

# Data Science and Covid. Group Delivery.

23/07/2020 to 26/08/2020

— Group **B**

**Jose M<sup>a</sup> González, Cristina Segura, Javier A. Yestera**

**Data Science**  
**The Bridge**

## **General Vision**

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

## **Goals**

We aim to reach option A. Once finished all C and B requirements, we have also saved a collection of plots in different folders for each country. We have also used distributed modules for each functionality, analyzed and explained the consequences of alarm state measures taken in each country and plotted the progression measured in 10 days lapses.

## **Specifications**

### **Software**

Python 3 (Including Numpy, Pandas, Matplotlib libraries), Visual Studio Code, Adobe Acrobat or any .pdf Reader, Google Chrome or any suitable internet browser. Works in Windows, Mac and Linux Os

### **Hardware**

AMD Ryzen 3 with 8Gb of Ram memory MacBook Pro i5 processor and 8Gb Ram memory.

### **Requirements**

We need an internet connection in order to update our dataset from <https://ourworldindata.org> where data is free and we won't need to register.

## Steps

- Research the context

We had to research the news and documents to find dates when every country declared the emergency state. Some countries also finished this state and others are still under quarantine measures, so we have been following the track in order to detect changes on this subject.

- Get Data / Data Mining / Clean Data

We get data from the url <https://ourworldindata.org/coronavirus-source-data> . Emergency state dates are taken from Spanish government (Spain), Wikipedia (India), Peruvian embassy in Spain and the USA embassy in Peru (Peru), CNN news (USA). This research is located in our documentation folder.

We have tried to automatize all steps using functions distributed in modules.

- **open\_csv(url)** function to read and convert into dataframe the dataset contained in the given url. It also changes date data to date type and returns the Dataframe. Located in **folders\_tb**.
- **seleccion\_paises(dataframe.paises)** function from **mining\_data\_tb** module to select our assigned countries from the complete dataframe, extracting the desired values with a loc method. Args and kwargs are the dataframe and a list of selected countries.
- **seleccion\_columnas(dataframe, col1, col2)** function in **mining\_data\_tb** module used to select the columns to analyze giving the columns position.
- **borrar\_previo(dataframe)** is used to search the first detected case and delete previous values. Then, using pivot table it returns the dataframe sorted by date, starting when first infection was detected and ending with the last.
- **mining\_data\_tb** module also **has\_cero\_nan(dataframe)** function to change Nan values for zeros by applying the fillna(0) method to the complete dataframe.
- We have detected several **outliers** in all countries data but we have kept them because countries are re defining the data collection. We have also plotted data in a month frequency in order to smooth these outliers,

## Data Wrangling / Visualization and Backup

We have automatized data wrangling and visualization using functions that parse every dataframe column to analyze, selecting a country or a group of countries. Several plots are also saved in independent folders named as countries.

All these functions are located in **visualization\_tb** module.

- **paises\_juntos()** function compares the evolution of several selected countries in a timeline, creating a pivot table for each column, and then plotting, printing and saving a copy of plots in “`../resources/plots/paises_juntos`”. It also updates the previous image if exists. Parameters are the Dataframe and a dictionary of countries.
- **paises\_juntos\_mes()** function is a variation of paises\_juntos with a month period between samples. Saved plots are named as monthly.
- **emergencias()** from **visualization\_tb** plots different trends for each column of selected country. It also plots with vertical lines the start end end (if declared) of the alarm state. Function parses a list made with selected columns. Other parameters are the known dates of emergency activation and deactivation. Plots include grouped countries bars plot, and individual countries plots. All files are saved and overwrite older versions in each country’s folder inside  
“`..resources/plots`”
- **emergencias\_mes()** function is a variation of emergencias with a month period between samples. Saved plots are named as monthly
- **total\_paises()** from **visualization\_tb** function compares the ranking of countries by number of infected and number of death. Both absolute and by million of inhabitants. Parameters are Dataframe, selected columns, and the countries dictionary (containing names and colours). Plots are printed and then saved as “`barras_columnname`” in “`../resources/plots/paises_juntos`”
- **total\_paises\_orden()** function returns the same as total\_paises in reversed order so position of the countries is clearer. Plots are printed and then saved as “`barras_orden_columnname`” in “`../resources/plots/paises_juntos`” .

軍機處奏摺存目表

# DATA SCIENCE GROUP RESEARCH

## COVID 19

Cristina Segura  
Jalex Yestera  
Jose M<sup>a</sup> 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://ourworldindata.org/>.

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.

### 2. 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.

### 3. 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.

### 4. Plotted Trends:

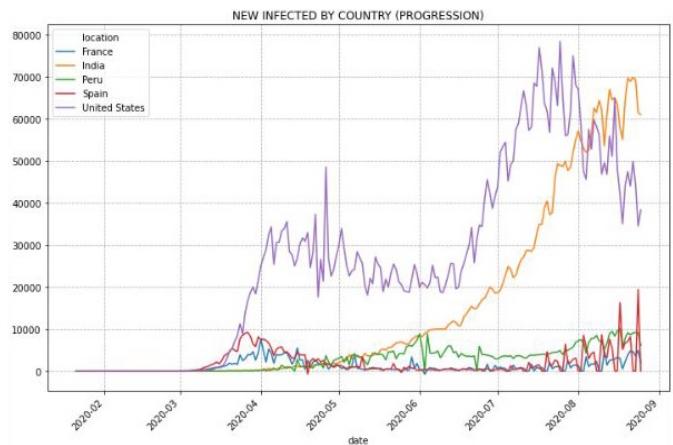
- New daily registered infections by country.
- Total number of infected people by country.
- Proportion of infected who died.
- Percentage of infected and recovered from disease.
- Infections registered during quarantine in each country (start to end of quarantine).
- Outliers.
- Amount of daily new deaths 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.

# COVID 19 - A WORLDWIDE PANDEMIC

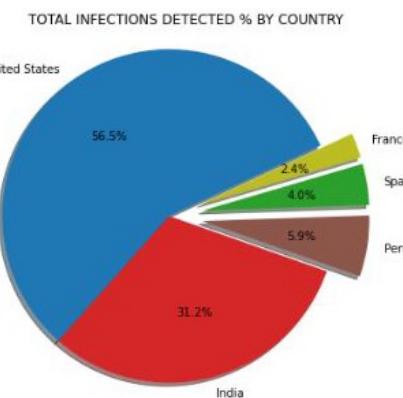
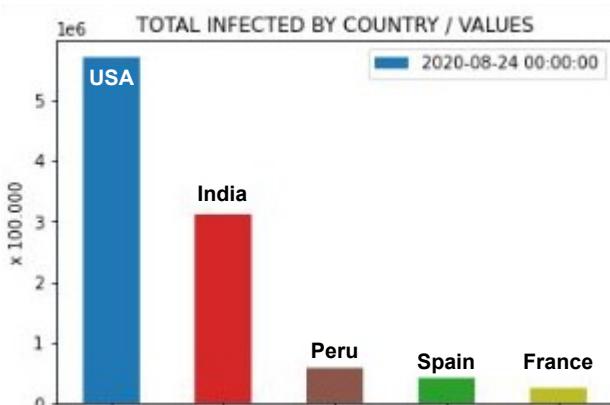
The pandemic data register started growing on March 2020.

Infections are still increasing in some countries where the infection trend hasn't changed.

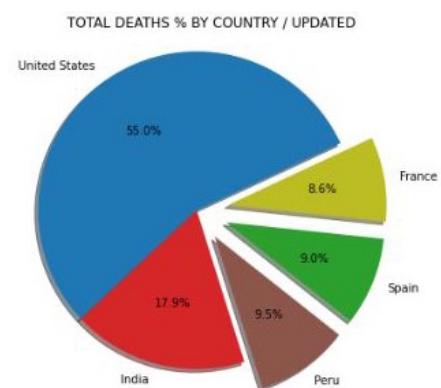
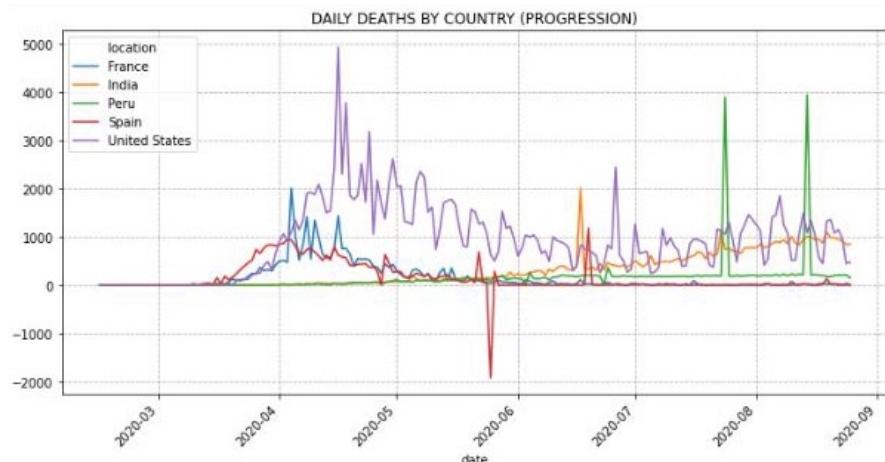
Evolution of infections is being different in each country due to several remarkable factors, such as population size, emergency measures taken by governments and national health system presence.



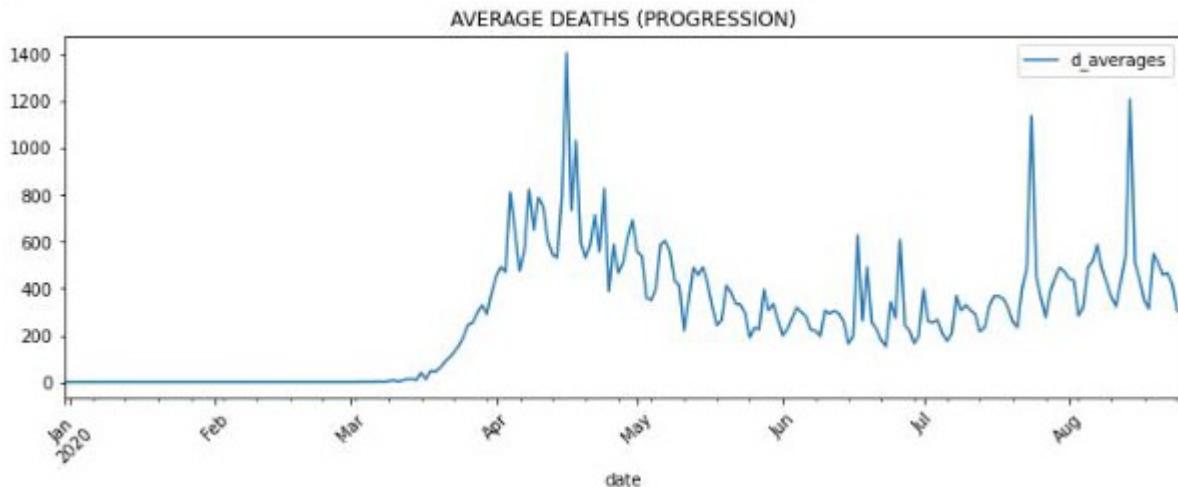
Total amount of infected people is related to country's population size



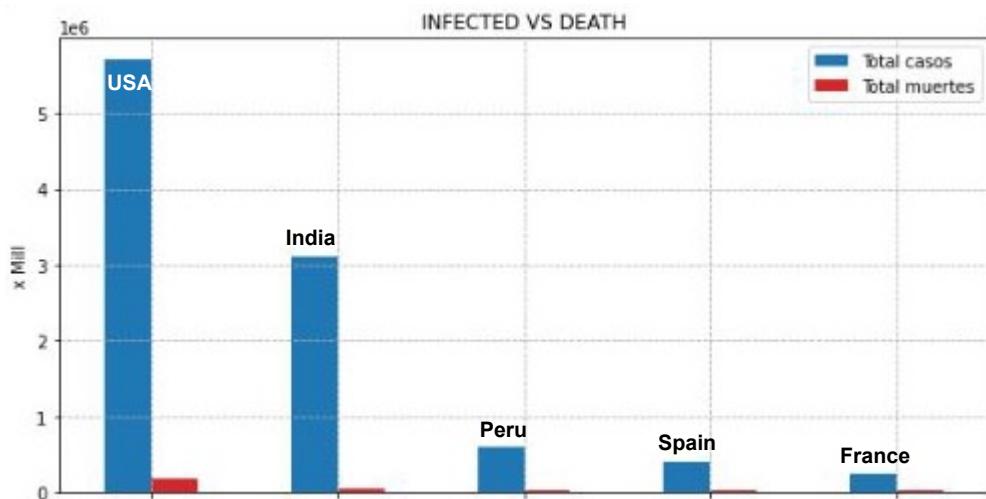
Amount of death people is related to all mentioned factors and population's mean age.



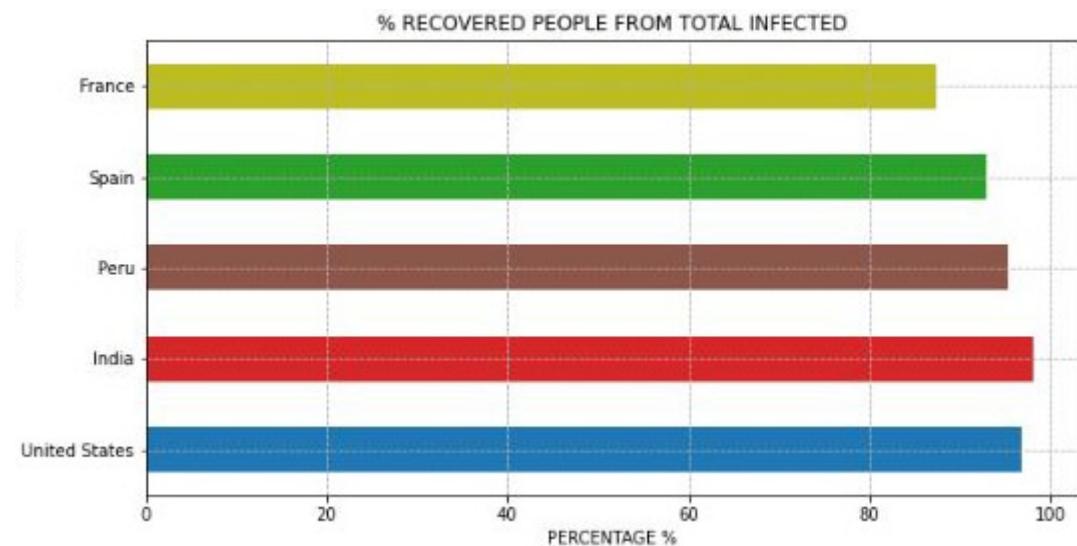
After the first wave, frequency of deaths remains stable and far from being stopped.



Death ratio among infected is low.



Percentage of infected people that recovers from disease is high.



# FRANCE

## Disease evolution.

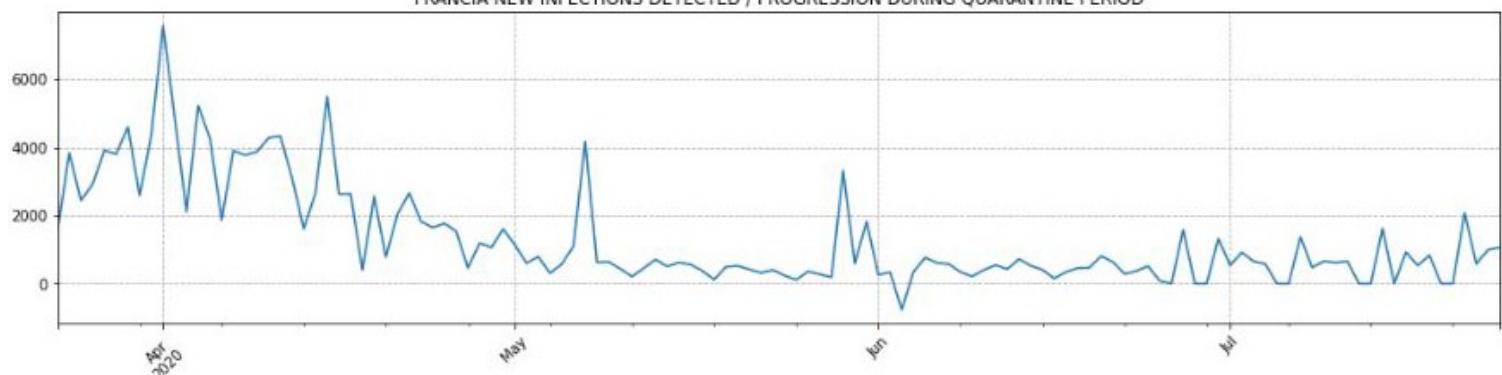
**Emergency state declared:  
March 23rd to July 10th**

Quarantine measures taken during emergency state slowed down the growing number of deaths.

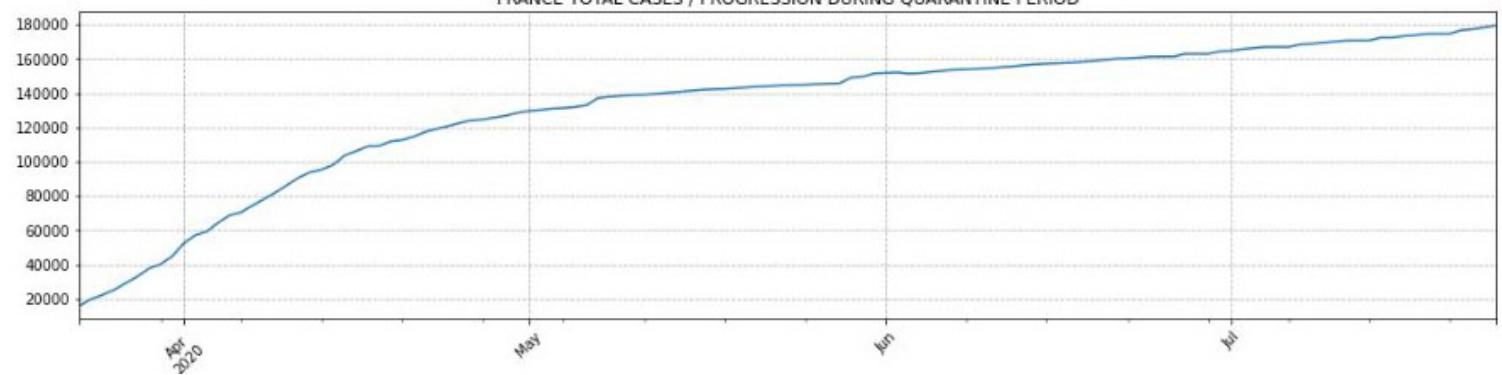
Emergency state finished after a long term without trend changes.



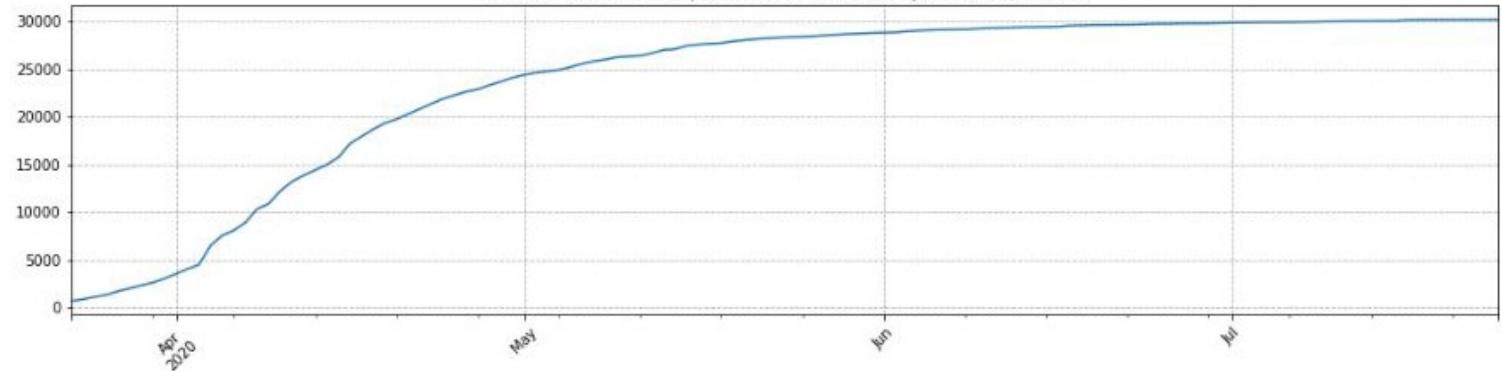
FRANCIA NEW INFECTIONS DETECTED / PROGRESSION DURING QUARANTINE PERIOD



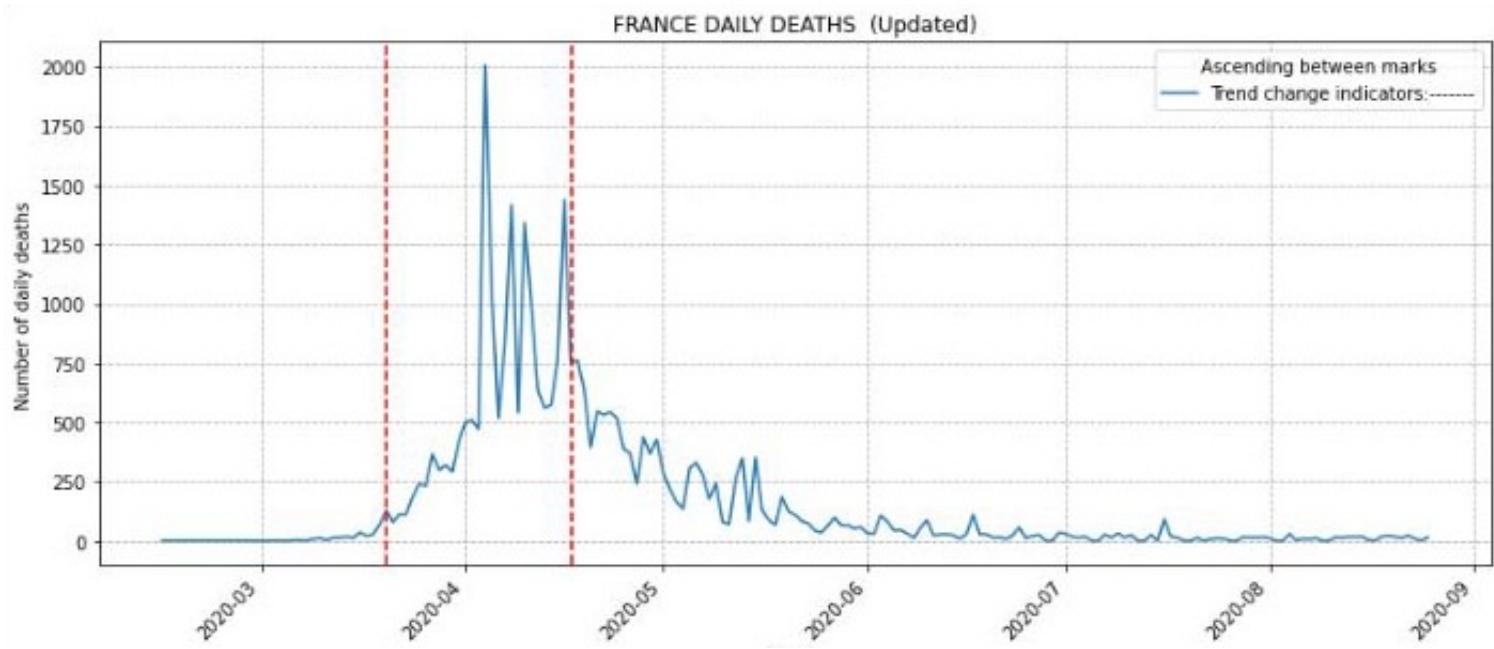
FRANCE TOTAL CASES / PROGRESSION DURING QUARANTINE PERIOD



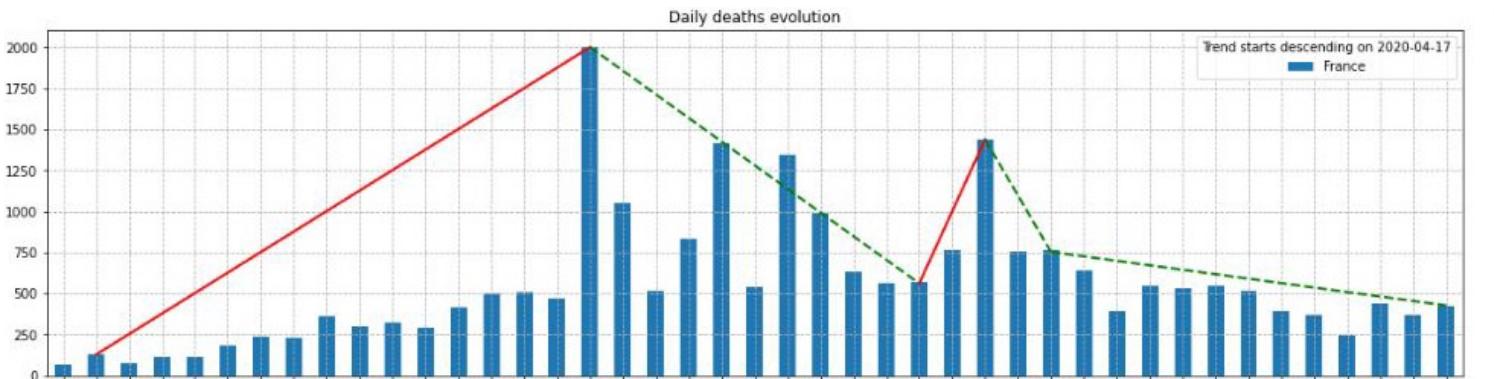
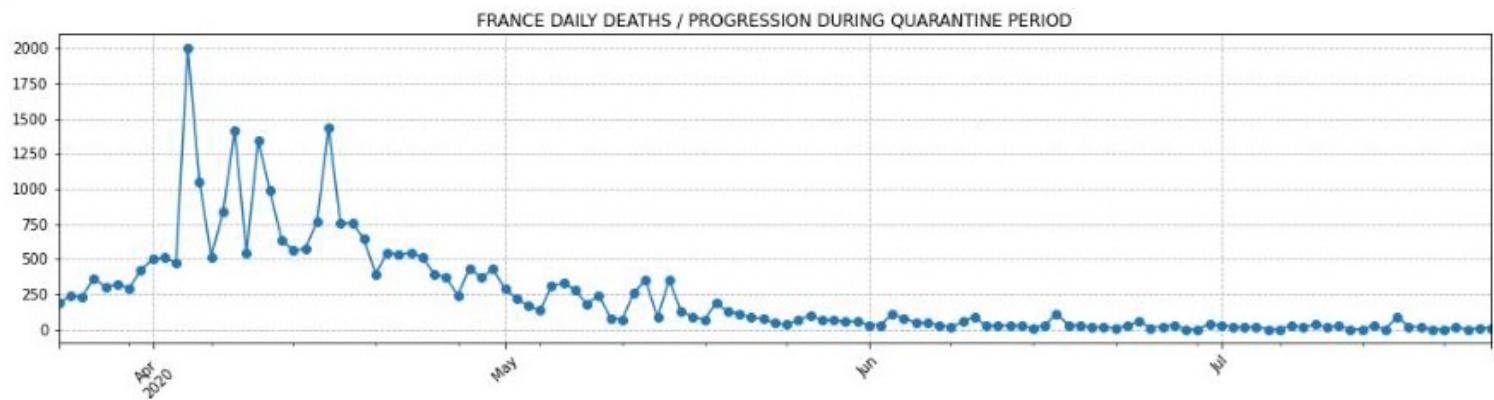
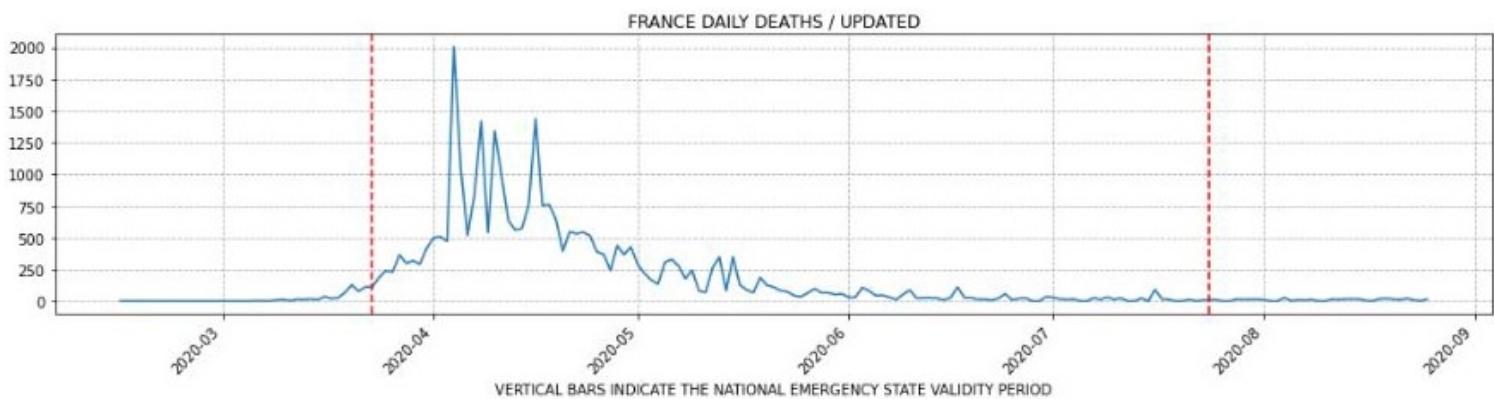
FRANCE TOTAL DEATHS / PROGRESSION DURING QUARANTINE PERIOD



# FRANCE: Death by covid evolution.



## Evolution with emergency state measures.

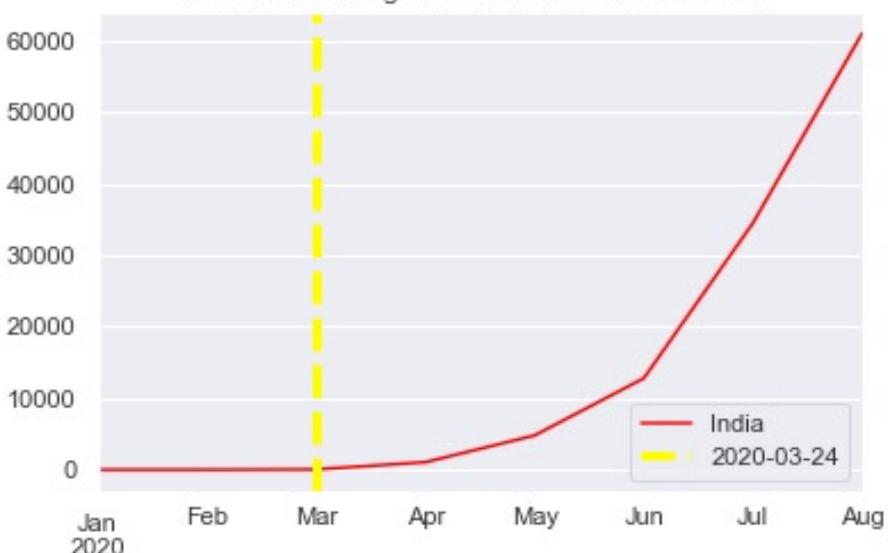


# INDIA

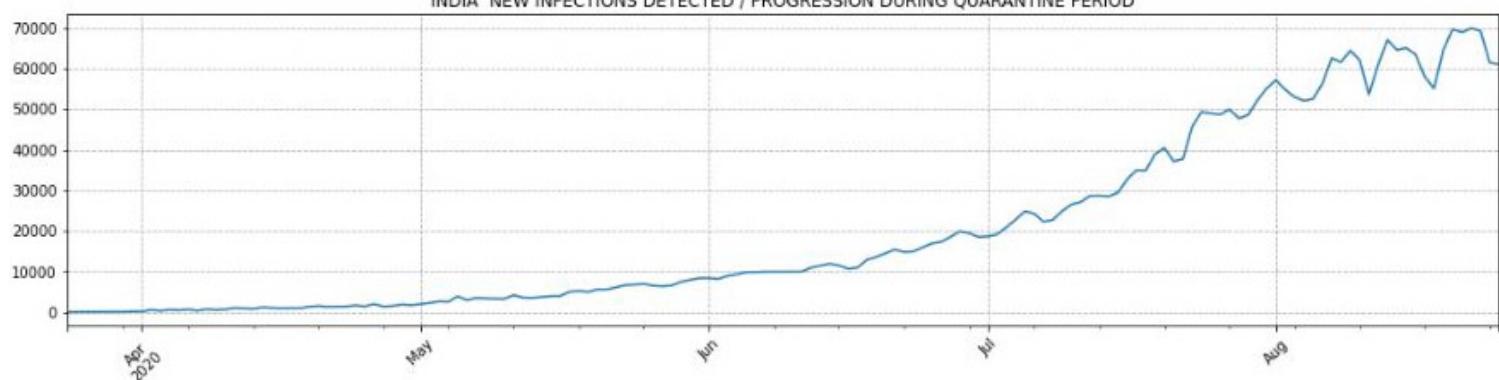
Disease evolution.  
Emergency state declared:  
March 24th - actually on

Emergency measures taken are not being useful. All numbers are growing and no changes in trends have been detected.

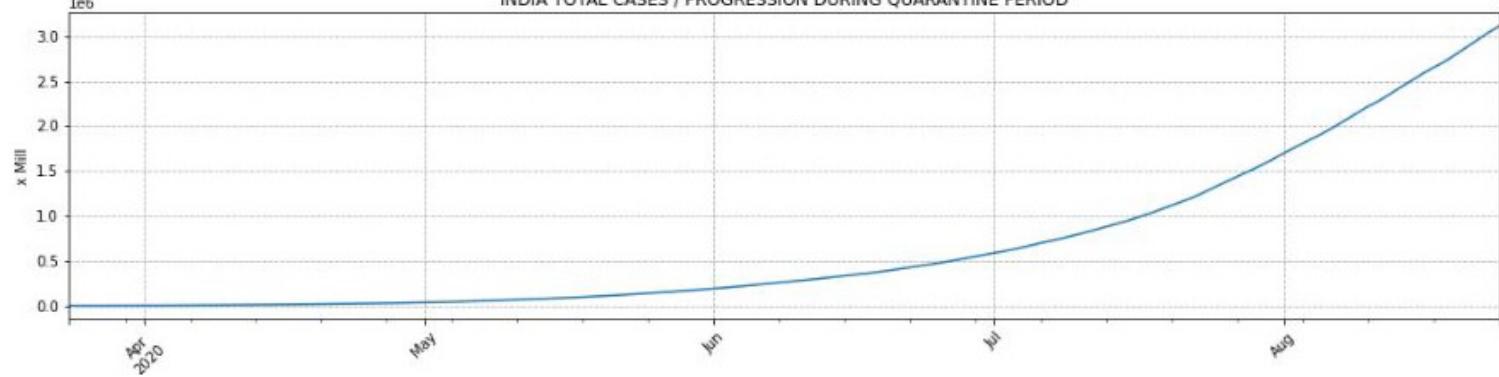
Emergency state is still valid.



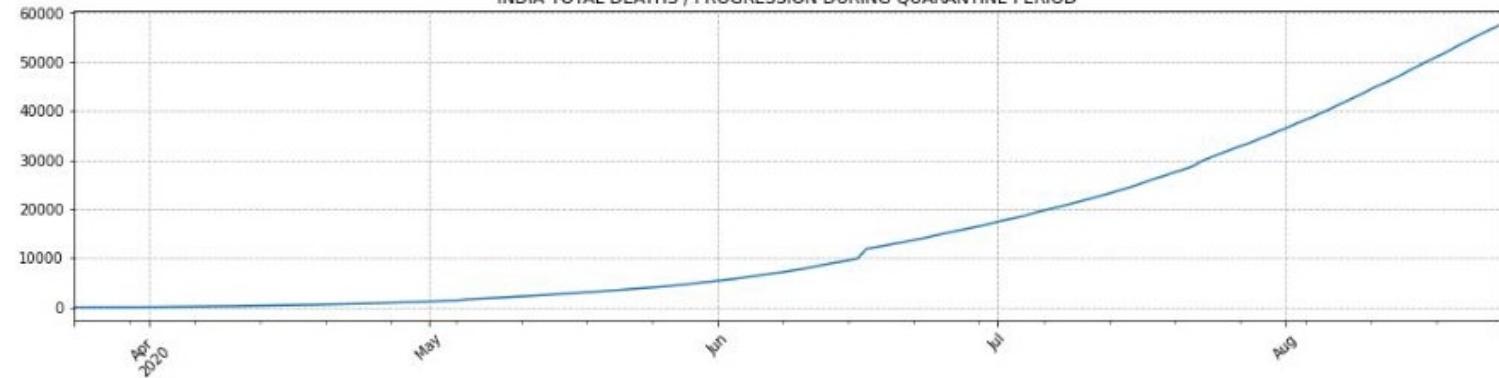
INDIA NEW INFECTIONS DETECTED / PROGRESSION DURING QUARANTINE PERIOD



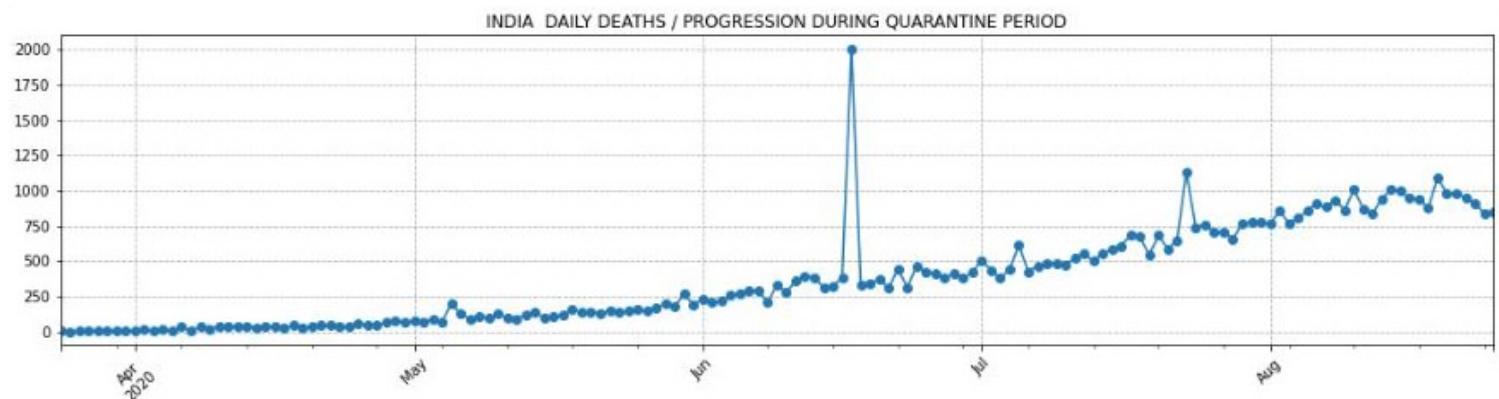
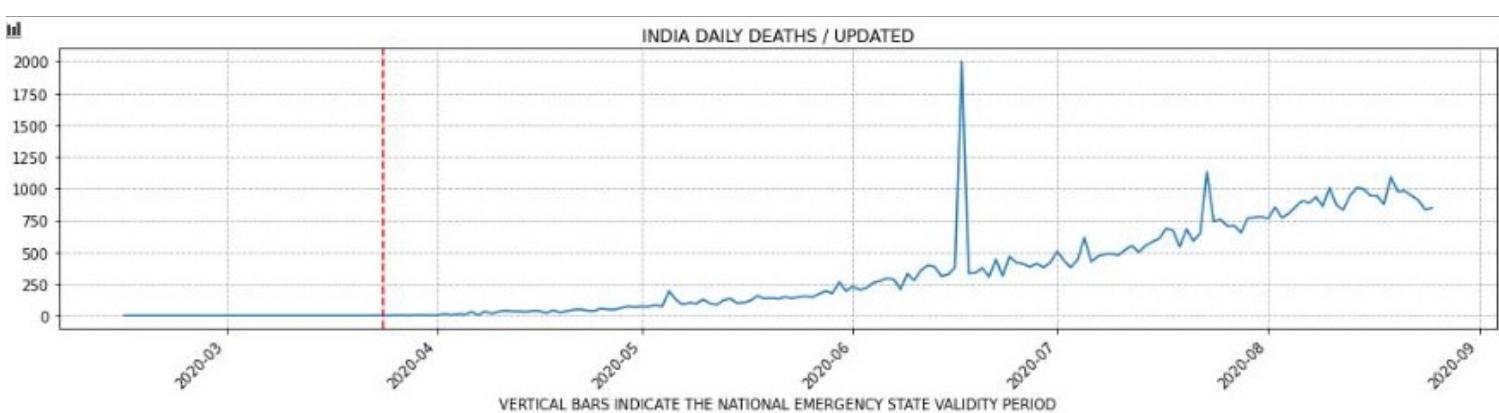
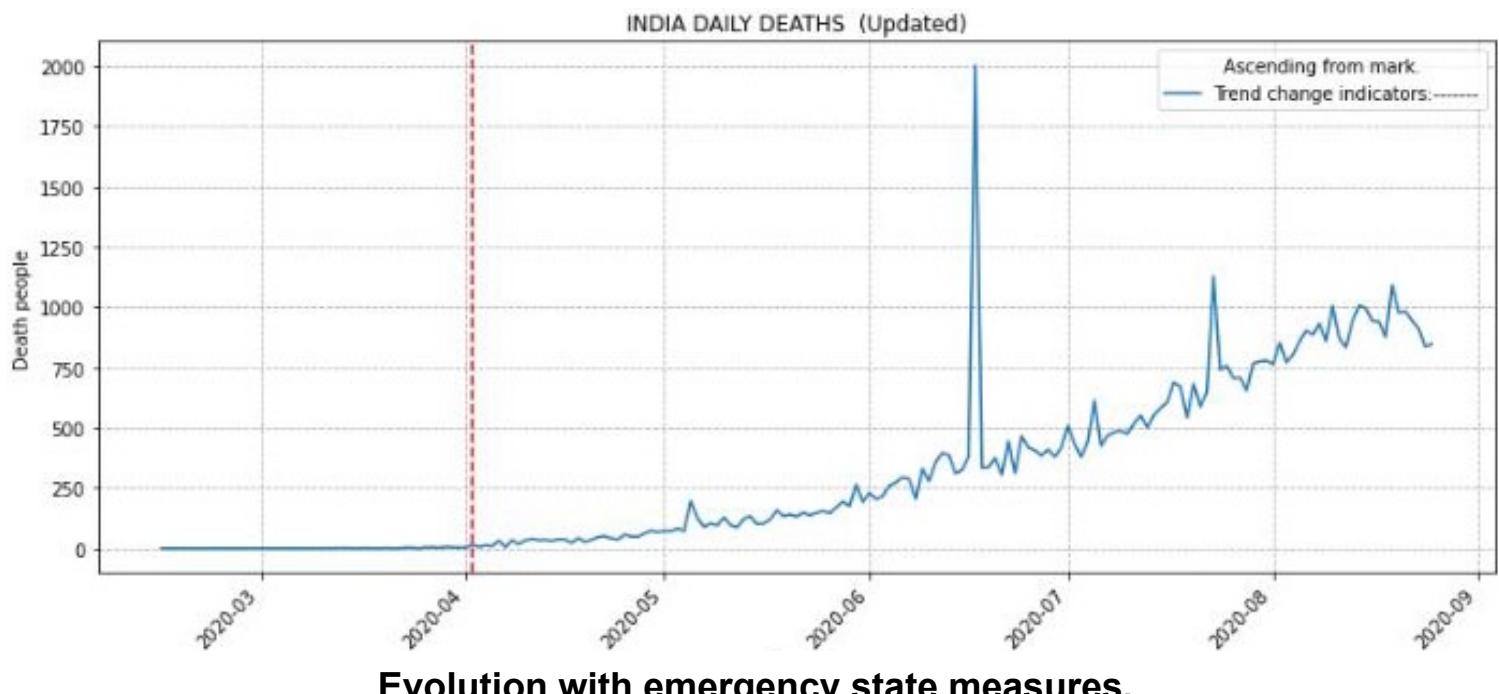
INDIA TOTAL CASES / PROGRESSION DURING QUARANTINE PERIOD



INDIA TOTAL DEATHS / PROGRESSION DURING QUARANTINE PERIOD



# INDIA : Death by covid evolution.



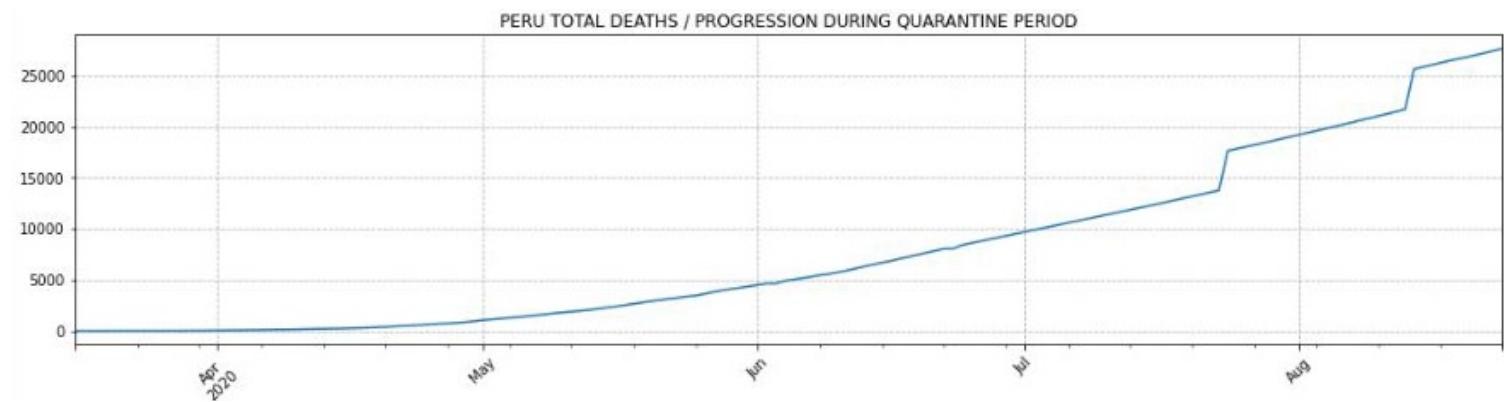
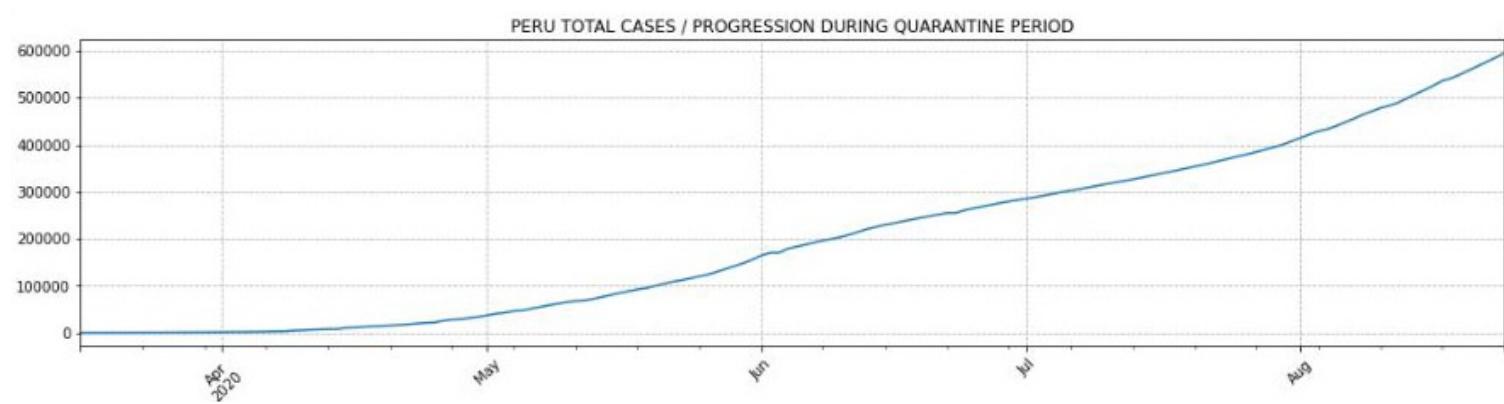
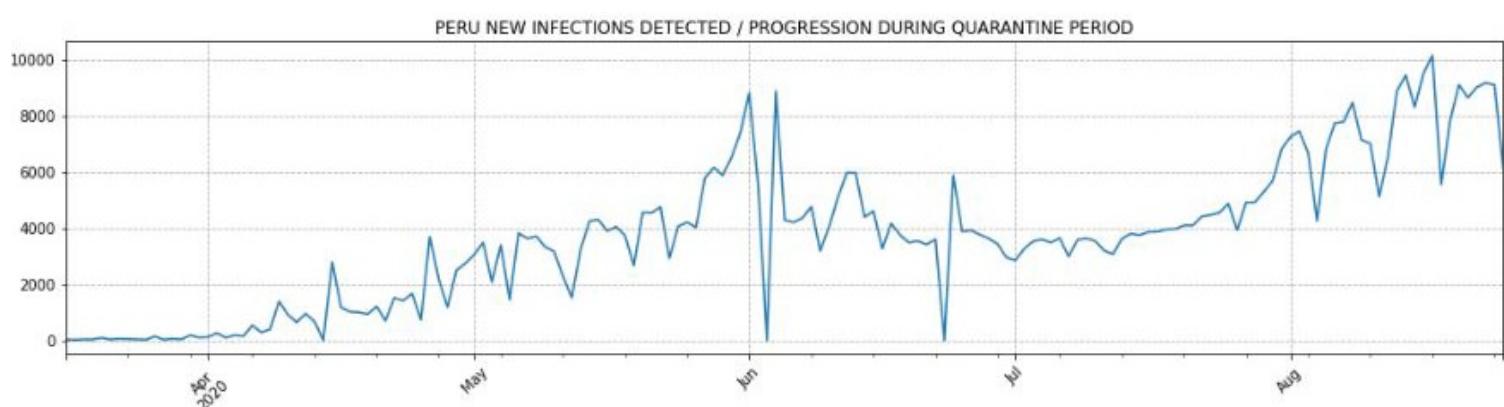
# PERU

## Disease evolution.

**Emergency state declared:  
March 15th to August 31st**

Emergency measures taken are not being useful. All numbers are growing. Growth slowed from May to July.

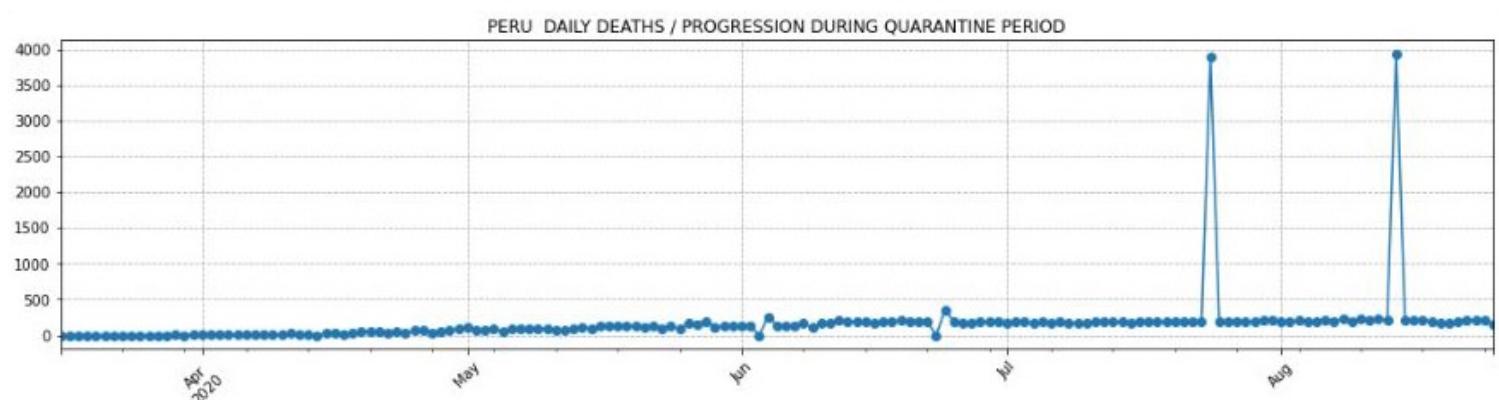
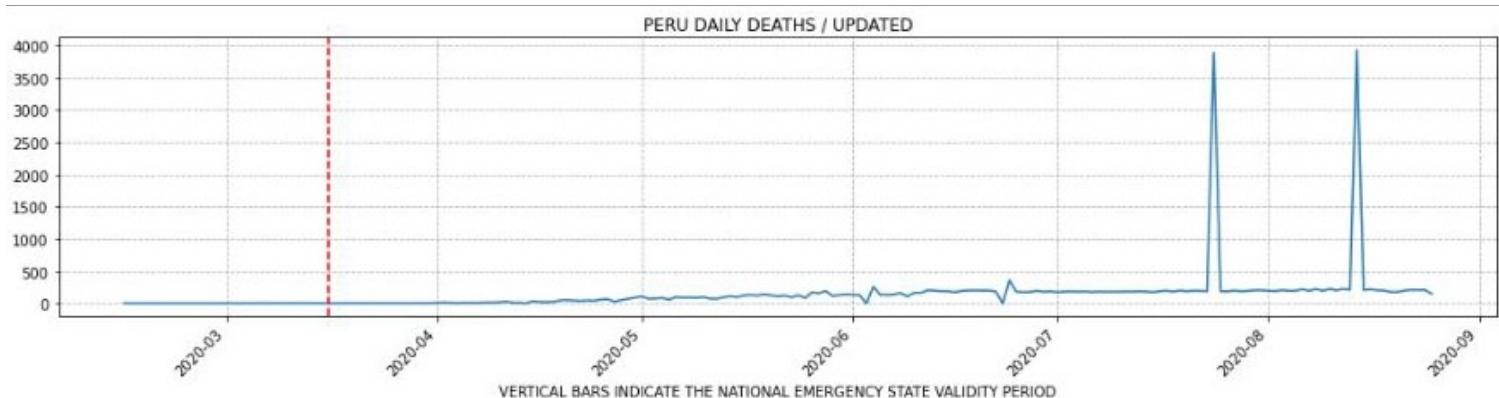
Emergency state is still valid.



# PERU : Death by covid evolution.



**Evolution with emergency state measures.**



# SPAIN

## Disease evolution.

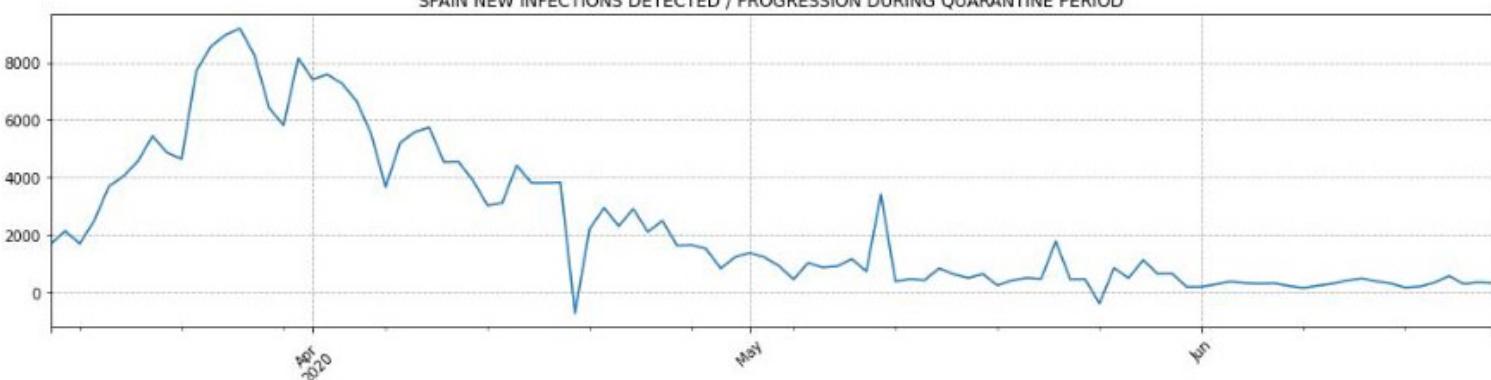
**Emergency state declared:  
March 14th to June 21st**

Quarantine measures taken during emergency state slowed down the growing number of infections.

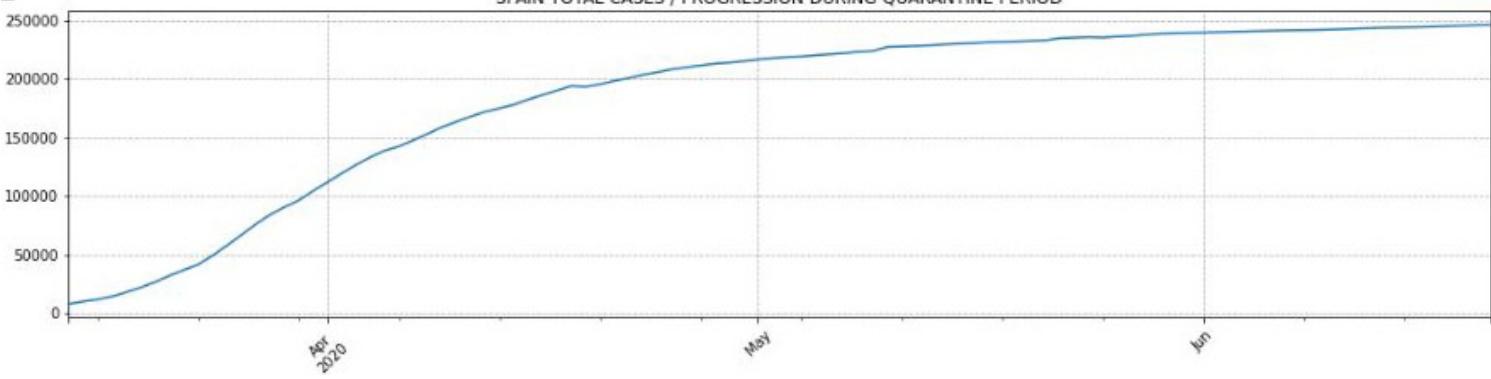
After emergency state was finished number of infections started growing. Possible new emergency declaration.



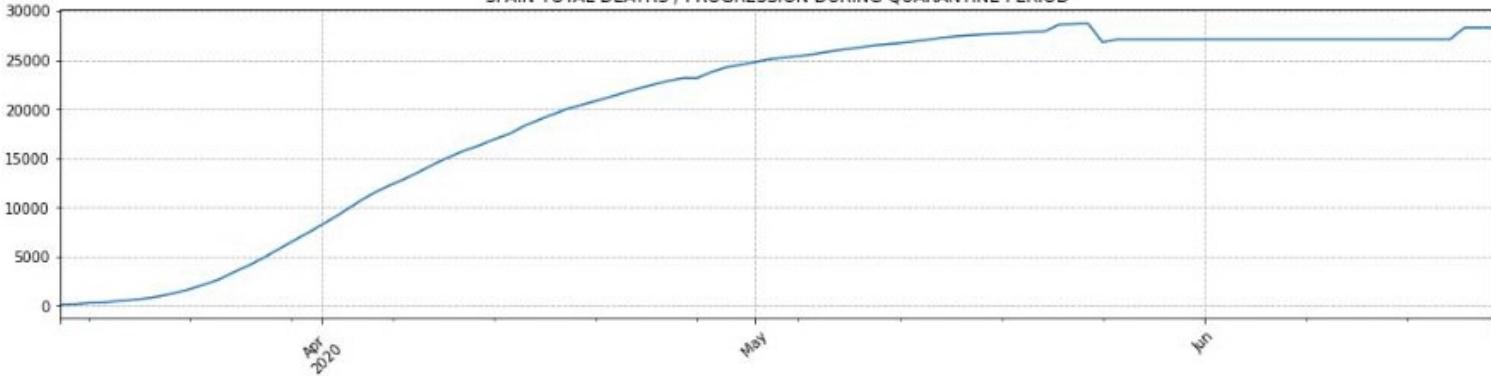
SPAIN NEW INFECTIONS DETECTED / PROGRESSION DURING QUARANTINE PERIOD



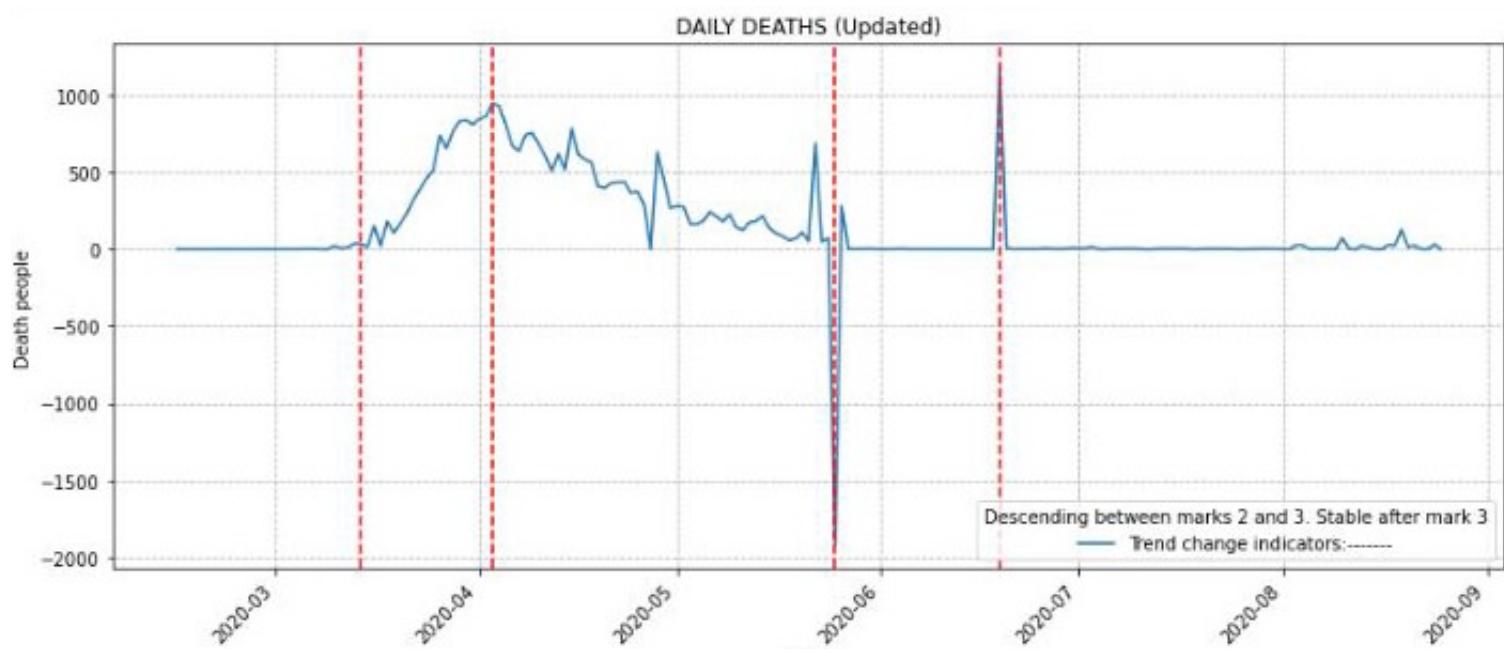
SPAIN TOTAL CASES / PROGRESSION DURING QUARANTINE PERIOD



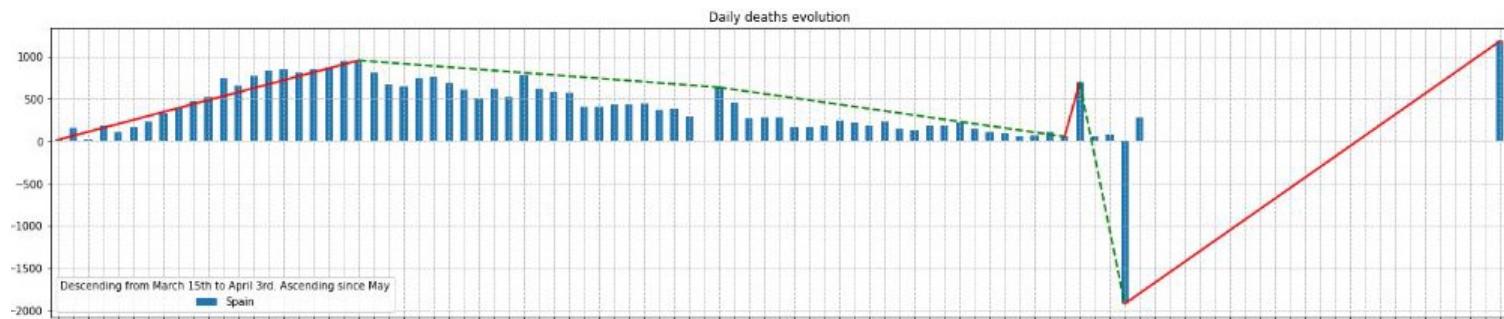
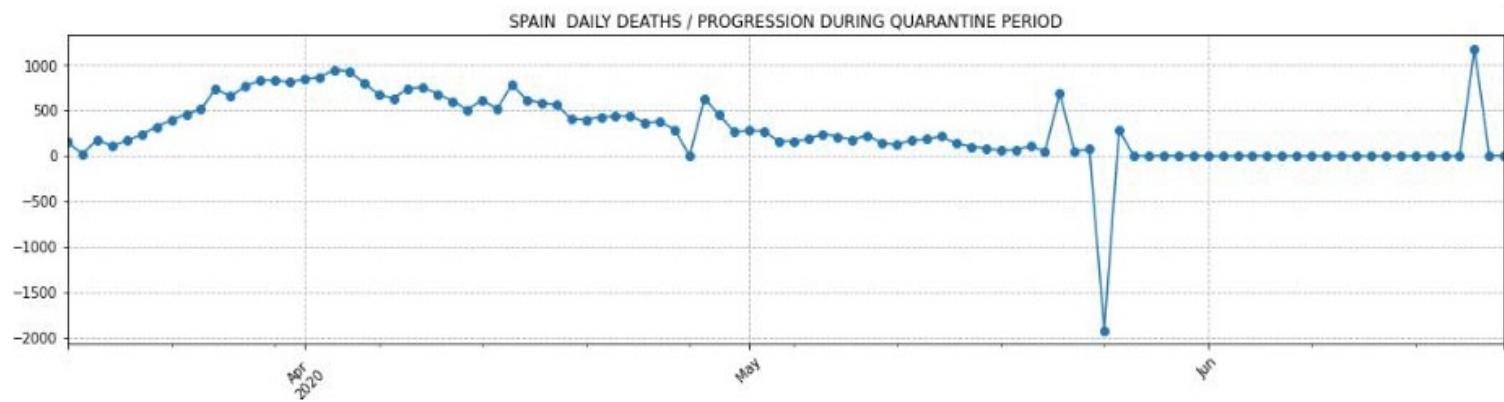
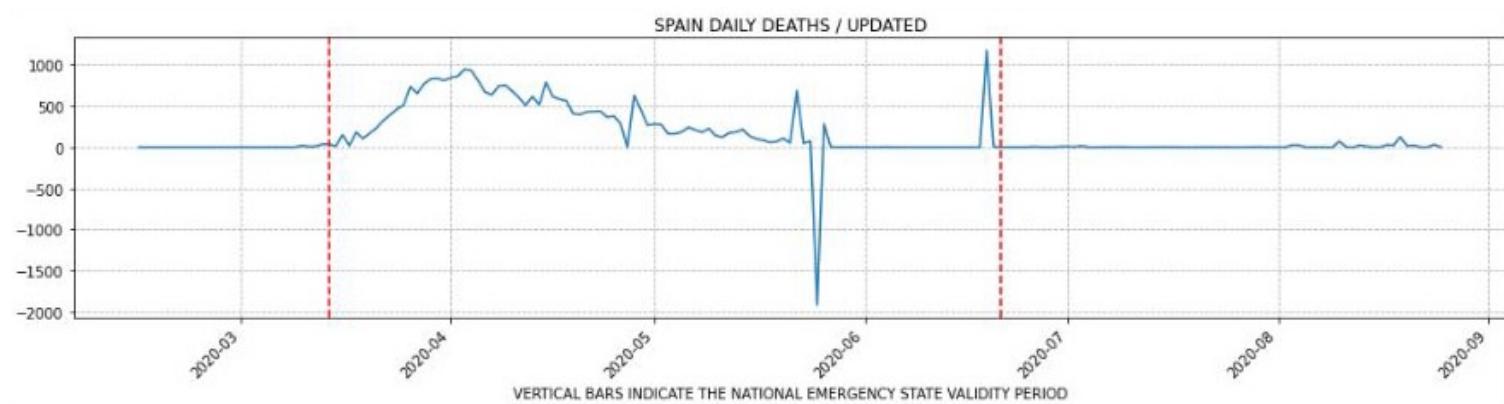
SPAIN TOTAL DEATHS / PROGRESSION DURING QUARANTINE PERIOD



# SPAIN: Death by covid evolution.



## Evolution with emergency state measures.



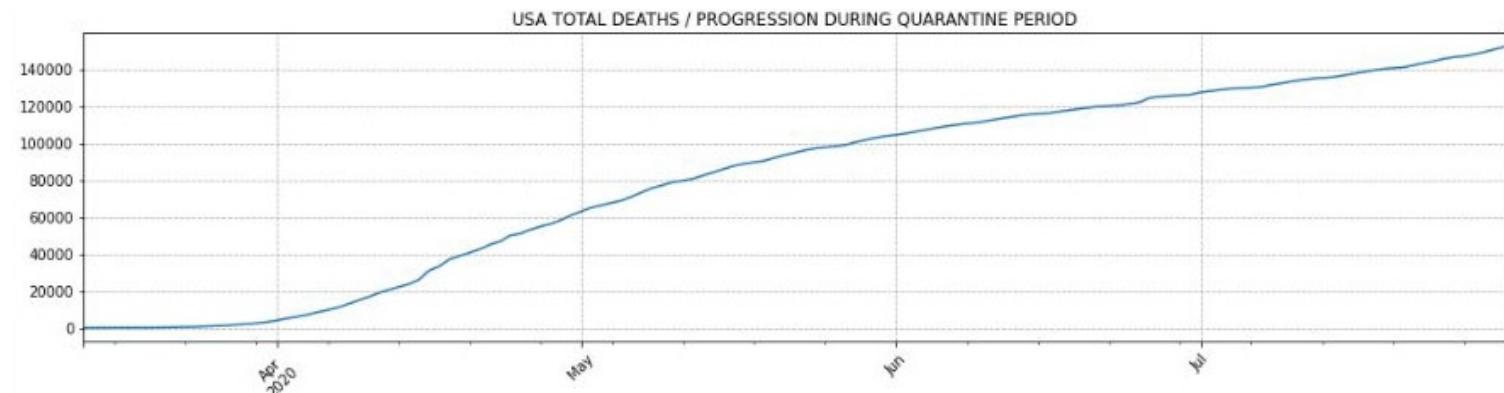
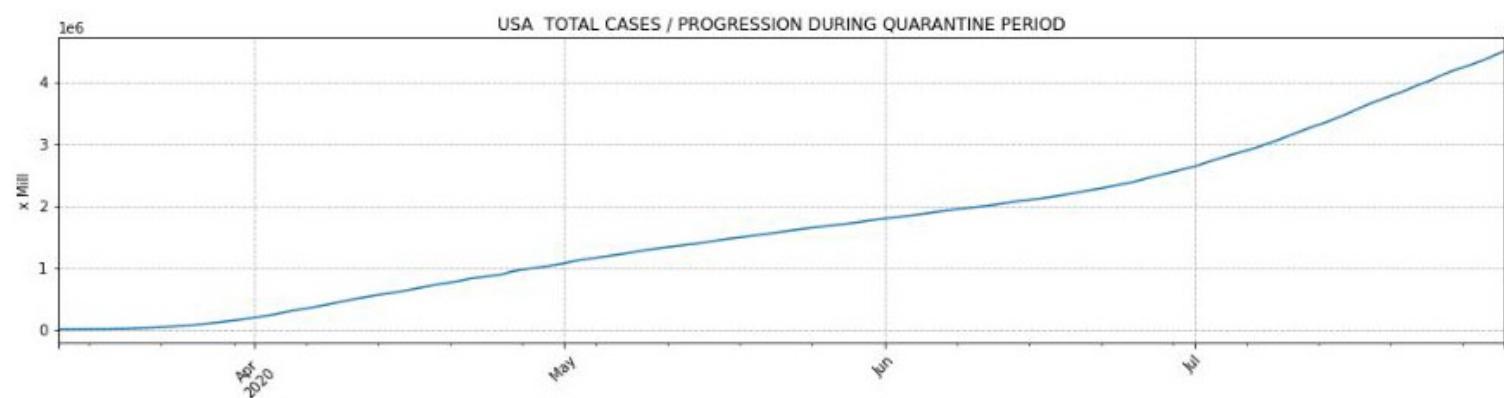
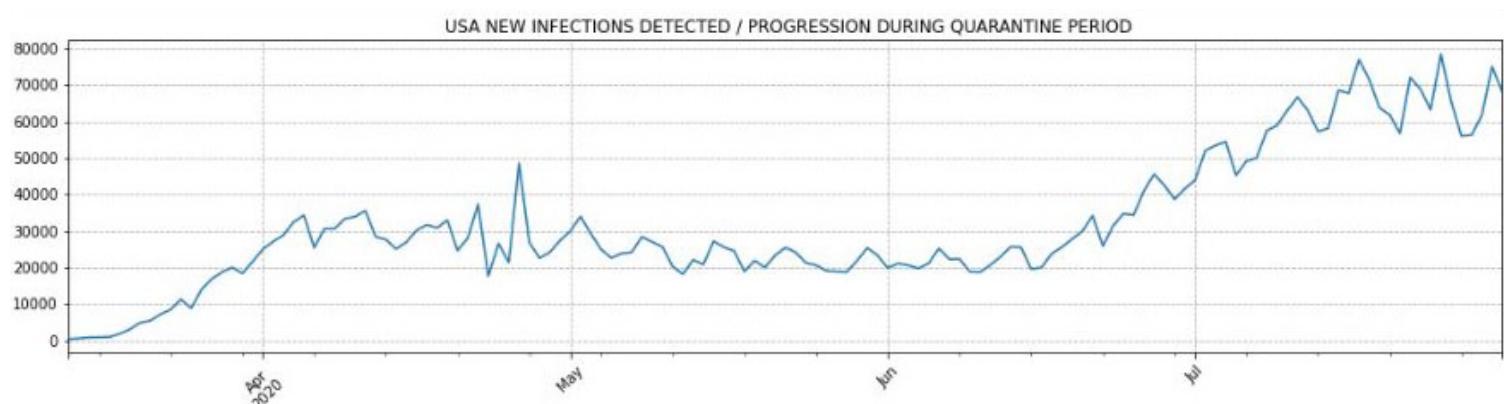
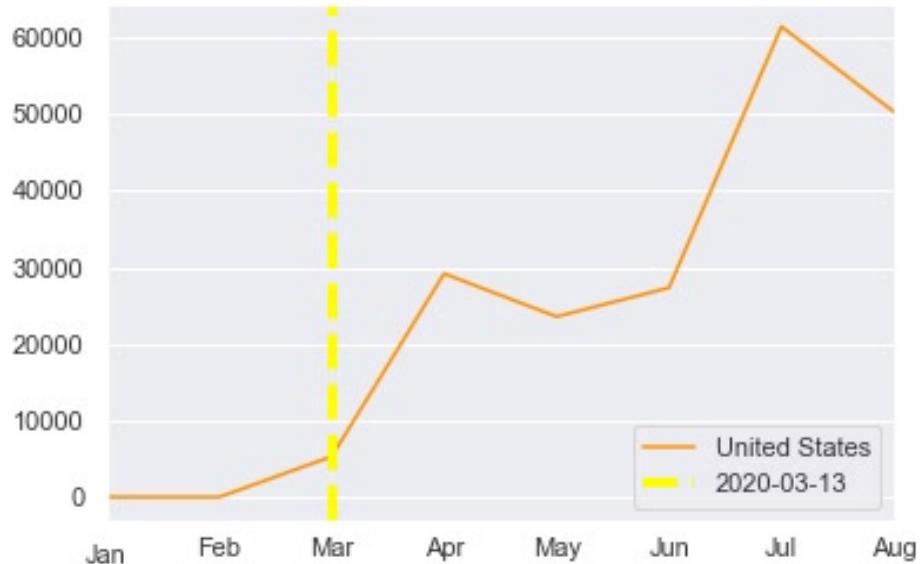
# USA

## Disease evolution.

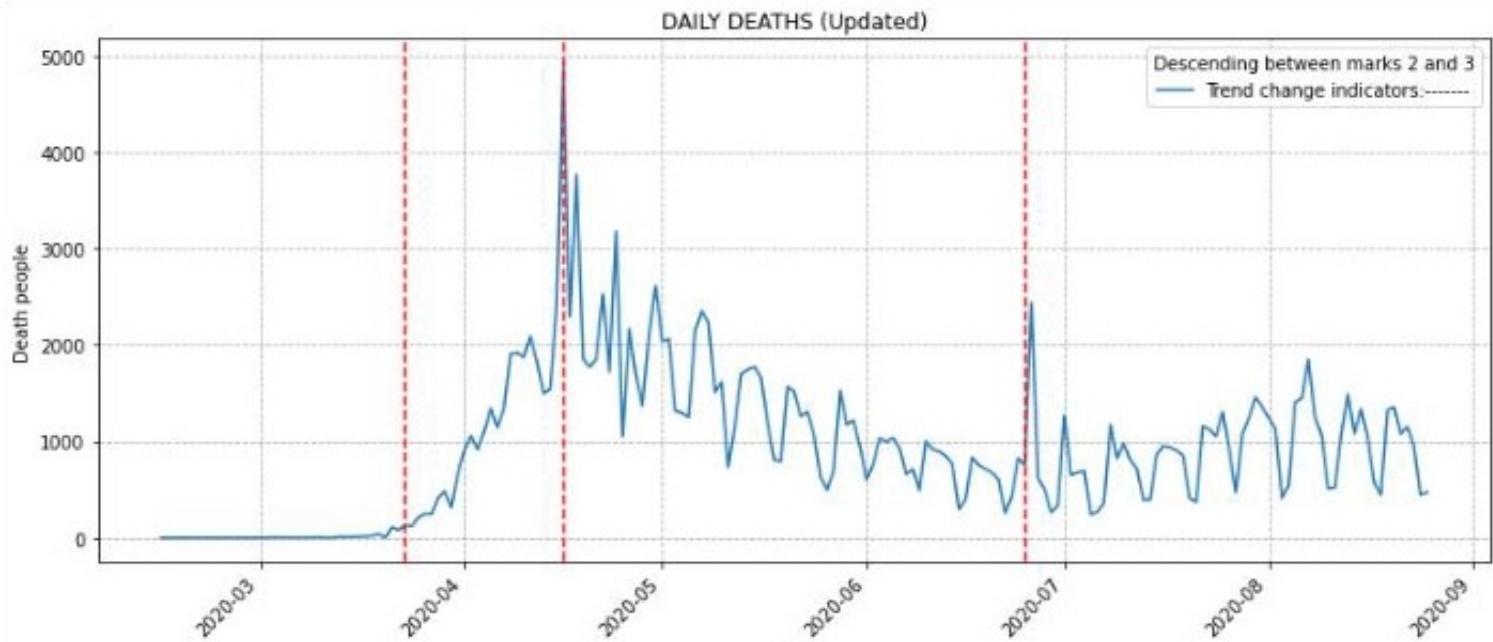
**Emergency state declared:  
March 13rd - Actually on**

Emergency measures taken are not being useful. All numbers are growing. Growth trends changed in April, May and July.

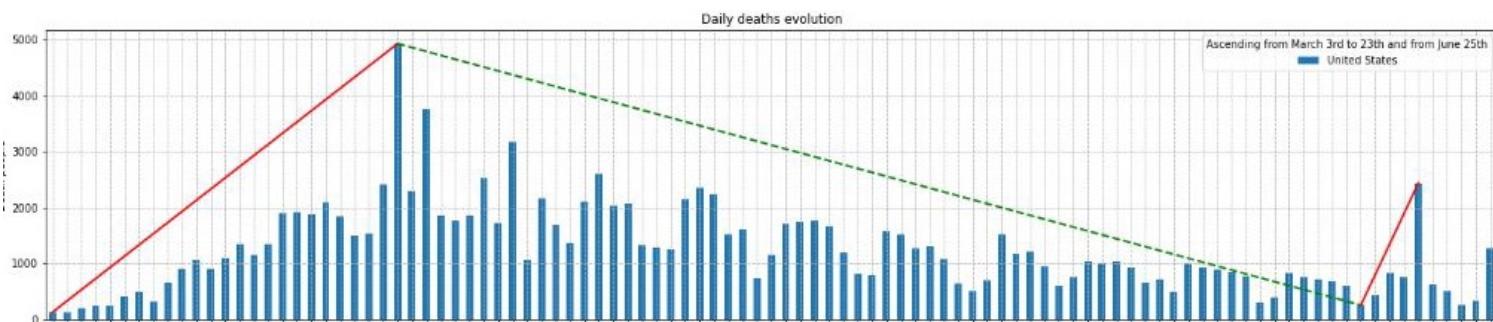
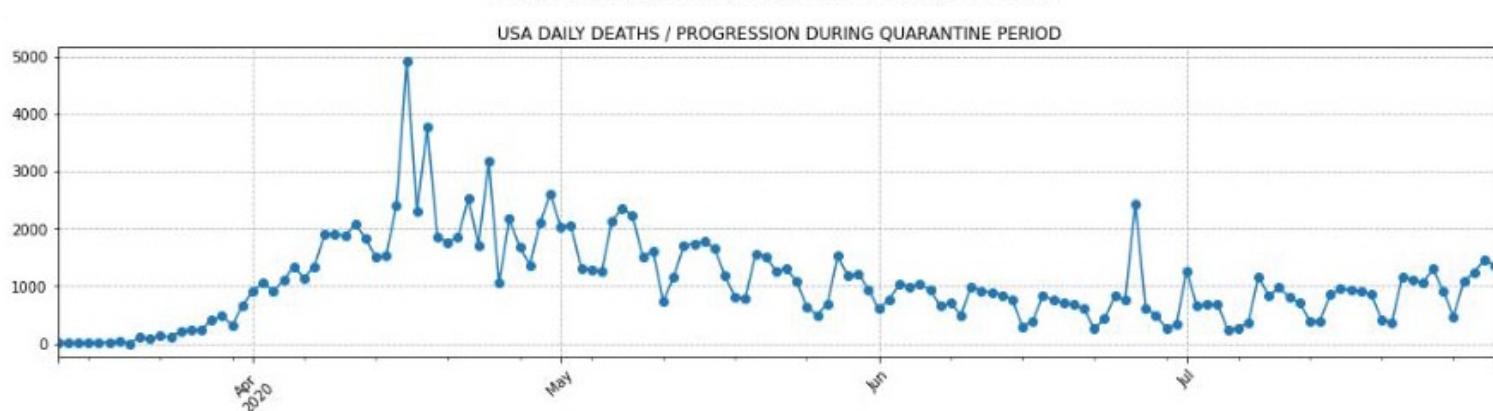
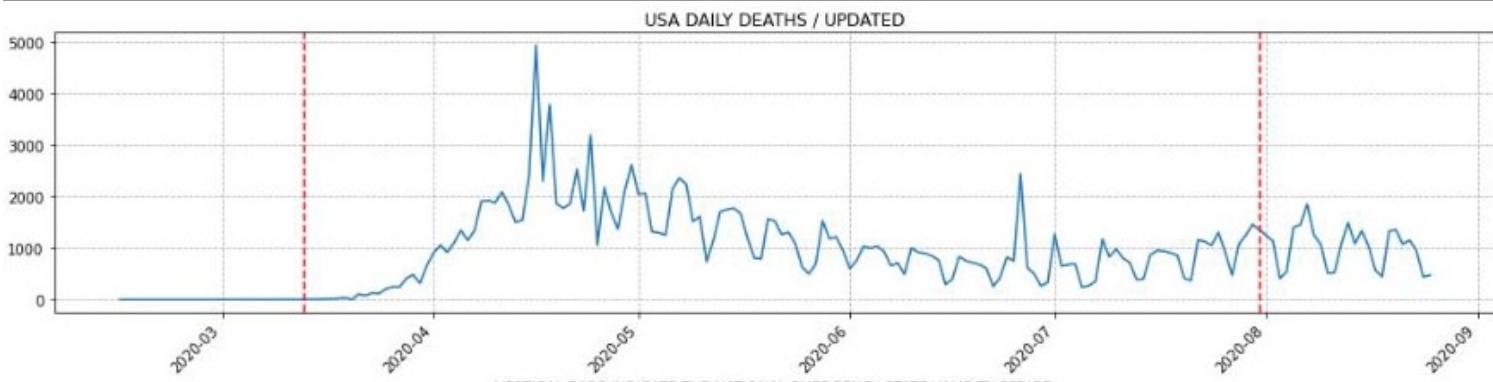
Emergency state is still valid.



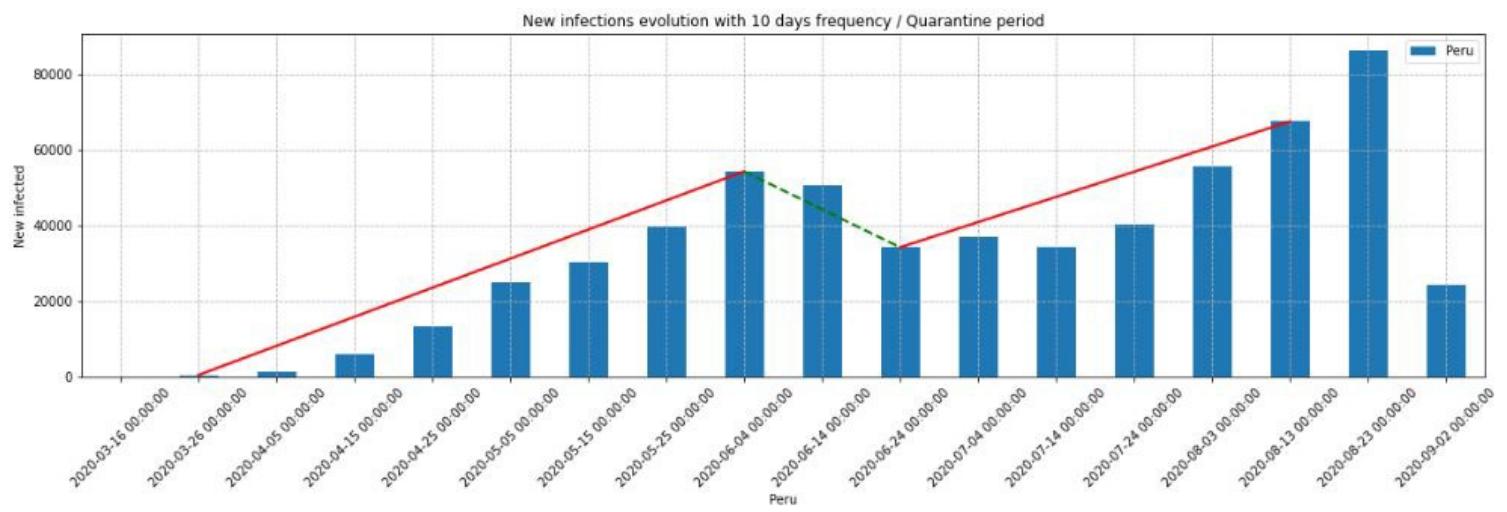
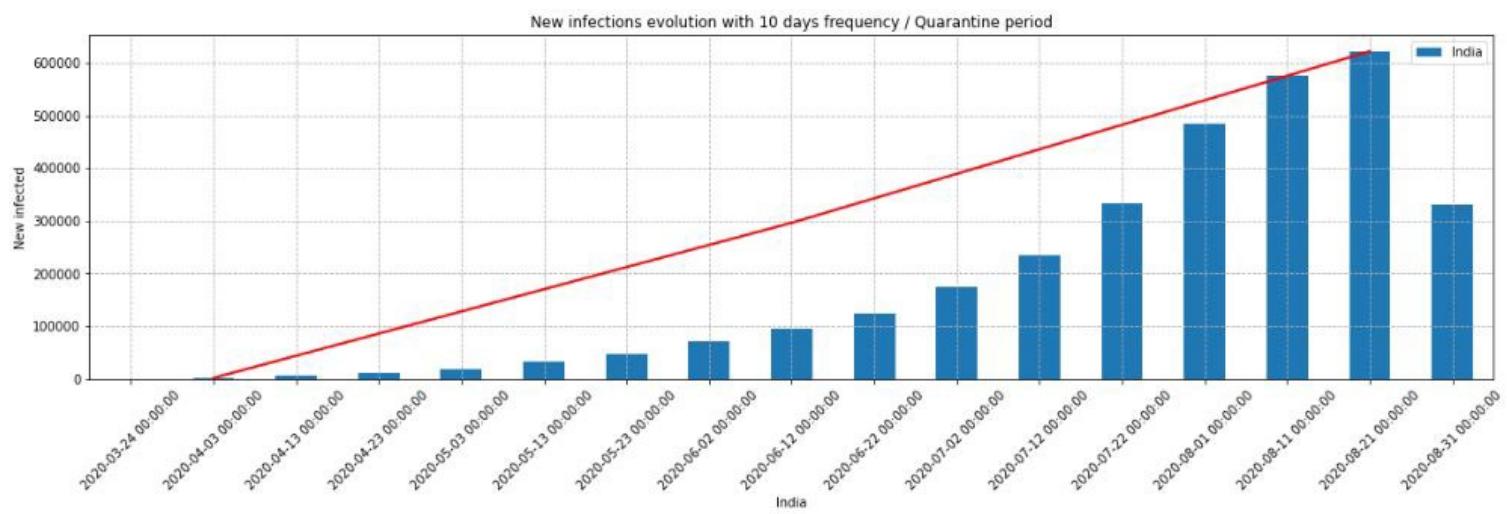
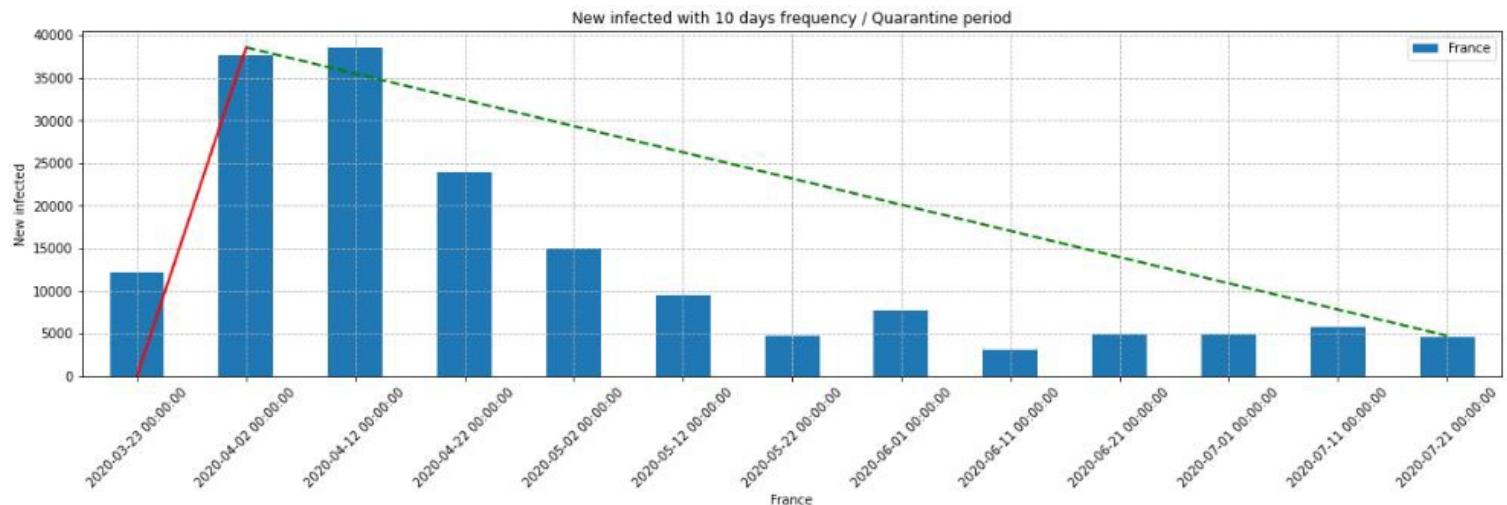
# USA: Death by covid evolution.



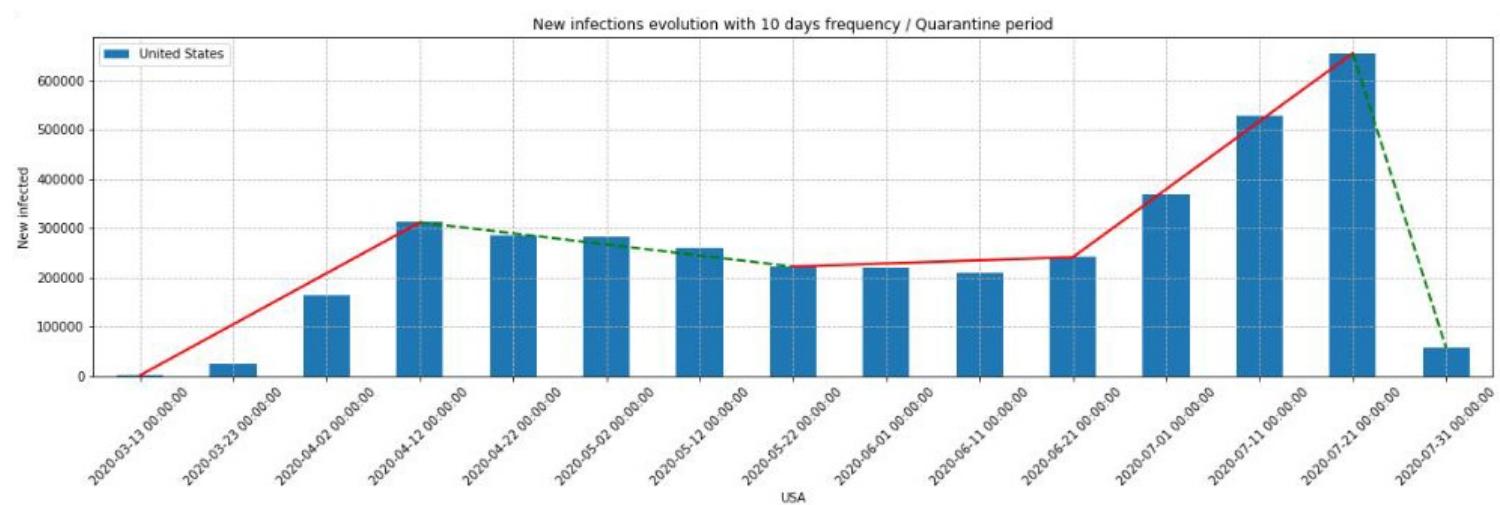
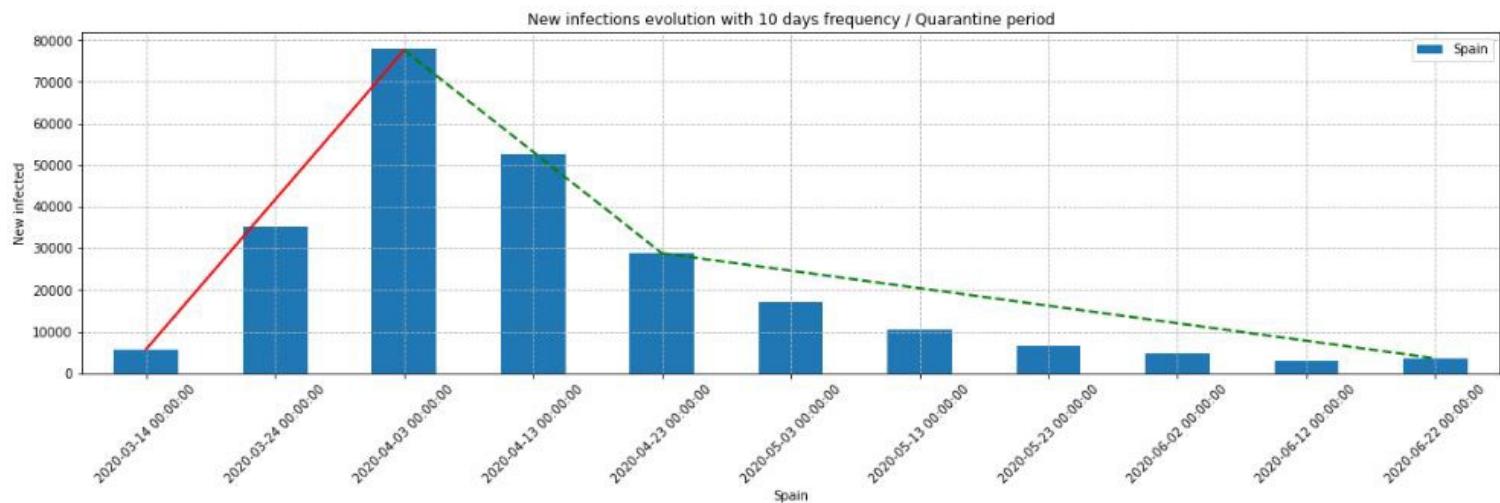
## Evolution with emergency state measures.



## BONUS: PLOTTED TENDENCES WITH 10 DAYS FRECUENCIES

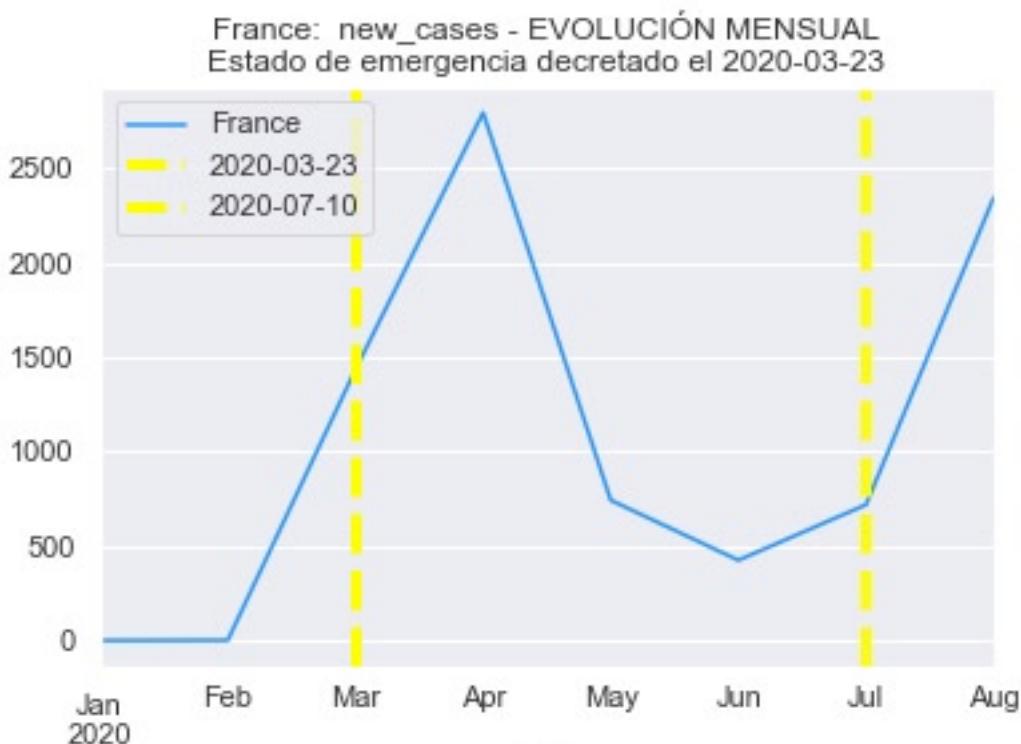


## BONUS: PLOTTED TENDENCES WITH 10 DAYS FRECUENCIES



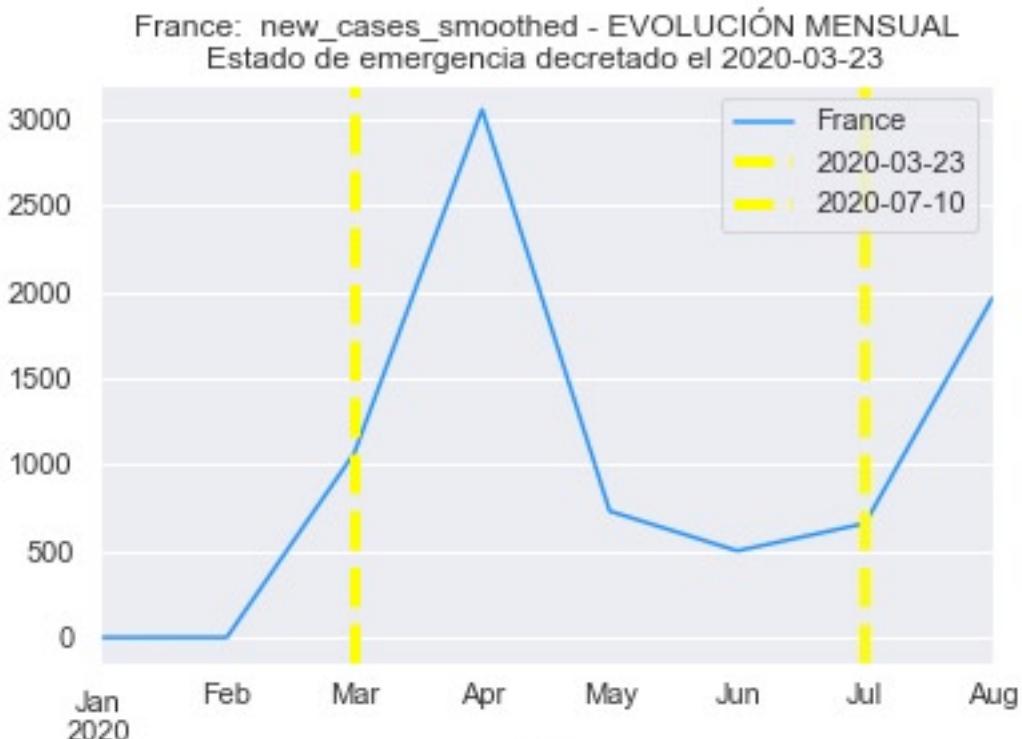
## COLLECTION OF SAVED PLOTS

### NEW DAILY INFECTIONS REGISTERED IN FRANCE (Updated 2020-08-23)



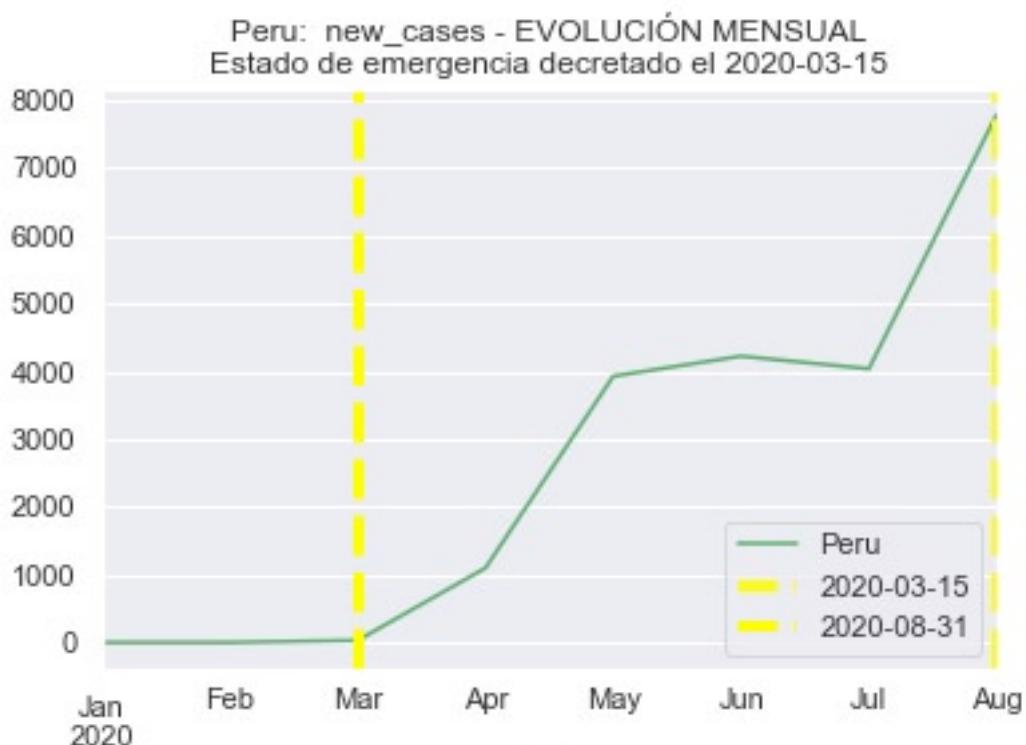
Amount of new cases reaches 2.600

### SMOOTHED NEW DAILY INFECTIONS REGISTERED IN FRANCE



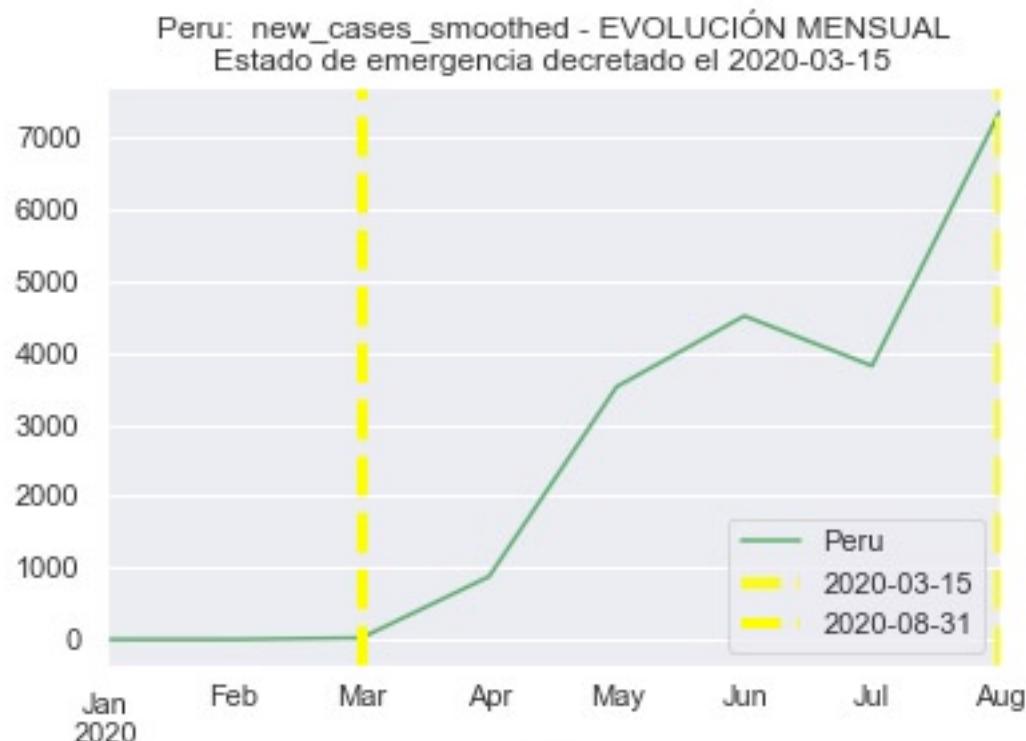
## COLLECTION OF SAVED PLOTS

### NEW DAILY INFECTIONS REGISTERED IN PERÚ (Updated 2020-08-23)



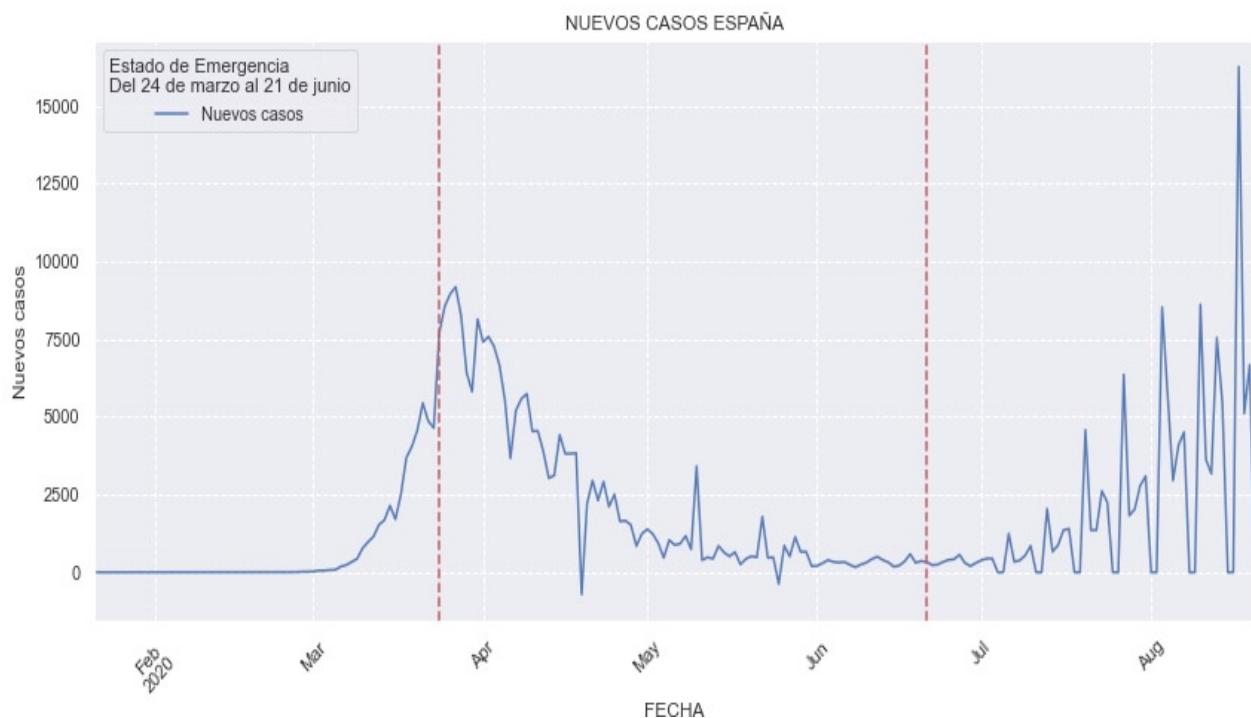
Amount of new cases reaches 8.000

### SMOOTHED NEW DAILY INFECTIONS REGISTERED IN PERÚ



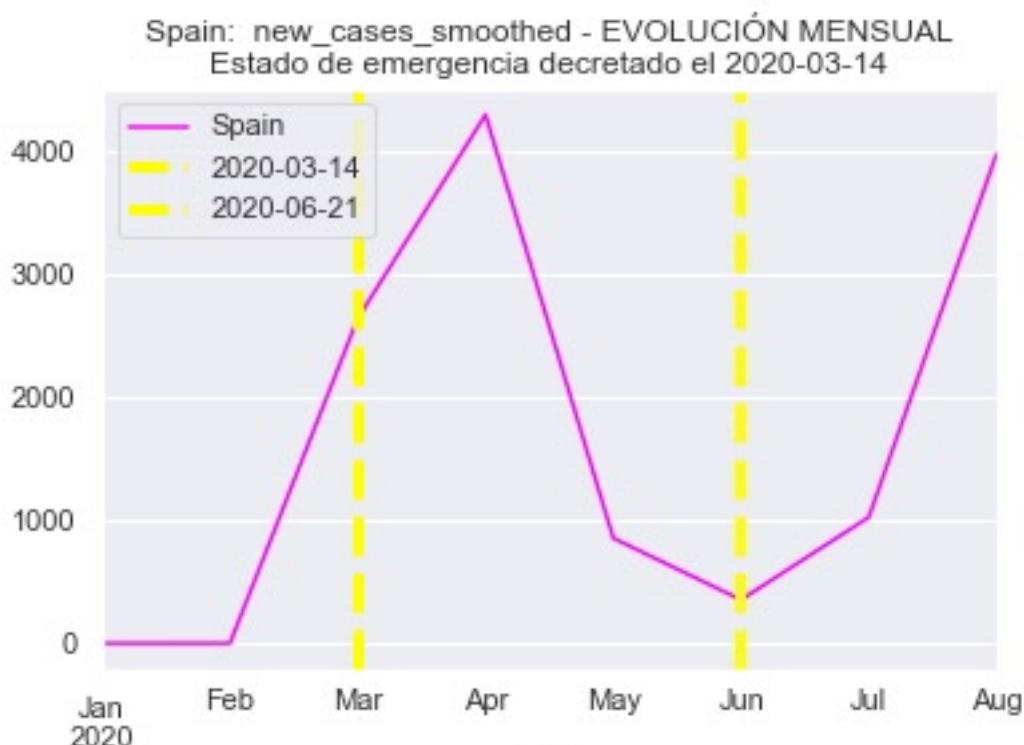
## COLLECTION OF SAVED PLOTS

### NEW DAILY INFECTIONS REGISTERED IN SPAIN (Updated 2020-08-23)



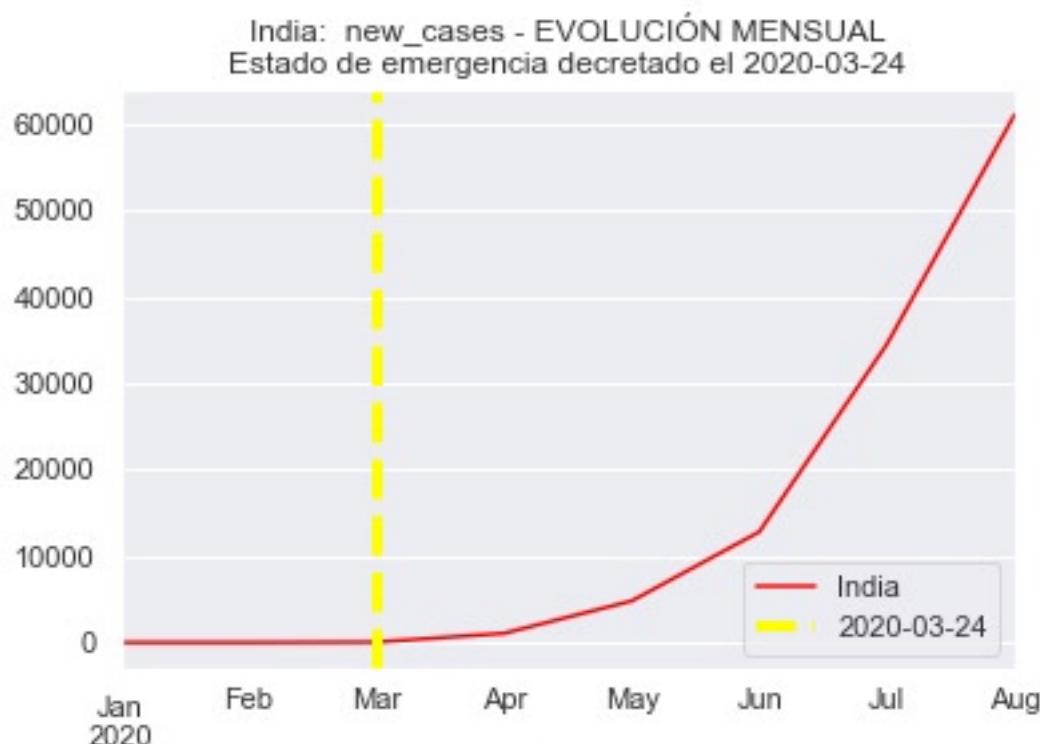
Amount of new cases reaches 15.000

### SMOOTHED NEW DAILY INFECTIONS REGISTERED IN SPAIN



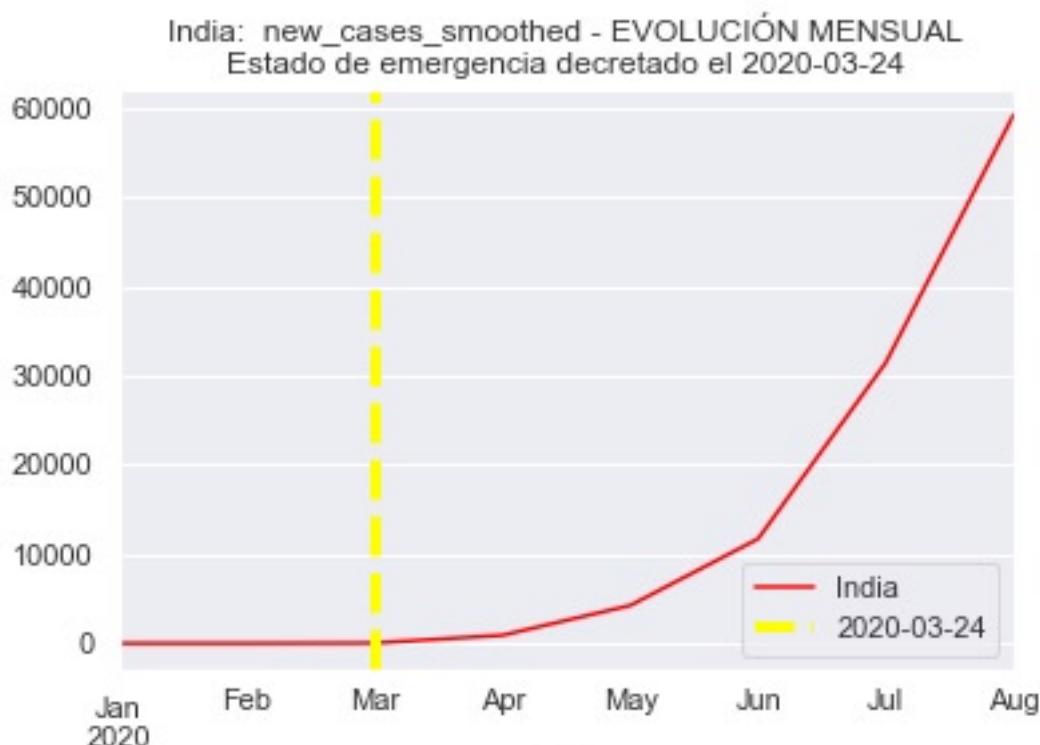
## COLLECTION OF SAVED PLOTS

### NEW DAILY INFECTIONS REGISTERED IN INDIA (Updated 2020-08-23)



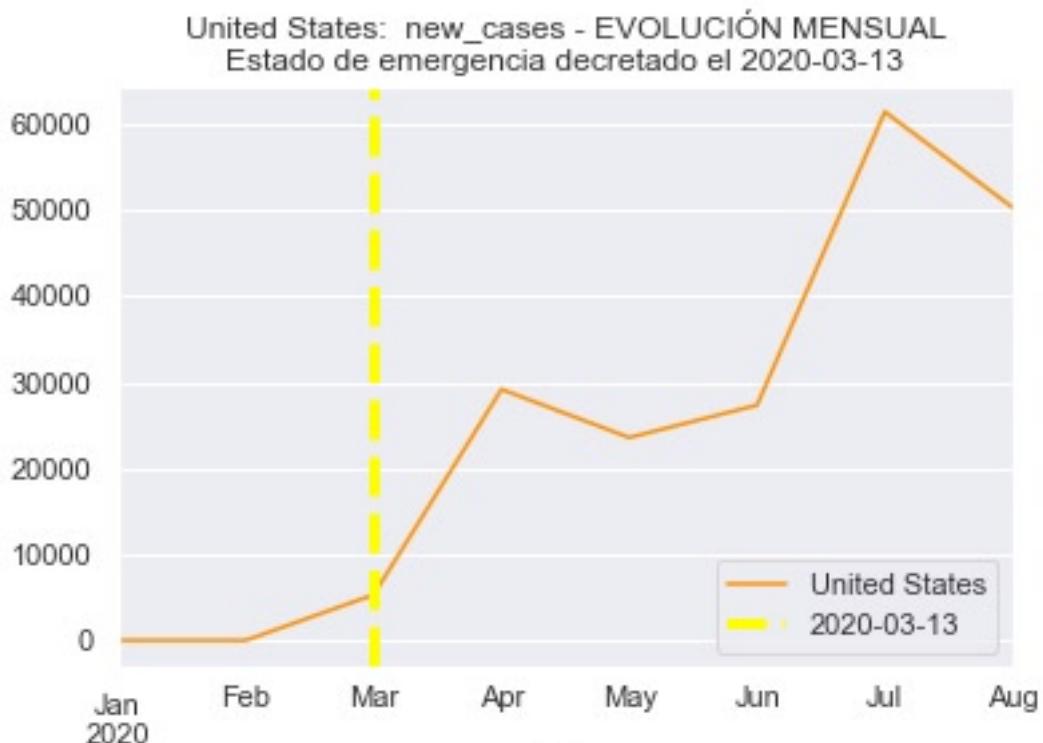
Amount of new cases reaches 60.000

### SMOOTHED NEW DAILY INFECTIONS REGISTERED IN INDIA



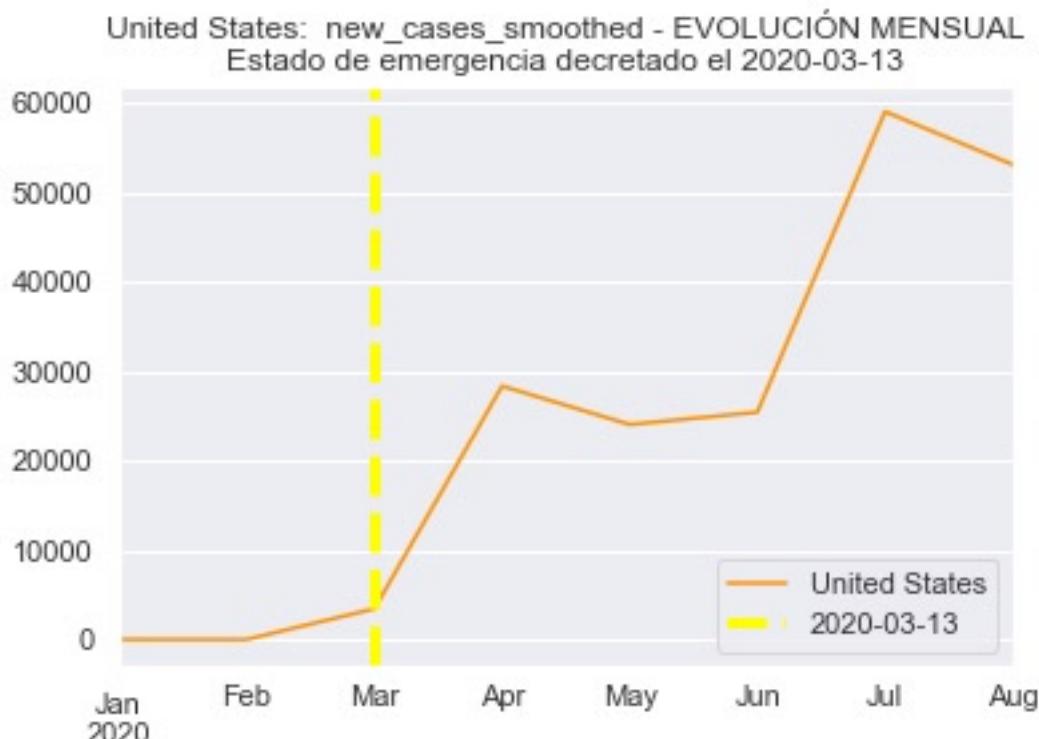
## COLLECTION OF SAVED PLOTS

### NEW DAILY INFECTIONS REGISTERED IN USA (Updated 2020-08-23)



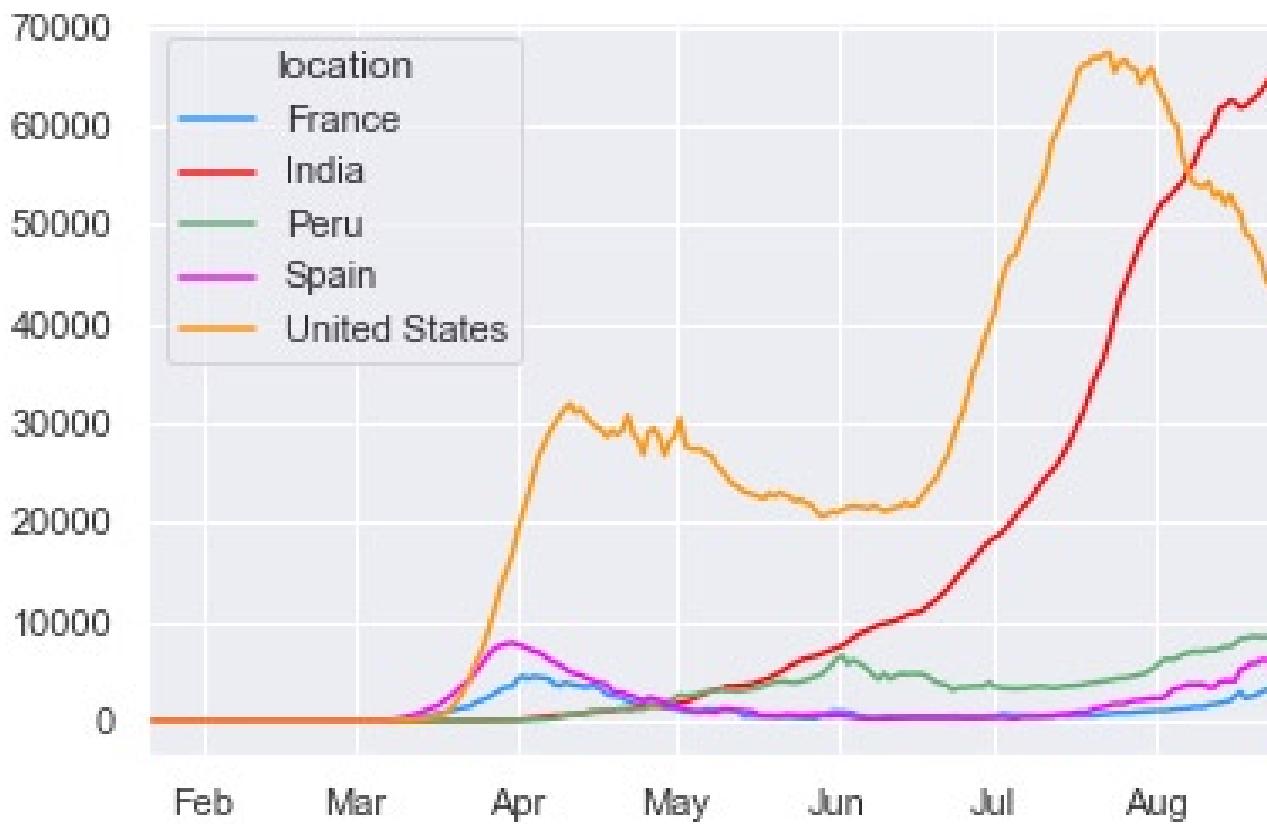
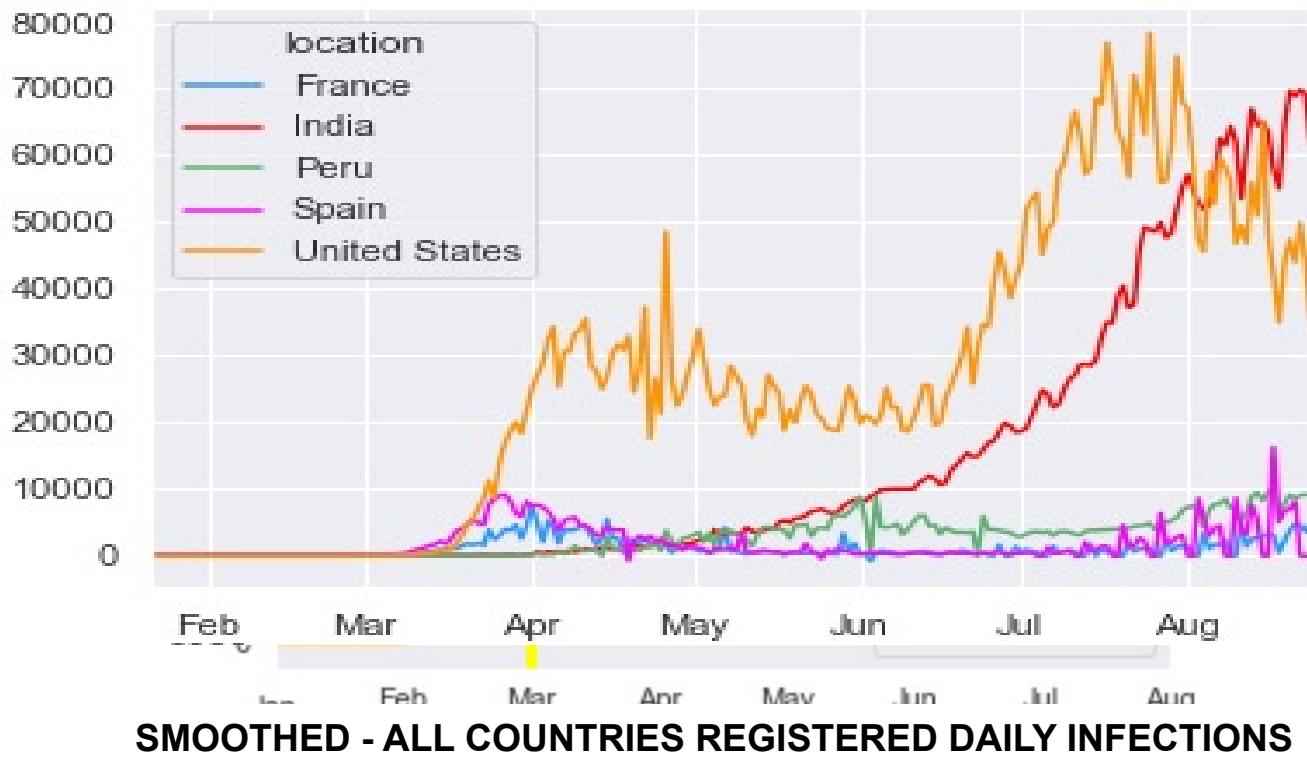
Amount of new cases (y axis) reaches 15.000

### SMOOTHED NEW DAILY INFECTIONS REGISTERED IN USA



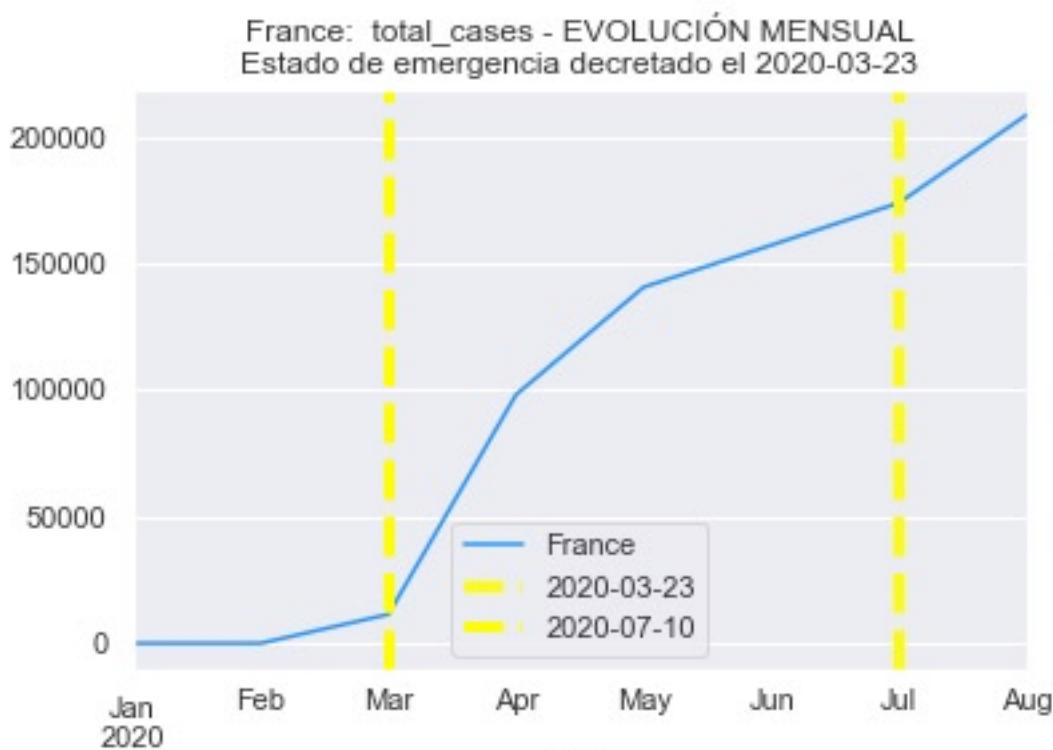
## COLLECTION OF SAVED PLOTS

### ALL COUNTRIES REGISTERED DAILY INFECTIONS (Updated 2020-08-23)



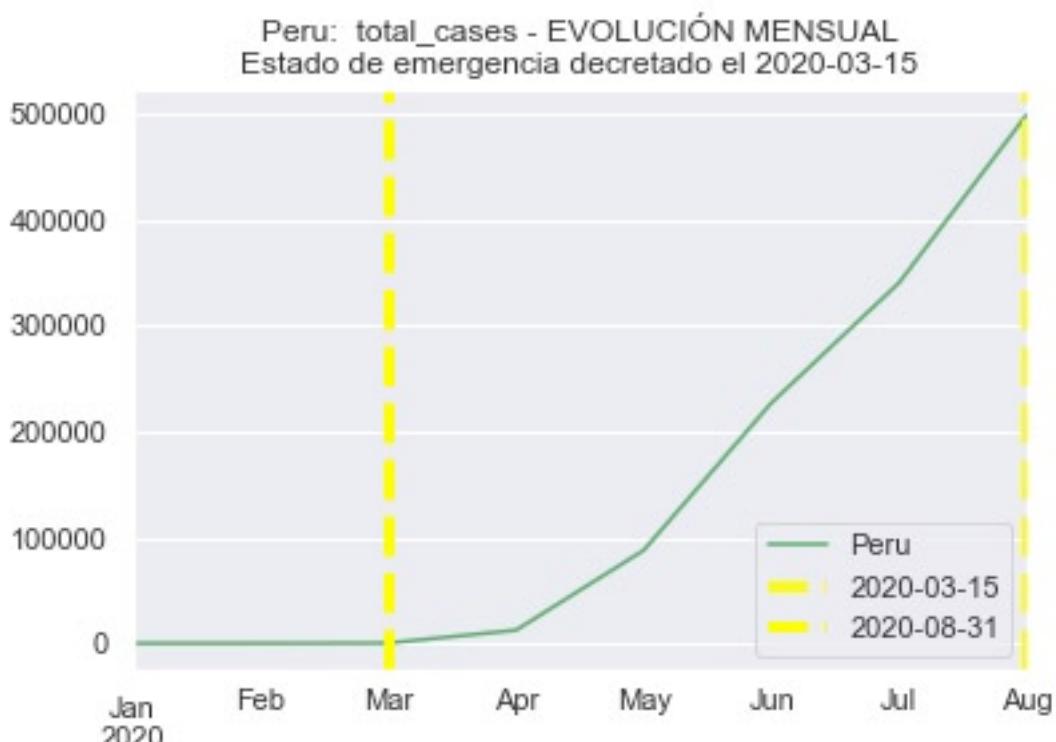
## COLLECTION OF SAVED PLOTS

### TOTAL INFECTED PEOPLE REGISTERED IN FRANCE (Updated 2020-08-23)



Amount of new cases reaches 250.000

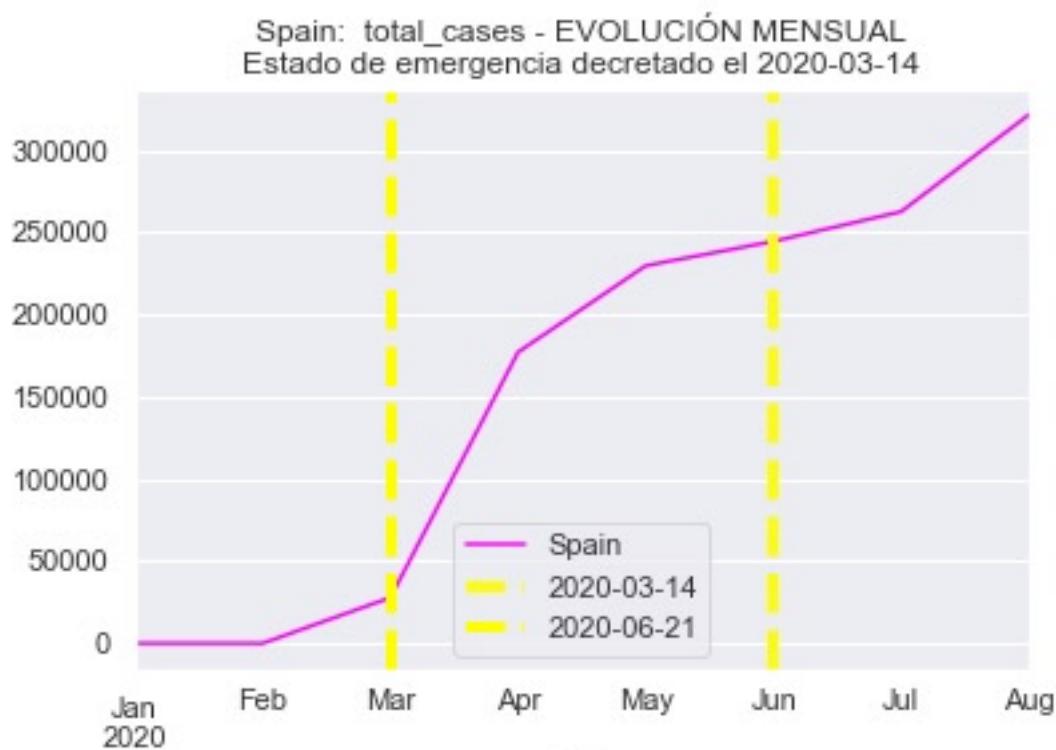
### TOTAL INFECTED PEOPLE REGISTERED IN PERÚ (Updated 2020-08-23)



Amount of new cases reaches 500.000

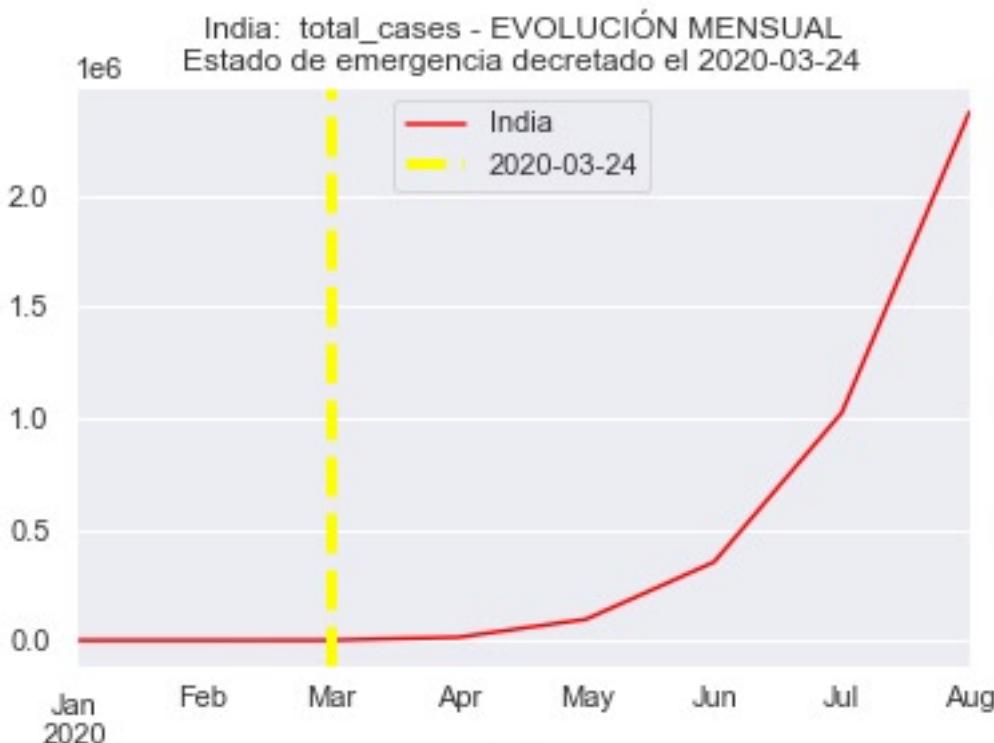
## COLLECTION OF SAVED PLOTS

### TOTAL INFECTED PEOPLE REGISTERED IN SPAIN (Updated 2020-08-23)



Amount of new cases reaches 300.000

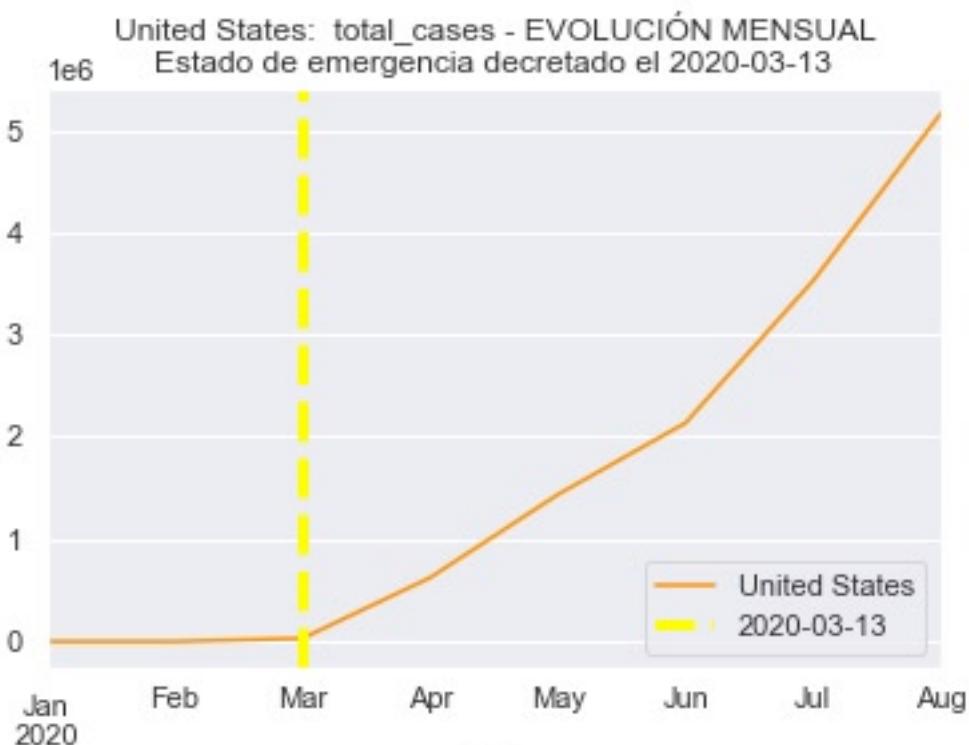
### TOTAL INFECTED PEOPLE REGISTERED IN INDIA (Updated 2020-08-23)



Amount of new cases reaches 3.0 x100K

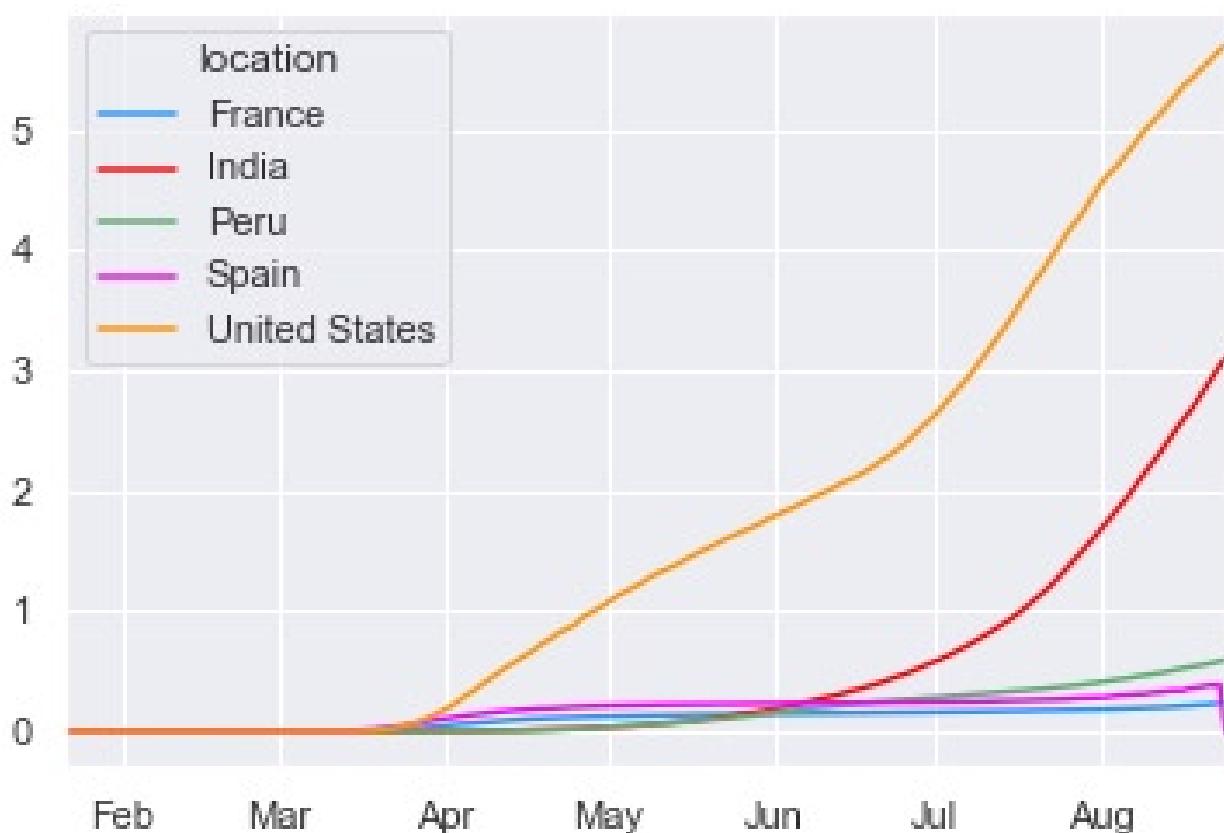
## COLLECTION OF SAVED PLOTS

### TOTAL INFECTED PEOPLE REGISTERED IN USA (Updated 2020-08-23)



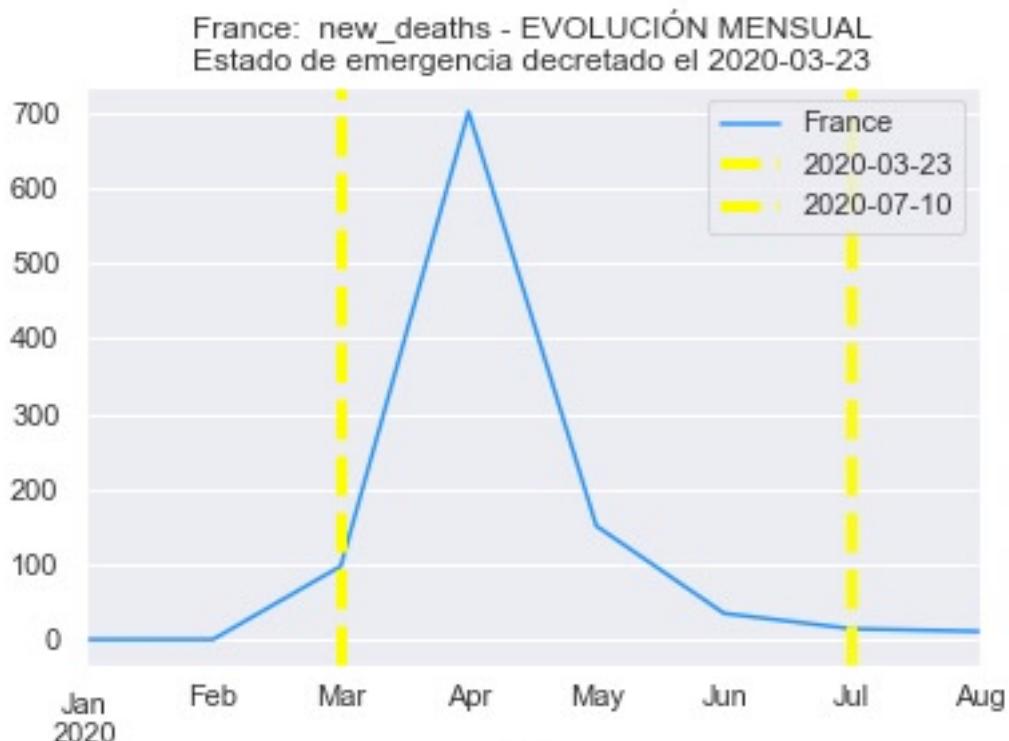
Amount of new cases reaches 5.0 X 100K

### SMOOTHED - ALL COUNTRIES TOTAL REGISTERED INFECTIONS



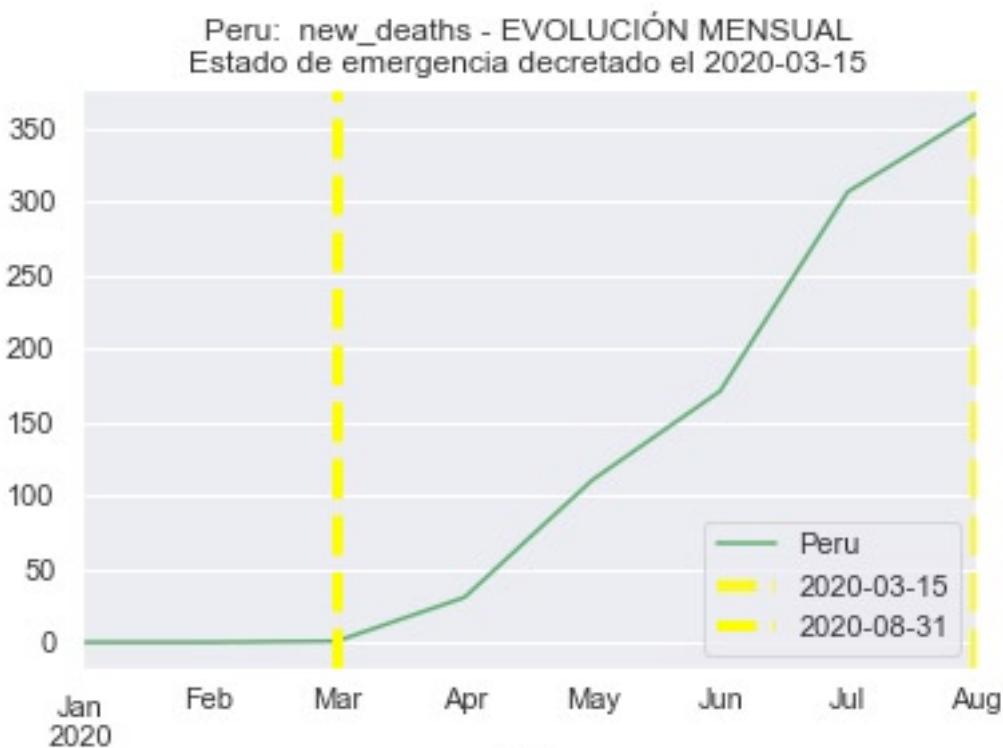
## COLLECTION OF SAVED PLOTS

### DAILY DEATHS REGISTERED IN FRANCE (Updated 2020-08-23)



Amount reaches 700

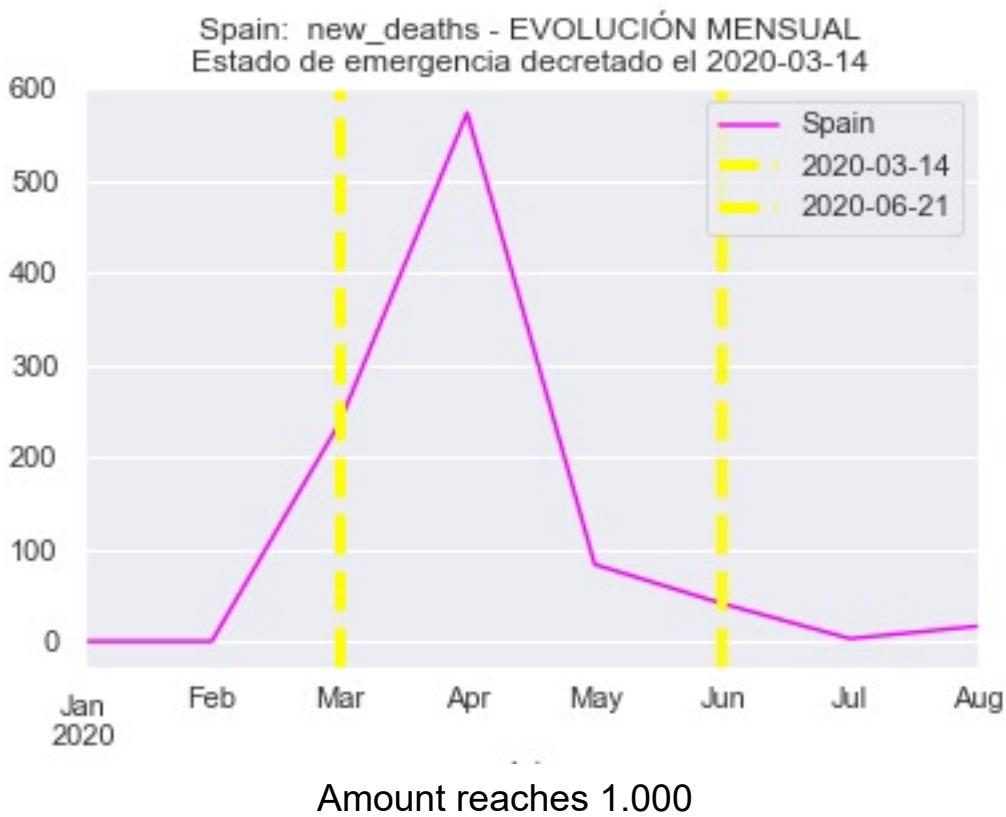
### DAILY DEATHS REGISTERED IN PERU (Updated 2020-08-23)



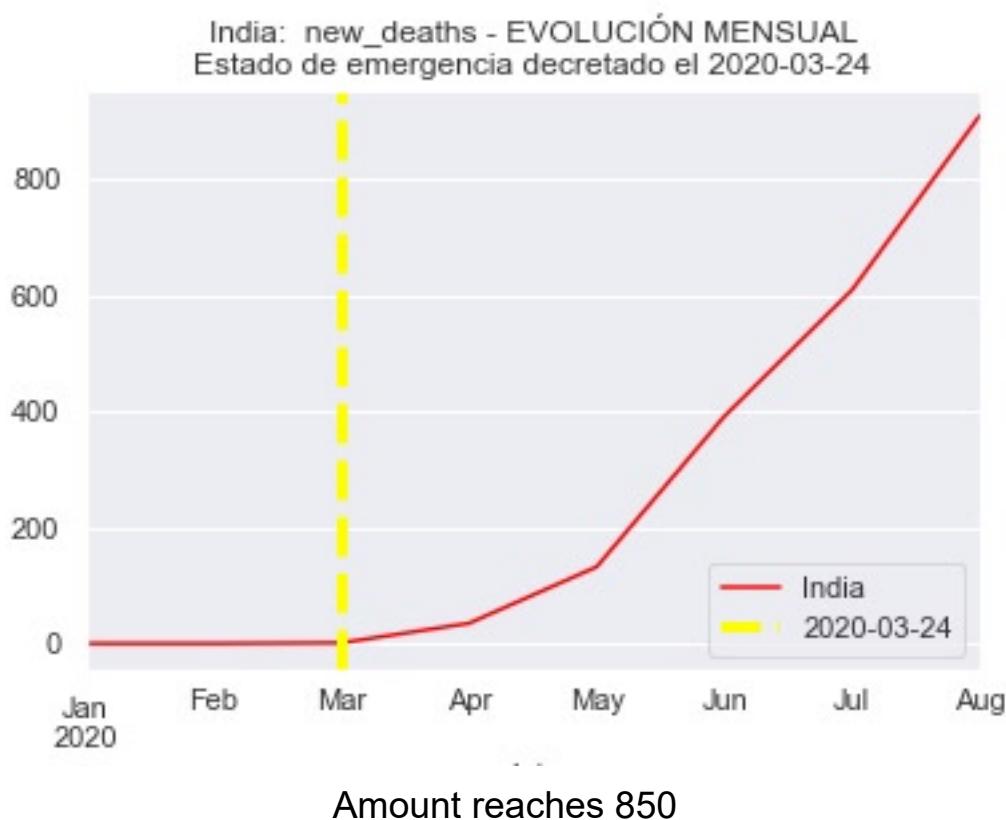
Amount reaches 350

## COLLECTION OF SAVED PLOTS

### DAILY DEATHS REGISTERED IN SPAIN (Updated 2020-08-23)

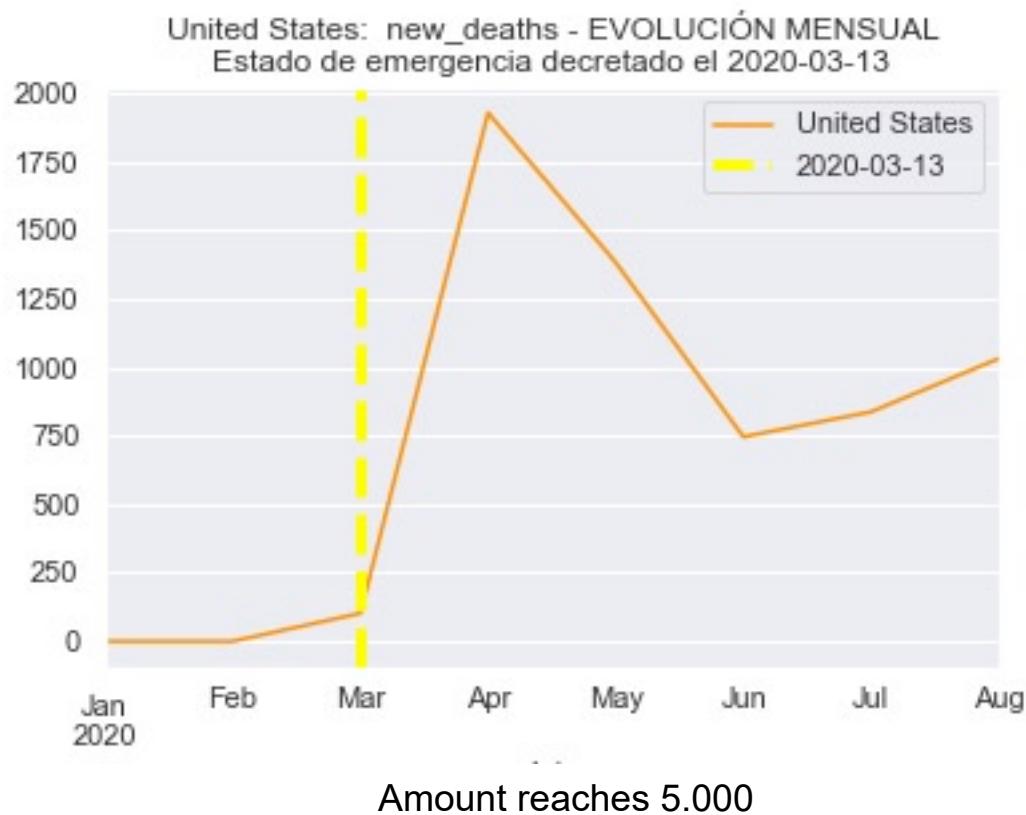


### DAILY DEATHS REGISTERED IN INDIA (Updated 2020-08-23)

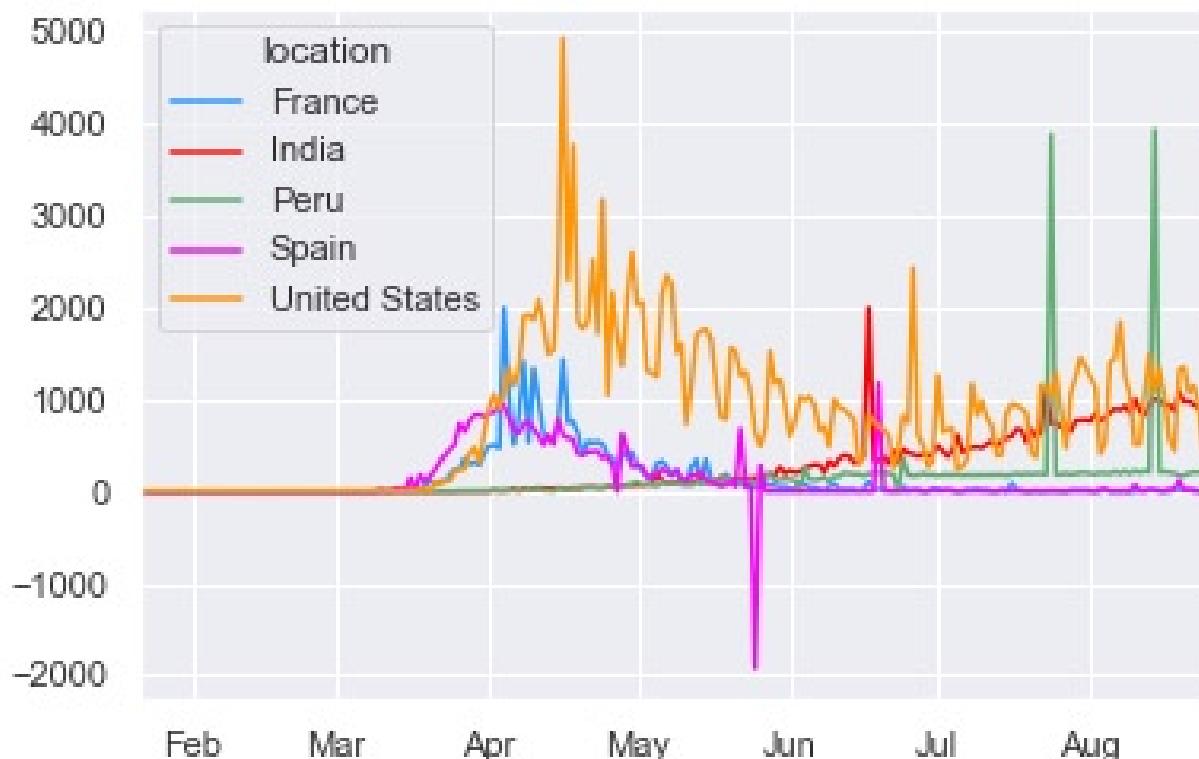


## COLLECTION OF SAVED PLOTS

### DAILY DEATHS REGISTERED IN USA (Updated 2020-08-23)

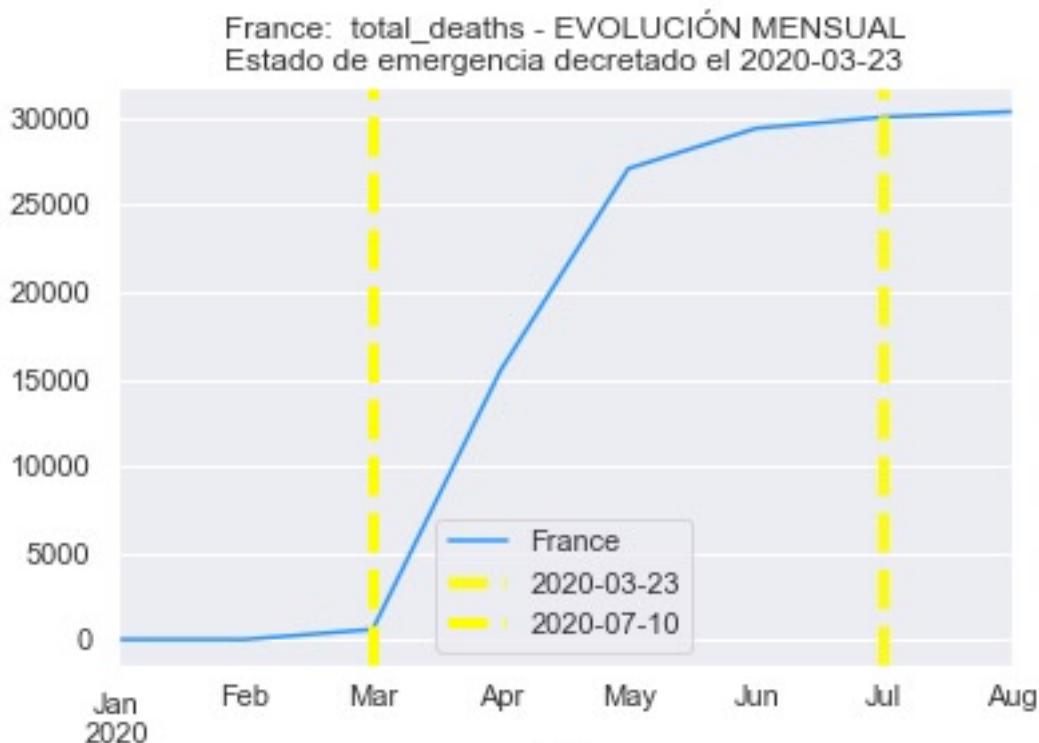


### DAILY DEATHS REGISTERED IN ALL COUNTRIES (Updated 2020-08-23)



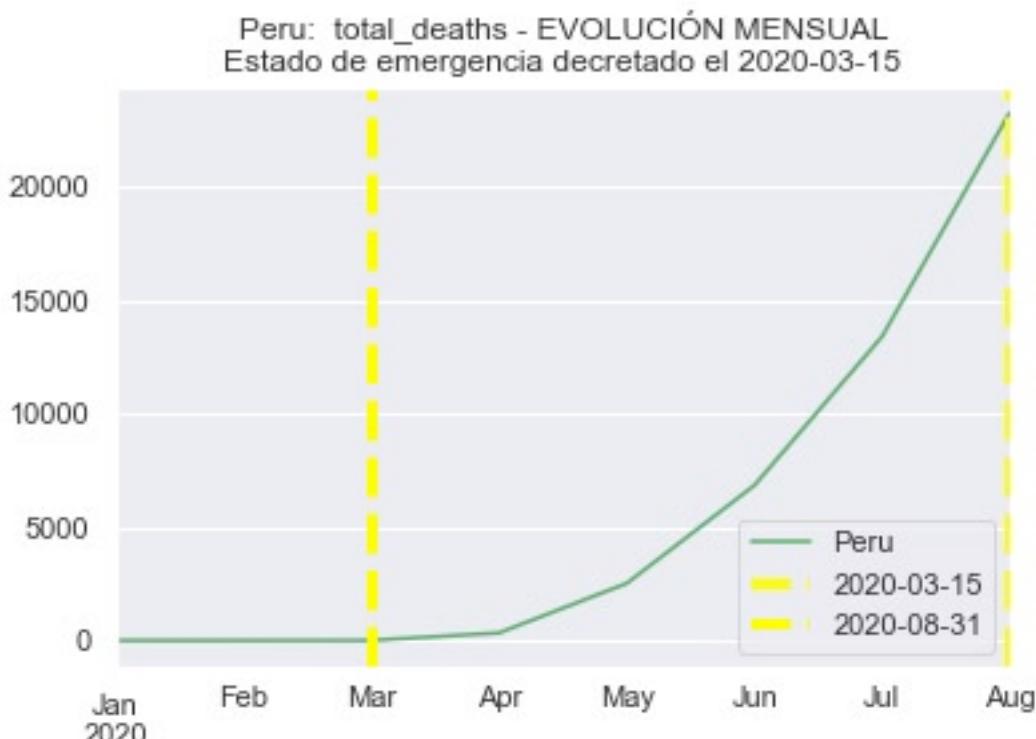
## COLLECTION OF SAVED PLOTS

### TOTAL DEATHS REGISTERED IN FRANCE (Updated 2020-08-23)



Amount reaches 30.000

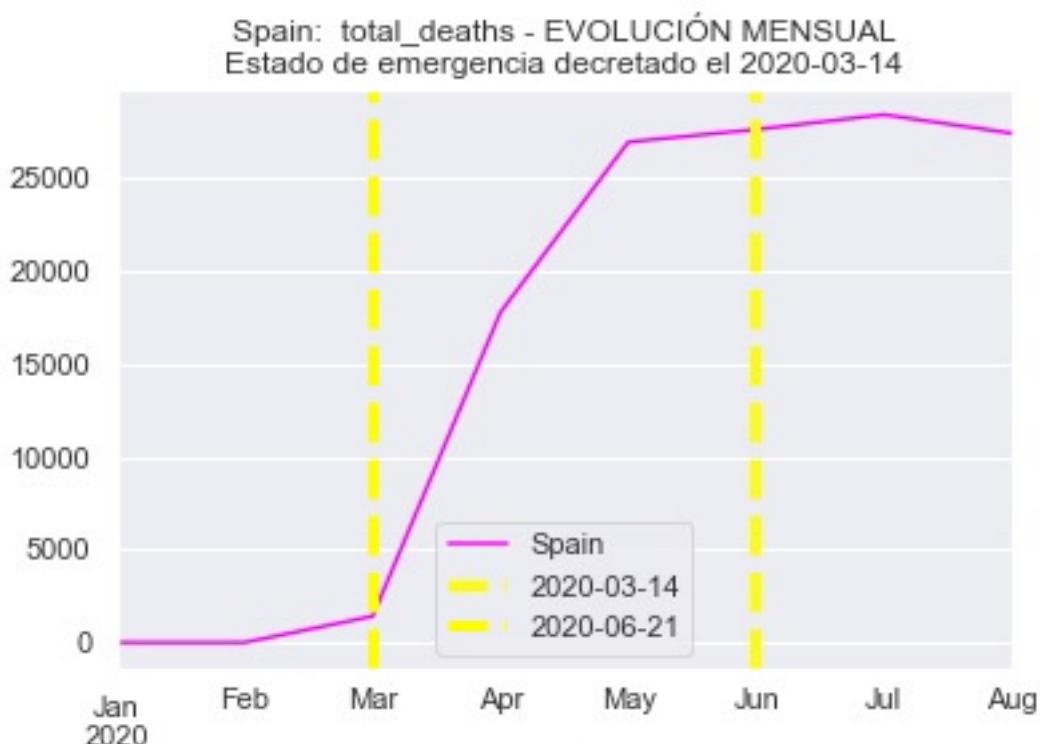
### TOTAL DEATHS REGISTERED IN PERÚ (Updated 2020-08-23)



Amount reaches 24.000

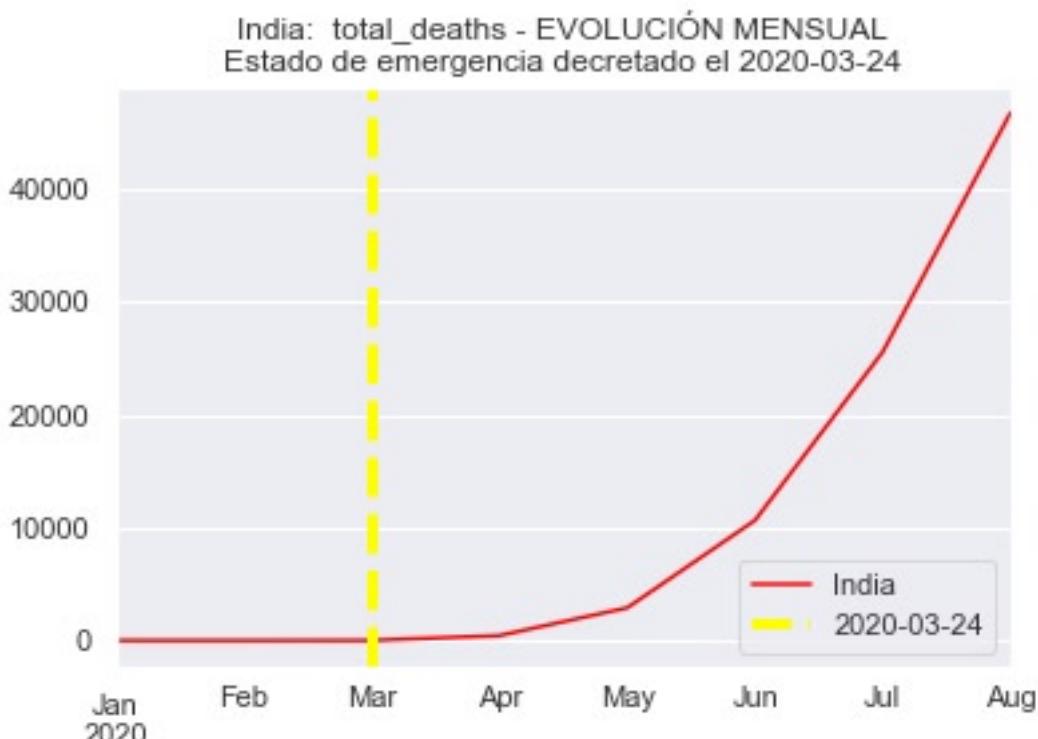
## COLLECTION OF SAVED PLOTS

### TOTAL DEATHS REGISTERED IN SPAIN (Updated 2020-08-23)



Amount reaches 27.000

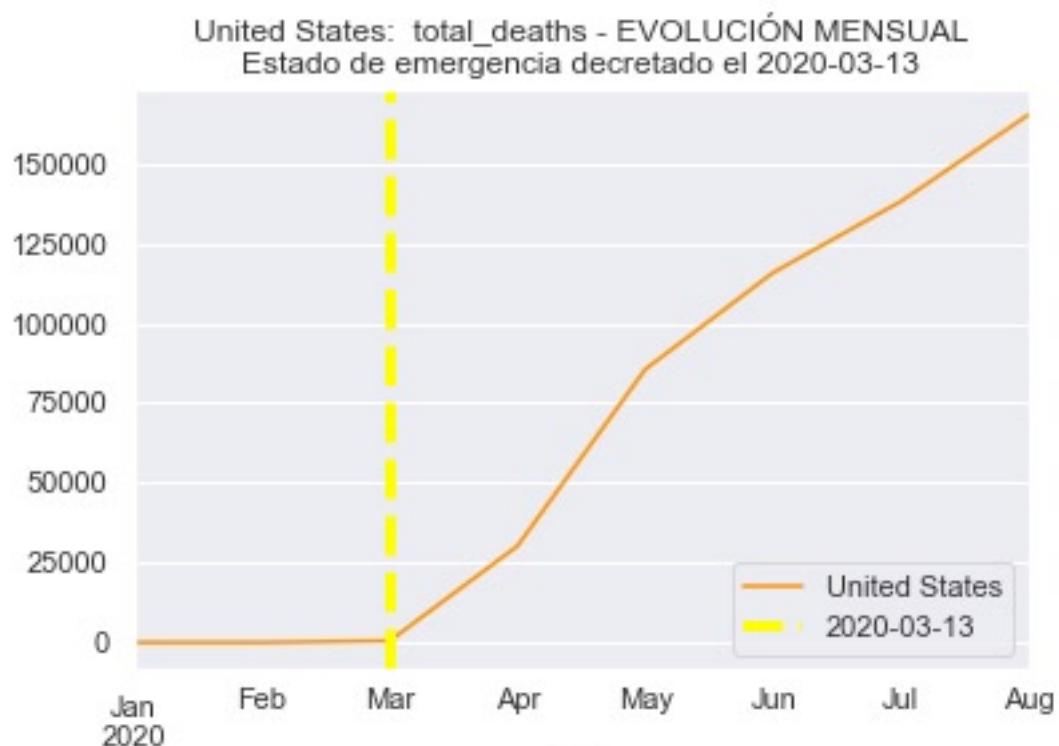
### TOTAL DEATHS REGISTERED IN INDIA (Updated 2020-08-23)



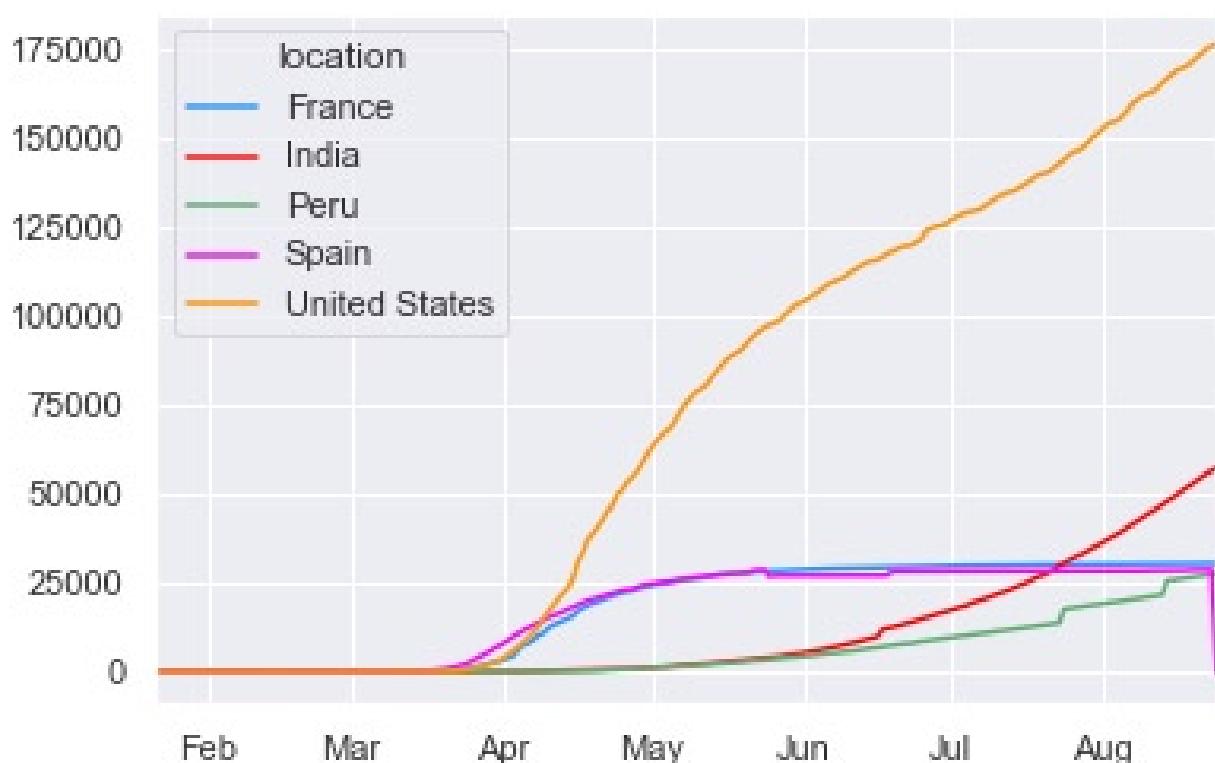
Amount reaches 50.000

## COLLECTION OF SAVED PLOTS

### TOTAL DEATHS REGISTERED IN USA (Updated 2020-08-23)



### TOTAL DEATHS REGISTERED IN ALL COUNTRIES (Updated 2020-08-23)



# COLLECTION OF SAVED PLOTS - OTHER SAVED PLOTS



## COMPARISON

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 %

## 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 France 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 in Spain and France while alarm state was activated.  
Measures taken in India, Perú and USA (not including lockdown) have been proved ineffective.

## OUTLIERS

- We have found a great decrease in total number of death people in Spain dated in May 25th. The total number of deaths decreased by 1.918 people.  
Spanish government said those deaths were not caused by covid and that was the reason told in the news.

## B- GROUP API

This Api returns a json file using a token and group\_id to verify the user's identity with a GET method.

The returned file <d\_averages.json> contains data about the daily mean of new deaths in the selected countries for group B

Original data source: <https://covid.ourworldindata.org/data/owid-covid-data.csv>

Countries: India, Peru, US, Francia, Spain.

---

**User's identification.**

[\*\*/get\\_token as endpoint\*\*](#)

**Args & kwargs(id):**

In both steps of this function we have labelled kwargs as 'id'

id = group\_id (A previously given string of characters)

**Returns:**

If the given group\_id matches the stored value, program returns a dictionary with an only key labeled 'token' and another string of characters as value.

---

**Print .json file in your browser.**

[\*\*/get\\_json as endpoint\*\*](#)

**Args & kwargs(id):**

In both steps of this function we have labelled kwargs as 'id'

id = token (The previously returned string of characters)

**Returns:**

If the given token matches the stored value, api returns a json file with updated info about the daily mean of new deaths in India, Peru, US, Francia and Spain.

---

**Import .json file to jupiter or python file.**

With [\*\*/get\\_json\*\*](#) and [\*\*token\\_id\*\*](#) as endpoint you can read the .json in your python file using the url.

**Quick save .json file to your computer.**

When json file is in screen you can select and copy all text.

Open a txt editor, paste your content and save it as d\_averages.json so you can use it later.