Histogram Analysis for Insurance Data Set

Submitted by,

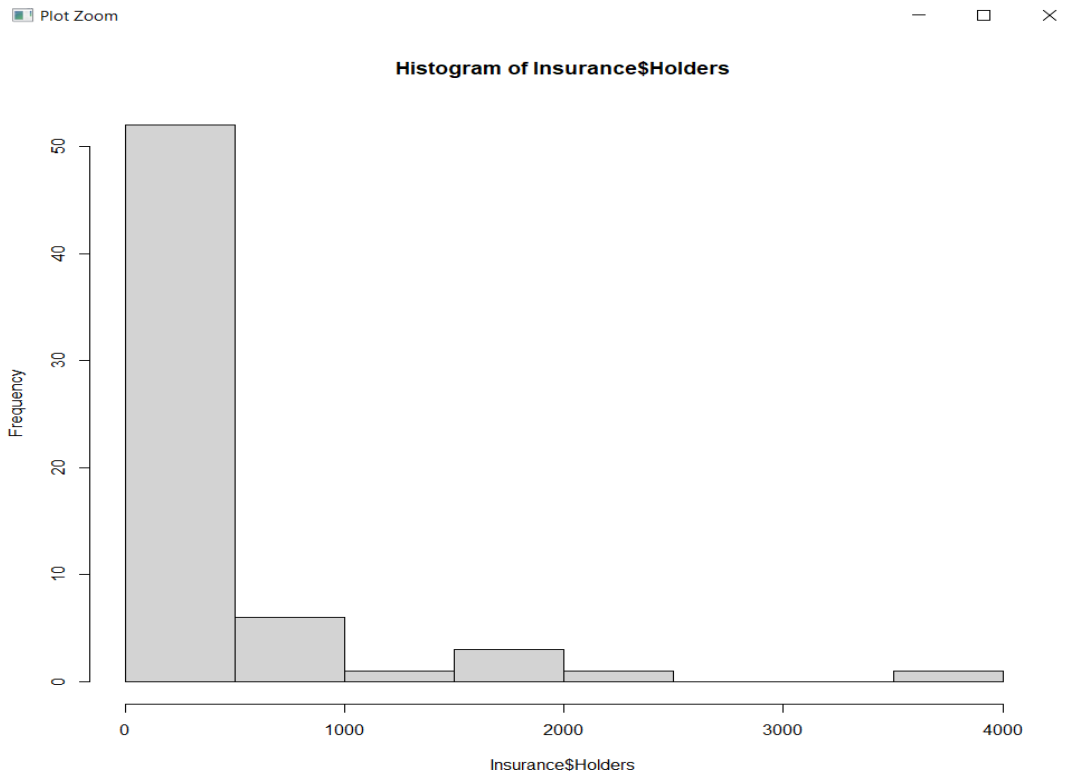
Manoj.K

20CSEG16

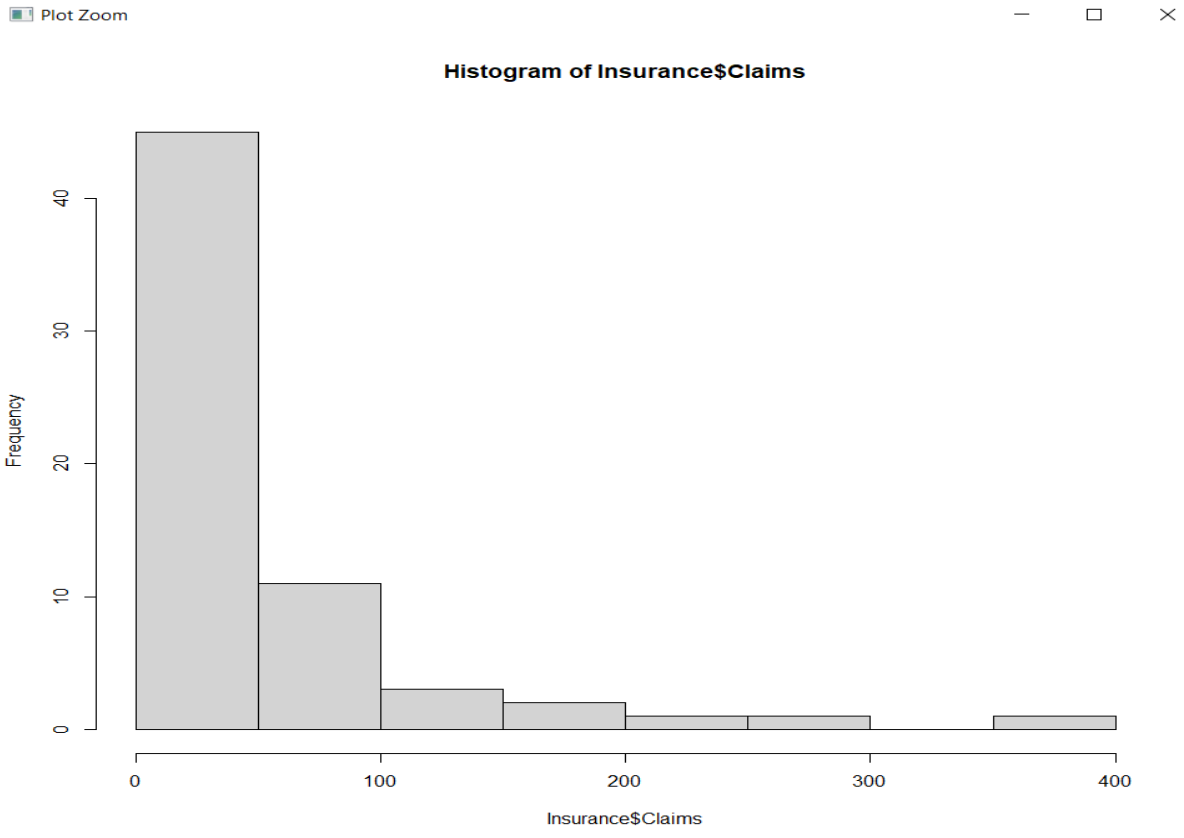
R-Code:

```
Console  Terminal  Jobs
~/
> library(MASS)
> View(Insurance)
> data("Insurance")
> summary(Insurance)
 District    Group       Age        Holders            Claims
 1:16     <1l   :16   <25  :16   Min.   :   3.00   Min.   :  0.00
 2:16     1-1.5l:16   25-29:16   1st Qu.:  46.75   1st Qu.:  9.50
 3:16     1.5-2l:16   30-35:16   Median : 136.00   Median : 22.00
 4:16     >2l   :16   >35  :16   Mean   : 364.98   Mean   : 49.23
                                 3rd Qu.: 327.50   3rd Qu.: 55.50
                                 Max.   :3582.00   Max.   :400.00
> nrow(Insurance)
[1] 64
> ncol(Insurance)
[1] 5
> str(Insurance)
'data.frame':   64 obs. of  5 variables:
 $ District: Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
 $ Group   : Ord.factor w/ 4 levels "<1l"<"1-1.5l"<..: 1 1 1 1 2 2 2 2 3 3 ...
 $ Age     : Ord.factor w/ 4 levels "<25"<"25-29"<..: 1 2 3 4 1 2 3 4 1 2 ...
 $ Holders : int  197 264 246 1680 284 536 696 3582 133 286 ...
 $ Claims  : int  38 35 20 156 63 84 89 400 19 52 ...
> head(Insurance)
  District  Group   Age Holders Claims
1        1    <1l   <25     197     38
2        1    <1l 25-29     264     35
3        1    <1l 30-35     246     20
4        1    <1l   >35    1680    156
5        1 1-1.5l   <25     284     63
6        1 1-1.5l 25-29     536     84
> tail(Insurance)
   District  Group   Age Holders Claims
59        4 1.5-2l 30-35      68     16
60        4 1.5-2l   >35     344     63
61        4    >2l   <25       3      0
62        4    >2l 25-29      16      6
63        4    >2l 30-35      25      8
64        4    >2l   >35     114     33
> hist(Insurance$Holders)
> hist(Insurance$Claims)
> hist(Insurance$Holders,breaks = 30,col = "blue",lwd=5)
> abline(v = mean(Insurance$Holders),col="red",lwd=4)
> abline(v = median(Insurance$Holders),col="green",lwd=3)
> hist(Insurance$Claims,breaks = 30,col = "blue",lwd=3)
> abline(v = mean(Insurance$Claims),col="red",lwd=4)
> abline(v = median(Insurance$Claims),col="brown",lwd=4)
> library(lattice)
> histogram(~Holders|Age,data = Insurance,breaks = 50,main="Holders vs Age",col=c(2,4))
> histogram(~Holders|District,data = Insurance,breaks = 50,main="Holders vs District",col=c(2,4))
> histogram(~Holders|Groupt,data = Insurance,breaks = 50,main="Holders vs Group",col=c(2,4))
Error in eval(modelRHS.vars[[i]], data, env) : object 'Groupt' not found
> histogram(~Holders|Group,data = Insurance,breaks = 50,main="Holders vs Group",col=c(2,4))
> histogram(~Claims|Group,data = Insurance,breaks = 50,main="Holders vs Group",col=c(2,4))
> histogram(~Claims|District,data = Insurance,breaks = 50,main="Claims vs District",col=c(2,4))
> histogram(~Claims|Group,data = Insurance,breaks = 50,main="Claims vs Group",col=c(2,4))
> histogram(~Claims|Age,data = Insurance,breaks = 50,main="Claims vs Age",col=c(2,4))
>
```

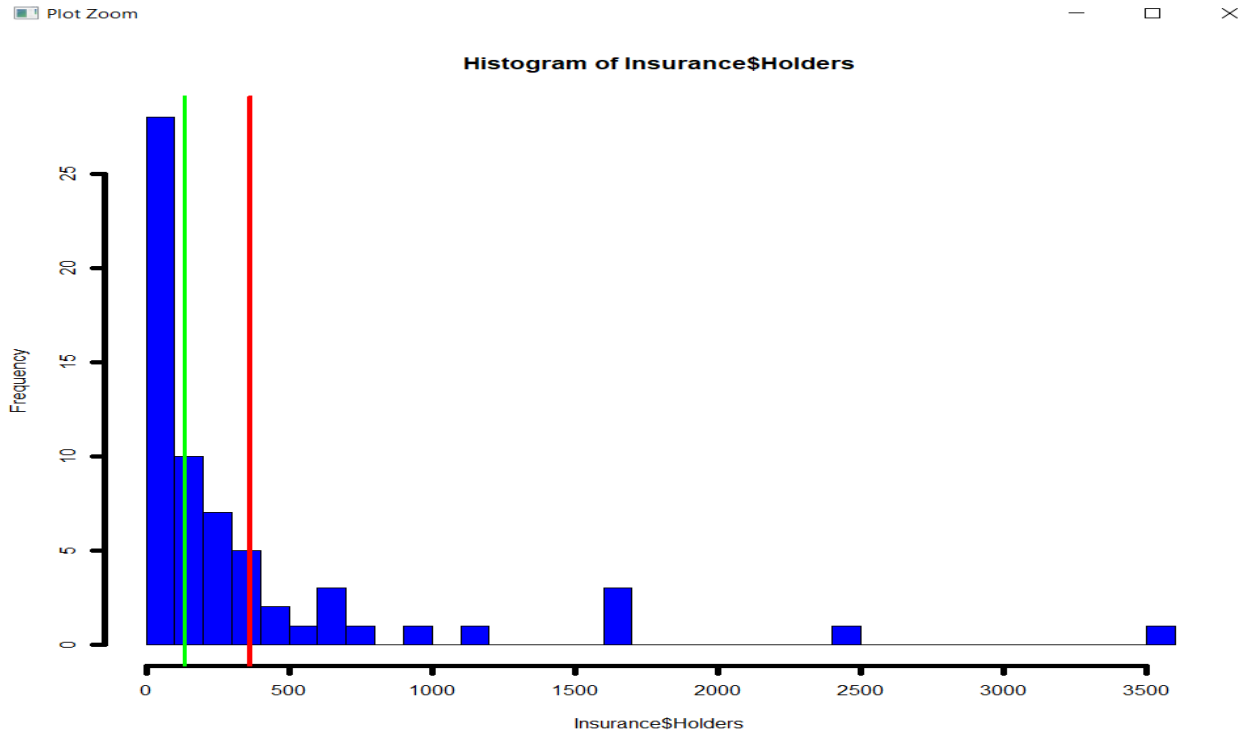
Histogram Plots:



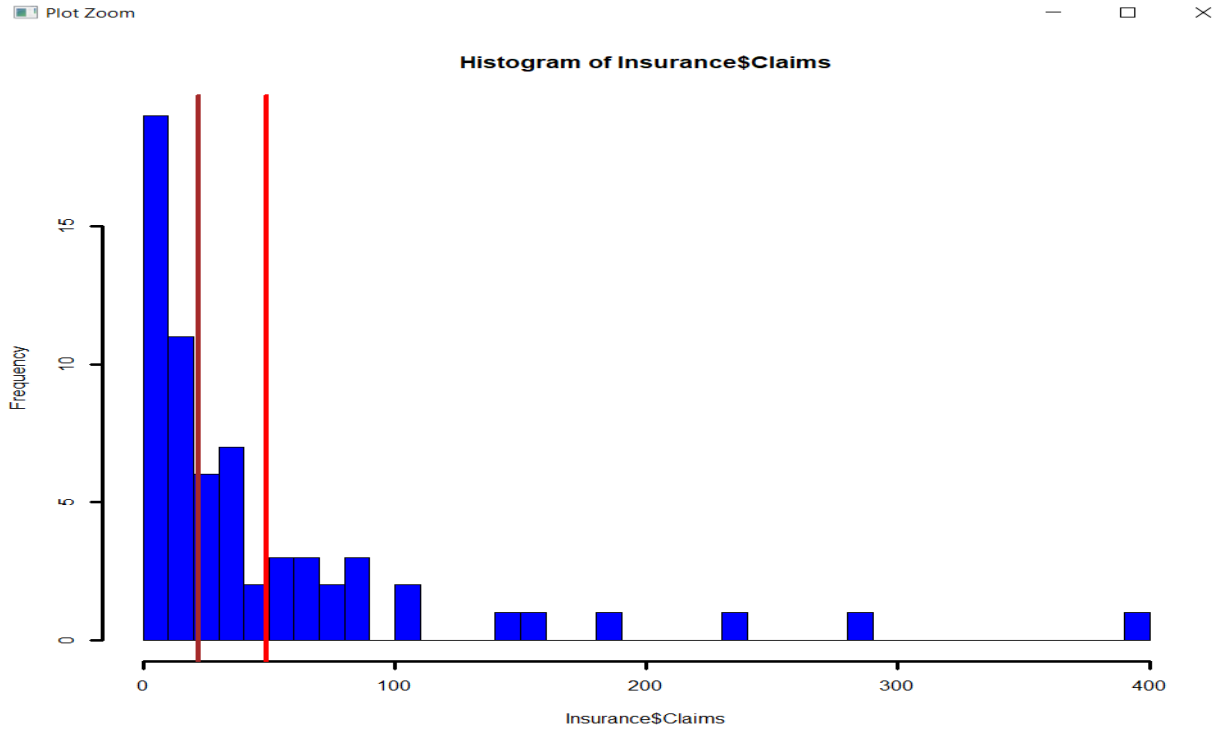
The above plot is about the number of holders in Insurance policy.



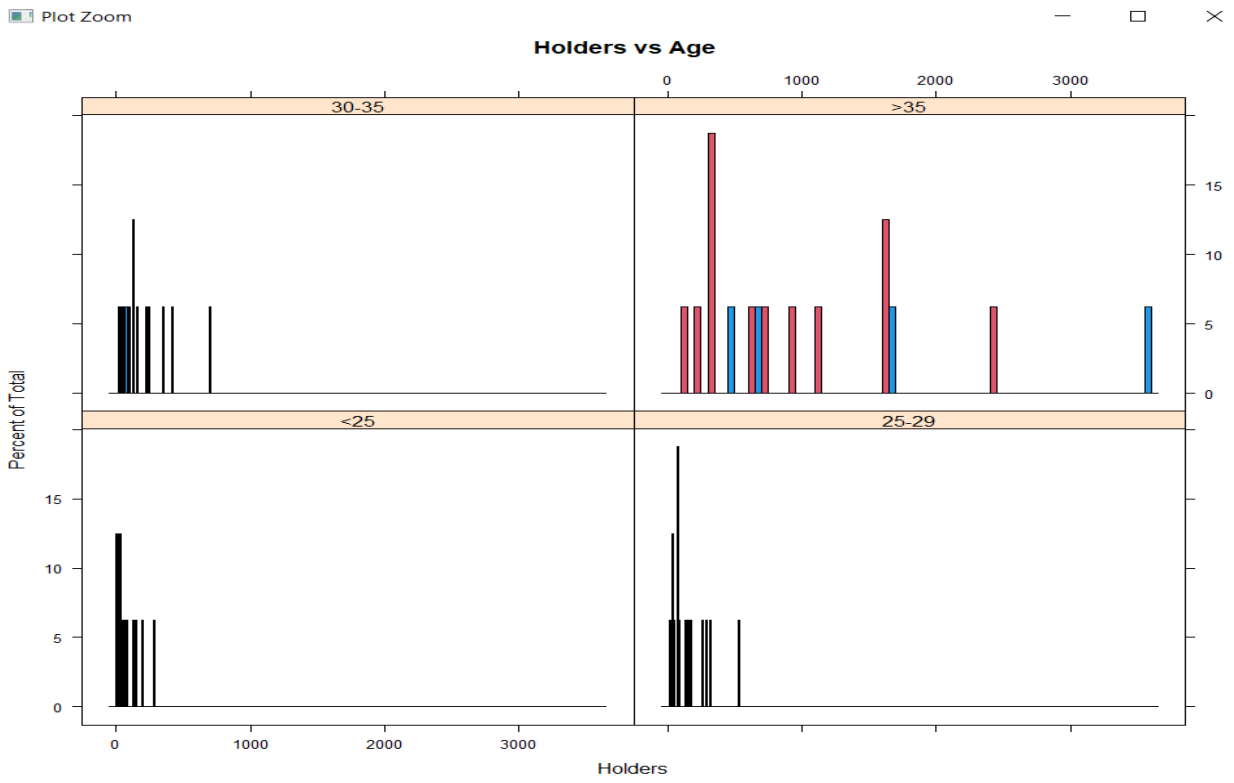
The above plot is about the number of claims in Insurance policy.



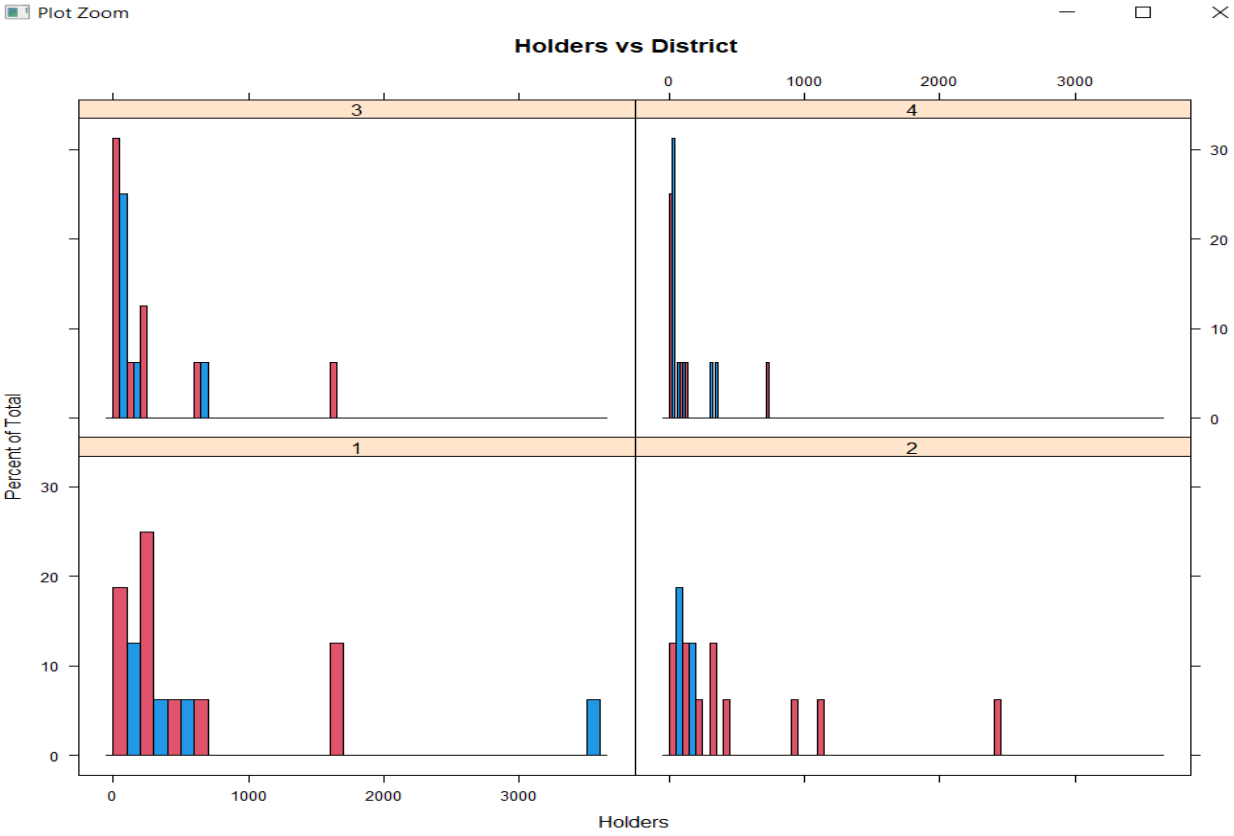
The mean and median of the holders are shown in the plot.



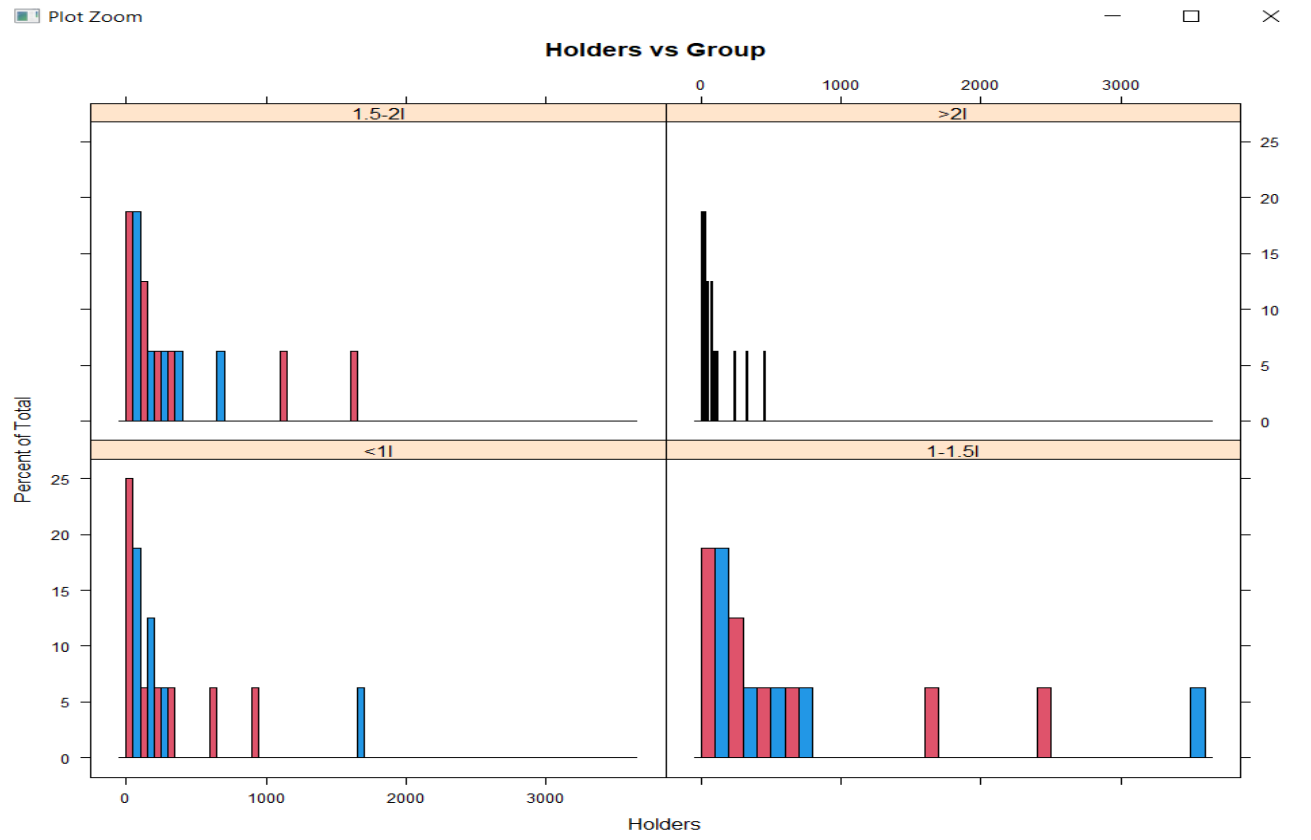
The mean and median of the claims is shown in the plot.



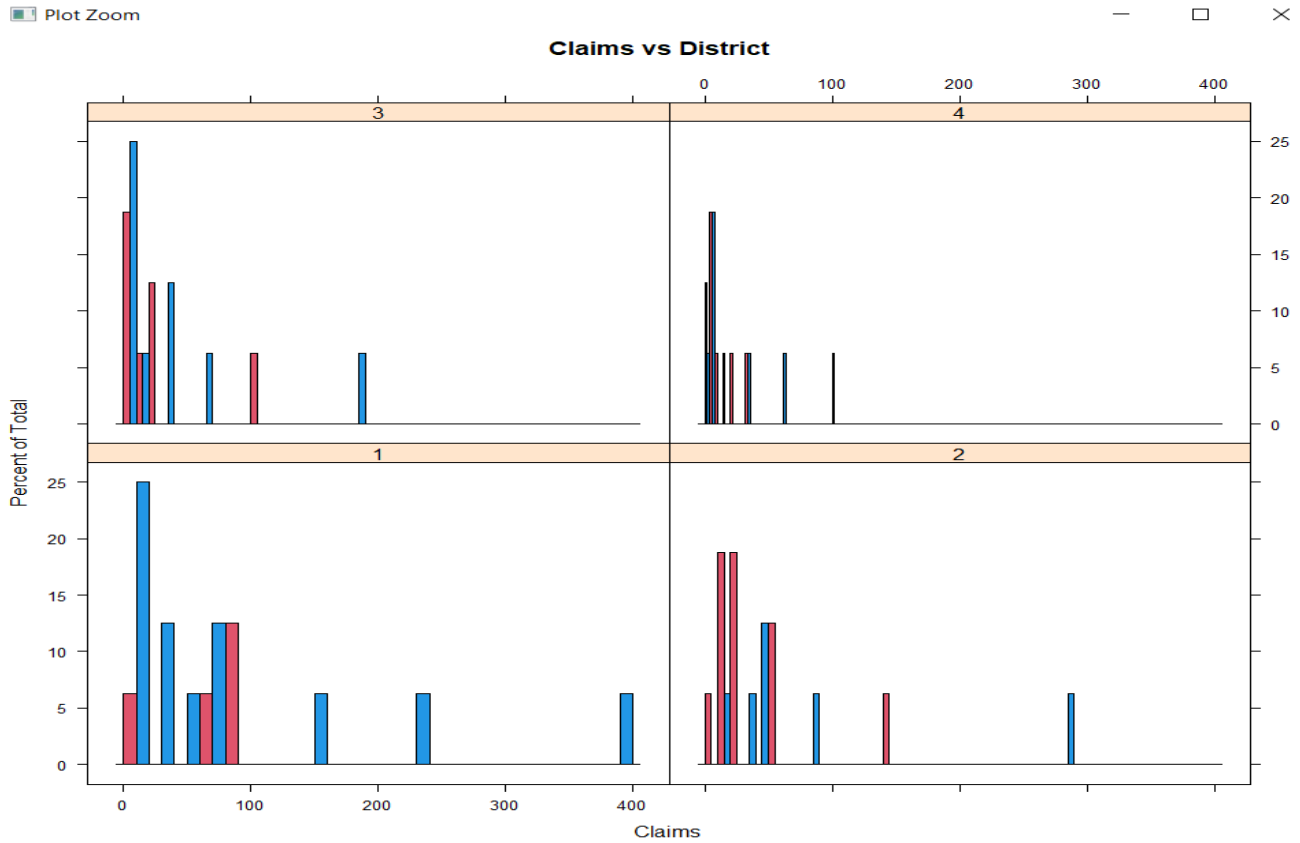
The histogram is plotted between the holders and the different age.



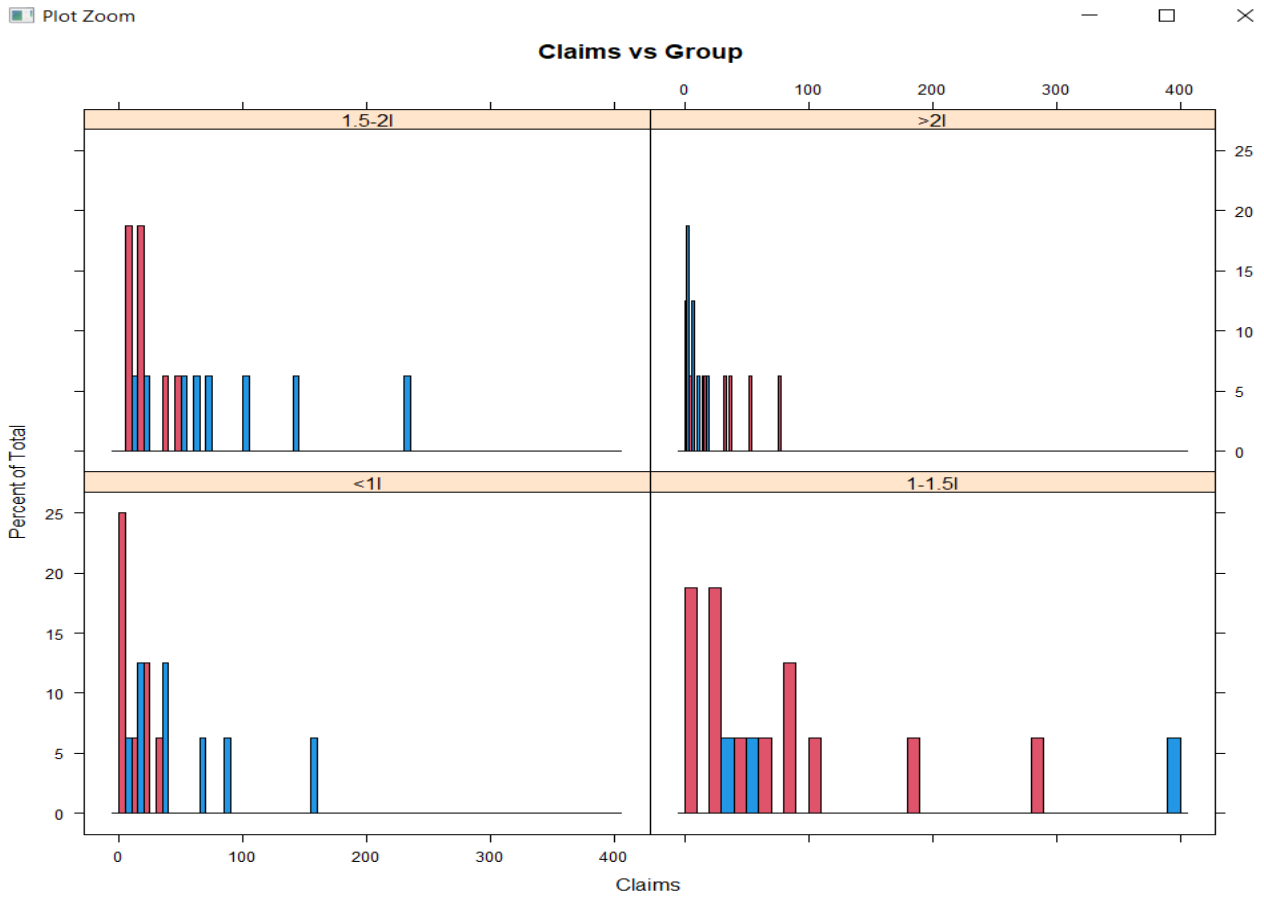
The histogram is drawn between the holders and district.



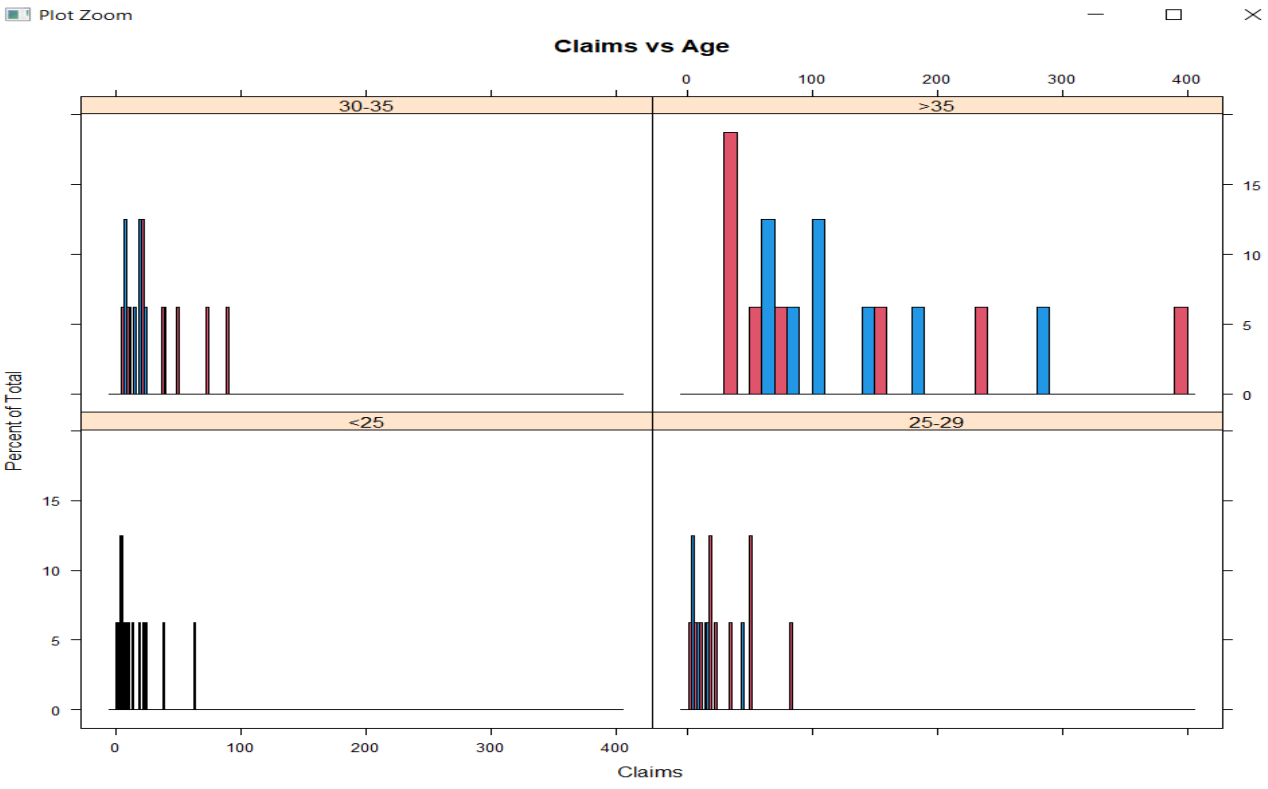
The histogram is plotted between the holders and group.



The histogram is drawn between the Claims and district.



The histogram is plotted between the Claims and group.



The histogram is plotted between the number of claims and Age.