7PAM2000 Applied Data Science 2

Assignment 2

In the given report pandas dataframe has been used to access the extensive world Bank database. Wbgapi has been used to call the data and use and access it. In this assignment the arable land data in hectares for which the indicator is AG.LND.ARBL.HA and the total area covered by forest of three countries namely India, USA and Europe has been accessed over the time period of 10 years.

At first the libraries are being called through the code

import pandas as pd
import matplotlib.pyplot as plt
import wbgapi as wb

wb.source.info()
wb.series.info(q='arable')
wb.series.info(q='forest')

The information about the indicators are being collected through these codes.

Arable= wb.data.DataFrame('AG.LND.ARBL.HA', wb.region.members('EMU'))
Forest= wb.data.DataFrame('AG.LND.FRST.K2', wb.region.members('EMU'))

The data has been arranged in a table through the following code

indicator=['AG.LND.ARBL.HA','AG.LND.FRST.K2']

country=['GBR','USA','IND']

my_dataframe=wb.data.DataFrame(indicator,country, mrv=5)

t[18]:

economy	series	YR2016	YR2017	YR2018	YR2019	YR2020
GBR	AG.LND.ARBL.HA	6026010.5	6082879.4	6037787.4	NaN	NaN
	AG.LND.FRST.K2	31590.0	31640.0	31730.0	31820.0	31900.0
IND	AG.LND.ARBL.HA	156416000.0	156416000.0	156416000.0	NaN	NaN
	AG.LND.FRST.K2	710944.0	713608.0	716272.0	718936.0	721600.0
USA	AG.LND.ARBL.HA	157191000.0	157736800.0	157736800.0	NaN	NaN
	AG.LND.FRST.K2	3100950.0	3097950.0	3097950.0	3097950.0	3097950.0

Data retrieval

De = wb.data.DataFrame('AG.LND.FRST.K2', ['GBR','USA','IND'], time=range(2011,2020,1))
De = wb.data.DataFrame('AG.LND.FRST.K2', ['GBR','USA','IND'], time=range(2011,2020,1))

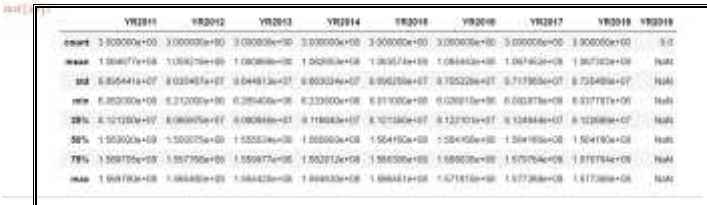
out[20]:

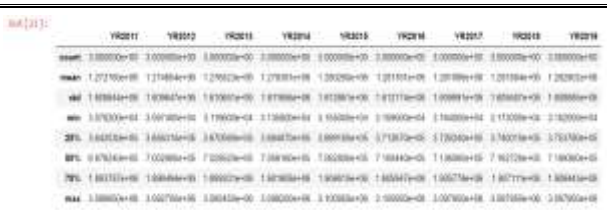
economy	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018	YR2019
GBR	30702.0	30974.0	31196.0	31358.0	31550.0	31590.0	31640.0	31730.0	31820.0
IND	667624.0	700288.0	702952.0	705616.0	708280.0	710944.0	713608.0	716272.0	718936.0
USA	3099950.0	3092700.0	3095450.0	3098200.0	3100950.0	3100950.0	3097950.0	3097950.0	3097950.0

out[21]:

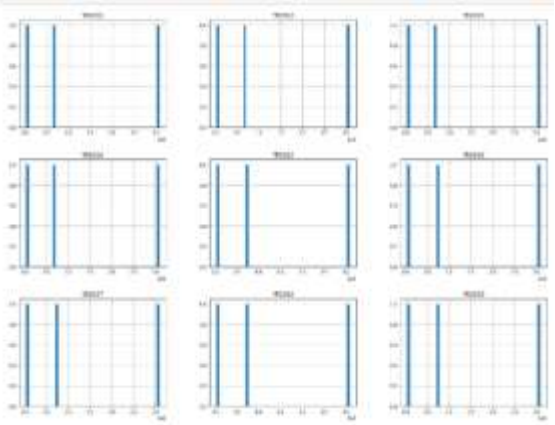
economy	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018	YR2019
GBR	6082879.4	6037787.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN
IND	156416000.0	156416000.0	156416000.0	156416000.0	156416000.0	156416000.0	156416000.0	156416000.0	NaN
USA	157736800.0	157736800.0	157736800.0	157736800.0	157736800.0	157736800.0	157736800.0	157736800.0	NaN

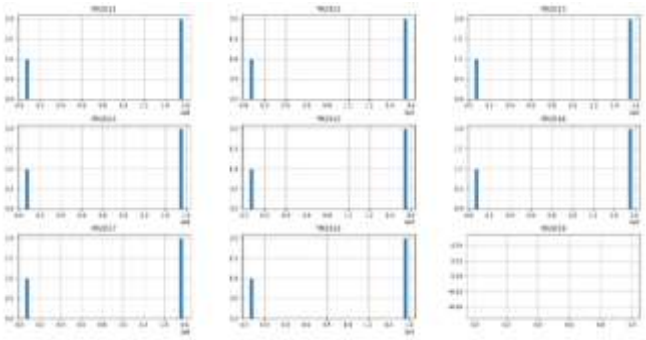
De.describe()
ar.describe()





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





Plot
wb.data.DataFrame('AG.LND.ARBL.HA',

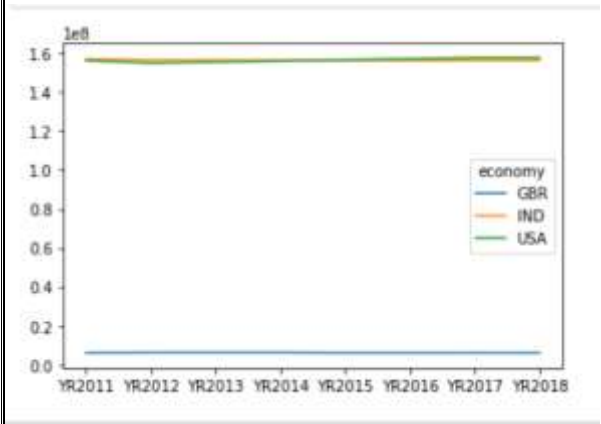
['GBR','USA','IND'],

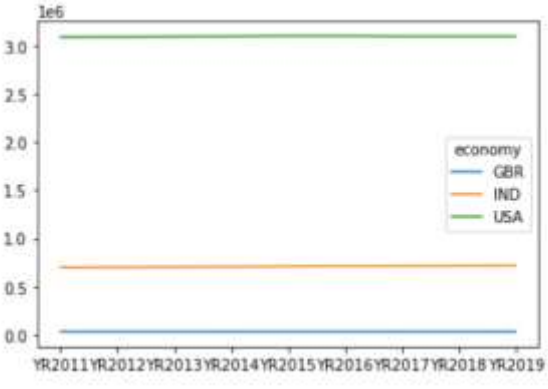
time=range(2011,2020,1)).T.plot();

wb.data.DataFrame('AG.LND.FRST.K2',

['GBR','USA','IND'],

time=range(2011,2020,1)).T.plot();





Arable land

Forest covered

From the above figure it can be observed that for INDIA the forest covered land has not changed much over the period of 10 years. It has not changed for all the three countries.