Visualizing Data

November 13, 2021

[13]:
```
import pandas as pd
import matplotlib.pyplot as plt
import dataframe_image as dfi
```

[14]:
```
mergeTable = pd.read_csv('mergedTable.csv')
```

[15]:
```
mergeTable.drop('Unnamed: 0', axis=1, inplace=True)
```

[16]:
```
mergeTable.head()
```

[16]:

	Entity	Year	Animal protein	Plant protein	Total Fat	Total Carbohydrates	Total Protein	Vegetable Oil	Health Care Index	Deaths	VO%
0	France	2007	285.96	156.44	1443.69	1579.91	442.40	507.945205	65.38	112.366845	35.183814
1	France	2008	285.12	165.20	1505.97	1594.71	450.32	517.808219	65.38	110.082662	34.383701
2	France	2009	285.36	159.40	1485.99	1599.25	444.76	530.136986	65.38	107.888487	35.675677
3	France	2010	283.84	168.80	1471.05	1612.31	452.64	520.273973	65.38	104.153629	35.367525
4	France	2011	281.64	165.64	1452.42	1614.30	447.28	517.808219	65.38	101.182683	35.651411

[17]:
```
mt = mergeTable.copy()
```

[18]:
```
mt.head()
```

[18]:

	Entity	Year	Animal protein	Plant protein	Total Fat
0	France	2007	285.96	156.44	1443.69
1	France	2008	285.12	165.20	1505.97

```
2  France  2009          285.36         159.40    1485.99
3  France  2010          283.84         168.80    1471.05
4  France  2011          281.64         165.64    1452.42

   Total Carbohydrates  Total Protein  Vegetable Oil  Health Care Index  \
0              1579.91         442.40     507.945205              65.38
1              1594.71         450.32     517.808219              65.38
2              1599.25         444.76     530.136986              65.38
3              1612.31         452.64     520.273973              65.38
4              1614.30         447.28     517.808219              65.38

       Deaths        VO%
0  112.366845  35.183814
1  110.082662  34.383701
2  107.888487  35.675677
3  104.153629  35.367525
4  101.182683  35.651411
```

Pivoting Data Data: dataframe

Values: values to summarize

Index: Rows

[21]:
```
mt.head()
```

[21]:
```
   Entity  Year  Animal protein  Plant protein  Total Fat  \
0  France  2007          285.96         156.44    1443.69
1  France  2008          285.12         165.20    1505.97
2  France  2009          285.36         159.40    1485.99
3  France  2010          283.84         168.80    1471.05
4  France  2011          281.64         165.64    1452.42

   Total Carbohydrates  Total Protein  Vegetable Oil  Health Care Index  \
0              1579.91         442.40     507.945205              65.38
1              1594.71         450.32     517.808219              65.38
2              1599.25         444.76     530.136986              65.38
3              1612.31         452.64     520.273973              65.38
4              1614.30         447.28     517.808219              65.38

       Deaths        VO%
0  112.366845  35.183814
1  110.082662  34.383701
2  107.888487  35.675677
3  104.153629  35.367525
4  101.182683  35.651411
```

[27]:
```
vog = mt.drop(mt.columns.difference(['Entity', 'Year', 'Vegetable Oil']),␣
↪axis=1)
```

[28]:
```
vog.head()
```

[28]:
```
   Entity  Year  Vegetable Oil
0  France  2007     507.945205
1  France  2008     517.808219
2  France  2009     530.136986
3  France  2010     520.273973
4  France  2011     517.808219
```

[24]:
```
vog = pd.melt(vog, id_vars=['Year'])
```

[25]:
```
vog.head()
```

[25]:
```
   Year variable   value
0  2007   Entity  France
1  2008   Entity  France
2  2009   Entity  France
3  2010   Entity  France
4  2011   Entity  France
```

[9]:
```
vog.reset_index(inplace=True)
```

[31]:
```
vog.head()
```

[31]:
```
   Entity  Year  Vegetable Oil
0  France  2007     507.945205
1  France  2008     517.808219
2  France  2009     530.136986
3  France  2010     520.273973
4  France  2011     517.808219
```

[32]:
```
vcal = vog.pivot(index='Year', columns='Entity', values='Vegetable Oil')
```

Reshaped Data for line plot

[33]:
```
vcal.head()
```

[33]:
```
Entity      France     Germany       Italy  Netherlands      Poland  \
Year
2007    507.945205  451.232877  673.150685   389.589041  276.164384
2008    517.808219  461.095890  675.616438   399.452055  293.424658
2009    530.136986  446.301370  658.356164   355.068493  320.547945
2010    520.273973  433.972603  680.547945   392.054795  290.958904
2011    517.808219  429.041096  690.410959   352.602740  330.410959
```

```
Entity      Russia       Spain  Switzerland      Turkey  United Kingdom
Year
2007    298.356164  660.821918   453.698630  584.383562      414.246575
2008    298.356164  687.945205   485.753425  552.328767      451.232877
2009    305.753425  739.726027   483.287671  576.986301      451.232877
2010    320.547945  764.383562   488.219178  567.123288      461.095890
2011    323.013699  821.095890   495.616438  613.972603      431.506849
```

[34]:
```
vcal.index
```

[34]:
```
Int64Index([2007, 2008, 2009, 2010, 2011], dtype='int64', name='Year')
```

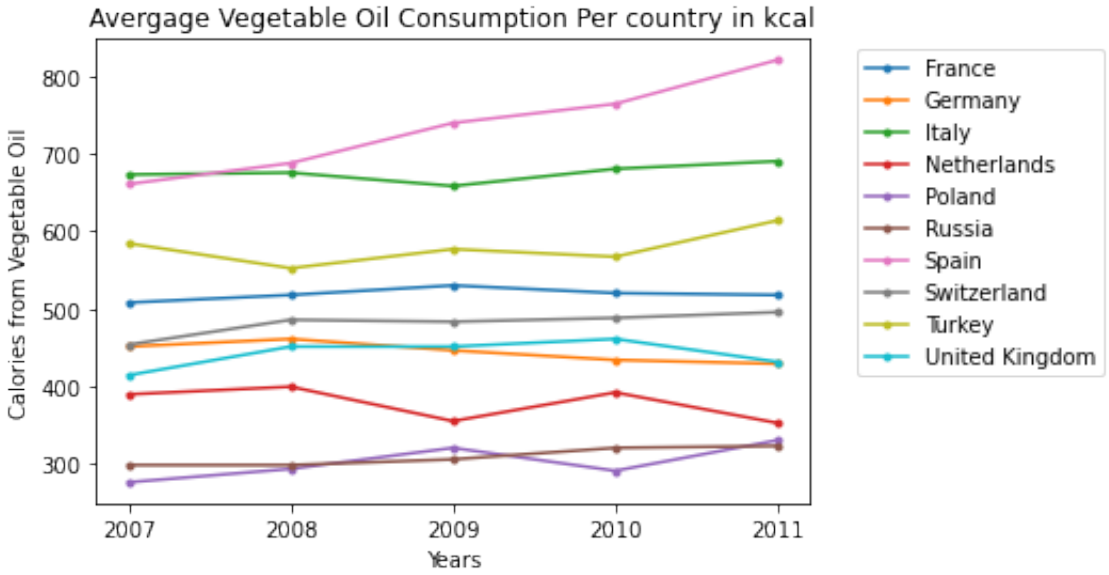
[35]:
```
vcal.columns
```

[35]:
```
Index(['France', 'Germany', 'Italy', 'Netherlands', 'Poland', 'Russia',
       'Spain', 'Switzerland', 'Turkey', 'United Kingdom'],
      dtype='object', name='Entity')
```

[ ]:
```
dfi.export(vcal, 'VOconsumptionPerYear.png')
```

[36]:
```
# plt.figure(figsize=(10,5))

plt.title('Avergage Vegetable Oil Consumption Per country in kcal')
plt.ylabel('Calories from Vegetable Oil')
plt.xlabel('Years')
plt.xticks(vcal.index)

for country in vcal:
    plt.plot(vcal.index, vcal[country], marker='.', label=country)

plt.legend(bbox_to_anchor=(1.05, 1))

# DPI argument means resolution
# plt.savefig('vegetableOilConsumptionPerCountryInKcal.png', dpi=200,␣
 ↪bbox_inches='tight')
```

[36]:
```
<matplotlib.legend.Legend at 0x21edaf57a60>
```


Avergage Vegetable Oil Consumption Per country in kcal
Calories from Vegetable Oil
800
700
600
500
400
300
2007
2008
2009
2010
2011
Years
France
Germany
Italy
Netherlands
Poland
Russia
Spain
Switzerland
Turkey
United Kingdom

[ ]: