# INFO 5502 Assignment 2

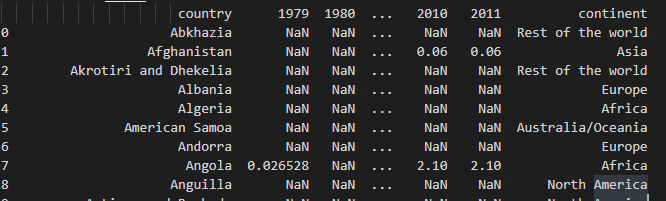
### **Venkata Athreya Uppaluri**

### **11338457**

1. Renamed the excel file with dataset to “ estimated\_hiv\_prevelance\_indicator.xlsx”.
2. Used pandas, pycountry\_convert python libraries to read/write files and assign continents codes based on countries. Below are the steps followed:
3. Imported dataset into python code from below attached file.



1. Using pycountry\_convert library assigned continent names to various countries in the newly created column.



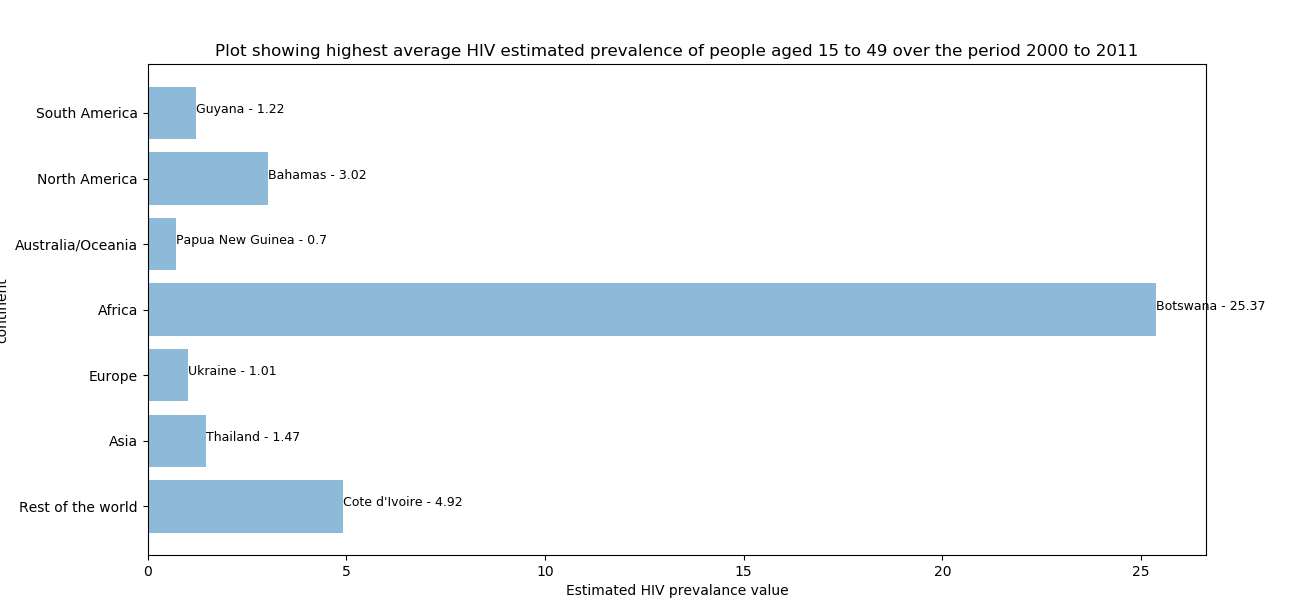
1. Validated for any missing data after modifying the dataset



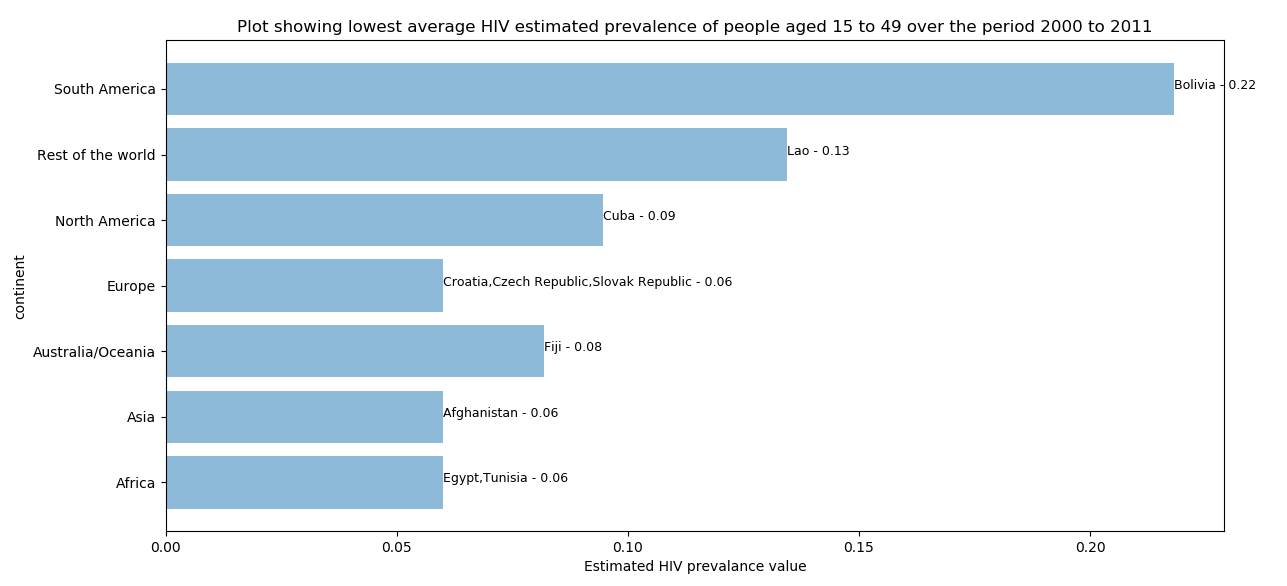
1. Exported the modified dataset as a new csv file.



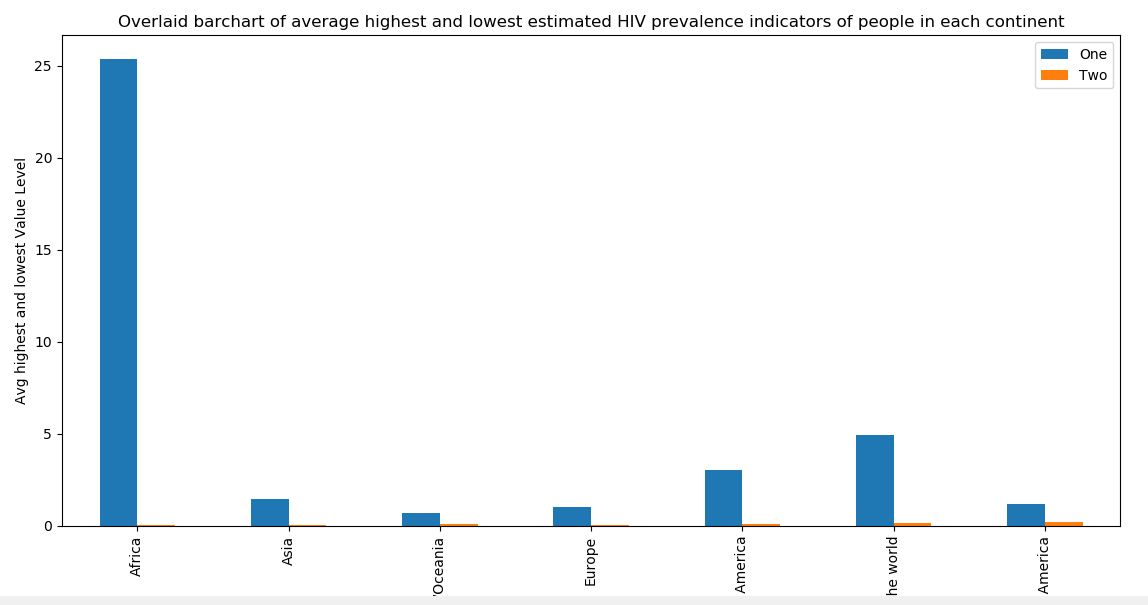
1. Below are the steps in implementing this question
2. Average of HIV prevalence indicator value is calculated for each country.
3. Finding the country with highest HIV prevalence indicator value from each continent which is as given below.



1. Finding the country with lowest HIV prevalence indicator value from each continent which is as given below.



1. Created an overlaid bar chart to show the highest and lowest average HIV estimated prevalence of people ages from 15 to 49 of from year 2000 to 2011 in each continent.



1. Selecting a country from each continent which has a different average HIV prevalence indicator that is not highest and lowest. Plotting an overlaid line chart for the selected Country.

