Title of the Project

Cause of Death

Submitted by

Name of the Candidate: Muskan Sureka

Internship Batch Number: 34

Supervised by

Name of the SME: Khushboo Garg

Month & Year of Submission: February,2023.

Student's Declaration

I hereby declare that the Project Work with the title (in block letters) "CAUSE OF DEATH" submitted by me for the project allocated to Internship batch no.: 34 by FLIPROBO TECHNOLOGIES as a part of my internship phase of my PG DIPLOMA COURSE OF DATA SCIENCE AND NEURAL NETWORKS BY DATATRAINED INSITITUE is my original work and has not been submitted earlier to any other Institution for the fulfilment of the requirement for any course of study.

I also declare that no chapter of this manuscript in whole or in part has been incorporated in this report from any earlier work done by others or by me. However, extracts of any literature which has been used for this report has been duly acknowledged providing details of such literature in the references.

Signature: Muskan Sureka

Name: Muskan Sureka

Address: Alipore Residency,
3 Burdwan Road

Place: Kolkata

Internship batch No.: 34

Date: 09/02/2023

ACKNOWLEDGEMENT

The success of the project is the result of hard work and endeavor of not only one but rather numerous people. It is the outcome of the exceptional support of many individuals.

Firstly, I take immense pleasure in thanking DataTrained Institute and Flip Robo Technologies for furnishing me with a chance to conduct a research project by making it a part of the curriculum.

I also express my sincere indebtedness and profound gratitude to my teachers: Dr. Deepika Sharma, Mr. Ravikesh Pandey and my SME: Khushboo Garg.

I would also like to express my deep sense of gratitude to my project SME Khushboo Garg for her continuous guidance and support, which has helped me tremendously to complete this project with the best of my abilities. Without her help, completing this research would not have been possible.

Also, I would like to express my sincere gratitude towards my friends and family for providing me with valuable and significant ideas throughout the course of the project.

I hence express my sincere gratitude and appreciation to all the individuals who have helped in any possible way and contributed in any manner. Their constant support and assistance has been significant and of great value.

INDEX

1. INTRODUCTION

1.1 Introduction

Around 56 million people die each year.

What caused their death? How did the causes of death change over time and differ between different countries and world regions? And what are the risk factors that lead to early death? These are the big questions we are answering here.

Definitions: Cause of death vs risk factors

It is important to understand what is meant by the *cause* of death and the *risk factor* associated with a premature death:

In the epidemiological framework of the Global Burden of Disease study each death has *one* specific cause. In their own words: 'each death is attributed to a single underlying cause — the cause that initiated the series of events leading to death'.[2]

This is different from the deaths that happened due to risk factors. These deaths are an estimation of the reduction of the number of deaths that would be achieved if the risk factors to which a population is exposed would be eliminated (in the case of tobacco smoking, for example) or reduced to an optimal, healthy level (in the case of body-mass index).

What do people die from?

56 million people died in 2017. What did they die from?

The *Global Burden of Disease* is a major global study on the causes of death and disease published in the medical journal *The Lancet*.[4] These estimates of the annual number of deaths by cause are shown here.

This is shown for deaths worldwide. But you can explore data on the annual number of deaths by cause for any country or region using the "change country" toggle.

Non-communicable diseases (NCDs) not only dominate mortality figures at a global level, but also account for the majority of deaths in high-income countries.

Deaths from causes such as infectious disease, malnutrition, nutritional deficiencies, neonatal and maternal deaths are common – and in some cases dominant – across low- and middle-income nations. In Kenya, for example, the leading cause of death remains diarrheal diseases. In South Africa and Botswana, the leading cause of death is HIV/AIDS. In high-income countries however the share of deaths caused by these is very low.

Using the timeline on the chart you can also explore how deaths by cause have changed over time.

Death rates related to disease, illness and other health factors tend to change relatively slowly over time. Whilst death rates may fall or decline from year-to-year as part of a general trend, dramatic changes in such deaths are typically rare. Natural disaster and terrorism-related deaths are an important exception to this rule, as they can vary significantly between countries. This can make the annual comparison of deaths and death rates between health-related factors and volatile events more challenging. Understanding the relative risk of these events can require a longer-term overview of high and low-mortality years.

1.2 Cause of death by Category

The share of deaths from infectious diseases are declining; a larger share is dying from NCDs

In the visualization we see the distribution of global deaths broken down by three broad categories:

- 1 – in yellow: Injuries caused by road accidents, homicides, conflict deaths, drowning, fire-related accidents, natural disasters and suicides.
- 2 – in blue: Non-communicable diseases. These are often chronic, long-term illnesses and include cardiovascular diseases (including stroke), cancers, diabetes and *chronic* respiratory diseases (such as chronic pulmonary disease and asthma, but excluding infectious respiratory diseases such as tuberculosis and influenza).
- 3 – In red: Communicable diseases (i.e. infectious diseases) such as HIV/AIDS, malaria, and tuberculosis together with maternal deaths, neonatal deaths and deaths from malnutrition.

This is shown for global deaths as the default, but can be viewed for any country or region using the "change country" toggle on the interactive chart.

At a global level we see that the majority of deaths are caused by non-communicable diseases (NCDs). Collectively NCDs account for more than 73% of global deaths. As the world is making progress in the fight against many infectious diseases, and as populations age, we expect that NCDs will become increasingly dominant as the cause of death.

1.3 Cause of death by Age

Breakdown of deaths by age

Fewer people die at a young age

In this chart we see the breakdown of deaths by age bracket. Globally fewer and fewer people die at a young age.

In 2017, there were 56.5 million deaths globally; just over half of these were people who were 70 years or older; 26% were between 50 and 69 years old; 13% were between 15 and 49; only 1% were older than 5 and younger than 14; and almost 9% were children under the age of 5.

The age at which people die has changed significantly since 1990. Fewer people die at a young age. In 1990 nearly one-quarter of all deaths were in children younger than 5. In 2019, this had declined to just under 9%. In contrast, the share of deaths in the over-70s age bracket has increased from a third to half of all deaths over this period.

It is possible to change this chart to any other country or region in the world. In countries with good health the share dying at a young age is very low. In Japan more than 85% are 70 years or older.

Causes of deaths of children younger than 5

This chart shows the number of deaths in children under 5 years old by cause.

Through the combination of neonatal (newborn infants less than 28 days old) disorders, infections and congenital (from birth) defects, we see that the largest share of deaths in under-5s arises from complications at birth or in the first few weeks of life. Under-5s are also highly susceptible to lower respiratory infections, infectious diseases, diarrheal infections, malnutrition and nutritional deficiencies.

This is shown for deaths worldwide. But you can explore data on the annual number of deaths by cause for any country or region using the “change country” toggle.

Death rates in under-5s are typically much lower in high-income countries, and the nature of these deaths is different from lower incomes. In the United Kingdom, for example, child deaths tend to be highly dominated by neonatal complications. Deaths from infectious and diarrheal diseases and malnutrition is very low. In contrast, infectious diseases and nutritional deficiencies are large causes of death in lower-income countries.

Causes of deaths for children between 5 and 14

This visualization shows the causes of deaths of children who died between the age of 5 and 14 year.

Globally, deaths in the 5-14 year old age bracket account for a small percentage of the total (1-2%).

There are six dominant causes of deaths in this age category. The leading causes globally in 5-14 year olds are road accidents, cancers and malaria. Lower respiratory infections, HIV/AIDS, diarrheal diseases, and drowning are all dominant causes typically in the range of 40,000-50,000 deaths in 2017.

Again, this distribution varies by country. In the United States, for example, cancers are the leading cause of death. In India, it's diarrheal diseases; in Bangladesh and China it's drowning; and in South Africa HIV/AIDS.

Causes of deaths for 15 to 49 year olds

This visualization shows the causes of deaths of those who died between the age of 15 and 49.

In the 15 to 49 years old category, we see that non-communicable diseases (NCDs) begin to become dominant. Globally the leading cause of death in this age group is cardiovascular disease, followed cancers which both account for more than one million deaths. Road accidents, HIV/AIDS and suicide are all significant within this group.

For some countries, such as South Africa, by far the dominant cause of death is HIV/AIDS in 15 to 49 year olds. In a number of countries (in particular across Latin America, including Brazil and Mexico), homicide is the dominant cause for 15-49 years old.

Causes of deaths for 50 to 69 year olds

This visualization shows the causes of deaths of those who died between the age of 50 and 69.

In 50 to 69 year olds, non-communicable diseases (NCDs) are strongly dominant — here cardiovascular disease, cancers, respiratory disease and diabetes are the top causes. With the exception of HIV/AIDS and tuberculosis which for some countries climb into the top causes, the global variability in death causes for 50-69 year olds is much lower than that of younger age categories.

Causes of deaths for people who were older than 69 years

This visualization shows the causes of death of people who were 70 years and older at the time of their death.

For the oldest age category (70 years and older), non-communicable diseases (NCDs) still dominate, however other death causes including Alzheimer's/dementias, and diarrheal diseases also become dominant. Diarrheal diseases remain within the few leading causes of deaths in 70+ year olds for many low-income countries, despite being relatively low at higher incomes.

Risk factors for death

It is important to understand what is meant by the *cause* of death and the *risk factor* associated with a premature death:

In the epidemiological framework of the Global Burden of Disease study each death has *one* specific cause. In their own words: 'each death is attributed to a single underlying cause — the cause that initiated the series of events leading to death'.

This is different from the deaths that happened due to risk factors. These deaths are an estimation of the reduction of the number of deaths that would be achieved if the risk factors to which a population is exposed would be eliminated (in the case of tobacco smoking, for example) or reduced to an optimal, healthy level (in the case of body-mass index). Risk factors can be grouped into four broad categories: behavioral risks, environmental risks, occupational risks, and metabolic risks.

All of these estimates are developed independently. This means that we cannot sum all 'attributed deaths' and conclude that this is the actual number of deaths. The attributed number of deaths by risk factor in many cases exceeds that of those by cause of death.

Below, in our section on Measurement, we describe in more detail how the epidemiologists of the Global Burden of Disease study attribute risk factors to mortality.

The number of deaths by risk factor

The estimates shown in this visualization show the numbers of deaths attributed to specific risk factors in 2017.

Here we see that there are several dominant risk factors for death: notably, those related to dietary and activity lifestyle factors (including blood pressure, physical activity, body-mass index, blood sugar, and dietary intake); smoking; air pollution (both outdoor and indoor); environmental factors including clean water and sanitation; and safe sex (for the prevention of HIV/AIDS).

This is shown for deaths worldwide. But you can explore data on the annual number of deaths by cause for any country or region using the "change country" toggle. The contribution of specific risk factors varies significantly by country.

For most high-income countries, the dominant risk factors are those related to healthy diets, smoking and alcohol intake. Other risk factors such as clean water, sanitation, and child wasting or stunting are very low. In low-income countries the inverse is true: in Sierra Leone for example, the top risk factors include child wasting, household air pollution, unsafe water source, poor sanitation, and the lack of access to

handwashing facilities. For countries where HIV/AIDS is a major health burden, such as South Africa and Kenya, unsafe sex is the top risk factor.

Cardiovascular diseases

Cardiovascular disease (CVD) is a term used to refer to the range of diseases which affect the heart and blood vessels. These include hypertension (high blood pressure); coronary heart disease (heart attack); cerebrovascular disease (stroke); heart failure; and other heart diseases.

Cardiovascular disease is the top cause of death globally.

In the map we see death rates from cardiovascular diseases across the world.

Overall we see a strong East-West divide in CVD death rates. Rates across North America and Western/Northern Europe tend to be significantly lower than those across Eastern Europe, Asia and Africa. Across most of Latin America, these rates are moderate. In France, for example, the age-standardized rate was around 86 per 100,000 in 2017; across Eastern Europe this rate was around 5 times higher at 400-500 per 100,000. At the highest end of the scale, Uzbekistan had a rate of 724 per 100,000.

Cancers

Cancers are defined by the National Cancer Institute as a collection of diseases in which abnormal cells can divide and spread to nearby tissue. Cancers can arise in many parts of the body – leading to a range of cancer types, as shown below – and in some cases spread to other parts of the body through the blood and lymph systems.

Dementia

Dementia comprises several forms — the most common being Alzheimer's disease — is an illness which results in a deterioration of cognitive capacity and function beyond what is expect from the normal ageing process. It can occur either in a chronic or progressive form. It affects several cognitive functions including memory, comprehension, judgement, language and learning capacity.

In the map we see death rates from dementia across the world. Note that these rates have been age-standardized which aims to correct for differences in the age structure of a population (which are different between countries and change over time). This therefore allows us to compare the likelihood that any given individual will die from dementia across countries and through time.

Across most countries, the death rate from dementia-related illness is below 55 per 100,000 individuals. Dementia rates in some countries have changed slightly since 1990, but significantly less so than other disease burdens.

Diarrheal diseases

Diarrheal diseases are caused primarily by viral and bacterial pathogens. They are particularly dominant at lower incomes where there is poor access to safe sanitation, drinking water and hygiene facilities. Diarrheal diseases are a leading cause of death in children.

Tuberculosis

Tuberculosis (TB) is an illness caused by the ingestion of bacteria (Mycobacterium tuberculosis) which affects the lungs. The World Health Organization (WHO) estimate that up to one-quarter of the global population has latent TB, meaning they have been infected with the disease but are not ill with the disease (although this does not inhibit it from becoming active in the future).

People with compromised immune systems, such as those suffering from malnutrition, diabetes, or are smokers are more likely to become ill with TB. There is a strong link between HIV/AIDS and TB: those infected with HIV are 20-30 times more likely to develop active tuberculosis.

Across most countries, the death rate from TB is below 5 per 100,000. Rates in 2017 across Eastern Europe were slightly higher, between 5-10 per 100,000. Across South Asia, these reach 25-50 per 100,000, with highest rates across Sub-Saharan Africa ranging from 50 to over 250 per 100,000.

Malnutrition

Malnutrition arises in various forms, with the broad definition capturing undernourishment, micronutrient deficiencies and obesity. In this case, we refer to 'protein-energy malnutrition' (PEM) which refers to energy or protein deficiency caused by insufficient food intake. Protein-energy deficiency can also be exacerbated by infection or disease, which can have the effect of increasing nutritional needs, and/or reducing the body's ability to retain energy or nutrients. You can find more information on hunger and undernourishment in our entry.

The highest rates are seen across across Sub-Saharan Africa, which are typically in the range of 10-100 per 100,000 individuals. For most countries, this rate is below 5 per 100,000. In North Korea during its famine period, rates reached over 400 per 100,000.

HIV/AIDS

An infection with HIV (human immunodeficiency virus) can lead to AIDS (acquired immunodeficiency syndrome). AIDS results in a gradual and persistent decline and failure of the immune system, resulting in heightened risk of life-threatening infection and cancers.

In the majority of cases, HIV is a sexually-transmitted infection. However, HIV can also be transmitted from a mother to her child, during pregnancy or childbirth, or

through breastfeeding. Non-sexual transmission can also occur through the sharing of injection equipment such as needles.

Most countries have a rate of less than 10 deaths per 100,000 – often much lower, below 5 per 100,000. Across Europe the death rate is less than one per 100,000.

Across Sub-Saharan Africa the rates are much higher. Most countries in the South of the region had rates greater than 100 per 100,000. In South Africa and Mozambique, it was over 200 per 100,000.

Malaria

Malaria is a disease that is transmitted from person to person by infected mosquitoes. The bite of an infected Anopheles mosquito transmits a parasite that enters the victim's blood system and travels into the person's liver where the parasite reproduces. There the parasite causes a high fever that involves shaking chills and pain. In the worst cases malaria leads to coma and death.

Smoking

Tobacco smoking is not a direct cause of death, but it nonetheless one of the world's largest health problems.

Smoking is one of the world's leading risk factors for premature death. Tobacco a risk factor for several of the world's leading causes of death, including lung and other forms of cancer, heart disease, and respiratory diseases.

Suicide

Every suicide is a tragedy. With timely, evidence-based interventions, suicides can be prevented.

Homicides

Intentional homicides are defined as "an unlawful death deliberately inflicted on one person by another person".[7] Civilian and military deaths during interstate wars, civil wars and genocides are not counted as homicides – but *Our World in Data* presents the evidence on deaths in the linked articles.

Natural disasters

Natural disasters can occur in many forms – ranging from earthquakes and tsunamis, to extreme weather events, and heatwaves.

The largest disaster events are often infrequent, but high-impact meaning there is significant variability in deaths from year-to-year.

Road incidents

Road incident deaths include those of drivers – motor vehicles and motorcyclists – in addition to cyclists and pedestrian deaths.

Death rates are typically lowest across Western Europe and Japan, with less than 5 deaths per 100,000 individuals. Across the Americas, rates are typically slightly higher at 5 to 20; most countries in Asia lie between 15 and 30; and rates are typically highest across Sub-Saharan Africa with over 25 per 100,000.

Drowning

The World Health Organization (WHO) emphasises that drowning is one of the most overlooked, preventable causes of death across the world.[8] For every country in the world, drowning is among the top 10 killers for children. In some countries, such as Bangladesh, it is the top mortality cause for children under 15 years old.

In 2016, death rates were highest in Papua New Guinea and Seychelles, between 10 to 16 deaths per 100,000. Rates were also high in countries such as Bangladesh, Central African Republic, Vietnam, and Haiti.

If we look at death rates we see a significant decline since 1990 — especially in low to middle-income countries. In Bangladesh and China, for example, rates have fallen by more than two-thirds over this period.

Fire

Most countries across the Americas, Western Europe, East Asia and Oceania average death rates below 2 per 100,000. Rates across other regions are typically higher at 2-6 per 100,000. When viewed through time we see a notable decline in fire death rates, particularly across Sub-Saharan Africa and Eastern Europe.

Terrorism

Terrorism is defined in the Oxford Dictionary as "the unlawful use of violence and intimidation, especially against civilians, in the pursuit of political aims." We quickly see that this definition is unspecific and subjective.[9] In our full article on *Terrorism* we look at adopted definitions, and how it's distinguished from other forms of violence.

Deaths by animal

Mosquitoes are by far the world's deadliest animal

Around 1.5 million people are killed by animals every year.

More than half a million are killed by other humans – in war, homicides, and terrorism. And close to a million people are killed by other animals in any given year.[10]

Mosquitoes are, by far, the world's deadliest animal for humans: at estimated 780,000 died from the transmission of disease from mosquitoes in 2016. Mosquito deaths are

the sum of deaths (in order, highest to lowest) from: Malaria, Dengue fever, Japanese encephalitis, Yellow fever, Zika virus, Chikungunya, West Nile virus, and Lymphatic filariasis, for which it is the vector.

1.4 RESEARCH OBJECTIVES

This research paper has the following objectives:

- **Data Scientists have to apply their analytical skills to give findings and conclusions in detailed data analysis written in jupyter notebook . Only data analysis is required.**

1.5 RESEARCH METHODOLOGY

❖ RESEARCH DESIGN

The research design is the conceptual framework around which the survey is undertaken. Here a part of the research undertaken is a **Exploratory Research** as it is describing the perception of the respondents.

❖ SAMPLE SIZE

The Sample Size is **6120**. Data has been collected and analysed on the basis of the responses. The research was conducted with an aim of getting respondents from all across the world.

❖ PERIOD OF STUDY

The period of study was more than **a week.**

❖ DATA COLLECTION

1. **Primary Data-** Dataset and excel sheet provided by Flip Robo technologies for project completion.
2. **Secondary Data-**The data includes reference from previously published research papers, journals, books and articles.

❖ METHOD OF ANALYSIS

In this research, tables and various types of graphs such as bar graph and count plot have been used. This helped to present the data in a meaningful way and making it easily understandable.

1.6 RESEARCH LIMITATIONS

- **Non Representative Sample**: The research project was based on the dataset received for only 6120 rows. Hence, such sample size cannot be said to be genuine and actual representative of the people.

- **Shortage of Time**: The time frame of the study was restricted and limited. It therefore becomes difficult to have detailed study on project work. The period was not enough for the proper study and investigation on the project.

- **Use of only Virtual Research Methods**: The survey was based on the data collected by questionnaire which was circulated virtually having limited questions which in turn resulted in collection of insufficient data due to which there was inadequacy in the research.

- **Lack of Scientific Method**: The absence of use of scientific and logical training in research approach turned out to be a great hindrance in the exploration programme.

2. STEPS USED IN EDA OF CAUSE OF DEATH DATASET:

2.1 Steps:

1. Identification of variables and data types

2. Analyzing the basic metrics

3. Non-Graphical Univariate Analysis

4. Missing value treatment

5. Graphical Univariate Analysis

6. Bivariate Analysis

7. Multivariate Analysis using correlation

8. Correlation using heat map

9. Describing the dataset

10. Dataset Info

*3. **ANALYSIS**: DATA FINDINGS AND INTERPRETATION*

3.1 Data shape and Data Types

The shape property returns a tuple containing the shape of the DataFrame.

The shape is the number of rows and columns of the DataFrame

When doing data analysis, it is important to make sure you are using the correct data types; otherwise you may get unexpected results or errors. In the case of pandas, it will correctly infer data types in many cases and you can move on with your analysis without any further thought on the topic.

Despite how well pandas works, at some point in your data analysis processes, you will likely need to explicitly convert data from one type to another. This article will discuss the basic pandas data types (aka dtypes), how they map to python and numpy data types and the options for converting from one pandas type to another.

Pandas Data Types

A data type is essentially an internal construct that a programming language uses to understand how to store and manipulate data. For instance, a program needs to understand that you can add two numbers together like 5 + 10 to get 15. Or, if you have two strings such as “cat” and “hat” you could concatenate (add) them together to get “cathat.”

A possible confusing point about pandas data types is that there is some overlap between pandas, python and numpy. This table summarizes the key points:

Pandas dtype mapping

Pandas dtype	**Python type**	**NumPy type**	**Usage**
object	str or mixed	string_, unicode_, mixed types	Text or mixed numeric and non-numeric values
int64	Int	int_, int8, int16, int32, int64, uint8, uint16, uint32, uint64	Integer numbers
float64	Float	float_, float16, float32, float64	Floating point numbers
bool	Bool	bool_	True/False values

Pandas dtype mapping

Pandas dtype	**Python type**	**NumPy type**	**Usage**
datetime64	NA	datetime64[ns]	Date and time values
timedelta[ns]	NA	NA	Differences between two datetimes
category	NA	NA	Finite list of text values

For the most part, there is no need to worry about determining if you should try to explicitly force the pandas type to a corresponding to NumPy type. Most of the time, using pandas default int64 and float64 types will work. The only reason I included in this table is that sometimes you may see the numpy types pop up on-line or in your own analysis.

For this article, I will focus on the follow pandas types:

- object
- int64
- float64
- datetime64
- bool

The category and timedelta types are better served in an article of their own if there is interest. However, the basic approaches outlined in this article apply to these types as well.

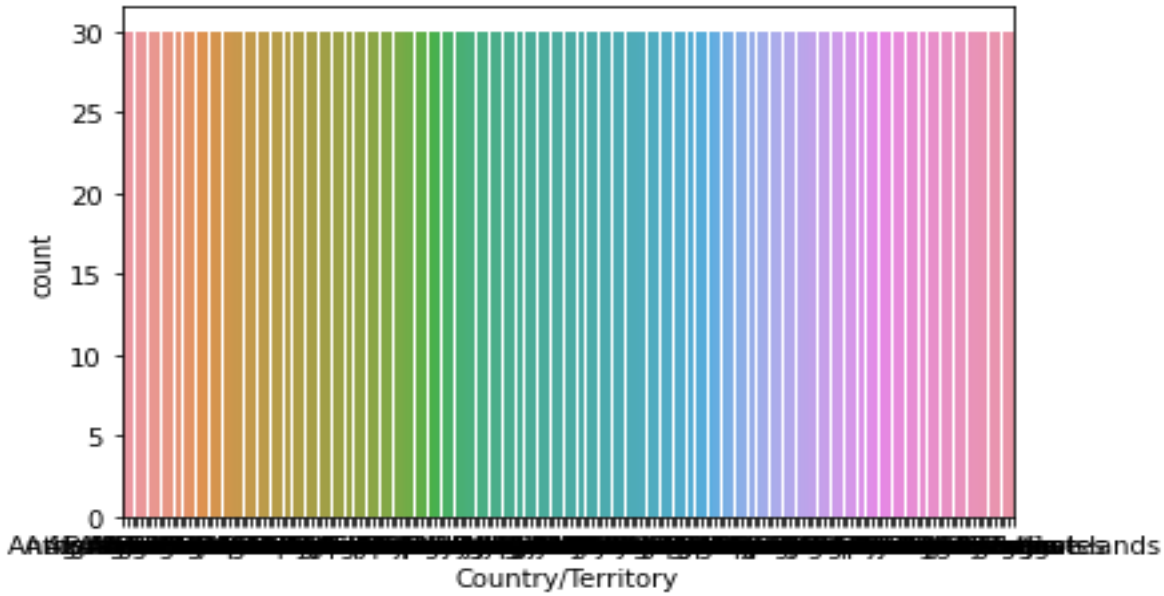
One other item I want to highlight is that the object data type can actually contain multiple different types. For instance, the a column could include integers, floats and strings which collectively are labeled as an object . Therefore, you may need some additional techniques to handle mixed data types in object columns.

3.2 DATA VISUALIZATION

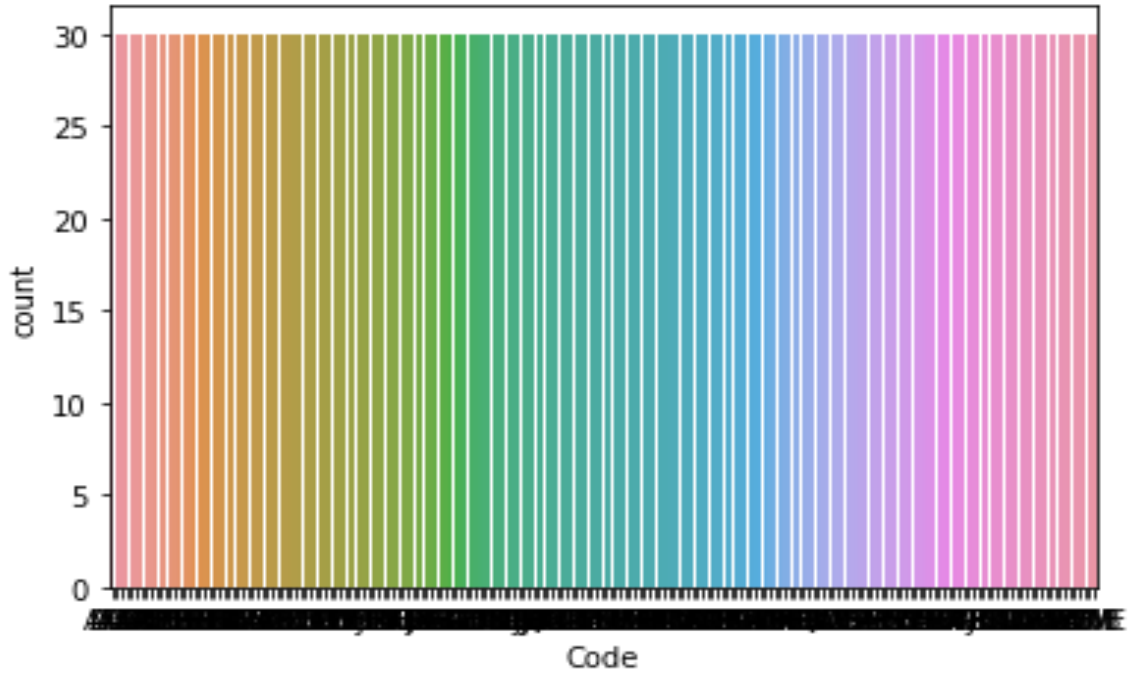
We will use countplot,distplot for univariate analysis and scatterplot for bivariate analysis. For multivariate analysis we will use correlation and heatmap.

UNIVARIATE ANALYSIS-

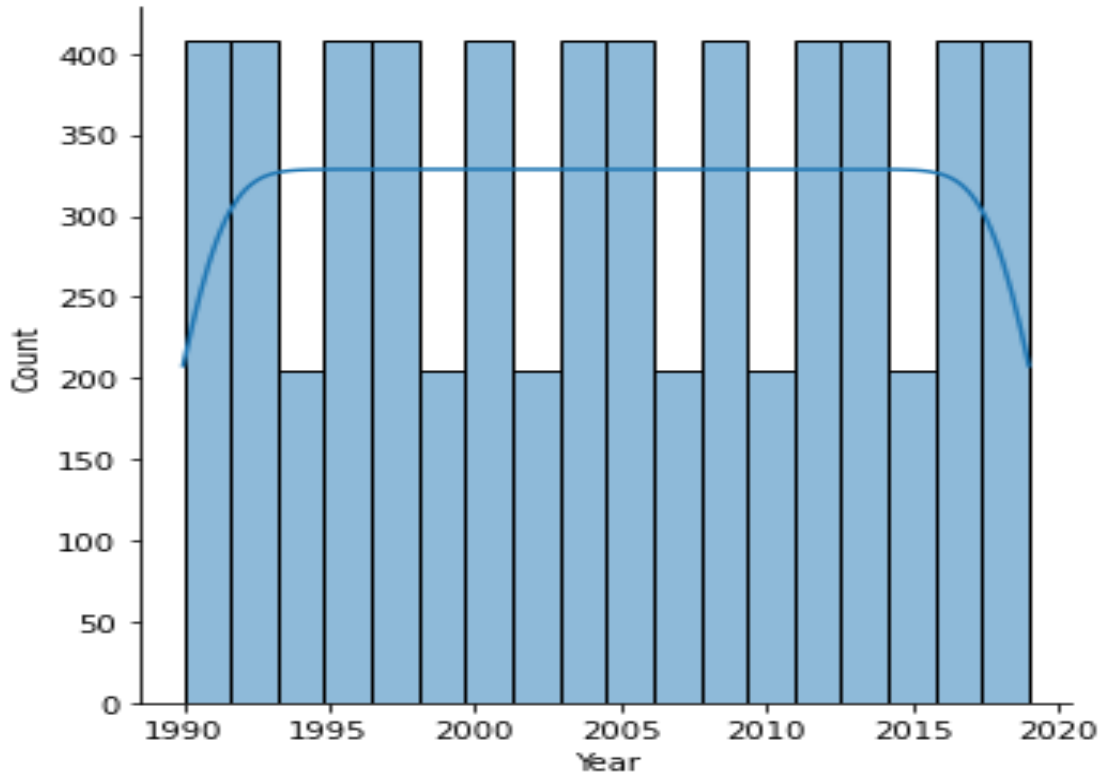
COUNTRY/TERRITORY-



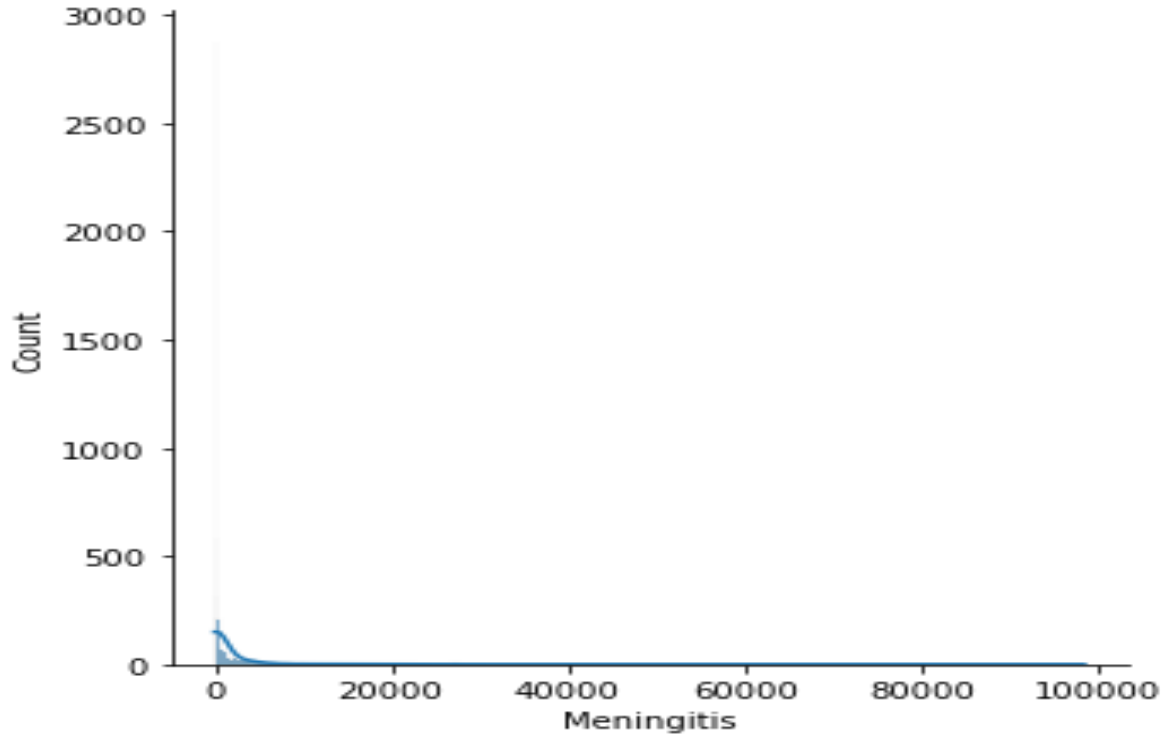
CODE



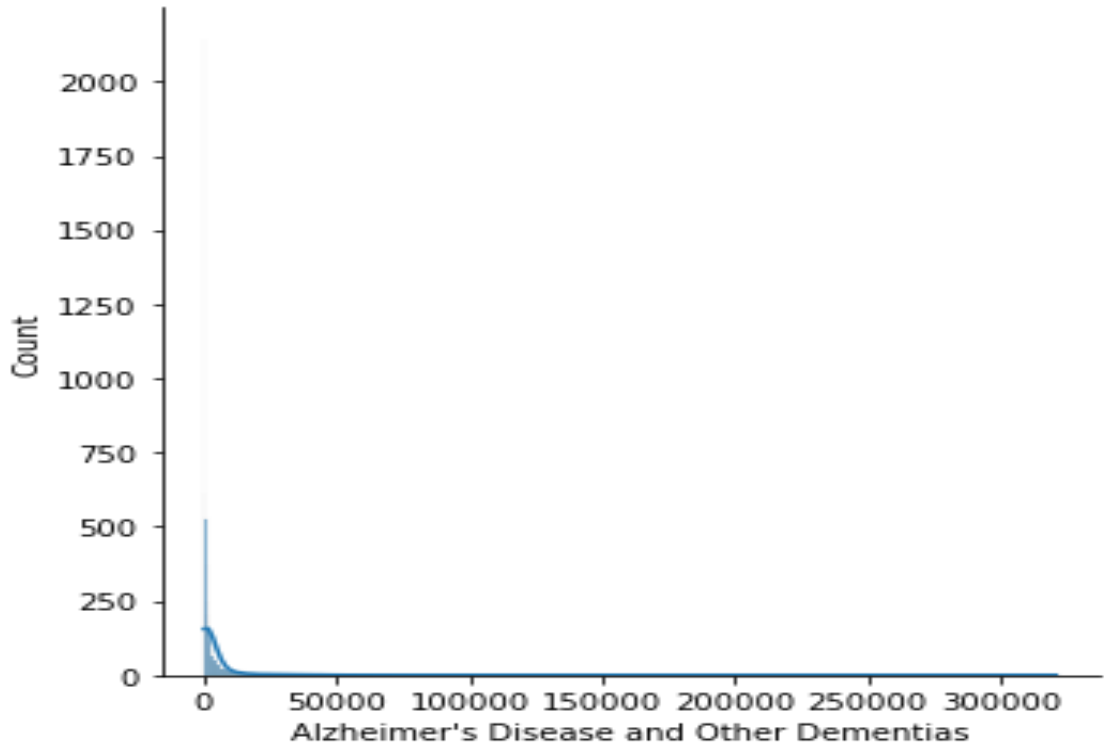
YEAR



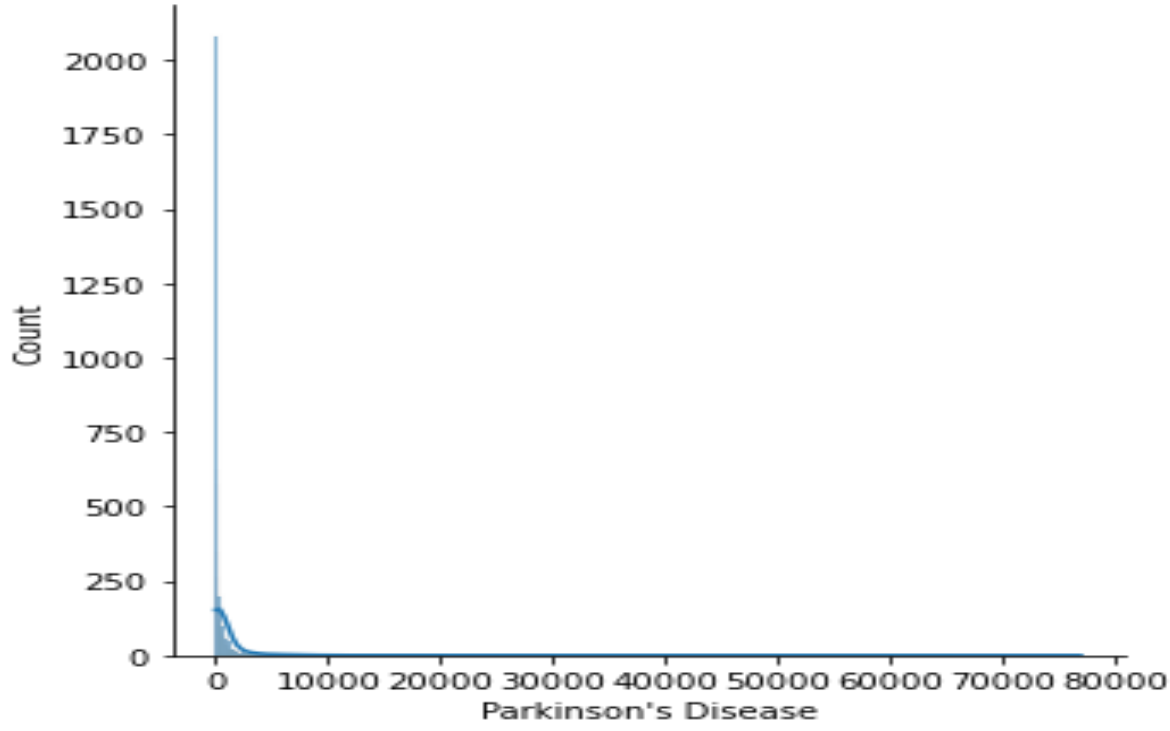
MENINGITIS



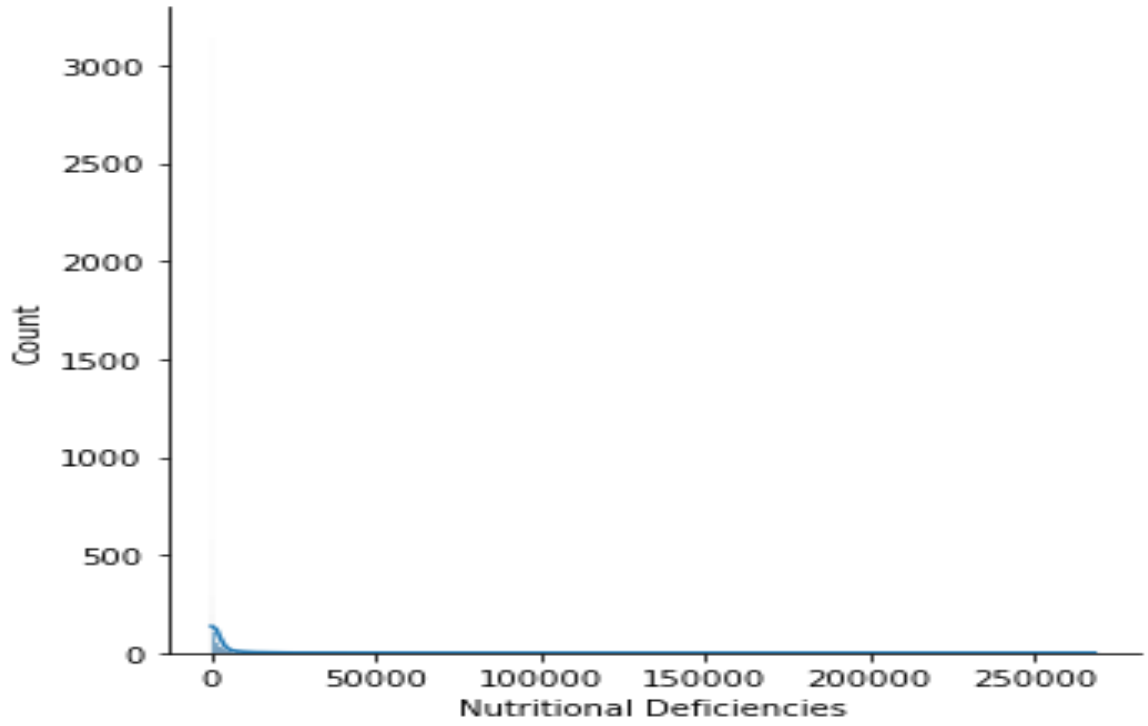
ALZHEIMER’S DISEASE AND OTHER DEMENTIS-



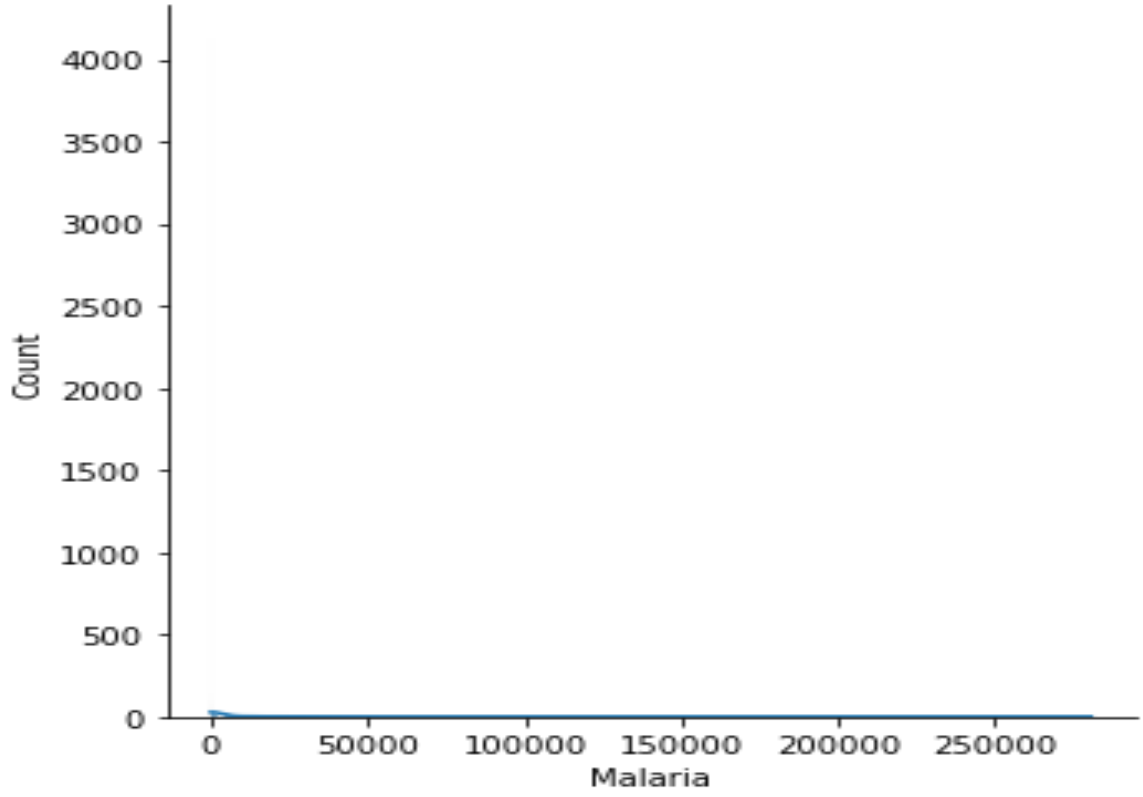
PARKINSON’S DISEASE-



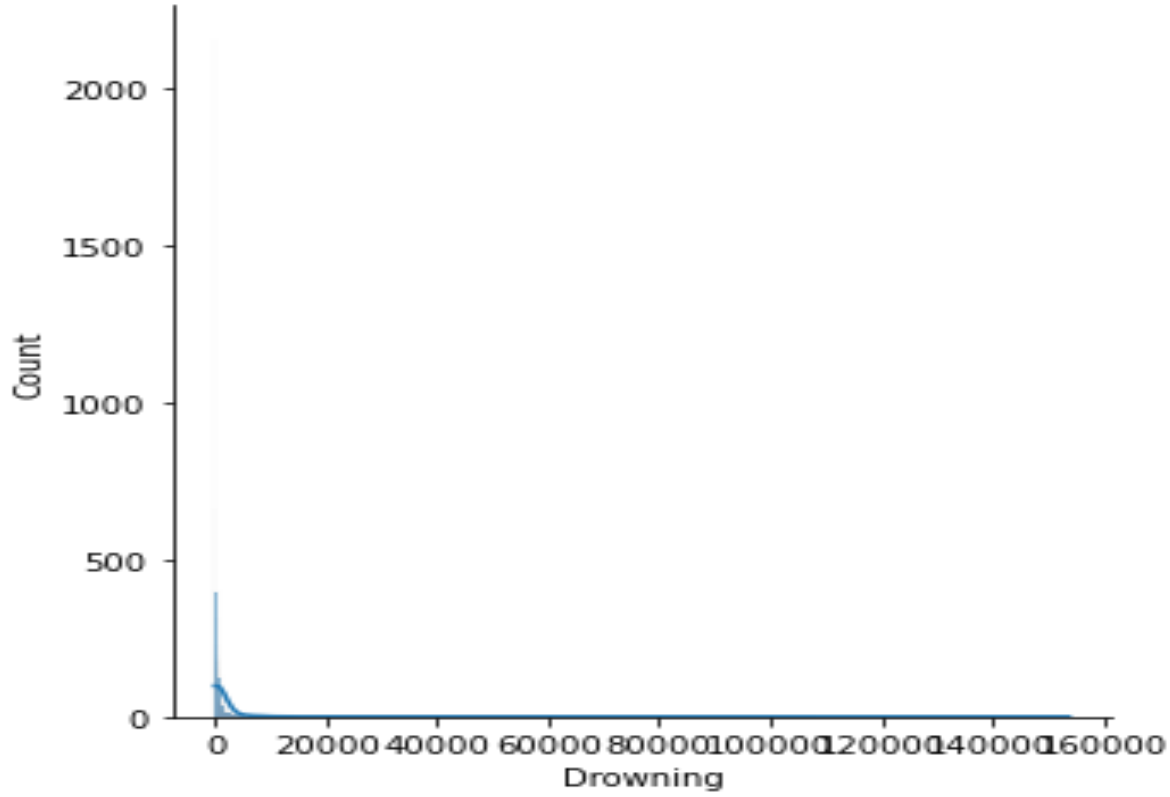
NUTRITIONAL DEFICIENCIES-



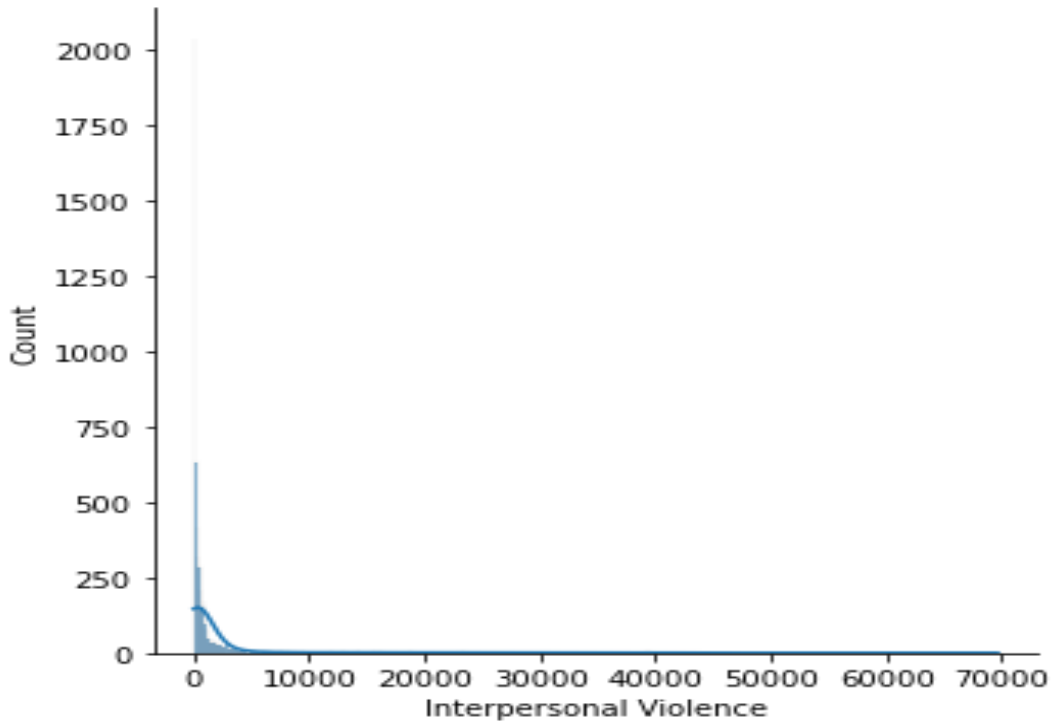
MALARIA-



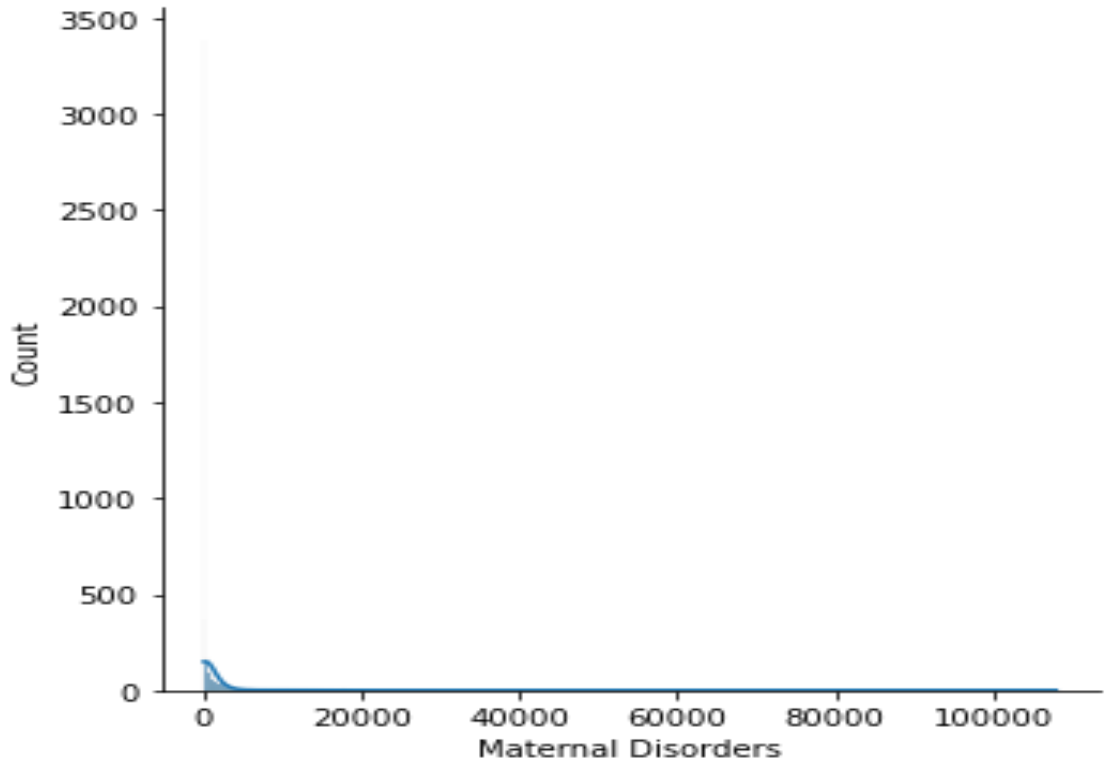
DROWNING-



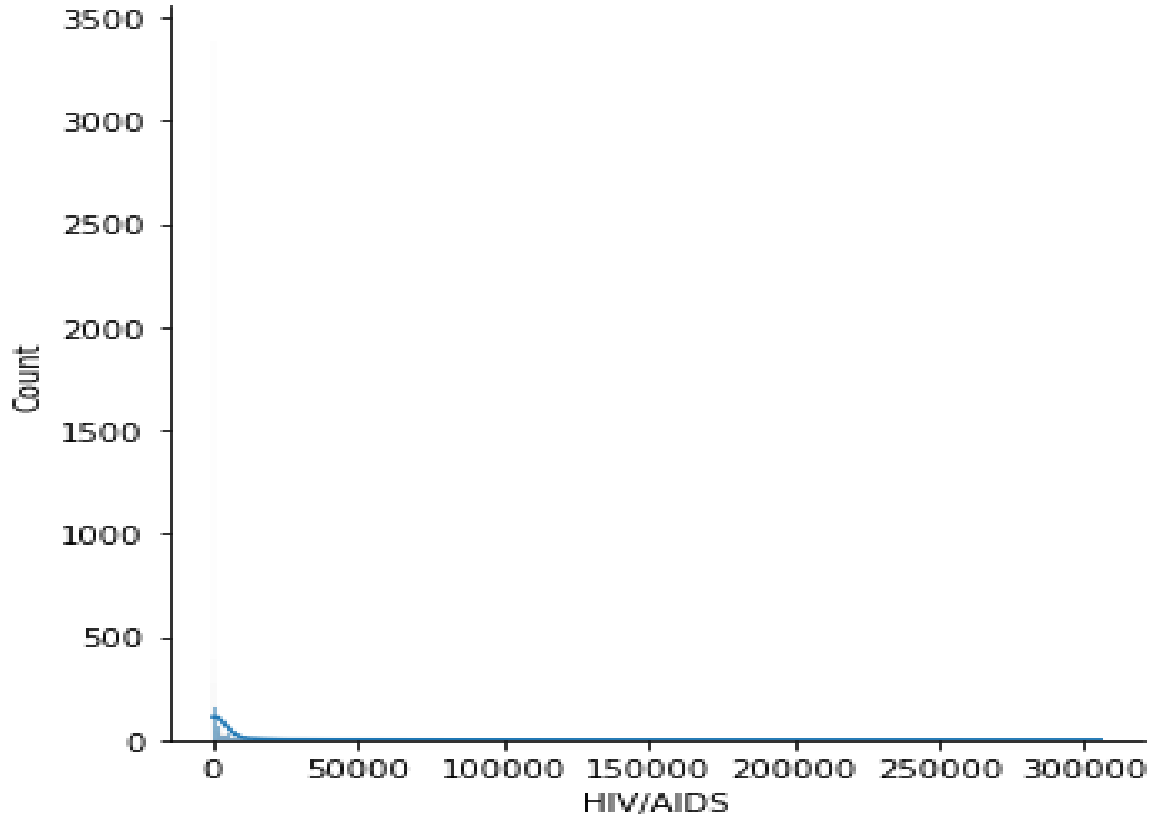
INTERPERSONAL VIOLENCE-



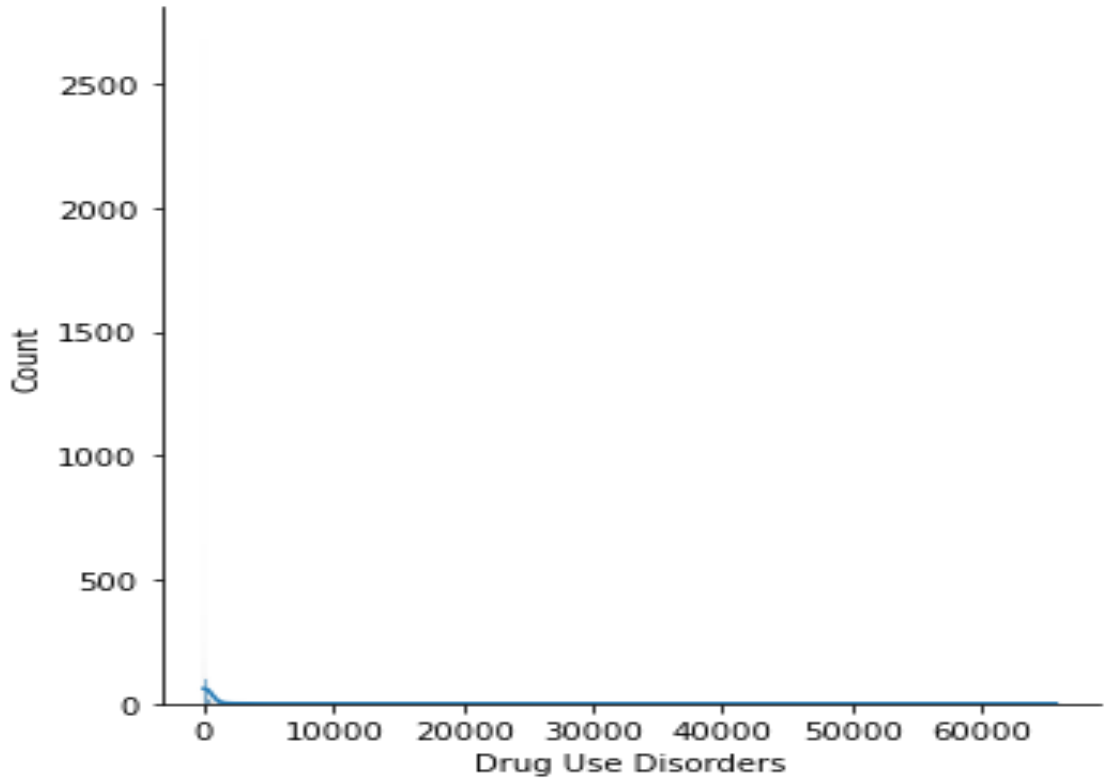
MATERNAL DISORDERS-



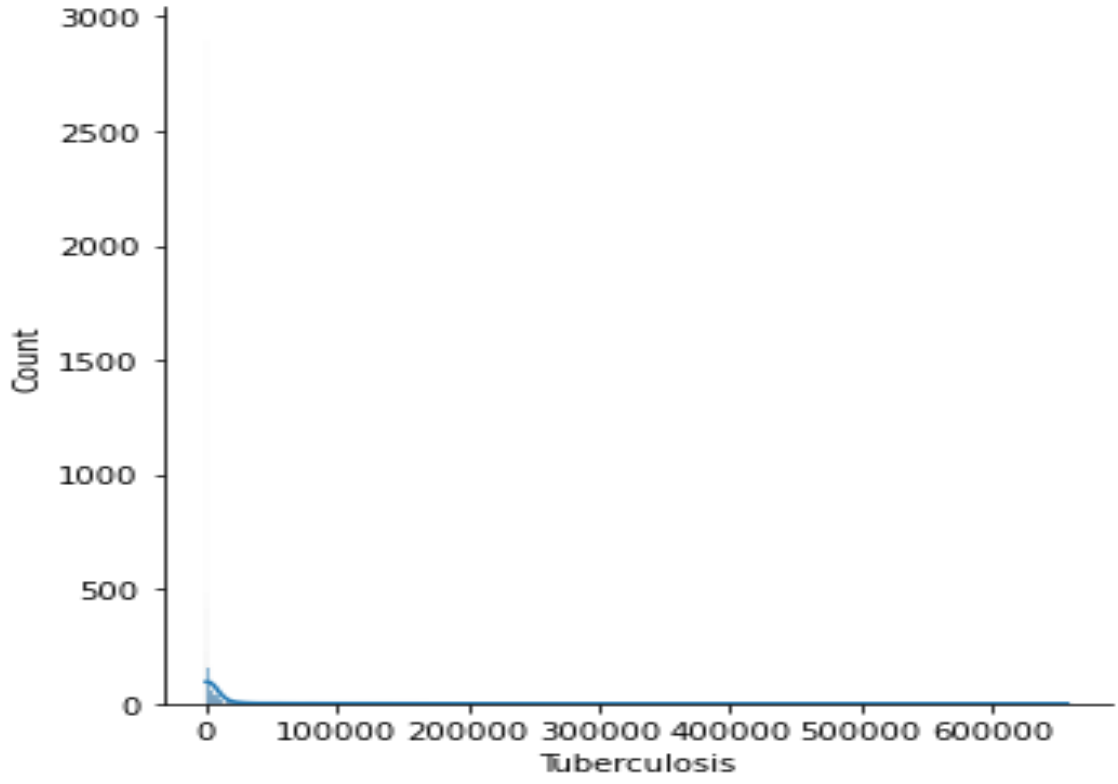
HIV/AIDS-



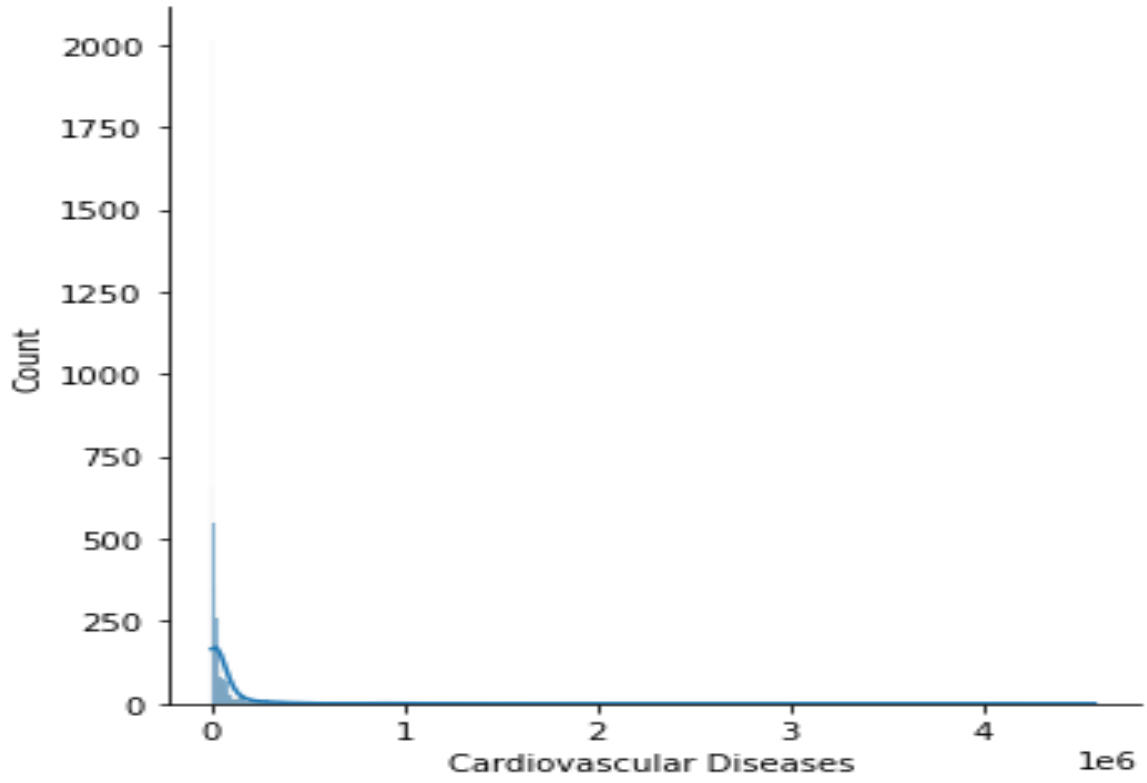
DRUG USE DISORDERS-



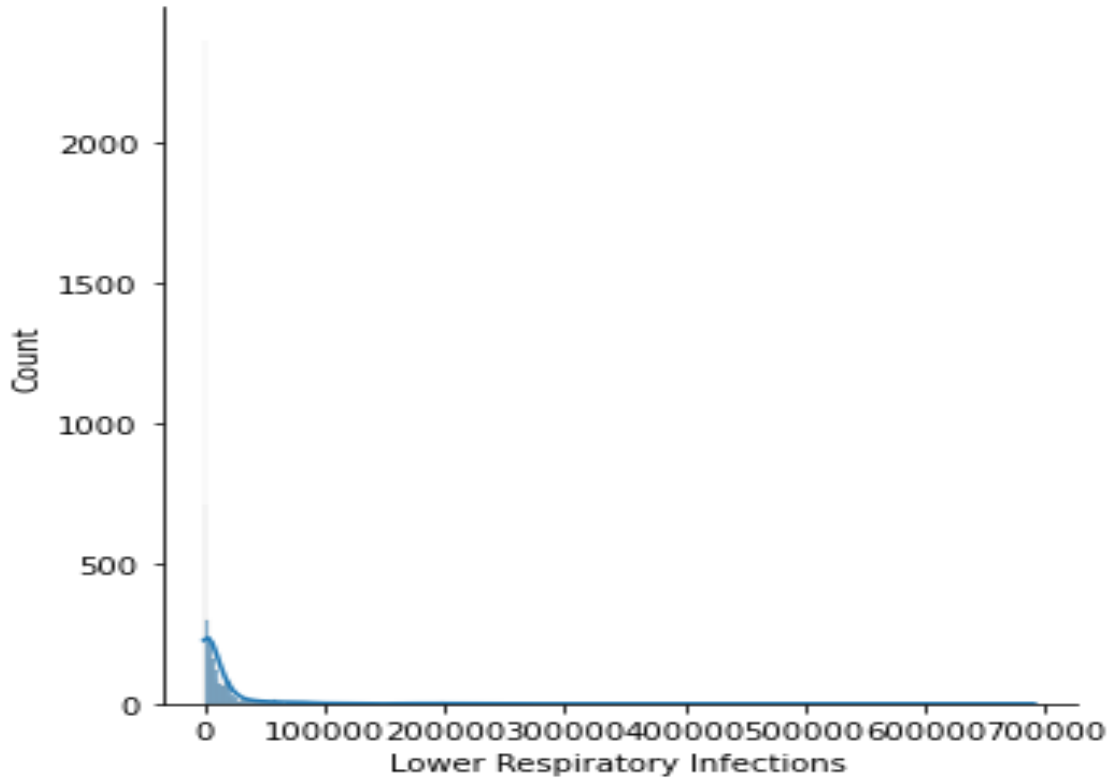
TUBERCULOSIS-



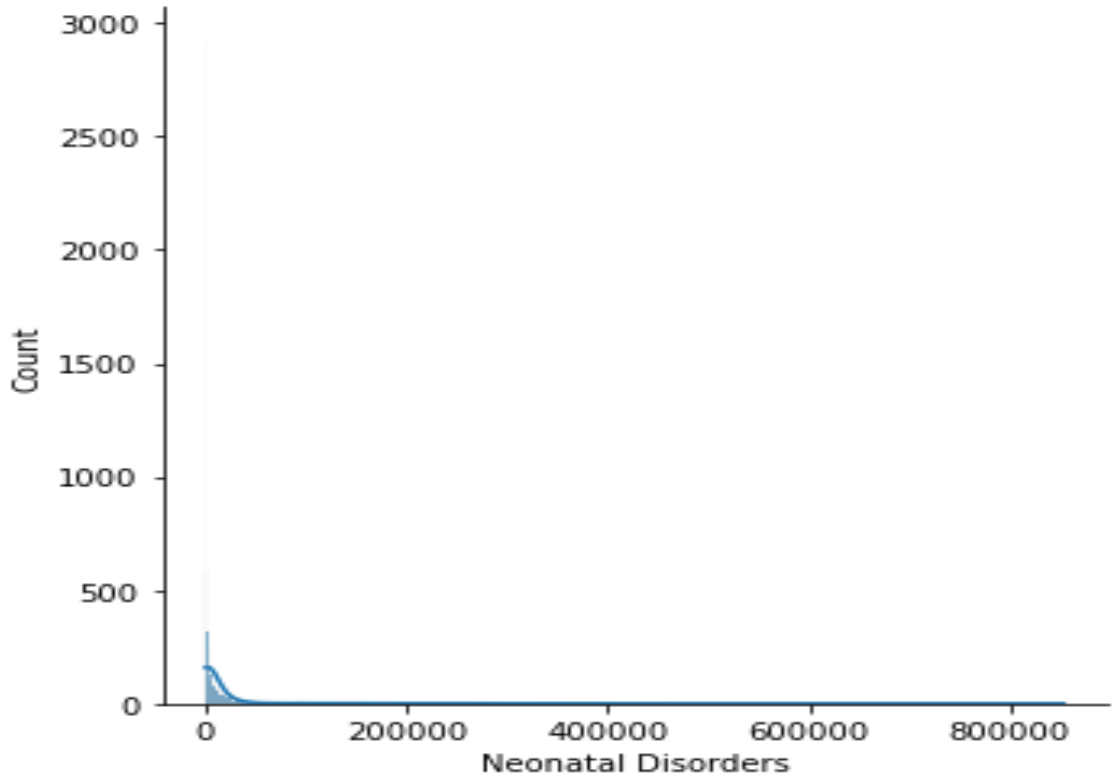
CARDIOVASCULAR DISEASES-



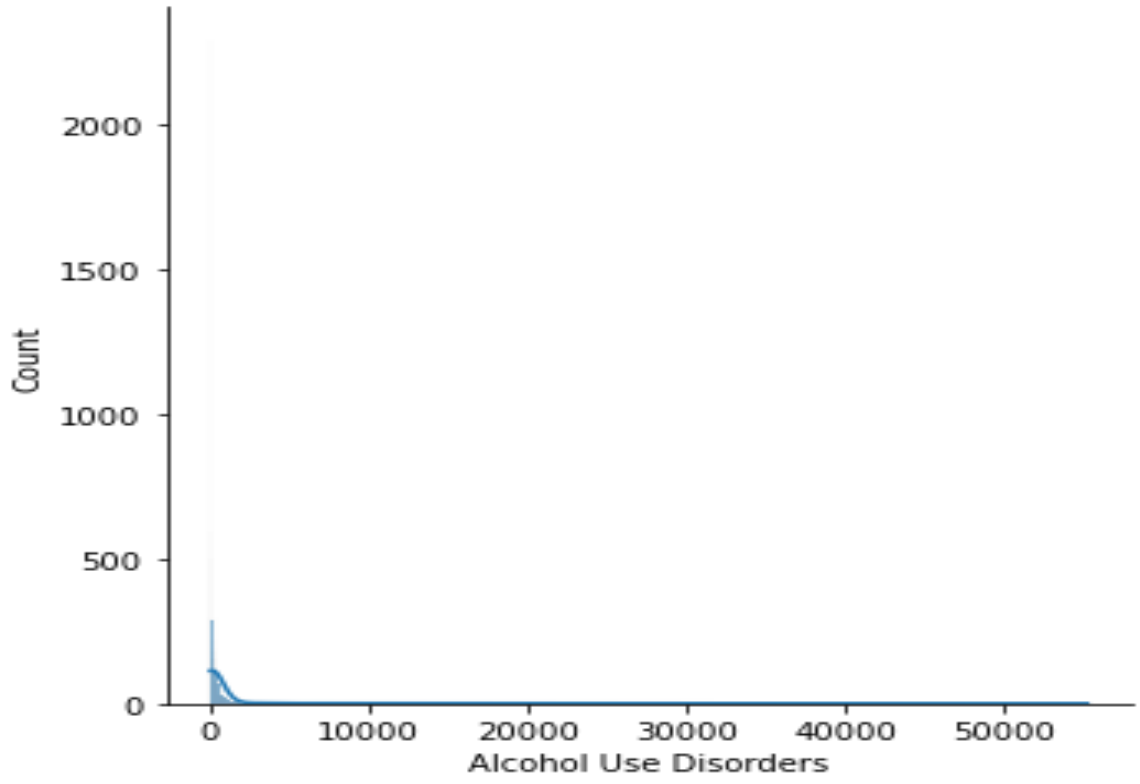
LOWER RESPIRATORY INFECTIONS-



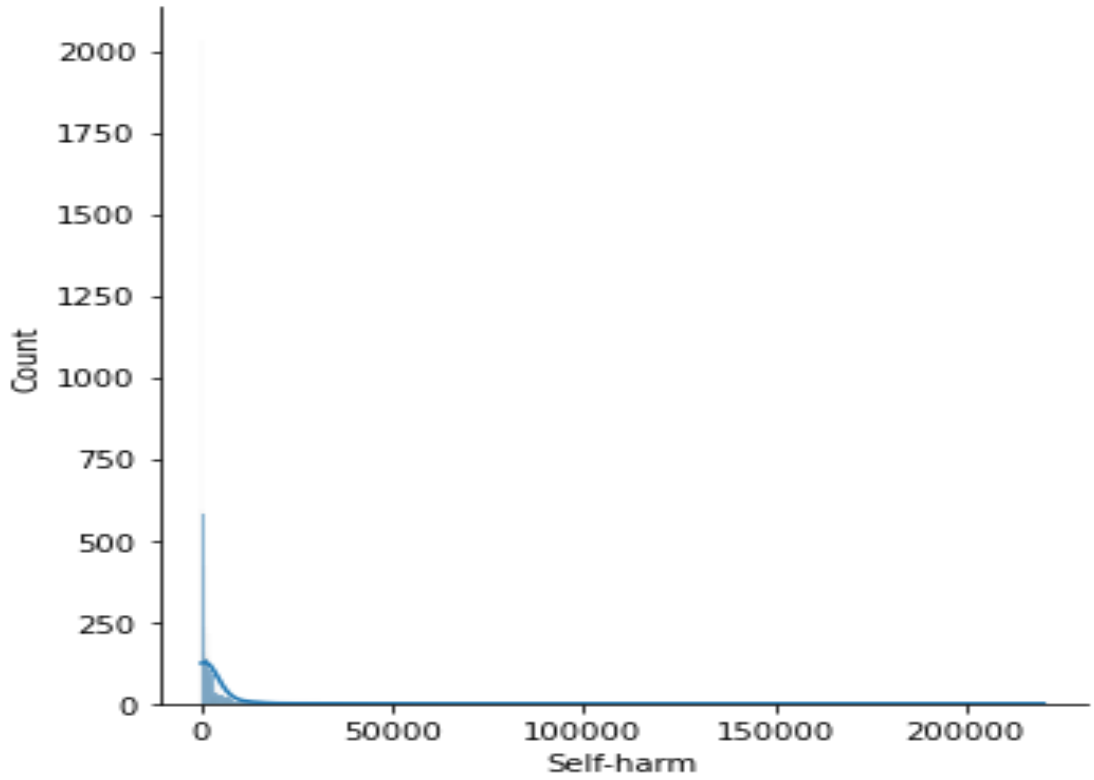
NEONATAL DISORDERS-



ALCOHOL USE DISORDERS-



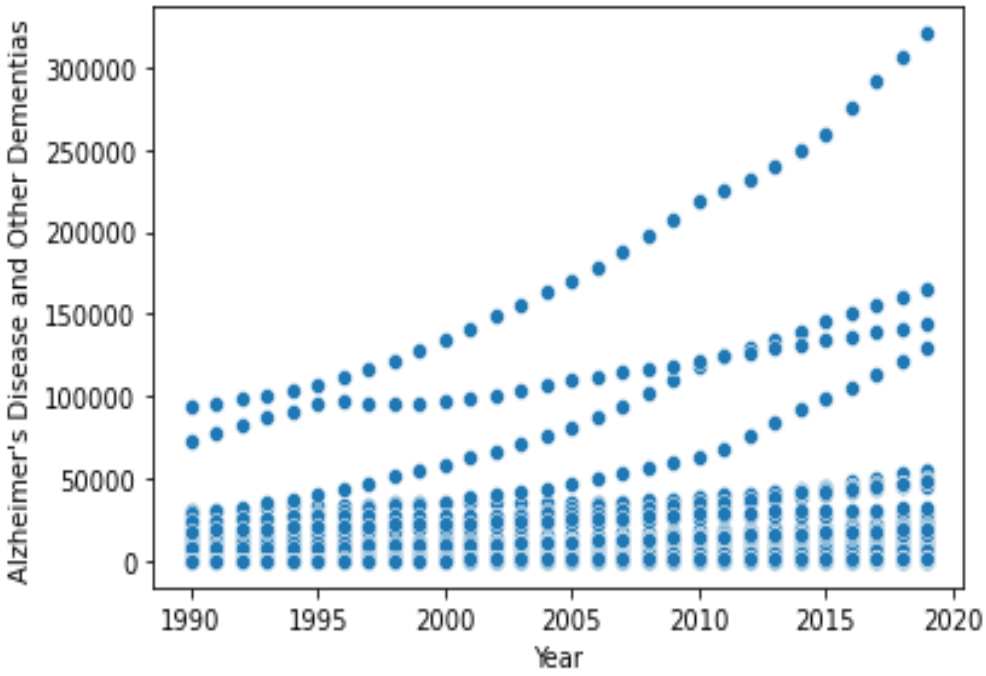
SELF-HARM-



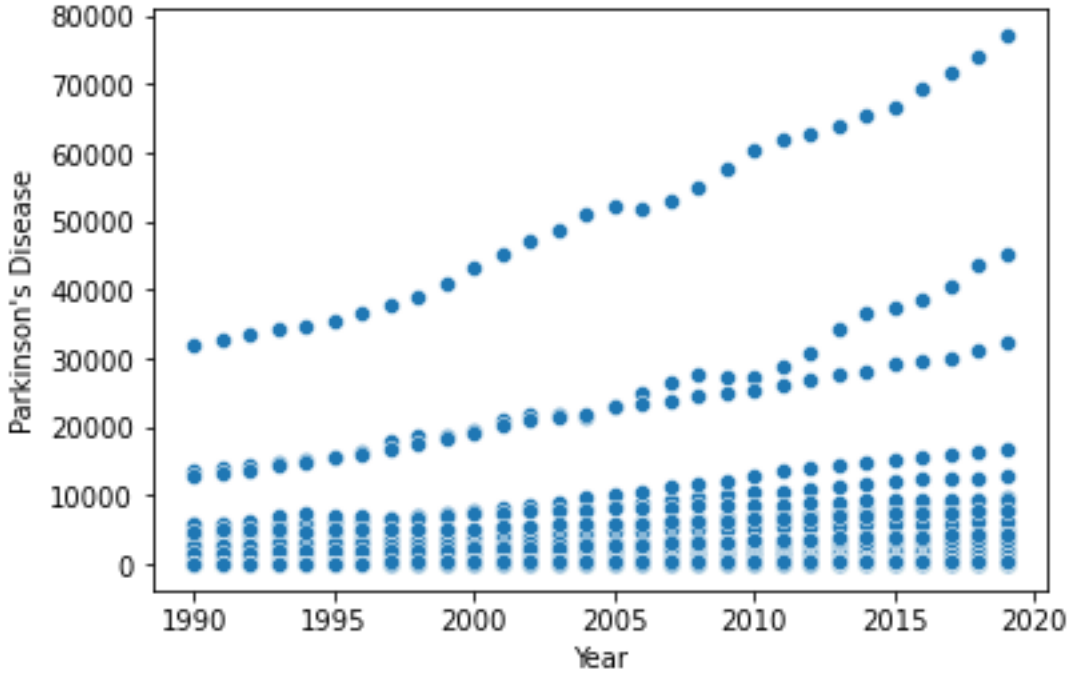
BIVARIATE ANALYSIS-

We will use all the column names as y axis and Year column as x axis to interpret the cause of death in relation to the year.

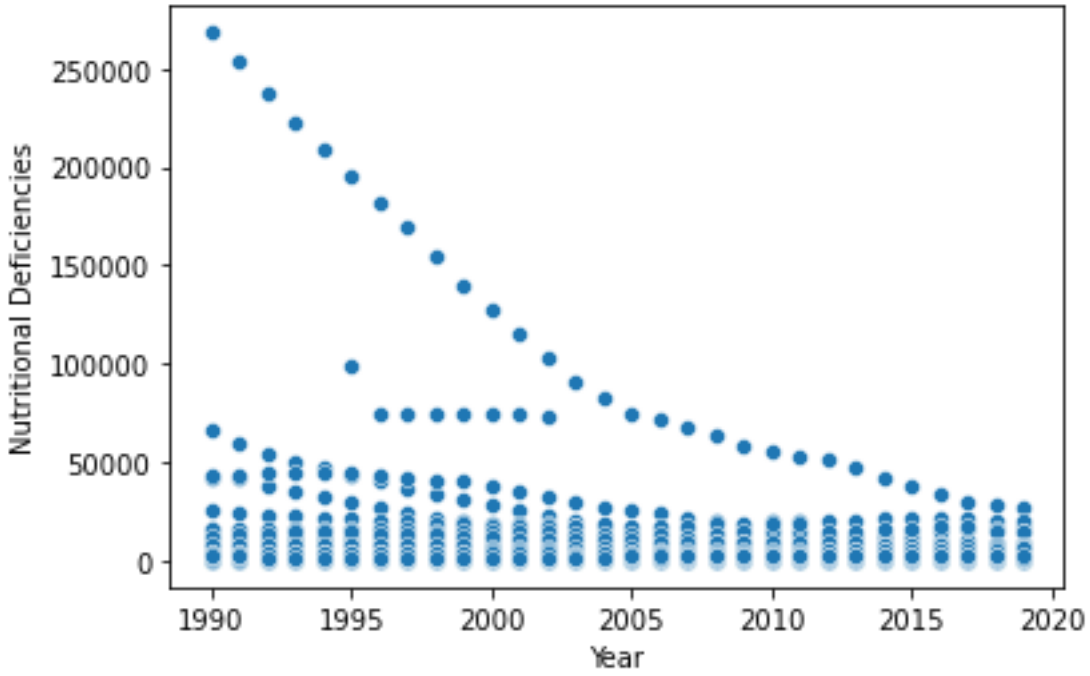
ALZHEIMER'S DISEASE AND OTHER DEMENTIAS-



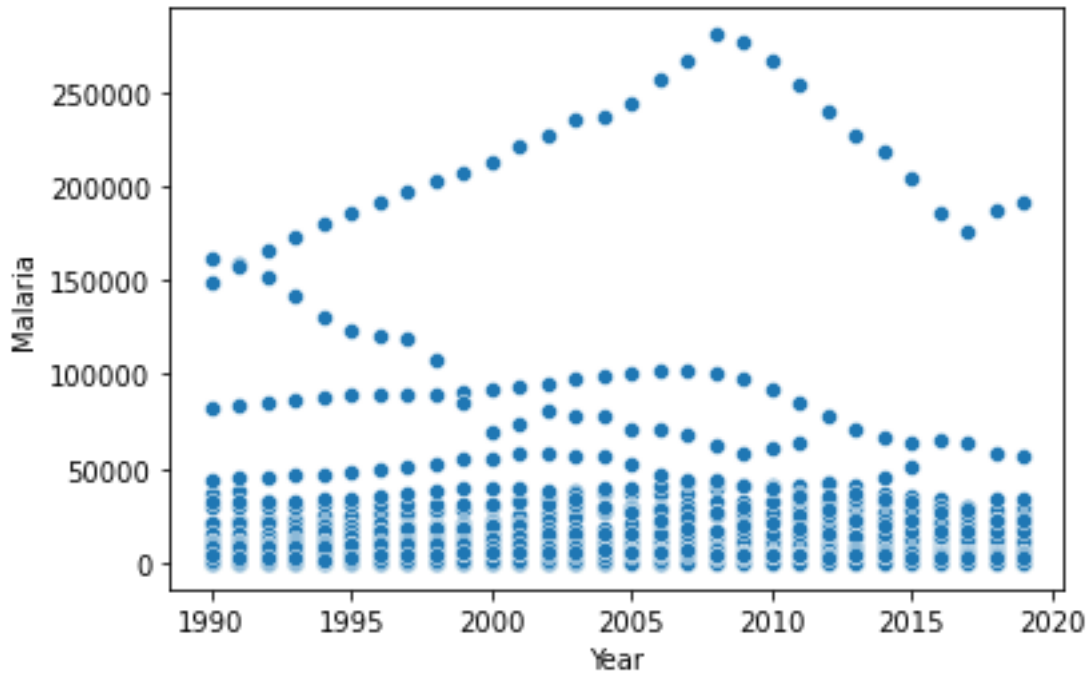
PARKINSON'S DISEASE-



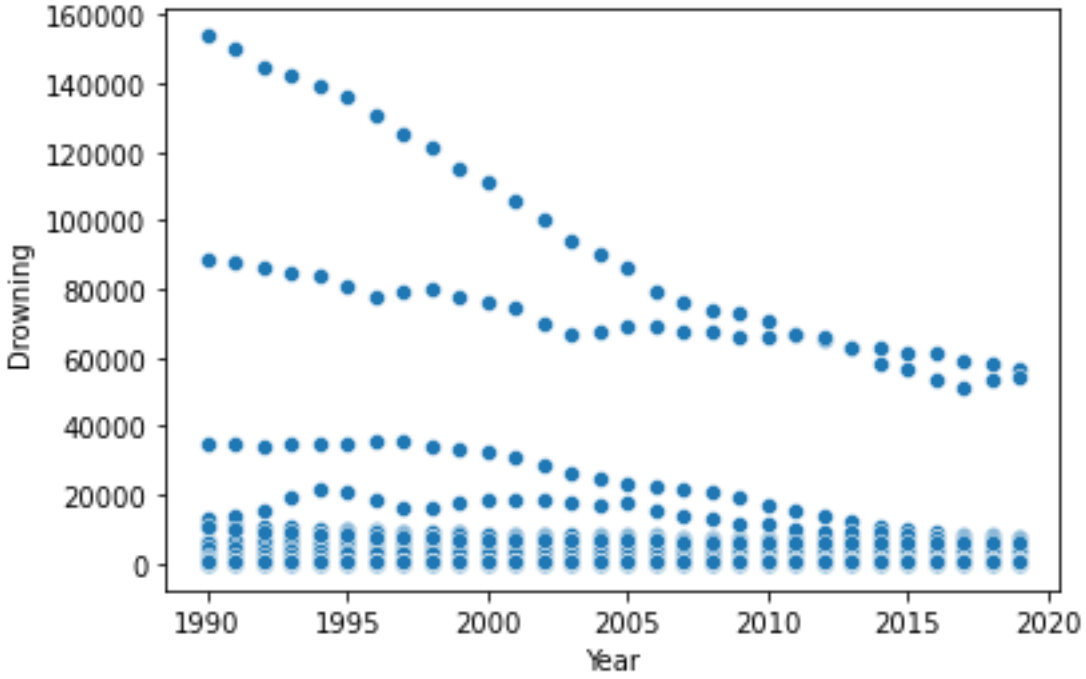
NUTRITIONAL DEFICIENCIES –



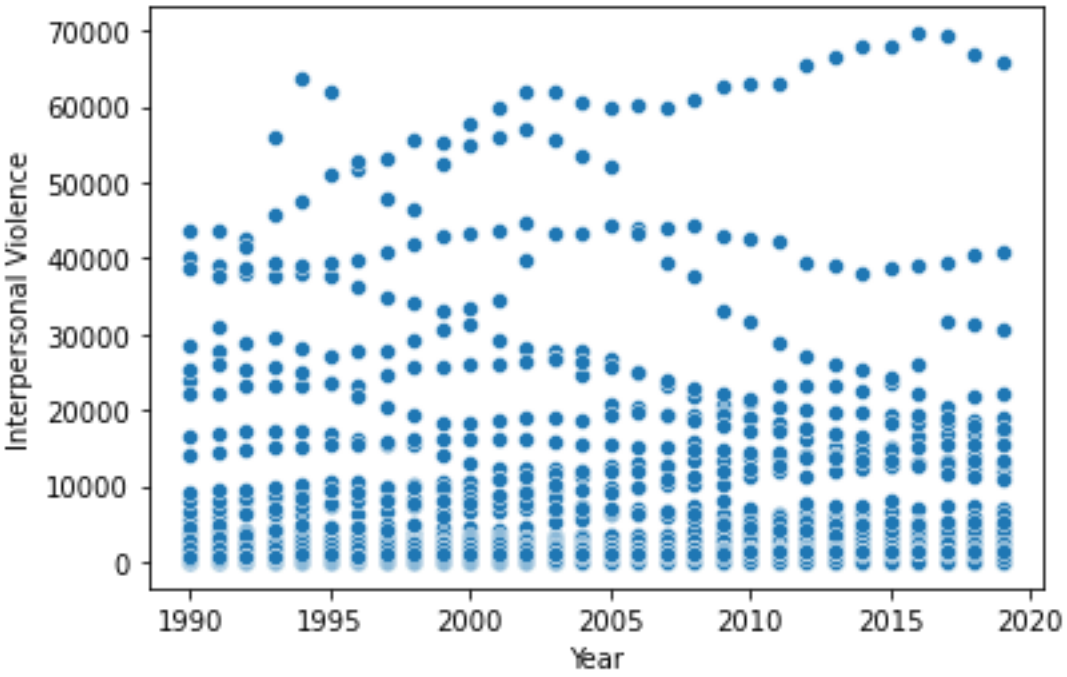
MALARIA-



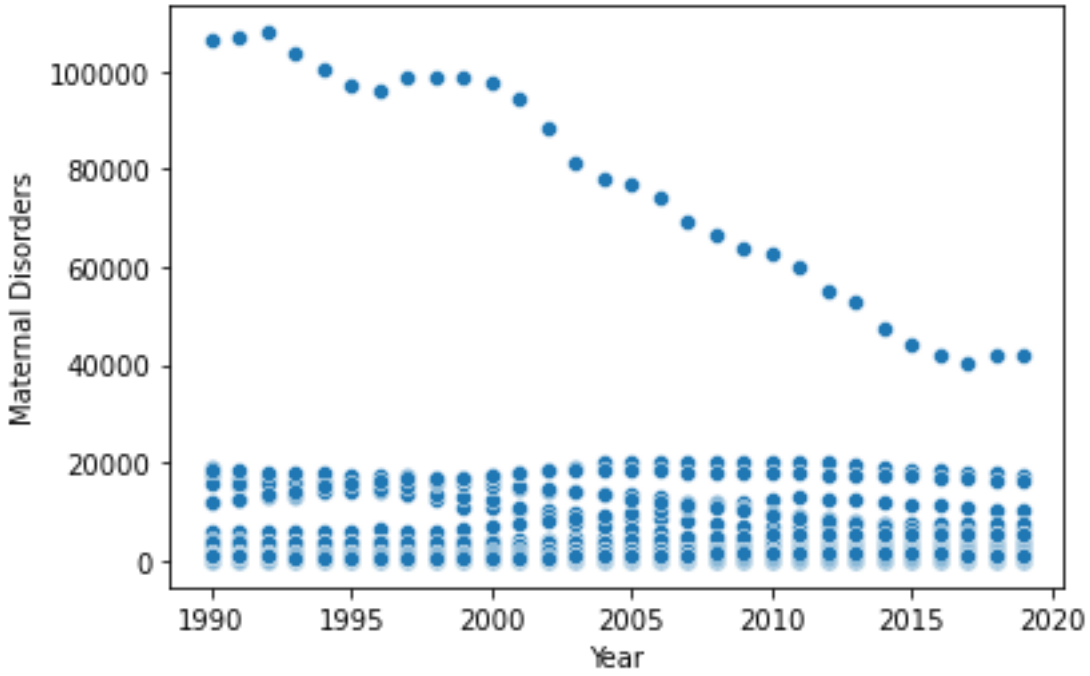
DROWNING-



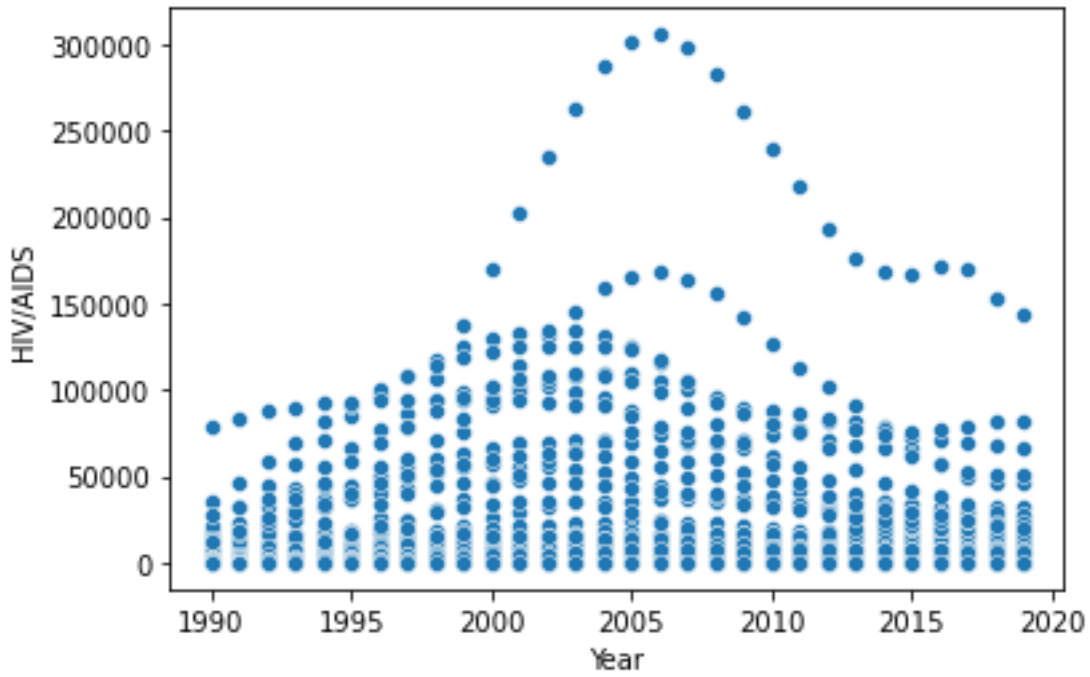
INTERPERSONAL VIOLENCE-



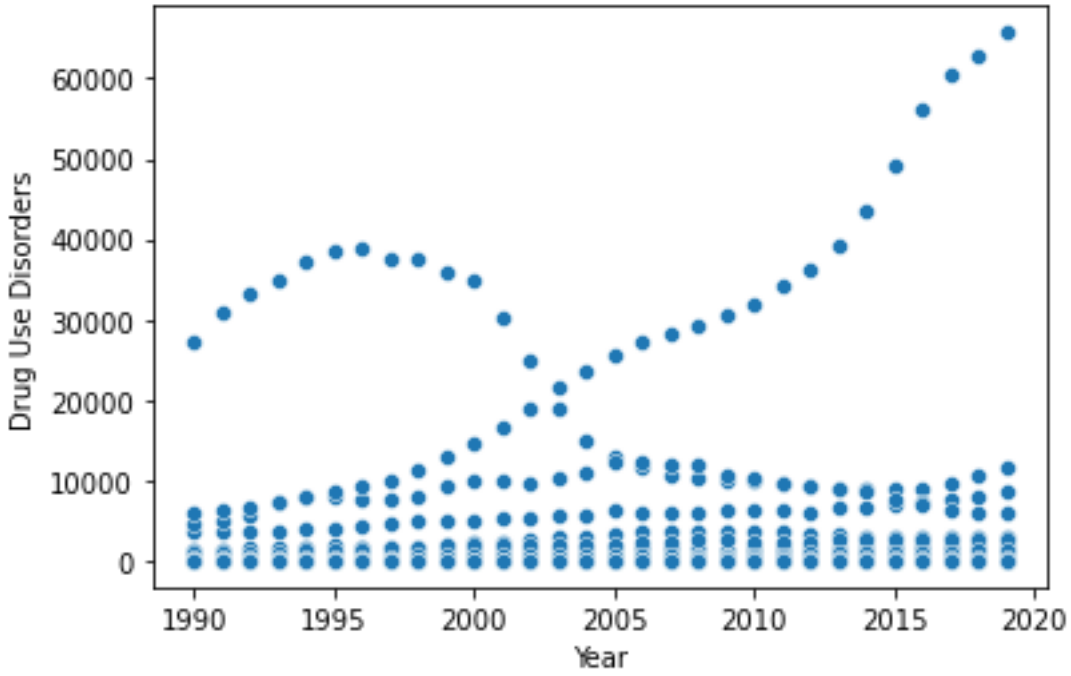
MATERNAL DISORDERS-



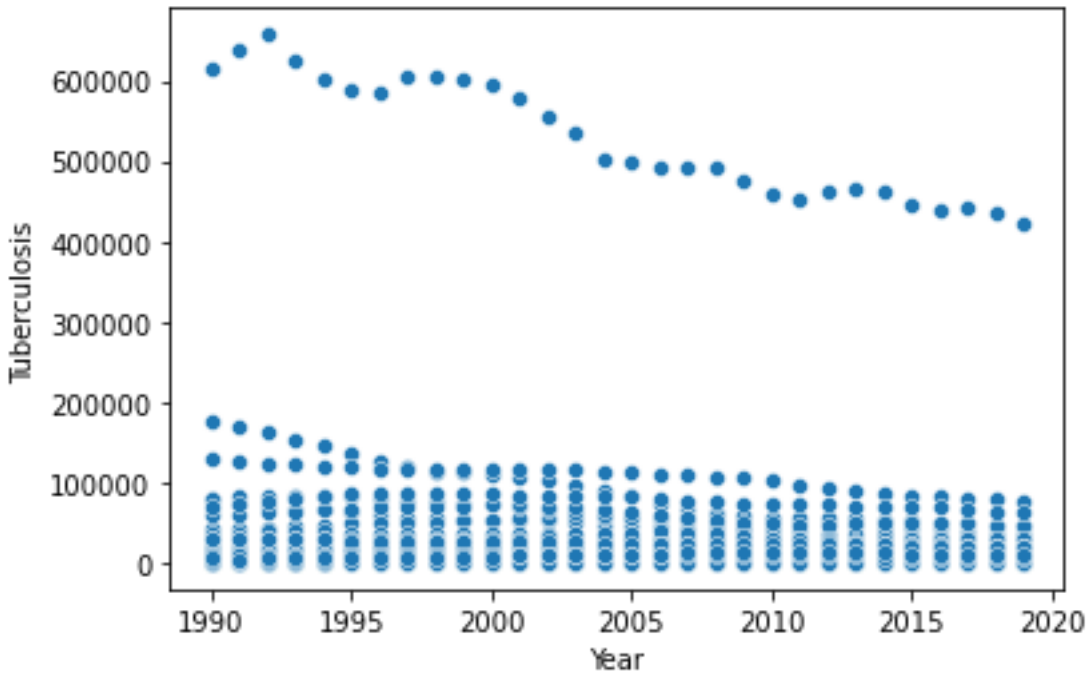
HIV/AIDS-



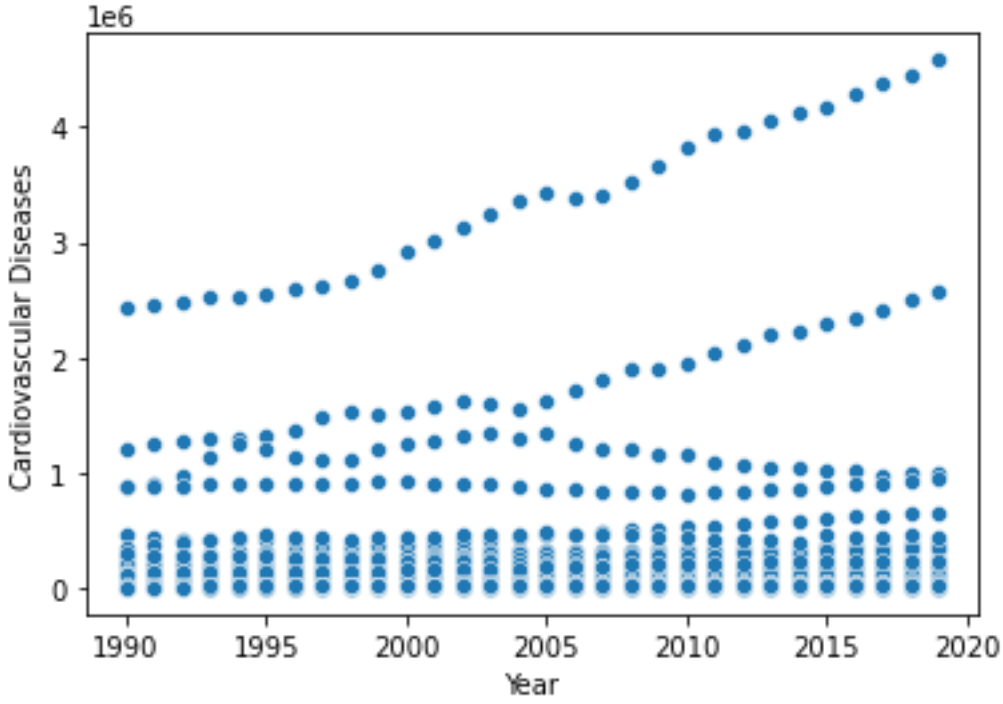
DRUG USE DISORDERS-



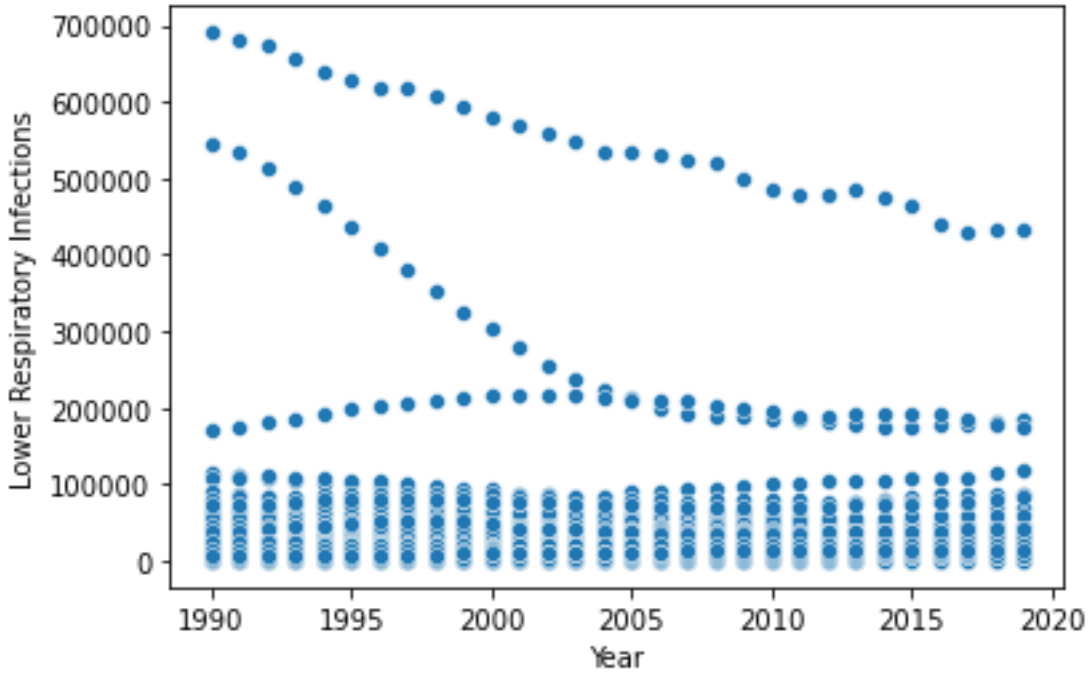
TUBERCULOSIS-



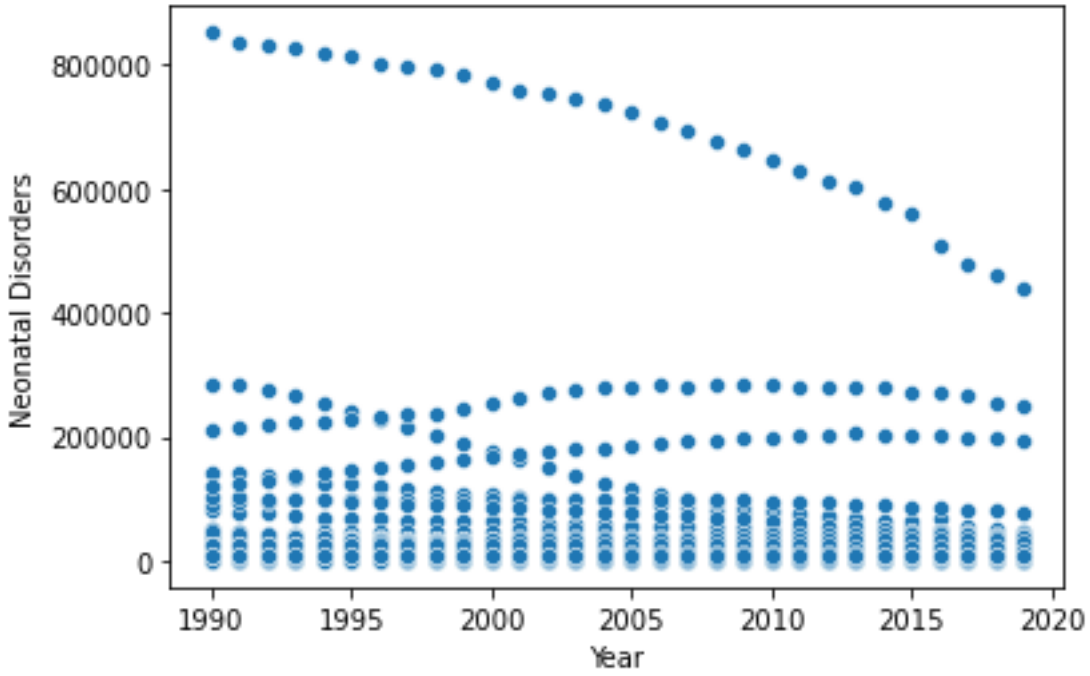
CARDIOVASCULAR DISEASES-



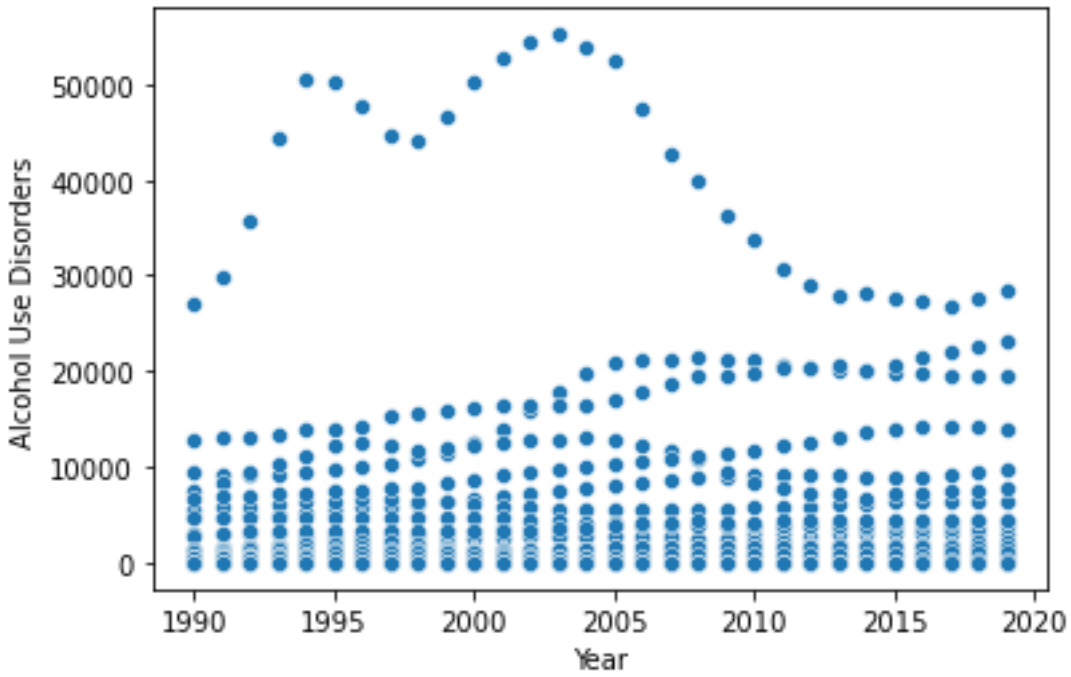
LOWER RESPIRATORY INFECTIONS-



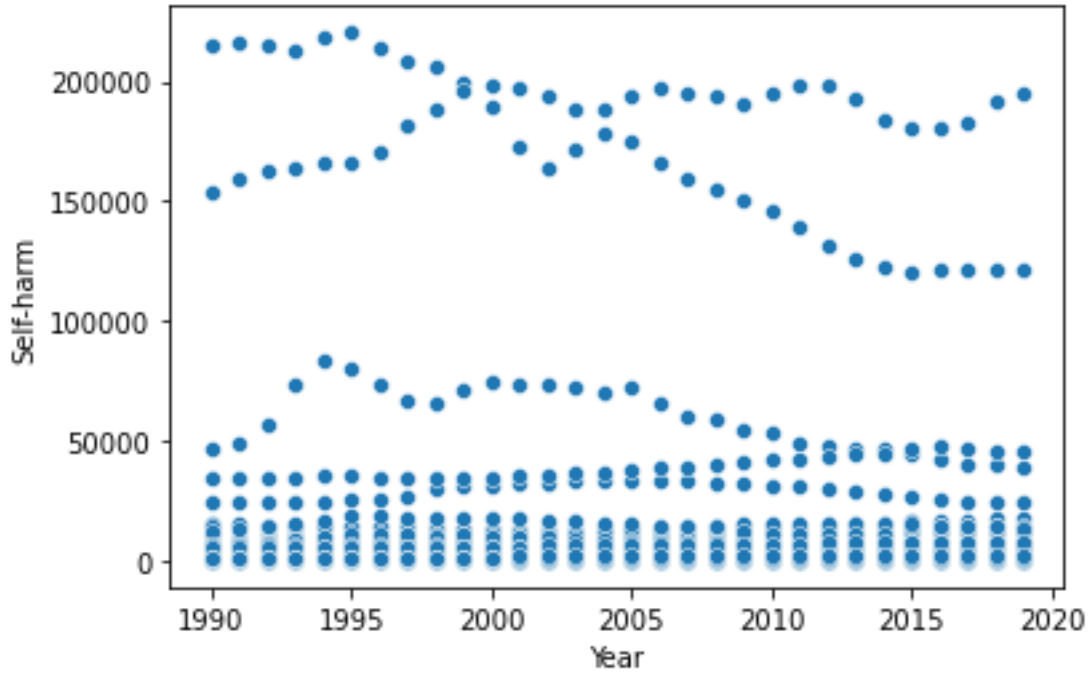
NEONATAL DISORDERS-



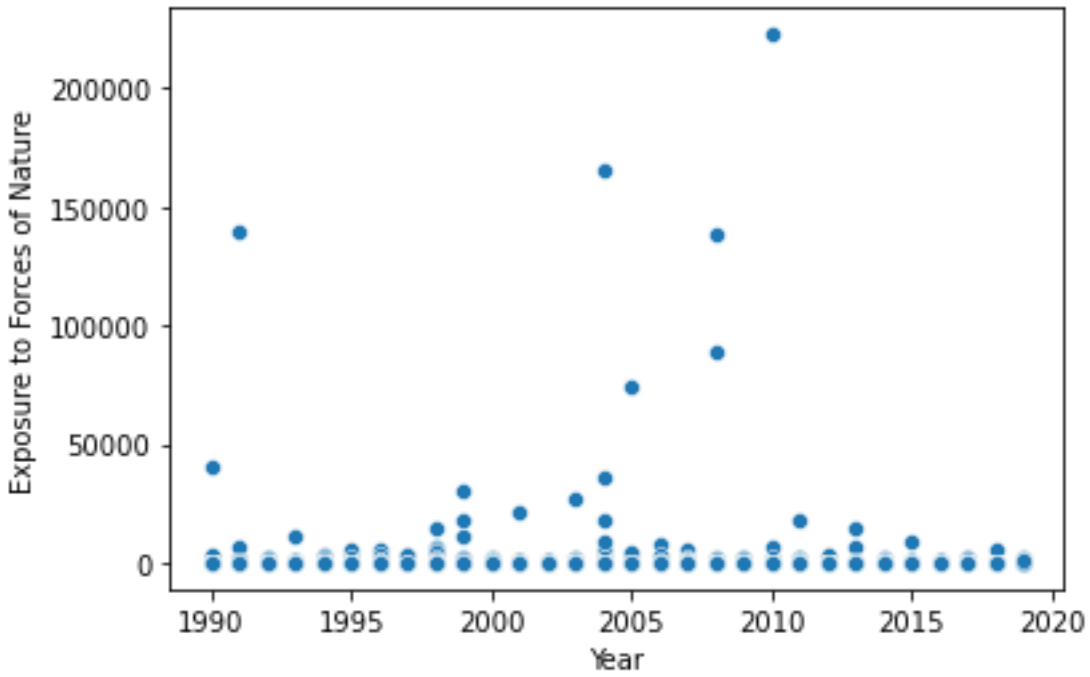
ALCOHOL USE DISORDERS-



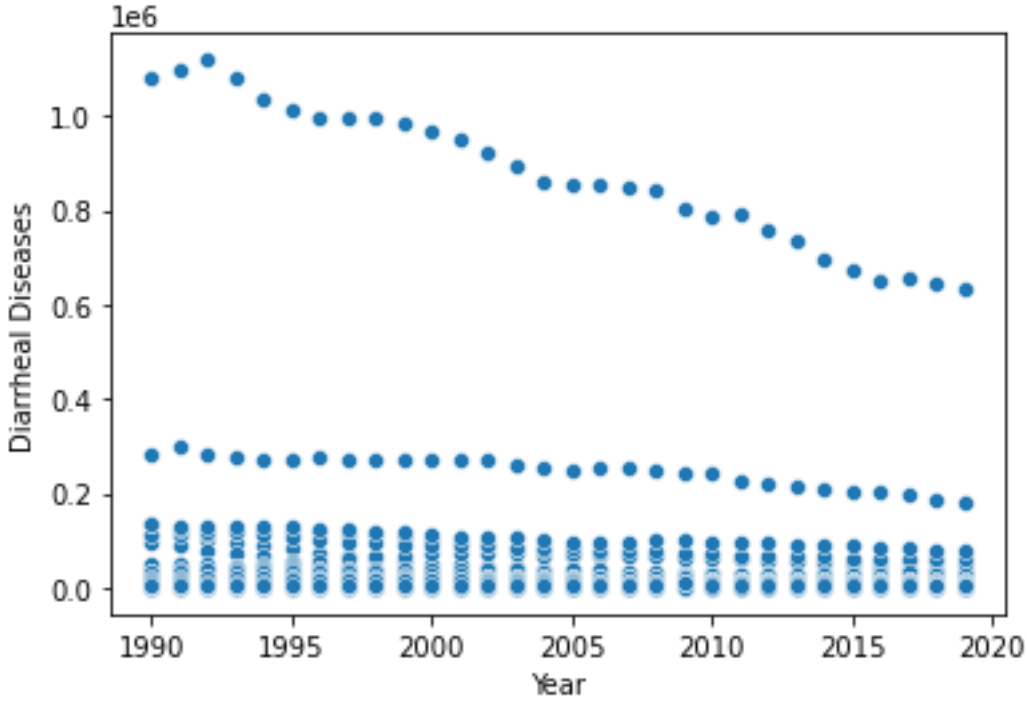
SELF-HARM-



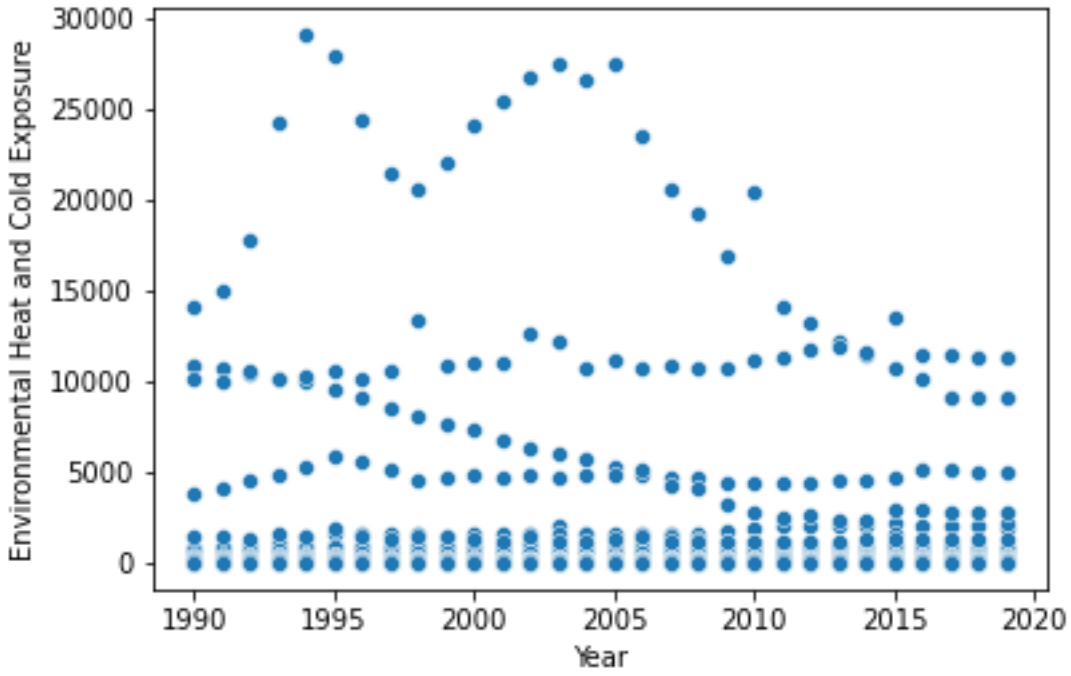
EXPOSURE TO FORCES OF NATURE-



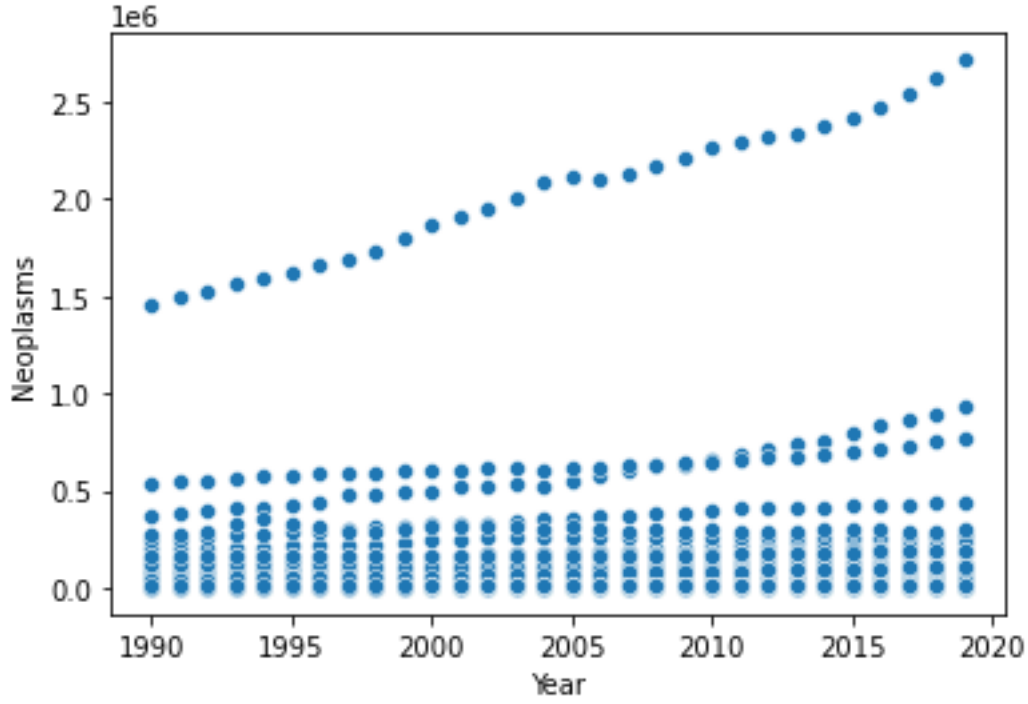
DIARRHEAL DISEASES-



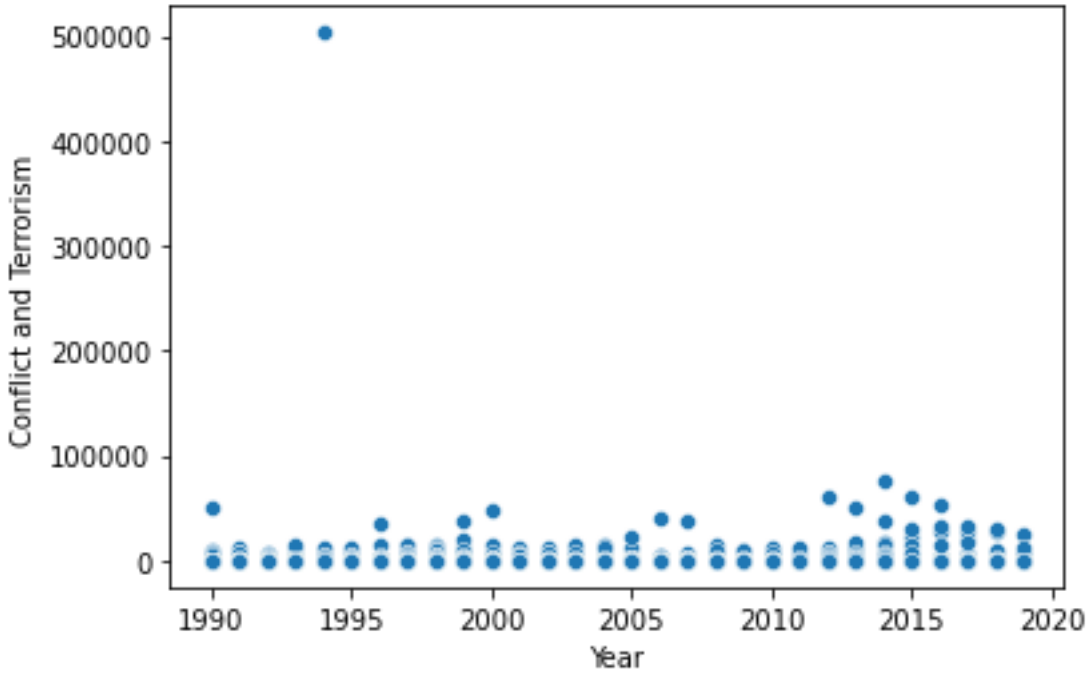
ENVIRONMENTAL HEAT AND COLD EXPOSURE-



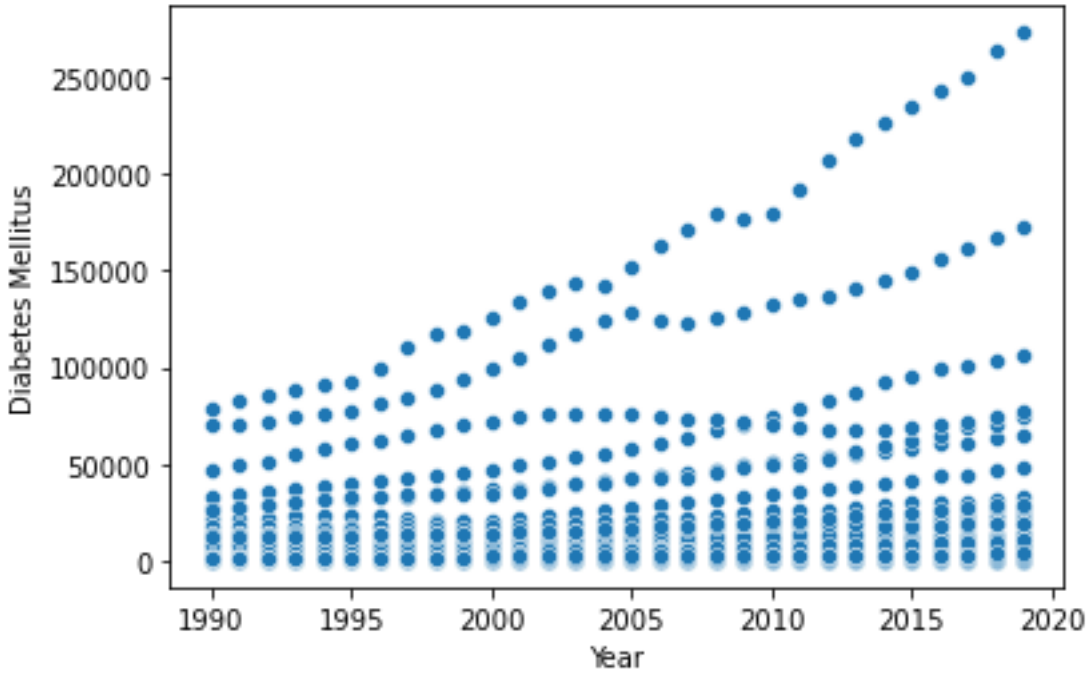
NEOPLASMS-



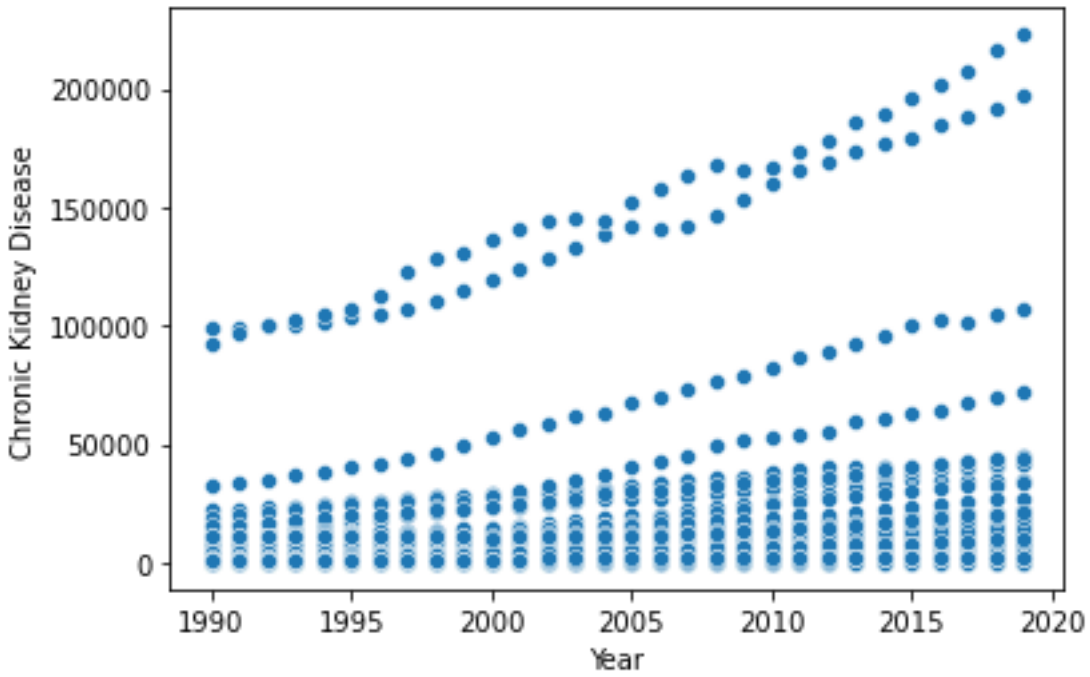
CONFLICT AND TERRORISM-



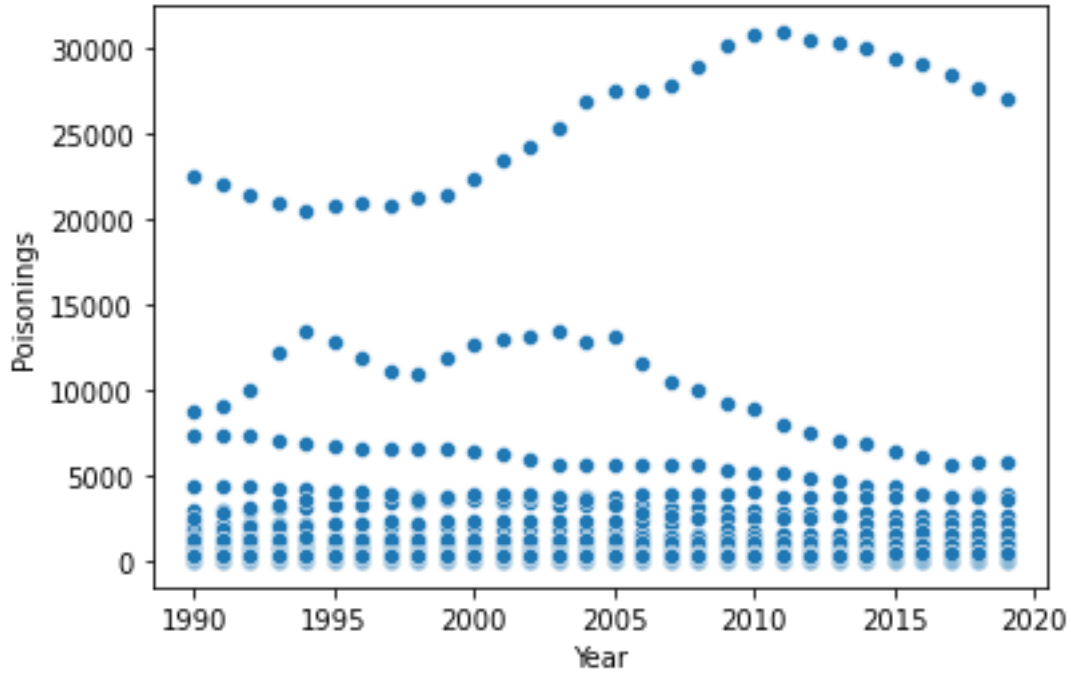
DIABETES MELLITUS-



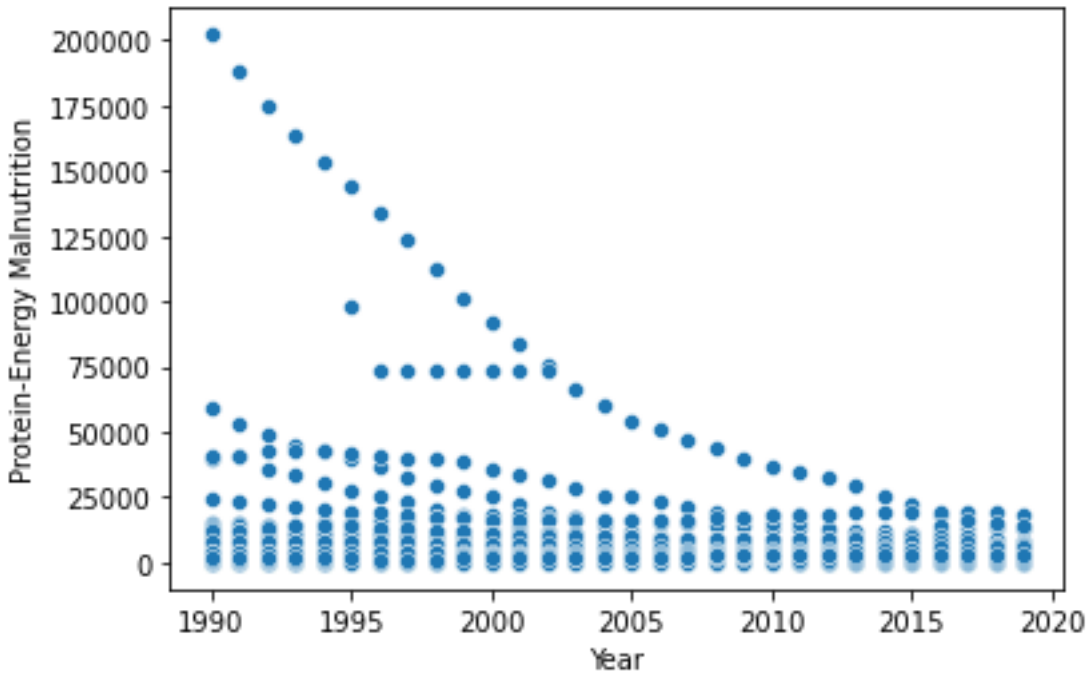
CHRONIC KIDNEY DISEASE-



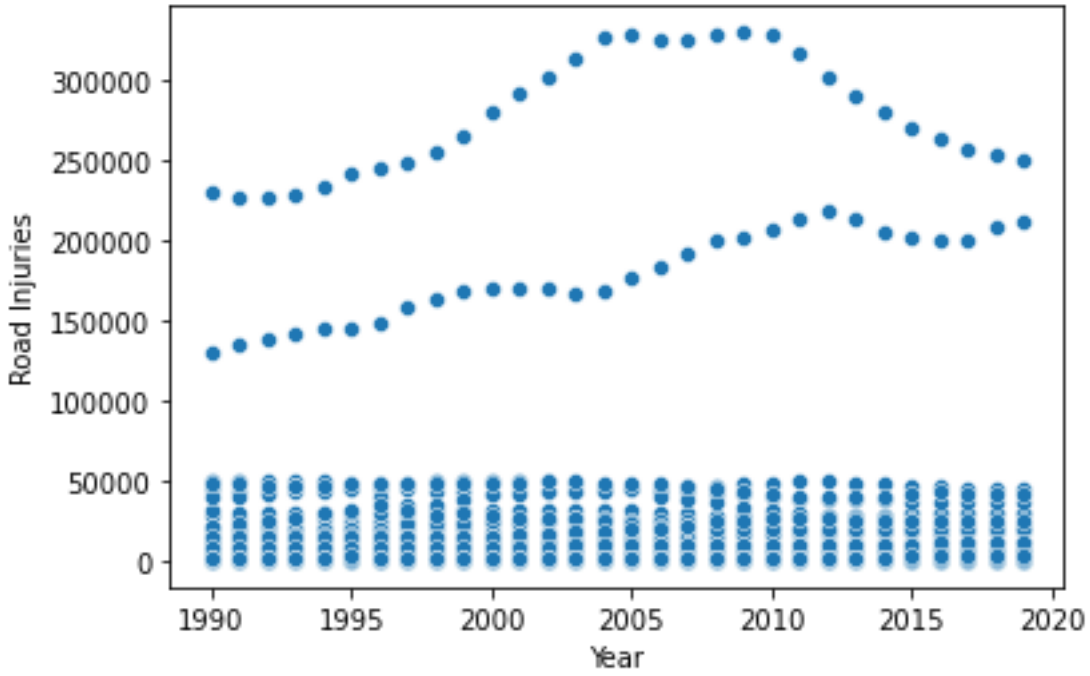
POISONINGS-



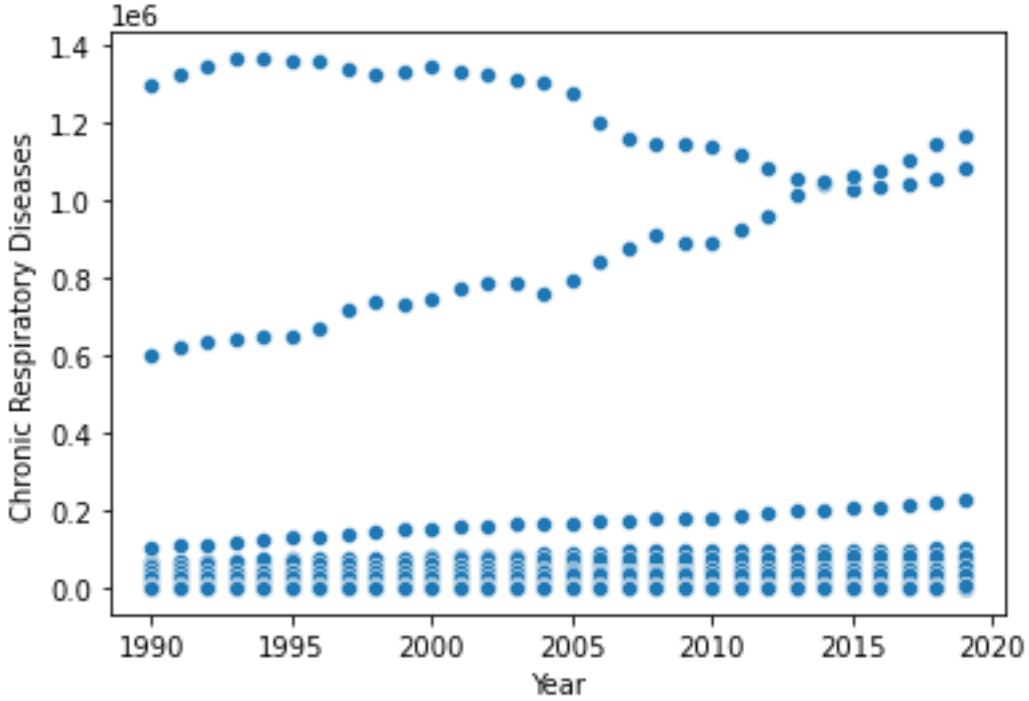
PROTEIN-ENERGY MALNUTRITION-



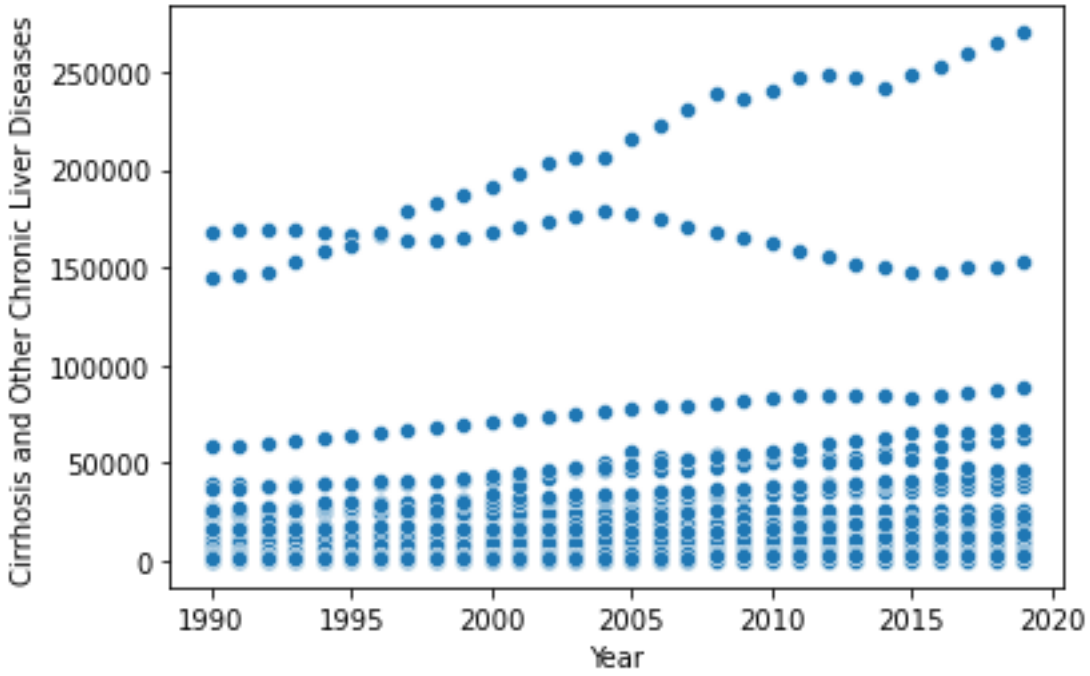
ROAD INJURIES-



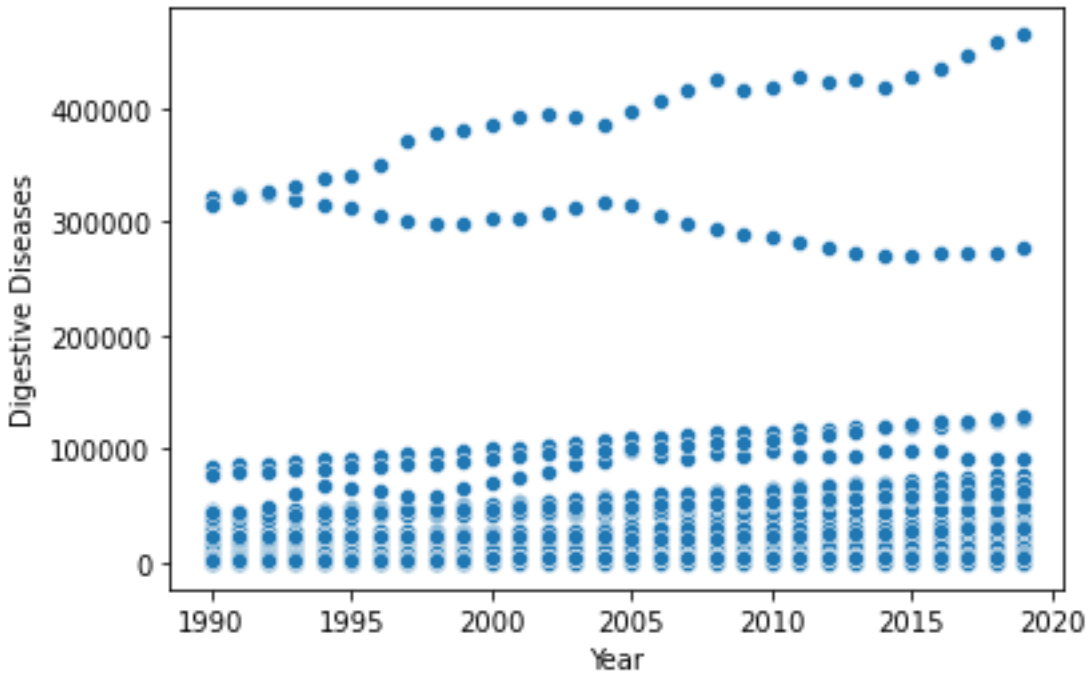
CHRONIC RESPIRATORY DISEASES-



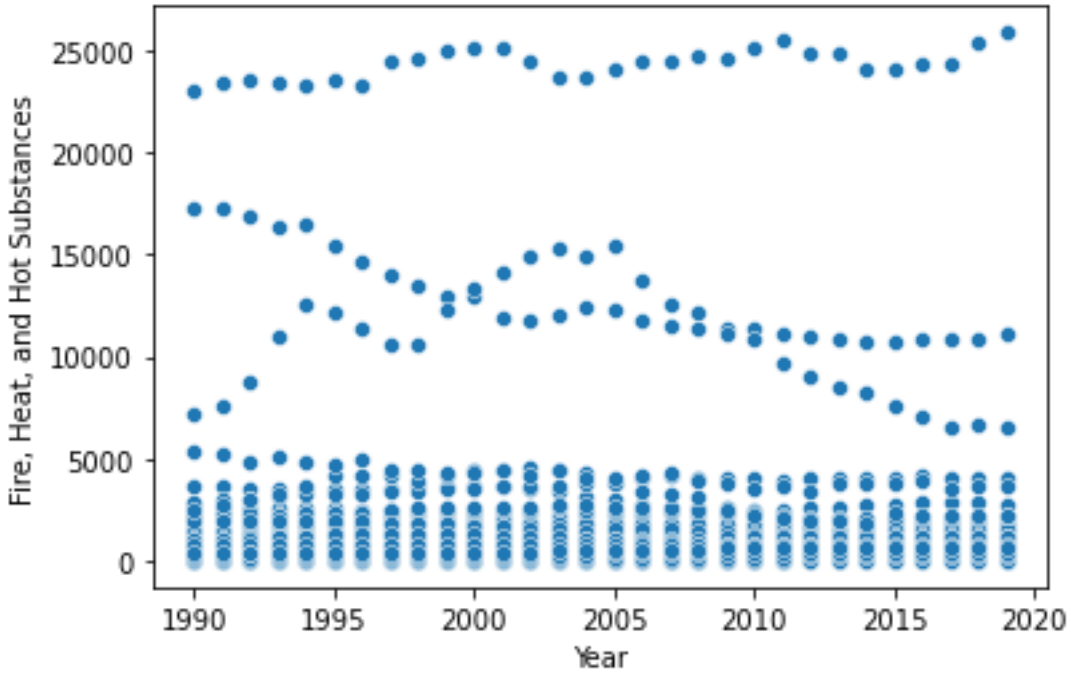
CIRRHOSIS AND OTHER CHRONIC LIVER DISEASES-



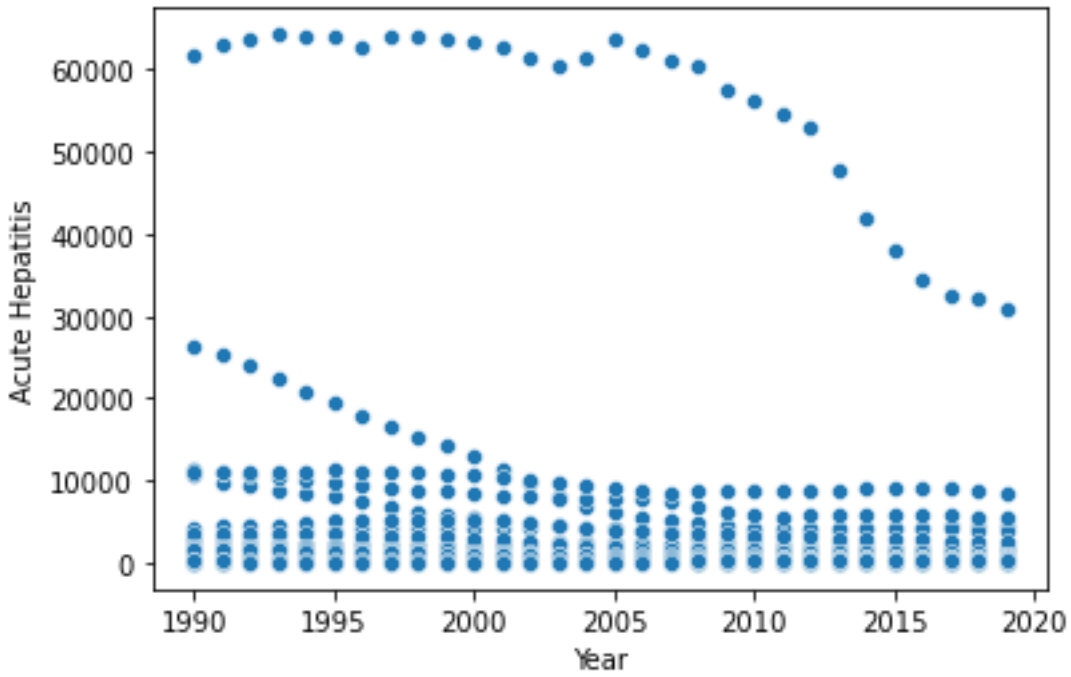
DIGESTIVE DISEASES-



FIRE,HEAT, AND HOT SUBSTANCES-



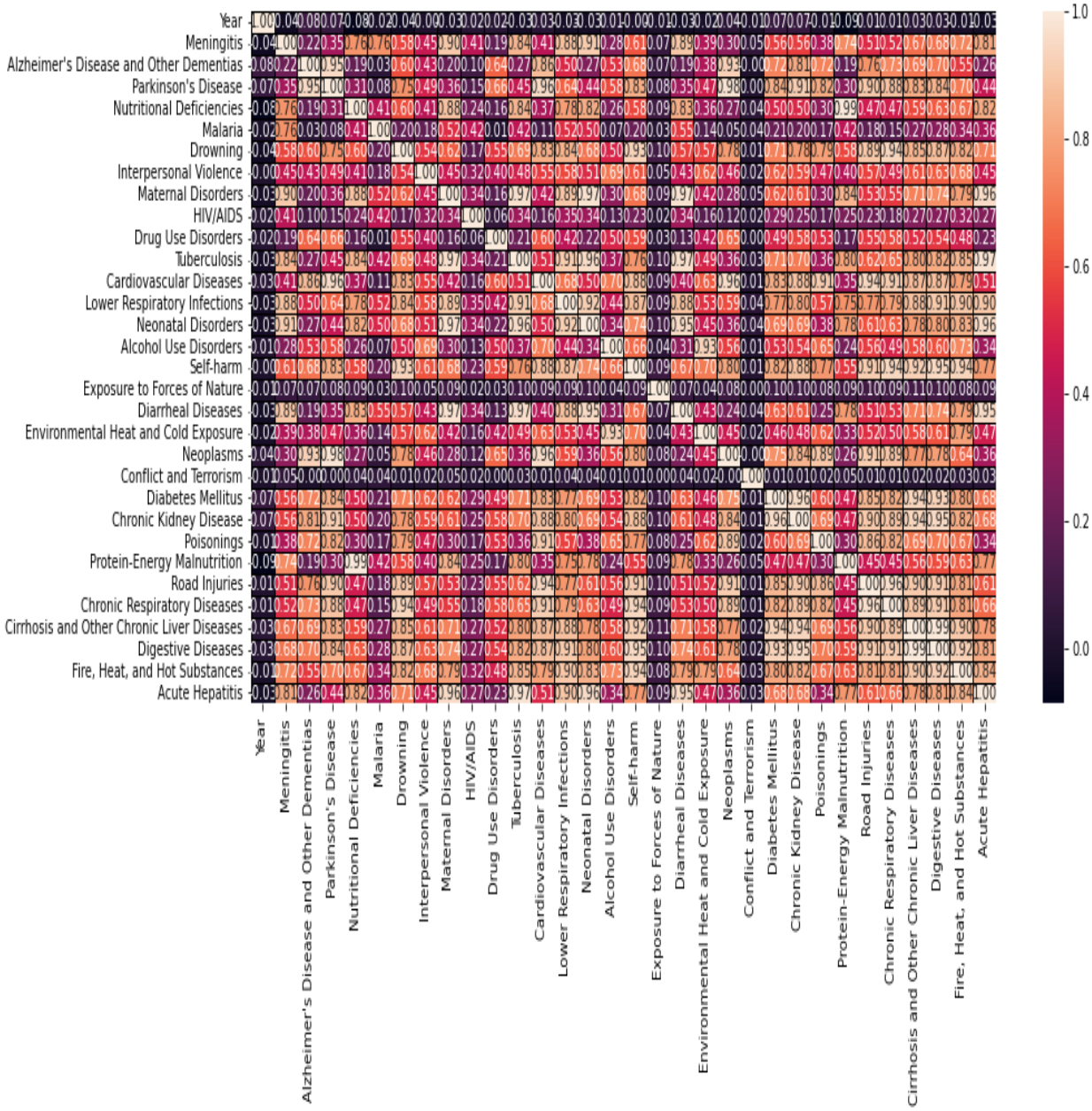
ACUTE HEPATITIS-



MULTIVARIATE ANALYSIS-

We will use correlation between all columns for this.

CORRELATION USING HEATMAP-

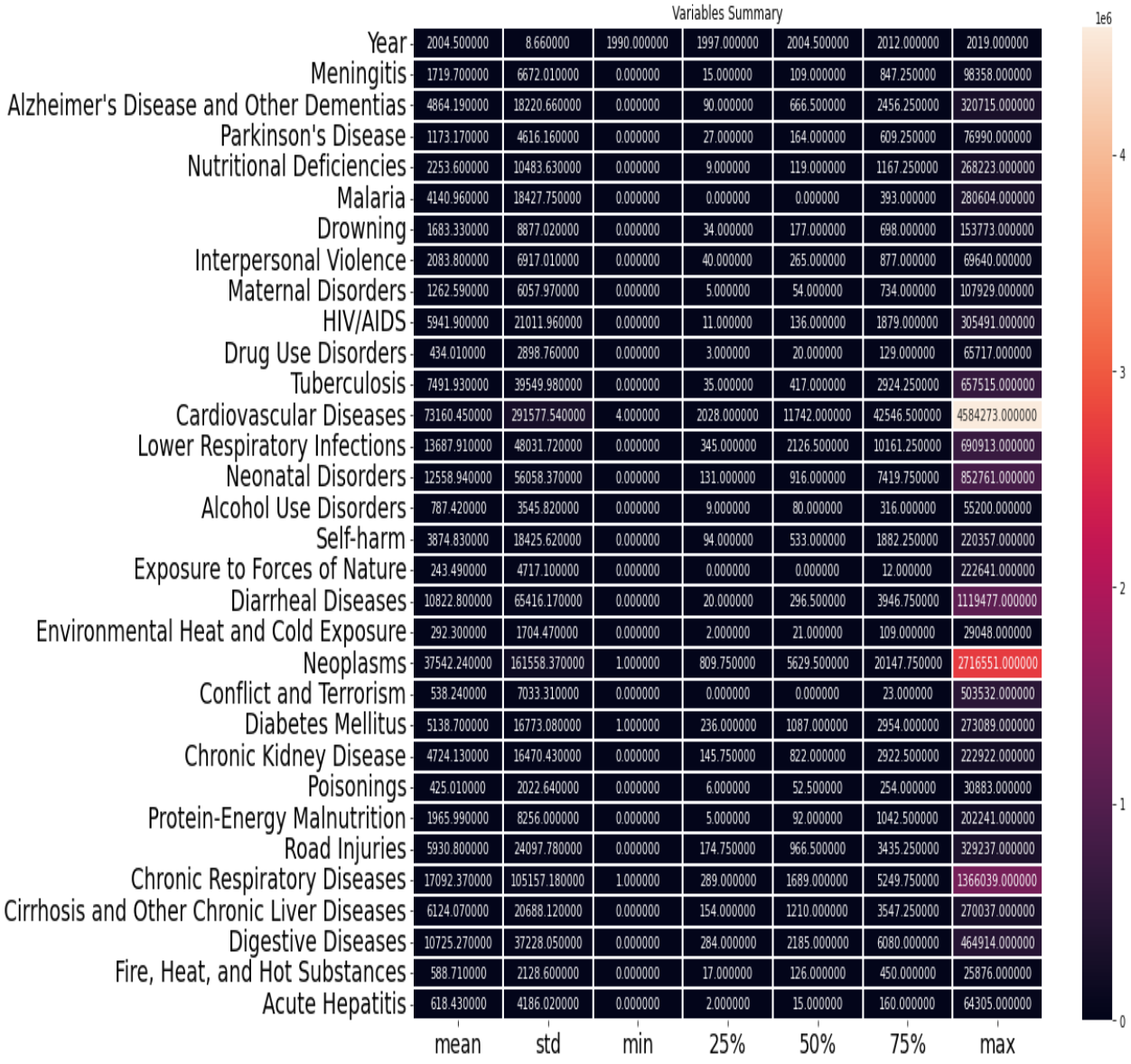


3.3 DESCRIBING THE DATASET

Describing the dataset gives us an understanding of various values like count of the attribute,mean of that attribute, standard deviation, minimum, 25th percentile, median/50th percentile, 75th percentile and the maximum value of that attribute.

We also use heatmap visualization for checking the relation between the described data.

Here, the mean is greater than the median so, the data is right skewed.



4. CONCLUSION

4.1 DATASET CONTENT

Country Or Region : Countries in the world or Regions
Code : Country ISO Code
Year : Year
Number of executions (Amnesty International) : Executions count by the countries
Deaths - Meningitis : Meningitis -> Meningitis is an inflammation (swelling) of the protective membranes covering the brain and spinal cord
Deaths - Neoplasms : Tumors -> Any growth that develops inside or on the body
Deaths - Fire, heat, and hot substances : Death by the high temperature (Not the blood temparature, Depends the environment)
Deaths - Malaria : Malaria -> Malaria is a serious and sometimes fatal disease caused by a parasite that commonly infects a certain type of mosquito which feeds on humans
Deaths - Drowning : Drowning -> Drowning is the process of experiencing respiratory impairment from submersion or immersion in liquid.
Deaths - Interpersonal violence : Interpersonal violence -> The intentional use of physical force or power against other persons by an individual or small group of individuals
Deaths - HIV/AIDS : HIV is a virus that attacks the body's immune system. If HIV is not treated, it can lead to AIDS (acquired immunodeficiency syndrome)
Deaths - Drug use disorders : Death, depends Drug use disorders
Deaths - Tuberculosis : Tuberculosis is caused by a bacterium. The bacteria usually attack the lungs, but TB bacteria can attack any part of the body such as the kidney, spine, and brain
Deaths - Road injuries : Road traffic injuries are defined as "fatal or non-fatal injuries incurred as a result of a road traffic crash
Deaths - Maternal disorders : Maternal Mental Health (MMH) disorders include a range of disorders and symptoms, including but not limited to depression, anxiety and psychosis
Deaths - Lower respiratory infections : Lower respiratory tract infections are any infections in the lungs or below the voice box.
Deaths - Neonatal disorders : Neonatal disorders mean disturbance of normal state of body, organs and abnormal function of a newborn
Deaths - Alcohol use disorders : Death, depends Alcohol
Deaths - Exposure to forces of nature : Definition A force that is beyond human control. This Includes earthquakes, volcanic eruptions, avalanches, storms, lightning strikes, and floods

Deaths - Diarrheal diseases : Diarrheal diseases are a collection of diseases caused by multiple viral, bacterial, and parasitic organisms that share the common symptom of diarrhea
Deaths - Environmental heat and cold exposure : Environmental heat and cold exposure
Deaths - Nutritional deficiencies : Nutritional deficiency occurs when the body is not getting enough nutrients such as vitamins and minerals
Deaths - Self-harm : Self-harm. Like a suicide
Deaths - Conflict and terrorism : Death, depends conflict and terrorism
Deaths - Diabetes mellitus : Diabetes mellitus refers to a group of diseases that affect how your body uses blood sugar (glucose)
Deaths - Poisonings : A poison is any substance that is harmful to your body. You might swallow it, inhale it, inject it, or absorb it through your skin
Deaths - Protein-energy malnutrition : An imbalance between the supply of protein and energy and the body's demand for them to ensure optimal growth and function
Deaths - Cardiovascular diseases : Cardiovascular disease (CVD) is a general term for conditions affecting the heart or blood vessels
Deaths - Chronic kidney disease : Chronic kidney disease is a long-term condition where the kidneys don't work as well as they should. It's a common condition often associated with getting older
Deaths - Chronic respiratory diseases : Chronic respiratory disease is an umbrella term to describe diseases that affect the lungs and airways. Common types include: asthma, cystic fibrosis, lung cancer and sleep apnea
Deaths - Cirrhosis and other chronic liver diseases : Cirrhosis is when scar tissue replaces healthy liver tissue. This stops the liver from working normally. Cirrhosis is a long-term (chronic) liver disease
Deaths - Digestive diseases : A digestive disease is any health problem that occurs in the digestive tract
Deaths - Acute hepatitis : Acute hepatitis is a term used to describe a wide variety of conditions characterized by acute inflammation of the hepatic parenchyma or injury to hepatocytes resulting in elevated liver function indices
Deaths - Alzheimer's disease and other dementias : Dementia is a general term for a decline in mental ability severe enough to interfere with daily life. Alzheimer's is the most common cause of dementia. Alzheimer's is a specific disease
Deaths - Parkinson's disease : Parkinson's disease is a brain disorder that leads to shaking, stiffness, and difficulty with walking, balance, and coordination.

4.2 FINDINGS

We see that, in this dataset the cause of death happened around the world has been due to many natural and unnatural reasons. We have to be mainly careful and precautionary regarding the diseases with upward sloping graphs like- diabetes, kidney malfunction, road injuries etc. We see due to strict Government rules and regulations, many causes have come under control like hunger, natural calamity, fire injuries, poisoning etc.

5. BIBLIOGRAPHY

References:

- https://pbpython.com/pandas_dtypes.html
- https://www.datasciencemadesimple.com/encode-decode-column-dataframe-python/#:~:text=encode%20%28%29%20function%20with%20codec%20%E2%80%98base64%E2%80%99%20and%20error,be%20Decode%20a%20column%20of%20dataframe%20in%20python%3A
- https://www.kaggle.com/code/spscientist/a-simple-tutorial-on-exploratory-data-analysis
- https://en.wikipedia.org/wiki/Exploratory_data_analysis#:~:text=In%20statistics%2C%20exploratory%20data%20analysis,and%20other%20data%20visualization%20methods.
- https://ourworldindata.org/causes-of-death

THE END