



R course for beginners

Session 4: Summary and Outlook



Session 1 – Recap commands

Working directory	I/O	Check data	Create data	Mathematical operations
getwd()	read.delim()	data[row,column]	rbind()	+
setwd()	write.table()	colnames()	<-	-
dir()		rownames()	c()	*
		length()	1:10	/
			seq()	^
			rep()	sum()
			array()	mean()
				sd()



Session 2 – Recap commands

Descriptive statistics	Statistical tests	In-build data sets	Variance	Plotting
mean()	t.test()	mtcars	cor()	plot()
median()	chisq.test()	USArrests	prcomp()	biplot()
min()	fisher.test()		hclust()	
max()			kmeans()	
quantile()				
range()				
sd()				
summary()				



Session 3 – Recap commands

Plots	Parameters	Functions	Load a library
plot() hist() abline() density() dotchart() barplot() pie() boxplot() stripchart() corrplot()	cex lwd lty pch col main xlab ylab oma par mar	title() text() legend() lines()	library(name)



Customizing R: Finding packages

R packages can be found in Bioconductor as well as CRAN

1) Bioconductor

Tools for analysis and comprehension of high-throughput genomic data

Supportive community (mailing list)

- Install bioconductor via
`source("http://bioconductor.org/biocLite.R")`
`biocLite()`
- Install packages via
`source("http://bioconductor.org/biocLite.R")`
`biocLite("edgeR")`



Customizing R: Finding packages

R packages can be found in Bioconductor as well as CRAN

2) CRAN

If you know the package name (e.g. ggplot2), type

```
install.packages("ggplot2")
```

And choose a mirror for downloading the package

You can also download the package, then direct R to the location of the package and install it



Customizing R: Update packages, remove packages

Commands: **update.packages()**

remove.packages(package)

Update.packages() can be used without arguments – all packages will be updated, or with specified library – only this library will be updated

Remove.packages(package) removes the package „package“.



R: Where to find help and ideas

Bioconductor

<http://www.bioconductor.org/>

Repository for packages and tutorials

Very active community! Sending questions via mailinglist to experts

R tutor

<http://www.r-tutor.com/>

Introduction to R programming, many examples



R: Where to find help and ideas

Quick R

<http://www.statmethods.net/>

Tutorials by Robert I. Kabacoff, Ph.D., with short examples

Mattick lab

<http://matticklab.com/>

Prof. John Mattick's homepage, providing basic protocols for e.g. analysis of microarrays



R: Coding!

Imagine...

.... you have 10 datasets...

.... pca...

.... comparison (e.g. edgeR)...

Do I have to load 10 datasets by hand??! Do those steps 10x by hand?!

----> let your computer do so!

R: Coding!

Imagine...

.... you have 10 data

.... pca...

.... comparison (e

Do I have to load

----> let your comp

```
# Read in a file containing gene names and cpm over different libraries #
# EdgeR                                                                    #
#####
library(edgeR)
root <- "my_path"
infile <- paste(paste(root, "infile.csv", sep="/"))
outdir <- paste(paste(root, "edger", sep="/"))
outdirpca <- paste(paste(root, "PCA_mean", sep="/"))
dir.create(outdir, showWarnings = TRUE, recursive = FALSE, mode = "0777")
dir.create(outdirpca, showWarnings = TRUE, recursive = FALSE, mode="0777")

edgerplot <- paste(paste(outdir, "edger.pdf", sep="/"))
mdsplot <- paste(paste(outdir, "mdsplot.pdf", sep="/"))
colors <- c("orange", "orange", "violet", "violet", "violet", "darkviolet", "darkviolet", "green", "green", "green", "red", "red", "red")
pchs <- c(3,3,15,15,15,16,16,17,17,17,18,18,18)
cols <- c("sample1", "sample2", "sample3", "sample4", "sample5")
name <- "my_results"

#####
## Main
#####

counts <- as.matrix(read.table(infile, header = TRUE, sep = "\t", row.names = 1, as.is = TRUE))
cpms <- cpm(counts)
keep <- rowSums(cpms>1)>=2
counts <- counts[keep,]

## PCA
print ("pca")
pca <- prcomp(t(counts), retx=TRUE, scale=TRUE, center=TRUE)
sd <- pca$sdev
loadings <- pca$rotation
scores <- pca$x
sum <- summary(pca)
centers <- pca$center
scales <- pca$scale

loadingsw = data.frame(rownames(loadings), loadings)
scoresw = data.frame(rownames(scores), scores)
sumw=data.frame(rownames(sum$importance), sum$importance)
centersw=data.frame(names(centers), centers)
scalesw=data.frame(names(scales), scales)

name1 <- paste(paste(name, "sd.txt", sep="_"))
out <- paste(paste(outdirpca, name1, sep="/"))
write.table(sd,out,sep="\t",row=F,quote=F)

name1 <- paste(paste(name, "rotation.txt", sep="_"))
out <- paste(paste(outdirpca, name1, sep="/"))
write.table(loadingsw,out,sep="\t",row=F,quote=F)

name1 <- paste(paste(name, "scores.txt", sep="_"))
out <- paste(paste(outdirpca, name1, sep="/"))
write.table(scoresw,out,sep="\t",row=F,quote=F)

name1 <- paste(paste(name, "summary.txt", sep="_"))
out <- paste(paste(outdirpca, name1, sep="/"))
write.table(sumw,out,sep="\t",row=F,quote=F)

name1 <- paste(paste(name, "centers.txt", sep=" "))
out <- paste(paste(outdirpca, name1, sep="/"))
write.table(centersw,out,sep="\t",row=F,quote=F)
```

Chri: