

# 7PAM2000 Applied Data Science 2

## Assignment 2

In the following assignment a world bank database has been analysed using pandas dataframe. The Co2 emission of the countries and the GDP of the countries has been accessed through it. The values have been plotted for five countries and the descriptive statistics has been displayed. Through the help of the plots the presence any linear relationship has been accessed in the data.

Countries chosen: DEU','FRA','ESP','GBR','USA  
Germany, France, Spain, United Kingdom, United States.

At first the library has been imported

```
import pandas as pd
import matplotlib.pyplot as plt
import wbapi as wb
```

Then the source info has been obtained with **wb.source.info()**

The emission data has been obtained through **emission = wb.data.DataFrame('EN.ATM.CO2E.KT', wb.region.members('EMU'))**

**emission**

Then the data of the countries and the statistical summary has been obtained through

```
ren = wb.data.DataFrame('EN.ATM.CO2E.KT',
                        ['DEU','FRA','ESP','GBR','USA'],
                        time=range(2000,2020,1))
ren.describe()
```

Out[9]:

	YR2000	YR2001	YR2002	YR2003	YR2004	YR2005	YR2006	YR2007	YR2008	YR2009	YR2010	YR2011	YR2012	YR2013
country														
DEU	83000.0	84700.0	83100.0	83500.0	82000.0	82100.0	81400.0	78300.0	78550.0	75400.0	77200.0	74000.0	70000.0	77700.0
ESP	29310.0	29470.0	34750.0	34990.0	35590.0	35650.0	34700.0	35460.0	32490.0	30740.0	27250.0	27400.0	29670.0	24000.0
FRA	57240.0	57660.0	57100.0	57840.0	57760.0	58020.0	57900.0	58200.0	58200.0	54300.0	53300.0	53500.0	53800.0	53800.0
GBR	53100.0	54480.0	53020.0	54200.0	54300.0	54300.0	54940.0	52910.0	51410.0	46100.0	48100.0	44400.0	46700.0	45300.0
USA	57840.0	57400.0	55410.0	55590.0	57400.0	57600.0	56950.0	57400.0	56910.0	51650.0	52600.0	51700.0	49000.0	50800.0

Out[10]:

	YR2000	YR2001	YR2002	YR2003	YR2004	YR2005	YR2006	YR2007	YR2008	YR2009	YR2010	YR2011	YR2012	YR2013
count	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00
mean	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01
std	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01	2.300000e+01
min	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01
25%	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01
50%	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01
75%	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01	1.800000e+01
max	5.784000e+01	5.740000e+01	5.541000e+01	5.559000e+01	5.740000e+01	5.760000e+01	5.695000e+01	5.740000e+01	5.691000e+01	5.165000e+01	5.260000e+01	5.170000e+01	4.900000e+01	5.080000e+01

The data has been taken over the period of 10 years taking 1 year as a gap.

Similarly the data and the statistical summary has been obtained for GDP

Out[9]:

	YR2000	YR2001	YR2002	YR2003	YR2004	YR2005	YR2006	YR2007	YR2008	YR2009	YR2010	YR2011	YR2012	YR2013
economy														
DEU	1.94145e+12	1.94107e+12	1.93862e+12	1.94013e+12	1.93918e+12	1.94003e+12	1.93910e+12	1.93910e+12	1.93910e+12	1.93910e+12	1.93910e+12	1.93910e+12	1.93910e+12	1.93910e+12
ESP	5.98077e+11	6.27286e+11	1.03304e+12	1.03402e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12	1.03703e+12
FRA	1.35234e+12	1.37465e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12	1.36400e+12
GBR	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12
USA	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12	1.02523e+12

Out[10]:

	YR2000	YR2001	YR2002	YR2003	YR2004	YR2005	YR2006	YR2007	YR2008	YR2009	YR2010	YR2011	YR2012	YR2013
count	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00	5.000000e+00
mean	3.163349e+12	3.234718e+12	3.397760e+12	3.751488e+12	4.125513e+12	4.355337e+12	4.620361e+12	5.185072e+12	5.185072e+12	4.889966e+12	4.889966e+12	4.889966e+12	4.889966e+12	4.889966e+12
std	3.964595e+12	4.136083e+12	4.244820e+12	4.347201e+12	4.567484e+12	4.894818e+12	5.181754e+12	5.323271e+12	5.379390e+12	5.387515e+12	5.387515e+12	5.387515e+12	5.387515e+12	5.387515e+12
min	5.980770e+11	6.272860e+11	7.053043e+11	9.054921e+11	1.067093e+12	1.153286e+12	1.259344e+12	1.472131e+12	1.626225e+12	1.485583e+12	1.485583e+12	1.485583e+12	1.485583e+12	1.485583e+12
25%	1.362249e+12	1.376485e+12	1.494287e+12	1.840481e+12	2.115742e+12	2.196126e+12	2.318594e+12	2.657213e+12	2.918383e+12	2.425798e+12	2.425798e+12	2.425798e+12	2.425798e+12	2.425798e+12
50%	1.662127e+12	1.643908e+12	1.784077e+12	2.057094e+12	2.421814e+12	2.544829e+12	2.717060e+12	3.106182e+12	2.938882e+12	2.690222e+12	2.690222e+12	2.690222e+12	2.690222e+12	2.690222e+12
75%	1.943145e+12	1.944107e+12	2.068624e+12	2.496129e+12	2.809188e+12	2.845803e+12	2.992197e+12	3.421229e+12	3.730028e+12	3.397791e+12	3.397791e+12	3.397791e+12	3.397791e+12	3.397791e+12
max	1.025235e+13	1.058182e+13	1.093842e+13	1.145824e+13	1.221373e+13	1.303964e+13	1.381461e+13	1.445186e+13	1.471284e+13	1.444893e+13	1.444893e+13	1.444893e+13	1.444893e+13	1.444893e+13

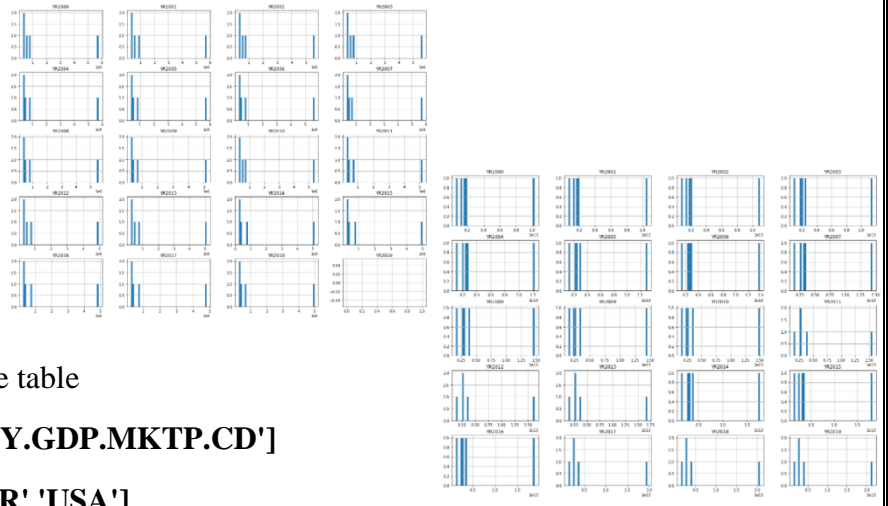
histogram of the countries has been obtained by

```
ren.hist(bins=50, figsize=(20, 15))
plt.savefig('numeric_attributes.png')
```

The

`plt.show()`

`gdp.hist(bins=50, figsize=(20, 15))`  
`plt.savefig('numeric_attributes.png')`  
`plt.show()`



Both the indicators are arranged in a single table

`indicator_ids=['EN.ATM.CO2E.KT','NY.GDP.MKTP.CD']`

`country_code=['DEU','FRA','ESP','GBR','USA']`

`my_dataframe=wb.data.DataFrame(indicator_ids,country_code,mrv=5)`

`my_dataframe`

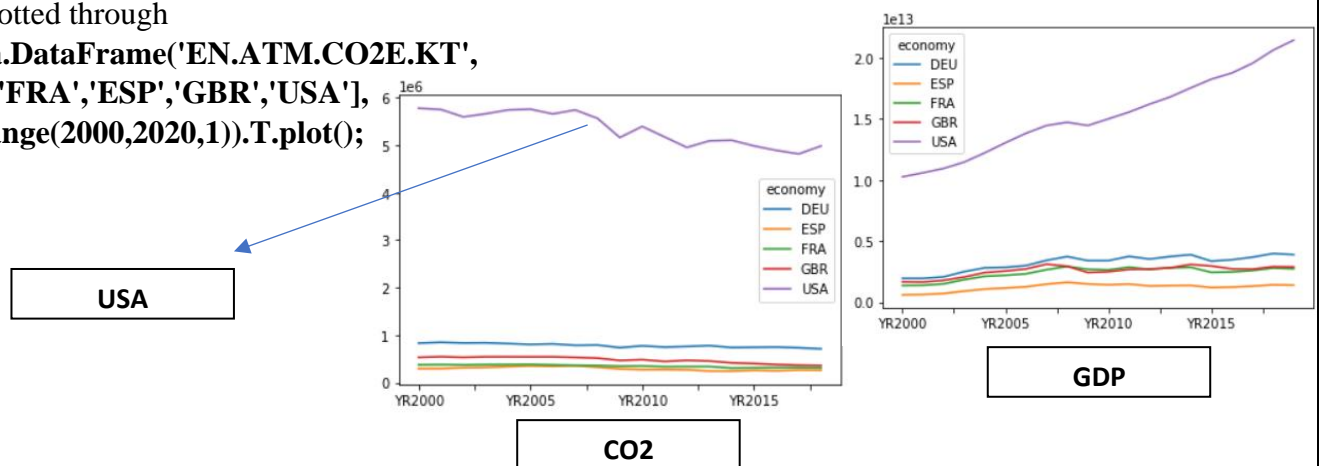
Out[18]:

		YR2016	YR2017	YR2018	YR2019	YR2020
economy	series					
DEU	EN.ATM.CO2E.KT	7.471500e+05	7.322000e+05	7.095400e+05	NaN	NaN
	NY.GDP.MKTP.CD	3.467498e+12	3.681733e+12	3.975347e+12	3.888327e+12	3.846414e+12
ESP	EN.ATM.CO2E.KT	2.470300e+05	2.634400e+05	2.583400e+05	NaN	NaN
	NY.GDP.MKTP.CD	1.232076e+12	1.309297e+12	1.420300e+12	1.393046e+12	1.281485e+12
FRA	EN.ATM.CO2E.KT	3.127100e+05	3.163600e+05	3.099600e+05	NaN	NaN
	NY.GDP.MKTP.CD	2.471286e+12	2.588741e+12	2.789594e+12	2.728870e+12	2.630318e+12
GBR	EN.ATM.CO2E.KT	3.788900e+05	3.663800e+05	3.588000e+05	NaN	NaN
	NY.GDP.MKTP.CD	2.722852e+12	2.699017e+12	2.900791e+12	2.878674e+12	2.759804e+12
USA	EN.ATM.CO2E.KT	4.888640e+06	4.813720e+06	4.981300e+06	NaN	NaN
	NY.GDP.MKTP.CD	1.874508e+13	1.954298e+13	2.061186e+13	2.143322e+13	2.095303e+13

## Correlation:

In order to find the correlation among variables at first the CO2 emission and the GDP of the countries are being plotted through

`wb.data.DataFrame('EN.ATM.CO2E.KT',`  
`['DEU','FRA','ESP','GBR','USA'],`  
`time=range(2000,2020,1)).T.plot();`



A negative correlation can be observed among all the countries. As the GDP of the country of all the countries increases the Co2 emission decreases. It can be evident in the case of USA where there is a significant negative correlation among GDP and CO2 emission. Thus it can be inferred that the country has managed to curb their pollution over time with the increase in GDO. However the emission in USA was already a lot when compared to the emission in the rest of the countries of the world.