Package ATPR for Statistical Analyses of Men's Professional Tennis

Stephanie Kovalchik[1,*]

1. RAND Corporation
*Contact author: skovalch@rand.org

Keywords: XML parsing, Web scraping, Sports, Tennis,

Since the publication of *Moneyball*, the use of statistical analysis to address challenging problems in sports has been more popular than ever. At the same time, professional sports organizations are collecting and publicizing more data about athletic events than ever. Although these data should ostensibly support quantitative analysis of sports, their extraction is frequently a barrier because the data needed for a complete analysis are often embedded within a web markup language and spread across multiple web pages. In this talk, I present a package ATPR that makes numerous data about the Association of Tennis Professionals (ATP) World Tour readily accessible in the R environment. The package consists of a collection of specialized `HTML`/`XML` parsing tools that enable users to extract data about tournament results, match statistics, and player rankings for more than 20 years of singles play on Tour without the need of manual web scraping. In this presentation, I will provide an overview of these tools and the specific tennis data they can be used to gather. It is hoped that the availability of ATPR will encourage more statistical research on tennis and also serve as a useful resource for teachers of statistics.