

Workshop Preparations

Resul Umit

19 May 2021

This document includes instructions to prepare your computer for the workshop on automated web scraping.

One of the R packages that we will use for the workshop, **RSelenium**, is a package that requires an unusual setup. This setup might be short, but it frequently goes wrong.

To avoid having to deal with potential issues during the workshop, it would be great if you could please follow the instructions below at your earliest opportunity.

If you encounter any issues while following these instructions, 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 restarting your computer.

2. RSelenium

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

After the restart, and while in R or RStudio, please install `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, try starting a browser from within R. The R code below should lead to a Chrome window, as shown in Figure 1, popping up. It should not disappear on its own.

If this works, congratulations, you are ready for the workshop.

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

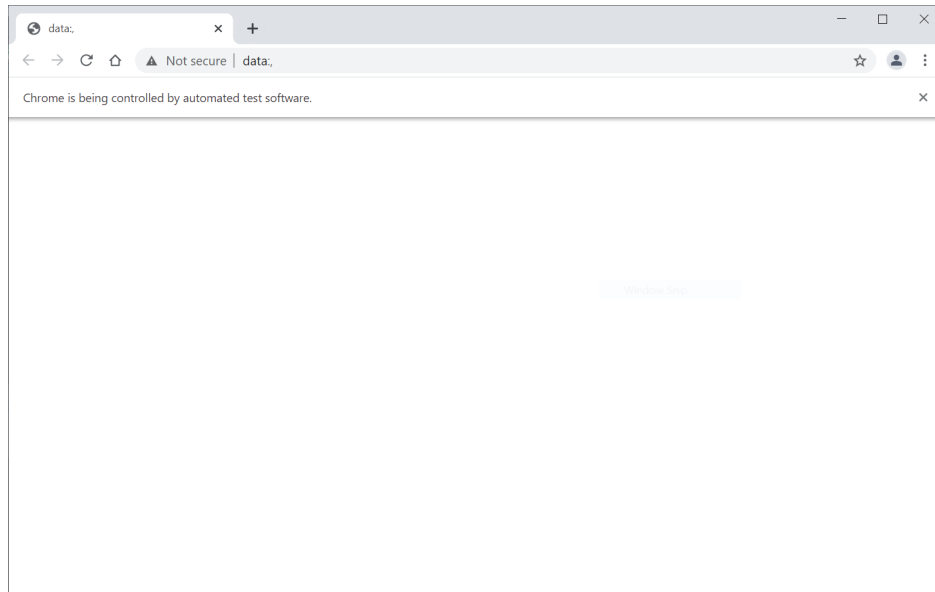


Figure 1: If you see this window popping up, and it stays still until you close it, you are ready for the workshop.

If it does not work as expected — it happens — please see the next section.

3. Troubleshooting

The instructions above may not work, 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 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 `chromever` argument.

For me, it is the middle number that works. This might be different for you. Please note that you should try *the numbers that you get, not the numbers on this document*.

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

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

If googling does not solve the issue, please send me an email with as much information as possible, including the exact error message.

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