

TED HALEY

M.D.S., B.A.Sc.



tedhaley.ca



[linkedin.com/in/Ted-Haley](https://www.linkedin.com/in/Ted-Haley)



github.com/TedHaley



teddyhaley1@gmail.com



604-704-9467

I am a data engineer with experience in data science and software development. As someone who finds gratification in designing solutions to complex problems, I am interested in finding a position that will challenge me and allow me to continue to improve my skills as a data engineer and project manager.

EDUCATION

Masters of Data Science (M.D.S.)

University of British Columbia (UBC)

September 2017 – June 2018

- Student Representative
- Sun Run organizer and team captain
- 4.0/4.3 GPA

Data Analytics

Brainstation

March 2017 – May 2017

Bachelors of Applied Science in Civil Engineering

(Specialization in Management & Transportation Engineering)

University of British Columbia (UBC)

September 2011 – April 2016

- Completed the co-op program
- Member of the Institute of Transportation Engineers
- Member of Engineers Without Borders

SKILLS

Hard Skills:

Software Development,
Data Analysis,
Data Wrangling,
Data Visualization,
Machine Learning,
Statistics,
Project Management

Soft Skills:

Strong Communication,
Project Management,
Problem Solving,
AGILE, 6 Sigma

Frameworks:

Linux, AWS, Google Cloud,
Databricks

Programming Languages:

Python, R, SQL, NoSQL,
GIT, BASH, HTML, CSS,
Javascript

Software:

Terminal, PyCharm,
RStudio, Atom, Jupyter
Notebook, Slack, Docker,
TravisCI, Pip, Anaconda,
Tableau, Postgres, MySQL,
MongoDB

Packages:

Numpy, Pandas, Matplotlib,
DASH, Django, Flask, Luigi,
Threading, Scikit,
Tensorflow, PyTorch, Keras,
Tidyverse

WORK EXPERIENCE

Berlinguette Research - UBC

Applied Software & Data Engineer

September 2018 - Present

At UBC, I have been working on developing the autonomous software capabilities of a materials testing robot that utilizes machine learning to accelerate green material design and discovery ([Project Ada](#)). I manage a small team of software developers that operate and develop new features for the robotics systems alongside a group of chemists who design the experiments. [My work](#) and responsibilities include:

- Communicate with chemists and project stakeholders to plan for new features and to further develop the capabilities of the existing framework.
- Design and develop the data management system for experimental data (Postgres Database, REST APIs, Python Data Pipelines).
- Develop data validation, cleaning, and analysis scripts to promote machine learning and optimization workflows.
- Data dashboards for analyzing and querying data.
- Developing user interfaces to communicate data with the robot, database, and dashboard.
- Creating modular services using Docker for various services, including ETL pipelines, databases, REST APIs, and dashboards.

BGC Engineering Ltd. (BGC)

Data Science Intern

April 2018 – July 2018

I worked at BGC as a Data Science intern as a part of my Masters of Data Science program. There, I was tasked with developing an anomaly detection system and a flood forecasting model. The project included:

- Data wrangling, cleaning, and analysis in Python and R.
- Developing machine learning models for time-series and geographical data that can be scalable for use on thousands of gauges in real-time.
- Measure model performance and compare against the existing system.
- Optimize the ML model to improve accuracy.

PROJECTS

For the most up-to-date projects, please visit my website.

Project Ada - Berlinguette Research

September 2018 - Present: [Open Source](#)

Project Ada is an automated robotics platform that uses machine learning to accelerate the rate of material discovery. I joined the Berlinguette Research group to lead the development of the data management capabilities of the platform. I work alongside chemists, hardware engineers, and other software engineers to understand the software requirements of the project. Through analysing the relationships of material creation and testing, I was able to design a data management framework that consists of modular ETL data pipelines, Postgres databases, REST APIs, data validation and cleaning pipelines, and data visualization through a Dash web interface. All of the services run in Docker images on a Linux virtual machine that is hosted by Compute Canada.

VANquish Collisions - Hackathon

September 2018: [Open Source](#)

The VANquish Collisions hackathon was a 3-day event hosted by the City of Vancouver that is intended to develop solutions to achieve the city's goal of zero traffic related fatalities and serious injuries as part of the Moving Towards Zero Strategy. Using collision and infrastructure data, my team developed a web application to identify dangerous road crossings and provide an infrastructure upgrade to aide the most vulnerable road user at that location. To identify dangerous locations by road user, we used Scikit learn's K-Nearest Neighbours classifier in conjunction with a Grid Search optimization to find the best model parameters.

SCOOT - Personal Project

July 2018: [Open Source](#)

SCOOT is an interactive real-time transportation mapping web application I built using RStudio's Shiny web framework. The intended purpose is to allow the user to get transportation information for walking, cycling, transit, driving, and vehicle sharing platforms. SCOOT queries the location of the user (with their permission), and then queries that information against the Translink API and Google Maps API to get the best route for the users desired mode.

Flow Anomaly Detection System - BGC Engineering

June 2018: Closed Source

I developed a real-time anomaly detection system for river flow gauges across North America for BGC Engineering. The system is implemented using a Robust Mean LOESS method in R and Python. This system was able to classify anomalies with an F1 score of 83% on a dataset of over 33 million observations.

2-Day Flood Forecasting System - BGC Engineering

June 2018: Closed Source

I developed a 2-day flood forecasting system using Python and Sklearn for BGC Engineering. This system uses historic flow, precipitation, and spatial data to train a Random Forest model. This model was able to classify flood events with an F1 score of 74% on a dataset of over 33 million observations.

Pythia / Karl - UBC

April 2018: [Open Source](#)

Pythia and Karl are Python and R packages that take a dataset with a continuous response variable and various continuous explanatory variables, and provide the user with several functions to build a linear regression model. The fit of the regression can be tested using the diagnosis (residual) plots returned by the plotting function.