Report

Part 1

A CSV file with all countries mapped to respective continent is acquired online. This file is loaded into a data science table and joined to the Input table. The joining is based on the common column Country. And the countries are prepared in the same format as the input table. So, Continents are mapped to each country in the input table

Here a new column Average2000_2011 is calculated by taking mean of all values from columns 2000 to 2011 across all rows. This is used in the next task.

Country	Continent	Average2000_2011	1979	1980	1981	1982	1983	1984	1985	1986	1987	1990	1991	1992	1993	1994	1995	1996
Swaziland	AFRICA	25.58	nan	nan	nan	0.01097	nan	nan	nan	nan	nan	2.3	3.2	4.4	6.1	8.1	10.6	13.3
Botswana	AFRICA	25.02	0.105976	nan	nan	nan	nan	nan	nan	nan	nan	3.5	5.1	7.3	10.1	13.3	16.6	19.7
Lesotho	AFRICA	23.59	nan	nan	nan	nan	nan	0.0110793	nan	nan	nan	0.8	1.7	3.3	5.9	9.8	14.3	18.5
South Africa	AFRICA	17.77	nan	nan	0.0142086	nan	nan	nan	nan	nan	nan	0.7	1.2	1.8	2.9	4.3	6.1	8.4
Zimbabwe	AFRICA	17.6	0.01	nan	nan	nan	nan	nan	nan	nan	nan	10.1	13.6	17.2	20.6	23.3	25.1	26.2

Part 2

Create new tables for Highest and Lowest Average value, by selecting columns Continent Country and Average2000_2011, and group it per continent to find maximum and minimum values per continent from Average2000_2011 calculated in the previous step. This table is appended to input table as Average2000_2011 min and Average2000_2011 max columns and year fields from 1979 to 1999 are dropped.

The List of Countries with Maximum Average HIV Estimator in each Continent

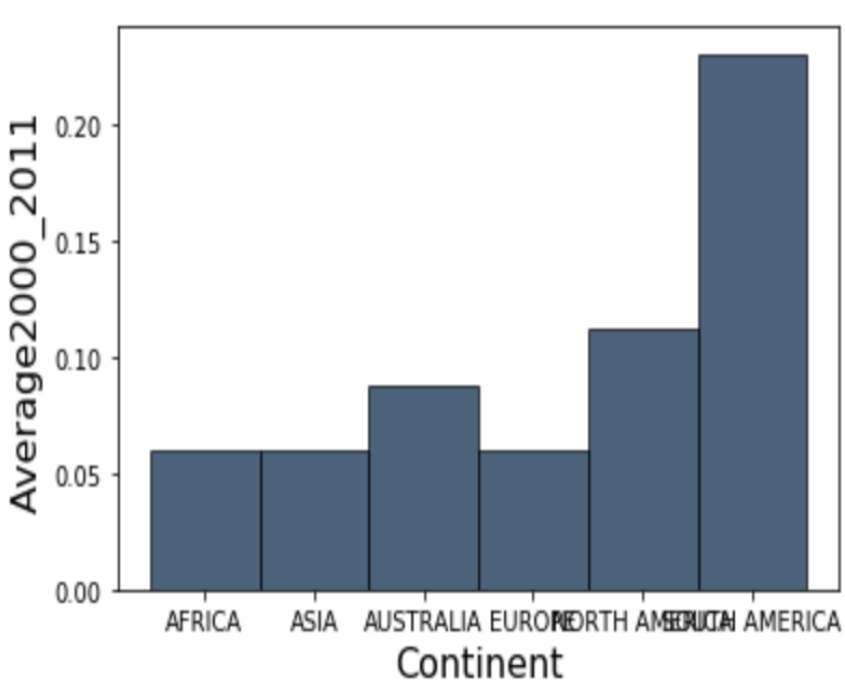
Continent	Country	Average2000_2011
AFRICA	Swaziland	25.58
ASIA	Thailand	1.39
AUSTRALIA	Papua New Guinea	0.75
EUROPE	Estonia	1.1
NORTH AMERICA	Bahamas	2.97
SOUTH AMERICA	Guyana	1.16

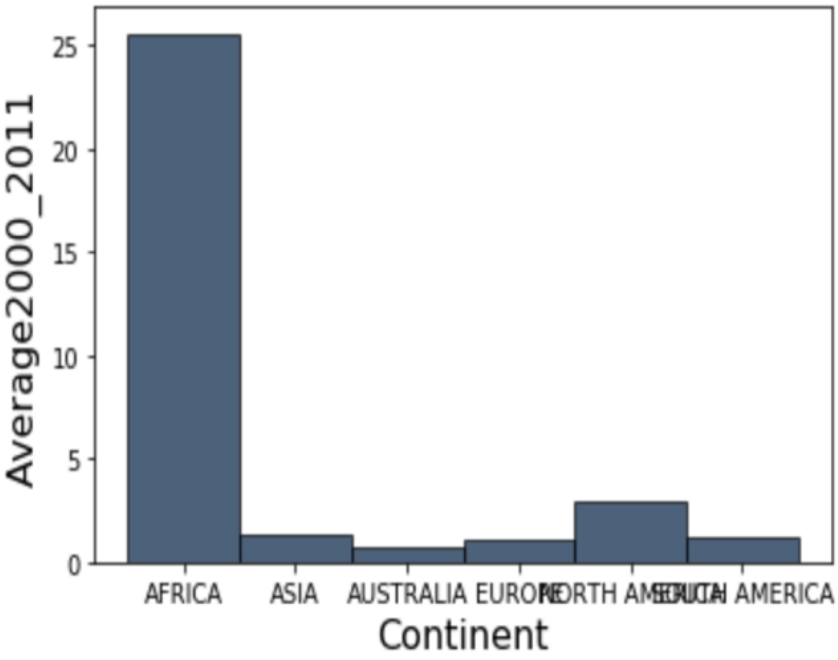
Continent	Country	Average2000_2011
AFRICA	Tunisia	0.06
AFRICA	Egypt	0.06
ASIA	Turkey	0.06
ASIA	Sri Lanka	0.06
ASIA	South Korea	0.06
ASIA	Qatar	0.06
ASIA	Philippines	0.06
ASIA	Mongolia	0.06
ASIA	Maldives	0.06
ASIA	Japan	0.06

ASIA	China	0.06
ASIA	Bangladesh	0.06
ASIA	Afghanistan	0.06
AUSTRALIA	Fiji	0.088
EUROPE	Slovak Republic	0.06
EUROPE	Czech Republic	0.06
EUROPE	Croatia	0.06
NORTH AMERICA	Cuba	0.112
SOUTH AMERICA	Bolivia	0.23

Country with the highest is found by comparing values of average and highest average and only those records are retained in the new table.

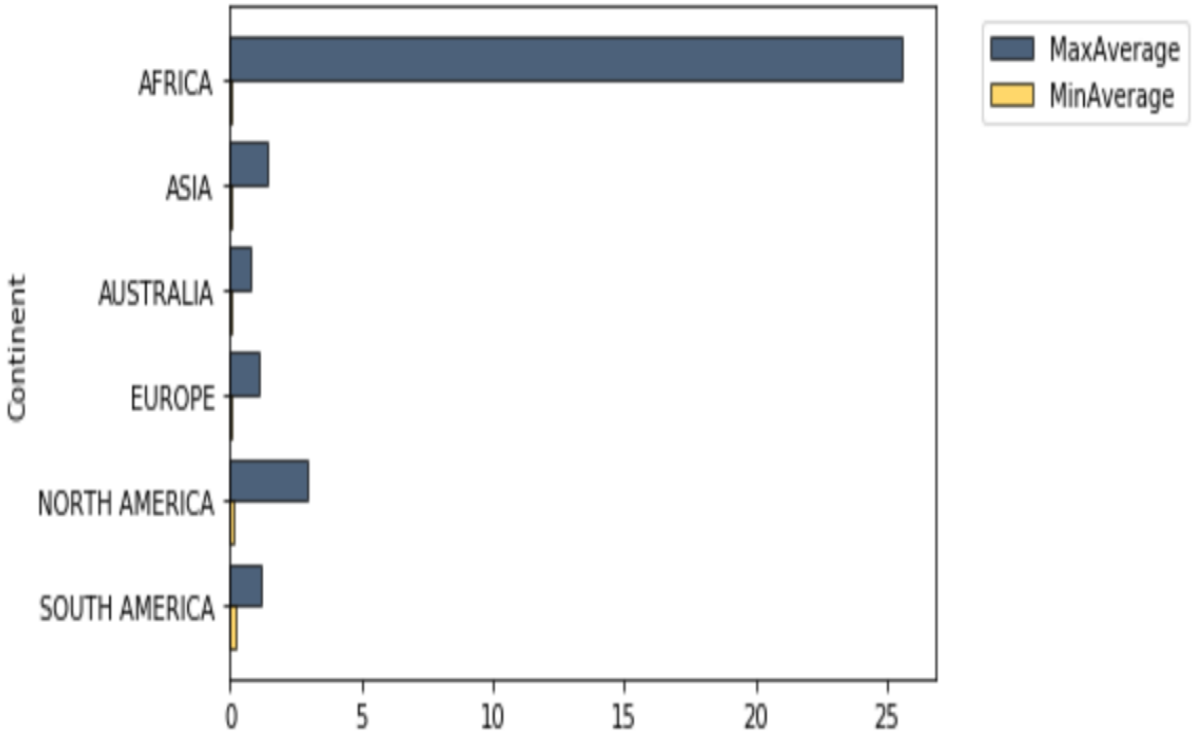
Same is done for Lowest average value. As there are multiple countries with same low values output table has multiple entries one for each country-continent pair





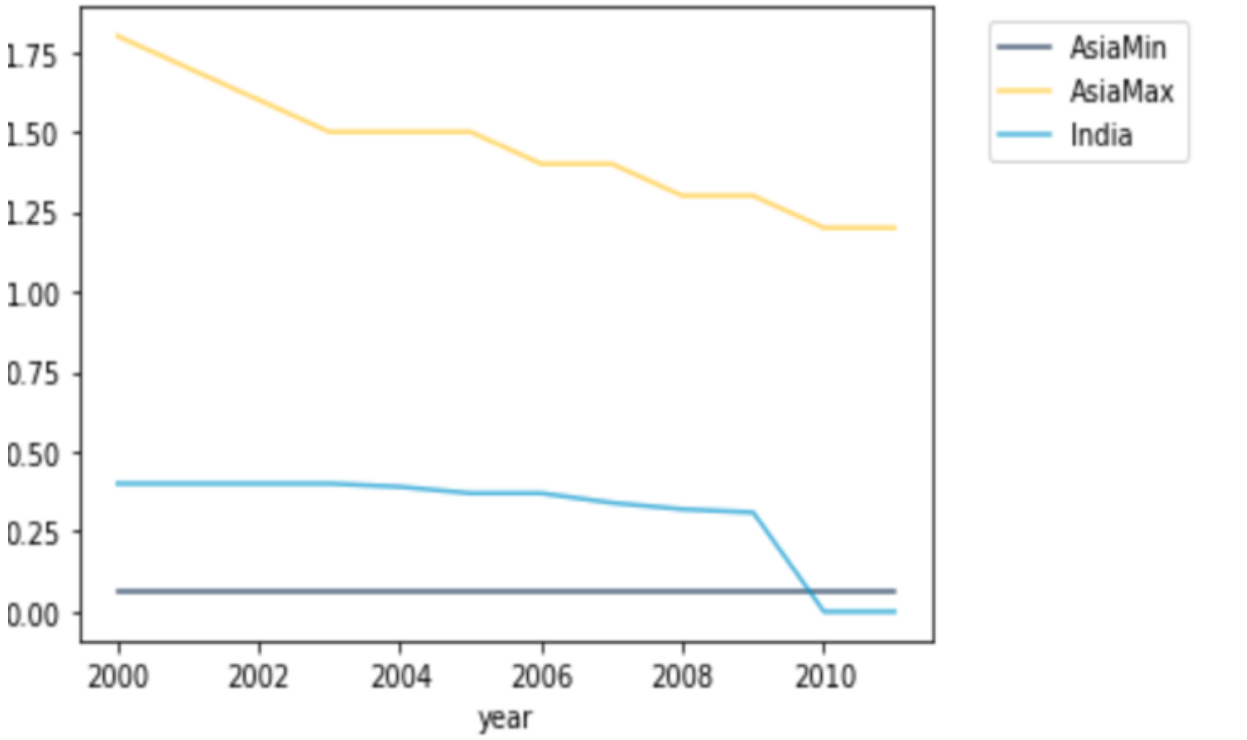
The new table created here is used to plot highest average HIV prevalence per continent. For Highest Average value, it is straight forward. For lowest we need to select only one among the multiple lowest value. This is done by grouping again only based on Continent and selecting min.

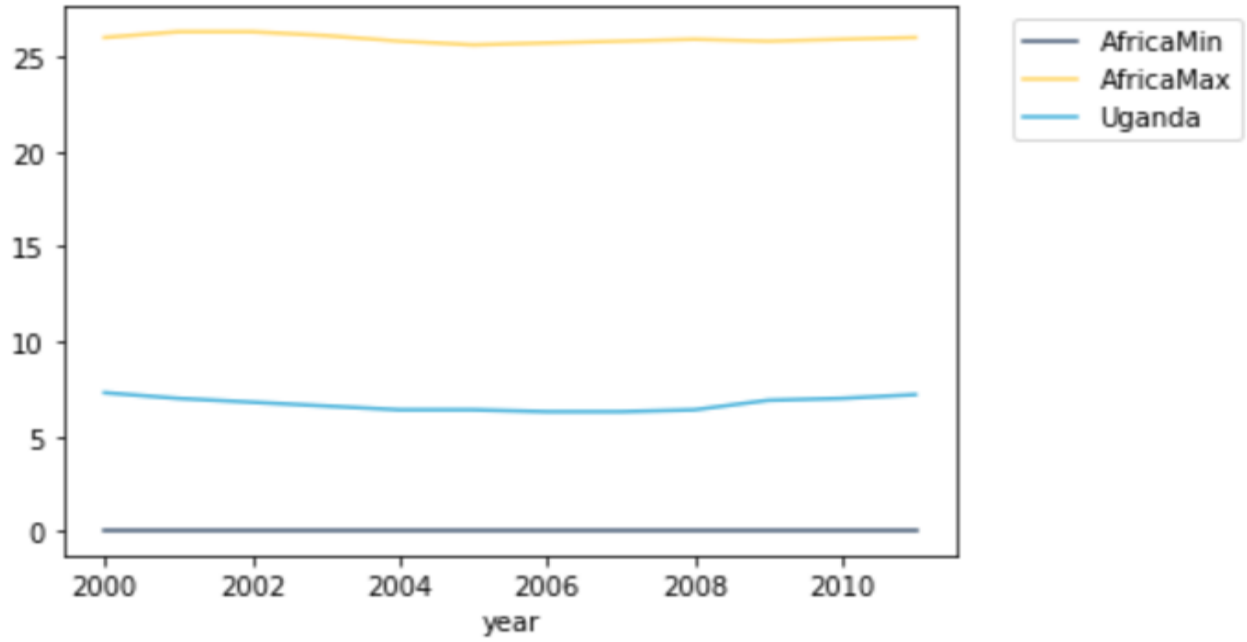
The new tables created are joined on column Continent and used to plot minimum, maximum overlaid bar-chart across each continent

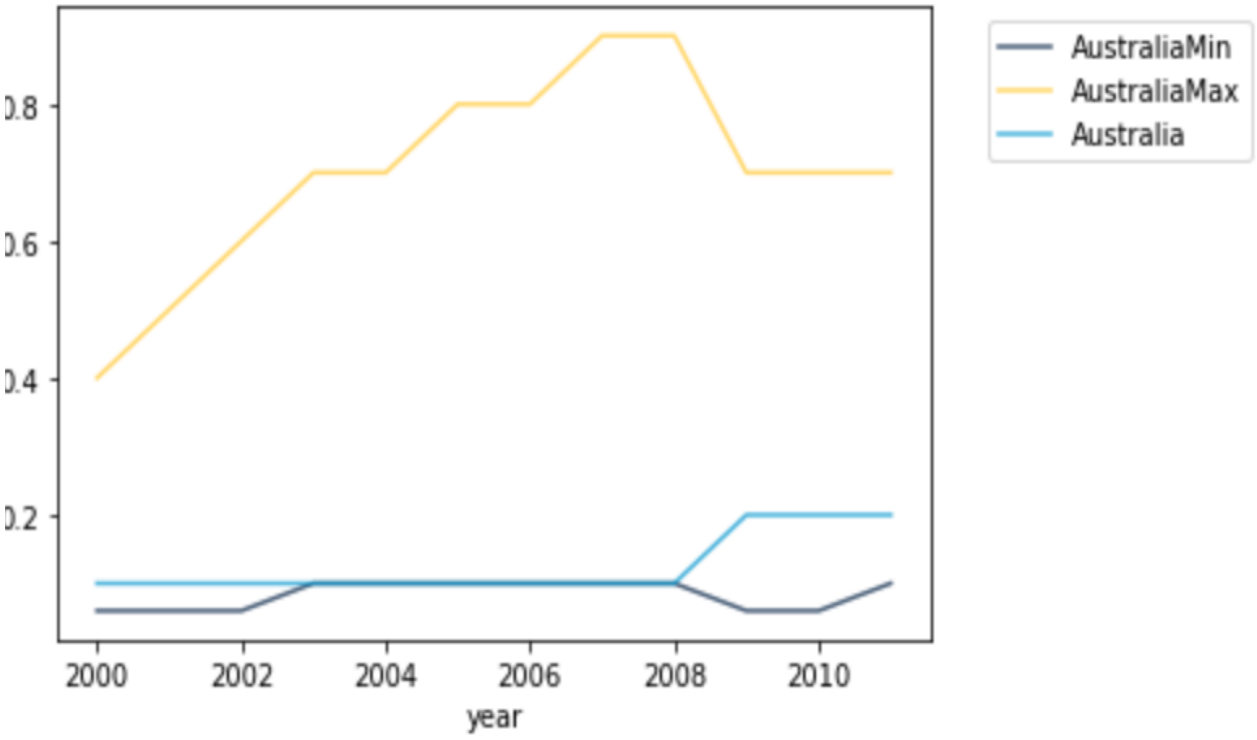


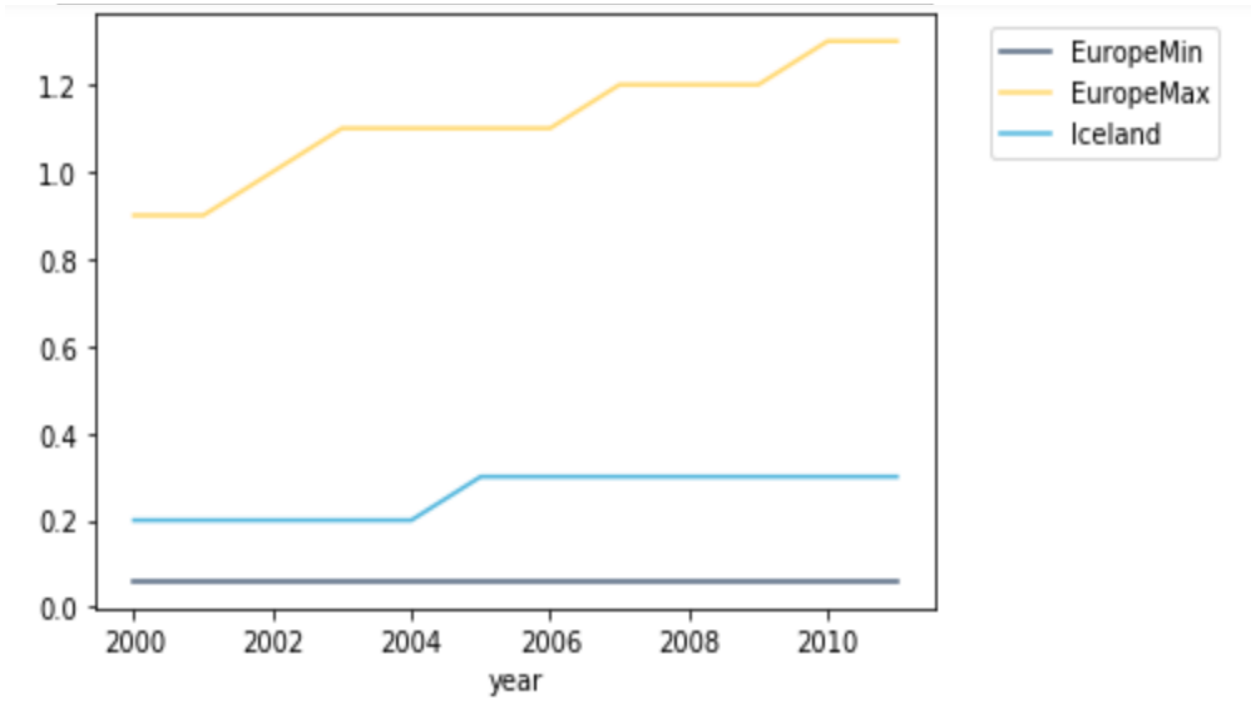
One country is selected from each continent and that country's row is selected with just values from year 2000 to 2011 country, continent, and min/max of each continent columns

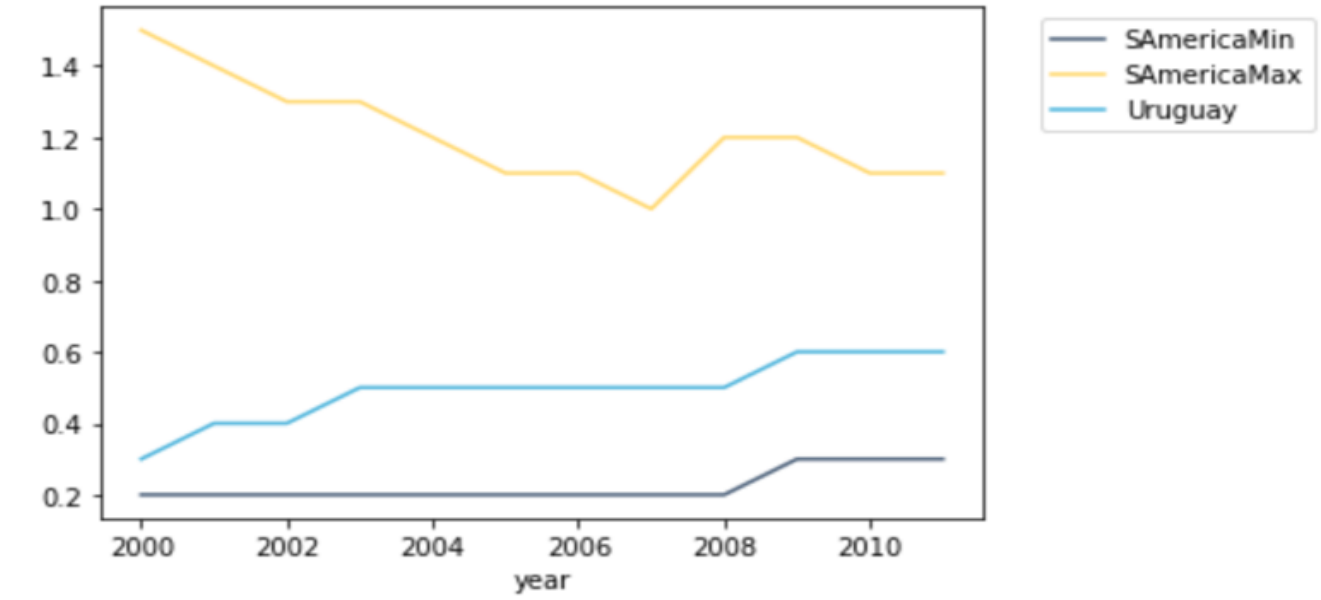
A new Table is created with each of these country as columns. This table is plotted to get minimum, maximum per continent and the selected country in that continent

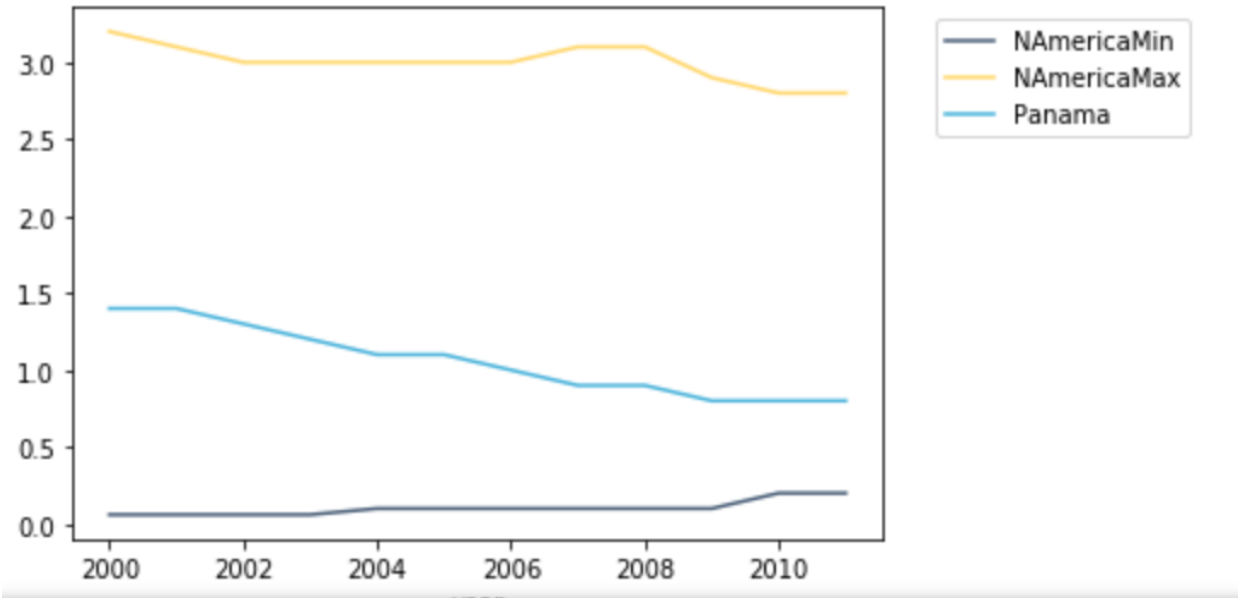

1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2000
2002
2004
2006
2008
2010
year
AsiaMin
AsiaMax
India


25
20
15
10
5
0
2000
2002
2004
2006
2008
2010
year
AfricaMin
AfricaMax
Uganda


AustraliaMin
AustraliaMax
Australia
0.8
0.6
0.4
0.2
2000
2002
2004
2006
2008
2010
year

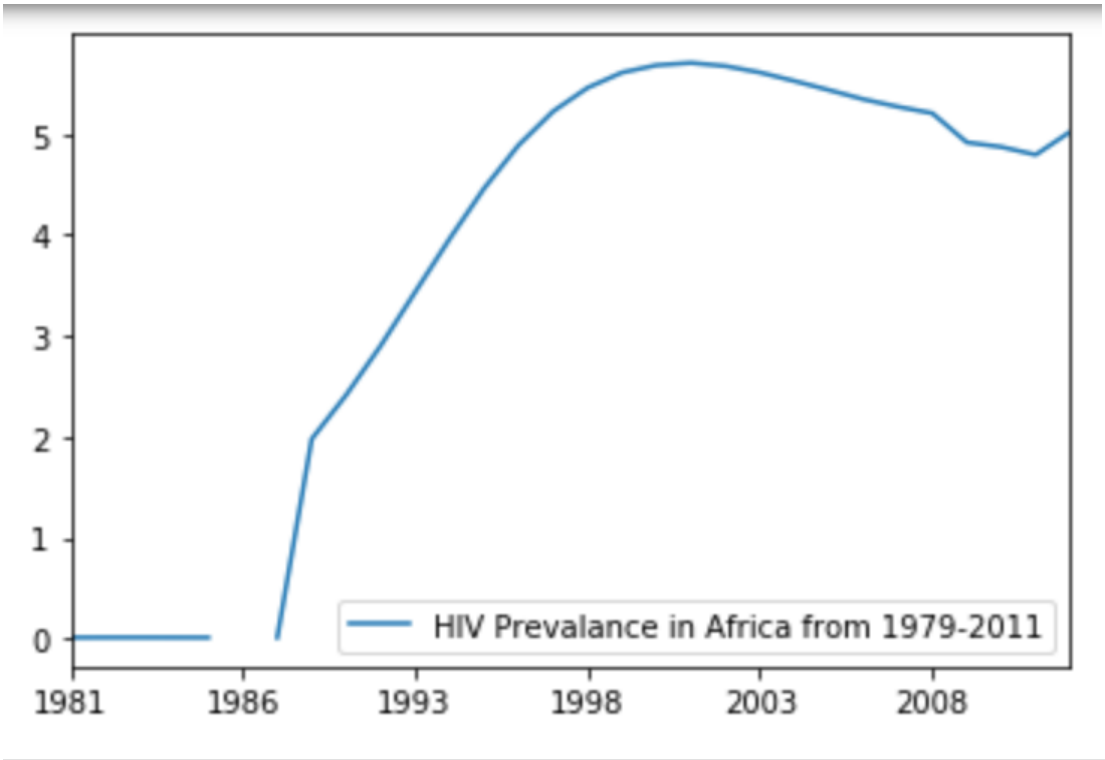

EuropeMin
EuropeMax
Iceland
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2000
2002
2004
2006
2008
2010
year

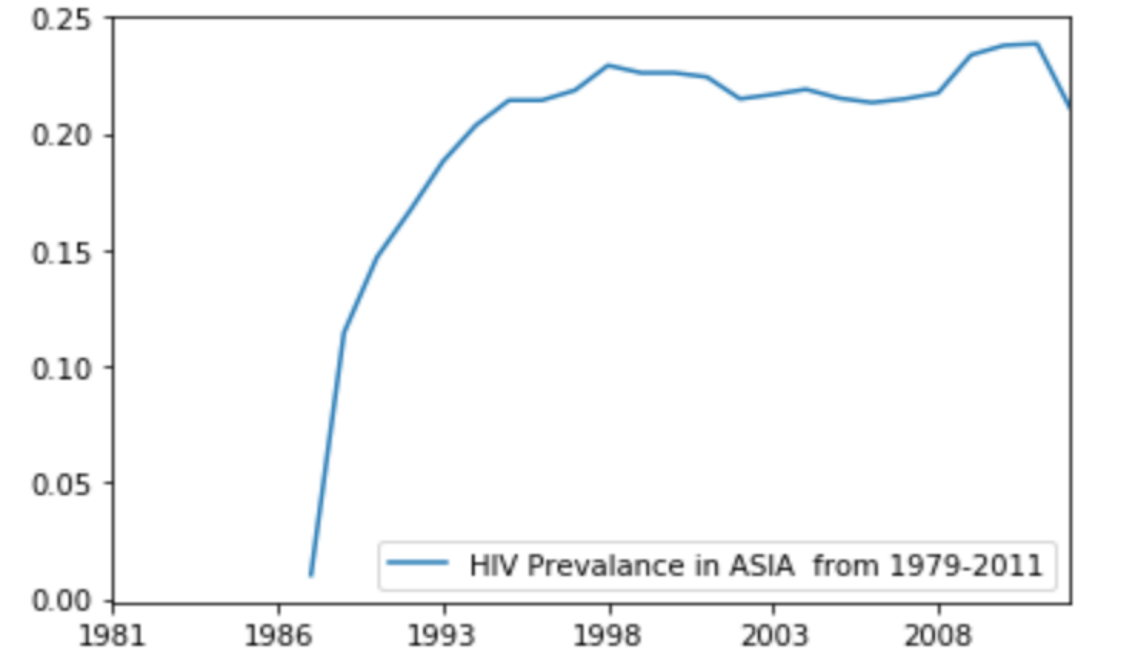


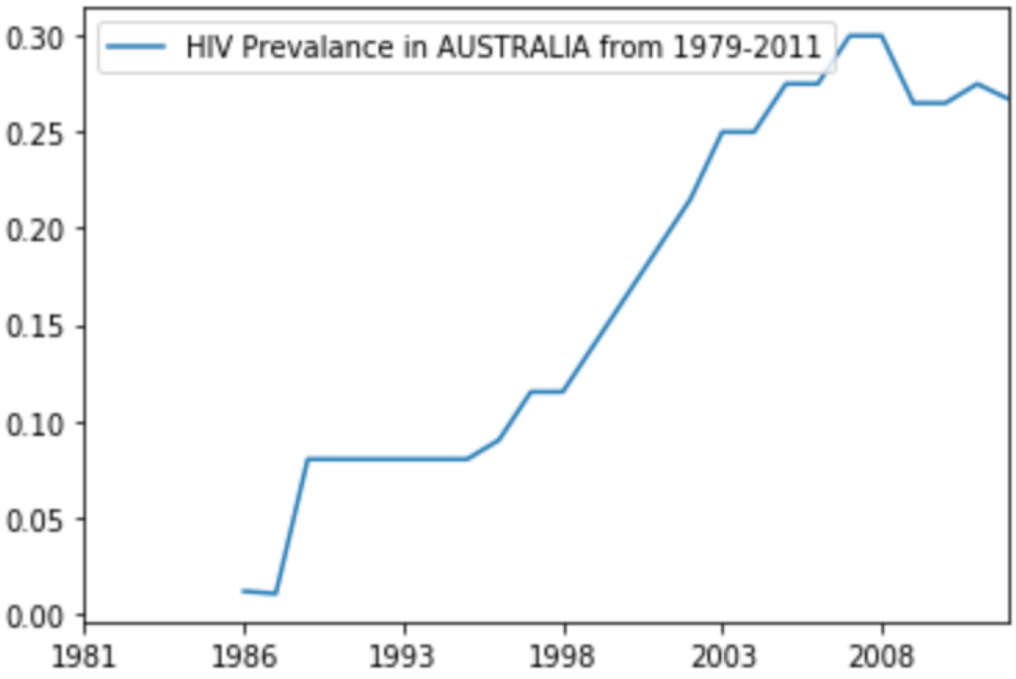


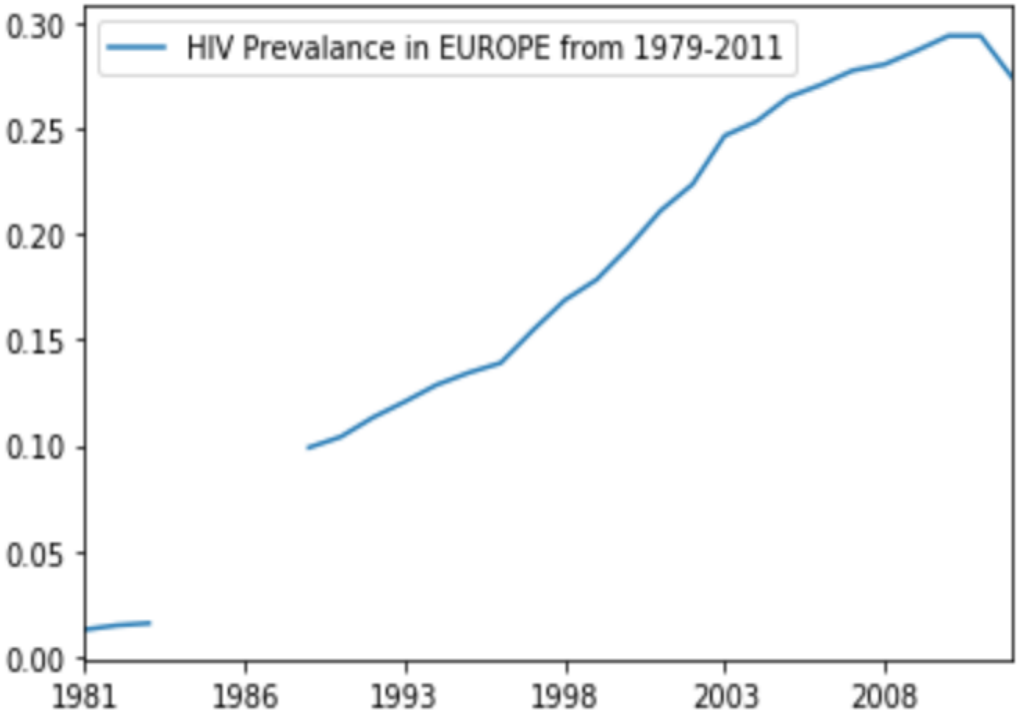
Part 3

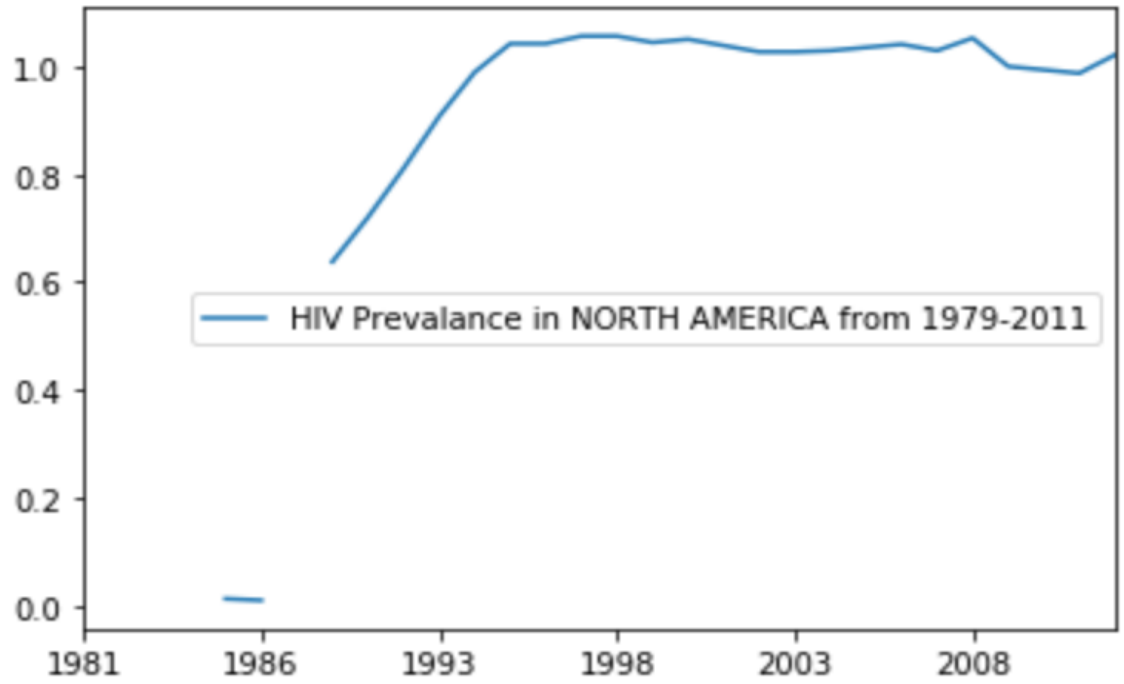
New/Separate data frame is created for each continent and mean is calculated on axis =1 for 1979 to 2011 to get average of each year. For each continent this table is plotted separately and for the overlaid chart, these are plotted in a single plot.

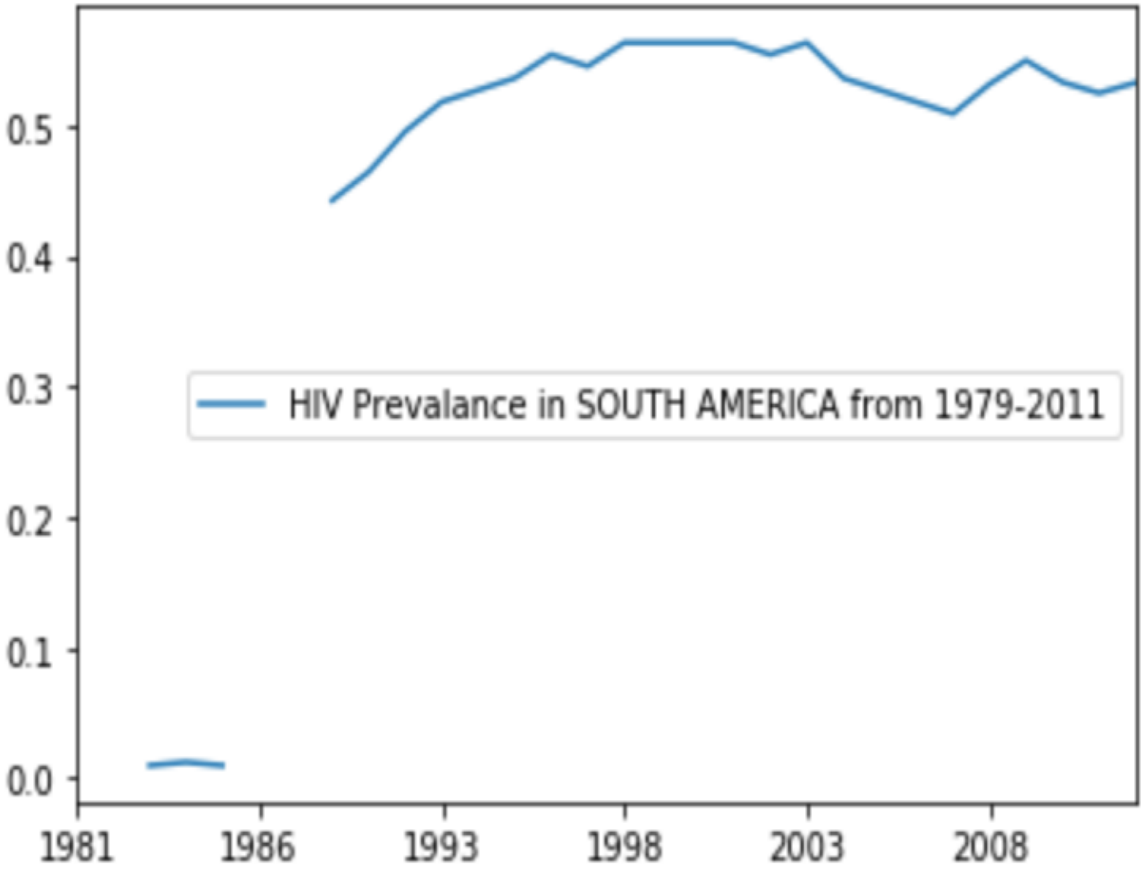

5
4
3
2
1
0
1981
1986
1993
1998
2003
2008
HIV Prevalance in Africa from 1979-2011

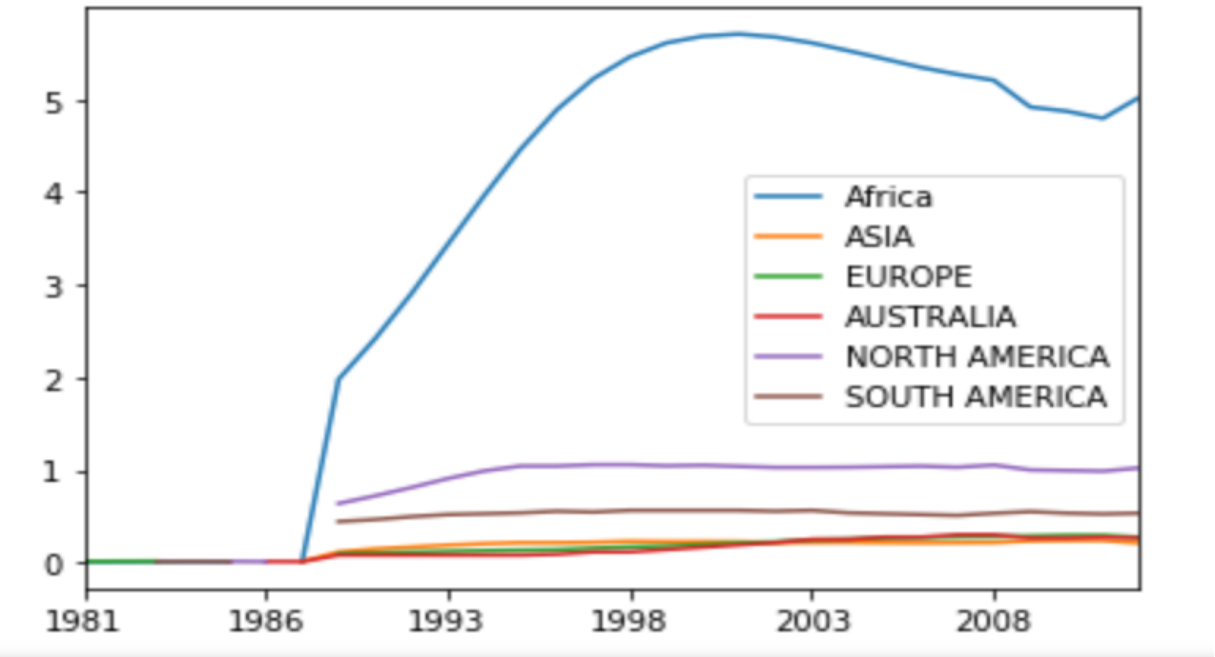

0.25
0.20
0.15
0.10
0.05
0.00
1981
1986
1993
1998
2003
2008
HIV Prevalance in ASIA from 1979-2011


HIV Prevalance in AUSTRALIA from 1979-2011
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1981
1986
1993
1998
2003
2008


HIV Prevalance in EUROPE from 1979-2011
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1981
1986
1993
1998
2003
2008

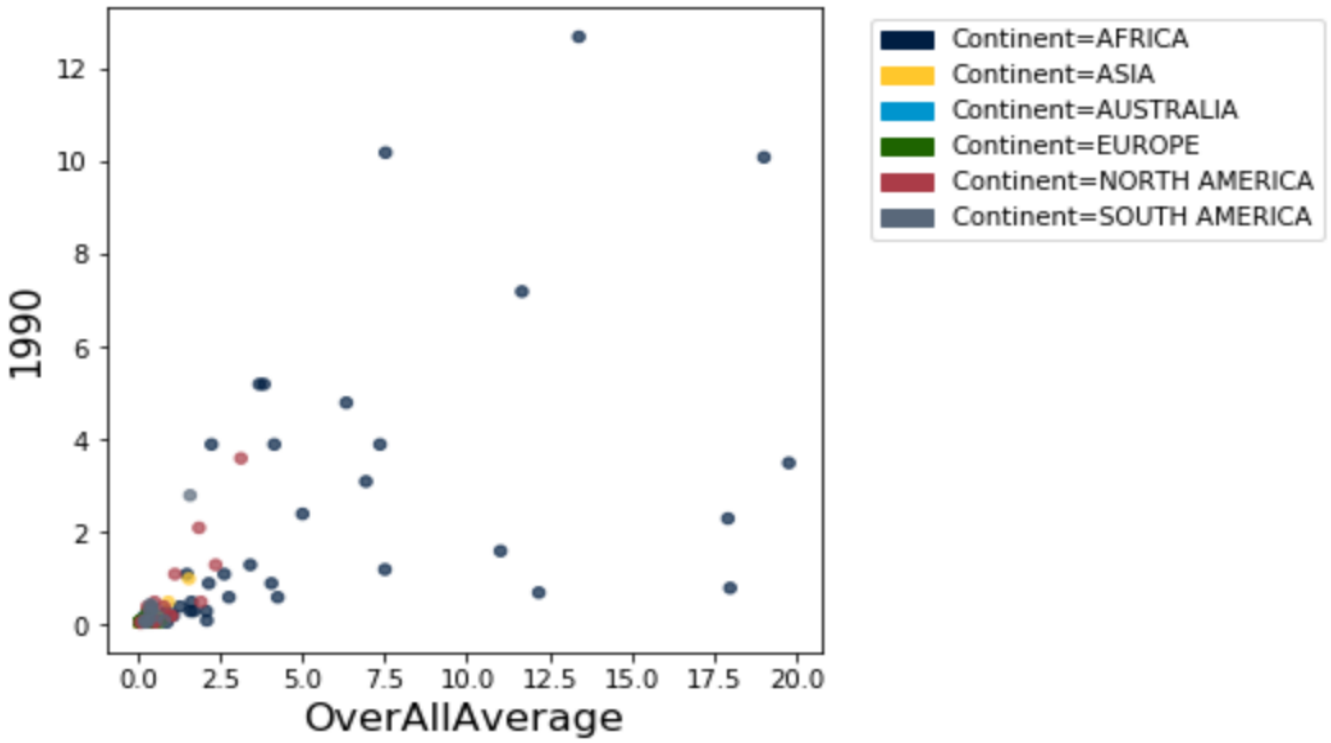

1.0
0.8
0.6
0.4
0.2
0.0
HIV Prevalance in NORTH AMERICA from 1979-2011
1981
1986
1993
1998
2003
2008

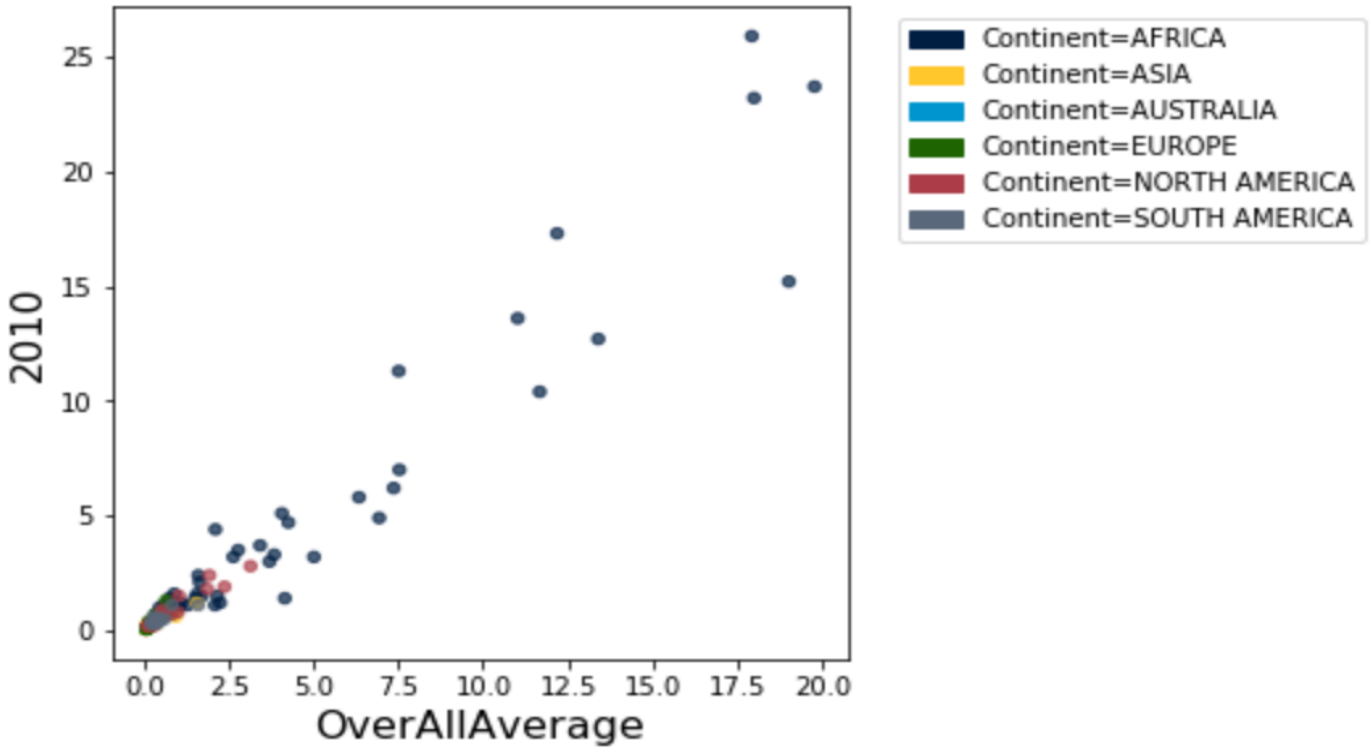

0.5
0.4
0.3
0.2
0.1
0.0
HIV Prevalance in SOUTH AMERICA from 1979-2011
1981
1986
1993
1998
2003
2008



Part 4

The overall average of each country is plotted against the HIV estimated prevalence data for year 1990 and in second scatter the overall average of each country is plotted against the HIV estimated prevalence data for year 2010.





It is to be noted that in year 1990 the number of countries with HIV estimated prevalence significantly different from the overall average was more. But in 2010 both these values are almost same.

Secondly, overall average did not increase significantly in 2010, but African countries are showing an drastic increase compared to 1990.The overall average remained almost same as the HIV estimated prevalence in other continents decreased.