# Course:

URL: <https://www.coursera.org/learn/reproducible-research/>

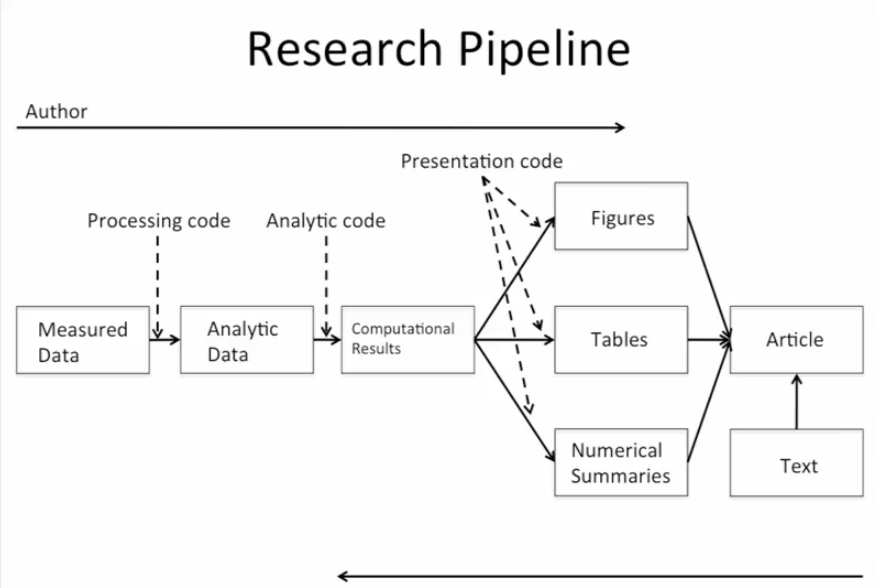
# Week 1:

Replication is very important – with different data, investigators, labs, etc.

However, it’s getting more difficult due to volumes of data, money, uniqueness etc.

Reproducibility lies between replication and doing nothing – make the data, the code (methods) available and let people run it themselves – essentially a validation of the data analysis.

* The author moves from left to right
* The reader from right to left



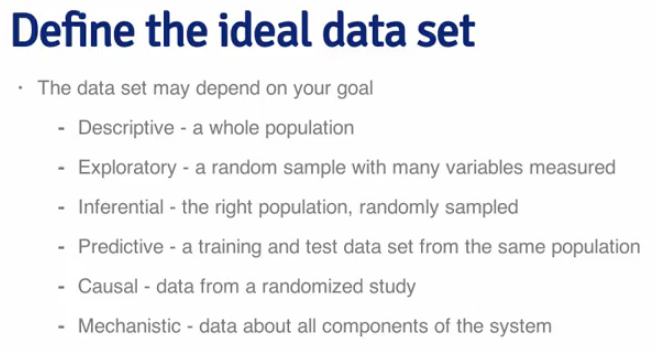
What’s required?

* The analytic data (not the raw)
* The code
* Documentation of both
* Standard means of distribution (accessibility)

Literate (statistical) programming allows you to weave natural language (human readable) into code (machine readable) – examples include Latex and Knitr.

Script Everything – write everything down

1. Define the question
   1. Narrow it down in order to simplify the research
   2. Don’t randomly apply statistical methods to data
2. Define ideal dataset



1. Determine what data you can access
2. Get the data
   1. Try to get the raw data
   2. Reference the source (url, time and date)
3. Clean data
4. Explore
   1. Look at summaries
   2. Check for missing data
   3. Create exploratory plots
   4. Perform exploratory analysis (e.g. clustering)
5. Statistical prediction / model
6. Interpret
   1. Use the right language – describes, correlates, leads to, predicts etc.
   2. Interpret coefficients
   3. Interpret measures of uncertainty
7. Challenge the results
8. Synthesize / write up results
9. Create reproducible code

Data Analysis Files

1. Data – raw and processed
2. Figures – exploratory and final
3. R code – raw, final, markdown
4. Text – readme, analysis

# Week 2:

Coding Standards in R

* Use text editor, save as text file
* Indent code (at least 4, ideally 8)
* Limit the width of code (80 columns)
* Limit the length of functions (1 basic activity)

Markdown

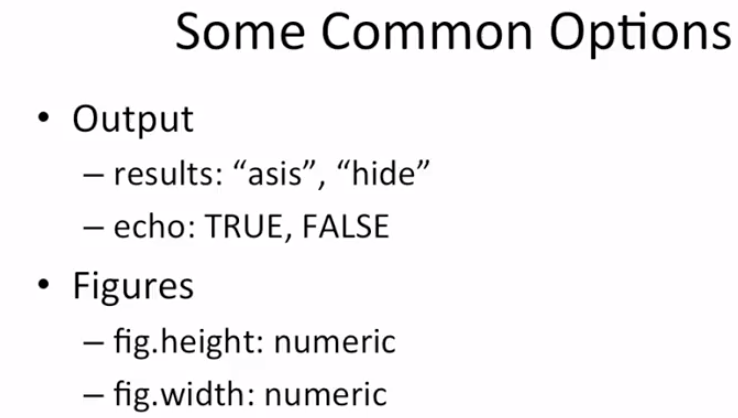
* Simplified mark-up language
* Text-to-HTML conversion tool
* Syntax – italics, bold, headings, unordered lists, ordered, links, newlines,
* <https://guides.github.com/features/mastering-markdown/>
* <https://daringfireball.net/projects/markdown/>

R Markdown

* Live R code in a document
  + R markdown converts to markdown via Knitr
  + Markdown package converts markdown to HTML
  + Slidify converts markdown to slides
* R markdown -> Markdown (Knitr) -> HTML (markdown) or -> Slides (Slidify)

Knitr

* Latex and HTML are “markup” languages, where you take regular text and add tags, annotation
* “Markdown” is simplified – just a few formatting elements
* All the elements are at: <http://goo.gl/MUt9i5> (<http://daringfireball.net/projects/markdown/basics>)
* Good for: manuals, regular reports, etc.
* Not good for: long research documents (single document), time-consuming computations, precise formatting
* Processing: .Rmd -> .md -> .html
* Hints:
  + Can set code to not dispay with echo=false
  + Results to not display with results=hide
* Common options:



* You can cache computations if they are particularly complex, long-running. It runs the code first time, but stored on the disk for the next time you run the computation