

Web scraping in R using the rvest package

R Ladies Meetup

Kathryn Morrison, Co-Founder

Precision Analytics Inc

March 15, 2018

What is web scraping and why is it useful?

Say you want to download the Olympics data to do some analysis:

VIDEO | SCHEDULE | RESULTS | MEDALS | NEWS | SPORTS | VR

Q | FR

Overall

Medal winners

Rank	Country	G	S	B	Total	
1	 Norway	14	14	11	39	>
2	 Germany	14	10	7	31	>
3	 Canada	11	8	10	29	>
4	 United States	9	8	6	23	>
5	 Netherlands	8	6	6	20	>
6	 South Korea	5	8	4	17	>
7	 Olympic Athlete from Russia	2	6	9	17	>
8	 Switzerland	5	6	4	15	>
9	 France	5	4	6	15	>
10	 Sweden	7	6	1	14	>
11	 Austria	5	3	6	14	>

You could manually copy and paste the data into a spreadsheet, but it would take a long time.

What is web scraping and why is it useful?

Web scraping is basically automating the collection of unstructured data from online sources into a format you can use, e.g., plain text, csv, JSON, etc

You can web scrape in almost any language (e.g., python, java, and R!)

Examples of webscraping with rvest

Olympics data