

Penn Football testing dashboard

# Building a dataviz pipeline for athlete evaluation



Anna Duan | MUSA Practicum | May 5, 2025

Professor Xiaojiang Li, Weitzman School of Design

Coach Miguel Rivera, Associate Strength and Conditioning Coach, Head of Football Performance

# Agenda

1. Motivation: leveraging testing data for athlete evaluation
2. Ideation: let's not oversell the numbers
3. Data preparation: balancing reproducibility vs practicality
4. Data visualization & design: prioritizing functionality over minimalism
5. Demo time
6. Acknowledgements

# 01 Motivation



Leveraging testing data for athlete evaluation



# Can testing data drive better decisions?

## Motivation

- Building a football roster takes years of decisions
  - Positions require distinct athletic profiles
  - Decisions are not always data-driven
- 5 years of annual testing data
- How can this data be scaffolded to simplify coaching decisions?

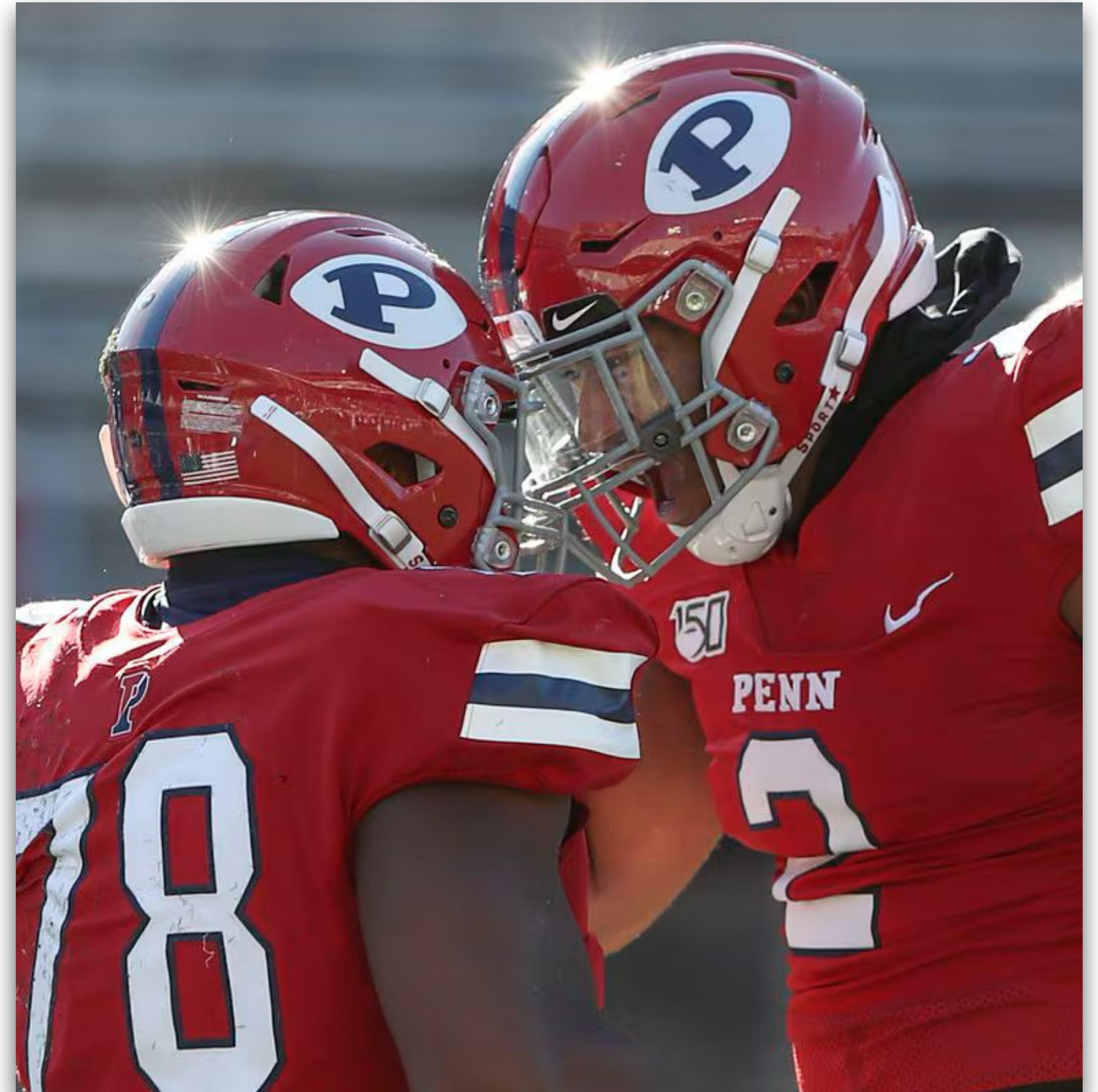




# The data

## Motivation

- Demographics
- Anthropometrics
- Testing
  - Strength
  - Power
  - Agility
  - Speed



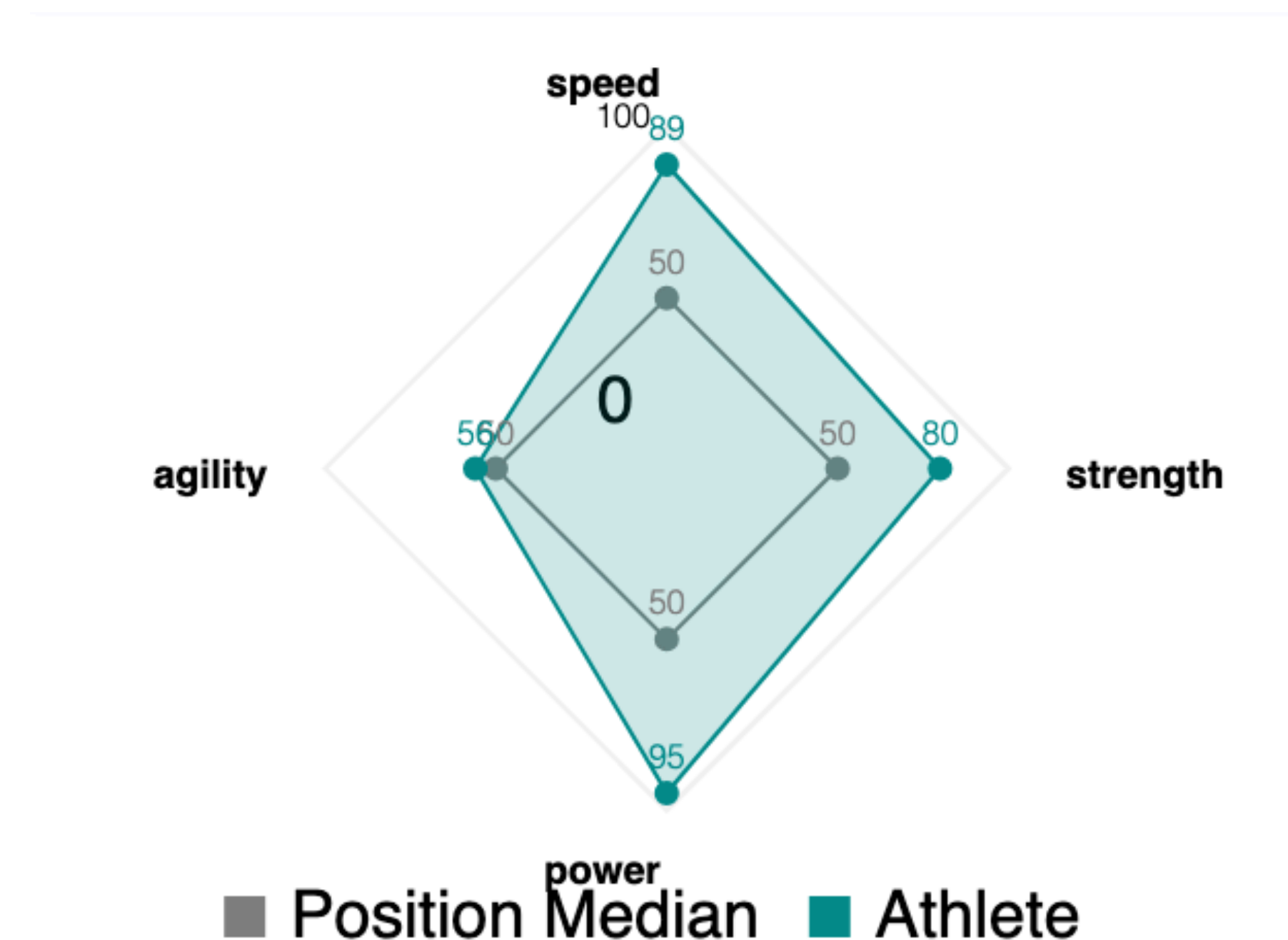
# 02 Ideation

Let's not oversell the numbers

# What should a testing dashboard prioritize?

Ideation

- Benchmark athletes within position
- Compare athletes 1:1
- Facilitate better data collection, storage, and decision-making
- Avoid prescribing solutions and overselling the data



An athlete's profile relative to his position group median



# 03 Data Processing 🧹

Balancing reproducibility and practicality



# 3 layers of data preparation

Data processing

1. Understanding football data conventions
2. Detecting errors
3. Unexpected NAs\*

```
29 # 4 digit height notation
30 parse_height_notation <- function(height_str) {
31   feet <- as.numeric(substr(height_str, 1, 1))
32   inches_full <- as.numeric(substr(height_str, 2, 3))
33   inches_frac <- as.numeric(substr(height_str, 4, 4))/10
34   height_inches <- (feet * 12) + inches_full + inches_frac
35   return(height_inches)
36 }
```

Converting 6015 to 73.5 inches

# \* Fancy quotes ⚠

Data processing

- 20% of heights were missing after reformatting
- Word, Powerpoint, Google Sheets like “fancy quotes”
- Code editors like "straight quotes"
- Fancy quotes create hard-to-find NAs for length-based stats
- Detecting and removing “” was a regex nightmare
- Eliminated missing height values

‘	straight single quote
”	straight double quote
‘	opening single quote
’	closing single quote
“	opening double quote
”	closing double quote

# Verifying athlete records

Data processing

- Roster data as ground truth
- Investigating failed joins
  - Name typos
  - Transfer students
  - Quitting
  - Pandemic super-seniors
  - Changed positions / numbers
- Unique IDs: name + position

2025 Football Roster					
<a href="#">Go To Coaching Staff</a>					
<a href="#">Print</a>				Roster V	
#	NAME	POS.	HT.	WT.	ACADEMIC YEAR
1	<a href="#">Vhito DeCapria</a>	Captain	-		R-Fr.
2	<a href="#">Kadari Machen</a>	LB	6-0	205	Sr.
2	<a href="#">Liam O'Brien</a>	QB	6-0	195	Sr.
3	<a href="#">Davis Ellis</a>	WR	6-0	190	Sr.
4	<a href="#">Devin Malloy</a>	DB	5-11	180	Jr.
4	<a href="#">Sage Webb</a>	WR	6-0	195	Sr.
5	<a href="#">Sean Williams, Jr.</a>	RB	5-10	200	Jr.
6	<a href="#">Karson Siqueiros-Lasky</a>	QB	6-1	190	So.
6	<a href="#">Alec Wills</a>	DB	5-11	175	Sr.
7	<a href="#">Bisi Owens</a>	WR	6-4	215	Sr.
8	<a href="#">Cadin Olsen</a>	QB	6-5	230	Jr.
10	<a href="#">Paul Jennings</a>	DL	6-4	255	Sr.
11	<a href="#">Alex Haight</a>	WR	6-2	190	Sr.
12	<a href="#">Jayden Drayton</a>	WR	5-10	175	Jr.

Penn Football 2025 roster



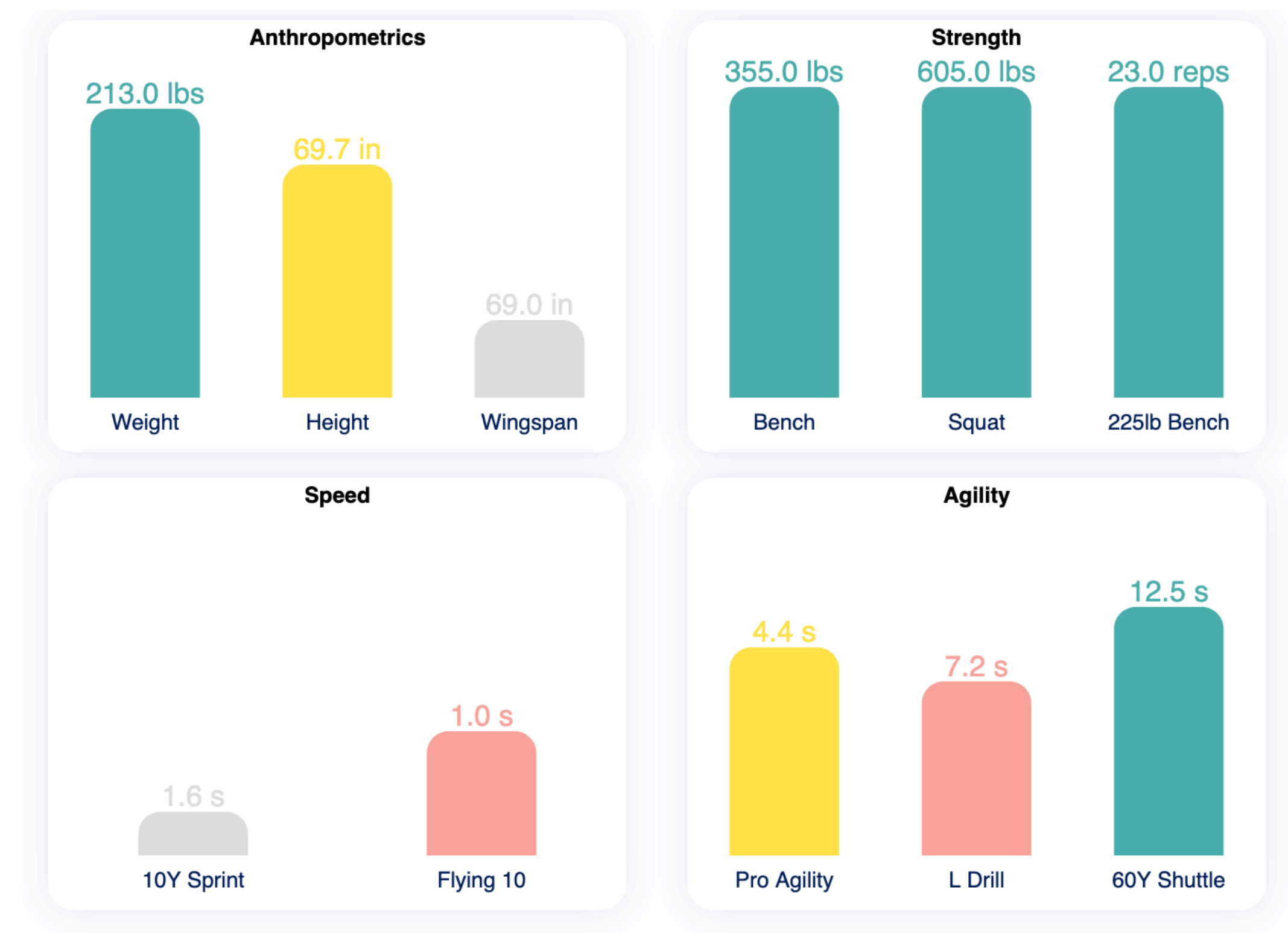
# 04 Visualization & Design

Functionality > minimalism

# Visualizing relative strengths/weaknesses

## Visualization & Design

- Percentiles vs absolute
- Color palette
  - Universal colors
  - Is red too alarming?
- Sometimes, more is more

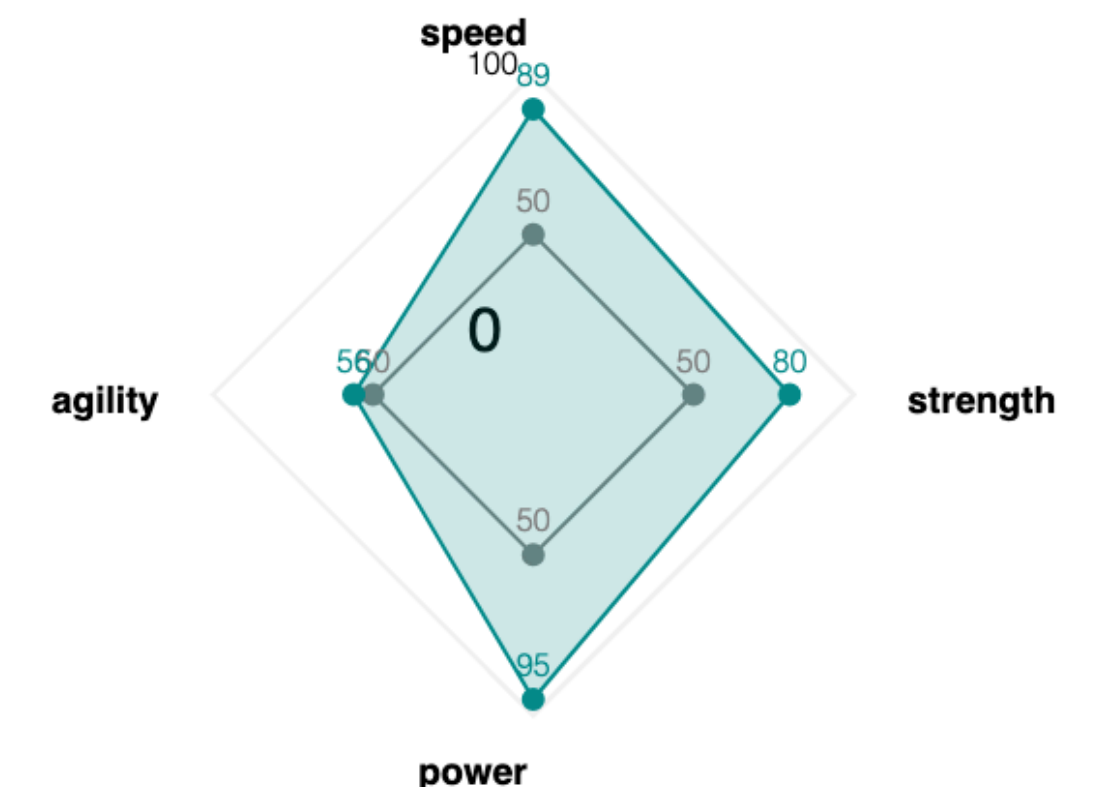
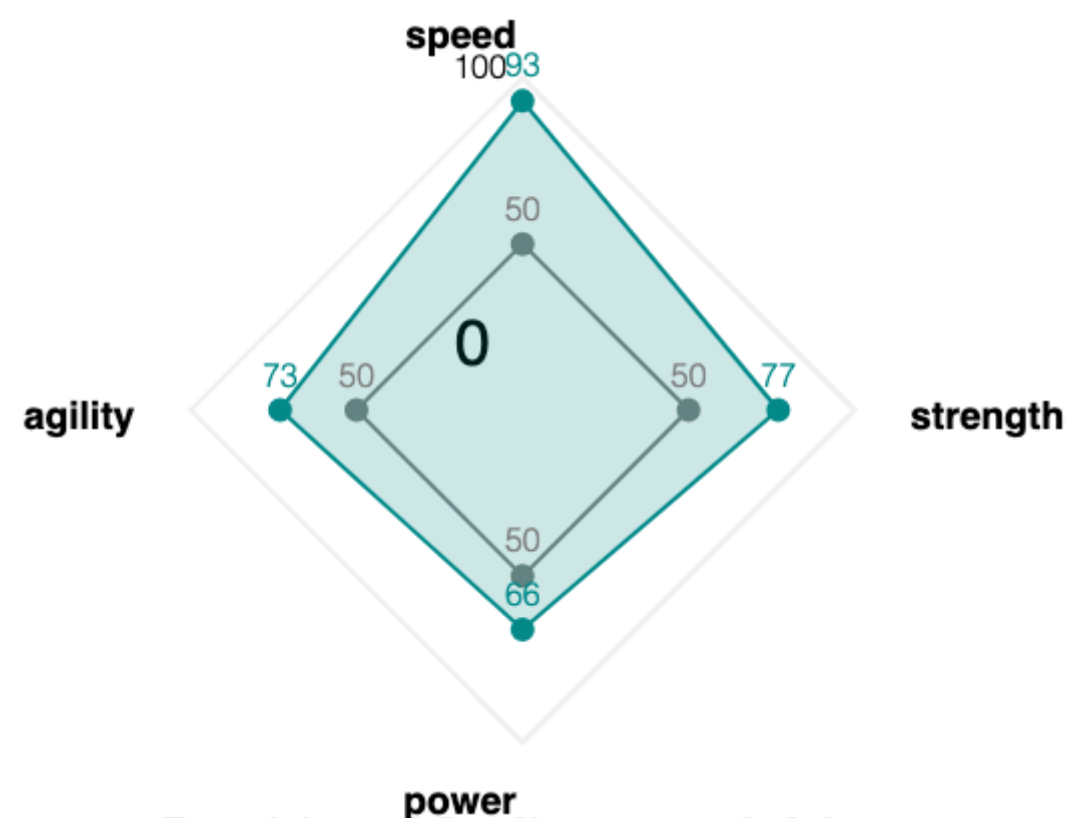
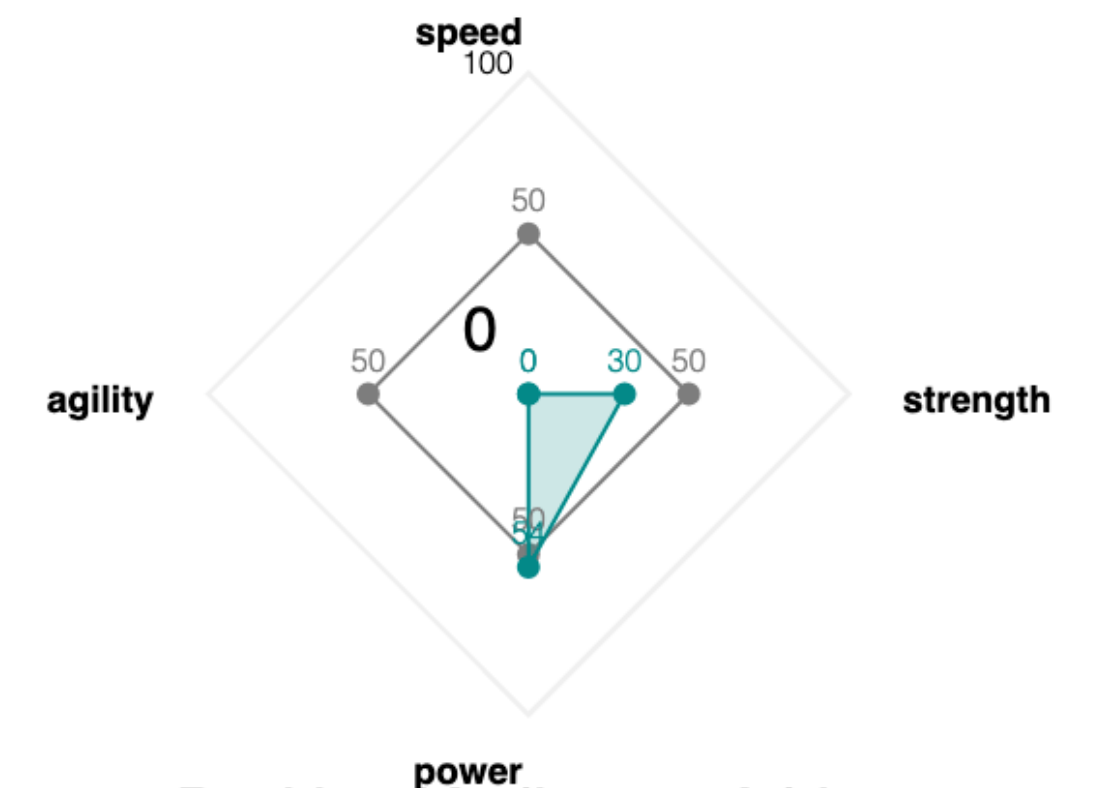
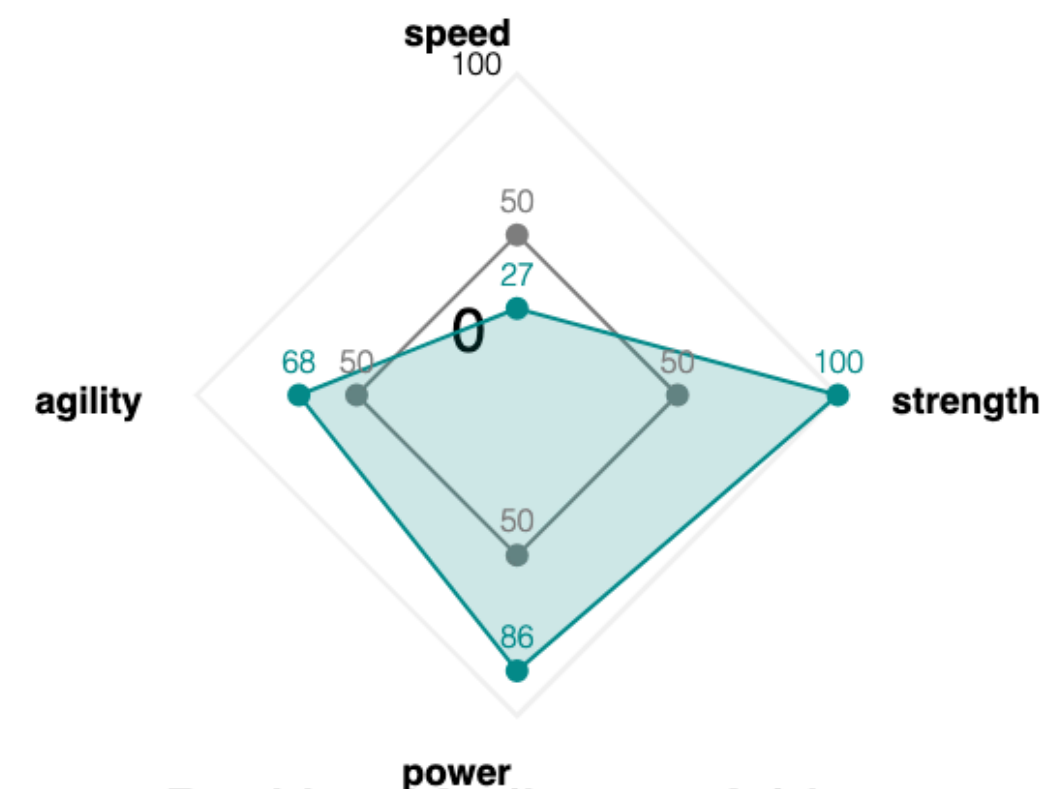


Dashboard bar charts

# Quick visual diagnostics

## Visualization & Design

- Radar chart
  - above or below median?
  - visual symbol of athlete typologies



4 athlete profiles



**05 Demo time**



# Penn Football Testing Dashboard

Add Data

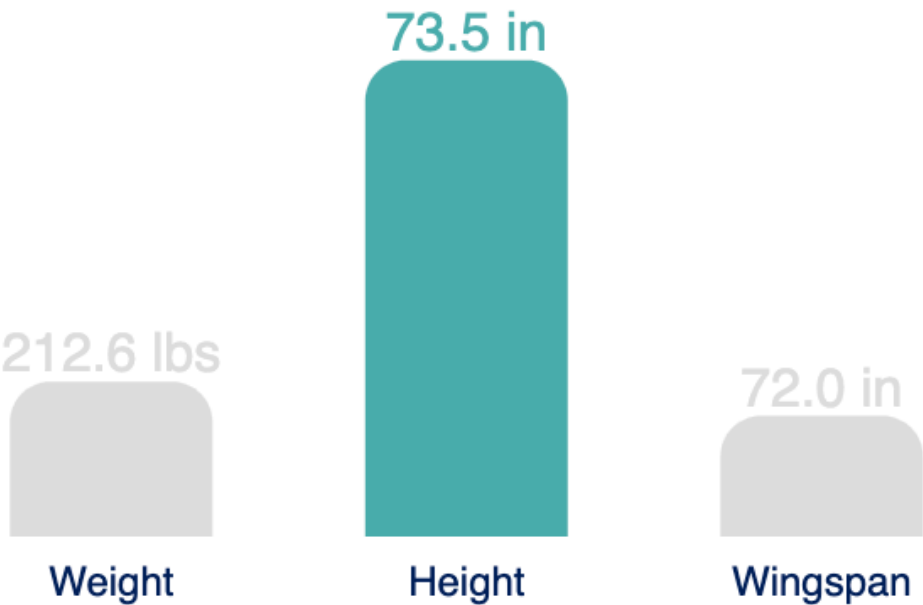
Athlete Report

Compare Athletes

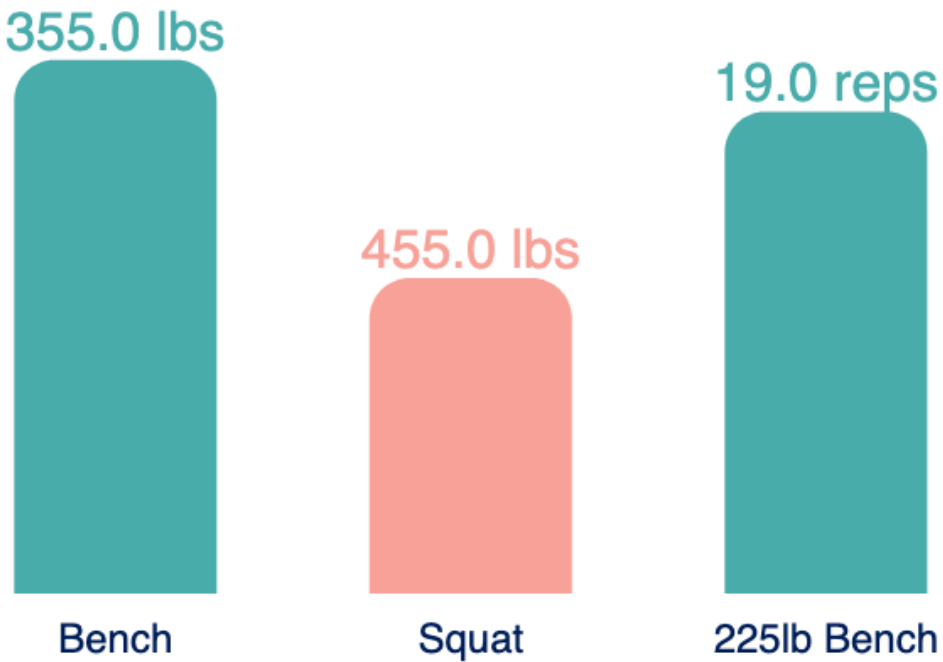
About this tool

Aidan Nash LB SE 38 RSI: 1.9 [Pick Athlete](#)

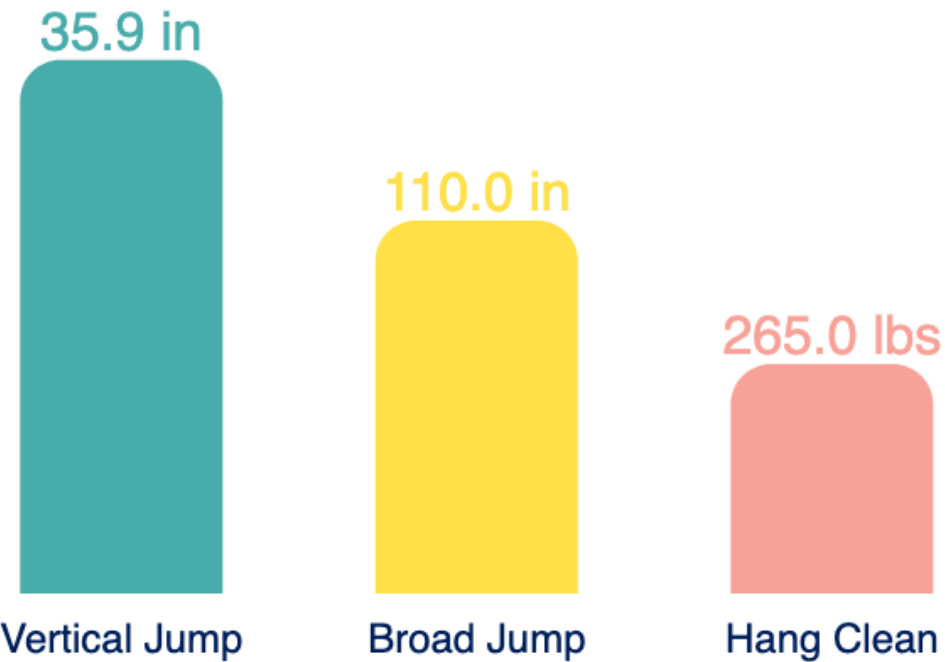
## Anthropometrics



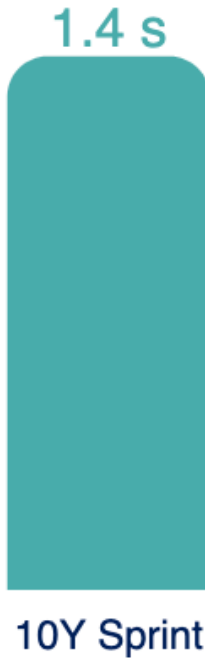
## Strength



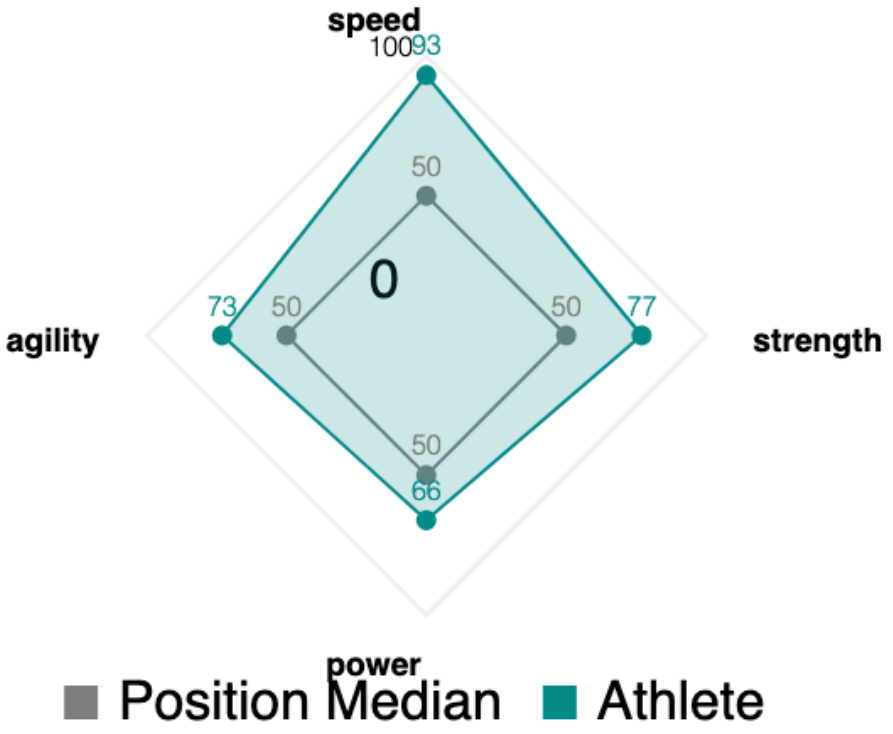
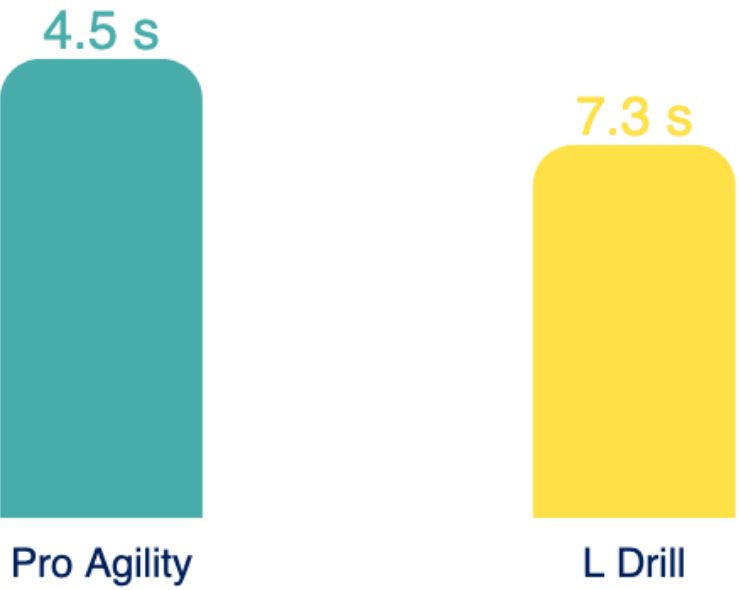
## Power



## Speed



## Agility



Charts visualize athletes' percentiles for each stat within their position groups. For example, 80th percentile indicates that an athlete scored better in a stat than 80% of his position group. The number above each bar is the absolute value of each stat. Hover over each bar to see the position group median and the athlete's percentile.

 **With gratitude to**

**Professors Xiaojiang Li and Mjumbe Poe**, for your technical and conceptual guidance, and your openness to an unconventional capstone project.

**Coach Miguel Rivera**, for trusting me with the data, and your continued support collaboration and feedback.

**MUSA '25**, for providing comments, ideas, and inspiration over two semesters' worth of iterations.







# Next steps

## Questions

- How to create safeguards around user interactions with the data?
- How can this tool track and visualize athletes' progress over time?
- Is it worth automating data processing/upload?

## Improvements/features

- PDF download
- Sorted/filtered athlete search
- Refining web design