[Edx] Mini_Porject

August 12, 2019

1 Let's compare KOREA to JAPAN with Economy Indices.

```
[71]: %matplotlib inline
      import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
      import re
      plt.style.use('ggplot')
```

1.1 Add data with newest dataset (WDI 2019)

now_data is given data by edX
new_data is newest data from worldbank
       However, Format of the dataset's table is differnt.
so It was difficult to join two table simply.
Let's check how differnt two tables.

```
[72]: path = '../Week-5-Visualization/world-development-indicators/Indicators.csv'
      now_data = pd.read_csv(path)
```

```
[73]: p = './WDI_csv/WDIData.csv'
      new_data = pd.read_csv(p)
```

```
[74]: now_data.head()
```

```
[74]:   CountryName CountryCode                                 IndicatorName  \
      0   Arab World         ARB  Adolescent fertility rate (births per 1,000 wo...
      1   Arab World         ARB  Age dependency ratio (% of working-age populat...
      2   Arab World         ARB  Age dependency ratio, old (% of working-age po...
      3   Arab World         ARB  Age dependency ratio, young (% of working-age ...
      4   Arab World         ARB       Arms exports (SIPRI trend indicator values)

          IndicatorCode  Year         Value
      0     SP.ADO.TFRT  1960  1.335609e+02
      1     SP.POP.DPND  1960  8.779760e+01
      2  SP.POP.DPND.OL  1960  6.634579e+00
      3  SP.POP.DPND.YG  1960  8.102333e+01
      4  MS.MIL.XPRT.KD  1960  3.000000e+06
```

As you can see now_data looks like above.

Otherwise, new_data does not have Year field, but it has many fields by each years like below.

```
[75]: new_data.head()
```

```
[75]:   Country Name Country Code  \
     0   Arab World          ARB
     1   Arab World          ARB
     2   Arab World          ARB
     3   Arab World          ARB
     4   Arab World          ARB


                                        Indicator Name     Indicator Code  1960  \
     0  2005 PPP conversion factor, GDP (LCU per inter...      PA.NUS.PPP.05   NaN
     1  2005 PPP conversion factor, private consumptio...   PA.NUS.PRVT.PP.05  NaN
     2  Access to clean fuels and technologies for coo...      EG.CFT.ACCS.ZS   NaN
     3            Access to electricity (% of population)     EG.ELC.ACCS.ZS   NaN
     4  Access to electricity, rural (% of rural popul...   EG.ELC.ACCS.RU.ZS  NaN

         1961  1962  1963  1964  1965  ...        2010       2011       2012  \
     0    NaN   NaN   NaN   NaN   NaN  ...        NaN        NaN        NaN
     1    NaN   NaN   NaN   NaN   NaN  ...        NaN        NaN        NaN
     2    NaN   NaN   NaN   NaN   NaN  ...  82.368101  82.783289  83.120303
     3    NaN   NaN   NaN   NaN   NaN  ...  86.007620  86.428272  87.070576
     4    NaN   NaN   NaN   NaN   NaN  ...  73.466653  73.942103  75.244104

             2013       2014       2015       2016       2017  2018  Unnamed: 63
     0        NaN        NaN        NaN        NaN        NaN   NaN          NaN
     1        NaN        NaN        NaN        NaN        NaN   NaN          NaN
     2  83.533457  83.897596  84.171599  84.510171        NaN   NaN          NaN
     3  88.176836  87.342739  89.130121  89.678685  90.273687   NaN          NaN
     4  77.162305  75.538976  78.741152  79.665635  80.749293   NaN          NaN

     [5 rows x 64 columns]
```

This is how i add new datas.

```
[76]: def add_data(now_data,new_data):
          dict_list =[]
          for i in picked_indicators_code :
              nowdata = now_data[now_data['IndicatorCode'] == i]
              add_year = nowdata['Year'].max() + 1
              last = len(now_data)
              while add_year < 2019 :
                  filter1 = new_data['Indicator Code'] == i
                  filter2 = new_data['Country Name'].isin(now_data['CountryName'])
                  target = new_data[filter1 & filter2]
                  theyear = int(add_year)
                  try:
                      added_data ={
```

```
                'CountryName' : target['Country Name'].values[0],
                'CountryCode': target['Country Code'].values[0],
                'IndicatorName':target['Indicator Name'].values[0],
                'IndicatorCode': target['Indicator Code'].values[0],
                'Year':theyear,
                'Value':target[str(add_year)].values[0]
            }
        except:
            added_data ={
                'CountryName' : None,
                'CountryCode': None,
                'IndicatorName':None,
                'IndicatorCode': None,
                'Year':None,
                'Value':None
            }
        dict_list.append(added_data)
        add_year = add_year + 1
    now_data = now_data.append(dict_list,ignore_index=True)
    return now_data
```

1.2 Picking the correct indicators to explore

Need to explorer only a particular area of interest …

The database is so rich with so many indicators that it is desirable to have a better way of picking required indicators.

So I have created a new Indicator list using which specific topics like for eg: Health, Food, Energy etc. can be searched for. Then within each topic required indicators can be more easily picked.

Change IndicatorNames Create a new list of IndicatorName and IndicatorCode such that special characters like "(", ")", "," are replaced just by spaces This new list (modified_indicators) can be used to search for specific topics as done below

```
[77]: # Create list of unique indicators, indicator codes
      Indicator_array =  df[['IndicatorName','IndicatorCode']].drop_duplicates().
      ↪values
      Indicator_array
```

```
[77]: array([['Adolescent fertility rate (births per 1,000 women ages 15-19)',
              'SP.ADO.TFRT'],
             ['Age dependency ratio (% of working-age population)',
              'SP.POP.DPND'],
             ['Age dependency ratio, old (% of working-age population)',
              'SP.POP.DPND.OL'],
             ...,
             ['Fish species, threatened', 'EN.FSH.THRD.NO'],
             ['Mammal species, threatened', 'EN.MAM.THRD.NO'],
             ['Plant species (higher), threatened', 'EN.HPT.THRD.NO']],
            dtype=object)
```

```
[78]: modified_indicators = []
      unique_indicator_codes = []
      for ele in Indicator_array:
          indicator = ele[0]
          indicator_code = ele[1].strip()
          if indicator_code not in unique_indicator_codes:
              # delete , ( ) from the IndicatorNames
              new_indicator = re.sub('[,()]',"",indicator).lower()
              # replace - with "to" and make all words into lower case
              new_indicator = re.sub('-'," to ",new_indicator).lower()
              modified_indicators.append([new_indicator,indicator_code])
              unique_indicator_codes.append(indicator_code)

      Indicators = pd.
       →DataFrame(modified_indicators,columns=['IndicatorName','IndicatorCode'])
      Indicators = Indicators.drop_duplicates()
      Indicators.shape
```

[78]: (1344, 2)

```
[79]: # dict ,   .
      key_word_dict = {}
      key_word_dict['Demography'] =␣
       →['population','birth','death','fertility','mortality','expectancy']
      key_word_dict['Food'] = ['food','grain','nutrition','calories']
      key_word_dict['Trade'] =␣
       →['trade','import','export','good','shipping','shipment']
      key_word_dict['Health'] = ['health','desease','hospital','mortality','doctor']
      key_word_dict['Economy'] =␣
       →['GNI','income','gdp','gni','deficit','budget','market','stock','bond','infrastructure',␣
       →'saving', 'household', 'debt']
      key_word_dict['Energy'] =␣
       →['fuel','energy','power','emission','electric','electricity']
      key_word_dict['Education'] = ['education','literacy']
      key_word_dict['Employment']␣
       →=['employed','employment','umemployed','unemployment','female']
      key_word_dict['Rural'] = ['rural','village']
      key_word_dict['Urban'] = ['urban','city']
      key_word_dict['Tech'] = ['technology','research','intellectual','r&d']
```

1.2.1 Pick required fields

Now within specific topics we cah chose what ever indicators we are interested in

```
[80]: #    (  )
      feature = 'Economy'
      for indicator_ele in Indicators.values:
          for ele in key_word_dict[feature]:
```

```
        word_list = indicator_ele[0].split()
        if ele in word_list or ele+'s' in word_list:
            # Uncomment this line to print the indicators explicitely
            print(indicator_ele)
            break
```

```
['international migrant stock % of population' 'SM.POP.TOTL.ZS']
['international migrant stock total' 'SM.POP.TOTL']
['merchandise exports to high to income economies % of total merchandise
exports'
 'TX.VAL.MRCH.HI.ZS']
['merchandise imports from high to income economies % of total merchandise
imports'
 'TM.VAL.MRCH.HI.ZS']
['merchandise trade % of gdp' 'TG.VAL.TOTL.GD.ZS']
['gdp at market prices current us$' 'NY.GDP.MKTP.CD']
['gdp per capita current us$' 'NY.GDP.PCAP.CD']
['gni current us$' 'NY.GNP.MKTP.CD']
['net oda received % of gni' 'DT.ODA.ODAT.GN.ZS']
['co2 emissions kg per 2005 us$ of gdp' 'EN.ATM.CO2E.KD.GD']
['exports of goods and services % of gdp' 'NE.EXP.GNFS.ZS']
['external balance on goods and services % of gdp' 'NE.RSB.GNFS.ZS']
['gdp at market prices constant 2005 us$' 'NY.GDP.MKTP.KD']
['gdp per capita constant 2005 us$' 'NY.GDP.PCAP.KD']
['general government final consumption expenditure % of gdp'
 'NE.CON.GOVT.ZS']
['gni constant 2005 us$' 'NY.GNP.MKTP.KD']
['gni per capita constant 2005 us$' 'NY.GNP.PCAP.KD']
['gross domestic income constant 2005 us$' 'NY.GDY.TOTL.KD']
['gross fixed capital formation % of gdp' 'NE.GDI.FTOT.ZS']
['imports of goods and services % of gdp' 'NE.IMP.GNFS.ZS']
['trade % of gdp' 'NE.TRD.GNFS.ZS']
['agriculture value added % of gdp' 'NV.AGR.TOTL.ZS']
['gross capital formation % of gdp' 'NE.GDI.TOTL.ZS']
['gross domestic savings % of gdp' 'NY.GDS.TOTL.ZS']
['gross domestic savings current us$' 'NY.GDS.TOTL.CD']
['household final consumption expenditure current us$' 'NE.CON.PRVT.CD']
['household final consumption expenditure etc. % of gdp' 'NE.CON.PETC.ZS']
['household final consumption expenditure etc. current us$'
 'NE.CON.PETC.CD']
['industry value added % of gdp' 'NV.IND.TOTL.ZS']
['manufacturing value added % of gdp' 'NV.IND.MANF.ZS']
['services etc. value added % of gdp' 'NV.SRV.TETC.ZS']
['household final consumption expenditure constant 2005 us$'
 'NE.CON.PRVT.KD']
['household final consumption expenditure per capita constant 2005 us$'
 'NE.CON.PRVT.PC.KD']
```

```
['household final consumption expenditure etc. constant 2005 us$'
 'NE.CON.PETC.KD']
['gross fixed capital formation private sector % of gdp' 'NE.GDI.FPRV.ZS']
['final consumption expenditure etc. % of gdp' 'NE.CON.TETC.ZS']
['gdp current lcu' 'NY.GDP.MKTP.CN']
['gdp per capita current lcu' 'NY.GDP.PCAP.CN']
['gni current lcu' 'NY.GNP.MKTP.CN']
['gni per capita current lcu' 'NY.GNP.PCAP.CN']
['gross domestic savings current lcu' 'NY.GDS.TOTL.CN']
['gross national expenditure % of gdp' 'NE.DAB.TOTL.ZS']
['household final consumption expenditure current lcu' 'NE.CON.PRVT.CN']
['net income from abroad current lcu' 'NY.GSR.NFCY.CN']
['net income from abroad current us$' 'NY.GSR.NFCY.CD']
['discrepancy in expenditure estimate of gdp constant lcu'
 'NY.GDP.DISC.KN']
['discrepancy in expenditure estimate of gdp current lcu' 'NY.GDP.DISC.CN']
['gdp constant lcu' 'NY.GDP.MKTP.KN']
['gdp deflator base year varies by country' 'NY.GDP.DEFL.ZS']
['gdp per capita constant lcu' 'NY.GDP.PCAP.KN']
['gni constant lcu' 'NY.GNP.MKTP.KN']
['gni per capita constant lcu' 'NY.GNP.PCAP.KN']
['gross domestic income constant lcu' 'NY.GDY.TOTL.KN']
['gross domestic savings constant lcu' 'NY.GDS.TOTL.KN']
['household final consumption expenditure constant lcu' 'NE.CON.PRVT.KN']
['household final consumption expenditure etc. constant lcu'
 'NE.CON.PETC.KN']
['household final consumption expenditure etc. current lcu'
 'NE.CON.PETC.CN']
['adjusted savings: gross savings % of gni' 'NY.ADJ.ICTR.GN.ZS']
['gross savings % of gdp' 'NY.GNS.ICTR.ZS']
['gross savings % of gni' 'NY.GNS.ICTR.GN.ZS']
['gross savings current lcu' 'NY.GNS.ICTR.CN']
['gross savings current us$' 'NY.GNS.ICTR.CD']
['net secondary income bop current us$' 'BN.TRF.CURR.CD']
['net income from abroad constant lcu' 'NY.GSR.NFCY.KN']
['gdp growth annual %' 'NY.GDP.MKTP.KD.ZG']
['gdp per capita growth annual %' 'NY.GDP.PCAP.KD.ZG']
['gni growth annual %' 'NY.GNP.MKTP.KD.ZG']
['gni per capita growth annual %' 'NY.GNP.PCAP.KD.ZG']
['inflation gdp deflator annual %' 'NY.GDP.DEFL.KD.ZG']
['household final consumption expenditure annual % growth'
 'NE.CON.PRVT.KD.ZG']
['household final consumption expenditure per capita growth annual %'
 'NE.CON.PRVT.PC.KD.ZG']
['household final consumption expenditure etc. annual % growth'
 'NE.CON.PETC.KD.ZG']
['gni per capita atlas method current us$' 'NY.GNP.PCAP.CD']
['gni atlas method current us$' 'NY.GNP.ATLS.CD']
```

['water productivity total constant 2005 us$ gdp per cubic meter of total
freshwater withdrawal'
 'ER.GDP.FWTL.M3.KD']
['adjusted savings: carbon dioxide damage % of gni' 'NY.ADJ.DCO2.GN.ZS']
['adjusted savings: consumption of fixed capital % of gni'
 'NY.ADJ.DKAP.GN.ZS']
['adjusted savings: education expenditure % of gni' 'NY.ADJ.AEDU.GN.ZS']
['adjusted savings: energy depletion % of gni' 'NY.ADJ.DNGY.GN.ZS']
['adjusted savings: mineral depletion % of gni' 'NY.ADJ.DMIN.GN.ZS']
['adjusted savings: natural resources depletion % of gni'
 'NY.ADJ.DRES.GN.ZS']
['adjusted savings: net forest depletion % of gni' 'NY.ADJ.DFOR.GN.ZS']
['coal rents % of gdp' 'NY.GDP.COAL.RT.ZS']
['foreign direct investment net inflows % of gdp' 'BX.KLT.DINV.WD.GD.ZS']
['forest rents % of gdp' 'NY.GDP.FRST.RT.ZS']
['mineral rents % of gdp' 'NY.GDP.MINR.RT.ZS']
['natural gas rents % of gdp' 'NY.GDP.NGAS.RT.ZS']
['oil rents % of gdp' 'NY.GDP.PETR.RT.ZS']
['public and publicly guaranteed debt service % of gni'
 'DT.TDS.DPPG.GN.ZS']
['total debt service % of gni' 'DT.TDS.DECT.GN.ZS']
['total natural resources rents % of gdp' 'NY.GDP.TOTL.RT.ZS']
['adjusted net national income current us$' 'NY.ADJ.NNTY.CD']
['adjusted net national income per capita current us$' 'NY.ADJ.NNTY.PC.CD']
['adjusted net national income constant 2005 us$' 'NY.ADJ.NNTY.KD']
['adjusted net national income per capita constant 2005 us$'
 'NY.ADJ.NNTY.PC.KD']
['average grace period on new external debt commitments years'
 'DT.GPA.DPPG']
['average grace period on new external debt commitments official years'
 'DT.GPA.OFFT']
['average grace period on new external debt commitments private years'
 'DT.GPA.PRVT']
['average grant element on new external debt commitments %' 'DT.GRE.DPPG']
['average grant element on new external debt commitments official %'
 'DT.GRE.OFFT']
['average grant element on new external debt commitments private %'
 'DT.GRE.PRVT']
['average interest on new external debt commitments %' 'DT.INR.DPPG']
['average interest on new external debt commitments official %'
 'DT.INR.OFFT']
['average interest on new external debt commitments private %'
 'DT.INR.PRVT']
['average maturity on new external debt commitments years' 'DT.MAT.DPPG']
['average maturity on new external debt commitments official years'
 'DT.MAT.OFFT']
['average maturity on new external debt commitments private years'
 'DT.MAT.PRVT']

['concessional debt % of total external debt' 'DT.DOD.ALLC.ZS']
['currency composition of ppg debt all other currencies %'
 'DT.CUR.OTHC.ZS']
['currency composition of ppg debt deutsche mark %' 'DT.CUR.DMAK.ZS']
['currency composition of ppg debt french franc %' 'DT.CUR.FFRC.ZS']
['currency composition of ppg debt japanese yen %' 'DT.CUR.JYEN.ZS']
['currency composition of ppg debt multiple currencies %' 'DT.CUR.MULC.ZS']
['currency composition of ppg debt pound sterling %' 'DT.CUR.UKPS.ZS']
['currency composition of ppg debt sdr %' 'DT.CUR.SDRW.ZS']
['currency composition of ppg debt swiss franc %' 'DT.CUR.SWFR.ZS']
['currency composition of ppg debt u.s. dollars %' 'DT.CUR.USDL.ZS']
['debt forgiveness grants current us$' 'DT.DOD.MDRI.CD']
['debt service on external debt long to term tds current us$'
 'DT.TDS.DLXF.CD']
['debt service on external debt private nonguaranteed png tds current us$'
 'DT.TDS.DPNG.CD']
['debt service on external debt public and publicly guaranteed ppg tds current
us$'
 'DT.TDS.DPPG.CD']
['debt service on external debt total tds current us$' 'DT.TDS.DECT.CD']
['disbursements on external debt long to term dis current us$'
 'DT.DIS.DLXF.CD']
['disbursements on external debt long to term + imf dis current us$'
 'DT.DIS.DLTF.CD']
['disbursements on external debt private nonguaranteed png dis current us$'
 'DT.DIS.DPNG.CD']
['disbursements on external debt public and publicly guaranteed ppg dis current
us$'
 'DT.DIS.DPPG.CD']
['external debt stocks % of gni' 'DT.DOD.DECT.GN.ZS']
['external debt stocks concessional dod current us$' 'DT.DOD.ALLC.CD']
['external debt stocks long to term dod current us$' 'DT.DOD.DLXF.CD']
['external debt stocks long to term private sector dod current us$'
 'DT.DOD.PRVS.CD']
['external debt stocks long to term public sector dod current us$'
 'DT.DOD.PUBS.CD']
['external debt stocks private nonguaranteed png dod current us$'
 'DT.DOD.DPNG.CD']
['external debt stocks public and publicly guaranteed ppg dod current us$'
 'DT.DOD.DPPG.CD']
['external debt stocks short to term dod current us$' 'DT.DOD.DSTC.CD']
['external debt stocks total dod current us$' 'DT.DOD.DECT.CD']
['external debt stocks variable rate dod current us$' 'DT.DOD.VTOT.CD']
['interest payments on external debt % of gni' 'DT.INT.DECT.GN.ZS']
['interest payments on external debt long to term int current us$'
 'DT.INT.DLXF.CD']
['interest payments on external debt private nonguaranteed png int current us$'
 'DT.INT.DPNG.CD']

```
['interest payments on external debt public and publicly guaranteed ppg int
current us$'
 'DT.INT.DPPG.CD']
['interest payments on external debt short to term int current us$'
 'DT.INT.DSTC.CD']
['interest payments on external debt total int current us$'
 'DT.INT.DECT.CD']
['multilateral debt % of total external debt' 'DT.DOD.MLAT.ZS']
['multilateral debt service % of public and publicly guaranteed debt service'
 'DT.TDS.MLAT.PG.ZS']
['multilateral debt service tds current us$' 'DT.TDS.MLAT.CD']
['net flows on external debt long to term nfl current us$'
 'DT.NFL.DLXF.CD']
['net flows on external debt private nonguaranteed png nfl current us$'
 'DT.NFL.DPNG.CD']
['net flows on external debt public and publicly guaranteed ppg nfl current us$'
 'DT.NFL.DPPG.CD']
['net flows on external debt short to term nfl current us$'
 'DT.NFL.DSTC.CD']
['net flows on external debt total nfl current us$' 'DT.NFL.DECT.CD']
['net transfers on external debt long to term ntr current us$'
 'DT.NTR.DLXF.CD']
['net transfers on external debt private nonguaranteed png ntr current us$'
 'DT.NTR.DPNG.CD']
['net transfers on external debt public and publicly guaranteed ppg ntr current
us$'
 'DT.NTR.DPPG.CD']
['net transfers on external debt total ntr current us$' 'DT.NTR.DECT.CD']
['png bonds amt current us$' 'DT.AMT.PNGB.CD']
['png bonds dis current us$' 'DT.DIS.PNGB.CD']
['png bonds dod current us$' 'DT.DOD.PNGB.CD']
['png bonds int current us$' 'DT.INT.PNGB.CD']
['png bonds nfl current us$' 'DT.NFL.PNGB.CD']
['png bonds ntr current us$' 'DT.NTR.PNGB.CD']
['png bonds tds current us$' 'DT.TDS.PNGB.CD']
['portfolio investment bonds ppg + png nfl current us$' 'DT.NFL.BOND.CD']
['ppg bonds amt current us$' 'DT.AMT.PBND.CD']
['ppg bonds dis current us$' 'DT.DIS.PBND.CD']
['ppg bonds dod current us$' 'DT.DOD.PBND.CD']
['ppg bonds int current us$' 'DT.INT.PBND.CD']
['ppg bonds nfl current us$' 'DT.NFL.PBND.CD']
['ppg bonds ntr current us$' 'DT.NTR.PBND.CD']
['ppg bonds tds current us$' 'DT.TDS.PBND.CD']
['primary income on fdi payments current us$' 'BX.KLT.DREM.CD.DT']
['principal repayments on external debt long to term amt current us$'
 'DT.AMT.DLXF.CD']
['principal repayments on external debt long to term + imf amt current us$'
 'DT.AMT.DLTF.CD']
```

['principal repayments on external debt private nonguaranteed png amt current us$'
 'DT.AMT.DPNG.CD']
['principal repayments on external debt public and publicly guaranteed ppg amt current us$'
 'DT.AMT.DPPG.CD']
['short to term debt % of total external debt' 'DT.DOD.DSTC.ZS']
['short to term debt % of total reserves' 'DT.DOD.DSTC.IR.ZS']
['undisbursed external debt official creditors und current us$'
 'DT.UND.OFFT.CD']
['undisbursed external debt private creditors und current us$'
 'DT.UND.PRVT.CD']
['undisbursed external debt total und current us$' 'DT.UND.DPPG.CD']
['adjusted net savings excluding particulate emission damage % of gni'
 'NY.ADJ.SVNX.GN.ZS']
['adjusted savings: net national savings % of gni' 'NY.ADJ.NNAT.GN.ZS']
['personal remittances received % of gdp' 'BX.TRF.PWKR.DT.GD.ZS']
['government expenditure on education as % of gdp %' 'SE.XPD.TOTL.GD.ZS']
['adjusted net savings excluding particulate emission damage current us$'
 'NY.ADJ.SVNX.CD']
['adjusted savings: net national savings current us$' 'NY.ADJ.NNAT.CD']
['debt service ppg and imf only % of exports of goods services and primary income'
 'DT.TDS.DPPF.XP.ZS']
['external debt stocks % of exports of goods services and primary income'
 'DT.DOD.DECT.EX.ZS']
['interest payments on external debt % of exports of goods services and primary income'
 'DT.INT.DECT.EX.ZS']
['total debt service % of exports of goods services and primary income'
 'DT.TDS.DECT.EX.ZS']
['total reserves % of total external debt' 'FI.RES.TOTL.DT.ZS']
['adjusted net national income annual % growth' 'NY.ADJ.NNTY.KD.ZG']
['adjusted net national income per capita annual % growth'
 'NY.ADJ.NNTY.PC.KD.ZG']
['government expenditure per secondary student as % of gdp per capita %'
 'SE.XPD.SECO.PC.ZS']
['government expenditure per primary student as % of gdp per capita %'
 'SE.XPD.PRIM.PC.ZS']
['market capitalization of listed domestic companies % of gdp'
 'CM.MKT.LCAP.GD.ZS']
['market capitalization of listed domestic companies current us$'
 'CM.MKT.LCAP.CD']
['stocks traded total value % of gdp' 'CM.MKT.TRAD.GD.ZS']
['stocks traded total value current us$' 'CM.MKT.TRAD.CD']
['foreign direct investment net outflows % of gdp' 'BM.KLT.DINV.GD.ZS']
['stocks traded turnover ratio of domestic shares %' 'CM.MKT.TRNR']
['gdp per person employed constant 1990 ppp $' 'SL.GDP.PCAP.EM.KD']

```
['income share held by fourth 20%' 'SI.DST.04TH.20']
['income share held by highest 10%' 'SI.DST.10TH.10']
['income share held by highest 20%' 'SI.DST.05TH.20']
['income share held by lowest 10%' 'SI.DST.FRST.10']
['income share held by lowest 20%' 'SI.DST.FRST.20']
['income share held by second 20%' 'SI.DST.02ND.20']
['income share held by third 20%' 'SI.DST.03RD.20']
['government expenditure per tertiary student as % of gdp per capita %'
 'SE.XPD.TERT.PC.ZS']
['military expenditure % of gdp' 'MS.MIL.XPND.GD.ZS']
['debt buyback current us$' 'DT.DSB.DPPG.CD']
['debt forgiveness or reduction current us$' 'DT.DFR.DPPG.CD']
['debt stock reduction current us$' 'DT.DSF.DPPG.CD']
['debt stock rescheduled current us$' 'DT.DXR.DPPG.CD']
['residual debt stock to flow reconciliation current us$' 'DT.DOD.RSDL.CD']
['total amount of debt rescheduled current us$' 'DT.TXR.DPPG.CD']
['total change in external debt stocks current us$' 'DT.DOD.DECT.CD.CG']
['adjusted net savings including particulate emission damage % of gni'
 'NY.ADJ.SVNG.GN.ZS']
['adjusted savings: particulate emission damage % of gni'
 'NY.ADJ.DPEM.GN.ZS']
['co2 emissions kg per 2011 ppp $ of gdp' 'EN.ATM.CO2E.PP.GD.KD']
['co2 emissions kg per ppp $ of gdp' 'EN.ATM.CO2E.PP.GD']
['energy intensity level of primary energy mj/$2011 ppp gdp'
 'EG.EGY.PRIM.PP.KD']
['energy use kg of oil equivalent per $1000 gdp constant 2011 ppp'
 'EG.USE.COMM.GD.PP.KD']
['gdp per capita ppp constant 2011 international $' 'NY.GDP.PCAP.PP.KD']
['gdp per capita ppp current international $' 'NY.GDP.PCAP.PP.CD']
['gdp per unit of energy use constant 2011 ppp $ per kg of oil equivalent'
 'EG.GDP.PUSE.KO.PP.KD']
['gdp per unit of energy use ppp $ per kg of oil equivalent'
 'EG.GDP.PUSE.KO.PP']
['gdp ppp constant 2011 international $' 'NY.GDP.MKTP.PP.KD']
['gdp ppp current international $' 'NY.GDP.MKTP.PP.CD']
['gni per capita ppp current international $' 'NY.GNP.PCAP.PP.CD']
['gni ppp current international $' 'NY.GNP.MKTP.PP.CD']
['household final consumption expenditure ppp current international $'
 'NE.CON.PRVT.PP.CD']
['household final consumption expenditure ppp constant 2011 international $'
 'NE.CON.PRVT.PP.KD']
['gni per capita ppp constant 2011 international $' 'NY.GNP.PCAP.PP.KD']
['gni ppp constant 2011 international $' 'NY.GNP.MKTP.PP.KD']
['cash surplus/deficit % of gdp' 'GC.BAL.CASH.GD.ZS']
['expense % of gdp' 'GC.XPN.TOTL.GD.ZS']
['revenue excluding grants % of gdp' 'GC.REV.XGRT.GD.ZS']
['tax revenue % of gdp' 'GC.TAX.TOTL.GD.ZS']
['taxes on income profits and capital gains % of revenue'
```

```
'GC.TAX.YPKG.RV.ZS']
['central government debt total % of gdp' 'GC.DOD.TOTL.GD.ZS']
['net incurrence of liabilities domestic % of gdp' 'GC.FIN.DOMS.GD.ZS']
['net incurrence of liabilities foreign % of gdp' 'GC.FIN.FRGN.GD.ZS']
['adjusted net savings including particulate emission damage current us$'
 'NY.ADJ.SVNG.CD']
['ppp conversion factor gdp lcu per international $' 'PA.NUS.PPP']
['price level ratio of ppp conversion factor gdp to market exchange rate'
 'PA.NUS.PPPC.RF']
['central government debt total current lcu' 'GC.DOD.TOTL.CN']
['taxes on income profits and capital gains % of total taxes'
 'GC.TAX.YPKG.ZS']
['taxes on income profits and capital gains current lcu' 'GC.TAX.YPKG.CN']
['female headed households % of households with a female head'
 'SP.HOU.FEMA.ZS']
['depth of the food deficit kilocalories per person per day' 'SN.ITK.DFCT']
['consumption of iodized salt % of households' 'SN.ITK.SALT.ZS']
['health expenditure private % of gdp' 'SH.XPD.PRIV.ZS']
['health expenditure public % of gdp' 'SH.XPD.PUBL.ZS']
['health expenditure total % of gdp' 'SH.XPD.TOTL.ZS']
['research and development expenditure % of gdp' 'GB.XPD.RSDV.GD.ZS']
['broad money % of gdp' 'FM.LBL.BMNY.GD.ZS']
['claims on central government etc. % gdp' 'FS.AST.CGOV.GD.ZS']
['claims on other sectors of the domestic economy % of gdp'
 'FS.AST.DOMO.GD.ZS']
['domestic credit provided by financial sector % of gdp'
 'FS.AST.DOMS.GD.ZS']
['domestic credit to private sector % of gdp' 'FS.AST.PRVT.GD.ZS']
['domestic credit to private sector by banks % of gdp' 'FD.AST.PRVT.GD.ZS']
['liquid liabilities m3 as % of gdp' 'FS.LBL.LIQU.GD.ZS']
['money and quasi money m2 as % of gdp' 'FM.LBL.MQMY.GD.ZS']
['quasi to liquid liabilities % of gdp' 'FS.LBL.QLIQ.GD.ZS']
['currency composition of ppg debt euro %' 'DT.CUR.EURO.ZS']
['cost of business start to up procedures % of gni per capita'
 'IC.REG.COST.PC.ZS']
['survey mean consumption or income per capita bottom 40% of population 2011 ppp
$ per day'
 'SI.SPR.PC40']
['survey mean consumption or income per capita total population 2011 ppp $ per
day'
 'SI.SPR.PCAP']
['adequacy of social insurance programs % of total welfare of beneficiary
households'
 'per_si_allsi.adq_pop_tot']
['adequacy of social protection and labor programs % of total welfare of
beneficiary households'
 'per_allsp.adq_pop_tot']
['adequacy of unemployment benefits and almp % of total welfare of beneficiary
```

households'
 'per_lm_alllm.adq_pop_tot']
['benefits incidence in poorest quintile % to all labor market'
 'per_lm_alllm.ben_q1_tot']
['coverage % to all labor market' 'per_lm_alllm.cov_pop_tot']
['adequacy of social safety net programs % of total welfare of beneficiary
households'
 'per_sa_allsa.adq_pop_tot']
['cpia debt policy rating 1=low to 6=high' 'IQ.CPA.DEBT.XQ']
['exports of goods services and primary income bop current us$'
 'BX.GSR.TOTL.CD']
['imports of goods services and primary income bop current us$'
 'BM.GSR.TOTL.CD']
['net oda received % of imports of goods services and primary income'
 'DT.ODA.ODAT.MP.ZS']
['primary income payments bop current us$' 'BM.GSR.FCTY.CD']
['primary income receipts bop current us$' 'BX.GSR.FCTY.CD']
['secondary income other sectors payments bop current us$'
 'BM.TRF.PRVT.CD']
['trade in services % of gdp' 'BG.GSR.NFSV.GD.ZS']
['secondary income receipts bop current us$' 'BX.TRF.CURR.CD']
['public and publicly guaranteed debt service % of exports of goods services and
primary income'
 'DT.TDS.DPPG.XP.ZS']
['short to term debt % of exports of goods services and primary income'
 'DT.DOD.DSTC.XP.ZS']
['2005 ppp conversion factor gdp lcu per international $' 'PA.NUS.PPP.05']
['current account balance % of gdp' 'BN.CAB.XOKA.GD.ZS']
['net primary income bop current us$' 'BN.GSR.FCTY.CD']
['survey mean consumption or income per capita bottom 40% of population 2005 ppp
$ per day'
 'SI.SPR.PC40.05']
['survey mean consumption or income per capita total population 2005 ppp $ per
day'
 'SI.SPR.PCAP.05']
['logistics performance index: quality of trade and transport to related
infrastructure 1=low to 5=high'
 'LP.LPI.INFR.XQ']
['quality of port infrastructure wef 1=extremely underdeveloped to 7=well
developed and efficient by international standards'
 'IQ.WEF.PORT.XQ']
['annualized average growth rate in per capita real survey mean consumption or
income bottom 40% of population %'
 'SI.SPR.PC40.ZG']
['annualized average growth rate in per capita real survey mean consumption or
income total population %'
 'SI.SPR.PCAP.ZG']
['present value of external debt current us$' 'DT.DOD.PVLX.CD']

```
['present value of external debt % of gni' 'DT.DOD.PVLX.GN.ZS']
['present value of external debt % of exports of goods services and primary
income'
 'DT.DOD.PVLX.EX.ZS']
```

1.2.2 Important Features

```python
[82]: picked_indicators = [
          #economy
          ['gdp per capita current us$' 'NY.GDP.PCAP.CD'],
          ['gdp growth annual %' 'NY.GDP.MKTP.KD.ZG'],
          ['gni per capita ppp current international $' 'NY.GNP.PCAP.PP.CD'],
          ['foreign direct investment net inflows % of gdp' 'BX.KLT.DINV.WD.GD.ZS'],

          # tech
          ['high to technology exports % of manufactured exports' 'TX.VAL.TECH.MF.
       ↪ZS'],
          ['high to technology exports current us$' 'TX.VAL.TECH.CD'],
          ['researchers in r&d per million people' 'SP.POP.SCIE.RD.P6'],
          ['technicians in r&d per million people' 'SP.POP.TECH.RD.P6'],

          # employment
          ['unemployment total % of total labor force' 'SL.UEM.TOTL.ZS'],
          ['unemployment total % of total labor force national estimate' 'SL.UEM.TOTL.
       ↪NE.ZS'],
          ['employers total % of employment' 'SL.EMP.MPYR.ZS'],

          # Social
          ['labor force participation rate female % of female population ages 15+␣
       ↪national estimate' 'SL.TLF.CACT.FE.NE.ZS'],
          ['age dependency ratio old % of working to age population' 'SP.POP.DPND.
       ↪OL'],
          ['birth rate crude per 1000 people' 'SP.DYN.CBRT.IN'],

           # Government
          ['central government debt total % of gdp' 'GC.DOD.TOTL.GD.ZS'],
          ['gross savings current us$' 'NY.GNS.ICTR.CD'],


      ]
```

1.2.3 Seperate only Indicator Code from above list.

```python
[83]: picked_indicators_code = []
      for ele in picked_indicators :
          code = re.findall('([A-Z.]\S*)',ele[0])
```

```
        picked_indicators_code.append(code[0])

picked_indicators_code
```

[83]:
```
['NY.GDP.PCAP.CD',
 'NY.GDP.MKTP.KD.ZG',
 'NY.GNP.PCAP.PP.CD',
 'BX.KLT.DINV.WD.GD.ZS',
 'TX.VAL.TECH.MF.ZS',
 'TX.VAL.TECH.CD',
 'SP.POP.SCIE.RD.P6',
 'SP.POP.TECH.RD.P6',
 'SL.UEM.TOTL.ZS',
 'SL.UEM.TOTL.NE.ZS',
 'SL.EMP.MPYR.ZS',
 'SL.TLF.CACT.FE.NE.ZS',
 'SP.POP.DPND.OL',
 'SP.DYN.CBRT.IN',
 'GC.DOD.TOTL.GD.ZS',
 'NY.GNS.ICTR.CD']
```

1.2.4 Subset of data with the required features alone

1.2.5 Chose only India and China for Analysis

[84]:
```
# Subset of data with the required features alone
df_subset = df[df['IndicatorCode'].isin(picked_indicators_code)]

# Chose only India and China for Analysis
df_korea = df_subset[df['CountryName'].str.contains("Korea, Rep")]
df_japan = df_subset[df['CountryName'].str.contains("Japan")]
```

/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:5: UserWarning:
Boolean Series key will be reindexed to match DataFrame index.
  """
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:6: UserWarning:
Boolean Series key will be reindexed to match DataFrame index.

[85]:
```
df_korea = add_data(df_korea,new_data)
df_japan = add_data(df_japan,new_data)
```

[86]:
```
def plot_indicator(indicator, delta=10):
    # df
    ds_korea =␣
 ↪df_korea[['IndicatorName','Year','Value']][df_korea['IndicatorCode']==indicator]
    ds_japan =␣
 ↪df_japan[['IndicatorName','Year','Value']][df_japan['IndicatorCode']==indicator]
```
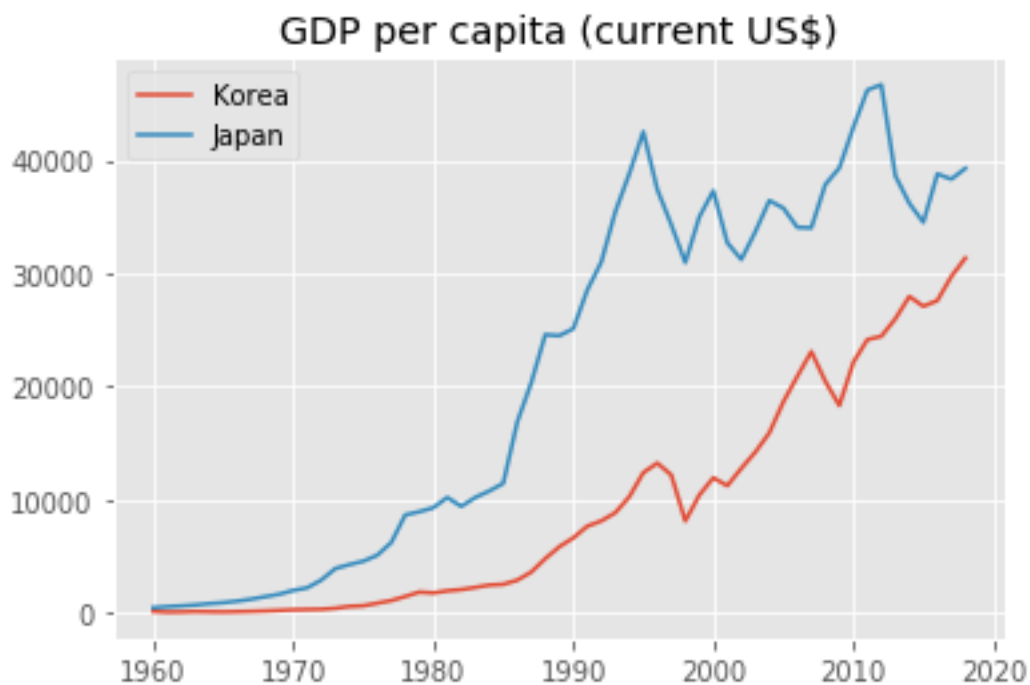
```
    try:
        title = ds_korea['IndicatorName'].iloc[0]
    except:
        title = "None"

    ykorea = ds_korea['Value'].values
    yjapan = ds_japan['Value'].values
    xkorea = ds_korea['Year'].values
    xjapan = ds_japan['Year'].values

    plt.plot(xkorea,ykorea,label='Korea')
    plt.plot(xjapan,yjapan,label='Japan')
    plt.title(title)
    plt.legend(loc=2)
```

[87]:
```
plot_indicator(picked_indicators_code[0])
```


GDP per capita (current US$)

[ ]: