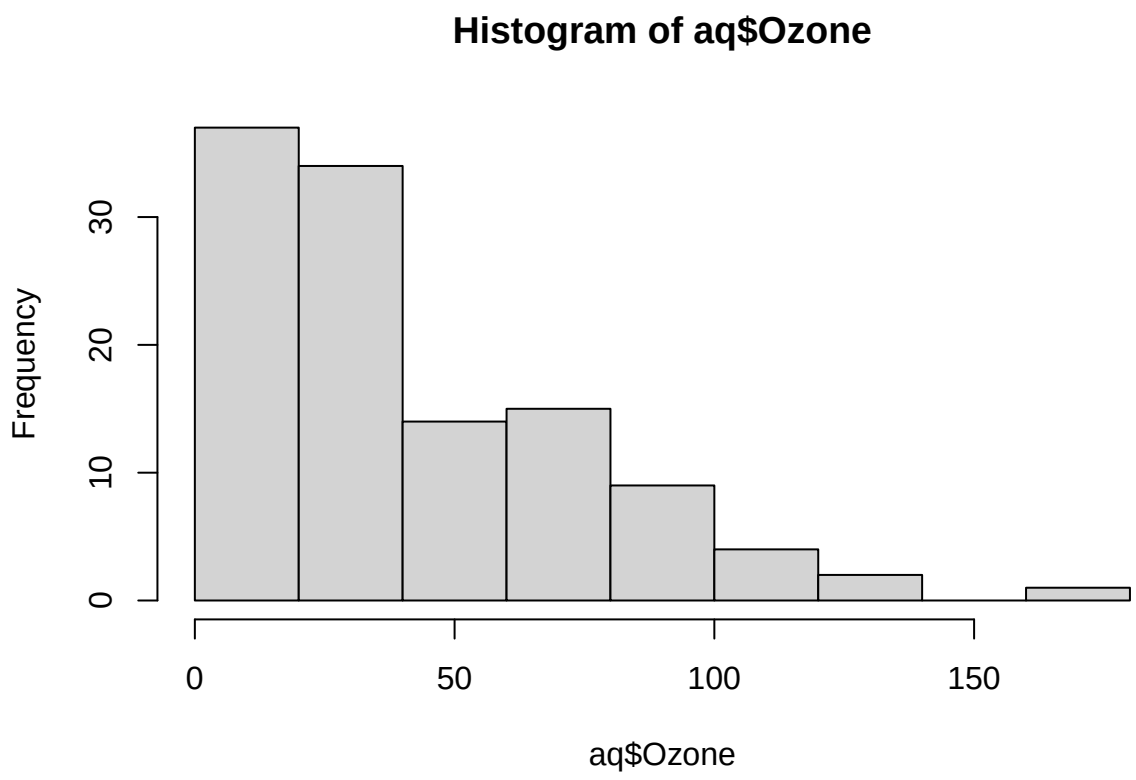
Session-14_15.R

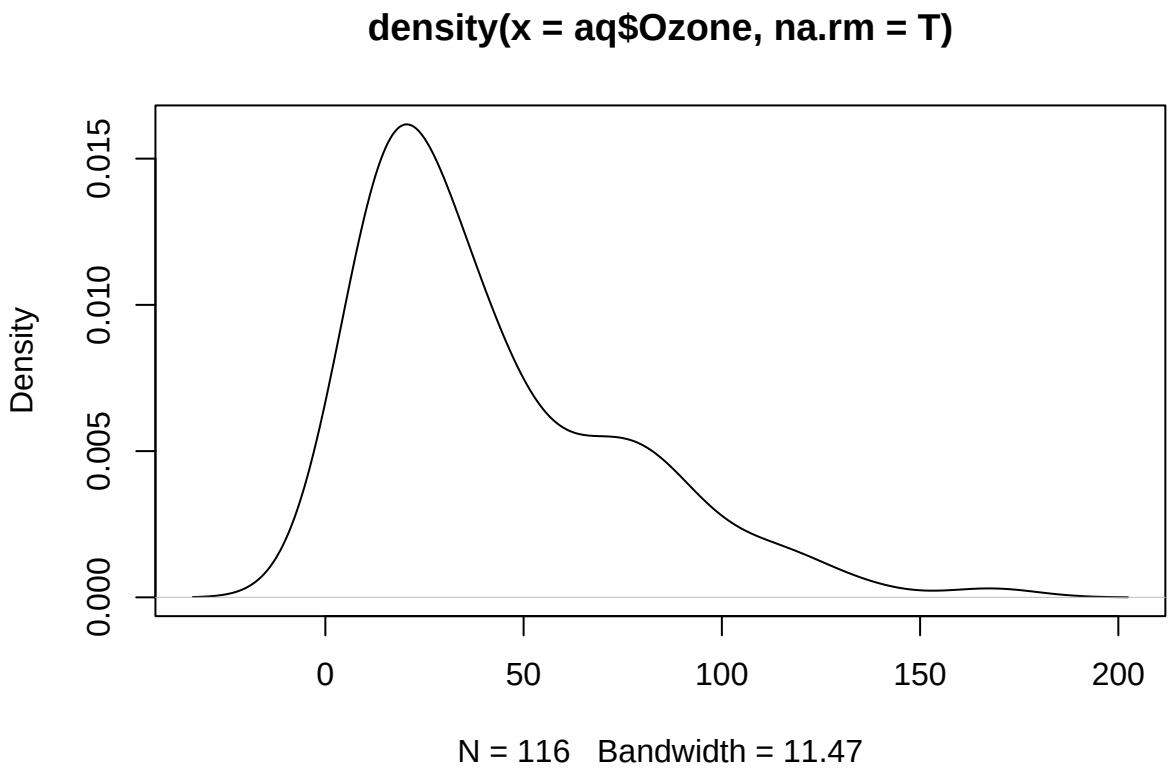
SumanPaudel

2024-04-12

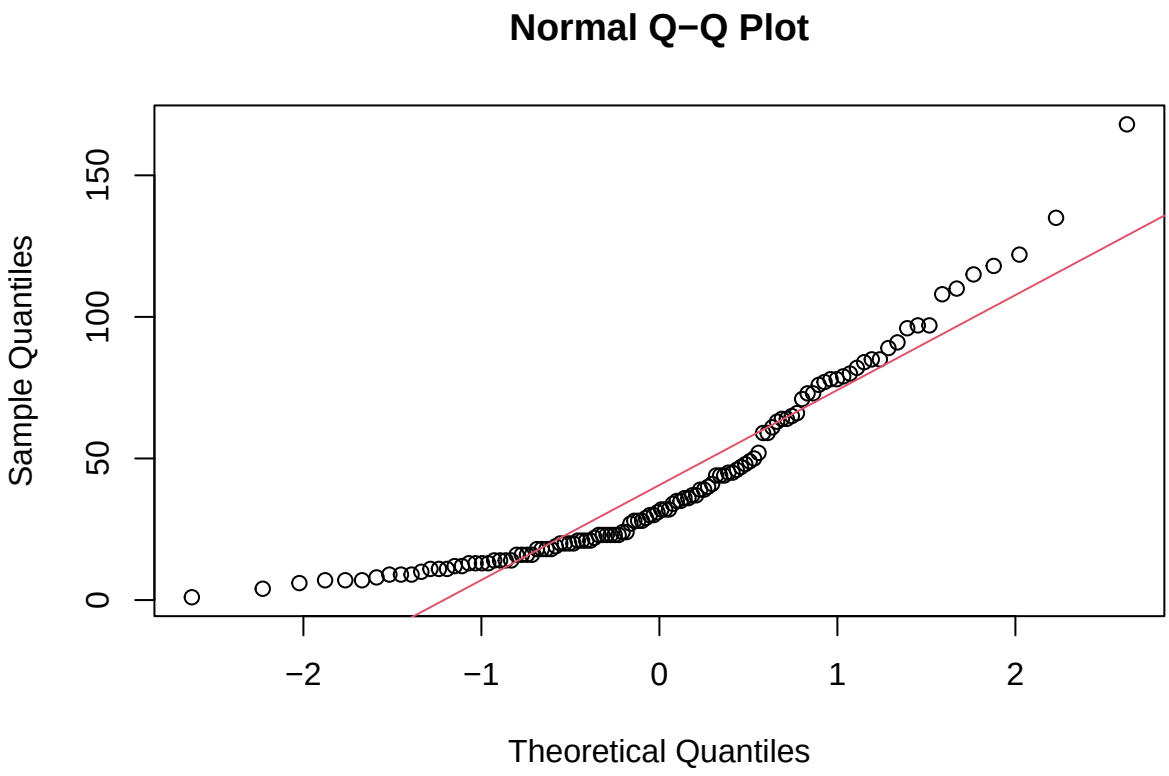
```
aq <- airquality

hist(aq$Ozone)
```




```
plot(density(aq$Ozone,na.rm=T))
```



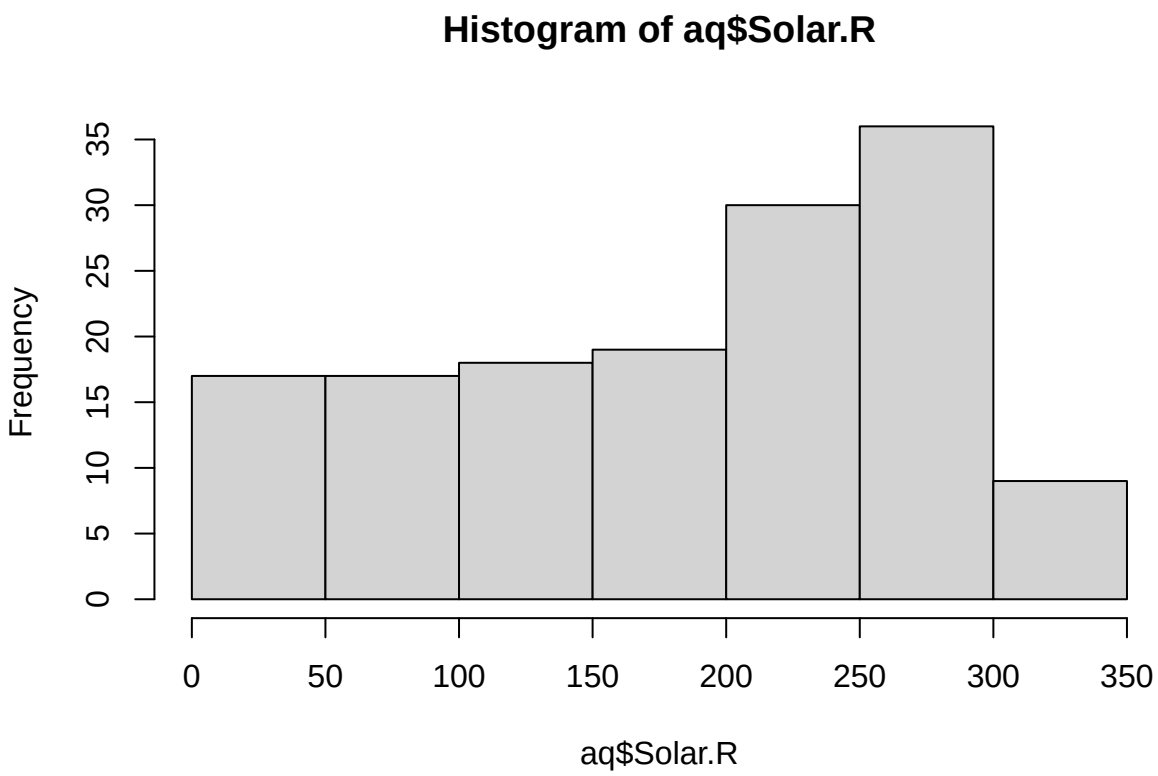

```
qqnorm(aq$Ozone)
qqline(aq$Ozone, col=2)
```



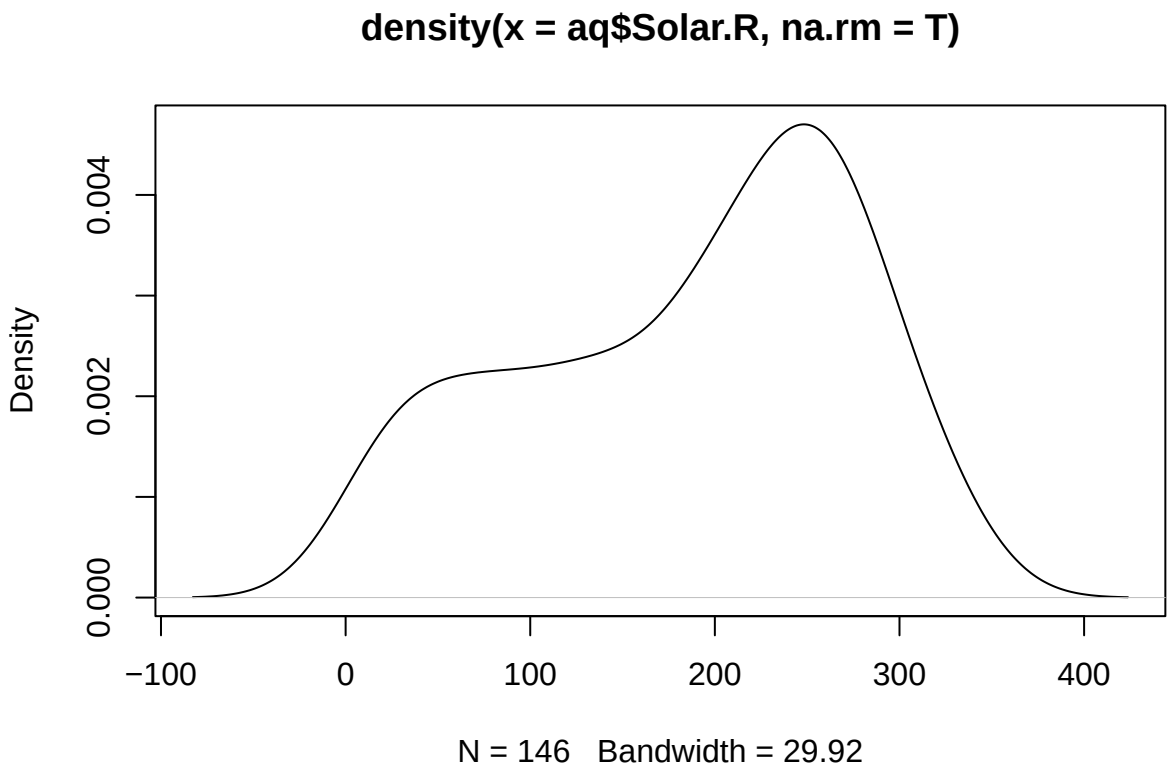

```
shapiro.test(aq$Ozone)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  aq$Ozone
## W = 0.87867, p-value = 2.79e-08
```

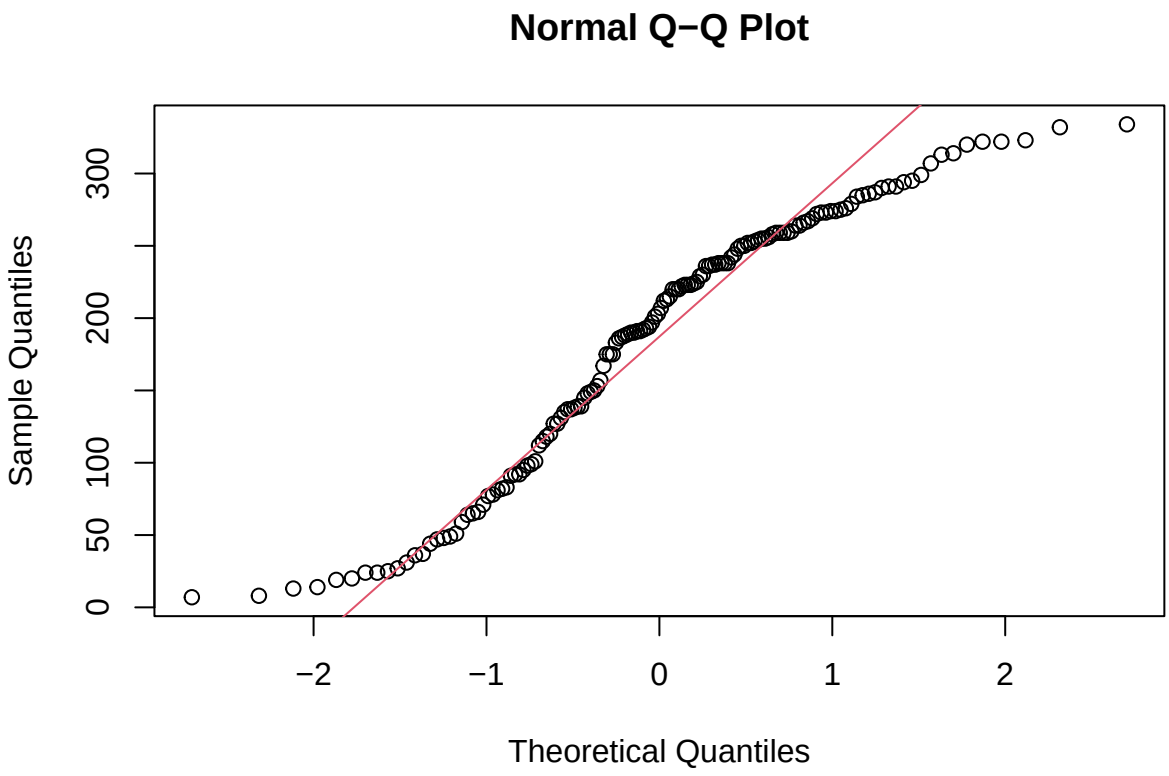
```
hist(aq$Solar.R)
```




```
plot(density(aq$Solar.R,na.rm=T))
```



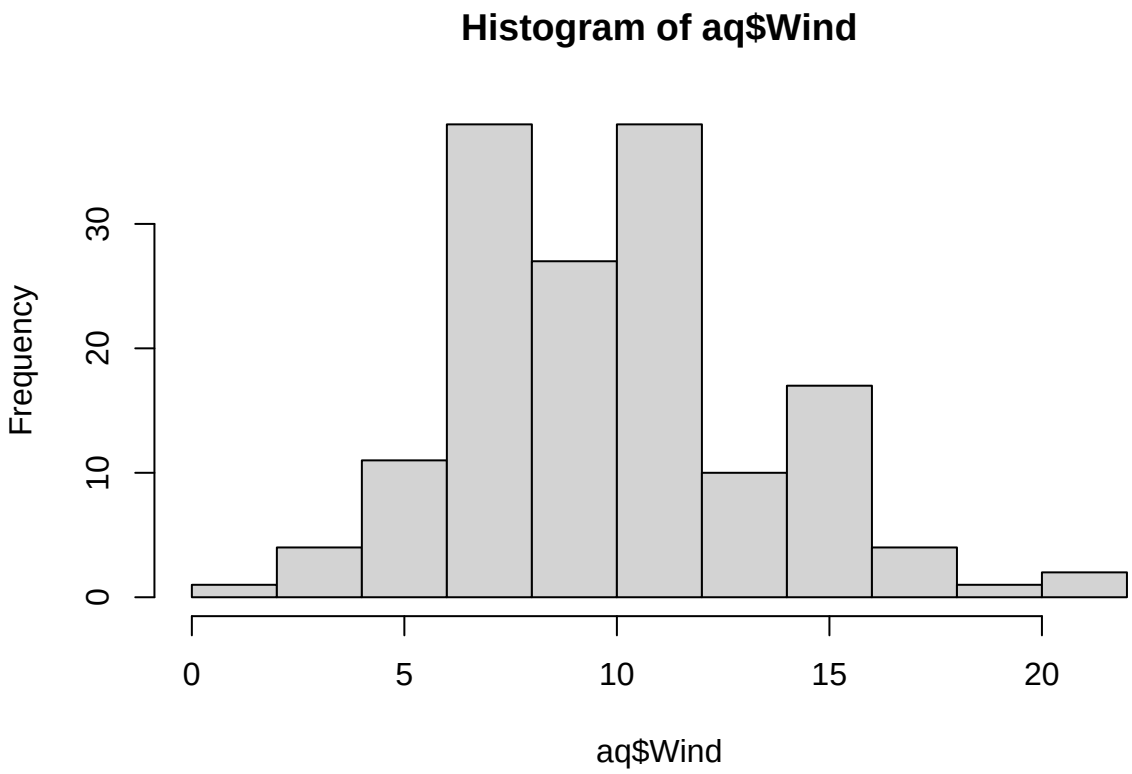

```
qqnorm(aq$Solar.R)
qqline(aq$Solar.R, col=2)
```



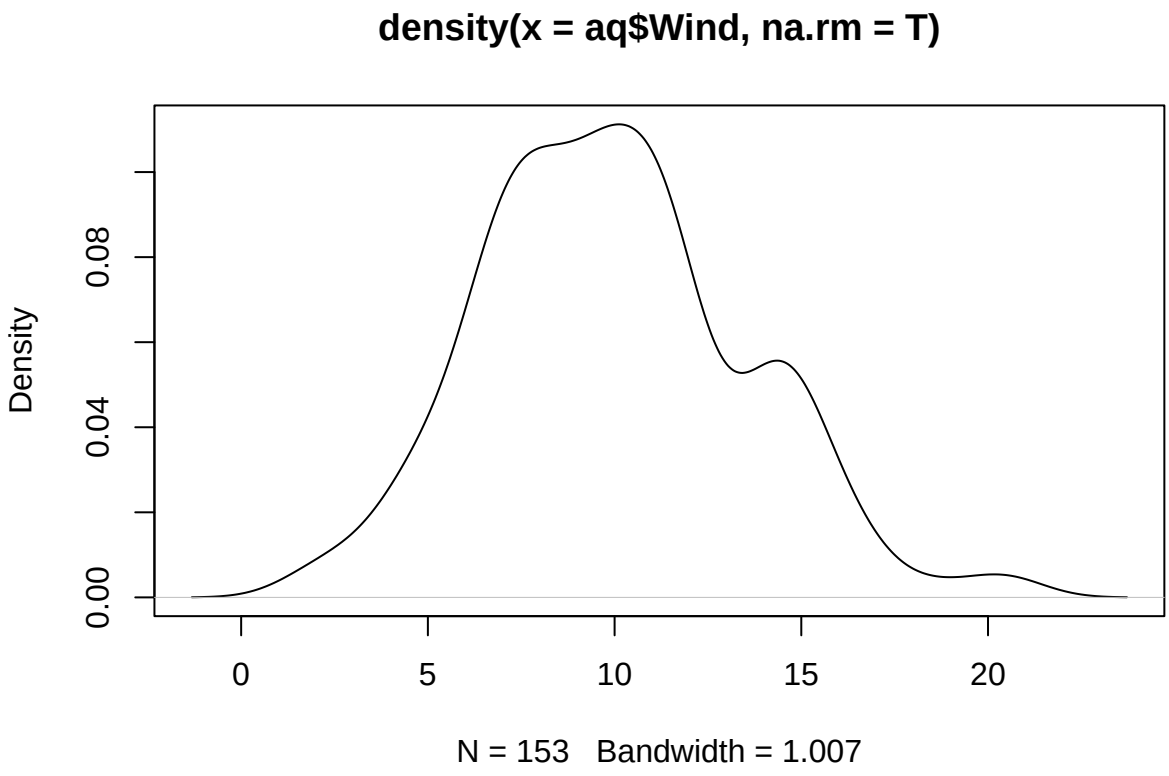

```
shapiro.test(aq$Solar.R)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  aq$Solar.R
## W = 0.94183, p-value = 9.492e-06
```

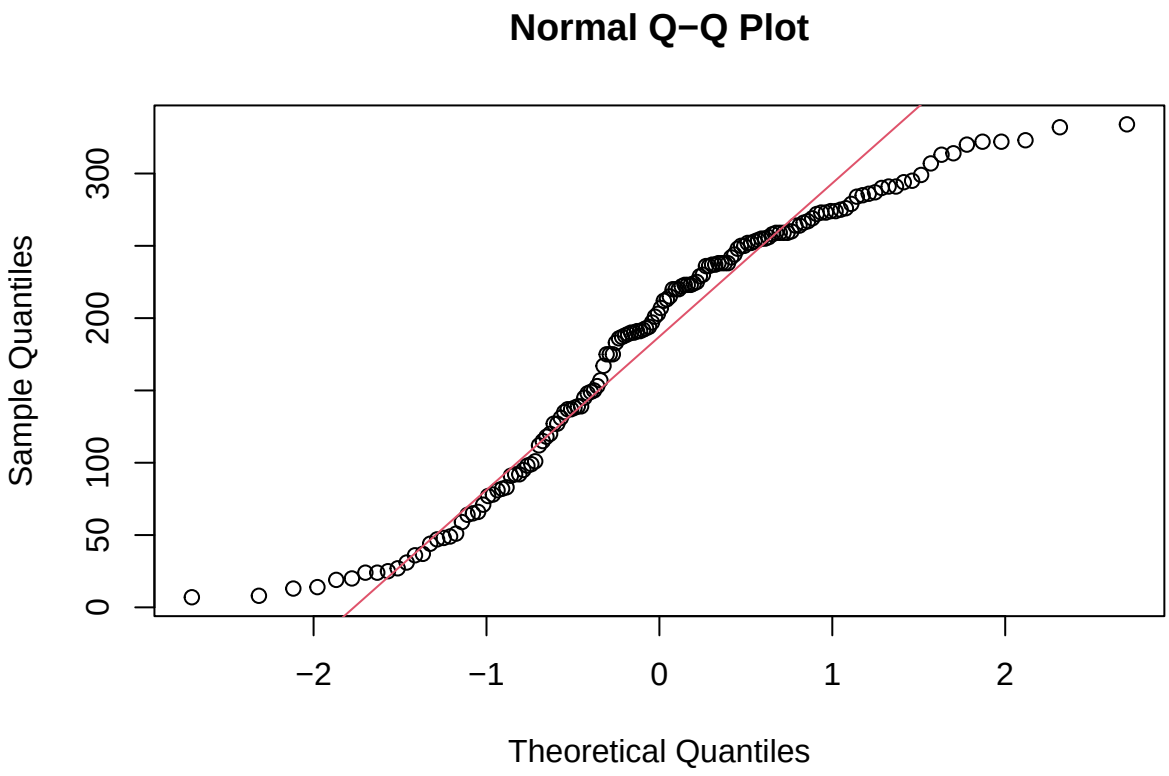
```
# testing normality for wind
hist(aq$Wind)
```




```
plot(density(aq$Wind,na.rm=T))
```



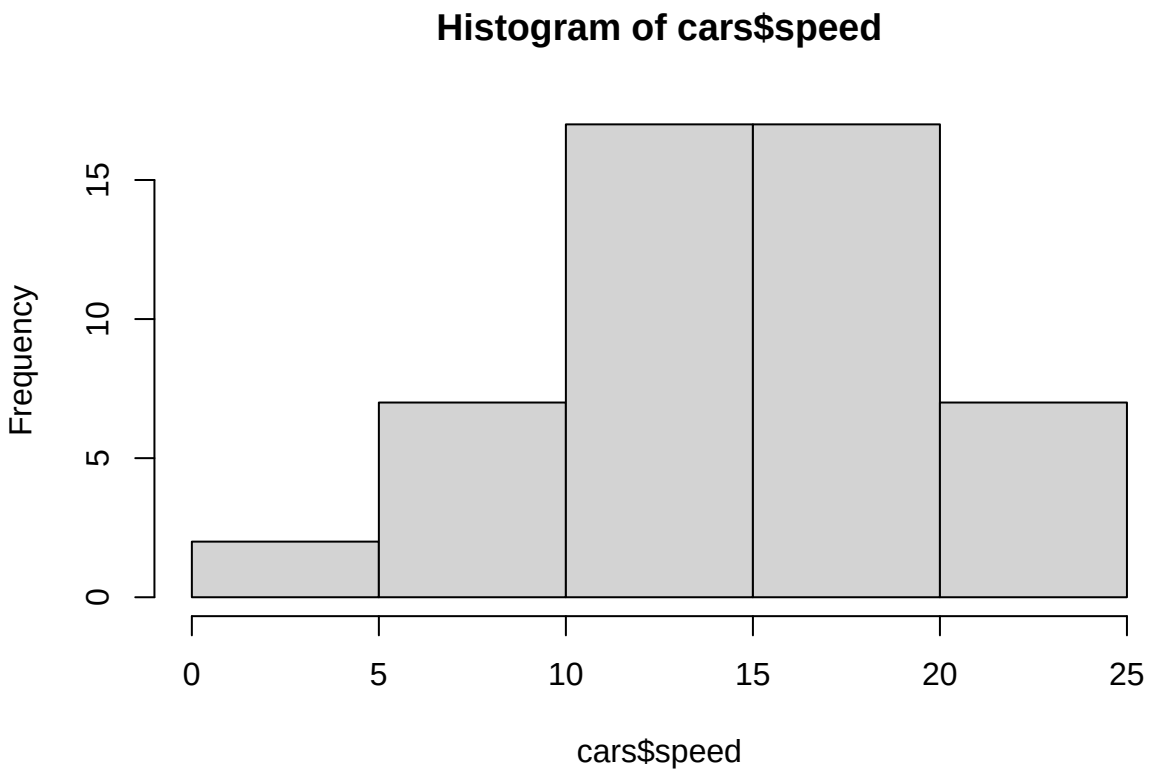

```
qqnorm(aq$Solar.R)
qqline(aq$Solar.R, col=2)
```



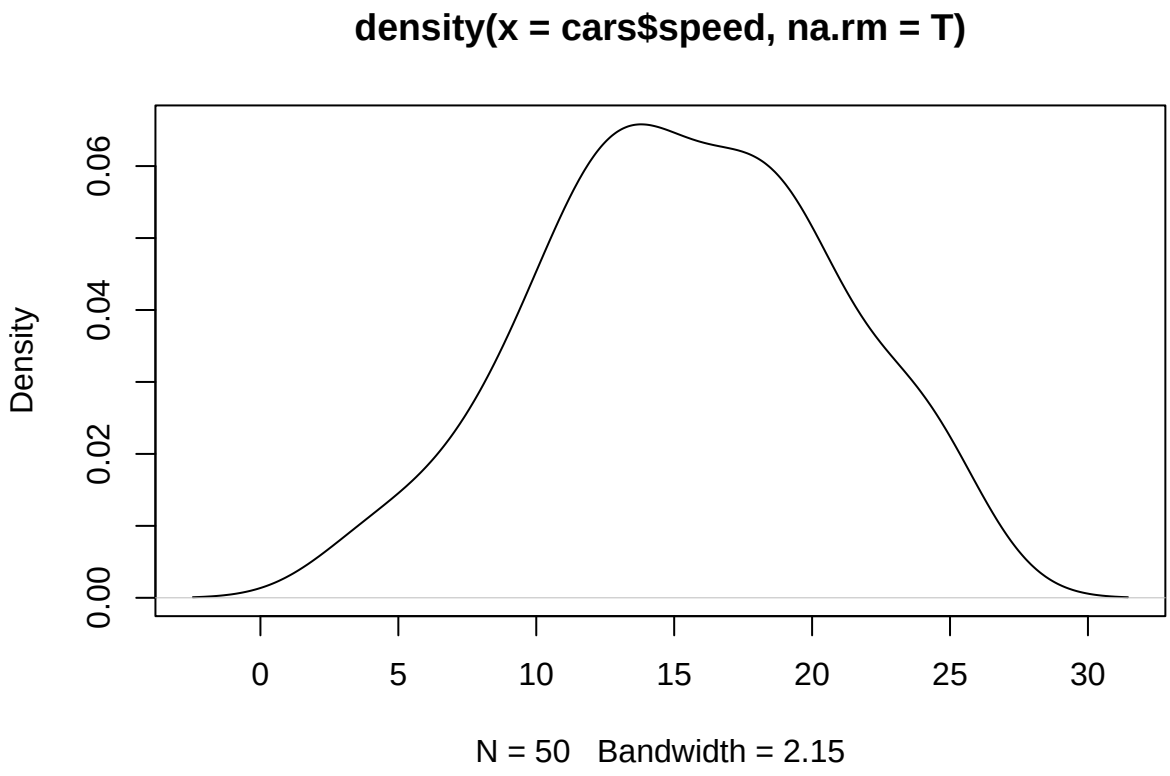

```
shapiro.test(aq$Wind)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  aq$Wind
## W = 0.98575, p-value = 0.1178
```

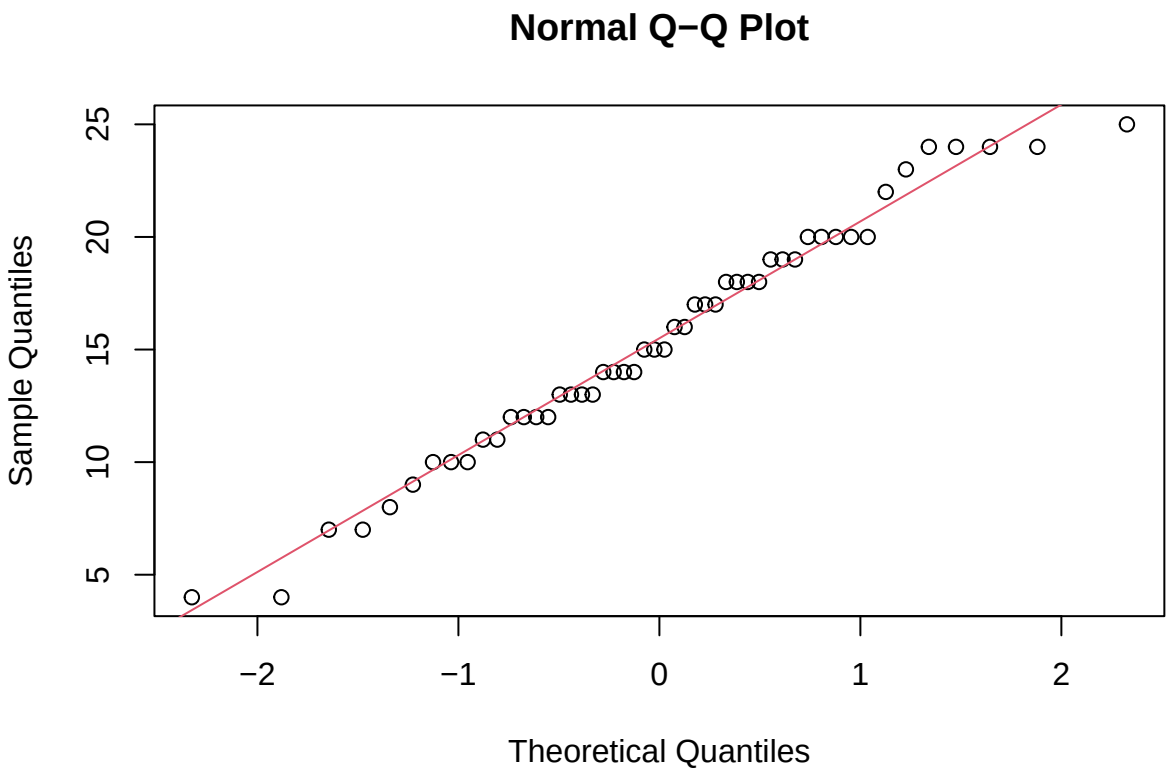
```
# testing normality for wind
# goodness of fit
hist(cars$speed)
```




```
plot(density(cars$speed,na.rm=T))
```




```
qqnorm(cars$speed)
qqline(cars$speed,col=2)
```

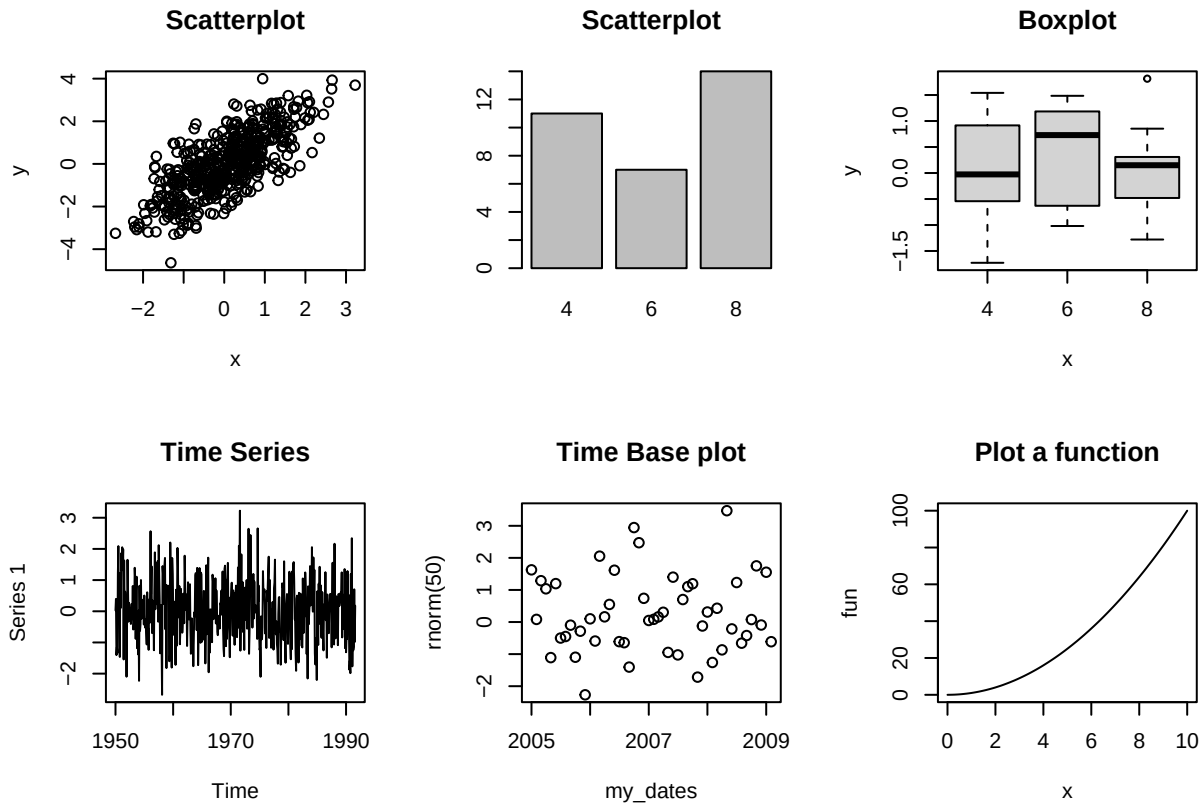



```
shapiro.test(cars$speed)
```

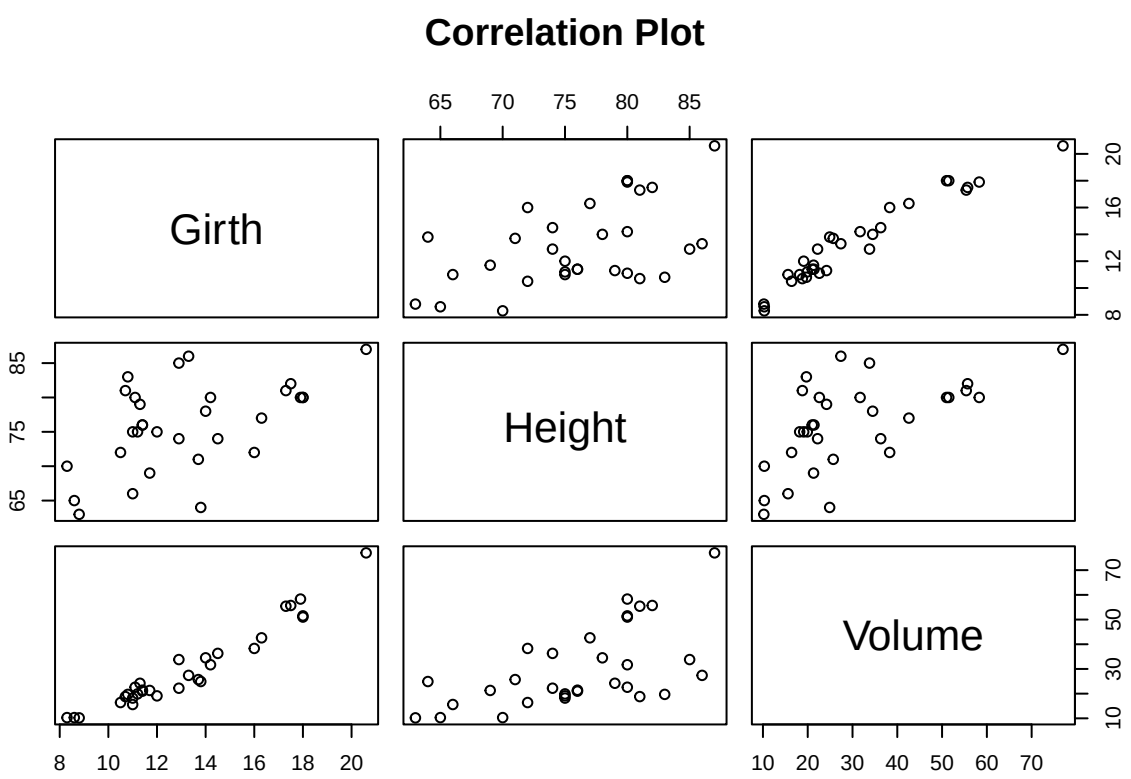
```
##
##  Shapiro-Wilk normality test
##
## data:  cars$speed
## W = 0.97765, p-value = 0.4576
```

```
# test of normality: types
# sekwness and kurtosis based
# # jarque-bera test of normality

# larges sample based
# kolmogorav-smirnov test of normality
# ks.test(cars$speed)

# small sample based
# shapiro wil test of normality


# multiple graph in single wind

x <- rnorm(500)
y <- x + rnorm(500)
```

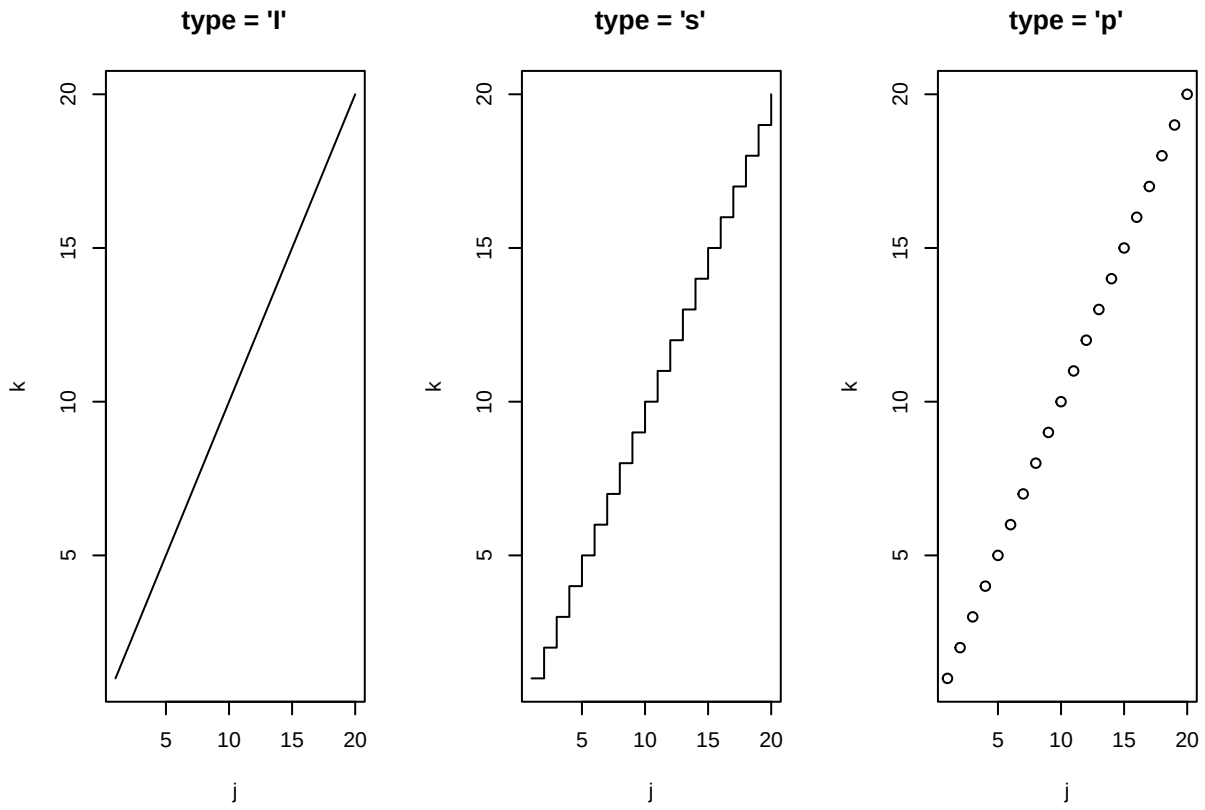
```r
my_ts <-
  ts(
    matrix(x,
    nrow = 500,
    ncol = 1),
    start = c(1950, 1),
    frequency = 12
  )

my_dates <- seq(as.Date("2005/1/1"),by="month",length=50)


my_factor <- factor(mtcars$cyl)

fun <- function(x) x^2

par(mfrow=c(2,3))
plot(x,y,main="Scatterplot")
plot(my_factor,main="Scatterplot")
plot(my_factor, rnorm(32), main="Boxplot")
plot(my_ts, main="Time Series")
plot(my_dates, rnorm(50), main= "Time Base plot")
plot(fun, 0, 10, main="Plot a function")
```



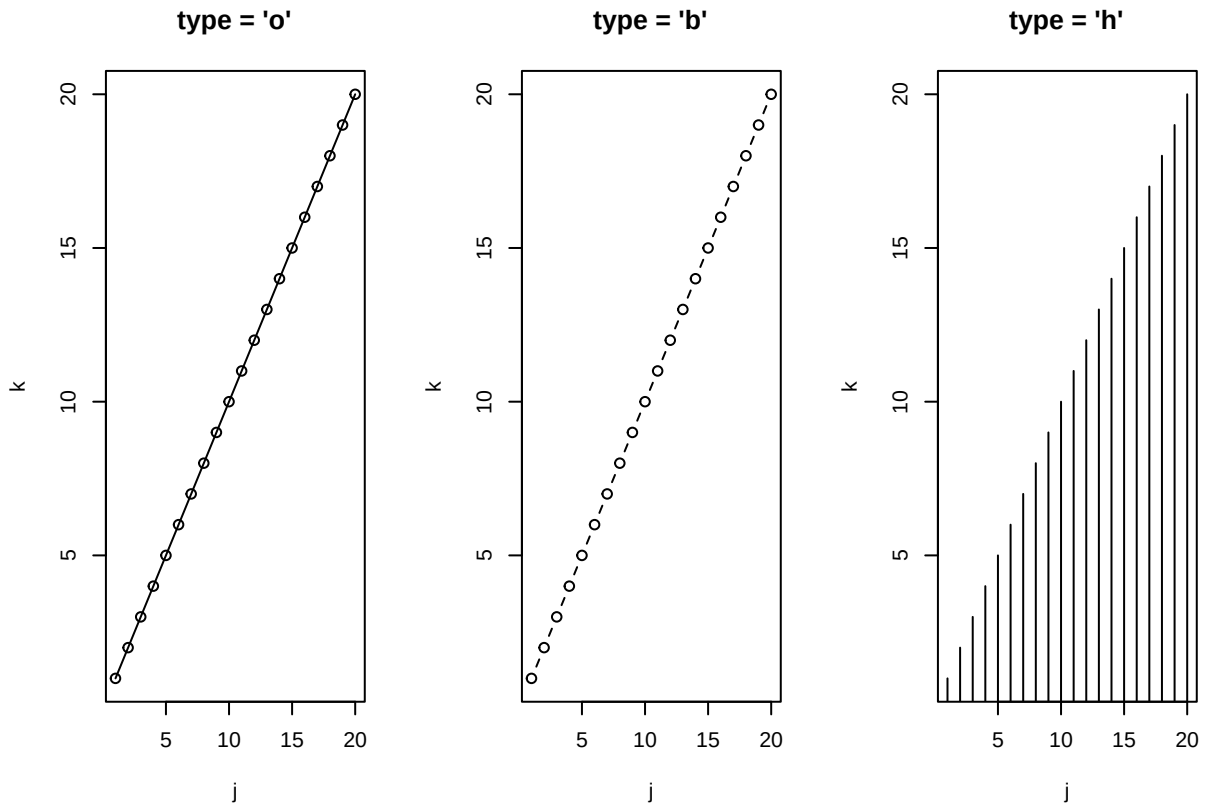
```r
par(mfrow=c(1,1))
plot(trees[,1:3], main="Correlation Plot")
```



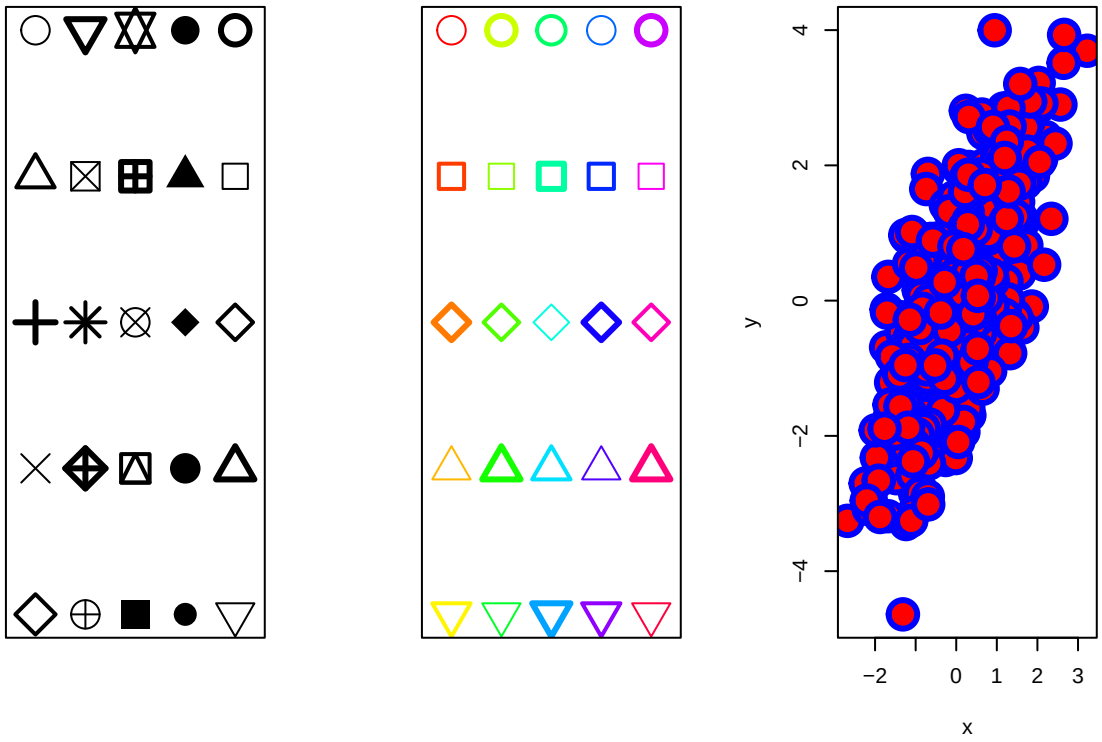

```r
# if linear use pearson coeffiecent
# if not linear use spearman coeffiecent


# bi serial
# for testing both categorical and continuous

j <- 1:20
k <- j
par(mfrow=c(1,3))

plot(j,k,type="l",main="type = 'I'")

plot(j,k,type="s",main="type = 's'")
plot(j,k,type="p",main="type = 'p'")
```




```
par(mfrow=c(1,3))

plot(j,k,type="o",main="type = 'o'")

plot(j,k,type="b",main="type = 'b'")
plot(j,k,type="h",main="type = 'h'")
```




```
r <- c(sapply(seq(5, 25, 5), function(i) rep(i,5)))

t <- rep(seq(25, 5, -5), 5)

plot(r, t, pch=1:25, cex=3, yaxt="n", xaxt="n",ann=F,xlim=c(3,27),lwd=1:3)

plot(
  r,
  t,
  pch = 21:25,
  cex = 3,
  yaxt = "n",
  xaxt = "n",
  ann = F,
  xlim = c(3, 27),
  lwd = 1:3,
  col = rainbow(25)
)


plot(x,y, pch=21, bg="red",col="blue",cex=3,lwd=3)
```




```
# why values in the scatterplot lie in the range of -4 to 4
# x <- rnom(5)

plot(x,y, main=latex2exp::TeX('$\\beta^3,\\beta\\in 1 \\ | dots 10$'))


# networks
# graph analysis

# Session 15
# SNA Basics
library(graph)
```

```
## Loading required package: BiocGenerics

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
```

```
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min
```

```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 4.3.3
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:graph':
##
##     degree, edges, intersection, union
```

```
## The following objects are masked from 'package:BiocGenerics':
##
##     normalize, path, union
```

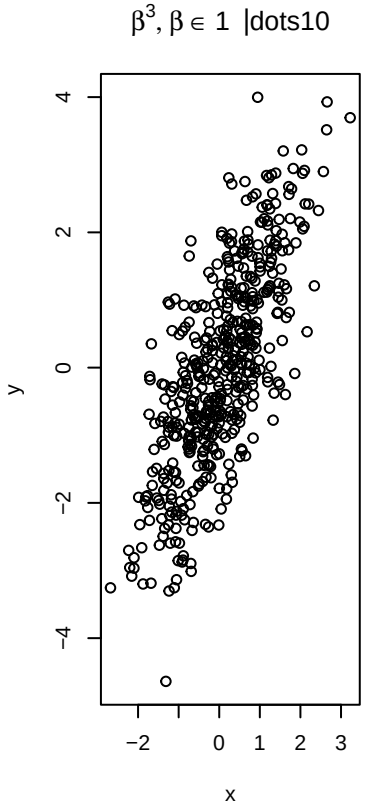
```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

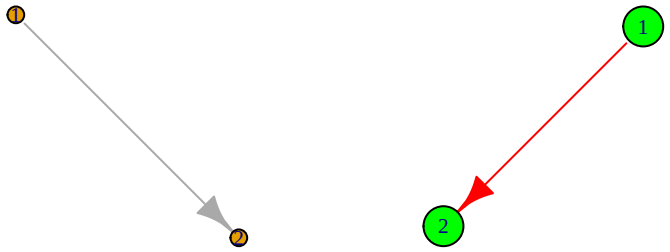
```
## The following object is masked from 'package:base':
##
##     union
```

```
library(Rgraphviz)
```

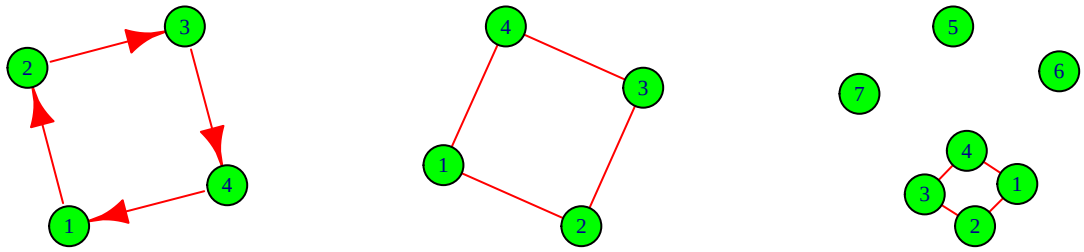
```
## Loading required package: grid
```

```
g <- graph((c(1,2)))
plot(g)
plot(g, vertex.color='green',vertex.size=40,edge.color='red', edge.size=20)
```






```
g <- graph(c(1,2,2,3,3,4,4,1))
plot(g, vertex.color='green',vertex.size=40,edge.color='red', edge.size=20)


g <- graph(c(1,2,2,3,3,4,4,1),directed = F)
plot(g, vertex.color='green',vertex.size=40,edge.color='red', edge.size=20)

g <- graph(c(1, 2, 2, 3, 3, 4, 4, 1), directed = F, n = 7)
plot(g, vertex.color='green',vertex.size=40,edge.color='red', edge.size=20)
```




```
# defining nodes with text data

g1 <- graph(c("Sita","Ram","Ram","Rita","Rita","Sita","Sita","Rita","Anju","Ram"))
plot(g1, vertex.color='green',vertex.size=40,edge.color='red', edge.size=20)

igraph::degree(g1)
```

```
## Sita  Ram Rita Anju
##    3    3    3    1
```

```
igraph::degree(g1, mode="in")
```

```
## Sita  Ram Rita Anju
##    1    2    2    0
```

```
diameter(g1, directed = F)
```

```
## [1] 2
```

```
edge_density(g1, loops = F)
```

```
## [1] 0.4166667
```

```r
# edge density
ecount(g1) / (vcount(g1)*(vcount(g1-1)))
```

```
## [1] 0.4166667
```

```r
reciprocity(g1)
```

```
## [1] 0.4
```

```r
closeness(g1, mode='all', weights=NA)
```

```
##      Sita       Ram      Rita      Anju
## 0.2500000 0.3333333 0.2500000 0.2000000
```

```r
betweenness(g1, directed = T, weights = NA)
```

```
## Sita  Ram Rita Anju
##    1    2    2    0
```