

Patient Count data

```
###Importing required packages
```

```
#install. packages("tidyverse")
```

```
#install. packages("corrplot")
```

```
library(tidyverse)
```

```
library(corrplot)
```

```
library(ggplot2)
```

```
## Loading the data
```

```
Patients_count<- read_csv(file.choose())
```

```
attach (Patients_count)
```

```
View (Patients_count)
```

```
##Exploring and preparing the data ----
```

```
summary (Patients_count) ### explains statistical information of the data
```

Year	Count
Min. :2014	Min. :4236176
1st Qu.:2015	1st Qu.:4794370
Median :2016	Median :5037669
Mean :2016	Mean :5195765
3rd Qu.:2017	3rd Qu.:5782699
Max. :2018	Max. :6127911

```
str(Patients_count)
```

```
##Column Names
```

```
colnames(Patients_count)
```

```
##Check the null values
```

```
is.null(Patients_count) ## No Null values
```

```
#table
```

```
table(Patients_count$Count)
```

Patient Count data

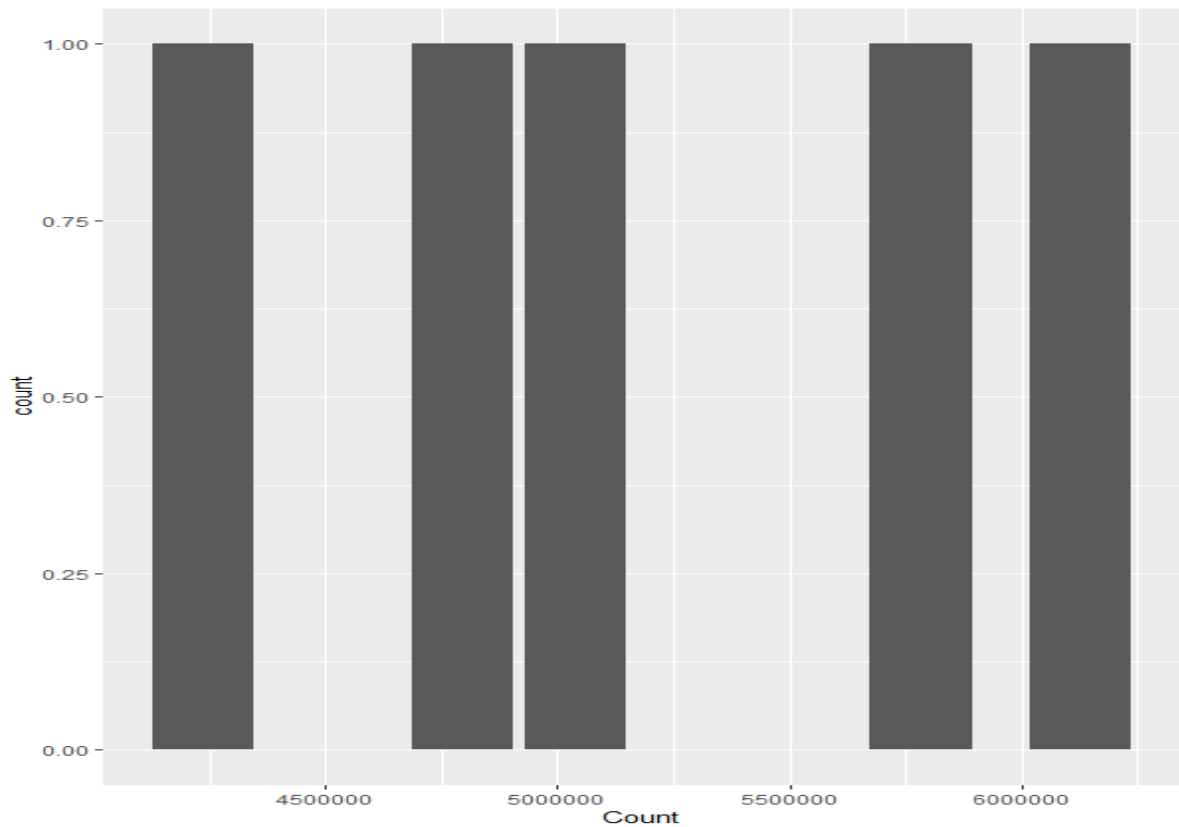
```
# Some EDA by Visualization
```

```
library(tidyr)
```

```
# Bar Plot
```

```
ggplot(data = Patients_count) +
```

```
  geom_bar(mapping = aes(x = Count))
```



```
# plot the distribution of Count with modified colors and labels
```

```
ggplot(Patients_count, aes(x = Year)) +
```

```
  geom_bar(fill = "cornflowerblue",
```

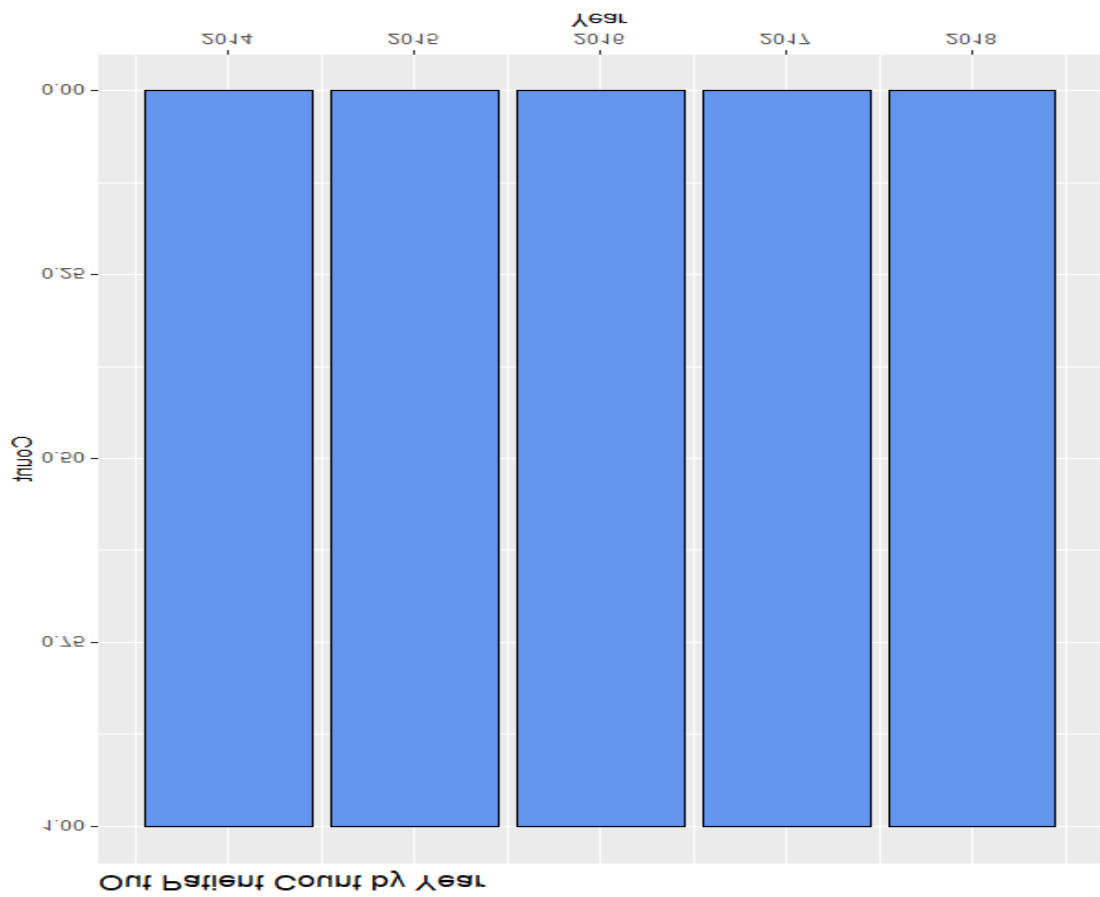
```
    color="black") +
```

```
  labs(x = "Year",
```

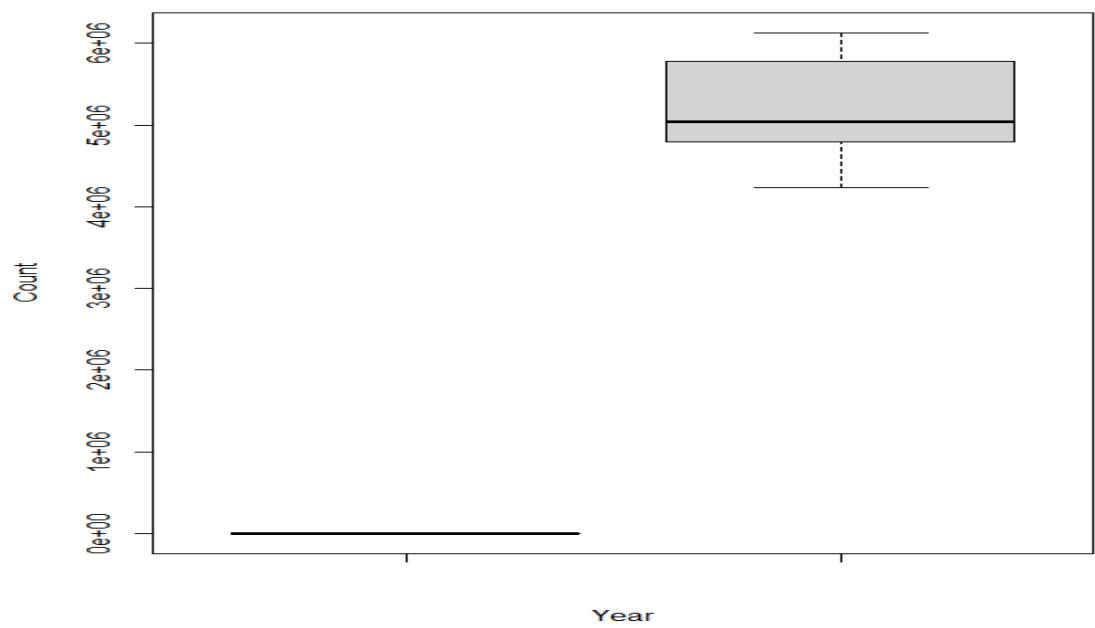
```
    y = "Count",
```

```
    title = "Outpatient Count by Year")
```

Patient Count data



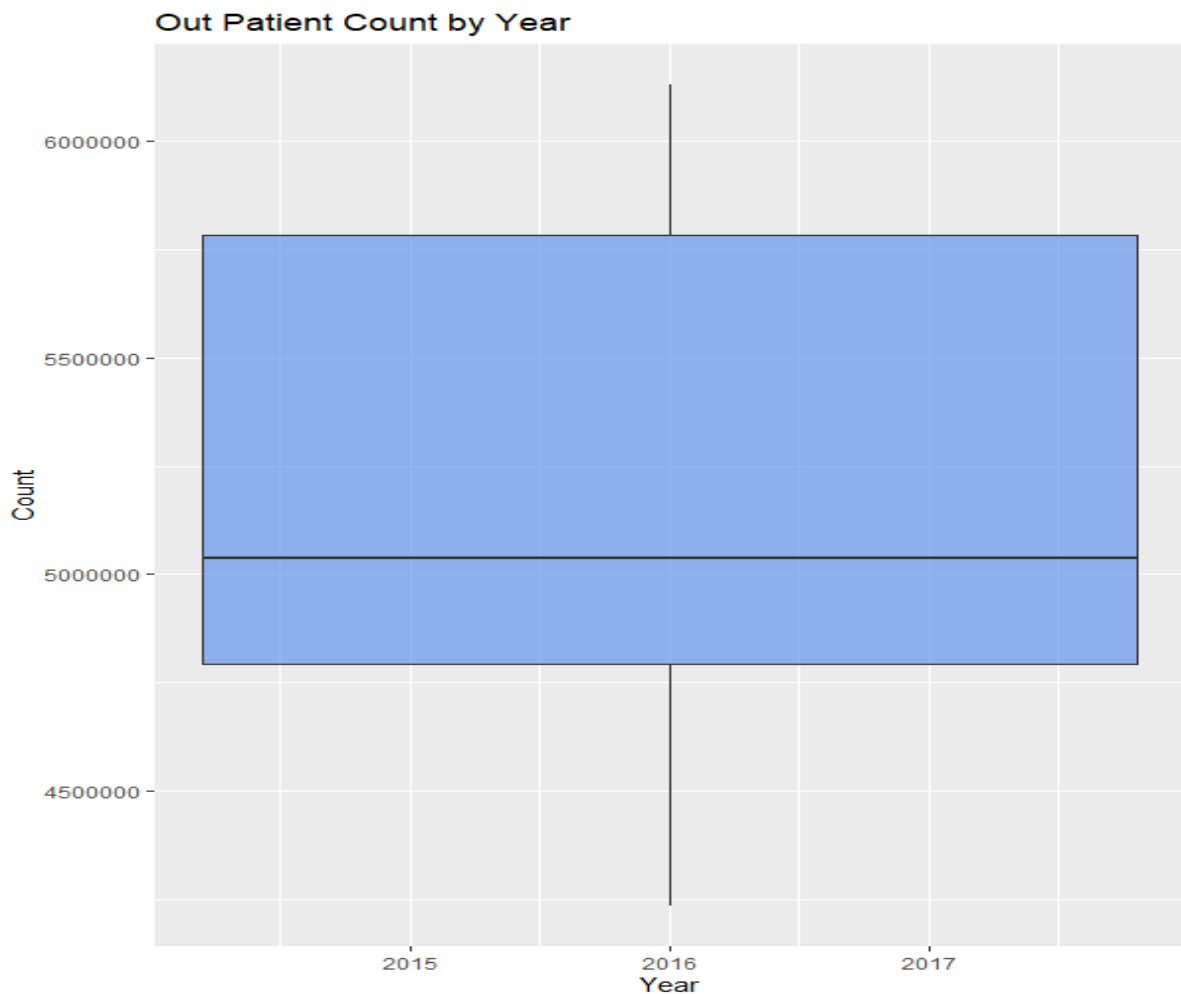
```
## Box Plot
boxplot(Patients_count$Count)
boxplot(Year,Count,xlab="Year",ylab="Count",title="Patients count data")
```



Patient Count data

plot the distribution of Count by year using boxplots

```
ggplot(Patients_count,  
  aes(x = Year,  
    y = Count)) +  
  geom_boxplot(fill = "cornflowerblue",  
    alpha = .7) +  
  labs(title = "Out Patient Count by Year")
```

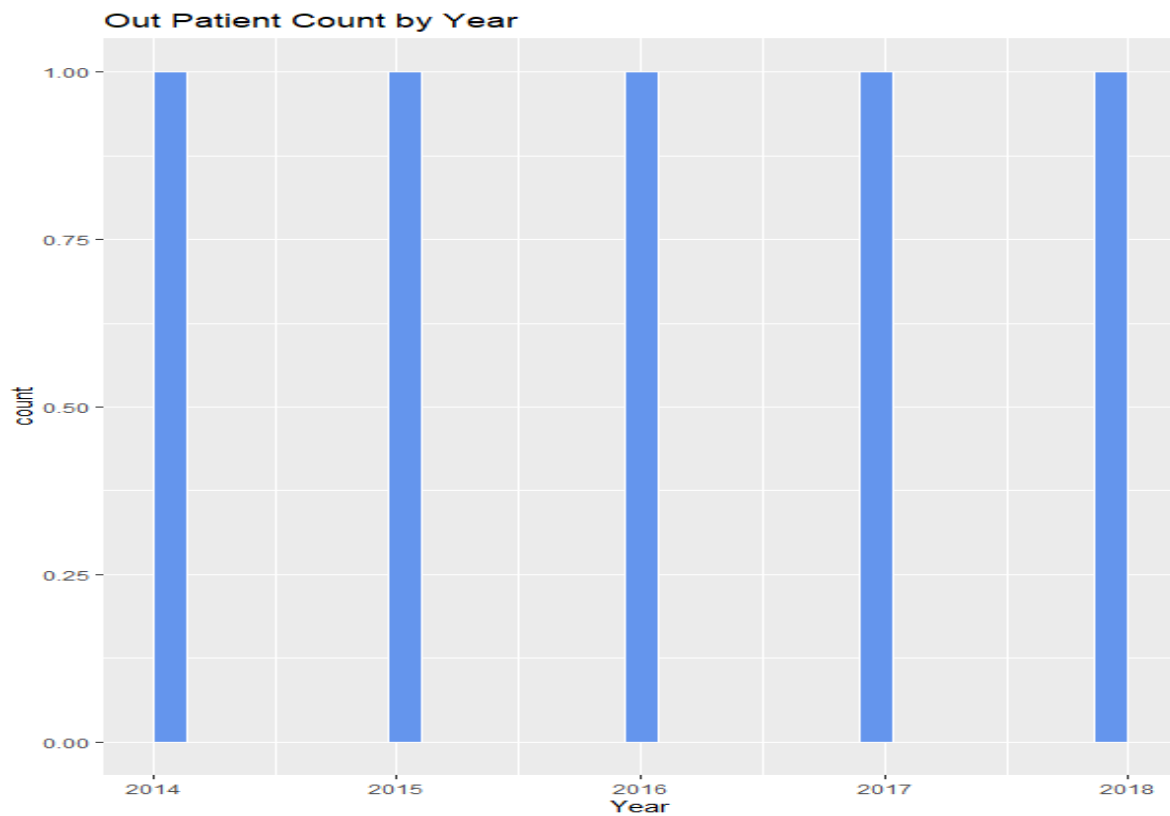


###Histogram simple

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

Patient Count data

x = "Year")



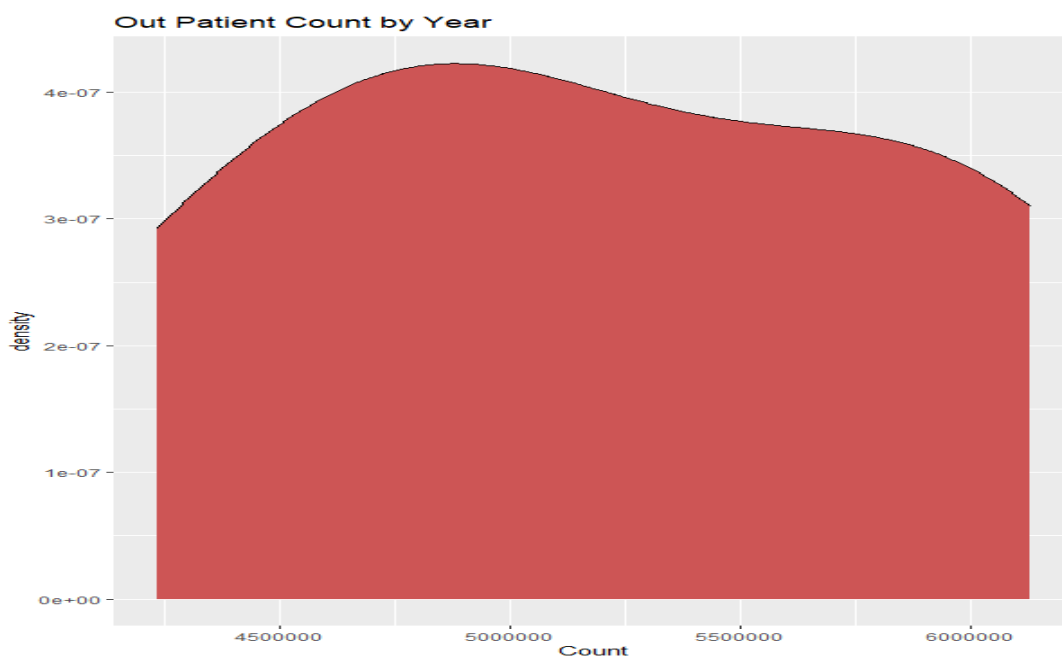
#An alternative to a histogram is the kernel density plot.

Create a kernel density plot of Count

```
ggplot(Patients_count, aes(x = Count)) +
```

```
geom_density(fill = "indianred3") +
```

```
labs(title = "Out Patient Count by Year")
```

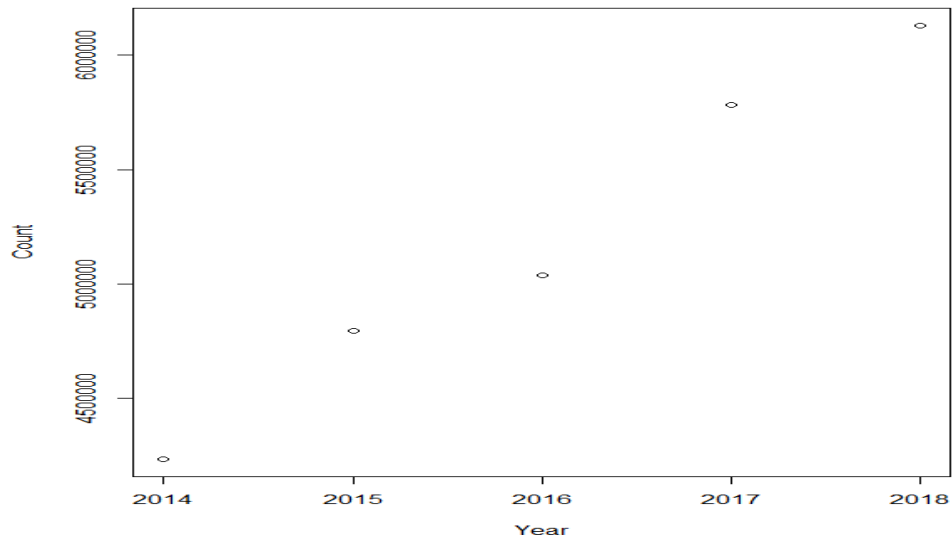


Patient Count data

```
## simple Scatter plot
```

```
plot(Year,Count)
```

```
abline(lm(Year~Count))
```



```
# redraw scatterplot wiht X and Y axis
```

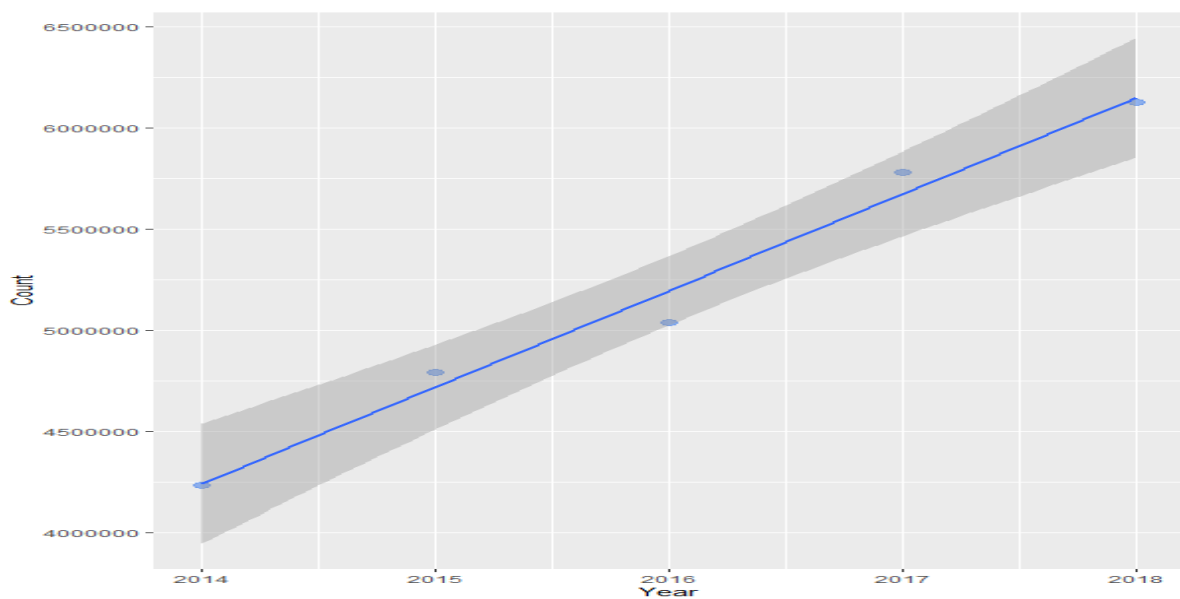
```
ggplot(data = Patients_count,
```

```
  mapping = aes(x = Year, y = Count)) +
```

```
  geom_point(color = "cornflowerblue",
```

```
    alpha = .7,
```

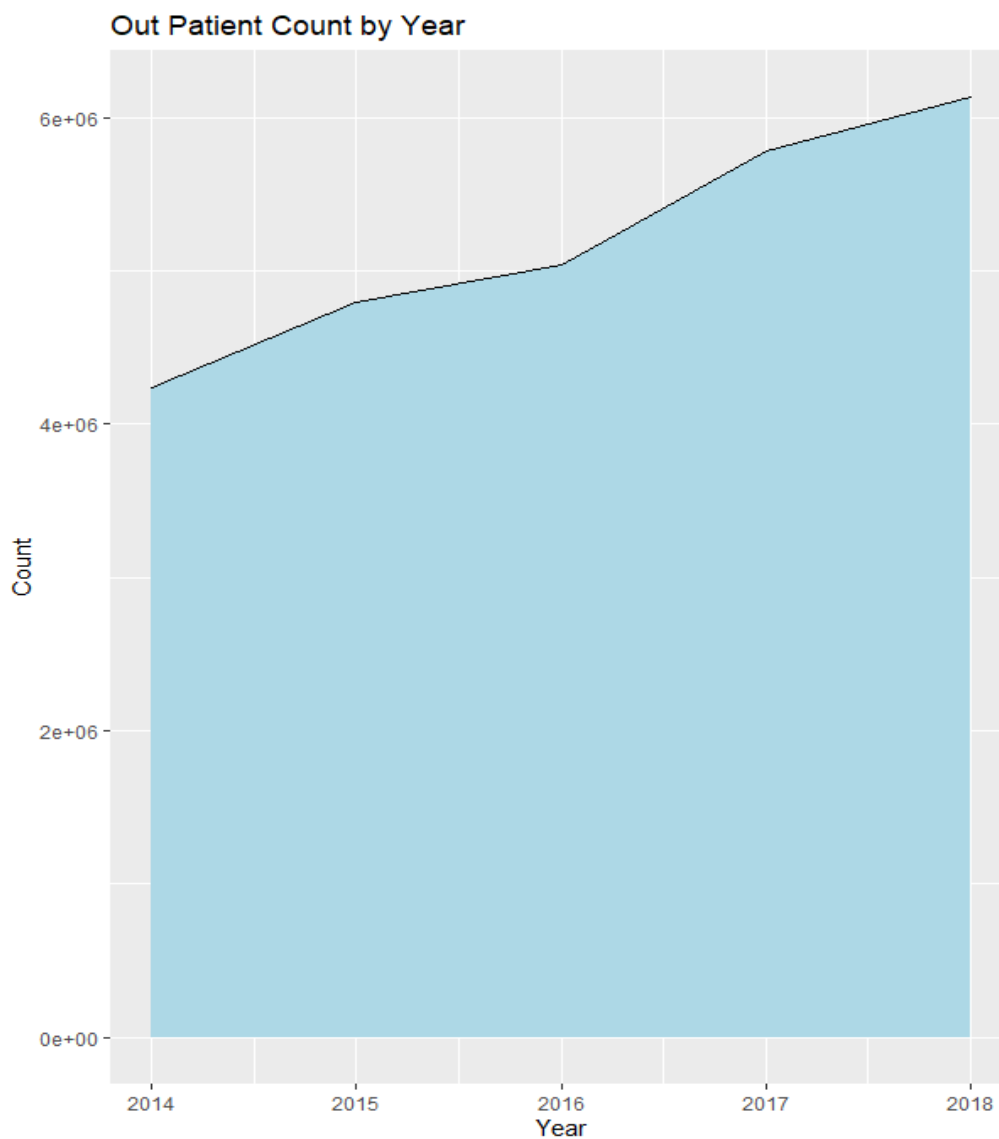
```
    size = 3) + geom_smooth(method = "lm")### There is a increase in the count year wise and  
linear trend with pattern
```



Patient Count data

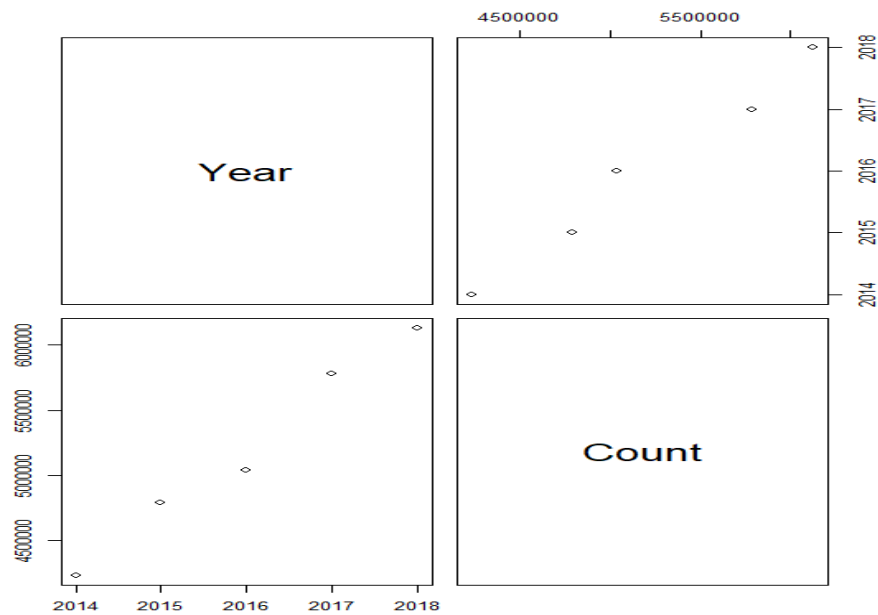
basic area chart

```
ggplot(Patients_count, aes(x = Year, y = Count)) +  
  geom_area(fill="lightblue", color="black") +  
  labs(title = "Out Patient Count by Year",  
        x = "Year",  
        y = "Count")## clearly explains the pattern of the data
```



Patient Count data

```
pairs(Patients_count)
```



```
# sorted heat map
```

```
install.packages("superheat")
```

```
library(superheat)
```

```
superheat(Patients_count,row.dendrogram = TRUE )
```

