

DATA SCIENCE GROUP RESEARCH

COVID 19

Cristina Segura
Jalex Yestera
Jose M^a González

As part of our bootcamp training, we have done some investigation about the actual pandemic happening all around the planet. Countries were previously selected and assigned to groups, so we have focused in *India, Perú, France, Spain* and *USA*.

1. Data sources:

All used data is taken from <https://covidly.com>.

Most countries data is daily updated except Spanish data which usually takes a one day delay in updating. Due to this circumstances, this research is updated to the day before the current one, so today we'd get yesterday's values and so.

1. Data cleaning:

In order to avoid NaN values and focus on useful data, we have taken out from our countries dataset every '0' before the first infection was registered. This way we avoid plotting a time period where nothing happens (or the virus was not detected).

We haven't removed any outlier values as we could explain them, adding some relevant information which was also required by this investigation task.

1. Data plotting:

We have plotted the pandemic status in our selected countries so we can research trends and get a more accurate view of how the problem is being carried in these nations.

Plotted Trends:

- New daily registered infections by country.
- Total number of infected people by country.
- Infections registered during quarantine in each country (start to end of quarantine).
- Outliers.
- Amount of daily new deaths supposedly caused by covid.
- Total number of death people caused by covid19 since the beginning of the pandemic.
- Total number of deaths by covid19 registered in quarantine period in each country.
- Total number of recovered people in each country.
- % of infected people who died.
- % of infected people that is now recovered from the disease.

Results:

Countries ranking by amount of infected people *(2020/08/07).

Country	Total infected people
USA	4.823.891
India	1.964.536
Perú	447.624
Spain	309.855
France	194.029

Countries ranking by amount of death people *(2020/08/07).

Country	Total death people	% of infected who died
USA	158.256	3,28 %
India	40.699	2,07 %
France	30.305	15,62 %
Spain	28.500	9,20 %
Perú	20.228	4,52 %

Countries ranking by amount of recovered people *(2020/08/07).

Country	Total infected people
USA	4.665.635
India	1.923.837
Perú	427.396
Spain	281.355
France	163.724

Conclusions:

- Number of infected and death people is proportional to each country's population.
- Europe was hit by the virus in March while other countries registered first cases in April or May.
- While maximum infection rates in Europe were reached in May, we can see these infection rates are still growing in other countries.
- Amount of daily deaths is not growing in Francia and Spain, but it keeps growing in USA, India and Perú.
- It is remarkable to notice that death ratio is much lower in India and Perú. These are countries where health system is supposed to be worse, but we think the cause is the population's mean age in these countries is lower and we know the disease kills older people.
- This mean age difference is also reflected in the amount of recovered people.
- Lack of foresight on the part of authorities of each country is detected when it comes to taking social distancing and isolation measures. These measures have been proved as effective containing the infection ratios.

Outliers:

- We have found a great decrease in total number of death people in Spain dated in May 25th, which is the day quarantine was over in this country. The total number of deaths decreased by 1.918 people. We can't believe 1.918 spanish zombies are out in the streets now. Spanish government said those deaths were not caused by covid and that was the reason told in the news. *(We keep the right to believe it or not, as an example of how data is manipulated by authorities).*