**Assignment 2**

**Code:**

# In[1]:

get\_ipython().system('pip install pycountry\_convert')

import pandas as pd

import matplotlib.pyplot as plot

import pycountry\_convert as pc

import numpy as np

# In[2]:

mn = pd.read\_excel(r'E:\5502\HIV.xlsx',encoding = "ISO-8859-1")

# In[3]:

continents = []

for country in mn['Estimated HIV Prevalence% - (Ages 15-49)']:

try:

country\_code = pc.country\_name\_to\_country\_alpha2(country, cn\_name\_format="default")

continent\_name = pc.country\_alpha2\_to\_continent\_code(country\_code)

continents.append(continent\_name)

except:

continents.append("Unknown")

# In[4]:

mn['continents']=continents

mn\_new=mn.iloc[:,[22,23,24,25,26,27,28,29,30,31,32,33]]

mn\_new=(mn\_new.sum(axis=1))/12

mn['Average']=mn\_new

mn.to\_csv("E:\\5502\\results.csv")

# In[5]:

Highavg = mn.groupby(['continents']).apply(lambda t: t[t.Average == t.Average.max()])

lowavg = mn.groupby(['continents']).apply(lambda t : t[t.Average == t.Average.min()])

# In[6]:

Highavg.plot(x='continents',y='Average',kind="bar")

ax=Highavg.plot(x='continents',y='Average',kind="bar")

lowavg.plot(x='continents',y='Average',kind="bar")

lowavg.plot(x='continents',y='Average',kind="bar",ax=ax)

mn1=mn.sample(frac=.12)

print(mn1)

bx=Highavg.plot(x='continents',y='Average',kind="line")

lowavg.plot(x='continents',y='Average',kind="line",ax=bx)

mn1.plot(x='continents',y='Average',kind="line",ax=bx)

plot.show()

# In[7]:

mn2=mn.groupby(['continents'])

asia=mn2.get\_group('AS')

mnasia = asia.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnasiaavg=(mnasia.sum(axis=0))/(len(asia))

mnasiaavg.plot(x='year',y='Average',color='red',kind='line')

# In[8]:

africa=mn2.get\_group('AF')

mnafrica = africa.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnafricaavg=(mnafrica.sum(axis=0))/(len(africa))

mnafricaavg.plot(x='year',y='Average',color='blue',kind='line')

# In[9]:

europe=mn2.get\_group('EU')

mneurope = europe.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mneuropeavg=(mneurope.sum(axis=0))/(len(europe))

mneuropeavg.plot(x='year',y='Average',color='green',kind='line')

# In[10]:

unknown=mn2.get\_group('Unknown')

mnunknown = unknown.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnunknownavg=(mnunknown.sum(axis=0))/(len(unknown))

mnunknownavg.plot(x='year',y='Average',color='violet',kind='line')

# In[11]:

northamerica = mn2.get\_group('NA')

mnna = northamerica.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnnaavg=(mnna.sum(axis=0))/(len(northamerica))

mnnaavg.plot(x='year',y='Average',color='black',kind='line')

# In[12]:

southamerica = mn2.get\_group('SA')

mnsa = southamerica.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnsaavg=(mnsa.sum(axis=0))/(len(southamerica))

mnsaavg.plot(x='year',y='Average',color='yellow',kind='line')

# In[13]:

aus = mn2.get\_group('OC')

mnaus = aus.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33]]

mnausavg=(mnaus.sum(axis=0))/(len(aus))

mnausavg.plot(x='year',y='Average',color='orange',kind='line')

# In[14]:

mn3=mn.groupby(['continents','Estimated HIV Prevalence% - (Ages 15-49)'],as\_index=False).mean()

colors=np.random.RandomState(0)

co=colors.rand(275)

fig=plot.figure(figsize=(20,10))

sp1991=fig.add\_subplot(1,2,1)

sp2010=fig.add\_subplot(1,2,2)

sp1991.scatter(mn['continents'],mn[1990],s=100,c=co)

sp2010.scatter(mn['continents'],mn['2010'],s=100,c=co)

sp1991.set\_xlabel("Continents")

sp1991.set\_ylabel("Avg")

sp1991.set\_title(" 1990 Year Matplot ")

sp2010.set\_xlabel("Continents")

sp2010.set\_ylabel("Avg")

sp2010.set\_title(" 2010 Year Matplot ")

plot.show()

Question 1:

Imported Pycountry and Pandas modules to get continents of the given countries and the column named Country as mentioned.

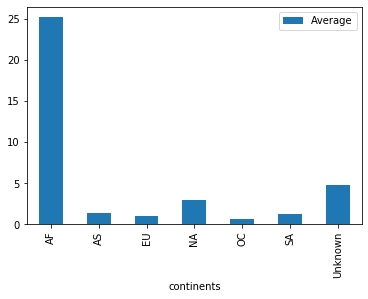
All the countries in the pycountry module are converted to country code and country codes to continent.

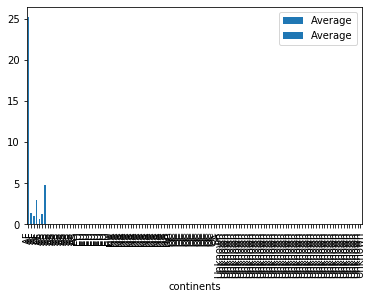
The results are stored in the results.csv file

Question 2:

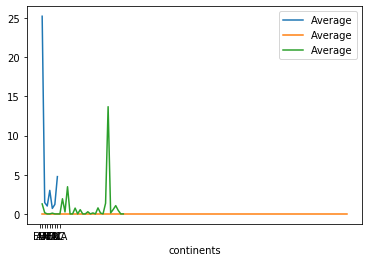
**Barchart:**

[33 rows x 36 columns]

****

****

****

****

The average of all countries from 1990 – 2011 is calculated and group by group method is used to find the highest and lowest average continent based country values

The barcharts shows the highest and lowest average values.

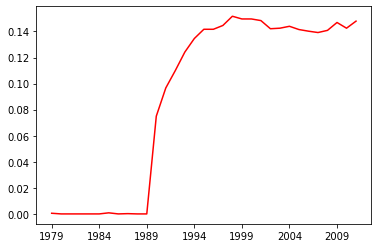
Question 3:

Performed mean function and group by function and after transposed rows and columns and plotted line charts for each continents as shown below.

Output:

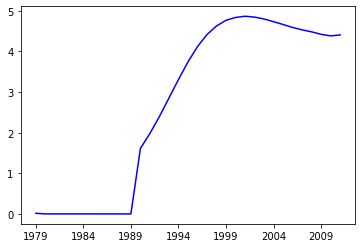
For Asia

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee38e7d9c8>



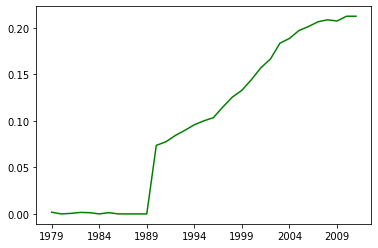
For Africa

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee38edf548>



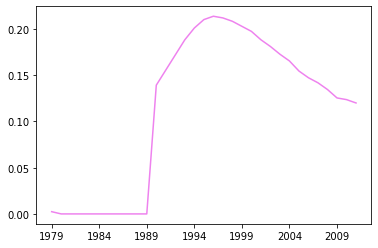
For Europe

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee38f46f88>



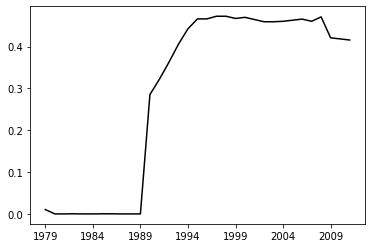
For Unknown

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee38fb0148>



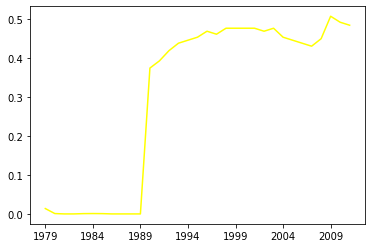
For North America

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee3795e308>



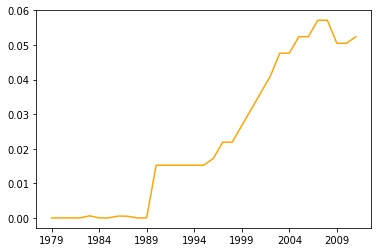
For South America

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee38dd7a48>

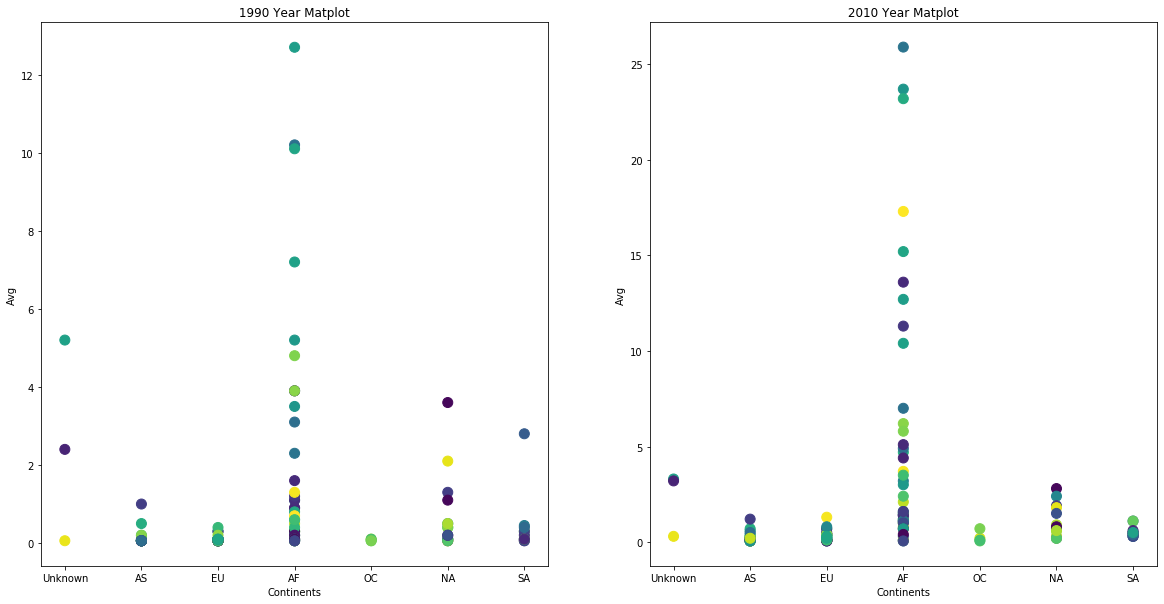


For Australia

<matplotlib.axes.\_subplots.AxesSubplot at 0x1ee37d73d08>



Question 4:



The above figures shows scatter plots for 1990 year and 2010 year for each continent.