

Datafest 2019

Margaret Reed

What is a datafest/ our data set



The problem posed to us:

Fatigue

What we were given:

- Self reported wellness data
- Self reported workload data
- GPS data, where the players were located during the games

The players were de-identified

We need more data!

But how?

---> we decided to try to re-identify the players as to recover more of their data that we could scrape from the web

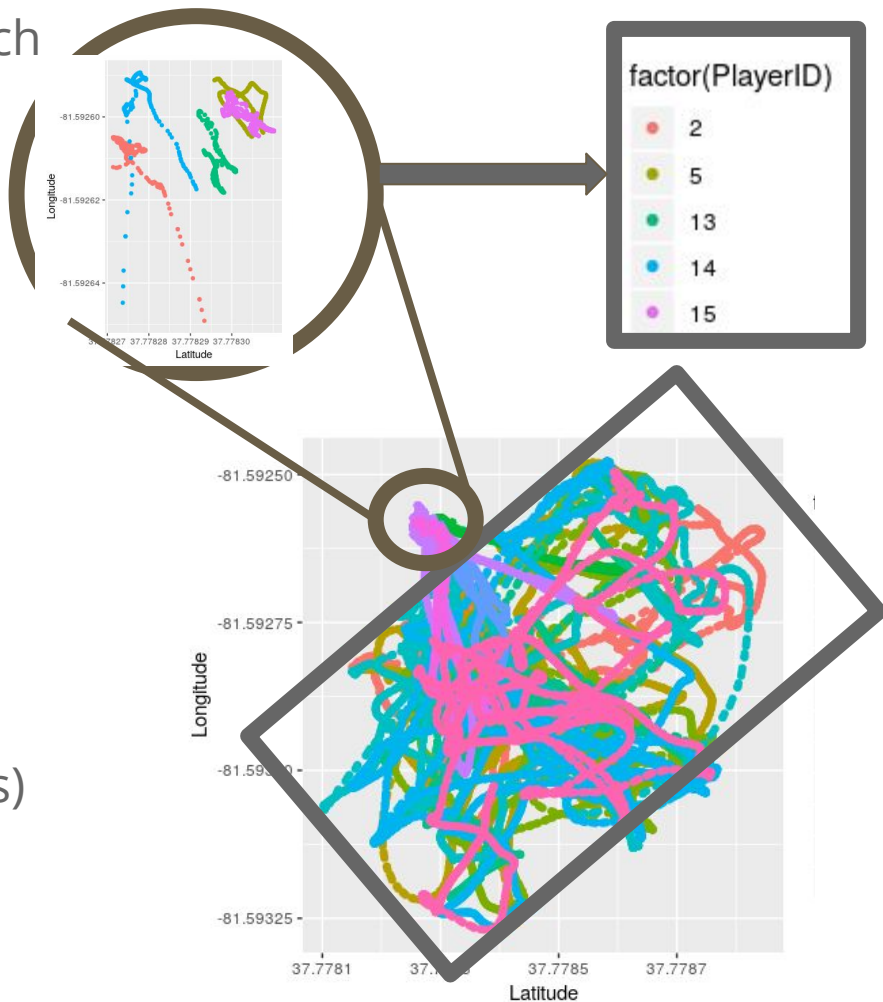
We need more data!

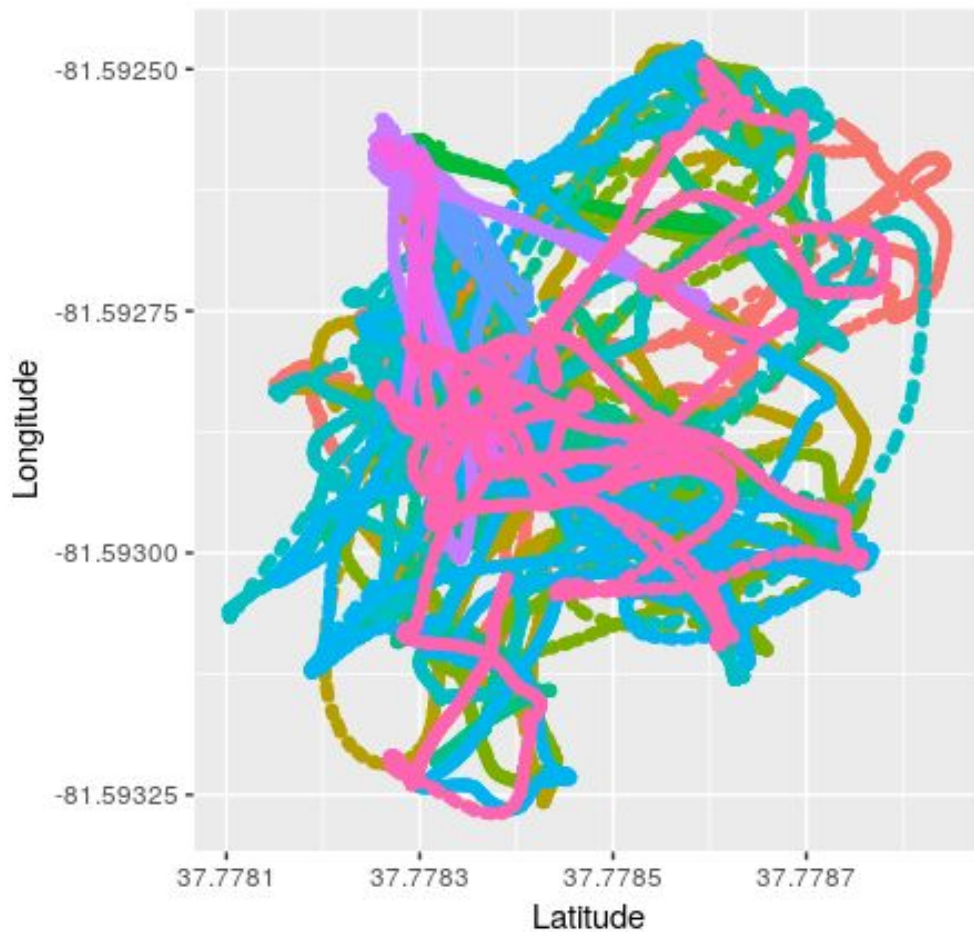
Successfully identified players based on substitution and try data from RWC Sevens

Data we recovered

- Age
- Weight/height
- Average Points Scored (in RWC Sevens)

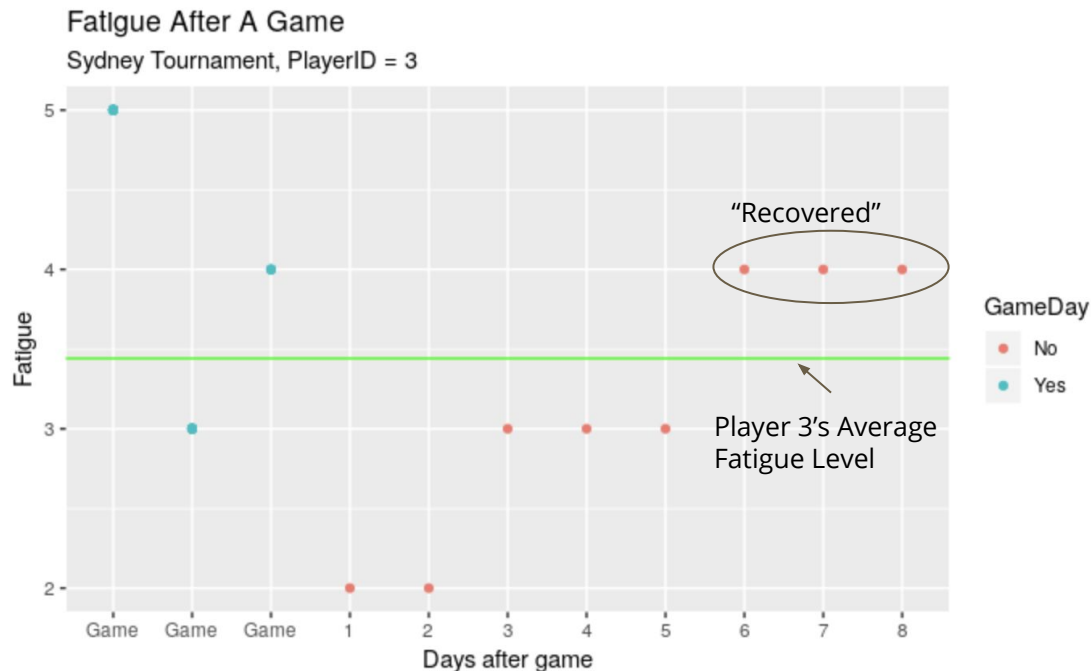
The Bench





ggplot2 +
the gps
data!

Making the Recovery Variable



Mean fatigue for each player

=

Threshold for "recovery"

Average days to recover = 3.38

We had to drop player 9

We also had to drop the
Kitakyushu tournament

Inference

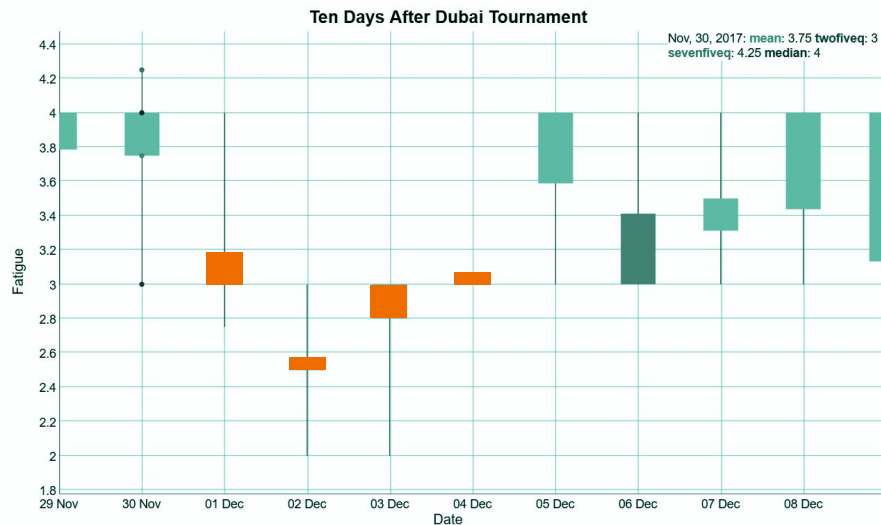
Response:

- Average Days to Recover

Explanatory:

- SleepHours, SleepQuality, Pain, Illness, Menstruation, Nutrition, USG*, TrainingReadiness, Duration, Age, Average Points Scored

*Measure of dehydration



What affects the recovery time?