COVID-19 Data Analysis and Forecasting

Mark

2025-04-15

## Install and Load Required Packages

# install.packages("plotly")  
library(tidyverse)  
library(dplyr)  
library(lubridate)  
library(forecast)  
library(ggplot2)  
library(scales)  
library(tseries)

## Load the Dataset

covid\_data <- read.table('https://catalog.ourworldindata.org/garden/covid/latest/compact/compact.csv', sep=',', head=TRUE)

## Explore the Dataset

head(covid\_data, 5)

## country date total\_cases new\_cases new\_cases\_smoothed  
## 1 Afghanistan 2020-01-01 NA NA NA  
## 2 Afghanistan 2020-01-02 NA NA NA  
## 3 Afghanistan 2020-01-03 NA NA NA  
## 4 Afghanistan 2020-01-04 0 0 NA  
## 5 Afghanistan 2020-01-05 0 0 NA  
## total\_cases\_per\_million new\_cases\_per\_million new\_cases\_smoothed\_per\_million  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 0 0 NA  
## 5 0 0 NA  
## total\_deaths new\_deaths new\_deaths\_smoothed total\_deaths\_per\_million  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 0 0 NA 0  
## 5 0 0 NA 0  
## new\_deaths\_per\_million new\_deaths\_smoothed\_per\_million excess\_mortality  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 0 NA NA  
## 5 0 NA NA  
## excess\_mortality\_cumulative excess\_mortality\_cumulative\_absolute  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## excess\_mortality\_cumulative\_per\_million hosp\_patients  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## hosp\_patients\_per\_million weekly\_hosp\_admissions  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## weekly\_hosp\_admissions\_per\_million icu\_patients icu\_patients\_per\_million  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## weekly\_icu\_admissions weekly\_icu\_admissions\_per\_million stringency\_index  
## 1 NA NA 0  
## 2 NA NA 0  
## 3 NA NA 0  
## 4 NA NA 0  
## 5 NA NA 0  
## reproduction\_rate total\_tests new\_tests total\_tests\_per\_thousand  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## new\_tests\_per\_thousand new\_tests\_smoothed new\_tests\_smoothed\_per\_thousand  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## positive\_rate tests\_per\_case total\_vaccinations people\_vaccinated  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## people\_fully\_vaccinated total\_boosters new\_vaccinations  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## new\_vaccinations\_smoothed total\_vaccinations\_per\_hundred  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## people\_vaccinated\_per\_hundred people\_fully\_vaccinated\_per\_hundred  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## total\_boosters\_per\_hundred new\_vaccinations\_smoothed\_per\_million  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## new\_people\_vaccinated\_smoothed new\_people\_vaccinated\_smoothed\_per\_hundred  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## code continent population population\_density median\_age life\_expectancy  
## 1 AFG Asia 40578801 62.21548 16.752 NA  
## 2 AFG Asia 40578801 62.21548 16.752 NA  
## 3 AFG Asia 40578801 62.21548 16.752 NA  
## 4 AFG Asia 40578801 62.21548 16.752 NA  
## 5 AFG Asia 40578801 62.21548 16.752 NA  
## gdp\_per\_capita extreme\_poverty diabetes\_prevalence handwashing\_facilities  
## 1 1516.273 NA 10.9 48.21469  
## 2 1516.273 NA 10.9 48.21469  
## 3 1516.273 NA 10.9 48.21469  
## 4 1516.273 NA 10.9 48.21469  
## 5 1516.273 NA 10.9 48.21469  
## hospital\_beds\_per\_thousand human\_development\_index  
## 1 0.39 0.462  
## 2 0.39 0.462  
## 3 0.39 0.462  
## 4 0.39 0.462  
## 5 0.39 0.462

tail(covid\_data, 10)

## country date total\_cases new\_cases new\_cases\_smoothed  
## 491236 Zimbabwe 2025-03-21 266410 0 0  
## 491237 Zimbabwe 2025-03-22 266410 0 0  
## 491238 Zimbabwe 2025-03-23 266410 0 0  
## 491239 Zimbabwe 2025-03-24 266410 0 0  
## 491240 Zimbabwe 2025-03-25 266410 0 0  
## 491241 Zimbabwe 2025-03-26 266410 0 0  
## 491242 Zimbabwe 2025-03-27 266410 0 0  
## 491243 Zimbabwe 2025-03-28 266410 0 0  
## 491244 Zimbabwe 2025-03-29 266410 0 0  
## 491245 Zimbabwe 2025-03-30 266410 0 0  
## total\_cases\_per\_million new\_cases\_per\_million  
## 491236 16579.12 0  
## 491237 16579.12 0  
## 491238 16579.12 0  
## 491239 16579.12 0  
## 491240 16579.12 0  
## 491241 16579.12 0  
## 491242 16579.12 0  
## 491243 16579.12 0  
## 491244 16579.12 0  
## 491245 16579.12 0  
## new\_cases\_smoothed\_per\_million total\_deaths new\_deaths  
## 491236 0 5740 0  
## 491237 0 5740 0  
## 491238 0 5740 0  
## 491239 0 5740 0  
## 491240 0 5740 0  
## 491241 0 5740 0  
## 491242 0 5740 0  
## 491243 0 5740 0  
## 491244 0 5740 0  
## 491245 0 5740 0  
## new\_deaths\_smoothed total\_deaths\_per\_million new\_deaths\_per\_million  
## 491236 0 357.2093 0  
## 491237 0 357.2093 0  
## 491238 0 357.2093 0  
## 491239 0 357.2093 0  
## 491240 0 357.2093 0  
## 491241 0 357.2093 0  
## 491242 0 357.2093 0  
## 491243 0 357.2093 0  
## 491244 0 357.2093 0  
## 491245 0 357.2093 0  
## new\_deaths\_smoothed\_per\_million excess\_mortality  
## 491236 0 NA  
## 491237 0 NA  
## 491238 0 NA  
## 491239 0 NA  
## 491240 0 NA  
## 491241 0 NA  
## 491242 0 NA  
## 491243 0 NA  
## 491244 0 NA  
## 491245 0 NA  
## excess\_mortality\_cumulative excess\_mortality\_cumulative\_absolute  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## excess\_mortality\_cumulative\_per\_million hosp\_patients  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## hosp\_patients\_per\_million weekly\_hosp\_admissions  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## weekly\_hosp\_admissions\_per\_million icu\_patients icu\_patients\_per\_million  
## 491236 NA NA NA  
## 491237 NA NA NA  
## 491238 NA NA NA  
## 491239 NA NA NA  
## 491240 NA NA NA  
## 491241 NA NA NA  
## 491242 NA NA NA  
## 491243 NA NA NA  
## 491244 NA NA NA  
## 491245 NA NA NA  
## weekly\_icu\_admissions weekly\_icu\_admissions\_per\_million stringency\_index  
## 491236 NA NA NA  
## 491237 NA NA NA  
## 491238 NA NA NA  
## 491239 NA NA NA  
## 491240 NA NA NA  
## 491241 NA NA NA  
## 491242 NA NA NA  
## 491243 NA NA NA  
## 491244 NA NA NA  
## 491245 NA NA NA  
## reproduction\_rate total\_tests new\_tests total\_tests\_per\_thousand  
## 491236 NA NA NA NA  
## 491237 NA NA NA NA  
## 491238 NA NA NA NA  
## 491239 NA NA NA NA  
## 491240 NA NA NA NA  
## 491241 NA NA NA NA  
## 491242 NA NA NA NA  
## 491243 NA NA NA NA  
## 491244 NA NA NA NA  
## 491245 NA NA NA NA  
## new\_tests\_per\_thousand new\_tests\_smoothed  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## new\_tests\_smoothed\_per\_thousand positive\_rate tests\_per\_case  
## 491236 NA NA NA  
## 491237 NA NA NA  
## 491238 NA NA NA  
## 491239 NA NA NA  
## 491240 NA NA NA  
## 491241 NA NA NA  
## 491242 NA NA NA  
## 491243 NA NA NA  
## 491244 NA NA NA  
## 491245 NA NA NA  
## total\_vaccinations people\_vaccinated people\_fully\_vaccinated  
## 491236 NA NA NA  
## 491237 NA NA NA  
## 491238 NA NA NA  
## 491239 NA NA NA  
## 491240 NA NA NA  
## 491241 NA NA NA  
## 491242 NA NA NA  
## 491243 NA NA NA  
## 491244 NA NA NA  
## 491245 NA NA NA  
## total\_boosters new\_vaccinations new\_vaccinations\_smoothed  
## 491236 NA NA NA  
## 491237 NA NA NA  
## 491238 NA NA NA  
## 491239 NA NA NA  
## 491240 NA NA NA  
## 491241 NA NA NA  
## 491242 NA NA NA  
## 491243 NA NA NA  
## 491244 NA NA NA  
## 491245 NA NA NA  
## total\_vaccinations\_per\_hundred people\_vaccinated\_per\_hundred  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## people\_fully\_vaccinated\_per\_hundred total\_boosters\_per\_hundred  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## new\_vaccinations\_smoothed\_per\_million new\_people\_vaccinated\_smoothed  
## 491236 NA NA  
## 491237 NA NA  
## 491238 NA NA  
## 491239 NA NA  
## 491240 NA NA  
## 491241 NA NA  
## 491242 NA NA  
## 491243 NA NA  
## 491244 NA NA  
## 491245 NA NA  
## new\_people\_vaccinated\_smoothed\_per\_hundred code continent population  
## 491236 NA ZWE Africa 16069010  
## 491237 NA ZWE Africa 16069010  
## 491238 NA ZWE Africa 16069010  
## 491239 NA ZWE Africa 16069010  
## 491240 NA ZWE Africa 16069010  
## 491241 NA ZWE Africa 16069010  
## 491242 NA ZWE Africa 16069010  
## 491243 NA ZWE Africa 16069010  
## 491244 NA ZWE Africa 16069010  
## 491245 NA ZWE Africa 16069010  
## population\_density median\_age life\_expectancy gdp\_per\_capita  
## 491236 41.53809 17.683 NA 2207.957  
## 491237 41.53809 17.683 NA 2207.957  
## 491238 41.53809 17.683 NA 2207.957  
## 491239 41.53809 17.683 NA 2207.957  
## 491240 41.53809 17.683 NA 2207.957  
## 491241 41.53809 17.683 NA 2207.957  
## 491242 41.53809 17.683 NA 2207.957  
## 491243 41.53809 17.683 NA 2207.957  
## 491244 41.53809 17.683 NA 2207.957  
## 491245 41.53809 17.683 NA 2207.957  
## extreme\_poverty diabetes\_prevalence handwashing\_facilities  
## 491236 39.75453 2.1 42.45609  
## 491237 39.75453 2.1 42.45609  
## 491238 39.75453 2.1 42.45609  
## 491239 39.75453 2.1 42.45609  
## 491240 39.75453 2.1 42.45609  
## 491241 39.75453 2.1 42.45609  
## 491242 39.75453 2.1 42.45609  
## 491243 39.75453 2.1 42.45609  
## 491244 39.75453 2.1 42.45609  
## 491245 39.75453 2.1 42.45609  
## hospital\_beds\_per\_thousand human\_development\_index  
## 491236 1.7 0.55  
## 491237 1.7 0.55  
## 491238 1.7 0.55  
## 491239 1.7 0.55  
## 491240 1.7 0.55  
## 491241 1.7 0.55  
## 491242 1.7 0.55  
## 491243 1.7 0.55  
## 491244 1.7 0.55  
## 491245 1.7 0.55

#sample(covid\_data)  
dim(covid\_data)

## [1] 491245 61

names(covid\_data)

## [1] "country"   
## [2] "date"   
## [3] "total\_cases"   
## [4] "new\_cases"   
## [5] "new\_cases\_smoothed"   
## [6] "total\_cases\_per\_million"   
## [7] "new\_cases\_per\_million"   
## [8] "new\_cases\_smoothed\_per\_million"   
## [9] "total\_deaths"   
## [10] "new\_deaths"   
## [11] "new\_deaths\_smoothed"   
## [12] "total\_deaths\_per\_million"   
## [13] "new\_deaths\_per\_million"   
## [14] "new\_deaths\_smoothed\_per\_million"   
## [15] "excess\_mortality"   
## [16] "excess\_mortality\_cumulative"   
## [17] "excess\_mortality\_cumulative\_absolute"   
## [18] "excess\_mortality\_cumulative\_per\_million"   
## [19] "hosp\_patients"   
## [20] "hosp\_patients\_per\_million"   
## [21] "weekly\_hosp\_admissions"   
## [22] "weekly\_hosp\_admissions\_per\_million"   
## [23] "icu\_patients"   
## [24] "icu\_patients\_per\_million"   
## [25] "weekly\_icu\_admissions"   
## [26] "weekly\_icu\_admissions\_per\_million"   
## [27] "stringency\_index"   
## [28] "reproduction\_rate"   
## [29] "total\_tests"   
## [30] "new\_tests"   
## [31] "total\_tests\_per\_thousand"   
## [32] "new\_tests\_per\_thousand"   
## [33] "new\_tests\_smoothed"   
## [34] "new\_tests\_smoothed\_per\_thousand"   
## [35] "positive\_rate"   
## [36] "tests\_per\_case"   
## [37] "total\_vaccinations"   
## [38] "people\_vaccinated"   
## [39] "people\_fully\_vaccinated"   
## [40] "total\_boosters"   
## [41] "new\_vaccinations"   
## [42] "new\_vaccinations\_smoothed"   
## [43] "total\_vaccinations\_per\_hundred"   
## [44] "people\_vaccinated\_per\_hundred"   
## [45] "people\_fully\_vaccinated\_per\_hundred"   
## [46] "total\_boosters\_per\_hundred"   
## [47] "new\_vaccinations\_smoothed\_per\_million"   
## [48] "new\_people\_vaccinated\_smoothed"   
## [49] "new\_people\_vaccinated\_smoothed\_per\_hundred"  
## [50] "code"   
## [51] "continent"   
## [52] "population"   
## [53] "population\_density"   
## [54] "median\_age"   
## [55] "life\_expectancy"   
## [56] "gdp\_per\_capita"   
## [57] "extreme\_poverty"   
## [58] "diabetes\_prevalence"   
## [59] "handwashing\_facilities"   
## [60] "hospital\_beds\_per\_thousand"   
## [61] "human\_development\_index"

glimpse(covid\_data)

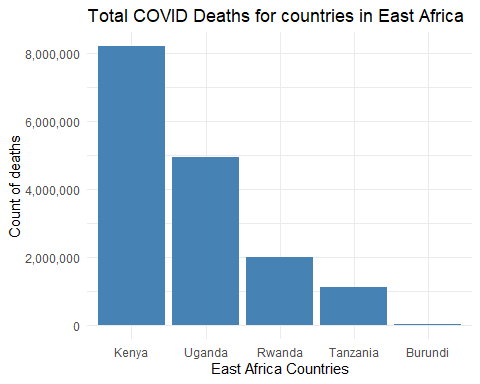
## Rows: 491,245  
## Columns: 61  
## $ country <chr> "Afghanistan", "Afghanistan…  
## $ date <chr> "2020-01-01", "2020-01-02",…  
## $ total\_cases <int> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_cases <int> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_cases\_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_cases\_per\_million <dbl> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_cases\_per\_million <dbl> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_cases\_smoothed\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_deaths <int> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_deaths <int> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_deaths\_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_deaths\_per\_million <dbl> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_deaths\_per\_million <dbl> NA, NA, NA, 0, 0, 0, 0, 0, …  
## $ new\_deaths\_smoothed\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ excess\_mortality <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ excess\_mortality\_cumulative <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ excess\_mortality\_cumulative\_absolute <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ excess\_mortality\_cumulative\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ hosp\_patients <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ hosp\_patients\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ weekly\_hosp\_admissions <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ weekly\_hosp\_admissions\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ icu\_patients <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ icu\_patients\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ weekly\_icu\_admissions <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ weekly\_icu\_admissions\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ stringency\_index <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ reproduction\_rate <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_tests <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_tests <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_tests\_per\_thousand <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_tests\_per\_thousand <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_tests\_smoothed <int> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_tests\_smoothed\_per\_thousand <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ positive\_rate <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ tests\_per\_case <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_vaccinations <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ people\_vaccinated <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ people\_fully\_vaccinated <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_boosters <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_vaccinations <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_vaccinations\_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_vaccinations\_per\_hundred <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ people\_vaccinated\_per\_hundred <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ people\_fully\_vaccinated\_per\_hundred <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ total\_boosters\_per\_hundred <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_vaccinations\_smoothed\_per\_million <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_people\_vaccinated\_smoothed <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ new\_people\_vaccinated\_smoothed\_per\_hundred <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ code <chr> "AFG", "AFG", "AFG", "AFG",…  
## $ continent <chr> "Asia", "Asia", "Asia", "As…  
## $ population <dbl> 40578801, 40578801, 4057880…  
## $ population\_density <dbl> 62.21548, 62.21548, 62.2154…  
## $ median\_age <dbl> 16.752, 16.752, 16.752, 16.…  
## $ life\_expectancy <lgl> NA, NA, NA, NA, NA, NA, NA,…  
## $ gdp\_per\_capita <dbl> 1516.273, 1516.273, 1516.27…  
## $ extreme\_poverty <dbl> NA, NA, NA, NA, NA, NA, NA,…  
## $ diabetes\_prevalence <dbl> 10.9, 10.9, 10.9, 10.9, 10.…  
## $ handwashing\_facilities <dbl> 48.21469, 48.21469, 48.2146…  
## $ hospital\_beds\_per\_thousand <dbl> 0.39, 0.39, 0.39, 0.39, 0.3…  
## $ human\_development\_index <dbl> 0.462, 0.462, 0.462, 0.462,…

## Clean the Data

covid\_data$date <- as.Date(covid\_data$date, format = "%Y-%m-%d")  
complete\_covid\_data <- covid\_data[!is.na(covid\_data$total\_cases), ]

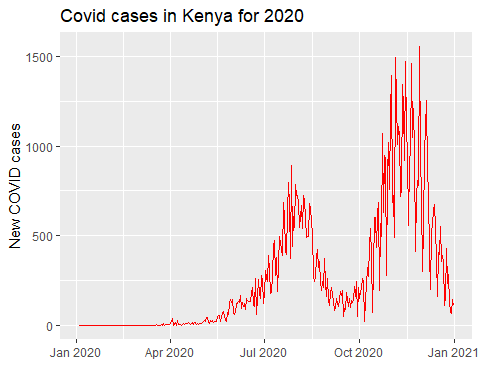
## Deaths in East Africa

complete\_covid\_data %>%   
 filter(country %in% c('Kenya','Uganda', 'Tanzania', 'Rwanda', 'Burundi')) %>%   
 ggplot(aes(x = reorder(country, -total\_deaths), y = total\_deaths)) +  
 geom\_bar(stat = "identity", fill = "steelblue") +  
 labs(title = "Total COVID Deaths for countries in East Africa", x = "East Africa Countries", y = "Count of deaths") +  
 scale\_y\_continuous(labels = comma) +  
 theme\_minimal()



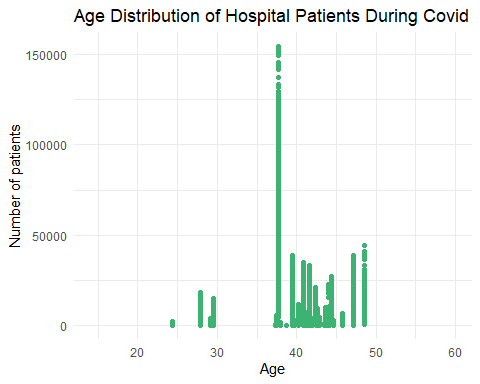
## COVID-19 Cases in Kenya (2020)

complete\_covid\_data %>%   
 filter(country == "Kenya", date >= "2020-01-01" & date <= "2020-12-31") %>%   
 ggplot(aes(x = date, y = new\_cases)) +  
 geom\_line(linewidth = 0.5, color = 'red') +  
 labs(title = "Covid cases in Kenya for 2020", x = "", y = "New COVID cases")



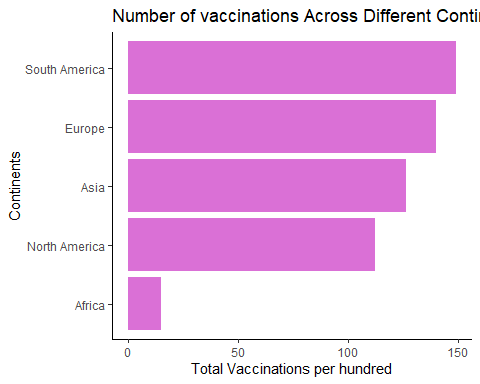
## Age Distribution of Hospital Patients During COVID

complete\_covid\_data %>%   
 ggplot(aes(x = median\_age, y = hosp\_patients)) +  
 geom\_point(color = 'mediumseagreen') +  
 labs(title = "Age Distribution of Hospital Patients During Covid", x = 'Age', y = 'Number of patients') +  
 theme\_minimal()



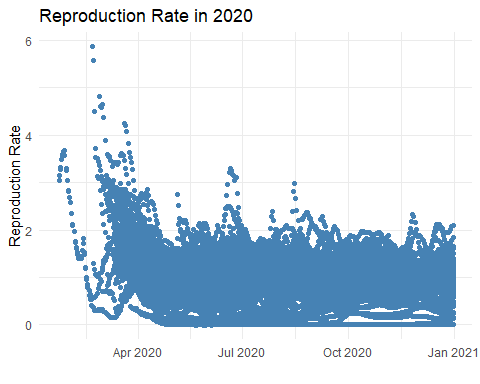
## Vaccinations Across Continents

complete\_covid\_data %>%   
 filter(continent %in% c("Africa", "Asia", "Europe", "North America", "South America"), !is.na(total\_vaccinations\_per\_hundred)) %>%   
 group\_by(continent) %>%  
 summarise(med\_vaccinations = median(total\_vaccinations\_per\_hundred, na.rm = TRUE)) %>%   
 ggplot(aes(x = reorder(continent, med\_vaccinations), y = med\_vaccinations)) +  
 geom\_bar(stat = "identity", fill = 'orchid') +  
 labs(title = "Number of vaccinations Across Different Continents", x = "Continents", y = "Total Vaccinations per hundred") +  
 coord\_flip() +  
 scale\_y\_continuous(labels = comma) +  
 theme\_classic()



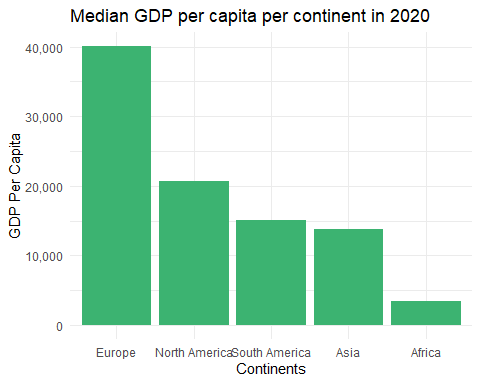
## Reproduction Rate in 2020

complete\_covid\_data %>%   
 filter(date >= "2020-01-01" & date <= "2020-12-31") %>%   
 filter(!is.na(reproduction\_rate)) %>%   
 ggplot(aes(x = date, y = reproduction\_rate)) +  
 geom\_point(color = "steelblue") +  
 labs(title = "Reproduction Rate in 2020", x = '', y = 'Reproduction Rate') +  
 theme\_minimal()



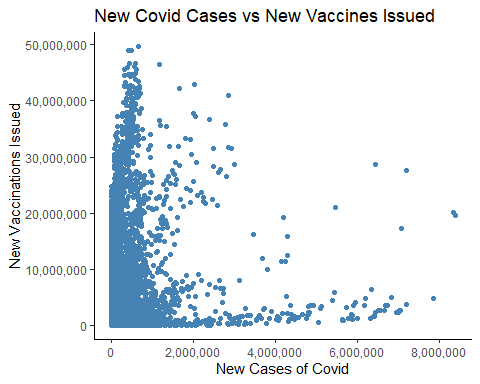
## GDP per Capita by Continent

complete\_covid\_data %>%   
 filter(continent %in% c("Africa", "Asia", "Europe", "North America", "South America")) %>%  
 group\_by(continent) %>%   
 summarise(med\_gdp = median(gdp\_per\_capita, na.rm = TRUE)) %>%  
 ggplot(aes(x = reorder(continent, -med\_gdp), y = med\_gdp)) +  
 geom\_bar(stat = "identity", fill = 'mediumseagreen') +  
 labs(title = "Median GDP per capita per continent in 2020", x = 'Continents', y = 'GDP Per Capita') +  
 scale\_y\_continuous(labels = comma) +  
 theme\_minimal()



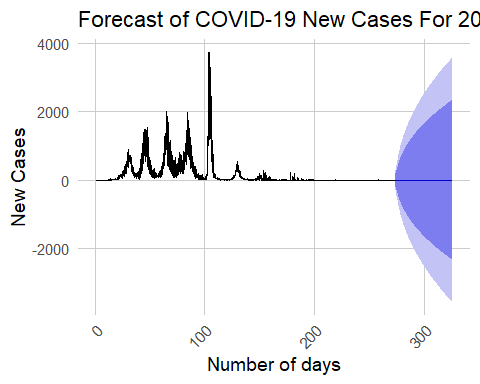
## New COVID Cases vs New Vaccinations

complete\_covid\_data %>%   
 ggplot(aes(x = new\_cases, y = new\_vaccinations)) +  
 geom\_point(color = 'steelblue') +  
 labs(title = "New Covid Cases vs New Vaccines Issued", x = 'New Cases of Covid', y = 'New Vaccinations Issued') +  
 scale\_y\_continuous(labels = comma) +  
 scale\_x\_continuous(labels = comma) +  
 theme\_classic()



## Forecasting COVID-19 Cases for Kenya (2025)

kenya\_cases <- complete\_covid\_data %>%  
 filter(country == "Kenya") %>%  
 select(date, new\_cases) %>%  
 drop\_na()  
  
cases\_ts <- ts(kenya\_cases$new\_cases, frequency = 7)  
model <- auto.arima(cases\_ts)  
forecast\_data <- forecast(model, h = 365)  
  
autoplot(forecast\_data) +  
 labs(title = "Forecast of COVID-19 New Cases For 2025", y = "New Cases", x = "Number of days") +  
 theme\_minimal(base\_size = 14) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1),  
 panel.grid.major = element\_line(color = "gray80"),  
 panel.grid.minor = element\_blank())



## Export Forecast to CSV

forecast\_df <- data.frame(  
 date = seq(from = max(kenya\_cases$date) + 1, by = "day", length.out = 365),  
 predicted\_cases = as.numeric(forecast\_data$mean),  
 lower\_80 = forecast\_data$lower[, 1],  
 upper\_80 = forecast\_data$upper[, 1],  
 lower\_95 = forecast\_data$lower[, 2],  
 upper\_95 = forecast\_data$upper[, 2]  
)  
  
write.csv(forecast\_df, "kenya\_covid\_forecast\_2025.csv", row.names = FALSE)