Reproducible Research Overview

Jeffrey Leek

May 17, 2016

Reproducible Research Content

- Structure of a Data Analysis
- Organizing a Data Analysis
- Markdown
- LaTeX
- R Markdown
- Evidence-based data analysis
- RPubs

Steps in a data analysis

- Define the question
- Define the ideal data set
- Determine what data you can access
- Obtain the data
- Clean the data
- Exploratory data analysis
- Statistical prediction/modeling
- Interpret results
- Challenge results
- Synthesize/write up results
- Create reproducible code

Data analysis files

- Data
- Raw data
- Processed data
- Figures
- Exploratory figures
- Final figures
- R code
- Raw scripts
- Final scripts
- R Markdown files (optional)
- Text
- Readme files
- Text of analysis

Define the ideal data set

- The data set may depend on your goal
- Descriptive - a whole population
- Exploratory - a random sample with many variables measured
- Inferential - the right population, randomly sampled
- Predictive - a training and test data set from the same population
- Causal - data from a randomized study
- Mechanistic - data about all components of the system