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
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
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@bicaster.org

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, mocCluster, 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
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$^A$, Wolfgang Viechtbauer$^A$

$^A$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
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
Mélina Gallopin, Benjamin Auder, Annie Glatigny, Marie-Hélène Mucchielli

Modeling the assembly process of a protein complex

PPI network

Identification of all proteins involved in a biological process

List of proteins of a complex or of a biological process

IKI3
ELP2
ELP3
ELP4
IKI1
ELP6

Intact
BioGRID
String

PPI Databases
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

Garbouge Hadhami¹, Santagostini Pierre², Charrier Aurélie³, Demilly Didier³, and Rousseau David¹

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

Best (Index 1)

Worst (Index 7)

K₁ K₂

Clustering

Empirical ordinal classification
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.

• Bioconductor
• Genetics
• Gap
• Pheatmap
• ReCon Packages

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

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

\[
\begin{align*}
R &= 0.89 \\
\end{align*}
\]
Quantifying the impact of tree choice in metagenomics differential abundance studies with R

Antoine Bichat, Christophe Ambroise, Mahendra Mariadassou, Jonathan Plassais

<table>
<thead>
<tr>
<th>Genus</th>
<th>S001</th>
<th>S002</th>
<th>S003</th>
<th>S004</th>
<th>S005</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atopobium</td>
<td>84</td>
<td>0</td>
<td>12</td>
<td>54</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Eggerthella</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Prevotella</td>
<td>525</td>
<td>7</td>
<td>134</td>
<td>753</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Lactobacillus</td>
<td>88</td>
<td>1770</td>
<td>1490</td>
<td>119</td>
<td>2136</td>
<td>...</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>0</td>
<td>0</td>
<td>138</td>
<td>4</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Dialister</td>
<td>152</td>
<td>4</td>
<td>2</td>
<td>192</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Megasphaera</td>
<td>402</td>
<td>0</td>
<td>4</td>
<td>102</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Sneathia</td>
<td>302</td>
<td>0</td>
<td>35</td>
<td>272</td>
<td>0</td>
<td>...</td>
</tr>
</tbody>
</table>
Which is the best method for metagenomics differential analysis?

**META*DAF* PACKAGE: EASY BENCHMARK AND COMPARISONS**

### Scientific Context
- **Commensal**
- **Pathogenic**
- Cancer
- Diabetes
- IBD, …

### Methods & Results
**Input data**
- `build_DAF_data()`
- **Choose method**
- `select_process()`
- **Run & visualize**
- `run_DAF_analysis()`
- **Keywords**
  - Metagenomics
  - Biomarker discovery
  - Differential abundance
  - Negative binomial & zero inflated distribution
  - R package
  - Reproducible research
  - In-silico benchmark

**Choose method**
- DESeq2
- mbzinb
- ALDEx2
- RAIDEx
- mSeq
- voom
- wilcox
- zibseq
- edgeR

**Input data**
- Standardization
- Generate artificial datasets

**Scientific Context & Methods & Results**
- Icons made by Freepik from www.flaticon.com
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\textsuperscript{a}, Primož KOCBEK\textsuperscript{a}, Gregor ŠTIGLIC\textsuperscript{a,b}

\textsuperscript{a}Faculty of Health Sciences, University of Maribor, Žitna ulica 15, 2000 Maribor, Slovenia
\textsuperscript{b}Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova ulica 17, 2000 Maribor, Slovenia

\textbf{Methods:} EHR in 5 SLO Healthcare institutions, 2073 individuals, FINDRISC model.

\textbf{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.

\textbf{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
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

**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*?

**Where?**
- Estimands are a central part of late phase clinical trials

Randomisation

A

B

- Estimands may also be useful to you!
OTrecod: An R package for data integration using Optimal Transportation theory

Gregory Guernec, 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:
- Y in B
- Z in A

**How solve this problem?**
- Framework of statistical matching
- Could be seen as a specific problem of missing data imputation

**What's new with OTrecod?**
- Most frequently used packages: StatMatch, mice
- 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.

Conclusions:
The Stacking Meta-Model gives a prediction model with better performances for the development of alternative approaches in safety evaluation of chemicals the each of 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)
What R we doing at the International Agency for Research on Cancer?

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

- Data Management
  - Harmonization
- Cancer Surveillance
  - Incidence trends
- Etiologic Research
  - Mendelian randomization
- Risk Prediction
  - Lung cancer risk from biomarkers
- Clinical Applications
  - Cancer screening

**Tools:**
- lubridate
- haven
- Segmentied Epi Nordpred
- TwoSampleMR
- pROC
- glmnet
- survival
- dplyr
- ggplot2
- markdown
- tidyverse

International Agency for Research on Cancer
World Health Organization
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)

Open Analytics
Analyzing Energy-Research in Europe

Patrick Wolf, Tobias Buchmann

**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*)
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

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

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

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

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

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!
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

Using only 4 R commands!
Comparison of HPC Techniques for Storing and Accessing Sim. BigData for Bayesian Inference and Designs for Survival 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

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

Filter documents from collections

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

Graph traversal

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

Add relationships into the graph

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

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

useR! 2019 Toulouse – France   11/07/2019
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.

### 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()`

---

**Introduction**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mpg</strong></td>
<td><strong>cyl</strong></td>
</tr>
<tr>
<td>Mazda RX4</td>
<td>21.0</td>
</tr>
<tr>
<td>Mazda RX4 Wgn</td>
<td>21.0</td>
</tr>
<tr>
<td>Datsun 710</td>
<td>22.8</td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td>21.4</td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td>18.7</td>
</tr>
<tr>
<td>Valiant</td>
<td>18.1</td>
</tr>
</tbody>
</table>

**Introduction**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>displ</strong></td>
<td><strong>hp</strong></td>
</tr>
<tr>
<td>Mazda RX4</td>
<td>160.0</td>
</tr>
<tr>
<td>Mazda RX4 Wgn</td>
<td>160.0</td>
</tr>
<tr>
<td>Datsun 710</td>
<td>108.0</td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td>258.0</td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td>306.0</td>
</tr>
<tr>
<td>Valiant</td>
<td>223.0</td>
</tr>
</tbody>
</table>

**Introduction**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>drivetrain</strong></td>
<td><strong>comb</strong></td>
</tr>
<tr>
<td>Mazda RX4</td>
<td>3.9</td>
</tr>
<tr>
<td>Mazda RX4 Wgn</td>
<td>3.9</td>
</tr>
<tr>
<td>Datsun 710</td>
<td>3.8</td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td>3.1</td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td>3.1</td>
</tr>
<tr>
<td>Valiant</td>
<td>2.8</td>
</tr>
</tbody>
</table>

---

**Theo Boutaris, Clemens Zauchner, Dana Jomar**
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
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:
  
  [BRGM Logo]
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.
- 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](image)

**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](image)

**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$^{1,2,3}$, Julien Randon-Furling$^{1,3}$, William Clark$^4$

1. SAMM, Université Paris 1 Panthéon-Sorbonne
2. MaIAGE, INRA
3. Institut des Migrations
4. Department of Geography, University of California, Los Angeles

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

- 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

Local densities of Chinese and Algerian communities (log scale)

Distortion coefficients maps

runDTW: Detect Query Patterns in Long Time Series

Query:  
How?

Closest!

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

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

---

Cross-referencing catchment data

- Catchment boundaries
- Climate/Morphology/Landcover
- Data at basin scale
- Hydrology
- Catchments boundaries
- Climate/Morphology/Landcover

---

- Hydrological model
- Flood prediction

---

Main characteristics of catchments
airGR & airGRteaching

two packages for rainfall-runoff modeling & teaching hydrology

Olivier Delaigue¹, Guillaume Thirel¹
Laurent Coron², Pierre Brigode³

(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 electricity system needs to remain balanced at all times

Could household appliances play a role?
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** - Guyliaann 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

- **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

---

Céline Monteil¹, Fabrice Zaoui¹, Frédéric Hendrickx¹, Nicolas Le Moine²

¹ **EDF R&D LNHE** - Laboratoire National d'Hydraulique et Environnement, Chatou, France,
² UMR 7619 Metis (SU/CNRS/EPHE), Sorbonne Université, Paris, France
BAYESDEF: Graphical Interface for the Bayesian Analysis of DSD
Víctor Aguirre, Statistics Department, ITAM; Sofia Huerta, UV; Edgar López, ITAM


Table 1. DSD for m = 6 factors

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

\[
y_i = 20 + C + F + \varepsilon_i
\]

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

\[
P(A)/(1 - P(A))
\]

JMP 11.1.1.

A, B, C, D, F, AC, AF, CD, CF, and BB
Gaussian process (GP) models under linear inequality constraints

Let $Y$ be a GP on $\mathbb{R}$ given by
\[
Y(x) \sim \mathcal{GP}(m, k),
\]
with mean function $m$ and covariance function $k$.

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

![Unconstrained GP](a) Unconstrained GP  
![Constrained GP](b) Constrained GP  

**Figure 1:** Examples GP regression models.

Main functionalities as S3 methods

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

References


Further Implementations

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

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](image1)

![Data Set 2](image2)
# The tool: sandbox

**Applications:**
- Geochemistry
- Age determination

**Discussions:**
- Speed things up
- Inverse problems
- Solving diff. eqs.

**Parameters (specific/plot):**

- **Parameter:** Parameters are used to describe populations. They can be seen as "normalor nuclear" definitions.
- **Rules:** Rules are used to describe how parameters behave with depth. They can be regarded as "virtual exploitation" of the sediment deposits.
- **Rule Book:** A rule book is the combination of parameters ("normalor nuclear") definitions and rules ("virtual exploitation") into a coherent reference book.

**Analysis Functions:** Show a virtual sediment deposit is defined by a rule book. This deposit can be "recovered". Analysis functions use information stored for each grain.

**Building a section:**

- **Lyme section (Austria)**: with updated geochemical data, modeling of 3D and 4D elements (described by distribution functions), and streamlined for easy interpretation.
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

Some limitations

“Multispecies Individual-based model (IBM) which focuses on fish species”

CRAN Packages

- population
- BiodiversityR
  - vegan
  - Ecotroph
- sdmpredictors
  - sdm
- SDMTools
- biomod2
- ibm
- osmose
- bgmfiles
- atlantistools
- rbgm

CRAN GitHub

configuration

- model
- output

run_osmose()

read_osmose()

getVar()

S3 OOP classes & methods

H71
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)

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

AVM appraisal

Regular appraisal

Banking sector
- Accounting
- Compliance
Dynamic allocation optimization in A/B tests using classification-based preprocessing

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.

Goals: limit losses / maximize rewards and identify the best (conditional) treatment.

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** (WEighted Means 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
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

G. Engels & Ph. Grosjean
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
Transitioning from Academia to Industry: Using R skills for Career 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://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

globe r-toulouse.netlify.com

twitter @RUG_Toulouse

e-mail rug.toulouse@gmail.com

e-mail groupes.renater.fr/sympa/info/r-toulouse

French speaking group  ---  Code of conduct: 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
StakeholdeRs
Mapping People Impacted by R

Who is Growing the most?

How are they extracted?

Filippo Chiarello, PhD
filippochiarello.com
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

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
BECOMING A MISSION-DRIVEN R-GANIZATION
Examples of R solutions in the non-profit sector
Mikaela Miller, Lory Fehlig, & Shawn Rouen

88% of expenses went directly to programs

200,000 children and youth participating

67 community centers operated worldwide

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

2018 STATS

88% of expenses went directly to programs

200,000 children and youth participating

67 community centers operated worldwide

PROGRAM AREAS

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

mmiller@children.org

www.children.org

useR! Conference 2019, Toulouse, France
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.

**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)*

**Laurent Pantera**

CEA Cadarache, laurent.pantera@cea.fr

**Marion Savanier**

GE Healthcare, marion.savanier@ge.com
R PACKAGE DEVELOPMENT USING GITLAB CI/CD

BUILD YOUR OWN "--as-cran" PIPELINE

L 93

And More...
RCocoa and R.NET
A Consistent Interface Across macOS and Windows

UNLEASH the POWER of R in macOS apps

Embed R (run R code) in your macOS app

Simplify interaction with underlying R types

RCocoa

github.com/stattag/rcocoa

R.NET

github.com/rdotnet/rdotnet
https://github.com/Azure/AzureR
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
Data analysis workflow to track running, cycling & swimming

Fully interactive **shiny** app

Leverages the power of **trackeR**

No data trails

GPL3 R package in CRAN

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

[github.com/trackerproject/trackeRapp](https://github.com/trackerproject/trackeRapp)
Our Addins

A world full of little Addins

#M99

addinit

esquisse

prefixer

And other ...
A shiny app to identify latent research themes in published abstracts over time

Jaynal Abedin\textsuperscript{1} and John Newell\textsuperscript{1,2}

\textsuperscript{1}Insight Centre for Data Analytics, NUI Galway
\textsuperscript{2}School of Mathematics, Statistics and Applied Mathematics, NUI Galway
time-series ExploreR: Interactive time series analysis for data science in Shiny App@M103
guide to Historical Water Reservoirs in the Vicinity of Banská Štiavnica

Veronika Soldanová, Milan Cíty

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 information
- 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

shinyApps.io