Lab4_EDA

Micheline Casey

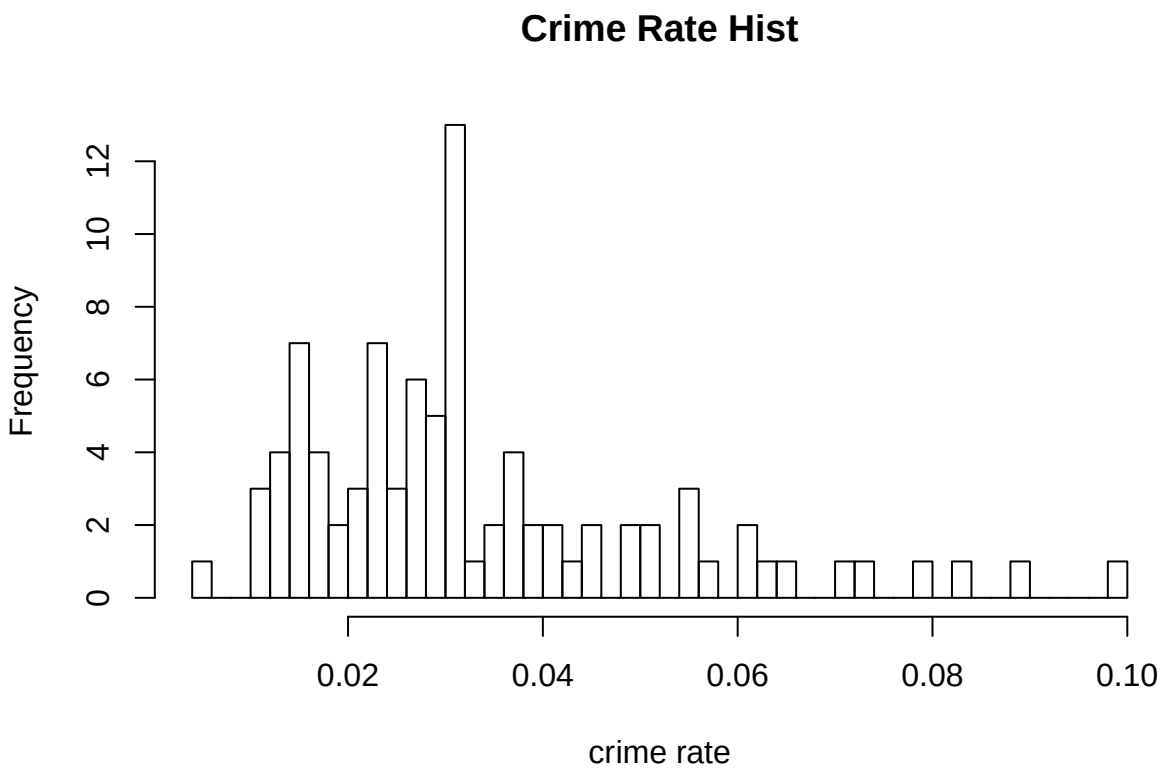
8/12/2017

```
df <- read.csv("crime.csv", stringsAsFactors = T)
#str(df)
#describe(df)
#unique(df)
summary(df)
```

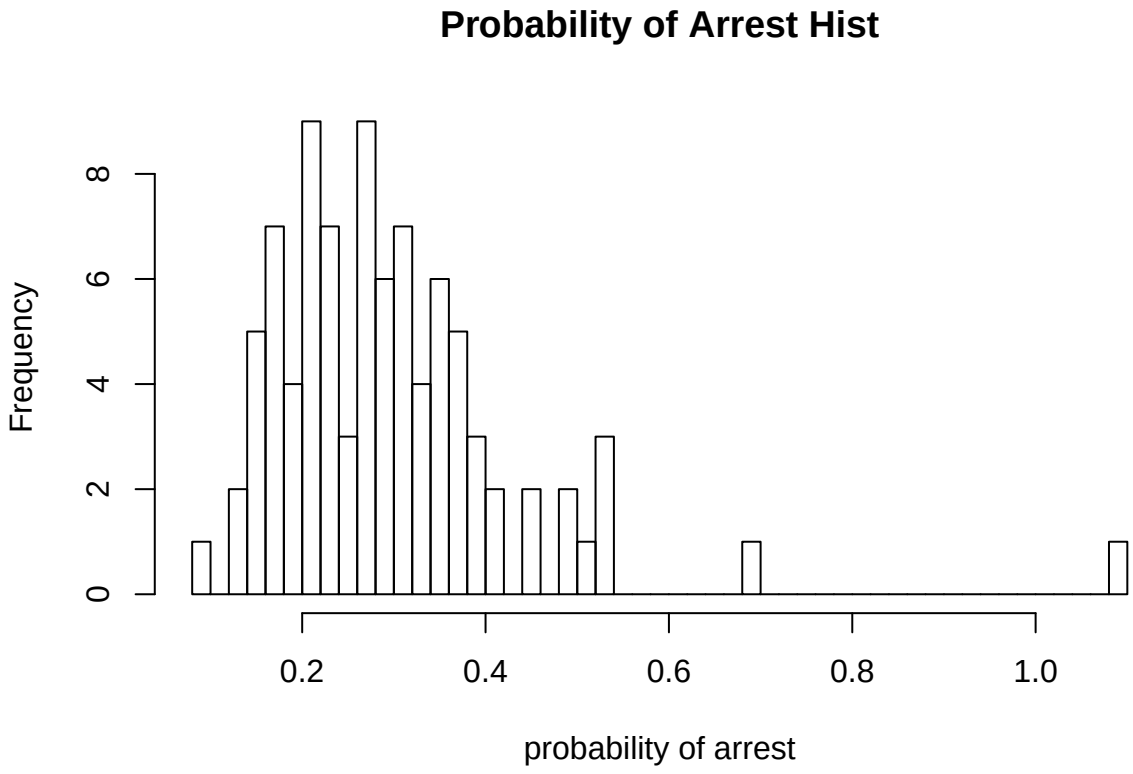
```
##        X             county           year        crmrte
##  Min.   : 1.00   Min.   :  1.0   Min.   :87   Min.   :0.005533
##  1st Qu.:23.25   1st Qu.: 51.5   1st Qu.:87   1st Qu.:0.020604
##  Median :45.50   Median :103.0   Median :87   Median :0.030002
##  Mean   :45.50   Mean   :100.6   Mean   :87   Mean   :0.033510
##  3rd Qu.:67.75   3rd Qu.:150.5   3rd Qu.:87   3rd Qu.:0.040249
##  Max.   :90.00   Max.   :197.0   Max.   :87   Max.   :0.098966
##      prbarr           prbconv           prbpris           avgsen
##  Min.   :0.09277   Min.   :0.06838   Min.   :0.1500   Min.   : 5.380
##  1st Qu.:0.20495   1st Qu.:0.34422   1st Qu.:0.3642   1st Qu.: 7.375
##  Median :0.27146   Median :0.45170   Median :0.4222   Median : 9.110
##  Mean   :0.29524   Mean   :0.55086   Mean   :0.4106   Mean   : 9.689
##  3rd Qu.:0.34487   3rd Qu.:0.58513   3rd Qu.:0.4576   3rd Qu.:11.465
##  Max.   :1.09091   Max.   :2.12121   Max.   :0.6000   Max.   :20.700
##      polpc              density           taxpc             west
##  Min.   :0.0007459   Min.   :0.2034   Min.   : 25.69   Min.   :0.0000
##  1st Qu.:0.0012378   1st Qu.:0.5472   1st Qu.: 30.73   1st Qu.:0.0000
##  Median :0.0014897   Median :0.9792   Median : 34.92   Median :0.0000
##  Mean   :0.0017080   Mean   :1.4379   Mean   : 38.16   Mean   :0.2333
##  3rd Qu.:0.0018856   3rd Qu.:1.5693   3rd Qu.: 41.01   3rd Qu.:0.0000
##  Max.   :0.0090543   Max.   :8.8277   Max.   :119.76   Max.   :1.0000
##     central            urban            pctmin80           wcon
##  Min.   :0.0000   Min.   :0.00000   Min.   : 1.284   Min.   :193.6
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:10.024   1st Qu.:250.8
##  Median :0.0000   Median :0.00000   Median :24.852   Median :281.2
##  Mean   :0.3778   Mean   :0.08889   Mean   :25.713   Mean   :285.4
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:38.183   3rd Qu.:315.0
##  Max.   :1.0000   Max.   :1.00000   Max.   :64.348   Max.   :436.8
##       wtuc             wtrd             wfir             wser
##  Min.   :187.6   Min.   :154.2   Min.   :170.9   Min.   : 133.0
##  1st Qu.:374.3   1st Qu.:190.7   1st Qu.:285.6   1st Qu.: 229.3
##  Median :404.8   Median :203.0   Median :317.1   Median : 253.1
##  Mean   :410.9   Mean   :210.9   Mean   :321.6   Mean   : 275.3
##  3rd Qu.:440.7   3rd Qu.:224.3   3rd Qu.:342.6   3rd Qu.: 277.6
##  Max.   :613.2   Max.   :354.7   Max.   :509.5   Max.   :2177.1
##       wmfg             wfed             wsta             wloc
##  Min.   :157.4   Min.   :326.1   Min.   :258.3   Min.   :239.2
##  1st Qu.:288.6   1st Qu.:398.8   1st Qu.:329.3   1st Qu.:297.2
##  Median :321.1   Median :448.9   Median :358.4   Median :307.6
##  Mean   :336.0   Mean   :442.6   Mean   :357.7   Mean   :312.3
##  3rd Qu.:359.9   3rd Qu.:478.3   3rd Qu.:383.2   3rd Qu.:328.8
```

```
##  Max.   :646.9   Max.   :598.0   Max.   :499.6   Max.   :388.1
##       mix             pctymle
##  Min.   :0.01961   Min.   :0.06216
##  1st Qu.:0.08060   1st Qu.:0.07437
##  Median :0.10095   Median :0.07770
##  Mean   :0.12905   Mean   :0.08403
##  3rd Qu.:0.15206   3rd Qu.:0.08352
##  Max.   :0.46512   Max.   :0.24871
```

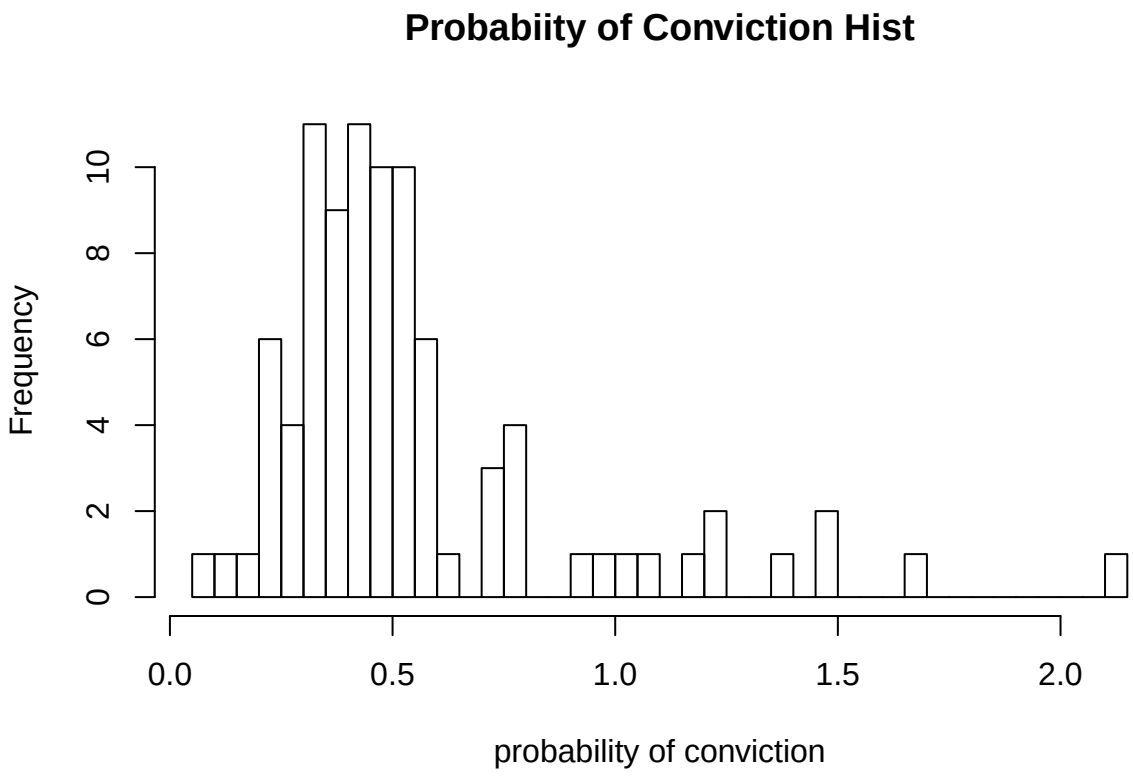
```
hist(df$crmrte, breaks = 50, xlab = "crime rate", ylab = "Frequency", main = "Crime Rate Hist")
```



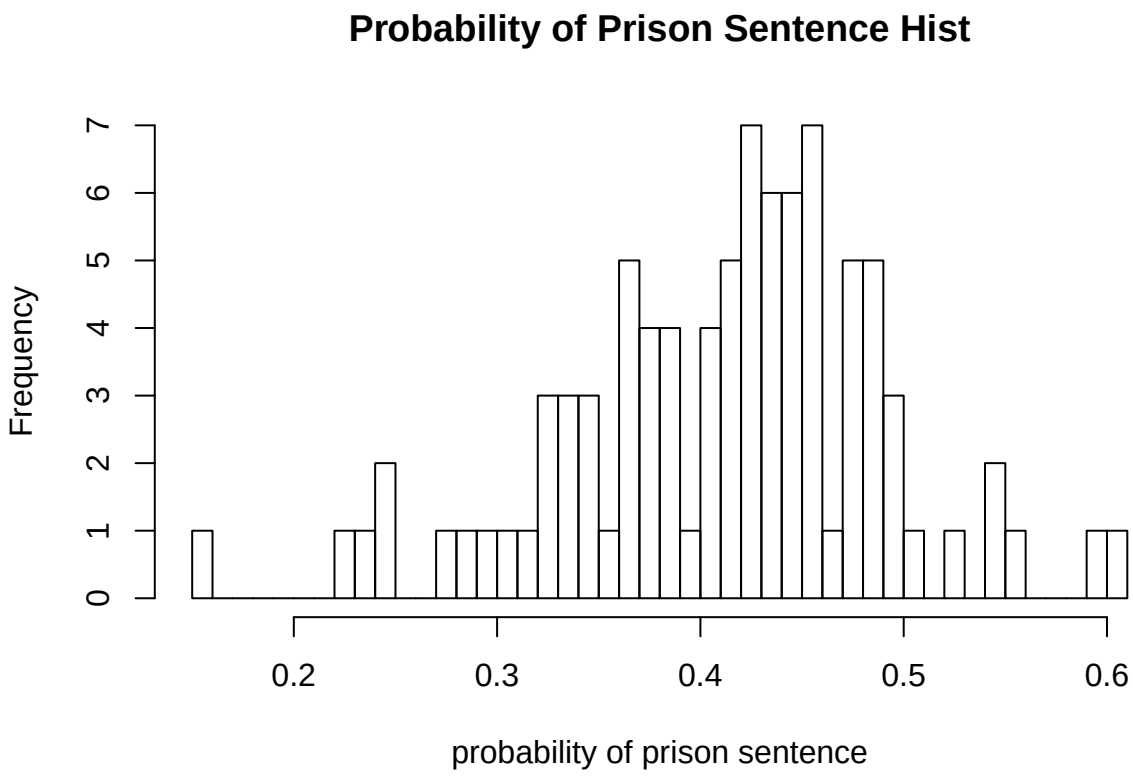

```
hist(df$prbarr, breaks = 50, xlab = "probability of arrest", ylab = "Frequency", main = "Probability of
```



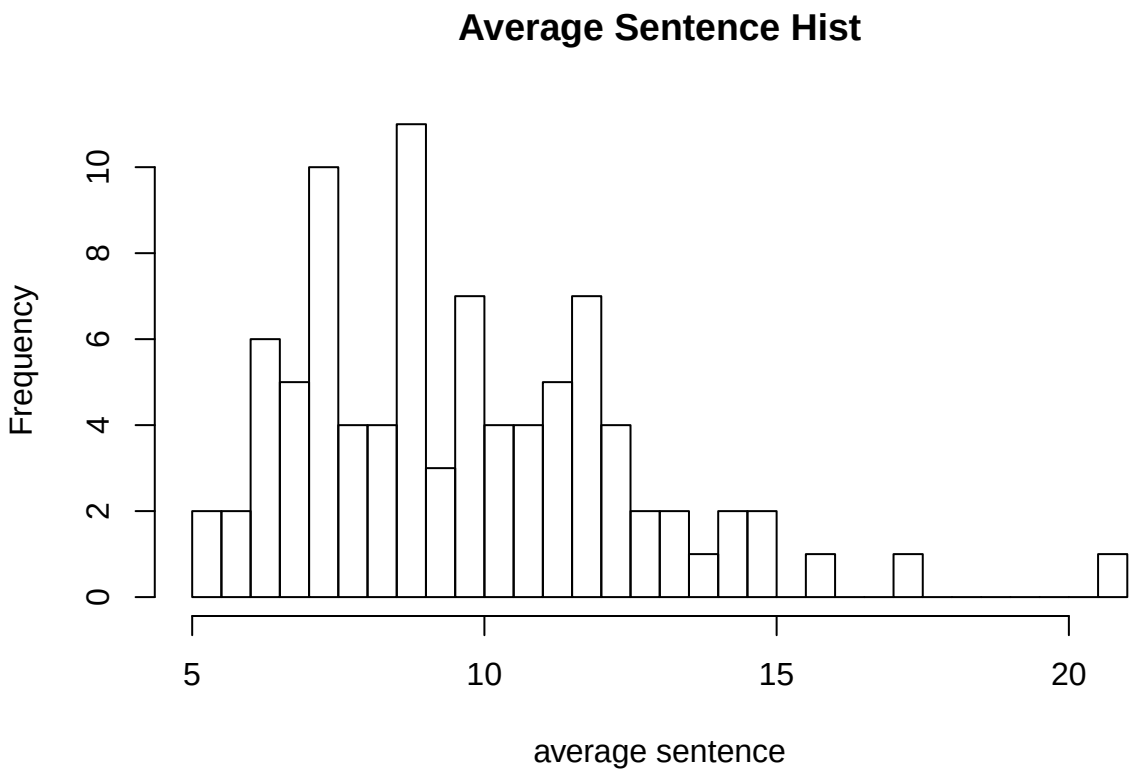

```
hist(df$prbconv, breaks = 50, xlab = "probability of conviction", ylab = "Frequency", main = "Probabiity
```



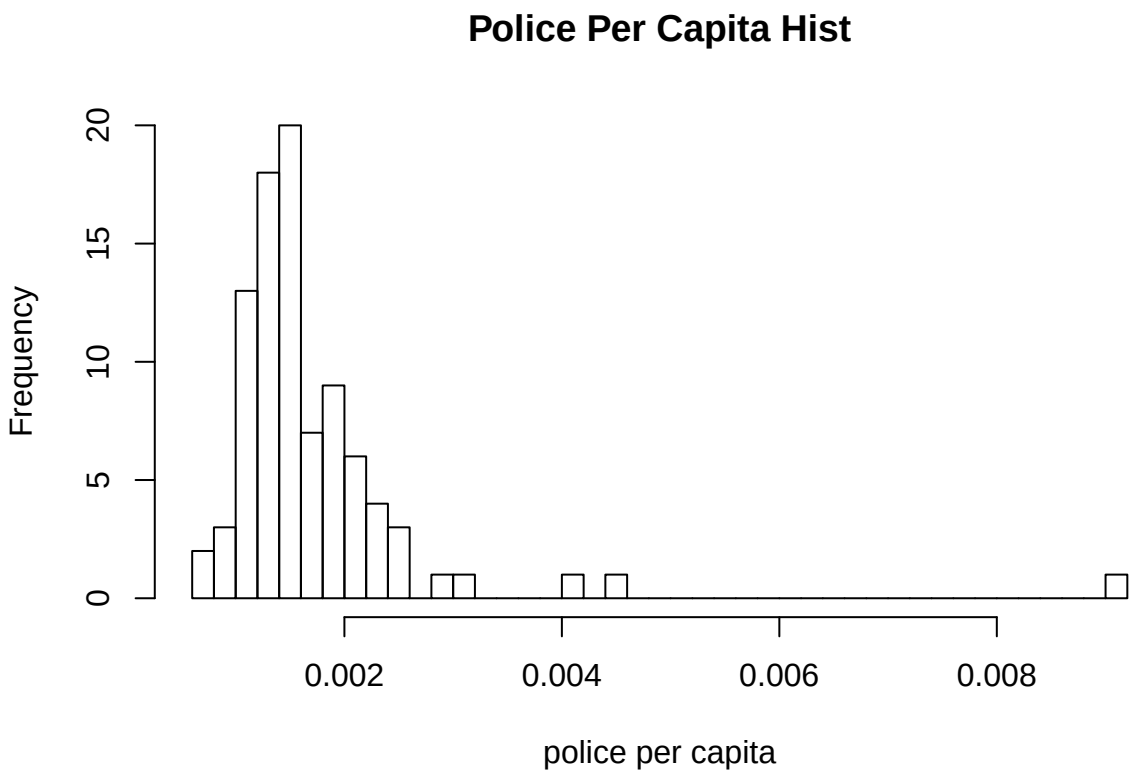

```
hist(df$prbpris, breaks = 50, xlab = "probability of prison sentence", ylab = "Frequency", main = "Proba
```



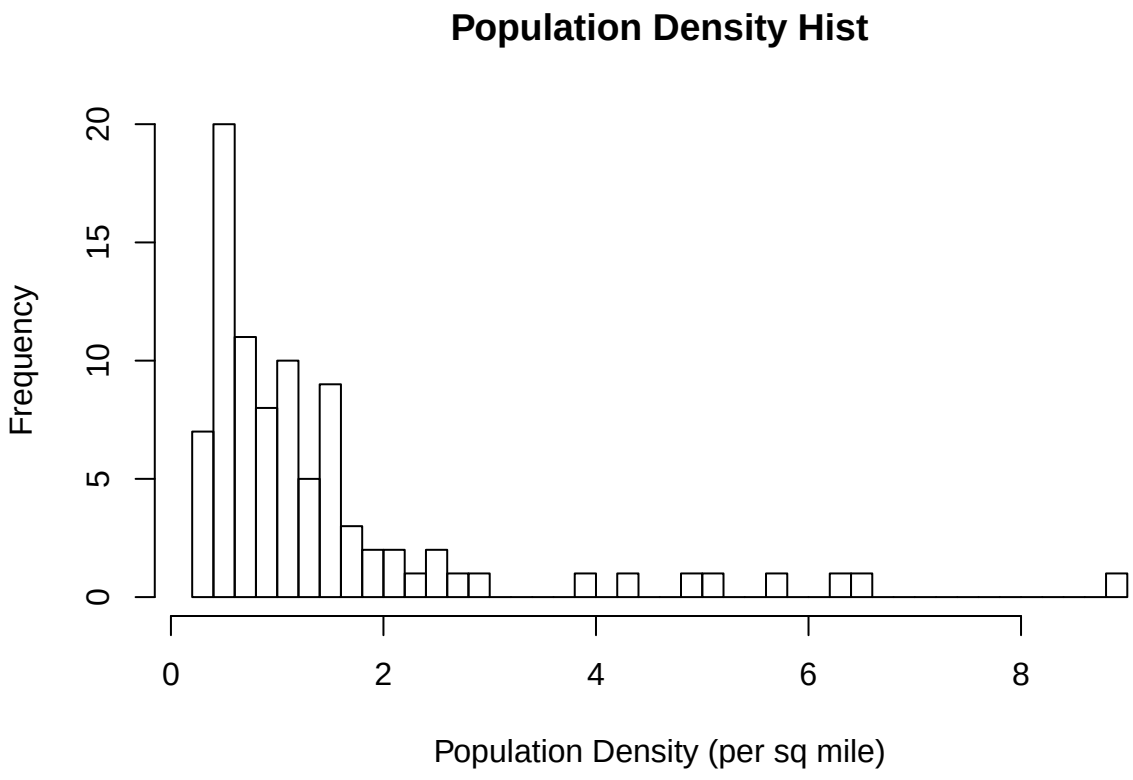

```
hist(df$avgsen, breaks = 50, xlab = "average sentence", ylab = "Frequency", main = "Average Sentence His
```



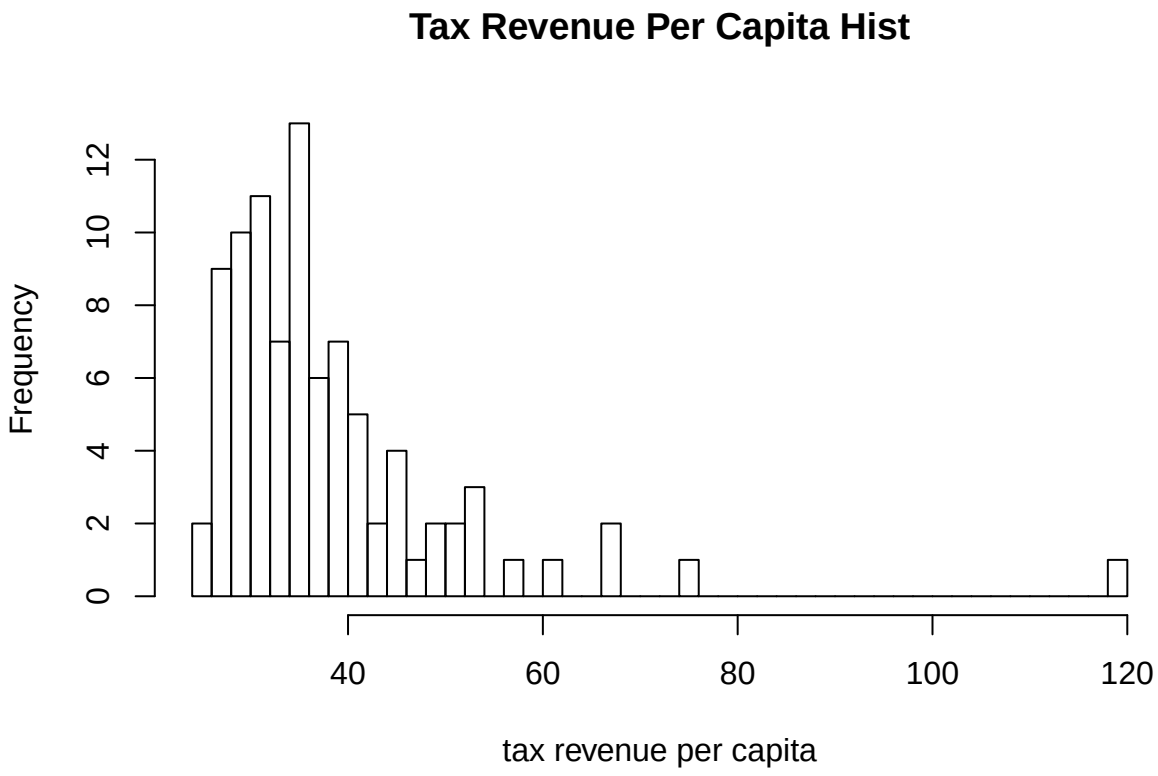

```
hist(df$polpc, breaks = 50, xlab = "police per capita", ylab = "Frequency", main = "Police Per Capita Hi
```



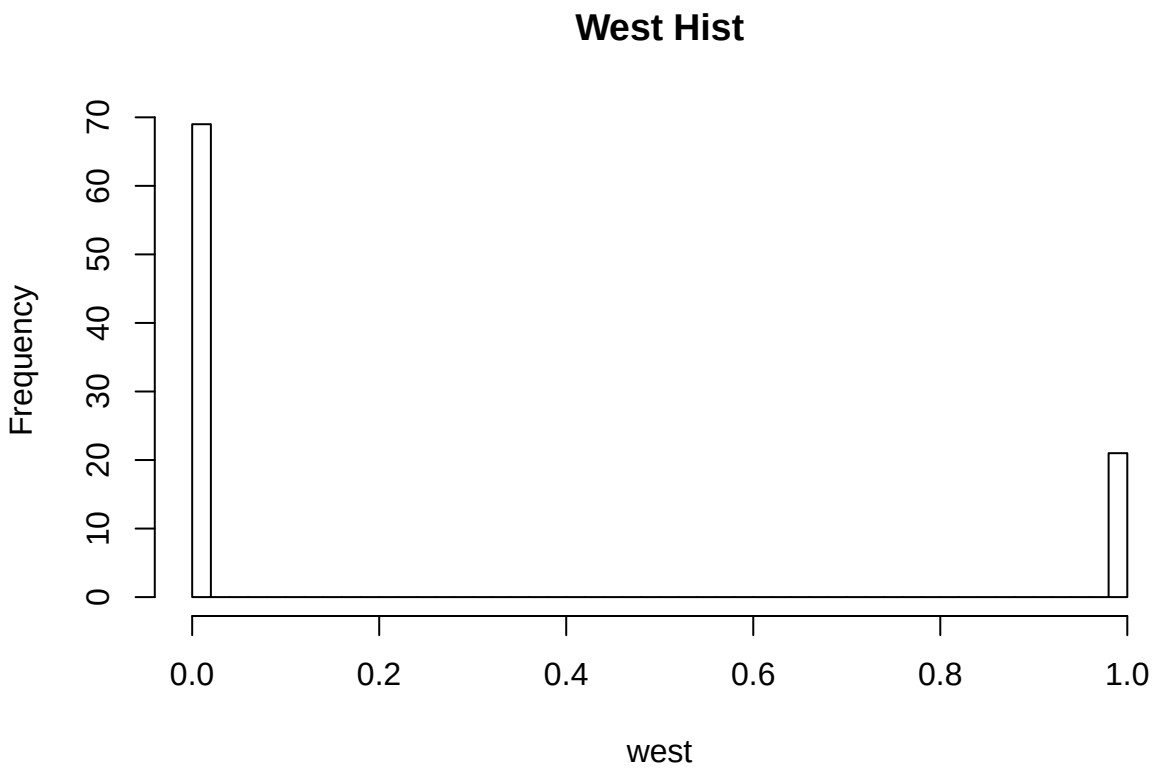

```
hist(df$density, breaks = 50, xlab = "Population Density (per sq mile)", ylab = "Frequency", main = "Po
```



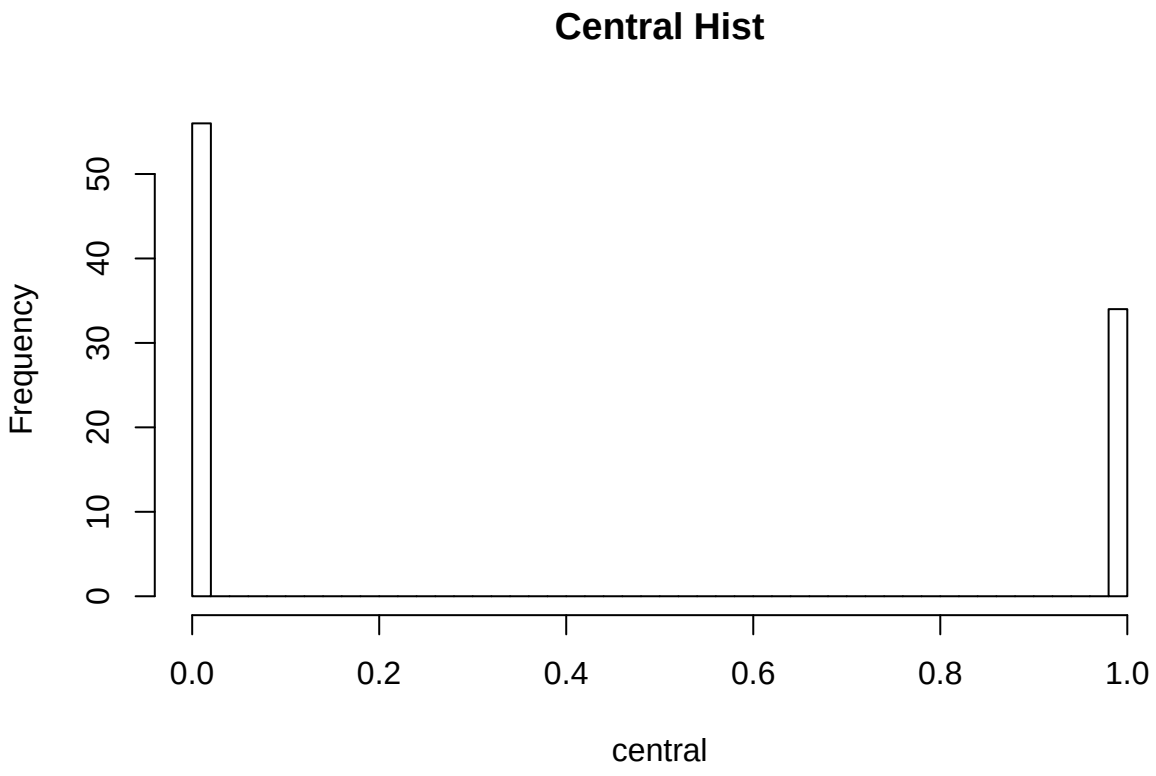

```
hist(df$taxpc, breaks = 50, xlab = "tax revenue per capita", ylab = "Frequency", main = "Tax Revenue Pe
```



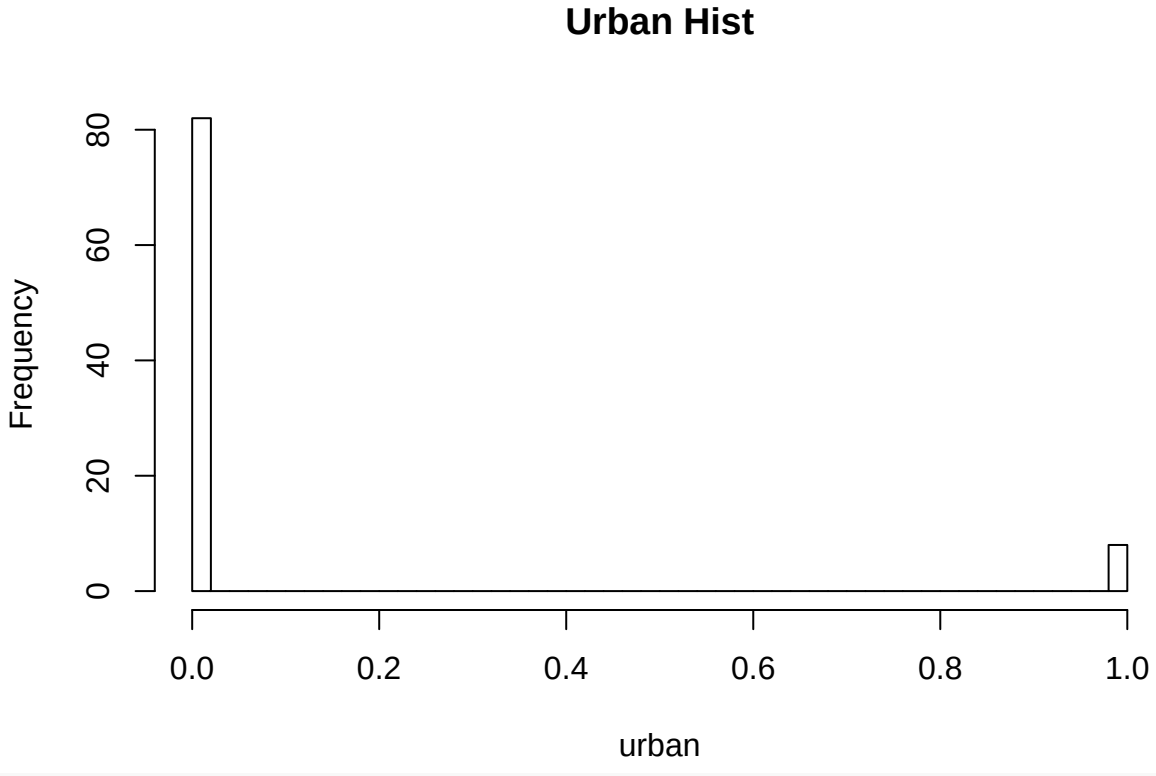

```r
hist(df$west, breaks = 50, xlab = "west", ylab = "Frequency", main = "West Hist")
```



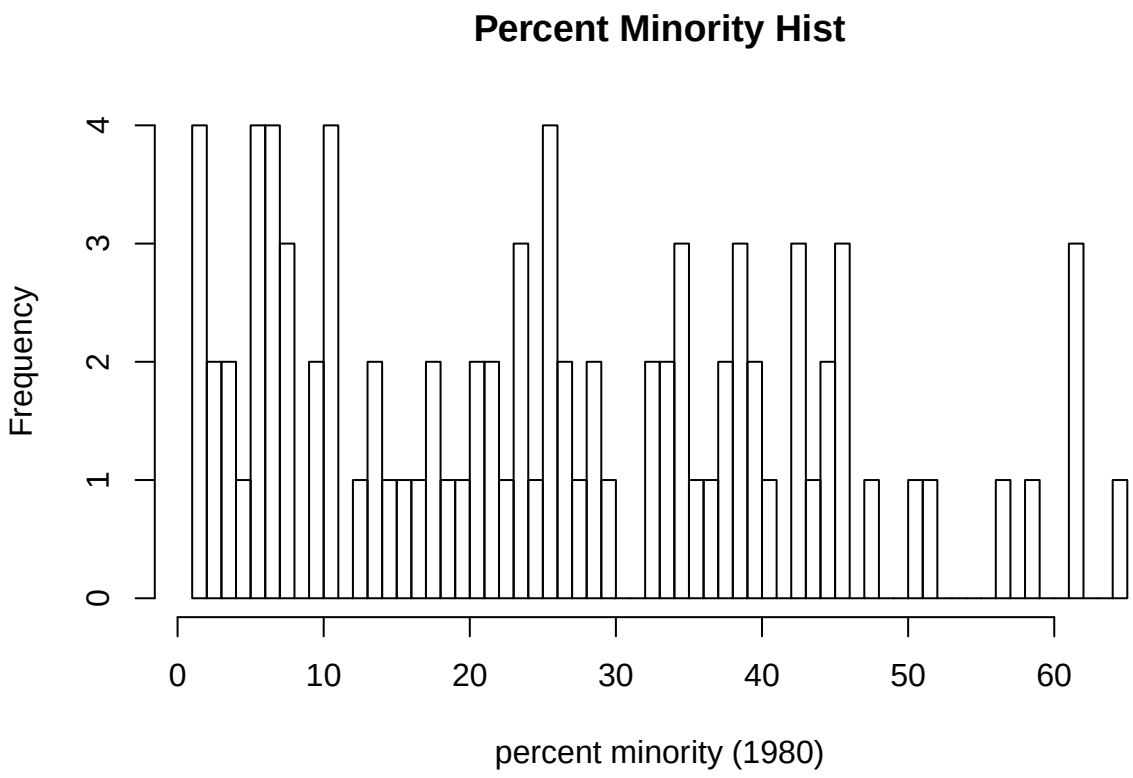

```r
hist(df$central, breaks = 50, xlab = "central", ylab = "Frequency", main = "Central Hist")
```



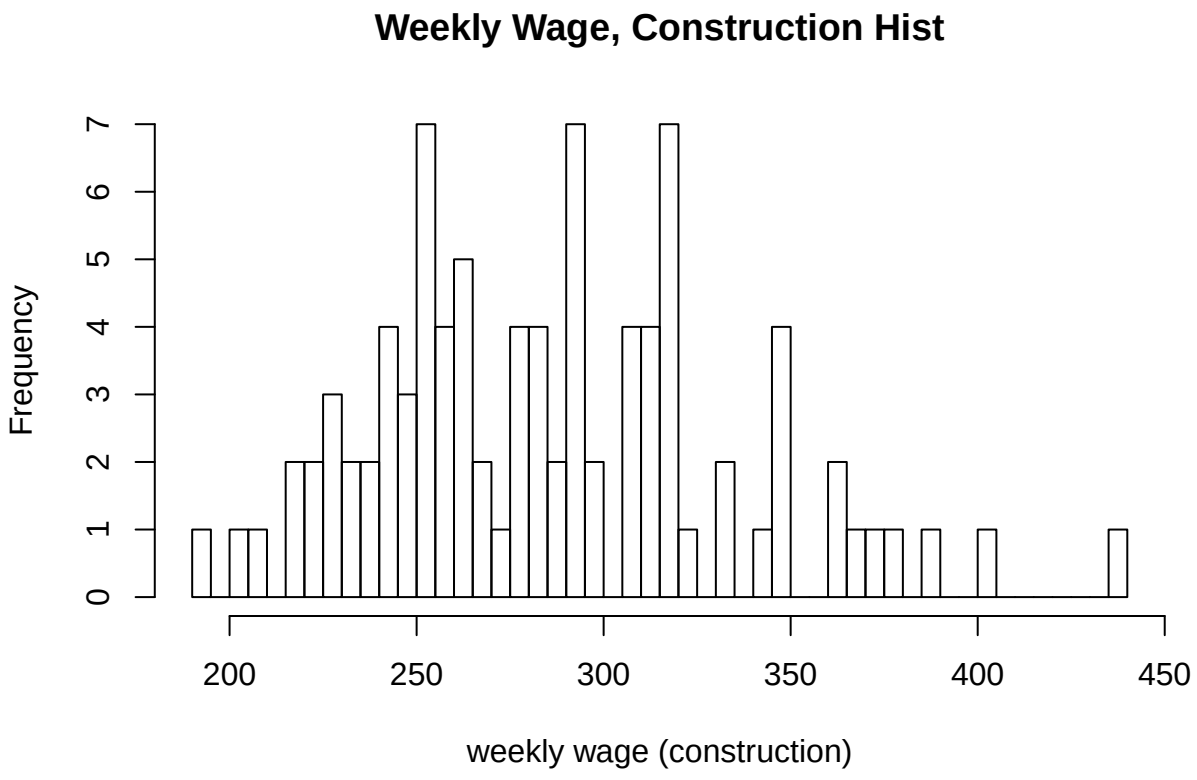

```
hist(df$urban, breaks = 50, xlab = "urban", ylab = "Frequency", main = "Urban Hist")
```



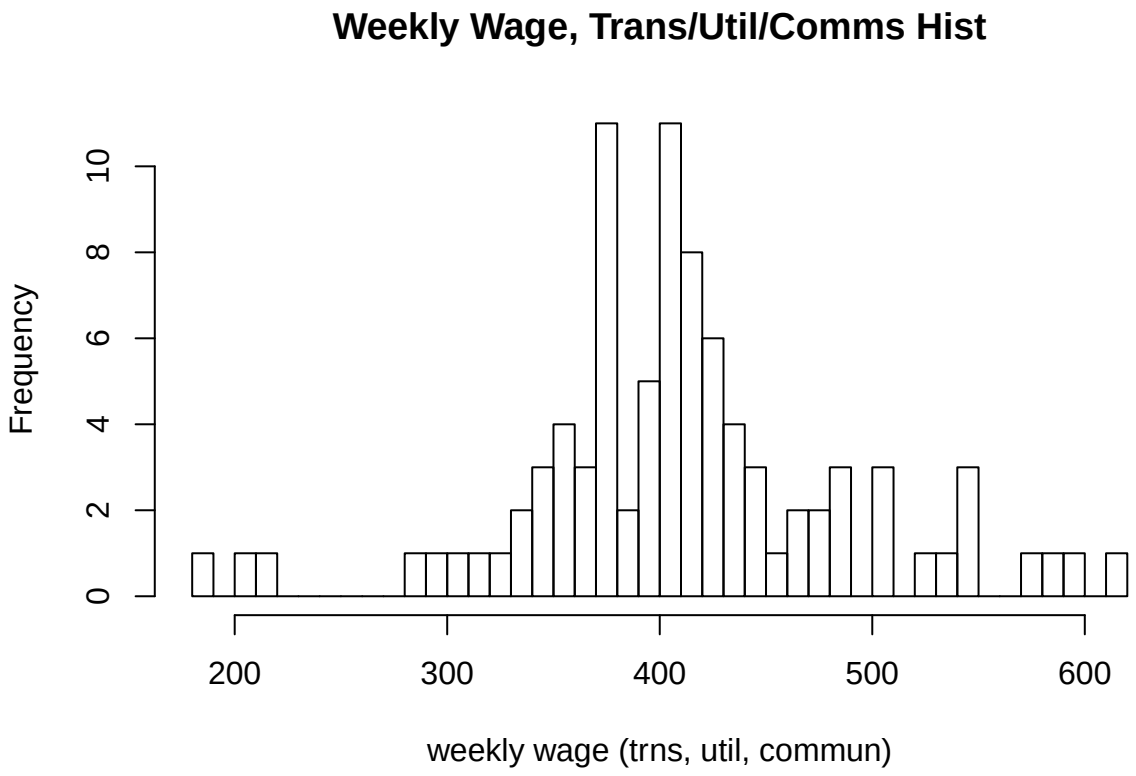

```
hist(df$pctmin80, breaks = 50, xlab = "percent minority (1980)", ylab = "Frequency", main = "Percent Mi
```



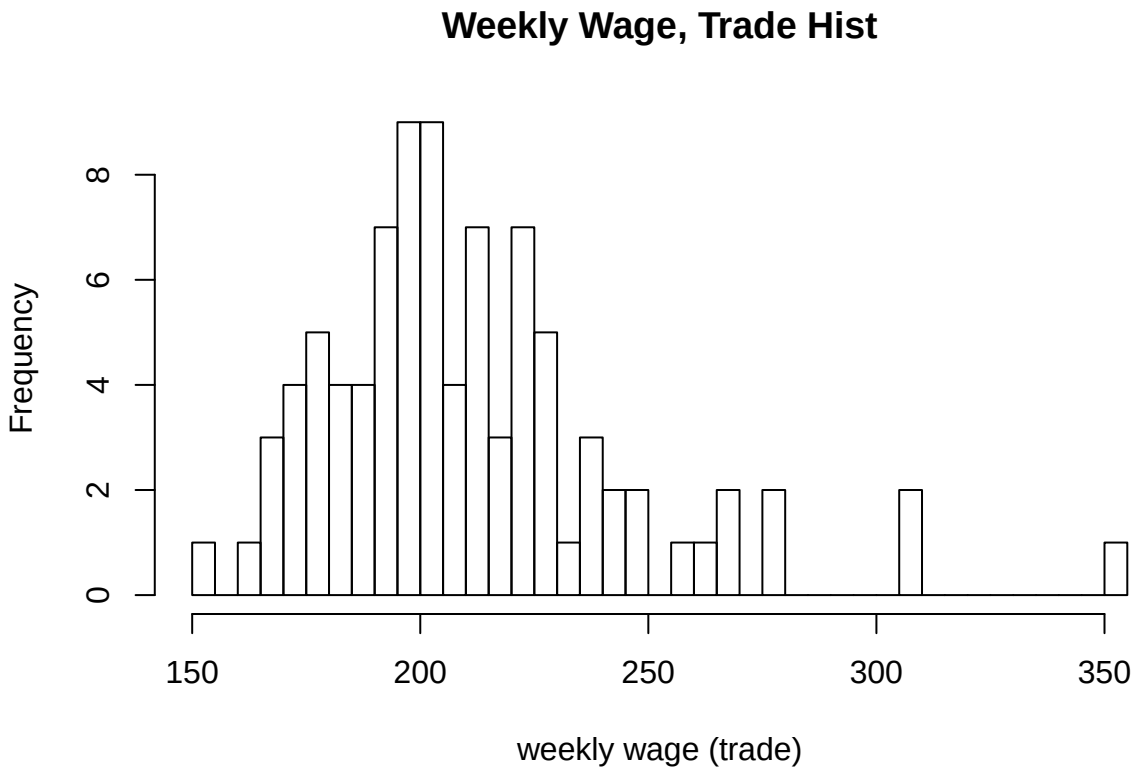

```
hist(df$wcon, breaks = 50, xlab = "weekly wage (construction)", ylab = "Frequency", main = "Weekly Wage
```



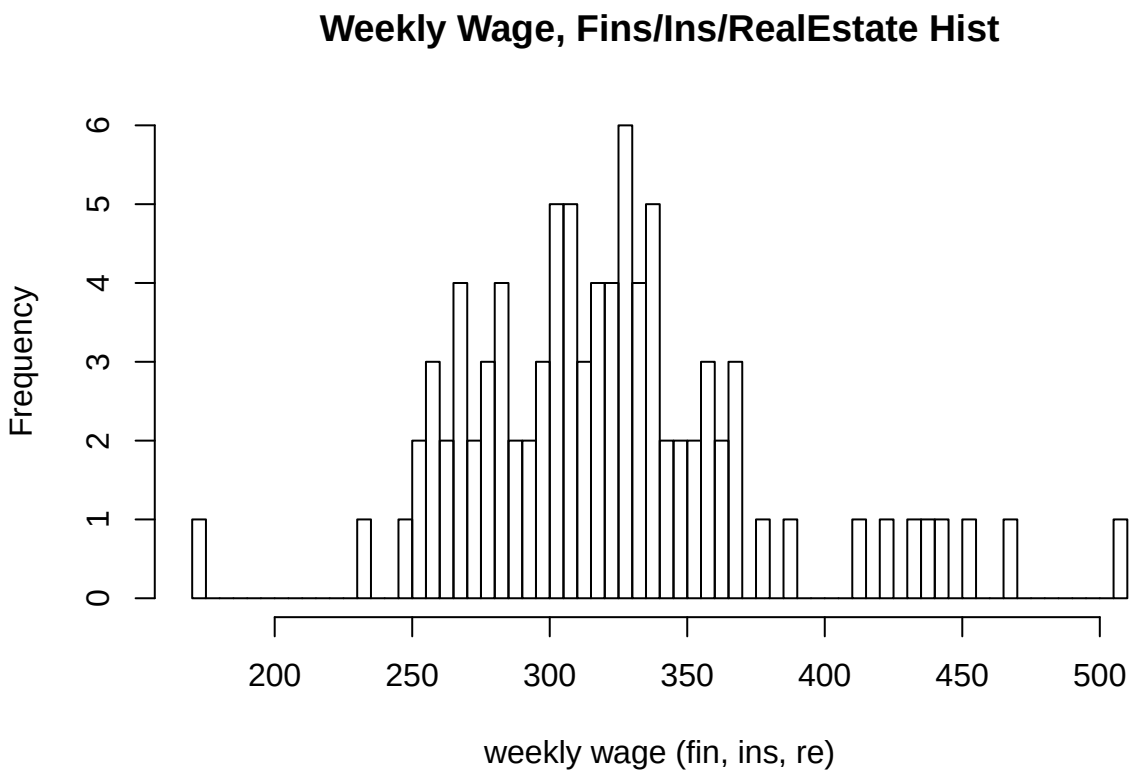

```
hist(df$wtuc, breaks = 50, xlab = "weekly wage (trns, util, commun)", ylab = "Frequency", main = "Weekl
```




```
hist(df$wtrd, breaks = 50, xlab = "weekly wage (trade)", ylab = "Frequency", main = "Weekly Wage, Trade
```



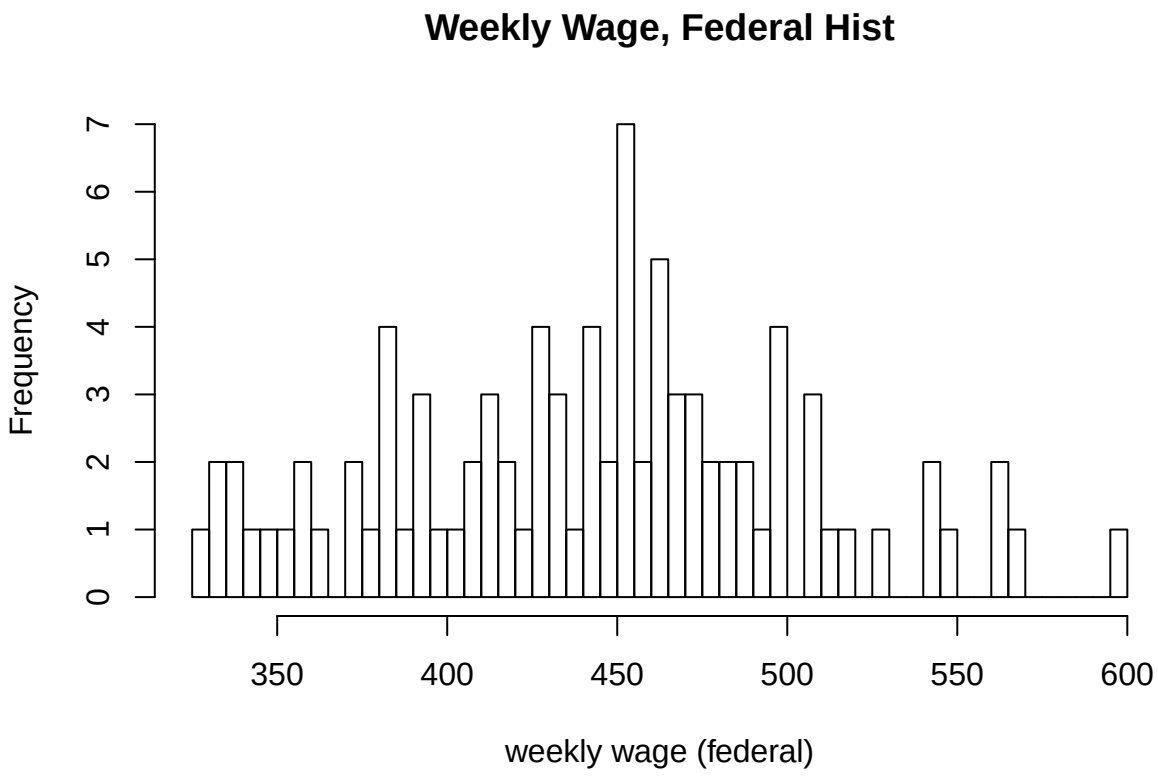

```
hist(df$wfir, breaks = 50, xlab = "weekly wage (fin, ins, re)", ylab = "Frequency", main = "Weekly Wage
```



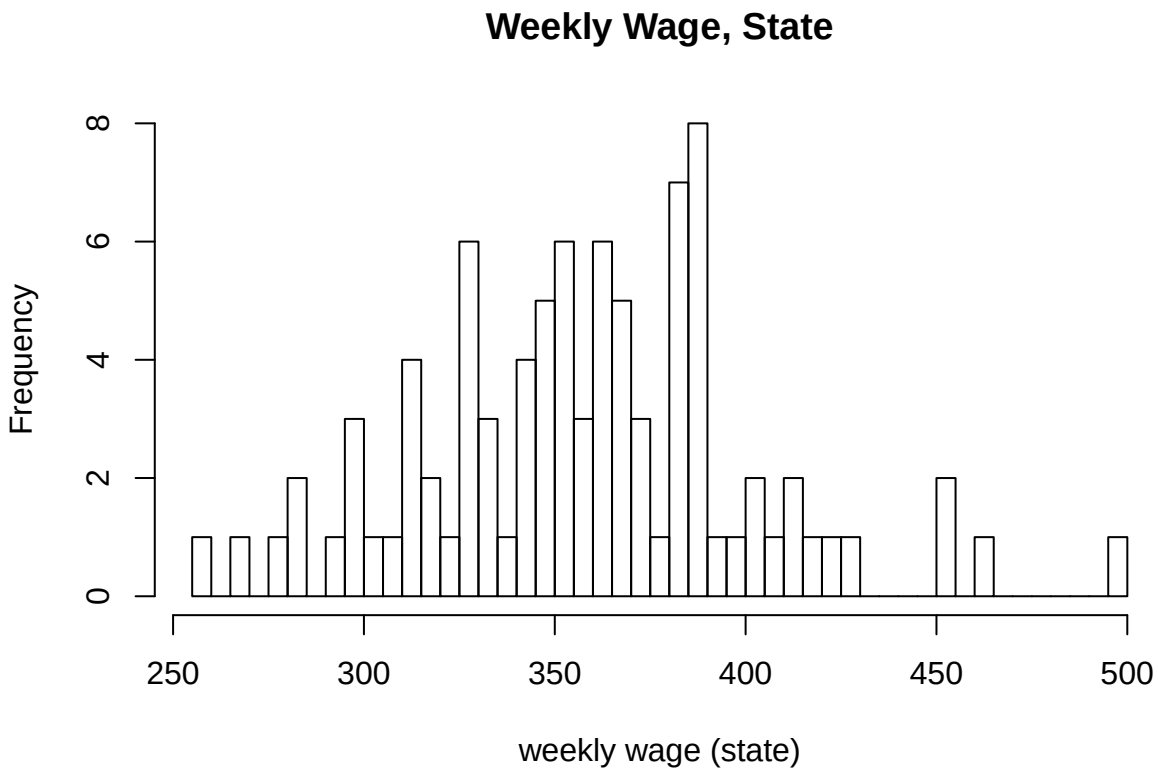

```
hist(df$wser, breaks = 50, xlab = "weekly wage (service)", ylab = "Frequency", main = "Weekly Wage, Ser
```


```
hist(df$wmfg, breaks = 50, xlab = "weekly wage (manufacturing)", ylab = "Frequency", main = "Weekly Wag
```

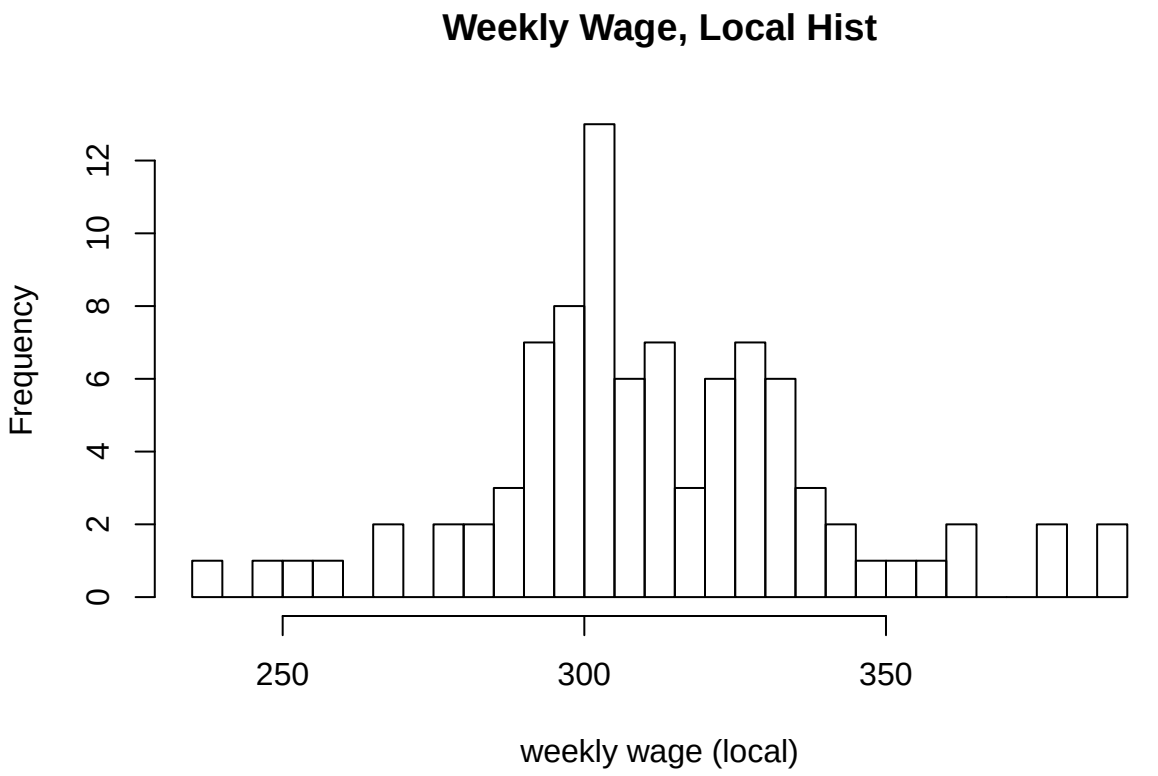

```
hist(df$wfed, breaks = 50, xlab = "weekly wage (federal)", ylab = "Frequency", main = "Weekly Wage, Fede
```



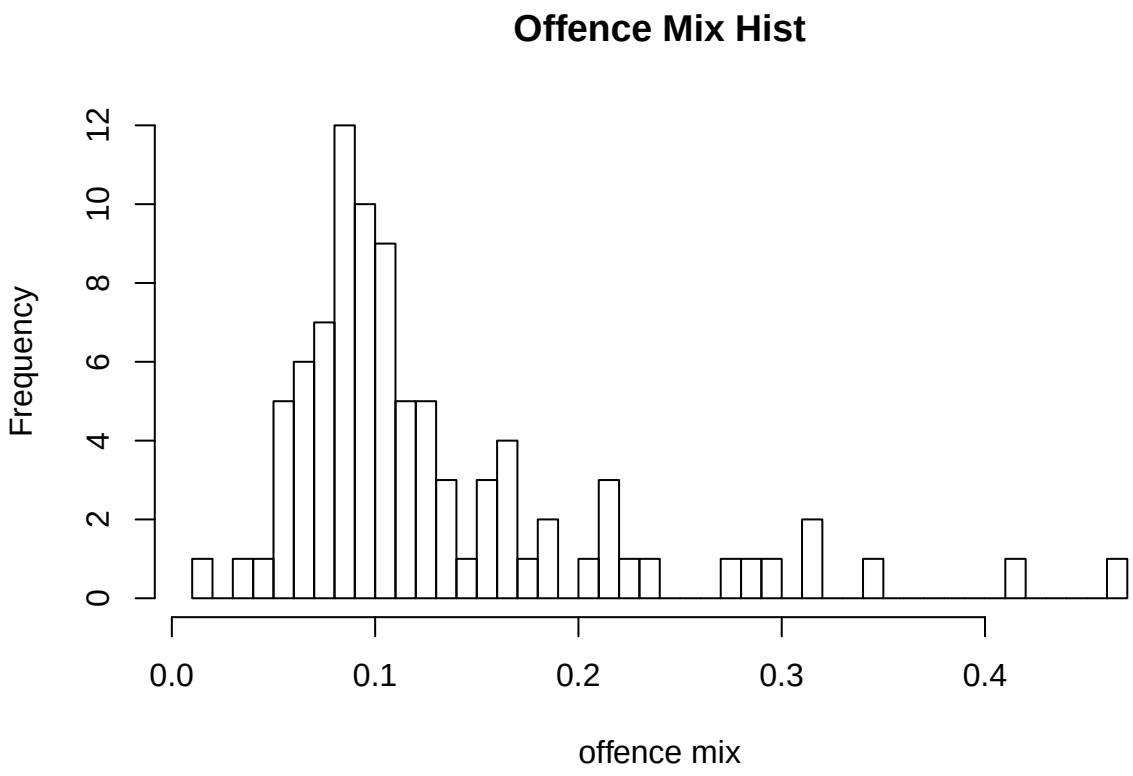

```
hist(df$wsta, breaks = 50, xlab = "weekly wage (state)", ylab = "Frequency", main = "Weekly Wage, State
```



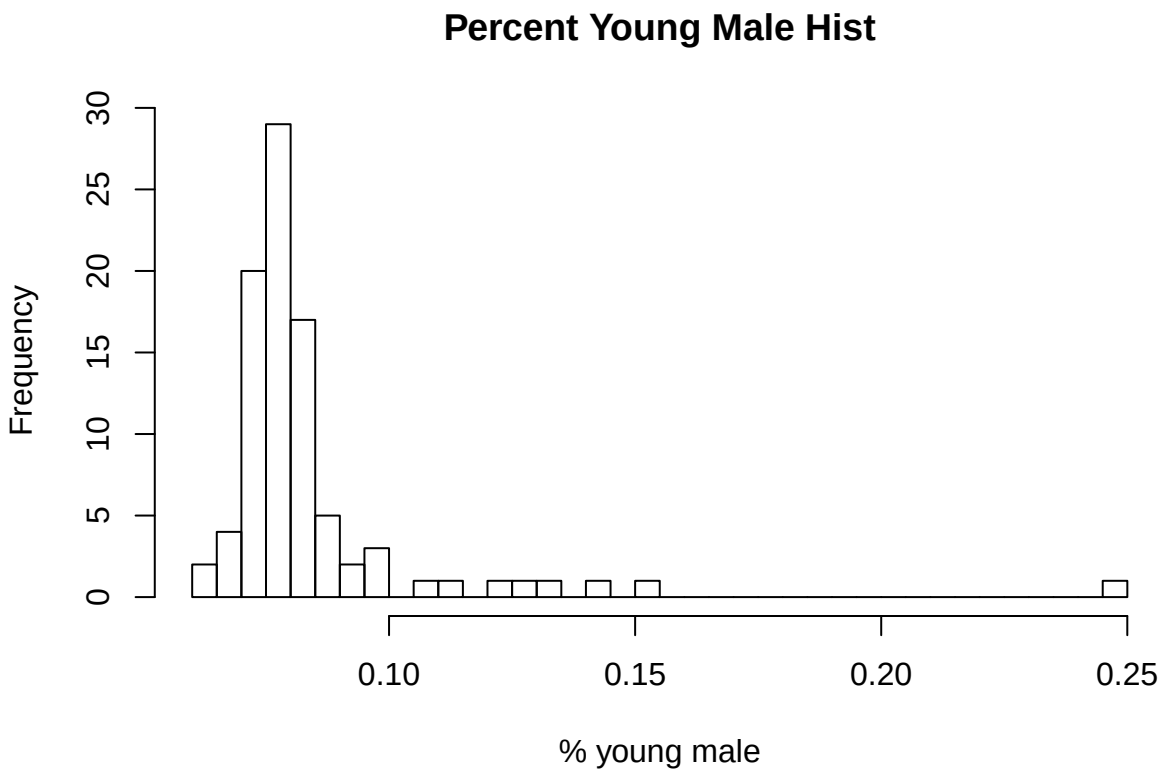

```
hist(df$wloc, breaks = 50, xlab = "weekly wage (local)", ylab = "Frequency", main = "Weekly Wage, Local
```



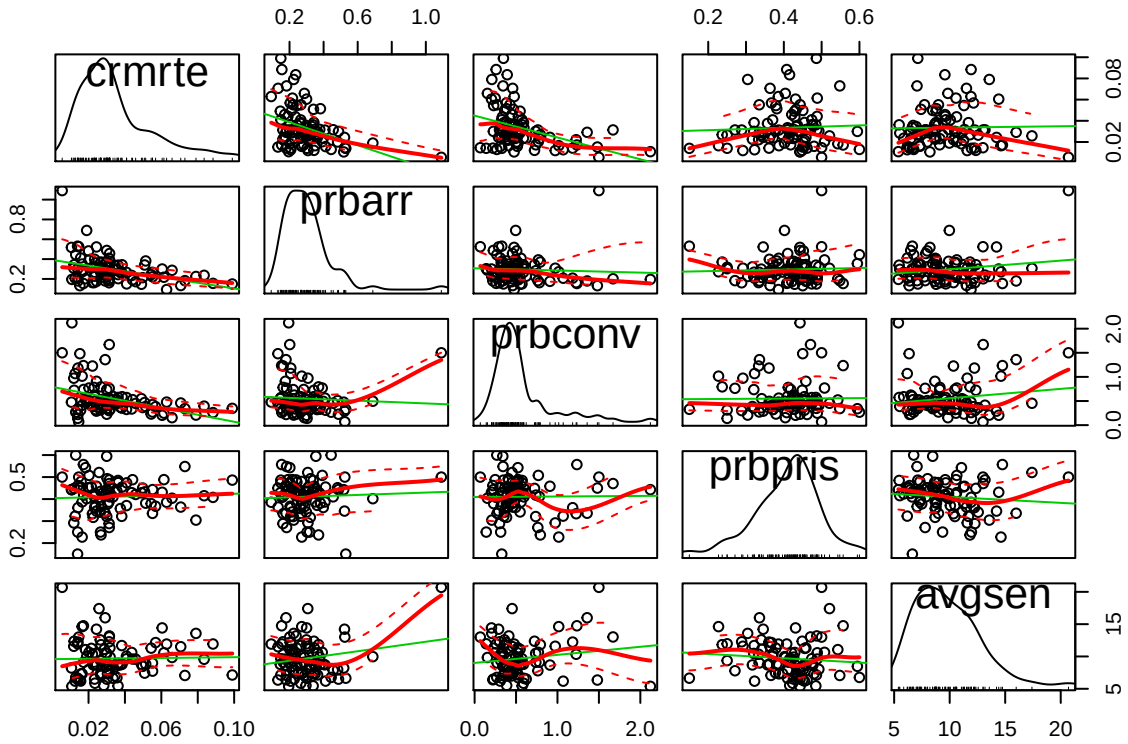

```
hist(df$mix, breaks = 50, xlab = "offence mix", ylab = "Frequency", main = "Offence Mix Hist")
```



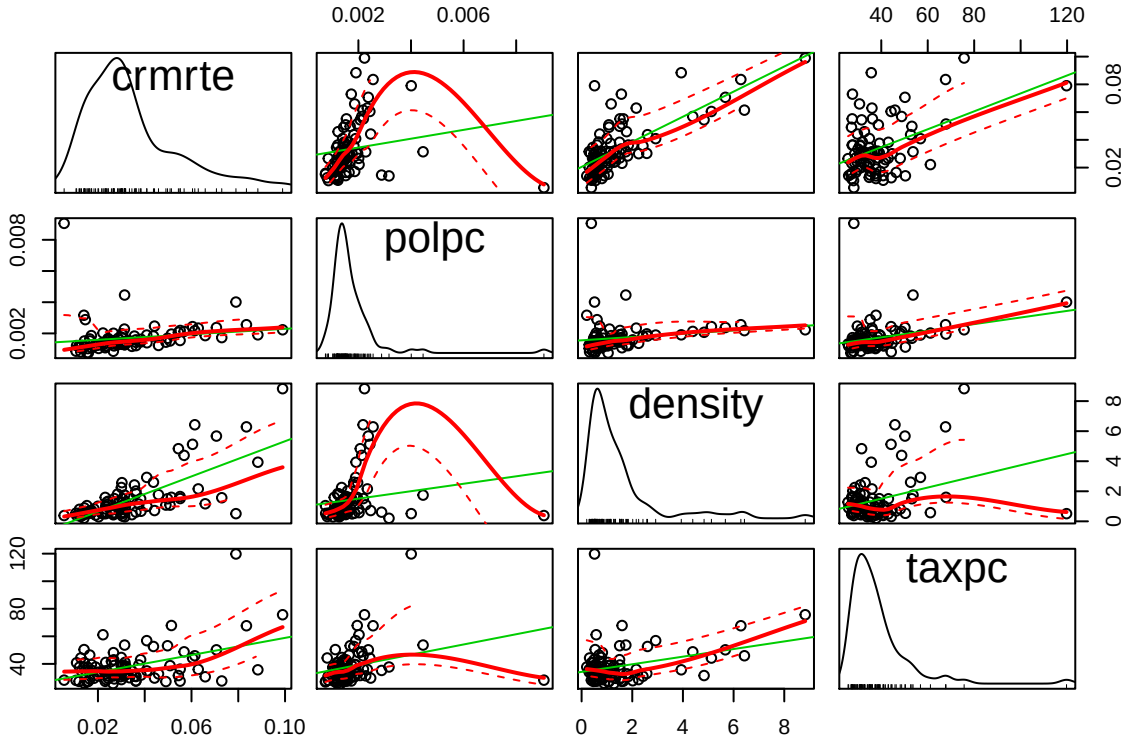

```
hist(df$pctymle, breaks = 50, xlab = "% young male", ylab = "Frequency", main = "Percent Young Male His
```



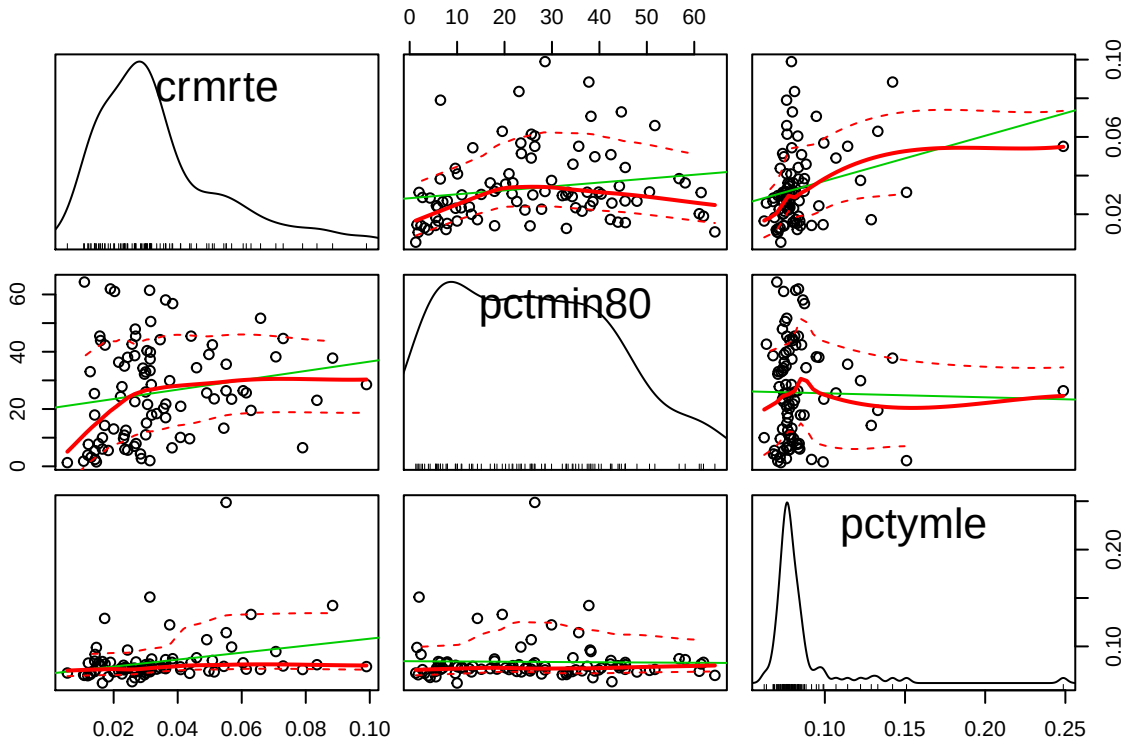

```
scatterplotMatrix(~crmrte + prbarr + prbconv + prbpris + avgsen, data= df )
```



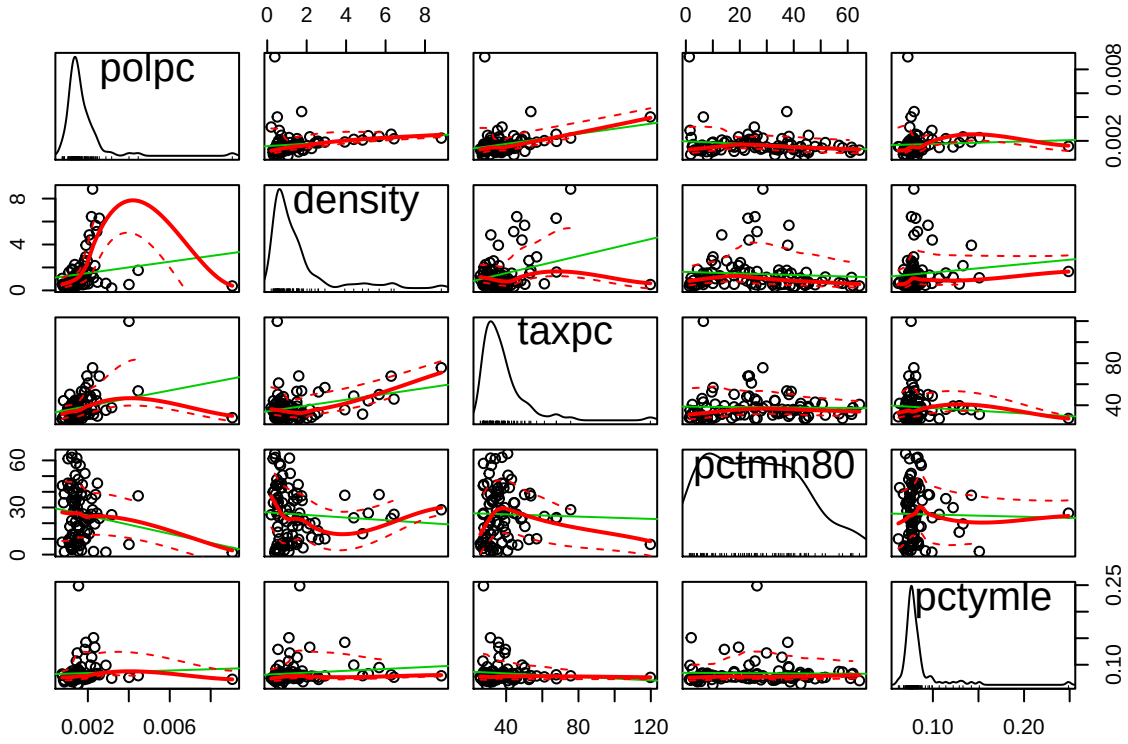

```
scatterplotMatrix(~crmrte + polpc + density + taxpc, data= df )
```




```
scatterplotMatrix(~crmrte + pctmin80 + pctymle, data= df )
```




```
scatterplotMatrix(~polpc + density + taxpc + pctmin80 + pctymle, data= df )
```



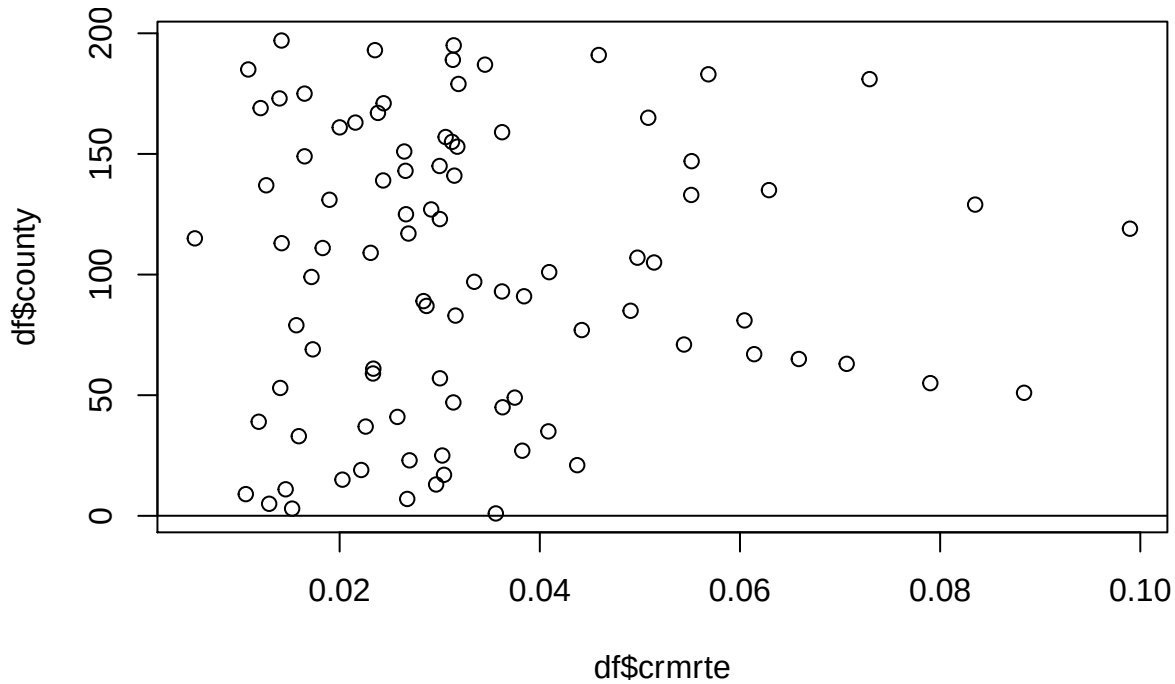

```
#df[df$crmrte >= 0.0950 | df$pctymle >= 0.20,]  #county:119,133
#df[df$prbarr >= 0.80, ]   #county:115
#df[df$prbconv >= 1.6, ]   #county:185, 195
#df[df$prbpris >= 0.575, ] #county:5,145
#df[df$avgsen >= 15, ]     #county:41,115,127
```

```
#df[df$polpc >= 0.004, ]    #county:55,115,195
#df[df$density >= 6, ]  #county:67,119,129
#df[df$taxpc >= 60, ]  #county:19,55,105,119,129
#df[df$taxpc <= 27, ] #county:3,151,197
#df[df$pctmin80 >= 60, ] #county:15,131,155,185
#df[df$pctmin80 <= 3, ] #county:9,11,87,115,189
#df[df$pctmin80 >=55 | df$pctmin80<= 10, ]
#df[df$urban ==1, ]  #county:21,51,63,67,81,119,129,183
#df[df$west ==1 | df$urban ==1 | df$central ==1, ]  #57 total
#df[df$west ==0 & df$urban ==0 & df$central ==0, ]  #county:13,15,17,19,41,47,49,53,55,61,65,79,83,85,9
```

```
plot(df$crmrte, df$county)
model1 <- lm(crmrte~county, data = df)
model1
```

```
##
## Call:
## lm(formula = crmrte ~ county, data = df)
##
## Coefficients:
## (Intercept)       county
##   3.283e-02    6.734e-06
```

```
abline(model1)
```




```
summary(model1) #Multiple R-squared:  0.0004323; p: 0.846
```

```
##
## Call:
## lm(formula = crmrte ~ county, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -0.028074 -0.012578 -0.003471  0.006879  0.065332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.283e-02  4.008e-03   8.191 1.88e-12 ***
## county      6.734e-06  3.452e-05   0.195    0.846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01899 on 88 degrees of freedom
## Multiple R-squared:  0.0004323,  Adjusted R-squared:  -0.01093
## F-statistic: 0.03805 on 1 and 88 DF,  p-value: 0.8458
```

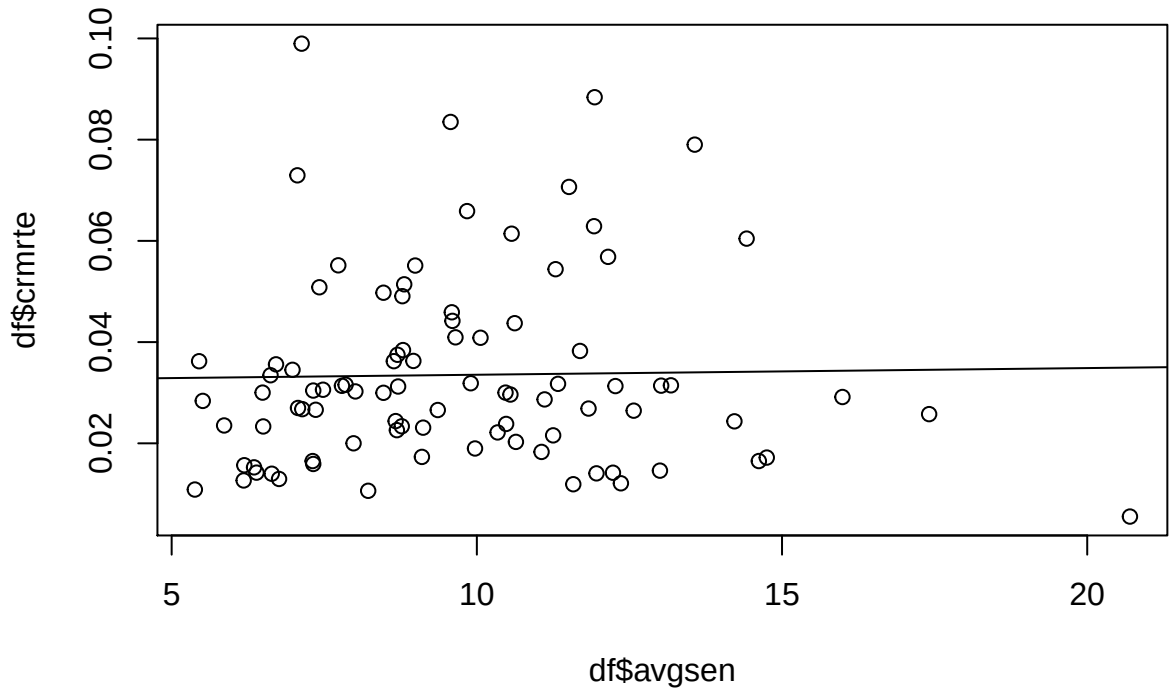
```
cor(df$crmrte, df$county, use="pairwise.complete.obs")   #0.02079069
```

```
## [1] 0.02079069
```

```
plot(df$avgsen, df$crmrte)
model2 <- lm(crmrte~avgsen, data = df)
model2
```

```
##
## Call:
## lm(formula = crmrte ~ avgsen, data = df)
##
## Coefficients:
## (Intercept)       avgsen
##   0.0322318    0.0001319
```

```
abline(model2)
```




```
summary(model2)  #R-squared:  0.0003919; p:0.853
```

```
##
## Call:
```

```
## lm(formula = crmrte ~ avgsen, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.029429 -0.012994 -0.003477  0.006732  0.065794
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0322318  0.0071665   4.498 2.09e-05 ***
## avgsen      0.0001319  0.0007102   0.186    0.853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01899 on 88 degrees of freedom
## Multiple R-squared:  0.0003919,  Adjusted R-squared:  -0.01097
## F-statistic: 0.0345 on 1 and 88 DF,  p-value: 0.8531
```

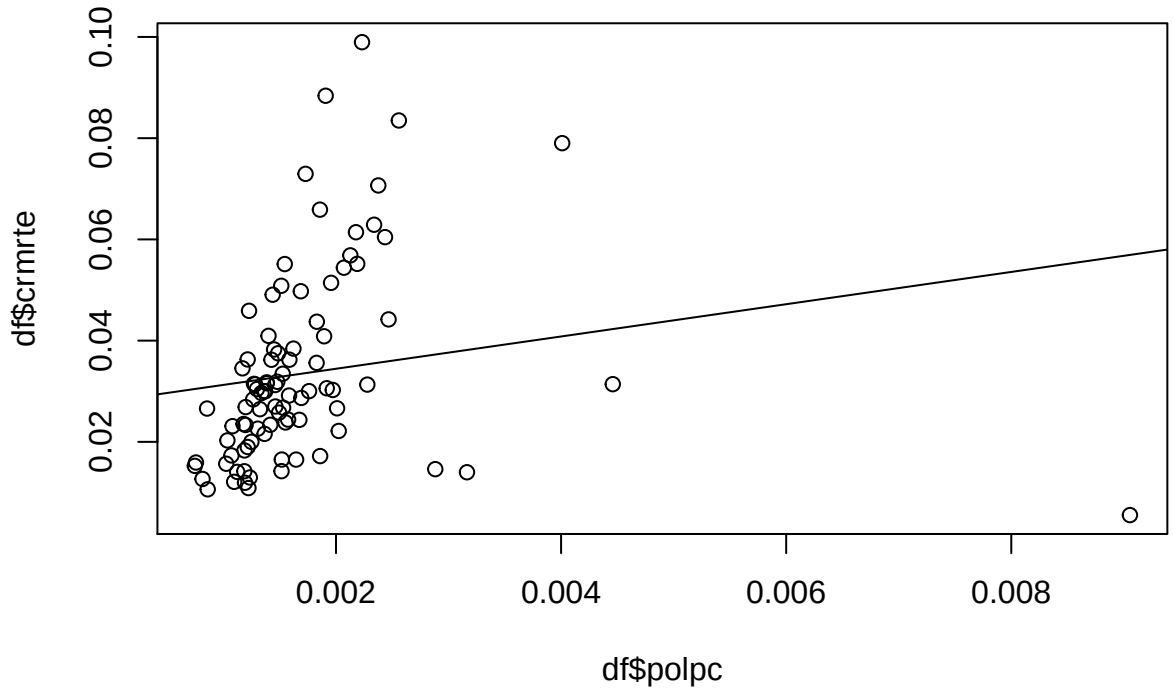
```
cor(df$crmrte, df$avgsen, use="pairwise.complete.obs")   #0.01979653
```

```
## [1] 0.01979653
```

```
plot(df$polpc, df$crmrte)
model3 <- lm(crmrte~polpc, data = df)
model3
```

```
##
## Call:
## lm(formula = crmrte ~ polpc, data = df)
##
## Coefficients:
## (Intercept)        polpc
##     0.02806      3.18839
```

```
abline(model3)
```



```r
summary(model3)  #R-squared:  0.02798; p:0.115
```

```
##
## Call:
## lm(formula = crmrte ~ polpc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.051400 -0.011799 -0.003837  0.006455  0.063787
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02806    0.00395   7.105 2.99e-10 ***
## polpc        3.18839    2.00318   1.592    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01873 on 88 degrees of freedom
## Multiple R-squared:  0.02798,    Adjusted R-squared:  0.01694
## F-statistic: 2.533 on 1 and 88 DF,  p-value: 0.115
```

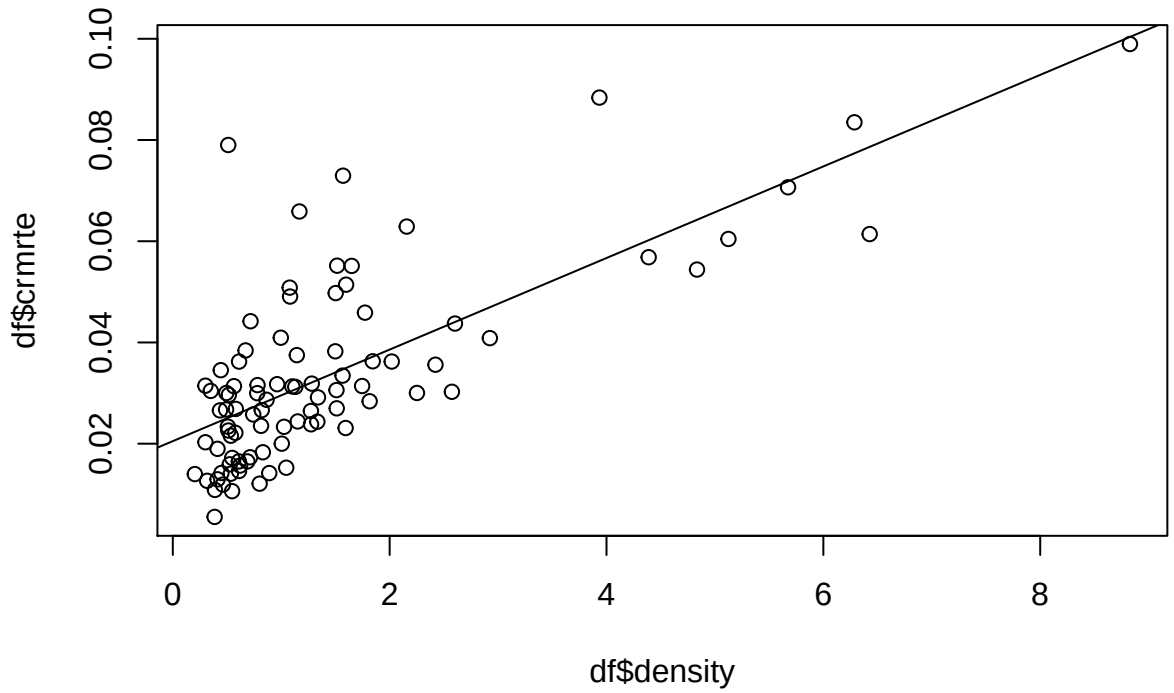
```r
cor(df$crmrte, df$polpc, use="pairwise.complete.obs")   #0.1672816
```

```
## [1] 0.1672816
```

```r
plot(df$density, df$crmrte)
model4 <- lm(crmrte~density, data = df)
model4
```

```
##
## Call:
## lm(formula = crmrte ~ density, data = df)
##
## Coefficients:
## (Intercept)      density
##    0.020503     0.009046
```

```r
abline(model4)
```




```
summary(model4)    #R-squared:  0.5297; p:4.45e-16
```

```
##
## Call:
## lm(formula = crmrte ~ density, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.018459 -0.009471 -0.002741  0.004902  0.053887
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0205027  0.0018954  10.817  < 2e-16 ***
## density     0.0090458  0.0009087   9.955 4.45e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01303 on 88 degrees of freedom
## Multiple R-squared:  0.5297, Adjusted R-squared:  0.5243
## F-statistic:  99.1 on 1 and 88 DF,  p-value: 4.45e-16
```

```
cor(df$crmrte, df$density, use="pairwise.complete.obs")   #0.7277783
```

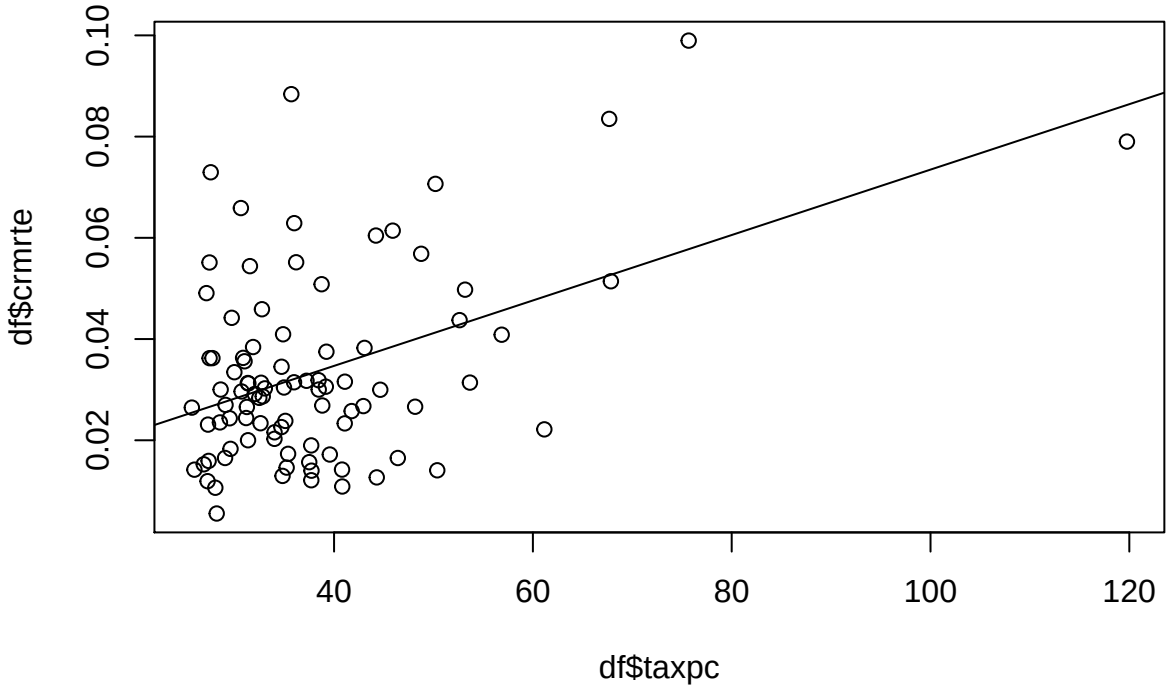
```
## [1] 0.7277783
```

```
plot(df$taxpc, df$crmrte)
model5 <- lm(crmrte~taxpc, data = df)
model5
```

```
##
## Call:
## lm(formula = crmrte ~ taxpc, data = df)
##
## Coefficients:
```

```
## (Intercept)        taxpc
##   0.0088444    0.0006464
```

```
abline(model5)
```




```
summary(model5)   #R-squared:  0.2013; p:p-value: 9.181e-06
```

```
##
## Call:
## lm(formula = crmrte ~ taxpc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027343 -0.010886 -0.002133  0.006679  0.056466
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0088444  0.0055339   1.598    0.114
## taxpc       0.0006464  0.0001372   4.710 9.18e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01697 on 88 degrees of freedom
## Multiple R-squared:  0.2013, Adjusted R-squared:  0.1923
## F-statistic: 22.19 on 1 and 88 DF,  p-value: 9.181e-06
```

```
cor(df$crmrte, df$taxpc, use="pairwise.complete.obs")   #0.4487151
```

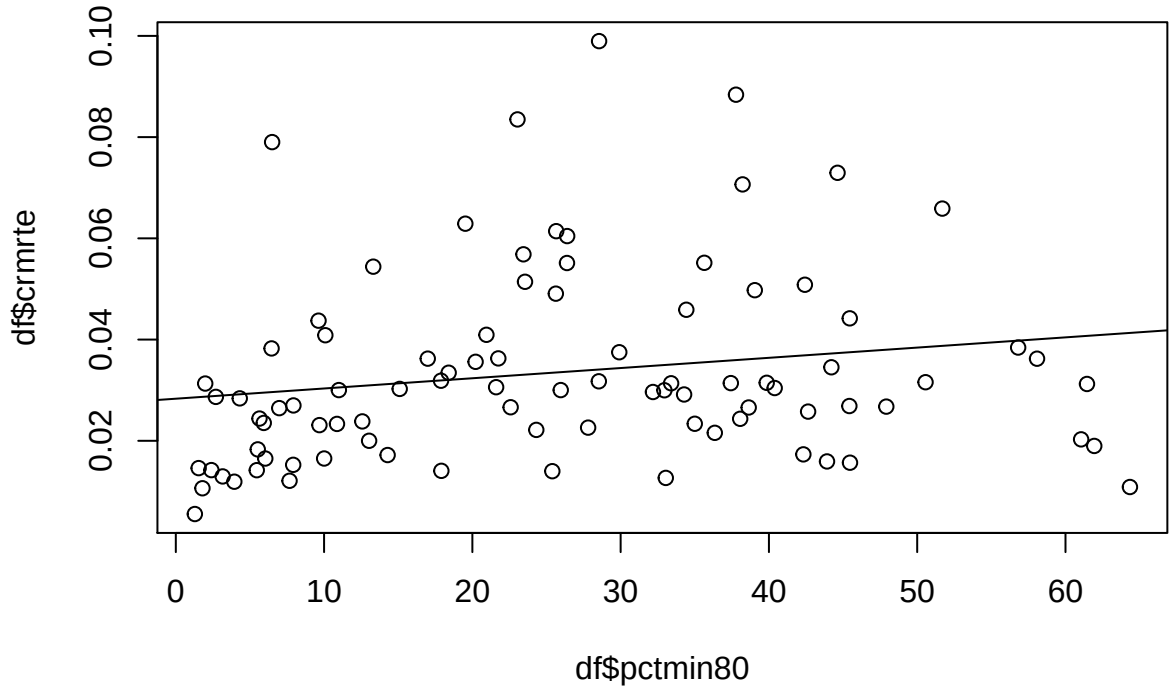
```
## [1] 0.4487151
```

```
plot(df$pctmin80, df$crmrte)
model6 <- lm(crmrte~pctmin80, data = df)
model6
```

```
##
```

```
## Call:
## lm(formula = crmrte ~ pctmin80, data = df)
##
## Coefficients:
## (Intercept)     pctmin80
##    0.028316     0.000202
```

```
abline(model6)
```




```
summary(model6)  #R-squared:  0.033; p:p-value: 0.08662
```

```
##
## Call:
## lm(formula = crmrte ~ pctmin80, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.030444 -0.011928 -0.004692  0.007969  0.064884
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0283161  0.0035861   7.896 7.55e-12 ***
## pctmin80    0.0002020  0.0001166   1.733   0.0866 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01868 on 88 degrees of freedom
## Multiple R-squared:  0.033,  Adjusted R-squared:  0.02201
## F-statistic: 3.003 on 1 and 88 DF,  p-value: 0.08662
```

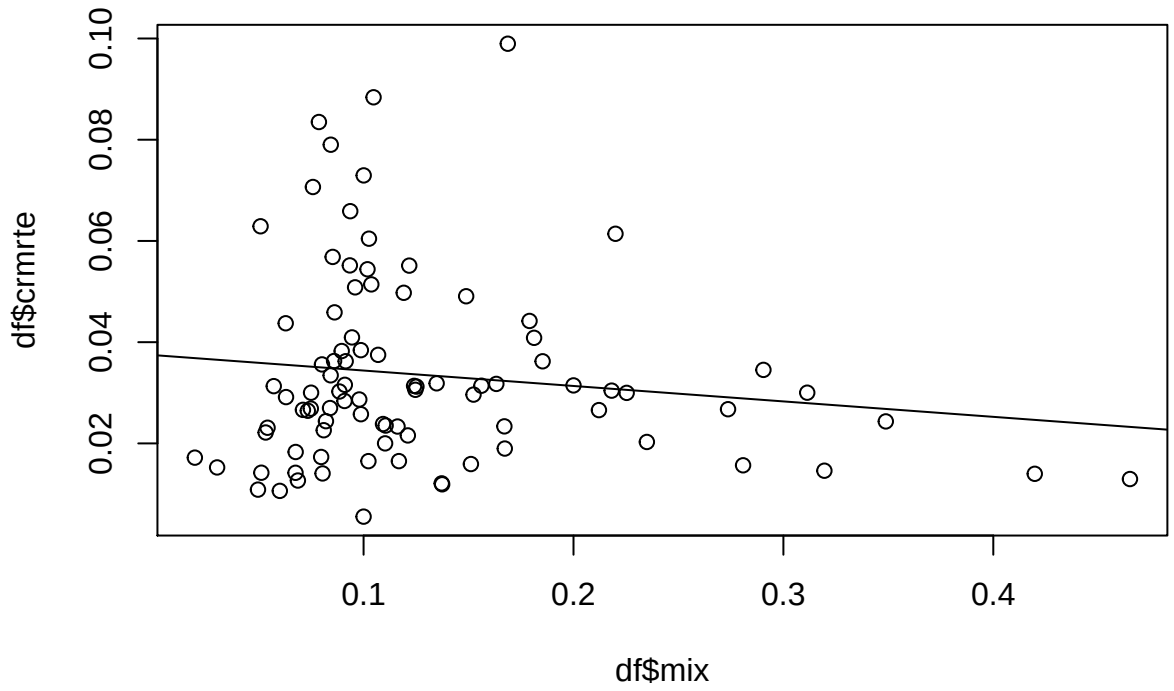
```
cor(df$crmrte, df$pctmin80, use="pairwise.complete.obs")   #0.1816506
```

```
## [1] 0.1816506
```

```r
plot(df$mix, df$crmrte)
model7 <- lm(crmrte~mix, data = df)
model7
```

```
##
## Call:
## lm(formula = crmrte ~ mix, data = df)
##
## Coefficients:
## (Intercept)          mix
##     0.03744     -0.03049
```

```r
abline(model7)  #negative slope/correlation
```




```r
summary(model7)    #R-squared:  0.01742; p:p-value: 0.2149
```

```
##
## Call:
## lm(formula = crmrte ~ mix, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.028862 -0.012325 -0.003125  0.006266  0.066665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.037445   0.003723  10.058 2.73e-16 ***
## mix         -0.030491   0.024409  -1.249    0.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01883 on 88 degrees of freedom
## Multiple R-squared:  0.01742,    Adjusted R-squared:  0.006258
```

```
## F-statistic: 1.561 on 1 and 88 DF,  p-value: 0.2149
```

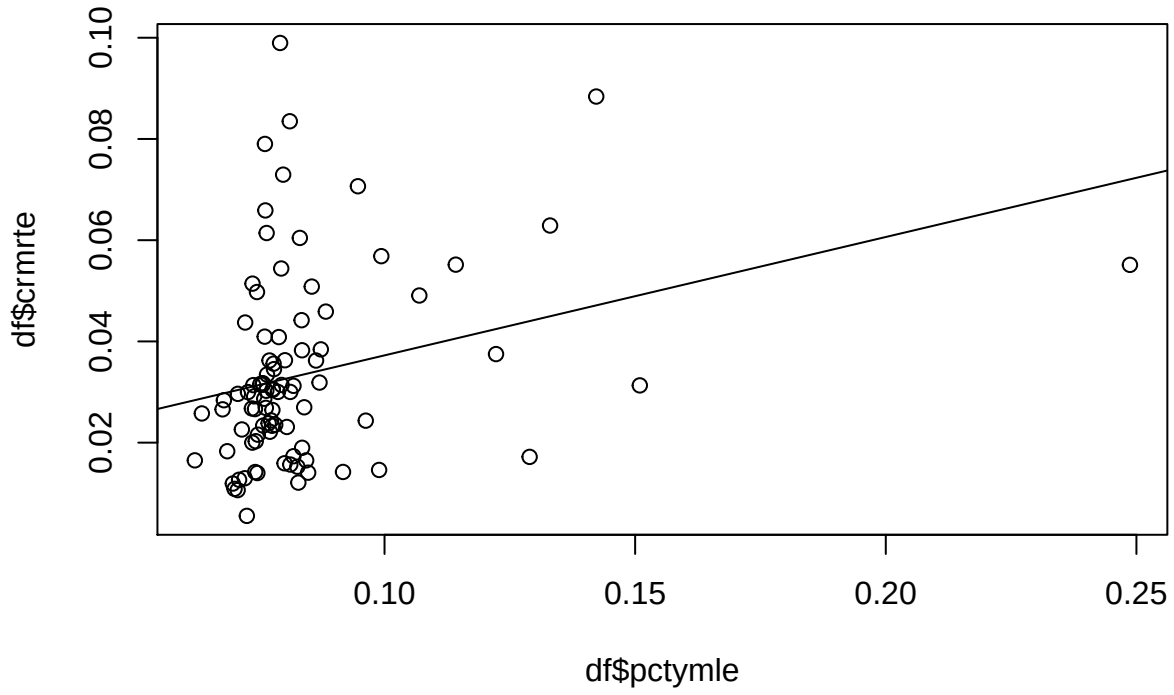
```
cor(df$crmrte, df$mix, use="pairwise.complete.obs")   #-0.1320004
```

```
## [1] -0.1320004
```

```
plot(df$pctymle, df$crmrte)
model8 <- lm(crmrte~pctymle, data = df)
model8
```

```
##
## Call:
## lm(formula = crmrte ~ pctymle, data = df)
##
## Coefficients:
## (Intercept)      pctymle
##     0.01386      0.23385
```

```
abline(model8)
```




```
summary(model8)   #R-squared:  0.0843; p:0.005504
```

```
##
## Call:
## lm(formula = crmrte ~ pctymle, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026828 -0.011823 -0.002585  0.004728  0.066593
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.013860   0.007165   1.935   0.0563 .
## pctymle     0.233851   0.082162   2.846   0.0055 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01818 on 88 degrees of freedom
## Multiple R-squared:  0.0843, Adjusted R-squared:  0.07389
## F-statistic: 8.101 on 1 and 88 DF,  p-value: 0.005504
```

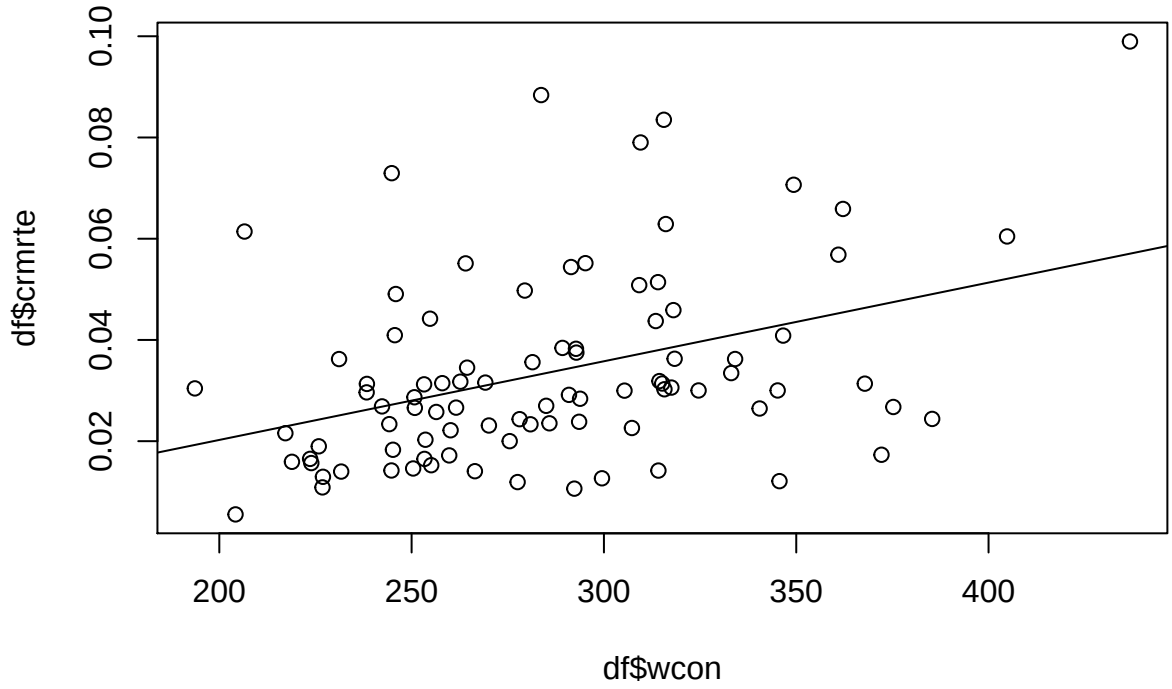
```
cor(df$crmrte, df$pctymle, use="pairwise.complete.obs")  #0.2903397
```

```
## [1] 0.2903397
```

```
plot(df$wcon, df$crmrte)
model9 <- lm(crmrte~wcon, data = df)
model9
```

```
##
## Call:
## lm(formula = crmrte ~ wcon, data = df)
##
## Coefficients:
## (Intercept)         wcon
##  -0.0108405    0.0001554
```

```
abline(model9)
```




```
summary(model9)   #R-squared:  0.1544; p:0.0001275
```

```
##
## Call:
## lm(formula = crmrte ~ wcon, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.030776 -0.011121 -0.005039  0.006936  0.055137
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.084e-02  1.122e-02  -0.967 0.336409
## wcon         1.554e-04  3.877e-05   4.009 0.000128 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01747 on 88 degrees of freedom
## Multiple R-squared:  0.1544, Adjusted R-squared:  0.1448
## F-statistic: 16.07 on 1 and 88 DF,  p-value: 0.0001275
```

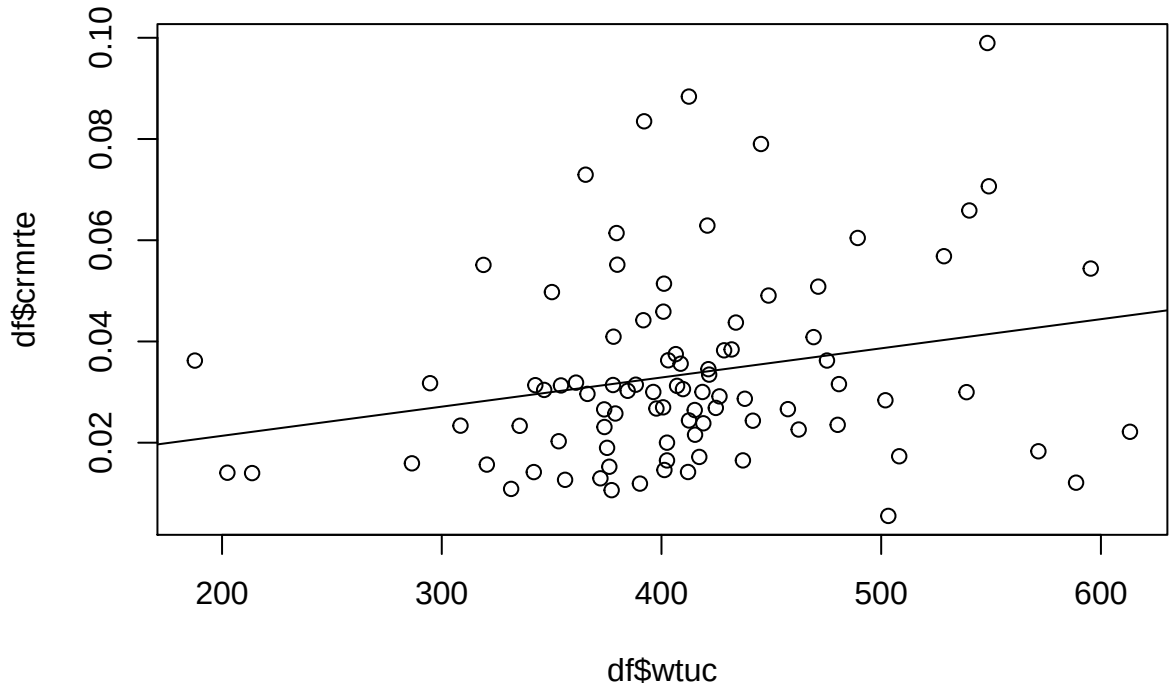
```
cor(df$crmrte, df$wcon, use="pairwise.complete.obs")    #0.3929616
```

```
## [1] 0.3929616
```

```
plot(df$wtuc, df$crmrte)
model10 <- lm(crmrte~wtuc, data = df)
model10
```

```
##
## Call:
## lm(formula = crmrte ~ wtuc, data = df)
##
## Coefficients:
## (Intercept)         wtuc
##   9.833e-03    5.762e-05
```

```
abline(model10)
```




```
summary(model10)    #R-squared:  0.05569; p:0.02514
```

```
##
## Call:
## lm(formula = crmrte ~ wtuc, data = df)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -0.033297 -0.011875 -0.004088  0.007913  0.057538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.833e-03  1.057e-02   0.930   0.3549
## wtuc        5.762e-05  2.529e-05   2.278   0.0251 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01846 on 88 degrees of freedom
## Multiple R-squared:  0.05569,    Adjusted R-squared:  0.04496
## F-statistic:  5.19 on 1 and 88 DF,  p-value: 0.02514
```

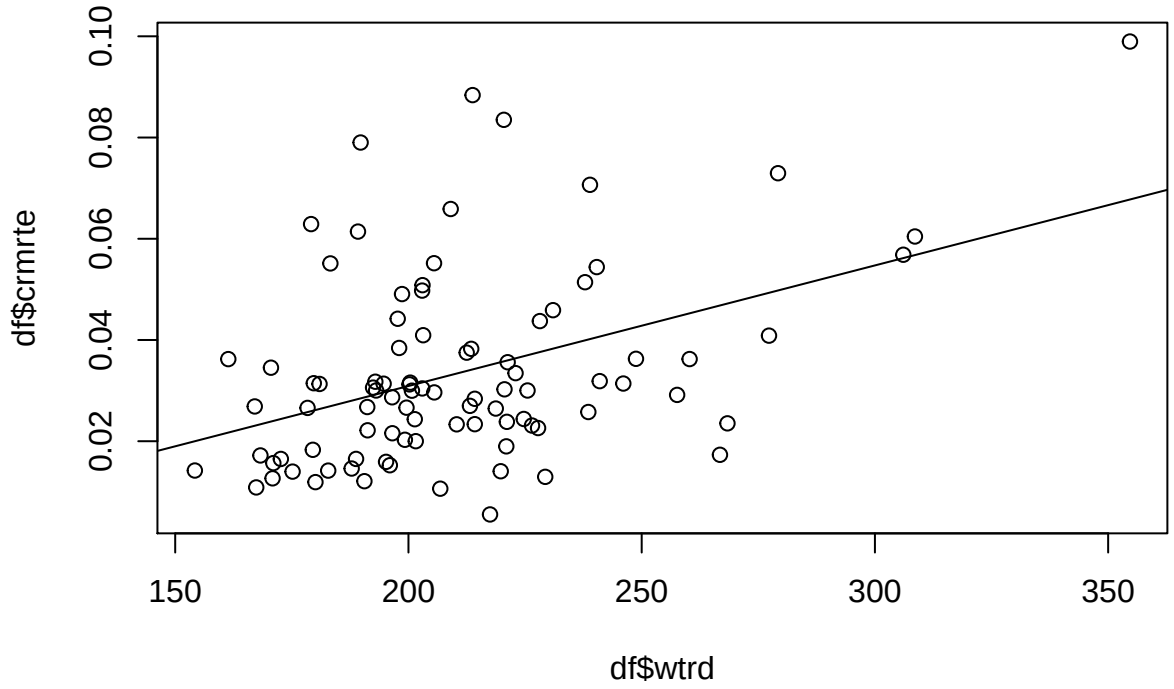
```
cor(df$crmrte, df$wtuc, use="pairwise.complete.obs")   #0.2359957
```

```
## [1] 0.2359957
```

```
plot(df$wtrd, df$crmrte)
model11 <- lm(crmrte~wtrd, data = df)
model11
```

```
##
## Call:
## lm(formula = crmrte ~ wtrd, data = df)
##
## Coefficients:
## (Intercept)         wtrd
##  -0.0167380    0.0002382
```

```
abline(model11)
```




```
summary(model11)   #R-squared:  0.1825; p:2.675e-05
```

```
##
```

```
## Call:
## lm(formula = crmrte ~ wtrd, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.029542 -0.011205 -0.004870  0.005934  0.054201
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.674e-02  1.148e-02  -1.458    0.148
## wtrd         2.382e-04  5.375e-05   4.433 2.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01717 on 88 degrees of freedom
## Multiple R-squared:  0.1825, Adjusted R-squared:  0.1732
## F-statistic: 19.65 on 1 and 88 DF,  p-value: 2.675e-05
```

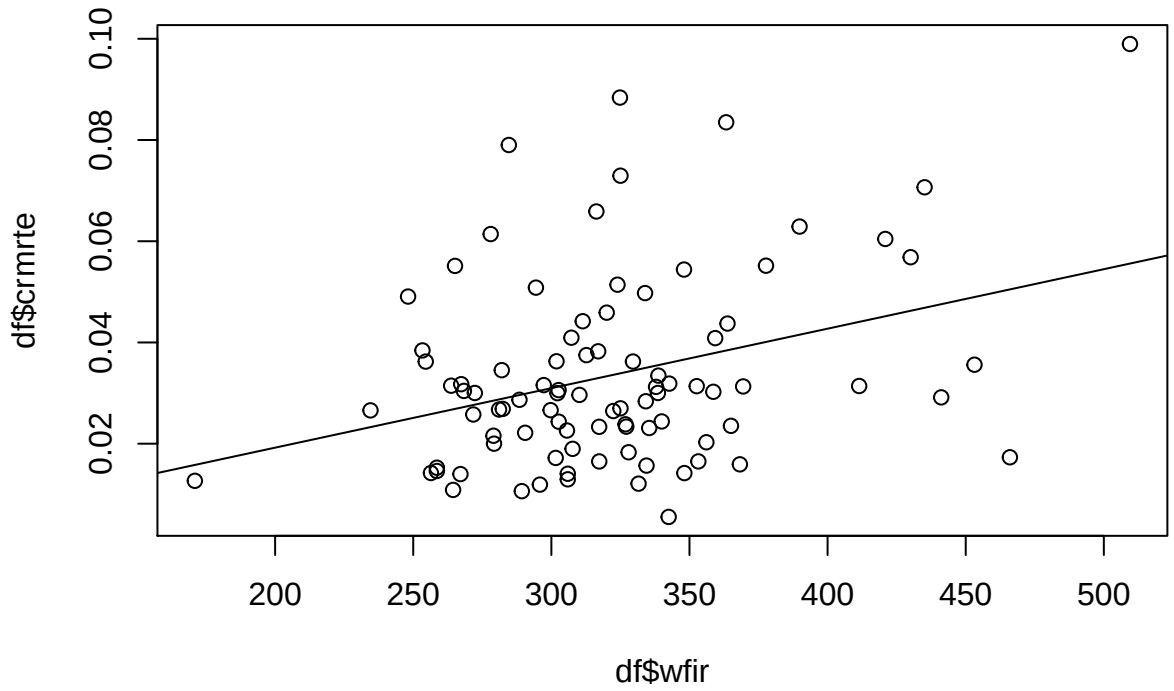
```
cor(df$crmrte, df$wtrd, use="pairwise.complete.obs")   #0.4272226
```

```
## [1] 0.4272226
```

```
plot(df$wfir, df$crmrte)
model12 <- lm(crmrte~wfir, data = df)
model12
```

```
##
## Call:
## lm(formula = crmrte ~ wfir, data = df)
##
## Coefficients:
## (Intercept)         wfir
##  -0.0042906    0.0001175
```

```
abline(model12)
```




```
summary(model12)   #R-squared:  0.1129; p:0.001204
```

```
##
## Call:
## lm(formula = crmrte ~ wfir, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.033163 -0.011934 -0.003624  0.008255  0.054497
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.291e-03  1.145e-02  -0.375   0.7088
## wfir         1.175e-04  3.512e-05   3.347   0.0012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01789 on 88 degrees of freedom
## Multiple R-squared:  0.1129, Adjusted R-squared:  0.1028
## F-statistic:  11.2 on 1 and 88 DF,  p-value: 0.001204
```

```
cor(df$crmrte, df$wfir, use="pairwise.complete.obs")   #0.3360261
```

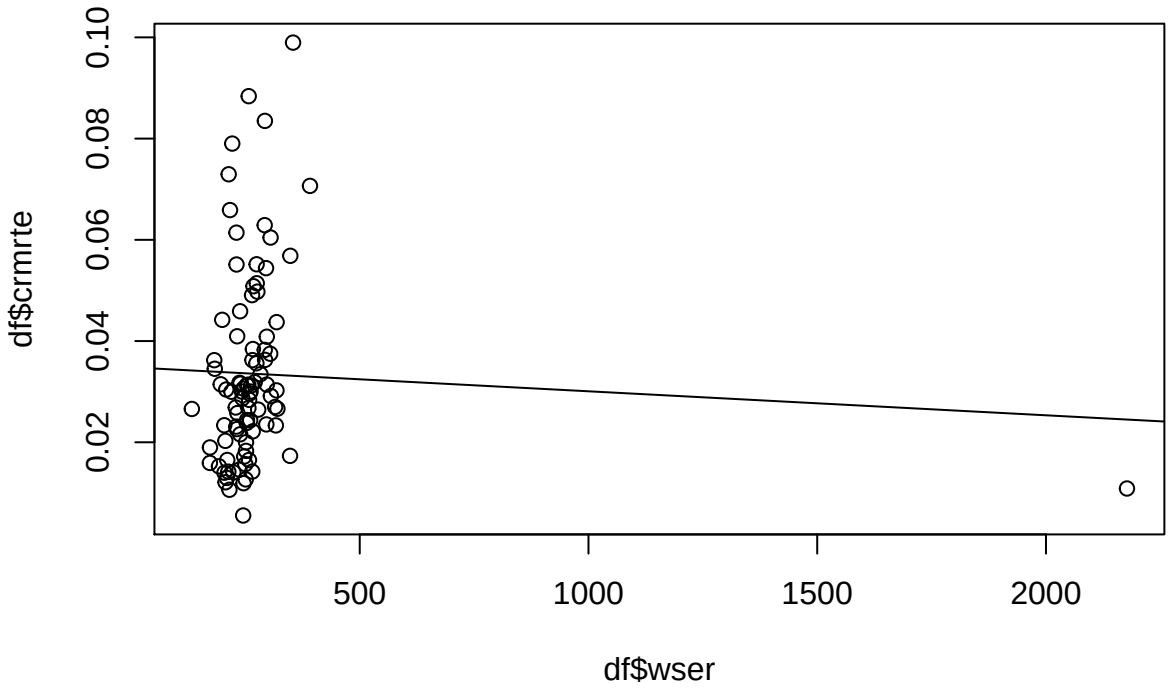
```
## [1] 0.3360261
```

```
plot(df$wser, df$crmrte)
model13 <- lm(crmrte~wser, data = df)
model13
```

```
##
## Call:
## lm(formula = crmrte ~ wser, data = df)
##
## Coefficients:
```

```
## (Intercept)         wser
##   3.482e-02   -4.742e-06
```

```
abline(model13)
```




```
summary(model13)   #R-squared:  0.002711; p:0.626
```

```
##
## Call:
## lm(formula = crmrte ~ wser, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.028120 -0.013195 -0.003718  0.006638  0.065830
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.482e-02  3.335e-03  10.439   <2e-16 ***
## wser        -4.742e-06  9.695e-06  -0.489    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01897 on 88 degrees of freedom
## Multiple R-squared:  0.002711,   Adjusted R-squared:  -0.008622
## F-statistic: 0.2392 on 1 and 88 DF,  p-value: 0.626
```

```
cor(df$crmrte, df$wser, use="pairwise.complete.obs")   #-0.05206995
```

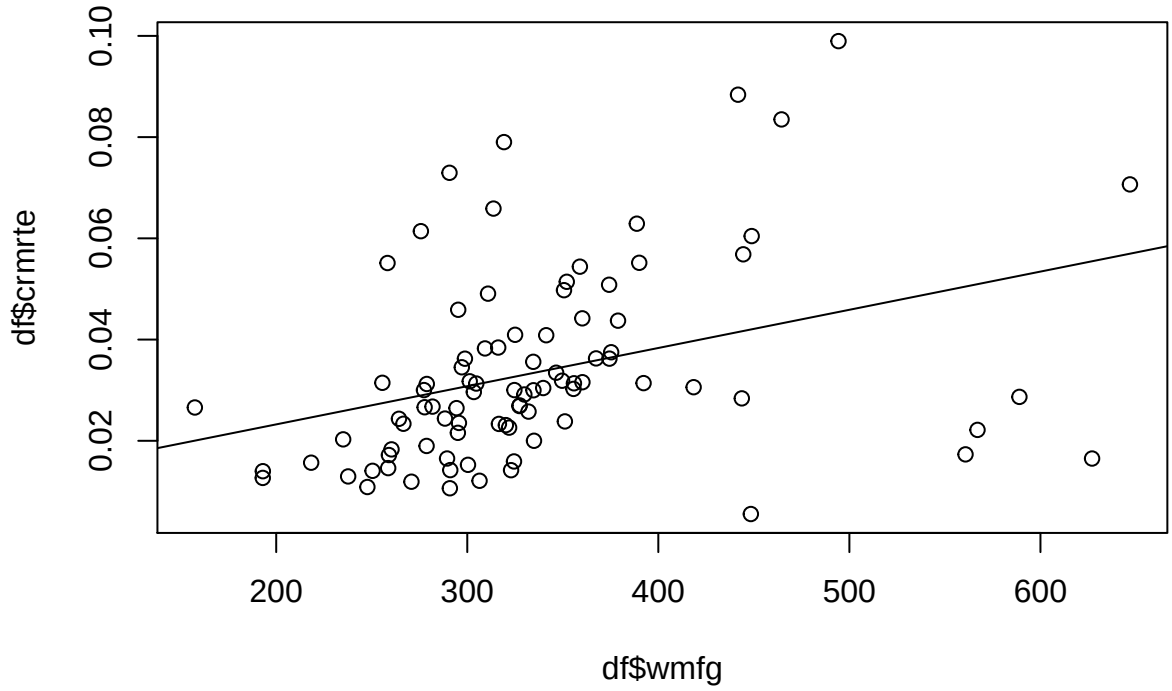
```
## [1] -0.05206995
```

```
plot(df$wmfg, df$crmrte)
model14 <- lm(crmrte~wmfg, data = df)
model14
```

```
##
```

```
## Call:
## lm(formula = crmrte ~ wmfg, data = df)
##
## Coefficients:
## (Intercept)         wmfg
##   8.149e-03    7.547e-05
```

```
abline(model14)
```




```
summary(model14)   #R-squared:  0.1243; p:0.0006543
```

```
##
## Call:
## lm(formula = crmrte ~ wmfg, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.038978 -0.009995 -0.003531  0.006901  0.053511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.149e-03  7.416e-03   1.099 0.274829
## wmfg        7.547e-05  2.135e-05   3.534 0.000654 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01777 on 88 degrees of freedom
## Multiple R-squared:  0.1243, Adjusted R-squared:  0.1143
## F-statistic: 12.49 on 1 and 88 DF,  p-value: 0.0006543
```

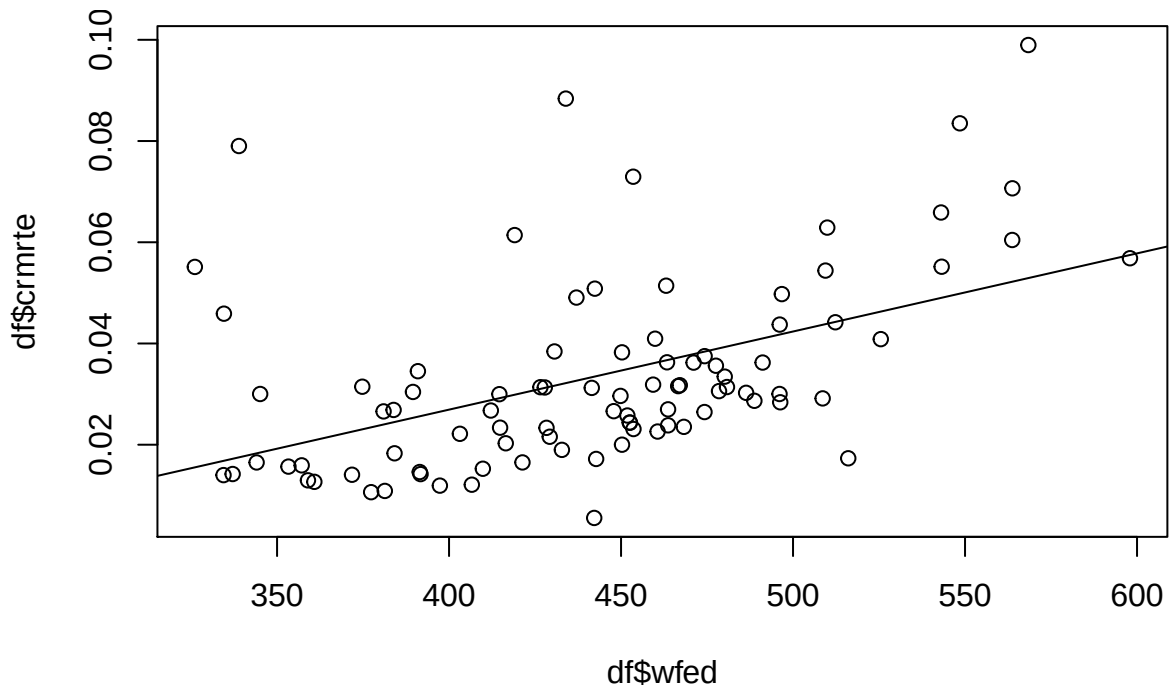
```
cor(df$crmrte, df$wmfg, use="pairwise.complete.obs")   #0.3525612
```

```
## [1] 0.3525612
```

```
plot(df$wfed, df$crmrte)
model15 <- lm(crmrte~wfed, data = df)
model15
```

```
##
## Call:
## lm(formula = crmrte ~ wfed, data = df)
##
## Coefficients:
## (Intercept)         wfed
##  -0.0348051    0.0001543
```

```
abline(model15)
```




```
summary(model15)   #R-squared:  0.24; p:9.54e-07
```

```
##
## Call:
## lm(formula = crmrte ~ wfed, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027912 -0.010687 -0.004585  0.005028  0.061513
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.481e-02  1.308e-02  -2.662  0.00924 **
## wfed         1.543e-04  2.928e-05   5.272 9.54e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01656 on 88 degrees of freedom
## Multiple R-squared:   0.24,  Adjusted R-squared:  0.2314
```

```
## F-statistic: 27.79 on 1 and 88 DF,  p-value: 9.54e-07
```

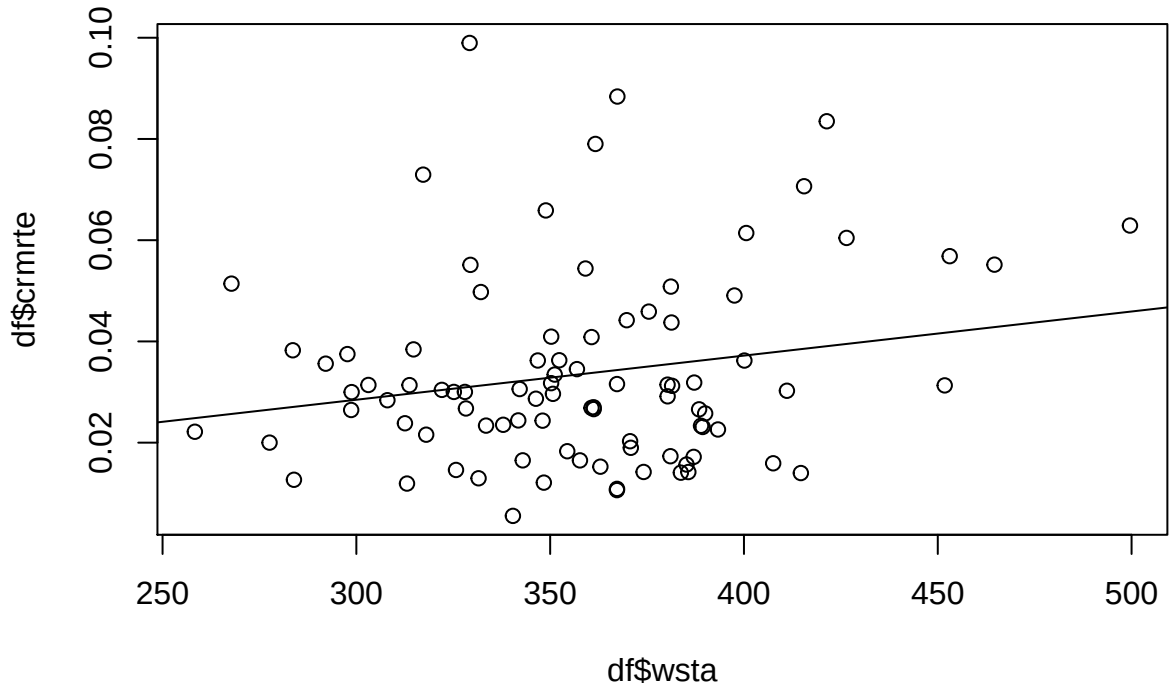
```
cor(df$crmrte, df$wfed, use="pairwise.complete.obs")   #0.4899163
```

```
## [1] 0.4899163
```

```
plot(df$wsta, df$crmrte)
model16 <- lm(crmrte~wsta, data = df)
model16
```

```
##
## Call:
## lm(formula = crmrte ~ wsta, data = df)
##
## Coefficients:
## (Intercept)         wsta
##   2.321e-03    8.718e-05
```

```
abline(model16)
```




```
summary(model16)   #R-squared:  0.03994; p:0.05896
```

```
##
## Call:
## lm(formula = crmrte ~ wsta, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.026464 -0.013797 -0.003547  0.008541  0.067942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.321e-03  1.642e-02   0.141    0.888
## wsta        8.718e-05  4.557e-05   1.913    0.059 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01861 on 88 degrees of freedom
## Multiple R-squared:  0.03994,    Adjusted R-squared:  0.02903
## F-statistic: 3.661 on 1 and 88 DF,  p-value: 0.05896
```

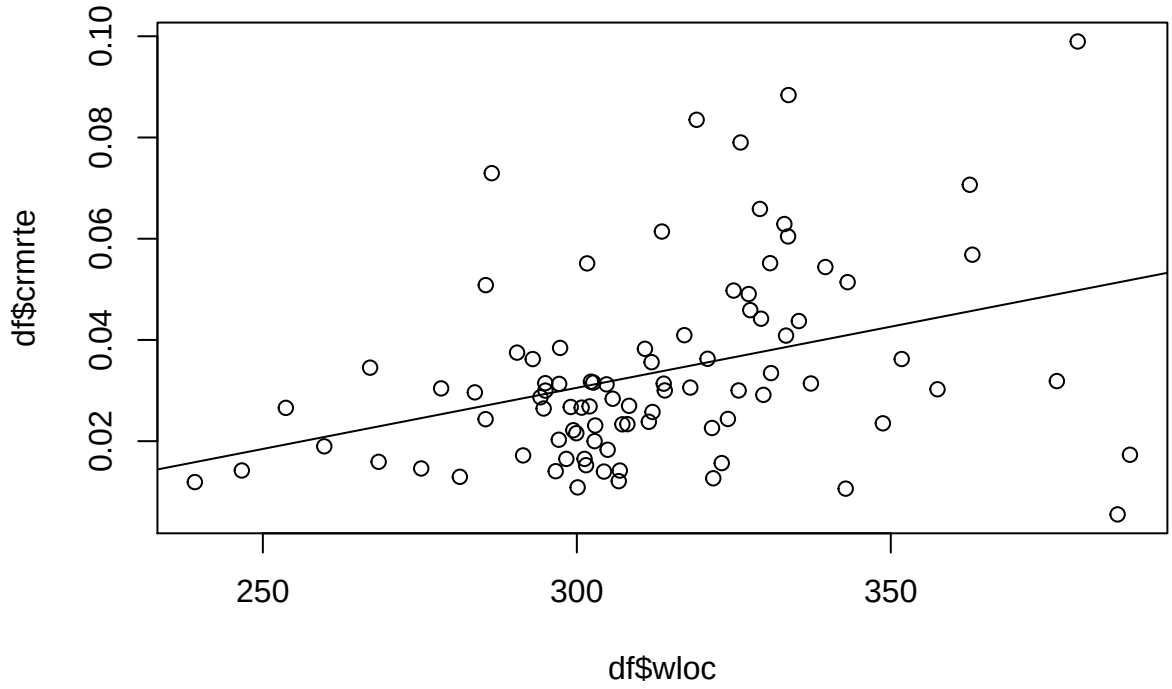
```
cor(df$crmrte, df$wsta, use="pairwise.complete.obs")   #0.1998467
```

```
## [1] 0.1998467
```

```
plot(df$wloc, df$crmrte)
model17 <- lm(crmrte~wloc, data = df)
model17
```

```
##
## Call:
## lm(formula = crmrte ~ wloc, data = df)
##
## Coefficients:
## (Intercept)         wloc
##  -0.0419300    0.0002416
```

```
abline(model17)
```




```
summary(model17)   #R-squared:  0.1295; p:0.0004951
```

```
##
## Call:
## lm(formula = crmrte ~ wloc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.045815 -0.009855 -0.003494  0.007339  0.049698
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.193e-02  2.094e-02  -2.003 0.048278 *
## wloc         2.416e-04  6.677e-05   3.618 0.000495 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01772 on 88 degrees of freedom
## Multiple R-squared:  0.1295, Adjusted R-squared:  0.1196
## F-statistic: 13.09 on 1 and 88 DF,  p-value: 0.0004951
```

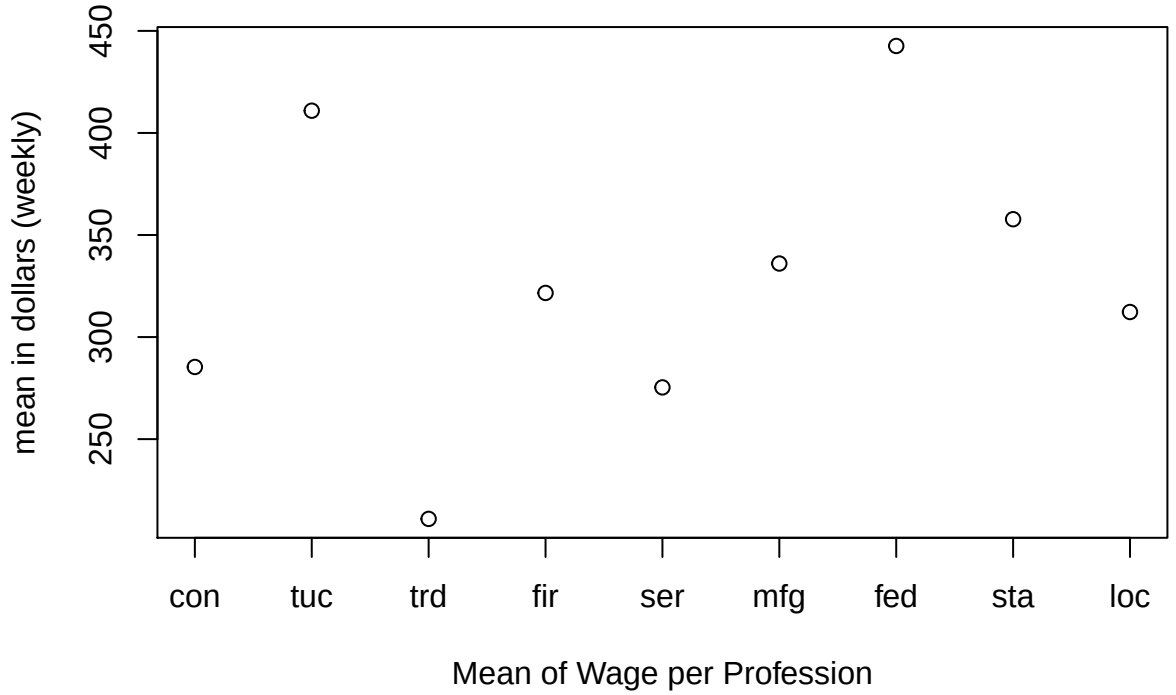
```
cor(df$crmrte, df$wloc, use="pairwise.complete.obs")   #0.3598293
```

```
## [1] 0.3598293
```

```
mean1<- apply(df[16:24], 2, mean)
mean(mean1)   #328.0903
```

```
## [1] 328.0903
```

```
plot(mean1, xaxt = "n", xlab='Mean of Wage per Profession', ylab = "mean in dollars (weekly)")
axis(1, at = 1:9, labels = c("con", "tuc", "trd", "fir", "ser", "mfg", "fed", "sta", "loc" ))
```




```
model20 <- lm(crmrte~wloc + wsta + wfed, data = df)
summary(model20)   #R-squared:  0.2652; Adjusted R-squared:  0.2396
```

```
##
## Call:
## lm(formula = crmrte ~ wloc + wsta + wfed, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.033839 -0.009107 -0.005086  0.006279  0.057169
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.659e-02  2.263e-02  -2.943 0.004180 **
## wloc         9.059e-05  7.285e-05   1.243 0.217093
## wsta         4.459e-05  4.119e-05   1.083 0.281981
## wfed         1.262e-04  3.433e-05   3.676 0.000412 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01647 on 86 degrees of freedom
## Multiple R-squared:  0.2652, Adjusted R-squared:  0.2396
## F-statistic: 10.35 on 3 and 86 DF,  p-value: 6.954e-06
```

```
model21 <- lm(crmrte~polpc + density + taxpc, data = df)
summary(model21)   #R-squared:  0.5814; Adjusted R-squared:  0.5668
```

```
##
## Call:
## lm(formula = crmrte ~ polpc + density + taxpc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.016453 -0.007955 -0.002938  0.003917  0.042047
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0087751  0.0042403   2.069  0.04150 *
## polpc       -0.1086169  1.3896572  -0.078  0.93788
## density      0.0080949  0.0009183   8.816 1.17e-13 ***
## taxpc        0.0003480  0.0001094   3.181  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01243 on 86 degrees of freedom
## Multiple R-squared:  0.5814, Adjusted R-squared:  0.5668
## F-statistic: 39.82 on 3 and 86 DF,  p-value: 3.122e-16
```

```
model22 <- lm(crmrte~polpc + density, data = df)
summary(model22)   #R-squared:  0.5322, Adjusted R-squared:  0.5214
```

```
##
## Call:
## lm(formula = crmrte ~ polpc + density, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.025716 -0.009161 -0.002382  0.004857  0.051552
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0189880  0.0029109   6.523 4.38e-09 ***
## polpc       0.9731434  1.4162336   0.687    0.494
## density     0.0089433  0.0009235   9.684 1.79e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.01307 on 87 degrees of freedom
## Multiple R-squared:  0.5322, Adjusted R-squared:  0.5214
## F-statistic: 49.49 on 2 and 87 DF,  p-value: 4.442e-15
```

```
model23 <- lm(crmrte~polpc + taxpc, data = df)
summary(model23)   #R-squared:  0.2032, Adjusted R-squared:  0.1849
```

```
##
## Call:
## lm(formula = crmrte ~ polpc + taxpc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.028006 -0.010275 -0.001730  0.006745  0.056250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0080747  0.0058157   1.388    0.169
## polpc       0.8563592  1.9003738   0.451    0.653
## taxpc       0.0006282  0.0001436   4.374 3.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01705 on 87 degrees of freedom
## Multiple R-squared:  0.2032, Adjusted R-squared:  0.1849
## F-statistic: 11.09 on 2 and 87 DF,  p-value: 5.112e-05
```

```
model24 <- lm(crmrte~density + taxpc, data = df)
summary(model24)   #R-squared:  0.5814, Adjusted R-squared:  0.5718
```

```
##
## Call:
## lm(formula = crmrte ~ density + taxpc, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.016369 -0.007906 -0.002897  0.003928  0.042023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0086776  0.0040294   2.154   0.0340 *
## density     0.0080892  0.0009102   8.888 7.63e-14 ***
## taxpc       0.0003459  0.0001055   3.279   0.0015 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01236 on 87 degrees of freedom
## Multiple R-squared:  0.5814, Adjusted R-squared:  0.5718
## F-statistic: 60.42 on 2 and 87 DF,  p-value: < 2.2e-16
```

```
model25 <- lm(crmrte~wcon + wtuc + wtrd + wfir + wser + wmfg, data = df)
summary(model25)   #R-squared: 0.2517,  Adjusted R-squared:  0.1976
```

```
##
## Call:
## lm(formula = crmrte ~ wcon + wtuc + wtrd + wfir + wser + wmfg,
```

```
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.03885 -0.01042 -0.00423  0.00718  0.04963
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.761e-02  1.396e-02  -1.978   0.0512 .
## wcon         7.685e-05  4.787e-05   1.605   0.1122
## wtuc        -7.180e-06  2.761e-05  -0.260   0.7955
## wtrd         1.517e-04  7.682e-05   1.975   0.0516 .
## wfir        -1.354e-05  4.832e-05  -0.280   0.7800
## wser        -4.176e-06  8.657e-06  -0.482   0.6308
## wmfg         4.655e-05  2.528e-05   1.842   0.0691 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01692 on 83 degrees of freedom
## Multiple R-squared:  0.2517, Adjusted R-squared:  0.1976
## F-statistic: 4.653 on 6 and 83 DF,  p-value: 0.0004001
```

```
model26 <- lm(crmrte~wcon + wtuc + wtrd + wfir, data = df)
summary(model26)   #R-squared:  0.2192, Adjusted R-squared:  0.1825
```

```
##
## Call:
## lm(formula = crmrte ~ wcon + wtuc + wtrd + wfir, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.034736 -0.009661 -0.006039  0.006656  0.054520
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.023e-02  1.383e-02  -2.186   0.0316 *
## wcon         7.990e-05  4.829e-05   1.655   0.1017
## wtuc         1.090e-05  2.609e-05   0.418   0.6770
## wtrd         1.499e-04  7.748e-05   1.935   0.0563 .
## wfir         1.504e-05  4.600e-05   0.327   0.7446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01708 on 85 degrees of freedom
## Multiple R-squared:  0.2192, Adjusted R-squared:  0.1825
## F-statistic: 5.967 on 4 and 85 DF,  p-value: 0.0002792
```

```
model27 <- lm(crmrte~wcon + wtuc + wser + wmfg, data = df)
summary(model27)   #R-squared:  0.2101, Adjusted R-squared:  0.1729
```

```
##
## Call:
## lm(formula = crmrte ~ wcon + wtuc + wser + wmfg, data = df)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -0.038282 -0.010411 -0.004362  0.006482  0.049353
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.721e-02  1.276e-02  -1.348  0.18111
## wcon         1.222e-04  4.259e-05   2.870  0.00518 **
## wtuc        -2.095e-06  2.792e-05  -0.075  0.94037
## wser        -4.591e-06  8.783e-06  -0.523  0.60255
## wmfg         5.344e-05  2.382e-05   2.244  0.02743 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01718 on 85 degrees of freedom
## Multiple R-squared:  0.2101, Adjusted R-squared:  0.1729
## F-statistic: 5.651 on 4 and 85 DF,  p-value: 0.0004408
```

```
model28 <- lm(crmrte~wtrd + wfir + wser + wmfg, data = df)
summary(model28)   #R-squared:  0.2284, Adjusted R-squared:  0.1921
```

```
##
## Call:
## lm(formula = crmrte ~ wtrd + wfir + wser + wmfg, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.037237 -0.011049 -0.003738  0.007366  0.050084
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.159e-02  1.248e-02  -1.730  0.08722 .
## wtrd         1.938e-04  7.160e-05   2.707  0.00821 **
## wfir        -3.227e-06  4.806e-05  -0.067  0.94662
## wser        -4.261e-06  8.685e-06  -0.491  0.62496
## wmfg         4.891e-05  2.356e-05   2.076  0.04091 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01698 on 85 degrees of freedom
## Multiple R-squared:  0.2284, Adjusted R-squared:  0.1921
## F-statistic: 6.289 on 4 and 85 DF,  p-value: 0.0001757
```

```
model29 <- lm(crmrte~pctmin80 + pctymle + density, data = df)
summary(model29)   #R-squared: 0.6301,  Adjusted R-squared:  0.6172
```

```
##
## Call:
## lm(formula = crmrte ~ pctmin80 + pctymle + density, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.021145 -0.006220 -0.001704  0.003676  0.060281
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -5.920e-04  5.072e-03  -0.117  0.90735
## pctmin80     2.665e-04  7.314e-05   3.643  0.00046 ***
## pctymle      1.709e-01  5.318e-02   3.213  0.00185 **
## density      8.966e-03  8.228e-04  10.897  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01169 on 86 degrees of freedom
## Multiple R-squared:  0.6301, Adjusted R-squared:  0.6172
## F-statistic: 48.83 on 3 and 86 DF,  p-value: < 2.2e-16
```

```
model30 <- lm(crmrte~pctmin80 + pctymle, data = df)
summary(model30)   #R-squared:  0.1194, Adjusted R-squared:  0.09913
```

```
##
## Call:
## lm(formula = crmrte ~ pctmin80 + pctymle, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.027386 -0.012567 -0.003621  0.006580  0.066017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0082605  0.0076801   1.076  0.28509
## pctmin80    0.0002083  0.0001119   1.861  0.06606 .
## pctymle     0.2367562  0.0810496   2.921  0.00444 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01793 on 87 degrees of freedom
## Multiple R-squared:  0.1194, Adjusted R-squared:  0.09913
## F-statistic: 5.896 on 2 and 87 DF,  p-value: 0.003968
```

```
model31 <- lm(crmrte~pctmin80 + density, data = df)
summary(model31)   #R-squared:  0.5857, Adjusted R-squared:  0.5762
```

```
##
## Call:
## lm(formula = crmrte ~ pctmin80 + density, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.023115 -0.007062 -0.002932  0.003894  0.059163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.340e-02  2.737e-03   4.896 4.47e-06 ***
## pctmin80    2.639e-04  7.695e-05   3.430 0.000926 ***
## density     9.266e-03  8.601e-04  10.773  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0123 on 87 degrees of freedom
## Multiple R-squared:  0.5857, Adjusted R-squared:  0.5762
```

```
## F-statistic: 61.49 on 2 and 87 DF,  p-value: < 2.2e-16
```

```
model32 <- lm(crmrte~pctymle + density, data = df)
summary(model32)   #R-squared:  0.573,  Adjusted R-squared:  0.5632
```

```
##
## Call:
## lm(formula = crmrte ~ pctymle + density, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.016834 -0.008635 -0.002138  0.005083  0.054941
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0067496  0.0049719   1.358  0.17811
## pctymle     0.1687921  0.0568024   2.972  0.00383 **
## density     0.0087468  0.0008766   9.979 4.47e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01248 on 87 degrees of freedom
## Multiple R-squared:  0.573,  Adjusted R-squared:  0.5632
## F-statistic: 58.37 on 2 and 87 DF,  p-value: < 2.2e-16
```