**Reproducible Research: Concepts and Ideas**

* Replication
  + - The ultimate standard for strengthening scientific evidence is replication of findings and conducting studies with different
      * Investigators
      * Data
      * Analytical methods
      * Laboratories
      * Instruments
* What’s wrong with replication?
  + - Some studies cannot be replicated
    - No money
    - Unique
  + 🡺 reproducible research: make analytic data and code available so that others may reproduce findings
* Bridge the gap: Reproducibility
* Why?
  + - More complex data
    - Existing databases can be merged into new “megadatabases”
    - Computing power is greatly increased, allowing more sophisticated analyses
    - For every field “x” there is a field “computational x”
* What do we need?
  + - Analytic data
    - Analytic code
    - Documentation of code and data
    - Standard means of distribution
* Who are the players?
  + - Authors
      * Want to be reproducible
      * Want tools for RR
    - Readers
      * Want to reproduce
      * Want tools for RR
* In reality
  + - Authors
      * Put everything on the web
    - Readers
      * Download everything and (try to) figure things out
* Literate (statistical) programming
  + - An article is a stream of texts and code
    - Analysis code is divided into text and code “chunks”
    - Each code chunks loads data and computes results

**Scripting your analysis**

* RULE: scripting everything

**Structure of a Data Analysis**

* Steps
  + - Define the question
    - Define the ideal dataset
    - Determine what data can be accessed
    - Obtain the data
    - Cleaning the data
    - Exploratory data analysis
    - Statistical prediction/modelling
    - Interpret results
    - Challenge results
    - Synthesize/write up results
    - Create reproducible code