



Introduction to R programming

Rita Giordano

rgiordano@gmx.com

R introduction

R is a programming language for statistical computing.

Useful website:

- <https://www.r-project.org>
- <https://cran.r-project.org>
- <https://www.r-bloggers.com>
- https://en.wikibooks.org/wiki/R_Programming

Packages

- DISP (D. Waterman): to analyse X-ray images from CCD or Pilatus detector.
- CRy (J. Foadi): to analyse mtz file...work in progress!!!
- bio3d - protein structure analysis
<https://thegrantlab.org/bio3d/index.php>
- ggplot2 - grammar for graphics

How to assign values to a variable and read data from a file

```
x <- 5  
x
```

```
## [1] 5
```

<- this symbol is the assignment operator

read.table("file_path")

Control flow

- if -else:

```
if (condition) {expression}  
else
```

- for:

```
for (var in seq) {expression}
```

- while:

```
while(condition) {expression}
```

Control flow

- repeat:

```
repeat {expression}
```

- break: to stop loop
- next: to skip iteration
- return: to exit from a loop and return a value

R function

An R function is defined as global environment:
.GlobalEnv.

```
add <- function(x,y)
{
  x+y
}
```

```
environment(add)
```

```
## <environment: R_GlobalEnv>
```

Introduction

Welcome to this introduction to R programming for crystallography

Look which is your working directory.

```
getwd()
```

```
## [1] "/Users/ritagiordano/Documents/ECM30"
```

Now list all files present in your working directory.

```
dir()
```

```
## [1] "File.R"           "Fscopelv2.png"
## [3] "Fscopelv2.svg"    "Introduction.Rmd"
## [5] "Introduction.html" "R.jpg"
## [7] "R_for_crystallography" "Rpicture.pdf"
## [9] "Rpicture.svg"      "Rprogramming.key"
## [11] "Tutorial"          "ecm30-logo.png"
## [13] "mytest.R"          "rlogo.png"
## [15] "rprog.png"         "test"
## [17] "test1"             "test_xray"
```

Create a new directory test_xray in your working directory.

```
dir.create("test_xray")
```

```
## Warning in dir.create("test_xray"): 'test_xray' already exists
```

List all objects in your workspace

```
ls()
```

```
## [1] "course_dir"      "dest_dir"      "destrmd"      "initcode"
## [5] "initpath"        "install_course" "keep_rmd"      "les"
## [9] "lessonPath"      "meta"          "open_html"     "out"
## [13] "quiet"           "rmd_filename"  "unit"
```

change the current working directiry in the new one that you already created.

```
setwd('test_xray')
```

Create now a new R file in your new directory.

```
file.create("mytest.R")
```



```
## [1] TRUE
```

Open now your new file for editing.

```
file.edit("mytest.R")
```

Read Data with R

Open the help for reading tabular data.

```
?read.table
```

Open an XDS file using the function read.table().

```
xds=read.table('Files/XDS.HKL', comment.char='!')
```

Rename the column name of the file.

```
colnames(xds) <- c('H', 'K', 'L', 'Iobs', 'SigmaIobs', 'Xd', 'Yd', 'Zd', 'RPL', 'PEAK', 'CORR', 'PSI')
```

Create a new data frame with only H, K, L, Iobs and SigmaIobs.

```
xds_new <- subset(xds, select=c('H', 'K', 'L', 'Iobs', 'SigmaIobs'))
```

Write the new data frame in a new file called XDS_new.HKL.

```
write.table(xds_new, 'Files/XDS_new.HKL')
```

Open a csv file called spot, which contains the XY coordinates, the peak values and other variables.

```
spot<-read.csv('Files/spot.csv')
```

Open again the csv file spot.csv and skip the first column

```
spot<-read.csv('Files/spot.csv', colClass=c('NULL', NA, NA, NA, NA, NA, NA, NA))
```

Rename the column of the file

```
colnames(spot)<-c('X', 'Y', 'Peak', 'Xa', 'Ya', 'Xb', 'Yb')
```

Simple statistics with R

Use the functions names to see the column names of xds_new

```
names(xds_new)
```

```
## [1] "H" "K" "L" "Iobs" "SigmaIobs"
```

Calculate the mean value of Iobs

```
mean(xds_new$Iobs)
```

```
## [1] 92.01138
```

Calculate the mean value for each columns

```
colMeans(xds_new)
```

```
##           H           K           L           Iobs           SigmaIobs
## -0.005375955 -0.002483065 -0.007569731 92.011379778  5.759163220
```

Calculate the sum for each columns

```
colSums(xds_new)
```

```
##           H           K           L           Iobs SigmaIobs
##    -223.0    -103.0    -314.0 3816724.0  238895.8
```

Display the summary statics for xds_new

```
summary(xds_new)
```

```
##           H           K           L
##  Min.      : -14.000000  Min.      : -14.000000  Min.      : -25.00000
## 1st Qu.:  -5.000000  1st Qu.:  -5.000000  1st Qu.:  -9.00000
## Median :   0.000000  Median :   0.000000  Median :   0.00000
## Mean    :  -0.005376  Mean    :  -0.002483  Mean    :  -0.00757
## 3rd Qu.:   5.000000  3rd Qu.:   5.000000  3rd Qu.:   9.00000
## Max.    :  14.000000  Max.    :  14.000000  Max.    :  25.00000
##           Iobs           SigmaIobs
##  Min.      : -473.80  Min.      : -174.200
## 1st Qu.:   13.85  1st Qu.:    2.154
## Median :   41.66  Median :    3.803
## Mean    :   92.01  Mean    :    5.759
## 3rd Qu.:  101.10  3rd Qu.:    6.584
## Max.    : 3795.00  Max.      : 155.300
```

Create the histograms for the variable lobs using 500 breaks

```
hist(xds_new$Iobs,breaks=500)
```

