Web scraping exercises from DataCamp course

# Intrduction to HTML and CSS

1. The corresponding HTML is provided to you in the html\_excerpt\_raw variable.

- In this exercise you’ll use rvest to read it into R, so you can work with it.

- Use the xml\_structure() function to get a better overview of the tag hierarchy of the HTML excerpt.

html\_excerpt\_raw <- ’

<html>

<body>

<h1>Web scraping is cool</h1>

<p>It involves writing code – be it R or Python.</p>

<p><a href=“https://datacamp.com”>DataCamp</a>

has courses on it.\</p\>

</body>

</html>’

# Turn the raw excerpt into an HTML document R understands

html\_excerpt <- read\_html(html\_excerpt\_raw)

html\_excerpt

# Print the HTML excerpt with the xml\_structure() function