

Guilherme Vituri

Email: vituri.vituri@gmail.com
GitHub: github.com/vituri

PROFESSIONAL SUMMARY

Data scientist with a strong background in **R, Julia, and mathematics**. Experienced in delivering quick data solutions and contributing to open-source and academic communities. Skilled with interactive dashboards in R/Shiny, data pipelines, machine learning models, and scientific research. Publishes on Topological Data Analysis (TDA) and geospatial analytics.

EDUCATION

PhD in Topological Data Analysis

Universidade Estadual Paulista (Unesp) / Ohio State University (OSU)

2016 – 2020

- Research on generalizing Vietoris-Rips and Cech constructions using motifs, resulting in new clustering methods for graphs/networks.
- Six-month research stay at OSU under Prof. [Facundo Mémoli](#).
- Dissertation: [Motivic constructions on graphs and networks with stability results](#).

Master's Degree – Algebraic Topology

Universidade Estadual Paulista (Unesp)

2014 – 2015

- Studied the Gottlieb group, a subgroup of the fundamental group.
- Dissertation: [Sobre os grupos de Gottlieb](#) (in Portuguese).

Bachelor's in Mathematics

Universidade Estadual Paulista (Unesp)

2010 – 2013

WORK EXPERIENCE

CONSULTING

I offer consulting on R/Shiny, data science, development and machine learning in general. Some projects I've worked:

- Geospatial analysis: helping with code and technical writing of books and papers.
- API creation with R and plumber, database modeling, authorization methods and LLMs.
- Dashboard creation using R and Shiny for market research companies.

SENIOR DATA SCIENTIST

[Cielo \(Remote\)](#)

September 2025 – Present

Cielo is a major Brazilian financial technology company, specializing in electronic payment processing and providing a diverse range of e-commerce solutions.

- Created and refined machine learning models with Python on big data to predict failures, using tree models and neural networks on noisy time series; presented results to stakeholders with pros-and-cons.
- Automated reports on Databricks to alert clients about needed repairs on their terminals.

R/SHINY ENGINEER

[Appsilon \(Remote\)](#)

Feb 2025 – July 2025

Appsilon provides data science services to the biggest pharma companies in the world, focusing on R/Shiny and quick development.

- Contributed to [open source packages](#) and collaborated with Posit to feature applications in the Connect Gallery.
- Delivered a proof-of-concept using AI/LLMs and R/Shiny, automating clinical trial data extraction and reducing processing time from weeks to hours.
- Developed and maintained R/Shiny applications using Rhino; fixed issues and reviewed PRs for other open-source packages.
- Contributed to internal knowledge bases and authored [blog posts](#).

HEAD OF INTELLIGENCE

[Argus Solutions](#)

Jan 2020 – Jan 2025

Argus provides solutions for fatigue and distraction detection, control towers, and telemetry in transport operations around the world.

- Founded and led the Data team, growing it to over 12 members, including developers, data scientists, engineers, and mathematicians. Managed hiring, mentoring, and professional development for the team.
- Responsible for creating, testing and deploying 30+ R/Shiny applications for the entire company and clients, used by 300+ different users daily and 20+ users simultaneously. Some of these apps are used by the control tower 24/7 to analyze photos and videos.
- Automated the pipeline of data from hundreds of Excel files to a MariaDB database on AWS RDS, which has 10+ million new entries daily.
- Automated daily/weekly reporting with RMarkdown and Sendgrid, sending 10k+ emails monthly and saving over 20 hours of manual work daily.
- Developed a machine learning model to [predict driver drowsiness](#) with >80% accuracy on noisy data, using ensemble methods with [tidymodels](#).
- Created a high-performance Julia webserver using [Oxygen.jl](#), reducing data ingestion time from four minutes (with R) to five seconds (with Julia) per iteration.
- Led computer vision projects (cellphone detection, drowsiness, pothole, and gesture detection) using Keras and YOLO in Python.

TECHNICAL SKILLS

- **R:** Advanced – tidyverse, shiny, geospatial, APIs, package development, optimization
- **Julia:** Advanced – TDA, performance, APIs, parallelism, documentation
- **SQL:** Advanced – MariaDB/Postgres, database design, optimization (indexation, normalization), dbplyr within R
- **AWS:** Intermediate – EC2, S3, RDS, ECS, Docker, Rekognition
- **Python:** Intermediate – numpy, pandas, polars, scikit-learn
- **Databricks:** Beginner – delta lake, notebooks, PySpark

- **Data Science:** Machine learning, computer vision, reporting, dashboards
- **Technical Writing:** Workshops, blog posts, and academic papers

SELECTED PROJECTS & OPEN SOURCE

- **JuliaTDA Organization:** Owner and main contributor. Wrote the [Mapper and Ball Mapper](#) and [ToMATo](#) implementations in Julia.
- **TidierOrg:** Contributor to [TidierIteration.jl](#), a Julia version of R's purrr package.
- **QuartoDocBuilder.jl:** Created a package to facilitate Julia documentation generation with Quarto ([docs](#)).
- **Blog:** [julia-for-r-users](#) and other posts on R, Julia, and data science. The online version of this CV is [here](#).

PUBLICATIONS, TALKS & WORKSHOPS

- **Paper:** [Remote sensing to quantify potential aquifer recharge as a complementary tool for groundwater monitoring](#) (Hydrological Sciences Journal, 2024). Co-author; led the R/terra analysis and data pipeline.
- **Workshop:** [Topological Data Analysis workshop](#) (at [XXIII Brazilian Topology Meeting](#), 2024), using Julia.
- **Talk:** [Topology meets the real world: how geometry can help us analyze finite metric spaces](#) (at [Workshop of Algebraic Topology and Applications](#), 2023).
- **Paper:** [Motivic clustering schemes for directed graphs](#) (with Facundo Mémoli, arXiv, 2020).

ADDITIONAL INFORMATION

- **Languages:** Portuguese (native), English (fluent), Basic Italian and Russian.
- **Writing:** Author of the forthcoming book, [Topological Data Analysis with Julia](#).
- **Interests:** R and Julia, open source, topological data analysis, algorithms and performance, technical writing, scientific communication, and mentoring.