Fargo, ND
didier.statistician@gmail.com
didiermurillo.com
linkedin.com/in/damurillof
github.com/DidierMurilloF

# Didier Murillo

## Agricultural Data Scientist

### SUMMARY

Agricultural Data Scientist who builds statistical and machine learning models for genomic and phenomic prediction in plant breeding. Expert in phenotypic analysis of field trials, from experimental design to linear mixed models with spatial adjustment for field variation. I turn models into tools that research teams use: R packages (FielDHub, on CRAN), R Shiny and React applications, REST APIs, PostgreSQL databases and Docker deployments. I also bring production experience from a regulated industry, building tested R Shiny applications for clinical trial data review as a consultant to Daiichi Sankyo.

### EXPERIENCE

**Research Statistician Developer** Feb 2022 – Present
North Dakota State University, Ag Data Analytics · Fargo, ND

- Build statistical and machine learning models for genomic and phenomic prediction workflows in plant breeding.
- Lead phenotypic analysis of field trials, from experimental design to linear mixed models with spatial adjustment and genotype rankings.
- Process satellite imagery for analysis of field trials and agricultural data.
- Develop R Shiny, React and JavaScript applications backed by REST APIs and PostgreSQL databases, deployed with Docker.
- Develop R packages for web API integration.
- Mentor junior statisticians in agricultural statistics and Shiny development, and lead workshops and conference talks on experimental design and statistical analysis in agriculture.

**R Shiny Developer (Consultant)** Oct 2023 – Dec 2025
Daiichi Sankyo · Remote

- Designed, built and maintained R Shiny applications for clinical trial data review from CDISC SDTM/ADaM data, with traceable data handling and reproducible outputs.
- Implemented oncology tumor response visualizations (waterfall, spider and swimmer plots) within the review workflows.
- Owned the full delivery lifecycle in a regulated environment: modular architecture, testing, debugging across data, UI and reactivity layers, and deployment support.

**Data Scientist (Contract)** Jan 2024 – Dec 2024
CodeStack Systems · Remote

- Built the data backbone of a satellite crop-monitoring app: Python services in Docker that ingest daily satellite imagery, vegetation indices (NDVI, NDRE, NDMI, RECI, MSAVI) and 10-day weather forecasts for each field from the EOS satellite API, into Google Cloud Storage and PostgreSQL.
- Built a service that onboards newly registered fields automatically, syncing field boundaries from MySQL and backfilling their vegetation indices.
- Developed an R Plumber REST API that forecasts vegetation indices per field from crop growth curves (polynomial regression on days since planting), and alerts growers to possible crop stress in corn and soybeans using crop-specific thresholds.
- Prototyped ARIMA time-series models for NDVI forecasting, and wrote the system and data-pipeline documentation.

**R Shiny Developer** Oct 2019 – Jan 2022

North Dakota State University · Fargo, ND

- Created FielDHub, an R package and Shiny app on CRAN for randomizing complex field experiments in plant breeding, agronomy, forestry, animal science and biology, and published it in a peer-reviewed paper.
- Led 12+ statistical projects spanning experimental design, mixed models, machine learning and data governance.
- Automated time-consuming analysis tasks with Shiny apps and JMP add-ins.

**Earlier experience**

**Statistician Assistant**, University of Puerto Rico, Mayagüez · 2017 – 2018

Generalized linear and mixed model analyses in R and SAS for educational research data.

**Statistician**, National Army Healthcare System, Bogotá · 2014 – 2015

Automated processing and analysis of medical information system data with R and Excel.

## TECHNICAL SKILLS

| | |
|---|---|
| **Statistics & modeling** | Experimental design · Linear mixed models (incl. spatial) · GAMs · Genomic & phenomic prediction · Machine learning · Time-series & growth-curve forecasting |
| **Agricultural data** | Field trial phenotypic analysis · Plant breeding · Satellite image processing · Vegetation indices (NDVI, NDRE, NDMI) |
| **Software** | R (Shiny, Golem, Plumber, tidyverse, ggplot2, plotly) · Python · R package development · React · JavaScript · REST APIs |
| **Data & deployment** | PostgreSQL · MySQL · BigQuery · Google Cloud Storage · Docker · AWS · Git |
| **Also** | ASReml · SAS · JMP · CDISC SDTM/ADaM |

## OPEN SOURCE

**FielDHub** (Developer): Shiny app and R package for design of experiments in life sciences (CRAN).

**FIELDimageR** (Contributor): Analysis of orthomosaic images from agricultural field trials.

**MrBean** (Contributor): Shiny app for large-scale plant breeding trials with linear mixed models.

**ASRgwas** (Developer): Preparing and auditing phenotypic and genomic data.

## WRITING

Likelihood vs Prediction Accuracy in Mixed Models

Smoothers in the Field: GAMs for Spatial Trends in Plant Breeding Trials

## EDUCATION

**M.S. in Statistics**, University of Puerto Rico, Mayagüez · 2019

**B.S. in Mathematics**, University of Tolima, Ibagué, Colombia · 2014

## LANGUAGES

Spanish (native) · English (advanced)