Introduction to Web Scraping with R

Introduction

Simon Munzert | IPSDS

Welcome!

Welcome!

Thank you for your interest in the course!

Welcome!

Thank you for your interest in the course!

I'm glad that I will have the opportunity to give you a gentle introduction to web scraping with R.

Welcome!

Thank you for your interest in the course!

I'm glad that I will have the opportunity to give you a gentle introduction to web scraping with R.

Just a few words on my personal background:

- Lecturer in Political Data Science at the Hertie School of Governance, Berlin
- political scientist by training, focus on public opinion and political communication
- working with web-based data since about 2010
- fascinated by the new opportunities that web-based data provide, also when it comes to measuring public opinion, awareness, or political behavior
- as a result, I have become a somewhat seasoned gatherer of web data

Prerequisites, goals, and outline

Prerequisites

I expect that you . . .

- have an interest in coding, that is programmatic means to solve problems
- are familiar with the statistical software R. Knowledge about the "tidyverse" packages, in particular `dplyr` and `magrittr`, is of help.

Prerequisites

I expect that you ...

- have an interest in coding, that is programmatic means to solve problems
- are familiar with the statistical software R. Knowledge about the "tidyverse" packages, in particular `dplyr` and `magrittr`, is of help.

If you are familiar with R but have no experience in working with these packages, the best place to learn them is the primary reading "R for Data Science". It's freely accessible here: http://r4ds.had.co.nz/

Goals

After attending this course, . . .

Goals

After attending this course, . . .

- you have acquired basic knowledge of web technologies

Goals

After attending this course, . . .

- you have acquired basic knowledge of web technologies
- you are able to scrape information from static and dynamic websites using R

Goals

After attending this course, . . .

- you have acquired basic knowledge of web technologies
- you are able to scrape information from static and dynamic websites using R
- you are able to access web services (APIs) with R

Goals

After attending this course, . . .

- you have acquired basic knowledge of web technologies
- you are able to scrape information from static and dynamic websites using R
- you are able to access web services (APIs) with R
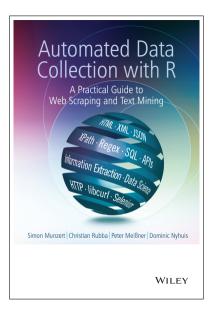- you can build up and maintain your own original sets of web-based data

Course outline

Unit	Topic
1	Introduction – Web technologies
2	Scraping static webpages
3	Scraping dynamic webpages and good practice
4	Tapping APIs

The accompanying book

- contains most of which I tell you during the course (but much more, and at times more accurate)
- written between 2012 and 2014 → not entirely up-to-date anymore, but I will provide updated material during the course
- homepage with materials: www.r-datacollection.com
- if you find any errors in the book, please let me know!