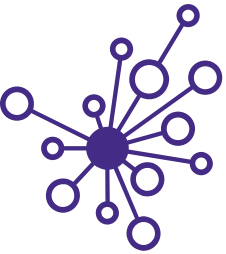


UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY of WASHINGTON

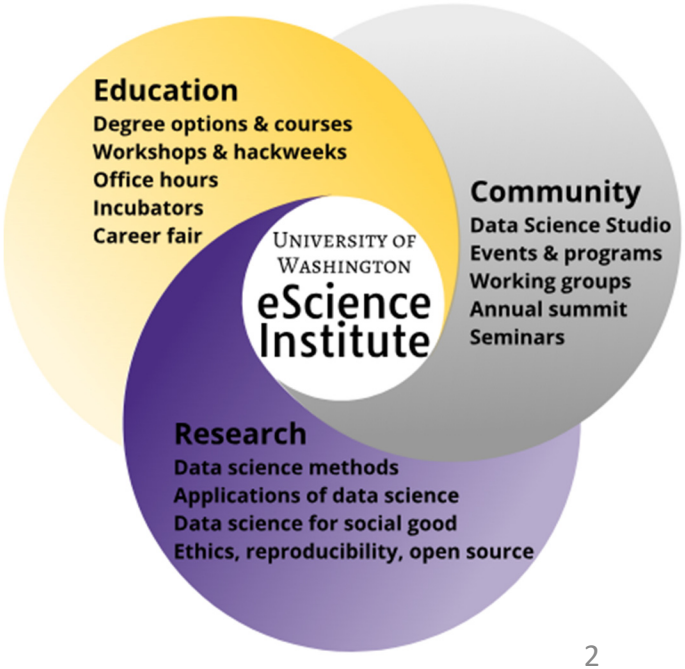
Data Science Specializations in PhD & Undergrad Programs at UW

Magdalena Balazinska
Professor, Paul G. Allen School of Computer Science & Eng
Director, eScience Institute
Associate Vice-Provost for Data Science
University of Washington

UW eScience Institute Since 2008

http://escience.washington.edu/

The eScience Institute **empowers** researchers and students in all fields to answer fundamental questions through the use of large, complex, and noisy data. As the **hub** of data-intensive discovery on campus, we lead a **community** of innovators in the techniques, technologies, and best practices of data science and the fields that depend on them



10 years!

Who We Are?

Founding Director

Ed Lazowska
Ph.D. Computer Science

Director

Magdalena Balazinska
Ph.D. Computer Science

Executive Directors

Micaela Parker
Ph.D. Oceanography

Sarah Stone
Ph.D. Oceanography

Director of Research

David Beck
Ph.D. Medicinal Chemistry,
Biomolecular Struct. & Design

Events

Rachael Murray
Program Specialist

Communications

Robin Brooks
Communications
Associate

Cloud Solutions

Rob Fatland
Director of Cloud &
Data Solutions
Ph.D. Geophysics

Amanda Tan
Cloud Technology
Lead Developer
Ph.D. Civil & Env. Eng.

Data Scientists

Bernease Herman
B.S. Statistics
Formerly SE at Amazon

Ariel Rokem
Ph.D. Neuroscience

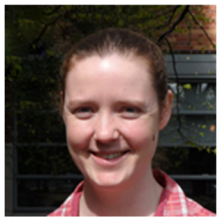
Valentina Staneva
Ph.D. Applied
Mathematics and
Statistics

Jake VanderPlas
Director of Open
Software
Ph.D. Astronomy

Jose Hernandez
PhD Measurement
& Statistics

Research Scientists

Anthony Arendt
Ph.D. Geophysics
APL



Bryna Hazelton
Ph.D. Astrophysics
Physics

Joe Hellerstein
Ph.D. Computer Science
IBM Research, Microsoft Research, Google (ret.)

Vaughn Iverson
Ph.D. Oceanography
Oceanography

Executive Committee and 100+ affiliated faculty from around UW

Cecilia Aragon
Director, Human-Centered DS Lab; Prof., HCDE

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Ginger Armbrust
Director, School of Oceanography

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Magdalena Balazinska
Director; IGERT Director

Data Science Fellow, Directorship, Executive Committee, Steering Committee

READ MORE >

David Beck
Director of Research; Research Assistant Prof., Chemical Engineering

Data Science Fellow, Directorship, Executive Committee, Research Staff, Steering Committee

READ MORE >

Ed Lazowska
Founding Director

Data Science Fellow, Directorship, Executive Committee, Research Staff, Steering Committee

READ MORE >

Randy LeVeque
Prof., Applied Mathematics, Earth & Space Sciences

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Tyler McCormick
Associate Prof., Statistics, Sociology

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Micaela Parker
Executive Director

Administrative Staff, Data Science Fellow, Directorship, Executive Committee, Steering Committee

READ MORE >

Andy Connolly
Prof., Astronomy

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Tom Daniel
Joan & Richard Komen Endowed Chair, Biology

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Rob Fatland
Director of Cloud and Data Solutions

Data Science Fellow, Directorship, Executive Committee, Research Staff, Steering Committee

READ MORE >

Bill Howe
Former Associate Director; Associate Prof., iSchool

Data Science Fellow, Executive Committee, Steering Committee

READ MORE >

Sarah Stone
Executive Director

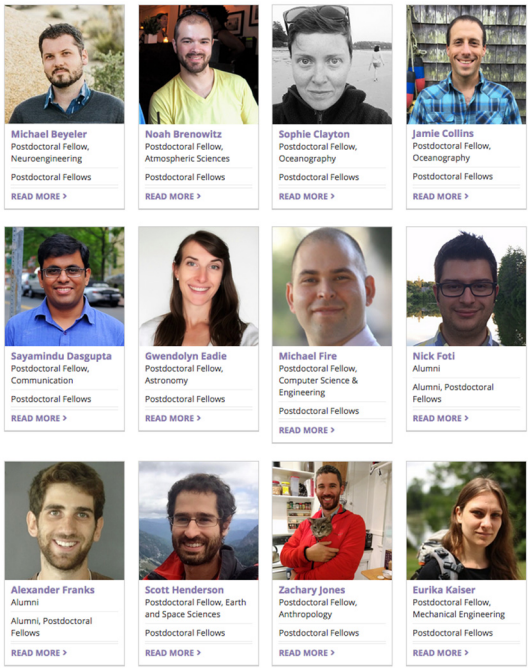
Administrative Staff, Data Science Fellow, Directorship, Executive Committee, Steering Committee

READ MORE >

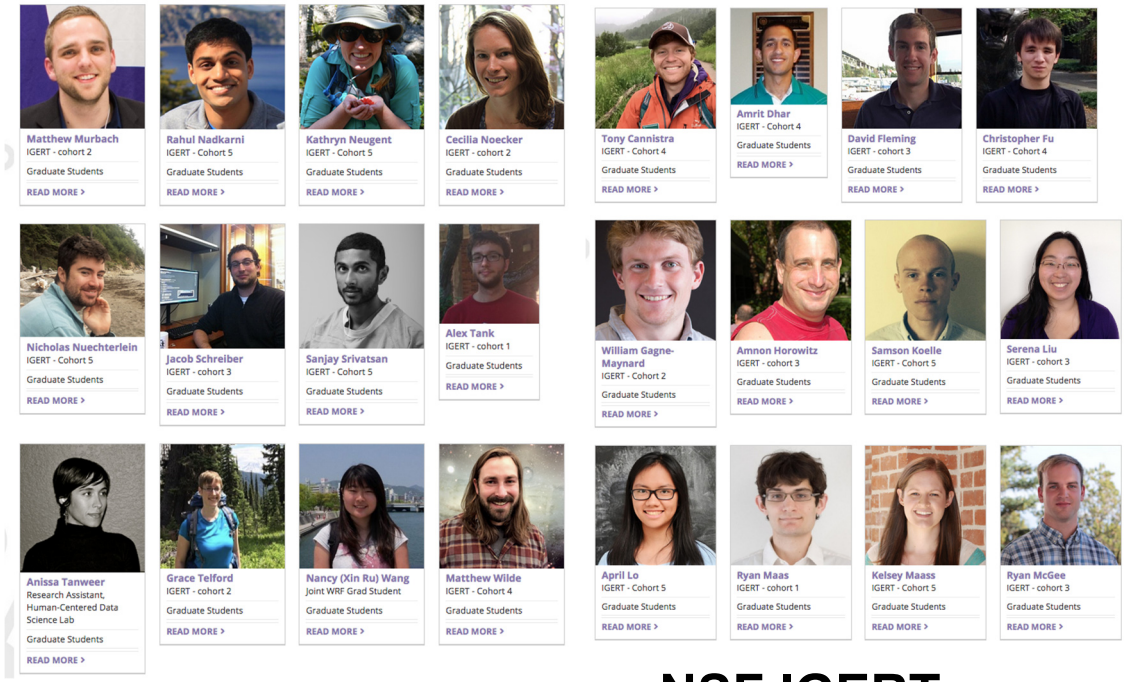
Jake VanderPlas
Director of Open Software; Senior Data Scientist

Data Science Fellow, Directorship, Executive Committee, Research Staff, Steering Committee

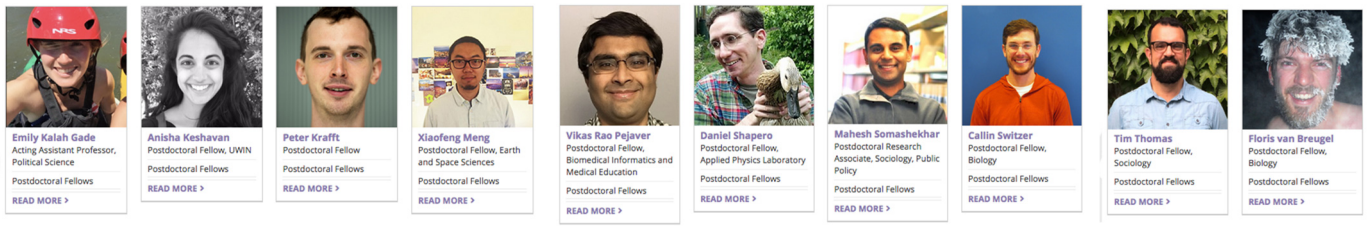
READ MORE >



Postdocs and PhD Students



NSF IGERT (awarded in 2013)



Moore/Sloan & WRF Innovation in Data Science Postdoctoral Fellows

eScience Education Working Group

- Inter-disciplinary working group
- Members from large variety of schools and departments
  - Anyone interested in data science education is welcome!
- Monthly meetings
- Defined common framework for data science specializations:
  - What does data science mean?
  - What are minimal core components of a data science specializations?
  - What courses are missing and who should teach them?

"Data Science for ALL"

Formal Programs

Short Courses, Workshops, & Seminars

eScience Education Working Group

Educational objectives

Goal 1: Educate ALL UW students in data science

- Many need to learn how to **use** data science tools
- Some need to learn how to **build** data science tools

Goal 2: Create a community of data science students

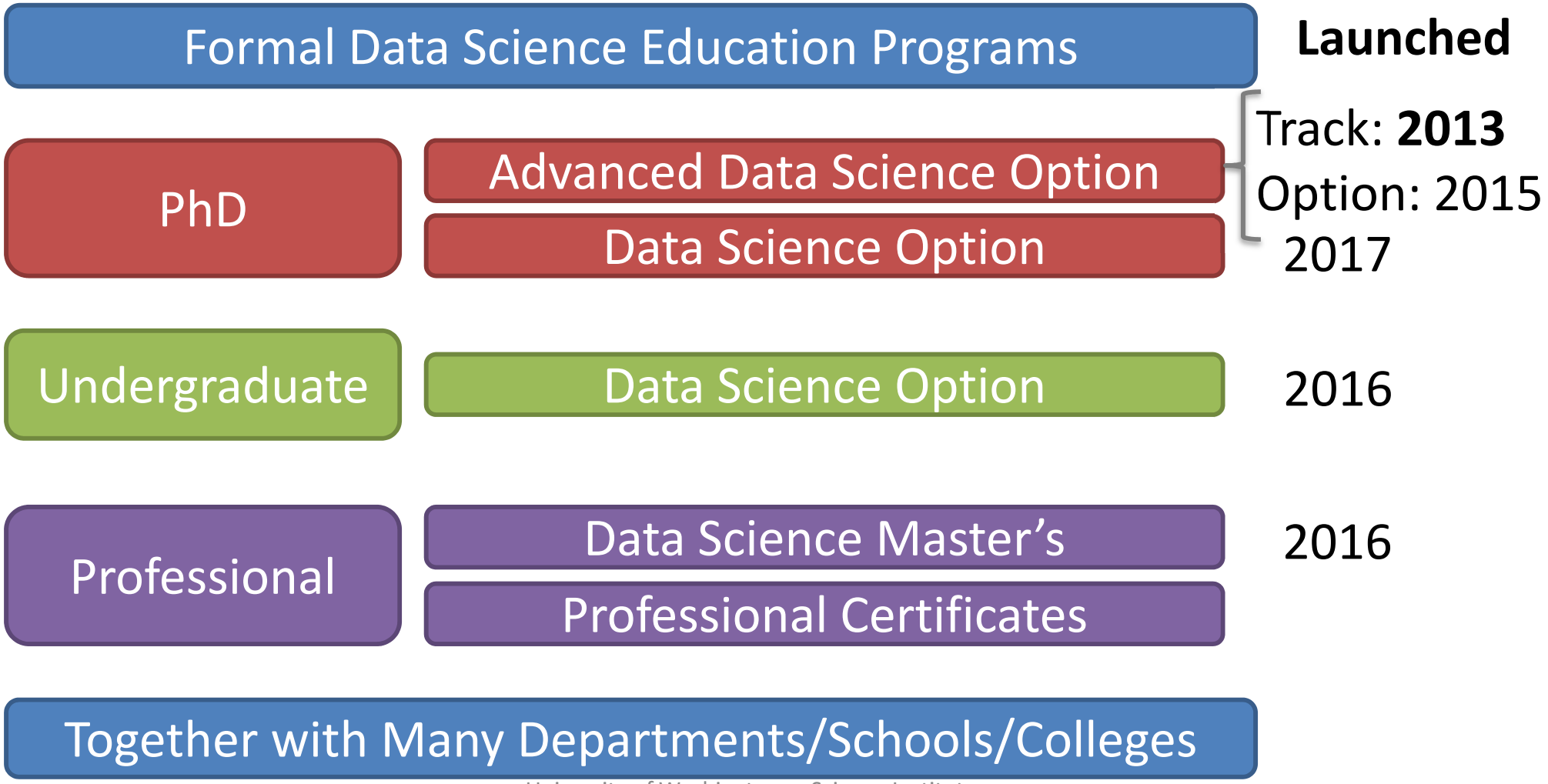
- Inter-disciplinary community of students
- Tied to eScience Institute community

"Data Science for ALL"

Formal Programs

Short Courses, Workshops, & Seminars

eScience Education Working Group

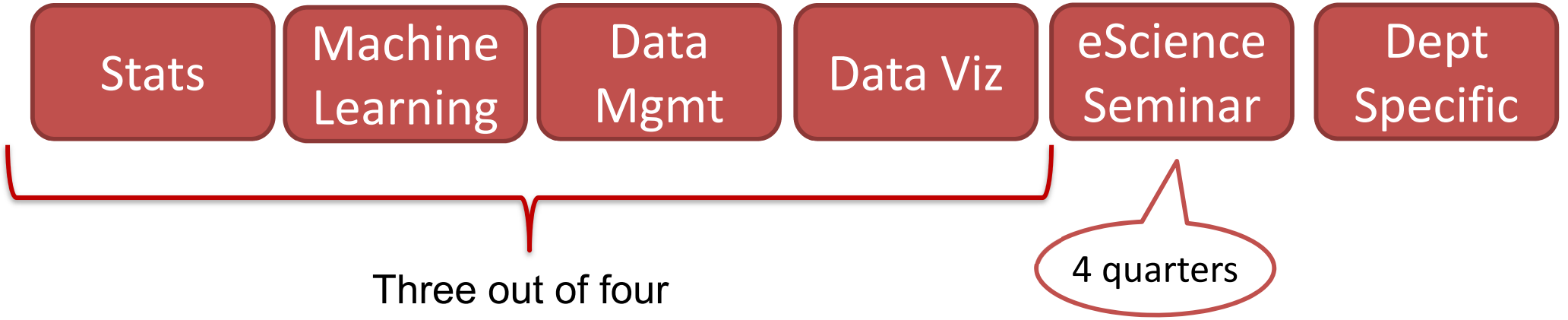

Formal Data Science Education Programs
Launched
PhD
Advanced Data Science Option
Track: 2013
Option: 2015
Data Science Option
2017
Undergraduate
Data Science Option
2016
Professional
Data Science Master's
2016
Professional Certificates
Together with Many Departments/Schools/Colleges

PhD Data Science Options

- Option: specialization within existing major
- Student first admitted to participating department
- Student elects to pursue the option
- Single framework under eScience umbrella
- Central steering committee
  - One representative from each department/school
- But otherwise managed by individual departments

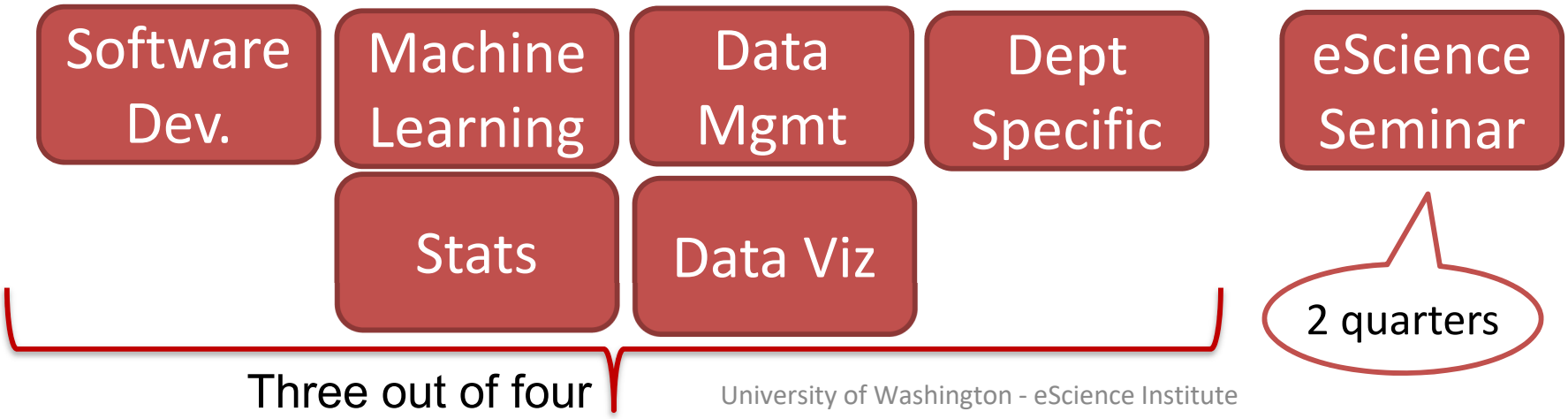
Advanced Data Science Option (Transcriptable)

- Bootstrapped with an NSF IGERT grant
- Designed for tool builders
- Grad courses in CSE & Stats + Seminar + Specialized



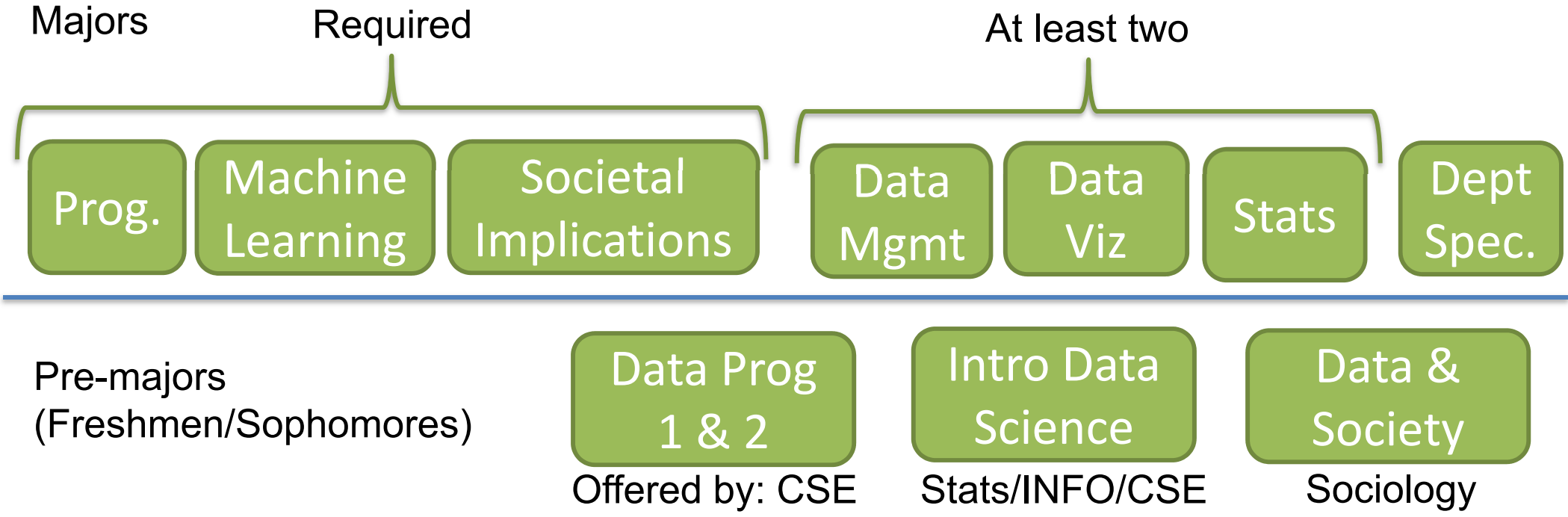
Data Science Option (Transcriptable)

- Common framework with specialized instantiations
- Designed for tool users
  - Courses with limited pre-reqs, designed for everyone
  - Courses by CSE, iSchool, HCDE, Stats, ChemE, other



Data Science for Undergraduate Students

- Data Science Option (Transcriptable)
  - Common framework with specialized instantiations



Community Building through eScience Seminars

eScience Community Seminar

The eScience Community Seminar serves as an informal environment for presentations and discussions on research that is relevant to the eScience Institute and the Data Science Environment. Topics will span science, methods, and technology across the mission of the eScience Institute.

Spring 2018 Schedule

Thursdays, 4:30 p.m., Physics Astronomy Auditorium (PAA), room A110

Date	Title	Speaker	Notes
3/29/18	machine and deep learning applications to voice data for Parkinson's disease	Reza Hosseini Ghomi	
4/5/18	Life of a Data Scientist in a Metabolic Engineering Startup	Janet Matsen	
4/12/18	Computational Approaches to Privacy, Bias, Transparency in Data Science	Bernease Herman	
4/19/18	Assured information distillation in social sensing	Dong Wang	
4/26/18	Algorithmic Curation: Putting Open Science Data to Work	Bill Howe	
5/3/18	Data-based sustainable agriculture: machine learning groundwater models and irrigation development in Africa	Tess Russo	
5/10/18	When Can We Say a Rumor Has Grown Out of Proportion? A Computational Approach	Peter Krafft	
5/17/18	Cancelled	Cancelled	
5/24/18	Predicting Substitution Profiles for B Cell Receptor Sequences using Large-Scale Repertoire Data	Amrit Dhar	
5/31/18	TBD	Andrej Blejec	

Weekly seminar * 3 quarters
Typically, approximately 40-45 attendees
Approximately 1000 attendee-hours

UW Data Science Seminar

Analysis, Visualization & Discovery

The Data Science Seminar is a university-wide effort bringing together thought-leading speakers and researchers across campus to discuss topics related to data analysis, visualization and applications to domain sciences.

All seminars are free and open to the public.

Spring 2018 Schedule

Wednesdays, 3:30 to 4:20 p.m., Physics/Astronomy Auditorium, room A102

Mar 28		**"Vulnerabilities in a sociotechnical society"** danah boyd, Principal Researcher *Microsoft Research*
Apr 11		**"Making data science training resources FAIR"** John Darrell Van Horn, the USC Mark and Mary Stevens Neuroimaging and Informatics Institute and the Laboratory of Neuro Imaging
Apr 25		**"Artificial unintelligence: How computers misunderstand the world"** Meredith Broussard, Assistant Professor, Arthur L. Carter Journalism Institute of New York University
May 23		**"Data science, AI, and big data analytics for global good and social impact"** Uyi Stewart, Director of Global Development Strategy, Data, & Analytics
May 30		**"Mining personal, dense, dynamic data clouds to enhance health and drive discovery"** Nathan Price, Professor & Associate Director of the Institute for Systems Biology

9 speakers in 2017-18 for a total of nearly 500 attendee-hours

Community Building through eScience Seminars

eScience Community Seminar

The eScience Community Seminar serves as an informal environment for presentations and discussions on res
eScience Institute and the Data Science Environment. Topics will span science, methods, and technology acros
Institute.

Spring 2018 Schedule

Thursdays, 4:30 p.m., Physics Astronomy Auditorium (PAA), room A110

Date	Title	Sp
3/29/18	machine and deep learning applications to voice data for Parkinson's disease	Re Gh
4/5/18	Life of a Data Scientist in a Metabolic Engineering Startup	Jar
4/12/18	Computational Approaches to Privacy, Bias, Transparency in Data Science	Be He
4/19/18	Assured information distillation in social sensing	Do
4/26/18	Algorithmic Curation: Putting Open Science Data to Work	Bil
5/3/18	Data-based sustainable agriculture: machine learning groundwater models and irrigation development in Africa	Te
5/10/18	When Can We Say a Rumor Has Grown Out of Proportion? A Computational Approach	Pe
5/17/18	Cancelled	Ca
5/24/18	Predicting Substitution Profiles for B Cell Receptor Sequences using Large-Scale Repertoire Data	An
5/31/18	TBD	An

Topics in Data Science Seminar

A Series of Tutorials and Demos on Topics of Interest to Data Scientists

The Topics in Data Science seminar is a new offering held this spring at the eScience institute. The goal is to provide a closer look at useful techniques in data science in a long-form interactive setting. Over the quarter, we'll hear from a variety of speakers discussing tools and techniques that go beyond the fundamentals presented in previous courses. The seminars are partly interactive, so bring your laptops!

Spring 2018 Schedule

Time: Wednesdays, 10:00 – 11:30 a.m.

Location: Data Science Studio Seminar Room, 6th Floor Physics/Astronomy Tower

The seminars are partly interactive, so bring your laptops!

Date	Title	Speaker	Affiliation
April 20 (Friday)	Tools for Data Cleaning and Munging (files)	Ryan Maas	Staff scientist, UW eScience
April 25	Tensor Decompositions in Machine Learning (github repo)	Valentina Staneva	Data Scientist, UW eScience
May 2	Sequential Experimental Design	Kevin Jamieson	Professor, UW CSE
May 8	Vega-Lite – A Grammar of Interactive Graphics	Dominik Moritz	Ph.D. Candidate, UW CSE
May 16	Text Classification: From Logistic Regression to Neural Networks (data/slides)	Yangfeng Ji	Postdoctoral Researcher, UW CSE
May 23	Evaluation Methods for Machine Learning	Bernease Herman	Data Scientist, UW eScience
May 30	Convex Modeling and Optimization	Maryam Fazel	Professor, UW EE
June 6	Theoretical Insights on Deep Learning	Zaid Harchaoui	Professor, UW Statistics

UW Data Science Seminar

Analysis, Visualization & Discovery

The Data Science Seminar is a university-wide effort bringing together thought-leading speakers and researchers across campus to discuss

om A102

iety"

s FAIR"
Mary Stevens Neuroimaging and
f Neuro Imaging

rs misunderstand the world"
Arthur L. Carter Journalism Institute of

s for global good and social impact"
ent Strategy, Data, & Analytics

clouds to enhance health and drive

tor of the Institute for Systems Biology

~15-25 attendees
each week in first offering

Career Building

Data Science Career Fair

Share

The 2017 UW Data Science Career Fair will be held on Monday, October 23 from noon-5pm in the HUB Ballroom.

Multiple UW departments are working together to create this opportunity for undergraduates, graduate students, and postdocs working in data science to connect with companies and research labs looking for permanent employees and summer interns in data science positions. Please register in advance for the event **at this link**.

Nearly 900 students and 45 companies

UW Data Science Summit

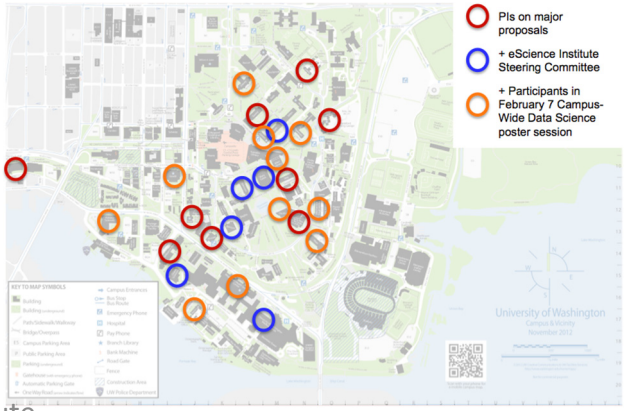
April 3 & 4 in the HUB

Join us for the inaugural **2018 UW Data Science Summit**! This event will be held **Tuesday, Apr. 3** and **Wednesday, Apr. 4** in the **Husky Union Building** (HUB). A unique educational opportunity for students, faculty, staff and community members, the Summit will feature:

- Prominent experts discussing data science
- Tutorials, break-out sessions, working groups, a poster session, and lightning and industry talks
- An instructional on how a department can make its own data science option
- Prizes for best-of posters and talks
- Networking opportunities and celebratory receptions

Preregistration is required and space is limited! Register now on EventBrite: **https://www.eventbrite.com/e/uw-data-science-summit-tickets-43217951055**!

255 attendees, although only 216 spots!



Integration with eScience a Key to Success

- Three eScience seminar series
- eScience short courses & hack weeks
- eScience social events
- Interactions with eScience postdocs
- Interactions with eScience research and data scientists
- Use of Data Science Studio for activities
- **Natural & neutral hub for the program**

"Data Science for ALL"

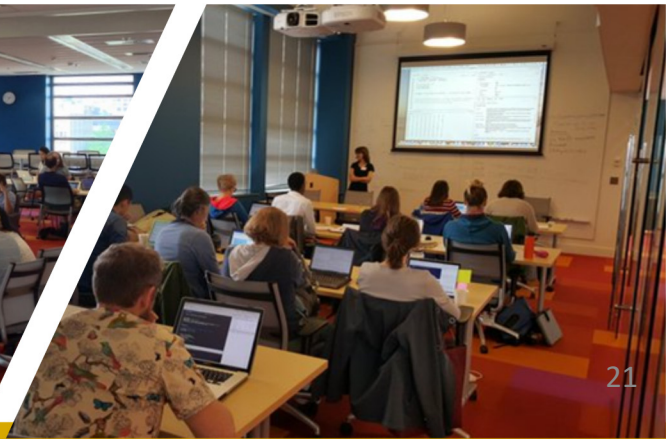
Formal Programs

Short Courses, Workshops, & Seminars

eScience Education Working Group

Informal Data Science Education

- **Open Office Hours**
- **Seminars**
- **Tutorials, workshops, and hack weeks**
  - Astrohack, neurohack, geohack, oceanhack,...
  - Software & data carpentry
  - Cloud 101
  - …



Various short courses and workshops

Scientific Computing with Python

Graduate seminar course offered jointly through the UW Astronomy & Applied Math departments designed as a comprehensive introduction to scientific computing in Python, geared toward graduate students, postdocs, and researchers in scientific fields which depend on analysis of large datasets.

Learn More about Scientific Computing with Python

Software Carpentry

Software Carpentry is a volunteer organization whose goal is to make scientists more productive, and their work more reliable, by teaching them basic computing skills. Founded in 1998, it runs short, intensive workshops that cover program design, version control, testing, and task automation. The Software Carpentry Foundation was created in October 2014 to act as a governing body for the project.

LEARN MORE ABOUT SOFTWARE CARPENTRY >

Community Data Science Workshops (CDSW)

The CDSWs are a series of workshops designed to introduce some of the basic tools of programming and analysis of data from online communities to absolute beginners. The CDSW have been held four times in Seattle in 2014 and 2015. So far, more than 80 people have volunteered their weekends to teach more than 350 people to program in Python, to build datasets from Web APIs, and to ask and answer questions using these data.

LEARN MORE ABOUT CDSWS >

Documentation for Cloud Computing and Data Science

> **Summary:** Cloud computing tutorials, case studies, pathways into data science technology; presenting the public cloud as opportunity for innovation. This page is intended to accelerate research; developed in close association with the [UW eScience Institute](http://escience.washington.edu). We have two goals - Reduce friction, i.e. help with what you already do with computers, streamline your analysis and pathways to publication. Promote new ideas and methods, i.e. We want to make it easy for you to share data, collaborate, sandbox, and learn about new methods in data science.

Help us improve this website, take a short survey here.

Announcements & News

- Cloud 101 for Google Cloud Platform will be held on Feb. 13, 2018 from 9am to 12pm. Registration now open.
- Cloud 101 for Amazon Web Services, Microsoft Azure & Google Cloud Platform will be held from Nov. 13 - Nov. 15, 2017 ** REGISTRATION IS NOW CLOSED. WE HAVE FILLED TO CAPACITY**
- GeohackWeek was held from Sept 11-15, 2017
- Cloud 101 for the Foster School of Business was presented on May 25, 2017.
- Research Computing Cloud 101 Immersion Day was April 4, 2017.

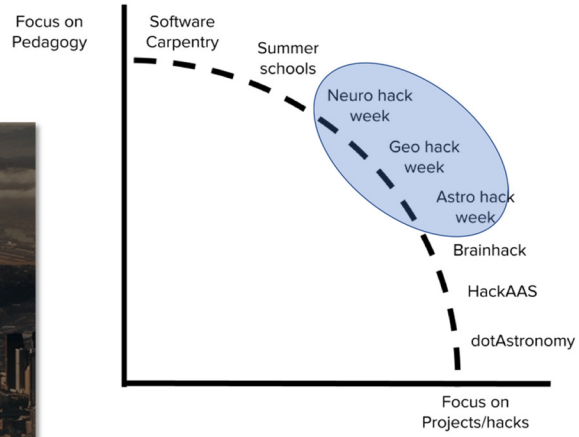
Courses are typically 10-60 people

Total for this academic year over 12 events and ~400 attendees

Hack Weeks


ASTRO HACK WEEK 2018
AUG 6 — AUG 10, 2018
LORENTZ CENTER@OORT, LEIDEN, THE NETHERLANDS
APPLY HERE


GEOHACKWEEK 2018
WORKSHOP ON GEOSPATIAL DATA SCIENCE
UNIVERSITY OF WASHINGTON ESCIENCE INSTITUTE
SEPT 10 - 14, 2018
SIGN UP


Focus on Pedagogy
Software Carpentry
Summer schools
Neuro hack week
Geo hack week
Astro hack week
Brainhack
HackAAS
dotAstronomy
Focus on Projects/hacks


NEUROHACKADEMY
ABOUT INSTRUCTORS APPLY FAQS CONTACT
NEUROHACKADEMY
SUMMER SCHOOL IN NEUROIMAGING AND DATA SCIENCE
University of Washington eScience Institute
July 30th - August 10th, 2018

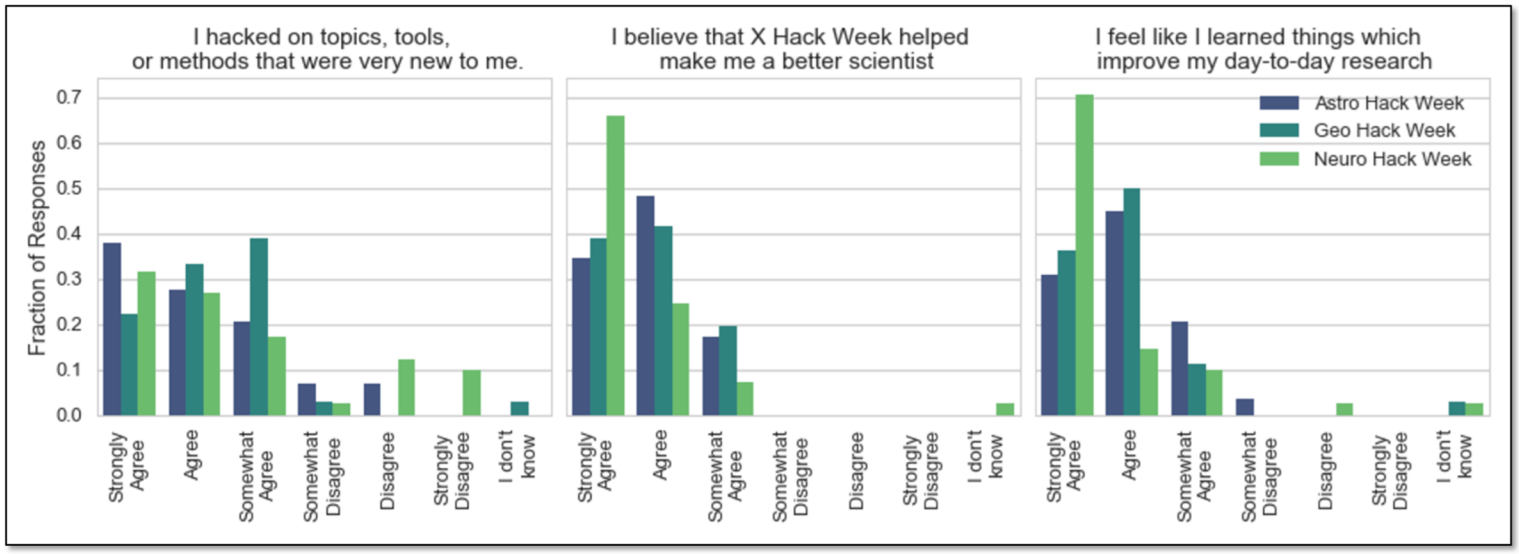

OCEANHACKWEEK 2018
UNIVERSITY OF WASHINGTON
AUG. 20 - 24, 2018
APPLICATION COMING SOON.

Hack Weeks

https://github.com/uwescience/HackWeek-Writeup

Hack Weeks as a model for Data Science Education and Collaboration

Daniela Huppenkothen[a,b,c,1], Anthony Arendt[d,e], David W. Hogg[b,a,f,g], Karthik Ram[h], Jake VanderPlas[e], and Ariel Rokem[e]



Participants

- AstroHack: ~45
- NeuroHack: ~50
- GeoHack: ~50
- OceanHack:
  - Cabled-array hack: 25

Special Interest Groups

Career Paths and Alternative Metrics

Working Group Lead: Tyler McCormick

Identify "data science fellows" who might otherwise slip through the cracks or go to industry, and groom them for a new breed of faculty position by creating new roles for data science professionals on campus that are not subject to "publish or perish," and that are equipped to pursue pragmatic, high-impact software-oriented data science projects.

LEARN MORE ABOUT CAREER PATHS AND ALTERNATIVE METRICS >

Education and Training

Working Group Leads: Bing Brunton & Steve Brunton

Establish alternative mechanisms that are free from departmental politics and conventional structures: boot camps, summer schools, tutorials.

LEARN MORE ABOUT EDUCATION AND TRAINING >

Software Tools, Environments, and Support

Working Group Lead: Jake Vanderplas

Successful projects are characterized by a balance between specialization and generality: sufficiently focused to actually solve a problem, but with the ability to scale to enough users or enough domains to amortize the cost of the initial development. We seek to institutionalize these patterns of success to help deliver the "next 100 Sloan Digital Sky Surveys."

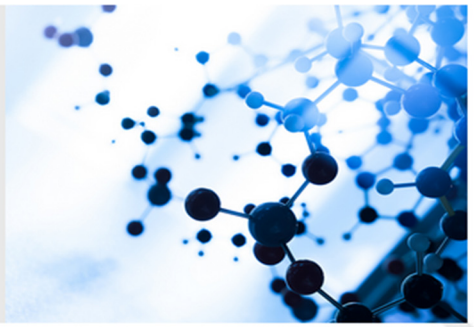
LEARN MORE ABOUT SOFTWARE TOOLS, ENVIRONMENTS, AND SUPPORT >

Reproducibility and Open Science

Working Group Lead: Ariel Rokem

Establish a culture of reproducibility and open science, and develop tools to support an environment where researchers find both the tools and the best practices for their research to be openly accessible to society and fully reproducible, more effectively feeding a productive cycle of research.

LEARN MORE ABOUT REPRODUCIBILITY AND OPEN SCIENCE >



Data Science Studies

Working Group Leads: Anissa Tanweer & Cecilia Aragon

UW Data Science Studies is a group of cross-disciplinary researchers studying the sociocultural and organizational processes around the emerging practice of data science.

LEARN MORE ABOUT DATA SCIENCE STUDIES >

Working Spaces and Culture

Working Group Leads: Micaela Parker & Sarah Stone

Establish new physical spaces on our campuses, specifically designed to meet the new requirements of data science activities, which in many cases will flourish best outside of traditional departmental boundaries.

LEARN MORE ABOUT WORKING SPACES AND CULTURE >

New Research-Themed Special Interest Groups

Neuroinformatics

Working Group Leads: Ariel Rokem

The neuroinformatics working group at the University of Washington (UW) eScience Institute and the University of Washington Institute for Neuroengineering (UWIN) focuses on neuroinformatics methods and their role in understanding the brain.

LEARN MORE ABOUT NEUROINFORMATICS >

NSF Tripods

ADSI
Algorithmic Foundations for Data Science Institute

Algorithmic Foundations for Data Science Institute (ADSI) Working Group

Working Group Lead: John Thickstun

Seek new algorithms and design principles that unify ideas and provide a common language for addressing contemporary data science challenges at this working group.

LEARN MORE ABOUT THE ADSI WORKING GROUP >

Efforts Specific to ECE (1/2)

- Working on adding Data Science Options
- Many data science courses
  - Convex Modeling and Optimization, cross listed by CSE/AA/ME, by Maryam Fazel
  - Data Science for Sequencing by Sreeram Kannan
  - Principles of Representation Learning by Sreeram Kannan
  - Data Science for Power Systems by Baosen Zhang

Efforts Specific to ECE (2/2)

- Algorithmic Data Science Institute
  - Funded by an NSF TRIPODS grant
  - ECE Prof. Maryam Fazel is co-director
  - Research and education activities
    - New courses, seminars, and summer schools
    - Integrating theoretical ML topics into domain-specific hack weeks
- eScience Executive Committee
  - Maryam is Senior Data Science Fellow & Ex Committee eScience
- Inter-disciplinary Data Science Lab in new CSE bldg

Conclusion

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

- Data science specializations
  - To ensure “data science for all” goal
  - Enables individual departments/schools to advertise
- Formal education with extra activities
- Focus on community, networking, and career
- Data science institute, eScience, core of program
  - eScience is natural home, core, and leader of program