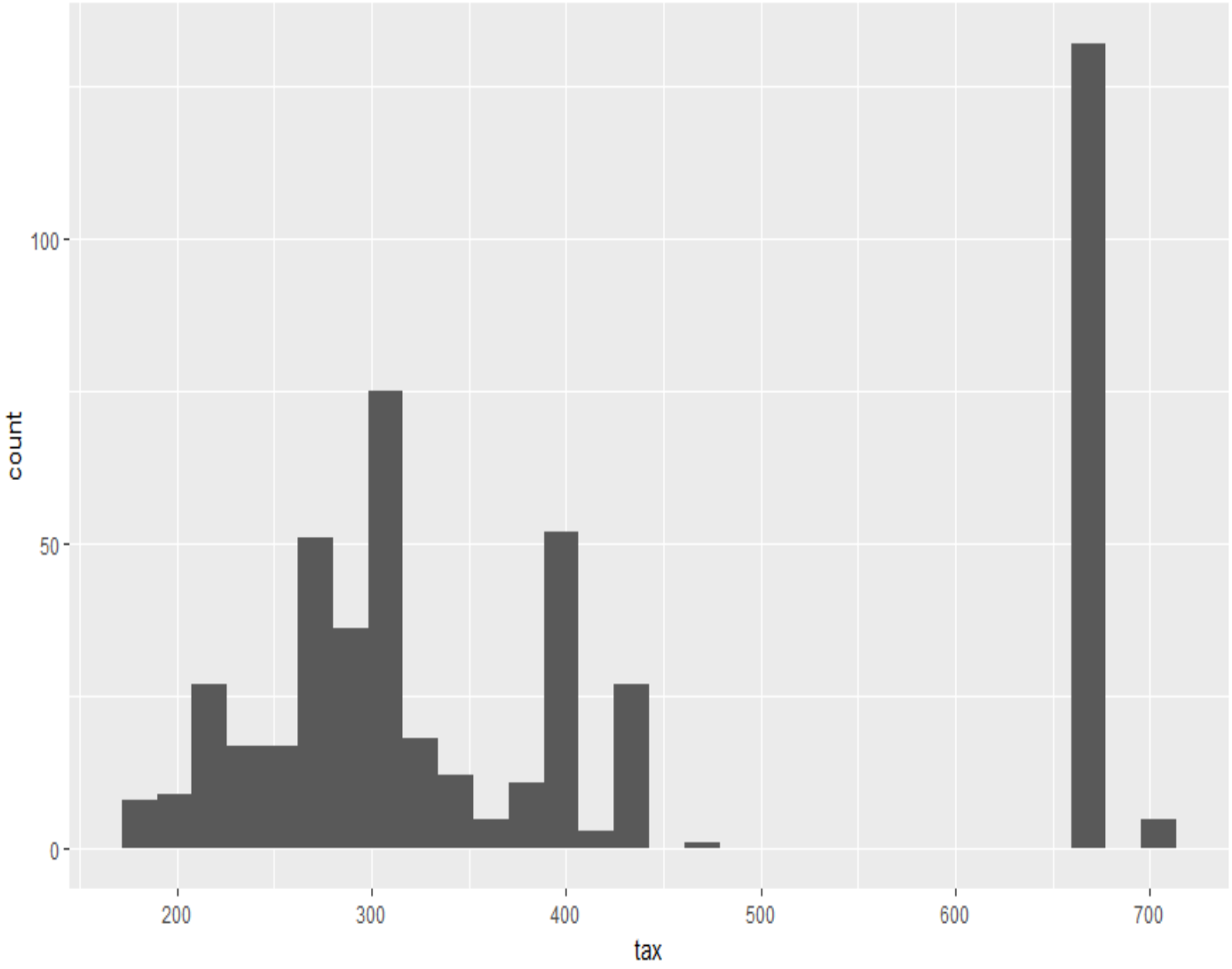
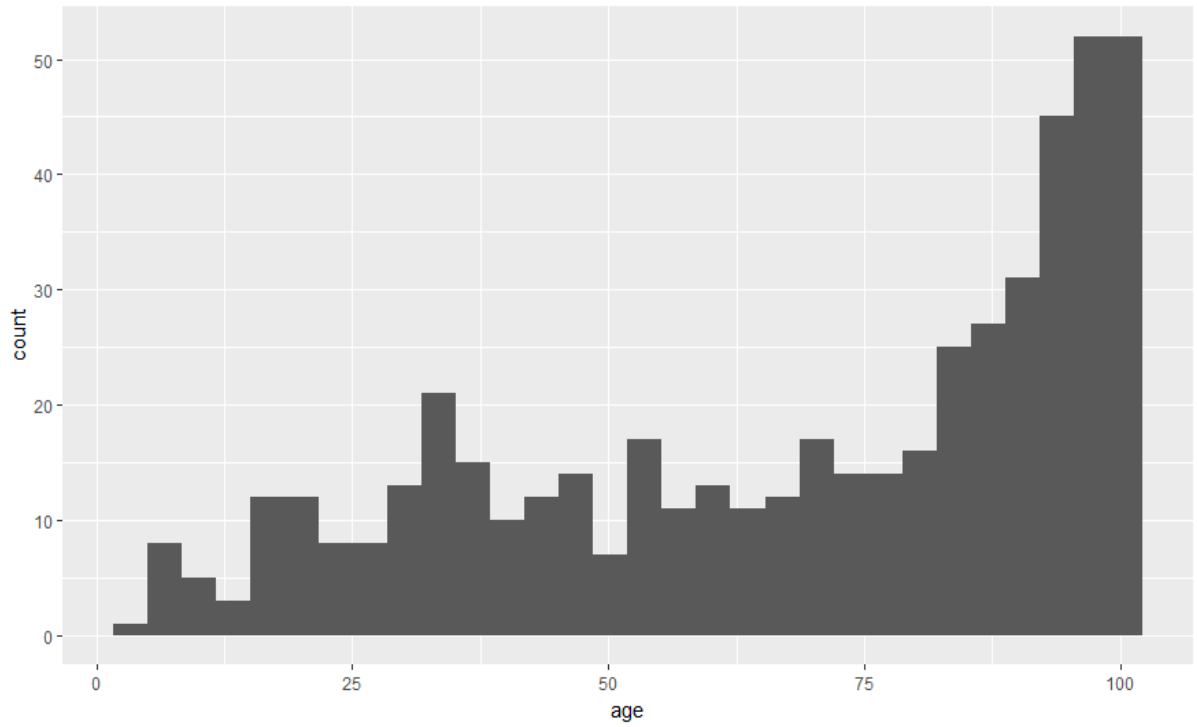
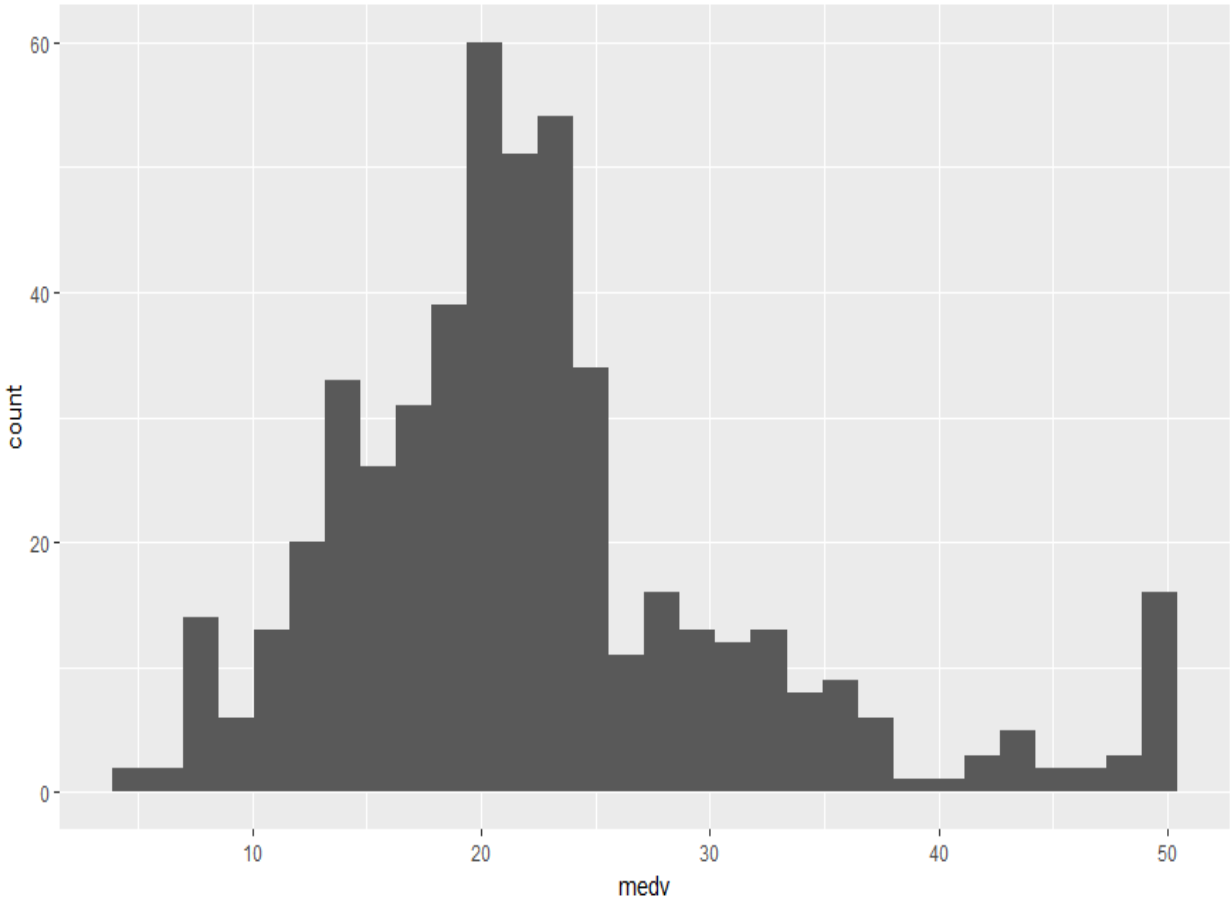
2.)

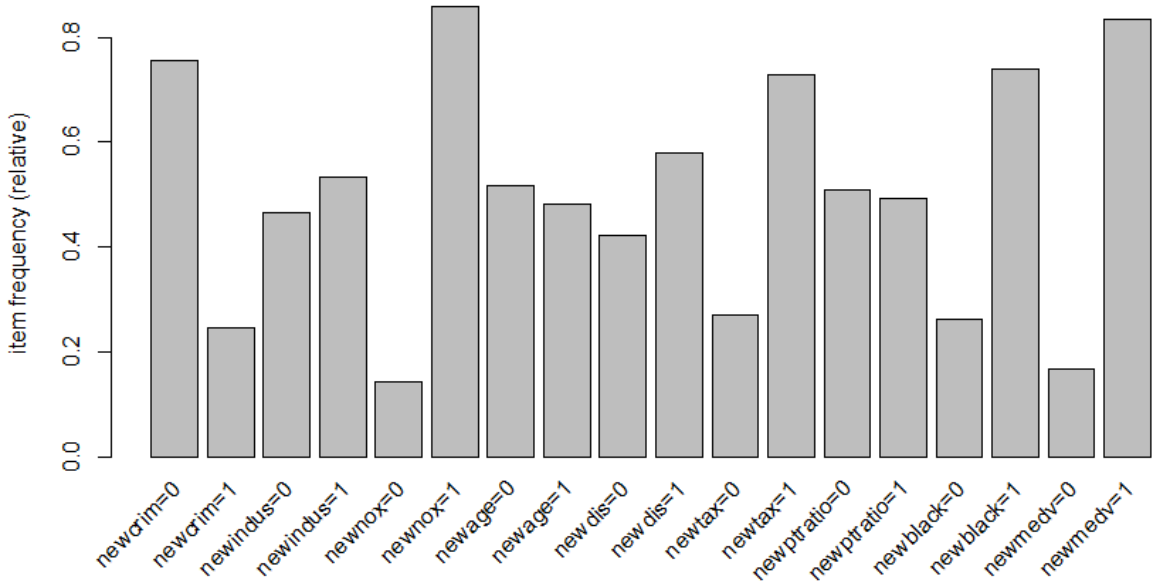
a.) Histograms are plotted for different variables in the dataset and can be observed in the file:




count
50
40
30
20
10
0
0
25
50
75
100
age


count
60
40
20
0
10
20
30
40
50
medv

Following is the frequency plot obtained:



From the graph above we can see that newnox and newmedv have more frequency.

c.)

```
    lhs                                                rhs          support    confidence lift     count
[1] {newdis=0,newblack=1,newmedv=0}                 => {newcrim=1} 0.05533597 0.8        3.264516 28
[2] {newage=1,newdis=0,newblack=1,newmedv=0}        => {newcrim=1} 0.05533597 0.8        3.264516 28
[3] {newdis=0,newptratio=1,newblack=1,newmedv=0}    => {newcrim=1} 0.05533597 0.8        3.264516 28
[4] {newindus=1,newdis=0,newblack=1,newmedv=0}      => {newcrim=1} 0.05533597 0.8        3.264516 28
[5] {newdis=0,newtax=1,newblack=1,newmedv=0}        => {newcrim=1} 0.05533597 0.8        3.264516 28
[6] {newnox=1,newdis=0,newblack=1,newmedv=0}        => {newcrim=1} 0.05533597 0.8        3.264516 28
```

From this it can be observed that for the crime to be low, it is better to opt for a place that is far away from the employment center.

```
    lhs                                rhs          support   confidence lift     count
[1] {newnox=0}                      => {newdis=1} 0.1422925 1          1.726962 72
[2] {newnox=0,newtax=0}             => {newdis=1} 0.1106719 1          1.726962 56
[3] {newindus=0,newnox=0}           => {newdis=1} 0.1422925 1          1.726962 72
[4] {newnox=0,newptratio=0}         => {newdis=1} 0.1106719 1          1.726962 56
[5] {newnox=0,newage=0}             => {newdis=1} 0.1422925 1          1.726962 72
[6] {newnox=0,newblack=1}           => {newdis=1} 0.1146245 1          1.726962 58
```

d.)

```
    lhs                                    rhs             support    confidence lift     count
[1] {newcrim=1,newblack=1,newmedv=0}   => {newptratio=1} 0.05928854 1          2.032129 30
[2] {newdis=0,newblack=1,newmedv=0}    => {newptratio=1} 0.06916996 1          2.032129 35
[3] {newage=1,newblack=1,newmedv=0}    => {newptratio=1} 0.08300395 1          2.032129 42
[4] {newindus=1,newdis=1,newmedv=0}    => {newptratio=1} 0.05928854 1          2.032129 30
[5] {newindus=1,newblack=1,newmedv=0} => {newptratio=1} 0.11462451 1          2.032129 58
[6] {newdis=1,newtax=1,newmedv=0}      => {newptratio=1} 0.06916996 1          2.032129 35
```

It is better to opt for a place where the crime is less, black proportion is less and distance to employment center is less.

Regression Model:

```
Call:
lm(formula = newptratio ~ ., data = subset)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97694 -0.07909  0.07719  0.32456  0.63415

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.30758    0.09197   3.344 0.000887 ***
newcrim      0.12589    0.04585   2.746 0.006255 **
newindus    -0.12126    0.05040  -2.406 0.016505 *
newnox      -0.15629    0.05520  -2.831 0.004823 **
newage       0.12954    0.04666   2.776 0.005706 **
newdis       0.07302    0.04677   1.561 0.119119
newtax       0.69147    0.05135  13.466  < 2e-16 ***
newblack     0.05416    0.03792   1.428 0.153800
newmedv     -0.35566    0.04769  -7.457 3.96e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3696 on 497 degrees of freedom
Multiple R-squared:  0.4631,    Adjusted R-squared:  0.4544
F-statistic: 53.58 on 8 and 497 DF,  p-value: < 2.2e-16
```

Also regression follows the similar trend where crime is less, black proportion is less and distance to employment center is less. Regression is generally preferred when the features are mostly non-categorical.