

Team 26 Stat 3355 Project

Kenan Stredic, Roopali Kallem

2022-11-02

importing necessary libraries

```
library(ggplot2)
```

```
library(gridExtra)
```

reading the csv files into data frames

```
# reading the files into data frames
```

```
world_covid <- read.csv( "WHO C-19 cummulative.csv" )  
air_traffic <- read.csv( "c-19 impact on air traffic.csv" )  
world_vacc <- read.csv( "vaccination-data.csv" )
```

Data Pre-Processing

```
# converting dates in the data frames from character type to date type.
```

```
# Factor Date_reported in world_covid
```

```
world_covid_Date_reported_factored <- as.factor( world_covid$Date_reported )
```

```
# Convert Date_reported from character to POSIXlt in world_covid
```

```
world_covid_Date_reported_POSIXlt <- strptime( world_covid_Date_reported_factored, format = "%m/%d/%Y" )
```

```
# Convert Date_reported from POSIXlt to Date in world_covid
```

```
world_covid$Date_reported <- as.Date( world_covid_Date_reported_POSIXlt, format = "%Y-%m-%d" )
```

```
str( world_covid ) # printing to see the data type of each variable
```

```
## 'data.frame': 245058 obs. of 8 variables:  
## $ Date_reported : Date, format: "2020-01-03" "2020-01-04" ...  
## $ Country_code : chr "AF" "AF" "AF" "AF" ...  
## $ Country : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...  
## $ WHO_region : chr "EMRO" "EMRO" "EMRO" "EMRO" ...  
## $ New_cases : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Cumulative_cases : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ New_deaths : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Cumulative_deaths: int 0 0 0 0 0 0 0 0 0 0 ...
```

```
# Removing un-necessary columns from world_covid
```

```
world_covid <- world_covid[ -c( 2, 4 ) ]
```

```
str( world_covid ) # printing again to see if correct columns are removed and to confirm data type of e
```

```
## 'data.frame': 245058 obs. of 6 variables:  
## $ Date_reported : Date, format: "2020-01-03" "2020-01-04" ...
```

```

## $ Country      : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ New_cases    : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Cumulative_cases : int   0 0 0 0 0 0 0 0 0 0 ...
## $ New_deaths    : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Cumulative_deaths: int   0 0 0 0 0 0 0 0 0 0 ...

# checking for any possible null values

sum( is.na( world_covid$Date_reported ) )

## [1] 0

sum( is.na( world_covid$Country ) )

## [1] 0

sum( is.na( world_covid$New_cases ) )

## [1] 0

sum( is.na( world_covid$Cumulative_cases ) )

## [1] 0

sum( is.na( world_covid$New_deaths ) )

## [1] 0

sum( is.na( world_covid$Cumulative_deaths ) )

## [1] 0

# converting dates in the data frames from character type to date type.

# Factor Date_reported in air_traffic
air_traffic_Date_factored <- as.factor( air_traffic$Date )

# Convert Date from character to POSIXlt in air_traffic
air_traffic_Date_POSIXlt <- strptime( air_traffic_Date_factored, format = "%m/%d/%Y" )

# Convert Date from POSIXlt to Date in air_traffic
air_traffic$Date <- as.Date( air_traffic_Date_POSIXlt, format = "%Y-%m-%d" )

str( air_traffic ) # printing to see the data type of each variable

## 'data.frame': 7247 obs. of 11 variables:
## $ AggregationMethod: chr  "Daily" "Daily" "Daily" "Daily" ...
## $ Date              : Date, format: "2020-04-03" "2020-04-13" ...
## $ Version           : int   1 1 1 1 1 1 1 1 1 1 ...
## $ AirportName       : chr  "Kingsford Smith" "Kingsford Smith" "Kingsford Smith" "Kingsford Smith" ...
## $ PercentOfBaseline: int   64 29 54 18 22 59 59 48 20 27 ...
## $ Centroid          : chr  "POINT(151.180087713813 -33.9459774986125)" "POINT(151.180087713813 -33.9459774986125)" ...
## $ City              : chr  "Sydney" "Sydney" "Sydney" "Sydney" ...
## $ State             : chr  "New South Wales" "New South Wales" "New South Wales" "New South Wales" ...
## $ ISO_3166_2        : chr  "AU" "AU" "AU" "AU" ...
## $ Country           : chr  "Australia" "Australia" "Australia" "Australia" ...
## $ Geography         : chr  "POLYGON((151.164354085922 -33.9301772341877, 151.163324117661 -33.931485117661, 151.163324117661 -33.931485117661, 151.164354085922 -33.9301772341877, 151.163324117661 -33.931485117661, 151.163324117661 -33.931485117661, 151.164354085922 -33.9301772341877, 151.163324117661 -33.931485117661, 151.163324117661 -33.931485117661, 151.164354085922 -33.9301772341877))" ...

# Removing un-necessary columns from air_traffic
air_traffic <- air_traffic[ -c( 1, 3, 4, 6, 9, 11 ) ]

```

```

str( air_traffic ) # printing again to see if correct columns are removed and to confirm data type of

## 'data.frame': 7247 obs. of 5 variables:
## $ Date : Date, format: "2020-04-03" "2020-04-13" ...
## $ PercentOfBaseline: int 64 29 54 18 22 59 59 48 20 27 ...
## $ City : chr "Sydney" "Sydney" "Sydney" "Sydney" ...
## $ State : chr "New South Wales" "New South Wales" "New South Wales" "New South Wales" ...
## $ Country : chr "Australia" "Australia" "Australia" "Australia" ...

# checking for any possible null values

sum( is.na( air_traffic$Date ) )

## [1] 0

sum( is.na( air_traffic$Country ) )

## [1] 0

sum( is.na( air_traffic$City ) )

## [1] 0

sum( is.na( air_traffic$State ) )

## [1] 0

# Factor Date_updated in world_vacc
world_vacc_Date_Updated_factored <- as.factor( world_vacc$DATE_UPDATED )

# Convert Date_updated from character to POSIXlt in world_vacc
world_vacc_Date_Updated_POSIXlt <- strptime( world_vacc_Date_Updated_factored, format = "%m/%d/%Y" )

# Convert Date_updated from POSIXlt to Date in world_vacc
world_vacc$DATE_UPDATED <- as.Date( world_vacc_Date_Updated_POSIXlt, format = "%Y-%m-%d" )

# Factor FIRST_VACCINE_DATE in world_vacc
world_vacc_FIRST_VACCINE_DATE_factored <- as.factor( world_vacc$FIRST_VACCINE_DATE )

# Convert FIRST_VACCINE_DATE from character to POSIXlt in world_vacc
world_vacc_FIRST_VACCINE_DATE_POSIXlt <- strptime( world_vacc_FIRST_VACCINE_DATE_factored, format = "%m/%d/%Y" )

# Convert FIRST_VACCINE_DATE from POSIXlt to Date in world_vacc
world_vacc$FIRST_VACCINE_DATE <- as.Date( world_vacc_FIRST_VACCINE_DATE_POSIXlt, format = "%Y-%m-%d" )

str( world_vacc ) # printing to see the data type of each variable

## 'data.frame': 229 obs. of 16 variables:
## $ COUNTRY : chr "Afghanistan" "Albania" "Algeria" "American Samoa" ...
## $ ISO3 : chr "AFG" "ALB" "DZA" "ASM" ...
## $ WHO_REGION : chr "EMRO" "EURO" "AFRO" "WPRO" ...
## $ DATA_SOURCE : chr "REPORTING" "REPORTING" "REPORTING" "REPORTING" ...
## $ DATE_UPDATED : Date, format: "2022-11-02" "2022-10-16" ...
## $ TOTAL_VACCINATIONS : num 12140367 2991576 15267442 111316 154320 ...
## $ PERSONS_VACCINATED_1PLUS_DOSE : int 11165700 1339100 7840131 44885 57898 14220830 10852 64 ...
## $ TOTAL_VACCINATIONS_PER100 : num 31.2 104 34.8 201.7 199.7 ...

```

```
## $ PERSONS_VACCINATED_1PLUS_DOSE_PER100: num 28.7 47.1 17.9 81.3 76 ...
## $ PERSONS_FULLY_VACCINATED : int 10458947 1265900 6481186 41423 53482 7814121 10366 623
## $ PERSONS_FULLY_VACCINATED_PER100 : num 26.9 44.5 14.8 75 70.2 ...
## $ VACCINES_USED : chr "AstraZeneca - Vaxzevria,Beijing CNBG - BBIBP-CorV,Bha
## $ FIRST_VACCINE_DATE : Date, format: "2021-02-22" "2021-01-13" ...
## $ NUMBER_VACCINES_TYPES_USED : int 11 5 4 3 3 1 2 6 7 8 ...
## $ PERSONS_BOOSTER_ADD_DOSE : int NA 363122 575651 24160 42940 1127156 2998 9838 3081018
## $ PERSONS_BOOSTER_ADD_DOSE_PER100 : num NA 12.76 1.31 43.77 56.37 ...
```

```
# Removing un-necessary columns from world_covid
```

```
world_vacc <- world_vacc[ -c( 2:4, 12:16 ) ]
```

```
str( world_vacc ) # printing again to see if correct columns are removed and to confirm data type of e
```

```
## 'data.frame': 229 obs. of 8 variables:
```

```
## $ COUNTRY : chr "Afghanistan" "Albania" "Algeria" "American Samoa" ...
## $ DATE_UPDATED : Date, format: "2022-11-02" "2022-10-16" ...
## $ TOTAL_VACCINATIONS : num 12140367 2991576 15267442 111316 154320 ...
## $ PERSONS_VACCINATED_1PLUS_DOSE : int 11165700 1339100 7840131 44885 57898 14220830 10852 64
## $ TOTAL_VACCINATIONS_PER100 : num 31.2 104 34.8 201.7 199.7 ...
## $ PERSONS_VACCINATED_1PLUS_DOSE_PER100: num 28.7 47.1 17.9 81.3 76 ...
## $ PERSONS_FULLY_VACCINATED : int 10458947 1265900 6481186 41423 53482 7814121 10366 623
## $ PERSONS_FULLY_VACCINATED_PER100 : num 26.9 44.5 14.8 75 70.2 ...
```

```
# checking for any possible null values
```

```
sum( is.na( world_vacc$DATE_UPDATED ) )
```

```
## [1] 0
```

```
sum( is.na( world_vacc$Country ) )
```

```
## [1] 0
```

```
sum( is.na( world_vacc$TOTAL_VACCINATIONS ) )
```

```
## [1] 1
```

```
sum( is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE ) )
```

```
## [1] 1
```

```
sum( is.na( world_vacc$TOTAL_VACCINATIONS_PER100 ) )
```

```
## [1] 2
```

```
sum( is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE_PER100 ) )
```

```
## [1] 1
```

```
sum( is.na( world_vacc$PERSONS_FULLY_VACCINATED ) )
```

```
## [1] 1
```

```
sum( is.na( world_vacc$PERSONS_FULLY_VACCINATED_PER100 ) )
```

```
## [1] 1
```

```
# replacing the Null values with the mean of the column
```

```
world_vacc$TOTAL_VACCINATIONS[ is.na( world_vacc$TOTAL_VACCINATIONS ) ] <- mean( world_vacc$TOTAL_VACCI
```

```

world_vacc$PERSONS_VACCINATED_1PLUS_DOSE[ is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE ) ] <- mean( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE )
world_vacc$TOTAL_VACCINATIONS_PER100[ is.na( world_vacc$TOTAL_VACCINATIONS_PER100 ) ] <- mean( world_vacc$TOTAL_VACCINATIONS_PER100 )
world_vacc$PERSONS_VACCINATED_1PLUS_DOSE_PER100[ is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE_PER100 ) ] <- mean( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE_PER100 )
world_vacc$PERSONS_FULLY_VACCINATED[ is.na( world_vacc$PERSONS_FULLY_VACCINATED ) ] <- mean( world_vacc$PERSONS_FULLY_VACCINATED )
world_vacc$PERSONS_FULLY_VACCINATED_PER100[ is.na( world_vacc$PERSONS_FULLY_VACCINATED_PER100 ) ] <- mean( world_vacc$PERSONS_FULLY_VACCINATED_PER100 )

# rechecking to see if any more null values

sum( is.na( world_vacc$DATE_UPDATED ) )

## [1] 0

sum( is.na( world_vacc$Country ) )

## [1] 0

sum( is.na( world_vacc$TOTAL_VACCINATIONS ) )

## [1] 0

sum( is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE ) )

## [1] 0

sum( is.na( world_vacc$TOTAL_VACCINATIONS_PER100 ) )

## [1] 0

sum( is.na( world_vacc$PERSONS_VACCINATED_1PLUS_DOSE_PER100 ) )

## [1] 0

sum( is.na( world_vacc$PERSONS_FULLY_VACCINATED ) )

## [1] 0

sum( is.na( world_vacc$PERSONS_FULLY_VACCINATED_PER100 ) )

## [1] 0

```

Data Visualization

Between the two years is there any pattern between the number of cases and the number of deaths

Between the two years is there any pattern between the number of cases and the number of deaths in US
Instead of using just the smooth line graph, we showed the line and point graph next to the smooth line graph

```

usa_covid <- world_covid[ which( world_covid$Country == "United States of America" ), ]

usa_deaths1 <- ggplot( data = usa_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in the USA" )

usa_deaths2 <- ggplot( data = usa_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in the USA" )

```

```

usa_cases1 <- ggplot( data = usa_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in the USA" )

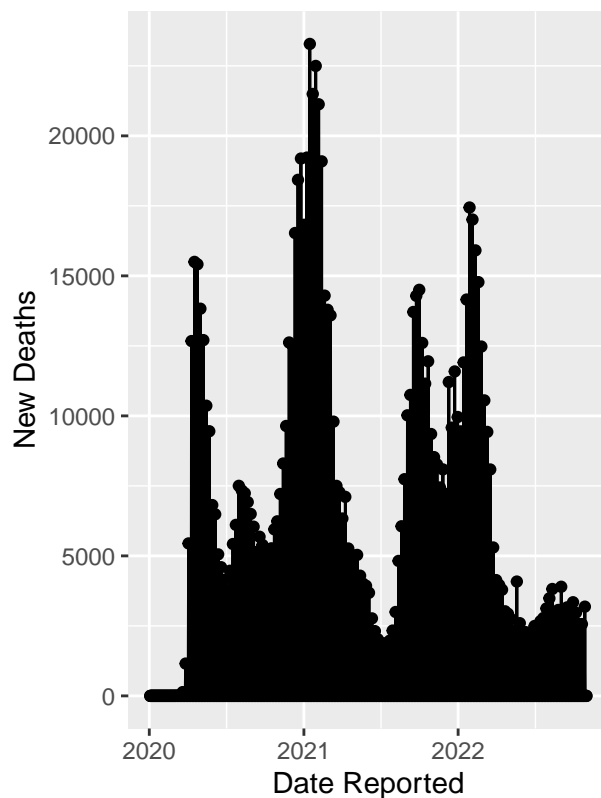
usa_cases2 <- ggplot( data = usa_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in the USA" )

grid.arrange( usa_deaths1, usa_deaths2, nrow = 1 )

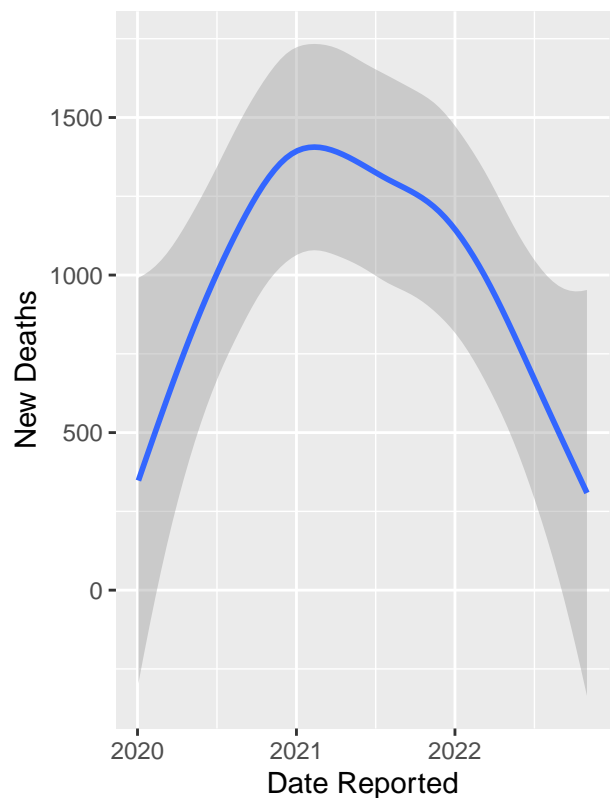
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

Covid-19 Deaths in the USA



Covid-19 Deaths in the USA



```

grid.arrange( usa_cases1, usa_cases2, nrow = 1 )

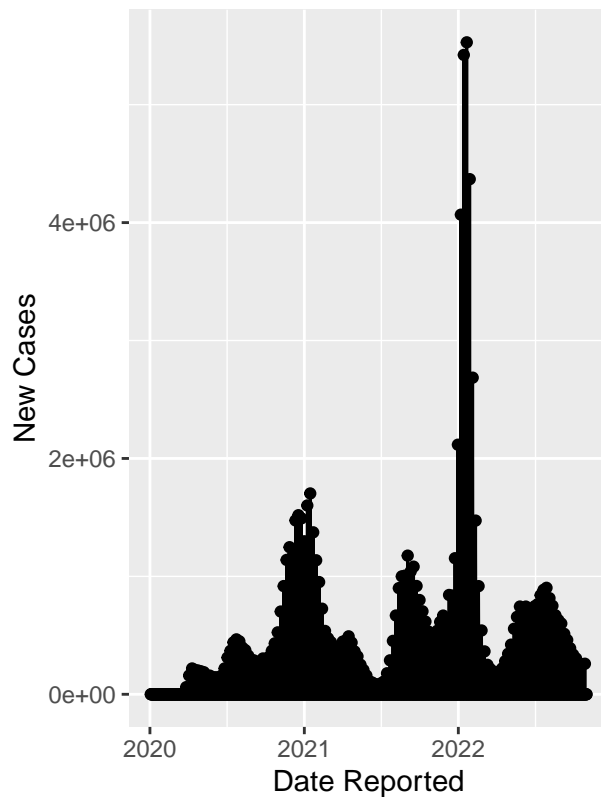
```

```

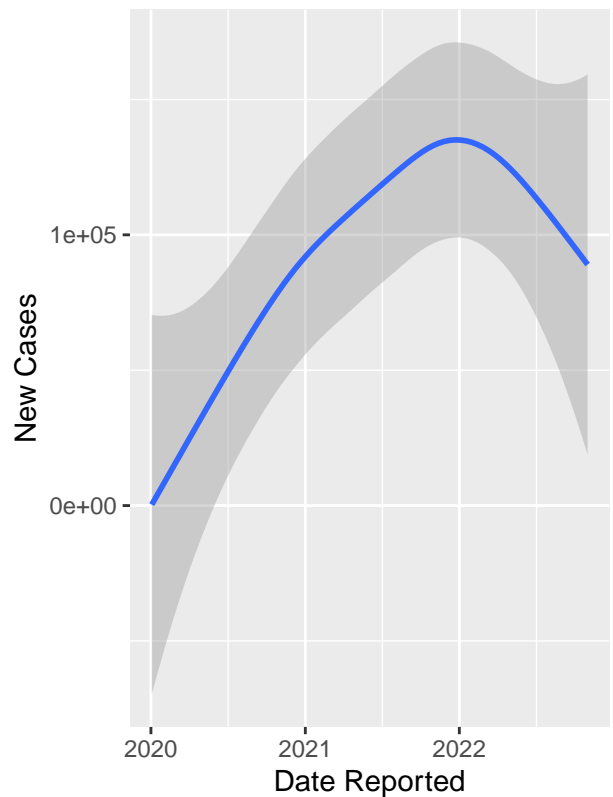
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

Covid-19 Cases in the USA



Covid-19 Cases in the USA



Between the two years is there any pattern between the number of cases and the number of deaths in India

```
india_covid <- world_covid[ which( world_covid$Country == "India" ), ]

india_deaths1 <- ggplot( data = india_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in India" )

india_deaths2 <- ggplot( data = india_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in India" )

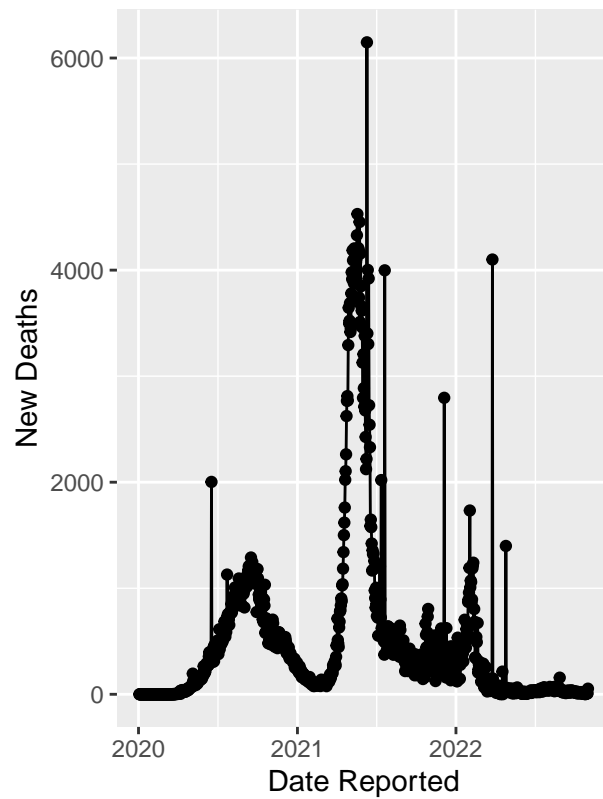
india_cases1 <- ggplot( data = india_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in India" )

india_cases2 <- ggplot( data = india_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in India" )

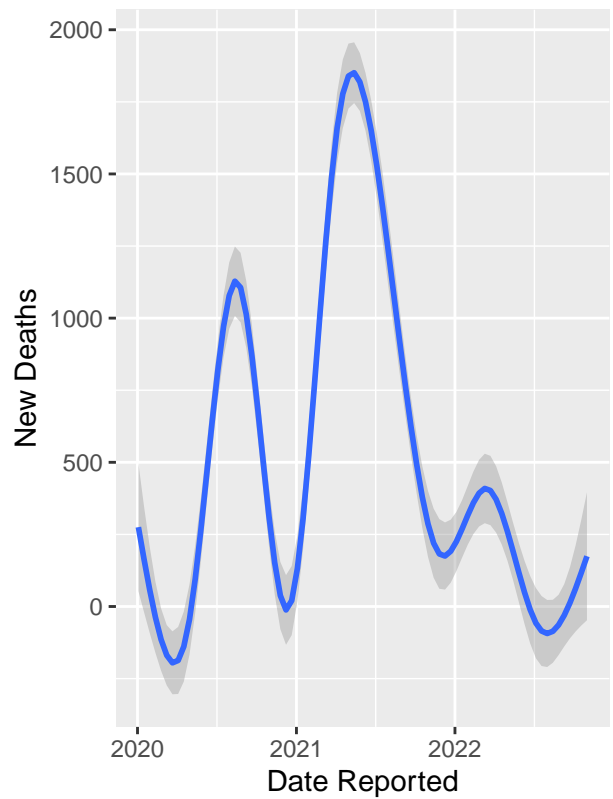
grid.arrange( india_deaths1, india_deaths2, nrow = 1 )

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Covid-19 Deaths in India



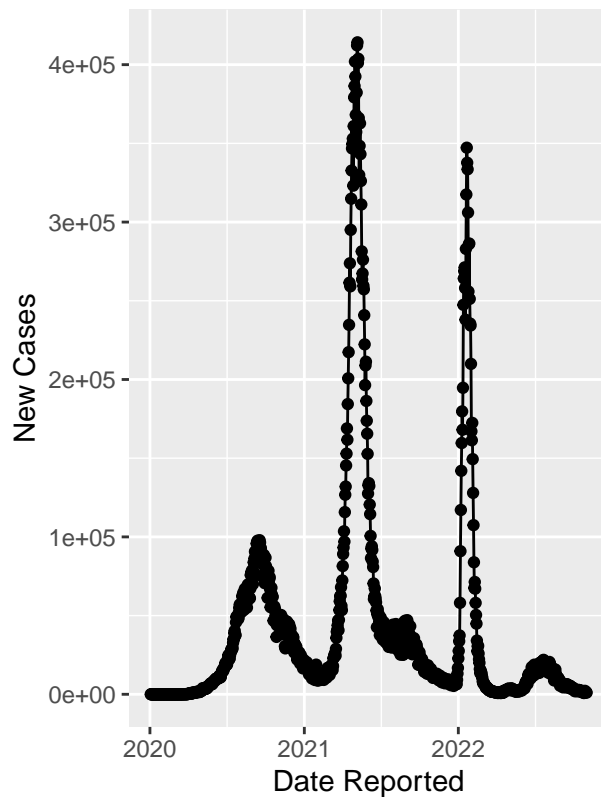
Covid-19 Deaths in India



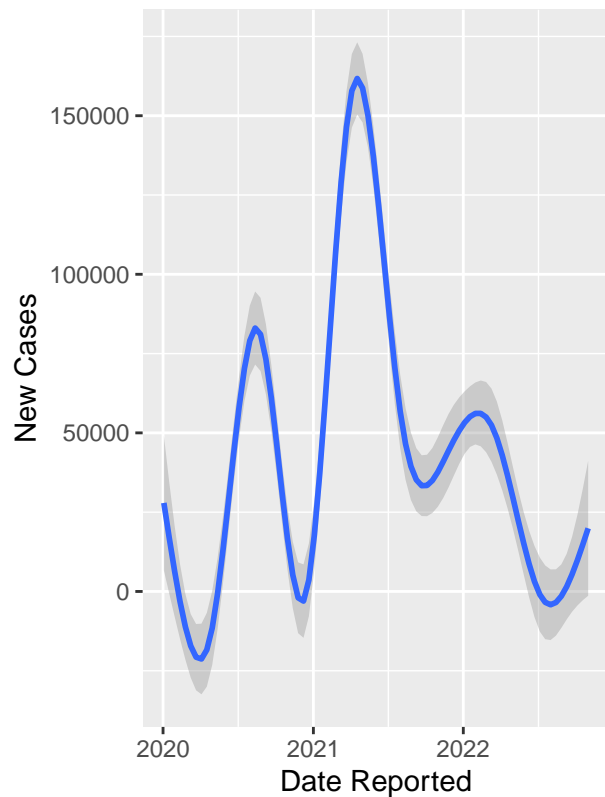
```
grid.arrange( india_cases1, india_cases2, nrow = 1 )
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


Covid-19 Cases in India



Covid-19 Cases in India



Between the two years is there any pattern between the number of cases and the number of deaths in Germany?

```
germany_covid <- world_covid[ which( world_covid$Country == "Germany" ), ]

germany_deaths1 <- ggplot( data = germany_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in Germany" )

germany_deaths2 <- ggplot( data = germany_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in Germany" )

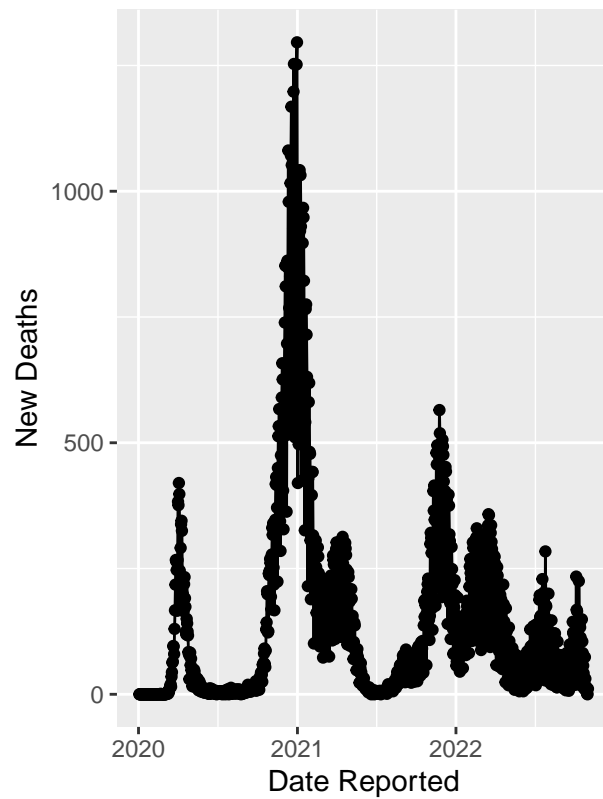
germany_cases1 <- ggplot( data = germany_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in Germany" )

germany_cases2 <- ggplot( data = germany_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in Germany" )

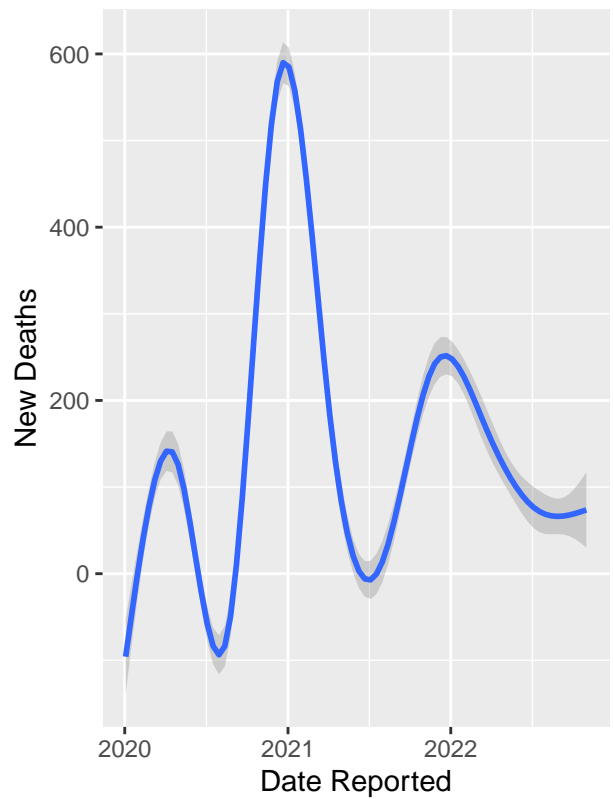
grid.arrange( germany_deaths1, germany_deaths2, nrow = 1 )

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Covid-19 Deaths in Germany



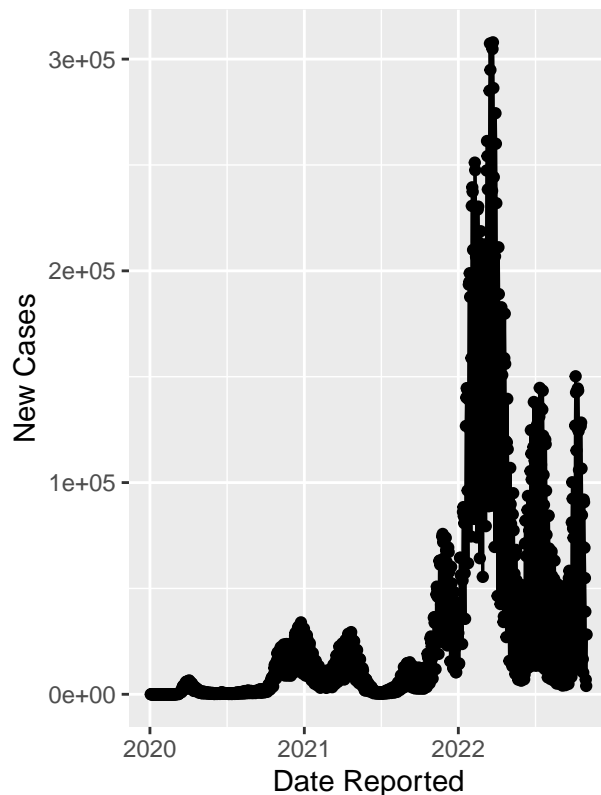
Covid-19 Deaths in Germany



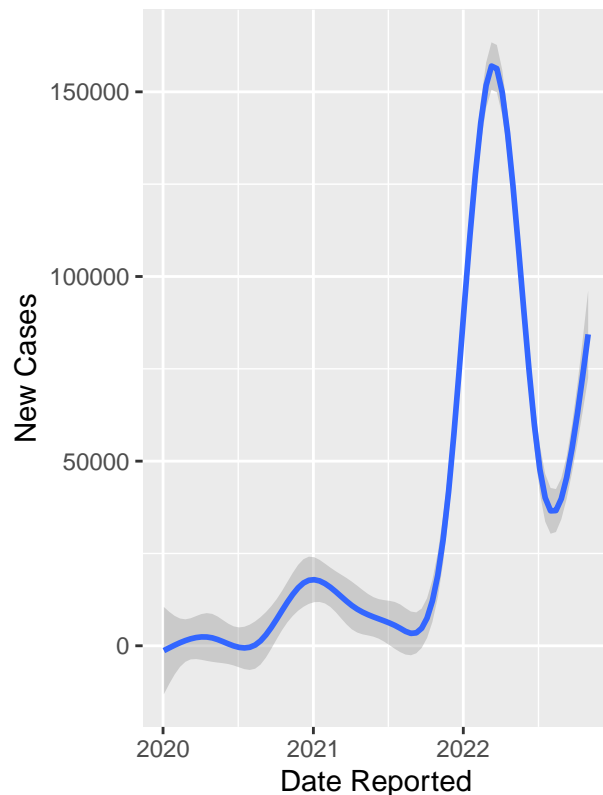
```
grid.arrange( germany_cases1, germany_cases2, nrow = 1 )
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Covid-19 Cases in Germany



Covid-19 Cases in Germany



Between the two years is there any pattern between the number of cases and the number of deaths in France

```
france_covid <- world_covid[ which( world_covid$Country == "France" ), ]
```

```
france_deaths1 <- ggplot( data = france_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in France" )
```

```
france_deaths2 <- ggplot( data = france_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in France" )
```

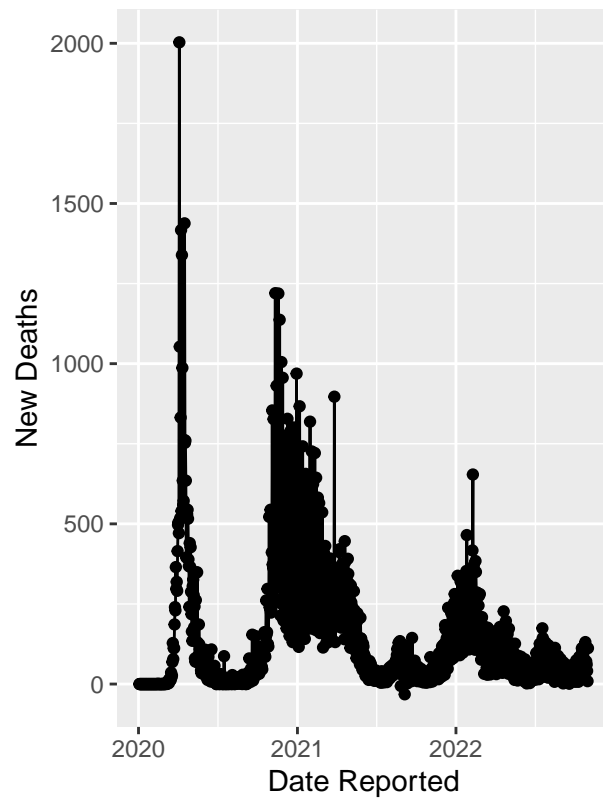
```
france_cases1 <- ggplot( data = france_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in France" )
```

```
france_cases2 <- ggplot( data = france_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in France" )
```

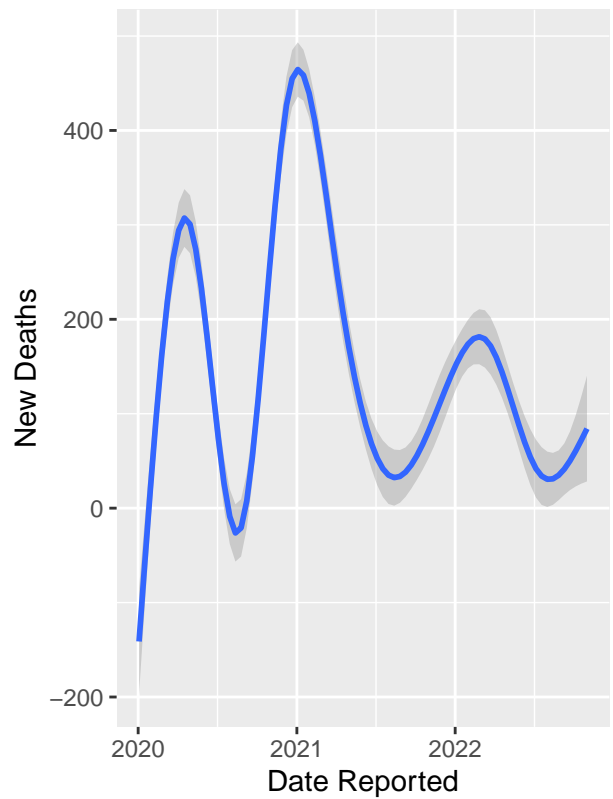
```
grid.arrange( france_deaths1, france_deaths2, nrow = 1 )
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Covid-19 Deaths in France

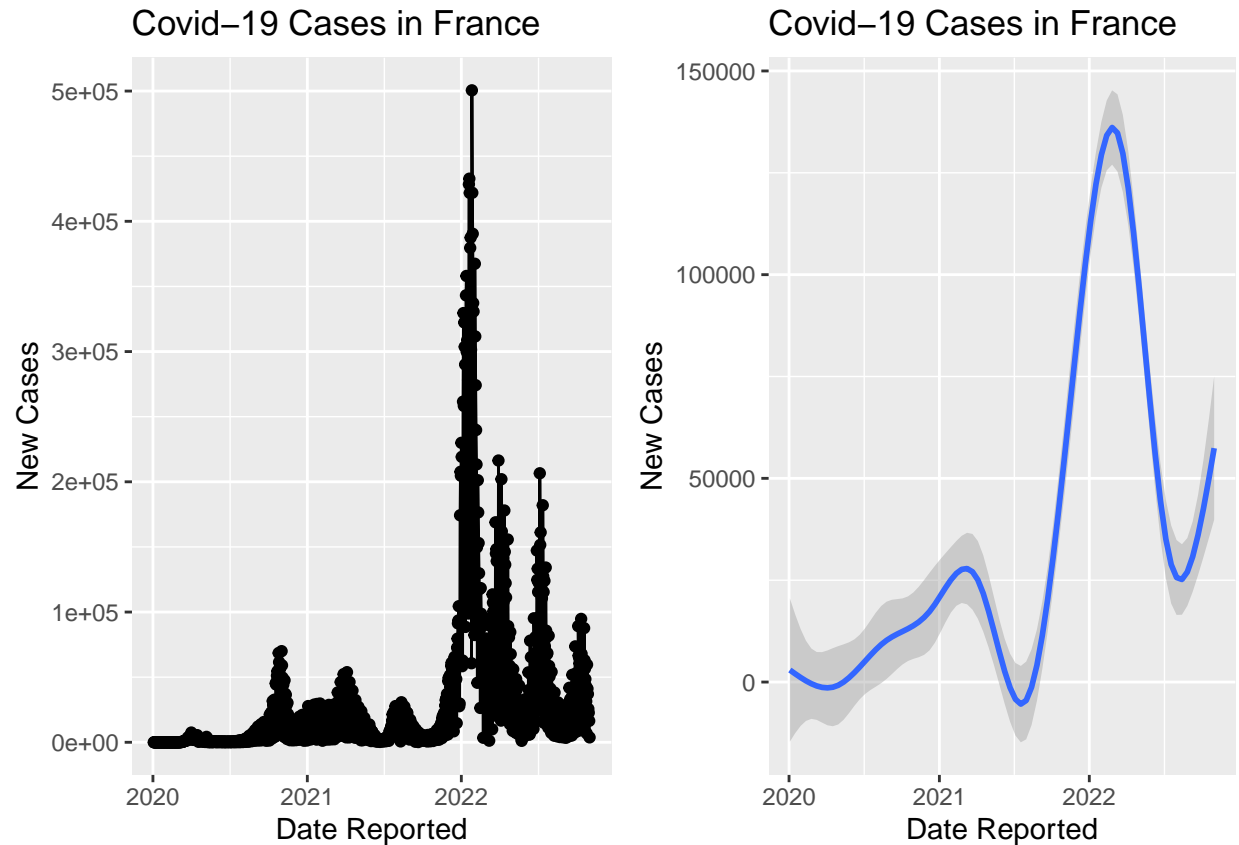


Covid-19 Deaths in France



```
grid.arrange( france_cases1, france_cases2, nrow = 1 )
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Between the two years is there any pattern between the number of cases and the number of deaths in Br

```
brazil_covid <- world_covid[ which( world_covid$Country == "Brazil" ), ]

brazil_deaths1 <- ggplot( data = brazil_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in Brazil" )

brazil_deaths2 <- ggplot( data = brazil_covid, mapping = aes( x = Date_reported, y = New_deaths ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Deaths", title = "Covid-19 Deaths in Brazil" )

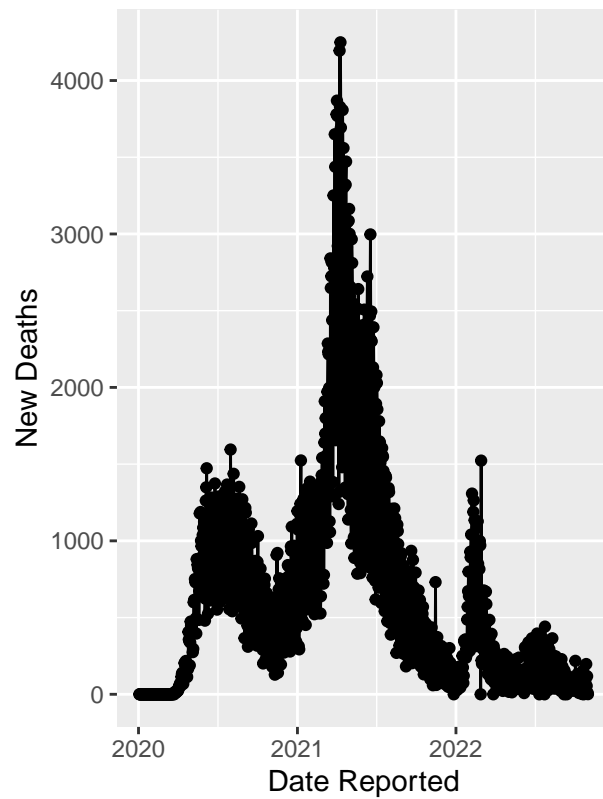
brazil_cases1 <- ggplot( data = brazil_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_line() +
  geom_point() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in Brazil" )

brazil_cases2 <- ggplot( data = brazil_covid, mapping = aes( x = Date_reported, y = New_cases ) ) +
  geom_smooth() +
  labs( x = "Date Reported", y = "New Cases", title = "Covid-19 Cases in Brazil" )

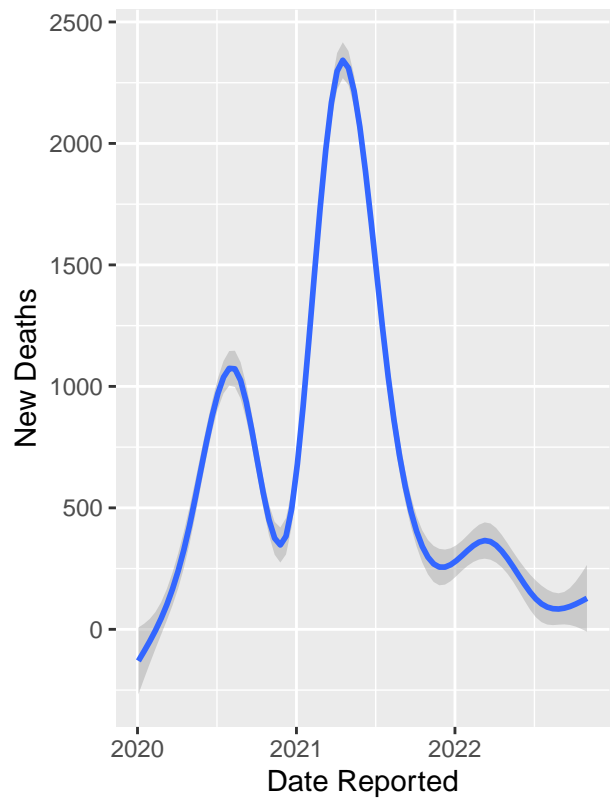
grid.arrange( brazil_deaths1, brazil_deaths2, nrow = 1 )

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Covid-19 Deaths in Brazil

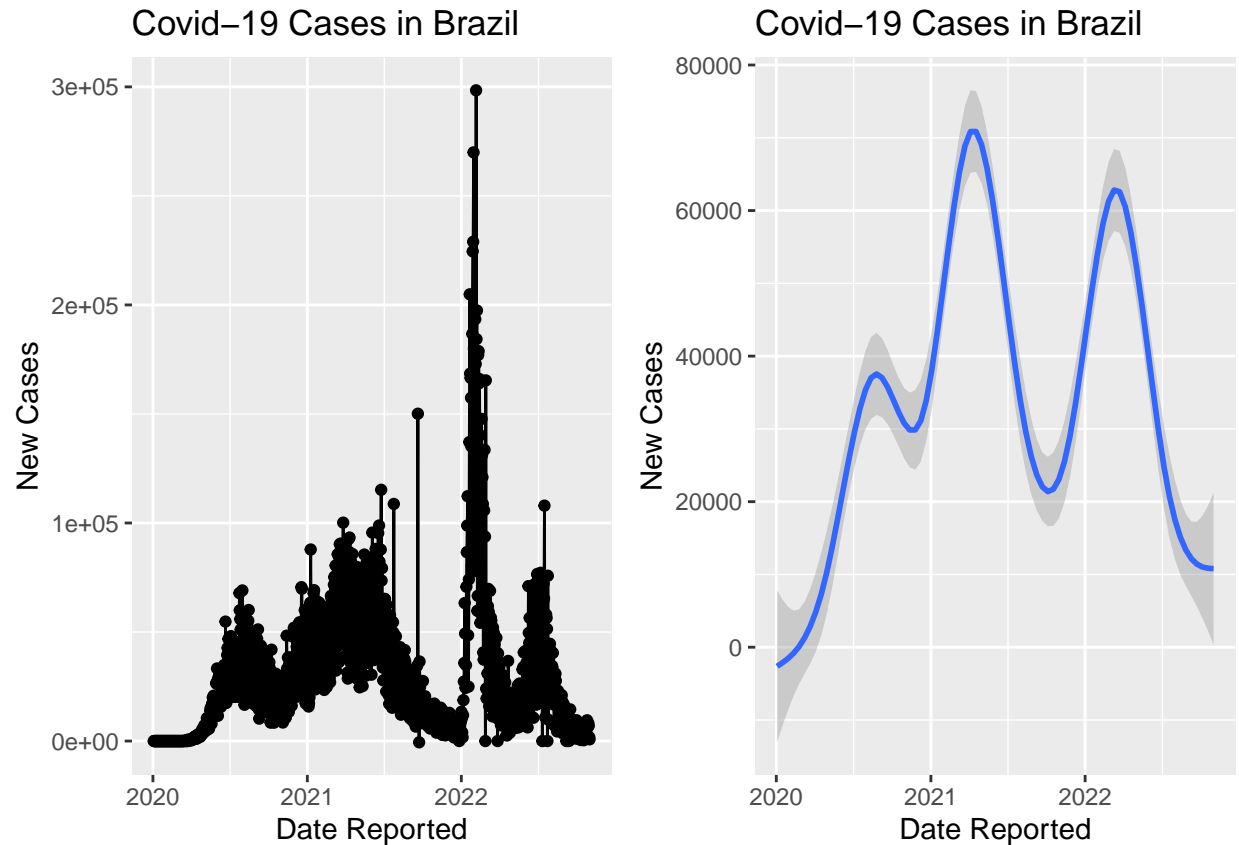


Covid-19 Deaths in Brazil



```
grid.arrange( brazil_cases1, brazil_cases2, nrow = 1 )
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Did air traffic play a role in the increase of covid cases at the start of the pandemic (2020)?

```
library( dplyr )

## Warning: package 'dplyr' was built under R version 4.2.2
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:gridExtra':
##
##   combine
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

usa_air_traffic <- air_traffic[ which( air_traffic$Country == 'United States of America (the)' ), ]
usa_air_traffic <- usa_air_traffic %>% arrange( Date )

usa_air_traffic[, "Cumulative_flights" ] <- cumsum( usa_air_traffic$PercentOfBaseline )

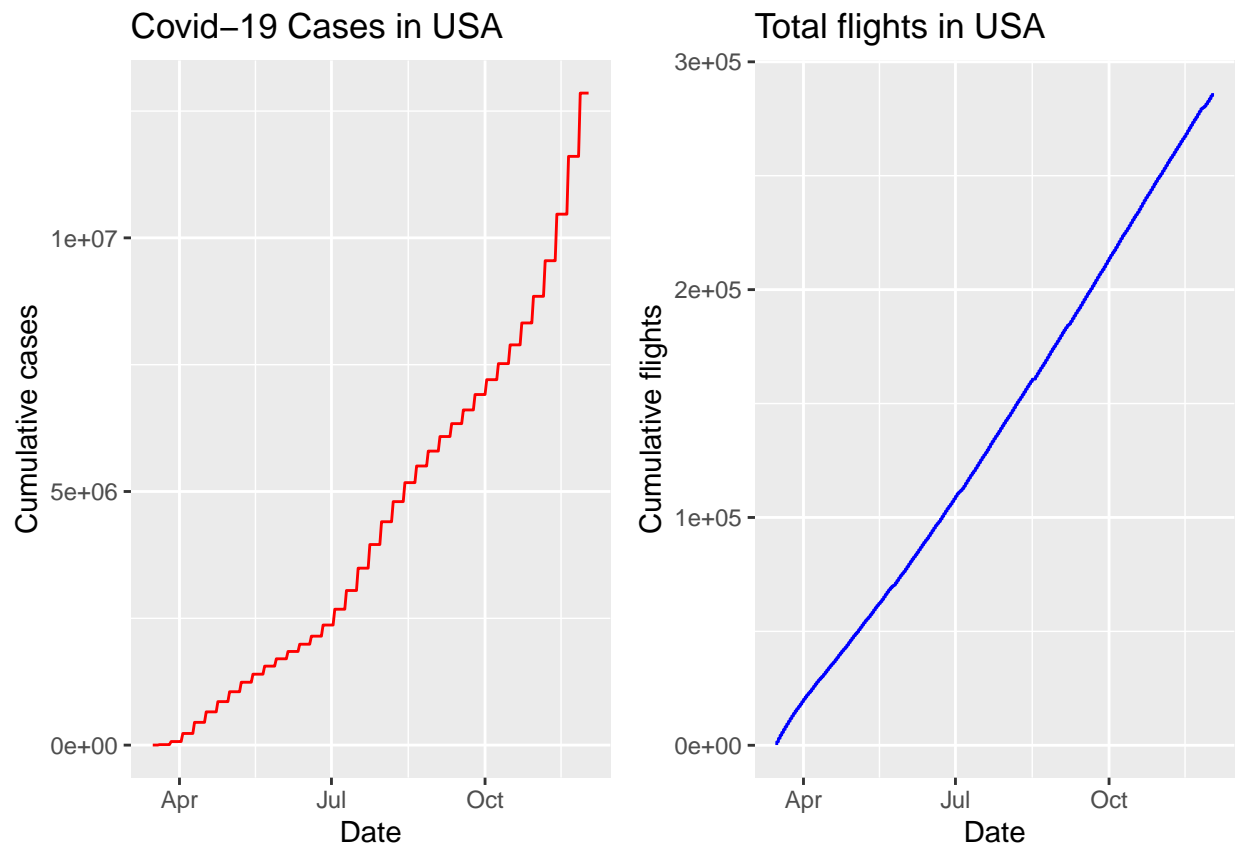
usa_20 <- usa_covid %>% filter(between(Date_reported, as.Date('2020-03-16'), as.Date('2020-12-02')))
```

```
# Changed from bar plots to line plots
```

```
usa_cases <- ggplot( data = usa_20, aes( x = Date_reported, y = Cumulative_cases ) ) +  
  geom_line( stat = "identity", color = "red" ) +  
  labs( x = "Date", y = "Cumulative cases", title = "Covid-19 Cases in USA" )
```

```
usa_traffic <- ggplot( data = usa_air_traffic, aes( x = Date, y = Cumulative_flights ) ) +  
  geom_line( stat = "identity", color = "blue" ) +  
  labs( x = "Date", y = "Cumulative flights", title = "Total flights in USA" )
```

```
grid.arrange( usa_cases, usa_traffic, nrow = 1 )
```



checking if covid spread or cases increased in summer season and festival season due to more travel and family gatherings

```
# Get the library.
```

```
library( plotrix )
```

```
#library(scales)
```

```
# summer holidays in USA mid may to about mid August
```

```
usa_covid$year <- format( usa_covid$Date_reported, "%y" ) # Extract year
```

```
usa_covid_20 <- usa_covid[which(usa_covid$year == '20'),]
```

```
usa_covid_21 <- usa_covid[ which( usa_covid$year == '21' ), ]
```



```

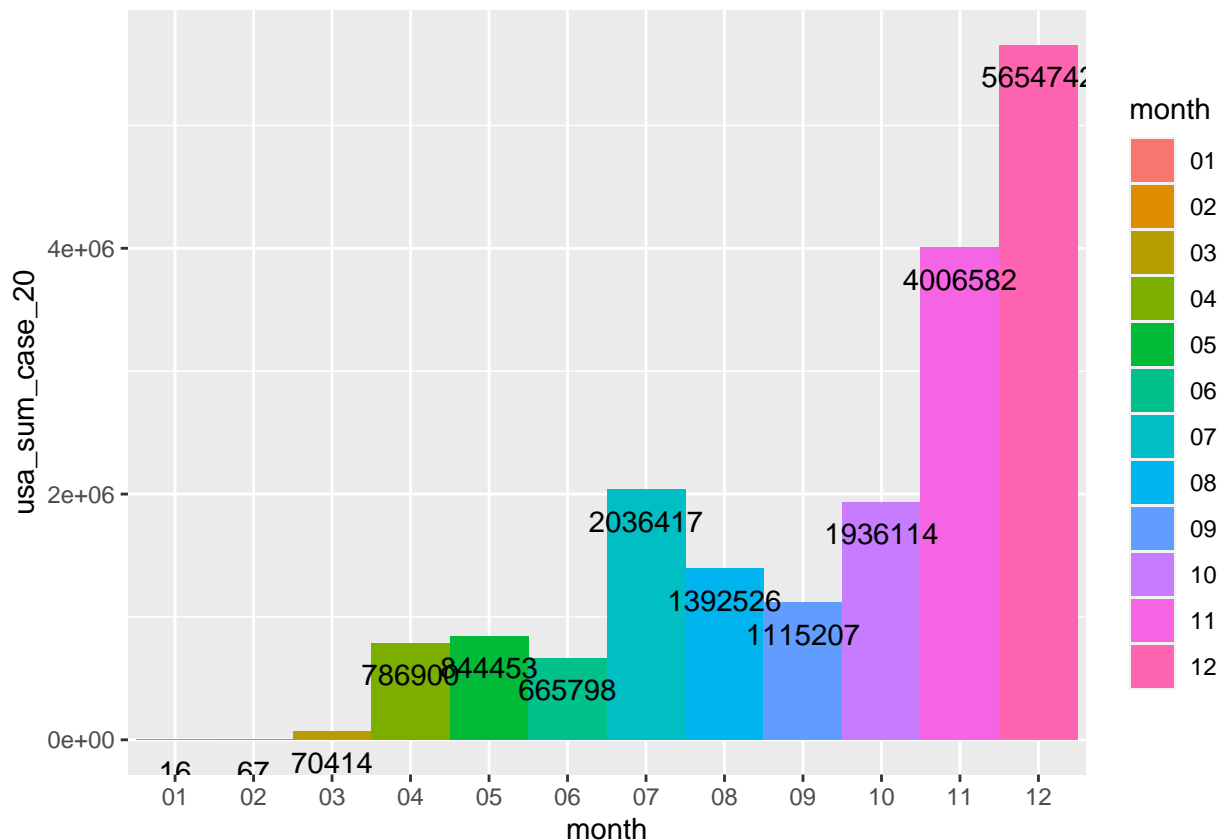
usa_covid_22 <- usa_covid[ which( usa_covid$year == '22' ), ]

usa_covid_20$month <- format( usa_covid_20$Date_reported, "%m" )      # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_sum_20 <- usa_covid_20 %>% group_by( month ) %>% summarise( usa_sum_case_20 = sum( New_cases )

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( usa_covid_sum_20, aes( x = month, y = usa_sum_case_20, fill = month, width = 1 )) + geom_bar( stat = "sum",

```



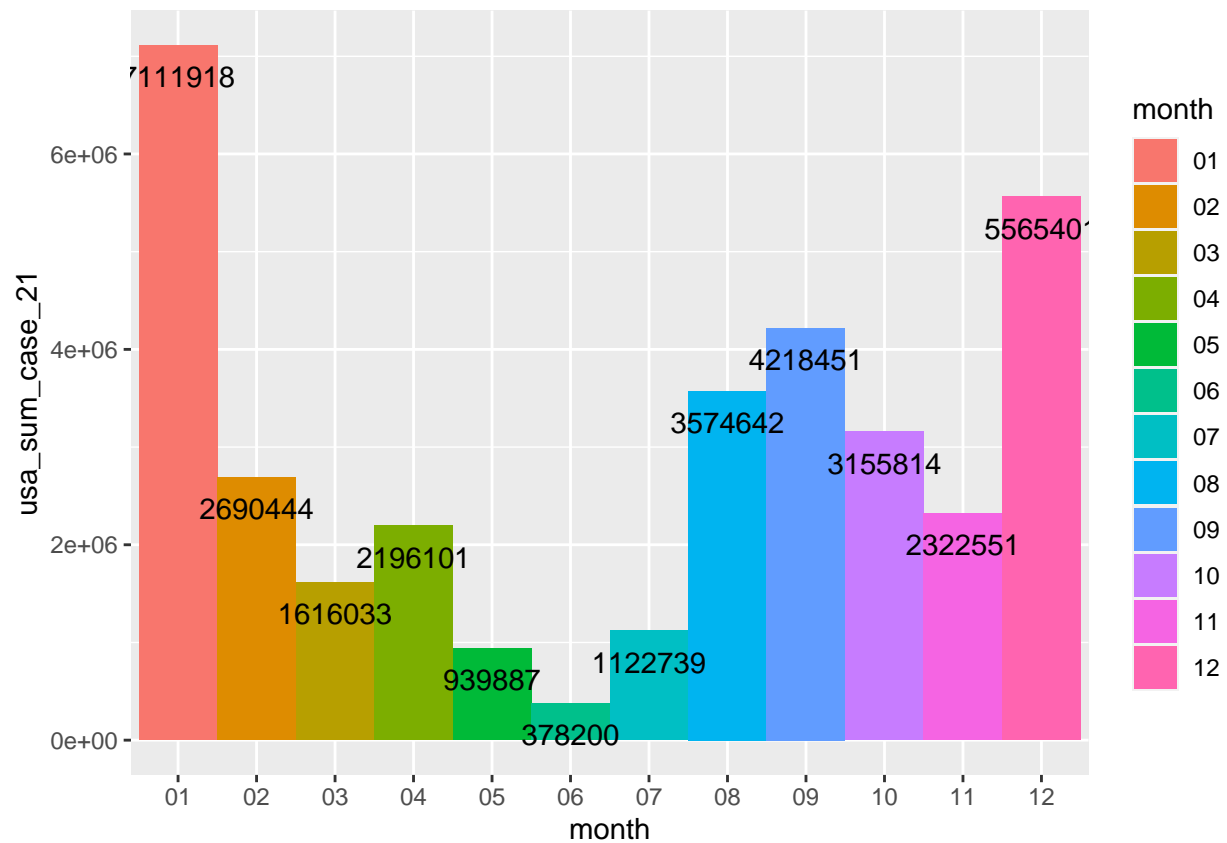
```

usa_covid_21$month <- format( usa_covid_21$Date_reported, "%m" )      # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_sum_21 <- usa_covid_21 %>% group_by( month ) %>% summarise( usa_sum_case_21 = sum( New_cases )

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( usa_covid_sum_21, aes( x = month, y = usa_sum_case_21, fill = month, width = 1 )) + geom_bar( stat = "sum",

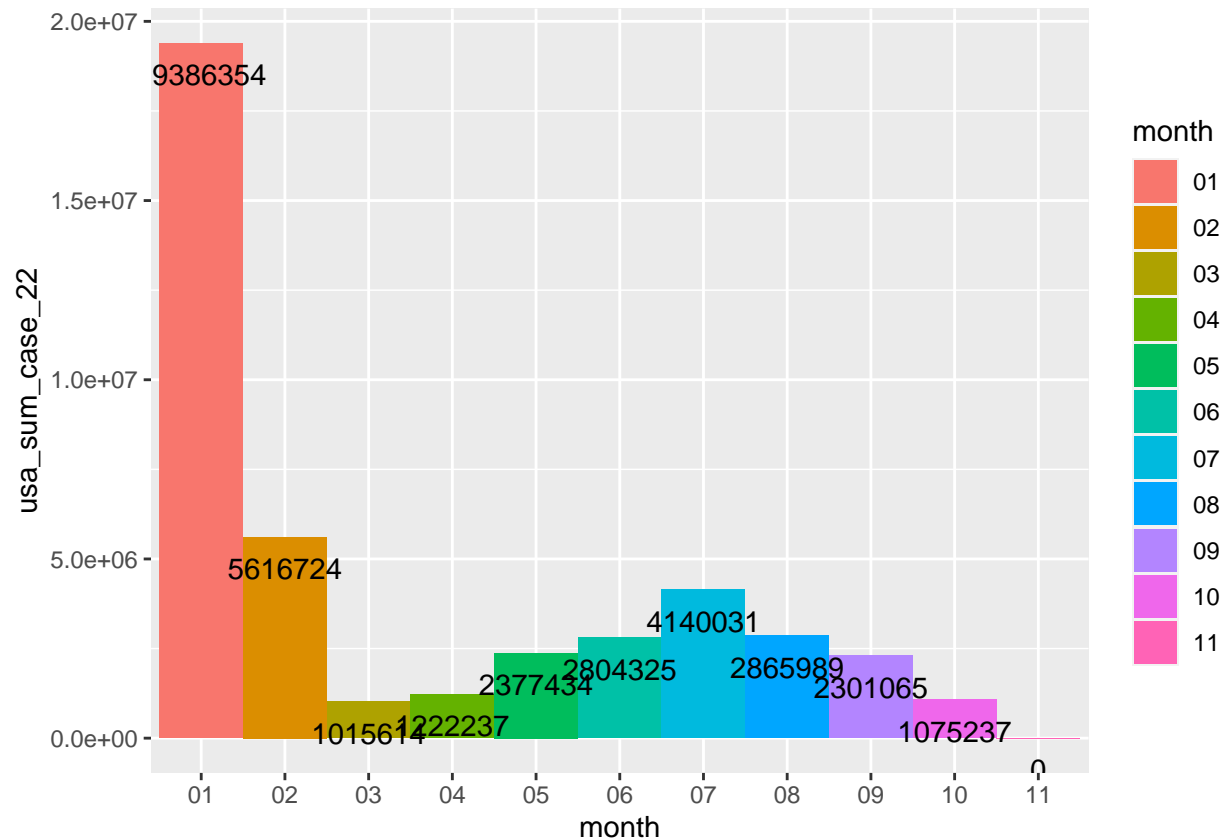
```



```
usa_covid_22$month <- format( usa_covid_22$Date_reported, "%m" ) # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_sum_22 <- usa_covid_22 %>% group_by( month ) %>% summarise( usa_sum_case_22 = sum( New_cases )

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( usa_covid_sum_22, aes( x = month, y = usa_sum_case_22, fill = month, width = 1 )) + geom_bar( s
```



```
# -----

# summer holidays in India Mid may to July 1st

india_covid$year <- format( india_covid$Date_reported, "%y" )           # Extract year

india_covid_20 <- india_covid[which(india_covid$year == '20'),]

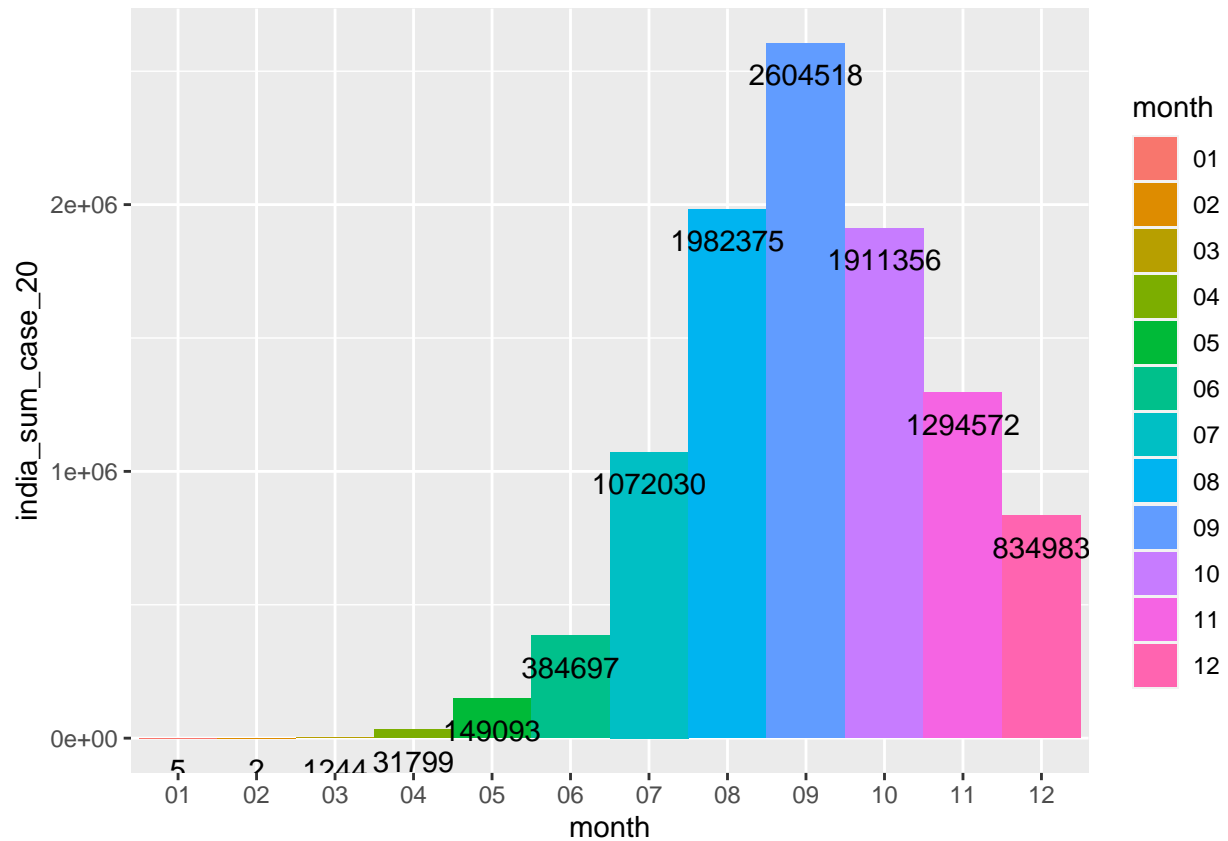
india_covid_21 <- india_covid[ which( india_covid$year == '21' ), ]

india_covid_22 <- india_covid[ which( india_covid$year == '22' ), ]


india_covid_20$month <- format( india_covid_20$Date_reported, "%m" )     # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_sum_20 <- india_covid_20 %>% group_by( month ) %>% summarise( india_sum_case_20 = sum( New_

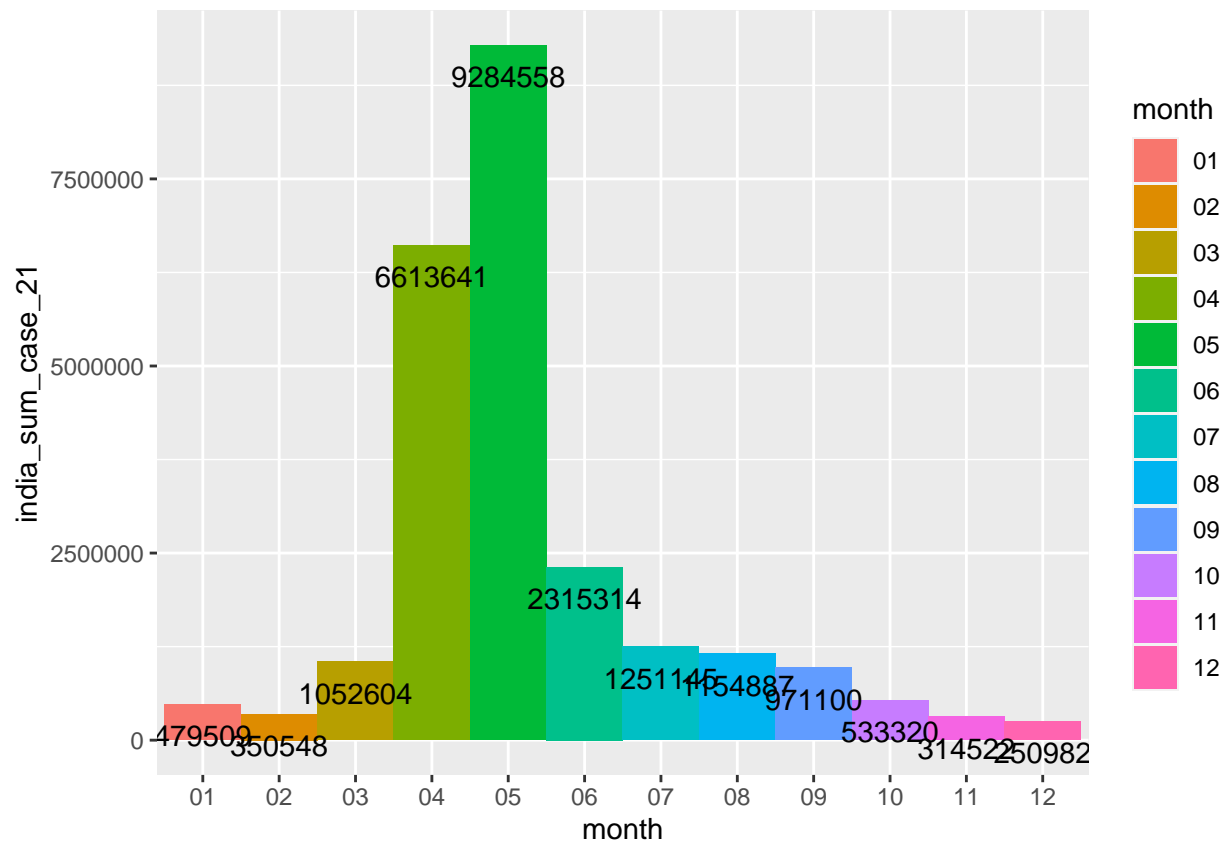
# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( india_covid_sum_20, aes( x = month, y = india_sum_case_20, fill = month, width = 1 )) + geom_bar
```



```
india_covid_21$month <- format( india_covid_21$Date_reported, "%m" ) # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_sum_21 <- india_covid_21 %>% group_by( month ) %>% summarise( india_sum_case_21 = sum( New_

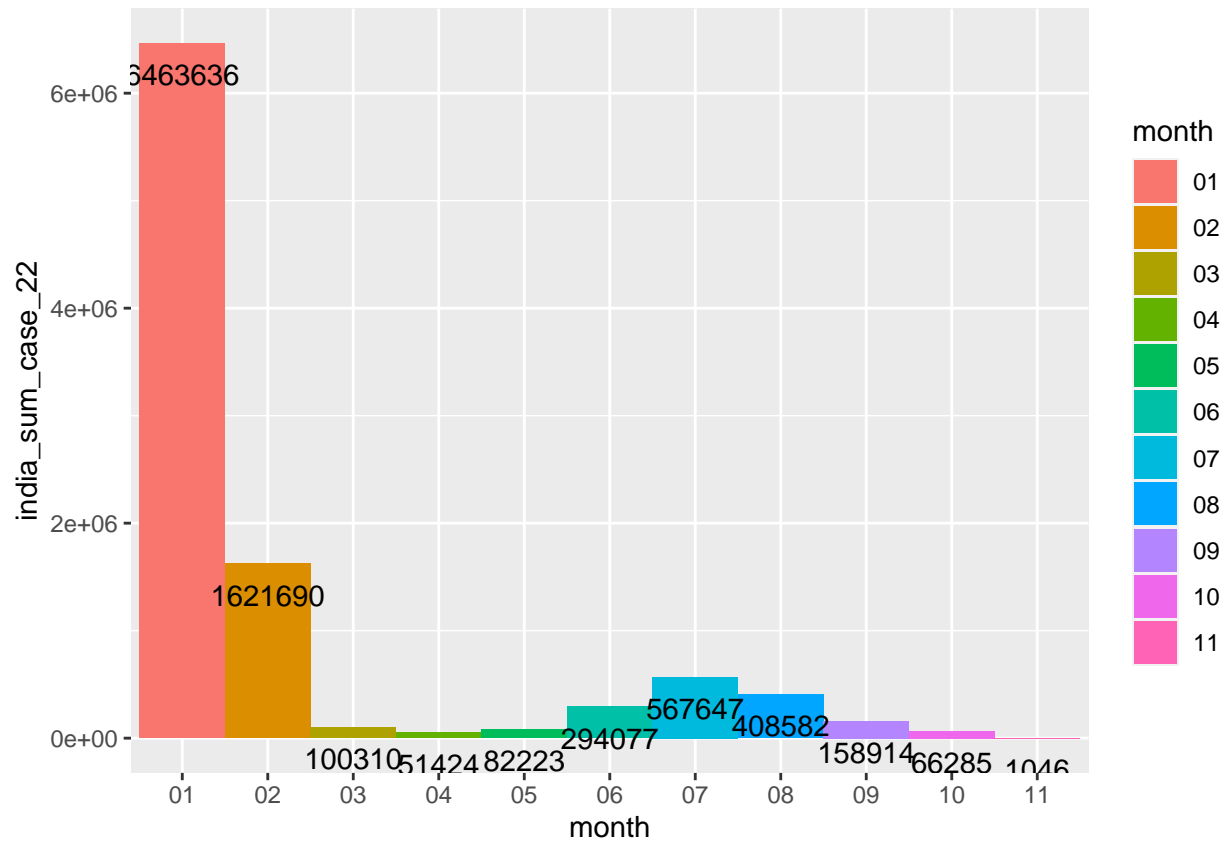
# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( india_covid_sum_21, aes( x = month, y = india_sum_case_21, fill = month, width = 1 )) + geom_bar(
```



```
india_covid_22$month <- format( india_covid_22$Date_reported, "%m" ) # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_sum_22 <- india_covid_22 %>% group_by( month ) %>% summarise( india_sum_case_22 = sum( New_

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( india_covid_sum_22, aes( x = month, y = india_sum_case_22, fill = month, width = 1 )) + geom_bar(
```



```
# -----

germany_covid$year <- format( germany_covid$Date_reported, "%y" )           # Extract year

germany_covid_20 <- germany_covid[ which( germany_covid$year == '20' ), ]

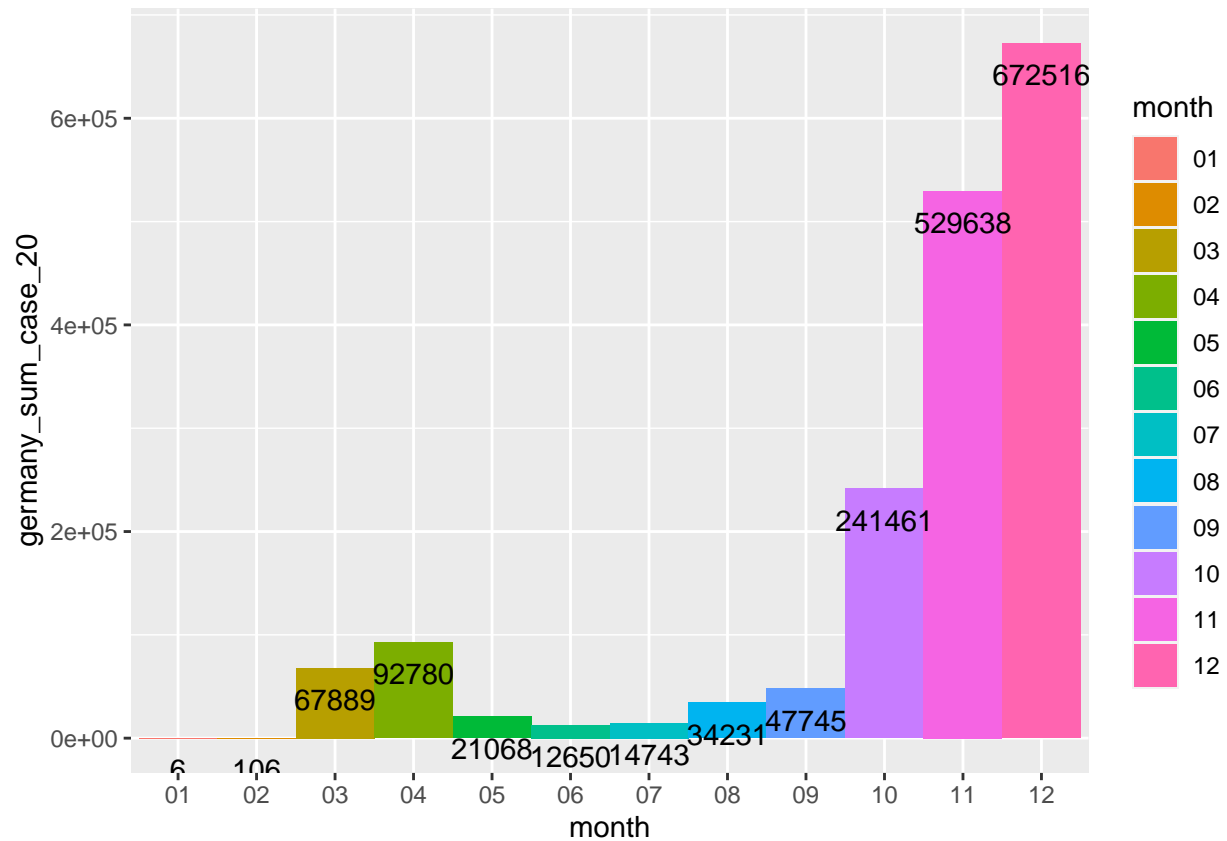
germany_covid_21 <- germany_covid[ which( germany_covid$year == '21' ), ]

germany_covid_22 <- germany_covid[ which( germany_covid$year == '22' ), ]


germany_covid_20$month <- format( germany_covid_20$Date_reported, "%m" )   # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_sum_20 <- germany_covid_20 %>% group_by( month ) %>% summarise( germany_sum_case_20 = sum

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( germany_covid_sum_20, aes( x = month, y = germany_sum_case_20, fill = month, width = 1 )) + geom_bar( )
```



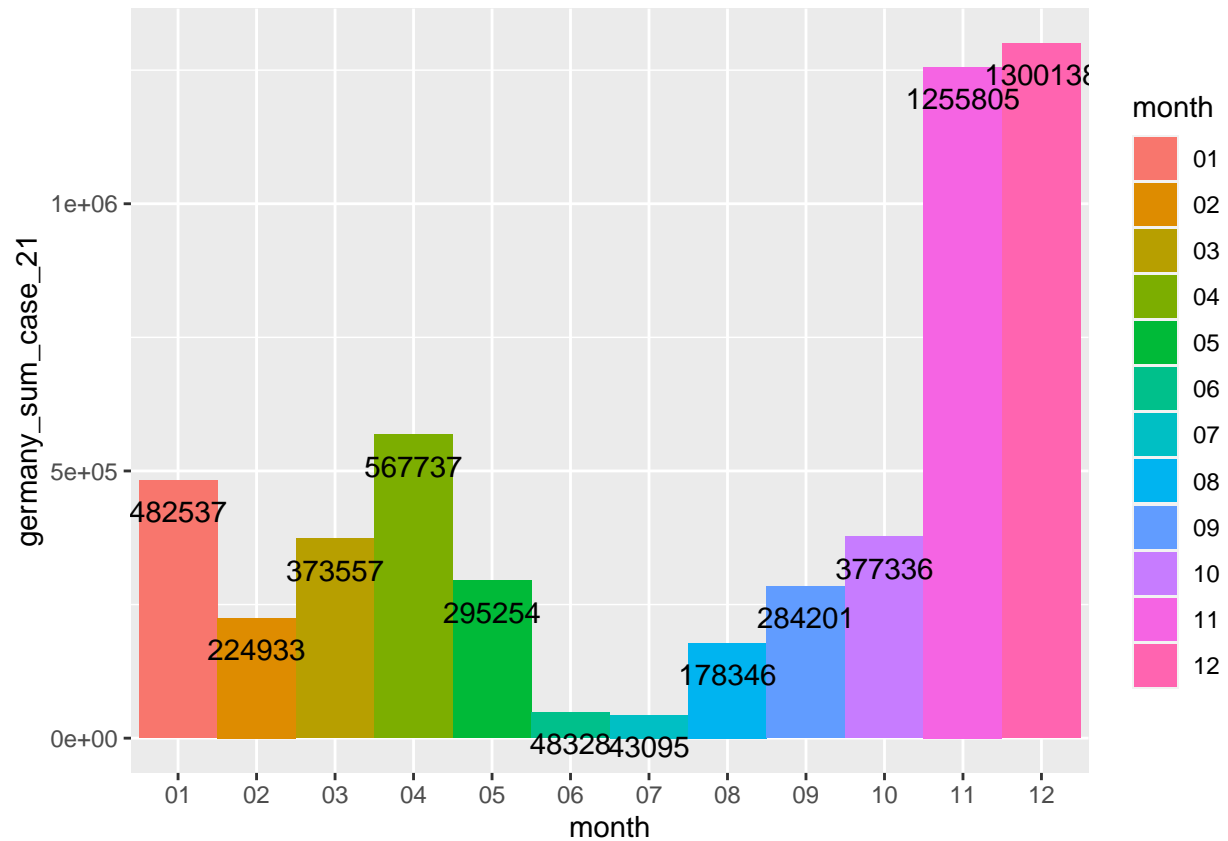
```

germany_covid_21$month <- format( germany_covid_21$Date_reported, "%m" ) # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_sum_21 <- germany_covid_21 %>% group_by( month ) %>% summarise( germany_sum_case_21 = sum

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( germany_covid_sum_21, aes( x = month, y = germany_sum_case_21, fill = month, width = 1 )) + geom

```



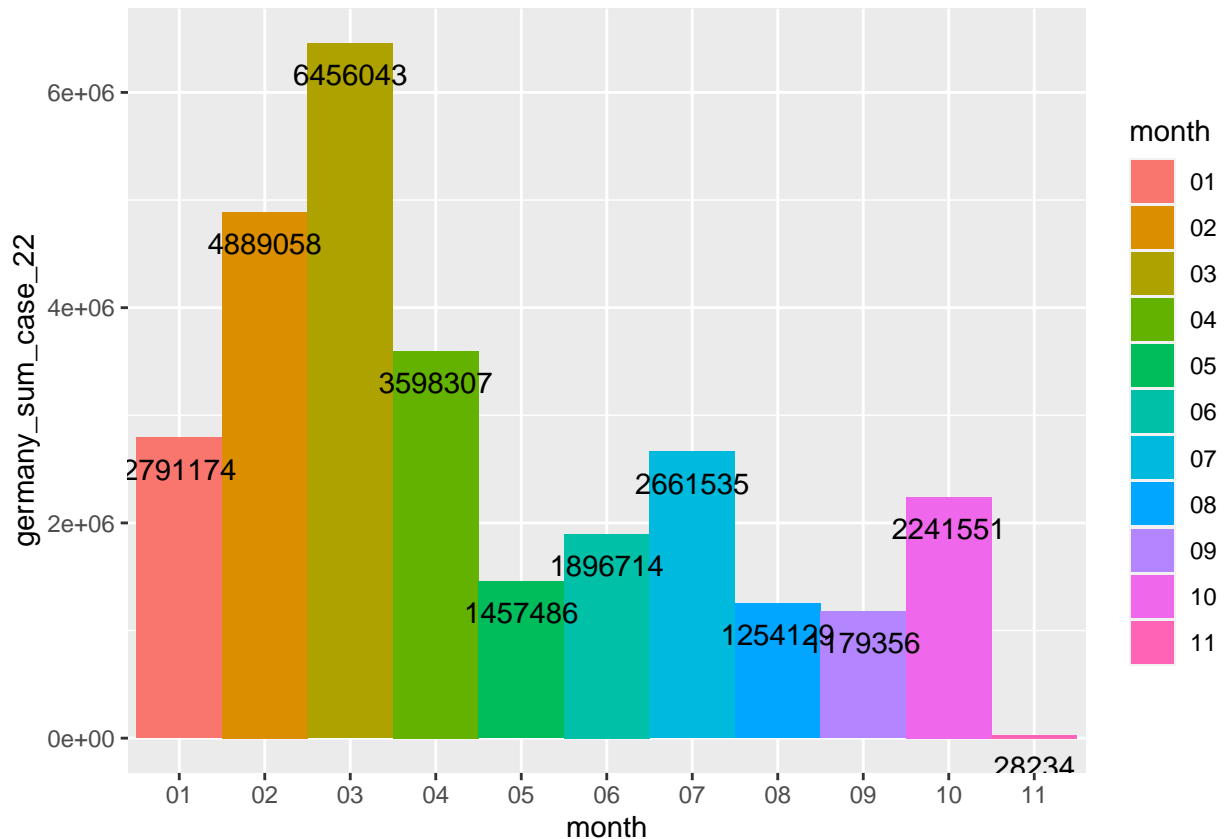
```

germany_covid_22$month <- format( germany_covid_22$Date_reported, "%m" ) # Extract month

# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_sum_22 <- germany_covid_22 %>% group_by( month ) %>% summarise( germany_sum_case_22 = sum

# plotting bar graphs by passing in month as x and cases as y
# setting each bar width to 1 and writing the value in black
ggplot( germany_covid_sum_22, aes( x = month, y = germany_sum_case_22, fill = month, width = 1 )) + geom

```

Did the total number of vaccines affect the total amount of deaths in a country?

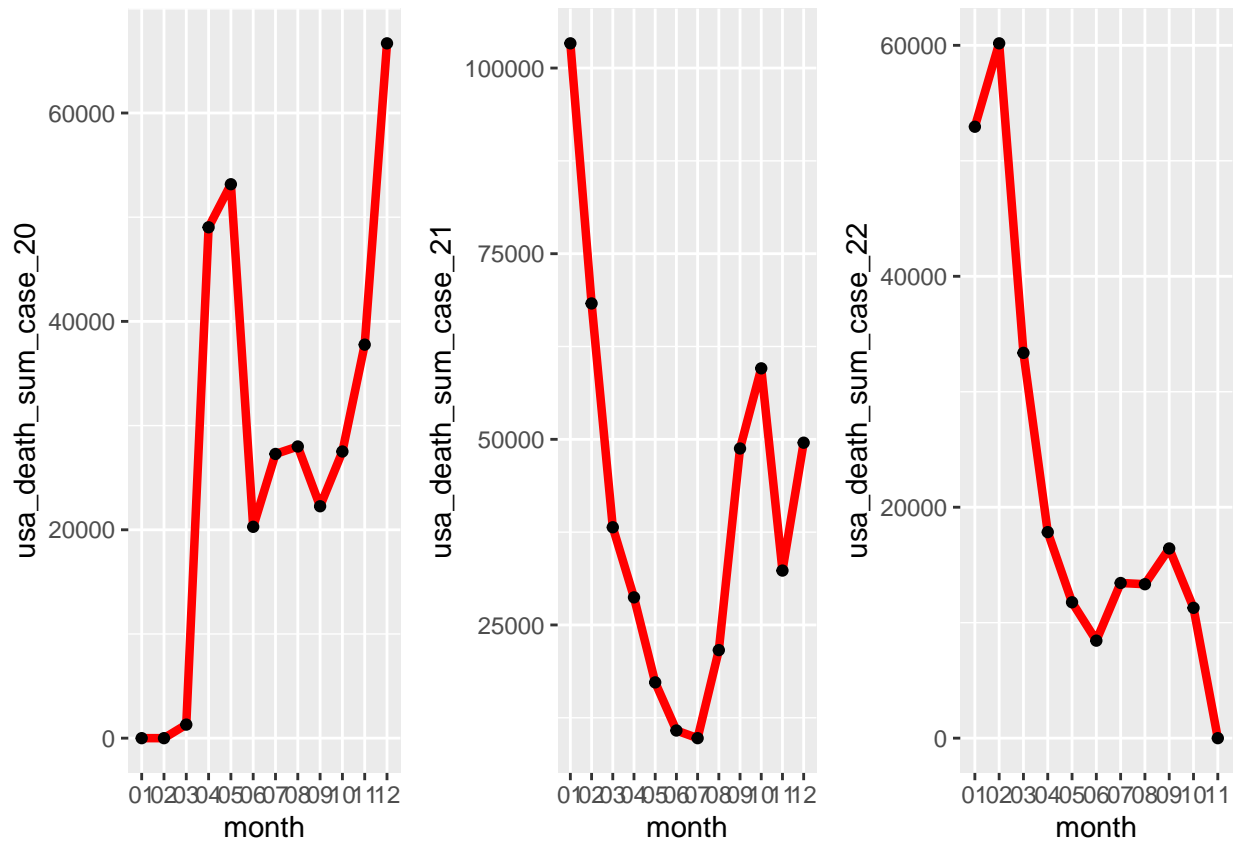
```
# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_death_sum_20 <- usa_covid_20 %>% group_by( month ) %>% summarise( usa_death_sum_case_20 = sum(
  cases ) )

# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_death_sum_21 <- usa_covid_21 %>% group_by( month ) %>% summarise( usa_death_sum_case_21 = sum(
  cases ) )

# grouping by month and adding the covid cases to get the total cases registered in that whole month
usa_covid_death_sum_22 <- usa_covid_22 %>% group_by( month ) %>% summarise( usa_death_sum_case_22 = sum(
  cases ) )

# plotting line graphs by passing in month as x and deaths as y
# setting each bar width to 1 and drawing lines in red for this country
usa_20 <- ggplot( data = usa_covid_death_sum_20, aes( x = month, y = usa_death_sum_case_20, group = 1 ) )
usa_21 <- ggplot( data = usa_covid_death_sum_21, aes( x = month, y = usa_death_sum_case_21, group = 1 ) )
usa_22 <- ggplot( data = usa_covid_death_sum_22, aes( x = month, y = usa_death_sum_case_22, group = 1 ) )

grid.arrange( usa_20, usa_21, usa_22, nrow = 1 )
```



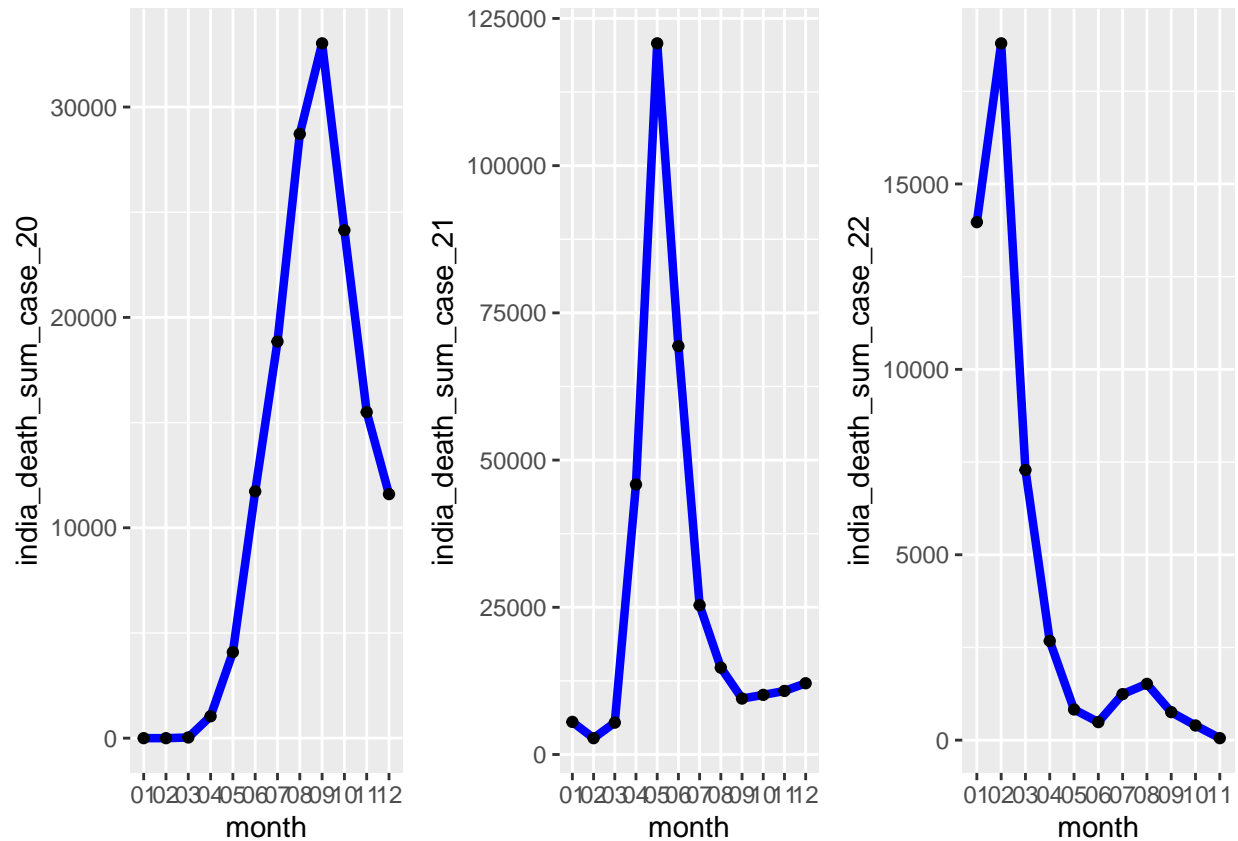
```
# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_death_sum_20 <- india_covid_20 %>% group_by( month ) %>% summarise( india_death_sum_case_20

# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_death_sum_21 <- india_covid_21 %>% group_by( month ) %>% summarise( india_death_sum_case_21

# grouping by month and adding the covid cases to get the total cases registered in that whole month
india_covid_death_sum_22 <- india_covid_22 %>% group_by( month ) %>% summarise( india_death_sum_case_22

# plotting line graphs by passing in month as x and deaths as y
# setting each bar width to 1 and drawing lines in blue for this country
india_20 <- ggplot( data = india_covid_death_sum_20, aes( x = month, y = india_death_sum_case_20, group
india_21 <- ggplot( data = india_covid_death_sum_21, aes( x = month, y = india_death_sum_case_21, group
india_22 <- ggplot( data = india_covid_death_sum_22, aes( x = month, y = india_death_sum_case_22, group

grid.arrange( india_20, india_21, india_22, nrow = 1 )
```



```
# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_death_sum_20 <- germany_covid_20 %>% group_by( month ) %>% summarise( germany_death_sum_case_20 = sum( cases ) )

# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_death_sum_21 <- germany_covid_21 %>% group_by( month ) %>% summarise( germany_death_sum_case_21 = sum( cases ) )

# grouping by month and adding the covid cases to get the total cases registered in that whole month
germany_covid_death_sum_22 <- germany_covid_22 %>% group_by( month ) %>% summarise( germany_death_sum_case_22 = sum( cases ) )

# plotting line graphs by passing in month as x and deaths as y
# setting each bar width to 1 and drawing lines in green for this country
germany_20 <- ggplot( data = germany_covid_death_sum_20, aes( x = month, y = germany_death_sum_case_20, color = "green" ) )
germany_21 <- ggplot( data = germany_covid_death_sum_21, aes( x = month, y = germany_death_sum_case_21, color = "green" ) )
germany_22 <- ggplot( data = germany_covid_death_sum_22, aes( x = month, y = germany_death_sum_case_22, color = "green" ) )

grid.arrange( germany_20, germany_21, germany_22, nrow = 1 )
```

