

# 18/10/2022

Εθνικό Μετσόβιο Πολυτεχνείο



Σχολή Αγρονόμων & Τοπογράφων Μηχανικών

## Αναλυτικές Μέθοδοι στη Γεωπληροφορική Εισαγωγή στο λογισμικό R

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

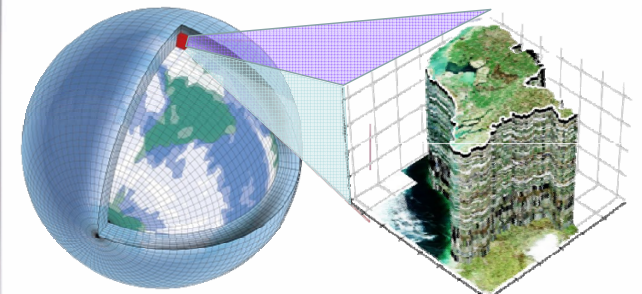
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

## Σημερινή ενότητα του μαθήματος

- Μετατροπή πρωτογενών (αδόμητων ή δομημένων) δεδομένων σε κατανόηση, διορατικότητα και γνώση → **Επιστήμη Δεδομένων (Data Science)**
- Μια συνοπτική αναφορά στο ρόλο και τη χρησιμότητα του λογισμικού R (και της γλώσσας προγραμματισμού r) στις σημερινές ανάγκες της **Επιστήμης Δεδομένων (Data Science)**
- R: Μια πρώτη ματιά στο περιβάλλον και τα εργαλεία του

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



- Ενώ σήμερα έχουμε πρόσβαση σε περισσότερα δεδομένα από ποτέ, η **περιορισμένη ικανότητα ανάλυσής τους** συχνά **παρεμποδίζει τον αντίκτυπο** αυτών των δεδομένων στη βελτίωση των καθημερινών πρακτικών ή προβλημάτων ενδιαφέροντος

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



## Πόσα δεδομένα δημιουργούνται καθημερινά ;

- Το πλήθος των διαθέσιμων δεδομένων είναι τεράστιο και αυξάνεται εκθετικά κάθε μέρα.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

## Δεδομένα παντού

- Επιστημονικά πειράματα
- Το διαδίκτυο των πραγμάτων
- Δεδομένα Ιστού, ηλεκτρονικό εμπόριο, χρηματοοικονομικές συναλλαγές
- Αμέτρητες 'έξυπνες' συσκευές και ασύρματα δίκτυα αισθητήρων
- Κοινωνικά δίκτυα
- Διακίνηση δισεκατομμυρίων email καθημερινά
- Ίχνη κινούμενων οχημάτων και σκαφών
- Διακίνηση οπτικοακουστικού περιεχομένου
- ..... πολλά περισσότερα!



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'



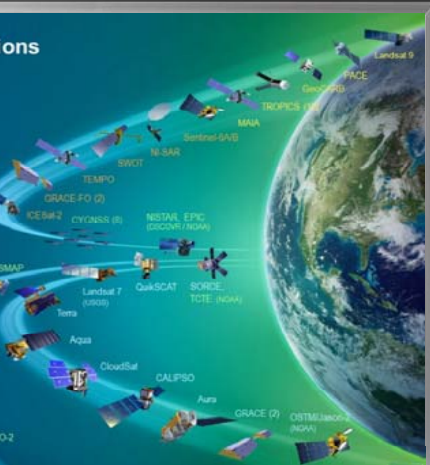
Κάθε μέρα δημιουργούνται τουλάχιστον 2.5 quintillion [ $\times 10^{18}$ ] bytes δεδομένων  
Μέχρι το 2020, ο όγκος των δεδομένων παγκοσμίως υπολογίστηκε σε  $44 \times 10^{21}$  bytes

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

## Earth Science Missions

Κατά το 2021 μόνο τα δεδομένα της NASA για την παρατήρηση της Γης αυξήθηκαν από 41.8 σε 59.2 Petabyte (1 PB=1000 TB) δεδομένων, από σχεδόν 100 ενεργές αποστολές και χιλιάδες αισθητήρες και συστήματα στη Γη και το διάστημα



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

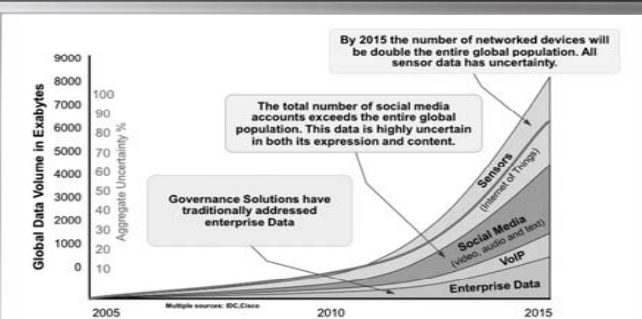
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Σήμερα τα διαθέσιμα γεωδεδομένα προέρχονται από παντού ...

- Αισθητήρες που χρησιμοποιούνται για τη συλλογή πληροφοριών σχετικά με το κλίμα
- Γεωγραφικές πληροφορίες που εμπεριέχονται σε αναρτήσεις σε ιστοτόπους κοινωνικών μέσων
- Ψηφιακές εικόνες και βίντεο που διακινούνται ηλεκτρονικά
- Αρχεία από σαρωτές Lidar, UAVs (drones) / UUVs (υποβρύχια οχήματα) και από δορυφορικά σήματα εντοπισμού GNSS (κινητά τηλέφωνα), ...
- Αυτά είναι τα δεδομένα που στον κόσμο της τεχνολογίας συχνά αποκαλούνται με τον όρο **'Μεγάλα Δεδομένα' (Big Data)**

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

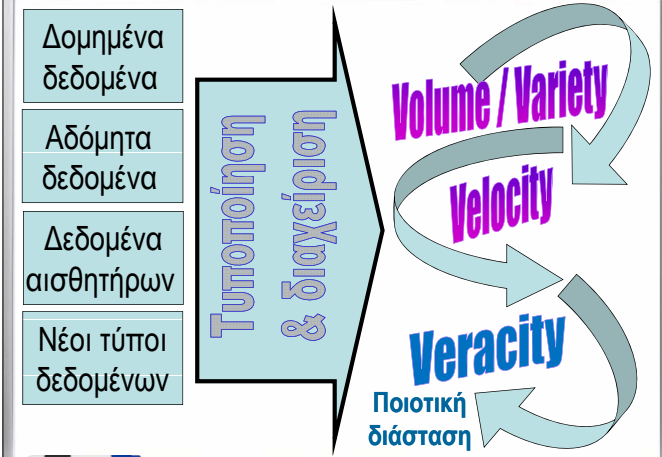
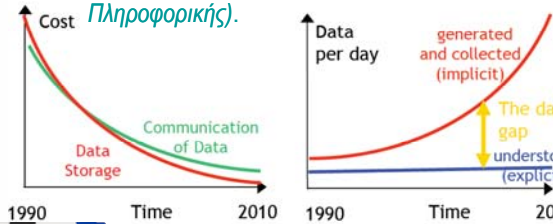


- Μερικές φορές, μαζί με το αυξανόμενο μέγεθος των διαθέσιμων δεδομένων, συχνά αλλάζει απότομα και η **αβεβαιότητα των ίδιων των δεδομένων**, όπως και η **αβεβαιότητα στη μοντελοποίηση και στην επεξεργασία τους** μεταβάλλεται επίσης πολύ σημαντικά.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΙΚΗ'

- Πρόσφατες μελέτες → το νέο πληροφοριακό περιεχόμενο υπερδιπλασιάζεται σε ετήσια βάση, ο χρόνος που δαπανάται για την αξιοποίησή του παραμένει σταθερός (μόνο 5-10% των διαθέσιμων δεδομένων τυχάνει κάποιας πλήρους ανάλυσης) → Συνεχώς αυξανόμενο αποτύπωμα των ψηφιακών δεδομένων + εκθετική αύξηση της υπολογιστικής ισχύος → **φαινόμενο Big Data** (Πρόβλημα της Πληροφορικής).



## ΕΤΔ: Επιστήμη των Δεδομένων / Data Science



- Ένας νέος όρος που ήλθε να αντικαταστήσει προγενέστερους όρους, όπως
  - Ανακάλυψη Γνώσης (*Knowledge Discovery*)
  - Εξόρυξη Δεδομένων (*Data Mining*)
- Συχνά και οι τρεις όροι χρησιμοποιούνται σχεδόν ως ταυτόσημοι για να περιγράψουν τεχνικές ανακάλυψης μοτίβων ή μη αναμενόμενων σχέσεων μεταξύ των δεδομένων, και να τα συνοψίσουν με τρόπους κατανοητούς στους χρήστες

- Η ΕΤΔ συχνά χαρακτηρίζεται ως το επόμενο σύνορο για την καινοτομία
- Αποτελεί εξέλιξη και συνδυασμό άλλων επιστημονικών κλάδων, κυρίως της Στατιστικής και της Πληροφορικής.
  - Με βασικό αντικείμενο την εξαγωγή γνώσης και προβλέψεων από ετερογενείς, συνεχώς μεταβαλλόμενους μεγάλους όγκους δεδομένων



- **Παράδειγμα:** Η ναυτιλιακή βιομηχανία υφίσταται αναπόφευκτα μια τεράστια αλλά ευεργετική αλλαγή που οδηγείται από τις δυνατότητες Μεγάλων Δεδομένων σε διάφορους τομείς

- Παράδειγμα: στον κλάδο της ναυτιλίας ...**
- Το **Πεδίο Ενδιαφέροντος** είναι η βέλτιστη μετακίνηση των πλοίων, τα διακινούμενα φορτία, τα καιρικά μοτίβα, ... δεδομένου ότι αυτή η γνώση δίνει τη δυνατότητα στις ναυτιλιακές εταιρείες να βελτιστοποιήσουν τις δραστηριότητές τους και να μειώσουν το κόστος.
  - Το **Ψηφιακό Αποτύπωμα** είναι τα γεωγραφικά και επιχειρησιακά δεδομένα για τα πλοία (π.χ., η τρέχουσα θέση και ο προορισμός τους), καθώς και τα δεδομένα που σχετίζονται με την κατάσταση των θαλασσών (κύματα, ρεύματα, παλίρροιες) και τις επικρατούσες συνθήκες εν πλω.
  - Οι **Καταγεγραμμένες Πληροφορίες** περιορίζονται συνήθως σε γεωγραφικά και επιχειρησιακά δεδομένα από πολλές ναυτιλιακές εταιρείες, αλλά και ποικίλα δεδομένα από δορυφόρους ή επίγειους σταθμούς.

- Παράδειγμα: στον κλάδο της ναυτιλίας ...**
- Το **Ενδιάμεσο Στάδιο** είναι όλες εκείνες οι υποδομές για την αποθήκευση και το χειρισμό μεγάλου όγκου διαφορετικών δεδομένων, αξιολόγησης της ποιότητας / ακρίβειας των δεδομένων, καθώς και τη διαχείριση μεγάλων ροών δεδομένων σε πραγματικό χρόνο.
  - Η **Ανθρώπινη και Τεχνητή Νοημοσύνη** που προέρχεται από την μεγάλη συλλογή δεδομένων είναι ο συσχετισμός παραμέτρων και μεγεθών μεταξύ θαλάσσιων μετακινήσεων, της δυναμικής κατάστασης των θαλασσών και των επιχειρησιακών χαρακτηριστικών της εκάστοτε θαλάσσιας δραστηριότητας.

Παράδειγμα: στον κλάδο της ναυτιλίας ...

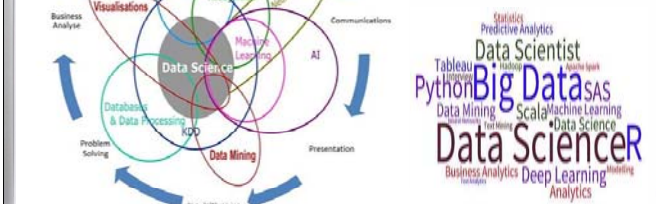
- Οι **οργανωτικές και άλλες εφαρμογές** παρέχουν αυξημένη βεβαιότητα στις προβλέψεις και τη χρήση των καυσίμων, βελτιστοποιημένες διαδρομές πλεύσης και πλοήγησης, καθώς και συστηματική παρακολούθηση για την αποφυγή συμβάντων λόγω αντίξων καιρικών συνθηκών.
- Οι επιπτώσεις και ο **αντίκτυπος στην αξία** από τη χρήση των δεδομένων με αυτόν τον επικεντρωμένο τρόπο είναι το χαμηλότερο κόστος, οι χαμηλότεροι λειτουργικοί κίνδυνοι και η μεγαλύτερη αξιοποίηση του διαθέσιμου δυναμικού και πόρων που θα μπορούσε να επιτρέψει περισσότερα έσοδα.

## Παράδειγμα: στις θαλάσσιες επιστήμες



## Η Επιστήμη των δεδομένων

... ως διεπιστημονική περιοχή, εκτός της πληροφορικής και των μαθηματικών, έχει ως βασικό πυλώνα και τη γνώση του εκάστοτε πεδίου εφαρμογής της



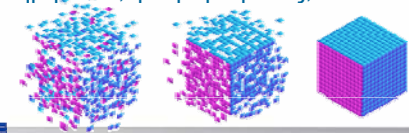
- Η **Επιστήμη των Δεδομένων** είναι χρήσιμη μόνο όταν τα δεδομένα (structured / unstructured) χρησιμοποιούνται για να απαντήσουν συγκεκριμένα ερωτήματα, συνήθως επιστημονικού, οικονομικού, ... ενδιαφέροντος



– Το πρόβλημα με αυτή τη άποψη της ΕτΔ είναι ότι είναι πολύ πιο δύσκολο από την άποψη που επικεντρώνεται απλά στο μέγεθος ή στα εργαλεία των δεδομένων

- Η **Ανάλυση Δεδομένων** είναι ένας χαλαρός όρος που αναφέρεται σε τεχνικές που εμπλέκουν, συνδυάζουν και χρησιμοποιούν δεδομένα, όπως και σε κάθε προσπάθεια κατανόησης των διαθέσιμων δεδομένων, με στόχο τη δημιουργία μοντέλων με δυνατότητες πρόγνωσης.

– π.χ., εξειδικευμένες στατιστικές τεχνικές ανάλυσης για ενδελεχή και εμπειριστατωμένη επεξεργασία δεδομένων, προηγμένες τεχνικές συνόρθωσης παρατηρήσεων, φιλτραρίσματα, ...



- Οι εκάστοτε διαδικασίες συλλογής και ανάλυσης στοιχείων απαιτούν τεχνολογική και υπολογιστική υποδομή ικανή να υποστηρίξει όλες τις επί μέρους διεργασίες (εκτίμησης, απόδοσης, ποιότητας,...) μιας ολοκληρωμένης έρευνας



- Η **Μηχανική Μάθηση** αναφέρεται στη δημιουργία υπολογιστικών διαδικασιών που είναι αρκετά ευφυείς ώστε να μαθαίνουν από τα ίδια τα δεδομένα.

– Αφορούν κυρίως στο σχεδιασμό και ανάπτυξη αλγορίθμων που δέχονται ως είσοδο εμπειρικά δεδομένα (π.χ. από αισθητήρες ή βάσεις δεδομένων) → δίνουν ως έξοδο σχετικές προβλέψεις για τους μηχανισμούς που δημιούργησαν τα δεδομένα

- Συνήθως, αυτό είναι το τελευταίο στάδιο της εξόρυξης δεδομένων

## Αναγκαία Εργαλεία για την Επιστήμη των Δεδομένων;



- Αυτά συνεχώς εξελίσσονται. Ωστόσο, υπάρχουν δύο χαρακτηριστικές κατηγορίες αναδυόμενων εργαλείων:
  - Εργαλεία αυτοεξυπηρέτησης που απαιτούν τεχνική εμπειρογνομosύνη (δεξιότητες προγραμματισμού και κατανόηση στατιστικών αναλύσεων και πληροφορικής)
  - Εργαλεία που επιτρέπουν στους χρήστες να αυτοματοποιούν την ανάλυση δεδομένων με ανοιχτό, συνεργατικό και τυποποιημένο τρόπο

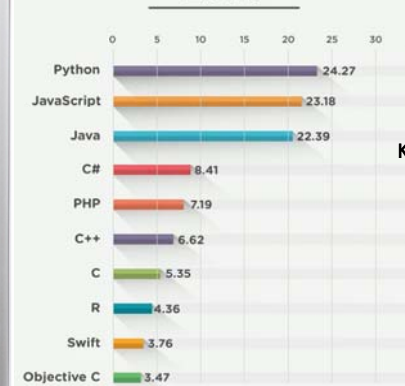
## Προγραμματιστικά εργαλεία στην Ετά



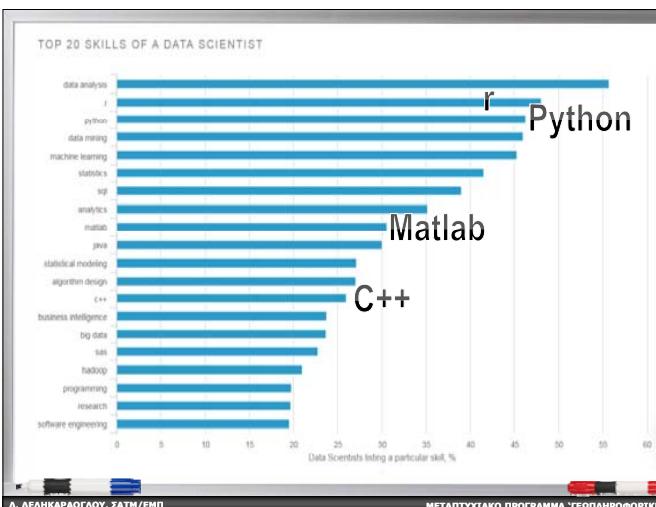
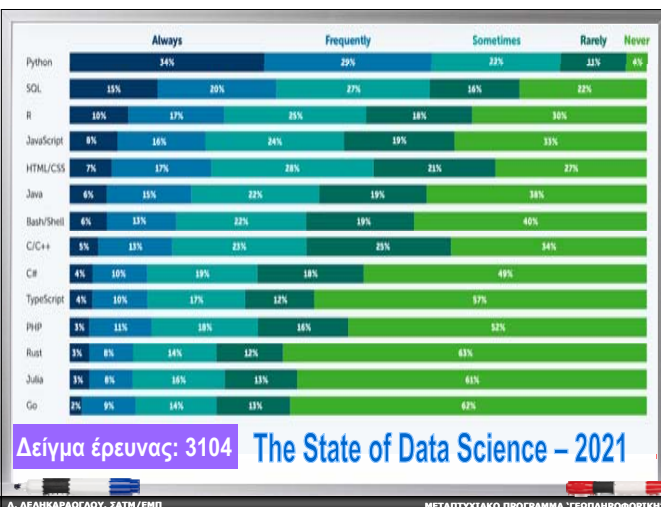
## Η πλατφόρμα του R – Με μια σύντομη ματιά –

- Ένα πολύτιμο εργαλείο διαχείρισης δεδομένων, κυρίως στατιστικής ανάλυσης (αλλά όχι μόνο) που περιλαμβάνει σύγχρονες μεθοδολογίες ανάλυσης μεγάλου όγκου δεδομένων
  - Μια γλώσσα προγραμματισμού και λογισμικό ανοικτού κώδικα που με τη χρήση τους, καθένας μπορεί να συνεισφέρει στην ανάπτυξη βιβλιοθηκών που υλοποιούν στατιστικές μεθοδολογίες και εκτεταμένες αναλύσεις δεδομένων + Υψηλής ποιότητας γραφικά → σημαντικό εργαλείο στην οπτικοποίηση δεδομένων

2019



Από το 2014 η πλατφόρμα R συγκαταλέγεται στην πρώτη δεκάδα στην κατάταξη δημοτικότητας μεταξύ των κυρίαρχων γλωσσών προγραμματισμού, αντανάκλωντας την αυξανόμενη σημασία της για την ανάλυση δεδομένων



## Τι ακριβώς είναι το R;



περιβάλλον προγραμματισμού (Linux, Mac OS και Windows) + μια γλώσσα προγραμματισμού, ερμηνείας και εκτέλεσης εντολών = προηγμένο λογισμικό ανοικτού κώδικα

## Επιστήμη των δεδομένων και το λογισμικό R;



## Επιστήμη των δεδομένων και το λογισμικό R;



- Εισαγωγή δεδομένων ...
  - από ένα αρχείο, μια βάση δεδομένων, ή έναν ιστότοπο, και αποθηκεύονται με κατάλληλο τρόπο στο R για περαιτέρω επεξεργασία
  - Το R έχει την πιο εκτεταμένη σειρά εργαλείων για την εισαγωγή δεδομένων, από διαφορετικές πηγές, ποικίλες δομές και μορφότυπους, άλλα λογισμικά, ...

## Επιστήμη των δεδομένων και το λογισμικό R;



- Τακτοποίηση/μορφοποίηση των στοιχείων των δεδομένων ...
  - αποθήκευση σε συνεκτικές μορφές/δομές που ταιριάζουν με τη σημασιολογία του συνόλου των δεδομένων
    - π.χ., κάθε στήλη να είναι μια μεταβλητή, και κάθε σειρά να είναι μια παρατήρηση

## Επιστήμη των δεδομένων και το λογισμικό R;



### • Μετατροπή των δεδομένων ...

– π.χ., εξαγωγή ενός υποσυνόλου στοιχείων ενδιαφέροντος, εισαγωγή νέων μεταβλητών που είναι συναρτήσεις πρότερων μεταβλητών, συνοπτικά στατιστικά μέτρα, ... με το R

- Effective analyses can be scripted
- Scripting also required for repeatability
  - Transparency, Sharing, Insight

## Επιστήμη των δεδομένων και το λογισμικό R;



### • Δύο κύριοι μηχανισμοί παραγωγής γνώσης ...

– Οπτικοποίηση και μοντελοποίηση

- συμπληρωματικά πλεονεκτήματα και μειονεκτήματα

– οποιαδήποτε ανάλυση πραγματικών δεδομένων επιβάλλει το συνδυασμό και εναλλαγή μεταξύ τους

## Επιστήμη των δεδομένων και το λογισμικό R;



### • Οπτικοποίηση των δεδομένων ...

– Μια καλή απεικόνιση μπορεί να αναδείξει μη αναμενόμενα χαρακτηριστικά, να προκαλέσει νέα ερωτήματα σχετικά με τα δεδομένα ή να επιβάλει τη συλλογή νέων στοιχείων.



## Επιστήμη των δεδομένων και το λογισμικό R;



### • Μοντελοποίηση των δεδομένων ...

– συμπληρωματική διαδικασία της οπτικοποίησης

• Ένα καλό μοντέλο μπορεί να περιγράψει με ακρίβεια μια φυσική διαδικασία, ένα φαινόμενο ή ένα σύστημα

– κάθε μοντέλο στηρίζεται στις δικές του υποθέσεις, και εκ της φύσεως του δεν μπορεί να αμφισβητεί τις υποθέσεις του

## Επιστήμη των δεδομένων και το λογισμικό R;



• Επικοινωνία των συμπερασμάτων από την ανάλυση των δεδομένων ... το πλέον κρίσιμο μέρος κάθε ανάλυσης δεδομένων.

• Δεν έχει σημασία πόσο καλά είναι τα μοντέλα και πως η οπτικοποίηση των αποτελεσμάτων έχει οδηγήσει στο να κατανοήσουμε τα δεδομένα, εκτός εάν μπορούμε επίσης να επικοινωνήσουμε τα αποτελέσματα των αναλύσεων μας σε άλλους

– R + Markdown (authoring syntax) + Latex = Authoring R Presentations (slides, HTML5, ...)

## Η εργαλειοθήκη του λογισμικού R;

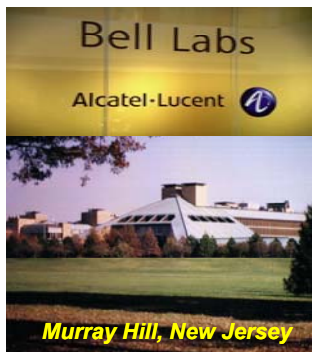


Η χρήση του λογισμικού και της γλώσσας προγραμματισμού R καλύπτει εγκάρσια, με πληθώρα εργαλείων, όλα τα βασικά στάδια της ανάλυσης δεδομένων (από Small σε Big Data)

– αν υπάρχει μια παλαιότερη ή νέα στατιστική τεχνική ή μεθοδολογία ανάλυσης δεδομένων, οι πιθανότητες είναι ότι υπάρχει ήδη για αυτήν ένα διαθέσιμο πακέτο R για άμεση χρήση ... !!!

## Η αφετηρία ανάπτυξης του R

- Η ανάπτυξη του έχει βάση στη δημιουργία της γλώσσας προγραμματισμού S που αναπτύχθηκε, γύρω στο 1976, στα εργαστήρια της εταιρίας τηλεπικοινωνιών Bell (πρώην AT&T) για την ανάλυση δεδομένων
- Το εμπορικό προϊόν, εμφανίστηκε, το 1988, με την ονομασία: **S-plus**.



## Η προέλευση του R

### Οι 'γονείς' του R



- Το διαδραστικό περιβάλλον S, και
- η αλγοριθμική γλώσσα **Scheme**, που αναπτύχθηκε στο MIT, και είναι μια από τις δύο βασικές διαλέκτους της **Lisp** → κοινοποιήθηκε μέσα από τα περίωνυμα "Lambda Papers"

## Λογισμός λ (λ-Calculus)

από όπου και η ονομασία των 'λ-papers'



- Ένας formalismos σχεδιασμένος, τη δεκαετία του '70, για τη διερεύνηση ορισμών, εφαρμογών συναρτήσεων και αναδρομής συναρτήσεων
- Αναπτύχθηκε αρχικά από τον **Alonzo Church** στις αρχές της δεκαετίας του 1930, πολύ πριν αρχίσουν να χρησιμοποιούνται οι Η/Υ

## Οι εμπνευστές και δημιουργοί του R

- Αρχικά δημιουργήθηκε και διανεμήθηκε ως ελεύθερο λογισμικό από τους **Robert Gentleman** και **Ross Ihaka** στο Πανεπιστήμιο του Auckland, Νέα Ζηλανδία, στη δεκαετία του '90. Στόχος τους ήταν η δημιουργία ενός φιλικού προγραμματιστικού περιβάλλοντος για τη διδασκαλία της Στατιστικής



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Η ονομασία R – ένα απλό λογοπαίγνιο

- " We have named our language R -in part to acknowledge the influence of S and in part to celebrate our own efforts."

R. Gentleman και R. Ihaka

Scheme's semantics + S' syntax = R



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Η διαδρομή εξέλιξης του R

- Το περίφημο άρθρο από όπου ξεκίνησε η διαδρομή του R...
  - *Ihaka R. & Gentleman R. 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5: 299–314*
- Το 1995, εμφανίζεται η πρώτη έκδοση του λογισμικού R, ως ανοικτό πακέτο εργαλείων για τη διδασκαλία ενός εισαγωγικού μαθήματος Στατιστικής.
- Το 1996, στη Βιέννη, δημιουργείται το πρώτο αποθετήριο προγραμμάτων στη γλώσσα r από τους ίδιους τους τότε πρώτους (και λίγους) χρήστες

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## ... και η εξέλιξη του R

- Από το 1997: αναπτύσσεται από μια διεθνή επιστημονική ομάδα (**R-Core Team**) & χιλιάδες ενεργούς προγραμματιστές, στατιστολόγους, χρήστες κ.ά. που μοιράζονται τις βιβλιοθήκες και τα εργαλεία τους !!



- **Contributed functionality** → Παρέχει πλήρη πρόσβαση σε αλγόριθμους, ανοιχτά και προσβάσιμα εργαλεία και τις εφαρμογές τους που σε μεγάλο βαθμό δεν είναι ακόμη διαθέσιμα σε άλλα εμπορικά λογισμικά.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## ... και η εξέλιξη του R

- Η σχέση του R με τη γλώσσα S: παραμένει εξ ίσου ισχυρή μέχρι σήμερα !!
- Η R μπορεί να θεωρηθεί ως μια διαφορετική υλοποίηση (εφαρμογή) της S
  - Αν και υπάρχουν ορισμένες σημαντικές διαφορές μεταξύ της S και της r, κυρίως ως προς τον τρόπο που διαχειρίζονται την μνήμη του Η/Υ
  - Ένα μεγάλο μέρος του κώδικα της S έχει προστεθεί αυτούσιο στην r, ενώ
  - κώδικας που γράφεται για το R τρέχει σχεδόν αμετάβλητος και στις σημερινές εκδόσεις της S-Plus

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Πρόσβαση στο R;



- Είναι ελεύθερα προσβάσιμο από ένα κεντρικό δίκτυο διανομής στο Διαδίκτυο (**Comprehensive R Archive Network, CRAN**)
    - [cran.r-project.org](http://cran.r-project.org) → ~90+ sites
    - [ftp.cc.uoc.gr/mirrors/CRAN/](http://ftp.cc.uoc.gr/mirrors/CRAN/)
    - [www.r-project.org](http://www.r-project.org)
- Πλήθος διακομιστών (mirror CRAN sites) που διανέμονται σε όλο τον κόσμο χρησιμοποιούνται για τη διανομή του R και των πακέτων του.

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Πρόσβαση στο R;



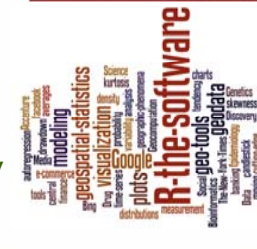
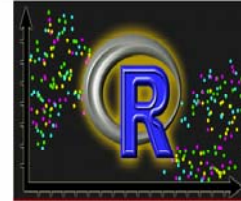
- Ο κύκλος ανάπτυξης/ανανέωσης του R είναι πολύ γρήγορος
  - Σχεδόν κάθε 6 μήνες (Απρίλιο & Οκτώβριο) → νέες εκδόσεις
  - Στο ενδιάμεσο διάστημα → εκδόσεις *alpha* και *beta* αναρτώνται (στο κεντρικό αποθετήριο και στους ιστότοπους-είδωλα του CRAN για έλεγχο και δοκιμές από τους χρήστες)
- Οι πηγαίοι και οι εκτελέσιμοι κώδικες του R μπορούν να ληφθούν για διάφορα λειτουργικά συστήματα (**Windows, Linux, MacOS, ...**)

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Το R (το λογισμικό)

- Αποτελεί μια απaráμιλλη πλατφόρμα για τον προγραμματισμό νέων στατιστικών μεθόδων με έναν εύκολο και απλό τρόπο.
- **State-of-the-art** δυνατότητες γραφικών που συναγωνίζεται μόνο η *Matlab*



Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## Το R (το λογισμικό)



- Αποτελεί ένα όχημα εφαρμογής στατιστικών (και όχι μόνο) αναλύσεων, παρέχοντας μέσω επεκτάσιμου ανοικτού κώδικα κατάλληλες διαδρομές για τη λήψη διερευνητικών αποφάσεων
- Καλύπτει ευρύ φάσμα μεθόδων, όπως
  - γραμμικά και μη γραμμικά μοντέλα,
  - στατιστικούς ελέγχους υποθέσεων και ποιότητας,
  - ανάλυσης χρονοσειρών, ταξινόμησης, ομαδοποίησης, ...καθώς και δημιουργίας γραφικών και απεικονίσεων

Α. ΔΕΛΗΚΑΡΑΓΛΟΥ, ΣΑΤΜ/ΕΜΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ 'ΓΕΩΠΛΗΡΟΦΟΡΪΚΗ'

## To R (το λογισμικό)



- Δίνει τη δυνατότητα βελτίωσης ή/και επέκτασης των διαθέσιμων πακέτων (βιβλιοθήκες συναρτήσεων και σύνολα δεδομένων) και εργαλείων που παρέχονται ελεύθερα, και βελτιώνονται συνεχώς
  - Επεκτάσιμες βιβλιοθήκες, υπό τον πλήρη έλεγχο του εκάστοτε χρήστη
    - Ενσωμάτωσή τους σε καινούργιες βιβλιοθήκες εύκολα και αποτελεσματικά → ραγδαία αύξηση και ανάπτυξη

## To R (το λογισμικό)



- Ευκολία παραγωγής τεκμηριωμένων δημοσιεύσεων, συμπεριλαμβανομένων των μαθηματικών συμβόλων και τύπων, όπου χρειάζεται.

## To R (το λογισμικό)



- Πλείστα *User forums* και *mailing lists* προάγουν τη *συνεργασία ερευνητών και αμοιβαία υποστήριξη χρηστών* → διερεύνηση και επέκταση των πλέον προηγμένων 'κλασικών' και 'σύγχρονων' μεθόδων και των πρακτικών εφαρμογών για την ανάλυση δεδομένων
  - Σε ένα πλήρως σχεδιασμένο, συνεκτικό περιβάλλον



## H r (η γλώσσα)



- Είναι μια **διερμηνευόμενη γλώσσα προγραμματισμού** (*interpreted computer language*) → *γλώσσα εντολών*
  - <http://cran.r-project.org/doc/manuals/R-lang.html>
  - <http://cran.r-project.org/doc/manuals/R-ints.html>
- Με debugger, code profiling (*Rprof*) ...
  - <http://cran.r-project.org/doc/manuals/R-exts.html>
- Υποστηρίζει παράλληλους υπολογισμούς
  - <http://cran.r-project.org/web/views/HighPerformanceComputing.html>

## H r (η γλώσσα)



- Είναι μια **συναρτησιακή γλώσσα** που επωφελείται από στοιχεία αντικειμενοστρεφούς προγραμματισμού (*Functional, Object-oriented language*)
  - Οι περισσότερες (στατιστικές και άλλες) συναρτήσεις που 'βλέπουν' οι χρήστες είναι γραμμένες στην ίδια τη γλώσσα του R, καλώντας ένα μικρότερο σύνολο εσωτερικών αρχετύπων (*internal primitives*)
    - + Αυτό καθιστά πολύ εύκολο στο χρήστη να κατανοήσει πως πραγματικά λειτουργούν, ενώ επιπλέον
    - + δίνεται η δυνατότητα δημιουργίας νέων πακέτων R
      - <http://cran.r-project.org/doc/contrib/Leisch-CreatingPackages.pdf>

## H r (η γλώσσα)

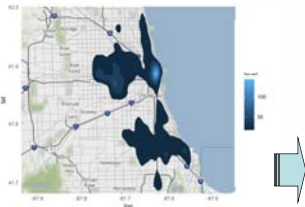


- Για κάποιες υπολογιστικά εντατικές εφαρμογές, η *r* μπορεί να θεωρείται 'αργή' (τουλάχιστον για κάποιους).
- Ωστόσο, σε τέτοιες περιπτώσεις, είναι δυνατή η διασύνδεση εντολών στην R με διαδικασίες γραμμένες σε *C*, *C++*, *FORTRAN*, *Java*, *Python* κ.ά. για μεγαλύτερη αποτελεσματικότητα, και δημιουργία επιπλέον υψηλού επιπέδου αρχετύπων
  - *System commands can be called from within R*
  - Μπορεί να χρησιμοποιηθούν είτε με κατευθείαν εντολές, είτε με προγράμματα τα οποία μπορούν να αναπτυχθούν και να καλούνται για απευθείας εκτέλεση

## To R ως GIS



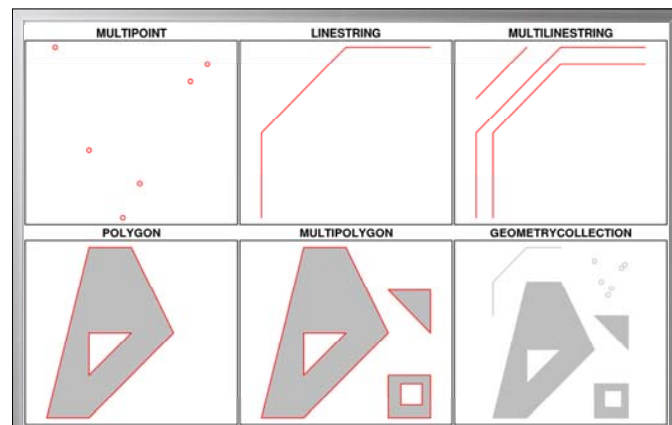
- Το λογισμικό R αυξάνει συνεχώς τις GIS δυνατότητες του, και είναι πλέον αξιόπιστο για τις περισσότερες (αν όχι όλες) τις χωρικές αναλύσεις που συνήθως κάνουμε με τα παραδοσιακά λογισμικά GIS
  - εξαγωγή και εισαγωγή raster και vector δεδομένων, σχεδίαση, ανάλυση και κατασκευή χαρτών, ...



## To R ως GIS



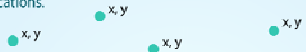
- **Κύρια διαφορά:** ενώ οι χρήστες των εργαλείων GIS αλληλεπιδρούν τυπικά με το λογισμικό μέσω point-and-click γραφικής διασύνδεσης, το R απαιτεί scripts (δηλ. γραμμές εντολών).
- Στο R, πολλά πακέτα χωρικής ανάλυσης (π.χ., πακέτο *sf*), επωφελούνται από ένα υιοθετημένο πρότυπο που ονομάζεται "**Simple Features**" το οποίο καθορίζει ένα συγκεκριμένο μοντέλο αποθήκευσης και πρόσβασης για χωρικές διανυσματικές γεωμετρίες → *συμβατότητα με χωρικές βάσεις δεδομένων (π.χ., PostGIS), αλλά και ανοικτούς τυποποιημένους μορφότυπους (π.χ. GeoJSON).*



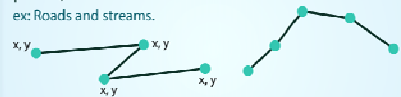
Common 'simple features' as native objects in R

### POINTS: Individual $x, y$ locations.

ex: Center point of plot locations, tower locations, sampling locations.

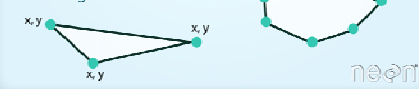


**LINES:** Composed of many (at least 2) vertices, or points, that are connected.  
ex: Roads and streams.



**POLYGONS:** 3 or more vertices that are connected and closed.

ex: Building boundaries and lakes.



- Εισαγωγή στο R αρχείων shape (*shapefiles*) επιτρέπει πλήρη χρήση διανυσματικών δεδομένων (*σημεία, γραμμές ή πολύγωνα*)
- Πακέτα *sp, rgdal, ...*

## Το R ως GIS



- Το R είναι ιδιαίτερα ευέλικτο για τη δημιουργία χαρτών → από στατικούς χάρτες υψηλής ποιότητας για εκδόσεις, μέχρι διαδραστικούς χάρτες σε ιστότοπους
- Το R ενσωματώνεται εύκολα με άλλα εργαλεία GIS ανοιχτού κώδικα όπως το SAGA, το QGIS και το GRASS



## Η δύναμη του R

- Πέρα από τις ευκολίες που παρέχονται από τη δική του καλά ανεπτυγμένη, αλλά απλή γλώσσα προγραμματισμού, πηγάζει επιπλέον από ...
- Πλούσιες πηγές υλικού για ατομική εκμάθηση:
  - <http://faculty.washington.edu/tlumley/Rcourse/>
  - <https://stats.idre.ucla.edu/r/>
  - <https://learnrthegreekway.wordpress.com/>
  - ...
- Εξαιρετικά βιβλία, οδηγοί χρήσης, άρθρα κ.ά. προσβάσιμα στο Διαδίκτυο και από την ψηφιακή βιβλιοθήκη του μαθήματος

- Το R χαρακτηρίζεται από εύκολη **επεκτασιμότητα** καθώς
  - ο κάθε ένας χρήστης μπορεί να δημιουργήσει τις δικές του συναρτήσεις και να τις μορφοποιήσει σε μια βιβλιοθήκη, η οποία με τη σειρά της
  - μπορεί να δημοσιοποιηθεί και να χρησιμοποιηθεί αργότερα από άλλους χρήστες με παρόμοιες ανάγκες
- Η **κεντρική πλατφόρμα R-Forge** προσφέρεται για τη συνεργατική ανάπτυξη των πακέτων R, όπου και φιλοξενούνται επίσης πολλά αδημοσίευτα πακέτα στην Beta έκδοσή τους
  - <http://r-forge.r-project.org/>

- Επιτρέπει την ταχεία ανάπτυξη νέων εργαλείων και προγραμμάτων σύμφωνα με τις εκάστοτε ανάγκες των χρηστών
  - Εξαιρετικό για την προσομοίωση, προγραμματισμό, και αναλύσεις δεδομένων που απαιτούν υψηλό υπολογιστικό φόρτο
  - Το R λειτουργεί ως αυτόνομη εργαλειοθήκη διαχείρισης πινακοποιημένων δεδομένων, με απόδοση συγκρίσιμη με εκείνη της Matlab
- Εξαναγκάζει το χρήστη του να σκεφθεί για τη ροή της εκάστοτε ανάλυσης που πρόκειται να εφαρμόσει.
- Διασυνδέσεις με λογισμικά αποθήκευσης δεδομένων (SQL)

- Το υβριδικό περιβάλλον του R (**προγραμματισμός + ανάλυση + γραφικά**) προσφέρει απεριόριστες δυνατότητες διαχείρισης και **στατιστικής ανάλυσης δεδομένων** καθώς και δυνατότητες κατασκευής γραφημάτων → **εξαιρετικές δυνατότητες ερμηνείας των δεδομένων**
  - Η χρήση του για αναλύσεις δεδομένων βελτιώνει γρήγορα την ουσιαστική ικανότητα του ερευνητή αλλά και την βαθύτερη αντίληψη της αλγοριθμικής θεώρησης καθώς ο χρήστης αντιλαμβάνεται τον ακριβή τρόπο με τον οποίο επεξεργάζεται τα δεδομένα

## Τι ΔΕΝ είναι το R (αδυναμίες ή πλεονεκτήματα ?)

- ΔΕΝ ΕΙΝΑΙ ένα από τα πολλά άλλα διαθέσιμα λογισμικά Στατιστικής Ανάλυσης
  - Δεν αποτελεί απλά ένα σύνολο πολύ συγκεκριμένων και άκαμπτων εργαλείων, όπως συμβαίνει συχνά σε άλλα συναφή (εμπορικά και μη) λογισμικά ανάλυσης δεδομένων
- ΔΕΝ ΕΙΝΑΙ εύκολο για το χρήστη (*τουλάχιστον αρχικά*) να εξοικειωθεί και να αξιοποιήσει πλήρως και γρήγορα όλες τις δυνατότητες του
  - Εύκολο στην εκμάθηση, αλλά απαιτεί σημαντική ενασχόληση με τη σύνταξη και τη χρήση εντολών, πακέτων κλπ. για να γίνει πραγματικά οικείο σε κάποιον

## Τι ΔΕΝ είναι το R (αδυναμίες ή πλεονεκτήματα ?)

- ΔΕΝ ΕΙΝΑΙ Menu-driven
  - Υπολείπεται ενός φιλικού παραθυρικού περιβάλλοντος χρήσης, αλλά συνδέεται εύκολα με εξαιρετικά περιβάλλοντα διεπαφής GUI
    - Η επικοινωνία του χρήστη με το R μέσω εντολών, αρχικά φαίνεται τουλάχιστον 'επίπονη'
- ΔΕΝ ΠΑΡΕΧΕΙ πλούσια γραφική διεπαφή
  - αλλά συνδέεται με Java, και διαθέτει ποικιλία συναρτήσεων που παρέχουν ένα ευέλικτο γραφικό περιβάλλον για την οπτικοποίηση δεδομένων
- ΔΕΝ ΛΕΙΤΟΥΡΓΕΙ ως βάση δεδομένων
  - αλλά συνδέεται με συστήματα διαχείρισης βάσεων δεδομένων (DBMS)

## Πως είναι διαθέσιμο;

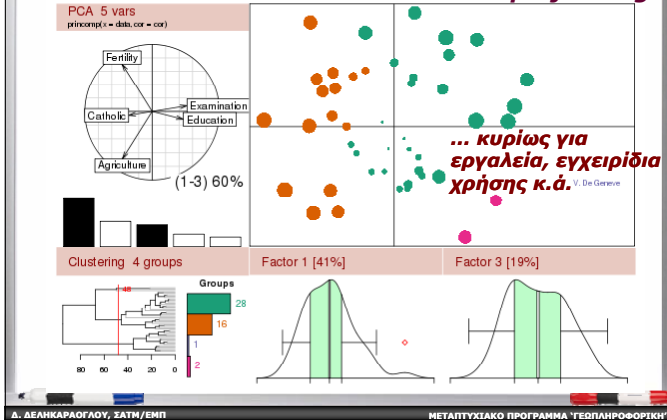


- Διατίθεται ελεύθερα
  - Η βασική σουίτα του R και τα περισσότερα εξειδικευμένα πακέτα του είναι διαθέσιμα από το δίκτυο διανομής CRAN - [cran.r-project.org](http://cran.r-project.org)
    - Mirror sites σε πολλές χώρες (και στην Ελλάδα)
    - Νέες κυκλοφορίες του βασικού λογισμικού R παρέχονται σε συχνά διαστήματα (τυπικά περίπου κάθε 1-3 μήνες !!!), ανάλογα και με το λειτουργικό σύστημα (*Windows, MacOS, Linux*)
  - Πλείστα user blogs, tutorials, user guides ...
    - Η κοινότητα των χρηστών του R είναι πολύ ενεργή



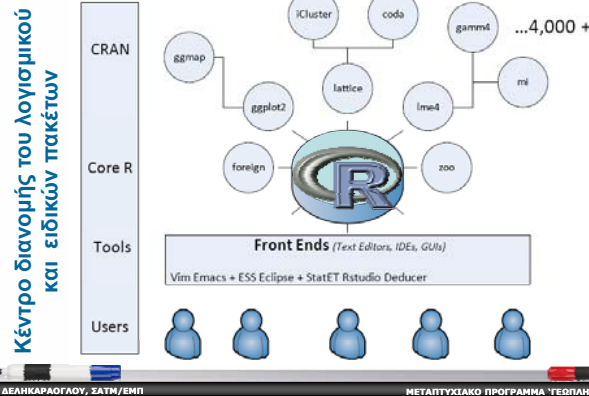
# R Project for Statistical Computing

www.r-project.org

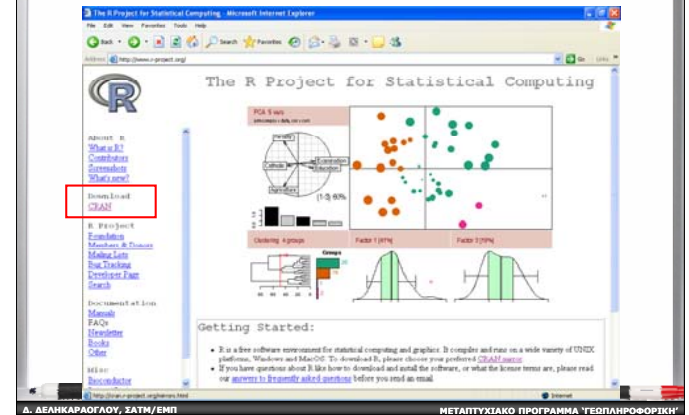


# Comprehensive R Archive Network

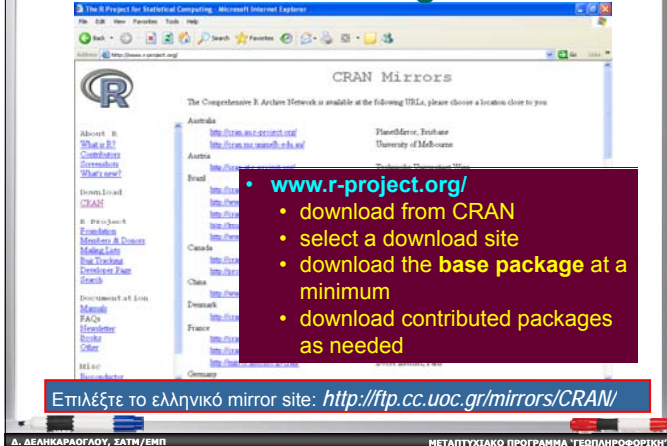
cran.r-project.org



# Downloading R

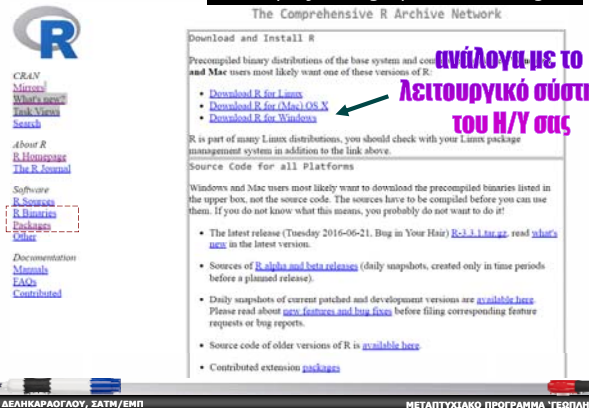


# Downloading R



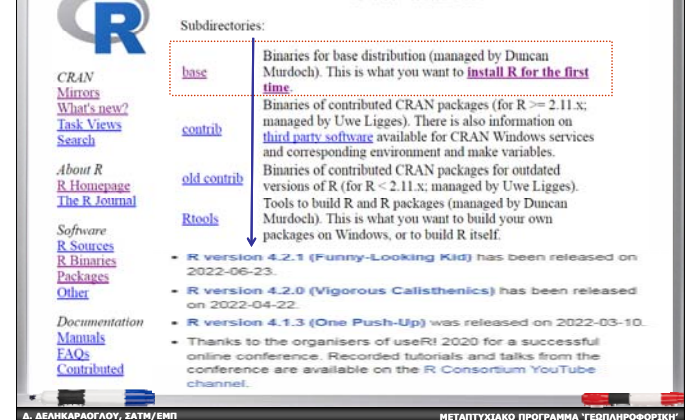
# Ιστότοπος διανομής του R και πακέτων

cran.r-project.org ή cran.cc.uoc.gr

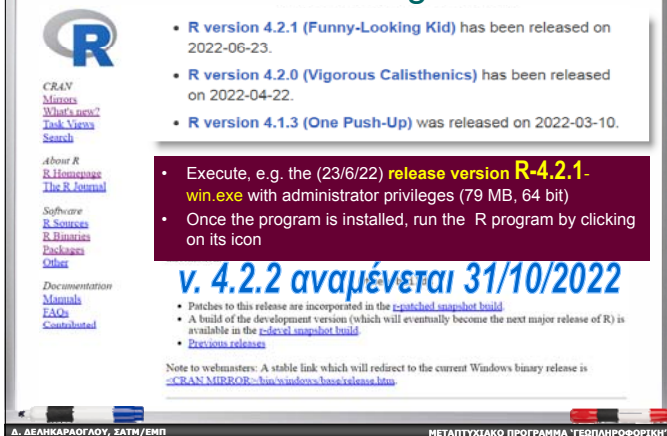


# Downloading R

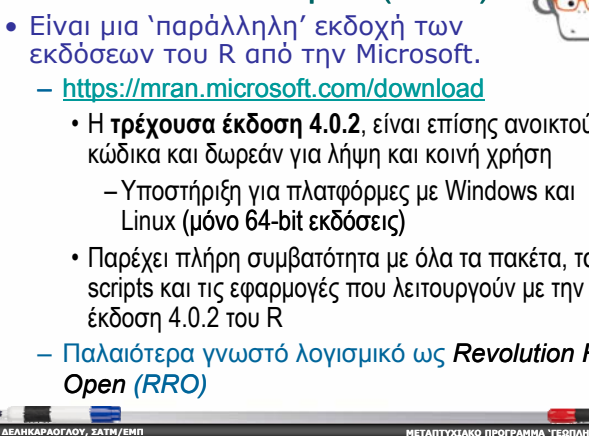
R for Windows



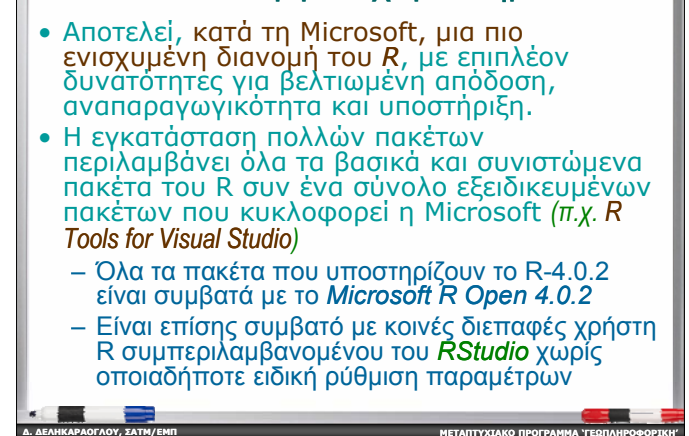
# Downloading R



# Microsoft R Open (MRO)



# MRO λειτουργικά χαρακτηριστικά



- Εφαρμογές που περιλαμβάνουν τη δυνατότητα λειτουργιών του R είναι επίσης συμβατές με το Microsoft R Open.
  - Οι χρήστες μπορούν να ενσωματώνουν τις δυνατότητες του R στις δικές τους εφαρμογές, μέσω του *DeployR Open*
  - Με την εγκατάστασή του, το MRO χρησιμοποιεί ένα προεπιλεγμένο στιγμιότυπο του αποθετηρίου CRAN που παρέχει ένα σταθερό και στατικό σύνολο πακέτων σε όλους τους χρήστες της Microsoft R Open
    - Αυτό καθιστά ευκολότερη την κοινή χρήση R κώδικα που βασίζεται στα πακέτα R και μειώνει την πιθανότητα εγκατάστασης ασύμβατων R πακέτων στο ίδιο σύστημα.

- Για βελτίωση της απόδοσης του R, το Microsoft R Open περιλαμβάνει βιβλιοθήκες μαθηματικών συναρτήσεων (modules) πολλαπλών νημάτων/σπειρωμάτων (*multi-threaded math libraries*)
  - Παράλληλοι υπολογισμοί και να χρησιμοποιείται όλη η διαθέσιμη ισχύ επεξεργασίας για τη μείωση του χρόνου υπολογισμού περίπλοκων αλγορίθμων.
- Επιπρόσθετα διατίθεται το MRS (*Microsoft R Server*) που επιλύει δύο προβλήματα που σχετίζονται με τη χρήση του R: χωρητικότητα (χειρισμός του μεγέθους συνόλων δεδομένων και μοντέλων) και ταχύτητα

## Φορητή έκδοση του R (και RStudio)

- Το R (*όπως και το RStudio*) είναι επίσης δυνατόν να εκτελεστούν από μια φορητή μονάδα USB αντί να τα εγκατασταθούν προηγουμένως σε κάποιο Η/Υ.
  - Χρήσιμο π.χ. όταν δεν έχετε δικαιώματα διαχειριστή σε κάποιον υπολογιστή
  - Τα δεδομένα και οι ρυθμίσεις θα παραμείνουν στον τοπικό Η/Υ
- <https://sourceforge.net/projects/rportable/>, έκδοση 4.2.0
- [https://github.com/ClaudiaBrauer/A-very-short-introduction-to-R/blob/master/documents/Portable\\_versions\\_of\\_R\\_and\\_RStudio.pdf](https://github.com/ClaudiaBrauer/A-very-short-introduction-to-R/blob/master/documents/Portable_versions_of_R_and_RStudio.pdf)

## Έκδοση pqR (Pretty Quick R)

- Μια νέα έκδοση του *R interpreter* (*pqR-2020-07-23*, για *Linux, Unix, Mac, Windows*). Βασίζεται στις νεώτερες εκδόσεις (>2.15.0) που διανέμονται από το αποθετήριο CRAN του R
  - <http://www.pqr-project.org/>
- Επιταχύνει πολλές από τις λειτουργίες του R, καθώς και τη γενική λειτουργία του διερμηνέα του.
- Αυτόματα εκτελεί ορισμένους αριθμητικούς υπολογισμούς παράλληλα σε επεξεργαστές πολλαπλών πυρήνων.
- Επεκτείνει τη γλώσσα *r*.

## Άμεσα βοηθήματα για τη R

<https://cran.r-project.org/other-docs.html>

English Documents

Documents with more than 100 pages:

- “Visual Statistics. Use R!” by Alexey Shipunov (PDF, 2016-06-06, 301 pages). All book materials are accessible from [Alexey Shipunov's English R page](#).
- “Using R for Data Analysis and Graphics - Introduction, Examples and Commentary” by John Maindonald (PDF, data sets and scripts are available at [JM's homepage](#)).
- “Practical Regression and Anova using R” by Julian Faraway (PDF, data sets and scripts are available at the [book homepage](#)).
- The [Web Appendix](#) to the book “An R Companion to Applied Regression” (second edition) by John Fox and Sanford Weisberg contains information about R to fit a variety of regression models.
- “An Introduction to S and the Hmisc and Design Libraries” by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. (PDF)

<https://cran.r-project.org/manuals.html>

R-release

An Introduction to R is based on the former “Notes on R”, gives an introduction to the language and how to use R for doing statistical analysis and graphics. [HTML](#) | [PDF](#) | [EPUB](#)

R Data Import/Export describes the import and export facilities available either in R itself or via packages which are available from CRAN. [HTML](#) | [PDF](#) | [EPUB](#)

R Installation and Administration [HTML](#) | [PDF](#) | [EPUB](#)

Writing R Extensions covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces. [HTML](#) | [PDF](#) | [EPUB](#)

A draft of The R language definition documents the language *per se*. That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions. [HTML](#) | [PDF](#) | [EPUB](#)

R Internals: a guide to the internal structures of R and coding standards for the core team working on R itself. [HTML](#) | [PDF](#) | [EPUB](#)

The R Reference Index: contains all help files of the R standard and recommended packages in printable form. (9MB, approx. 3500 pages) [PDF](#)

## Άμεσα βοηθήματα για τη R

<https://cran.r-project.org/other-docs.html>

- Short Documents and Reference Cards:
- “R reference card” by Jonathan Baron (PDF).
  - “R and Octave” by Robin Hankin (Text), a reference sheet translating between the most common *Octave* (or *Matlab*) and R commands.
  - A “time series reference card” (PDF) and a “regression reference card” (PDF) by Vito Ricci.
  - “R reference card” by Tom Short (PDF, LaTeX source).
  - “R reference card data mining” by Yanchang Zhao (PDF, 2013-01-05).
  - “A (very) short Introduction to R” by Paul Torfs and Claudia Brauer (PDF, 2014-03-18).
  - “R reference card v2” by Matt Baggott (PDF, 2013-01-05).
- Για περισσότερα βοηθήματα συμβουλευτείτε την ψηφιακή βιβλιοθήκη του μαθήματος ...

## Άμεσα βοηθήματα για τη R

Greek <https://cran.r-project.org/other-docs.html>

- “Μια (πολύ) σύντομη εισαγωγή στην R” by Paul Torfs and Claudia Brauer (PDF, 2015-11-06), a Greek translation by George Kaffezas of “A (very) short Introduction to R”.
- “Lecture Notes for Introduction to R” by Konstantinos Fokianos and Charalambos Charalambous (PDF), a semester course notes for introductory statistical theory and computations by R.
- “Statistical Analysis with R” by Anastasios Katsileros (PDF), data analysis of agronomic experiments for Greek speaking R users.

## Άλλες πηγές βοήθειας για το R ...

- Οδηγίες εγκατάστασης του R, στο Youtube
  - <http://www.youtube.com/watch?v=ZoPJGmpYJzw&feature=related>
- r-podcast* περιέχει συνδέσμους για βίντεοσυζητήσεις που σχετίζονται με το R
  - <http://www.r-podcast.org/>
- Μέσω της μηχανής αναζήτησης RSeek
  - <http://www.rseek.org/>
- Το forum StackOverflow είναι μια καλή πηγή για την αναζήτηση απαντήσεων σε ερωτήματα που σχετίζονται με το R
  - <http://stackoverflow.com/questions/tagged/r>

# TutoRials

- Από τον ιστοχώρο του R στην ενότητα "Contributed Documentation"
  - Tutorials and manuals created by R users
    - Simple R
    - R for Beginners
    - Practical Regression and ANOVA Using R
  - R FAQ
  - Mailing Lists (listserv)
    - r-help

## Το καλύτερο χαρακτηριστικό του R!!! Η τεράστια ποικιλία διαθέσιμων πακέτων του

```
install.packages(" ")
```

και μεταξύ τους πολλά  
γεωπεμπιστημονικού  
ενδιαφέροντος:

- Ecological and Environmental Data
- ...

< 18725 πακέτα, μέχρι  
πρόσφατα  
(7300 το 2015, 8900 το  
2016, ..., 14700 το  
2019) !!!

<http://cran.r-project.org/web/views/>

- Κάθε πακέτο είναι μια συλλογή από αντικείμενα που αναγνωρίζει και μπορεί να χρησιμοποιήσει το R
  - Μπορεί να αποτελείται από συναρτήσεις, σύνολα δεδομένων, ή άλλου τύπου αντικείμενα, όπως δυναμικά προσβάσιμες βιβλιοθήκες έτοιμου προς εκτέλεση κώδικα εντολών
- Παρόλο που *το R συνοδεύεται από ένα σύνολο πακέτων από προεπιλογή* (δηλ. με την εγκατάστασή του), υπάρχουν πολλά ακόμα πακέτα που μπορούν να προστεθούν για να επεκτείνουν τις δυνατότητες του R

## 14/10/2022 διαθέσιμα ~18725 πακέτα!!

Μπορούν να αναζητηθούν με την ονομασία τους, συγγραφέα, ή τίτλο ...

## 18725 CRAN packages ... and counting

Date	Package	Title
2018-08-09	<a href="#">Comintake</a>	Flexible Tools for Multiple Imputation
2018-08-09	<a href="#">EedData</a>	Functions to Automate Downloading Geospatial Data Available from Several Federated Data Sources
2018-08-09	<a href="#">HMDHEdplus</a>	Read Human Mortality Database and Human Fertility Database Data from the Web
2018-08-09	<a href="#">logKDE</a>	Computing Log-Transformed Kernel Density Estimates
2018-08-09	<a href="#">nlqs</a>	Semi-Confirmatory Structural Equation Modeling via Penalized Likelihood
2018-08-09	<a href="#">relsurv</a>	Relative Survival
2018-08-09	<a href="#">stlPacales</a>	Spatio-Temporal Uncertainty Propagation Methods for Multiple Scales
2018-08-09	<a href="#">tmIrr</a>	The Integration Between IBM COGNOS TM and R
2018-08-09	<a href="#">VariableScreening</a>	High-Dimensional Screening for Semiparametric Models
2018-08-08	<a href="#">anGRtschinge</a>	Teaching Hydrological Modelling with the GR Rainfall-Runoff Models ('Shiny' Interface Included)
2018-08-08	<a href="#">hWGB</a>	Bayesian Whole-Genome Regression
2018-08-08	<a href="#">CarletonStats</a>	Functions for Statistics Classes at Carleton College
2018-08-08	<a href="#">cOde</a>	Automated C Code Generation for 'deSolve', 'bvpSolve' and 'Sundials'

Μπορούν να αναζητηθούν με την ονομασία τους, συγγραφέα, ή τίτλο ... και με ημερομηνία δημοσίευσής

## Μια πρώτη γεύση χρήσης της r

- Χωρίς να είναι απαραίτητο να έχετε εγκαταστήσει το R στον Η/Υ σας
- Χρησιμοποιήστε έναν από τους WebServers του R
  - <http://users.stat.umn.edu/~geyer/r.html#rweb>
  - <https://rweb.webapps.cla.umn.edu/Rweb/Rweb.general.html>
  - <http://claree.univ-lille1.fr/Rweb/>
- Για δοκιμή: εισάγετε τις απλές εντολές από το αρχείο *my1stRexample.R* (από τις ιστοσελίδες του μαθήματος)
  - Ετοιμαστείτε να σχολιάσετε τα πρώτα σας αποτελέσματα ή απορίες σας, στο επόμενο μάθημα

## Συμπερασματικά ...

- Η στατιστική γλώσσα *r* και το ανοικτό λογισμικό *R* αποτελούν **πολύτιμα εργαλεία διαχείρισης δεδομένων**, συμπεριλαμβανομένης της ανάλυσης και επεξεργασίας χωροχρονικών δεδομένων. Εξελίσσονται με φρενήρη ρυθμό ανάπτυξης και εξυπηρετούν κυρίως
  - τις ανάγκες στατιστικής ανάλυσης, αλλά και
  - την εφαρμογή σύγχρονων μεθοδολογιών αιχμής για την ανάλυση μεγάλου όγκου δεδομένων, καθώς και
  - την οπτικοποίηση δεδομένων μέσω γραφημάτων υψηλής αισθητικής

## Την επόμενη φορά ...

- Θα αναφερθούμε λεπτομερέστερα στο αυτόνομο περιβάλλον χρήσης του *R*, την κονσόλα του, και ...
- Στα εναλλακτικά περιβάλλοντα εργασίας (GUI), RGui, RCommander, RStudio, ...
  - Στοιχεία σύνταξης εντολών στο *R*
  - ...



## Μέχρι την επόμενη φορά ...

- Ξεκινήστε τη δική σας διαδρομή εξοικείωσης με το R
- Κατεβάστε από το κεντρικό αποθετήριο *CRAN* και εγκαταστήστε την τρέχουσα έκδοση του *R* στον Η/Υ σας
- Κατανοήστε το οικοσύστημα των πακέτων του *R* επισκεπτόμενοι το *CRAN Task Views*
- Ανατρέξτε στα πρώτα βοηθήματα σας από τη βιβλιογραφία ή τις ιστοσελίδες του μαθήματος

