

# Kaleb Coberly

---

📍 Bellingham, WA, USA

✉ [kaleb.coberly@gmail.com](mailto:kaleb.coberly@gmail.com)

🌐 [kalebcoberly.com](http://kalebcoberly.com)

## Summary

Data engineer, backend software engineer with a background in research and analysis. Builds and deploys tools, analytics, and pipelines that help others reliably use data better. Supports users and developers with documentation, dev tools, issue triage, hands-on training, and debugging. Improves systems with a focus on clarity and maintainability. Prioritizes data quality, reproducibility, validity, and meaningful results.

## Skills

- *Programming & Scripting*: Python, R, Bash, SQL, NoSQL.
- *DevOps & Tooling*: Git, Make, Docker, GitHub Actions, Jenkins, CI/CD, Python packaging, debugging, QC and testing frameworks, pytest, pytype, black, flake8, jake.
- *Data Engineering*: ETL pipelines, RDBMS, database design, data modeling, data cleaning, data validation, methodological implementation, data versioning, MySQL, MariaDB, MongoDB, Elasticsearch, XML, JSON, CSV.
- *ML Engineering*: ML pipelines, feature engineering, performance evaluation, model validation, scikit-learn, caret.
- *Analysis & Visualization*: EDA, Pandas, NumPy, matplotlib, seaborn, ggplot2, datatable, dplyr, knitr, Jupyter Notebook, RStudio, R Markdown, mermaid, HTML, Excel, Kibana.
- *Documentation & Support*: Sphinx, reStructuredText, user training, contributor onboarding, issue triage.
- *Platforms & Environments*: Linux, MacOS, Windows, HPC (Slurm, UGE), conda, Kubernetes (k8s), Bitbucket/Stash, VS Code, RStudio, JIRA.

## Experience

### Advisory Board Member and Citizen Science Volunteer Engineer, Cascade STEAM, April 2025–Present

- [cascadesteam.org](http://cascadesteam.org)
- Leading the Citizen Science group's development of computer-vision ETL tool for stormwater monitoring field observation datasheets:
  - [github.com/crickets-and-comb/stormwater\\_monitoring\\_datasheet\\_extraction](https://github.com/crickets-and-comb/stormwater_monitoring_datasheet_extraction)

- Integrating DevOps tools and workflows with Cascade STEAM infrastructure.
- Consulting on selection of collaborative and project-management platforms.

### **Founder and Lead Engineer, Crickets and Comb, Nov 2024–Present**

- Implemented and maintain reusable DevOps tools and workflows for local + Docker development and GitHub CI/CD.
- Onboard developers with documentation, presentations, and hands-on support.
- Automated part of Bellingham Food Bank's delivery operations, reducing several staff-hours with each weekly use:
  - [cricketsandcomb.org/#projects](https://cricketsandcomb.org/#projects)
  - Designed and implemented an ETL CLI integrated with staff workflows and 3<sup>rd</sup>-party API.
  - Achieved high test coverage, by line count, problem space, and abstraction level.
  - Led incremental design, rollout, and user training.
  - Wrote and deployed technical and non-technical documentation.
  - Maintain ongoing user support and dependency management.
  - Released as open-source project; lead collaborative development at all skill levels.

### **Research Engineer, IHME – Central Computation (GBD), Nov 2022–Oct 2024**

- [www.healthdata.org/research-analysis/gbd](https://www.healthdata.org/research-analysis/gbd)
- Developed and maintained internal Python and dev tools for data modeling, versioning, access, and dev workflows.
- Contributed to a large in-house Python ecosystem: features, bugfixes, major test coverage increases, and documentation.
- Maintained Jenkins builds across ~100 packages; debugged issues across DBs, Python APIs/CLIs, Docker, k8s, and HPC.
- Directly supported research teams: Investigated failed model runs, data access issues, and tooling gaps.
- Analyzed system usage and tool performance via Elasticsearch; ran ad-hoc reports and built automated report.

### **Data Analyst, IHME – Pandemics Team, Nov 2021–Dec 2022**

- Owned multiple data pipelines for COVID-19 production; reduced key runtime by 65%.
- Refactored vaccine and booster models for maintainability and extensibility.
- Built diagnostics, with automated reporting for data QC; inspected and cleaned data.
- Led early UGE-to-Slurm transition for production jobs.
- Built ETL pipeline in R of national data into crosswalked, versioned output.
- Produced executive reports for world leaders, running ETL into R Markdown for PDFs; adapted to changing model outputs and feature requests:
  - [www.healthdata.org/research-analysis/diseases-injuries/covid/policy-briefings](https://www.healthdata.org/research-analysis/diseases-injuries/covid/policy-briefings)

## Education

B.S., Database Management and Data Analytics – Western Governors University

B.A., Critical Studies and Pedagogy – The Evergreen State College

## Certifications

- Quantum Computing (Qiskit & Python) – Udemy, Apr 2025
- Mathematics for ML & Data Science – DeepLearning.AI, Jan 2025
- Oracle Database SQL Certified Associate – Oracle, Jul 2020
- CompTIA A+, Network+, Project+ – CompTIA, 2019–2020