

Lab 5

Janice Chong See Wai

2022-04-14

Web Scraping

```
library(rvest)

# Reading the HTML code from the website
webpage = read_html("https://en.wikipedia.org/wiki/Data_science")

# Using CSS selectors to scrape the heading section
heading = html_node(webpage, '#siteSub , #firstHeading')

# Converting the heading data to text
text = html_text(heading)
print(text)
```

```
## [1] "Data science"
```

```
# Using CSS selectors to scrape
# all the paragraph section
# Note that we use html_nodes() here
paragraph = html_nodes(webpage, 'p:nth-child(10) , .nowraplinks+ p')

# Converting the heading data to text
pText = html_text(paragraph)

# Print the top 6 data
print(head(pText))
```

```
## [1] "Data science is an interdisciplinary field that uses scientific methods, processes, algorithms a
## [2] "A data scientist is someone who creates programming code and combines it with statistical knowl
```