

Preperations for the Workshop on Web Scraping

Resul Umit

17 May 2021

One of the R packages that we will use for this workshop, **RSelenium**, is a package that requires a specific setup. This setup might be short, but it frequently goes wrong, which can easily derail any workshop. To avoid having to deal with related issues during the workshop, it would be great if you could please follow a set of instructions on this document at your earliest opportunity.

If you encounter any issues while following the instructions on this document, please do not hesitate to email me. I would be happy to hear from you, and I am confident that we can solve those issues — provided that you have administrative rights (e.g., 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, which will ensure that we will all be working with exactly the same set of tools during the workshop.

Once you install them, I would recommend re-starting your *computer*.

2. RSelenium

After the re-start, and while in R or RStudio, please install the R package **RSelenium** to your computer.

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

Once the package is installed, first restart your *R session*, and then load the package.

```
library(RSelenium)
```

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

If this works, congratulations, you are ready for the workshop. You can close the browser and R or RStudio.

```
rD <- rsDriver(port = sample(4500:4600, 1))  
remDr <- rD[["client"]]
```

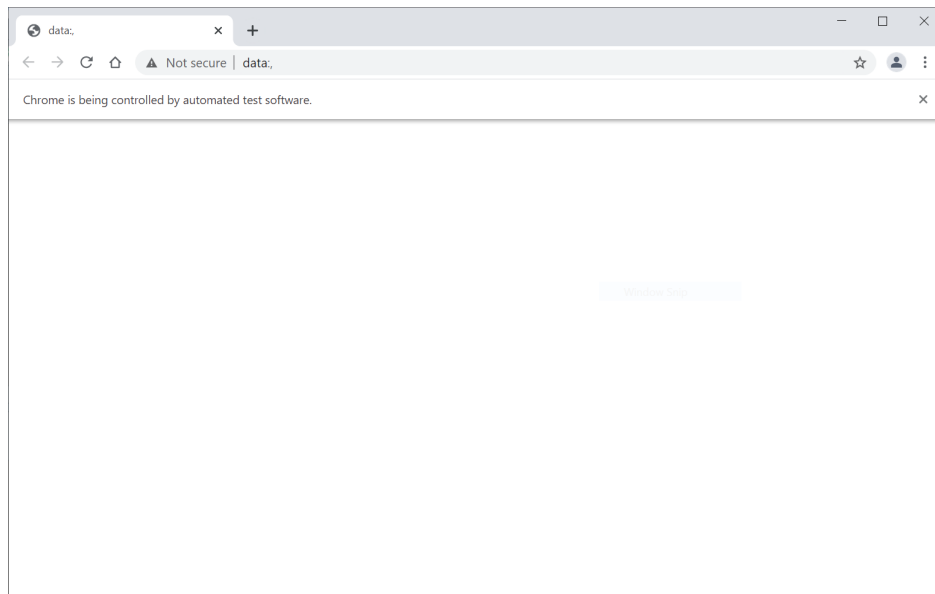


Figure 1: If you see this window popping up, and it stays still until you close it, you are ready for the workshop.

If it does not work as expected — it happens (a lot!) — please see the next section.

3. Troubleshooting

The instructions above may not work for everyone — this is perfectly normal for *RSelenium*.

- It might be that the window pops up for a split second, and then disappears without any error messages.
- It might also be that you get an error message similar to:

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

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

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

## $win32
## [1] "89.0.4389.23" "90.0.4430.24" "91.0.4472.19"
```

One of the resulting numbers that *you* get is what you possibly need, but we do not know which one. It is a trial-and-error situation. Try supplying them *one by one* to the `rsDriver` function, with the `chromeversion` argument.

For me, it is the middle number that works. Please note that you should use *the numbers that you get, not the numbers on this document*.

```
rD <- rsDriver(port = sample(4500:4600, 1),
               chromeversion = "90.0.4430.24")
remDr <- rD[["client"]]
```

If this does not help, and especially if you get a different error message, please try to google the message to see if there is a solution available on the internet.

If googling doesn't solve your problem, please send me an email with as much information as possible about the problem, including the exact error message.

If we cannot solve the problem via emails, we can set up a pre-workshop meeting.