

# Next on stage :

## Group A: Bioinformatics

- A01** - Christelle Hennequet-Antier
- A02** - Cécile Chauvel
- A03** - Oluwafemi Olusoji
- A04** - Olajumoke Evangelina Owokotomo
- A05** - Ozan Cinar
- A06** - Gaëlle Lefort
- A07** - Mélina Gallopin



### Get ready ! Groups B, C

On stage in 3.5 minutes

- B09** - Santagostini Pierre
- B10** - Paul Bouchequet
- B11** - Joseph Oladokun
- B12** - Jacob Bergstedt
- B13** - Antoine Bichat
- B15** - Magali Berland
- B16** - Claire Dandine-Roulland
- C18** - Gregor Stiglic
- C19** - Yao Nie

## Poster A1

# ViSEAGO: Easier data mining of biological functions organized into clusters using Gene Ontology and semantic similarity

<https://forgemia.inra.fr/umr-boa/viseago>

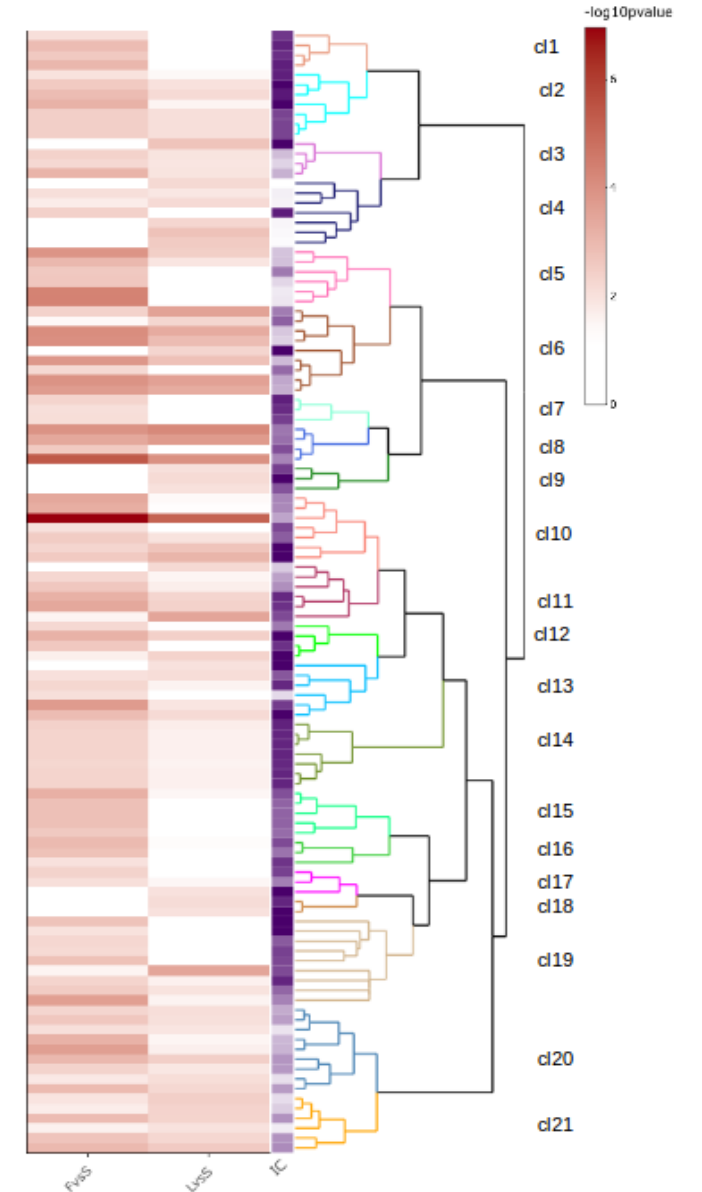
Christelle HENNEQUET-ANTIER, Aurélien BRIONNE, Amélie JUANCHICH

### Concepts

Gene Ontology  
Semantic similarity  
Clustering of GO terms

### R / Bioconductor packages

topGO  
GOSemSim  
ViSEAGO



Clustering heatmap of GO terms



# Evaluation of integrative clustering methods for the analysis of multi-omics data

Cécile Chauvel\*, Alexei Novoloaca\*, Pierre Veyre, Frédéric Reynier and Jérémie Becker

Corresponding author: Jérémie Becker, BIOASTER Research Institute, 40 avenue Tony Garnier, 69007 Lyon, France. Tel.: +33 4 69 85 19 21;

Fax: +33 4 72 70 48 2; E-mail: jeremie.becker@bioaster.org

\*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

## Abstract

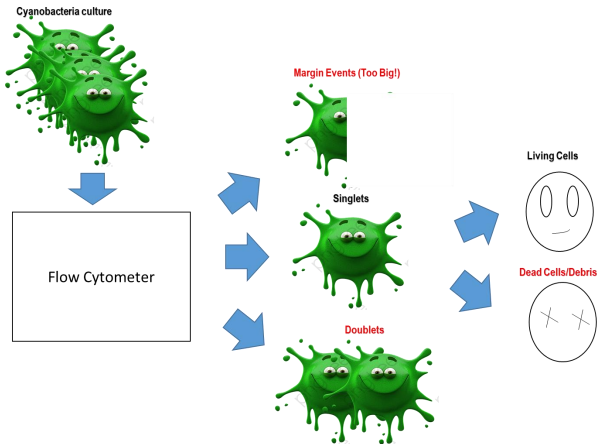
Recent advances in sequencing, mass spectrometry and cytometry technologies have enabled researchers to collect large-scale omics data from the same set of biological samples. The joint analysis of multiple omics offers the opportunity to uncover coordinated cellular processes acting across different omic layers. In this work, we present a thorough comparison of a selection of recent integrative clustering approaches, including Bayesian (BCC and MDI) and matrix factorization approaches (iCluster, moCluster, JIVE and iNMF). Based on simulations, the methods were evaluated on their sensitivity and their ability to recover both the correct number of clusters and the simulated clustering at the common and data-specific levels. Standard non-integrative approaches were also included to quantify the added value of integrative methods. For most matrix factorization methods and one Bayesian approach (BCC), the shared and specific structures were successfully recovered with high and moderate accuracy, respectively. An opposite behavior was observed on non-integrative approaches, i.e. high performances on specific structures only. Finally, we applied the methods on the Cancer Genome Atlas breast cancer data set to check whether results based on experimental data were consistent with those obtained in the simulations.

**Key words:** benchmark; clustering; data integration; multi-omics; unsupervised analysis

POSTER  
A2

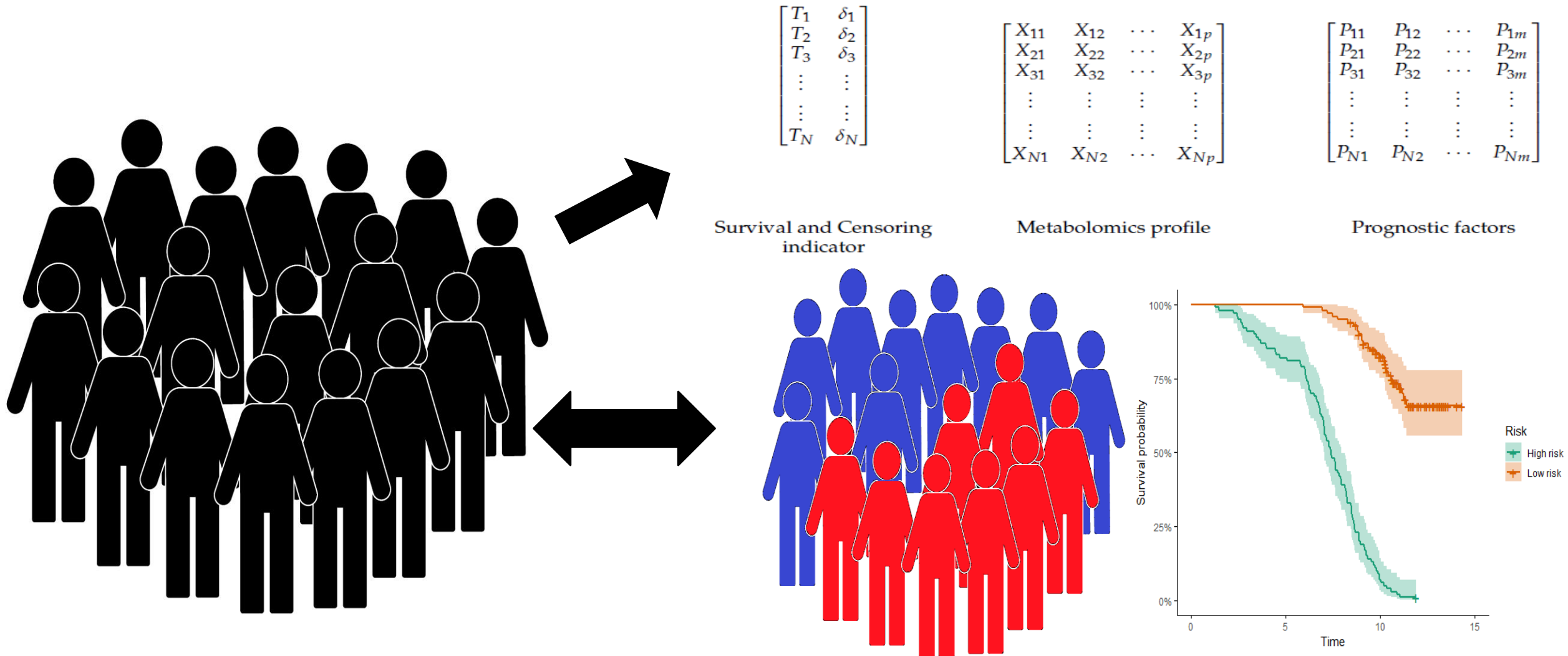


# cyanoFilter, An Automated Framework for identifying picocyanobacteria populations obtained via flow cytometry



How about a package for this?

# Predicting Risk Groups for Survival of Cancer Patients Using a Robust Metabolomic Signature: The MetabolicSurv R package.



Olajumoke Evangelina Owokotomo & Ziv Shkedy

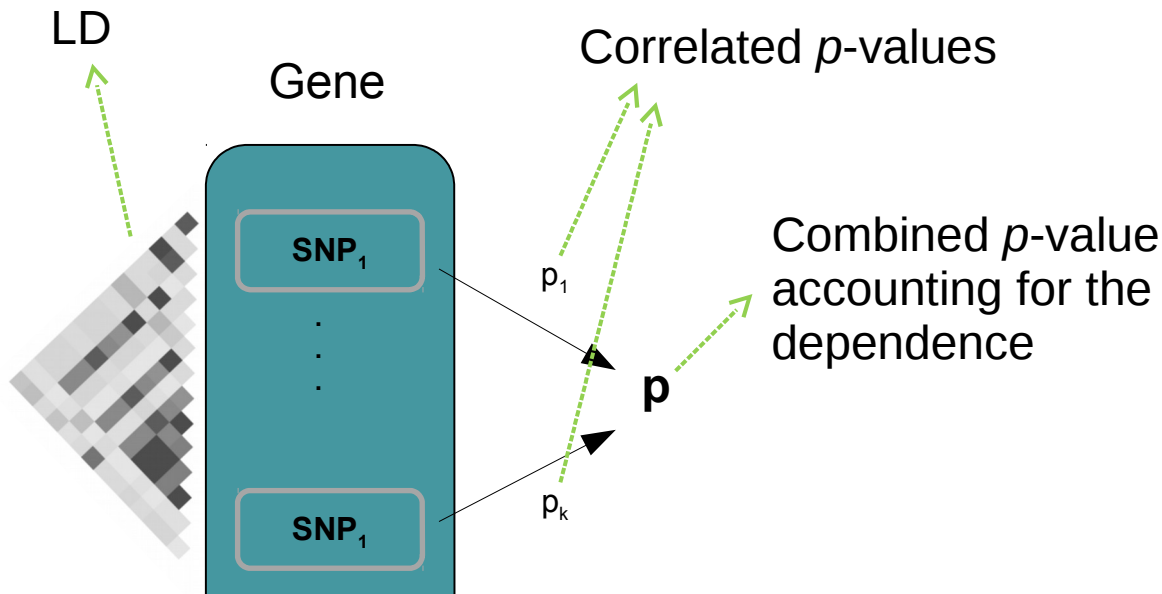
Center for Statistics, Statistical Bioinformatics, Hasselt University, 3590 Diepenbeek, Belgium

# The Current Status of Methods for Combining Dependent $p$ -Values and Extending them with a Novel Package, poolR

Ozan Çınar<sup>A</sup>, Wolfgang Viechtbauer<sup>A</sup>

<sup>A</sup> Department of Psychiatry and Neuropsychology, Maastricht University

- Combining  $p$ -values where the tests are dependent
- An example: Gene-Based Testing in Genome-Wide Association Studies
  - Linkage Disequilibrium (LD) among the Single-Nucleotide Polymorphisms  
=> Correlated tests (i.e., correlated  $p$ -values)



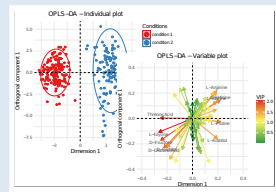
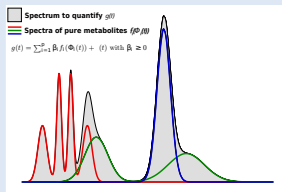
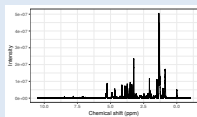
- poolR implements and extends methods for combining dependent  $p$ -values by using
  - Effective number of tests
  - Empirical distributions
  - Test-statistics under dependence

Poster A05

# Poster A6

## ASICS: a new R package for identification and quantification of metabolites in complex 1H NMR spectra

G. Lefort, L. Liaubet, H. Quesnel, C. Canlet, N. Vialaneix and R. Servien



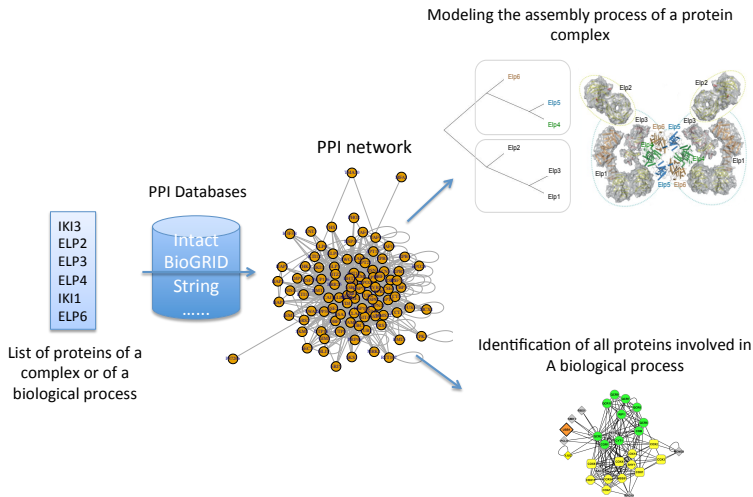
Automatic Statistical Identification (and quantification) in Complex Spectra

Complete workflow for 1H NMR spectra

Available on Bioconductor

# Appinetwork Package for Protein-Protein Interaction

Méline Gallopin, Benjamin Auder, Annie Glatigny, Marie-Hélène Mucchielli





# Next on stage :

Groups B & C: Bioinformatics/Biostatistics/Epidemiology



**B09** - Santagostini Pierre  
**B10** - Paul Bouchequet  
**B11** - Joseph Oladokun  
**B12** - Jacob Bergstedt  
**B13** - Antoine Bichat  
**B15** - Magali Berland  
**B16** - Claire Dandine-Roulland  
**C18** - Gregor Stiglic  
**C19** - Yao Nie

## Get ready ! Groups D, E

On stage in 4.5 minutes

**D29** - Kristoffer Segerstrøm Mørk

**D30** - Gregory Guernec

**D31** - Hicham Nocairi

**D32** - Michael Sachs

**E33** - Shama Virani, Daniela Mariosa, Florence Guida, Hilary Robbins

**E34** - Laure Cougnaud

**E35** - Patrick Wolf, Tobias Buchmann

**E36** - Özge Igde

**E37** - Sigrid Keydana

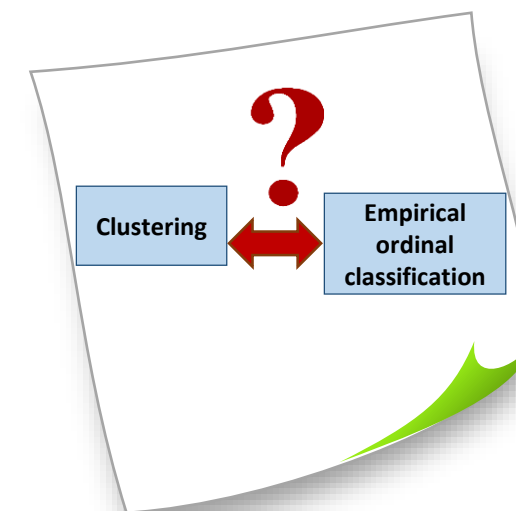
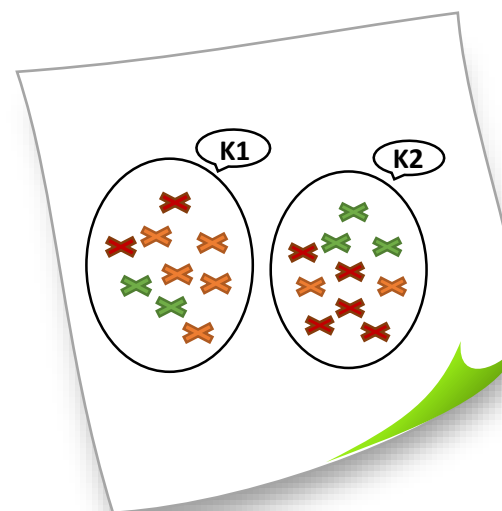
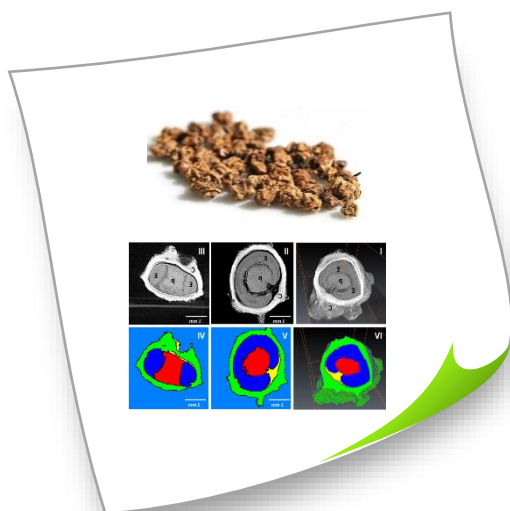
# Ordinal clustering of seed populations with data extracted from RGB imaging and X-ray tomography

Garbougé Hadhami<sup>1</sup>, Santagostini Pierre<sup>2</sup>, Charrier Aurélie<sup>3</sup>, Demilly Didier<sup>3</sup>, and Rousseau David<sup>1</sup>

1-LARIS, UMR INRA IRHS, Université d'Angers, 62 Avenue Notre Dame du Lac, 49000 Angers

2-IRHS, Agrocampus Ouest, INRA, Université d'Angers, SFR 4207 QuaSav, Beaucouzé, France

3- GEVES, Station Nationale d'Essais de Semences (SNES), France



# Using R for automatic sleep analysis as a regular part of the clinical process

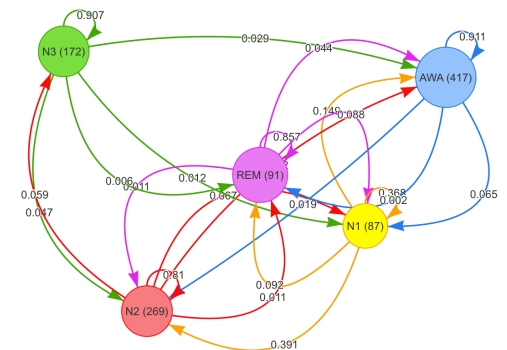
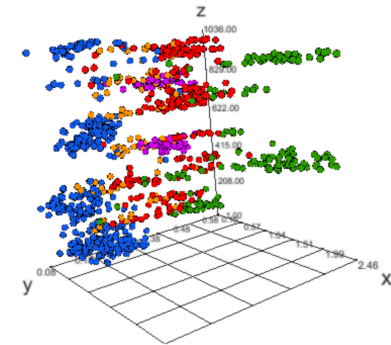
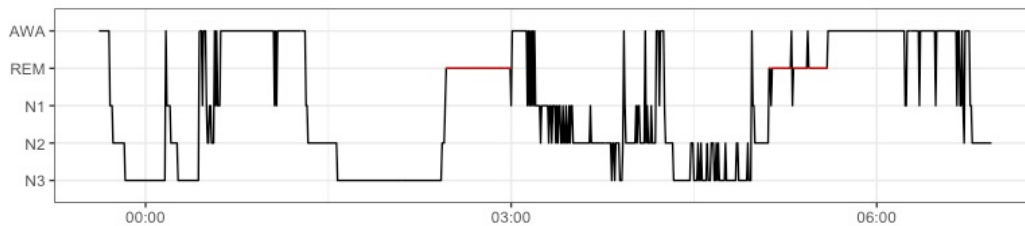
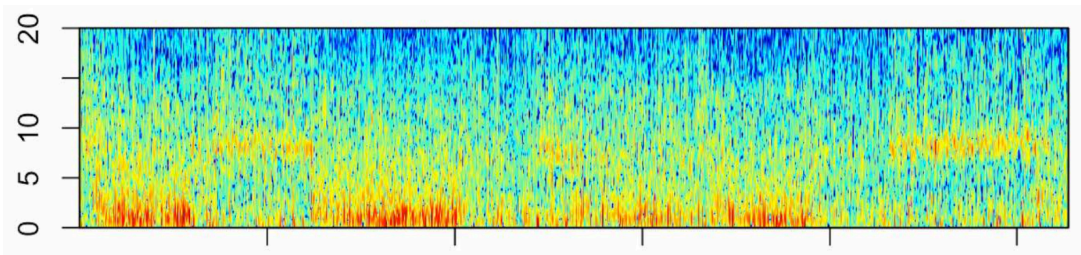
Paul Bouchequet

Geoffroy Solelhac

Damien Léger

Centre du Sommeil et de la Vigilance, Hôtel Dieu, Paris

Université de Paris



# R SUPPORT FOR PUBLIC HEALTH AND BIOINFORMATICS

Joseph Oladokun



## BACKGROUND

- Analysis of Public health data helps increase emergency response
- It also helps in developing solutions that addresses public health problems
- With data, we are able to reach hard-to-reach communities with affordable healthcare



## GROWING NUMBER OF PACKAGES FOR PUBLIC HEALTH



There are some packages in R that are very important when working with genomic and public health data. These packages help us in analyzing and making sense of data which are crucial in responding to emergency and building solutions that support and solve these problems.

- Bioconductors
- Genetics
- Gap
- Pheatmap
- ReCon Packages

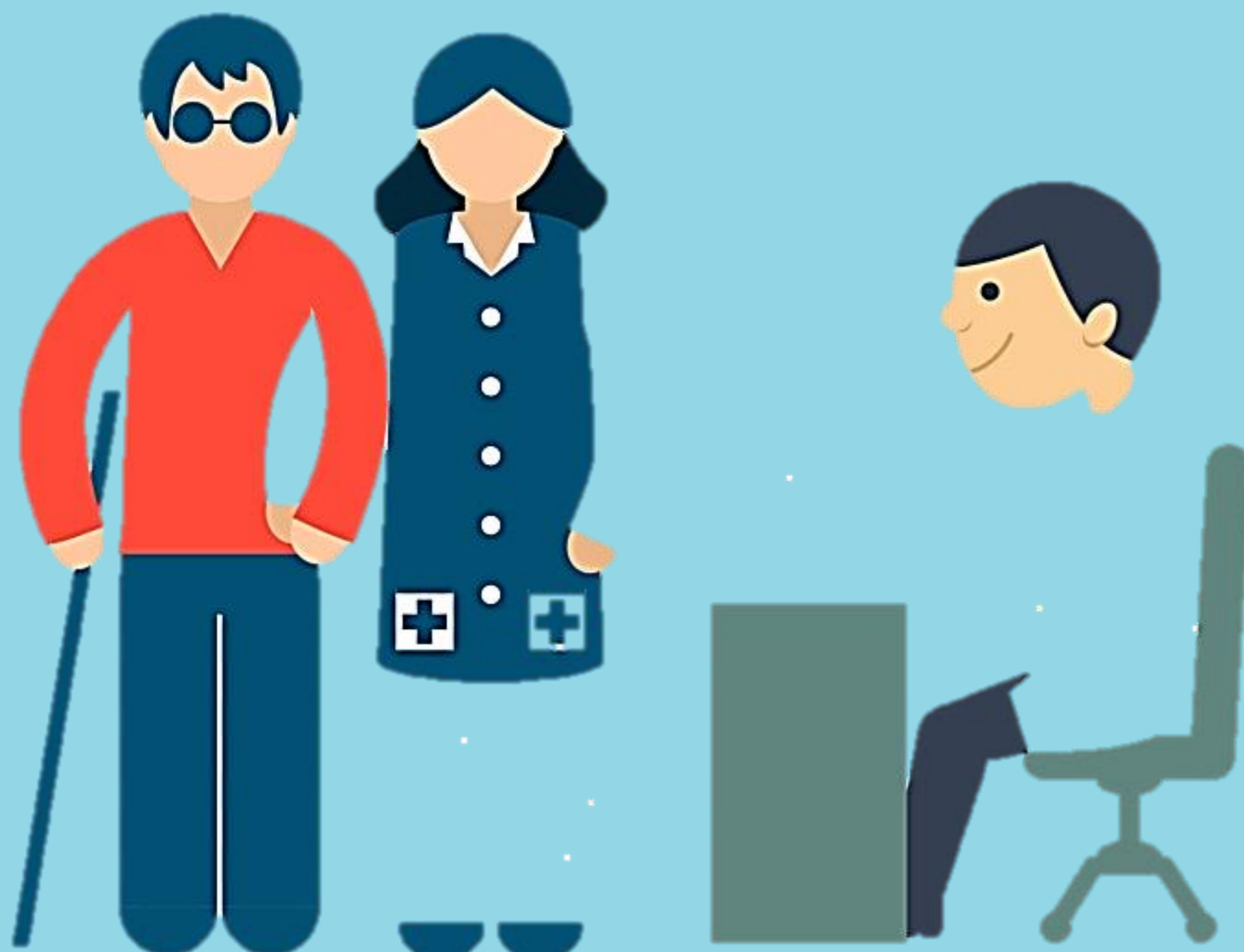


Using the right tool to analyze data has aided our emergency response and development of new products in healthtech industry.



## KEY FUNCTIONS

- Bioconductor package is provides access to a broad range of powerful statistical and graphical methods for the analysis of genomic data and facilitate the inclusion of biological metadata in the analysis of genomic data.
- ReCon Packages includes packages specifically designed for handling, visualizing, and analyzing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption, and system infrastructure. Their packages include projection, incidence, earlyR etc



**Joseph Oladokun**

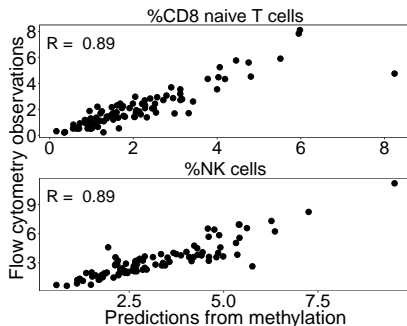
eHealth Africa and Curators  
University

@Godskid\_CFC

oladokunjoseph2@gmail.com

# Using R to estimate blood cell composition

- Concentration of different immune cells in blood is a critical biomarker
- Measuring immune cells is invasive and labor intensive
- New method estimates 25 immune cell proportions with high accuracy from DNA methylation

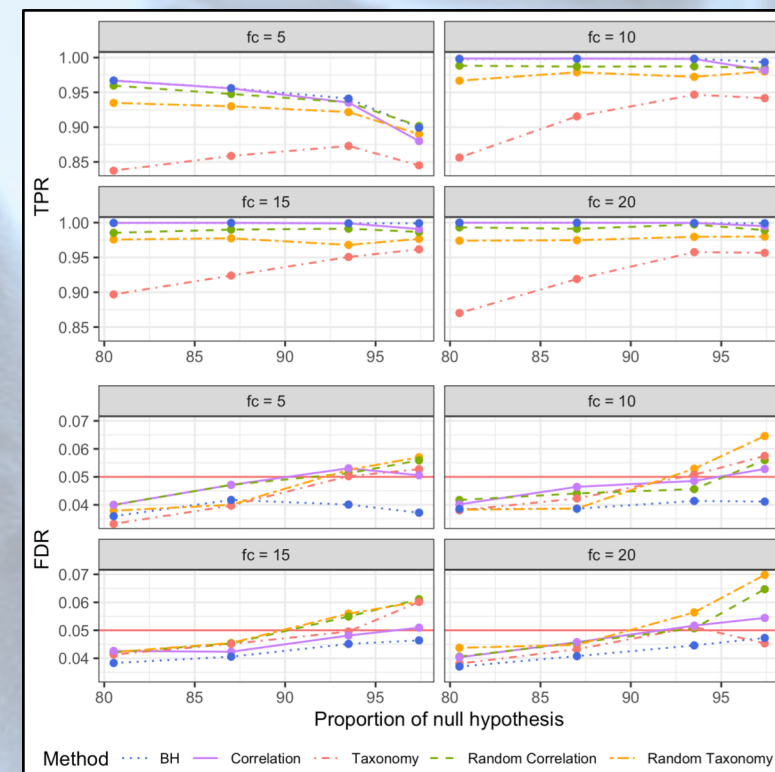
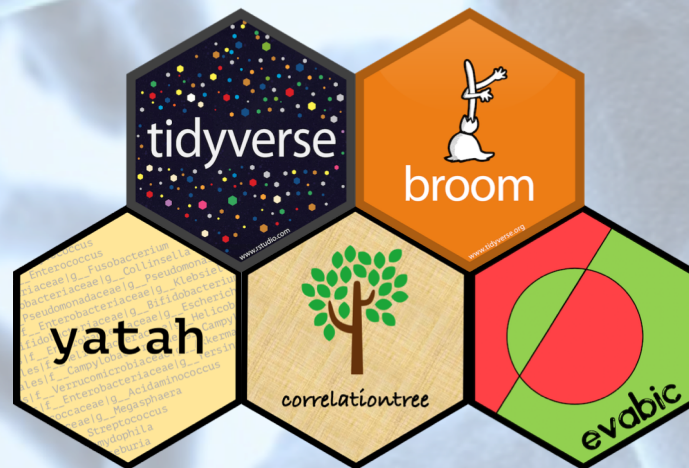


# Quantifying the impact of tree choice in metagenomics differential abundance studies with R

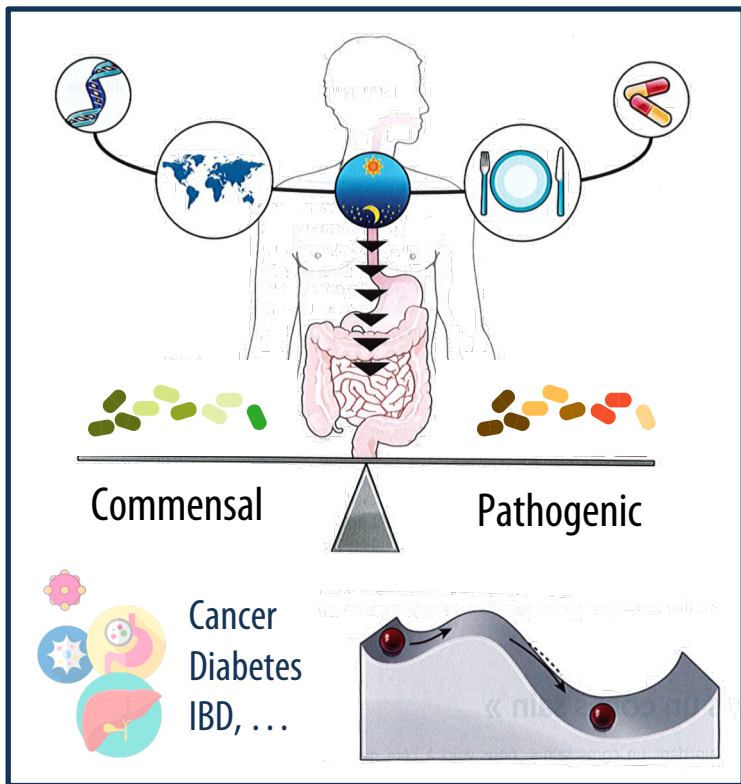
B13

Antoine Bichat, Christophe Ambroise, Mahendra Mariadassou, Jonathan Plassais

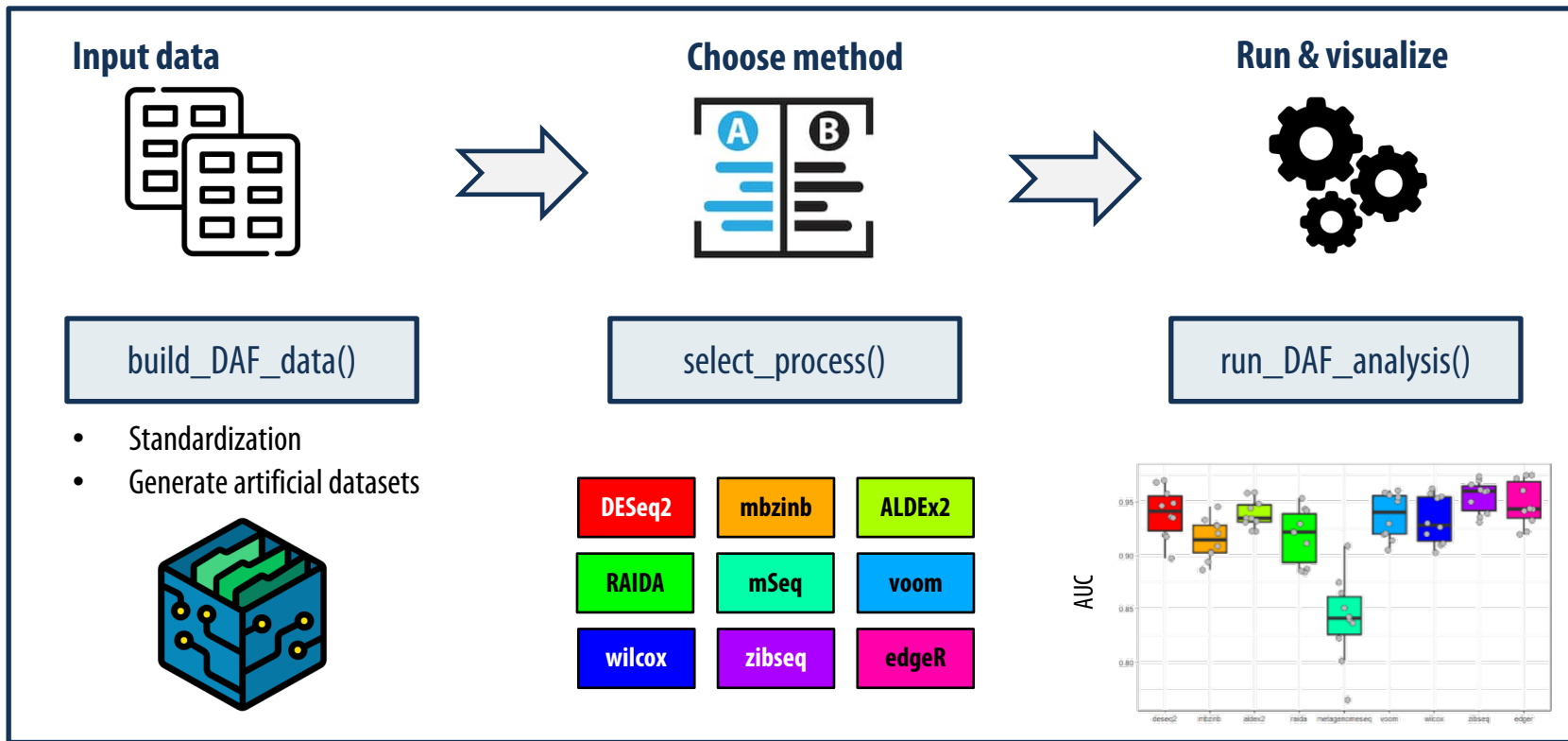
Genus	S001	S002	S003	S004	S005	...
Atopobium	84	0	12	54	0	...
Eggerthella	2	0	0	7	0	...
Prevotella	525	7	134	753	0	...
Lactobacillus	88	1770	1490	119	2136	...
Streptococcus	0	0	138	4	0	...
Dialister	152	4	2	192	0	...
Megasphaera	402	0	4	102	0	...
Sneathia	302	0	35	272	0	...



### SCIENTIFIC CONTEXT



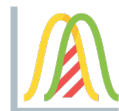
### METHODS & RESULTS



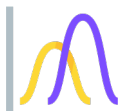
### KEYWORDS



METAGENOMICS  
BIOMARKER DISCOVERY



DIFFERENTIAL  
ABUNDANCE



NEGATIVE BINOMIAL &  
ZERO INFLATED DISTRIBUTION



R PACKAGE



REPRODUCIBLE RESEARCH



IN SILICO BENCHMARK

## Genetic Data Manipulation

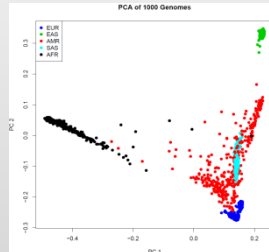
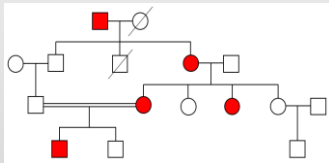
## Quality Control



- Callrate
- Frequency
- Hardy-Weinberg Equilibrium
- ...

## Descriptive Analyses

## Familial and/or Population Structure

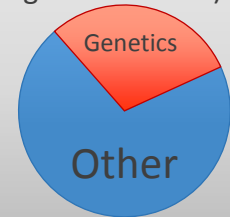


## Linkage Disequilibrium

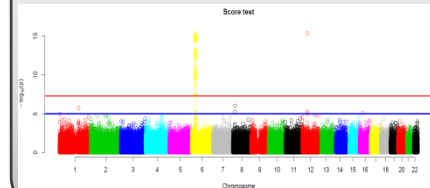
Genetic Analyses  
Mixed or Classical  
models

## Heritability

(Phenotype variance  
proportion explained by  
genetic effects)



## Association Studies





# Early detection of type 2 diabetes mellitus

Leona CILAR<sup>a</sup>, Primož KOCBEK<sup>a</sup>, Gregor ŠTIGLIC<sup>a,b</sup>

<sup>a</sup>Faculty of Health Sciences, University of Maribor, Žitna ulica 15, 2000 Maribor, Slovenia

<sup>b</sup>Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova ulica 17, 2000 Maribor, Slovenia



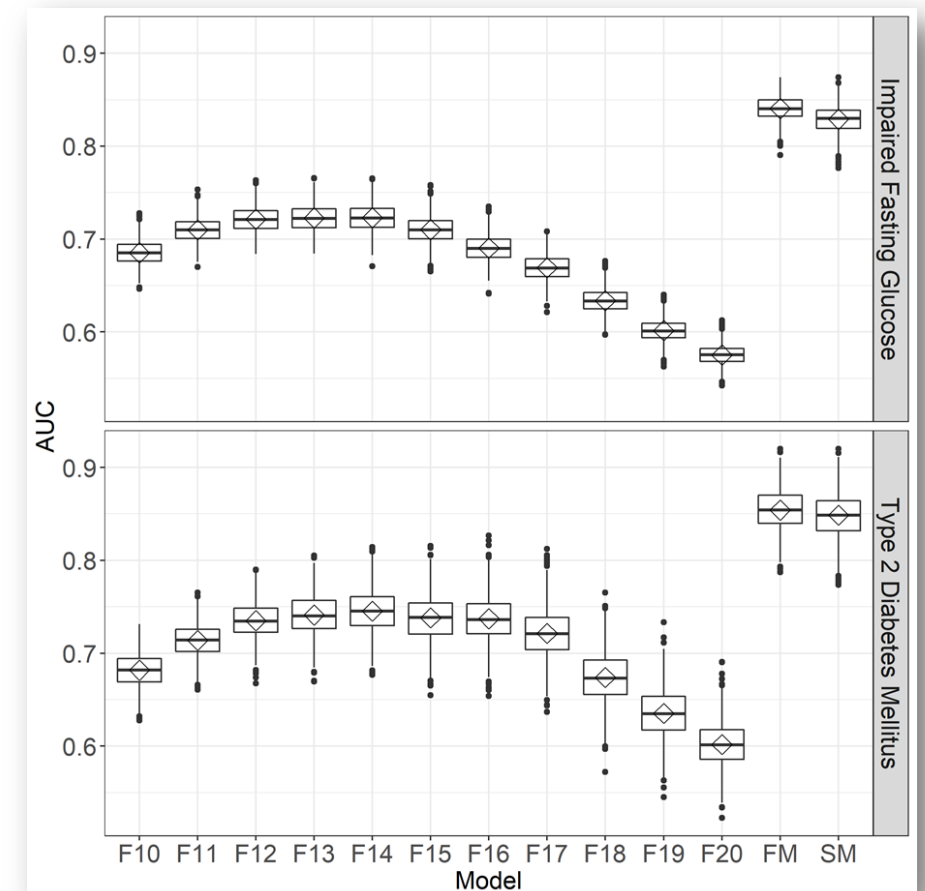
Faculty of Health Sciences

**Methods:** EHR in 5 SLO Healthcare institutions, 2073 individuals, FINDRISC model.

**Results:** The AUC of 0.851 was achieved by a simplified version of a screening test predicting undiagnosed T2DM with 0.840 when predicting undiagnosed IFG. The final model contained 5 questions for undiagnosed T2DM prediction and 6 for IFG prediction model.

**Implementation:** Study proposed two simplified prediction models for screening of undiagnosed T2DM and IFG. A significant improvement in performance was demonstrated in comparison to the original FINDRISC questionnaire.

SLORISK tool has given healthcare professionals the opportunity to demonstrate how a change in lifestyles can influence the risk over time.



# Small Area Estimation (SAE) of All-Cause Mortality and Life Expectancy in British Columbia (BC), Canada, 2000-2017

**Yao Nie, MSc; Bonnie Henry, MD, MPH, FRCPC;  
Kim Reimer, BSc, BEd, BA; Aijun Yang, MSc; Xibiao Ye, PhD**

Office of the Provincial Health Officer, BC Ministry of Health



Office of the  
Provincial Health Officer

# Next on stage :

Groups D & E: Methods/Case studies/Machine learning



**D29** - Kristoffer Segerstrøm Mørk

**D30** - Gregory Guernec

**D31** - Hicham Nocairi

**D32** - Michael Sachs

**E33** - Shama Virani, Daniela Mariosa, Florence Guida,  
Hilary Robbins

**E34** - Laure Cougnaud

**E35** - Patrick Wolf, Tobias Buchmann

**E36** - Özge Igde

**E37** - Sigrid Keydana

## Get ready ! Group F

On stage in 4.5 minutes

**F41** - Job Spijker

**F42** - Ivan Navarro

**F43** - Gabriele Galatolo

**F44** - Caroline Buridant, Camille Gaal, Antoine Menard,  
Sébastien Lê

**F45** - Johannes Burkhardt, Matthias Bannert

**F46** - Clemens Zauchner, Theo Boutaris, Dana Jomar

# Estimate, Estimator, Estimand?

## What?

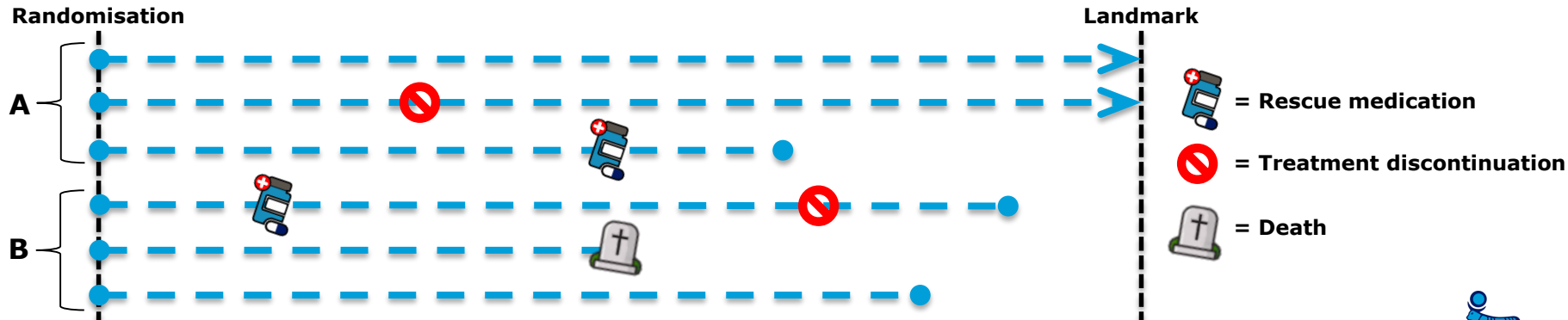
- A description of what has to be estimated in order to answer a scientific question of interest

## Where?

- Estimands are a central part of late phase clinical trials

## Why?

- To have a clear and standard way to communicate what is being estimated
- To be aware of what question is actually answered
  - *What is the treatment effect of A versus B when use of rescue medication is ignored?*
  - *What is the treatment effect of A versus B if use of rescue medication did not occur?*



- Estimands may also be useful to you!

# OTrecod: An R package for data integration using Optimal Transportation theory



Gregory Guerneq,

Pierre Navaro, Valerie Gares, Jeremy Omer,  
Philippe Saint-Pierre, Nicolas Savy



## Context

« Combining information available in 2 independent databases referred to the same target population »

Database (DB) A   DB B  

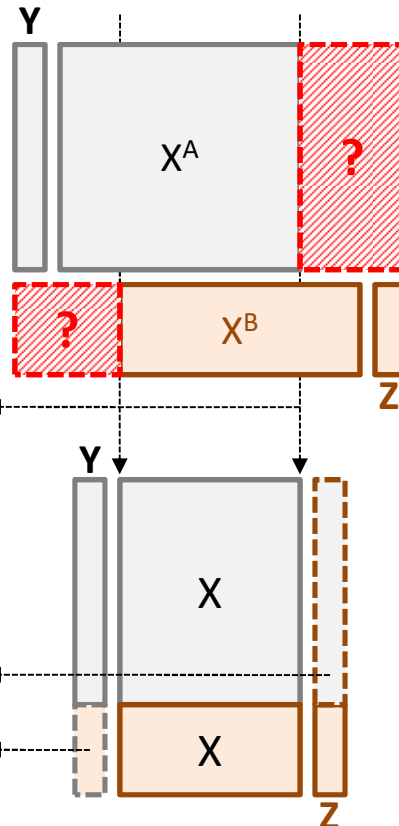
⇒ Y and Z not jointly observed, can summarize a same information coded in distinct scalings

No overlapping part

⇒ Using a remaining set X of common covariates

## Goal

Generate a single complete database with adapted recodings:  
Z in A  
Y in B



## How solve this problem ?

⇒ Framework of statistical matching

⇒ Could be seen as a specific problem of missing data imputation



Most frequently used packages

- StatMatch
- mice

## What's new with OTrecod ?

⇒ Recoding Y and Z using Optimal Transportation theory ensures the respect of conditional distributions

⇒ Give promising accuracies compared to other methods (MICE and Hot Deck especially) on simulated and real datasets

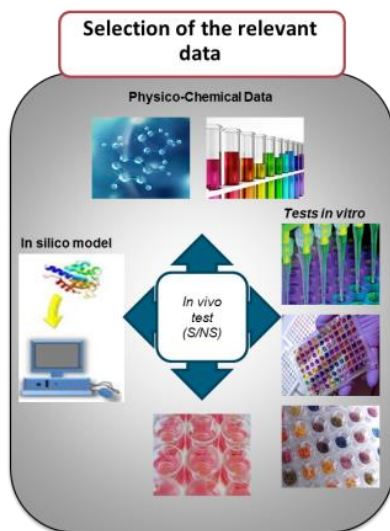
⇒ Give an R users access to this original theory from 2 independent databases

## Statutory context:

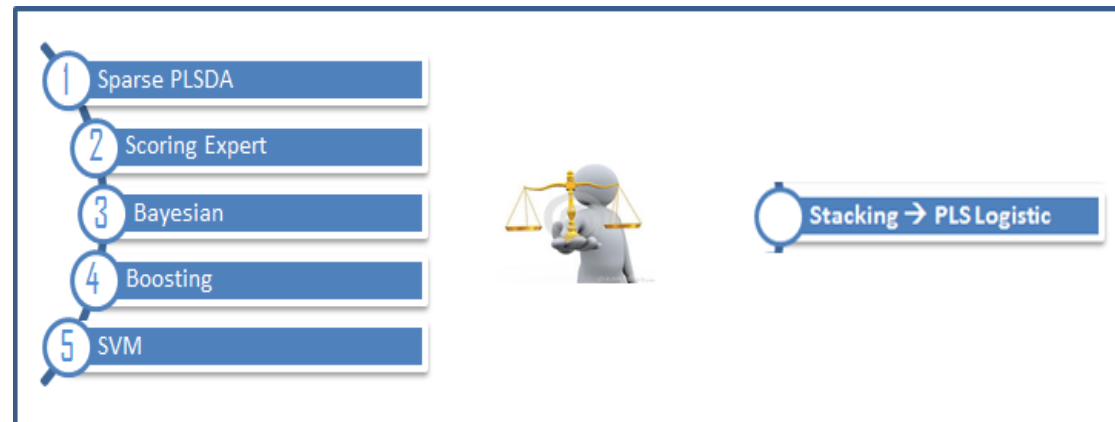
The 7th Amendment of the European Cosmetic Directive has banned the *in vivo* tests on animals for the safety evaluation of ingredients. L'Oréal has thus developed several types of *in vitro*, *in silico* methods and collected other kinds of information on its chemicals like *physico-chemical* data. Due to the complexity of the skin sensitization process (Sensitizer or not), it is now agreed that it is necessary to use all these information to predict safety.

The statistical objective in this case is to predict the *in vivo* tests results realized before their ban, by using *in vitro in silico* data and *physico-chemical* data.

## Data Preparation



## Stacking meta-Model



## Conclusions :

The Stacking Meta-Model gives a prediction model with better performances for the development of alternative approaches in safety evaluation of chemicals than each of the initial five models separately.

Existing **machine-learning** methods can be used for classification of **censored event-history data** by using **pseudo-observations** and these tools for **estimating and optimizing the AUC**.

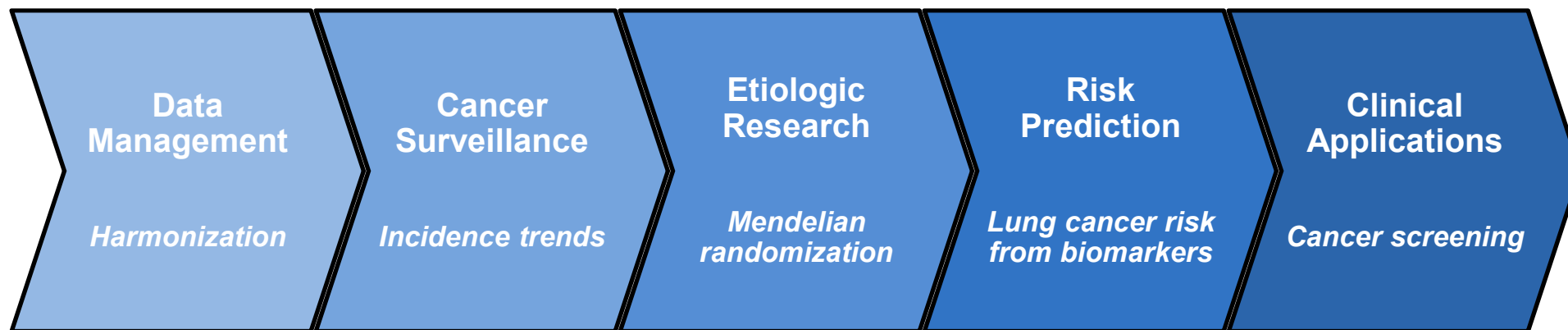


[github.com/sachsmc/pseudoloss](https://github.com/sachsmc/pseudoloss)

D32

# What R we doing at the International Agency for Research on Cancer?

Karine Alcalá, Florence Guida, Daniela Mariosa, Hilary Robbins, Shama Virani



International Agency for Research on Cancer

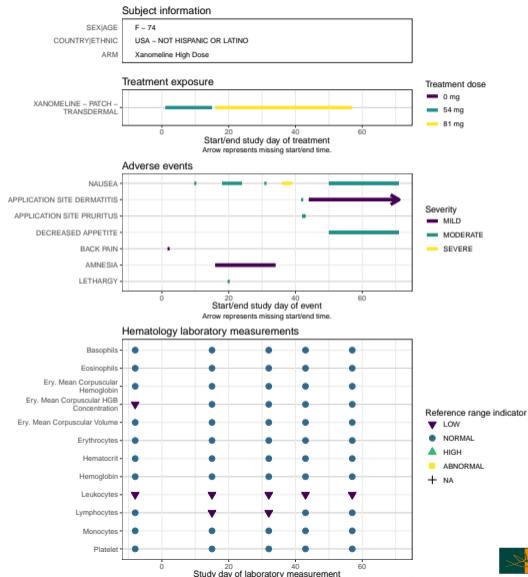




# Patient profile visualization

Laure Cougnaud

- Early detection/diagnostic/monitoring in a **clinical trial**
- **Combined view** of patient data along trial timelines:
  - patient metadata
  - treatment exposure
  - adverse events
  - concomitant medication
  - laboratory measurements
- **Modular approach**: event/interval/line/text
- Creation of trial report + Shiny application
- Enhanced by **standard data format** (CDISC **ADaM**, **SDTM**)



# Analyzing Energy-Research in Europe

Patrick Wolf, Tobias Buchmann

SPONSORED BY THE



Federal Ministry  
of Education  
and Research



## Goal

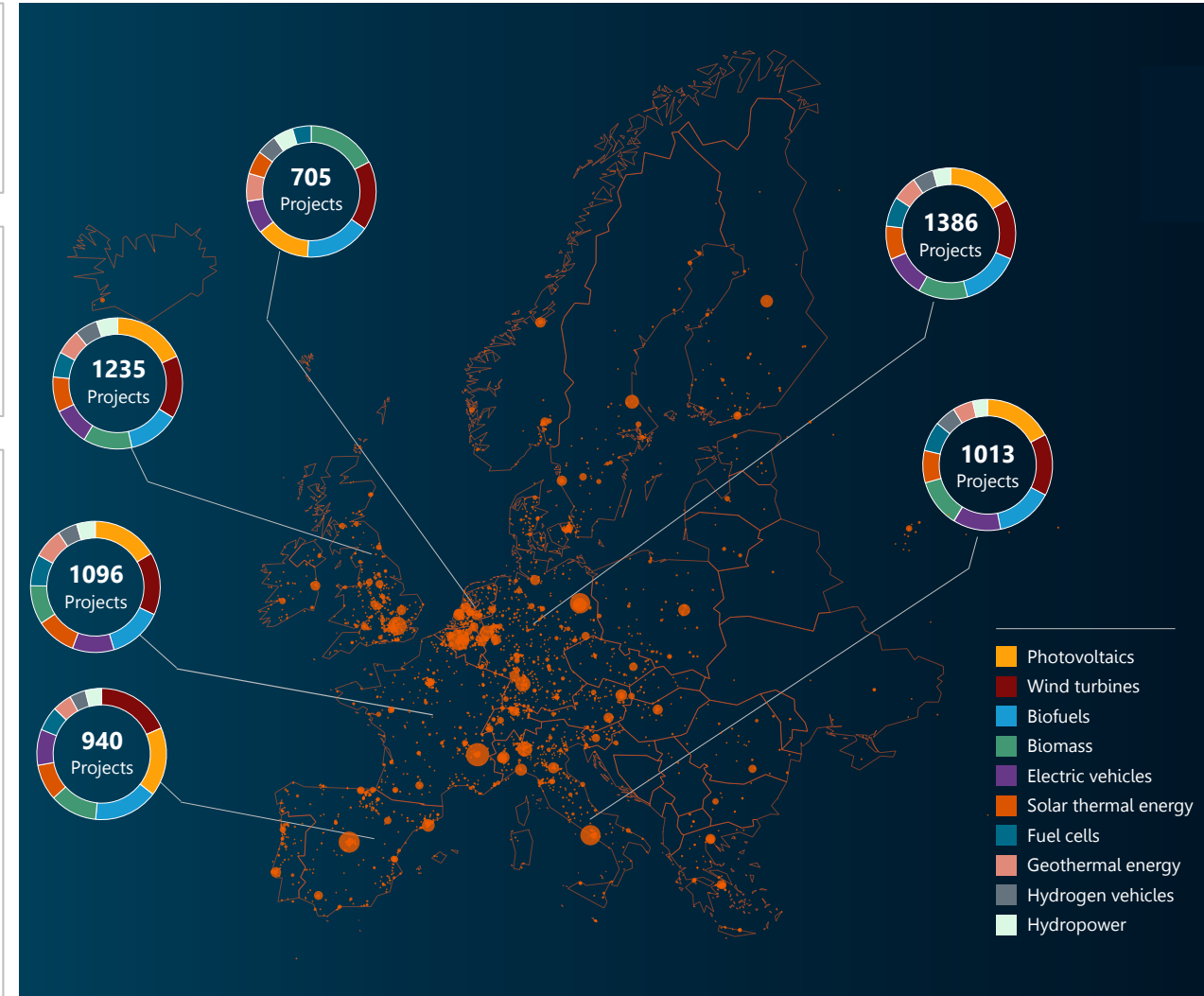
- Analysing innovation processes and networks in the field of energy-research in Europe

## Data

- CORDIS database of the European Commission, containing information about research projects funded by the European Union

## Methods & Procedure

- 1) Information retrieval via text- and data mining methods (*rvest*)
- 2) Effortful data processing and preparation via matching techniques and distance metrics (*base*, *stringdist*, *taRifx.geo*)
- 3) Identifying energy-related projects via multi-label text classification methods (*ruimtehol*)
- 4) Topic Modelling approach for further concretization of the research topics (*ldatuning*, *text2vec*, *LDAvis*)
- 5) Analyze and illustrate data (*base*, *rgexf*)



# NETWORK ANALYSIS: DETECTION OF RISK GROUPS WITHIN BANKING SYSTEM OF TURKEY

## STEP 1

Tagging risk group names and citizenship/tax ids

1. Starting with a Citizenship/Tax ID
2. Getting all risk group names declared for this ID by different banks
3. Other Citizenship/Tax IDs for the risk group names declared by banks in the second item
4. Iterating second and third item
5. Stop the iteration when there is no Citizenship/Tax ID and no risk group name to be connected
6. Defining the composed risk group set as Flag with name "RG1"
7. Starting next risk group set with another Citizenship/Tax ID

## STEP 2

Singularization of risk group names for each identification number

1. Deleting "GROUP" and special characters from risk group names
2. Creating the list of words and their similar words in the risk group name text by using "agrep" function
3. Finding the most frequently used risk group name
4. Choosing the longest risk group name by comparing found risk group name with most frequently used risk group name
5. Adding new risk group names into dataset

## STEP 3

Clustering the network structure in one Flag

1. Creating the two-mode network
2. Converting two-mode network to undirected one-mode network
3. Allocating flags which have many vertices
4. Defining sub-risk groups in flags by "cluster label proportion" function
5. Detecting the possibly incorrect records which connect different sub-risk groups in one flag

## STEP 4

Clustering the network of flags with many nodes

1. Clustering the flags which have many nodes by using hierarchical clustering
2. Cutting the tree where Dice dissimilarity score is 0.8
3. Adding labels of sub-risk group names like "RG1-1"
4. Getting the singularized risk group names for the subgroups of large sized flags
5. Repeating hierarchical clustering for every similar sub-risk group names

## STEP 5

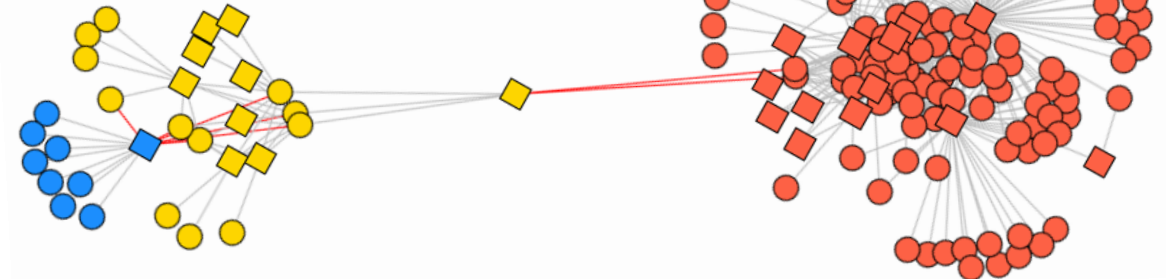
Building shiny application

1. Two tabs for big-sized and normal-sized flags
2. Side panel including the choices for the visualization of the network structure
3. Red marks on the possible incorrect records
4. Tables giving details of detected subgroups such as group members, their credits and non-performing loans

## STEP 6

Deployment of the application via shiny server

1. Installing the Shiny Server on a Linux
2. Deploying the Shiny Application
3. Viewing app via browser



# Because we're skeptics

... make it all probabilistic:

**tfprobability**: R interface to **TensorFlow Probability**

- Deep probabilistic neural networks (Keras layers)
- Markov Chain Monte Carlo and Variational Inference for hierarchical models
- Dynamic linear models and time series
- Optimizer and linear algebra extensions
- All the distributions and transformations you could dream of

... integrated with **TensorFlow 2.0** and running highly performant on **GPU**!



1 skeptischer Hamster zu verkaufen

20 €

25899 Niebüll >

Art

Hamster

Er guckt einen skeptisch an, als würde man nichts richtig machen.

Es macht mich wahnsinnig, ich kann diesen vorwurfsvollen Blick nicht länger ertragen. Sein Name ist Olaf.

Thanks Richard McElreath for the hamster!

# Next on stage :

## Group F: Data management/Data mining



**F41** - Job Spijker

**F42** - Ivan Navarro

**F43** - Gabriele Galatolo

**F44** - Caroline Buridant, Camille Gaal, Antoine Menard, Sébastien Lê

**F45** - Johannes Burkhardt, Matthias Bannert

**F46** - Clemens Zauchner, Theo Boutaris, Dana Jomar

### Get ready ! Groups I, G

On stage in 3 minutes

**I52** - Gi-Seop Lee

**I53** - Nicolas Raillard

**I54** - Wayne Jones

**I55** - Bénédicte Fontez

**I56** - Madalina Olteanu

**G59** - Maximilian Leodolter

**G60** - Benoît Génot

**G61** - Olivier Delaigue

**G63** - Ellen Webborn

**G64** - Ria Van Hecke



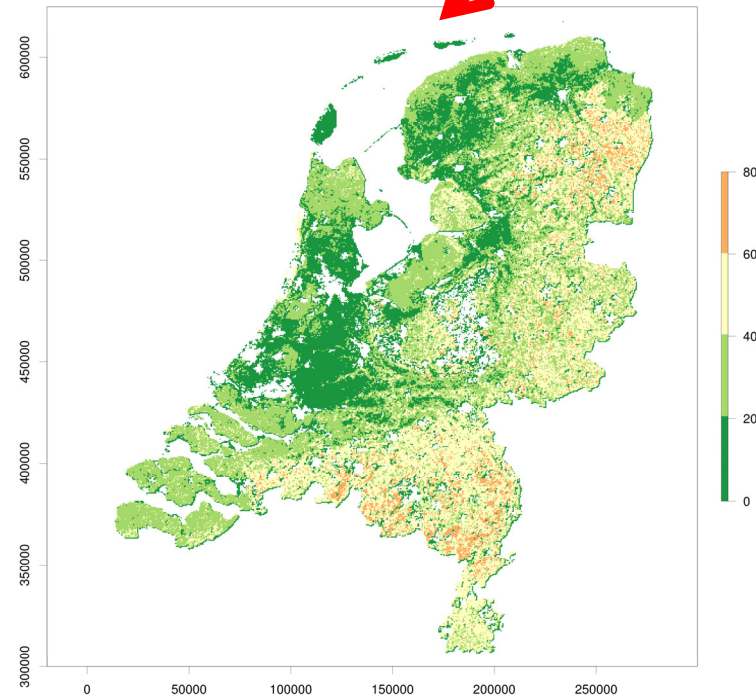
# Using a Data Cube to Efficiently Manage Data for Machine Learning

Job Spijker

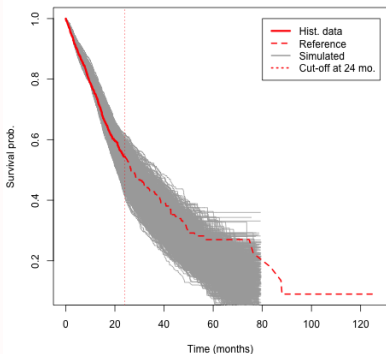
- ✓ Project management
- ✓ Data repository
- ✓ Team work
- ✓ Data lineage
- ✓ Revision History
- ✓ Reusable
- ✓ Reproducible

**F41**

*Using only 4 R  
commands !*



Simulated Survival Curves from Hist. Data



## • Interim Survival Data

- ▶ 1000 patients
- ▶ 2 treatment arms

## • Modeling

- ▶ Parametric (Exp. Wei.)
- ▶ Non-parametric (P-W Exp.)

## • (Objective) Sim. Final Analysis

- ▶ Up to 2K patients (+1K)
- ▶ Impute censored and new patients
- ▶ 10K posterior draws, 10K completions each
- ▶ About 1.5 TByte of simulated data

## Context and Challenge

- High precision in event time predictions is crucial during interim stages of a clinical trial.
- Bayesian inference can be very expensive computationally.
- **We propose:** Storage and use of large R object under constrained computational resources.

# Add new documents into the db

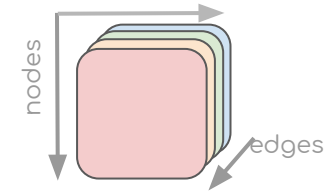
```
persons %>%
  document_insert("john.doe") %>%
  document_set(age=30, birthday="17/01/1989") %>%
  collection_update()
```

# Filter documents from collections

```
filtered.cities <-
  cities %>%
  collection_filter(country="UK", position.latitude %gt% 52.0)
```

## Graph traversal

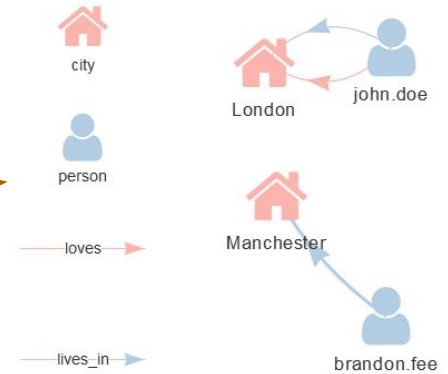
```
london.residence <- residenceGraph %>%
  traversal(vertices = c(all.cities$London), depth = 2)
nodes
```



# aRangodb (and much more!)

```
residenceGraph <-
  residenceGraph %>%
  define_edge("person", "lives_in", "city")
residenceGraph <-
  residenceGraph %>%
  add_edges("lives_in" %owns% edge(all.persons$john.doe %->% all.cities$London)) %>%
  add_edges("lives_in" %owns% edge(all.persons$brandon.fee %->% all.cities$Manchester, since="09/01/2016"))
```

## Add relationships into the graph



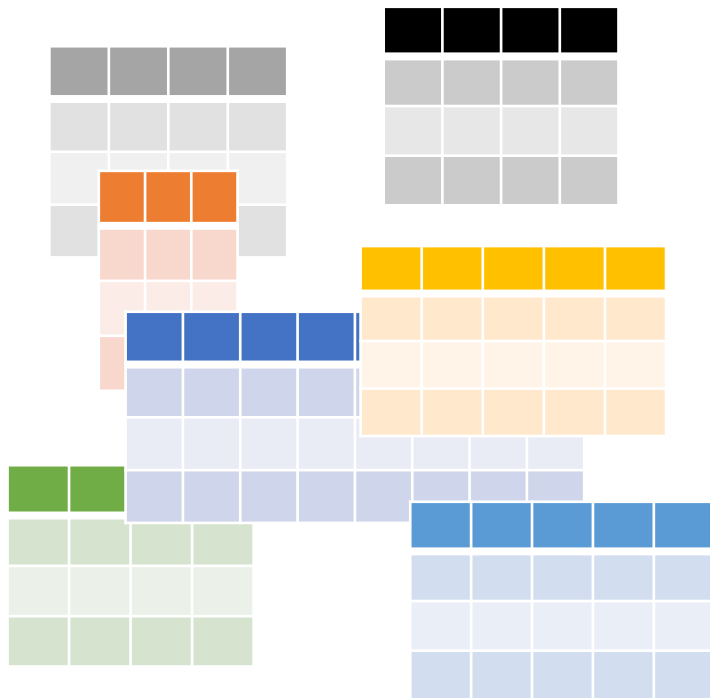
## Visualize data



# An unsupervised classification methodology of heterogeneous datasets based on MFA

Caroline Buridant, Camille Gaal, Antoine Menard, Arthur David, Sébastien Lê

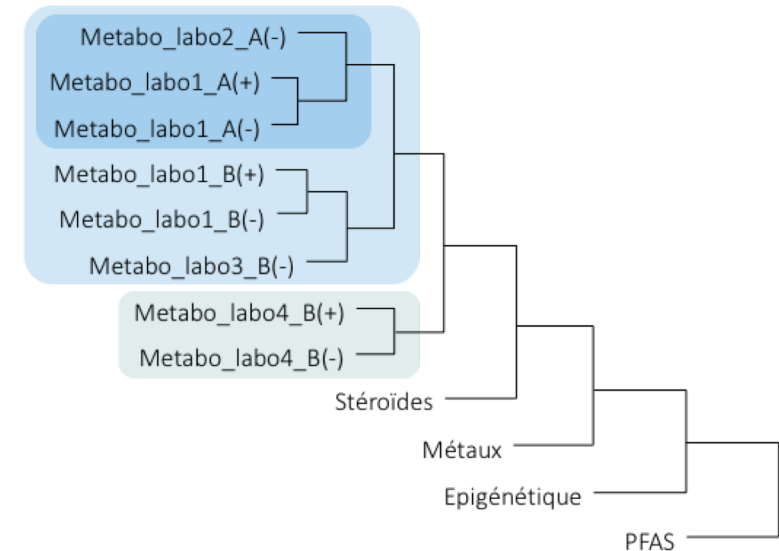
Multisource data is more and  
more frequent



## Datasets clustering

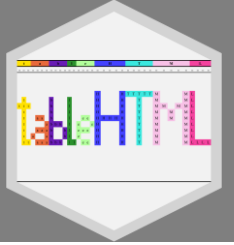
- ① Distance
- ② Aggregation criteria
- ③ Clustering algorithm

Application to epidemiological  
datasets



# RAdwords - Utilizing the Google Ads API with R

- R Interface for the Google Ads API
- Tool box for data driven online marketing
- R package on CRAN
- Poster Location F45



# Level up your tables with **tableHTML** in R

## Introduction

- This is a package for building CSS-ible HTML tables in an easy and intuitive way.
- These are compatible with any application that accepts HTML (e.g. shiny, rmarkdown).
- The main function `tableHTML` will convert a `data.frame` or matrix or any other object that can be converted into a `data.frame` into an HTML table
- Using the function on R Studio will print the table on the viewer otherwise it will use the default browser.
- The default tables are built without any CSS in order to allow for full flexibility in design
- The package has been developed so that the functions are chained with the `%>%` (pipe) operator which comes from the `magrittr` package

	mpg	cyl	disp	hp	drat
Mazda RX4	21	6	160	110	3.9
Mazda RX4 Wag	21	6	160	110	3.9
Datsun 710	22.8	4	108	93	3.8
Hornet 4 Drive	21.4	6	258	110	3.1
Hornet Sportabout	18.7	8	360	175	3.1
Valiant	18.1	6	225	105	2.8

	mpg	cyl	disp	hp	drat
Mazda RX4	21	6	160	110	3.9
Mazda RX4 Wag	21	6	160	110	3.9
Datsun 710	22.8	4	108	93	3.8
Hornet 4 Drive	21.4	6	258	110	3.1
Hornet Sportabout	18.7	8	360	175	3.1
Valiant	18.1	6	225	105	2.8

	mpg	cyl	disp	hp	drat
Mazda RX4	21	6	160	110	3.9
Mazda RX4 Wag	21	6	160	110	3.9
Datsun 710	22.8	4	108	93	3.8
Hornet 4 Drive	21.4	6	258	110	3.1
Hornet Sportabout	18.7	8	360	175	3.1
Valiant	18.1	6	225	105	2.8

## Features

- Adding themes
  - There are three categories of themes to give the users a way to build nice looking tables very fast
- Adjusting the appearance
  - You can add / remove `rownames`
  - Add a specific CSS-class
  - Change the `width` of the columns
  - Add `second_headers` and `row_groups`
  - Round numeric columns in the table
  - Provide a replacement string for NA values
  - Add a `caption` / `footer`
  - Remove the `borders`
- Apply CSS without writing CSS
  - Using the `add_css_*()` family of functions
  - Including conditional formatting on columns
- Can be used with Rmarkdown and Shiny
- The tables can be exported as images using `tableHTML_to_image()`

# Next on stage :

Groups I & G: Environment/spatial/Ecology/resource management



**I52** - Gi-Seop Lee

**I53** - Nicolas Raillard

**I54** - Wayne Jones

**I55** - Bénédicte Fontez

**I56** - Madalina Olteanu

**G59** - Maximilian Leodolter

**G60** - Benoît Génot

**G61** - Olivier Delaigue

**G63** - Ellen Webborn

**G64** - Ria Van Hecke

## Get ready ! Groups H, J

On stage in 5 minutes

**H65** - Celine Monteil, Fabrice Zaoui

**H67** - Victor Aguirre

**H68** - Andrés Lopez-Lopera

**H69** - Sanela Omerovic

**H70** - Michael Dietze

**H71** - Criscely Luján

**J75** - Yeonjeong Kim

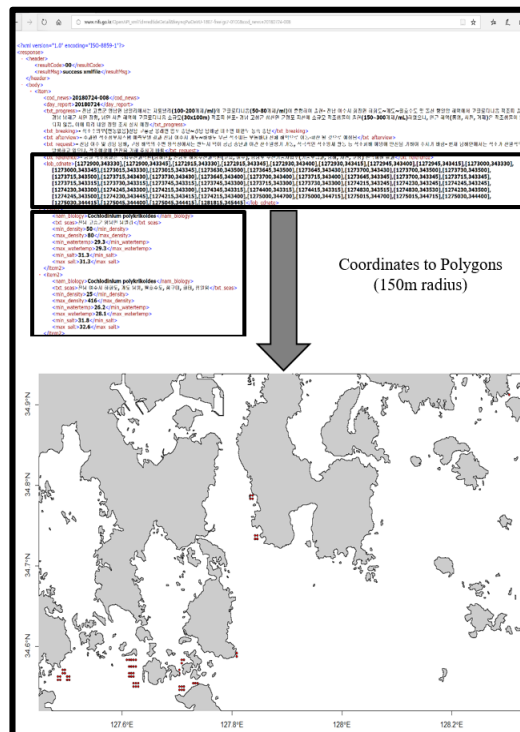
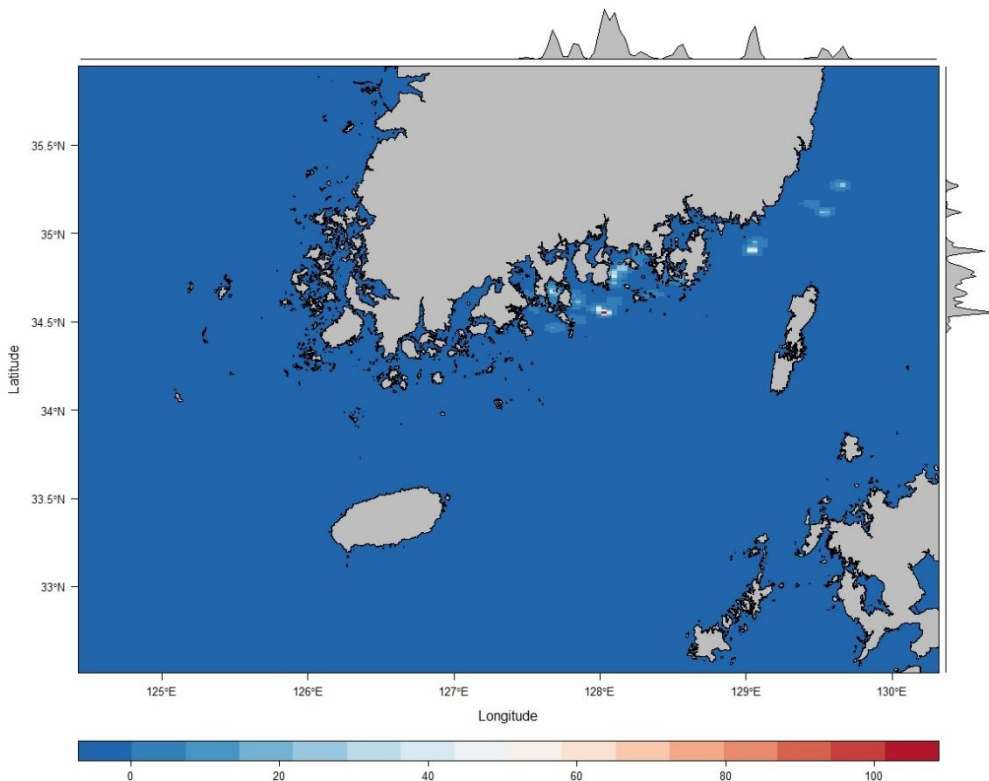
**J76** - Emilio L. Cano

**J78** - Emmanuelle Claeys, Myriam Maumy-Bertrand

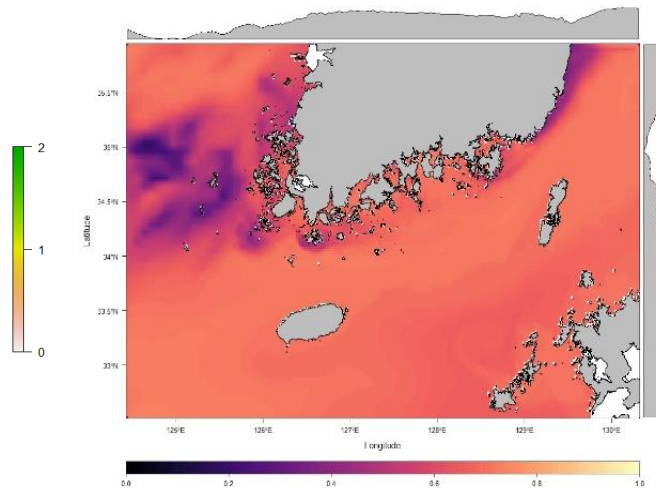
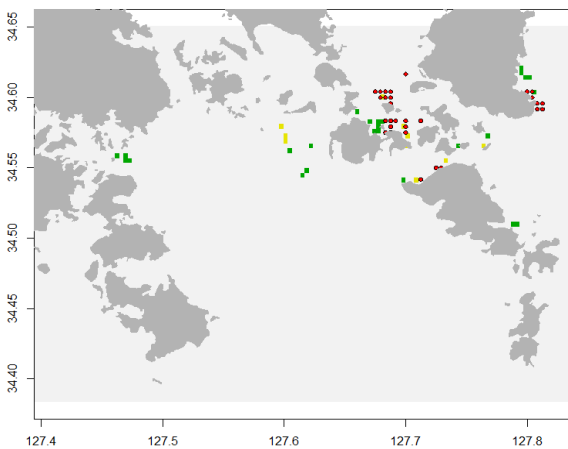
**J79** - Teresa Gonzalez-Arteaga

# Integrated Operational Modeling of the Harmful Algal Blooms(HABs) Using R

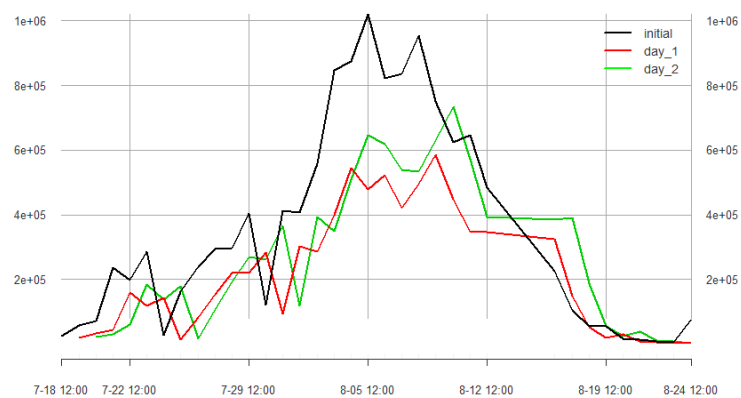
- G.S. Lee, H.Y. Cho, J.Y. Choi, D.H. K



20180801 Data(Polygon) & Prediction(raster)



Time Series of Number of Particles 2013 Jul-Aug 2013-07-18 12:00:00 / 2013-08-24 12:00:00



# Using R and Shiny to access and analyse hindcast database of sea-states

Nicolas Raillard, Marc Prevosto

Laboratoire Comportement des Structures en Mer, IFREMER, Brest

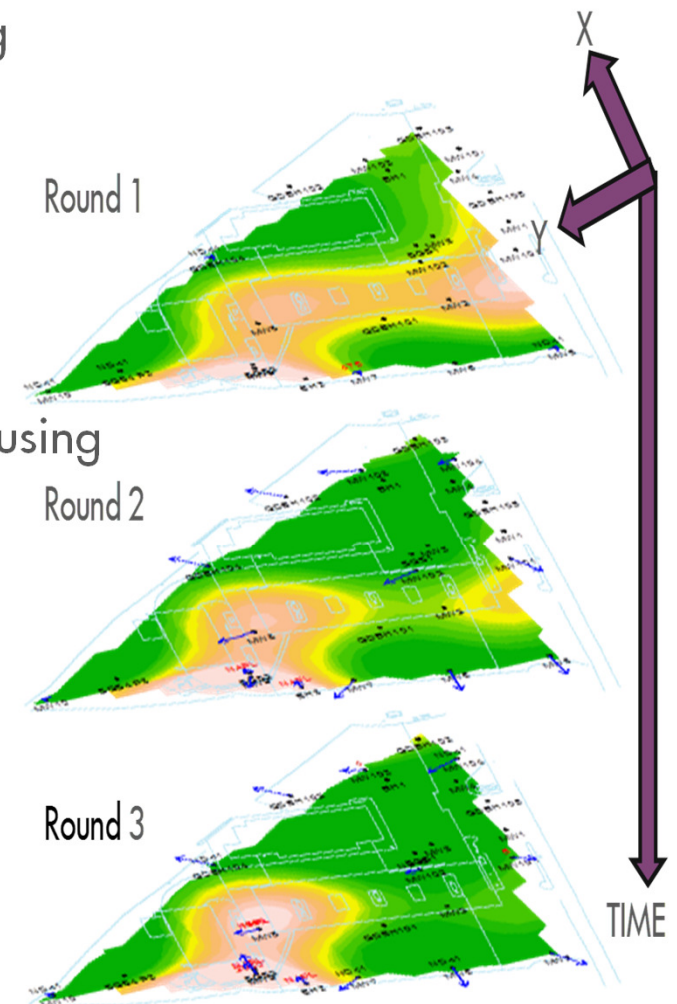
- HOMERE data base:
  - Hindcast numerical model;
  - 110 000 data points;
  - 4 To of data;
  - 1 NetCDF files per month : difficult to extract and study time series;
  - Many user-developed tools (Matlab, R , python...)
- Objective :
  - provide a user-friendly web interface to data;
  - Unifying access to data.
- Application contents:
  - **Exploratory data analysis**
    - Time series plots;
    - Histograms, scatter plots...
  - **Conditional distributions;**
    - Interface with gamlss package
  - **Extreme Values analysis;**
    - Interface with POT package
- Joint work with :





## GWSDAT – GroundWater Spatiotemporal Data Analysis Tool

- A decision support tool for the analysis and reporting of groundwater monitoring data.
- Available as an R package on CRAN and GitHub.
- Online Shiny App hosted at [www.gwsdat.net](http://www.gwsdat.net).
- Primary functionality:
  - Analysis of historic trends in solute concentrations using smoothing statistics and significance tests.
  - Groundwater flow velocity estimation.
  - Animations of solute plume dynamics with mass, concentration and area tracked through time.
- Spatiotemporal analysis gives more information for fewer sampling points compared to spatial methods.



# Determination of the correlation of different factors observed in the field and by Sentinel2 using the getSpatialData package under R

- ▶ Request to the Copernicus web service using the "getSpatialData" package
- ▶ Management and computation using the "tidyverse" philosophy on Spatial Data
- ▶ Comparison between Sentinel-2 and hand-collection data

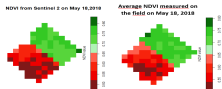


Figure 1: Comparison of the NDVI measured and calculated using sentinel on May 18, 2018

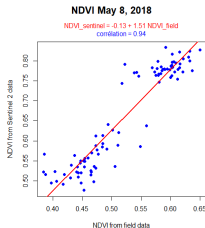


Figure 2: Correlation of the NDVI measured and calculated using sentinel on May 8, 2018

- ▶ Measurement of different indexes such as NDVI, NDWI or Moisture Index



# Focal distances and distortion coefficients : assessing the individual perception of multiscalar segregation

Madalina Olteanu<sup>1,2,3</sup>, Julien Randon-Furling<sup>1,3</sup>, William Clark<sup>4</sup>

<sup>1</sup> SAMM, Université Paris 1 Panthéon-Sorbonne

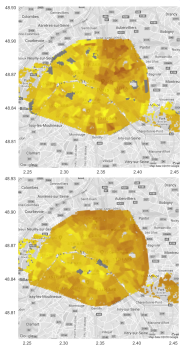
<sup>2</sup> MaIAGE, INRA

<sup>3</sup> Institut des Migrations

<sup>4</sup> Department of Geography, University of California, Los Angeles

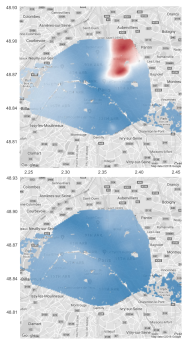


A new tool for visualizing segregation across space and at all scales!



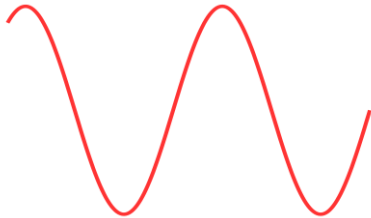
Local densities of Chinese and Algerian communities (log scale)

- Quantify how distorted a city looks like, from each location, and at all scales
- Build individual *trajectories* encoding the aggregated distribution of some variable of interest (multiscalar *fingerprint* of the city)
- Eventually, all *trajectories* converge to the city average
- Compute *focal distances* : the instants of convergence of each trajectory
- Integrate *focal distances* over all convergence thresholds and get a *distortion coefficient* for each trajectory



Distortion coefficients maps

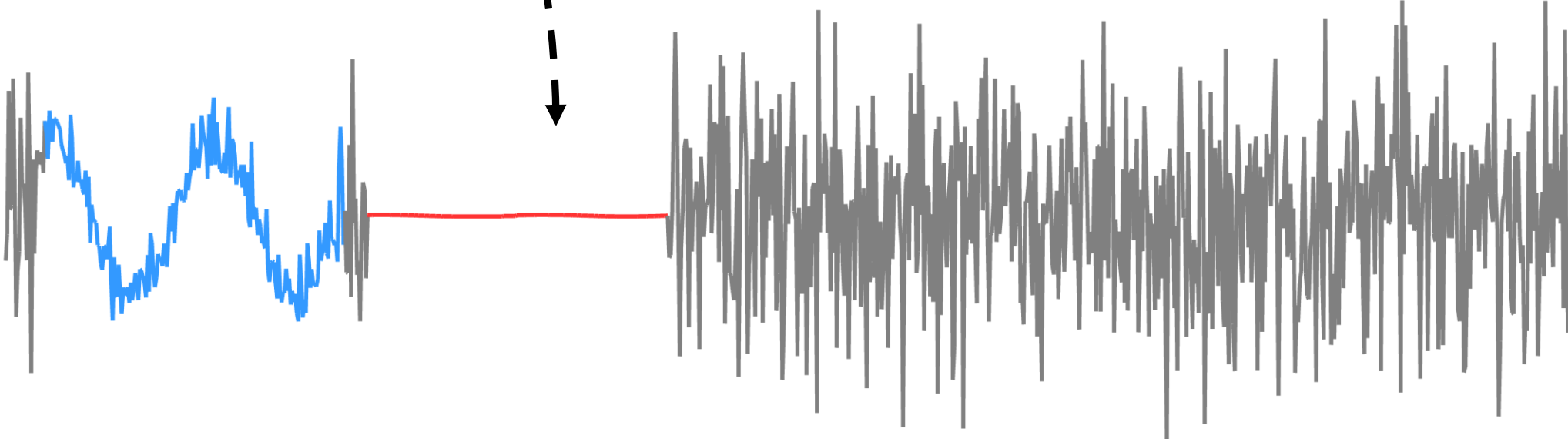
# runDTW: Detect Query Patterns in Long Time Series

Query: 

runDTW → Smartphone Accelerometer:  
Subway or Walk or ...?

Closest!  
How?

Long Time Series:





# Cross-referencing catchment data

how R can provide essential tools for the development of hydrological models for flood prediction

IRSTEA – Hydrology Research Group (HYCAR) – Antony, France



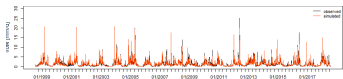
Hydrology

Catchments boundaries

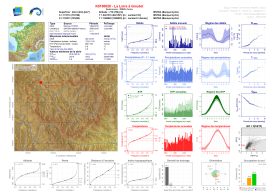
Climate/Morphology/Landcover

Data at basin scale

Hydrological model



Flood prediction



Main characteristics of catchments



# airGR & airGRteaching

two packages for rainfall-runoff modeling & teaching hydrology

Olivier Delaigue<sup>1</sup>, Guillaume Thirel<sup>1</sup>

Laurent Coron<sup>2</sup>, Pierre Brigode<sup>3</sup>

(1) IRSTEA, Hydrology Research Group, Antony, France

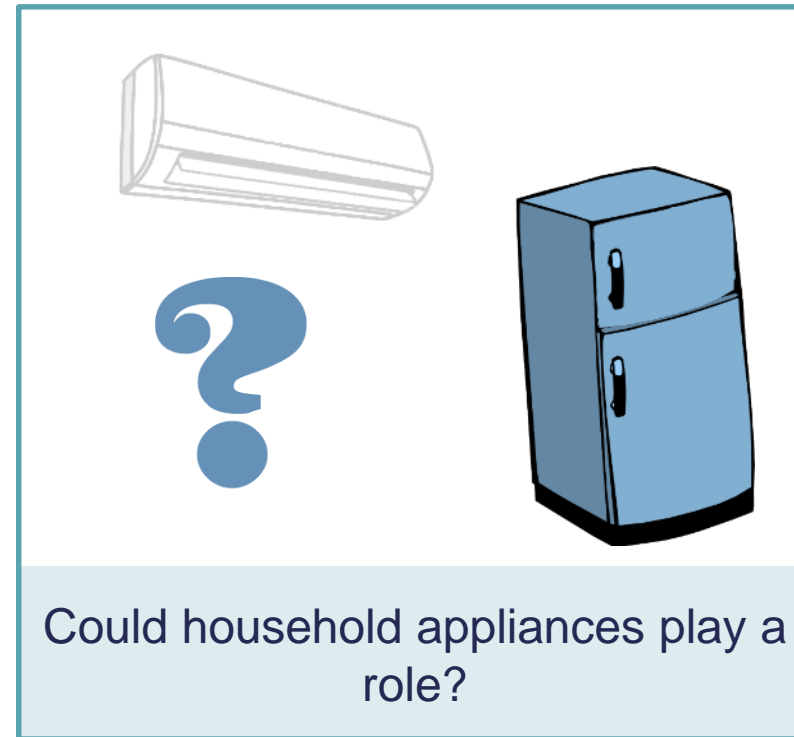
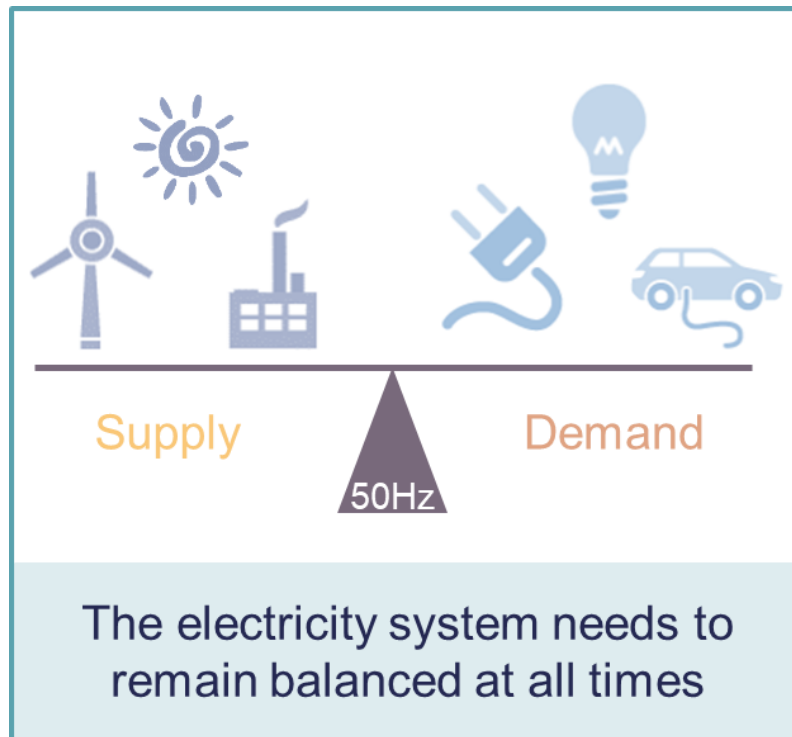
(2) EDF – PMC Hydrometeorological Center – Toulouse, France

(3) Nice-Sophia-Antipolis University – Géoazur UMR 7329 – Sophia-Antipolis, France

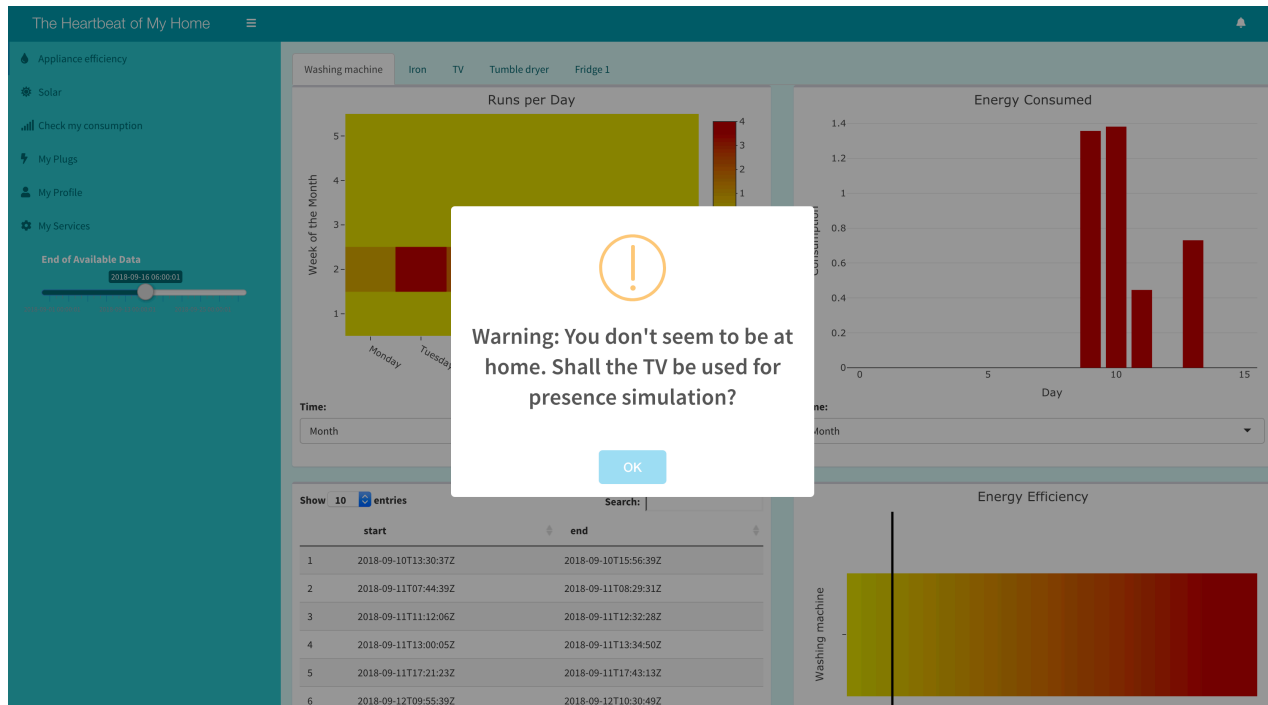


# Using Fridges to Balance the Electricity Grid

Ellen Webborn



# The Heartbeat of My Home – A Shiny App Making Households Smarter with Modern Classification and State Detection Algorithms



- State detection of smart plug data by clustering
- Device classification via Random Forest
- Energy breakdown of household
- Alarming system
- Similar home comparison

# Next on stage :

Groups H & J: Statistical Packages/Methods/economics



**H65** - Celine Monteil, Fabrice Zaoui

**H67** - Victor Aguirre

**H68** - Andrés Lopez-Lopera

**H69** - Sanela Omerovic

**H70** - Michael Dietze

**H71** - Criscely Luján

**J75** - Yeonjeong Kim

**J76** - Emilio L. Cano

**J78** - Emmanuelle Claeys, Myriam Maumy-Bertrand

**J79** - Teresa Gonzalez-Arteaga

## Get ready ! Group K

On stage in 5 minutes

**K81** - Guyliann Engels

**K82** - Henna Kettunen

**K83** - David Granjon

**K84** - Katie Sasso-Schafer

**K85** - Matthias Gehrke

**K86** - Guillaume Devailly

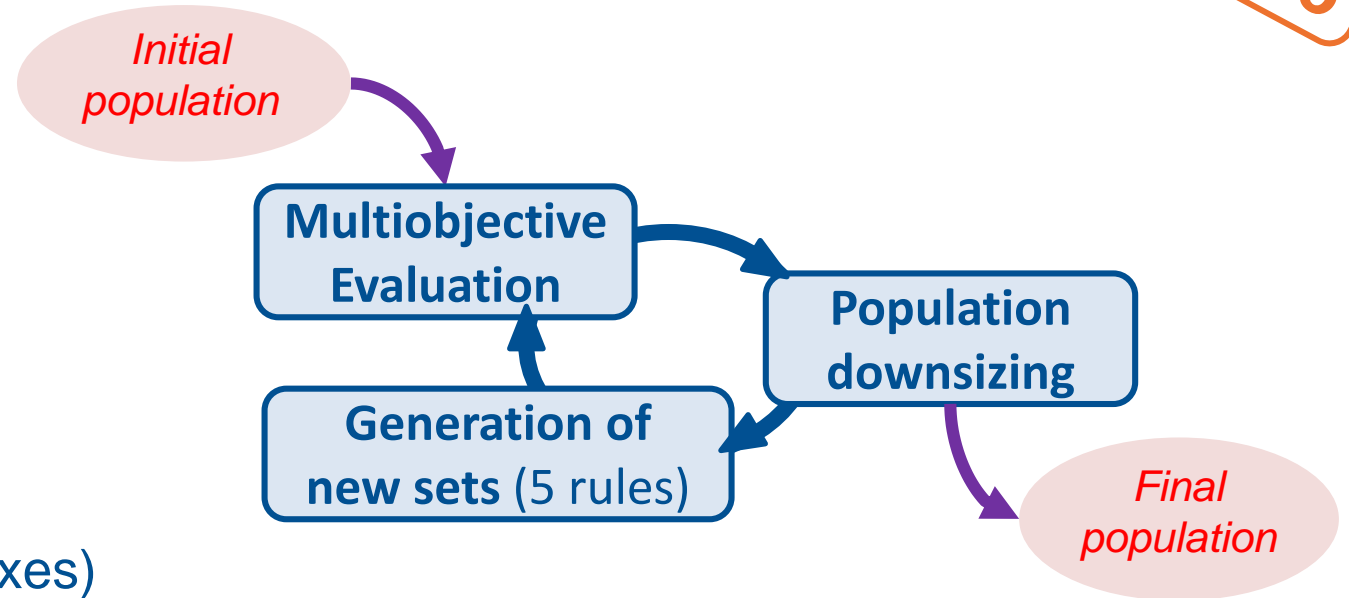
**K87** - Chia-Yi Yen, Mia Huai-Wen Chang, Chung-Hong Chan

**K88** - Filippo Chiarello

# A fast multi-objective optimizer: the **caRamel** R package

Poster H65

- **Multi-objective** optimizer
- Designed for **calibration** of **complex model**  
(Hydrology, Environmental Modelling, ...)
- Algorithm combines two methods:
  - **MEAS** (directional search based on simplexes)
  - **NSGA-II** (vectors classification)

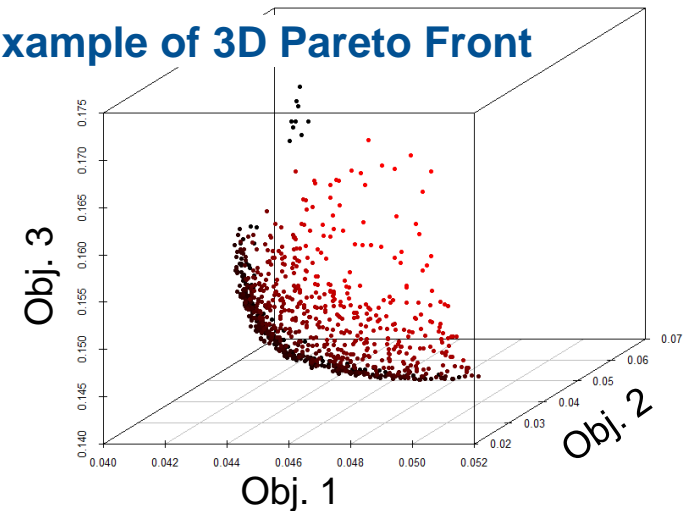


✓ **Package availability**

<https://cran.r-project.org/package=caRamel>

<https://github.com/fzao/caRamel>

Example of 3D Pareto Front





# H67

## BAYESDEF: Graphical Interface for the Bayesian Analysis of DSD

Víctor Aguirre, Statistics Department, ITAM; Sofia Huerta, UV; Edgar López, ITAM

Run	$x_1(A)$	$x_2(B)$	$x_3(C)$	$x_4(D)$	$x_5(E)$	$x_6(F)$
1	0	1	-1	-1	-1	-1
2	0	-1	1	1	1	1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	-1	-1	0	1	-1	-1
6	1	1	0	-1	1	1
7	-1	1	1	0	1	-1
8	1	-1	-1	0	-1	1
9	1	-1	1	-1	0	-1
10	-1	1	-1	1	0	1
11	1	1	1	1	-1	0
12	-1	-1	-1	-1	1	0
13	0	0	0	0	0	0

Table 1. DSD for  $m = 6$  factors

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^m \beta_{jk} x_{ij} x_{ik} + \sum_{j=1}^m \beta_{jj} x_{ij}^2 + \varepsilon_i$$

$$y_i = 20 + C + F + \varepsilon_i$$

$$\varepsilon_i \text{ i.i.d. } N(0, 1)$$

Stepwise Fit for Y2

Stepwise Regression Control

Stopping Rule: P-value Threshold

Probto Enter: 0.1  
Probto Leave: 0.1

Direction: Forward

Rules: Combine

JMP 11.1.1

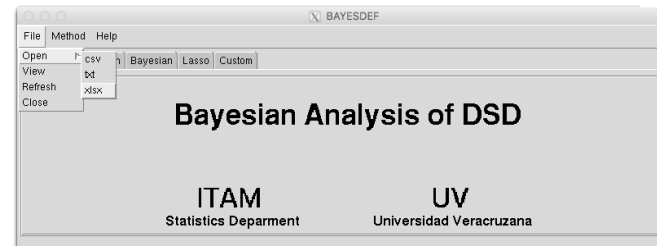
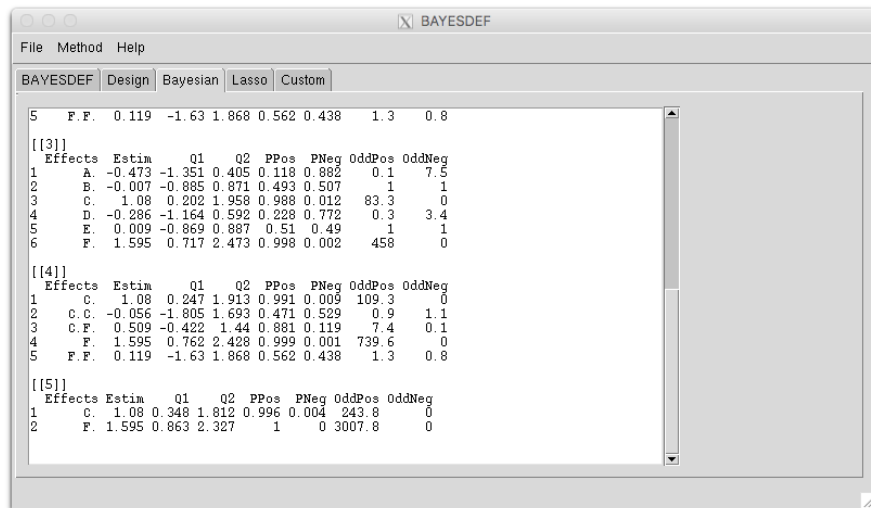
SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
0.0011051	2	0.0235066	1.0000	0.9999		11		-54.1739

Current Estimates

Lock	Entered	Parameter	Estimate	DF	SS	F Ratio	Prob>F
		Intercept	18.870468	1	0	0.000	1
		X1	-0.473	3	3.424532	2066.654	0.00048
		X2	-0.007	2	1.851487	1675.369	0.0006
		X3	1.08	4	17.52868	7930.656	0.00013
		X4	-0.286	2	1.345914	1217.638	0.00062
		X5	0	1	0.00081	2.746	0.34573
		X6	1.595	3	30.35222	18310.03	5.46e-5
		X1*X2	0	1	0.000295	0.364	0.65427
		X1*X3	0.58329268	1	0.978908	1771.584	0.00056
		X1*X4	0	1	0.000295	0.364	0.65427
		X1*X5	0	2	0.001105		
		X1*X6	0.2777439	1	0.171542	310.449	0.00321
		X2*X3	0	1	0.000295	0.364	0.65427
		X2*X4	0	1	0.000295	0.364	0.65427
		X2*X5	0	0			
		X2*X6	0	1	0.000295	0.364	0.65427
		X3*X4	-0.4872561	1	0.527954	955.488	0.00104
		X3*X5	0	2	0.001105		
		X3*X6	1.21670732	1	4.259329	7708.343	0.00013
		X4*X5	0	2	0.001105		
		X4*X6	0	1	0.000295	0.364	0.65427
		X5*X6	0	2	0.001105		
		X1*X1	0	1	0.000295	0.364	0.65427
		X2*X2	1.48363669	1	1.850997	3349.851	0.0003
		X3*X3	0	1	0.000295	0.364	0.65427
		X4*X4	0	1	0.000295	0.364	0.65427
		X5*X5	0	2	0.001105		
		X6*X6	0	1	0.000295	0.364	0.65427

???

A, B, C, D, F, AC, AF, CD, CF, and BB



$$P(A)/(1 - P(A))$$

1. Aguirre, V., (2016) "Bayesian analysis of definitive screening designs when the response is nonnormal", Applied Stochastic Models in Business and Industry 32: 440-452.
2. Aguirre VM, Huerta NS, Lopez E (2017). BAYESDEF: Bayesian Analysis of Definitive Screening Designs with R. R package version 0.1.0, URL <https://CRAN.R-project.org/package=BAYESDEF>
3. JMP (2013) Statistical Software version 11.1.1. SAS Institute Inc.
4. Jones, B. and Nachtshiem, C. J. (2011). "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects" Journal of Quality Technology 43:1-14.
5. R core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2014. ISBN 3-900051-07-0, <https://www.r-project.org/>.
6. Tibshirani R. (1966) "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society B 1996; 58:267-288.
7. Verzani J. (2014a). gWidgets: gWidgets API for building toolkit-independent, interactive GUIs. Based on the iwidgets code of Simon Urbanek, suggestions by Simon Urbanek, Philippe Grosjean and Michael Lawrence. R package version 0.0-54. \URL <https://CRAN.R-project.org/package=gWidgets>.

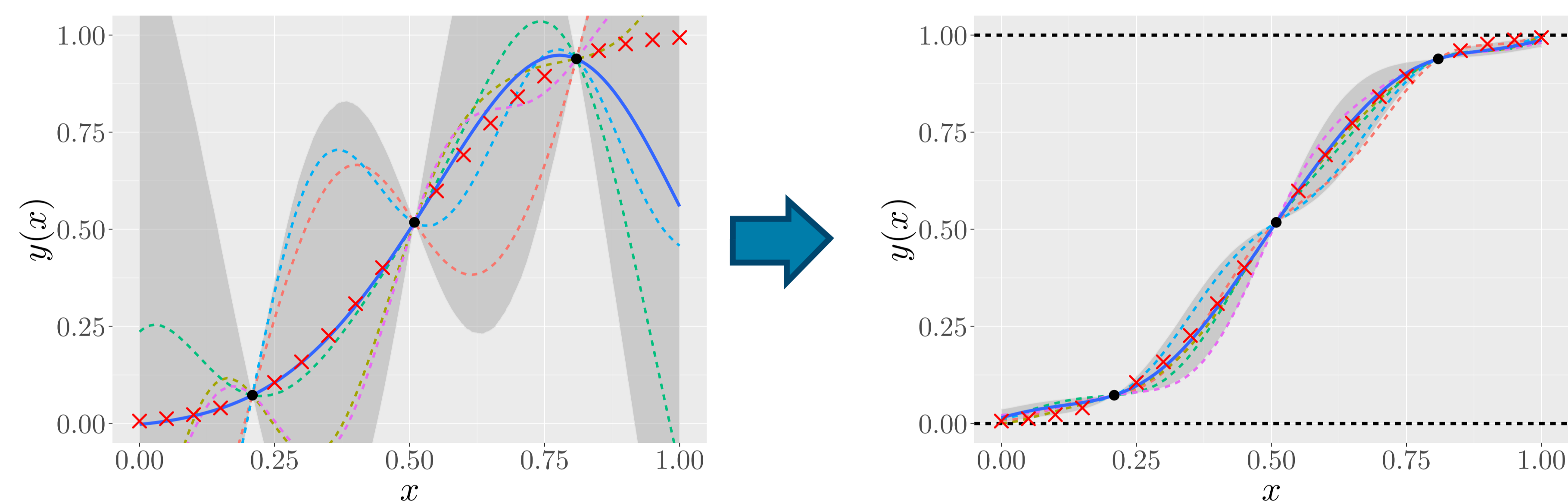
## Gaussian process (GP) models under linear inequality constraints

Let  $Y$  be a GP on  $\mathbb{R}$  given by

$$Y(x) \sim \mathcal{GP}(m, k), \quad (1)$$

with mean function  $m$  and covariance function  $k$ .

**Goal:** To ensure linear inequality constraints (e.g. boundedness, monotonicity, convexity) everywhere in the space



(a) Unconstrained GP

(b) Constrained GP

Figure 1: Examples GP regression models.

## Main functionalities as S3 methods

Method Name	Description
<code>create</code>	Creation function of GP models under inequality constraints.
<code>lineqGPOptim</code>	Covariance parameter estimation under inequality constraints.
<code>predict</code>	Prediction of the objective function at new points.
<code>simulate</code>	Simulation of GP models under inequality constraints.
<code>plot, ggplot</code>	Plot for a constrained GP models.

## References

- [1] H. Maatouk and X. Bay, "Gaussian process emulators for computer experiments with inequality constraints," *Mathematical Geosciences*, vol. 49, no. 5, pp. 557–582, 2017.
- [2] A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant, "Finite-dimensional Gaussian approximation with linear inequality constraints," *SIAM/ASA J. on Uncertainty Quantification*, vol. 6, no. 3, pp. 1224–1255, 2018.
- [3] F. Bachoc, A. Lagnoux, and A. F. López-Lopera, "Maximum likelihood estimation for Gaussian processes under inequality constraints," *ArXiv e-prints*, Apr. 2018.

## 2D Nuclear Criticality Example

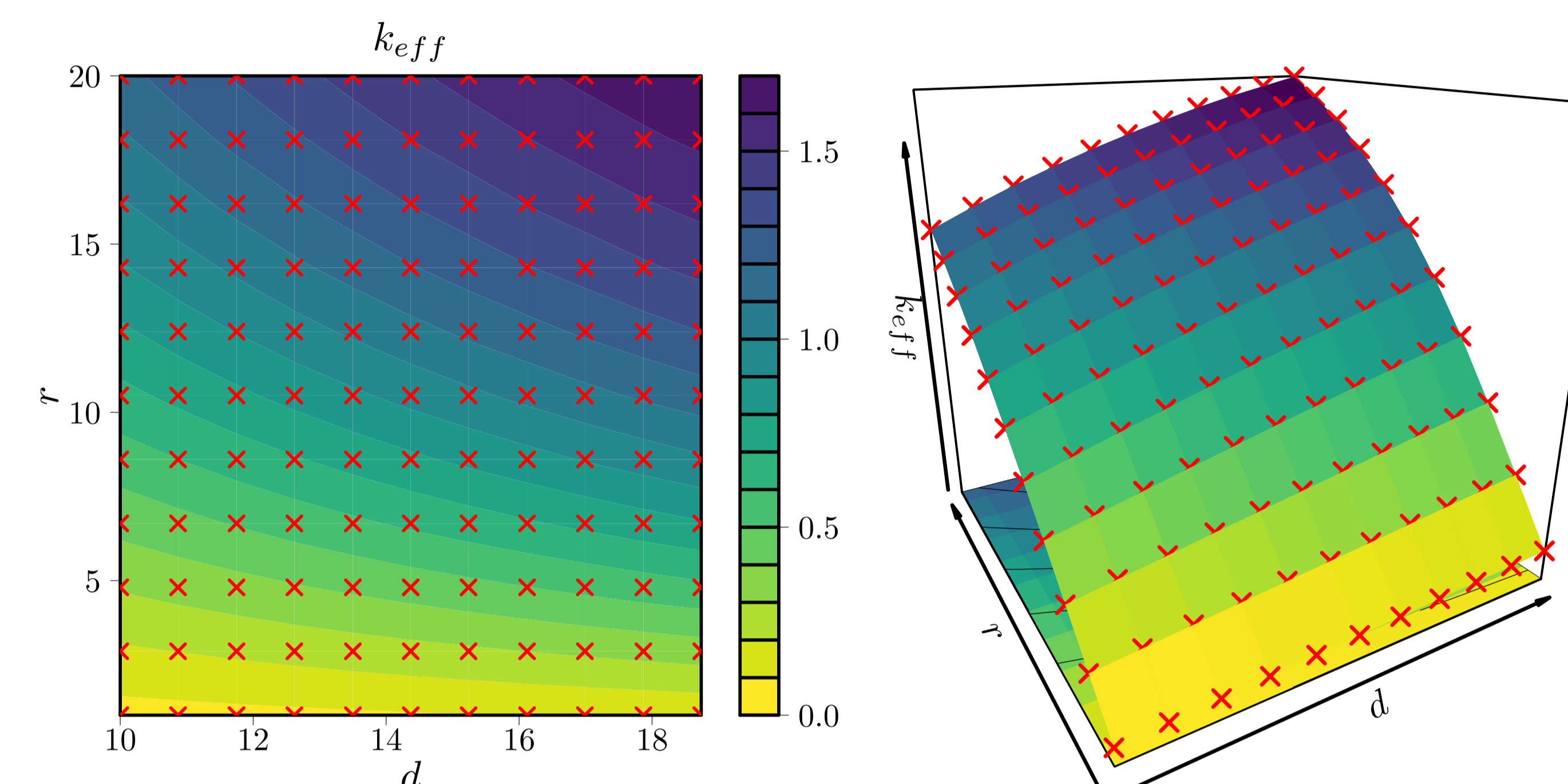
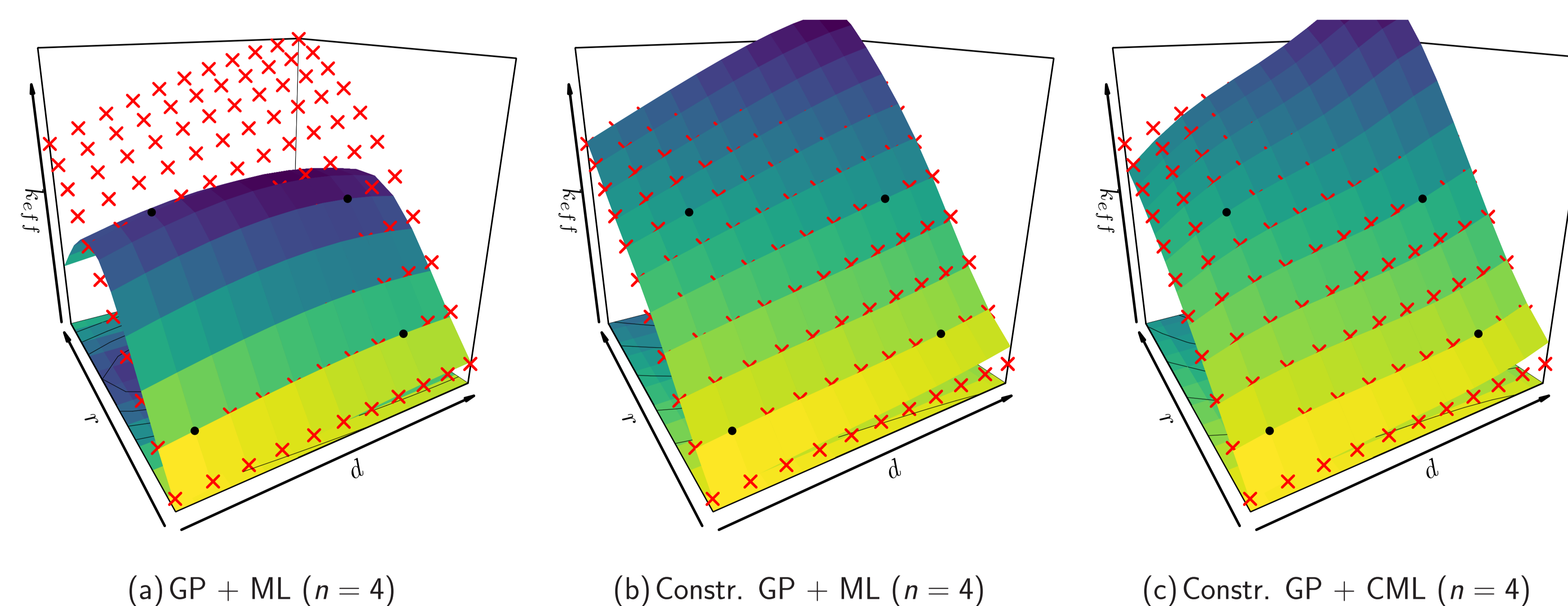


Figure 2: Nuclear criticality safety dataset.  $k_{eff}$  is positive and non-decreasing.



(a) GP + ML ( $n = 4$ )

(b) Constr. GP + ML ( $n = 4$ )

(c) Constr. GP + CML ( $n = 4$ )

Figure 3: 2D GP regression models using  $n = 4$  training points. ML: Maximum Likelihood. CML: Constrained Maximum Likelihood.

## Further Implementations

GP Model	Description
<code>lineqDGP</code>	Framework proposed in (Maatouk et al., 2017) with <b>derivative information</b> .
<code>lineqGP</code>	<b>Derivative-free</b> framework proposed in (López-Lopera et al., 2018).
<code>lineqAGP</code>	<b>Additive GP models</b> under linear inequality constraints.

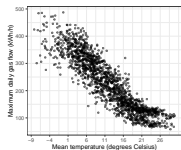
## Acknowledgment

- This work was funded by the chair of applied mathematics OQUAIDO.

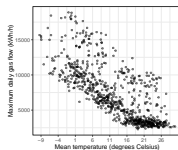
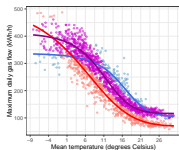
# flexmixNL: an R package for mixtures of Generalized Nonlinear Models

Sanela Omerovic and Herwig Friedl (Graz University of Technology, Austria)

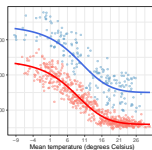
- Deal with heterogeneity in data due to latent classes with nonlinear mean patterns
- Nonlinear mean functions as symbolic description compatible with `nls()` and `gnm()`
- Gamma and normal distribution available as error distributions
- Extension of package **flexmix** (developed by Bettina Grün and Friedrich Leisch)
- Two- and Three-Component Gamma Mixture Models with Sigmoid Mean Functions:



Data Set 1



Data Set 2



### H70

# The tool: sandbox

# Assited by: EMMAgeo

**Parameters (specific/global)**

- GRAINSIZE**  
type = "rnorm"  
mean = f(d)  
sd = f(d)
- DENSITY**  
type = "exact"  
value = f(d)
- DOSE RATE**  
type = "rnorm"  
mean = f(d)  
sd = f(d)
- WATER CONTENT**  
type = "rnorm"  
mean = f(d)  
sd = f(d)

**Grain**

**Populations**

- P1
- P2
- P3

**Rule Book**

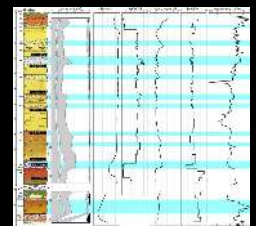
**Rules**

**Analysis functions**

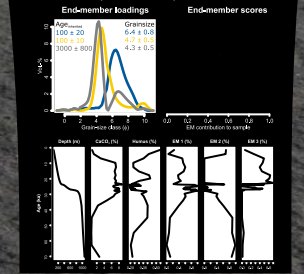
- MAKE\_SAMPLE()
- MAKE\_SIEVING()
- MAKE\_ALIQUOT()
- MAKE\_IRRADIATE()



### Building a section



Loess section Gleina (Germany) with updated laser based grain-size data, modelled by 3 end-members (described by distribution functions), and interpolated to equal intervals.



**Process end-members** → **Archive composition** → **The simulated sediment** → **EMM member** → **EMM loadings**

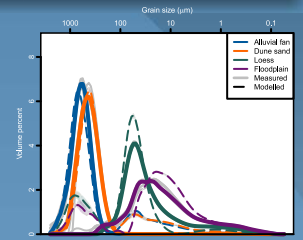
**Process end-members**: Relative frequency vs Grain-size class (µ)

**Archive composition**: Heatmap of relative frequency vs Grain-size class (µ)

**The simulated sediment**: Grain size distributions vs Grain-size class (µ)

**EMM member**: Sample composition heatmap vs Grain-size class (µ)

**EMM loadings**: Generic distributions heatmap vs Grain-size class (µ)



**Population...** The most basic, coherent element in the entire model. A set of sediment grains that share a common characteristic.

**Parameter...** Parameters are used to describe populations. They can be seen as "horizontal" or "thematic" definition.

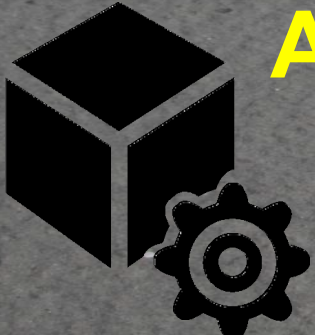
**Rule...** Rules are used to describe how parameters behave with depth. They can be regarded as "vertical" definition of a sediment deposit.

**Rule Book...** A rule book is the combination of parameters ("horizontal" definition) and rules ("vertical" definition) into one coherent reference book.

**Analysis functions...** Once a virtual sediment deposit is defined by a rule book, this deposit can be "harvested". Analysis function use information stored for each grain.

# Applications:

Geochemistry  
Age determination



# Discussions:

Speed things up  
Inverse problems  
Solving diff. eqs.

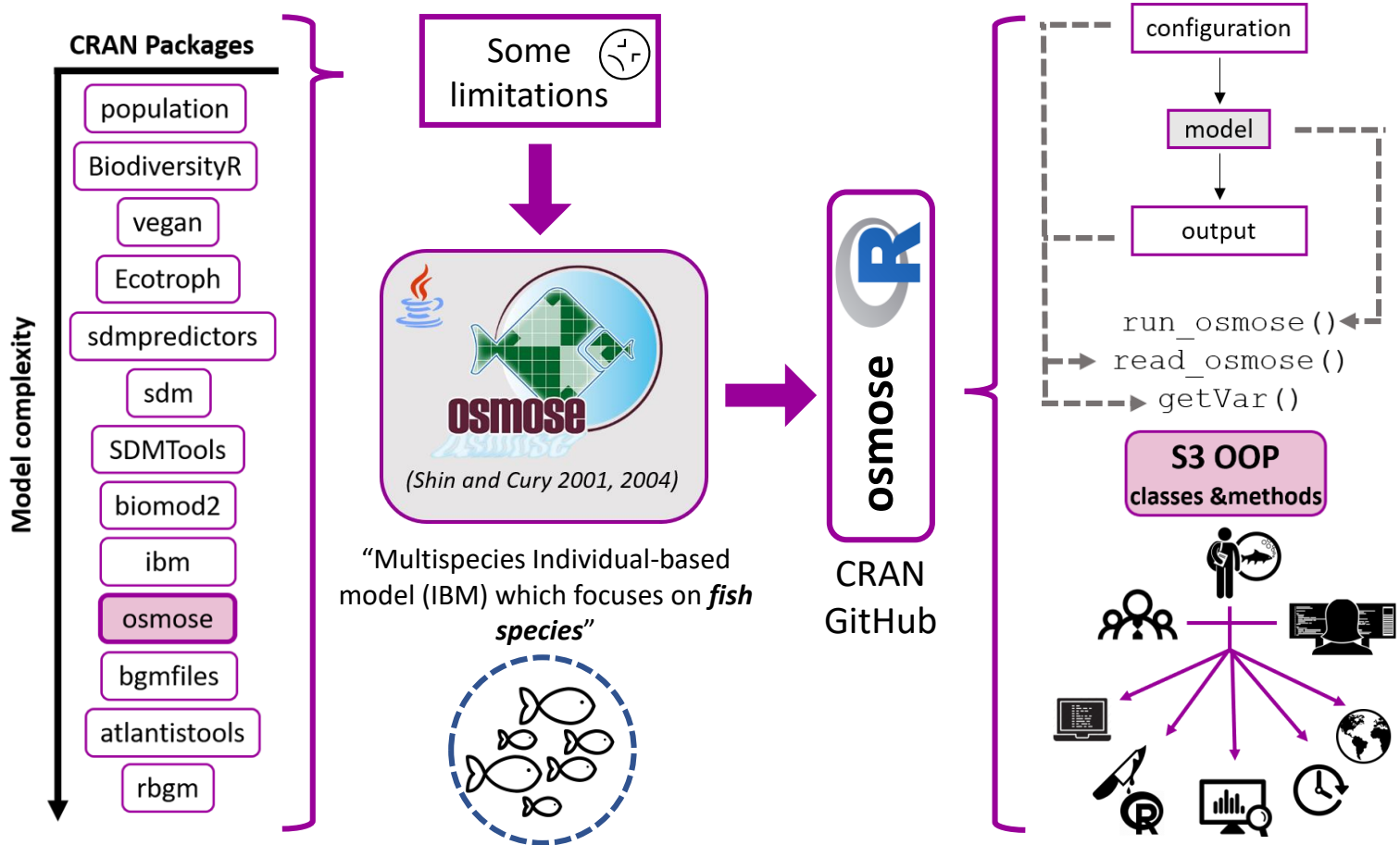
# An S3 approach to the analysis of computer simulations: an illustration with the marine ecosystem model OSMOSE

Criscely Luján, R. Oliveros-Ramos, N. Barrier and Y. Shin



@CriscelyLP

H71



Eunhyung Lee / Myungji Ko / Jeongmin Park / Yeonjeong Kim\*



# Forecasting and Visualizing Churn data

1. EDA
2. MODELING
3. R SHINY

# R and Shiny to support real estate appraisers: An expert algorithm implementation for Automated Valuation Models (AVM)

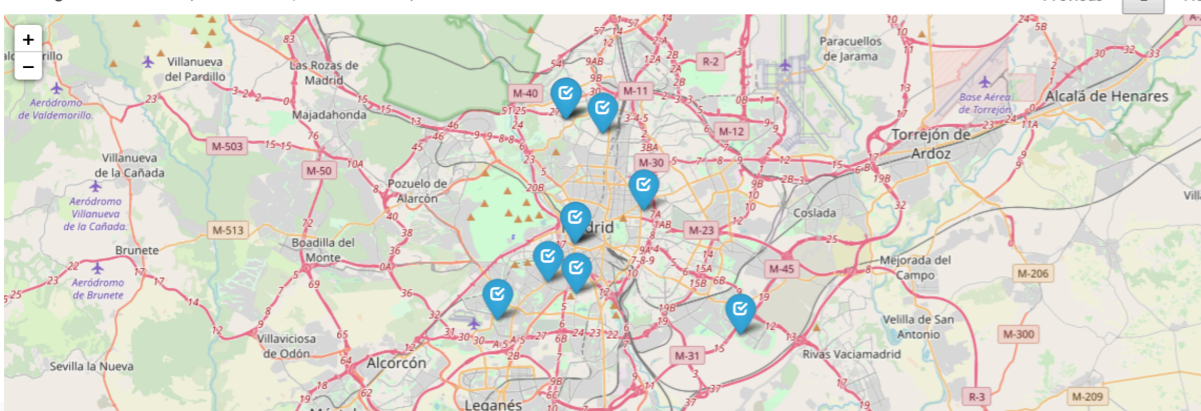
## AVM - Estimación de una vivienda

Configuración y código | Viviendas a estimar | Estimación vivienda | Cluster | Resumen estimadas | Estimadas municipio

Show 10 entries Search:

id_vivienda	unitario	n_dormitorios	n_banos	sup_construida	id_aire	id_calefaccion	ascensores	id_estado_conservacion
6244	2,874.41	3	2	97	No tiene	Sí tiene	Sí tiene ascensor	Adecuado a su edad

Showing 1 to 1 of 1 entries (filtered from 1,207 total entries) Previous 1 Next



Regular appraisal

## AVM appraisal

### Key features

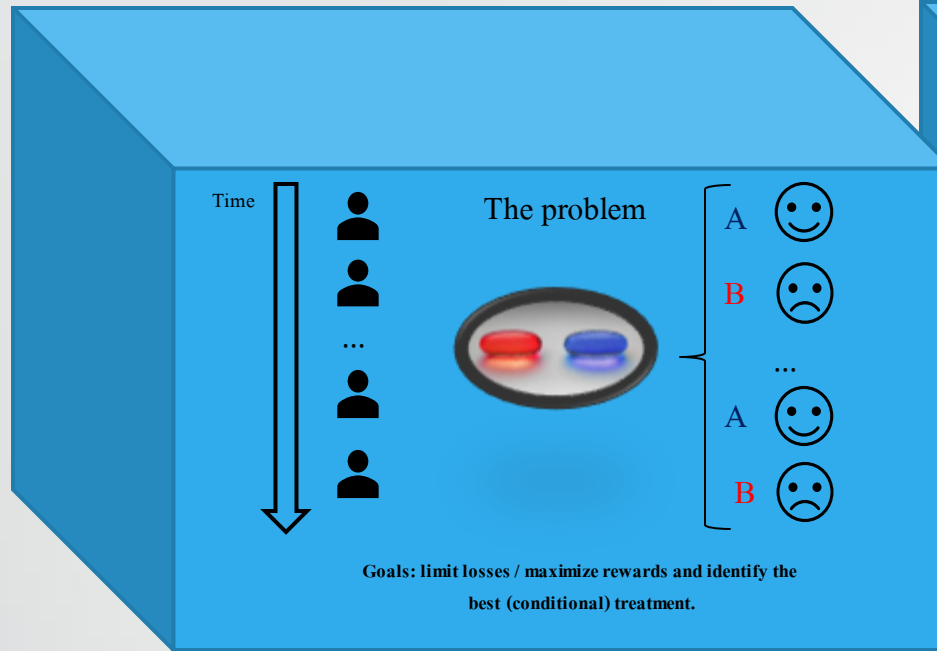
- Config files: steps mimicking appraisers
- Individual assessment
- Search comparables
- Summarise portfolio errors
- Clustering

```
for (e in as.character(etapas$etapa)){
  econf <- etapas %>% filter(etapa == e) %>% as.list()
  if (econf$usar){
    suppressWarnings(
      testigos <- testigos %>%
        bind_rows(datos %>%
          filter(!(id_vivienda %in% testigos$id_vivienda & id_provincia
            if (econf$id_estado_conservacion) id_estado_conservacion == 1 else TRUE,
            if (econf$geo_fiable) fiabilidad_geo == 1 else TRUE,
            if (econf$id_provincia) id_provincia == vivienda$id_provincia,
            if (econf$id_municipio) id_municipio == vivienda$id_municipio,
            if (econf$cod_postal) cod_postal == vivienda$cod_postal,
            if (econf$id_aparcamiento) id_aparcamiento == vivienda$id_aparcamiento,
            if (econf$id_sit_especial_altura & !is.na(vivienda$id_sit_especial_altura)) id_sit_especial_altura == vivienda$id_sit_especial_altura,
            if (econf$id_trastero) id_trastero == vivienda$id_trastero,
            if (econf$id_calefaccion) id_calefaccion == vivienda$id_calefaccion,
            if (econf$ascensores) ascensores == vivienda$ascensores,
            if (econf$id_estado_conservacion) id_estado_conservacion == vivienda$id_estado_conservacion)
        )
    )
  }
}
```



### Banking sector

- Accounting
- Compliance



Multi armed bandit (*Lai and Robbins, 1952*)

- Early stopping and limit the regret.
- Explore and exploit with reinforcement learning algorithms.
- Many algorithms are available depending on the context.

**Our Package : `bandit4abtest`**

Global and contextual strategies.

Our new algorithm : CTREE-UCB adapted for AB test.

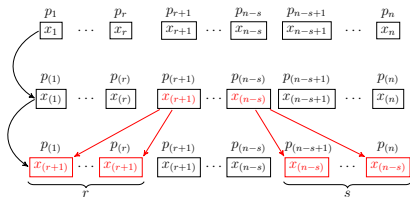




# An R package to deal with generalizations of weighted means and OWA operators

T. González-Arteaga, B. Llamazares and R. de Andrés Calle

- The R package **WEMOWA** (**WE**ighted **M**eans and **OWA** operators) is introduced to do computations in relation to Winsorized weighted means, SUOWA and Semi-SUOWA operators
- **Winsorized weighted means** are presented in this poster



# Next on stage :

## Group K: Community/education

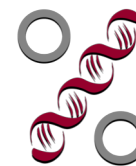
- K81** - Guyliann Engels
- K82** - Henna Kettunen
- K83** - David Granjon
- K84** - Katie Sasso-Schafer
- K85** - Matthias Gehrke
- K86** - Guillaume Devailly
- K87** - Chia-Yi Yen, Mia Huai-Wen Chang, Chung-Hong Chan
- K88** - Filippo Chiarello



### Get ready ! Group L

On stage in 4 minutes

- L89** - Brandon Allen
- L90** - Mikaela Miller
- L92** - Laurent Pantera
- L93** - Jean-François Rey
- L94** - Luke Rasmussen, Eric Whitley
- L96** - Hong Ooi



Collaborative  
works



Fully configured  
virtual machine



Online  
course



## Better learning of data science in a biology curriculum by using R, RStudio, learnr & Github Classroom

<http://biodatascience-course.sciviews.org>

43 students

187 Github repositories

High participation rate

High success rate

# Teaching data analysis with R as a part of an upper secondary school science project



*Henna Kettunen, Ville Tilvis, Elisa Mehtälä*  
*Helsingin matematiikkalukio*

## WHAT?

- We taught basics of data analysis with R to a group of 24 mathematically gifted Finnish upper secondary school students.

## HOW?

- The students collected an empirical ecological data set and analysed it with a linear regression model in R during a summer school week in June 2018.

## WHY?

- The students got a first hands-on experience of empirical research and statistical analysis, demonstrating the important role of statistics in ecological research.
- Data analysis skills are increasingly needed in post-secondary studies in all fields.



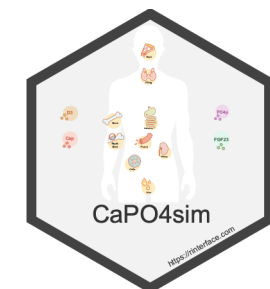
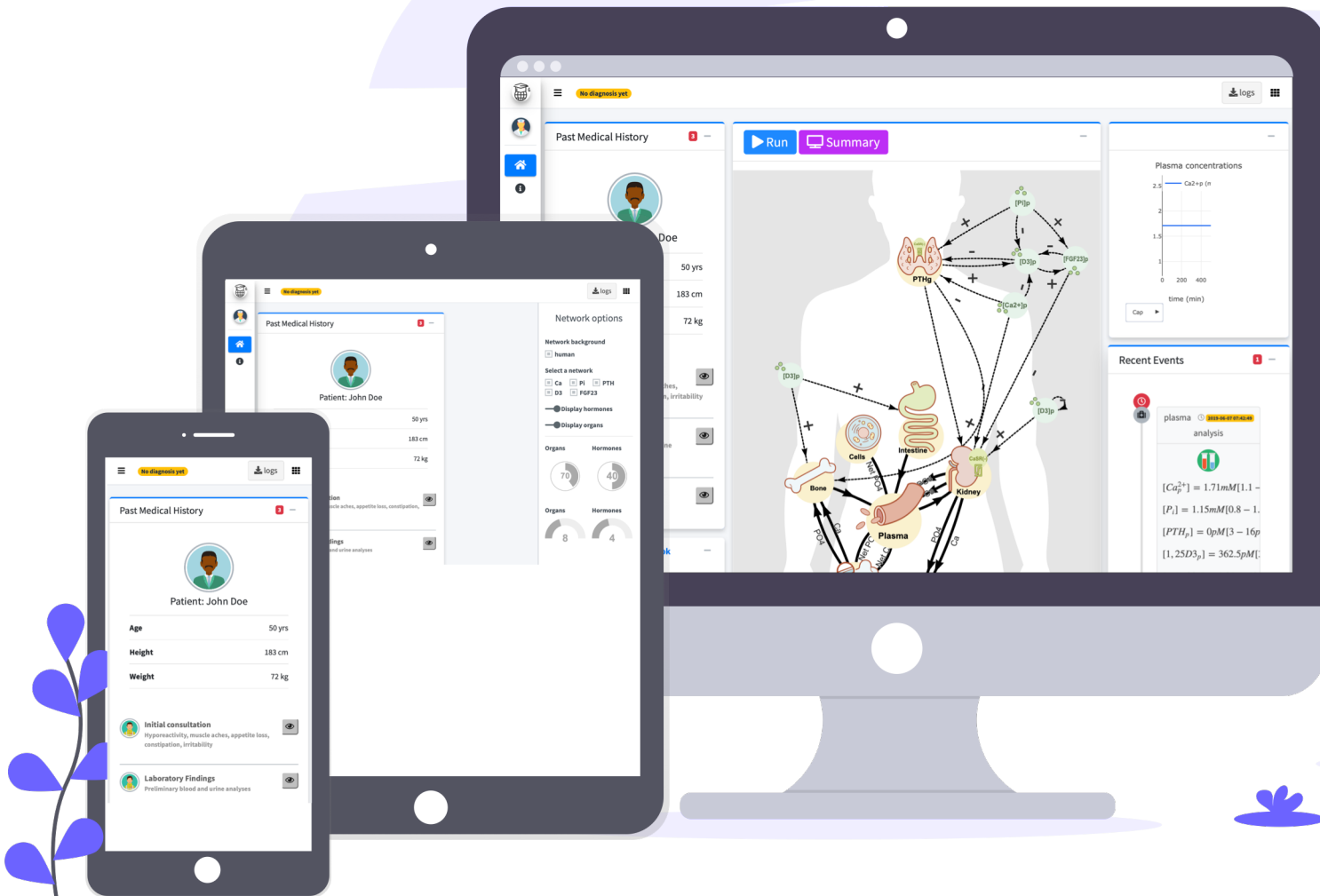
# Shiny Powered Virtual Patients

David Granjon, Olivier Bonny, François Verrey,  
Vartan Kurtcuoglu and Diane de Zélicourt

R Studio Connect



## RinteRface



# Transitioning from Academia to Industry: Using R skills for CaRReeR Success

## Key Steps & Concepts

- **CV --> Resume**
  - Research experience = Data Science Skills
  - Abstract away the subject matter
  - Mention tools, stats methods, data manipulation techniques
- **Online Presence**
  - LinkedIn, Website, Github
  - Concise repeat of resume highlights (languages, stats methods)
  - Don't forget the soft skills!
- **Key R Packages**
  - Pagedown
  - Caret
  - Recipes
  - Keras/TensorFlow
- **Know the “Lingo”!**
  - Key machine learning buzzwords and business acronyms

<https://www.burtchworks.com/2018/07/09/2018-data-scientist-salary-report-highlights/>

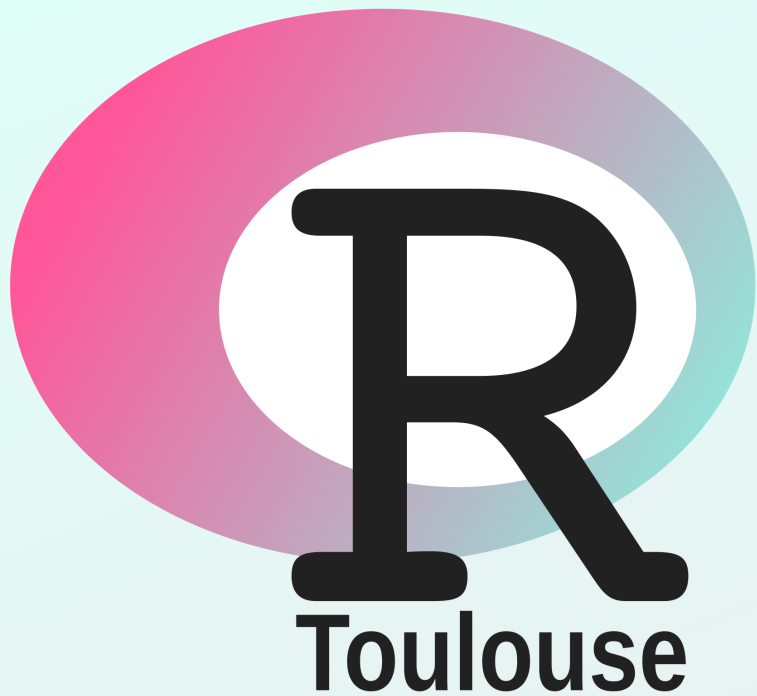
<https://versatilephd.com/>

<https://www.erdosinstitute.org/>

<https://rladies.org/>



- 
- ☹️ Fear of teaching cargo cult statistics in IntroStats?
  - 🤔 Looking for ideas how to help students to think and work with data?
  - 👁️ Poster K85: *Teaching Statistics for Data Literacy by the means of R mosaic and R Markdown*



# R user group Toulouse

Free R events, open to all without registration



[r-toulouse.netlify.com](https://r-toulouse.netlify.com)



[@RUG\\_Toulouse](https://twitter.com/RUG_Toulouse)



[rug.toulouse@gmail.com](mailto:rug.toulouse@gmail.com)



[groupes.renater.fr/sympa/info/r-toulouse](https://groupes.renater.fr/sympa/info/r-toulouse)

**French speaking group** --- Code of conduct: [r-toulouse.netlify.com/code-de-conduite/](https://r-toulouse.netlify.com/code-de-conduite/)

*We are looking for speakers!*





A Computational Analysis of

## K87: the Dynamics of R Style

Based on 94 Million Lines of Code from All CRAN Packages in the Past 20 Years

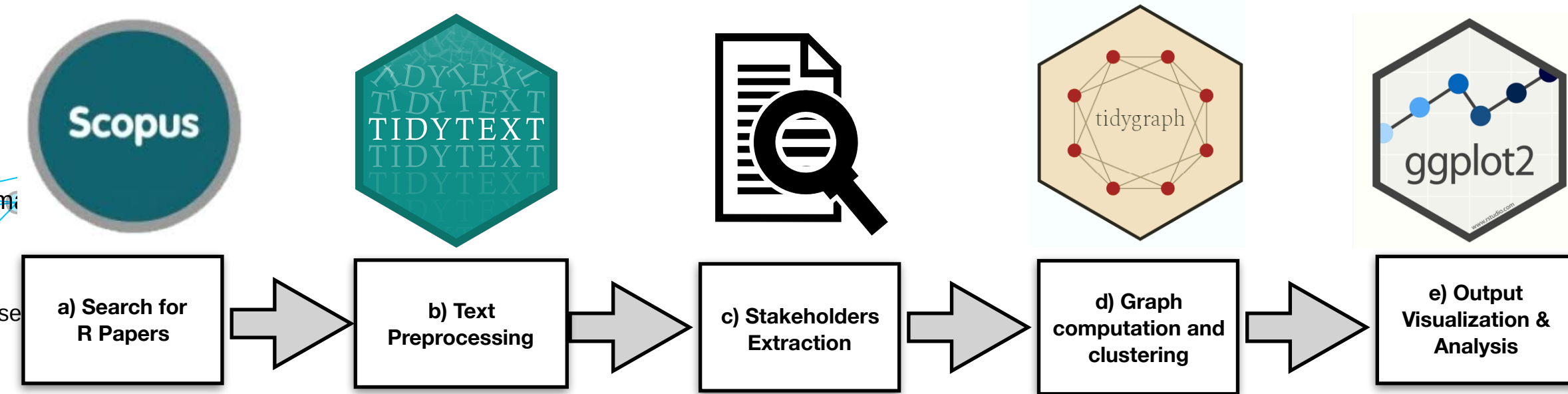
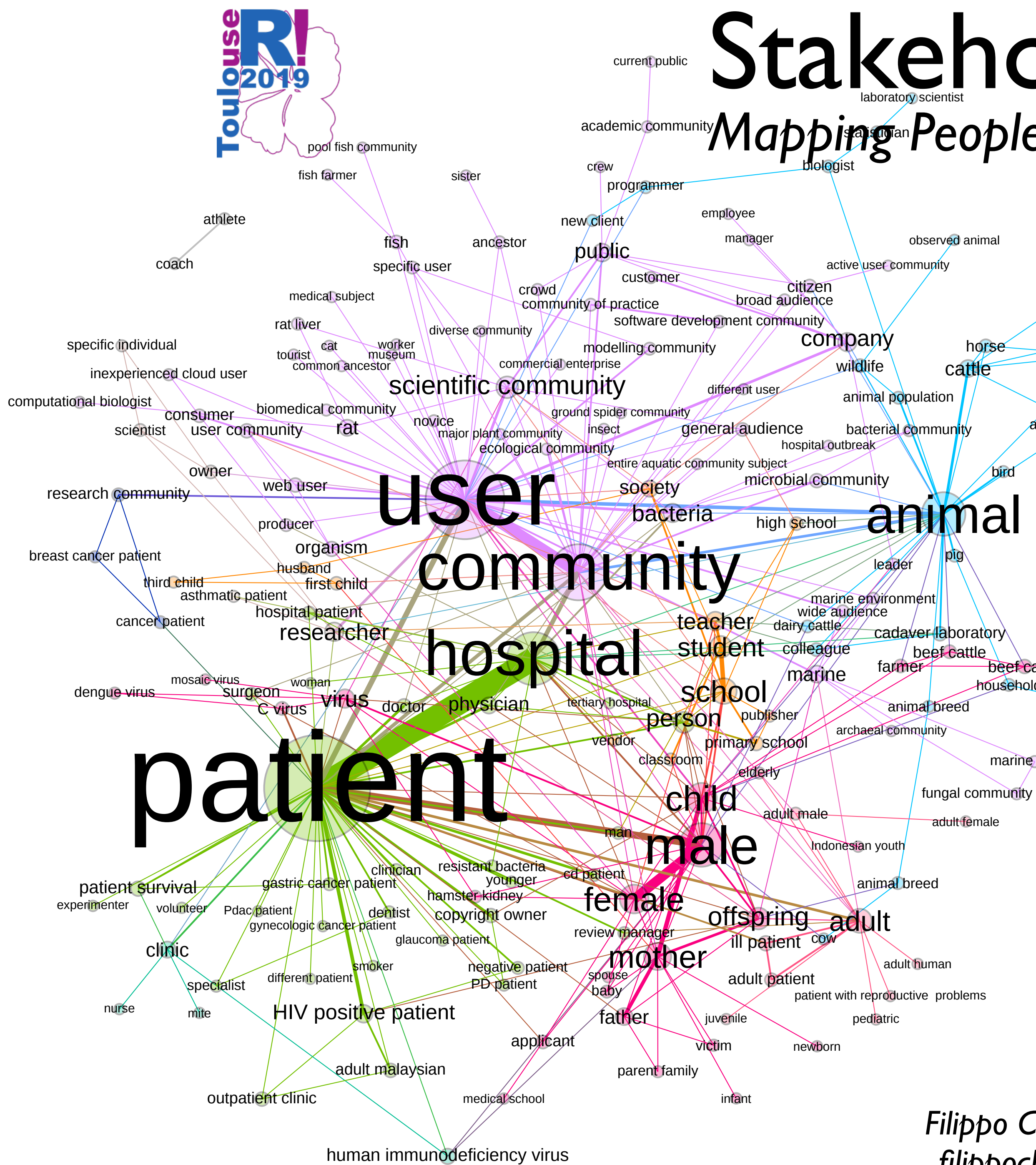
“ Given no dominant style guide in the past 20 years,  
what happens to the **programming styles** in R community then?

- **What are the “ins & outs?”** Least-agreed style elements in R
- **Who’s the “Naughty, Naughty?”** Unusual styles adopted in some communities
- **Is there an UNIVERSAL style guide in R?** A consensus-based one maybe



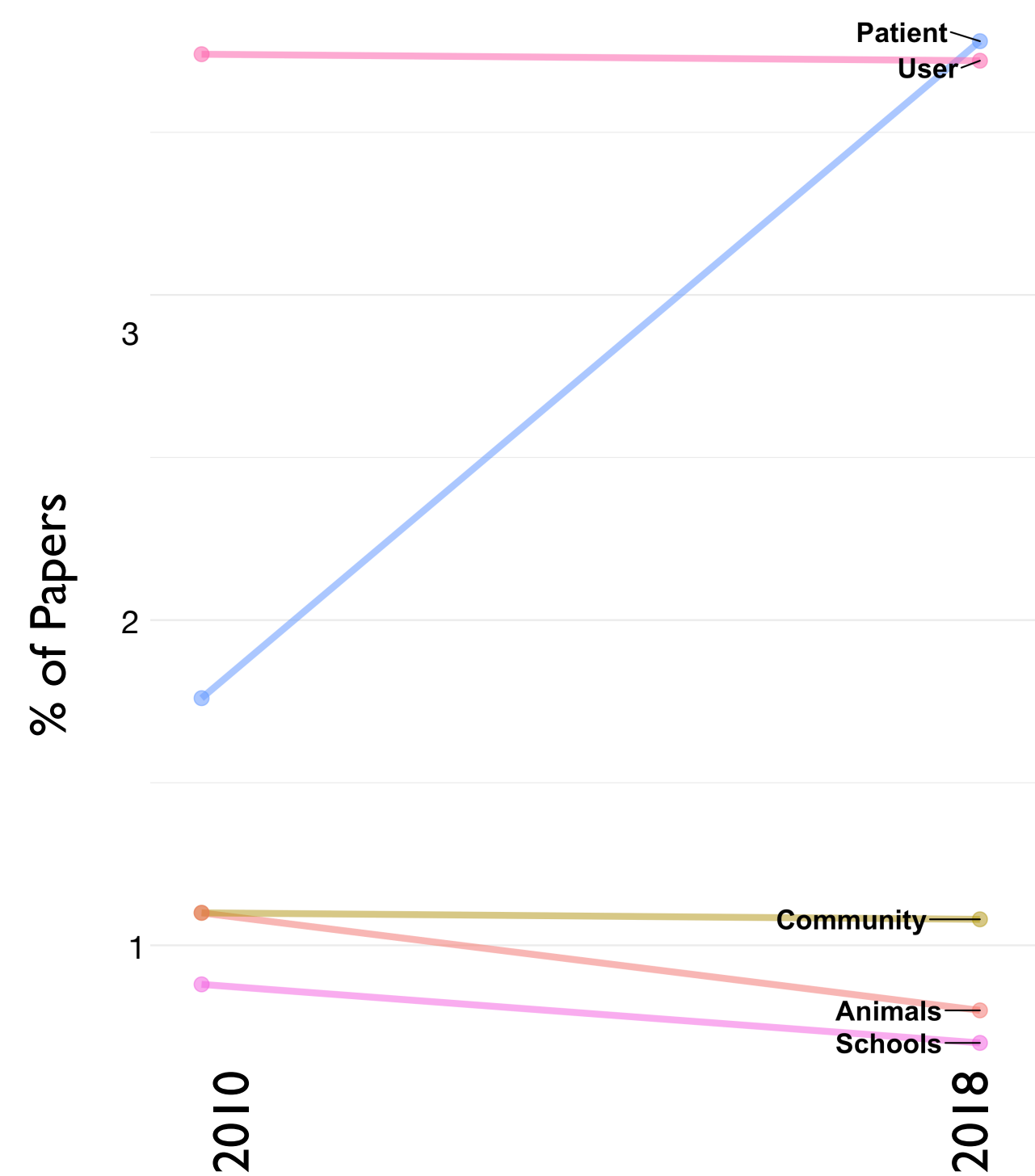
# StakeholderR

## Mapping People Impacted by R



**How are they extracted?**

**Who is Growing the most?**



# Next on stage :

## Group L: Operations

**L89** - Brandon Allen

**L90** - Mikaela Miller

**L92** - Laurent Pantera

**L93** - Jean-François Rey

**L94** - Luke Rasmussen, Eric Whitley

**L96** - Hong Ooi



### Get ready ! Group M

On stage in 3 minutes

**M97** - Ioannis Kosmidis

**M99** - Fanny Meyer, Victor Perrier

**M101** - Jaynal Abedin

**M103** - Yuki Hira, Kohei Kawai

**M104** - Veronika Soldánová

# A Significant Difference:

How the Institutional Research & Strategic Insights Team at the University of Chicago Booth School of Business uses R to Optimize and Institutionalize Analytics

Brandon Allen | University of Chicago Booth School of Business

# CHICAGO



The University of Chicago  
Booth School of Business

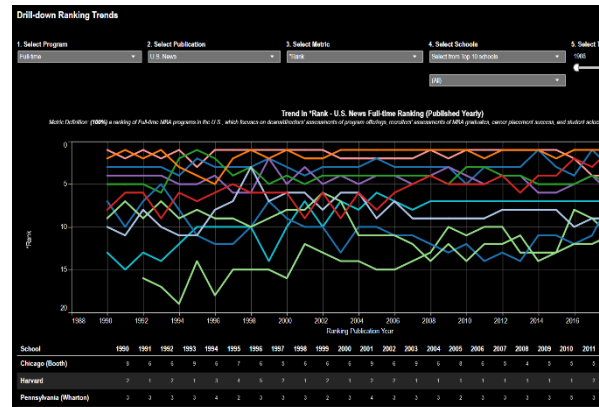
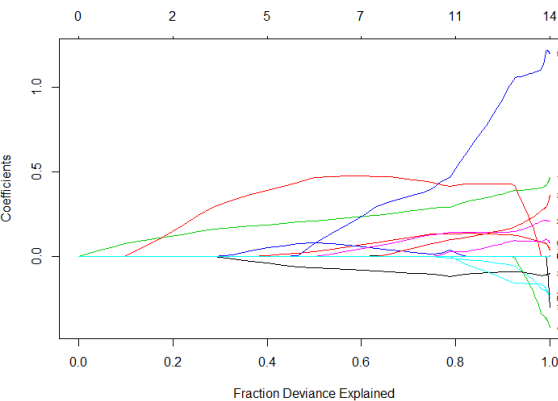
Leveraging R has saved Chicago Booth **thousands of hours** and **millions of dollars**, and it has allowed us to positively impact **tens of thousands** of students, faculty, and alumni.

This success relies on using R for:

## Statistics

## Institutionalization

## Optimization



```
l_test_variable_year_program_question = factor(variable_list_dictionary[, year, question, string, program, list])
df = data.frame(question_number = character(), year = character(), program = character(), variable_name = character(), count_one = character(), count_two = integer(), count = integer(), label = double(), year = double(), value = double(), error = character(), stringsAsFactors = FALSE)
for (row in 1:length(question_string)) {
  question_number = question_string[row]
  program_name = all_character(program_name_and_year_list[row, "program_name"])
  year = as.numeric(program_name_and_year_list[row, "year"])
  variable_name = variable_list_list[[row]]
  variable = variable_list_list[[row]]
  if (variable == "gender") {
    result = result[0, ]
  } else {
    count_one = length(dictionary[[question_number]][eval(parse(text=variable))=="A"year-year])
    count_two = length(dictionary[[question_number]][eval(parse(text=variable))=="B"year-year])
    result = rbind(result, list(question_number, year, program_name, variable, "A", count_one, count_two, NA, NA, NA, paste("warning", warning), warning, error = function() {}))
    result = rbind(result, list(question_number, year, program_name, variable, "B", count_one, count_two, NA, NA, NA, paste("error", error), error))
  }
  if (row == length(result)) {
    result = result
  } else {
    count_one = length(dictionary[[question_number]][eval(parse(text=variable))=="A"program-program_name-year])
    count_two = length(dictionary[[question_number]][eval(parse(text=variable))=="B"program-program_name-year])
    if (count_one > 0 & count_two > 0) {
      result = rbind(result, list(question_number, year, program_name, variable, "A", count_one, count_two, NA, NA, NA, paste("warning", warning), warning, error = function() {}))
      result = rbind(result, list(question_number, year, program_name, variable, "B", count_one, count_two, NA, NA, NA, paste("error", error), error))
    }
  }
}
```

# BECOMING A MISSION-DRIVEN R-ORGANIZATION

Examples of R solutions in the non-profit sector

Mikaela Miller, Lory Fehlig, & Shawn Rouen



## WHO WE ARE

*We connect people around the world in the fight to end poverty. Working together, we invest in the lives of children and youth, build the healthy environments they need to thrive, and empower them to create lasting change in their own lives and communities.*

## WHERE WE WORK



PROGRAM AREAS

## 2018 STATS



Health

Education

Empowerment

Employment

## HOW WE ARE USING R AT CHILDREN INTERNATIONAL

MARKETING  
xgboost models

*Challenge*  
Effective marketing and fundraising strategies

*Solution*  
Develop predictive models to maximize revenue for recurring appeals

GLOBAL PROGRAMS  
RMD Parameterized Reports

*Challenge*  
With 13 agencies in 10 countries, reporting is time-consuming and redundant

*Solution*  
Build reports for year-end results by agency with single .rmd file

SPONSOR RELATIONS  
R Shiny Apps

*Challenge*  
All sponsor/child correspondence is manually reviewed for content

*Solution*  
Upload correspondence file to Shiny app that detects and flags prohibited content



L92

## A generic HMI for physics experimentalists developed in R, rJava, htmlwidgets, plotly and shiny tools

Laurent Pantera CEA Cadarache, laurent.pantera@cea.fr  
Marion Savanier GE Healthcare, marion.savanier@ge.com

### Context and objective

- A small laboratory with some experimentalists in nuclear science engineering, all using R for data processing and interactive visualisation
- The number of new features grows with every passing year and we would like to make the developed tools (i.e. the R scripts) available through a light Human-Machine Interface

### Idea and adopted solution

Experimentalists will only have to write their scripts with:

- **htmlwidgets**
- **plotly**
- **shiny**

*outside the HMI*

*HTML files created by R scripts*

microphone-analysis.html  
top-onset-calculation.html  
new-feature.html.....

**Generic GUI**  
developed in the  
framework of  
**rJava and JavaFX**  
(Web Engine  
functionality)

# R PACKAGE DEVELOPMENT USING GITLAB CI/CD

BUILD YOUR OWN "--as-cran" PIPELINE



L 93



VirtualBox



docker



And More...

BIO/Π

L94

# RCocoa and R.NET

StatTag

A Consistent Interface Across macOS and Windows

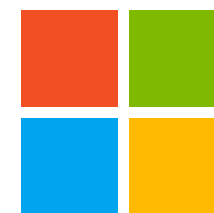
RCocoa



UNLEASH the  
**POWER** of **R**  
in macOS apps

[github.com/stattag/rcocoa](https://github.com/stattag/rcocoa)

R.NET



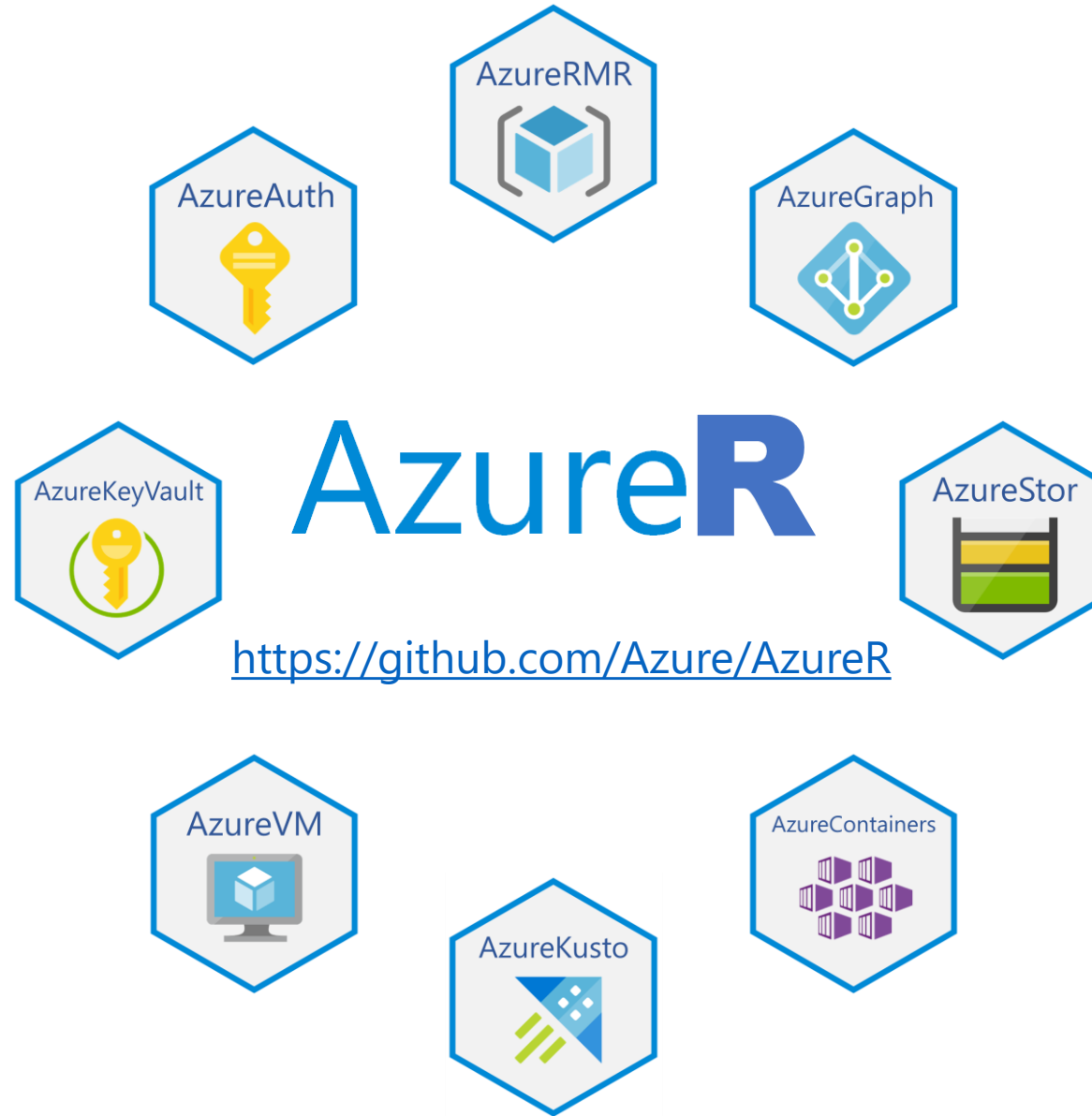
[github.com/rdotnet/rdotnet](https://github.com/rdotnet/rdotnet)

Embed R (run R code)  
in your macOS app



Simplify interaction  
with underlying R types





# Next on stage :

Group M: Web apps/tools



**M97** - Ioannis Kosmidis

**M99** - Fanny Meyer, Victor Perrier

**M101** - Jaynal Abedin

**M103** - Yuki Hira, Kohei Kawai

**M104** - Veronika Soldánová

## Get ready !

1. To applause
2. Go to see the **Poster session** in Caravelle 1, Greenhouse L and Greenhouse R



Ioannis Kosmidis  
Robin Horňák



The  
Alan Turing  
Institute



Data analysis workflow to track running, cycling & swimming



Fully interactive **shiny** app



Leverages the power of **trackerR**



No data trails



GPL3 R package in CRAN



[trackerapp.com](https://trackerapp.com)

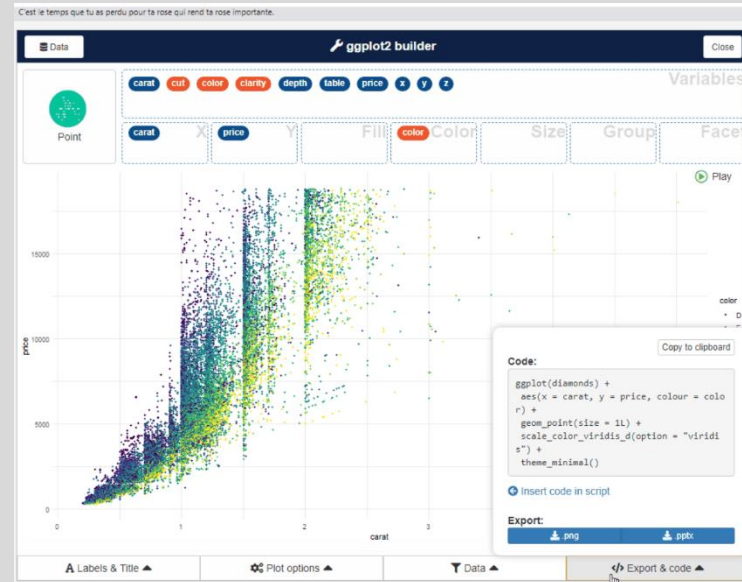


[github.com/trackerproject/trackerRapp](https://github.com/trackerproject/trackerRapp)

# Our Addins

A world full of little Addins

#M99



addinit

esquisse

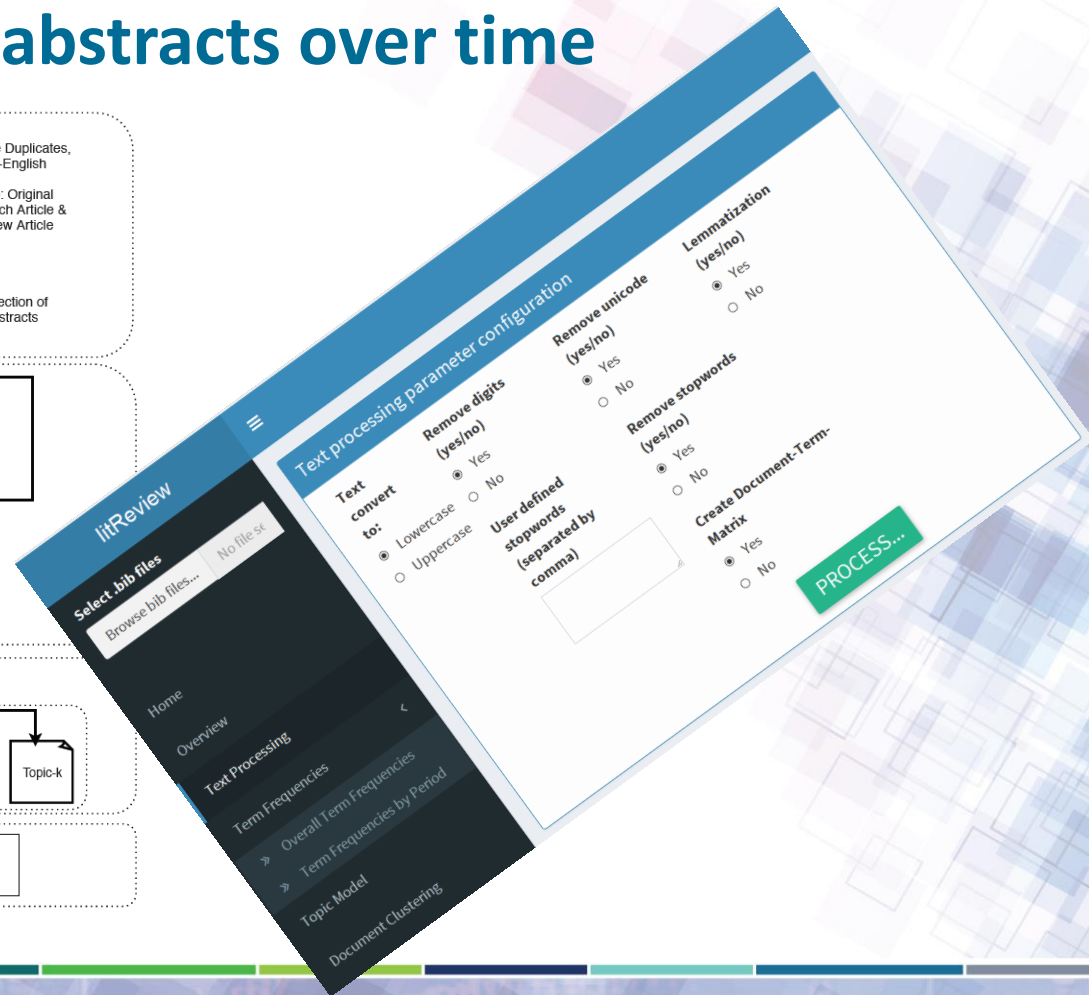
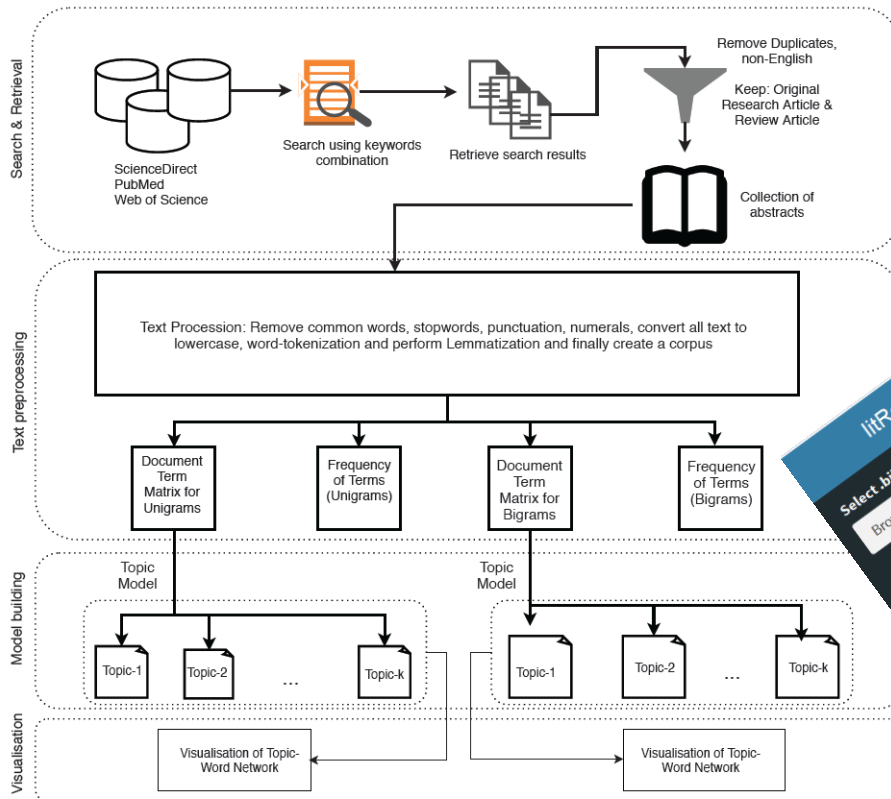
prefixer

And other ...



CRAN

# A shiny app to identify latent research themes in published abstracts over time

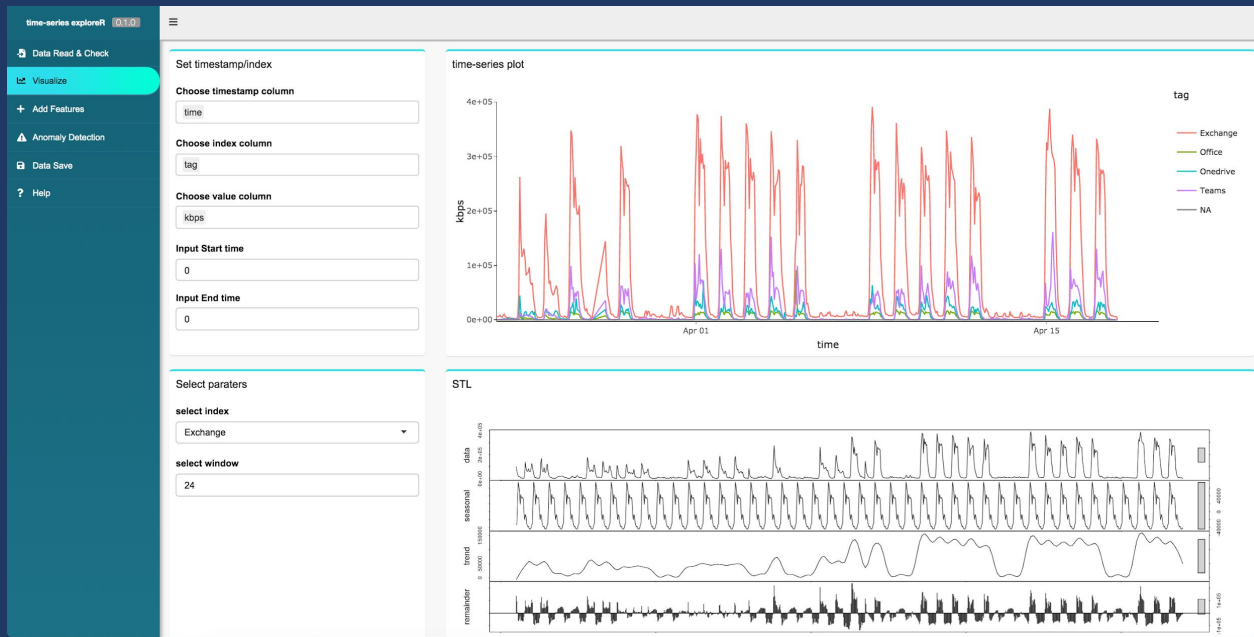


Jaynal Abedin<sup>1</sup> and John Newell<sup>1,2</sup>

<sup>1</sup>Insight Centre for Data Analytics, NUI Galway

<sup>2</sup>School of Mathematics, Statistics and Applied Mathematics, NUI Galway

# time-series ExploreR: Interactive time series analysis for data science in Shiny App@M103



time-series



Exploratory Data Analysis



Shiny Web App



# Guide to Historical Water Reservoirs in the Vicinity of Banská Štiavnica

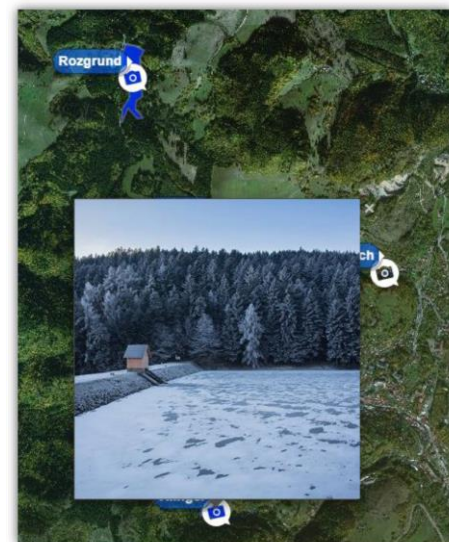
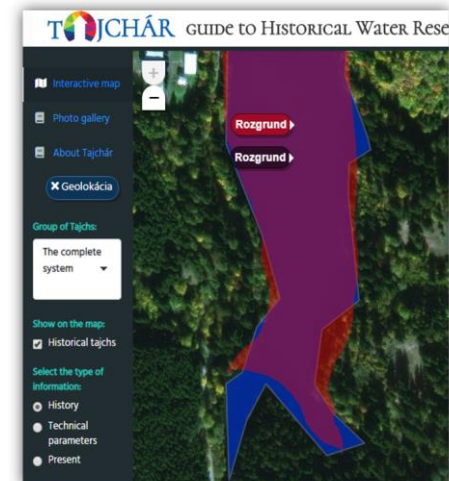
GUIDE TO HISTORICAL WATER RESERVOIRS IN THE VICINITY OF BANSKÁ ŠTIAVNICA



More than 66 of reservoirs were built in the Štiavnica Mountains at the end of the 17th and the beginning of the 18th centuries. Many of them are completely forgotten; some of them have been backfilled or otherwise destroyed. 24 of them are functional, serving as venues for leisure activities, fishing, or as water sources.



## Interactive Map Application



### History

- Historical group
- Year of construction
  - Builder
- Historical name
- Main purpose

...

### Technical Parameters

- Catchments
  - Volume
  - Dams informations
- Water surface area
- Coordinates

### Present

- Photos
- World heritage
- ICOLD
- Fishing Ground

...

### Tools

- Questionnaire
- Emotional Map
- Photogallery
- Geolocation

...

The domestic inhabitant's opinions on the map

