# R Workflows in Azure Machine Learning for Athletic Data Analysis

Emily Kraschel

UNIVERSITY OF OREGON

Wu Tsai Human Performance Alliance
UNIVERSITY OF OREGON

# Data for High Performance Athletics

**Trying to keep athletes**

- Healthy
- Available
- Performing well

**Measured by**

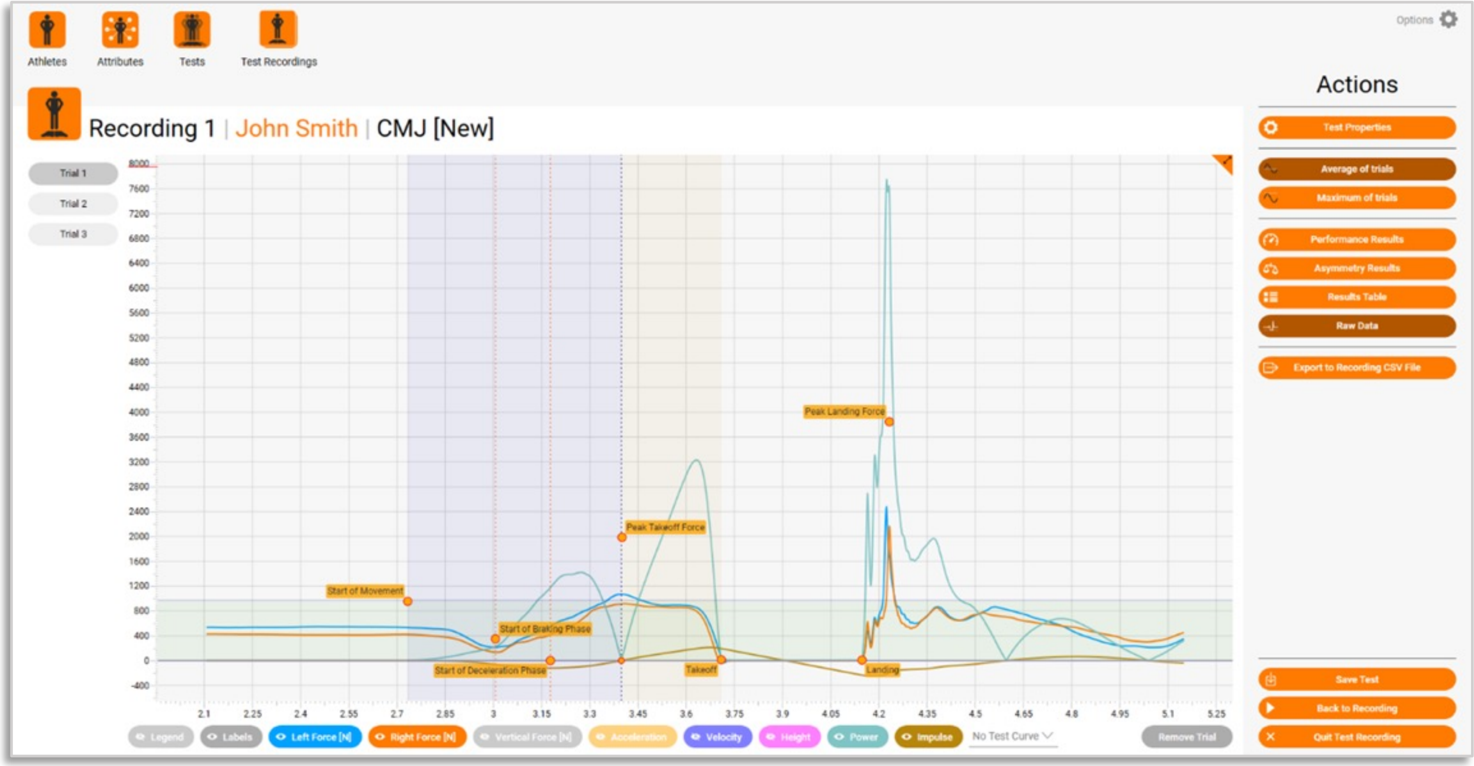
- Load & volume
- Performance metrics & tests
- Injuries

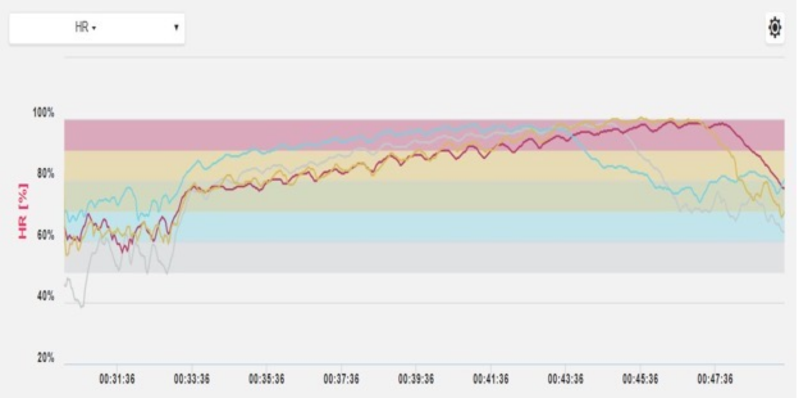
Online platforms host & automate reports
Data available to export manually to Excel or through API to R
Data used by Sports science, S&C, sport coaches, trainers and more

# Working Fast

## Day-to-Day Athletics Operations

- Data collection
    - Live, daily, weekly, longer
- Storage
- Basic analysis & reporting
- Decision making




# Working Fast

## Day-to-Day Athletics Operations

- Data collection
    - Live, daily, weekly, longer
- Storage
- Basic analysis & reporting
- Decision making

# Working Slow

## Data Science, Research, Long-Term Goals

- Data cleaning
- In-depth reports & aggregation
- Advanced analysis
- Connecting the Dots

**Conflicting Needs**

Changing data, constant additions, too many versions

Can’t have controlled experiment

Too many bottlenecks and stopping points

# Old Framework

**FAST**

- Automated reports via dashboard

**Difficult for Slow Work**

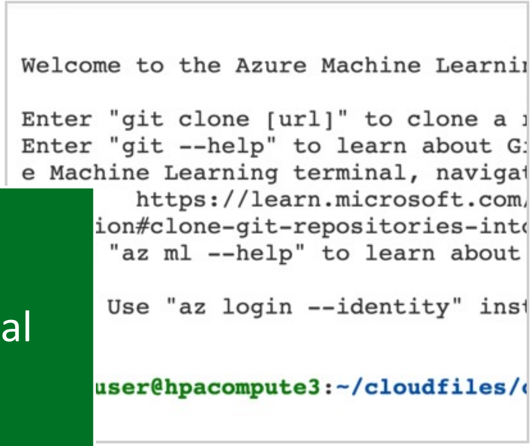
- What do we put on the dashboard? New metrics?
- No central source for raw instrument data
- Variety of IDs from different instrument interfaces
- Data in different versions, different places, different files
- CSV creation and pull was slow, too many CSVs….

# New Framework Goals

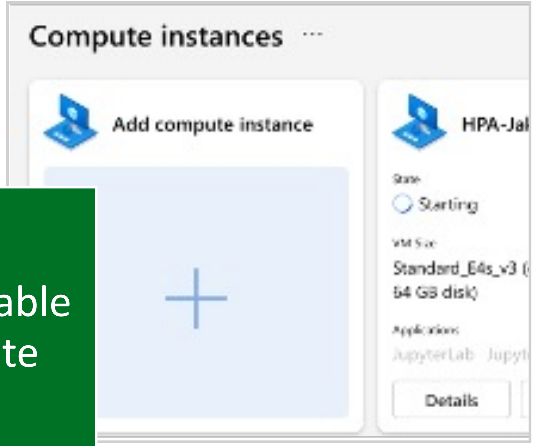
- Improve outcomes (more complete, valid, actionable data)
  - Lower barriers to more complex data science & analytics
- Aggregate data in one place
  - Easier data cleaning, validation, analysis
  - Unify player ID
- More compute power
  - Ability to collaborate & compute easier
- Less bottlenecks- duplication, updates, incomplete data
- Ability to share findings more easily

# What is Azure?

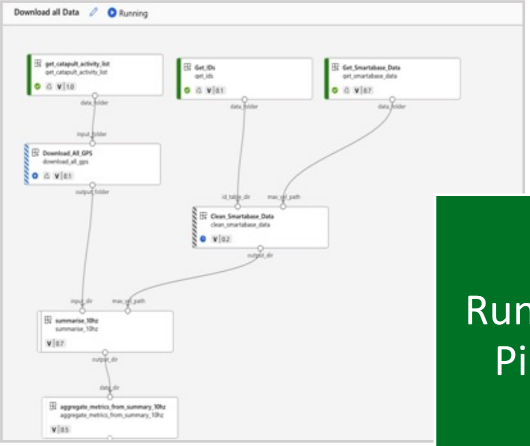

Analysis in Jupyter Notebook (R, Python, Julia)



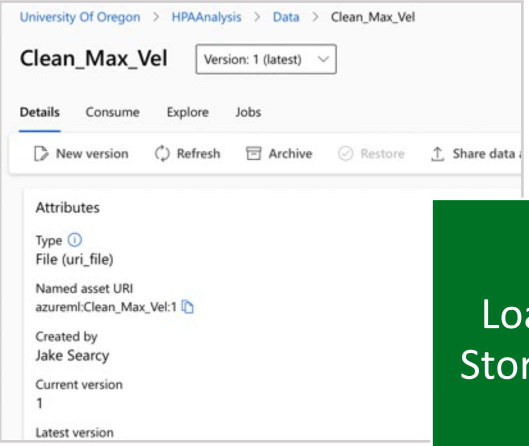

Terminal



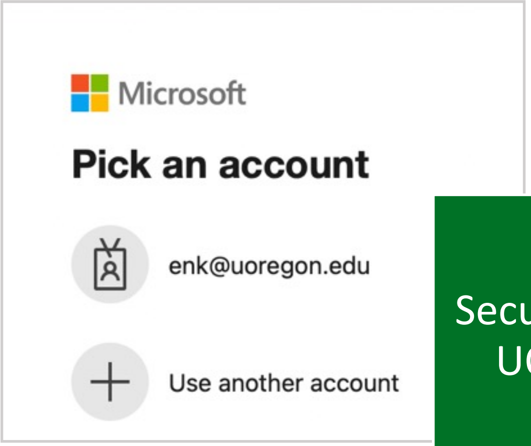

Configurable Compute




Running Job Pipelines




Loading & Storing Data




Secured with UO login

# New Framework

**FAST**

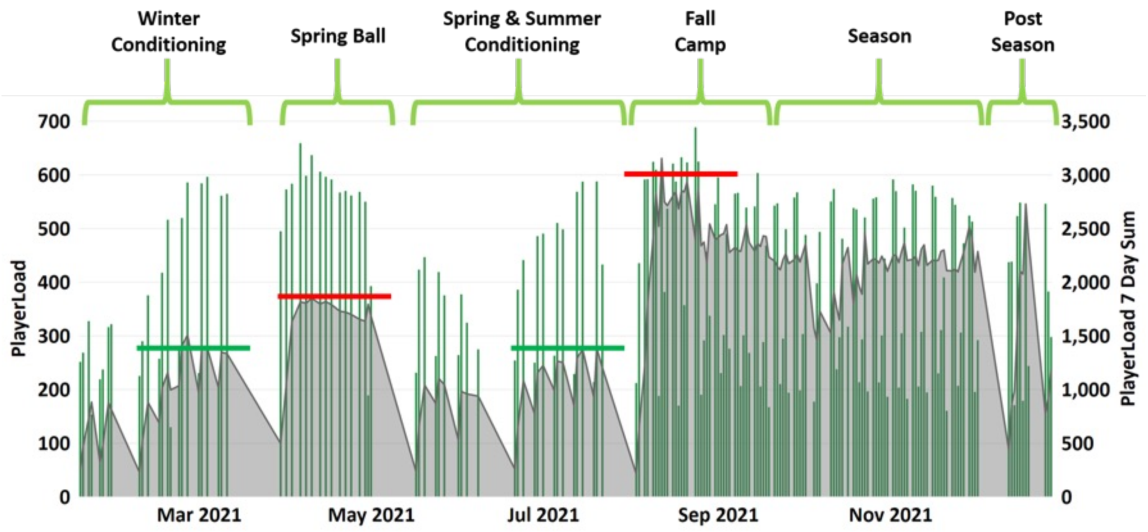
- Automated reports via dashboard → Still in use

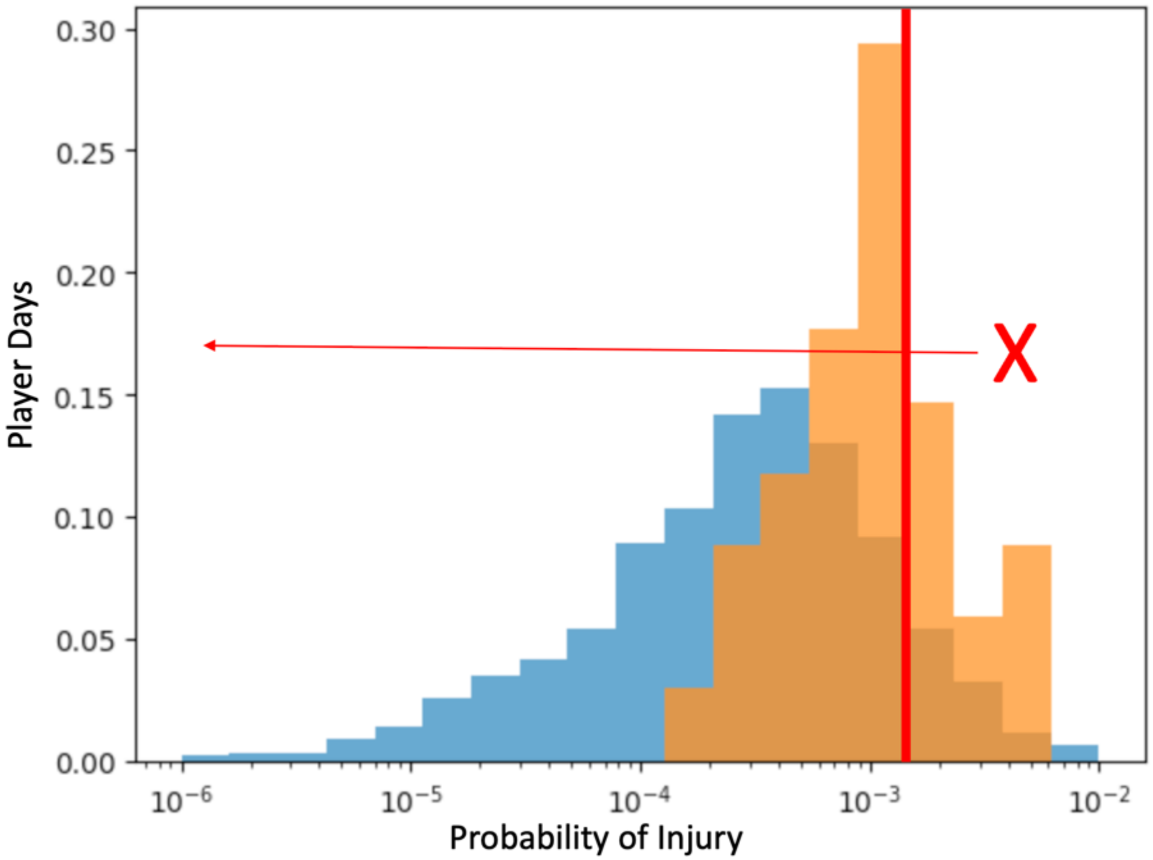
**Better for Slow Work**

- What do we put on the dashboard? New metrics? → Able to step back from constant stream of data
- No central source for raw instrument data → Able to pull from all APIs
- Variety of IDs from different instrument interfaces → Easier merging
- Data in different versions, different places, different files →Data Lives in one place
- CSV creation and pull was slow, too many CSVs…. → Easier to maintain the most recent version, quicker due to higher computing power

# What we are looking at now

- Hamstring injury predictability
- Discrete time survival models
- Differences between genders
    - Relationships between metrics
    - Response to training



# Countermovement Jump Analysis

<u>**Kinetic Variables of Interest**</u>

**Eccentric Rate of Force Development**

An athlete’s ability to generate force during the eccentric phase of the movement

**Jump Height**

The global output of the stretch shortening cycle

# Countermovement Jump Analysis

<u>**Kinetic Variables of Interest**</u>

**Eccentric Rate of Force Development**

An athlete's ability to generate force during the eccentric phase of the movement

**Jump Height**

The global output of the stretch shortening cycle

<u>**Research Questions**</u>

Is there a difference in relationship between the two variables of interest?

Can we change Jump Height or Eccentric Rate of Force Development over time as a result of training?

# Gender Differences

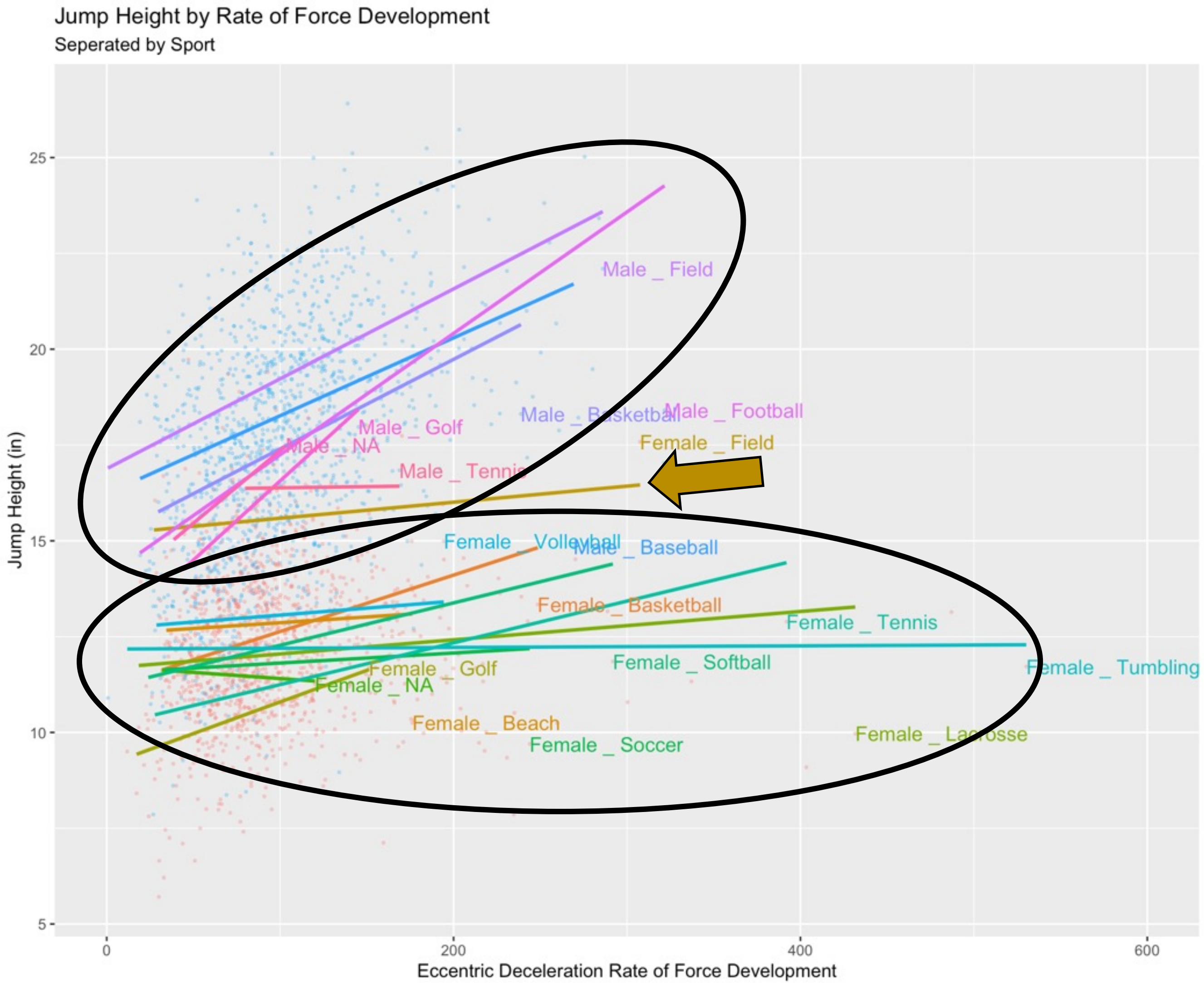
**Relationship between Eccentric Deceleration Rate of Force Development and Jump Height**

Higher Eccentric Deceleration Rate of Force Development (RFD) has stronger correlation with higher jump height for men than for women

# Regression Model

Jump Height regressed by RFD, sport (separated by gender) and interaction term between gender & RFD

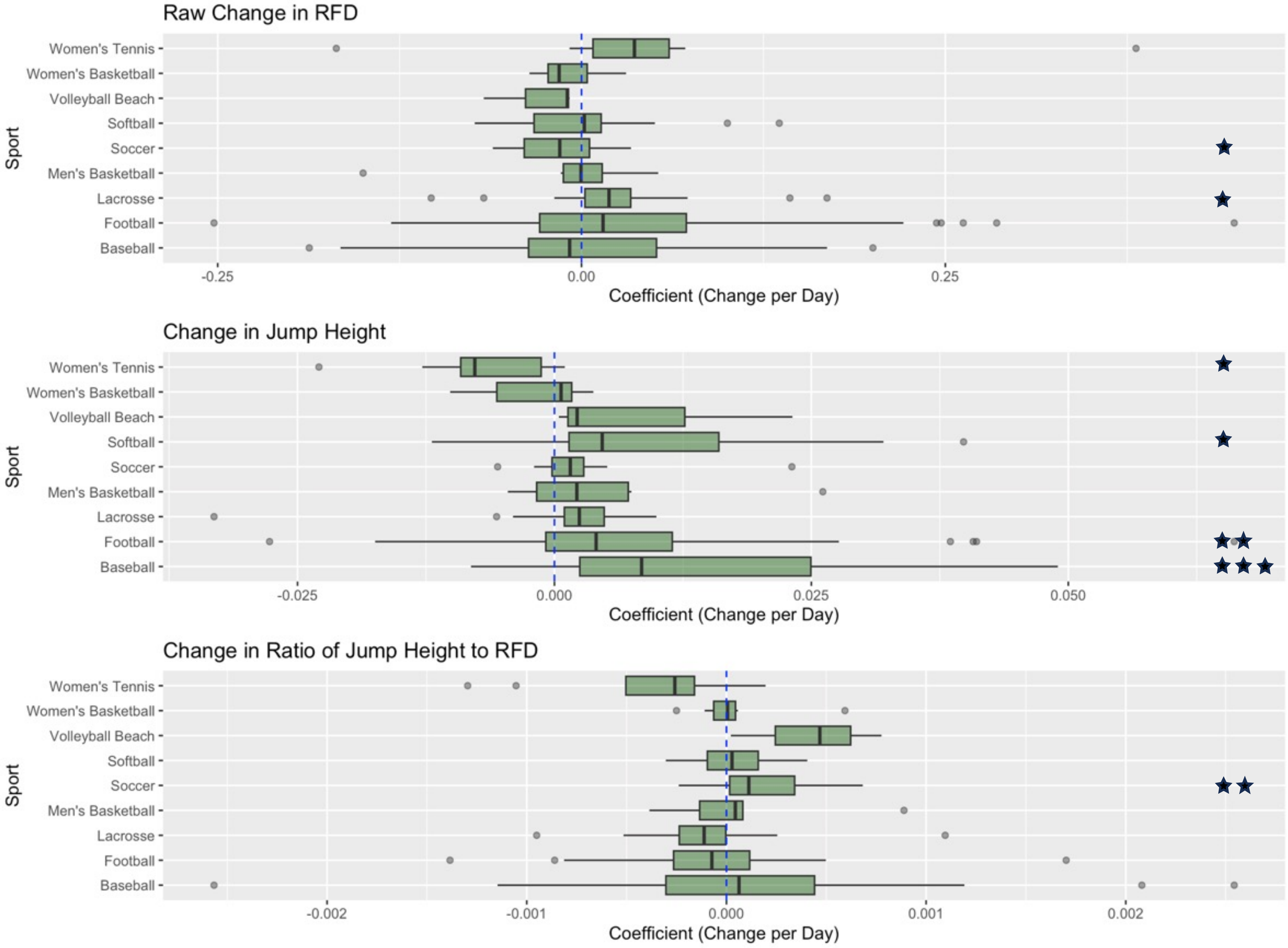
- Men’s sports generally have a significantly higher coefficient than the baseline, with a few exceptions
- Interested in sports and individuals that break this patten or improve over time
  - Can we train this?



# Training Response

Individual change regressed over their entire time at Oregon

- Trained two years at Oregon, 10 tests
- RFD over time
  - Decrease for Soccer, increase for Lacrosse
- Jump Height over time
  - Decrease for Women's Tennis, increase for Softball, Football, Baseball
- Ratio of Jump Height/RFD over time
  - Increase for Soccer



# Acknowledgements



Jordan Troester


Jake Searcy


Ben McKay


Lillian Aoki


Ethan Martin

UNIVERSITY OF OREGON

University of Oregon

Department of Data Science

Athletics

Wu Tsai Human Performance Alliance

Wu Tsai Human Performance Alliance