In [49]:

```python
print("Name : Lakshya Singh Chauhan")
print("Clean the data and show which top 10 countries has the highest Undernourished rat
print("Show a comparison between Meat and Vegetables consumption across the countries")
```

Name : Lakshya Singh Chauhan
Clean the data and show which top 10 countries has the highest Undernouri
shed rate as compared with their Population
Show a comparison between Meat and Vegetables consumption across the coun
tries

In [50]:

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

Out[50]:

Activity - 1 Clean the data and show which top 10 countries has the highest Undernourished rate as compared with their Population
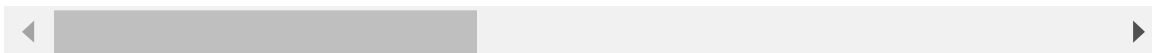
In [51]:

```python
#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")
new_df = df.dropna()
df
```

Out[51]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fi Excl
0	Afghanistan	0.0000	9.7523	0.0277	0.0000	35.9771	0.4067	0.0647	0
1	Albania	0.1840	27.7469	0.0711	0.0000	14.2331	1.8069	0.6274	1
2	Algeria	0.0323	13.8360	0.0054	0.0000	26.5633	1.2916	0.6350	1
3	Angola	0.6285	15.2311	0.0277	0.0000	20.3882	0.1756	5.4436	1
4	Antigua and Barbuda	0.1535	33.1901	0.1289	0.0000	10.5108	0.4850	8.2146	1
...	...	...	...	...	...	...	...	...	
165	Venezuela (Bolivarian Republic of)	0.1955	22.5411	0.1244	0.0000	21.6526	0.8707	2.6477	1
166	Vietnam	0.1555	20.4466	0.1555	0.0056	18.5247	0.7665	5.7435	0
167	Yemen	0.0000	10.0122	0.0188	0.0000	35.1179	0.4320	0.9392	0
168	Zambia	0.4824	9.8925	0.0338	0.0000	28.5182	0.5839	3.0126	0
169	Zimbabwe	0.2929	11.3443	0.0391	0.0000	33.1934	0.5077	1.0837	0

170 rows × 32 columns

In [52]:

```python
#Cleaning data
new_df['Undernourished'].astype(float)
clean_data = new_df["Undernourished"]
clean_data.replace("<2.5", int(3), inplace=True)
new_df.dropna()
clean_data
```

C:\Users\laksh\AppData\Local\Temp\ipykernel_14300\1003417287.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  clean_data.replace("<2.5", int(3), inplace=True)

Out[52]:

```
0      29.8
1       6.2
2       3.9
3      25.0
5       4.6
       ...
165    21.2
166     9.3
167    38.9
168    46.7
169    51.3
Name: Undernourished, Length: 154, dtype: float64
```
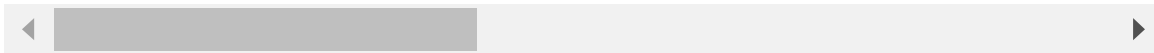
In [53]:

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

Out[53]:

	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fr Excl
0	Afghanistan	0.0000	9.7523	0.0277	0.0000	35.9771	0.4067	0.0647	0
1	Albania	0.1840	27.7469	0.0711	0.0000	14.2331	1.8069	0.6274	1
2	Algeria	0.0323	13.8360	0.0054	0.0000	26.5633	1.2916	0.6350	1
3	Angola	0.6285	15.2311	0.0277	0.0000	20.3882	0.1756	5.4436	1
4	Antigua and Barbuda	0.1535	33.1901	0.1289	0.0000	10.5108	0.4850	8.2146	1
...	...	...	...	...	...	...	...	...	
165	Venezuela (Bolivarian Republic of)	0.1955	22.5411	0.1244	0.0000	21.6526	0.8707	2.6477	1
166	Vietnam	0.1555	20.4466	0.1555	0.0056	18.5247	0.7665	5.7435	0
167	Yemen	0.0000	10.0122	0.0188	0.0000	35.1179	0.4320	0.9392	0
168	Zambia	0.4824	9.8925	0.0338	0.0000	28.5182	0.5839	3.0126	0
169	Zimbabwe	0.2929	11.3443	0.0391	0.0000	33.1934	0.5077	1.0837	0

170 rows × 32 columns

In [54]:

```
#Groupby country and apply sum on Undernourished and Population column and create a new
group_country = df.groupby(["Undernourished", "Population"])["Country"].sum().reset_inde
group_country
```

Out[54]:

	Undernourished	Population	Country
0	2.5	32781000.0	Malaysia
1	2.6	9778000.0	United Arab Emirates
2	2.7	125000.0	Kiribati
3	2.7	200000.0	Samoa
4	2.7	19470000.0	Chile
...	...	...	...
158	46.7	18384000.0	Zambia
159	47.8	25779000.0	Korea, North
160	49.3	11403000.0	Haiti
161	51.3	14863000.0	Zimbabwe
162	59.6	4830000.0	Central African Republic

163 rows × 3 columns

In [55]:

```
#Sort the new dataframe as per Undernourished column
sort1 = group_country.sort_values(by="Undernourished", ascending=False)
sort1
```

Out[55]:

	Undernourished	Population	Country
162	59.6	4830000.0	Central African Republic
161	51.3	14863000.0	Zimbabwe
160	49.3	11403000.0	Haiti
159	47.8	25779000.0	Korea, North
158	46.7	18384000.0	Zambia
...	...	...	...
4	2.7	19470000.0	Chile
2	2.7	125000.0	Kiribati
3	2.7	200000.0	Samoa
1	2.6	9778000.0	United Arab Emirates
0	2.5	32781000.0	Malaysia

163 rows × 3 columns

In [56]:

```python
#Get the top 10 countries from the sorted data
top_10 = sort1.head(10)
top_10
#Convert Undernourished percentage to number and add a new column to the top 10 countrie
top_10["percentage_number"] = top_10["Undernourished"]/100 * top_10["Population"]
top_10
```

C:\Users\laksh\AppData\Local\Temp\ipykernel_14300\1366676030.py:5: Settin
gWithCopyWarning:
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-do
cs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (http
s://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
ng-a-view-versus-a-copy)
  top_10["percentage_number"] = top_10["Undernourished"]/100 * top_10["Po
pulation"]

Out[56]:

	Undernourished	Population	Country	percentage_number
162	59.6	4830000.0	Central African Republic	2878680.0
161	51.3	14863000.0	Zimbabwe	7624719.0
160	49.3	11403000.0	Haiti	5621679.0
159	47.8	25779000.0	Korea, North	12322362.0
158	46.7	18384000.0	Zambia	8585328.0
157	44.4	27691000.0	Madagascar	12294804.0
156	41.0	45741000.0	Uganda	18753810.0
155	40.3	5518000.0	Congo	2223754.0
154	38.9	29826000.0	Yemen	11602314.0
153	37.5	16877000.0	Chad	6328875.0

In [57]:

```python
#Plot a stacked bar graph for showing the Population vs Undernourished rate
# Plot the data
ax = df.plot(kind='bar', x='Country', stacked=True)

# Add some labels and legends
ax.set_ylabel('Population / Undernourished')
ax.set_title('Population vs Undernourished Rate by Country')

# Display the plot
plt.show()
```
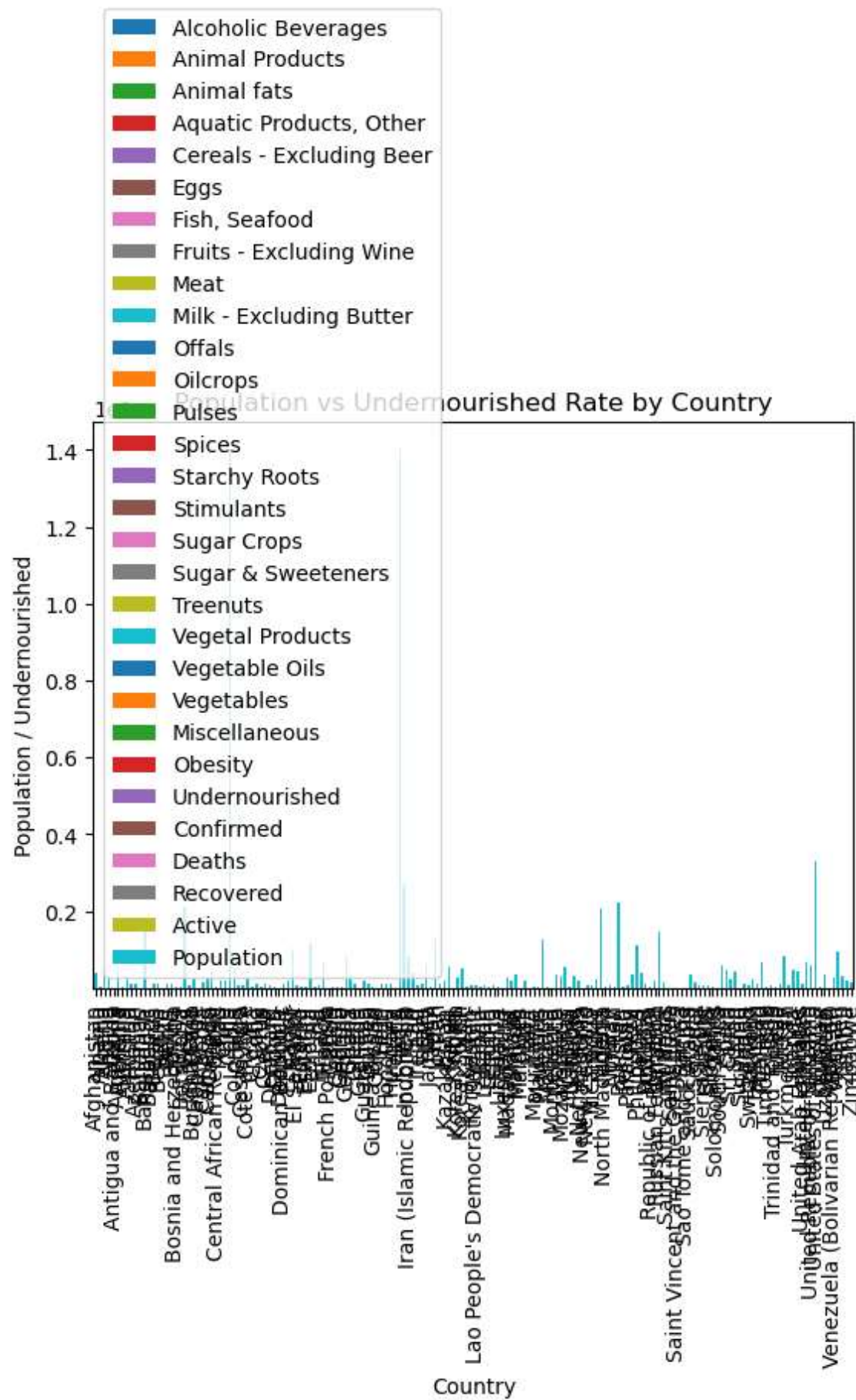
Conclusion -

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

In [58]:

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

Out[58]:

|	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fr Excl
5	Argentina	0.1704	31.9799	0.0097	0.0000	13.6702	2.0593	1.0223	0
10	Bahamas	0.0325	34.1684	0.0895	0.0000	7.8081	1.6755	5.7340	1
101	Mongolia	0.2367	34.0697	0.1092	0.0000	13.0902	0.7768	0.1032	0
128	Saint Kitts and Nevis	0.3183	32.0169	0.1104	0.0000	10.9516	0.7795	5.1770	0
131	Samoa	0.1866	30.9096	0.1341	0.1341	8.4665	0.4023	7.3761	1
...	...	...	...	...	...	...	...	...	
142	Sri Lanka	0.0699	14.5498	0.0000	0.0000	23.7396	1.0176	7.4963	0
54	Gambia	0.2116	13.0026	0.0078	0.0000	26.6870	0.3213	6.7403	0
48	Ethiopia	0.1676	4.4561	0.0076	0.0000	29.7303	0.0990	0.1066	0
11	Bangladesh	0.0000	9.9195	0.0083	0.0000	31.7243	0.7216	5.8306	0
68	India	0.0077	11.2582	0.0306	0.0000	25.2794	0.7271	1.5537	0

170 rows × 32 columns
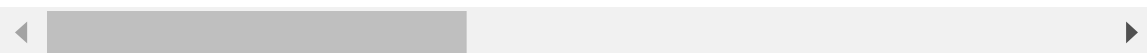
In [59]:

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

Out[59]:

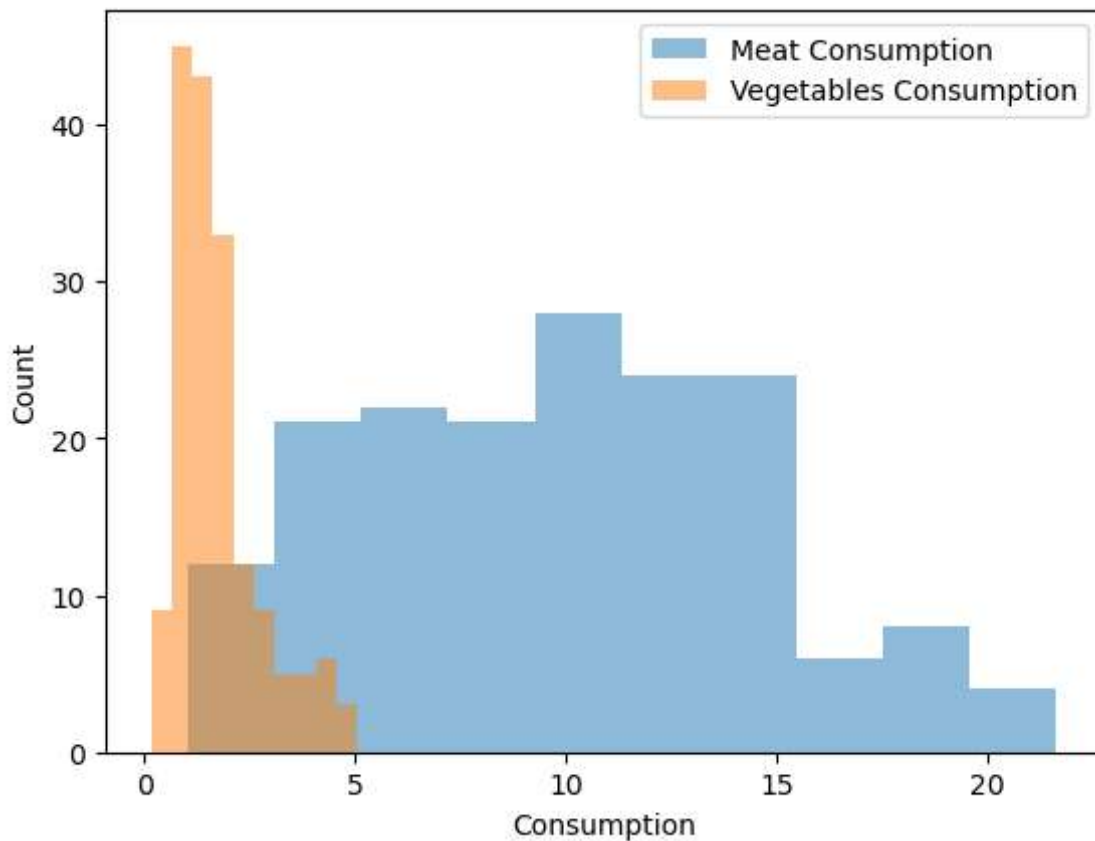|	Country	Alcoholic Beverages	Animal Products	Animal fats	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fr Exclu
148	Tajikistan	0.0090	12.1401	0.0000	0.0000	28.2246	0.6499	0.0903	0
6	Armenia	0.0411	22.8846	0.1438	0.0000	18.2738	1.7560	0.8831	1
81	Korea, North	0.0287	10.0832	0.0000	0.0000	23.9070	1.3872	2.9561	0
63	Guyana	0.1537	22.1489	0.0114	0.0000	15.7922	0.2902	4.6153	0
85	Lao People's Democratic Republic	0.0541	12.0786	0.1082	0.0000	25.5153	0.3365	4.5310	1
...	...	...	...	...	...	...	...	...	
131	Samoa	0.1866	30.9096	0.1341	0.1341	8.4665	0.4023	7.3761	1
139	Solomon Islands	0.1398	14.7782	0.0839	0.0186	15.0205	0.3168	9.1129	0
48	Ethiopia	0.1676	4.4561	0.0076	0.0000	29.7303	0.0990	0.1066	0
22	Burkina Faso	1.3660	7.1846	0.0066	0.0000	29.1784	0.4466	1.4842	0
28	Chad	0.0567	15.4353	0.0425	0.0000	22.6744	0.0709	2.2689	0

170 rows × 32 columns
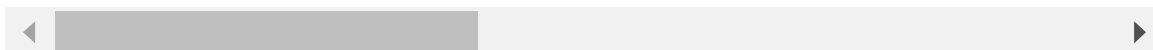
In [60]:

```python
#Plot a histogram showing the Meat consumption vs Vegetables consumption across countrie

# create a figure and axis object
fig, ax = plt.subplots()

# create two histograms on the same axis
ax.hist(df["Meat"], bins=10, alpha=0.5, label="Meat Consumption")
ax.hist(df["Vegetables"], bins=10, alpha=0.5, label="Vegetables Consumption")

# add axis labels and a legend
ax.set_xlabel("Consumption")
ax.set_ylabel("Count")
ax.legend()

# show the plot
plt.show()
```


Conclusion -

In [ ]: