**Death Mortality Rate Analysis of Different Country**

**By:**

**Ayush Pandey**

**Abstract:**

Death is the end of a life in an organism.We all are blessed with life in us and no matter what we do not like to give up on it at any point of time. However in some situations at some point in our lives, we may be face to face with death and in such situations, it is indeed a battle between life and death.

Early death for people is most commonly due to illnesses caused by smoking, high blood pressure, high sugar, alcohol misuse, less physical activity,etc.

There are various causes of death, so this analysis focuses on the Death Rate of each country due to various risk factors and on what area the people and the government need to work in order to reduce the cause of death due to that particular reason.

**About Data:**

The data set consists of 6,468 record and 32 different features containing the data like Entity, Year, Unsafe water source, Alcohol use and much more of 231 entity’s including Country and Island from the year 1990 to 2017.

**Approach:**

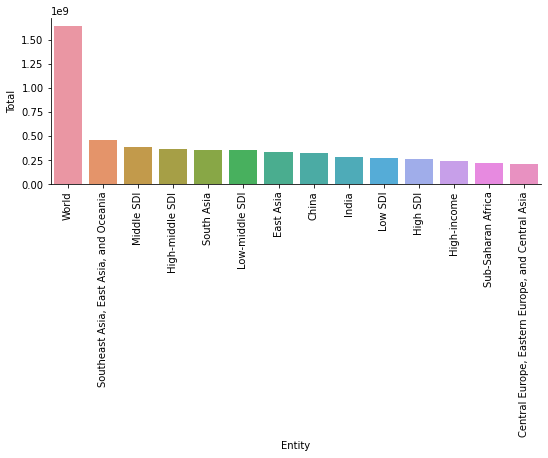
* Various python libraries such as NumPy, Pandas, Matplotlib, Seaborn are used for the purpose of mathematical calculations, extraction and visualization of the data.
* The extracted dataset was found to have null values which are then effectively handled to achieve accurate analysis.
* The datasets to be used for analysis was found to be highly skewed (Right Skewed). Although skewness will not have much effect on descriptive analysis, it is handled for the purpose of predictive analysis to accomplish better decisions from the model.

**ANALYSIS AND VISUALIZATION**

**1)Country where number of death is greater than 20,00,00,000 :**

**Purpose:**

* To get the number of death due to different features such as Unsafe water source, Smoking, Air pollution,etc. for each country .



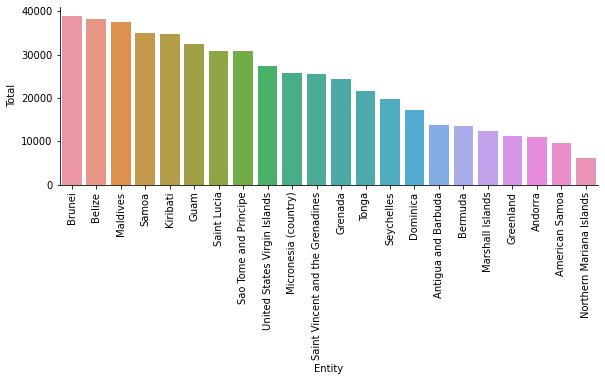
**Inference:**

* From the above visualization it is clear that the country having maximum number of death due to all the factors from the year 1990 to 2017 is in China with more than 320 million (323072368.80) deaths followed by India with more than 280 million (282281596.42) deaths.
* The total number of death worldwide due to all the factors is more than 1.6 billion (1645135262.86).
* The total number of death in Southeast Asia,East Asia and Oceania due to all the factors is more than 460 million (460875240.51).

**2)Country where number of death is less than 50,000 :**

**Purpose:**

* To know which country or Island has recorded the least number of death due to all the different features.



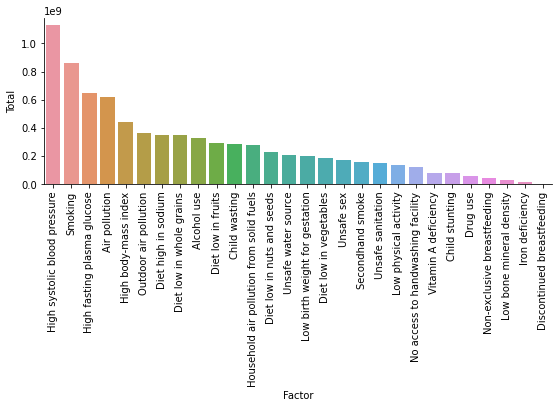
**Inference:**

* From the above visualization it is clear that the number of death recorded from 1990 to 2017 is least in Northern Mariana Islands with less than 6,200 (6141.96) deaths followed by American Samoa with around 9,600 (9659.87) deaths.
* Beacuse of the small size and population of Island as compared to a country, the number of deaths recorded in Island is much less as compared to a country.

**3)Major factors leading to the cause of death worldwide:**

**Purpose:**

* For analysing the total number of death due to all the different factors and to get an idea of the major factors leading to the cause of death worldwide.



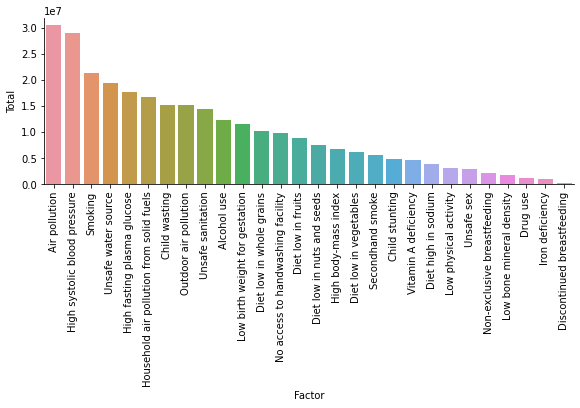
**Inference:**

* From the above visualization it is clear that the major cause of death worldwide is due to High Systolic blood pressure and from 1990 to 2017 more than 1.1 billion (1127919446.38) people have died, followed by Smoking which is the second major cause of death and has taken more than 863 million (863790716.22) lives.

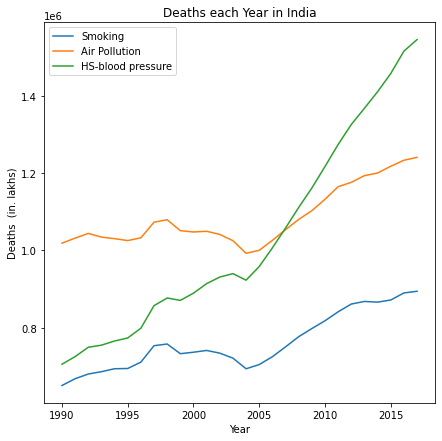
1. **Major factor leading to the cause of death in India:**

Purpose:

* For analysing the total number of death due to all the different factors and to get an idea of the major factors leading to the cause of death in India.



**Figure1**



**Figure 2**

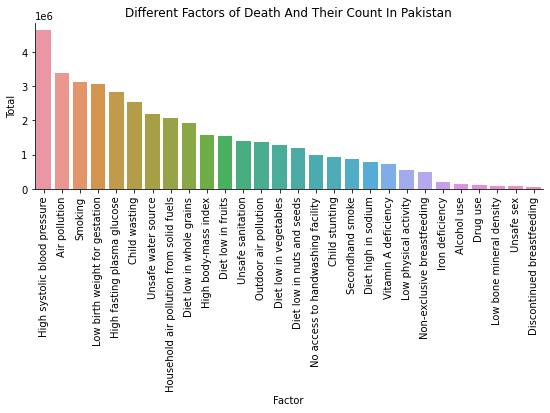
**Inference:**

* From the above visualization it is clear that the major cause of death in India is due to Air Pollution and from 1990 to 2017 around 30 million (30394991.85) people have died, followed by High systolic blood presure which is the second major cause of death and has taken around 29 million (28881676.30) lives.
* From figure 2,it is clear that the rate of increase of death due to High systolic blood pressure has increased exponentially from the year 2005, so major steps has to be taken in order to reduce the death from High systolic blood pressure and Air pollution.

1. **Major factor leading to the cause of death in Pakistan:**

Purpose:

* For analysing the total number of death due to all the different factors and to get an idea of the major factors leading to the cause of death in Pakistan.



**Figure 3**

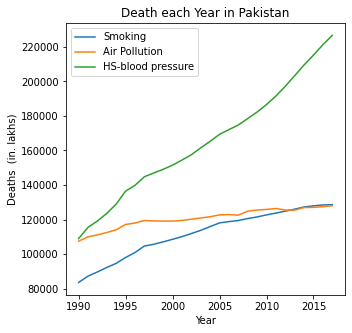


Figure 4

**Inference:**

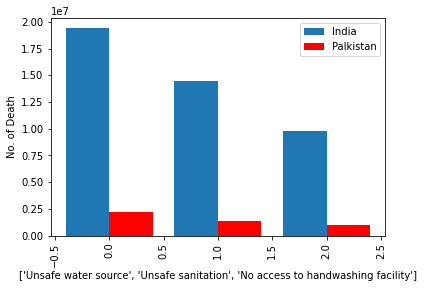
* From the above visualization it is clear that the major cause of death in Pakistan is due to High systolic blood pressure and from 1990 to 2017 more than 4.6 million (4631167.65) people have died, followed by Smoking which is the second major cause of death and has taken more than 3.3 million (3380837.05) lives.
* Both High systolic blood pressure and Air pollution are the main cause of death in Pakistan and from Figure 4, the rate of death due to Smoking is increasing in much faster rate as compared to Air pollution and from the year 2012, the rate of death reported due to Smoking is more as compared to Air pollution.

**6)Comparing the ‘Unsafe water source’, ‘Unsafe sanitation’,**

**‘No access to handwashing facility’ in India And Pakistan**.

**Purpose:**

* To get an idea of the number of deaths in both the countries based on ‘Unsafe water source’, ‘Unsafe sanitation’, ‘No access to handwashing facility factors’.



**Inference:**

* From the above visualization it is clear that the number of death due to ‘Unsafe water source’, ‘Unsafe sanitation’ and ‘No access to handwashing facility’ factors in India is more as compared To Pakistan.
* Number of death in India due to ‘Unsafe water source’ is more than 19 million (19381797.52) and in Pakistan is more than 2.1 million (2180227.56).
* Number of death in India due to ‘Unsafe sanitation’ is more than 14 million (14427951.31) and in Pakistan is more than 1.3 million (1396614.61).
* Number of death in India due to ‘No access to handwashing facility’ is more than 9.7 million (9789742.32) and in Pakistan close to 1 million (988069.29).

**7)Multiple Regression between ‘Unsafe water source’, ‘Unsafe sanitation’, ‘No access to handwashing facility’ worldwide:**

**Purpose:**

* To predict ‘No access to handwashing facility’ worldwide with the help of ‘Unsafe water source’, ‘Unsafe sanitation’.

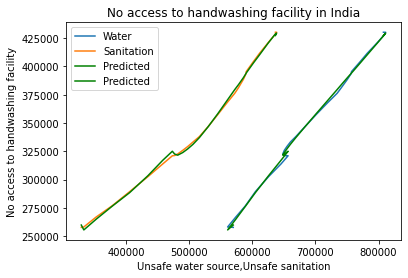
**Inference:**

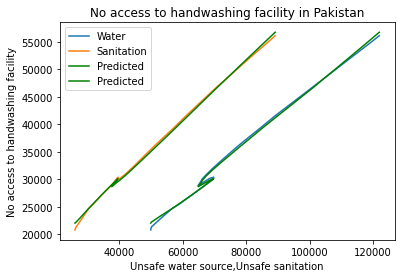
* The regression coefficient of Unsafe water source is 0.297 and for Unsafe sanitation is 0.384 and by this it is clear that the Unsafe sanitation is contributing much to no access to handwashing facility as compared to Unsafe water source that is with 1% decrease in the sanitation facility leads to 38% (38.44%) decrease in access to handwashing facility.
* The performance of the multiple regression model in predicting the No access to handwashing facility worldwide with the help of unsafe water source and unsafe sanitation is 99.11%

**8)Multiple Regression between ‘Unsafe water source’, ‘Unsafe sanitation’, ‘No access to handwashing facility’ for the country India and Pakistan:**

**Purpose:**

* To predict ‘No access to handwashing facility’ in both the countries with the help of ‘Unsafe water source’, ‘Unsafe sanitation’.





**Inference:**

* In India, the regression coefficient of Unsafe water source is 0.539 and for Unsafe sanitation is 0.126 and by this it is clear that the Unsafe water source is contributing much to no access to handwashing facility as compared to Unsafe sanitisation that is with 1% decrease in the water souce purity leads to 54% (53.94%) decrease in access to handwashing facility.
* The performance of the multiple regression model in predicting the No access to handwashing facility in India with the help of Unsafe water source and unsafe sanitation is is 99.87%.
* In Pakistan, the regression coefficient of Unsafe water source is 0.146 and for Unsafe sanitation is 0.383 and by this it is clear that the Unsafe sanitation is contributing much to no access to handwashing facility as compared to Unsafe water source that is with 1% decrease in the sanitation facility leads to 38% (38.34%) decrease in access to handwashing facility.
* The performance of the multiple regression model in predicting the No access to handwashing facility in Pakistan with the help of Unsafe water source and unsafe sanitation is is 99.83%.

**9)Multiple Regression between 'Smoking','High fasting plasma glucose' and ‘High systolic blood pressure’ worldwide:**

**Purpose:**

* The nicotine in cigarette smoke causes rise in blood pressure and heart rate, narrows the arteries and hardens their walls, and makes the blood more likely to clot which can lead to a heart attack or stroke.
* As blood with high glucose levels travels through the body, it can cause widespread damage, including to the blood vessels and kidneys. These organs play a key role in maintaining healthy blood pressure. If they experience damage, blood pressure can rise, increasing the risk of further harm and complications.
* To predict the ‘High systolic blood pressure’ worldwide with the help of ‘Smoking’ and ‘High fasting plasma glucose’.

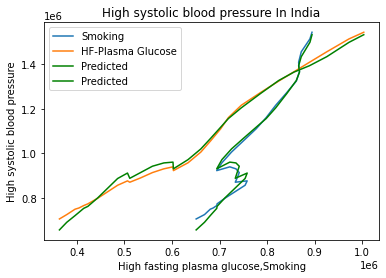
**Inference:**

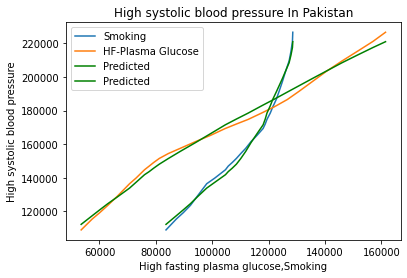
* The regression coefficient of Smoking is 0.685 and for High fasting plasma glucose is 0.832 and by this it is clear that High fasting plasma glucose is contributing much to High systolic blood pressure as compared to Smoking that is with 1% decrease in High fasting plasma glucose rate leads to 83% (83.2%) decrease in High systolic blood pressure.
* The performance of the multiple regression model in predicting the High systolic blood pressure worldwide with the help of Smoking and High fasting plasma glucose is 99.10%.

**10)Multiple Regression between 'Smoking','High fasting plasma glucose' and ‘High systolic blood pressure’ for the country India and Pakistan**

**Purpose:**

* To predict the ‘High systolic blood pressure’ in both the countries with the help of ‘Smoking’ and ‘High fasting plasma glucose’.





**Inference:**

* In India, the regression coefficient of Smoking is 1.124 and for High fasting plasma glucose is 0.943 and by this it is clear that Smoking is contributing much to High systolic blood pressure as compared to High fasting plasma glucose that is with 1% decrease in the Smoking rate leads to 112% (112.4%) decrease in High systolic blood pressure.
* The performance of the multiple regression model in predicting the High systolic blood pressure in India with the help of Smoking and High fasting plasma glucose is 99.34%.
* In Pakistan, the regression coefficient of Smoking is 0.589 and for High fasting plasma glucose is 0.761 and by this it is clear that High fasting plasma glucose is contributing much to High systolic blood pressure as compared to Smoking that is with 1% decrease in High fasting plasma glucose rate leads to 76% (76.10%) decrease in High systolic blood pressure.
* The performance of the multiple regression model in predicting the High systolic blood pressure in Pakistan with the help of Smoking and High fasting plasma glucose is 99.18%.

**11)Multiple Regression between 'Vitamin A deficiency','Child Stunting','Low birth weight for gestation' and ‘Child Wasting’ worldwide:**

**Purpose:**

* Children who are thin for their height is because of acute food shortages, deficiency and diseases.
* Deficiency of Vitamin A and Low birth weight for gestation are associated with significant morbidity and mortality from common childhood infections, and is one of the major cause for Child wasting.
* To predict the low weight for height (Child Wasting) with the help of 'Vitamin A deficiency','Child stunting' and 'Low birth weight for gestation'.

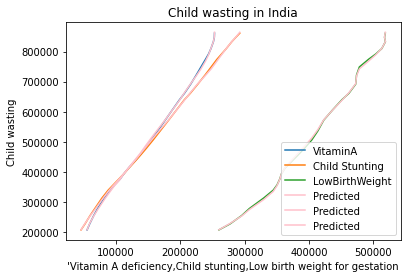
**Inference:**

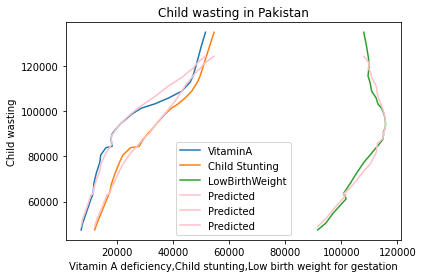
* The regression coefficient of Vitamin A deficiency is 1.757, for Child stunting is 0.717 and for Low birth weight for gestation is 0.442,and by this it is clear that the Vitamin A deficiency is contributing much to low weight for height as compared to Child Stunting and Low birth weight for gestation that is with 1 unit downfall in death due to deficiency in Vitamin A leads to 1.75 unit decrease in death due to Child wasting.
* The performance of the multiple regression model in predicting the Child wasting worldwide with the help of child stunting and Low birth weight for gestation is 99.71%.

**12)Multiple Regression between 'Vitamin A deficiency','Child Stunting','Low birth weight for gestation' and ‘Child Wasting’ for the country India and Paksitan:**

**Purpose:**

* To predict the low weight for height (Child Wasting) in both the countries with the help of 'Vitamin A deficiency','Child stunting' and 'Low birth weight for gestation'.





**Inference:**

* For India, the regression coefficient of Vitamin A deficiency is -0.40, for Child stunting is 2.567 and for Low birth weight for gestation is 0.397,and by this it is clear that Child stunting is contributing much to low weight for height as compared to Vitamin A deficiency and Low birth weight for gestation that is with one unit change in Child stunting leads to 2.57 unit change in death rate due to Child wasting.
* The performance of the multiple regression model in predicting the Child wasting in India with the help of child stunting and Low birth weight for gestation is 99.97%.
* For Pakistan, the regression coefficient of Vitamin A deficiency is 0.722, for Child stunting is 0.698 and for Low birth weight for gestation is 0.826,and by this it is clear that Low birth weight for gestation is contributing much to low weight for height as compared to Vitamin A deficiency and child stunting that is with 1 unit change in Low birth weight for gestation leads to 0.826 unit change in death rate due to Child wasting.
* The performance of the multiple regression model in predicting the Child wasting in Pakistan with the help of child stunting and Low birth weight for gestation is 97.79%.

**13)Multiple Regression between Household air pollution from solid fuels,Smoking,Outdoor air pollution and Air Pollution worldwide:**

**Purpose:**

* Cigarette smoking causes environmental pollution by releasing toxic air pollutants into the atmosphere. The cigarette butts also litter the environment, and the toxic chemicals in the residues seep into soils and waterways, thereby causing soil and water pollution, respectively.
* Most indoor air pollution comes from sources that release gases or particles into the air. Things such as building materials and air fresheners give off pollution constantly. Other sources such as tobacco smoke and wood-burning stoves also cause air pollution..
* To predict the Air pollution with the help of ‘Household air pollution from solid fuels’, ’Smoking’, ‘Outdoor air pollution’.

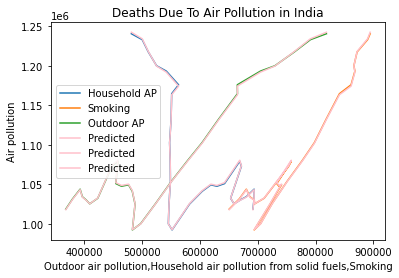
**Inference:**

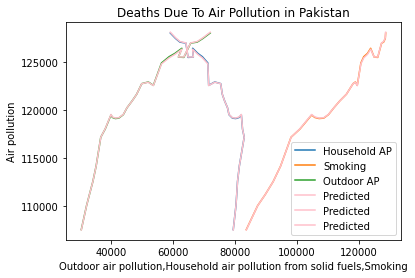
* The regression coefficient of Household air pollution from solid fuelsdeficiency is 0.966, for Smoking is 0.011 and for Outdoor air pollution is 0.938,and by this it is clear that the ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ are contributing much to Air pollution as compared to Smoking that is with 1% decrease in ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ leads to 96.6% and 93.8% decrease in Air pollution respectively.
* The performance of the multiple regression model in predicting the Air pollution worldwide with the help of ‘Household air pollution from solid fuels’, ‘Smoking’, ‘Outdoor air pollution’ is 99.98%.

**14)Multiple Regression between Household air pollution from solid fuels,Smoking,Outdoor air pollution and Air Pollution for the country India and Pakistan:**

**Purpose:**

* To predict the Air pollution in both the countries with the help of ‘Household air pollution from solid fuels’, ’Smoking’, ‘Outdoor air pollution’.





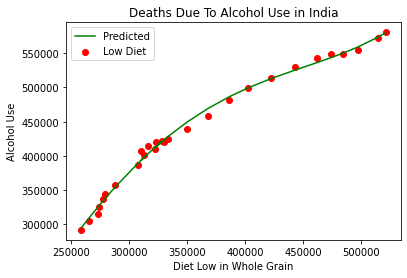
**Inference:**

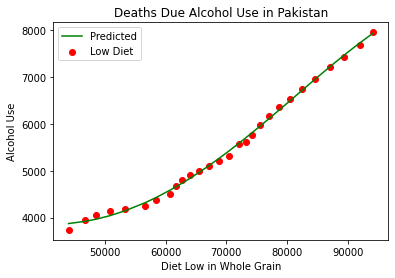
* For India, the regression coefficient of Household air pollution from solid fuelsdeficiency is 0.966, for Smoking is -0.019 and for Outdoor air pollution is 0.959,and by this it is clear that the ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ are contributing much to Air pollution as compared to Smoking that is with 1% decrease in ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ leads to 96.6% and 96% decrease in Air pollution respectively.
* The performance of the multiple regression model in predicting the Air pollution in India with the help of ‘Household air pollution from solid fuels’, ‘Smoking’, ‘Outdoor air pollution’ is 99.98%.
* Air pollution is the major cause of death in India so steps has to be taken in order to reduce the Air pollution.
* For Pakistan, the regression coefficient of Household air pollution from solid fuelsdeficiency is 0.967, for Smoking is 0.020 and for Outdoor air pollution is 0.946,and by this it is clear that the ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ are contributing much to Air pollution as compared to Smoking that is with 1% decrease in ‘Household air pollution from solid fuels’ and ‘Outdoor air pollution’ leads to 96.7% and 94.6% decrease in Air pollution respectively.
* The performance of the multiple regression model in predicting the Air pollution in Pakistan with the help of ‘Household air pollution from solid fuels’, ‘Smoking’, ‘Outdoor air pollution’ is 99.99%.

**15)Polynomial Regression between Diet Low in whole grains and Alcohol use for the country India and Pakistan:**

**Purpose:**

* Many alcoholic people are malnourished,as they ingest too little of essential nutrients (e.g., carbohydrates, proteins). As a result, alcoholic people frequently experience deficiencies in proteins.
* To predict the death due to Alcohol use in both the countries using ‘Diet low in whole grains’ data.





**Inference:**

* The performance of the polynomial regression model in prediting the death due to Alcohol use in India with the help of ‘Diet low in whole grains’ is 99.55%.
* The performance of the polynomial regression model in prediting the death due to Alcohol use in Pakistan with the help of ‘Diet low in whole grains’ is 99.74%.

**Conclusion:**

* The total number of death worldwide due to all the factors is more than 1.6 billion.
* Maximum number of death due to all the factors is in China with more than 320 million (323072368.80) deaths followed by India with more than 280 million (282281596.42) deaths.
* The major cause of death worldwide is due to High Systolic blood pressure and from 1990 to 2017 more than 1.1 billion people have died, followed by Smoking which is the second major cause of death and has taken more than 863 million lives.
* The major cause of death in India is due to Air Pollution and from 1990 to 2017 around 30 million people have died.
* In India, Household air pollution from solid fuels and Outdoor air pollution are contributing much to Air pollution, so more efforts has to be done to improve the Household air pollution from solid fuels and Outdoor air pollution in order to reduce the number of death due to Air pollution.
* The number of death due to ‘Unsafe water source’, ‘Unsafe sanitation’, ‘No access to handwashing facility’ factors in India is more as compared to Pakistan.It is because of India’s large size and population as compared to Pakistan.
* In India, unsafe water affect more on access to handwash as compared to sanitation. So in order to increase the handwashing facility in India more efforts has to be done to improve the water sources.
* In Pakistan, improper sanitation affect more on access to handwash as compared to water sources. So in order to increase the handwashing facility in Pakistan more efforts has to be done to improve the sanitation process in the country.
* In India, Smoking is contributing more to High systolic blood pressure as compared to High fasting plasma glucose. So in order to decrease the death rate due to High systolic blood pressure in India more efforts has to be done to reduce the use of Smoking.
* In Pakistan,High fasting plasma glucose is contributing more to High systolic blood pressure as compared to Smoking that is with. So in order to decrease the death rate due to High systolic blood pressure in Pakistan more efforts has to be done to reduce High fasting plasma glucose.