

# Flux de travail intégré avec R

**Hadrien Commenges**

Séminaire RUSS

Paris, 24 novembre 2016

# R et le flux de travail

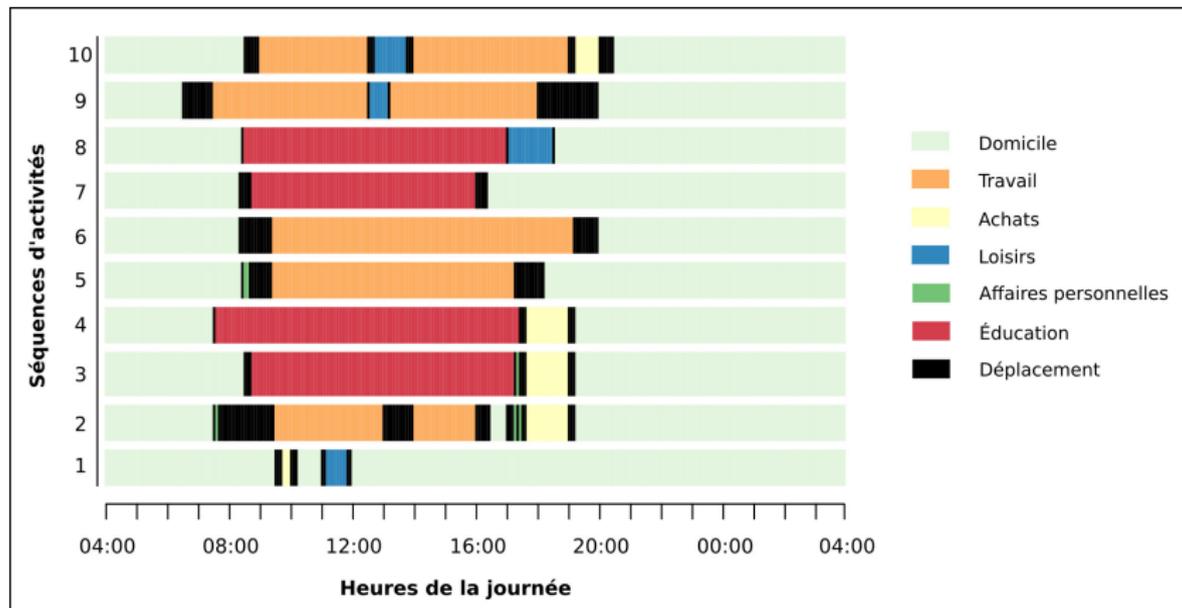
- ① Intégration des tâches dans un flux de travail individuel
- ② Articulation et suivi dans un flux de travail collectif
- ③ Production d'outils labo *longue traîne*

(1)

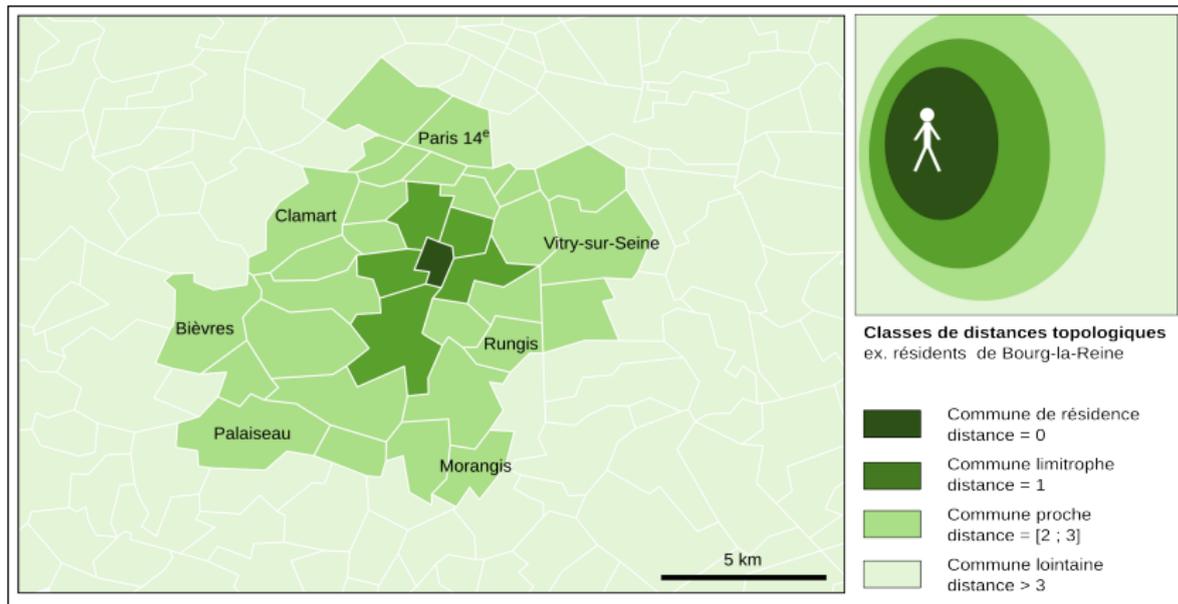
# Intégration des tâches dans un flux de travail individuel

# Exemple de flux de travail

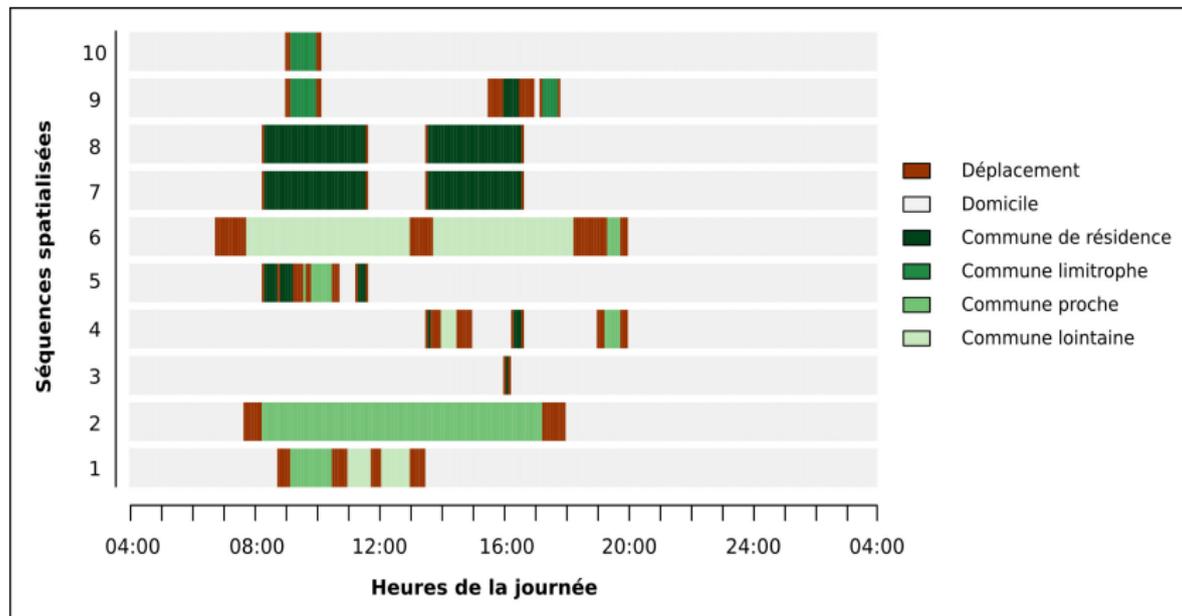
Commenges H. (2014) "La mobilité comme variabilité temporelle de la présence spatiale", *Flux*, n°95, pp.41–55.



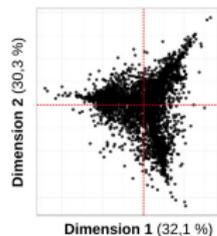
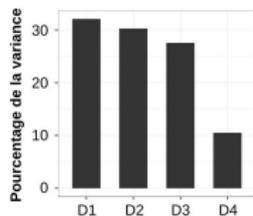
# Exemple de flux de travail



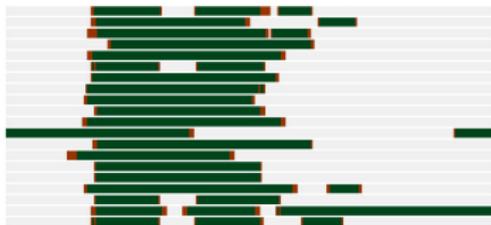
# Exemple de flux de travail



# Exemple de flux de travail



**Groupe 1 (n = 4305)**



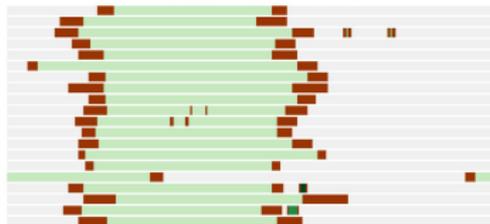
**Groupe 2 (n = 2136)**



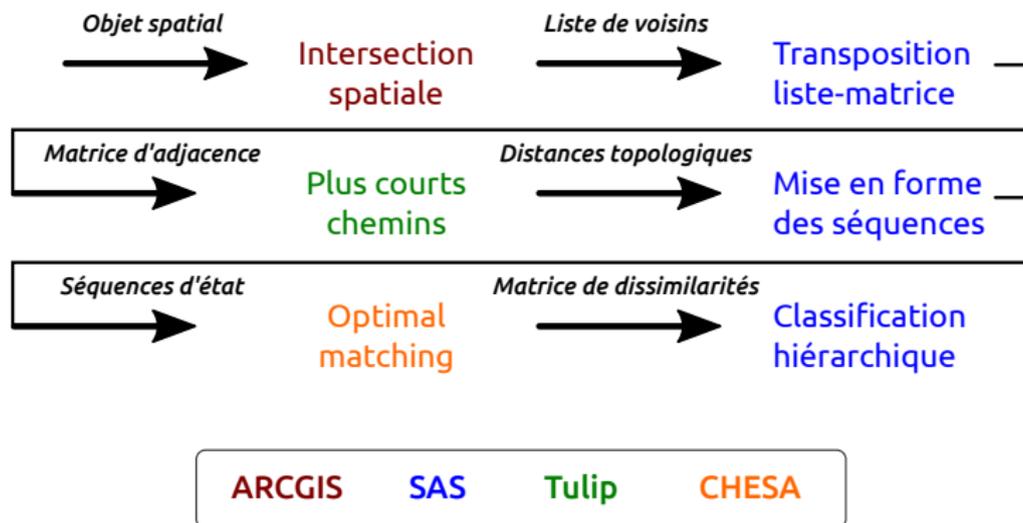
**Groupe 3 (n = 3100)**



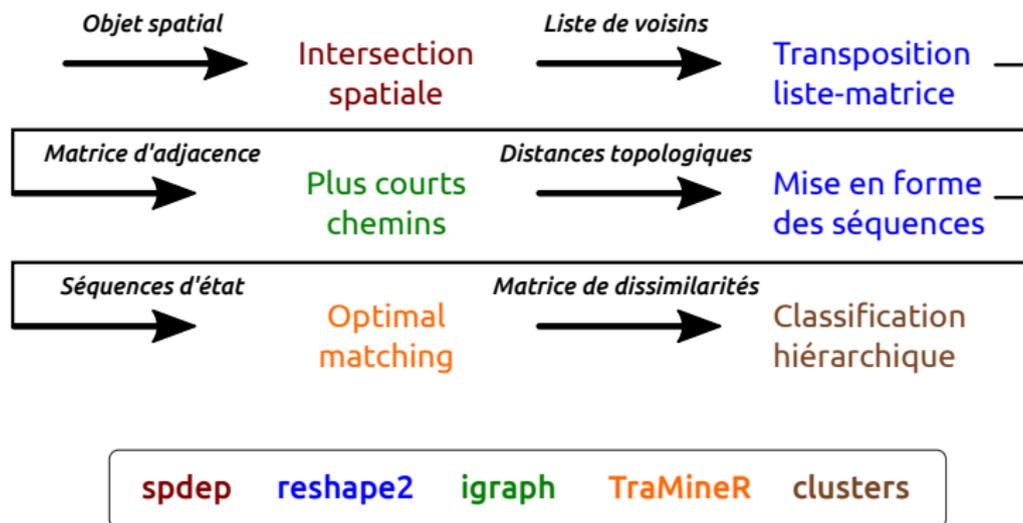
**Groupe 4 (n = 4753)**



# Le flux de travail avant R



# Le flux de travail avec R



# Le flux de travail avec R

Quels bénéfices ?

Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrngh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#&\$@*&!!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohool!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

Type: Ph.D Thesis Modified: too many times Copyright: Jorge Cham www.phdcomics.com

# Le flux de travail avec R

## Quels bénéfices ?

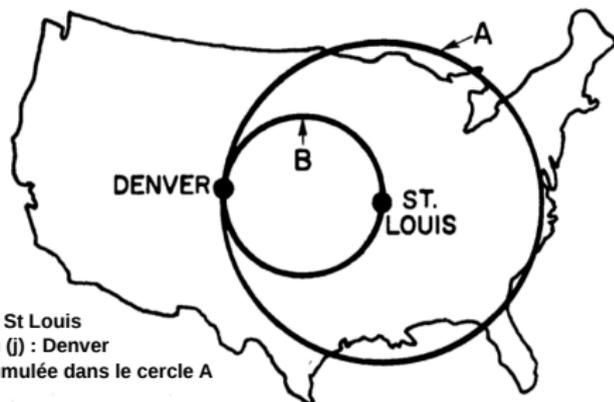
- **Mémoire** : trace complète des traitements effectués (pas de journal).
- **Échange** : les traitements sont échangeables (équipe, aide).
- **Reproductibilité** : les traitements peuvent être reproduits sur n'importe quelles données de même structure.
- **Reproductibilité encore** : le résultat est vérifiable.

# Et pour l'analyse exploratoire ?

Commenges H. (2016) "Modèle de radiation et modèle gravitaire ..."

$$T_{ij} = k \frac{P_i P_j}{D_{ij}^2}$$

$$T_{ij} = T_i \frac{P_i P_j}{(P_i + S_{ij})(P_i + P_j + S_{ij})}$$

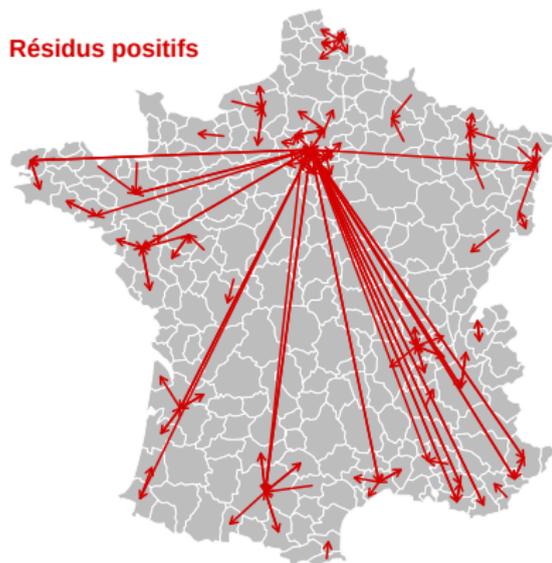


Origine (i) : St Louis  
Destination (j) : Denver  
 $S_{ij}$  : pop. cumulée dans le cercle A

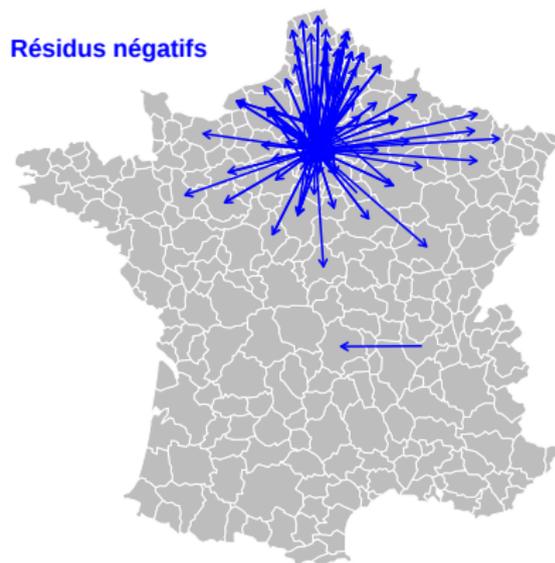
Source : Stouffer S.A. (1960) "Intervening opportunities and competing migrants", *J. Reg. Sc.*

# Et pour l'analyse exploratoire ?

Mobilité résidentielle - Zones d'emploi - **Gravitaire** (Poisson)



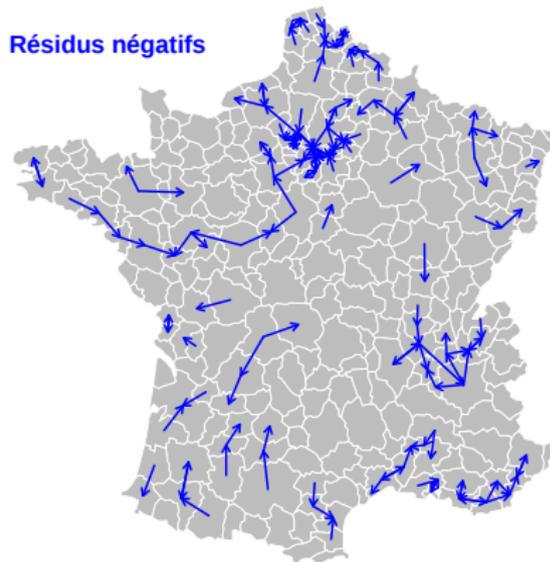
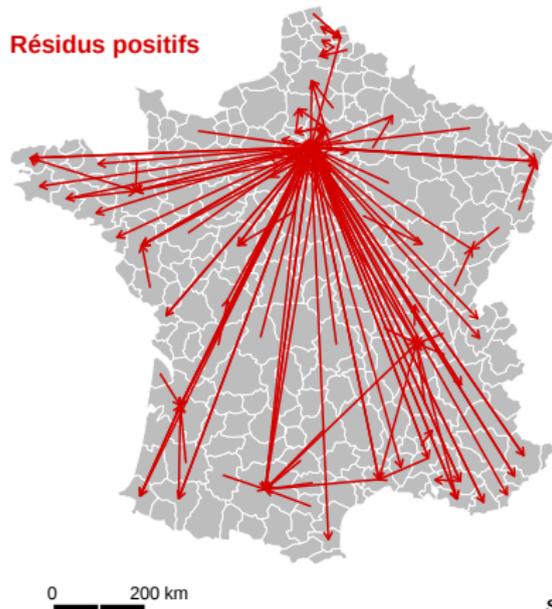
0 200 km



Source : Insee, Zones d'emploi 2010 & Insee, RP, Mobilités résidentielles 2010

# Et pour l'analyse exploratoire ?

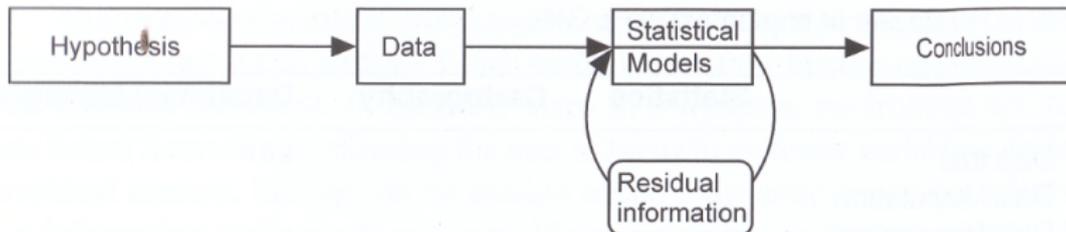
## Mobilité résidentielle - Zones d'emploi - **Radiation**



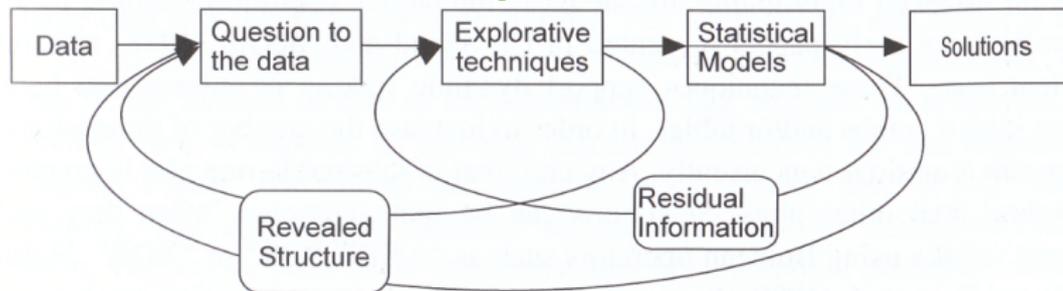
Source : Insee, Zones d'emploi 2010 & Insee, RP, Mobilités résidentielles 2010

# Et pour l'analyse exploratoire ?

## Flux de travail déductif

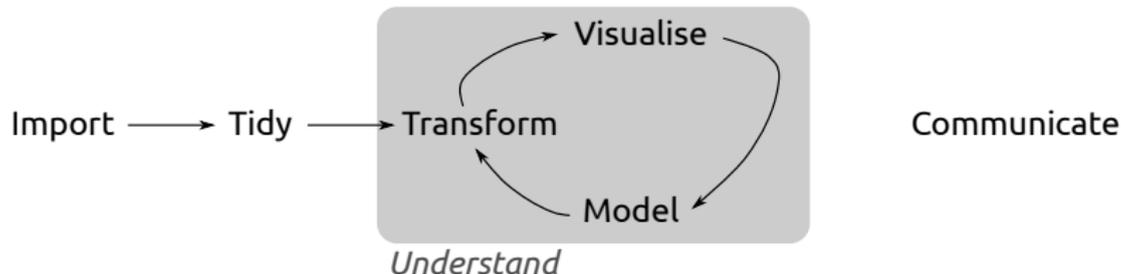


## Flux de travail abductif

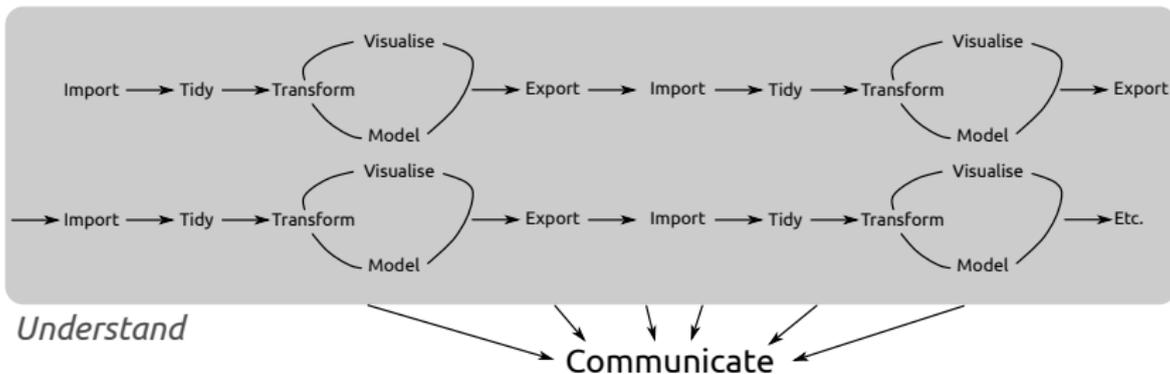


# Et pour l'analyse exploratoire ?

## EDA d'après Hadley Wickham



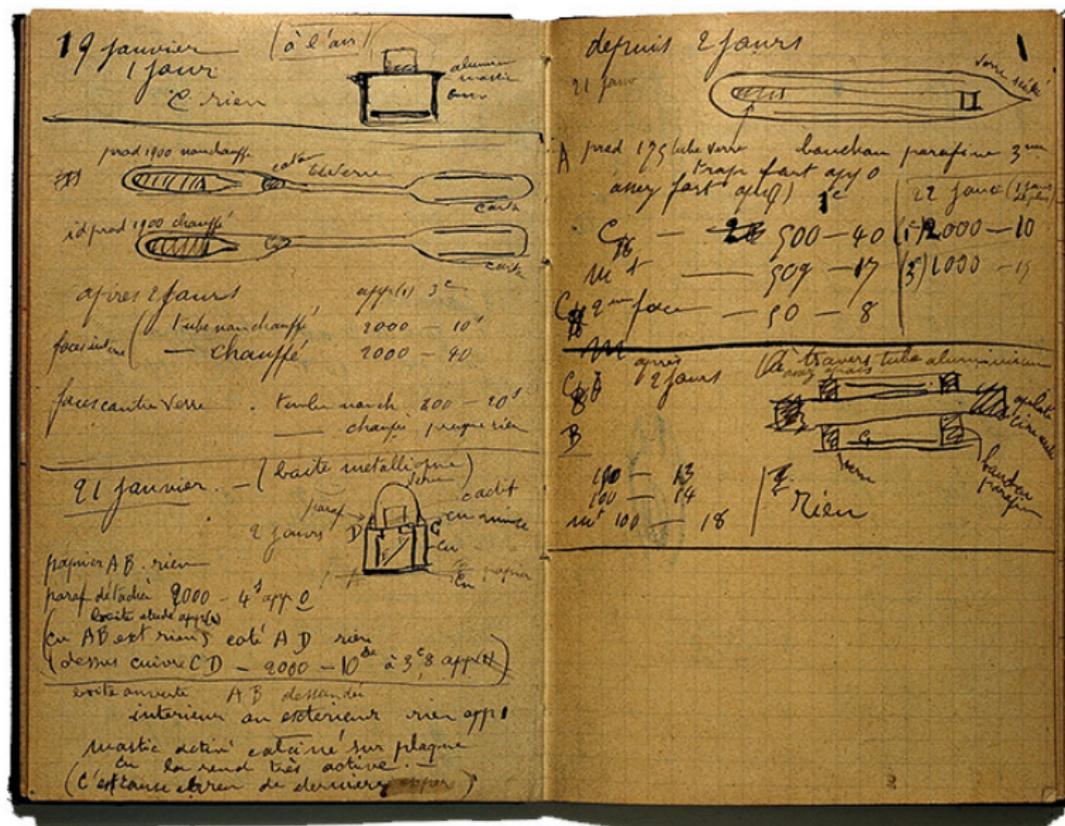
## EDA en géographie avant R



(2)

## Articulation et suivi dans un flux de travail collectif

# Le cahier de laboratoire



# Le cahier de laboratoire

Année : \_\_\_\_\_

Numéro : \_\_\_\_\_

Cahier de laboratoire

**Cahier de laboratoire**  
Laboratory notebook

Numéro : \_\_\_\_\_  
Number: \_\_\_\_\_

Unité de recherche : \_\_\_\_\_  
Research unit: \_\_\_\_\_

Prénom et nom du ou des utilisateur(s) : \_\_\_\_\_  
First name and last name of the user(s): \_\_\_\_\_

Signature(s) : \_\_\_\_\_  
Signature(s): \_\_\_\_\_

Date de début : \_\_\_\_\_ Date de fin : \_\_\_\_\_  
Start date: \_\_\_\_\_ End date: \_\_\_\_\_

 Avec la collaboration de l'INPI

 Institut national de recherche en informatique et automatique

# Le cahier de laboratoire

- Le CdL, élaboré par le Ministère de l'ESR, en collaboration avec l'Institut National de la Propriété Industrielle (INPI)
- Sur le site de l'INPI [<https://www.inpi.fr>] : le CdL se trouve dans la section "Protéger vos innovations"

Pourquoi utiliser ce cahier de laboratoire ?

- **Garantir** la traçabilité des résultats de recherche : identification de la date et de l'auteur des travaux
- **Capitaliser** le savoir-faire du laboratoire et faciliter la transmission des connaissances
- **S'adapter** aux exigences de la recherche internationale
- **Professionaliser** les pratiques liées à la recherche
- **Accompagner** une démarche qualité

# Dynamic reporting

**Avec le *package* `knitr` il est devenu très facile de produire des documents dynamiques avec :**

- Interprétation du code R inséré dans des *chunks*
- Interprétation du texte en Markdown ou Latex

Les documents dynamiques ont une vaste gamme d'applications (reproductibilité, transfert, supports de formation).

# Le cahier de laboratoire

Exemple du LabNotebook au sein du Boettiger Group (UC Berkeley) :  
<http://www.carlboettiger.info>

## Coding

- cboettig pushed to master at [cboettig/cboettig.github.io](https://github.com/cboettig/cboettig): *Site updated from circleci* 03:04 2016/10/22
- cboettig commented on issue [codemeta/codemeta#90](https://github.com/codemeta/codemeta#90): *@mbjones Correct me if I'm wrong, but didn't we discuss including a namespace (or maybe that's the wrong term for it) for the identifie.* 10:52 2016/10/21
- cboettig pushed to master at [cboettig/cboettig.github.io](https://github.com/cboettig/cboettig): *Site updated from circleci* 08:46 2016/10/21
- cboettin pushed to source at

## Discussing

- that time when @\_inundata and I end up in Berkeley photo studio on a steep slope lit t projected #rstats code... <https://t.co/KKApSXDkiW> 05:01 2016/10/21
- RT @pisastero: 5 jobs open @nceas: Dep Director, Sci Programmer, Sci web developer, Training & Outreach Coord, Projects Data Coord [https://...](https://t.co/045720161021) 04:57 2016/10/21
- Is science only for the rich? : Nature News & Comment <https://t.co/wsINXcaPt> 08:09 2016/10/20

## Reading

- A survey of POMDP solution techniques: Environment (2000). Volume: 2, Issue: September. Pages: X3. K P Murphy et al. 08:04 2016/07/25
- Advances in marine conservation: the role of marine protected areas.: Trends in ecology & evolution (1994). Volume: 9, Issue: 7. Pages: 267-70. M Tundi Agardy, I Tundi Agardy et al. 08:04 2016/07/25
- Adaptive optimization and the harvest of biological populations: Mathematical Biosciences (1996). Volume: 136, Issue: 1 Pages: 1-20. Avron K. Williams et al. 08:04

## Recent Entries

- Reflecting on five years of the open lab notebook: 2015/12/31
- Docker Workflows: 2015/12/17
- Download and Parse Github Issues: 2015/11/20
- Nimble Model Construction: 2015/08/27
- a draft repository for rOpenSci (draft for ropensci blog): 2015/08/03

## Notebooks

- 2015 Notebook
- 2014 Notebook
- 2013 Notebook
- 2012 Notebook
- 2011 Notebook
- 2010 Notebook
- 2009 Notebook
- 2008 Notebook

## RSS feeds

- All posts
- Blog posts only
- R posts

# Le cahier de laboratoire

Exemple du LabNotebook au sein du Boettiger Group (UC Berkeley) :  
<http://www.carlboettiger.info>

## Mdptoolbox Allen Model

29 Jul 2015

```
library("MDPtoolbox", quietly = TRUE)
library("ggplot2", quietly = TRUE)
```

```
K <- 150 # state space limit
states <- 0:K # Vector of all possible states
actions <- states # Vector
```

[Read more](#)

## Fastgp Explore

29 Jul 2015

```
library(FastGP)
library(mvtnorm)
library(MASS)
library(rbenchmark)
```

### Demo of elliptical slice sampling

Relevant Parameters:

```
A <- 1 #amplitude of the si
```

[Read more](#)

## Mdptoolbox Ex 2

14 Jul 2015

Adapted from Marescot et al. appendix 5, to Reed optimal control problem, including direct comparison against (semi) analytic optimum.

### step 1: define objectives

This is a conceptual step which does not require coding

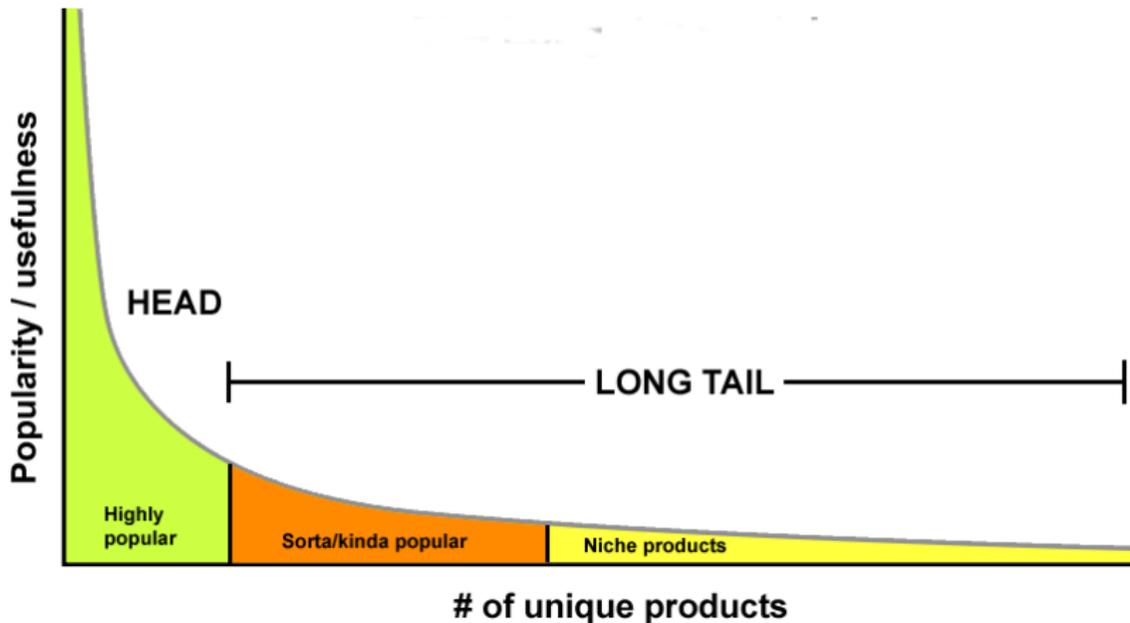
### step 2: define states

[Read more](#)

(3)

Production d'outils labo  
*longue traîne*

# Outils labo *longue traîne*



# Outils labo *longue traîne*

Le modèle de labo technicien-ingénieur-chercheur avec spécialisation et hiérarchie est souvent inadapté. Avec le framework R+shiny :

- Figure du développeur-utilisateur : ni spécialisé ni hiérarchisé
- Développement presque instantané
- Vaste palettes d'applications (recherche et enseignement)

**Outils labo Géographie-cités** : <http://shiny.parisgeo.cnrs.fr>

- Exemple de SLIDER : <http://cybergegeo.revues.org/26530>
- Exemple d'ExaminatR

**Galleries d'applications shiny** :

- RStudio (avec code) : <http://shiny.rstudio.com/gallery>
- ShowMeShiny : <http://www.showmeshiny.com>