Lab13a

Amir Jafarzadeh

09/11/2021

install.packages("GGally")

```
setwd ("U:/econometrics/lab 13/lab13a")



load ("dt_wages.RData")


list.files()
```

```
## [1] "dt_wages.RData"
## [2] "Lab Home Assignment 3.pdf"
## [3] "Lab02_mkm_Econometrics1_RodBased_DataHandling.pdf"
## [4] "Lab13a.html"
## [5] "Lab13a.Rmd"
```

```
#b 526 obs, 24 variables



#Reproduce the Data-Handling steps shown in the slides (and dt.wages)

#Data Structure



#Structure of the Data

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>
```

```
#Peek at the data

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     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
## 524  4.67   15    13     18        0      0       1      3    0        0     0
## 525 11.56   16     5      1        0      0       1      0    0        0     0
## 526  3.50   14     5      4        1      1       0      2    0        0     0
##     west construc ndurman trcommpu trade services profserv profocc clerocc
## 524    1        1       0        0     0        0        0       1       0
## 525    1        0       1        0     0        0        0       0       0
## 526    1        0       0        0     0        0        1       0       1
##     servocc    lwage expersq tenursq
## 524       0 1.541159     169     324
## 525       0 2.447551      25       1
## 526       0 1.252763      25      16
```

```
#Descriptive Statistics



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
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 4.1.1
```

```
##
## 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(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.1.1
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.1.1
```

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

```
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) # With interquartil
```

```
##
## ===============================================================
## 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
## ----------------------------------------------------------------
```

```
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)])
```

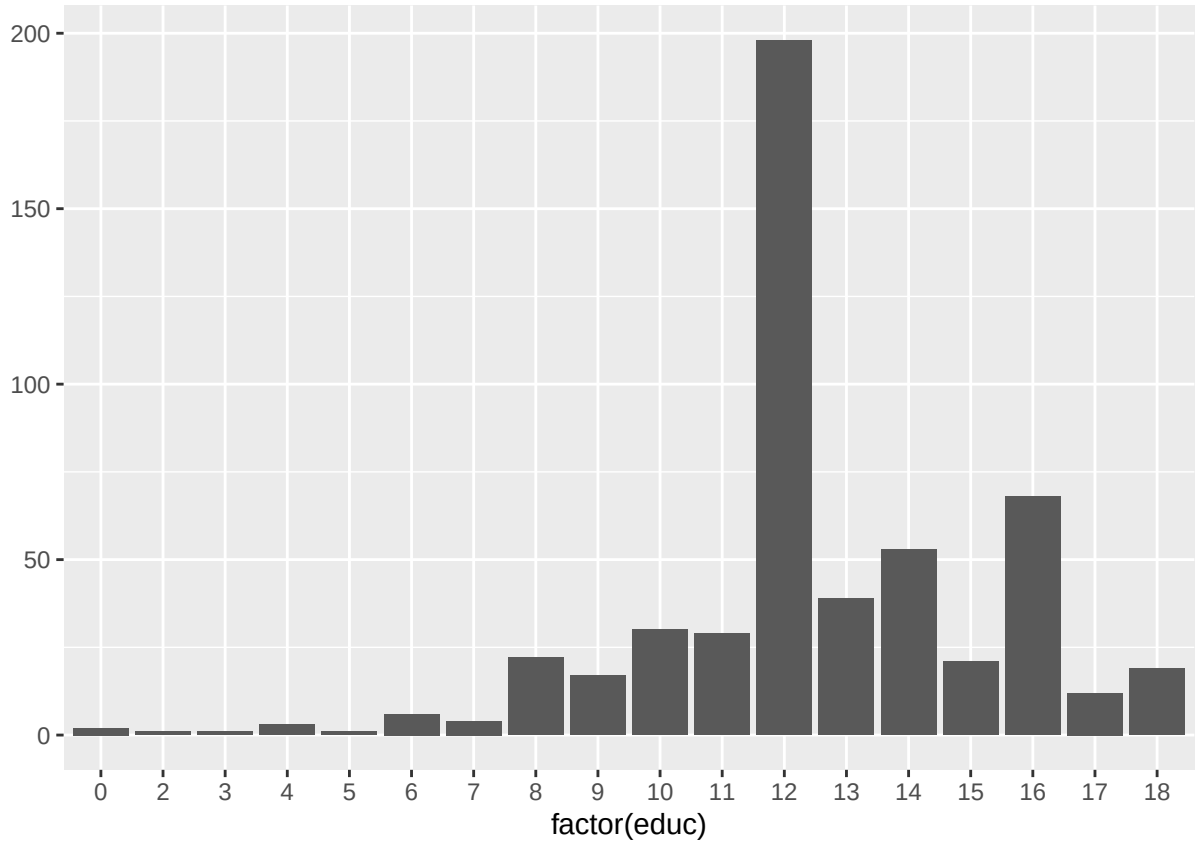
```
##       tenure
## female  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
##      0 75 19 32 22  8 19  7 11 11  5  7  5  4  8  1  2  2  1  3  2  2  4  2  3
##      1 88 32 31 20 19 11  8  4  4  4  6  4  4  0  1  4  2  2  0  0  3  1  0  0
##       tenure
## female 24 25 26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

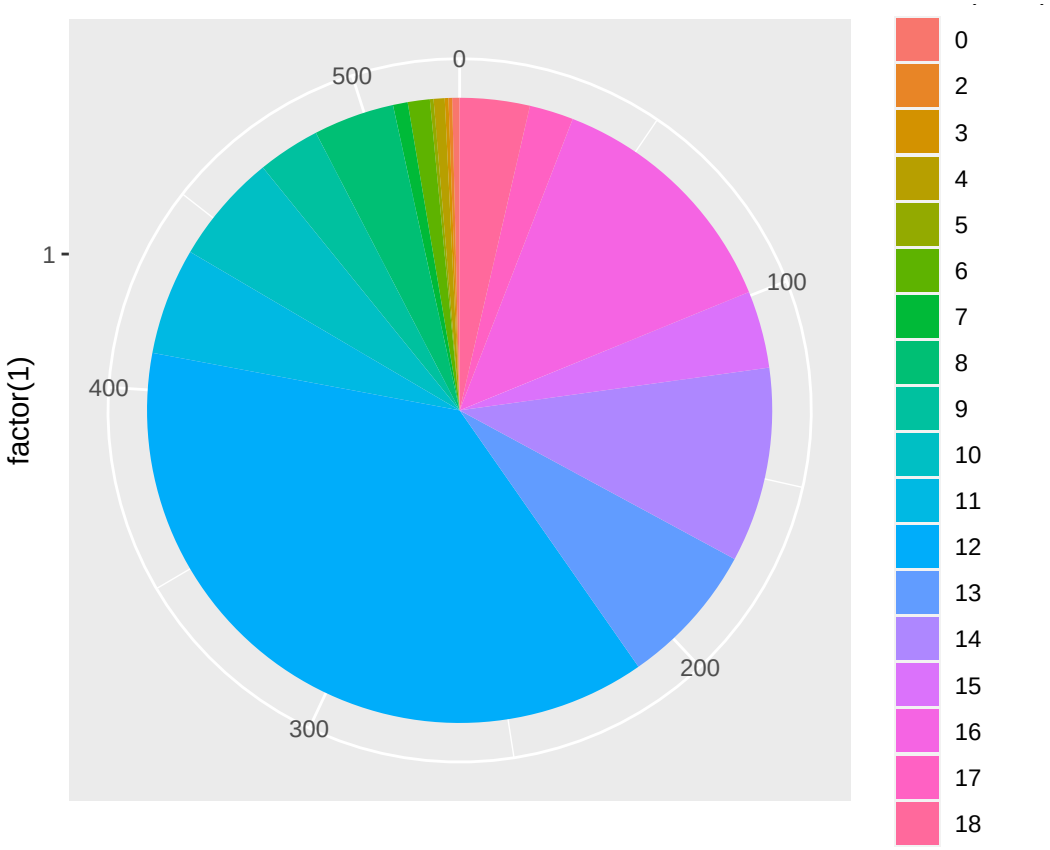
```
##         tenure
## nonwhite   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
##        0 151  48  54  35  24  26  13  13  14   6  12   9   8   7   2   5   4
##        1  12   3   9   7   3   4   2   2   1   3   1   0   0   1   0   1   0
##         tenure
```

```
## 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
```

```
#Graphical Analysis



#Categorial Data: Bar Plot

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

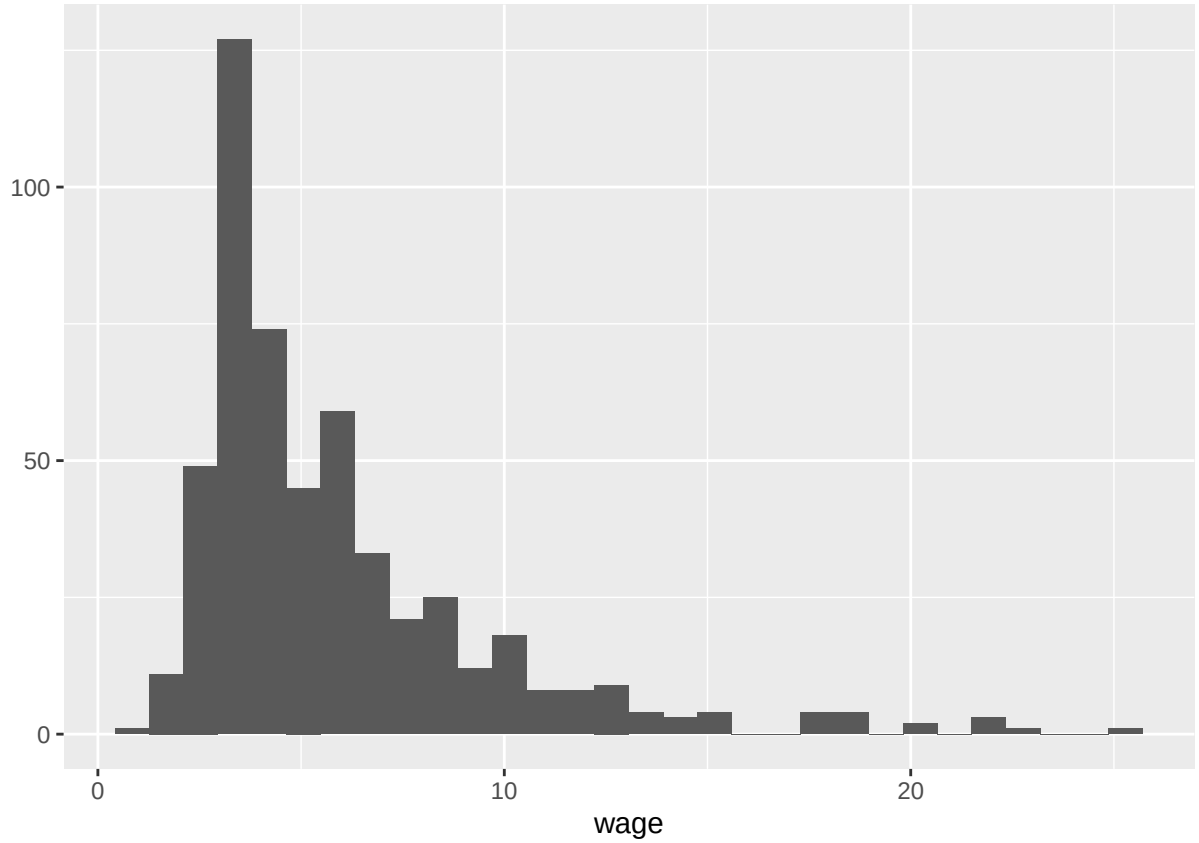



```
#Categorial Data: Pie Chart



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

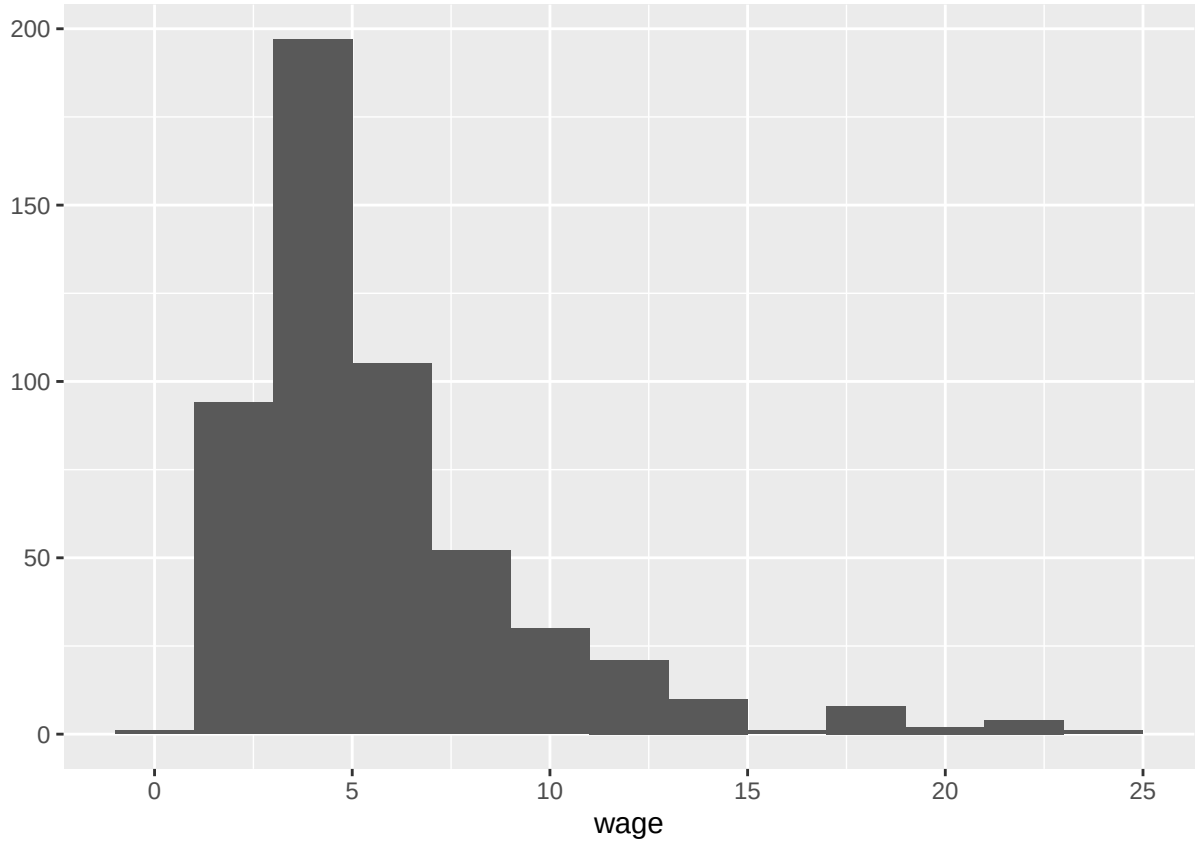



```
#Continus Data: Histogram


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

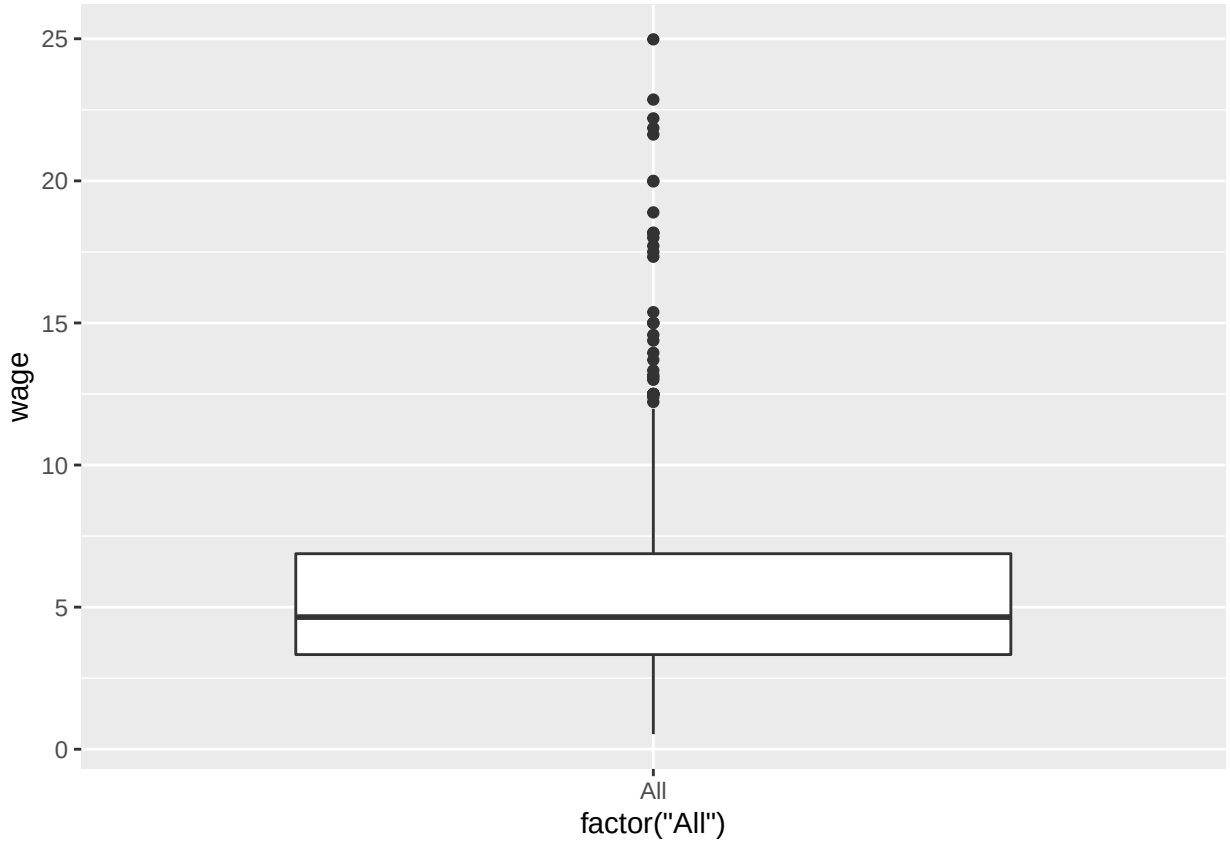
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



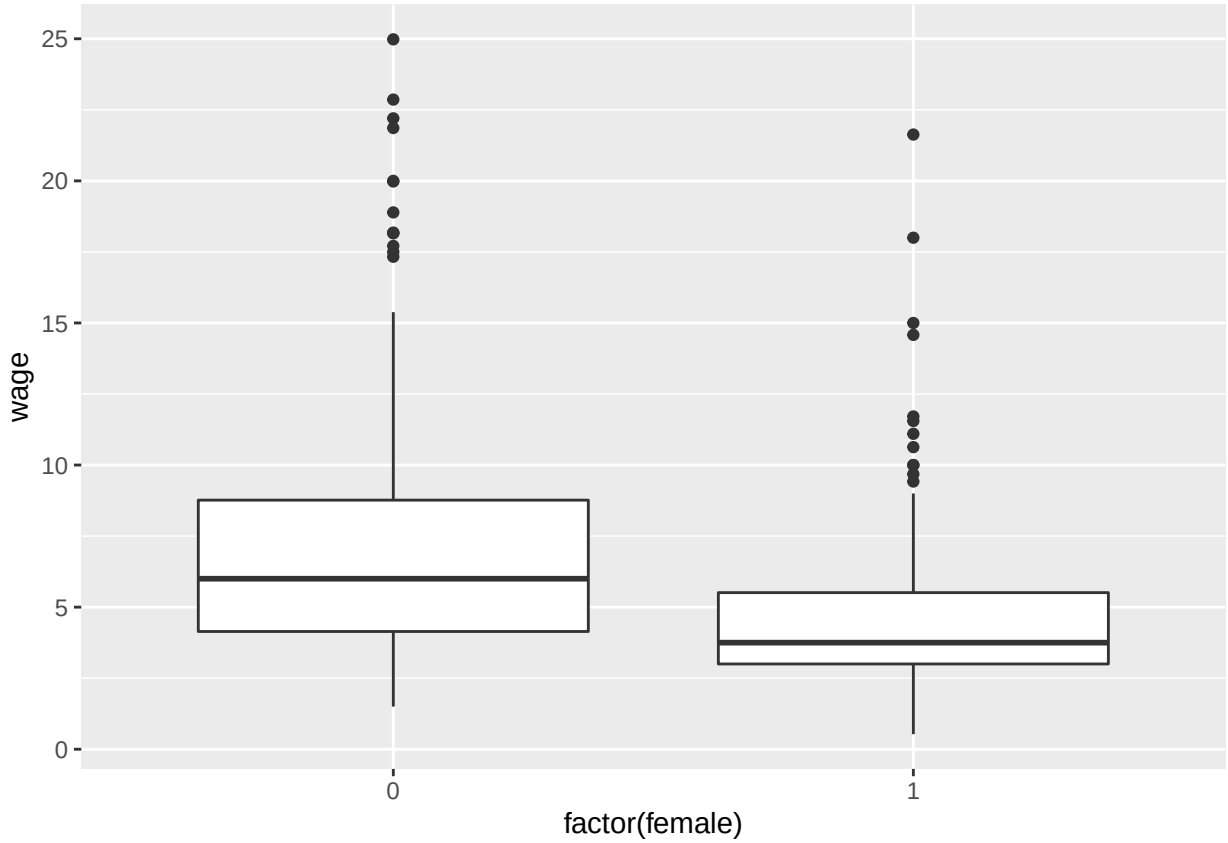

```
qplot (wage, data = dt.wages, geom = "histogram", binwidth=2)
```



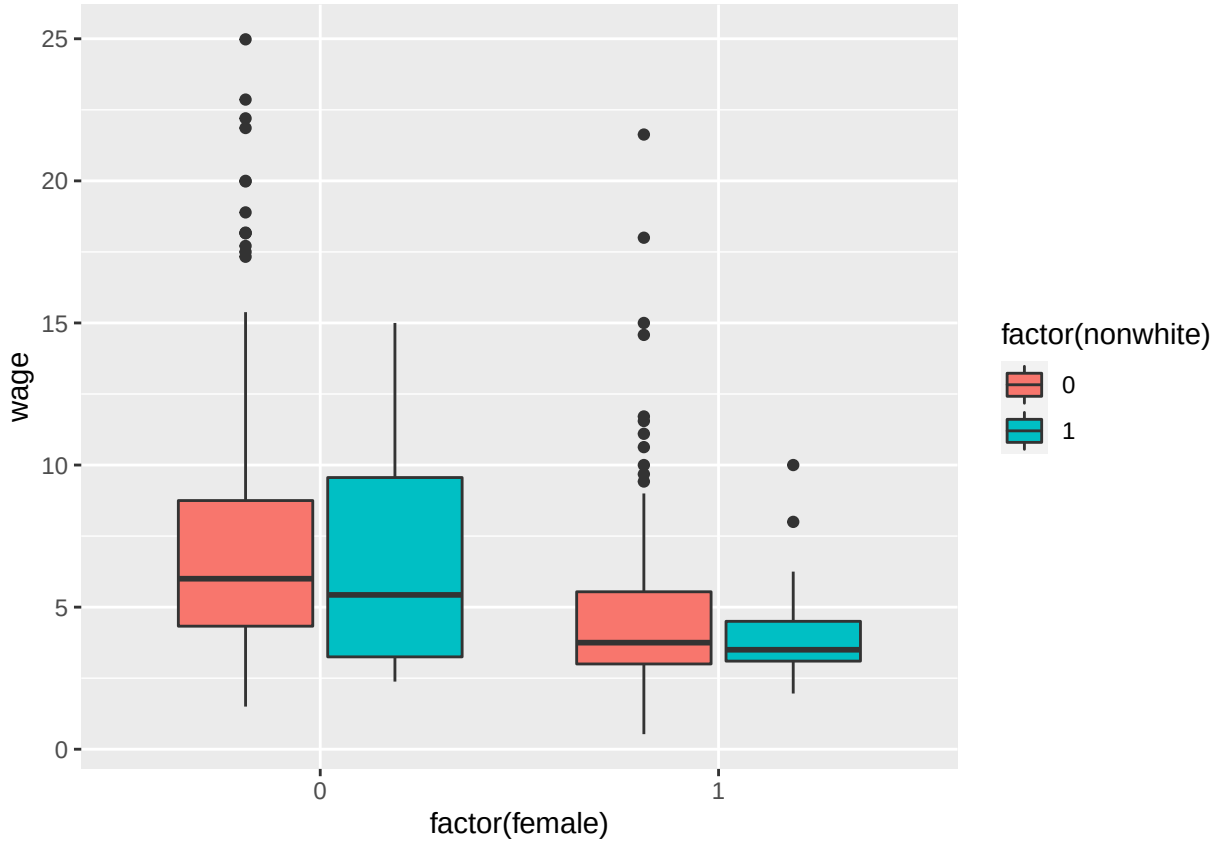

```
#Continuous Data: Boxplot


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



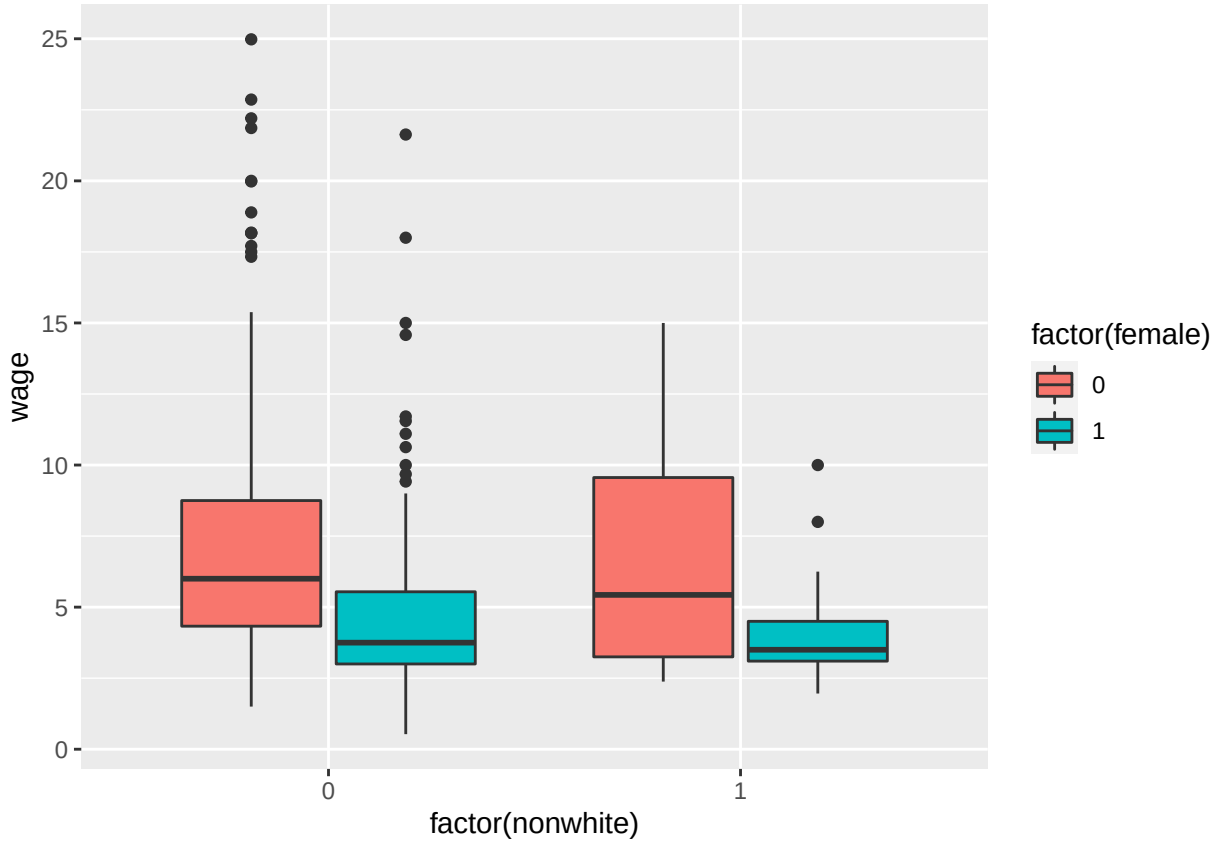

```
#Continuous vs CategoricalData: Boxplot


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



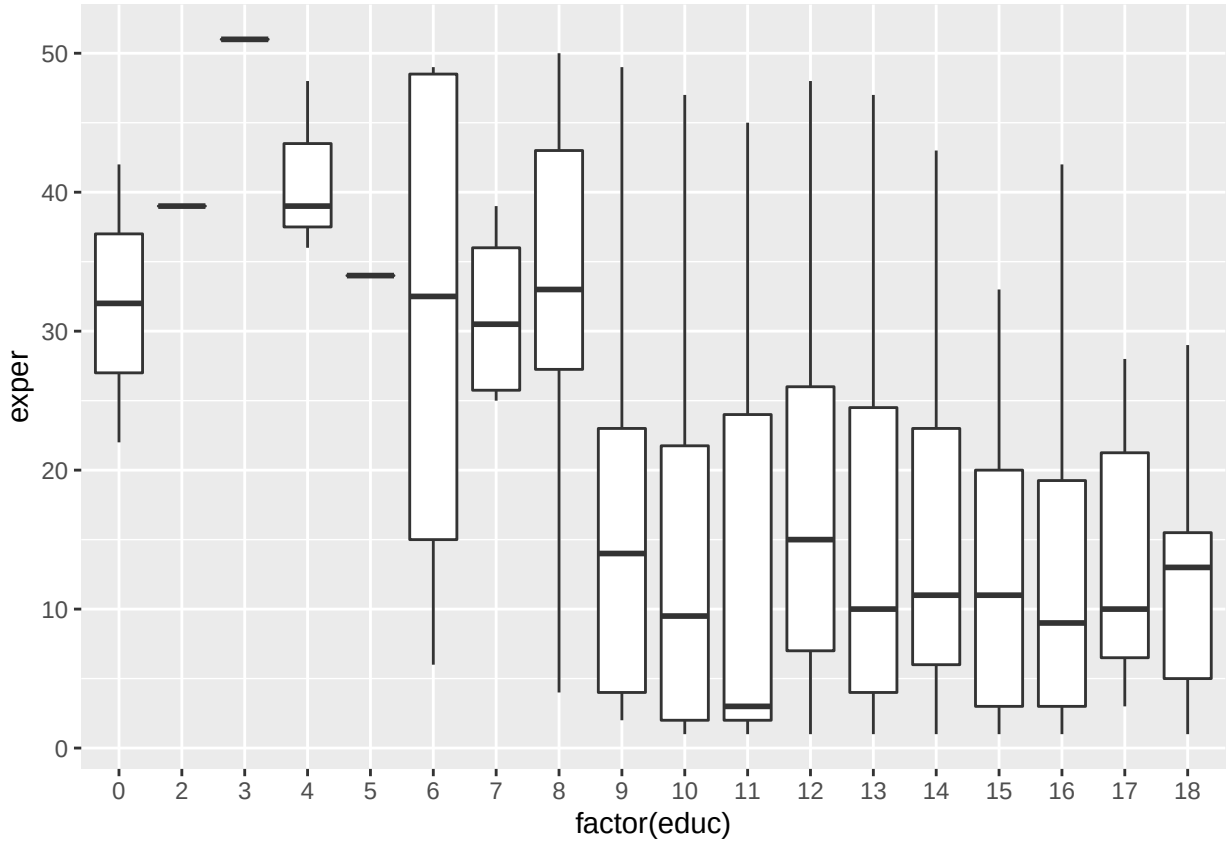

```
qplot(factor(female), wage, fill= factor(nonwhite), data=dt.wages, geom = "boxplot")
```



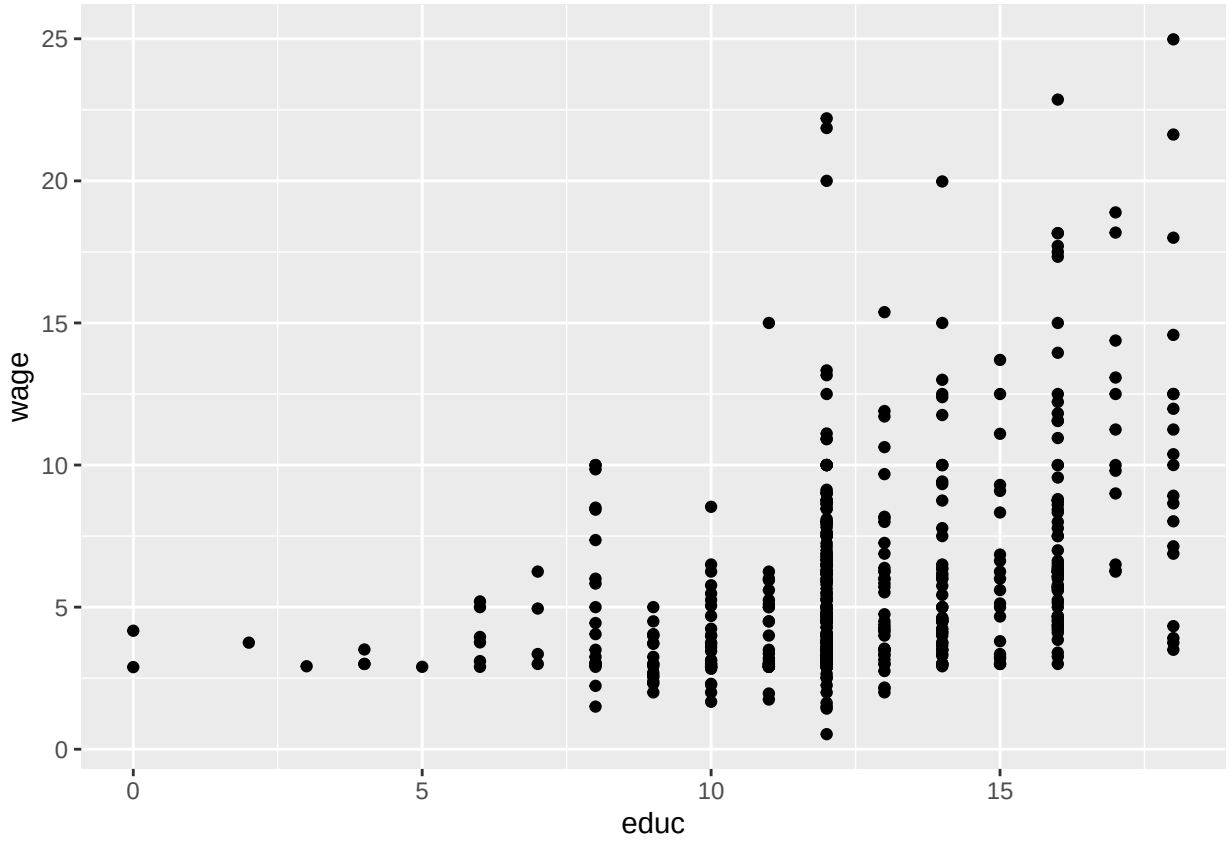

```
qplot(factor(nonwhite), wage, fill= factor(female), data=dt.wages, geom = "boxplot")
```



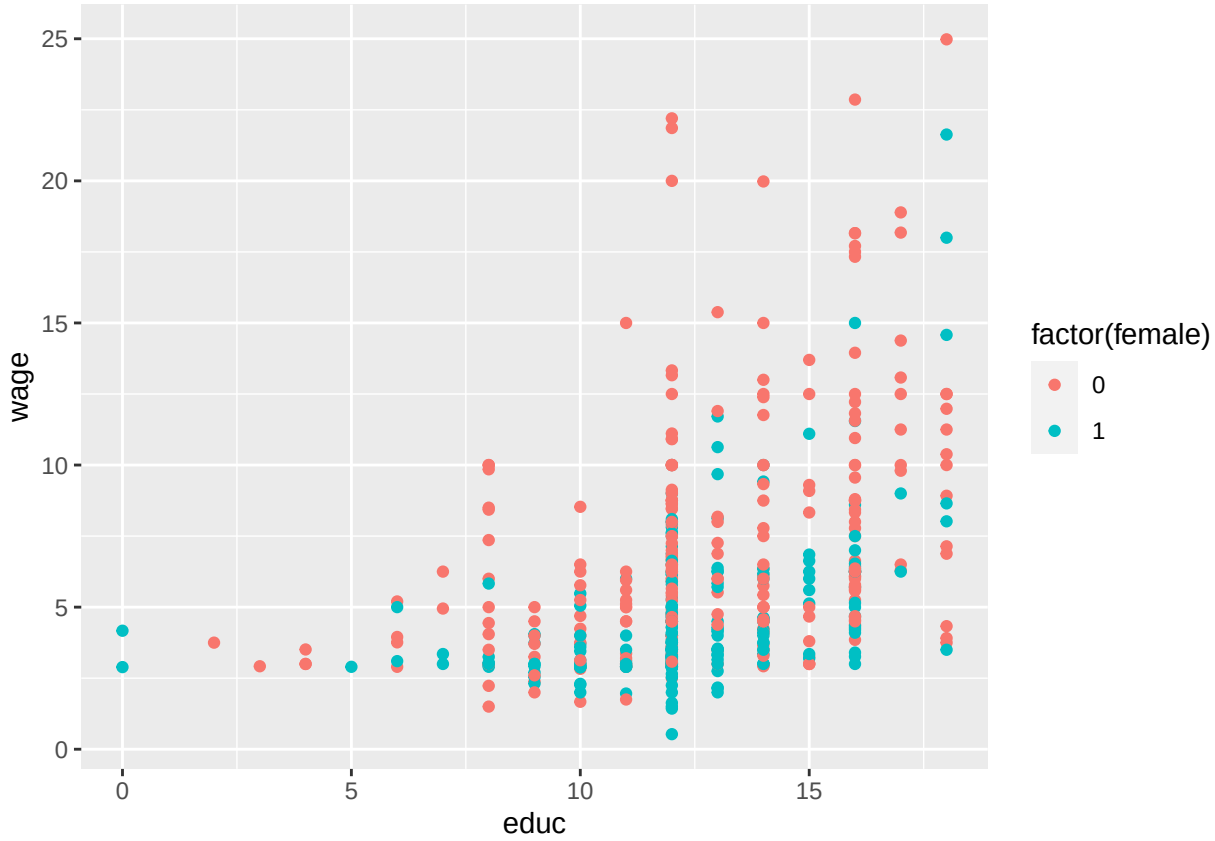

```
ggplot(dt.wages) + geom_boxplot(aes(factor(educ), exper))
```



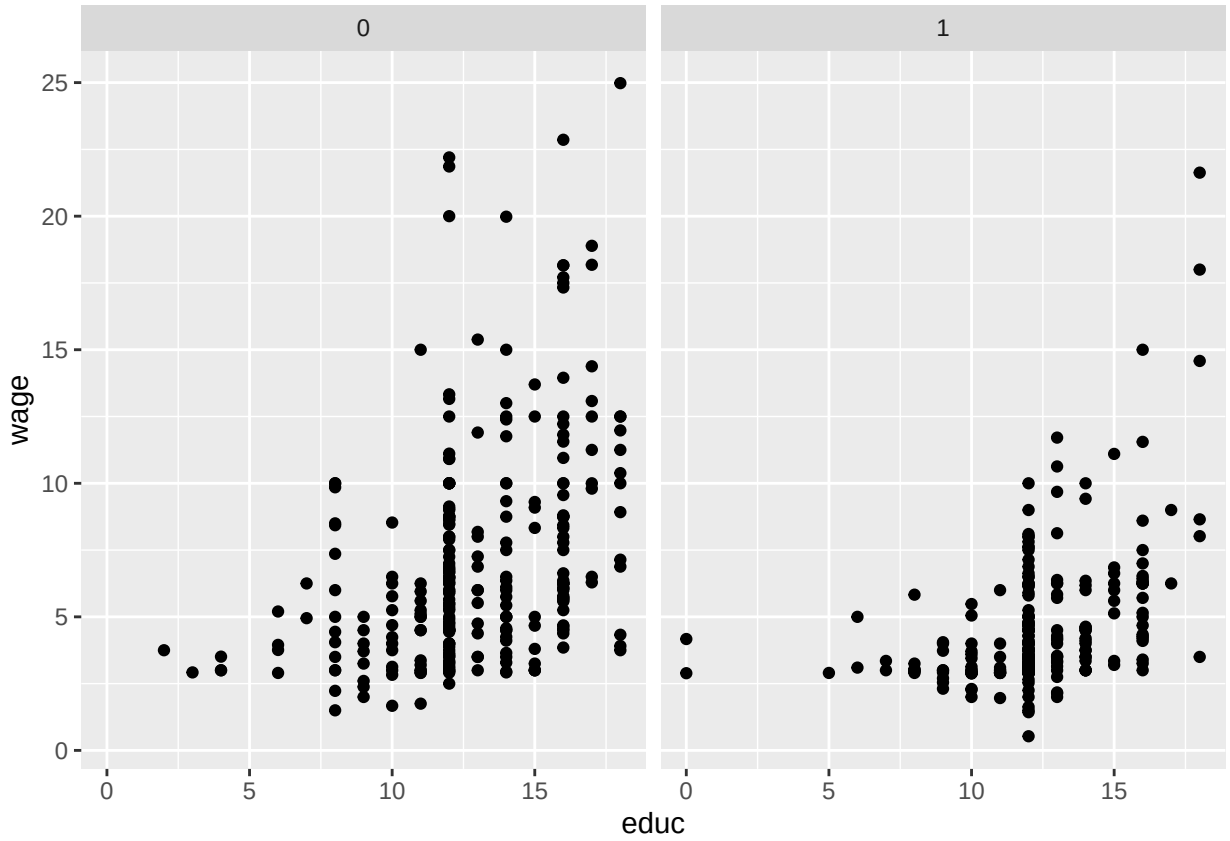

```
#Descriptive Statistics - Scatter


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



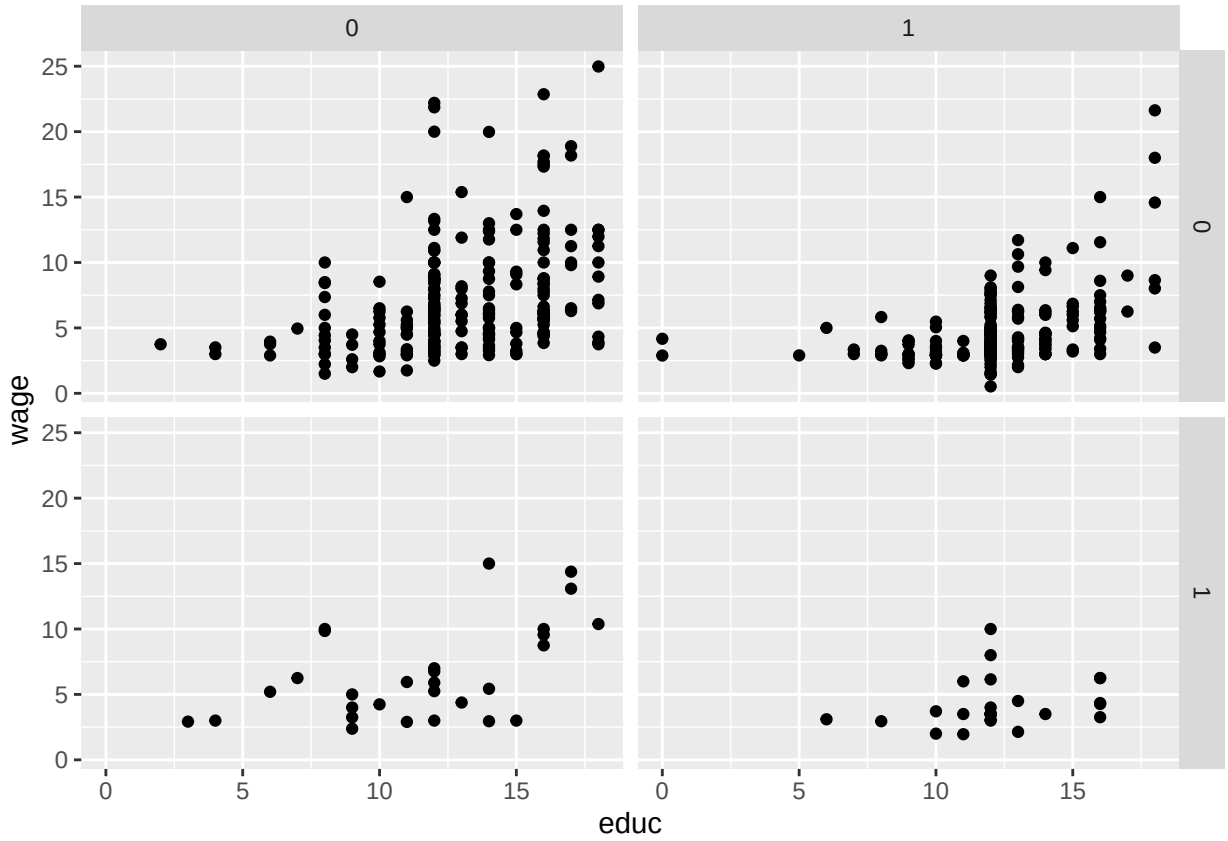

```
qplot(educ,wage, color=factor(female), data = dt.wages, geom = "point")
```



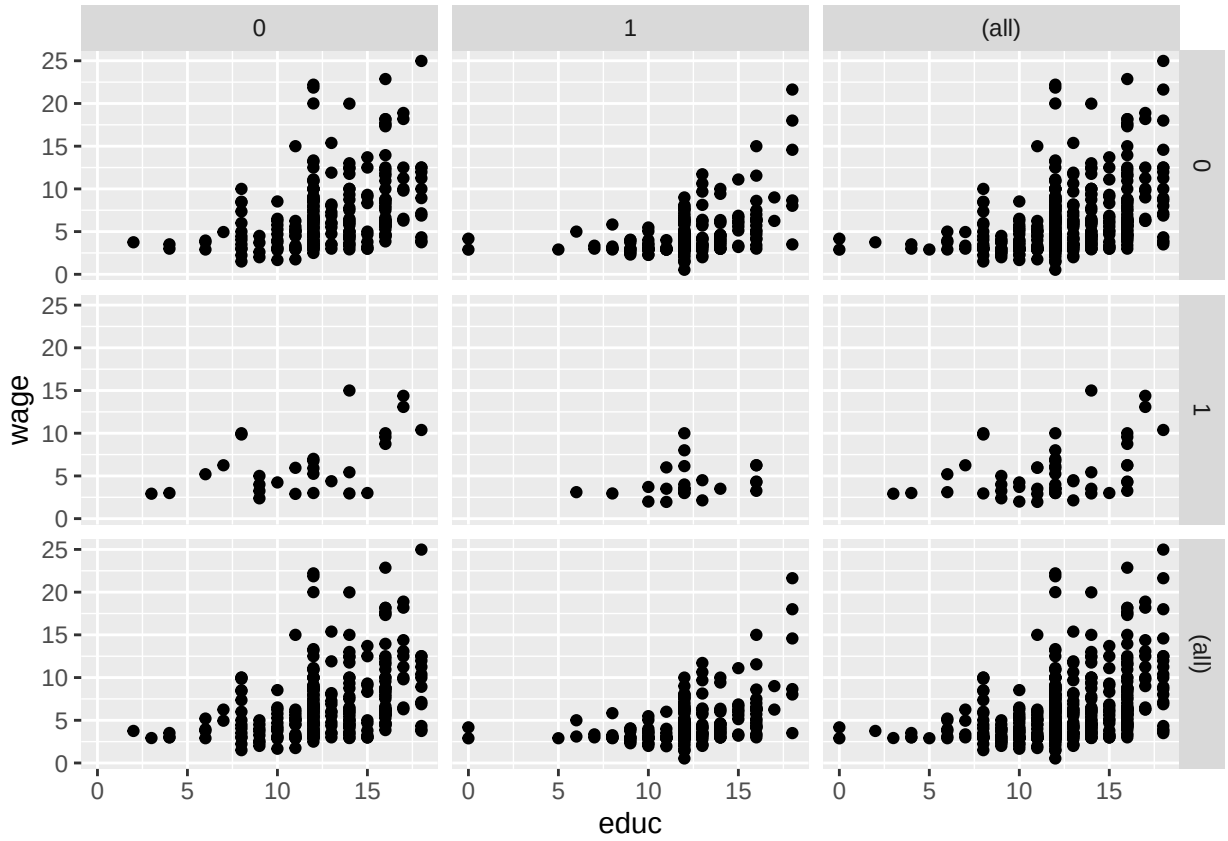

```
#Scatter plot


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



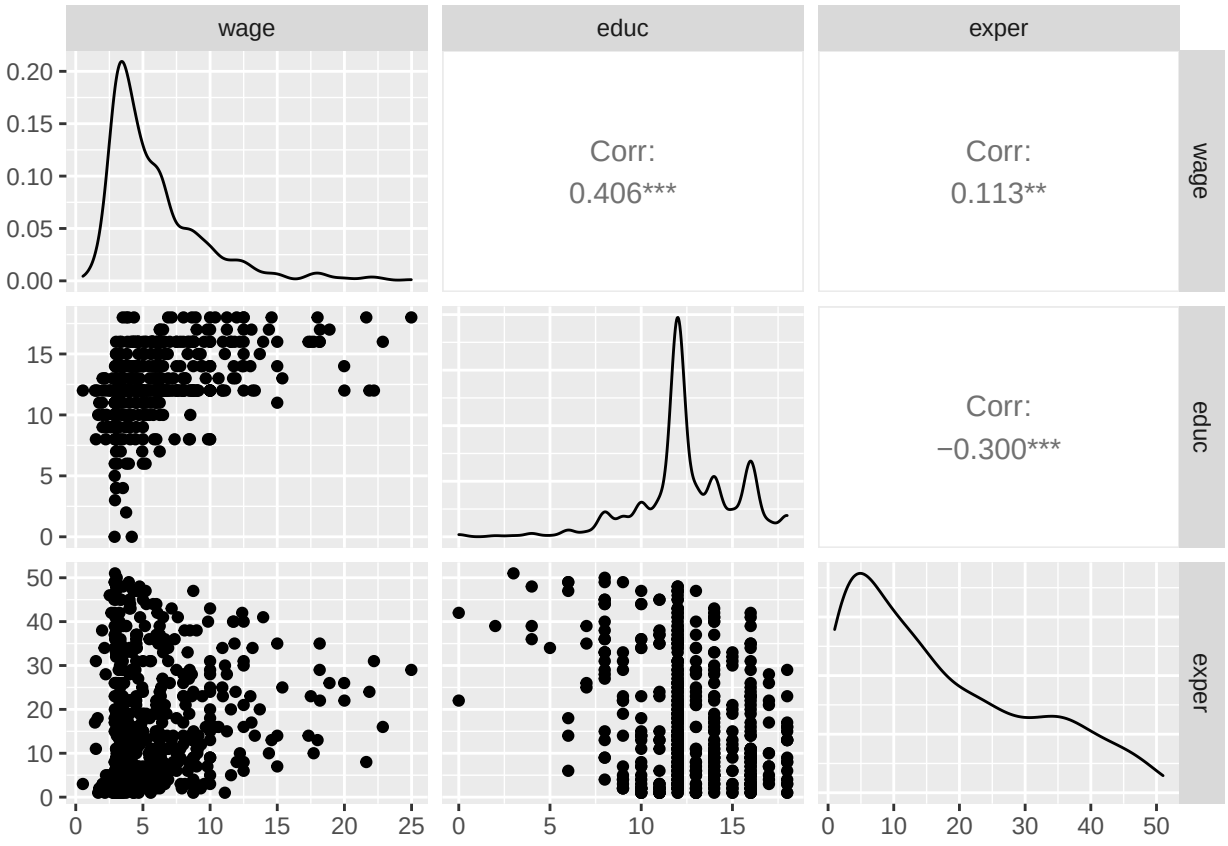

```
ggplot(dt.wages) + geom_point(aes(educ, wage)) + facet_grid ( nonwhite ~ female)
```



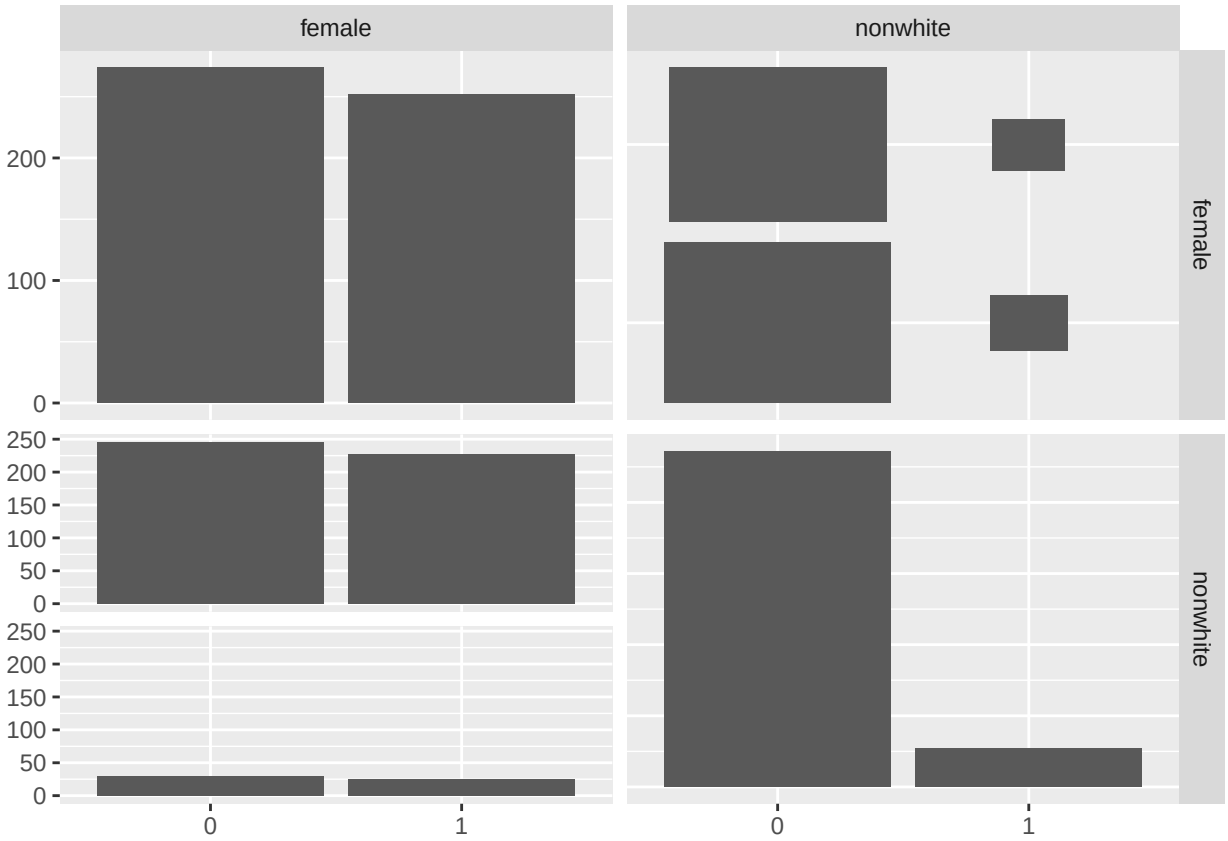

```
ggplot(dt.wages) + geom_point(aes(educ, wage)) + facet_grid ( nonwhite ~ female, margins = TRUE)
```




```
#GGpairs

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



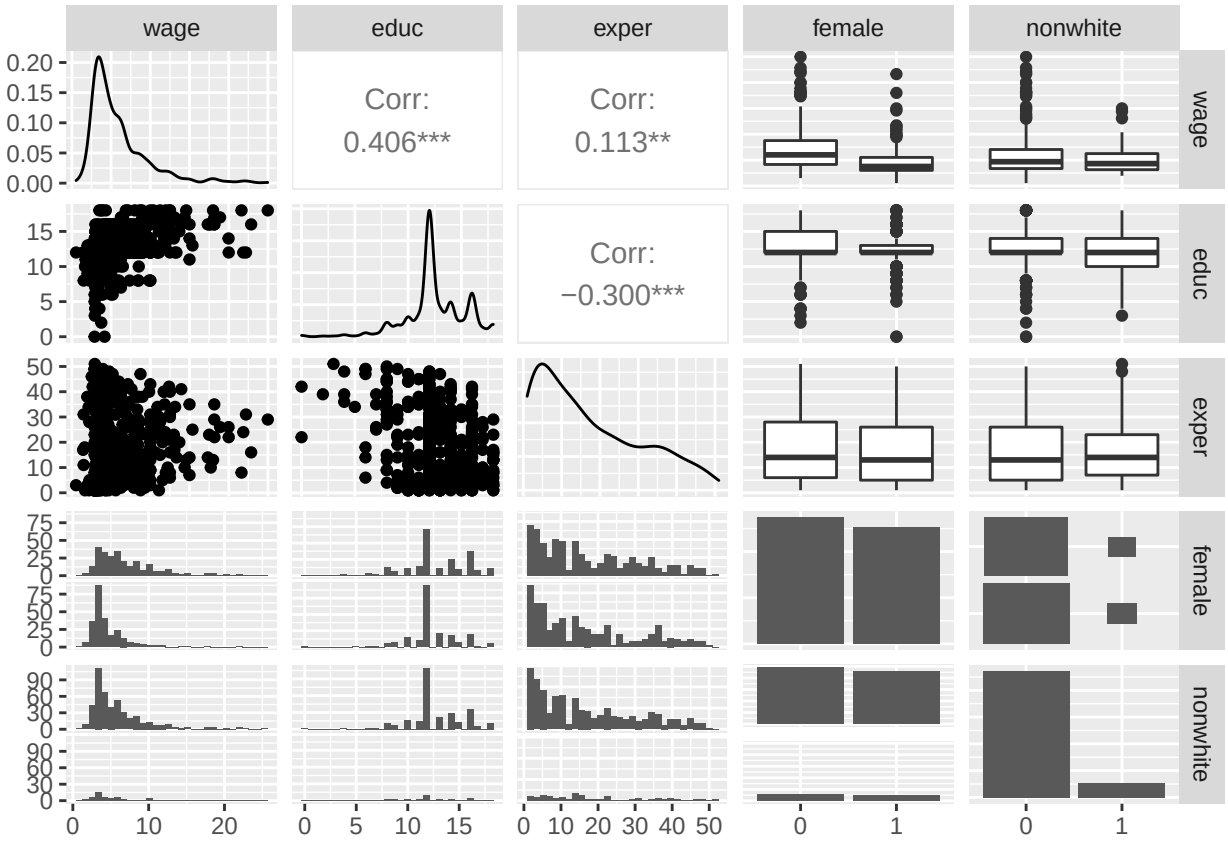

```
ggpairs(dt.wages[, list( female=factor(female), nonwhite=factor(nonwhite))])
```




```
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`.
```

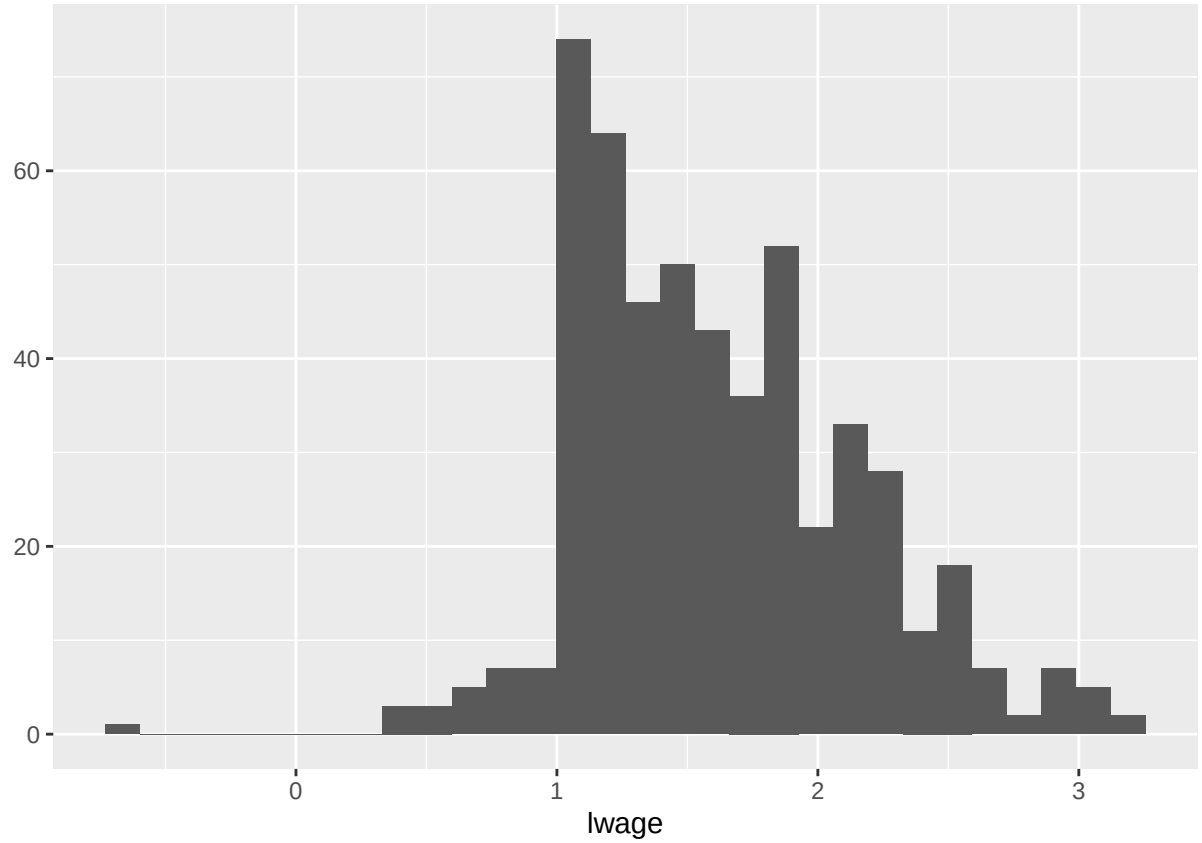



```
#Outliers

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
```