```
# Instalar los paquetes si no están ya instalados
> install.packages("rvest")
WARNING: Rtools is required to build R packages but is not currently inst
alled. Please download and install the appropriate version of Rtools befo
re proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/DELL/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/rvest_1
.0.4.zip'
Content type 'application/zip' length 308590 bytes (301 KB)
downloaded 301 KB

package 'rvest' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\DELL\AppData\Local\Temp\RtmpueXxkt\downloaded_packages
> install.packages("dplyr")
WARNING: Rtools is required to build R packages but is not currently inst
alled. Please download and install the appropriate version of Rtools befo
re proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/DELL/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/dplyr_1
.1.4.zip'
Content type 'application/zip' length 1581291 bytes (1.5 MB)
downloaded 1.5 MB

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\DELL\AppData\Local\Temp\RtmpueXxkt\downloaded_packages
>
> # Cargar los paquetes
> library(rvest)
> library(dplyr)

Adjuntando el paquete: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

>
> # URL de la página de Wikipedia sobre R
> url <- "https://en.wikipedia.org/wiki/R_(programming_language)"
>
> # Leer la página web
> webpage <- read_html(url)
>
> # Extraer el título de la página
> title <- webpage %>% html_node("h1") %>% html_text()
> print(title)
[1] "R (programming language)"
>
> # Extraer todos los párrafos
> paragraphs <- webpage %>% html_nodes("p") %>% html_text()
> head(paragraphs)
```

```
[1] "\n"
[2] "R is a programming language for statistical computing and data visua
lization. It has been adopted in the fields of data mining, bioinformatic
s, and data analysis.[8]"
[3] "The core R language is augmented by a large number of extension pack
ages, containing reusable code, documentation, and sample data.\n"
[4] "R software is open-source and free software. It is licensed by the G
NU Project and available under the GNU General Public License.[3] It is w
ritten primarily in C, Fortran, and R itself. Precompiled executables are
provided for various operating systems.\n"
[5] "As an interpreted language, R has a native command line interface. M
oreover, multiple third-party graphical user interfaces are available, su
ch as RStudio—an integrated development environment—and Jupyter—a noteboo
k interface.\n"
[6] "R was started by professors Ross Ihaka and Robert Gentleman as a pro
gramming language to teach introductory statistics at the University of A
uckland.[9] The language was inspired by the S programming language, with
most S programs able to run unaltered in R.[6] The language was also insp
ired by Scheme's lexical scoping, allowing for local variables.[1]"
>
> # Extraer todas las tablas de la página
> tables <- webpage %>% html_nodes("table")
>
> # Convertir la primera tabla a un data frame
> if(length(tables) > 0) {
+     first_table <- tables[[1]] %>% html_table(fill = TRUE)
+     print(first_table)
+ }
# A tibble: 19 × 2
   X1                          X2
   <chr>                       <chr>
 1 ""                          ""
 2 "R terminal"                "R terminal"
 3 "Paradigms"                 "Multi-paradigm: procedural, object-orien
ted, functional, reflec…
 4 "Designed by"               "Ross Ihaka and Robert Gentleman"
 5 "Developer"                 "R Core Team"
 6 "First appeared"            "August 1993; 30 years ago (1993-08)"
 7 ""                          ""
 8 "Stable release"            "4.4.1[2] \n   / 14 June 2024; 1 day ag
o (14 June 2024)"
 9 ""                          ""
10 "Typing discipline"         "Dynamic"
11 "Platform"                  "arm64 and x86-64"
12 "License"                   "GNU GPL v2[3]"
13 "Filename extensions"       ".mw-parser-output .plainlist ol,.mw-pars
er-output .plainlist ul…
14 "Website"                   "www.r-project.org"
15 "Influenced by"             "Influenced by"
16 "Lisp\nS[6]\nScheme[1]"     "Lisp\nS[6]\nScheme[1]"
17 "Influenced"                "Influenced"
18 "Julia[7]"                  "Julia[7]"
19 "R Programming at Wikibooks" "R Programming at Wikibooks"
>
> # Guardar los párrafos en un archivo de texto
> writeLines(paragraphs, 'paragraphs.txt')
>
> # Guardar la primera tabla en un archivo CSV (si existe)
> if(exists("first_table")) {
+     write.csv(first_table, 'first_table.csv', row.names = FALSE)
+ }
>
> # Extraer el primer párrafo
> first_paragraph <- webpage %>% html_node('p') %>% html_text()
> print(first_paragraph)
```

```
[1] "\n"
>
> # Extraer la tabla de información (infobox)
> infobox <- webpage %>% html_node('.infobox') %>% html_table()
> print(infobox)
# A tibble: 19 × 2
   X1                          X2
   <chr>                       <chr>
 1 ""                          ""
 2 "R terminal"                "R terminal"
 3 "Paradigms"                 "Multi-paradigm: procedural, object-orien
ted, functional, reflec…
 4 "Designed by"               "Ross Ihaka and Robert Gentleman"
 5 "Developer"                 "R Core Team"
 6 "First appeared"            "August 1993; 30 years ago (1993-08)"
 7 ""                          ""
 8 "Stable release"            "4.4.1[2] \n   / 14 June 2024; 1 day ag
o (14 June 2024)"
 9 ""                          ""
10 "Typing discipline"         "Dynamic"
11 "Platform"                  "arm64 and x86-64"
12 "License"                   "GNU GPL v2[3]"
13 "Filename extensions"       ".mw-parser-output .plainlist ol,.mw-pars
er-output .plainlist ul…
14 "Website"                   "www.r-project.org"
15 "Influenced by"             "Influenced by"
16 "Lisp\nS[6]\nScheme[1]"     "Lisp\nS[6]\nScheme[1]"
17 "Influenced"                "Influenced"
18 "Julia[7]"                  "Julia[7]"
19 "R Programming at Wikibooks" "R Programming at Wikibooks"
>
> # Limpiar y estructurar la tabla de información
> infobox_clean <- infobox %>%
+     rename(Attribute = 1, Value = 2) %>%
+     filter(!is.na(Attribute))
> print(infobox_clean)
# A tibble: 19 × 2
   Attribute                   Value
   <chr>                       <chr>
 1 ""                          ""
 2 "R terminal"                "R terminal"
 3 "Paradigms"                 "Multi-paradigm: procedural, object-orien
ted, functional, reflec…
 4 "Designed by"               "Ross Ihaka and Robert Gentleman"
 5 "Developer"                 "R Core Team"
 6 "First appeared"            "August 1993; 30 years ago (1993-08)"
 7 ""                          ""
 8 "Stable release"            "4.4.1[2] \n   / 14 June 2024; 1 day ag
o (14 June 2024)"
 9 ""                          ""
10 "Typing discipline"         "Dynamic"
11 "Platform"                  "arm64 and x86-64"
12 "License"                   "GNU GPL v2[3]"
13 "Filename extensions"       ".mw-parser-output .plainlist ol,.mw-pars
er-output .plainlist ul…
14 "Website"                   "www.r-project.org"
15 "Influenced by"             "Influenced by"
16 "Lisp\nS[6]\nScheme[1]"     "Lisp\nS[6]\nScheme[1]"
17 "Influenced"                "Influenced"
18 "Julia[7]"                  "Julia[7]"
19 "R Programming at Wikibooks" "R Programming at Wikibooks"
>
> # Agregar una columna ficticia de datos numéricos para el análisis
> set.seed(123) # Para reproducibilidad
```

```
> infobox_clean$NumericValue <- sample(1:100, nrow(infobox_clean), replac
e = TRUE)
> print(infobox_clean)
# A tibble: 19 × 3
   Attribute                    Value
NumericValue
   <chr>                        <chr>
<int>
 1 ""                           ""
31
 2 "R terminal"                 "R terminal"
79
 3 "Paradigms"                  "Multi-paradigm: procedural, object-orien
ted, funct…           51
 4 "Designed by"                "Ross Ihaka and Robert Gentleman"
14
 5 "Developer"                  "R Core Team"
67
 6 "First appeared"             "August 1993; 30 years ago (1993-08)"
42
 7 ""                           ""
50
 8 "Stable release"             "4.4.1[2] \n   / 14 June 2024; 1 day ag
o (14 June 2…           43
 9 ""                           ""
14
10 "Typing discipline"          "Dynamic"
25
11 "Platform"                   "arm64 and x86-64"
90
12 "License"                    "GNU GPL v2[3]"
91
13 "Filename extensions"        ".mw-parser-output .plainlist ol,.mw-pars
er-output …           69
14 "Website"                    "www.r-project.org"
91
15 "Influenced by"              "Influenced by"
57
16 "Lisp\nS[6]\nScheme[1]"      "Lisp\nS[6]\nScheme[1]"
92
17 "Influenced"                 "Influenced"
9
18 "Julia[7]"                   "Julia[7]"
93
19 "R Programming at Wikibooks" "R Programming at Wikibooks"
99
>
> # Calcular medidas de resumen estadístico
> summary_stats <- infobox_clean %>%
+     summarise(
+         Mean = mean(NumericValue),
+         Median = median(NumericValue),
+         SD = sd(NumericValue),
+         Min = min(NumericValue),
+         Max = max(NumericValue)
+     )
> print(summary_stats)
# A tibble: 1 × 5
   Mean Median    SD   Min   Max
  <dbl>  <int> <dbl> <int> <int>
1  58.3     57  30.4     9    99
>
> # Guardar las medidas de resumen estadístico en un archivo CSV
> write.csv(summary_stats, 'summary_stats.csv', row.names = FALSE)
>
```

```
> # Guardar el primer párrafo en un archivo de texto
> writeLines(first_paragraph_clean, 'primer_parrafo.txt')
>
```