

Technical preparations

for the workshop on web scraping with R

Resul Umit

02 June 2022

This document includes instructions for preparing your computer for the workshop on web scraping with R.

If you encounter issues while following these instructions, please do not hesitate to email me at resulmit@gmail.com — you can expect a prompt, welcoming reply.

I am confident that we can solve any issues that you might encounter — provided that (1) you contact me as early as possible and that (2) you have administrative rights (that is, you can install software) on your computer.

1. Software

Please download and install the following software to your computer:

- **R**: <https://cloud.r-project.org/>
- **RStudio**: <https://www.rstudio.com/products/rstudio/download/>
- **Java**: <https://www.java.com/en/download/>
- **Chrome**: <https://www.google.com/chrome/>

You may already have one or more of them installed. In that case, please consider updating them to their latest versions, which will ensure that we will all be working with exactly the same set of tools during the workshop.

Once you install the above, I recommend restarting your computer.

RSelenium

RSelenium is a great package for, among others, visiting a website, clicking on a button, or scrolling down a page — all controlled from within R. We will use it to scrape *dynamic* websites.

After the restart, and while in R or RStudio, please install the RSelenium package.

```
install.packages("RSelenium")
```

Once the package is installed, first restart your R session (from the RStudio menu, follow **Session -> Restart R**), and then load the package.

```
library(RSelenium)
```

Finally, try starting Chrome from within R. The code below should lead to a Chrome window, as shown in Figure 1, popping up. The window should not disappear on its own.

```
driver <- rsDriver(port = sample(4500:4600, 1))
```

If this works, congratulations, your computer is ready for the workshop. If it does not work as expected, please see the next section.

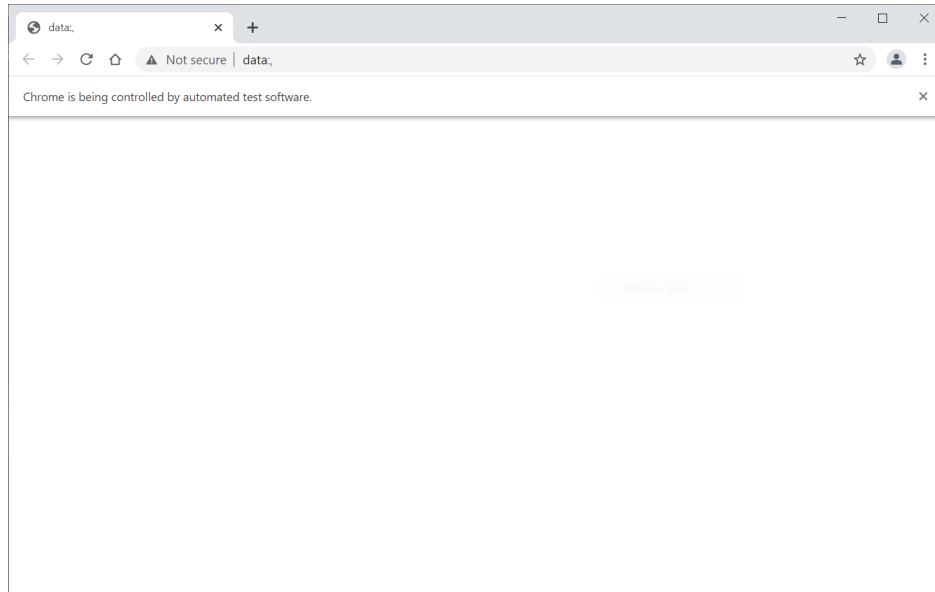


Figure 1: If you see this window popping up, and it stays still until you close it, you are ready for the workshop.

2. Troubleshooting

The instructions above may not work on all computers, for various reasons. In this section, I will provide you with a solution to one of the most common problems: a mismatch between the versions of *ChromeDriver* that comes with *RSelenium* and the *Chrome* browser on your computer. There are two main symptoms of this problem:

- a window pops up for a split second, and then disappears without any error messages
- nothing pops up, leading to an error message as follows:

```
Selenium message:session not created: This version of ChromeDriver  
only supports Chrome version XX
```

If either of these is the case for you, please run the following code in R Console. This will result in a list of numbers (driver versions):

```
binman::list_versions("chromedriver")
```

```
## $win32  
## [1] "100.0.4896.20" "100.0.4896.60" "102.0.5005.27" "102.0.5005.61"  
## [5] "103.0.5060.24" "89.0.4389.23" "90.0.4430.24" "91.0.4472.19"  
## [9] "99.0.4844.35" "99.0.4844.51"
```

One of the resulting numbers that you see *in your R Console* is possibly what you need to get it to work, but we do not know which one that might be. It is a trial-and-error situation. Try supplying them *one by one* to the `rsDriver` function, with the `chromeever` argument, as shown below. For me, it is the third number that works. This might be different for you.

```
driver <- rsDriver(port = sample(4500:4600, 1), chromeever = "102.0.5005.27")
```

If this does not help, please google the error message that you have to see if there is a solution available on the internet.

If googling does not help you solve the issue, you may wish to use a different browser, such as Firefox. It would be great if all of us use the same browser (that is, Chrome) during the workshop, but this is not absolutely necessary. You would also be fine by using Firefox.

Assuming that you have Firefox installed on your computer, try running the following code. Again, if a browser pops up as a result, and it stays on, you are ready for the workshop.

```
driver <- rsDriver(port = sample(4500:4600, 1), browser = "firefox")
```

If none of these solutions help, please send me an email at resulmit@gmail.com with as much information as possible, including exact error message that you get. I would be happy to hear from you.