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

Sahir Rai Bhatnagar¹

May 14, 2019

¹https://github.com/sahirbhatnagar/ragc

Remerciements

Recherche Reproductible (RR)

Pourquoi? 001-exemp

Commenc

ET_EX RStudio

knitr RMarkdov

Evemnle

Exemple

working-exa

003-model-outpu

004-ngures

presentati

parameter 007-sensitivity analysis-many

008-larg

009-rmarkdowr

010-rmarkdowi presentation

Final Remarks

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





Avis #1

Recherche Reproductib (RR)

Pourquoi?

Commeno

EleX

knitr

Exemple

working-exampl

003-model-out

004-figures

005 -----

presentatio

006-sensitiv

007-sensitivit

paramete

documents

009-rmarkdo

010-rmarkdow

Final Remarks

 Ceci est une introduction au outils pour la recherche reproductible

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**

Avis #1

Recherche Reproductible (RR)

Pourquoi?

Commend

ET_EX PStudio

knitr

RMarkdov

Excilipite

003-model-output

004-figures

004-пgures 005-beamer-

006-sensitivit analysis-one-

007-sensitivit analysis-man

008-large-

documents

010-rmarkdow

presentation

Final Remari

- 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

Avis #2



R Markdown v2







Je n'ai aucune relation commerciale avec ces logiciels.

Recherche Reproductible

Reproduction [RR] Quoi? Pourquoi? 001-exemplejustificatif

ET_EX RStudio

knitr RMarkdown

Exemple

working-exampl

004-figures

005-beamer-

analysis-oneparameter

parameters

documents 009-rmarkdow

010-rmarkdown presentation

rillat Reliidi KS

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:





Eat Your Own Dog Food

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

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

Recherche Reproductible (RR)

Recherche Reproductible (RR)

C'est quoi la science?

Recherche Reproductib (RR)

Quoi?

001-exemple-

Commenc

ETEX

RStudi

KHILLE

RMarkdov

Exemple

002-minimum-

003-model-outo

004-figures

005 6----

presentatio

006-sensitiv

parameter

007-sensitivi analysis-mai

008-lan

documents

010-rmarkdow

010-rmarkdow presentation

Final Remark

Recherche Reproductible

Quoi?

Pourquoi? 001-exemp

Commence

FT-X

RStud

RMarkdow

Exemples

002-minim

working-example

003-model-outp

004-figures

005-beam

006-sensitiv

007-sensitivit

parameters

documents

009-rmarkdow

010-rmarkdown presentation

Final Remarks

C'est quoi la science?

Selon l'American Physical Society:

La science est l'entreprise systématique consistant à rassembler des connaissances sur l'univers et à les organiser et les condenser en lois et **théories vérifiables**...

ble :he

Recherche Reproductible (RR)

Quoi?

ourquoi?

Commeno

ETEX

RStudi

RMarkdov

Exemple

002-min

002-minimum working-exam

003-model-out

004-figures

005-beamer-

006-sensitivit analysis-one

007-sensitivit analysis-man

008-large-

documents

009-rmarkdow 010-rmarkdow

010-rmarkdowi presentation

Final Kemark

C'est quoi la science?

Selon l'American Physical Society:

La science est l'entreprise systématique consistant à rassembler des connaissances sur l'univers et à les organiser et les condenser en lois et **théories vérifiables**...

Le succès et la crédibilité de la science sont fondés sur la volonté des scientifiques d'exposer leurs idées et leurs résultats à des tests indépendants et à leur reproduction par d'autres scientifiques.

RR: Une norme minimale pour vérifier les résultats scientifiques

Recherche Reproductibl (RR)

Quoi?

Pourquoi? 001-exemple

Commence

ETEX

RStudi

DMarkdow

C.

exempte:

working-exampl

003-model-outp

004-figure:

005-beam

006-sensiti

analysis-on

007-sensitivit analysis-mar

008-larg

documents

010-rmarkdow

010-rmarkdowi presentation

mat remain

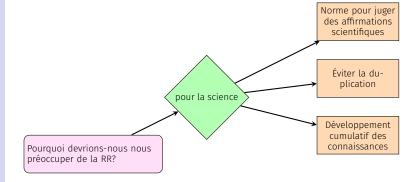
Ouoi?

RR: Une norme minimale pour vérifier les résultats scientifiques

Recherche reproductible (RR) dans la science des données

Les données et le code utilisés pour effectuer une constatation sont disponibles et suffisent à un chercheur indépendant pour recréer la constatation

Pourquoi?



Recherche Reproductible (RR)

Оної

Pourquoi? 001-exemp

Commence

Commenc

RStudio

RMarkdow

Exemple

Excilipite

002-modol-outpu

003-model-outpu

004-figures

presentation

006-sensitivi analysis-one

007-sensitivity

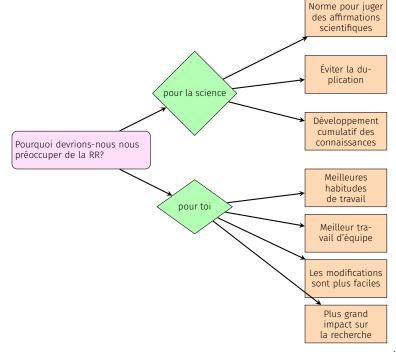
008-large-

documents

009-rmarkdov

010-rmarkdow presentation

Final Remarks



Un exemple justificatif

001-exemplejustificatif

Démo: 001-motivating-example

Commencer

Commencer

Outils pour la diffusion rapide de la recherche

Outils pour la recherche reproductible²

Logiciel gratuit et « open source »

- RStudio: Créer, gérer, compiler des documents
- धा-X: langage de balisage pour la composition d'un document
- R: Langage d'analyse statistique
- knitr: Intègre le code धाःXet le code R. La version moderne de Sweave du professeur Friedrich Leisch
- RMarkdown: Intègre le code Markdown et le code R

²http://onepager.togaware.com/

Comparaison

impossible to do effort and time consuption ET_EX MS Word LaTeX document complexity and size Figure 1: Comparison

- 町EX a une plus grande courbe d'apprentissage
- De nombreuses tâches sont très difficiles ou impossibles (la plupart des cas) à effectuer dans MS Word ou Libre Office

Final Remark

La philosophie derrière **ET**FX

ET_EX

Figure 2: Adam Smith, l'auteur de The Wealth of Nations (1776), dans lequel il conceptualise la notion de division du travail

Division du travail

La composition et la structuration logique du texte constituent la contribution spécifique de l'auteur à la production d'un texte imprimé.

La philosophie derrière **M**EX

Recherche Reproductible (RR) Quoi?

Commence

ETEX

RStudio

knitr RMarkdo

RMarkdo

Exemple

working-example

003-model-outpu

005-heam

presentati

analysis-one-

008-largedocuments

009-rmarkdown

Final Rem



Figure 2: Adam Smith, l'auteur de The Wealth of Nations (1776), dans lequel il conceptualise la notion de division du travail

Division du travail

La composition et la structuration logique du texte constituent la contribution spécifique de **l'auteur** à la production d'un texte imprimé.

Des questions telles que le choix de la famille de polices, les en-têtes de section doivent-ils être en caractères gras ou en petites capitales? Doivent-ils être alignés à gauche ou centrés? Le texte doit-il être justifié ou non? Les notes doivent-elles apparaître au bas de la page ou à la fin? Le texte doit-il être placé dans une colonne ou deux? et ainsi de suite, est l'affaire de la typographe

Le génie derrière **ET**EX



Figure 3: The T_EX 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).

Reproductibl (RR)

Quoi?

001-exemp

Commend

ET_EX RStuc

RMarkdov

Exemples

000 --:-:

003-model-outp

004-figures

005-beam

006-sensitivit analysis-one

007-sensitivit analysis-man

008-larg

009-rmarkdow

presentation

rillat Reillatks

Integrated Development Environment (IDE)

Recherche Reproductibl (RR)

Quoi

001-exempl

Commence

Commence

RStudio

RStud

Knitr

200

Exemple

working-example

003-model-outp

004-figures

nne hoam

005-beam

006-sensiti

parameter

analysis-mai

008-lare

documents

009-rmarkdown

010-rmarkdow presentation

Final Remarks

Recherche Reproductible (RR)

Quoi?

Pourquoi? 001-exempl

Commenc

ET-Y

RStudio

RMarkdow

Exemple

working-example

004-figures

005-beam presentati

006-sensitiv analysis-one

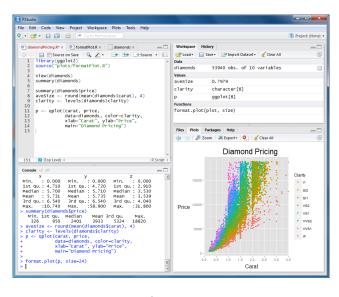
007-sensitivi analysis-mar parameters

008-large documen

009-rmarkdown 010-rmarkdown presentation

Tillat Italiiai

Integrated Development Environment (IDE)



Demonstrate: Explore RStudio

Recherche Reproductible (RR)

Quoi

Pourquoi? 001-exem

Commenc

Commenc

knitr

KIIICI

Exemple

002-minimu

002 model out

ous-modet-outp

004-figures

presentatio

006-sensitiv

007-sensitivit

008-large-

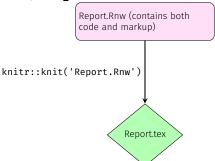
documents

010-rmarkdow

Final Remarks

Que fait knitr

Exemple **LTEX**:



knitr

Que fait knitr

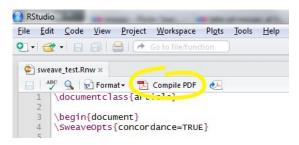
Exemple LTFX: Report.Rnw (contains both code and markup) knitr::knit('Report.Rnw') Report.tex latex2pdf('Report.tex') Report.pdf

Compiler un document . Rnw

Les deux étapes de la diapositive précédente peuvent être exécutées en une seule commande:

knitr::knit2pdf()

ou dans **RStudio**:



knitr

Recherche Reproductible

Reproductible (RR) Quoi?

Pourquoi? 001-exemple

Commence

ETEX

knitr

RMarkdov

Exemple

working-example 003-model-outpu 004-figures

005-beamerpresentation 006-sensitivity analysis-one-

007-sensitivity analysis-many parameters 008-large-

009-rmarkdow 010-rmarkdowi presentation

Final Remarks

Incorporer le code R

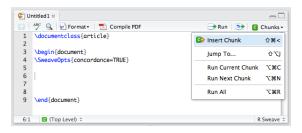
• Insérer le code R dans un morceau de code commençant par



et se terminant par



Dans **RStudio**:



```
Outils pour la
diffusion rapide
et reproductible
de la recherche
```

Exemple 1

knitr

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

produces

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

[1] 0.12

```
Outils pour la
diffusion rapide
et reproductible
de la recherche
```

Exemple 2

knitr

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

produces

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

```
Outils pour la
diffusion rapide
et reproductible
de la recherche
```

Example 2.2

```
Recherche
Reproductibl
(RR)
```

Quoi?

Commence

ETEX

RStudio

knitr

Exemple

003-model-outpu

004-figures

005-beamerpresentation

006-sensitivit analysis-one

007-sensitivity analysis-man parameters

documents

009-rmarkdown 010-rmarkdown

010-rmarkdow presentation

Final Remarks

```
<<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

```
Reproductib
(RR)
Quoi?
```

001-exempl justificatif

Commenc

RStudio

RStudio

RMarkdo

Exemple

002-minim

003-model-outp

003-model-outp

004-figures

presentatio

analysis-one-

007-sensitivi analysis-mar parameters

008-lar

and an add

010-rmarkdow

010-rmarkdow presentation

Final Remarks

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

produces

Démo: Essayez vous-même

Recherche Reproductible (RR)

(RR)
Quoi?
Pourquoi?

Commence

ETEX

knitr

RMarkd

002-mini

003-model-o

004-figures

005-beame

006-sensitivit analysis-oneparameter

007-sensitivity analysis-man parameters

008-large documen

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

Recherche Reproductible (RR) Quoi?

justificatif

Commence

BIEX

knitr

TUTTUT

Exemple

working-examp

003-model-outp

004-figures

005-beame presentation

analysis-oneparameter

parameters

008-large-

documents

010-rmarkdowr

Final Remar

Final Remarks

Include a Figure

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

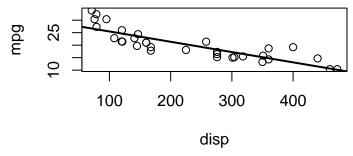


Figure 4: Linear regression

```
Outils pour la
diffusion rapide
et reproductible
de la recherche
```

knitr

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

Markdown: HTML without knowing HTML

example.Rmd × RStudio: Preview HTML Knit HTML Chunks + Preview: ~/example.html Header 1 Header 1 This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages. This is an R Markdown document, Markdown i simple formatting syntax for authoring web page Use an asterisk mark, to provide emphasis such as Use an asterisk mark, to provide emphasis suc *italics* and **bold**. as italics and bold. RMarkdown Create lists with a dash: Create lists with a dash: Item 1 - Item 1 Item 2 - Item 2 Item 3 10 - Ttem 3 You can write in-line code with a back-tick. 11 12 You can write `in-line` code with a back-tick. Code blocks display 13 with fixed-width font 14 15 Code blocks display Blockquotes are offset 16 with fixed-width font 17

18 19

20

> Blockquotes are offset

Recherche Reproductible (RR)

Quoi?

Pourquoi? 001-exemple

Commence

ETEX

RMarkdown

Exemple

OO2 minin

003-model-outpi

004-figures

005-beam

006-sensitivit

007-sensitivit analysis-man

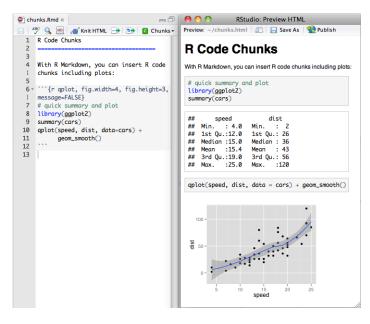
008-large

009-rmarkdov

presentation

nal Remarks

R + Markdown = RMarkdown



Recherche Reproductible (RR)

Quoi

Pourquoi: 001-exem

Commenc

Commenc

DC*****

RSTUDIO

RMarkdown

Evomple

Exemple

working-examp

003-model-ou

004-figures

005-heam

006-sensiti

analysis-on

007-sensitivit analysis-man

008-large

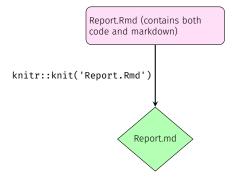
documents

010-rmarkdov

Final Remarks

What rmarkdown does

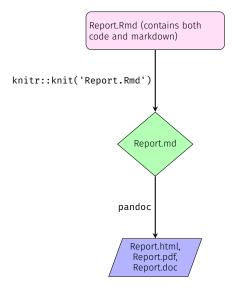
RMarkdown example:



RMarkdown

What **rmarkdown** does

RMarkdown example:

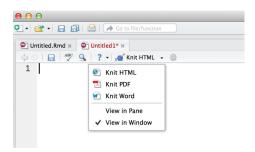


Compiling a . Rmd document

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

rmarkdown::render()

or in **RStudio**:



RMarkdown

How to choose between **ETFX** and Markdown?

Recherche Reproductible (RR)

Quoi? Pourquoi?

Commenc

RStudio

knitr RMarkdown

Evomple

Exemple

working-example 003-model-outs

003-model-outpt

005-beam

006-sensitivi analysis-one

007-sensitivity analysis-many

008-large-

documents

010-rmarkdow

presentation

Final Remark

math/stat symbols beamer presentations customized documents publish to journals, arXiv

Markdown

quick and easy reports use javascript libraries interactive plots publish to websites

Exemples

Exemples

Minimum Working Example

002-minimumworking-example

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

Extracting output from Regression Models

003-model-output

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

Figures

Recherche Reproductib (RR)

Quoi?

001-exempl

Commence

COMMICTIC

RStud

KHILL

RMarkdo

Exemp

002-minimun

003-model-outpu

004-figures

005-beame

presentatio

analysis-one parameter

analysis-ma

008-larg

documents

010-rmarkdow

010-rmarkdow presentation

Final Remarks

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

Beamer Presentations

Recherche Reproductib (RR)

Quoi

001-exempl

Commenc

Commenc

RStudi

knitr

RMarkdo

Exemp

002-minimu

working-exa

003-model-outp

004-figures

005-beamer-

presentation

analysis-oneparameter

007-sensitivi analysis-mar

008-larg

documents

010-rmarkdou

010-rmarkdow presentation

Final Remarks

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

Changing one Parameter in an Analysis

Recherche Reproductib (RR)

Quoi?

001-exempl

Commence

Commenc

RStudio

knitr

RMarkdo

Exemp

002-minim

002-model-out

004-figures

005-beamer

006-sensitivity-

analysis-oneparameter

007-sensitivity analysis-man

008-large-

documents

009-rmarkdov

010-rmarkdow presentation

Final Remarks

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

Changing Many Parameters in an Analysis

007-sensitivityanalysis-manyparameters

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

Large Documents

008-large-

documents

https://github.com/sahirbhatnagar/raqc/tree/master/008-largedocuments

HTML Reports

Recherche Reproductib (RR)

Quoi

001-exemp

Commence

Commenc

RStudi

knitr

RMarkdo

Exemp

002-minimu

002 model outs

004-figures

005-beam

presentatio

006-sensitiv analysis-on

007-sensitivit

paramete

documents

009-rmarkdown

010-rmarkdov presentation

Final Remarks

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

HTML Presentations

Recherche Reproductib (RR)

Quoi

001-exempl

Commence

ICT V

RStudio

Knitr

RMarkdo

Exemp

002-minim

working-example

003-1110061-1

004-ngure

005-beam

006-sensiti

narameter

007-sensitivi

paramete

documents

009-rmarkdov

010-rmarkdownpresentation

presentation

Final Remarks

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

Recherche Reproductible (RR)

Quoi?

001-exemple

commenc

RStudio

knitr

RMarkdow

Evennles

Litemples

working-example

002-model-oute

004-figures

005-beamer

presentation

006-sensitivi

007-sensitivi

analysis-mar parameters

008-large

documents

010-rmarkdov

presentation

Final Remarks

Final Remarks

Recherche Reproductibl (RR)

Quoi?

001-exemp

Commenc

RStudi

knitr RMarkdow

Evemnle

Exempte:

working-example

003-model-outpi

005-hoamor

presentatio

analysis-one

007-sensitivit analysis-man

008-large

documents

010-rmarkdow

010-rmarkdow presentation

Final Remarks

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

Jeffrey T. Leek^{a,1} and Roger D. Peng^b

^aAssociate Professor of Biostatistics and Oncology and ^bAssociate 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 ...

Final Remarks

Reproducibility \propto copy paste

Is the juice worth the squeeze?

Recherche Reproductib (RR)

Quoi?

001-exempl

Commenc

ETEX

RStudi

RMarkdou

Exemple

002-minimum-

oos moder our

004-figures

nne hoomer

000 -----

analysis-one

007-sensitivity

parame

documents

009-rmarkdo

010-rmarkd

presentation

Final Remarks

