

Lucas Cardozo

Data Scientist

✉ lucasecardozo@gmail.com **in** [linkedin.com/in/lucasecardozo](https://www.linkedin.com/in/lucasecardozo)

 github.com/lecardozo  lecardozo.github.io

Data scientist with background in bioinformatics and experience in high throughput genomic and gene expression data pipelines, analyses and visualization.

Experience

Research Assistant

Computational Systems Biology Laboratory - INOVA USP

- Analysed genomic and transcriptomic data in the context of infectious diseases and vaccines.
- Built bioinformatics data pipelines and tools to standardize common analyses.
- Managed a four node cluster used by bioinformaticians and data scientists.

Education

University of Sao Paulo, M.Sc. in Pharmacy

Idealized and implemented the whole data pipeline (collection, transformation, analysis and visualization) of the Sickness Positioning System (<https://sipos.fcf.usp.br>) to help epidemiologists make better data driven decisions when creating public policies for malaria vector control.

University of Sao Paulo, B.Sc. in Biological Sciences

Analysed gene expression data in order to understand the molecular mechanisms behind the immune protection acquired with the radiation-attenuated cercariae vaccine in mice. Co-created the CEMiTool R package, for gene co-expression data analysis and created the webCEMiTool.

Side Projects

rworker - Celery worker for R tasks (github.com/lecardozo/rworker)

rabbitr - RabbitMQ client for R (github.com/lecardozo/rabbitr)

bioboard - Interactive dashboard for bioinformaticians (b.ioboard.org)

Skills

Python - pandas, numpy, scikit-learn, scipy, seaborn

R - dplyr, tidyr, data.table, ggplot2

SQL - MySQL, PostgreSQL

Web - React, d3.js, vx

2016 - 2018

2012 - 2016