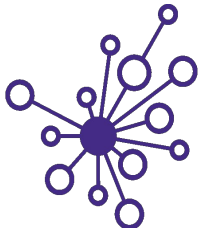
Data Science for Social Good

Information session for prospective student applicants

Anissa Tanweer
Research Scientist, eScience
Program Chair, DSSG

Info session outline

- Introduction to the eScience Institute
- Data Science for Social Good (DSSG)
  - Program overview
  - Application process
  - Previous projects
- Questions



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

Our Mission

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.

eScience Research Team

Director of Research

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

Data Scientists

Noah Benson
Ph.D. Biomedical & Health Informatics

Bernease Herman
B.S. Statistics
Formerly SE at Amazon

Jose Hernandez
Ph.D. Measurement & Statistics

Valentina Staneva
Ph.D. Applied Mathematics and Statistics

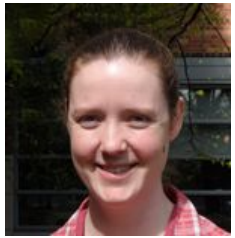
Amanda Tan
Ph.D. Civil & Env. Engineering

Anissa Tanweer
Ph.D. Communication

Research Scientists

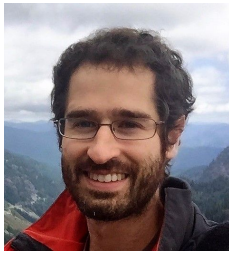
Anthony Arendt
Ph.D. Geophysics
APL

Nicoleta Cristea
Ph.D. Environmental Engineering



Bryna Hazelton
Ph.D. Astrophysics Physics

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



Scott Henderson
Ph.D. Geological Sciences

Vaughn Iverson
Ph.D. Oceanography

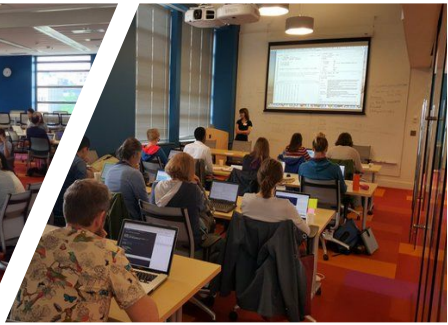
Spencer Wood
Ph.D. Zoology

Technical areas of eScience expertise

- New platforms, new algorithms, new methods, new datasets
- Working with large, heterogeneous, and noisy datasets
- Scalable analytics and predictive models
- Interactive visualization
- Code review, publishing, and reproducibility
- Online teaching materials, tutorials

We disseminate data science expertise and best practices

- Open Office Hours
- UW Data Science Seminar
- Tutorials, bootcamps, workshops, and hack weeks
  - Neurohackademy, Geohack, Waterhack, Oceanhack, etc.
  - Software carpentry (> 400 participants since we started counting in 2015)
- Winter Incubator
- Summer DSSG



DSSG Program

Modeled after similar programs with elements from our own Data Science Incubator.

DSSG Program Goals

- Figure out what it means to do “good” with data science
- Train students in data science methods
- Increase data science capacity across fields and organizations
- Positively impact society

Team Composition

- Student Fellows (4-5)
- eScience Data Scientist Leads (1-2)
- Project Leads (1-2)

Examples of Project Lead Affiliations

- **Academia**: E.g. University of Washington
  - Washington State Transportation Center
  - Disaster Data Science Lab
  - Architecture Department
- **Government**: E.g. Seattle Department of Transportation
- **Nonprofits**: E.g. Conservation International
- **Industry**: E.g. Bell Labs

What we expect from Project Leads

- Scoping meetings in preparation
- Co-presence 16 hrs/wk on average
    * Probably more during first 2 weeks
- Participation in program-wide sessions & meetings
- Domain expertise
- Stakeholder engagement
- Ability to discuss and promote work
- Open & reproducible when possible (Github)
- Description of project on our website
- Acknowledgment of the program in publications & authorship for team members

What we expect from Student Fellows

- 40 hours/week ($7,000 stipend)
- Current student, grad and advanced undergrad
- Baseline programming and stats knowledge
- Eligible to work in US (can’t support visas)
- Strong personal statement
- Team player

Micron Opportunity Scholarship

- Optional and separate application process
- Supplemental award
- For students facing barriers to participation
- https://escience.washington.edu/micron-opportunity-scholarship

What you can expect from eScience

- Committed project leads doing impactful work
- Data scientists highly experienced in cross-disciplinary mentoring
- Expertise in (non-exhaustive):

Machine learning	Statistical inference
Databases	GIS
Modeling	Optimization
Visualization	Cloud computing

- Best practices in version control, reproducibility and human-centered design
- Holistic data science curriculum
- Professional networking and career exposure
- Support in promoting and disseminating your work

Previous Tutorials

- Intro to Git & GitHub
- Git and Git Workflow
- Team Management Processes
- Pandas, Geopandas, and SQL
- Python Coding Standards and Documentation
- Unit Tests
- Project Organization, Virtualization, Continuous Integration
- Pair Programming
- Object Oriented Python
- Software Design
- Machine Learning
- Web Design and Web Apps
- Cython/Dask/High Performance Python
- Vega/Altair
- Data Visualization with Tableau
- An Introduction to Visual Communication

Previous Workshops

- Introduction to Data Science for Social Good
- Team Development
- Preparing for Stakeholder Engagement
- Stakeholder Analysis & Speculative Ethics
- Ethical Agency in Data-Rich Organizations
- Best Practices in Public Speaking

Other Activities

- Project Spotlights
- Career Conversation
- Social Theory Talks
- Docathons
- Field research and stakeholder engagement
- Social events

Overview of DSSG Program Structure

First Two Weeks

- Mandatory team development workshops
- Front-loaded tutorials

Rest of Summer

- Occasional tutorials (can be on-demand)
- Weekly “project spotlight” meetings and program check-in
- Bi-weekly leadership meeting with all PL’s, DS’s and administrators
- Visits and calls with stakeholders
- Project work and regularly scheduled team meetings

End of Summer

- Final presentations and celebration

Running DSSG Remotely

- High degrees of satisfaction with last year's remote offering
- Zoom, Github, Slack, Google Calendar, and other technologies
- Schedule program-wide meetings with consideration for multiple time zones
- Flexibility in coordinating teamwork necessary
- Intentionality in cultivating social connectedness
- Support for students in challenging physical environments

Call for Applications is open now!

General Info:
https://escience.washington.edu/dssg-proposal

FAQ's:
https://escience.washington.edu/dssg-student-faq/

Application Form:
https://form.jotform.com/203355090666153

Questions:

We encourage you to reach out *after* reading all program materials and FAQs. with any further questions: Program Coordinator Emily Keller <efkeller@uw.edu>

Selection Process

- Initial screening of applications for threshold criteria
- Committee review of threshold candidates
- Interviews with top candidates (~20% of total applicant pool) - Late March
- Advancement to short-list (~50% of interviewees)
- Solicitation of further info from short-listed candidates - Early April
- Notification of admission offer or waitlist - Mid April

What We Look for in DSSG Fellows

In individuals:

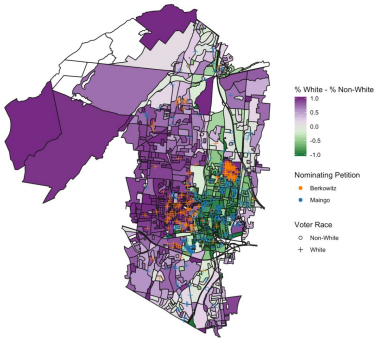
- Motivation for wanting to participate
- Baseline programming and research methods training
- Strong teamwork
- Experience with research and “social good”
- Commitment and contribution to diversity and inclusion

Across the cohort:

- Range of disciplinary backgrounds and expertise
- Range of technical abilities
- Range of educational experience levels

We take a broad view of what counts as data science

Proof of Concept ⟷ Polished Product



Mining Online Data for Early Identification of Unsafe Food Products

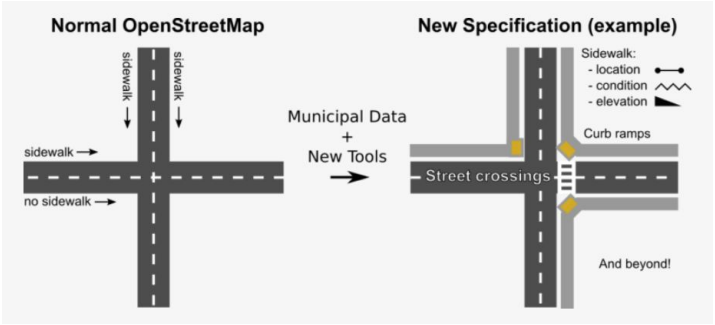
- DSSG 2016

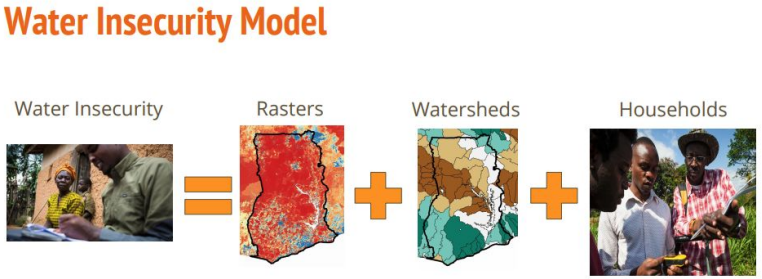
Detection of Vote Dilution: New tools and methods for protecting voting rights

- DSSG 2020

Proof of Concept
Infrastructural

Polished Product
Analytical





Global Open Sidewalks: Creating a shared open data layer and an OpenStreetMap data standard for sidewalks

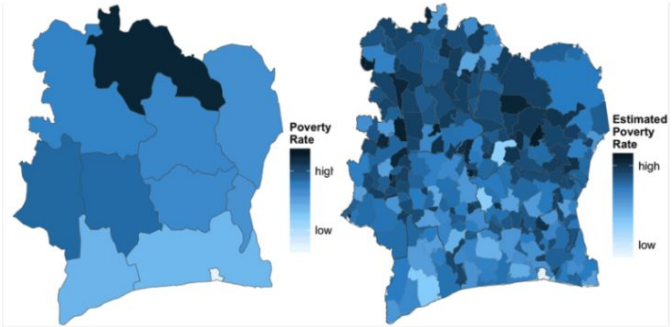
- DSSG 2016

Strengthening Capacities, Knowledge and Data Sharing Platforms for Sustainable Development

- DSSG 2017

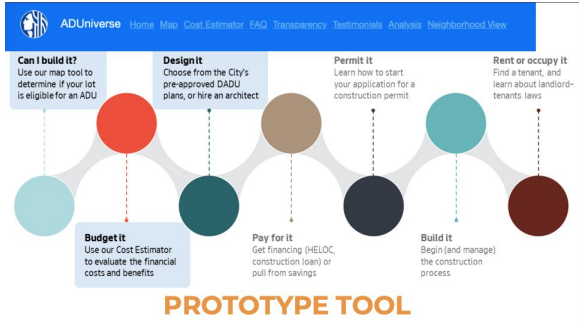
Proof of Concept
Infrastructural
Novel Data

Polished Product
Analytical
Traditional Data



CrowdSensing Census: A heterogeneous-based tool for estimating poverty

- DSSG 2016

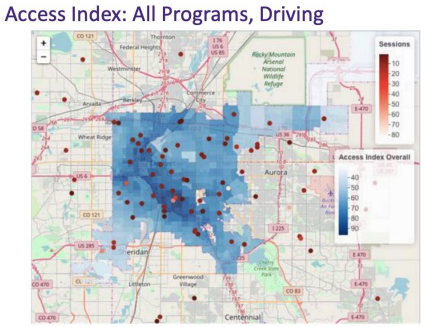


ADUniverse: Evaluating the Feasibility of (Affordable) Accessory Dwelling Units in Seattle

- DSSG 2019

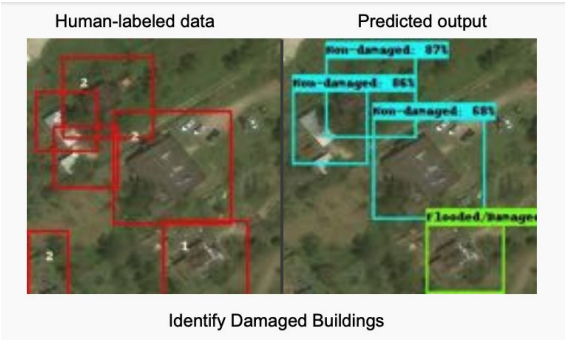
Proof of Concept
Infrastructural
Novel Data
Inference

Polished Product
Analytical
Traditional Data
Automation



Access to Out-of-School Opportunities and Student Outcomes

- DSSG 2018



Automatic Damage Annotation on Post-Hurricane Satellite Imagery

- DSSG 2018

Proof of Concept	⟷	Polished Product
Infrastructural	⟷	Analytical
Novel Data	⟷	Traditional Data
Inference	⟷	Automation

Important Dates

Now	Student applications open
Jan. 26	Student Info Session
Feb. 15	Student Fellow Applications due ***
Mid March	Notification of interview invitation
Late March	Interviews held
Early April	Notification of shortlisting
Mid April	Notification of admission offer or waitlist status ***
June 14	First day of program
August 21	Last day of program

Questions?

UNIVERSITY of WASHINGTON
eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

Anissa Tanweer
Research Scientist, eScience
Program Chair, DSSG