

## 39\_Project2

Tilak Poudel

2025-03-29

Import covnep\_252days.csv file into R studio as covnep\_252days data frame

```
# load the csv as data frame
covnep_252days <- read.csv("covnep_252days.csv")
```

```
# Display the first few rows of the data frame
head(covnep_252days)
```

```
##      date totalCases newCases totalRecoveries newRecoveries totalDeaths
## 1 1/23/2020         1         1              0              0            0
## 2 1/24/2020         0         0              0              0            0
## 3 1/25/2020         0         0              0              0            0
## 4 1/26/2020         0         0              0              0            0
## 5 1/27/2020         0         0              0              0            0
## 6 1/28/2020         0         0              0              0            0
## newDeaths
## 1         0
## 2         0
## 3         0
## 4         0
## 5         0
## 6         0
```

```
# Check the structure of the data frame
str(covnep_252days)
```

```
## 'data.frame':    252 obs. of  7 variables:
## $ date          : chr  "1/23/2020" "1/24/2020" "1/25/2020" "1/26/2020" ...
## $ totalCases    : int  1 0 0 0 0 0 0 0 0 0 ...
## $ newCases      : int  1 0 0 0 0 0 0 0 0 0 ...
## $ totalRecoveries: int  0 0 0 0 0 0 0 0 1 1 ...
## $ newRecoveries : int  0 0 0 0 0 0 0 0 1 0 ...
## $ totalDeaths   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ newDeaths     : int  0 0 0 0 0 0 0 0 0 0 ...
```

First, we loads the data from the CSV file into a dataframe called “covnep\_252days”. The head() function then displays the first six(default) rows of this dataframe, giving a nature of the dataset’s structure. The str() function provides information about the structure of the dataframe, such as number of variables(7), the data types and the number of observations(252).

Covert the date (character date) variable as date variable (date2) using as.Date function (covnep\_252days data frame)

```
# Convert the date variable to Date type
covnep_252days$date2 <- as.Date(covnep_252days$date, format = "%m/%d/%Y")

head(covnep_252days)
```

```
##      date totalCases newCases totalRecoveries newRecoveries totalDeaths
## 1 1/23/2020         1         1             0             0             0
## 2 1/24/2020         0         0             0             0             0
## 3 1/25/2020         0         0             0             0             0
## 4 1/26/2020         0         0             0             0             0
## 5 1/27/2020         0         0             0             0             0
## 6 1/28/2020         0         0             0             0             0
##   newDeaths      date2
## 1         0 2020-01-23
## 2         0 2020-01-24
## 3         0 2020-01-25
## 4         0 2020-01-26
## 5         0 2020-01-27
## 6         0 2020-01-28
```

```
# check the structure of the data frame again
str(covnep_252days)
```

```
## 'data.frame':   252 obs. of  8 variables:
## $ date          : chr  "1/23/2020" "1/24/2020" "1/25/2020" "1/26/2020" ...
## $ totalCases    : int   1 0 0 0 0 0 0 0 0 0 ...
## $ newCases      : int   1 0 0 0 0 0 0 0 0 0 ...
## $ totalRecoveries: int   0 0 0 0 0 0 0 0 1 1 ...
## $ newRecoveries : int   0 0 0 0 0 0 0 0 1 0 ...
## $ totalDeaths   : int   0 0 0 0 0 0 0 0 0 0 ...
## $ newDeaths     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ date2         : Date, format: "2020-01-23" "2020-01-24" ...
```

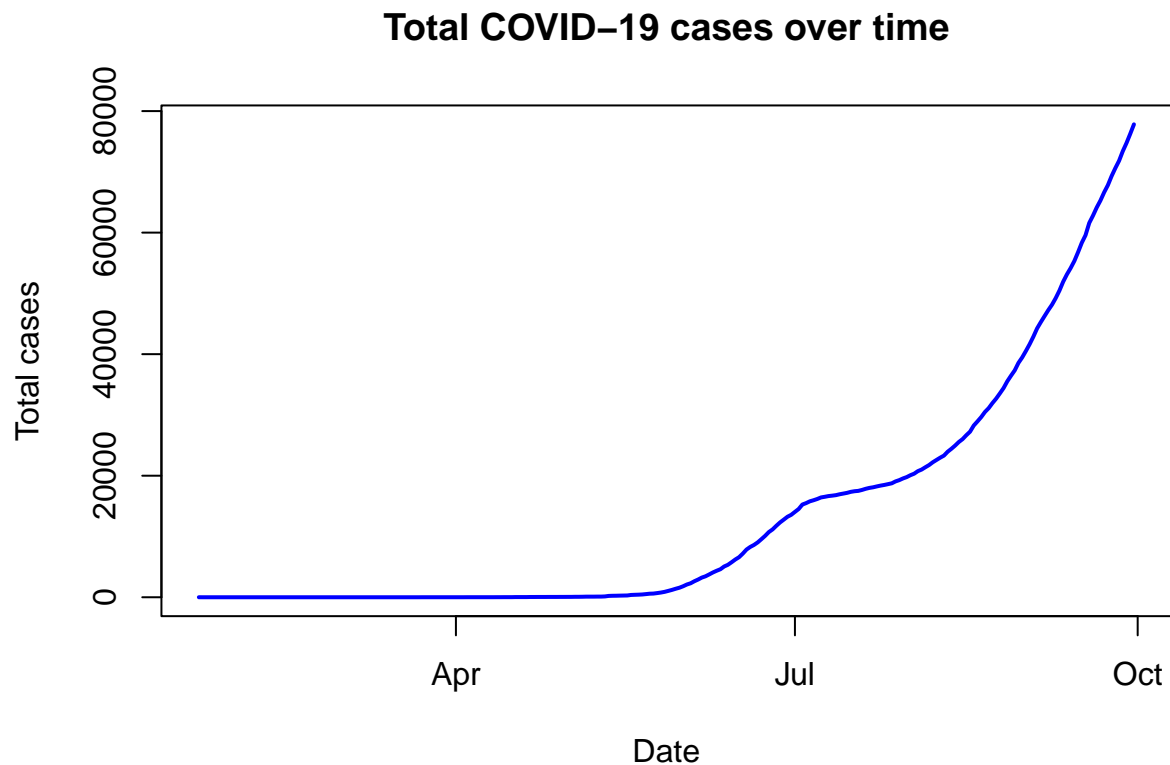
Now, we created a new column date2 of type Date converting from data type char. We can see the type of date2 is Date.

Create line chart of date2 and totalCases variables and interpret it carefully (covnep\_252days data frame)

```
Sys.setlocale("LC_TIME", "en_US.UTF-8") # set locale to english
```

```
## [1] "en_US.UTF-8"
```

```
plot(
  covnep_252days$date2,
  covnep_252days$totalCases,
  type="l",
  col="blue",
  lwd=2,
  xlab="Date",
  ylab="Total cases",
  main="Total COVID-19 cases over time",
)
```



We can observe that the case seems to be almost none up to May, and has increased rapidly from end of August.

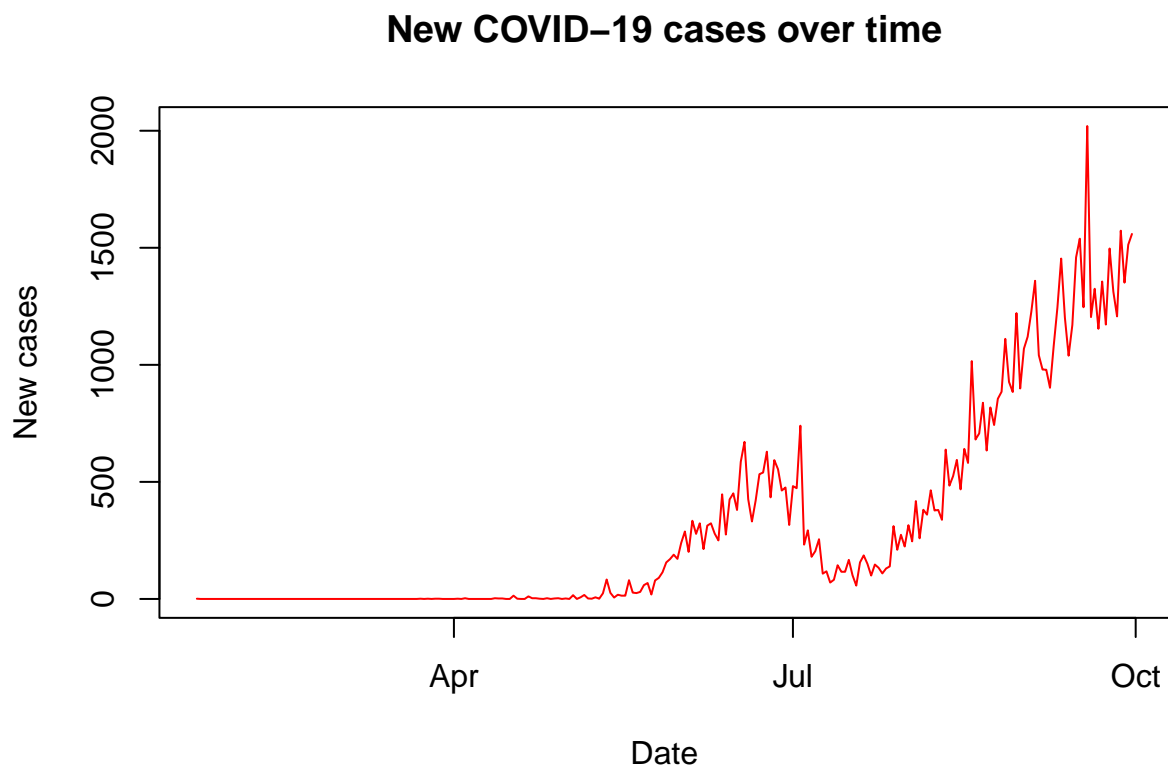
Create line chart of date2 and newCases variables and interpret it carefully (cownep\_252days data frame)

```
plot(
  covnep_252days$date2,
  covnep_252days$newCases,
  type="l",
  col="red",
  lwd=1,
```

```

xlab="Date",
ylab="New cases",
main="New COVID-19 cases over time",
)

```

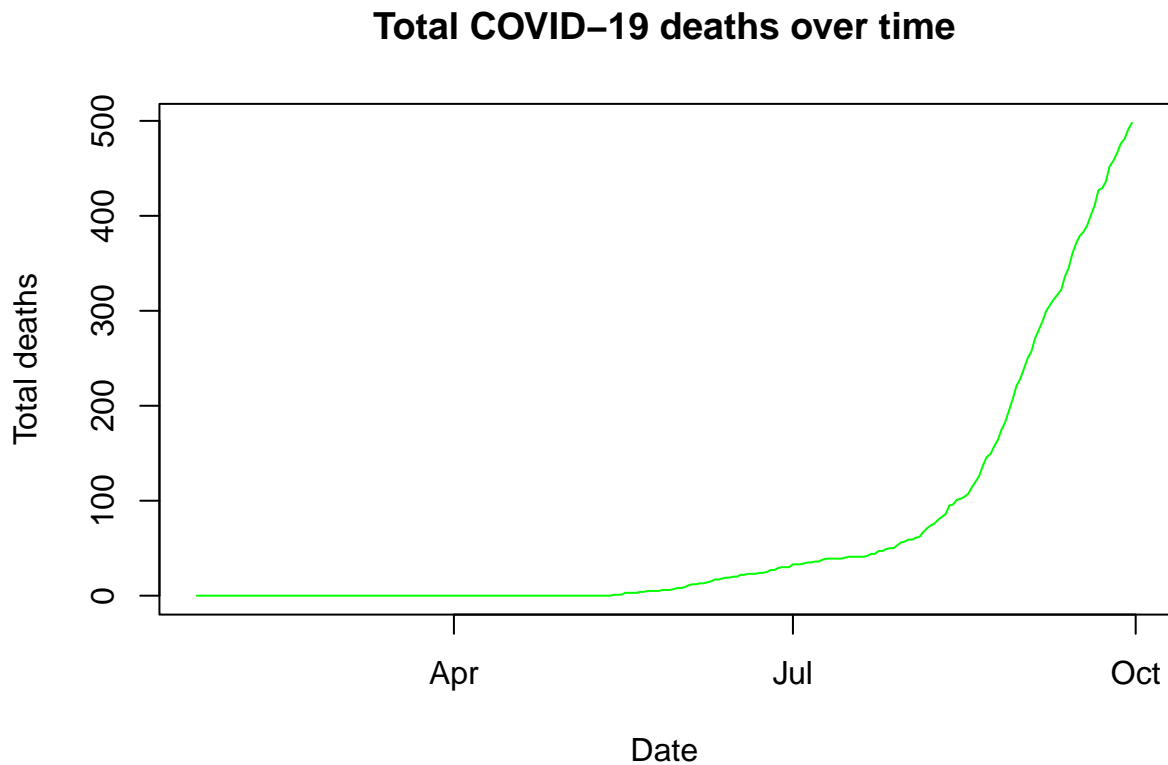


**Interpretation:**

1. Early Period (Before April)
  - Very few or no reported new cases.
  - Possibly due to low transmission, limited testing, or early containment.
2. Gradual Increase (April – July)
  - The number of new cases starts rising.
  - Fluctuations suggest intermittent outbreaks.
3. Rapid Surge (After July – October)
  - A sharp increase in new cases, indicating widespread transmission.
  - Peaks and dips suggest waves of infections.
  - The highest peak exceeds 2000 cases per day, signaling a significant outbreak.

Create line chart of date2 and totalDeaths variables and interpret it carefully (covnep\_252days data frame)

```
plot(  
  covnep_252days$date2,  
  covnep_252days$totalDeaths,  
  type="l",  
  col="green",  
  lwd=1,  
  xlab="Date",  
  ylab="Total deaths",  
  main="Total COVID-19 deaths over time",  
)
```

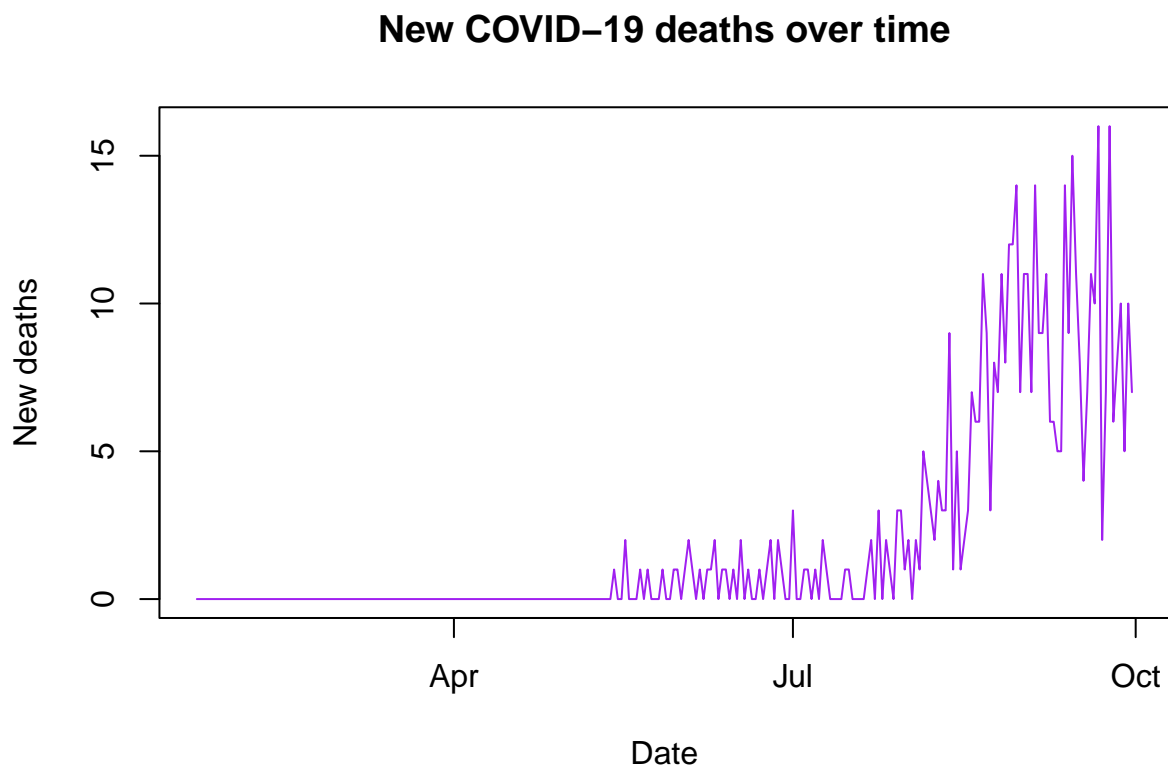


Interpretation:

We can say that the number of deaths are none till may and gradually increase till July. The total deaths seems to rise rapidly after July reaching more than 500 per day.

Create line chart of date2 and newDeaths variable and interpret it carefully (covnep\_252days data frame)

```
plot(
  covnep_252days$date2,
  covnep_252days$newDeaths,
  type="l",
  col="purple",
  lwd=1,
  xlab="Date",
  ylab="New deaths",
  main="New COVID-19 deaths over time",
)
```



**Key Observations from the plot** It can be interpreted in 3 phases:

1. Early Period (Before April)
  - No recorded COVID-19 deaths.
  - Likely due to low infections, effective containment, or delays in reporting fatalities.
2. Gradual Increase (April – July)
  - A small but noticeable rise in deaths.
  - The fluctuations suggest periodic outbreaks, possibly linked to localized surges.

### 3. Significant Surge (After July – October)

- A sharp increase in deaths, correlating with the earlier observed rise in new cases.
- The highest peaks exceed 15 daily deaths, indicating a worsening outbreak.
- Large fluctuations suggest variability in fatality rates, possibly due to hospital capacity, treatment improvements, or reporting delays.

**Compute summary measures of totalCases, newCases, totalRecoveries, newRecoveries, totalDeaths and newDeaths variables using an appropriate apply family of functions (covnep\_252days data frame)**

```
summary_measures <- sapply(covnep_252days[, c("totalCases", "newCases", "totalRecoveries", "newRecoveries", "totalDeaths", "newDeaths")],  
  c(  
    mean = mean(x, na.rm = TRUE),  
    median = median(x, na.rm = TRUE),  
    sd = sd(x, na.rm = TRUE),  
    min = min(x, na.rm = TRUE),  
    max = max(x, na.rm = TRUE)  
  )  
)  
summary_measures
```

##	totalCases	newCases	totalRecoveries	newRecoveries	totalDeaths	newDeaths
## mean	13376.33	308.7976	8380.341	223.3413	66.6746	1.976190
## median	963.00	82.5000	182.000	3.5000	6.0000	0.000000
## sd	19629.60	439.2922	13785.458	424.2460	122.7278	3.625857
## min	0.00	0.0000	0.000	0.0000	0.0000	0.000000
## max	77816.00	2020.0000	56282.000	2287.0000	498.0000	16.000000

Here `sapply()` function from `apply` function is used to compute the summary of the measures of `totalCases`, `newCases`, `totalRecoveries`, `newRecoveries`, `totalDeaths` and `newDeaths`. The null values are removed with `na.rm` and there after the mean, median, standard deviation, min and max is computed.

The data shows significant variation in total cases, recoveries, and deaths across observations, with a highly skewed distribution. While the average total cases are 13,376, the median is much lower (963), indicating that a few high-case regions are inflating the mean. Recoveries follow a similar trend, averaging 8,380 but with a median of 182. Deaths are relatively low, with an average of 67 but a median of 6, and most observations report zero new deaths. The high standard deviations confirm substantial disparities, with some locations experiencing extreme spikes in cases, recoveries, and deaths.