

# A 20 Year Analysis of Crimes on Women in India

Md. Tanvir Rahman

2024-09-17

```
# Import 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(ggplot2)
```

```
library(forcats)
```

```
# Import dataset stored as csv file format
```

```
getwd()
```

```
## [1] "D:/r learning/self learning"
```

```
setwd("D:\\r learning\\self learning")
```

```
crimes_on_women <- read.csv("crimes_on_women.csv")
```

```
# Data explore and check if any value is missing
```

```
View(crimes_on_women)
```

```
head(crimes_on_women)
```

```
##   X      State Year Rape K.A DD  AoW  AoM  DV WT
## 1 0    ANDHRA PRADESH 2001  871  765 420 3544 2271 5791 7
## 2 1 ARUNACHAL PRADESH 2001   33   55   0   78    3   11  0
## 3 2              ASSAM 2001  817 1070  59  850    4 1248  0
## 4 3              BIHAR 2001  888  518 859  562   21 1558 83
## 5 4    CHHATTISGARH 2001  959  171  70 1763  161  840  0
## 6 5              GOA  2001   12    6   2   17    7   11  0
```

```
tail(crimes_on_women)
```

```
##      X      State Year Rape  K.A  DD  AoW AoM  DV WT
## 731 730 Chandigarh 2021   3   23   0   7   0   6  0
## 732 731 D&N Haveli 2021 1250 4083 141 2068 417 4731 4
## 733 732 Daman & Diu 2021  315  904  16 1851  10  501  1
## 734 733   Delhi UT 2021   2    1   0   5   1   9  0
## 735 734 Lakshadweep 2021   0    0   0   1   1   3  0
## 736 735 Puducherry 2021   2    0   2  31   3  12  0
```

```
str(crimes_on_women)
```

```
## 'data.frame':   736 obs. of  10 variables:
## $ X      : int  0 1 2 3 4 5 6 7 8 9 ...
## $ State: chr  "ANDHRA PRADESH" "ARUNACHAL PRADESH" "ASSAM" "BIHAR" ...
## $ Year  : int  2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 ...
## $ Rape  : int  871 33 817 888 959 12 286 398 124 169 ...
## $ K.A   : int  765 55 1070 518 171 6 857 297 105 504 ...
## $ DD    : int  420 0 59 859 70 2 67 285 10 13 ...
## $ AoW   : int  3544 78 850 562 1763 17 756 478 310 622 ...
## $ AoM   : int  2271 3 4 21 161 7 111 401 14 288 ...
## $ DV    : int  5791 11 1248 1558 840 11 3667 1513 317 50 ...
## $ WT    : int  7 0 0 83 0 0 0 0 0 0 ...
```

```
names(crimes_on_women)
```

```
## [1] "X"      "State" "Year"  "Rape"  "K.A"   "DD"    "AoW"   "AoM"   "DV"
## [10] "WT"
```

```
sum(is.na(crimes_on_women)) #check missing values
```

```
## [1] 0
```

```
# Convert all entries of the 'State' column to uppercase
```

```
crimes_on_women <- crimes_on_women %>%
  mutate(State = toupper(State))
```

```
# Data Analysis begins from here
```

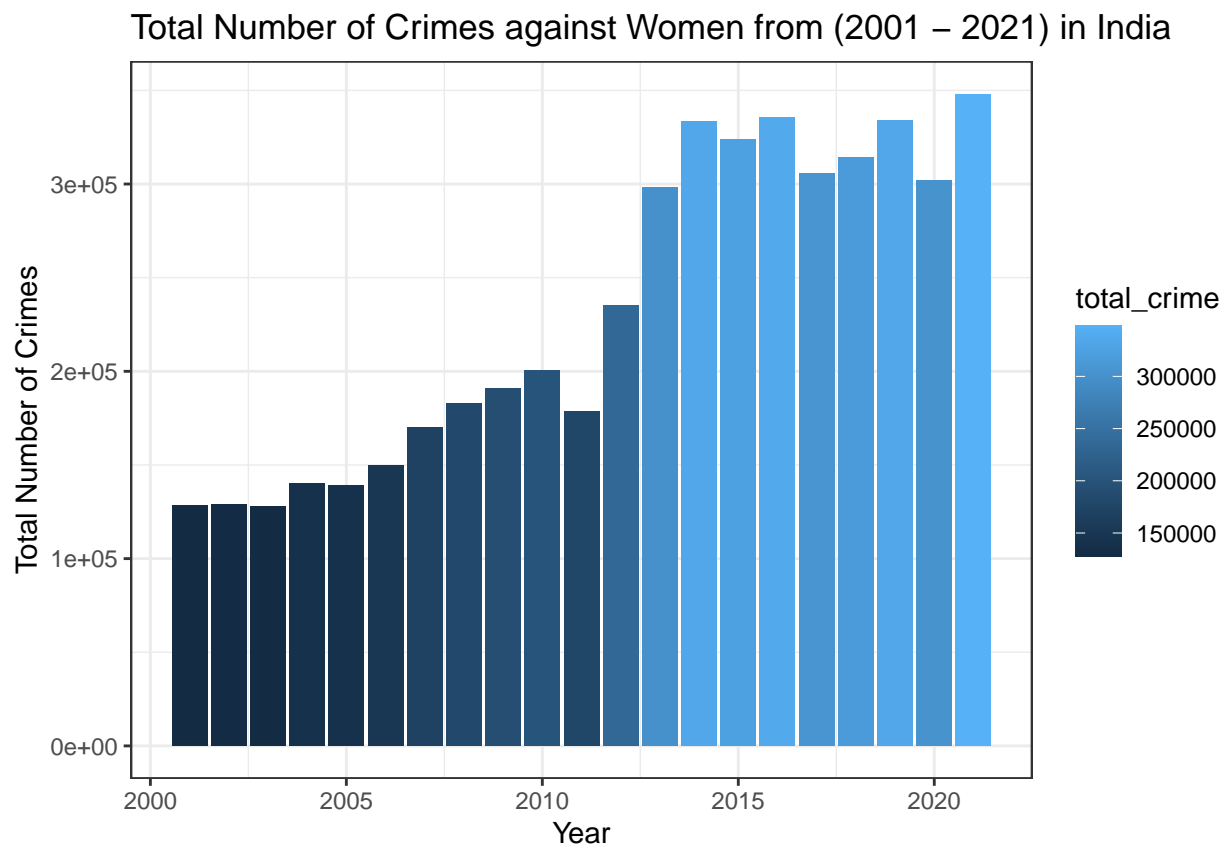
```
## total number of crimes by year
```

```
total_crime_yr <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_crime = sum(Rape + K.A + DD + AoW + AoM + DV + WT)) %>%
  arrange(desc(total_crime))
print(total_crime_yr)
```

```
## # A tibble: 21 x 2
##   Year total_crime
##   <int>     <int>
## 1  2021     348092
## 2  2016     335769
## 3  2019     333717
## 4  2014     333216
## 5  2015     323852
## 6  2018     314093
## 7  2017     305897
## 8  2020     302186
## 9  2013     298444
## 10 2012     235025
## # i 11 more rows
```

```
# Display total number of crimes by year in bar plot
```

```
total_crime_yr %>%
  ggplot(aes(Year, total_crime, fill = total_crime)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "Total Number of Crimes",
       title = "Total Number of Crimes against Women from (2001 - 2021) in India") +
  theme_bw()
```



```
# Find and print sum total of various crimes
```

```
total_rape <- sum(crimes_on_women$Rape)
total_K.A <- sum(crimes_on_women$K.A)
total_DD <- sum(crimes_on_women$DD)
total_Aow <- sum(crimes_on_women$AoW)
total_AoM <- sum(crimes_on_women$AoM)
total_DV <- sum(crimes_on_women$DV)
total_WT <- sum(crimes_on_women$WT)
```

```
total_rape
```

```
## [1] 535702
```

```
total_K.A
```

```
## [1] 835023
```

```
total_DD
```

```
## [1] 158750
```

```
total_Aow
```

```
## [1] 1162229
```

```
total_Aow
```

```
## [1] 1162229
```

```
total_DV
```

```
## [1] 1909978
```

```
total_WT
```

```
## [1] 21156
```

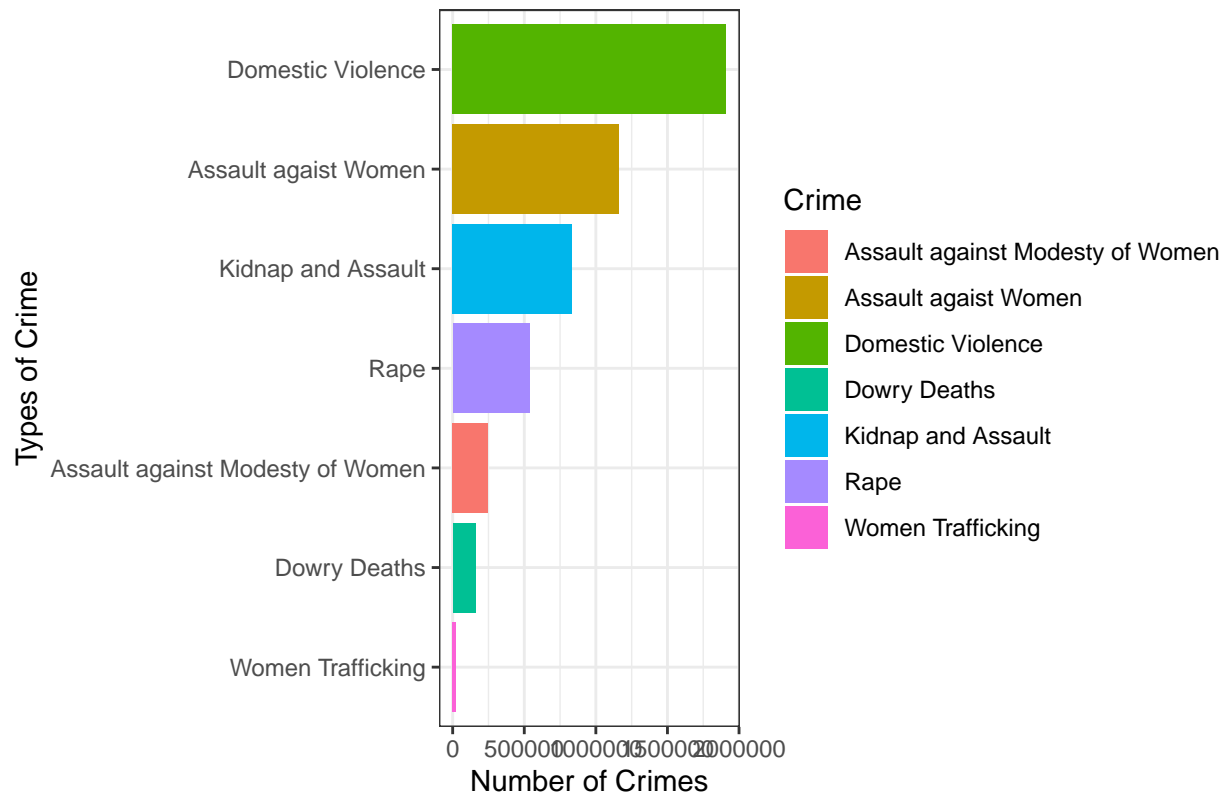
```
## Create a summary data frame with total crimes
```

```
crime_totals <- data.frame(
  Crime = c("Rape", "Kidnap and Assault", "Dowry Deaths", "Assault against Women", "Assault against Mod",
  Total = c(total_rape, total_K.A, total_DD, total_Aow, total_AoM, total_DV, total_WT))
```

```
## Display total number of various crimes in bar chart
```

```
crime_totals %>%
  ggplot(aes(fct_reorder(Crime, Total), Total, fill = Crime)) +
  geom_bar(stat = "identity") +
  theme_bw() +
  coord_flip()+
  labs(x = "Types of Crime", y = "Number of Crimes",
       title = "Total Number of Different Crimes in 20 years (2001 - 2021) in India")
```

## Total Number of Different Crimes in 20 years (2001 –



```
# Incident of Various crimes in years
```

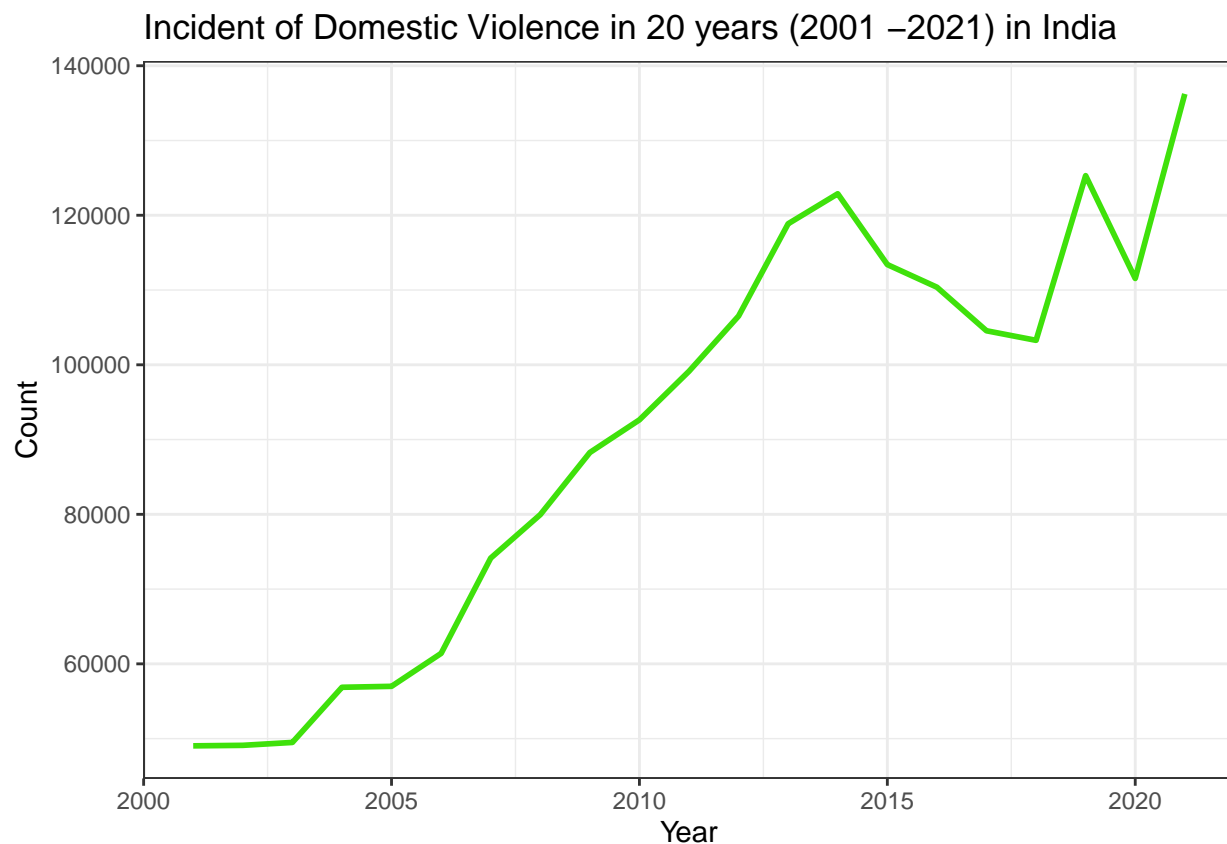
```
## Incident of Domestic Violence in years
```

```
DV_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_DV = sum(DV, na.rm = T)) %>%
  arrange(desc(total_DV))
print(DV_by_year)
```

```
## # A tibble: 21 x 2
##   Year total_DV
##   <int>   <int>
## 1  2021  136234
## 2  2019  125298
## 3  2014  122877
## 4  2013  118866
## 5  2015  113403
## 6  2020  111549
## 7  2016  110378
## 8  2012  106527
## 9  2017  104551
## 10 2018  103272
## # i 11 more rows
```

```
### Display incident of Domestic Violence in years
```

```
DV_by_year %>%
  ggplot(aes(Year, total_DV))+
  geom_smooth(stat = "identity", color = "#3fe10b") +
  labs(x = "Year", y = "Count",
        title = "Incident of Domestic Violence in 20 years (2001 -2021) in India")+
  theme_bw()
```



```
## Incident of Assault Against Women in years
```

```
AoW_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_AoW = sum(AoW, na.rm = T)) %>%
  arrange(desc(total_AoW))
print(AoW_by_year)
```

```
## # A tibble: 21 x 2
##   Year total_AoW
##   <int>     <int>
## 1  2021     89200
## 2  2018     89097
## 3  2019     88367
## 4  2017     86001
## 5  2020     85392
```

```
## 6 2016      84746
## 7 2015      82422
## 8 2014      82235
## 9 2013      70739
## 10 2012     45344
## # i 11 more rows
```

```
### Display average Assault Against Women in years
```

```
AoW_by_year %>%
  ggplot(aes(Year, total_AoW)) +
  geom_smooth(stat = "identity", color = "#a6ba26") +
  labs(x = "Year", y = "Count",
       title = "Incident of Assault Against Women in 20 years (2001 -2021) in India") +
  theme_bw()
```



```
## Incident of Kidnap and Assault
```

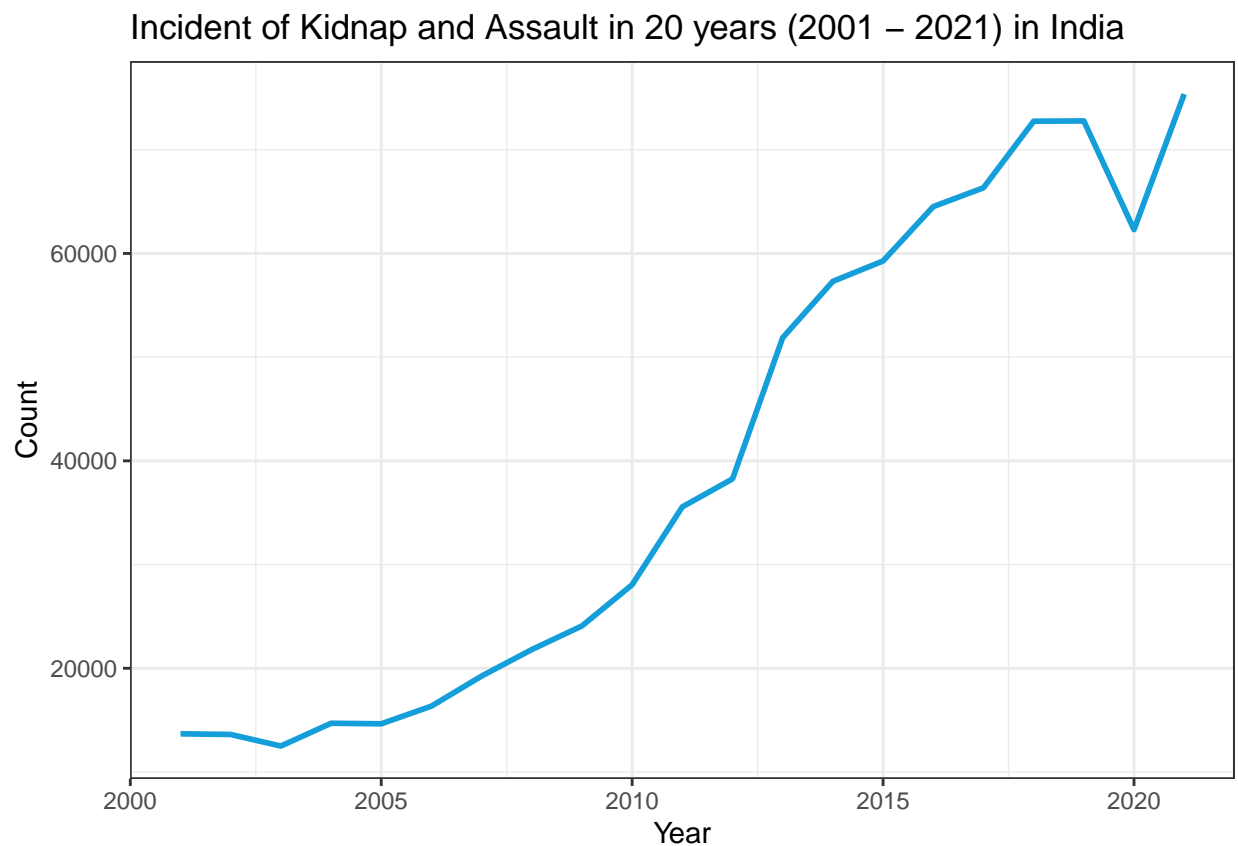
```
K.A_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_K.A = sum(K.A, na.rm = T)) %>%
  arrange(desc(total_K.A))
print(K.A_by_year)
```

```
## # A tibble: 21 x 2
```

```
##      Year total_K.A
##      <int>      <int>
## 1  2021      75369
## 2  2019      72780
## 3  2018      72751
## 4  2017      66333
## 5  2016      64519
## 6  2020      62300
## 7  2015      59277
## 8  2014      57311
## 9  2013      51881
## 10 2012      38262
## # i 11 more rows
```

```
### Display incident of kidnap and assault
```

```
K.A_by_year %>%
  ggplot(aes(Year, total_K.A)) +
  geom_smooth(stat = "identity", color = "#149eda") +
  labs(x = "Year", y = "Count",
       title = "Incident of Kidnap and Assault in 20 years (2001 - 2021) in India") +
  theme_bw()
```



```
## Incident of Rape in years
```



```

rapes_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_rape = sum(Rape, na.rm = T)) %>%
  arrange(desc(total_rape))
print(rapes_by_year)

```

```

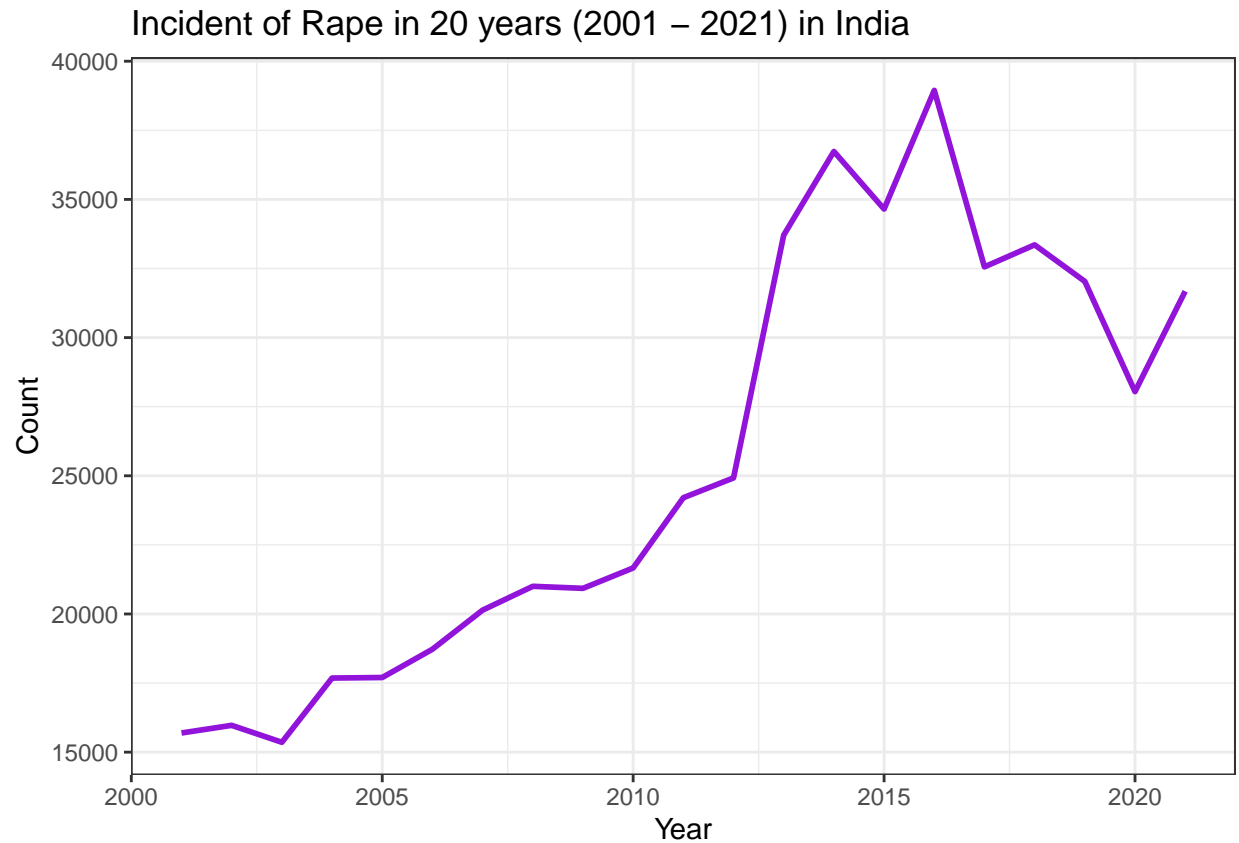
## # A tibble: 21 x 2
##   Year total_rape
##   <int>      <int>
## 1  2016      38947
## 2  2014      36735
## 3  2015      34651
## 4  2013      33707
## 5  2018      33356
## 6  2017      32559
## 7  2019      32033
## 8  2021      31677
## 9  2020      28046
## 10 2012      24923
## # i 11 more rows

```

```

### Display incident of Rape in years
rapes_by_year %>%
  ggplot(aes(Year, total_rape)) +
    geom_smooth(stat = "identity", color = "#9214da") +
    labs(x = "Year", y = "Count",
         title = "Incident of Rape in 20 years (2001 - 2021) in India") +
    theme_bw()

```



```
## Incident of Assault against modesty of women by year
```

```
AoM_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_AoM = sum(AoM, na.rm = T)) %>%
  arrange(desc(total_AoM))
print(AoM_by_year)
```

```
## # A tibble: 21 x 2
##   Year total_AoM
##   <int>     <int>
## 1  2016     27344
## 2  2015     24041
## 3  2014     21938
## 4  2013     12589
## 5  2003     12220
## 6  2008     12084
## 7  2009     10891
## 8  2007     10783
## 9  2002      10027
## 10 2010       9881
## # i 11 more rows
```

```
### Display incident of Assault against modesty of Women by year
AoM_by_year %>%
```

```
ggplot(aes(Year, total_AoM)) +
  geom_smooth(stat = "identity", color = '#d34c0c') +
  labs(x = "Year", y = "Count",
       title = "Incident of Assault against Modesty of Women in 20 years (2001 - 2021) in India") +
  theme_bw()
```



```
## Incident of Dowry Death by year
```

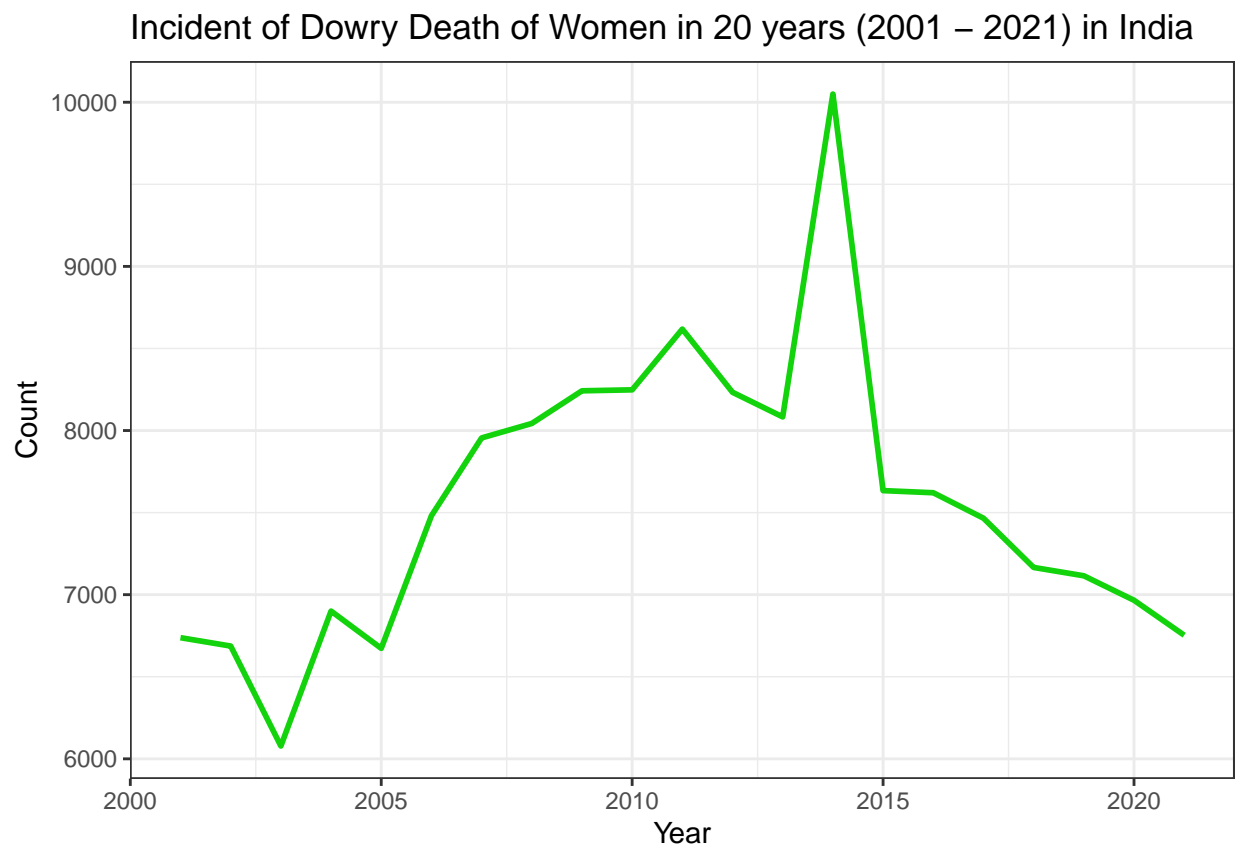
```
DD_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_DD = sum(DD, na.rm = T)) %>%
  arrange(desc(total_DD))
print(DD_by_year)
```

```
## # A tibble: 21 x 2
##   Year total_DD
##   <int>   <int>
## 1  2014    10050
## 2  2011     8618
## 3  2010     8248
## 4  2009     8242
## 5  2012     8233
## 6  2013     8083
## 7  2008     8043
## 8  2007     7955
```

```
## 9 2015 7634
## 10 2016 7621
## # i 11 more rows
```

```
### Display incident of Dowry Death by year
```

```
DD_by_year %>%
  group_by(Year) %>%
  ggplot(aes(Year, total_DD)) +
  geom_smooth(stat = "identity", color = "#12d30c") +
  labs(x = "Year", y = "Count",
       title = "Incident of Dowry Death of Women in 20 years (2001 - 2021) in India") +
  theme_bw()
```



```
## Incident of Women trafficking by year
```

```
WT_by_year <- crimes_on_women %>%
  group_by(Year) %>%
  summarize(total_WT = sum(WT, na.rm = T)) %>%
  arrange(desc(total_WT))
print(WT_by_year)
```

```
## # A tibble: 21 x 2
##   Year total_WT
##   <int>   <int>
## 1 2013    2579
```

```
## 2 2012 2563
## 3 2011 2435
## 4 2015 2424
## 5 2016 2214
## 6 2014 2070
## 7 2017 1536
## 8 2018 1459
## 9 2019 1185
## 10 2021 1071
## # i 11 more rows
```

```
### Display incident of Women Trafficking by year
```

```
WT_by_year %>%
  group_by(Year) %>%
  ggplot(aes(Year, total_WT)) +
  geom_smooth(stat = "identity", color = "#ee3ec6") +
  labs(x = "Year", y = "Count",
       title = "Incident of Women Trafficking in 20 years (2001 - 2021) in India") +
  theme_bw()
```



```
# Average number of crimes in different states
```

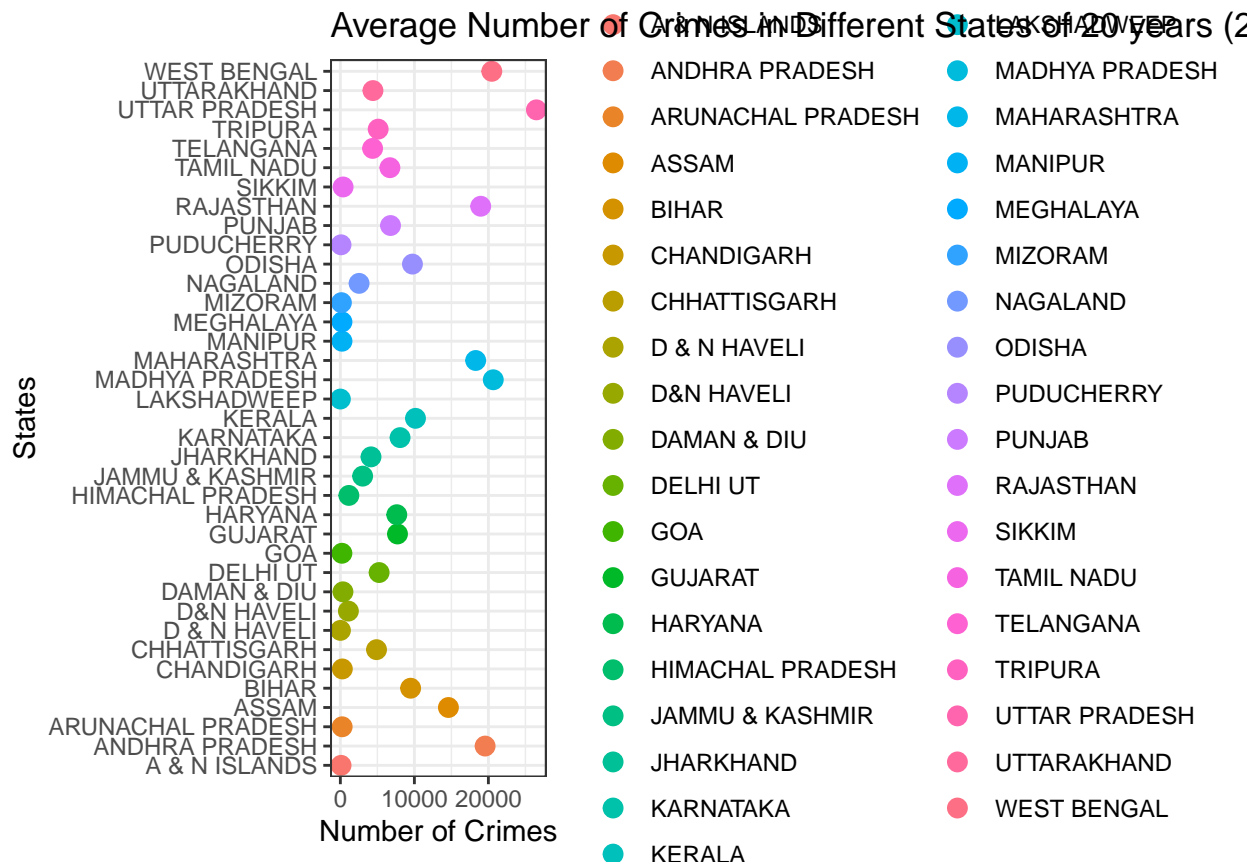
```
avg_crime_state <- crimes_on_women %>%
```

```
group_by(State) %>%
summarize(avg_crime = sum((Rape + K.A + DD + AoW + AoM + DV +WT)/20)) %>%
arrange(desc(avg_crime))

print(avg_crime_state)
```

```
## # A tibble: 37 x 2
##   State      avg_crime
##   <chr>      <dbl>
## 1 UTTAR PRADESH 26487.
## 2 MADHYA PRADESH 20658.
## 3 WEST BENGAL 20462.
## 4 ANDHRA PRADESH 19556.
## 5 RAJASTHAN 18963.
## 6 MAHARASHTRA 18282.
## 7 ASSAM 14599
## 8 KERALA 10149.
## 9 ODISHA 9739.
## 10 BIHAR 9494.
## # i 27 more rows
```

```
## Display total number of crimes by States in Dot plot
avg_crime_state %>%
  ggplot(aes(avg_crime, State))+
  geom_point(aes(colour = State), size = 3)+
  labs(x = "Number of Crimes", y = "States",
       title = "Average Number of Crimes in Different States of 20 years (2001-2021) in India")+
  theme_bw()
```



```
# Top 10 states with highest number of various crimes
```

```
## Top 10 states with highest number of domestic violence
```

```
DV_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_DV = sum(DV, na.rm = T)) %>%
  arrange(desc(total_DV))

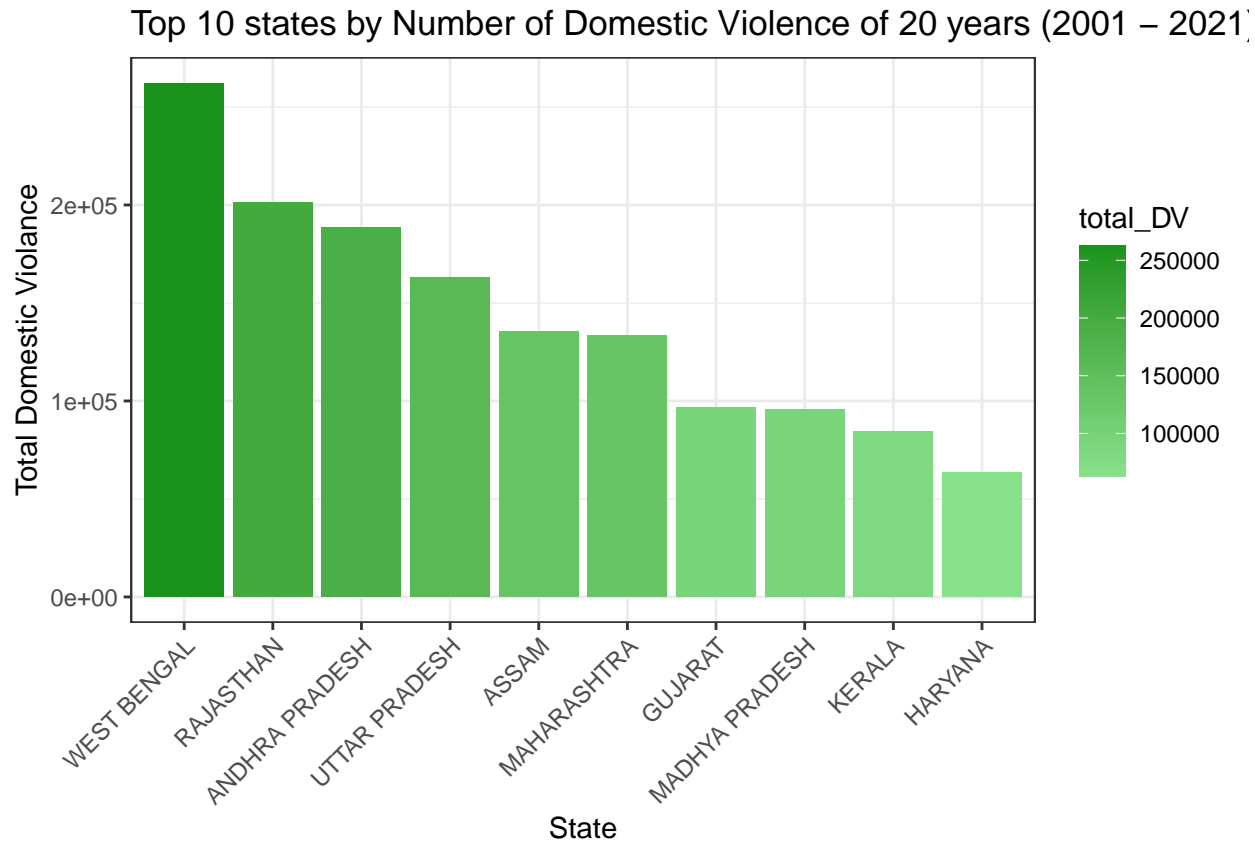
print(DV_by_state[1:10, ])
```

```
## # A tibble: 10 x 2
##   State      total_DV
##   <chr>      <int>
## 1 WEST BENGAL 262235
## 2 RAJASTHAN   201488
## 3 ANDHRA PRADESH 188511
## 4 UTTAR PRADESH 163062
## 5 ASSAM      135415
## 6 MAHARASHTRA 133357
## 7 GUJARAT     96480
## 8 MADHYA PRADESH 95521
## 9 KERALA      84277
## 10 HARYANA     63266
```

```

### Display top 10 states with highest number of domestic violence
DV_by_state[1:10, ] %>%
  ggplot(aes(reorder(State, -total_DV), total_DV, fill = total_DV)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "#88e18b", high = "#19931c") +
  theme_bw() +
  labs(x = "State", y = "Total Domestic Violence",
       title = "Top 10 states by Number of Domestic Violence of 20 years (2001 - 2021) in India") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



*## Top 10 states with highest number of Assault Against Women*

```

AoW_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_Aow = sum(AoW, na.rm = T)) %>%
  arrange(desc(total_Aow))

print(AoW_by_state[1:10, ])

```

```

## # A tibble: 10 x 2
##   State      total_Aow
##   <chr>      <int>
## 1 MADHYA PRADESH 152000
## 2 MAHARASHTRA   107519
## 3 UTTAR PRADESH  99650

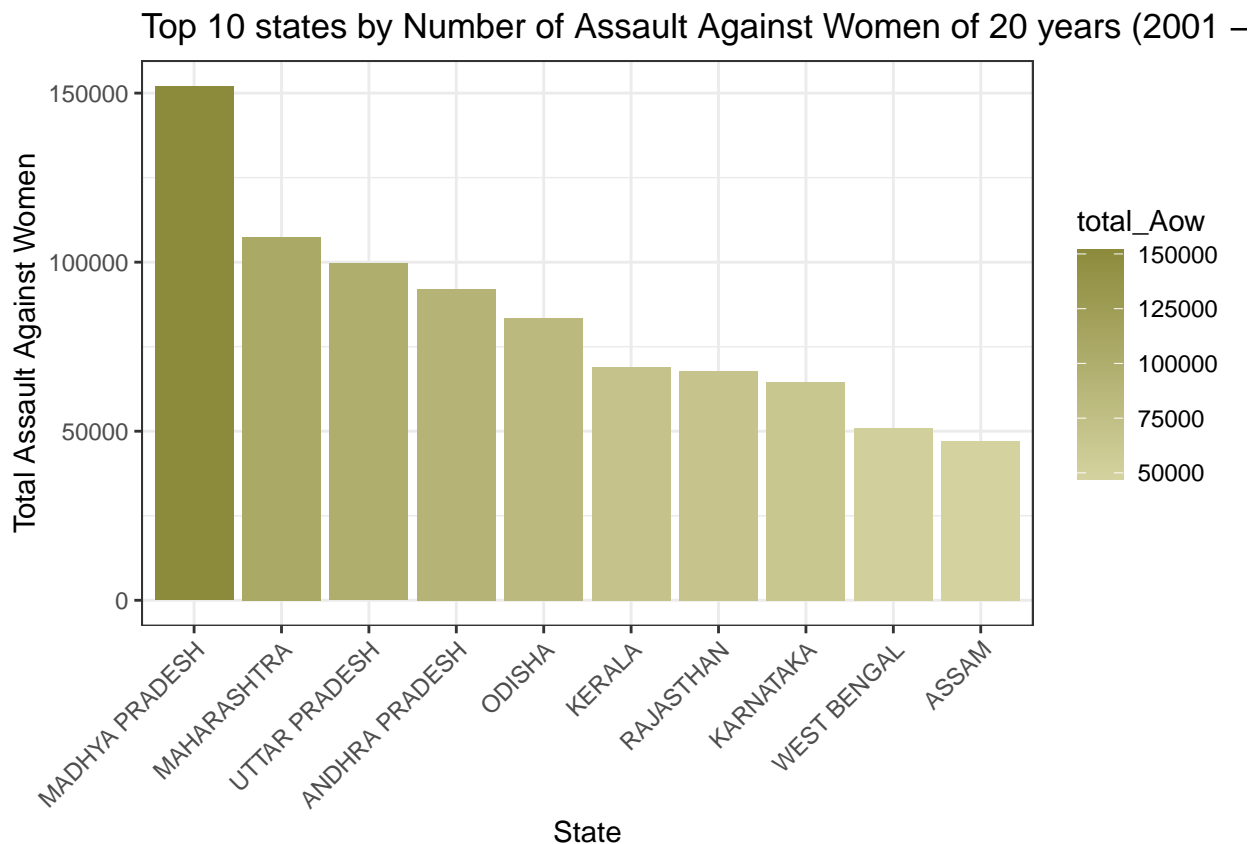
```



```
## 4 ANDHRA PRADESH      92084
## 5 ODISHA              83473
## 6 KERALA              69081
## 7 RAJASTHAN           67855
## 8 KARNATAKA           64492
## 9 WEST BENGAL         51014
## 10 ASSAM              47115
```

```
### Display top 10 states with highest number of Assault Against Women
```

```
AoW_by_state[1:10, ] %>%
  ggplot(aes(reorder(State, -total_Aow), total_Aow, fill = total_Aow)) +
  geom_bar(stat = "identity") +
  theme_bw() +
  scale_fill_gradient(low = "#d4d3a0", high = "#8c8b3c") +
  labs(x = "State", y = "Total Assault Against Women",
       title = "Top 10 states by Number of Assault Against Women of 20 years (2001 - 2021) in India",
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
## Top 10 states with highest number of Kidnap and Assault
```

```
K.A_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_K.A = sum(K.A, na.rm = T)) %>%
  arrange(desc(total_K.A))
```

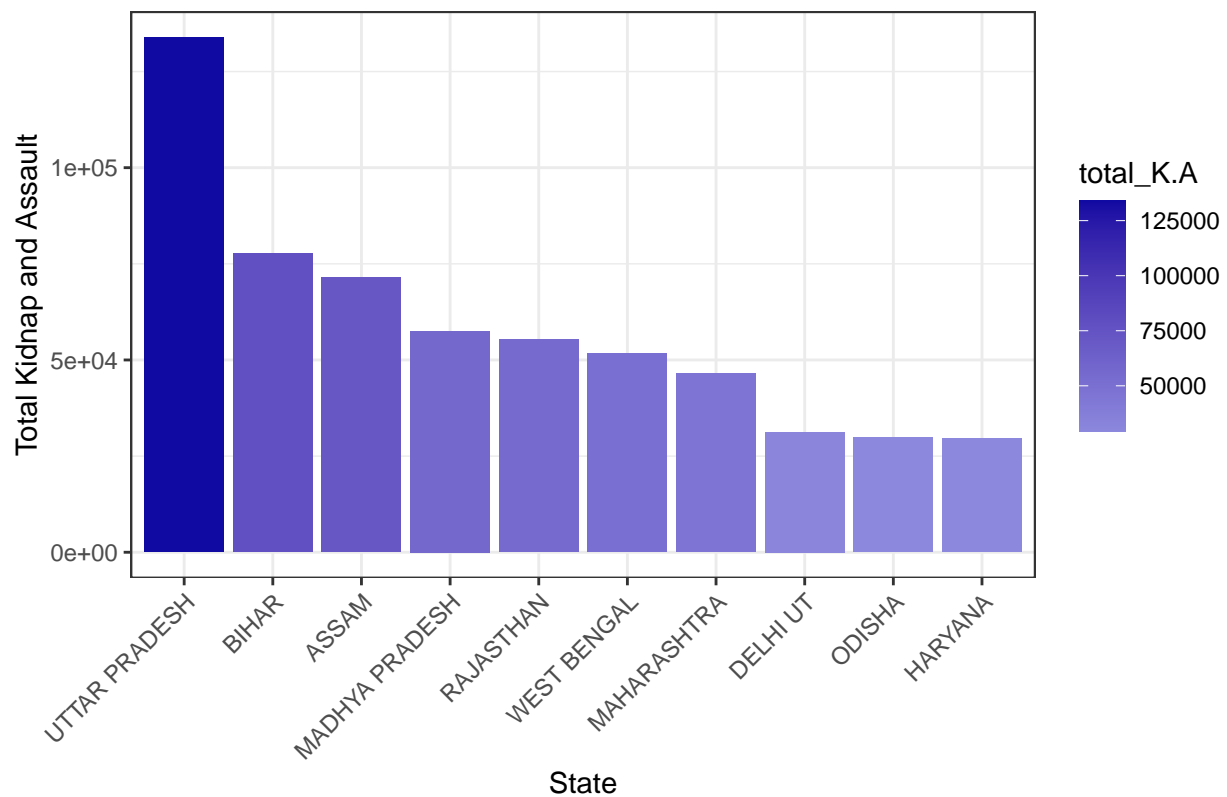
```
print(K.A_by_state[1:10, ])
```

```
## # A tibble: 10 x 2
##   State      total_K.A
##   <chr>      <int>
## 1 UTTAR PRADESH 133856
## 2 BIHAR        77687
## 3 ASSAM        71462
## 4 MADHYA PRADESH 57485
## 5 RAJASTHAN    55320
## 6 WEST BENGAL  51742
## 7 MAHARASHTRA  46522
## 8 DELHI UT     31251
## 9 ODISHA       29847
## 10 HARYANA     29593
```

```
## Display top 10 states with highest number of Kidnap and Assault
```

```
K.A_by_state[1:10, ] %>%
  ggplot(aes(reorder(State, -total_K.A), total_K.A, fill = total_K.A)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "#8c88dd", high = "#1009a2") +
  theme_bw() +
  labs(x = "State", y = "Total Kidnap and Assault",
       title = "Top 10 states by Number of Kidnap and Assault of 20 years (2001 - 2021) in India") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 10 states by Number of Kidnap and Assault of 20 years (2001 – 2021)



*## Top 10 states with highest number of rapes*

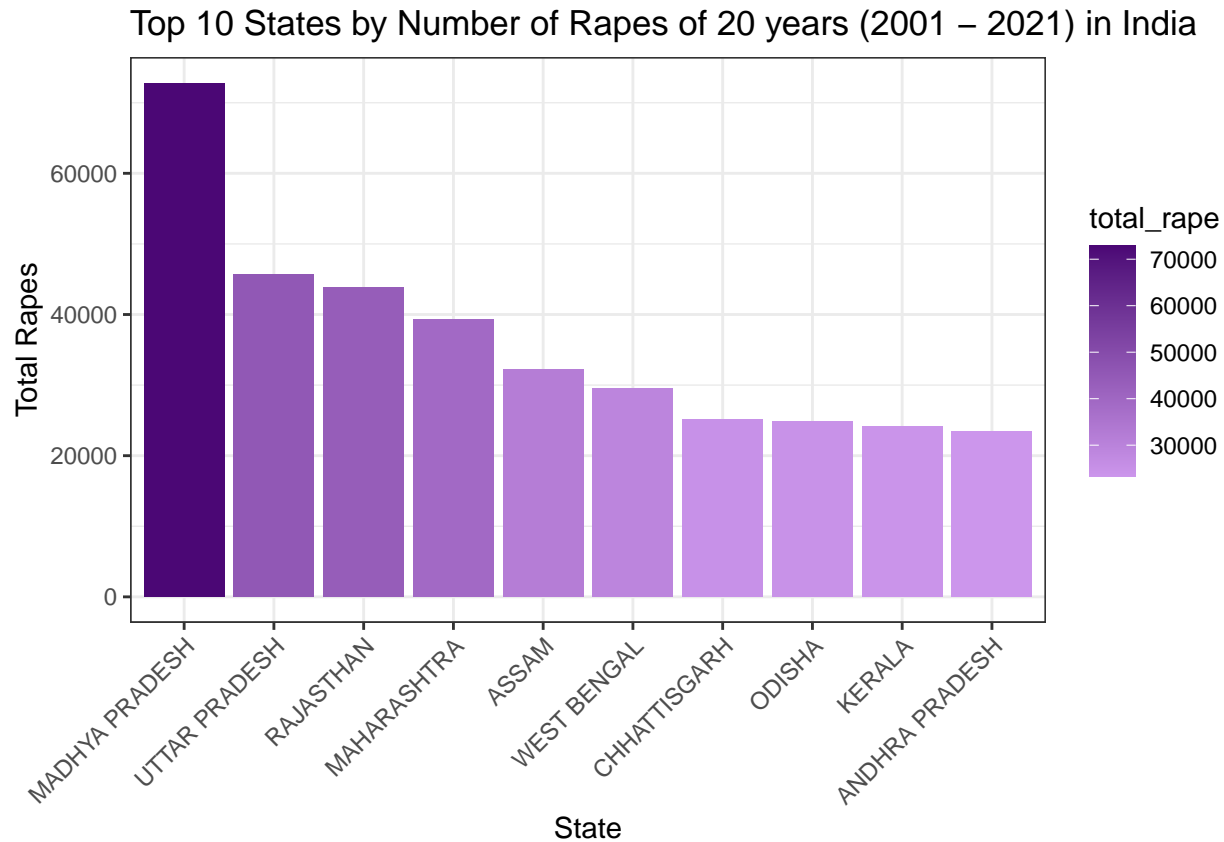
```
rape_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_rape = sum(Rape, na.rm = T)) %>%
  arrange(desc(total_rape))

print(rape_by_state[1:10, ])
```

```
## # A tibble: 10 x 2
##   State      total_rape
##   <chr>      <int>
## 1 MADHYA PRADESH    72808
## 2 UTTAR PRADESH    45694
## 3 RAJASTHAN        43799
## 4 MAHARASHTRA      39252
## 5 ASSAM            32190
## 6 WEST BENGAL      29486
## 7 CHHATTISGARH     25178
## 8 ODISHA           24810
## 9 KERALA           24184
## 10 ANDHRA PRADESH   23424
```

*### Display top 10 states with highest number of rape*

```
rape_by_state[1:10, ] %>%
  ggplot(aes(x = reorder(State, -total_rape), y = total_rape, fill = total_rape)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "#cc96ec", high = "#4b0775") +
  theme_bw() +
  labs(x = "State", y = "Total Rapes", title = "Top 10 States by Number of Rapes of 20 years (2001 - 2021) in India") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



*## Top 10 states with highest number of Assault against Modesty of Women*

```
AoM_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_AoM = sum(AoM, na.rm = T)) %>%
  arrange(desc(total_AoM))

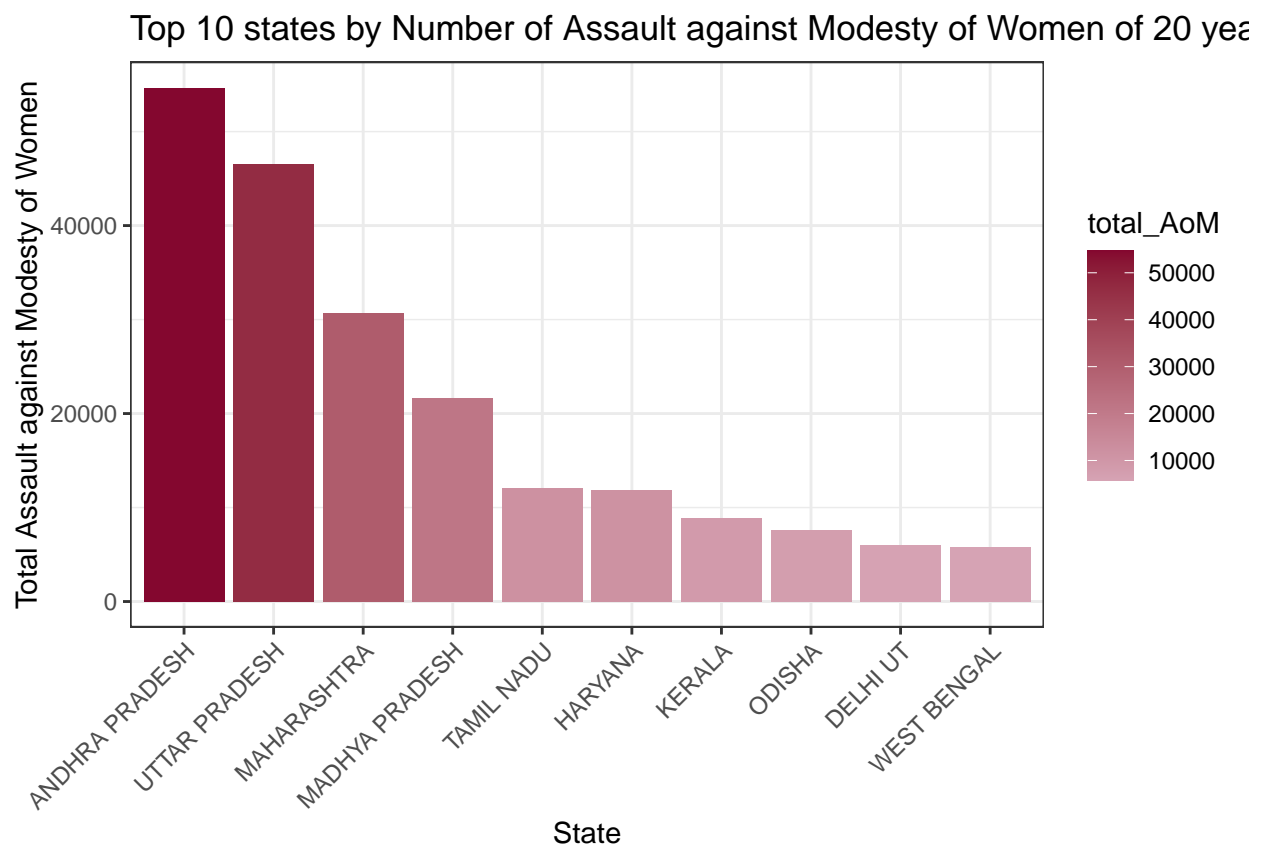
print(AoM_by_state[1:10, ])
```

```
## # A tibble: 10 x 2
##   State      total_AoM
##   <chr>      <int>
## 1 ANDHRA PRADESH    54684
## 2 UTTAR PRADESH    46524
## 3 MAHARASHTRA      30693
## 4 MADHYA PRADESH   21631
## 5 TAMIL NADU       12084
```

```
## 6 HARYANA 11834
## 7 KERALA 8873
## 8 ODISHA 7631
## 9 DELHI UT 6046
## 10 WEST BENGAL 5750
```

```
### Display top 10 states with highest number Assault against Modesty of Women
```

```
AoM_by_state[1:10, ] %>%
  ggplot(aes(reorder(State, -total_AoM), total_AoM, fill = total_AoM)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "#d6a3b4", high = "#84072f") +
  theme_bw() +
  labs(x = "State", y = "Total Assault against Modesty of Women",
       title = "Top 10 states by Number of Assault against Modesty of Women of 20 years (2001 - 2021)",
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
## Top 10 states with highest number of Dowry Death
```

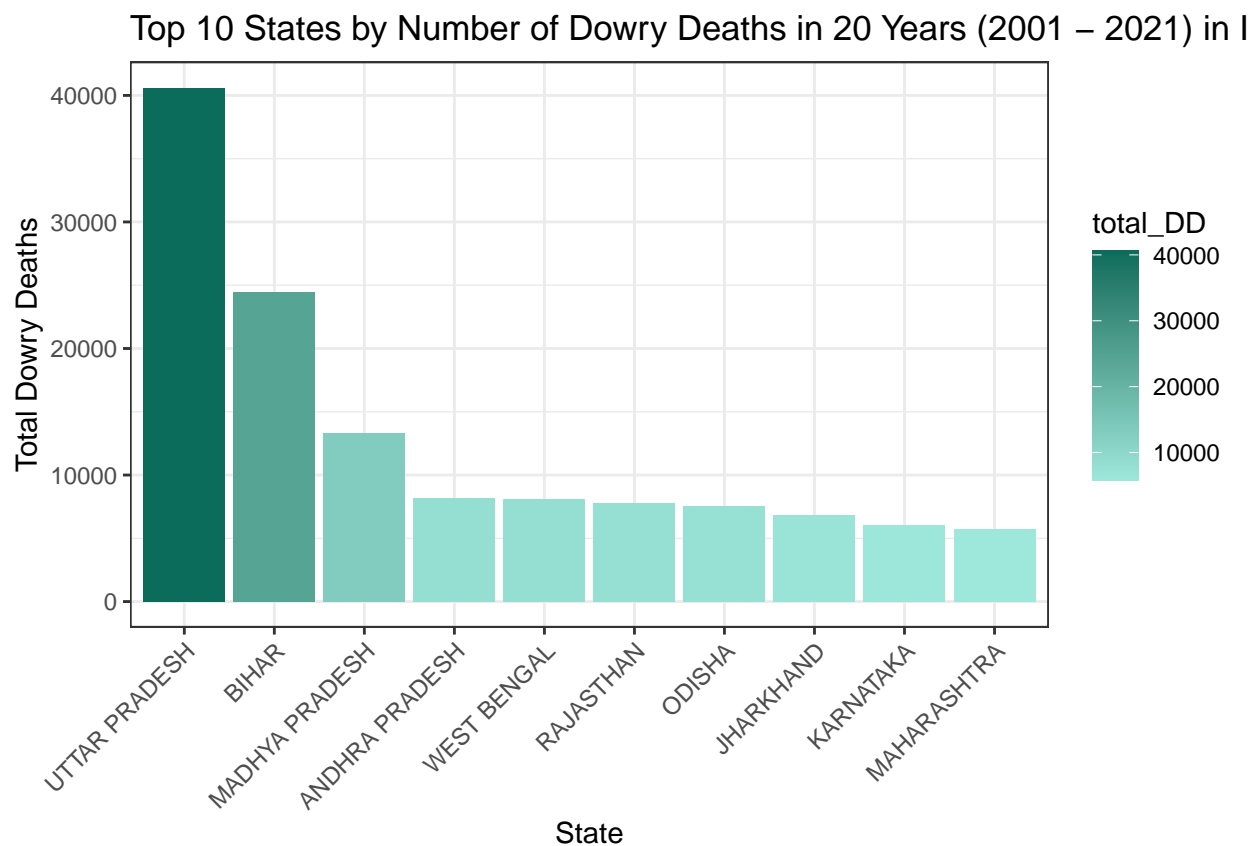
```
DD_by_state <- crimes_on_women %>%
  group_by(State) %>%
  summarize(total_DD = sum(DD, na.rm = T)) %>%
  arrange(desc(total_DD))

print(DD_by_state[1:10, ])
```

```
## # A tibble: 10 x 2
##   State      total_DD
##   <chr>      <int>
## 1 UTTAR PRADESH    40615
## 2 BIHAR            24428
## 3 MADHYA PRADESH  13316
## 4 ANDHRA PRADESH   8165
## 5 WEST BENGAL      8075
## 6 RAJASTHAN        7769
## 7 ODISHA           7521
## 8 JHARKHAND        6814
## 9 KARNATAKA        6028
## 10 MAHARASHTRA     5756
```

### Display top 10 states with highest number of Dowry Deaths

```
DD_by_state[1:10, ] %>%
  ggplot(aes(x = reorder(State, -total_DD), y = total_DD, fill = total_DD)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "#9ee8db", high = "#0b6c5b") + # Custom color gradient
  theme_bw() +
  labs(x = "State", y = "Total Dowry Deaths",
       title = "Top 10 States by Number of Dowry Deaths in 20 Years (2001 - 2021) in India") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
## Top 10 states with highest number of Women Trafficking
```

```
WT_by_state <- crimes_on_women %>%  
  group_by(State) %>%  
  summarize(total_WT = sum(WT, na.rm = T)) %>%  
  arrange(desc(total_WT))  
  
print(WT_by_state[1:10, ])
```

```
## # A tibble: 10 x 2  
##   State      total_WT  
##   <chr>      <int>  
## 1 TAMIL NADU      4041  
## 2 KARNATAKA       2799  
## 3 ANDHRA PRADESH  2548  
## 4 MAHARASHTRA    2533  
## 5 KERALA          990  
## 6 WEST BENGAL     940  
## 7 TELANGANA       866  
## 8 BIHAR           825  
## 9 RAJASTHAN       668  
## 10 SIKKIM         613
```

```
### Display top 10 states with highest number of Women Trafficking
```

```
WT_by_state[1:10, ] %>%  
  ggplot(aes(reorder(State, -total_WT), total_WT, fill = total_WT)) +  
  geom_bar(stat = "identity") +  
  scale_fill_gradient(low = "#e3a2de", high = "#780870") +  
  theme_bw() +  
  labs(x = "State", y = "Number of Women Trafficking",  
       title = "Top 10 States by Number of Women Trafficking in 20 Years (2001 - 2021) in India") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 10 States by Number of Women Trafficking in 20 Years (2001 – 2021)

