



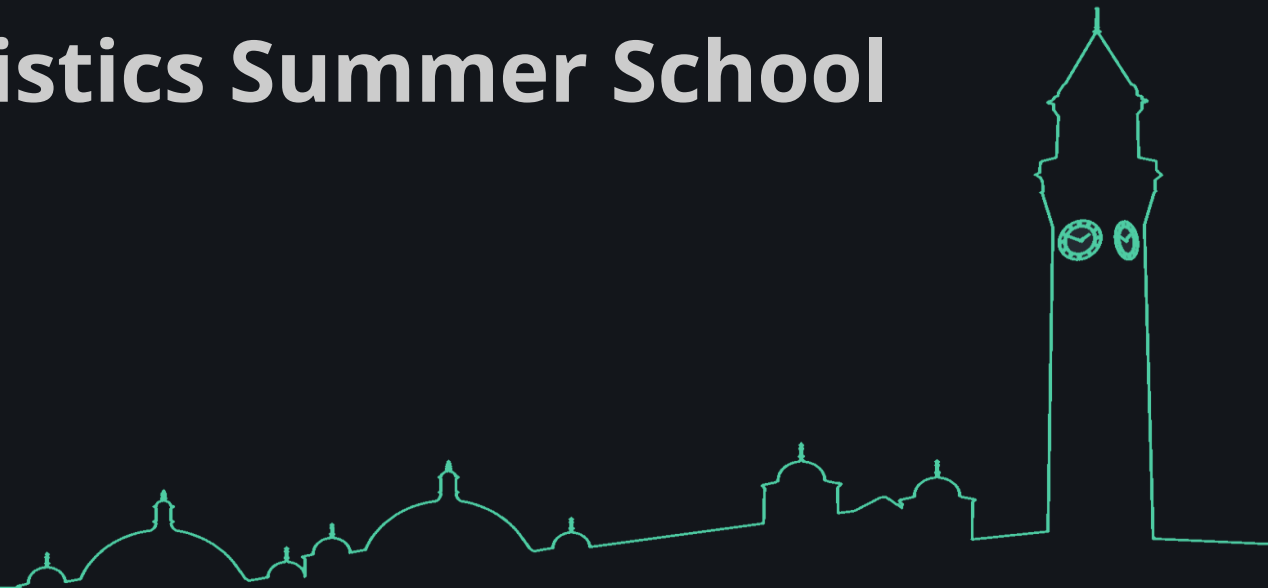
UNIVERSITY OF
BIRMINGHAM

Webscraping with R

Part 2: Intro to HTML & CSS

Birmingham Corpus Linguistics Summer School

Jason Grafmiller





WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Current events](#)
[Random article](#)
[About Wikipedia](#)
[Contact us](#)
[Donate](#)

[Contribute](#)

[Help](#)
[Learn to edit](#)
[Community portal](#)
[Recent changes](#)
[Upload file](#)

[Tools](#)

[What links here](#)
[Related changes](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Cite this page](#)
[Wikidata item](#)

[Print/export](#)

[Download as PDF](#)
[Printable version](#)

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#)



Parasitic gap

From Wikipedia, the free encyclopedia

In [generative grammar](#), a **parasitic gap** is a construction in which one "gap" appears to be dependent on another "gap", that is, the one gap can appear only by virtue of the appearance of the other gap, hence the former is said to be "parasitic" on the latter, e.g. *Which explanation did you reject __₁ without first really considering __₂?* While parasitic gaps are present in English and some related Germanic languages, e.g. Swedish (see [Engdahl 1983](#)), their appearance is much more restricted in other, closely related languages, e.g. German and the Romance languages.^[1] Japanese linguistic scholar Fumikazu Niinuma has attempted to differentiate between parasitic gaps and [coordination](#) in his research, as he believes the two are often confused.^[2] An aspect of parasitic gaps that makes them particularly mysterious is the fact they usually appear inside [islands to extraction](#). Although the study of parasitic gaps began in the late 1970s, no consensus has yet been reached about the best analysis.^[3]

Contents [hide]

- [1 The phenomenon](#)
- [2 Some historical notes](#)
- [3 Some traits of parasitic gaps](#)
 - [3.1 Optional parasitic gaps](#)
 - [3.2 Obligatory parasitic gaps](#)
 - [3.3 Missing object constructions](#)
 - [3.4 The role of parallelism](#)
- [4 Theoretical controversy](#)
- [5 See also](#)
- [6 Notes](#)
- [7 References](#)

The phenomenon [[edit](#)]

Optional parasitic gaps [\[edit \]](#)

Many parasitic gaps appear optionally. They are in non-complementary distribution with a pronoun, meaning that the speaker has the choice whether to employ the gap or not, e.g.

- a. **Which dish** did you order ___ after you tried ____p? – Parasitic gap present
- b. **Which dish** did you order ___ after you tried **it**? – Parasitic gap absent

- a. **Which movie** will they like ___ as soon as they see ____p? – Parasitic gap present
- b. **Which movie** will they like ___ as soon as they see **it**? – Parasitic gap absent

The a-sentence contain typical parasitic gaps, whereas the b-sentence choose to use a pronoun instead of the gap. In other words, the parasitic gap in the a-sentences is occurring optionally. Optionality like this suggests an analysis of parasitism in terms of [ellipsis](#), since optionality is the primary trait of known ellipsis mechanisms.

Obligatory parasitic gaps [\[edit \]](#)

While many parasitic gaps occur optionally as just illustrated, other parasitic gaps occur obligatorily. This can be the case when the parasitic gap precedes the "real" gap, e.g.

- a. The rumor about **the girl** annoyed **her**. – No gap at all present
- b. **Which girl** did [the rumor about ____p] annoy ___? – Parasitic gap precedes "real" gap
- c. ??**Which girl** did the rumor about **her** annoy ___? – Real gap hardly possible unless the parasitic gap is present; weak crossover has occurred

- a. If you get to know **him**, you will like **Bill**. – No gap at all present
- b. Bill is the type of guy **who** [if you get to know ____p], you will like ___. – Parasitic gap precedes "real" gap
- c. ??Bill is the type of guy **who** if you get to know **him**, you will like ___. – Real gap hardly possible unless the parasitic gap is present; weak crossover has occurred

These examples illustrate a couple of important facts about parasitic gaps. The b-sentences demonstrate that the parasitic gap can indeed precede the "real" gap, and the strong marginality of the c-sentences shows that in a sense, the real gap can also be dependent on the parasitic gap. Note that we know that the first gap (the leftmost gap) in the b-sentences is parasitic on the following gap because it, i.e. the leftmost gap, appears inside what is normally an extraction island (marked with square brackets). The aspect of parasitic gaps illustrated with

line wrap ☐

```

1 <!DOCTYPE html>
2 <html class="client-nojs" lang="en" dir="ltr">
3 <head>
4 <meta charset="UTF-8"/>
5 <title>Parasitic gap - Wikipedia</title>
6 <script>document.documentElement.className="client-js";RLCONF={"wgBreakFrames":!1,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],
7 "wgRestrictionMove":[],"wgFlaggedRevsParams":{"tags":{"status":{"levels":-1}}},"wgMediaViewerOnClick":!0,"wgMediaViewerEnabledByDefault":!0,"wgPopupsF
8 "ext.visualEditor.desktopArticleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","wikibase.client.init":"ready"}
9 <script>(RLQ=window.RLQ||[]).push(function(){mw.loader.implement("user.options@1hzgi",function($,jQuery,require,module){/*@nomin*/mw.user.tokens.set({
10 });});</script>
11 <link rel="stylesheet" href="/w/load.php?lang=en&modules=ext.cite.styles%7Cext.uls.interlanguage%7Cext.visualEditor.desktopArticleTarget.noscript%
12 <script async="" src="/w/load.php?lang=en&modules=startup&only=scripts&raw=1&skin=vector"></script>
13 <meta name="ResourceLoaderDynamicStyles" content="">
14 <link rel="stylesheet" href="/w/load.php?lang=en&modules=site.styles&only=styles&skin=vector"/>
15 <meta name="generator" content="MediaWiki 1.37.0-wmf.4"/>
16 <meta name="referrer" content="origin"/>
17 <meta name="referrer" content="origin-when-crossorigin"/>
18 <meta name="referrer" content="origin-when-cross-origin"/>
19 <meta property="og:title" content="Parasitic gap - Wikipedia"/>
20 <meta property="og:type" content="website"/>
21 <link rel="alternate" media="only screen and (max-width: 720px)" href="//en.m.wikipedia.org/wiki/Parasitic_gap"/>
22 <link rel="alternate" type="application/x-wiki" title="Edit this page" href="/w/index.php?title=Parasitic_gap&action=edit"/>
23 <link rel="edit" title="Edit this page" href="/w/index.php?title=Parasitic_gap&action=edit"/>
24 <link rel="apple-touch-icon" href="/static/apple-touch/wikipedia.png"/>
25 <link rel="shortcut icon" href="/static/favicon/wikipedia.ico"/>
26 <link rel="search" type="application/opensearchdescription+xml" href="/w/opensearch_desc.php" title="Wikipedia (en)"/>
27 <link rel="EditURI" type="application/rsd+xml" href="//en.wikipedia.org/w/api.php?action=rsd"/>
28 <link rel="license" href="//creativecommons.org/licenses/by-sa/3.0/">
29 <link rel="canonical" href="https://en.wikipedia.org/wiki/Parasitic_gap"/>
30 <link rel="dns-prefetch" href="//login.wikimedia.org"/>
31 <link rel="dns-prefetch" href="//meta.wikimedia.org" />
32 </head>
33 <body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject mw-editable page-Parasitic_gap rootpage-Parasitic_gap skin-vector action-view
34 <div id="mw-head-base" class="noprint"></div>
35 <div id="content" class="mw-body" role="main">
36 <a id="top"></a>
37 <div id="siteNotice" class="mw-body-content"><!-- CentralNotice --></div>
38 <div class="mw-indicators mw-body-content">
39 </div>
40 <h1 id="firstHeading" class="firstHeading">Parasitic gap</h1>
41 <div id="bodyContent" class="mw-body-content">
42 <div id="siteSub" class="noprint">From Wikipedia, the free encyclopedia</div>
43 <div id="contentSub"></div>
44 <div id="contentSub2"></div>
45

```

Webpages behind the scenes

All web pages have an expected kind of structure, which usually consists of three types of code

- **Markup**: provides annotations, i.e. “markup”, describing the structure and organization of the information contained in a document (e.g. HTML, XML)

Webpages behind the scenes

All web pages have an expected kind of structure, which usually consists of three types of code

- **Markup**: provides annotations, i.e. “markup”, describing the structure and organization of the information contained in a document
- **CSS**: controls the presentation, formatting, and layout of the page that you see in your browser

Webpages behind the scenes

All web pages have an expected kind of structure, which usually consists of three types of code

- **Markup**: provides annotations, i.e. “markup”, describing the structure and organization of the information contained in a document
- **CSS**: controls the presentation, formatting, and layout of the page that you see in your browser
- **JavaScript**: controls how things behave when you interact with them

**Content &
structure**

**Components of a
web page**

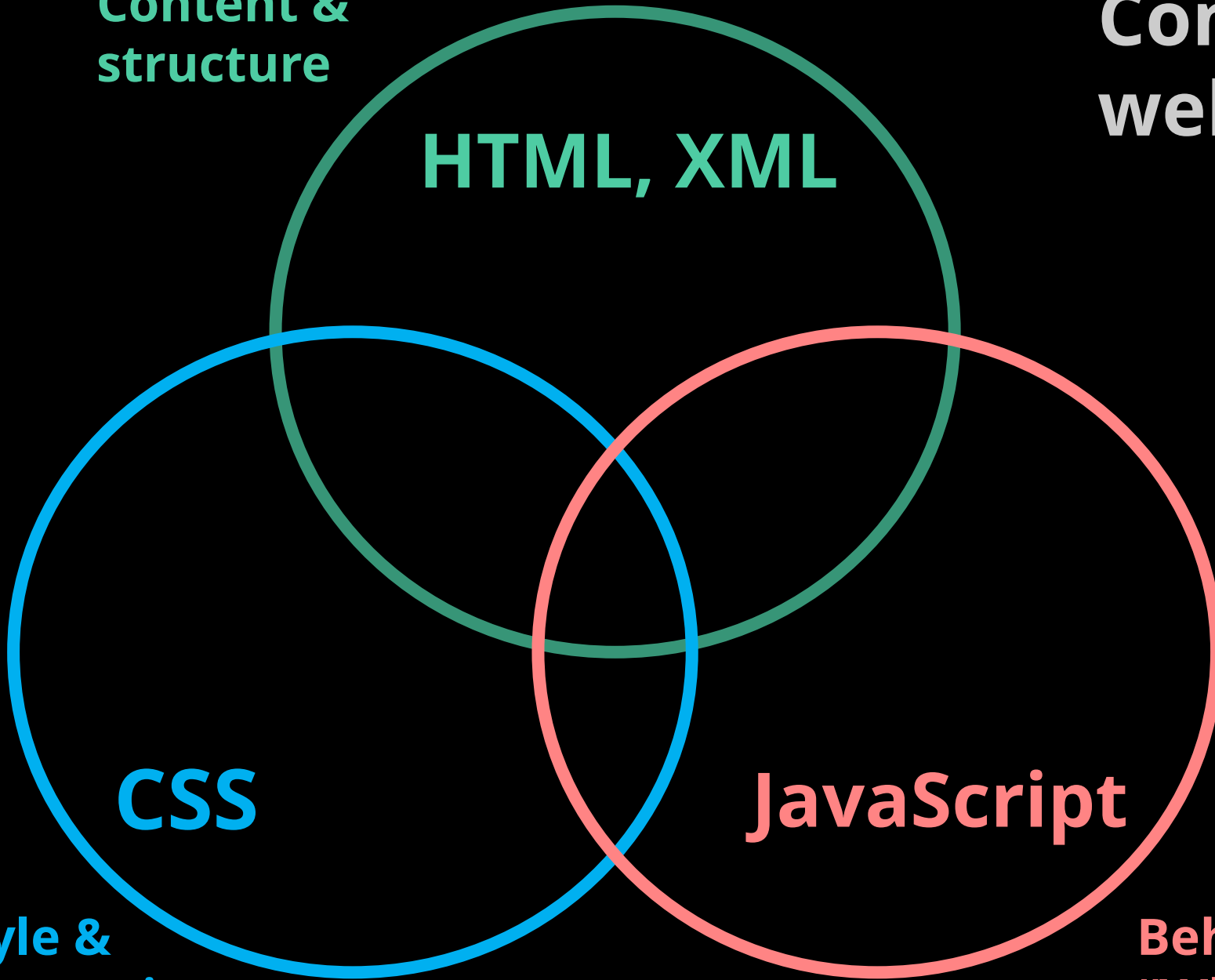
HTML, XML

CSS

JavaScript

**Style &
Formatting**

**Behaviour & Functionality
"What it does"**



Markup code: HTML and XML

HTML (HyperText Markup Language) is designed to display data, with focus on how data *looks*

XML (eXtensible Markup Language) is designed to organize and store data, with a focus on what data *is*

Markup code: HTML and XML

```
<html>
```

```
<body>
```

```
<p>
```

The 'p' tag is used to indicate a paragraph.

Other general purpose tags include `<div>...</div>`, and
`...`

for creating divisions and text span elements.

```
</p>
```

```
</body>
```

```
</html>
```

Markup code: HTML and XML

```
<html>
```

```
<body>
```

```
<p>
```

opening tags

The 'p' tag is used to indicate a paragraph.

Other general purpose tags include `<div>...</div>`, and `...`

for creating divisions and text span elements.

```
</p>
```

```
</body>
```

```
</html>
```

closing tags

Markup code: Lists

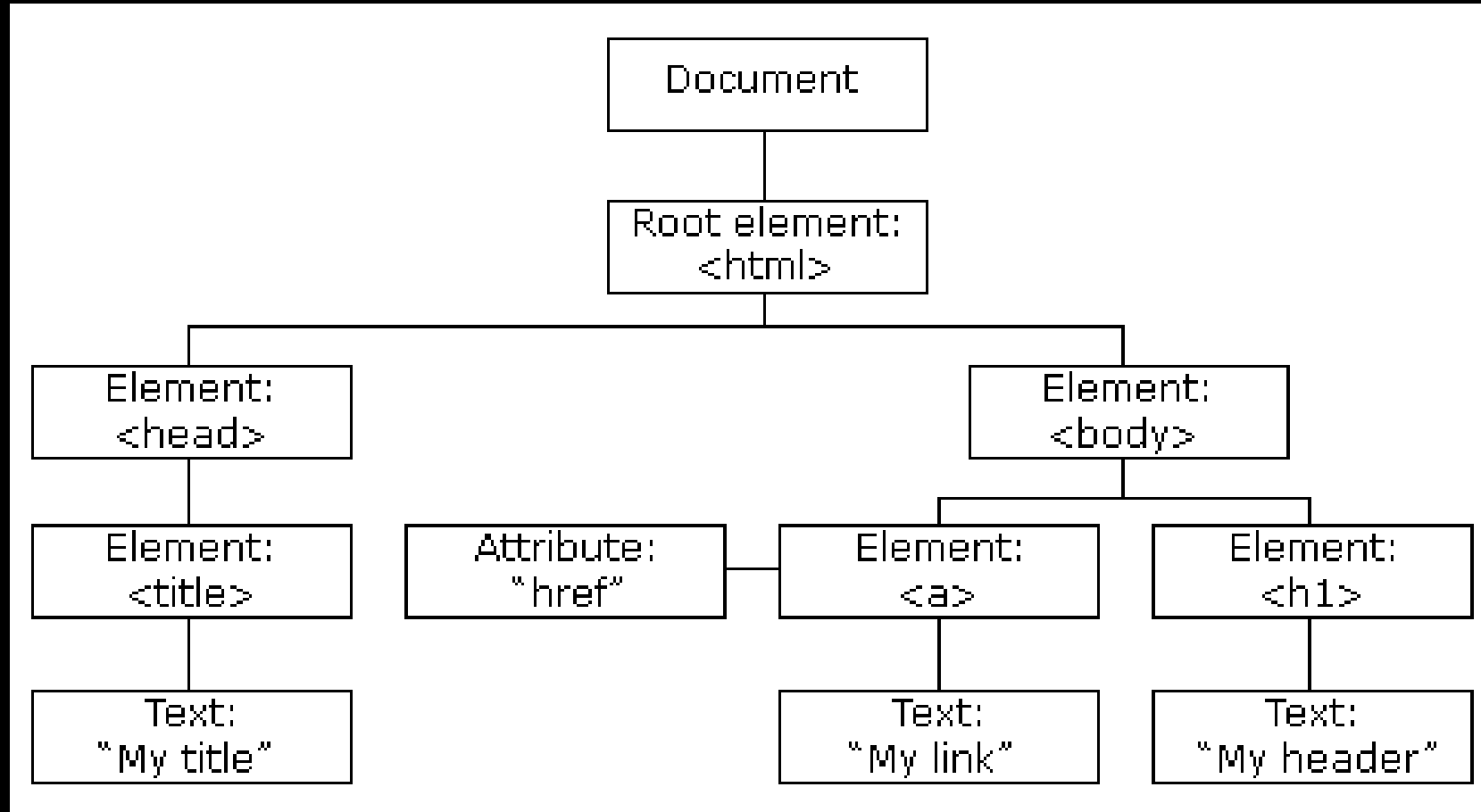
```
<ul>
```

```
<li>The 'ul' tag indicates an unordered  
list.</li>
```

```
<li>The 'li' tag demarcates an individual list  
item.</li>
```

```
</ul>
```

Markup is *structured*



Common tags

Tag	Description
<html> ... </html>	Declares the Web page to be written in HTML
<head> ... </head>	Delimits the page's head
<title> ... </title>	Defines the title (not displayed on the page)
<body> ... </body>	Delimits the page's body
<h <i>n</i> > ... </h <i>n</i> >	Delimits a level <i>n</i> heading
 ... 	Set ... in boldface
<i> ... </i>	Set ... in italics
<center> ... </center>	Center ... on the page horizontally
 ... 	Brackets an unordered (bulleted) list
 ... 	Brackets a numbered list
 ... 	Brackets an item in an ordered or numbered list
 	Forces a line break here
<p>	Starts a paragraph
<hr>	Inserts a horizontal rule
	Displays an image here
 ... 	Defines a hyperlink

Attributes

```
<p id="paragraph1">
```

This is a unique paragraph which can be referred to with the id "#paragraph1".

```
</p>
```

```
<p class="red-right-hand">
```

This paragraph is one of many potential elements of class ".red-right-hand".

```
</p>
```

Markup code: Links

```
<a href="url">link text</a>
```

```
<a href="https://www.birmingham.ac.uk">  
University of Birmingham main page  
</a>
```


Cascading Style Sheets (CSS)

```
p {  
    text-align: center;  
}
```

CSS code declaring all <p> elements to be center-aligned

```
li {  
    font-style: italic;  
}
```

CSS code declaring all (list item) elements to be italicised

Cascading Style Sheets (CSS)

```
.red-right-hand {  
    text-align: right;  
    color: red;  
}
```

CSS code declaring all elements of class="red-right-hand" to be right-aligned and red

```
#paragraph1 {  
    font-weight: bold;  
    font-size: 16pt;  
}
```

CSS code declaring the *specific element* with id="paragraph" to be bold and size 16pt font

CSS selectors

Selectors identify HTML element(s) according to their attributes, tags, and structure.

Selectors can be very complex

CSS selectors

`p.red-right-hand`

Selects only `<p>` elements with `class="red-right-hand"`

`div > p`

Selects only `<p>` elements that are children of a `<div>` element

`p.red-right-hand li`

Selects only `` elements that are descendants of a `<p>` element with `class="red-right-hand"`

Further practice

I highly recommend the tutorials on [w3schools.com](https://www.w3schools.com).

Just the basics should be enough for what we'll be doing.