Reproducible Research practical

Contents

Choose `File` and then `R Markdown` give *Title*, *Author* and press `OK`.

A Rmarkdown template file (*Untitled1*) is generated use the `Knit` button to create your first reproducible document (you will be requested to give the file a name and to store it somewhere).

Now remove everything except the header (yaml code between the `---`) and follow the sections below:

Adding code chunks

Now add the following code chunks and type a small description of what kind of analysis is performed.

describe code …

```
library(airway)
library(DESeq2)
data("airway")
airway$dex <- relevel(airway$dex, "untrt")
dds <- DESeqDataSet(airway, design = ~ cell + dex) #add formula
dds
```

describe results…

Filtering un- or lowly expressed genes using counts per million is advocated by the developers of edgeR[@] another package for the differential expression analysis (see section 2.6 Filtering).

```
cpm <- 1e6*counts(dds)/colSums(counts(dds))
keep <- rowSums(cpm>1) >= 4
dds <- dds[keep, ]
dds
```

describe results…

describe code …

```
dds <- DESeq(dds)
res <- results(dds, alpha = 0.05)
res$Symbol <- mapIds(org.Hs.eg.db, rownames(res), "SYMBOL", "ENSEMBL")
res[order(res$padj),]
```

describe results…

Adding a figure

Adding a figure is as easy as a code chunk!

describe code …

```r
library(vsn)
library(ggplot2)
rld <- rlog(dds, blind = FALSE)
pcaData <- plotPCA(rld, intgroup = c( "dex", "cell"), returnData = TRUE)
pcaData
percentVar <- round(100 * attr(pcaData, "percentVar"))
ggplot(pcaData, aes(x = PC1, y = PC2, color = dex, shape = cell)) +
  geom_point(size =3) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  coord_fixed()
```

describe results…

R and package versions

It is advisable to let R tell exactly which R version was used and which packages and there version numbers. Every half year there is a new R version and packages can change even more frequent. By attaching this information to your analysis you know exactly to to reproduce it.

Add something like this to the end of your analysis:

```r
sessionInfo()
```

Actually, there are ways to store the R version and all the libraries that are used in a particular analysis to make your analysis even more reproducible e.g. see packrat.

Here is an example rmarkdown file and generated pdf document.

Tip of the iceberg

This is really the *tip of the iceberg* check out the rmarkdown website or the knitr website that show many applications of `rmarkdown` i.e. how to write complete books, website, scientific manuscripts and much more!