

# Web scraping

M.G

This is a simple example of web scraping using HTML and the rvest and broom library. The website that I have used is: <https://www.managementtoday.co.uk/top100entrepreneurs-2015-list>

```
library(rvest)
```

```
## Loading required package: xml2
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

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

```
##
```

```
## filter, lag
```

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

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
library(broom)
```

```
url <- paste0("https://www.managementtoday.co.uk/top100entrepreneurs-2015-list")
```

```
h <- read_html(url)
```

```
#tab holds the html code for only the tables in the page
```

```
tab <- h %>% html_nodes("table")
```

```
#let us convert the first table in a data frame and view it
```

```
tab <- tab[[2]] %>% html_table
```

```
#we can see that the header is taken as a row so we copy it to header
```

```
names(tab) <- tab[1,]
```

```
#remove the first row
```

```
tab = tab[-1,]
```

Now we have the data frame ready to use and we can see that it needs some data wrangling but we will leave that for later.