



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Ανάλυση και Σχεδιασμός Μεταφορών I Εισαγωγή στην R

Παναγιώτης Παπαντωνίου
Δρ. Πολιτικός Μηχανικός, Συγκοινωνιολόγος

Πάτρα, 2017

Τι είναι η R;

- Η R είναι μία διάλεκτος της γλώσσας S

Τι είναι η S;

- Η S είναι μια γλώσσα που ξεκίνησε να αναπτύσσεται το 1976 από τον John Chambers και άλλους στα εργαστήρια της Bell με βάση τη γλώσσα Fortran.
- Το 1988 ξαναγράφηκε με βάση τη γλώσσα C



Εισαγωγή

- Ξεκίνησε το 1991, ανακοινώθηκε στο κοινό το 1993
- Πήρε το όνομα της από τα αρχικά των δημιουργών της Ross Ihaka και Robert Gentleman από το Πανεπιστήμιο του Όκλαντ, Νέα Ζηλανδία
- Ίσως και για να θυμίζει λίγο τη γλώσσα S
- Πλέον η R αναπτύσσεται από την R Development Core Team, στην οποία ο Chambers είναι μέλος



Μειονεκτήματα της R

- Βασίζεται (απαραίτητα) σε τεχνολογία 40 ετών
- Μικρή ενσωματωμένη υποστήριξη για δυναμικά ή 3D γραφικά
- Η λειτουργικότητα βασίζεται στη ζήτηση και στη συνεισφορά των χρηστών. Αν κανένας δεν ενδιαφέρεται για αυτό που κάνεις ή την αγαπημένη σου μέθοδο, τότε είναι δική σου δουλειά να το κάνεις να δουλέψει!
- Όχι ιδανικό για όλες τις πιθανές καταστάσεις



Ξεκινώντας με την R

- Πληροφορίες για την R
<http://www.r-project.org>
- Εγκατάσταση της R:
<http://cran.r-project.org>
- Εγκατάσταση του RStudio:
<http://www.rstudio.com>



R studio

The screenshot displays the RStudio environment with the following components:

- Code Editor:** Contains R code for data loading and visualization. The code includes comments in Greek and English, and uses functions like `read.csv()`, `ggplot()`, `geom_point()`, `t.test()`, `var.test()`, `plot()`, `abline()`, and `lines()`.
- Console:** Shows the R startup message for version 3.2.0 (2015-04-16) and the first line of code being executed: `New_Car_Sales_counter <- read.csv("New_Car_Sales.csv", sep=";")`.
- Environment Pane:** Shows the `Data` environment with a single object: `New_Car_Sales_count` (136 observations of 54 variables).
- Package Manager:** Lists installed and available packages in the User Library, including `abind`, `acepack`, `aplpack`, `bitops`, `car`, `caTools`, `colorspace`, `corpcor`, `corrplot`, `d3Network`, `dichromat`, `digest`, `e1071`, and `effects`.



Ονόματα μεταβλητών

- Κάθε πράξη μπορεί να γραφτεί σαν μία μεταβλητή ($c=a+b*2$)
- Οι μεταβλητές μπορούν να γραφούν σαν ένας οποιοσδήποτε συνδυασμός γραμμάτων, αριθμών, τελειών και κάτω παύλας(Kykl_Fortos.1)
- Τα ονόματα των μεταβλητών δεν θα πρέπει να ξεκινούν με αριθμούς ή σύμβολα (1Kykl_Fortos,_KyklFortos1)
- Τα ονόματα των μεταβλητών δεν θα πρέπει να περιέχουν κενά (Kykl Fortos 1)
- Η R αντιλαμβάνεται τα κεφαλαία ως διαφορετικούς χαρακτήρες [case sensitive] (KyklFortos \neq kyklfortos)



Διανύσματα

- Αριθμητικά
- Λογικά
- Χαρακτήρες
- Σύνθετα

Χρήση εντολή `c()` για τη δημιουργία διανυσμάτων

```
Console ~/ ↩
> x<-c(0.5,0.6)      ## numeric
> x<-c(TRUE,FALSE)  ## logical
> x<-c(T,F)         ## logical
> x<-c(9:29)        ## interger
> x<-c(1+1i,2-4i)   ## complex
> |
```



Πίνακες

- Οι πίνακες είναι διανύσματα με διαστασιακή λειτουργία

```
Console ~/ ↵
> m<-matrix(nrow=2,ncol=3)
> m
      [,1] [,2] [,3]
[1,]  NA  NA  NA
[2,]  NA  NA  NA
> dim(m)
[1] 2 3
> m<-matrix(1:6,nrow=2,ncol=3)
> m
      [,1] [,2] [,3]
[1,]   1   3   5
[2,]   2   4   6
> m<-1:10
> dim(m)<-c(2,5)
> m
      [,1] [,2] [,3] [,4] [,5]
[1,]   1   3   5   7   9
[2,]   2   4   6   8  10
> dim(m)<-c(5,2)
> m
      [,1] [,2]
[1,]   1   6
[2,]   2   7
[3,]   3   8
[4,]   4   9
[5,]   5  10
> |
```



Βρόγχοι

- Με τους βρόγχους (loops) μπορούμε εκτελούμε κάποια διαδικασία για έναν αριθμό επαναλήψεων (iterations).

```
> x<-1:10
> new_v<-NULL
> counter=1:length(x)
> for (i in counter)
+ new_v[i]=x[i]^2
> new_v
[1]  1  4  9 16 25 36 49 64 81 100
> |
```

```
> class<-NULL
> for (i in counter)
+ {
+   if(new_v[i]<=50)
+   {
+     class[i]<-"Small"
+   } else{
+     class[i]<-"Big"
+   }
+ }
> class
[1] "Small" "Small" "Small" "Small" "Small" "Small" "Small" "Small" "Big"  "Big"  "Big"
>
```

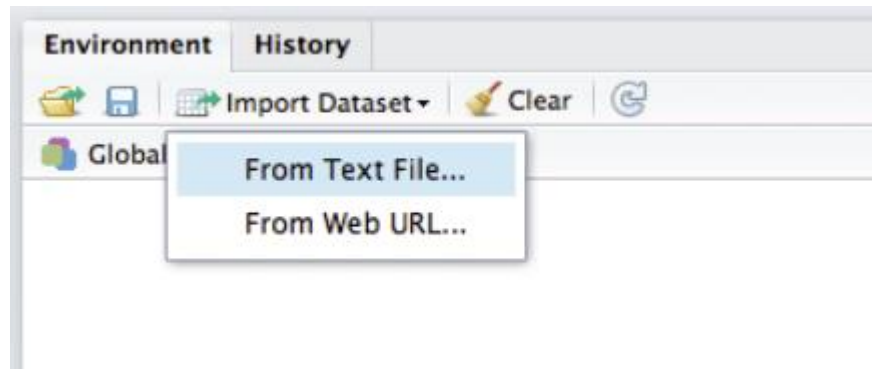


Εισαγωγή δεδομένων

- Η R δουλεύει πολύ εύκολα με δεδομένα που αποθηκεύονται σαν αρχεία κειμένου. Οι τιμές σε αυτά τα αρχεία χωρίζονται μεταξύ τους είτε με κόμμα είτε με κενό

```
Console ~/ ↵  
> #Comma Separated Values (CSV)  
> dat.csv<-read.csv("http://www.ats.ucla.edu/stat/data/hsb2.csv")  
> #Tab Separated Values  
> dat.tab<-read.table("http://www.ats.ucla.edu/stat/data/hsb2.txt",header=TRUE,sep="\t")  
>
```

- Στο RStudio η εισαγωγή μπορεί να γίνει και με πιο εύκολο τρόπο



Πίνακας τελεστών

Περιγραφή	R σύμβολο	Παράδειγμα
Σχόλια	#	# αυτό είναι σχόλιο
Απόδοση τιμής	<-	x<-sin(pi/2)
Τελεστής παράθεσης	c	c(1,3.5,2.8)
Στοιχείο επί στοιχείο πολλαπλασιασμός	*	a*b
Ύψωση σε δύναμη	^	a^b
Υπόλοιπο διαίρεσης ακεραίων x/y	%%	17%%5
Πηλίκο διαίρεσης ακεραίων x/y	%/%	17%/5
Ακολουθία αριθμών από a έως b ανά h	seq	seq(0,20,2)
Ακολουθία αριθμών ανά 1	:	0:20



Πίνακας συνήθων συναρτήσεων

Περιγραφή	R σύμβολο	Παράδειγμα
ημίτονο, συνημίτονο, εφαπτομένη	<code>sin, cos, tan</code>	<code>sin(pi/3) (=0.866..)</code>
τόξο ημιτόνου, συνημιτόνου, εφαπτομένης	<code>asin, acos, atan</code>	<code>atan(seq(0,1,.25))/pi</code>
τετραγωνική ρίζα	<code>sqrt</code>	<code>sqrt(x)</code>
Εκθετική συνάρτηση e^x	<code>e^x</code>	<code>exp(x)</code>
παραγοντικό	<code>n!</code>	<code>factorial(n)</code>
τυχαίοι αριθμοί στο (0,1)	<code>runif</code>	<code>runif(100)</code>
τυχαίοι κανονικοί αριθμοί	<code>rnorm</code>	<code>u=rnorm(100000,2,4)</code>
κανονική κατανομή	<code>pnorm, dnorm</code>	<code>pnorm(1,2,4)</code>
βαθμίδες, διάταξη	<code>rank, sort</code>	<code>z=floor(10*runif(10));z; rank(z);sort(z)</code>
διασπορά, συνδιασπορά	<code>var, cov</code>	<code>var(u), cov(x,y)</code>
τυπ. απόκλιση, συντ. συσχέτισης	<code>sd, cor</code>	<code>sd(u), cor(x,y)</code>



Βασικές εντολές γραφημάτων

Όνομα	Παράμετροι	Γράφημα
plot	x, y	Σχηματίζει το διάγραμμα διασποράς της y ως προς x.
lines points		Προσθέτει γραμμές στο προηγούμενο γράφημα. Καθορίζει το χαρακτήρα που τυπώνει στα σημεία. Υπάρχουν 18 χαρακτήρες (pch=n, n=1,2,...,18)
hist	x	Σχηματίζει το ιστόγραμμα του x
barplot	height	Σχηματίζει ραβδόγραμμα με ύψη σύμφωνα με το height
tsplot tspoints tslines	x	Σχηματίζει το γράφημα της χρονοσειράς x Καθορίζονται τα σημεία χρονοσειράς Καθορίζεται ο τύπος των γραμμών της χρονοσειράς
matplot	X, Y	Σχηματίζει διαγράμματα διασποράς των στηλών του πίνακα X προς τις αντίστοιχες του Y. Αρχικά διακρίνει τα σημεία με "1" για τις α' στήλες, "2" για τις β' κλπ.
pairs	X	Σχηματίζει όλα τα ανά ζεύγη στηλών του πίνακα X διαγράμματα διασποράς, για συνολική εποπτεία
abline	a, b reg	Προσθέτει στο προηγούμενο μία γραμμή με εξίσωση $y=a+bx$ της οποίας δίνονται τα a και b. Προσθέτει στο διάγραμμα διασποράς τη γραμμή παλινδρόμησης της y στην x, όπου <code>reg <- lsfit(x, y)</code>
box	n	Περικλείει το προηγούμενο γράφημα σε ορθογώνιο με γραμμές τόσο έντονες όσο μεγαλύτερο το n
boxplot stem	x	Σχηματίζεται το θηκόγραμμα του x Σχηματίζεται το φυλλογράφημα του x
pie	x	Σχηματίζει κυκλικό διάγραμμα (πίττα) για το x
qqplot qqnorm	x, y x	Σχηματίζεται κανονικό πιθανοτικό διάγραμμα, δηλαδή διάγραμμα σε normal probability paper.
usa()		Σχηματίζει το χάρτη των ηνωμένων πολιτειών





ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Ανάλυση και Σχεδιασμός Μεταφορών I Εισαγωγή στην R

Παναγιώτης Παπαντωνίου
Δρ. Πολιτικός Μηχανικός, Συγκοινωνιολόγος

Πάτρα, 2017