DATA INSTITUTE

UNIVERSITY OF SAN FRANCISCO

Talk Organization

- About me
- Introduction of USF's Data Institute
  - USF's Masters of Data Science and the Practicum
- Practicum Case Studies
  - LACC
  - WRI ☹
  - Swiftly
  - SFCTA
  - Valor Water, PowWow and SunRun

Nicholas Ross

- ncross@usfca.edu
- Feel free to reach out with any questions!
- Director of the Practicum, Asst. Director of Partnerships and Asst. Professor of Data Science
- Before USF:
  - Director of Analytics at Sega
  - Director of Analytics at TinyCo (Video Game Startup)
- Disclaimer

DATA INSTITUTE

UNIVERSITY OF SAN FRANCISCO

About the Data Institute

- Umbrella organization that houses data initiatives at USF
- Mission:
  - Build an inclusive community of data scientists in the heart of San Francisco to advance research in data science
  - Continue to create innovative curriculum to support the training of the next generation of data scientists
  - Partner with nonprofit and civic organizations to seek data-driven solutions to address pressing social, economic and environmental challenges
  - Foster a new paradigm between industry and academia to tackle industrial data science problems

Broad Strokes

- Undergraduate Data Science and Masters of Data Science (more information later)
- Certificate courses on nights and weekends
- Trainings with local companies
- Consulting Services
- Conference
- Seminar Series

UPCOMING DATA INSTITUTE CERTIFICATES:

- Deep Learning
- Design and Analysis of Experienments
- Exploratory Data Analysis
- Intro to SQL
- Management of Data Analytics Teams

MS in Data Science

UNIVERSITY OF SAN FRANCISCO

An accelerated one-year program that delivers a rigorous curriculum focused on mathematical and computational techniques in data science

UNIVERSITY OF SAN FRANCISCO

What is MSDS?

- This is a **12-month** program: six seven-week modules and a two-week winter intersession course.
- Intensive: Outside work is not permitted while in the program. *Some weeks exceed **70+ hours** of work*.
- Begins **July 9, 2018** and ends the following June.
- It's a **35-credit** program. All classes are one or two credits.
- Supporting faculty come from **Computer Science**, **Mathematics**, **Statistics** and **Business**.

What is the curriculum?

- Boot Camp Experience
- Statistical Modeling
- Machine Learning
- Data Acquisition, Exploration, Management, Visualization
- Business: Strategy, Communication, App Development
- Practicum Experience

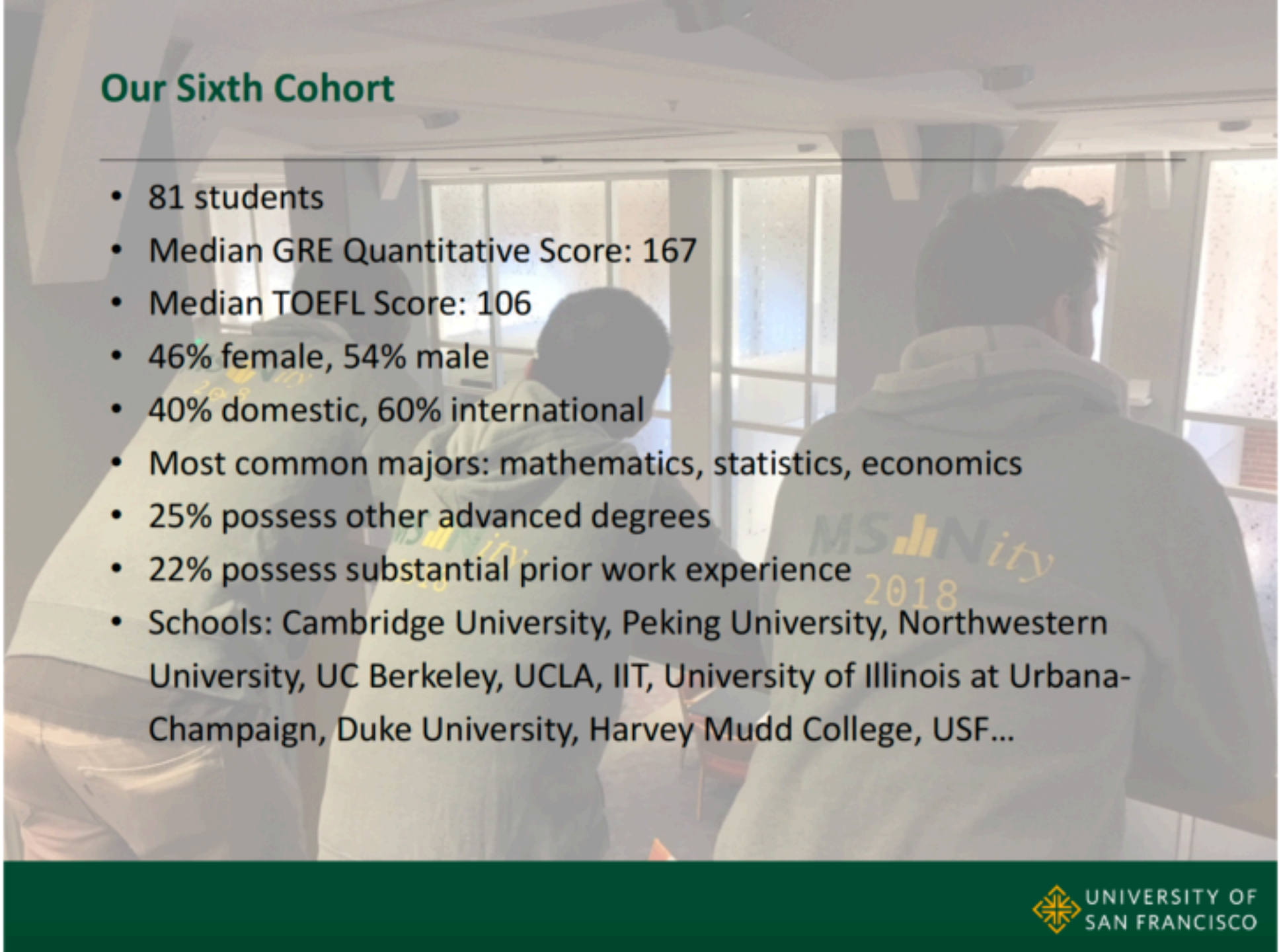
Employment Outcomes

98% of graduates from our first 4 cohorts received an offer of employment within three months of graduation.

97% of those actively looking for positions in our 2016-2017 cohort have received an offer of employment.

Class of 2017 Outcomes:

- Median salary: **$110K**
- Median salary, international students: **$115K**
- Median salary, women: **$110K**
- Median salary, Bay Area: **$115K**
- Median deferred compensation: **$19,625**



Our Sixth Cohort

- 81 students
- Median GRE Quantitative Score: 167
- Median TOEFL Score: 106
- 46% female, 54% male
- 40% domestic, 60% international
- Most common majors: mathematics, statistics, economics
- 25% possess other advanced degrees
- 22% possess substantial prior work experience
- Schools: Cambridge University, Peking University, Northwestern University, UC Berkeley, UCLA, IIT, University of Illinois at Urbana-Champaign, Duke University, Harvey Mudd College, USF…

The Practicum Program

- Similar to an internship, but with faculty mentorship.
- Faculty develop relationships with Bay Area companies.
- Students start practicums in the third module (mid-October).
- Teams are formed and assigned a mentor at the company and a faculty mentor at USF.
- Most students stay with the same company for 35 weeks. Some students change companies/projects once during the program.

Practicum Outcomes

- **21%** of our most recent cohort received an employment offer from their practicum company.
- **32%** believed they would have received an offer if they had interest in pursuing employment there.
- **15%** accepted full-time positions with their practicum company.

Who are some of our past practicum partners?

Who are some of our current practicum partners?

SEGA

Eventbrite

C1 Capital One Labs

UCSF University of California San Francisco

xoom A PayPal Service

UnitedHealthcare

mozilla

Nextdoor

Houston Astros

First Republic Bank It's a privilege to serve you®

FLYR

Recology Waste Zero

zipcar

49ers

kiva

Under Armour

Undergraduate Data Science

- One of USF's newest majors
- Includes a "Baby Practicum"

Practicum Projects of Interest

- LA County
- WRI
- Swiftly
- SFCTA
- Valor Water, PowWow and SunRun

Los Angeles County Project

- Worked with Ben Uminsky, Interim Division Manager, Los Angeles County Register-Recorder.
- Improve the efficiency of voting precincts by:
  - Identifying clusters of voting locations and assigning them to easily accessible Check in Centers.
- Data:
  - Information on voting booths, current check-in-centers and the time it took to check-in

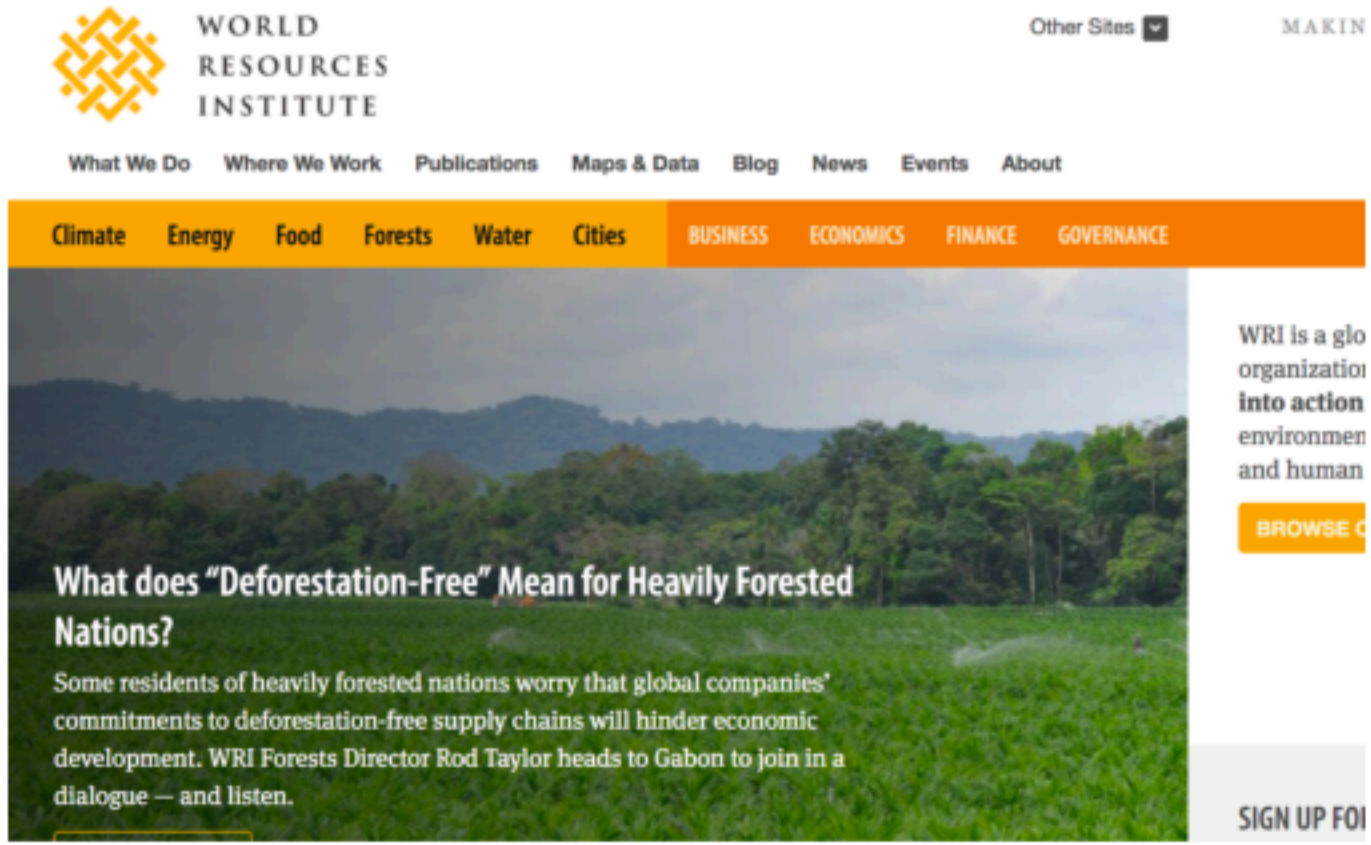
Los Angeles County Project (cont.)

- Check-in-centers serve multiple precincts and serve as locations where ballots are stored and counted after an election.
- Undergraduate students undertook the following project:
  - Mapped both CICs and voting precincts
  - Collected data on drive-times, number of precincts served and the amount of time it took to collect information
- Used k-means clustering (and some other techniques) to re-allocate the check-in-center locations to increase efficiency:
  - Lower the amount of time it takes to check-in
  - Lower the total distance travelled

Los Angeles County Project (cont.)

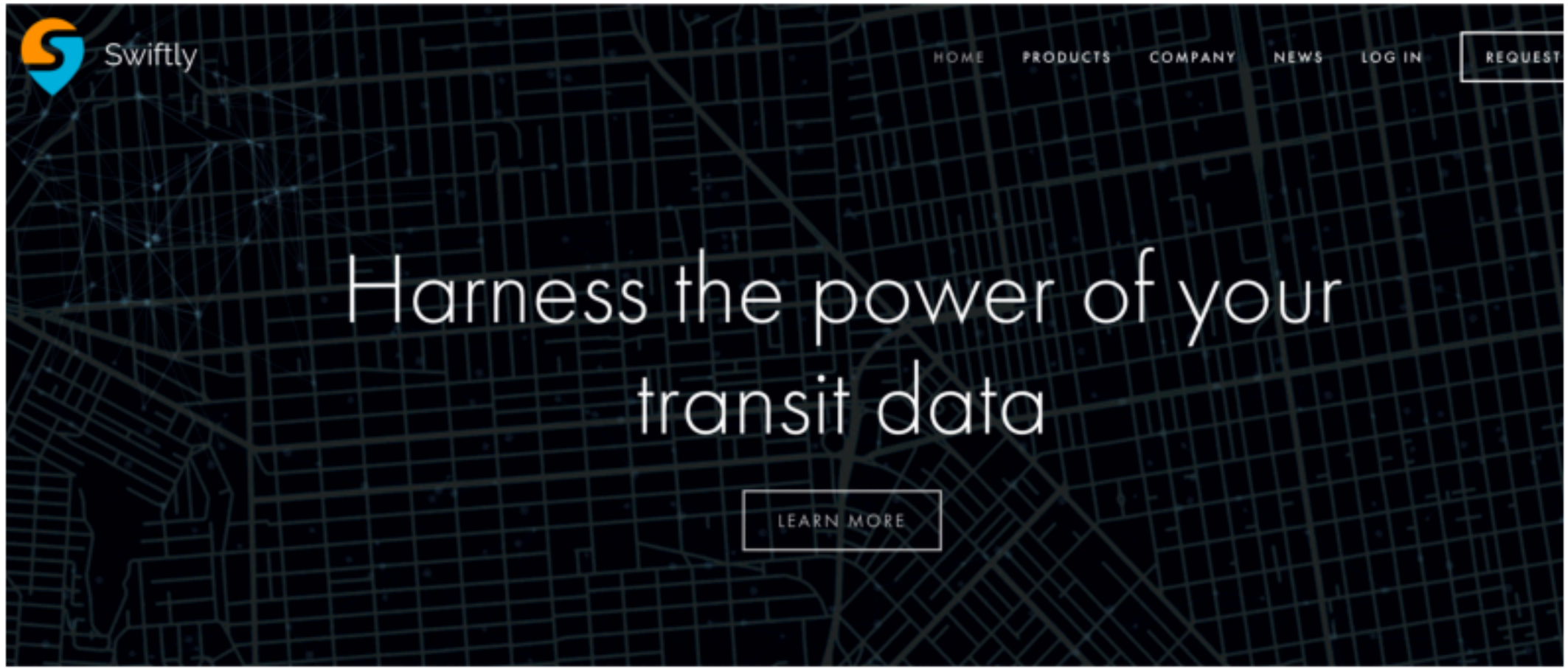
- Generate web page
- Demonstration

World Resources Institute



- Pitched a project, but was not able to staff
- Using deep learning and other neural networks techniques to identify deforestation

Swiftly


Swiftly
HOME
PRODUCTS
COMPANY
NEWS
LOG IN
REQUEST
Harness the power of your transit data
LEARN MORE

Swiftly

- Transportation startup which helps transit agencies be more efficient
- Example projects:
  - Identify how weather effects bus arrival times
  - Use Statistical Techniques to identify when a bus goes off route

San Francisco County Transportation Authority

- Congestion Management
  - http://tncstoday.sfcta.org/
- Clean Air
- Street Repair (some), pedestrian safety and some other transportation activities

Visualizing pedestrian safety

- Video Demonstration

Valor Water

- Start up which has access to information on water meters
- Goal:
  - Use anomaly detection techniques to predict if a water meter is using too much water

Sunrun

- Estimate the effect of snow on solar panels
- Previous method used a measure of Global Horizontal Irradiance, which has low accuracy with snow
- Using real-time snow reports and mapping information, the practicum team was able to better estimate how snow changes power output
- Hardest part – combining multiple sources of data

PowWow energy

- Predict water stress in trees using IOT data and aerial imagery
- Goal is to estimate the amount of stress on a tree – avoid a tree falling over by changing irrigation patterns

Technology Information

- All of the projects above used one of two stacks:
  1. R / R-shiny
  2. Python & Cloud technologies
- In the data science world, these represent >95% of the work being done
- Data formats that are proprietary and not available in these technologies face a significant disadvantage