CV **Nicholas M. Nunley** | nicknunley17@ucla.edu | GitHub | Linkedin | Los Angeles, CA

Data scientist, software engineer, molecular biologist, and entrepreneur with experience in computational biology, statistical analysis, and production-grade software development; passionate about information theory, mathematics, and general problem solving

Experience

Research Data Scientist & Software Engineer (Staff Research Associate)

University of California Los Angeles (2022 - 2025)

- Hybrid data scientist, software engineer, and molecular and cell biology researcher that helped lead computational biology, statistical analysis, and production-grade software development
- Built end-to-end computational pipelines and experimental research projects across transcriptomic analysis, metabolic profiling, and software development in two laboratories, resulting in three peer-reviewed publications, two open-source packages, and a poster:
    - Goldstein lab
    - Boutros lab
- Owned and led two distinct research projects in parallel and developed a third open-source software project spanning experimental design, data analysis, and software development
- Architected and released an R datasets package containing processed transcriptional data from over 200 samples across seven different studies at release:
    - Performed data curation and QA, detecting and correcting sample swaps in external datasets
    - Built an object-oriented S4 dataset class to expose metadata, sample information, and analysis methods
    - Added gene set enrichment analysis and meta-analysis features to the package
    - Implemented a gene-identifier mapping system for cross-species analysis without external databases
    - Resolved graphics rendering failures from memory limits and label collisions via custom downsampling and coordinate offsets
    - Extended lattice axes labeling to support logarithmic axes
    - Designed a custom colour-mapping engine for continuous and diverging gradients using affine RGB transformations
- Mentored junior-level colleagues in Software Development Lifecycle (SDLC) practices, statistical modeling, & aseptic cell culture techniques
- Conducted peer code reviews across pull-request-based and ad hoc workflows related to data quality assurance, quality control, reproducibility, & analytical correctness
- Shipped RAGToolBox: a modular python package for Retrieval-Augmented Generation (RAG) prototyping
    - Engineered 9 modules and 4 CLI entrypoints for loading, indexing, retrieval, and augmentation workflows
    - Shipped v0.1.0 with a tagged release and changelog
    - Deployed 3 CI/CD workflows via GitHub Actions to automate build, test, and release steps
    - Published the package to PyPI
    - Implemented continuous unit and integration testing to ~83% coverage
- Managed mammalian cell culture and functional assays across ~50 experiments and ~8 prostate cancer model systems, including lentiviral transduction, immunoblotting, metabolic tracing, and viability assays

Lead K-12 Mathematics Instructor

Mathnasium, LLC (2019 - 2021)

- Provided one-on-one and group tutoring in foundational mathematics (e.g. algebra, geometry, trigonometry, calculus) and other general science courses
- Managed learning center operations on weekends:
    - Administered assessments to onboard new students
    - Managed student learning plans to track progress
    - Followed up with leads to convert enrollments
    - Scheduled and matched instructors with students to maintain coverage

Education

BS: Biochemistry

University of California Los Angeles (2020-2022)

- Conducted biomedical research while completing undergraduate coursework
- Completed a graduate-level course in mass spectrometry proteomics, and upper-division electives in mathematics and machine learning

AA (honors): Social and Behavioral Science

Santa Monica College (2015-2020)

- Explored multiple academic tracks before transitioning to STEM
- Worked part-time in service roles and as a peer tutor for foundational science and math courses

Skills

- Programming & scripting
  - Python programming
  - SQL
  - TypeScript
  - R programming
  - Unix
- Data science & analytics
  - Machine learning
  - Statistics
  - Optimization
  - Model evaluation & validation
  - Data visualization
  - Data modeling
  - AI Engineering
  - NumPy
  - pandas
  - Reproducibility
  - RAG
  - Network analysis
  - Simulation-based inference
  - Applied mathematics
- Software engineering & pipelines
  - git
  - CI/CD
  - Automation
  - API development
  - React web development
  - Docker
  - Testing (unit & integration testing)
  - Pipeline development and applications
  - High-performance computing
- Scientific computing & bioinformatics
  - Nextflow
  - Transcriptomics
  - Metabolomics
- Leadership
  - Project management
  - Problem formulation
  - Cross-functional collaboration
  - Mentorship
  - Scientific communication
- Experimental biology techniques
  - Biomedical experimental design
  - Mammalian cell culture
  - Immunoblotting
  - Microscopy and IHC
  - Cellular respirometry