

David CARAYON

Statistician · 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

Experiences

Statistician

Bordeaux, France

ETTIS Research Unit

2019 - Present

- Managing data pipelines for numerous research projects, ranging from data collection to final reporting, including data management, cleaning, and statistical processing (EDA, modeling, multivariate analysis, machine learning).
- Transforming raw data into useful information and interactive solutions through packages, applications, and associated documentation.
- Providing methodological support to temporary staff (code reviews, data analysis strategy, statistics) and conducting internal training on best practices (Programming, Open Source, Open Science).

Biostatistician

Bordeaux, France

EABX Research Unit (Fixed-term contract)

2017-2019

- Handling large datasets (monitoring the ecological quality of rivers across France over 10 years) and implementing appropriate processing algorithms.
- Promoting results through R packages (indicator calculations), interactive maps (Leaflet), and writing technical reports, scientific articles, and presentations at international conferences.

Skills

Programming 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, Office Suite, QGIS

Languages English (bilingual), Spanish (basic)

Project Selection

Data.Interventions

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

ETL and Machine learning for drowning prediction

2023-Present

- Deploying a digital data collection solution from lifeguard interventions in Landes via Open Data Kit (ODK) into a PostgreSQL database (INRAE Datacenter).
- Developing a complete ETL pipeline collecting and shaping data in a SQLite database.
- Developing an R Shiny application for automated reporting of lifeguard activities.
- Training a parallel machine learning model (xGBoost) to predict high-risk days.

IDEATools

 davidcarayon/IDEATools

Sustainability Indicators for Agricultural Farms (IDEA4)

2019 - Present

- Developing the R package making the method operational and the associated web interface (Shiny).

Shiny Kubernetes Service (SK8)

 <https://sk8.inrae.fr>

Institutional service for scalable deployment of Shiny applications

2022 - Present

- Designing Shiny applications with Gitlab API interactions, handling containers (Docker) via CI/CD.

Education and Certifications

Datacamp Certification

Online

Data Scientist with Python and R

2024

University of Bordeaux

Bordeaux

Master in Biodiversity and Environmental Monitoring

2015-2017