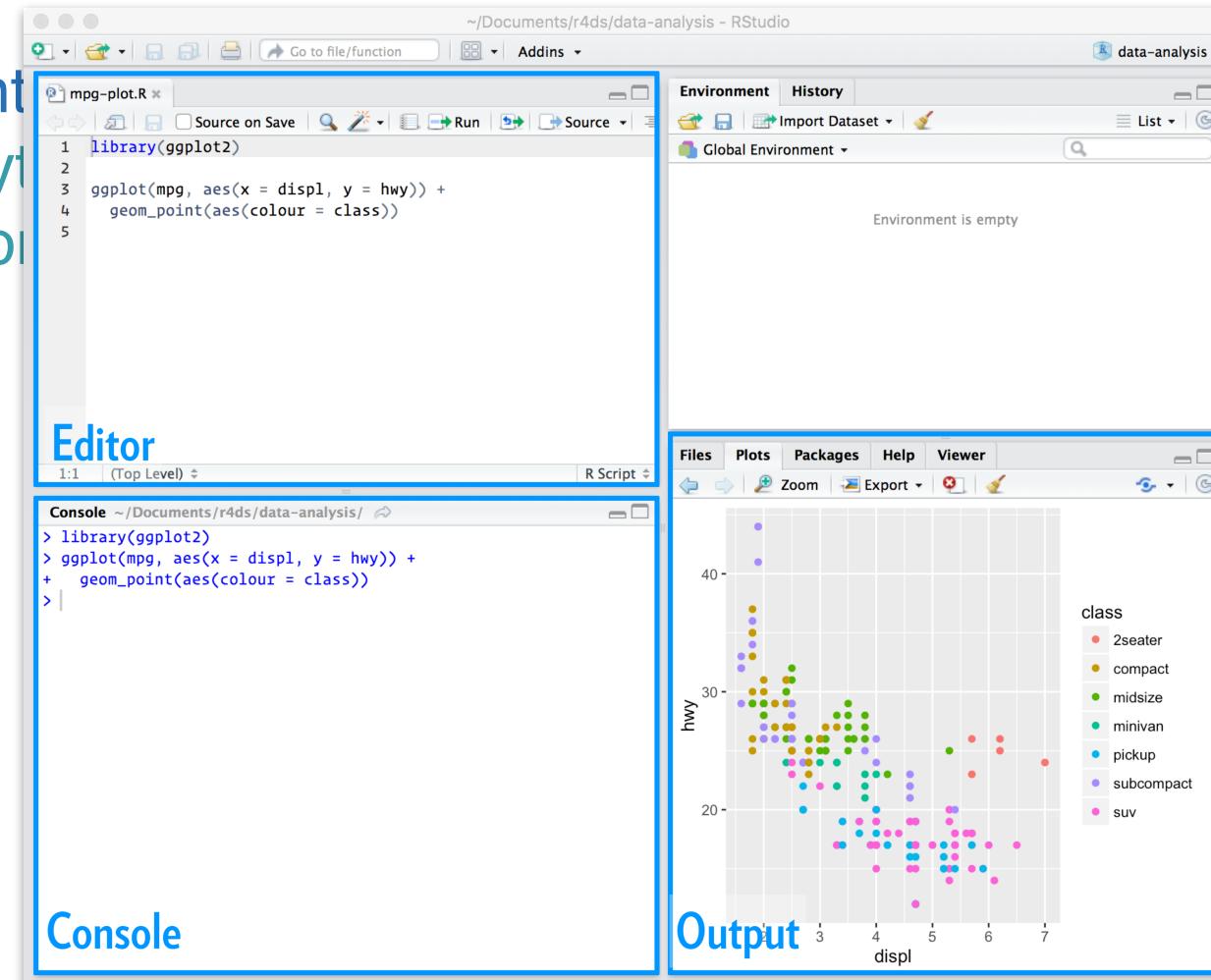
# Working in RStudio

- Rstudio is an Integrated Development Environment (IDE)
  - It's got everything you need to do data science in R
  - This IDE is one of the better reasons to use R ...



# Working in RStudio

- Rstudio is an IDE
  - It's got everything you need
  - This IDE is open source



R for Data Science



# Code

- Code is case sensitive
- There is no autocorrect
- Establish a variable naming convention
  - `this_is_snake_case`
  - `this.is.period.case`
  - `thisIsLowerCamelCase`
  - `ThisIsUpperCamelCase`
- Your names should match your regex skills
- Extensive documentation will save you headache



# Some perspective on code

- Treat your inputs (e.g. raw data) and code as “real”
  - Your results are created by input and code, and you can always reproduce your results from these if you need to
- Your code matters
  - It’s one of the most central ways you will communicate.
- Plan for mistakes
  - Write code that makes it easy to fix mistakes without breaking the rest of your analysis



# Some perspective on code

- Treat your inputs (e.g. raw data) and code as “real”
  - Your results are created by input and code, and you can always reproduce your results from these if you need to
- Your code matters
  - It's one of the  To: Jeff Goldsmith
- Plan for mistakes
  - Write code that handles errors in your analysis

Hey Jeff,

When I was running the analysis for the distributions I noticed that one of the Pittsburgh patients had an age listed as “-10”. I have now updated this to the correct age. Attached is the updated database.



# Text and code are important

- You spend a lot of your time communicating in writing
  - With collaborators, a general public, future you
  - About data cleaning, analyses, results
  - In formal reports, brief summaries, replies to questions



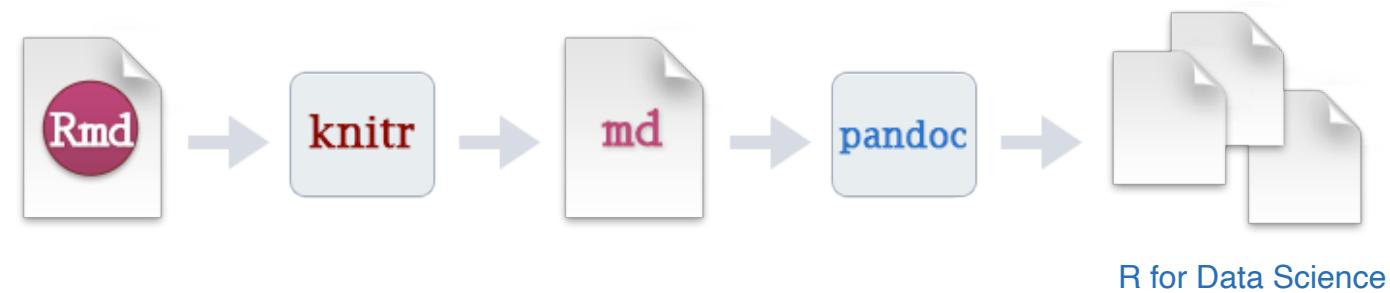
# Tools

- Code is necessary but not sufficient for clear communication
- Use tools that combine your code and text
- Greatly facilitates *reproducibility*, which is a big concept
  - In short, someone you don't know or work with should be able to reproduce each step of your analysis
  - As a part of this, they should understand why you did what you did
  - (Again, this someone is often future you)
- We'll use R Markdown to write reproducible reports



# R Markdown

- A “Markdown” language is a *lightweight syntax* that can be easily converted to another format (HTML, PDF, Word)
- R Markdown lets you combine formatted text with code chunks and the results of those chunks



- Having text and code in the same place, and having the combined output be user-friendly, is huge for your workflow



# R Markdown

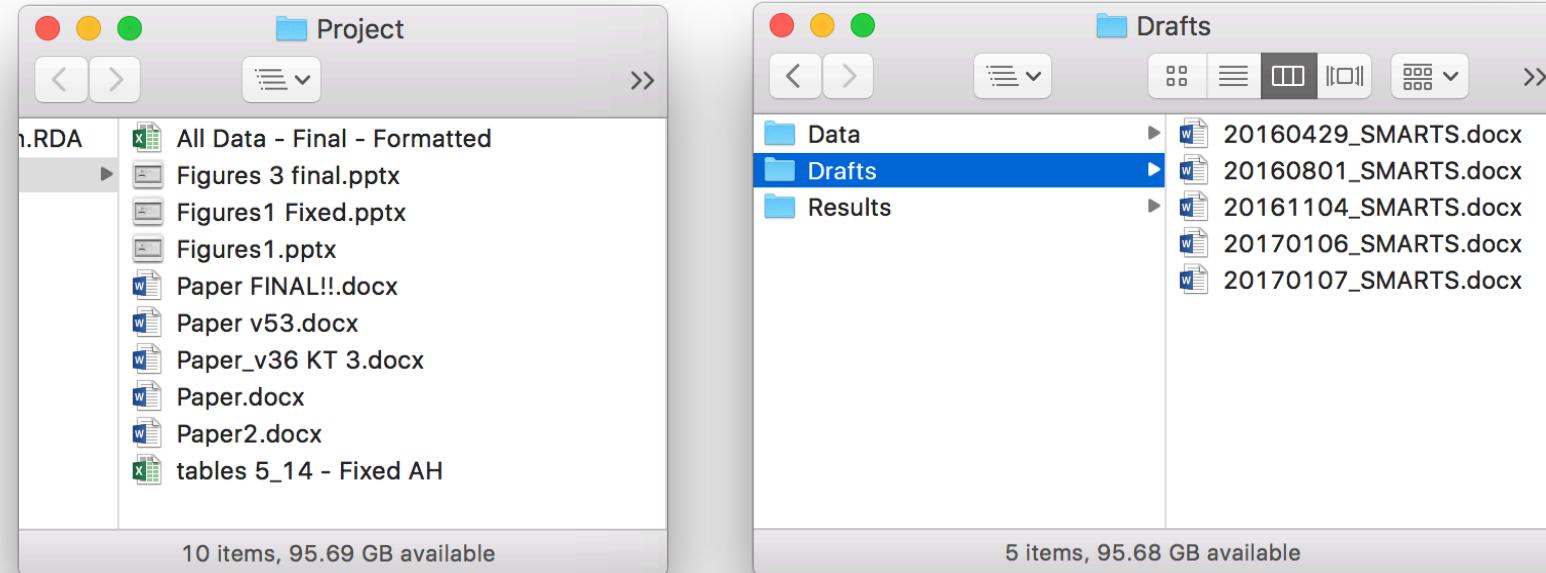
- A “Markdown” language is a *lightweight syntax* that can be easily converted to another format (HTML, PDF, Word)
- R Markdown lets you combine formatted text with code chunks and the results of those chunks



- Having text and code in the same place, and having the combined output be user-friendly, is huge for your workflow

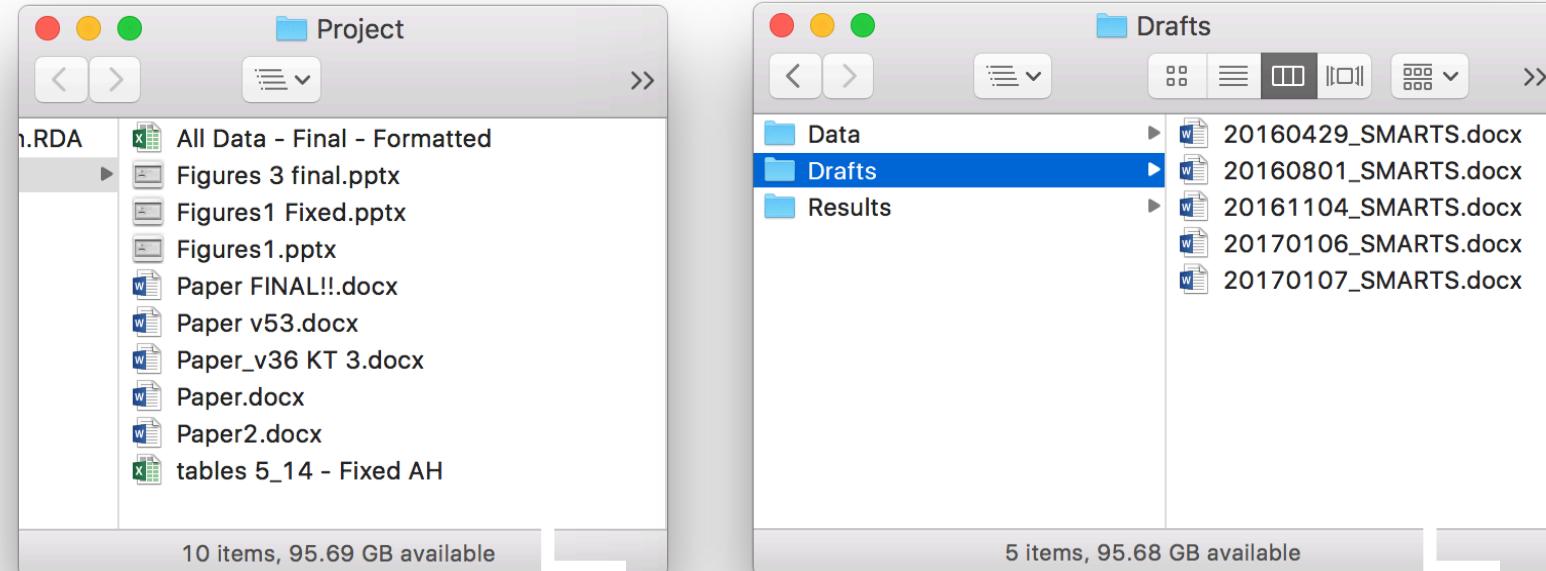


# Organizing files



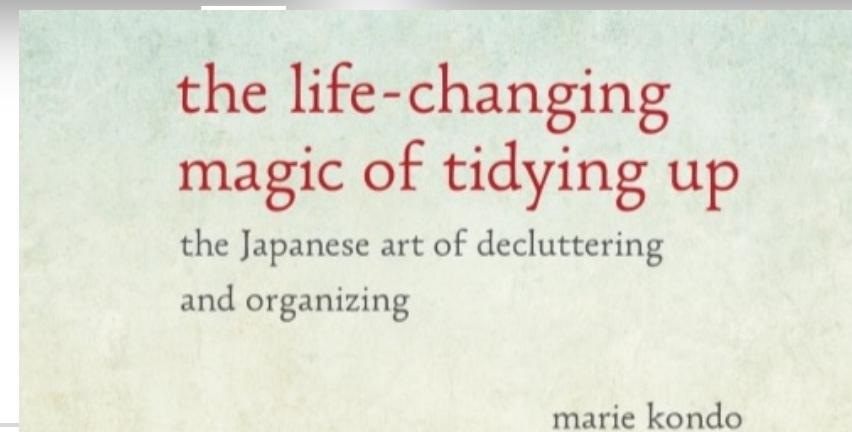
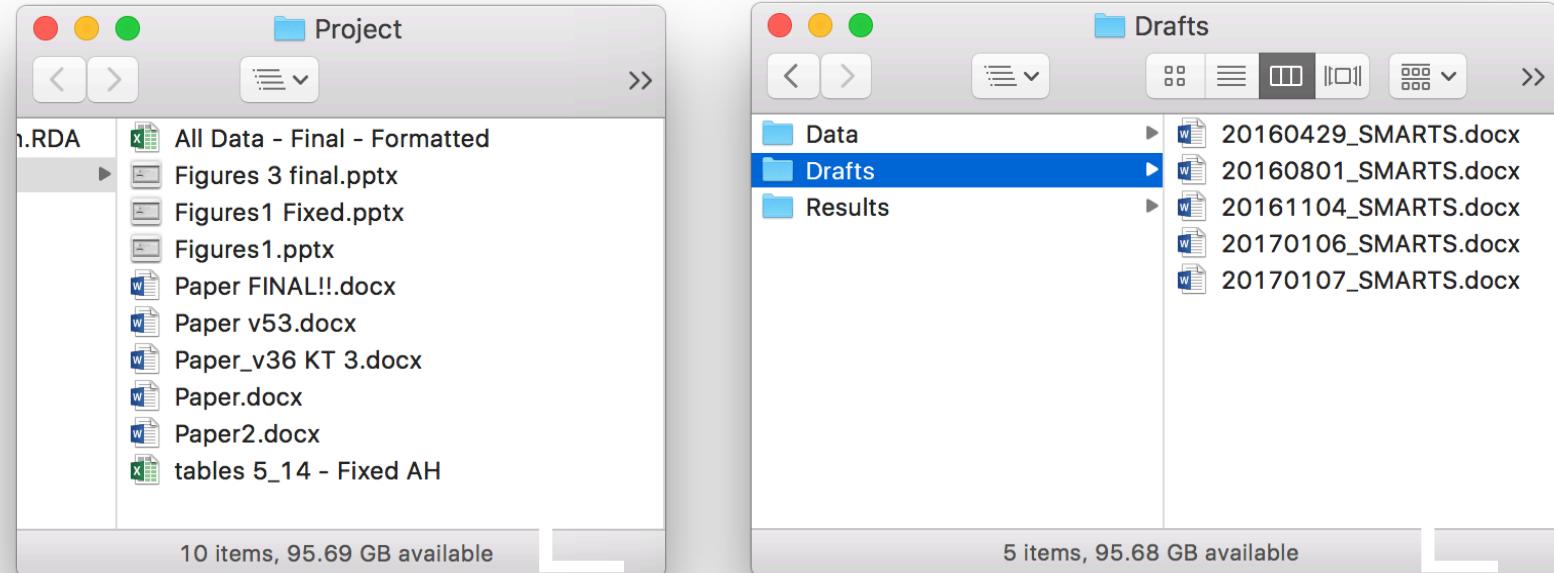


# Organizing files





# Organizing files





# Why organization matters

Being organized will *frequently* make your life easier

- “Your most frequent collaborator is you from six months ago, but you don’t reply to emails”<sup>1</sup>
- Eventually, someone other than you (or even future you) will need to reproduce your results
  - Be ready for that.

<sup>1</sup>. This version of the quote comes from Karl Broman, who traced it to a tweet: [http://bit.ly/motivate\\_git](http://bit.ly/motivate_git)

# Time to code!!

