In this project, I used the skills learned in data visualization to create a piece of data journalism about where the worst coronavirus situation in the world is.

I will present the findings within 15 minutes communicating important messages about the current impact of the disease across the world.

# Scenario

You work for a global public health body, and it has asked you to produce a piece of data journalism on, “***Where is the coronavirus situation the worst?***” to be published on the agency’s website to coincide with an important press conference.

# Data

The data source is available [here](https://github.com/owid/covid-19-data/tree/master/public/data). It is from [Our World in Data](https://ourworldindata.org/), a great resource for data and visualizations with a mission to “..make progress against the world’s largest problems.”

The Data Dictionary is available [here](https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-codebook.csv).

# Skills

I have applied Data Storytelling and Tableau skills such as data visualization best practices, how to communicate the data, how to create visuals, dashboards, and stories within Tableau, and how to create animations.

# Questions

Raised the following questions and clarifications:

**Strategic Context and Objectives**

1. **Why we need to identify where the coronavirus situation is worst:**Our primary responsibility is global health surveillance and coordinated response.

**Identifying hotspots** allows us to:

* Allocate limited resources (medical supplies, personnel, funding) where they're most urgently needed.
* Provide targeted technical assistance to overwhelmed health systems.
* Issue timely travel advisories and public health recommendations.
* Coordinate international support and humanitarian aid.
* Learn from both successful containment strategies and areas of concern.

1. **Actions we will take based on this information:**

* Immediate resource deployment to critical areas.
* Enhanced surveillance in emerging hotspot regions
* Technical support missions to countries with deteriorating situations
* Policy recommendations for international travel and trade
* Coordination with UN agencies for humanitarian response
* Public health emergency declarations where warranted

1. **Press conference context:**This is **our monthly global situation** briefing where we provide transparent updates to the international community on pandemic status, emerging threats, and coordinated response efforts. The timing coincides with the release of our quarterly global health security report.
2. **Key message for the press conference:**"While we see encouraging progress in some regions, the pandemic remains uneven globally. Our data shows that coordinated international action and continued vigilance are essential - no one is safe until everyone is safe."  
   The visualization will support this message by showing both areas of concern and success stories, emphasizing the interconnected nature of global health security.
3. **Target audiences:**Primary: Member state health ministers and policymakers

Secondary: International media and public health professionals

Tertiary: General public and civil society organizations

Internal: WHO country offices and regional directors

1. **Desired audience actions:**Governments: Increase health system preparedness and international cooperation

Media: Accurate reporting that avoids panic while maintaining public awareness

Public: Continued adherence to public health measures and support for global solidarity

Partners: Enhanced coordination and resource sharing

# Exploratory Data Analysis (EDA)

1. **What is the meaning of smoothed measures like new\_cases\_smoothed?**

A 7-day smoothed average calculated by taking the average of data from seven consecutive days – for example, from Monday to Sunday.

Daily case counts can fluctuate rapidly—spiking or dipping due to reporting delays or temporary anomalies—it’s hard to detect meaningful patterns just by looking at day-to-day numbers. That’s why we use the 7-day smoothed average: to reveal longer-term trends and provide a clearer picture of how cases are evolving over time.

Since daily case numbers can jump up and down, it’s tough to spot real trends. A 7-day smoothed average helps by showing the overall pattern more clearly.

**Technical Write-Up**

The 7-day smoothed average is the mean of reported values over a sliding window of seven days. For instance, Sunday’s smoothed value would be the average of data from the previous Monday through Sunday.

It’s like taking the average of a week’s worth of data to smooth out the daily noise and see the real trend—just like averaging your coffee intake over a week instead of reacting to one wild day.

1. **What is the purpose of measures calculated per million?**

By using metrics per million people, we can standardize the impact of the disease across countries, regardless of population size. This allows us to compare countries more fairly and make informed decisions about where support or interventions are most needed.

Measure per million people = Total of the measure divided by the population and multiply by 1 million.

For example, 100,000 cases / 50,000,000 population \* 1,000,000

However, we need to be mindful of country with small population which may skew interpretation of this measure.

1. **What are the iso\_code records which starts with ‘OWID\_’ value?**

This is the region defined by the Our World In Data organization.

## Data Quality

|  |  |
| --- | --- |
| **Column Name** | **Missing Data %** |
| iso\_code | 0 |
| continent | 0 |
| location | 0 |
| date | 0 |
| total\_cases | 3 |
| new\_cases | 3 |
| new\_cases\_smoothed | 4 |
| total\_deaths | 3 |
| new\_deaths | 3 |
| new\_deaths\_smoothed | 4 |
| total\_cases\_per\_million | 3 |
| new\_cases\_per\_million | 3 |
| new\_cases\_smoothed\_per\_million | 4 |
| total\_deaths\_per\_million | 3 |
| new\_deaths\_per\_million | 3 |
| new\_deaths\_smoothed\_per\_million | 4 |
| reproduction\_rate | 56 |
| icu\_patients | 91 |
| icu\_patients\_per\_million | 91 |
| hosp\_patients | 90 |
| hosp\_patients\_per\_million | 90 |
| weekly\_icu\_admissions | 97 |
| weekly\_icu\_admissions\_per\_million | 97 |
| weekly\_hosp\_admissions | 94 |
| weekly\_hosp\_admissions\_per\_million | 94 |
| total\_tests | 81 |
| new\_tests | 82 |
| total\_tests\_per\_thousand | 81 |
| new\_tests\_per\_thousand | 82 |
| new\_tests\_smoothed | 75 |
| new\_tests\_smoothed\_per\_thousand | 75 |
| positive\_rate | 77 |
| tests\_per\_case | 77 |
| tests\_units | 74 |
| total\_vaccinations | 81 |
| people\_vaccinated | 82 |
| people\_fully\_vaccinated | 83 |
| total\_booster | 89 |
| new\_vaccinations | 84 |
| new\_vaccinations\_smoothed | 55 |
| total\_vaccinations\_per\_hundred | 81 |
| people\_vaccinated\_per\_hundred | 82 |
| people\_fully\_vaccinated\_per\_hundred | 83 |
| total\_booster\_per\_hundred | 89 |
| new\_vaccinations\_smoothed\_per\_million | 55 |
| new\_people\_vaccinated\_smoothed | 55 |
| new\_people\_vaccinated\_smoothed\_per\_hundred | 55 |
| stringency\_index | 53 |
| population\_density | 14 |
| median\_age | 20 |
| aged\_65\_older | 23 |
| aged\_70\_older | 21 |
| gdp\_per\_capita | 21 |
| extreme\_poverty | 49 |
| cardiovasc\_death\_rate | 21 |
| diabetes\_prevalence | 17 |
| female\_smokers | 41 |
| male\_smokers | 42 |
| handwashing\_facilities | 62 |
| hospital\_beds\_per\_thousand | 31 |
| life\_expectancy | 7 |
| human\_development\_index | 24 |
| population | 0 |
| excess\_mortality\_cumulative\_absolute | 97 |
| excess\_mortality\_cumulative | 97 |
| excess\_mortality | 97 |
| excess\_mortality\_cumulative\_per\_million | 97 |

There are 5 continents in the data:

1. Africa
2. Asia
3. Europe
4. North America
5. Oceania
6. South America

## Data Cleaning

The data cleaning was done in Excel.

1. Updated the OWID\_AFR continent’s value as ‘Africa’.
2. Updated the OWID\_ASI continent’s value as ‘Asia’.
3. Updated the OWID\_OCE continent’s value as ‘Oceania’.
4. Updated the OWID\_NAM continent’s value as ‘North America’.
5. Updated the OWID\_SAM continent’s value as ‘South America’.
6. Removed the OWID\_HIC, OWID\_LIC, OWID\_LMC, OWID\_UMC and OWID\_WRL records.
   1. OWID\_HIC records refer to High-income countries aggregation defined by OWID organization.
   2. OWID\_LIC records refer to Low-income countries aggregation defined by OWID organization.
   3. OWID\_LMC records refer to Low-middle-income countries aggregation defined by OWID organization.
   4. OWID\_UMC records refer to Upper-middle-income countries aggregation defined by OWID organization.
   5. OWID\_WRL records refer to World countries aggregation defined by OWID organization.
   6. Removed a total of 13,430 records.
7. Removed the Macao records because it has no data for the key metrics like total\_deaths\_per\_million,
8. Merged the 2 separate sets of data reported in the same timeline for the OWID\_EUN records.
9. Merged the 2 separate sets of data reported in the same timeline for the Faroe Island records.
10. Updated the Fiji record with NULL value on 2023-10-15 for key metrics like new\_cases\_smoothed as 0.
11. Updated the Malawi record with NULL value on 2022-11-06 for key metrics like new\_cases\_smoothed as 0.
12. Updated the Maldives record with NULL value on 2020-03-29 for key metrics like new\_cases\_smoothed as 0.
13. Updated the Mauritania record with NULL value on 2023-02-12 for key metrics like new\_cases\_smoothed as 0.
14. Updated the Mayotte record with NULL value on 2021-05-23 for key metrics like new\_cases\_smoothed as 0.
15. Updated the Micronesia record with NULL value on 2024-06-02 for key metrics like new\_cases\_smoothed as 0.
16. Updated the Thailand record with NULL value on 2020-01-26 for key metrics like new\_cases\_smoothed as 0.

## Limitations

1. US total cases is until 2023-05-20.
2. East Timor records have no data for the key metrics like total\_deaths\_per\_million.
3. Macao records have no data for the key metrics like total\_deaths\_per\_million.
4. Taiwan records have no data for the key metrics like total\_deaths\_per\_million.
5. Hong Kong records have no data for the key metrics like total\_deaths\_per\_million.
6. Northern Cyprus records have no data for the key metrics like total\_deaths\_per\_million.
7. Northern Ireland records have no data for the key metrics like total\_deaths\_per\_million.
8. England records have no data for the key metrics like total\_deaths\_per\_million.
9. Wales records have no data for the key metrics like total\_deaths\_per\_million.
10. Scotland records have no data for the key metrics like total\_deaths\_per\_million.
11. Estonia total cases are until 2024-08-04.
12. Europe region total cases are until 2024-08-04.
13. The OWID\_EUN records have 2 separate sets of data reported in the same timeline.
    1. Row 121939
14. France data reported until 2023-07-01.
15. Germany data reported until 2023-07-09.
16. India, Italy, Lithuania, Malaysia, New Zealand data reported until 2024-08-04.
17. No Data from 2020-01-05 to 2020-01-09
18. The hospitalizations and intensive care unit admissions data is sourced from the European Centre for Disease Prevention and Control (ECDC) for selected European countries updated on a weekly basis. It is also sourced from the COVID Tracking Project for the United States and Canada. This explains the reason for the high number of empty records in the data.
    1. If we’ve this data from more countries, then it’ll help with resource planning for countries with strained healthcare facilities.

## Storyboard

1. Explain the context of measures used for the analysis.
   1. Cases per Million
   2. 7-day Smoothed Trends
2. Communicating the facts on:
   1. Cases Impact
   2. Deaths Impact
   3. Cases Trend
   4. Cases Per Million Trends
   5. Deaths Trend
   6. Deaths Per Million Trends
3. Zoom into Europe’s cases per million.
4. Zoom into South America’s deaths per million.
5. Summarize the findings.

# Learnings

Visualization Inspiration from

1. [Clayton Bolitho's](https://public.tableau.com/app/profile/clayton.bolitho/viz/COVID-19DataOWID/Timeline) [Covid-19 viz](https://public.tableau.com/app/profile/clayton.bolitho/viz/COVID-19DataOWID/Timeline)
2. <https://ourworldindata.org/coronavirus>

## Tableau Technical Documentation

Steps to create dynamic measure selection using Parameter and Calculated Field:

1. Created a Parameter as a list of measures to select.
   1. Cases
   2. Cases Per Million
   3. Deaths
   4. Deaths Per Million
   5. New Cases
   6. New Cases (7-day Smoothed)
   7. New Cases Per Million
   8. New Deaths
   9. New Deaths (7-day Smoothed)
   10. New Deaths Per Million
   11. New Deaths Per Million (7-day Smoothed)
2. Created a Calculated Field measure to apply the measure dynamically using the CASE statement.

Steps to create dynamic measure selection using Parameter and Calculated Field for response measures:

1. Created a Parameter as a list of measures to select.
   1. New Tests
   2. New Tests Per Thousand
   3. New Tests Smoothed
   4. New Tests Smoothed Per Thousand
   5. New Vaccinations
   6. New Vaccinations Smoothed
   7. New Vaccinations Smoothed Per Million
   8. People Fully Vaccinated
   9. People Fully Vaccinated Per Hundred
   10. People Vaccinated
   11. People Vaccinated Per Hundred

Changed the dimensions below to measures:

1. Hosp Patients
2. Hosp Patients Per Million
3. Weekly Hosp Admissions
4. Weekly Hosp Admissions Per Million
5. Weekly Icu Admissions
6. Weekly Icu Admissions Per Million