Report:

“Causes of Deaths - Worldwide (1990-2019)”

Università di Catania

Alessandro Platania
Gaia Sandra Schillaci

Prof. O. Tomarchio

A.Y. 2021-2022

Index

1. Presentation of the dataset

The proposed report is based on a dataset containing information about the most common causes of deaths worldwide from 1990 to 2019. The source of our data is https://www.kaggle.com/datasets/saleh846/causes-of-deaths-worldwide?resource=download.

There are 5 .csv files that compose the dataset:

- **above-age-70.csv**: This file contains the causes of deaths of people aged above 70 years (limit included).
- **age-between-50-and-69.csv**: This file contains the causes of deaths of people aged between 50 and 69 years (limits included).
- **age-between-15-and-49.csv**: This file contains the causes of deaths of people aged between 15 and 49 years (limits included).
- **age-between-5-and-14.csv**: This file contains the causes of deaths of people aged between 5 and 14 years (limits included).
- **under-age-5.csv**: This file contains the causes of deaths of people aged below 5 years.

Each table shares three variables:

Name	Type	Description
Country	*string*	Name of the country
Country Code	*string*	Code that represents the country
Year	*int*	The period in which the data collected are referred to

Below, all the other variables contained in the tables are listed: they are all *int* variables and show the number of people dead because of the specified reason.

above-age-70.csv

Self-harm	Chronic respiratory diseases
Interpersonal violence	Alzheimer's disease and other dementias
Exposure to forces of nature	Cardiovascular diseases
Drowning	Nutritional deficiencies
Environmental heat and cold exposure	Drug use disorders

Diarrheal diseases	Alcohol use disorders
Road injuries	Lower respiratory infections
Tuberculosis	Diabetes mellitus
HIV/AIDS	Protein-energy malnutrition
Parkinson's disease	Acute hepatitis
Malaria	Cirrhosis and other chronic liver diseases
Fire, heat, and hot substances	Digestive diseases
Neoplasms	Chronic kidney disease

age-between-50-and-69.csv

Self-harm	Chronic respiratory diseases
Interpersonal violence	Alzheimer's disease and other dementias
Exposure to forces of nature	Cardiovascular diseases
Drowning	Nutritional deficiencies
Environmental heat and cold exposure	Drug use disorders
Diarrheal diseases	Alcohol use disorders
Road injuries	Lower respiratory infections
Tuberculosis	Diabetes mellitus
HIV/AIDS	Protein-energy malnutrition
Parkinson's disease	Acute hepatitis
Malaria	Cirrhosis and other chronic liver diseases
Fire, heat, and hot substances	Digestive diseases
Chronic kidney disease	Neoplasms

age-between-15-and-49.csv

Self-harm	Chronic respiratory diseases
Interpersonal violence	Alzheimer's disease and other dementias
Exposure to forces of nature	Cardiovascular diseases
Drowning	Nutritional deficiencies
Environmental heat and cold exposure	Drug use disorders
Diarrheal diseases	Alcohol use disorders
Road injuries	Lower respiratory infections
Tuberculosis	Diabetes mellitus
HIV/AIDS	Protein-energy malnutrition
Parkinson's disease	Acute hepatitis
Malaria	Cirrhosis and other chronic liver diseases
Fire, heat, and hot substances	Digestive diseases
Chronic kidney disease	Neoplasms

age-between-5-and-14.csv

Self-harm	Chronic kidney disease
Interpersonal violence	Cardiovascular diseases
Drowning	Lower respiratory infections
Malaria	Nutritional deficiencies
Fire, heat, and hot substances	Diabetes mellitus
Digestive diseases	Protein-energy malnutrition
Neoplasms	Exposure to forces of nature
Cirrhosis and other chronic liver diseases	Environmental heat and cold exposure
Chronic respiratory diseases	Diarrheal diseases
Acute hepatitis	Road injuries
HIV/AIDS	Tuberculosis

under-age-5.csv

Invasive Non-typhoidal Salmonella (iNTS)	Congenital birth defects
Interpersonal violence	Lower respiratory infections
Nutritional deficiencies	Neonatal preterm birth
Acute hepatitis	Environmental heat/cold exposure
Neoplasms	HIV/AIDS
Measles	Exposure to forces of nature
Digestive diseases	Diabetes mellitus
Cirrhosis and other chronic liver diseases	Drowning
Chronic kidney disease	Meningitis
Cardiovascular diseases	Other neonatal disorders
Road injuries	Whooping cough
Tuberculosis	Diarrheal diseases
Malaria	Fire, heat, and hot substances
Neonatal encephalopathy due to birth asphyxia	Syphilis
Neonatal sepsis and other neonatal infections	

2. Business questions

According to the data provided, our focus is on 2 groups through which we want to draw some insights: the age groups and the WHO regions (it refers to the division of the world provided by the World Health Organization for the purpose of reporting, analysis and administration).

1) **Considering the WHO regions**:
    a) Which is the most common cause of death?
    b) In which WHO regions do we observe the highest number of deaths?
    c) How did the number of deaths change from 1990 to 2019?
    d) Taking into account the time frame, can we draw some insights?

2) **Considering the age groups:**
    a) Which is the most common cause of death?
    b) In which age group do we observe the highest number of deaths?
    c) Where do we observe the highest number of deaths?
    d) Taking into account the time frame, can we draw some insights?

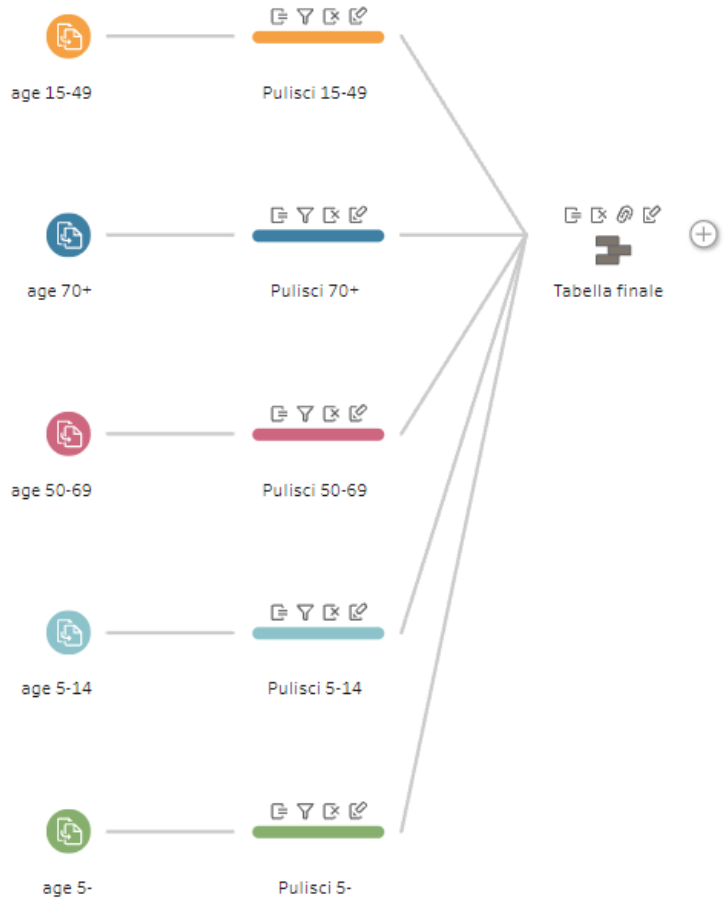
3) **Which is the general trend? Do we notice a decrease or an increase in deaths from 1990 to 2019?**

3. ETL

Before proceeding with the analysis, a preparation phase is mandatory to organize the dataset.

ETL is a process that allows us to do this through the extraction, integration and cleaning of data from operational sources to eventually feed the data warehouse and it consists in three phases: Extraction, Transformation, Loading.

The tool used is Tableau Prep Builder.



I. Cleaning

The original inputs are .csv files associated to a particular age group. Each of them contains the different causes of deaths for every country (and WHO region) between 1990 and 2019.

For each table:

- According to our business questions we removed some values in the variable *Country*: East Asia and Pacific (WB), Europe and Central Asia (WB), G20, Latin America and Caribbean (WB), Middle East and North Africa (WB), North America (WB), OECD countries, Sub-Saharan Africa (WB), South Asia (WB), World Bank High Income, World Bank Low Income, World Bank Lower Middle Income, World Bank Upper Middle Income.

- We renamed *Country* and *Country Code* in *Area* and *Area Code*, since these variables also include the data of the 6 WHO regions (so to have a more general reference name).

- Since the 6 WHO regions and some countries did not have values in *Area Code*, we created a calculated field in which we put the values of the original *Area Code* variable and added the codes missing. This new calculated field override the original variable.

```
IF [Area] = 'African Region (WHO)' THEN
(IFNULL([Area Code], 'AFR'))
ELSEIF [Area] = 'Region of the Americas (WHO)' THEN
(IFNULL([Area Code], 'AMR'))
ELSEIF [Area] = 'South-East Asia Region (WHO)' THEN
(IFNULL([Area Code], 'SEAR'))
ELSEIF [Area] = 'European Region (WHO)' THEN
(IFNULL([Area Code], 'EUR'))
ELSEIF [Area] = 'Eastern Mediterranean Region (WHO)'
THEN
(IFNULL([Area Code], 'EMR'))
ELSEIF [Area] = 'Western Pacific Region (WHO)' THEN
(IFNULL([Area Code], 'WPR'))
ELSEIF [Area] = 'England' THEN
(IFNULL([Area Code], 'ENG'))
ELSEIF [Area] = 'Scotland' THEN
(IFNULL([Area Code], 'SCO'))
ELSEIF [Area] = 'Wales' THEN
(IFNULL([Area Code], 'WAL'))
ELSEIF [Area] = 'Northern Ireland' THEN
(IFNULL([Area Code], 'NI'))
ELSE [Area Code]
END
```

Even though the *Area Code* variable is not strictly helpful for our analysis (we will not exploit it since our focus is more on the *Area* variable), we decided to maintain it with the purpose of giving more consistency to our dataset (actually we wanted to show how it's possible to align this field).

II. Union

After having performed the cleaning operations, we used the union to properly link all the tables.

- As supposed, there are some variables that are not shared between all the tables. Therefore, it yields to have some *null* values. We decided to transform these values in numeric values (= 0). The variables that presented this problem are: *Measles, Congenital birth defects, Neonatal sepsis and other neonatal infections, Neonatal encephalopathy due to birth asphyxia and trauma, Meningitis, Other mental disorders, Whooping cough, Syphilis, Self-harm, Parkinson's diseases, Chronic respiratory diseases, Alzheimer's disease and other dementias, Drug use disorders, Alcohol use disorders, Protein-energy malnutrition.*
  Below, it is reported just an example of how this change has been done for all the variables.

```
IFNULL([Chronic respiratory diseases],0)
```

- Unifying our tables, a variable called *Table Names* has been automatically created, containing as values the name of the original .csv files. We changed the variable name into *Age* and each value has been renamed according to the age group of reference (i.e. the value "age-between-50-and-69.csv" now is "50-69").

- In the final stage, we noticed that the dataset contained as *Area* "United Kingdom" but also "England", "Scotland", "Wales" and "Northern Ireland". Since we later realized that the default Tableau map works considering United Kingdom as default, we decided to cut the values associated to "England", "Scotland", "Wales" and "Northern Ireland" (even though in a previous step we fixed their *Area Code*, we decided to maintain here in the report the exact order in which we performed the ETL operations).

4. Dimensional Fact Model

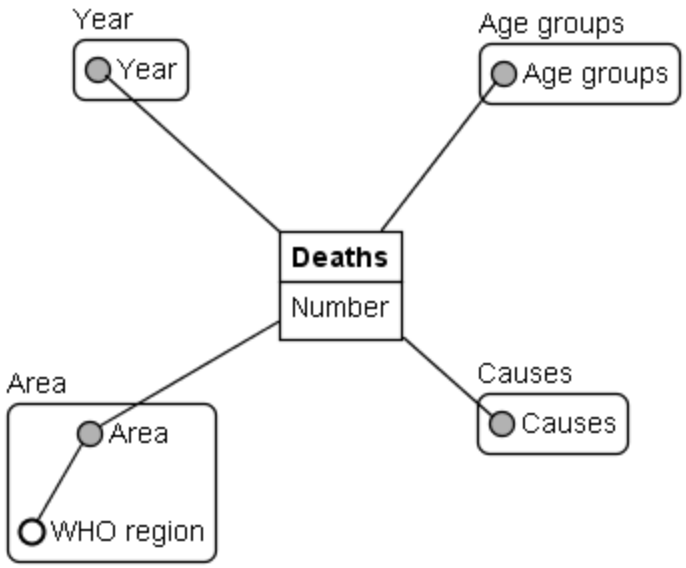
Published in 1998, the Dimensional Fact Model (DFM) by Golfarelli, Maio and Rizzi represents a graphical conceptual model for data warehouses. DFM is extremely intuitive and can be used by analysts and non-technical users as well.

The basic concepts are:

- Fact → is a concept relevant to decision-making processes.
- Measure → is a numerical property of a fact that describes a quantitative attribute that is relevant to analysis.
- Dimension → is a property, with a finite domain, that describes an analysis coordinate of the fact.
- Dimensional attribute → is a property, with a finite domain, of a dimension.
- Hierarchy → is a directed tree whose nodes are dimensional attributes and whose arcs model many-to-one associations between dimensional attribute pairs

In our dataset, the fact is represented by *Deaths*, while its measure is *Number*. The dimensions are represented by *Year*, *Area*, *Age groups* and *Causes*. For the *Area* dimension we recognize a hierarchy composed of it and the dimensional attribute *WHO region.*

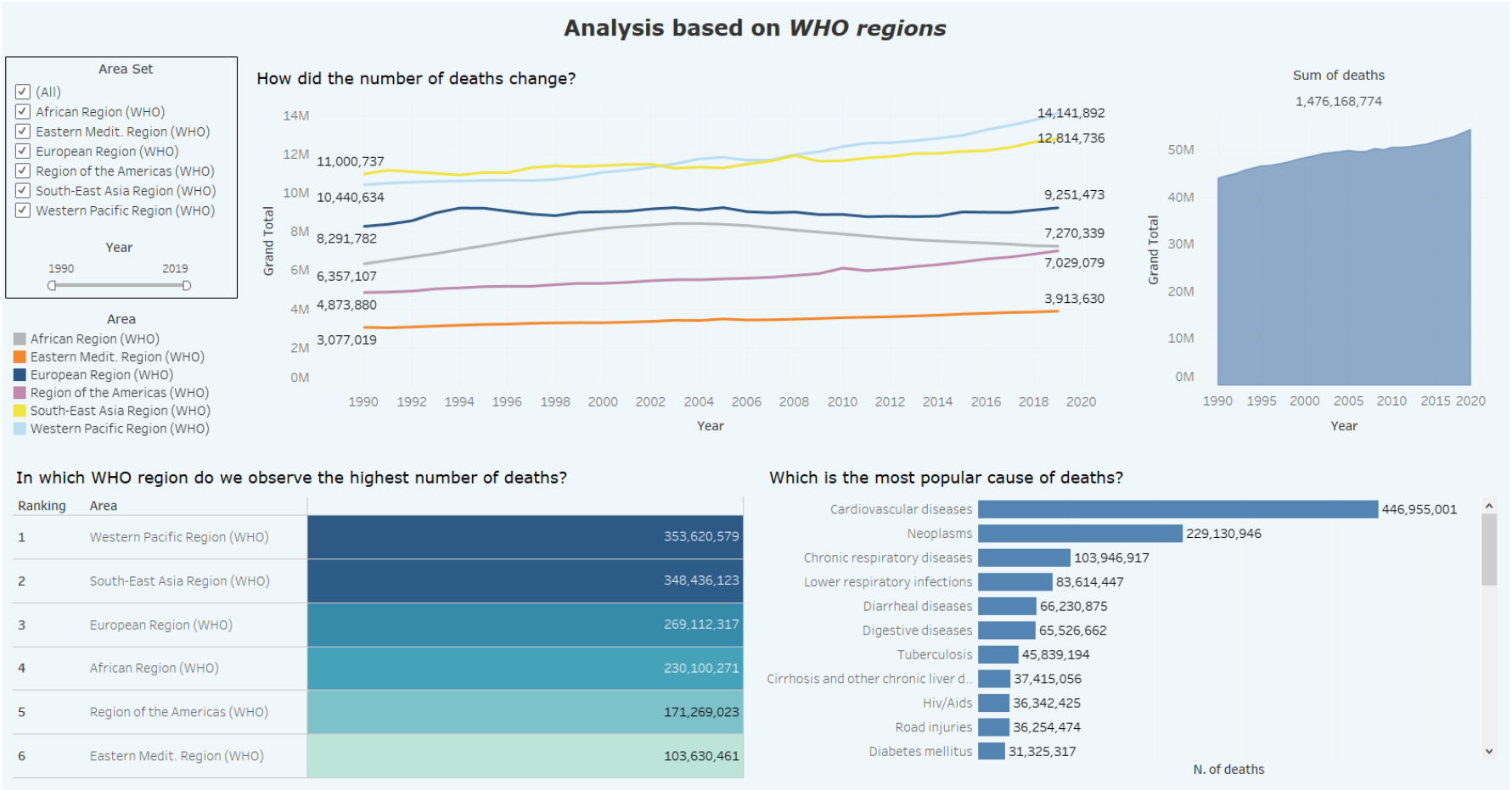
Area → *WHO region*



5. Dashboards

In order to answer the business questions, we realized two dashboards: the first one is based on the division of the world according to WHO (6 regions), while the second one refers to the age groups (5 age group). The main colour is blue since it is a good choice for colour blind people.

I. Dashboard – WHO regions



This dashboard aims to analyse the trend of the causes of deaths according to the WHO regions. There are five sheets: *N. of Deaths per Region, Causes of deaths, Changes of deaths throughout the years, Sum of deaths, Sum of deaths per year*.

The box in the top left corner is the tool through which the dashboard is managed. We can select the regions that we want to analyse (one or more). Moreover, there is a slider that allows to filter the analysis according to some specific years.

Single sheets in detail

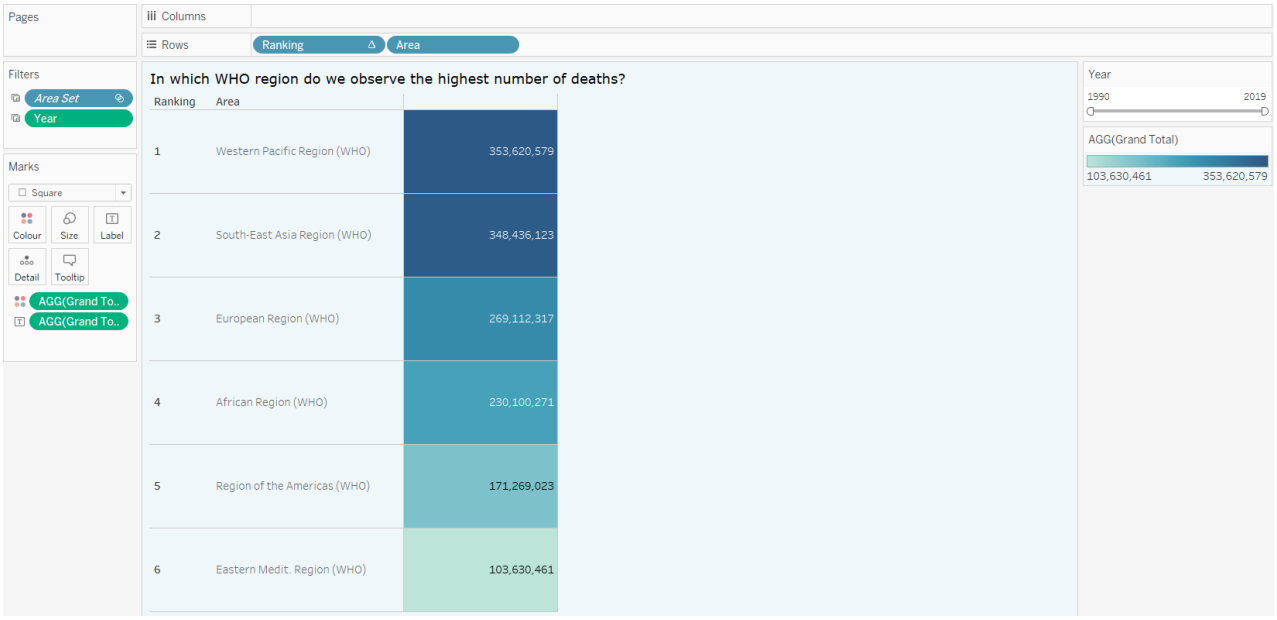
- **N. of deaths per Region**

The *N. of deaths per Region* sheet shows each region that have been selected by the filter with the associated number of deaths. The regions are sorted considering this value and the colour applied (blue) changes its intensity through the full range, so that the darker the colour, the higher the number of deaths is.

The measure used for this chart is obtained by setting the calculated field "Grand Total":

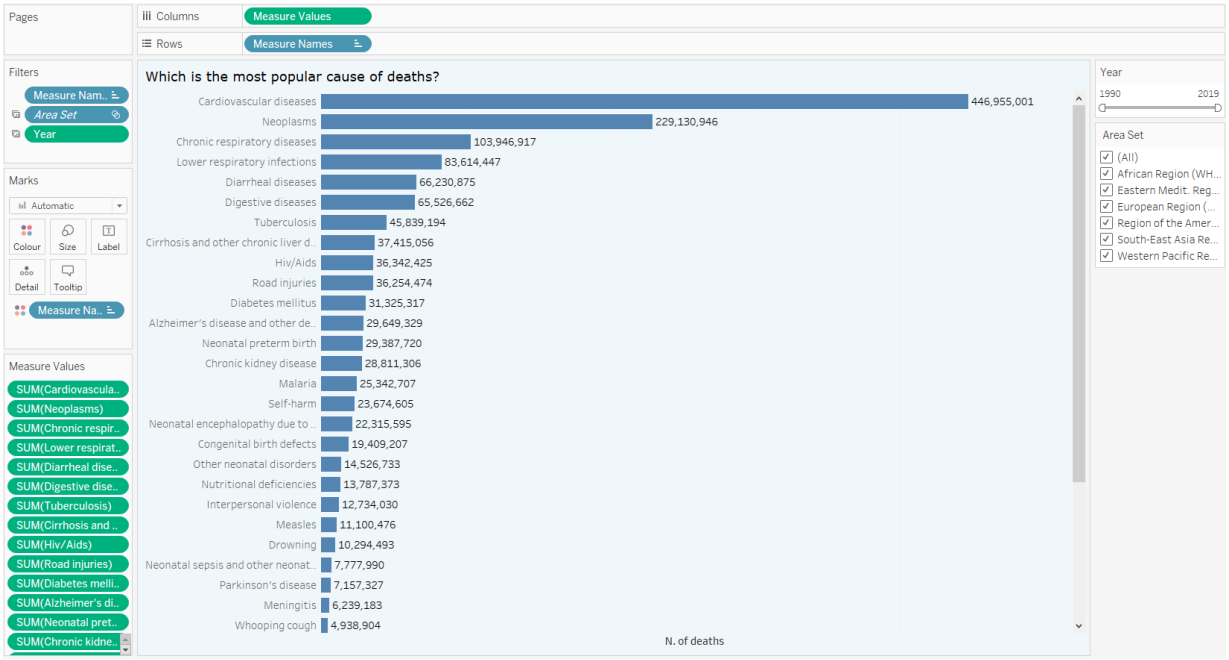
```
SUM([Acute hepatitis])+SUM([Alcohol use disorders])+SUM([Alzheimer's disease and other dementias])+SUM([Cardiovascular diseases])+SUM([Chronic kidney disease])+SUM([Chronic respiratory diseases])+SUM([Cirrhosis and other chronic liver diseases])+SUM([Congenital birth defects])+SUM([Diabetes mellitus])+SUM([Diarrheal diseases])+SUM([Digestive diseases])+SUM([Drowning])+SUM([Drug use disorders])+SUM([Environmental heat and cold exposure])+SUM([Exposure to forces of nature])+SUM([Fire, heat, and hot substances])+SUM([Hiv/Aids])+SUM([Interpersonal violence])+SUM([Invasive Non-typhoidal Salmonella (iNTS)])+SUM([Lower respiratory infections])+SUM([Malaria])+SUM([Measles])+SUM([Meningitis])+SUM([Neonatal encephalopathy due to birth asphyxia and trauma])+SUM([Neonatal preterm birth])+SUM([Neonatal sepsis and other neonatal infections])+SUM([Neoplasms])+SUM([Nutritional deficiencies])+SUM([Other neonatal disorders])+SUM([Parkinson's disease])+SUM([Protein-energy malnutrition])+SUM([Road injuries])+SUM([Self-harm])+SUM([Syphilis])+SUM([Tuberculosis])+SUM([Whooping cough])
```

In this way, we found the global total of deaths and we can filter this result for each region.



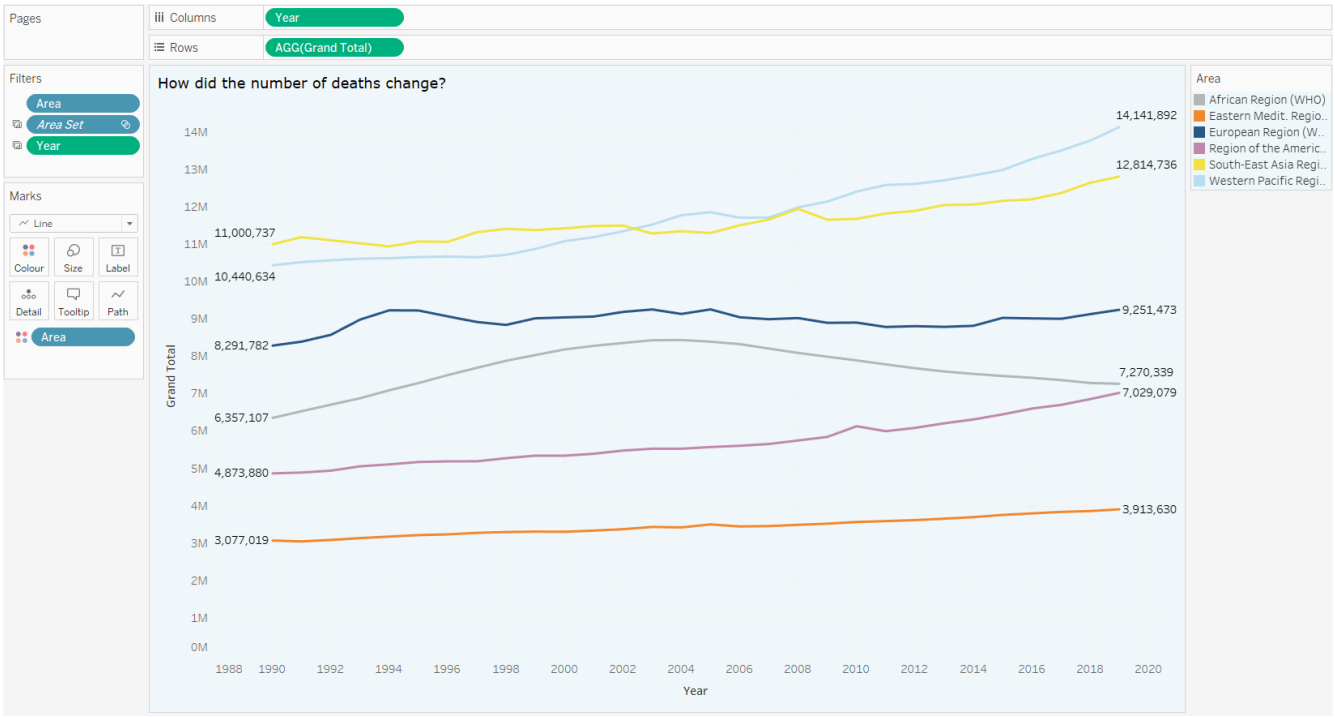
- **Causes of deaths**

The *Causes of deaths* sheet is composed of a bar chart that shows the causes of deaths from the most popular to the least. We used just a single tone of blue since we found it better to visualize the available data. Moreover, to highlight just the six WHO regions, we created a set called *Area Set* that is composed just of these regions (otherwise if we had used *Area* as filter, we would have seen all the countries and the WHO regions as possible choices in the filter). The option to show the mark labels is ticked.



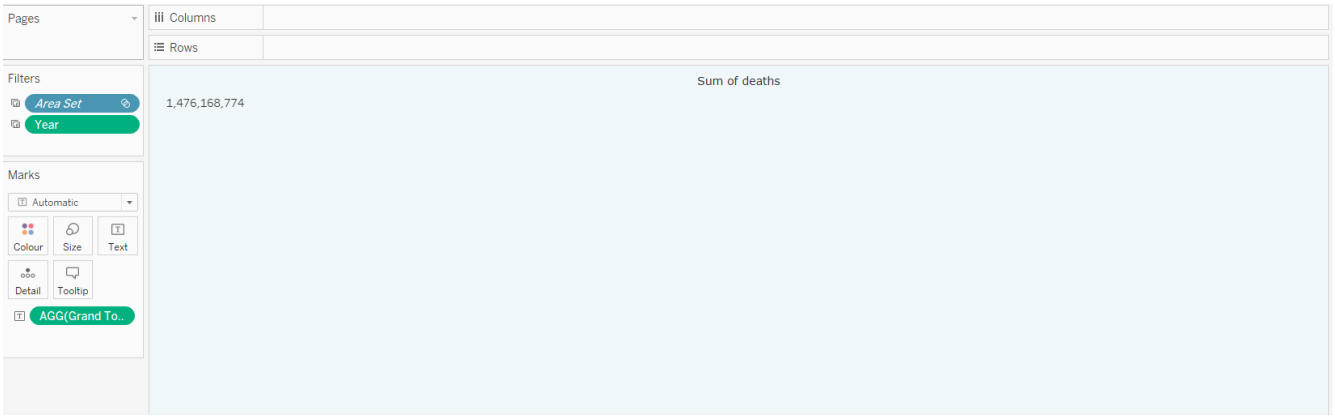
- **Changes of deaths throughout the years**

The *Changes of deaths throughout the years* sheet shows how the number of deaths changed over time. Each colour is associated with a different region. We opted for a line chart since our goal was to focus the attention on the changes over the period. As we did before, we used the Area Set to highlight just the WHO regions.



- **Sum of deaths**

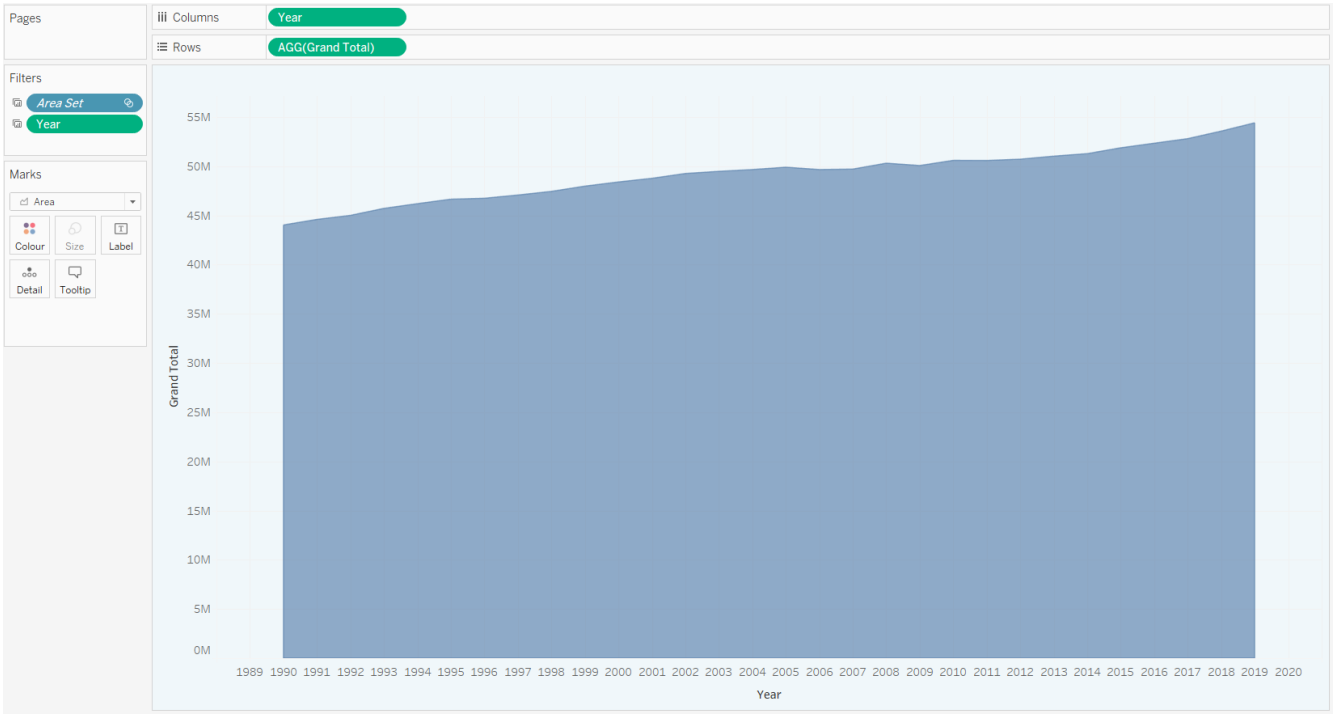
This sheet just shows the global sum of deaths, obviously according to the values set in the filters (area and period).



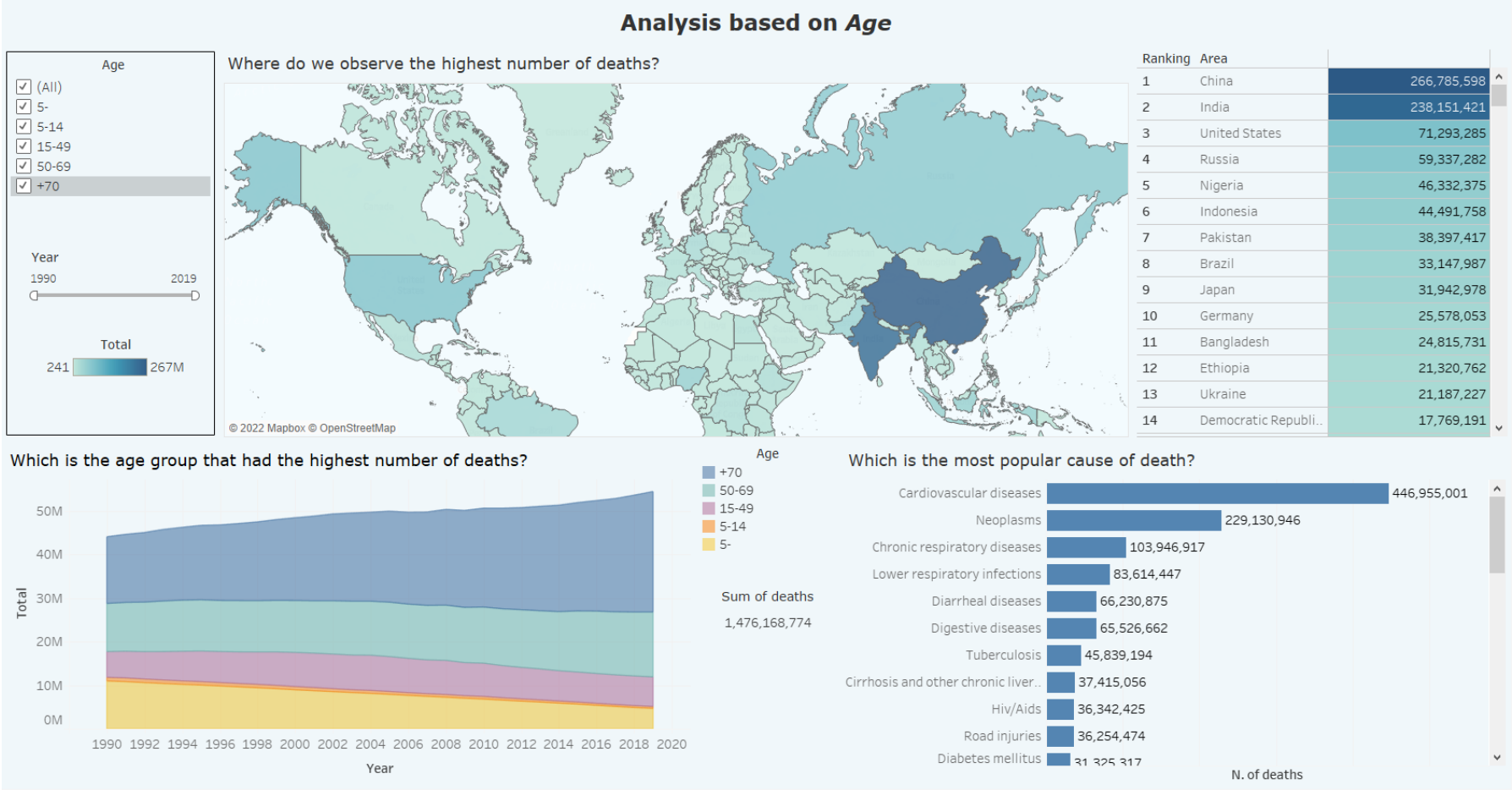
- **Sum of deaths per year**

This sheet shows the number of deaths per year, according to the values of the filters. Differently from the previous sheets, here the aim is to underline the changes over the period of the sum of the number of deaths of the regions considered.

Hovering the pointer in a part of the graph would lead to see the sum of deaths of the WHO regions marked in a specific year.



II. Dashboard – Age



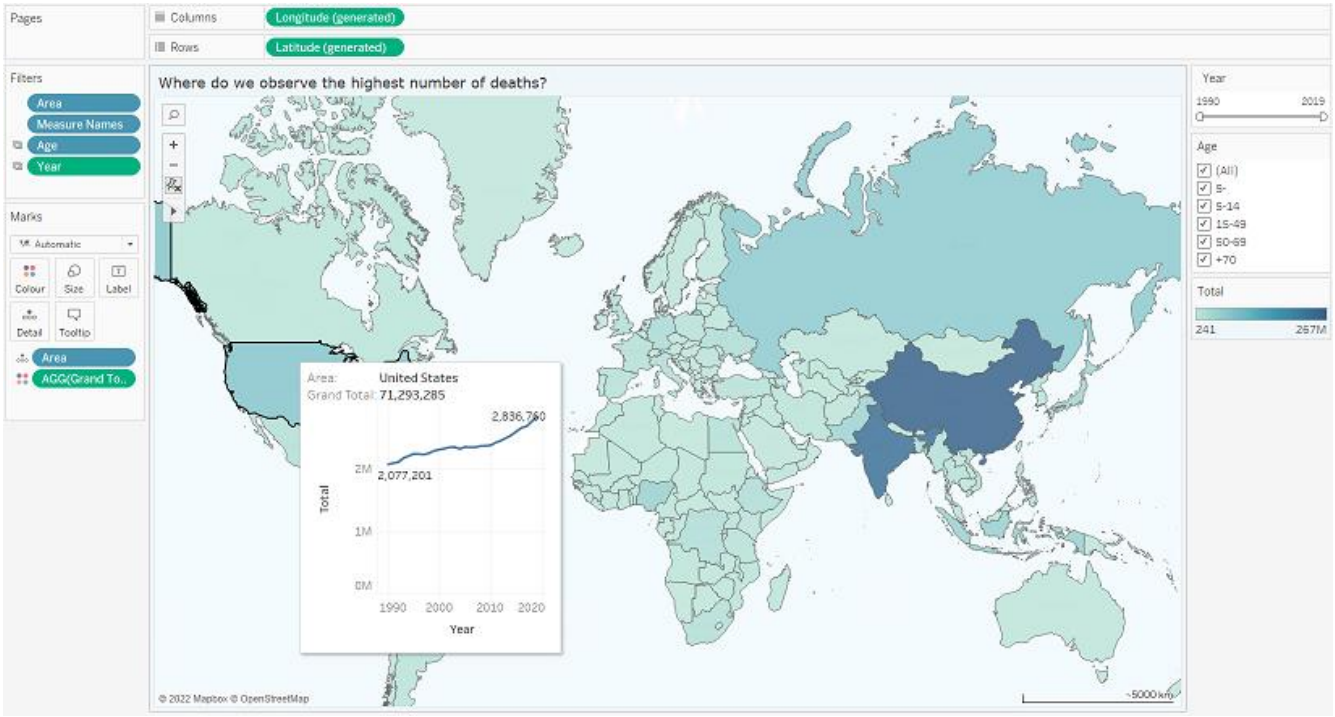
This dashboard aims to analyse the trend of the causes of deaths according to the age groups. There are six sheets: *Map of Deaths, Causes of deaths per age, N. of Deaths per Age area chart, N. of deaths Rank country, N. of deaths line chart country tooltip, Sum of Deaths per age*.

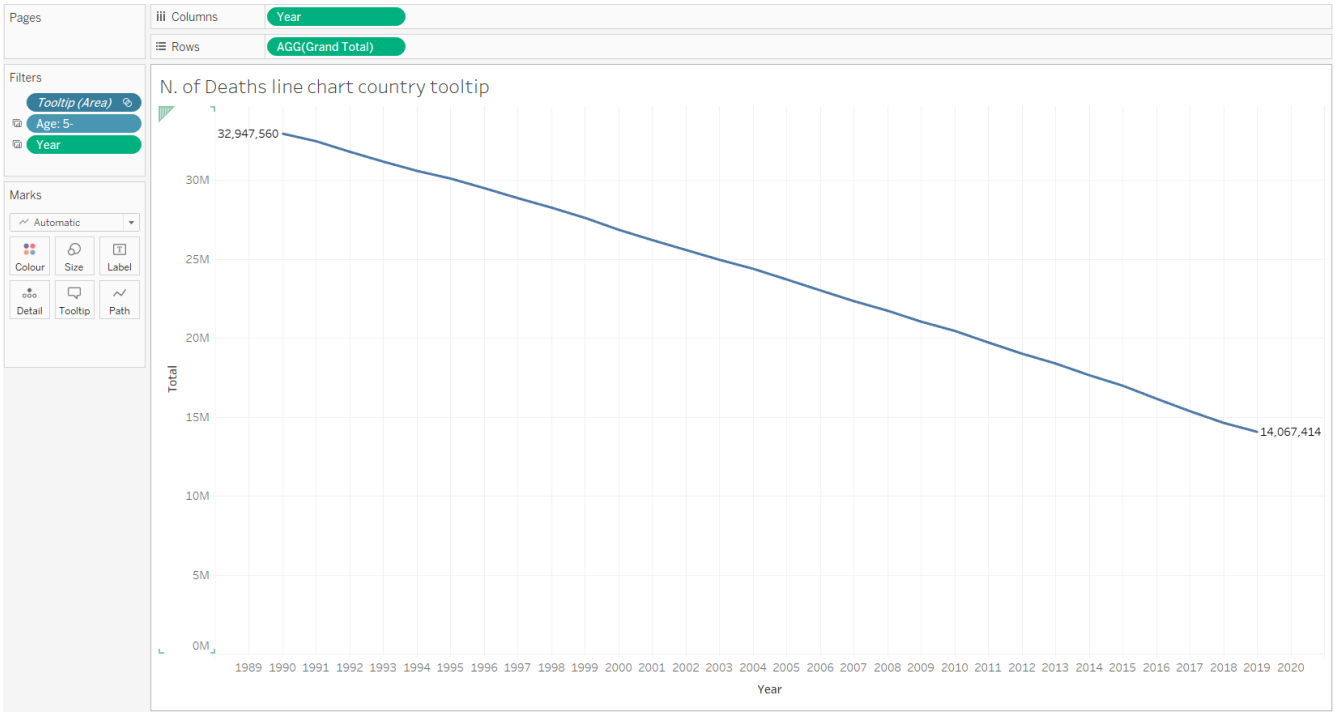
The box in the top left corner is the tool through which the dashboard is managed. We can select the age groups that we want to analyse (one or more). Moreover, there is a slider that allows to filter the analysis according to some specific years.

Single sheets in detail

- **Map of deaths + N. of deaths line chart country tooltip**

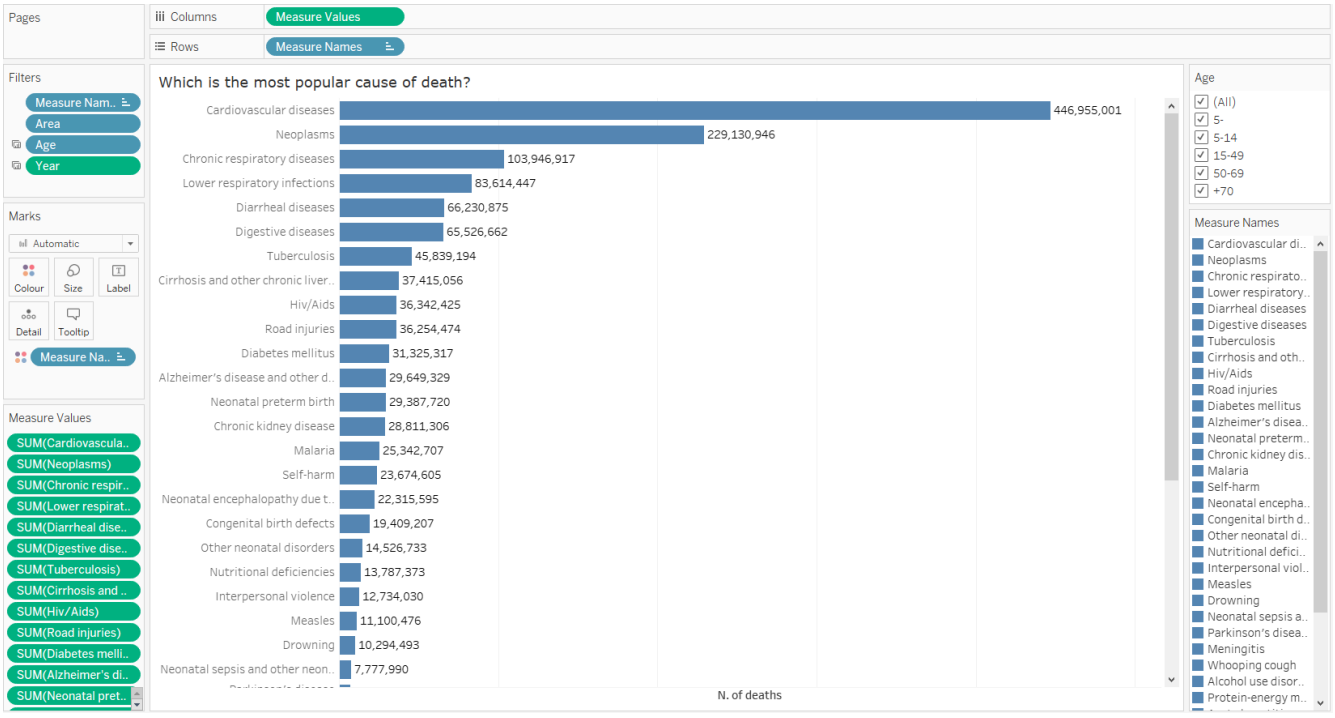
This sheet is based on a geographical map in which, according to the filters, we can observe how the number of deaths for each country changes. Moreover, if we move the pointer on a country, a specific tooltip containing some interesting information is available: apart from showing the total of deaths for that country, there is a line chart that gives info about the changes over the years of the number of deaths (obviously even this chart is managed according to the period considered).





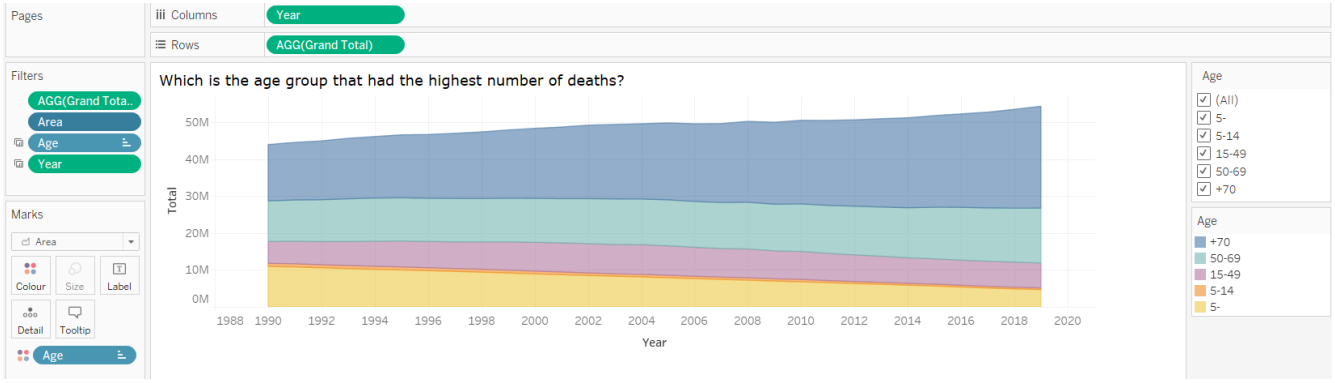
- **Causes of deaths per age**

This sheet is composed of a bar chart that shows the causes of deaths from the most popular to the least. We used just a single tone of blue since we found it better to visualize the available data. The option to show the mark labels is ticked.



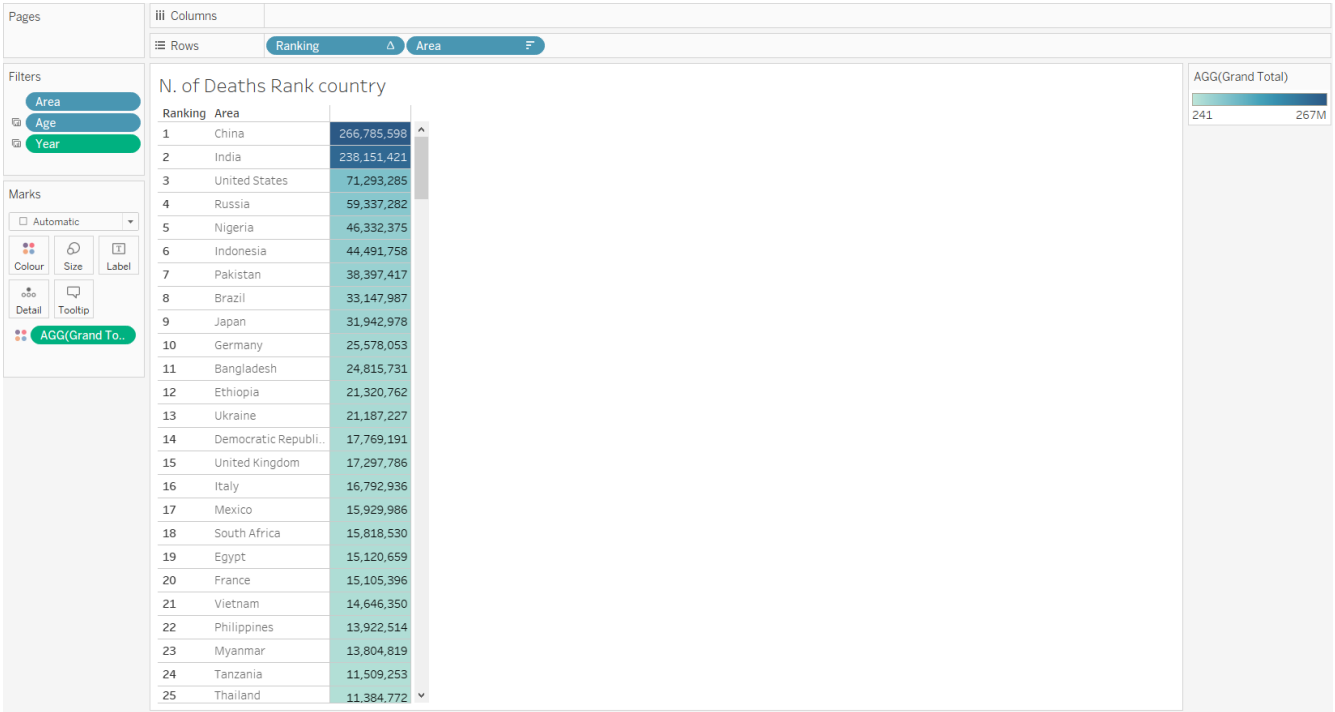
- **N. of Deaths per Age Area chart**

This sheet shows the number of deaths for each age group over the period considered. We used a stacked area chart to track both the total value of the groups considered and to understand the breakdown of that total by groups. Comparing the heights of each segment of the curve allows us to get a general idea of how each subgroup compares to the others in their contributions to the total.



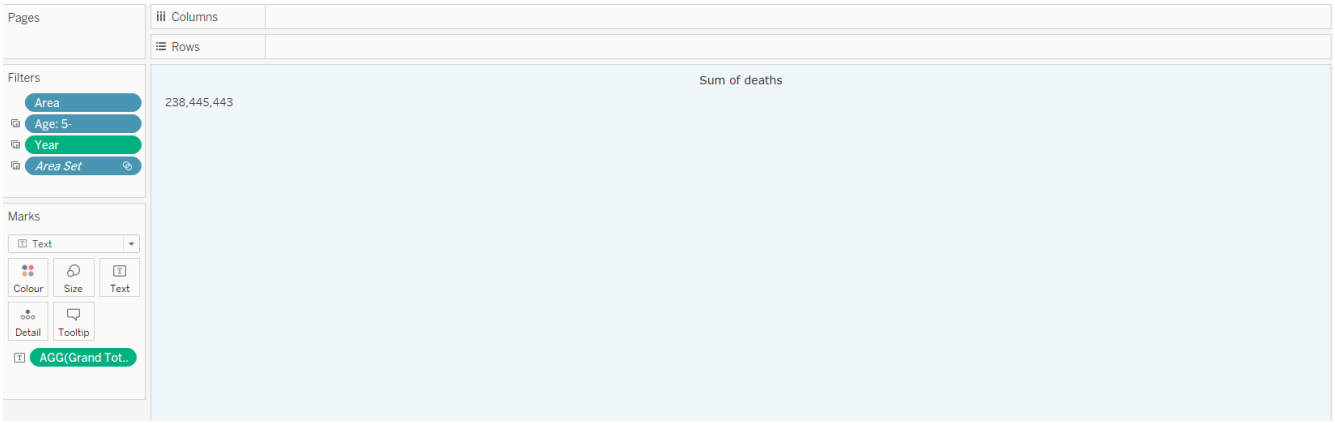
- **N. of deaths rank country**

This sheet focuses on showing in terms of numerical values the rank of the countries in which the highest number of people died. A gradient of blue has been used to underline the countries that are ranked from the top to the bottom.



- **Sum of deaths per age**

This sheet shows the global total number of deaths according to the filters.

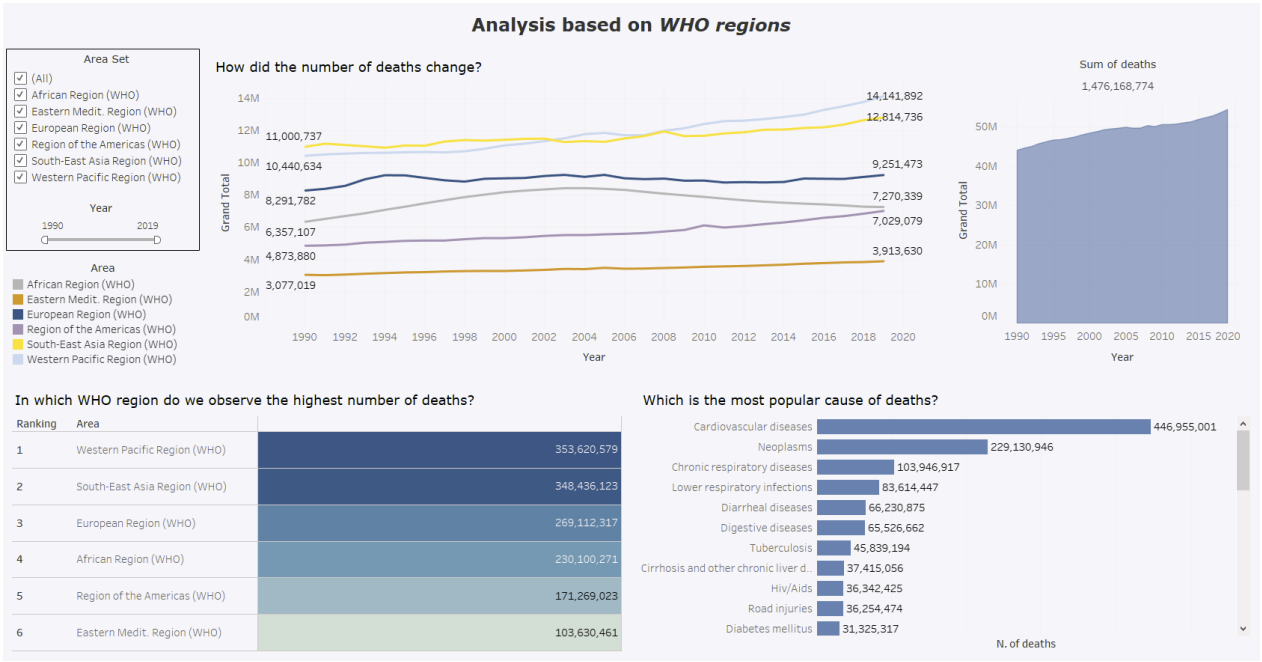


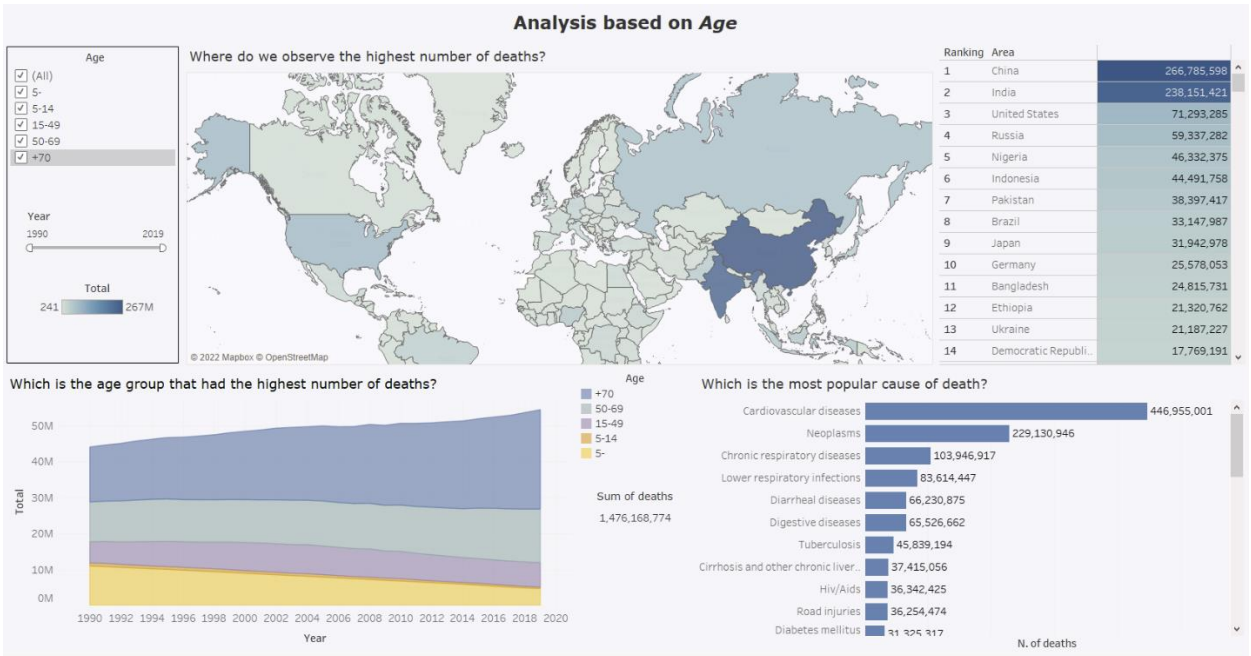
6. Colour Blindness Test

We performed an analysis to understand if our dashboards would suit also for colour blind people. We used [Color Blind Vision Simulator | Color Blind Glasses Simulator (pilestone.com)] as simulator. Overall, the results obtained are pretty satisfying: we can highlight some difficulties just for people with Monochromacy.

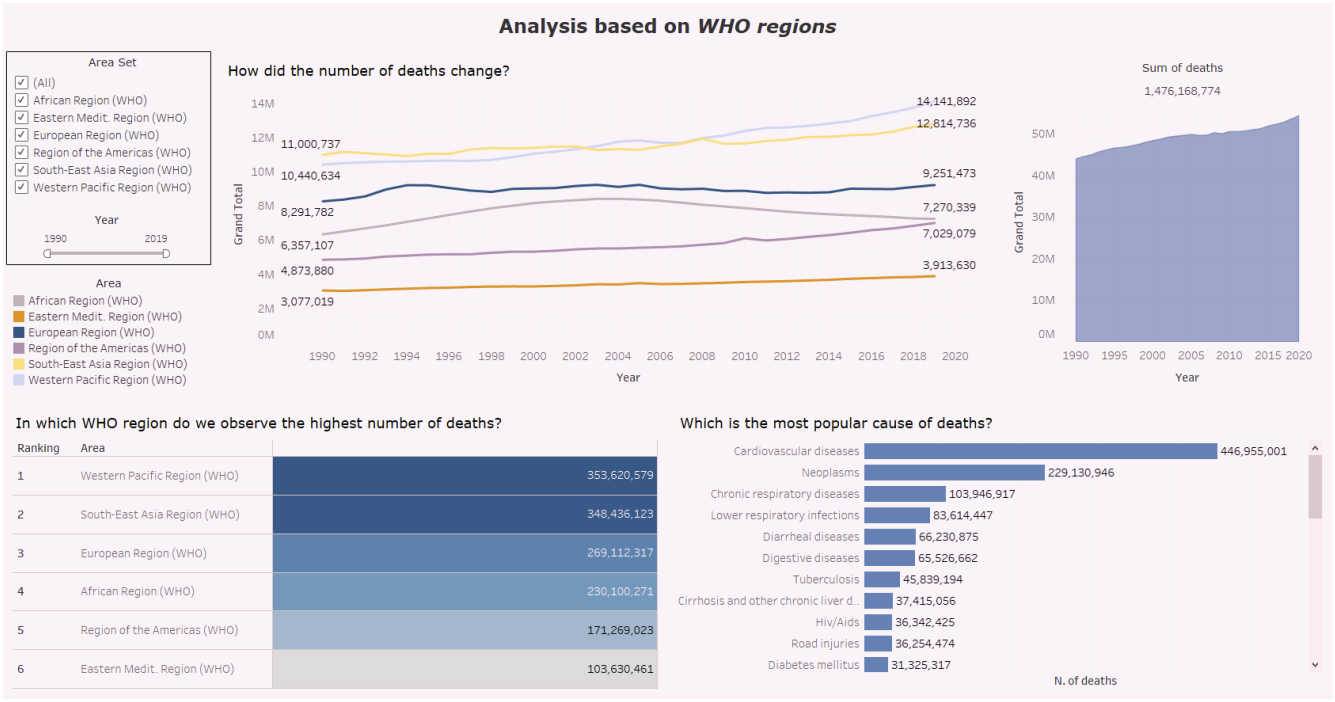
Anomalous Trichromacy

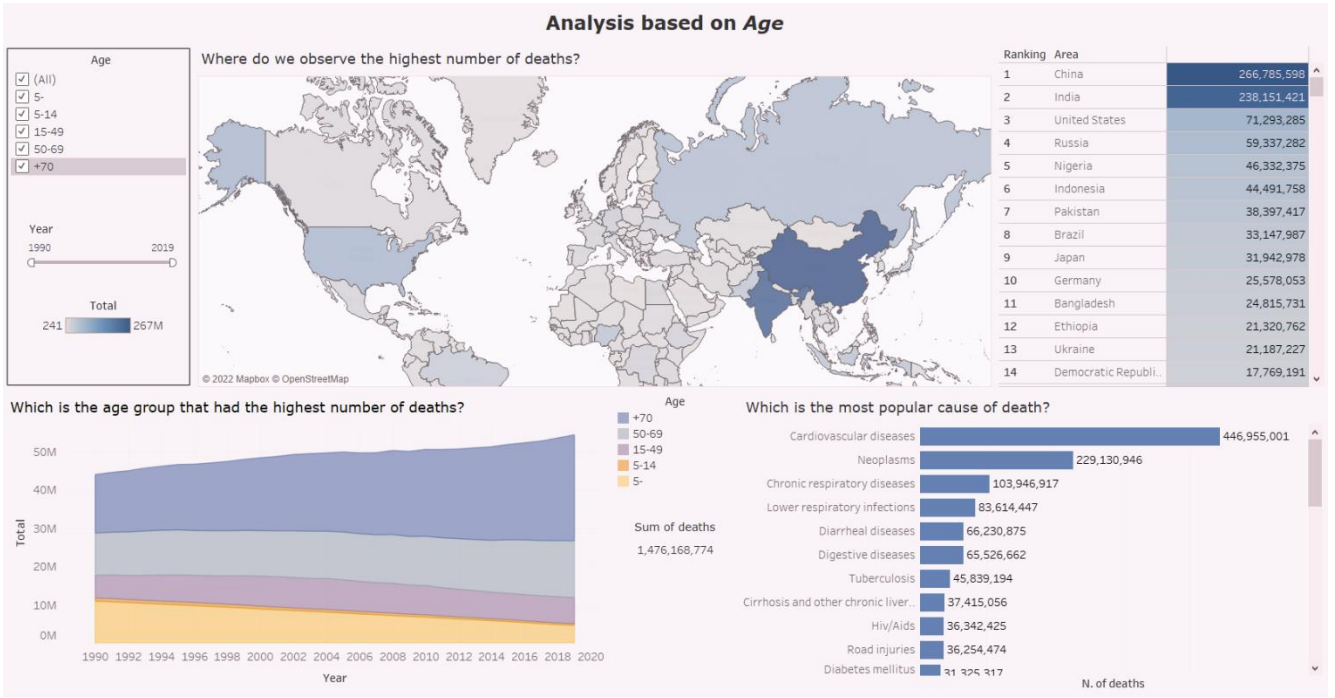
Red-Weak



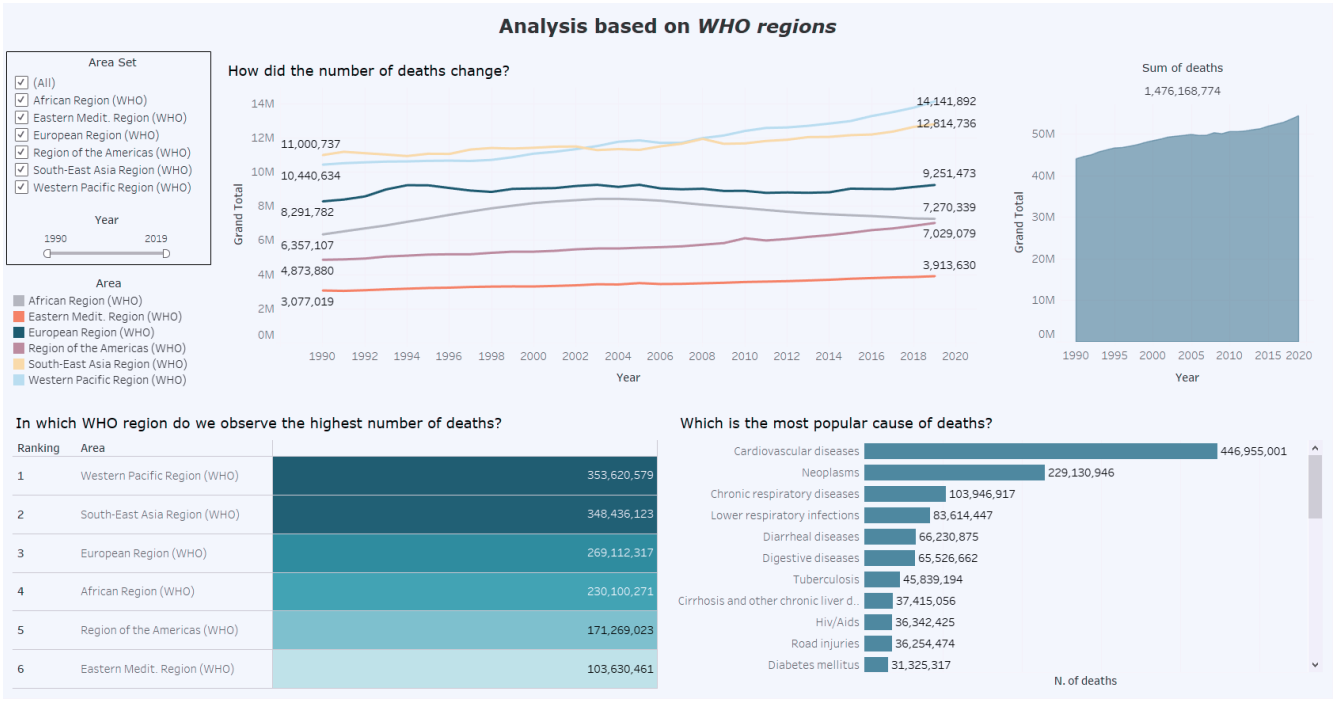


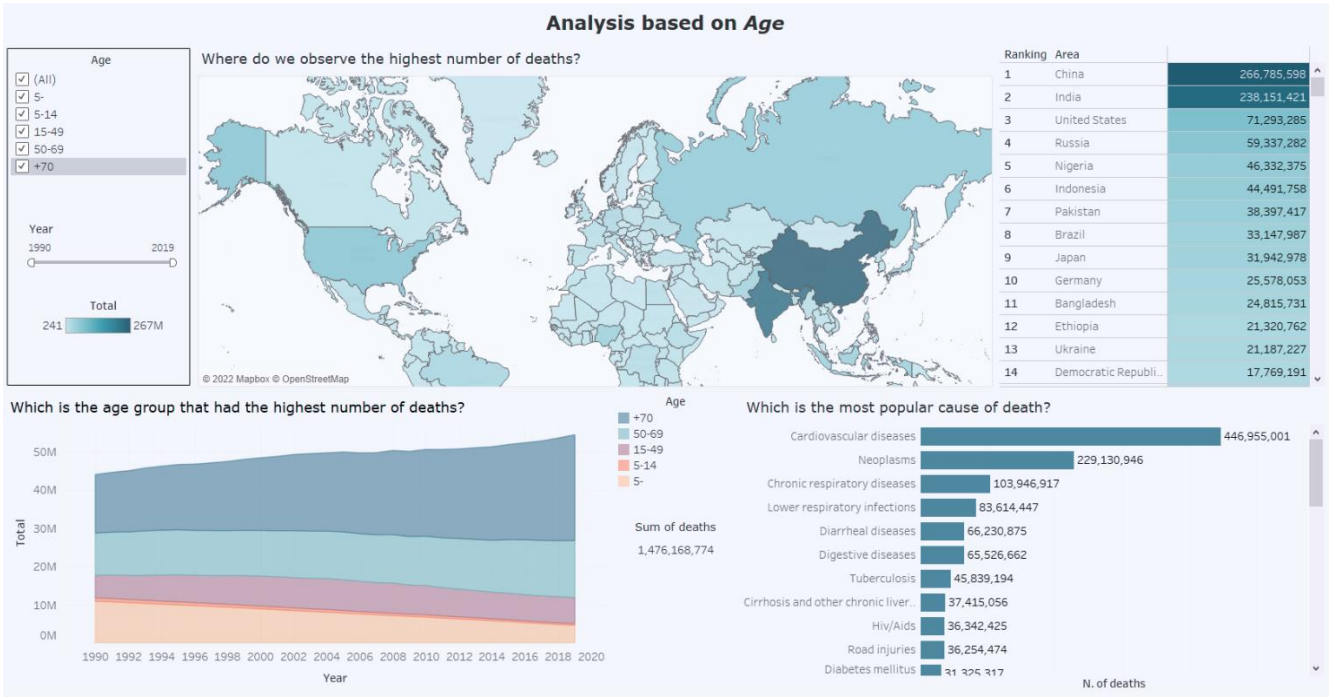
Green-Weak


Analysis based on WHO regions
Area Set
(All)
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
Year
1990
2019
Area
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
How did the number of deaths change?
Grand Total
Year
14,141,892
12,814,736
11,000,737
10,440,634
9,251,473
8,291,782
7,270,339
7,029,079
6,357,107
4,873,880
3,913,630
3,077,019
Sum of deaths
1,476,168,774
In which WHO region do we observe the highest number of deaths?
Ranking
Area
1 Western Pacific Region (WHO) 353,620,579
2 South-East Asia Region (WHO) 348,436,123
3 European Region (WHO) 269,112,317
4 African Region (WHO) 230,100,271
5 Region of the Americas (WHO) 171,269,023
6 Eastern Medit. Region (WHO) 103,630,461
Which is the most popular cause of deaths?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver d.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths


Analysis based on Age
Age
(All)
5-
5-14
15-49
50-69
+70
Year
1990
2019
Total
241
267M
Where do we observe the highest number of deaths?
© 2022 Mapbox © OpenStreetMap
Ranking
Area
1 China 266,785,598
2 India 238,151,421
3 United States 71,293,285
4 Russia 59,337,282
5 Nigeria 46,332,375
6 Indonesia 44,491,758
7 Pakistan 38,397,417
8 Brazil 33,147,987
9 Japan 31,942,978
10 Germany 25,578,053
11 Bangladesh 24,815,731
12 Ethiopia 21,320,762
13 Ukraine 21,187,227
14 Democratic Republi.. 17,769,191
Which is the age group that had the highest number of deaths?
Total
Year
Age
+70
50-69
15-49
5-14
5-
Sum of deaths
1,476,168,774
Which is the most popular cause of death?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths

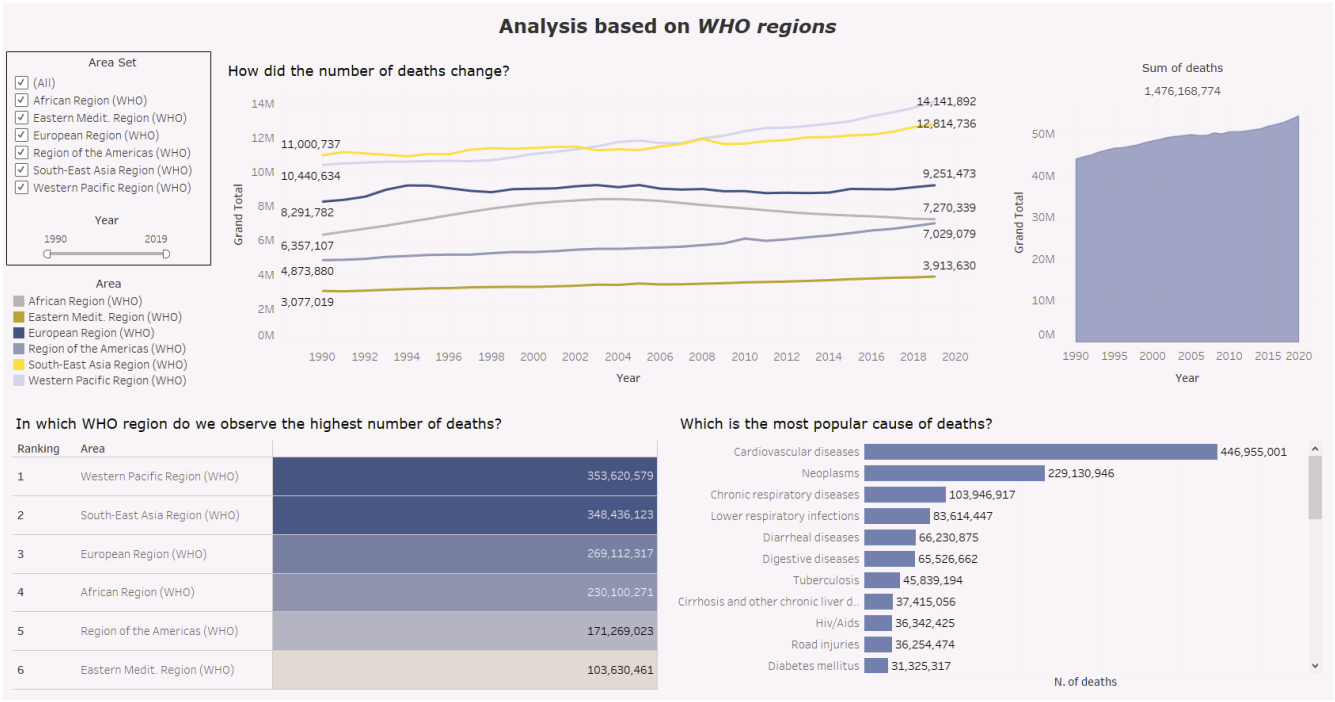
Blue-Weak

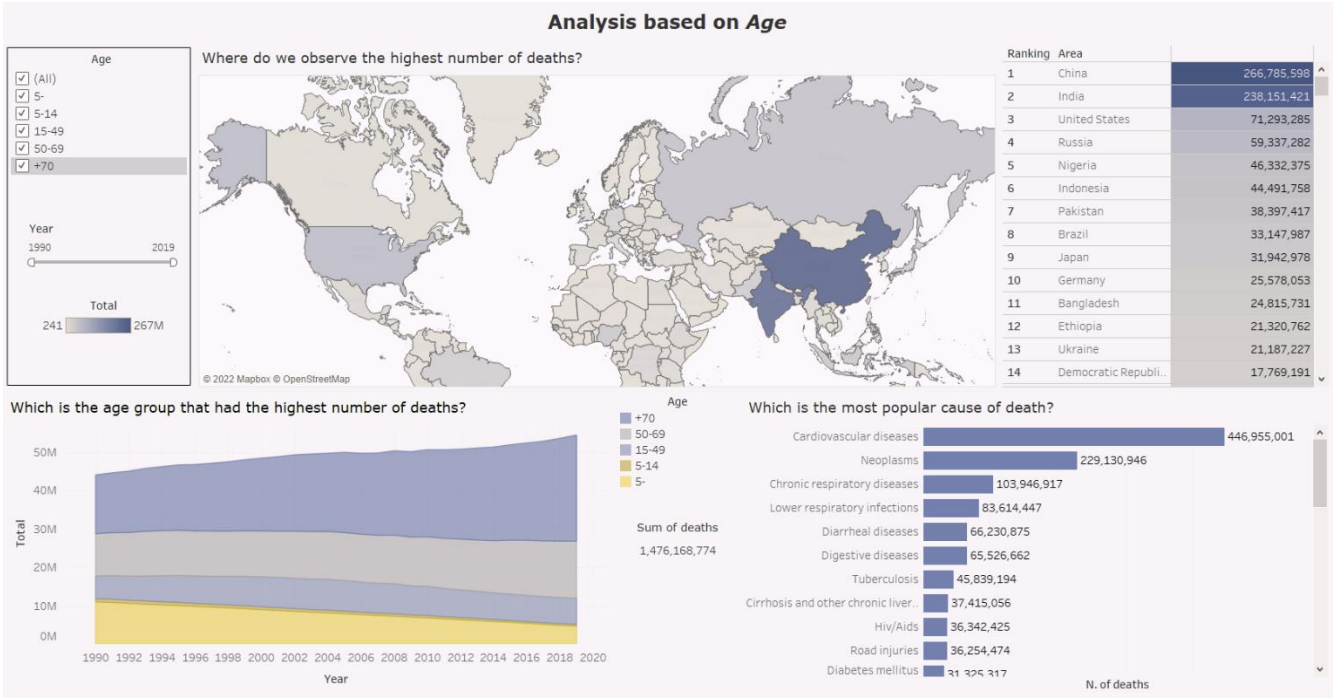

Analysis based on WHO regions
Area Set
(All)
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
Year
1990
2019
Area
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
How did the number of deaths change?
Grand Total
Year
14,141,892
12,814,736
11,000,737
10,440,634
9,251,473
8,291,782
7,270,339
7,029,079
6,357,107
4,873,880
3,913,630
3,077,019
Sum of deaths
1,476,168,774
Grand Total
Year
In which WHO region do we observe the highest number of deaths?
Ranking
Area
1 Western Pacific Region (WHO) 353,620,579
2 South-East Asia Region (WHO) 348,436,123
3 European Region (WHO) 269,112,317
4 African Region (WHO) 230,100,271
5 Region of the Americas (WHO) 171,269,023
6 Eastern Medit. Region (WHO) 103,630,461
Which is the most popular cause of deaths?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver d.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths


Analysis based on Age
Age
(All)
5-
5-14
15-49
50-69
+70
Year
1990
2019
Total
241
267M
Where do we observe the highest number of deaths?
© 2022 Mapbox © OpenStreetMap
Ranking
Area
1 China 266,785,598
2 India 238,151,421
3 United States 71,293,285
4 Russia 59,337,282
5 Nigeria 46,332,375
6 Indonesia 44,491,758
7 Pakistan 38,397,417
8 Brazil 33,147,987
9 Japan 31,942,978
10 Germany 25,578,053
11 Bangladesh 24,815,731
12 Ethiopia 21,320,762
13 Ukraine 21,187,227
14 Democratic Republi.. 17,769,191
Which is the age group that had the highest number of deaths?
Total
Year
Age
+70
50-69
15-49
5-14
5-
Sum of deaths
1,476,168,774
Which is the most popular cause of death?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths

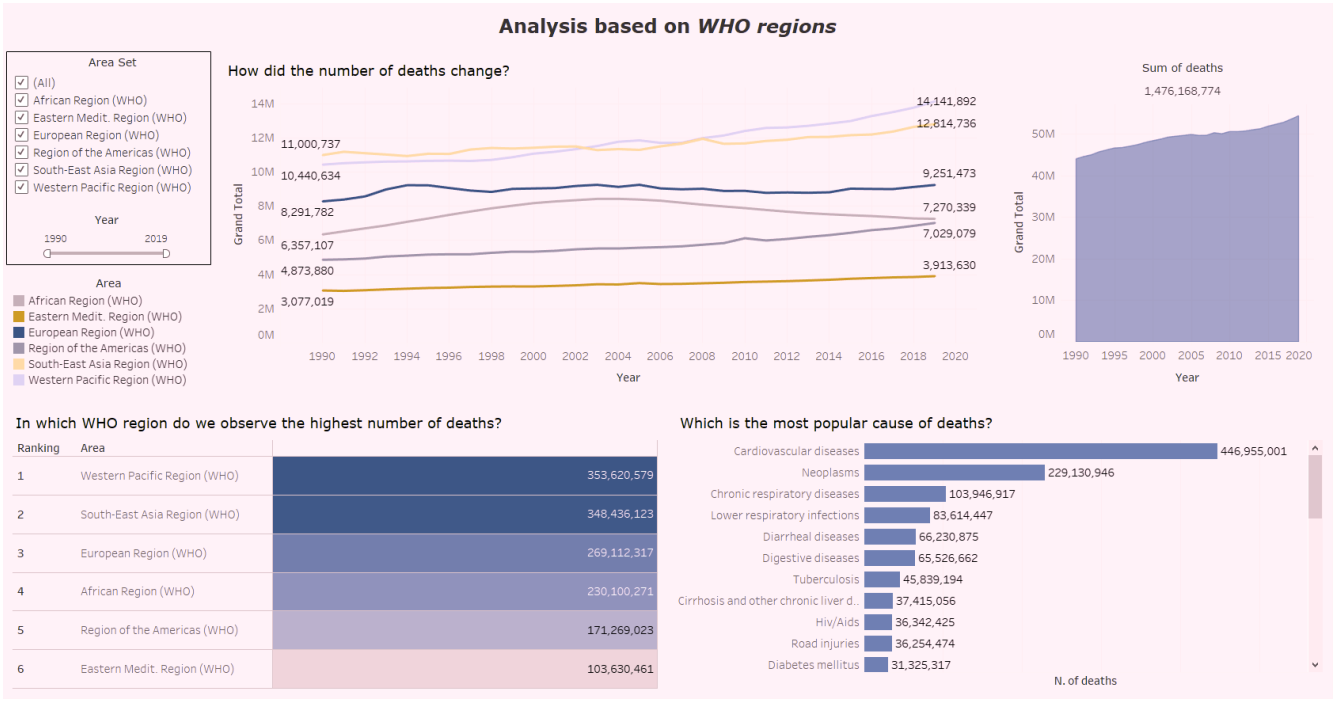
Dichromatic view

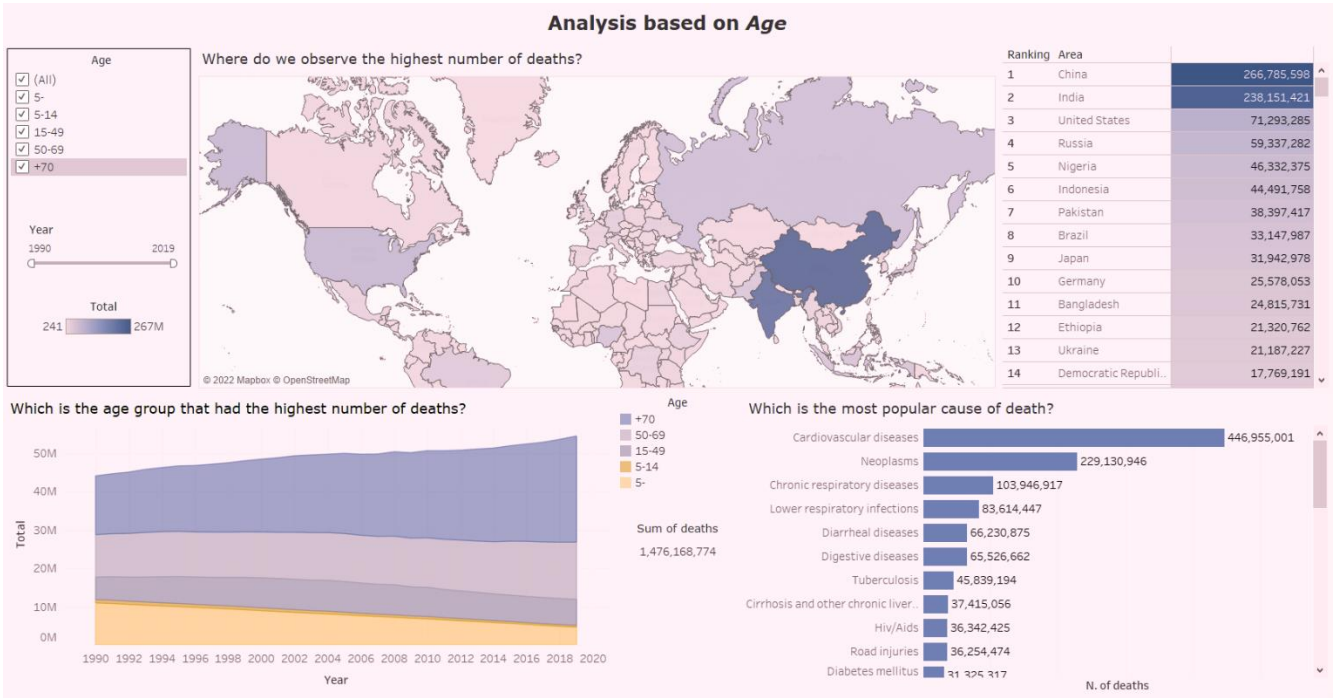
Red-Blind



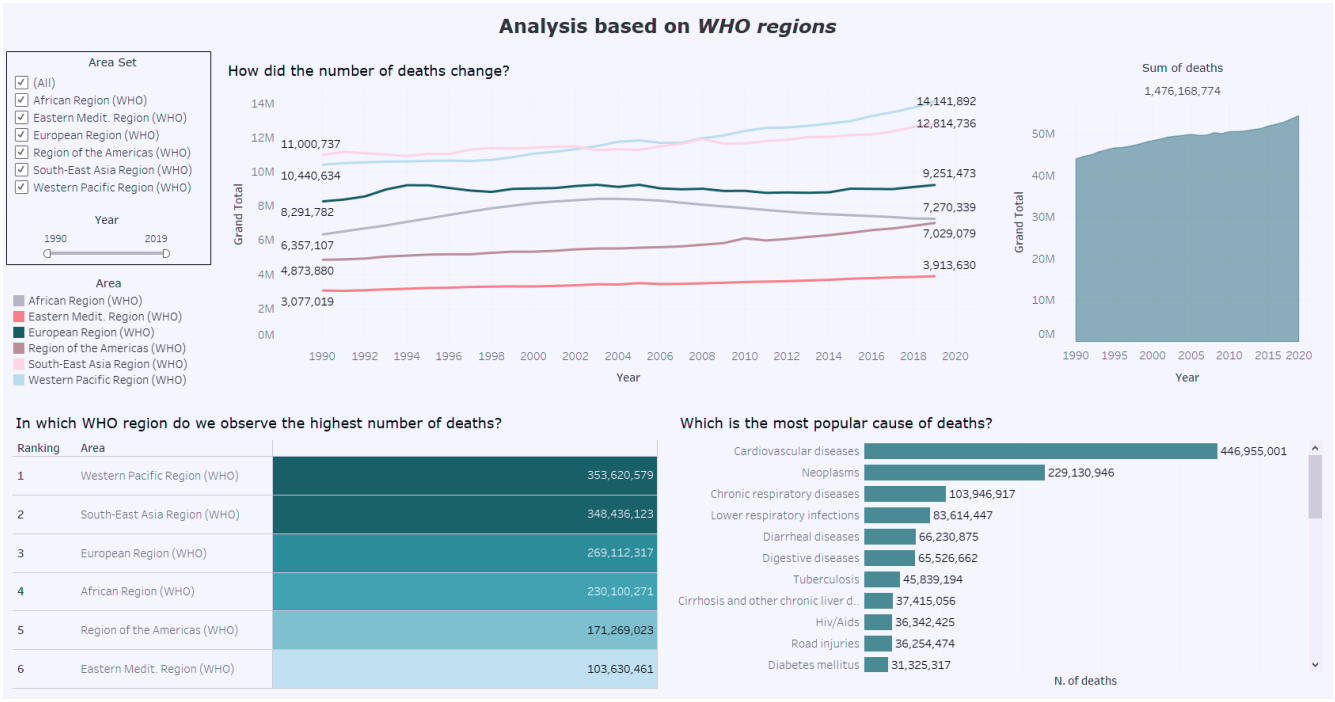


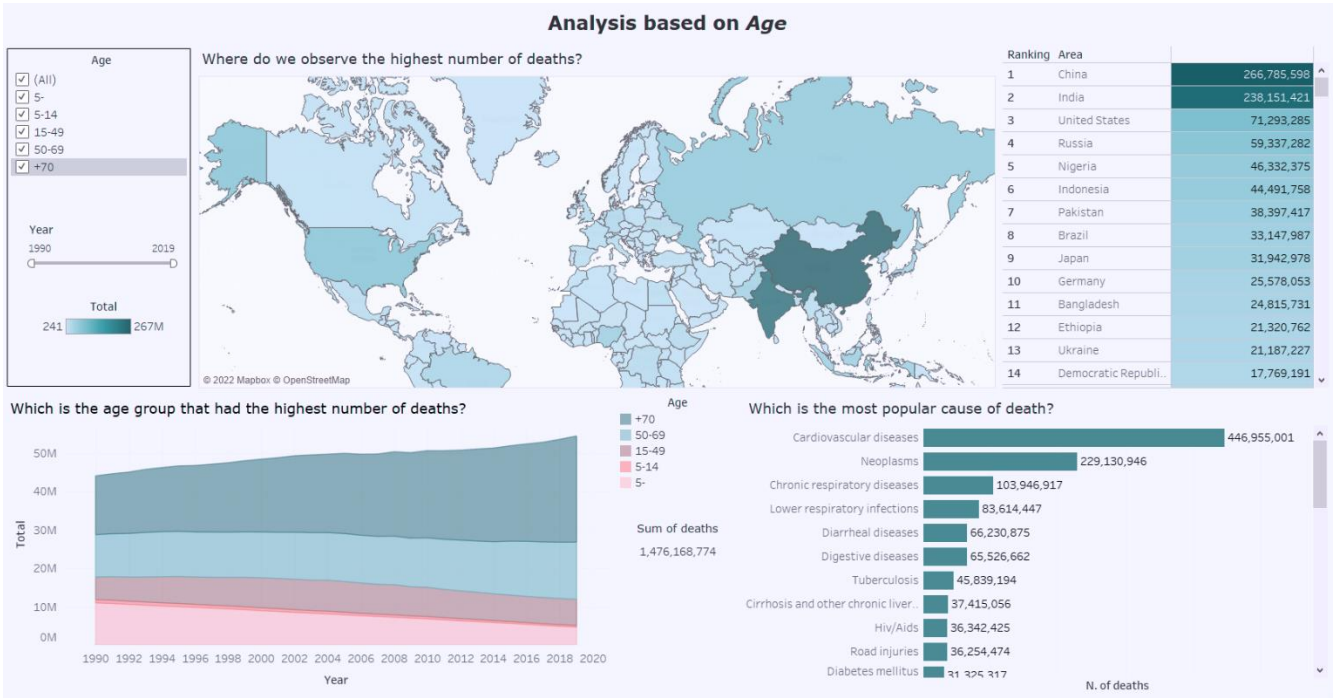
Green-Blind


Analysis based on WHO regions
Area Set
(All)
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
Year
1990
2019
Area
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
How did the number of deaths change?
Grand Total
Year
14,141,892
12,814,736
11,000,737
10,440,634
9,251,473
8,291,782
7,270,339
7,029,079
6,357,107
4,873,880
3,913,630
3,077,019
Sum of deaths
1,476,168,774
Grand Total
Year
In which WHO region do we observe the highest number of deaths?
Ranking
Area
1 Western Pacific Region (WHO) 353,620,579
2 South-East Asia Region (WHO) 348,436,123
3 European Region (WHO) 269,112,317
4 African Region (WHO) 230,100,271
5 Region of the Americas (WHO) 171,269,023
6 Eastern Medit. Region (WHO) 103,630,461
Which is the most popular cause of deaths?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver d.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths


Analysis based on Age
Age
(All)
5-
5-14
15-49
50-69
+70
Year
1990
2019
Total
241
267M
Where do we observe the highest number of deaths?
© 2022 Mapbox © OpenStreetMap
Ranking
Area
1 China 266,785,598
2 India 238,151,421
3 United States 71,293,285
4 Russia 59,337,282
5 Nigeria 46,332,375
6 Indonesia 44,491,758
7 Pakistan 38,397,417
8 Brazil 33,147,987
9 Japan 31,942,978
10 Germany 25,578,053
11 Bangladesh 24,815,731
12 Ethiopia 21,320,762
13 Ukraine 21,187,227
14 Democratic Republi.. 17,769,191
Which is the age group that had the highest number of deaths?
Total
Year
Age
+70
50-69
15-49
5-14
5-
Sum of deaths
1,476,168,774
Which is the most popular cause of death?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths

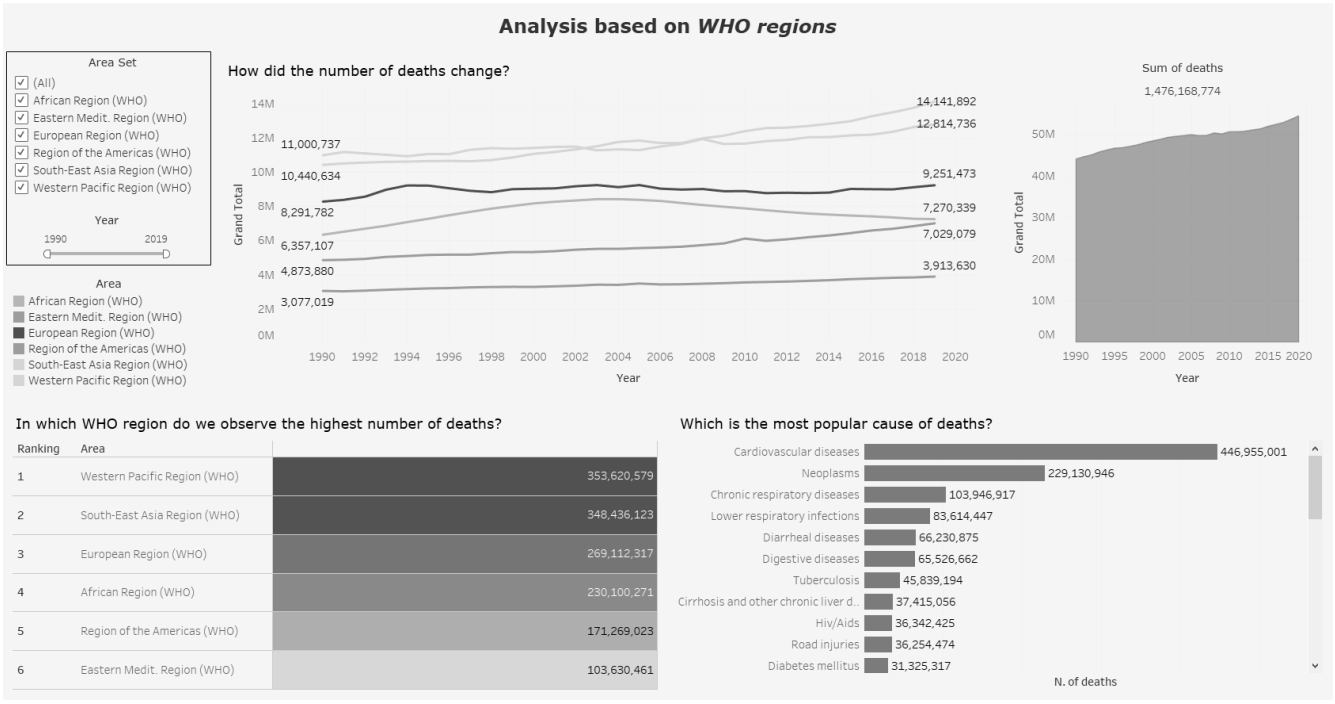
Blue-blind

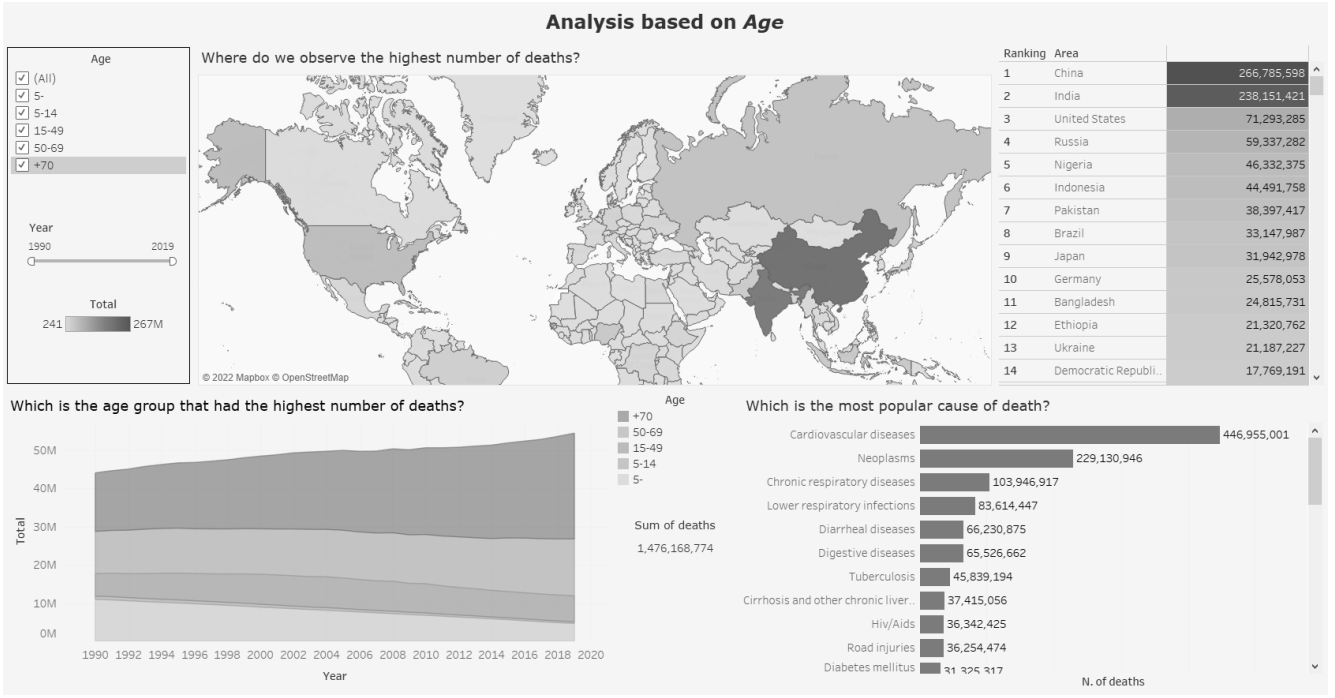

Analysis based on WHO regions
Area Set
(All)
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
Year
1990
2019
Area
African Region (WHO)
Eastern Medit. Region (WHO)
European Region (WHO)
Region of the Americas (WHO)
South-East Asia Region (WHO)
Western Pacific Region (WHO)
How did the number of deaths change?
Grand Total
Year
14,141,892
12,814,736
11,000,737
10,440,634
9,251,473
8,291,782
7,270,339
7,029,079
6,357,107
4,873,880
3,913,630
3,077,019
Sum of deaths
1,476,168,774
In which WHO region do we observe the highest number of deaths?
Ranking
Area
1 Western Pacific Region (WHO) 353,620,579
2 South-East Asia Region (WHO) 348,436,123
3 European Region (WHO) 269,112,317
4 African Region (WHO) 230,100,271
5 Region of the Americas (WHO) 171,269,023
6 Eastern Medit. Region (WHO) 103,630,461
Which is the most popular cause of deaths?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver d.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths


Analysis based on Age
Age
(All)
5-
5-14
15-49
50-69
+70
Year
1990
2019
Total
241
267M
Where do we observe the highest number of deaths?
© 2022 Mapbox © OpenStreetMap
Ranking
Area
1 China 266,785,598
2 India 238,151,421
3 United States 71,293,285
4 Russia 59,337,282
5 Nigeria 46,332,375
6 Indonesia 44,491,758
7 Pakistan 38,397,417
8 Brazil 33,147,987
9 Japan 31,942,978
10 Germany 25,578,053
11 Bangladesh 24,815,731
12 Ethiopia 21,320,762
13 Ukraine 21,187,227
14 Democratic Republi.. 17,769,191
Which is the age group that had the highest number of deaths?
Total
Year
Age
+70
50-69
15-49
5-14
5-
Sum of deaths
1,476,168,774
Which is the most popular cause of death?
Cardiovascular diseases 446,955,001
Neoplasms 229,130,946
Chronic respiratory diseases 103,946,917
Lower respiratory infections 83,614,447
Diarrheal diseases 66,230,875
Digestive diseases 65,526,662
Tuberculosis 45,839,194
Cirrhosis and other chronic liver.. 37,415,056
Hiv/Aids 36,342,425
Road injuries 36,254,474
Diabetes mellitus 31,325,317
N. of deaths

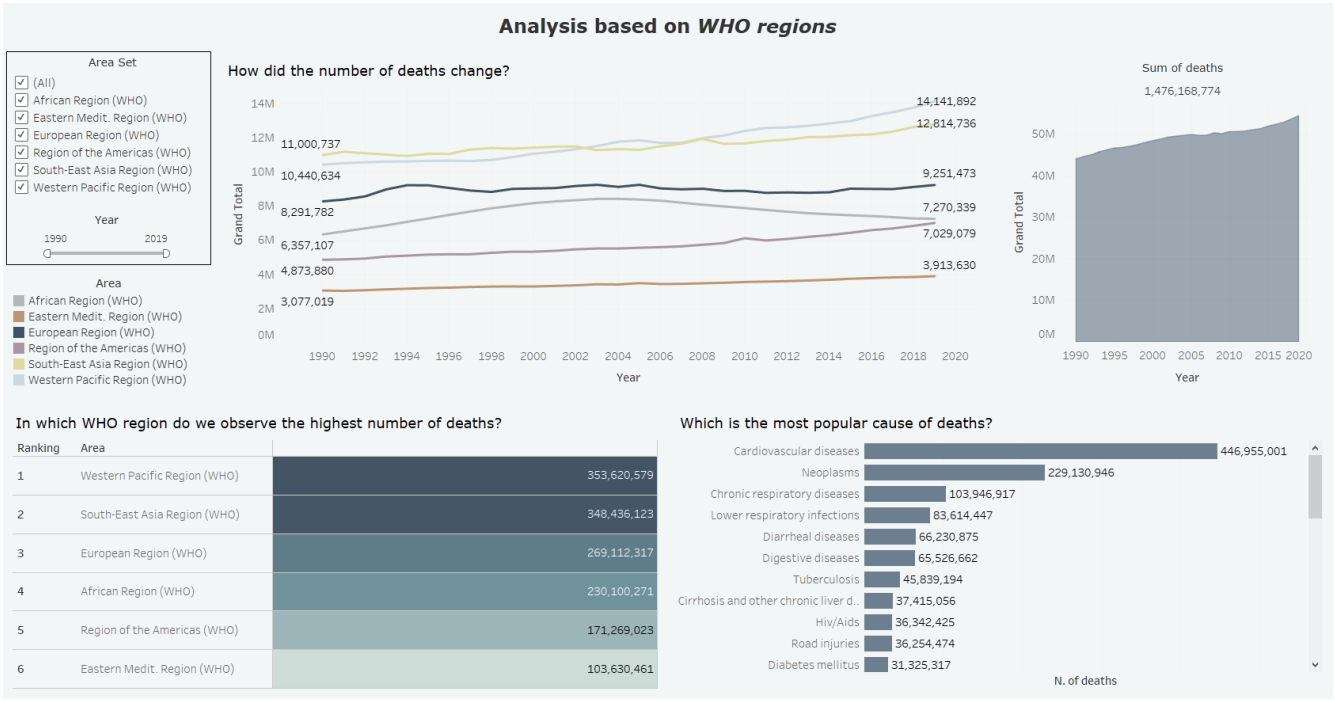
Monochromatic view

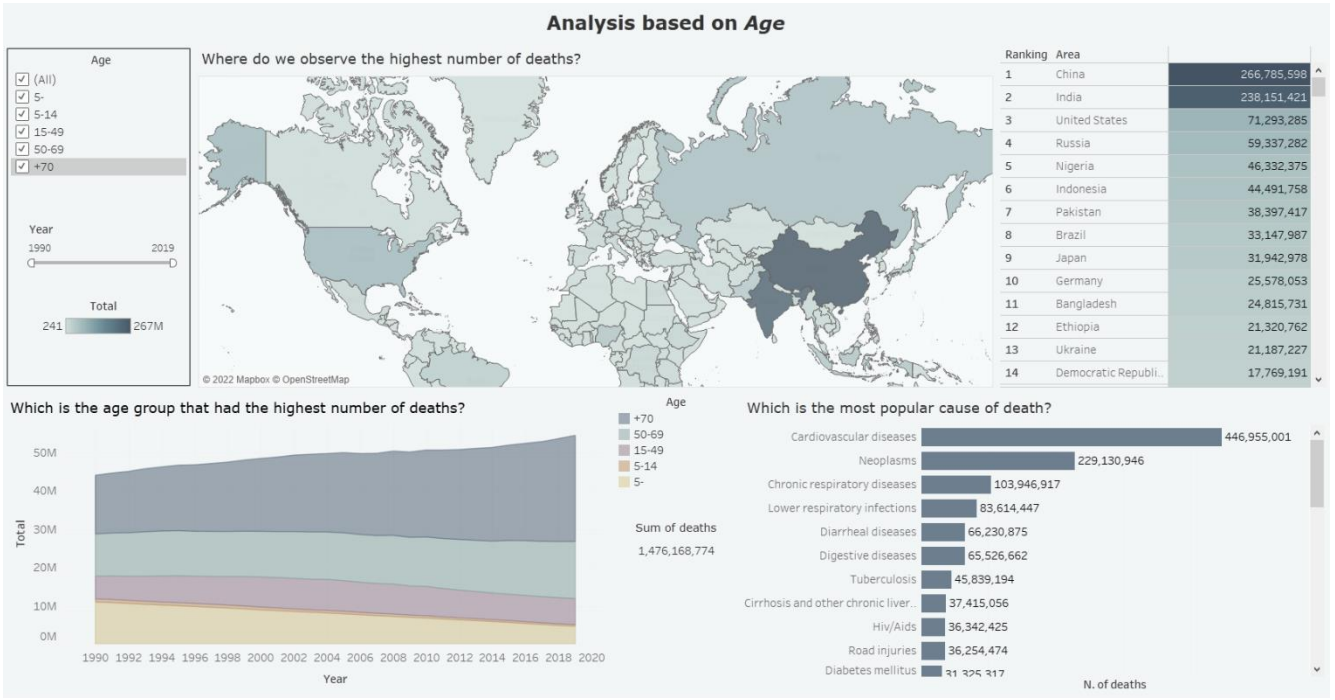
Monochromacy





Blue Cone Monochromacy





7. Answers to business questions

Now that we have all set, we can answer to the business questions we have pointed out at the beginning of the report.

1) **Considering the WHO regions**:

a) Which is the most common cause of death?

Considering the last 30 years, we can observe that in the African Region *Hiv/Aids* has been the most popular cause of death, while for the others five WHO regions *cardiovascular diseases* represent by far the biggest reason why people die. However, if we restrict the time frame, we can notice that even in the African Region in the last ten years (2010-2019) *cardiovascular diseases* have been preponderant, while *HIV/Aids* in the years between 1990 and 2009 was at the top of the chart.

b) In which WHO regions do we observe the highest number of deaths?

The Western Pacific Region has the highest number of deaths (353.620.579), followed by South-East Asia Region with 348.436.123. Over the years, Western Pacific Region and South-East Asia Region have always been at the top of the chart, inverting their position in 2002 and having very close values between 2006 and 2008.

c) How did the number of deaths change from 1990 to 2019?

Generally speaking, we observe for every region an increase in terms of number of deaths over these 30 years (except the African and the European region where a small decrease is detectable). The highest increase in percentage is that of the Region of the Americas, passing from 4.8 M in 1990 to 7 M in 2019.

d) Taking into account the time frame, can we draw some insights?

Some interesting considerations have been made in the previous answers. According to the needs, the flexibility of the dashboards allows to see more in depth how over the time each element changed.

2) **Considering the age groups:**

a) Which is the most common cause of death?

For the age group 5- the most common cause of death are *lower respiratory infections*, followed by some problems that may arise for new-born children. For the age group 5-14 *road injuries* is at the top, with *diarrheal diseases* and *drowning* ranked second and third. For all the others age groups we observe that *cardiovascular diseases* and *neoplasms* are at the top; Hiv/Aids are in the third position for people between 15 and 49, while for people older than 50 *chronic respiratory diseases* occupy that position.

b) In which age group do we observe the highest number of deaths?

The highest number of deaths is related to people that are older than 70, with a relevant positive trend over the years. Considering the others age group, we can observe that there has been a decrease of deaths, showing that the quality of life is increasing and therefore people tend to live longer.

c) Where do we observe the highest number of deaths?

For the age group 5-, we observe the highest number of deaths in India (more than 50 million), followed by Nigeria (almost 27 million) in second place and China and Pakistan with very close values (more than 14 million). For the age group 5-14 the ranking shows India, China and Bangladesh as top 3, with India leading with more than the double of deaths compared to the second place (5 million and 2.2 million). For the age group 15-49 India, China and Russia have the highest values even though Russia is far from the numbers of the other two countries (42.6 M vs 32 M vs 9.7 M). For the age group 50-69 China India and Russia still lead the ranking, while for the age group +70 in the third position United States takes the place of Russia.

Hovering the pointer above the countries in the dashboard allows to see the trend of each state according to the age group selected and we can make some notes. For instance, in the age group 5- is interesting to see how Nigeria had a relevant increase in deaths until the first years of the new century, followed by a decrease that led the country in 2019 in a situation like that of 1990.

d) Taking into account the time frame, can we draw some insights?

The flexibility given by the years slider allows to have the opportunity to go in depth in our analysis at will. For instance, considering the age group 5- and

modifying the time frame to visualize just the last 10 years (2010-2019) we can see how countries like China have cut the death toll by more than half compared to the previous decade (China passed from 3.5 M in the period 2000-2009 to 1.5M between 2010 and 2019). Considering the age group 70+, for most countries we see an increase in number of deaths but there are some exceptions: while China, Japan, India experienced a relevant increase in the last decade (from 47.2M to 58.1M for China, from 7.7M to 10.3M for Japan, from 23.6M to 31.6M for India, so an increase of around +33% for the first two and +23% for the last one), US had a small increase (from 15.4M to 16.7M, +7%) and Russia had even a decrease (from 10.2M to 9.8M, -4.4%).

3) **Which is the general trend? Do we notice a decrease or an increase in deaths from 1990 to 2019?**

   The number of total deaths of the world has increased with an almost steady positive trend, passing from around 44 million in 1990 to more than 54 million in 2019.

Some extra considerations

Our dashboards allow to have both a general point of view and an in-depth knowledge of which are the causes of deaths and how their frequency changes according to the age groups and WHO regions. An extra analysis that would have been interesting to perform is related to how some diseases/injuries changed over the period in percentage: to move accordingly, we would have needed the number of people who lived in each country for each year. As general knowledge we know that the world population is constantly increasing: therefore, some considerations on the decrease or increase of some of the causes of deaths would have had a different perspective.