

Understanding the structure of web pages

Web scraping for social scientists

What is a web page?

It is a document which can be displayed in a web browser (e.g., Firefox, Safari etc).

A **website** is a collection of web pages that are connected in various ways.

A **web server** is a computer that hosts/stores a website on the Internet.

A **URL** (uniform resource locator) is the location of a web page on the internet.

([Mozilla, 2021](#))

How are web pages structured?

Web pages are written in a language called **Hyper Text Markup Language (HTML)**.

HTML describes the *nested* structure of a web page.

HTML consists of a series of elements, which are distinguished using tags.

HTML elements tell the browser how to display the content (e.g., fonts, colours, sections).

Exercise

Mary's Meals

<https://www.marysmeals.org/what-we-do/our-impact>

Using the six steps from earlier, write a solution for scraping information about how many children are fed every day by this charity.