MARIA PATTERSON
Email: maria.t.patterson_at_gmail.com
LinkedIn: http://linkedin.com/in/mariatpatterson
Website: http://mtpatter.github.io
GitHub: http://github.com/mtpatter

WORK EXPERIENCE

High Alpha
(approx 40 employees total, 3 in Data Science group)
Indianapolis, IN
Machine Learning Engineer - 12/2019 to present - 40 hours/week
Data Scientist - 10/2018 to 12/2019 - 40 hours/week

High Alpha is a venture studio focused on a new model for entrepreneurship that combines company building and traditional venture capital to design, create, launch, and fund Software-as-a-Service companies in the Business-to-Business space. "Upon launching a new company, we surround the startup with world-class expertise in every discipline required to build a market-leading company. We compress the amount of time it takes to move from an idea to a world-class business. Our dedicated teams span practice areas, such as brand and design, product and engineering, talent and HR, sales and marketing, finance, and data science." Our motto is "Dream Big. Expect More. Move Fast."

DUTIES AND RESPONSIBILITIES
Leading the design of architecture for machine learning pipelines and cloud-based data analysis for early-stage SaaS companies to augment their software products and to improve business operations with data science.
- Machine learning and data pipeline projects:
    - Statistical analysis of ads to recommend creative optimizations for clients of an AI marketing platform company.
    - Automated machine learning detection of financial transactions for clients of a software management company.
    - Similar people, skills, and mentor recommendations systems for clients of a company that provides an intelligent, data-rich, internal personnel directory and communications platform.
    - Classification and prediction system for identifying high potential qualified leads for a product-led growth company.
    - End-of-period forecasting for sales rep deals management software.
    - Churn prediction for customer retention models.
- Collaboration on cross-functional teams:

- ○ Collaborate with product design teams and company leadership to set goals and metrics of success for data science projects.
- ○ Collaborate with engineering teams to incorporate analysis results into pipelines and end product software.
- Entrepreneurship:
  - ○ Participated in 5 "Sprint Weeks," which are High Alpha's forcing function behind startup launch - compressing market research, customer validation, product and brand design, revenue and go-to-market strategy planning, and investor pitches into one week.
- Presentations and thought leadership:
  - ○ Authored two blog posts on streaming data architecture with combined 43k+ views since April 2019.
    - ■ https://medium.com/high-alpha/data-stream-processing-for-newbies-with-kafka-ksql-and-postgres-c30309cfaaf8
    - ■ https://medium.com/high-alpha/streaming-data-from-the-universe-with-apache-kafka-3b6b54dee6a9
  - ○ Delivered an accepted technical talk called "Cloud architecture for the data scientist: Deploying machine learning pipelines to production" at Indy Cloud Conf 2020.
  - ○ Delivered an accepted use case talk called "Building a newsfeed from the Universe: Data streams in astronomy" at Kafka Summit 2019.
    - ■ https://www.confluent.io/kafka-summit-san-francisco-2019/building-a-newsfeed-from-the-universe-data-streams-in-astronomy/

TECHNICAL SKILLS
Utilizing machine learning Python packages (sklearn, pandas, Jupyter notebook) and collaborative management software (GitHub, comet.ml) for data science analysis. Utilizing collaborative software for Kanban board style task management and building a team knowledgebase (Notion, Google Docs and Sheets). Deploying data analysis pipelines on cloud platforms with scalable technologies (Amazon Web Services - ec2; Google Cloud Platform - Pub/Sub, BigQuery; Docker tools including Kubernetes).

Data Intensive Research in Astrophysics and Cosmology (DIRAC) Institute, University of Washington
(approx 30 researchers in DIRAC institute)
Seattle, WA
Research Scientist III, Data Management Group, Large Synoptic Survey Telescope - 08/2016 to 09/2018 - 40 hours/week

"DIRAC's objective is…to think of software as the chief instrument for exploring the universe. We are creating a new model of data intensive, computationally-driven science – one that is profoundly interdisciplinary, uniting computer scientists, statisticians, astrophysicists and

cosmologists to develop the computational solutions to problems presented by massive data streams. [Focused research teams] develop new approaches for managing, storing, and accessing petabyte data sets and running analyses at scale and within these databases. The Institute is the culmination of ten years of crosscutting research, strategic hiring, interdisciplinary partnerships, and proven collaborative relationships with technology companies like Microsoft and Google."

DUTIES AND RESPONSIBILITIES

Responsible for building open source software for the science pipelines team behind the analysis of data from the Zwicky Transient Facility and the upcoming petabyte scale Vera Rubin Observatory (formerly Large Synoptic Survey Telescope). Leading the design and prototyping of alert distribution systems, streaming telescope data on newly detected changing objects in the sky out to the community in real time.

- Data pipelines:
    - Architected a real-time data processing and analysis framework for terabytes of streaming astronomical data.
    - Led the production deployment of said framework for the Zwicky Transient Facility's (ZTF) Survey Alert Stream Distribution System (ZADS), which processes data from one million changing objects in the sky each night and is hardened at 10x scale.
- Technical system design:
    - Collaborated on system requirements documents in an Agile environment and led a cross-team data product review effort to ensure both consistency and achievement of scientific system requirements.
    - Deployed, benchmarked, and optimized software on simulated data at terabyte scale.
- Presentations and thought leadership:
    - First-author on one peer-reviewed paper describing technical system design.
        - https://mtpatter.github.io/new-paper-zads/
    - Delivered technical talks at several internal collaboration meetings. Delivered an accepted talk called "Building a community fountain around your data stream" at PyData 2017.

TECHNICAL SKILLS

Utilized open source software including Python, GitHub, Kafka, Avro, Docker, and Prometheus for product base, testing, and collaboration. Utilized large scale computational resources for benchmarking our software products (Docker for AWS, Kubernetes).

Center for Data Intensive Science (now Center for Translational Data Science), University of Chicago
(approx 40 team members in CTDS)

Chicago, IL
Research Professional (Director of the Open Science Data Cloud) - Mar 2014 to Jul 2016 - 40 hours/week
Research Professional - Sep 2013 to Jul 2016 - 40 hours/week

"Our work centers around developing instruments to integrate commons of complex data with cloud computing technology. We architect large scale commons of research data, computing resources, applications, tools, services. Through this approach, we can more effectively use data at scale to study and pursue scientific inquiry in the areas of biology, medicine, healthcare, and the environment. We are leaders in data sharing, expanding opportunities for dissemination among the research community and accelerating discovery. Our leadership emerged with the launch of the first open-source cloud-based computational research platform recognized as an NIH Trusted Partner, achieving rigorous data quality and data management service requirements. Today we offer over seven petabytes of rich research data through our data commons platforms."

DUTIES AND RESPONSIBILITIES
Program manager for a petabyte scale "data commons", managing on-prem cloud and cloud-based data science projects and services, and interfacing with systems and software engineers in a matrix organization, as well as leading research projects.
- Cross-functional program management:
    - Organized an international data science and cloud computing summer fellowship program and supervised student projects.
    - Developed a response and proposal to NOAA's Big Data Project (BDP), an open government initiative in collaboration with Presidential Innovation Fellows (PIF Round 3), Request for Information leading to a Cooperative Research and Development Agreement with the Open Commons Consortium.
- Data pipelines and technical architecture:
    - Maintained an end-to-end automated analytic pipeline of machine learning algorithms for cloud processing and analysis of daily acquired satellite data for NASA.
    - Architected a 300 TB data distribution hub for academic and scientific researchers as Technical Lead for the Open Commons Consortium on the NOAA BDP.
- Data science research projects:
    - Developed a water and land type classifier to process daily satellite images for said pipeline.
    - Developed a method for detecting spatial patterns in geo-coded medical records.
    - Modeling data storage systems using Monte Carlo methods to analyze hardware performance issues.
TECHNICAL SKILLS
- Utilized Python, Hadoop, Accumulo, and Storm for streaming pipelines and R for data science research.

University of Edinburgh, Royal Observatory & School of Informatics
Edinburgh, Scotland
Visiting Fellow, as participant in a National Science Foundation Partnerships for International Research and Education (PIRE) run by the University of Chicago's Open Science Data Cloud project - Jun 2013 to Aug 2013 40 hours/week

"The Open Science Data Cloud (OSDC) NSF-sponsored Partnerships for International Research and Education (PIRE) program is driven by two main goals: 1) Use the Open Science Data Cloud to train the next generation of scientists (graduate students and early career researchers) in data science and to sponsor their travel so that they can collaborate on data science research projects with OSDC partners in the United Kingdom, Namibia, the Netherlands, Japan, China, Brazil, and other countries. 2) Perform research to improve scientific clouds, including their software stack, the software service they run, and the algorithms that support these services and to use the resulting knowledge to improve the OSDC.

DUTIES AND RESPONSIBILITIES
- Programmed 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.

Sapling Learning, Inc. (Macmillan Learning - parent company)
Remote/Online
Astronomy Content Author and Reviewer - Apr 2013 to Nov 2013 - 15 hours/week

"Created by and for educators, Sapling Learning online homework drives student success with a variety of questions that include wrong-answer feedback, hints, and solutions as well as time-saving tools for educators."

DUTIES AND RESPONSIBILITIES
- Created questions and solutions of varying difficulty to test students' understanding of astronomical concepts in an online interactive system for higher education course homework.

EDUCATION
New Mexico State University, Las Cruces, NM
PhD, Astronomy - May 2013
DUTIES AND RESPONSIBILITIES
- 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, Chicago, IL
BA, with Honors, Physics with a specialization in Astrophysics - May 2007

DUTIES AND RESPONSIBILITIES
- Built Monte Carlo models to investigate optimal telescope arrangements for science use cases.

SPECIALTIES

Data science at scale - Analysis - Machine learning Python and relevant tools (pandas, scikit-learn, matplotlib, dash) - Production quality data engineering - Cross-matrix organizations - Public speaking - University teaching - Scientific writing and presentations - Public outreach - Mentoring - Computing languages: Python, R, SQL - Open source: Docker, Kafka, Avro - Clouds: AWS (EC2, S3, Docker for AWS), Google Cloud Platform - GitHub - Docker (Compose, Swarm, Kubernetes)

PROFESSIONAL ENGAGEMENT

Conference Presentations
- Speaker, "Streaming your science: Analytics for real time" - Women in Analytics (2020 postponed, accepted for 2021) - Columbus, OH
- Speaker, "Cloud architecture for the data scientist" - Indy Cloud Conf 2020 - Indianapolis, IN - Jun 2020
- Speaker, "Building a newsfeed from the Universe" - Kafka Summit 2019 - San Francisco, CA - Oct 2019
- Invited Panelist, Careers in Science - Field Museum's 5th Annual Chicagoland Women in Science Mixer - Chicago, IL - Oct 2019
- Panelist (accepted proposal), "Befriending failure" - Tapia Celebration of Diversity in Computing - Atlanta, GA - Sep 2017
- Speaker, "Building a community fountain around your data stream" - PyData 2017 - Seattle, WA - Jul 2017
- Panelist (accepted proposal), "Turning Big Data into Big Opportunities" - 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
- Speaker, "Big Data vs the Scientist" - ACM Meetup Group - Chicago, IL - Jun 2016

- Lecturer - NSF PIRE Data Intensive Science and Cloud Computing Workshops - Amsterdam - 2014, 2015 Topics - NASA satellite data analysis in the cloud, reproducibility and collaborative tools
- 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

ADVISORY AND COMMITTEE SERVICE

Committee Member - AAS's Committee on the Status of Women in Astronomy (CSWA) - Jun 2017 - present
- Charged with providing recommendations to the American Astronomical Society Board of Trustees for practical measures that can be taken to improve the status of women in astronomy and encourage entry into the field.

Sub-group committee member - CSWA Strategic Plan - Oct 2019 - present
- Charged with writing a cascade model strategic plan, providing direction for the next 10 years of focus areas important to women in the astronomical community.

Sub-group committee member - CSWA Astro2020 Decadal Survey White Papers - Jun 2019 - present
- Co-authored two science white papers on the state of the profession to the Astro2020 Decadal Survey on Astronomy and Astrophysics organized by the National Academy of Sciences.

Planning Committee - PyData Seattle Conference - Jun/Jul 2017
- Reviewed and assessed conference talk proposals from speaker applications.

Diversity Committee - PyData Seattle Conference - Jul 2017
- Responsible for finding funding for and organizing the awarding of scholarships for attending the conference.

Roundtable Contributor - White House OSTP Open Data Roundtable Series - Jun 2016
- Contributed to the White House Office of Science and Technology Policy and Center for Open Data Enterprise Open Data Roundtable on Public-Private Collaboration, representing the academic and not-for-profit research perspective and resulting in "The Global Impact of Open Data" report.

VOLUNTEER EXPERIENCE AND EXTRACURRICULARS

STEM gallery volunteer - The Children's Museum of Indianapolis, Beyond Spaceship Earth exhibit - Aug 2019 - present (3 hour shifts, every 2 weeks)
- Teach families about the International Space Station and assist with A/V for shows

Advisor - #BuiltByGirls WAVE Program - Jun 2019 - present
- Advisor to female college undergraduates and recent graduates on a career path in tech.

Editor - AASWomen's Women in Astronomy Newsletter - Aug 2016 - present
- Editing a weekly newsletter distributed to 1,500+ astronomers with articles and content relevant to women in science.

Volunteer - UChicago Alumni Schools Committee - Aug 2009 - present
- Interviewing prospective students for the University admissions process.

Co-Organizer - PyLadies Seattle and PyData Seattle Meetup Groups - Jan 2017 - Oct 2018
- Organized monthly Technical Talk Nights for Meet-up groups around the Python programming language and women.

Mentor - New York Academy of Sciences NEXT Scholar Program - Mar 2017 - Aug 2017
- Mentored and provided support for a female undergraduate in STEM.

Distance Runner - 20 races including 1 marathon and 7 half marathons - https://www.athlinks.com/athletes/219901730/

Amateur woodworker - Small furniture items and hand-tool wood carving

HONORS AND AWARDS

20th Century Fox and PepsiCo's "Search for Hidden Figures" in STEM, Hidden Figure Awardee in Professionals category - 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

SELECT PUBLICATIONS

Findings and Recommendations from the AAS Committee on the Status of Women in Astronomy: Advancing the Career Development of Women in Astronomy
N. Zellner, J. McBride, N. Morrison, A. Olmstead, M.T. Patterson, G. Rudnick, A. Venkatesan, H. Flewelling, D. Grinspoon, J. D. Mink, C. Richey, A. Speck, C. A. Thomas, S. E. Tuttle
Published in: Bulletin of the American Astronomical Society, Vol. 51, Issue 7, id. 170 (2019)

Findings and Recommendations from the AAS Committee on the Status of Women in Astronomy: Towards Eliminating Harassment in Astronomy

N. Zellner, J. McBride, N. Morrison, A. Olmstead, M.T. Patterson, G. Rudnick, A. Venkatesan, H. Flewelling, D. Grinspoon, J. D. Mink, C. Richey, A. Speck, C. A. Thomas, S. E. Tuttle
Published in: Bulletin of the American Astronomical Society, Vol. 51, Issue 7, id. 169 (2019)

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
 Published in: 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
 Published in: 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
Published in: Computing in Science Engineering special issue: Science as a Service, Sep 2016, Vol. 18, Issue 5