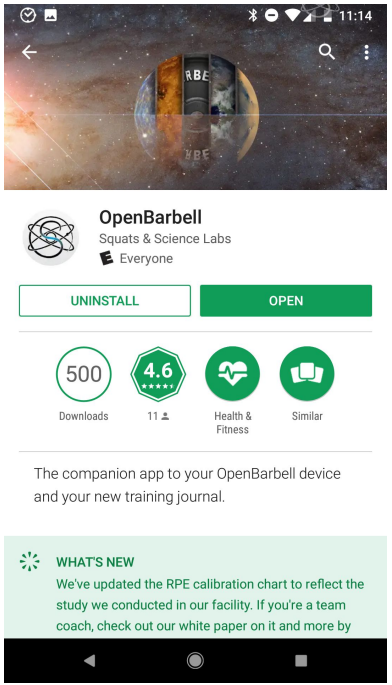
OpenBarbell Project

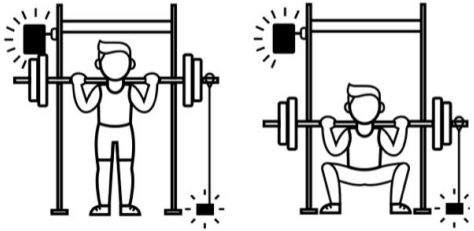
W207 - Final Project
Summer 2018

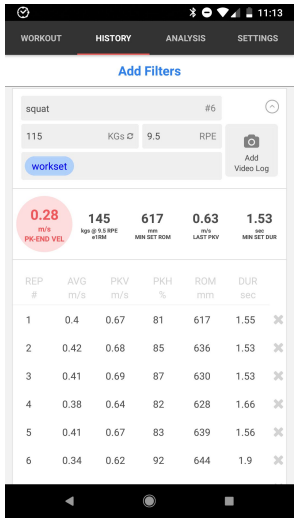
Yulia Zamriy
Renzee Reyes
Tim Witthoefft
Jack Workman

Berkeley
UNIVERSITY OF CALIFORNIA

What is OpenBarbell?



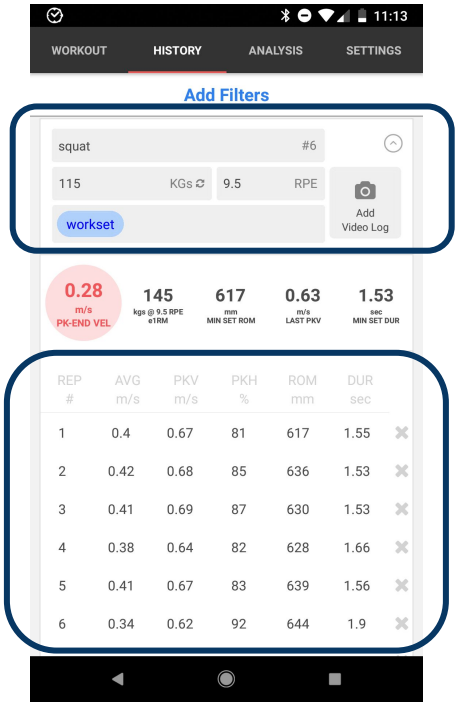




Berkeley
UNIVERSITY OF CALIFORNIA

What does it track?

Set-level user inputs



Rep-level device captured metrics

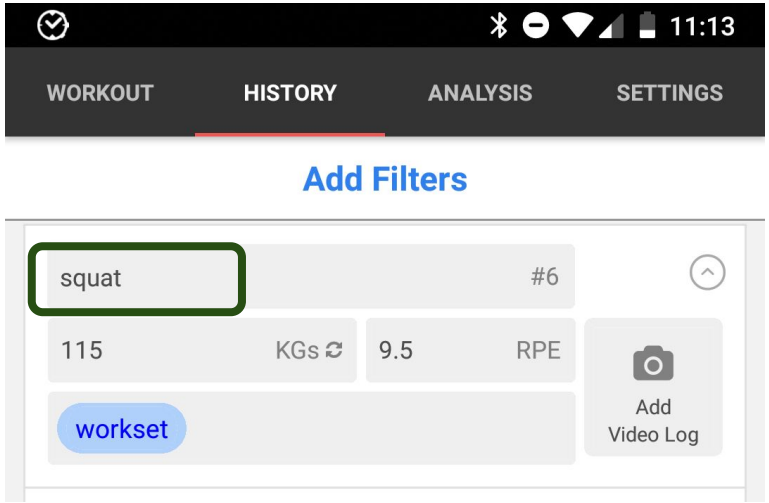
- Rep number
- Average velocity
- Peak velocity
- Peak velocity height
- Range of motion
- Duration
- Other…

Berkeley
UNIVERSITY OF CALIFORNIA

What do we want to achieve?

Part 1.
Exercise Classification

Goal:
Automatically classify the exercise type based on the measurements taken by the OpenBarbell device



What do we want to achieve?

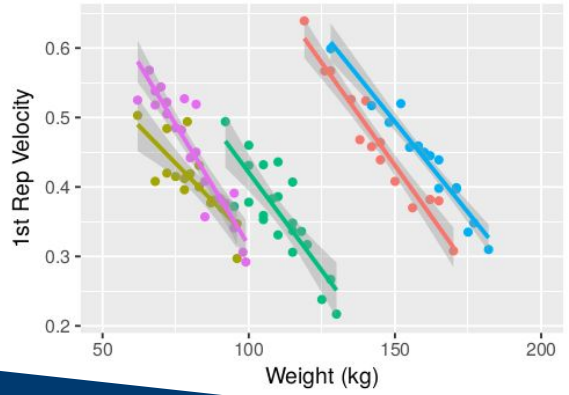
Part 2.
Lifter Clustering

Goal:

Group lifters based on their lifting parameters (velocity, range of motion, weight) to help them calibrate their training

Table 4a. OpenBarbell Squat Peak-End Velocity vs. RPE

Reps per Set	RPE 6.0	RPE 6.5	RPE 7.0	RPE 7.5	RPE 8.0	RPE 8.5	RPE 9.0	RPE 9.5	RPE 10
1	0.46	0.44	0.41	0.38	0.35	0.32	0.29	0.26	0.23
2	0.47	0.44	0.41	0.38	0.35	0.32	0.29	0.26	0.23
3	0.48	0.45	0.42	0.39	0.36	0.33	0.30	0.27	0.24
4	0.49	0.46	0.43	0.40	0.37	0.34	0.31	0.28	0.25
5	0.49	0.46	0.43	0.40	0.37	0.34	0.32	0.29	0.26
6	0.50	0.47	0.44	0.41	0.38	0.35	0.32	0.29	0.26
7	0.51	0.48	0.45	0.42	0.39	0.36	0.33	0.30	0.27
8	0.51	0.49	0.46	0.43	0.40	0.37	0.34	0.31	0.28
9	0.52	0.49	0.46	0.43	0.40	0.37	0.34	0.31	0.28
10	0.53	0.50	0.47	0.44	0.41	0.38	0.35	0.32	0.29



Source Data

- Over 100,000 exercises recorded
- Over 200,000 repetitions performed
- Extracted and shared as a JSON

Data Dictionary

- Schema openly shared by OpenBarbell

Metadata	User Input	Exercise in Motion
Set ID	Exercise Label	Average Velocity
Rep Number	Weight	Range of Motion
App Version	Rate of Perceived Exertion	Peak Velocity
Timestamp		Peak Velocity Location
		Rep Duration

Data Pipeline

Data Preprocessing	Model Classification	Post-Model Tasks
Unpack Inner Data Structures	Partition Training, Test, and Dev Data	Rep Consensus Among Exercise Sets
Handle Unclean & Invalid Data	Model Training & Tuning	Ensembles of Multiple Models
Select & Transform Key Features	Produce Visualizations	Translating Results for End Users

Part 1. Exercise Classification

Lifting 101. Know the Big 3!

Bench press

Squat

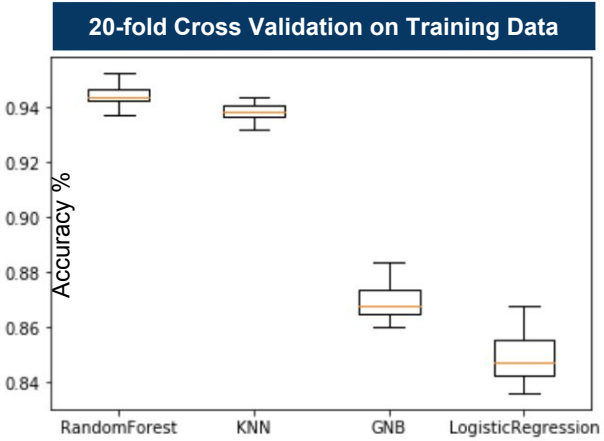
Sumo Deadlift

Approach: Exercise Classification

- Split into train (80%) development (10%) and test (10%)
- Standardize data (StandardScaler)
- Classification models:
  - Gaussian Naive Bayes
  - Logistic Regression
  - KNN
  - Random Forest
- Evaluation:
  - Confusion Matrix
  - Classification Report
  - 20-fold cross validation

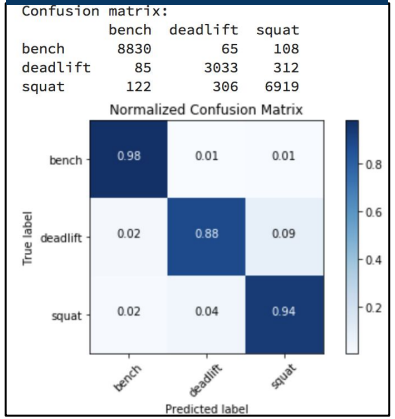
Findings

- Random Forest and KNN performed the best at ~94% accuracy
  - 20-fold cross validation shows low variation on training data splits
  - Both models classified the bench press the most consistently, at 98%
  - Similarly, both models had trouble with the deadlift
- A set ensemble - the collection of multiple reps together - boosts accuracy further
  - Simple voting between the reps of a set increased accuracy to 96+%



Random Forest

KNN



Berkeley
UNIVERSITY OF CALIFORNIA

Learnings

- **This could work!** The team feels confident that a single model or ensemble model could be used in the OpenBarbell app

- **Range of Motion** (ROM) is the most important feature. Future hardware updates to OpenBarbell plan to include 3D location tracking → better classification
  - User characteristics such as **height, body shape, and gender** should be added to the model (not available to our team) to further improve the ROM measurement

Feature importance
(Random Forest)

	features	importance
1	ROM	0.397768
2	PeakVel	0.172415
5	weight_lbs	0.152851
3	PeakVelLoc	0.152157
4	RepDur	0.075787
0	AvgVel	0.049022

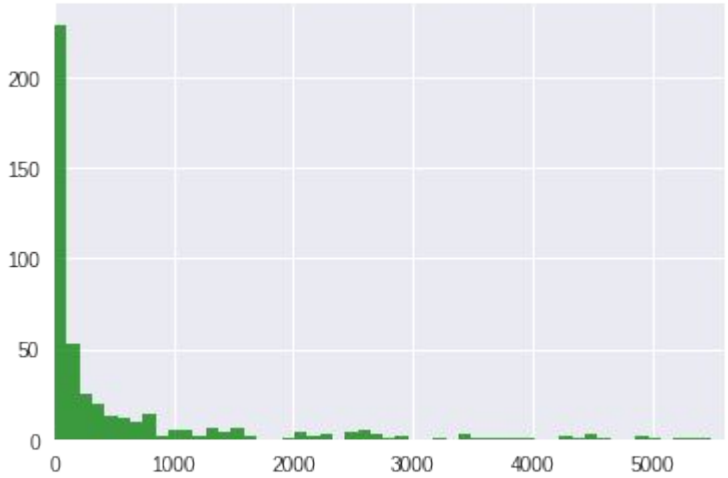
Part 2. Lifter Clustering

Approach: Steps

1. Feature analysis:
    - identify natural clusters
2. Hierarchical clustering:
    - How many clusters?
3. Cluster Visualization:
    - Are their distinct patterns between clusters?
4. Revise Hierarchical Clustering:
    - Can we make clusters more distinct?
5. Practical Application of Clustering:
    - User membership prediction

Approach: Setup

User ID Rep Count Histogram

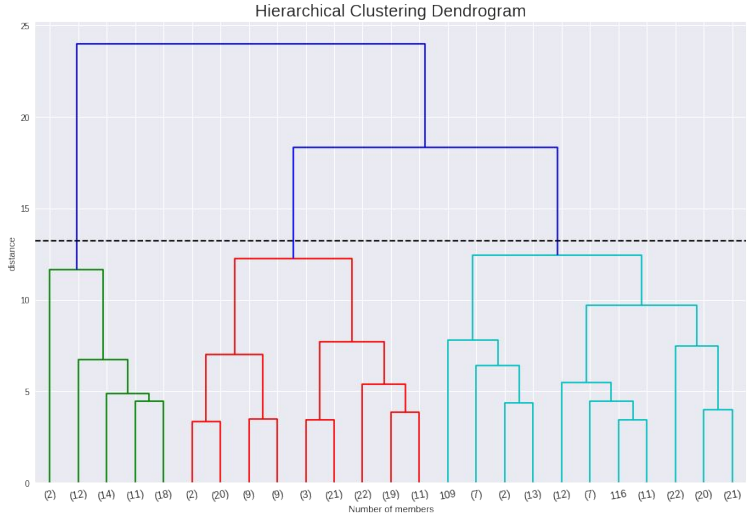


Exercise Distribution

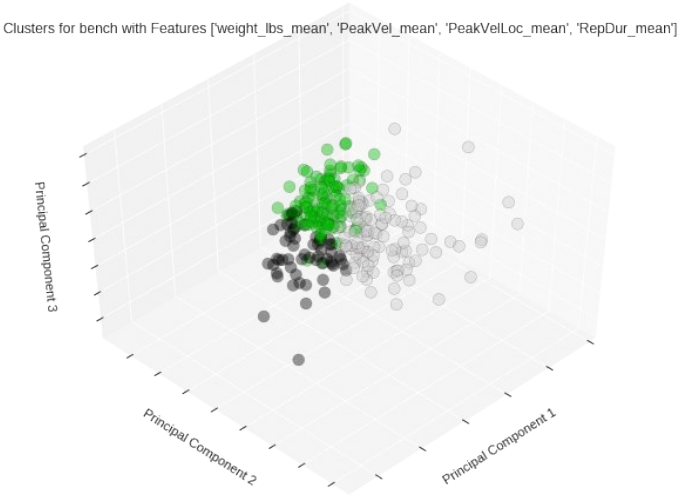
Exercise	Number of users
Bench Press	323
Squat	311
Deadlift	247

Hierarchical Clustering

Dendrogram for Bench Press



PCA on Bench Press Clusters

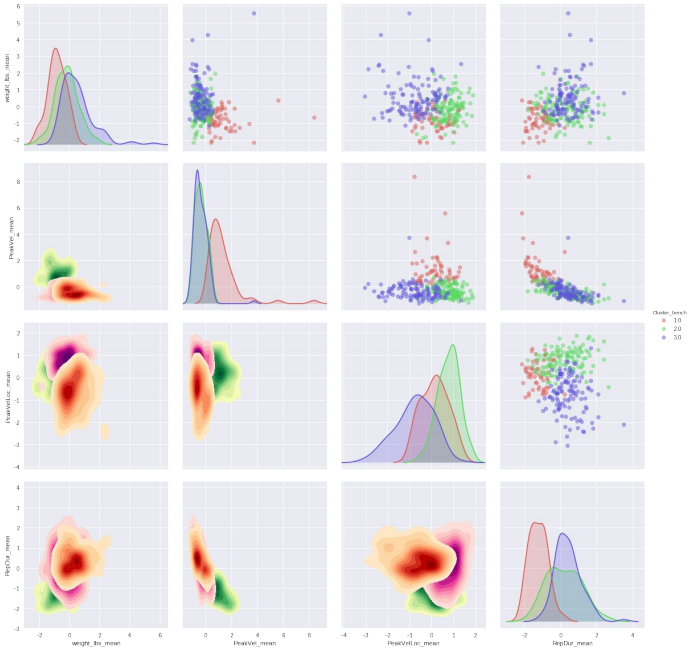


Berkeley
UNIVERSITY OF CALIFORNIA

Cluster Analysis

Bench Press Cluster Number	Number of users
1	57
2	116
3	117

Bench Press Clusters



Cluster Fine-Tuning

Created tool with Google Colab that enabled rapid iteration on model testing

Select Exercise

exercise: squat

Select Feature Columns

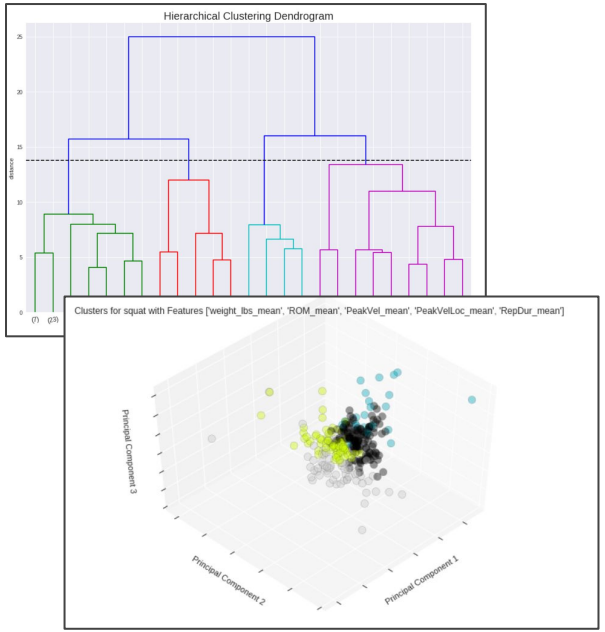
- rpe_num: ☐
- weight_lbs: ☑
- RepCount: ☐
- ROM: ☑
- AvgVel: ☐
- PeakVel: ☑

Select Feature Aggregators

- np_mean: ☑
- np_min: ☐
- np_max: ☐
- np_std: ☐

Hierarchical Clustering Setup

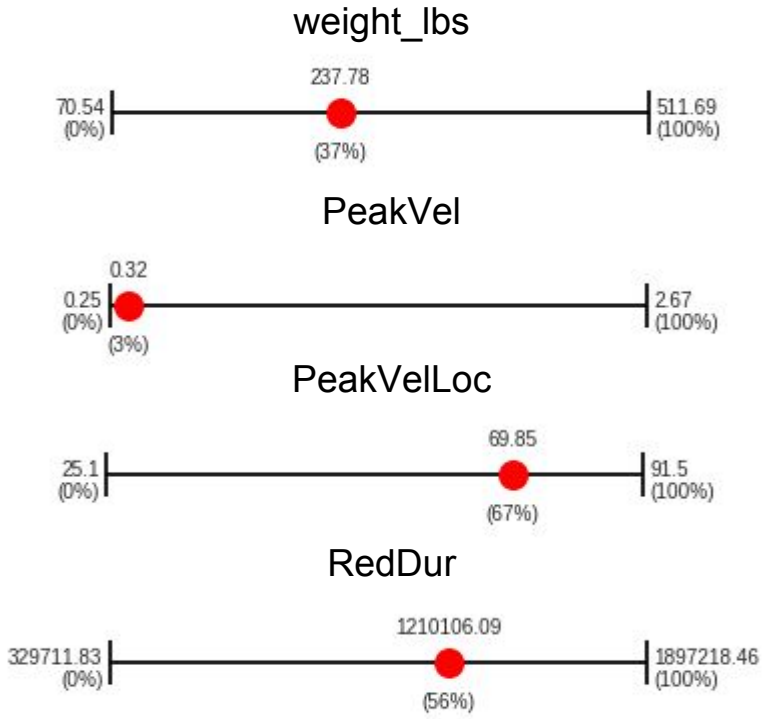
- cutoff: 0.55
- linkage_algo: ward
- distance_algo: euclidean



Berkeley
UNIVERSITY OF CALIFORNIA

User Prediction

- Example random user
- Used KNN to find user's cluster
- Plotted user's parameters relative to those in their cluster



Findings

- Top clustering features vary by exercise:
  - Weight, Rep Duration, Peak Velocity Location
- Number of clusters by exercise highly dependent on features:
  - Outliers skew results
- Requires a lot of iterations to come to an applicable solution

Learnings

- Lifter segmentation is not visually apparent when there is a couple hundred of lifters
- Instead of approaching it from exploratory standpoint, we need to do it with a clear intention:
  - What metrics are important for programming design? What can be actually actionable?

What's next for OpenBarbell?

Next Steps for ML @ OpenBarbell

- Use exercise predictions to increase the amount of data for lifter segmentation
- Combine segments across exercises

OpenBarbell No More!

- RepOne is OB on steroids, heavily reliant on data science
- Automated exercise classification will increase the amount of data usable for analytics
- Lifter segmentation will allow creation of custom templates for lifters

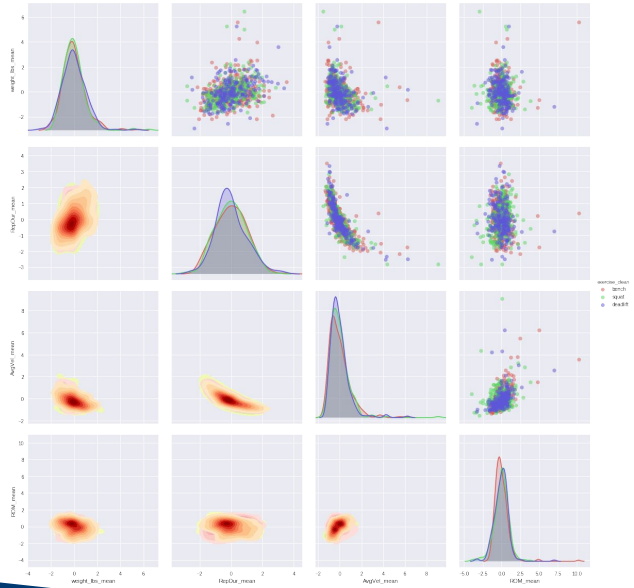
Appendix

Lifter Clustering: Motivation

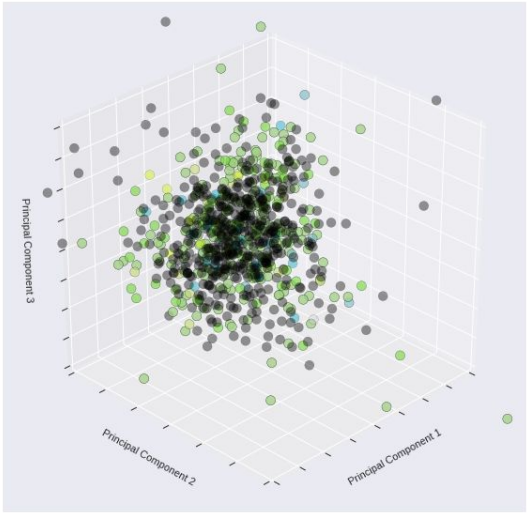
- Yulia:
  - Tall ⇒ *large* range of motion on squat and bench
  - Sumo puller ⇒ *smaller* range of motion on deadlift
  - I can grind on squat ⇒ *low* average velocity
  - But not on bench ⇒ *high* average velocity
  - I fail my deadlift off the floor ⇒ peak velocity height is *at the top*
  - I fail my bench off the chest and struggle with a lockout ⇒ peak velocity height is somewhere in the *middle*
- I am not unique, but I am not standard either

Feature Analysis

Pairwise Feature Plots



PCA on Features



Cluster Overlap

		Squat											
		Deadlift											
		1			2			3			4		
		1	2	3	1	2	3	1	2	3	1	2	3
Bench	1	0	0	3	0	0	0	3	0	1	9	0	3
	2	5	12	20	0	1	3	9	2	8	1	0	4
	3	5	17	16	0	1	5	5	0	3	3	0	1