

Sub-problem 1

```
if(!require("RSelenium"))  
  ↪ install.packages("RSelenium")  
library(RSelenium)  
if(!require("rvest")) install.packages("rvest")  
library(rvest) #for read_html(), html_elements()...  
#Free all ports  
  system("taskkill /im java.exe /f", intern=FALSE,  
    ↪ ignore.stdout=FALSE)  
#Start a selenium & Assign client to an R-object  
  rD <- rsDriver(port = 4561L, browser = "firefox")  
  remDr <- rD[["client"]]  
  #remDr$quit
```

Define savepage()

```
#Load url & return content as r-object
savepage <- function(url){
  #Navigate to starting page
  remDr$navigate(url)
  #Wait until page is loaded
  Sys.sleep(abs(rnorm(1, 2, 1)))
  #Save content to an R-object
  remDr$getPageSource(header = TRUE)[[1]] %>%
    read_html() %>%
    return()
}
```

Note: [[1]] behind getPageSource() unlist the output -> makes it searchable

Sub-problem 2

#navigate to url & save content as r-object

page <-

↪ savepage(["https://www.bundestag.de/webarchiv/Ausschuess](https://www.bundestag.de/webarchiv/Ausschuess)

Sub-problem 3

page

```
## {html_document}  
## <html xml:lang="de" dir="ltr" class="detection-firefox"  
## [1] <head>\n<meta http-equiv="Content-Type" content="tex  
## [2] <body class="bt-archived-page">\n  <div class="bt-ar
```

Sub-problem 4

```
html_element(page, xpath=  
    ↪ "head/meta[@property='og:image']")
```

```
## {html_node}
```

```
## <meta property="og:image" content="https://www.bundestag
```