Statistical Models for Sport in R

UseR! Brisbane
July 2018
Instructor: Stephanie Kovalchik

Overview:

I am pleased you will be participating in the *Statistical Models for Sport in R* tutorial. The tutorial will cover a number of skills and statistical models that are common in sports statistics and show how each can be implemented in **R**. The workshop will introduce participants to a range of **R** packages and real sports examples will be used throughout. It will be very hands-on!

After completing this workshop, participants will be able to gather and clean public sports data more effectively, explore data with graphics, and apply some common models used in sport.

A brief outline of the workshop is given below.

Tutorial Outline:

- Web scraping
- Sports models and real applications
  - *Bradley-Terry paired comparison models*. Application: Which team is the strongest in the English Football Association League?
  - *Pythagorean Theorem*. Application: How can we predict an NBA teams expected wins for the season?

- *Generalized Additive Models*. Application: How can we model the strike zone from baseball tracking data?
- *Forecasting*. Application: What is the chance that a tennis serve will be a service winner?

Preparing:

Please bring a laptop with R installed and install the following libraries via CRAN: dplyr, tidyr, ggplot2, rvest, jsonlite, stringr, mgcv, rjags, BradleyTerry2, lubridate, pitchRx

These additional libraries should be installed via github using devtools:

- Rselenium (`https://github.com/ropensci/RSelenium`)
- deuce (`https://github.com/skoval/deuce`)

There will also be part of the Web scraping material that will require Docker, which you can install here: `https://docs.docker.com/install/`

Contact:

If you have questions before or after the conference, please send an e-mail to me (Stephanie Kovalchik) at `s.a.kovalchik@gmail.com`.