

## Table of Contents

Analyzing India's COVID-19 Data.....	2
Introduction .....	2
Data Analysis Workflow .....	2
Applied Techniques/Concepts .....	2
Implemented R Script.....	3
Screenshots of Outputs .....	5
Conclusion.....	20

## **Analyzing India's COVID-19 Data**

### **Introduction**

The India COVID-19 data set revolves around information about COVID-19 cases in India between 2020 and 2021. A total of 18,110 observations are collected and recorded in this data set, and it contains variables such as serial number, date and time of observation and states or union territories as well as the cumulative number of COVID-19 deaths, cured and confirmed Indian and foreign national cases in India. One assumption is that there are possibly missing values in the raw data set.

### **Data Analysis Workflow**

#### ***Applied Techniques/Concepts***

There are a few steps that need to be done before the data analysis process commences. The first step revolves around data import, cleaning and pre-processing which includes loading the raw data set into the IDE and managing missing values. The second step is exploratory data analysis (EDA) to better understand the data set and the variable distribution through visualization. The third step is data manipulation which includes creating new variables, changing column names and sub-setting columns of interest. The final step is data analysis which includes descriptive and correlational analysis and showing relevant outputs through statistical tables.

## Implemented R Script

```

1 # VICTOR HEW XIN KAI
2 # TP078400
3
4 ## TASK 1
5 # Install relevant packages
6 install.packages("dplyr")
7 install.packages("magrittr")
8 install.packages("tidyverse")
9 install.packages("skimr")
10 install.packages("ggplot2")
11 install.packages("stringr")
12
13 # Load relevant packages
14 library(dplyr)
15 library(magrittr)
16 library(tidyverse)
17 library(skimmr)
18 library(ggplot2)
19 library(stringr)
20
21 # Read the data set from the CSV file and name the data frame as "COVID19_India"
22 COVID19_India <- read.csv(file = "covid_19_india.csv", header = TRUE)
23
24 # View data
25 View(COVID19_India)
26
27 # Change data types of cumulative number of confirmed Indian & foreign nationals and replace '-' values with
28 # missing values labelled as "NA"
29 COVID19_India <- COVID19_India %>%
30   mutate(ConfirmedIndianNational = as.numeric(ConfirmedIndianNational, na.rm = TRUE)) %>%
31   mutate(ConfirmedForeignNational = as.numeric(ConfirmedForeignNational, na.rm = TRUE)) %>%
32   # Sort the updated table by State/Union Territory
33   arrange(State.UnionTerritory)
34
35 # Check if there are any missing values
36 anyNA(COVID19_India)
37 # and remove them if any
38 COVID19_India <- drop_na(COVID19_India)
39
40 # Check data types of each column
41 skim(COVID19_India)
42
43
44 ## TASK 2
45 # Summary for the data set
46 summary(COVID19_India)
47
48 # Total number of records in the data set, labelled as "total_record"
49 total_record <- as.numeric(nrow(COVID19_India)) # 446
50
51 # The day with the highest number of confirmed and cured cases
52 Filtered <- COVID19_India %>%
53   filter(Confirmed == max(Confirmed), Cured == max(Cured)) # Filter to find the highest confirmed and cured cases
54
55 Day_Highest_Cases <- Filtered[, "Date"] # Extract "Date" column from Filtered data frame
56
57 # Visualization - Distribution of Confirmed COVID-19 cases in India using histogram
58 ggplot(COVID19_India, mapping = aes(Confirmed, fill = 'red')) +
59   geom_histogram(binwidth = 2) +
60   labs(title = "Distribution of Confirmed COVID-19 Cases in India")
61
62 # Visualization - Distribution of Confirmed COVID-19 cases on Locals in India using histogram
63 ggplot(COVID19_India, mapping = aes(ConfirmedIndianNational, fill = 'red')) +
64   geom_histogram(binwidth = 2) +
65   labs(title = "Distribution of Cumulative Confirmed COVID-19 Cases on Indian Nationals in India")
66
67 # Visualization - Distribution of Confirmed COVID-19 cases on Foreigners in India using histogram
68 ggplot(COVID19_India, mapping = aes(ConfirmedForeignNational, fill = 'red')) +
69   geom_histogram(binwidth = 2) +
70   labs(title = "Distribution of Cumulative Confirmed COVID-19 Cases on Foreign Nationals in India")
71
72 # Visualization - Distribution of Cured COVID-19 cases in India using histogram
73 ggplot(COVID19_India, mapping = aes(Cured, fill = 'red')) +
74   geom_histogram(binwidth = 2) +
75   labs(title = "Distribution of Cured COVID-19 Cases in India")
76
77 # Visualization - Distribution of Death COVID-19 cases in India using histogram
78 ggplot(COVID19_India, mapping = aes(Deaths, fill = 'red')) +
79   geom_histogram(binwidth = 2) +
80   labs(title = "Distribution of Death COVID-19 Cases in India")
81
82

```

```

82
83 ## TASK 3
84 # Extract "Cured", "Deaths" and "Confirmed" columns from "COVID19_India" data frame and form new one labelled "Avg_Cases"
85 Avg_Cases <- COVID19_India %>%
86   subset(select = c(Cured, Deaths, Confirmed))
87
88 # Add new variables to "COVID19_India" data frame
89 COVID19_India <- COVID19_India %>%
90   mutate(Percentage_Cured = (Cured / Confirmed) * 100) %>% # % of cured cases
91   mutate(Percentage_Death = (Deaths / Confirmed) * 100) %>% # % of death cases
92   mutate(Percentage_Confirmed_Indian_National = (ConfirmedIndianNational / Confirmed) * 100) %>% # % of confirmed cases (Indians)
93   mutate(Percentage_Confirmed_Foreign_National = (ConfirmedForeignNational / Confirmed) * 100) %>% # % of confirmed cases (Foreigners)
94
95 # Separate variable "Date" into 3 levels - Jan, Feb, Mar
96 COVID19_India <- COVID19_India %>%
97   mutate(Date = case_when(
98     str_detect(Date, "/1/") ~ "Jan", # If there's "/1/" in any Date value, then classified as "Jan"
99     str_detect(Date, "/2/") ~ "Feb", # If there's "/2/" in any Date value, then classified as "Feb"
100    str_detect(Date, "/3/") ~ "Mar", # If there's "/3/" in any Date value, then classified as "Mar"
101    TRUE ~ Date # If there's any other value, then remain as Date
102  )) %>%
103   rename(Month = Date) # rename column name Date to Month
104

```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Drafts

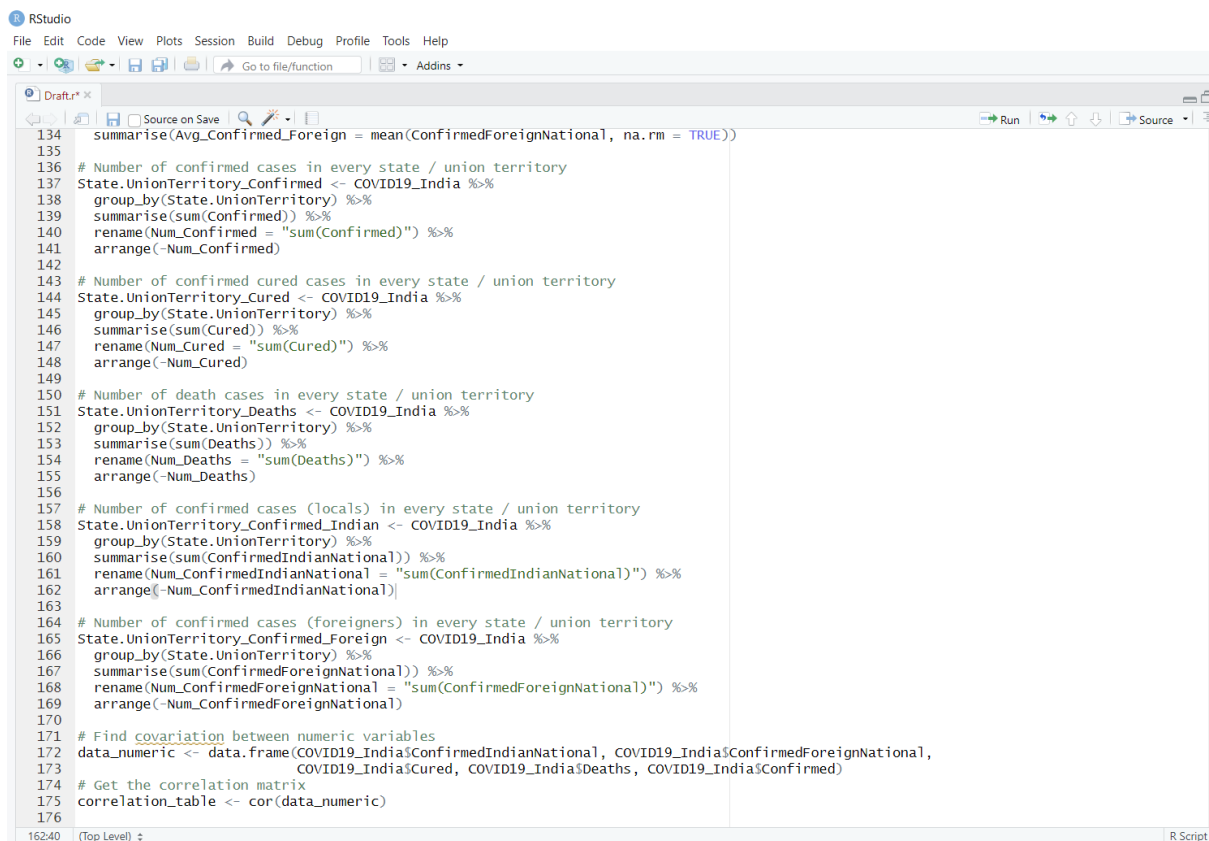
Source on Save Run Source

```

106 ## TASK 4
107 # Average, total and standard deviation number of cured, deaths, and confirmed cases all over the states/union territory
108 Avg_Cases <- Avg_Cases %>%
109   summarise(across(where(is.numeric), list(mean = mean, sum = sum, sd = sd)))
110
111 # Average number of confirmed cases for each month
112 Month_Avg_Confirmed <- COVID19_India %>%
113   group_by(Month) %>%
114   summarise(Avg_Confirmed = mean(Confirmed, na.rm = TRUE))
115
116 # Average number of cured cases for each month
117 Month_Avg_Cured <- COVID19_India %>%
118   group_by(Month) %>%
119   summarise(Avg_Cured = mean(Cured, na.rm = TRUE))
120
121 # Average number of death cases for each month
122 Month_Avg_Death <- COVID19_India %>%
123   group_by(Month) %>%
124   summarise(Avg_Deaths = mean(Deaths, na.rm = TRUE))
125
126 # Average number of confirmed cases on locals for each month
127 Month_Avg_ConfirmedIndian <- COVID19_India %>%
128   group_by(Month) %>%
129   summarise(Avg_Confirmed_Indian = mean(ConfirmedIndianNational, na.rm = TRUE))
130
131 # Average number of confirmed cases on foreigners for each month
132 Month_Avg_ConfirmedForeign <- COVID19_India %>%
133   group_by(Month) %>%
134   summarise(Avg_Confirmed_Foreign = mean(ConfirmedForeignNational, na.rm = TRUE))
135
136 # Number of confirmed cases in every state / union territory
137 State.UnionTerritory_Confirmed <- COVID19_India %>%
138   group_by(State.UnionTerritory) %>%
139   summarise(sum(Confirmed)) %>%
140   rename(Num_Confirmed = "sum(Confirmed)") %>%
141   arrange(-Num_Confirmed)
142
143 # Number of confirmed cured cases in every state / union territory
144 State.UnionTerritory_Cured <- COVID19_India %>%
145   group_by(State.UnionTerritory) %>%
146   summarise(sum(Cured)) %>%
147   rename(Num_Cured = "sum(Cured)") %>%
148   arrange(-Num_Cured)

```

162:40 (Top Level) R Script

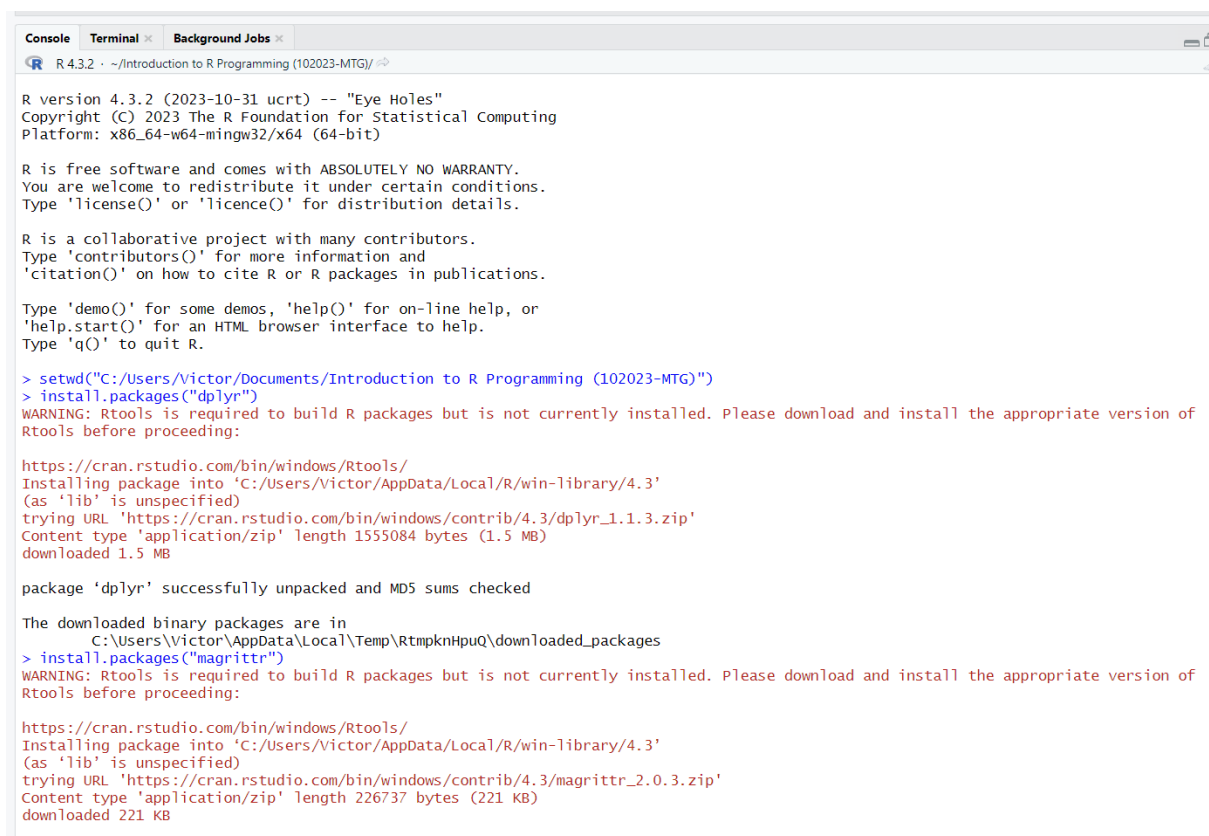


```

134 summarise(Avg_Confirmed_Foreign = mean(ConfirmedForeignNational, na.rm = TRUE))
135
136 # Number of confirmed cases in every state / union territory
137 State.UnionTerritory_Confirmed <- COVID19_India %>%
138   group_by(State.UnionTerritory) %>%
139   summarise(sum(Confirmed)) %>%
140   rename(Num_Confirmed = "sum(Confirmed)") %>%
141   arrange(-Num_Confirmed)
142
143 # Number of confirmed cured cases in every state / union territory
144 State.UnionTerritory_Cured <- COVID19_India %>%
145   group_by(State.UnionTerritory) %>%
146   summarise(sum(Cured)) %>%
147   rename(Num_Cured = "sum(Cured)") %>%
148   arrange(-Num_Cured)
149
150 # Number of death cases in every state / union territory
151 State.UnionTerritory_Deaths <- COVID19_India %>%
152   group_by(State.UnionTerritory) %>%
153   summarise(sum(Deaths)) %>%
154   rename(Num_Deaths = "sum(Deaths)") %>%
155   arrange(-Num_Deaths)
156
157 # Number of confirmed cases (locals) in every state / union territory
158 State.UnionTerritory_Confirmed_Indian <- COVID19_India %>%
159   group_by(State.UnionTerritory) %>%
160   summarise(sum(ConfirmedIndianNational)) %>%
161   rename(Num_ConfirmedIndianNational = "sum(ConfirmedIndianNational)") %>%
162   arrange(-Num_ConfirmedIndianNational)
163
164 # Number of confirmed cases (foreigners) in every state / union territory
165 State.UnionTerritory_Confirmed_Foreign <- COVID19_India %>%
166   group_by(State.UnionTerritory) %>%
167   summarise(sum(ConfirmedForeignNational)) %>%
168   rename(Num_ConfirmedForeignNational = "sum(ConfirmedForeignNational)") %>%
169   arrange(-Num_ConfirmedForeignNational)
170
171 # Find covariation between numeric variables
172 data_numeric <- data.frame(COVID19_India$ConfirmedIndianNational, COVID19_India$ConfirmedForeignNational,
173   COVID19_India$Cured, COVID19_India$Deaths, COVID19_India$Confirmed)
174 # Get the correlation matrix
175 correlation_table <- cor(data_numeric)
176
162:40 (Top Level)
R Script

```

## Screenshots of Outputs



```

R version 4.3.2 (2023-10-31 ucrt) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd("C:/Users/Victor/Documents/Introduction to R Programming (102023-MTG)")
> install.packages("dplyr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/dplyr_1.1.3.zip'
Content type 'application/zip' length 1553084 bytes (1.5 MB)
downloaded 1.5 MB

package 'dplyr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\Victor\AppData\Local\Temp\RtmpkNHpuQ\downloaded_packages
> install.packages("magrittr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/magrittr_2.0.3.zip'
Content type 'application/zip' length 226737 bytes (221 KB)
downloaded 221 KB

```

```

R 4.3.2 · ~/Introduction to R Programming (102023-MTG)/
package 'magrittr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/tidyverse_2.0.0.zip'
Content type 'application/zip' length 430790 bytes (420 KB)
downloaded 420 KB

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
> install.packages("skimmer")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/skimmer_2.1.5.zip'
Content type 'application/zip' length 1237094 bytes (1.2 MB)
downloaded 1.2 MB

package 'skimmer' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/ggplot2_3.4.4.zip'
Content type 'application/zip' length 4299437 bytes (4.1 MB)
downloaded 4.1 MB

```

```

R 4.3.2 · ~/Introduction to R Programming (102023-MTG)/
The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/ggplot2_3.4.4.zip'
Content type 'application/zip' length 4299437 bytes (4.1 MB)
downloaded 4.1 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
> install.packages("stringr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/stringr_1.5.0.zip'
Content type 'application/zip' length 318458 bytes (310 KB)
downloaded 310 KB

package 'stringr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Victor\AppData\Local\Temp\RtmpknHpuQ\downloaded_packages
>
> # Load relevant packages
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':

```

```

R 4.3.2 · ~/Introduction to R Programming (102023-MTG)/
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Victor/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/stringr_1.5.0.zip'
Content type 'application/zip' length 318458 bytes (310 KB)
downloaded 310 KB

package 'stringr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/Victor/AppData/Local/Temp/RtmpknhpuQ/downloaded_packages
>
> # Load relevant packages
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

> library(magrittr)
> library(tidyverse)
— Attaching core tidyverse packages — tidyverse 2.0.0 —
✓ forcats 1.0.0   ✓ readr 2.1.4
✓ ggplot2 3.4.4   ✓ stringr 1.5.0
✓ lubridate 1.9.3 ✓ tibble 3.2.1
✓ purrr 1.0.2    ✓ tidyr 1.3.0
— Conflicts — tidyverse_conflicts() —
✖ tidy::extract() masks magrittr::extract()
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag() masks stats::lag()
✖ purrr::set_names() masks magrittr::set_names()
i Use the conflicted package to force all conflicts to become errors
> library(skimr)
> library(ggplot2)
> library(stringr)
>

```

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R 4.3.2 · ~/Introduction to R Programming (102023-MTG)/
> COVID19_India <- read.csv(file = "covid_19_india.csv", header = TRUE)
> View(COVID19_India)
> View(COVID19_India)
> COVID19_India <- COVID19_India %>%
+   mutate(ConfirmedIndianNational = as.numeric(ConfirmedIndianNational, na.rm = TRUE)) %>%
+   mutate(ConfirmedForeignNational = as.numeric(ConfirmedForeignNational, na.rm = TRUE)) %>%
+   # Sort the updated table by State/Union Territory
+   arrange(State.UnionTerritory)
Warning messages:
1: There was 1 warning in `mutate()`.
i In argument: `ConfirmedIndianNational = as.numeric(ConfirmedIndianNational, na.rm = TRUE)`.
Caused by warning:
! NAs introduced by coercion
2: There was 1 warning in `mutate()`.
i In argument: `ConfirmedForeignNational = as.numeric(ConfirmedForeignNational, na.rm = TRUE)`.
Caused by warning:
! NAs introduced by coercion
> anyNA(COVID19_India)
[1] TRUE
> # and remove them if any
> COVID19_India <- drop_na(COVID19_India)
>

```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal Background Jobs

R 4.3.2 - ~/Introduction to R Programming (102023-MTG)/

```

1: There was 1 warning in `mutate()`.
i In argument: `ConfirmedIndianNational = as.numeric(ConfirmedIndianNational, na.rm = TRUE)`.
Caused by warning:
! NAs introduced by coercion
2: There was 1 warning in `mutate()`.
i In argument: `ConfirmedForeignNational = as.numeric(ConfirmedForeignNational, na.rm = TRUE)`.
Caused by warning:
! NAs introduced by coercion
> anyNA(COVID19_India)
[1] TRUE
> # and remove them if any
> COVID19_India <- drop_na(COVID19_India)
>
> # Check data types of each column
> skim(COVID19_India)
— Data Summary —

```

Name	Values
COVID19_India	
Number of rows	446
Number of columns	9

```

Column type frequency:
character      3
numeric        6

Group variables:      None

— Variable type: character —
skim_variable      n_missing complete_rate min max empty n_unique whitespace
1 Date              0              1 8 9      0      59      0
2 Time              0              1 7 8      0       2      0
3 State.UnionTerritory 0              1 3 27     0      27      0

— Variable type: numeric —
skim_variable      n_missing complete_rate mean sd p0 p25 p50 p75 p100 hist
1 Sno              0              1 224. 129. 1 112. 224. 335. 446
2 ConfirmedIndianNational 0              1 12.2 21.6 0 1 3 13 177
3 ConfirmedForeignNational 0              1 1.50 3.58 0 0 0 1 14
4 Cured            0              1 1.09 2.64 0 0 0 1 25
5 Deaths          0              1 0.244 0.597 0 0 0 0 5
6 Confirmed        0              1 13.7 22.6 1 2 4 15 180

```

> View(COVID19\_India)

>

File Debug Profile Tools Help

to file/function Addins

Project: (None)

Environment History Connections Tutorial

Import Dataset 217 MiB

R Global Environment

Data

Object	Description
Avg_Cases	1 obs. of 9 variables
correlation_table	num [1:5, 1:5] 1 0.194 0.589 0.589 0.988 ...
COVID19_India	446 obs. of 13 variables
data_numeric	446 obs. of 5 variables
Filtered	1 obs. of 9 variables
Month_Avg_Confirmed	3 obs. of 2 variables
Month_Avg_ConfirmedForeign	3 obs. of 2 variables
Month_Avg_ConfirmedIndian	3 obs. of 2 variables
Month_Avg_Cured	3 obs. of 2 variables
Month_Avg_Death	3 obs. of 2 variables
State.UnionTerritory_Confirmed	27 obs. of 2 variables
State.UnionTerritory_Confirmed_Foreign	27 obs. of 2 variables
State.UnionTerritory_Confirmed_Indian	27 obs. of 2 variables
State.UnionTerritory_Cured	27 obs. of 2 variables
State.UnionTerritory_Deaths	27 obs. of 2 variables

Values

Object	Value
Day_Highest_Cases	"28/3/2020"
total_record	446



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

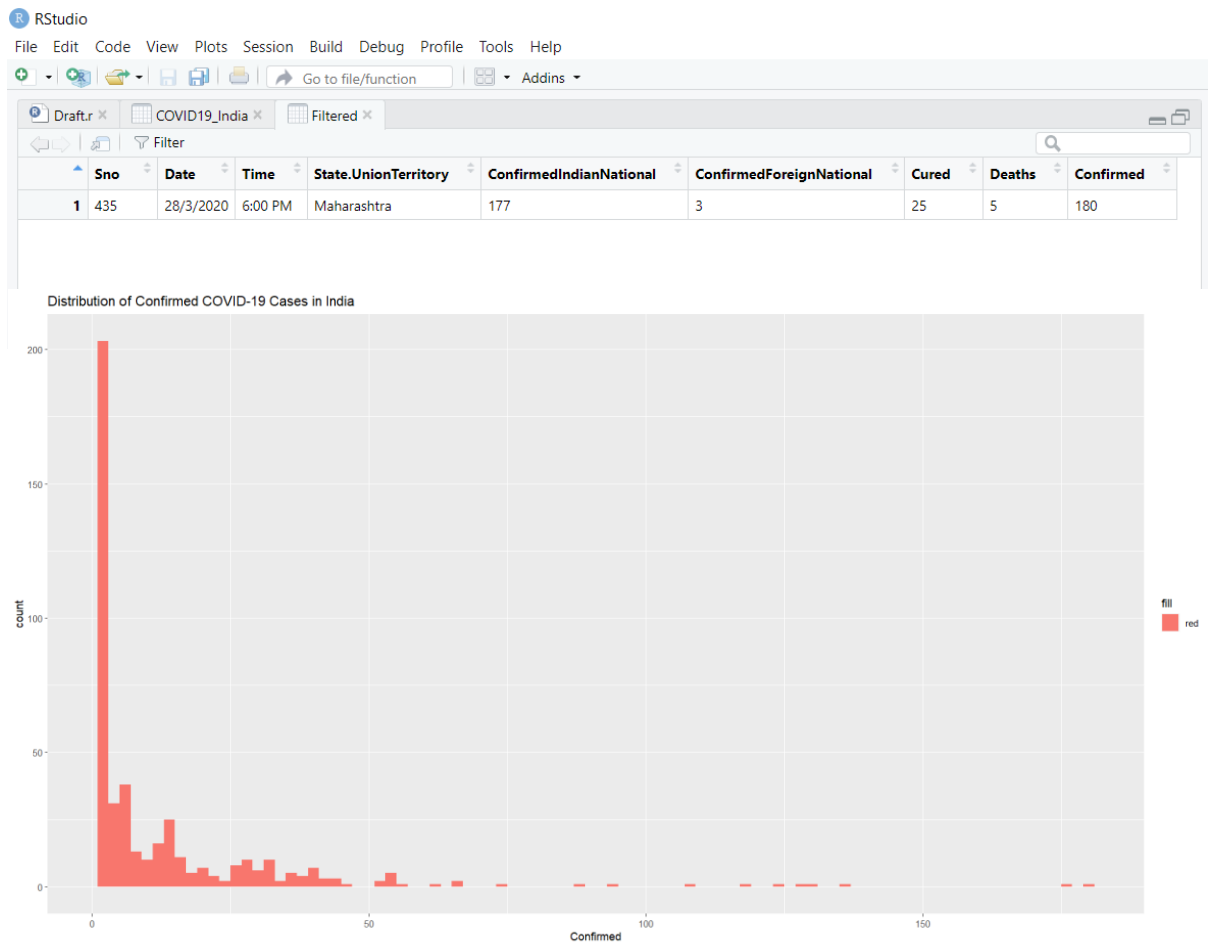
Go to file/function Addins

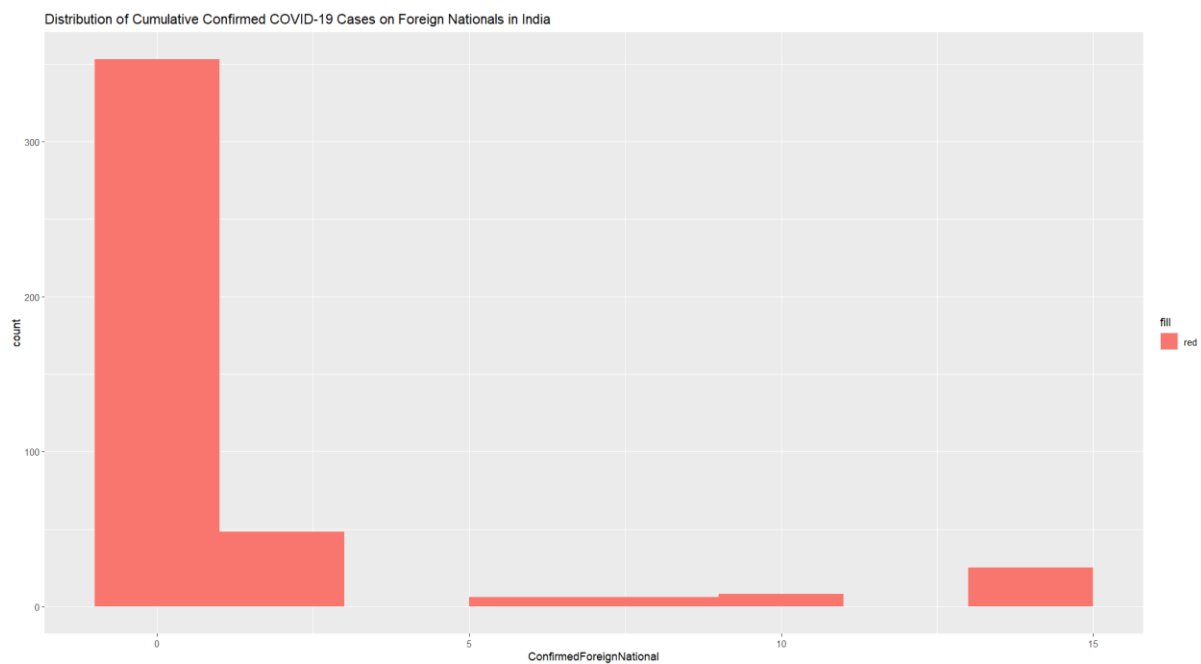
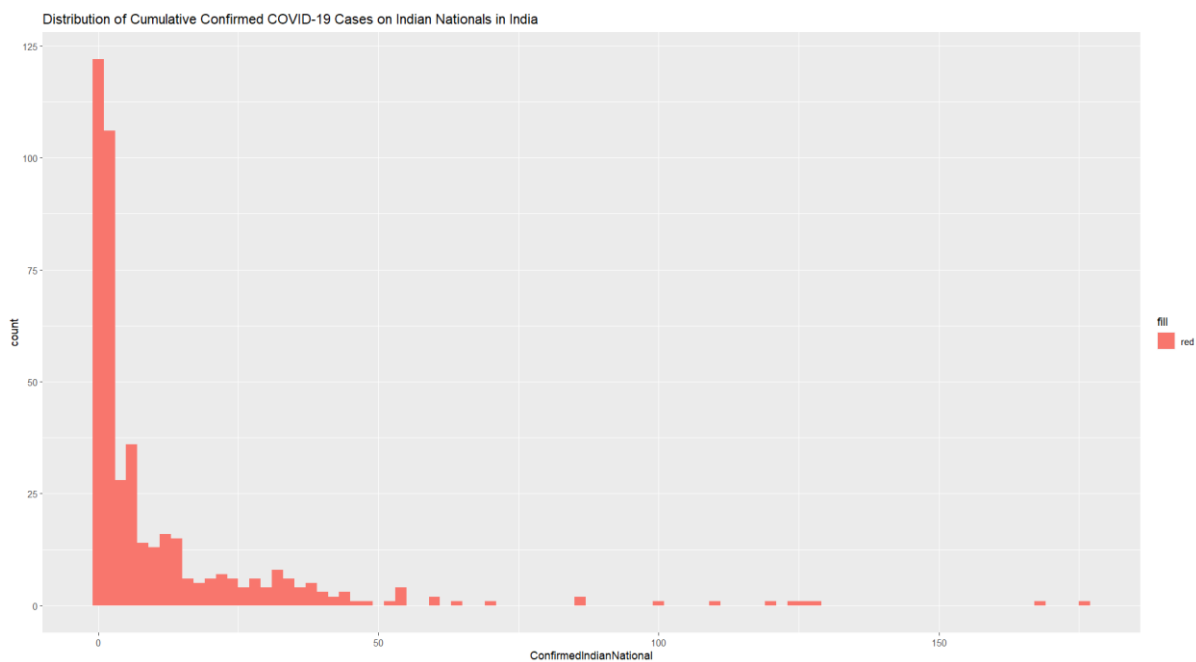
DraftLr COVID19\_India

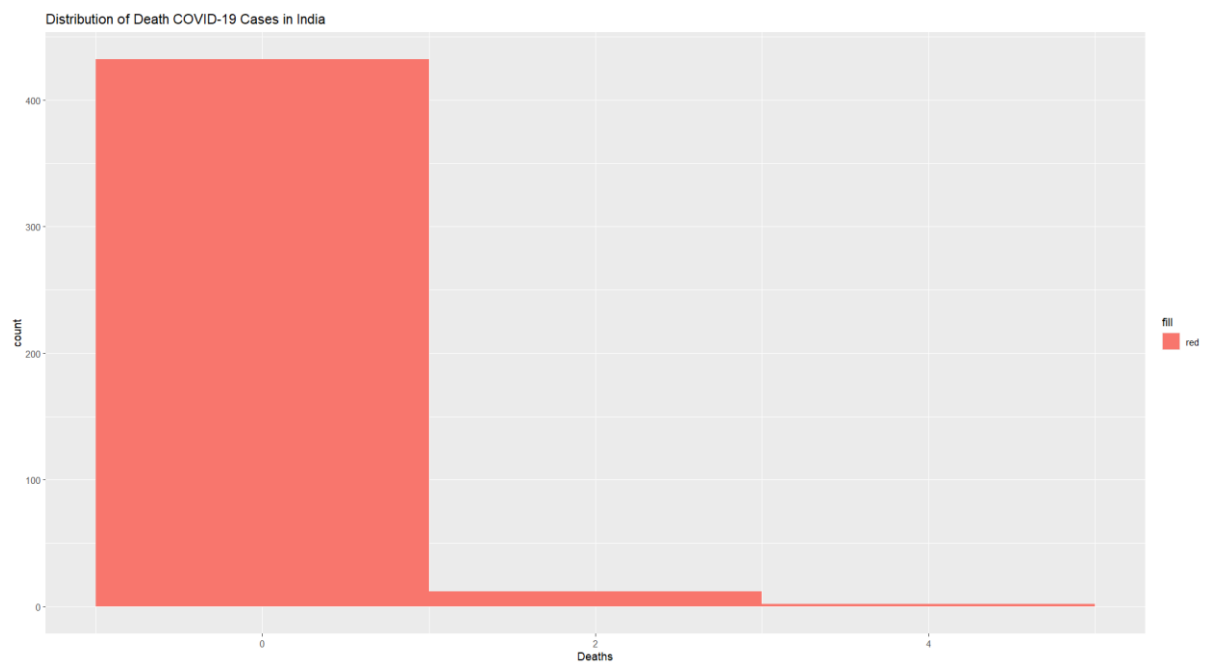
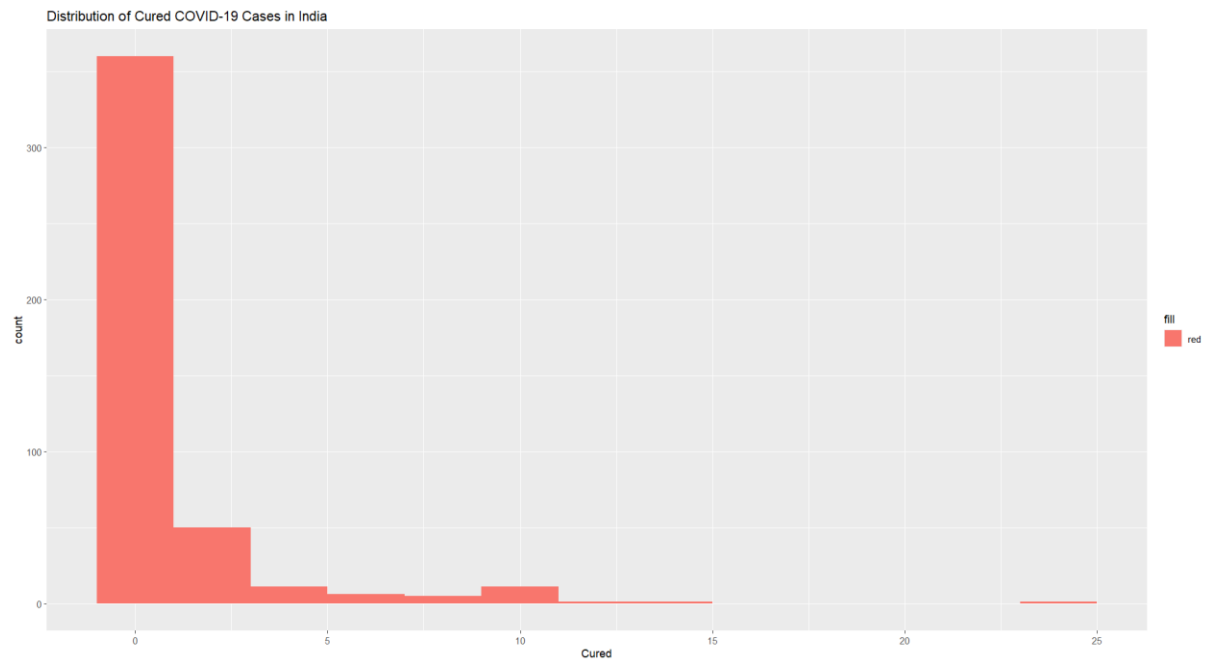
Filter

Sno	Date	Time	State.UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
1	30/1/2020	6:00 PM	Kerala	1	0	0	0	1
2	31/1/2020	6:00 PM	Kerala	1	0	0	0	1
3	1/2/2020	6:00 PM	Kerala	2	0	0	0	2
4	2/2/2020	6:00 PM	Kerala	3	0	0	0	3
5	3/2/2020	6:00 PM	Kerala	3	0	0	0	3
6	4/2/2020	6:00 PM	Kerala	3	0	0	0	3
7	5/2/2020	6:00 PM	Kerala	3	0	0	0	3
8	6/2/2020	6:00 PM	Kerala	3	0	0	0	3
9	7/2/2020	6:00 PM	Kerala	3	0	0	0	3
10	8/2/2020	6:00 PM	Kerala	3	0	0	0	3
11	9/2/2020	6:00 PM	Kerala	3	0	0	0	3
12	10/2/2020	6:00 PM	Kerala	3	0	0	0	3
13	11/2/2020	6:00 PM	Kerala	3	0	0	0	3
14	12/2/2020	6:00 PM	Kerala	3	0	0	0	3
15	13/2/2020	6:00 PM	Kerala	3	0	0	0	3
16	14/2/2020	6:00 PM	Kerala	3	0	0	0	3
17	15/2/2020	6:00 PM	Kerala	3	0	0	0	3
18	16/2/2020	6:00 PM	Kerala	3	0	0	0	3
19	17/2/2020	6:00 PM	Kerala	3	0	0	0	3
20	18/2/2020	6:00 PM	Kerala	3	0	0	0	3
21	19/2/2020	6:00 PM	Kerala	3	0	0	0	3
22	20/2/2020	6:00 PM	Kerala	3	0	0	0	3
23	21/2/2020	6:00 PM	Kerala	3	0	0	0	3
24	22/2/2020	6:00 PM	Kerala	3	0	0	0	3
25	23/2/2020	6:00 PM	Kerala	3	0	0	0	3
26	24/2/2020	6:00 PM	Kerala	3	0	0	0	3
27	25/2/2020	6:00 PM	Kerala	3	0	0	0	3
28	26/2/2020	6:00 PM	Kerala	3	0	0	0	3

Showing 1 to 29 of 18,110 entries, 9 total columns







RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r COVID19\_India Filtered Avg\_Cases

Filter

	Cured	Deaths	Confirmed
1	0	0	1
2	0	0	1
3	0	0	6
4	0	0	1
5	0	0	1
6	0	0	1
7	0	0	1
8	0	0	1
9	0	0	1
10	0	0	1
11	0	0	2
12	0	0	3
13	0	0	3
14	0	0	5
15	0	0	7
16	0	0	8
17	1	0	9
18	1	0	11
19	1	0	12
20	1	0	14
21	0	1	2
22	0	1	2
23	0	1	3
24	0	1	4
25	0	1	6
26	0	1	6
27	0	1	9
28	0	0	1

Showing 1 to 29 of 446 entries, 3 total columns

Console

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

COVID19\_India x Filtered x Avg\_Cases x

Filter

Cured	Deaths	Confirmed	Percentage_Cured	Percentage_Death	Percentage_Confirmed_Indian_National	Percentage_Confirmed_Foreign_National
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	6	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	1	0.000000	0.000000	100.000000	0.000000
0	0	2	0.000000	0.000000	100.000000	0.000000
0	0	3	0.000000	0.000000	100.000000	0.000000
0	0	3	0.000000	0.000000	100.000000	0.000000
0	0	5	0.000000	0.000000	100.000000	0.000000
0	0	7	0.000000	0.000000	100.000000	0.000000
0	0	8	0.000000	0.000000	100.000000	0.000000
1	0	9	11.111111	0.000000	100.000000	0.000000
1	0	11	9.090909	0.000000	100.000000	0.000000
1	0	12	8.333333	0.000000	100.000000	0.000000
1	0	14	7.142857	0.000000	100.000000	0.000000
0	1	2	0.000000	50.000000	100.000000	0.000000
0	1	2	0.000000	50.000000	100.000000	0.000000
0	1	3	0.000000	33.333333	100.000000	0.000000
0	1	4	0.000000	25.000000	100.000000	0.000000
0	1	6	0.000000	16.666667	100.000000	0.000000
0	1	6	0.000000	16.666667	100.000000	0.000000
0	1	9	0.000000	11.111111	100.000000	0.000000

Showing 1 to 28 of 446 entries, 13 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r COVID19\_India Filtered Avg\_Cases

Filter

	Sno	Month	Time	State.UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed	Percent
1	366	Mar	6:00 PM	Andaman and Nicobar Islands	1	0	0	0	1	
2	393	Mar	10:00 AM	Andaman and Nicobar Islands	1	0	0	0	1	
3	421	Mar	6:00 PM	Andaman and Nicobar Islands	6	0	0	0	6	
4	122	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
5	135	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
6	148	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
7	149	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
8	163	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
9	178	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
10	193	Mar	6:00 PM	Andhra Pradesh	1	0	0	0	1	
11	210	Mar	6:00 PM	Andhra Pradesh	2	0	0	0	2	
12	229	Mar	6:00 PM	Andhra Pradesh	3	0	0	0	3	
13	249	Mar	6:00 PM	Andhra Pradesh	3	0	0	0	3	
14	271	Mar	6:00 PM	Andhra Pradesh	5	0	0	0	5	
15	294	Mar	6:00 PM	Andhra Pradesh	7	0	0	0	7	
16	317	Mar	6:00 PM	Andhra Pradesh	8	0	0	0	8	
17	341	Mar	6:00 PM	Andhra Pradesh	9	0	1	0	9	
18	367	Mar	6:00 PM	Andhra Pradesh	11	0	1	0	11	
19	394	Mar	10:00 AM	Andhra Pradesh	12	0	1	0	12	
20	420	Mar	6:00 PM	Andhra Pradesh	14	0	1	0	14	
21	272	Mar	6:00 PM	Bihar	2	0	0	1	2	
22	295	Mar	6:00 PM	Bihar	2	0	0	1	2	
23	318	Mar	6:00 PM	Bihar	3	0	0	1	3	
24	342	Mar	6:00 PM	Bihar	4	0	0	1	4	
25	368	Mar	6:00 PM	Bihar	6	0	0	1	6	
26	395	Mar	10:00 AM	Bihar	6	0	0	1	6	
27	422	Mar	6:00 PM	Bihar	9	0	0	1	9	

Showing 1 to 28 of 446 entries. 13 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r COVID19\_India Filtered Avg\_Cases

Filter

	Cured_mean	Cured_sum	Cured_sd	Deaths_mean	Deaths_sum	Deaths_sd	Confirmed_mean	Confirmed_sum	Confirmed_sd
1	1.089686	486	2.642952	0.2443946	109	0.5965462	13.68386	6103	22.55094

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r Month\_Avg\_Confirmed COVID19\_India Filtered Avg\_Cases

Filter

	Month	Avg_Confirmed
1	Feb	2.965517
2	Jan	1.000000
3	Mar	14.493976

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r x Month\_Avg\_Cured x Month\_Avg\_Confirmed x COVID19\_India x Filtered x Avg\_Cases x

Filter

	Month	Avg_Cured
1	Feb	0.000000
2	Jan	0.000000
3	Mar	1.171084

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r x Month\_Avg\_Death x Month\_Avg\_Cured x Month\_Avg\_Confirmed x COVID19\_India x Filtered x Avg\_Cases x

Filter

	Month	Avg_Deaths
1	Feb	0.0000000
2	Jan	0.0000000
3	Mar	0.2626506

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r x Month\_Avg\_ConfirmedIndian x Month\_Avg\_Death x Month\_Avg\_Cured x Month\_Avg\_Confirmed x COVID19\_India x

Filter

	Month	Avg_Confirmed_Indian
1	Feb	2.965517
2	Jan	1.000000
3	Mar	12.886747

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Draft.r x Month\_Avg\_ConfirmedForeign x Month\_Avg\_ConfirmedIndian x Month\_Avg\_Death x Month\_Avg\_Cured x Month\_Avg\_Confirmed x

Filter

	Month	Avg_Confirmed_Foreign
1	Feb	0.000000
2	Jan	0.000000
3	Mar	1.607229

RStudio

File Edit Code View Plots Session Build Debug Profile

Go to file/function

Draft.r x State.UnionTerritory\_Confirmed x Month

Filter

	State.UnionTerritory	Num_Confirmed
1	Kerala	1160
2	Maharashtra	1147
3	Uttar Pradesh	479
4	Rajasthan	431
5	Karnataka	405
6	Haryana	377
7	Delhi	363
8	Telangana	343
9	Gujarat	261
10	Punjab	231
11	Tamil Nadu	164
12	Ladakh	162
13	Madhya Pradesh	105
14	Jammu and Kashmir	95
15	Andhra Pradesh	81
16	West Bengal	71
17	Chandigarh	50
18	Uttarakhand	40
19	Bihar	32
20	Chhattisgarh	25
21	Odisha	23
22	Himachal Pradesh	21
23	Puducherry	11
24	Goa	9
25	Andaman and Nicobar Islands	8
26	Manipur	5
27	Mizoram	4

Showing 1 to 27 of 27 entries, 2 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools

Go to file/function

Draft.r x State.UnionTerritory\_Confirmed\_Foreign x State.Unio

Filter

	State.UnionTerritory	Num_ConfirmedForeignNational
1	Haryana	266
2	Rajasthan	135
3	Telangana	97
4	Kerala	69
5	Maharashtra	36
6	Tamil Nadu	26
7	Uttar Pradesh	17
8	Delhi	11
9	Gujarat	5
10	Uttarakhand	5
11	Andaman and Nicobar Islands	0
12	Andhra Pradesh	0
13	Bihar	0
14	Chandigarh	0
15	Chhattisgarh	0
16	Goa	0
17	Himachal Pradesh	0
18	Jammu and Kashmir	0
19	Karnataka	0
20	Ladakh	0
21	Madhya Pradesh	0
22	Manipur	0
23	Mizoram	0
24	Odisha	0
25	Puducherry	0
26	Punjab	0
27	West Bengal	0

Showing 1 to 27 of 27 entries, 2 total columns



RStudio

File Edit Code View Plots Session Build Debug Profile Tools

Go to file/function

Draft.r\* x State.UnionTerritory\_Confirmed\_Foreign x State.Uni

Filter

	State.UnionTerritory	Num_ConfirmedIndianNational
1	Maharashtra	1111
2	Kerala	1091
3	Uttar Pradesh	462
4	Karnataka	405
5	Delhi	352
6	Rajasthan	296
7	Gujarat	256
8	Telangana	246
9	Punjab	231
10	Ladakh	162
11	Tamil Nadu	138
12	Haryana	111
13	Madhya Pradesh	105
14	Jammu and Kashmir	95
15	Andhra Pradesh	81
16	West Bengal	71
17	Chandigarh	50
18	Uttarakhand	35
19	Bihar	32
20	Chhattisgarh	25
21	Odisha	23
22	Himachal Pradesh	21
23	Puducherry	11
24	Goa	9
25	Andaman and Nicobar Islands	8
26	Manipur	5
27	Mizoram	4

Showing 1 to 27 of 27 entries, 2 total columns

RStudio

File Edit Code View Plots Session Build Debug

Go to file/function

Draft.r\* x State.UnionTerritory\_Confirmed\_Foreign

Filter

	State.UnionTerritory	Num_Cured
1	Uttar Pradesh	123
2	Kerala	99
3	Delhi	62
4	Haryana	56
5	Rajasthan	43
6	Maharashtra	42
7	Karnataka	21
8	Telangana	14
9	Tamil Nadu	11
10	Ladakh	6
11	Andhra Pradesh	4
12	Jammu and Kashmir	4
13	Punjab	1
14	Andaman and Nicobar Islands	0
15	Bihar	0
16	Chandigarh	0
17	Chhattisgarh	0
18	Goa	0
19	Gujarat	0
20	Himachal Pradesh	0
21	Madhya Pradesh	0
22	Manipur	0
23	Mizoram	0
24	Odisha	0
25	Puducherry	0
26	Uttarakhand	0
27	West Bengal	0

Showing 1 to 27 of 27 entries, 2 total columns

RStudio

File Edit Code View Plots Session Build Debug

Go to file/function

Draft.r\* x State.UnionTerritory\_Confirmed\_Foreign

Filter

	State.UnionTerritory	Num_Deaths
1	Maharashtra	26
2	Karnataka	19
3	Delhi	15
4	Gujarat	13
5	Punjab	10
6	Bihar	7
7	Himachal Pradesh	5
8	West Bengal	5
9	Madhya Pradesh	4
10	Tamil Nadu	3
11	Jammu and Kashmir	2
12	Andaman and Nicobar Islands	0
13	Andhra Pradesh	0
14	Chandigarh	0
15	Chhattisgarh	0
16	Goa	0
17	Haryana	0
18	Kerala	0
19	Ladakh	0
20	Manipur	0
21	Mizoram	0
22	Odisha	0
23	Puducherry	0
24	Rajasthan	0
25	Telengana	0
26	Uttar Pradesh	0
27	Uttarakhand	0

Showing 1 to 27 of 27 entries, 2 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Filter

	COVID19_India.ConfirmedIndianNational	COVID19_India.ConfirmedForeignNational	COVID19_India.Cured	COVID19_India.Deaths	COVID19_India.Confirmed
1	1	0	0	0	1
2	1	0	0	0	1
3	6	0	0	0	6
4	1	0	0	0	1
5	1	0	0	0	1
6	1	0	0	0	1
7	1	0	0	0	1
8	1	0	0	0	1
9	1	0	0	0	1
10	1	0	0	0	1
11	2	0	0	0	2
12	3	0	0	0	3
13	3	0	0	0	3
14	5	0	0	0	5
15	7	0	0	0	7
16	8	0	0	0	8
17	9	0	1	0	9
18	11	0	1	0	11
19	12	0	1	0	12
20	14	0	1	0	14
21	2	0	0	1	2
22	2	0	0	1	2
23	3	0	0	1	3
24	4	0	0	1	4
25	6	0	0	1	6
26	6	0	0	1	6
27	9	0	0	1	9
28	1	0	0	0	1

Showing 1 to 29 of 446 entries, 5 total columns

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Filter

	COVID19_India.ConfirmedIndianNational	COVID19_India.ConfirmedForeignNational	COVID19_India.Cured	COVID19_India.Deaths	COVID19_India.Confirmed
COVID19_India.ConfirmedIndianNational	1.0000000	0.1940886	0.5887546	0.5888112	0.9878246
COVID19_India.ConfirmedForeignNational	0.1940886	1.0000000	0.2261414	-0.04741149	0.3443388
COVID19_India.Cured	0.5887546	0.2261414	1.0000000	0.24404604	0.5993276
COVID19_India.Deaths	0.5888112	-0.04741149	0.2440460	1.0000000	0.5559997
COVID19_India.Confirmed	0.9878246	0.34433876	0.5993276	0.55599975	1.0000000

## Conclusion

The results indicate that there are a total of 6,103 confirmed COVID-19 cases, which include 486 cured cases and 109 death cases in total in the first quarter of 2020. The overall average number of cases in March 2020, including the confirmed (Indian and foreign nationals), cured and death cases, is greater than the cases in January and February 2020. The strongest positive correlation here is the relationship between the cumulative number of confirmed cases and confirmed Indian national cases, followed by the cumulative number of confirmed cases and cured people as well as the cumulative number of confirmed Indian national cases and deaths.

Besides that, the finding also indicates that Kerala has the highest number of confirmed cases ( $n = 1160$ ), followed by Maharashtra ( $n = 1147$ ) and Uttar Pradesh ( $n = 479$ ). In terms of number of cured cases, the top three states or union territory are Uttar Pradesh ( $n = 123$ ), Kerala ( $n = 99$ ) and Delhi ( $n = 62$ ). For deaths, the top three are Maharashtra ( $n = 26$ ), Karnataka ( $n = 19$ ), and Delhi ( $n = 15$ ). For confirmed Indian nationals, the top three are Maharashtra ( $n = 1111$ ), Kerala ( $n = 1091$ ) and Uttar Pradesh ( $n = 462$ ). For confirmed foreigners, Haryana has the highest number ( $n = 266$ ), followed by Rajasthan ( $n = 135$ ) and Telangana ( $n = 97$ ).

Overall, in terms of discussion, these findings are likely to have less generalizability because the number of records in March 2020 in the cleaned data set, which has only records in the first quarter of 2020, is significantly greater than in the other first two months, so the findings will be more inclined towards March 2020.