Avoir une présence en ligne
Outils pour la diffusion rapide et reproductible de la recherche

Sahir Rai Bhatnagar[1]

May 14, 2019

---

[1]https://github.com/sahirbhatnagar/raqc

Outils pour la
diffusion rapide
et reproductible
de la recherche

Remerciements

- La comité organisateur
- Pierre Racine et Sophie Baillargeon
- Don Knuth (TEX)
- Friedrich Leisch (Sweave)
- Yihui Xie (knitr)
- Vous

Outils pour la
diffusion rapide
et reproductible
de la recherche

Avis #1

- Ceci est une **introduction** au outils pour la recherche reproductible

Outils pour la
diffusion rapide
et reproductible
de la recherche

Avis #1

- Ceci est une **introduction** au outils pour la recherche reproductible
- Le niveau de cet atelier est "intermédiaire" et suppose des connaissances de base en R ainsi que de l'environnement RStudio

Outils pour la
diffusion rapide
et reproductible
de la recherche

Avis #1

- Ceci est une **introduction** au outils pour la recherche reproductible
- Le niveau de cet atelier est "intermédiaire" et suppose des connaissances de base en R ainsi que de l'environnement RStudio
- N'hésitez pas à posez des questions

Outils pour la
diffusion rapide
et reproductible
de la recherche

Avis #2

Je n'ai aucune relation commerciale avec ces logiciels.

Outils pour la
diffusion rapide
et reproductible
de la recherche

Avis #3

- Le matériel pour cet atelier est basé sur plusieurs ressources
- Voir ce lien pour une liste complète de références:
  https://github.com/sahirbhatnagar/raqc
- Une grande partie du contenu de ces diapositives est basée sur ces deux livres:
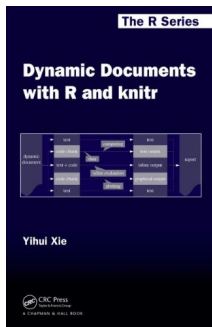
Outils pour la
diffusion rapide
et reproductible
de la recherche

Eat Your Own Dog Food

- Ces diapositives sont reproductibles
- code:
  https://github.com/sahirbhatnagar/raqc/tree/master/slides

Outils pour la
diffusion rapide
et reproductible
de la recherche

Le programme de l'atelier

- **8h30 à 10h00**: Introduction aux raports reproductibles avec knitr et RMarkdown
- **10h00 à 10h30**: Pause
- **10h30 à 12h00**:
- **13h30 à 15h00**: Créer un siteweb avec blogdown
- **15h00 à 15h30**: Pause
- **15h30 à 17h**:
- **17h**: Fin de l'atelier

Outils pour la
diffusion rapide
et reproductible
de la recherche

What is Science Anyway?

Outils pour la
diffusion rapide
et reproductible
de la recherche

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*

Outils pour la
diffusion rapide
et reproductible
de la recherche

RR: A Minimum Standard to Verify Scientific Findings

Outils pour la
diffusion rapide
et reproductible
de la recherche
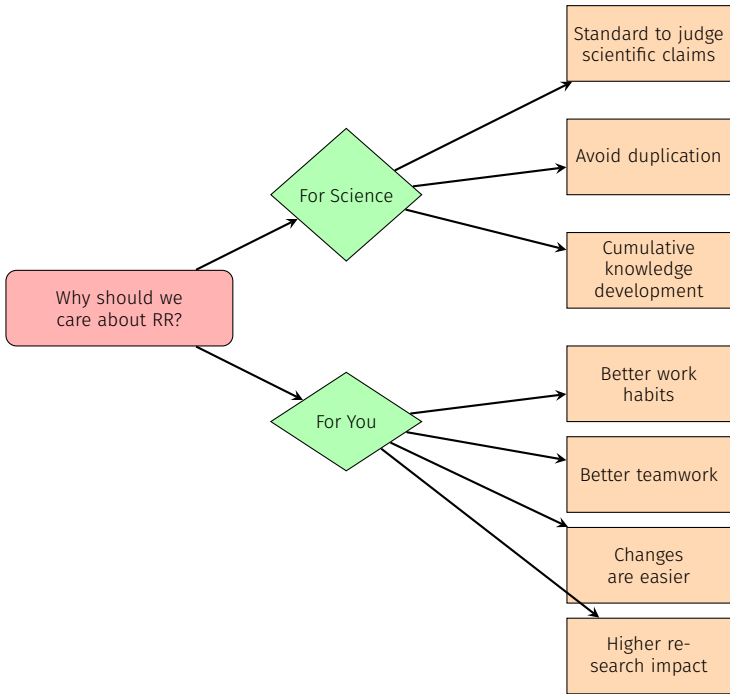
RR: A Minimum Standard to Verify Scientific Findings

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
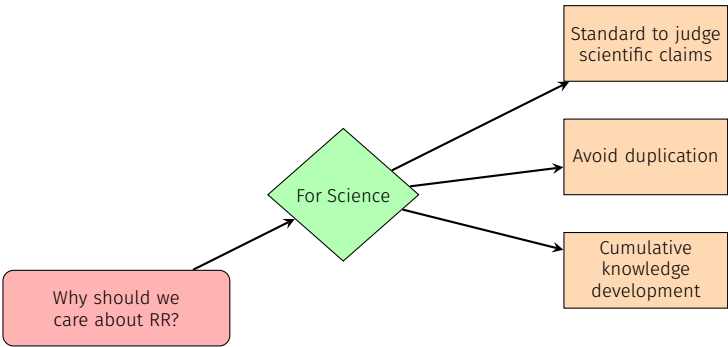
Outils pour la
diffusion rapide
et reproductible
de la recherche

Why should we care about RR? → For Science → Standard to judge scientific claims

Avoid duplication

Cumulative knowledge development

Outils pour la
diffusion rapide
et reproductible
de la recherche

Outils pour la
diffusion rapide
et reproductible
de la recherche

A Motivating Example

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

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 LaTeX and R code. Based on Prof. Friedrich Leisch's Sweave

---

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Comparison

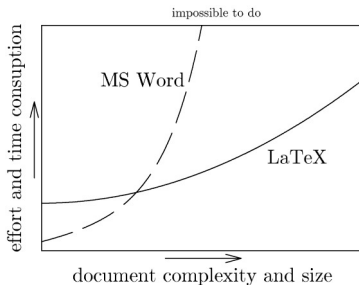
Figure 1: Comparison

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

Outils pour la
diffusion rapide
et reproductible
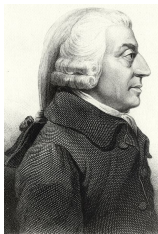de la recherche

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

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).

Outils pour la
diffusion rapide
et reproductible
de la recherche

Integrated Development Environment (IDE)

Outils pour la
diffusion rapide
et reproductible
de la recherche

Integrated Development Environment (IDE)

Demonstrate: Explore `RStudio`

Outils pour la
diffusion rapide
et reproductible
de la recherche

What `knitr` does

LaTeX example:

Outils pour la
diffusion rapide
et reproductible
de la recherche

What `knitr` does

LaTeX example:


```
Report.Rnw
(contains
both code
and markup)
```

`knitr::knit('Report.Rnw')`

```
Report.tex
```

`latex2pdf('Report.tex')`

```
Report.pdf
```

Outils pour la
diffusion rapide
et reproductible
de la recherche

Recherche
Reproductible
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
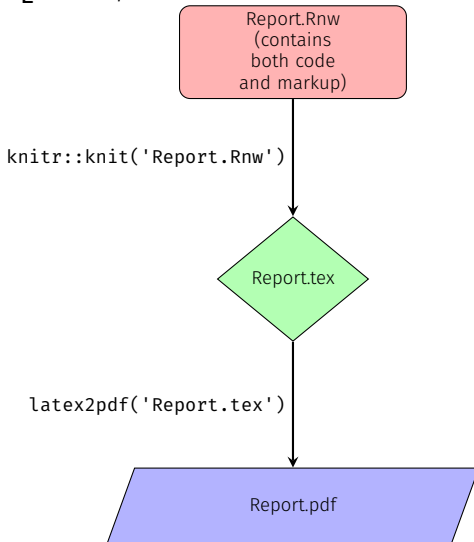007-sensitivity-
analysis-many-
parameters
008-large-
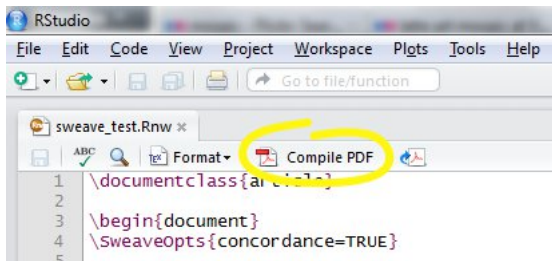documents
009-rmarkdown
010-rmarkdown-
presentation

Final Remarks

Compiling a `.Rnw` document

The two steps on previous slide can be executed in one command:
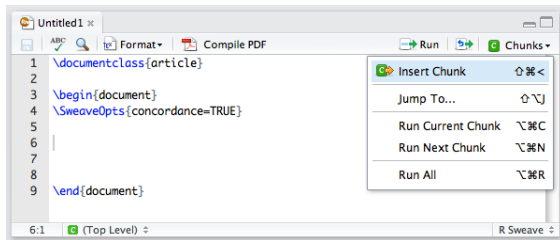
```
knitr::knit2pdf()
```

or in `RStudio`:

Outils pour la
diffusion rapide
et reproductible
de la recherche

Incorporating **R** code

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

$$<< \quad >>=$$

and ending with

$$@$$

In `RStudio`:

Outils pour la
diffusion rapide
et reproductible
de la recherche

Recherche
Reproductible
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
```

Outils pour la
diffusion rapide
et reproductible
de la recherche

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
```

Outils pour la
diffusion rapide
et reproductible
de la recherche

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
```

Outils pour la
diffusion rapide
et reproductible
de la recherche

Example 2.3

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

produces
Demonstrate: Try it yourself

Outils pour la
diffusion rapide
et reproductible
de la recherche

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

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)
@
```
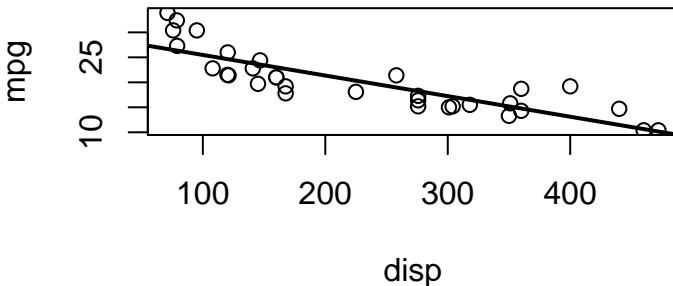

Figure 4: Linear regression

Include a Table

```
<<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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Minimum Working Example

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Extracting output from Regression Models

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Figures

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Beamer Presentations

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

Changing one Parameter in an Analysis

Outils pour la
diffusion rapide
et reproductible
de la recherche

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Changing Many Parameters in an Analysis

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Large Documents

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

HTML Reports

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

HTML Presentations

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

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.

Outils pour la
diffusion rapide
et reproductible
de la recherche

Always Remember …

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

Outils pour la
diffusion rapide
et reproductible
de la recherche

Is the juice worth the squeeze?