Clustering Race Horse Movement Profiles to Discover Trends in Injured Horses

Sara Colando, Jonathan Pipping, Kris Wilson

Motivation

Churchill Downs to suspend all racing operations to further evaluate safety measures amid increase in horse deaths

By Homero De la Fuente and Zoe Sottile, CNN
Updated 6:37 PM EDT, Fri June 2, 2023

via CNN

Project Goals

1 Identify horses who under-raced between 2019 and 2021

2 Cluster movement profiles for horses who raced in New York in 2019

3 Discover whether certain movement clusters are more associated with injured horses

Our Data Sources

2019-2020 Severe Horse Injury Data

2019-2022 NYRA Start Lists

2019 NYRA Tracking Data

Under-Racing Data Sources

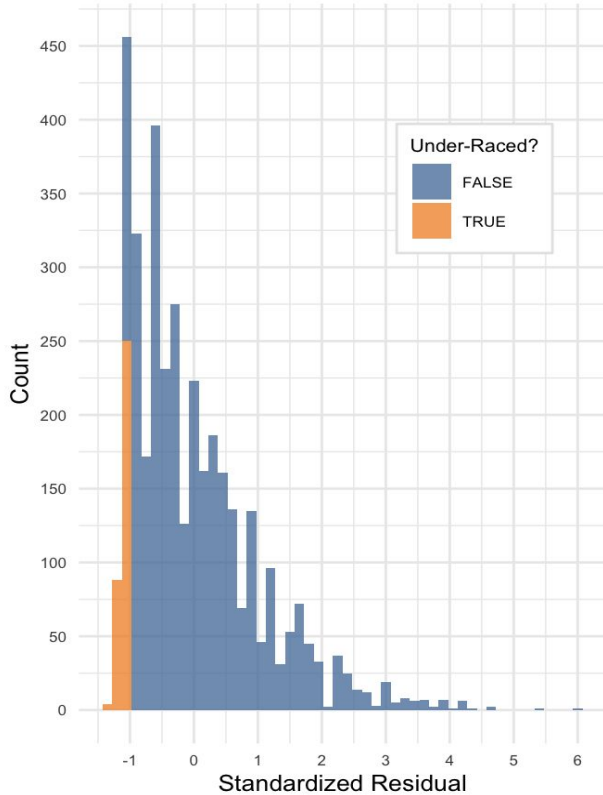| 2019-2020 Severe Horse Injury Data	**+**	2019-2022 NYRA Start Lists

2019 NYRA Tracking Data

3678 rows x 6 columns

horse_id	age_year	race_year	n_races	if_injury_reported	fatal
350	4	2019	4	TRUE	No
350	5	2019	8	TRUE	No
350	6	2020	3	TRUE	No
350	6	2021	1	TRUE	No
358	3	2019	2	TRUE	No

Under-Racing Horses

Evaluation Visualizations for Expected Race Count Model
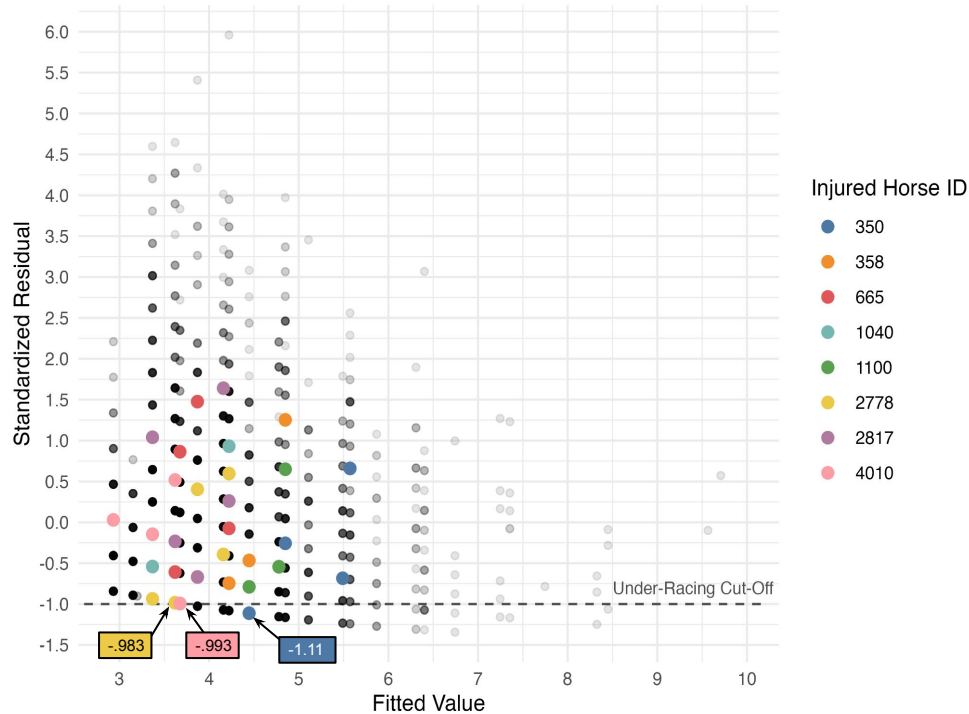
- **27%** (251/931) of horses under-raced between 2019 and 2021

- **29.9%** (75/251) of horses who under-raced did so more than once
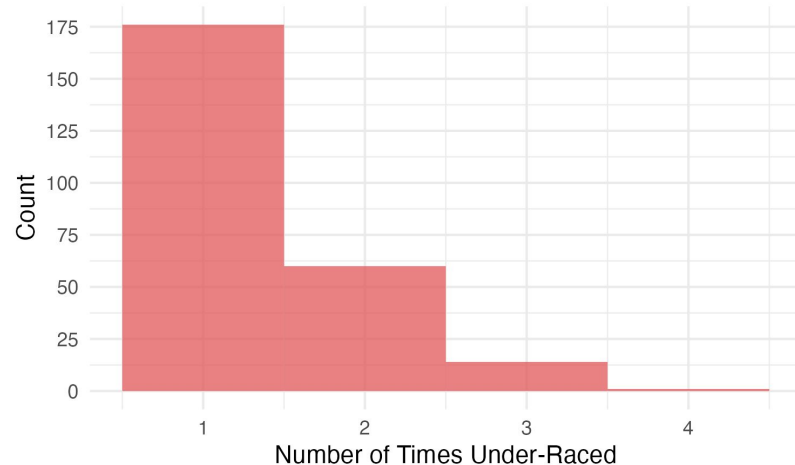
Under-Racing vs. Severely Injured Horses

Standardized Residuals versus Fitted Values for Race Count given the Horse's Age and the Calendar Year

Only 1 horse who was severely injured in 2019 under-raced and it was in 2021

So, what are the standardized residuals for the horses who were severely injured in 2019?

Severely injured horses were more likely to *race more than expected* in 2019 and *less than expected* in 2020 and 2021

Actual versus Expected Race Count for Horses who became Severely Injured in 2019

Clustering Data Source

2019-2020 Severe Horse Injury Data

2019-2022 NYRA Start Lists
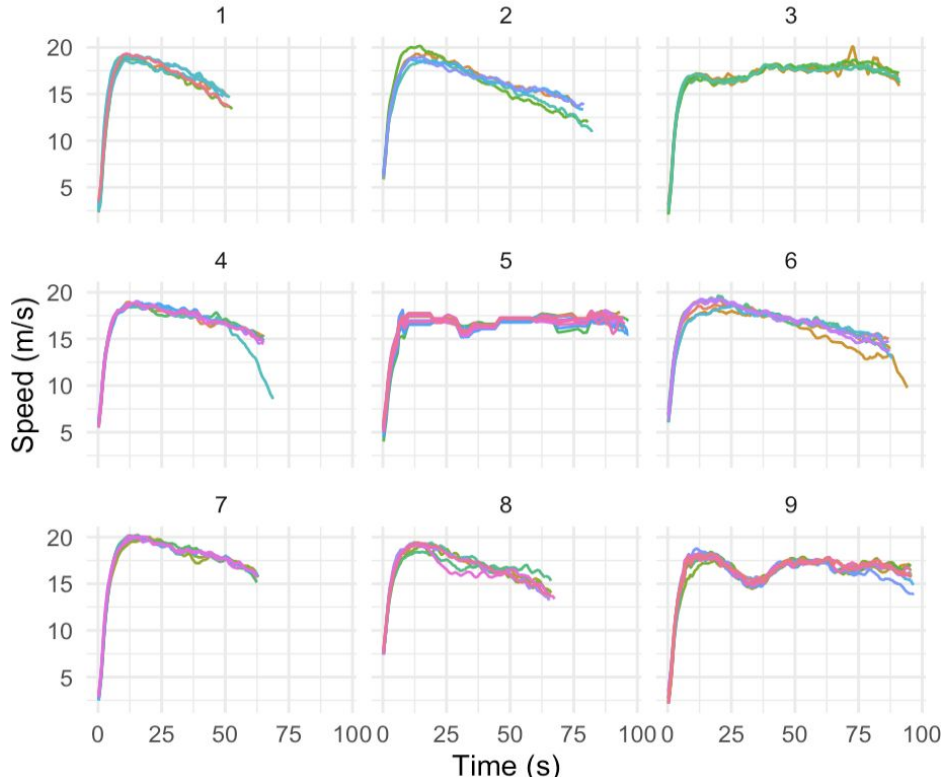
2019 NYRA Tracking Data

4,965,717 rows x 41 columns

Visualizing NYRA Tracking Data
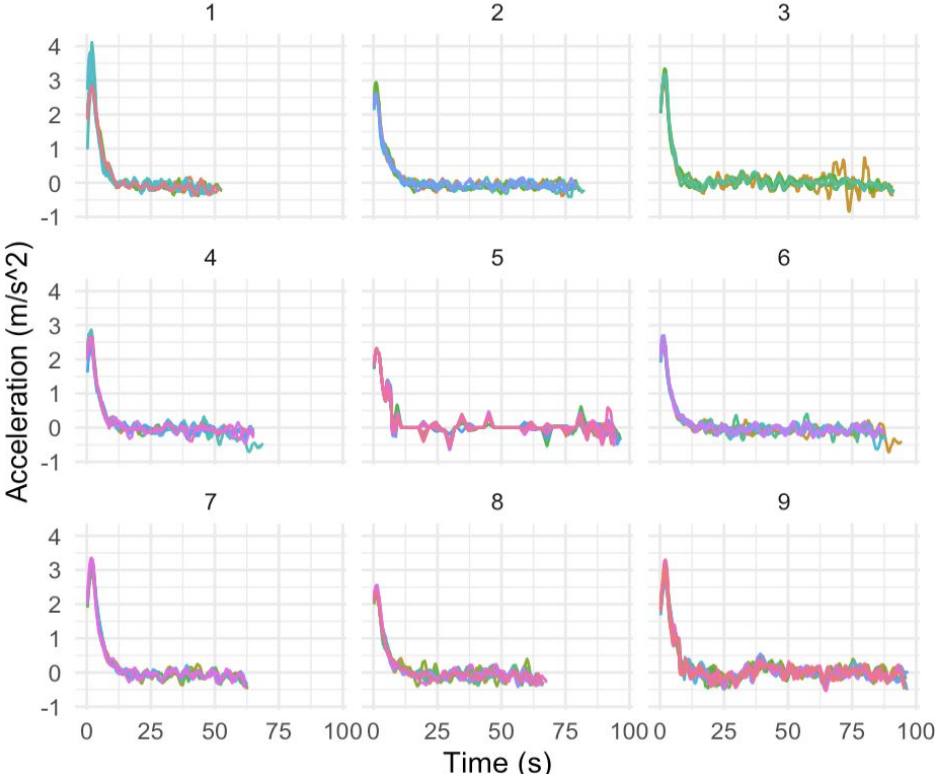

The Aqueduct on April 19th, 2019: Race 1
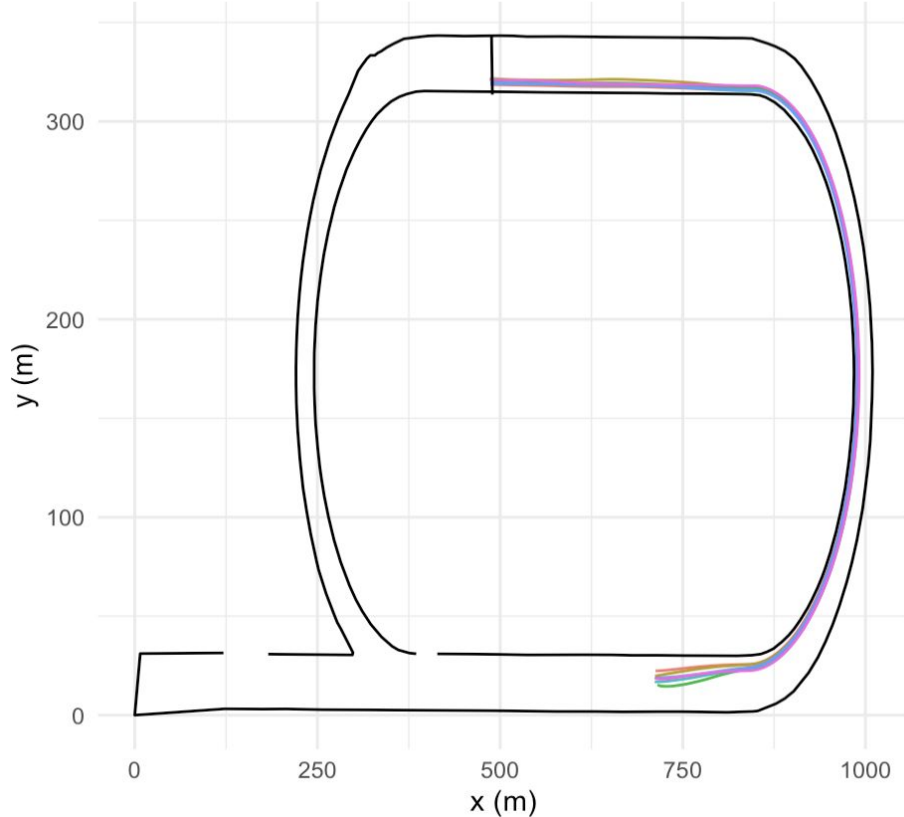
Visualizing NYRA Tracking Data

Speed vs Time for Each Horse
At the Aqueduct on April 19th, 2019

Visualizing NYRA Tracking Data

Acceleration vs Time for Each Horse
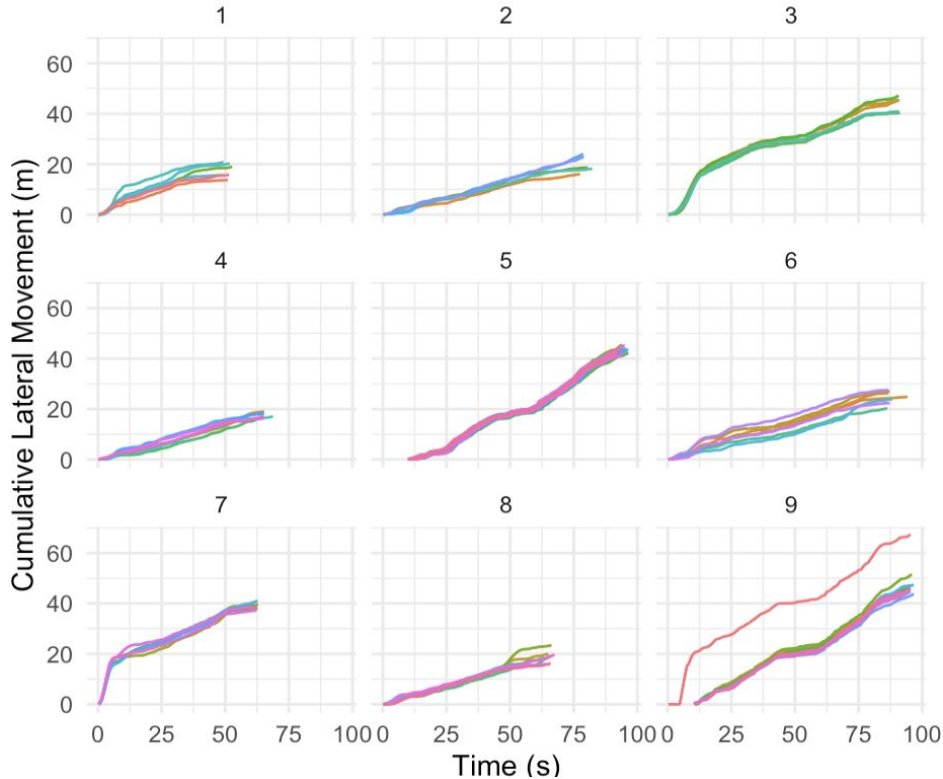At the Aqueduct on April 19th, 2019

Visualizing NYRA Tracking Data

Lateral Movement vs Time for Each Horse
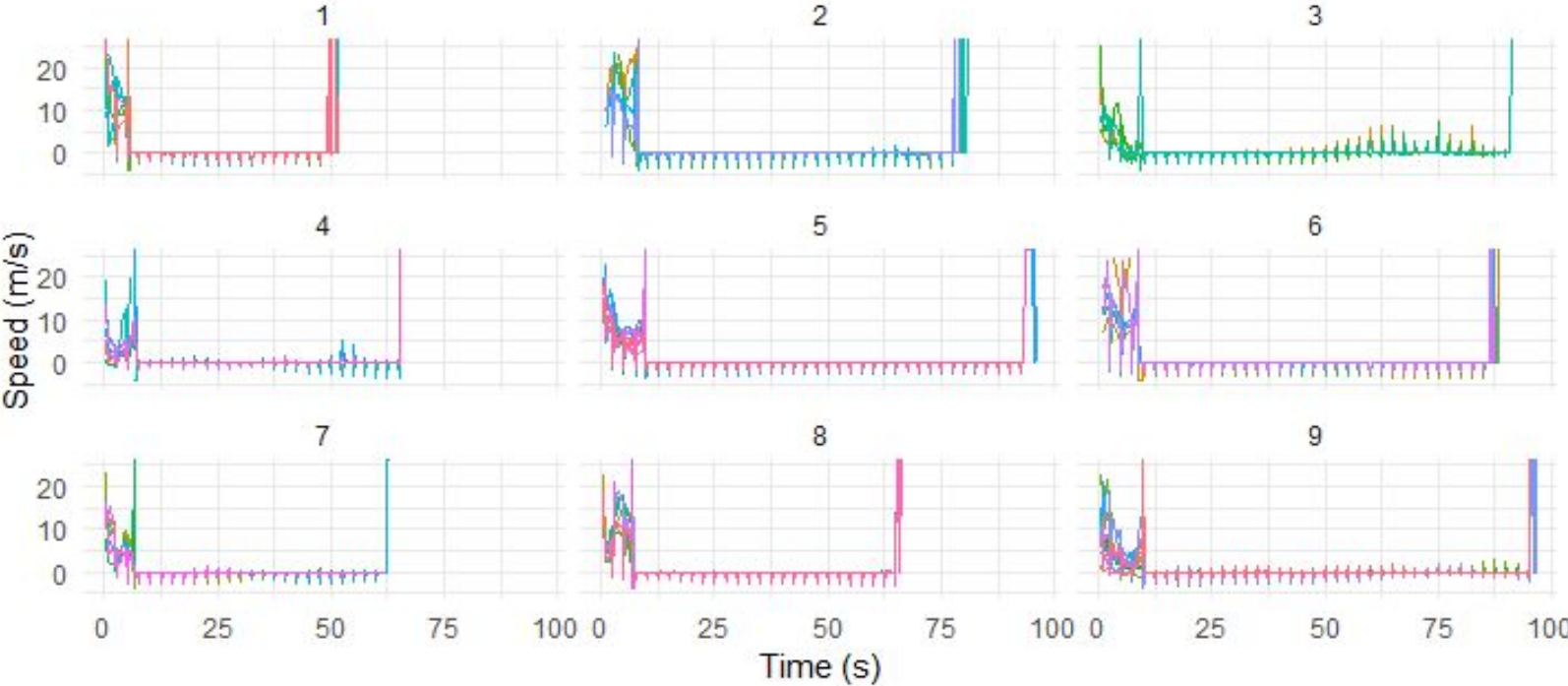At the Aqueduct on April 19th, 2019
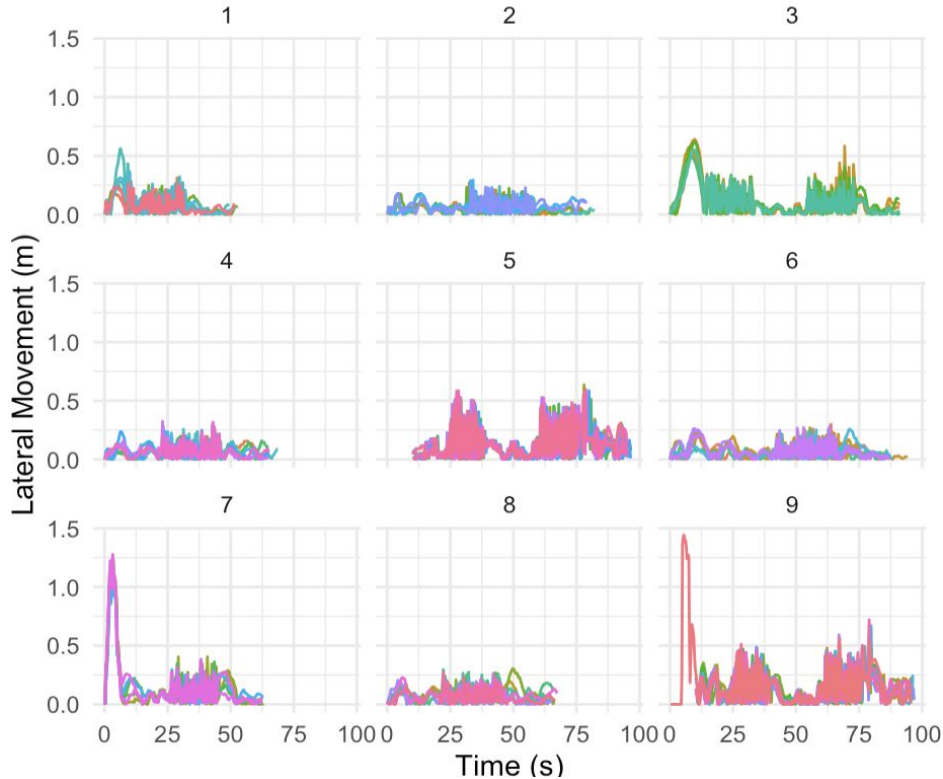
Visualizing NYRA Tracking Data

Cumulative Lateral Movement vs Time for Each Horse
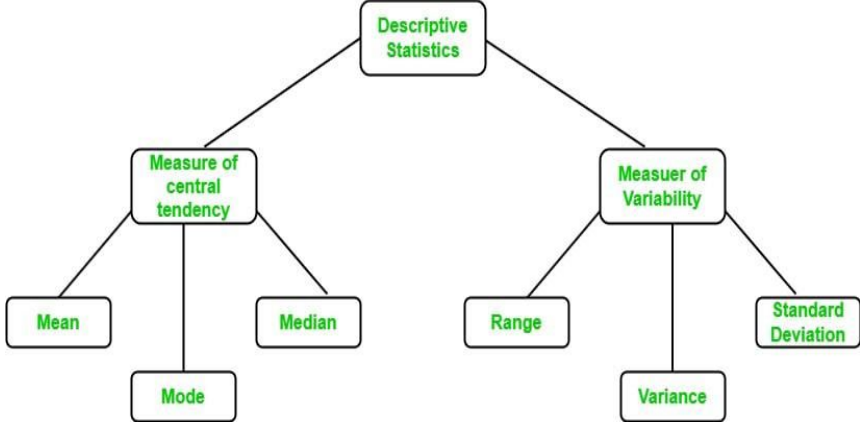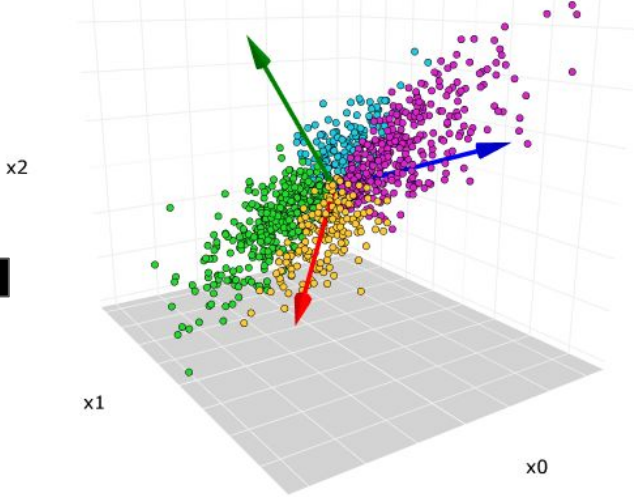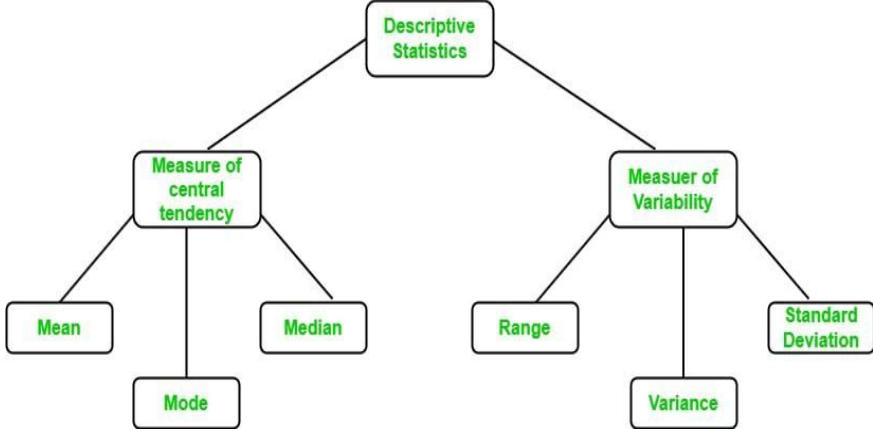At the Aqueduct on April 19th, 2019

Visualizing NYRA Tracking Data
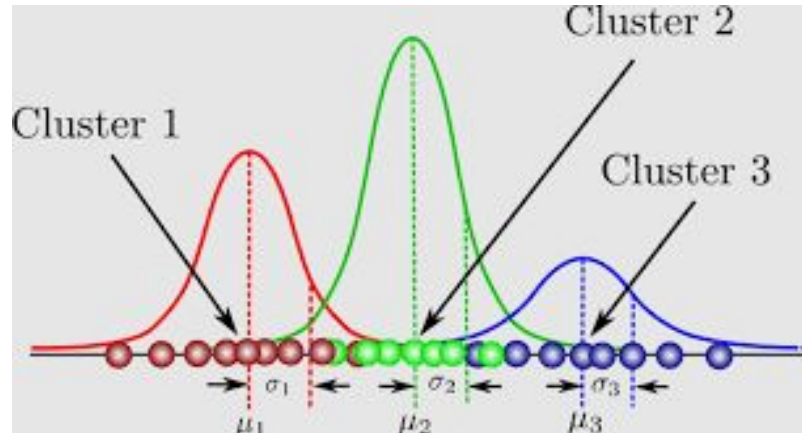

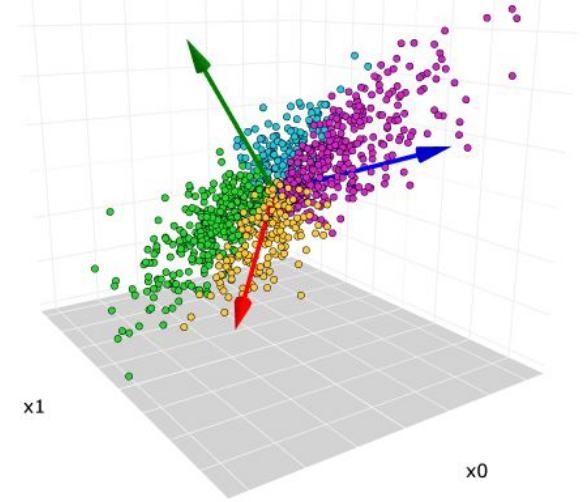Average Strain vs Time for Each Horse
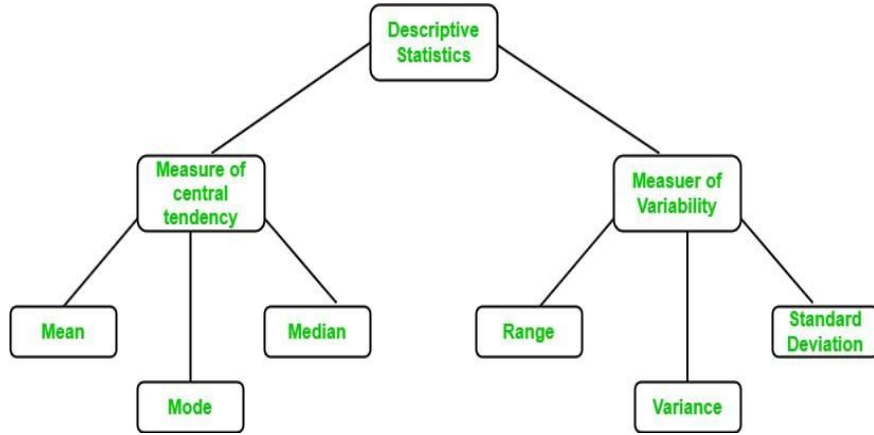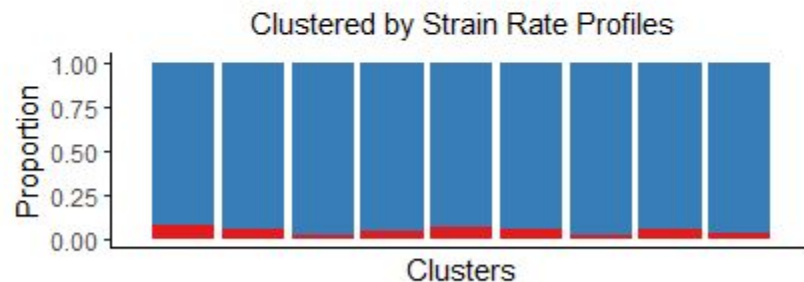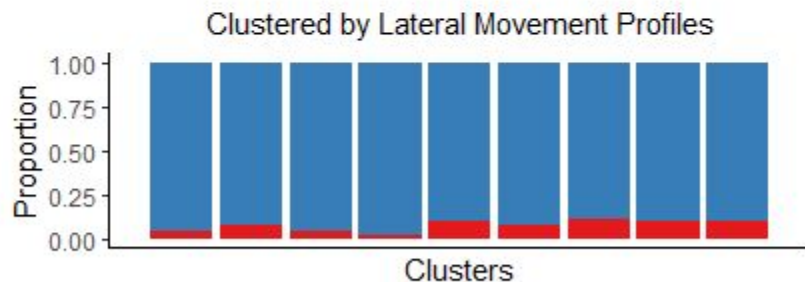At the Aqueduct on April 19th, 2019

Summarizing NYRA Tracking Data

Summarizing NYRA Tracking Data

Summarizing NYRA Tracking Data

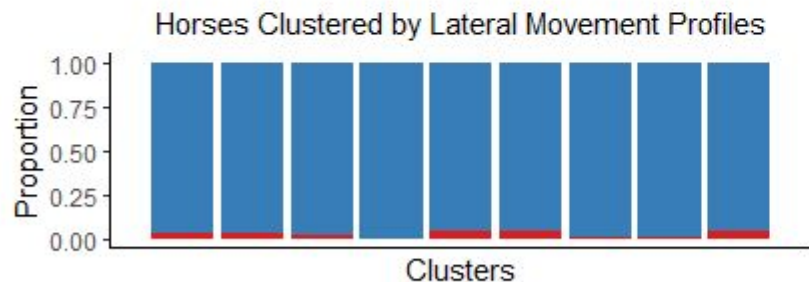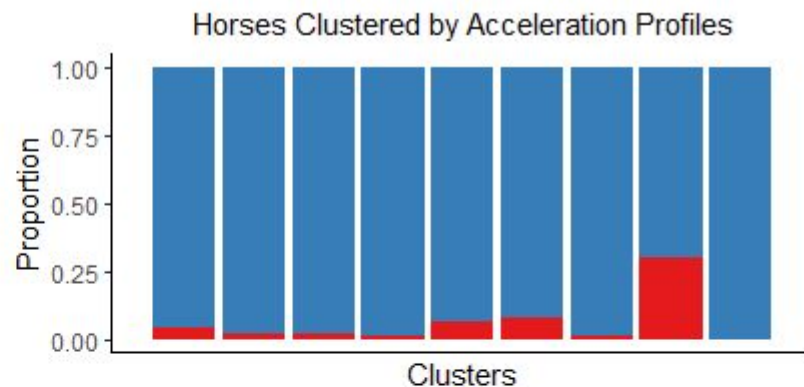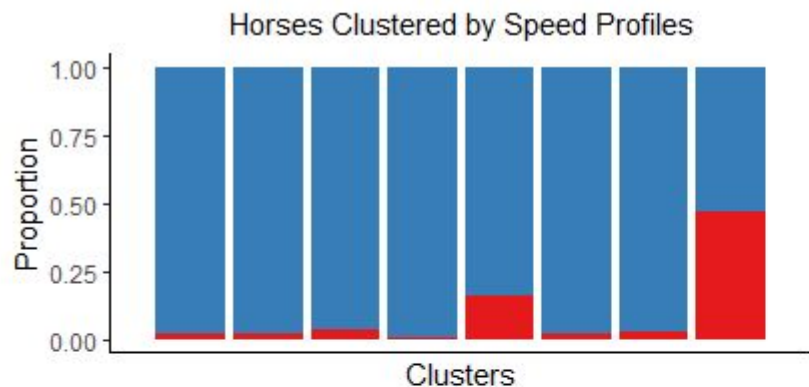# Putting it Together: Investigating Under-Racing

Putting it together: Investigating Injuries

Putting it Together: Investigating Injuries (cont.)

Cluster	Average Median Speed (m/s)	Average Minimum Speed (m/s)	Average Maximum Speed (m/s)	Average CV Speed
8	16.10	1.24	21.13	0.44
5	16.32	4.39	19.32	0.17
3	16.82	2.70	19.49	0.13
1	17.48	3.91	19.84	0.12
7	17.10	4.29	22.16	0.12
6	16.67	6.42	19.24	0.11
2	17.10	3.67	18.93	0.10
4	17.39	7.56	18.98	0.09

Main Findings

- Many horses raced less than expected, given their age and the calendar year
- Speed and acceleration clusters have substantial spikes in the ratio of injured to non-injured horses

Future Work

- Collecting start information for all the horses tracked in 2019 by NYRA
- Computing lateral movement for every race distance
- Multilevel modeling to account for individual horses