

R + LATEX

Une très brève introduction

Objectifs & Philosophie

Objectifs & Philosophie

- ▶ Installation: RStudio + R + L^AT_EX

Objectifs & Philosophie

- ▶ Installation: RStudio + \mathbb{R} + \LaTeX
- ▶ Résumé de la journée: \mathbb{R} et \LaTeX
 1. Le monde de l'open source
 2. Quelques bases de programmation \mathbb{R}
 3. Faire des graphiques en \mathbb{R} avec \LaTeX

Objectifs & Philosophie

- ▶ Installation: RStudio + \mathbb{R} + \LaTeX
- ▶ Résumé de la journée: \mathbb{R} et \LaTeX
 1. Le monde de l'open source
 2. Quelques bases de programmation \mathbb{R}
 3. Faire des graphiques en \mathbb{R} avec \LaTeX
- ▶ Philosophie: Les outils *avant* la méthode

Installation

2 Languages, 1 Software



Installation

2 Languages, 1 Software



1. R : www.cran.rstudio.com

Installation

2 Languages, 1 Software



1. R : www.cran.rstudio.com
2. L^AT_EX: www.latex-project.org/get/

Installation

2 Languages, 1 Software



1. R : www.cran.rstudio.com
2. L^AT_EX: www.latex-project.org/get/
3. RStudio: www.rstudio.com/products/rstudio/download/

Installation

2 Languages, 1 Software

The screenshot shows the RStudio interface with the following details:

- Project:** (None)
- Environment:** Shows "Environment is empty".
- Files:** Displays several files: JfMorin.tex*, Class3.tex, beamerthemeCement_WorkshopR.sty, CreateMaps.R, Class2.tex, and mapWorld.
- R Script:** The main pane contains R code for mapping:

```
## MAP THE ADDITIVE SCALE
library(RColorBrewer)
library(mapprojtools)
library(ggplot2)
library(rworldmap)

# Replace Badly coded countries in df
Data$region[Data$region == "Russian Federation"] <- "Russia"
Data$region[Data$region == "United Kingdom"] <- "UK"
Data$region[Data$region == "Congo"] <- "Republic of Congo"
Data$region[Data$region == "DR Congo"] <- "Democratic Republic of the Congo"

# 
mapWorld = map_data(map='world')
mapWorld = merge(Data, mapWorld, by='region', all.y=TRUE)
mapWorld = mapWorld[order(mapWorld$order), ] # <---

# Fix missing code in World
mapWorld$region[mapWorld$region == "Denmark"] <- "Greenland"

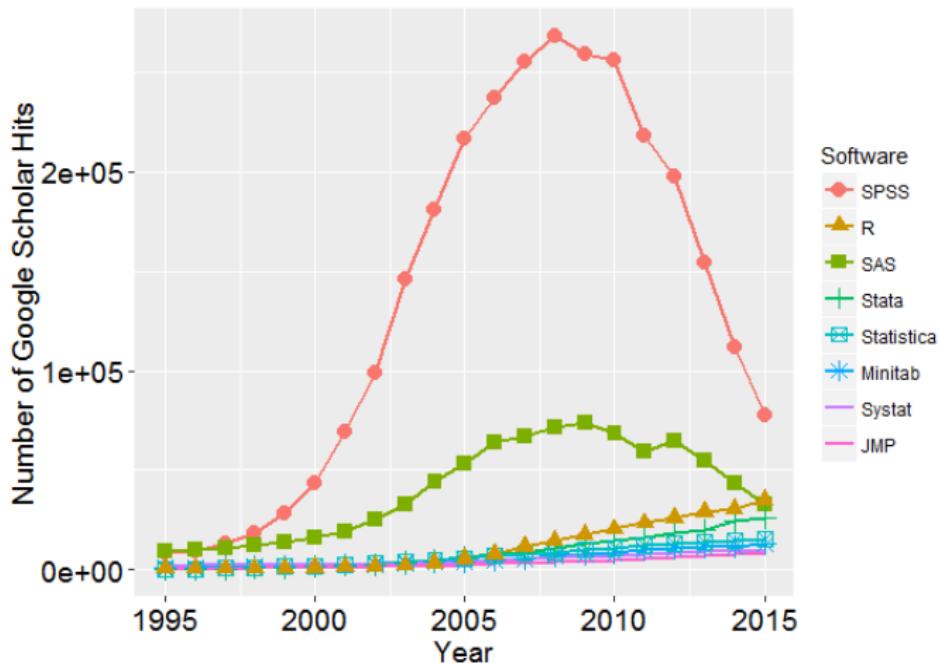
96:53  (Untitled) ±
```
- Console:** Shows the command `./CreateMaps.R` being run, with the output:

```
>
>
>
>
>
>
>
>
> 1+1
[1] 2
> |
```

Le monde de l'Open Source

Why R ?

Why R ?



Pourquoi R ? Les raisons de l'aimer

1. Gratisssss
2. Disponible pour tous les systèmes d'exploitation
3. Graphiques + L^AT_EX
4. Popularité + Packages
5. *Open source*: Développer par et pour les chercheurs

Pourquoi R ? Les raisons de détester

1. Programmer du code = Courbe d'apprentissage raide
2. Développement éclectique. Par moment... chaotique

Why L^AT_EX?

Why L^AT_EX? Reasons to Love

- ▶ Bibliography: BIBT_EX
- ▶ Table of content, tables, etc.
- ▶ Deals automatically with stuff like tables, figures, etc.
- ▶ Pretty templates
- ▶ Code + *Open source* = A large expert community on the web

Why L^AT_EX? Reasons to Hate

- ▶ Tough to learn... I mean very tough. But the basics are easy
- ▶ Incompatible with MS Word
- ▶ ~~No spell checker~~
- ▶ No “Change trackers” and stuff like that
- ▶ Final document available only after code compilation
- ▶ Some journals do not accept L^AT_EX submission... others strongly encourage it

L^AT_EX: A Nice Table

Table 1. Length of Bananas and Apples

Quantile	Bananas	Apples
0%	59	44
50%	69	64
100%	77	71

L^AT_EX: The Nice Table Code

```
\begin{table}
  \centering
  \caption{Length of Bananas and Apples}
  \begin{tabular}{lrr}
    Quantile & Bananas & Apples\\ \hline
    0\%      & 59      & 44 \\
    50\%     & 69      & 64 \\
    100\%    & 77      & 71 \\
  \end{tabular}
  \label{tab:bananasapples}
\end{table}
```

L^AT_EX: It's a joke, right?!



LATEX: Nope



LATEX

Tableau 1: Tests des hypothèses

	Vote pour le NPD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Évaluation du chef NPD					3.87*** (0.22)	3.81*** (0.24)	3.17*** (0.52)
Droite idéologique			-2.86*** (0.46)	-3.24*** (0.53)			-2.66** (0.57)
Québec	0.69*** (0.09)	0.61*** (0.16)		0.92** (0.34)		0.56** (0.17)	0.93** (0.35)
Femme		0.05 (0.09)		-0.08 (0.19)		-0.03 (0.10)	-0.08 (0.20)
Francophone		-0.02 (0.17)		-0.37 (0.35)		-0.29 (0.18)	-0.63 (0.37)
allophone		-0.17 (0.15)		-0.38 (0.34)		-0.18 (0.17)	-0.22 (0.36)
Moins de 34 ans		-0.03 (0.15)		-0.17 (0.34)		-0.13 (0.16)	-0.26 (0.36)
Plus de 55 ans		-0.23* (0.10)		-0.33 (0.21)		-0.24* (0.11)	-0.23 (0.22)
Haut revenu		-0.33** (0.12)		-0.36 (0.24)		-0.30* (0.13)	-0.32 (0.25)
Faible revenu		0.30* (0.15)		0.33 (0.31)		0.40* (0.17)	0.49 (0.33)
Pas de diplôme secondaire		-0.23 (0.15)		0.04 (0.36)		-0.12 (0.17)	0.03 (0.38)
Diplôme universitaire		0.13 (0.10)		-0.61** (0.21)		-0.12 (0.11)	-0.79** (0.22)
-constante	-1.05*** (0.05)	-0.86*** (0.11)	0.34 (0.20)	0.96** (0.35)	-3.17*** (0.15)	-2.95*** (0.19)	-1.21* (0.51)
N	2,745	2,464	655	610	2,636	2,381	602
Log Likelihood	-1,650.11	-1,487.30	-383.02	-346.16	-1,412.88	-1,276.31	-317.77
AIC	3,304.22	2,996.60	770.04	716.31	2,829.77	2,576.62	661.54

Source : Étude électorale canadienne, 2011.

Note : Régression logistique binomiale.

*p<0.05 ; **p<0.01 ; ***p<0.001

L^AT_EX: Code (Part 1)

<...>

```
1 % Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac
2 % Date and time: Wed, Jan 07, 2015 - 22:20:00
3 \begin{table} [!htbp] \centering
4 \caption{Tests des hypothèses}
5 \label{}
6 \scriptsize
7 \begin{tabular}{@{\extracolsep{5pt}}lcccccc}
8 \\[-1.8ex]\hline \\[-1.8ex]
9 & \multicolumn{7}{c}{Vote pour le NPD} \\
10 & (1) & (2) & (3) & (4) & (5) & (6) & (7) \\
11 \hline \\[-1.8ex]
12 & Évaluation du chef NPD & & & & 3.87^{***} & 3.81^{***} & 3.17^{***} \\
13 & & & & & (0.22) & (0.24) & (0.52) \\
14 & Droite idéologique & & & & $-2.86^{***}$ & $-3.24^{***}$ & $-2.66^{***}$ \\
15 & & & & & (0.46) & (0.53) & (0.57) \\
16 & Québec & 0.69^{***} & 0.61^{***} & 0.92^{**} & 0.56^{**} & 0.93^{**} \\
17 & & (0.09) & (0.16) & (0.34) & (0.17) & (0.35) \\
18 & Femme & 0.05 & $-0.08$ & $-0.03$ & $-0.08$ \\
19 & & (0.09) & (0.19) & (0.10) & (0.20) \\
20 & Francophone & $-0.02$ & $-0.37$ & $-0.29$ & $-0.63$ \\
21 & & (0.17) & (0.35) & (0.18) & (0.37) \\
22 & Allophone & $-0.17$ & $-0.38$ & $-0.18$ & $-0.22$ \\
23 & & (0.15) & (0.34) & (0.17) & (0.36) \\
```

LATEX: Code (Part 2)

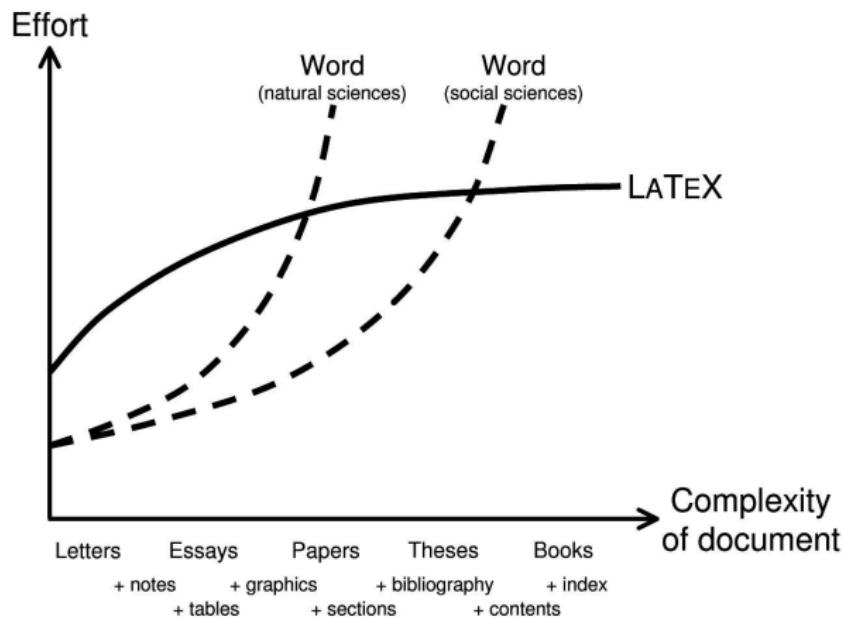
<...>

```
1 Moins de 34 ans & & $-$0.03 & & $-$0.17 & & $-$0.13 & $-$0.26 \\
2 & & (0.15) & & (0.34) & & (0.16) & (0.36) \\
3 Plus de 55 ans & & $-$0.23$^{*} \$ & & $-$0.33 & & $-$0.24$^{*} \$ & $-$0.23 \\
4 & & (0.10) & & (0.21) & & (0.11) & (0.22) \\
5 Haut revenu & & $-$0.33$^{**} \$ & & $-$0.36 & & $-$0.30$^{*} \$ & $-$0.32 \\
6 & & (0.12) & & (0.24) & & (0.13) & (0.25) \\
7 Faible revenu & & 0.30$^{*} \$ & & 0.33 & & 0.40$^{*} \$ & 0.49 \\
8 & & (0.15) & & (0.31) & & (0.17) & (0.33) \\
9 Pas de diplôme secondaire & & $-$0.23 & & 0.04 & & $-$0.12 & 0.03 \\
10 & & (0.15) & & (0.36) & & (0.17) & (0.38) \\
11 Diplôme universitaire & & 0.13 & & $-$0.61$^{*} \$ & & $-$0.12 & $-$0.79$^{***} \$ \\
12 & & (0.10) & & (0.21) & & (0.11) & (0.22) \\
13 \_constante & $-$1.05$^{***} \$ & & $-$0.86$^{***} \$ & 0.34 & 0.96$^{**} \$ & $-$3.17$^{***} \$ & $-$1.276 \\
14 & & (0.05) & & (0.11) & & (0.20) & (0.35) & (0.15) & (0.19) & (0.51) \\
15 N & 2,745 & 2,464 & 655 & 610 & 2,636 & 2,381 & 602 \\
16 Log Likelihood & $-$1,650.11 & $-$1,487.30 & $-$383.02 & $-$346.16 & $-$1,412.88 & $-$1.276 \\
17 AIC & 3,304.22 & 2,996.60 & 770.04 & 716.31 & 2,829.77 & 2,576.62 & 661.54 \\
18 \\hline \\[-1.8ex] \\
19 \\multicolumn{8}{l}{\\emph{Source}: Étude électorale canadienne, 2011.} \\
20 \\multicolumn{8}{l}{\\emph{Note}: Régression logistique binomiale.} \\
21 \\multicolumn{8}{l}{\$^{*} \$p\$ < \$0.05; \$^{**} \$p\$ < \$0.01; \$^{***} \$p\$ < \$0.001} \\
22 \\end{tabular} \\
23 \\end{table}
```

LATEX



LATEX



$\mathbb{R} + \text{\LaTeX}$

<...>

```
1 | stargazer(model1, model2, model3, model4, model5, model6, model7)
```

R + LATEX

Tableau 1: Tests des hypothèses

	Vote pour le NPD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Évaluation du chef NPD					3.87*** (0.22)	3.81*** (0.24)	3.17*** (0.52)
Droite idéologique			-2.86*** (0.46)	-3.24*** (0.53)			-2.66** (0.57)
Québec	0.69*** (0.09)	0.61*** (0.16)		0.92** (0.34)		0.56** (0.17)	0.93** (0.35)
Femme		0.05 (0.09)		-0.08 (0.19)		-0.03 (0.10)	-0.08 (0.20)
Francophone		-0.02 (0.17)		-0.37 (0.35)		-0.29 (0.18)	-0.63 (0.37)
allophone		-0.17 (0.15)		-0.38 (0.34)		-0.18 (0.17)	-0.22 (0.36)
Moins de 34 ans		-0.03 (0.15)		-0.17 (0.34)		-0.13 (0.16)	-0.26 (0.36)
Plus de 55 ans		-0.23* (0.10)		-0.33 (0.21)		-0.24* (0.11)	-0.23 (0.22)
Haut revenu		-0.33** (0.12)		-0.36 (0.24)		-0.30* (0.13)	-0.32 (0.25)
Faible revenu		0.30* (0.15)		0.33 (0.31)		0.40* (0.17)	0.49 (0.33)
Pas de diplôme secondaire		-0.23 (0.15)		0.04 (0.36)		-0.12 (0.17)	0.03 (0.38)
Diplôme universitaire		0.13 (0.10)		-0.61** (0.21)		-0.12 (0.11)	-0.79** (0.22)
-constante	-1.05*** (0.05)	-0.86*** (0.11)	0.34 (0.20)	0.96** (0.35)	-3.17*** (0.15)	-2.95*** (0.19)	-1.21* (0.51)
N	2,745	2,464	655	610	2,636	2,381	602
Log Likelihood	-1,650.11	-1,487.30	-383.02	-346.16	-1,412.88	-1,276.31	-317.77
AIC	3,304.22	2,996.60	770.04	716.31	2,829.77	2,576.62	661.54

Source : Étude électorale canadienne, 2011.

Note : Régression logistique binomiale.

*p<0.05 ; **p<0.01 ; ***p<0.001

R Programming Bases

R = Programming Language

- ▶ Calculation operators
- ▶ Assignment operators
- ▶ Logical operators
- ▶ Control instructions

Calculation Operators

- +
- -
- /
- %%

Logical Operators

- ==
- !=
- >=
- <=
- <
- >
- &
- |
- %in%

Control Instructions

- ▶ if... else
- ▶ for loop

R Data Structure

Data Structure

- ▶ Constants

Data Structure

- ▶ Constants
- ▶ Vectors

Data Structure

- ▶ Constants
- ▶ Vectors
- ▶ Data frames

Constants

<...>

```
1 variableString <- "Banana"  
2 variableNumerical <- 1492  
3 variableBoolean <- TRUE
```

Vecteurs



```
1 vecteurString <- c(variableString, "Apple", "Orange", "Sand Paper")
2 vecteurNumerical <- c(variableNumerical, 1604, 2011, 0328424)
3 vecteurBoolean <- c(variableBoolean, FALSE, TRUE, TRUE)
```

Data Frames

<...>

```
1 Data <- data.frame(vectorString, vectorNumerical, vectorBoolean, c(23,17,32,56))
```

Constantes

Vecteurs

Data frames

Constantes

Vecteurs

Data frames

1 

Constantes

1 

Vecteurs



Data frames

Constantes

1 

Vecteurs

1	
2	
3	
4	
5	
6	

Data frames

	v1	v2	v3	v4
1				
2				
3				
4				
5				
6				

aFruit <- “banana”

Constantes



Vecteurs

1	
2	
3	
4	
5	
6	

Data frames

	v1	v2	v3	v4
1				
2				
3				
4				
5				
6				

fruits[1] <- “banana”

Constantes



Vecteurs

1	A yellow banana icon inside a small square box.
2	
3	
4	
5	
6	

Data frames

	v1	v2	v3	v4
1				
2				
3				
4				
5				
6				

Data[1,1] <- “banana”

Constantes



Vecteurs

1	A yellow banana icon inside a small square box.
2	
3	
4	
5	
6	

Data frames

	v1	v2	v3	v4
1	A yellow banana icon inside a small square box.			
2				
3				
4				
5				
6				

Data\$fruits[1] <- “banana”

Constantes



Vecteurs

1	A yellow banana icon inside a small square box.
2	
3	
4	
5	
6	

Data frames

	v1	v2	v3	v4
1	A yellow banana icon inside a small square box.			
2				
3				
4				
5				
6				

Functions

R Base Functions

- ▶ `length()`
- ▶ `min()`
- ▶ `max()`
- ▶ `sum()`
- ▶ `median()`
- ▶ `mean()`

R Function: mean()

<...>

```
1 mean(yourVector)
```

Creating a Function in R

R Function: meanGirls()



R Function: meanGirls()

<...>

```
1 meanGirls <- function(Data){  
2     result <- sum(Data$age[Data$woman==1])/length(Data$age[Data$woman==1])  
3     return(result)  
4 }
```

R Function: meanGirlsPlus()

<...>

```
1 meanGirlsPlus <- function(Data, star=FALSE){  
2     if(star == FALSE){  
3         result <- sum(Data$age[Data$woman==1])/length(Data$age[Data$woman==1])  
4     } else {  
5         result <- sum(Data$age[Data$woman==1])/length(Data$age[Data$woman==1])  
6         result <- paste("*****", result, "*****")  
7     }  
8     return(result)  
9 }
```

Then?

Then? More R Functions...

```
meanGirls()  
meanGirlsPlus()  
meanBoys()  
meanBoysPlus()
```

Then? A \mathbb{R} Package

MeanSexPak

```
meanGirls()  
meanGirlsPlus()  
meanBoys()  
meanBoysPlus()
```

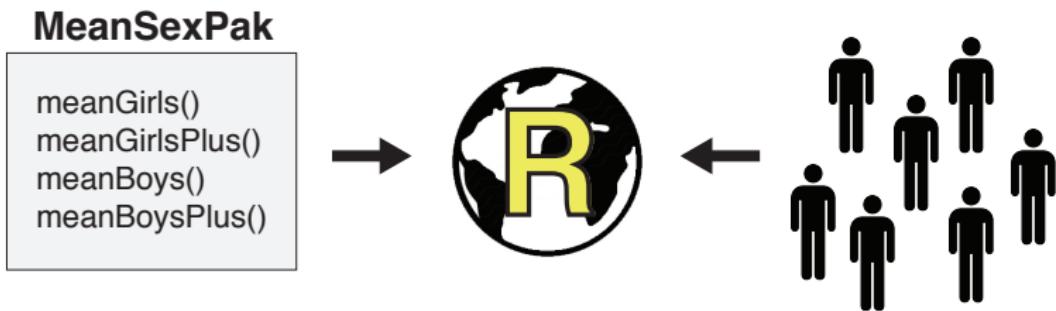
Then? Package Publication

MeanSexPak

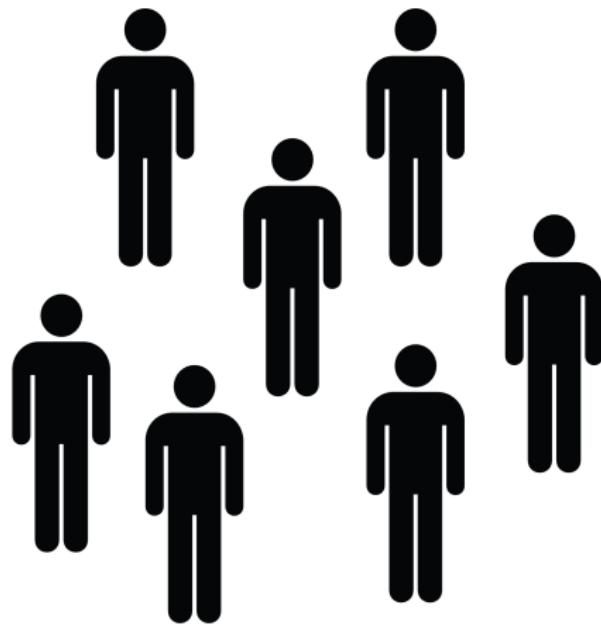
```
meanGirls()  
meanGirlsPlus()  
meanBoys()  
meanBoysPlus()
```



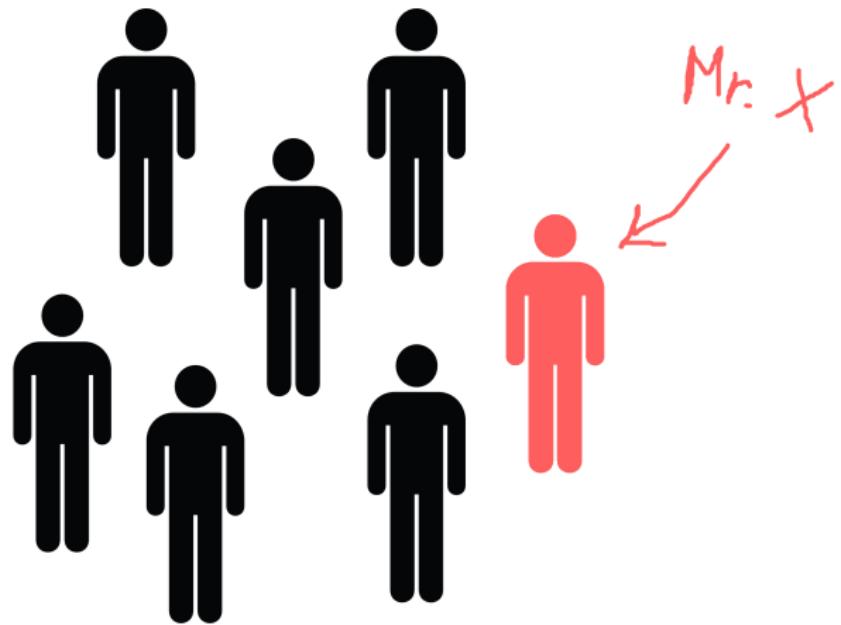
Then? Diffusion to the Community



R Community



Mr. X



Then? Mr. X Installs the Package

<...>

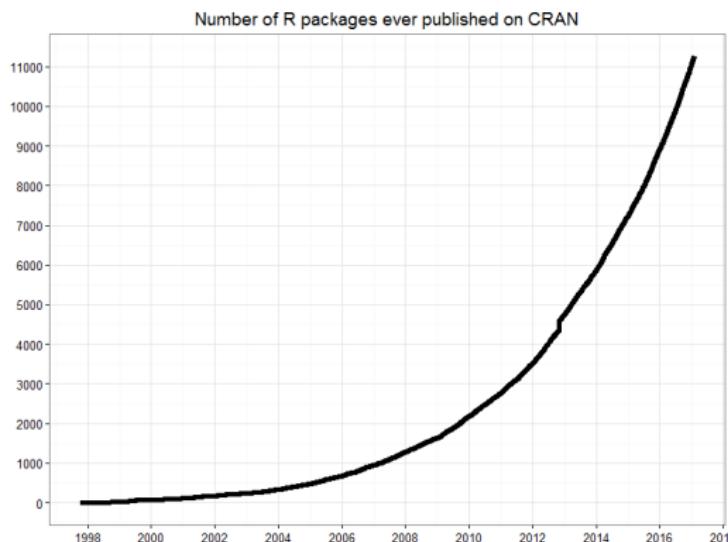
```
1 ]install.packages("MeanSexPak")
```

Ensuite? Mr. X Uses the Package

<...>

```
1 library(MeanSexPak)
2
3 # Calculate the mean age of the girls
4 girlsMeanAge <- meanGirls(MrXOwnData)
```

Number of R Packages



**Enough blabla...
Let's Code!**