exercise2

Lee Rui(BA19008042)

2020/3/24

clean the current workspace

```
rm(list = ls())
```

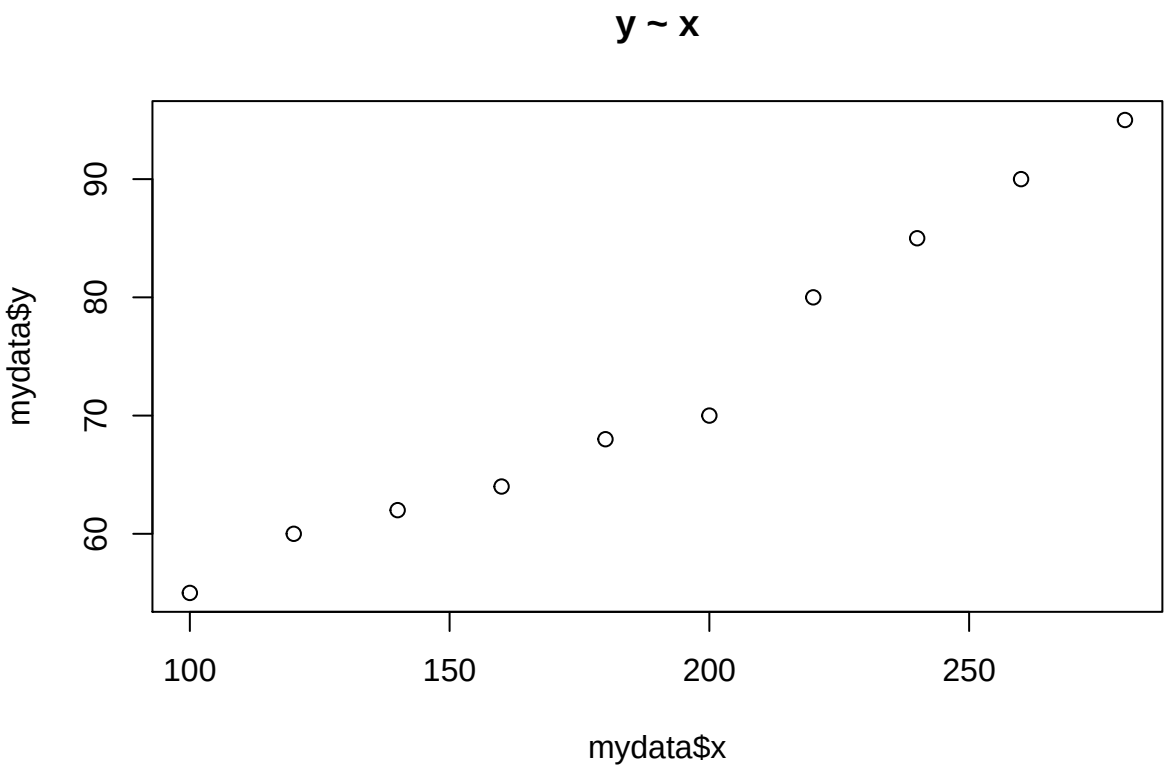
read data

```
mydata <- read.csv("/cloud/project/ecology/exercise2/xy.csv")
head(mydata)
```

```
##     x  y
## 1 100 55
## 2 120 60
## 3 140 62
## 4 160 64
## 5 180 68
## 6 200 70
```

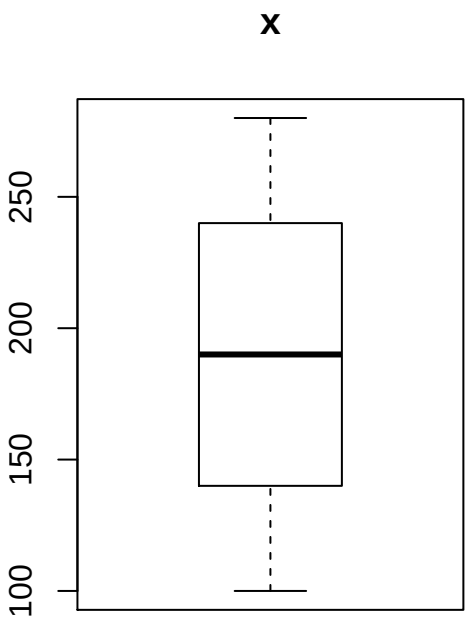
check what relationship between x and y by scatter plot

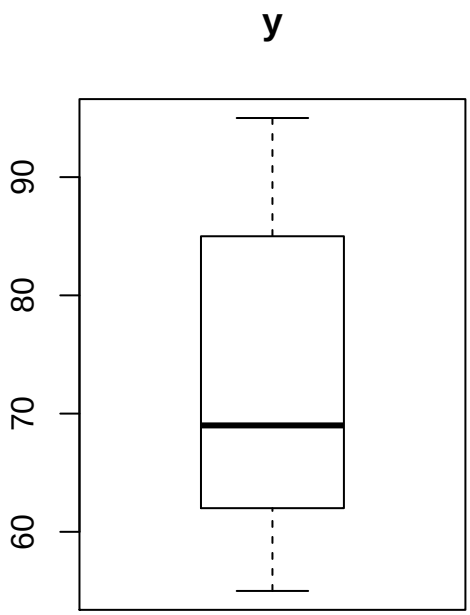
```
plot(mydata$y ~ mydata$x, data = mydata, main="y ~ x")
```



check whether there are outliers

```
par(mfrow=c(1, 2)) # set outplay of figure pannel
boxplot(mydata$x, main="x", sub=paste("Outlier rows: ", boxplot.stats(mydata$x)$out))
boxplot(mydata$y, main="y", sub=paste("Outlier rows: ", boxplot.stats(mydata$y)$out))
```



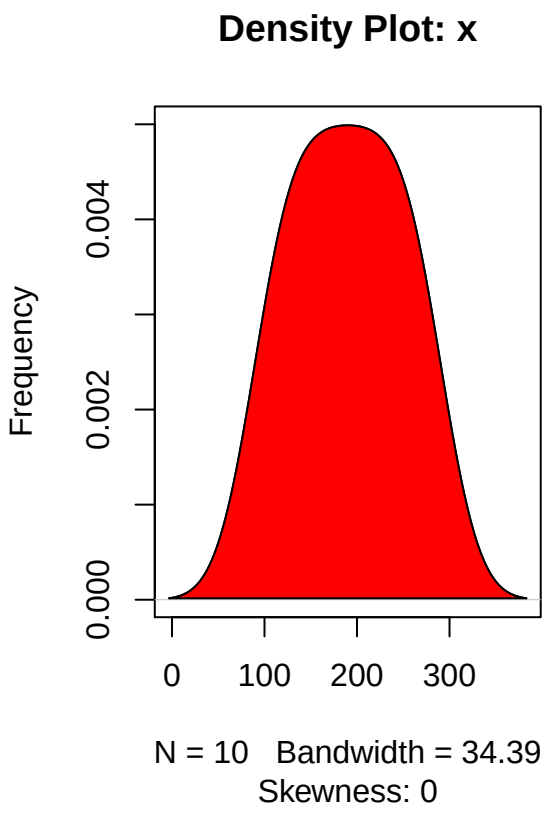


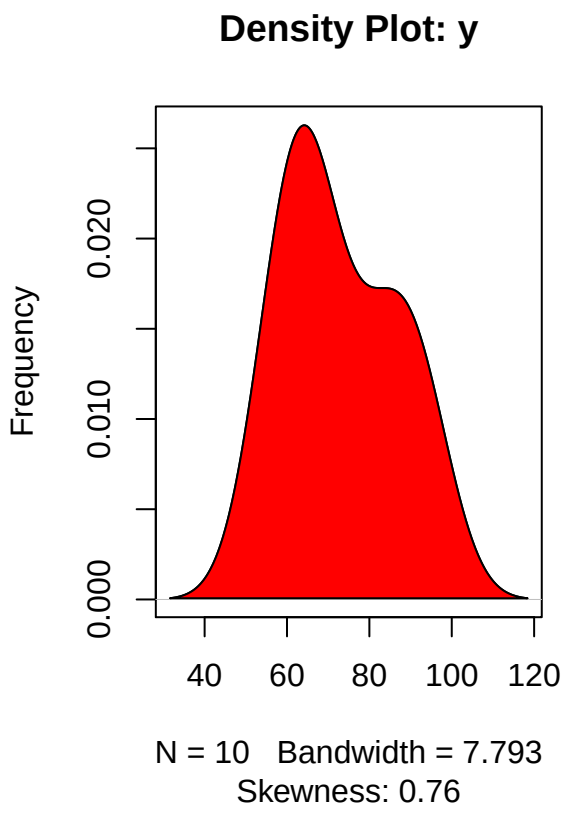
Outlier rows:

Outlier rows:

check whether data meet normal distribution

```
library(e1071)
par(mfrow=c(1, 2))
plot(density(mydata$x), main="Density Plot: x", ylab="Frequency", sub=paste("Skewness:", round(e1071::sk
polygon(density(mydata$x), col="red")
plot(density(mydata$y), main="Density Plot: y", ylab="Frequency", sub=paste("Skewness:", round(e1071::sk
polygon(density(mydata$y), col="red")
```





calculate coefficient

```
cor(mydata$x, mydata$y)
```

```
## [1] 0.9827846
```

build a linear model

```
linearMod <- lm(y ~ x, data= mydata)
print(linearMod)
```

```
##
## Call:
## lm(formula = y ~ x, data = mydata)
##
## Coefficients:
## (Intercept)            x
##     30.5818       0.2227
```

check of statistic significance

```
summary(linearMod)
```

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

```
## lm(formula = y ~ x, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1273 -1.6045  0.6909  1.9182  2.6909
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.5818     2.9384   10.41 6.29e-06 ***
## x             0.2227     0.0148   15.04 3.76e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.689 on 8 degrees of freedom
## Multiple R-squared:  0.9659, Adjusted R-squared:  0.9616
## F-statistic: 226.4 on 1 and 8 DF,  p-value: 3.764e-07
```