Maria Patterson

@ maria.t.patterson@gmail.com
in linkedin.com/in/mariatpatterson
🔗 mtpatter.github.io ⌥ github.com/mtpatter

BIG DATA ■ PROGRAM MANAGEMENT ■ DATA ANALYSIS ■ PROGRAMMING

EXPERIENCE

HIGH ALPHA | Data Scientist
Oct 2018 – present | 📍 Indianapolis, IN

- Working with B2B SaaS startups in the High Alpha portfolio to augment their software products with data science.

UNIVERSITY OF WASHINGTON, DEPT OF ASTRONOMY | Research Scientist III, Data Management
Aug 2016 – Oct 2018 | 📍 Seattle, WA

- Developed a Python module for real-time analysis of terabytes of streaming astronomical data using Apache Kafka and Avro, driving pipeline design, writing code, and assisting with system requirements documents in an Agile environment.
- Deployed, benchmarked, and optimized said pipeline using Docker and Prometheus to simulate analytics running on real-time telescope data from 200 sensors sending and receiving 10 million messages per night.
- Led a cross-team data product review effort to ensure both consistency and achievement of scientific system requirements.

CENTER FOR DATA INTENSIVE SCIENCE, UNIVERSITY OF CHICAGO, DEPT OF MEDICINE
Research Professional (Director of the Open Science Data Cloud)
Mar 2014 – Jul 2016 | 📍 Chicago, IL

- Program manager for a petabyte scale "community science cloud", managing several cloud-based data science projects and services, interfacing with systems and software engineers, and leading training for multi-disciplinary groups of scientists.
- Led the development of a 300 TB data distribution hub for climate data under the NOAA Big Data open gov project.
- Organized a data science and cloud computing summer fellowship program and supervised projects for students.

Research Professional (Data Scientist Equivalent)
Sep 2013 – Jul 2016 | 📍 Chicago, IL

- Maintained an automated machine learning pipeline for real-time cloud processing and analysis of daily acquired NASA satellite image data using Python, Hadoop, Accumulo, and Storm.
- Developed a water and land type classifier to process daily satellite images for said pipeline in Python using scikit-learn.
- Developed an algorithm, Neighbor-based Bootstrapping, to detect spatial patterns in geo-coded medical records using R.
- Modeled data storage systems in R using Monte Carlo methods to analyze hardware performance issues.

UNIVERSITY OF EDINBURGH | Visiting Fellow, Royal Observatory & School of Informatics
Jun 2013 – Aug 2013 | 📍 Edinburgh, Scotland

- Developed a Python tool to test query speeds and compare system utilization in row-oriented vs. column-oriented SQL database implementations for a large scale astronomical dataset.

EDUCATION

NEW MEXICO STATE UNIVERSITY | PhD, Astronomy
May 2013 | 📍 Las Cruces, NM

UNIVERSITY OF CHICAGO | BA, with Honors, Physics with a specialization in Astrophysics
May 2007 | 📍 Chicago, IL

COMPUTING SKILLS

PROGRAMMING LANGUAGES
Experienced:
Python: pandas, matplotlib, scikit-learn
R: caret, ggplot, parallel, knitr
Familiar:
C ▪ Fortran ▪ LATEX ▪ d3.js ▪ HTML/CSS ▪ Jekyll

TECHNOLOGIES
Experienced:
Linux ▪ Apache Kafka ▪ Apache Avro ▪ Apache Spark ▪ Docker ▪ GitHub ▪ Jira ▪ AWS (EC2, S3, Docker for AWS)
Familiar:
Kubernetes ▪ SQL ▪ Hadoop ▪ Google Cloud Platform

PROFESSIONAL DEVELOPMENT AND ENGAGEMENT

Speaker, "Building a Newsfeed from the Universe" — Kafka Summit | San Francisco, CA | Oct 2019
Panelist, "Befriending failure" — Tapia Celebration of Diversity in Computing | Atlanta, GA | Sep 2017
Speaker, "Building a Community Fountain around your Data Stream" — PyData | Seattle, WA | Jul 2017
Panelist, "Turning Big Data into Big Opportunities" — Tapia Celebration of Diversity in Computing | Austin, TX | Sep 2016
Invited Participant — White House Office of Science and Technology Policy Open Data Roundtable | Washington DC | Jun 2016
Speaker, "Big Data vs the Scientist" — ACM Meetup Group | Chicago, IL | Jun 2016
Lecturer — NSF PIRE Data Intensive Science and Cloud Computing Workshop | Amsterdam | 2014, 2015

HONORS AND AWARDS

20th Century Fox and PepsiCo's "Search for Hidden Figures" in STEM Contest, Winner | 2017
Murrell Award for Professional Development and Research Accomplishment | 2013
NMSU Outstanding Graduate Assistant Award (awarded twice) | 2009 & 2013
Pegasus Award for Excellence in Teaching | 2010

VOLUNTEER EXPERIENCE AND EXTRACURRICULARS

STEM gallery volunteer — The Children's Museum of Indianapolis, Beyond Spaceship Earth exhibit | Aug 2019 - present
Advisor — #BuiltByGirls WAVE Program | Jun 2019 - present
Member — American Astronomical Society's Committee on the Status of Women in Astronomy | Jun 2017 - present
Editor — AASWomen's Women in Astronomy Newsletter | Aug 2016 - present
Volunteer — UChicago Alumni Schools Committee | Aug 2009 - present
Co-Organizer — PyLadies Seattle and PyData Seattle | Jan 2017 - Oct 2018
Mentor — New York Academy of Sciences NEXT Scholar Program | Mar 2017 - Aug 2017
Volunteer — PyData Seattle, Diversity Committee | Jul 2017
Panelist — "Taking the Next Step" Career Program, University of Chicago | 2015, 2019
Distance Runner — 20 races including 1 marathon and 7 half marathons | runningaverage.com

SELECT PUBLICATIONS

The Zwicky Transient Facility Alert Distribution System
M.T. Patterson, E.C. Bellm, B. Rusholme, F.J. Masci, M. Juric, K.S. Krughoff, V.Z. Golkhou, M.J. Graham, S.R. Kulkarni, G. Helou, Zwicky Transient Facility Collaboration
Publications of the Astronomical Society of the Pacific, Nov 2018, Vol. 131, Iss. 995

Detecting Spatial Patterns of Disease in Large Collections of Electronic Medical Records Using Neighbor-Based Bootstrapping
M.T. Patterson and R.L. Grossman
Big Data, Sep 2017, Vol. 5, No. 3

The Case for Data Commons: Towards Data Science as a Service
R.L. Grossman, A. Heath, M. Murphy, M.T. Patterson, W. Wells
Computing in Science Engineering special issue: Science as a Service, Sep 2016, Vol. 18, Issue 5