

13/12/2022

Εθνικό Μετσόβιο Πολυτεχνείο



Αναλυτικές Μέθοδοι στη Γεωπληροφορική Δημιουργία γραφικών στο R

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Σύνδεση με τα προηγούμενα

Κυριότερες εργασίες / στάδια για την εξερεύνηση δεδομένων

- Διάθεση και προσθήκη συνόλων δεδομένων, μετασχηματισμοί και εμπλουτισμός με βοηθητικά δεδομένα
- Μορφοποίηση και εισαγωγή συνόλων δεδομένων, εξερεύνηση και ανάλυση
- Απεικόνιση δεδομένων και αποτελεσμάτων ανάλυσης με διαφορετικούς τρόπους (για επικοινωνία, δημοσίευση, ...)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

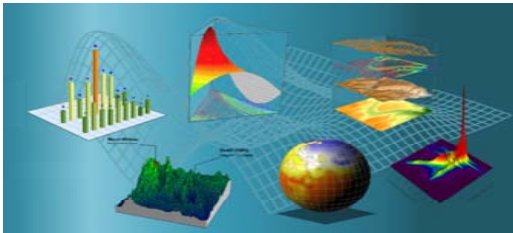
Σημερινά θέματα μαθήματος

- Γιατί να μας απασχολεί η δημιουργία γραφημάτων;
- Δυνατότητες γραφικών με το R
 - Βασικά γραφήματα (*Base R graphics*)
 - Γρήγορη και ευέλικτη διερεύνηση δεδομένων με οπτικοποίηση
 - Προβολή πολυμεταβλητών σχέσεων → πλέγμα γραφημάτων (*Lattice graphics*)
 - Λογική πίσω από τη γραμματική της γραφικής αντίληψης (*ggplot2 graphics*)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

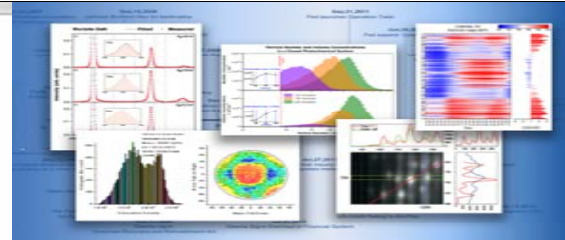
Γραφική απεικόνιση δεδομένων



- Γραφήματα και χάρτες αποτελούν το πιο θεμελιώδες συστατικό της σύγχρονης έρευνας, για την αναπαράσταση και αναφορά των δεδομένων και την καλύτερη κατανόηση των αποτελεσμάτων τους.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Πολλά γραφικά μέσα, όπως ιστογράμματα, γραφήματα διασποράς, εικονογράμματα, απόψεων επιφάνειας κ.ά.
- Είναι αποτελεσματικά για την εμφάνιση και τη σύνοψη μεγάλων ποσοτήτων αριθμητικών δεδομένων και είναι χρήσιμα για την εμφάνιση τάσεων, προτύπων και σχέσεων μεταξύ μεταβλητών

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Κύριες ανάγκες για τη δημιουργία γραφημάτων

- Διερευνητική ερμηνεία/αντίληψη – Σε πολλούς τομείς εφαρμογής, για τη λήψη αποφάσεων, ανάλυση ή αντιμετώπιση καταστάσεων έκτακτης ανάγκης είναι βασικό να εξαχθούν γρήγορα οι δυνητικά χρήσιμες πληροφορίες από τα διαθέσιμα δεδομένα.
- Παρουσίαση – Για την κοινοποίηση των αποτελεσμάτων της επιστημονικής έρευνας σε ευρύτερο κοινό χρειάζεται ευελιξία για την ικανοποίηση των απαιτήσεων διάδοσης των πληροφοριών

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Ένας από τους κύριους λόγους για το πρακτικό ενδιαφέρον που προσελκύει το R (ή/και τη γλώσσα r) για την ανάλυση δεδομένων είναι για το περιβάλλον και τις ισχυρές δυνατότητες δημιουργίας γραφικών τις οποίες παρέχει, χρησιμοποιώντας ...
 - το παραδοσιακό σύστημα γραφικών, καθώς και
 - μια εκτεταμένη σειρά πρόσθετων πακέτων και το *ggplot2*

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

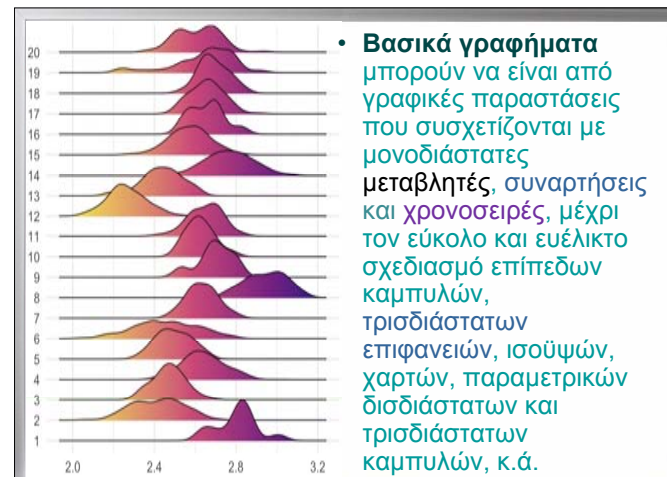
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Η δημιουργία και αποθήκευση γραφημάτων στο R περιλαμβάνει διάφορους τρόπους οπτικής απεικόνισης δεδομένων που μπορούν να καθοδηγήσουν στην διαδικασία μοντελοποίησης και αξιολόγησης μιας ανάλυσης στατιστικών ή άλλων στοιχείων (π.χ. χώρο-χρονικά δεδομένα, εικόνες, χάρτες, ...)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

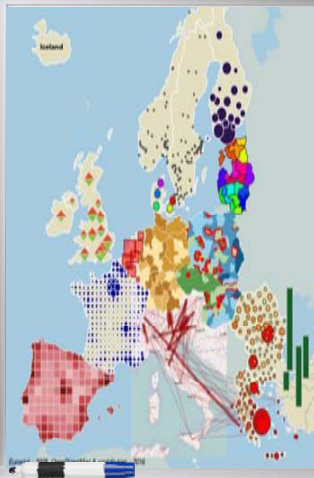
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Βασικά γραφήματα μπορούν να είναι από γραφικές παραστάσεις που συσχετίζονται με μονοδιάστατες μεταβλητές, συναρτήσεις και χρονοσειρές, μέχρι τον εύκολο και ευέλικτο σχεδιασμό επίπεδων καμπυλών, τρισδιάστατων επιφανειών, ισουψών, χαρτών, παραμετρικών δισδιάστατων και τρισδιάστατων καμπυλών, κ.ά.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

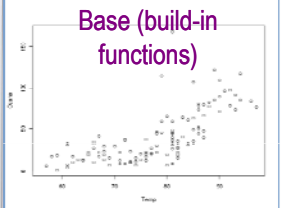
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



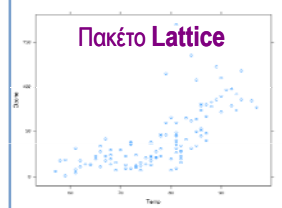
- Προηγμένες δυνατότητες μπορούν να είναι από απεικονίσεις χωρικών δεδομένων και αναλύσεις (π.χ. πρόβλεψη), χαρτογραφικές παραστάσεις (όπως αναλογικά σύμβολα) και χρήσιμα χαρακτηριστικά (όπως χαρτογραφικές παλέτες), μέχρι την πρόσβαση σε χαρτογραφικό API για να διευκολύνεται η γραφική παρουσίαση χαρτών.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

Base (build-in functions)




Πακέτο Lattice



- το R κατανοεί 3 κύριες "διαλέκτους" για τη δημιουργία γραφικών

Πακέτο ggplot2



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

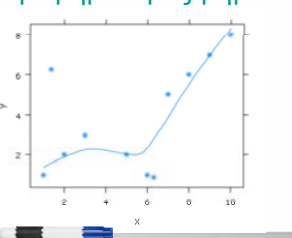
Βασικές επιλογές γραφικών στο R

- Η τυποποιημένη σουίτα πακέτων γραφικών → *παραδοσιακά γραφήματα* (High and low level plot functions, Annotations, Multipanel plots, Custom-made functions, ...)
- Τα *γραφήματα πλέγματος* (lattice graphics) για πολυπαραγοντική απεικόνιση δεδομένων → καλύτερες προεπιλογές και ικανότητα να εμφανίζονται εύκολα πολυμεταβλητές σχέσεις
- Το *πακέτο ggplot2*, η *Γραμματική των Γραφικών Παραστάσεων* → κανόνες για την ευκολότερη επεξεργασία και ανάλυση δεδομένων για την απεικόνιση των στοιχείων τους με ενημερωτικό και ελκυστικό τρόπο

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

Γραφικά χαμηλού επιπέδου δυσκολίας → υψηλής χρηστικότητας

- Τέτοια γραφήματα παράγονται στο R μέσα από βασικές δομικές μονάδες (συναρτήσεις γραφικών εντολών) οι οποίες μπορούν να χρησιμοποιηθούν για τη δημιουργία γραφικών με βήμα-προς-βήμα διαδραστικές διαδικασίες




Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

Απλά γραφήματα στο R

```
# Creating a Graph
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")
```

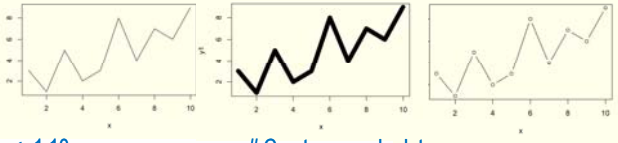
- Στο R, τα περισσότερα γραφικά δημιουργούνται συνήθως αλληλεπιδραστικά



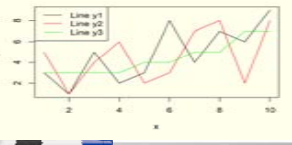
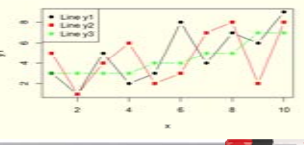

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

```
# Create Plots/Graphs with lines, points, legends, colors, ...
x <- 1:10 # Create example data
y1 <- c(3, 1, 5, 2, 3, 8, 4, 7, 6, 9) # Basic line plot in R
plot(x, y1, type = "l") # Change main title & axis labels
main = "This is my Line Plot",
xlab = "My X-Values",
ylab = "My Y-Values")
plot(x, y1, type = "l", lwd = 10) # Change thickness of line
plot(x, y1, type = "b") # Add symbols to points
y2 <- c(5, 1, 4, 6, 2, 3, 7, 8, 2, 8) # Create more example data
y3 <- c(3, 3, 3, 3, 4, 4, 5, 5, 7, 7)
plot(x, y1, type = "l") # Draw first line
lines(x, y2, type = "l", col = "red") # Add second line
lines(x, y3, type = "l", col = "green") # Add third line
legend("topleft", legend = c("Line y1", "Line y2", "Line y3"), col = c("black", "red", "green"), lty = 1)
```

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

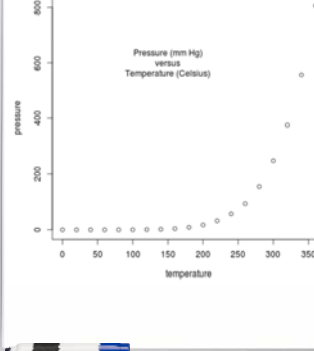


```
x <- 1:10 # Create example data
y1 <- c(3, 1, 5, 2, 3, 8, 4, 7, 6, 9) # Basic line plot in R
plot(x, y1, type = "l") # Change color of line
plot(x, y1, type = "l", col = "blue") # Change line thickness
plot(x, y1, type = "l", lwd = 10)
```

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

```
plot(pressure)
text(150, 600, "Pressure (mm Hg)\nversus\nTemperature (Celsius)")
```



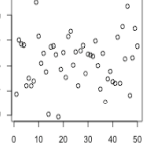
- Η βασική συνάρτηση για απλά γραφήματα είναι η συνάρτηση **plot()**, η οποία έχει πολλές δυνατότητες και μπορεί να πάρει διάφορες γραφικές παραμέτρους για ορισμάτα
- Η συνάρτηση **text()** προσθέτει τον τίτλο στην επιθυμητή θέση (150, 600) εντός του γραφήματος.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

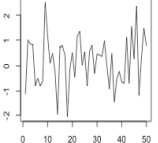
Titles, legends and annotations in plot()

```
x1 <- rnorm(50)
png("plottype.png")
par(mfrow = c(2,2))
plot(x1, type = "p", main = "points", ylab = "", xlab = "")
plot(x1, type = "l", main = "lines", ylab = "", xlab = "")
plot(x1, type = "b", main = "both", ylab = "", xlab = "")
plot(x1, type = "o", main = "both overplot", ylab = "", xlab = "")
dev.off()
```

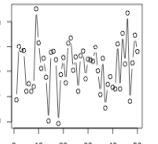
points



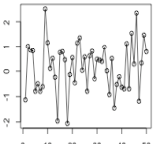
lines



both



both overplot



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΓΡΑΦΟΡΙΚΗ'

Προσαρμογές στα γραφήματα

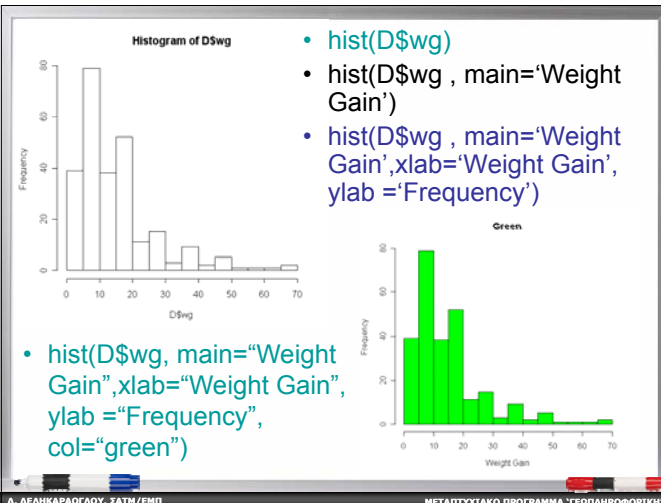
- Επιτυγχάνονται με γραφικές παραμέτρους που περιγράφουν τον τρόπο αλλαγής των συμβόλων, των γραμματοσειρών, των χρωμάτων και των γραμμών ενός γραφήματος.
- Επιλογές για τους άξονες και κείμενα επιτρέπουν την προσαρμογή των αξόνων ενός γραφήματος, την προσθήκη γραμμών αναφοράς, κειμένου σχολιασμών και λεζάντες.
- Ο συνδυασμός γραφημάτων επιτρέπει την οργάνωση πολλαπλών γραφικών παραστάσεων σε ένα μόνο γράφημα.

Μερικές σημαντικές συναρτήσεις για απλά ή πιο σύνθετα γραφήματα

```
plot: make a scatterplot, or other type of plot
      depending on the class of the object being plotted
lines: add lines to a plot, given a vector x values
      and a corresponding vector of y values (or a 2-column matrix);
      this function just connects the dots
points: add points to a plot
text: add text labels to a plot using specified x, y coordinates
title: add annotations to x, y axis labels, title, subtitle,
      outer margin
mtext: add arbitrary text to the margins (inner or outer)
      of the plot
axis: adding axis ticks/labels
```

Παράμετροι γραφικών συναρτήσεων

- Οι συναρτήσεις υψηλών και χαμηλών επιπέδων γραφικών λαμβάνουν παραμέτρους που επιτρέπουν να αλλαχθεί η εμφάνιση των γραφημάτων.
 - π.χ., για να τροποποιηθούν οι ετικέτες αξόνων, τίτλοι αντιστοίχισης (κύρια, δευτερεύουσα), η κλίμακα των αξόνων, κ.λπ.
- Στις συναρτήσεις υψηλού επιπέδου γραφικών, είναι δυνατόν να οριστούν ορισμένες γενικές παράμετροι, μέσω της συνάρτησης **par()**, και οι επιλογές να ισχύουν μόνιμα (σε μια συνεδρία του R) ή να αποκατασταθούν πρότερα επιλεγμένες τιμές.

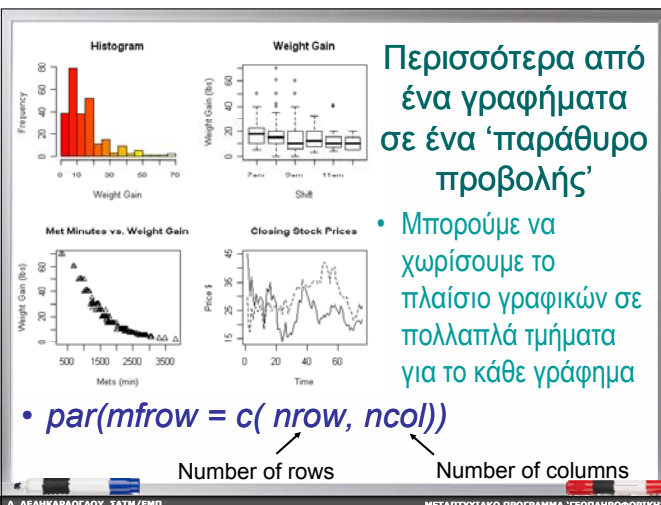
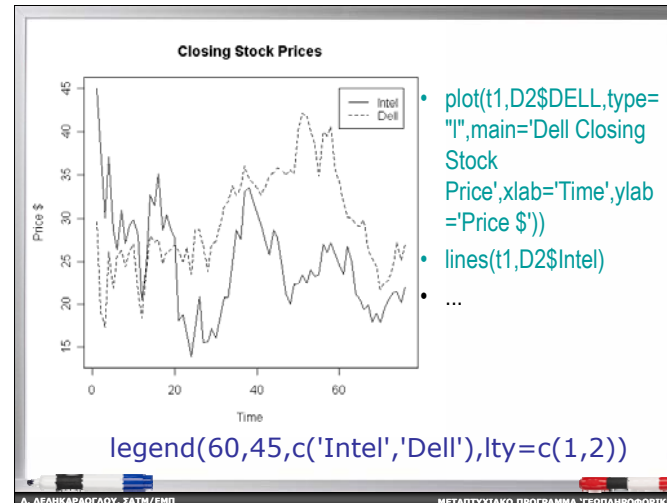
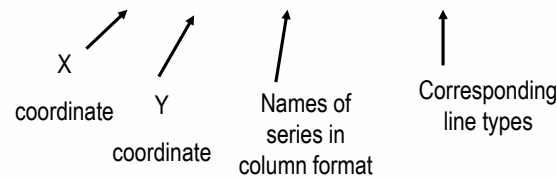


Σε απλά γραφήματα ...

- Η τοποθέτηση μια λεζάντας είναι σχετικά εύκολη.

Syntax

- `legend(x, y, names, line types)`



Set a graphical parameter using par()

```
par() # view current settings
opar <- par() # make a copy of current settings
par(col.lab="red") # red x and y labels
hist(mtcars$mpg) # create a plot with these new settings
par(opar) # restore original settings
```

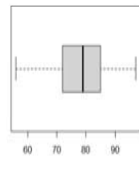
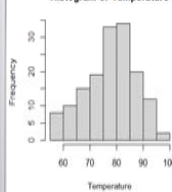
- Ένας τρόπος για να καθοριστούν οι επιλογές των παραμέτρων ενός γραφήματος είναι μέσω της συνάρτησης `par(optionname = value , optionname = value , ...)`.
 - Οι τιμές των παραμέτρων, θα ισχύουν για το υπόλοιπο της συνεδρίας ή μέχρι να αλλαχθούν ξανά.

- Η συνάρτηση `par()` χρησιμοποιείται για να καθορίσει τις συνολικές (global) παραμέτρους που επηρεάζουν όλα τα γραφικά σε μια συνεδρία του R.
- Αυτές οι παράμετροι μπορεί συχνά να αναδιαμορφώνονται με την είσοδο τους σε συγκεκριμένες συναρτήσεις γραφικών → αλλαγή στο πεδίο εφαρμογής των παραμέτρων των γραφικών (*scope*)
 - οι τιμές παραμέτρων που ορίζονται σε συναρτήσεις υψηλού επιπέδου μπορούν να έχουν διαφορετική επίδραση από εκείνες που ορίζονται στην `par()` και διαφορετική από άλλες συναρτήσεις υψηλού επιπέδου, π.χ. `par(col="red")` και `plot(x, col="red")`

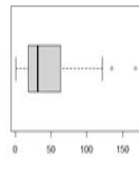
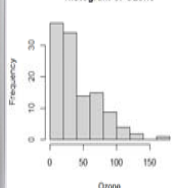
Typical parameters in par()

pch: the plotting symbol (default is open circle)
lty: the line type (default is solid line),
can be dashed, dotted, ...
lwd: the line width, specified as an integer multiple
col: the plotting color, specified as a number, string, or hex
code; the colors function gives you a vector
of colors by name
las: the orientation of the axis labels on the plot
bg: the background color
mar: the margin size
oma: the outer margin size (default is 0 for all sides)
mfrow: number of plots per row, column
(plots are filled row-wise)
mfcpl: number of plots per row, column
(plots are filled column-wise)

Histogram of Temperature

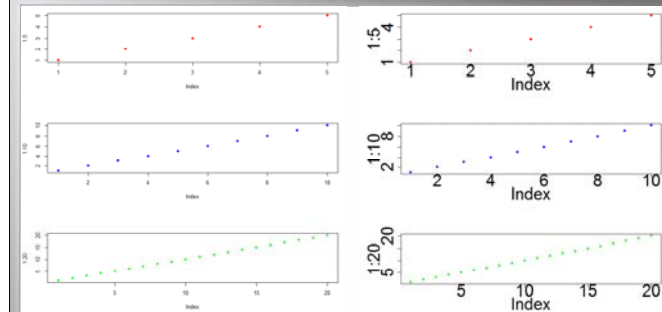


Histogram of Ozone



πολλαπλά γραφήματα σε μια γραφική παράσταση
ορίζοντας την παράμετρο mfrow της par()

```
Temperature <- airquality$Temp  
Ozone <- airquality$Ozone  
par(mfrow=c(2,2))  
hist(Temperature)  
boxplot(Temperature, horizontal=TRUE)  
hist(Ozone)  
boxplot(Ozone, horizontal=TRUE)
```



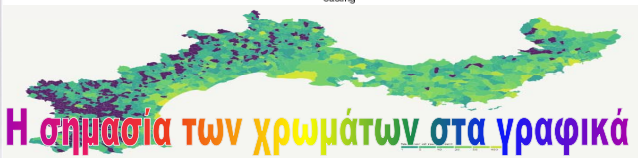
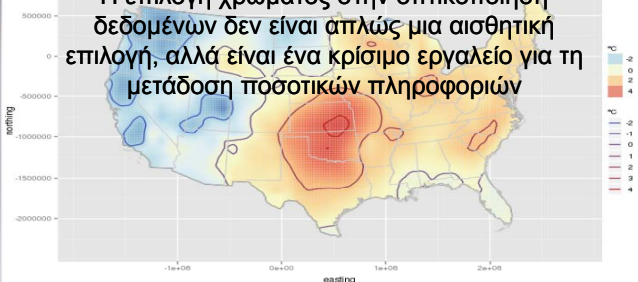
#define plot area as three rows and one column

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

#define plot area with large axis labels

```
par(mfrow = c(3, 1),  
mar = c(5, 10, 4, 1), cex.axis = 3, cex.lab = 3)
```

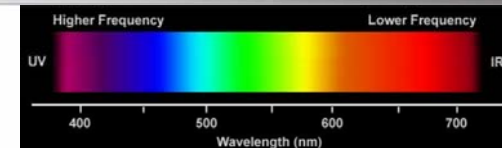
Η επιλογή χρώματος στην οπτικοποίηση δεδομένων δεν είναι απλώς μια αισθητική επιλογή, αλλά είναι ένα κρίσιμο εργαλείο για τη μετάδοση ποσοτικών πληροφοριών



Η σημασία των χρωμάτων στα γραφικά

Η σημασία των χρωμάτων σε γραφήματα

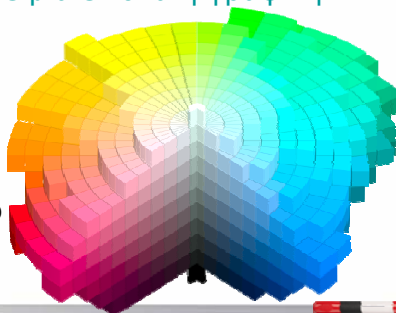
- Παρέχουν ένα μέτρο του ρεαλισμού (π.χ. στους χάρτες) → αντίληψη και αισθητική
- Η χρήση χρωμάτων είναι ένα πολυσύνθετο θέμα, που αφορά τη φυσική, τη φυσιολογία, την ψυχολογία, την τέχνη, και τη γραφιστική
- Παράγοντες της οπτικής αίσθησης χρώματος – διάκριση μεταξύ χρωμάτων, καθαρότητα και αντιληπτή ένταση του χρώματος, ποσοτικοποίηση της αντίληψης



- Η απόχρωση (hue) συνήθως σχετίζεται με ονόματα χρωμάτων, π.χ. κόκκινο, πράσινο, μπλε, κίτρινο → Συνδέεται με το κυρίαρχο μήκος κύματος: από μεγάλο μήκος κύματος (κόκκινο) σε κύμα μικρού μήκους (μπλε) και τη μετάβαση από κόκκινο σε πορτοκαλί έως κίτρινο, και έως πράσινο έως μπλε.
 - Μωβ (purple) αποχρώσεις προκύπτουν από την ανάμιξη των αντίθετων άκρων του φάσματος μαζί.

- Η ελαφρότητα (Lightness) είναι ένα σχετικό μέτρο που περιγράφει πόσο φως φαίνεται να αντανακλάται από ένα αντικείμενο σε σύγκριση με αυτό που μοιάζει λευκό σε μια εικόνα ή γραφική παράσταση.

- Αλλάζει την αντίληψή μας, όταν το ίδιο χρώμα τοποθετείται σε διαφορετικό χρώμα φόντου.



Ο κορεσμός (saturation) μετρά τη ζωντάνια ενός χρώματος

- Οι οθόνες των Η/Υ χρησιμοποιούν ένα μειωμένο σύνολο πρωτευόντων συνδυασμών των χρωμάτων του φωτός για την ανάμιξη/διαβάθμιση όλων των αποχρώσεων: **κόκκινο, πράσινο και μπλε (RGB)**.



- Η διαφάνεια (transparency) επιτρέπει τη μορφοποίηση του φόντου χρωμάτων – από το συμπαγές, μέχρι το ημιδιαφανές χρώμα ή άλλες διαβαθμίσεις τους → ένας καλός τρόπος για να δημιουργηθεί οπτικό ενδιαφέρον, π.χ. σε χαρτογραφικές οπτικοποιήσεις και απεικονίσεις (ποσοτικοποίηση της αντίληψης)

Στο R, το χρώμα μπορεί να καθοριστεί με διαφορετικούς τρόπους:

- RGB κωδικοί
- | | | |
|------------|---------|-------------|
| Blue | #0000FF | 0,0,255 |
| BlueViolet | #6A2BE2 | 138,43,226 |
| Brown | #A52A2A | 165,42,42 |
| BurlyWood | #DEB887 | 222,184,135 |

- Ένα όνομα, π.χ. "red", "green", ...
- Συναρτήσεις δημιουργίας χρωμάτων, π.χ. *gray*, *rgb*, *CMYK*, ...
- Χρωματικά μοντέλα (παλέτες) που παράγουν ένα συνεκτικό σύνολο χρωμάτων: *heat.colors*, *rainbow*, *colorRamp*, *terrain.colors*, *topo*, ...
- Τα πακέτα *RColorBrewer* και *colorspaces*.

R color cheatsheet

Finding a good color scheme for presenting data can be challenging. This color cheatsheet will help! **R uses hexadecimal to represent colors.** Hexadecimal is a base-16 number system used to describe color. Red, green, and blue are each represented by two characters (#RRGGBB). Each character has 16 possible symbols: 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F.

"00" can be interpreted as 0.0 and "FF" as 1.0 i.e. red = #FF0000, black = #000000, white = #FFFFFF

Two additional characters (with the same scale) can be added to the end to describe transparency (#RRGGBBAA)

R has 657 built in color names Example: `peachpuff4`

To see a list of names: `colors()`

These colors are displayed on P. 3.

R translates various color models to hex, e.g.:

- RGB (red, green, blue): The default intensity scale in R ranges from 0-1; but another commonly used scale is 0-255. This is obtained in R using `maxColorValue=255`.
- alpha is an optional argument for transparency, with the same intensity scale.

- `rgb(r, g, b, maxColorValue=255, alpha=255)`
- HSV (hue, saturation, value): values range from 0-1, with optional alpha argument `hsv(h, s, v, alpha)`

<https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf>

R Color Palettes

This is for all of you who don't know anything about color theory, and don't care but want some nice colors on your map or figure...NOW!

TIP: When it comes to selecting a color palette, **DO NOT** try to handpick individual colors! You will waste a lot of time and the result will probably not be all that great. R has some good packages for color palettes. Here are some of the options

Packages: grDevices and colorRamps

`grDevices` comes with the base installation and `colorRamps` must be installed. Each palette's function has an argument for the number of colors and transparency (`alpha`):

```
heat.colors(4, alpha=1)
> #FF0000FF "#FF8000FF" "#FFFFFF00FF" "#FFFFFF00FF"
```

For the `rainbow` palette you can also select start/end color (red = 0, yellow = 1/6, green = 2/6, cyan = 3/6, blue = 4/6 and magenta = 5/6) and saturation (s) and value (v):

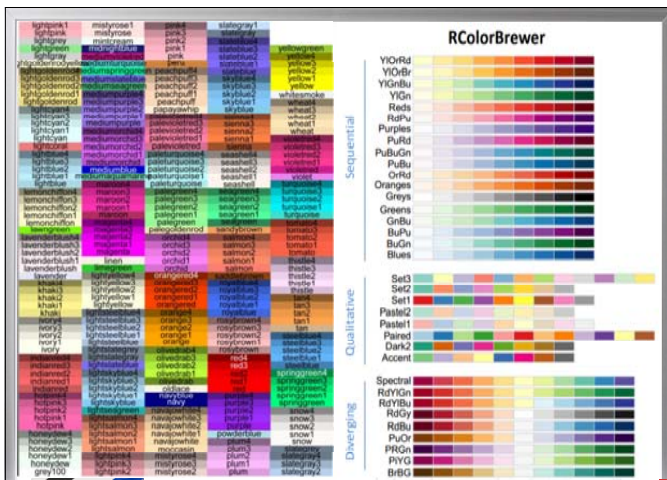
```
rainbow(n, s = 1, v = 1, start = 0, end = max(1, n - 1), alpha = 1)
```

Package: RColorBrewer
This function has an argument for the number of colors

Named Colours

<https://www.stat.auckland.ac.nz/~ihaka/787/R-colours-a4.pdf>

Brown	Dark Red	Firebrick	Gainsboro
Indian Red	Light Coral	Red	Rosy Brown
Snow	White Smoke	Misty Rose	Salmon
Tomato	Dark Salmon	Coral	Orange Red
Light Salmon	Sienna	Sea Shell	Chocolate
Sandy Brown	Sandy Brown	Peach Puff	Peru
Linen	Bisque	Dark Orange	Burlywood
Antique White	Tan	Navajo White	Bianched Almond
Papaya Whip	Moccasin	Orange	Wheat
Old Lace	Floral White	Dark Goldenrod	Goldenrod
Corn Silk	Light Goldenrod	Gold	Lemon Chiffon
Khaki	Pale Goldenrod	Dark Khaki	Beige
Ivory	Light Goldenrod Yellow	Light Yellow	Yellow
Olive Drab	Yellow Green	Dark Olive Green	Green Yellow
Light Green	Dark Green	Dark Sea Green	Dark Sea Green

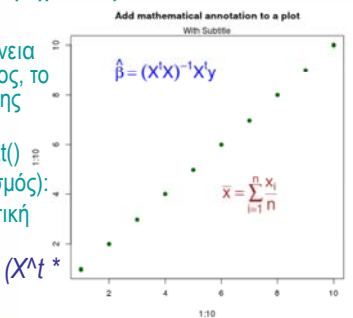


Ετικέτες, τίτλοι, ... στα γραφικά

- Με χρήση της συνάρτησης `text()` μπορεί να τοποθετηθεί κείμενο οπουδήποτε στην περιοχή σχεδίασης ενός γραφήματος

Κύριες επιλογές:

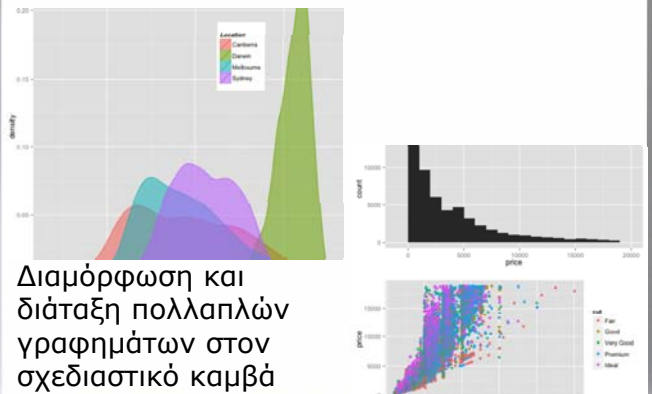
- Control (έλεγχος): η οικογένεια γραμματοσειρών, το μέγεθος, το στυλ και το χρώμα, μέσω της `par()`
- Margins (περιθώρια): `mtext()`
- Orientation (Προσανατολισμός):
- Math.expression (Μαθηματική έκφραση)
 $expression(\hat{\beta} = (X^t X)^{-1} X^t y)$



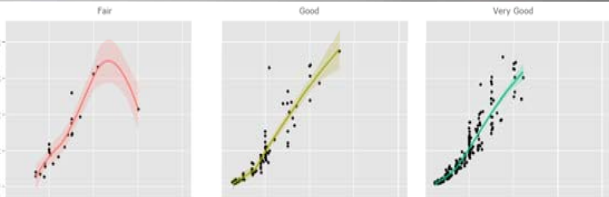
Κύρια βήματα για τη δημιουργία γραφικών απεικονίσεων

- `plot.new()` - Ορισμός της περιοχής (καμβά) του γραφήματος
- `plot.window()` - Ρύθμιση του συστήματος συντεταγμένων (`xlim`, `ylim`)
- `box()` - Σχεδιασμός ορθογώνιου πλαισίου γύρω από το γράφημα
- `axis` - Σχεδίαση αξόνων, σημείων, τίτλων, ετικετών και άλλων κειμένων, ...
- `usr` - παρέχει / καθορίζει το σύστημα συντεταγμένων

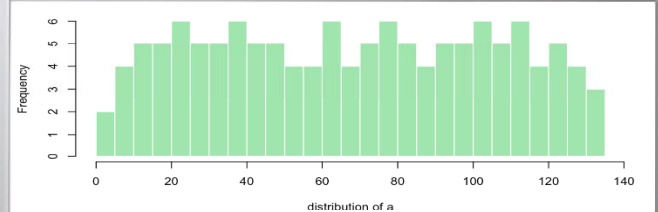
Πολλαπλά γραφήματα



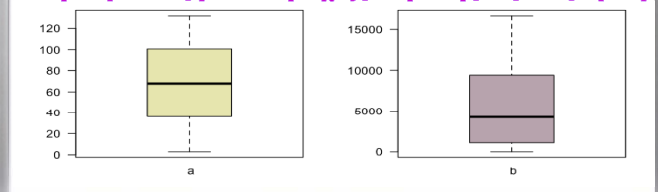
Διαμόρφωση και διάταξη πολλαπλών γραφημάτων στον σχεδιαστικό καμβά

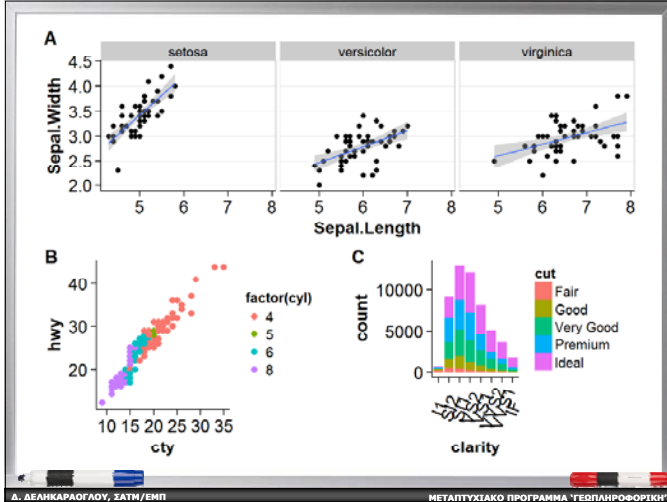


- Η παράμετρος `mfrow` "κόβει" την περιοχή σε πλέγμα κανάβου
- Η συνάρτηση `layout()` παρέχει μια διάταξη πλέγματος με πιθανά άνισα ύψη και πλάτη και μπορεί να τοποθετήσει ένα γράφημα σε μια ορθογώνια περιοχή που καλύπτει περισσότερες από μια ενότητες του πλέγματος



Διάρθρωση του καμβά σε περιοχές με προσαρμοσμένα μεγέθη





```
# Create data
my_variable=rnorm(1000, 0, 2),
rnorm(1000, 9, 2))

# Layout to split the screen
layout(mat = matrix(c(1,2), 2,1, byrow=TRUE),
height = c(1,8))

# Draw the boxplot and the histogram
par(mar=c(0, 3.1, 1.1, 2.1))
boxplot(my_variable, horizontal=TRUE,
ylim=c(-10,20), xaxt="n", col=rgb(0.0,0.0,0.5),
frame=F)
par(mar=c(4, 3.1, 1.1, 2.1))
hist(my_variable, breaks=40, col=rgb(0.2,0.0,0.5,0.5),
border=F, main="", xlab="value of the variable",
xlim=c(-10,20))
```

• Η διαρρύθμιση ενός γραφήματος μπορεί να τακτοποιηθεί έτσι ώστε να παρουσιάζει περισσότερες από μια γραφικές παραστάσεις

```
#Creating data
Ixos=rnorm(4000,
Primadur=Ixos+rnorm(4000, 0, 30))

#Divide the screen in 1 line and 2 columns
par(mfrow=c(1,2), oma = c(0, 0, 2, 0))

#Make the margin around each graph a bit smaller
par(mar=c(4, 2, 2, 2))

#Classical histogram and plot
hist(Ixos, main="", breaks=30, col=rgb(0.3,0.5,1,0.4),
xlab="height", ylab="nbr of plants")
plot(Ixos, Primadur, main="", pch=20, cex=0.4,
col=rgb(0.3,0.5,1,0.4), xlab="primadur", ylab="Ixos")

#And I add only ONE title:
mtext("Primadur : Distribution and correlation with Ixos",
outer = TRUE, cex = 1.5, font=4, col=rgb(0.1,0.3,0.5,0.5))
```

```
# Create data
Iris=rnorm(4000, 100, 30)
Dafodil=rnorm(4000, 200, 30)

# Represent separately first
par(mfrow=c(1,2))
hist(Iris, breaks=30, xlim=c(0,300), col=rgb(1,0,0,0.5),
xlab="height", ylab="nbr of plants", main="")
hist(Dafodil, breaks=30, xlim=c(0,300), col=rgb(0,0,1,0.5),
xlab="height", ylab="nbr of plants", main="")

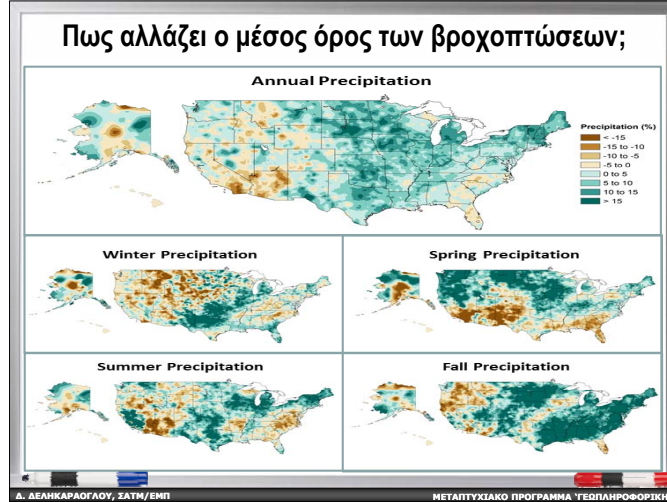
# It would be more interesting to have both distribution on the same graph,
# with transparency permitting to see the whole distribution:
hist(Iris, breaks=30, xlim=c(0,300), col=rgb(1,0,0,0.5),
xlab="height", ylab="nbr of plants",
main="distribution of height of 2 plant varieties")
hist(Dafodil, breaks=30, xlim=c(0,300), col=rgb(0,0,1,0.5), add=T)
legend("topright", legend=c("Iris", "Dafodil"),
col=c(rgb(1,0,0,0.5), rgb(0,0,1,0.5)), pt.cex=2, pch=15)
```

Διατεταγμένα γραφήματα (panels side-by-side)

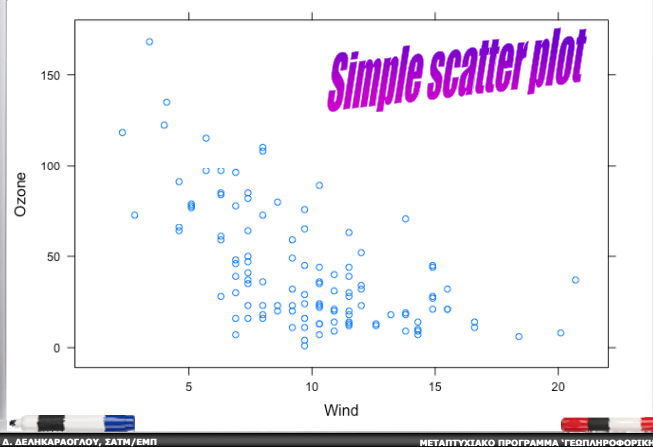
Lattice graphics / Γραφικά πλέγματος

- Ο όρος υποδηλώνει ένα σύνολο πολλαπλών γραφημάτων με κάποια τακτική δομή ή διάταξη σε έναν ορισμένο ευκλειδείο χώρο R^n (συντά το επίπεδο ή ο 3-D χώρος)
- Γραφήματα που εμφανίζουν μια μεταβλητή ή τη σχέση μεταξύ μεταβλητών, που εξαρτώνται από μια ή περισσότερες άλλες μεταβλητές → πολυπαραγοντική απεικόνιση δεδομένων

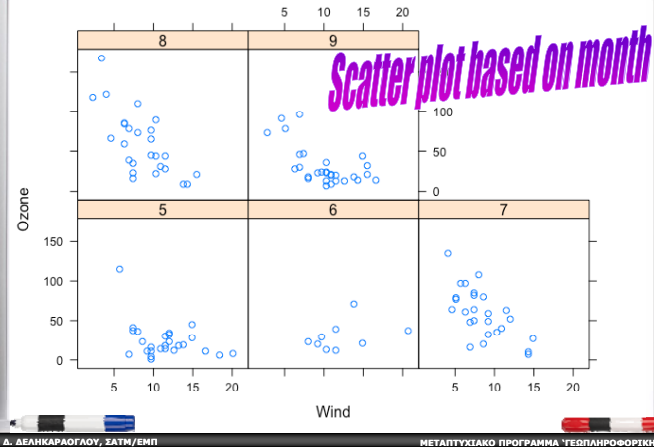
- Ένα πλέγμα ή προβολή γραφημάτων είναι ένα εναλλακτικό σύστημα γραφικών που προστέθηκε αργότερα στο R.
- Η μεγάλη διαφορά μεταξύ του πλέγματος και του αρχικού συστήματος γραφικών βάσης του R είναι ότι το πλέγμα επιτρέπει τη δημιουργία πολλών περιοχών, που ονομάζονται παράθυρα προβολής, σε μία σελίδα γραφικών.
 - για σύγκριση και αντιπαραβολή, ή/και
 - για την παρακολούθηση τάσεων και τον εντοπισμό μοτίβων σε ένα σύνολο δεδομένων



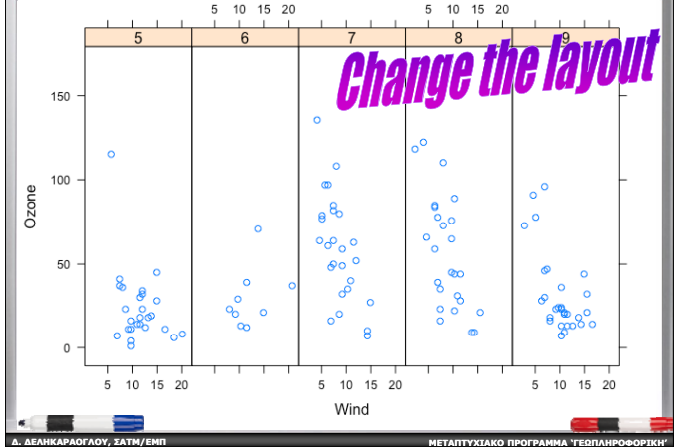
xyplot(Ozone~Wind, data=airquality)



xyplot(Ozone~Wind, data=airquality)



xyplot(Ozone~Wind, data=airquality, layout=c(5,1))



Lattice graphics

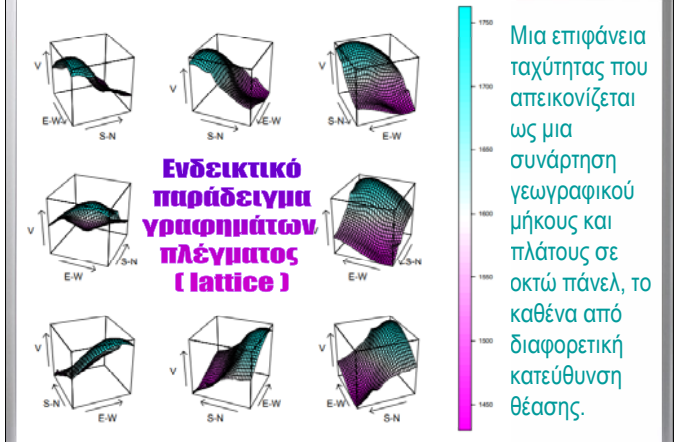
```
# Install
install.packages("lattice")
# Load
library("lattice")
```

- Το πακέτο πλέγματος γραφημάτων (***lattice package***) παρέχει αναγκαίες συναρτήσεις για τη δημιουργία όλων των τυποποιημένων γραφικών πλέγματος και επιπλέον:
 - Προεπιλογές για γραφήματα υψηλής οπτικής αισθητικής
 - Δυνατότητες δημιουργίας και τροποποίησης θεματικών γραφικών, καθώς και συλλογές απλούστερων γραφημάτων από υποσύνολα δεδομένων

- Γραφήματα πλέγματος εμφανίστηκαν αρχικά με την S και Splus (~1980), με την ονομασία **γραφικά TRELLIS**.

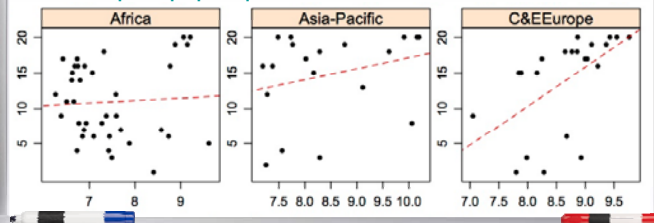


- Ο όρος χρησιμοποιήθηκε για να υποδηλώσει ένα πλαίσιο συνολικής οπτικοποίησης δεδομένων που θυμίζει **πέργκολα κήπου**, στο οποίο τα γραφήματα είναι τοποθετημένα σε σειρές, στήλες και σελίδες → **Μια εξαιρετικά χρήσιμη προσέγγιση για την εμφάνιση συμπαγών σχέσεων σε μεγάλα σύνολα δεδομένων**



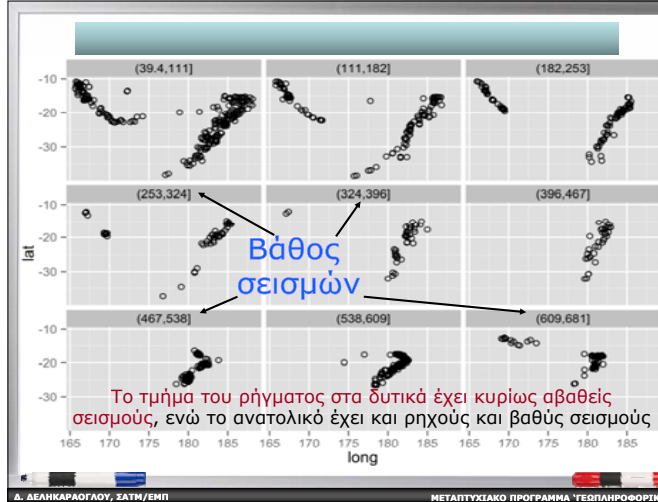
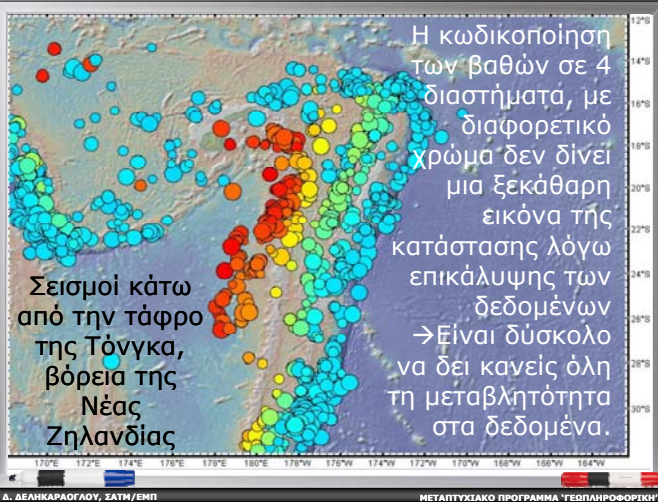
- Το κεντρικό πλεονέκτημα της χρήσης πλέγματος γραφημάτων είναι ότι επιτρέπουν
 - Την οπτική διάκριση μεταξύ διαφορετικών ομάδων / σύνθετων πολυπαραγοντικών δεδομένων (*με πολλαπλές μεταβλητές*), π.χ.,
 - για να εξεταστούν πτυχές της δομής των δεδομένων μέσα σε κάθε πάνελ γραφημάτων,
 - για να εξακριβωθεί πώς αλλάζει η σχέση μεταξύ δύο μεταβλητών με παραλλαγές της μιας
 - All plots can be described by **plot formulas** → **graphics functions**
 - High-level functions handle **panel layout** → **panel functions**

- Σύνθετα δεδομένα πολλών μεταβλητών μπορούν συχνά να απεικονιστούν καλύτερα μέσω **σύγκρισης στην ίδια κλίμακα (conditioning)** & **ομαδοποίησης (grouping)**
- Δείχνουν πως κάποια σχέση στα δεδομένα αλλάζει σε σχέση με άλλες μεταβλητές
- Χωριστά πάνελ γραφημάτων, διατεταγμένα για οπτική σύγκριση



Σύγκριση σε κοινές κλίμακες (Conditioning)

- Κάθε πάνελ περιέχει μια γραφική παράσταση της οποίας τα δεδομένα είναι "υπό όρους" βάσει τιμών που προέρχονται από την μεταβλητή στην οποία αναφέρεται το συγκεκριμένο γράφημα
 - Στην περίπτωση μιας κατηγορικής μεταβλητής, αυτό σημαίνει απεικόνιση σε ίδιου τύπου γράφημα για υποσύνολα δεδομένων που αντιστοιχούν σε καθένα από τα επίπεδα αυτής της μεταβλητής.
 - Στην περίπτωση μιας αριθμητικής μεταβλητής, σημαίνει δημιουργία του ίδιου γραφήματος για υποσύνολα δεδομένων που αντιστοιχούν σε πολλαπλά διαστήματα αυτής της μεταβλητής.



- Μια ομάδα δεδομένων περιγράφεται ως "ζώνη" (*shingle*, στην ορολογία του R)
- Στα δεδομένα της τάφρου Tonga, υποθέτοντας ότι τα βάθη του σεισμού περιέχονται στη μεταβλητή 'βάθος' (*depth*), το *shingle* δημιουργείται με τις εντολές R
 - > `depth = earthquakes$depth`
 - > `Depth = equal.count(depth, number=8, overlap=.1)`
- Το *shingle* που έχει εκχωρηθεί στη μεταβλητή *Depth* περιλαμβάνει 8 διαστήματα με γειτονικά διαστήματα να έχουν το 10% των τιμών τους ως επικάλυψη

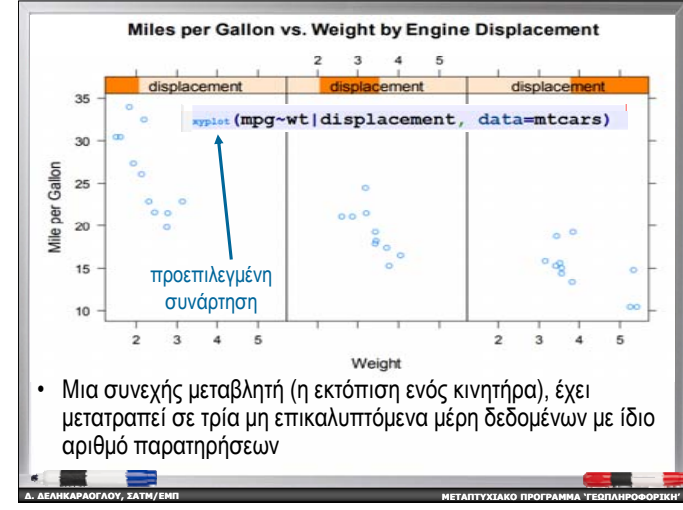
- Μια "ζώνη" (*shingle*) περιέχει τις αριθμητικές τιμές της ρυθμιστικής μεταβλητής από τις οποίες δημιουργήθηκε και μπορεί να αντιμετωπιστεί ως αντίγραφο αυτής της μεταβλητής, π.χ.
 - > `range(Depth)`
[1] 40 680
 - > `range(depth)`
[1] 40 680
- Αυτό περιέχει επίσης τις πληροφορίες που επισυνάπτονται σε αυτό (το εύρος βαθών) που εμφανίζονται με μια απλή εντολή εκτύπωσης
 - > `plot(Depth)`



- Οι ρυθμιστικές μεταβλητές (*conditioning variables*) ορίζουν τα "panel" (πίνακα/κάναβο) γραφικών σε μια συνολική γραφική παράσταση
 - Αυτά μπορούν να τοποθετηθούν σε μια "σελίδα" με διάφορους τρόπους
 - Οι συναρτήσεις στα πάνελ του πίνακα αντιστοιχούν σε ένα υποσύνολο δεδομένων και "δίνουν" τη γραφική παράστασή τους
 - `panel functions` get the data for a subset and "render" (plot) it
 - Συναρτήσεις υψηλού επιπέδου (*high-level functions*) διαχειρίζονται τη διάταξη των πάνελ του πίνακα και καλούν τις συναρτήσεις τους

- π.χ., η απεικόνιση των σεισμών στην τάφρο Tonga παράγεται από το συνάρτηση *xyplot*, η οποία είναι μια Lattice παραλλαγή της βασικής συνάρτησης του R για τη δημιουργία γραφημάτων διασποράς σημείων (*scatterplot*)
 - > `Depth = equal.count(quakes$depth, number = 8, overlap = .1)` # δημιουργία του *shingle*
 - > `xyplot(lat ~ long | Depth, data = quakes, xlab = "Longitude", ylab = "Latitude")`
 - # δημιουργία του πλέγματος γραφημάτων
- Η *xyplot* μπορεί επίσης να δημιουργήσει ένα "ενιαίο" γράφημα (*unconditional plot*)
 - > `xyplot(lat ~ long, data = quakes, xlab = "Longitude", ylab = "Latitude")`

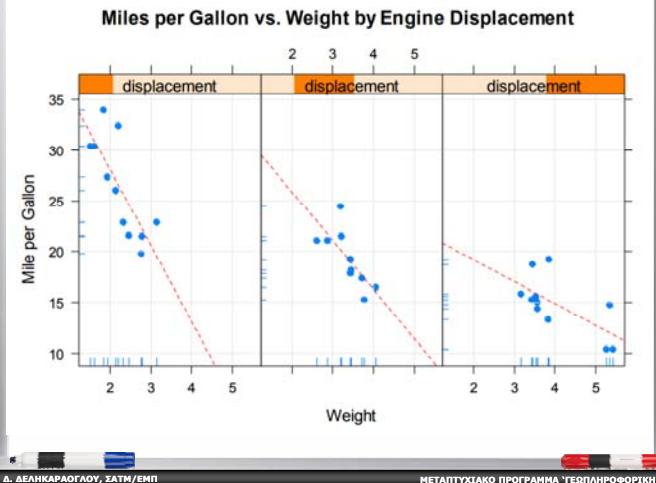
- Προσαρμογή (*Customize*) - Κάθε μία από τις συναρτήσεις γραφικής παράστασης υψηλού επιπέδου χρησιμοποιεί μια προεπιλεγμένη συνάρτηση για την σχεδίαση των γραφικών πλαισίων → της μορφής *panel.graph_function*
- Ένα ισχυρό χαρακτηριστικό των γραφικών πλέγματος είναι ότι επιτρέπεται στον χρήστη να αντικαταστήσει την προεπιλεγμένη συνάρτηση με μια προσαρμοσμένη συνάρτηση για τις εκάστοτε δικές του γραφικές ανάγκες → μεγαλύτερη ευελιξία στην υλοποίηση των γραφικών
 - Γραφικά "θέματα" γενικεύουν τη συνάρτηση *par()*
 - Combine multiple panel functions → write new ones.



- Μια συνεχής μεταβλητή (η εκτόπιση ενός κινητήρα), έχει μετατραπεί σε τρία μη επικαλυπτόμενα μέρη δεδομένων με ίδιο αριθμό παρατηρήσεων


```
library(lattice)
displacement <- equal.count(mtcars$disp, number=3, overlap=0)
# Using four separate building block functions
# (to plot regression lines, plots, and grid lines)
mypanel <- function(x, y) {
  panel.xyplot(x, y, pch=19)
  panel.rug(x, y)
  panel.grid(h=-1, v=-1)
  panel.lmline(x, y, col="red", lwd=1, lty=2)
}

# xyplot with custom panel function
xyplot(mpg~wt|displacement, data=mtcars,
  layout=c(3, 1),
  aspect=1.5,
  main = "Miles per Gallon vs. Weight by Engine Displacement",
  xlab = "Weight",
  ylab = "Miles per Gallon",
  panel = mypanel)
```



High-level functions in 'lattice' package

Function	Default Display
histogram()	Histogram
densityplot()	Kernel Density Plot
qqmath()	Theoretical Quantile Plot
qq()	Two-sample Quantile Plot
stripplot()	Stripchart (Comparative 1-D Scatter Plots)
bwplot()	Comparative Box-and-Whisker Plots
barchart()	Bar Plot
dotplot()	Cleveland Dot Plot
xyplot()	Scatter Plot
sploM()	Scatter-Plot Matrix
contourplot()	Contour Plot of Surfaces
levelplot()	False Color Level Plot of Surfaces
wireframe()	Three-dimensional Perspective Plot of Surfaces
cloud()	Three-dimensional Scatter Plot
parallel()	Parallel Coordinates Plot

- Ο πίνακας/ κάναβος των γραφημάτων καθορίζεται αναφορικά με
 - Τον τύπο απεικόνισης (ιστόγραμμα, διαγράμματα διασποράς, ραβδογράμματα, κ.λπ.)
 - Μεταβλητές με συγκεκριμένους ρόλους
- Τυπικοί ρόλοι για μεταβλητές
 - Πρωτεύουσες μεταβλητές (για την κύρια γραφική απεικόνιση)
 - Προσαρμοσμένες ή ρυθμιστικές μεταβλητές (για τη διαίρεση των δεδομένων σε υποομάδες και αντιπαραβολή (ρύθμιση πολλαπλών πλαισίων, multipanel conditioning))
 - Μεταβλητές ομαδοποίησης για τη διαίρεση των δεδομένων σε υποομάδες και υπέρθεση τους

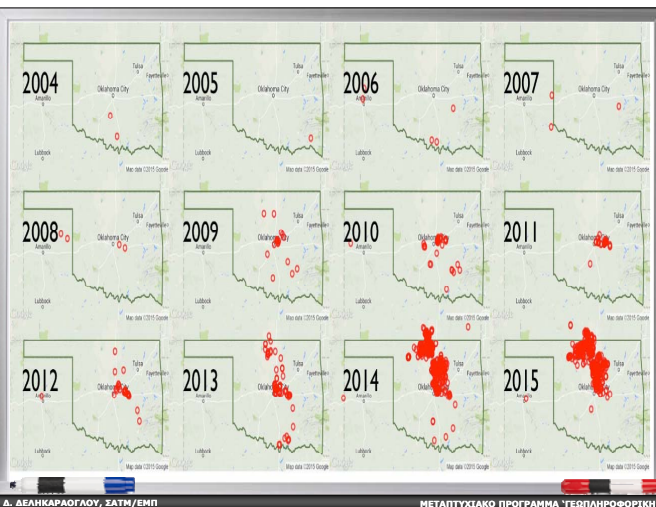
Βασικές συναρτήσεις πλέγματος γραφικών

- Βασικός τρόπος για κλήση συνάρτησης πλέγματος: `function.name(type,...)`
- Η γενική διάταξη ενός τύπου γραφήματος σε μια συνάρτηση πλέγματος γραφικών είναι:
 - `vertical.axis.variable ~ horizontal.axis.variable`
 - ο τελεστής `tilde` (δηλ. `~`) πρέπει να χρησιμοποιηθεί σε μια κλήση συνάρτησης πλέγματος, ακόμα και αν το γράφημα χρησιμοποιεί μια μόνο μεταβλητή.
- π.χ., `histogram(~data$x)` ή `xyplot(data$y ~ data$x)`
- Γραφικά πολλαπλών πλαισίων (panel) απαιτούν ελαφρώς πιο πολύπλοκες εκφράσεις, όπως π.χ. `histogram(~data$x | data$z)`

Βασικές συναρτήσεις πλέγματος γραφικών και αντικείμενα του R

- Με αυστηρούς όρους, μια κλήση σε μια συνάρτηση πλέγματος γραφικών (π.χ. `histogram(~x)`) δεν δημιουργεί το ιστόγραμμα, αλλά ένα αντικείμενο που περιέχει το γράφημα
- Με την κλήση της συνάρτησης για να δημιουργηθεί το γράφημα, παρέχεται σιωπηρά το όνομα του αντικειμένου
 - Στο R, η πληκτρολόγηση του ονόματος ενός αντικειμένου προκαλεί την εκτύπωση του περιεχομένου του αντικειμένου (δηλ. το γράφημα σε αυτήν την περίπτωση)

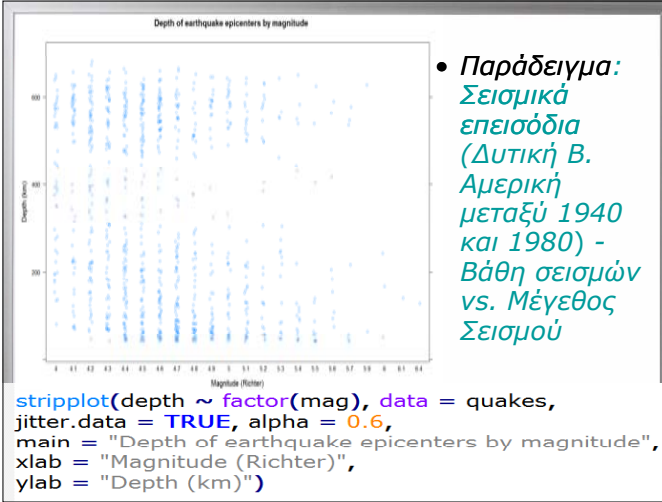
- Εναλλακτικά, το γράφημα μπορεί να αντιστοιχιστεί / εκχωρηθεί σε ένα αντικείμενο: `my.graph <- histogram(~x)`
 - Το αντικείμενο που περιέχει το γράφημα μπορούμε να το χειριστούμε όπως κάθε άλλο αντικείμενο R
 - Για παράδειγμα, πληκτρολογώντας το όνομα του αντικειμένου του γραφήματος στη γραμμή εντολών θα ανοίξει ένα παράθυρο γραφικών και θα εκτυπωθεί το ιστόγραμμα που δημιουργήθηκε.
- Τα αντικείμενα γραφικών πλέγματος είναι ιδιαίτερα χρήσιμα για την τοποθέτηση πολλαπλών γραφημάτων σε μια μόνο σελίδα



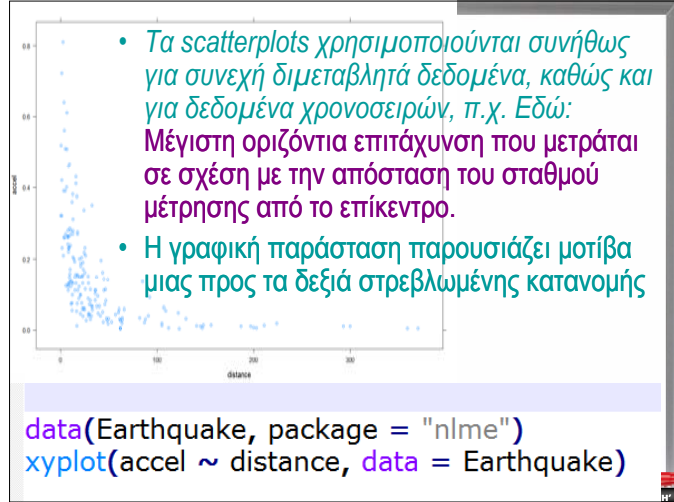
Συναρτήσεις γραφικών πλέγματος (Panel functions)

- Η δημιουργία ενός γραφήματος πλέγματος είναι μια διαδικασία δύο σταδίων (αν και η διάκριση μεταξύ των δύο βημάτων είναι συχνά άορατες για τον χρήστη)
 - Η κλήση της συνάρτησης (π.χ. η έκδοση της εντολής `histogram()`) ενεργοποιεί τη συνάρτηση "γενικού πάνελ" για το γράφημα → αυτό ρυθμίζει τα εξωτερικά στοιχεία του πάνελ (ορθογώνιο πλαίσιο, κλίμακα, άξονες με σημεία επισήμανσης, ετικέτες, κ.λπ.).
 - Η γενική συνάρτηση απεικόνισης καλεί τη 'συνάρτηση πλέγματος' που δημιουργεί όλα τα ζητούμενα γραφήματα, τα οποία τοποθετούνται στην περιοχή σχεδίασης του γραφήματος

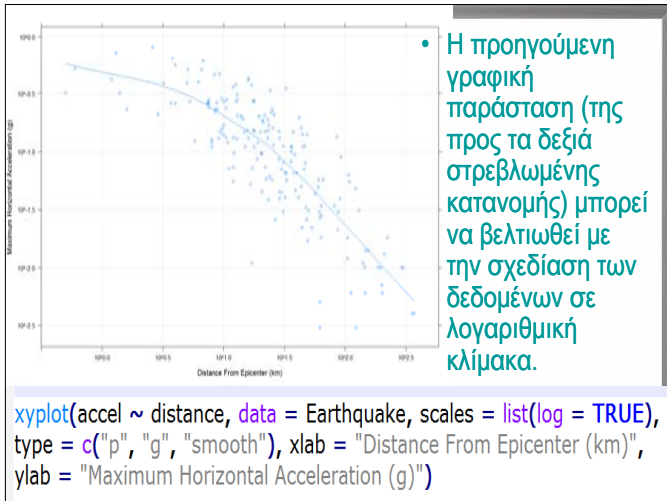
- Κάθε συνάρτηση γενικής απεικόνισης στο πλέγμα έχει μια αντίστοιχα προεπιλεγμένη συνάρτηση πλέγματος, συνήθως με το όνομα της γενικής συνάρτησης με το πρόθεμα *panel*.
 - Για παράδειγμα, η προεπιλεγμένη συνάρτηση πλέγματος για τη γενική συνάρτηση απεικόνισης, `histogram()`, ονομάζεται `panel.histogram()`.
- Η ίδια η συνάρτηση του πλέγματος καλείται από τη γενική συνάρτηση απεικόνισης με την παράμετρο `panel`, π.χ., η εντολή `xyplot(y ~ x, data = dataset)` είναι ισοδύναμη με την εντολή: `xyplot(y~x, data=dataset, panel = panel.xyplot)`



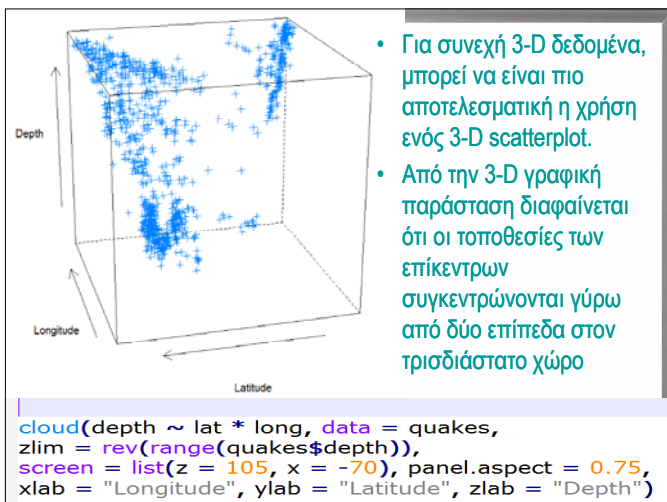
- Παράδειγμα: Σεισμικά επεισόδια (Δυτική Β. Αμερική μεταξύ 1940 και 1980) - Βάθη σεισμών vs. Μέγεθος Σεισμού



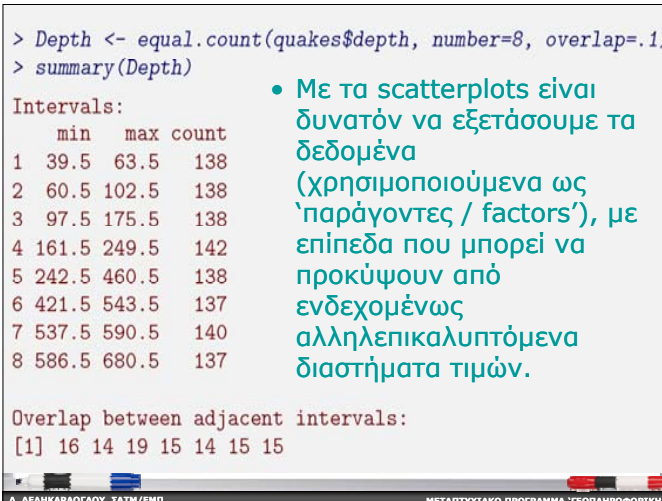
- Τα scatterplots χρησιμοποιούνται συνήθως για συνεχή διμεταβλητά δεδομένα, καθώς και για δεδομένα χρονοσειρών, π.χ. Εδώ: Μέγιστη οριζόντια επιτάχυνση που μετράται σε σχέση με την απόσταση του σταθμού μέτρησης από το επίκεντρο.
- Η γραφική παράσταση παρουσιάζει μοτίβα μιας προς τα δεξιά στρεβλωμένης κατανομής



- Η προηγούμενη γραφική παράσταση (της προς τα δεξιά στρεβλωμένης κατανομής) μπορεί να βελτιωθεί με την σχεδίαση των δεδομένων σε λογαριθμική κλίμακα.



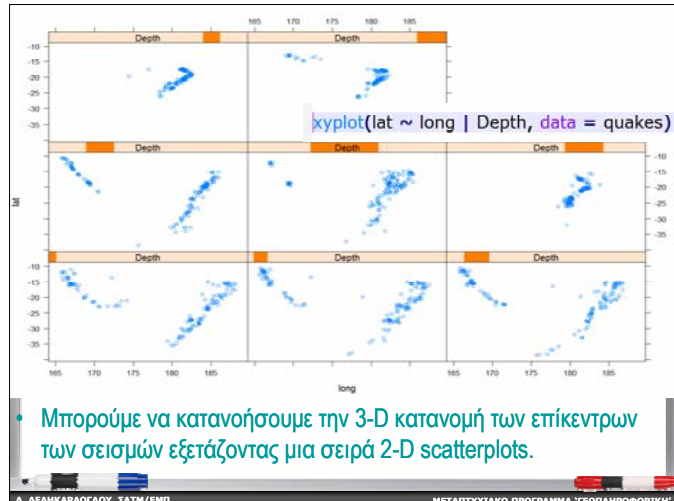
- Για συνεχή 3-D δεδομένα, μπορεί να είναι πιο αποτελεσματική η χρήση ενός 3-D scatterplot.
- Από την 3-D γραφική παράσταση διαφαίνεται ότι οι τοποθεσίες των επίκεντρων συγκεντρώνονται γύρω από δύο επίπεδα στον τρισδιάστατο χώρο



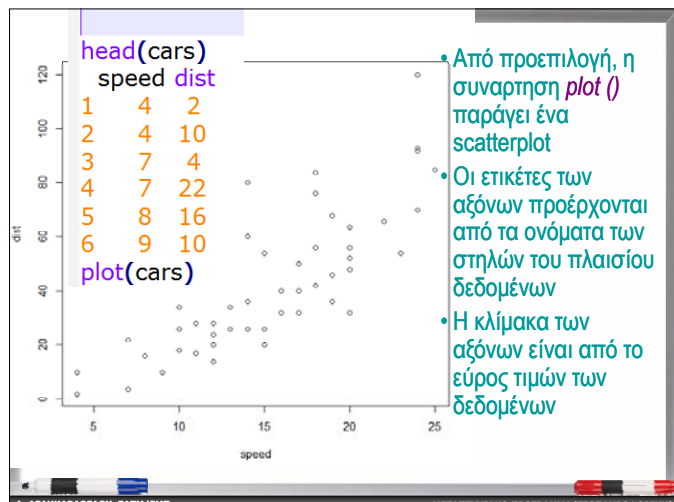
- Με τα scatterplots είναι δυνατόν να εξετάσουμε τα δεδομένα (χρησιμοποιούμενα ως 'παράγοντες / factors'), με επίπεδα που μπορεί να προκύψουν από ενδεχομένως αλληλεπικαλυπτόμενα διαστήματα τιμών.



- Στατικά τρισδιάστατα scatterplots δεν είναι πολύ χρήσιμα λόγω της ισχυρής επίδρασης της κατεύθυνσης προβολής.
- Δυστυχώς, τα γραφικά πλέγματος δεν επιτρέπουν διαδραστικό χειρισμό της κατεύθυνσης προβολής των δεδομένων.

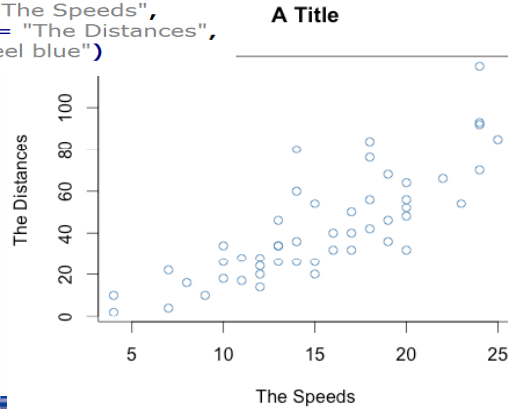


- Μπορούμε να κατανοήσουμε την 3-D κατανομή των επίκεντρων των σεισμών εξετάζοντας μια σειρά 2-D scatterplots.



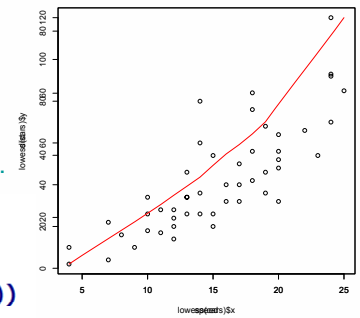
- Από προεπιλογή, η συνάρτηση `plot()` παράγει ένα scatterplot
- Οι ετικέτες των αξόνων προέρχονται από τα ονόματα των στηλών του πλαισίου δεδομένων
- Η κλίμακα των αξόνων είναι από το εύρος τιμών των δεδομένων


```
plot(cars$speed,
cars$dist,
main = "A Title",
xlab = "The Speeds",
ylab = "The Distances",
col="steel blue")
```



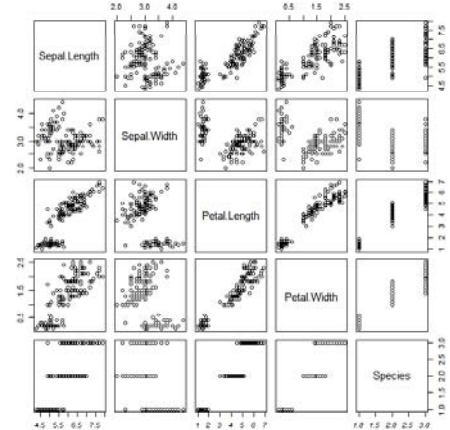
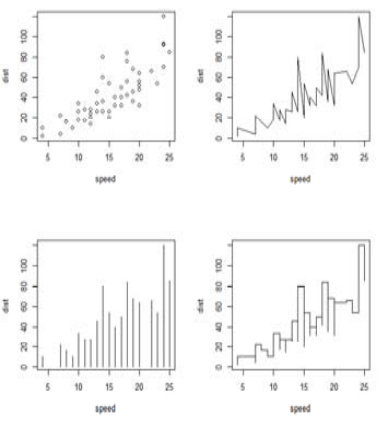
- Κάθε κλήση της `plot()` δημιουργεί άξονες, ετικέτες κ.λπ.
- Συνδυασμός πολλαπλών κλήσεων της `plot()` μπορεί να δημιουργήσει προβλήματα
- π.χ., οι ετικέτες στους άξονες,
- όπως (αν και όχι τόσο εμφανές) οι κλίμακες στους άξονες διαφέρουν.

```
plot(cars)
par(new=TRUE)
plot((lowess(cars),
type="l", col="red"))
```

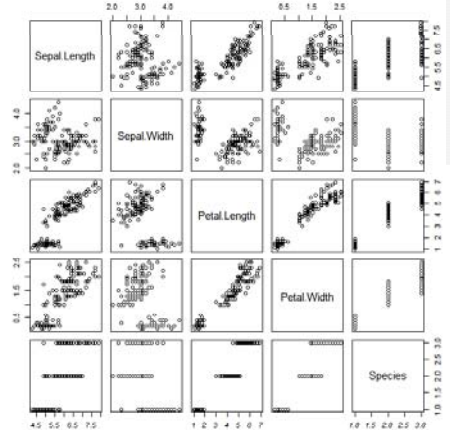


- Πολλαπλά γραφήματα στο ίδιο παράθυρο γραφικών με την εντολή `par()` και την παράμετρο `mfrow`

```
plot(cars)
par(mfrow=c(2, 2))
plot(cars, type="p")
plot(cars, type="l")
plot(cars, type="h")
plot(cars, type="s")
```



`plot(iris)`
Ο προεπιλεγμένος τύπος γραφικής παράστασης για ένα πλαίσιο δεδομένων είναι ένα (ή πολλαπλά) scatterplot

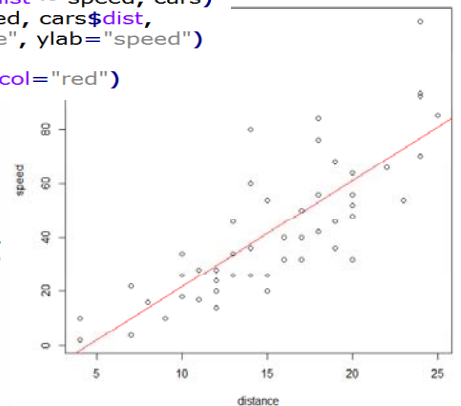


```
png(filename = "iris.png")
plot(iris)
dev.off()
```

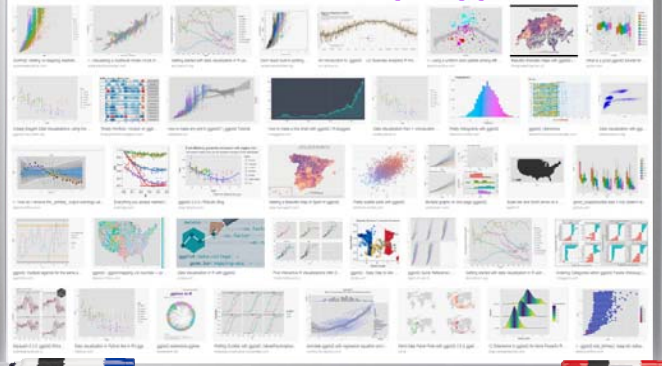
Το γράφημα μπορεί επίσης να προωθηθεί σε κάποιο αρχείο

```
lm(dist ~ speed, cars)
car.fit <- lm(dist ~ speed, cars)
plot(cars$speed, cars$dist,
xlab="distance", ylab="speed")
abline(car.fit, col="red")
```

Δεδομένα + αποτέλεσμα παλινδρόμησης
Plot data fit



Αναζήτηση στο Google: ggplot2



Η φιλοσοφία του ggplot2



- Διατυπώθηκε από τον καθ. Hadley Wickham (Rice Univ., Houston, Texas) επεκτείνοντας την "Γραμματική των Γραφικών": Όλα τα γραφήματα μπορούν να κατασκευαστούν συνδυάζοντας προδιαγραφές με δεδομένα.
- Μια προδιαγραφή είναι ένας δομημένος τρόπος για να περιγραφεί ο τρόπος κατασκευής ενός γραφήματος από γεωμετρικά αντικείμενα (σημεία, γραμμές κ.λπ.) που προβάλλονται σε κλίμακες (x, y, χρώμα, μέγεθος κ.λπ.)

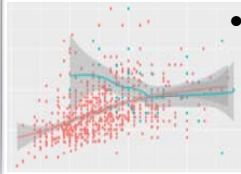
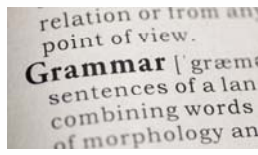
Η γραμματική των γραφικών απεικονίσεων

- Ο όρος χρησιμοποιείται για να υποδηλώσει ένα εσωτερικευμένο σύστημα κανόνων που επιτρέπει να κατανοούμε τη δομή και τη σύνταξη γραφημάτων, και συνεπώς να περιγράψουμε και να ερμηνεύουμε τα γραφικά δεδομένα όπως διαισθητικά κάνουμε με τις γραπτές προτάσεις.
- Στόχος της οπτικοποίησης δεδομένων είναι να αποδώσει με συνοπτικό/συνεκτικό και αισθητικό τρόπο κάποιου είδους πληροφορίες ώστε να γίνουν ευκολότερα κατανοητές (όπως, ένα βιβλίο → σε ταινία)



Τι ακριβώς είναι η γραμματική των γραφικών απεικονίσεων;

- Η γραμματική μιας γλώσσας ορίζει τους κανόνες δόμησης λέξεων και φράσεων σε εκφράσεις με νόημα.



- Μια γραμματική γραφικών ορίζει τους κανόνες της δόμησης των μαθηματικών και αισθητικών στοιχείων σε ένα γράφημα με νόημα.

Το ανάλογο του 'παντογράμματος'

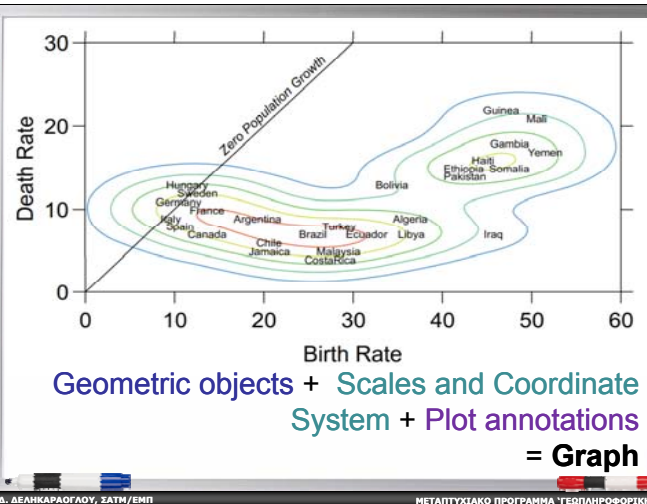
(μια πρόταση στην οποία χρησιμοποιούνται όλα τα γράμματα ενός αλφάβητου)

- *The quick brown fox jumps over the lazy dog*
- *A rapid red fox bit the friendly dog.*
- *The friendly dog was bitten by a rabid red fox*

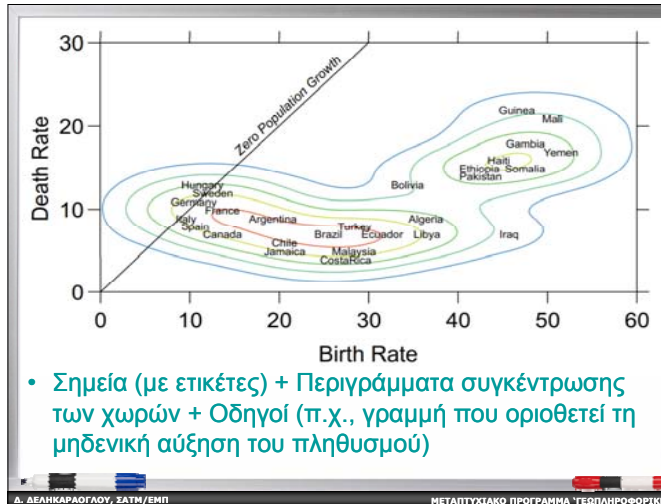
Η δομή της πρότασης υπαγορεύει τη γραμματική

- Αναδιατάσσοντας γραμματικά στοιχεία σε κείμενο (σημεία στίξης, σειρά ρήματος) και χρησιμοποιώντας συγκεκριμένες λέξεις (επίθετα, επιρρήματα), μπορούμε να ελέγξουμε με ευκρίνεια το μήνυμα που θέλουμε να επικοινωνήσουμε. Το ίδιο ισχύει και για τις οπτικοποιήσεις δεδομένων.

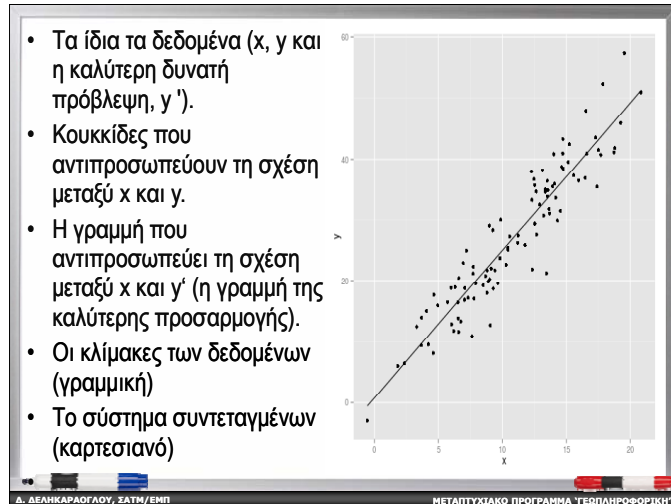
Όπως οι προτάσεις είναι κάπως συνθέσεις προσεκτικά επιλεγμένων γραμματικών στοιχείων που μεταφέρουν ακριβή και καθαρά μηνύματα (*The quick brown fox jumps over the lazy dog*), η 'πλοκή' των γραφικών εξαρτάται από κάποια βασικά χαρακτηριστικά: **DATA** - συναρτήσεις που απαιτούνται για τη δημιουργία μεταβλητών από δεδομένα ♦ **TRANS** - μετατροπές εάν υπάρχουν, που πρέπει να εφαρμοστούν στις μεταβλητές ♦ **FRAME** - Το γραφικό περιεχόμενο της αναπαράστασης των δεδομένων, π.χ. αλγεβρική έκφραση (παλινδρόμηση), ♦ **SCALE** - Διαστάσεις στις οποίες προσανατολίζονται τα γραφικά (*mapping between values and visual properties*)



Geometric objects + Scales and Coordinate System + Plot annotations = Graph



- Σημεία (με ετικέτες) + Περιγράμματα συγκέντρωσης των χωρών + Οδηγόι (π.χ., γραμμή που οριοθετεί τη μηδενική αύξηση του πληθυσμού)



- Τα ίδια τα δεδομένα (x, y και η καλύτερη δυνατή πρόβλεψη, y').
- Κουκκίδες που αντιπροσωπεύουν τη σχέση μεταξύ x και y.
- Η γραμμή που αντιπροσωπεύει τη σχέση μεταξύ x και y' (η γραμμή της καλύτερης προσαρμογής).
- Οι κλίμακες των δεδομένων (γραμμική)
- Το σύστημα συντεταγμένων (καρτεσιανό)

Γραφικές απεικονίσεις με το ggplot2

- **ggplot2** → Ένα εργαλείο ανοικτού κώδικα και πηγή υλοποίησης (φορμαλισμός) της γραμματικής γραφικών
 - Ένα από τα καλύτερα συντηρημένα, πιο σημαντικά και πραγματικά καλοδουλεμένα πακέτα του R
 - Συνέργεια με άλλα δημοφιλή εργαλεία γραφικών απεικονίσεων ... **Python, React, Jupyter, MATLAB, και sagemath**
- ▲ Ένα σύστημα σχεδίασης με καλές προεπιλογές για ένα μεγάλο σύνολο στοιχείων που μπορούν να συνδυαστούν με ευέλικτους και δημιουργικούς τρόπους (π.χ., σε υψηλό επίπεδο αφαίρεσης, ...)

Το πακέτο ggplot2

- Το ggplot2 δεν αποτελεί μέρος της κανονικής κατανομής του R. Συνεπώς πρέπει να γίνει λήψη του πακέτου από το CRAN και να εγκατασταθεί στο R → `install.packages("ggplot2")`
- Προσοχή: Ίσως κάπως συγκεχυμένη, η πιο σημαντική λειτουργία σε αυτό το πακέτο είναι η συνάρτηση `ggplot()`



Εγκατάσταση του ggplot2 στο R

- # The easiest way is to install the whole **tidyverse**: `install.packages("tidyverse")`
- # Alternatively, install just ggplot2: `install.packages("ggplot2")`
- # Or the development version from GitHub:
- # `install.packages("devtools")`
`devtools::install_github("tidyverse/ggplot2")`
- Get the useful cheat sheet: <https://github.com/rstudio/cheatsheets/blob/main/data-visualization-2.1.pdf> → **Download**

Το ggplot2 ...

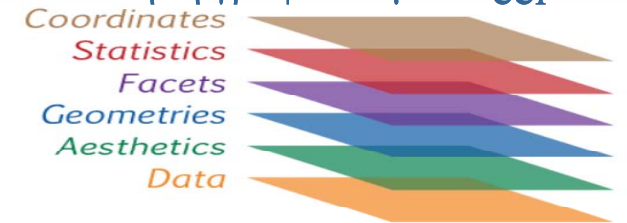
- Παράγει πολυεπίπεδα γραφικά
- Χρησιμοποιεί μια υποκείμενη “γραμματική” για τη δημιουργία γραφημάτων *επίπεδο-προς-επίπεδο* αντί να παρέχει προκατασκευασμένα γραφήματα
- Είναι αρκετά εύκολο στη χρήση χωρίς καμία έκθεση του χρήστη στην υποκείμενη γραμματική, αλλά είναι ακόμα πιο εύκολο στη χρήση όταν ο χρήστης είναι γνώστης της γραμματικής
- Επιτρέπει στον χρήστη να δημιουργήσει ένα γράφημα από σημασιολογικές έννοιες αντί για ανάκληση εντολών και επιμέρους επιλογών

Το ggplot2 ...

- Παράγει πολυεπίπεδα γραφικά
- Χρησιμοποιεί μια υποκείμενη “γραμματική” για τη δημιουργία γραφημάτων *επίπεδο-προς-επίπεδο* αντί να παρέχει προκατασκευασμένα γραφήματα
- Είναι αρκετά εύκολο στη χρήση χωρίς καμία έκθεση του χρήστη στην υποκείμενη γραμματική, αλλά είναι ακόμα πιο εύκολο στη χρήση όταν ο χρήστης είναι γνώστης της γραμματικής
- Επιτρέπει στον χρήστη να δημιουργήσει ένα γράφημα από σημασιολογικές έννοιες αντί για ανάκληση εντολών και επιμέρους επιλογών



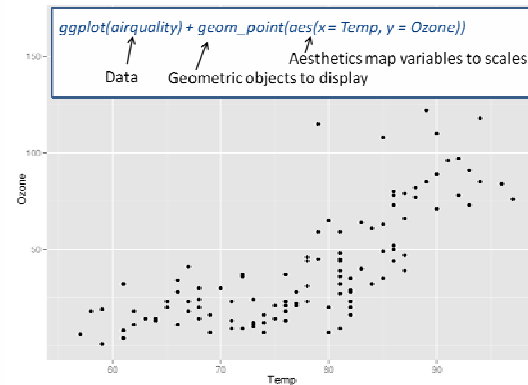
Υλοποίηση γραφικών με το ggplot2



Κάθε γράφημα αποτελείται από τα βασικά συστατικά μέρη του

- **data** - δεδομένα για απεικόνιση
- **aesthetics** / αισθητική – μεταβλητές που αντιστοιχούνται σε αισθητικές ιδιότητες
- **geometric objects** / γεωμετρία – Αντιληπτά γεωμετρικά αντικείμενα (points, bars, κ.λπ.)
- **scales** – κλίμακες αντιστοιχούν τιμές από δεδομένα στον αισθητικό χώρο
- **faceting** – υποδιαίρεση των δεδομένων για να εμφανιστούν σε πολλαπλά γραφήματα
- **statistical transformations** – σύνοψη δεδομένων
- **coordinate systems** θέτουν τα δεδομένα στο επίπεδο της γραφικής παράστασης

Τυπικό γράφημα από το ggplot2



- Όλα τα γραφικά ξεκινούν με την κλήση της συνάρτησης **ggplot()**
- καθορίζουμε το σύνολο δεδομένων που περιέχει τις μεταβλητές που θα αντιστοιχίσουμε με την **αισθητική**, δηλ. τις οπτικές ιδιότητες του γραφήματος: **ggplot(data, aes(x=xvar, y=yvar))**
- **data**: το όνομα του αντικείμενου των δεδομένων
- **x** και **y**: αισθητική που τοποθετεί αντικείμενα στο γράφημα
- **xvar** και **yvar**: ονόματα μεταβλητών στα δεδομένα που αντιστοιχίζονται σε τιμές δεδομένων x και y
- Η αισθητική **aes()** που ορίζεται στο εσωτερικό της ggplot() **κληρονομείται** στα επόμενα στρώματα

ggplot2 : data

ggplot(**data = <DATA>**, mapping = **aes(<MAPPINGS>)**) + **<GEOM_FUNCTION>()**

- **Περιγραφή**: Μέσω αυτού του συστατικού ορίζουμε το σύνολο δεδομένων εισόδου που θα χρησιμοποιηθεί στην απεικόνιση →
- **Σύνταξη**: Το όνομα του συνόλου δεδομένων που ορίζεται στη συνάρτηση **ggplot()**, αρχικοποιεί ένα νέο γράφημα και παρέχει μια πιο προγραμματική διεπαφή για τον καθορισμό των μεταβλητών που θα απεικονιστούν. EAN δεν προσδιορίζεται, πρέπει να παρέχεται σε κάθε στρώμα που προστίθεται στο γράφημα.

ggplot2 : data

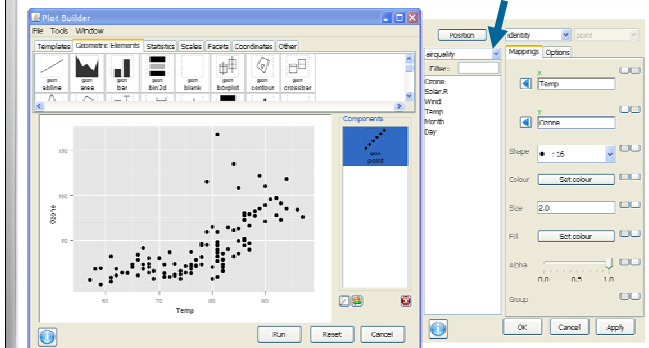
- Το ggplot2 αναμένει τα δεδομένα ως **data.frame**:
 - Rows: observations
 - Columns: variables

```
diamonds <- data.frame(carat, cut, price)
  carat  cut  price
1 0.23  Ideal  326
2 0.21  Premium 326
3 0.23   Good  327
4 0.29  Premium  33
```

- Διάφορα επίπεδα γραφικών μπορούν να χρησιμοποιήσουν διαφορετικά δεδομένα (π.χ., μια προ-υπολογισμένη σύνοψη στοιχείων σε ένα διαφορετικό πλαίσιο δεδομένων)

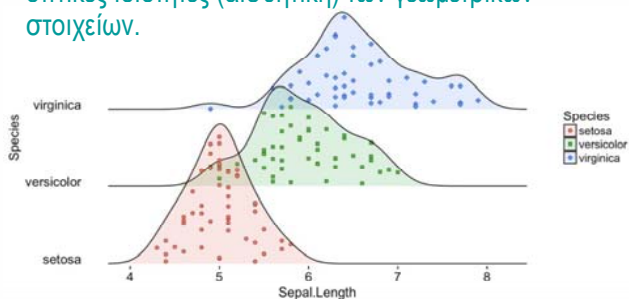
ggplot2 : data

Δεδομένα (data) στο Deducer



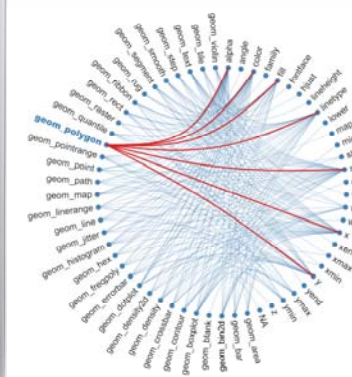
ggplot2 : aesthetics

- Η έννοια της αισθητικής των γραφικών υποδηλώνει πως οι μεταβλητές στα δεδομένα μετατρέπονται στις οπτικές ιδιότητες (αισθητική) των γεωμετρικών στοιχείων.



Καθορισμός αισθητικής

- Ποια αισθητικές ιδιότητες απαιτούνται και ποια χαρακτηριστικά επιτρέπονται ανάλογα με τα γεωμετρικά στοιχεία, το `geom()`.



Κοινές παράμετροι αισθητικής

- `x, y`: τοποθέτηση κατά μήκος των αξόνων x και y
- `color`: χρώμα αντικειμένων. Για 2-d αντικείμενα, το χρώμα του περιγράμματος του αντικειμένου
- `fill`: γέμισμα χρώματος 2-d αντικειμένων
- `linetype`: πώς πρέπει να σχεδιάζονται οι γραμμές (συμπαγείς, διακεκομμένες, με σημεία, κ.λπ.)
- `shape`: σχήμα συμβόλων απεικόνισης αντικειμένων δεδομένων (π.χ., σε διαγράμματα διασποράς)
- `size`: μέγεθος εμφάνισης αντικειμένων απεικόνισης
- `alpha`: διαφάνεια αντικειμένων (μεταξύ 0=διαφανές και 1=αδιαφανές)

```

aes(x = mpg, y = wt)
# Aesthetic mapping:
# * `x` -> `mpg`
# * `y` -> `wt`
aes(mpg, wt)
# Aesthetic mapping:
# * `x` -> `mpg`
# * `y` -> `wt`
    
```

x, y position (the first two by default), color, fill, shape, linetype, size, group, (depending on the geom)

```

# You can also map aesthetics to functions of variables
aes(x = mpg ^ 2, y = wt / cyl)
# Aesthetic mapping:
# * `x` -> `mpg^2`
# * `y` -> `wt/cyl`

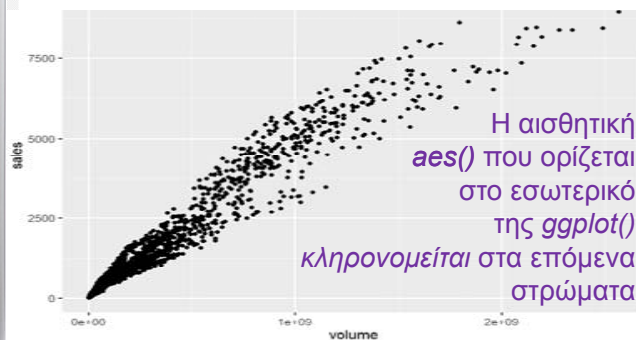
# Or to constants
aes(x = 1, colour = "smooth")
# Aesthetic mapping:
# * `x` -> 1
# * `colour` -> "smooth"
    
```

Αισθητικές ιδιότητες των γεωμετρικών στοιχείων:

scatter plot of volume vs sales

```

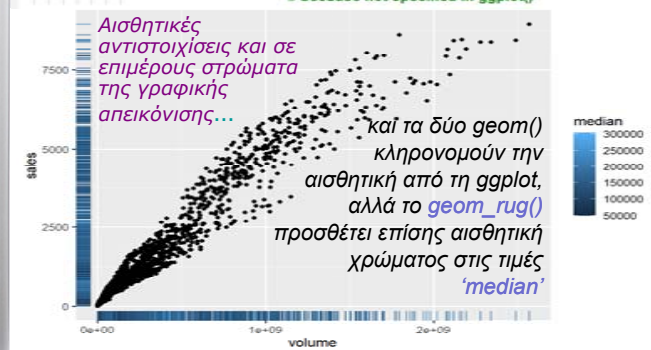
ggplot(txhousing, aes(x=xvariable, y=yvariable)) +
geom_point()
    
```



Η αισθητική `aes()` που ορίζεται στο εσωτερικό της `ggplot()` κληρονομείται στα επόμενα στρώματα

```

# scatter plot of volume vs sales
# with rug plot colored by median sale price
ggplot(txhousing, aes(x=volume, y=sales)) +
  geom_point() +
  geom_rug(aes(color=median))
# color will only apply to the rug plot
# because not specified in ggplot()
    
```

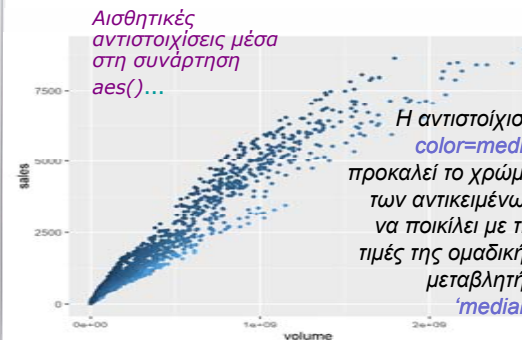


Αισθητικές αντιστοιχίσεις και σε επιμέρους στρώματα της γραφικής απεικόνισης... και τα δύο `geom()` κληρονομούν την αισθητική από τη `ggplot`, αλλά το `geom_rug()` προσθέτει επίσης αισθητική χρώματος στις τιμές `'median'`

mapping color to median inside of aes()

```

ggplot(txhousing, aes(x=volume, y=sales)) +
geom_point(aes(color=median))
    
```

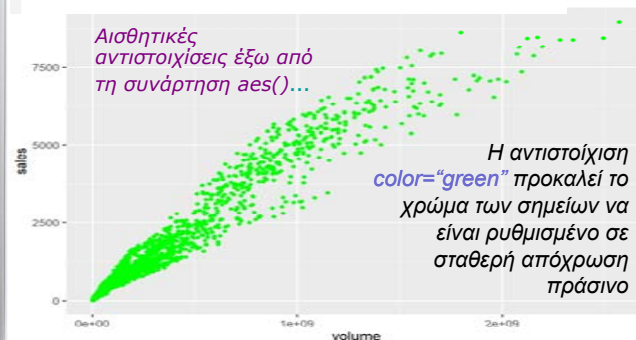


Αισθητικές αντιστοιχίσεις μέσα στη συνάρτηση `aes()`... Η αντιστοιχία `color=median` προκαλεί το χρώμα των αντικειμένων να ποικίλει με τις τιμές της ομαδικής μεταβλητής `'median'`

setting color to green outside of aes()

```

ggplot(txhousing, aes(x=volume, y=sales)) +
geom_point(color="green")
    
```



Αισθητικές αντιστοιχίσεις έξω από τη συνάρτηση `aes()`... Η αντιστοιχία `color="green"` προκαλεί το χρώμα των σημείων να είναι ρυθμισμένο σε σταθερή απόχρωση πράσινο

```

# color="green" inside of aes()
# geom_point() cannot find a variable called "green" and
# uses a default color instead
ggplot(txhousing, aes(x=volume, y=sales)) +
geom_point(aes(color="green"))
    
```



Η ρύθμιση μιας αισθητικής ως σταθερά μέσα στη συνάρτηση `aes()`... μπορεί να οδηγήσει σε απροσδόκητα αποτελέσματα, καθώς η αισθητική ορίζεται στη συνέχεια σε μια προεπιλεγμένη τιμή αντί για την καθορισμένη τιμή.

ggplot2 : aesthetics

geom_point()

X

Y

Shape

Colour

Size

Fill

Alpha

Group

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

ggplot2 : aesthetics

geom_point()

X

Y

Shape

Colour

Size

Fill

Alpha

Group

Points & lines

Areas (inside Polygons)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- **Mapping:** variable ↔ visual property
 - Εντός της εντολής κλήσης της `aes(x, y, ...)`
`ggplot(data = airquality) + geom_point(aes(x = Temp, y = Ozone, color = Month))`
 - *Color is mapped to month*
 - **Setting:** fixed value → visual property
 - Εκτός της εντολής κλήσης της `aes(x, y, ...)`
`ggplot(data = airquality) + geom_point(aes(x = Temp, y = Ozone), color = "red")`
 - *Color is set to "red" – δεν σχετίζεται με κάποια μεταβλητή με το όνομα "red"*
- Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

mapping vs setting

Column of buttons switch between states

These two are being mapped

Remainder are set (using default settings)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

ggplot2 : geometric objects

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

ggplot2 : geometric objects

The `ggplot()` function

The data parameter

The `aes()` function

```
ggplot(data = , aes(x = , y = )) + geom_line()
```

The geometric object we want to draw (i.e., the geom)

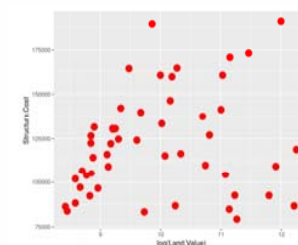
geom() := οπτικές δομές για την οπτικοποίηση δεδομένων

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Οι συναρτήσεις `geom_()` διαφέρουν ως προς τα γεωμετρικά σχήματα που χρησιμοποιούνται για την 'πλοκή' ενός γραφήματος. Μερικά παραδείγματα:
- `geom_bar()`: ράβδοι με βάσεις στον άξονα x
 - `geom_boxplot()`: κουτιά και όρια
 - `geom_errorbar()`: Γραμμές σφάλματος σε σχήμα T
 - `geom_density()`: γραφήματα πυκνότητας
 - `geom_histogram()`: ιστόγραμμα
 - `geom_line()`: γραμμές
 - `geom_point()`: σημεία (scatterplot)
 - `geom_ribbon()`: ζώνες που εκτείνονται σε τιμές y σε ένα εύρος τιμών x
 - `geom_smooth()`: εξομαλυνόμενες τιμές
 - `geom_text()`: κείμενο
- Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- **Περιγραφή:** καθορίζουμε τον τρόπο με τον οποίο οι τιμές των μεταβλητών θα εμφανιστούν στο γράφημα (όπως διάγραμμα ράβδων, ιστόγραμμα, διάγραμμα σκέδασης, ...) και κάθε τύπος ορίζεται για συγκεκριμένο αριθμό και τύπους μεταβλητών.

- **Σύνταξη:** Η σύνταξη ξεκινά με το `geom_*` και το όνομα του εκάστοτε χρησιμοποιούμενου σχήματος (π.χ., `geom_point`)



Κανόνες αισθητικής των γραφικών:

- Κάθε γεωμετρικό σχήμα μπορεί να οριστεί με το **δικό του σύνολο δεδομένων και αισθητική** που θα ισχύει μόνο για αυτό το σχήμα.
 - Σε αυτή την περίπτωση το πλαίσιο δεδομένων και `aes()` ορίζονται στη συνάρτηση `geom_*()`.
 - Κάθε σχήμα έχει τα δικά του **συγκεκριμένα αισθητικά χαρακτηριστικά**: π.χ., `hjust` και `vjust` είναι για το `geom_text()` και το `linetype` για γραμμές.
 - Μπορούμε επίσης και να συνδυάσουμε γεωμετρικά σχήματα
- #two geom shapes - geom_line and geom_point used on one graph**
- ```
ggplot(GOT, aes(x=Episode,y=Number_of_viewers, colour=Season, group=Season)) + geom_line()+geom_point()
```
- Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ  
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

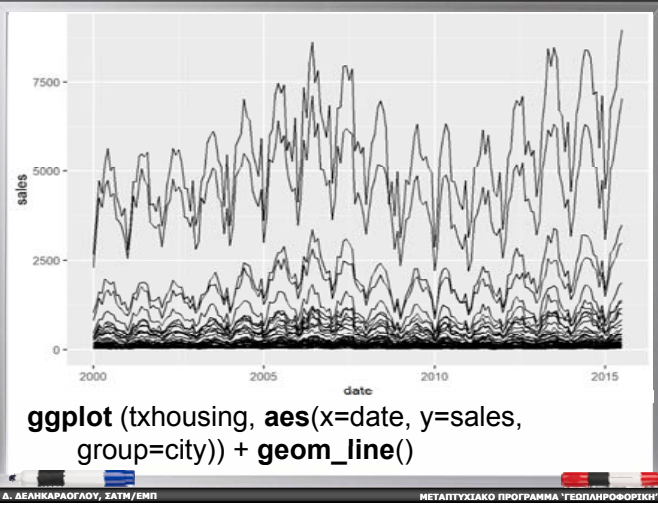
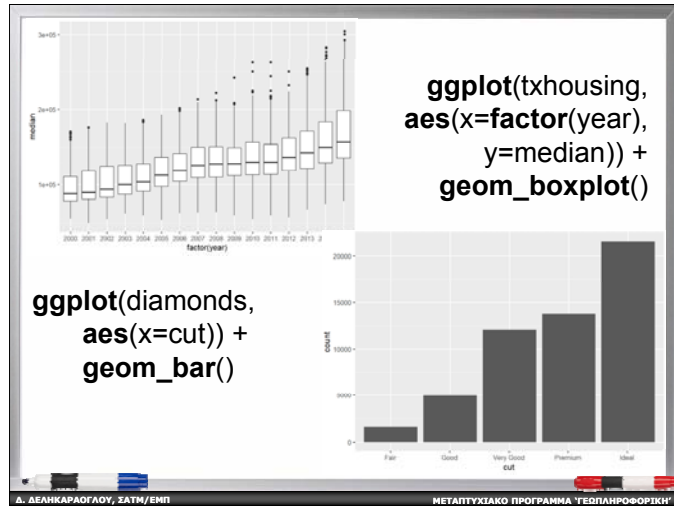
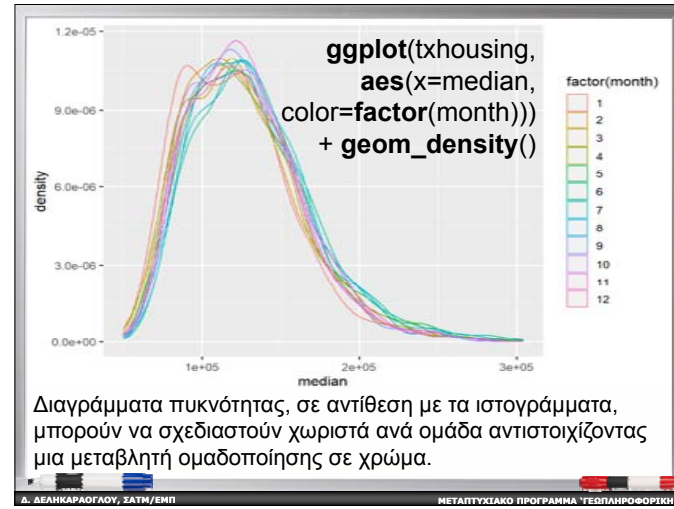
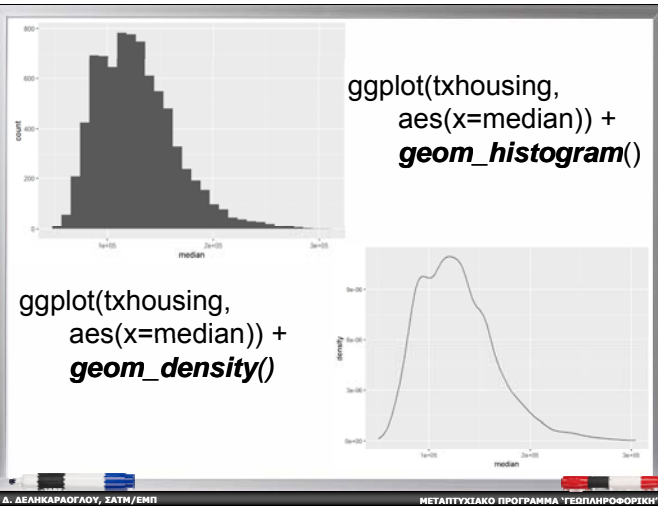
| Number/type of variables                   | Geometrical shape | Plot type                                                         |
|--------------------------------------------|-------------------|-------------------------------------------------------------------|
| One continuous                             | geom_area()       | area plot                                                         |
| One continuous                             | geom_density()    | smoothed density estimate                                         |
| One continuous or X discrete, Y continuous | geom_dotplot()    | stack individual points into a dot plot                           |
| One continuous                             | geom_histogram()  | bin and count continuous variable, display with bars              |
| One discrete or X discrete, Y continuous   | geom_bar()        | display distribution of discrete variable                         |
| Two continuous                             | geom_blank()      | display nothing. Most useful for adjusting axes limits using data |
| Two continuous                             | geom_jitter()     | randomly jitter overlapping points                                |
| Two continuous                             | geom_point()      | points                                                            |
| Two continuous                             | geom_quantile()   | smoothed quantile regression                                      |
| Two continuous                             | geom_rug()        | marginal rug plots                                                |
| Two continuous                             | geom_smooth()     | smoothed line of best fit                                         |
| Two continuous                             | geom_text()       | text labels                                                       |
| X discrete, Y continuous                   | geom_boxplot()    | boxplots                                                          |
| X discrete, Y continuous                   | geom_violin()     | show density of values in each group                              |
| X time, Y continuous                       | geom_area()       | Area plot                                                         |
| X time, Y continuous                       | geom_line()       | Line plot                                                         |
| X time, Y continuous                       | geom_step()       | Step plot                                                         |

### ggplot2 : *geometric objects*

- geom\_abline**: Line, specified by slope and intercept
- geom\_area**: Area plots
- geom\_bin2d**: Add heatmap of 2d bin counts
- geom\_boxplot**: Box and whiskers plot
- geom\_crossbar**: Hollow bar with middle indicated by horizontal line
- geom\_density**: Display a smooth density estimate
- geom\_density2d**: Contours from a 2d density estimate
- geom\_errorbar**: Horizontal error bars
- geom\_frug**: Marginal rug plots
- geom\_jitter**: Points, jittered to reduce overplotting
- geom\_line**: Line, horizontal
- geom\_line**: Line, vertical
- geom\_linerange**: An interval represented by a vertical line
- geom\_point**: Points, as for a scatterplot
- geom\_polygon**: Polygon, a filled path
- geom\_rect**: 2d rectangles
- geom\_rug**: Marginal rug plots
- geom\_smooth**: Add a smoothed condition mean
- geom\_text**: Textual annotations
- geom\_vline**: Line, vertical
- geom\_path**: Connect observations, in original order
- geom\_pointrange**: An interval represented by a vertical line, with a point in the middle
- geom\_quantile**: Add quantile lines from a quantile regression
- geom\_ribbon**: Add quantile lines from a continuous x values
- geom\_sagittae**: Single line segments
- geom\_step**: Connect observations by stairs
- geom\_tile**: Tile plot as densely as possible, assuming that every tile is the same size.

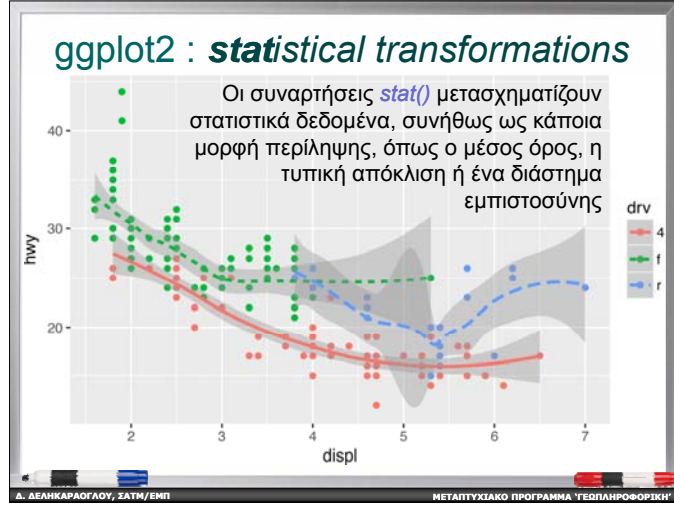
### ggplot2 : *geom και αισθητική*

- Κάθε geom() ορίζεται από την αισθητική που απαιτείται για την απόδοσή του, π.χ.
  - geom\_point() απαιτεί x και y ως ελάχιστη προδιαγραφή για ένα scatterplot
- Τα geom() διαφέρουν ως προς την αισθητική που δέχονται ως ορίσματα
  - geom\_point() δέχεται την αισθητική shape, η οποία ορίζει τα σχήματα των σημείων στο γράφημα, ενώ geom\_bar() δεν δέχεται shape.
- ΣΥΝΙΣΤΑΤΑΙ να συμβουλευέστε τη βοήθεια του R για τη συναρτήσεις *geom()* ως προς τις απαιτήσεις και την κατανόηση της εκάστοτε αισθητικής.



### ggplot2 : *statistical transformations*

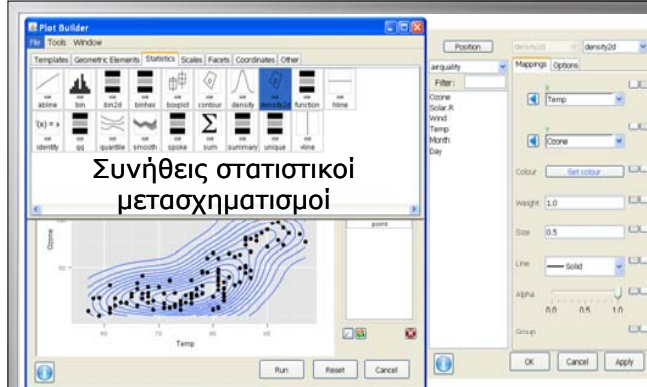
- Μερικοί τύποι γραφημάτων (scatterplots) δεν απαιτούν μετασχηματισμούς - κάθε σημείο σχεδιάζεται με τις συντεταγμένες του x και y ίσες με την αρχική τιμή τους. Άλλα γραφήματα, π.χ. τα boxplots, τα ιστογράμματα, οι γραμμές πρόβλεψης κλπ. απαιτούν στατιστικούς μετασχηματισμούς:
  - για ένα boxplot οι τιμές y πρέπει να μετατραπούν στο διάμεσο και στο 1.5 (IQR, Interquartile range)
  - για ομαλότερη εξομάλυνση, οι τιμές y πρέπει να μετατραπούν σε προβλεπόμενες τιμές





## ggplot2 : **statistical transformations**

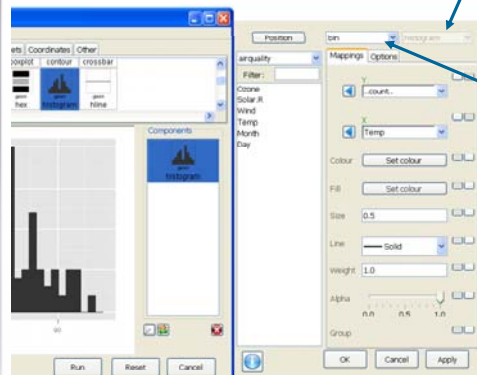
- **Περιγραφή:** Χρησιμοποιούνται για τη μετατροπή των δεδομένων πριν από την απεικόνιση. Πολλοί από αυτούς τους μετασχηματισμούς χρησιμοποιούνται "πίσω από τη σκηνή" κατά τη δημιουργία γεωμετρικών σχημάτων. Συχνά δεν τους ορίζουμε άμεσα, το ggplot2 το κάνει αυτό μόνο του.
- **Σύνταξη:** Η σύνταξη εξαρτάται από τον εκάστοτε μετασχηματισμό που χρησιμοποιείται → διάφορα στατιστικά μεγέθη



Κάθε συνάρτηση `stat()` συσχετίζεται με ένα προεπιλεγμένο `geom()`, επομένως δεν απαιτείται `geom()` για την απόδοση των σχημάτων.

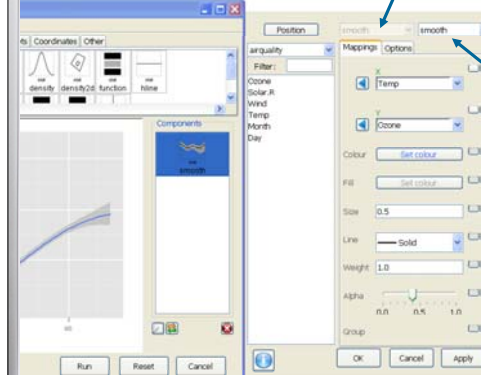
| Statistical transformation | Used by Geometrical shape             | Description                                                                                                             |
|----------------------------|---------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| <code>stat_count</code>    | <code>geom_bar</code>                 | It counts number of case for each category of discrete variable x                                                       |
| <code>stat_identity</code> | <code>geom_col</code>                 | Leaves columns as is                                                                                                    |
| <code>stat_bin</code>      | <code>geom_bar, geom_histogram</code> | Creates additional data frame which tells us how many points are in each bin                                            |
| <code>stat_boxplot</code>  | <code>geom_boxplot</code>             | Creates additional variables that are necessary for boxplot visualization (mean, median, mod, min value, max value etc) |
| <code>stat_density</code>  | <code>geom_density</code>             | Computes kernel density estimate, which is a smoothed version of the histogram.                                         |
| <code>stat_summary</code>  | <code>geom_point</code> (can be used) | Summarise y values at unique/binned x (for example mean, median per each bin)                                           |
| <code>stat_unique</code>   | Several geometrical shapes            | Remove duplicate rows from a data frame                                                                                 |

Αν καθοριστεί ένα γεωμετρικό αντικείμενο



Μπορεί να αλλάξει ο στατιστικός μετασχηματισμός

Αν καθοριστεί ο στατιστικός μετασχηματισμός



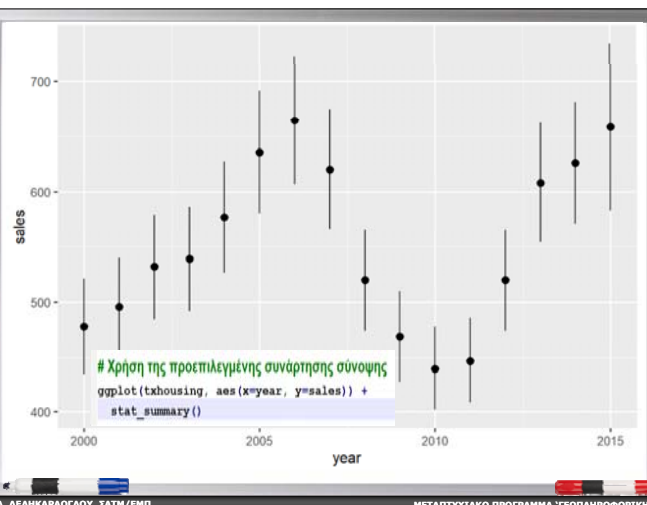
Μπορεί να αλλάξει το γεωμετρικό αντικείμενο

- Με στατιστικά στοιχεία δημιουργούνται **πρόσθετες μεταβλητές** (συνήθως μερικές συγκεντρωτικές ή παρόμοιες τιμές). Για την απεικόνιση αυτών των δεδομένων **πρέπει να χρησιμοποιήσουμε κάποια συνάρτηση `geom_*()`**. Διαφορετικά, οι νέες μεταβλητές δεν θα είναι ορατές στην οθόνη → Αυτό γίνεται με δύο τρόπους

```
two geom shapes - geom_line and geom_point used on one graph
ggplot(GOT, aes(x=Episode, y=Number_of_viewers,
 colour=Season, group=Season)) + geom_line()+geom_point()
```

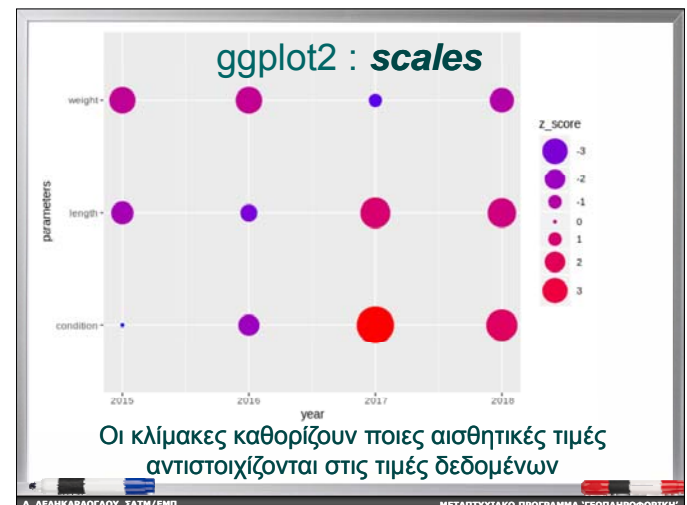
```
define stat_* () και το όρισμα geom μέσα σε αυτή τη συνάρτηση
ggplot(input_data, aes (col1, col2)) + geom_point () +
 stat_summary (geom = "point" , fun.y =)
```

```
define geom_* () και το όρισμα stat μέσα σε αυτή τη συνάρτηση
ggplot(input_data, aes (col1, col2)) + geom_point () +
 geom_point (stat = "summary" , fun.y = "mean")
```



## ggplot2 : **scales**

- Με την αισθητική αντιστοίχιση (δηλ., με τη χρήση της `aes()`) δεν καθορίζεται *πως* θα συμβεί μια μεταβλητή ενδιαφέροντος να αντιστοιχηθεί σε μια αισθητική.
  - Π.χ., κατά την αντιστοίχιση μιας μεταβλητής σε *σχήμα*, `aes(shape = x)`, δεν προσδιορίζεται *ποια* σχήματα θα πρέπει να χρησιμοποιηθούν. Ομοίως, `aes(color = z)` δεν λέει *τι* χρώματα πρέπει να χρησιμοποιηθούν.
- Η περιγραφή *ποια* χρώματα/σχήματα/μεγέθη κ.λπ. θα χρησιμοποιηθούν στις γραφικές αναπαραστάσεις γίνεται με αντίστοιχες **κλίμακες**



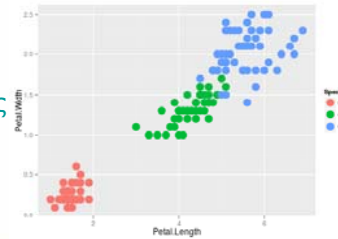
Οι κλίμακες καθορίζουν ποιες αισθητικές τιμές αντιστοιχίζονται στις τιμές δεδομένων

# ggplot2 : scales

- scale\_alpha**  
Alpha scale for continuous variable
- scale\_continuous**  
Continuous position scale
- scale\_datetime**  
Position scale, date time
- scale\_gradient**  
Smooth gradient between two colours
- scale\_gradientn**  
Smooth gradient between n colours
- scale\_hue**  
Qualitative colour scale with evenly spaced hues
- scale\_linetype**  
Scale for line patterns
- scale\_shape**  
Scale for shapes, aka glyphs
- scale\_brewer**  
Sequential, diverging and qualitative colour scales from colorbrewer.org
- scale\_date**  
4/10/1972: Position scale, date
- scale\_discrete**  
Discrete position scale
- scale\_gradient2**  
Smooth gradient between three colours (high, low and midpoints)
- scale\_identity**  
Sequential grey colour scale
- scale\_manual**  
Create your own discrete scale
- scale\_size**  
Size scale for continuous variable
- scale\_shape**  
Scale for shapes, aka glyphs

• **Περιγραφή:** Οι κλίμακες παρέχουν τα εργαλεία που μας επιτρέπουν να κατανοήσουμε το γράφημα: τους άξονες και τις λεζάντες (μπορούμε να προσαρμόσουμε τους τίτλους των αξόνων, τις ετικέτες και τις θέσεις τους).

• Το ggplot2 δημιουργεί αυτόματα προεπιλεγμένες κλίμακες για κάθε αισθητική που ορίζουμε. Ωστόσο, για να προσαρμόσουμε διάφορες κλίμακες μπορούμε να τροποποιήσουμε κάθε στοιχείο κλίμακας.

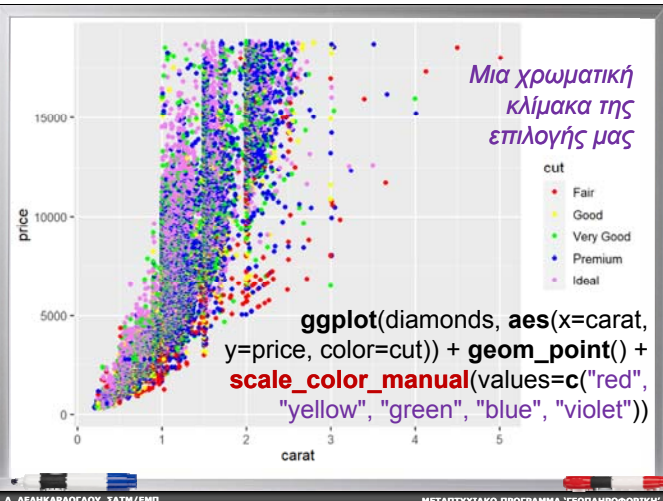
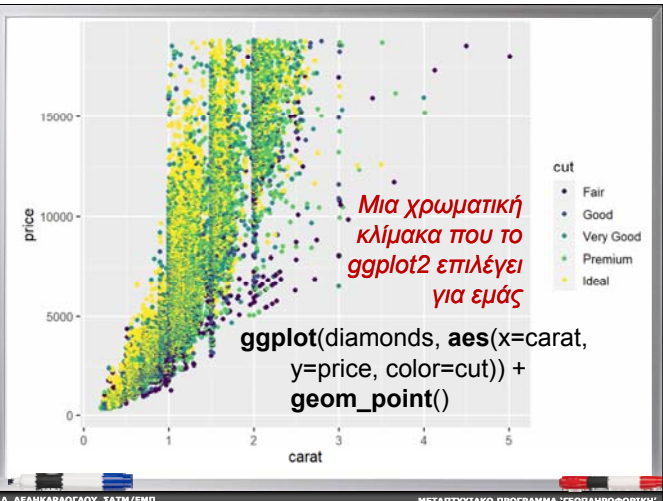


scale\_x\_continuous(Names=,Breaks=,Labels= ...)

**Βασική σύνταξη για τις κλίμακες**

• Δεν υπάρχουν ειδικοί κανόνες – απλά πρέπει να επιλεγεί η κατάλληλη συνάρτηση. Η σύνταξη για τις κλίμακες είναι λίγο πιο περίπλοκη επειδή για κάθε κλίμακα αισθητικής πρέπει να γνωρίζουμε το όνομα της αισθητικής (x, y, χρώμα, μέγεθος, σχήμα), τον τύπο μιας μεταβλητής (συνεχής ή διακριτή) και τις παραμέτρους για κάθε συνάρτηση καθορισμού κλίμακας αισθητικής.

| Scale *                                        | Description                                                                                                                       | Arguments                                                                                                                                                                                                                                                                  |
|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| scale_x_continuous()                           | Map continuous values to visual values. It can be used for any aesthetic: alpha, color, fill, linetype, shape, size.              | Name – title name of axis or legend<br>breaks, labels – define marks on axis<br>na.value – missing values can be replaced with special value<br>position – left or right for vertical axis, top or bottom for horizontal axis<br>limits – define limits on axis (min, max) |
| scale_x_discrete()                             | map discrete values to visual values. It can be used for any aesthetic: alpha, color, fill, linetype, shape, size                 | Same arguments as in scale_x_continuous                                                                                                                                                                                                                                    |
| scale_x_identity()                             | use data values as visual values. It can be used for any aesthetic: alpha, color, fill, linetype, shape, size                     | Usually without arguments                                                                                                                                                                                                                                                  |
| scale_x_manual()                               | map discrete values to manually chosen visual values. It can be used for any aesthetic: alpha, color, fill, linetype, shape, size | values – define color for each category<br>breaks – define which category you want to see<br>limits – define which category you want to see on the legend box                                                                                                              |
| scale_fill_brewer(), scale_colour_brewer()     | Scale used in a combination with fill/color aesthetics. Predefined color scheme from ColorBrewer.                                 | Type – One of seq (sequential), qual (qualitative) or qual (qualitative) string, with use that named palette. If a number, will index into the list of palettes of appropriate type<br>direction – defines order of colors on the scale                                    |
| scale_fill_grey(), scale_colour_grey()         | Scale used in a combination with fill/color aesthetics. Gray scale will represent data on the graph                               | start, gray value at low end of palette<br>end, gray value at high end of palette<br>na.value Colour to use for missing values                                                                                                                                             |
| scale_fill_gradient(), scale_colour_gradient() | Scale used in a combination with fill/color aesthetics (two colors, gradient (low-high) is created)                               | name, limits, breaks, labels, na.values as in other scales<br>low, high - Colours for low and high ends of the gradient.                                                                                                                                                   |
| scale_x_log10(...)                             | Plot x on log10 scale                                                                                                             |                                                                                                                                                                                                                                                                            |
| scale_x_date(...)                              | Plot values as dates.                                                                                                             | Labels – how dates will be visible on axis<br>Breaks – which dates will be visible on axis                                                                                                                                                                                 |

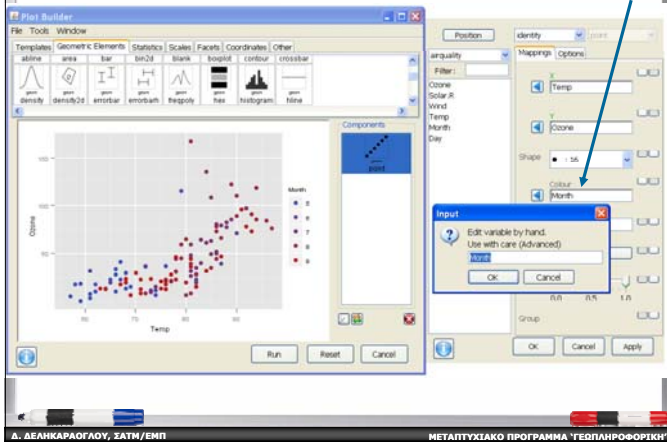


π.χ., αν θέλουμε να αναδείξουμε τα δεδομένα κάθε μήνα

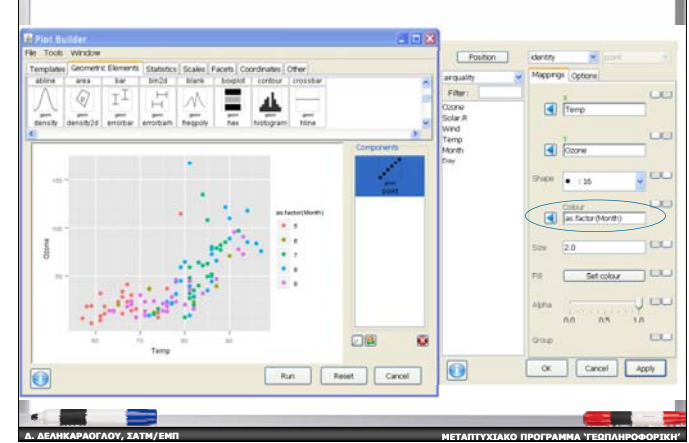
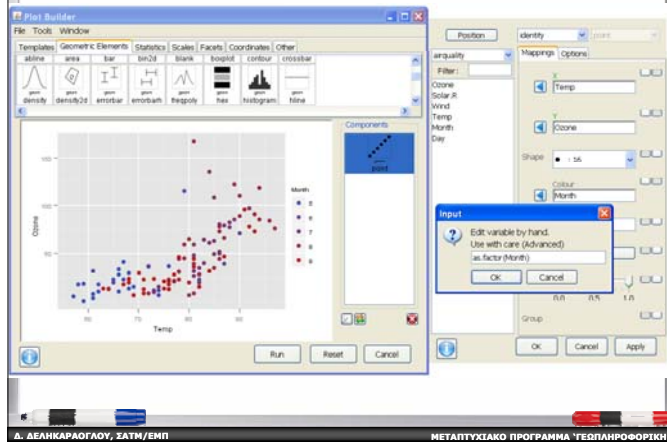
By default – συνεχείς μεταβλητές αντιστοιχίζονται σε χρωματικές αποχρώσεις



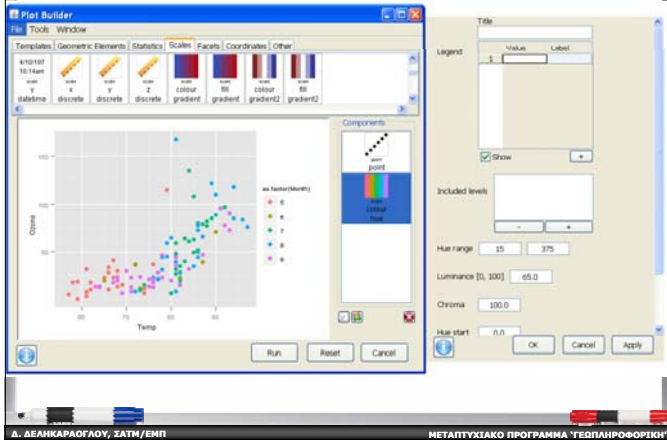
right-click & edit



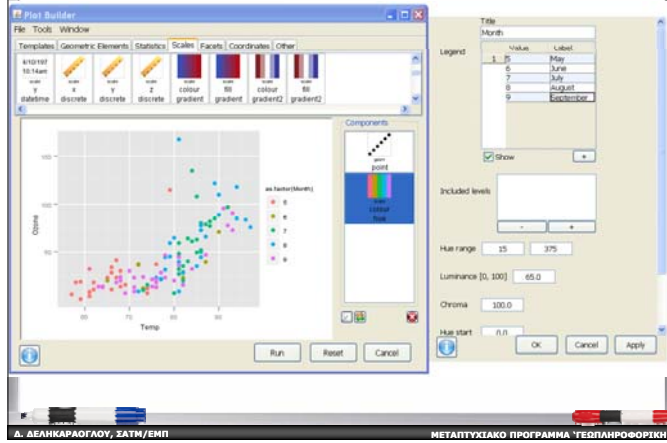
το R αποθηκεύει κατηγορικές μεταβλητές ως παράγοντες



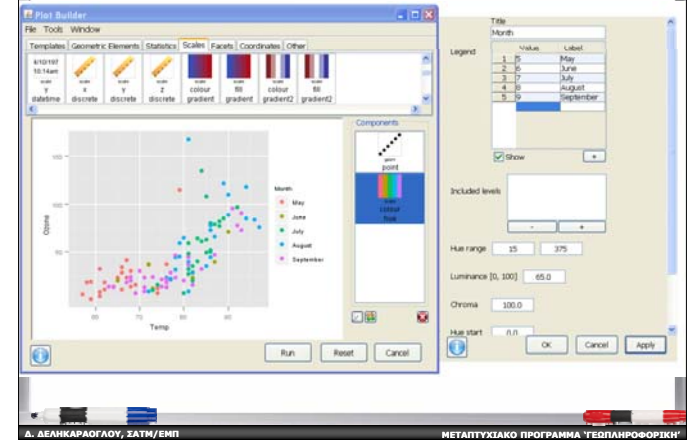
color scale for discrete vars – "colour hue"



Menus allow to fix the title and specify meaningful labels

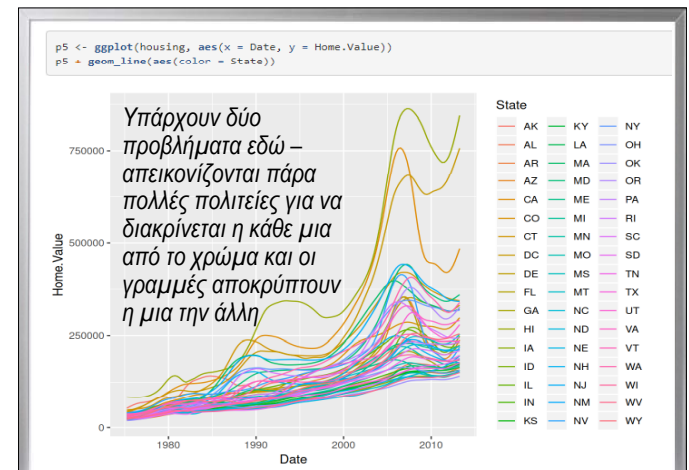
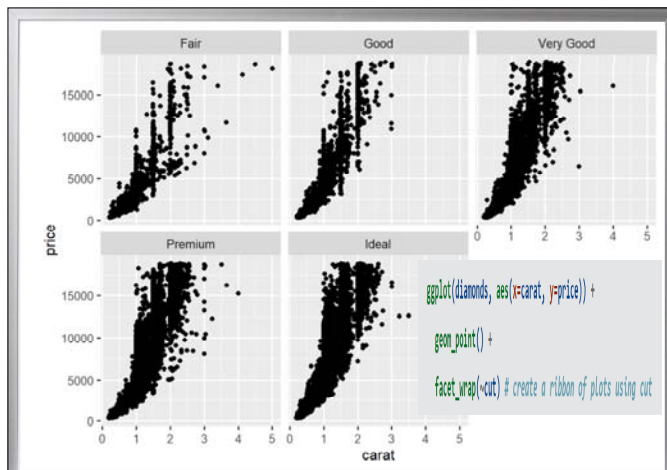


Τελικό γράφημα !!



## ggplot2 : faceting

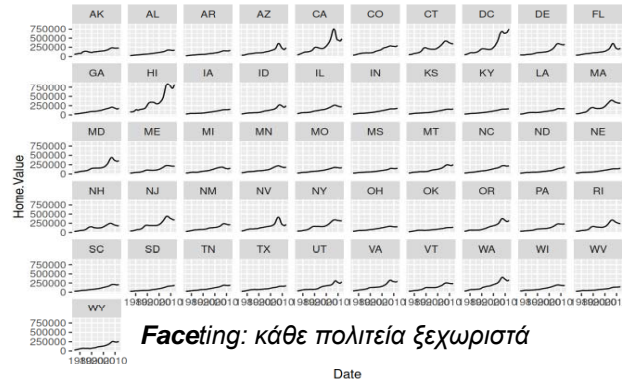
- Υποδηλώνει τη διαδικασία δημιουργίας ξεχωριστών γραφημάτων για μικρά υποσύνολα των δεδομένων
- Το ggplot2 προσφέρει δύο συναρτήσεις για τη δημιουργία μικρών υποσυνόλων:
  - `facet_wrap()`: ορίζουν υποσύνολα ως τα επίπεδα ομαδοποίησης μιας μόνο μεταβλητής
  - `facet_grid()`: ορίζει υποσύνολα ως την επικάλυψη ομαδοποίησης δύο μεταβλητών
- Διευκολύνει τη σύγκριση ανάμεσα σε γραφήματα, όχι μόνο ως γεωμετρικά σχήματα σε γραφήματα



## Χρήση “Θεμάτων”

- Τα **Θέματα** ελέγχουν στοιχεία του γραφήματος που δεν σχετίζονται με τα δεδομένα, όπως:
  - **χρώμα του φόντου**
  - **μέγεθος γραμματοσειρών**
  - **γραμμές πλέγματος**
  - **χρώμα των ετικετών**
- Για να τα τροποποιήσουμε, χρησιμοποιούμε τη συνάρτηση `theme()`, η οποία έχει μεγάλο αριθμό ορισμάτων, τα **στοιχεία θέματος**.

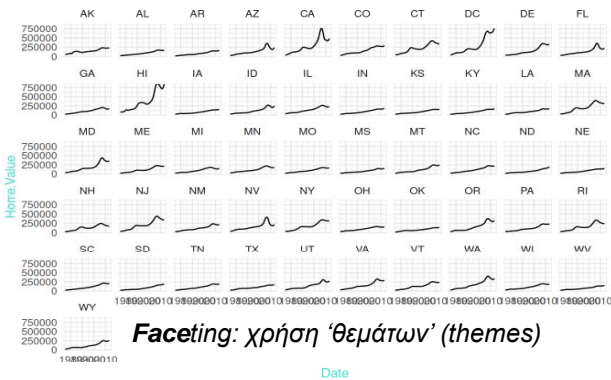
```
(p5 <- p5 + geom_line() +
 facet_wrap(~State, ncol = 10))
```



```
p5 + theme_linedraw()
```



```
p5 + theme_minimal() +
 theme(text = element_text(color = "turquoise"))
```



| Theme name                   | Description                      |
|------------------------------|----------------------------------|
| <code>theme_bw()</code>      | White background with grid lines |
| <code>theme_grey()</code>    | Grey background (default theme)  |
| <code>theme_classic()</code> | White background, no gridlines   |
| <code>theme_minimal()</code> | Minimal theme                    |

Partic function name

Specific arguments

Theme (title = element\_text(...), axis.title.x=element\_text(...), axis.title.y=element\_text(...), legend.background=element\_rect(...), legend.text=element\_text(...), ...)

Each argument is defined via element functions

| Element function             | Description                          | Used with arguments (several examples)                                                           |
|------------------------------|--------------------------------------|--------------------------------------------------------------------------------------------------|
| <code>element_blank()</code> | draws nothing, and assigns no space. | axis.title.x, axis.title.y, plot.title, ...                                                      |
| <code>element_rect()</code>  | borders and backgrounds.             | plot.background, legend.background, panel.background, legend.key, ...                            |
| <code>element_line()</code>  | lines                                | panel.grid.minor.x, panel.grid.minor.y, axis.line.x, axis.line.y, ...                            |
| <code>element_text()</code>  | text                                 | axis.title.x, axis.title.y, axis.text.x, axis.text.y, legend.text, legend.title, plot.title, ... |

```
ggplot(data_set_name, aes(col1,col2)) + geom_point() +
 theme_bw() +
 # panel background is used with element_rect()
 theme(panel.background = element_rect(fill = "white", colour = "grey"))
```

## Συμπερασματικά

### ggplot2 και η γραμματική των γραφικών

- Μπορεί να φαίνονται σύνθετα στην αρχή, επειδή υπάρχουν πολλοί κανόνες και λεπτομέρειες που πρέπει να κατανοηθούν.
- Καταρχήν πρέπει να κατανοηθεί κάθε γραφικό συστατικό ξεχωριστά - έννοια, σύνταξη και κανόνες για το καθένα από αυτά ανεξάρτητα
- Στη συνέχεια, πως να συνδυαστούν σωστά αυτά τα στοιχεία σε μια ενιαία οντότητα → το γράφημα
  - Υπάρχει πολλή θεωρία πίσω από τη σκηνή, αλλά με εξάσκηση σχεδόν κάθε ιδέα απεικόνισης μπορεί να μεταφερθεί με ακρίβεια στην οθόνη

## Επόμενες ενότητες ...

**Πιθανοθεωρητικές συναρτήσεις,  
Στατιστικά Μέτρα και Διαστήματα Εμπιστοσύνης  
για στατιστικές αναλύσεις δεδομένων στο R**

