

# DeCovarT, a generative model for the deconvolution of heterogeneous transcriptomic samples

#RR2023



17/06/2023

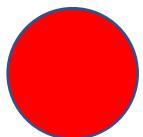
Bastien CHASSAGNOL  
Gregory NUEL, Etienne BECHT, Yufei LUO

**SERVIER**   
*moved by you*

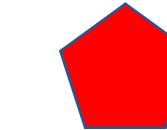
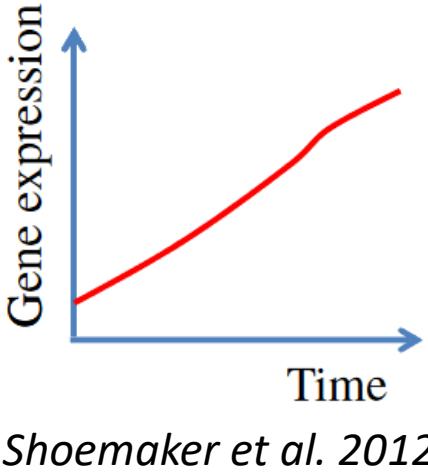
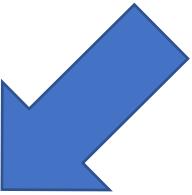
# Confusing biological noise



resting cell population 1

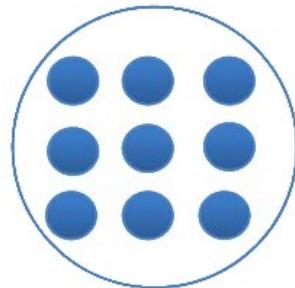


activated cell population 1

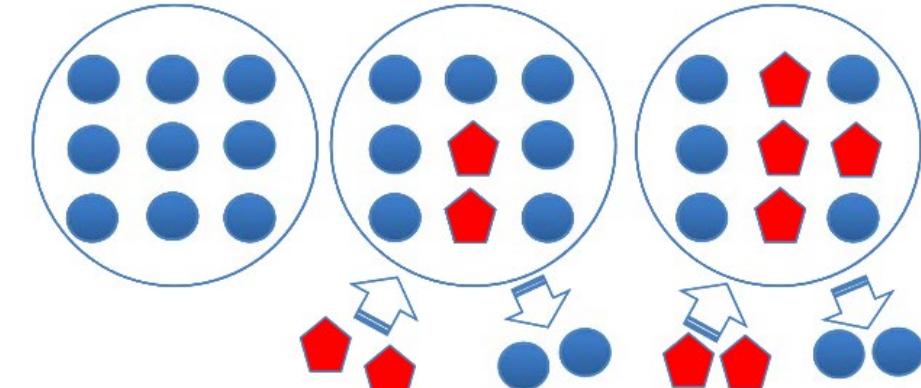


activated cell population 2

**B**



Time

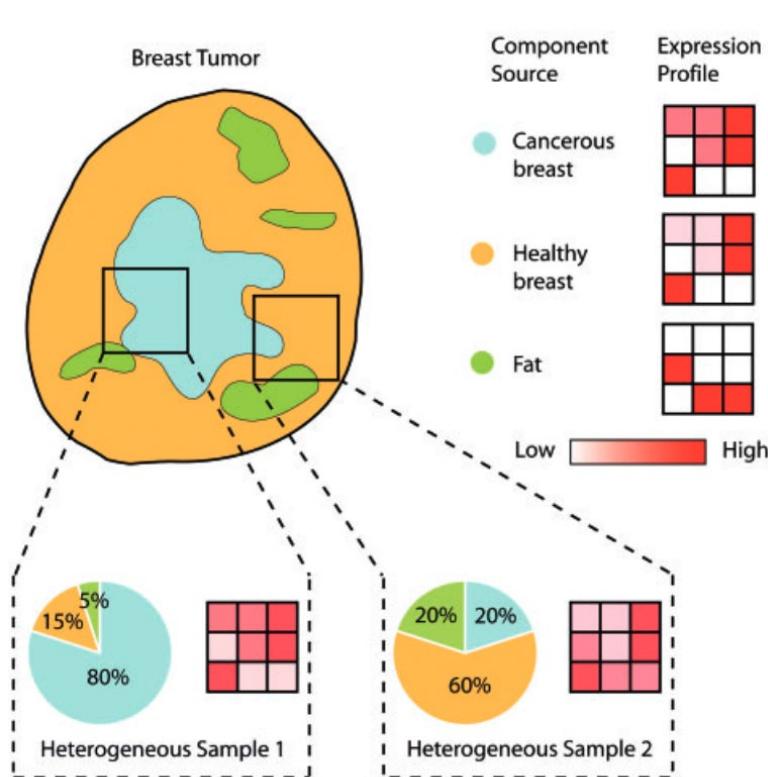


Time

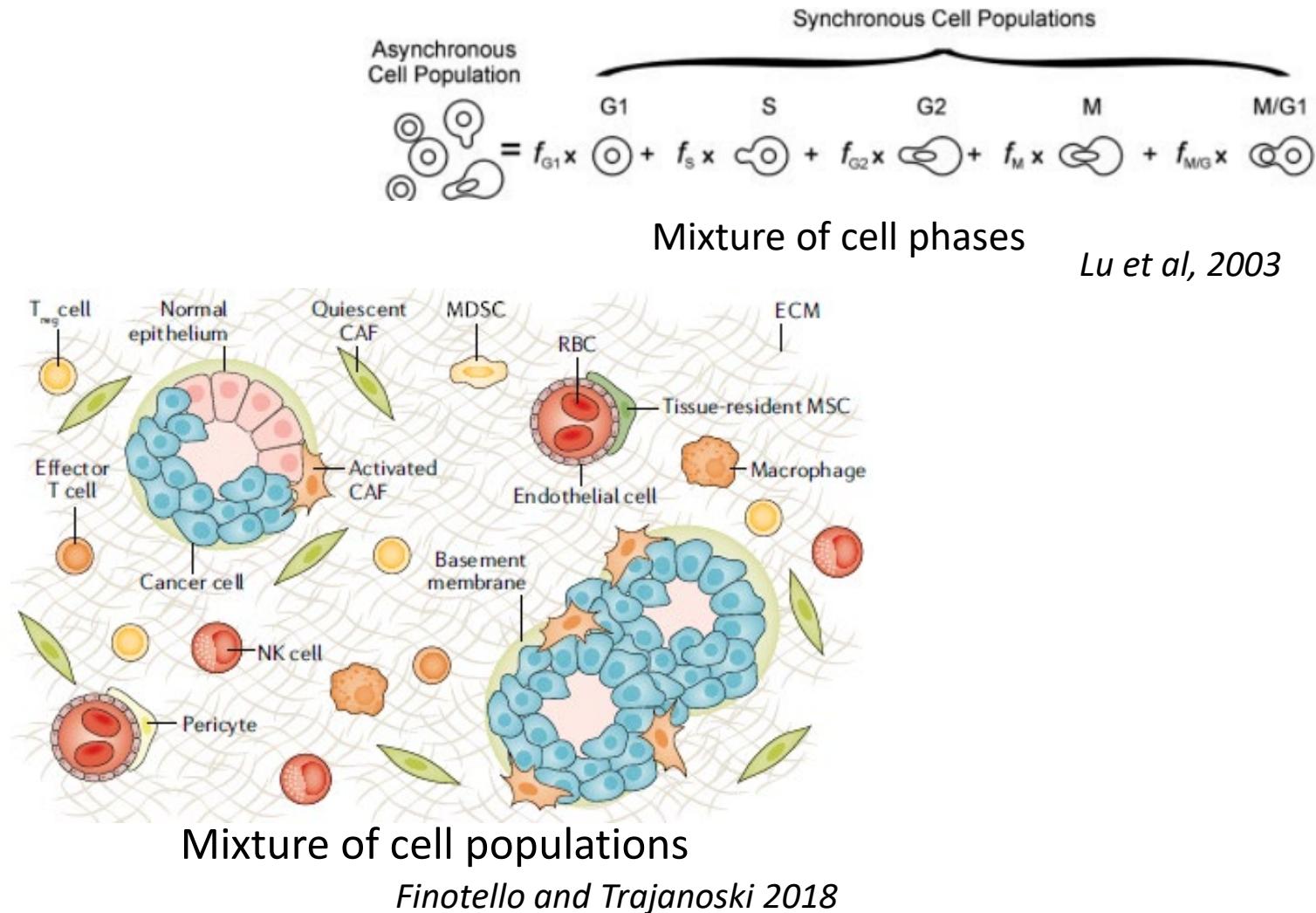
Scenario A: increase of the gene expression is generated by an **activation** of cell population 1

Scenario B: the gene expression increases due to the **arrival** of a **new** cell population 2

# Heterogeneity of tissues



Mixture of tissues  
Quon and Morris, 2009



# Deconvolution inputs

$$\begin{pmatrix} x_{1,1} & \dots & x_{1,J} \\ \vdots & \ddots & \vdots \\ x_{G,1} & \dots & x_{G,J} \end{pmatrix} \times \begin{pmatrix} p_{1,1} & \dots & p_{1,N} \\ \vdots & \ddots & \vdots \\ p_{J,1} & \dots & p_{J,N} \end{pmatrix} = \begin{pmatrix} y_{1,1} & \dots & y_{1,N} \\ \vdots & \ddots & \vdots \\ y_{G,1} & \dots & y_{G,N} \end{pmatrix}$$

$\mathbf{X}$  stores purified cellular expression profiles

$$\begin{pmatrix} x_{G_1,1} & \dots & 0 \\ 0 & x_{G_2,2} & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_{G_k,k} \end{pmatrix}$$

Marker-based

$\mathbf{p}$  is the individual vector of cell ratios

$\mathbf{Y}$  stores the resulting bulk expression values

$$\mathbf{X} \mathbf{p}_i = \mathbf{y}_i$$

$$\left\{ \begin{array}{l} \sum_{j=1}^J p_{ji} = 1 \\ \forall j \in \{1, \dots, J\}, \quad p_{ji} \geq 0 \end{array} \right.$$

# Deconvolution ecosystem

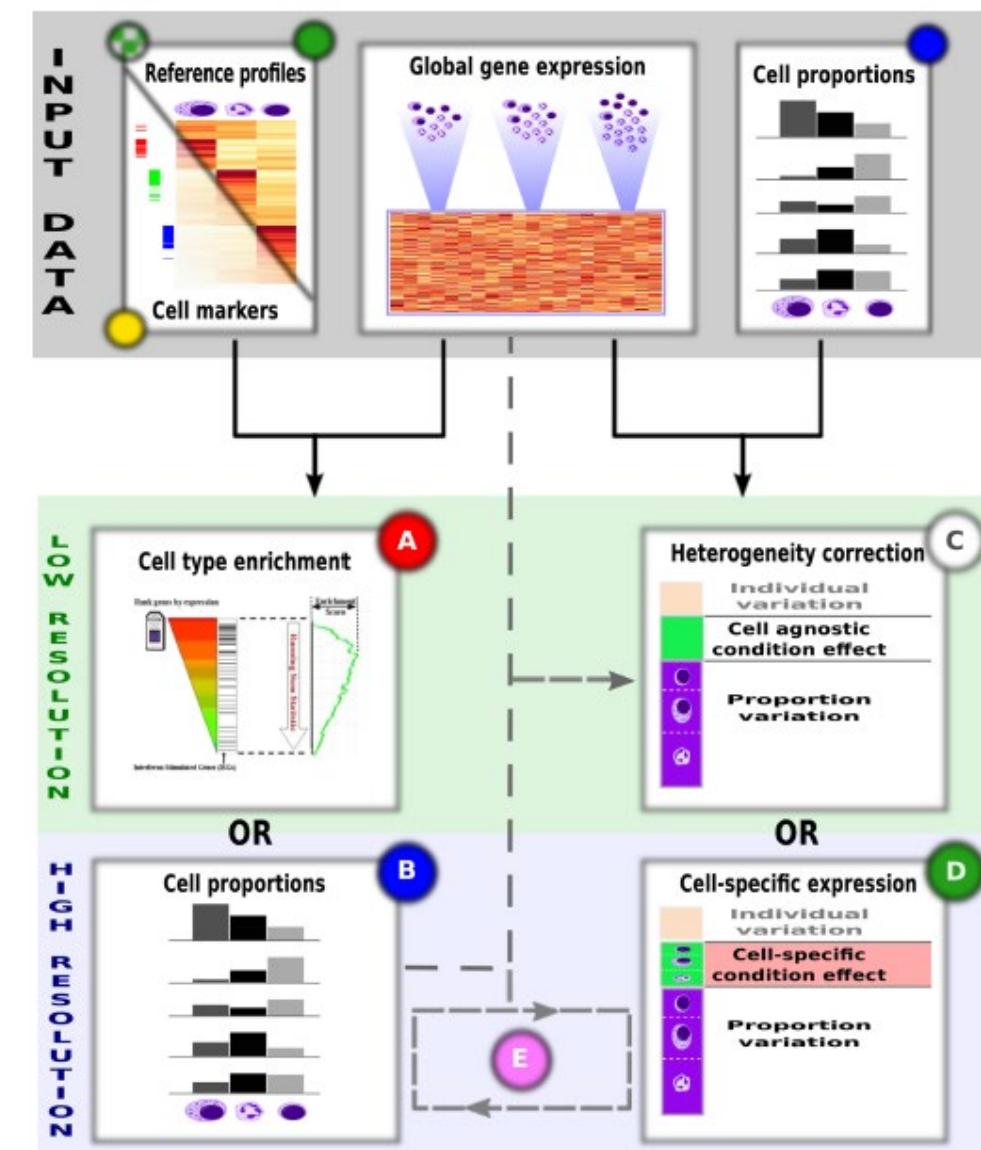
Partial deconvolution

- Estimate the ratios  $p$  for all individuals with the purified cell signature  $\mathbf{X}$  and bulk mixture  $\mathbf{y}$ .

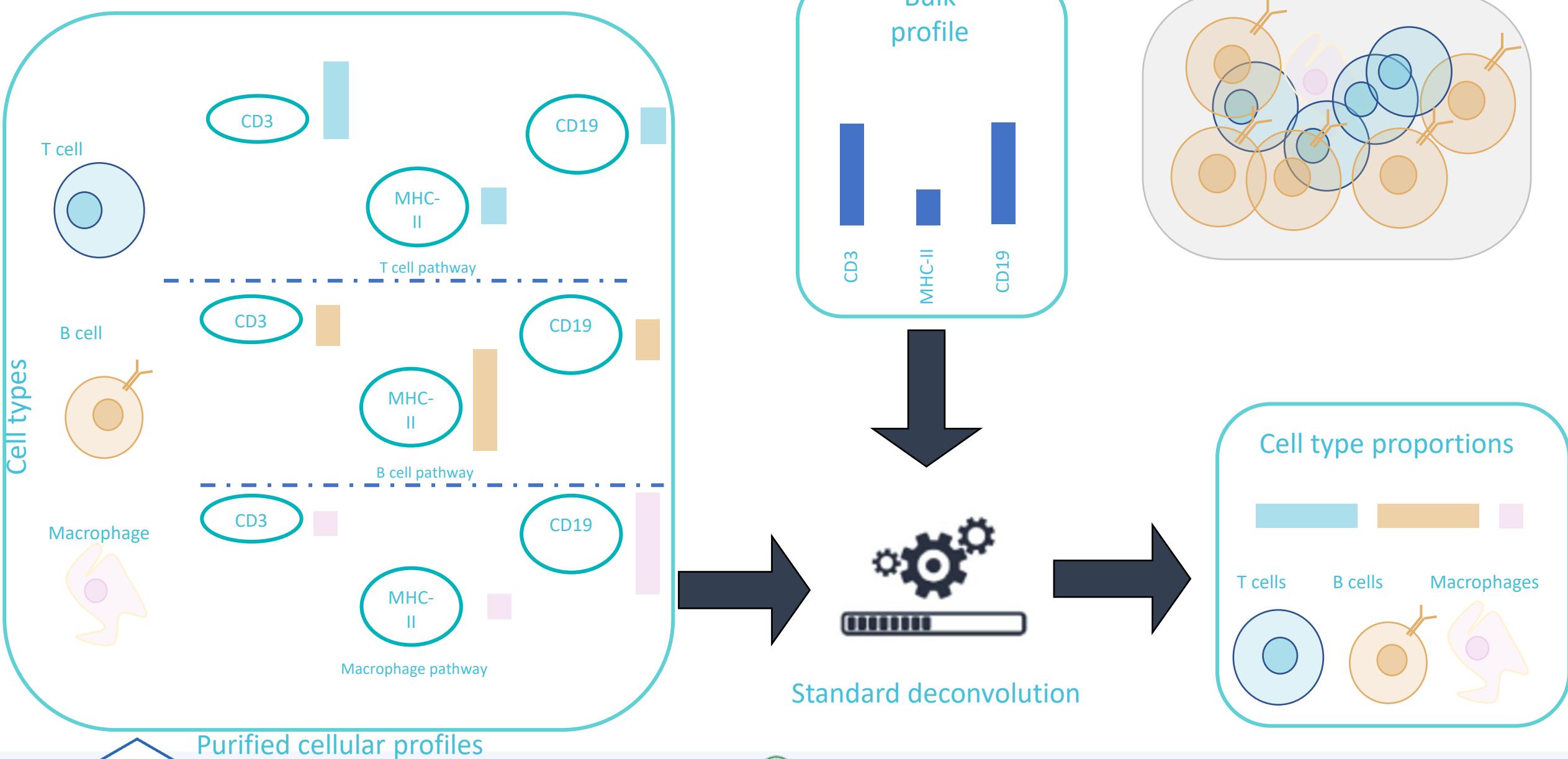
- Try to infer cell specific expression profiles  $\mathbf{X}$  based on  $\mathbf{p}$  and  $\mathbf{y}$ .

- Try to infer alternatively both  $\mathbf{p}$  and  $\mathbf{X}$  (unsupervised, reference-free methods).  
Undetermined problem without prior.

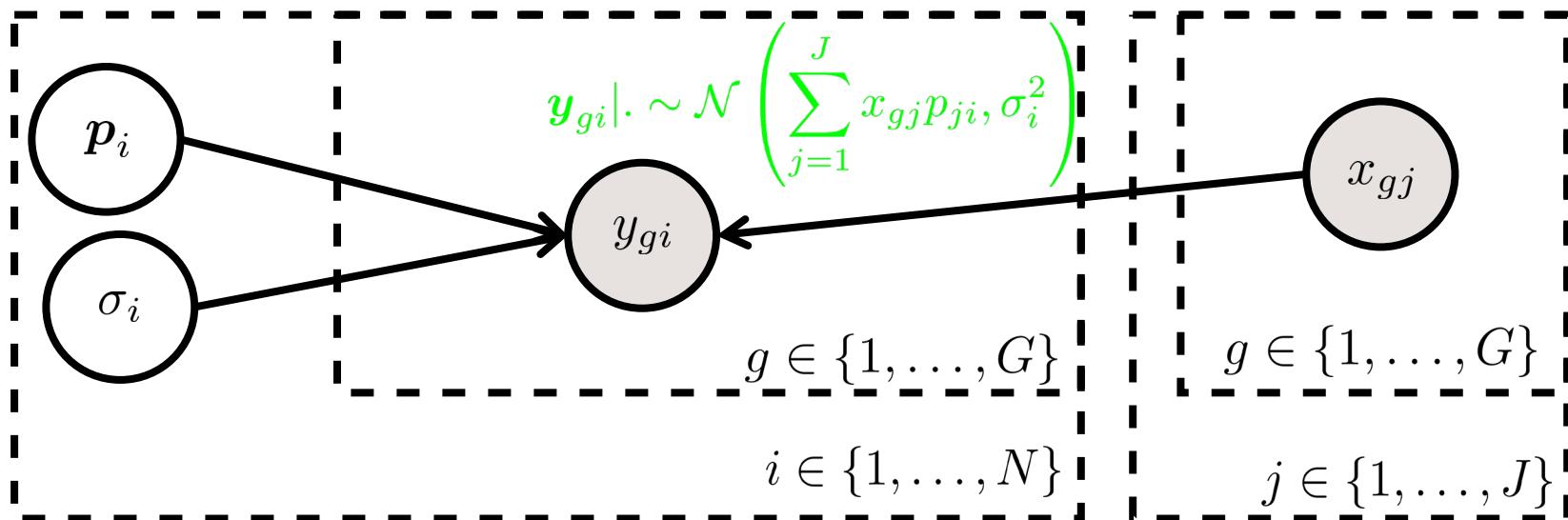
Complete deconvolution



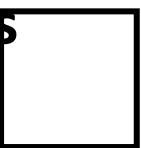
# Standard deconvolution framework



# Graphical model of linear regression



Nodes



Constant



Stochastic variable



Deterministic variable

Observations

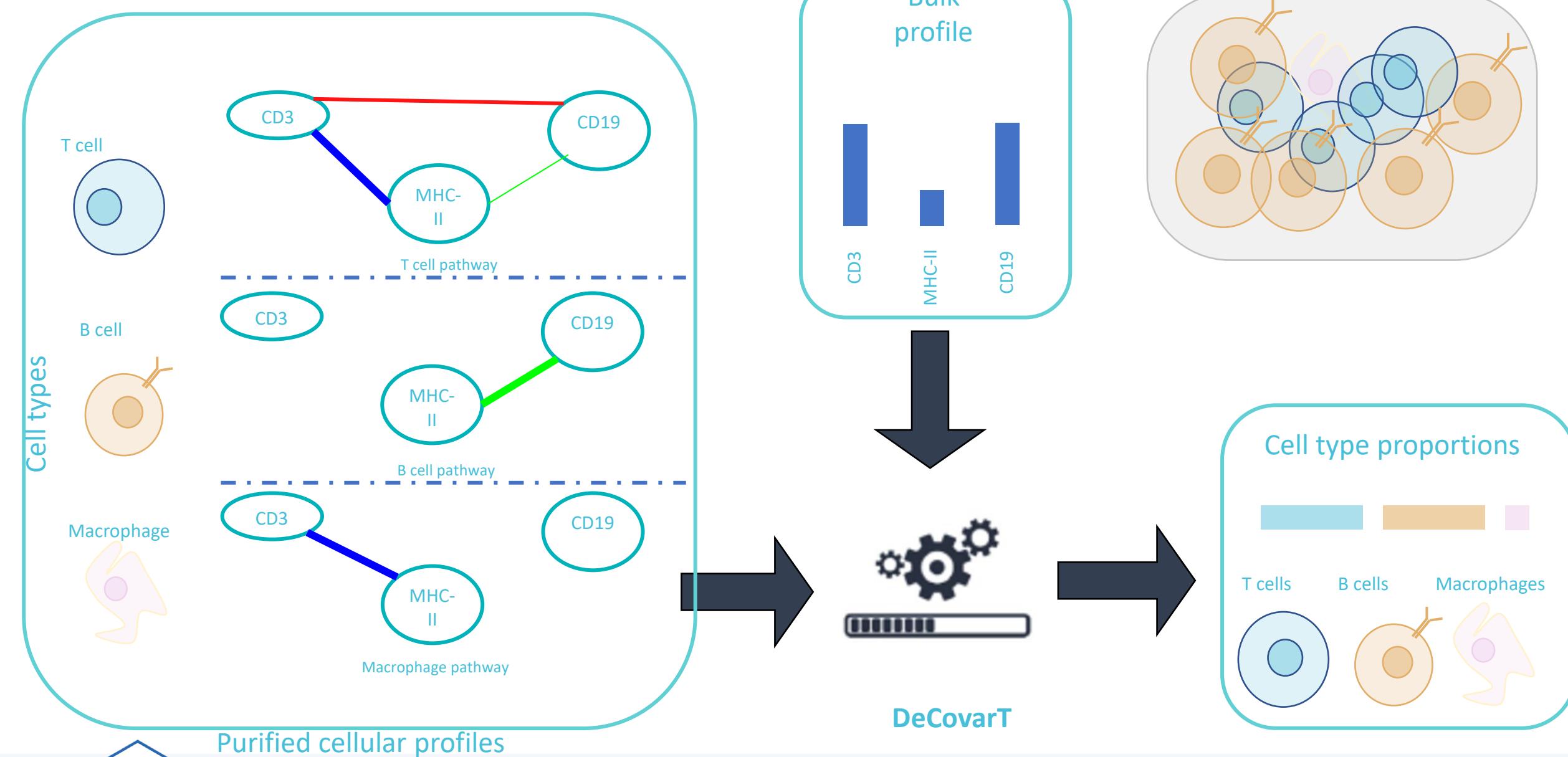
Parameters

Estimated parameters  
 $\theta = (\mathbf{p}, \sigma)$

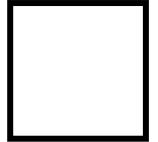
Distribution probabilities

Likelihood Laws  
 $f(\mathcal{D}|\theta)$

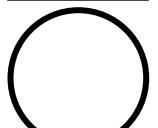
# DeCovarT framework



## Nodes



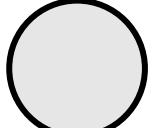
Constant



Stochastic variable



Deterministic variable



Observations

$$\boldsymbol{p}_i = \psi(\rho_i) = \begin{cases} p_{ji} = \frac{e^{\rho_{ji}}}{\sum_{j' < J} e^{\rho_{j'i}} + 1}, & j < J \\ p_J = \frac{1}{\sum_{k < J} e^{\rho_{jk}} + 1} \end{cases}$$

$$\boldsymbol{\Sigma}_i = \sum_{j=1}^J p_{ji}^2 \boldsymbol{\Sigma}_j$$

$$\mathbf{y}_i | \cdot \sim \mathcal{N}_G(\mathbf{X}\boldsymbol{p}_i, \boldsymbol{\Sigma}_i)$$

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{p}_i$$

$$i \in \{1, \dots, N\}$$

$$\rho_i$$

$$\mathbf{y}_i$$

$$\mathbf{x}_j \sim \mathcal{N}_G(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

$$\boldsymbol{\mu}_j$$

$$\boldsymbol{\Sigma}_j$$

$$j \in \{1, \dots, J\}$$

## Graphical model of DeCoVarT

### Parameters

Prior parameters

$$\xi = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Estimated parameters

$$\theta = (\boldsymbol{p}, \mathbf{X})$$

### Distribution probabilities

Prior laws

$$f(\theta|\xi)$$

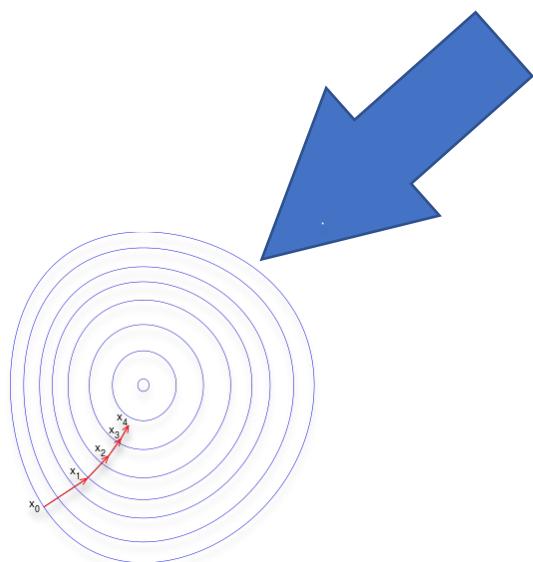
Likelihood Laws

$$f(\mathcal{D}|\theta)$$

# Optimisation algorithms in R

$$\ell_{\mathbf{y}|\mathbf{X},\Sigma}(\mathbf{p}) = C + \log \left( \det \left( \sum_{j=1}^J p_j^2 \Sigma_j \right)^{-1} \right) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\mathbf{p})^\top \left( \sum_{j=1}^J p_j^2 \Sigma_j \right)^{-1} (\mathbf{y} - \mathbf{X}\mathbf{p})$$

Log-likelihood associated to the model = log-likelihood of a multivariate Gaussian distribution



Steepest descent method

$$\theta^{(q+1)} = \theta^{(q)} + \gamma \nabla \ell(\theta^{(q)})$$



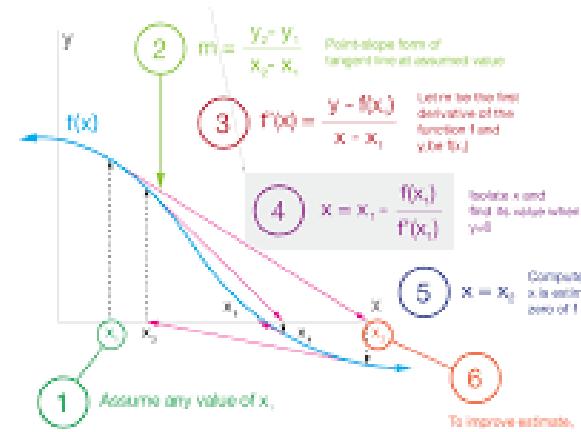
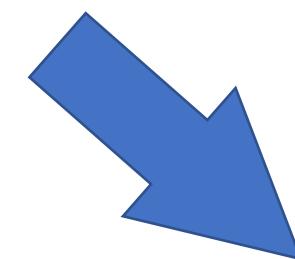
$$\theta^{(q+1)} = \theta^{(q)} + \gamma_q \frac{\nabla \ell(\theta^{(q)})}{\tilde{\mathbf{H}}_\ell(\theta^{(q)})}$$



Updated Marquardt-Levenberg,

in package [marqLevAlg](#)

BP Hejblum, 2021, R Journal



Newton Raphson method

$$\theta^{(q+1)} = \theta^{(q)} + \frac{\nabla \ell(\theta^{(q)})}{\mathbf{H}_\ell(\theta^{(q)})}$$

SERVIER \*

# Aider de nouveaux utilisateurs dans la création de graphiques

## {CGI} – Charte Graphique de l'Insee



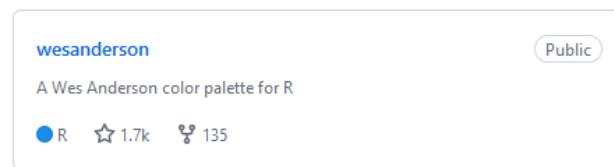
RENCONTRES R - JUIN 2023

- L'Insee a changé sa charte graphique en **2021**
  - La transition de **SAS** vers **R** (et **Python**) a été actée en **2022**
- Proposer une solution simple pour créer des graphiques respectant les contraintes de la charte sans devoir connaître toutes les subtilités de **ggplot2**

Deux principales inspirations pour l'architecture et les fonctionnalités :



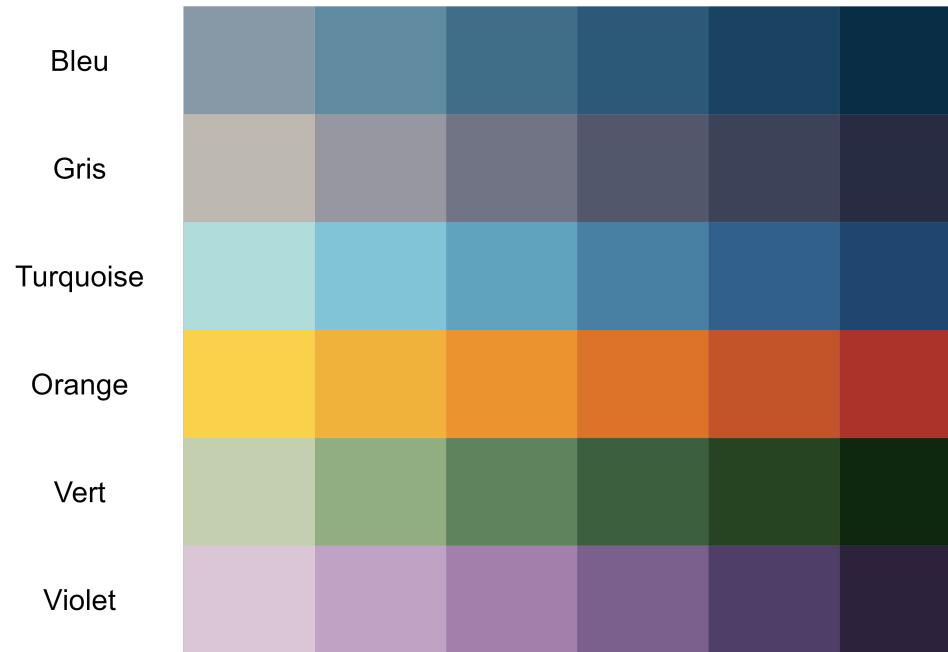
<https://github.com/BlakeRMills/MetBrewer>



<https://github.com/karthik/wesanderson>

## La génération de palettes

- La fonction centrale : [insee.pal](#)



## La génération de palettes

- La fonction centrale : **insee.pal**
- L'affichage automatique

```
● ● ●  
print.palette <- function(x, ...) {  
  
  n <- length(x)  
  old <- par(mar = c(0.5, 0.5, 0.5, 0.5))  
  on.exit(par(old))  
  
  image(1:n, 1, as.matrix(1:n), col = x, xlab = "x",  
        ylab = "", xaxt = "n", yaxt = "n", bty = "n")  
  
  rect(0, 0.94, n + 1, 1.06, col = rgb(1, 1, 1, 0.8), border = NA)  
  text(seq(1,n,1), 1, labels = x, cex = seq(1.7,0.6,-.1)[n], family = "sans")  
}
```



## La génération de palettes

- La fonction centrale : **insee.pal**
- L'affichage automatique
- La copie automatique dans le presse-papier

```
print.palette <- function(x, ...){  
  
  if(!is.nullgetOption("copie.palette")) && getOption("copie.palette")){
    writeLines(
      paste0("c(",  
        paste0(paste0('"' ,x,'"' ),  
               collapse = ", " ),  
        ")"),  
    con = "clipboard",
    sep = ""
  )
  if(requireNamespace("rlang", quietly = T)){
    rlang:::warn("La palette a \u00e9t\u00e9 automatiquement copi\u00e9e dans le presse-papiers. Pour changer ce  
comportement, faites options(copie.palette = FALSE)",  
            .frequency = "regularly",
            .frequency_id = "autocopy_cgt")
  }
}
```

```
.onLoad <- function(libname, pkgname) {  
  
  options(copie.palette = TRUE)  
  
}  
  
.onAttach <- function(libname, pkgname) {  
  
  options(copie.palette = TRUE)  
  
}
```

## Des ajouts maison

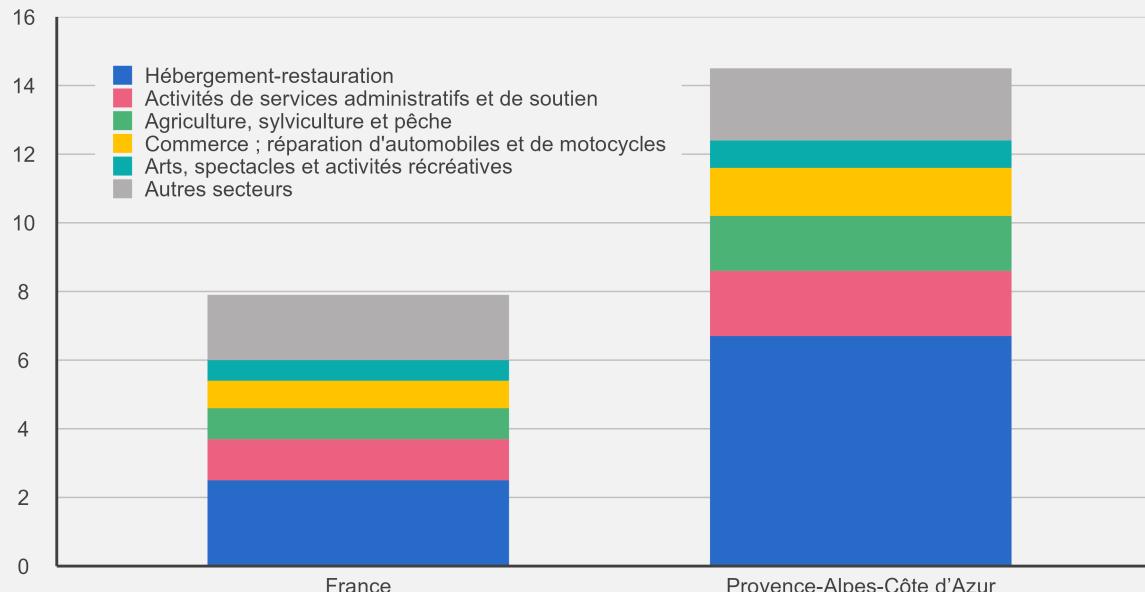
- Une gestion de la **note de lecture** et de la **position des titres** des graphiques
- Création de **palettes sur mesure**



## Un thème pour ggplot2

Répartition de la part d'emplois saisonniers ETP par secteur d'activité

Nombre d'ETP saisonniers pour 1000 ETP saisonniers



Champ : emplois hors intérimaires, hors fonction publique et hors salariés des particuliers employeurs.  
Source : Insee, base Tous salariés 2017.



```
theme_insee <- function(taille.texte = 6,
                        axe.y = FALSE,
                        legende = TRUE,
                        axes.secondaires = FALSE,
                        axes.ticks = FALSE,
                        fond.gris = TRUE)

{

  if(axe.y == TRUE){
    b_axis.line.y = ggplot2::element_line()
  } else {
    b_axis.line.y = ggplot2::element_blank()
  }

  ...

  ggplot2::theme(
    text = ggplot2::element_text(
      family = "sans",
      colour = "#3F3F3F",
      size = taille.texte
    ),
    line = ggplot2::element_line(
      color = "#3F3F3F",
      linewidth = 0.5
    ),
    ...
  )
}
```

## Proposer de nouvelles fonctionnalités

- Permettre d'avoir des entrées personnalisées dans la note de lecture ...
- Ajouter de nouveaux thèmes (notamment pour les cartes et les tableaux) ...
- ... mais au prix de dépendances ? (**grid**, **ggtext**, **gt**, ...)

## Proposer de nouvelles fonctionnalités

- Permettre d'avoir des entrées personnalisées dans la note de lecture ...
- Ajouter de nouveaux thèmes (notamment pour les cartes et les tableaux) ...
- ... mais au prix de dépendances ? (**grid**, **ggtext**, **gt**, ...)

## Un besoin de contributions

- Continuer à s'inspirer de ce qui se fait ailleurs (**gouvdown**, **bbplot**, ...)
- Étendre son utilisation en interne pour tester les fonctionnalités et recenser les besoins

**Merci pour votre attention !**

[insee.fr](http://insee.fr)



Jean DUPIN

Direction Régionale de Bourgogne-Franche-Comté

RENCONTRES R - JUIN 2023

# **Applications Shiny pour le suivi de systèmes agricoles et environnementaux**

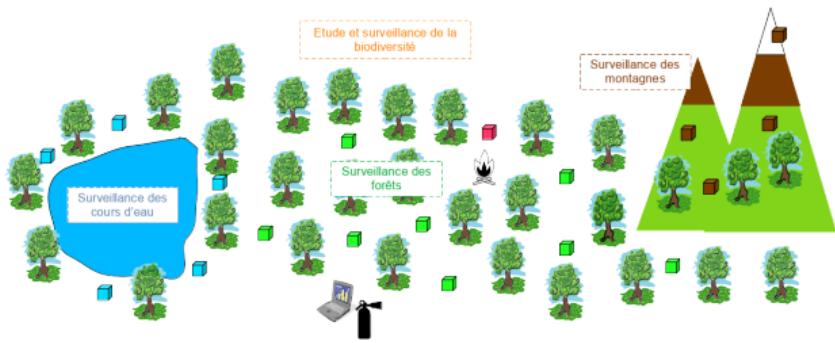
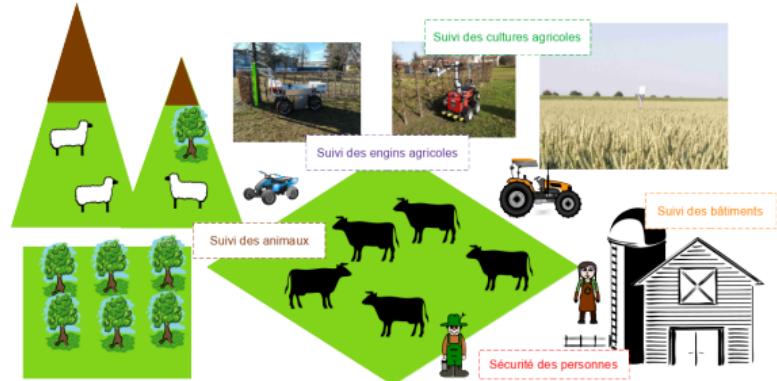
**Journées R 2023, Avignon**

Loris Croce   Thierry Faure   Gil De Sousa

21/06/2023



# Introduction



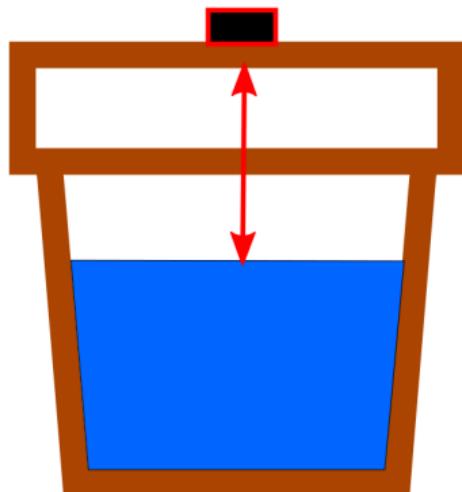
# IoT et systèmes agricoles et environnementaux

- Besoin de surveillance des cultures
  - Agroécologie :
    - Économiser les ressources en eau
    - Dispositif “low-tech” : réutiliser, minimiser les dispositifs
  - Dispositif extérieur :
    - Communication longue portée
    - Autonomie d'énergie
- Visualisation et contrôle à distance

# Projet CoSWoT

- CoSWoT pour *Constrained Semantic Web of Things*
  - Développement de composants logiciels pour systèmes de contraints (ESP32) à très contraints (Arduino Uno)
    - Logiciel “Servient” : composant logiciel client/serveur Web, composé de différents modules dont un gestionnaire d’entrée/sortie et un raisonneur
  - Différents cas d’usage :
    - Bâtiment intelligent
    - Agriculture

# Cas d'études irrigation



**Figure 1:** Niveau d'eau



**Figure 2:** Installation

# Cas d'études irrigation

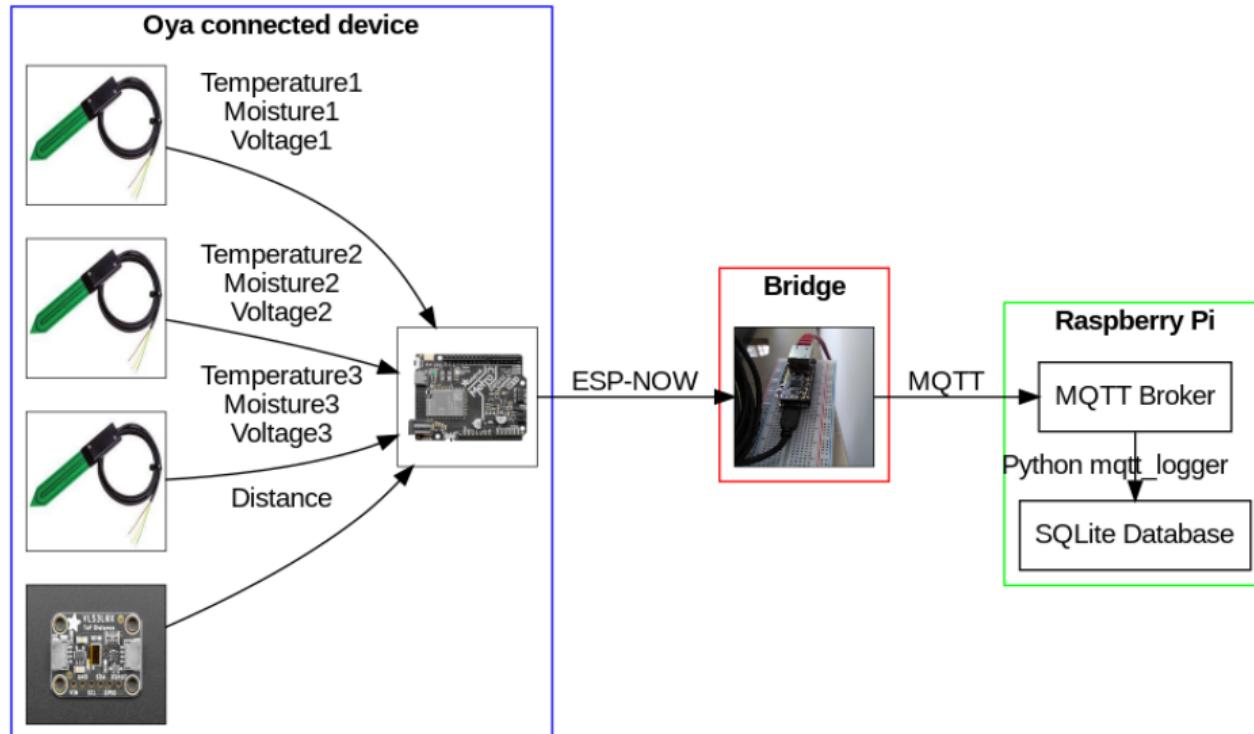


Figure 3: Schéma de l'architecture

# Cas d'études irrigation



# Cas d'études surveillance verger

- Verger de conservation de variétés anciennes d'arbres fruitiers
- Plusieurs sites INRAE
- Caméra posée sur un mat de 15m
  - détection des stades : floraison, poussée & chute de feuilles

# Cas d'études surveillance verger

Surveillance du verger de Crouel

Welcome !

Cette application permet de visualiser les données récoltées sur le verger de conservation de Crouel.

Verger Crouel

- Introduction
- Visualisation
- Recalage caméra
- Détection Floraison
- A Propos

2023-05-12

2023-05-12

© - LISC PIAF - 2023

# Conclusion et perspectives

- R Shiny est un bon outil pour créer des dashboards IoT pour l'agriculture et l'environnement
  - Les dashboards peuvent être organisés par lieu d'expérimentation, par cas d'étude voire par scénario avec des paramètres dédiés
  - Les dashboards permettent de conserver une "reproductibilité" des traitements appliqués tout en conservant les données "brutes" historisées (c'est-à-dire ne pouvant être ni modifiées ni supprimées)
- Il est possible de l'embarquer sur des modèles constraints type Raspberry Pi
- Il permet d'intégrer directement des modèles pour les données IoT

# Camtrapviz, une interface Shiny pour visualiser les données de pièges photographiques



Lisa Nicvert

Laboratoire de Biométrie et Biologie Évolutive,  
Villeurbanne (France)

Stéphane Dray

Laboratoire de Biométrie et Biologie Évolutive,  
Villeurbanne (France)

Hervé Fritz

REHABS International Research Laboratory,  
George (Afrique du Sud)

Sustainability Research Unit, George (Afrique  
du Sud)

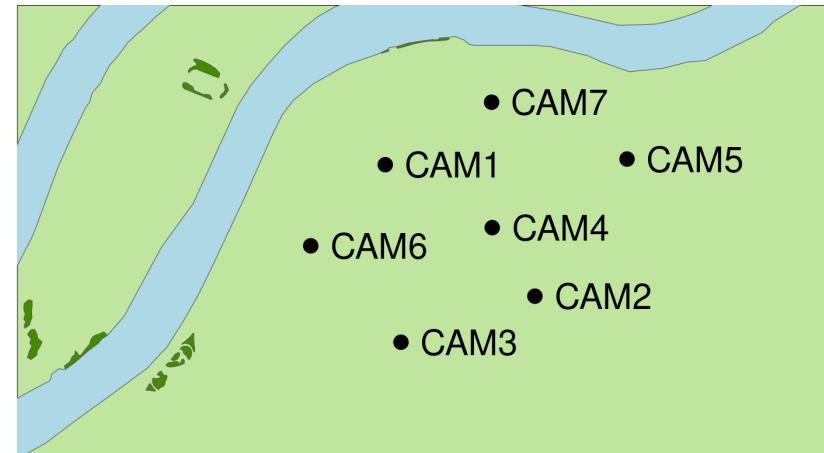
21 juin 2023 – Rencontres R, Avignon

# Pièges photos (camtraps)



Piège photo sur le terrain

- Appareils photos à déclenchement automatique
- Savoir quelles espèces sont présentes, quand et où



Déploiement de pièges photos

# Données

Photos → données tabulées :

## Observations

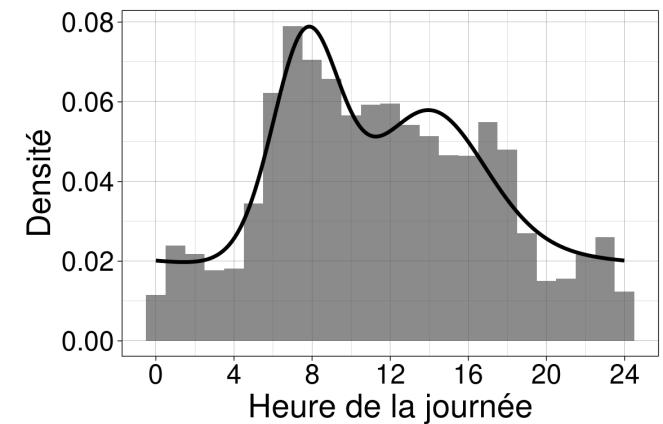
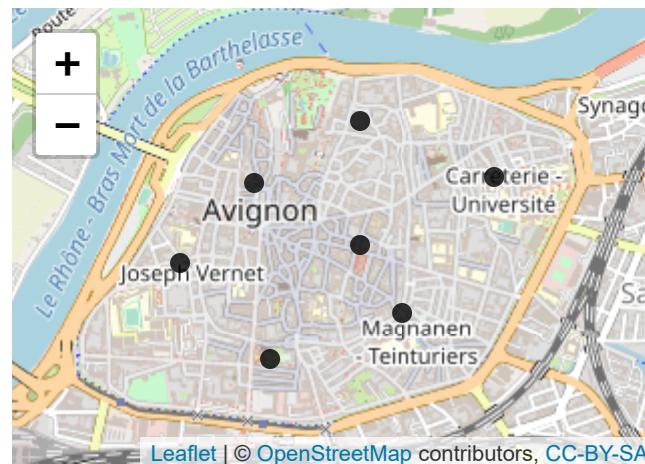
Espèce	Date	Heure	Piège photo	...
elephant	2023-21-06	14:57:12	CAM1	
elephant	2023-24-06	06:06:42	CAM2	
gnou	2023-24-06	07:42:11	CAM1	

## Pièges photos

Piège photo	Latitude	Longitude	...
CAM1	43.9790	4.8909	
CAM2	43.9500	4.8159	
CAM3	43.9521	4.8074	

# Analyses

- **Indicateurs descriptifs** : nombre de pièges photos actifs
- **Graphiques** : cartes
- **Analyses spécifiques aux pièges photos** : activité temporelle des espèces



# Camtrapviz, c'est une app Shiny

The screenshot shows the user interface of the Camtrapviz Shiny application. At the top, there is a dark teal header bar with the title "Camtrapviz" on the left and a menu icon (three horizontal lines) on the right. Below the header is a sidebar on the left containing several menu items, each with an icon and text: "Data import", "Select data", "Data overview", "All species", and "One species". The "Data import" and "One species" items are highlighted with a red rectangular border. To the right of the sidebar is a main content area with a white background. The word "Module" is displayed in large, bold, teal letters at the top of this area. A thin teal horizontal line runs across the page below "Module". Underneath "Module", the text "Contenu du module" is visible.

# Camtrapviz, c'est une app Shiny

## Data import

How to you want to import data?

- Load example file  Upload File

### Records table

Choose records table

#### Camera i

cameralD

#### Species i

snapshotName

#### Date / time column(s)

- Date and time  Timestamp

#### Date i

eventDate

#### Time i

eventTime

#### Count (optional) i

countMax

### Cameras table

Import cameras table

Choose cameras table

#### Camera i

cameralD

#### Latitude/y i

Lat\_Y

#### Longitude/x i

Long\_X

Setup.Time

Lat\_Y

Long\_X

(CRS) i

Lat\_Y

Long\_X

Cam.Brand

Flash

Cam.Serial.No

Height

Fixture

Not present in  
data

## One species analyses

### Choose species

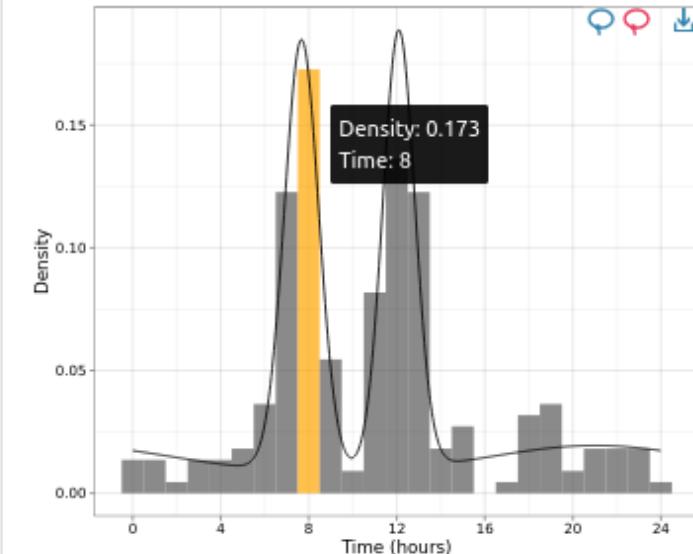
gemsbok

### Activity plot

Number of mixture  
components

3

</> Show code

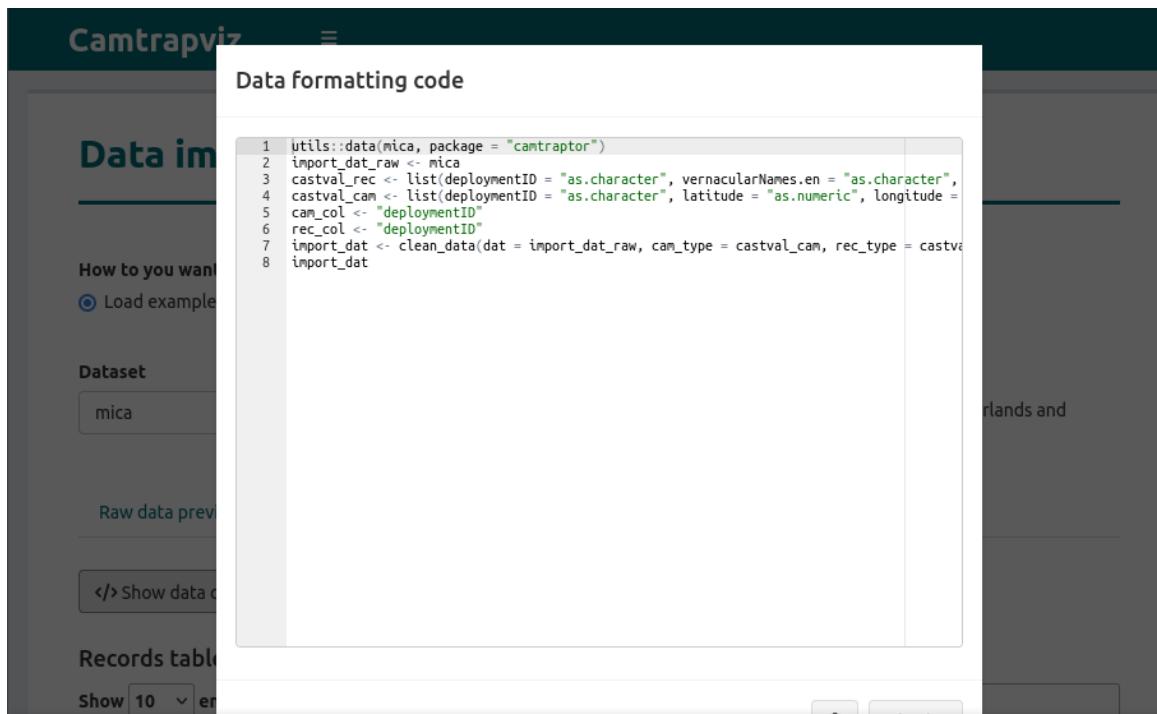


# Exporter le code depuis l'app Shiny

Le package **Shinymeta** permet de générer le code R pour reproduire l'analyse Shiny

`</>` Show data cleaning code

 Download script



The screenshot shows the Camtrapviz shiny application interface. On the left, there's a sidebar with options like 'Data import', 'How to you want to use it?', and 'Dataset' (set to 'mica'). A button labeled '</> Show data cleaning code' is visible. In the center, a modal window titled 'Data formatting code' displays the following R code:

```
1 utils::data(mica, package = "camtraptor")
2 import_dat_raw <- mica
3 castval_rec <- list(deploymentID = "as.character", vernacularNames.en = "as.character",
4 castval_cam <- list(deploymentID = "as.character", latitude = "as.numeric", longitude =
5 cam_col <- "deploymentID"
6 rec_col <- "deploymentID"
7 import_dat <- clean_data(dat = import_dat_raw, cam_type = castval_cam, rec_type = castva
8 import_dat
```



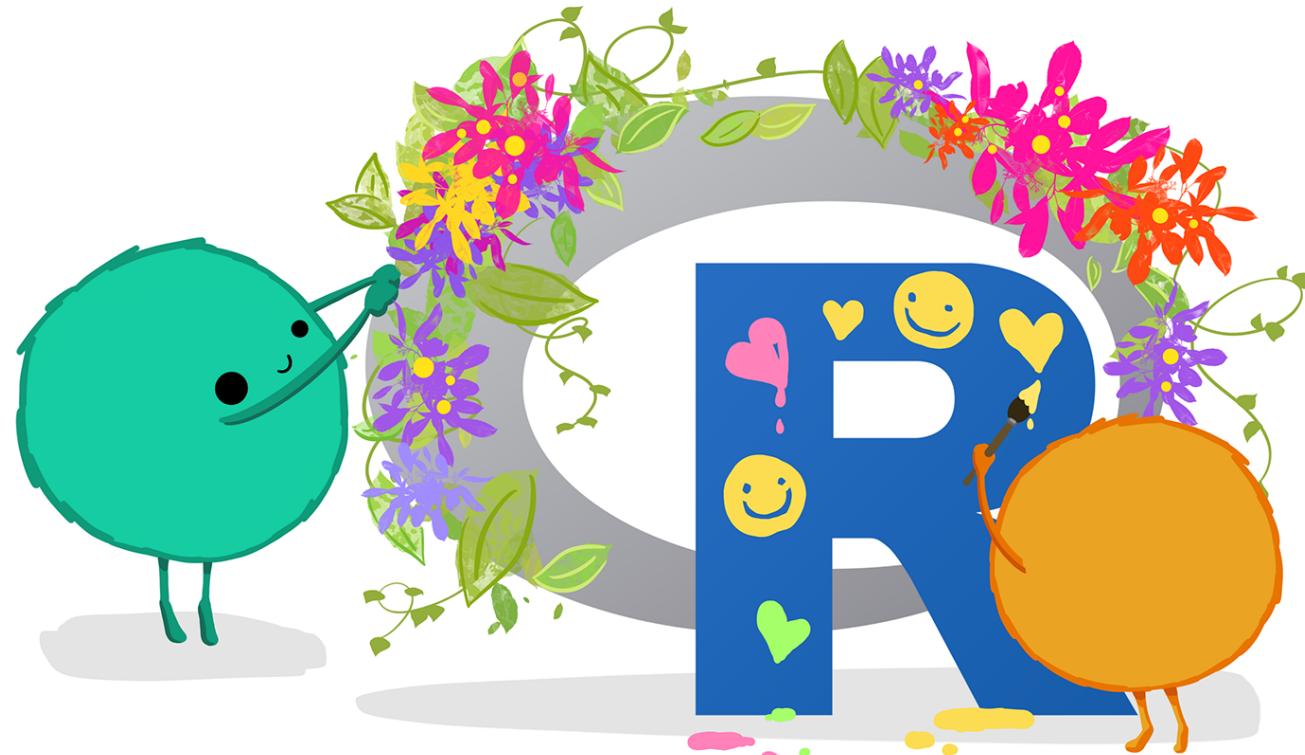
# Camtrapviz, c'est aussi un package

- Utilisation en dehors de l'interface Shiny
- Installation facile
- Documentation



# Pourquoi cette app ?

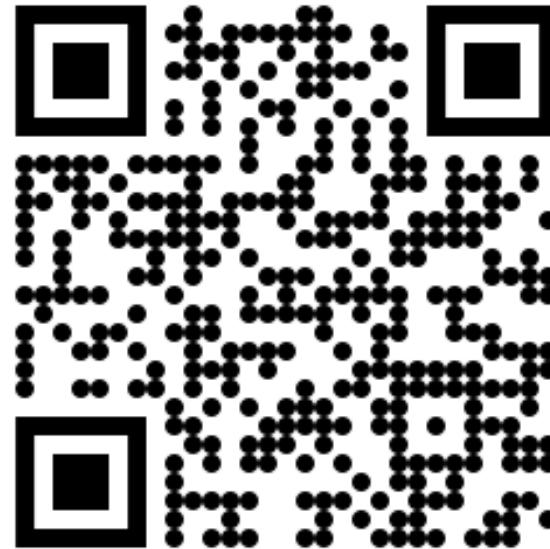
- Visualiser facilement les données de pièges photos
- En s'appuyant sur l'écosystème R en écologie



# Merci !

Code disponible sur GitHub:

<https://github.com/LisaNicvert/camtrapviz>



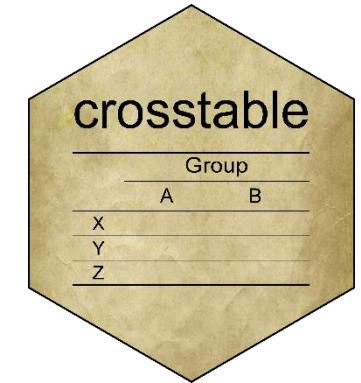
LisaNicvert



@lisanicvert

Photo : steenbok (Snapshot Safari, modifié par Lisa Nicvert)





<https://github.com/DanChaltiel/crosstable>

# Crosstable

L'analyse descriptive pour les nuls  
(et les autres)



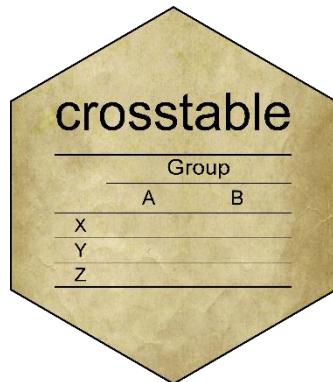
Dan Chaltiel  
Biostatisticien  
Institut Gustave Roussy



Oncostat



Centre de recherche en épidémiologie et santé des populations



# Crosstable: exemple 1

```
library(crosstable)
ct1 = crosstable(mtcars2, c(disp, vs), by=am, total="both",
                 percent_pattern="{n} ({p_row}/{p_col})", percent_digits=0) %>%
  as_flextable()
ct1
```

label	variable	Transmission		Total
		auto	manual	
Displacement (cu.in.)	Min / Max	120.1 / 472.0	71.1 / 351.0	71.1 / 472.0
	Med [IQR]	275.8 [196.3;360.0]	120.3 [79.0;160.0]	196.3 [120.8;326.0]
	Mean (std)	290.4 (110.2)	143.5 (87.2)	230.7 (123.9)
	N (NA)	19 (0)	13 (0)	32 (0)
	straight	7 (50%/37%)	7 (50%/54%)	14 (44%)
	vshaped	12 (67%/63%)	6 (33%/46%)	18 (56%)
Total		19 (59%)	13 (41%)	32 (100%)

total="both" (col)

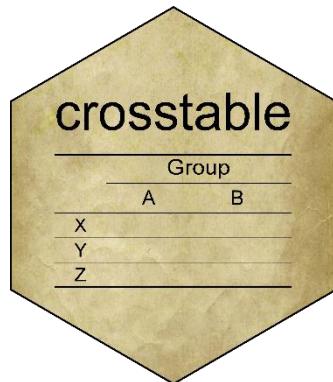
percent\_pattern="{n} ({p\_row}/{p\_col})«  
percent\_digits=0

by=am

total="both" (row)

c(disp, vs)

Engine



# Crosstable: exemple 1

```
library(crosstable)
ct1 = crosstable(mtcars2, c(disp, vs), by=am, total="both",
                 percent_pattern="{n} ({p_row}/{p_col})", percent_digits=0) %>%
  as_flextable()
ct1
```

label	variable	Transmission		Total
		auto	manual	
Displacement (cu.in.)	Min / Max	120.1 / 472.0	71.1 / 351.0	71.1 / 472.0
	Med [IQR]	275.8 [196.3;360.0]	120.3 [79.0;160.0]	196.3 [120.8;326.0]
	Mean (std)	290.4 (110.2)	143.5 (87.2)	230.7 (123.9)
	N (NA)	19 (0)	13 (0)	32 (0)
	straight	7 (50%/37%)	7 (50%/54%)	14 (44%)
	vshaped	12 (67%/63%)	6 (33%/46%)	18 (56%)
	Total	19 (59%)	13 (41%)	32 (100%)

total="both" (col)

percent\_pattern="{n} ({p\_row}/{p\_col})«  
percent\_digits=0

by=am

total="both" (row)

c(disp, vs)

label

variable

Transmission

Total

Displacement (cu.in.)

Engine

Min / Max

Med [IQR]

Mean (std)

N (NA)

straight

vshaped

Total

auto

manual

71.1 / 351.0

120.3 [79.0;160.0]

71.1 / 472.0

275.8 [196.3;360.0]

143.5 (87.2)

196.3 [120.8;326.0]

290.4 (110.2)

230.7 (123.9)

19 (0)

13 (0)

32 (0)

7 (50%/37%)

7 (50%/54%)

14 (44%)

12 (67%/63%)

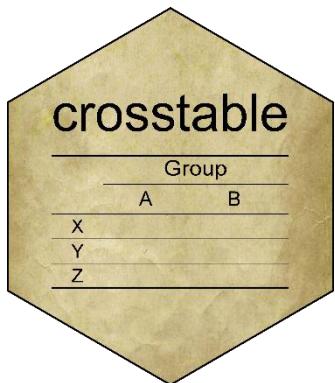
6 (33%/46%)

18 (56%)

19 (59%)

13 (41%)

32 (100%)



# Crosstable: exemple 2

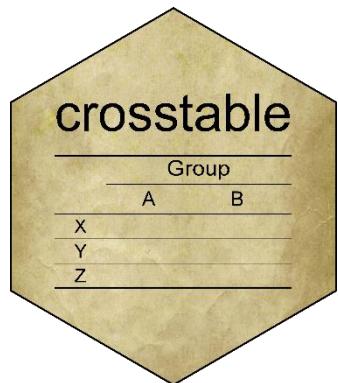
```
ct2 = crosstable(mtcars2, c(starts_with("cy"), ends_with("at")), by=c(am, vs),
                  label=FALSE, num_digits=3, funs=c(mean, quantile),
                  funs_arg=list(probs=c(.25,.75))) %>%
  as_flextable(compact=TRUE, header_show_n=1:2)
ct2
```

**header\_show\_n=1:2**  
Shows the group size

variable	vs=straight (n=14)		vs=vshaped (n=18)	
	am=auto (n=7)	am=manual (n=7)	am=auto (n=12)	am=manual (n=6)
cyl				
4	3 (27.27%)	7 (63.64%)	0 (0%)	1 (9.09%)
6	4 (57.14%)	0 (0%)	0 (0%)	3 (42.86%)
8	0 (0%)	0 (0%)	12 (85.71%)	2 (14.29%)
drat				
mean	3.570	4.149	3.121	3.935
quantile 25%	3.385	3.965	3.052	3.690
quantile 75%	3.920	4.165	3.165	4.140

Annotations pointing to specific parts of the code and the resulting table:

- by=c(am, vs) - Points to the 'by' argument in the code and the 'am' and 'vs' columns in the table.
- starts\_with("cy") - Points to the 'starts\_with("cy")' argument in the code and the 'cyl' row in the table.
- ends\_with("at") - Points to the 'ends\_with("at")' argument in the code and the 'drat' row in the table.
- functs=c(mean, quantile) - Points to the 'funs' argument in the code and the 'mean' and 'quantile' rows in the 'drat' section of the table.
- functs\_arg=list(probs=c(.25,.75)) - Points to the 'funs\_arg' argument in the code and the '25%' and '75%' quantiles in the 'drat' section of the table.
- header\_show\_n=1:2 - Points to the 'header\_show\_n' argument in the code and the header row of the table.
- num\_digits=3 - Points to the 'num\_digits' argument in the code and the numerical values in the table.



# Crosstable: exemple 2

```
ct2 = crosstable(mtcars2, c(starts_with("cy"), ends_with("at")), by=c(am, vs),
                 label=FALSE, num_digits=3, funs=c(mean, quantile),
                 funs_arg=list(probs=c(.25,.75))) %>%
  as_flextable(compact=TRUE, header_show_n=1:2)
ct2
```

**header\_show\_n=1:2**  
Shows the group size

variable	vs=straight (n=14)		vs=vshaped (n=18)	
	am=auto (n=7)	am=manual (n=7)	am=auto (n=12)	am=manual (n=6)
cyl				
4	3 (27.27%)	7 (63.64%)	0 (0%)	1 (9.09%)
6	4 (57.14%)	0 (0%)	0 (0%)	3 (42.86%)
8	0 (0%)	0 (0%)	12 (85.71%)	2 (14.29%)
drat				
mean	3.570	4.149	3.121	3.935
quantile 25%	3.385	3.965	3.052	3.690
quantile 75%	3.920	4.165	3.165	4.140

Annotations pointing to specific parts of the code and the resulting table:

- by=c(am, vs) - Points to the 'by' argument in the code and the 'am' and 'vs' columns in the table.
- starts\_with("cy") - Points to the 'starts\_with("cy")' argument in the code and the 'cyl' row in the table.
- ends\_with("at") - Points to the 'ends\_with("at")' argument in the code and the 'drat' row in the table.
- functs=c(mean, quantile) - Points to the 'funs' argument in the code and the 'mean' and 'quantile' rows in the 'drat' section of the table.
- functs\_arg=list(probs=c(.25,.75)) - Points to the 'funs\_arg' argument in the code and the '25%' and '75%' quantiles in the 'drat' section of the table.
- header\_show\_n=1:2 - Points to the 'header\_show\_n' argument in the code and the header row of the table.
- num\_digits=3 - Points to the 'num\_digits' argument in the code and the numerical values in the table.

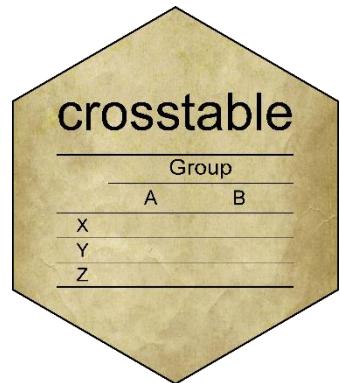


# Crosstable: résumé

- Alternative à {table1} et {gtsummary}
- Tidyverse FTW!
  - Tidyselection sur les colonnes
  - Lambda-fonctions pour funs
- Tout est paramétrable via `crosstable_options()`
- Tout est sur github:
  - Code : <https://github.com/DanChaltiel/crosstable>
  - Doc : <https://danchaltiel.github.io/crosstable/reference/crosstable.html>

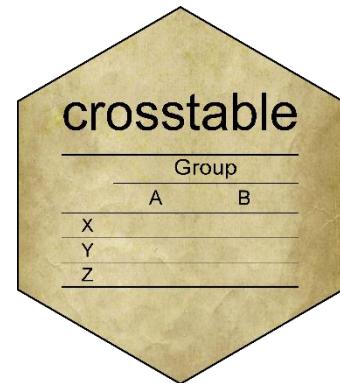
## Usage

```
crosstable(  
  data,  
  cols = everything(),  
  ...,  
  by = NULL,  
  total = c("none", "row", "column", "both"),  
  percent_pattern = "{n} ({p_row})",  
  percent_digits = 2,  
  num_digits = 1,  
  showNA = c("ifany", "always", "no"),  
  label = TRUE,  
  funs = c(` ` = cross_summary),  
  funs_arg = list(),  
  cor_method = c("pearson", "kendall", "spearman"),  
  drop_levels = FALSE,  
  unique_numeric = 3,  
  date_format = NULL,  
  times = NULL,  
  followup = FALSE,  
  test = FALSE,  
  test_args = crosstable_test_args(),  
  effect = FALSE,  
  effect_args = crosstable_effect_args(),  
  margin = deprecated(),  
  .vars = deprecated()  
)
```





# Crosstable: rapports



- Interface avec {officer} pour l'édition de rapports automatiques

```
1 library(crosstable)
2 library(officer)
3
4 ct1 = crosstable(iris2, by=Species)
5
6 crosstable_options(
7   crosstable_fontsize_body=8,
8   crosstable_padding_v=0,
9 )
10
11 doc = read_docx() %>%
12   body_add_title("Dataset iris (nrow={nrow(iris2)}", 1) %>%
13   body_add_title("Not compacted", 2) %>%
14   body_add_normal("Table @ref(ct1) is an *example*.") %>%
15   body_add_crosstable(ct1) %>%
16   body_add_table_legend("Example table",
17                         bookmark="ct1")
18
19 write_and_open(doc)
```

## 1. Dataset iris (nrow=150)

### 1.1. Not compacted

Table 1 is an *example*.

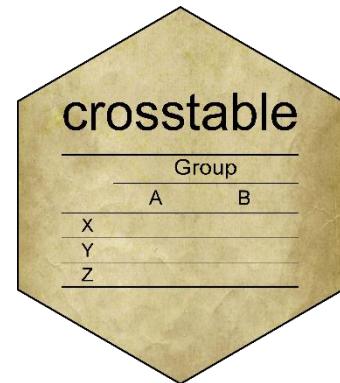
label	variable	Specie		
		setosa	versicolor	virginica
Length of Sepal	Min / Max	4.3 / 5.8	4.9 / 7.0	4.9 / 7.9
	Med [IQR]	5.0 [4.8;5.2]	5.9 [5.6;6.3]	6.5 [6.2;6.9]
	Mean (std)	5.0 (0.4)	5.9 (0.5)	6.6 (0.6)
	N (NA)	50 (0)	50 (0)	50 (0)
Width of Sepal	Min / Max	2.3 / 4.4	2.0 / 3.4	2.2 / 3.8
	Med [IQR]	3.4 [3.2;3.7]	2.8 [2.5;3.0]	3.0 [2.8;3.2]
	Mean (std)	3.4 (0.4)	2.8 (0.3)	3.0 (0.3)
	N (NA)	50 (0)	50 (0)	50 (0)
Length of Petal	Min / Max	1.0 / 1.9	3.0 / 5.1	4.5 / 6.9
	Med [IQR]	1.5 [1.4;1.6]	4.3 [4.0;4.6]	5.5 [5.1;5.9]
	Mean (std)	1.5 (0.2)	4.3 (0.5)	5.6 (0.6)
	N (NA)	50 (0)	50 (0)	50 (0)
Width of Petal	Min / Max	0.1 / 0.6	1.0 / 1.8	1.4 / 2.5
	Med [IQR]	0.2 [0.2;0.3]	1.3 [1.2;1.5]	2.0 [1.8;2.3]
	Mean (std)	0.2 (0.1)	1.3 (0.2)	2.0 (0.3)
	N (NA)	50 (0)	50 (0)	50 (0)

Table 1: Example table



#RR2023

# Crosstable: rapports



- Interface avec {officer} pour l'édition de rapports automatiques

```
1 library(crosstable)
2 library(officer)
3
4 ct1 = crosstable(iris2, by=Species)
5
6 crosstable_options(
7   crosstable_fontsize_body=8,
8   crosstable_padding_v=0,
9 )
10
11 doc = read_docx() %>%
12   body_add_title("Dataset iris (nrow={nrow(iris2)}", 1) %>%
13   body_add_title("Not compacted", 2) %>%
14   body_add_normal("Table @ref(ct1) is an *example*.") %>%
15   body_add_crosstable(ct1) %>%
16   body_add_table_legend("Example table",
17                         bookmark="ct1")
18
19 write_and_open(doc)
```

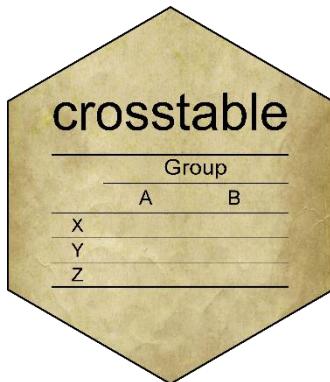
## 1. Dataset iris (nrow=150)

### 1.1. Not compacted

Table 1 is an *example*.

label	variable	Specie		
		setosa	versicolor	virginica
Length of Sepal	Min / Max	4.3 / 5.8	4.9 / 7.0	4.9 / 7.9
	Med [IQR]	5.0 [4.8;5.2]	5.9 [5.6;6.3]	6.5 [6.2;6.9]
	Mean (std)	5.0 (0.4)	5.9 (0.5)	6.6 (0.6)
	N (NA)	50 (0)	50 (0)	50 (0)
Width of Sepal	Min / Max	2.3 / 4.4	2.0 / 3.4	2.2 / 3.8
	Med [IQR]	3.4 [3.2;3.7]	2.8 [2.5;3.0]	3.0 [2.8;3.2]
	Mean (std)	3.4 (0.4)	2.8 (0.3)	3.0 (0.3)
	N (NA)	50 (0)	50 (0)	50 (0)
Length of Petal	Min / Max	1.0 / 1.9	3.0 / 5.1	4.5 / 6.9
	Med [IQR]	1.5 [1.4;1.6]	4.3 [4.0;4.6]	5.5 [5.1;5.9]
	Mean (std)	1.5 (0.2)	4.3 (0.5)	5.6 (0.6)
	N (NA)	50 (0)	50 (0)	50 (0)
Width of Petal	Min / Max	0.1 / 0.6	1.0 / 1.8	1.4 / 2.5
	Med [IQR]	0.2 [0.2;0.3]	1.3 [1.2;1.5]	2.0 [1.8;2.3]
	Mean (std)	0.2 (0.1)	1.3 (0.2)	2.0 (0.3)
	N (NA)	50 (0)	50 (0)	50 (0)

Table 1: Example table



# Crosstable: pivot/transpose

Pivot : pivot\_wider() sur les niveaux

```
ctl = crosstable(mtcars2, c(mpg, drat, wt, qsec), label=F,
                 funs=c(mean, sd))
```

```
ctl %>% as_flextable()
```

```
pivot_crosstable(ctl) %>%
  as_flextable()
```

	label	variable	value
mpg	mean	20.1	
	sd	6.0	
drat	mean	3.6	
	sd	0.5	
wt	mean	3.2	
	sd	1.0	
qsec	mean	17.8	
	sd	1.8	



	label	variable	
		mean	sd
	mpg	20.1	6.0
	drat	3.6	0.5
	wt	3.2	1.0
	qsec	17.8	1.8

Transpose : inverse cols & by

```
ct2 = crosstable(mtcars2, c(mpg, drat), by=am, label=F)
```

```
transpose_crosstable(ct2) %>%
  as_flextable()
```

am	variable	label	
		mpg	drat
auto	Min / Max	10.4 / 24.4	2.8 / 3.9
	Med [IQR]	17.3 [14.9;19.2]	3.1 [3.1;3.7]
	Mean (std)	17.1 (3.8)	3.3 (0.4)
	N (NA)	19 (0)	19 (0)
manual	Min / Max	15.0 / 33.9	3.5 / 4.9
	Med [IQR]	22.8 [21.0;30.4]	4.1 [3.8;4.2]
	Mean (std)	24.4 (6.2)	4.0 (0.4)
	N (NA)	13 (0)	13 (0)



# Crosstable: misc

- Plein d'autres fonctions 😊
- Autres formats d'output
  - `as_gt()`, `as_workbook()`
- Gestion des labels
  - `set_label()`, `apply_label()`
  - `rename_with_labels()`
- Helpers pour officer
  - `body_add_normal()`, `body_add_gg2()`
  - `generate_autofit_macro()`

Lien vers la documentation



Merci de votre attention

