Reproducible

What? Why? 001-motivating-

etting Starte

LATEX

knitr

Examples

002-minimumworking-example

003-model-

004-figure

005-beamerpresentation

006-sensitivity-

analysis-one-

007-sensitivityanalysis-many-

008-large-

document

010-rmarkdow

Final Remarks

## Reproducible Research

An Introduction to knitr

Sahir Rai Bhatnagar<sup>1</sup>

May 28, 2014

<sup>&</sup>lt;sup>1</sup>https://github.com/sahirbhatnagar/knitr-tutorial

#### RR: Intro to knitr

006-sensitivity-

007-sensitivityanalysis-many-

# Acknowledgements

- Dr. Erica Moodie
- Maxime Turgeon (Windows)
- Kevin McGregor (Mac)
- Greg Voisin
- Don Knuth (TFX)
- Friedrich Leisch (Sweave)
- Yihui Xie (knitr)
- You





Reproducible

What? Why? 001-motivatir

etting Starte

RStudi

Evamples

Lxampics

working-e

outnut

004-figure

005-beamerpresentation 006-sensitivityanalysis-oneparameter

007-sensitivityanalysis-manyparameters

008-large-

009-rmarkdow 010-rmarkdow

Final Remarks

# Disclaimer #1

- Feel free to Ask questions
- Interrupt me often
- You don't need to raise your hand to speak

Reproducible Research

What? Why? 001-motivating

LAT<sub>E</sub>X RStudi

RStudi knitr

#### Examples

working-exam

output

004-figure

005-beam

presentatio

006-sensitivity-

parameter 007-sensitivity-

analysis-many-

008-large-

documents

010-rmarkdo

Final Remark

# Disclaimer #2











I don't work for, nor am I an author of any of these packages. I'm just a messenger.

RR: Intro to knitr

LAT<sub>E</sub>X

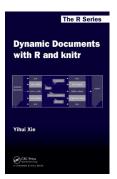
006-sensitivity-007-sensitivityanalysis-many-

# Disclaimer #3

• Material for this tutorial comes from many sources. For a complete list see:

https://github.com/sahirbhatnagar/knitr-tutorial

Alot of the content in these slides are based on these two books





Reproducible

What? Why? 001-motivatin

Getting Starte

LATEX RStud knitr

Example

002-minimumworking-exampl 003-modeloutput 004-figures 005-beamerpresentation 006-sensitivityanalysis-oneparameter 007-sensitivityaparameters 008-largedocuments

009-rmarkdowr 010-rmarkdowr presentation

Final Remarks

# Eat Your Own Dog Food

- These slides are reproducible
- Source code: https://github.com/sahirbhatnagar/knitrtutorial/tree/master/slides

#### Reproducible

What? Why? 001-motivating-

Getting Start

RStudi

#### Examples

002-minimumworking-exampl 003-model-

output

004-figure

presentation 006-sensitivity-

analysis-oneparameter

007-sensitivityanalysis-many-

008-largedocuments

009-rmarkdov

\_\_\_\_\_

# Main objective for today





# What is Science Anyway?

Reproducible

#### What?

Why? 001-motivating-

atting Storte

LATEX

LAT<sub>E</sub>X

kniti

#### Examples

002-mini

working-examp 003-model-

output

004-figure

005-beame

presentation

006-sensitivityanalysis-oneparameter

007-sensitivityanalysis-many-

008-large-

documents 009-rmarkdow

010-rmarkdown presentation

Final Remark

#### RR: Intro to knitr

# What?

006-sensitivity-007-sensitivity-

# What is Science Anyway?

### According to the American Physical Society:

Science is the systematic enterprise of gathering knowledge about the universe and organizing and condensing that knowledge into testable laws and theories. The success and credibility of science are anchored in the willingness of scientists to expose their ideas and results to independent testing and replication by other scientists



#### What?

Why? 001-motivatingexample

Getting Start

LAT<sub>E</sub>X RStudio

Exampl

#### Lxample

working-examp 003-model-

004-figure

005-beam

presentation 006-sensitivity-

parameter 007-sensitivityanalysis-many-

008-large

documents

009-rmarkdow 010-rmarkdow

Final Remarks

# RR: A Minimum Standard to Verify Scientific Findings



### What?

Why? 001-motivating example

LATEX RStudio

Example

002-minimu working-exa

003-modeloutput

004-figures 005-beamer

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-many-

008-largedocuments

009-rmarkdowi 010-rmarkdowi presentation

Final Remarks

# RR: A Minimum Standard to Verify Scientific Findings

### Reproducible Research (RR) in Computational Sciences

The data and the code used to make a finding are available and they are sufficient for an independent researcher to recreate the finding

Reproducible

What?

Why?

001-motivatingexample

etting Started

LATEX

knitr

Example

working-examp 003-model-

output

004-tigur

presentation 006-sensitivity-

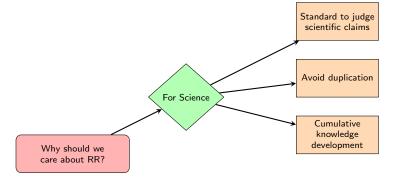
analysis-oneparameter 007-sensitivity-

paramete

documents

010-rmarkdov presentation

Final Remarks



Reproducible

What?

Why?

001-motivating-

Getting Starte

RStud knitr

Examples

002-minimumworking-exampl

003-model

004-figures

005-beamer-

006-sensitivity-

parameter

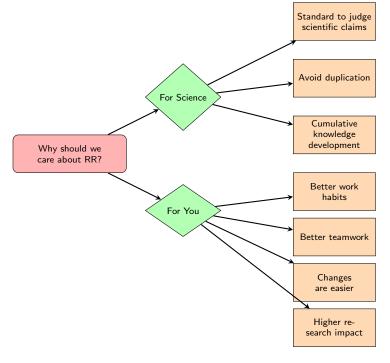
007-sensitivityanalysis-many-

000 1----

document

009-rmarkdowi 010-rmarkdowi

Final Remarks



Reproducible

What? Why?

001-motivatingexample

Getting Start

IATEX

RStudio

Example

002 min

003-mode

004-figures 005-beamer presentation

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-many-

008-large-

009-rmarkdowi

Final Remarks

# A Motivating Example

Demonstrate: 001-motivating-example

*Survey:* https://www.surveymonkey.com/s/CDVXW3C

What? Why? 001-motivatin example

#### Getting Started

LATEX RStudi

Evamo

002-minimumworking-examp 003-model-

output 004-figures 005-beamerpresentation

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-manyparameters 008-large-

009-rmarkdowi

inal Remarks

# Tools for Reproducible Research<sup>2</sup>

### Free and Open Source Software

- RStudio: Creating, managing, compiling documents
- LATEX: Markup language for typesetting a document
- R: Statistical analysis language
- knitr: Integrate ATEXand R code. Based on Prof. Friedrich Leisch's Sweave

<sup>&</sup>lt;sup>2</sup>http://onepager.togaware.com/

Reproducible

What? Why? 001-motiva

ting Starte

### **LATEX**

knit:

Examp

Lxampi

working-

output

004-figures

005-beamer

presentation 006-sensitivity-

analysis-oneparameter 007-sensitivity-

analysis-many-

008-large-

009-rmarkdow 010-rmarkdow

inal Remarks

## Comparison

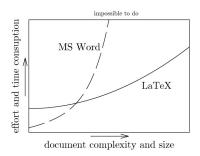


Figure 1 : Comparison

- LATEX has a greater learning curve
- Many tasks are very tedious or impossible (most cases) to do in MS Word or Libre Office

# The Philosophy behind LATEX

Research
What?
Why?
001-motivating-example

Getting Star

RStud knitr

Examples 002-mini working-e

working-exampl 003-modeloutput 004-figures 005-beamerpresentation 006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-manyparameters 008-largedocuments

documents 009-rmarkdown 010-rmarkdown presentation

inal Remarks



Figure 2: Adam Smith, author of *The Wealth of Nations* (1776), in which he conceptualizes the notion of the division of labour

### Division of Labour

Composition and logical structuring of text is the author's specific contribution to the production of a printed text. Matters such as the choice of the font family, should section headings be in bold face or small capitals? Should they be flush left or centered? Should the text be justified or not? Should the notes appear at the foot of the page or at the end? Should the text be set in one column or two? and so on, is the typesetter's business

Reproducible

What? Why? 001-motivating

Getting Star

LAT<sub>E</sub>X RStudi

RStudio knitr

Exampl

working-exampl 003-modeloutput 004-figures 005-beamerpresentation 006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-manyparameters 008-largedocuments

010-rmarkdown presentation

Final Remar

# The Genius Behind LATEX



Figure 3: The TEX project was started in 1978 by Donald Knuth (Stanford). He planned for 6 months, but it took him nearly 10 years to complete. Coined the term "Literate programming": mixture of code and text segments that are "human" readable. Recipient of the Turing Award (1974) and the Kyoto Prize (1996).



Research

What? Why?

001-motivatingexample

etting Start

IAT<sub>E</sub>X RStudio

knita

Exampl

Lxampic

002-minimumworking-examp

003-modeloutput

004-figure

005-beam

presentation 006-sensitivity-

analysis-oneparameter 007-sensitivity-

analysis-manyparameters

008-largedocuments

009-rmarkdov 010-rmarkdov presentation

Final Remarks

# Integrated Development Environment (IDE)

eproducible

What? Why? 001-motivating

Getting Star

RStudio

Example

working-exam

004-figures 005-beamerpresentation 006-sensitivity-

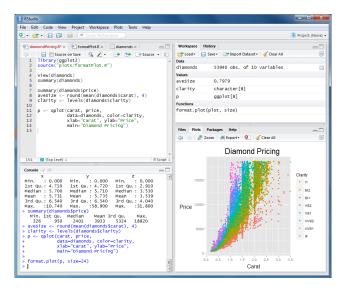
007-sensitivityanalysis-many-

008-largedocuments

010-rmarkdow presentation

inal Remarks

# Integrated Development Environment (IDE)



Demonstrate: Explore RStudio

#### Reproducible

What? Why? 001-motivating-

Getting Starte

IAT<sub>E</sub>X

knitr

Evample

Example

working-example

output

004-figure

presentation

006-sensitivity-

007-sensitivityanalysis-many-

008-large

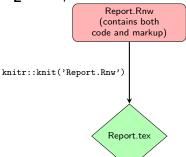
documents

009-rmarkdowi 010-rmarkdowi

Final Romarke

### What knitr does

### **LEX** example:



RR: Intro to knitr

001-motivating-

LAT<sub>E</sub>X

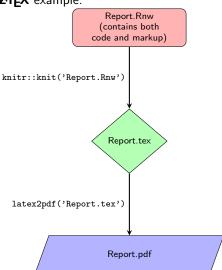
knitr

006-sensitivity-

007-sensitivityanalysis-many-

### What knitr does

### **LATEX** example:





What? Why? 001-motivating-

001-motivatin example

RStud

knitr

Evample

Example

working

003-mode output

004-figure

005-beamerpresentation 006-sensitivity-

parameter 007-sensitivityanalysis-many-

008-large-

009-rmarkdow 010-rmarkdow

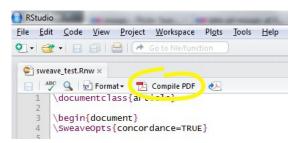
Final Remark

# Compiling a .Rnw document

The two steps on previous slide can be executed in one command:

knitr::knit2pdf()

or in RStudio:





knitr

006-sensitivity-

007-sensitivityanalysis-many-

# Incorporating R code

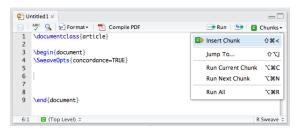
• Insert R code in a Code Chunk starting with

<< >>=

and ending with

@

#### In RStudio:



```
RR: Intro to knitr
```

What? Why? 001-motivating-

Getting Starter

knitr

knitr

#### Example

002-minim

003-model-

output

004-figures

005-beamerpresentation

analysis-oneparameter 007-sensitivity-

analysis-manyparameters

documents

009-rmarkdow 010-rmarkdow presentation

Final Remarks

# Example 1

```
<<example-code-chunk-name, echo=TRUE>>=
library(magrittr)
rnorm(50) %>% mean
@
```

#### produces

```
library(magrittr)
rnorm(50) %>% mean
```

```
## [1] 0.031
```

```
RR: Intro to
   knitr
```

001-motivating-

LAT<sub>E</sub>X

#### knitr

006-sensitivity-

007-sensitivityanalysis-many-

# Example 2

```
<<example-code-chunk-name2, echo=TRUE, tidy=TRUE>>=
for(i in 1:5){ (i+3) %>% print}
0
```

### produces

```
for (i in 1:5) {
    (i + 3) \% print
  [1] 8
```

```
RR: Intro to knitr
```

What? Why? 001-motivatingexample

IAT<sub>E</sub>X RStudi

knitr

#### Examples

#### Example.

working-example 003-model-

004-figures

005-beamerpresentation 006-sensitivity-

analysis-oneparameter

007-sensitivityanalysis-manyparameters

008-larg

009-rmarkdown 010-rmarkdown presentation

Final Remarks

# Example 2.2

```
<<example-code-chunk-name3, echo=FALSE>>=
for(i in 1:5){ (i+3) %>% print}
@
```

### produces

## [1] ·

## [1]

## [1] '

## [1] 8

```
RR: Intro to knitr
```

What? Why? 001-motivating

Getting Started

RStudio

Examples

002-min

working-example

103-model-

OUTPUT

005-beamerpresentation

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-many-

008-large-

documents 009-rmarkdo

010-rmarkdow presentation

Final Remarks

# Example 2.3

```
<<example-code-chunk-name4, echo=FALSE, eval=FALSE>>=
for(i in 1:5){ (i+3) %>% print}
@
```

#### produces

Demonstrate: Try it yourself

Reproducible Research

What? Why? 001-motivating example

LAT<sub>E</sub>X RStudi

knitr

E.....

002-minimu working-exa

003-modeloutput

004-figures 005-beame presentatio

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-manyparameters

documents 009-rmarkdow

Final Remarks

## R output within the text

- Include R output within the text
- We can do that with "S-expressions" using the command \Sexpr{...}

### Example:

The iris dataset has \Sexpr{nrow(iris)} rows and \Sexpr{ncol(iris)} columns

produces

The iris dataset has 150 rows and 5 columns

```
RR: Intro to
```

What? Why? 001-motivatingexample

example
Getting Starte

IAT<sub>E</sub>X RStud

#### knitr

#### Example

002-minimumworking-exam 003-model-

output 004-figures 005-beamer-

006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-many-

008-large-

009-rmarkdow 010-rmarkdow presentation

Final Remarks

# Include a Figure

```
<<fig.ex, fig.cap='Linear Regression',fig.height=3,fig.width=3>>=
plot(mtcars[ , c('disp','mpg')])
lm(mpg ~ disp , data = mtcars) %>%
abline(lwd=2)
@
```

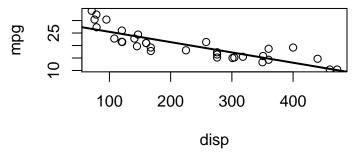


Figure 4: Linear regression

```
RR: Intro to
```

What? Why? 001-motivating-

example

LAT<sub>E</sub>X

knitr

Kniti

Example

002-minim

003-modeloutput

005-beamerpresentation 006-sensitivity-

parameter 007-sensitivityanalysis-many-

008-large-

009-rmarkdow 010-rmarkdow presentation

Final Romarks

### Include a Table

```
<<table.ex, results='asis'>>=
library(xtable)
iris[1:5,1:5] %>%
rins[1:5,1:5] %>%
print(include.rownames=FALSE)
@
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.10	3.50	1.40	0.20	setosa
4.90	3.00	1.40	0.20	setosa
4.70	3.20	1.30	0.20	setosa
4.60	3.10	1.50	0.20	setosa
5.00	3.60	1.40	0.20	setosa

Table 1: Sample of Iris data



What? Why? 001-motivating-

tting Starte

IAT<sub>E</sub>X RStudi

RStudi knitr

Example

002-minimum-

working-example 003-modeloutput

005-beamerpresentation

oub-sensitivityanalysis-oneparameter 007-sensitivity-

007-sensitivityanalysis-manyparameters

documents

009-rmarkdow 010-rmarkdow

Final Remarks

# Minimum Working Example

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/002-minimum-working-example

Reproducible

What? Why? 001-motivating-

tting Starte

LAT<sub>E</sub>X RStud

knitr

Examples

002-minis

working-example

003-mc

output
004-figures
005-beamerpresentation
006-sensitivityanalysis-oneparameter
007-sensitivityanalysis-manyparameters
008-largedocuments
009-rmarkdown
010-rmarkdown

Final Remarks

# Extracting output from Regression Models

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/003-model-output

### RR: Intro to knitr

Reproducible

What? Why? 001-motivating-

etting Starte

LAT<sub>E</sub>X RStud

knitr

Examples

002-minimumworking-example 003-model-

004-figures

005-beamerpresentation 006-sensitivityanalysis-oneparameter 007-sensitivityanalysis-manyparameters 008-largedocuments

presentation

# **Figures**

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/004-figures

#### RR: Intro to knitr

001-motivating-

LAT<sub>E</sub>X

005-beamer-

presentation

006-sensitivity-007-sensitivityanalysis-many-

documents

presentation

### Beamer Presentations

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/005beamer-presentation

Reproducible

What? Why? 001-motivat

tting Starte

IAT<sub>E</sub>X RStudio

KHIGI

Example

working-

003-mod

004-figure

005-beam

presentation 006-sensitivity-

#### analysis-oneparameter

007-sensitivityanalysis-manyparameters

008-large-

009-rmarkdow

010-rmarkdowr presentation

Final Remarks

# Changing one Parameter in an Analysis

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/006-sensitivity-analysis-one-parameter

Reproducible

What? Why? 001-motivation

tting Starte

LAT<sub>E</sub>X RStudio

knitr

Example

000 --:-

003-mod

output

005-beam

presentation 006-sensitivity-

parameter 007-sensitivity-

analysis-manyparameters

008-largedocuments 009-rmarkdow

inal Remarks

# Changing Many Parameters in an Analysis

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/007-sensitivity-analysis-many-parameters

Reproducible

What? Why? 001-motivating-

etting Starte

IAT<sub>E</sub>X RStud

RStud knitr

Example

003-model output 004-figure

005-beamerpresentation 006-sensitivityanalysis-one-

007-sensitivityanalysis-manyparameters

008-largedocuments

009-rmarkdowr 010-rmarkdowr presentation

Final Remarks

# Large Documents

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/008-large-documents

Reproducible

What? Why? 001-motivating-

etting Starte

LAT<sub>E</sub>X RStud

knitr

Example

003-modeloutput 004-figures 005-beamerpresentation 006-sensitivityanalysis-one-

parameter 007-sensitivityanalysis-manyparameters 008-large-

documents 009-rmarkdown 010-rmarkdownpresentation

Final Remarks

# **HTML** Reports

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/009-rmarkdown

#### RR: Intro to knitr

001-motivating-

LAT<sub>E</sub>X

006-sensitivity-

007-sensitivityanalysis-many-

documents

010-rmarkdownpresentation

### **HTML** Presentations

https://github.com/sahirbhatnagar/knitr-tutorial/tree/master/010rmarkdown-presentation

Reproducible

What? Why? 001-motivating

etting Started

LATEX RStudion

Examples

002-minimumworking-exampl

004-figures

005-beamerpresentation 006-sensitivity-

analysis-oneparameter 007-sensitivity-

analysis-ma

008-large

010-rmarkdow

Final Remarks

# Opinion: Reproducible research can still be wrong: Adopting a prevention approach

Jeffrey T. Leek<sup>a,1</sup> and Roger D. Peng<sup>b</sup>

<sup>a</sup>Associate Professor of Biostatistics and Oncology and <sup>b</sup>Associate Professor of Biostatistics, Johns Hopkins University, Baltimore, MD computational tools such as knitr, iPython notebook, LONI, and Galaxy (8) have simplified the process of distributing reproducible data analyses.

#### Reproducible

What? Why? 001-motivating-

Getting Starte

LAT<sub>E</sub>X

RStudi knitr

#### Example

002-minimumworking-examp 003-model-

output

004-figure

presentation

006-sensitivityanalysis-one-

007-sensitivityanalysis-many-

008-large-

documents

009-rmarkdow 010-rmarkdow

Final Remarks

# Is the juice worth the squeeze?

