R programming

Basics of R language

About me: Jan Gorecki

Computer Science @ Warsaw School of Information Technology, Poland

Since then slowly moving from Databases to Data Analytics

2009+	Databases
2011+	Data Warehousing, Business Intelligence
2013+	Data Analytics, R programming
2015+	Open source R development: **data.table**, **H2O**

What is R?

- Programming language and environment for statistical computing
- First released in 1993
- As a statistical software it is difficult to use
- As a programming language it is easy to use

Features

- Vector as atomic data type
- Open source, free to use and extend without asking for permission (GPL-2)
- Interactive - no compiler
- Community - thousands of R packages (CRAN, Bioconductor, GitHub, others)
- Visualisation (graphics, lattice; R packages: ggplot2, rgl, others)
- Computing on the language (metaprogramming)

Limitations

- Memory management
- Security

R extensions (packages) capabilities

- Import and export data from/to various file formats and databases
- Efficient data cleansing and transformation
- Plotting multidimensional data using multi panel charts or 3D graphs
- Native support for missing values
- Statistical modeling
- Signal processing
- Distributed parallel computing
- Machine learning
- Time series data support
- Spatial data support
- much more...

Install R

- R-project website download
- Optional RStudio IDE download

Start R

- Windows: "C:\Program Files\R\R-3.3.2\bin\x64\Rgui.exe"
- MacOSX: R
- Linux: R

Assignment and basic vector examples

```
x <- 1
y <- 5
sum(x, y)
x + y
length(x)
z <- c(1, 5)
length(z)
sum(x, z)
x + z # element wise with recycling
sum(z, z)
z + z # element wise
```

Atomic data types

```
# integer
1L
# real (numeric)
1.5
# string (character)
"a"
# logical
TRUE
# complex (imaginary numbers)
1i
# raw (binary type)
as.raw(10)
```

Sequences

```
# integer
x <- c(1, 2, 3, 4, 5)
x <- seq(1, 5)
y <- 6:10
x * y
# numeric
x <- c(1, 1.5, 2, 2.5, 3)
x <- seq(1, 3, by = 0.5)
# logical
lgc <- c(TRUE, FALSE, TRUE)
# character
chr <- letters # R built-in, same as c("a", "b", ..., "z")
```

Operations on vectors

```
-x

lgc <- x < 2
!x < 2 # negation
x >= 2

lgc & !lgc # AND
lgc | !lgc # OR

chr <-  c("a", "b", "c", "d", "e")
paste(x, chr) # element wise
```

Subsetting using integer type

```
x[1]
x[2:4]
x[-(2:4)]
x[2:10]
x[c(1, 2, 3, 5, 4)]

lgc[2:4]
lgc[-5]

chr[1:3]
chr[10]
chr[c(1, 2, 3, 2, 1)]
```

Subsetting using logical type

```
x > 2
x[x > 2]
x[c(FALSE, TRUE, TRUE, FALSE, FALSE)]
x[c(FALSE, TRUE, TRUE)] # unexpected result due to recycling
x[x < 2 | x >= 3]

# %in% operator
chr %in% c("d", "e", "f")
chr[chr %in% c("d", "e", "f")]
chr[chr >= "d"] # OS locale specific!
```

Names and subsetting using character type (names)

```
x
chr
names(x) <- chr
x
x <- c(a = 1.0, b = 1.5, c = 2.0, d = 2.5, e = 3.0)

x["a"]
x[c("d", "e")]
x[c("d", "f")]
```

Modify elements in vector (sub-assign)

```
y[1] <- 100
y
y[c(8, 10)] <- c(5, 6)
length(y)
sum(y)
is.na(y)
y[!is.na(y)]
y <- c(y, 7)
y[y > 50] <- NA
y <- y[!is.na(y)]
```

Matrices and arrays

```
mx <- matrix(1:25, 5, 5)
mx
mx <- 1:25
dim(mx) <- c(5, 5)
mx
mx[1:2, 1:3]

ar <- 1:27
dim(ar) <- c(3, 3, 3)
ar
str(ar)
ar[1:2, 1:3, 2:3]
ar[1, 1:3, 2:3, drop = FALSE]
```

Lists and data frames

```
lst <- list(1:5, letters[1:3], c(TRUE, FALSE))
lst
str(lst)
names(lst) <- c("a1", "a2", "a3")
lst$a1

df <- data.frame(c1 = 1:5, c2 = letters[1:5], c3 = c(TRUE,
FALSE, TRUE, TRUE, TRUE))
df <- rbind(df, df, df, df, df)
df
str(df)
head(df)
```

Base plot

```
x <- rnorm(50)
y <- rnorm(50)
plot(x, y)

class(mtcars)
head(mtcars)
attach(mtcars)
plot(wt, mpg)
abline(lm(mpg ~ wt))
title("Regression of MPG on Weight")
detach(mtcars)
```

Getting help

- Function manuals, use question mark in front of function name: `?sum`, `?"["`
- R manuals: R-intro, Manuals
- R packages vignettes (tutorials)
- Post question on stackoverflow.com, use R tag, make reproducible example
- Read examples in blog posts - R blogs aggregator: r-bloggers.com
- Read R mailing lists

Questions?

Jan Gorecki: github.com/jangorecki

Contact: **jan61ji@gmail.com**