

State/UT-wise Details of Crime against Women

Statistical Analysis Report

R-Script

Name:-

Muhibullah Noorzad (69)

Asim Hussain (14)

Deepak Kumar Dev (20)

Course Name:

Master of Computer Application (MCA)

Subject Name:

Mathematical Techniques for computer Applications

Submitted to:-

Dr. Vasudha Bhatnagar

Department of computer science

University of Delhi

Delhi-110007, India

```
#install.packages('dplyr')
```

```
#install.packages('ggplot2')
```

```
df<-read.csv("crimes_against_women_2001-2014.csv",sep="," ,header=TRUE,)
```

```
head(df,3)
```

```
#2.Clean and prepare the data dealing with missing values and zero value columns
```

```
#removing the null value
```

```
data<-na.omit(data)
```

```
head(df,3)
```

```
#print the number of row and columns
```

```
cat("Number of columns : ",ncol(df))
```

```
cat("Number of rows : ",nrow(df))
```

```
# for print the minimum value of the data kidnapping columns
```

```
cat("Minimum type crime Reported : ",min(df$Kidnapping.and.Abduction))
```

```
# for print the maximum value of the data kidnapping columns
```

```
cat("Maximum type crime Reported ",max(df$Kidnapping.and.Abduction))
```

```
#for print the the number mean kidnapping columns  
cat("Mean : ",mean(df$Kidnapping.and.Abduction))
```

```
#for print the the number median kidnapping columns  
cat("Median : ",median(df$Kidnapping.and.Abduction))
```

```
#for print the quantile 0.25  
quantile(df$Kidnapping.and.Abduction, 0.25)
```

```
#for print the quantile 0.50  
quantile(df$Kidnapping.and.Abduction, 0.50)
```

```
#for print the quantile 0.75  
quantile(df$Kidnapping.and.Abduction, 0.75)
```

```
#for print the fivenum  
fivenum(df$Dowry.Deaths)
```

```
#for printing the number of row and columns
```

```
dim(df)
```

```
#for print the summary of the data
```

```
summary(df)
```

```
#import the dplyr library
```

```
library(dplyr)
```

```
#Data cleaning the remove the extra columns
```

```
df=select(df, -X0)
```

```
df=select(df, -DISTRICT)
```

```
#for print the columns name of our dataset
```

```
colnames(df)
```

```
#find the total number crime from each state
```

```
state = df %>% group_by(df$STATE.UT) %>% summarise(
```

```
  total_Rape = sum(Rape),
```

```
  total_kinapping= sum(df$Kidnapping.and.Abduction),
```

```
  total_dowry = sum(df$Dowry.Deaths),
```

```
  total_assault = sum(df$Assault.on.women.with.intent.to.outrage.her.modesty),
```

```
  total_insult = sum(df$Insult.to.modesty.of.Women),
```

```
  total_cruelty = sum(df$Cruelty.by.Husband.or.his.Relatives),
```

```
  total_importation = sum(df$Importation.of.Girls),
```

```
  .groups = 'drop')
```

state

#importing the ggplot library

library(ggplot2)

#drawing the line plot the every

ggplot(data=df,aes(x=Year,y=df\$Dowry.Deaths,color=STATE.UT))+geom_line()+

ggtitle("Dowry Deaths of each state ")+

xlab("Year ")+

ylab("Dowry Deaths ")+

theme_classic()

#fine the value that less then mean

df %>% filter(df\$Kidnapping.and.Abduction<mean(df\$Kidnapping.and.Abduction))->gmddf

gmddf

#we chose these state for comparing the best and worse state

dh<-subset(df,df\$STATE.UT=="HIMACHAL PRADESH" | df\$STATE.UT=="KERALA" |
df\$STATE.UT=="RAJASTHAN" | df\$STATE.UT=="ASSAM" | df\$STATE.UT=="GUJARAT" |
df\$STATE.UT=="UTTAR PRADESH" | df\$STATE.UT=="BIHAR")

#box plot of best and worst state

```

ggplot(data=dh ,aes(x=dh$STATE.UT,y=Dowry.Deaths,fill=STATE.UT))+geom_boxplot()+
  ggtitle("Dowry Deaths  of each state ")+
  xlab(" State ")+
  ylab("value ")+
  theme_classic()
dh %>% filter(dh$Dowry.Deaths<100)->dh

```

#boxplot of best and worst state

```

ggplot(data=dh ,aes(x=dh$STATE.UT,y=Dowry.Deaths,fill=STATE.UT))+geom_boxplot()+
  ggtitle("Dowry Deaths  of each state ")+
  xlab(" State ")+
  ylab("value ")+
  theme_classic()

```

#to find the total number crime state wise

```

qs1<-sum(df$Rape[df$STATE.UT=="Delhi UT"])
qs2<-sum(df$Rape[df$STATE.UT=="BIHAR"])
qs3<-sum(df$Rape[df$STATE.UT=="UTTAR PRADESH"])
qs4<-sum(df$Rape[df$STATE.UT=="GUJARAT"])
qs5<-sum(df$Rape[df$STATE.UT=="JHARKHAND"])
pies<-c(qs1,qs2,qs3,qs4,qs5)
names(pies)<-c("Delhi UT","BIHAR","UTTAR PRADESH","GUJARAT","JHARKHAND")
lb=c("Delhi UT","BIHAR","UTTAR PRADESH","GUJARAT","JHARKHAND")
#install.packages("plotrix")
#importing the plotrix
library(plotrix)
#drawing pie for total number of crime

```

```
pie3D(pies,labels = lb, explode = 0.2, theta = 1.5)
```

```
install.packages('psych')
```

```
#delete the outlier of data
```

```
df %>% filter(df$Kidnapping.and.Abduction<200)->df
```

```
#drawing the histogram of the data
```

```
hist(df$Kidnapping.and.Abduction, col = 'blue', border = "green")
```

```
#printing a histogram for kidnapping and Rape
```

```
ggplot(data=dh ,aes(x=dh$Kidnapping.and.Abduction,y=dh$Rape,color=STATE.UT))+geom_point()+
```

```
  ggtitle("Correlation of kidnapping and rape ")+
```

```
  xlab(" Rape ")+
```

```
  ylab(" Kidnapping ")+
```

```
  theme_classic()
```

```
#histogram of Dowry Death and
```

```
ggplot(data=dh,aes(x=dh$Dowry.Deaths))+geom_histogram(fill='green',col='orange')+
```

```
  ggtitle("histogram of Dowry Deaths from all data ")+
```

```
  xlab("Dowry Death ")+
```

```
  ylab("Count")+
```

```
  theme_classic()
```

```
#plot the total number of rape case in for all state
```

```
plot(state$total_Rape, type='s')
```

```
# for find the the maxximum rage case of each state
```

```
statemax = df %>% group_by(df$STATE.UT) %>% summarise(
```

```
  max_Rape = max(Rape),
```

```
  max_kinapping= max(df$Kidnapping.and.Abduction),
```

```
  max_dowry = max(df$Dowry.Deaths),
```

```
  max_assault = max(df$Assault.on.women.with.intent.to.outrage.her.modesty),
```

```
  max_insult = max(df$Insult.to.modesty.of.Women),
```

```
  max_cruelty = max(df$Cruelty.by.Husband.or.his.Relatives),
```

```
  max_importation = max(df$Importation.of.Girls),
```

```
  .groups = 'drop')
```

```
statemax
```

```
#bar for the maximum Repe of all state
```

```
barplot(statemax$max_Rape,names.arg= statemax$`df$STATE.UT`,ylab="Rape",col="blue",
```

```
  main="Maximum Repe case of each state ",border="red")
```

```
install.packages("GGally")
```

```
library(GGally)
```

```
#print the correlation of the columns
```

```
ggpairs(dh
```