In [1]:

```
print("Name: ")
print("Clean the data and show which top 10 countries has the highest Undernourished rate as compared with their Population")
print("Show a comparison between Meat and Vegetables consumption across the countries")
```

In [2]:

```
#predefine code for image
from IPython.display import Image
Image(filename='healthy.jpg')
#predefine code end
```

Out[2]:

Activity - 1 Clean the data and show which top 10 countries has the highest Undernourished rate as compared with their Population

In [3]:

```
#import the required packages
import pandas as pd
import matplotlib.pyplot as plt

#Read the csv file.
df = pd.read_csv("COVID-19 Healthy Diet Dataset.csv")
df
```

Out[3]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetables	Miscellaneous	Obesity	Undernourished	Confirmed	Deaths	Recovered
0	Afghanistan	0.0000	9.7523	0.0277	0.0000	35.9771	0.4067	0.0647	0.5824	3.1337	...	1.1370	0.0462	4.5	29.8	0.138283	0.005970	0.116713
1	Albania	0.1840	27.7469	0.0711	0.0000	14.2331	1.8069	0.6274	1.2757	7.6582	...	3.2456	0.0544	22.3	6.2	2.347956	0.044574	1.396230
2	Algeria	0.0323	13.8360	0.0054	0.0000	26.5633	1.2916	0.6350	1.1624	3.5088	...	3.1267	0.1399	26.6	3.9	0.233066	0.006373	0.158234
3	Angola	0.6285	15.2311	0.0277	0.0000	20.3882	0.1756	5.4436	1.2754	7.6248	...	0.8133	0.0924	6.8	25	0.057435	0.001316	0.049566
4	Antigua and Barbuda	0.1535	33.1901	0.1289	0.0000	10.5108	0.4850	8.2146	1.2586	16.0670	...	1.6024	0.2947	19.1	NaN	0.187755	0.006122	0.159184
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165	Venezuela (Bolivarian Republic of)	0.1955	22.5411	0.1244	0.0000	21.6526	0.8707	2.6477	1.0662	11.8347	...	1.0129	0.0267	25.2	21.2	0.414928	0.003823	0.392110
166	Vietnam	0.1555	20.4466	0.1555	0.0056	18.5247	0.7665	5.7435	0.7165	11.0426	...	3.7216	0.0389	2.1	9.3	0.001597	0.000036	0.001434
167	Yemen	0.0000	10.0122	0.0188	0.0000	35.1179	0.4320	0.9392	0.4884	5.9453	...	0.5448	0.0564	14.1	38.9	0.007078	0.002052	0.004758
168	Zambia	0.4824	9.8925	0.0338	0.0000	28.5182	0.5839	3.0126	0.0931	4.3158	...	0.8039	0.0592	6.5	46.7	0.186456	0.002867	0.131119
169	Zimbabwe	0.2929	11.3443	0.0391	0.0000	33.1934	0.5077	1.0837	0.2636	6.6582	...	0.5955	0.0586	12.3	51.3	0.175664	0.004481	0.103707

170 rows × 32 columns

In [4]:

```
#Cleaning data
df.replace("<2.5", int(3), inplace=True)
df
```

Out[4]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetables	Miscellaneous	Obesity	Undernourished	Confirmed	Deaths	Recovered
0	Afghanistan	0.0000	9.7523	0.0277	0.0000	35.9771	0.4067	0.0647	0.5824	3.1337	...	1.1370	0.0462	4.5	29.8	0.138283	0.005970	0.116713
1	Albania	0.1840	27.7469	0.0711	0.0000	14.2331	1.8069	0.6274	1.2757	7.6582	...	3.2456	0.0544	22.3	6.2	2.347956	0.044574	1.396230
2	Algeria	0.0323	13.8360	0.0054	0.0000	26.5633	1.2916	0.6350	1.1624	3.5088	...	3.1267	0.1399	26.6	3.9	0.233066	0.006373	0.158234
3	Angola	0.6285	15.2311	0.0277	0.0000	20.3882	0.1756	5.4436	1.2754	7.6248	...	0.8133	0.0924	6.8	25	0.057435	0.001316	0.049566
4	Antigua and Barbuda	0.1535	33.1901	0.1289	0.0000	10.5108	0.4850	8.2146	1.2586	16.0670	...	1.6024	0.2947	19.1	NaN	0.187755	0.006122	0.159184
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165	Venezuela (Bolivarian Republic of)	0.1955	22.5411	0.1244	0.0000	21.6526	0.8707	2.6477	1.0662	11.8347	...	1.0129	0.0267	25.2	21.2	0.414928	0.003823	0.392110
166	Vietnam	0.1555	20.4466	0.1555	0.0056	18.5247	0.7665	5.7435	0.7165	11.0426	...	3.7216	0.0389	2.1	9.3	0.001597	0.000036	0.001434
167	Yemen	0.0000	10.0122	0.0188	0.0000	35.1179	0.4320	0.9392	0.4884	5.9453	...	0.5448	0.0564	14.1	38.9	0.007078	0.002052	0.004758
168	Zambia	0.4824	9.8925	0.0338	0.0000	28.5182	0.5839	3.0126	0.0931	4.3158	...	0.8039	0.0592	6.5	46.7	0.186456	0.002867	0.131119
169	Zimbabwe	0.2929	11.3443	0.0391	0.0000	33.1934	0.5077	1.0837	0.2636	6.6582	...	0.5955	0.0586	12.3	51.3	0.175664	0.004481	0.103707

170 rows × 32 columns

In [5]:

```
#Converting object datatype column to float datatype
df['Undernourished'] = df['Undernourished'].astype(float)
df
```

Out[5]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetables	Miscellaneous	Obesity	Undernourished	Confirmed	Deaths	Recovered
0	Afghanistan	0.0000	9.7523	0.0277	0.0000	35.9771	0.4067	0.0647	0.5824	3.1337	...	1.1370	0.0462	4.5	29.8	0.138283	0.005970	0.116713
1	Albania	0.1840	27.7469	0.0711	0.0000	14.2331	1.8069	0.6274	1.2757	7.6582	...	3.2456	0.0544	22.3	6.2	2.347956	0.044574	1.396230
2	Algeria	0.0323	13.8360	0.0054	0.0000	26.5633	1.2916	0.6350	1.1624	3.5088	...	3.1267	0.1399	26.6	3.9	0.233066	0.006373	0.158234
3	Angola	0.6285	15.2311	0.0277	0.0000	20.3882	0.1756	5.4436	1.2754	7.6248	...	0.8133	0.0924	6.8	25.0	0.057435	0.001316	0.049566
4	Antigua and Barbuda	0.1535	33.1901	0.1289	0.0000	10.5108	0.4850	8.2146	1.2586	16.0670	...	1.6024	0.2947	19.1	NaN	0.187755	0.006122	0.159184
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165	Venezuela (Bolivarian Republic of)	0.1955	22.5411	0.1244	0.0000	21.6526	0.8707	2.6477	1.0662	11.8347	...	1.0129	0.0267	25.2	21.2	0.414928	0.003823	0.392110
166	Vietnam	0.1555	20.4466	0.1555	0.0056	18.5247	0.7665	5.7435	0.7165	11.0426	...	3.7216	0.0389	2.1	9.3	0.001597	0.000036	0.001434
167	Yemen	0.0000	10.0122	0.0188	0.0000	35.1179	0.4320	0.9392	0.4884	5.9453	...	0.5448	0.0564	14.1	38.9	0.007078	0.002052	0.004758
168	Zambia	0.4824	9.8925	0.0338	0.0000	28.5182	0.5839	3.0126	0.0931	4.3158	...	0.8039	0.0592	6.5	46.7	0.186456	0.002867	0.131119
169	Zimbabwe	0.2929	11.3443	0.0391	0.0000	33.1934	0.5077	1.0837	0.2636	6.6582	...	0.5955	0.0586	12.3	51.3	0.175664	0.004481	0.103707

170 rows × 32 columns

In [6]:

```
#Groupby country and apply sum on Undernourished and Population column and create a new dataframe out of it
group_country = df.groupby('Country')[['Undernourished','Population']].sum().reset_index()
group_country
```

Out[6]:

	Country	Undernourished	Population
0	Afghanistan	29.8	38928000.0
1	Albania	6.2	2838000.0
2	Algeria	3.9	44357000.0
3	Angola	25.0	32522000.0
4	Antigua and Barbuda	0.0	98000.0
...	...	...	...
165	Venezuela (Bolivarian Republic of)	21.2	28645000.0
166	Vietnam	9.3	96209000.0
167	Yemen	38.9	29826000.0
168	Zambia	46.7	18384000.0
169	Zimbabwe	51.3	14863000.0

170 rows × 3 columns

In [7]:

```
#Sort the new dataframe as per Undernourished column
sorted_by_undernourished = group_country.sort_values(by=['Undernourished'], ascending=False)
sorted_by_undernourished
```

Out[7]:

	Country	Undernourished	Population
27	Central African Republic	59.6	4830000.0
169	Zimbabwe	51.3	14863000.0
64	Haiti	49.3	11403000.0
81	Korea, North	47.8	25779000.0
168	Zambia	46.7	18384000.0
...	...	...	...
10	Bahamas	0.0	393000.0
129	Saint Lucia	0.0	182000.0
128	Saint Kitts and Nevis	0.0	54000.0
148	Tajikistan	0.0	9429000.0
124	Republic of Moldova	0.0	3535000.0

170 rows × 3 columns

In [8]:

```
#Get the top 10 countries from the sorted data
top_10 = sorted_by_undernourished.head(10)

#Convert Undernourished percentage to number and add a new column to the top 10 countries dataframe
top_10['percentage_number'] = top_10['Undernourished']/100 * top_10['Population']
top_10
```

```
<ipython-input-8-d8668cb8659a>:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  top_10['percentage_number'] = top_10['Undernourished']/100 * top_10['Population']
```

Out[8]:

	Country	Undernourished	Population	percentage_number
27	Central African Republic	59.6	4830000.0	2878680.0
169	Zimbabwe	51.3	14863000.0	7624719.0
64	Haiti	49.3	11403000.0	5621679.0
81	Korea, North	47.8	25779000.0	12322362.0
168	Zambia	46.7	18384000.0	8585328.0
92	Madagascar	44.4	27691000.0	12294804.0
156	Uganda	41.0	45741000.0	18753810.0
32	Congo	40.3	5518000.0	2223754.0
167	Yemen	38.9	29826000.0	11602314.0
28	Chad	37.5	16877000.0	6328875.0

In [10]:

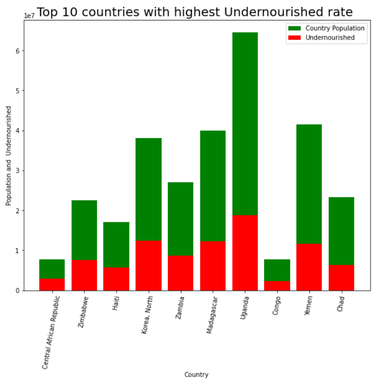
```
#Plot a stacked bar graph for showing the Population vs Undernourished rate

total_population = top_10['Population']
percent_number = top_10['percentage_number']

index = top_10['Country']
plt.subplots(figsize=(10,8))
plt.title('Top 10 countries with highest Undernourished rate ', fontsize=20)

plt.bar(index, total_population, bottom=percent_number , color='green', label='Country Population')
plt.bar(index, percent_number ,  color='red', label = 'Undernourished')
plt.xticks(rotation=80)
plt.legend()
plt.xlabel("Country")
plt.ylabel("Population and  Undernourished")

plt.show()
```



Conclusion - Among the top 10 countries with highest Undernourished rate, Uganda has the highest Undernourished rate

Activity - 2 Show a comparison between Meat and Vegetables consumption across the countries

In [12]:

```
#Sort the big dataframe as per Meat consumption
sorted_meat = df.sort_values(by=['Meat'], ascending=False)
sorted_meat
```

Out[12]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetables	Miscellaneous	Obesity	Undernourished	Confirmed	Deaths	Recovered
5	Argentina	0.1704	31.9799	0.0097	0.0000	13.6702	2.0593	1.0223	0.5209	21.6250	...	1.0516	0.0000	28.5	4.6	3.929407	0.099669	3.442191
10	Bahamas	0.0325	34.1684	0.0895	0.0000	7.8081	1.6755	5.7340	1.8788	20.4311	...	2.6759	1.1305	32.1	NaN	2.040967	0.044529	1.655725
101	Mongolia	0.2367	34.0697	0.1092	0.0000	13.0902	0.7768	0.1032	0.0850	19.8082	...	0.8618	0.1699	19.6	13.4	0.044987	0.000060	0.027046
128	Saint Kitts and Nevis	0.3183	32.0169	0.1104	0.0000	10.9516	0.7795	5.1770	0.8444	19.6622	...	0.8574	0.5781	23.1	NaN	0.062963	0.000000	0.059259
131	Samoa	0.1866	30.9096	0.1341	0.1341	8.4665	0.4023	7.3761	1.6968	19.3878	...	0.5364	0.3032	45.5	2.7	0.001000	0.000000	0.001000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
142	Sri Lanka	0.0699	14.5498	0.0000	0.0000	23.7396	1.0176	7.4963	0.5438	2.4314	...	1.5148	0.0388	5.4	9.0	0.235192	0.001162	0.201755
54	Gambia	0.2116	13.0026	0.0078	0.0000	26.6870	0.3213	6.7403	0.0705	2.3513	...	0.6427	0.2116	8.7	10.2	0.161233	0.005254	0.152627
48	Ethiopia	0.1676	4.4561	0.0076	0.0000	29.7303	0.0990	0.1066	0.1143	1.6834	...	0.4570	0.0076	3.6	20.6	0.113410	0.001760	0.100446
11	Bangladesh	0.0000	9.9195	0.0083	0.0000	31.7243	0.7216	5.8306	0.3069	1.2773	...	1.2026	0.0249	3.4	14.7	0.310045	0.004630	0.277443
68	India	0.0077	11.2582	0.0306	0.0000	25.2794	0.7271	1.5537	0.6659	1.0179	...	2.3343	0.0000	3.8	14.5	0.753006	0.010863	0.727071

170 rows × 32 columns

In [13]:

```
#Sort the big dataframe as per Vegetables consumption
sorted_vegetables = df.sort_values(by=['Vegetables'], ascending=False)
sorted_vegetables
```

Out[13]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Vegetables	Miscellaneous	Obesity	Undernourished	Confirmed	Deaths	Recovered
148	Tajikistan	0.0090	12.1401	0.0000	0.0000	28.2246	0.6499	0.0903	0.5777	5.5601	...	5.0546	0.1083	12.6	NaN	0.141139	0.000955	0.140185
6	Armenia	0.0411	22.8846	0.1438	0.0000	18.2738	1.7560	0.8831	1.1142	8.5233	...	4.6621	0.1078	20.9	4.3	5.547091	0.100609	5.152909
81	Korea, North	0.0287	10.0832	0.0000	0.0000	23.9070	1.3872	2.9561	0.6218	5.0990	...	4.6398	0.0287	7.1	47.8	NaN	NaN	NaN
63	Guyana	0.1537	22.1489	0.0114	0.0000	15.7922	0.2902	4.6153	0.8821	8.5306	...	4.4560	0.2219	19.2	8.1	0.864676	0.021601	0.772935
85	Lao People's Democratic Republic	0.0541	12.0786	0.1082	0.0000	25.5153	0.3365	4.5310	1.9710	5.9912	...	4.4408	0.1622	4.5	16.5	0.000566	0.000000	0.000566
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
131	Samoa	0.1866	30.9096	0.1341	0.1341	8.4665	0.4023	7.3761	1.6968	19.3878	...	0.5364	0.3032	45.5	2.7	0.001000	0.000000	0.001000
139	Solomon Islands	0.1398	14.7782	0.0839	0.0186	15.0205	0.3168	9.1129	0.5498	3.9415	...	0.4752	0.2329	20.5	8.9	0.002378	0.000000	0.001399
48	Ethiopia	0.1676	4.4561	0.0076	0.0000	29.7303	0.0990	0.1066	0.1143	1.6834	...	0.4570	0.0076	3.6	20.6	0.113410	0.001760	0.100446
22	Burkina Faso	1.3660	7.1846	0.0066	0.0000	29.1784	0.4466	1.4842	0.0591	2.9750	...	0.3809	0.0394	4.5	20.0	0.042492	0.000464	0.033617
28	Chad	0.0567	15.4353	0.0425	0.0000	22.6744	0.0709	2.2689	0.0993	7.4872	...	0.1631	0.0142	4.8	37.5	0.016632	0.000658	0.012324

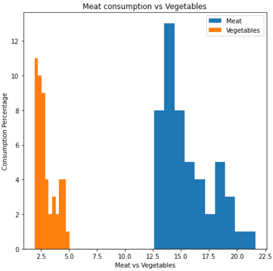
170 rows × 32 columns

In [14]:

```
#Plot a histogram showing the Meat consumption vs Vegetables consumption across countries
plt.figure(figsize=(7,7))

plt.hist(sorted_meat['Meat'].head(50), bins=10, label="Meat")
plt.hist(sorted_vegetables['Vegetables'].head(50), bins=10, label = "Vegetables")

plt.title("Meat consumption vs Vegetables")
plt.ylabel("Consumption Percentage")
plt.xlabel("Meat vs Vegetables")
plt.legend()
plt.show()
```



Conclusion - The Meat consumption is higher then the Vegetable consumption across countries

In [ ]: