

# REPRODUCIBLE RESEARCH WITH R

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

*Maja Založnik*



The Oxford Institute of  
Population Ageing

OXFORD – 19<sup>th</sup> November 2015

- Introduction
- Reproducibility
- RStudio
- Version control with git/GitHub
- Literate programming with knitr & R Markdown
- Dissemination with RPubS
- Interactivity with Shiny

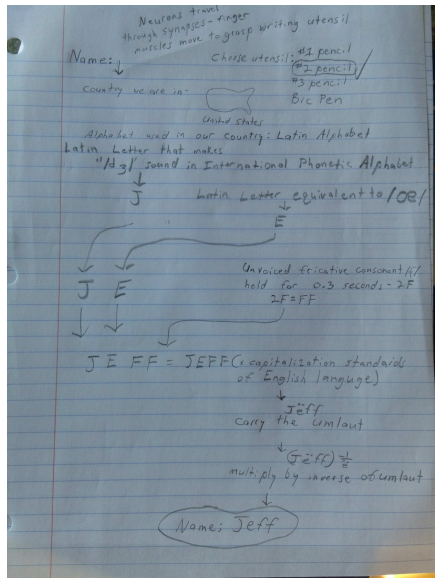
# REPRODUCIBILITY OF RESEARCH

- **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."*



# REPRODUCIBILITY: DON'T USE EXCEL®?

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

FIGURE : Reinhart and Rogoff's Excel Spreadsheet (Source: [gz.com](http://gz.com))

# GROWING POPULARITY OF R

- The proportion of analytic professionals using R continues to grow
  - Since 2010, R has been the #1 most-used data mining tool
- An increasing number of analytic professionals also select R as their primary tool
  - Since 2013, R has been #1 in primary tool rankings

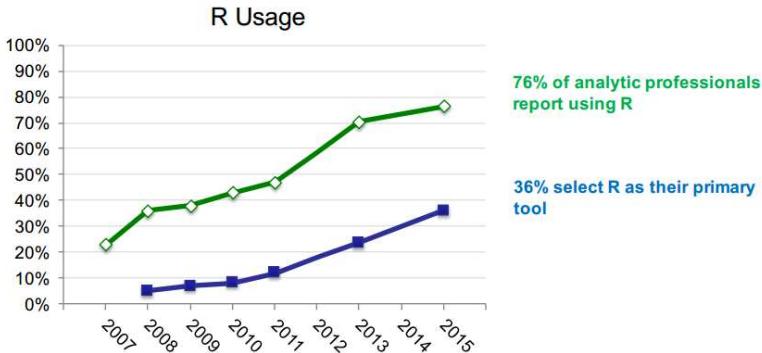
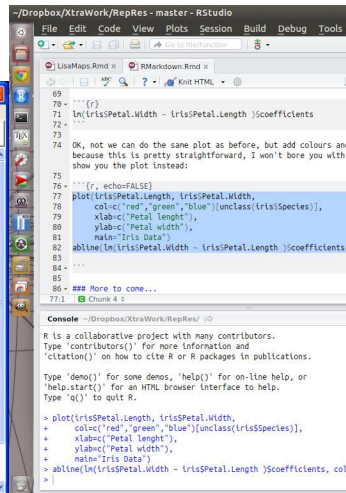
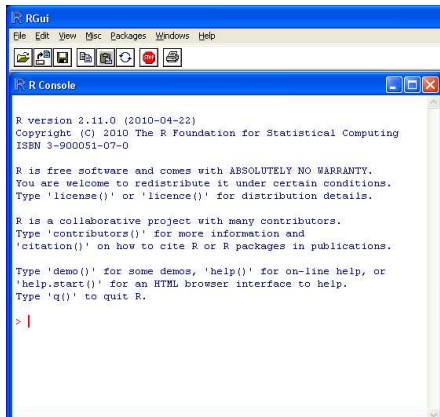


FIGURE : 2015 Data Science Survey Results - N=1,220 (Source: Karl Rexer 2015)

# R IDEs THEN AND NOW



- 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



FIGURE : [xkcd](#)

# 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

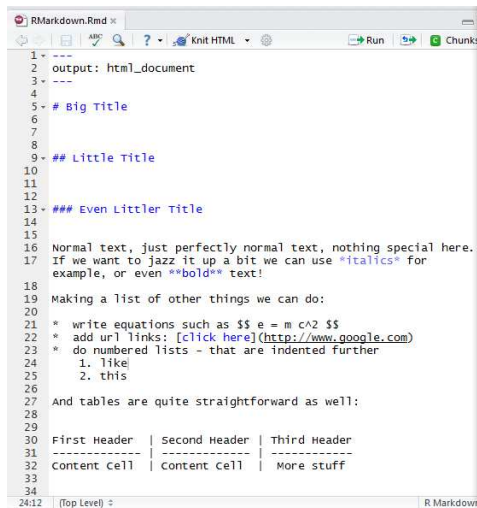
*“Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”*

Donald Knuth (1984)

# LITERATE PROGRAMMING AND KNITR

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

# KNITTING WITH MARKDOWN AND R



```
1 ---
2 output: html_document
3 ---
4
5 # Big Title
6
7
8
9 ## Little Title
10
11
12
13 ### Even Littler Title
14
15
16 Normal text, just perfectly normal text, nothing special here.
17 If we want to jazz it up a bit we can use italics for
18 example, or even bold text!
19
20 Making a list of other things we can do:
21
22 * write equations such as  $e = mc^2$ 
23 * add url links: [click here](http://www.google.com)
24 * do numbered lists - that are indented further
25   1. like
26   2. this
27
28 And tables are quite straightforward as well:
29
30 First Header | Second Header | Third Header
31 -----|-----|-----
32 content cell | Content Cell | More stuff
33
34
```

24:12 (Top Level) R Markdown

[Publish on RPubS](#)

# INTERACTIVE GRAPHICS WITH R

SHINY

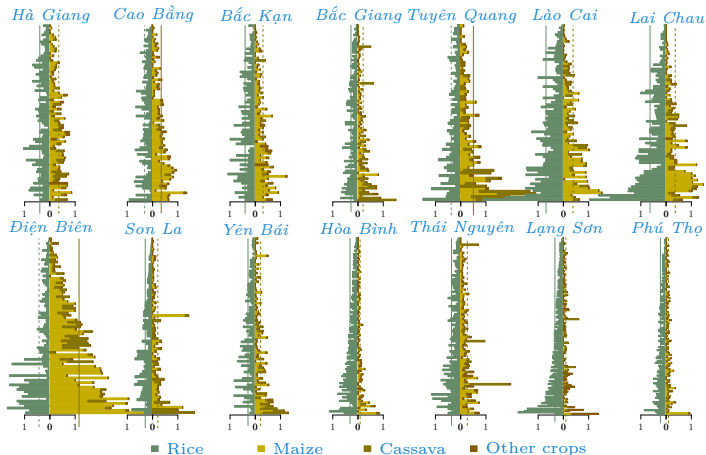
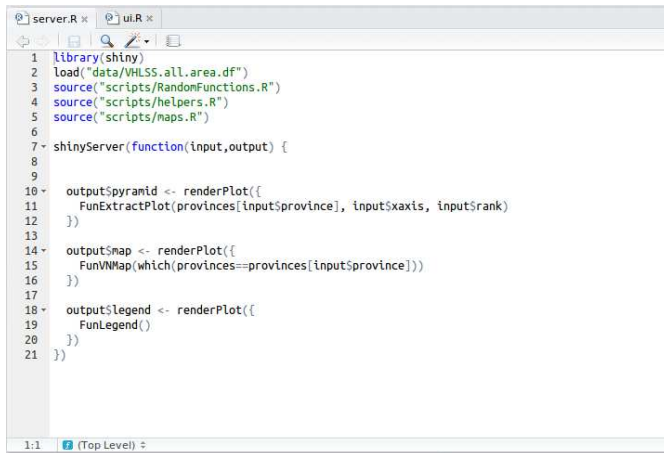


FIGURE : Areas of rice planted (left) and other crops (right) on individual farms for each province (in ha) (data: [VHLSS 2012](#))

# INTERACTIVE GRAPHICS WITH R

## SHINY - SERVER.R



```
1 library(shiny)
2 load("data/VHLSS.all.area.df")
3 source("scripts/RandomFunctions.R")
4 source("scripts/helpers.R")
5 source("scripts/maps.R")
6
7 shinyServer(function(input,output) {
8
9
10  output$pyramid <- renderPlot({
11    FunExtractPlot(provinces[input$province], input$xaxis, input$rank)
12  })
13
14  output$map <- renderPlot({
15    FunVMap(which(provinces==provinces[input$province]))
16  })
17
18  output$legend <- renderPlot({
19    FunLegend()
20  })
21 })
```

FIGURE : Content of `server.R` file for VHLSS shiny app



# INTERACTIVE GRAPHICS WITH R

SHINY - UI.R

# INTERACTIVE GRAPHICS WITH R

- [Iris example again](#)
- [3D rendering of mortality data](#)
- [“Hans Rosling” style chart](#)
- [showmeshiny.com](http://showmeshiny.com)

# 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](#)
- [RPresentations](#) & [Slidify](#)
- [Shiny](#) tutorial
- [R-bloggers](#)
- [Stackoverflow](#)
- [This presentation](#) on github (RepRes)