

EDUCATION

- **PhD in Biological Sciences.**
FCEyN, University of Buenos Aires. Buenos Aires, Argentina. 2010-2015
US equivalency: PhD in Biological Sciences.
- **Licentiate in Biological Sciences.** Area of concentration: genetics and evolution.
US equivalence: MS in Biological Sciences.
FCEyN, University of Buenos Aires. Buenos Aires, Argentina. 2003-2010

EXPERIENCE

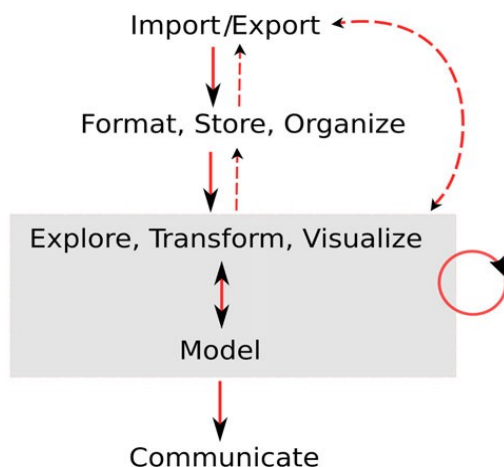
- **Data scientist.** Softtek. Buenos Aires, Argentina. 2019-current (7 months)
Performed tasks: Analysis of large scale data using Azure and Databricks + Pyspark/Sparklyr. Development of Machine-Learning models with Tensorflow/Keras and scikit-Learn. Time series forecasting. Anomaly detection. Computer vision. Development of software with R, Python and SQL. Sentiment analysis. Natural Language Processing. Regular expressions, Modeling of categorical/continuous variables, neural networks, random forests, CART and generalized additive models, support vector machines, among others. Power BI. Rmarkdown. Data visualization with ggplot2, matplotlib, Seaborn and Plotly.
- **Researcher/Data scientist.** Elson S. Floyd College of Medicine, Washington State University. Spokane, WA. 2018-2019
Performed tasks: Analysis and interpretation of biological data. Data manipulation with R, Python, Bash and C++. Cleaning, modeling and visualization (diverse statistical and ML models, as multivariate models, linear mixed models, generalized additive models, random forests, regularized regression and different architectures of multilayer neural networks). Development of software packages. Rmarkdown. Shiny.
- **Researcher/Data scientist.** IIB-INTECH, National University of San Martin. Buenos Aires, Argentina. 2016-2018
Performed tasks: Analysis and interpretation of biological data. Data manipulation using R, Python, Bash, C++, and SQL. Cleaning, modeling and visualization (diverse analytical methods, as multivariate methods, use of advanced graphical tools as ggplot2, matplotlib, Seaborn, Plotly, etc.). Statistical design. Development of software packages to structure and make reproducible the already refined analyses. Development of a graphical application for efficient data manipulation and analysis using Shiny.
- **Lecturer, Advanced Biostatistics.** Faculty of Medical Sciences, Favaloro University. Buenos Aires, Argentina. 2017
- **Teaching Assistant, Human Genetics.** IIB-INTECH, National University of San Martin. Buenos Aires, Argentina. 2016-2017
- **Teaching Assistant, Introduction to the Analysis of Genetic Data with R.**
FCEyN, University of Buenos Aires. Buenos Aires, Argentina. 2012, 2014

DATA SCIENCE SKILLS

- **Programming languages:** R, Shell, C++, Python, SQL, HTML/CSS/JavaScript.
- Software development, high-performance computing, parallel-computing, Git, Unix/Linux, Azure, Databricks, Spark.
- Software packaging and distribution, unit testing, continuous integration (Travis CI, AppVeyor), code coverage testing (Codecov), code documentation.
- Strong background in statistics and machine-learning.

A toolbox for the whole data analysis cycle

- 1) **Data wrangling:** Using dplyr and Pandas, or custom R or Python code
- 2) **Data structuring and querying:** SQL, HDF5
- 3) **Data exploration via multivariate algorithms:** K-means, K-medoids, fuzzy-clustering, Principal Component Analysis (PCA), Redundancy Analysis (RDA), t-SNE, hierarchical clustering
- 4) **Data exploration via visualization methods:** Box and violin-plots, bar-plots, network-visualization, charts, word-clouds, heatmaps, etc. Use of visualization packages as ggplot2, matplotlib, Seaborn, Plotly
- 5) **Data modeling:** Frequentist statistics, Bayesian statistics (Stan, PyMC3, selection of statistical distributions), machine-learning (Tensorflow/Keras + scikit-learn and R: regularized regression, random-forests, neural networks, as multilevel models, RNN, LSTM, convolutional neural networks, etc., cross-validation (K-fold, bootstrap, etc), ANOVA, MANOVA, linear models, linear mixed models, Generalized Linear Models (GLMs), Generalized Linear Mixed Models (GLMMs), Discriminant Analysis (DA), text-mining, regular expressions, NLP, sentiment analysis
- 6) **Technical reports and dashboards:** Interactive and static reports using R Markdown, Shiny and Power BI

**OPEN SOURCE SOFTWARE IN SCIENTIFIC REPOSITORIES**

- **chunkR:** Read Tables in Chunks. R / C++.
Website: <https://CRAN.R-project.org/package=chunkR>
- **FastqCleaner:** A Shiny web-app for cleaning FASTQ files with R. R/ HTML + JavaScript + CSS/ C++.
Website: <https://bioconductor.org/packages/FastqCleaner>.
- **EcoGenetics:** Spatial Analysis of Phenotypic, Genotypic and Environmental Data. R.
Website: <https://CRAN.R-project.org/package=EcoGenetics>, Tutorial: <https://leandroroser.github.io/EcoGenetics-Tutorial>

PUBLICATIONS

- **Roser L**, Agüero F, Sánchez D (2019). FastqCleaner: an interactive Bioconductor application for quality-control, filtering and trimming of FASTQ files. *BMC Bioinformatics*
- Ingiosi A, Wintler T, Schoch H, Singletary K, Righelli D, **Roser L**, Risso D, Frank M, Peixoto L (2019). Shank3 Modulates Sleep and Expression of Circadian Transcription Factors, *eLife* 8:e42819
- Bañuelos C, Levy G, Nittolo A, **Roser L**, Tekiel V, Sánchez D. The Trypanosoma brucei RNA-Binding Protein Tb RRM 1 is Involved in the Transcription of a Subset of RNA Pol II-Dependent Genes. *Journal of Eukaryotic Microbiology*. DOI: 10.1111/jeu.12716
- **Roser L**, Ferreyra L, Saidman B, Vilardi J (2017). EcoGenetics: an R package for the management and exploratory analysis of spatial data in landscape genetics. *Molecular Ecology Resources*, 17:e241-e250
- **Roser L**, Ferreyra L, Ewens M, Vilardi J; Saidman B (2017). Isolation by distance and stochastic migration in a fragmented population of *Prosopis alba*. *American journal of botany*, 104:313-321
- **Roser L**, Ferreyra L, Ewens M, Vilardi J; Saidman B (2014). Genetic and morphometric characterization of clones of *Prosopis alba*, Algarobia, selected for salt tolerance. *Tree genetics & genomes*, 10:383-397

SELECTED COURSES

• Data & AI track. Microsoft Partner Experience Tour, Argentina.	2019
• Human genomic analysis for precision medicine with GATK4 Best Practices and FireCloud. Viña del Mar, Chile.	2018
• Introduction to remote sensing and its application in geology. FCEyN-UBA. Buenos Aires, Argentina.	2013
• Bayesian statistics applied to neuroscience and molecular and cellular biology. FCEyN-UBA. Buenos Aires, Argentina.	2012
• Mathematical Analysis I (course from the Licentiate in mathematics, multivariate mathematical analysis). FCEyN-UBA. Buenos Aires, Argentina.	2011
• Introduction to the analysis of genetic data with R. FCEyN-UBA. Buenos Aires, Argentina.	2011
• Algebra I (course from the Licentiate in mathematics). FCEyN-UBA. Buenos Aires, Argentina.	2010

LANGUAGES

- Spanish (fluent), English (advanced), German (novice, Goethe-Institut A1 and A2 levels).