Data Analytics

Assignment - 1

Name: - Ashish
Roll no.: - 17BCS006

CODE:

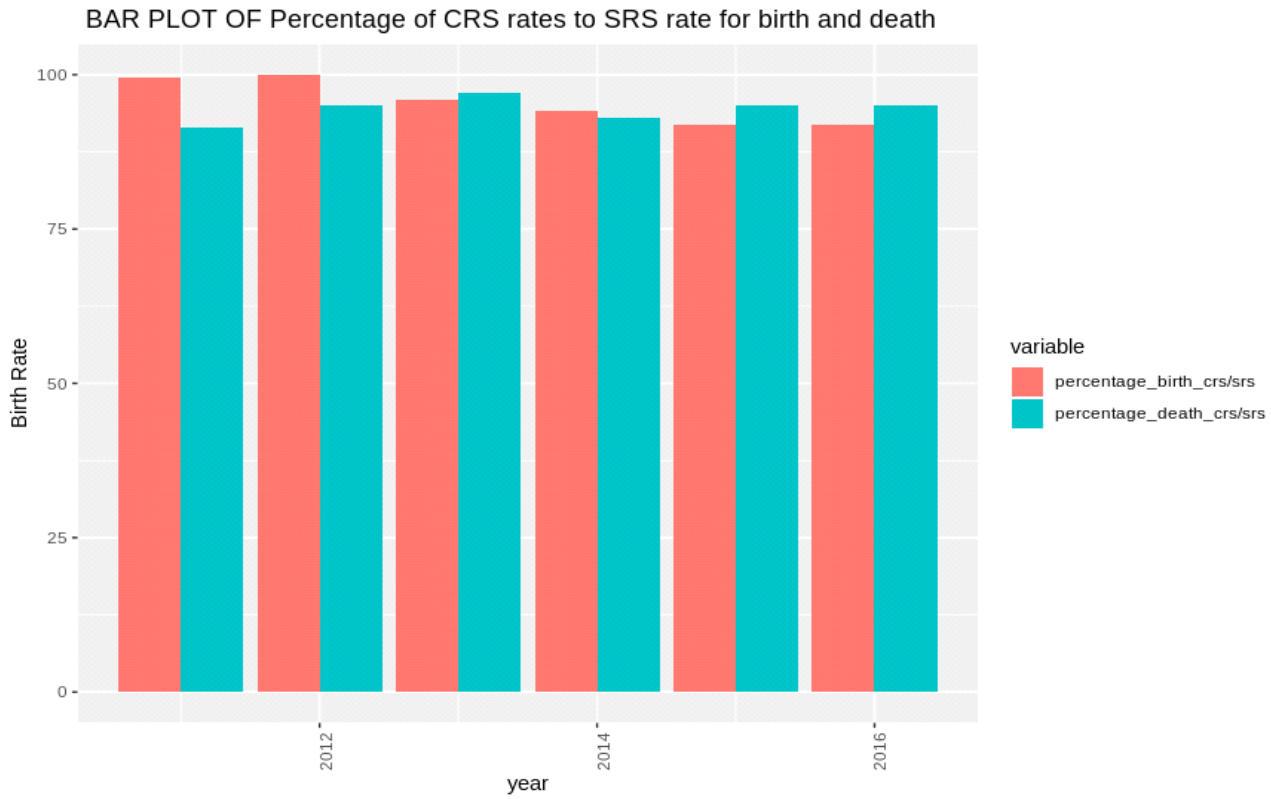
```
data <- read.table("/home/Ashish/DA_lab/lab1/data.csv", sep = ",", header = TRUE)

# Find Min
min(data$Live.Births)
min(data$X.Still.Births)
min(data$Death)
min(data$Birth.rate)
min(data$Death.rate)
min(data$Birth..)
min(data$Death..)

# Find Max
max(data$Live.Births)
max(data$X.Still.Births)
max(data$Death)
max(data$Birth.rate)
max(data$Death.rate)
max(data$Birth..)
max(data$Death..)

# Find Mean.
mean(data$Live.Births)
mean(data$X.Still.Births)
mean(data$Death)
mean(data$Birth.rate)
mean(data$Death.rate)
mean(data$Birth..)
mean(data$Death..)

#Find Median
median(data$Live.Births)
median(data$X.Still.Births)
median(data$Death)
median(data$Birth.rate)
median(data$Death.rate)
median(data$Birth..)
median(data$Death..)

#Find Mode
mode(data$Live.Births)
mode(data$X.Still.Births)
mode(data$Death)
mode(data$Birth.rate)
mode(data$Death.rate)
mode(data$Birth..)
mode(data$Death..)

#Find Variance
```

```
var(data$Live.Births)
var(data$X.Still.Births)
var(data$Death)
var(data$Birth.rate)
var(data$Death.rate)
var(data$Birth..)
var(data$Death..)

#Find Standard Deviation
sd(data$Live.Births)
sd(data$X.Still.Births)
sd(data$Death)
sd(data$Birth.rate)
sd(data$Death.rate)
sd(data$Birth..)
sd(data$Death..)

#Find IQR
IQR(data$Live.Births)
IQR(data$X.Still.Births)
IQR(data$Death)
IQR(data$Birth.rate)
IQR(data$Death.rate)
IQR(data$Birth..)
IQR(data$Death..)

#Detecting outliers in data
#data <- read.table("data.csv", sep = ",", header = TRUE)
outlierKD <- function(dt, var) {
  var_name <- eval(substitute(var),eval(dt))
  tot <- sum(!is.na(var_name))
  na1 <- sum(is.na(var_name))
  m1 <- mean(var_name, na.rm = T)
  par(mfrow=c(2, 2), oma=c(0,0,3,0))
  #dev.new(width=5, height=4, unit="in")
  boxplot(var_name, main="With outliers")
  hist(var_name, main="With outliers", xlab=NA, ylab=NA)
  outlier <- boxplot.stats(var_name)$out
  mo <- mean(outlier)
  var_name <- ifelse(var_name %in% outlier, NA, var_name)
  boxplot(var_name, main="Without outliers")
  hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
  title("Outlier Check", outer=TRUE)
  na2 <- sum(is.na(var_name))
  message("Outliers identified: ", na2 - na1, " from ", tot, " observations")
  message("Proportion (%) of outliers: ", (na2 - na1) / tot*100)
  message("Mean of the outliers: ", mo)
  m2 <- mean(var_name, na.rm = T)
  message("Mean without removing outliers: ", m1)
  message("Mean if we remove outliers: ", m2)
  response <- readline(prompt="Do you want to remove outliers and to replace with NA? [yes/no]: ")
  if(response == "y" | response == "yes"){
    dt[as.character(substitute(var))] <- invisible(var_name)
    assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)
    message("Outliers successfully removed", "\n")
    return(invisible(dt))
  } else{
    message("Nothing changed", "\n")
    return(invisible(var_name))
  }
```

```
}
outlierKD(data,data$Live.Births)
outlierKD(data,data$X.Still.Births)
outlierKD(data,data$Death)
outlierKD(data,data$Birth.rate)
outlierKD(data,data$Death.rate)
outlierKD(data,data$Birth..)
outlierKD(data,data$Death..)
```

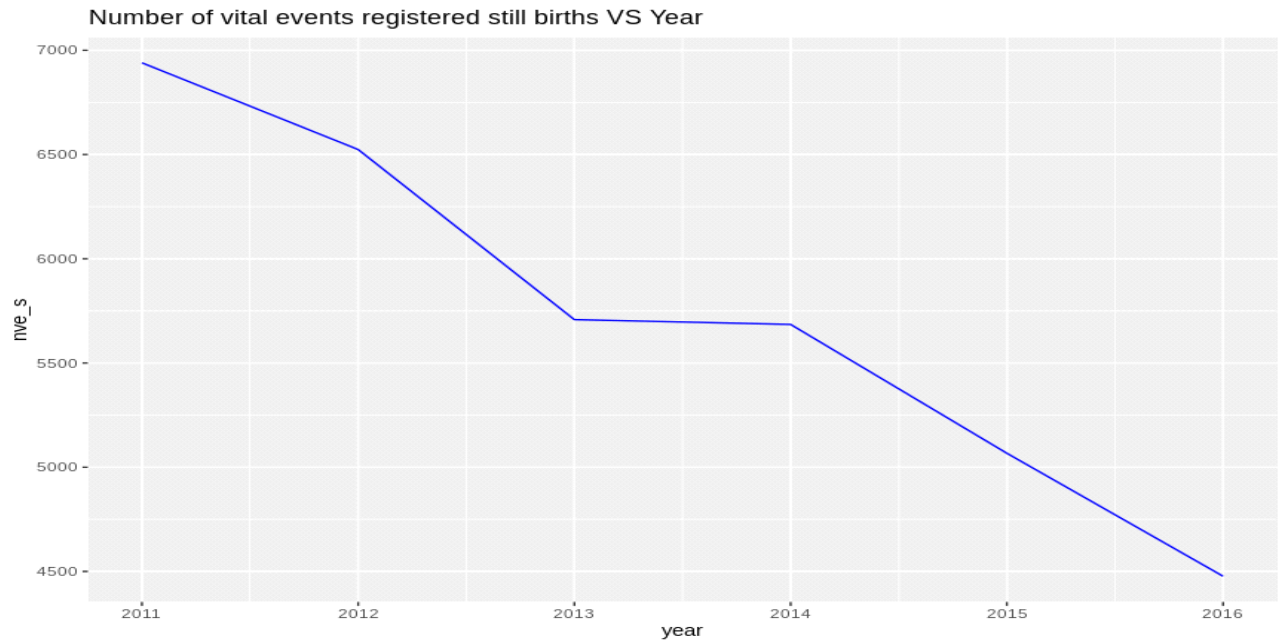
BAR PLOT OF Percentage of CRS rates to SRS rate for birth and death

```
mg <- data.frame(data$year,data$per_colnames(mg)<-c("year","percentage_birth_
mg <- melt(mg,id.vars = "year") ggplot(mg,aes(x = year , y = value,fill = variable))+ylab("Birth Rate") +
geom_bar(stat = "identity",position = "dodge") + theme(axis.text.x = element_text(angle = 90, hjust =
1))+ggtitle (" BAR PLOT OF Percentage of CRS rates to SRS rate for birth and death")
```



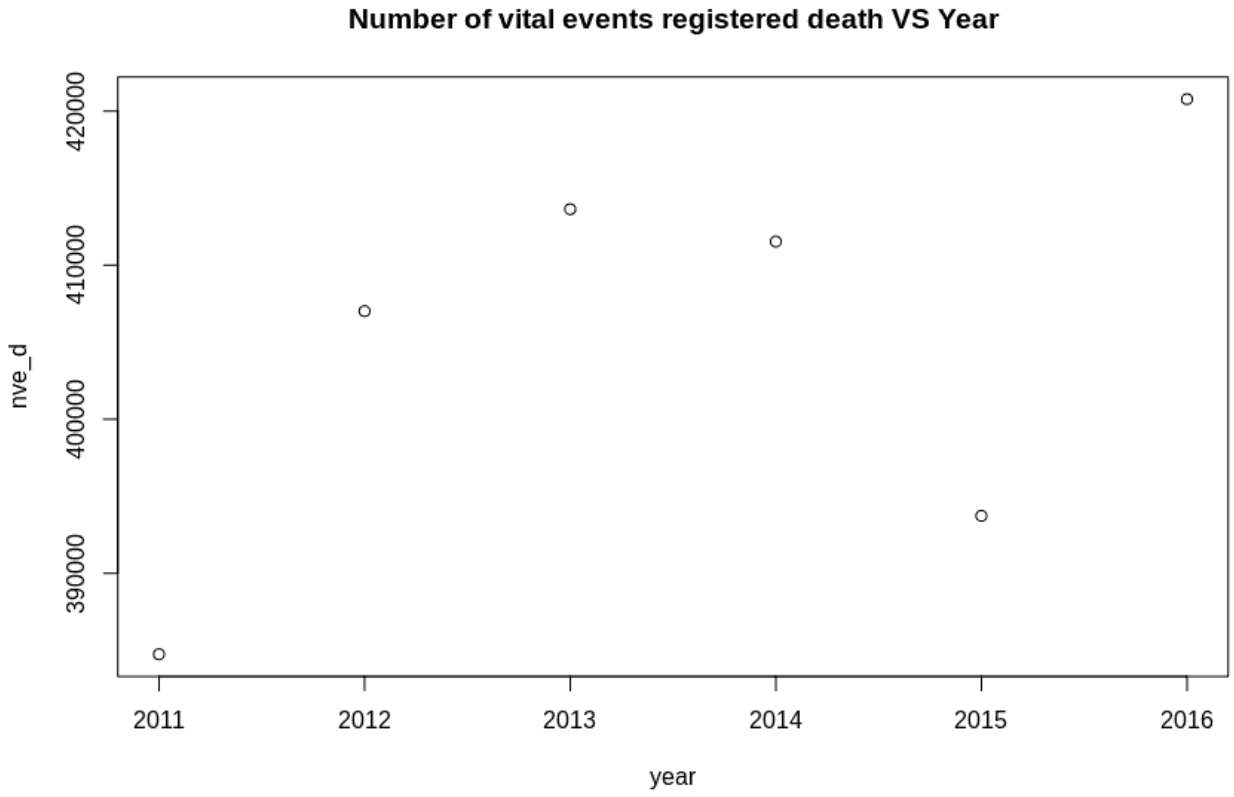
Line Graph of Number of vital events registered still births VS Year

```
ggplot(data,aes(x = year,y = nve_s)) + geom_line(color = "blue") + ggtitle("Number of vital events registered
still births VS Year")
```



scatter plot of Number of vital events registered death VS Year

with (df2, plot (year,nve_d,main = "Number of vital events registered death VS Year"))



Comparision for rural and urban

ubru <- data.frame(urban$Birth_rate,
colnames(ubru) <- c("Urban","Rural","Districts")

```
mubru <- melt(ubru,id.vars= "Districts")
ggplot(mubru,aes(x= Districts , y = value,fill = variable))+ylab("Birth Rate") + geom_bar(stat =
"identity",position = "dodge") + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```