

# Kristin Henderson

Leesburg, VA • [kdhenderson@gmail.com](mailto:kdhenderson@gmail.com) • <https://kdhenderson.github.io>  
<https://github.com/kdhenderson> • <https://www.linkedin.com/in/kdhende>

---

## TECHNICAL SKILLS

- **Programming:** Python, R, SAS, MATLAB, SQL, Bash
  - **Data & Analysis Libraries:** NumPy, pandas, dplyr, tidyverse, scikit-learn, XGBoost, PyTorch
  - **Machine Learning & AI:** RAG, FAISS, agentic workflows, prompt engineering, LoRA fine-tuning, contrastive learning, transfer learning, Hugging Face, LangChain
  - **Data Visualization:** matplotlib, seaborn, ggplot2, D3.js, Tableau, R Shiny
  - **Data Formats & Infrastructure:** Parquet, PostgreSQL, MySQL, AWS (RDS, EC2), LSF clusters
  - **Documentation & Version Control:** Git, GitHub, Jupyter notebooks, Quarto, Markdown
- 

## SELECTED PROJECTS

- Built a decision analysis framework in R to score the investment opportunity of emerging brands for a national retail brokerage / distribution firm, processing a large proprietary Parquet dataset with dplyr and benchmarking rankings with ordinal logistic regression.
  - Developed a Retrieval-Augmented Generation (RAG) pipeline in Python with contrastive-trained embedding models, a LoRA fine-tuned language model, and cross-encoder reranking.
  - Built an agentic workflow using LangChain and OpenAI SDK that routes queries to specialized RAG agents, retrieves context, and generates formatted responses.
  - Built a graph-based system using NetworkX to analyze migration and species co-occurrence.
  - Built Python scripts to download and consolidate electron microscopy image datasets and metadata from multiple public repositories for ML pipeline development.
  - Applied time-series forecasting methods (ARMA, VAR, MLP, ensemble) in R to model multi-horizon river flow rates for resource planning.
  - Authored an open-source Quarto-based statistics reference linking theory to applied data science: [kdhenderson.github.io/stats-notes-for-ds](https://kdhenderson.github.io/stats-notes-for-ds)
- 

## WORK EXPERIENCE

### **Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA**

Research Technician III, Jayaraman Lab. Dec 2021 – Present

- Build semi-automated Python pipelines to collect and process 100+ GB/day of behavioral video data for downstream statistical analysis and modeling.
- Apply statistical analysis and machine learning methods to identify patterns and behavioral structure from video tracking data.
- Collaborate with scientists and engineers to define analytical questions, assess data availability, and develop quantitative analyses and visualizations.
- Document hardware, software, and analysis workflows to support data quality, reproducibility, and cross-team use.

Research Technician II, Keleman Lab. Jan 2017 – Dec 2021

- Led data analysis and experiments studying learning and memory; prepared datasets, figures, and results for publication.

Research Technician I, Project Technical Resources. June 2015 – Dec 2016

- Supported high-precision microscopy data collection for downstream quantitative analyses.
- 

## EDUCATION

Master of Science in Data Science, May 2026, Southern Methodist University, Dallas, TX

Bachelor of Science, emphasis in Biology, The Evergreen State College, Olympia, WA