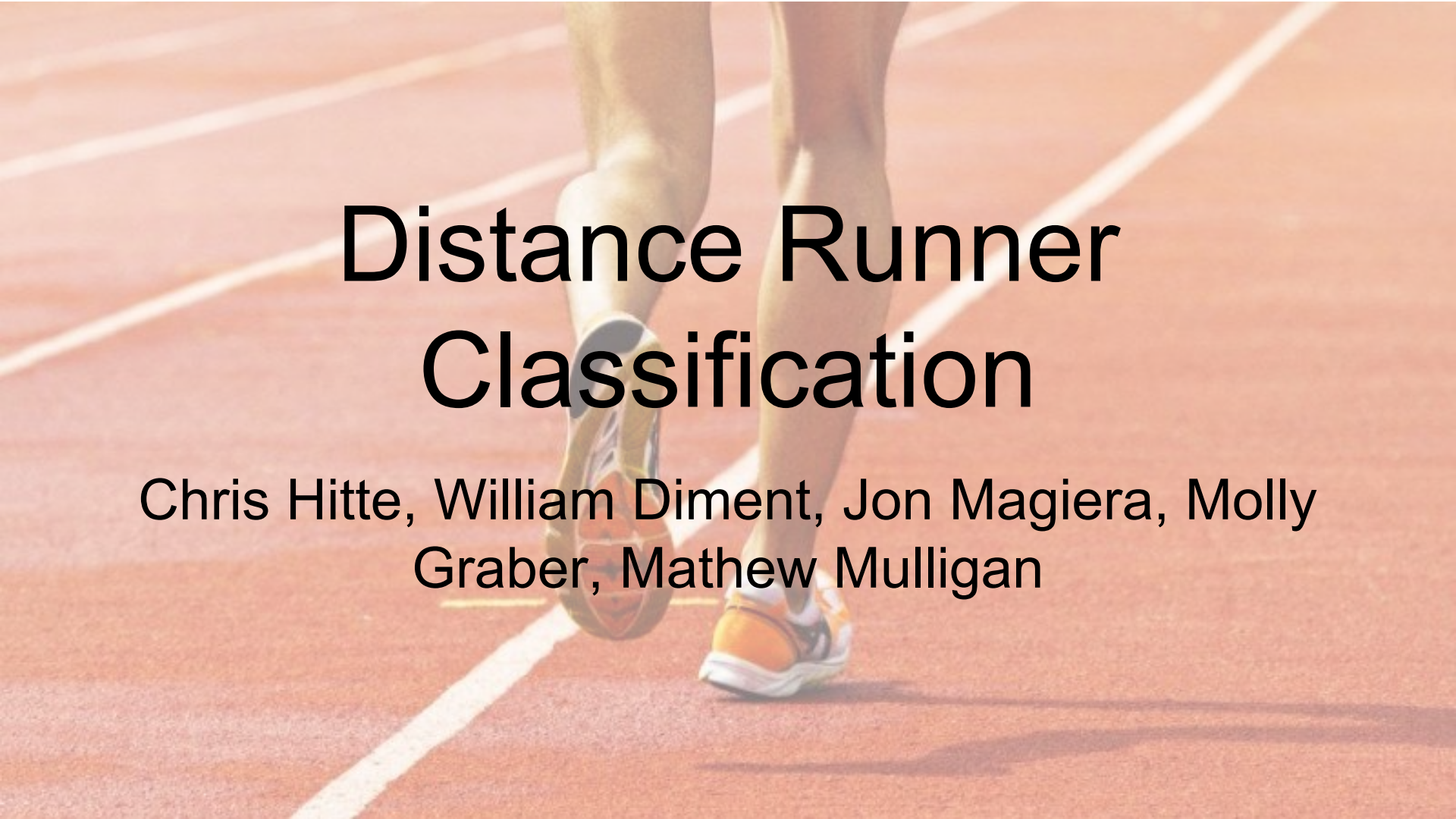

Distance Runner Classification

Chris Hitte, William Diment, Jon Magiera, Molly Graber, Mathew Mulligan

Background

- Distance runners tend to excel at either middle or longer distance events.
- This difference is due to different physiological requirements for each category.
  - Shorter distances require more anaerobic training
  - Longer distances requires aerobic training
- We intend to train a model that will estimate a given athlete's performance across a range of long and middle distance events based on their relationship to other athletes' races times.

Data

The data for our project will come from the Track and Field Results Reporting System (TFRRS) which holds the event times for thousands of athletes over the last several years.

While the data is not published as a concise dataset, we will use a scraper to generate one based on the publicly available data on the site.

This data consists of run times for a given athlete, event, distance, division, gender and date.

Method

Having scraped a sufficiently large dataset from the TFRRS, we will perform K-Means classification on the data.

Our goal is to generate clusters from runner performance, in order to categorize “types of runners.” We anticipate that these clusters will be indicative of a runner's relative strengths -- short, middle, or long distances.

Using these clusters, we will then be able to estimate a runners performance in events he or she has not directly competed in, based on the trends apparent from runners within their same cluster.