Covid 19 Data Analysis

# COVID 19 Dataset

We have taken a small dataset of Covid-19, just for your understaning purpose. You have to work on the original dataset which contains about 19000 rows. The data used here is till 29-April-2020 and has records as on 29-April-2020. This data is available as a CSV file, downloaded from Kaggle. We will analyze this data.

head(covid)

## # A tibble: 6 × 6  
## Date State Region Confirmed Deaths Recovered  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 4/29/2020 <NA> Afghanistan 1939 60 252  
## 2 4/29/2020 <NA> Albania 766 30 455  
## 3 4/29/2020 <NA> Algeria 3848 444 1702  
## 4 4/29/2020 <NA> Andorra 743 42 423  
## 5 4/29/2020 <NA> Angola 27 2 7  
## 6 4/29/2020 <NA> Antigua and Barbuda 24 3 11

describe(covid)

## vars n mean sd median trimmed mad min max range  
## Date\* 1 321 1.00 0.00 1.0 1.00 0.00 1 1 0  
## State\* 2 140 68.66 39.66 68.5 68.57 51.15 1 137 136  
## Region\* 3 321 107.02 58.45 104.0 109.74 85.99 1 187 186  
## Confirmed 4 321 9949.80 31923.85 653.0 2719.53 947.38 0 299691 299691  
## Deaths 5 321 709.15 3236.16 12.0 91.24 17.79 0 27682 27682  
## Recovered 6 321 3030.28 14364.87 73.0 312.08 108.23 0 132929 132929  
## skew kurtosis se  
## Date\* NaN NaN 0.00  
## State\* 0.01 -1.24 3.35  
## Region\* -0.17 -1.32 3.26  
## Confirmed 5.63 36.67 1781.82  
## Deaths 6.87 48.90 180.62  
## Recovered 6.87 50.76 801.77

covid |> colnames()

## [1] "Date" "State" "Region" "Confirmed" "Deaths" "Recovered"

covid |> str()

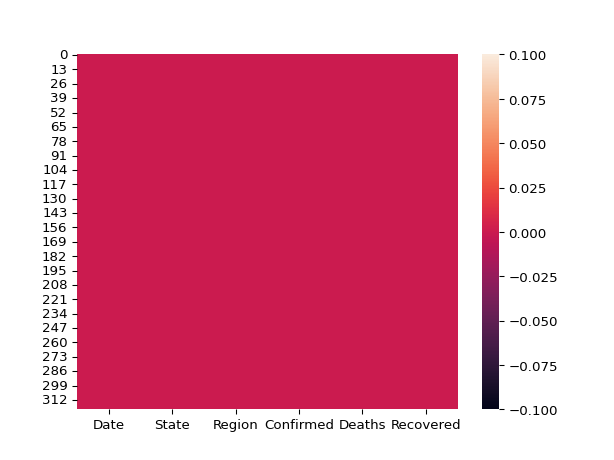
## spc\_tbl\_ [321 × 6] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Date : chr [1:321] "4/29/2020" "4/29/2020" "4/29/2020" "4/29/2020" ...  
## $ State : chr [1:321] NA NA NA NA ...  
## $ Region : chr [1:321] "Afghanistan" "Albania" "Algeria" "Andorra" ...  
## $ Confirmed: num [1:321] 1939 766 3848 743 27 ...  
## $ Deaths : num [1:321] 60 30 444 42 2 3 214 30 580 23 ...  
## $ Recovered: num [1:321] 252 455 1702 423 7 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Date = col\_character(),  
## .. State = col\_character(),  
## .. Region = col\_character(),  
## .. Confirmed = col\_double(),  
## .. Deaths = col\_double(),  
## .. Recovered = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

is.na(covid) |> colSums()

## Date State Region Confirmed Deaths Recovered   
## 0 181 0 0 0 0

## ### Lets run some Analysis to view somw certain Updates. Firstly we would want to see the heatmap of the null values.

import seaborn as sns  
import matplotlib.pyplot as plt  
sns.heatmap(r.covid.isnull())  
plt.show()

 ## Lets go into solving real life Questions. ### Q.1 ) Show the number of Confirmed , Deaths and Recovered cases in each Region.

covid |> group\_by(Region) |> reframe(COnfirmed = sum(Confirmed), Deaths = sum(Deaths), Recovered = sum(Recovered))

## # A tibble: 187 × 4  
## Region COnfirmed Deaths Recovered  
## <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 1939 60 252  
## 2 Albania 766 30 455  
## 3 Algeria 3848 444 1702  
## 4 Andorra 743 42 423  
## 5 Angola 27 2 7  
## 6 Antigua and Barbuda 24 3 11  
## 7 Argentina 4285 214 1192  
## 8 Armenia 1932 30 900  
## 9 Australia 6752 91 5715  
## 10 Austria 15402 580 12779  
## # ℹ 177 more rows

### Q2) Remove all the records where Confirmed Cases is Less Than 10.

covids <- covid |> filter(Confirmed > 10)

covids

## # A tibble: 301 × 6  
## Date State Region Confirmed Deaths Recovered  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 4/29/2020 <NA> Afghanistan 1939 60 252  
## 2 4/29/2020 <NA> Albania 766 30 455  
## 3 4/29/2020 <NA> Algeria 3848 444 1702  
## 4 4/29/2020 <NA> Andorra 743 42 423  
## 5 4/29/2020 <NA> Angola 27 2 7  
## 6 4/29/2020 <NA> Antigua and Barbuda 24 3 11  
## 7 4/29/2020 <NA> Argentina 4285 214 1192  
## 8 4/29/2020 <NA> Armenia 1932 30 900  
## 9 4/29/2020 <NA> Austria 15402 580 12779  
## 10 4/29/2020 <NA> Azerbaijan 1766 23 1267  
## # ℹ 291 more rows

### Q.3) In which Region, maximum number of Confirmed cases were recorded ?

covid |> group\_by(Region) |> summarise(Confirmed = sum(Confirmed)) |> arrange(desc(Confirmed))

## # A tibble: 187 × 2  
## Region Confirmed  
## <chr> <dbl>  
## 1 US 1039909  
## 2 Spain 236899  
## 3 Italy 203591  
## 4 France 166543  
## 5 UK 166441  
## 6 Germany 161539  
## 7 Turkey 117589  
## 8 Russia 99399  
## 9 Iran 93657  
## 10 Mainland China 82862  
## # ℹ 177 more rows

### Q.4) In which Region, minimum number of Deaths cases were recorded ?

covid |> group\_by(Region) |> summarise(Deaths = sum(Deaths)) |> arrange((Deaths)) |> print(n = 5)

## # A tibble: 187 × 2  
## Region Deaths  
## <chr> <dbl>  
## 1 Bhutan 0  
## 2 Cambodia 0  
## 3 Central African Republic 0  
## 4 Dominica 0  
## 5 Eritrea 0  
## # ℹ 182 more rows

### Q.5) How many Confirmed , Deaths & Recovered cases were reported from India till 29 April 2020?

covid |> filter(Region == 'India')

## # A tibble: 1 × 6  
## Date State Region Confirmed Deaths Recovered  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 4/29/2020 <NA> India 33062 1079 8437

### Q. 6-A ) Sort the entire data wrt No. of Confirmed cases in ascending order

covids |> arrange(Confirmed) |> print(n = 5)

## # A tibble: 301 × 6  
## Date State Region Confirmed Deaths Recovered  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 4/29/2020 <NA> Burundi 11 1 4  
## 2 4/29/2020 <NA> Seychelles 11 0 6  
## 3 4/29/2020 Greenland Denmark 11 0 11  
## 4 4/29/2020 Montserrat UK 11 1 2  
## 5 4/29/2020 Yukon Canada 11 0 0  
## # ℹ 296 more rows

### Q. 6-B ) Sort the entire data wrt No. of Recovered cases in descending order.

covid |> arrange(desc(Recovered)) |> print(n = 10)

## # A tibble: 321 × 6  
## Date State Region Confirmed Deaths Recovered  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 4/29/2020 <NA> Spain 236899 24275 132929  
## 2 4/29/2020 Recovered US 0 0 120720  
## 3 4/29/2020 <NA> Germany 161539 6467 120400  
## 4 4/29/2020 <NA> Iran 93657 5957 73791  
## 5 4/29/2020 <NA> Italy 203591 27682 71252  
## 6 4/29/2020 Hubei Mainland China 68128 4512 63616  
## 7 4/29/2020 <NA> France 165093 24087 48228  
## 8 4/29/2020 <NA> Turkey 117589 3081 44040  
## 9 4/29/2020 <NA> Brazil 79685 5513 34132  
## 10 4/29/2020 <NA> Switzerland 29407 1716 22600  
## # ℹ 311 more rows

## ### We have come to an End.