

## Κεφάλαιο 4

# Γραφήματα

Τα *γραφήματα* είναι πολύ χρήσιμα για την οπτική αναπαράσταση των δεδομένων και καθοδηγούν τον στατιστικό στην διαδικασία της μοντελοποίησης και αξιολόγησης της ανάλυσης. Το κεφάλαιο αυτό περιγράφει μερικές χρήσιμες συναρτήσεις γραφημάτων που υπάρχουν στην R και κάνει εισαγωγή στις διάφορες γραφικές παραμέτρους όπως την εισαγωγή πληροφοριών στο γράφημα αλλά και την εποπτική συσχέτιση. Όπως θα δούμε, η R δίνει ένα πολύ ισχυρό περιβάλλον για τη δημιουργία γραφημάτων.

Εκτός από τα γραφήματα και τα χαρακτηριστικά τους τα οποία θα δούμε πιο κάτω, η R περιλαμβάνει και τη βιβλιοθήκη Trellis Graphics. Τα γραφήματα Trellis έχουν περισσότερη ευελιξία και μπορούν να χρησιμοποιηθούν για πολλαπλά γραφήματα και βελτιωμένες τρισδιάστατες αναπαραστάσεις.

### 4.1 Απλά Γραφήματα

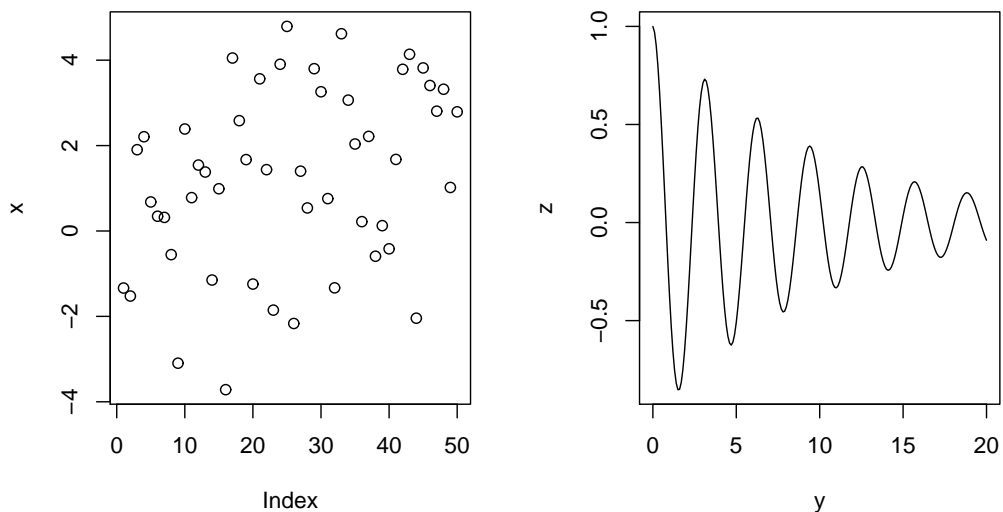
Τα πιο απλά γραφήματα είναι οι γραφικές παραστάσεις που συσχετίζονται με μονοδιάστατη τυχαία μεταβλητή, οι γραφικές παραστάσεις συναρτήσεων και οι γραφικές παραστάσεις χρονοσειρών. Η βασική εντολή για γραφική παράσταση είναι η εντολή `plot`, η οποία έχει πολλές δυνατότητες και μπορεί να πάρει διάφορες γραφικές παραμέτρους για ορίσματα. Ακολουθούν μερικά απλά παραδείγματα.

```
> x <- rnorm(50, mean=1, sd=2)
> plot(x)
> y <- seq(0,20, .1)
> z <- exp(-y/10)*cos(2*y)
```

---

```
> plot(y,z, type="l")
```

Η τέταρτη εντολή δίνει το γράφημα της  $f(y) = e^{-\frac{y}{10}} \cos 2y$ . Με αυτόν τον τρόπο δουλεύουμε συνήθως όταν θέλουμε να δημιουργήσουμε γραφικές παραστάσεις συναρτήσεων. Τα αντίστοιχα γραφήματα παρουσιάζονται στο Σχήμα 4.1.



Σχήμα 4.1: Απλά γραφήματα

Όπως φαίνεται στο δεύτερο παράδειγμα, ένα διάγραμμα διασποράς (scatter plot) μπορεί να κατασκευαστεί στην R εφαρμόζοντας την εντολή `plot` σε ένα ζεύγος διανυσμάτων της ίδιας διάστασης, ή σε μια λίστα με συνιστώσες `x` και `y`.

```
> x1 <- rnorm(50)
> plot(x1,x)
> plot(cbind(x1,x))
```

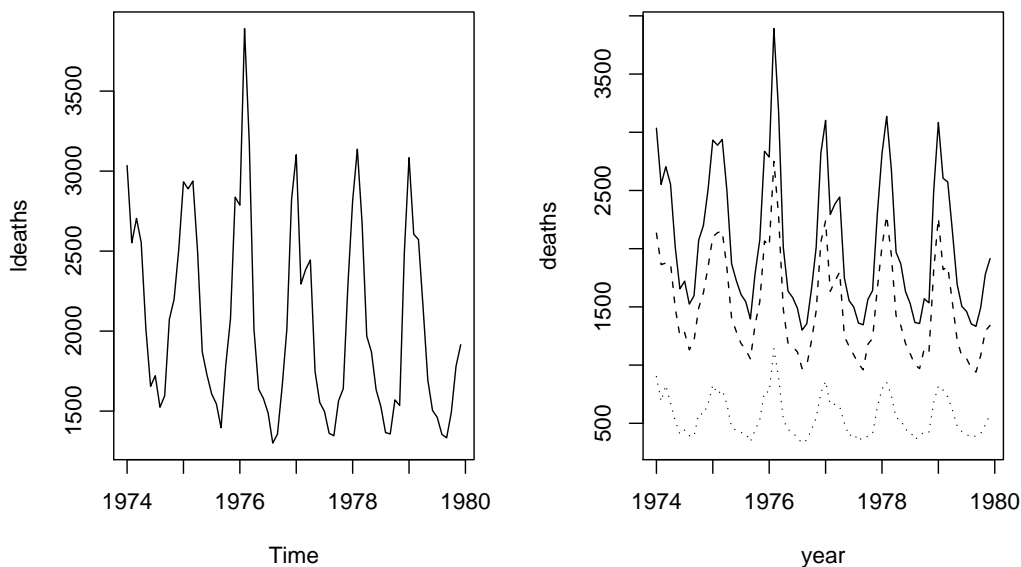
Όταν τα δεδομένα παρατηρούνται διαδοχικά στον χρόνο, είναι φυσικό να γίνει η γραφική παράσταση των δεδομένων σε συνάρτηση με τον χρόνο (χρονοσειρές). Στην R αυτό γίνεται χρησιμοποιώντας την εντολή `ts.plot`. Για παράδειγμα, έστω τα πλαίσια δεδομένων `ldeaths`, `mdeaths` και `fdeaths` που ανήκουν στην R και αναφέρονται στους μηνιαίους θανάτους από καρκίνο του πνεύμονα στο Ηνωμένο

---

Βασιλείο κατά την περίοδο από το 1974 ως το 1979 συνολικά, στους άντρες και στις γυναίκες, αντίστοιχα.

```
> ts.plot(ldeaths)
> ts.plot(ldeaths,mdeaths,fdeaths,gpars=list(xlab="year",ylab="deaths",lty=1:3))
```

Οι γραφικές παραστάσεις φαίνονται στο Σχήμα 4.2 και παρατηρείται ότι στο δεύτερο γράφημα κάθε χρονοσειρά παρουσιάζεται με διαφορετικό είδος γραμμής.



Σχήμα 4.2: Γραφήματα χρονοσειρών.

## 4.2 Γραφικές Δυνατότητες

Υπάρχουν πολλές γραφικές δυνατότητες αλλά η παρουσίαση θα περιοριστεί μόνο σε μερικές οι οποίες στο τέλος είναι πιο χρήσιμες από τις υπόλοιπες. Η διαρρύθμιση του γραφήματος μπορεί να τακτοποιηθεί έτσι ώστε να παρουσιάζει περισσότερες από μια γραφική παράσταση (βλέπε Σχήμα 4.3).

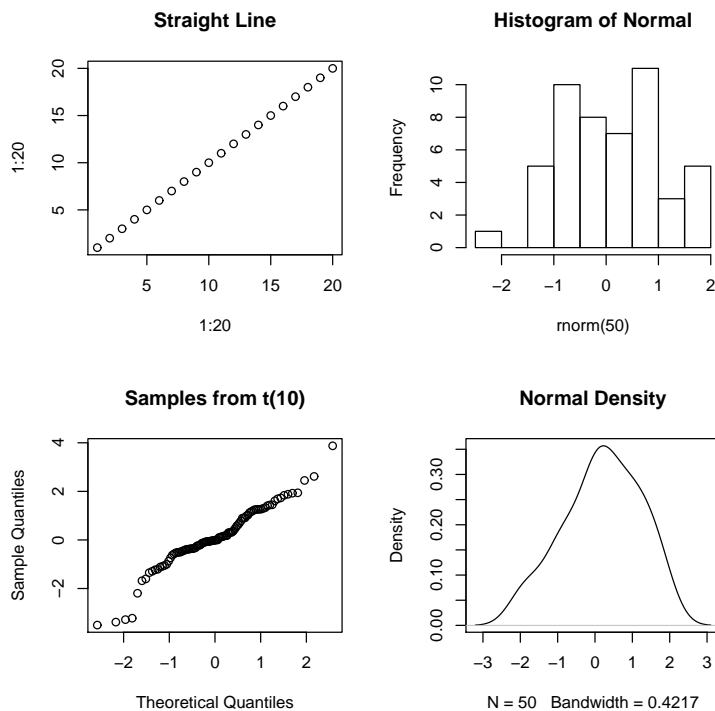
```
> par(mfrow=c(2,2)) #a 2x2 plot
```

```

> plot(1:20,1:20,main="Straight Line")
> hist(rnorm(50),main="Histogram of Normal")
> qqnorm(rt(100,10),main="Samples from t(10)")
> plot(density(rnorm(50)),main="Normal Density")
> par(mfrow=c(1,1))

```

Το πρώτο γράφημα (Σχήμα 4.3, πάνω αριστερά) δίνει τη γραφική παράσταση της  $f(x) = x, x \in [1, 20]$ . Το δεύτερο γράφημα (Σχήμα 4.3, πάνω δεξιά) μας δίνει το ιστόγραμμα 50 τυχαίων παρατηρήσεων από την τυπική κανονική. Παρόμοια, το τρίτο γράφημα (Σχήμα 4.3, κάτω αριστερά) μας δίνει την γραφική παράσταση των δειγματικών ποσοστημορίων από την  $t$  κατανομή με 10 βαθμούς ελευθερίας ως προς τα θεωρητικά ποσοστημόρια της τυπικής κανονικής. Το τελευταίο γράφημα απεικονίζει τη μη παραμετρική εκτιμήτρια συνάρτησης πυκνότητας πιθανότητας από 50 παρατηρήσεις της τυπικής κανονικής. Παρόμοια διάταξη γραφημάτων μπορεί να επιτευχθεί με την εντολή `split.screen`.



Σχήμα 4.3: 2x2 διαρύθμιση γραφημάτων

---

Έχοντας σαν βάση το προηγούμενο παράδειγμα, είναι αρκετά εύκολο να εισαχθεί κύριος τίτλος ή υπότιτλος σε μια γραφική παράσταση.

```
> plot(x,main="Sample From Normal")
> plot(x,sub="Mean 1 and variance 4")
> plot(x, main="Sample from Normal", sub="Mean 1 and Variance 4")
> plot(x)
> title(main="Sample from Normal", sub="Mean 1 and Variance 4")
```

Επίσης, μπορούν να δοθούν ονόματα στους άξονες χρησιμοποιώντας τα ορίσματα `xlab` και `ylab`, όπως το επόμενο παράδειγμα.

```
> plot(x, xlab="Index", ylab="Sample from Normal")
> plot(x, xlab="", ylab="") #no axis labels
> title(xlab="Index", ylab="Sample from Normal")
```

Η δεύτερη εντολή δε δίνει ονόματα στους άξονες. Οι εντολές `xlim` και `ylim` χρησιμεύουν στο να αλλάξουν τα όρια των αξόνων, θέτοντας αυτά που κάνουν το γράφημα πιο εύκολο για κατανόηση.

### 4.3 Είδη και Γραμμές Γραφικής Παράστασης

Στην R τα δεδομένα μπορούν να απεικονιστούν σε γράφημα με διάφορους τρόπους. Αυτό επιτυγχάνεται με το όρισμα `type` στην εντολή `plot`. Αυτοί οι τρόποι φαίνονται στον πιο κάτω πίνακα.

Σύμβολο	Είδος (Type)
"p"	Σημεία
"l"	Γραμμή
"b"	Γραμμή και Σημεία
"c"	Γραμμή με κενό στα σημεία
"o"	Γραμμή και Σημεία ενωμένα
"h"	Κάθετες γραμμές για κάθε σημείο
"s"	Με Βήμα
"n"	Τίποτα

Πίνακας 4.1: Είδη Γραφικής Παράστασης

---

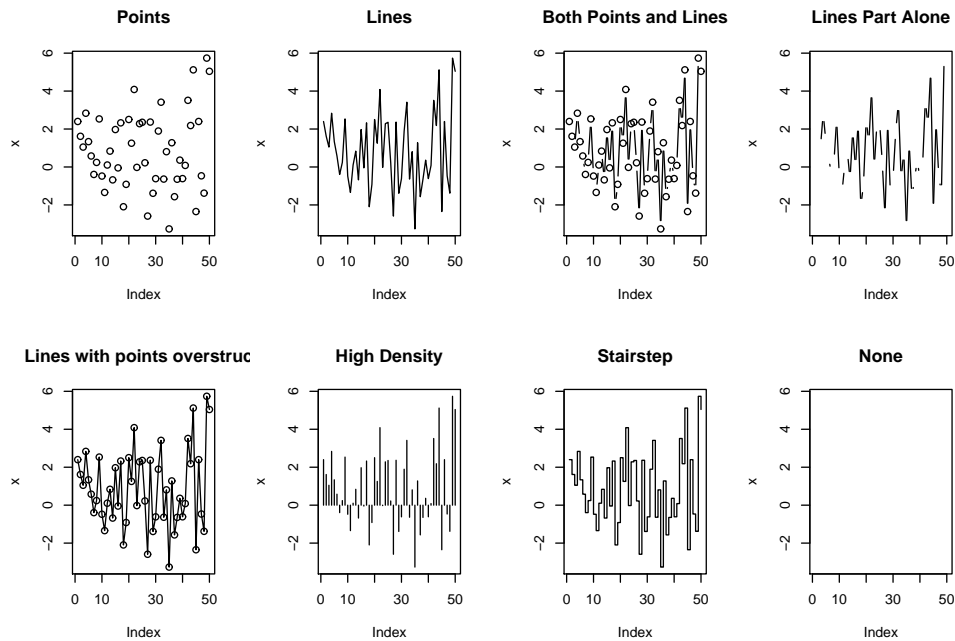
Ακολουθεί ένα παράδειγμα για το πώς χρησιμοποιούνται τα πιο πάνω. Τα γραφήματα φαίνονται στο Σχήμα 4.4. Δημιουργούνται από τα αριστερά προς δεξιά ανά γραμμή.

```
> par(mfrow=c(2,4))
> plot(x, type="p")
> title(main="Points")
> plot(x, type="l")
> title(main="Lines")
> plot(x, type="b")
> title(main="Both Points and Lines")
> plot(x, type="c")
> title(main="Lines Part Alone")
> plot(x, type="o")
> title(main="Lines with points overstruck")
> plot(x, type="h")
> title(main="High Density")
> plot(x, type="s")
> title(main="Stairstep")
> plot(x, type="n")
> title(main="None")
> par(mfrow=c(1,1))
```

Όταν το είδος της γραφικής παράστασης περιλαμβάνει γραμμές, τότε μπορεί να επιλεγεί διαφορετικό είδος γραμμής δίνοντας διάφορους αριθμούς στο όρισμα `lty`. Για παράδειγμα, η διακεκομμένη γραμμή με παύλες συμβολίζεται με `lty=2`. Το εξ' ορισμού είδος γραμμής είναι η συνεχής γραμμή. Υπάρχουν οκτώ διαφορετικά είδη γραμμής. Επιπρόσθετα, μπορούμε να δώσουμε χρώμα στο είδος γραφικής παράστασης δίνοντας αριθμούς ή σε εισαγωγικά τα αγγλικά ονόματα των χρωμάτων στο όρισμα `col` της εντολής `plot` (π.χ. `col="green"` για πράσινο χρώμα).

## 4.4 Προσθήκη Πληροφοριών σε Γράφημα

Σε μερικές περιπτώσεις, είναι αναγκαίο να υποδειχθούν οι απομακρυσμένες τιμές, να προστεθεί ένα κείμενο ή άλλες πληροφορίες σε ένα γράφημα με διαδραστικό τρόπο. Υπάρχουν διάφοροι τρόποι με τους οποίους η R μπορεί να το κάνει αυτό αλληλεπιδρώντας με το χρήστη. Ακολουθεί ένα παράδειγμα για το πώς μπορεί να υποδειχθεί μια απομακρυσμένη τιμή (βλ. Σχήμα 4.5).



Σχήμα 4.4: Είδη γραφικών παραστάσεων.

```

> x <- runif(20)
> y <- 6*x+rnorm(20)
> x <- c(x,3)
> y <- c(y,4)
> plot(x,y)
> identify(x,y, n=1) #R waits until you click the mouse on the selected point
[1] 21

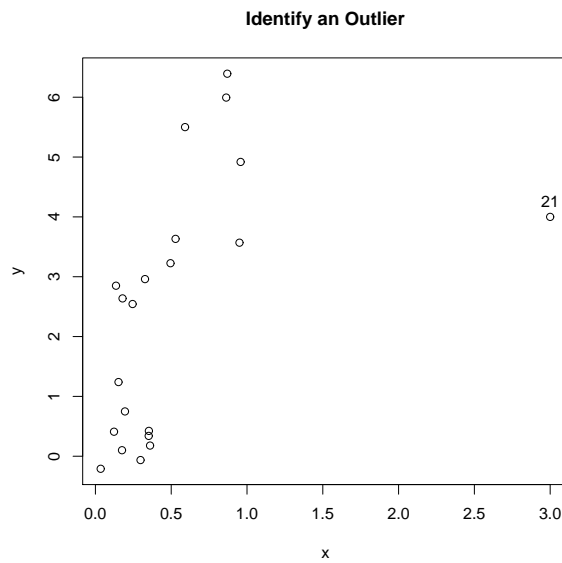
```

Γενικά, μπορούν να υποδειχθούν όσα σημεία επιθυμεί ο χρήστης με την εντολή `identify(x,y, n=k)`, όπου  $k$  είναι ο αριθμός των σημείων που θα υποδειχθούν. Τα ακόλουθα βοηθούν στο να γίνει κατανοητός ο τρόπος που μπορεί να προστεθεί η ευθεία ελαχίστων τετραγώνων, αλλά και καινούργια σημεία ή ευθείες σε μια γραφική παράσταση (βλ. Σχήμα 4.6).

```

> plot(x,y)
> abline(lm(y~x), lty=2)
> plot(y^2, type="l", xlab="", ylab="Square of Y")
> lines(y, lty=2)

```



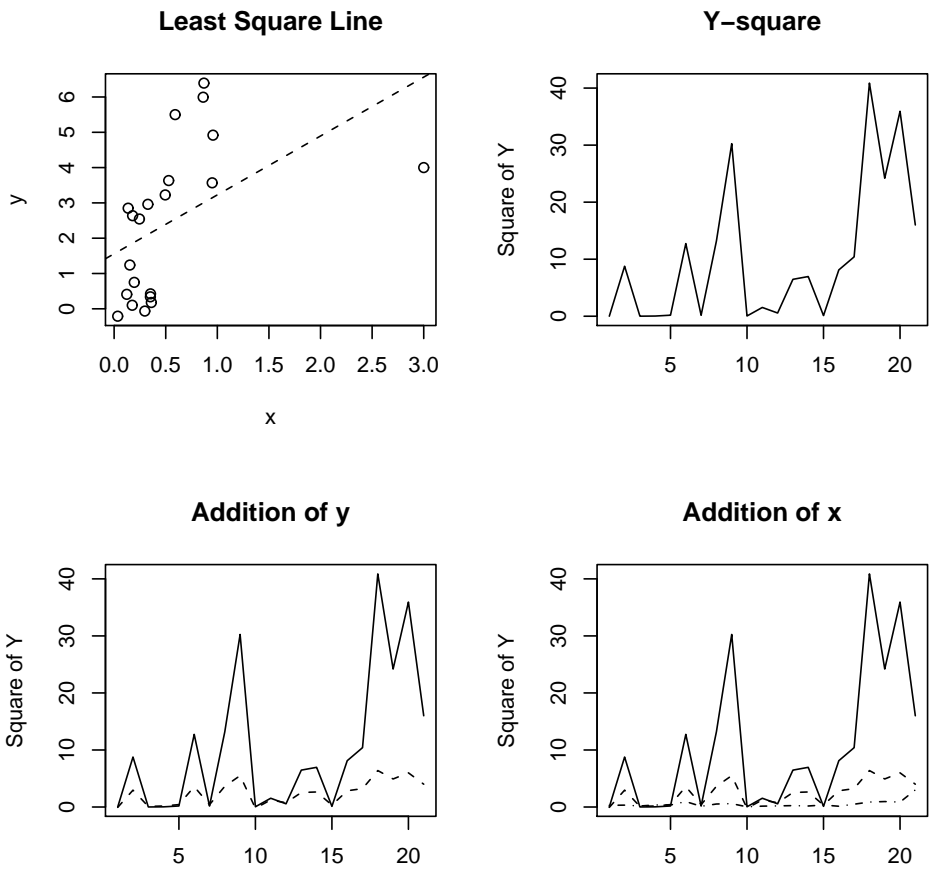
Σχήμα 4.5: Υπόδειξη απομακρυσμένης τιμής.

```
> lines(x, lty=4)
```

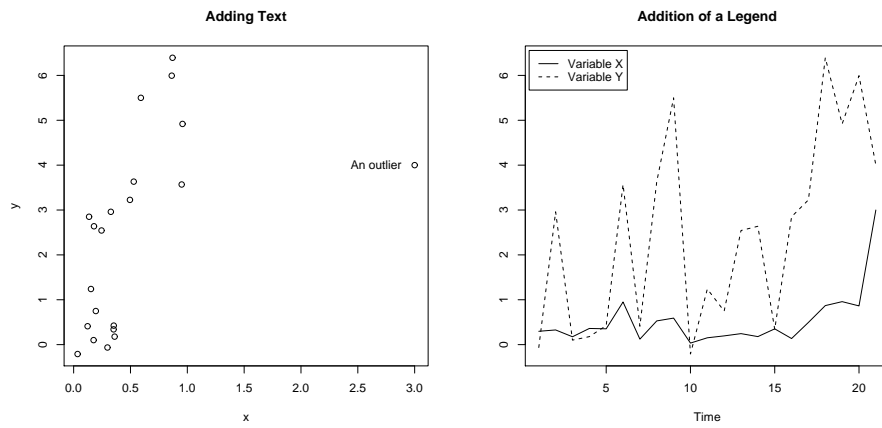
Είναι δυνατόν επίσης να προστεθεί κείμενο και υπόμνημα στη γραφική παράσταση (βλ. Σχήμα 4.7).

```
> plot(x, y, main="Adding Text")
> text(locator(1), "An outlier") #click the mouse to place the text
> ts.plot(ts(x),ts(y),gpars=list(lty=1:2)) #time series plot of both x and y
> leg.names <- c("Variable X","Variable Y")
> legend(locator(1), leg.names, lty=1:2) #click the mouse to place the legend
```





Σχήμα 4.6: Εισαγωγή γραμμών.



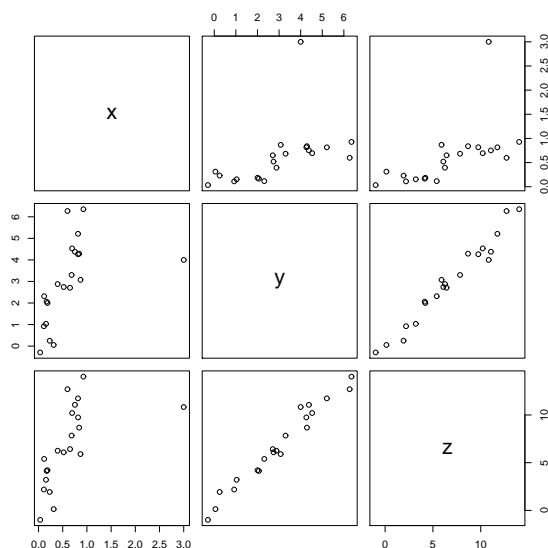
Σχήμα 4.7: Προσθήκη κειμένου και υπομνήματος.

---

## 4.5 Γραφήματα Σε Μεγαλύτερες Διαστάσεις

Για να κατασκευαστούν γραφικές παραστάσεις πολυδιάστατων τυχαίων μεταβλητών, ένας τρόπος είναι να κατασκευαστούν διαγράμματα διασπορών για κάθε ζευγάρι μεταβλητών, ξεχωριστά. Για παράδειγμα, έστω ότι ορίζεται η μεταβλητή  $Z$  από τις προϋπάρχουσες μεταβλητές  $X$  και  $Y$  και κατασκευάζεται πίνακας με τρεις στήλες. Για να ερευνηθεί η συσχέτιση μεταξύ των τριών μεταβλητών χρησιμοποιείται η εντολή `pairs()`, δηλαδή

```
> z <- x+2*y+rnorm(21)
> pairs(cbind(x,y,z))
```



Σχήμα 4.8: Ζευγάρια διαγραμμάτων διασποράς.

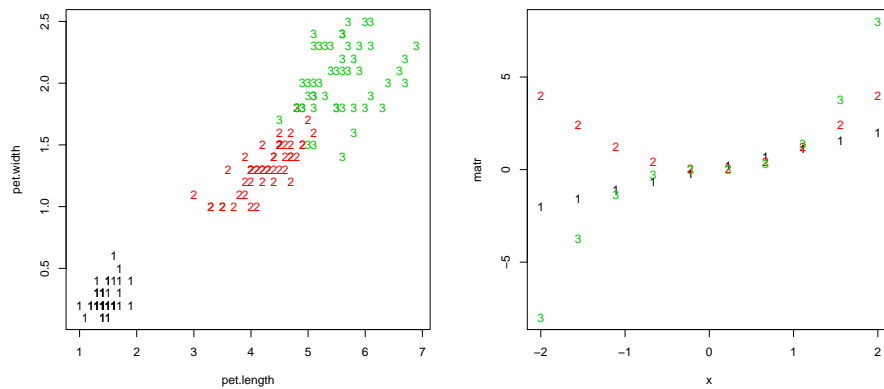
Όταν πρέπει να παρουσιαστούν διάφορα διανύσματα δεδομένων ή πολυδιάστατα δεδομένα στο ίδιο γράφημα, τότε μπορεί να εφαρμοστεί η εντολή `matplot` η οποία κατασκευάζει γραφική παράσταση των στηλών ενός πίνακα συναρτήσει των στηλών κάποιου άλλου. Για σκοπό παραδείγματος, έστω το πλαίσιο δεδομένων `iris3` από το οποίο εξάγουμε το `pet.length` και το `pet.width`. Ο πίνακας `pet.length` περιλαμβάνει 50 παρατηρήσεις (γραμμές) του μήκους του πετάλου τριών ειδών ίριδων (στήλες): *Setosa*, *Versicolor* και *Virginica*. Ο πίνακας `pet.width` περιλαμβάνει 50 παρατηρήσεις (γραμμές) του πλάτους του πετάλου για καθένα από

τα τρία είδη ίριδων. Για να ερευνηθεί γραφικά η συσχέτιση μεταξύ του μήκους και του πλάτους των πετάλων, χρησιμοποιείται η εντολή `matplot` για να παρουσιάσει το μήκος συναρτήσει του πλάτους και για τα τρία είδη σε ένα γράφημα.

```
> pet.length <- iris3[,3,]
> pet.width <- iris3[,4,]
> matplot(pet.length,pet.width)
```

Ακόμη ένα παράδειγμα για τη χρήση της εντολής `matplot` είναι το ακόλουθο:

```
x<-seq(-2,2,length=10)
y<-x^2
z<-x^3
matr<-cbind(x,y,z)
matplot(x,matr)
```



Σχήμα 4.9: Γράφημα `matplot`.

Πολλά είδη δεδομένων μπορούν να παρουσιαστούν σε μορφή επιφάνειας που παράγεται από συναρτήσεις δύο μεταβλητών. Η R παρέχει τρεις εντολές για παρουσίαση τέτοιων δεδομένων. Η πιο απλή, `contour`, παρουσιάζει την επιφάνεια σε διάγραμμα ισοψών. Η προοπτική απεικόνιση της επιφάνειας γίνεται με την εντολή `persp`. Τέλος η εντολή `image` παρουσιάζει την επιφάνεια με βοήθεια χρωμάτων ή αποχρώσεις του γκριζου. Και οι τρεις εντολές έχουν τα ίδια ορίσματα: το διάνυσμα των συντεταγμένων του  $x$ , το διάνυσμα των συντεταγμένων του  $y$ , και ένα πίνακα με τις τιμές του  $z$  με διαστάσεις το μήκος του  $x$  και το μήκος του  $y$ . Τέτοια παραδείγματα μπορείτε να δείτε με την εντολή `demo("persp")`.