

REPRODUCIBLE RESEARCH WITH R

STATISTICAL ANALYSIS WITH R USING RSTUDIO, GITHUB, KNITR AND SHINY

Maja Založnik



The Oxford Institute of
Population Ageing

OXFORD – 19th March 2015

OUTLINE

- Introduction
- Reproducibility
- RStudio
- GitHub
- knitr
- RPubS
- Shiny

- **Introduction**
- Reproducibility
- RStudio
- GitHub
- knitr
- RPubS
- Shiny

OUTLINE

- Introduction
- **Reproducibility**
- RStudio
- GitHub
- knitr
- RPubS
- Shiny

OUTLINE

- Introduction
- Reproducibility
- **RStudio**
- GitHub
- knitr
- RPubS
- Shiny

OUTLINE

- Introduction
- Reproducibility
- RStudio
- **GitHub**
- knitr
- RPubS
- Shiny

OUTLINE

- Introduction
- Reproducibility
- RStudio
- GitHub
- **knitr**
- **RPubs**
- Shiny

OUTLINE

- Introduction
- Reproducibility
- RStudio
- GitHub
- knitr
- RPubS
- **Shiny**

REPRODUCIBILITY OF RESEARCH

- **Reproducibility vs Replicability of research?**

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- “*The confirmation of results and conclusions from one study obtained independently in another*” (Jasny et al. 2011)

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- **Levels of Replication**

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- **Levels of Replication**
 1. Re-ask the question

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- **Levels of Replication**
 1. Re-ask the question
 2. Re-do the experiment

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- **Levels of Replication**
 1. Re-ask the question
 2. Re-do the experiment
 3. Re-analyse the data

REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- **Levels of Replication**
 1. Re-ask the question
 2. Re-do the experiment
 3. Re-analyse the data
 4. Reproduce the analysis

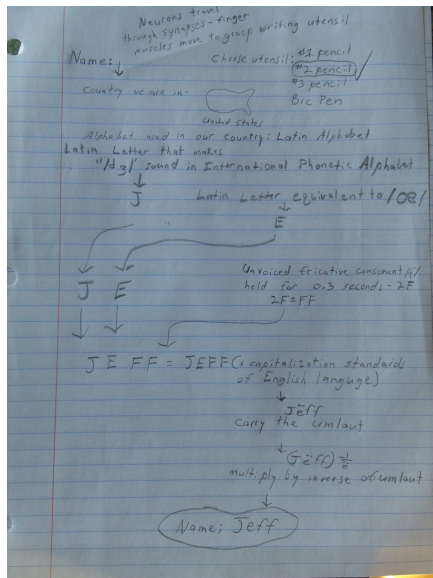
REPRODUCIBILITY OF RESEARCH

- Reproducibility vs Replicability of research?
- *“The confirmation of results and conclusions from one study obtained independently in another”* (Jasny et al. 2011)
- *“[T]he independent verification of prior findings”* (Santer et al. 2011)
- Levels of Replication
 1. Re-ask the question
 2. Re-do the experiment
 3. Re-analyse the data
 4. **Reproduce the analysis**

REPRODUCIBILITY: SHOW YOUR WORK!

imgur user TVsJeff:

"A math teacher took points off for not showing all of my work. The next homework assignment i turned in looked like this. It was 45 pages long."



- Probably the most popular IDE for R
- Launched February 2011
- January 2012 - Project system and Version control integration (git/SVN)
- May 2012 - knitr & R Markdown publishing tools added
- June 2012 - publish to RPubs integration
- December 2013 - Shiny integration
- October 2014 - direct publishing to shinyapps.io

VERSION CONTROL: GIT

THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.



GIT/GITHUB FOR REPRODUCIBLE RESEARCH

GIT/GITHub FOR REPRODUCIBLE RESEARCH

- Full documentation

GIT/GITHUB FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration

GIT/GITHUB FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration
- Dissemination

GIT/GITHUB FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration
- Dissemination
- Backup

GIT/GITHUB FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration
- Dissemination
- Backup
- [RStudio integration](#)

GIT/GitHub FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration
- Dissemination
- Backup
- [RStudio integration](#)
- [GitHub](#) - the Facebook of code

GIT/GITHUB FOR REPRODUCIBLE RESEARCH

- Full documentation
- Collaboration
- Dissemination
- Backup
- [RStudio integration](#)
- [GitHub](#) - the Facebook of code
- But [click here for five free private repos!](#)

LITERATE PROGRAMMING AND KNITR

“Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”

Donald Knuth (1984)

LITERATE PROGRAMMING AND KNITR

“Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”

Donald Knuth (1984)



LITERATE PROGRAMMING AND KNITR

- Human readable

LITERATE PROGRAMMING AND KNITR

- Human readable
 - Pure code: WHAT & HOW but not WHY
 - Pure text: WHAT & WHY but not HOW

LITERATE PROGRAMMING AND KNITR

- Human readable
 - Pure code: WHAT & HOW but not WHY
 - Pure text: WHAT & WHY but not HOW
- Consistent coding style e.g.:
 - [Google style guide](#)
 - [Hadley Wickham's style guide](#)

LITERATE PROGRAMMING AND KNITR

- Human readable
 - Pure code: WHAT & HOW but not WHY
 - Pure text: WHAT & WHY but not HOW
- Consistent coding style e.g.:
 - [Google style guide](#)
 - [Hadley Wickham's style guide](#)

LITERATE PROGRAMMING AND KNITR

- Human readable
 - Pure code: WHAT & HOW but not WHY
 - Pure text: WHAT & WHY but not HOW
- Consistent coding style e.g.:
 - [Google style guide](#)
 - [Hadley Wickham's style guide](#)
- Commenting

LITERATE PROGRAMMING AND KNITR

- Human readable
 - Pure code: WHAT & HOW but not WHY
 - Pure text: WHAT & WHY but not HOW
- Consistent coding style e.g.:
 - [Google style guide](#)
 - [Hadley Wickham's style guide](#)
- Commenting
- knitting

LIST OF HELPFUL LINKS AND FREE RESOURCES

- Christopher Gandrud's *Reproducible Research with R and RStudio*
- Coursera [Data Science Specialisation](#)
- [GitHub](#) & [academic discount link](#)
- [R markdown and knitr resources](#)
- [RPods](#)
- [Shiny tutorial](#)