# Hawaiʻi Data Science Institute
# Data Science Fellows Program

Sept. 2, 2022

# Agenda

- Welcome!

- Introductions
  - Instructors and staff

- Overview of the program
  - Fall/Spring Semester activities
  - Summer immersion - CITRUS

- Onboarding tasks

- Ice breaker – intro to SAGE3

# Introductions
# Instructors and staff

- Workshop Instructors
  - Jason Leigh, Professor, Computer Science
    - Data Vizualization
  - Sean Cleveland, Assoc Director, Cyberinfrastructure
    - Software engineering
  - Alex Stokes, Asst Prof, Cell & Molecular Cell Bio/JABSOM
    - Data Science
  - Peter Sadowski, Asst. Prof Computer Science
    - Data Science – Machine learning/AI
  - Mahdi Belcaid, Asst. Prof Computer Science
    - Data Science – Bioinformatics
  - Philip Johnson, Prof. Computer Science
    - Software engineering, pedagogy
  - Laura Tipton, Asst Professor – Chaminade
    - Data Science/Microbiome
- Staff
  - Michelle Choe, Maria Dumanlang, Shivani Tanaka

HAWAIʻI DATA SCIENCE

CHANGE HAWAIʻI

HAWAIʻI EPSCoR

UNIVERSITY OF HAWAIʻI

# Program Overview

- Mentored research project
  - Domain science + data science

- Workshops and training – CI/Data Science Skills
  - Co-development with a mentor
  - Morea framework – teaching pedagogy

- Professional Development activities
  - Resume development, social media, linked-in

- Research Showcase
  - Present your work, meet industry partners

datascience.hawaii.edu



# Data Science Fellows Alumni 2018 - 2021

## PROGRAM BY THE NUMBERS

**31** STUDENTS

**32%** WORK IN INDUSTRY

**68%** WORK IN ACADEMIA

## FELLOWS TRAINED

- Kyle Hart, Database Engineer at Institute for Astronomy UHM
- Jonathan Bennet, UH Cancer Center
- Oscar Ramfelt, Dept. Microbial Oceanography UHM
- Moseli Motsoehli, Dept. Information & Computer Sciences UHM
- Katherine Ackerman, Dept. Atmospheric Sciences UHM
- Nanami Kimoto, John A. Burns School of Medicine
- Sofia Ferreira, Aqualink Community Manager, San Francisco, CA
- Yannik Glasser, Dept. Information & Computer Sciences UHM
- Anton Nikolaev, Machine Learning Scientist at Saks Fifth Avenue, NY
- Alberto Gonzalez Martinez, Entrepreneur & Founder RendezView Inc
- Troy Wooton, Hawaii State Energy Office
- Matt Lucas, Climate Data Analyst, UHM
- Elizabeth Fink, Materials Management Data Analyst at Hawaii Pacific Health
- Kaitlyn Terayama, Medical Data Entry Specialist at Rhodes Group
- Justin Fletcher, Chief Engineer at Maui High Performance Computing Center
- Amihan Camson, Project Manager at First Hawaiian Bank
- Paul Spooren, Dept. Information Technology Services UHM
- Layne Fujioka, Belle 2 Team, KEK Tsukuba, Japan
- Charlotte Smith, Dept. Natural Resources & Environmental Management UHM
- Trista Mckenzie, Dept. Geology & Geophysics UHM
- Cedric Arisdakessian, Dept. Information & Computer Sciences UHM
- Jaclyn Lee, Federal Reserve DC
- Charles Dickens, Dept. Computer Science & Engineering UC Santa Cruz
- Sean Takafuji, Microsoft

# Intro to Morea

- We will be using the Morea framework for managing assets and developing workshops

- https://change-hi.github.io/

- https://change-hi.github.io/modules/

# Fall/Spring Semester Schedule

**HAWAI'I DATA SCIENCE**

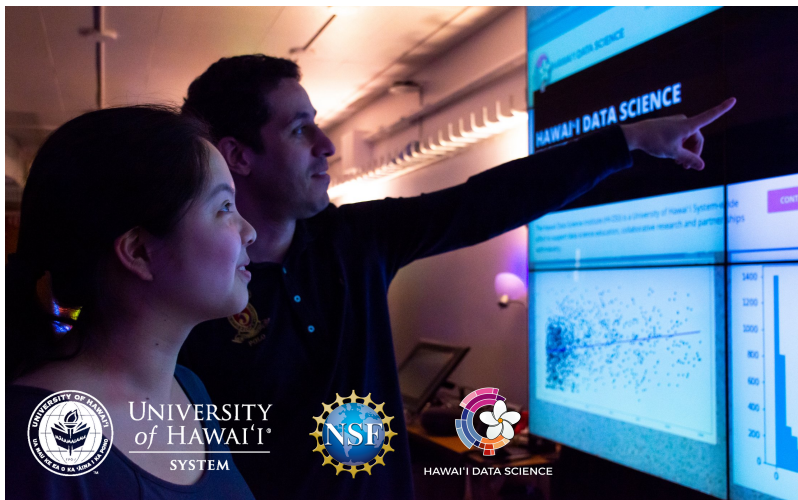| Date | Meeting Agenda |
|---|---|
| Fri. 9/2 | Introduction/Icebreaker & Kick-Off Meeting / Onboarding Exercise / Semester Expectations |
| Fri. 9/9 | Fellows Meeting: Workshop Repository & Practice Instructions |
| Fri. 9/16 | Fellows Meeting: Professional Development: Science Communications / Introduction to Research Showcase Requirements |
| Fri. 9/23 | Fellows Meeting: Workshop Check-In & Work Session / SAGE3 and HIDSI Activities |
| Fri. 9/30 | Seminar |
| Fri. 10/7 | Seminar - Kamuela Enos (hold for 2-4) |
| Fri. 10/14 | Fellows Meeting: Research Showcase Practice Presentations |
| Fri. 10/21 | Workshop Slot (FAIR Data Management Security and Ethics) |
| Fri. 10/28 | Workshop Slot (Scientific Software Basics) |
| Fri. 11/4 | Workshop Slot (HPC) |
| Fri. 11/11 | No Meeting: Veteran's Day |
| Fri. 11/18 | Fellows Meeting: Research Showcase Practice Presentations |
| Fri. 11/25 | No Meeting: Non-Instructional Day |
| Fri. 12/2 | Workshop Slot (Data Movement Dissemination and Archiving) |
| Fri. 12/9 | Research Showcase (2:00pm-5:30pm) |
| Fri. 12/16 | No Meeting: Final Exams |

| Fri. 1/13 | Kick-Off Meeting |
|---|---|
| Fri. 1/20 | Seminar |
| Fri. 1/27 | Workshop Slot |
| Fri. 2/3 | Workshop Slot (Data Wrangling) |
| Fri. 2/10 | Seminar |
| Fri. 2/17 | Workshop Slot (Machine Learning) |
| Fri. 2/24 | Workshop Slot (Visualization) |
| Fri. 3/3 | Workshop Slot |
| Fri. 3/10 | Fellows Meeting: Research Showcase Practice Presentations |
| Fri. 3/17 | No Meeting: Spring Recess |
| Fri. 3/24 | Workshop Slot (Smart Sensor Data Collection) |
| Fri. 3/31 | Workshop Slot (Creative Thinking) |
| Fri. 4/7 | No Meeting: Good Friday |
| Fri. 4/14 | Fellows Meeting: Research Showcase Practice Presentations |
| Fri. 4/21 | Workshop Slot (Scientific Workflows and Gateways) |
| Fri. 4/28 | Fellows Meeting: Professional Development |
| Fri. 5/5 | Research Showcase (2:00pm-5:30pm) |
| Fri. 5/12 | No Meeting: Final Exams |

CHANGE HAWAI'I

HAWAI'I EPSCoR

UNIVERSITY OF HAWAI'I

## HAWAI'I DATA SCIENCE

# HAWAI'I DATA SCIENCE INSTITUTE

# CITRUS DATA SCIENCE SUMMER PROGRAM

**Apply!**

Apply for a one-month summer program for undergraduates to develop their skills in data science, analytics and visualization.

## June 6 to 30, 2022
## 9:00 am - 3:00 pm

Online pre-training sessions (4 hours max) weeks of May 23 and May 30, 2022

**Stipend**
**$3,250**

**Eligibility**
Undergraduate students from Hawai'i & the Pacific at any institution (all majors are eligible, non-computing majors are encouraged to apply)

**Apply by April 15, 2022**
http://go.hawaii.edu/3pV

**Questions?**  email: datasci@hawaii.edu

CHANGE HAWAI'I

HAWAI'I EPSCoR

UNIVERSITY OF HAWAI'I

# LIFECYCLE OF A
# DATA SCIENCE PROJECT

Ever heard the phrase "Here's some data, can you find some insights?" Right? Too often stakeholders approach Data Scientists with vague or even undefined goals. Understanding the end goal is very important and sets up the rest of the project for success.

By far, everybody's least favourite stage, but perhaps the most important one. Data can come from many sources, be in the wrong format, have anomalies and a myriad of other problems. A single mistake in this stage can render the rest of the analysis useless.

Creating models, performing data mining, running text analytics, setting up simulations - the list goes on! This is the most fun and exciting part and if the previous stages have been done correctly, analyzing the data and deriving insights will feel like a breeze.

**1 IDENTIFY THE PROBLEM** — 10%

**2 PREPARE THE DATA** — 70%

**5 PRESENT FINDINGS** — +80%

**3 ANALYSE THE DATA** — 10%

**4 VISUALIZE INSIGHTS** — 10%

We've reached 100% the project is over! Actually, not yet. Presenting findings is a whole separate "Bonus" stage. You need to not only convey the insights in your audience's language but also get buy-in from them to take action based on those insights. This is an art in its own right.

Visualizing comes hand-in-hand with analyzing. This is a very powerful technique as seeing the data in various forms and shapes can help uncover insights that are otherwise not evident. Also some projects such as BI dashboards don't require much analysis but rely heavily on visualization instead.

Join our Facebook group

SuperDataScience

# Onboarding

- Access all onboarding materials here:

- https://change-hi.github.io/morea/intro-ci-tracs/experience-onboarding.html

# Ice-breaker – intro to SAGE3

- Let's get to know each other!

UNIVERSITY OF HAWAI'I