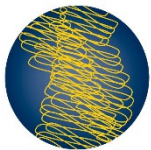


Introduction to R concepts

R in command line environments



Genomics for
Social Scientists
NATIONAL INSTITUTE ON AGING



R background



- R is a language and an environment for statistical computing
- Based on the S language and environment which was developed at Bell Laboratories by John Chambers and colleagues
- Designed around a true computer language
- Implemented on PC, Mac, or Linux/Unix
- Open source and free!



R background

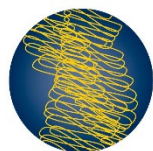
- *Package* are written by users and available to download and load into your *R distribution*
- Several basic packages load by default

```
> getOption("defaultPackages")  
[1] "datasets" "utils" "grDevices" "graphics" "stats" "methods"
```



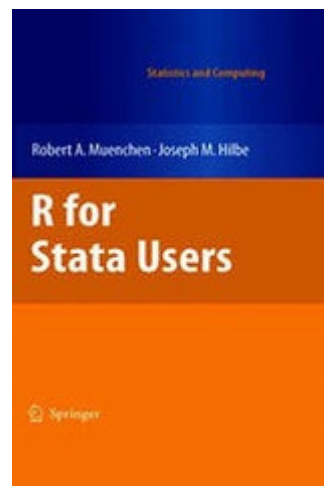
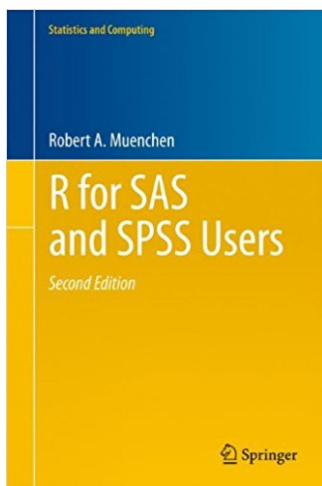
R background

- Packages are available from the Comprehensive R Archive Network (CRAN)
- Bioinformatic packages housed on Bioconductor
 - <http://www.bioconductor.org/>



Learning the R language

- Many resources



The screenshot shows the RStudio environment with the following components highlighted:

- 1- Code Editor:** The main window for writing R scripts, showing code for loading the 'ggplot2' library, viewing and summarizing the 'diamonds' dataset, and creating a plot.
- 2- R Console:** The terminal window showing the execution of the code, including the output of 'summary(diamonds\$price)' and the execution of 'p <- ggplot(carat, price, data=diamonds, color=clarity, xlab="Carat", ylab="Price", main="Diamond Pricing")'.
- 3- Workspace and History:** The top-right panel showing the current workspace with the 'diamonds' dataset (53,940 observations) and the 'aveSize' variable.
- 4- Plots and files:** The bottom-right panel displaying a scatter plot titled 'Diamond Pricing' showing the relationship between 'Carat' (x-axis) and 'Price' (y-axis), colored by 'clarity'.



General syntax of R code

- At its very core, as we have seen, R is a programming language.
- As such, formatted similarly
- For instance,

Action,
Program,
Function,
Command

How,
Argument

Options,
Modifier,
Flag,
Switch

File(s)
Input,
Output,
Destination

```
$ head -n 1 big.data.file.txt
```

```
> head(dataset, n = 1)
```



General syntax of R code

- Basic syntax - generalized

- `function(input/output, flag=argument, flag=argument,...)`

- YOU can also store things (data, results) in objects

- `object1<-
function(input/output, flag=argument, flag=argument,...)`


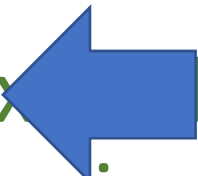
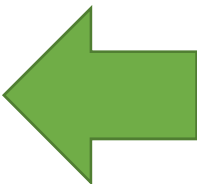





Breaking down code

```
#Plot the data in a scatter plot with titles  
plot(x=individual$F_MISS,y=individual$N_  
MISS, main = "Plot of number and percent  
missing by individual", xlab="Frequency  
missing", ylab="number missing")
```



Breaking down code


#Plot the data in a scatter plot with titles
`plot(x=individual$F_MISS,` 
`y=individual$N_MISS,` 
`main = "Plot of number of missing values and percent` 
`missing by individual",`
`xlab="Frequency of missing values",` 
`ylab="number missing by individual".` 

➤ `function(input/output, flag=argument, flag=argument,...)`



Breaking down code

```
#Read in the individual missing data  
individual<-read.table(  
"/home/username/data/EA.chr19.Quality.imis  
s",sep=" ",header=T)
```



Breaking down code

```
#Read in the individual missing data
individual <- read.table(
  "/home/username/data/EA.chr1p.quality.imiss",
  sep = "\t",
  header = TRUE)
```

The diagram illustrates the breakdown of the R code above. A large black arrow points to the first line. A purple arrow points to the variable `individual`. A blue arrow points to the function `read.table`. A green arrow points to the file path. A yellow arrow points to the separator `sep = "\t"`. An orange arrow points to the `header = TRUE` argument.

➤ `object1 <- function(input/output, flag=argument, flag=argument, ...)`



Help building commands

- Help pages exist in the form of documentation
- Accessible from online or the R window

Search: R CRAN *function*

➤ **help** (**mean**)

➤ **?mean**

➤ **? "mean"**



Arithmetic Mean

mean {base}

Description

Generic function for the (trimmed) arithmetic mean.

Usage

```
mean(x, ...)
```

```
## Default S3 method:
```

```
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

`x`

An R object. Currently there are methods for numeric/logical vectors and [date](#), [date-time](#) and [time interval](#) objects. Complex vectors are allowed for `trim = 0`, only.

`trim`

the fraction (0 to 0.5) of observations to be trimmed from each end of `x` before the mean is computed. Values of `trim` outside that range are taken as the nearest endpoint.

`na.rm`

a logical value indicating whether NA values should be stripped before the computation proceeds.

...

further arguments passed to or from other methods.



Value

If `trim` is zero (the default), the arithmetic mean of the values in `x` is computed, as a numeric or complex vector of length one. If `x` is not logical (coerced to numeric), numeric (including integer) or complex, `NA_real_` is returned, with a warning.

If `trim` is non-zero, a symmetrically trimmed mean is computed with a fraction of `trim` observations deleted from each end before the mean is computed.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

See Also

[weighted.mean](#), [mean.POSIX](#) for row and column means.

Examples

```
x <- c(0:10, 50)
xm <- mean(x)
c(xm, mean(x, trim = 0.10))
```

[Package *base* version 3.6.0 [Index](#)]



Summary

Action,
Program,
Function,
Command

How,
Argument

Options,
Modifier,
Flag,
Switch

File(s)
Input,
Output,
Destination

➤ `function(input/output, flag=argument, flag=argument, ...)`

- Case sensitivity!
- Help is available (documentation)



Funding support for this module

- This work is supported by funding from:
 - The National Institute on Aging to MPIs: Kardia, Mitchell, Faul (R25 AG053227)
 - The Michigan Center on the Demography of Aging at the University of Michigan, which is sponsored by the National Institute on Aging, Centers on the Demography and Economics of Aging (P30AG012846).
- We gratefully acknowledge the use of the services from:
 - The Population Studies Center at the University of Michigan, funded by the National Institute of Child Health and Human Development Center Grant (P2CHD041028).



Genomics for
Social Scientists
NATIONAL INSTITUTE ON AGING

