Lab 10

Nikita Grabher-Meyer

11/11/2020

Matrix, Figures and Experiments

1) Starting from file Lab_09.RScript, try to run yourself what you saw in "Matrix_03" the primer session

Identity matrix

```
I1 <- diag(1)
I1

##      [,1]
## [1,]    1

I2 <- diag(2)
I2

##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1

A <- matrix(1:4, nrow=2, ncol = 2)
A

##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4

AI <- A%*%I2
AI

##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

Symmetric matrix

```
S <- cbind(c(1,2,3),c(2,1,4), c(3,4,3))
S

##      [,1] [,2] [,3]
## [1,]    1    2    3
```

```
## [2,]    2    1    4
## [3,]    3    4    3

trS <- t(S)
trS

##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    2    1    4
## [3,]    3    4    3
```

Basic properties

```
X<- cbind(c(1,2), c(3,4), c(5,6))
Y <- cbind(c(7,8), c(9,10), c(11,12))
X+Y

##      [,1] [,2] [,3]
## [1,]    8   12   16
## [2,]   10   14   18

Z <- cbind(c(1,2), c(3,4))
W <- cbind(c(5,6), c(8,8))
ZW1 <- Z%*%W
ZW2 <- W%*%Z
ZW1

##      [,1] [,2]
## [1,]   23   32
## [2,]   34   48

ZW2

##      [,1] [,2]
## [1,]   21   47
## [2,]   22   50

trZW1 <- t(ZW1)
trZ <- t(Z)
trW <- t(W)
trWtrZ <- trW%*%trZ
trZW1

##      [,1] [,2]
## [1,]   23   34
## [2,]   32   48

trWtrZ

##      [,1] [,2]
## [1,]   23   34
## [2,]   32   48
```

```
trtrZ <- t(trZ)
trtrZ
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```
Z
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

Inverse matrix

```
C <- matrix(c(1,0,3,1), nrow=2, ncol = 2)
C
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    0    1
```

```
invC <- solve(C)
invC
```

```
##      [,1] [,2]
## [1,]    1   -3
## [2,]    0    1
```

```
CinvC <- C%*%invC
CinvC
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

2) Using the file Lab02_mkm_Econometrics1_RodBased_DataHandling:

a.Read dt.wages and apply 'data.table'

```
setwd("/Users/nikitagrabher-meyer/Desktop/PHD/Econometrics/Labs/Lab 10, Homework")

library(data.table)
library(ggplot2)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

library(Metrics)

load("dt_wages.RData")

dt.wages <- data.table(dt.wages)
```

b.Get basic summary statistics. How many observations, how many variables?

```
str(dt.wages)

## Classes 'data.table' and 'data.frame':   526 obs. of  24 variables:
##  $ wage    : num  3.1 3.24 3 6 5.3 ...
##  $ educ    : int  11 12 11 8 12 16 18 12 12 17 ...
##  $ exper   : int  2 22 2 44 7 9 15 5 26 22 ...
##  $ tenure  : int  0 2 0 28 2 8 7 3 4 21 ...
##  $ nonwhite: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ female  : int  1 1 0 0 0 0 0 1 1 0 ...
##  $ married : int  0 1 0 1 1 1 0 0 0 1 ...
##  $ numdep  : int  2 3 2 0 1 0 0 0 2 0 ...
##  $ smsa    : int  1 1 0 1 0 1 1 1 1 1 ...
##  $ northcen: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ south   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ west    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ construc: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ndurman : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ trcommpu: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ trade   : int  0 0 1 0 0 0 1 0 1 0 ...
##  $ services: int  0 1 0 0 0 0 0 0 0 0 ...
##  $ profserv: int  0 0 0 0 0 1 0 0 0 0 ...
##  $ profocc : int  0 0 0 0 0 1 1 1 1 1 ...
##  $ clerocc : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ servocc : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ lwage   : num  1.13 1.18 1.1 1.79 1.67 ...
##  $ expersq : int  4 484 4 1936 49 81 225 25 676 484 ...
##  $ tenursq : int  0 4 0 784 4 64 49 9 16 441 ...
##  - attr(*, ".internal.selfref")=<externalptr>

head(dt.wages, 3)

##    wage educ exper tenure nonwhite female married numdep smsa northcen
south
## 1: 3.10   11     2      0        0      1       0      2    1        0
0
## 2: 3.24   12    22      2        0      1       1      3    1        0
0
## 3: 3.00   11     2      0        0      0       0      2    0        0
```

```
##    west construc ndurman trcommpu trade services profserv profocc clerocc
## 1:    1        0       0        0     0        0        0       0       0
## 2:    1        0       0        0     0        1        0       0       0
## 3:    1        0       0        0     1        0        0       0       0
##    servocc    lwage expersq tenursq
## 1:       0 1.131402       4       0
## 2:       1 1.175573     484       4
## 3:       0 1.098612       4       0
```

```
tail(dt.wages, 3)
```

```
##     wage educ exper tenure nonwhite female married numdep smsa northcen
south
## 1:  4.67   15    13     18        0      0       1      3    0        0
0
## 2: 11.56   16     5      1        0      0       1      0    0        0
0
## 3:  3.50   14     5      4        1      1       0      2    0        0
0
##    west construc ndurman trcommpu trade services profserv profocc clerocc
## 1:    1        1       0        0     0        0        0       1       0
## 2:    1        0       1        0     0        0        0       0       0
## 3:    1        0       0        0     0        0        1       0       1
##    servocc    lwage expersq tenursq
## 1:       0 1.541159     169     324
## 2:       0 2.447551      25       1
## 3:       0 1.252763      25      16
```

```
summary(dt.wages)
```

```
##       wage             educ           exper           tenure
##  Min.   : 0.530   Min.   : 0.00   Min.   : 1.00   Min.   : 0.000
##  1st Qu.: 3.330   1st Qu.:12.00   1st Qu.: 5.00   1st Qu.: 0.000
##  Median : 4.650   Median :12.00   Median :13.50   Median : 2.000
##  Mean   : 5.896   Mean   :12.56   Mean   :17.02   Mean   : 5.105
##  3rd Qu.: 6.880   3rd Qu.:14.00   3rd Qu.:26.00   3rd Qu.: 7.000
##  Max.   :24.980   Max.   :18.00   Max.   :51.00   Max.   :44.000
##     nonwhite          female          married           numdep
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.0000   Median :1.0000   Median :1.000
##  Mean   :0.1027   Mean   :0.4791   Mean   :0.6084   Mean   :1.044
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:2.000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :6.000
##       smsa           northcen         south             west
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.7224   Mean   :0.251   Mean   :0.3555   Mean   :0.1692
##  3rd Qu.:1.0000   3rd Qu.:0.750   3rd Qu.:1.0000   3rd Qu.:0.0000
```

```
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##     construc          ndurman          trcommpu          trade
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.04563   Mean   :0.1141   Mean   :0.04373   Mean   :0.2871
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##     services         profserv         profocc          clerocc
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1008   Mean   :0.2586   Mean   :0.3669   Mean   :0.1673
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##     servocc           lwage            expersq          tenursq
##  Min.   :0.0000   Min.   :-0.6349   Min.   :   1.0   Min.   :   0.00
##  1st Qu.:0.0000   1st Qu.: 1.2030   1st Qu.:  25.0   1st Qu.:   0.00
##  Median :0.0000   Median : 1.5369   Median : 182.5   Median :   4.00
##  Mean   :0.1407   Mean   : 1.6233   Mean   : 473.4   Mean   :  78.15
##  3rd Qu.:0.0000   3rd Qu.: 1.9286   3rd Qu.: 676.0   3rd Qu.:  49.00
##  Max.   :1.0000   Max.   : 3.2181   Max.   :2601.0   Max.   :1936.00
```

```
stargazer (dt.wages, type="text")
```

```
##
## ===============================================================
## Statistic  N   Mean   St. Dev.  Min   Pctl(25) Pctl(75)  Max
## ---------------------------------------------------------------
## wage      526  5.896   3.693   0.530   3.330    6.880   24.980
## educ      526 12.563   2.769     0       12       14      18
## exper     526 17.017   13.572    1       5        26      51
## tenure    526  5.105   7.224     0       0        7       44
## nonwhite  526  0.103   0.304     0       0        0       1
## female    526  0.479   0.500     0       0        1       1
## married   526  0.608   0.489     0       0        1       1
## numdep    526  1.044   1.262     0       0        2       6
## smsa      526  0.722   0.448     0       0        1       1
## northcen  526  0.251   0.434     0       0       0.8      1
## south     526  0.356   0.479     0       0        1       1
## west      526  0.169   0.375     0       0        0       1
## construc  526  0.046   0.209     0       0        0       1
## ndurman   526  0.114   0.318     0       0        0       1
## trcommpu  526  0.044   0.205     0       0        0       1
## trade     526  0.287   0.453     0       0        1       1
## services  526  0.101   0.301     0       0        0       1
## profserv  526  0.259   0.438     0       0        1       1
## profocc   526  0.367   0.482     0       0        1       1
## clerocc   526  0.167   0.374     0       0        0       1
## servocc   526  0.141   0.348     0       0        0       1
```

```
## lwage     526  1.623   0.532   -0.635  1.203    1.929   3.218
## expersq   526 473.435 616.045    1       25      676    2,601
## tenursq   526 78.150  199.435    0       0        49    1,936
## ----------------------------------------------------------

stargazer (dt.wages, type="text", iqr=TRUE)

##
## ==========================================================
## Statistic  N   Mean   St. Dev.  Min   Pctl(25) Pctl(75)  Max
## ----------------------------------------------------------
## wage      526  5.896   3.693   0.530   3.330    6.880   24.980
## educ      526 12.563   2.769     0       12       14      18
## exper     526 17.017   13.572    1       5        26      51
## tenure    526  5.105   7.224     0       0        7       44
## nonwhite  526  0.103   0.304     0       0        0       1
## female    526  0.479   0.500     0       0        1       1
## married   526  0.608   0.489     0       0        1       1
## numdep    526  1.044   1.262     0       0        2       6
## smsa      526  0.722   0.448     0       0        1       1
## northcen  526  0.251   0.434     0       0       0.8      1
## south     526  0.356   0.479     0       0        1       1
## west      526  0.169   0.375     0       0        0       1
## construc  526  0.046   0.209     0       0        0       1
## ndurman   526  0.114   0.318     0       0        0       1
## trcommpu  526  0.044   0.205     0       0        0       1
## trade     526  0.287   0.453     0       0        1       1
## services  526  0.101   0.301     0       0        0       1
## profserv  526  0.259   0.438     0       0        1       1
## profocc   526  0.367   0.482     0       0        1       1
## clerocc   526  0.167   0.374     0       0        0       1
## servocc   526  0.141   0.348     0       0        0       1
## lwage     526  1.623   0.532   -0.635  1.203    1.929   3.218
## expersq   526 473.435 616.045    1       25      676    2,601
## tenursq   526 78.150  199.435    0       0        49    1,936
## ----------------------------------------------------------

ncol(dt.wages)

## [1] 24

table(dt.wages[, list(female, nonwhite)])

##        nonwhite
## female   0   1
##      0 245  29
##      1 227  25

table(dt.wages[, list (female, nonwhite, south)])

## , , south = 0
##
```

```
##        nonwhite
## female   0   1
##      0 158  13
##      1 154  14
##
## , , south = 1
##
##        nonwhite
## female   0   1
##      0  87  16
##      1  73  11
```

```
table(dt.wages[, list (female , tenure)])
```

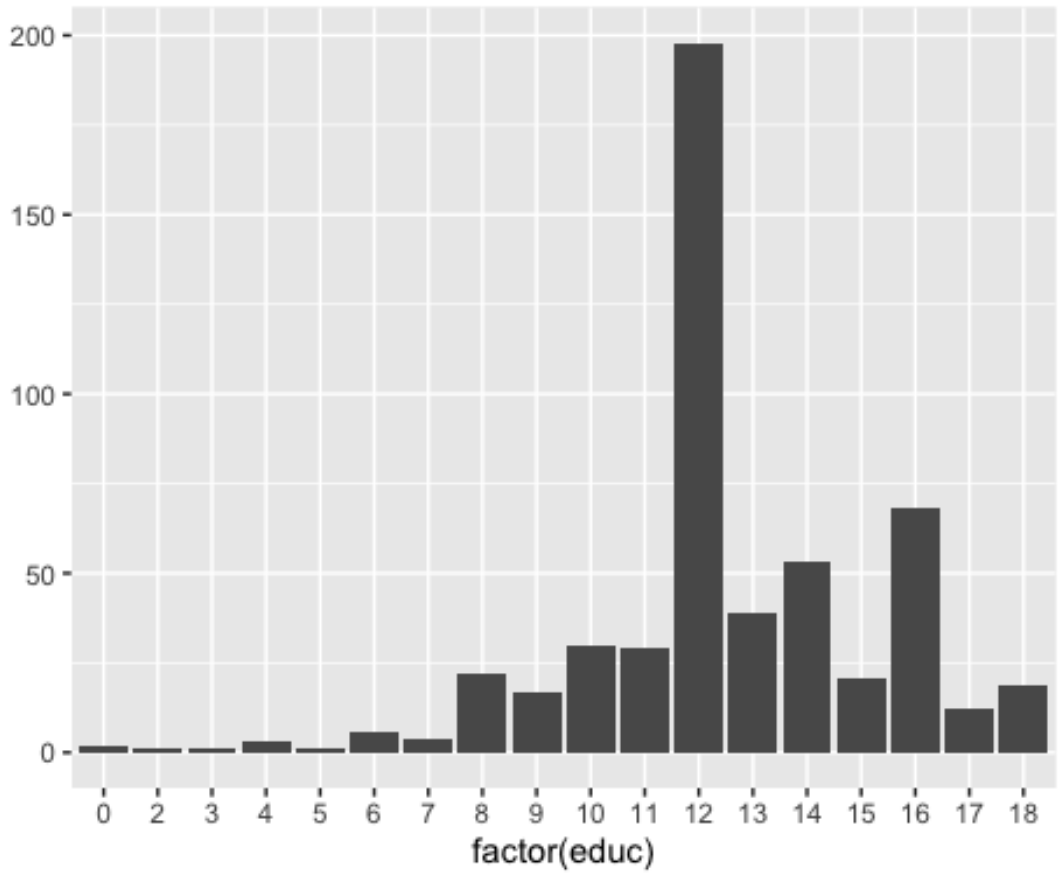
```
##        tenure
## female  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21
22 23
##      0 75 19 32 22  8 19  7 11 11  5  7  5  4  8  1  2  2  1  3  2  2  4
2  3
##      1 88 32 31 20 19 11  8  4  4  4  6  4  4  0  1  4  2  2  0  0  3  1
0  0
##        tenure
## female 24 25 26 28 30 31 33 34 39 44
##      0  4  3  2  2  3  2  1  0  1  1
##      1  1  1  1  0  0  0  0  1  0  0
```

```
table(dt.wages[, list (nonwhite , tenure)])
```

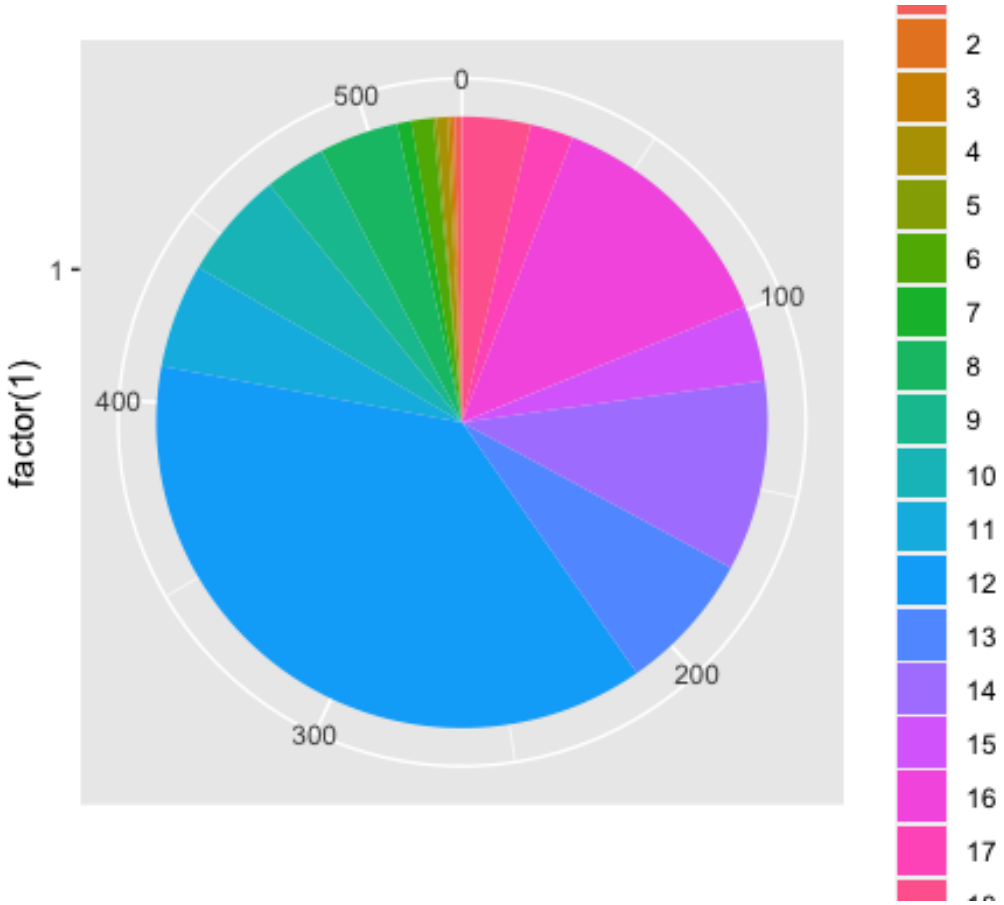
```
##         tenure
## nonwhite   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
16
##        0 151  48  54  35  24  26  13  13  14   6  12   9   8   7   2   5
4
##        1  12   3   9   7   3   4   2   2   1   3   1   0   0   1   0   1
0
##         tenure
## nonwhite  17  18  19  20  21  22  23  24  25  26  28  30  31  33  34  39
44
##        0   3   2   2   5   4   2   2   4   4   3   2   2   2   1   1   1
1
##        1   0   1   0   0   1   0   1   1   0   0   0   1   0   0   0   0
0
```

c.Reproduce the data-handling steps shown in the slides (and dt.wages)

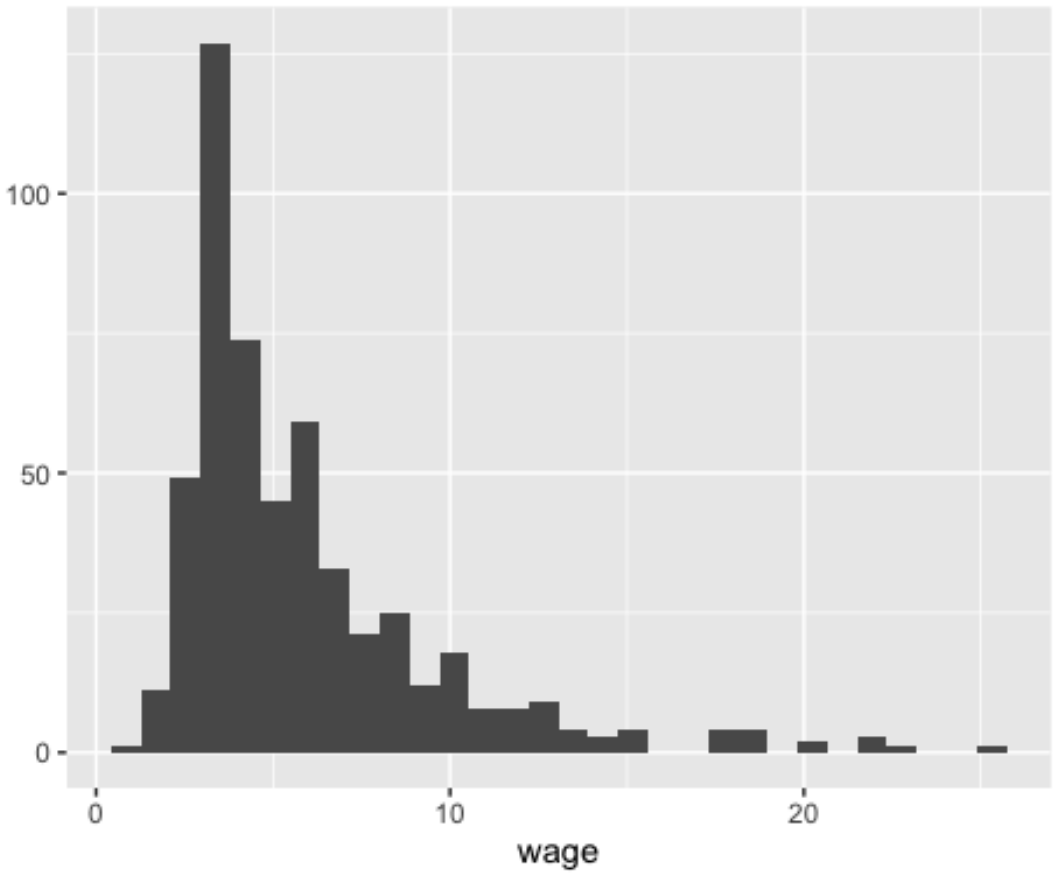
```
qplot(factor(educ),data=dt.wages,geom="bar")
```



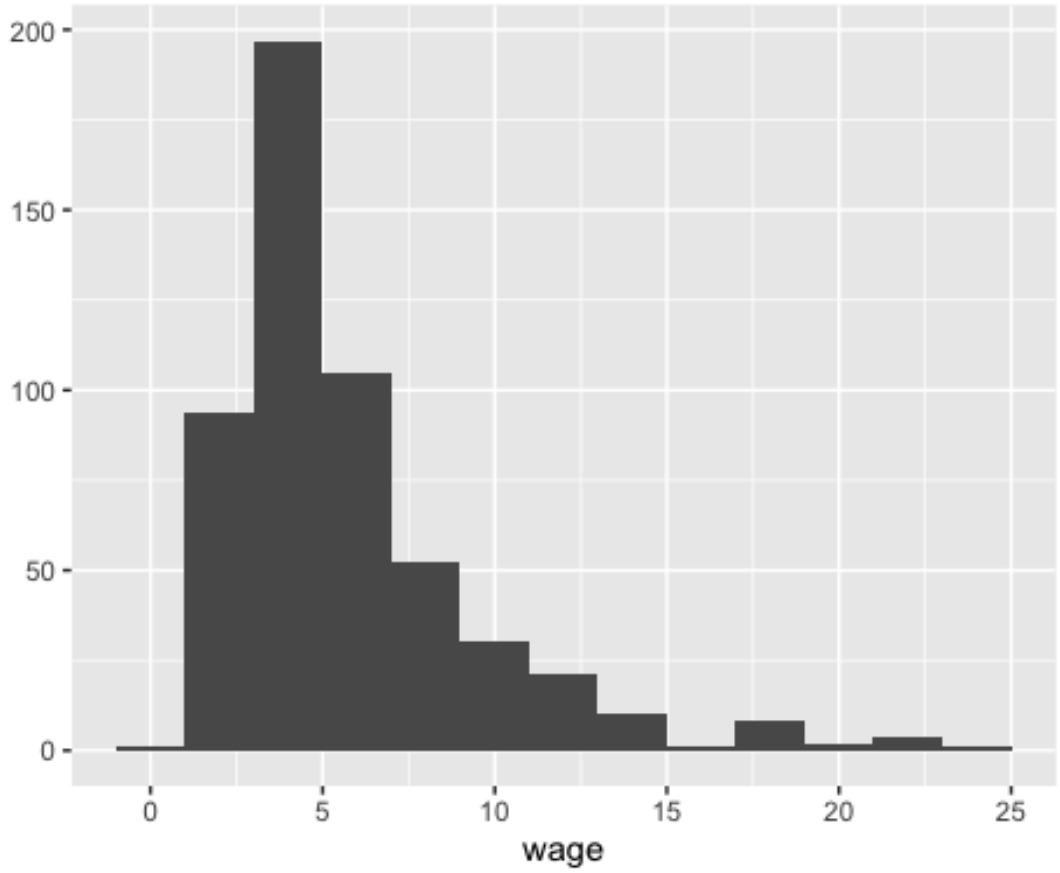

```
qplot(x=factor(1),fill=factor(educ),data=dt.wages,geom="bar") +
coord_polar(theta="y")
```



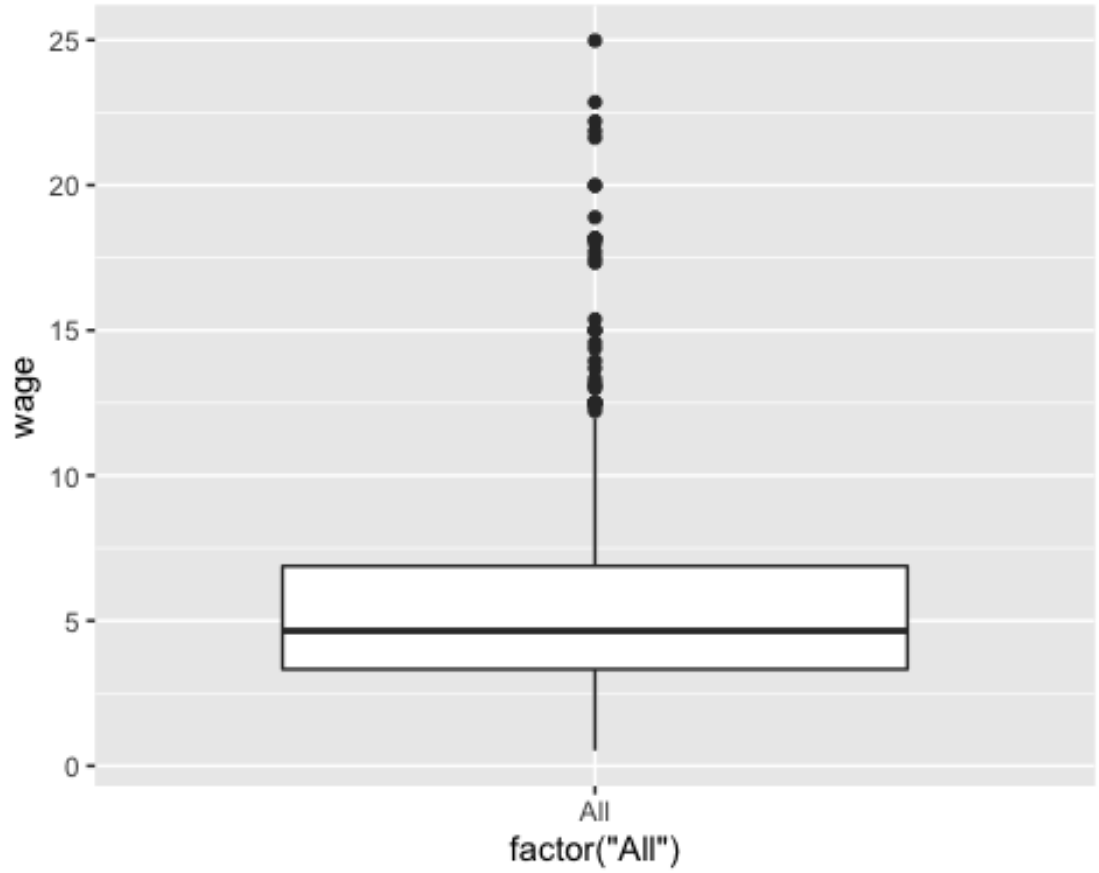

```
qplot(wage,data=dt.wages,geom="histogram")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



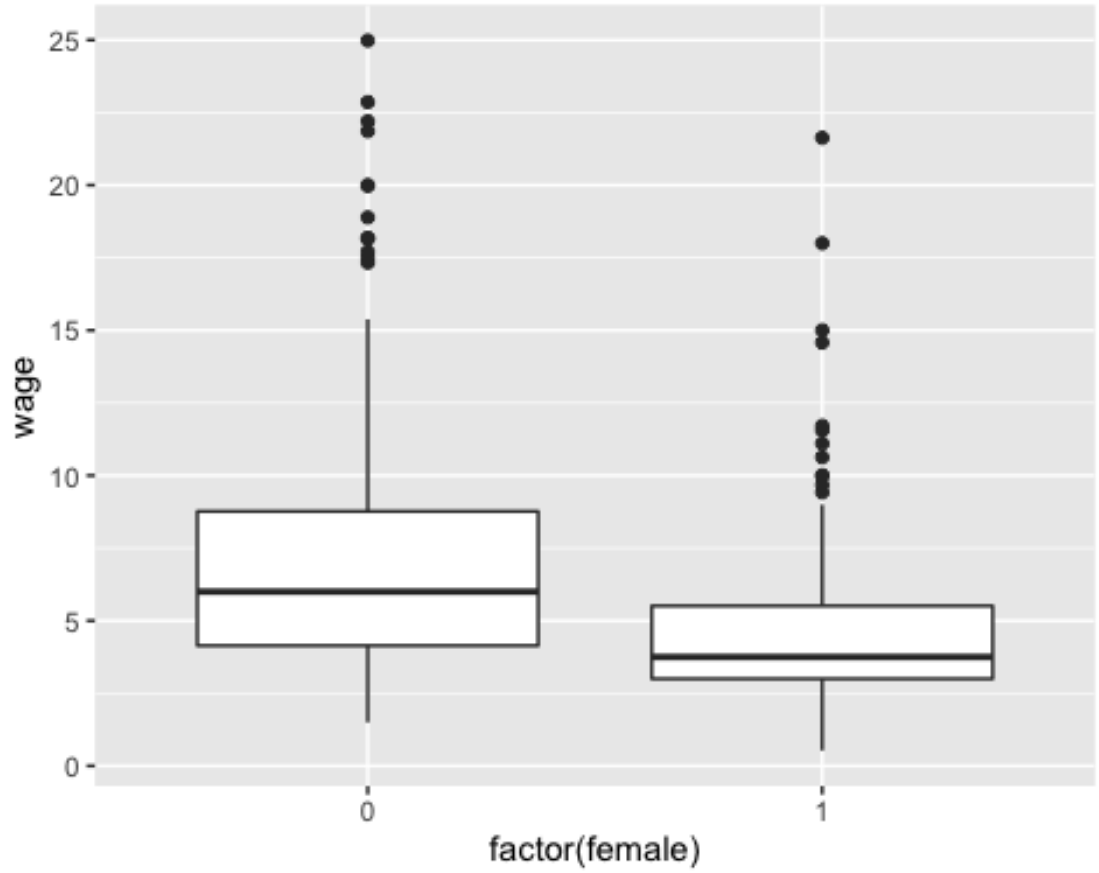

```
qplot(wage,data=dt.wages,geom="histogram",binwidth=2)
```



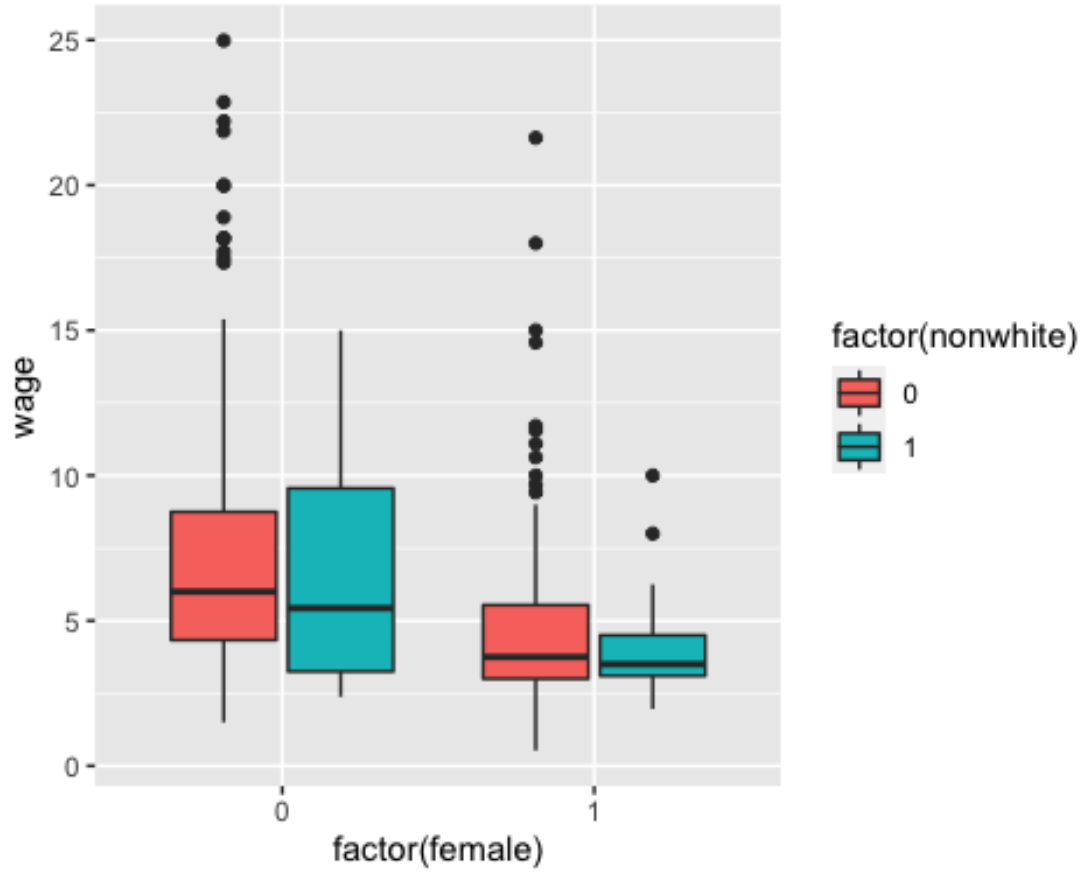

```
qplot(x=factor("All"),y=wage,data=dt.wages,geom="boxplot")
```



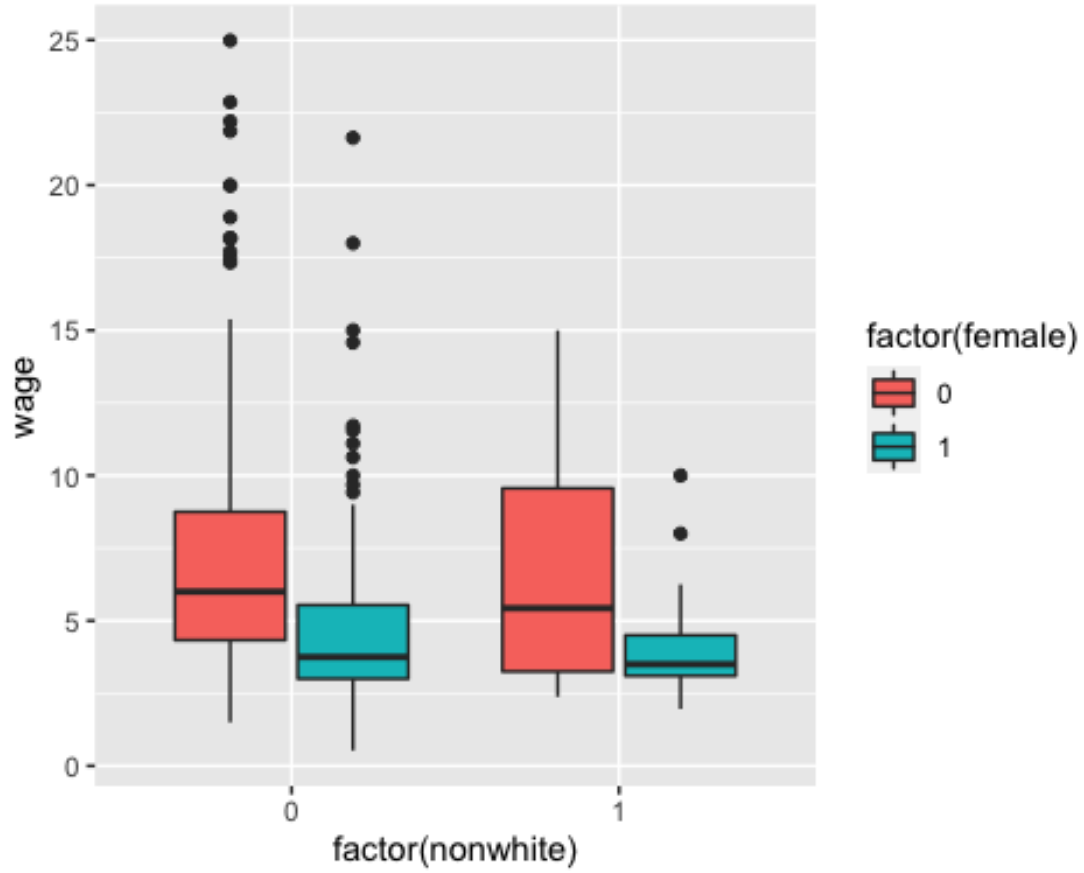

```
qplot(factor(female),wage,data=dt.wages,geom="boxplot")
```



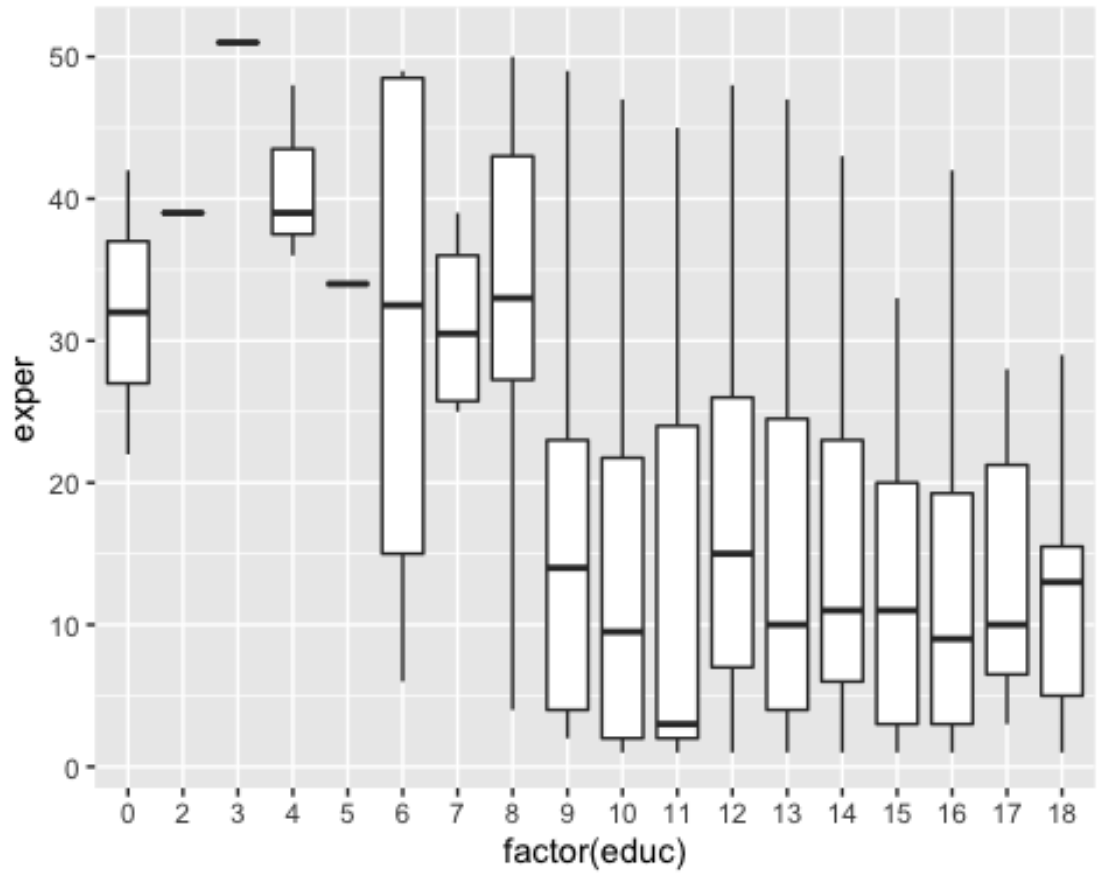

```
qplot(factor(female),wage,fill=factor(nonwhite),data=dt.wages,geom="boxplot")
```



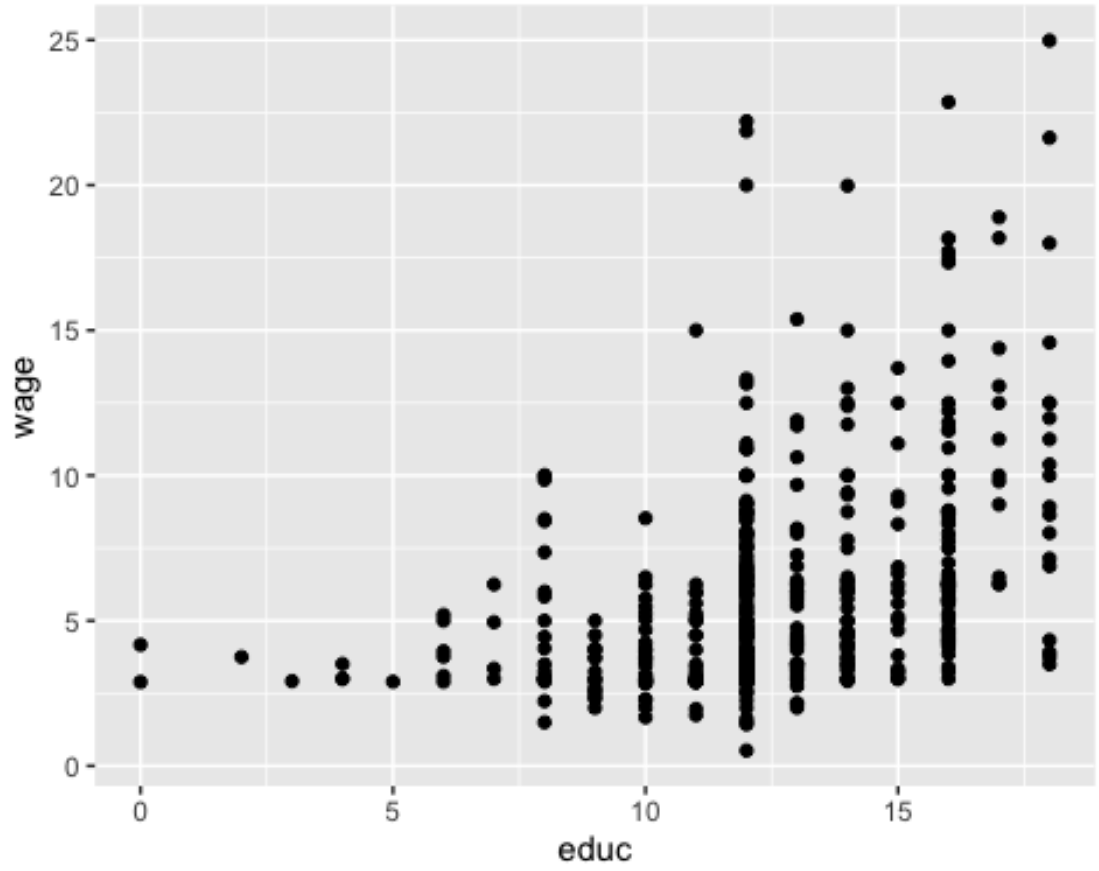

```
qplot(factor(nonwhite),wage,fill=factor(female),data=dt.wages,geom="boxplot")
```



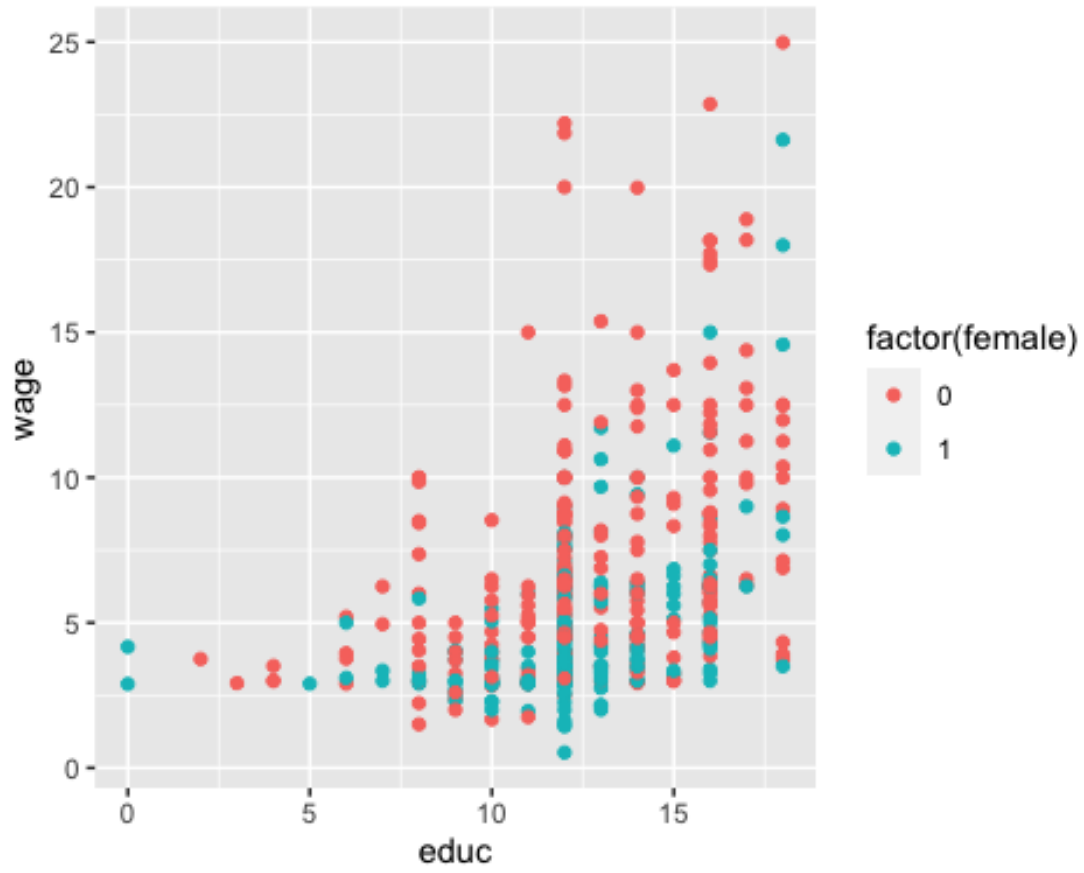

```
ggplot(dt.wages)+geom_boxplot(aes(factor(educ),exper))
```



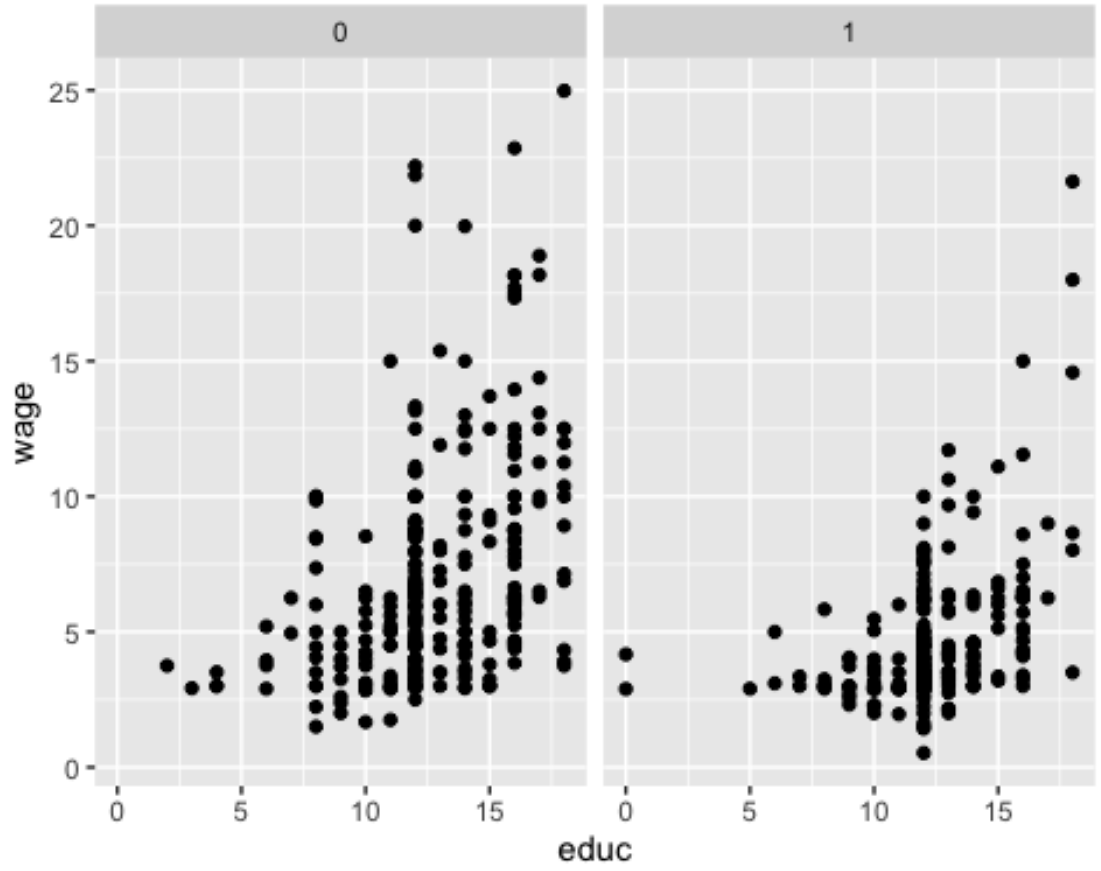

```
qplot(educ,wage,data=dt.wages,geom="point")
```



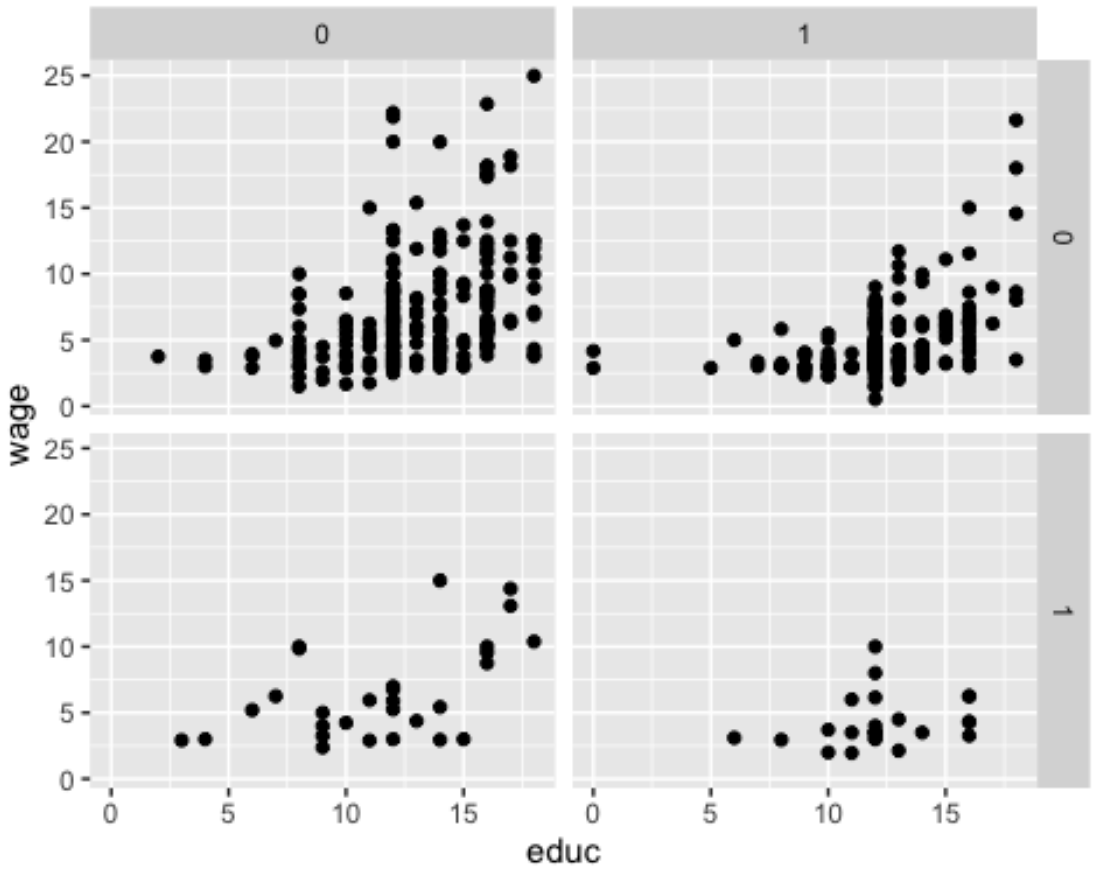

```
qplot(educ,wage,color=factor(female),data=dt.wages,geom="point")
```



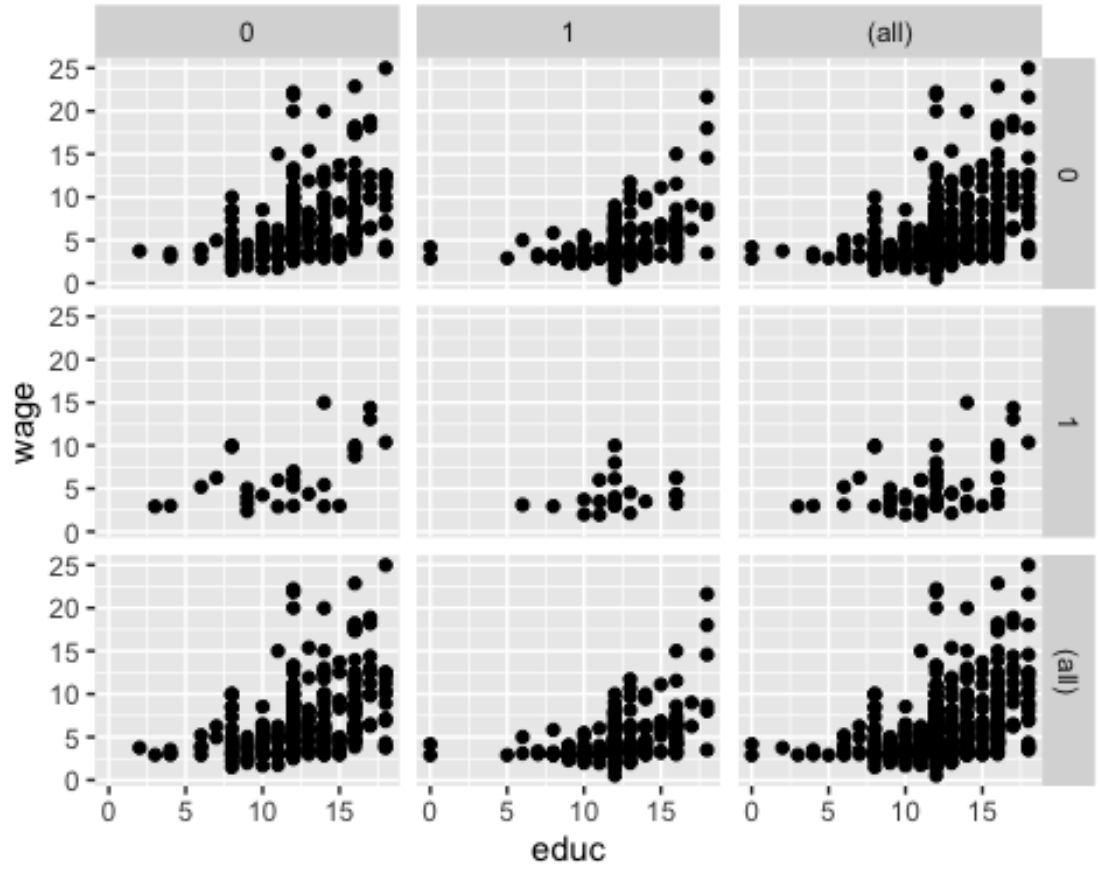

```
ggplot(dt.wages)+geom_point(aes(educ,wage))+facet_grid(~female)
```



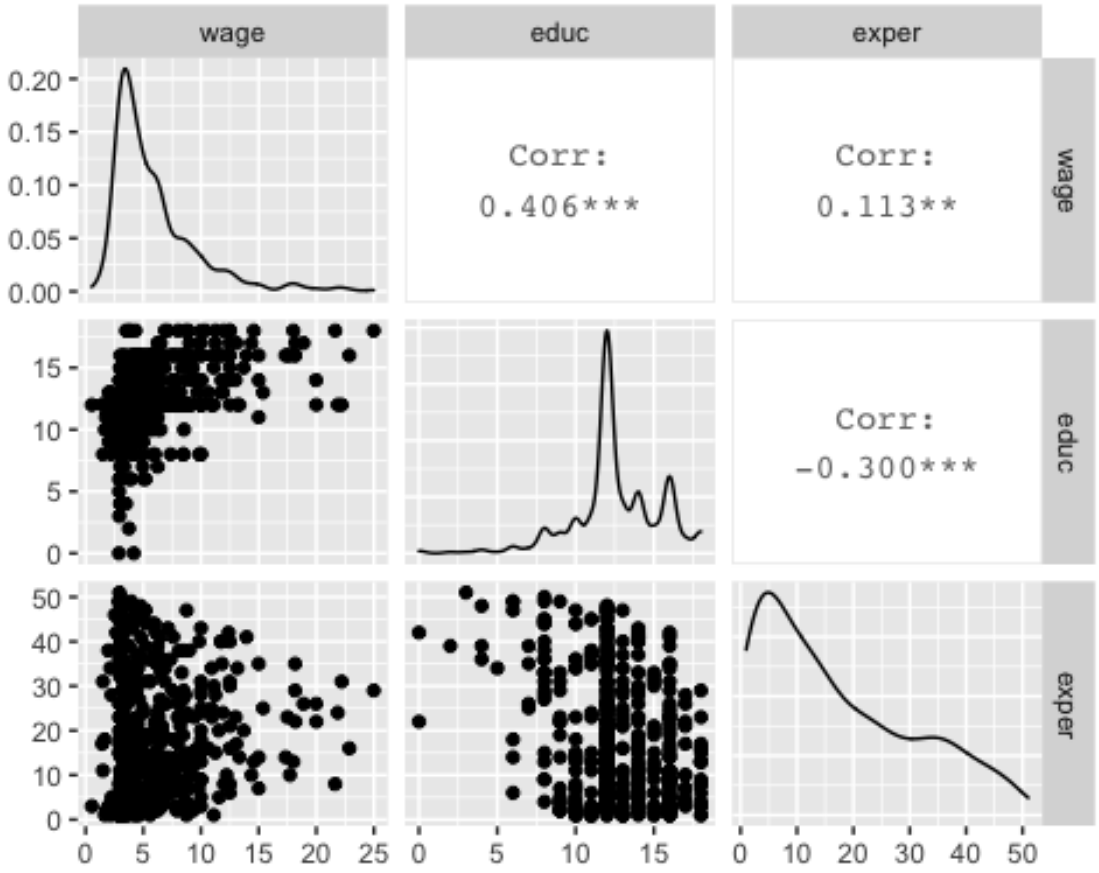

```
ggplot(dt.wages)+geom_point(aes(educ,wage))+facet_grid(nonwhite~female)
```



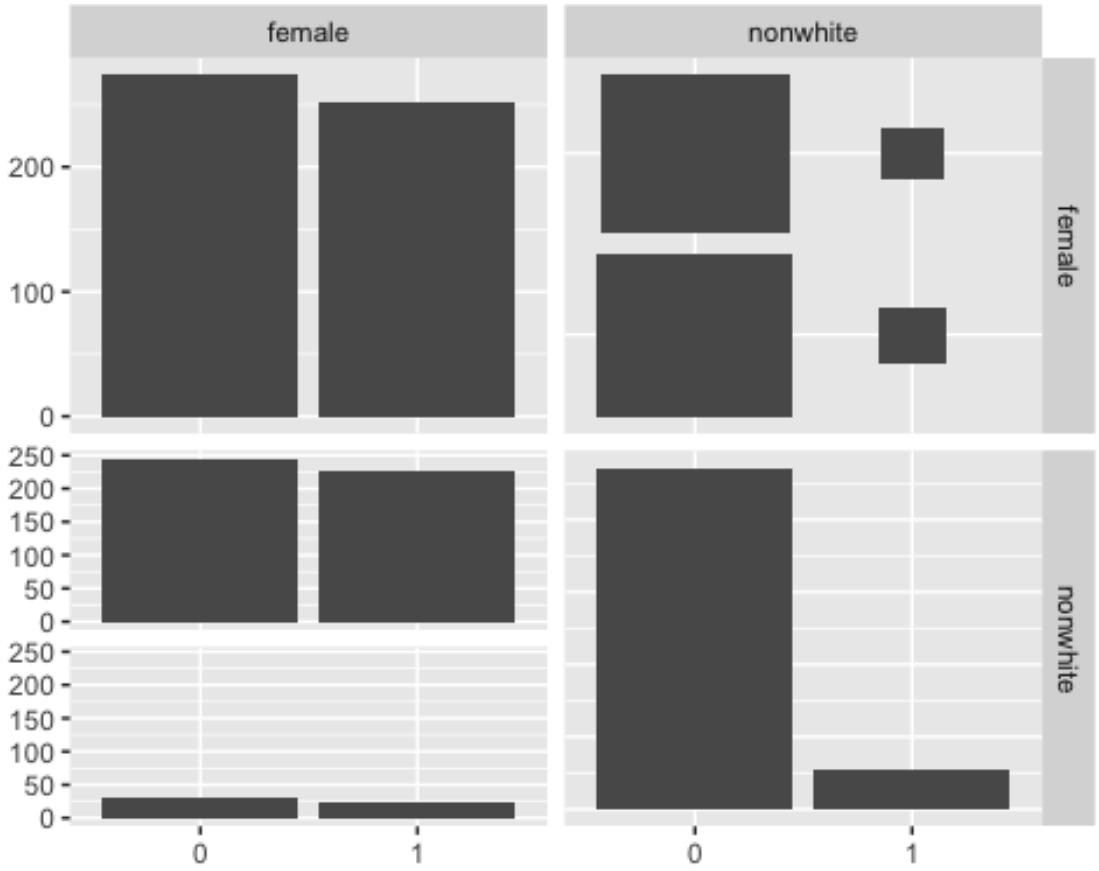

```
ggplot(dt.wages)+geom_point(aes(educ,wage))+facet_grid(nonwhite~female,
margins=TRUE)
```




```
ggpairs(dt.wages[,list(wage,educ,exper)])
```

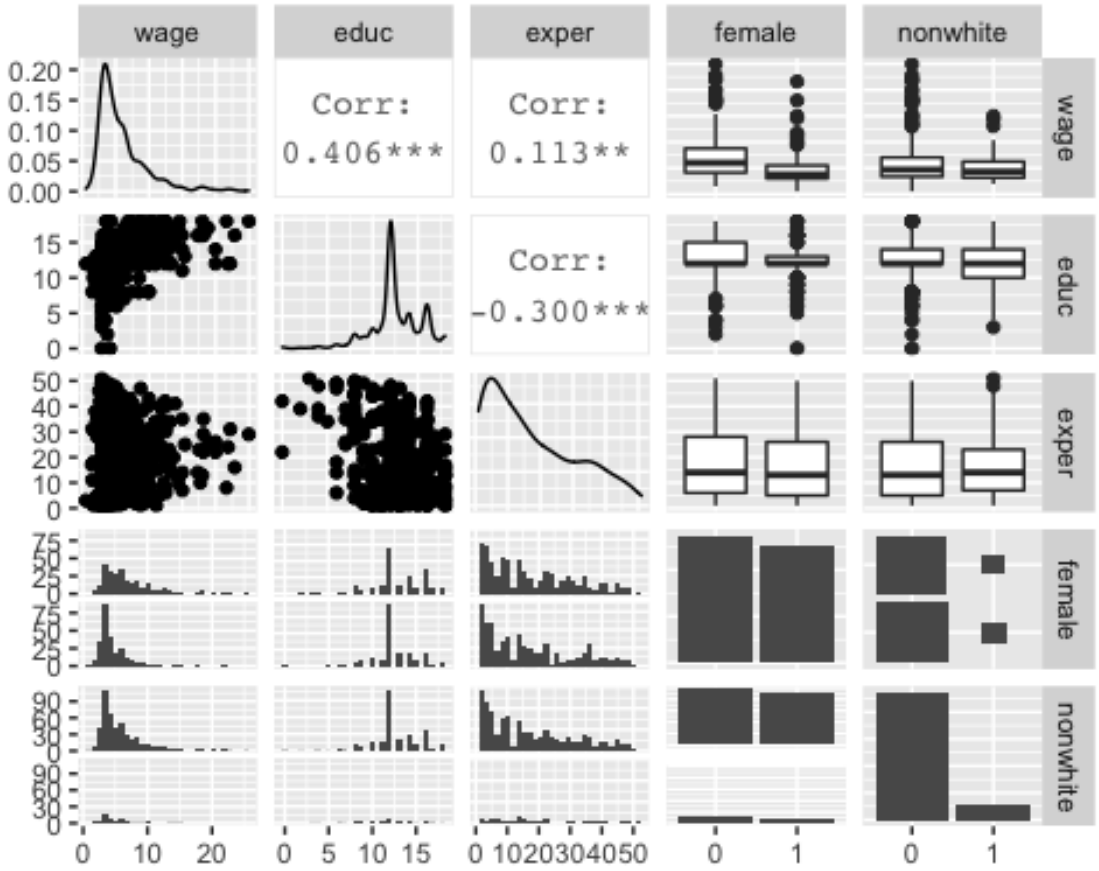



```
ggpairs(dt.wages[,list(female=factor(female),nonwhite=factor(nonwhite))])
```




```
ggpairs(dt.wages[,list(wage,educ,exper,female=factor(female),nonwhite=factor(
nonwhite))])
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



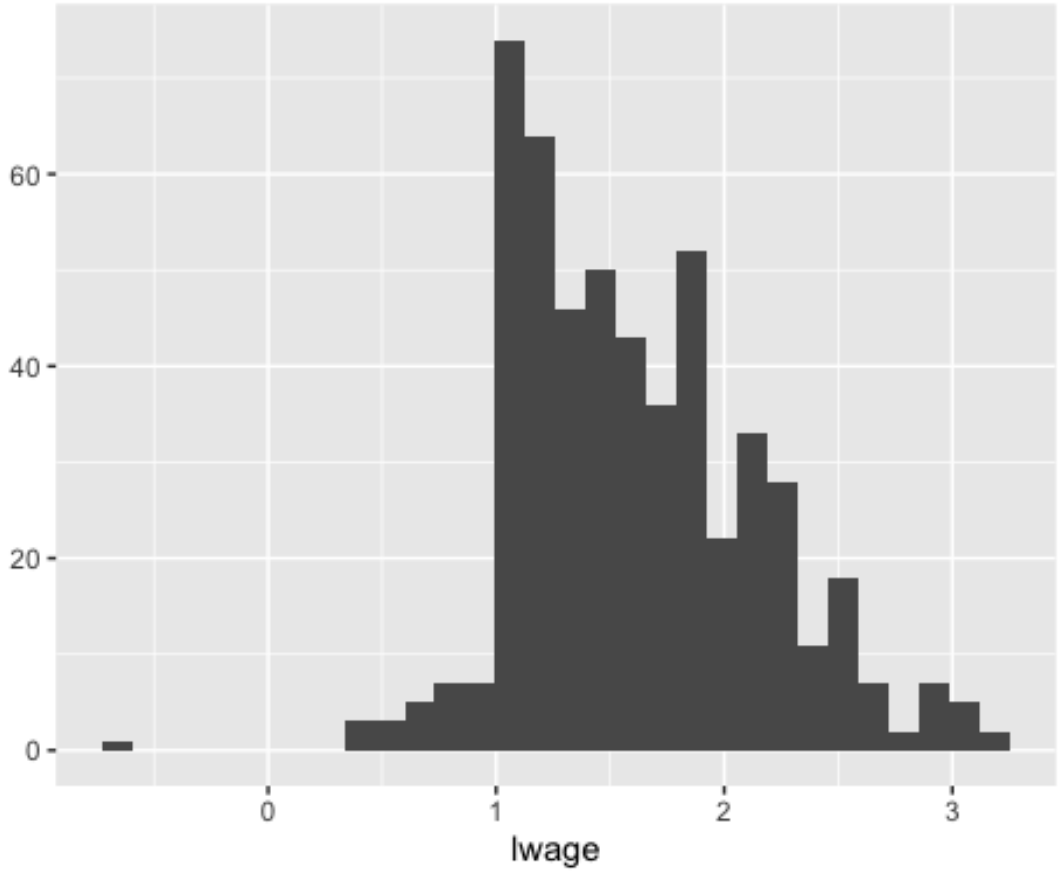

```
dt.wages[wage>20,][order(-wage)]
```

```
##     wage educ exper tenure nonwhite female married numdep smsa northcen
south
## 1: 24.98   18    29     25        0      0       1      0    1        0
0
## 2: 22.86   16    16      7        0      0       1      2    1        0
0
## 3: 22.20   12    31     15        0      0       1      1    1        0
0
## 4: 21.86   12    24     16        0      0       1      3    1        1
0
## 5: 21.63   18     8      8        0      1       0      0    1        0
0
##    west construc ndurman trcommpu trade services profserv profocc clerocc
## 1:    0        0       0        0     0        0        0       1       0
## 2:    0        0       0        0     0        0        0       1       0
## 3:    1        0       0        0     0        0        0       1       0
## 4:    0        0       0        0     1        0        0       1       0
## 5:    0        0       0        0     0        0        1       1       0
##    servocc    lwage expersq tenursq
## 1:       0 3.218076     841     625
## 2:       0 3.129389     256      49
```

```
## 3:       0 3.100092     961     225
## 4:       0 3.084659     576     256
## 5:       0 3.074081      64      64
```

```
qplot(lwage,data=dt.wages,geom="histogram")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
dt.wages[lwage< 0,]
```

```
##    wage educ exper tenure nonwhite female married numdep smsa northcen
south
## 1: 0.53   12     3      1        0      1       0      0    1        0
0
##    west construc ndurman trcommpu trade services profserv profocc clerocc
## 1:    1        0       0        0     0        1        0       0       0
##    servocc      lwage expersq tenursq
## 1:       1 -0.6348783       9       1
```

3) Let's do a thought experiment: Using the slide-deck UEA_ecoR2PhD CoreLect_06 ATENT_Match_Stk and dt.wages, let's pretend, for a moment "south," is a randomly assigned treatment in an experiment:

a.Compute a difference-in-means estimator when treatment is "south," and the outcome is wage

```
dt.wages[south==1, mean(wage)] - dt.wages[south==0, mean(wage)]
```

```
## [1] -0.7900927
```

b.Now focus on race and gender as control variables (in "x") and run a regression estimation of treatment effects

```
reg1 <- lm(wage ~ south + nonwhite + female, data=dt.wages)
stargazer(reg1, type = "text")
```

```
##
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                                 wage
## ------------------------------------------------
## south                          -0.884***
##                                 (0.317)
##
## nonwhite                        -0.372
##                                 (0.499)
##
## female                         -2.552***
##                                 (0.302)
##
## Constant                       7.471***
##                                 (0.243)
##
## ------------------------------------------------
## Observations                      526
## R2                               0.130
## Adjusted R2                      0.125
## Residual Std. Error       3.454 (df = 522)
## F Statistic           26.104*** (df = 3; 522)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

c.Now try to estimate the regression and account for potentially heterogeneous treatment effects.

```
reg2 <- lm(wage ~ south + nonwhite + female + nonwhite*south + female*south,
data=dt.wages)
stargazer(reg2, type = "text")
```

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                  wage
## ------------------------------------------------
## south                           -1.288***
##                                  (0.447)
##
## nonwhite                         0.155
##                                  (0.691)
##
## female                          -2.953***
##                                  (0.374)
##
## south:nonwhite                   -1.047
##                                  (0.996)
##
## south:female                     1.117*
##                                  (0.630)
##
## Constant                        7.628***
##                                  (0.269)
##
## ------------------------------------------------
## Observations                      526
## R2                                0.138
## Adjusted R2                       0.129
## Residual Std. Error       3.446 (df = 520)
## F Statistic           16.591*** (df = 5; 520)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

d.Try to implement a 2-step fitted regression.

```
south.yes <- subset(dt.wages, south==1)
south.no <- subset(dt.wages, south==0)

reg3.1 <- lm(wage ~ nonwhite + female, data=south.yes)
reg3.2 <- lm(wage ~ nonwhite + female, data=south.no)

stargazer(reg3.1, reg3.2, type = "text")
```

```
##
## ===================================================================
##                                   Dependent variable:
##                      ---------------------------------------------
##                                          wage
##                                (1)                    (2)
## -------------------------------------------------------------------
## nonwhite                      -0.892                  0.155
```

```
##                             (0.617)                 (0.739)
##
## female                     -1.836***               -2.953***
##                             (0.436)                 (0.400)
##
## Constant                    6.341***                7.628***
##                             (0.307)                 (0.287)
##
## ---------------------------------------------------------------------
## Observations                  187                     339
## R2                           0.096                   0.139
## Adjusted R2                  0.086                   0.134
## Residual Std. Error    2.963 (df = 184)        3.684 (df = 336)
## F Statistic         9.724*** (df = 2; 184) 27.234*** (df = 2; 336)
## =====================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

```
dt.wages$y_hat1 <- predict(reg3.1, newdata=dt.wages)
dt.wages$y_hat0 <- predict(reg3.2, newdata=dt.wages)
```