Maria Patterson

linkedin.com/in/mariatpatterson
mtpatter.github.io github.com/mtpatter
@ maria.t.patterson@gmail.com 812.345.2654

BIG DATA ■ PROGRAM MANAGEMENT ■ DATA ANALYSIS ■ PROGRAMMING

EXPERIENCE

UNIVERSITY OF WASHINGTON, DEPT OF ASTRONOMY | Research Scientist III, Data Management

Aug 2016 – present | 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.
- Developed a Python module for detecting anomalies in time series data by implementing a bootstrapping method.

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 scientific 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 system 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

- Principal Investigator, translating telescope image data of galaxies to insights on galaxy mergers and star formation.
- Built 3D models of rotating galaxy disks for comparison to deep observations of gaseous galaxy halos using Python.
- Teaching assistant, preparing laboratory equipment, leading weekly meetings to plan exercises, and teaching undergrad labs.

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

CONFERENCES

Panelist on befriending failure, Tapia Celebration of Diversity in Computing — Atlanta, GA | Sep 2017
Speaker on open data streams in astronomy, PyData — Seattle, WA | Jul 2017
Panelist on big data, Tapia Celebration of Diversity in Computing — Austin, TX | Sep 2016
Participant, White House Office of Science and Technology Policy Open Data Roundtable — Washington DC | Jun 2016
Network Research Exhibition presenter, Supercomputing — Austin, TX, 2015| New Orleans, LA, 2014
Partnerships for International Research and Education, Big Data and Cloud Computing Workshops
- Lecturer on NASA satellite data analysis in the cloud — University of Amsterdam | Jun 2015
- Lecturer on reproducibility and collaborative tools — University of Amsterdam | Jun 2014
Target Scholar, Grace Hopper Celebration of Women in Computing — Phoenix, AZ | Oct 2014
Participant, New Mexico Celebration of Women in Computing — Las Cruces, NM | Nov 2012

CONTINUED LEARNING EFFORTS

Summer School in Statistics for Astronomers IX — Penn State | Jun 2013
8 data science related online classes — Coursera | 2013 - 2017

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

EXTRACURRICULARS AND VOLUNTEER EXPERIENCE

Member – American Astronomical Society's Committee on the Status of Women in Astronomy — Jun 2017 - present
Technical Talk Night Organizer – Pyladies Seattle and PyData Seattle — Jan 2017 - present
Editor – AASWomen's Women in Astronomy Newsletter — Aug 2016 - present
Volunteer – UChicago Alumni Schools Committee — Aug 2009 - present
Mentor – New York Academy of Sciences NEXT Scholar Program — Mar 2017 - Aug 2017
Volunteer – PyData Seattle, Diversity Committee — Jul 2017
Kaggle data science competitor – 7-time participant, two top 20% placements — kaggle.com/mtpatter
Distance Runner – 18 races including 1 marathon and 6 half marathons — runningaverage.com
Blog contributor – programming tutorials — opensciencecafe.org

SELECT PUBLICATIONS

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 Matsu Wheel: a reanalysis framework for Earth satellite imagery in data commons
M.T. Patterson, N. Anderson, C. Bennett, J. Bruggemann, R.L. Grossman, M. Handy, V. Ly, D.J. Mandl, S. Pederson, J. Pivarski, R. Powell, J. Spring, W. Wells, J. Xia
International Journal of Data Science and Analytics, Mar 2017, https://doi.org/10.1007/s41060-017-0052-3

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