**Web Scraping for Sports Data**

**Outline:**

Ever want to get involved in the sports analytics community? “Where can I find data?” is usually the first question. It’s an important question and learning to collect data from trusted open online sources is critical. Web scraping is such a process that can automate the extraction of data in an efficient way. This workshop aims to show beginners how to scrape useful information from both static and dynamic web pages, and how to convert unstructured data from the websites into structured data for further analysis in R. Topics covered include: high level overview of different web scraping techniques, exposure to several related R packages (rvest, RSelenum, deuce), and exercises on creating scraping functions. Following this, a full case study on scraping a list of sports web links along with further data analysis will be practiced.

A person wearing glasses and smiling at the camera

Description automatically generated**Instructor bio:**

Wanwan Xu is currently a Ph.D. student in statistics at the University of Connecticut. Her research interests include variable selection in high dimensional data and multimodal data fusion. She is currently working as a research assistant at the UConn Health Center providing technical support on medical claims data, as well as applying statistical methods to improve the identification of patients with suicide risk. She was the instructor and teaching assistant of several undergraduate and graduate level statistics courses, with experience coaching statistical software (R, Minitab) and tutoring data analysis techniques.

**Prerequisite:** Previous experience with R; a laptop with R and RStudio preinstalled.