

15/11/2022

Εθνικό Μετσόβιο Πολυτεχνείο



Σχολή Αγρονόμων & Τοπογράφων Μηχανικών

Αναλυτικές Μέθοδοι στη Γεωπληροφορική Χρήση πακέτων στο R

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Σύνδεση ... με τα προηγούμενα

- **Πακέτα** (*packages*) στο R
- Τι ακριβώς είναι; Πως είναι διαθέσιμα;
- Ποιες δυνατότητες επιλογής έχει ο χρήστης;
- Πως χρησιμοποιούνται;
 - Εγκατάσταση διαθέσιμων πακέτων
 - Χρήση και προβολή εγκαταστημένων πακέτων
 - Ενημέρωση και ανανέωση πακέτων
 - Δημιουργία και διάθεση νέων πακέτων



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Τι είναι τα πακέτα και γιατί πρέπει να ενσωματώσετε τη χρήση τους στην εμπειρία σας στο R;

- Γενικά - **Πακέτα | Packages**: = μια συλλογή επαναχρησιμοποιήσιμου, καλά ελεγμένου λειτουργικού κώδικα εντολών R και δειγμάτων δεδομένων που έχουν αναπτυχθεί κυρίως από την κοινότητα των χρηστών του R.
- Αυξάνουν τη δύναμη του R βελτιώνοντας τις υπάρχουσες βασικές λειτουργίες του ή προσθέτοντας νέες.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Τα πακέτα είναι θεμελιώδεις μονάδες κοινόχρηστου μεταγλωττισμένου κώδικα σε μια καλά καθορισμένη μορφή (όχι απαραίτητα μόνο στη γλώσσα R), συμπεριλαμβανομένων ποικίλων συναρτήσεων R, περιγραφές του τρόπου χρήσης τους και δείγματα δεδομένων για τεκμηρίωση, δοκιμές και εξοικείωση.

Packages in R

- Εφαρμόζουν την φιλοσοφία του R: **anything that can be automated, should be automated.**



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Η ομαδοποίηση ανοικτού κώδικα σε ένα πακέτο που διανέμεται ελεύθερα διευκολύνει τη χρήση του από άλλους χρήστες → Τα πακέτα R παρέχουν έναν βολικό και τυποποιημένο μηχανισμό για τη διανομή του κώδικα R σε ένα ευρύ κοινό
 - Εξοικονόμηση χρόνου για την ανάπτυξη νέου κώδικα και εφαρμογών που αντιπροσωπεύουν έναν συγκεκριμένο σκοπό ανάλυσης δεδομένων



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

CRAN και R πακέτα

- Τα πακέτα R και ο διαδικτυακός κόμβος | αποθετήριο CRAN (*Comprehensive R Archive Network*) συνιστούν βασικές συνιστώσες του R
 - Παρέχουν έναν απλό τρόπο διανομής και τεκμηρίωσης της συνολικής λειτουργίας επαναχρησιμοποιήσιμου κώδικα R
 - Τα πακέτα στο CRAN κατασκευάζονται, εγκαθίστανται και δοκιμάζονται τακτικά σε πολλαπλά συστήματα (*Windows, Linux, Mac,...*) → είναι εγγυημένα για την εύρυθμη λειτουργία τους

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Ο πυρήνας του R + πακέτα;

- Η βασική διανομή του R περιλαμβάνει τον πυρήνα (*core*) του λογισμικού που επίσης περιέχει πλήθος πακέτων (*packages*) και βιβλιοθηκών (*libraries*)
- Ο κεντρικός πυρήνας του R περιέχει ότι απαιτείται ώστε ο χρήστης να μπορεί να εκτελέσει ένα μεγάλο αριθμό στατιστικών διαδικασιών ευρείας χρήσης:
 - *linear and generalized linear models*
 - *nonlinear regression models*
 - *time series analysis*
 - *parametric and nonparametric tests*
 - *clustering, smoothing, ...*

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Με την εγκατάσταση του R, από προεπιλογή ενεργοποιούνται αυτόματα ~30 πακέτα που έχουν συνήθως κάποιο σημείο κοινής αναφοράς, όπως:
 - να επιτρέπουν τη βασική λειτουργικότητα του R, την εκτέλεση τυπικών συναρτήσεων διαχείρισης, επεξεργασίας και ανάλυσης δεδομένων, τη δημιουργία γραφικών, ...

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Όλα τα πακέτα στο R, πληρούν σχετικά αυστηρές προδιαγραφές και συμβάσεις προτυποποίησης που στοχεύουν να μεγιστοποιηθεί η χρηστικότητα και η μακροπρόθεσμη σταθερότητα των πακέτων για τους τελικούς χρήστες
 - Τα πακέτα υπόκεινται σε ποικίλες δοκιμές διασφάλισης της λειτουργικότητά τους, καθώς και σε συγκεκριμένες διαδικασίες υποβολής και επανεξέτασης / ανανέωσης
- Αποθηκεύονται σε ειδικά "αποθετήρια" (*repositories*), από όπου, συνήθως διαδικτυακά, είναι εύκολα προσβάσιμα από όλους τους εν δυνάμει χρήστες τους.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

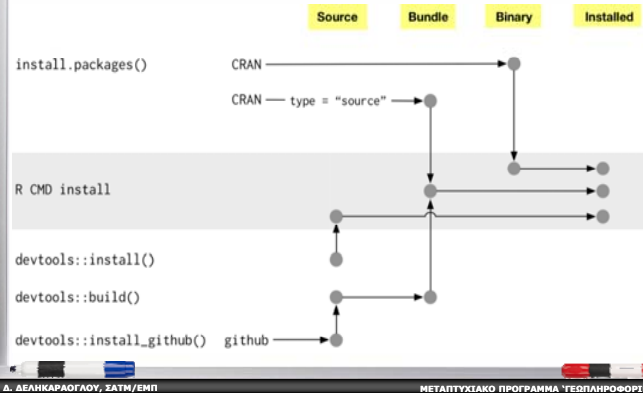
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

Αποθετήρια για πακέτα R

Ανάμεσα στα πιο δημοφιλή συγκαταλέγονται:

- **CRAN**: Ένα ολοκληρωμένο δίκτυο ftp και διακομιστών ιστού που διατηρούνται από την κοινότητα R σε όλο τον κόσμο.
- **Github**: Παγκοσμίως η μεγαλύτερη και πιο προηγμένη πλατφόρμα ανάπτυξης λογισμικού ανοικτού κώδικα που επιτρέπει τη συνεργασία ομάδων προγραμματιστών σε πραγματικό χρόνο !!!
- **Bioconductor**: Ένα ειδικό αποθετήριο για πακέτα που προορίζονται κυρίως για εφαρμογές Βιοπληροφορικής.

Καταστάσεις πακέτων στο R



Διαφορές: πακέτα και βιβλιοθήκες

- Δυστυχώς, στον κόσμο του R, θα συναντήσετε συχνά την μπερδεμένη χρήση των λέξεων "πακέτο | package" και "βιβλιοθήκη | library".
- Η ορολογία του R αναμφισβήτητα έρχεται σε αντίθεση με τις ευρύτερες συμβάσεις προγραμματισμού, όπου η συνήθης έννοια της "βιβλιοθήκης" είναι πιο κοντά σε αυτό που εννοούμε ως "πακέτο" → αυτό το σφάλμα λεξιλογίου είναι συχνά ακίνδυνο
- Ωστόσο, η διάκριση μεταξύ των δύο είναι σημαντική και χρήσιμη στις διεργασίες ανάπτυξης πακέτων

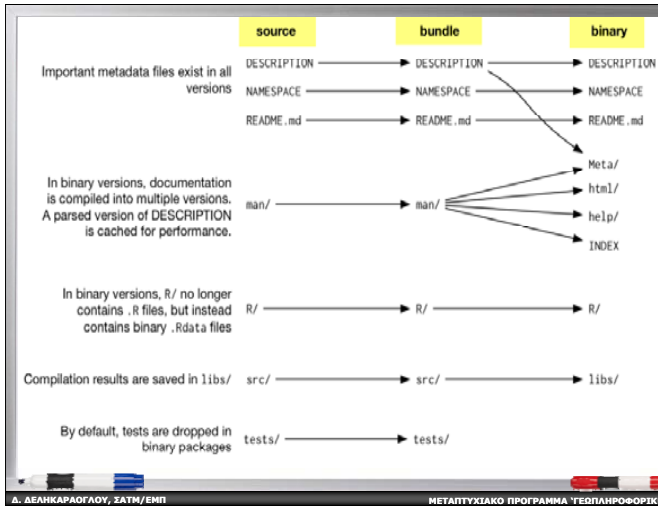
- Στο R, ένα πακέτο είναι μια συλλογή από συναρτήσεις, δεδομένα και μεταγλωττισμένο κώδικα R.

Τα πακέτα μπορούν να διανεμηθούν ως

- ✓ **SOURCE** (αρχεία πηγαίου κώδικα → ένας κατάλογος αρχείων με συγκεκριμένη δομή, π.χ., αρχεία .R, DESCRIPTION,...),
- ✓ **BINARY** (αρχεία σε συγκεκριμένη μορφή (.tar, ...) για το εκάστοτε λειτουργικό σύστημα) ή
- ✓ **BUNDLE** (συμπίεσμένο αρχείο (.gz, ...) που περιέχει στοιχεία του πακέτου, παρόμοια με την πηγή).

- Στο R, η τοποθεσία όπου αποθηκεύονται τα πακέτα ονομάζεται βιβλιοθήκη (library) → ένας κατάλογος που περιέχει εγκατεστημένα πακέτα, ... κάτι σαν μια βιβλιοθήκη για βιβλία

- Εάν υπάρχει μια συγκεκριμένη λειτουργικότητα που χρειάζεστε για μια ανάλυση δεδομένων ...
 - Μπορείτε να κάνετε λήψη του πακέτου από την κατάλληλη τοποθεσία όπου έχει αποθηκευτεί τοπικά στον Η/Υ σας και να είναι διαθέσιμο για άμεση χρήση
 - Στη συνέχεια, απλώς καλείτε τις κατάλληλες συναρτήσεις του πακέτου κ.λπ.

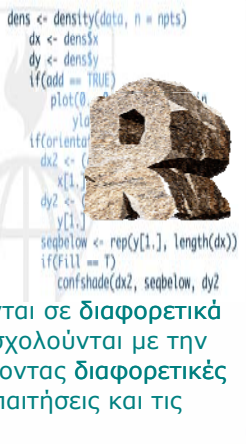


- Όλα τα επιπλέον διαθέσιμα πακέτα και οι βιβλιοθήκες του R περιέχουν άλλες συναρτήσεις που δεν είναι διαθέσιμες στον κεντρικό πυρήνα του R

- Για αυτά απαιτείται ξεχωριστή διαδικασία ενεργοποίησης τους προκειμένου να επεκταθούν οι δυνατότητες του βασικού πυρήνα του R (ή/και κάποιου ολοκληρωμένου περιβάλλοντος χρήσης του, π.χ. το RStudio)

- Η εύκολη πρόσβαση και ευρεία διάθεση πακέτων που αναπτύσσονται από έμπειρους προγραμματιστές είναι ένας από τους κύριους λόγους για την προώθηση της ευρείας αποδοχής της γλώσσας R (και του R) στην Επιστήμη των Δεδομένων

- Τα πακέτα του R απευθύνονται σε διαφορετικά πεδία και εφαρμογές που ασχολούνται με την ανάλυση δεδομένων, παρέχοντας διαφορετικές δυνατότητες με βάση τις απαιτήσεις και τις προτιμήσεις των χρηστών



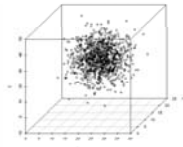
Πληθώρα διαθέσιμων πακέτων

- Είναι ένα τεράστιο πλεονέκτημα / R's strong package ecosystem
 - διαθέσιμες νέες τεχνικές, χωρίς καθυστέρηση, οι οποίες μπορούν να υλοποιηθούν με τη χρήση της γλώσσας R
- Η ελεύθερη διάθεση τους επιτρέπει στους χρήστες, να οικοδομήσουν πολυσύνθετα προγράμματα κατάλληλα για τις δικές τους ανάγκες → build own tools and methods for analyzing data

EXAMPLE :
 # If we want to draw a 3d scatterplot, we need the 3d scatterplot package,
 # and if this is not there in the library then you can't run the commands
 x <- rnorm(1000, mean=20, sd=5)
 y <- rnorm(1000, mean=5, sd=3)
 z <- rnorm(1000, mean=25, sd=8)
 scatterplot3d(x, y, z) # we will get an error (for not finding the function)

Hence, we have to install that particular package
 install.packages("scatterplot3d", repos = "cran.rstudio.com")
 library(scatterplot3d) # this attaches it to the core packages
 scatterplot3d(x, y, z)

Voilà ... το
 ζητούμενο
 3d scatterplot



- **Σχετικό μειονέκτημα** = καθώς ο αριθμός των πακέτων του R μεγαλώνει ταχύτατα, γίνεται δύσκολο να επιλεγεί το καλύτερο πακέτο για τις εκάστοτε ανάγκες, και ο αυστηρότερος ποιοτικός έλεγχος των διαθέσιμων πακέτων επαφίεται στο χρήστη.
- Η φόρτωση και η προσάρτηση πολλών πακέτων ενίοτε οδηγεί σε συγκρούσεις με τις ονοματολογίες πανομοιότυπων μεταβλητών
- Η φόρτωση κάποιων πακέτων μπορεί να είναι αργή → ένας λόγος για να φορτώνετε μόνο τα πακέτα που χρειάζεστε κάθε φορά

Install packages using the automated process from within R

Από το μενού στο R

Λίστα επιλογής προεπιλεγμένου

ΒΑΣΙΚΕΣ ΕΝΤΟΛΕΣ για τη χρήση 'πακέτων' στο R
 install.packages("package") # Εγκατάσταση ενός πακέτου με την ονομασία του
 # Εγκατάσταση περισσότερων από ένα πακέτο τη φορά,
 # ως διάνυσμα χαρακτήρων στο πρώτο όρισμα της συνάρτησης install.packages
 install.packages(c("ggplot2", "MASS"))
 # Εγκατάσταση από κάποιο mirror site του κεντρικού αποθετηρίου CRAN
 install.packages("ggplot2", repo = "https://lib.ugent.be/CRAN/")
 # Επισκόπηση των συναρτήσεων και δεδομένων που περιέχονται σε ένα πακέτο
 help(package = "ggplot2")
 remove.packages("ggplot2") # Απεγκατάσταση ενός πακέτου
 old.packages() # Ελέγξτε ποια πακέτα χρειάζονται ενημέρωση
 update.packages() # Ενημερώστε όλα τα πακέτα (ή κάποια από αυτά)
 # Για σποραδική χρήση μερικών συναρτήσεων ή δεδομένων ενός πακέτου
 packagename::functionname() # π.χ., babynames::births
 library(ggplot2) # Φόρτωση πακέτου σε μια συνεδρία R για χρήση
 library() # Η εντολή χωρίς ορίσματα παρέχει τη λίστα των πακέτων στον Η/Υ σας

You can install from the R command line, like this:
 install.packages("NameOfTheLibraryToInstall", dependencies = TRUE)
 library(NameOfTheLibraryToLoad)
 or
 require(NameOfTheLibraryToLoad)
 Παραδείγματα:
 install.packages("stats")
 library(stats)
 install.views("stats")
 ...
 library(glsme) # for applying the Generalized Least Squares Method
 library(lmtest) # for testing Linear Regression Models

«Βιβλιοθήκη» (library) είναι η τοποθεσία όπου αποθηκεύονται τα πακέτα. Η εντολή library(package) καθιστά το πακέτο διαθέσιμο προς χρήση.

Για να γίνει χρήση μιας συνάρτησης ή δεδομένων από ένα πακέτο, αυτό πρέπει προηγουμένως να φορτωθεί στο R. Οι σχετικές εντολές (στην πλήρη μορφή τους) είναι ...
 library(package, help, pos = 2, lib.loc = NULL, character.only = FALSE, logical.return = FALSE, warn.conflicts = TRUE, quietly = FALSE, verbose = getOption("verbose"))
 Χρήση της εντολής library() συνιστάται δεδομένου ότι αυτή δίνει ξεκάθαρα και νωρίς προειδοποιητικά μηνύματα (εάν το πακέτο δεν είναι εγκατεστημένο ή δεν μπορεί να φορτωθεί)

Για να γίνει χρήση μιας συνάρτησης ή δεδομένων από ένα πακέτο, αυτό πρέπει προηγουμένως να φορτωθεί στο R. Οι σχετικές εντολές είναι ...
 Αντίθετα, η εντολή require() δεν δημιουργεί άμεσα κάποιο σφάλμα -απλά δίνει μια ένδειξη FALSE και ο κώδικας συνήθως αποτυγχάνει αργότερα ...
 require(package, lib.loc = NULL, quietly = FALSE, warn.conflicts = TRUE, character.only = FALSE)

Χρήσιμες εντολές για τη διαχείριση πακέτων
 library() # list all available packages in current session
 library(lib.loc = .Library) # list all packages in the default library
 library(help = anypack) # seek help on package 'anypack'
 help(package=anypack) # ditto, help on package 'anypack'
 library(anypack) # attach package 'anypack'
 require(anypack) # does the same
 search() # "anypack", too
 detach("package:anypack")
 # if the package name is in a character vector, use
 pkg <- "anypack"
 library(pkg, character.only = TRUE)
 detach(pos = match(paste("package", pkg, sep = ":"), search()))
 require(pkg, character.only = TRUE)
 detach(pos = match(paste("package", pkg, sep = ":"), search()))
 require(nonexistent) # FALSE

Μερικά από τα πλέον χρησιμοποιούμενα/χρήσιμα διαθέσιμα πακέτα ...
 To load data
 RODBC, RMySQL, RPostgreSQL, RSQLite # To read in data from a database
 XLConnect, xlsx # To read and write MS Excel files from/in R
 foreign # To read from other programs, e.g. SAS, SPSS data sets
 To manipulate data
 dplyr # Essential shortcuts for subsetting, summarizing, rearranging, and joining together data sets
 tidyr # Tools for changing the layout of your data sets, e.g. to convert data into the tidy format, that R likes best.
 stringr # Easy to learn tools for regular expressions and character strings.
 lubridate # Tools that make working with dates and times easier.
 datasets # R's comprehensive ready to load and use datasets
 To visualize data
 ggplot2 # For making graphics, and layered, customizable plots.
 ggviz # For interactive, web based graphics.
 rgl # For interactive 3D visualizations with R
 htmlwidgets # For interactive (javascript based) visualizations with R.
 googleVis # Let's you use Google Chart tools to visualize data in R.

Μερικά από τα πλέον χρησιμοποιούμενα/χρήσιμα διαθέσιμα πακέτα ...

```
To model data
car      # For Anova statistical analyses.
mgcv     # Generalized Additive Models
lme4/nlme # Linear and Non-linear mixed effects models
randomForest # Random forest methods from machine learning
multcomp # Tools for multiple comparison testing
vcd      # Visualization tools and tests for categorical data
survival # Tools for survival analysis
caret    # Tools for training regression and classification models

For spatial analysis
sp, maptools # Tools for loading and using spatial data (e.g. shapefiles).
maps         # Easy to use map polygons for plots.
ggmap        # Download street maps straight from Google maps.

For Time Series and Financial data
zoo          # Formatting and saving time series objects in R.
xts         # Flexible tools for manipulating time series data sets.
quantmod    # Tools for downloading financial data, plotting common charts,
            # and doing technical analysis.
```

Η εντολή **installed.packages()** παρέχει ένα πίνακα πληροφοριών με το όνομα των πακέτων ως 'ονόματα σειρών' και τα πεδία περιγραφής τους ως 'ονόματα στηλών'. Οι πληροφορίες περιέχονται σε ένα DESCRIPTION file

```
packages = installed.packages()
rownames(packages) # to see all installed packages
[1] "abind" "ade4" "adegenet"
...

colnames(packages) # to see all fields
[1] "Package" "LibPath" "Version" "Priority"
"Depends" "Imports"
[7] "LinkingTo" "Suggests" "Enhances"
"OS_type" "License" "Archs"
[13] "Built"

packages["sp",c("Package","Version","License")]
Package Version License
"sp" "0.9-72" "GPL (>= 2)"
```

```
packageDescription(pkg, lib.loc = NULL, fields =
NULL, drop = TRUE, encoding = "")
packageVersion(pkg, lib.loc = NULL)
```

Examples

```
packageDescription("stats")
packageDescription("stats", fields = c("Package",
"Version"))
```

```
packageDescription("stats", fields = "Version")
packageDescription("stats", fields = "Version",
drop = FALSE)
```

```
if(packageVersion("MASS") < "7.3.29")
message("you need to update 'MASS'")
```

Ο πυρήνας του R + πακέτα;

```
# Για να εξακριβωθεί ποια πακέτα έχουν φορτωθεί στο R,
# η σχετική εντολή είναι ...
installed.packages(lib.loc = NULL, priority = NULL,
noCache = FALSE, fields = NULL,
subarch = .Platform$arch)
```

Τι υποδηλώνουν οι διαφορετικές διαβαθμίσεις της παραμέτρου 'priority'?

```
# installed.packages scans the 'DESCRIPTION' files of each package
# found along lib.loc and returns a matrix of package names,
# library paths and version numbers
```

	Package	LibPath	Version	Priority
base	"base"	"C:/PROGRA~1/r/R-301~1.1/library"	"2.11.1"	"base"
boot	"boot"	"C:/PROGRA~1/r/R-301~1.1/library"	"1.2-42"	"recommended"
car	"car"	"C:/PROGRA~1/r/R-301~1.1/library"	"2.0-2"	NA
class	"class"	"C:/PROGRA~1/r/R-301~1.1/library"	"7.3-2"	"recommended"
cluster	"cluster"	"C:/PROGRA~1/r/R-301~1.1/library"	"1.12.3"	"recommended"

```
# installed.packages ... also returns of package dependencies,
# imported packages and their links, ... licence type, and version
```

	Depends	Imports	LinkingTo
base	NA	NA	NA
boot	"R (>= 2.9.0), graphics, stats"	NA	NA
car	"R (>= 2.1.1), stats, graphics, MASS, nnet, survival"	NA	NA
class	"R (>= 2.5.0), stats, utils"	"MASS"	NA
cluster	"R (>= 2.9.0), stats, graphics, utils"	NA	NA

	Suggests	Enhances	OS_type	License	Built
base	NA	NA	NA	"Part of R 2.11.1"	"2.11.1"
boot	"survival"	NA	NA	"Unlimited"	"2.11.1"
car	"alr3, leaps, lmtree, sandwich, mgcv, rgl"	NA	NA	"GPL (>= 2)"	"2.11.1"
class	NA	NA	NA	"GPL-2 GPL-3"	"2.11.1"
cluster	NA	NA	NA	"GPL (>= 2)"	"2.11.1"

Install Packages from Repositories or Local Files

```
install.packages(pkgs, lib, repos = getOption("repos"),
contriburl = contrib.url(repos, type),
method, available = NULL, destdir = NULL,
dependencies = NA, type = getOption("pkgType"),
configure.args = getOption("configure.args"),
configure.vars = getOption("configure.vars"),
clean = FALSE, Ncpus = getOption("Ncpus", 1L),
verbose = getOption("verbose"),
libs_only = FALSE, INSTALL_opts, quiet = FALSE,
keep_outputs = FALSE, ...)
```

```
a<-installed.packages()
packages<-a[,1]
is.element("boot", packages)
[1] TRUE
```

```
p <- available.packages()
```

```
dim(p)
[1] 2553 12
```

Ο πυρήνας του R + πακέτα;

- Στα διάφορα πακέτα του R περιλαμβάνονται επίσης και μια μεγάλη σειρά από
 - αρθρώματα (modules) και συναρτήσεις που επεκτείνουν το υπάρχον περιβάλλον R με τις γραφικές δυνατότητες που παρέχουν
 - ένα ευέλικτο γραφικό περιβάλλον για τη δημιουργία διαφόρων ειδών παρουσιάσεις εκτεταμένων και πολύμορφων δεδομένων και καθοδηγούν στην διαδικασία της μοντελοποίησης και αξιολόγησης της ανάλυσής τους

Βιβλιοθήκες/πακέτα

```
install.packages('dplyr')
```

Λήψη και εγκατάσταση βιβλιοθηκών από το CRAN.

```
library(dplyr)
```

Φόρτωση βιβλιοθήκης/πακέτου, για να χρησιμοποιήσετε όλες τις συναρτήσεις της.

```
dplyr::select
```

Κλήση συγκεκριμένης συνάρτησης από μια βιβλιοθήκη.

```
data(iris)
```

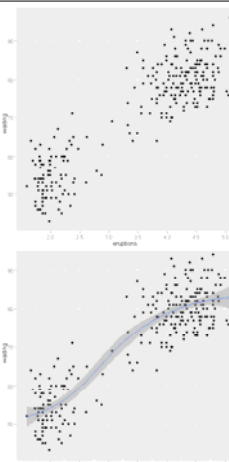
Εισαγωγή ενός ενσωματωμένου συνόλου δεδομένων στο περιβάλλον.

Μερικά ενδιαφέροντα πακέτα του R

- **foreign**: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase, ..
- **ggplot2**: An implementation of the *Grammar of Graphics*
- **GGally**: Extension to ggplot2, ... templates for different plots
- **gdistance**: distances and routes on geographical grids
- **FKF**: Fast Kalman Filter Toolbox
- **gplots**: Various R programming tools for plotting data

Το πακέτο ggplot2

- The fundamental building block of a plot is based on **aesthetics** and **facets**
- Aesthetics are graphical attributes that effect how the data are displayed. Color, Size, Shape
- Facets are subdivisions of graphical data.
- The graph is realized by adding layers, geoms ("point", "smooth", ...), and statistics.



Εγκατάσταση πακέτων του R;

- Για Windows και Mac, υπάρχουν binary αρχεία των πακέτων για εύκολη εγκατάσταση στο R .
 - Οι προκαθορισμένες (default) τιμές είναι οι συνθηέστερες και αν δεν έχετε κάποιο άλλο λόγο, είναι γενικά μια καλή ιδέα να τις αποδεχθείτε.
 - Για εγκατάσταση σε Linux ... *ψάξτε το*
 - Για τα Windows, Mac και Linux, αν έχετε τα κατάλληλα εργαλεία, μπορείτε να δημιουργήσετε το R από τον πηγαίο κώδικα. Οι σχετικές οδηγίες → <http://cran.r-project.org/doc/manuals/R-admin.html>

Εγκατάσταση πακέτων του R;

- `install.packages("packagename")`
- Μπορείτε να προσδιορίσετε **mirror CRAN sites**
`install.packages("packagename", repos = "http://cran.stat.ucla.edu/")`
- `update.packages("packagename")`
- **Code to load or install then load**

```
# Foreign ships with R so no need to install
require(xlsx) || {
  install.packages("xlsx")
  require(xlsx)
}

## [1] TRUE
```

Εγκατάσταση πακέτων του R;

- Για να χρησιμοποιήσετε επιμέρους πακέτα του R , πρέπει πρώτα να τα εγκαταστήσετε χρησιμοποιώντας τη συνάρτηση `install.packages`, η οποία
 - κατεβάζει το ζητούμενο πακέτο από τον ιστότοπο CRAN και
 - το εγκαθιστά για χρήση στον Η/Υ σας

```
#install.packages("ggplot2")
#install.packages("GGally")
```

Εγκατάσταση πακέτων του R;

- Αν θα χρειαστείτε ένα συγκεκριμένο πακέτο για την τρέχουσα χρήση του R, πρέπει να το φορτώσετε χρησιμοποιώντας τη συνάρτηση `require`

> `require(foreign)`

Loading required package: foreign

- Σε μερικές περιπτώσεις, η καλή λειτουργία μιας βιβλιοθήκης εξαρτάται από τα ήδη εγκατεστημένα πακέτα. Για το λόγο αυτό, καλύτερα να χρησιμοποιείτε την επιλογή `dep=TRUE`, έτσι ώστε να γίνεται πλήρη εγκατάσταση όλων των πιθανών εξαρτήσεων.

> `install.packages("foreign", dep=TRUE)`

Εγκατάσταση πακέτων του R;

- Αν απαιτούνται επιπλέον πακέτα για τη χρήση ενός πακέτου, αυτά φορτώνονται ή προσαρτώνται αυτόματα ...

```
require(GGally)

## Loading required package: GGally Loading required package: reshape Loading
## required package: plyr
##
## Attaching package: 'reshape'
##
## The following object is masked from 'package:plyr':
##   rename, round_any
##
## The following object is masked from 'package:reshape2':
##   colsplit, melt, recast
```

Εγκατάσταση πακέτων του R;

```
r <- getOption("repos")
r["CRAN"] <- "http://cran.cc.uoc.gr"
options(repos = r)

pkgInstall <- function(x)
{
  if ( x %in% rownames(installed.packages()) == TRUE )
  {
    cat(paste(x, "already installed\n\n"))
  }
  else
  {
    install.packages(x, dep=TRUE)
  }
}

pkgInstall("timeSeries")
pkgInstall("CircStats")
pkgInstall("frontier")
pkgInstall("forecast")
pkgInstall("quantmod")
```

- Χρήσιμο R script για την εγκατάσταση πακέτων μετά από κάποιο στοιχειώδη έλεγχο μη προηγούμενης εγκατάστασής τους (πηγή: <http://stavrakoudis.econ.uoi.gr/stavrakoudis/?iid=221>)

Εγκατάσταση πακέτων του R; Package.maps

```
> install.packages("maps")
> library("maps")
> map("world")

> map("world2")
```



Εγκατάσταση πακέτων του R;

```
sessionInfo()

## R version 3.0.1 (2013-05-16)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=c
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] grid      stats      graphics  grDevices  utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] vcd 1.2-13      colorspace 1.2-2 MASS 7.3-27      ggally 0.4.3
## [5] reshape 0.8.4  plyr 1.8      ggplot2 0.9.3.1 reshape2 1.2.2
## [9] xlsx 0.5.1     xlsxjars 0.5.0 rJava 0.9-4    foreign 0.8-54
## [13] knitr 1.3
##
## loaded via a namespace (and not attached):
## [1] dichromat 2.0-0 digest 0.6.3 evaluate 0.4.4
## [4] formatR 0.5  grabc 0.1-2 labeling 0.2
## [7] munsell 0.4.2  proto 0.3-10 RColorBrewer 1.0-5
## [10] scales 0.2.3  stringr 0.6.2 tools 3.0.1
```

• Τι χρησιμοποιεί στην τρέχουσα συνεδρία (session) του R

Εγκατάσταση πακέτων του R;

- 1 Open the general help, and look at the "Introduction to R" section.
- 2 Install the *mypack* package (any package of your choice).
- 3 List the functions and datasets available in this package.
- 4 Load the package, and read the description of any of its datasets, e.g. *expldata*.
- 5 Print out the *expldata* dataset (entering the name of an object will list it).
- 6 Run the example that comes with the *expldata* dataset.
- 7 Quit

```
> help.start()
> install.packages("mypack")
> help(package="mypack")
> library(mypack)
> help(expldata)
> expldata
> example(expldata)
> q()
```

← Δοκιμάστε τις ενδεικτικές εντολές ... ως εξάσκηση

```
# A package is a collection or group of objects that R can use.
# A package may contain functions, data frames, or other objects
# (e.g. dynamically loaded libraries of compiled code).
}
# To see which packages have available in current session?
library()
# To load a package?
library("package_name")
# How to see the documentation for a particular package?
library(help="package_name")
help(package="package_name")
# How to see the help file for a specific function?
help("function_name")
?function_name
# How to save your work, in an .RData file?
save(file="d:/file_name.RData")
save.image("d:/file_name.RData")
# You can also access this .RData file through the file menu:
# File, Save workspace, browse to the folder where you want
# to save the file and supply the file name of your choice
# How to retrieve the saved file?
load("d:/file_name.RData")
# You can also access this .RData file through the file menu:
# File, Load workspace, browse to the folder where you saved
# the .RData file and click open
```

Τυπικές εντολές διαχείρισης πακέτων ...

```
# To retrieve the work that has been saved in .Rhistory file ?
loadhistory("d:/file_name.Rhistory")
```

You can also access this .Rhistory file through the file menu:
File, Load history, then browse to the folder where you saved the .Rhistory file and click open

```
# To use a script of commands and functions saved in a text file?
source("d:/file_name.txt")
```

```
# and to get R to echo back the commands and functions in a script
source("d:/file_name.txt", echo=T)
```

```
# To see a list of the objects that are currently available?
```

```
objects()
ls()
# To remove unwanted objects and functions?
rm(object_name1, object_names2, etc.)
rm(function_name1, function_name2, etc.)
```

... Τυπικές εντολές διαχείρισης πακέτων

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Αγρονόμων & Τοπογράφων Μηχανικών
Αναλυτικές Μέθοδοι στη Γεωπληροφορική
Χειρισμός ημερομηνιών και ώρας στο R

Τυπικά παραδείγματα σύγχυσης στις ημερομηνίες και την ώρα

Σκεφτείτε αυτές τις τρεις φαινομενικά απλές ερωτήσεις: Κάθε έτος έχει 365 ημέρες; • Κάθε μέρα έχει 24 ώρες; • Κάθε λεπτό έχει 60 δευτερόλεπτα;

Γενικά, ο χειρισμός ημερομηνιών και ωρών στις αναλύσεις δεδομένων παρουσιάζει ποικίλες προκλήσεις, καθώς αυτά πρέπει να συμβιβάζουν δύο φυσικά φαινόμενα (την περιστροφή της Γης και την τροχιά της γύρω από τον Ήλιο) με μια ολόκληρη σειρά γεωπολιτικών συμβάσεων (π.χ., ζώνες ώρας, θερινή ώρα, ...).

Χρονική σήμανση δεδομένων στο R

- Το R παρέχει πολλές επιλογές για την αντιμετώπιση δεδομένων ημερομηνίας ή/και ώρας (π.χ., χρονοσφραγίδας παρατηρήσεων).

2013-03-13T02:11:47.592883024-0500

Year, Month, Day, Hour, Minute, Second, NanoSecond, Timezone Offset (or Z for UTC)

Τυπικές ανάγκες χρονοσήμανσης δεδομένων στο R

- Πως λαμβάνουμε την τρέχουσα ημερομηνία και ώρα
- Πως χειριζόμαστε διαστήματα, διάρκειες και περιόδους χρόνου
- Πως το R χειρίζεται την ημερομηνία/ώρα εσωτερικά, σε μορφές ISO 8601, ...
- Πως μετατρέπουμε μεταξύ τους τις διάφορες μορφές χρονοσήμανσης (έτος, μήνας, ημέρα, ώρα, λεπτό κ.λπ., θερινή ώρα, ζώνες ώρας, ...)

Προαπαιτούμενα για τον χειρισμό ημερομηνιών / ωρών στο R

- Πρωτίστως απαιτείται το πακέτο **lubridate**
 - Δεν αποτελεί μέρος του βασικού πυρήνα του R γιατί χρειάζεται μόνο όταν αναλύονται δεδομένα με χρονική σήμανση
 - Διευκολύνει για το χειρισμό τριών τύπων δεδομένων ημερομηνίας/ώρας που αναφέρονται σε μια χρονική στιγμή: Μια **ημερομηνία** <date> • Μια **ώρα** μέσα σε μια μέρα <time> • Συνδυαστικά μια ημερομηνία και μια ώρα που προσδιορίζουν μοναδικά μια στιγμή στο χρόνο (συνήθως στο πλησιέστερο δευτερόλεπτο), <dtm>

Σήμανση τρέχουσας ημερομηνίας / ώρας

- → Sys.Date() / today() > Sys.Date() [1] "2022-11-13"
- Sys.time() / now() > today() [1] "2022-11-13"
- am() > Sys.time() [1] "2022-11-13 17:43:39 +03"
- pm() > now() [1] "2022-11-13 17:43:39 +03"
- leap_year()
- Συνήθως, χρονοσημάνσεις στο R αποθηκεύονται ως χαρακτήρες - μεταβλητές κλάσης/κατηγορίας (class) → Απαραίτητο να μετατρέπονται σε μια **κλάση | κατηγορία [Date-Time]**, ώστε το R να μπορεί να τις διαχειριστεί ως αριθμητικά δεδομένα

```
> library("lubridate")# nextto R που διασφαλίζει το χειρισμό ημερομηνιών και ωρών
> my_dt <- c("2012-03-28 11:59:22", "2022-10-03 09:26:50 AM", "2012-03-28 23:59:22")
> my_dt
[1] "2012-03-28 11:59:22" "2022-10-03 09:26:50 AM" "2012-03-28 23:59:22"
> class(my_dt)
[1] "character"
> am(my_dt)
[1] TRUE TRUE FALSE
> pm(my_dt)
[1] FALSE FALSE TRUE
> leap_year(my_dt)
[1] TRUE FALSE TRUE
> ## heterogeneous formats in a single vector:
> x <- c("20220101", "2022-01-02", "2022 01 03", "2022-1-4",
+ "2022-1, 5", "Created on 2022 1 6", "202201 !!! 07")
> ymd(x)
[1] "2022-01-01" "2022-01-02" "2022-01-03" "2022-01-04" "2022-01-05" "2022-01-06"
[7] "2022-01-07"
> y <- c("01012022", "01-02-2022", "01 03 2022", "1-4-2022",
+ "1-5, 2022", "Created on 1 6 2022", "0107 !!! 2022")
> mdy(y)
[1] "2022-01-01" "2022-01-02" "2022-01-03" "2022-01-04" "2022-01-05" "2022-01-06"
[7] "2022-01-07"
> z <- c("01012022", "02-01-2022", "03 01 2022", "4-1-2022",
+ "5-1, 2022", "Created on 6 1 2022", "0701 !!! 2022")
> dmy(z)
[1] "2022-01-01" "2022-01-02" "2022-01-03" "2022-01-04" "2022-01-05" "2022-01-06"
[7] "2022-01-07"
```

Στη σήμανση ημερομηνιών είναι σημαντικό να προσδιορίζεται τη σειρά με την οποία εμφανίζεται το έτος, ο μήνας και η ημέρα.

```
> my_dt <- c("5/02/2022 11:59:22 PM", # Create example dates & times
+ "7/12/2024 07:15:59 AM",
+ "3/10/2022 10:12:03 PM",
+ "5/09/2023 08:11:27 PM")
> my_dt
[1] "5/02/2022 11:59:22 PM" "7/12/2024 07:15:59 AM" "3/10/2022 10:12:03 PM"
[4] "5/09/2023 08:11:27 PM"
> # Print example dates & times
> my_dt_new <- parse_date_time(my_dt, # Parse as a POSIXct class
+ "%d/%m/%y %H:%M:%S %p")
> my_dt_new
[1] "2022-02-05 23:59:22 UTC" "2024-12-07 07:15:59 UTC" "2022-10-03 22:12:03 UTC"
[4] "2023-09-05 20:11:27 UTC"
> # Έλεγχος κατηγορίας δεδομένων
class(my_dt_new)
[1] "POSIXct" "POSIXt"
> # Create example dates & times
my_dt_zone <- dmy_hms(c("5/02/2022 11:59:22 PM",
+ "7/12/2024 07:15:59 AM",
+ "3/10/2022 10:12:03 PM",
+ "5/09/2023 08:11:27 PM"), tz = "Pacific/Auckland")
> # Print example dates & times
my_dt_zone
[1] "2022-02-05 23:59:22 NZDT" "2024-12-07 07:15:59 NZDT" "2022-10-03 22:12:03 NZDT"
[4] "2023-09-05 20:11:27 NZST"
```

Απομάκρυνση am (π.μ.), pm (μ.μ.)

```
as.Date(x, ...)
# S3 method for character
as.Date(x, format,
+ tryFormats = c("%Y-%m-%d", "%Y/%m/%d"),
+ optional = FALSE, ...)
# S3 method for numeric
as.Date(x, origin, ...)
# S3 method for POSIXct
as.Date(x, tz = "UTC", ...)
# S3 method for Date
format(x, ...)
# S3 method for Date
as.character(x, ...)
```

- Επιτρέπει ποικιλία μορφών εισαγωγής μέσω του ορίσματος **format=**
- Η προεπιλεγμένη μορφή είναι ένα **τετραψήφιο έτος**, ακολουθούμενο από έναν **μήνα** και μετά μια **ημέρα**

Code	Value
%d	Day of the month (decimal number)
%m	Month (decimal number)
%b	Month (abbreviated)
%B	Month (full name)
%y	Year (2 digit)
%Y	Year (4 digit)

```
> as.Date("1/15/2001",format="%m/%d/%Y")
[1] "2001-01-15"
> as.Date("April 26, 2001",format="%B %d, %Y")
[1] "2001-04-26"
# %y is system-specific; use with caution
> as.Date("22JUN01",format="%d%b%y")
[1] "2001-06-22"
```

- Στο R, η συνάρτηση **as.Date()** μετατρέπει μεταξύ αναπαραστάσεων κλάσης χαρακτήρων και αντικειμένων που αντιπροσωπεύουν ημερομηνίες
- ```
as.Date("2020-12-10") # Basic R syntax of as.Date function
> my_dt_new
[1] "2022-02-05 23:59:22 UTC" "2024-12-07 07:15:59 UTC"
[3] "2022-10-03 22:12:03 UTC" "2023-09-05 20:11:27 UTC"
> as.Date(my_dt_new) # Αγνοούνται τα στοιχεία ώρας
[1] "2022-02-05" "2024-12-07" "2022-10-03" "2023-09-05"
```

```
Χειρισμός αντικειμένων κλάσεων "POSIXct" και "POSIXlt"
για την αναπαράσταση ημερολογιακών ημερομηνιών και ωρών
as.POSIXct(x, tz = "", ...)
as.POSIXlt(x, tz = "", ...)
S3 method for character
as.POSIXlt(x, tz = "", format,
+ tryFormats = c("%Y-%m-%d %H:%M:%OS",
+ "%Y/%m/%d %H:%M:%OS",
+ "%Y-%m-%d %H:%M",
+ "%Y/%m/%d %H:%M",
+ "%Y-%m-%d",
+ "%Y/%m/%d"),
+ optional = FALSE, ...)
S3 method for default
as.POSIXlt(x, tz = "",
+ optional = FALSE, ...)
S3 method for numeric
as.POSIXlt(x, tz = "", origin, ...)
S3 method for POSIXlt
as.double(x, ...)
```

Fractional seconds  
Στο R, οι συναρτήσεις **as.POSIXct()** και **as.POSIXlt()** μετατρέπουν μια συμβολοσειρά (χαρακτήρες) ημερομηνίας-ώρας, αντίστοιχα στις **κλάσεις POSIXct** (ημερολογιακή ώρα) και **POSIXlt** (ημερολογιακή τοπική ώρα)

- Από προεπιλογή (**format=""**), ως ζώνη ώρας, η συνάρτηση **as.POSIXct()** αποδίδει την τοπική ζώνη ώρας στην οποία έχει ρυθμιστεί ο Η/Υ σας
  - Η προδιαγραφή ζώνης ώρας "GMT" είναι UTC (**Universal Time Coordinated**). Μη έγκυρες τιμές αντιμετωπίζονται συνήθως ως UTC (με προειδοποίηση σε ορισμένες πλατφόρμες).
- ```
> my_dt_nozone <- c("2022-02-05 23:59:22", "2024-12-07 07:15:59", "2022-10-03 22:12:03")
> timeDatect<- as.POSIXct(my_dt_nozone)
> timeDatect
[1] "2022-02-05 23:59:22 +03" "2024-12-07 07:15:59 +03" "2022-10-03 22:12:03 +03"
```

```

> my_dt <- c("2012-03-28 11:59:22", "2022-10-05 09:26:50 AM", "2012-03-28 23:59:22")
> my_dt
[1] "2012-03-28 11:59:22" "2022-10-05 09:26:50 AM" "2012-03-28 23:59:22"
> class(my_dt)
[1] "character"
>
> my_datetime <- as.POSIXct(my_dt)
> str(my_datetime)
POSIXct[1:3], format: "2012-03-28 11:59:22" "2022-10-05 09:26:50" "2012-03-28 23:59:22"
> my_datetime
[1] "2012-03-28 11:59:22 EEST" "2022-10-05 09:26:50 +03" "2012-03-28 23:59:22 EEST"

```

- Στο R, το αποτέλεσμα της συνάρτησης **as.POSIXt()** εξακολουθεί να μοιάζει με τη **POSIXct** μορφή.
- Ωστόσο, η συνάρτηση **as.POSIXt** χρησιμοποιείται όταν θέλουμε να εξαγάγουμε γρήγορα κάποιο τμήμα των δεδομένων (π.χ. μήνες)

```

> my_datetime <- as.POSIXct(my_dt)
> str(my_datetime)
POSIXct[1:3], format: "2012-03-28 11:59:22" "2022-10-05 09:26:50" "2012-03-28 23:59:22"
> my_datetime
[1] "2012-03-28 11:59:22 EEST" "2022-10-05 09:26:50 +03" "2012-03-28 23:59:22 EEST"

```

Αποθήκευση ημερομηνίας / ώρας στο R

- Με τη χρήση των κλάσεων **Date** και **POSIXct**, δεδομένα ημερομηνίας και ώρας αποθηκεύονται **εσωτερικά** ως αριθμός δευτερολέπτων, η μέτρηση των οποίων ξεκινά τα μεσάνυχτα της 1ης Ιανουαρίου 1970.

– Για την αποθήκευση ημερομηνιών πριν από το 1970 χρησιμοποιούνται αρνητικοί αριθμοί.

```

> now <- Sys.time()
> datetime1 <- as.POSIXct(now)
> unclass(datetime1) # πώς αποθηκεύεται το αντικείμενο R
[1] 1668362314

```

- Η κλάση **POSIXt** αποθηκεύει ημερομηνίες και ώρες ως λίστα στοιχείων: δευτερόλεπτο, λεπτό, ώρα, ημέρα του μήνα, μήνα, έτος, αριθμητική ημέρα του έτους, ζώνη ώρας κ.λπ.

```
> datetime2 <- as.POSIXct(now)
```

```
> unclass(datetime2) # πώς αποθηκεύεται το αντικείμενο R
```

```

$sec      $year      # years since 1900
[1] 18.72318 [1] 122
$min      $yday      # 0-6 day of the week, from Sunday
[1] 38      [1] 0
$hour     $yday      $gmtoff
[1] 21      [1] 316    [1] 10800
$mday     $isdst     # Daylight Saving Time flag
[1] 13      [1] 0      attr(,"tzone")
$mon      $zone      [1] ""    "+03"    "EEST"
[1] 10      [1] "+03"

```

Format codes for dates

Code	Value
m	Month (decimal number)
d	Day of the month (decimal number)
y	Year (4 digit)
mon	Month (abbreviated)
month	Month (full name)

Format codes for times

Code	Value
h	Hour
m	Minute
s	Second

Οι προεπιλεγμένες μορφές εισαγωγής είναι έτος-μήνας-ημέρα ώρα:λεπτά: δευτερόλεπτα ή έτος/μήνας/ημέρα ώρα:λεπτά:δευτερόλεπτα

%a	Name of day of the week, an abbreviation.
%A	Same as %a, but the full name
%b	Month name, an abbreviation.
%B	Same as %b, but the full name
%c	Date/time, in the format of the locale.
%C	Century number [00-99]. Calculates the year if a two-digit year is used.
%d	Day of the month [1-31].
%D	Date format, same as %m/%d/%y.
%e	Same as %d.
%g	2 digit year portion of ISO week date [00-99].
%G	4 digit year portion of ISO week date. Can be negative.
%h	Same as %b.
%H	Hour in 24-hour format [0-23].
%I	Hour in 12-hour format [1-12].
%j	Day of the year [1-366].
%m	Month [1-12].
%M	Minute [0-59].

Εάν η μορφή εισόδου δεν είναι τυπική, ...

Μπορούμε να αντιστοιχίσουμε την επιθυμητή μορφή μέσω του ορίσματος **format=**

```

%n Skip all whitespaces until a newline character is found.
%p AM or PM string, used for calculating the hour if 12-hour format is used.
%r Time in AM/PM format of the locale. If not available in the locale time format, defaults to the POSIX time AM/PM format: %I:%M:%S %p.
%R 24-hour time format without seconds, same as %H:%M.
%S Second [00-61]. The range for seconds allows for a leap second and a double leap second.
%t Skip all whitespaces until a tab character is found.
%T 24 hour time format with seconds, same as %H:%M:%S .
%u Weekday [1-7]. Monday is 1 and Sunday is 7.
%U Week number of the year [0-53], Sunday is the first day of the week. Used in calculating the day of the year.
%V ISO week number of the year [1-53]. Monday is the first day of the week. If the week containing January 1st has four or more days in the new year, it is considered week 1. Otherwise, it is the last week of the previous year, and the next week is week 1 of the new year. Used in calculating the day of the year.

```

```

%w Weekday [0 -6]. Sunday is 0.
%W Week number of the year [0-53]. Monday is the first day of the week. Used in calculating the day of the year.
%x Date in the format of the locale.
%X Time in the format of the locale.
%y 2-digit year [0-99].
%Y 4-digit year. Can be negative.
%Z UTC offset. Output is a string with format +HHMM or -HHMM, where + indicates east of GMT, - indicates west of GMT, HH indicates the number of hours from GMT, and MM indicates the number of minutes from GMT.
%Z Time zone name.
%% % character.

```

Λεπτομέρειες χρήσης δεδομένων χρόνου

- Οι συναρτήσεις **as.POSIX*** μετατρέπουν ένα αντικείμενο της άλλης κλάσης και της κλάσης "Date" σε αυτές τις κλάσεις
 - Ημερομηνίες χωρίς ώρα θεωρούνται ότι είναι μεσάνυχτα UTC
 - Η εισαγωγή χαρακτήρων μετατρέπεται πρώτα σε κλάση **"POSIXt"** με τη συνάρτηση **strptime()**. Η εισόδος αριθμητικών δεδομένων χρόνου μετατρέπεται πρώτα σε κλάση **"POSIXct"**.
- Μπορούν επίσης να μετατρέψουν συμβολοσειρές χαρακτήρων των μορφών π.χ., "2001-02-03" και "2001/02/03". Το ίδιο με ώρες στη μορφή π.χ. "11:52" ή "11:52:03".


```
> now <- Sys.time()
> now
[1] "2022-11-14 00:38:53 +03"
> options(digits.secs=4) #How many decimal points
> now <- Sys.time()
> now
[1] "2022-11-14 00:39:02.9988 +03"
> datetime1 <-
  as.POSIXct(now,format='%d/%b/%Y:%H:%M:%OS3')
> datetime1
[1] "2022-11-14 00:39:02.9988 +03"
> datetime2 <-
  as.POSIXlt(now,format='%d/%b/%Y:%H:%M:%OS3')
> datetime2
[1] "2022-11-14 00:39:02.9988 +03"
```

Μερικές ιδιαιτερότητες ...

- Οι τιμές μήνα που είναι αποθηκευμένες σε ένα **POSIXit** αντικείμενο χρησιμοποιούν **zero-based indexing** → ο μήνας #1 (Ιανουάριος) αποθηκεύεται ως 0, ο μήνας #2 (Φεβρουάριος) ως 1, ..., ο Οκτώβριος ως ο 9ος μήνας (\$mon = 9)
- Τα έτη αποθηκεύονται επίσης διαφορετικά σε ένα **POSIXit** αντικείμενο. Οι τιμές του έτους αποθηκεύονται χρησιμοποιώντας ως εποχή αναφοράς το 1900/01/01. Έτσι, το 2022 αποθηκεύεται ως 122 (\$year = 122, δηλ. 122 χρόνια από το 1900).

```
IAIATEPOTHTEE (MINNE KAI ETH ETIE EYNAPTHHEIT POSIXot KAI POSIXlt)
> # Convert character data to POSIXit date and time
> timeDate1<- as.POSIXlt("2022-10-19 10:15")
> str(timeDate1)
POSIXlt[1:1], format: "2022-10-19 10:15:00"
>
> timeDate1
[1] "2022-10-19 10:15:00 +03"
>
> unclass(timeDate1)
$sec
[1] 0
$min
[1] 15
$mon
[1] 9
$year
[1] 122
```

Η ταξινόμηση ημερομηνιών ως ξεχωριστών στοιχείων καθιστά τη συνάρτηση **as.POSIXlt()** υπολογιστικά πιο εντατική στη χρήση δεδομένων που αποθηκεύονται ως πλαίσια δεδομένων. Αυτό τυπικά επιβραδύνει τους υπολογισμούς!

Πακέτο **chron** ...

- Παρέχει επιπλέον συναρτήσεις μετατροπής δεδομένων χρόνου σε ποικίλα αντικείμενα διαφορετικών χειρισμών ημερομηνίας και ώρας
 - **.Holidays**, ..., **as.character.dates**, ..., **as.chron.Date**, ..., **day.of.week**, ..., **month.day.year**
 - Επιτρέπουν αριθμητικές πράξεις (π.χ., αφαίρεση Day1-Day2, σταθερή πρόσθεση Day1+σταθερές μέρες), όλες τις λογικές συγκρίσεις, περιλήψεις **min()**, **max()** και **range()** (που απορρίπτουν NA τιμές από προεπιλογή).

Τυπική σύνταξη εντολών:

- ```
chron(dates., times.,
 format = c(dates = "m/d/y", times = "h:m:s"),
 out.format, origin.)
```
- Χρησιμοποιώντας ως είσοδο οποιοδήποτε **chron** αντικείμενο μπορείτε να εξάγετε το αντίστοιχο χρονικό διάστημα μέσω κατάλληλων συναρτήσεων, όπως
    - **days()**, **months()**, **quarters()**, **years()**, **weeks()**, **weekdays()**, **hours()**, **minutes()**, και **seconds()**
  - Η τρέχουσα υλοποίηση **chron** αντικειμένων δεν χειρίζεται ζώνες ώρας ούτε θερινή ώρα

- Η μορφή ημερομηνιών μπορεί να είναι οποιαδήποτε μετάθεση των χαρακτήρων "d", "m" ή "y" οριοθετημένη με ή ενδεχομένως χωρίς διαχωριστικό, π.χ., "m/d/y", "d-m-y", "ymd", είναι όλες έγκυρες μορφές.
- Η μορφή **format=** μπορεί επίσης να είναι μεταθέσεις των λέξεων "day", "month" και "year" (με κενό ως διαχωριστικό), που παράγει την ημερομηνία με την χρησιμοποιούμενη διάταξη
  - **"month day year"** → **February 27 2017**
  - **"day mon year"** → **27 February 2017**
- Η τρέχουσα υλοποίηση **chron** αντικειμένων δεν χειρίζεται ζώνες ώρας ούτε θερινή ώρα

```
> dts <- dates(c("02/27/17", "05/31/17", "06/03/17"))
> tms<- times(c("23:03:20", "22:29:56", "01:03:30"))
> x <- chron(dates = dts, times = tms)
> x
[1] (02/27/17 23:03:20) (05/31/17 22:29:56) (06/03/17 01:03:30)
>
> y <- chron(dates = dts, format = "month day year")
> y
[1] February 27 2017 May 31 2017 June 03 2017
>
> z <- t(chron(dates = dts, format = "day mon year"))
> z
[1,] [2,] [3,]
[1,] 27 Feb 2017 31 May 2017 03 Jun 2017
>
> dts[3] = dts[1]
Error: unexpected input in "dts[3] ="
> c(dts[1], dts[1] + 10)
[1] 02/27/17 03/09/17
> dts[3]-dts[1]
Time in days:
[1] 96
> c(dts[1], dts[1] + 10)
[1] 02/27/17 03/09/17
```

Μπορείτε επίσης να εξαγάγετε το έτος, το τρίμηνο, το μήνα, την ημέρα (εντός του μήνα) ή την ημέρα της εβδομάδας της ημερομηνίας χρησιμοποιώντας άλλες χρήσιμες συναρτήσεις, όπως: **weekdays()**, **months()**, **years()**, **quarters()**

```
> library("lubridate")
> ymd_hms("2017-06-03 23:59:59")
[1] "2017-06-03 23:59:59 UTC"
> ld <- mdy_hms("05/31/2017 23:59:59")
> ld + seconds(1)
[1] "2017-06-01 UTC"
> ld
[1] "2017-05-31 23:59:59 UTC"
> ld + seconds(1)
[1] "2017-06-01 UTC"
>
> now()
[1] "2022-11-20 19:36:07 +03"
> # Using ceiling_date() takes a date-time object and rounds it up
> # to the nearest integer value of the specified time unit
> ceiling_date(now(), unit = "month")
[1] "2022-12-01 +03"
> ceiling_date(now(), unit = "day")
[1] "2022-11-21 +03"
> ceiling_date(now(), unit = "month") - days(1)
[1] "2022-11-30 +03"
> am(now())#is it morning?
[1] FALSE
```

```
Extract parts from dates
> dts
[1] 02/27/17 05/31/17 06/03/17
> weekdays(dts)
[1] Mon Wed Sat
Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
> months(dts)
[1] Feb May Jun
12 Levels: Jan < Feb < Mar < Apr < May < Jun < Jul < Aug < Sep < ... < Dec
> years(dts)
[1] "17224y 0m 0d 0H 0M 0S" "17317y 0m 0d 0H 0M 0S" "17320y 0m 0d 0H 0M 0S"
> quarters(dts)
[1] 1Q 2Q 2Q
Levels: 1Q < 2Q < 3Q < 4Q
>
> out(dts, "months")
[1] Feb 17 May 17 Jun 17
Levels: Feb 17 < Mar 17 < Apr 17 < May 17 < Jun 17
> out(dts, "days")
[1] day 1 day 94 day 97
97 Levels: day 1 < day 2 < day 3 < day 4 < day 5 < day 6 < day 7 < ... < day 97
```

```
> # Use each accessor function to set the components of a date/time
```

```
>
> datetime <- ymd_hms("2016-07-08 12:34:56")
> datetime
```

```
[1] "2016-07-08 12:34:56 UTC"
```

```
>
> year(datetime) <- 2020
> datetime
```

```
[1] "2020-07-08 12:34:56 UTC"
```

```
>
> month(datetime) <- 01
> datetime
```

```
[1] "2020-01-08 12:34:56 UTC"
```

```
>
> hour(datetime) <- hour(datetime) + 1
> datetime
```

```
[1] "2020-01-08 13:34:56 UTC"
```

```
> # Modifying in place, you can create a new date-time with update()
> # If values are too big, they will roll-over:
```

```
> update(ymd("2015-02-01"), mday = 30)
[1] "2015-03-02"
```

```
> update(ymd("2015-02-01"), hour = 400)
[1] "2015-02-17 16:00:00 UTC"
```

## Στη συνέχεια ...

- Συμβουλευτείτε το συναφές διδακτικό υλικό από τις ιστοσελίδες του μαθήματος
- Rscripts για τις πρώτες δοκιμές και εξοικείωση σας με τις διαδικασίες διαχείρισης πακέτων του R



- Περισσότερα μέσω των ασκήσεων ...