

Websurfing with R

Tom Theile, Max-Planck-Institute for demographic research, theile@demogr.mpg.de

July 2022

Reading a website with R

Requesting webpages can be done not only by using the webbrowser, but also by using any programming language. We will use R to do this. And we will use the R-package `rvest`, because it makes this very convenient.

Please google `rvest`, and you will get the installation-instructions:

```
# The easiest way to get rvest is to install the whole tidyverse:
install.packages("tidyverse")

# Alternatively, install just rvest:
install.packages("rvest")
```

Now you can download any webpage with the simple command `read_html`:

```
library(rvest)
simple_page <- read_html("https://user.demogr.mpg.de/theile/Files/edsd/simple.html")
print(paste(simple_page))
```

```
## [1] "<!DOCTYPE html>\n<html><body>\r\n\r\n<h1>Hello Barcelona!</h1>\r\n\r\n<p>This is some text.</p>
```

```
<!DOCTYPE html>
<html>
<body>

<h1>Hello Barcelona!</h1>

<p>This is some text.</p>

<a href="https://www.example.com">This is a link</a>
<a href="./simple2.html">This is also a link</a>
</body>
</html>
```