Data 230: Data Visualization

Extinguished Lights: A Visual deep dive into Infant Mortality

Name: Eshita Gupta
Student ID: 016978741
Email: eshita.gupta@sjsu.edu
GitHub ID: eshita1991

Table of Contents

1 Introduction

"Infant mortality" is a term encapsulating the heartbreaking reality of babies succumbing to death before they reach the age of one. This unfortunate phenomenon transcends geographical and racial boundaries, impacting societies worldwide irrespective of their socio-economic standing. Despite remarkable strides in modern medicine leading to a decline in such cases, infant mortality remains a critical metric reflecting the overall health and well-being of a community.

Delving into a detailed study on the causes of infant mortality holds profound significance. Such an inquiry not only sheds light on the immediate factors leading to infant deaths but also unravels deeper-seated issues. The study can reveal intricate patterns, region-specific determinants, socio-economic influences, and disparities in healthcare access that contribute to the occurrence of infant mortality. This understanding is extremely critical in formulating targeted interventions and policies aimed at mitigating the root causes.

With access to extensive geographical data and comprehensive statistics for countries worldwide, there is an opportunity to transform raw information into actionable insights. By leveraging this data, the goal is to conduct a detailed analysis that goes beyond merely identifying regions with higher infant mortality rates. The aim is to discern the various factors at play, recognizing the circumstances that contribute to the loss of infant lives.

The ultimate objective of this research endeavor is to unravel the major contributing factors to infant mortality. Armed with an understanding, policymakers, healthcare professionals, and communities can collaboratively implement strategies to address these factors directly. This approach not only aims at reducing the immediate impact of infant mortality but also seeks to foster long-term improvements in global healthcare practices, ensuring a healthier start for the world's youngest and most vulnerable members.

2 Overview

In this project, the primary objective is to conduct a thorough analysis, exploration, and visualization of data obtained from various sources concerning "Infant Mortality." The goal is to discern the factors influencing the survival of infants under one year old.

The concept revolves around presenting the data in a visual format to facilitate a comprehensive study. By doing so, we intend to delve deep into global trends, root causes, region-specific elements, socio-economic influences, and the accessibility of healthcare. The visualization of raw data serves as a powerful tool to raise awareness about the issue, telling a compelling visual story that underscores its severity.

To achieve this, the process involves cleaning, processing, and organizing the raw data. This refined data can then be utilized to generate trend analyses, identify patterns, and develop

heat maps. These visual representations are instrumental in preparing indicators that can shed light on potential interventions with a positive impact on reducing infant mortality causes.

In essence, this project aims to leverage visualizations to convey a compelling narrative. The intention is not only to understand the factors contributing to infant mortality but also to effectively communicate the urgency of addressing this critical issue through visually impactful and accessible means.

3 Data Source

Source	Name and URL	License	Comments
Kaggle	Name: Child and Infant Mortality https://www.kaggle.com/datasets/programmerrdai/child-and-infant-mortality	CC0: Public Domain	Primary Data Set
Kaggle	Name: World Population https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset	CC0: Public Domain	World Countries Data
Kaggle	Name: GDP per Capita for all countries https://www.kaggle.com/datasets/nitishabharathi/gdp-per-capita-all-countries	CC0: Public Domain	World GDP Data

4 Project Deliverables

Deliverable	Delivery Format
Project Report	DOCX or PDF format
Project Presentation	PPTX
Raw Dataset	CSV or TXT format
Cleaned Dataset	CSV or Tableau Hyper format
Dashboard & Story	Tableau
Code	GitHub

5 Code Repository

Source	GitHub ID	Group	Project URL
GitHub	eshita1991	sjsu-data230	https://github.com/eshita1991/DATA-230-PROJECT/tree/main/Infant%20Mortality

6 Tools Used

Name
Microsoft Word
Microsoft Excel
Microsoft PowerPoint
Tableau Prep Builder
Tableau
Python

7 Data Overview

I used data from 3 different datasets for my overall analysis. My primary data set was "Child and Infant Mortality." This dataset gives us various statistics on infant mortality for different years and countries in CSV format. There are multiple files (35 files) in the original data set, however, for my analysis I will focus only on the following 5 files. In addition to these, I used two additional files from different datasets for country specific data.

- Causes of Death in Children
- Child Deaths by Age
- Child Mortality around the World
- Child Mortality by Sex
- Per Capita Health Expenditure vs Infant Mortality
- Country wise GDP
- World population

I did a quick analysis into each sheet to check the layout and get some basic information about each sheet using Python. I simply imported each sheet into Python using pandas and checked the basic structure and areas that needed cleaning. I did not do the cleanup in Python as I wanted to visually see what I was modifying while joining the different data sheets. I preferred to use Tableau Prep builder for that activity, I will go into detail on the same during **Section 8, Data cleaning**.

7.1 Sheet 1 - Causes of Death in Children

The data in this sheet represented statistics related to causes of death among children under the age of five (both sexes) for the years 1990 to 2019. Each row corresponds to a specific year, and the columns include information on the number of deaths attributed to various causes. Here's a breakdown of the different columns:

```
df.head()
```

✓ 0.0s Python

	Entity	Code	Year	Deaths - Malaria - Sex: Both - Age: Under 5 (Number)	Deaths - HIV/AIDS - Sex: Both - Age: Under 5 (Number)	Deaths - Meningitis - Sex: Both - Age: Under 5 (Number)	Deaths - Nutritional deficiencies - Sex: Both - Age: Under 5 (Number)	Deaths - Other neonatal disorders - Sex: Both - Age: Under 5 (Number)	Deaths - Whooping cough - Sex: Both - Age: Under 5 (Number)	Deaths - Lower respiratory infections - Sex: Both - Age: Under 5 (Number)	Deaths - Congenital birth defects - Sex: Both - Age: Under 5 (Number)	Deaths - Measles - Sex: Both - Age: Under 5 (Number)	Deaths - Neonatal sepsis and other neonatal infections - Sex: Both - Age: Under 5 (Number)	er
0	Afghanistan	AFG	1990	21	10	1709	1779	7112	2455	20224	12850	8649	420	
1	Afghanistan	AFG	1991	41	12	1743	1822	7574	2385	20879	13701	8669	520	
2	Afghanistan	AFG	1992	51	13	1954	2069	8614	2370	23585	15812	8539	662	
3	Afghanistan	AFG	1993	24	16	2252	2427	9458	2659	27116	17855	8949	723	
4	Afghanistan	AFG	1994	52	19	2446	2649	9823	3187	29271	18835	10642	736	

- **Entity**: The country name for which the data is reported
- **Code**: A country code or identifier for each country
- **Year**: The year for which the data is recorded (1990 - 2019). There are different rows for each country and year combination.
- **Deaths - Malaria**: Number of deaths under the age of five attributed to Malaria.
- **Deaths - HIV/AIDS**: Number of deaths under the age of five attributed to HIV/AIDS.
- **Deaths - Meningitis**: Number of deaths under the age of five attributed to Meningitis.
- **Deaths - Nutritional deficiencies**: Number of deaths under the age of five attributed to Nutritional deficiencies.
- **Deaths - Other neonatal disorders:** Number of deaths under the age of five attributed to Other neonatal disorders.
- **Deaths - Whooping cough:** Number of deaths under the age of five attributed to Whooping cough.
- **Deaths - Lower respiratory infections:** Number of deaths under the age of five attributed to Lower respiratory infections.
- **Deaths - Congenital birth defects:** Number of deaths under the age of five attributed to Congenital birth defects.
- **Deaths - Measles:** Number of deaths under the age of five attributed to Measles.
- **Deaths - Neonatal sepsis and other neonatal infections:** Number of deaths under the age of five attributed to Neonatal sepsis and other neonatal infections.
- **Deaths - Neonatal encephalopathy due to birth asphyxia and trauma:** Number of deaths under the age of five attributed to Neonatal encephalopathy due to birth asphyxia and trauma.
- **Deaths - Drowning:** Number of deaths under the age of five attributed to Drowning.
- **Deaths - Tuberculosis:** Number of deaths under the age of five attributed to Tuberculosis.
- **Deaths - Neonatal preterm birth:** Number of deaths under the age of five attributed to Neonatal preterm birth.
- **Deaths - Diarrheal diseases:** Number of deaths under the age of five attributed to Diarrheal diseases.

- **Deaths - Neoplasms:** Number of deaths under the age of five attributed to Neoplasms.
- **Deaths - Syphilis:** Number of deaths under the age of five attributed to Syphilis.

This data provides a detailed breakdown of the causes of death among infants and young children for different countries split by years from 1990-2019, allowing for a comprehensive analysis of the health challenges faced by each demographic.

We see that we need to change the names of columns as they are too long, and it is difficult to use them as-is.

```
df = pd.read_csv("IM - Sheet 1 - Causes of Death in Children.csv")
print(df.columns)
```

✓ 0.1s

```
Index(['Entity', 'Code', 'Year',
       'Deaths - Malaria - Sex: Both - Age: Under 5 (Number)',
       'Deaths - HIV/AIDS - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Meningitis - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Nutritional deficiencies - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Other neonatal disorders - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Whooping cough - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Lower respiratory infections - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Congenital birth defects - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Measles - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Neonatal sepsis and other neonatal infections - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Neonatal encephalopathy due to birth asphyxia and trauma - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Drowning - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Tuberculosis - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Neonatal preterm birth - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Diarrheal diseases - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Neoplasms - Sex: Both - Age: Under 5 (Number)',
       'Deaths - Syphilis - Sex: Both - Age: Under 5 (Number)'],
      dtype='object')
```

Data was mostly in integer format, so I didn't worry too much about changing formats for fields as that can cause issues.

```
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6840 entries, 0 to 6839
Data columns (total 20 columns):
 #   Column                                                                                                  Non-Null Count  Dtype
---  ------                                                                                                  --------------  -----
 0   Entity                                                                                                  6840 non-null   object
 1   Code                                                                                                    6150 non-null   object
 2   Year                                                                                                    6840 non-null   int64
 3   Deaths - Malaria - Sex: Both - Age: Under 5 (Number)                                                    6840 non-null   int64
 4   Deaths - HIV/AIDS - Sex: Both - Age: Under 5 (Number)                                                   6840 non-null   int64
 5   Deaths - Meningitis - Sex: Both - Age: Under 5 (Number)                                                 6840 non-null   int64
 6   Deaths - Nutritional deficiencies - Sex: Both - Age: Under 5 (Number)                                   6840 non-null   int64
 7   Deaths - Other neonatal disorders - Sex: Both - Age: Under 5 (Number)                                   6840 non-null   int64
 8   Deaths - Whooping cough - Sex: Both - Age: Under 5 (Number)                                             6840 non-null   int64
 9   Deaths - Lower respiratory infections - Sex: Both - Age: Under 5 (Number)                               6840 non-null   int64
 10  Deaths - Congenital birth defects - Sex: Both - Age: Under 5 (Number)                                   6840 non-null   int64
 11  Deaths - Measles - Sex: Both - Age: Under 5 (Number)                                                    6840 non-null   int64
 12  Deaths - Neonatal sepsis and other neonatal infections - Sex: Both - Age: Under 5 (Number)              6840 non-null   int64
 13  Deaths - Neonatal encephalopathy due to birth asphyxia and trauma - Sex: Both - Age: Under 5 (Number)  6840 non-null   int64
 14  Deaths - Drowning - Sex: Both - Age: Under 5 (Number)                                                   6840 non-null   int64
 15  Deaths - Tuberculosis - Sex: Both - Age: Under 5 (Number)                                               6840 non-null   int64
 16  Deaths - Neonatal preterm birth - Sex: Both - Age: Under 5 (Number)                                     6840 non-null   int64
 17  Deaths - Diarrheal diseases - Sex: Both - Age: Under 5 (Number)                                         6840 non-null   int64
 18  Deaths - Neoplasms - Sex: Both - Age: Under 5 (Number)                                                  6840 non-null   int64
 19  Deaths - Syphilis - Sex: Both - Age: Under 5 (Number)                                                   6840 non-null   int64
dtypes: int64(18), object(2)
memory usage: 1.0+ MB
```

I saw a lot of Null values in the "Code" field, and these had to be removed in the cleaned dataset.

```
df.isnull().sum()
```
✓ 0.0s Python

```
Entity                                                                                  0
Code                                                                                  690
Year                                                                                    0
Deaths - Malaria - Sex: Both - Age: Under 5 (Number)                                    0
Deaths - HIV/AIDS - Sex: Both - Age: Under 5 (Number)                                   0
Deaths - Meningitis - Sex: Both - Age: Under 5 (Number)                                 0
Deaths - Nutritional deficiencies - Sex: Both - Age: Under 5 (Number)                   0
Deaths - Other neonatal disorders - Sex: Both - Age: Under 5 (Number)                   0
Deaths - Whooping cough - Sex: Both - Age: Under 5 (Number)                             0
Deaths - Lower respiratory infections - Sex: Both - Age: Under 5 (Number)               0
Deaths - Congenital birth defects - Sex: Both - Age: Under 5 (Number)                   0
Deaths - Measles - Sex: Both - Age: Under 5 (Number)                                    0
Deaths - Neonatal sepsis and other neonatal infections - Sex: Both - Age: Under 5 (Number)  0
Deaths - Neonatal encephalopathy due to birth asphyxia and trauma - Sex: Both - Age: Under 5 (Number)  0
Deaths - Drowning - Sex: Both - Age: Under 5 (Number)                                   0
Deaths - Tuberculosis - Sex: Both - Age: Under 5 (Number)                               0
Deaths - Neonatal preterm birth - Sex: Both - Age: Under 5 (Number)                     0
Deaths - Diarrheal diseases - Sex: Both - Age: Under 5 (Number)                         0
Deaths - Neoplasms - Sex: Both - Age: Under 5 (Number)                                  0
Deaths - Syphilis - Sex: Both - Age: Under 5 (Number)                                   0
dtype: int64
```

I saw certain invalid countries in the dataset. These were regions and not countries, however, since my analysis was to be based on different countries, I didn't need this information.

```
df["Entity"].unique()
```
✓ 0.0s

```
array(['Afghanistan', 'African Region (WHO)', 'Albania', 'Algeria',
       'American Samoa', 'Andorra', 'Angola', 'Antigua and Barbuda',
       'Argentina', 'Armenia', 'Australia', 'Austria', 'Azerbaijan',
       'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados', 'Belarus',
       'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan', 'Bolivia',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil', 'Brunei',
       'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon',
       'Canada', 'Cape Verde', 'Central African Republic', 'Chad',
       'Chile', 'China', 'Colombia', 'Comoros', 'Congo', 'Cook Islands',
       'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Cyprus',
       'Czechia', 'Democratic Republic of Congo', 'Denmark', 'Djibouti',
       'Dominica', 'Dominican Republic', 'East Asia & Pacific (WB)',
       'Eastern Mediterranean Region (WHO)', 'Ecuador', 'Egypt',
       'El Salvador', 'England', 'Equatorial Guinea', 'Eritrea',
       'Estonia', 'Eswatini', 'Ethiopia', 'Europe & Central Asia (WB)',
       'European Region (WHO)', 'Fiji', 'Finland', 'France', 'G20',
       'Gabon', 'Gambia', 'Georgia', 'Germany', 'Ghana', 'Greece',
       'Greenland', 'Grenada', 'Guam', 'Guatemala', 'Guinea',
       'Guinea-Bissau', 'Guyana', 'Haiti', 'Honduras', 'Hungary',
       'Iceland', 'India', 'Indonesia', 'Iran', 'Iraq', 'Ireland',
       'Israel', 'Italy', 'Jamaica', 'Japan', 'Jordan', 'Kazakhstan',
       'Kenya', 'Kiribati', 'Kuwait', 'Kyrgyzstan', 'Laos',
       'Latin America & Caribbean (WB)', 'Latvia', 'Lebanon', 'Lesotho',
       'Liberia', 'Libya', 'Lithuania', 'Luxembourg', 'Madagascar',
       'Malawi', 'Malaysia', 'Maldives', 'Mali', 'Malta',
...
       'Uruguay', 'Uzbekistan', 'Vanuatu', 'Venezuela', 'Vietnam',
       'Wales', 'Western Pacific Region (WHO)', 'World',
       'World Bank High Income', 'World Bank Low Income',
       'World Bank Lower Middle Income', 'World Bank Upper Middle Income',
       'Yemen', 'Zambia', 'Zimbabwe'], dtype=object)
```

7.2 Sheet 2 - Child Deaths by Age

The data in this sheet represented the number of deaths for different age groups in different countries over a span of several years. Here's a breakdown of the columns:

```
df.head()
```
✓ 0.0s Python

	Entity	Code	Year	Deaths - All causes - Sex: Both - Age: 7-27 days (Number)	Deaths - All causes - Sex: Both - Age: 1-4 years (Number)	Deaths - All causes - Sex: Both - Age: 0-6 days (Number)	Deaths - All causes - Sex: Both - Age: 28-364 days (Number)
0	Afghanistan	AFG	1990	9869	21812	15560	30929
1	Afghanistan	AFG	1991	10963	21565	17665	32841
2	Afghanistan	AFG	1992	12905	21929	21124	37901
3	Afghanistan	AFG	1993	14399	24292	23522	44325
4	Afghanistan	AFG	1994	15015	28068	24633	47618

- **Entity**: The country or region for which the data is reported
- **Code**: A country code or identifier
- **Year**: The year for which the data is recorded (1990-2019).
- **Deaths - All causes - Sex: Both - Age: 7-27 days (Number):** Number of deaths within the age range of 7 to 27 days, regardless of gender, attributed to all causes.
- **Deaths - All causes - Sex: Both - Age: 1-4 years (Number)**: Number of deaths within the age range of 1 to 4 years, regardless of gender, attributed to all causes.
- **Deaths - All causes - Sex: Both - Age: 0-6 days (Number)**: Number of deaths within the age range of 0 to 6 days, regardless of gender, attributed to all causes.
- **Deaths - All causes - Sex: Both - Age: 28-364 days (Number):** Number of deaths within the age range of 28 to 364 days (approximately 1 month to 1 year), regardless of gender, attributed to all causes.

Performed similar quick analysis using Python on this sheet and my observations were similar. The name of columns was too long and had to be shortened.

```
df = pd.read_csv("IM - Sheet 2 - Child Deaths by Age.csv")
print(df.columns)
```
✓ 0.0s

```
Index(['Entity', 'Code', 'Year',
       'Deaths - All causes - Sex: Both - Age: 7-27 days (Number)',
       'Deaths - All causes - Sex: Both - Age: 1-4 years (Number)',
       'Deaths - All causes - Sex: Both - Age: 0-6 days (Number)',
       'Deaths - All causes - Sex: Both - Age: 28-364 days (Number)'],
      dtype='object')
```

This sheet also had similar NULL counts and invalid countries as observed in the previous sheet. I had to clean these before joining this to the main sheet. Important point to note is that this sheet had the same three columns "**Entity, Code, Year**", that define the basic structure of the previous sheet. So, this helped me to easily join this with the previous sheet.

```
df.isnull().sum()
```
✓ 0.0s

```
Entity                                                    0
Code                                                    690
Year                                                      0
Deaths - All causes - Sex: Both - Age: 7-27 days (Number)       0
Deaths - All causes - Sex: Both - Age: 1-4 years (Number)       0
Deaths - All causes - Sex: Both - Age: 0-6 days (Number)        0
Deaths - All causes - Sex: Both - Age: 28-364 days (Number)     0
dtype: int64
```

```
df["Entity"].unique()
```
✓ 0.0s

```
array(['Afghanistan', 'African Region (WHO)', 'Albania', 'Algeria',
       'American Samoa', 'Andorra', 'Angola', 'Antigua and Barbuda',
       'Argentina', 'Armenia', 'Australia', 'Austria', 'Azerbaijan',
       'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados', 'Belarus',
       'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan', 'Bolivia',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil', 'Brunei',
       'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon',
       'Canada', 'Cape Verde', 'Central African Republic', 'Chad',
       'Chile', 'China', 'Colombia', 'Comoros', 'Congo', 'Cook Islands',
       'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Cyprus',
       'Czechia', 'Democratic Republic of Congo', 'Denmark', 'Djibouti',
       'Dominica', 'Dominican Republic', 'East Asia & Pacific (WB)',
       'Eastern Mediterranean Region (WHO)', 'Ecuador', 'Egypt',
       'El Salvador', 'England', 'Equatorial Guinea', 'Eritrea',
       'Estonia', 'Eswatini', 'Ethiopia', 'Europe & Central Asia (WB)',
```

7.3 Sheet 3 - Child Mortality around the World

The data in this sheet represented estimates of under-five mortality rates for different countries from the year 1950 to 2019. Here's a breakdown of the columns:

```
df.head()
```
✓ 0.0s Python

	Entity	Code	Year	Estimates, 1950 - 2020: Annually interpolated demographic indicators - Under-five mortality (deaths under age 5 per 1,000 live births)
0	Afghanistan	AFG	1950	41.6952
1	Afghanistan	AFG	1951	41.4045
2	Afghanistan	AFG	1952	40.8172
3	Afghanistan	AFG	1953	40.2182
4	Afghanistan	AFG	1954	39.6077

- **Entity**: The country or region for which the data is reported
- **Code**: A country code or identifier
- **Year**: The specific year for which the data is recorded
- "**Estimates, 1950 - 2020: Annually interpolated demographic indicators - Under-five mortality (deaths under age 5 per 1,000 live births)**": This column provides estimates of the under-five mortality rate, which represents the number of deaths of children under the age of five per 1,000 live births each year.

This type of data is crucial for understanding and tracking the health and well-being of children in a population. A decreasing under-five mortality rate over time is often

considered an indicator of improvements in healthcare, nutrition, and overall living conditions.

As observed earlier, "**Entity, Code and Year**" fields were common and was used join the data. Number of NULL values in this sheet was much higher, this was because this sheet had data for more years. Cleaning strategy remained the same for this sheet as well. However, an additional filter was needed to exclude data for unwanted years.

```
df.isnull().sum()
```

✓ 0.0s Python

```
tity                                                                                    0
de                                                                                   3010
ar                                                                                      0
timates, 1950 - 2020: Annually interpolated demographic indicators - Under-five mortality (deaths under age 5 per 1,000 live births)    0
ype: int64
```

7.4 Sheet 4 - Child Mortality by Sex

The data in this sheet represented mortality rates for children under the age of 5 in different regions split by sex. Here's a breakdown of the columns:

```
df.head()
```

✓ 0.0s Python

	Entity	Code	Year	Mortality rate, under-5, female (per 1,000 live births)	Mortality rate, under-5, male (per 1,000 live births)	Population (historical estimates)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	1961	34.370001	35.979999	9169406.0	NaN
2	Afghanistan	AFG	1962	33.810001	35.400000	9351442.0	NaN
3	Afghanistan	AFG	1963	33.250000	34.829999	9543200.0	NaN
4	Afghanistan	AFG	1964	32.729999	34.260001	9744772.0	NaN

- **Entity**: The name of the region or country
- **Code**: A code or identifier associated with the region or country.
- **Year**: The specific year for which the data is recorded.
- **Mortality rate, under-5, female (per 1,000 live births)**: The mortality rate for females under the age of 5, expressed as the number of deaths per 1,000 live births.
- **Mortality rate, under-5, male (per 1,000 live births)**: The mortality rate for males under the age of 5, expressed as the number of deaths per 1,000 live births.
- **Population (historical estimates):** The estimated population for the given year and region.
- **Continent**: The continent to which the region or country belongs.

This data can be valuable for assessing the health and well-being of children in different regions over time. It allows for the examination of gender-specific under-5 mortality rates, which can be crucial for identifying disparities and informing targeted interventions to

improve child health outcomes. The inclusion of population estimates, and continent information adds further context to the data.

I analyzed this further to see if cleanup was needed. As expected based on other sheets, I had to shorten the column names.

```
df = pd.read_csv("IM - Sheet 4 - Child Mortality by Sex.csv")
print(df.columns)
```

✓ 0.0s

```
Index(['Entity', 'Code', 'Year',
       'Mortality rate, under-5, female (per 1,000 live births)',
       'Mortality rate, under-5, male (per 1,000 live births)',
       'Population (historical estimates)', 'Continent'],
      dtype='object')
```

Also, this sheet had a lot of NULL columns, so I had to drop those in the cleaned dataset. Continent information was mostly missing in this dataset, So I removed that.

```
df.isnull().sum()
```

✓ 0.0s

```
Entity                                                  0
Code                                                 2029
Year                                                    0
Mortality rate, under-5, female (per 1,000 live births)    44950
Mortality rate, under-5, male (per 1,000 live births)      44950
Population (historical estimates)                         452
Continent                                               55823
dtype: int64
```

7.5 Sheet 5 - Per Capita Health Expenditure vs Infant Mortality

The data in this sheet included information related to under-5 mortality rates, health expenditure, population estimates, and the continent for various regions or countries. Here's a breakdown of the columns:

```
df.head()
```

✓ 0.0s Python

	Entity	Code	Year	Mortality rate, under-5 (per 1,000 live births)	Current health expenditure per capita, PPP (current international $)	Population (historical estimates)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	1960	35.820001	NaN	8996967.0	NaN
2	Afghanistan	AFG	1961	35.220001	NaN	9169406.0	NaN
3	Afghanistan	AFG	1962	34.629999	NaN	9351442.0	NaN
4	Afghanistan	AFG	1963	34.060001	NaN	9543200.0	NaN

- **Entity**: The name of the region or country
- **Code**: A code or identifier associated with the region or country.
- **Year**: The specific year for which the data is recorded.
- **Mortality rate, under-5 (per 1,000 live births):** The mortality rate for children under the age of 5, expressed as the number of deaths per 1,000 live births.
- **Current health expenditure per capita, PPP (current international $):** The current health expenditure per capita in purchasing power parity (PPP) terms, measured in current international dollars.
- **Population (historical estimates)**: The estimated population for the given year and region.
- **Continent**: The continent to which the region or country belongs.

The health expenditure information provides insights into the financial resources allocated to health services per capita. Overall, this data provides a foundation for understanding the relationship between health expenditure, population health, and under-5 mortality rates over time.

The data in this sheet also had a lot of NULL values. So, I dropped the Continent field in this one as well. Also, it had data since 1960, so a filter was needed as my analysis was only for 1990-2019.

```
df.isnull().sum()
```

✓ 0.0s

```
Entity                                                          0
Code                                                         2029
Year                                                            0
Mortality rate, under-5 (per 1,000 live births)             44882
Current health expenditure per capita, PPP (current international $)    52074
Population (historical estimates)                             452
Continent                                                    55823
dtype: int64
```

7.6 Sheet 6 - Country Wise GDP

This sheet provided information about the Gross Domestic Product (GDP) per capita for different countries over multiple years. Here's a breakdown of the columns:

```
df.head()
```

✓ 0.0s Python

	Country	Country Code	1990	1991	1992	1993	1994	1995	1996	1997	...
0	Aruba	ABW	24101.109430	25870.755940	26533.343900	27430.752400	28656.520210	28648.990020	28499.089430	30215.949230	...
1	Afghanistan	AFG	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
2	Angola	AGO	3089.683369	3120.356148	2908.160798	2190.768160	2195.532289	2496.199493	2794.896906	2953.342709	...
3	Albania	ALB	2549.473022	1909.114038	1823.307673	2057.449657	2289.873135	2665.764906	2980.066288	2717.362124	...
4	Arab World	ARB	6808.206995	6872.273195	7255.328362	7458.647059	7645.682856	7774.207360	8094.149842	8397.515692	...

5 rows × 32 columns

- **Country**: The name of the country.
- **Country Code**: A code or identifier associated with each country.
- **1990 to 2019**: Columns representing the GDP per capita for each year. The values are expressed in international dollars.

This type of data is crucial for understanding the economic well-being of a country and how it changes over time. Analyzing trends in GDP per capita can provide insights into the economic growth or decline of a nation and can be used for comparative analyses between countries.

I did see a lot of NULL values in the data and had to consider these while joining my data. Also, an important thing to observe was that the data was not consistent to the structure of the previous files. So, I had to use the Pivot function to bring the data in the same format as expected by the previous files.

```
df.isnull().sum()
```
✓ 0.0s

```
Country          0
Country Code     0
1990            51
1991            49
1992            47
1993            45
1994            42
1995            36
1996            36
1997            35
1998            34
1999            33
```

7.7 Sheet 7 – World Population

This sheet provided information about various countries and territories, including population data, geographical details, and demographic indicators. Here's a breakdown of the columns:

```
df.head()
```
✓ 0.0s Python

	Rank	Code	Entity	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	
0	36	AFG	Afghanistan	Kabul	Asia	41128771	38972230	33753499	28189672	19542982	10694796	12486631	10752971	6
1	138	ALB	Albania	Tirana	Europe	2842321	2866849	2882481	2913399	3182021	3295066	2941651	2324731	
2	34	DZA	Algeria	Algiers	Africa	44903225	43451666	39543154	35856344	30774621	25518074	18739378	13795915	23
3	213	ASM	American Samoa	Pago Pago	Oceania	44273	46189	51368	54849	58230	47818	32886	27075	
4	203	AND	Andorra	Andorra la Vella	Europe	79824	77700	71746	71519	66097	53569	35611	19860	

- **Rank**: The rank of the entity based on a certain criterion (not explicitly defined in the data, but possibly related to population size or other factors).
- **Code**: A code or identifier associated with each country or territory.
- **Entity**: The name of the country or territory.
- **Capital**: The capital city of the country or territory.
- **Continent**: The continent to which the country or territory belongs.
- **2022 Population**: The estimated population for the year 2022.
- **2020 Population**: The estimated population for the year 2020.
- **2015 Population**: The estimated population for the year 2015.
- **2010 Population**: The estimated population for the year 2010.
- **2000 Population**: The estimated population for the year 2000.
- **1990 Population**: The estimated population for the year 1990.
- **1980 Population**: The estimated population for the year 1980.
- **1970 Population**: The estimated population for the year 1970.
- **Area (km²):** The land area of the country or territory in square kilometers.
- **Density (per km²):** The population density, calculated as the number of people per square kilometer.
- **Growth Rate**: The population growth rate, indicating the percentage change in population over a specified period.
- **World Population Percentage**: The percentage of the world's total population that the entity represents.

This dataset allows for an exploration of population dynamics, geographical details, and demographic indicators for various countries and territories. It can be used for comparative analyses, studying population growth trends, and understanding the distribution of the world's population.

This data was mostly clean and did not need a lot of cleaning.

```
df.isnull().sum()
```

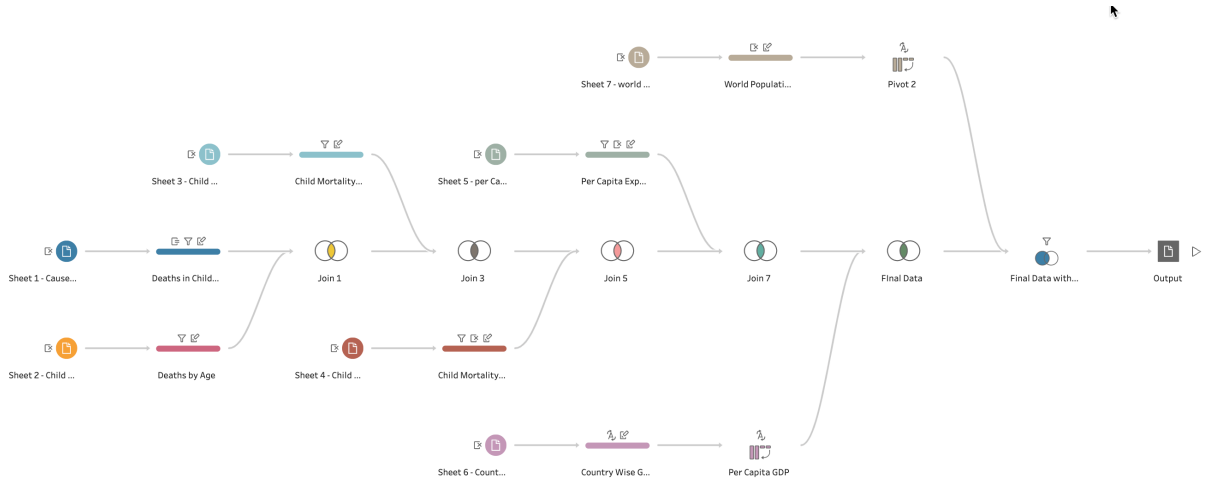
✓ 0.0s

```
Rank                            0
Code                            0
Entity                          0
Capital                         0
Continent                       0
2022 Population                 0
2020 Population                 0
2015 Population                 0
2010 Population                 0
2000 Population                 0
1990 Population                 0
1980 Population                 0
1970 Population                 0
Area (km²)                      0
Density (per km²)               0
Growth Rate                     0
World Population Percentage     0
dtype: int64
```

8 Data Cleansing & Preparation

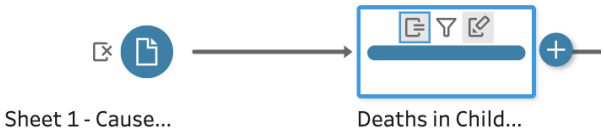
I analyzed details about the basic structure of each file and type of cleaning required for each of them before starting with the visualizations. I decided to use the Tableau Prep builder over Python for the same as it would be easier to visualize my data while cleaning and joining it.

My final data flow from Tableau prep builder is as follows. I generated a Tableau specific **(.hyper)** file to be used an input data source to Tableau for my visualizations.

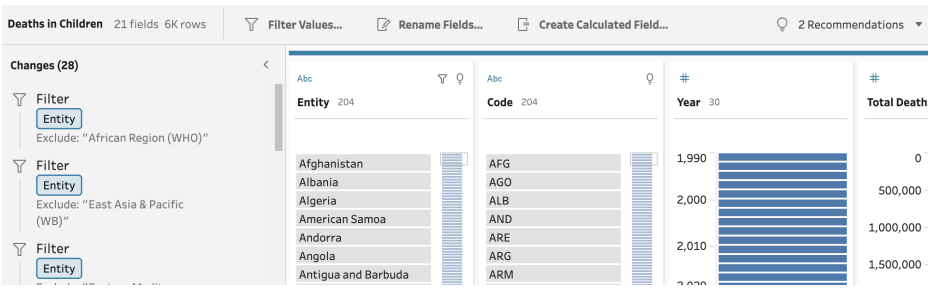


Following are the details about the activities performed on each file.

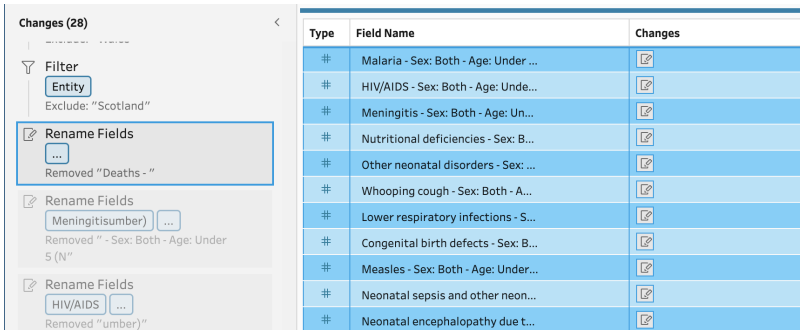
- **Sheet 1 - Causes of Death in Children**



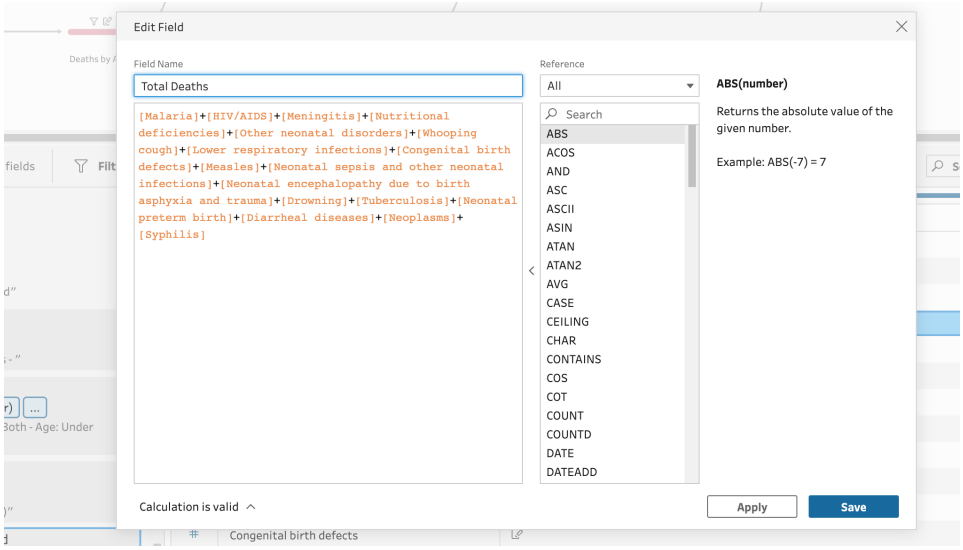
  - Removed all invalid countries, example "**African Region (WHO)**" and filtered the field code to remove NULL values.



  - Renamed all the fields to have removed unnecessary text.

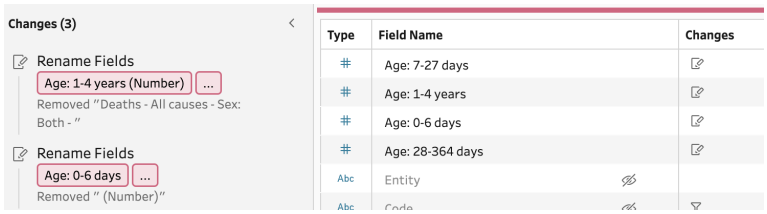


- Created a calculated Field for Total Death Counts

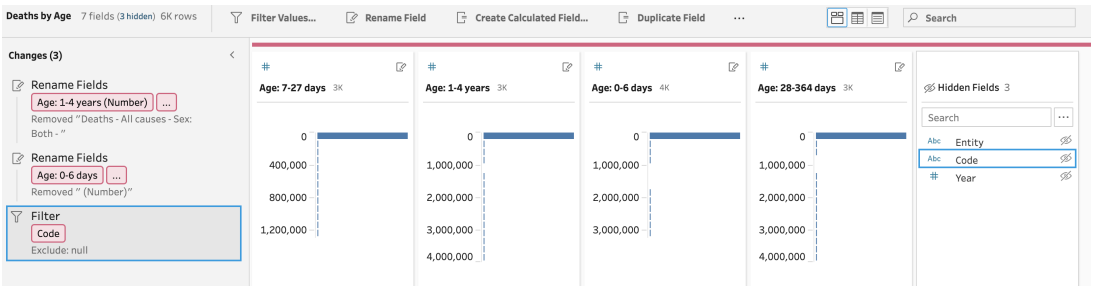


- **Sheet 2 - Child Deaths by Age**

- Renamed the fields to remove unnecessary text



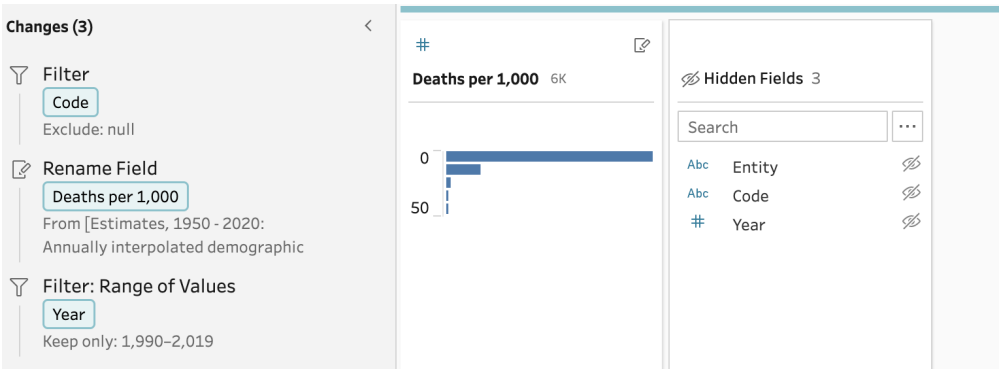
- Filtered the column “Code” to exclude NULL Values and marked the fields “Entity, Code, Year” as hidden.



- **Sheet 3 - Child Mortality around the World**

- Filtered the field “Code” to exclude all NULL values, renamed the fields to remove unnecessary text and marked “Entity, Code and Year” as hidden.

  Also, filtered the data to only have data from 1990-2019.

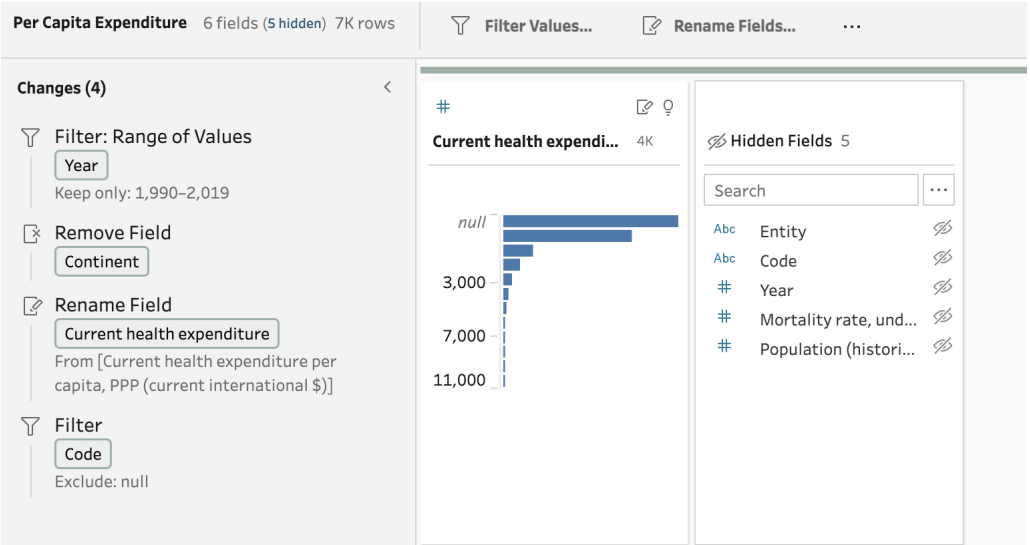


- **Sheet 4 - Child Mortality by Sex**

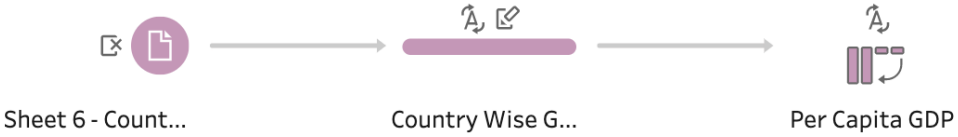
- Renamed the fields, excluded NULL values from Code field and filtered the data only for year 1990-2019. Also, marked “Entity, Code and Year” as hidden

- **Sheet 5 - Per Capita Health Expenditure vs Infant Mortality**

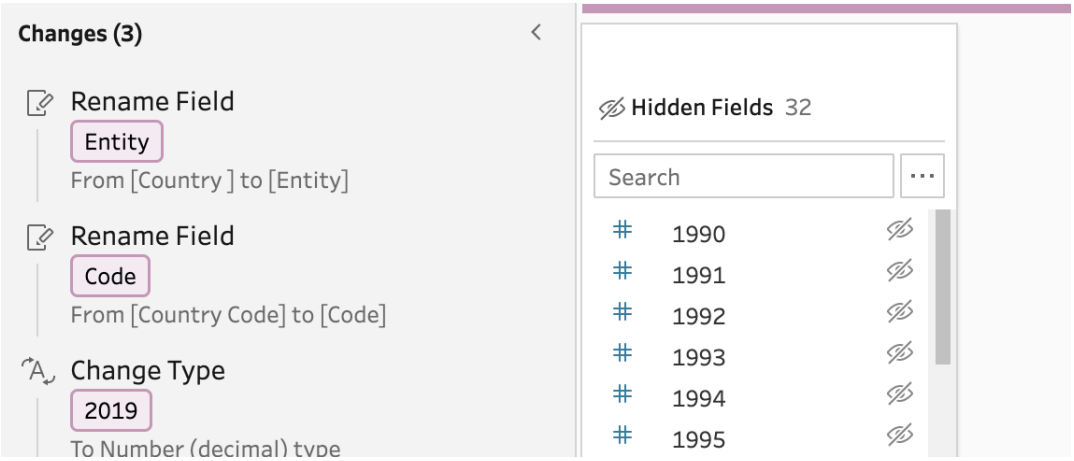
  - Excluded NULL values from Code, Removed the field Continent as it was mostly blank. Also filtered the data for year 1990-2019. Renamed the fields as needed.



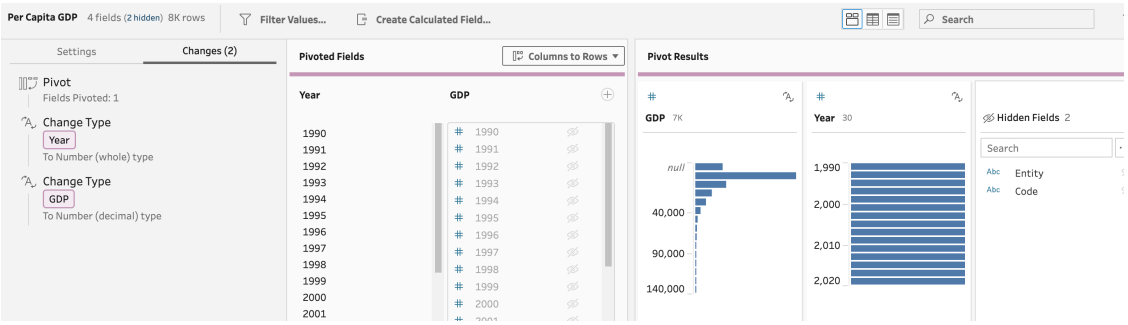
- **Sheet 6 - Country wise GDP**



- Renamed the fields to match with the other sheets in the dataset.



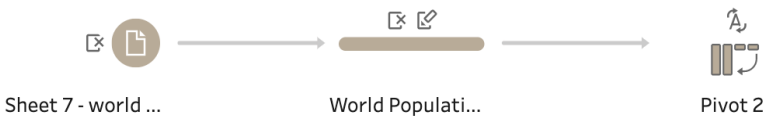
- Created a pivot view to have Year wise data for each country in the same structural format as the other sheets.



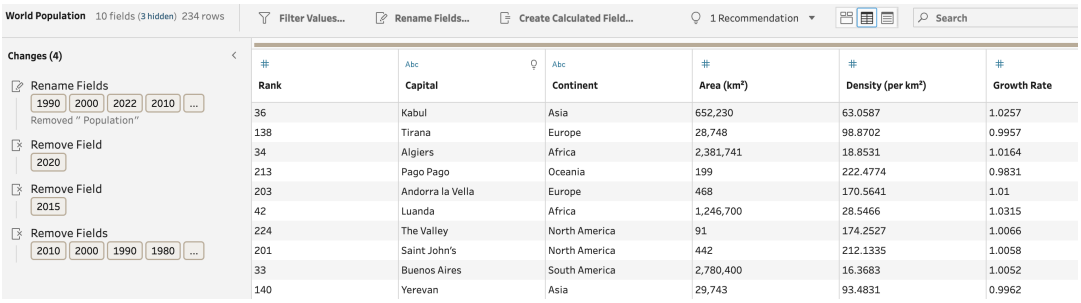
Pivot Results

GDP	Year	Entity	Code
24,101.10943	1,990		
25,870.75594	1,991		
26,533.3439	1,992		
27,430.7524	1,993		
28,656.52021	1,994		
28,648.99002	1,995		
28,499.08943	1,996		
30,215.94923	1,997		
30,512.68391	1,998		

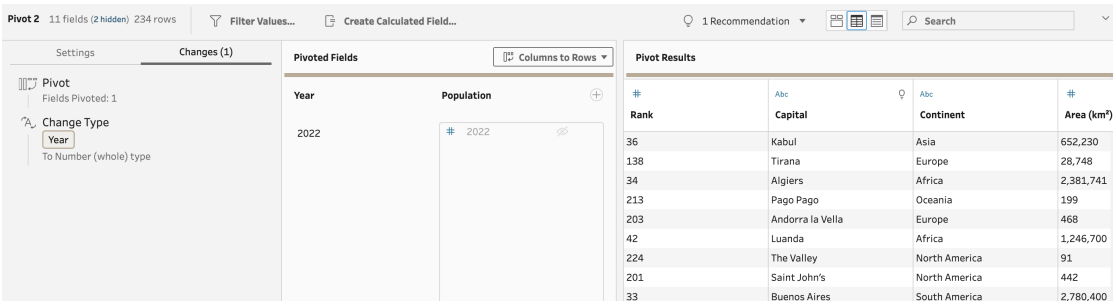
- **Sheet 7 - World population**



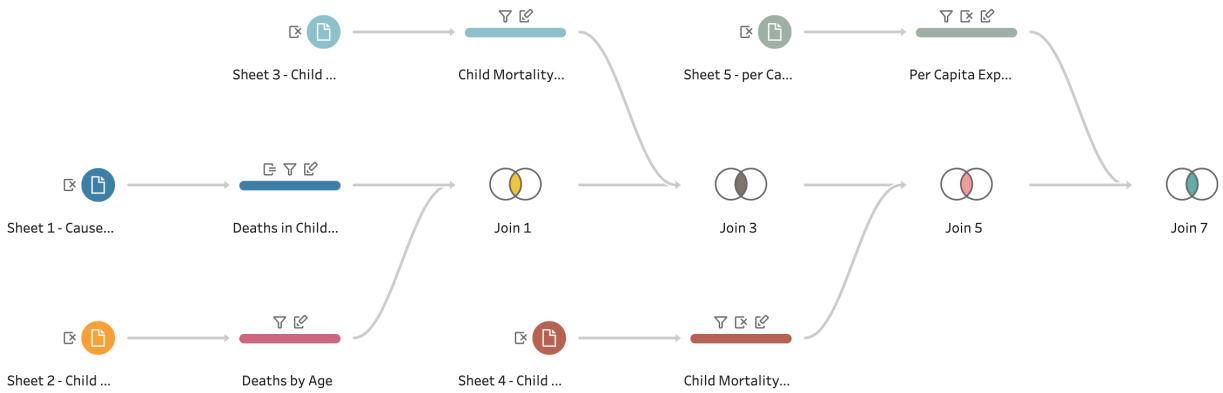
- Removed unnecessary fields.



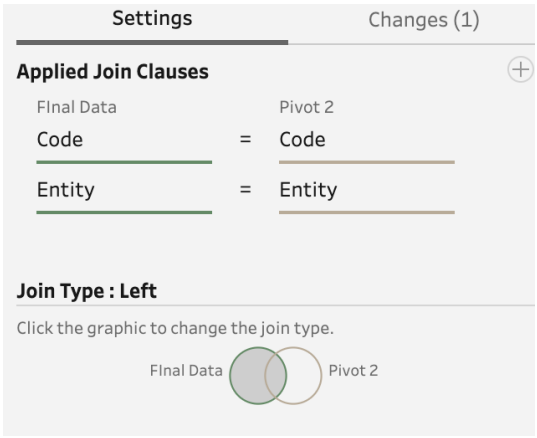
- Created a pivot view to have the data in the format aligned with previous sheets

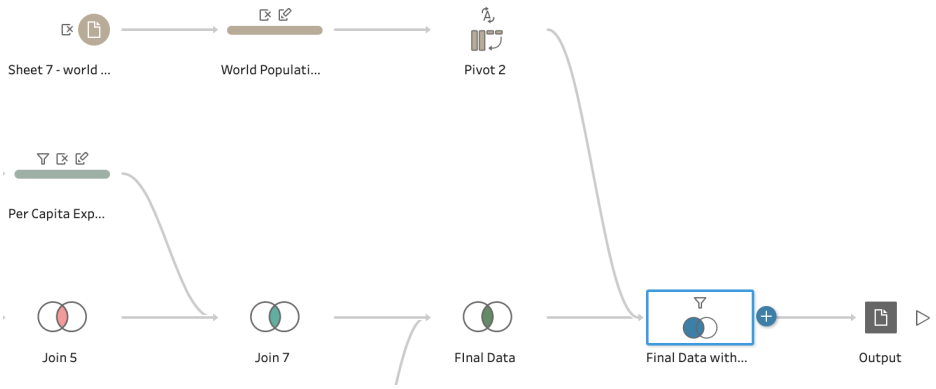


Finally joined all the sheets together using simple JOINS in Tableu prep.

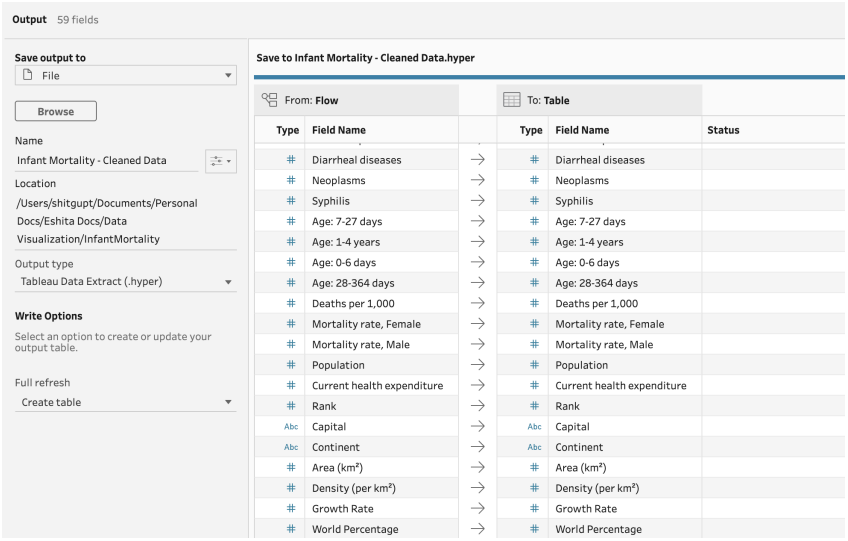


All the sheets were joined using simple inner joins. However, the Sheet 7 - world population did not have year wise data. It only had data for 2019, so had to create an outer join and ensure all my data before my current join was visible and I joined it using only 2 conditions instead of 3. This ensured all my rows were populated with the same details for the fields like “Continent, Capital, Density, Area etc.”. There would be minor differences in these fields if compared to data from 1990, however, to keep it consistent, I used the current data.





The final output of this flow was as follows. It had all the data I needed for my visualizations along with some additional data. With this, I was ready to start visualizing the data and extract meaningful insights.



9 Data Visualization

Infant mortality, a critical indicator of a society's health and well-being, continues to be a focal point of global concern. In this series of dashboards, we delve into the intricate landscape of infant mortality, employing a comprehensive approach that combines data from various sources. My visualizations aim to illuminate trends, patterns, and critical factors that influence infant mortality rates across regions and over time.

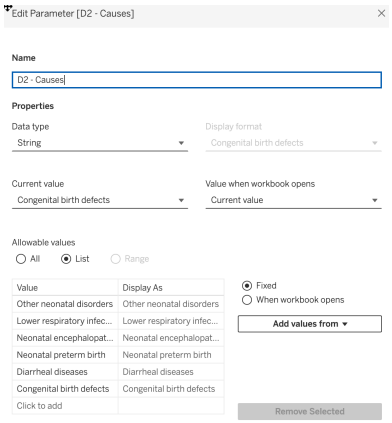
Following are the details about the dashboards and independent visualizations on each of these dashboards.

9.1 Parameters and Calculated Fields

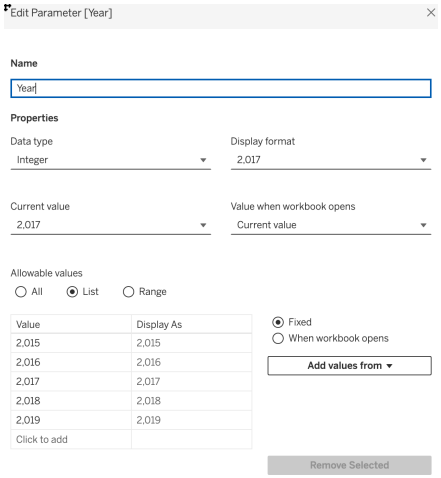
I created a set of parameters and Calculated fields to aid in my analysis and visual display of data. Following is a quick snapshot for the different calculated fields.

9.1.1 Parameters

- Causes – Parameter to filter on different top causes of infant mortality rates.



- Year – Listing of last 5 years from the data set



9.1.2 Calculated Fields

Total Deaths Filtered by Year parameter	Total Deaths (Yearly)
	`if [Year] = [Parameters].[Year]`
	`then [Total Deaths]`
	`end`

Total Deaths due to Top Causes	Total Deaths - Top Causes
	`SUM([Congenital birth defects])+`
	`SUM([Other neonatal disorders])+`
	`SUM([Neonatal encephalopathy])+`
	`SUM([Neonatal preterm birth])+`
	`SUM([Diarrheal diseases])+`
	`SUM([Lower respiratory infections])`

Total Deaths due to Other Causes	Total Deaths - Other Causes
	`SUM([Drowning])+`
	`sum([Whooping cough])+`
	`sum([Measles])+`
	`sum([Meningitis])+`
	`sum([Neonatal sepsis])+`
	`sum([Neoplasms])+`
	`sum([Tuberculosis])+`
	`sum([Syphilis])+`
	`sum([Hiv/Aids])+`
	`sum([Malaria])+`
	`sum([Nutritional deficiencies])`

Population Filtered by Year	Population
	`if [Year]= [Parameters].[Year]`
	`then [Population Overall]`
	`end`

Total Death count for all causes	SumCauses
	`IF [Causes]='Drowning' then SUM([Drowning])`
	`ELSEIF [Causes] = 'Malaria' then SUM([Malaria])`
	`ELSEIF [Causes] = 'HIV/AIDS' then SUM([Hiv/Aids])`
	`ELSEIF [Causes] = 'Meningitis' then SUM([Meningitis])`
	`ELSEIF [Causes] = 'Nutritional deficiencies' then SUM([Nutrition`
	`ELSEIF [Causes] = 'Other neonatal disorders' then SUM([Other neo`
	`ELSEIF [Causes] = 'Whooping cough' then SUM([Whooping cough])`
	`ELSEIF [Causes] = 'Lower respiratory infections' then SUM([Lower`
	`ELSEIF [Causes] = 'Congenital birth defects' then SUM([Congenita`
	`ELSEIF [Causes] = 'Measles' then SUM([Measles])`
	`ELSEIF [Causes] = 'Neonatal sepsis and other neonatal infections`
	`ELSEIF [Causes] = 'Neonatal encephalopathy due to birth asphyxia`
	`ELSEIF [Causes] = 'Tuberculosis' then SUM([Tuberculosis])`
	`ELSEIF [Causes] = 'Neonatal preterm birth' then SUM([Neonatal pr`
	`ELSEIF [Causes] = 'Diarrheal diseases' then SUM([Diarrheal disea`

Top Causes filtered by Year	D2 - Congenital birth defects
	`if [Year] = [Parameters].[Year]`
	`then [Congenital birth defects]`
	`end`

	D2 - Diarrheal Diseases
	`If [Year] = [Parameters].[Year] then [Diarrheal diseases] end`
	D2 - Neonatal Encephalopathy
	`if [Year] = [Parameters].[Year] then [Neonatal encephalopathy] end`
	D2 - Other Neonatal Disorders
	`if [Year] = [Parameters].[Year] then [Other neonatal disorders] end`
	D2 - Respiratory Infections
	`If [Year]= [Parameters].[Year] then [Lower respiratory infections] end`

Total Deaths counts filtered by Top Causes	D2 - Sum Causes
	`IF [D2 – Causes] = 'Other neonatal disorders' then SUM([D2 – Othe`
	`ELSEIF [D2 – Causes] = 'Lower respiratory infections' then SUM([D`
	`ELSEIF [D2 – Causes] = 'Congenital birth defects' then SUM([D2 –`
	`ELSEIF [D2 – Causes] = 'Neonatal encephalopathy due to birth asph`
	`ELSEIF [D2 – Causes] = 'Neonatal preterm birth' then SUM([D2 – Ne`
	`ELSEIF [D2 – Causes] = 'Diarrheal diseases' then SUM([D2 – Diarrh`
	`END`

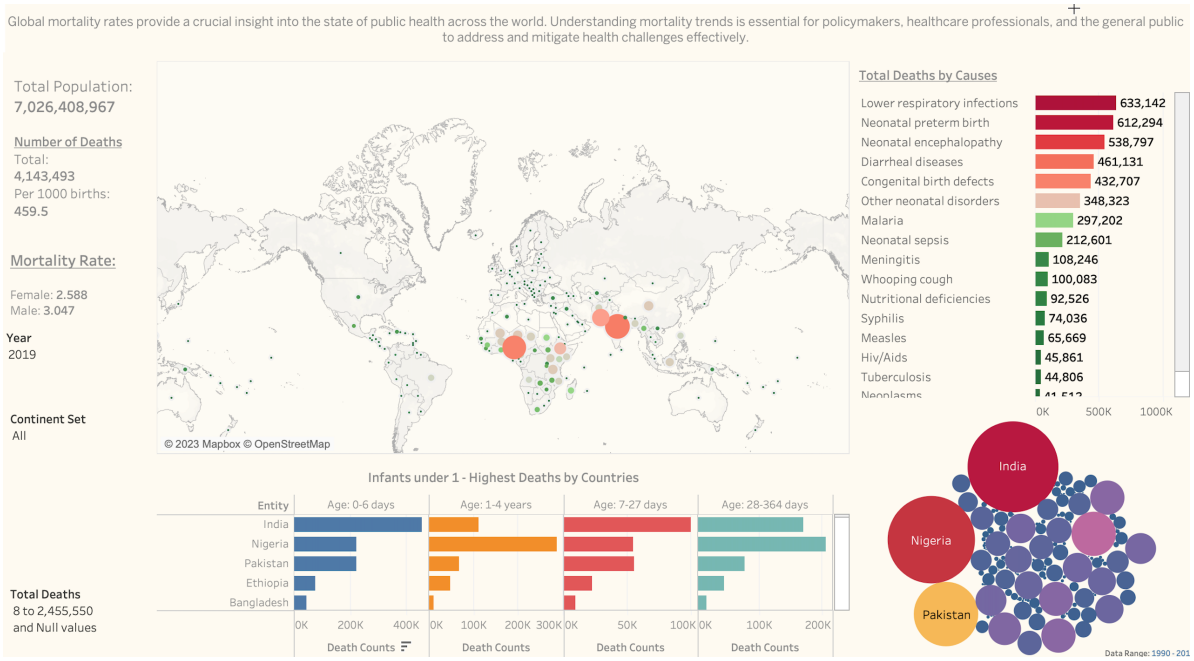
9.2 Dashboards

I will now deep dive into each dashboard and explain the individual visualizations created for each dashboard. I will also try to explain the thought process behind each visualization and how it can help us understand the data better.

9.2.1 Dashboard 1: Global Infant Mortality Statistics

This dashboard offers a comprehensive overview of infant mortality rates worldwide, providing a panoramic perspective. It illuminates the complex patterns and discrepancies that characterize the global scenario of infant mortality. With an intuitive interface, users can effortlessly explore fluctuations in rates and numbers across various years and regions. The visualizations go beyond mere graphics; they serve as insightful portals into the struggles and successes of nations confronting this vital public health issue.

Author: Eshita Gupta
SJSU ID: 016978741



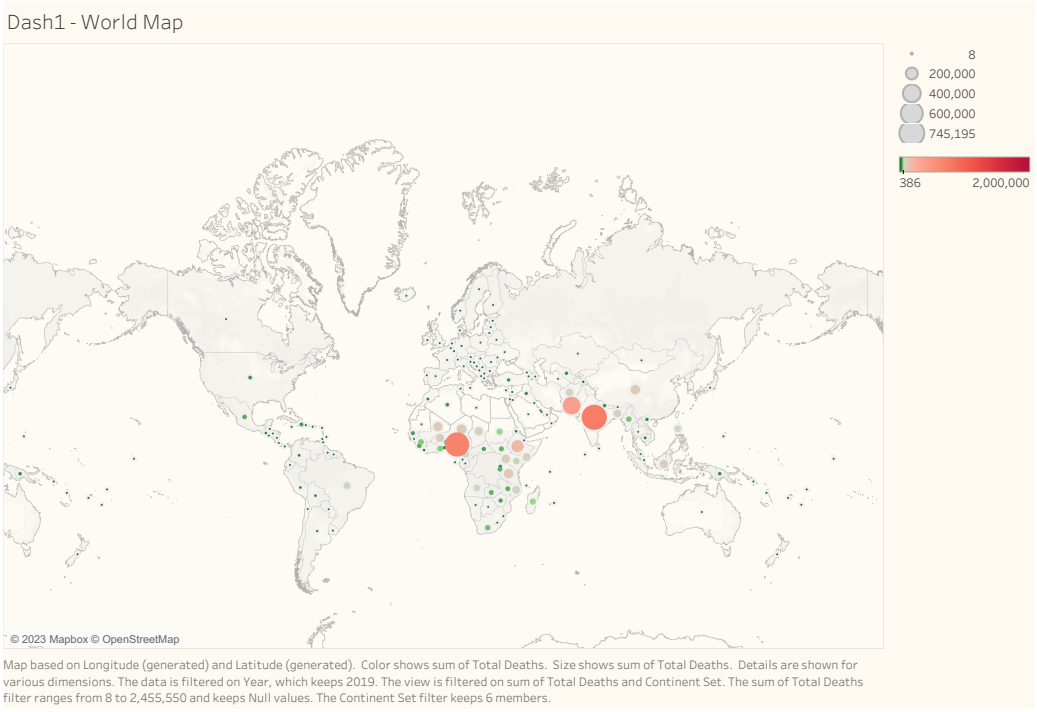
9.2.1.1 *Total Deaths*

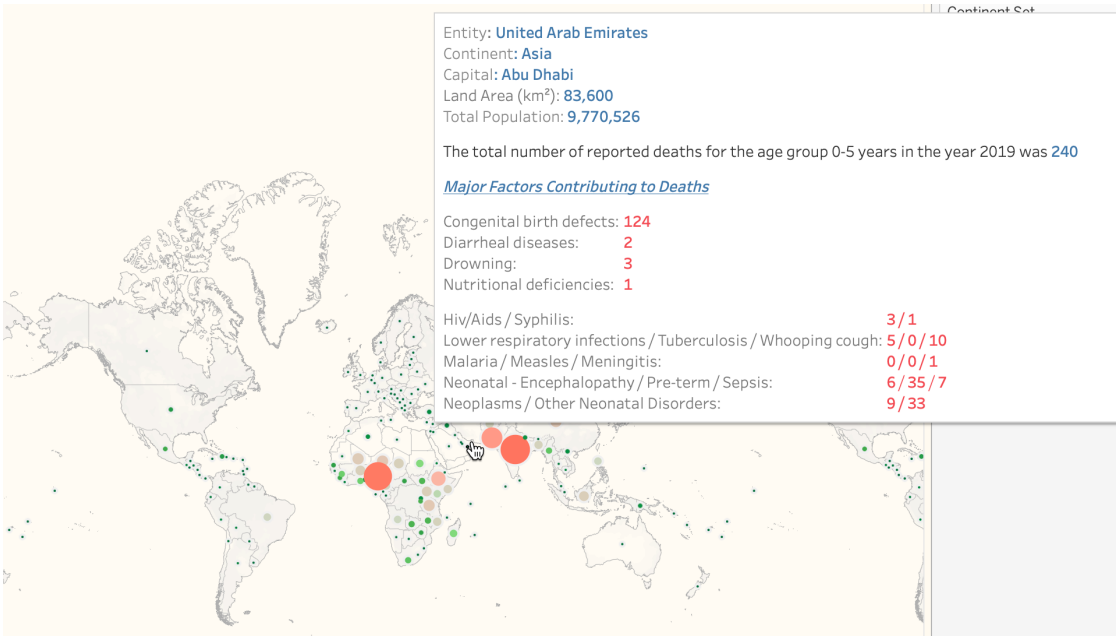
This visual provides a concise summary of population and mortality metrics.

- Total Population:
  The total population for the analyzed period is calculated by summing the individual populations across all relevant countries.

- Number of Deaths:
  Total: The total number of deaths recorded during the specified timeframe is represented by the sum of individual death counts.
  Per 1000 births: To provide a standardized measure, the number of deaths per 1,000 births is calculated by summing individual death rates and converting them to a per 1,000 births basis.

9.2.1.2 World Stats





This map provides a detailed insight into the infant mortality scenario for different countries in the specified year. It offers a geographic perspective on various critical indicators, emphasizing both demographic and health-related aspects.

- Entity: Name of the Country
- Continent: Region the country belongs to

- Capital: Capital for the country
- Land Area (km²): Total land area of the country
- Total Population: Total population of the country

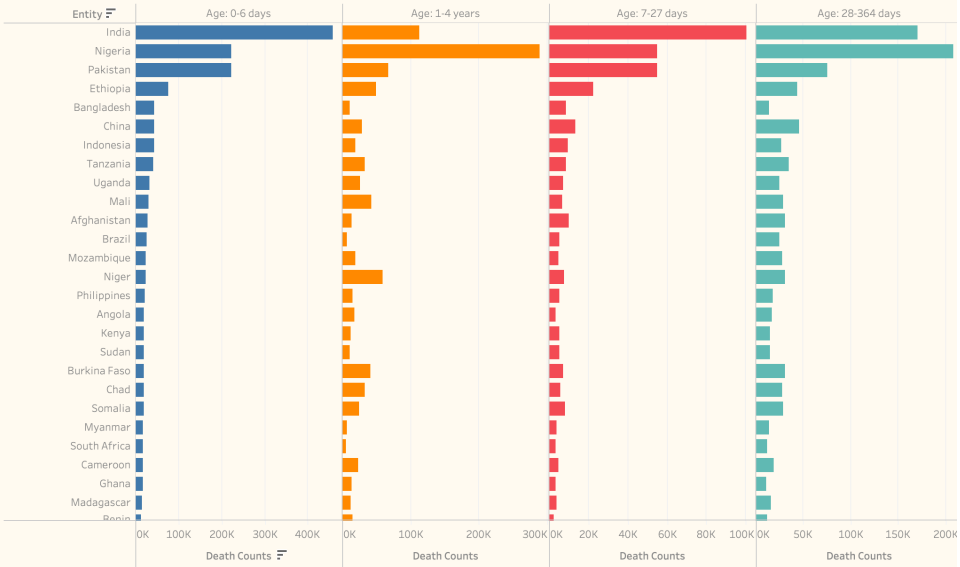
It also shows the total number of reported deaths within the age group of 0-5 years for different countries along with the contributing causes. The map visualizes the significant factors contributing to infant mortality, categorized as follows:

- **Congenital Birth Defects** - Anomalies present at birth, often resulting from genetic factors, environmental influences, or a combination of both.
- **Diarrheal Diseases** - Infections or conditions that lead to frequent, loose, or watery stools, often caused by bacteria, viruses, or parasites.
- **Drowning** - Submersion in liquid, resulting in respiratory impairment due to the intake of water.
- **Nutritional Deficiencies** - Insufficient intake of essential nutrients necessary for proper growth and development
- **HIV/AIDS** - A viral infection that weakens the immune system.
- **Syphilis** - A sexually transmitted infection caused by the bacterium Treponema pallidum.
- **Lower Respiratory Infections** - Infections affecting the lungs and airways.
- **Tuberculosis** - An infectious disease primarily affecting the lungs.
- **Whooping Cough** - A highly contagious respiratory infection
- **Malaria** - A mosquito-borne infectious disease
- **Measles** - A highly contagious viral infection
- **Meningitis** - Inflammation of the protective membranes covering the brain and spinal cord.
- **Neonatal Encephalopathy** - Brain dysfunction in the newborn period
- **Pre-term** - Birth occurring before 37 weeks of pregnancy completion.
- **Sepsis** - A severe infection affecting the bloodstream.
- **Neoplasms** - Abnormal growth of cells, often leading to tumors.
- **Other Neonatal Disorders** - Various health conditions affecting newborns.

Each region on the map is color-coded or symbolized to represent the severity or prevalence of these contributing factors. Darker shades or larger symbols may indicate higher mortality rates associated with specific causes.

This visualization aims to facilitate a quick understanding of the geographic distribution of infant mortality and its key contributing factors.

9.2.1.3 Countries with Highest Deaths



This visual encapsulates the distribution of death counts, providing an understanding of infant mortality patterns across different age ranges within each country. The data is segmented into four critical age brackets:

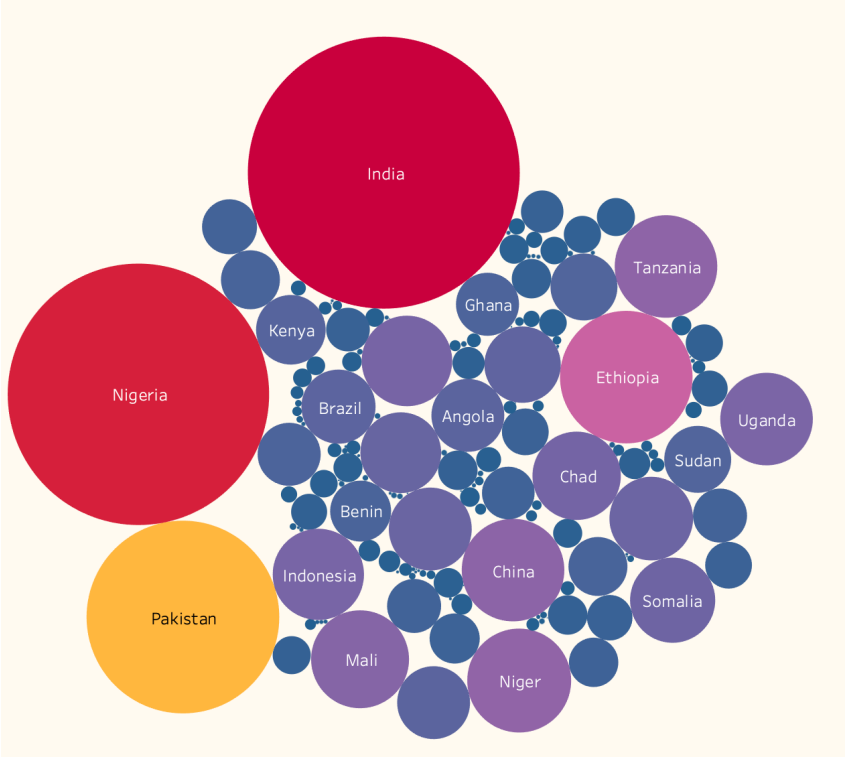
- **0-6 Days**: Reflects mortality within the first week of life.
- **7-27 Days**: Captures deaths occurring during the initial month.
- **28-364 Days**: Encompasses deaths within the first year, excluding the neonatal period.
- **1-4 Years**: Represents deaths in early childhood.

The visual enables a quick assessment of how death counts vary from country to country across the specified age ranges. Look for notable trends or outliers.

By categorizing deaths into distinct age ranges, it becomes possible to identify critical periods of vulnerability and tailor interventions accordingly.

Focus on the 0-6 days and 7-27-days categories for insights into neonatal mortality, a crucial aspect of infant health. Explore the 28-364 days and 1-4 years categories to understand mortality beyond the neonatal period, shedding light on health challenges faced during early childhood.

9.2.1.4 Country-wise Deaths Bubble



This visualization presents a comprehensive overview of infant mortality on a global scale, utilizing a bubble chart to illustrate total death counts by country. Each bubble represents a nation, with the size of the bubble indicating the magnitude of total deaths.

The size of each bubble correlates directly with the total death count for the corresponding country. Larger bubbles denote higher mortality rates, while smaller bubbles represent relatively lower death counts. The varying sizes of bubbles underscore the global disparities in infant mortality rates, emphasizing the need for targeted interventions in regions with higher mortality burdens.

9.2.1.5 Average Mortality Rates by Sex

Dash1 - Avg Mortality Rate

Mortality Rate:

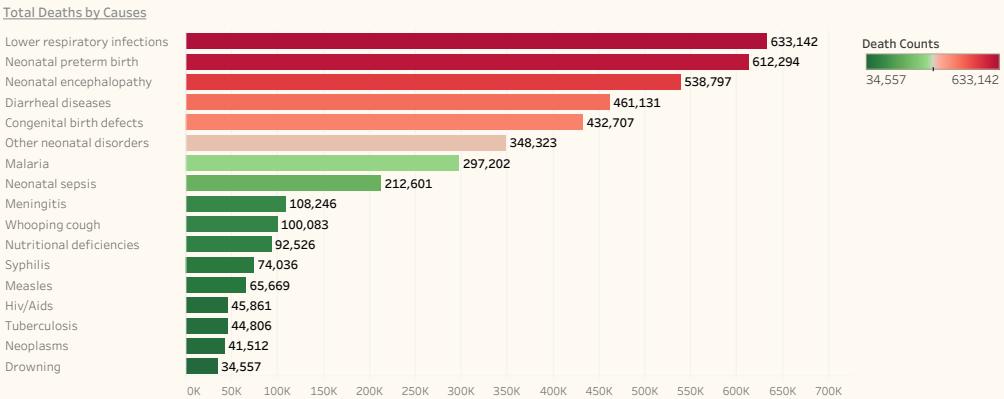
Female: 2.588
Male: 3.047

Average of Mortality rate, Female and average of Mortality rate, Male. The data is filtered on Year and Continent Set. The Year filter keeps 2019. The Continent Set filter keeps 6 members.

The mortality rates presented in this report are segmented by gender, providing insights into the average infant mortality rate for females and males.

These metrics offer a gender-specific perspective on mortality trends, aiding in the identification of potential gender-based disparities in health outcomes. The calculated averages provide a summary measure, allowing for a quick comparison of mortality rates between the two genders.

9.2.1.6 Cause-wise Death Counts



This bar chart offers a comprehensive insight into the leading causes of infant mortality, presenting a visual representation of the frequency of each cause. The chart is based on the following measures:

- Prevalence of Lower Respiratory Infections and Neonatal Preterm Birth: These two causes stand out with the highest frequency, indicating their significant impact on infant mortality.

- Neonatal Health Challenges: Causes such as Neonatal Encephalopathy, Other Neonatal Disorders, and Neonatal Sepsis highlight challenges during the early stages of life.
- Infectious Diseases: Diarrheal Diseases, Malaria, Meningitis, Whooping Cough, Syphilis, Measles, HIV/AIDS, and Tuberculosis collectively contribute to a substantial portion of infant mortality.
- Non-Communicable Causes: Nutritional Deficiencies, Neoplasms, and Drowning represent non-communicable factors influencing mortality.

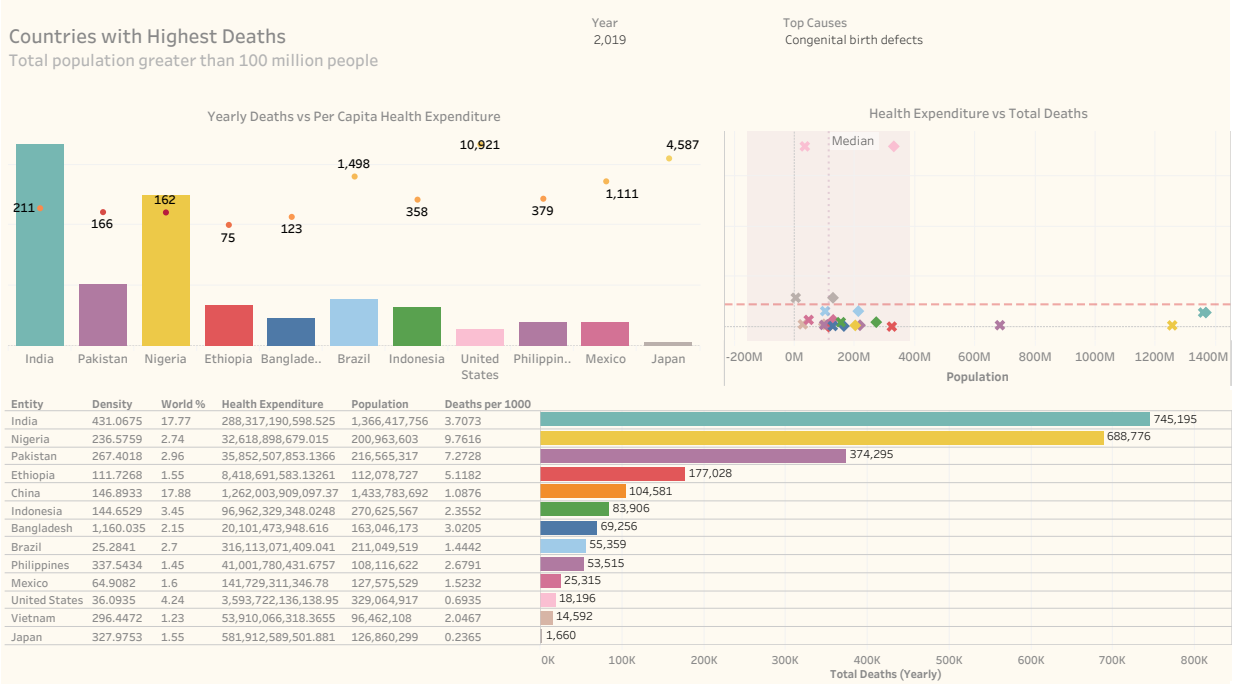
Based on the above chart I used the following 6 causes as the basis for my analysis into country wise causes and trends. I will explain about the dashboard on causes in the next section.

- Lower Respiratory Infections
- Neonatal Preterm Birth
- Neonatal Encephalopathy
- Other Neonatal Disorders
- Diarrheal Diseases
- Congenial Birth Defects.

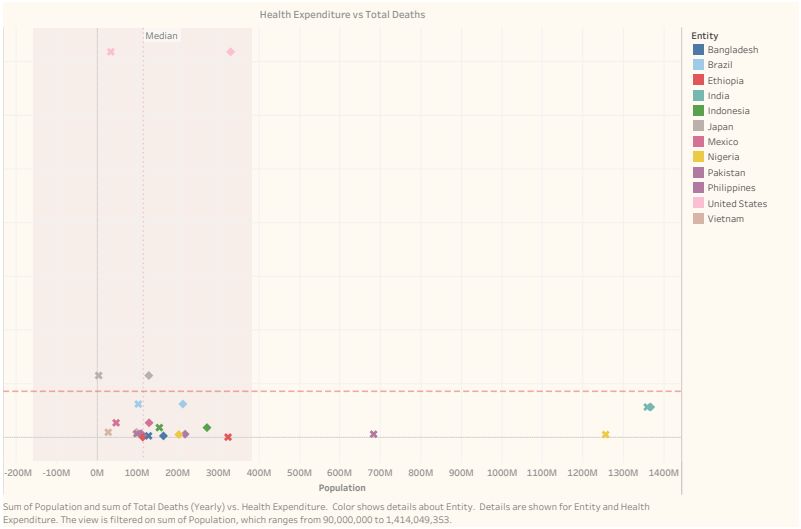
9.2.2 Dashboard 2: Exploring Causes and Correlations

In this dynamic dashboard, my attention has shifted towards a more profound exploration of the factors influencing infant mortality and the intricate correlations among various variables. By compiling data from a 5-year span (2015-2019) and mapping country-specific information against total death counts, I aimed to visualize the patterns that contribute to persistently higher mortality rates in certain nations.

The primary objective of this dashboard is to illuminate the impact of country-specific health expenditure on overall community health. Additionally, it delves into the top causes identified in the previous dashboard, which have consistently been key contributors to elevated mortality rates over the years. It is crucial to discern the countries with the highest death counts attributed to these causes and understand their relationship with factors such as population size, density, and overall health expenditure.



9.2.2.1 Healthcare Spend vs Mortality

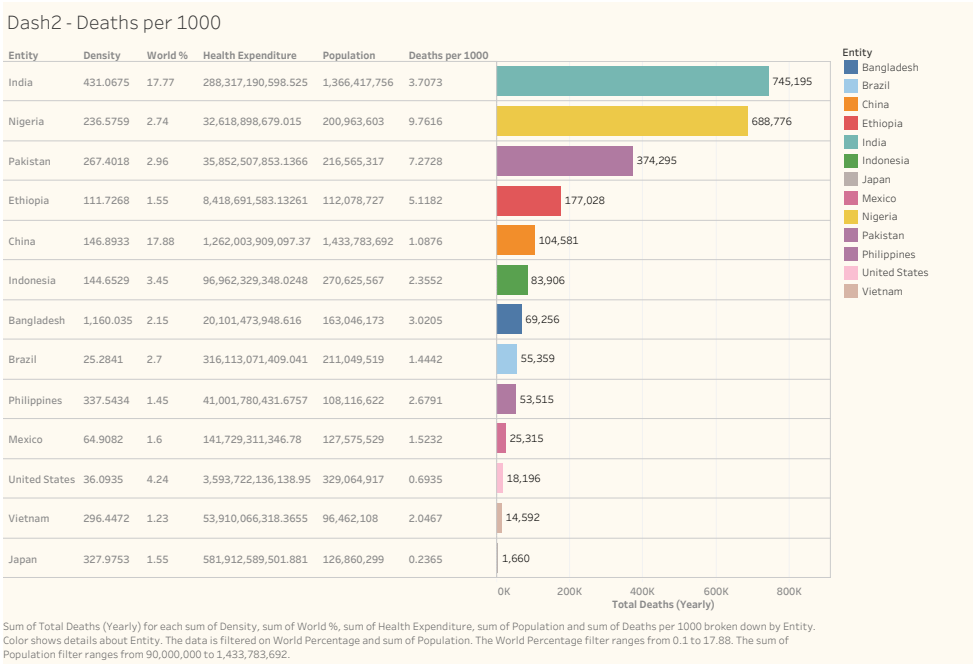


This scatter chart intricately weaves together the dimensions of healthcare spending and mortality, shedding light on the complex interplay between these critical factors. Focused on countries with the highest death counts, the chart examines three key statistics—Total Death Counts, Total Health Expenditure, and Population—to unravel insights into the relationship between financial investment in health and its impact on mortality rates.

- Each data point marked as X on the scatter chart represents a country, with the position along the vertical axis indicating the Total Death Counts attributed to various causes.
- The horizontal axis measures Total Health Expenditure, providing a comparative view of how much each country invests in healthcare.
- Population is also plotted on the vertical axis to give a dual representation, adding an additional layer to the analysis. This dimension aids in understanding the scale of healthcare challenges faced by populous nations.

As per the charts, it's evident that countries with the highest population and low per capita expenditure struggle to keep the mortality rates down. On the other hand, countries with extremely high expenditure rates do much better when it comes to managing the rates.

9.2.2.2 Country Specific Factors and Death Counts

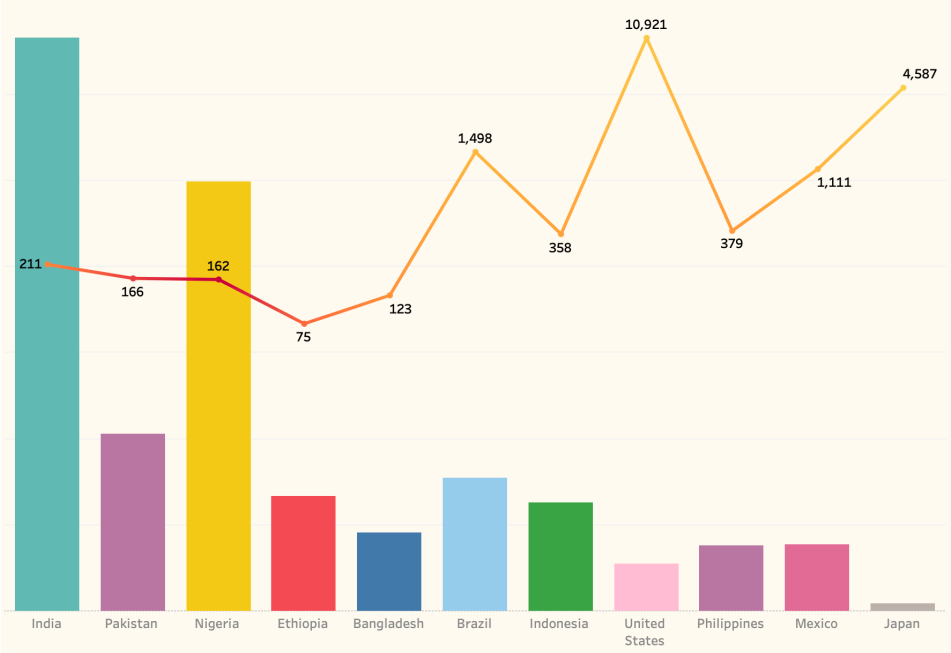


This bar chart aims to dissects the health profiles of countries by presenting key metrics in a visually impactful manner. Each country is represented with discrete data points for Density, Total Health Expenditure, Total Population, Deaths Per 1000 births, and Total Death Counts. The chart is designed to emphasize the magnitude of Total Death Counts, which is visually depicted through the height of the bars. Again, the data is specific for the last 5 years and aims to look at only the countries with the highest death counts.

Few important observations that can be made from this chart.

- Countries with high population size are generally always higher when it comes to the overall mortality rates.
- You can see a correlation to the density of the country to the total death counts. Example. India, Nigeria, and Pakistan all have very high densities along with high overall population counts. This can be one important factor to consider as the spread of infections is easier in countries with high population density and poor standard of living.
- Total health expenditure as outlined earlier plays a major part in the overall health of the country. It's evident from the chart that the countries with lower total health expenditure and higher population have significantly higher mortality rates.
- It is also interesting to note that country like China which is having a comparable population to a country like India, is not featured in the top 10. This can be attributed to higher per capita health expenditure and lower population density per sq km.

9.2.2.3 Top Causes vs Healthcare Spending



This perceptive bar chart carefully examines the intricacies of infant mortality dynamics, centering its attention on prominent causes across diverse countries. On the X-axis, individual countries are depicted, while the Y-axis quantifies the Total Death Counts associated with distinct causes—Lower Respiratory Infections, Neonatal Preterm Birth, Neonatal Encephalopathy, Other Neonatal Disorders, Diarrheal Diseases, and Congenital Birth Defects

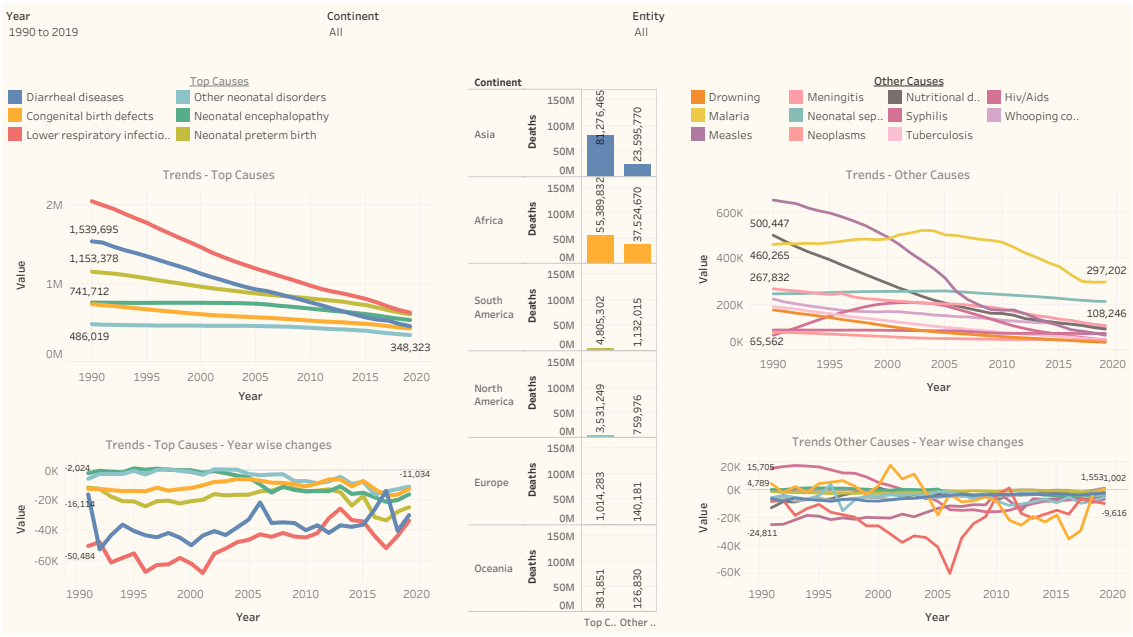
- The bars showcase the Total Death Counts, offering a visual comparison of the impact of each cause across different countries.
- The chart is strategically filtered to highlight the contribution of specific causes to mortality, aiding in a focused analysis.
- The trend line overlay provides an additional layer of insight, depicting the trend in per capita health expenditure for each country. This allows for a simultaneous examination of healthcare spending patterns.

Each set of bars corresponds to a specific cause, allowing for a targeted analysis of mortality dynamics related to Lower Respiratory Infections, Neonatal Preterm Birth, Neonatal Encephalopathy, Other Neonatal Disorders, Diarrheal Diseases, and Congenital Birth Defects. This also helps to identify the correlation between health expenditure and mortality rates using this chart.

9.2.3 Dashboard 3: Time Series Analysis

This dashboard focuses on the evolution of key metrics and trends over a specified timeframe, offering a comprehensive understanding of how infant mortality rates have changed over the years. The dashboard provides a granular view of yearly variations in death counts. Explores spikes, declines, and patterns, enabling a nuanced understanding of the temporal dynamics of infant mortality.

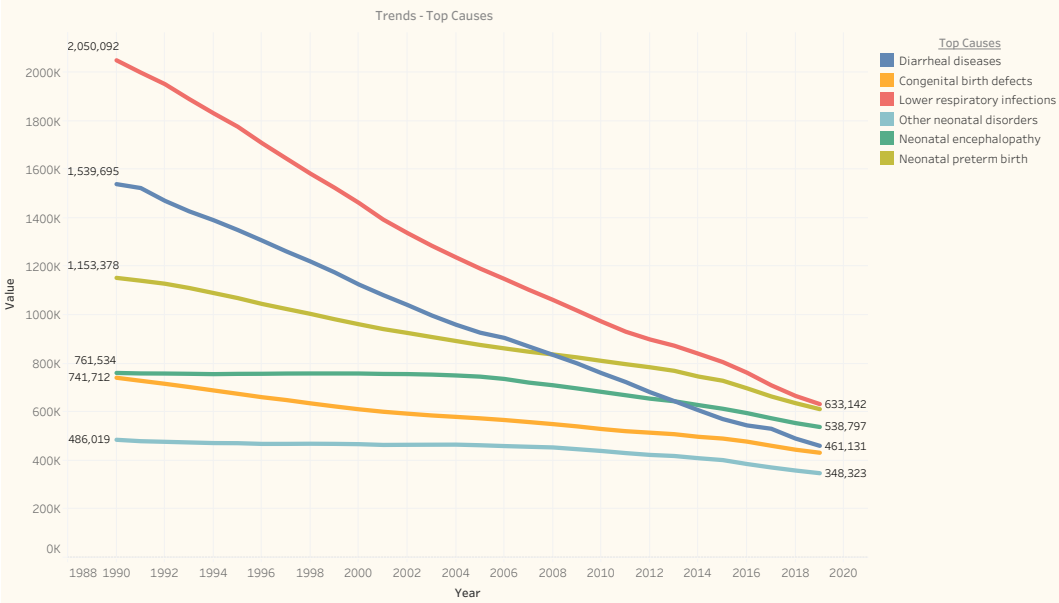
Gain region-specific insights by exploring how different regions have experienced shifts in infant mortality rates over the selected time period Identify regions that have made progress and those facing persistent challenges.



Left Side of the dashboard is focused on the trends for the Top causes as identifier with previous dashboards. Right side focusses on the other causes. Even though they don't contribute as significantly to the overall death counts, they are still statistically significant.

Middle chart focusses on region (continent) specific numbers for both top and other causes over the years.

9.2.3.1 Trends -– Top Causes



The trends of Congenital birth defects, Diarrheal diseases, Lower respiratory infections, Neonatal encephalopathy, Neonatal preterm birth and Other neonatal disorders for Year. Color shows details about Congenital birth defects, Diarrheal diseases, Lower respiratory infections, Neonatal encephalopathy, Neonatal preterm birth and Other neonatal disorders. The data is filtered on Entity and Continent. The Entity filter keeps 161 of 161 members. The Continent filter keeps 6 of 6 members. The view is filtered on Year, which ranges from 1990 to 2019.

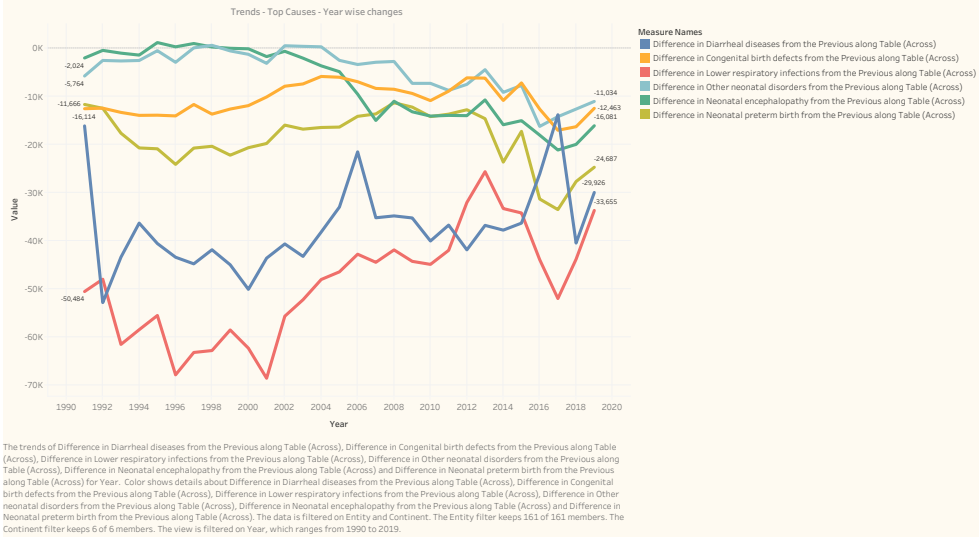
This line chart offers a chronological journey through the shifting landscape of infant mortality, focusing on six pivotal causes that have shaped the trajectory of global health. Spanning the years from 1990 to 2019, the chart provides a visual narrative of how the prevalence of specific causes has evolved over this critical period.

- Each line on the chart represents one of the six top causes of infant mortality: Lower Respiratory Infections, Neonatal Preterm Birth, Neonatal Encephalopathy, Other Neonatal Disorders, Diarrheal Diseases, and Congenital Birth Defects
- The horizontal axis captures the time dimension, spanning nearly three decades from 1990 to 2019. This timeline enables a comprehensive examination of trends and variations over the years.
- The vertical axis quantifies mortality rates, showcasing the prevalence of each cause per year. The Y-axis values provide insights into the relative impact of different causes on infant mortality rates.

This chart clearly shows that over the years the number of infant deaths has reduced significantly. This can be attributed to advancements in medical technology, advancements in field of Medicine and the overall focus on the health and well-being of the society.

Even though the numbers are significantly lower as compared to 1990, there is still a lot of work to do as there are certain specific regions that contribute to this.

9.2.3.2 Year on Year Difference – Top Causes

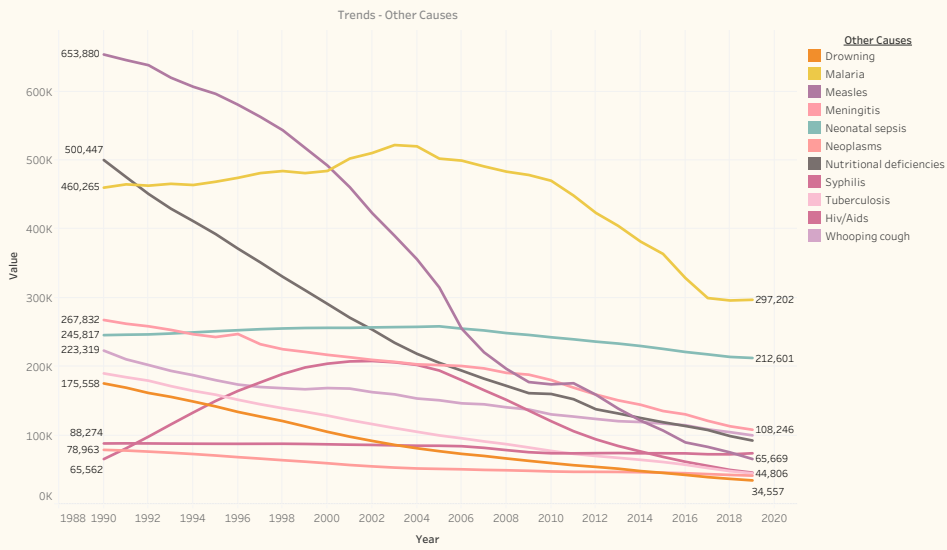


This line chart presents the annual variations in the prevalence of the six leading causes of infant mortality—Lower Respiratory Infections, Neonatal Preterm Birth, Neonatal Encephalopathy, Other Neonatal Disorders, Diarrheal Diseases, and Congenital Birth Defects—over the extensive period from 1990 to 2019.

- Instead of absolute total death counts, each line on this chart signifies the year-over-year change in the prevalence of a specific cause. It calculates the difference compared to the preceding year.
- The horizontal axis spans the entire timeline from 1990 to 2019, providing a comprehensive view of the yearly fluctuations in the impact of these causes.
- The vertical axis quantifies the annual change in prevalence, showcasing positive or negative values. A positive trend indicates an increase, while a negative trend suggests a decrease compared to the previous year.

Most of the top causes are on downward trend year to year. However, there is an increasing trend in years 2018-2020. This can be attributed to the fact that we had a global pandemic (COVID) and the overall condition of the healthcare system was in a bad shape. We can see a major spike in respiratory infections while minor spikes for the other top causes.

9.2.3.3 Trends -– Other Causes

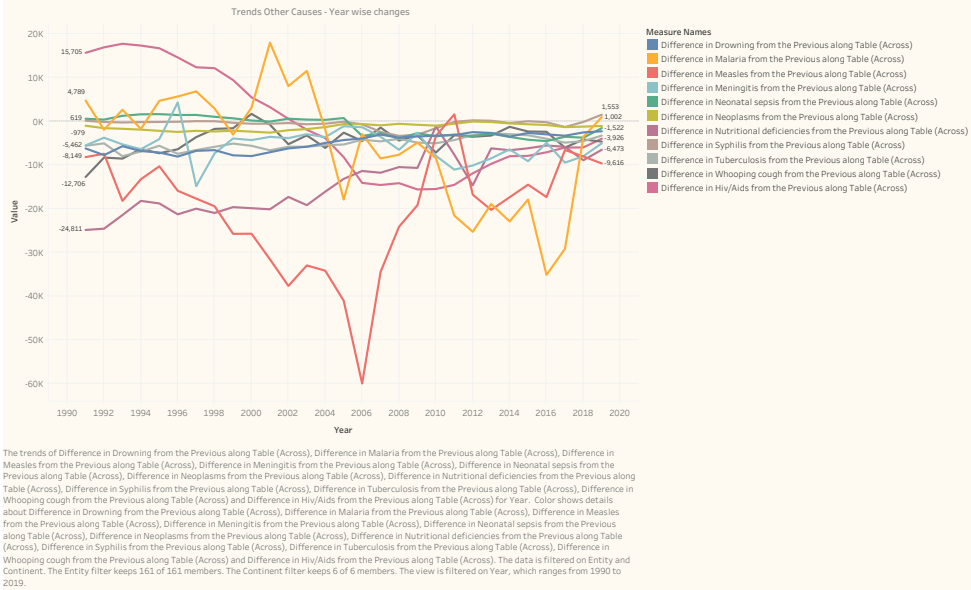


This line chart examines the yearly variations in causes beyond the identified top six. Covering the expansive period from 1990 to 2019, this chart sheds light on the evolving impact of secondary causes.

- Unlike the previous chart focusing on the top causes, this chart explores a broader spectrum of contributing factors to infant mortality. Each line represents the year-over-year change in prevalence for causes outside the primary six.
- The X-axis spans the timeline from 1990 to 2019, capturing the annual fluctuations in the prevalence of these additional causes.
- The Y-axis quantifies the annual change in prevalence, showcasing whether specific causes are experiencing an increase or decrease compared to the preceding year.

We see that the deaths due to these causes are also on a downward trend. However, there have been spikes that are noticed during specific periods. Example, there was a spike in deaths due to Malaria during the year 2003 and 2004, while there was a big spike during 1999-2004 for deaths due to HIV/AIDS. Malaria has been primary cause of concern when it comes to these additional causes, and it is mostly prevalent in countries with poor living conditions or countries with higher population densities.

9.2.3.4 Year on Year Difference – Other Causes

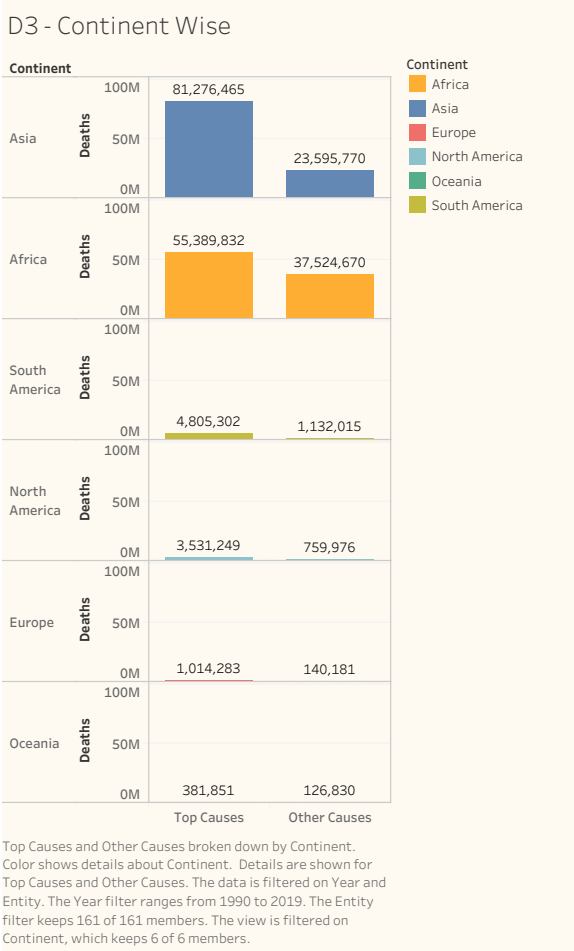


This line chart continues the exploration of the year-over-year differences in prevalence for causes beyond the primary six. Spanning the years from 1990 to 2019, this chart offers a nuanced perspective on how the impact of secondary causes has evolved over time.

- Like the previous chart, each line represents the year-over-year difference in prevalence for specific causes outside the primary six. It calculates how each cause's impact changes compared to the previous year.
- The X-axis encapsulates the entire timeline from 1990 to 2019, providing a comprehensive view of the annual fluctuations in the differences in prevalence for these additional causes.
- The Y-axis quantifies the yearly difference, showcasing positive or negative values. A positive trend indicates an increase in prevalence compared to the preceding year, while a negative trend suggests a decrease.

We see sharp spikes in Malaria and Measles, while there are equally significant drops as well. HIV/AIDS has been on a downward trend year by year. Overall, these trends show year by year decreasing trends, however, we again see significant rise in the year 2018-2020. It again could be due to the global pandemic; however, we don't have sufficient data to dive deeper into the same.

9.2.3.5 Regional Death counts vs Population



This bar chart provides a continent-level overview of infant mortality by categorizing total death counts into two segments: those attributed to the identified top causes and those stemming from other contributing factors. By breaking down the data in this manner, the chart offers insights into the distribution of infant deaths across different continents and the relative impact of primary and secondary causes.

- The chart categorizes total death counts for each continent into two distinct segments—deaths caused by the top identified causes and deaths resulting from other contributing factors.
- Each continent is represented by a set of bars, with different colors distinguishing between total death counts attributed to the top causes and those associated with other contributing factors.

When the complete data range (1990-2019) is selected, this chart showcases a harrowing metric. The total number of infants that have died in Asia and Africa alone is close to 200 million. This is a gut-wrenching number as so many innocent lives are lost due to different reasons and most of these deaths are centered around two regions.

We see that is a divide between Asia/Africa and other continents by a large margin when it comes to infant deaths. Majority of the developing/under-developed nations are in Asia/Africa region with poor living conditions and limited access to affordable health care. Add to that, the highest population densities, and lower per capita health expenditure, we see metrics that correlate directly to our findings.

While taking immediate steps is the need of the hour, however, there have been significant positive trends we have examined. Developing / Under-Developed nations grapple with population overflow and improving access to health care in such regions is a

mammoth task. However, steps are being taken by World Health Organization (WHO) and the governments of these specific regions to guide, educate and aid people from these regions.

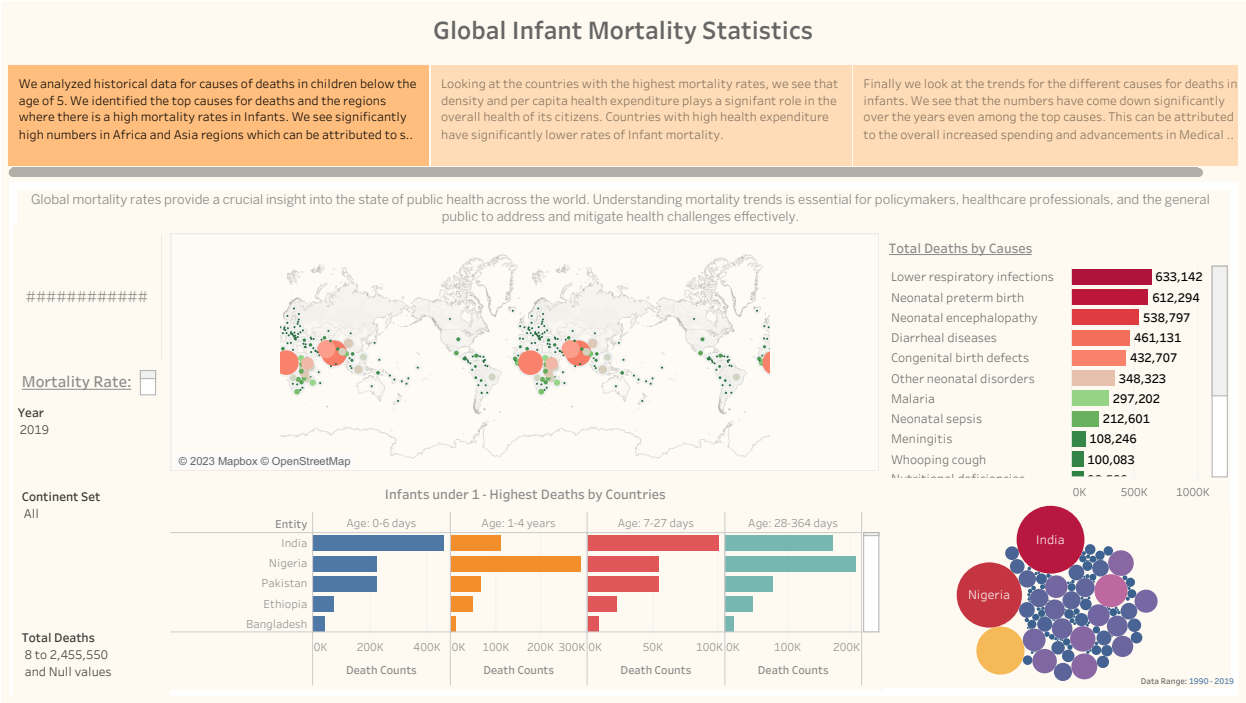
9.3 Final Story

This tableau story unfolds across three interconnected dashboards, each designed to provide a distinct perspective on the causes, correlations, and temporal dynamics shaping infant health globally.

Each dashboard is having a varying set of visualizations that focus on a combination of metrics. Following is a summary about each dashboard.

Dashboard 1: Global Infant Mortality Statistics

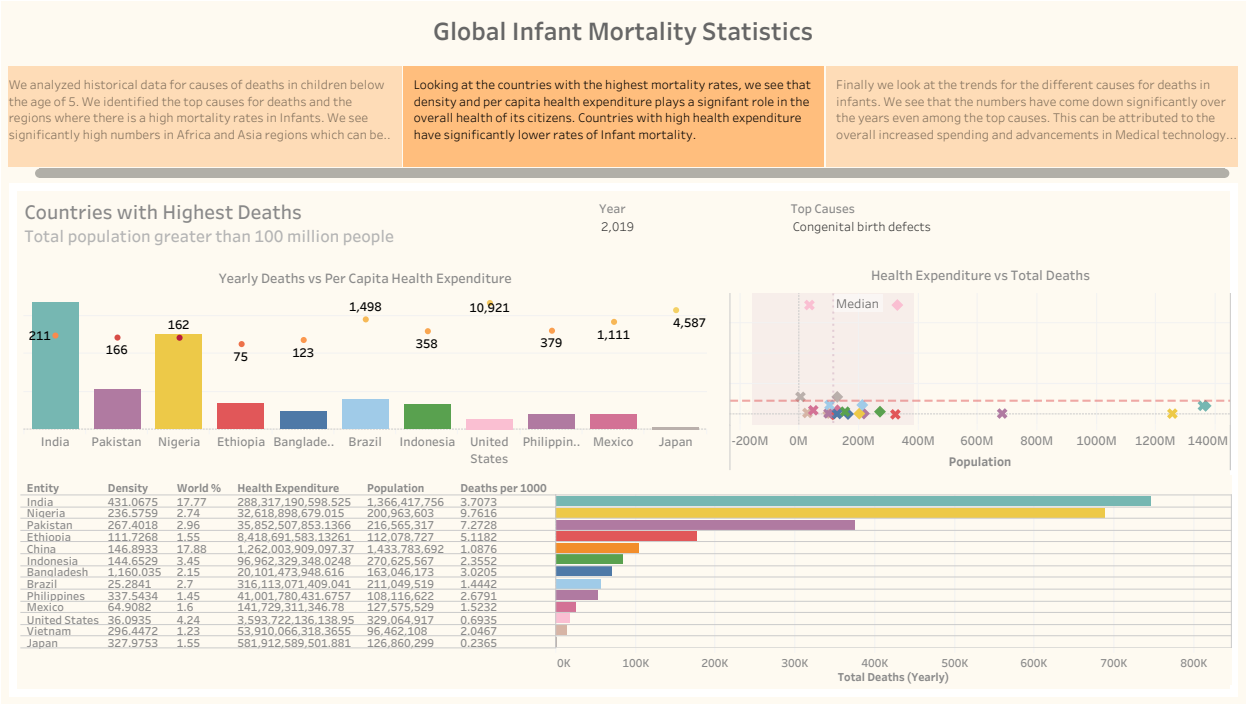
The first dashboard offers a panoramic view of infant mortality rates worldwide. Utilizing the cleaned data that was prepared, I presented a global map that vividly illustrates the regional distribution of the death counts. Through interactive features, users can explore how these rates have evolved over the years and gain insights into the relative disparities between nations. This also highlights the major causes for the deaths across different regions and provides insights into the countries with the highest death counts.



Dashboard 2: Exploring Causes and Correlations

Moving beyond the surface, my second dashboard dives into the top causes and correlations associated with infant mortality. Leveraging the cleaned data, I dissected the leading causes of infant deaths and their correlation with socio-economic factors, healthcare spending, population, density, and regional influences. This dashboard

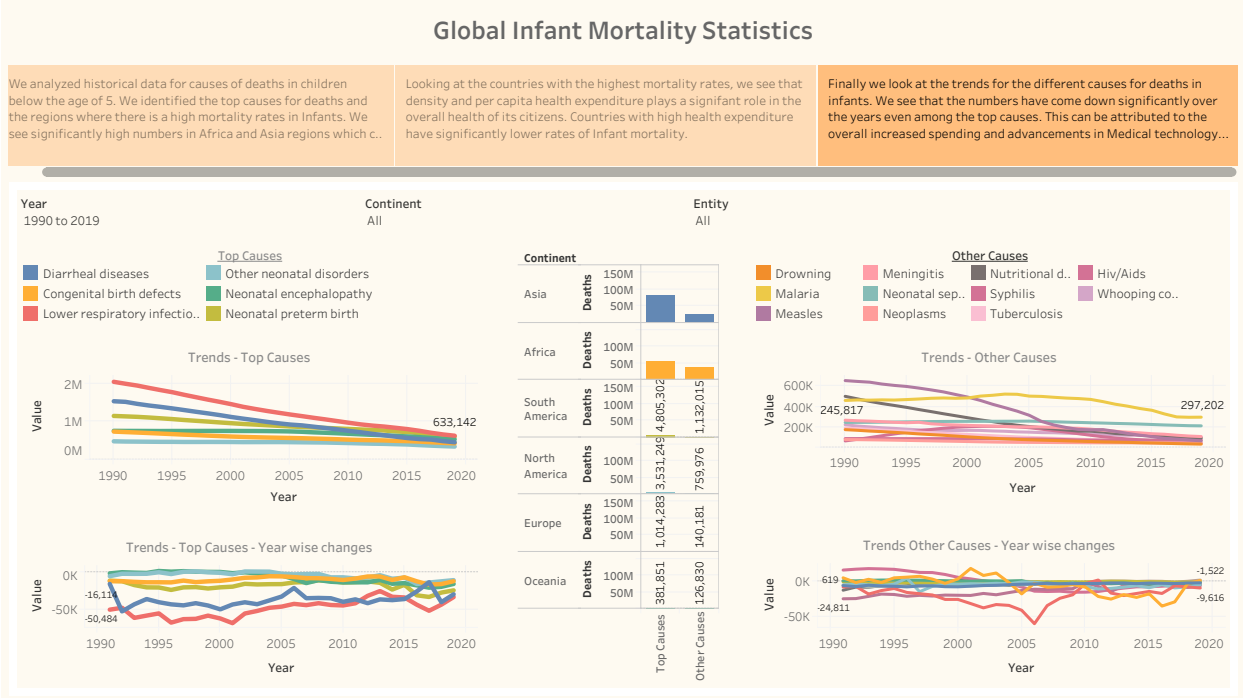
facilitates a deeper understanding of the dynamics that contribute to infant mortality, empowering stakeholders to identify targeted interventions.



Dashboard 3: Time Series Analysis

The third dashboard invites users on a journey through the last 5 years, unraveling the historical trajectory of infant mortality split by Top causes and other causes. By using the yearly data, I tried to show the difference in numbers year by year. I presented a dynamic time series analysis that captures how infant mortality rates have fluctuated globally and within specific regions. This dashboard serves as a valuable resource for policymakers, researchers, and healthcare professionals seeking to discern long-term trends and formulate evidence-based strategies.

Author: Eshita Gupta
SJSU ID: 016978741



10 Conclusion

My tableau story concludes with a profound understanding of infant mortality, unearthing critical insights that resonate with the global healthcare landscape. The journey through three interconnected dashboards has illuminated not only the challenges but also the opportunities for positive change.

As we explored the global statistics, we witnessed the nuanced patterns that define infant mortality rates, transcending borders and socio-economic boundaries. Moving to the exploration of causes and correlations, my analysis revealed a striking correlation between population density, lower per capita health expenditure, and higher mortality rates.

The temporal analysis across three decades offered a dynamic perspective, helping to identify trends, shifts, and emerging challenges in infant mortality causes. It is within this historical context that I found not just a story of challenges but also one of resilience and progress, where data-driven decisions can chart a course toward improved infant health outcomes.

In essence, this tableau story is not merely a collection of visualizations; it is a testament to the potential for positive change in global health. By understanding the co-relation between population, density, lower per capita health expenditure, and higher mortality rates, we are poised to shape interventions that are not only evidence-based but also targeted and effective.