

25 April 2024

Parsing Datable from HTML

```
[4]: # Finds all tables and saves them to a variable
all_tables = soup.find_all('table')
# Find the number of tables
print("Total number of tables: {}".format(len(all_tables)))

Total number of tables: 3

In [5]: # Finds header table
headers = soup.find("table", {"id": "fixedTable"})
print(type(headers))

<class 'bs4.element.Tag'>

In [6]: # Finds all tr tags within the header table
levels = headers.find_all('tr')
levels

Out[6]: <tr class="level0"><th class="first"></th><th class="separator" colspan="2"><div class="spacer">a data-text="Metadata:Forest area" href="javascript:void(0)" onclick="loadWIMetaData('AG.LND.FRST.K2','S','Series','Forest area','Forest area@Mammal species, threatened@Bird species, threatened@Fish species, threatened@Plant species (higher), threatened@Terrestrial protected areas@Marine protected areas@','AG.LND.FRST.K2@EN.MAM.THRO.NO@ER.BIR.THRD.NO@ER.FSH.THRD.NO@ER.HPT.THRD.NO@ER.LND.PTLD.ZS@ER.MRN.PTMR.ZS@')">Forest area</div></th><th class="separator" colspan="4"><div class="spacer">Threatened species</div></th><th class="separator" colspan="1"><div class="spacer">a data-text="Metadata:Terrestrial protected areas" href="javascript:void(0)" onclick="loadWIMetaData('ER.LND.PTLD.ZS','S','Serie s','Terrestrial protected areas','Forest area@Mammal species, threatened@Bird species, threatened@Fish species (higher), threatened@Terrestrial protected areas@Marine protected areas@s">s">">a data-text="Metadata:Marine protected areas" href="javascript:void(0)" onclick="loadWIMetaData('ER.MRN.PTMR.ZS','S','S','Series','Marine protected areas','Forest area@Mammal species, threaten ed@Bird species, threatened@Fish species, threatened@Plant species (higher), threatened@Terrestrial protected areas@Marine protected areas@','AG.LND.FRST.K2@EN.MAM.THRO.NO@ER.BIR.THRD.NO@ER.FSH.THRD.NO@E N.HPT.THRD.NO@ER.LND.PTLD.ZS@ER.MRN.PTMR.ZS@')">Marine protected areas</a></div></th></tr>,
<tr class="level1"><th class="first"></th><th class="separator" colspan="2"><div class="spacer"></div></th><th class="" colspan="1"><div class="spacer">a data-text="Metadata:Mammals" href="javascript: void(0)" onclick="loadWIMMetaData('EN.MAM.THRO.NO','S','Series','Mammal species, threatened','Forest area@Mammal species, threatened@Bird species, threatened@Fish species, threatened@Plant species (high er), threatened@Terrestrial protected areas@Marine protected areas@','AG.LND.FRST.K2@EN.MAM.THRO.NO@ER.BIR.THRD.NO@ER.FSH.THRD.NO@ER.HPT.THRD.NO@ER.LND.PTLD.ZS@ER.MRN.PTMR.ZS@')">Mammals</a></div></th><t h class"" colspan="1"><div class="spacer">a data-text="Metadata:Birds" href="javascript:void(0)" onclick="loadWIMetaData('EN.BIR.THRD.NO','S','Series','Bird species, threatened','Forest area@Mammal sp ecies, threatened@Bird species, threatened@Fish species, threatened@Plant species (higher), threatened@Terrestrial protected areas@Marine protected areas@','AG.LND.FRST.K2@EN.MAM.THRO.NO@ER.BIR.THRD.NO@ E.N.FSH.THRD.NO@ER.HPT.THRD.NO@ER.LND.PTLD.ZS@ER.MRN.PTMR.ZS@')">Birds</a></div></th><th class="" colspan="1"><div class="spacer">a data-text="Metadata:Fishes" href="javascript:void(0)" onclick="loadWIMEt aData('EN.FSH.THRD.NO','S','Series','Fish species, threatened','Forest area@Mammal species, threatened@Bird species, threatened@Fish species, threatened@Plant species (higher), threatened@Terrest rial protected areas@Marine protected areas@','AG.LND.FRST.K2@EN.MAM.THRO.NO@ER.BIR.THRD.NO@E N.FSH.THRD.NO@ER.HPT.THRD.NO@ER.LND.PTLD.ZS@ER.MRN.PTMR.ZS@')">Fishes</a></div></th><th class="separator" colspan="1"><div class="space r">Higher plants</a></div></th><th class="separator" colspan="1"><div class="spacer"></div></th><th class="separator" colspan="1"><div class="s pacer"></div></th></tr>,
<tr class="level2"><th class="first"></th><th class="separator" colspan="2"><div class="spacer">$q km thousands</div></th><th class="" colspan="1"><div class="spacer"></div></th><th class="" colspn ="1"><div class="spacer"></div></th><th class="" colspan="1"><div class="spacer"></div></th><th class="separator" colspan="1"><div class="spacer"></div></th><th class="separator" colspan="1"><div cla ss="spacer">% of total land area</div></th><th class="separator" colspan="1"><div class="spacer">% of territorial waters</div></th></tr>,
<tr class="level3"><th class="first"></th><th><div class="spacer">1990</div></th><th class="separator"><div class="spacer">2021</div></th><th><div class="spacer">2018</div></th><th><div class="space r">2018</div></th><th><div class="spacer">2018</div></th><th class="separator"><div class="spacer">2022</div></th><th class="separator"><div class="spacer">2022</div></th></tr>
```

```
In [7]: header_levels2 = [[div.get_text().strip() for div in tr.find_all('div')] for tr in levels]
header_levels2
```

```
Out[7]: [['Forest area',
'Threatened species',
'Terrestrial protected areas',
'Marine protected areas'],
[['Mammals', 'Birds', 'Fishes', 'Higher plants', '', ''],
['$q. km thousands',
'',
'',
'',
'% of total land area',
'% of territorial waters'],
['1990', '2021', '2018', '2018', '2018', '2018', '2022', '2022']]
```

```
In [8]: # Finds datatable that contains needed data
data_table = soup.find("table", {"id": "scrollTable"})
print(type(data_table))

<class 'bs4.element.Tag'>

In [9]: # Finds all tr data in data table
row = data_table.find_all('tr')
row
```

[illegible]

[illegible]

[illegible]

```


```

[illegible]

[illegible]

```
In [10]: # Finds text from each row of the data table
data_rows1 = [[td.get_text().strip() for td in tr.findAll('td')] for tr in row]
data_rows1
```



```
Out[10]: [['Afghanistan', '12', '12', '11', '16', '4', '5', '3.6', '...'],
['Albania', '8', '8', '3', '8', '44', '4', '18.6', '2.8'],
['Algeria', '17', '20', '14', '15', '41', '22', '4.6', '0.1'],
['American Samoa', '0', '0', '1', '8', '12', '1', '15.8', '8.7'],
['Andorra', '0', '0', '2', '3', '0', '0', '26.9', '...'],
['Angola', '793', '661', '18', '32', '53', '34', '7.0', '0.0'],
['Antigua and Barbuda', '0', '0', '2', '3', '31', '4', '19.9', '0.3'],
['Argentina', '352', '285', '38', '52', '42', '70', '8.7', '11.8'],
['Armenia', '3', '3', '9', '14', '3', '74', '24.7', '...'],
['Aruba', '0', '0', '2', '2', '24', '2', '18.9', '0.0'],
['Australia', '1,339', '1,340', '63', '52', '125', '108', '20.4', '44.3'],
['Austria', '38', '39', '3', '13', '11', '17', '29.5', '...'],
['Azerbaijan', '10', '11', '8', '17', '14', '44', '10.2', '0.4'],
['Bahamas, The', '5', '5', '5', '10', '43', '7', '36.6', '7.9'],
['Bahrain', '0', '0', '3', '7', '14', '0', '13.0', '21.1'],
['Bangladesh', '19', '19', '37', '36', '29', '23', '4.6', '5.4'],
['Barbados', '0', '0', '3', '4', '29', '3', '1.3', '0.0'],
['Belarus', '88', '4', '9', '2', '1', '13.8', '...'],
['Belgium', '7', '2', '8', '13', '0', '15.6', '38.1'],
['Belize', '16', '13', '10', '6', '45', '46', '37.5', '11.0'],
['Benin', '48', '31', '13', '12', '40', '20', '29.6', '0.0'],
['Bermuda', '0', '0', '4', '3', '27', '8', '2.1', '0.0'],
['Bhutan', '25', '27', '25', '21', '3', '43', '49.7', '...'],
['Bolivia', '578', '506', '21', '55', '8', '106', '30.9', '...'],
['Bosnia and Herzegovina', '22', '22', '4', '7', '36', '3', '4.1', '0.0'],
['Botswana', '188', '151', '11', '16', '2', '3', '29.1', '...'],
['Brazil', '5,889', '4,954', '80', '175', '93', '558', '30.3', '26.8'],
['Brunei Darussalam', '4', '4', '33', '31', '14', '127', '46.9', '0.2'],
['Bulgaria', '33', '39', '8', '17', '22', '9', '41.0', '8.1'],
['Burkina Faso', '77', '62', '9', '12', '4', '4', '16.4', '...'],
['Burundi', '3', '3', '14', '15', '17', '89', '7.6', '...'],
['Cabo Verde', '0', '0', '4', '7', '35', '51', '2.9', '0.0'],
['Cambodia', '110', '79', '39', '31', '48', '37', '39.7', '1.4'],
['Cameroon', '225', '203', '46', '29', '122', '555', '11.0', '10.9'],
['Canada', '3,483', '3,469', '18', '24', '44', '18', '12.7', '9.1'],
['Cayman Islands', '0', '0', '1', '3', '31', '22', '10.8', '0.1'],
['Central African Republic',
'232',
'223',
'16',
'16',
'3',
'26',
'18.1',
'...'],
['Chad', '67', '42', '16', '16', '1', '6', '21.0', '...'],
['Channel Islands', '0', '0', '...', '...', '...', '...', '...', '...'],
['Chile', '152', '183', '19', '35', '27', '73', '21.0', '41.4'],
['China', '1,571', '2,219', '73', '96', '136', '593', '15.6', '5.5'],
['Hong Kong SAR, China', '...', '...', '4', '21', '15', '10', '41.9', '0.0'],
['Macao SAR, China', '...', '...', '0', '4', '9', '0', '...', '...'],
['Colombia', '650', '589', '58', '126', '99', '268', '16.4', '17.1'],
['Comoros', '0', '0', '5', '14', '11', '9', '33.8', '0.4'],
['Congo, Dem. Rep.',
'1,506',
'1,251',
'32',
'42',
'94',
'148',
'13.9',
'0.2'],
['Congo, Rep.', '223', '219', '15', '7', '57', '46', '36.8', '3.0'],
['Costa Rica', '29', '31', '11', '27', '64', '143', '26.6', '29.0'],
['Cote d'Ivoire', '79', '27', '31', '25', '59', '118', '23.0', '0.1'],
['Croatia', '19', '19', '9', '14', '64', '8', '38.5', '9.0'],
['Cuba', '21', '32', '10', '19', '44', '179', '16.2', '3.9'],
['Curacao', '...', '0', '3', '2', '28', '2', '15.7', '0.0'],
['Cyprus', '2', '2', '6', '7', '24', '22', '38.7', '8.6'],
['Czechia', '26', '27', '3', '9', '2', '26', '22.2', '...'],
['Denmark', '5', '6', '2', '9', '18', '1', '17.0', '18.3'],
['Djibouti', '0', '0', '11', '12', '23', '3', '1.6', '0.2'],
['Dominica', '1', '0', '3', '7', '31', '12', '22.0', '0.0'],
['Dominican Republic', '16', '22', '7', '17', '29', '47', '26.4', '17.3'],
['Ecuador', '146', '124', '47', '106', '61', '1,859', '23.4', '19.1'],
['Egypt, Arab Rep.', '0', '0', '18', '14', '58', '8', '13.1', '5.0'],
['El Salvador', '7', '6', '6', '7', '17', '33', '8.6', '0.7'],
['Equatorial Guinea', '27', '24', '23', '6', '44', '98', '19.3', '0.2'],
['Eritrea', '11', '11', '15', '21', '29', '6', '0.0', '0.0'],
['Estonia', '22', '24', '1', '9', '5', '0', '21.4', '18.8'],
['Eswatini', '5', '5', '9', '13', '5', '13', '4.3', '...'],
['Ethiopia', '204', '170', '34', '35', '14', '64', '17.0', '...'],
['Faroe Islands', '0', '0', '4', '6', '13', '0', '2.3', '0.0'],
['Fiji', '9', '11', '7', '14', '20', '78', '5.4', '0.9'],
['Finland', '219', '224', '2', '11', '6', '2', '13.4', '12.0'],
['France', '144', '173', '9', '16', '53', '47', '28.4', '49.8'],
['French Polynesia', '1', '1', '0', '35', '28', '48', '2.0', '0.0'],
['Gabon', '238', '235', '18', '7', '72', '173', '22.4', '28.8'],
['Gambia, The', '4', '2', '10', '14', '35', '6', '7.7', '0.6'],
['Georgia', '28', '28', '9', '14', '11', '63', '11.1', '0.7'],
['Germany', '113', '114', '5', '11', '24', '36', '37.6', '45.5'],
['Ghana', '99', '80', '21', '23', '58', '119', '14.8', '0.1'],
['Greece', '33', '39', '10', '17', '80', '72', '35.2', '4.5'],
['Greenland', '0', '0', '9', '6', '9', '1', '41.1', '4.5'],
['Grenada', '0', '0', '3', '2', '29', '3', '9.5', '0.1'],
['Guam', '0', '0', '3', '14', '14', '5', '4.5', '0.0'],
['Guatemala', '48', '35', '15', '17', '41', '120', '20.1', '0.8'],
['Guinea', '73', '61', '30', '20', '76', '70', '37.6', '0.5'],
['Guinea-Bissau', '22', '20', '14', '12', '41', '6', '27.1', '9.0'],
['Guyana', '186', '184', '11', '16', '33', '30', '8.5', '0.0'],
['Haiti', '4', '3', '4', '17', '30', '93', '8.6', '1.5'],
['Honduras', '70', '63', '8', '14', '45', '134', '23.5', '4.6'],
['Hungary', '18', '21', '3', '13', '9', '43', '22.6', '...'],
['Iceland', '0', '1', '6', '7', '16', '0', '20.8', '0.4'],
['India', '639', '724', '93', '93', '227', '396', '7.5', '0.2'],
['Indonesia', '1,185', '915', '191', '160', '166', '458', '12.2', '3.1'],
['Iran, Islamic Rep.', '91', '108', '18', '28', '47', '7', '8.6', '0.8'],
['Iraq', '8', '8', '13', '17', '19', '2', '1.5', '0.0'],
['Ireland', '5', '8', '5', '9', '28', '3', '14.4', '2.3'],
['Isle of Man', '0', '0', '0', '0', '2', '0', '5.5', '0.0'],
['Israel', '1', '1', '15', '18', '50', '23', '24.5', '0.0'],
['Italy', '76', '96', '8', '17', '52', '102', '21.6', '10.6'],
['Jamaica', '5', '6', '7', '11', '30', '215', '20.0', '0.8'],
['Japan', '250', '249', '29', '49', '77', '50', '29.7', '13.9'],
['Jordan', '1', '1', '13', '14', '20', '8', '4.5', '1.0'],
['Kazakhstan', '32', '35', '16', '27', '14', '14', '10.0', '50.7'],
['Kenya', '39', '36', '30', '44', '74', '243', '12.2', '0.7'],
['Kiribati', '0', '0', '2', '6', '14', '0', '22.4', '11.8'],
['Korea, Dem. People's Rep.',
'69',
'60',
'10',
'29',
'20',
'17',
```

'2.4',
'0.0'],
['Korea, Rep.', '66', '63', '12', '33', '28', '31', '17.0', '2.5'],
['Kosovo', '...', '...', '...', '...', '...', '...', '...'],
['Kuwait', '0', '0', '7', '11', '18', '0', '17.1', '1.4'],
['Kyrgyz Republic', '11', '13', '5', '15', '3', '13', '6.7', '...'],
['Lao PDR', '178', '166', '45', '29', '55', '56', '18.7', '...'],
['Latvia', '32', '34', '1', '11', '6', '0', '18.2', '16.0'],
['Lebanon', '1', '1', '10', '11', '28', '24', '1.9', '0.2'],
['Lesotho', '0', '0', '4', '8', '1', '4', '0.5', '...'],
['Liberia', '85', '76', '24', '14', '64', '53', '4.1', '0.1'],
['Libya', '2', '2', '10', '8', '32', '7', '0.1', '0.6'],
['Liechtenstein', '0', '0', '0', '2', '0', '0', '42.6', '...'],
['Lithuania', '20', '22', '2', '10', '6', '1', '17.1', '25.6'],
['Luxembourg', '...', '1', '0', '3', '1', '0', '55.8', '...'],
['North Macedonia', '9', '10', '6', '13', '13', '4', '15.4', '...'],
['Madagascar', '137', '124', '121', '37', '111', '1,111', '7.5', '0.9'],
['Malawi', '35', '22', '10', '19', '36', '34', '22.9', '...'],
['Malaysia', '206', '191', '71', '63', '87', '727', '13.3', '5.6'],
['Maldives', '0', '0', '2', '0', '24', '0', '2.3', '0.1'],
['Mali', '133', '133', '14', '17', '2', '12', '7.5', '...'],
['Malta', '0', '0', '2', '5', '22', '4', '30.6', '7.4'],
['Marshall Islands', '0', '0', '1', '4', '18', '0', '11.9', '0.3'],
['Mauritania', '5', '3', '18', '19', '45', '0', '0.6', '4.2'],
['Mauritius', '0', '0', '7', '12', '21', '91', '4.7', '0.0'],
['Mexico', '706', '656', '96', '71', '181', '484', '14.6', '21.6'],
['Micronesia, Fed. Sts.', '1', '1', '5', '12', '25', '5', '0.0', '0.0'],
['Moldova', '3', '4', '4', '11', '8', '2', '11.4', '...'],
['Monaco', '0', '0', '3', '0', '15', '1', '10.2', '99.7'],
['Mongolia', '144', '142', '11', '24', '2', '0', '19.8', '...'],
['Montenegro', '...', '8', '6', '13', '32', '3', '13.9', '0.6'],
['Morocco', '55', '58', '18', '18', '54', '52', '2.2', '0.7'],
['Mozambique', '434', '365', '17', '30', '72', '145', '29.5', '2.1'],
['Myanmar', '392', '283', '49', '56', '53', '61', '6.6', '0.5'],
['Namibia', '88', '66', '15', '32', '33', '27', '37.9', '1.7'],
['Nepal', '57', '60', '29', '38', '7', '18', '23.6', '...'],
['Netherlands', '3', '4', '3', '10', '15', '1', '22.5', '26.9'],
['New Caledonia', '8', '8', '9', '17', '37', '350', '59.7', '96.3'],
['New Zealand', '94', '99', '9', '69', '38', '21', '33.4', '30.4'],
['Nicaragua', '64', '33', '7', '17', '40', '50', '21.3', '3.4'],
['Niger', '19', '11', '13', '13', '4', '4', '18.2', '...'],
['Nigeria', '265', '215', '31', '21', '74', '205', '13.9', '0.0'],
['Northern Mariana Islands', '0', '0', '3', '17', '14', '6', '7.7', '33.2'],
['Norway', '121', '122', '8', '11', '23', '10', '30.5', '9.8'],
['Oman', '0', '0', '10', '13', '36', '6', '3.6', '0.3'],
['Pakistan', '50', '37', '25', '33', '46', '12', '12.3', '0.8'],
['Palau', '0', '0', '4', '6', '21', '5', '44.2', '100.0'],
['Panama', '46', '42', '17', '25', '57', '212', '31.4', '26.8'],
['Papua New Guinea', '364', '358', '41', '39', '61', '179', '3.7', '0.1'],
['Paraguay', '255', '158', '10', '27', '0', '20', '14.3', '...'],
['Peru', '764', '722', '53', '119', '52', '328', '22.5', '8.0'],
['Philippines', '78', '72', '38', '93', '91', '254', '15.9', '1.7'],
['Poland', '89', '95', '4', '11', '8', '11', '39.6', '22.6'],
['Portugal', '34', '33', '13', '15', '67', '101', '22.9', '16.9'],
['Puerto Rico', '3', '5', '2', '11', '32', '64', '7.4', '1.8'],
['Qatar', '0', '0', '4', '9', '17', '0', '15.8', '2.3'],
['Romania', '64', '69', '9', '17', '22', '7', '24.5', '23.1'],
['Russian Federation',
'8,090',
'8,153',
'34',
'57',
'39',
'56',
'11.5',
'2.2'],
['Rwanda', '3', '3', '24', '20', '7', '41', '9.1', '...'],
['Samoa', '2', '2', '2', '6', '16', '2', '8.2', '0.1'],
['San Marino', '0', '0', '0', '0', '0', '0', '...', '...'],
['Sao Tome and Principe', '1', '1', '4', '12', '25', '49', '29.3', '0.0'],
['Saudi Arabia', '10', '10', '11', '18', '41', '4', '4.8', '2.5'],
['Senegal', '93', '80', '19', '19', '58', '14', '26.4', '1.9'],
['Serbia', '...', '27', '6', '13', '15', '6', '8.1', '...'],
['Seychelles', '0', '0', '6', '13', '24', '61', '61.5', '32.8'],
['Sierra Leone', '31', '25', '22', '16', '57', '72', '12.7', '1.6'],
['Singapore', '0', '0', '14', '22', '29', '62', '5.6', '0.0'],
['Sint Maarten (Dutch part)', '...', '0', '2', '1', '31', '3', '0.7', '8.7'],
['Slovak Republic', '19', '19', '4', '12', '5', '25', '37.6', '...'],
['Slovenia', '12', '12', '6', '10', '33', '7', '40.4', '2.3'],
['Solomon Islands', '15', '25', '19', '24', '24', '16', '2.0', '0.1'],
['Somalia', '83', '59', '16', '21', '31', '49', '0.0', '0.0'],
['South Africa', '181', '170', '30', '54', '121', '153', '9.3', '15.5'],
['South Sudan', '...', '72', '14', '21', '0', '17', '15.5', '...'],
['Spain', '139', '186', '17', '19', '83', '247', '28.1', '12.8'],
['Sri Lanka', '24', '21', '30', '16', '57', '297', '29.9', '0.1'],
['St. Kitts and Nevis', '0', '0', '2', '3', '30', '2', '22.9', '4.0'],
['St. Lucia', '0', '0', '2', '7', '30', '7', '18.7', '0.2'],
['St. Martin (French part)', '...', '0', '2', '1', '3', '12.8', '96.4'],
['St. Vincent and the Grenadines',
'0',
'0',
'2',
'4',
'29',
'6',
'22.4',
'0.2'],
['Sudan', '307', '182', '16', '26', '32', '17', '2.3', '16.0'],
['Suriname', '154', '152', '9', '9', '32', '27', '14.5', '1.5'],
['Sweden', '281', '280', '1', '11', '15', '5', '15.4', '15.8'],
['Switzerland', '12', '13', '3', '9', '4', '12.1', '...'],
['Syrian Arab Republic', '4', '5', '14', '17', '51', '26', '0.7', '0.2'],
['Tajikistan', '4', '4', '7', '15', '5', '12', '22.3', '...'],
['Tanzania', '574', '453', '41', '49', '179', '644', '38.4', '3.0'],
['Thailand', '194', '198', '59', '62', '106', '159', '18.5', '4.4'],
['Timor-Leste', '10', '9', '5', '6', '13', '2', '16.0', '1.4'],
['Togo', '14', '12', '13', '13', '38', '13', '28.0', '0.2'],
['Tonga', '0', '0', '2', '5', '16', '5', '12.6', '0.1'],
['Trinidad and Tobago', '2', '2', '2', '5', '37', '50', '30.6', '0.0'],
['Tunisia', '6', '7', '14', '11', '39', '9', '7.9', '1.0'],
['Turkiye', '198', '224', '19', '20', '131', '113', '7.0', '1.8'],
['Turkmenistan', '41', '41', '10', '19', '13', '4', '3.2', '3.0'],
['Turks and Caicos Islands', '0', '0', '2', '4', '30', '9', '44.4', '0.1'],
['Tuvalu', '0', '0', '2', '1', '13', '0', '13.2', '0.0'],
['Uganda', '36', '23', '31', '30', '60', '64', '16.1', '...'],
['Ukraine', '93', '97', '11', '17', '24', '22', '13.0', '9.2'],
['United Arab Emirates', '2', '3', '8', '13', '20', '0', '19.4', '11.5'],
['United Kingdom', '28', '32', '5', '11', '48', '42', '27.8', '44.1'],
['United States', '3,025', '3,098', '40', '91', '251', '510', '13.0', '19.0'],
['Uruguay', '8', '21', '10', '22', '45', '22', '3.7', '0.8'],
['Uzbekistan', '26', '37', '10', '19', '7', '16', '5.8', '...'],
['Vanuatu', '4', '4', '8', '8', '18', '10', '4.2', '0.0'],
['Venezuela, RB', '520', '462', '35', '52', '45', '86', '56.9', '4.4'],
['Vietnam', '94', '147', '56', '52', '83', '231', '7.6', '0.6'],
['Virgin Islands (U.S.)', '0', '0', '1', '3', '29', '17', '14.4', '0.9'],

```
['West Bank and Gaza', '0', '0', '4', '15', '2', '6', '8.4', '..'],
['Yemen, Rep.', '5', '5', '11', '16', '40', '163', '0.8', '0.5'],
['Zambia', '474', '446', '13', '20', '20', '23', '41.3', '..'],
['Zimbabwe', '188', '174', '10', '19', '3', '52', '27.2', '..'],
['World',
 '42,034',
 '40,449',
 '3,434',
 '4,584',
 '8,233',
 '15,735',
 '15.3',
 '11.9'],
['East Asia & Pacific',
 '6,398',
 '6,649',
 '934',
 '1,224',
 '1,630',
 '3,799',
 '17.1',
 '18.3'],
['Europe & Central Asia',
 '10,232',
 '10,576',
 '350',
 '678',
 '1,239',
 '1,306',
 '14.2',
 '10.7'],
['Latin America & Caribbean',
 '10,700',
 '9,296',
 '629',
 '1,117',
 '1,716',
 '5,439',
 '24.1',
 '19.4'],
['Middle East & North Africa',
 '205',
 '230',
 '228',
 '290',
 '672',
 '374',
 '5.1',
 '1.3'],
['North America',
 '6,507',
 '6,567',
 '62',
 '118',
 '322',
 '536',
 '12.3',
 '12.8'],
['South Asia', '826', '900', '252', '253', '397', '794', '8.7', '0.5'],
['Sub-Saharan Africa',
 '7,340',
 '6,232',
 '967',
 '993',
 '2,064',
 '4,862',
 '16.4',
 '..'],
['Low income', '3,524', '2,971', '584', '578', '963', '2,291', '12.2', '..'],
['Lower middle income',
 '6,286',
 '5,824',
 '1,255',
 '1,436',
 '2,408',
 '4,886',
 '14.1',
 '1.6'],
['Upper middle income',
 '21,415',
 '20,679',
 '980',
 '1,502',
 '2,179',
 '6,952',
 '15.6',
 '11.0'],
['High income',
 '10,110',
 '10,513',
 '568',
 '1,105',
 '2,445',
 '2,895',
 '15.7',
 '19.3']]
```

```
In [11]: # Create data frame of development indicator data
development_data = pd.DataFrame(data_rows1)
development_data
```

Out[11]:

		0	1	2	3	4	5	6	7	8
0	Afghanistan	12	12	11	16	4	5	3.6	..	
1	Albania	8	8	3	8	44	4	18.6	2.8	
2	Algeria	17	20	14	15	41	22	4.6	0.1	
3	American Samoa	0	0	1	8	12	1	15.8	8.7	
4	Andorra	0	0	2	3	0	0	26.9	..	
...	
221	Sub-Saharan Africa	7,340	6,232	967	993	2,064	4,862	16.4	..	
222	Low income	3,524	2,971	584	578	963	2,291	12.2	..	
223	Lower middle income	6,286	5,824	1,255	1,436	2,408	4,886	14.1	1.6	
224	Upper middle income	21,415	20,679	980	1,502	2,179	6,952	15.6	11.0	
225	High income	10,110	10,513	568	1,105	2,445	2,895	15.7	19.3	

226 rows × 9 columns

Transformation 1- Add column names

I tried repeatedly to use hierarchical indexing for this table because that's how it appeared on the website, but was unsuccessful. I was able to create the multi-index but then had a hard time correctly accessing needed elements of the dataframe. In the end, I decided it was better to just name the columns as needed rather than using hierarchical indexing.

```
In [12]: # Add column names
development_data.columns = ['Country', 'Forest Area (sq.km thousands) 1990', 'Forest Area (sq.km thousands) 2021', 'Threatened Mammals', 'Threatened Birds', 'Threatened Fishes', 'Threatened Higher Plants',
development_data

Out[12]:
```

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Threatened Mammals	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Terrestrial protected areas (% of total land area) 2022	Marine protected areas (% of total territorial waters) 2022
0	Afghanistan	12	12	11	16	4	5	3.6	..
1	Albania	8	8	3	8	44	4	18.6	2.8
2	Algeria	17	20	14	15	41	22	4.6	0.1
3	American Samoa	0	0	1	8	12	1	15.8	8.7
4	Andorra	0	0	2	3	0	0	26.9	..
...
221	Sub-Saharan Africa	7,340	6,232	967	993	2,064	4,862	16.4	..
222	Low income	3,524	2,971	584	578	963	2,291	12.2	..
223	Lower middle income	6,286	5,824	1,255	1,436	2,408	4,886	14.1	1.6
224	Upper middle income	21,415	20,679	980	1,502	2,179	6,952	15.6	11.0
225	High income	10,110	10,513	568	1,105	2,445	2,895	15.7	19.3

226 rows x 9 columns

Transformation 2- Remove income levels from country column

In my other datasets, information is filtered by country. Information on income level is not important to me in this investigation.

```
In [13]: # Sorts index
development_data = development_data.sort_index(axis=1)
development_data

Out[13]:
```

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
0	Afghanistan	12	12	..	3.6	16	4	5	11
1	Albania	8	8	2.8	18.6	8	44	4	3
2	Algeria	17	20	0.1	4.6	15	41	22	14
3	American Samoa	0	0	8.7	15.8	8	12	1	1
4	Andorra	0	0	..	26.9	3	0	0	2
...
221	Sub-Saharan Africa	7,340	6,232	..	16.4	993	2,064	4,862	967
222	Low income	3,524	2,971	..	12.2	578	963	2,291	584
223	Lower middle income	6,286	5,824	1.6	14.1	1,436	2,408	4,886	1,255
224	Upper middle income	21,415	20,679	11.0	15.6	1,502	2,179	6,952	980
225	High income	10,110	10,513	19.3	15.7	1,105	2,445	2,895	568

226 rows x 9 columns

```
In [14]: development_data.set_index(development_data['Country'], inplace=True)
development_data = development_data.drop('Country', axis=1)
development_data

Out[14]:
```

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
	Afghanistan	12	12	..	3.6	16	4	5	11
	Albania	8	8	2.8	18.6	8	44	4	3
	Algeria	17	20	0.1	4.6	15	41	22	14
	American Samoa	0	0	8.7	15.8	8	12	1	1
	Andorra	0	0	..	26.9	3	0	0	2

	Sub-Saharan Africa	7,340	6,232	..	16.4	993	2,064	4,862	967
	Low income	3,524	2,971	..	12.2	578	963	2,291	584
	Lower middle income	6,286	5,824	1.6	14.1	1,436	2,408	4,886	1,255
	Upper middle income	21,415	20,679	11.0	15.6	1,502	2,179	6,952	980
	High income	10,110	10,513	19.3	15.7	1,105	2,445	2,895	568

226 rows x 8 columns

```
In [15]: # Check shape of dataframe
development_data.shape

Out[15]: (226, 8)

In [16]: # Drop rows containing income level data
development_data = development_data.drop(index=['Low income', 'Lower middle income', 'Upper middle income', 'High income'])
development_data
```

Out[16]:

	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
Country								
Afghanistan	12	12	..	3.6	16	4	5	11
Albania	8	8	2.8	18.6	8	44	4	3
Algeria	17	20	0.1	4.6	15	41	22	14
American Samoa	0	0	8.7	15.8	8	12	1	1
Andorra	0	0	..	26.9	3	0	0	2
...
Latin America & Caribbean	10,700	9,296	19.4	24.1	1,117	1,716	5,439	629
Middle East & North Africa	205	230	1.3	5.1	290	672	374	228
North America	6,507	6,567	12.8	12.3	118	322	536	62
South Asia	826	900	0.5	8.7	253	397	794	252
Sub-Saharan Africa	7,340	6,232	..	16.4	993	2,064	4,862	967

222 rows × 8 columns

In [17]:

```
# Checks shape of dataframe to confirm rows were dropped
development_data.shape
```

Out[17]: (222, 8)

In [18]:

```
# Resets index
development_data = development_data.reset_index()
development_data
```

Out[18]:

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
0	Afghanistan	12	12	..	3.6	16	4	5	11
1	Albania	8	8	2.8	18.6	8	44	4	3
2	Algeria	17	20	0.1	4.6	15	41	22	14
3	American Samoa	0	0	8.7	15.8	8	12	1	1
4	Andorra	0	0	..	26.9	3	0	0	2
...
217	Latin America & Caribbean	10,700	9,296	19.4	24.1	1,117	1,716	5,439	629
218	Middle East & North Africa	205	230	1.3	5.1	290	672	374	228
219	North America	6,507	6,567	12.8	12.3	118	322	536	62
220	South Asia	826	900	0.5	8.7	253	397	794	252
221	Sub-Saharan Africa	7,340	6,232	..	16.4	993	2,064	4,862	967

222 rows × 9 columns

Transformation 3- Fix blank values

In [19]:

```
# Check number of Na values in each column
development_data.isna().sum()
```

Out[19]:

```
Country                0
Forest Area (sq.km thousands) 1990    0
Forest Area (sq.km thousands) 2021    0
Marine protected areas (% of total territorial waters) 2022    0
Terrestrial protected areas (% of total land area) 2022    0
Threatened Birds        0
Threatened Fishes       0
Threatened Higher Plants 0
Threatened Mammals      0
dtype: int64
```

This dataset does not contain NA values, but missing values are filled with two periods (..), I need to fill these with NA values so data can be sorted correctly.

In [20]:

```
# Replace '..' values with NaN
development_data = development_data.replace('..', np.nan)
development_data
```

Out[20]:

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
0	Afghanistan	12	12	NaN	3.6	16	4	5	11
1	Albania	8	8	2.8	18.6	8	44	4	3
2	Algeria	17	20	0.1	4.6	15	41	22	14
3	American Samoa	0	0	8.7	15.8	8	12	1	1
4	Andorra	0	0	NaN	26.9	3	0	0	2
...
217	Latin America & Caribbean	10,700	9,296	19.4	24.1	1,117	1,716	5,439	629
218	Middle East & North Africa	205	230	1.3	5.1	290	672	374	228
219	North America	6,507	6,567	12.8	12.3	118	322	536	62
220	South Asia	826	900	0.5	8.7	253	397	794	252
221	Sub-Saharan Africa	7,340	6,232	NaN	16.4	993	2,064	4,862	967

222 rows × 9 columns

In [21]:

```
# Find number of NaN values after fixing dataframe
development_data.isna().sum()
```

```
Out[21]: Country                                0
Forest Area (sq.km thousands) 1990           11
Forest Area (sq.km thousands) 2021            3
Marine protected areas (% of total territorial waters) 2022  45
Terrestrial protected areas (% of total land area) 2022      4
Threatened Birds                                           2
Threatened Fishes                                          2
Threatened Higher Plants                                   2
Threatened Mammals                                         2
dtype: int64
```

Transformation 4- Remove rows that have NaN values

To work with this data effectively, I need to remove rows that contain NaN values.

```
In [22]: development_data.set_index(development_data['Country'], inplace=True)
development_data = development_data.drop('Country', axis=1)

# Drop rows with more than 1 Na value
development_data = development_data.dropna()
development_data
```

Out[22]:

	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
Country								
Albania	8	8	2.8	18.6	8	44	4	3
Algeria	17	20	0.1	4.6	15	41	22	14
American Samoa	0	0	8.7	15.8	8	12	1	1
Angola	793	661	0.0	7.0	32	53	34	18
Antigua and Barbuda	0	0	0.3	19.9	3	31	4	2
...
Europe & Central Asia	10,232	10,576	10.7	14.2	678	1,239	1,306	350
Latin America & Caribbean	10,700	9,296	19.4	24.1	1,117	1,716	5,439	629
Middle East & North Africa	205	230	1.3	5.1	290	672	374	228
North America	6,507	6,567	12.8	12.3	118	322	536	62
South Asia	826	900	0.5	8.7	253	397	794	252

171 rows × 8 columns

```
In [23]: development_data.isna().sum()
```

```
Out[23]: Forest Area (sq.km thousands) 1990           0
Forest Area (sq.km thousands) 2021           0
Marine protected areas (% of total territorial waters) 2022  0
Terrestrial protected areas (% of total land area) 2022      0
Threatened Birds                                           0
Threatened Fishes                                          0
Threatened Higher Plants                                   0
Threatened Mammals                                         0
dtype: int64
```

Transformation 5- Change data types of columns

```
In [24]: # Check data types
development_data.dtypes
```

```
Out[24]: Forest Area (sq.km thousands) 1990           object
Forest Area (sq.km thousands) 2021           object
Marine protected areas (% of total territorial waters) 2022  object
Terrestrial protected areas (% of total land area) 2022      object
Threatened Birds                                           object
Threatened Fishes                                          object
Threatened Higher Plants                                   object
Threatened Mammals                                         object
dtype: object
```

```
In [25]: # Loop through column in data frame to first change them to strings, then remove commas, and finally convert them to floats
for column in development_data:
    development_data[column] = development_data[column].astype(str)
    development_data[column] = development_data[column].str.replace(',', '')
    development_data[column] = development_data[column].astype(float)
development_data
```

Out[25]:

	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals
Country								
Albania	8.0	8.0	2.8	18.6	8.0	44.0	4.0	3.0
Algeria	17.0	20.0	0.1	4.6	15.0	41.0	22.0	14.0
American Samoa	0.0	0.0	8.7	15.8	8.0	12.0	1.0	1.0
Angola	793.0	661.0	0.0	7.0	32.0	53.0	34.0	18.0
Antigua and Barbuda	0.0	0.0	0.3	19.9	3.0	31.0	4.0	2.0
...
Europe & Central Asia	10232.0	10576.0	10.7	14.2	678.0	1239.0	1306.0	350.0
Latin America & Caribbean	10700.0	9296.0	19.4	24.1	1117.0	1716.0	5439.0	629.0
Middle East & North Africa	205.0	230.0	1.3	5.1	290.0	672.0	374.0	228.0
North America	6507.0	6567.0	12.8	12.3	118.0	322.0	536.0	62.0
South Asia	826.0	900.0	0.5	8.7	253.0	397.0	794.0	252.0

171 rows × 8 columns

Transformation 6: Add column for change in forest area from 1990 to 2021

```
In [26]: # Adds column while preforming needed calculation
development_data['Change in Forest Area 1990 to 2021'] = development_data['Forest Area (sq.km thousands) 1990'] - development_data['Forest Area (sq.km thousands) 2021']
development_data
```

Out[26]:

	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals	Change in Forest Area 1990 to 2021
Country									
Albania	8.0	8.0	2.8	18.6	8.0	44.0	4.0	3.0	0.0
Algeria	17.0	20.0	0.1	4.6	15.0	41.0	22.0	14.0	-3.0
American Samoa	0.0	0.0	8.7	15.8	8.0	12.0	1.0	1.0	0.0
Angola	793.0	661.0	0.0	7.0	32.0	53.0	34.0	18.0	132.0
Antigua and Barbuda	0.0	0.0	0.3	19.9	3.0	31.0	4.0	2.0	0.0
...
Europe & Central Asia	10232.0	10576.0	10.7	14.2	678.0	1239.0	1306.0	350.0	-344.0
Latin America & Caribbean	10700.0	9296.0	19.4	24.1	1117.0	1716.0	5439.0	629.0	1404.0
Middle East & North Africa	205.0	230.0	1.3	5.1	290.0	672.0	374.0	228.0	-25.0
North America	6507.0	6567.0	12.8	12.3	118.0	322.0	536.0	62.0	-60.0
South Asia	826.0	900.0	0.5	8.7	253.0	397.0	794.0	252.0	-74.0

171 rows × 9 columns

Writing final table to CSV file

```
In [27]: # Writing dataframe to a csv file
development_data.to_csv('DevelopmentData', sep=',', encoding='utf-8', index=True)
```

```
In [28]: # Checking that writing to file worked correctly
csvFile = pd.read_csv("C:/Users/kayly/OneDrive/Desktop/MSDS/DSC540/Tem Project/DevelopmentData")
csvFile
```

Out[28]:

	Country	Forest Area (sq.km thousands) 1990	Forest Area (sq.km thousands) 2021	Marine protected areas (% of total territorial waters) 2022	Terrestrial protected areas (% of total land area) 2022	Threatened Birds	Threatened Fishes	Threatened Higher Plants	Threatened Mammals	Change in Forest Area 1990 to 2021
0	Albania	8.0	8.0	2.8	18.6	8.0	44.0	4.0	3.0	0.0
1	Algeria	17.0	20.0	0.1	4.6	15.0	41.0	22.0	14.0	-3.0
2	American Samoa	0.0	0.0	8.7	15.8	8.0	12.0	1.0	1.0	0.0
3	Angola	793.0	661.0	0.0	7.0	32.0	53.0	34.0	18.0	132.0
4	Antigua and Barbuda	0.0	0.0	0.3	19.9	3.0	31.0	4.0	2.0	0.0
...
166	Europe & Central Asia	10232.0	10576.0	10.7	14.2	678.0	1239.0	1306.0	350.0	-344.0
167	Latin America & Caribbean	10700.0	9296.0	19.4	24.1	1117.0	1716.0	5439.0	629.0	1404.0
168	Middle East & North Africa	205.0	230.0	1.3	5.1	290.0	672.0	374.0	228.0	-25.0
169	North America	6507.0	6567.0	12.8	12.3	118.0	322.0	536.0	62.0	-60.0
170	South Asia	826.0	900.0	0.5	8.7	253.0	397.0	794.0	252.0	-74.0

171 rows × 10 columns

```
In [ ]:
```

Ethical Implications

This dataset was overall a pretty clean dataset. There were not many missing values and those that were missing were clearly labeled. I chose to remove NaN values in this dataset to make further processing easier down the line. Because I will be combining datasets based on the country, it is important that the countries I keep have all necessary data.

I did not search for outliers within this dataset. It's would be almost impossible to correctly categorize values as outliers because so much of this data is dependent on the country the values are gathered from. For example, comparing forest area of Angola and the United States would show that Angola has a small forest area while the United States has a massive forest area. This is dependent on the size of the country. Just because one value is massively bigger than the other, does not mean it's an outlier.

I chose to be conservative when transforming data and only changed obviously values. The main change I made was to convert _ to NaN and then drop those values from the dataset. I don't see this to be very risky because I didn't alter any values in the dataset. I did end up dropping about 50 countries or income levels from the dataset because of mising values. In this case, I see it as more ethically sound to remove these countries from the dataset rather than alter them. In other instances, this could be seen as skewing data.

My data was sourced form the World Data Bank which is a well known and reputable source of data. It often accumulates data from multiple studies for further evaluation by data scientists. I am unsure how the original data was collected, so I cannot be sure there are not ethical breeches there. As far as I know, there are no legal or regulatory implications in my data.