

















SUPPLEMENT: TAG INVESTIGATION

SEARCHING FOR TAGS

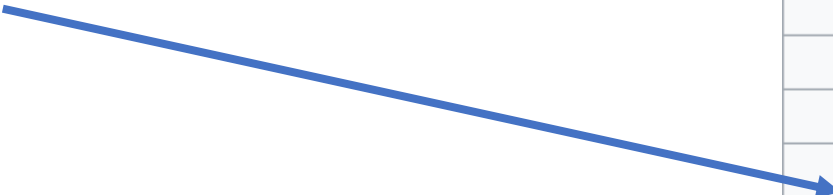


- In lab2, this website is to be scraped:

Country ^[17] ⇅	Fossil CO ₂ Emissions (Mt CO ₂ /yr)			Fossil CO ₂ Emissions		2017 – Fossil CO ₂ Emissions	
	1990 ⇅	2005 ⇅	2017 ⇅	2017 (% of world) ⇅	2017 vs 1990: change (%) ⇅	Per Land Area (t CO ₂ /km ² /yr) ⇅	Per Capita (t CO ₂ /cap/yr) ⇅
World	22,674.116	30,049.809	37,077.404	100.00%	63.5%	73	4.9
World – International Aviation	258.941	422.777	543.381	1.47%	109.8%	n/a	n/a
World – International Shipping	371.804	572.169	677.248	1.83%	82.2%	n/a	n/a
 Afghanistan	2.546	1.063	11.422	0.03%	348.6%	18	0.3
 Albania	6.583	4.196	5.026	0.01%	-23.7%	175	1.7
 Algeria	65.677	98.197	159.929	0.43%	143.5%	67	3.9
 Angola	5.851	15.975	30.876	0.08%	427.7%	25	1.0
 Anguilla	0.006	0.014	0.028	0.00%	366.7%	308	1.9
 Antigua and Barbuda	0.223	0.283	0.624	0.00%	179.8%	1,412	6.1
 Argentina	112.434	165.429	209.968	0.57%	86.7%	76	4.7
 Armenia	20.699	4.542	4.832	0.01%	-76.7%	162	1.6
 Aruba	0.297	0.470	0.959	0.00%	222.9%	5,328	9.1
 Australia	275.408	391.590	402.253	1.08%	46.1%	52	16.5
 Austria	62.918	80.994	72.249	0.19%	14.8%	861	8.3
 Azerbaijan	58.077	30.485	32.544	0.09%	-44.0%	376	3.3
 Bahamas	1.524	2.068	2.997	0.01%	96.7%	215	7.6
 Bahrain	11.988	23.388	35.775	0.10%	198.4%	46,643	24.0
 Bangladesh	13.868	38.834	84.546	0.23%	509.6%	573	0.5
 Barbados	0.776	2.162	3.172	0.01%	308.8%	7,377	11.1

Finding Specific Data









- Lab 2 calls for scraping the 2017 (% of world) column.
- This Wikipedia web page is very long. How can you find the data where the data is stored?



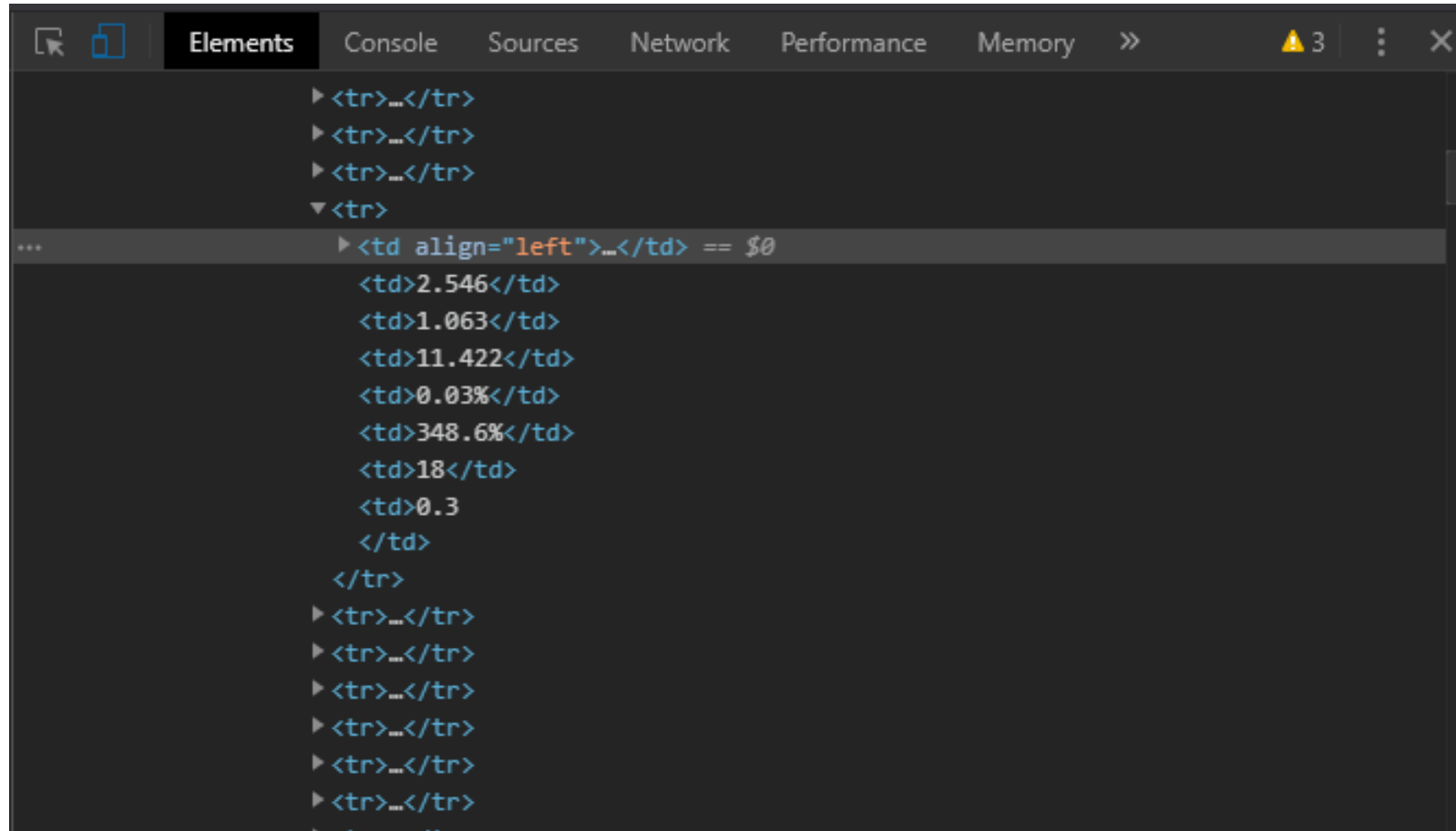
2017 (% of world) ↕
100.00%
1.47%
1.83%
0.03%
0.01%
0.43%
0.08%
0.00%
0.00%
0.57%
0.01%
0.00%
1.08%
0.19%
0.09%

Inspecting Website Elements

- Web browsers have the ability inspect webpage elements.
- First scroll to the webpage data you want to inspect.
- Then right-click on the element.
- This pop-up menu appears. Click 'Inspect'

	Fossil CO ₂ Emissions (Mt CO ₂ /yr)			Fossil CO ₂ Emissions	
				1990 vs 1990:	Change (%)
World					63.5%
World – In					109.8%
World – In					82.2%
 Afghanistan					348.6%
 Albania					-23.7%
 Algeria					143.5%
 Angola					427.7%
 Anguilla					366.7%
 Antigua and Barbuda					179.8%
 Argentina	112.434	165.429	209.968	0.57%	86.7%
 Armenia	20.699	4.542	4.832	0.01%	-76.7%

Inspect Window



Tags Used

- When one of the data columns of the countries is clicked, the tag data for it is revealed.

```
<td align="left"><span class="flagicon"><noscript></noscript>&nbsp;</span><a  
href="/wiki/Afghanistan" title="Afghanistan">Afghanistan</a></td>  
<td>2.546</td>  
<td>1.063</td>  
<td>11.422</td>  
<td>0.03%</td>  
<td>348.6%</td>  
<td>18</td>  
<td>0.3  
</td>
```

Tags to Scrape

```
href="/wiki/Afghanistan" title="Afghanistan">Afghanistan</a></td>
```

Tag: href.title.text = Country Name “Afghanistan”

These tags may not necessarily have the specified tags. So additional scraping using Regex may be required to isolate the Country Name.

Tags to Scrape

```
<td>2.546</td>
```

```
<td>1.063</td>
```

```
<td>11.422</td>
```

```
<td>0.03%</td>
```

```
<td>348.6%</td>
```

```
<td>18</td>
```

```
<td>0.3</td>
```

```
<td>0.03%</td>
```

The `<td>` tag with 0.03% corresponds to the 2017 column.

Note

- In the previous two pages, the tags `<td>` and `<href>` were scraped using the Webpage inspect utility.
- The inspect utility shows which tags are available to be scraped.
- Using BeautifulSoup in combination with Regex should allow you to isolate and scrape the data from the tags.
- The tags you use may not necessarily be `<td>` or `<href>`. These were used for example purposed only.