



www.fernandogelin.com



lnkdin.me/fergelin



github.com/fernandogelin



fernando@fernandogelin.com



802.310.6213

technical skills

programming languages python, r
database SQL, mongoDB **scripts**
JavaScript, CoffeeScript **web** HTML,
CSS, Bootstrap **platforms** linux,
macOS **data formats** csv, json, xml,
unstructured, python dict **IDE** jupyter,
r studio **cloud** AWS **other** bash, *nix
command line, tableau

data science tools

python pandas, numpy, scipy, scikit-
learn, matplotlib, bokeh, seaborn,
graphlab-create, spark **R** dplyr,
data.table, ggvis, ggplot, caret, shiny
JavaScript D3.js, json

other skills

languages portuguese, english,
spanish **project management** MS
Project, agile, scrum **bioinformatics**
high-throughput sequencing and
analysis, r bioconductor, biopython,
healthcare, GATK

Fernando Gelin

Data Scientist

Education

Project Management Certificate, Biotech Jan 2015 - Nov 2015

University of Washington, Seattle, WA / University of California, San Diego, CA

Ph.D. Biology Sep 2009 - May 2014

University of Vermont, Burlington, VT

Relevant coursework / Online training

The Data Scientist's Toolbox ● R Programming ● Getting and Cleaning Data ● Exploratory Data Analysis ● Reproducible Research ● Using the Command-Line for Analysis of High Throughput Sequence Data ● Machine Learning Foundations: A Case Study Approach ● Data Manipulation at Scale: Systems and Algorithms ● Machine Learning: Regression ● Machine Learning: Classification ● Project Management: Standards and Processes ● Project Management Within a Biotech/Research Environment ● Applying Project Management Principles to Biomedical and Pharmaceutical Product Development ● Accessing Biomedical Big Data ● Visualization of Biomedical Big Data

Experience

Senior Research Fellow Jun 2014 - Jun 2016

Department of Psychiatry and Behavioral Sciences - University of Washington, Seattle, WA

- Analyzed human genome sequences using cutting edge bioinformatics tools such as GATK, Picard, BWA.
- Developed Python scripts for data cleaning, processing, and visualizations. 🔄
- Created a R shiny app to explore and visualize variant data. 📄 🔄
- Created a web application to generate protocols and calculate reagent amount for targeted sequencing reactions. 📄 🔄

Other Projects

Pronto Cycle Share Data Challenge Nov 2015 📄 🔄

- Data analysis and visualization of Seattle's bike share service first year of operation.
- Tools: **R** dplyr, data.table, reshape2, jsonlite, leaflet, maptools **web** HTML, CSS, JS, D3.js.

Obesity Comorbidity Analysis Jan 2016 🔄

- Obesity comorbidity analysis using PubMed articles database.
- Tools: **Python** pandas, numpy, json, urllib2, sqlalchemy, biopython, entrez, bokeh **database** PostgreSQL **cloud** AWS RDS.

Brainspan Viz March 2016 - present 🔄

- Interactive visualization of gene expression by brain region and age using data from brainspan.org (in progress).
- Tools: **Python** pandas, numpy, pymongo, scipy, scikit-learn, bokeh **database** MongoDB.