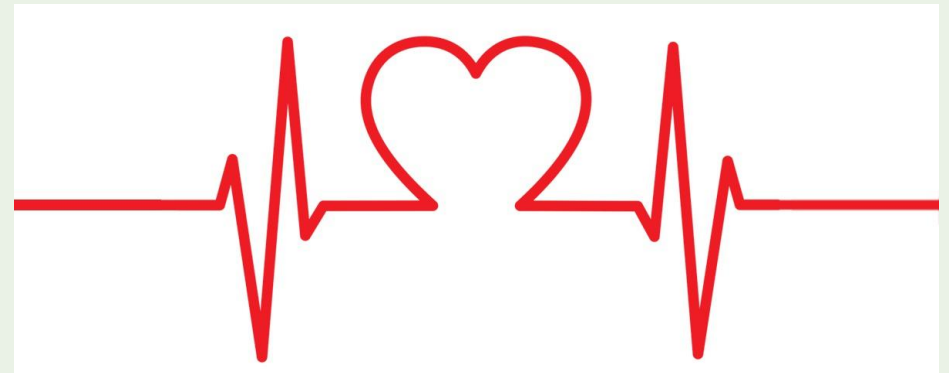


**WORLDWIDE DEATH  
ANALYSIS  
1990-2019**

**IMT 562**

**Exploratory Data Analysis**



# KEY QUESTIONS

---

1) Which are the topmost reasons for deaths across the world?

2) Which countries have been highly succumbed to Terrorism Attacks, Road Injuries, and Suicidal deaths?

3) Are people dying more than the population growth?

4) Countries Ranking by total number of deaths?

# ABOUT DATA

**Data Source 1** : Our World in Data -<https://ourworldindata.org/causes-of-death>

**About Dataset:** The **worldwide death dataset** has country wise death records from 1990-2019 due to 33 reasons, e.g., Cardiovascular, Malaria, Chronic kidney, diabetes, road injuries, terrorism, neoplasm(cancer), etc.

## Data Exploration and Cleaning

- Total Records or Rows: 8590
- Total Columns: 36 (33 death reasons, Entity, Code, Year)
- Entity has 304 unique values which includes different countries, income groups, world regions, etc.
- **Data Quality:** 8323 death records were either missing, unreported, or null for “Execution” reason, 5699 records for “terrorism”, and 370 records for rest of the 30 reasons. I chose not to drop them. Datatypes were matched as expected.
- **Dataset Transformation:** Renamed the column values, unpivoted the death reasons to single column, and filtered the entity by countries’ name only by joining the dataset with a dataset containing 195 countries (Data source 3).
- **Transformed Dataset Information**

Column	Count
Country	184 (Unique)
Year	30 (Unique)
Death Reasons:	33 (Unique)
Total Deaths	182160 (Total)

- Tools Used for EDA and Transformation: **Jupyter Notebook, Azure Data Factory**- Checked data quality using Pandas (Python Library), transformed data using Azure Data Factory and saved it into Azure Blob Storage.
- Tool Used for Visualization: **Tableau** – Tableau does not support connecting to Azure Blob Storage, therefore downloaded the transformed files in csv format and loaded into Tableau

# ABOUT DATA

**Data Source 2 :** The World Bank - <https://data.worldbank.org/indicator/SP.POP.TOTL>

**About Dataset:** The **worldwide population** dataset has country wise population from 1960-2020.

## Data Exploration and Cleaning

- Total Records or Rows: 266
- Total Columns: 65 (1960-2020 years, country name, country code, indicator name, and indicator code)
- Country name has 266 unique values which includes countries' names, income group, world regions, etc.
- **Data Quality:** Population for 2 data records ('Not Classified' and 'West Bank and Gaza') is missing for each of the years. Datatypes were matched as expected.
- **Dataset Transformation:** Unpivoted all the year columns to single column, filtered the year from 1990-2019, and excluded income group, world regions, etc. from 'country name' column by joining the dataset with a dataset containing 195 countries (Data source 3)
- **Transformed Dataset Information** has no null records.

Column Name	Record Count
Country	181 (Unique)
Population	5415 (Total)
Year	30 (Unique)

- Tools Used for EDA and Transformation: **Jupyter Notebook, Azure Data Factory**

**Data Source 3:** Worldometer- <https://www.worldometers.info/geography/alphabetical-list-of-countries/>

**About Dataset:** It has list of 195 countries.

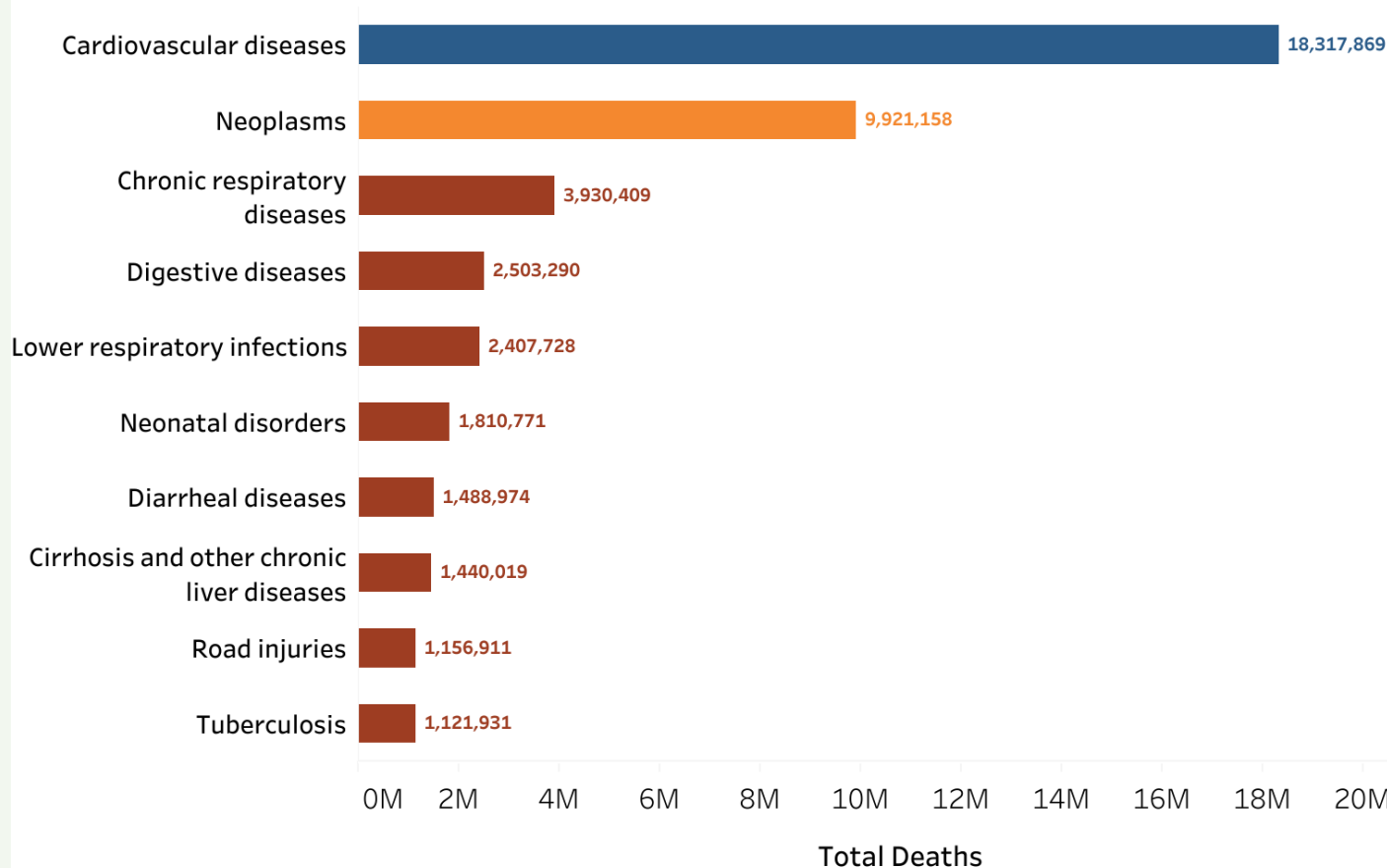
## WORKBOOK LINK:

[HTTPS://PUBLIC.TABLEAU.COM/APP/PROFILE/SAMAN.ATEEQ4840/VIZ/WORLDDWIDEDEATHANALYSIS/GEOGRAPHICALANALYSIS](https://public.tableau.com/app/profile/saman.ateeq4840/viz/worldwideDeathAnalysis/GEOGRAPHICALANALYSIS)

# EXPLORATORY VISUALIZATION 1

## What do People die from?

Data Source has death records (1990-2019) of 184 countries with 33 death reasons.  
Cardiovascular diseases is the leading cause for deaths all around the world. Cancer takes the next position.

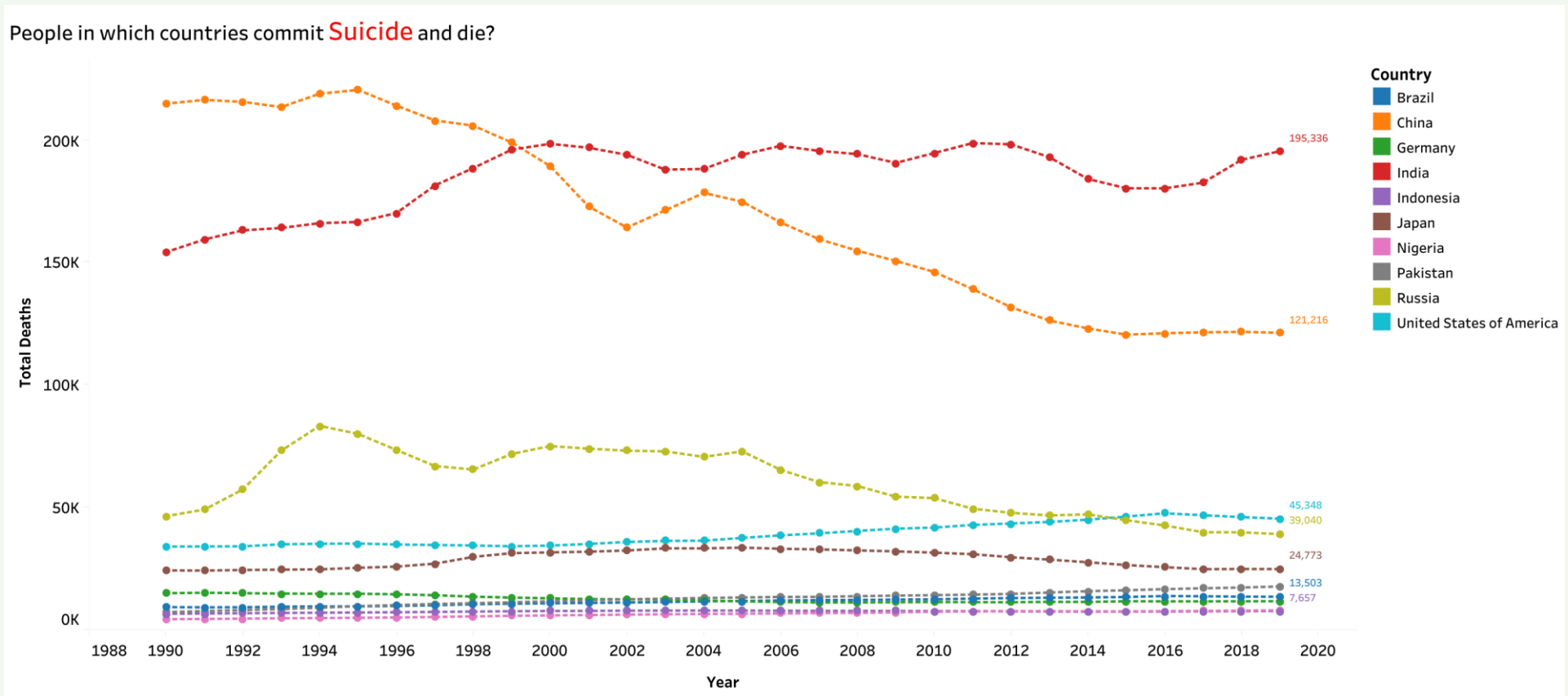


The snapshot is for **worldwide deaths** in the year of **2019**. There is a total of 33 reasons for death in the dataset and the visualization depicts top 10 reasons for worldwide deaths.

I chose a bar chart as length of bar chart can be easily compared with the number of deaths. I sorted the bar chart in descending order of total number of deaths.

By looking at the visualization, we can see that cardiovascular disease is the topmost reasons of worldwide death followed by neoplasm or cancer. To my surprise, Road Injuries is one of the top 10 reasons for death. The top 10 reasons of death can be further investigated for individual countries.

# EXPLORATORY VISUALIZATION 2: PART-1

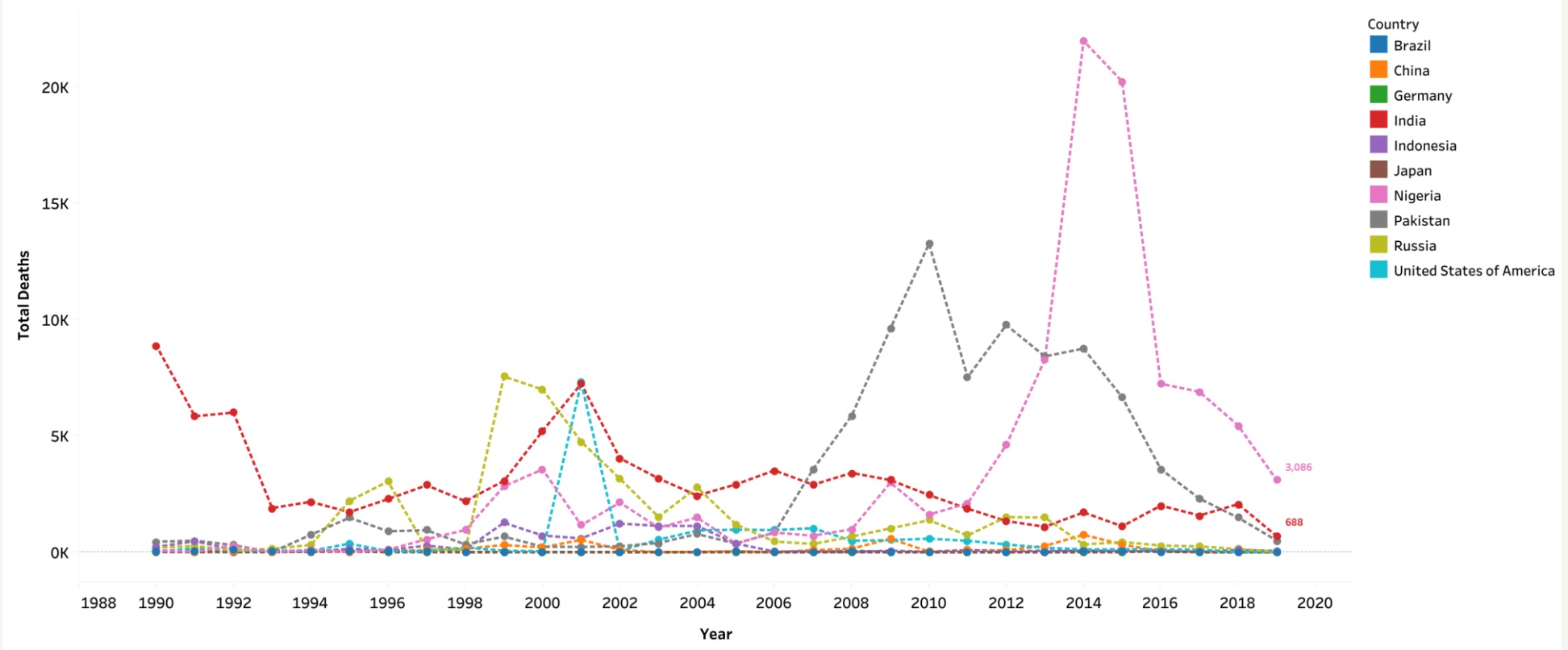


I chose a line chart for visualizing top 10 countries with the highest number of deaths due to self-harm or suicide as line-chart is well-suited for analyzing the trend over time. We can see that there has been decline in suicidal deaths in China, rest of the countries have almost constant suicidal deaths over years. Further analyzing these countries by population, I found that these all countries are one of the populous countries of the world where China and India has 3-4 times higher population than rest of the countries.

# EXPLORATORY VISUALIZATION 2: PART-2

Which 10 countries have been terrified by the **Terrorism and Conflicts** the most?

Nigeria has been hit badly by terrorism in the last few years

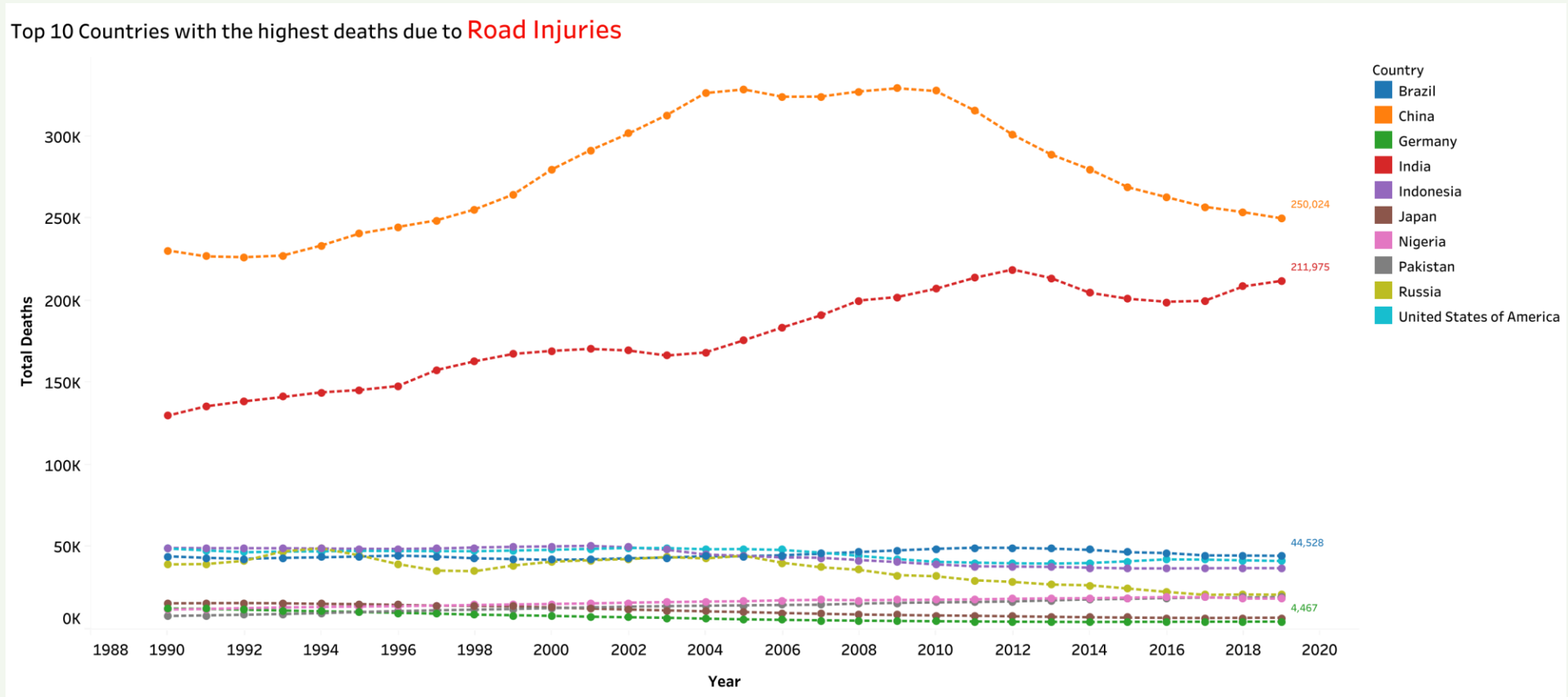


Here also, I chose a dashed line chart with connected dots for visualizing top 10 countries with the highest number of deaths due to terrorist attacks. We can see below facts.

- Nigeria has been hit very badly due to terrorist attacks in the last few years.
- Also, deaths have been at peak in USA and India in the year of 2001, it was the year of 9/11 attack by Al-Qaeda terrorist group in USA and Indian Parliament Attack, respectively.
- There has been a high terrorist activities in Pakistan during 2007-2016



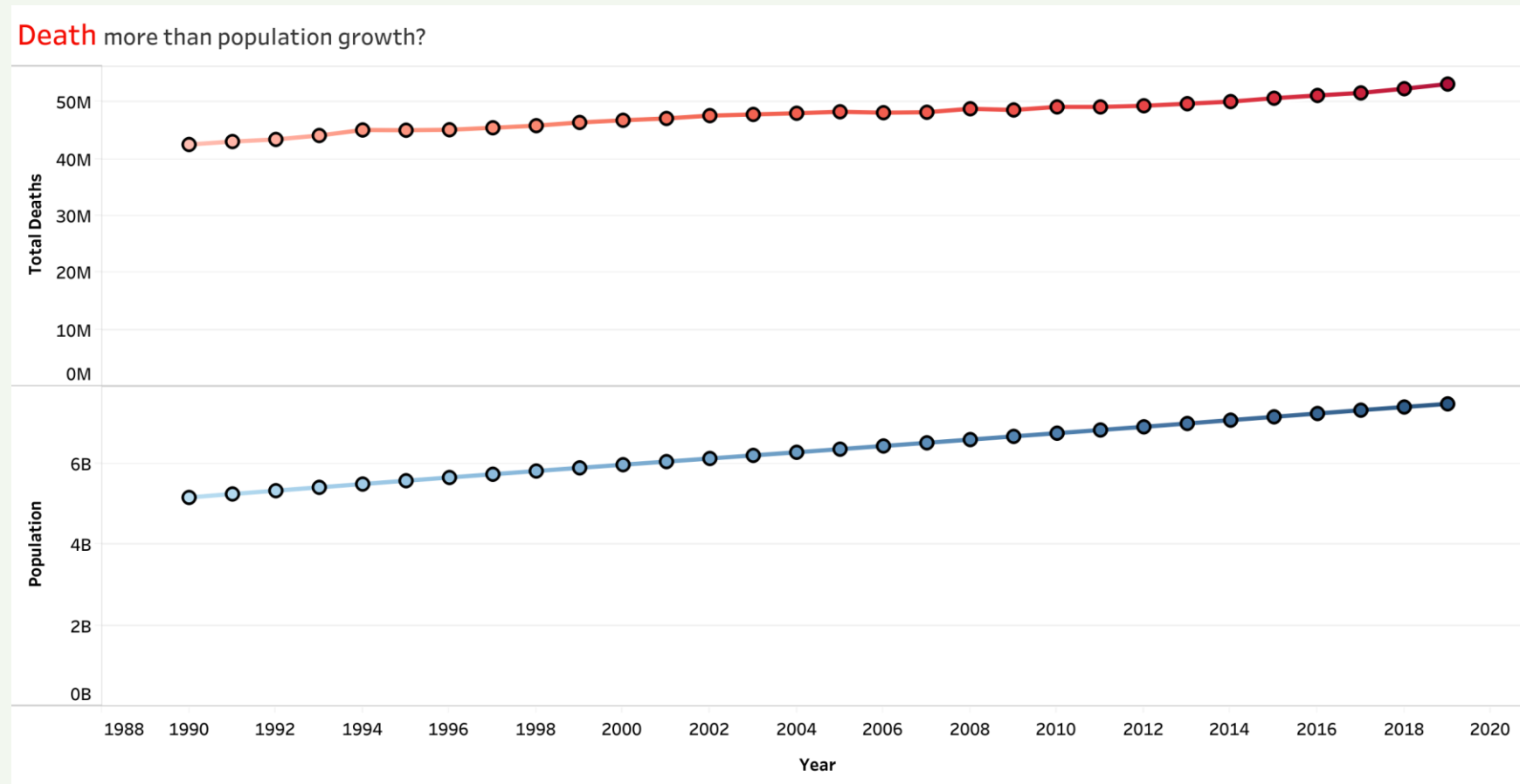
# EXPLORATORY VISUALIZATION 2: PART-3



The dots in the dashed line chart show the years and connectivity is shown to visualize trend of deaths due to road injuries.

We can observe that road injuries have been almost constant in these countries over the years. China is leading in the death numbers due to road accidents.

# EXPLORATORY VISUALIZATION 3

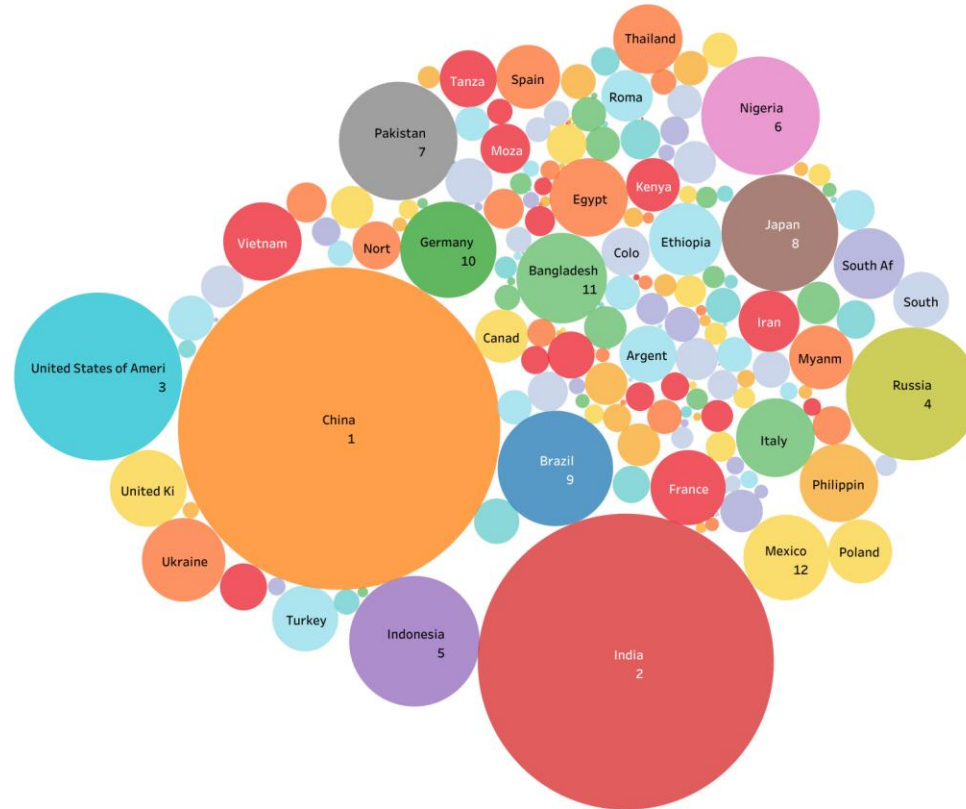


The visualization is for comparing the population growth and death over years. The x-axis (Year) has been set as same for both the categories and Y-axis range is different for deaths and population.

The snapshot is for worldwide deaths and population, and we can see that deaths and population both seem to be increasing over years. Further interacting with visualization, to my surprise I found that there has been surges in death count in **Syria**.

# EXPLORATORY VISUALIZATION 4

Countries Ranking By **Death**



I chose packed bubbles to visualize the countries' ranking by the total number of deaths with the tooltips enabled. Here, bubble size has been sorted by rank in ascending order. Labels are not visible for smaller bubbles, but details can be checked by using tooltip. The snapshot is for 2019 with all death reasons. The visualization can be filtered by Year and Death Reason.

# KEY INSIGHTS

The iteration in EDA helped me to deep delve into the questions I asked for analysis. Below are the major insights.

- People are more likely to die of cardiovascular, digestive, respiratory diseases, and Cancer, etc.
- Top 10 countries with the highest number of deaths are populous countries too.
- India, China, USA, Russia, Germany, etc. countries need to take strong measures for preventing the road accidents and people equally need to be careful on roads.
- Pakistan, India, Nigeria, etc. are the hot centers for terrorist attacks.
- People die from suicide therefore there should be enough emphasis on taking care of mental health.