



Recueil des résumés

Un événement de la Société Française de Statistique

SFds

Table des matières

Air, un nouveau formateur de code pour R, Henry Lionel	1
Apprendre R avec méthode et efficacité, Della-Vedova Claire	3
Approche morphométrique 3D par Analyse Procrustes Généralisée sous R, Neels Nicolas [et al.]	4
Automatisations du contrôle qualité de formulaire de recueil de données et de rapport de résultats avec les packages docextractr, flextable et officer, Le Guen Nelly [et al.]	5
Automatiser l'analyse de graphes d'hyperliens avec igraph, Viseur Robert	6
Basculer d'une application shiny (SaaS) vers des rapports Rmarkdown en Pharmacovigilance (RaaS): résumé des avantages et inconvénients, Van Holle Lionel	7
Calcul d'indicateurs qualité à la Haute Autorité de Santé : Mise en place d'une migration de SAS vers R sans arrêt de la production des indicateurs, Coquelin Anaëlle [et al.]	8
Comment traduire vos documents efficacement dans R, Salmon Maëlle [et al.]	10
Conception d'applications Shiny avec {golem}, Guyader Vincent [et al.]	12
Création d'un package d'échantillonnage de polytope en modélisation linéaire inverse, Brehelin Jacques [et al.]	13

Créer des applications RAG-LLM avancées avec R Shiny, Ihaddaden Mohamed El Fodil	14
De la carte aux modèles spatiaux, Visée Corentin	15
Des Learning Analytics qui ne manquent pas d’R, Armand Patrice [et al.]	17
DoReMIFaSol, un package pour simplifier l’usage des données en open-data de l’Insee, Lamarche Pierre	19
Dynamisez vos documents Quarto grâce à Observable JS, Bichat Antoine	20
Détection et quantification de manipulations comptables autour d’un seuil psychologique avec R, Chavent Marie [et al.]	21
Développement de workflows en R pour le nettoyage, la validation et la visualisation de données spatiales en écologie, Benrezkallah Jordan	22
Explorer le CRAN avec le package RWsearch, Kiener Patrice	23
Faciliter la prise en compte des risques économiques en agriculture grâce à R et Shiny, Topart Morgane [et al.]	24
Framework d’import de base de données en fichiers pour la recherche clinique : le package {EDCimport}, Chaltiel Dan	25
Introduction pratique à R : les fondamentaux pour organiser, analyser et visualiser, Della-Vedova Claire	26
La traduction du mlverse en français, Regouby Christophe	27
Le format Parquet et l’écosystème DuckDB: l’essayer c’est l’adopter!, Galiana Lino	28
Les élections européennes et législatives 2024, Kiener Patrice	29

Make a difference: fast and accurate numerical derivatives with 'pnd', Kostyrka Andreï V.	30
Méthodes quantitatives en géographie : régression, ACP et classification sur des objets spatiaux, Périlleux Hugo	32
Nettoyer son code en un clin d'oeil avec flint, Bacher Etienne	33
OLAF, la reine des Apps Shiny !, Bichat Antoine [et al.]	35
Outils et conseils pour la maintenance partagée de package R, Gruson Hugo	36
PATAPI : un outil de planification des pesées de bovins développé avec R, Legris Maxime	38
Pages d'aide simplifiées et en français dans R grâce à l'IA, Engels Guyliann [et al.]	40
PerRegMod : un nouveau package R pour le test et l'estimation des modèles de régression à coefficients périodiques, Slimane Regui [et al.]	42
Personnalisation de documents Quarto avec '_brand.yml', Dervieux Christophe	44
PhaochR : géocoder des adresses belges avec R, Périlleux Hugo [et al.]	46
R and CRAN on Ubuntu With r2u, Eddelbuettel Dirk	47
R au service du séquençage microbien : analyse et visualisation des résultats, Delacuvellerie Alice [et al.]	48
R et le déclin des pollinisateurs : des outils statistiques au service de l'écologie, Vanderplanck Maryse	49
R pour l'analyse de données vectorielles en analyse de la parole, Crouzet Olivier	50

R vous ment!, Fabri Antoine	51
R, un catalyseur d'interdisciplinarité : tour d'horizon du projet eGait sur l'analyse de la marche, Stamm Aymeric	52
RDepot - 100% open source enterprise management of R and Python repositories, Van Malder Jonas	53
RProtoBuf: sérialiser pour communiquer, Barthelemy Tanguy	55
Rajouter l'interactivité à vos ggplots avec animint2, Hocking Toby Dylan	57
Rescuelog : Collecte numérique et reporting des interventions de sauveteurs à l'océan avec (ru)ODK, Shiny et SK8, Carayon David [et al.]	58
Rlinguo - R en natif sur mobile, Fay Colin	60
Sharing data science artifacts across teams using Crane, Lionakis Lucianos	62
ShinyProxy: easily deploy your Shiny apps, De Koninck Tobia [et al.]	63
Simplifiez, fiabilisez et accélérez vos workflows R complexes avec {targets}, Guyader Vincent	64
Strengthening confidence in LLM-generated responses with TrustMe: a package for evaluating semantic proximities between responses generated by the NaileR package, Mahmoud Remi [et al.]	66
Tester une application Shiny : méthodologie et outils pratiques, Bréant Arthur	68
Traiter des verbatims et questions ouvertes à l'aide d'un modèle d'IA local, Vroylandt Thomas [et al.]	69
Traiter les données collectées à l'aide de formulaires pour partager des inventaires et des résultats d'enquêtes. Développement informatique et application aux pratiques agroécologiques en Europe, Vanwindekens Frédéric	71

Transition SAS à R en recherche clinique : traduction et validation de macros SAS, Ibrahim Nusaibah [et al.]	73
Truffles : une application Shiny pour cultiver ses données... et ses truffes !, Delmotte Murielle	74
Tutoriel {data.table}, Hocking Toby Dylan	75
Un exemple d'utilisation du package likert et l'interface R Shiny pour la visualisation et l'analyse de questionnaires d'auto-évaluation des pratiques de pédagogie inclusive à l'université, Nguyen Viet Chau Linh [et al.]	76
Utiliser les GitHub Actions pour automatiser des tâches : exemple de l'update de la version dev d'un package, Chaltiel Dan	78
Workflow automatisé d'intégration et d'analyse de données en Shiny, Demeure Mélanie [et al.]	79
audio.whisper – Speech Recognition in R, Wijffels Jan	81
dbreportR : Un package R pour la documentation automatisée des bases de données, Boutry Sébastien [et al.]	82
fonctionr : l'inférence statistique pour tou-te-s, Girès Joël [et al.]	84
gaussratiovegind : un package pour analyser la distribution du ratio de deux distributions normales – Application aux images de fluorescence de chlorophylle, Santagostini Pierre [et al.]	85
iteR : Un package R pour l'analyse de matrices de confusions issues d'expériences de perception phonétique., Crouzet Olivier	87
pkgdocs: a modular R package site generator, Seynaeve Daan [et al.]	89
rix: Reproductibilité d'Environnements De Développement R Avec Nix, Rodrigues Coelho Bruno André	90

shiny2docker : dockeriser une application Shiny en toute simplicité, Guyader Vincent	92
{microinverterdata}, la collecte de données de votre souveraineté énergétique, Regouby Christophe	93
{torch}, {tabnet} et l'apprentissage profond par l'usage, Regouby Christophe	94
Liste des participants	95
Liste des auteurs	99

Air, un nouveau formateur de code pour R

Lionel Henry*

Résumé (max 300 mots)

Les formateurs de code sont devenus des outils de développement indispensables pour de nombreux langages tels que Python, Rust, ou Go. Ils sont des compagnons d'écriture de code qui réorganisent leur présentation sans en changer le contenu. Invoqués automatiquement à chaque fois qu'un fichier est sauvé sur disque, ils garantissent l'homogénéité et la lisibilité du code source.

Dans cette présentation, vous découvrirez Air, un nouveau formateur de code pour R extrêmement rapide développé par Posit pour Positron et RStudio. Nous verrons comment intégrer Air dans d'autres outils de développement modernes grâce à la généricité et agnosticité de cet outil. Vous apprendrez ce qu'est un formateur, quels sont les principes que nous avons retenus pour standardiser les codes sources en R, comment l'utiliser de façon efficace, et quelles sont les perspectives de développement pour ce projet.

Mots-clefs (3 à 5) : Outil - Développement

Développement

Les formateurs de code sont devenus des outils de développement indispensables pour de nombreux langages tels que Python, Rust ou Go. Ils réorganisent la présentation du code sans en changer le contenu, garantissant ainsi l'homogénéité et la lisibilité du code source. Dans le monde du développement logiciel, des outils comme Black et Ruff pour Python, Rustfmt pour Rust, Prettier et Biome pour JavaScript, sont largement adoptés pour assurer une cohérence stylistique.

Dans la communauté R, le package `styler` (Walthert (n.d.)) a été une première étape importante pour apporter un formateur de code. Écrit lui-même en R, ce formateur est distribué de façon pratique via CRAN. Cependant il manque de certaines fonctionnalités telles que la réorganisation de code sur plusieurs lignes (ajout de nouvelles lignes) et de la rapidité nécessaire pour une adoption transparente dans des outils de développement tels que Positron. C'est dans ce contexte que nous avons développé Air, un formateur de code pour R qui vise à combler ces lacunes.

Air a été conçu pour être extrêmement rapide et s'intégrer facilement dans divers environnements de développement. Écrit en Rust sur base d'un framework fourni par le projet Biome (Contributors (n.d.)), le formateur est à la fois très performant et fonctionnel. Air est capable de formater des fichiers individuels en quelques millisecondes et des projets entiers en moins d'une seconde. Le formateur est exposé dans un outil de ligne de commande ainsi que dans un serveur de langage.

Air représente une avancée significative pour le développement en R. Le projet offre une solution rapide et moderne pour le formatage de code en R comparable à ce qui existe pour d'autres langages de programmation.

Références

Contributors, Biome.js. n.d. "Biome, Toolchain of the Web." <https://github.com/biomejs/>.

Walthert, Lorenz. n.d. "Styler: Non-Invasive Pretty Printing of r Code." <https://github.com/r-lib/styler/>.

*Posit PBC, lionel@posit.co

Apprendre R avec méthode et efficacité

Claire Della-Vedova*

Résumé

Que vous soyez chercheur, étudiant, chargé de projet ou statisticien, apprendre à utiliser R peut sembler complexe sans une approche adaptée.

Lors de cette présentation, je partagerai mon expérience en tant que formatrice spécialisée dans l'enseignement de R et présenterai les approches pédagogiques les plus efficaces pour aider les apprenants à progresser et à utiliser R de manière autonome.

Nous identifierons ensemble les principales difficultés rencontrées par les débutants : la prise en main de la syntaxe, la compréhension des concepts fondamentaux et l'adoption de bonnes pratiques de travail.

Je présenterai ensuite les solutions pédagogiques que j'ai mises en place, en m'appuyant sur une approche progressive et structurée avec le tidyverse, ainsi que sur l'utilisation d'outils comme Quarto, qui facilitent la structuration des analyses et assurent leur reproductibilité.

À partir de retours d'expérience concrets, j'aborderai les approches qui ont donné les meilleurs résultats, en mettant l'accent sur l'importance d'un accompagnement adapté et de formations pensées pour répondre aux besoins spécifiques des apprenants.

Enfin, je proposerai des conseils pratiques pour concevoir des formations claires et interactives, permettant un apprentissage progressif et efficace.

Mots-clefs : Approche pédagogique – tidyverse – Quarto – Formation

* DellaData (prestations et formations méthodologie / statistiques / langage R), claire@delladata.fr

Approche morphométrique 3D par Analyse Procrustes Généralisée sous R

Nicolas Neels*'

Kévin Tougeron*

Christophe Mallet'

Résumé

Le monde qui nous entoure est un trésor de biodiversité. Les organismes qui le composent peuvent présenter des formes très variées, même entre des espèces proches, et des différences intraspécifiques marquées peuvent également être mises en évidence. Cette richesse de formes peut être caractérisée grâce à la morphométrie 3D, plus précisément par généralisation de l'analyse Procrustes. Il s'agit d'une approche permettant de s'intéresser uniquement aux variations de forme d'un organisme ou d'une partie du corps de celui-ci, indépendamment de la taille de l'élément considéré, menant à une analyse plus détaillée des relations de causalités à l'origine de ces différences. Cela en fait donc un outil particulièrement intéressant en biologie, principalement pour les domaines de la paléontologie ou de la biologie évolutive.

Dans cette présentation, nous vulgariserons les grandes lignes de la morphométrie 3D^{2,3} et de l'Analyse Procrustes Généralisée (GPA)^{2,3} en expliquant comment implémenter ces analyses sous R, principalement grâce aux packages 'Morpho'¹ et 'Rvcg'¹. Nous présenterons ces approches au travers de l'étude d'os de rhinocéros fossiles de l'espèce *Teleoceras fossiger*.

Mots-clefs : Morphométrie 3D, Analyse Procrustes Généralisée (GPA), Biologie, Biostatistiques, Paléontologie

Références

1. **Schlager, S. (2017)**. Morpho and Rvcg – Shape Analysis in R. Dans Elsevier eBooks (p. 217-256). <https://doi.org/10.1016/b978-0-12-810493-4.00011-0>
2. **Mitteroecker, P., & Schaefer, K. (2022)**. Thirty years of geometric morphometrics : Achievements, challenges, and the ongoing quest for biological meaningfulness. *American Journal Of Physical Anthropology*, 178(S74), 181-210. <https://doi.org/10.1002/ajpa.24531>
3. **Rohlf, F. J., & Slice, D. (1990)**. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic zoology*, 39(1), 40-59. 9

* Université de Mons (UMons), service d'écologie des interactions et changements globaux, Institut des Biosciences, nicolas.neels@student.umons.ac.be, kevin.tougeron@umons.ac.be

' Institut Royal des Sciences Naturelles de Belgique (IRSNB), cmallet@naturalsciences.be

Automatisations du contrôle qualité de formulaire de recueil de données et de rapport de résultats avec les packages docxtractr, flextable et officer

Nelly Le Guen*, Agnès Solomiac*, Alwenna Salaün*, Caroline Prunet*, Anaëlle Coquelin*,
Anaïs Sitruk*, Sandrine Morin*

Résumé (max 300 mots)

La Haute Autorité de Santé (HAS) développe des indicateurs de qualité et de sécurité des soins (IQSS) en lien avec les professionnels de santé, les patients et les usagers. Ces indicateurs sont mesurés à partir d'informations issues de dossiers patients, saisies via des formulaires disponibles sur une plateforme dédiée. Ces formulaires sont spécifiés en amont dans des documents sous format Word. La richesse des packages du logiciel R offre la possibilité d'automatiser le traitement de données et la production de résultats statistiques.

Méthode :

- Travail de standardisation des formulaires (document Word) réalisé selon des règles spécifiques.
- Formations et accompagnement des chefs de projet scientifiques par les data-managers/statisticiens ;
- Recherche de packages répondant au besoin d'automatiser le contrôle qualité des données saisies sur les formulaires et les rapports de résultats :
 - o Lecture et extraction du texte des formulaires WORD : Package docxtractr
 - o Création automatique de rapports détaillant les statistiques descriptives : Packages flextable et officer

Résultats :

Ces packages ont permis d'automatiser :

- La création d'un catalogue de variables (noms codifiés, libellés, modalités et filtres) ;
- La validation des saisies via les formulaires (programme Rmarkdown et sorties HTML) ;
- La production d'un rapport statistique (tableaux et graphiques standardisés).

Ce processus a été déployé sur les formulaires dédiés à une dizaine d'indicateurs.

Conclusion :

L'automatisation de plusieurs tâches rendues possibles par R a contribué à améliorer :

- La fiabilité du catalogue de variables produit ;
- La sécurisation de la conformité des saisies ;
- La lecture et interprétation des productions statistiques ;

Le temps de programmation est diminué significativement grâce aux packages docxtractr, flextable et officer.

Ce processus d'automatisation est facilement transposable à l'ensemble des formulaires de recueil de données.

Perspective : Création de packages R à partir des fonctions développées.

Mots-clefs (3 à 5) : Automatisation - Formulaire de recueil – Contrôle qualité – Statistiques descriptives

* Haute Autorité de Santé, Service Evaluation et Outils pour la qualité et sécurité des soins -
contact.iqss@has-sante.fr

Automatiser l'analyse de graphes d'hyperliens avec igraph

Robert VISEUR*

Résumé

Spygraph¹ est un crawler configurable (*focused crawler* en anglais) développé en Python. Il permet d'explorer un ensemble de sites web, de manière plus ou moins ciblée, puis d'exporter les hyperliens et les domaines découverts. L'outil facilite l'exploration itérative d'écosystèmes d'affaires, au travers des sites web des membres, dans un premier temps de manière divergente (par exemple pour découvrir de nouveaux acteurs jusqu'à alors inconnus), dans un second temps de manière convergente (pour concentrer l'analyse sur un ensemble validé de sites web). L'exportation des résultats de l'exploration peut se faire sous la forme d'un fichier de tableur (CSV) et d'un fichier de graphe (DOT, GML). Une première version du logiciel exportait les graphes au format DOT uniquement. Les formats CSV et DOT permettaient notamment une analyse dans le logiciel libre [Gephi](#). Une version ultérieure permet l'exportation au format GML. Ce dernier est exploitable au sein du logiciel [igraph](#). igraph est un ensemble d'outils d'analyse de réseaux, libre et gratuit, programmable en R, Python, Mathematica et C/C++. Deux cas d'utilisation sont présentés : d'une part, l'utilisation, avec Spygraph, à l'aide du langage Python², d'autre part, l'analyse d'un fichier au format GML, exporté depuis Spygraph, directement dans R³. L'objectif est de générer, après le crawl, premièrement, un ensemble de métriques (p. ex. *degree*, *betweenness centrality* et *pagerank*), deuxièmement, une prévisualisation des communautés. Une comparaison des différentes approches d'analyse est proposée en guise de conclusion.

Mots-clefs : crawler - Spygraph - analyse de graphes - Gephi - igraph.

1 Cf. http://inforsid.fr/actes/2022/INFORSID_2022_p81-84.pdf.

2 Cf. <https://python.igraph.org/en/stable/>.

3 Cf. <https://r.igraph.org/articles/igraph.html>.

* UMONS, robert.viseur@umons.ac.be

Basculer d'une application shiny (SaaS) vers des rapports Rmarkdown en Pharmacovigilance (RaaS): résumé des avantages et inconvénients

Van Holle Lionel*

* Van Holle Lionel, OpenSourcePV, Belgium (lionel.vanholle@opensourcepv.com)

Résumé (max 300 mots)

Une application R-Shiny de Pharmacovigilance a été développée en mode Software-as-a-Service (SaaS) sur la base de données VAERS permettant aux utilisateurs d'explorer des signaux quantitatifs et de déterminer la présence d'un potentiel effet causal entre l'exposition à un vaccin et l'émergence d'effets secondaires.

Les requis pour ce modèle SaaS sont :

- le déploiement sécurisé d'une application shiny sur un serveur disponible 24/7
- une application shiny acceptée par les clients en terme de sécurité, compliance et validation
- l'établissement d'un contrat
- la formation des utilisateurs à l'accès, la navigation et l'interprétation des données

Ces requis amènent des contraintes de coûts, une scalabilité limitée et des opportunités ratées en terme de contrats.

Je vais présenter le basculement du mode SaaS vers le mode Reporting-as-a-Service (RaaS) ainsi que ses potentiels avantages et inconvénients. Le mode RaaS consiste en la génération automatique grâce à Rmarkdown d'articles en PDF qui, pour chaque signal quantitatif, fournit la description des différentes méthodes, les résultats associés, la caractérisation des données pour en retirer les éléments saillants et l'interprétation qui peut en être faite. Un listing des cas d'intérêt en appendice permet aux utilisateurs d'évaluer plus en profondeur si certains cas présentent des preuves de relation causale entre l'exposition au vaccin et l'effet secondaire.

Une fois les articles générés, ils sont mis à disposition dans un site d'e-commerce derrière un paywall avec comme description un abstract sur le même modèle que les journaux dans la littérature scientifique. De cette façon, le modèle RaaS court-circuite le déploiement d'une app Shiny, l'établissement d'un contrat, la formation des utilisateurs ainsi que quelques contraintes des départements IT ou procurement. Les utilisateurs peuvent choisir à la carte ce qui les intéresse et évitent les copier-coller d'une application shiny ainsi que les erreurs d'interprétation. Mais d'un autre côté, ils perdent aussi en capacité d'exploration.

Mots-clefs (3 à 5) : Pharmacovigilance – Shiny – Software-as-a-Service – Reporting-as-a-Service – Rmarkdown

Calcul d'indicateurs qualité à la Haute Autorité de Santé : Mise en place d'une migration de SAS vers R sans arrêt de la production des indicateurs

Anaëlle Coquelin*, Anaïs Sitruk*, Caroline Prunet*, Agnès Solomiac*, Nelly Le Guen*,
Alwenna Salaün*, Sandrine Morin*, Pierre-Alain Jachiet†

Résumé (max 300 mots)

La transformation numérique, marquée par la collecte massive de données transforme le secteur de la santé. Elle affecte les outils de programmation et d'analyse et leur usage au quotidien, passant de SAS® aux langages Open-Source R/Python. Situation identifiée dès 2021, la Haute Autorité de Santé engage en 2023 ce chantier de migration, tout en assurant ses missions de mesure d'indicateurs en parallèle, sur une période de 2 ans.

Méthode :

Une stratégie est élaborée par le service en charge des indicateurs et la mission data pour identifier les travaux, ressources et temps nécessaires. Les programmes SAS® sont cartographiés, les besoins identifiés

Résultats :

Etapes réalisées depuis 2023 :

1. Partage d'expérience sur la migration logicielle : échanges avec la DREES et l'INSEE (accompagnement par des formations, bonnes pratiques et externalisation).
2. Formations adaptées et auto-formation collaborative : migration facilitée par l'intensification d'«Ateliers R» en interne, existant depuis 2021 par des agents déjà formés. L'apprentissage de R est complété par des formations adaptées aux agents. Des binômes sont constitués pour encourager les échanges sur les bonnes pratiques de programmation et accélérer la prise en main.
3. Priorisation des programmes SAS®.
4. Programmation en R.
5. Pour accélérer le processus, renfort (2024) :
 - Développeur R/Python recruté ;
 - Prestation externe.

=> 70% des programmes ont été migrés.

Conclusion :

Les agents impactés par la migration ont suivi le module d'initiation, d'autres sont prévus en 2025.

* Haute Autorité de Santé, Service Evaluation et Outils pour la qualité et sécurité des soins (SEvOQSS) - contact.iqss@has-sante.fr

† Haute Autorité de Santé, Mission Data (MD) - data@has-sante.fr

SAS® et R cohabitent mais les nouveaux projets sont développés en R/Python. L'objectif est de finaliser la migration en T1/2026. Elle nécessite un investissement important en temps, formation, prestation et appropriation du logiciel.

Perspective : Cette migration modernise les outils de traitement de données ouvrant la voie à une adoption plus large de R. L'étape suivante est la mise en package des fonctions créées.

Mots-clefs (3 à 5) : Migration – SAS® – R – Python – Prestation – Formation

* Haute Autorité de Santé, Service Evaluation et Outils pour la qualité et sécurité des soins (SEvOQSS) - contact.iqss@has-sante.fr

† Haute Autorité de Santé, Mission Data (MD) - data@has-sante.fr

Comment traduire vos documents efficacement dans R

Maëlle Salmon* Yanina Bellini Saibene† Paola Corrales‡ Elio Campitelli§
Pascal Burkhard¶

Résumé (max 300 mots)

La communauté des utilisateur-riche-s de R est mondiale et donc multilingue. Les personnes qui produisent du contenu sur R, qu'il s'agisse de sites de documentation, de livres ou d'articles de blog, adaptent fréquemment leurs documents à des publics multilingues. À rOpenSci, nous avons un projet de publication multilingue actif, pour lequel nous avons traduit notre guide de développement de paquets et d'autres documents de l'anglais vers l'espagnol. Une traduction vers le portugais est également en cours. La réalisation de ce travail a nécessité le développement non seulement de processus humains, mais aussi de deux paquets R : babelquarto pour construire des sites et des livres Quarto multilingues, et babeldown pour utiliser l'API de traduction automatique DeepL afin d'accélérer le processus de traduction des documents Markdown. Les paquets babelquarto et babeldown se sont révélés utiles au-delà de notre projet. Dans cette communication, nous présenterons les deux paquets, leurs fonctionnalités et leurs limites.

Mots-clés (3 à 5) : Markdown - Communauté - Enseignement - Multilinguisme - Communication

Créer une documentation Quarto multilingue avec babelquarto

Notre guide de développement est publié avec Quarto, car il s'agit d'un outil moderne qui, contrairement à Hugo par exemple, est assez facile à apprendre et également utile pour d'autres applications telles que les rapports et les présentations. Malheureusement, Quarto ne permet pas de créer des sites et des livres multilingues, c'est-à-dire des sites et des livres où chaque page a un lien vers sa version dans d'autres langues. C'est pourquoi nous avons décidé de créer un paquet qui permet de construire de tels livres et sites avec Quarto. Basé sur des documents avec différentes extensions en fonction de la langue, et sur des champs de configuration spécifiques, babelquarto construit les différentes versions d'un livre ou d'un site, et ajoute les liens vers d'autres langues. Le paquet fonctionne parfaitement pour notre guide de développement de paquets. Il est également utilisé par d'autres communautés et projets, ce qui donne lieu à des contributions qui rendent le paquet encore plus collaboratif et adapté à d'autres utilisations.

Utiliser un outil de traduction avancée avec babeldown

Nous nous appliquons à réduire les efforts humains grâce à l'automatisation. Tout d'abord, au lieu de traduire des documents à partir de zéro, nous nous appuyons sur une traduction automatique réalisée par DeepL, ce qui nous permet de concentrer notre travail de traduction manuelle sur des aspects tels que le choix des mots les plus appropriés pour notre communauté ou l'utilisation de formulations non générées. Deuxièmement, afin d'automatiser le processus autant que possible, nous ne voulons pas faire de copier-coller à partir d'un outil externe. Troisièmement, nos documents sont préparés en Markdown et la traduction automatique doit donc respecter cette syntaxe. Sur la base de ces trois critères, nous avons créé le paquet R babeldown, qui envoie le contenu des documents Markdown à l'API de traduction DeepL et écrit le résultat dans des documents Markdown corrects. En outre, babeldown permet de mettre à jour les traductions en ne traduisant que les

*rOpenSci, mmaellesalmon@gmail.com

†rOpenSci, yabellini@ropensci.org

‡Monash University

§Monash University

¶Lycée Denis-de-Rougemont

parties qui ont été modifiées, ce qui fait gagner du temps aux traducteurs et de l'argent si on paie pour l'utilisation de l'API.

Leçons apprises

L'infrastructure créée s'intègre bien aux flux de travail R, Git et GitHub et a été utilisée avec succès par rOpenSci pour traduire du contenu en anglais, portugais, espagnol et français. Elle est également utilisée par d'autres projets tels que R4Epi, et le développement reste réactif aux demandes des utilisateur-riche-s.

Références

- <https://docs.ropensci.org/babelquarto/>
- <https://docs.ropensci.org/babeldown/>
- <https://ropensci.org/commcalls/nov2023-multilingual/>

Conception d'applications Shiny avec {golem}

Vincent Guyader* (assisté de Colin Fay et Laurent Spanu)

Résumé du tutoriel

- {golem} est un framework puissant qui permet de structurer le code Shiny, facilitant ainsi la création d'applications robustes, facilement déployables et maintenables sur le long terme. En transformant une application Shiny en package R, {golem} offre de nombreux avantages, tels que l'intégration d'outils de documentation, de tests unitaires et de gestion des dépendances, assurant ainsi une qualité et une fiabilité accrues du code.
-
-
- Ce tutoriel est pensé à la fois pour les débutants en Shiny et pour les utilisateurs expérimentés cherchant à adopter de meilleures pratiques de développement. Nous commencerons par un rappel des concepts fondamentaux de Shiny, avant de plonger dans les spécificités de {golem}. Les participants apprendront à structurer leur projet, à modulariser le code en séparant l'interface utilisateur et la logique serveur, et à utiliser les fonctionnalités avancées de {golem} pour améliorer la maintenabilité et la scalabilité de leurs applications.
-
- Ce tutoriel vise à fournir une méthode de travail rigoureuse, permettant de développer des applications Shiny sur lesquelles on peut compter, tout en favorisant une collaboration efficace au sein des équipes de développement.

Prérequis : maîtrise de base de la programmation en langage R.

Durée : 2h

À réaliser avant de venir au tutoriel :

Des instructions pour la connexion à l'environnement utilisé (bakacode) seront envoyées aux inscrits avant le début du tutoriel.

* CTO chez ThinkR, vincent@thinkr.fr

Création d'un package d'échantillonnage de polytope en modélisation linéaire inverse

Jacques Bréhélin ^{*} Valérie Girardin [†] Théo Grente [‡] Nathalie Niquil [§]
Philippe Regnault [¶]

Résumé

On appelle «réseaux trophiques», l'ensemble des interactions d'ordre alimentaire entre différentes espèces d'un écosystème. Celles-ci peuvent être modélisées sous forme de graphes pondérés et orientés, dans lesquels les arêtes (flux d'échanges de biomasse) sont soumises à des contraintes linéaires, cette modélisation est appelée «modélisation linéaire inverse» (LIM). L'ensemble de solutions satisfaisant ces contraintes constitue un polytope.

`Samplelim` est un package R permettant l'échantillonnage de ces polytopes de solutions en utilisant des méthodes de Monte-Carlo par chaînes de Markov (MCMC). Il s'agit d'une version améliorée d'un package plus ancien nommé `limSolve`, qui avait l'inconvénient d'avoir des temps d'exécution très élevés pour certains problèmes de grande dimension. `Samplelim` corrige ce défaut et permet des gains de temps considérables.

Cette présentation commencera par une brève présentation du package et se concentrera ensuite principalement sur les différentes étapes en vue de la soumission d'un package sur le CRAN (Comprehensive R Archive Network), le dépôt officiel des packages R. Parmi ces étapes, on retrouve notamment l'écriture d'un didacticiel pour le package à destination d'utilisateurs pouvant avoir des niveaux de connaissances très variés, mais aussi la réalisation de tests unitaires, afin de s'assurer du bon fonctionnement de l'ensemble des parties du code constituant le package et aussi de faciliter sa mise à jour. Ces tests peuvent notamment se faire avec le package `testthat`. Enfin, il est nécessaire de s'assurer que le package respecte un ensemble de critères stricts, définis par le CRAN, afin de garantir son admissibilité.

Mots-clefs : CRAN – MCMC – Package – Tests unitaires – Vignette

Références

- J. Bréhélin. Vérification et didacticiel d'un package d'échantillonnage de polytopes en analyse de réseaux trophiques. Mémoire de master informatique, Université de Caen Normandie. https://breheli241.users.lmno.cnrs.fr/rapport_M2.pdf, 2024.
- Grente T. Niquil N. et Regnault P. Girardin, V. Comparing and updating r packages using mcmc algorithms for linear inverse modeling of metabolic networks. *HAL Preprint*, 2024.
- K. Van Den Meersche., K. Soetaert., and D. Van Oevelen. `xsample()` : An r function for sampling linear inverse problems. *Journal of Statistical Software*, 2009.

^{*}Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Normandie, CNRS-6139, BP5186, Caen, 14032, France., breheli241@unicaen.fr

[†]Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Normandie, CNRS-6139, BP5186, Caen, 14032, France., valerie.girardin@unicaen.fr

[‡]GREYC, Université de Caen Normandie, CNRS 6072, Caen, 14000, France., theo.grente@unicaen.fr

[§]Laboratoire Biologie des Organismes et Ecosystemes Aquatiques, Université de Caen Normandie, CNRS-8067, Esplanade de la Paix, Caen, 14000, France., nathalie.niquil@unicaen.fr

[¶]Laboratoire de Mathématiques de Reims, Université de Reims Champagne-Ardenne, CNRS-9008, BP1039, Reims, 51687, France., philippe.regnault@univ-reims.fr

Créer des applications RAG-LLM avancées avec R Shiny

Mohamed El Fodil Ihaddaden*

Résumé (max 300 mots)

La génération augmentée par récupération (RAG) combine les capacités génératives des grands modèles de langage (LLMs) avec une récupération d'informations ciblée et contextuelle, permettant ainsi de développer des applications d'IA performantes. Bien que R propose certains packages pour interagir avec les LLMs, il ne dispose pas des outils nécessaires pour implémenter un pipeline RAG complet. En revanche, l'écosystème Python offre de nombreuses solutions pour le RAG, avec des bibliothèques phares comme LangChain et LlamaIndex. Dans cette session, je présenterai ce que je considère comme la meilleure approche pour intégrer les capacités RAG de Python au sein d'une application R Shiny. Spoiler : ce n'est pas reticulate.

Mots-clefs (3 à 5) : Shiny - IA - LLM - RAG

*HDI Global SE, mohamedelfodil.ihaddaden@hdi.de

De la carte aux modèles spatiaux

Corentin Visée^{1,2}

Résumé (max 300 mots)

À ses débuts, la production de cartes visait à représenter des territoires à des fins de navigation, d'administration et de gestion des ressources. Cet objectif demeure central à travers les siècles, en témoignent les travaux de van Deventer ou encore de Philippe Vandermaelen entre les XVIIe et XIXe siècles. La cartographie a évolué progressivement, adoptant de nouveaux objectifs, notamment l'analyse spatiale et la représentation thématique des données, avec notamment le travail de Jon Snow sur le choléra lors de l'épidémie de 1854 à Londres¹.

Aujourd'hui, l'augmentation des quantités de données géolocalisées, générées par divers produits technologiques, rend nécessaire leur présentation au moyen de cartes et d'infographies complexes. Cette augmentation des données rend nécessaires les agrégations et autres techniques de visualisation pour produire des cartes présentables et efficaces auprès du public visé.

Au cours de ma présentation, je présenterai différentes méthodes de représentation de l'information géographique sur des cartes à l'aide du package **tmap**². Je présenterai les biais et règles à maîtriser pour représenter ces données géolocalisées (Figure 1).

Je présenterai également les règles à maîtriser pour modéliser des données spatiales. En particulier, l'hypothèse d'indépendance des observations n'est pas valable dans le cadre d'observations géolocalisées, car les observations proches dans l'espace se ressemblent davantage que les observations éloignées : on parle d'autocorrélation spatiale. La seconde moitié de la présentation aura pour objectif d'étudier cet effet, de le quantifier et de proposer des pistes pour produire des modèles robustes à l'aide du package **spdep**³. Pour conclure, je présenterai quelques travaux utilisant R et les données spatiales du département de géographie de l'université de Namur, portant notamment sur l'épidémiologie spatiale⁴, la cartographie de la population⁵ ou encore la pénurie des médecins en zone rurale de Belgique.

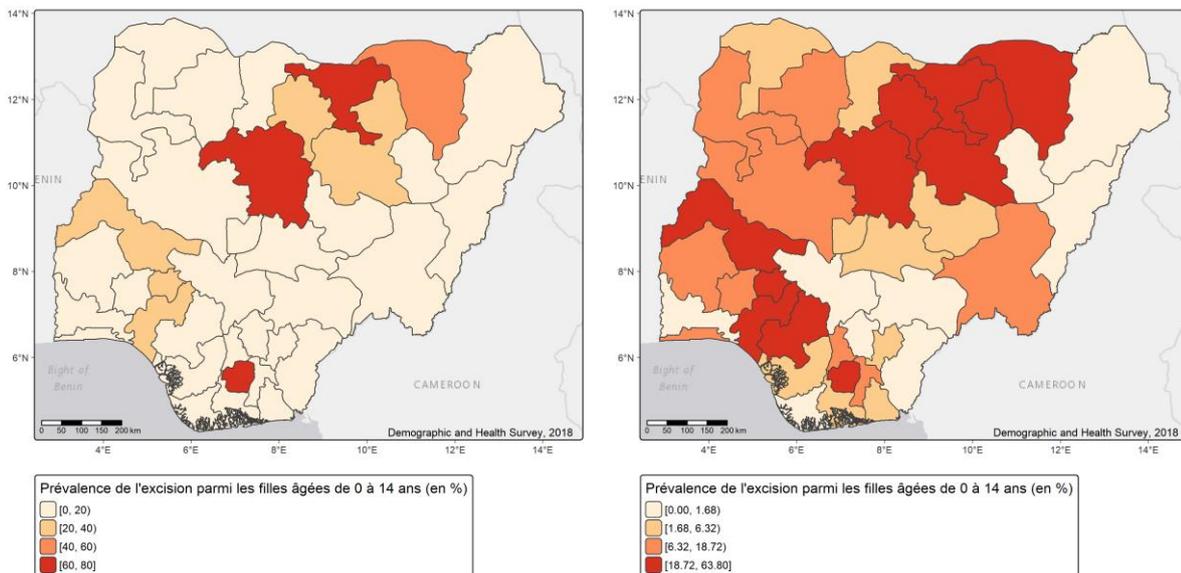


Figure 1 : Comparaison des méthodes de discrétisation entre deux données similaires au Nigéria^{6,7} (gauche : intervalles égaux, droite : quantiles).

Mots-clefs (3 à 5) : Cartographie – Géostatistiques – Géographie – Cartes

1 : Département de Géographie, Université de Namur

2 : Institute of Life, Earth and Environment, Université de Namur

Références :

- [1] Borak, Jonathan MD. The Ghost Map: The Story of London's Most Terrifying Epidemic and How It Changed Science, Cities, and the Modern World. *Journal of Occupational and Environmental Medicine* 49(12):p 1403-1406, December 2007. | DOI: 10.1097/JOM.0b013e3181578d88
- [2] Tennekes M (2018). "tmap: Thematic Maps in R." *Journal of Statistical Software*, 84(6), 1–39. doi:10.18637/jss.v084.i06.
- [3] Pebesma E, Bivand R (2023). *Spatial Data Science With Applications in R*. Chapman & Hall. <https://r-spatial.org/book/>.
- [4] Morlighem, Camille, Celia Chaiban, Stefanos Georganos, Oscar Brousse, Jonas Van de Walle, Nicole P. M. van Lipzig, Eléonore Wolff, Sébastien Dujardin, and Catherine Linard. 'The Multi-Satellite Environmental and Socioeconomic Predictors of Vector-Borne Diseases in African Cities: Malaria as an Example'. *Remote Sensing* 14, no. 21 (27 October 2022): 5381. <https://doi.org/10.3390/rs14215381>.
- [5] Visée, Corentin, Camille Morlighem, Catherine Linard, Abdoulaye Faty, Sabine Henry, and Sébastien Dujardin. 'Addressing Bias in National Population Density Models: Focusing on Rural Senegal'. *PLOS ONE* 19, no. 11 (12 November 2024): e0310809. <https://doi.org/10.1371/journal.pone.0310809>.
- [6] National Population Commission - NPC and ICF. 2019. Nigeria Demographic and Health Survey 2018 - Final Report. Abuja, Nigeria: NPC and ICF. Available at <http://dhsprogram.com/pubs/pdf/FR359/FR359.pdf>.
- [7] Visée, Corentin, Morlighem, Camille and Nnanatu, Chibuzor Christopher (2024) A robust cross-sectional assessment of the impacts of COVID-19 pandemic on the prevalence of female genital mutilation among 0 - 14 years old girls in Nigeria. *Women's Health*. (In Press)

Des Learning Analytics qui ne manquent pas d'R

Patrice Armand* Philippe Grosjean†

Résumé (max 300 mots)

Dans le cadre d'un projet de Learning Analytics initié à l'Université de Mons, nous avons mis en place une plateforme de reporting basée sur plumber, Shiny et ShinyProxy. Une première application a été développée. Elle permet aux enseignants d'avoir une vision globale de l'état de réalisation des quiz présents dans leur cours Moodle. Elle leur fournit également des informations sur l'évolution de la participation à ces différents quiz ainsi que des informations leur permettant d'identifier des étudiants en risque de décrochage. D'autres applications sont en cours de développement.

Mots-clefs : plumber – ShinyProxy – Learning Analytics – Moodle – UMONS

Développement

Dans le cadre d'un projet de Learning Analytics initié à l'Université de Mons (UMONS) sur fonds européens RRF (Recovery and Resilience Facility), nous avons mis en place une plateforme de reporting basée sur plumber, Shiny et ShinyProxy. Cette plateforme doit potentiellement être accessible à tous les enseignants, à tous les étudiants ainsi qu'à d'autres utilisateurs institutionnels.

R était un candidat possible parmi d'autres (python - Dash, javascript - 3D.js ou plotly.js, Power BI, ...) Nous avons écarté rapidement Power BI pour son coût et son manque de souplesse. R s'est imposé car il permettait de répondre complètement au cahier des charges. Par ailleurs, une première expérience de Learning Analytics utilisant R existait déjà à l'UMONS (<https://learnitr.github.io/learnitrdoc/>). Enfin, du temps avait déjà été investi dans l'acquisition de compétences en R au moment de faire le choix.

Concernant la plateforme de publication des rapports, nous avons dû exclure Posit Connect en raison de son coût trop élevé compte tenu du grand nombre d'utilisateurs potentiels. ShinyProxy a été préféré à l'utilisation d'un Shiny Server local afin de rencontrer notre besoin de scalabilité (<https://www.appsilon.com/post/alternatives-to-scaling-shiny>)

L'architecture de la plateforme est classique. C'est du n-tiers. L'utilisation d'une REST API est pratique. Il est prévu en effet de développer plusieurs applications et chacune d'entre elles devra récupérer des données depuis plusieurs sources (un LRS, la base de données de Moodle, ...). Cela autorise également le développement ultérieur d'autres clients. L'utilisation d'un serveur ShinyProxy quant à elle présente plusieurs avantages : ShinyProxy publie les applications en mode web, il est scalable et permet de faire du SSO avec Entra ID.

Nous pouvons signaler, comme premiers retours d'expériences, quelques difficultés liées à la sécurité réseau, essentiellement concernant l'autorisation des WebSockets. Les problèmes de performances ont quant à eux été résolus rapidement en adaptant le code de l'application (réutilisation de la connexion à la DB, affichage progressif des graphiques).

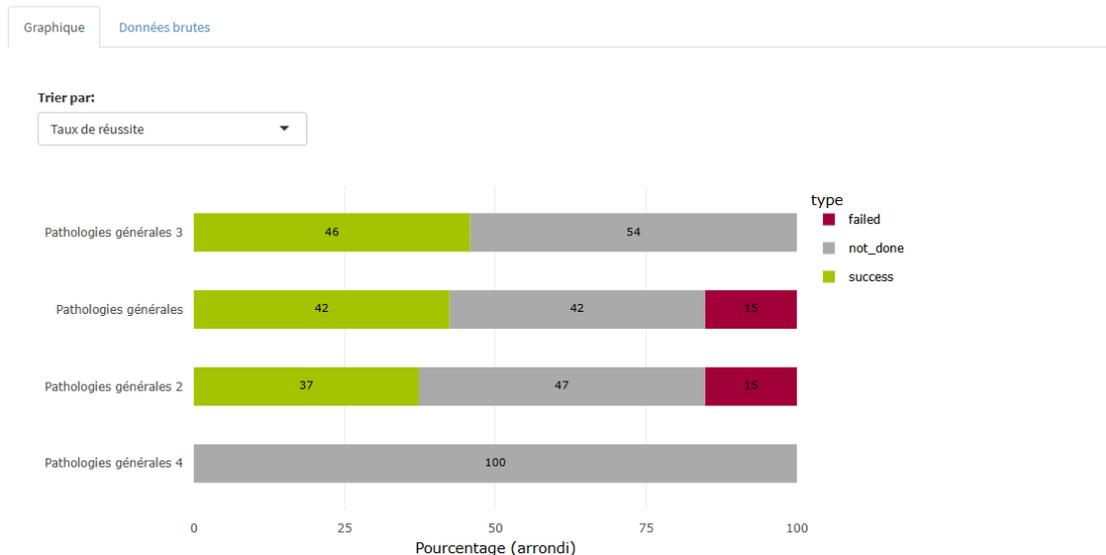
* Service d'appui pédagogique, Université de Mons, Belgique, patrice.armand@umons.ac.be

† Service d'Écologie numérique, Instituts Infortech et Complexys, Université de Mons, Belgique, Philippe.grosjean@umons.ac.be

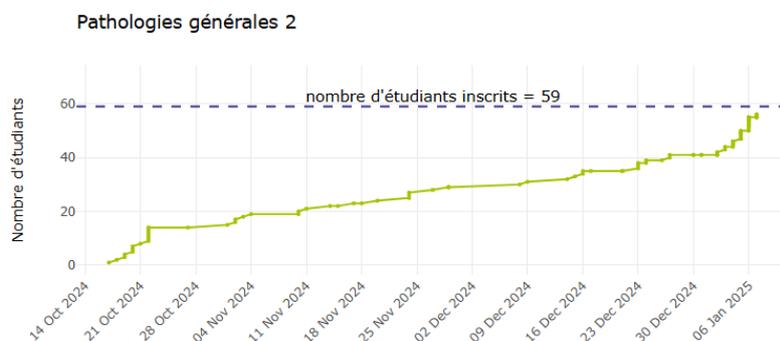
Un certain nombre de choses restent encore à faire. Parmi celles-ci, citons la sécurisation de l'API par l'utilisation du SSL et d'une Api Key (même si des restrictions d'accès ont déjà été configurées sur le firewall du serveur et le fait qu'un Web Application Firewall est présent dans le réseau), la mise en place du SSO, le codage d'un mécanisme d'autorisation plus élaboré, l'amélioration des performances du LRS qui sera bientôt utilisé par la plateforme.

Voici quelques graphiques présents dans notre première application. D'autres seront présentées sur le poster.

ETAT D'AVANCEMENT DE LA CLASSE



EVOLUTION DE LA PARTICIPATION :



Dans un avenir proche, nous envisageons d'améliorer l'application existante en l'élargissant à d'autres activités Moodle que les quiz. Nous souhaitons également créer une autre application exploitant les traces d'apprentissage générées par les activités Moodle de type « vidéos interactives H5P ».

Références

P. Grosjean and G. Engels: LearnIt::R plateforme pédagogique pour l'apprentissage de R, <https://learnitr.github.io/learnitrdoc/>, June 23, 2024

P. Przytula: Alternatives to Scaling Shiny with RStudio Connect or Custom Architecture, <https://www.appsilon.com/post/alternatives-to-scaling-shiny>, December 18, 2018

Pierre Lamarche*

Résumé (max. 300 mots)

Le package DoReMIFaSol, aujourd'hui disponible sur Github, est un package R s'appuyant sur les données mises à disposition en accès libre sur le site de l'Insee. L'objectif principal de DoReMIFaSol est de simplifier le processus de récupération, de traitement et d'analyse des données diffusées par l'Insee. Il permet aux utilisateurs de télécharger directement des jeux de données depuis le site de l'Insee, en automatisant les étapes fastidieuses de collecte et de mise en forme des données ; il permet de scripter la récupération de ces données, tout en cherchant à recenser les différents jeux de données disponibles et devient un outil de traçage de la source de la donnée, comme peuvent l'être d'autres packages tel le package eurostat.

Parmi les fonctionnalités du package, on trouve l'accès à différentes API, comme l'API Sirene, qui permet de requêter par ce biais la donnée relative au répertoire Sirene, le répertoire d'entreprises maintenu par l'Insee. Depuis peu, l'Insee a également développé une nouvelle offre incluant un catalogue de données mises à disposition sur son site Internet. Cette nouvelle offre, dénommée Mélodi, doit permettre de mieux référencer les jeux de données mis à disposition par l'Insee. DoReMIFaSol se connecte désormais à ce catalogue pour actualiser la liste de jeux de données que le package permet de générer, et ce de manière quotidienne. Cette mise à jour dynamique de l'information est possible grâce au SSP Cloud, une infrastructure de datascience ouverte maintenue par l'Insee, qui contient des outils permettant à l'utilisateur de construire de manière autonome des processus automatisés, entre autres choses. Le service sous forme d'API permet ainsi de satisfaire des besoins de manière agnostique, bénéficiant à des outils tiers construits de manière indépendante et décentralisée.

Mots-clés : Open data – Diffusion – Reproductibilité

* Insee, pierre.lamarche@insee.fr

Dynamisez vos documents Quarto grâce à Observable JS

Antoine Bichat*

Résumé

Depuis son lancement en 2022, Quarto (Allaire et al. 2024) permet de générer des publications techniques et scientifiques dans divers formats. Contrairement aux sorties PDF, qui produisent des documents figés, celles au format HTML favorisent une interactivité poussée entre le document et le lecteur final. Plusieurs formidables packages R, tels que `{ggiraph}` (Gohel and Skintzos 2024) pour les graphiques ou `{DT}` (Xie, Cheng, and Tan 2024) pour les tableaux profitent déjà de cette interface interactive. Toutefois, ces solutions se limitent souvent aux données précalculées et ne permettent pas d’explorer les jeux de données à la volée, comme avec un filtre par exemple.

Pour dépasser cette limitation, Quarto propose l’intégration de la technologie OJS (Observable JavaScript), qui exécute du code JavaScript réactif directement dans le navigateur et s’articule facilement avec R. On peut ainsi créer des documents capables de recalculer et d’actualiser leur contenu en temps réel. Cette solution ouvre la voie à des documents véritablement dynamiques, tout gardant la simplicité d’un document `.qmd`.

Mots-clefs : Quarto - Visualisation - Interactivité - Communication

Développement

Cette présentation se veut être un tutoriel pour découvrir OJS et créer ses premier documents avec. Nous commencerons avec un premier graphique puis détaillerons les différentes options de personnalisation ainsi que la grammaire des graphiques (Wickham 2010) implémentée dans OJS. Enfin, nous conclurons par une démonstration de l’intégration de widgets interactifs, permettant d’enrichir l’expérience utilisateur.

Références

- Allaire, J. J., Charles Teague, Carlos Scheidegger, Yihui Xie, Christophe Dervieux, and Gordon Woodhull. 2024. “Quarto.” <https://doi.org/10.5281/zenodo.5960048>.
- Gohel, David, and Panagiotis Skintzos. 2024. “Ggiraph: Make ‘Ggplot2’ Graphics Interactive.” <https://davidgohel.github.io/ggiraph/>.
- Wickham, Hadley. 2010. “A Layered Grammar of Graphics.” *Journal of Computational and Graphical Statistics* 19 (1): 3–28. <https://doi.org/10.1198/jcgs.2009.07098>.
- Xie, Yihui, Joe Cheng, and Xianying Tan. 2024. “DT: A Wrapper of the JavaScript Library ‘DataTables’” <https://CRAN.R-project.org/package=DT>.

*Spotlight Medical, antoine.bichat@spotlightmedical.com

Détection et quantification de manipulations comptables autour d'un seuil psychologique avec R

Marie Chavent* Véronique Darmendrail† Delphine Feral‡ Frédéric Pourtier§
Jérôme Saracco¶

Résumé

Cette communication présente l'implémentation dans R d'une nouvelle méthode statistique d'identification/détection et de quantification des manipulations de données comptables autour de seuils psychologiques, comme le résultat nul, la variation nulle de résultat ou, le résultat anticipé par le consensus des analystes et investisseurs. La méthode devrait aussi permettre l'analyse/comparaison des manipulations entre diverses sous-populations. Notons que, si l'illustration actuelle est faite avec des problématiques et des données comptables, les codes et méthodes développés peuvent être utilisés dans d'autres champs, pour l'étude des comportements face à des effets de seuils psychologiques.

Plus précisément, un algorithme de type EM a été proposé pour estimer les paramètres sous-jacents du modèle considéré, qui est un modèle de mélanges (impliquant une loi Exponentielle pour modéliser le comportement des manipulateurs, et un mélange de deux distributions Gaussiennes pour modéliser le comportement d'un individu n'ayant pas manipulé). Un test d'adéquation des données au modèle est disponible. La méthode fournit des intervalles de confiance Bootstrap pour les paramètres estimés du modèle. De plus, différentes représentations graphiques sont possibles via l'utilisation de la fonction `plot`. Après avoir évalué le bon comportement numérique de la méthodologie sur des données simulées, l'approche proposée est illustrée sur des données réelles pour étudier la gestion des bénéfices.

Les diverses fonctions R sont faciles à mettre en œuvre sur des benchmarks de performance financière ou extra-financière ou en dehors du domaine de la comptabilité. Un package R sera bientôt disponible publiquement.

Mots-clefs : algorithme EM – estimation de la densité – gestion de seuils – gestion du résultat comptable – modèle de mélange

*IMB, Université de Bordeaux & ASTRAL, Centre Inria de l'Univ. de Bordeaux, marie.chavent@u-bordeaux.fr

†IRGO, Université de Bordeaux, veronique.darmendrail@u-bordeaux.fr

‡IMB, Université de Bordeaux, delphine.feral@u-bordeaux.fr

§IRGO, Université de Bordeaux, frederic.pourtier@u-bordeaux.fr

¶ENSC, Bordeaux INP & ASTRAL, Centre Inria de l'Univ. de Bordeaux, jerome.saracco@bordeaux-inp.fr

Développement de workflows en R pour le nettoyage, la validation et la visualisation de données spatiales en écologie

Jordan Benrezkallah¹, Natasha de Manincor², Ahlam Sentil³, Kévin Tougeron⁴ & Denis Michez⁵

Résumé

Avec l'augmentation des volumes de données en sciences, il est essentiel de disposer de workflows reproductibles pour le nettoyage, l'exploitation et l'interopérabilité des données. En écologie, ces enjeux sont particulièrement cruciaux pour le suivi des espèces et la conservation. Dans ce cadre, nous avons développé un workflow en R pour traiter de grands jeux de données sur les pollinisateurs européens, en lien avec les projets **PULSE** et **SAFEGUARD**. Ces initiatives visent à alimenter la Liste Rouge européenne des abeilles et analyser les tendances des pollinisateurs à partir de bases de données consolidées. Nous avons ainsi normalisé 69 jeux de données issus de toute l'Europe, représentant plus de 5,5 millions d'enregistrements et couvrant **2081 espèces d'abeilles sauvages**.

Le **workflow** développé en **R** inclut les étapes suivantes :

1. **Importation et correction** des formats (encodage, caractères spéciaux, dates, coordonnées).
2. **Contrôle qualité** (unicité des identifiants, formats des coordonnées, cohérence temporelle).
3. **Vérification des noms d'espèces**, sur base d'une liste d'espèces notamment via des recherches approximatives (*fuzzy search*).
4. **Assignation de la classification taxonomique** en attribuant les rangs taxonomiques.
5. **Attribution des divisions administratives** aux enregistrements.
6. **Génération automatisée de cartes spatiales** pour validation par des experts taxonomistes.

Ce processus permet d'obtenir des données nettoyées et validées pour estimer l'aire de répartition des espèces et évaluer leur statut de conservation. Les données sont structurées selon le **standard Darwin Core**, garantissant leur interopérabilité avec d'autres bases de données et facilitant leur réutilisation. L'ensemble du workflow ainsi que les scripts sont accessibles sur un **dépôt GitHub**, suivant ainsi les principes **FAIR** (*Findable, Accessible, Interoperable, Reusable*).

Cette présentation mettra en avant les défis du traitement de données écologiques massives et illustrera la puissance de R pour leur standardisation et leur visualisation.

Mots-clefs : Biologie - Statistique Spatiale - Data - Reproductibilité

1 Laboratoire de Zoologie et laboratoire d'Écologie des Interactions et Changements globaux, Université de Mons, jordan.benrezkallah@umons.ac.be

2 Laboratoire de Zoologie, Université de Mons, natasha.demanincor@umons.ac.be

3 Laboratoire de Zoologie, Université de Mons, ahlam.sentil@umons.ac.be

4 Laboratoire d'Écologie des Interactions et Changements globaux, Université de Mons, kevin.tougeron@umons.ac.be

5 Laboratoire de Zoologie, Université de Mons, denis.michez@umons.ac.be

Explorer le CRAN avec le package RWsearch

Patrice Kiener 1*

Résumé

Chaque jour, le CRAN supprime en moyenne 1.6 package, accueille 6.1 nouveaux packages et en renouvelle 20.5. Sur un mois, cela équivaut à 48 packages supprimés, 184 nouveaux packages et 617 packages renouvelés. Le package **RWsearch** permet de détecter les nouveaux packages dès leur premier jour d'apparition sur le CRAN et fournit toutes les informations utiles pour les évaluer avec notamment un graphe des dépendances récursives qui montre quelquefois des centaines de packages, bien plus que ce que le mainteneur imagine. RWsearch permet des recherches par mots-clés pour se constituer ses propres fichiers de packages pertinents, une bibliographie exhaustive ou la constitution d'une task view. Je détaillerai une nouveauté (suggérée par Toby Hocking) qui sépare les packages qui peuvent être mis à jour immédiatement de ceux, compilés et chargés dans le namespace, qui requièrent une mise à jour ultérieure dans un R-vanilla. Retrouvez le package RWsearch sur le CRAN à l'adresse : <https://cran.r-project.org/package=RWsearch>

Mots-clefs : CRAN - Package - Reproductibilité - Bibliographie

*InModelia, patrice.kiener@inmodelia.com

Faciliter la prise en compte des risques économiques en agriculture grâce à R et Shiny

Morgane Topart¹ Aïcha Ronceux²

Résumé

Les agriculteurs font face à une forte incertitude économique liée aux variations des prix et des rendements. Pourtant, ces risques sont rarement intégrés dans les évaluations économiques, faute d'outils adaptés (M. Topart, 2019). Pour combler ce manque, le projet Strat&co a développé Oser, une application Shiny qui permet de visualiser simplement l'impact de ces fluctuations sur la rentabilité des exploitations.

Oser a été développé de manière itérative, en collaboration avec des utilisateurs potentiels. Initialement conçu comme un prototype simple, croisant rendements et prix via un produit cartésien pour donner une marge économique, il s'est complexifié pour répondre aux besoins d'accessibilité et de personnalisation. Golem a été utilisé pour structurer et modulariser l'application, tandis que bs4Dash a permis d'améliorer l'ergonomie et l'interface. Les fonctions du package Shelf ont été adaptées pour Oser et ainsi permettre aux non-experts en statistiques d'ajuster les paramètres selon leur propre contexte.

Les premiers retours sont positifs. L'interface, réactive et personnalisable, rend l'outil agréable à utiliser. Les résultats peuvent être visualisés sous forme de graphiques variés (histogramme, boîte à moustache, matrice de gain) et exportés dans un rapport Word. Grâce à son approche visuelle et interactive, Oser facilite l'évaluation des risques économiques agricoles et aide conseillers et agriculteurs à mieux anticiper les fluctuations des prix et des rendements.

Oser est disponible sur <https://mtopart.shinyapps.io/osser/>.

Mots-clefs : Accompagnement – Agriculture - Shiny - Visualisation - Variabilité

Références

M. Topart, Synthèse des enquêtes conseillers et agriculteurs – projet Strat&co, 2019

¹ Agro-Transfert Ressources et Territoires, m.topart@agro-transfert-rt.org

² Agro-Transfert Ressources et Territoires, a.ronceux@agro-transfert-rt.org

Framework d'import de base de données en fichiers pour la recherche clinique : le package {EDCimport}

Dan Chaltiel^{*†}

Résumé (max 300 mots)

En recherche clinique, les bases de données sont générées par des logiciels d'Electronic Data Capture (EDC) qui collectent et organisent des informations provenant de patients inscrits dans des études. Ces logiciels exportent généralement les données sous la forme d'une archive contenant une collection de fichiers, dans des formats tels que CSV, SAS, ou XPT. Chaque ligne de ces fichiers est une observation d'un patient, caractérisé par un identifiant unique, le SUBJID. L'importation manuelle de ces fichiers dans R est un processus laborieux, et qui ne prend pas en compte la relation entre les fichiers, centrée autour du SUBJID.

{EDCimport} vient apporter une solution en proposant une toolbox d'import pour les formats les plus courants, en gérant notamment l'application des labels et formats. Le package offre une gamme d'outils pour le traitement de ces bases de données: des fonctions de recherche globale, de standardisation, de vérification des données, ainsi que de visualisation graphique. Une application shiny interne au package permet également de visualiser l'ensemble de sa base de données de façon interactive, avec une interface plus riche et dynamique que celles disponibles dans les logiciels de programmation R.

{EDCimport} est un package encore en cours de développement, mais son utilisation permet de gagner un temps considérable lors de l'analyse des essais cliniques. En facilitant l'identification d'erreurs, il permet également d'améliorer la qualité et l'intégrité des données traitées.

Mots-clefs (3 à 5) : Database – Import – Biostatistique – Package

* Bureau de Biostatistique et d'Épidémiologie, Gustave Roussy, Université Paris-Saclay, Villejuif, France

† Oncostat, CESP U1018, Inserm, Équipe Labellisée Ligue Contre le Cancer, Université Paris-Saclay, Villejuif, France
dan.chaltiel@gustaveroussy.fr

Introduction pratique à R : les fondamentaux pour organiser, analyser et visualiser

Claire Della-Vedova*

Résumé du tutoriel

Ce tutoriel est destiné aux débutants sans expérience préalable ainsi qu'aux utilisateurs occasionnels souhaitant maîtriser les bases de l'analyse et de la visualisation de données. Durant ces deux heures, vous serez guidé étape par étape dans l'environnement RStudio, apprenant à manipuler vos données de manière simple et efficace.

Nous débuterons par une introduction à l'organisation des projets sous R, un outil essentiel pour structurer vos fichiers et scripts de manière optimale. Vous explorerez ensuite les quatre fenêtres principales de RStudio (console, script, environnement, graphiques) pour comprendre comment interagir avec vos données.

Vous découvrirez les premiers éléments du langage R, vous familiarisant avec la syntaxe de base : création de variables, utilisation de fonctions simples, et manipulation des objets essentiels tels que les vecteurs et les dataframes. Vous apprendrez également à importer des données au format CSV et Excel (XLSX), les formats les plus couramment utilisés dans l'analyse de données.

Enfin, vous serez initié à la création de vos premiers graphiques grâce au package *esquisse*, un outil interactif qui facilite la création de visualisations claires et intuitives, sans nécessiter une maîtrise approfondie de R.

Prérequis : Seules des connaissances de base de l'outil informatique sont nécessaires : création de dossier, navigation dans l'arborescence, utilisation du copier-coller, et du pack office.

Durée : 2h

À réaliser avant de venir au tutoriel :

- installer R
- installer RStudio
- installer l'application XQuartz pour celles et ceux qui sont sous macOS : <https://www.xquartz.org/>

* DellaData (prestations et formations méthodologie / statistiques / langage R), claire@delladata.fr

La traduction du mlverse en français

Christophe Regouby*

Résumé (max 300 mots)

Le mlverse est un univers de packages reliés de près ou de loin à {torch}. Voici un état de l'art de la traduction de 10 packages, de leurs messages, de leur documentation, de leur vignettes, et de l'outillage associé. Une conférence gesticulée comme appel à **utiliser** les packages en français, et peut-être à contribuer à l'effort de traduction.

Mots-clefs (3 à 5) : Enseignement - Accessibilité - Internationalisation

Références

mlverse
{luz.fr}

*pas d'affiliation, Christophe.regouby@free.fr

Le format Parquet et l'écosystème DuckDB: l'essayer c'est l'adopter!

Lino Galiana*

Résumé (max 300 mots)

Issu à l'origine de l'écosystème du *big data*, le format **Parquet**, par ses propriétés idéales pour l'analyse de données, s'est diffusé au-delà de ce cercle restreint. Léger, rapide, interopérable, fiable grâce à ses métadonnées et optimisé pour des opérations classiques d'analyse (filtres de colonnes ou de lignes), les avantages à ce format de données sont nombreux.

Ce format s'intègre dans un écosystème plus large, articulé autour des projets Arrow et DuckDB. Ces deux bibliothèques, très bien intégrées au *tidyverse*, proposent un ensemble d'outils pour lire et échanger des données dans ce format. Devenu une alternative viable au CSV pour les statisticiens et *data scientists*, ce format est l'une des pierres angulaires d'institutions diffusant des données: Insee, **data.gouv** ou encore **HuggingFace**.

Cette présentation évoquera les avantages du format Parquet, fera un panorama de l'écosystème associé et reviendra, à partir de plusieurs exemples, sur le gain de confort et de performance de ce format de données pour partager et analyser des données. Elle évoquera également la manière dont ce format s'intègre dans la stratégie de mise à disposition interne de données de l'Insee dans le cadre de la transition de **SAS** à **R** et dans sa politique de diffusion de données volumineuses.

Mots-clefs (3 à 5) : Data - Diffusion - Optimisation - Bonnes pratiques

*Insee, lino.galiana@insee.fr

Les élections européennes et législatives 2024

Patrice Kiener 1*

Résumé

Au soir des élections européennes du 9 juin 2024, Emmanuel Macron a dissout l'assemblée nationale et a appelé à des élections législatives le 30 juin 2024 pour le premier tour et le 7 juillet 2024 pour le deuxième tour. Il est raisonnable de penser que les électeurs n'ont pas changé d'opinion en 3 semaines. Sauf que les systèmes de vote pour le parlement européen et le parlement français sont différents et changent complètement la manière de voter des électeurs : vote éparpillé lors d'une élection proportionnelle à 1 tour contre vote concentré sur les premiers candidats lors d'une élection majoritaire à 2 tours.

Le vote électronique était institué dans les 11 circonscriptions des français de l'étranger. Il a permis une fraude dans au moins 4 circonscriptions, dont la circonscription *Bruzelles-Amsterdam-Luxembourg* où 29.143 nouveaux électeurs ont été inscrits en 3 semaines en provenance des électeurs radiés dans les circonscriptions de la métropole. Ces nouveaux inscrits ont largement favorisé quelques candidats.

« On n'organise pas des élections pour les perdre » ... Un dignitaire africain

Mots-clefs : Élections - Fraude - AC - ACP - Log-ratios centré (CLR)

*InModelia, patrice.kiener@inmodelia.com

Faites la différence : dérivées numériques rapides et précises avec `pnd`

Andreï V. KOSTYRKA*

28 février 2025

Résumé

Nous présentons un nouveau paquet pour le calcul des dérivées numériques par des approximations en différences finies, avec des capacités de parallélisation et une sélection optimale du pas pour la meilleure approximation. Le nouveau paquet, intitulé `pnd`, facilite le calcul efficace des dérivées, gradients, matrices Jacobiennes et Hessiennes, permettant davantage d'évaluations pour réduire les erreurs de méthode et de donnée. Conçu pour être compatible avec le paquet `numDeriv`, `pnd` introduit des fonctionnalités telles que le calcul des dérivées d'ordre arbitraire, l'approximation directe des matrices Hessiennes sans différences répétées, et la parallélisation des fonctions lentes sur tous les systèmes d'exploitation populaires.

Mots-clés : dérivées numériques ; différences finies ; algorithmes parallèles ; erreur d'arrondi.

Proposition de communication (260 mots)

Les dérivées numériques sont fondamentales dans les algorithmes d'optimisation basés sur le gradient, utilisés en statistiques, en apprentissage automatique et en calcul scientifique. Nous développons des algorithmes améliorés de différenciation numérique ainsi qu'un nouveau paquet open-source pour **R**, `pnd` (Parallel Numerical Derivatives), qui propose des implémentations rapides, précises et parallélisées de ces méthodes.

Le paquet `pnd` étend les méthodes classiques de différences finies en intégrant des algorithmes de sélection du pas de dérivation, permettant de trouver un équilibre entre l'erreur de troncature et l'erreur d'arrondi. Cela le rend particulièrement utile dans des domaines comme l'optimisation, les statistiques et les mathématiques appliquées, où les dérivées numériques jouent un rôle clé dans l'estimation et les algorithmes basés sur le gradient.

Nous commençons par présenter des méthodes pratiques pour sélectionner le pas optimale dans les dérivées numériques. Ensuite, nous explorons quatre algorithmes populaires de sélection de pas basés sur les données — jusqu'ici absents des logiciels open-source — et montrons comment nos améliorations renforcent leur rapidité et leur stabilité sur des systèmes multi-cœurs. Enfin, nous illustrons la syntaxe du paquet `pnd`, conçu comme un remplacement direct du paquet populaire `numDeriv`. Ses fonctions sont compatibles avec tout solveur ou optimiseur basé sur le gradient et permettent

*Université du Luxembourg — Faculté de Droit, Économie et Finances (FDEF).

un contrôle précis du compromis entre précision et vitesse. Nous montrons comment exploiter efficacement la puissance des machines multi-cœurs modernes pour un calcul rapide des gradients et des Jacobiennes des fonctions intensives en calcul. Des analyses comparatives de plusieurs logiciels mettent en évidence la précision et la robustesse de `pnd`, avec des visualisations soulignant l'importance d'une sélection appropriée du pas.

Références

- CURTIS, A. R., & REID, J. K. (1974). The Choice of Step Lengths When Using Differences to Approximate Jacobian Matrices. *IMA Journal of Applied Mathematics*, 13(1), 121-126. <https://doi.org/10.1093/imamat/13.1.121>
- DUMONTET, J., & VIGNES, J. (1977). Détermination du pas optimal dans le calcul des dérivées sur ordinateur. *RAIRO. Analyse numérique*, 11(1), 13-25. <https://doi.org/10.1051/m2an/1977110100131>
- GERBER, F., & FURRER, R. (2019). `optimParallel` : An R Package Providing a Parallel Version of the L-BFGS-B Optimization Method. *The R Journal*, 11(1), 352-358. <https://doi.org/10.32614/RJ-2019-030>
- GILBERT, P., & VARADHAN, R. (2019). *numDeriv : Accurate Numerical Derivatives* [R package version 2016.8-1.1]. <https://CRAN.R-project.org/package=numDeriv>
- MATHUR, R. (2012). *An Analytical Approach to Computing Step Sizes for Finite-Difference Derivatives* [thèse de doct., University of Texas at Austin]. <https://hdl.handle.net/2152/ETD-UT-2012-05-5275>
- SAUER, T. (2017). *Numerical analysis* (3^e éd.). Pearson Education.
- STEPLEMAN, R. S., & WINARSKY, N. D. (1979). Adaptive numerical differentiation. *Mathematics of Computation*, 33(148), 1257-1264. <https://doi.org/10.1090/s0025-5718-1979-0537969-8>

Méthodes quantitatives en géographie : régression, ACP et classification sur des objets spatiaux

Hugo Périlleux*

Résumé

Le matériel pédagogique pour les travaux pratiques du cours "Méthodes quantitatives en géographie" destiné aux étudiant-es en géographie de l'Université Libre de Bruxelles a été publié sous forme d'un livre Bookdown. La particularité est d'appliquer ces méthodes classiques en statistiques à des objets spatiaux (points ou polygones : secteurs statistiques, communes, arrondissements, annonces immobilières, etc.). Le livre est divisé en trois chapitres, chacun dédié à une méthode statistique particulière : 1) les régressions, 2) l'analyse en composantes principales (ACP), 3) les classifications. Pour chaque chapitre, une démonstration pratique est suivie d'un exercice permettant d'appliquer les méthodes. Les exemples et les exercices s'appuient sur de véritables données choisies de façon à rendre les méthodes pertinentes et pédagogiques : des données électorales françaises, des données de la structure de la propriété des logements loués à Bruxelles ou encore des annonces immobilières issues de la plateforme Immoweb pour Bruxelles. Pour réaliser les analyses et les visualisations, les packages R suivants sont utilisés : sf, mapsf, modelr, jtools, ggstats, performance, huxtable, FactoMineR, factoextra, corrplot et Jutils. L'objectif de la présentation est de présenter ce livre ainsi que les exercices associés.

Lien vers le livre: https://bookdown.org/hugoperilleux/manuel_geo_quanti/

Mots-clefs : Matériel pédagogique - géographie quantitative - Régression - ACP - Classification

Nettoyer son code en un clin d'oeil avec flint

Etienne Bacher*

Résumé (max 300 mots)

Si vous êtes ici, c'est que vous écrivez du code R. Vous êtes peut-être expert depuis 10 ans, ou vous avez créé vos premiers scripts le semestre dernier lors d'un cours à la fac. Dans tous les cas, il est très probable que votre code ne suive pas absolument toutes les meilleures pratiques du premier coup. Vous utilisez `apply()` mais êtes-vous sûr que c'est la fonction la plus rapide pour ce que vous voulez faire ? Est-ce que `testthat::expect_equal()` est vraiment la fonction la plus appropriée dans ce contexte ? Dans cette présentation, je vais parler d'un package que j'ai créé il y a un peu moins d'un an : `flint`. L'objectif de ce package est de détecter et corriger automatiquement toute une liste de motifs qui ne sont pas optimaux, que ce soit en termes d'efficacité, de facilité de lecture, ou qu'il s'agisse simplement de bugs faciles à écrire mais durs à détecter.

Mots-clefs (3 à 5) : Bonnes pratiques - Package - Dette technique

Développement

Beaucoup de projets utilisant R durent des mois, voire des années. Que ce soient des analyses pour écrire des articles académiques, des pipelines pour traiter les données quotidiennes, ou des paquets R, la plupart de ces projets vont souffrir à un moment ou à un autre de la **dette technique**.

Une des principales composantes de cette dette est l'écriture de code qui ne suit pas les meilleures pratiques. Ces meilleures pratiques peuvent être liées à :

- la performance du code. Par exemple, il est préférable d'utiliser `rowSums(x)` plutôt que `apply(x, 1, sum)`;
- la non-utilisation de fonctions plus spécialisées. Par exemple, on peut remplacer `testthat::expect_equal(length(x), 1)` par `testthat::expect_length(x, 1)`;
- la lecture difficile du code. Par exemple, il est plus facile de lire `!all(x %in% letters)` plutôt que `any(!x %in% letters)`.

*Sans institut, etienne.bacher@protonmail.com

Il existe quelques outils pour détecter ces mauvaises pratiques, le plus connu étant `lintr`. Ce paquet contient plus d'une centaine de règles et est extrêmement utile, mais souffre de deux défauts. D'abord, sa performance peut laisser à désirer quand les projets contiennent des milliers de lignes réparties dans des dizaines de fichiers. Ensuite, `lintr` indique les mauvaises pratiques mais ne peut pas les corriger automatiquement, laissant donc aux humains le soin de faire cette tâche assez pénible.

Voyant ces limites dans `lintr`, j'ai donc créé un package pour détecter et automatiquement corriger certaines de ces mauvaises pratiques : `flint`. Dans les coulisses, `flint` utilise le paquet `ast-grep` écrit en Rust, ce qui le rend bien plus rapide que `lintr`. `flint` est fourni avec environ 50 règles présentes dans `lintr`, et il est assez facilement possible d'ajouter ses propres règles.

Dans la présentation, je ferai une démonstration de `flint` et je montrerai comment l'appliquer automatiquement aux projets sur Github en utilisant Github Actions. En plus d'appliquer des règles prédéfinies, `flint` permet également aux utilisateurs d'écrire leurs propres règles. Je montrerai par exemple comment écrire une règle pour remplacer les vieilles fonctions de `dplyr` par les plus récentes.

OLAF, la reine des Apps Shiny !

Antoine Bichat^{*}, Anne-Sophie Pulcrano[†], [Sophie Courtade-Gaïani](#)[§]

Résumé

OLAF (Oncopanel Leukemia Analysis Facilitator) est une application Shiny développée pour faciliter et accélérer l'analyse des données pharmacologiques. Une campagne de criblage pharmacologique a été mise en œuvre pour évaluer le profil de sensibilité de plus de 200 traitements, administrés en agents simples ou en combinaison, sur un panel de 36 lignées de leucémie et 2 lignées contrôles. Cette application interactive et intuitive permet aux chercheurs d'analyser et d'interpréter les résultats pharmacologiques de manière efficace grâce à des visualisations appropriées. OLAF a pour vocation la comparaison des profils de réponse aux produits étudiés à travers des visualisations de type waterfall plots, boxplots, barplots et heatmaps, offrant une meilleure compréhension des effets des traitements sur les différentes lignées cellulaires étudiées.

OLAF constitue un outil puissant pour l'analyse et l'interprétation des données pharmacologiques sur les lignées AML. Il améliore l'efficacité de la recherche en offrant des visualisations claires pour réaliser des analyses approfondies, contribuant à une meilleure compréhension des mécanismes de réponse aux traitements dans la leucémie myéloïde aiguë.

Dans cette présentation, je vais présenter l'application et introduire plusieurs concepts essentiels du développement d'OLAF : le data management, le développement de l'appli shiny selon de standard Golem, la sélection des conditions à analyser, et des aspects de visualisation de données de type heatmap et waterfall plots.

Mots-clefs (3 à 5) : Shiny, heatmap, waterfall plot, analyse de données de pharmacologie

*Institut de Recherches Internationales Servier (IRIS), 22 route 128 – 91190 Gif sur Yvette - FRANCE : antoine.bichat@servier.com

† Keyrus Life Science, 5 rue Jacques Daguerre – 44300 Nantes - France , anne-sophie.pulcrano@keyrus.com

§ Institut de Recherches Servier (IdRS), 22 route 128 – 91190 Gif sur Yvette - FRANCE : sophie.courtade-gaiani@servier.com

Outils et conseils pour la maintenance partagée de package R

Hugo Gruson 1*

Résumé (max 300 mots)

Les règles du CRAN imposent un mainteneur unique clairement identifié pour chaque package. Mais, en pratique, il arrive que plusieurs personnes assurent la maintenance d'un package. Cela peut soit être temporaire lors de la transition entre deux mainteneurs, soit être permanent pour partager la charge de travail. Cette présentation relate mon expérience de maintenance partagée de package R, souligne des difficultés rencontrées et propose des solutions pour les surmonter. Ces solutions peuvent être de nature organisationnelle (archivage des discussions, documentation pour les développeurs) ou technique (par exemple, un tableau de bord pour la maintenance du package). En particulier, je m'appuie dans cette présentation sur l'exemple du package R pavo pour l'analyse de la couleur en biologie évolutive, que je co-maintiens avec Thomas White depuis 2018, l'exemple du package R fundiversity, que je co-maintiens avec Matthias Grenié depuis 2021, et sur l'exemple de la communauté de développement Epiverse.

Mots-clés (3 à 5) : Package R, maintenance, communauté, CRAN

Développement

Règles du CRAN et motivations

Les règles du CRAN imposent un mainteneur unique pour chaque package, indiqué dans le champ `Maintainer` ou le rôle `"cre"` dans le champ `Authors@R` du fichier `DESCRIPTION`. Il s'agit du seul rôle pour lequel un email est obligatoire. En particulier, il est contre les règles du CRAN de mettre une adresse email partagée pour le mainteneur.

Comment gérer le cas où plusieurs personnes assurent la maintenance d'un package ? En effet, ce cas de figure peut se présenter pour plusieurs raisons :

- transition entre deux mainteneurs. Dans ce cas, il est souvent désirable d'avoir une période durant laquelle les deux mainteneurs travaillent ensemble sur le projet. Cela permet au nouveau mainteneur de se familiariser avec le projet, tant en ayant l'occasion de poser des questions à l'ancien mainteneur.
- co-existence de mainteneurs multiples de façon pérenne. Cela peut être une bonne stratégie pour réduire la charge de travail, mais également dans la situation où les deux mainteneurs ont des champs d'expertise très différents. Il arrive par exemple qu'un mainteneur soit expert du sujet scientifique du package et l'autre mainteneur soit expert en programmation R.

Processus de développement et communication

Il est utile de définir des processus de développement et de communication clairs dès le début du projet pour éviter la perte d'informations. Les bonnes pratiques de développement habituelles, en particulier celles pour la durabilité du projet, restent valides et ne sont pas répétées ici. Les conseils suivants se concentrent sur des situations spécifiques à la maintenance partagée.

En particulier, il est utile d'archiver tous les emails échangés entre les mainteneurs et les utilisateurs, ou les administrateurs du CRAN. Dans des projets existants, nous avons constaté que l'existence d'un système de tickets (par exemple, sur GitHub) n'est pas suffisant par s'assurer de la centraliser de l'information. En effet, il reste courant que les utilisateurs contactent directement un des auteurs par email, et il est important de s'assurer que ces informations ne soient pas perdues si cet auteur quitte l'équipe de développement. De

*data.org, hugo.gruson@normalesup.org

la même manière, je recommande de s'assurer que les conversations au sein de l'équipe de développement sont préservées de manière fiable. Par exemple, dans le cas du package pavo, de nombreuses conversations informelles de l'équipe de développement ont eu lieu sur slack au cours des 10 dernières années et sont désormais inaccessibles suite à l'évolution de la politique commerciale de slack. Ces conversations contenaient de précieuses informations en terme d'idées pour le futur, ou de solutions à des problèmes techniques complexes.

Enfin, en cas d'utilisation de GitHub, il est utile d'intégrer tous les nouveaux changements via une Pull Request (PR) plutôt que de pousser directement sur la branche principale, même s'il n'y a pas de relecture du code. Cela permet au moins à l'ensemble de l'équipe de développement d'être informée des changements en cours, qui peuvent autrement passer inaperçus.

Documentation à l'usage des développeurs

De manière plus formelle, nous avons instauré l'utilisation d'une vignette spéciale à l'attention des futurs mainteneurs dans les packages du projet collaboratif Epiverse, guidé par notre expérience de maintenance partagée de fundiversity. Ces vignettes viennent s'ajouter au commentaire de code et enregistrent les décisions de design importantes, les choix de dépendances, les problèmes connus, et les solutions apportées.

Par exemple, cela peut apporter un éclairage spécifique à l'addition d'une dépendance justifié par la nécessité d'une performance accrue, ou présenter l'utilisation d'un schéma spécifique de nommage des fonctions.

Outils pour la maintenance partagée

Enfin, j'ai développé des outils spécifiques pour faciliter la maintenance partagée de package R, et réduire les points de défaillance unique. Par exemple, dans pavo, nous avons un automate qui ouvre une issue sur GitHub si le package est susceptible d'être archivé sur le CRAN. Cela supprime la nécessité pour le mainteneur officiel de communiquer cette information au reste de l'équipe, ce qui peut s'avérer impossible en cas d'absence, et au final mettre en péril la survie du package sur le CRAN.

Sur le même modèle, j'ai développé un tableau de bord pour Epiverse, qui permet à toute la communauté d'avoir une vue d'ensemble du statut de tous les packages sur le CRAN, et qui fournit un système de notification automatique en cas de risque d'archivage.

PATAPI : un outil de planification des pesées de bovins développé avec R

M. Legris *

A. Delpéuch †

Résumé

Le projet PATAPI s'inscrit dans le cadre de l'amélioration de la gestion des pesées des bovins dans les élevages français, un domaine clé pour la performance génétique et la gestion optimale des troupeaux. Ce projet vise à fournir un outil performant de planification des pesées, en utilisant des algorithmes d'optimisation développés principalement en langage R. Le package Rcpp a été utilisé pour intégrer du code C++, permettant d'accélérer les calculs, et le package GenSA a été utilisé pour trouver des solutions optimales grâce à la méthode du recuit simulé. Une application Shiny a été créée pour aider au développement de l'algorithme en facilitant la visualisation des résultats et en permettant des ajustements dynamiques des paramètres. Le moteur de calcul a été encapsulé dans un package R et il est en cours d'APIfication. Il permettra ainsi de générer des prévisions sur les dates de pesée idéales, offrant un outil efficace et réactif aux techniciens et éleveurs, dans le but de simplifier la planification des pesées et d'améliorer la gestion des élevages.

Mots-clés : Prévission - Optimisation - Package - Shiny

Développement

Contexte et objectif

En France, dans le cadre du Contrôle des Performances Viande (CPV), environ 400 000 poids à âge type (i.e. poids à 120 jours ou à 210 jours de vie de l'animal) sont calculés chaque année pour un peu plus d'un million de pesées, dont 80 % sont réalisées par des techniciens. Ces pesées sont utilisées pour conseiller l'éleveur dans sa conduite et alimentent l'évaluation génétique pour le calcul des index de croissance et de valeur laitière. Dans ce contexte, l'objectif du projet PATAPI est de développer un moteur de calcul permettant de planifier efficacement les visites des techniciens en élevage pour effectuer les pesées.

Les fonctionnalités de l'outil

Un premier travail préalable au projet PATAPI (Delpéuch (2021)) avait permis de définir des niveaux de fiabilité des estimations des poids à 120 jours et 210 jours en fonction de la position des dates de pesées dans la carrière de l'animal (interpolation/extrapolation, proximité des 120 jours ou 210 jours de l'animal...). L'outil développé dans le cadre de PATAPI devait proposer automatiquement des dates de pesées en élevage qui permettent d'estimer les poids des animaux à 120 et 210 jours avec un niveau de fiabilité suffisamment élevé. L'enjeu est de trouver un équilibre entre nombre de passage des techniciens en élevage et fiabilité de l'estimation des poids. Le moteur de calcul se base sur les dates de naissance des animaux (estimées ou réelles) et sur un ensemble de paramètres soit de performances souhaitées par l'éleveur/le technicien (ex : maximisation du poids à 120 jours, 210 jours ou les deux ; classes minimales de fiabilité...), soit liés aux contraintes de l'éleveur (ex : ne pas tenir compte d'animaux atypiques, définir des plages de dates interdites durant laquelle l'éleveur n'est pas disposé à peser, ne pas peser les animaux au-delà de l'âge moyen de sortie des animaux, fixer une date particulière de pesée...). L'outil final devait être à la fois un outil prévisionnel, i.e. qui permet d'avoir une simulation prévisionnelle des dates de passages en début de campagne, mais également un outil actualisé au fil de l'eau, i.e. qui permet d'ajuster les planifications en fonction des naissances réelles de l'élevage et en tenant compte des dates de pesées déjà réalisées.

*Institut de l'élevage, maxime.legris@idele.fr

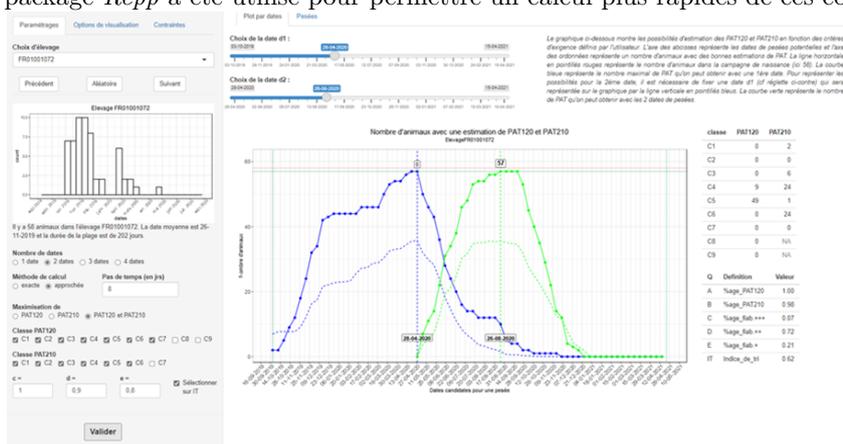
†Institut de l'élevage, arnaud.delpéuch@idele.fr

La fonction de coût et son optimisation via le package GenSA

Le travail dans le projet PATAPI consistait donc en la définition d'une fonction de coût, en son optimisation et en son intégration dans un dans un moteur de calcul robuste et unifié qui prend la forme d'un package R. Une premier travail a onsisté à réunir des experts du domaine pour définir une fonction de coût permettant de départager les solutions. Cette fonction de coût combine la proportion de poids bien estimés, le niveau de fiabilité de ces PAT et le nombre de visites nécessaires. Cette fonction de coût fournit un indice exprimé entre 0 (faible intérêt) et 1 (fort intérêt). La fonction de coût ainsi définie est une fonction avec de nombreux paramètres et contraintes qui définissent un espace de solution à plus de 3 dimensions, difficile à représenter graphiquement, et avec potentiellement de nombreux maxima locaux. Du fait de la complexité de cette fonction de coût, les techniques d'optimisation classique (descente de gradient, méthode de Newton...) ont été écartées au profit d'une méthode stochastique : le recuit simulé généralisé. Le package *GenSA* a été utilisé pour sa performance (Kirkpatrick, Gelatt, and Vecchi (1983), Xiang et al. (2013)).

L'application Shiny comme un démonstrateur

Tout au long du développement du moteur de calcul, l'utilisation d'une application *Shiny* a été une aide au développement de l'algorithme en facilitant la visualisation des résultats et en permettant des ajustements dynamiques des paramètres. Elle a permis de faire préciser le besoin des différents utilisateurs finaux de l'outil et a orienté le développement du moteur de calcul final vers deux modes de calcul : un mode *simplifié* avec uniquement un ensemble de dates de pesées optimales et un mode *complet* qui intègre en plus des courbes 2D. Le package *Rcpp* a été utilisé pour permettre un calcul plus rapides de ces courbes.



Un package R qui optimise les pesées

Le package R contenant le moteur de calcul développé dans PATAPI permet au final de proposer un ensemble de dates optimales pour les pesées et des résultats plus précis pour un pilotage des visites en élevage. Sur un échantillon d'élevages aux caractéristiques diverses (nombre d'animaux, étendue de la période de naissance), le moteur de calcul permet de réduire le nombre de visites et/ou d'améliorer le positionnement des dates de pesées. Les temps de calcul en mode simplifié sont raisonnables (inférieurs à 5 secondes pour la majorité des élevages) mais largement perfectibles et plusieurs pistes sont envisagées pour une deuxième version.

Références

- Delpuech, A. 2021. "Prise En Compte de La Qualité Des Données de Naissance Dans Le Calcul Des PAT Et Des Index IBOVAL." *Institut de l'Élevage*. <https://idele.fr/detail-article/prise-en-compte-de-la-qualite-des-donnees-de-naissance-dans-le-calcul-des-pat-et-des-index-iboval>.
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220 (4598): 671–80. <https://doi.org/10.1126/science.220.4598.671>.
- Xiang, Yang, Sylvain Gubian, Brian Suomela, and Julia Hoeng. 2013. "Generalized Simulated Annealing for Global Optimization: The GenSA Package." *The R Journal* 5 (1): 13–28.

Pages d'aide simplifiées et en français dans R grâce à l'IA

Guyliann Engels 1*

Philippe Grosjean 1†

Résumé (max 300 mots)

L'IA est déjà très présente dans R, notamment grâce à des package comme `{ellmer}`, `{tidyllm}` ou encore `{rollama}` qui proposent des connexions via API avec divers LLM (GPT-4o, Llama 3, Claude 3.5, Gemini 2.0...). Il existe également des packages qui facilitent l'interaction avec ces LLM et qui permettent d'accéder à des add-ins Shiny pour réviser et corriger du code, tels que `{chores}`, `{gander}` et `{shinychat}`. On peut aussi mentionner GitHub Copilot, pour lequel une intégration existe dans RStudio et Positron.

Dans R, une grande partie des pages d'aide est très technique pour une première approche d'une fonction. De plus, avec les fonctions génériques, il n'est pas toujours facile de trouver la page d'aide correspondante à la méthode utilisée. Cette complexité freine les nouveaux utilisateurs dans l'utilisation des pages d'aide. En outre, la barrière linguistique constitue un obstacle supplémentaire pour les utilisateurs non anglophones.

Dans ce contexte, nous proposons `{helpai}` qui permet, entre autres, de générer des pages d'aide simplifiées, éventuellement en français, grâce à l'IA choisie par l'utilisateur. Un addin Shiny permet de spécifier la fonction d'intérêt pour laquelle on souhaite une page d'aide simplifiée et la méthode, s'il s'agit d'une fonction générique. La page ainsi générée fournit une description non ou peu technique de la fonction en deux ou trois paragraphes, documente les principaux arguments et présente un exemple généralement exécutable. Enfin, cette page pointe vers la page d'aide "officielle" de ladite fonction suggérant, donc, à l'utilisateur de continuer à découvrir la fonction ciblée au travers de la documentation usuelle s'il le juge nécessaire.

Mots-clefs (3 à 5) : IA - Aide R - Enseignement - Shiny

Développement

Les outils d'intelligence artificielle (IA) sont de plus en plus répandus et simples d'accès. De très nombreux grands modèles de langage (LLM) existent. On peut trouver des versions open source ou propriétaires, ainsi que des modèles généralistes ou spécialisés. Le modèle le plus connu est probablement GPT-4. De nombreux packages R comme `{ellmer}` (Wickham, Cheng, and Jacobs 2025), `{tidyllm}` (Brüll 2024) et `{rollama}` (Gruber and Weber 2024) ont été développés pour permettre d'interagir avec ces LLM via des API. D'autres packages tels que `{gander}` (Couch 2025b) et `{chores}` (Couch 2025a) s'appuient sur les premiers pour implémenter des addins Shiny qui permettent d'utiliser les LLM directement dans les scripts R, les documents Quarto ou R Markdown.

Les pages d'aide dans R sont rédigées par les développeurs de packages qui détaillent minutieusement leurs travaux et proposent souvent des explications techniques approfondies sur les fonctions. Cette documentation riche peut cependant être trop technique pour une première approche. Elle peut aussi rebuter potentiellement par un nombre important de sections, d'arguments et d'exemples fournis. Les pages d'aide des fonctions `library()` ou `install.packages()` sont des exemples caractéristiques. De plus, il est parfois difficile de trouver la bonne page d'aide lorsqu'on utilise des fonctions génériques comme `summary()` ou `plot()` car il faut rechercher la page relative à la méthode que R va sélectionner dans le code, ce qui rajoute une difficulté supplémentaire pour l'utilisateur de R débutant. Enfin, l'anglais constitue un obstacle supplémentaire pour certains utilisateurs non anglophones. Ces éléments poussent souvent les utilisateurs à chercher des informations sur les fonctions en dehors des pages d'aide, alors qu'il s'agit pourtant de l'outil dédié dans R.

Dans ce contexte, nous proposons le package `{helpai}`, qui permet notamment de générer des pages d'aide simplifiées et adaptées à la langue de l'utilisateur grâce à l'IA. La description est simplifiée, ne présentant

*Service d'écologie numérique, Institut Complexys & Infotech, Université de Mons, Belgique, guyliann.engels@umons.ac.be

†Service d'écologie numérique, Institut Complexys & Infotech, Université de Mons, Belgique, philippe.grosjean@umons.ac.be

que les arguments essentiels, ainsi qu'un exemple concret et généralement exécutable, est proposé. Cette page d'aide simplifiée se présente de manière similaire et s'affiche au même endroit que les pages HTML de R, par exemple, dans l'onglet "Help" de RStudio.

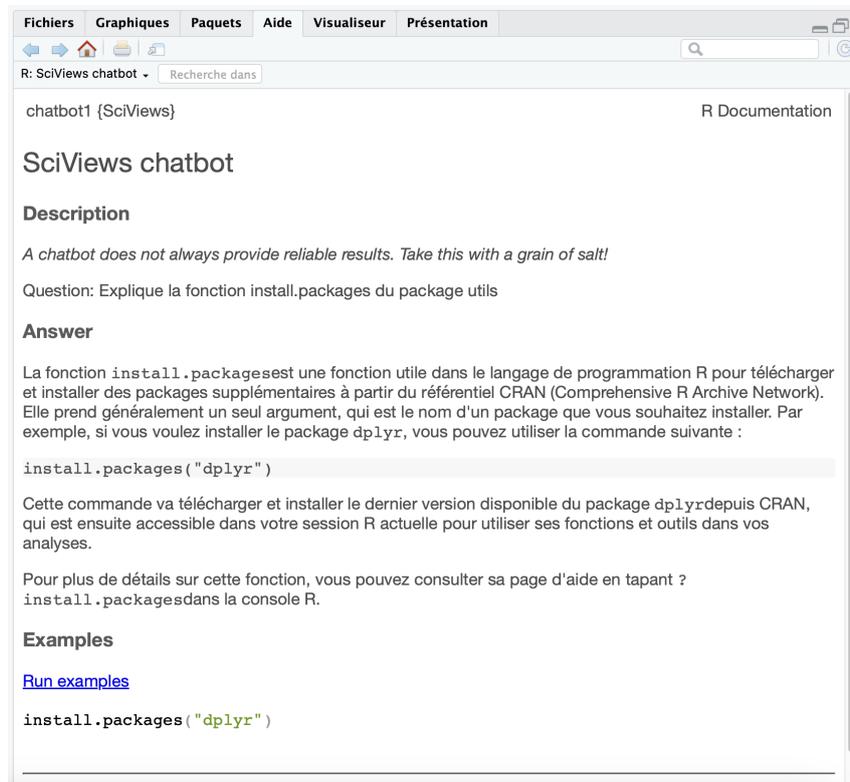


Figure 1: Exemple d'une page d'aide simplifiée en français pour la fonction `install.packages()`

En pratique, l'utilisateur sélectionne la fonction qui l'intéresse dans un script R, un document Quarto ou R Markdown et lance l'addin dédié dans RStudio. Il peut aussi accéder à cette fonctionnalité par une instruction R (dans un terminal, ou un autre IDE). Si la fonction est générique, l'ensemble des méthodes disponibles est présenté dans une liste de choix. Cette première étape permet de sélectionner la bonne fonction puis de générer la page d'aide simplifiée relative au contexte correct.

Ce package est actuellement en cours de test avec une bonne centaine d'étudiants qui suivent des cours de sciences de données biologiques à l'Université de Mons, en Belgique. Le modèle Open Source `codestral-22B`, spécialisé pour les tâches associées à la programmation et développé par la société Mistral, est employé sur des serveurs dédiés et connectés au package `{helpai}`.

Références

- Brüll, Eduard. 2024. "Tidyllm: Tidy Integration of Large Language Models." <https://CRAN.R-project.org/package=tidyllm>.
- Couch, Simon. 2025a. "Chores: A Collection of Large Language Model Assistants." <https://CRAN.R-project.org/package=chores>.
- . 2025b. "Gander: High Performance, Low Friction Large Language Model Chat." <https://CRAN.R-project.org/package=gander>.
- Gruber, Johannes B., and Maximilian Weber. 2024. "Rollama: Communicate with 'Ollama' to Run Large Language Models Locally." <https://CRAN.R-project.org/package=rollama>.
- Wickham, Hadley, Joe Cheng, and Aaron Jacobs. 2025. "Ellmer: Chat with Large Language Models." <https://CRAN.R-project.org/package=ellmer>.

PerRegMod : un nouveau package R pour le test et l'estimation des modèles de régression à coefficients périodiques.

Regui slimane *

Akharif Abdelhadi †

Mellouk Amal ‡

Résumé

Cette présentation introduit PerRegMod, un nouveau package R spécialement conçu pour l'ajustement de modèles de régression linéaire à coefficients périodiques. PerRegMod inclut un test pour détecter la périodicité des paramètres dans les modèles de régression et fournit des estimateurs pour estimer efficacement les coefficients périodiques.

Mots-clefs (3 à 5) : Modèle de régression à coefficients périodiques – test pseudo-gaussien – Estimateur des moindres carrés – R

Développement

Le modèle de régression à coefficients périodiques proposé par [Regui et al., 2024c] et [Regui et al., 2024a] surpasse à la fois le modèle de régression classique et le modèle de régression à coefficients aléatoires proposé par [Fihri et al., 2016] et [Fihri et al., 2020] pour les données périodiques. [Regui et al., 2024c] a introduit des tests paramétriques et non paramétriques pour détecter la périodicité des coefficients dans les modèles de régression simples, tandis que [Regui et al., 2024a] a développé des estimateurs des moindres carrés et adaptatifs pour les modèles de régression périodique.

Dans cette présentation, nous introduisons PerRegMod, un nouveau package R conçu pour détecter la périodicité et estimer les paramètres des modèles de régression périodique. Le package R PerRegMod, disponible sur CRAN à l'adresse <https://CRAN.R-project.org/package=PerRegMod> ou voir [Regui et al., 2024b], propose un test pseudo-gaussien permettant de détecter la périodicité des paramètres dans les modèles de régression multiple. De plus, il inclut deux estimateurs : l'estimateur des moindres carrés (LSE) et l'estimateur adaptatif (AE), conçus pour estimer les coefficients de modèle de régression périodique.

Références

Mohamed Fihri, Amal Mellouk, and Abdelhadi Akharif. Rank and signed-rank tests for random coefficient regression model. *Journal of Statistics Applications and Probability*, 5 :233–247, 2016.

Mohamed Fihri, Abdelhadi Akharif, Amal Mellouk, and Marc Hallin. Efficient pseudo-gaussian and rank-based detection of random regression coefficients. *Journal of Nonparametric Statistics*, 32(2) : 367–402, 2020.

Slimane Regui, Abdelhadi Akharif, and Amal Mellouk. Adaptive estimation of periodic regression model in short panel data. *Electronic Journal of Applied Statistical Analysis*, 17(3), 2024a. ISSN 2070-5948.

Slimane Regui, Abdelhadi Akharif, and Amal Mellouk. *PerRegMod : Fitting Periodic Coefficients Linear Regression Models*, 2024b. R package version 4.4.3.

*LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco, slimane.regui@etu.uae.ac.ma

†LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco, a.akharif@uae.ac.ma

‡Centre Régional des Métiers de l'Éducation et de la Formation, Tangier, Morocco, melloukamal@hotmail.com

Slimane Regui, Abdelhadi Akharif, and Amal Mellouk. Locally optimal tests against periodic linear regression in short panels. *Communications in Statistics - Simulation and Computation*, 0(0) :1–15, 2024c.

Personnalisation de documents Quarto avec `_brand.yml`

Christophe Dervieux*

Résumé (max 300 mots)

Dans le domaine de la recherche et de l'analyse de données, la présentation cohérente des résultats est cruciale pour transmettre efficacement des informations. Ce talk vous plongera dans la fonctionnalité innovante `_brand.yml` de Quarto, qui permet de personnaliser facilement l'apparence de vos documents tout en respectant les lignes directrices de votre organisation.

Que vous soyez utilisateur confirmé de Quarto ou que vous découvriez cet outil pour la première fois, cette session mettra en lumière les avantages pratiques d'utiliser `_brand.yml`. À travers des exemples concrets et des démonstrations, nous explorerons comment un simple fichier YAML peut intégrer logos, couleurs et typographies, facilitant ainsi la création de rapports, tableaux de bord et présentations attrayants et professionnels.

L'objectif est de montrer comment cette fonctionnalité peut transformer vos projets en rendant vos documents plus cohérents avec l'identité visuelle de votre organisation, tout en simplifiant le processus de personnalisation. Vous verrez comment `_brand.yml` permet non seulement d'harmoniser vos travaux existants, mais aussi de pouvoir réutiliser votre configuration pour d'autres types de project, `_brand.yml` étant supporté par d'autres outils (ex. Shiny).

En fin de compte, cette présentation vise à susciter votre curiosité pour essayer cette fonctionnalité dans vos propres projets et à découvrir comment Quarto peut améliorer votre flux de travail.

Mots-clefs (3 à 5) : Quarto - Personnalisation - Visualisation - Rapports - Reproductibilité

Développement

Introduction à Quarto et à `_brand.yml`

Cette présentation débutera par une introduction à Quarto (Allaire et al. (2024)), un outil open source qui facilite la création de documents dynamiques. La fonctionnalité `_brand.yml` (Posit (2024)) sera ensuite présentée de manière générale.

Comment utiliser `_brand.yml` avec Quarto

Une démonstration sera réalisée pour créer un fichier `_brand.yml` pour utiliser avec Quarto (Quarto (2024)). Les points suivants seront couverts :

- Les configurations possibles (logos, couleurs, typographies...)
- L'impact de ces personnalisations sur l'apparence des documents Quarto,
- La réutilisation de cette configuration pour différents formats et projets, soulignant la flexibilité offerte par cette approche.

Des exemples concrets montreront comment d'autres chercheurs et analystes utilisent `_brand.yml` pour rendre leurs documents plus cohérents et attrayants.

*Posit PBC, cderv@posit.co

Comment bénéficier au maximum de `_brand.yml`

Cette section explorera comment maximiser l'utilisation de `_brand.yml`. On discutera des meilleures pratiques pour structurer le fichier, ainsi que des options disponibles pour personnaliser les éléments d'identité visuelle. Il s'agira de voir comment intégrer dans différents projets pour bien réutiliser son identité visuelle.

Conclusion et perspectives

La présentation sera conclue par un récapitulatif des points essentiels, mettant en lumière l'importance de `_brand.yml` pour personnaliser de manière cohérente divers projets. Des pistes seront également suggérées pour explorer l'utilisation de cette fonctionnalité avec d'autres outils, telle que Shiny, favorisant une approche intégrée dans la communication via des publications.

Références

- Allaire, J. J., Charles Teague, Carlos Scheidegger, Yihui Xie, Christophe Dervieux, and Gordon Woodhull. 2024. "Quarto." <https://doi.org/10.5281/zenodo.5960048>.
- Posit. 2024. "Documentation for brand.yml Specification." 2024. <https://posit-dev.github.io/brand-yml/>.
- Quarto. 2024. "Branding with `_brand.yml` and Quarto." 2024. <https://quarto.org/docs/authoring/brand.html>.

PhaochR : géocoder des adresses belges avec R

Hugo Périlleux* Joël Gires†

Résumé

Nous présentons PhaochR : un géocodeur pour la Belgique publié sous forme de package R et d'une interface Shiny pour permettre son utilisation par les non-utilisateurs de R. Son principe est de produire de manière simple et rapide, à partir d'une liste d'adresses, une série d'informations nécessaires pour l'analyse spatiale : les coordonnées X-Y mais également d'autres informations utiles, comme le secteur statistique (la plus petite unité géographique belge). Le programme fonctionne sur base de BeSTAddress, des données publiques d'adresses quasi-exhaustives pour la Belgique, compilées à partir des listes régionales (Région de Bruxelles-Capitale, Région flamande et Région wallonne). Par ailleurs, le géocodage est réalisé entièrement en local, permettant une confidentialité maximale dans le cas de données d'adresses sensibles. PhaochR réalise également la cartographie des adresses localisées sur la base de données spatiales intégrées.

Le package a déjà été présenté lors des 9èmes rencontres R à Avignon en 2023. Il nous apparaît néanmoins pertinent de le représenter cette année à Mons pour expliquer la logique générale et les éléments d'attention à avoir en tête lorsqu'on l'utilise puisque le package fonctionne pour des adresses belges. Ce sera également l'occasion de présenter les mises-à-jour importantes qui ont été réalisées depuis lors.

Lien vers PhaochR: <https://phacochr.github.io/phacochrdev/>

Mots-clefs : Package – Géocodage – SIG

* Université Libre de Bruxelles, Géographie Appliquée et Géoinformation, hugo.perilleux@ulb.be

† Observatoire de la Santé et du Social de Bruxelles-Capitale, jgires@ccc.brussels

R and CRAN on Ubuntu With r2u

Dirk Eddelbuettel*

Résumé

We describe 'r2u', a repository for statistics, data science and machine learning that provides all CRAN packages as Ubuntu binaries suitable for fully automated analysis, processing, training as well as continuous integration development and testing. We highlight its key advantages for robustness, as well as speed and ease of use, with examples supporting the summarizing statement "r2u: Fast. Easy, Reliable. Pick All Three".

Mots-clefs : Repository – CRAN – Ubuntu – R package – Installation

* Department of Statistics, University of Illinois Urbana-Champaign, USA, edd@debian.org

R au service du séquençage microbien : analyse et visualisation des résultats

Alice Delacuvellerie, 1* Ruddy Wattiez 2†

* *Proteomics and Microbiology department, University of Mons, 20 place du parc, 7000 Mons, Belgium* (alice.delacuvellerie@umons.ac.be, ruddy.wattiez@umons.ac.be)

L'analyse des communautés bactériennes issues de séquençage 16S, 18S ou métagénomique est un enjeu clé pour comprendre ces écosystèmes. Qu'il s'agisse de l'étude des biofilms se développant sur les plastiques ou de l'exploration des microbiotes associés aux racines de tomates, R est un outil clé pour traiter, analyser et visualiser, en partie, ces données complexes.

Grâce à un large éventail de fonctions et de packages, R permet d'explorer la diversité bactérienne. En effet, la raréfaction des données, essentielle pour normaliser la profondeur de séquençage, s'effectue aisément avec des packages comme *vegan*. L'analyse statistique des différences de communautés en fonction des conditions expérimentales s'appuie sur des méthodes robustes telles que PERMANOVA (*vegan*), tandis que la représentation graphique des résultats est facilitée par la construction de heatmaps, boxplots, de diagrammes de Venn et par l'utilisation d'approches de simulation Monte-Carlo. En complément, l'intégration de tests statistiques classiques assure une évaluation rigoureuse des variations entre échantillons et conditions expérimentales.

En associant R à d'autres outils bioinformatiques, il devient possible d'obtenir une analyse approfondie et reproductible des données de séquençage bactérien, offrant ainsi une meilleure compréhension des communautés microbiennes dans divers environnements. Cette présentation mettra en avant des cas concrets d'analyse et de visualisation de données microbiologiques sous R, illustrant son rôle clé dans l'interprétation des résultats.

Mots-clefs (3 à 5) : Séquençage, biofilm, bactérie, ADN, analyses bioinformatiques

* *Proteomics and Microbiology department, University of Mons, 20 place du parc, 7000 Mons, Belgium* (alice.delacuvellerie@umons.ac.be)

† *Proteomics and Microbiology department, University of Mons, 20 place du parc, 7000 Mons, Belgium* (ruddy.wattiez@umons.ac.be)

R et le déclin des pollinisateurs : des outils statistiques au service de l'écologie

Maryse Vanderplanck*

Résumé

Dans le contexte actuel du déclin des pollinisateurs, les efforts de recherche se concentrent principalement sur les pesticides, les pathogènes et les changements climatiques. La réduction de la qualité de l'air est également un problème majeur, affectant la communication chimique entre plantes et pollinisateurs, et menaçant ainsi les réseaux d'interactions.

Plus particulièrement, les concentrations accrues en ozone, corollaire du réchauffement climatique, pourraient altérer les capacités cognitives des abeilles. Ces capacités cognitives sont essentielles pour que les pollinisateurs se nourrissent et se souviennent de l'emplacement des ressources florales les plus appropriées.

Une partie de mes recherches consiste à étudier l'impact des épisodes d'ozone sur la capacité des abeilles à détecter les signaux floraux et à les mémoriser. Dans le cadre de cet axe de recherche, j'établis divers modèles mixtes et matrices de contraste pour lesquels l'utilisation de R est particulièrement bien adaptée et très répandue.

Je présenterai dans un premier temps les analyses électrophysiologiques que j'effectue. A partir des électroantennogrammes, un code R établi par l'un des ingénieurs de mon laboratoire me permet d'extraire de manière automatisée les intensités de détection de différents composés volatils par des abeilles exposées ou non à différents épisodes d'ozone. L'analyse de ces données passent majoritairement par l'établissement de modèles linéaires.

Dans un second temps, je présenterai les tests d'apprentissages et de mémorisation de signaux colorés associés à des récompenses sucrées, toujours par des abeilles exposées ou non à différents épisodes d'ozone. A partir de données binaires, j'utilise cette fois des modèles binomiaux pour mes analyses ainsi que différents outils R afin d'établir des courbes d'apprentissage. J'aborderai les difficultés parfois rencontrées pour analyser ces jeux de données (surdispersion des données, occurrence de zéros) et insisterai sur l'importance de la prise en compte des analyses statistiques dès la conception des plans expérimentaux.

Mots-clefs : Insectes pollinisateurs – Changements climatiques – Électrophysiologie

* Centre d'Écologie Fonctionnelle & Évolutive, Université de Montpellier, CNRS,EPHE, IRD,
maryse.vanderplanck@cefe.cnrs.fr

R pour l'analyse de données vectorielles en analyse de la parole

Olivier Crouzet*

Résumé

En sciences humaines, l'analyse de la parole fait appel à de nombreuses données acoustiques dont le traitement peut être un véritable challenge. Notre exposé abordera l'intérêt de R pour développer des solutions de traitements de données phonétiques/acoustiques.

Nous illustrerons l'exploitation de données phonétiques vectorielles de type trajectoires formantiques (soit extraites du logiciel Praat, soit directement issues d'une requête R via une base de données Emu Speech avec le package EmuR (Winkelmann, Harrington & Jänsch, 2017) pour estimer des mesures phonétiques caractéristiques (vitesse de changement, locus formantique, équation de locus, target-locus scaling) et évoquerons aussi pour les non spécialistes la variabilité articulatoire et acoustique des signaux de parole et la nature des données de production avec lesquelles on travaille en phonétique.

Mots-clefs : Parole – Phonétique – Locus

Référence

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, 45, 392–410. <https://doi.org/10.1016/j.csl.2017.01.002>

* Enseignant-chercheur - Nantes Université - Département Sciences du Langage - UMR 6310
Laboratoire de Linguistique de Nantes - UFR Lettres et Langues, Olivier.Crouzet@univ-nantes.fr

R vous ment!

Antoine Fabri*

Résumé

Il peut être difficile de comprendre les objets dans R. `print()` et `str()` ne disent pas tout, et `dput()` a aussi ses problèmes: il est notamment incomplet, parfois inexact, généralement peu lisible. Ceci est problématique lorsqu'on souhaite explorer, comparer des objets, ou encore les capturer sous forme de code. Dans cette présentation je vous montrerai comment et pourquoi R dissimule certaines informations sur vos objets et je vous proposerai une solution sous la forme du package `{constructive}`.

Mots-clés : R - Objets - Package - Reproductibilité

Développement

Cette présentation a pour but de démystifier les objets R et de fournir un outil de débogage fiable. Elle se veut instructive, amusante et technique tout en restant accessible, afin que des utilisateurs de tout niveau puissent y apprendre quelque chose de nouveau.

La majeure partie de la présentation traitera des objets R et de leur représentation en général, sans évoquer `{constructive}`. On y montrera d'abord comment 7 objets différents peuvent imprimer "10" puis comment les homoglyphes (caractères différents s'imprimant identiquement) et les caractères "identiques" mais encodés différemment peuvent nous tromper. On y verra aussi comment R supporte partiellement le zéro négatif.

On présentera ensuite les outils intégrés à R pour imprimer ou inspecter les objets, et en particulier la fonction `dput()` dont on montrera les limitations.

Enfin, nous montrerons comment le package `{constructive}` répond à ces défis en l'appliquant aux exemples problématiques évoqués précédemment.

Cette présentation sera une traduction et révision de ce que j'ai eu le plaisir de présenter au useR! 2024 à Salzburg sous le nom de: "`{constructive}` : a nicer `dput()` using idiomatic constructors"

Le package `constructive` est en développement continu. `cynkra`, mon employeur à Zurich, soutient activement l'open source et encourage le développement de packages comme `{constructive}`.

Je partage personnellement mon temps entre Bruxelles et Zurich, suis belge, diplômé à l'ULB et aurai un grand plaisir à assister à une telle conférence dans mon pays.

*`cynkra`, `antoine@cynkra.com`

~~R, un catalyseur d'interdisciplinarité : tour d'horizon du projet eGait sur l'analyse de la marche~~

Aymeric Stamm*

Résumé

Une spécialiste de l'analyse des données multivariées, un spécialiste des statistiques sur des variétés, des spécialistes de l'analyse de forme, des spécialistes d'électronique embarquée, des neurologues, des neurochirurgiens, un anesthésiste, des gérontologues, un ingénieur du son, des ingénieurs statisticiens, des cliniciens. A Nantes, Vannes, Rennes, Toulon, Lyon, Florida State University. Des spécialités aux langages hétéroclites et éparpillés sur le globe. Nous proposerons une synthèse du projet eGait qui montrera comment l'écosystème autour de R, Shiny et Github a permis de trouver un langage commun et des outils collaboratifs efficaces pour travailler à la détection de troubles et à la rééducation fonctionnelle de la marche.

L'idée originale est d'exploiter la technologie *centrale inertielle* pour la captation du mouvement rotatif de la hanche au cours de la marche. L'ambition initiale est de suivre la démarche des individus de manière objective (de sorte que les individus ne se sentent pas observés) et quantitative (description de l'entièreté du mouvement rotatif de la hanche sur une foulée typique de l'individu). Le dispositif sert ainsi à la détection précoce des troubles de la marche mais également pourrait permettre de suivre l'évolution de la rééducation en cas de blessures. Dans cette optique, on pourrait même imaginer sonifier la marche de manière à inciter les patients à s'auto-corriger pour que le son produit soit plus *agréable*. Le langage R a permis à la fois la mise en œuvre d'outils permettant une collaboration efficace entre toutes les disciplines dont eGait est à la croisée mais également l'implication d'experts à l'international aux compétences transversales.

Mots-clefs : Captation du mouvement – Rééducation – Multidisciplinarité

* Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629, Nantes Université, École Centrale de Nantes, aymeric.stamm@cnrs.fr

RDepot - 100% open source enterprise management of R and Python repositories

Jonas Van Malder*

Résumé (max 300 mots)

[RDepot](#) is a solution for the management of R and Python package repositories in an enterprise environment. It allows to submit packages through a user interface or API and to automatically update and publish R and Python repositories. Multiple departments can manage their own repositories and different users can have different roles in the management of their packages. With continuous integration infrastructure for quality assurance on R and Python packages, package uploads can be automated. All configuration is declarative and RDepot can be set up as infrastructure as code, which is especially relevant in regulated contexts, since it makes validation activities much easier. Packages from publicly available R repositories such as [CRAN](#) and [Bioconductor](#) can be mirrored selectively in custom repositories for use behind a firewall, in internal networks and offline. Combined with [Crane](#), authentication and fine-grained authorization (using [OpenID Connect](#)) can be configured per repository, which offers extra security when dealing with sensitive data or sensitive methodology.

In this talk we will walk R users and developers through different features of RDepot and demonstrate how these can be useful in different scenarios. The logic of the different workflows will be explained and live demos will be given to see the open source solution in action. We will make sure to address needs ranging from small research groups sharing a handful of packages up to multinational companies managing their R (and Python) code across the globe.

Mots-clefs (3 à 5) : Open Source - Package Management - Infrastructure

*Open Analytics, jonas.vanmalder@openanalytics.eu

Développement

v2.5.0

Released on 31 January 2025.

New features:

- Support binary R packages (Linux)
- Support adding notes to package submissions

v2.4.0

Released on 7 October 2024.

New features:

- Notifications of new events since last logged in
- Show detailed repository publication status

v2.3.0

Released on 30 August 2024.

New features:

- Review and download multiple submissions at once
- Delete multiple packages at once

v2.2.0

Released on 9 July 2024.

New features:

- Provide detailed repository publication status
- Use non-root Docker containers by default

v2.1.0

Released on 4 June 2024.

New features:

- Disable replacing of packages via configuration
- Disable deleting of packages and repositories via configuration

v2.0.0

Released on 19 April 2024.

New features:

- Python package support
- new web UI and well-documented stable ReSTful API v2

Références

- RDepot website: <https://rdepot.io>

RProtoBuf: sérialiser pour communiquer

Tanguy BARTHELEMY*

Résumé

Comment sérialiser de manière optimisée afin de communiquer entre différents langages ?

Lorsqu'on travaille sur un projet, il est souvent nécessaire de transmettre des informations à d'autres outils ou langages. Cela peut se faire via des fichiers, des API, ou d'autres passerelles mais ces méthodes ne sont pas toujours optimisées et adaptées. Les **Protocol Buffers** (Protobuf - Google) sont des structures de données sérialisées et optimisées pour cet objectif précis. Cette présentation mettra en avant les avantages de cette approche et montrera comment l'implémenter facilement en R et avec le package **{RProtoBuf}**. J'illustrerai ces concepts à travers des exemples concrets, notamment les packages **R** du rjdverse qui utilisent cette technologie pour interagir avec les routines Java de JDemetra+ et profiter à la fois des performances de calcul de Java et de la flexibilité de R.

Mots-clefs (3 à 5) : Sérialisation - Data - Package

Développement

Introduction

Le package **{RProtoBuf}** (Eddelbuettel, Stokely, and Ooms (2016)) repose sur la technologie de **Protocol Buffers** (Protobuf - Google) pour sérialiser les données. La sérialisation consiste à transformer les données en un format facilement transportable quel que soit les objets en entrée. À l'inverse, ce format est moins lisible par l'humain. Un des principaux atouts du format Protobuf est qu'il est agnostique du langage, c'est-à-dire qu'il peut être utilisé indépendamment du langage de programmation.

Le rôle de **{RProtoBuf}** est d'interfacier ce format de données en permettant son écriture et sa lecture. Il fournit ainsi des fonctions permettant la conversion des fichiers Protobuf en objets **R** et inversement.

L'utilisation des Protocol Buffers présente plusieurs avantages dans la sérialisation des données. Tout d'abord, comme précisé précédemment, cette sérialisation est *universelle* et ne dépend pas d'un langage de programmation en particulier ce qui facilite la communication entre différents environnements (voir la section sur l'application Java). Ensuite, cette sérialisation est extrêmement efficace et rapide et particulièrement adapté pour des objets de grande taille. Enfin, grâce à **{RProtoBuf}**, les données sont automatiquement traduites en classes adaptées selon le langage de programmation cible.

Méthodes

La configuration des Protocol Buffers se fait au moyen de fichier `.proto`. Ces fichiers spécifient :

- le nom des objets transmis,
- leur contenu (chaînes de caractères, entiers, booléens...),
- et dans quel ordre les objets sont envoyés et reçus.

En R, **{RProtoBuf}** permet de lire ces fichiers `.proto` et génère automatiquement les fonctions qui permettront de convertir les buffers en objets S4 sous R ainsi que de créer les buffers à partir d'objets R.

*Insee, tanguy.barthelemy@insee.fr

Application

JDemetra+ (Smyk et al. (2025)) est un logiciel open source d'analyse des séries temporelles (ajustement saisonnier, détection d'outliers, nowcasting...). Ce logiciel est recommandé par Eurostat et à tous les instituts du système statistique européen (ESS) depuis 2015. Le développement de JDemetra+ a été motivé par l'envie de proposer aux utilisateurs une interface et un support stable de production et d'études des séries temporelles. Une galerie de packages *rjdverse* (Palate et al. (2025)) basé sur la version 3 de JDemetra+ a été développé. Ces packages permettent de faire le lien entre R et les algorithmes de JDemetra+.

Ces algorithmes sont écrits en Java. C'est pourquoi les packages R relatifs à JDemetra+ ont été développés à partir de cette structure et de ces routines de base pour ajouter une surcouche de connexion de JDemetra+ à l'environnement R. Cette connexion est effectuée grâce au Protocol buffer (le protocole de sérialisation de Google). Le Protobuf permet de faire le lien entre les programmes Java et l'environnement R. Il est nécessaire de décrire la structure de classe des programmes Java et d'expliquer un moyen de convertir chaque élément en objet R. Les objets R seront nativement écrits en S4 mais pourront être réécrits en S3 pour être plus facilement manipulables.

Nous allons présenter une application de ces buffers à la connexion entre Java et R. Cet interfaçage permet une connexion bien plus rapide que par l'utilisation du package `{rJava}` notamment pour des gros objets ou pour des aller-retours répétés entre R et Java.

Conclusion

`{RProtoBuf}` utilise les Protocol Buffers pour sérialiser efficacement les objets R. Comme présenté, cette approche peut aussi permettre de communiquer avec d'autres langage (ici Java).

Les principaux avantages de cette méthode sont :

- Réduction des temps de transmission des données,
- Prise en charge de gros objets,
- Facilité de développement, avec une conversion automatique en objets S4,
- Interopérabilité multi-langages, simplifiant les interactions entre R et d'autres environnements.

Références

- Eddelbuettel, Dirk, Murray Stokely, and Jeroen Ooms. 2016. "RProtoBuf: Efficient Cross-Language Data Serialization in r." *Journal of Statistical Software* 71 (2): 1–24. <https://doi.org/10.18637/jss.v071.i02>.
- Palate, Jean, Tanguy Barthelemy, Alain Quartier-la-Tente, Philippe Charles, Anna Smyk, and Corentin Lemasson. 2025. "Organisation Rjdverse." GitHub. <https://github.com/rjdverse>.
- Smyk, Anna, Tanguy Barthelemy, Alain Quartier-la-Tente, and Karsten Webel. 2025. *JDemetra+ Documentation*. jdemetra-new-documentation.netlify.app. <https://jdemetra-new-documentation.netlify.app/>.

Rajouter l'interactivité a vos ggplots avec animint2

Toby Dylan HOCKING*

Résumé

Dans la visualisation de données, les graphiques interactifs permettent de cliquer sur des éléments d'un graphique, pour ensuite changer ce qui est affiché dans un autre graphique. Ils sont utiles pour l'enseignement et la recherche, mais ils sont souvent difficiles à créer. Le package `animint2` permet la création rapide des graphiques interactifs, en utilisant la méthode "grammaire de graphiques" qui a été popularisée par `ggplot2` pour les graphiques non-interactifs. `animint2` est un logiciel mature, développé depuis 2013, disponible sur CRAN, avec beaucoup de documentation et tests [Sievert *et al.*, 2019]. Dans cette présentation, je vais expliquer comment vous pouvez utiliser `animint2` pour créer et partager les graphiques interactifs. Je vais aussi souligner quelques nouveautés permettant l'organisation et la documentation (les galeries, les visites guidées, les liens vers une vidéo et un code source).

Mots-clefs (3 à 5) : Package – Visualisation - interactivité

Références

C. Sievert, S. VanderPlas, J. Cai, K. Ferris, F. U. F. Khan, and T. D. Hocking. Extending `ggplot2` for linked and animated web graphics. *Journal of Computational and Graphical Statistics*, 28(2):299–308, 2019.

* Université de Sherbrooke, Québec, Canada, toby.dylan.hocking@usherbrooke.ca

Rescuelog : Collecte numérique et reporting des interventions de sauveteurs à l'océan avec (ru)ODK, Shiny et SK8

David CARAYON* Geoffrey DEHEZ† Bruno CASTELLE‡

Résumé (max 300 mots)

Les plages du Sud-Ouest de la France présentent des risques élevés de noyades, nécessitant des dispositifs de surveillance efficaces. Le projet Rescuelog modernise la collecte et l'analyse des données sur les interventions de sauveteurs grâce à un workflow entièrement intégré dans l'écosystème R.

Les interventions sont enregistrées via des tablettes Android utilisant ODK Collect, les données étant ensuite extraites avec **ruODK** (Mayer 2020) et structurées dans une base PostgreSQL après validation. Des données environnementales, incluant les prévisions météo et marines, sont enrichies via l'API de Météo-France grâce au package **httr2** (Wickham 2023). Ces données alimentent un modèle prédictif, construit avec **tidymodels** (Kuhn and Wickham 2020), permettant d'anticiper les jours à haut risque de noyade. Ce modèle est versionné et stocké grâce au package **vetiver** (Silge 2023), puis exploité par une application Shiny déployée via le service sécurisé **SK8** (SK8 Team 2024) d'INRAE.

Cette solution démontre la capacité des outils R à proposer des workflows robustes, sécurisés et scalables pour des problématiques opérationnelles critiques comme la prévention des noyades.

Mots-clefs (3 à 5) : ruODK, Shiny, machine learning, plages, sauveteurs

Développement

Contexte

Les plages du Sud-Ouest de la France attirent chaque année des millions de touristes, mais elles sont également marquées par des aléas naturels importants, notamment les courants d'arrachement (baïnes) et les vagues de shore-break (rouleaux de bords). Ces conditions exigent des dispositifs de surveillance réactifs, soutenus par des outils d'analyse performants, pour prévenir efficacement les noyades. Le système Rescuelog cherche à répondre à ce besoin en proposant un workflow digitalisé basé sur l'écosystème R, permettant de collecter, enrichir et exploiter des données d'intervention en temps réel.

Méthodologie

Le workflow (Figure 1) repose sur l'utilisation de plusieurs outils de l'écosystème R et sur une intégration étroite avec les infrastructures de calcul et de déploiement de l'INRAE. Les interventions des sauveteurs sont saisies quotidiennement sur tablette à l'aide d'ODK Collect avant d'être synchronisées sur un serveur central INRAE. Ces données, extraites et structurées via le package **ruODK**, sont d'abord stockées dans un bucket S3 avant d'être transférées, après validation, dans une base PostgreSQL pour archivage. L'enrichissement des données se fait par l'intégration des prévisions météorologiques et marines récupérées à l'aide de **httr2** depuis les API de Météo-France.

Un modèle prédictif est ensuite construit avec **tidymodels**, permettant notamment d'identifier les jours à haut risque de noyade. Ce modèle est versionné et stocké grâce au package **vetiver**. L'exploitation de ces données se fait à travers une application Shiny, développée avec **golem** (Fay et al. 2024) et stylisée avec

*INRAE | UR EABX, david.carayon@inrae.fr

†INRAE | UR ETTIS, jeoffrey.dehez@inrae.fr

‡CNRS | UMR EPOC, bruno.castelle@u-bordeaux.fr.fr

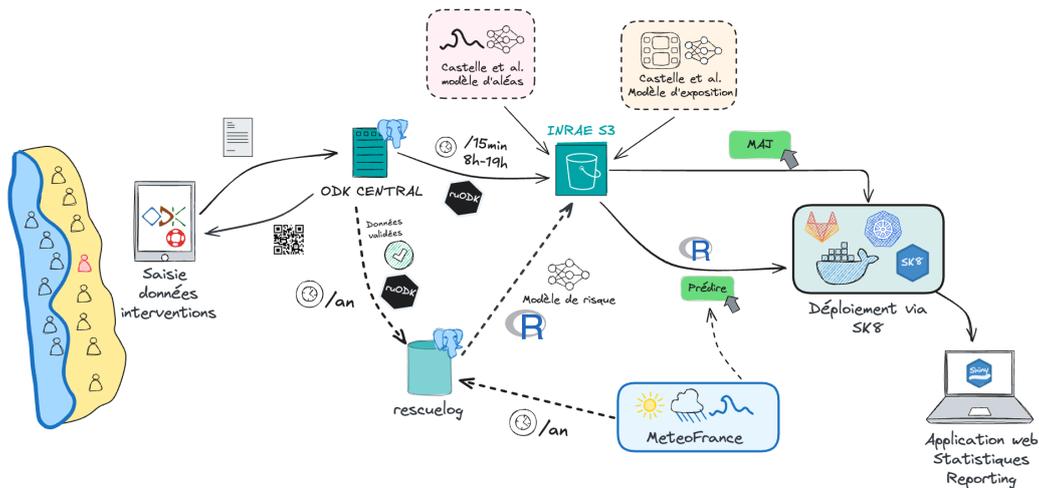


Figure 1: Workflow complet du système Rescuelog

bslib (Sievert, Cheng, and Aden-Buie 2025), qui fournit des visualisations et rapports interactifs adaptés (via quarto) aux besoins des sauveteurs et des organismes en charge de la sécurité. Enfin, cette application est déployée via **SK8**, un service de déploiement proposé par INRAE, qui garantit une solution sécurisée, scalable et maintenable.

Résultats

En 2024, Rescuelog a été utilisé avec succès sur plusieurs dizaines de plages du Sud-Ouest, permettant une collecte de données simplifiée et un accès en temps réel aux statistiques d'intervention. Le workflow a démontré son efficacité et sa fiabilité, avec un potentiel de déploiement à plus large échelle. En 2025, l'application sera déployée sur près de 80 plages, montrant l'adhésion des utilisateurs et les bénéfices opérationnels de cette solution.

Conclusion

Rescuelog illustre comment l'écosystème R, combinés à des technologies open-source telles qu'ODK et des infrastructures telles que SK8, peuvent transformer la gestion opérationnelle des risques en offrant des solutions sécurisées et flexibles.

Bibliographie

- Fay, Colin, Vincent Guyader, Sébastien Rochette, and Cervan Girard. 2024. *Golem: A Framework for Robust Shiny Applications*. <https://CRAN.R-project.org/package=golem>.
- Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. <https://www.tidymodels.org>.
- Mayer, Florian W. 2020. *ruODK: Client for the ODK Central API*. <https://github.com/ropensci/ruODK>.
- Sievert, Carson, Joe Cheng, and Garrick Aden-Buie. 2025. *Bslib: Custom 'Bootstrap' 'Sass' Themes for 'Shiny' and 'Rmarkdown'*. <https://rstudio.github.io/bslib/>.
- Silge, Julia. 2023. *Vetiver: Version, Share, Deploy, and Monitor Models*. <https://CRAN.R-project.org/package=vetiver>.
- SK8 Team. 2024. *SK8: Un Service Institutionnel de Gestion Et d'hébergement d'applications Shiny*. France: INRAE. <https://www.sk8.inrae.fr/>.
- Wickham, Hadley. 2023. *Httr2: Perform HTTP Requests and Process the Responses*. <https://CRAN.R-project.org/package=httr2>.

Rlinguo - R en natif sur mobile

Colin Fay*

Résumé

Rlinguo est une application mobile pionnière qui démontre la faisabilité d'exécuter R nativement sur des appareils mobiles grâce à **webR** (R compilé pour WebAssembly). Disponible sur l'App Store et Google Play, elle marque une avancée significative dans l'utilisation de R au-delà de ses contextes traditionnels.

Développée par ThinkR, Rlinguo est née d'une problématique récurrente : rendre accessible les fonctionnalités de R sur des plateformes mobiles, avec ou sans connexion réseau. Jusqu'à récemment, les limitations de R nécessitaient un serveur distant pour son exécution, excluant toute fonctionnalité hors ligne ou déploiement natif. Avec webR, nous avons surmonté ces obstacles, en publiant une application mobile native pleinement fonctionnelle.

Caractéristiques techniques clés :

- **Exécution native de R**, avec des fonctions R s'exécutent directement sur l'appareil.
- **Fonctionnement hors ligne**, garantissant l'accessibilité en toutes circonstances.
- **Intégration mobile avancée**, avec le stockage local des données en SQL, un traitement par R, et un affichage via JavaScript.
- **Compatibilité avec les packages R**, montrant la possibilité d'intégrer des packages de la communauté.
- **Déploiement sur les app stores**, ce qui est nouveau pour une application utilisant R.

Rlinguo est une application ludique, invitant les utilisateurs à associer des fonctions R à leurs packages. Bien que son objectif principal soit démonstratif, elle illustre de nouvelles possibilités pour les workflows de data science sur mobile, notamment dans des contextes où l'accès à un serveur est limité ou impossible.

Ce projet établit un précédent en matière d'utilisation de R, et ouvre la voie à de nouvelles applications professionnelles, scientifiques et pédagogiques. **Rlinguo** constitue une preuve de concept concrète du potentiel de R à s'intégrer dans des plateformes jusque-là inexplorées.

Mots-clefs (3 à 5) : webR, mobile, application, production

Développement

Rlinguo : une application mobile pionnière pour l'exécution native de R sur mobile

Rlinguo est une application mobile innovante qui explore une nouvelle frontière pour R : l'exécution native sur des appareils mobiles. Développée par ThinkR, cette application tire parti de **webR** — une version de R compilée pour WebAssembly — pour démontrer la faisabilité d'utiliser R sur mobile sans serveur distant. Disponible sur l'App Store et Google Play, Rlinguo ouvre la voie à une utilisation élargie de R dans des environnements où l'accès à un serveur est impossible ou peu pratique.

Depuis des années, R est largement utilisé pour le développement d'applications web interactives, notamment via {shiny}, et pour la création de tableaux de bord et d'API REST. Cependant, malgré les optimisations pour les navigateurs mobiles, ces solutions restent dépendantes de serveurs distants, limitant leur utilisation hors ligne et excluant un déploiement natif sur mobile. Les besoins croissants des utilisateurs,

*ThinkR, colin@thinkr.fr

en particulier ceux travaillant en conditions déconnectées, comme dans des zones rurales ou en pleine nature, ont motivé l'exploration de nouvelles solutions.

L'arrivée de **webR** a changé la donne. En permettant à R de s'exécuter directement dans un environnement JavaScript grâce à WebAssembly, webR a offert une opportunité inédite : intégrer R dans des applications mobiles natives, sans dépendre d'un serveur distant. Rlinguo est née d'une réflexion autour d'une question récurrente : « Peut-on transformer une application R en une application mobile native fonctionnelle ? ».

Le développement de l'application a ensuite été structuré autour de plusieurs axes techniques :

1. **Backend avec R et webR** : Rlinguo utilise webR pour exécuter des commandes R directement sur le téléphone.
2. **Traitement des données** : des données locales sont stockées en SQL, manipulées avec R, puis affichées via une interface JavaScript.
3. **Expérience utilisateur** : une attention particulière a été portée à l'interface, conçue avec Figma pour garantir une expérience intuitive.
4. **Déploiement natif** : l'application a été packagée avec un framework JavaScript pour être distribuée sur l'App Store et Google Play.

Rlinguo propose plusieurs avancées significatives pour la mise en production de R :

- **Exécution native de R sur mobile** : des commandes R classiques comme `sample()` ou `runif()`, ainsi que des fonctions définies sur mesure s'exécutent localement.
- **Fonctionnement hors ligne** : l'application est entièrement fonctionnelle sans connexion internet, un avantage clé pour les utilisateurs en conditions déconnectées.
- **Compatibilité avec les packages R** : l'intégration de packages comme `{praise}` illustre la capacité à reposer sur les packages de la communauté.
- **Utilisation de l'API mobile** : SQLite sur mobile, retours haptiques, etc.

Bien que présentée sous forme ludique, Rlinguo démontre des applications concrètes pour des domaines scientifiques, éducatifs et professionnels. Les workflows nécessitant un traitement local des données, comme dans des contextes de recherche de terrain, ou des solutions pédagogiques interactives pour l'apprentissage de R, pourraient bénéficier de cette approche. En prouvant la faisabilité de l'exécution native de R sur mobile, Rlinguo ouvre la porte à des applications plus complexes

Rlinguo démontre que R peut s'étendre au-delà des environnements traditionnels pour devenir un outil mobile accessible, puissant et fonctionnel hors ligne. Ce projet prouve non seulement la faisabilité technique de l'intégration de R dans des applications mobiles natives, mais ouvre également la voie à une nouvelle ère d'innovation pour les utilisateurs et développeurs R, et représente une première étape prometteuse dans l'évolution de R vers de nouveaux horizons.

Sharing data science artifacts across teams using Crane

Lucianos Lionakis*

Résumé (max 300 mots)

Does your organisation have to share many data science artifacts across teams? This is a problem for many data science organisations and can now be solved using a novel open source product Crane (<https://craneserver.net/>). Crane hosts data science artifacts such as data analysis reports, documentation sites, or even packages and libraries. All of these data science artifacts are kept under strict authentication and authorization using modern protocols (OIDC). In this talk, we walk you through the different features of Crane and provide a live demo to explain the concepts. We will discuss its configuration file and demonstrate that authentication (and authorization) in Crane is fully declarative and allows for fine-grained configuration (at user-level, group-level, network-level or using custom expressions (SpEL)) while still using an intuitive hierarchical tree that corresponds to the directory structure of the data. Next, we will show how artifacts can be accessed from or uploaded into Crane using the Crane API from R (e.g. to automate report updates, use data science artifacts in CI/CD) or using its customizable UI (e.g. for manual tasks and browsing). Further, we zoom in on audit logs to track operations on all files (e.g. for GxP purposes) and detail the different storage backends (a.o. S3 and local file system). To ensure Crane can perform in high security settings the code base has been tested using integration tests reaching a high code coverage of more than 70%. With this talk we want to teach any R user and developer the essentials of Crane and how it can be used to share his/her data science artifacts.

Mots-clefs (3 à 5) : Ingénierie - Data - Package - Reproductibilité

Références

<https://craneserver.net/>

*Open Analytics, lucianos.lionakis@openanalytics.eu

ShinyProxy: easily deploy your Shiny apps

Tobia De Koninck*

Tobias Verbeke†

Résumé (max 300 mots)

ShinyProxy (<https://shinyproxy.io/>) is a 100% open source framework to deploy Shiny (and other) apps or web-based IDEs (like RStudio). Because of its flexibility, ShinyProxy is being used by both small startups and large enterprises. Although ShinyProxy was originally tailored towards hosting Shiny apps, it can host virtually any web app. Since ShinyProxy makes it easy to make reproducible apps, even when using multiple R versions, it's often used by pharmaceutical companies. Nevertheless, it's used by financial and engineering companies as well. ShinyProxy seamlessly integrates with your existing infrastructure (such as authentication providers and databases). The purpose of this talk is to give an introduction to ShinyProxy, explain the use-cases of ShinyProxy and its unique advantages over other solutions. No deep technical knowledge (e.g. of Docker or Linux) is needed to follow this talk, however, the talk will give you enough information to start using ShinyProxy yourselves. As usual, the development of ShinyProxy has continued, therefore we'll also give a preview of upcoming features.

Mots-clefs (3 à 5) : Ingénieure - Data - Shiny - Reproductibilité

Références

<https://shinyproxy.io>

*Open Analytics NV, tdekoninck@openanalytics.eu

†Open Analytics NV, tverbeke@openanalytics.eu

Simplifiez, fiabilisez et accélérez vos workflows R complexes avec {targets}

Vincent Guyader*

Résumé

La gestion de workflows R complexes peut vite devenir laborieuse et source d'erreurs, surtout si vous devez tout relancer manuellement à chaque modification. Le package {targets} propose une solution efficace pour automatiser ces workflows tout en assurant un suivi reproductible. Il vous suffit de définir les étapes cibles (ou « targets ») de votre analyse et leurs dépendances. {targets} orchestre alors l'exécution : seules les étapes nécessaires sont recalculées en cas de modification des données ou du code, évitant ainsi tout calcul redondant.

Par exemple, vous définissez une cible pour importer un jeu de données, une pour le nettoyer et une pour générer un graphique. Si les données changent, {targets} ne recalculera que les étapes concernées (nettoyage, graphique) sans toucher au reste. Ce principe rappelle le package {memoise}, qui met en cache les résultats de fonctions pour éviter les recalculs inutiles ; de même, {targets} conserve les résultats intermédiaires de chaque étape. Cette mise en cache évite de répéter des calculs coûteux et améliore les performances. Vous gagnez du temps et profitez d'un workflow plus robuste et reproductible, avec moins de risques d'oubli ou d'erreur – vous pouvez ainsi vous concentrer sur l'essentiel : votre analyse.

Mots-clés : Automatisation - Reproductibilité - Workflow

Développement

Dans cette présentation, nous commencerons par introduire le concept de mise en cache en nous appuyant sur le package {memoise} comme modèle théorique. Ce package démontre comment le stockage temporaire des résultats d'une fonction permet d'éviter des recalculs inutiles, améliorant ainsi l'efficacité des analyses répétitives.

Nous passerons ensuite à {targets}, un outil qui étend ce principe à l'automatisation des workflows en R. À l'aide d'un exemple simple, nous illustrerons comment définir un pipeline de calcul composé de cibles interdépendantes : depuis l'importation des données, leur traitement, jusqu'à la génération d'une visualisation. L'idée est de suivre pas à pas l'exécution de ce workflow, en mettant en évidence comment {targets} détecte automatiquement les modifications et ne recalcule que les étapes concernées, garantissant ainsi une exécution rapide et reproductible.

La démonstration live permettra de voir concrètement le fonctionnement de ces mécanismes, en observant en temps réel le recalcul sélectif des cibles suite à une modification. Pour conclure, nous ouvrirons la discussion sur des cas d'usage plus complexes, afin de montrer que cette approche, bien que présentée via un exemple simplifié, est entièrement adaptable à des projets d'envergure nécessitant une gestion fine des dépendances et une optimisation des temps de calcul.

Références

1. Landau W. et al. (2023). *targets: A Pipeline Toolkit for Reproducible Workflows in R*. R package version 0.1.6. Disponible sur CRAN : <https://CRAN.R-project.org/package=targets>.

*ThinkR, vincent@thinkr.fr

-
2. Landau W. et al. (2023). *targets* – code source et documentation. Dépôt GitHub : <https://github.com/ropensci/targets>.
 3. Wickham H. (2023). *memoise: Memoization for R*. R package version 2.0.1. Disponible sur CRAN : <https://CRAN.R-project.org/package=memoise>.
 4. Wickham H. (2023). *memoise* – code source et documentation. Dépôt GitHub : <https://github.com/r-lib/memoise>.

Strengthening confidence in LLM-generated responses with TrustMe: a package for evaluating semantic proximities between responses generated by the NaileR package.

Rémi Mahmoud*

Sébastien Lê†

Résumé

Large Language Models (LLMs) are increasingly used across fields, but their fluency can mask factual inaccuracies, or hallucinations. Assessing the reliability of LLM-generated responses has therefore become a key challenge. A common and intuitive approach involves (Lin, Trivedi, and Sun (2024)):

1. Generating multiple responses to the same prompt,
2. Measuring the similarity between responses, and
3. Deriving an uncertainty score.

In previous work, we introduced [NaileR](#), an R package designed for interpreting continuous or categorical latent variables: typically, dimensions from an exploratory multivariate method, or a class variable from an unsupervised clustering algorithm. NaileR generates structured prompts based on the analysis results and leverages locally installed LLMs (e.g., Llama2, Llama3, GPT-3 *via* [Ollama](#)). However, the automatic prompt generation process introduces an additional layer of uncertainty, potentially reducing trust in the output. To address this issue, we propose TrustMe (*Turning to R for Understanding Semantic Topology from Model's Embeddings*).

TrustMe is an R package designed to *enhance confidence* in LLM-generated responses, particularly within the *NaileR ecosystem*. TrustMe offers different strategies to evaluate the reliability of LLM-generated responses.:

1. *Cross-model consistency*: Assesses the similarity of responses generated by *different LLMs* for the same prompt.
2. *Internal-model consistency*: Evaluates the stability of a model *via* its responses by comparing *multiple calls* to the same LLM with the same prompt.
3. *Data-model consistency*: Measures the similarity of responses generated from *siblings prompts* (possibly generated by NaileR).

For all these purposes, *TrustMe* provides several key outputs, mainly based on the embeddings of the responses (hence the name of the package):

- A “*central*” *response* that summarizes the set of generated responses.
- A *graphical representation* of responses’ variability.
- A set of *uncertainty scores* computed using metrics such as *eccentricity*, *eigenvalue sum of the graph Laplacian*, and *Krippendorff’s alpha*.

These outputs offer assessments of trustworthiness, helping users determine whether an LLM-generated response is reliable or potentially inconsistent.

*Institut Agro Rennes-Angers, remi.mahmoud@agrocampus-ouest.fr

†Institut Agro Rennes-Angers, sebastien.le@agrocampus-ouest.fr

Mots-clefs : LLM - Reproducibility - Repeatability - Multivariate exploratory analysis

Références

Lin, Zhen, Shubhendu Trivedi, and Jimeng Sun. 2024. “Generating with Confidence: Uncertainty Quantification for Black-Box Large Language Models.” <https://arxiv.org/abs/2305.19187>.

Tester une application Shiny : méthodologie et outils pratiques

Arthur Bréant*

Résumé (max 300 mots)

Qui n'a jamais développé une application Shiny sans prendre le temps d'écrire des tests ? On se dit souvent que “*ça marche bien comme ça*” et que l'ajout de tests viendra plus tard... Mais ce “plus tard” arrive rarement, et un jour, une modification anodine casse tout sans qu'on comprenne pourquoi.

Pourtant, dans d'autres domaines, cette approche nous semblerait risquée. Imaginez que vous construisiez une maison sans vérifier si les fondations tiennent. Ça semble absurde, non ? Et pourtant, c'est exactement ce qui se passe lorsqu'on développe une application Shiny sans tests.

Dans cette présentation, nous verrons comment tester une application Shiny efficacement à travers trois niveaux :

- **Les tests unitaires** : pour s'assurer que chaque brique fonctionne indépendamment.
- **Les tests d'intégration** : pour vérifier que les modules communiquent bien entre eux.
- **Les tests end-to-end (E2E)** : pour simuler l'expérience utilisateur et prévenir les bugs avant la mise en production.

À travers une application Shiny exemple, nous verrons comment mettre en place des tests unitaires et d'intégration avec `testthat`. Enfin, nous découvrirons **Playwright**, un outil puissant pour automatiser des tests *end-to-end* et garantir que l'application se comporte comme prévu, aussi bien en local qu'en CI/CD.

Tester une application Shiny, ce n'est pas une perte de temps. C'est une assurance pour éviter des bugs imprévus, faciliter la maintenance et, surtout, se développer sereinement. On vous montre comment faire, avec des outils simples et efficaces !

Mots-clefs : Shiny – Tests – Playwright – CI/CD – testthat

*ThinkR, arthur@thinkr.fr

Traiter des verbatims et questions ouvertes à l'aide d'un modèle d'IA local

Thomas Vroylandt*
Victoire Chatain^x
Emmanuel Herbepin[%]

Résumé

Il est fréquent de disposer lors d'enquêtes ou de dispositifs citoyens de données issues de champs libres ou de verbatims. L'analyse classique de ces données¹ passe par le compte des mots et des unités de textes, puis la classification de ces mots à l'aide de dictionnaires. Des packages comme `{tidytext}`² existent pour faciliter ces analyses. Enfin, une approche supervisée à l'aide de BERT³ donne en général de bons résultats, mais demande un travail préalable, parfois important.

De façon complémentaire, il est possible d'utiliser un modèle de langage (LLM) pour analyser de façon non supervisée ou peu supervisée ces verbatims, en identifier les objets principaux ou de les classer dans des catégories. Face à l'impact environnemental important, au coût financier non nul et aux risques de perte de la confidentialité des données, les modèles d'intelligence artificielle (IA) hébergés en local, peuvent être un bon compromis entre puissance et qualité.

Les packages `{mall}`⁴ et `{ollamar}`⁵ sur lequel il repose, proposent des fonctions plus faciles d'accès pour réaliser ces tâches d'analyse. Ils peuvent être complétés par `{ellmer}`⁶ pour les cas plus avancés. Un certain nombre de précautions, à la fois dans la préparation des données, les instructions (*prompts*) aux modèles ou la vérification des résultats, doivent être mis en place pour garantir un résultat de qualité.

Mots-clés: Text mining – IA – Enquêtes

Développement

Cette communication se fonde sur l'analyse de jeux de données disponibles sur data.gouv.fr contenant des verbatims issus de dispositifs citoyens de contributions, à la fois de façon classique et à l'aide du package `{mall}` (et dans une moindre mesure du package `{ellmer}`, qui est plus polyvalent). Ce package se fonde sur l'utilisation d'Ollama, qui permet d'héberger en local des modèles de langages de petite taille, et du package `{ollamar}` qui l'interface.

L'objectif est d'illustrer l'intérêt, le fonctionnement et les limites du recours à des modèles d'IA, en particulier local, pour l'analyse de données textuelles dans des champs libres, qui demandent souvent un traitement manuel lourd ou sont laissés de côté.

Sont notamment abordés :

- Le fonctionnement du package `{mall}` et son utilité
- La préparation des données
- Des précautions dans la rédaction des instructions

* Kantiles, [thomas\(at\)kantiles.com](mailto:thomas@kantiles.com)

^x Kantiles, [victoire\(at\)kantiles.com](mailto:victoire@kantiles.com)

[%] Kantiles, [emmanuel\(at\)kantiles.com](mailto:emmanuel@kantiles.com)

- Les vérifications à réaliser dans l'analyse des résultats
 - Les autres options possibles (et notamment l'usage de BERT)
-

Références

1. Lebart L., Salem A., *Statistique textuelle*, 1994
2. Silge J., Robinson D., *Text Mining with R*, 2017 : <https://www.tidytextmining.com/>
3. Devlin J., Chang M., Kenton L., Toutanova K., "BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv*, 2018 : <https://arxiv.org/abs/1810.04805>
4. {mall} package : <https://mlverse.github.io/mall/>
5. Lin H., Safi T., {ollamar} package : <https://hauselin.github.io/ollama-r/>
6. Wickham H., Cheng J., Jacobs A., {ellmer} package : <https://ellmer.tidyverse.org/>

Traiter les données collectées à l'aide de formulaires pour partager des inventaires et des résultats d'enquêtes. Développement informatique et application aux pratiques agroécologiques en Europe

Frédéric Vanwindekens*

Résumé

Les données requises pour l'étude de la gestion des systèmes socio-écologiques proviennent de sources très diversifiées. Les données écologiques, sur le sol, climat par exemple, sont en grande partie disponibles dans des bases de données spatialisées. D'autres données, sur l'utilisation du territoire, sur les structures économiques en présence, peuvent être disponibles à travers les services publics, les statistiques officielles. Ces types de données sont communément traitées en sciences des données, par les chercheurs, par les acteurs privés et dans les administrations. Mais la composante humaine de ces systèmes rend indispensable l'intégration de données décentralisées, liées aux connaissances des acteurs de terrain, leurs avis ou les pratiques de gestion qu'ils mettent concrètement en place.

Pour intégrer ces données essentielles à une démarche d'analyse systémique, les démarches scientifiques incluent principalement des enquêtes. Celles-ci peuvent être qualitatives, à la manière ethnographique, ou quantitative. Si les deux sont d'intérêt, nous nous focalisons ici dans le traitement des données d'enquêtes quantitatives collectées à travers des formulaires en ligne. Nous présentons les approches développées pour automatiser le traitement, l'analyse et le partage de ces données à travers des fonctions R, des packages R et des Shiny Apps. Nous proposons une illustration de ce travail dans le cadre de notre pratique de recherche quotidiennement centrées sur l'étude des pratiques agro-écologiques dans les systèmes agraires à différentes échelles (Wallonie, Europe).

Mots-clés : Data – Package – Shiny – Enquête – Systémique

Développement

La conduite d'enquêtes, qu'elles soient quantitatives ou qualitatives, est une pratique de recherche répandue en sciences sociales. Elle l'est moins dans les sciences de la nature. Elle retrouve par contre sa pertinence dans l'étude des systèmes socio-écologiques, dans lesquelles une société humaine interagit, gère voire exploite les services d'un écosystème, par exemple la forêt, une pêcherie, ou les plaines cultivées.

Ancrés dans l'étude des systèmes agraires, nos travaux ont eu pour objectif (i) de révéler les systèmes de pratiques des éleveurs en Ardenne et en Famenne (B) [Vanwindekens, 2014], (ii) de dresser les orientations des fermes laitières wallonnes (B) à la suite de l'arrêt des quotas laitiers européens [Dalcq et al., 2020], (iii) de quantifier les éléments favorisant ou freinant l'adoption de pratiques agro-écologiques dans la gestions des vignes [Garini et al., 2017], des systèmes de grandes cultures [Cornu et al., 2023] et de pratiques favorisant la fertilité, la qualité et la santé des sols en Europe (11 pays) [Heller et al., 2024] ou (iv) d'identifier les formes de connaissances que les chercheurs et les praticiens ont du concept de biodiversité des sols [Vanermen et al., 2020].

Les approches méthodologiques que nous avons développées pour la réalisation et du traitement des enquêtes qualitatives ont été présentées dans lors de conférences useR! précédentes à Bruxelles et à Toulouse [Vanwindekens et al., 2017] et publiée sous forme du package cogmapr disponible sur CRAN [Vanwindekens et al., 2019].

*Centre wallon de Recherches agronomiques, f.vanwindekens@cra.wallonie.be

Le traitement des données quantitatives collectées à travers des formulaires en ligne, préférentiellement LimeSurvey, ont fait l'objet de développements récent. En fonction des objectifs de l'étude, deux voies différentes sont privilégiées, via les fonctions dans deux packages en préparation.

Le package R `surveyvisualizr` qui vise à visualiser les résultats d'enquêtes quantitatives (données numériques, catégorielles, de type Liker) via la création de graphiques, de table de contingence, et par le biais d'une application ShinyApp.

Le package R `inventr` [Vanwindekens, 2022] quant à lui, vise à exploiter les données collectées dans le but de créer des inventaires. Un exemple concret est un inventaire européen des pratiques de bonne gestion des sols agricoles de manière à garantir la soutenabilité des services qu'ils rendent [Vanwindekens and Heller, 2024, Vanwindekens, 2024].

Références

- Marie-Aline Cornu et al. Identification and description of relationships between actors involved in crop diversification experiences across europe. *Agronomy for Sustainable Development*, 43(5) :67, 2023.
- Anne-Catherine Dalcq et al. The wallon farmers position differently their ideal dairy production system between a global-based intensive and a local-based extensive model of farm. *PLOS ONE*, 15(12) :1-22, 12 2020.
- C.S. Garini et al. Drivers of adoption of agroecological practices for winegrowers and influence from policies in the province of trento, italy. *Land Use Policy*, 68 :200 – 211, 2017.
- Olivier Heller et al. Towards enhanced adoption of soil-improving management practices in europe. *European Journal of Soil Science*, 75(2) :e13483, 2024.
- Iris Vanermen et al. What do scientists and managers know about soil biodiversity? comparative knowledge mapping for sustainable forest management. *Forest Policy and Economics*, 119 :102264, 2020. ISSN 1389-9341.
- Frédéric Vanwindekens. *Les pratiques dans la gestion des systèmes socioécologiques : développements méthodologiques & application à la gestion des prairies en région herbagère belge*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgique, 2014.
- Frédéric M Vanwindekens, Didier Stilmant, and Philippe V Baret. Using r in transdisciplinary approaches for visualiazing and analysing people's perceptions, knowledge and practices in complex social-ecological systems. In *The R User Conference, useR! 2017 July 4-7 2017 Brussels, Belgium*, page 251, 2017.
- Frederic M. Vanwindekens, Didier Stilmant, and Philippe V. Baret. *cogmapr : Cognitive Mapping Tools Based on Coding of Textual Sources*. Begium, 2019. URL <https://frdvnw.gitlab.io/cogmapr/>.
- Frédéric Vanwindekens. `inventr` - a r package for working with inventories, March 2022. URL <https://doi.org/10.5281/zenodo.6372365>.
- Frédéric Vanwindekens. Printable version of the inventory of soil management practices (i-sompe, version scientific paper), February 2024.
- Frédéric Vanwindekens and Olivier Heller. A web-app for exploring the inventory of innovative soil management practices, February 2024. URL <https://doi.org/10.5281/zenodo.10630668>.

Transition SAS à R en recherche clinique : traduction et validation de macros SAS

Nusaibah Ibrahim^{*}, Livia Pierotti[†], Dan Chaltiel[°]

Résumé

Au sein de l'institut Gustave Roussy, le service de biostatistiques et d'épidémiologie effectue des analyses dans la recherche clinique. Le service est mené à analyser et traiter des bases de données cliniques.

Comme beaucoup de services statistiques, nous sommes en transition de SAS vers R. Une grande partie de nos outils internes repose sur des macros SAS, qu'il faut désormais traduire en fonction R. Pour centraliser et structurer ce travail, nous avons créé un package collaboratif, {grstat} (actuellement en cours de développement), que nous développons sur GitHub.

L'un des enjeux majeurs concerne la génération des tableaux d'événements indésirables : ces tableaux standards sont complexes à produire et comportent un risque d'erreur élevé. Mais comment s'assurer de la fiabilité des résultats ? Pour cela, nous avons mis en place un processus de validation rigoureux.

Notre approche consiste à générer les outputs en SAS et en R, puis inclut une étape de prétraitement pour harmoniser les tables, suivie d'une comparaison automatique entre les résultats obtenus avec la macro SAS et la fonction R du package {grstat}. Le tout aboutit à un tableau récapitulatif facilitant l'identification des écarts éventuels.

Ce processus de validation a déjà été appliqué à plusieurs études réelles (4 essais cliniques, 6 560 observations). Aucune divergence n'a été détectée entre les tables SAS et R. Cette démarche est également en cours de réalisation sur des données simulées.

Ce travail nous permet de fiabiliser notre transition vers R en garantissant la conformité des analyses. Il s'inscrit dans une démarche plus large d'automatisation et de reproductibilité, avec pour objectif d'améliorer l'efficacité et la qualité de nos analyses statistiques.

Mots-clefs : Validation - Reproductibilité - Statistiques - Package - Recherche Clinique

* Institut Gustave Roussy, nusaibah.ibrahimi@gustaveroussy.fr

† Institut Gustave Roussy, livia.pierotti@gustaveroussy.fr

° Institut Gustave Roussy, dan.chaltiel@gustaveroussy.fr

Truffles : une application Shiny pour cultiver ses données... et ses truffes !

Murielle Delmotte*

Résumé

Les truffes, ça ne pousse pas par hasard ! La trufficulture repose sur un suivi minutieux des arbres, des sols et des récoltes. C'est en cherchant à simplifier la gestion d'une plantation de chênes truffiers que *Truffles* est né.

Conçu à l'origine pour aider un agriculteur passionné, ce package R est devenu une application Shiny interactive pour visualiser, cartographier et gérer les plantations de chênes truffiers en toute simplicité.

Grâce à Truffles, fini les carnets éparpillés et les relevés approximatifs au milieu du champs. Voici ce qu'il permet de faire :

- Une **carte interactive** pour voir d'un coup d'œil où se trouvent les arbres et suivre leur évolution.
- Des **fiches détaillées** sur chaque chêne truffier : date de plantation, rendement total et autres informations précieuses.
- L'enregistrement des **récoltes** directement dans l'application pour un suivi précis des truffes trouvées.
- Des **analyses visuelles** des rendements, avec des graphiques clairs pour mieux comprendre les tendances et optimiser la gestion de la truffière.

Cette présentation racontera la petite histoire derrière Truffles, son développement et comment il peut s'adapter aux besoins des agriculteurs et passionnés de trufficulture.

Mots-clefs : Shiny – Cartographie – Trufficulture – Gestion agricole – Suivi des récoltes

*ThinkR, murielle@thinkr.fr

Tutoriel {data.table}

Toby Dylan Hocking*

Résumé du tutoriel

{data.table} est un package R qui permet le traitement efficace de grands jeux de données. Il inclut des fonctions pour la lecture et l'écriture de fichiers CSV, l'agrégation, la restructuration, la jointure, etc. Par rapport aux autres logiciels avec ces fonctions, l'utilisation de {data.table} peut être 10x plus efficace en temps de calcul et en utilisation de mémoire. Il est aussi très facile à installer (il ne dépend pas d'autres packages) et à mettre à jour (les nouvelles versions de {data.table} sont compatibles avec les anciennes). Sur CRAN depuis 2006, il est très utilisé, avec plus de 1500 autres packages qui en dépendent.

Depuis 2024, il y a même des traductions en français, pour les messages d'erreur, ainsi que les vignettes.

Ce tutoriel va discuter de :

- l'utilisation du package (fonctions de base et avancées),
- son efficacité relative aux autres packages (tidyverse, pandas, etc),
- la manière d'écrire un package qui dépend de {data.table},
- quelques remarques sur la communauté de développeurs de {data.table} qui est très ouverte aux contributions du public.

Prérequis : maîtrise de base de la programmation en langage R.

Durée : 2h

À réaliser avant de venir au tutoriel :

Installer R version 3.3.0 ou supérieure et ensuite utiliser la commande

```
install.packages("data.table")
```

Informations complémentaires :

- Page web en anglais : <http://r-datatable.com/>
- Vignette d'introduction en français : <https://rdatatable.gitlab.io/data.table/articles/fr/datatable-intro.html>
- Code source du tutoriel : <https://github.com/tdhock/2023-10-LatinR-data.table>

* Professeur Agrégé, Département d'informatique, Université de Sherbrooke, Québec, Canada,
Toby.Dylan.Hocking@USherbrooke.ca

Un exemple d'utilisation du package *likert* et l'interface *R Shiny* pour la visualisation et l'analyse de questionnaires d'auto-évaluation des pratiques de pédagogie inclusive à l'université

Viet Chau Linh Nguyen^{*1}, Hiba Ghanem¹, Stéphanie Mazza², Amandine E Rey^{1,2}, Alice Gomez¹

Résumé (max 300 mots)

Les questionnaires représentent un outil de recherche pertinent pour répondre à certaines problématiques en sciences humaines et sociales. Cependant, la richesse et la complexité des informations apportées par ces questionnaires nécessitent des formats de visualisation graphique adaptés. Nous aborderons un exemple d'utilisation du package *likert* (Bryer, 2022) afin de visualiser les pourcentages de réponses concernant l'attitude des répondant.e.s vis-à-vis de certaines pratiques (de « Tout à fait en désaccord » à « Tout à fait d'accord »), et la fréquence d'application (de « Jamais » à « Toujours ») de ces pratiques par les répondant.e.s (s'ils ou elles sont enseignant.e.s) ou par leurs enseignant.e.s (s'ils ou elles sont étudiant.e.s). Afin de mieux comprendre les facteurs modulant les réponses et de faciliter la discussion de ces résultats avec des collaborateur.ice.s, nous visualisons de manière interactive les réponses des répondant.e.s ayant un profil spécifique (genre, établissement, niveau d'étude, besoin éducatif particulier, etc.) à l'aide de l'interface *R Shiny* (Chang et al., 2025).

Mots-clefs (3 à 5) : Questionnaire, échelle Likert, Shiny

Développement

Nous avons analysé 2213 réponses d'étudiant.e.s et 209 réponses d'enseignant.e.s sur la pertinence et l'adoption de 33 pratiques pédagogiques inclusives (varier les modalités de test, varier les modalités d'expression, permettre un délai supplémentaire si besoin, etc.) à une version adaptée du questionnaire Inclusive Teaching Strategies Inventory (ITSI, Lombardi et al., 2011). Pour un certain nombre de pratiques les résultats vont dans le sens d'un décalage entre les attitudes et les actions des enseignant.e.s, ces dernier(e)s n'adoptant pas systématiquement des pratiques qu'elles.ils jugent pertinentes. Ce décalage est aussi observé par les étudiant.e.s interrogé.e.s sur la pratique de leur enseignant.e.s. Les obstacles à la mise en œuvre de ces pratiques ainsi que les pistes d'amélioration seront discutés. Dans cette présentation, nous mettrons en valeur ces observations à l'aide du package *likert* et de l'interface *R Shiny*.

¹Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, Eduwell Team, F-69500, Bron, France

† ²Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, FORGETTING Team, F-69500, Bron, France

*Autrice intervenante : viet-chau-linh.nguyen@univ-lyon1.fr

Références

Bryer J (2022). *likert: Analysis and Visualization Likert Items*. <http://jason.bryer.org/likert>, <http://github.com/jbryer/likert>.

Lombardi, A. R., Murray, C., & Gerdes, H. (2011). College faculty and inclusive instruction: Self-reported attitudes and actions pertaining to Universal Design. *Journal of Diversity in Higher Education*, 4(4), 250–261. <https://doi.org/10.1037/a0024961>

Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2025). *shiny: Web Application Framework for R*. R package version 1.10.0.9000, <https://github.com/rstudio/shiny>, <https://shiny.posit.co/>.

Utiliser les GitHub Actions pour automatiser des tâches : exemple de l'update de la version dev d'un package

Dan Chaltiel^{*†}

Résumé (max 300 mots)

Le développement de packages R se fait de plus en plus via GitHub, qui propose un outil puissant d'automatisation : GitHub Actions (GHA). Cet outil permet d'exécuter du code automatiquement en réponse à des événements, comme un push sur la branche main. Dans la communauté R, certaines actions sont largement adoptées, telles que check-standard pour exécuter R-CMD-CHECK ou pkgdown pour générer un site de documentation. Le package usethis facilite leur mise en place sans nécessiter une compréhension approfondie de GHA. Cependant, la flexibilité de GitHub Actions va bien au-delà de ces usages courants : il est possible d'exécuter directement du code, y compris du R, pour automatiser des tâches répétitives.

Pour illustrer ceci, je vais montrer comment automatiser une tâche simple, mais très pénible : la mise à jour de la version de développement d'un package R. Traditionnellement, la version dev (4^{ème} chiffre de la version, supérieur à 9000) doit être mise à jour manuellement après chaque modification significative du code. Grâce à GHA, ce processus devient entièrement automatique : à chaque push, la version dev est incrémentée, la section "Installation" du Readme est mise à jour, et un tag est créé. Ainsi, l'installation via ``remotes::install_github()`` peut se faire avec une version identifiant précisément l'état du code.

Cette approche simplifie considérablement la collaboration avec une équipe de testeurs et peut être généralisée à d'autres tâches d'automatisation pour améliorer la gestion des packages en développement actif.

Mots-clefs (3 à 5) : GitHub Actions – Automatisation – Package – Versionning

* Bureau de Biostatistique et d'Épidémiologie, Gustave Roussy, Université Paris-Saclay, Villejuif, France

† Oncostat, CESP U1018, Inserm, Équipe Labellisée Ligue Contre le Cancer, Université Paris-Saclay, Villejuif, France
dan.chaltiel@gustaveroussy.fr

Workflow automatisé d'intégration et d'analyse de données en Shiny

Mélanie DEMEURE^{*†}, Véronique Kowandy^{*}, Emmanuelle MAUGER^{*}

Résumé

La capitalisation des données et l'automatisation des analyses statistiques sont cruciales pour les entreprises, leur permettant d'améliorer la prise de décision, d'optimiser les processus et de stimuler l'innovation. Chez Chanel, l'équipe Data de la Recherche Parfums Beauté développe des outils R Shiny pour faciliter l'exploitation des données par les équipes métiers. Nous présentons ici un workflow automatisé pour l'intégration et l'analyse des données, particulièrement utile lorsque les données biologiques et cliniques sont issues de formats variés.

Grâce à des interfaces Shiny, nous offrons une automatisation accessible sans nécessiter de compétences avancées en programmation. Le workflow comprend plusieurs applications : la première standardise les fichiers Excel hétérogènes en un format adapté à une base de données SQL ; la deuxième propose des fonctionnalités d'analyse statistique et de visualisation pour extraire des indicateurs clés, et génère automatiquement des rapports structurés ; la troisième application utilise les données de caractérisation des produits et les mesures instrumentales pour prédire la performance des produits.

Ce système améliore la traçabilité des données, réduit les erreurs humaines et accélère l'accès aux analyses. Son architecture modulaire s'adapte facilement à divers types de données et instruments. Notre approche illustre comment l'intégration de SQL, R Markdown et R Shiny peut optimiser la gestion des données scientifiques et industrielles en automatisant leur traitement, leur stockage et leur analyse.

Mots-clefs : Statistique – Capitalisation de données – Shiny – SQL – Markdown

Contacts :

Mélanie Demeure, Chanel & IT&M Stats melanie.demeure@chanel.com

Emmanuelle Mauger, Chanel emmanuelle.mauger@chanel.com

Développement

Lorsque des équipes métiers effectuent des mesures biologiques et cliniques, les formats de sortie de données sont divers. La mise en forme des données est une étape laborieuse pour qui souhaite les analyser. L'analyse elle-même des données peut aussi se révéler compliquée quand les connaissances en programmation statistiques sont limitées. Heureusement, il est possible d'automatiser un bon nombre de processus sur R, de manière accessible, sans confronter le personnel au code, grâce aux interfaces Shiny.

Pour aider les équipes métiers à capitaliser sur leurs données, nous avons créé un workflow de plusieurs applications Shiny, déployé sur nos serveurs en internes, permettant l'automatisation et la standardisation du traitement des fichiers issus d'instruments de mesure variés.

Une page d'accueil unique redirige les équipes sur 3 applications principales.

La première application permet l'intégration des données en transformant des fichiers Excel hétérogènes en un format standardisé, adapté à une base de données **SQL**. Cette transformation s'appuie sur des règles définies et une procédure SQL dédiée pour assurer la cohérence et l'intégrité des données stockées.

La seconde application permet l'exploitation des données en extrayant les informations depuis la base SQL. Elle offre des fonctionnalités d'analyse statistique et de visualisation standardisées en fonction du type de données, afin de fournir des indicateurs clés. De plus, elle génère automatiquement des **rapports** structurés pour faciliter l'interprétation et la prise de décision.

La troisième application permet de faire des prédictions de performance produit grâce à une utilisation combinée des données de caractérisation produits et les mesures instrumentales capitalisées dans la base.

Ce workflow améliore la traçabilité des données, réduit les erreurs liées aux manipulations manuelles et accélère l'accès aux analyses. Son architecture modulaire permet une adaptation facile à divers types de données et d'instruments. Le métier peut alors davantage porter son temps et énergie sur la qualité des données et l'interprétation des analyses. Notre approche illustre comment l'intégration de **R Shiny et SQL** peut optimiser la gestion des données scientifiques et industrielles en automatisant leur traitement, leur stockage et leur analyse.

audio.whisper – Speech Recognition in R

Jan Wijffels*

Résumé (max 300 mots)

audio.whisper is a wrapper around the whisper.cpp C++ library allowing to transcribe and translate wav files in multiple languages. It allows offline transcription of recorded audio files on your own infrastructure. It uses GPU acceleration but can also run on CPU. Whisper has low word error rate. It is integrated with existing R tooling for audio handling, and it is MIT-licensed.

Example use cases include transcribe of call centers or medical recordings. It performs Large Language Modelling on transcriptions to get structured data. It also provides guidance on the calls.

See <https://github.com/bnosac/audio.whisper>

* BNOSAC, <https://www.bnosac.be>, jwijffels@bnosac.be

dbreportR : Un package R pour la documentation automatisée des bases de données

Sébastien Boutry*

David CARAYON†

Résumé (max 300 mots)

La documentation exhaustive et standardisée des bases de données relationnelles est cruciale pour garantir leur maintenabilité, leur réutilisation et leur interopérabilité. Nous présentons ici dbreportR, un package R qui automatise la génération de documentation complète pour les bases de données PostgreSQL. dbreportR offre trois fonctionnalités principales :

1. la génération de dictionnaires de données détaillés sous forme de projets quarto
2. la création automatique de diagrammes entité-relation (ERD)
3. l'export de métadonnées dans des formats standards, tels que l'Ecological Metadata Language (EML) en écologie

Le package analyse automatiquement la structure de la base de données, incluant les schémas, tables, colonnes, clés primaires et étrangères, contraintes et index. Les dictionnaires de données générés incluent des descriptions détaillées de chaque élément, enrichies par les commentaires SQL existants. Les diagrammes ERD, créés via les packages **dm** (Schieferdecker, Müller, and Bergant 2024) et **DiagrammeR** (Iannone and Roy 2024), permettent une visualisation intuitive des relations entre tables, facilitant la compréhension de l'architecture de la base de données.

L'export aux formats standardisés permet l'intégration avec les standards de différents champs disciplinaires, favorisant l'interopérabilité avec d'autres outils et le partage de données dans la communauté scientifique.

dbreportR propose actuellement la documentation de bases de données **PostgreSQL** uniquement. Un portage vers d'autres systèmes de gestion de bases de données (MariaDB, SQLite) est prévu pour les prochaines versions.

Ce package répond aux besoins croissants de documentation standardisée dans les projets de données, particulièrement en écologie et en sciences de l'environnement, où la traçabilité et la reproductibilité sont essentielles.

Mots-clés (3 à 5) : R - PostgreSQL - Documentation - ERD - Ecologie

Développement

Le package {dbreportR} se connecte à une base de données PostgreSQL et en extrait automatiquement la structure (schémas, tables, colonnes, clés primaires et étrangères, contraintes, index). Il enrichit ces métadonnées en intégrant les commentaires SQL existants.

Ces informations sont principalement récoltées par le biais de fonctions présentes dans le package **dm** en ajoutant la prise en charge de liens *cross_schemas* et par le biais de requêtes SQL spécifiques exécutées via **DBI** (R Special Interest Group on Databases (R-SIG-DB), Wickham, and Müller 2024) pour les commentaires, par exemple. Les tables sont mises en forme par le biais de **gt** (Iannone et al. 2024).

La génération des dictionnaires de données diffusables et interactifs repose sur Quarto (Allaire and Dervieux 2024), un écosystème de publication scientifique très utilisé dans l'écosystème R, sous différents formats : **gfm** pour être affiché dans des README, PDF pour être imprimé ou encore HTML pour être déployé

*INRAE | UR EABX, sebastien.boutry@inrae.fr

†INRAE | UR EABX, david.carayon@inrae.fr

sur le web. Le package propose également une structure de pipeline CI/CD pour le déploiement automatisé sur une instance Gitlab.

Les diagrammes Entité-Relation (ERD) sont produits en utilisant le package **dm** qui les génère via **DiagrammeR**. Cette fonctionnalité permet une visualisation intuitive des relations entre tables / schémas avec un degré personnalisable de précision (types, intitulés complets, commentaires, etc.)

L'ensemble des fonctionnalités sont accessibles par une unique fonction `make_dictionnaire()` qui produit le dictionnaire complet. Les fonctions `get_metadata()` et `draw_dm_erd()` permettent quant à elles de récupérer certains éléments seulement (métadonnées, diagramme ERD).

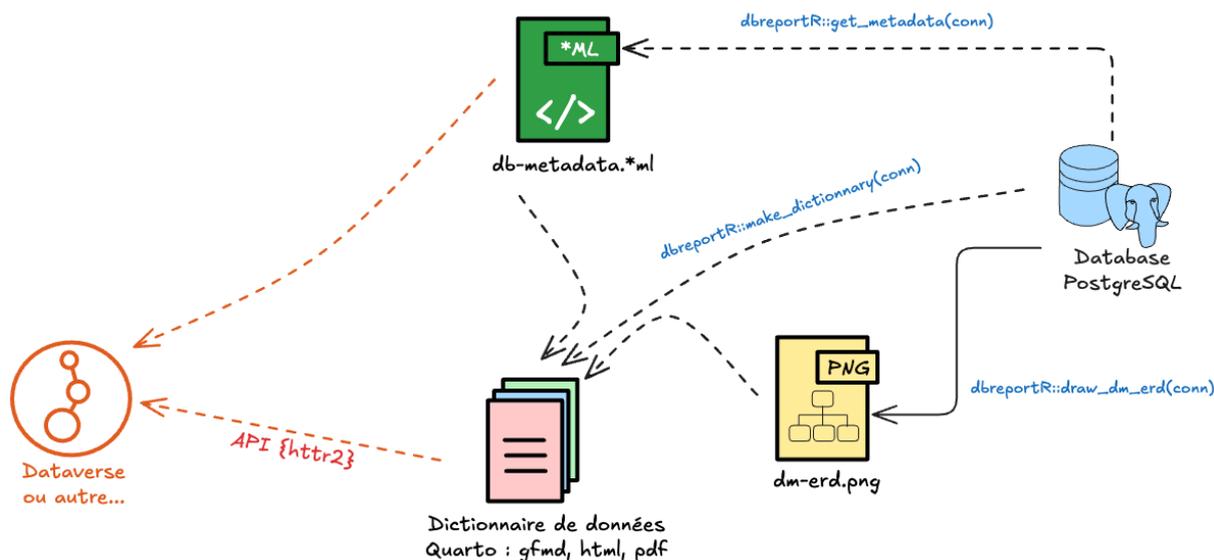


Figure 1: Fonctionnalités principales du package dbreportR

Enfin, l'export des métadonnées dans un format type "XML" garantit une interopérabilité avec d'autres plateformes de gestion des données environnementales et favorise le partage d'informations entre chercheurs. Notre utilisé de recherche traitant de questions en écologie et biodiversité, nous nous sommes concentrés sur le format EML, facilité par l'existence du package **EML** (Boettiger and Jones 2022)

Actuellement, `{dbreportR}` est compatible uniquement avec PostgreSQL. Toutefois, un portage vers d'autres systèmes de gestion de bases de données (MariaDB, SQLite) est en cours de développement. Le package est trouvable sur notre forge institutionnelle.

Références

- Allaire, JJ, and Christophe Dervieux. 2024. *Quarto: R Interface to Quarto Markdown Publishing System*. <https://github.com/quarto-dev/quarto-r>.
- Boettiger, Carl, and Matthew B. Jones. 2022. *EML: Read and Write Ecological Metadata Language Files*. <https://docs.ropensci.org/EML/>.
- Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. *Gt: Easily Create Presentation-Ready Display Tables*. <https://gt.rstudio.com>.
- Iannone, Richard, and Olivier Roy. 2024. *DiagrammeR: Graph/Network Visualization*. <https://rich-iannone.github.io/DiagrammeR/>.
- R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham, and Kirill Müller. 2024. *DBI: R Database Interface*. <https://dbi.r-dbi.org>.
- Schieferdecker, Tobias, Kirill Müller, and Darko Bergant. 2024. *Dm: Relational Data Models*. <https://dm.cynkra.com/>.

fonctionr : l'inférence statistique pour tou-te-s

Joël Girès* François Ghesquière†

fonctionr est un package R qui a pour but de faciliter l'analyse de données issues de sondages complexes, dont la manipulation est parfois laborieuse. fonctionr a été développé dans l'idée de réduire la taille du code à écrire pour produire les résultats descriptifs et d'analyses bivariées les plus courants. fonctionr a explicitement l'objectif de faciliter l'inférence statistique : outre des résultats descriptifs, le package produit des intervalles de confiance et des tests statistiques qui prennent en compte le design de l'enquête.

Le package a également été construit pour pouvoir utiliser directement les résultats produits dans un rapport. Ainsi, fonctionr produit en une seule fonction :

- Un dataframe de résultats avec les indicateurs voulus, les effectifs et leurs intervalles de confiance ;
- Un graphique reprenant ces résultats. Nous avons voulu les graphiques les plus esthétiques possibles, dans le but de pouvoir les inclure directement dans un rapport ou article (voir illustration ci-dessous). Par ailleurs, il s'agit d'objets ggplot modifiables ;
- Les différents résultats d'un test statistique lorsqu'il y en a un qui est pertinent.

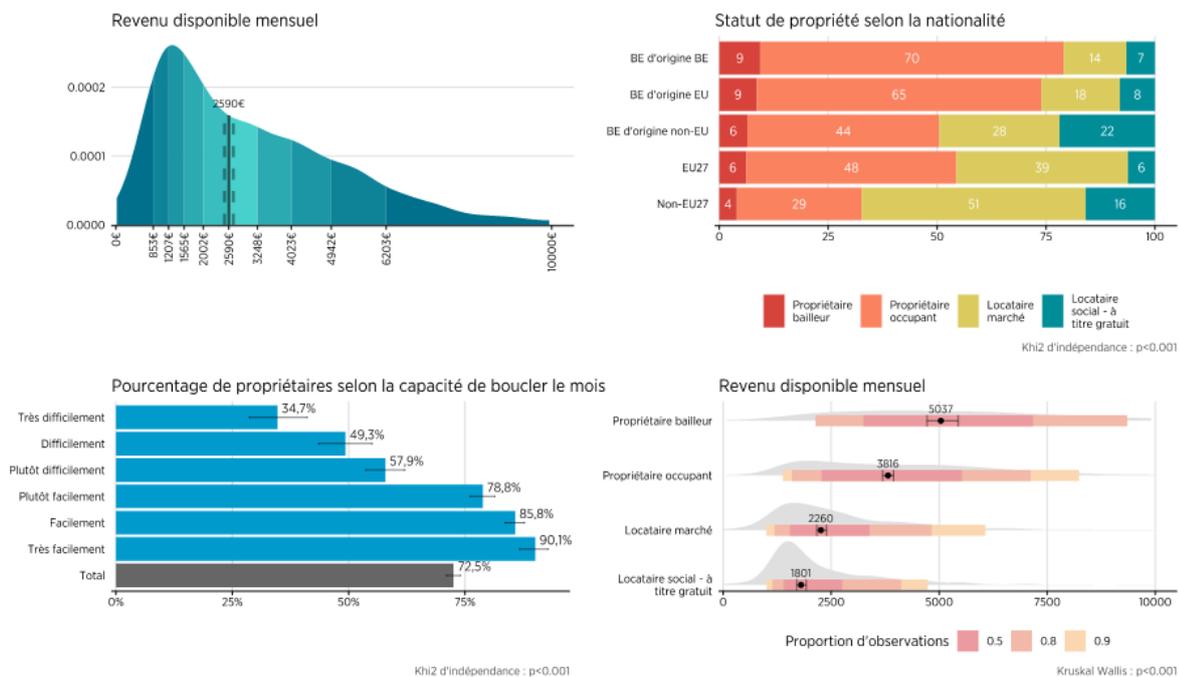


Figure 1 : Graphiques produits par fonctionr

Les fonctions permettent aussi d'exporter directement ces trois résultats dans un format Excel, dans l'idée de pouvoir facilement communiquer les résultats à une tierce personne, ou à un graphiste dans le cas du travail de mise en page d'un rapport.

Le package fonctionr et sa documentation complète sont disponibles à cette adresse : <https://jgires.github.io/fonctionr/>

Mots-clefs (3 à 5) : Inférence statistique – Graphiques – Sondage complexe – Enquêtes – Package

* Observatoire de la Santé et du Social de la Région de Bruxelles-Capitale, joel.gires@vivalis.brussels

† Institut Wallon de l'Évaluation de la Prospective et de la Statistique, f.ghesquiere@iweps.be

gaussratiovegind : un package pour analyser la distribution du ratio de deux distributions normales – Application aux images de fluorescence de chlorophylle

Pierre Santagostini* Angéline El Ghaziri† Nizar Bouhlel‡ David Rousseau§

Résumé

Il est bien connu que le rapport de deux variables aléatoires normales ne suit pas une loi normale. Par exemple, le quotient de deux variables normales centrées réduites suit une loi de Cauchy, dont la moyenne et la variance ne sont pas définies.

Cependant, les indices de végétation basés sur le rapport de deux mesures spectrales, en particulier l'indice de fluorescence de chlorophylle, sont souvent considérés comme suivant une loi normale, alors qu'ils résultent pourtant de ratios de variables gaussiennes. L'impact de cette hypothèse mérite d'être étudié. Pour cela, il est nécessaire de définir et d'estimer les paramètres de la distribution du quotient. Le cadre théorique a été développé par El Ghaziri et al. (2023).

Le package **gaussratiovegind** fournit des outils permettant de mettre en application les algorithmes développés et de manipuler ces rapports gaussiens dans R.

Mots-clés : Rapport gaussien - Estimation - Imagerie de fluorescence de chlorophylle - Epidémiologie végétale - Package

Imagerie de fluorescence de chlorophylle

À titre d'illustration, nous nous intéressons à des données de fluorescence de chlorophylle.

Il s'agit d'une technique d'imagerie utilisée pour le phénotypage des plantes. Des flashes lumineux sont envoyés sur une feuille, et la fluorescence émise est restituée par des images en niveaux de gris. Les images sont alors combinées pour obtenir des indices de fluorescence de chlorophylle (El Ghaziri et al. 2023).

On s'intéresse au maximum de fluorescence F_m et au minimum de fluorescence F_0 (voir Figure) et au rapport, connu comme le *maximum yield of photosystem II* (PSII) :

$$\frac{F_v}{F_m} = \frac{F_m - F_0}{F_m}$$

On peut vérifier que F_0 et F_m sont indépendants et normalement distribués. Donc $\frac{F_v}{F_m}$ est un rapport de deux distributions normales.

*Institut Agro, INRAE, Univ Angers, IRHS, SFR QUASAV, F-49000 Angers, France, pierre.santagostini@institut-agro.fr

†Institut Agro, INRAE, Univ Angers, IRHS, SFR QUASAV, F-49000 Angers, France, angelina.elGhaziri@institut-agro.fr

‡Institut Agro, INRAE, Univ Angers, IRHS, SFR QUASAV, F-49000 Angers, France, nizar.bouhlel@institut-agro.fr

§Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France

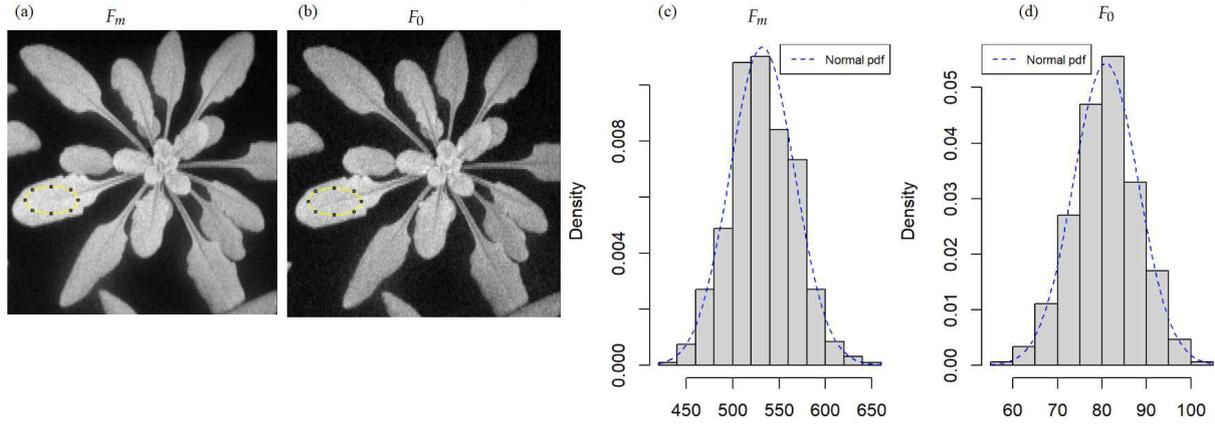


Figure 1: Images de fluorescence de chlorophylle d’*Arabidopsis thaliana* (plante saine) : (a) fluorescence maximale F_m et (b) fluorescence minimale F_0 . Les histogrammes (c,d) donnent la distribution de fréquence associée aux comptages de pixels dans la région d’intérêt dessinée par une ligne jaune continue dans (a,b), respectivement. La ligne bleue en pointillé dans des histogrammes est l’ajustement par une densité de probabilité normale (El Ghaziri et al. 2023).

Rapport gaussien

Soient $X \sim N(\mu_x, \sigma_x)$ et $Y \sim N(\mu_y, \sigma_y)$ deux variables aléatoires indépendantes, normalement distribuées. On définit les paramètres : $\beta = \frac{\mu_x}{\mu_y}$, $\rho = \frac{\sigma_y}{\sigma_x}$ et $\delta_y = \frac{\sigma_y}{\mu_y}$.

La densité de probabilité du rapport $Z = \frac{X}{Y}$ est donnée par :

$$f_Z(z; \beta, \rho, \delta_y) = \frac{\rho}{\pi(1 + \rho^2 z^2)} \exp\left(-\frac{\rho^2 \beta^2 + 1}{2\delta_y^2}\right) {}_1F_1\left(1, \frac{1}{2}; \frac{1}{2\delta_y} \frac{(1 + \beta\rho^2 z)^2}{1 + \rho^2 z^2}\right)$$

où ${}_1F_1$ est la fonction hypergéométrique confluente de Kummer, définie par : ${}_1F_1(a, b; z) = \sum_{n=0}^{+\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}$, $(\cdot)_n$ étant le symbole de Pochhammer : $(a)_n = a(a-1)\dots(a+n-1)$.

El Ghaziri et al. (2023) présente des méthodes d’estimation des paramètres β , ρ , δ_y .

Le package **gaussratiovegind** (Santagostini et al. 2025) fournit des outils pour l’analyse de ces distributions :

- **dnormratio**: fonction de densité de probabilité
- **rnormratio**: simulation d’échantillon
- **estparnormratio**: estimation des paramètres β , ρ et δ_y

À titre d’exemple, nous présenterons cela sur des images de fluorescence de chlorophylle.

Références

- El Ghaziri, A., N. Bouhlef, N. Sapoukhina, and D. Rousseau. 2023. “On the Importance of Non-Gaussianity in Chlorophyll Fluorescence Imaging.” *Remote Sensing* 15 (2): 528. <https://doi.org/10.3390/rs15020528>.
- Santagostini, P., A. El Ghaziri, N. Bouhlef, and D. Rousseau. 2025. *gaussratiovegind: Distribution of Gaussian Ratios*. <https://doi.org/10.32614/CRAN.package.gaussratiovegind>.

iteR : Un package R pour l'analyse de matrices de confusions issues d'expériences de perception phonétique.

Olivier CROUZET*

Résumé (max 300 mots)

Les travaux sur la perception de la parole peuvent parfois nécessiter l'analyse de *matrices de confusions* qui ont pour objet de servir de support à l'analyse des erreurs de perception. Une matrice de confusions est un tableau de N lignes par N colonnes qui associe les classes phonétiques présentées dans une expérience de perception (conventionnellement en lignes) avec les classes phonétiques perçues / identifiées par les participants (conventionnellement en colonnes). Chaque case du tableau indique le nombre d'observations associées à une intersection ligne / colonne (par exemple : le croisement entre la ligne /b/ et la colonne /m/ indique combien de fois les participants ont perçu /m/ lorsqu'on leur a fait écouter une séquence de parole contenant la consonne /b/). La diagonale principale de la matrice correspond aux identifications correctes alors que les autres cases correspondent aux erreurs d'identification. Mais ces confusions peuvent nous informer sur la nature des mécanismes perceptifs : si on confond le /b/ et le /m/, cette confusion pourrait impliquer une perturbation en termes de mode d'articulation ou de nasalité (occlusive orale vs. nasale) mais une bonne résistance en termes de perception de la place d'articulation et du voisement (les deux sont des bilabiales voisées). Il est alors utile de fusionner certaines classes afin de quantifier les taux de perturbation des traits spécifiques d'intérêt. Une fois la fusion de classes réalisée, il faut alors quantifier le *taux de transmission d'information* ou l'*incertitude (entropie mutuelle)* avec des méthodes de calcul adéquates. Dans cette perspective, nous proposons de présenter le travail réalisé autour d'une première version d'un package en développement (*iteR* pour *information transmission and entropy in R*) conçu pour faciliter la mise en oeuvre de ces procédures. Nous évoquerons également les extensions possibles vers d'autres problématiques en phonétique ou phonologie. **Mots-clefs (3 à 5)** : Phonétique - Perception de la parole - Matrices de confusions - Incertitude - Théorie de l'information

Références

- Christiansen, Thomas U., and Steven Greenberg. 2012. "Perceptual Confusions Among Consonants, Revisited: Cross-Spectral Integration of Phonetic-Feature Information and Consonant Recognition." *IEEE Transactions on Audio, Speech, and Language Processing* 20 (1): 147–61. <https://doi.org/10.1109/TASL.2011.2159202>.
- Christiansen, Thomas Ulrich, Torsten Dau, and Steven Greenberg. 2007. "Spectro-Temporal Processing of Speech – an Information-Theoretic Framework." In *Hearing – from Sensory Processing to Perception*. Springer. https://doi.org/10.1007/978-3-540-73009-5_55.
- Hume, Elizabeth, and Frédéric Mailhot. 2013. "The Role of Entropy and Surprisal in Phonologization and Language Change." In *Origins of Sound Change*, edited by Alan C. L. Yu, 29–48. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199573745.003.0002>.
- Iskarous, Khalil, Christine Mooshammer, Phil Hoole, Daniel Recasens, Christine H. Shadle, Elliot Saltzman, and D. H. Whalen. 2013. "The Coarticulation/Invariance Scale: Mutual Information as a Measure of Coarticulation Resistance, Motor Synergy, and Articulatory Invariance." *The Journal of the Acoustical Society of America* 134 (2): 1271–82. <https://doi.org/10.1121/1.4812855>.
- Miller, George A., and Patricia E. Nicely. 1955. "An Analysis of Perceptual Confusions Among Some English Consonants." *The Journal of the Acoustical Society of America* 27 (2): 338–52. <https://doi.org/10.1121/1.1907526>.

*Laboratoire de Linguistique de Nantes – UMR6310 – Nantes Université / CNRS, olivier.crouzet@univ-nantes.fr

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

pkgdocs: a modular R package site generator

Daan Seynaeve*

Anne-Katrin Hess†

Résumé

`pkgdocs` is a new R package to generate package documentation as markdown from an R source package. Compared to other tools, `pkgdocs` is not focused on generating a static website directly, but rather pages that can be included in a larger documentation site. A common pattern in big projects is to modularize development in several R packages. By just generating markdown and not a finished static site, combining documentation of multiple packages is made easier. `pkgdocs` was made to work well with *Hugo* and the *Docsy* theme, but the markdown output should also be useable with other markdown-based static site generators with minor changes.

Mots-clefs : Package - Documentation - Markdown - Static Site Generator

Développement

Building a documentation site with `pkgdocs` is a multi-step process:

1. “building”: generating markdown content from your package sources with `pkgdocs::build()`.
2. “publishing”: rendering the markdown content into a static web site with `hugo build`.
3. “deployment”: moving the static web assets to a location where it’s accessible by your target audience.

Between (1) and (2), the generated markdown content can be integrated with other content (including other package sites).

`pkgdocs` can be installed from the Open Analytics public R package repository:

```
install.packages("pkgdocs", repos = "https://repos.openanalytics.eu/repo/public")
```

A CRAN release is planned.

Références

*Open Analytics, daan.seynaeve@openanalytics.eu

†Open Analytics, ahess@openanalytics.eu

rix: Reproductibilité d'Environnements De Développement R

Avec Nix

Bruno Rodrigues*

Résumé

Le package **rix** propose une solution innovante pour garantir la reproductibilité des environnements de développement **R** en s'appuyant sur **Nix**, un gestionnaire de paquets spécialisé dans les builds reproductibles. Conçu pour répondre aux exigences des projets de recherche et de développement, **rix** permet de définir des environnements personnalisés en spécifiant précisément la version de **R**, les packages CRAN ou Bioconductor ainsi que l'ensemble des dépendances système (comme GDAL). Cette approche offre une reproductibilité totale, allant au-delà du simple snapshot des packages pour englober l'écosystème complet, ce qui s'avère crucial pour la fiabilité des analyses et la pérennité des résultats.

En simplifiant la création des expressions **Nix**—même à partir de fichiers `renv.lock`—**rix** rend l'utilisation de **Nix** accessible aux utilisateurs de **R**, qu'ils travaillent en mode interactif via des IDE tels que RStudio ou VS Code, ou qu'ils intègrent ces environnements dans des workflows CI/CD et des images Docker. La flexibilité offerte par **rix** se traduit par la possibilité de reproduire des environnements historiques ou de travailler avec des versions en développement de **R** et de ses packages. Ce faisant, il se positionne comme une alternative puissante aux solutions traditionnelles telles que **Docker** couplé à **renv** ou **Conda**, en proposant des environnements immuables et cohérents.

Ce poster présente le fonctionnement de **rix**, ses avantages en termes de reproductibilité et d'isolation des environnements, ainsi que des cas d'usage concrets pour des projets **R** en production et en recherche. L'objectif est de démontrer comment l'intégration de **Nix** dans l'écosystème **R** peut révolutionner la gestion des dépendances et la fiabilité des analyses dans un contexte scientifique.

Mots-clefs (3 à 5) : Package - Reproductibilité

Développement

Nix est un gestionnaire de paquets disponible pour les distributions Linux, macOS et, via WSL2, pour Windows. Il permet d'installer des logiciels de manière reproductible, en garantissant que chaque environnement est isolé et identique, indépendamment des variations du système sous-jacent. Cette approche est particulièrement utile pour les projets nécessitant des versions spécifiques de logiciels ou des configurations précises.

Pourquoi utiliser Nix avec R ?

Dans le contexte de **R**, assurer la reproductibilité implique de gérer non seulement les versions des packages **R** installés, mais aussi **R** lui-même, ainsi que les dépendances système nécessaires à certains de ces packages.

Traditionnellement, des outils comme **Docker** combinés à **renv** ou **groundhog** sont utilisés pour gérer ces aspects. Cependant, **Nix** offre une alternative en permettant de définir des environnements reproductibles qui incluent **R**, ses packages et toutes les dépendances système nécessaires par l'évaluation d'une seule expression écrite dans le langage **Nix**.

Utilisation de rix pour intégrer Nix avec R

* Ministère de la Recherche et de l'Enseignement Supérieur, bruno@brodrigues.co

Le package **rix** facilite l'utilisation de **Nix** pour les utilisateurs de **R** en simplifiant la création d'expressions Nix, qui définissent les environnements reproductibles. Avec **rix**, il est possible de :

- Générer des expressions **Nix** à partir de fichiers `renv.lock` existants.
- Travailler de manière interactive dans des environnements comme RStudio ou VS Code.
- Créer des images **Docker** contenant les packages appropriés.
- Faciliter le déploiement d'environnements logiciels en production.

Utilisation de rix pour créer des environnements reproductibles

Une fois **Nix** installé, **rix** peut être utilisé pour générer des expressions **Nix** spécifiques à un projet. Par exemple, pour créer un environnement contenant **R** avec les packages **dplyr** et **ggplot2**, la commande suivante peut être exécutée dans une session **R** :

```
rix::rix(  
  date = "2025-02-03",  
  r_pkgs = c("dplyr", "ggplot2"),  
  system_pkgs = NULL,  
  git_pkgs = NULL,  
  ide = "none",  
  project_path = ".",  
  overwrite = TRUE  
)
```

Cette commande génère un fichier `default.nix` qui définit l'environnement souhaité. Ensuite, l'environnement peut être construit et activé en utilisant les commandes **Nix** appropriées.

Avantages de l'approche Nix

L'utilisation de **Nix** présente plusieurs avantages :

- Reproductibilité complète : **Nix** garantit que chaque environnement est identique, éliminant les variations imprévues.
- Isolation des environnements : Chaque projet peut avoir son propre environnement, évitant les conflits entre différentes versions de packages ou de dépendances.
- Gestion centralisée des dépendances : **Nix** gère à la fois les packages **R** et leurs dépendances système, simplifiant la configuration des environnements.

Conclusion

L'intégration de **Nix** avec **R** via le package **rix** offre une solution puissante pour gérer des environnements reproductibles. En automatisant la configuration des dépendances et en assurant l'isolation des environnements, cette approche simplifie le développement, le déploiement et la maintenance des projets **R**.

Références

Dolstra, E. (2006). *The Purely Functional Software Deployment Model*. [Doctoral thesis 1 (Research UU / Graduation UU), Utrecht University].

shiny2docker : dockeriser une application Shiny en toute simplicité

Vincent Guyader*

Résumé

De nombreuses solutions de déploiement d'applications Shiny reposent sur l'utilisation de Docker. Toutefois, construire une image Docker adaptée est une étape essentielle qui demande des compétences spécifiques en containerisation – compétences que beaucoup de développeurs R ne maîtrisent pas toujours, et qui, même maîtrisées, restent fastidieuses à mettre en œuvre. Le package `{shiny2docker}` a été conçu pour lever cet obstacle en proposant une solution simple et robuste pour containeriser une application Shiny. En une seule commande, `{shiny2docker}` génère automatiquement un Dockerfile adapté à votre application, en s'appuyant sur `{renv}` pour capturer l'ensemble des dépendances R. Cette automatisation garantit une cohérence entre l'environnement de développement et celui de production, tout en évitant la rédaction manuelle de fichiers Docker complexes. Ainsi, grâce à `{shiny2docker}`, produire une image Docker directement déployable – que ce soit sur un serveur local ou via des solutions à grande échelle comme ShinyProxy ou Kubernetes – devient une réalité accessible à tous.

Mots-clés : Shiny - Docker - Conteneurisation - Déploiement - Reproductibilité

Développement

La dockerisation d'applications Shiny représente un véritable défi pour de nombreux développeurs, car elle exige de maîtriser des outils souvent éloignés de leurs compétences principales, tout en étant essentielle pour éviter le fameux “ça fonctionne sur ma machine”. Le package `{shiny2docker}` répond à ce problème en offrant une approche clé en main : une simple fonction R qui génère tous les fichiers nécessaires (Dockerfile, etc.) à la création de l'image Docker de l'application. L'outil prend en charge les dépendances R grâce à `{renv}`, et peut même fournir des configurations d'intégration continue (GitHub Actions, GitLab CI) pour automatiser la construction de l'image Docker.

Cette présentation illustrera l'utilisation de `{shiny2docker}` à travers un cas d'usage concrets qui montre comment passer d'un app.R en local à une image déployé en CI/CD pour ShinyProxy

En facilitant la création d'images Docker reproductibles, `{shiny2docker}` renforce la fiabilité des déploiements d'applications Shiny. Les développeurs peuvent ainsi se concentrer sur le contenu de leur application, tout en bénéficiant d'un processus de déploiement simplifié et d'environnements uniformes de bout en bout.

Références

1. Guyader V. (2025). *shiny2docker: Generate Dockerfiles for 'Shiny' Applications*. R package version 0.0.2. Disponible sur CRAN : <https://CRAN.R-project.org/package=shiny2docker>.
2. Guyader V. (2025). *shiny2docker* – code source et documentation. Dépôt GitHub : <https://github.com/VincentGuyader/shiny2docker>.

*ThinkR, vincent@thinkr.fr

{microinverterdata}, la collecte de données de votre souveraineté énergétique

Christophe Regouby*

Résumé (max 300 mots)

{microinverterdata} est un package de collecte de données locales de production d'électricité photovoltaïque. Une histoire étonnante qui part de l'idée de souveraineté énergétique : « *Je veux avoir un accès direct à mes données de production* » , et le goût de l'expérimentation des nouveautés : « *Mais comment fait-on des tests unitaires avec {httr2}* » et qui aboutit à un package à 600 téléchargements par mois sur le CRAN...

Mots-clefs (3 à 5) : Collection de données - Package - Série-temporelle

Créer un package de collecte de données

La présentation relate le parcours incrémental entre l'idée initiale et les bonnes pratiques et vise à inspirer l'audience à se lancer dans la création de packages de collecte de données. Elle porte sur l'appétit inattendu de la communauté pour ce packages (spoiler alerte : aucune explication à ce jour), sur les bonnes pratiques de développement, sur le développement en aveugle (ajoutant une pièce au concept du "hiddenverse"), et sur l'excitation de découvrir les packages récents {httr2}, {httptest2} par l'expérimentation.

Références

<https://cran.r-project.org/web/packages/microinverterdata/index.html>

*pas d'affiliation, Christophe.regouby@free.fr

~~{torch}, {tabnet} et l'apprentissage profond par l'usage~~

Christophe Regouby*

Résumé du tutoriel

R offre une interface native à l'environnement d'apprentissage profond torch. Et au-delà, au mlverse, un univers très complet de paquetages pour adapter les données tabulaires, les images, les textes, les modèles publics du hub huggingface, et bien d'autres.

Venez découvrir cet univers par l'usage, par un atelier pratique qui couvrira tout d'abord les généralités de torch avec les avantages et les inconvénients de cette librairie, incluant les arcanes de l'installation sur vos GPUs, la manipulation de tenseurs, etc.

On entrera en profondeur dans les modèles {tabnet} de classification ou de régression sur des données tabulaires, leur entraînement et leur optimisation, l'intégration dans l'environnement {tidymodels}, la gestion des valeurs manquantes...

On basculera ensuite dans le monde des modèles de langage avec le réentraînement de GPT-2 sur votre machine, l'usage et la modification des grands modèles de langage d'huggingface, le fine-tuning d'un LLM par l'ajout d'un LoRA.

On évoquera enfin le monde de la vision par ordinateur avec l'entraînement d'un modèle de segmentation d'image.

L'atelier inclut la théorie strictement nécessaire et beaucoup de pratique pour apprendre à tirer le meilleur parti des ressources matérielles disponibles tout en utilisant des modèles à l'état de l'art.

Prérequis : maîtrise intermédiaire de la programmation avec le langage R, et du machine-learning.

Durée : 2h

À réaliser avant de venir au tutoriel :

- Installer R 4.4.0 ou supérieur et un IDE confortable, par exemple RStudio desktop.
- Installer le package {usethis} : `install.packages("usethis")`
- Installer les exercices du tutoriel : `usethis::use_course("cregouby/Tutoriel_torch")`
- Installer les prérequis du tutoriel : `source("00_installation.R")`

Informations complémentaires :

- site web de {torch} : <https://torch.mlverse.org/>

* Airbus, christophe.regouby@airbus.com

Liste des participants

- Armand Patrice
- Aubert Julie
- Audinot Timothée
- Bacher Etienne
- Badiabo Ngongo Sandra Chancelvie
- Barbier Joseph
- Barthelemy Tanguy
- Bauer Hélène
- Bayon-Vicente Guillaume
- Beaufort Maël
- Benrezkallah Jordan
- Berger Vérane
- Bichat Antoine
- Boiveau Adrien
- Bonnand Samuel
- Boutry Sébastien
- Bréant Arthur
- Brehelin Jacques
- Cadoret Valentin
- Carayon David
- Cartelier Kévin
- Casse Sidy Mouhamed
- Chaltiel Dan
- Chatain Victoire
- Coquelin Anaëlle

- Cotton Thomas
- Courtade-Gaïani Sophie
- Crouzet Olivier
- De Koninck Tobia
- Deghorain Pauline
- Delacuvellerie Alice
- Della-Vedova Claire
- Delmotte Murielle
- Delvaux Véronique
- Demeure Mélanie
- Dervieux Christophe
- Drouilhet Remy
- Dirk Eddelbuettel
- Engels Guyliann
- Fabri Antoine
- Fagniard Sophie
- Fay Colin
- Flores Castillo Nicolas
- Galiana Lino
- Ghesquière François
- Gillard Malo
- Girard Antoine
- Girès Joël
- Godet Julien
- Goeseels Eva
- Goumri Samra
- Grosjean Philippe
- Gruson Hugo
- Guyader Vincent
- Henry Lionel
- Herbepin Emmanuel
- Hernando Isabelle

- Hocking Toby
- Hocquette Delphine
- Houée-Bigot Magalie
- Huet Kathy
- Husson François
- Ibrahimi Nusaibah
- Ihaddaden Mohamed El Fodil
- Julie Descamps
- Janeczek Sofiane
- Jean-Louis Eric
- Jonas Van Malder
- Kiener Patrice
- Kostyrka Andreï V.
- Lamarche Pierre
- Lamot Thomas
- Latreille Julie
- Le Guen Nelly
- Lechien Alain
- Legris Maxime
- Lionakis Lucianos
- Mahmoud Remi
- Mauger Emmanuelle
- Meyer Fanny
- Meyers Charlene
- Michel Philippe
- Millien Eléa
- Mohamed Saleh Amandine
- N'dala-Louika Isis
- Neels Nicolas
- Nguyen Viet Chau Linh
- Nopère Martin
- Ollivier Fanny

- Ozturk Dilhan
- Paul Lakin
- Panaiotis Thelma
- Périlleux Hugo
- Perrier Victor
- Piccaluga Myriam
- Pierotti Livia
- Plume Sara
- Pozzi Tiffani
- Prunet Caroline
- Quemar Thibaud
- Rapicault Clémence
- Regouby Christophe
- Remiche Marie-Ange
- Rey Jean-François
- Ricour Florian
- Rodberg Aubin
- Rodrigues Coelho Bruno André
- Roland Virginie
- Saadani Moetez
- Salmon Maëlle
- Santagostini Pierre
- Saracco Jerome
- Seynaeve Daan
- Siberchicot Aurélie
- Slimane Regui
- Sogalow Lise
- Spanu Laurent
- Stamm Aymeric
- Terny Lola
- Topart Morgane
- Ubertalli Maxence

- Ursache Ovidiu
- Van Holle Lionel
- Vanderplanck Maryse
- Vanwindekens Frédéric
- Visée Corentin
- Viseur Robert
- Vroylandt Thomas
- Wijffels Jan

Liste des auteurs

- Akharif Abdelhadi, 42, 43
Armand Patrice, 17, 18
- Bacher Etienne, 33, 34
Barthelemy Tanguy, 55, 56
Bellini Saibene Yanina, 10, 11
Benrezkallah Jordan, 22
Bichat Antoine, 20, 35
Bouhleb Nizar, 85, 86
Boutry Sébastien, 82, 83
Brehelin Jacques, 13
Bréant Arthur, 68
Burkhard Pascal, 10, 11
- Campitelli Elio, 10, 11
Carayon David, 58, 59, 82, 83
Castelle Bruno, 58, 59
Chaltiel Dan, 25, 73, 78
Chatain Victoire, 69, 70
Chavent Marie, 21
Coquelin Anaëlle, 5, 8, 9
Corrales Paola, 10, 11
Courtade-Gaïani Sophie, 35
Crouzet Olivier, 50, 87, 88
- Darmendrail Véronique, 21
De Koninck Tobia, 63
Dehez Geoffrey, 58, 59
Delacuvellerie Alice, 48
Della-Vedova Claire, 3, 26
Delmotte Murielle, 74
Demeure Mélanie, 79, 80
Dervieux Christophe, 44, 45
- Eddelbuettel Dirk, 47
El Ghaziri Angelina, 85, 86
Engels Guyliann, 40, 41
- Fabri Antoine, 51
Fay Colin, 12, 60, 61
Féral Delphine, 21
- Galiana Lino, 28
Ghanem Hiba, 76, 77
Ghesquière François, 84
- Girardin Valerie, 13
Girès Joël, 46, 84
Gomez Alice, 76, 77
Grente Théo, 13
Grosjean Philippe, 17, 18, 40, 41
Gruson Hugo, 36, 37
Guyader Vincent, 12, 64, 65, 92
- Henry Lionel, 2
Herbepin Emmanuel, 69, 70
Hess Anne-Katrin, 89
Hocking Toby Dylan, 57, 75
- Ibrahimi Nusaibah, 73
Ihaddaden Mohamed El Fodil, 14
- Jachiet Pierre-Alain, 8, 9
- Kiener Patrice, 23, 29
Kostyrka Andreï V., 30, 31
Kowandy Véronique, 79, 80
- Lamarche Pierre, 19
Le Guen Nelly, 5, 8, 9
Legris Maxime, 38, 39
Lionakis Lucianos, 62
Lê Sébastien, 66, 67
- Mahmoud Remi, 66, 67
Mallet Christophe, 4
Mauger Emmanuelle, 79, 80
Mazza Stéphanie, 76, 77
Mellouk Amal, 42, 43
Morin Sandrine, 5, 8, 9
- Neels Nicolas, 4
Nguyen Viet Chau Linh, 76, 77
Niquil Nathalie, 13
- Pierotti Livia, 73
Pourtier Frédéric, 21
Prunet Caroline, 5, 8, 9
Pulcrano Anne-Sophie, 35
Périlleux Hugo, 32, 46
- Regnault Philippe, 13

Regouby Christophe, 27, 93, 94
Rey Amandine Eve, 76, 77
Rodrigues Coelho Bruno André, 90, 91
Ronceux Aïcha, 24
Rousseau David, 85, 86

Salaün Alwenna, 5, 8, 9
Salmon Maëlle, 10, 11
Santagostini Pierre, 85, 86
Saracco Jerome, 21
Seynaeve Daan, 89
Sitruk Anaïs, 5, 8, 9
Slimane Regui, 42, 43
Solomiac Agnès, 5, 8, 9
Spanu Laurent, 12
Stamm Aymeric, 52

Topart Morgane, 24
Tougeron Kevin, 4

Van Holle Lionel, 7
Van Malder Jonas, 53, 54
Vanderplanck Maryse, 49
Vanwindekens Frédéric, 71, 72
Verbeke Tobias, 63
Viseur Robert, 6
Visée Corentin, 15, 16
Vroylandt Thomas, 69, 70

Wattiez Ruddy, 48
Wijffels Jan, 81