Time Bands Per Hospital

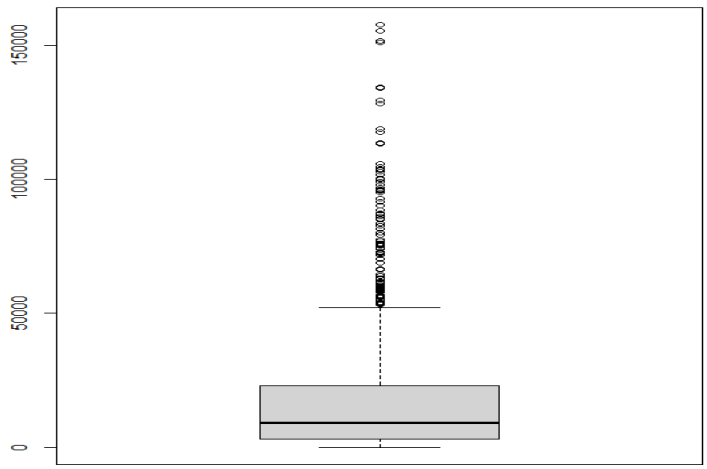
```
library(ggplot2)
#install.packages("ggthemes")
library(ggthemes)
#install.packages("forecast")
library(forecast)
#install.packages("tseries")
library(tseries)
#install.packages("ggfortify")
library(ggfortify)
#install.packages("summarytools")
library(summarytools)
library (autoplotly)
install.packages("dummies")
library(dummies)
#Import the data
TimeBands_Hospital<- read.csv(file.choose())
attach(TimeBands_Hospital)
View(TimeBands_Hospital)
```

```
      Year          Hospital          Time.Bands           Count
 Min.   :2014   Length:1504        Length:1504        Min.   :    1
 1st Qu.:2015   Class :character   Class :character   1st Qu.:  3019
 Median :2016   Mode  :character   Mode  :character   Median :  9171
 Mean   :2016                                         Mean   : 17273
 3rd Qu.:2017                                         3rd Qu.: 22983
 Max.   :2018                                         Max.   :157710
```

```
# Take a peek at the dataset
TimeBands_Hospital
```

Time Bands Per Hospital

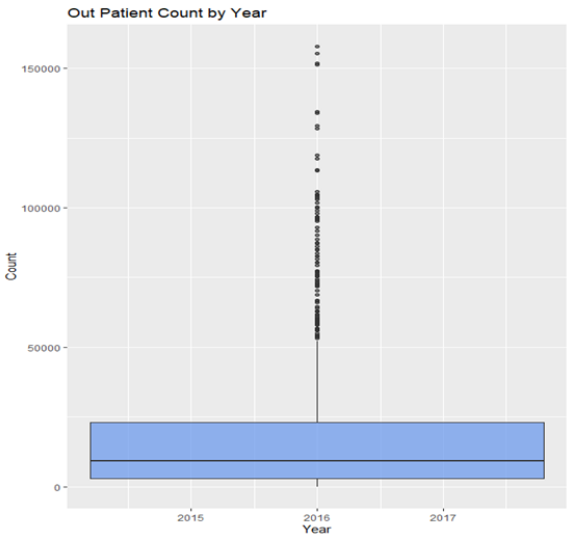
```
str(TimeBands_Hospital)

##Check the null values

is.null(TimeBands_Hospital)## No Null values

#table

table(TimeBands_Hospital$Count)

##Label encoding

#install.packages("superml")

library(superml)

lbl <- LabelEncoder$new()

lbl$fit(TimeBands_Hospital$Hospital)

TimeBands_Hospital$Hospital <- lbl$fit_transform(TimeBands_Hospital$Hospital)


lbl$fit(TimeBands_Hospital$Time.Bands)

TimeBands_Hospital$Time.Bands <- lbl$fit_transform(TimeBands_Hospital$Time.Bands)


# Some EDA By Visualization

library(tidyr)

## Box Plot

boxplot(TimeBands_Hospital$Count)
```



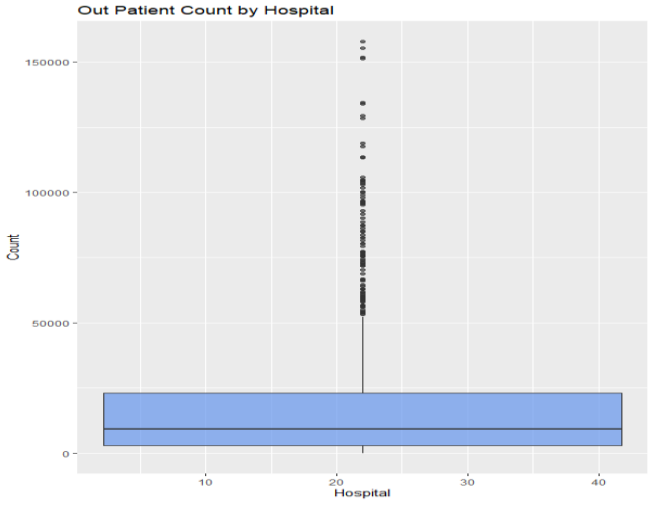
There are outliers present in the data

Time Bands Per Hospital

```
# plot the distribution of Count by year using boxplots
ggplot(TimeBands_Hospital,
       aes(x = Year,
           y = Count)) +
  geom_boxplot(fill = "cornflowerblue",
               alpha = .7) +
  labs(title = "Out Patient Count by Year")
```

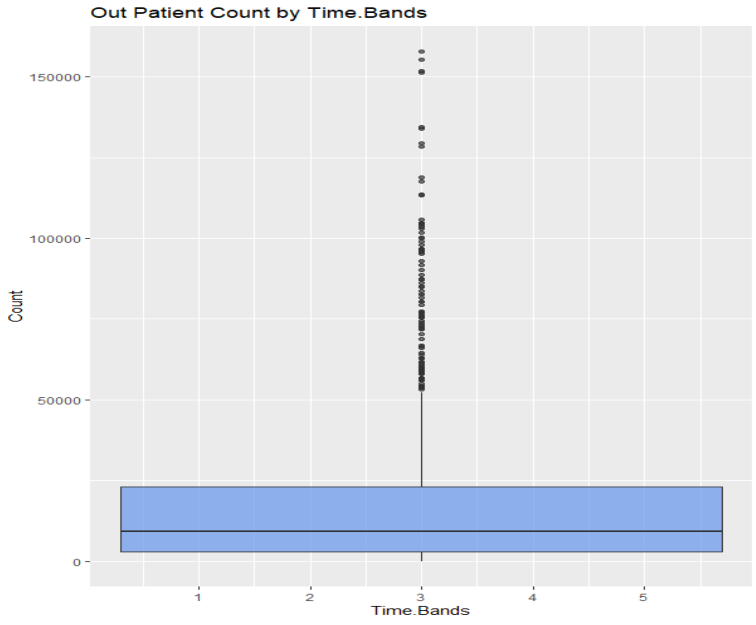



```
#plot the distribution of Count by Hospital using boxplots
ggplot(TimeBands_Hospital,
       aes(x = Hospital,
           y = Count)) +
  geom_boxplot(fill = "cornflowerblue",
               alpha = .7) +labs(title = "Out Patient Count by Hospital")
```

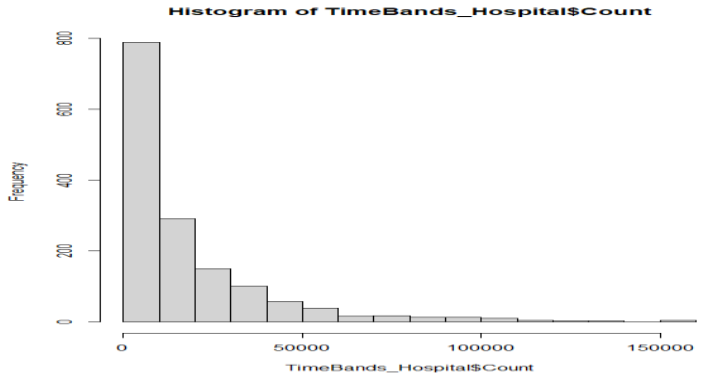


Time Bands Per Hospital

```
#plot the distribution of Count by Time.Bands using boxplots
ggplot(TimeBands_Hospital,
       aes(x = Time.Bands,
           y = Count)) +
  geom_boxplot(fill = "cornflowerblue",
               alpha = .7) +
  labs(title = "Out Patient Count by Time.Bands")
```

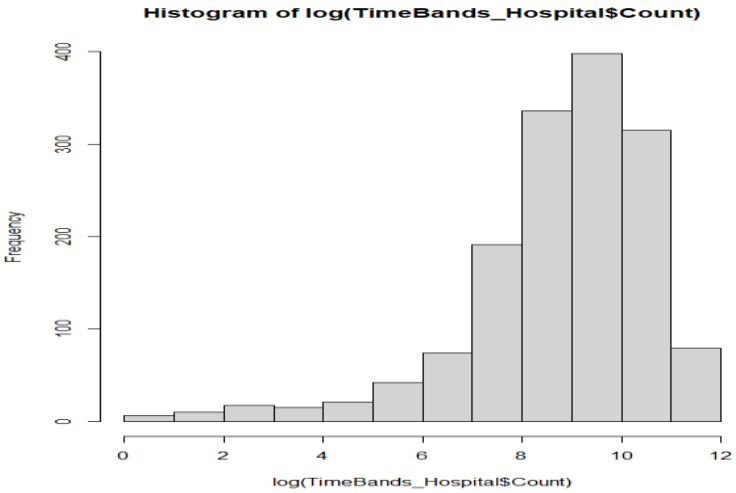



```
###Histogram simple
hist(TimeBands_Hospital$Count)##Data is Right/Positively skewed
```

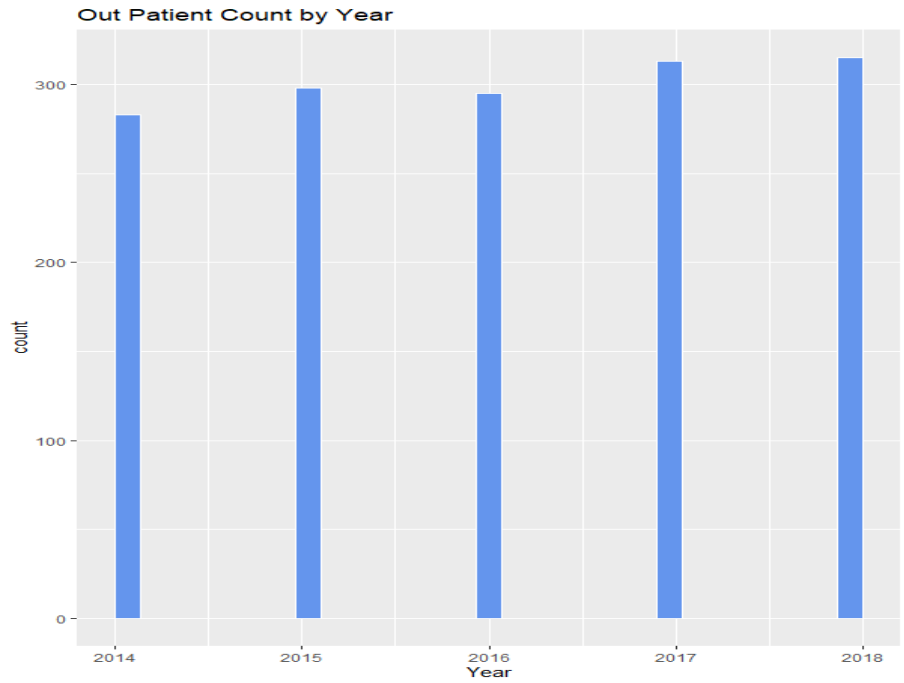


Time Bands Per Hospital

```
hist(log(TimeBands_Hospital$Count))
```



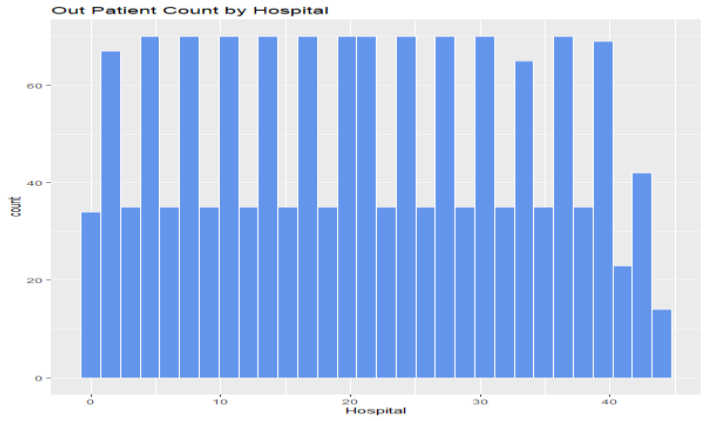

```
# plot the histogram with a binwidth of 5
ggplot(TimeBands_Hospital, aes(x = Year)) +
  geom_histogram(fill = "cornflowerblue",
                 color = "white",) +
  labs(title="Out Patient Count by Year",
       x = "Year")
```



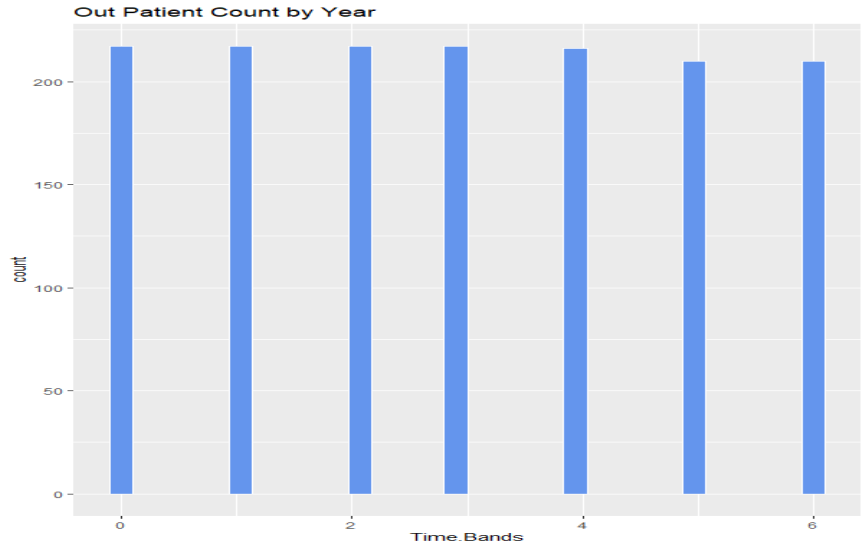

```
ggplot(TimeBands_Hospital, aes(x = Hospital)) +
  geom_histogram(fill = "cornflowerblue",
```

Time Bands Per Hospital

```
        color = "white",) +
  labs(title="Out Patient Count by Hospital",
       x = "Hospital")
```

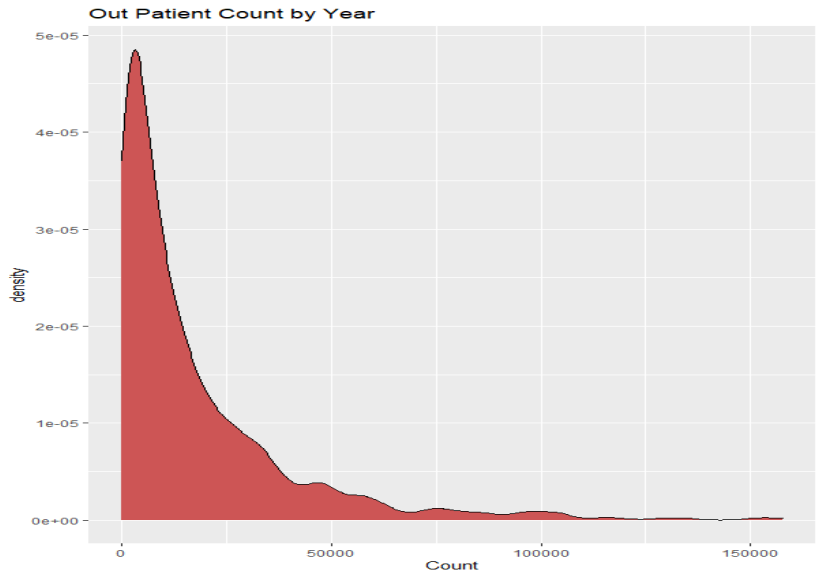



```
ggplot(TimeBands_Hospital, aes(x = Time.Bands)) +
  geom_histogram(fill = "cornflowerblue",
                 color = "white",) +   labs(title="Out Patient Count by Year",       x = "Time.Bands")
```




```
#An alternative to a histogram is the kernel density plot.
# Create a kernel density plot of Count
ggplot(TimeBands_Hospital, aes(x = Count)) +
  geom_density(fill = "indianred3") +
  labs(title = "Out Patient Count by Year")
```

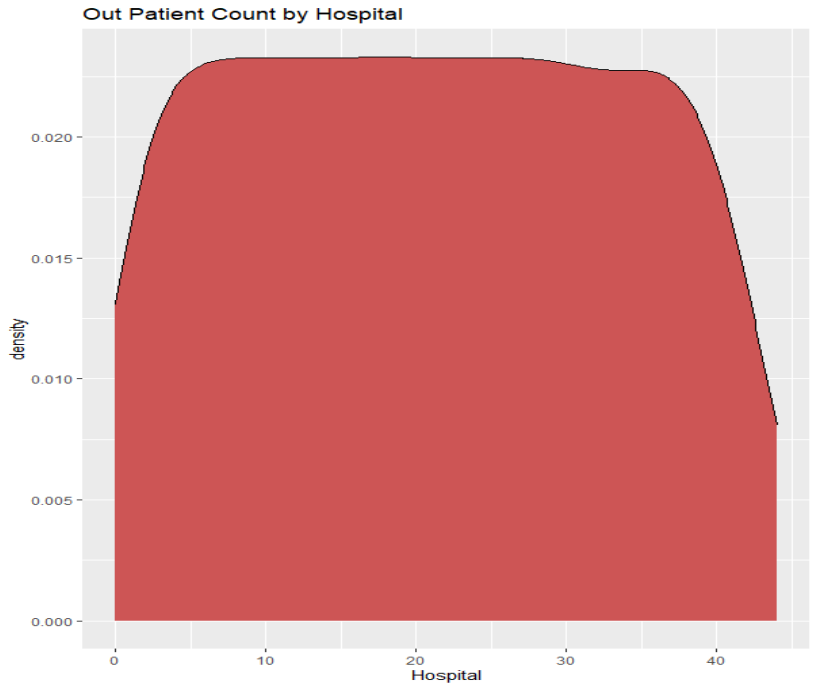
Time Bands Per Hospital



We can observe the data is right skewed.

##by Hopsital

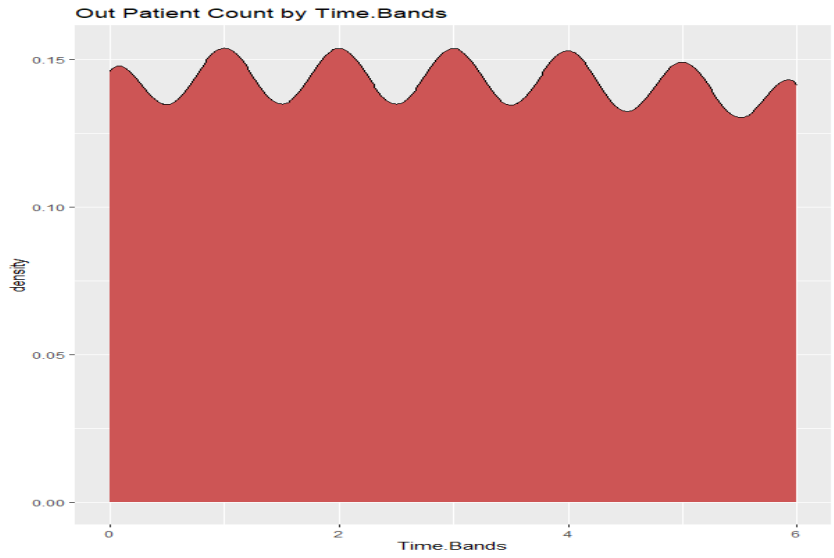
```
ggplot(TimeBands_Hospital, aes(x = Hospital)) +
  geom_density(fill = "indianred3") +
  labs(title = "Out Patient Count by Hospital")
```



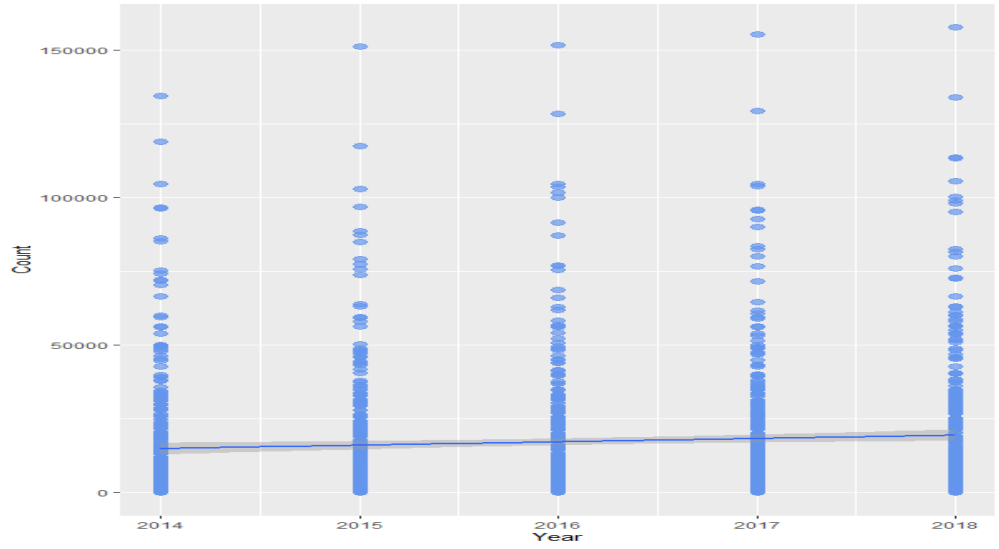
#By Time.Bands

Time Bands Per Hospital

```
ggplot(TimeBands_Hospital, aes(x = Time.Bands)) +
  geom_density(fill = "indianred3") +
  labs(title = "Out Patient Count by Time.Bands")
```




```
#scatterplot wiht X and Y axis
ggplot(data = TimeBands_Hospital,
       mapping = aes(x = Year, y = Count)) +
  geom_point(color = "cornflowerblue",
             alpha = .7,
             size = 3) +
  geom_smooth(method = "lm")### There is a increase in the count year wise and same trend for every two years. Patt
```



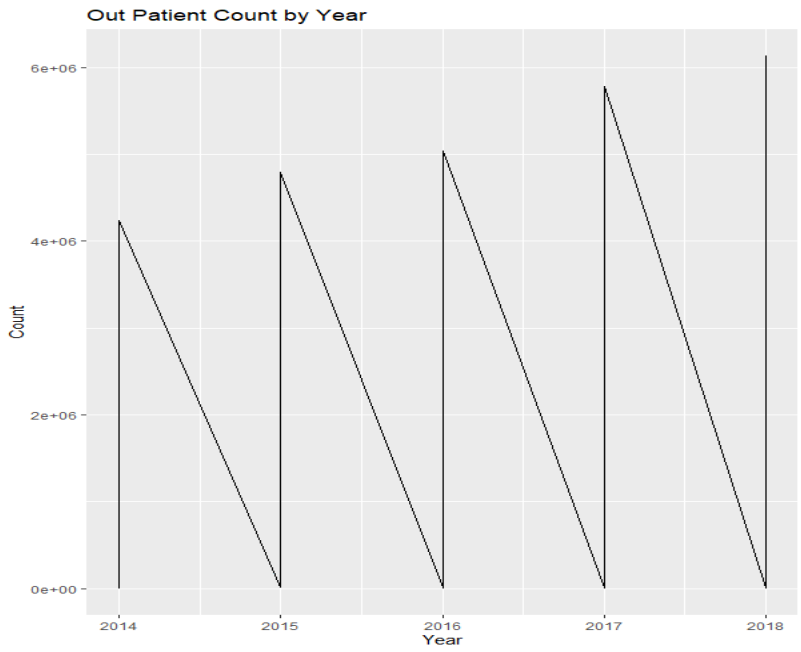
ern

Time Bands Per Hospital

#and we can observe there are outliers present in the data

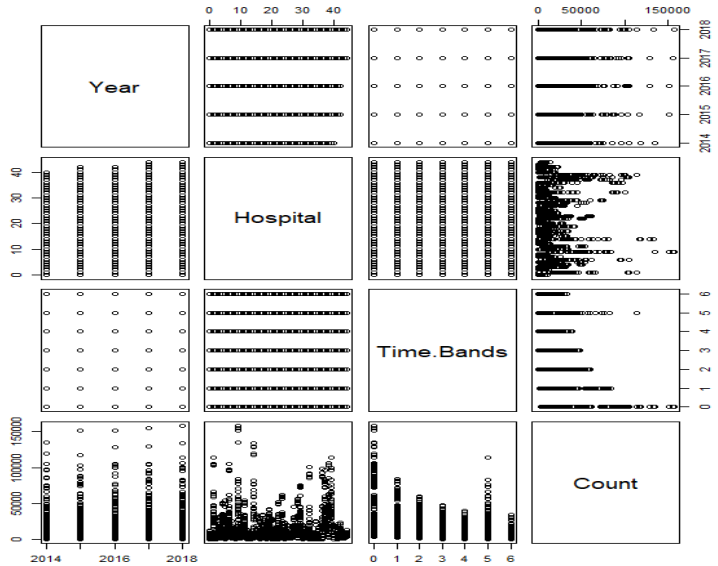
basic area chart

```
ggplot(TimeBands_Hospital, aes(x = Year, y = Count)) +
  geom_area(fill="lightblue", color="black") +
  labs(title = "Out Patient Count by Year",
       x = "Year",
       y = "Count)
```



clearly shows there is some pattern in the data. Year by year the count is increasing

```
pairs(TimeBands_Hospital)
```



Time Bands Per Hospital

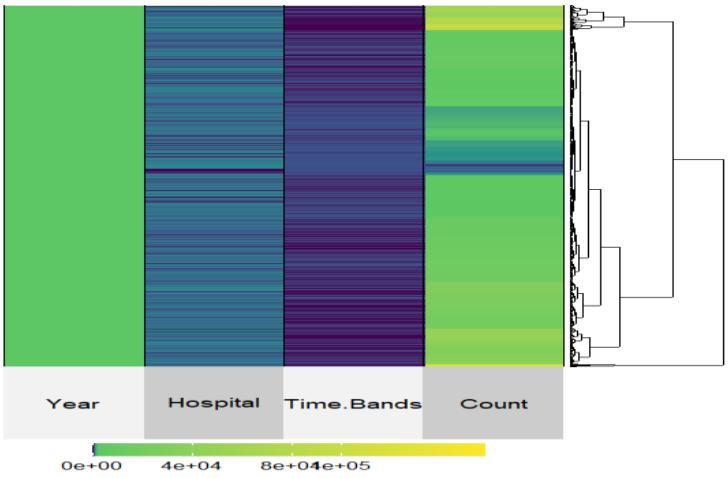
#Pair plot show the relationship between the variables.

sorted heat map

#install.packages("superheat")

library(superheat)

superheat(TimeBands_Hospital,row.dendrogram = TRUE)



###Outliers treatment by capping

x <- TimeBands_Hospital$Count

qnt <- quantile(x, probs=c(.25, .75), na.rm = T)

caps <- quantile(x, probs=c(.05, .95), na.rm = T)

H <- 1.5 * IQR(x, na.rm = T)

x[x < (qnt[1] - H)] <- caps[1]

x[x > (qnt[2] + H)] <- caps[2]