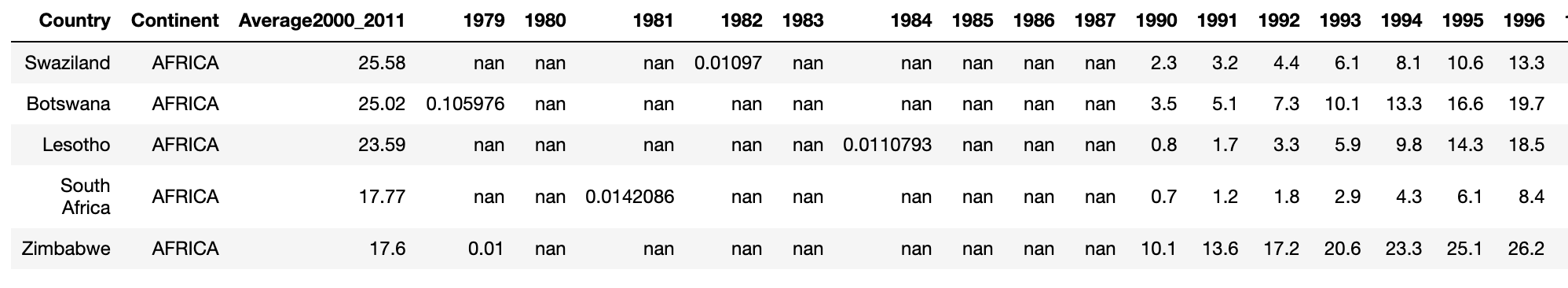
Report

**Part 1**

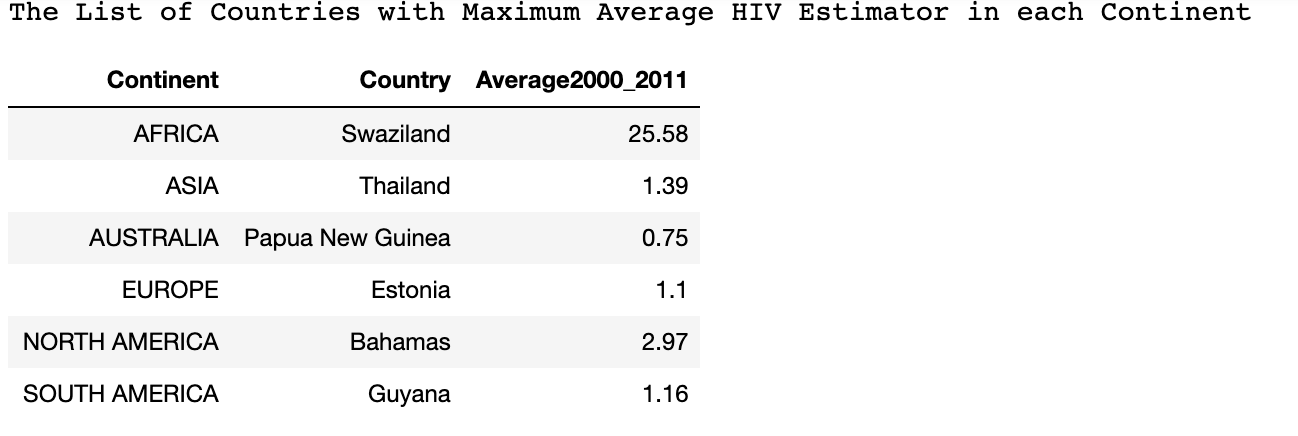
A CSV file with all countries mapped to respective continent is acquired online. This file is loaded into a data science table and joined to the Input table. The joining is based on the common column Country. And the countries are prepared in the same format as the input table. So, Continents are mapped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

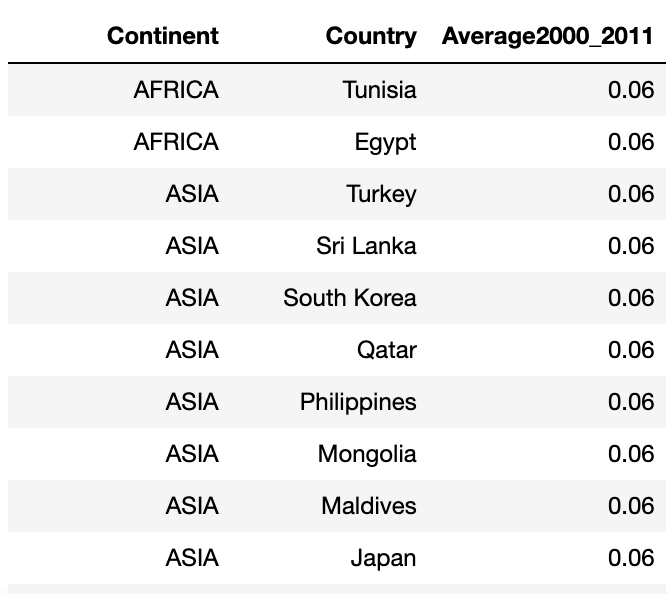
Here a new column Average2000\_2011 is calculated by taking mean of all values from columns 2000 to 2011 across all rows. This is used in the next task.

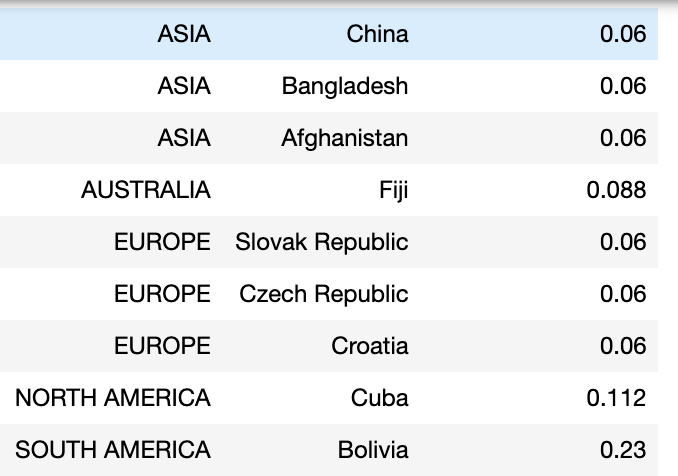


**Part 2**

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.

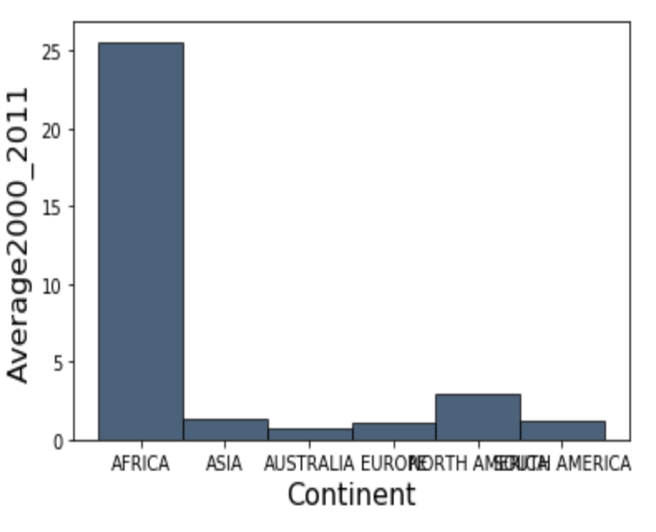
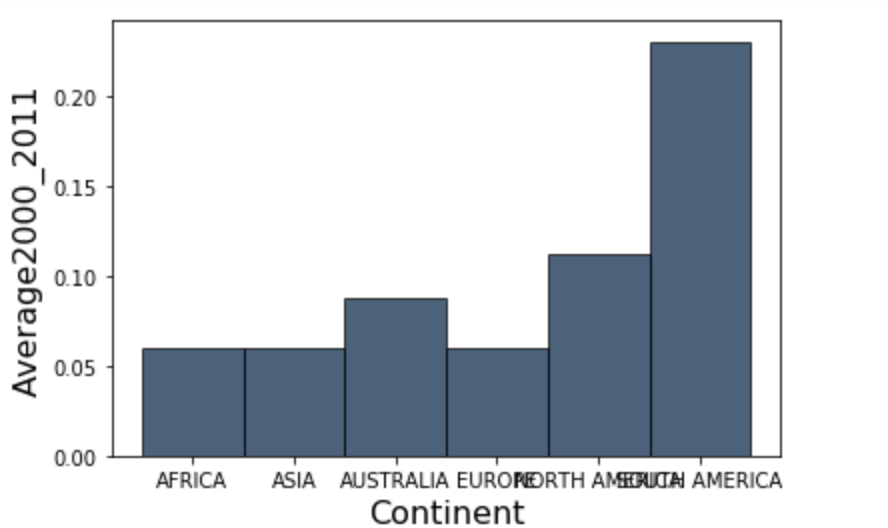






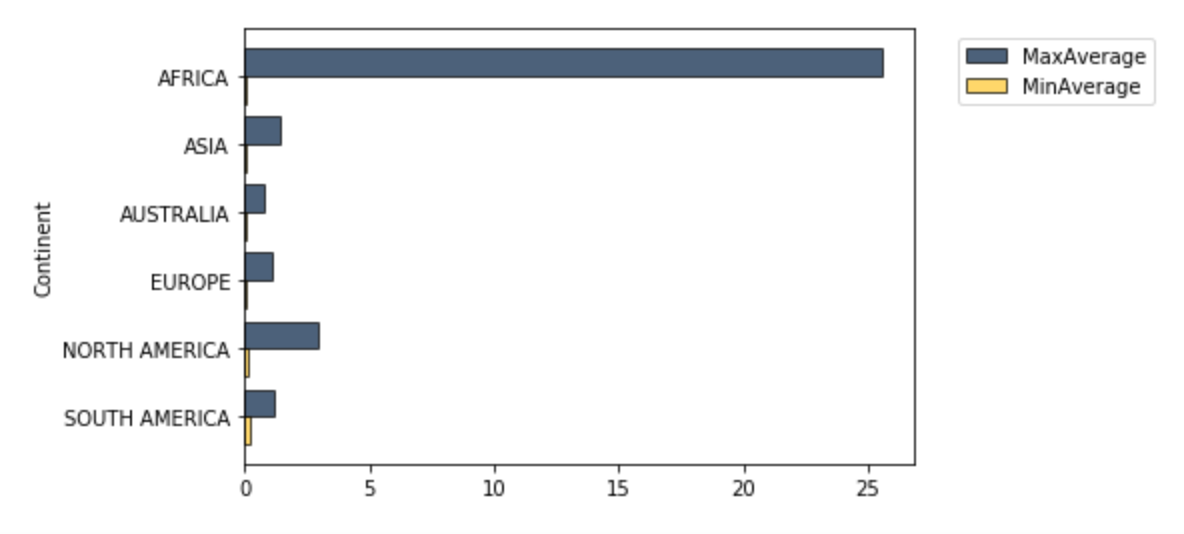
Country with the highest is found by comparing values of average and highest average and only those records are retained in the new table.

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



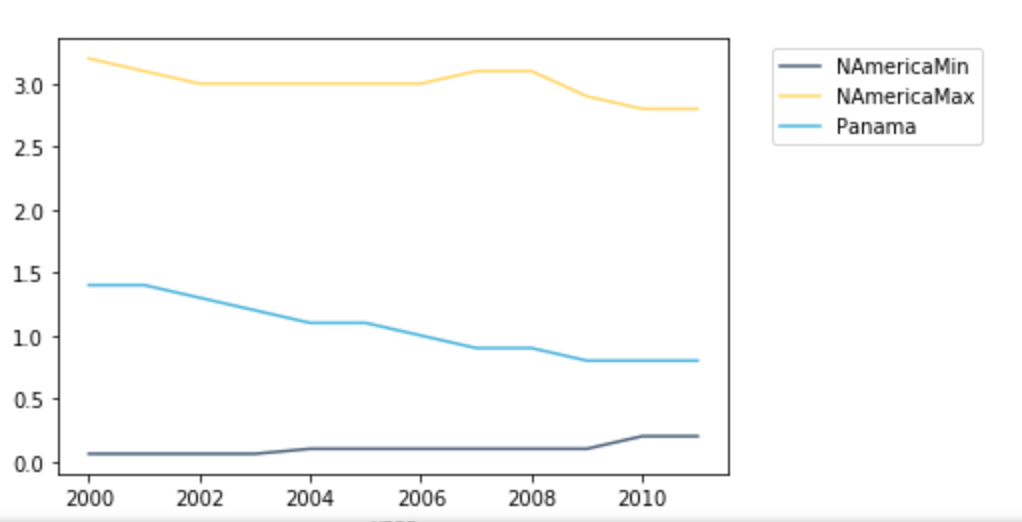
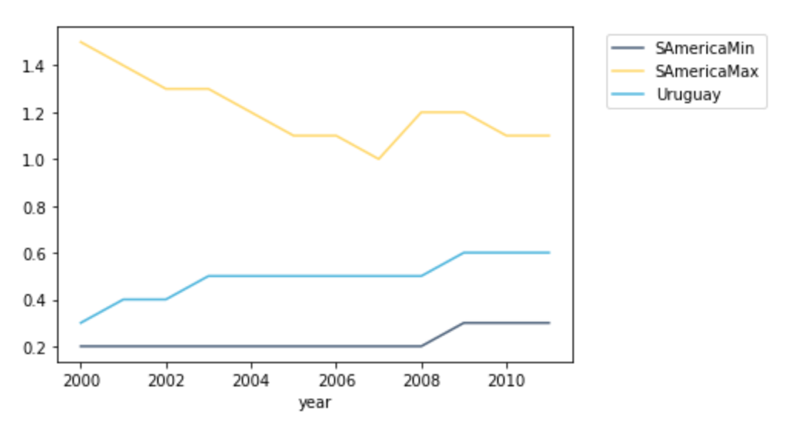
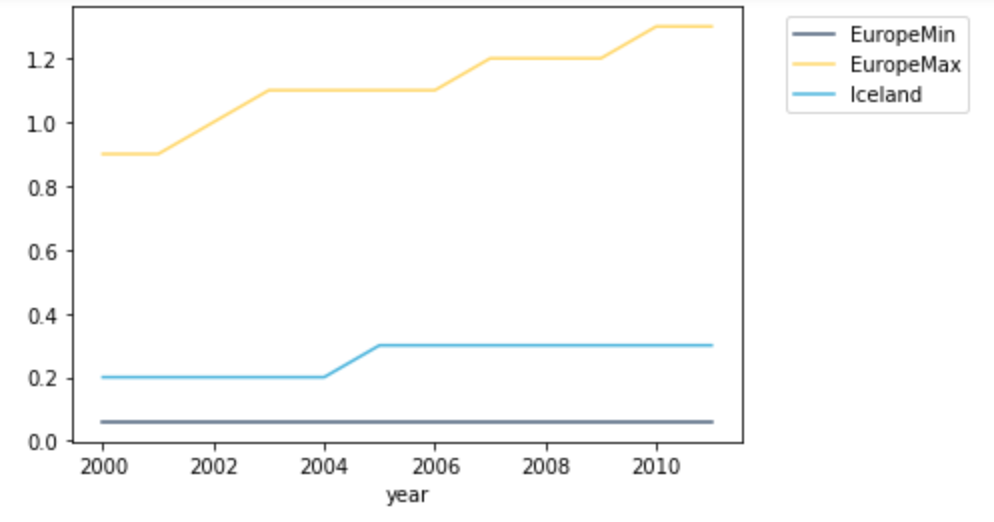
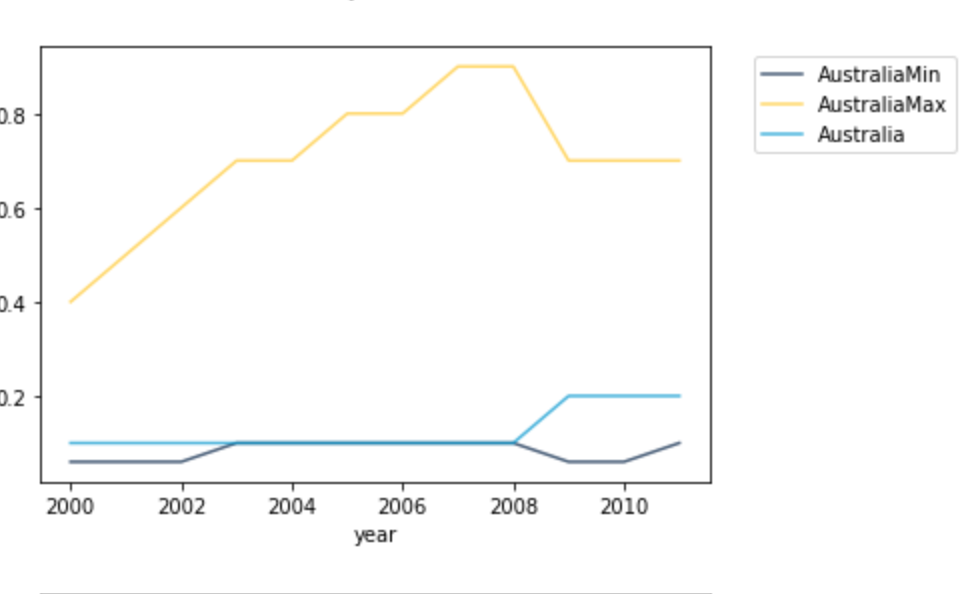
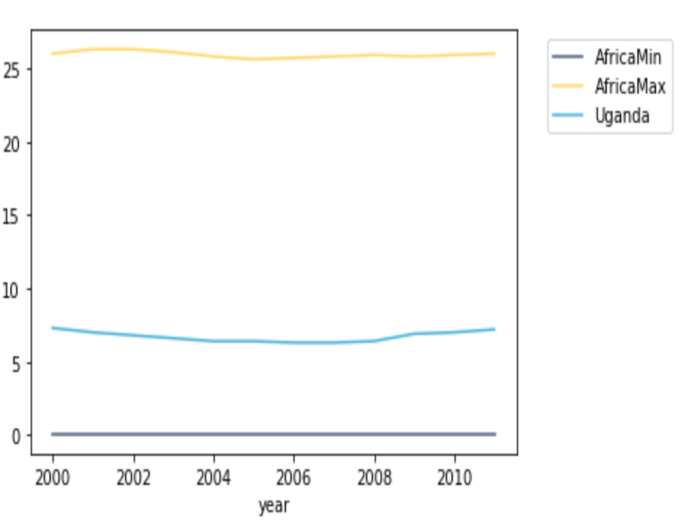
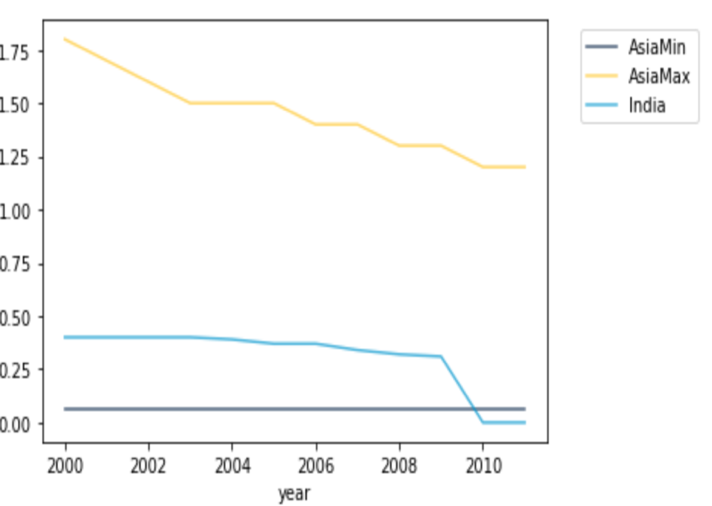
The new table created here is used to plot highest average HIV prevalence per continent. For Highest Average value, it is straight forward. For lowest we need to select only one among the multiple lowest value. This is done by grouping again only based on Continent and selecting min.

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



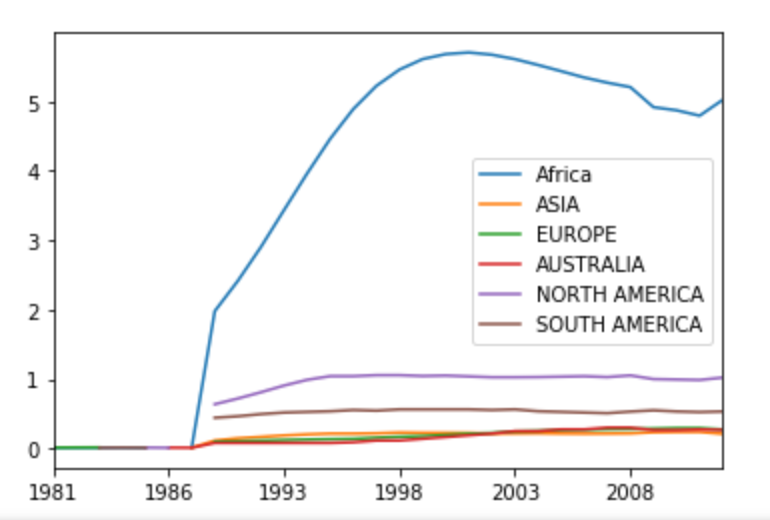
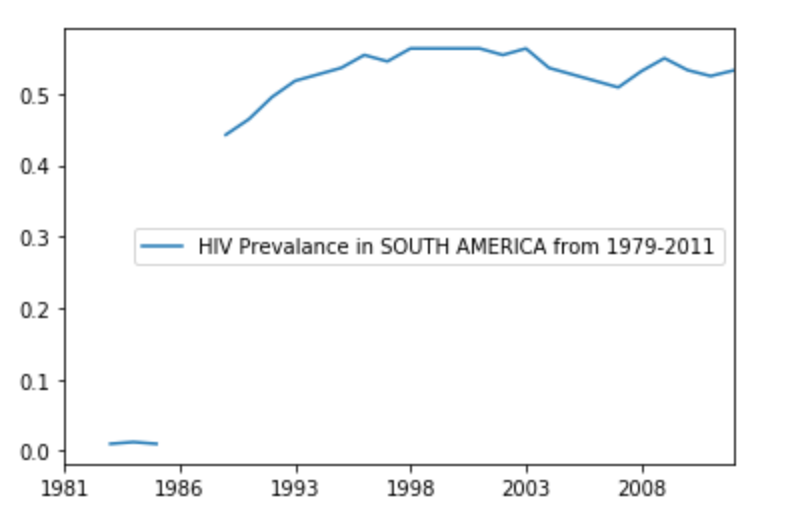
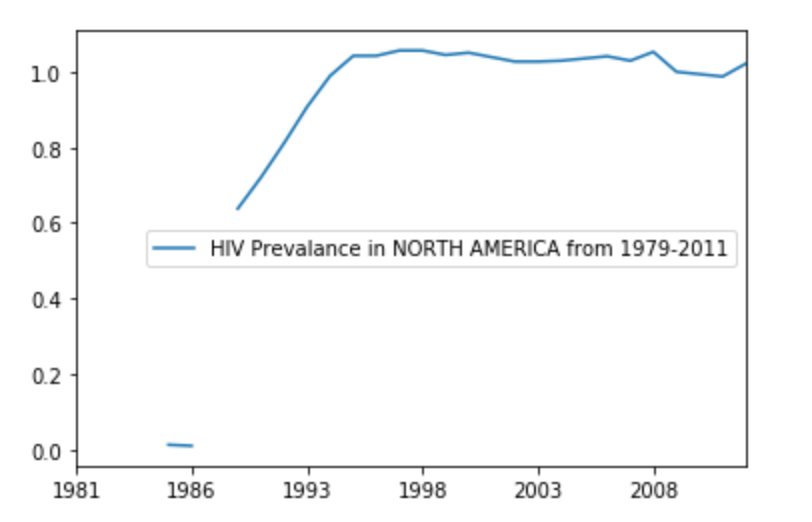
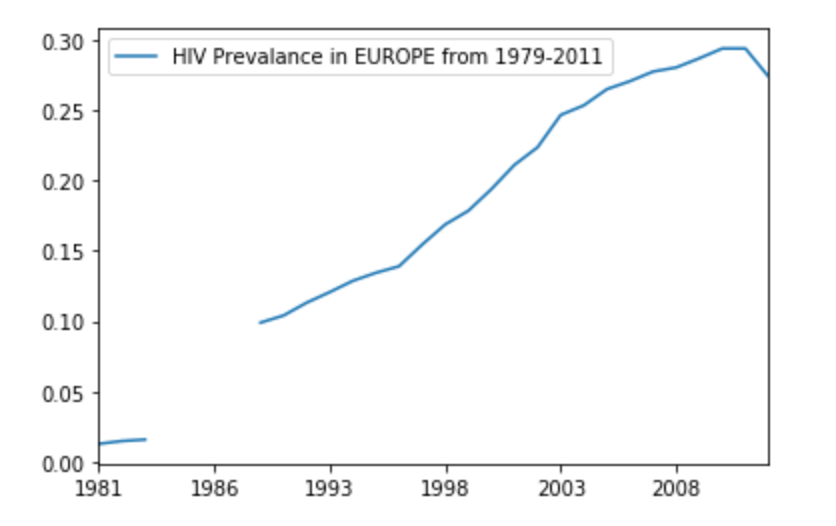
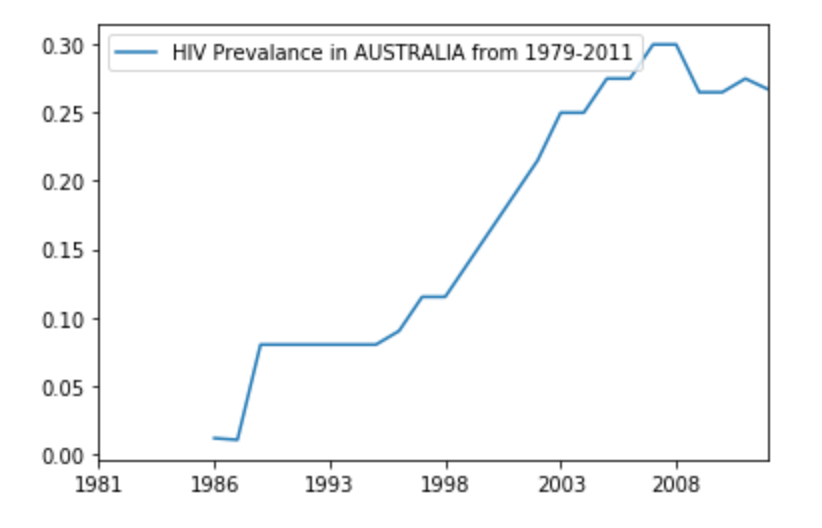
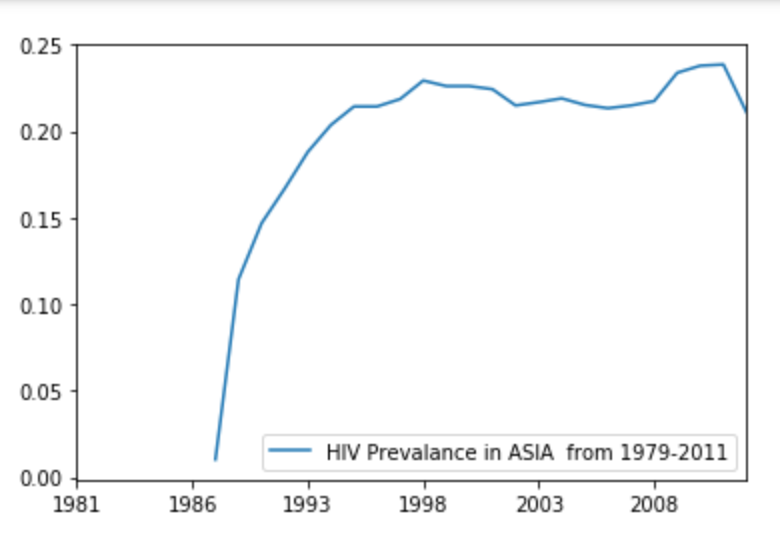
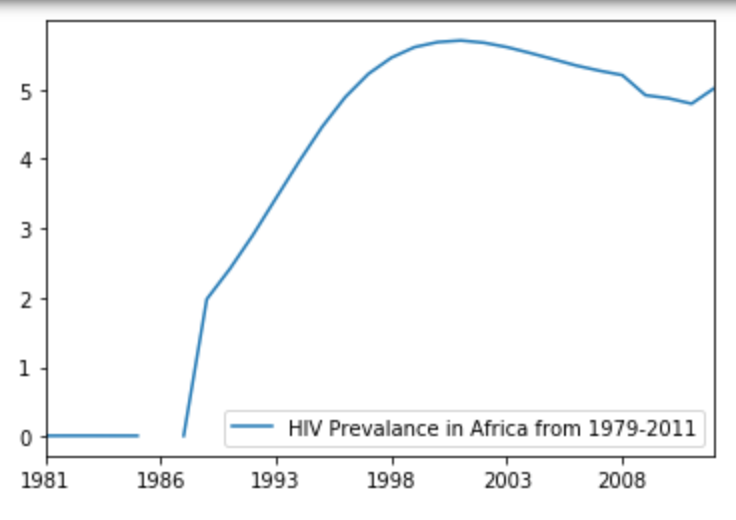
One country is selected from each continent and that country's row is selected with just values from year 2000 to 2011 country, continent, and min/max of each continent columns

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



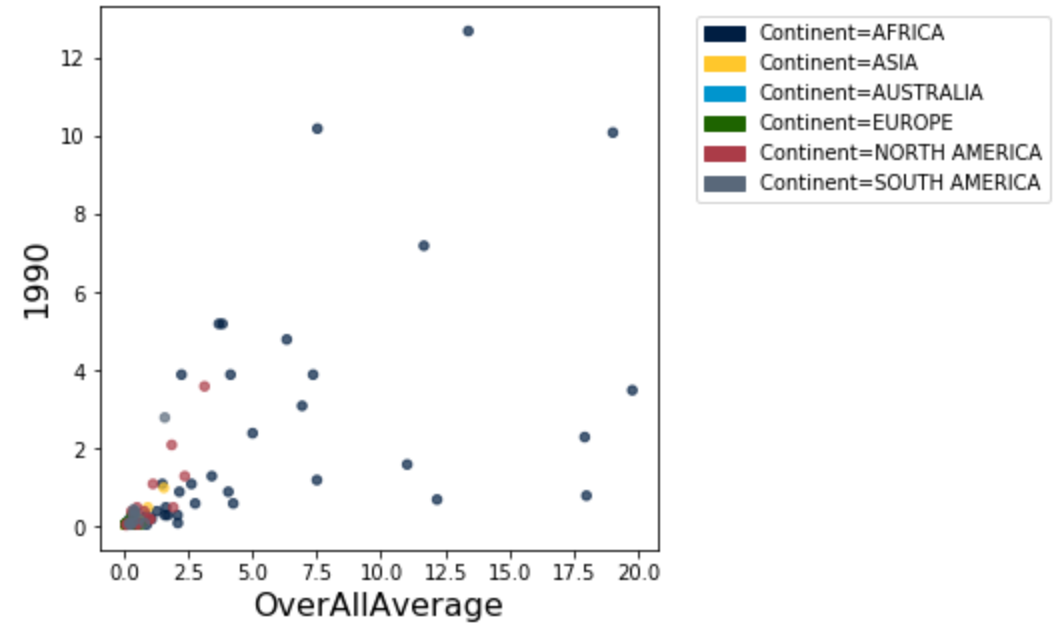
**Part 3**

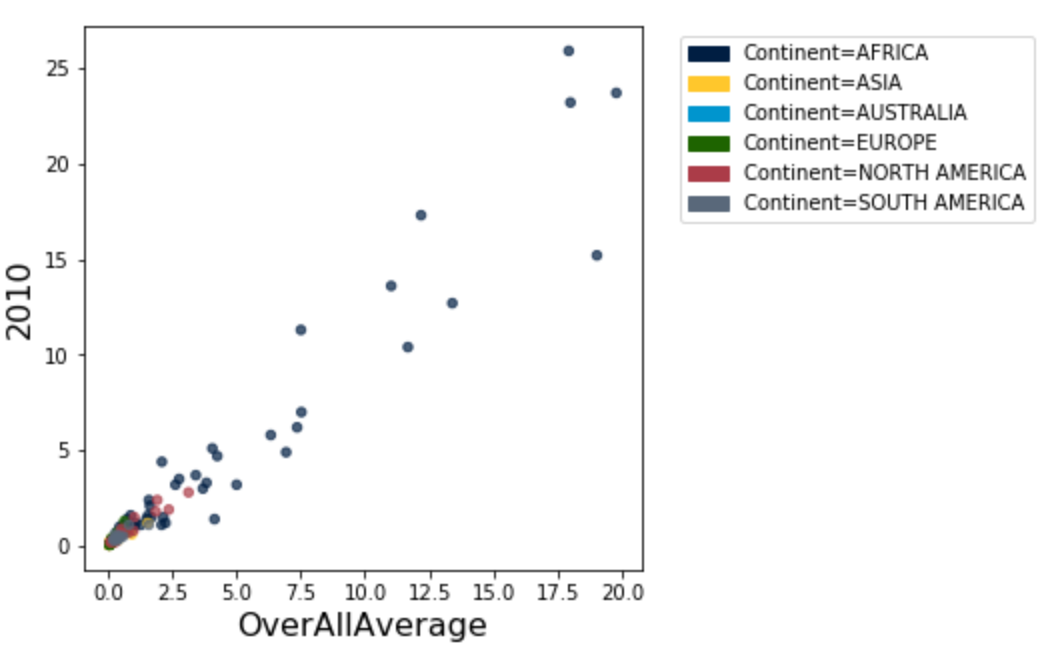
New/Separate data frame is created for each continent and mean is calculated on axis =1 for 1979 to 2011 to get average of each year. For each continent this table is plotted separately and for the overlaid chart, these are plotted in a single plot.



**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.