R-Language

INTRODUCTION TO R :

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

THE R ENVIRONMENT:

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,

- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term “environment” is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

R, like S, is designed around a true computer language, and it allows users to add additional functionality by defining new functions. Much of the system is itself written in the R dialect of S, which makes it easy for users to follow the algorithmic choices made. For computationally-intensive tasks, C, C++ and Fortran code can be linked and called at run time. Advanced users can write C code to manipulate R objects directly.

Many users think of R as a statistics system. We prefer to think of it as an environment within which statistical techniques are implemented. R can be extended (easily) via *packages*. There are about eight packages supplied with the R distribution and many more are available through the CRAN family of Internet sites covering a very wide range of modern statistics.

R has its own LaTeX-like documentation format, which is used to supply comprehensive documentation, both on-line in a number of formats and in hardcopy.

ABOUT DATASET :

CONTENT-

The dataset contains state/UT wise data about various crimes in India. It includes three folders:

- Cases against Police Personnels
- Escapes from Policy Custody
- Victims of Rape

Each folder consists of 3 csv files corresponding to each year from 2016-2018. Further, each csv file contains state/UT wise data about each of the crime based on various factors.

1) Cases against Police Personnels
It contains data about number of cases against police personnels for violation of various rules, number of the cases registered or ended, number of police personnels convicted, etc.

2) Escapes from Police Custody
It contains data about number of escapees that escaped from police custody or outside police custody, number of escapees that were re-arrested, etc.

3) Victims of Rape
It contains data about number of victims of rape - Child and Women which is further divided into age groups of below 18 or above 18.

Acknowledgements

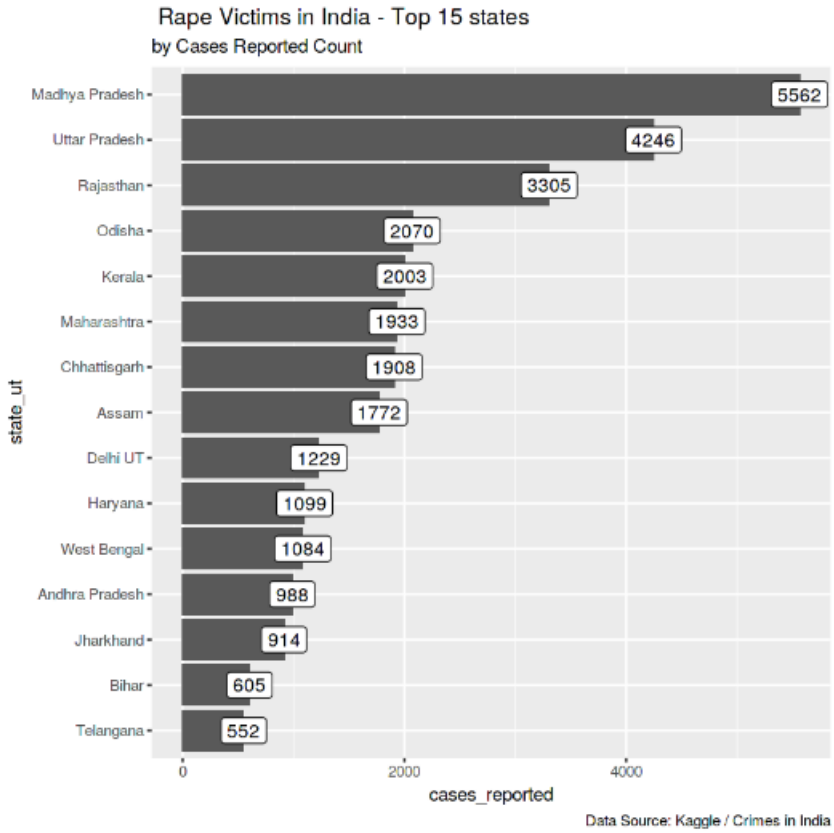
The data is collected from Open Goverment Data (OGD).

s_no	category	state_ut	cases_reported	child_victims_of_rape_below_18_yrs_below_6_years	child_victims_of_rape_below_18_yrs_6_years_above_
<dbl>	<chr>	<chr>	<dbl>	<dbl>	
1	State	Andhra Pradesh	988	14	
2	State	Arunachal Pradesh	59	1	
3	State	Assam	1772	2	
4	State	Bihar	605	0	
5	State	Chhattisgarh	1908	47	
6	State	Goa	76	4	

ANALYSIS

First we cleaned column name using clean_names() function of janitor library.Then we proceeded with our first analysis, top 15 states/UT with rape victims.

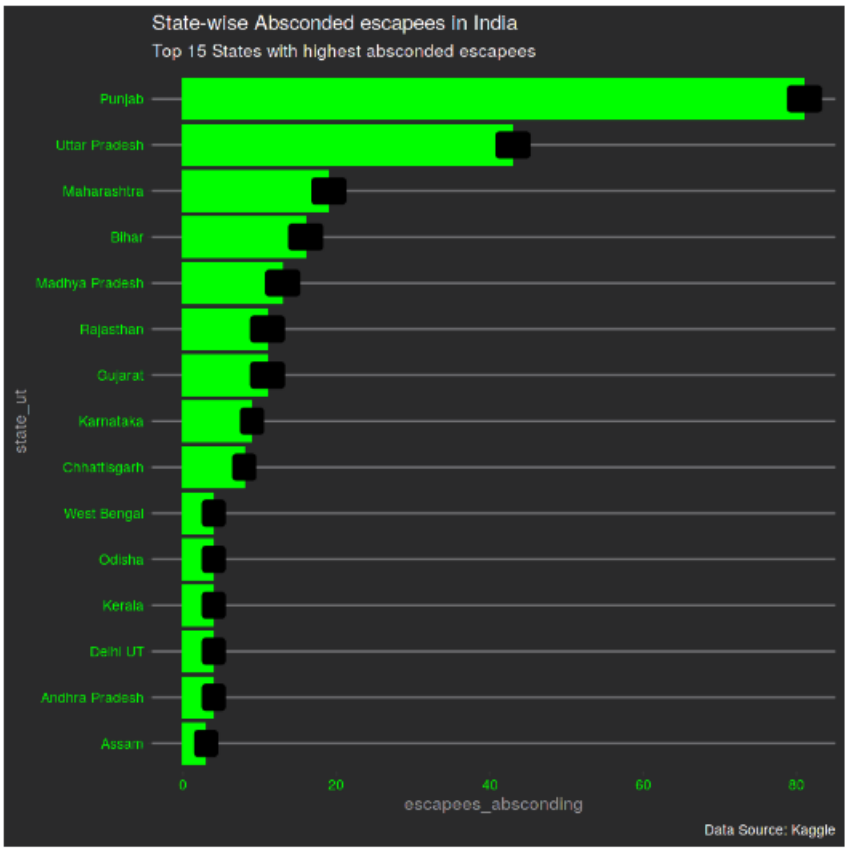
```r
crimeindia %>%
    select(one_of('state_ut','cases_reported')) %>%
    arrange(desc(cases_reported)) %>%
    head(15) %>%
    mutate(state_ut = fct_reorder(state_ut,cases_reported)) %>%
    ggplot() + geom_col(aes(y = state_ut,x =cases_reported)) +
    geom_label(aes(y = state_ut,x =cases_reported, label = cases_reported))+
    labs(title = ' Rape Victims in India - Top 15 states',
        subtitle = 'by Cases Reported Count',
        caption = 'Data Source: Kaggle / Crimes in India')
```



Then we analysed the top 15 states with highest absconded escapees in India

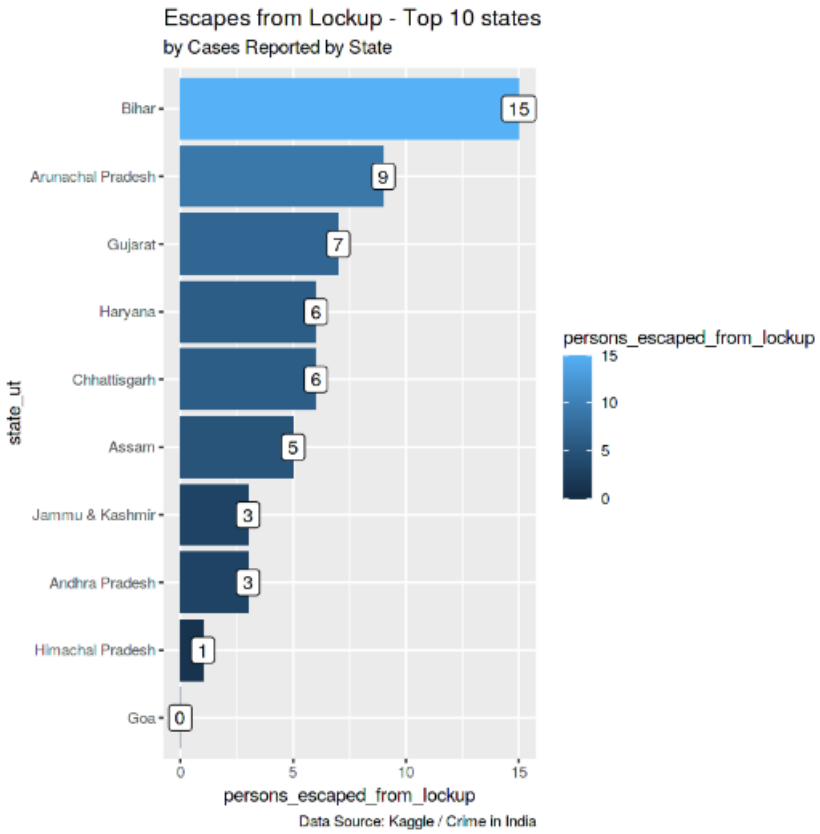
```
policecustodyescape %>%
    select(one_of('state_ut','escapees_absconding')) %>%
    arrange(desc(escapees_absconding)) %>%
    head(15) %>%
    mutate(state_ut = fct_reorder(state_ut,escapees_absconding)) %>%
    ggplot() + geom_col(aes(y = state_ut, x = escapees_absconding), fill = 'green') +
    geom_label(aes(y = state_ut, x = escapees_absconding,label = escapees_absconding), fill = 'black')

    labs(title = 'State-wise Absconded escapees in India',
        subtitle = 'Top 15 States with highest absconded escapees',
        caption = 'Data Source: Kaggle') +
ggthemes::theme_hc(
  base_size = 10,
  base_family = "sans",
  style = c("darkunica")
) +
theme(axis.text.x=element_text(colour="green"),
     axis.text.y=element_text(colour="green"))
```



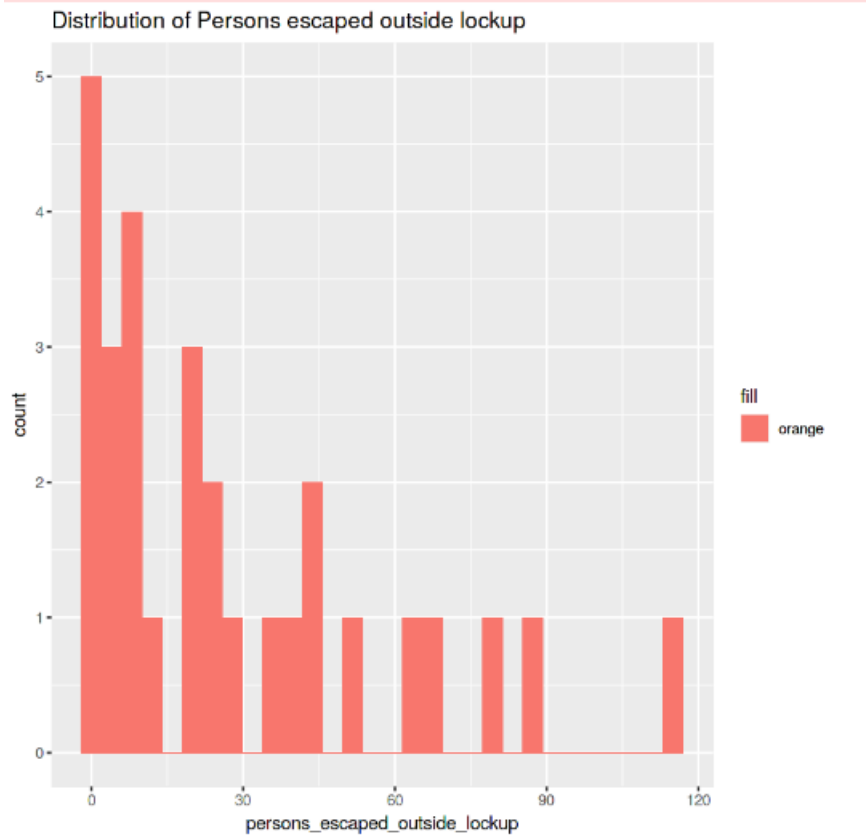
State-wise escapes from lockup in India:

```
policecustodyescape %>%
    select(one_of('state_ut','persons_escaped_from_lockup')) %>%
    arrange(desc('persons_escaped_from_lockup')) %>%
    head(10) %>%
    mutate(state_ut = fct_reorder(state_ut,persons_escaped_from_lockup )) %>%
    ggplot() + geom_col(aes(y = state_ut,x = persons_escaped_from_lockup , fill= persons_escaped_from_
    geom_label(aes(y = state_ut,x = persons_escaped_from_lockup , label = persons_escaped_from_lockup
    labs(title = 'Escapes from Lockup - Top 10 states',
        subtitle = 'by Cases Reported by State',
         caption = 'Data Source: Kaggle / Crime in India')
```



State-wise persons escaped outside lockup in india:

```
policecustodyescape %>%
filter(category == 'State') %>%
ggplot () + geom_histogram (aes(x=persons_escaped_outside_lockup, fill= 'orange'))+
labs(title='Distribution of Persons escaped outside lockup')
```


Distribution of Persons escaped outside lockup
count
0
1
2
3
4
5
0
30
60
90
120
persons_escaped_outside_lockup
fill
orange