

# R SHINY PROJECT REPORT

## TEAM 12

# Relationship between Covid-19 and weather

### Team Members:

1. Prajwal
2. Anitha
3. Amal
4. Bharath

**Objective:** This project describes the comparison between covid-19 and weather on a certain period of time.

### Introduction:

This is an analysis report of the Novel Coronavirus (COVID-19) around the world, to demonstrate data processing and visualisation with R, tidyverse and ggplot2.

### **What is a coronavirus?**

The coronavirus family causes illnesses ranging from the common cold to more severe diseases such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), according to the WHO.

They circulate in animals and some can be transmitted between animals and humans. Several coronaviruses are circulating in animals that have not yet infected humans.

The new coronavirus, the seventh known to affect humans, has been named COVID-19.

### **What are the symptoms?**

Common signs of infection include fever, coughing and breathing difficulties. In severe cases, it can cause pneumonia, multiple organ failure and death.

The incubation period of COVID-19 is thought to be between one and 14 days. It is contagious before symptoms appear, which is why so many people get infected.

Infected patients can be also asymptomatic, meaning they do not display any symptoms despite having the virus in their systems.

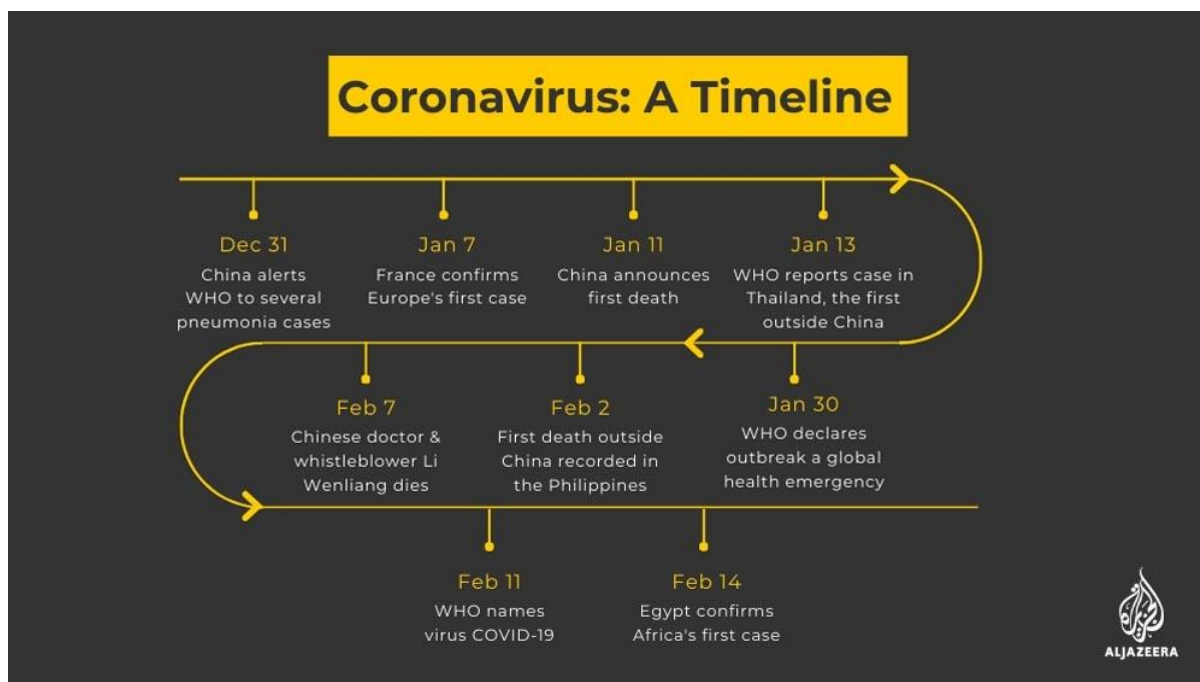
## Where did it come from?

China alerted the WHO to cases of unusual pneumonia in Wuhan on December 31.

COVID-19 is thought to have originated in a seafood market where wildlife was sold illegally.

On February 7, Chinese researchers said the virus could have spread from an infected animal to humans through illegally trafficked pangolins, prized in Asia for food and medicine.

Scientists have pointed to either bats or snakes as possible sources.



## Aim:

The aim of this project is to visualize :

- The impact of covid-19 on all the different countries of the world.
- The impact weather has on covid 19 and its correlation with covid-19.
- The impact of Covid-19 on India.

## Data Collection:

The data source used for this analysis is the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository built by the Center for Systems Science and Engineering, Johns Hopkins University. They are 3 csv files that are auto updated daily so that we can work with the current data and visualize it. These 3 csv files are downloaded and saved as local files and then are loaded into R.

## Loading Data:

```
#dataset files
DownloadCovidData <- function()
{
  #getwd()
  download.file(
    url      = "https://github.com/CSSEGISandData/COVID-19/
archive/master.zip",
    destfile = "F:/R-Shiny Project/COVID-19_Analysis/data/covid19JHU.zip"
  )

  data_path <- "COVID-19-master/csse_covid_19_data/
csse_covid_19_time_series/time_series_covid19_"
  unzip(
    zipfile  = "data/covid19JHU.zip",
    files    = paste0(data_path, c("confirmed_global.csv",
                                   "deaths_global.csv",
                                   "recovered_global.csv")),
    exdir    = "data",
    junkpaths = T
  )
}
```

```
#update function updates the data after a certain time stamp
UpdateCovidData <- function() {

  cur_time <- Sys.time()
  file_time <- file.info("data/covid19JHU.zip")$mtime
  time_diff <- difftime(cur_time, file_time, units = "hours")

  if (!dir.exists("data")) {
    dir.create('data')
    DownloadCovidData()
  } else if ((!file.exists("data/covid19JHU.zip")) || (time_diff > 12))
  {
    DownloadCovidData()
  }
}
```

```
#update with start of app
UpdateCovidData()

#reading downloaded files
data_confirmed <- read.csv("C:/Ani/~Education/Josephs/R_shiny_dashboard/
COVID-19_Analysis_New/data/
time_series_covid19_confirmed_global.csv")
data_deceased  <- read.csv("C:/Ani/~Education/Josephs/R_shiny_dashboard/
COVID-19_Analysis_New/data/
time_series_covid19_deaths_global.csv")
data_recovered <- read.csv("C:/Ani/~Education/Josephs/R_shiny_dashboard/
COVID-19_Analysis_New/data/
time_series_covid19_recovered_global.csv")
```

Each dataset has 266 rows, corresponding to country/region/province/state. It has 150 columns. Starting from column 5, each column corresponds to a single day. Here we have a look at the first 10 rows and the first 10 columns.

## Packages Used:

```
library(shiny)
library(shinydashboard)
library(dashboardthemes)
library(tidyverse)
library(leaflet)
library(plotly)
library(DT)
library(wbstats)
library(Hmisc)
library(reshape2)
library(ggplot2)
```

**Shiny package** - Shiny is an R package that makes it easy to build interactive web apps straight from R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards.

**Shiny Dashboard package** - shinydashboard is an R package whose job is to make it easier, as the name suggests, to build dashboards with Shiny.

**Dashboardthemes package** - Allows manual creation of themes and logos to be used in applications created using the 'shinydashboard' package. Removes the need to change the underlying css code by wrapping it into a set of convenient R functions.

**Tidyverse package** – This is a powerful package that allows us to wrangle, transform, import, manage and visualize the data with the help of sub packages like dplyr, tidyr, tibble and ggplot2.

**Leaflet package** – Leaflet is one of the most popular open-source libraries to make interactive maps. You can easily render spatial objects from the sp or sf packages, or data frames with latitude/longitude columns.

**Plotly package** – This package is an R package for creating interactive web-based graphs via the open source JavaScript graphing library [plotly.js](https://plot.ly).

**DT package** - The R package DT provides an R interface to the JavaScript library DataTables. R data objects (matrices or data frames) can be displayed as tables on HTML pages, and DataTables provides filtering, pagination, sorting, and many other features in the tables.

**Wbstats package** - The wbstats R-package allows researchers to quickly search and download the data of their particular interest in a programmatic and reproducible fashion; this facilitates a seamless integration into their workflow and allows analysis to be quickly rerun on different areas of interest and with realtime access to the data as well.

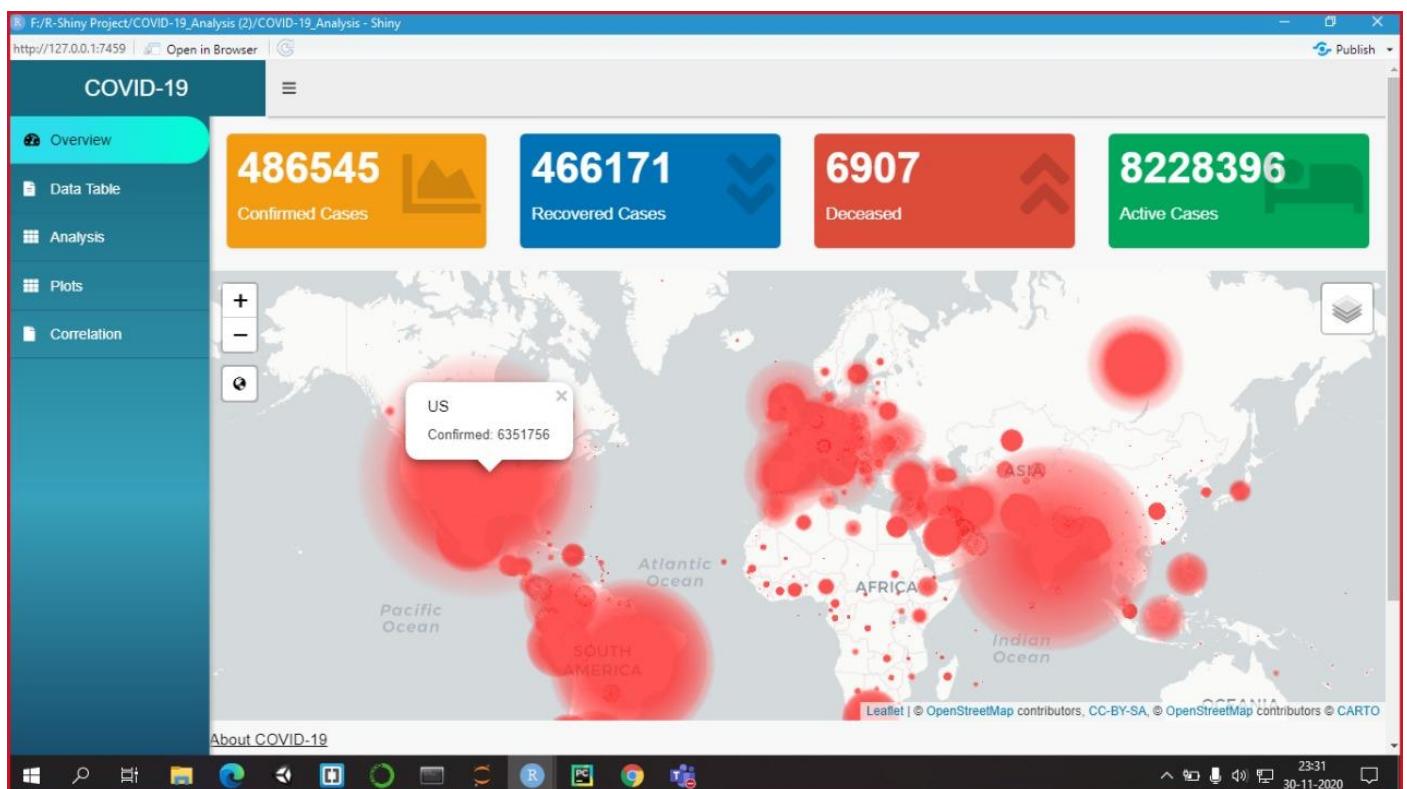
**Hmisc package** – This package contains many functions useful for data analysis, high-level graphics, utility operations, functions for computing sample size and power, importing and annotating datasets, imputing missing values, advanced table making, variable clustering, character string manipulation, conversion of R objects to LaTeX and html code, and recoding variables.

**Reshape2 Package** - reshape2 is an R package written by Hadley Wickham that makes it easy to transform data between wide and long formats.

**Ggplot2 Package** - ggplot2 is a plotting package that makes it simple to create complex plots from data in a data frame. It provides a more programmatic interface for specifying what variables to plot, how they are displayed, and general visual properties. Therefore, we only need minimal changes if the underlying data change or if we decide to change from a bar plot to a scatterplot. This helps in creating publication quality plots with minimal amounts of adjustments and tweaking.

# Dashboard

Our Covid-19 Dashboard has a side bar navigation orientation that consists of 5 tabs as you can see in the picture below:



## Overview Tab

The above picture shows the Overview tab.

This tab consists of value boxes that shows the total confirmed, recovered, deceased and active cases of the world. This tab also consists of a world map that depicts the total number of confirmed cases in every country.

## Data Table

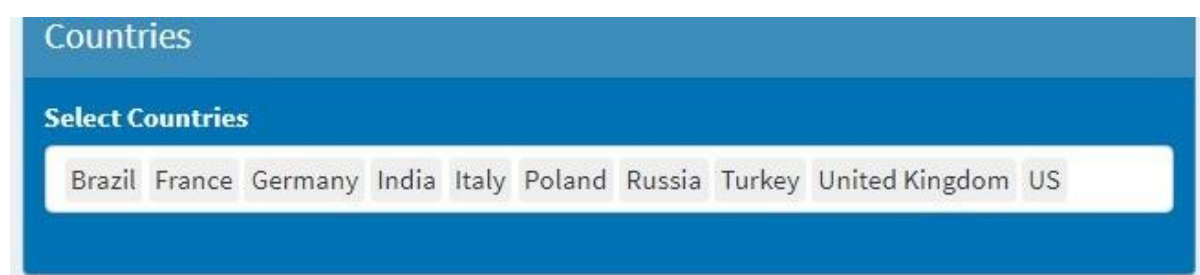
This tab consists of a Data table which summarizes all our datasets and shows us the most important columns that we took into consideration for our analysis and visualization. The Data Table consists of 15 columns and 190 rows.

This table also consists of a search bar through which you can type the name of a particular country and filter out the data for that country.

Search: <input type="text"/>										
Country	Total Confirmed	New Confirmed	Total Confirmed (per 100k)	Total Estimated Recoveries	New Estimated Recoveries	Total Deceased	New Deceased	Total Active	New Active	Total Active (per 100k)
Brazil	6314740	24468 (+0.39 %)	3014.64	5690260	13862 (+0.24 %)	172833	272 (+0.16 %)	451647	10334 (+2.34 %)	215.61
Italy	1585178	20646 (+1.32 %)	2623.52	1123625	33436 (+3.07 %)	54904	541 (+1.00 %)	406649	13331 (+3.39 %)	673.02
Russia	2249890	26390 (+1.19 %)	688.7	1871022	21862 (+1.18 %)	39127	451 (+1.17 %)	339741	4077 (+1.21 %)	104
Poland	985075	11482 (+1.18 %)	2594.03	695943	21571 (+3.20 %)	17029	283 (+1.69 %)	272103	10372 (+3.96 %)	716.54
Germany	1055691	3197 (+0.30 %)	1273.36	786640	3088 (+0.39 %)	16306	125 (+0.77 %)	252745	16 (+0.01 %)	304.86
United	12164	12164 (+0.76 %)	2432.54	12164	24762 (+1.92 %)	50000	215 (+0.37 %)	12164	12821	2432.54

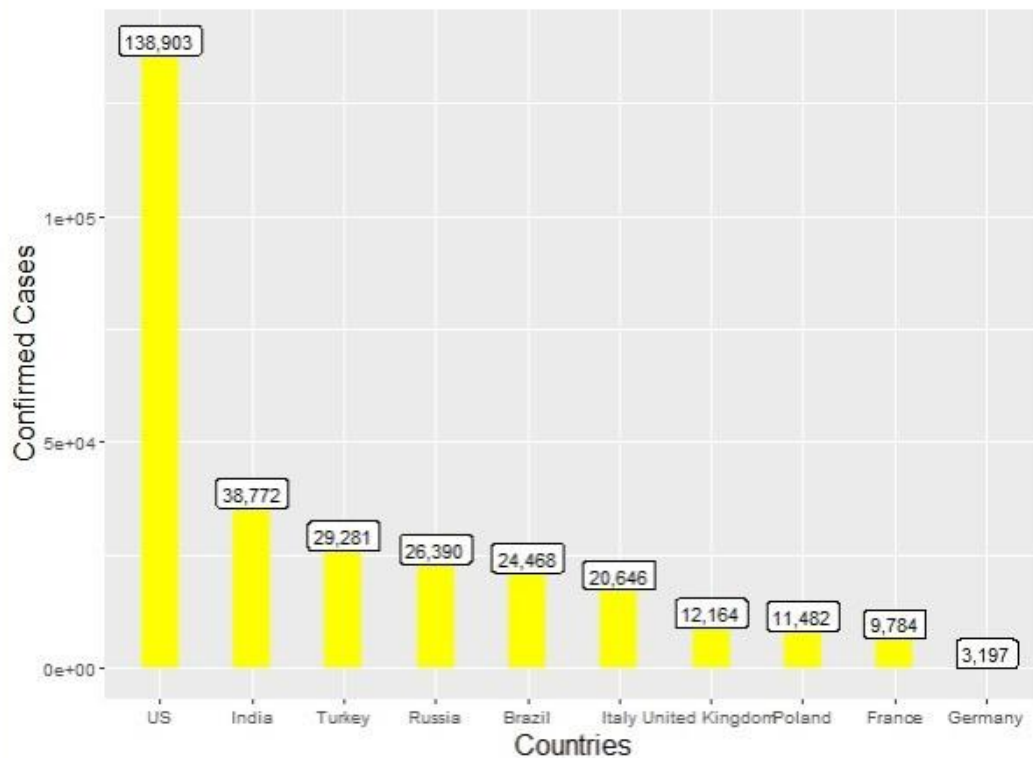
## Analysis Tab

This Tab consists of 6 different plots depicting the all the information contained in the data table. It consists of a select bar which lets you select any number of countries (with a default of 10) and shows you the visualized data of the countries selected.





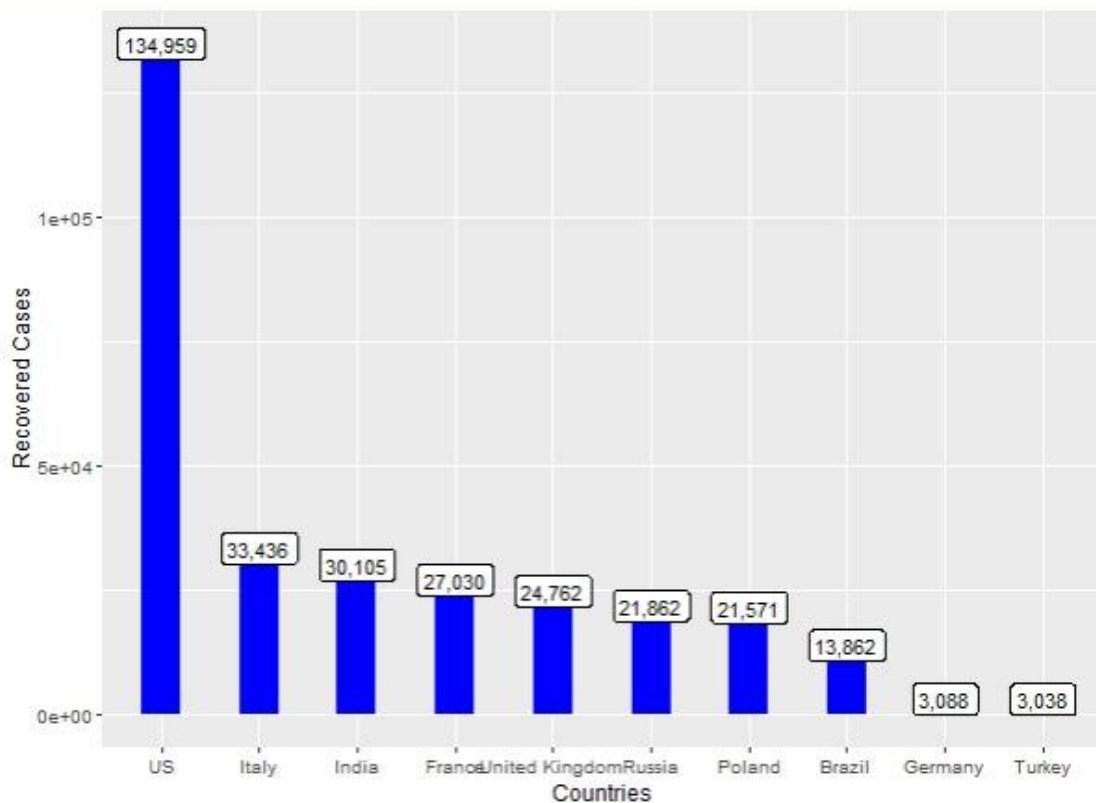
### Countries with New Confirmed Cases



1.

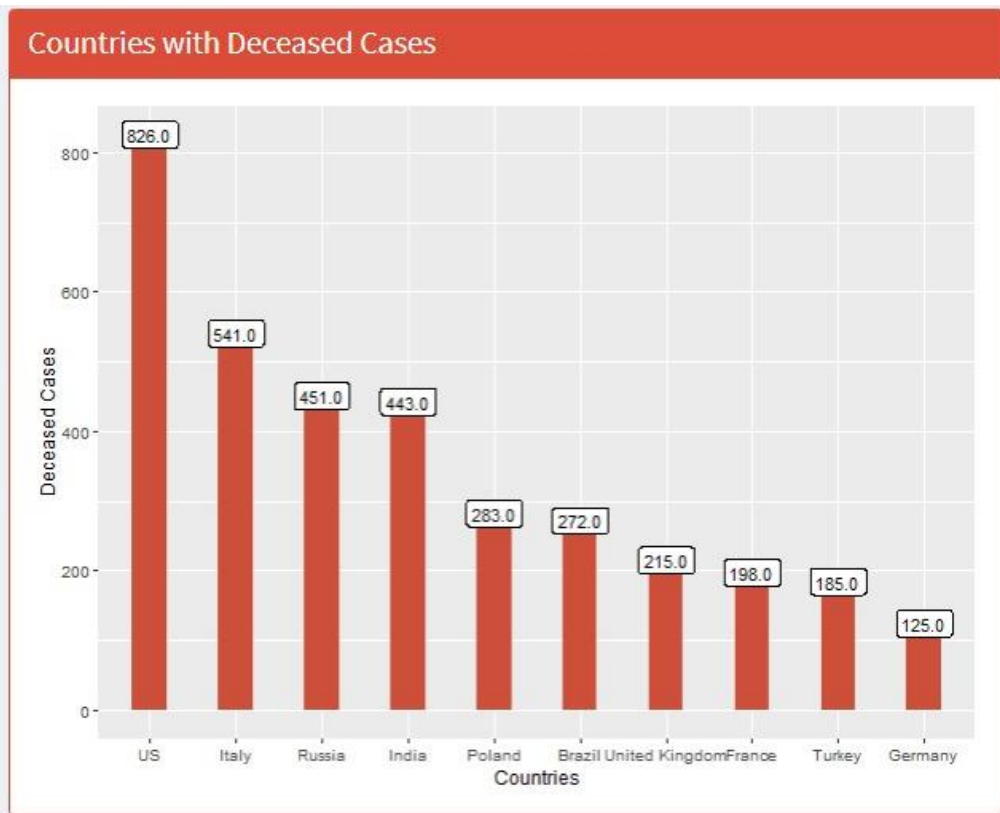
This bar chart shows the daily number of confirmed cases of the top 10 countries.

### Countries with Recovered Cases



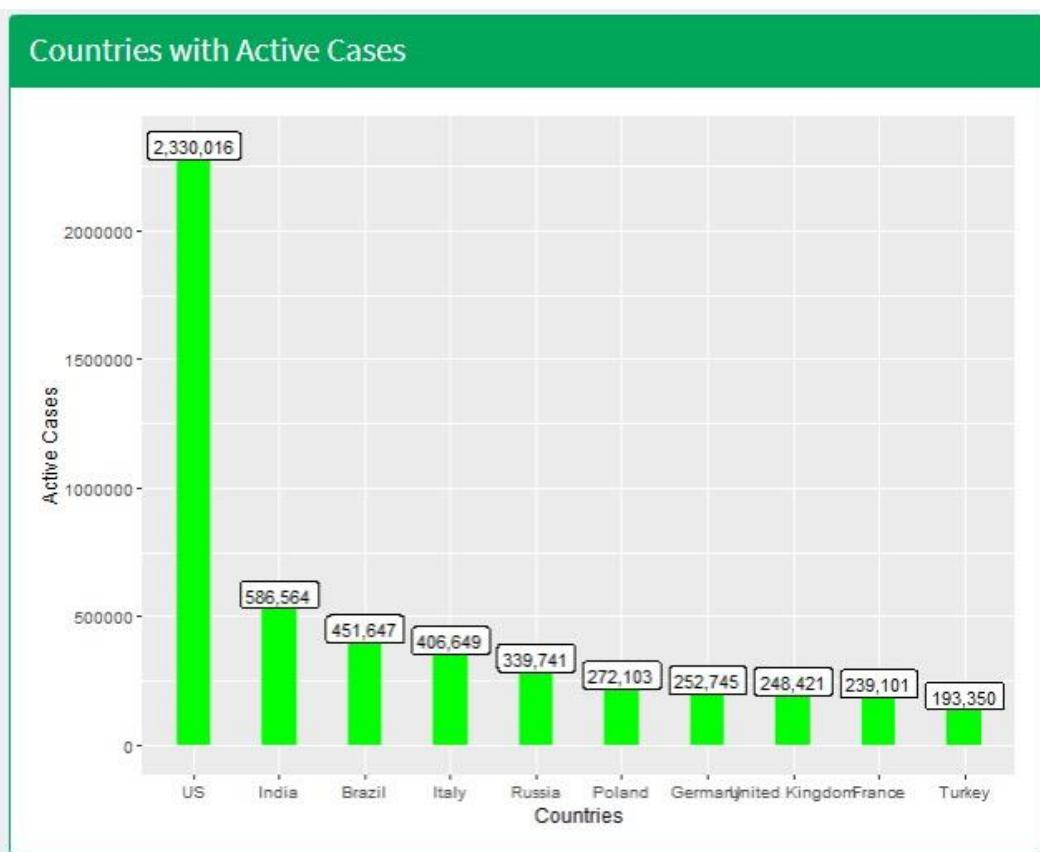
2.

This bar chart shows the daily number of recovered cases for the top 10 countries of the world.



3.

This chart shows us the daily number of deaths caused by covid-19 in the top 10 countries in the world.

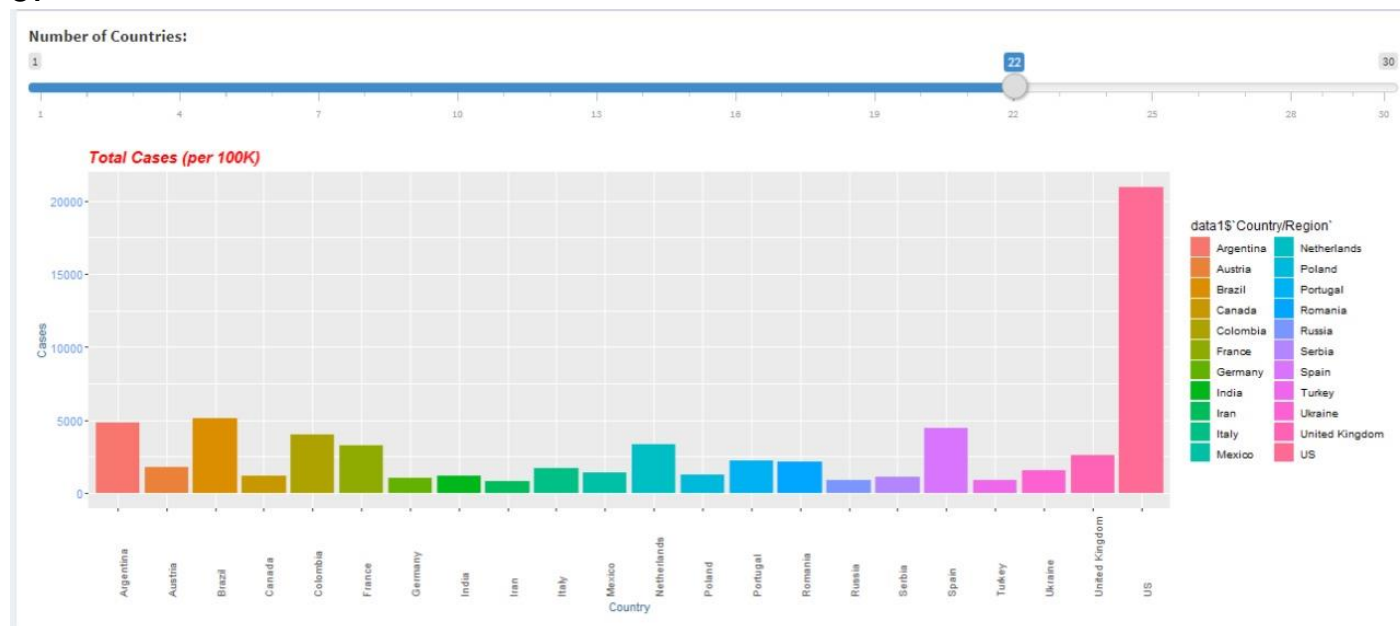


4.



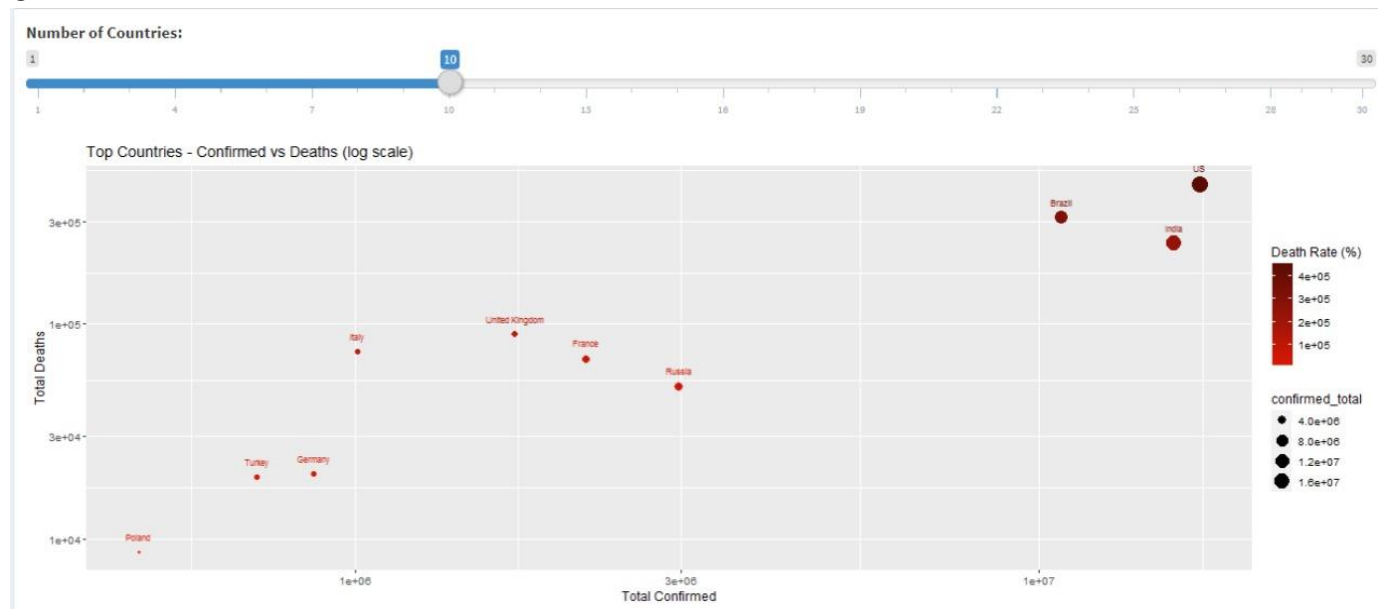
This bar chart shows us the total number of active cases in the top 10 countries in the world.

5.



This bar chart has a slider bar with which you can select between 1 to 30 different countries to compare the total number of cases per 100k of the population.

6.



This bar chart also has a slider bar with which you can select between 1 to 30 different countries to compare where they stand with respect to the total number of confirmed covid-19 cases and the total number of deaths. In this chart you can see that USA has the greatest number of confirmed covid-19 cases and deaths, followed by India and Brazil.

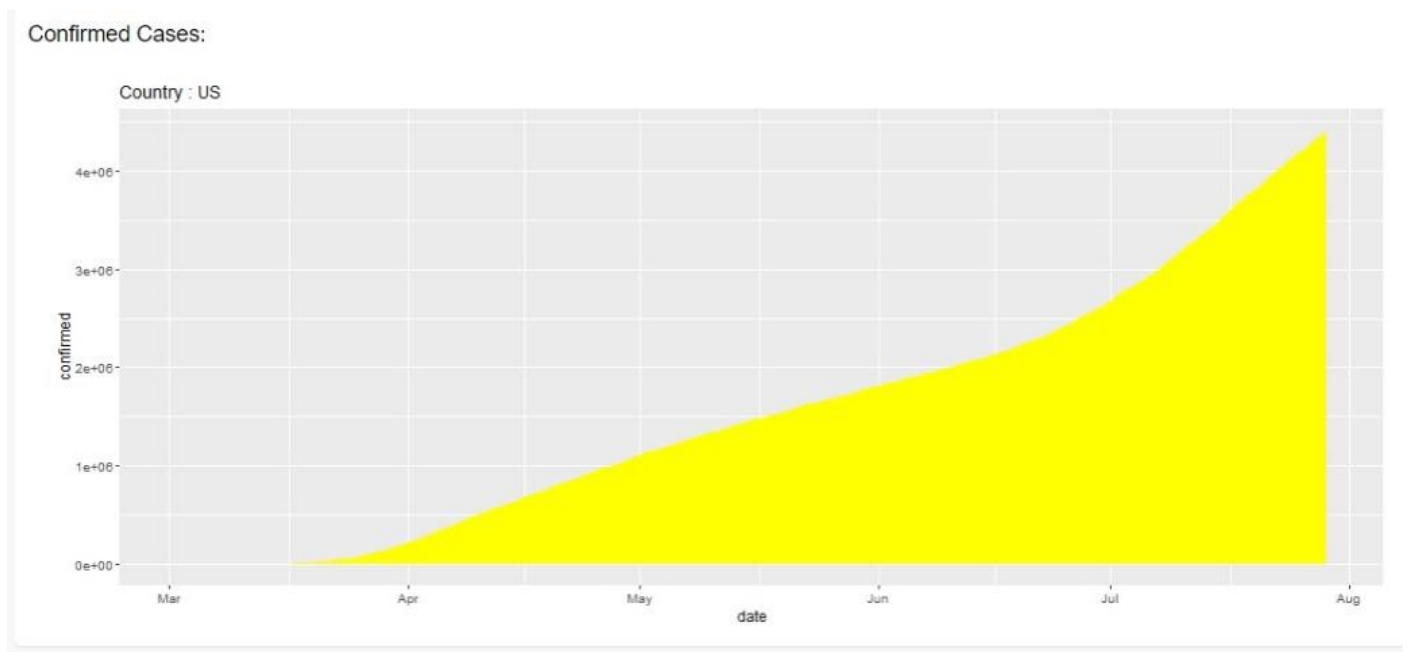
## Date Plot Tab

This tab consists of two date-range tabs, one start date and one end date, to select the time duration of the plots you want to see. This tab also contains a select tab to select the countries for which you want to see the visualizations.

Start Date

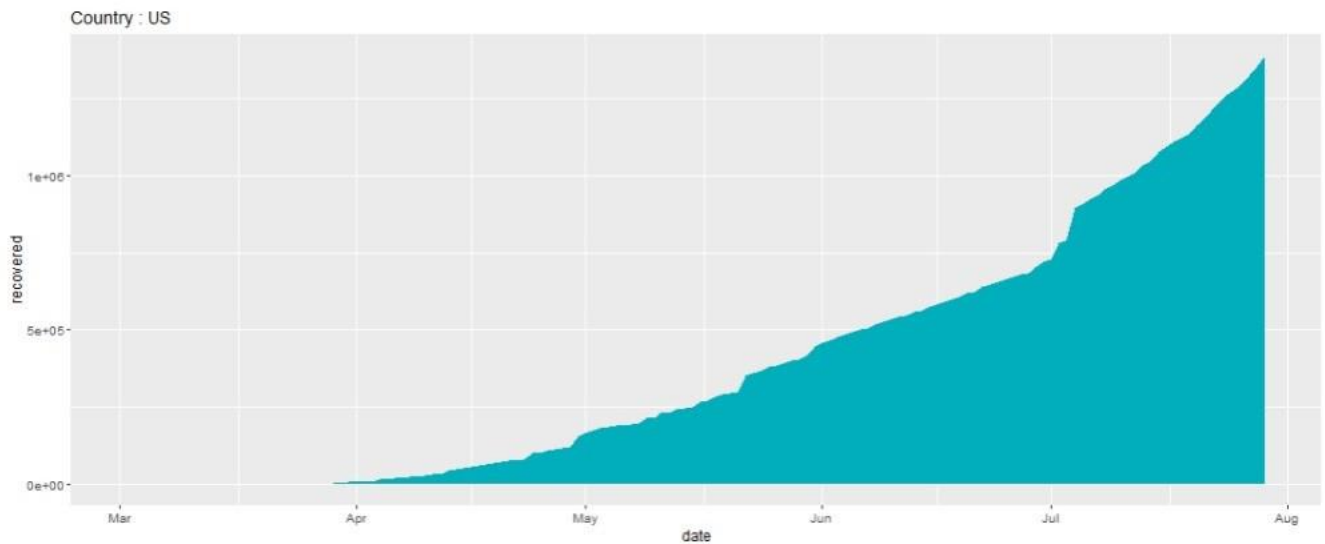
End Date

Select Countries  
Countries



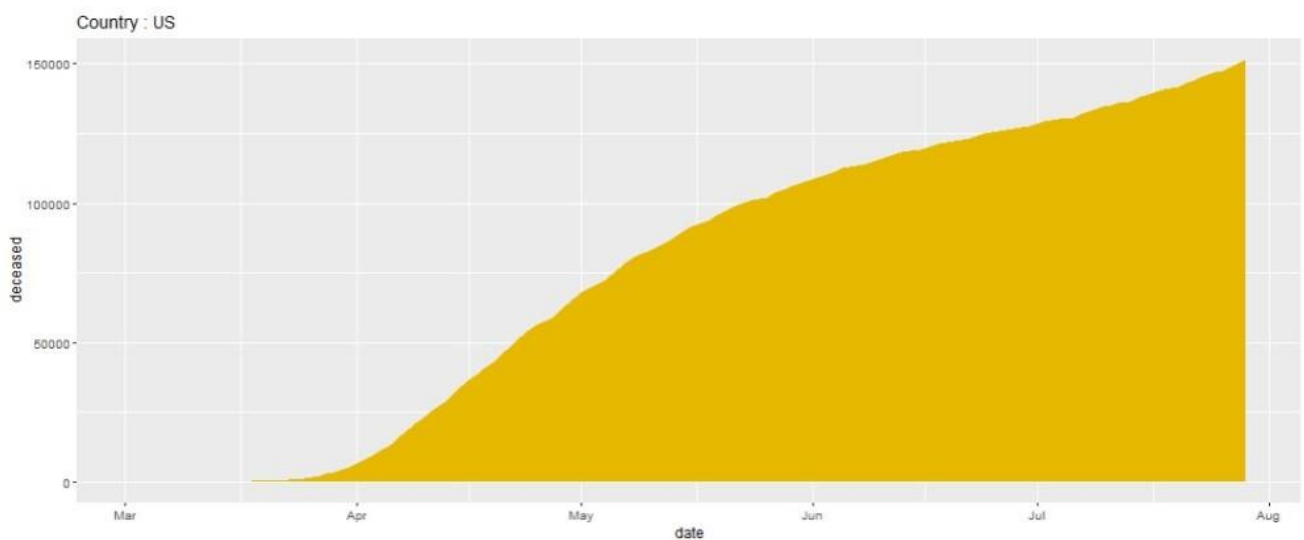
This area graph shows the total confirmed cases of covid-19 from the time cases were recorded for USA.

#### Active Cases:

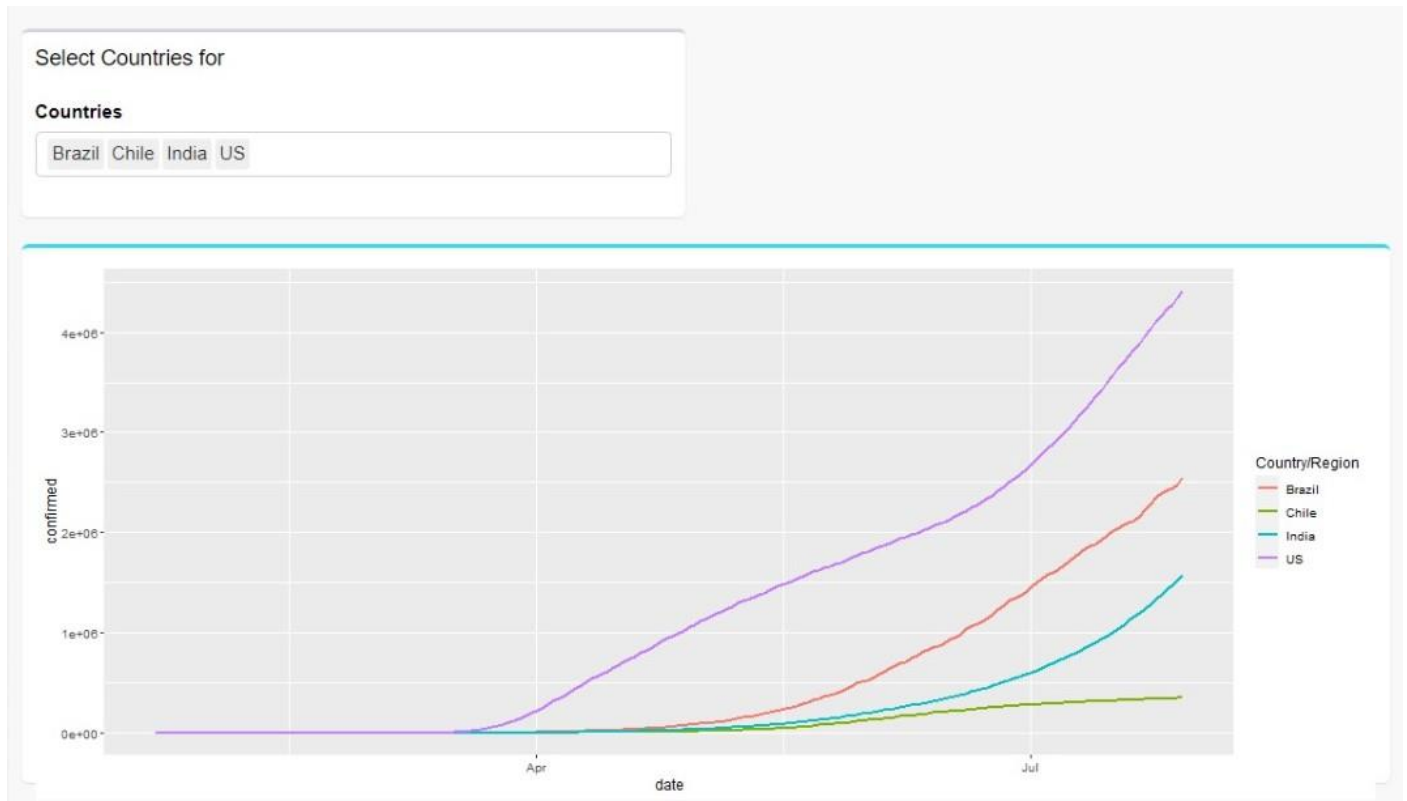


This area graph shows the total number of active cases of covid-19 from the time cases were recorded for USA.

#### Recovered Cases:



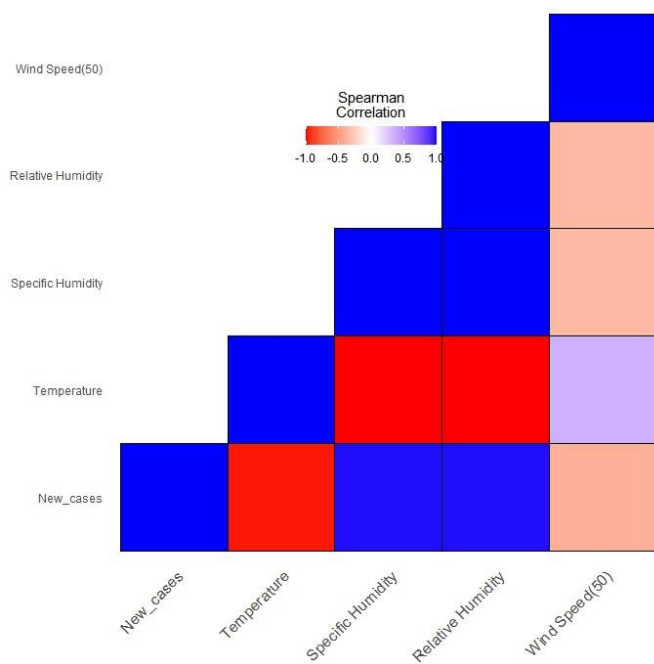
This area graph shows the total number of recovered cases of covid-19 from the time cases were recorded for USA.



This is a line chart that shows the gradual increase in Covid-19 cases for all the countries selected in the select bar.

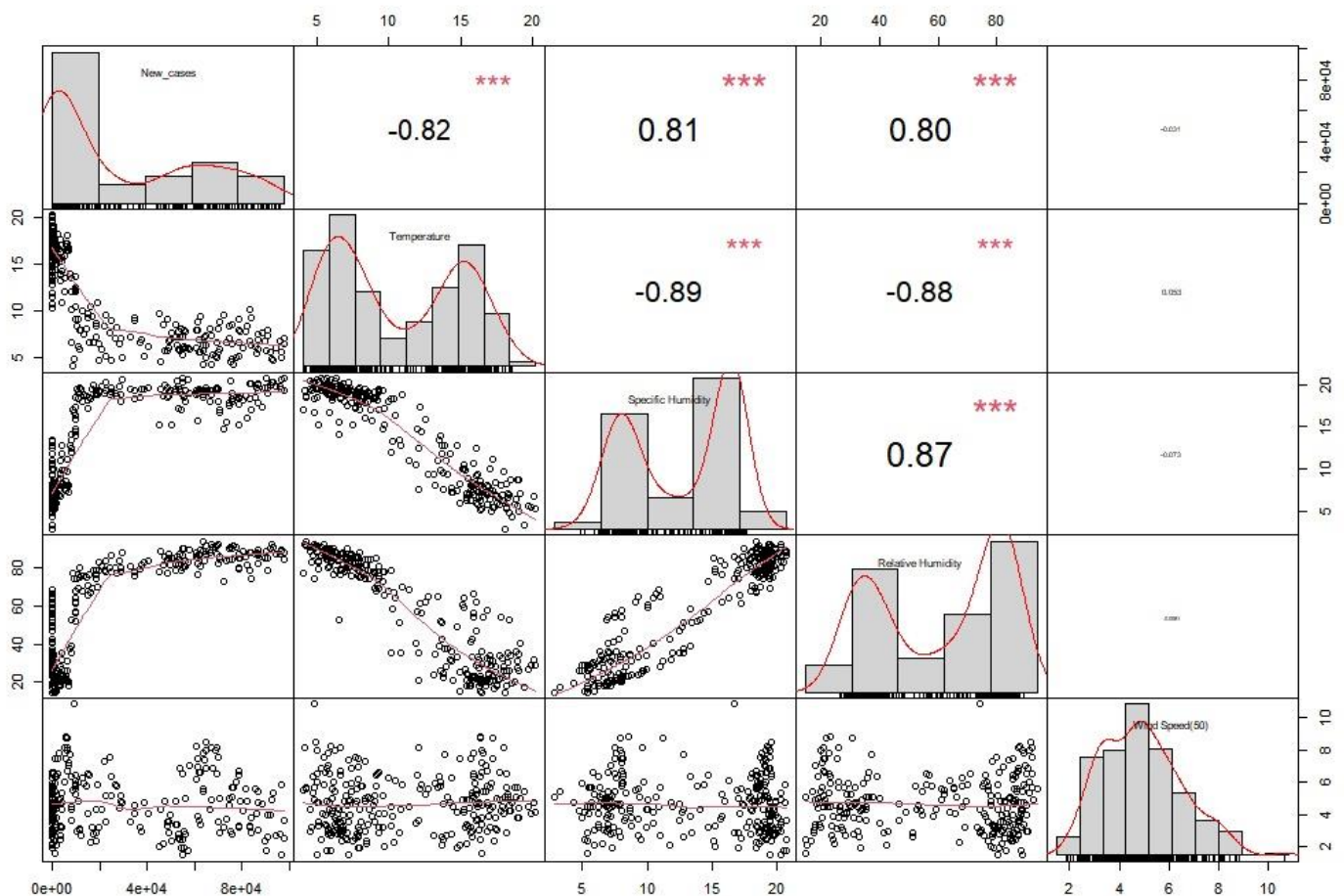
## Correlation Tab

### Heatmap



This is a heatmap showing the correlation between covid-19 cases and weather factors. Red color (-1) shows strong negative correlation and Blue color (+1) shows strong positive correlation. From the colors in the heatmap we can say that new cases and temperature perfect negative correlation and specific humidity and relative humidity shows perfect positive correlation and for wind speed it shows weak negative correlation.

## Correlation Matrix



In this table we can see that the correlation coefficient of new cases and temperature is -0.82 at 0.01 level of significance and correlation coefficient for specific humidity and relative humidity with new cases are 0.81 and 0.80 respectively at 0.01 level of significance and the correlation between new cases and windspeed is 0.031. From the below correlation values we can say that the temperature and covid-19 cases have strong negative correlation whereas specific humidity and relative humidity have strong positive correlation with covid-19 cases. Wind speed shows no correlation at all.

The histogram shows how the data is distributed. The red line does not form a bell shaped curve which indicates that our data doesn't follow normal distribution except wind speed.

From the scatter plot we can observe that the regression line goes from top left to bottom right for new cases vs temperature, which shows negative correlation. For both specific humidity and relative humidity with new cases the line goes from bottom left to top right which indicates positive correlation. For windspeed vs new cases, it is a horizontal line therefore shows no correlation.

## **Conclusion**

From this project we've learnt that there is a strong correlation between humidity and covid-19 cases. However, correlation between temperature and wind speed with covid-19 cases is very low. Through this, we can say that we can expect there to be an increase in covid-19 cases during summer seasons as humidity is more then.