Chapter 1

```r
knitr::opts_knit$set(root.dir = '~/Documents/Linear-Model-using-R')
```

Chapter 1 Introduction

1.1 Before you start

1.2 Initial Data Analysis

```r
# install.packages("faraway")
library(faraway)
```

import the dataset and have a first look at it

```r
# import the dataset called `pima` from the package
data(pima, package="faraway")
head(pima)
```

```
##   pregnant glucose diastolic triceps insulin  bmi diabetes age test
## 1        6     148        72      35       0 33.6    0.627  50    1
## 2        1      85        66      29       0 26.6    0.351  31    0
## 3        8     183        64       0       0 23.3    0.672  32    1
## 4        1      89        66      23      94 28.1    0.167  21    0
## 5        0     137        40      35     168 43.1    2.288  33    1
## 6        5     116        74       0       0 25.6    0.201  30    0
```

```r
# alternative way to import data
# makes all data and functions that are included in the package available
require(faraway)
head(pima)
```

```
##   pregnant glucose diastolic triceps insulin  bmi diabetes age test
## 1        6     148        72      35       0 33.6    0.627  50    1
## 2        1      85        66      29       0 26.6    0.351  31    0
## 3        8     183        64       0       0 23.3    0.672  32    1
## 4        1      89        66      23      94 28.1    0.167  21    0
## 5        0     137        40      35     168 43.1    2.288  33    1
## 6        5     116        74       0       0 25.6    0.201  30    0
```

make numerical summaries

At this stage, we are looking for anything unusual or unexpected e.g. data-entry error

```r
# obtain numerical summaries
# summary function gives the usual univariate summary information
summary(pima)
```

```
##     pregnant         glucose        diastolic        triceps
```

```
##  Min.   : 0.000   Min.   :  0.0   Min.   :  0.00   Min.   : 0.00  
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.00   1st Qu.: 0.00  
##  Median : 3.000   Median :117.0   Median : 72.00   Median :23.00  
##  Mean   : 3.845   Mean   :120.9   Mean   : 69.11   Mean   :20.54  
##  3rd Qu.: 6.000   3rd Qu.:140.2   3rd Qu.: 80.00   3rd Qu.:32.00  
##  Max.   :17.000   Max.   :199.0   Max.   :122.00   Max.   :99.00  
##     insulin           bmi           diabetes           age       
##  Min.   :  0.0   Min.   : 0.00   Min.   :0.0780   Min.   :21.00  
##  1st Qu.:  0.0   1st Qu.:27.30   1st Qu.:0.2437   1st Qu.:24.00  
##  Median : 30.5   Median :32.00   Median :0.3725   Median :29.00  
##  Mean   : 79.8   Mean   :31.99   Mean   :0.4719   Mean   :33.24  
##  3rd Qu.:127.2   3rd Qu.:36.60   3rd Qu.:0.6262   3rd Qu.:41.00  
##  Max.   :846.0   Max.   :67.10   Max.   :2.4200   Max.   :81.00  
##       test      
##  Min.   :0.000  
##  1st Qu.:0.000  
##  Median :0.000  
##  Mean   :0.349  
##  3rd Qu.:1.000  
##  Max.   :1.000
```

We first look at the maximum value in variable `pregnant`: 17 months is not impossible although highly unlikely.

The 0 values in `diastolic` should not be correct as it means no blood pressure. (During the data analysis part, we need to have basic domain knowledge on the area or some common sense) So, we need to further investigate the variable by sorting them:

```
sort(pima$diastolic)
```

```
##   [1]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##  [19]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  24
##  [37]  30  30  38  40  44  44  44  44  46  46  48  48  48  48  48  50  50  50
##  [55]  50  50  50  50  50  50  50  50  50  50  52  52  52  52  52  52  52  52
##  [73]  52  52  52  54  54  54  54  54  54  54  54  54  54  54  55  55  56  56
##  [91]  56  56  56  56  56  56  56  56  56  56  58  58  58  58  58  58  58  58
## [109]  58  58  58  58  58  58  58  58  58  58  58  58  58  60  60  60  60  60
## [127]  60  60  60  60  60  60  60  60  60  60  60  60  60  60  60  60  60  60
## [145]  60  60  60  60  60  60  60  60  60  60  60  60  60  60  61  62  62  62
## [163]  62  62  62  62  62  62  62  62  62  62  62  62  62  62  62  62  62  62
## [181]  62  62  62  62  62  62  62  62  62  62  62  62  62  64  64  64  64  64
## [199]  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64
## [217]  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64  64
## [235]  64  64  65  65  65  65  65  65  65  66  66  66  66  66  66  66  66  66
## [253]  66  66  66  66  66  66  66  66  66  66  66  66  66  66  66  66  66  66
## [271]  66  66  66  68  68  68  68  68  68  68  68  68  68  68  68  68  68  68
## [289]  68  68  68  68  68  68  68  68  68  68  68  68  68  68  68  68  68  68
## [307]  68  68  68  68  68  68  68  68  68  68  68  68  70  70  70  70  70  70
## [325]  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70
## [343]  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70
## [361]  70  70  70  70  70  70  70  70  70  70  70  70  70  70  70  72  72  72
## [379]  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72
## [397]  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72  72
## [415]  72  72  72  72  72  74  74  74  74  74  74  74  74  74  74  74  74  74
## [433]  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74
## [451]  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74  74
```

```
## [469]  74  74  74  75  75  75  75  75  75  75  75  76  76  76  76  76  76  76
## [487]  76  76  76  76  76  76  76  76  76  76  76  76  76  76  76  76  76  76
## [505]  76  76  76  76  76  76  76  76  76  76  76  76  76  76  78  78  78  78
## [523]  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78
## [541]  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78  78
## [559]  78  78  78  78  78  80  80  80  80  80  80  80  80  80  80  80  80  80
## [577]  80  80  80  80  80  80  80  80  80  80  80  80  80  80  80  80  80  80
## [595]  80  80  80  80  80  80  80  80  80  82  82  82  82  82  82  82  82  82
## [613]  82  82  82  82  82  82  82  82  82  82  82  82  82  82  82  82  82  82
## [631]  82  82  82  84  84  84  84  84  84  84  84  84  84  84  84  84  84  84
## [649]  84  84  84  84  84  84  84  84  85  85  85  85  85  85  86  86  86  86
## [667]  86  86  86  86  86  86  86  86  86  86  86  86  86  86  86  86  86  88
## [685]  88  88  88  88  88  88  88  88  88  88  88  88  88  88  88  88  88  88
## [703]  88  88  88  88  88  88  90  90  90  90  90  90  90  90  90  90  90  90
## [721]  90  90  90  90  90  90  90  90  90  90  92  92  92  92  92  92  92  92
## [739]  94  94  94  94  94  94  95  96  96  96  96  98  98  98 100 100 100 102
## [757] 104 104 106 106 106 108 108 110 110 110 114 122
```

From the data, it is obvious that 35 entries have value 0, which probably indicates that these 0s are used to indicate missing values. If we don't change the correct form for these missing values, these values may affect the analysis. So, we can convert these 0s to NA inferring that these values are missing.

Thus, we set all 0s in `diastolic`, `glucose`, `triceps`, `insulin` and `bmi` to NA.

```
# convert all the 0 values in 5 variables to NA
pima$diastolic[pima$diastolic==0] <- NA
pima$glucose[pima$glucose == 0] <- NA
pima$triceps[pima$triceps == 0] <- NA
pima$insulin[pima$insulin == 0] <- NA
pima$bmi[pima$bmi == 0] <- NA
```

Variable `test` is a categorical variable which means it is a `factor` variable.

```
# convert the `test` variable to a factor variable; otherwise, this variable will be treated as a
# quantitative variable
pima$test <- factor(pima$test)
# give a summary table for test variable
summary(pima$test)
```

```
##   0   1
## 500 268
```

We can convert the labels for factor variables to descriptive ones:

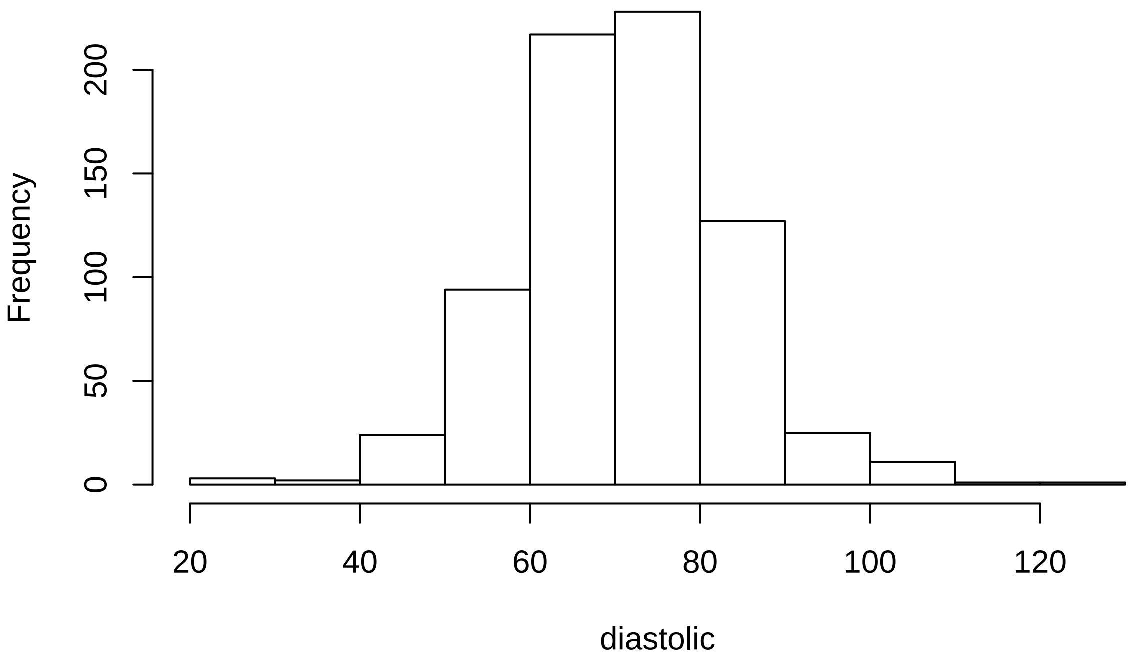
```
#convert the levels of `test` variable to descriptive labels
levels(pima$test) <- c("negative","positive")
summary(pima)
```

```
##     pregnant         glucose        diastolic         triceps
##  Min.   : 0.000   Min.   : 44.0   Min.   : 24.00   Min.   : 7.00
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 64.00   1st Qu.:22.00
##  Median : 3.000   Median :117.0   Median : 72.00   Median :29.00
##  Mean   : 3.845   Mean   :121.7   Mean   : 72.41   Mean   :29.15
##  3rd Qu.: 6.000   3rd Qu.:141.0   3rd Qu.: 80.00   3rd Qu.:36.00
##  Max.   :17.000   Max.   :199.0   Max.   :122.00   Max.   :99.00
##                   NA's   :5       NA's   :35       NA's   :227
##     insulin            bmi           diabetes           age
##  Min.   : 14.00   Min.   :18.20   Min.   :0.0780   Min.   :21.00
```

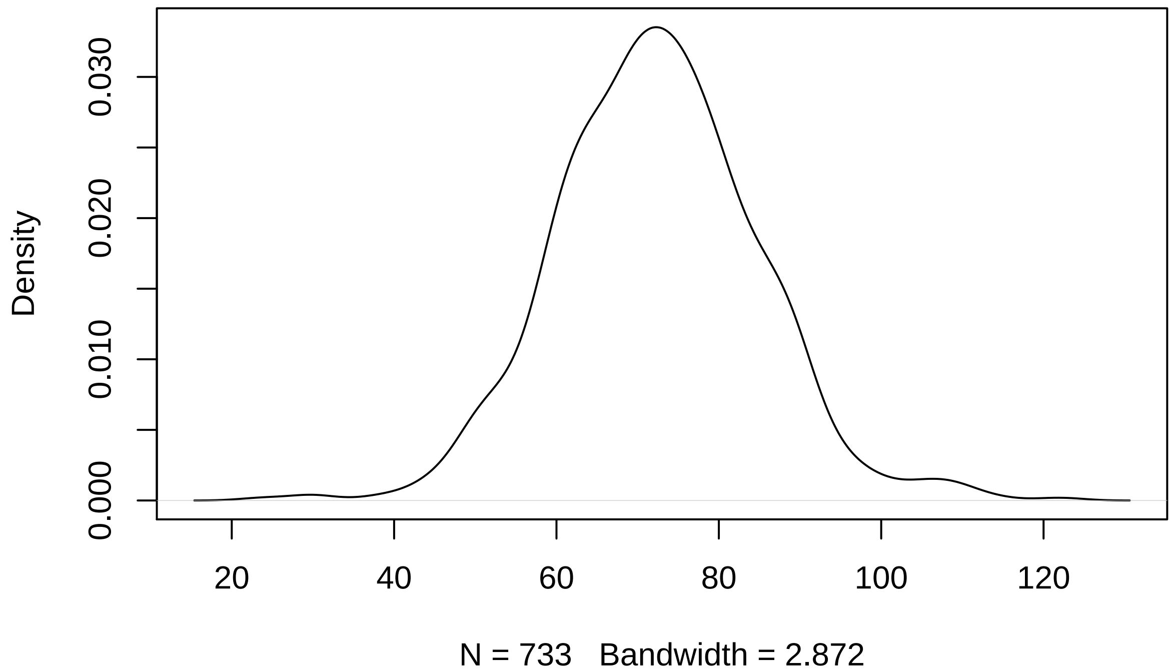
```
##  1st Qu.: 76.25   1st Qu.:27.50   1st Qu.:0.2437   1st Qu.:24.00
##  Median :125.00   Median :32.30   Median :0.3725   Median :29.00
##  Mean   :155.55   Mean   :32.46   Mean   :0.4719   Mean   :33.24
##  3rd Qu.:190.00   3rd Qu.:36.60   3rd Qu.:0.6262   3rd Qu.:41.00
##  Max.   :846.00   Max.   :67.10   Max.   :2.4200   Max.   :81.00
##  NA's   :374      NA's   :11
##        test
##  negative:500
##  positive:268
##
##
##
##
##
```

graphic descriptions

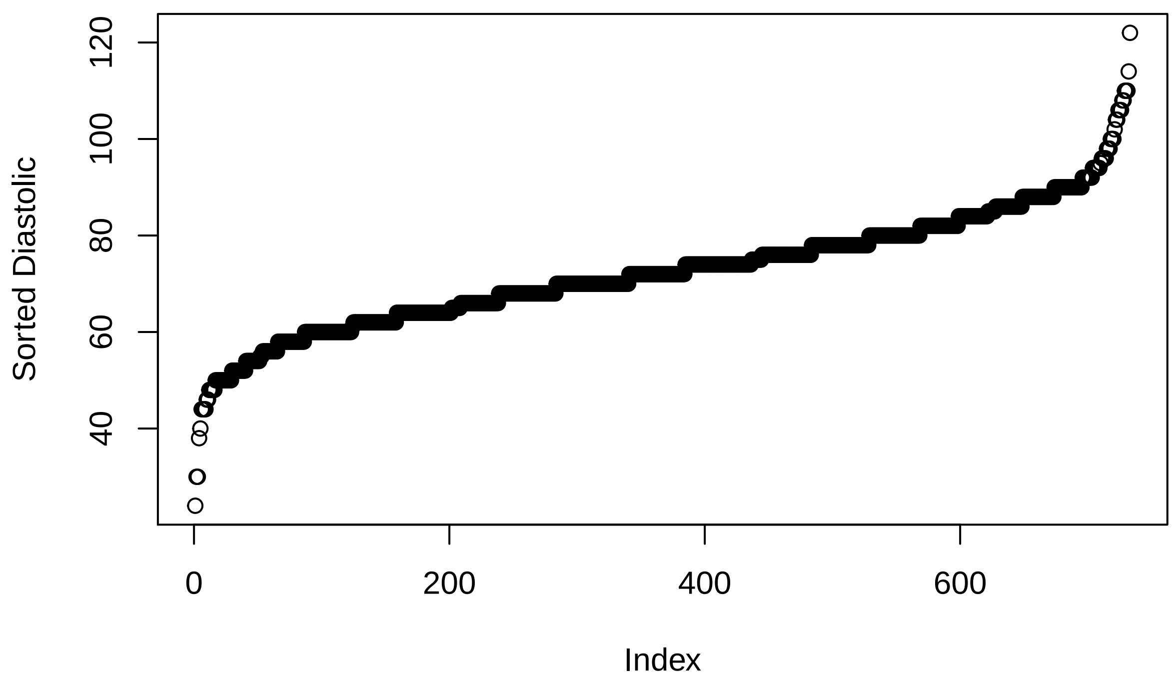
```r
#histogram with bins specified by r
hist(pima$diastolic, xlab="diastolic",main="")
```



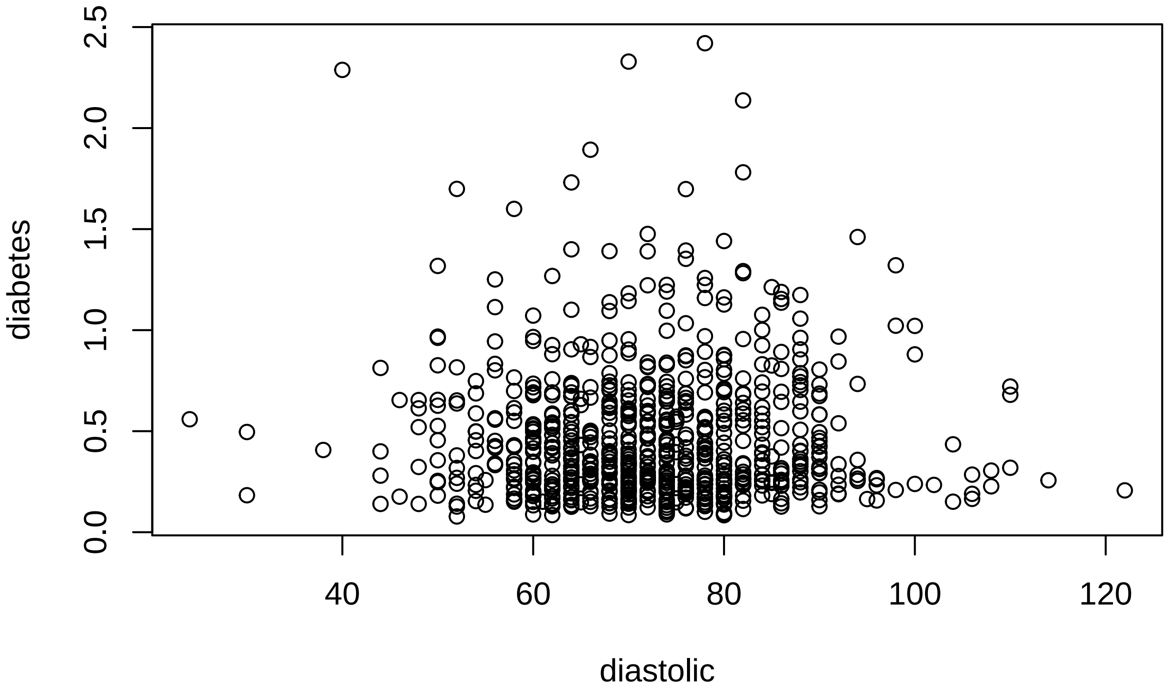

```r
# kernel estimate(density plot) of histogram
plot(density(pima$diastolic, na.rm = T),main="")
```



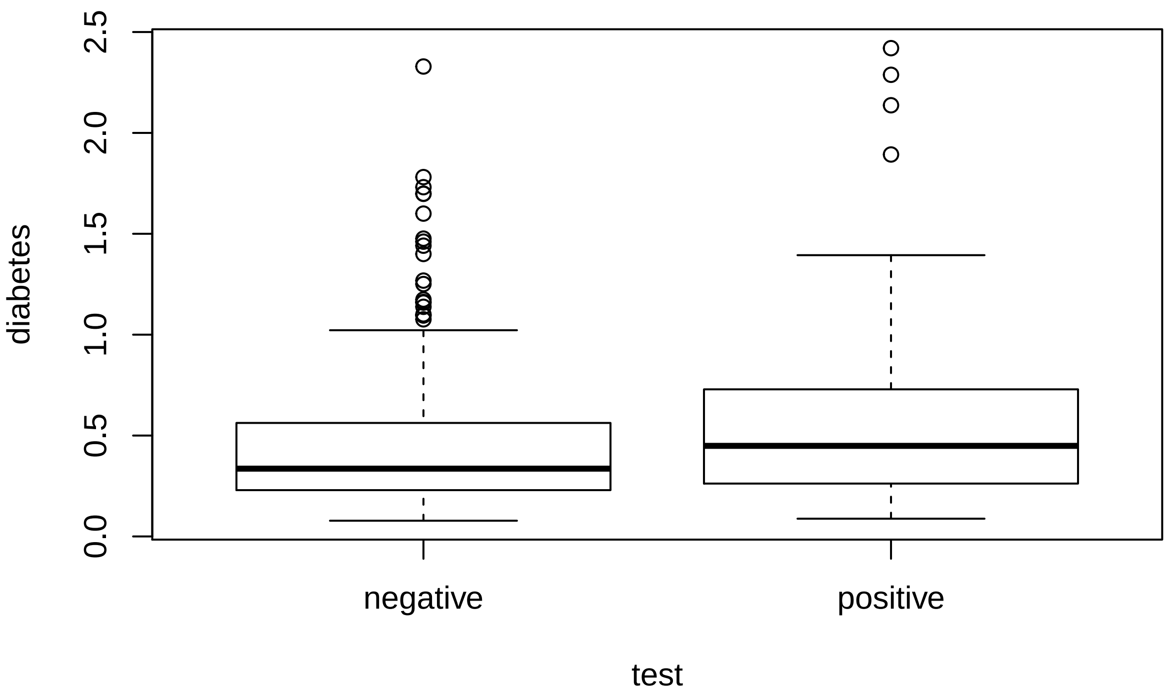

```
plot(sort(pima$diastolic),ylab="Sorted Diastolic")
```




```
# plot a scatterplot for diabetes against diastolic with the dataset pima
plot(diabetes~diastolic,pima)
```




```r
# plot two boxplots for diabetes against both positive and negative cases
plot(diabetes~test,pima)
```

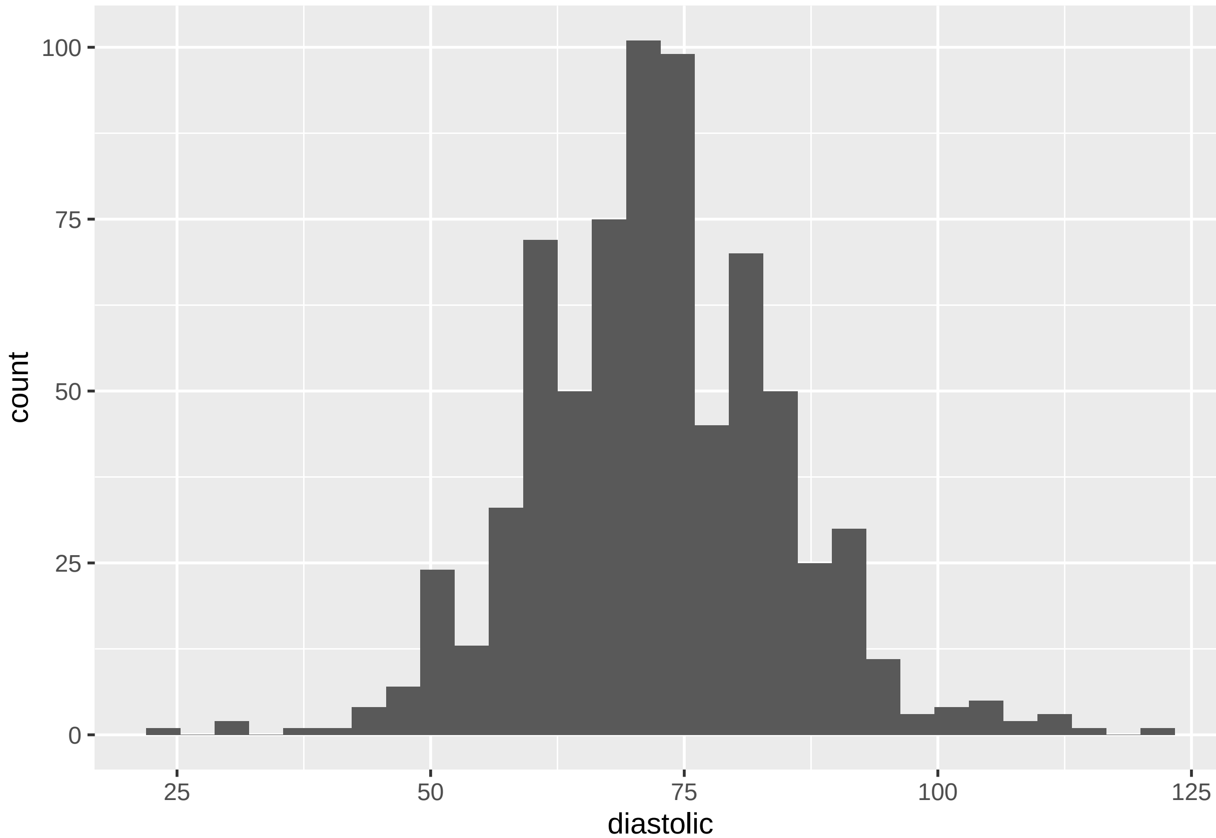



```r
# install.packages("ggplot2")
require(ggplot2)
```

```
## Loading required package: ggplot2
```

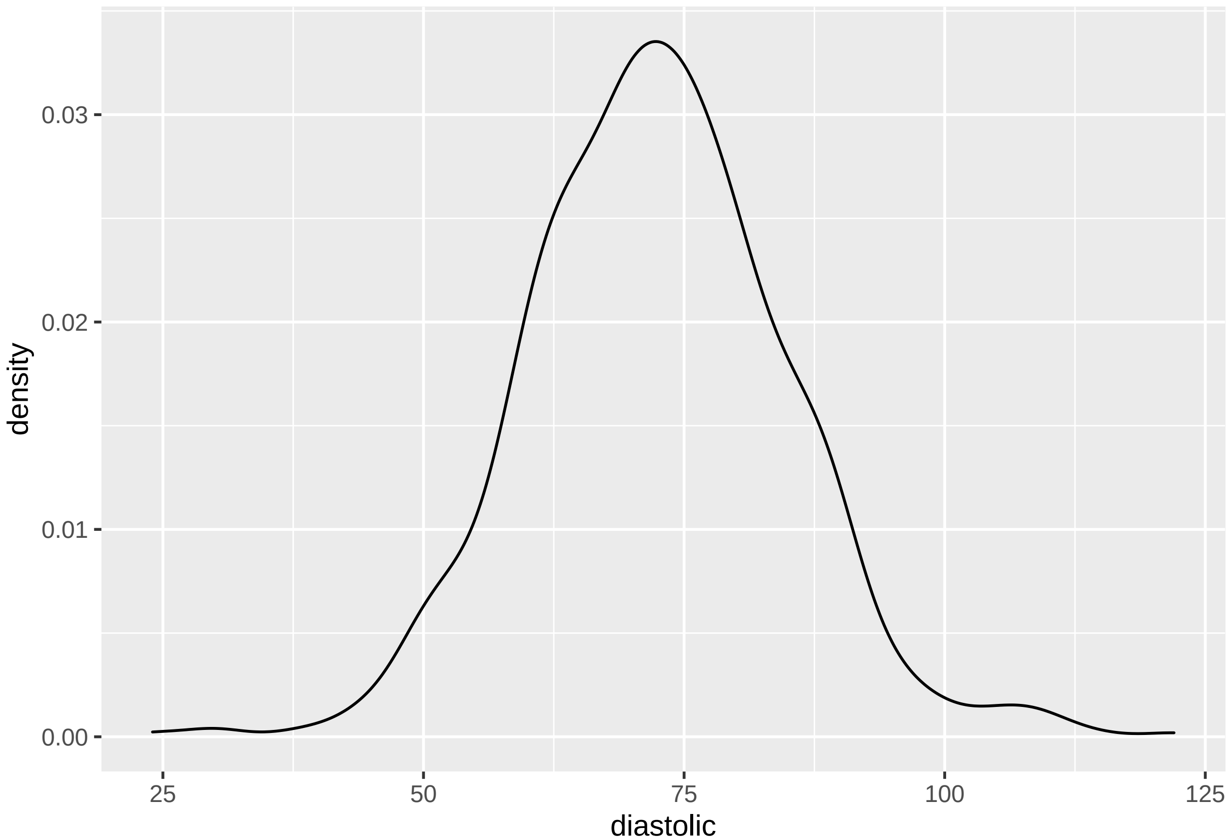
```r
# histogram
ggplot(pima, aes(x=diastolic))+geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 35 rows containing non-finite values (stat_bin).
```



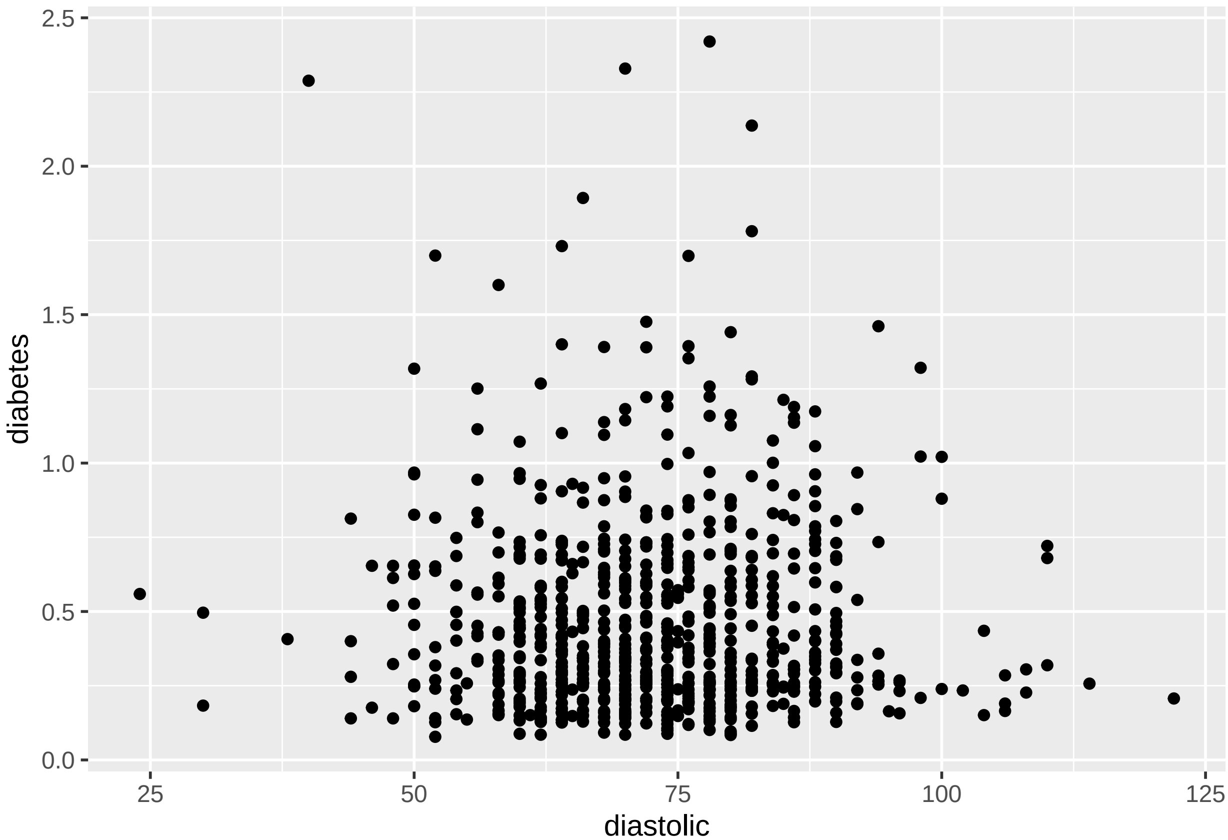

```r
# kernel estimate(density plot)
ggplot(pima, aes(x=diastolic))+geom_density()
```

```
## Warning: Removed 35 rows containing non-finite values (stat_density).
```




```
# scatterplot
ggplot(pima, aes(x=diastolic,y=diabetes))+geom_point()
```

```
## Warning: Removed 35 rows containing missing values (geom_point).
```

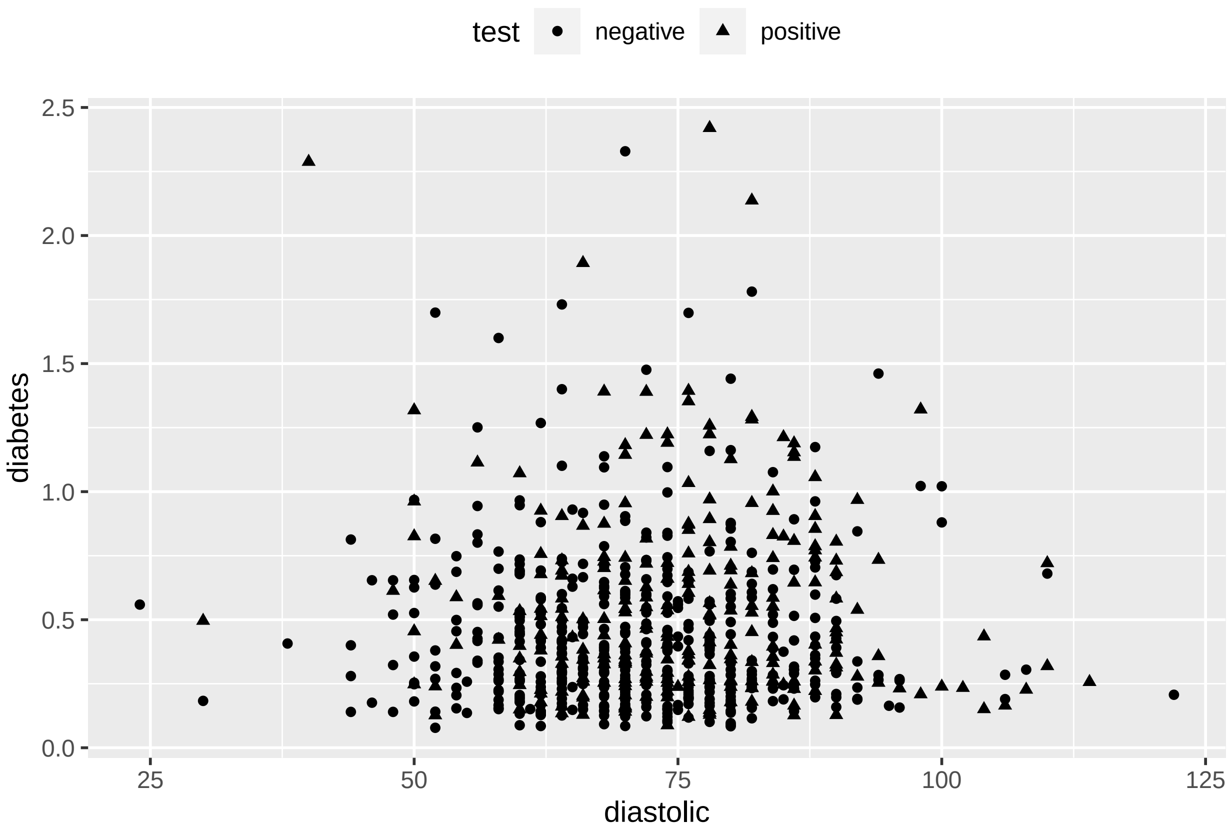


explanation of ggplot `ggplot(<dataset_name>, aes(x = <x_axis>, y = <y_axis>, colour=<colour_name>, ...))+ <geo_plots()>+theme()+facet_grid()`

ggplot arguments: - dataset - aesthetic: specify x-axis, y-axis, shape of points, colour of graph, etc - geom_plots: different sorts of plots - theme: specify options on apperance of the plot - facet_grid: form a matrix of panels defined by row and column faceting variables

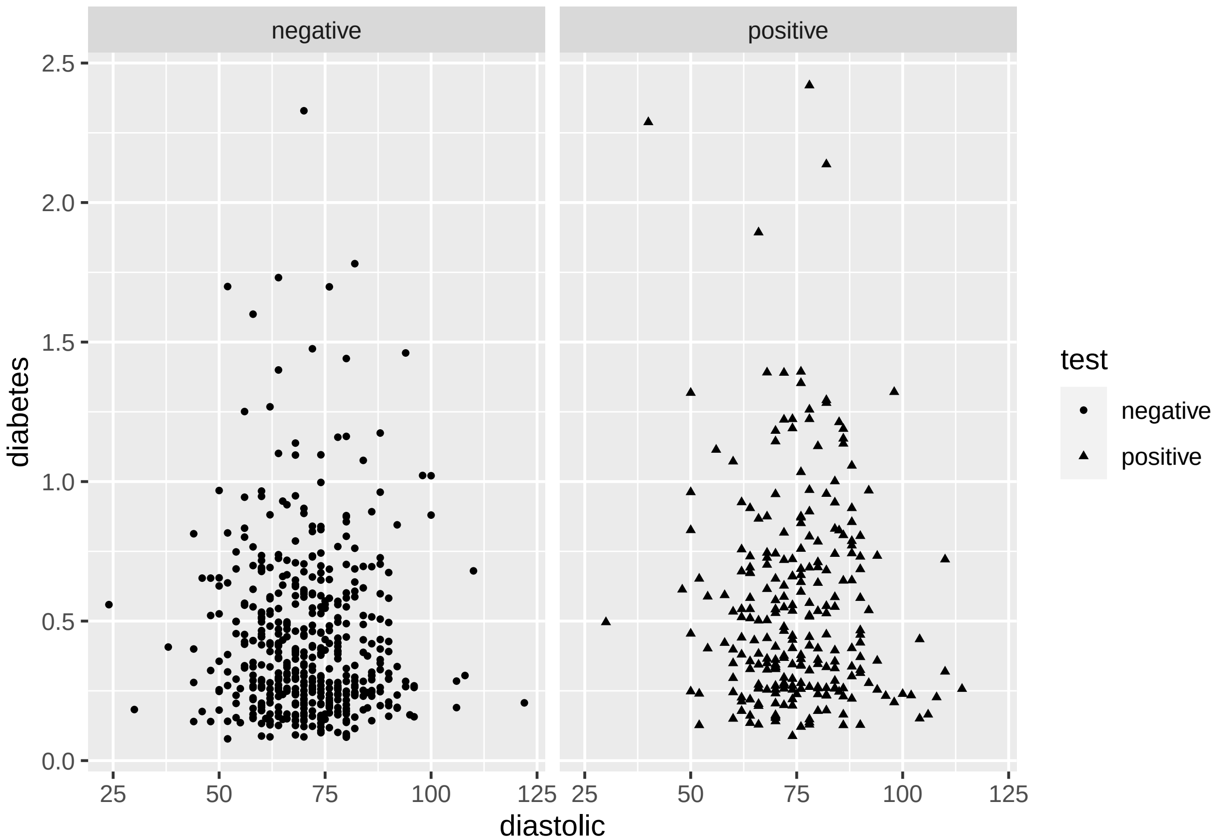
```
# scatterplot with a legend on the top
ggplot(pima, aes(x=diastolic, y=diabetes, shape=test))+geom_point()+
    theme(legend.position="top", legend.direction="horizontal")
```

```
## Warning: Removed 35 rows containing missing values (geom_point).
```




```
# shape in aes function change the shape of points based on the values of test variable
ggplot(pima, aes(x=diastolic, y=diabetes, shape=test))+geom_point(size=1)+facet_grid(~test)
```

```
## Warning: Removed 35 rows containing missing values (geom_point).
```



1.3 When to use linear models

Regression analysis have two main objectives:

1. prediction of future or unseen response given specified values of predictors.
2. Assessment of the effect of, or relations between, predictors and responses. If possible, we want to infer casual relationships.

1.4 History

Use linear regression to find coefficients in physical science

```
data(manilius, package="faraway")
head(manilius)
```

```
##        arc sinang  cosang group
## 1 13.16667 0.8836 -0.4682     1
## 2 13.13333 0.9996 -0.0282     1
## 3 13.20000 0.9899  0.1421     1
## 4 14.25000 0.2221  0.9750     3
## 5 14.70000 0.0006  1.0000     3
## 6 13.01667 0.9308 -0.3654     1
```

```
lm1 <- lm(arc~sinang+cosang,data=manilius)
lm1$coef
```

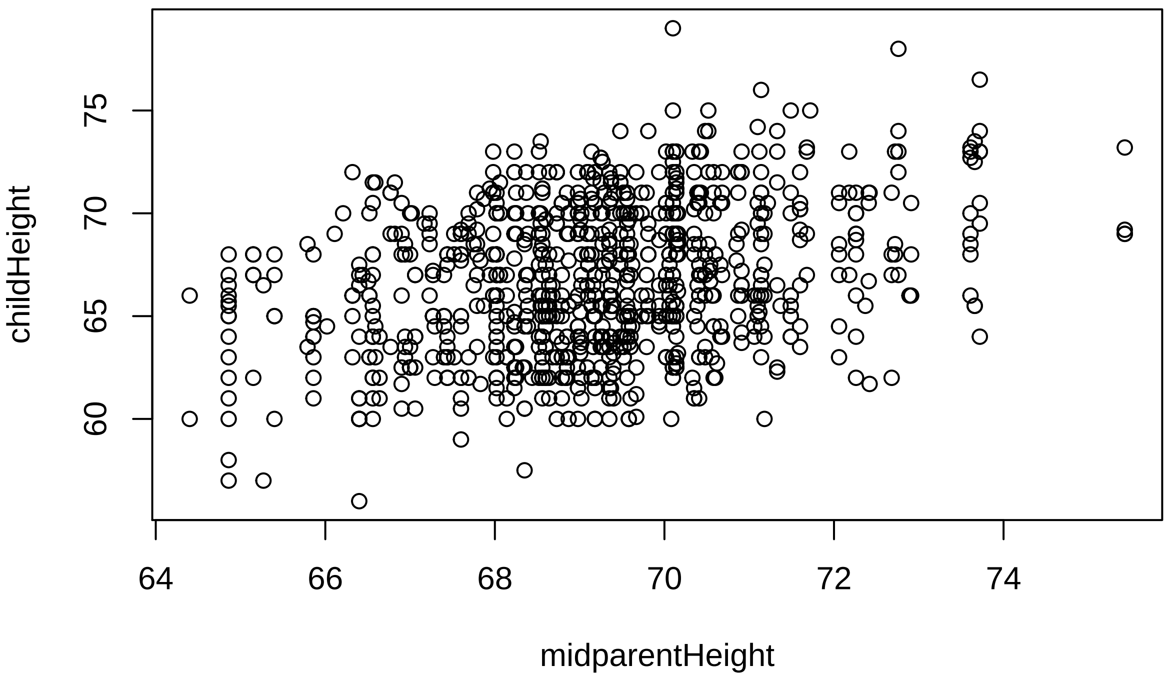
```
## (Intercept)      sinang      cosang
## 14.56162351 -1.50458123  0.09136504
```

Use of linear regression to find coefficients in social science

```
data(GaltonFamilies, package = "HistData")
head(GaltonFamilies)
```

```
##   family father mother midparentHeight children childNum gender childHeight
## 1    001   78.5   67.0           75.43        4        1   male        73.2
## 2    001   78.5   67.0           75.43        4        2 female        69.2
## 3    001   78.5   67.0           75.43        4        3 female        69.0
## 4    001   78.5   67.0           75.43        4        4 female        69.0
## 5    002   75.5   66.5           73.66        4        1   male        73.5
## 6    002   75.5   66.5           73.66        4        2   male        72.5
```

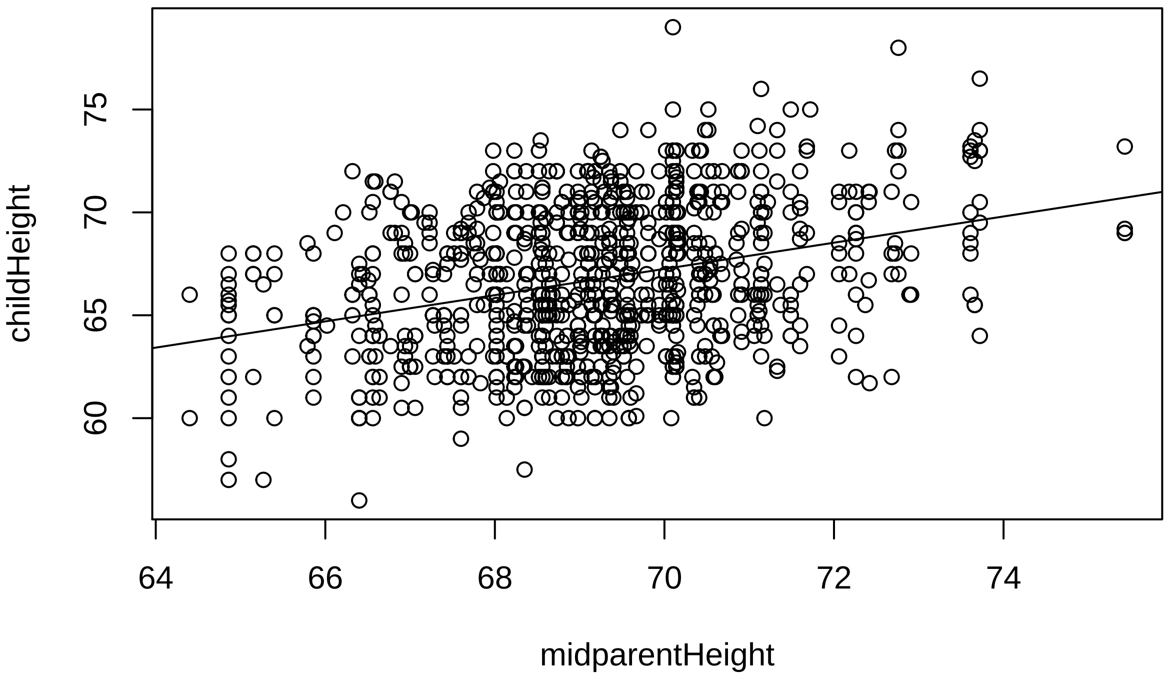
```
plot(childHeight~midparentHeight, GaltonFamilies)
```




```
lm2 <- lm(childHeight~GaltonFamilies$midparentHeight,data=GaltonFamilies)
coef(lm2)
```

```
##                    (Intercept) GaltonFamilies$midparentHeight
##                     22.6362405                      0.6373609
```

```
plot(childHeight~midparentHeight, GaltonFamilies)
abline(lm2)
```



Exercises

###1

```
data("teengamb", package="faraway")
head(teengamb)
```

```
##   sex status income verbal gamble
## 1   1     51   2.00      8    0.0
## 2   1     28   2.50      8    0.0
## 3   1     37   2.00      6    0.0
## 4   1     28   7.00      4    7.3
## 5   1     65   2.00      8   19.6
## 6   1     61   3.47      6    0.1
```

```
summary(teengamb)
```

numerical summary

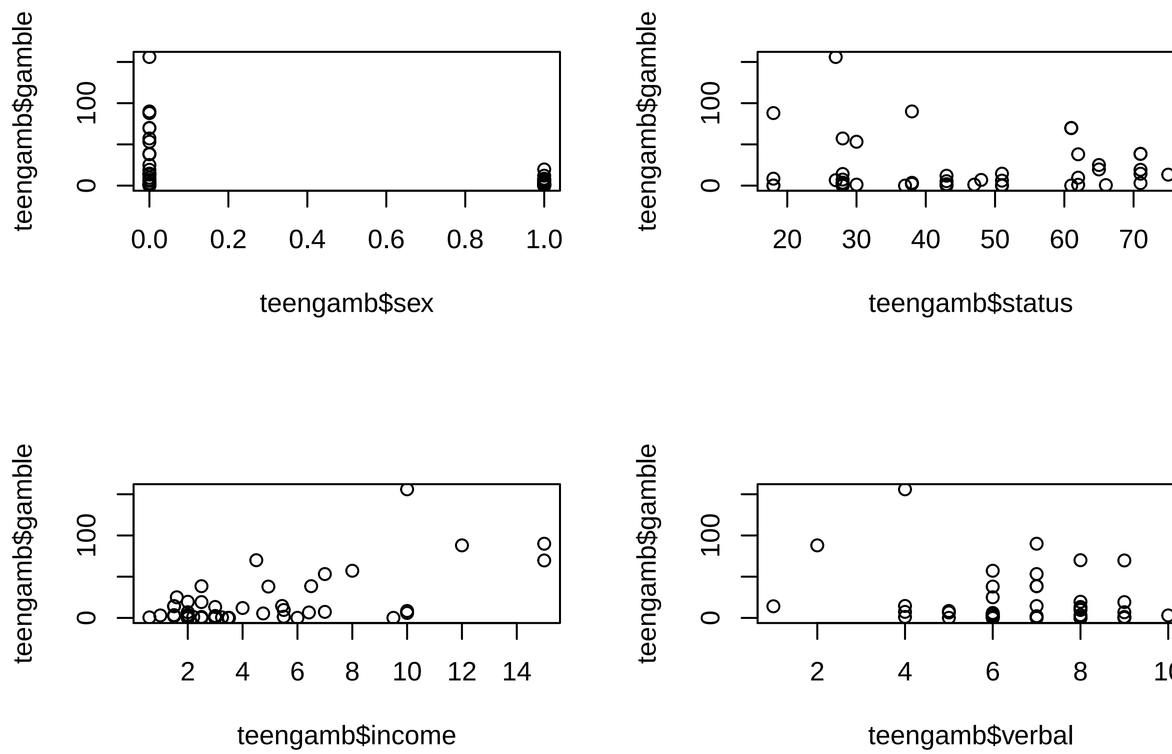
```
##       sex             status          income           verbal     
##  Min.   :0.0000   Min.   :18.00   Min.   : 0.600   Min.   : 1.00  
##  1st Qu.:0.0000   1st Qu.:28.00   1st Qu.: 2.000   1st Qu.: 6.00  
##  Median :0.0000   Median :43.00   Median : 3.250   Median : 7.00  
##  Mean   :0.4043   Mean   :45.23   Mean   : 4.642   Mean   : 6.66  
##  3rd Qu.:1.0000   3rd Qu.:61.50   3rd Qu.: 6.210   3rd Qu.: 8.00  
##  Max.   :1.0000   Max.   :75.00   Max.   :15.000   Max.   :10.00  
##      gamble    
##  Min.   :  0.0  
##  1st Qu.:  1.1  
##  Median :  6.0  
##  Mean   : 19.3  
##  3rd Qu.: 19.4  
```

```
##  Max.   :156.0
```

```r
table(teengamb$sex)
```

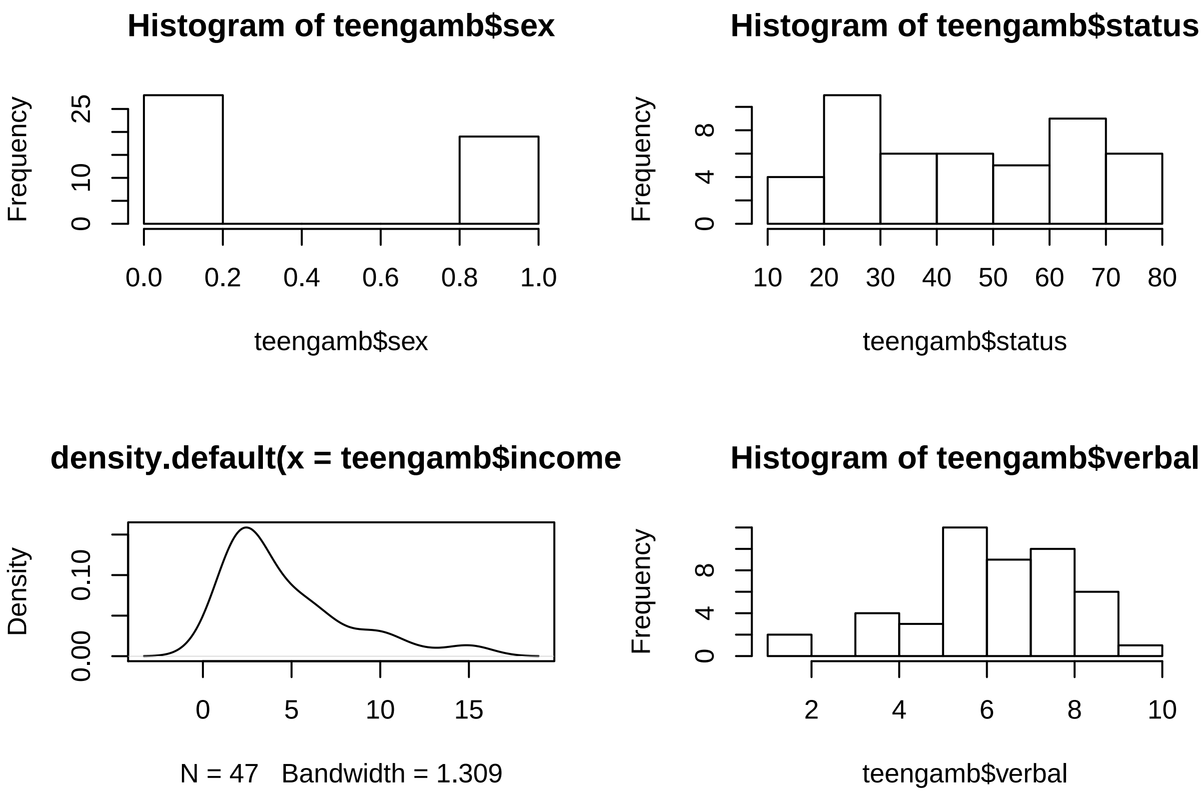
```
##
##  0  1
## 28 19
```

```r
par(mfrow=c(2,2))
plot(teengamb$sex,teengamb$gamble)
plot(teengamb$status,teengamb$gamble)
plot(teengamb$income,teengamb$gamble)
plot(teengamb$verbal,teengamb$gamble)
```



graphical summary

```r
par(mfrow=c(2,2))
hist(teengamb$sex)
hist(teengamb$status)
plot(density(teengamb$income))
hist(teengamb$verbal)
```



2

```r
data("uswages",package="faraway")
head(uswages)
```

```
##         wage educ exper race smsa ne mw so we pt
## 6085  771.60   18    18    0    1  1  0  0  0  0
## 23701 617.28   15    20    0    1  0  0  0  1  0
## 16208 957.83   16     9    0    1  0  0  1  0  0
## 2720  617.28   12    24    0    1  1  0  0  0  0
## 9723  902.18   14    12    0    1  0  1  0  0  0
## 22239 299.15   12    33    0    1  0  0  0  1  0
```

```r
summary(uswages)
```

numerical summary

```
##       wage              educ            exper            race      
##  Min.   :  50.39   Min.   : 0.00   Min.   :-2.00   Min.   :0.000  
##  1st Qu.: 308.64   1st Qu.:12.00   1st Qu.: 8.00   1st Qu.:0.000  
##  Median : 522.32   Median :12.00   Median :15.00   Median :0.000  
##  Mean   : 608.12   Mean   :13.11   Mean   :18.41   Mean   :0.078  
##  3rd Qu.: 783.48   3rd Qu.:16.00   3rd Qu.:27.00   3rd Qu.:0.000  
##  Max.   :7716.05   Max.   :18.00   Max.   :59.00   Max.   :1.000  
##       smsa             ne              mw               so        
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:1.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000  
##  Median :1.000   Median :0.000   Median :0.0000   Median :0.0000  
```

```
## Mean   :0.756   Mean   :0.229   Mean   :0.2485   Mean   :0.3125
## 3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:1.0000
## Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##       we              pt
## Min.   :0.00   Min.   :0.0000
## 1st Qu.:0.00   1st Qu.:0.0000
## Median :0.00   Median :0.0000
## Mean   :0.21   Mean   :0.0925
## 3rd Qu.:0.00   3rd Qu.:0.0000
## Max.   :1.00   Max.   :1.0000
```

```
# convert quatitative variables into qualitative
uswages$race <- factor(uswages$race)
uswages$smsa <- factor(uswages$smsa)
uswages$ne <- factor(uswages$ne)
uswages$mw <- factor(uswages$mw)
uswages$we <- factor(uswages$we)
uswages$so <- factor(uswages$so)
uswages$pt <- factor(uswages$pt)
uswages$educ <- factor(uswages$educ)
levels(uswages$race) <- c("White","Black")
levels(uswages$pt) <- c("full-time","part-time")
```

```
summary(uswages)
```

```
##      wage             educ         exper          race       smsa     ne
## Min.   :  50.39   12     :719   Min.   :-2.00   White:1844   0: 488   0:1542
## 1st Qu.: 308.64   16     :280   1st Qu.: 8.00   Black: 156   1:1512   1: 458
## Median : 522.32   14     :199   Median :15.00
## Mean   : 608.12   18     :197   Mean   :18.41
## 3rd Qu.: 783.48   13     :154   3rd Qu.:27.00
## Max.   :7716.05   15     : 86   Max.   :59.00
##                   (Other):365
## mw       so       we              pt
## 0:1503   0:1375   0:1580   full-time:1815
## 1: 497   1: 625   1: 420   part-time: 185
##
##
##
##
##
```

```
# summarise variable `ne`, `mw`, `we`, `so` into one variable `location`
uswages$location[uswages$ne==1] <- "North East"
uswages$location[uswages$mw==1] <- "Midwest"
uswages$location[uswages$we==1] <- "West"
uswages$location[uswages$so==1] <- "South"
uswages$location <- factor(uswages$location)
# levels(uswages$location) <- c("North East","Midwest","West","South")
# uswages$location
```

```
# sort(uswages$exper)
uswages$exper[uswages$exper<0] <- NA
sort(uswages$exper)
```

```
##    [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

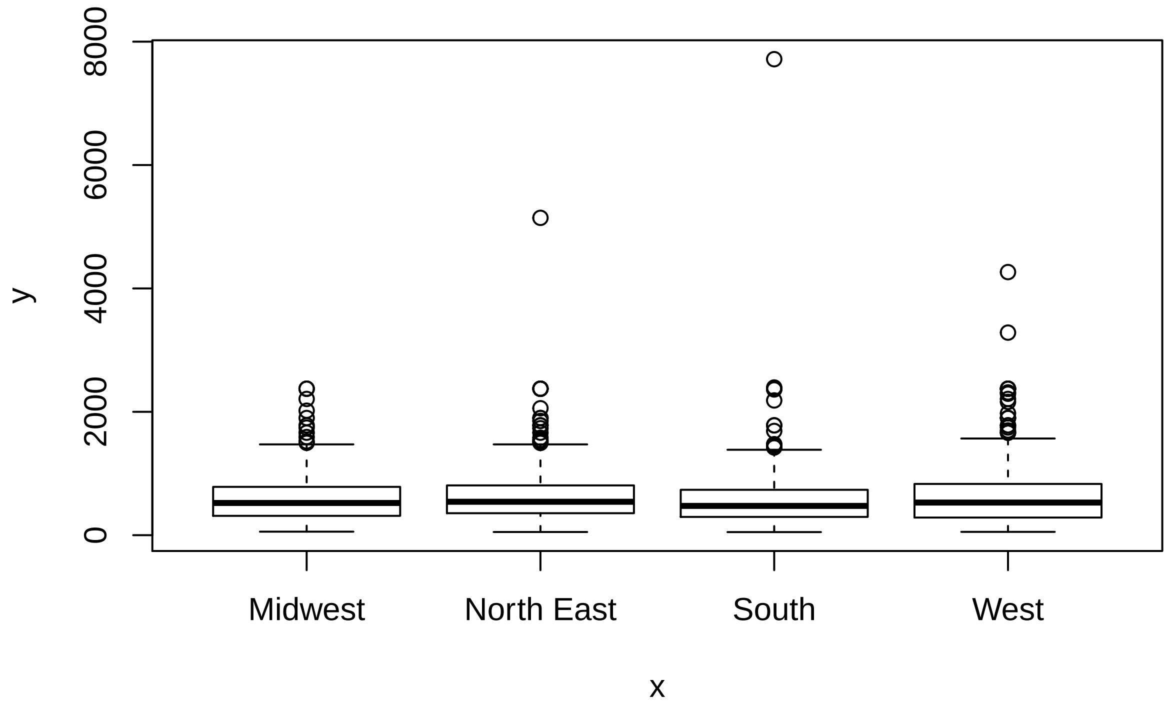
```
##    [25]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    [49]  0  0  0  0  0  0  0  0  0  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##    [73]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##    [97]  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
##   [121]  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3
##   [145]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
##   [169]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
##   [193]  3  3  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
##   [217]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
##   [241]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  5  5  5  5  5
##   [265]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
##   [289]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
##   [313]  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6
##   [337]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6
##   [361]  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7  7  7  7  7  7  7  7  7
##   [385]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
##   [409]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
##   [433]  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8
##   [457]  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
##   [481]  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
##   [505]  8  8  8  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9
##   [529]  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9
##   [553]  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9 10 10
##   [577] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
##   [601] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
##   [625] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 11
##   [649] 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
##   [673] 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
##   [697] 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
##   [721] 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
##   [745] 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
##   [769] 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
##   [793] 12 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13
##   [817] 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13
##   [841] 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13
##   [865] 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14
##   [889] 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14 14
##   [913] 14 14 14 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15
##   [937] 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15
##   [961] 15 15 15 15 15 15 15 15 15 15 16 16 16 16 16 16 16 16 16 16 16 16 16 16
##   [985] 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16
##  [1009] 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 17 17 17 17 17
##  [1033] 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17
##  [1057] 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 18 18 18 18 18 18
##  [1081] 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18
##  [1105] 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 19 19 19 19 19 19 19
##  [1129] 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19
##  [1153] 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 19 20 20
##  [1177] 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
##  [1201] 20 20 20 20 20 20 20 20 20 20 20 20 20 21 21 21 21 21 21 21 21 21 21 21
##  [1225] 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
##  [1249] 21 21 21 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22
##  [1273] 22 22 22 22 22 22 22 22 22 22 22 22 22 23 23 23 23 23 23 23 23 23 23 23
##  [1297] 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23
```

```
## [1321] 23 23 23 23 23 23 23 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24
## [1345] 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 25 25 25 25 25 25
## [1369] 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25
## [1393] 25 25 25 25 25 25 25 25 25 25 26 26 26 26 26 26 26 26 26 26 26 26 26 26
## [1417] 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 27
## [1441] 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27
## [1465] 27 27 27 27 27 27 27 27 27 27 27 27 28 28 28 28 28 28 28 28 28 28 28 28
## [1489] 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 29 29 29
## [1513] 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29
## [1537] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
## [1561] 30 30 30 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31 31
## [1585] 31 31 31 31 31 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32
## [1609] 32 32 32 32 32 32 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33
## [1633] 33 33 33 33 33 33 33 33 33 33 33 33 34 34 34 34 34 34 34 34 34 34 34 34
## [1657] 34 34 34 34 34 34 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35
## [1681] 35 35 35 35 35 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36
## [1705] 36 36 36 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
## [1729] 37 37 37 37 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38 38
## [1753] 38 38 38 38 38 38 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
## [1777] 39 39 39 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 41
## [1801] 41 41 41 41 41 41 41 41 41 42 42 42 42 42 42 42 42 42 42 42 42 42 42 42
## [1825] 42 42 42 42 42 42 42 42 42 42 42 43 43 43 43 43 43 43 43 43 43 43 43 43
## [1849] 43 43 43 43 43 43 43 43 43 43 43 43 44 44 44 44 44 44 44 44 44 44 44 44
## [1873] 44 44 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 45 46 46
## [1897] 46 46 46 46 46 46 46 46 46 46 47 47 47 47 47 47 47 47 47 47 47 47 47 47
## [1921] 47 47 47 48 48 48 48 48 48 48 48 48 48 48 48 49 49 49 49 50 50 50 51 51
## [1945] 51 52 52 52 52 52 52 52 53 53 54 54 54 54 55 55 55 55 56 56 57 58 59
```

```r
summary(uswages)
```

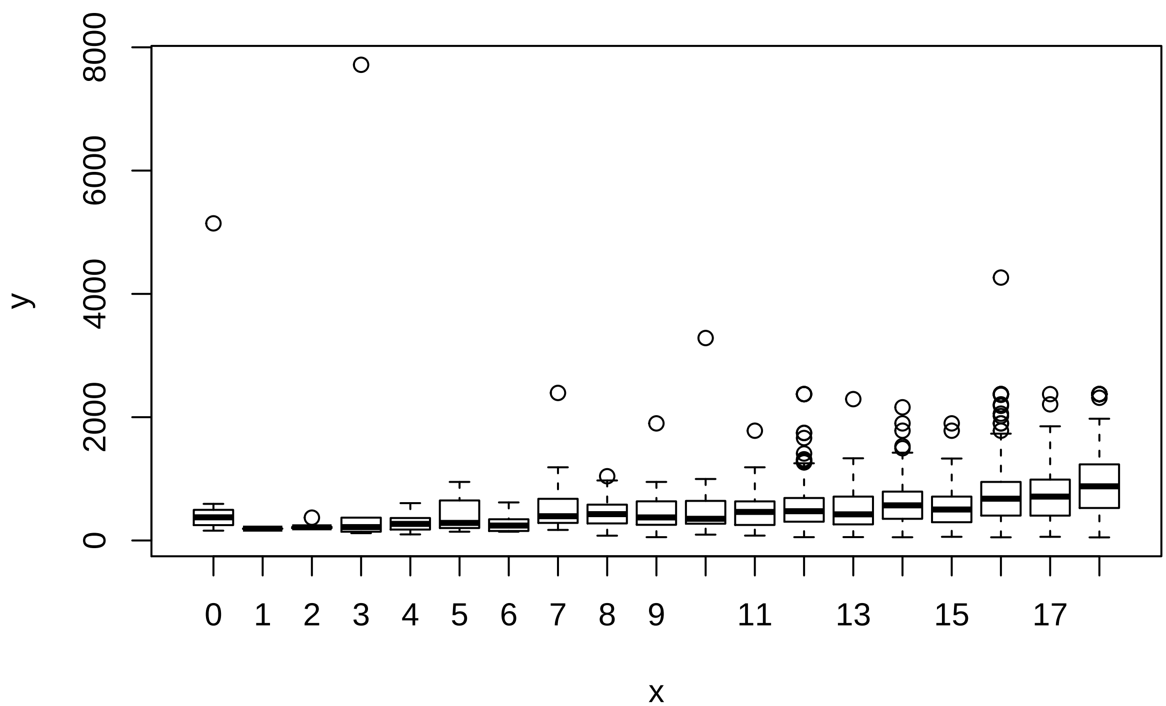
```
##       wage              educ          exper           race        smsa     ne
##  Min.   :  50.39   12     :719   Min.   : 0.00   White:1844   0: 488   0:1542
##  1st Qu.: 308.64   16     :280   1st Qu.: 8.00   Black: 156   1:1512   1: 458
##  Median : 522.32   14     :199   Median :16.00
##  Mean   : 608.12   18     :197   Mean   :18.74
##  3rd Qu.: 783.48   13     :154   3rd Qu.:27.00
##  Max.   :7716.05   15     : 86   Max.   :59.00
##                    (Other):365   NA's   :33
##  mw        so        we               pt              location
##  0:1503    0:1375    0:1580    full-time:1815   Midwest    :497
##  1: 497    1: 625    1: 420    part-time: 185   North East:458
##                                                 South      :625
##                                                 West       :420
##
##
##
```

```r
# par(mfrow=c(2,2))
plot(uswages$location, uswages$wage)
```

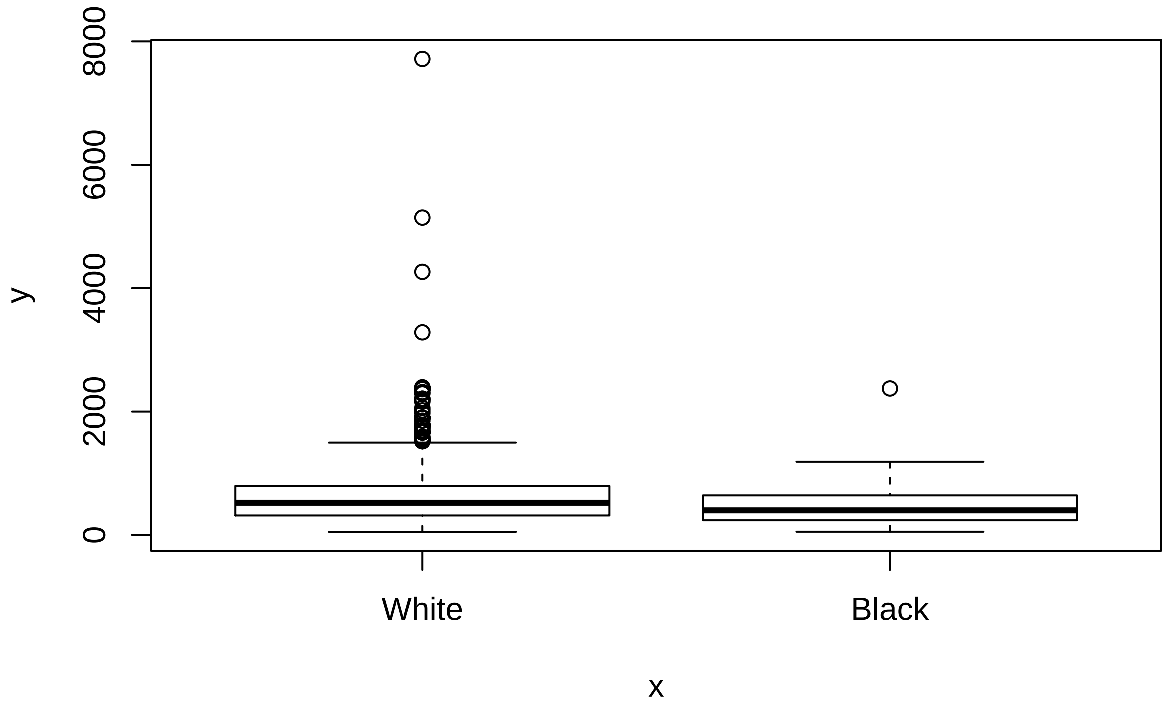


graphical summary

```r
plot(uswages$educ, uswages$wage)
```




```r
plot(uswages$race, uswages$wage)
```




```
plot(uswages$pt, uswages$wage)
```