Analysis of member countries of the health world organization (WHO)

How data can help fight coronavirus

1. INTRODUCTION

On the 31 of December 2019 things changed dramatically for everyone in the world. For the first time in modern times, we were faced with the challenges that carry a global pandemic, in a, each day more, totally integrated and interconnected world. The amount of interrelationships that our daily interactions have were placed into evidence, and once again humanity was faced with one of it's worst fears: The unknown.

Very little was known about this virus, and even less was known on how to fight it, and data was scarce. Fear quickly took place in people's hearts, governments and corporations were baffled, having no idea what to do. For the first time in a long time, humanity was reminded of it's fragility, and so we had two choices: Either work together and unite against this invisible enemy, or stand idle while our oldest generations were targeted by the virus, and most of the times, fell irrecoverably ill.

Science couldn't give us straight answers (this, after all, was a new threat we were facing). And so, governments relied on the only thing we had available that we could trust: Data.

Many important conclusions were drawn thanks to data. For instance, the mortality rate of the virus was actually less than people feared it was, around 0.2% for younger people, but it became increasingly dangerous towards elderly generations (20%, around 100 times deadlier, specially if the person had any underlying condition).

The mortality rate also gave a lot of feedback regarding each government policies and transparency regarding Coronavirus. Whenever an abnormal mortality rate was detected, reports of the government hiding cases (or not testing enough) soon showed up.

This was the case with Iran, which reported a mortality rate of over 60% for the virus, which didn't relate at all with the data from other parts of the world. Soon after, Iran was on the spotlight for hiding reports of coronavirus cases, and videos showing morgues and hospitals completely collapsed made it into the news. Governments became increasingly worried, watching as, one by one, each country's health system was tested to the limit.

This analysis studies the differences between each WHO regions and draws conclusions for the causes of these differences. Also, it will study the evolution of death/recovery rates over a period of time (from 22-01-2020 to 16-06-2020) for each of WHO´s member countries.

2. DATA

The original data required some cleaning, unfortunately every column had a general string format (everything from numbers to dates were labelled as strings), and some countries had their first letter in lowercase. Also, there were several columns which had no usefulness (like the “new cases” columns, which only showed the delta between one day and the other, tableau can easily calculate this value and there is no point in increasing the database´s size with this data). The data also came in CSV format, and needed to be loaded into a proper SQL database.

For this purpose I used SQL System integration services (SSIS) to create a pipeline in order to extract, transform and load the data into the database, as follows:

1) Source data: “full_grouped.csv”

2) SSMS database: “Covid19DB”

3) SSMS Table: “Covid19TB”

The SSIS pipeline takes the data from the “full_grouped.csv” file, transforms it, and uploads it into the “Covid19DB”, into the “Covid19TB” table, specifically.

The SSIS pipeline is shown in the following picture:

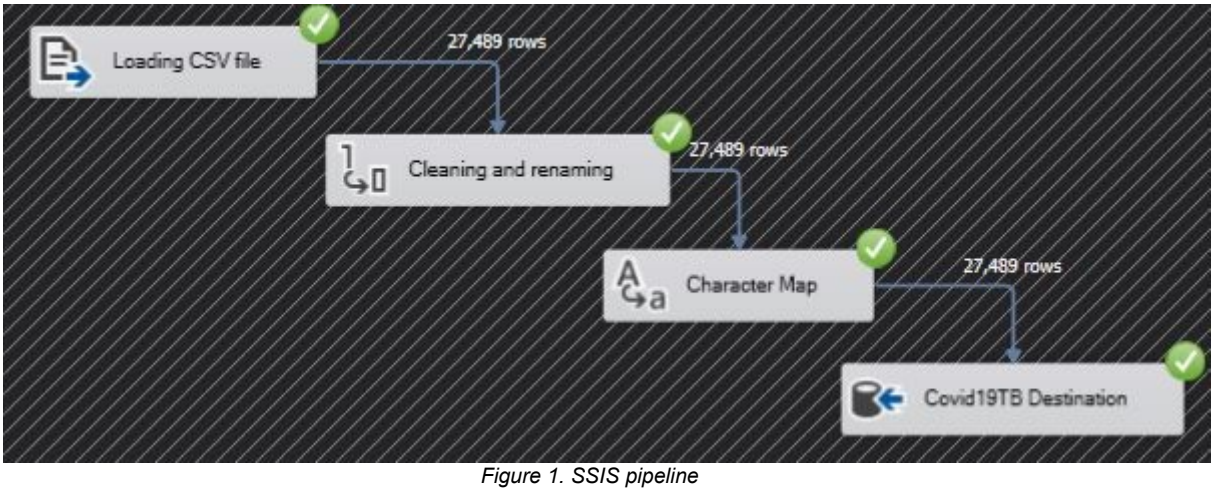

Figure 1. SSIS pipeline

Once the data is loaded into the DB, using SQL queries, we are able to extract additional information, like the lethality of the virus, or the recovery rate of each country.

```
SELECT WHO, SUM(Confirmed) AS Confirmed, SUM(Death) AS Death, SUM(Recovered) AS Recovered, SUM(Active) AS Active,
FORMAT(SUM(Death)/(SELECT SUM(Confirmed)*1.0),'P') AS Lethality, FORMAT(SUM(Recovered)/(SELECT SUM(Confirmed)*1.0),'P') AS RecoveryR
FROM Covid19TB
GROUP BY WHO
ORDER BY Death desc
```

Figure 2. SQL query used to obtain the Lethality/Recovery rate table

The table obtained is as follows:

	WHO	Confirmed	Death	Recovered	Active	Lethality	RecoveryR
1	Europe	1705534207	143029913	708037759	854466535	8.39%	41.51%
2	Americas	1880373365	104419042	497294928	1278659395	5.55%	26.45%
3	Eastern Mediterranean	329633772	9793329	176066683	143773760	2.97%	53.41%
4	Western Pacific	216516677	8018283	150566663	57931731	3.70%	69.54%
5	South-East Asia	150528495	4548076	57842252	88138167	3.02%	38.43%
6	Africa	58923657	1640171	24946051	32337435	2.78%	42.34%

Figure 3. Lethality/Recovery rate table

Our main data turns up being:

```
Date,Country/Region,Confirmed,Deaths,Recovered,Active,New cases,New deaths,New recovered,WHO Region
2020-01-22,Afghanistan,0,0,0,0,0,0,0,Eastern Mediterranean
2020-01-22,Albania,0,0,0,0,0,0,0,Europe
2020-01-22,Algeria,0,0,0,0,0,0,0,Africa
2020-01-22,Andorra,0,0,0,0,0,0,0,Europe
```

Figure 4. Data pre-pipeline

And after the data in passed through the pipeline, we end with:

	Date	Country	WHO	Confirmed	Death	Recovered	Active
1	2020-01-22	Afghanistan	Eastern Mediterranean	0	0	0	0
2	2020-01-22	Albania	Europe	0	0	0	0
3	2020-01-22	Algeria	Africa	0	0	0	0
4	2020-01-22	Andorra	Europe	0	0	0	0
5	2020-01-22	Angola	Africa	0	0	0	0
6	2020-01-22	Antigua and Barbuda	Americas	0	0	0	0
7	2020-01-22	Argentina	Americas	0	0	0	0

Figure 4. Data post-pipeline. This table has 357357 rows by 7 columns.

3. RESULTS

After loading all the data in tableau and coming up with the corresponding plots, we have some already known facts, but some interesting findings as well.

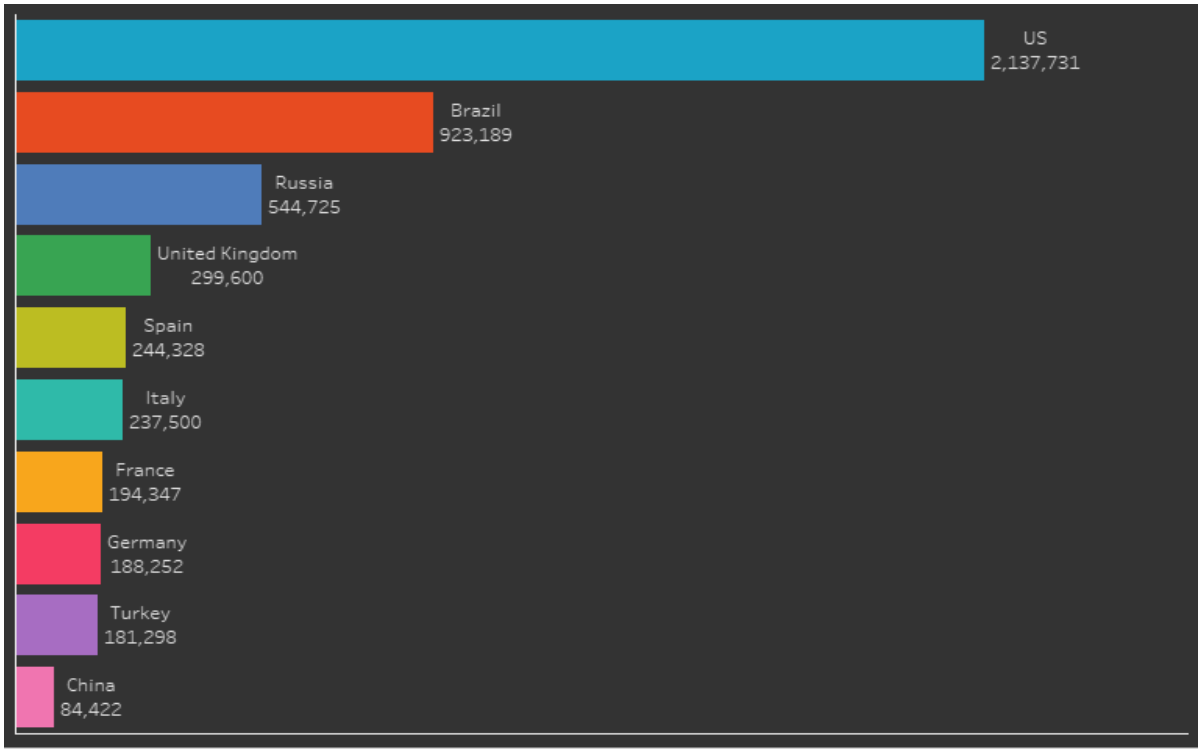

Figure 5. Top 10 countries with the highest Covid-19 related deaths.

To start off, the US has by far the highest death toll of any country, with 2.137.731 deaths as of June 16, 2020. This can be easily explained by the lack of a serious policy to fight the virus. Each state had the freedom to enforce (or not) each of their elected policies (if any), and Trump made it very clear that the US was not going to stop it's economical growth because of a quarantine, and so the virus was left unchecked. Cases promptly increased exponentially, and hospitals in key places like New York were collapsed in term of days.

Following the US steps, Brazil had a similar approach on how to fight the virus. Bolsonaro was against any quarantine whatsoever, even going as far as comparing the virus as a "normal cold" and publicly said that "if people died, then it was just bad luck".

This easily explains the abnormally high mortality rate in Brazil, as well as it's death toll, coming up to 923,189 people dying because of Covid-19.

Russia's case is special as well. Russian officials boasted of having a surprisingly low death rate, while the global average is 36 per million, Russia reported having a 13 deaths per million death-rate. This was extremely abnormal, specially coming from a country with an underfunded health system.

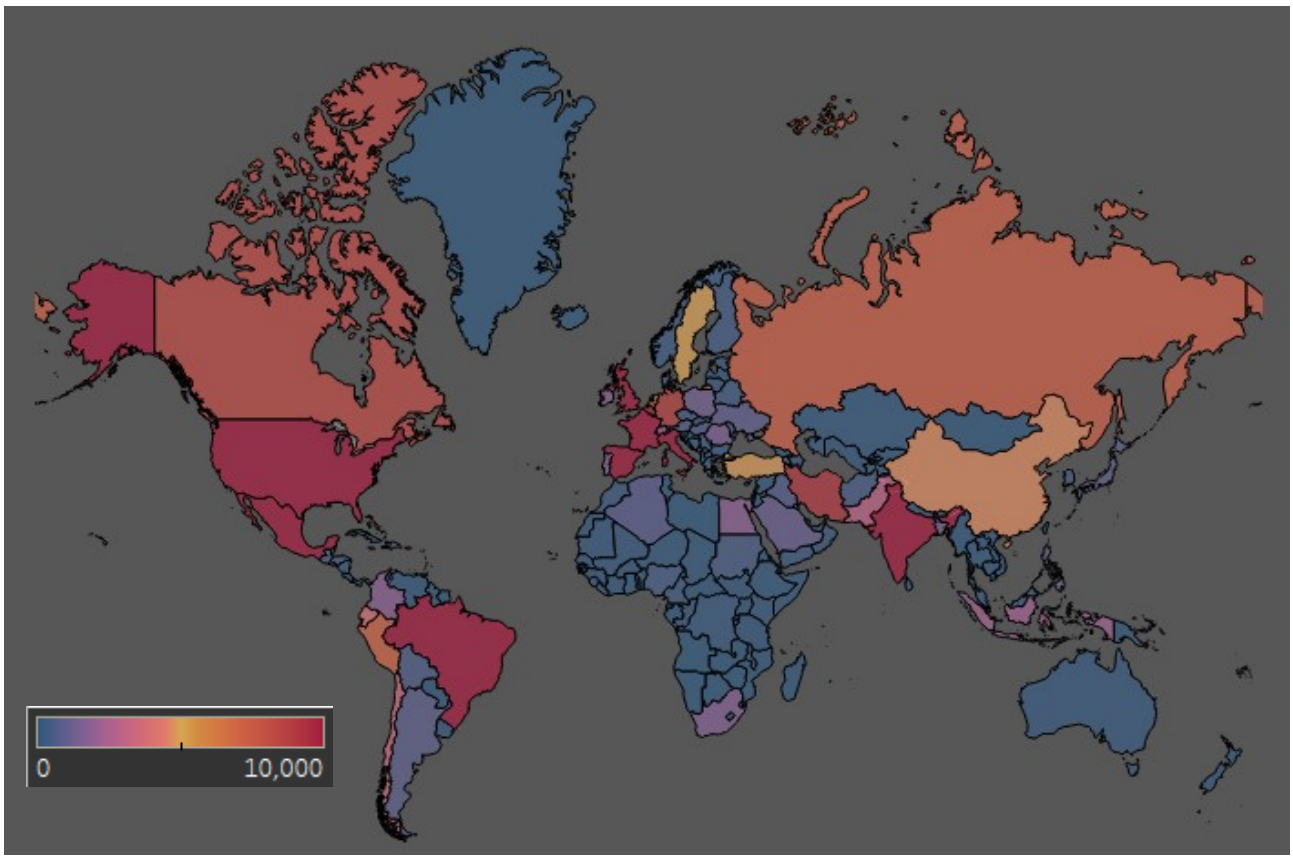

Figure 6. World map on Covid-19, with the countries in red having the most deaths per 100.000 habitants.

The case of India is easy to explain, unfortunately India has a clean water deficiency, making washing hands extremely hard for people in this country. The lack of access to clean drinking water, along with the country's poor general hygiene because of it's elevated poverty numbers and old infrastructure, made it easy for the virus to lay waste on it's population. The lack of a good health system did not contribute as well.

But what is eye-catching is the fact that, almost with no exception, the eastern countries seemed to dwell much better against the virus than their western counterparts. According to The Guardian, "Even the worst-hit central and eastern European countries have infection and death rates per million inhabitants much lower than western European nations, and in some the statistics are truly remarkable: Slovakia has recorded just 1,413 confirmed cases and 25 deaths. Neighbouring Austria, widely regarded as having tackled the challenge of the virus successfully, nevertheless has more than 10 times the number of infections and 20 times the deaths as Slovakia, with a population less than twice the size."

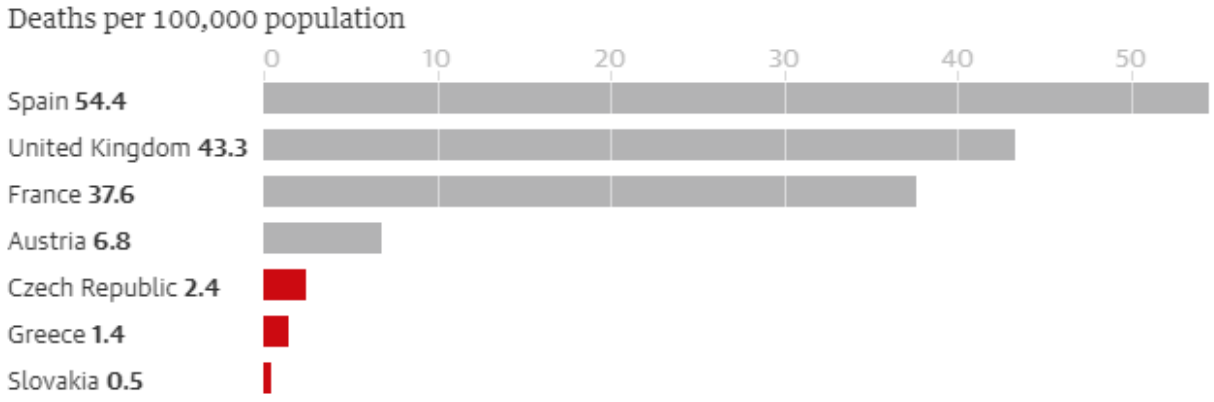

Figure 7. Deaths per 100.000 comparing eastern to western European countries

Numerous contributing causes have been mooted for the discrepancy in various individual countries: lower life expectancy meaning fewer vulnerable elderly people still alive, lower population density, fewer flights to China, lower testing rates or even just sheer luck. The obligatory wearing of masks outdoors, now common to much of Europe, was implemented very early on by the Czech Republic and Slovakia and may also have helped stop the spread.

The most important reason, however, seems to be the early lockdown implemented by almost all countries in the region. While in Britain and other western European nations, public events and gatherings were still going on in the second and third weeks of March, in central and eastern Europe, governments saw what was happening in Italy and implemented rapid lockdowns.
In many places, fear of underfunded and struggling healthcare systems being quickly overwhelmed helped with decisiveness.

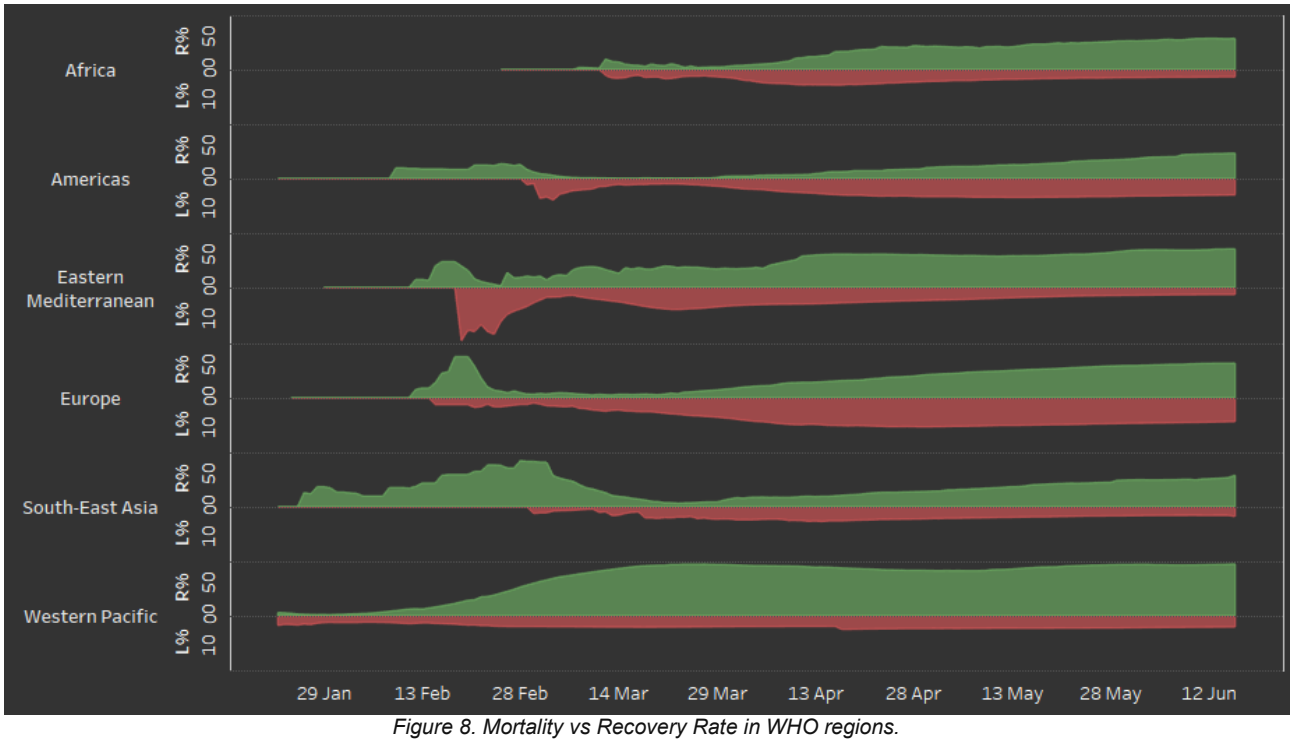

Figure 8. Mortality vs Recovery Rate in WHO regions.

The high recovery rate and low mortality of Western Pacific countries is no surprise. Out of the 5 countries that compose this region, only Brunei Darussalam is not an island. Being able to close the airports and basically isolate from the world, combined with rapid and good health policies regarding Covid-19, proves to be extremely effective (to the point where, to date, New Zealand has 0 active cases)

Figure N°8 shows something more interesting though. The mortality rate of the Eastern Mediterrean (with countries like Iran, Lebanon, Kuwit and Syria, amongst others) on the 18th of February saw an abrupt spike in it's mortality rate, raising to a maximum of 17%, which is by far the highest recorded to date.

To explain this, Dr Tedros Adhanom Ghebreyesus (the world health organization director), made a press conference on that same day addressing this issue. He explained that since December 1, 2019, dozens of health facilities have suspended services in the Idleb and Aleppo areas. Out of nearly 550 health facilities in northwest Syria, only about half are operational. Nearly 900,000 people where displaced, including half a million children. Children are particularly prone to hypothermia and respiratory tract infections, and due to lack of shelter, many of them slept in the open with their families, exposed to the elements.

WHO	Confirmed	Death	Recovery Rate	Lethality
Africa	59M	1.6M	42.34%	2.78%
Americas	1,880M	104.4M	26.45%	5.55%
Eastern Mediterr..	330M	9.8M	53.41%	2.97%
Europe	1,706M	143.0M	41.51%	8.39%
South-East Asia	151M	4.5M	38.43%	3.02%
Western Pacific	217M	8.0M	69.54%	3.70%

Figure 9. Confirmed cases, deaths, recovery rate and lethality of WHO regions..

4. CONCLUSION

The america's regions is one of the most affected by the pandemic, both because the virus has a higher than average lethality, as well as having a very low recovery rate. Their Confirmed cases and death statistics are amongst the worst as well. Europe is still struggling with the virus, having different countries have different strategies on how to deal with it. The high lethality rate of Europe was explained before, but it is also important to note that Europe is well known for having an inverted age pyramid, having most of it's citizen's being elder, who are much more fragile against this virus.

The Covid19 data has been extremely useful since the start of the pandemic, for it has given researchers valuable insight used to draw not only life saving conclusions, but to being able to bring some light into those countries which tend to hide information from the public.

Source:

Link to the tableau interactive dashboard:

https://public.tableau.com/profile/gabriel3881#!/vizhome/Covid-19_15929218379520/Dashboard1