ADV

Code

NAME - OM CHANDRA

UID - 2021700014

BATCH - L

EXPT - 4

Hide

```
dataset <- read.csv("C:\\Users\\Om Chandra\\Downloads\\39_Specific_purpose_of_kidnapping_and_
abduction.csv", header = TRUE, stringsAsFactors = FALSE)
```

Hide

```
head(dataset)
```

	Area_Name <chr>	Y… <int>	Group_Name <chr>	Sub_Group_Name <chr>	K_A_Cases_Reported <chr>
1	Andhra Pradesh	2001	Kidnap - For Adoption	01. For Adoption	8
2	Arunachal Pradesh	2001	Kidnap - For Adoption	01. For Adoption	0
3	Assam	2001	Kidnap - For Adoption	01. For Adoption	0
4	Bihar	2001	Kidnap - For Adoption	01. For Adoption	18
5	Chandigarh	2001	Kidnap - For Adoption	01. For Adoption	0
6	Chhattisgarh	2001	Kidnap - For Adoption	01. For Adoption	6

6 rows | 1-6 of 20 columns

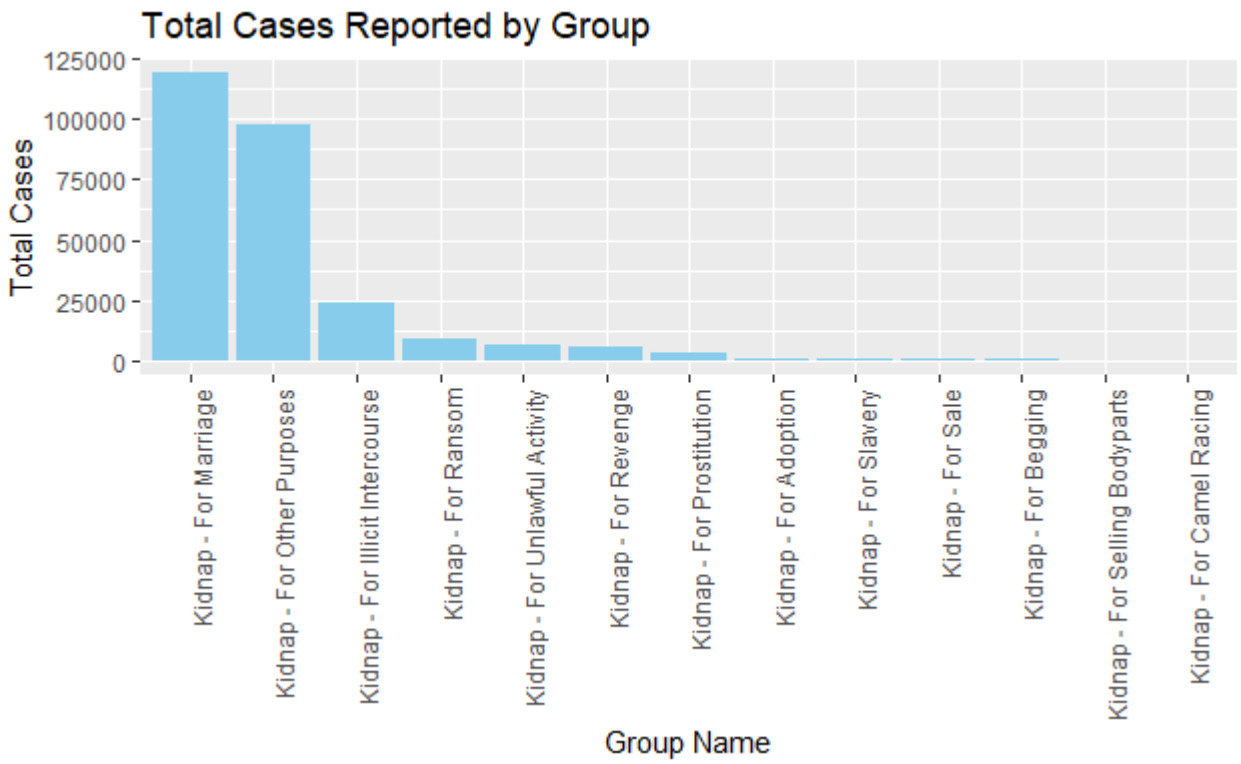
Hide

```
summary(dataset)
```

```
 Area_Name              Year       Group_Name        Sub_Group_Name
Length:3231        Min.   :2001   Length:3231        Length:3231
Class :character   1st Qu.:2003   Class :character   Class :character
Mode  :character   Median :2005   Mode  :character   Mode  :character
                   Mean   :2006
                   3rd Qu.:2008
                   Max.   :2010
K_A_Cases_Reported K_A_Female_10_15_Years K_A_Female_15_18_Years
Length:3231        Length:3231            Length:3231
Class :character   Class :character       Class :character
Mode  :character   Mode  :character       Mode  :character



K_A_Female_18_30_Years K_A_Female_30_50_Years K_A_Female_Above_50_Years
Length:3231            Length:3231            Length:3231
Class :character       Class :character       Class :character
Mode  :character       Mode  :character       Mode  :character



K_A_Female_Total   K_A_Female_Upto_10_Years K_A_Grand_Total
Length:3231        Length:3231              Length:3231
Class :character   Class :character         Class :character
Mode  :character   Mode  :character         Mode  :character



K_A_Male_10_15_Years K_A_Male_15_18_Years K_A_Male_18_30_Years
Length:3231          Length:3231          Length:3231
Class :character     Class :character     Class :character
Mode  :character     Mode  :character     Mode  :character



K_A_Male_30_50_Years K_A_Male_Above_50_Years K_A_Male_Total
Length:3231          Length:3231             Length:3231
Class :character     Class :character        Class :character
Mode  :character     Mode  :character        Mode  :character



K_A_Male_Upto_10_Years
Length:3231
Class :character
Mode  :character
```

Hide

```
library(ggplot2)

dataset$K_A_Grand_Total <- as.numeric(dataset$K_A_Grand_Total)
dataset$K_A_Grand_Total[is.na(dataset$K_A_Grand_Total)] <- 0

total_cases_by_group <- aggregate(K_A_Grand_Total ~ Group_Name, data = dataset, sum, na.rm =
TRUE)


ggplot(total_cases_by_group, aes(x = reorder(Group_Name, -K_A_Grand_Total), y = K_A_Grand_Tot
al)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Total Cases Reported by Group", x = "Group Name", y = "Total Cases") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



1.)Bar Chart - Total Cases Reported by Group

Observations:

i.)The group with the highest total number of cases reported is “Kidnap - For Marriage,” followed by other categories.

ii.)There is a significant difference in the number of cases reported among different groups, highlighting which categories are more and less prevalent.

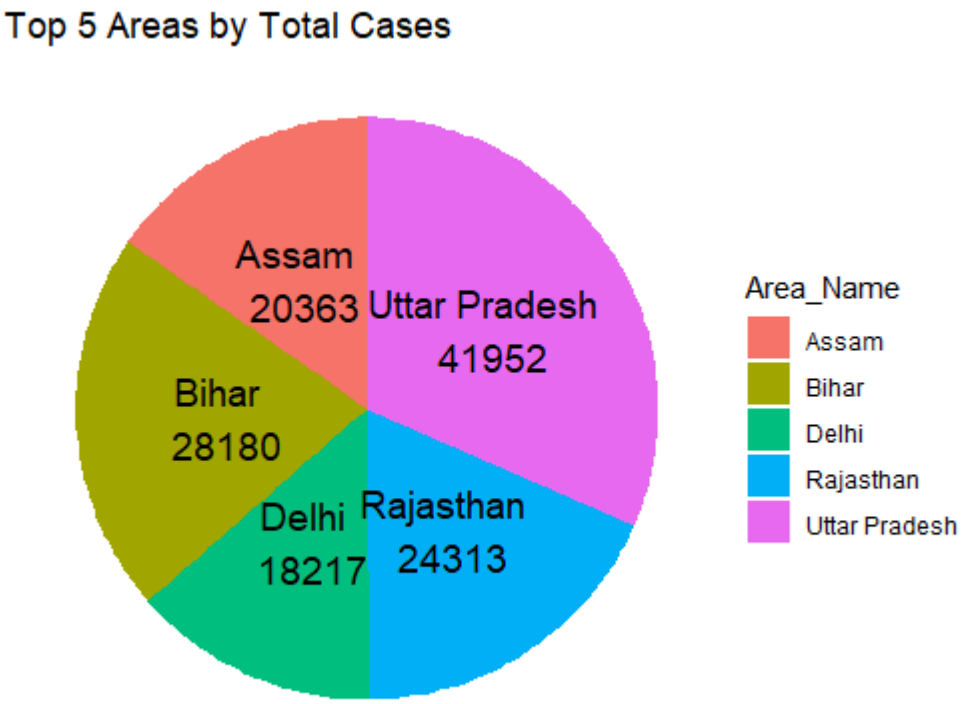
Questions Answered:

i.)Which group has the highest number of kidnapping cases reported?

ii.)How do the totals compare across different kidnapping categories?

Hide

```
total_cases_by_area <- aggregate(K_A_Grand_Total ~ Area_Name, dataset, sum, na.rm = TRUE)

total_cases_by_area$K_A_Grand_Total[is.na(total_cases_by_area$K_A_Grand_Total)] <- 0

top_areas <- total_cases_by_area %>%
  arrange(desc(K_A_Grand_Total)) %>%
  head(5)

ggplot(top_areas, aes(x = "", y = K_A_Grand_Total, fill = Area_Name)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste(Area_Name, "\n", K_A_Grand_Total)),
            position = position_stack(vjust = 0.5), size = 5) +
  labs(title = "Top 5 Areas by Total Cases", x = NULL, y = NULL) +
  theme_void()
```



2.)Pie Chart - Top 5 Areas by Total Cases

Observations:

i.)The top area with the highest number of kidnapping cases is Uttar Pradesh, with clear representation in the pie chart.

ii.)Area with the highest total(Uttar Pradesh) show a considerable difference compared to the lower-ranking areas.

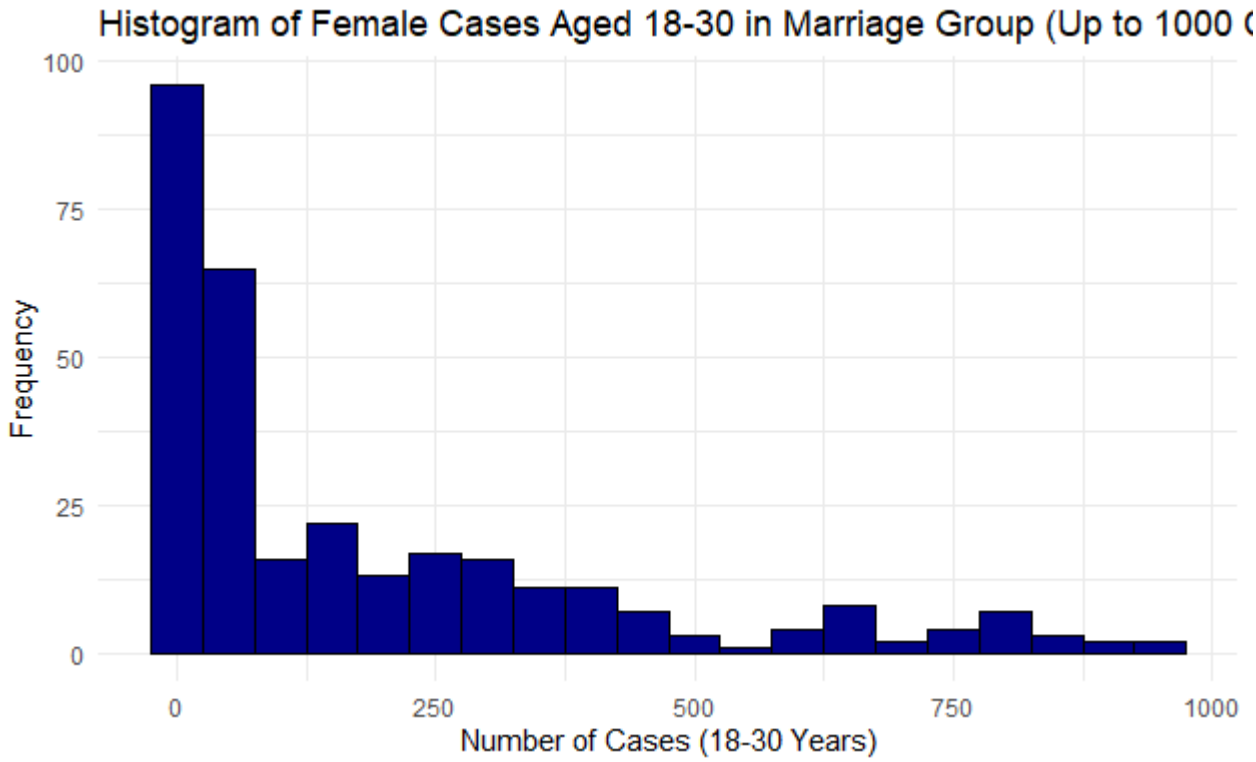
Questions Answered:

i.)Which five areas have the highest total number of kidnapping cases?

ii.)How does the distribution of cases compare across these top areas?

Hide

```
dataset$K_A_Female_18_30_Years <- as.numeric(gsub(",", "",dataset$K_A_Female_18_30_Years))

dataset$K_A_Female_18_30_Years[is.na(dataset$K_A_Female_18_30_Years)] <- 0

filtered_data <- subset(dataset, Group_Name == "Kidnap - For Marriage" & K_A_Female_18_30_Yea
rs <= 1000 & !is.na(K_A_Female_18_30_Years))

ggplot(filtered_data, aes(x = K_A_Female_18_30_Years)) +
  geom_histogram(binwidth = 50, fill = "darkblue", color = "black") +
  labs(title = "Histogram of Female Cases Aged 18-30 in Marriage Group (Up to 1000 Cases)",
       x = "Number of Cases (18-30 Years)",
       y = "Frequency") +
  theme_minimal()
```



3.)Histogram of Female Cases Aged 18-30 in Marriage Group

Observations:

i.)Most instances of female cases aged 18-30 in the “Kidnap - For Marriage” group are clustered within range 0-100, indicating common frequency level of 0-100 cases per year per State.

ii.)There are few instances where the number of cases exceeds 500 indicating instant measures to prevent it.

Questions Answered:

i.)What is the distribution of female cases aged 18-30 in the “Kidnap - For Marriage” group ?

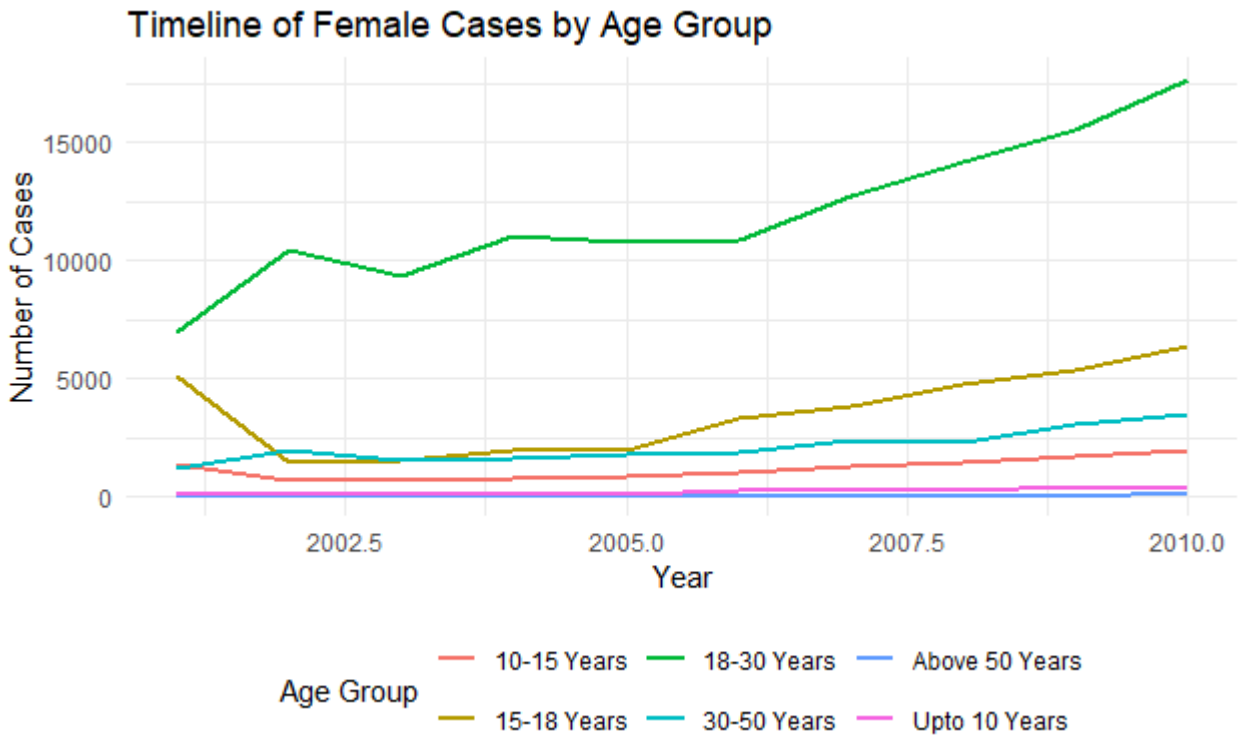
ii.)How frequent are these cases in the specified range?

Hide

```
install.packages("tidyr")
install.packages("dplyr")
```

Hide

```
library(ggplot2)
library(dplyr)
library(tidyr)

convert_to_numeric <- function(column) {
  as.numeric(as.character(column))
}

female_columns <- c("K_A_Female_Upto_10_Years", "K_A_Female_10_15_Years",
                    "K_A_Female_15_18_Years", "K_A_Female_18_30_Years",
                    "K_A_Female_30_50_Years", "K_A_Female_Above_50_Years")

for (col in female_columns) {
  dataset[[col]] <- convert_to_numeric(dataset[[col]])
  dataset[[col]][is.na(dataset[[col]])] <- 0
}

# Aggregate data by year and age group for females
female_data <- dataset %>%
  group_by(Year) %>%
  summarise(
    `Upto 10 Years` = sum(K_A_Female_Upto_10_Years, na.rm = TRUE),
    `10-15 Years` = sum(K_A_Female_10_15_Years, na.rm = TRUE),
    `15-18 Years` = sum(K_A_Female_15_18_Years, na.rm = TRUE),
    `18-30 Years` = sum(K_A_Female_18_30_Years, na.rm = TRUE),
    `30-50 Years` = sum(K_A_Female_30_50_Years, na.rm = TRUE),
    `Above 50 Years` = sum(K_A_Female_Above_50_Years, na.rm = TRUE)
  )

# Reshape data for plotting
female_long <- female_data %>%
  pivot_longer(cols = `Upto 10 Years`:`Above 50 Years`, names_to = "Age_Group", values_to =
"Cases")

# Plot timeline chart with solid lines
ggplot(female_long, aes(x = Year, y = Cases, color = Age_Group)) +
  geom_line(size = 1) +  # Default line type is solid
  labs(title = "Timeline of Female Cases by Age Group",
       x = "Year",
       y = "Number of Cases",
       color = "Age Group") +
  theme_minimal() +
  theme(legend.position = "bottom")
```



4.)Timeline chart - Timeline of Female Cases by Age Group

Observations:

i.)Trends in female cases across different age groups are observable, with variations over the years.Some age groups, such as “18-30 Years”,“15-18 years” show significant increases in cases post 2005

ii.)There is a significant differnce between categories indicating to take measures for a particular category.

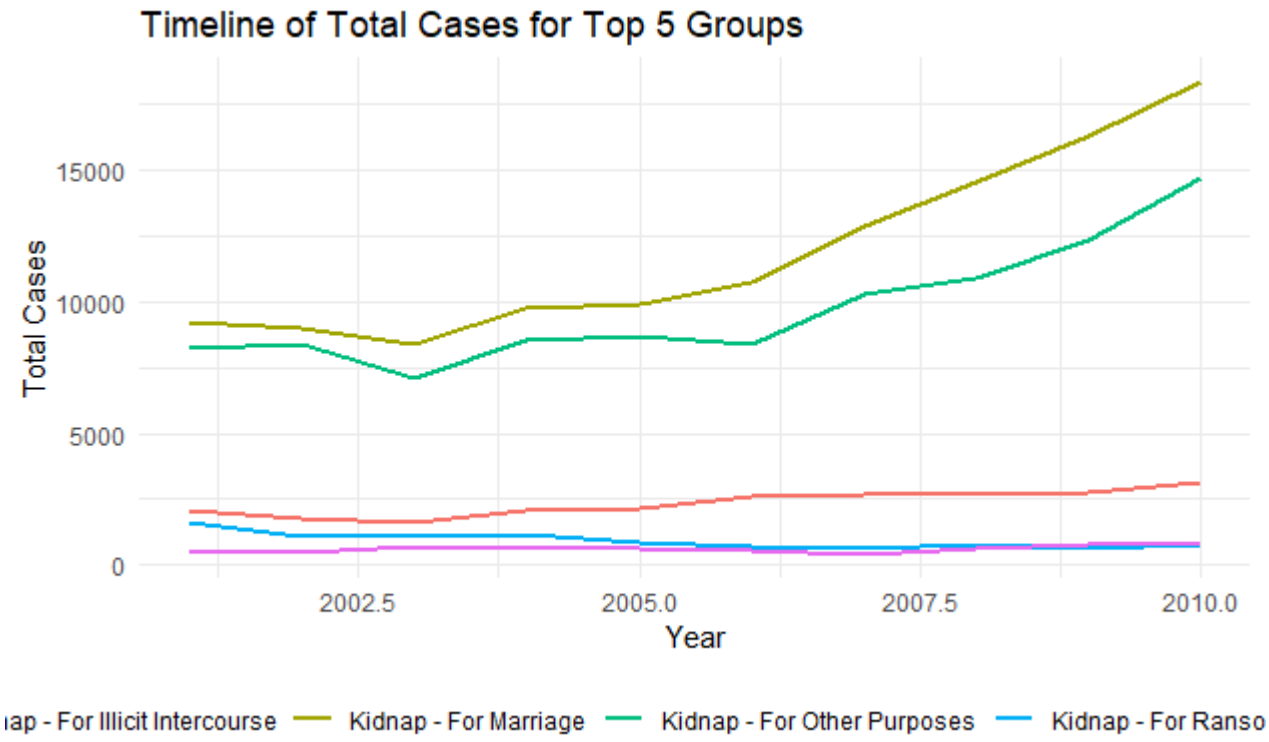
Questions Answered:

i.)How do the number of female cases vary by age group over the years?

ii.)Are there noticeable trends or patterns in case numbers across different age groups?

Hide

```
library(ggplot2)
library(dplyr)
library(tidyr)

dataset$K_A_Grand_Total <- as.numeric(gsub(",", "", dataset$K_A_Grand_Total))

group_total_cases <- dataset %>%
  group_by(Group_Name, Year) %>%
  summarise(Total_Cases = sum(K_A_Grand_Total, na.rm = TRUE), .groups = "drop")

top_5_groups <- group_total_cases %>%
  group_by(Group_Name) %>%
  summarise(Total_Cases = sum(Total_Cases, na.rm = TRUE)) %>%
  top_n(5, Total_Cases) %>%
  pull(Group_Name)

filtered_data <- group_total_cases %>%
  filter(Group_Name %in% top_5_groups)

ggplot(filtered_data, aes(x = Year, y = Total_Cases, color = Group_Name)) +
  geom_line(size = 1) +
  labs(title = "Timeline of Total Cases for Top 5 Groups",
       x = "Year",
       y = "Total Cases",
       color = "Group Name") +
  theme_minimal() +
  theme(legend.position = "bottom")
```



5.)Timeline chart -Timeline of Total Cases for Top 5 Groups

Observations:

i.)The top five groups exhibit distinct trends in the number of cases reported over time.

ii.)There is significant rise in kidnap cases for Marriage and other purposes post 2006

iii.)we can a downtrend in kidnap cases for Unlawful Activity.

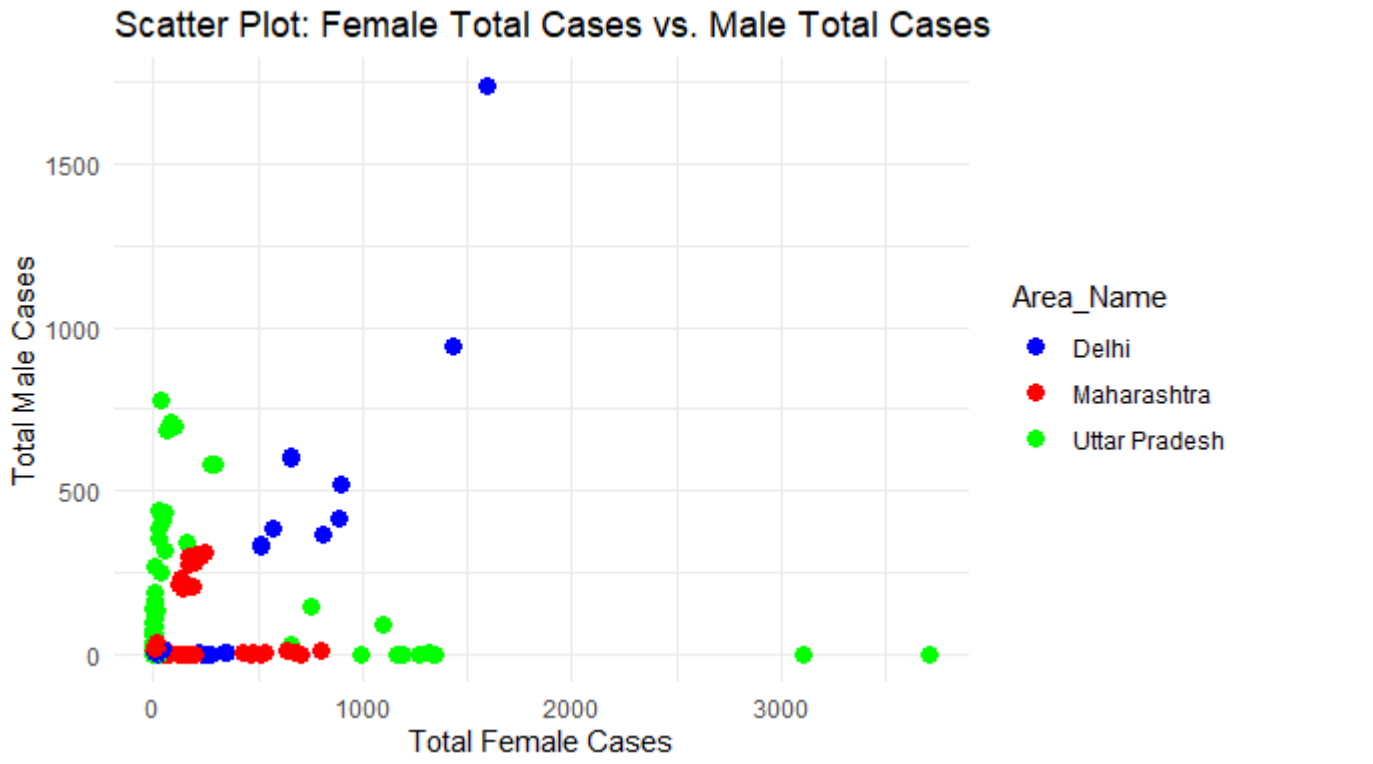
Questions Answered:

i.)What are the trends in total cases for the top five kidnapping groups over the years?

ii.)How do these trends vary among the top groups?

Hide

```
dataset$K_A_Female_Total <- as.numeric(as.character(dataset$K_A_Female_Total))
dataset$K_A_Male_Total <- as.numeric(as.character(dataset$K_A_Male_Total))

filtered_data <- subset(dataset, Area_Name %in% c("Delhi", "Maharashtra", "Uttar Pradesh") &
                                !is.na(K_A_Female_Total) &
                                !is.na(K_A_Male_Total))

ggplot(filtered_data, aes(x = K_A_Female_Total, y = K_A_Male_Total, color = Area_Name)) +
  geom_point(size = 3) +
  labs(title = "Scatter Plot: Female Total Cases vs. Male Total Cases",
       x = "Total Female Cases",
       y = "Total Male Cases") +
  scale_color_manual(values = c("Delhi" = "blue", "Maharashtra" = "red", "Uttar Pradesh" = "g
reen")) +
  theme_minimal()
```



6.)Scatter Plot - Female Total Cases vs. Male Total Cases in Specific States

Observations:

i.)Delhi have both Male and Female cases high leading to high total cases.

ii.)Maharashtra tend to have more female cases and Uttar Pradesh tends to have male and female cases both moderate.

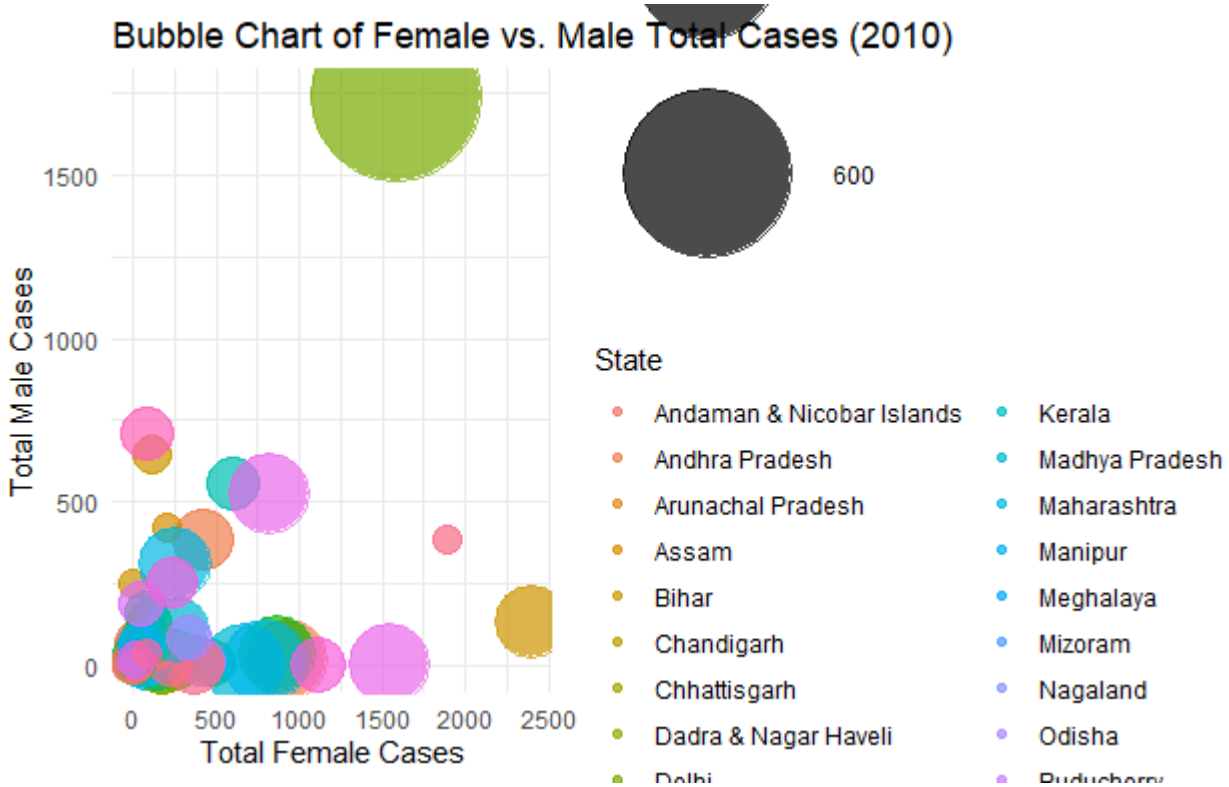
Questions Answered:

i.)How do total female cases compare to total male cases in different states?

ii.)Are there any states with exceptionally high or low case totals for either gender?

Hide

```
filtered_data_2010 <- filtered_data_2010 %>%
  mutate(Bubble_Size = as.numeric(as.character(K_A_Female_10_15_Years)))


# Create the bubble chart
ggplot(filtered_data_2010, aes(x = K_A_Female_Total, y = K_A_Male_Total, size = Bubble_Size,
color = Area_Name)) +
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(5, 30), name = "Cases Aged 10-15 Years") +
  labs(title = "Bubble Chart of Female vs. Male Total Cases (2010)",
       x = "Total Female Cases",
       y = "Total Male Cases",
       color = "State") +
  theme_minimal()
```



7.)Bubble Chart - Female vs. Male Total Cases in 2010 with female cases in 10-15 years representing size.

Observations:

i.)The bubble size represents the number of cases aged 10-15 years, showing a clear distribution across states.

ii.)The chart reveals states with high total cases for both genders and significant numbers of cases in the 10-15 age group.FOr example Delhi has high total cases as well as high cases in 10-15 age grp so Delhi must look into Child Safety.

Questions Answered:

i.)How do female and male total cases compare specifically in 2010?

ii.)What is the distribution of cases aged 10-15 years across different states?

iii.)Which State should look into Child Safety more?