

Webscraping with RSelenium

Reed Garvin and Francesca Giacco

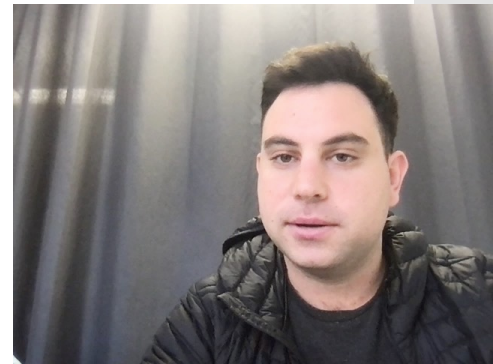


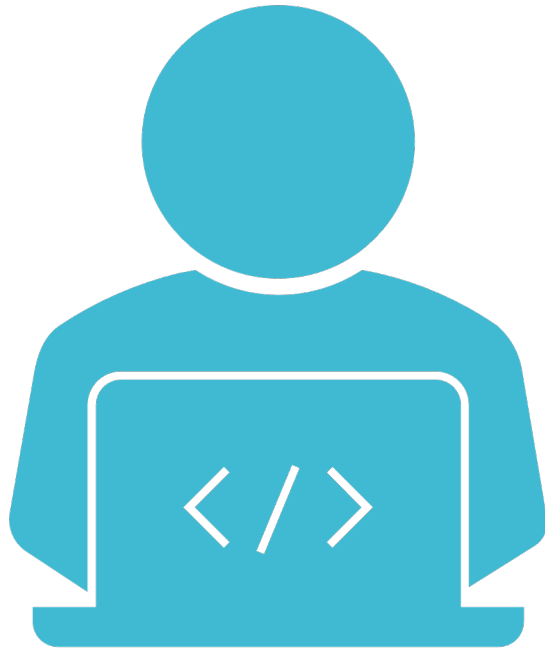
What is RSelenium?

- **Selenium** : open source tool that automates web browsers allowing the user to write code to be used on dynamic websites
- Can be used to write in different languages (Java, Python, C++)



- **RSelenium**: R package to use Selenium





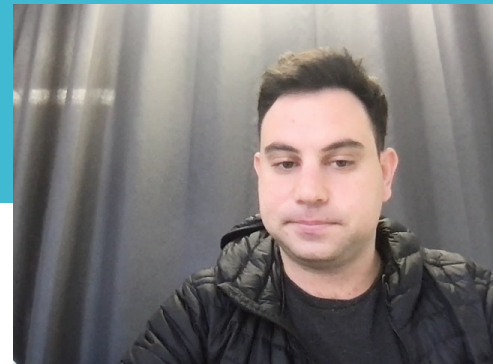
Why use it?

Dynamic webscraping

Allows interaction with webpages

Debugging of sites

Examples: clicking, typing, going back and forward, export html tables of dynamic web page info)

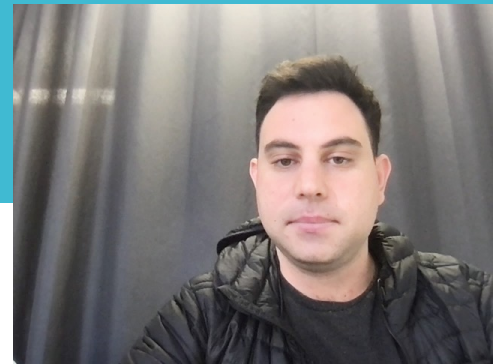


Different from static
webpages

Some elements appear only
if you interact with the page

Cannot be accessed from
the static html webscrape

Why do we
need it?





Features



How to set it up

Download Java



Download Selenium Driver



Download Docker



Download Firefox



How to get your selenium server running in R



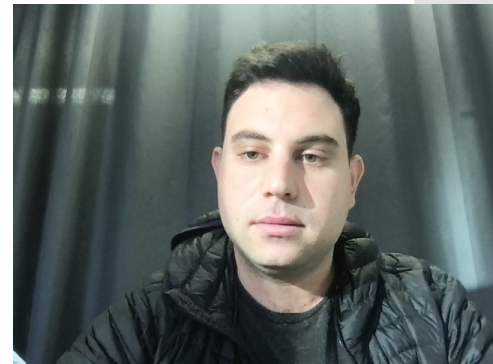
Download **RSelenium**

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



Load RSelenium

```
library(RSelenium)
```



Almost there..



Run Selenium Standalone from Docker

Use this code in docker to download the Selenium server

```
docker run -d -p 4444:4444 -p 7900:7900 --shm-size="2g"
```



cranky_joliot [selenium/stand...](#)

RUNNING PORT: 4444



Connect to Standalone

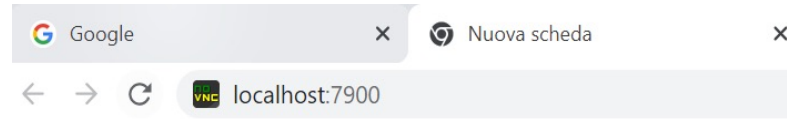
```
# initiate Selenium driver  
  
remDr <- RSelenium::remoteDriver  
(remoteServerAddr = "localhost",  
  port = 4444L,  
  browserName = "firefox")
```



Open



Open local host (use default password secret)



Connect

```
remDr$open()  
|
```

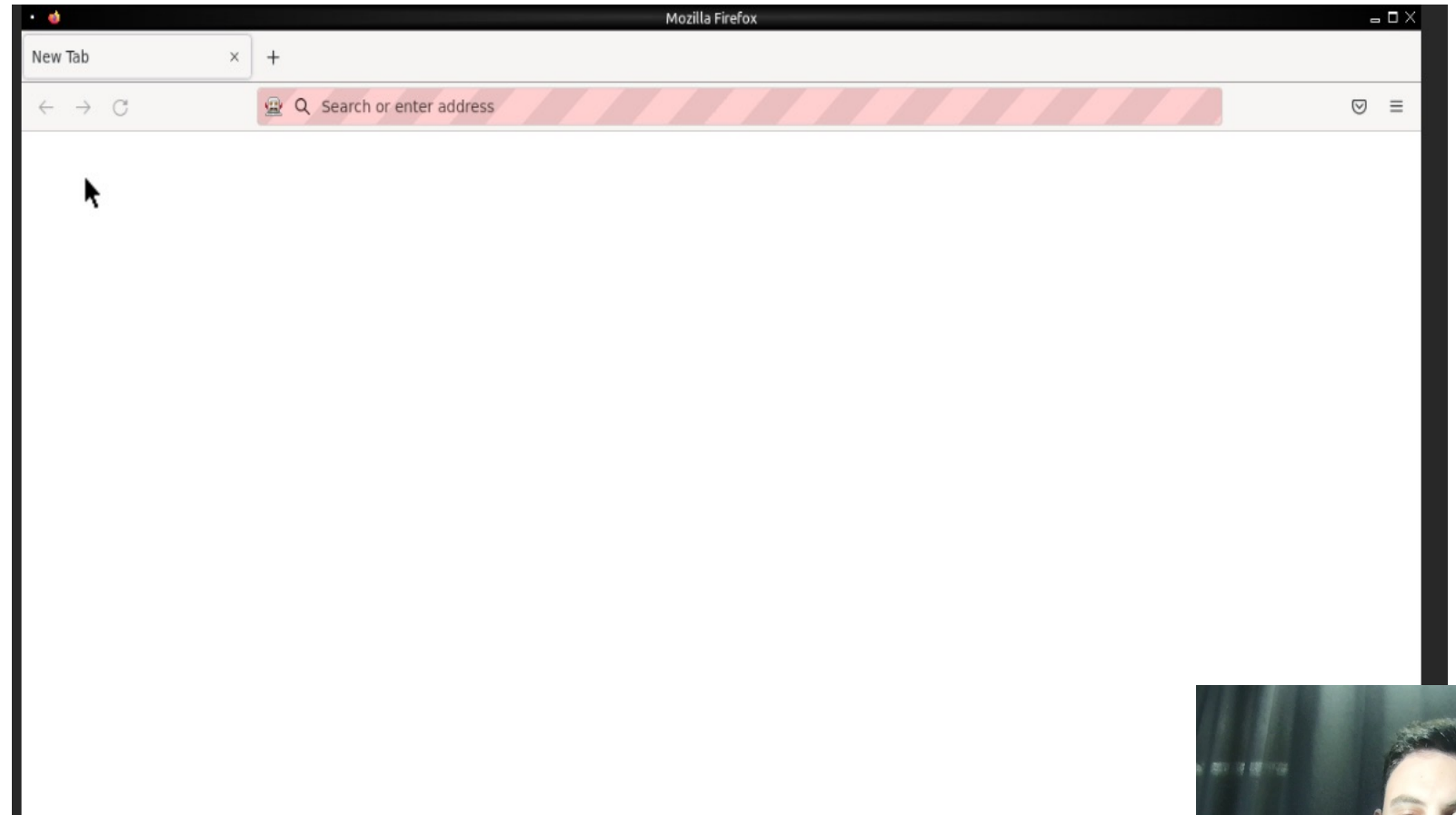


Check Local Host



Be proud of
yourself!!

If you get this:



Main functions

```
20 url<- "http://example.com"
21 remDr$navigate(url)
22
23 Elem<- remDr$findElement(using="xpath", value="[insert xpath]")
24 Elem2<- remDr$findElement(using="class", value="insert class name")
25
26 Elem$clickElement()
27 Elem2$sendKeysToElement(list("Write what you want"))
28
29 remDr$goBack()
30 remDR$goForward()
31 remDR$refresh()
32 remDr$closeServer()
```

Important: name all your variables with names Containing "Elem"
Example: searchElem, videoElem

