**Table of Contents**

[Analyzing India's COVID-19 Data 1](#_Toc174468077)

[Introduction 1](#_Toc174468078)

[Data Analysis Workflow 2](#_Toc174468079)

[Applied Techniques/Concepts 2](#_Toc174468080)

[Implemented R Script 2](#_Toc174468081)

[Screenshots of Outputs 5](#_Toc174468082)

[Conclusion 20](#_Toc174468083)

# 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.

### A screenshot of a computer Description automatically generatedImplemented R Script

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### Screenshots of Outputs

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

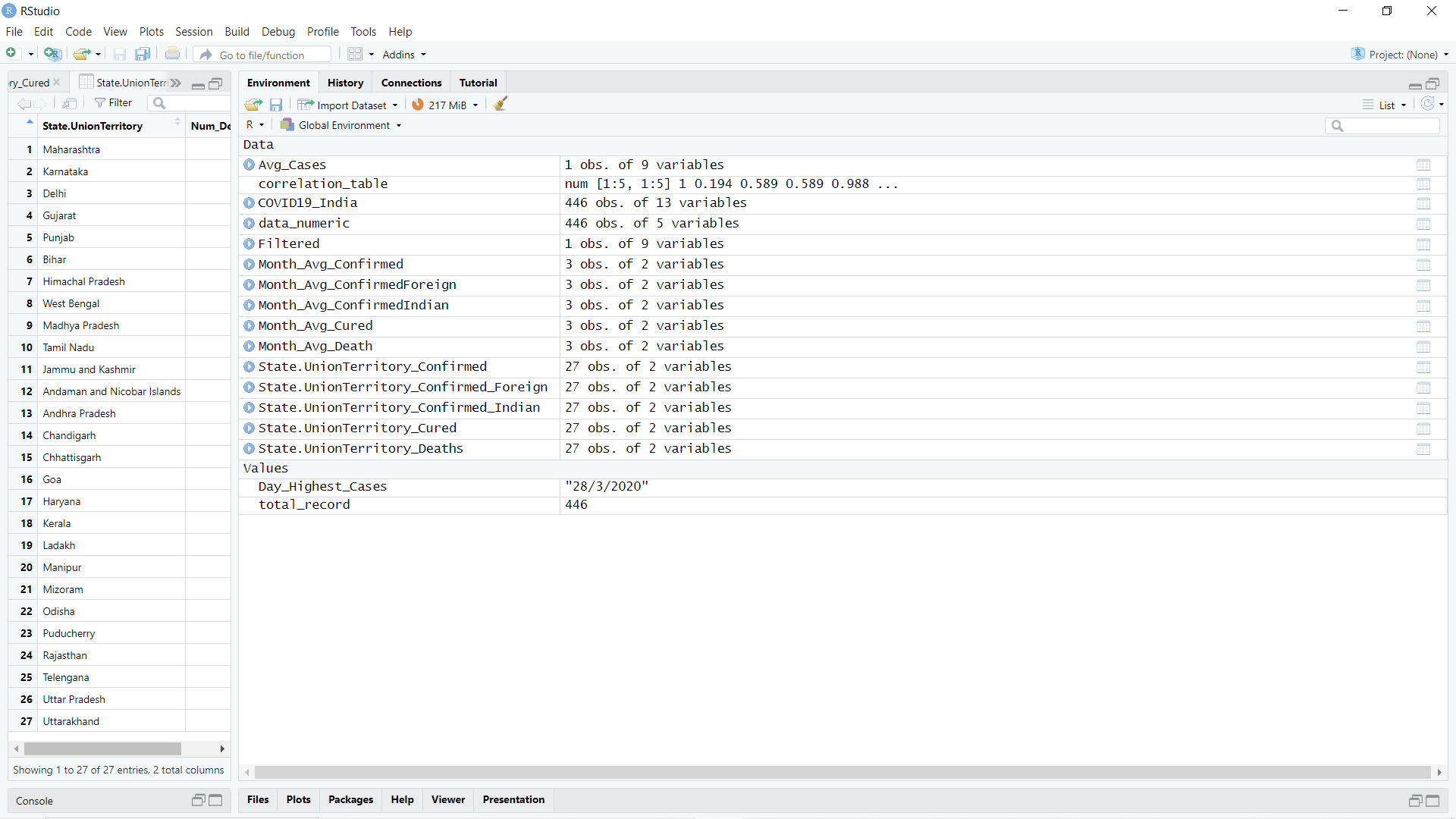
Description automatically generated

A screenshot of a computer

Description automatically generated

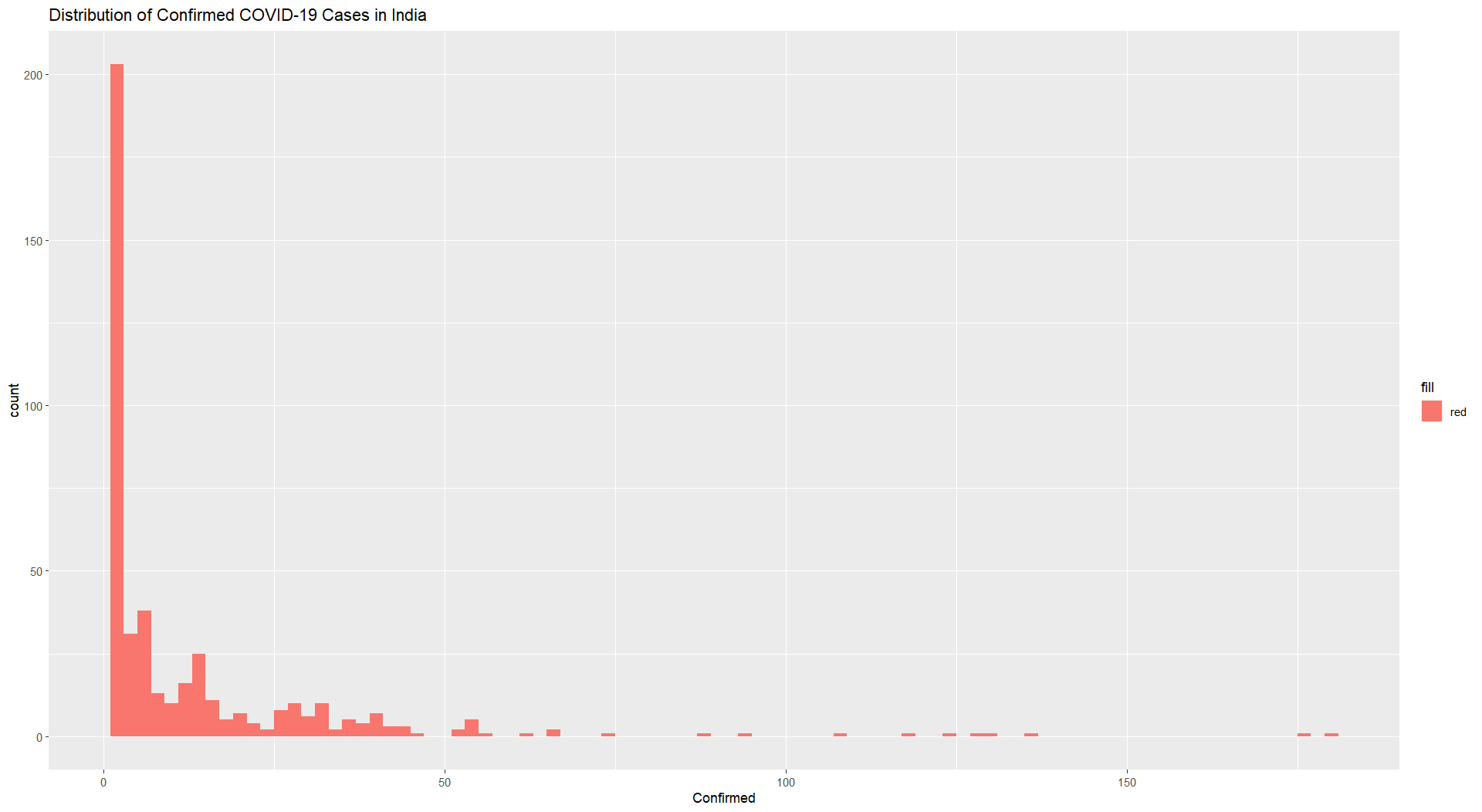
A screenshot of a computer

Description automatically generated

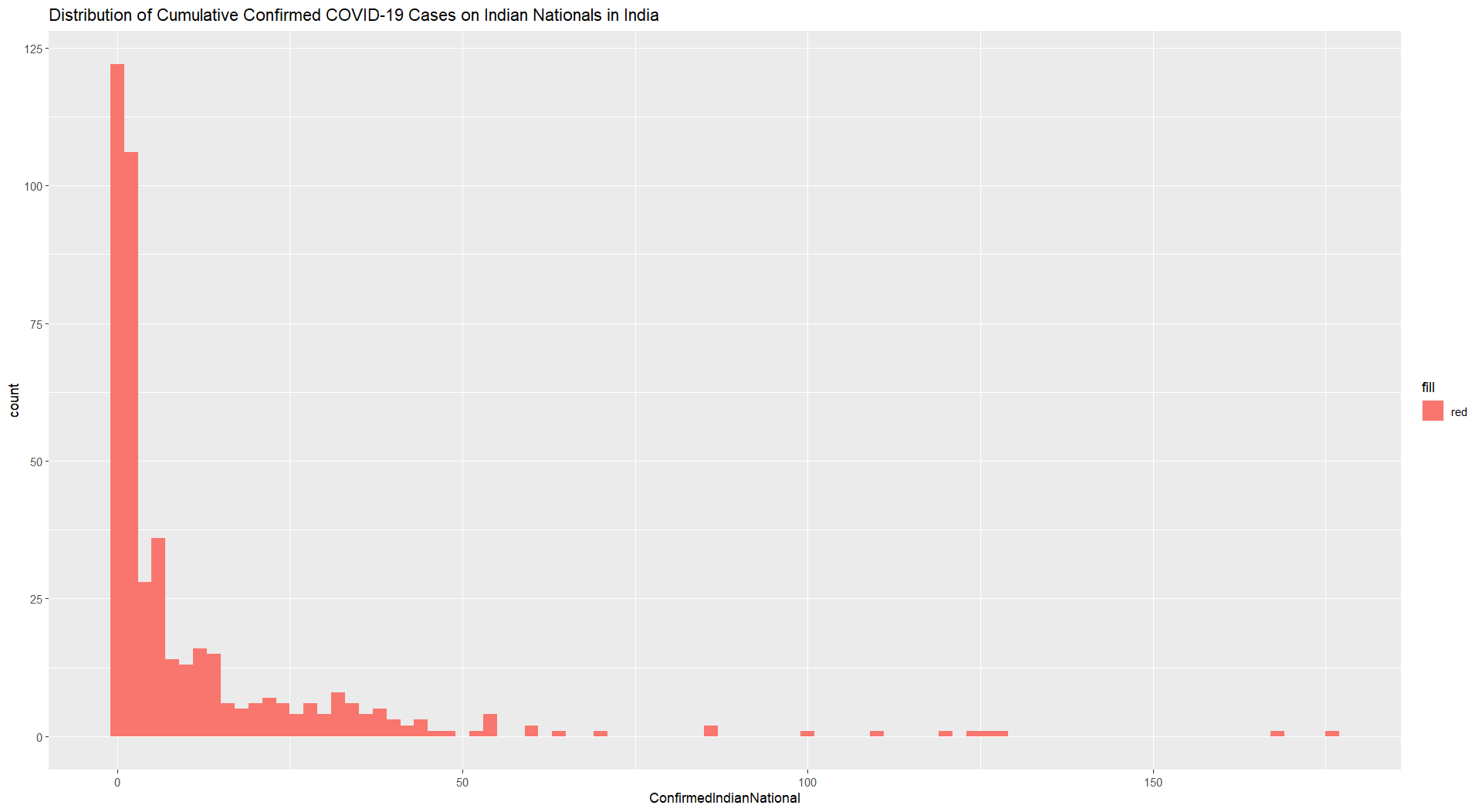


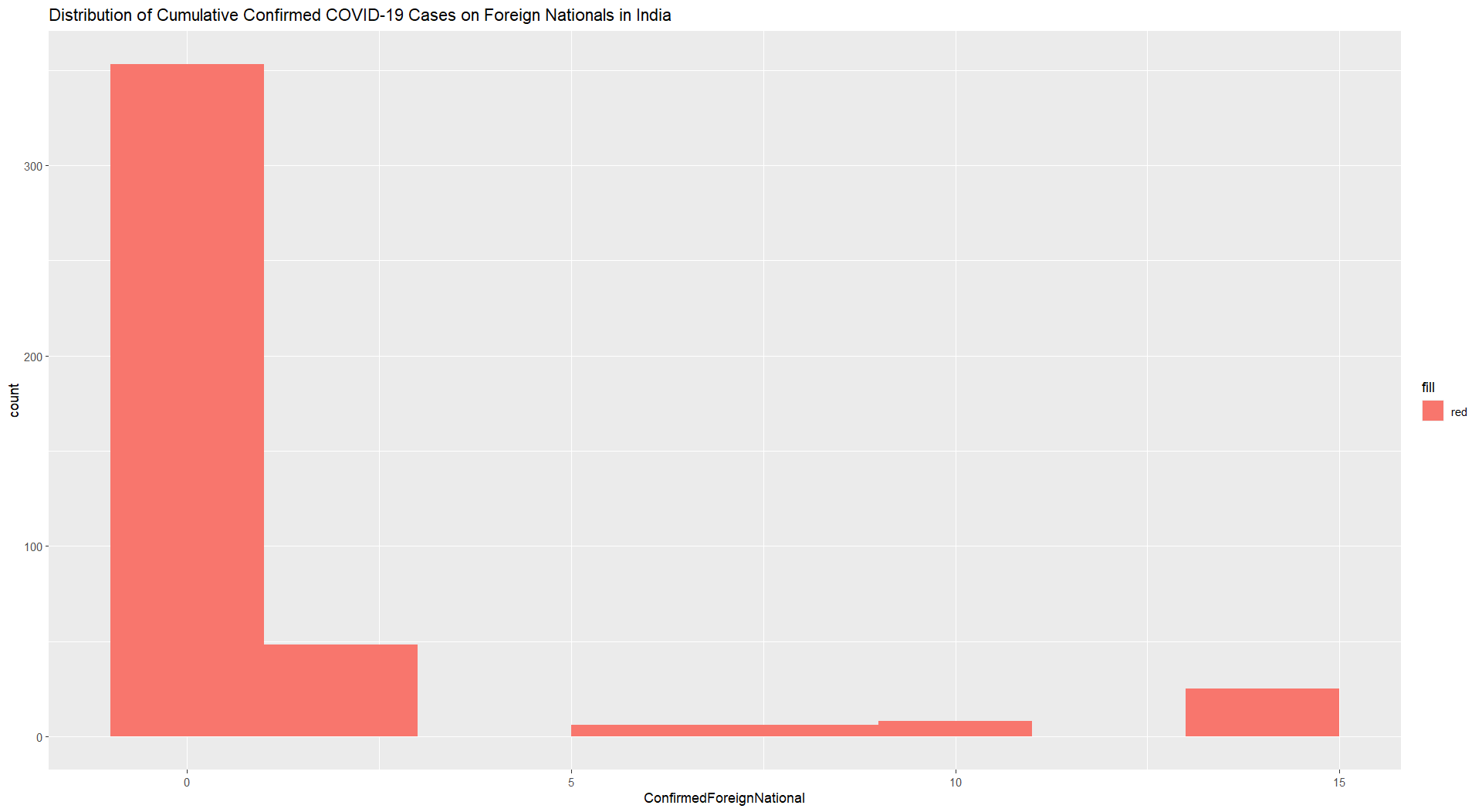
A screenshot of a computer

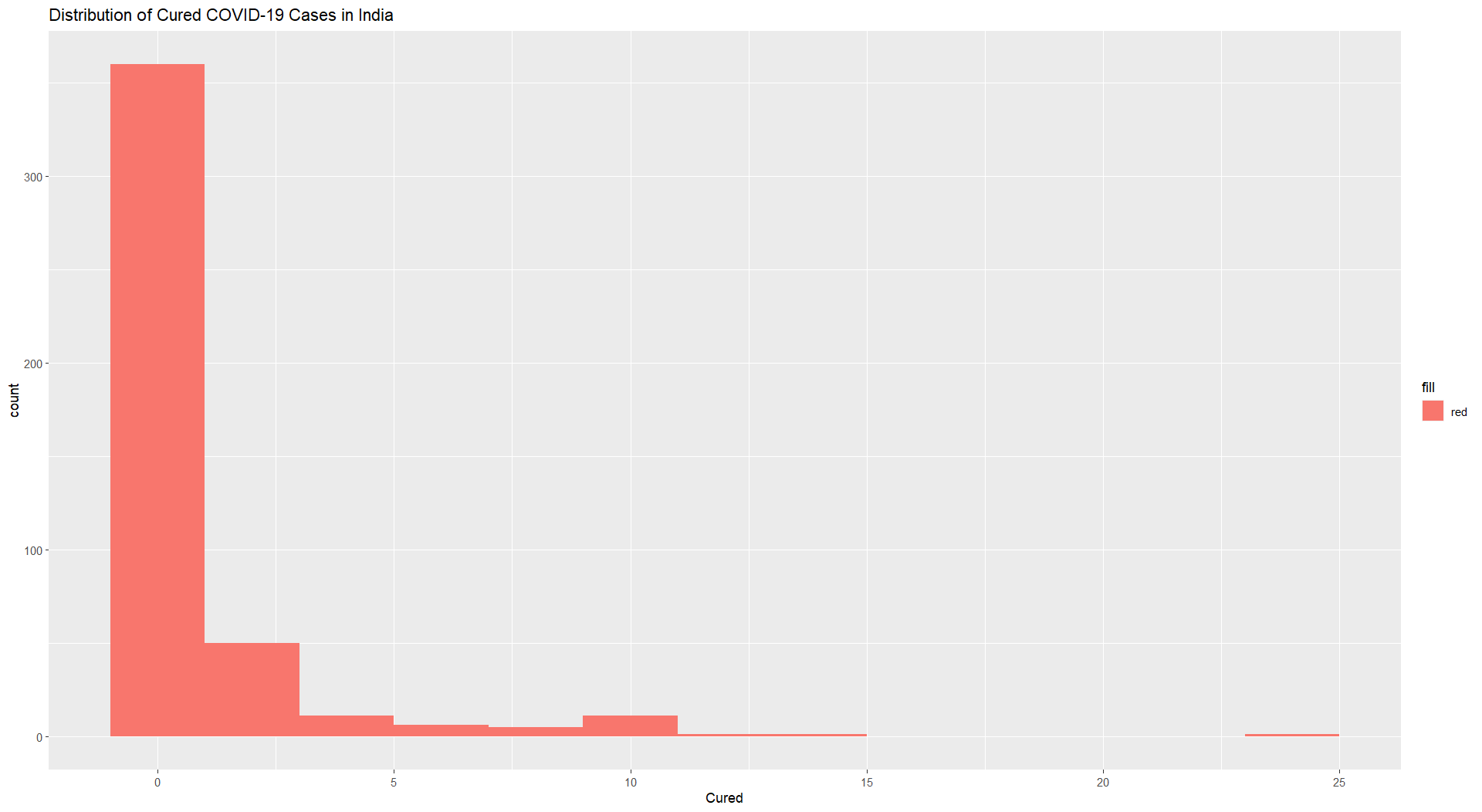
Description automatically generated

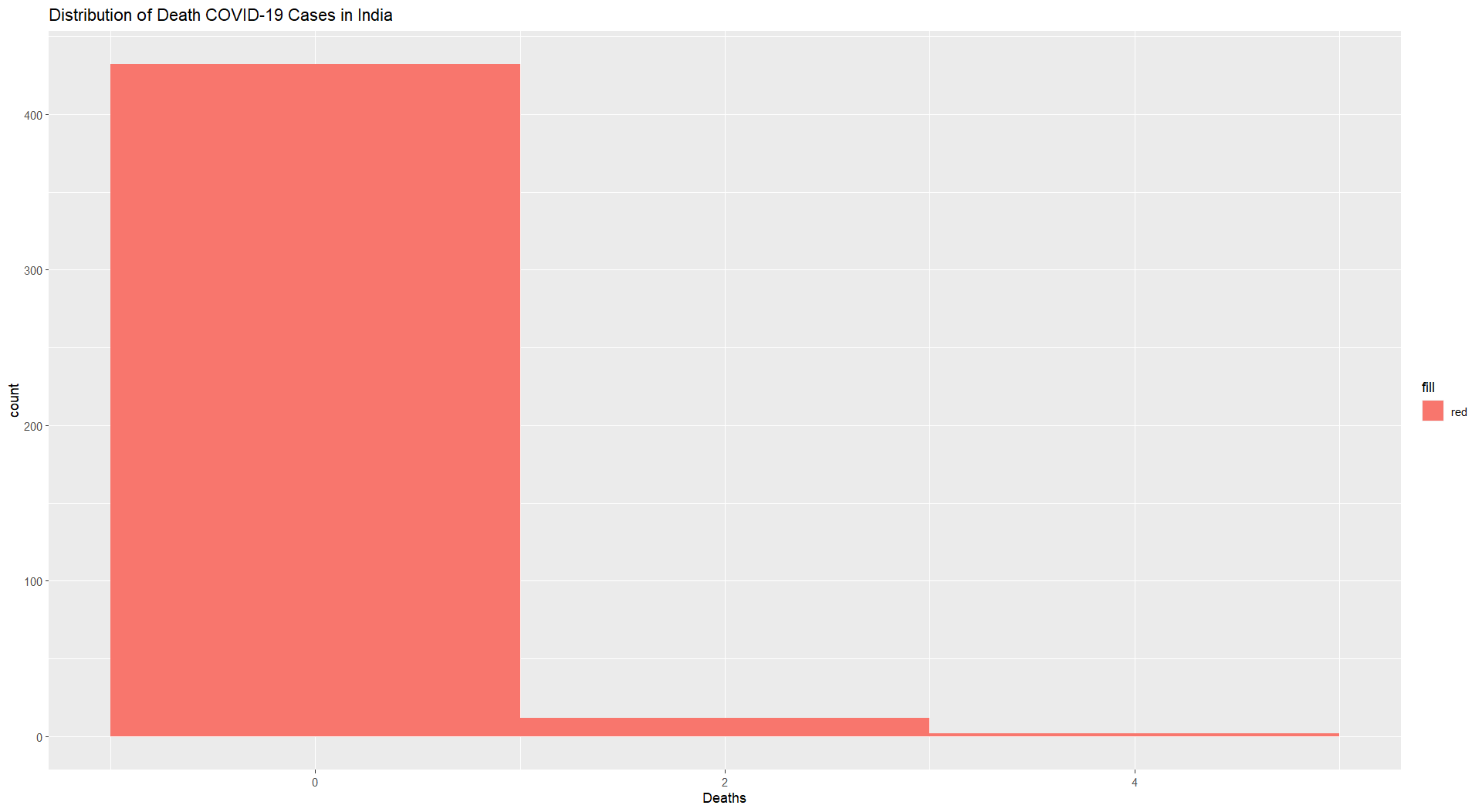
A screenshot of a computer

Description automatically generated









A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## 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.