

curlingR: Accessible and Detailed World Curling Data

September 6, 2024

1. Introduction

Preserving and accessing historical data is essential for researching and enthusiasts who seek to understand a game from a statistical perspective. The life-cycle of digital content can be unpredictable, so long term archival should always be in consideration, especially when working with a sport that does not have a significant digital presence to begin with. Providing accessible archived resources supports public analysis and democratizes information, making it available to a broader audience of researchers and others while preventing loss due to lack of maintenance, private data acquisition, or other risks.

This package **curlingR** is designed to facilitate the preservation and retrieval of world curling data, ensuring its longevity and usability. The package provides users with functions to access compiled player and team statistics, shot-by-shot event data, and other event logs in cleaned formats.

curlingR utilizes API access to DoubleTakeout, a world curling data website and blog, to archive men's and women's world rankings, game logs, shot logs, and other event info primarily for events from 2018 through present day, along with some game and shot log data dating back to 2001 for major events like the Olympics and World Championships. The package also features the site's win probability model which functions based on variables like team ranking, game time remaining, and score margin.

2. Installation

curlingR can be installed directly from its Github repository using the *install_github* function in the **devtools** package.

```
if (!requireNamespace('devtools', quietly = TRUE)) {install.packages('devtools')}  
devtools::install_github(repo = "jbrooksdata/curlingR")
```

3. Use

curlingR is equipped with functions to access men's and women's historical world team rankings, game logs for events including the Olympics, World Championships, and other major tournaments like the Grand Slam of Curling Players' Championship, as well as shot logs for each match within the event.

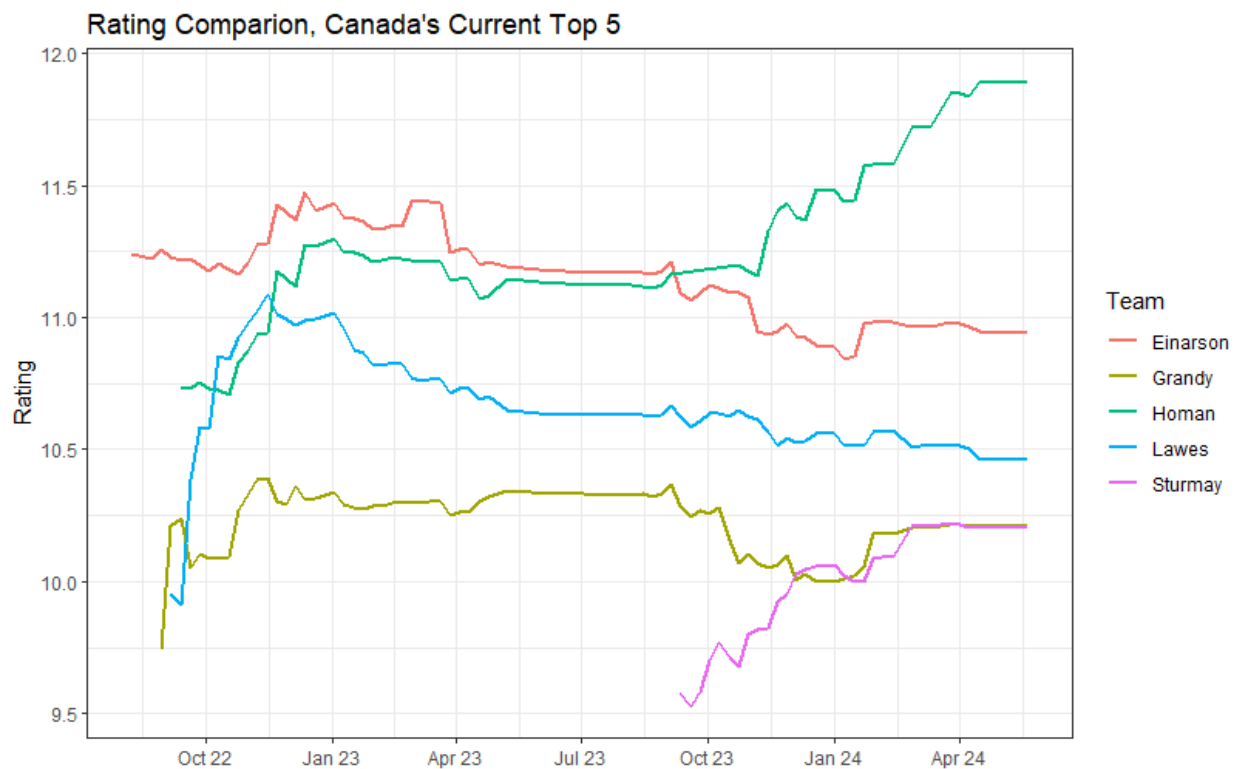
3.1 Team Rankings

DoubleTakeout world ranking data is typically updated a few times per month, and contains each team's overall record, record against other highly ranked teams, and change in team rank from the previous release. This data can be loaded using a single function call of *load_rankings_men()* or *load_rankings_women()*.

```
> load_rankings_women(2024) %>%
+   colnames()
[1] "x"      "Date"    "id"      "Rank"    "FirstName" "LastName"
[7] "Rating" "wins"    "Losses"  "Top10wins" "Top10Losses" "Top25wins"
[13] "Top25Losses" "Province" "Country" "Junior"   "RankDiff"   "RatingDiff"
[19] "season"
```

17-time Grand Slam winner and 2014 Olympic gold medalist Jennifer Jones, who TSN voted the greatest Canadian women's overall curler of all-time, announced prior to the 2024 Scotties tournament that she would be retiring from the sport following the conclusion of the event. Jones represented Canada in two Olympic games, including 2022 where the team finished 5th despite ending the tournament with the highest point total among all competing nations.

The example plot below shows how the current top Canadian women's teams' ratings have varied in the last two seasons, forecasting which captain may succeed Jones in representing Canada at the 2026 Olympics.



3.2 Investigating Event Logs and Player Stats

Basic information on major tournaments can be pulled using `load_events_men()` or `load_events_women()`

```
> load_events_men(2024) %>%
+   head()
  event_id event_name year gender
1     257   Baden Masters 2024   men
2     249 GSOC Players Championship 2024   men
3     251 world Championships 2024   men
4     245   USports Championships 2024   men
5     244       Brier 2024   men
6     253 Junior world Championships 2024   men
```

Each event includes some basic identifiers about its year and gender designation, as well as a unique `event_id` number that can be used to explore further. Using `load_events_games()`, users could examine how Sweden's men's team, the 2022 Olympic gold medalists, fared in the most recent edition of the World Championships.

```
  year event event_id game_id start_time round draw sheet
1 2024 world Championships 251 203 2024-04-07 15:00:00 <NA> Gold Medal Game <NA>
2 2024 world Championships 251 275 2024-04-07 10:00:00 <NA> Bronze Medal Game <NA>
3 2024 world Championships 251 276 2024-04-06 16:00:00 <NA> Semi-final B
4 2024 world Championships 251 277 2024-04-06 16:00:00 <NA> Semi-final D
5 2024 world Championships 251 278 2024-04-06 10:00:00 <NA> Qualification B
6 2024 world Championships 251 279 2024-04-06 10:00:00 <NA> Qualification D
  team1 team2 team1_score team2_score lsfe_team ee
1 CAN (Gushue) SWE (Edin) 5 6 SWE (Edin) 0
2 ITA (Retornaz) SCO (Mouat) 7 6 SCO (Mouat) 1
3 ITA (Retornaz) SWE (Edin) 3 5 SWE (Edin) 0
4 CAN (Gushue) SCO (Mouat) 9 4 CAN (Gushue) 0
5 SCO (Mouat) USA (Shuster) 8 4 SCO (Mouat) 0
6 GER (Muskatewitz) ITA (Retornaz) 3 8 GER (Muskatewitz) 0
```

The results show that Niklas Edin's Swedish squad won the gold medal game by a score of 6-5 over Canada, which may not be a surprising result when looking through the games list and seeing that they opened their tournament play with an 8-1 victory over the Netherlands.

`load_game_data()` can then be used to look at shot-by-shot stats of a given event. Shot-level data can help determine if a team was simply outclassed, or if there was a major momentum shift in a particular game that influenced the outcome.

Users can also call `load_events_stats()` to return overall player stats for each event. Looking into the player shot quality over expectation metric for the same event reveals that all four primary Swedish players had elite performances, with three finishing the tournament in the top 10.

```

> load_events_stats(251) %>%
+   filter(n > 100) %>%
+   mutate(avgOverExp = avg_pts - exp_pts) %>%
+   arrange(-avgOverExp)

```

	athlete	team	avg_pts	exp_pts	n	event_id	avgOverExp
1	Mark Nichols	CAN (Gushue)	3.6190	3.2491	252	251	0.3699
2	Brad Gushue	CAN (Gushue)	3.5280	3.2084	250	251	0.3196
3	Niklas Edin	SWE (Edin)	3.4802	3.1842	252	251	0.2960
4	Christoffer Sundgren	SWE (Edin)	3.7482	3.4524	274	251	0.2958
5	Bruce Mouat	SCO (Mouat)	3.4579	3.1716	273	251	0.2863
6	Amos Mosaner	ITA (Retornaz)	3.5176	3.2466	284	251	0.2710
7	Hammy McMillan	SCO (Mouat)	3.7007	3.4483	274	251	0.2524
8	Yannick Schwaller	SUI (Schwaller)	3.4912	3.2676	226	251	0.2236
9	Rasmus Wranaa	SWE (Edin)	3.5000	3.2793	274	251	0.2207
10	Schwarz-van Berkel	SUI (Schwaller)	3.3571	3.1389	224	251	0.2182

3.3 Win Probability

curlingR also features DoubleTakeout's True Win Probability data. *load_win_prob()* can be used for projections based on factors including team ranking, score margin, and ends remaining for either gender classification.

4. What's Next?

A next step for **curlingR** could be the inclusion of metrics outside of DoubleTakeout, such as the World Curling Tour team rankings.

DoubleTakeout describes the World Curling Tour's ranking system as conveying "how good a team's season has been from an accomplishment standpoint," as opposed to the pure "ability of a team" (DoubleTakeout, 2020). This is because some of a team's worst performances are dropped from consideration, meaning teams that participate in more events are at an advantage.

DoubleTakeout's rating system utilizes a Bradley-Terry model on two calendar years of data with a weight advantage given to more recent events. A two-year scope of data provides more stability, preventing outlier performances from impacting predictions.

5. Conclusion

curlingR seeks to preserve world curling data and maintain its usability for researchers and other enthusiasts in the sports data space. As media coverage of this sport is very limited, especially in the United States, this package can make it much easier for coaches, players, and all others to analyze games and deepen their understanding of the sport.

6. References

DoubleTakeout. 2021. "2. The Ratings." <https://doubletakeout.com/blog/2-the-ratings/>