Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Reproducible Research

An Introduction to knitr

Sahir Rai Bhatnagar[1]

May 28, 2014

[1] https://github.com/sahirbhatnagar/knitr-tutorial

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Acknowledgements

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
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-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

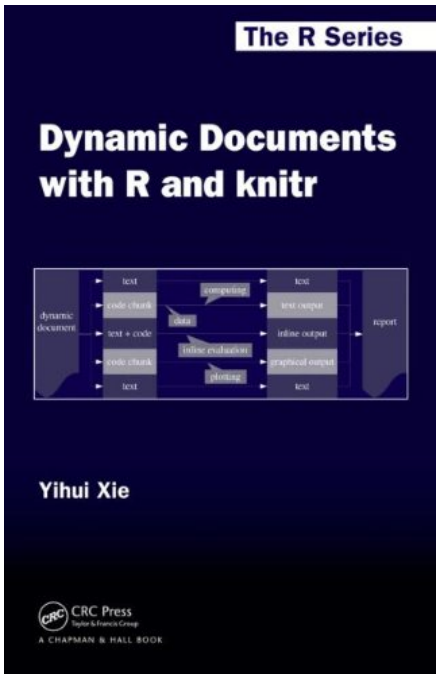
Disclaimer #2

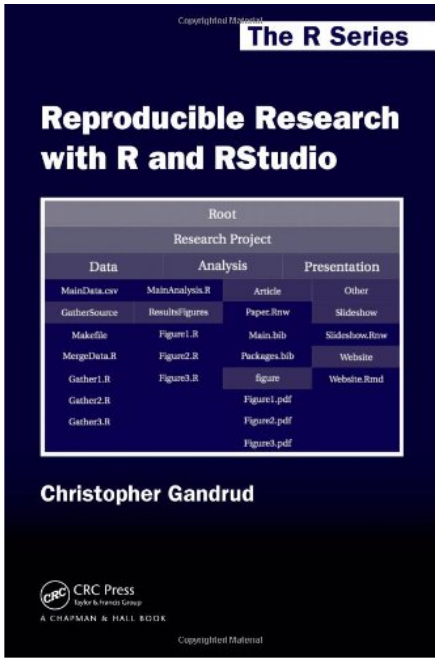
I don't work for, nor am I an author of any of these packages. I'm just a messenger.

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

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 Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Eat Your Own Dog Food

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Main objective for today

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

What is Science Anyway?

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

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*

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

RR: A Minimum Standard to Verify Scientific Findings

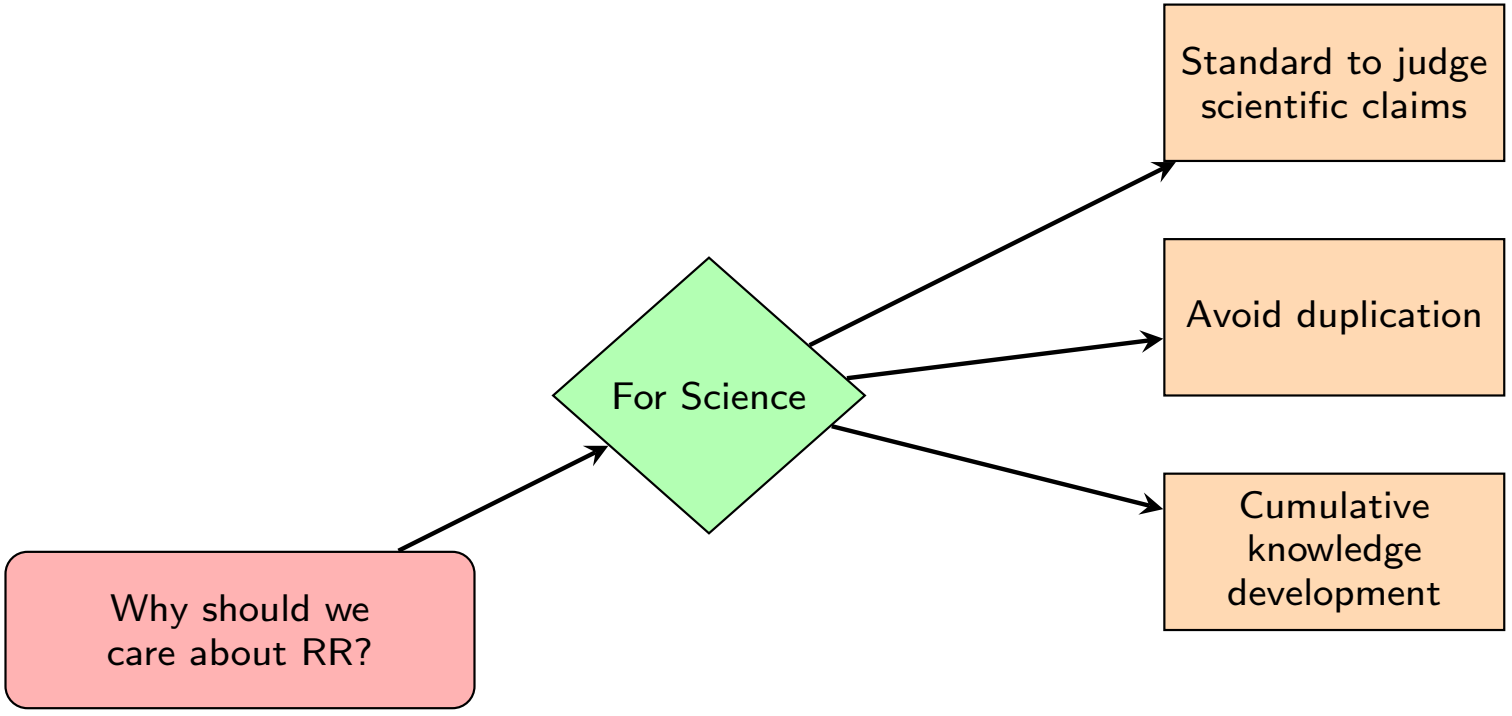
Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

RR: A Minimum Standard to Verify Scientific Findings

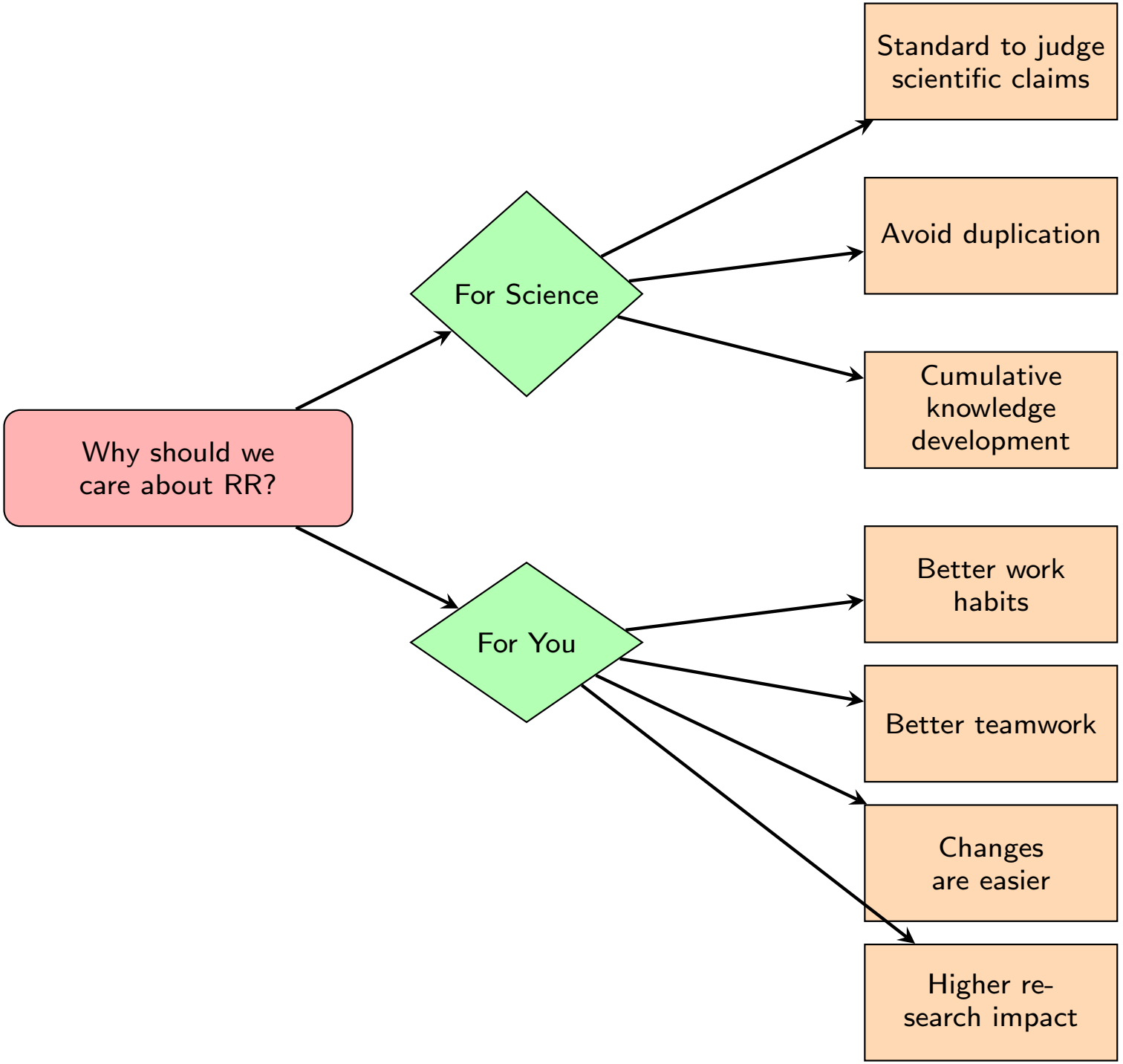
Reproducible Research (RR) in Computational Sciences

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 Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks



Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks



Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

A Motivating Example

Demonstrate: 001-motivating-example
Survey: https://www.surveymonkey.com/s/CDVXW3C

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Tools for Reproducible Research[2]

Free and Open Source Software

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

[2] http://onepager.togaware.com/

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Comparison

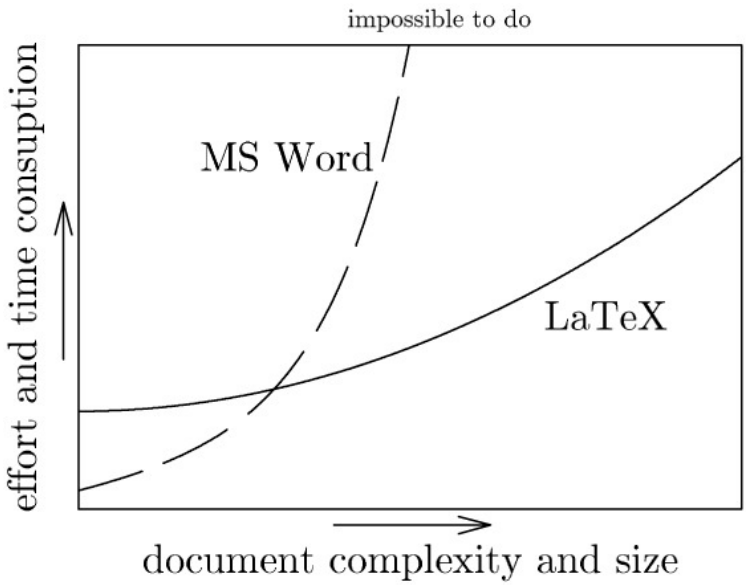


Figure 1: Comparison

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

The Philosophy behind LaTeX

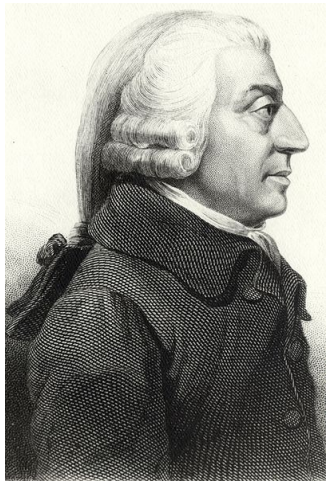


Figure 2: Adam Smith, author of *The Wealth of Nations* (1776), in which he conceptualizes the notion of the division of labour

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 Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

The Genius Behind LaTeX

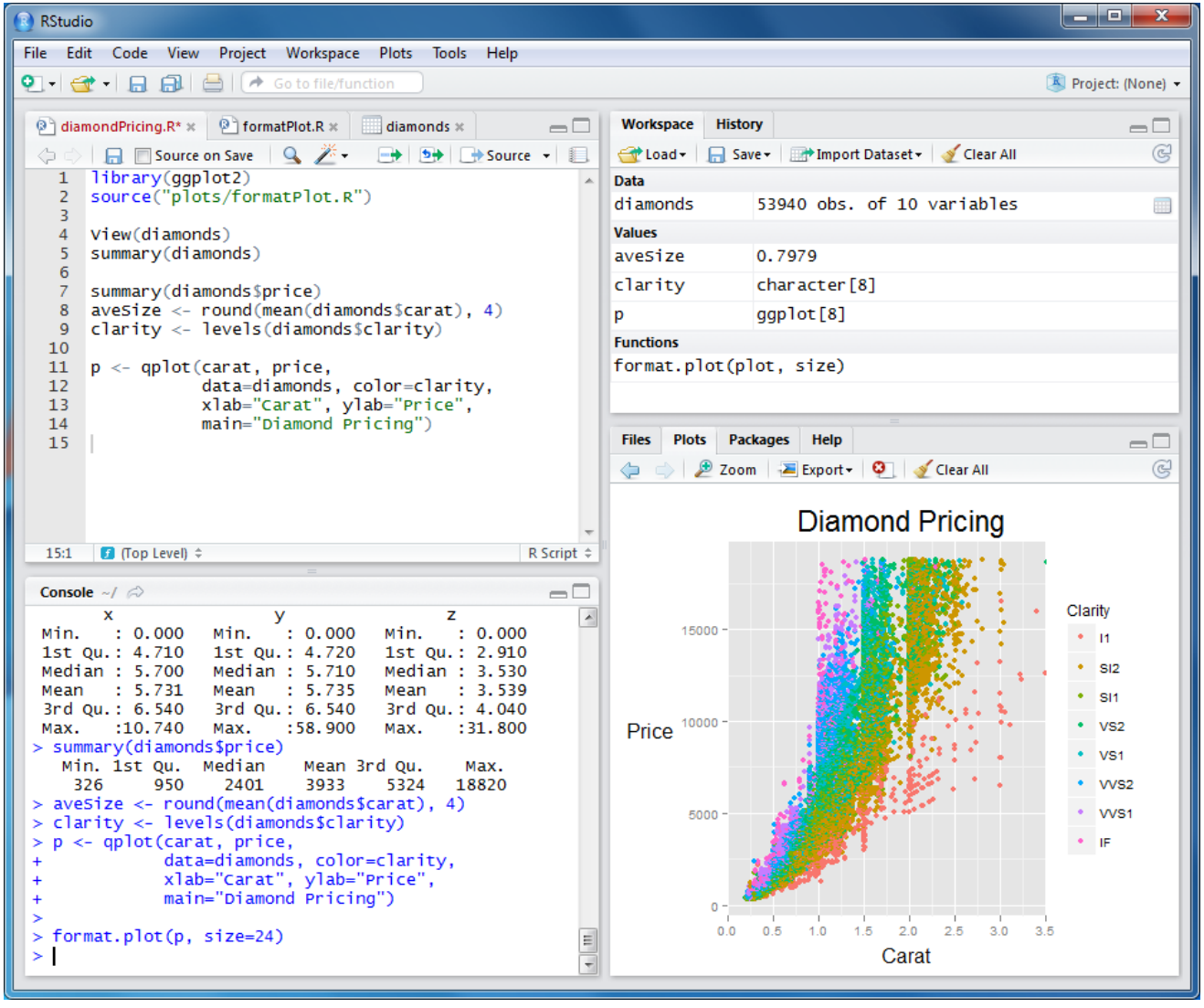
Figure 3: The TeX project was started in 1978 by Donald Knuth (Stanford). He planned for 6 months, but it took him nearly 10 years to complete. Coined the term "Literate programming": mixture of code and text segments that are "human" readable. Recipient of the Turing Award (1974) and the Kyoto Prize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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Integrated Development Environment (IDE)

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Integrated Development Environment (IDE)

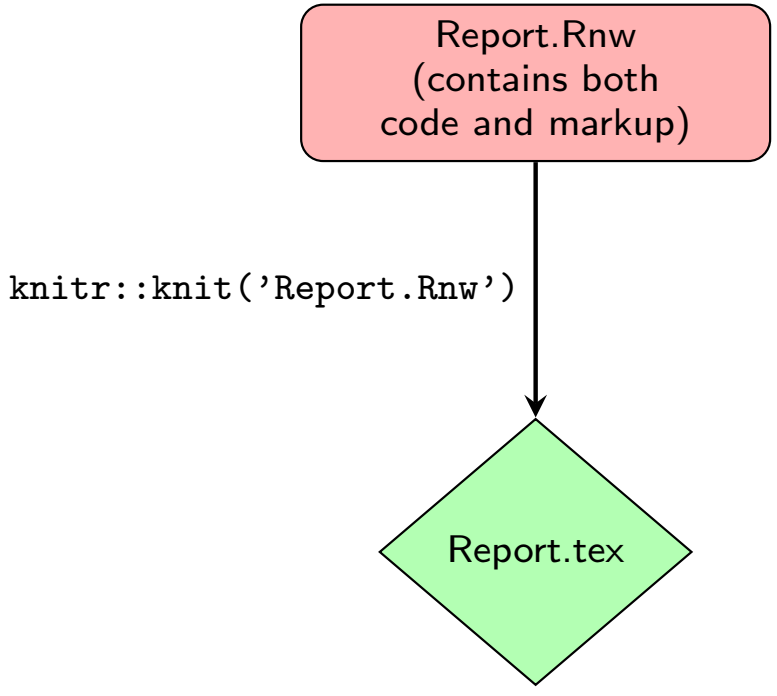


Demonstrate: Explore `RStudio`

What knitr does

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

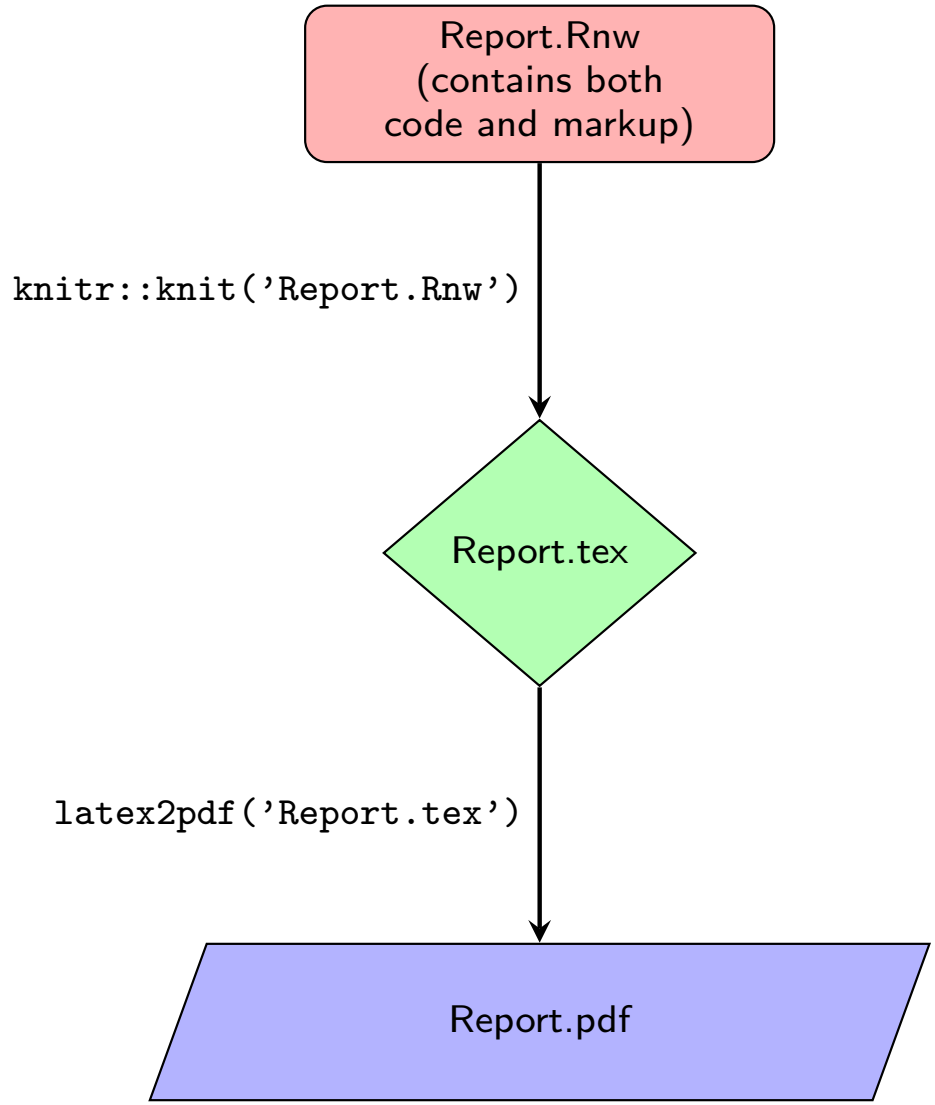
LaTeX example:



What knitr does

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

LaTeX example:



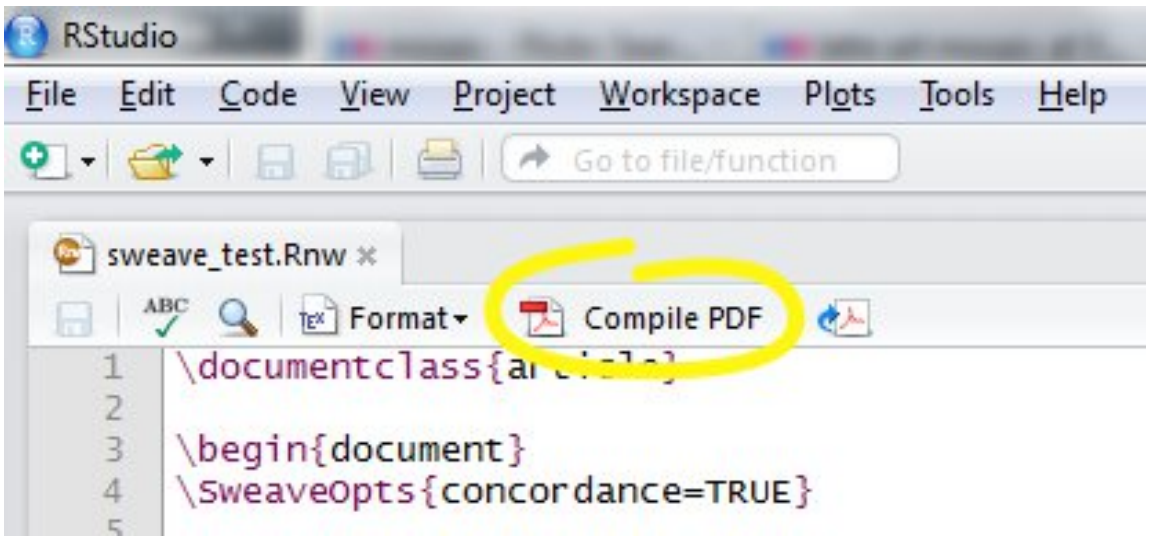
Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Compiling a `.Rnw` document

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

```
knitr::knit2pdf()
```

or in RStudio:



Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Incorporating R code

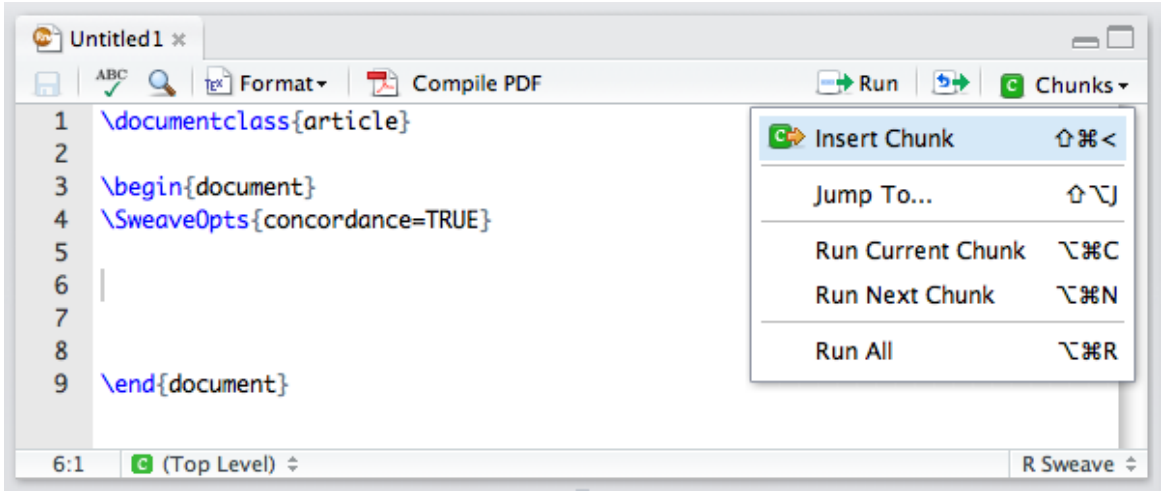
- Insert R code in a **Code Chunk** starting with

  << >>=

  and ending with

  @

In RStudio:



Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Example 1

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

produces

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

## [1] 0.12
```

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Example 2

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

produces

```
for (i in 1:5) {
    (i + 3) %>% print
}
```

```
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
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] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
```

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-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
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
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

Include a Figure

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

```
<<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)
@
```

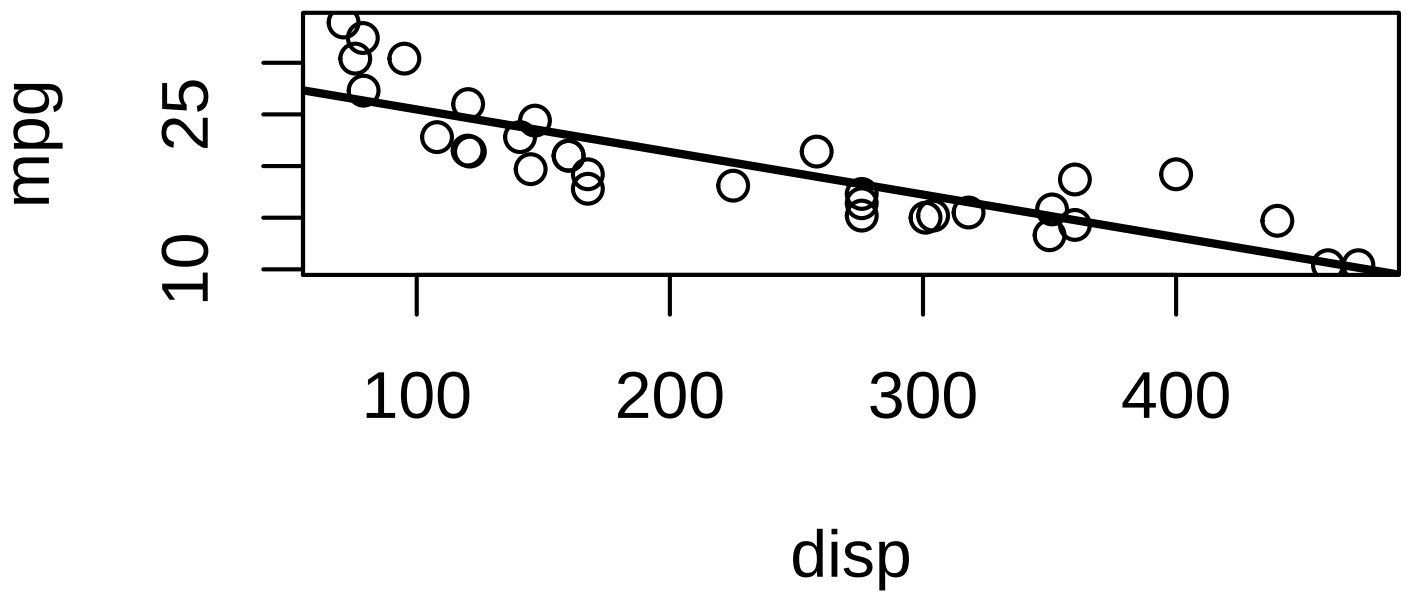


Figure 4: Linear regression

Include a Table

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

```
<<table.ex, results='asis'>>=
library(xtable)
iris[1:5,1:5] %>%
xtable(caption='Sample of Iris data') %>%
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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Minimum Working Example

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Extracting output from Regression Models

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Figures

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Beamer Presentations

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Changing one Parameter in an Analysis

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Changing Many Parameters in an Analysis

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Large Documents

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

HTML Reports

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

HTML Presentations

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

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

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

Jeffrey T. Leek[a,1] and Roger D. Peng[b]
[a]*Associate Professor of Biostatistics and Oncology and* [b]*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.

Always Remember ...

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

$$\text{Reproducibility} \propto \frac{1}{\text{copy paste}}$$

Reproducible Research
What?
Why?
001-motivating-example
Getting Started
LaTeX
RStudio
knitr
Examples
002-minimum-working-example
003-model-output
004-figures
005-beamer-presentation
006-sensitivity-analysis-one-parameter
007-sensitivity-analysis-many-parameters
008-large-documents
009-rmarkdown
010-rmarkdown-presentation
Final Remarks

Is the juice worth the squeeze?