

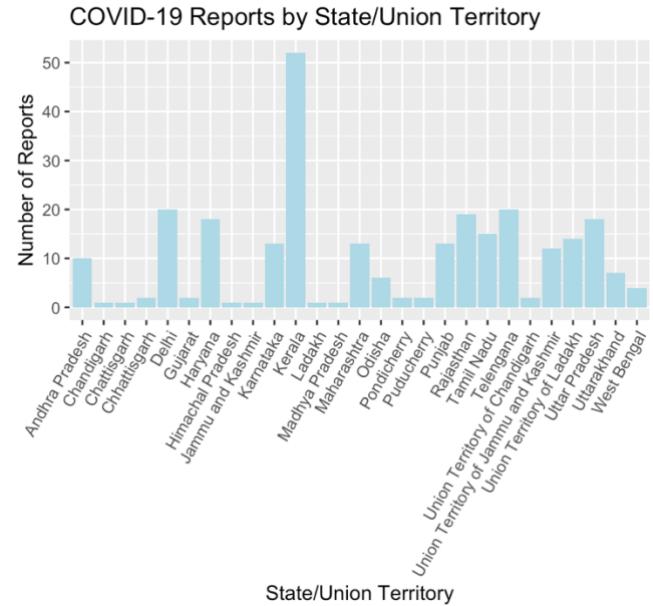
DATA ACTIVITY 4 – PLOTTING WITH R: COVID 19 IN INDIA

Task 4.1 - Create a bar chart showing the frequency of COVID-19 reports by state/union territory.

```
#load libraries
library(readr)
library(ggplot2)

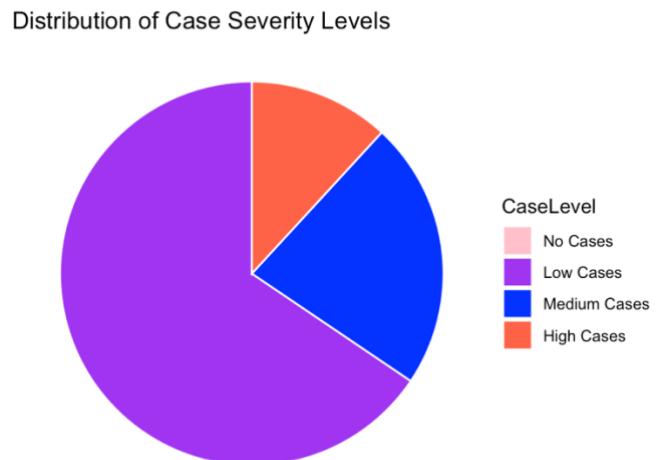
#load dataset
Cov19 <- read_csv("Covid19IndiaJan20Mar20.csv")

#create bar chart to show frequency of C-19 reports by state/union territory
ggplot(
  data=Cov19, aes(x=`State/UnionTerritory`))+
  geom_bar(fill="lightblue")+
  labs(title="COVID-19 Reports by State/Union Territory",
       x="State/Union Territory",
       y="Number of Reports")+
  theme(axis.text.x=element_text(angle=60,hjust=1)) #rotate labels
```



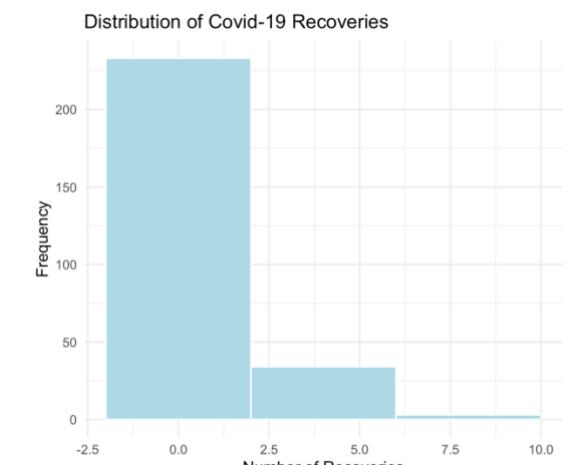
Task 4.2 - Using variables ConfirmedIndianNational and ConfirmedForeignNational, create a pie chart showing the distribution of case severity levels based on total confirmed cases.

```
#Pie chart showing distribution of cases by severity levels
#total cases
Cov19$Total <- Cov19$ConfirmedIndianNational+Cov19$ConfirmedForeignNational
#create categories for case levels
Cov19$CaseLevel <- cut(
  Cov19$Total,
  breaks=c(-Inf,0,5,15,Inf),
  labels=c("No Cases", "Low Cases", "Medium Cases", "High Cases")
)
#create frequency of each case level
Case_freq <- table(Cov19$CaseLevel)
Case_freq
#convert to data frame
casefreq_df <- as.data.frame(Case_freq)
colnames(casefreq_df) <- c("CaseLevel", "count")
#create pie chart with ggplot
ggplot(casefreq_df, aes(x="", y=count, fill=CaseLevel))+ 
  geom_col(width=1, color="white")+
  coord_polar(theta="y")+
  labs(title="Distribution of Case Severity Levels")+
  theme_void()+
  scale_fill_manual(values=c("pink","purple","blue","tomato"))
```



Task 4.3 - Create a histogram showing the distribution of recovery numbers

```
#Histogram showing distribution of recoveries
ggplot(Cov19, aes(x=Cured))+ 
  geom_histogram(binwidth=4, fill="lightblue", color="white")+
  labs(title="Distribution of Covid-19 Recoveries",
       x="Number of Recoveries",
       y="Frequency")+
  theme_minimal()
```



Task 4.4 - Create a line chart showing the trend of total cases over time

```
#Line graph showing total cases over time
#convert date column to date format
Cov19$Date <- as.Date(Cov19$Date, format="%d-%m-%y")
#aggregate total cases by date
daily_cases <- aggregate(Total ~ Date, data=Cov19, sum)
head(daily_cases)
#line graph in ggplot2
ggplot(daily_cases, aes(x=Date, y=Total))+  
  geom_line(color="red",size=1.2)+  
  labs(title="Trend of Total Covid-19 Cases Over Time",  
       x="Date",  
       y="Total Confirmed Cases")+  
  theme_minimal()
```

