

# David CARAYON

Ingénieur d'études · Statisticien · Data Scientist ·  ·  ·  ·  ·  · 

8 Lotissement L'entrada 33650 CABANAC-ET-VILLAGRAINS

 (+33) 6 64 66 90 60  david.carayon@inrae.fr  <https://dcarayon.fr>  davidcarayon  David CARAYON

## Expériences

### Statisticien

Bordeaux, France

Unité de recherche ETTIS (Fonctionnaire)

2019 - Present

- Pilotage de pipelines de données pour de nombreux projets de recherches, allant de la collecte de la donnée au rapport final en passant par sa gestion, son nettoyage et son traitement statistique (EDA, modélisation, analyses multivariées, machine learning).
- Transformation de données brutes en informations utiles et solutions interactives par le biais de packages, d'applications et de documentations associées.
- Appui méthodologique au personnel temporaire (revues de code, stratégie d'analyse de données, statistiques) et animation de formations internes aux bonnes pratiques (Programmation, Open Source, Open Science)

### Biostatisticien

Bordeaux, France

Unité de recherche EABX (CDD)

2017-2019

- Manipulation de larges jeux de données (suivi de qualité écologique des cours d'eau - France entière sur 10 ans) et mise en place d'algorithmes de traitement adaptés
- Valorisation des résultats via des packages R (calculs d'indicateurs), des cartes interactives (Leaflet) et par la rédaction de rapports techniques, d'articles scientifiques et de présentations lors de colloques internationaux

## Compétences

**Programmation** R, Python, SQL, Git, Github Actions, Gitlab CI/CD, Docker, ODK

**Frameworks** Tidyverse, Shiny, data.table, gt, targets, renv, sf

**Software** VSCode + Quarto, Rstudio, Jupyter, Zotero, Suite office, QGIS

**Langages parlés** Anglais (bilingue), Espagnol (scolaire)

## Sélection de projets

### Data.Interventions

 <https://projet-swym.fr/data.interventions>

ETL et Machine learning pour la prédiction des noyades

2023-Present

- Déploiement d'une solution de collecte numérique de données issues d'interventions de sauveteurs dans les Landes via Open Data Kit (ODK) dans une BDD PostgreSQL (Datacenter INRAE)
- Développement d'un pipeline ETL complet collectant la donnée puis la mettant en forme dans une BDD SQLite
- Développement d'une application R Shiny pour le reporting automatisé de l'activité des sauveteurs
- Entraînement en parallèle d'un modèle de machine learning (xGBoost) pour prédire les journées à haut risque

### IDEATools

 [davidcarayon/IDEATools](https://github.com/davidcarayon/IDEATools)

Indicateurs de Durabilité des Exploitations Agricoles (IDEA4)

2019 - Present

- Développement du package R rendant la méthode opérationnelle et de l'interface web (Shiny) associée

### Shiny Kubernetes Service (SK8)

 <https://sk8.inrae.fr>

Service institutionnel pour le déploiement scalable d'applications Shiny

2022 - Present

- Conceptions d'applications Shiny avec interactions API Gitlab, manipulation de containers (Docker) via CI/CD

## Formation et Certifications

### Certification Datacamp

En ligne

Data Scientist with Python and R

2024

### Université de Bordeaux

Bordeaux

Master Biodiversité et Suivis Environnementaux

2015-2017