

# Sample L<sup>A</sup>T<sub>E</sub>X knitr report

Sarah Gerster (sarah.gerster@isb-sib.ch)  
Frédéric Schütz (frederic.schutz@isb-sib.ch)  
Bioinformatics Core Facility  
SIB Swiss Institute of Bioinformatics  
Lausanne, Switzerland

March 11, 2014

## 1 Welcome to the workshop

**knitr** allows you to combine your data analysis and reporting in a single document. You can choose how much R code will actually be displayed in the generated report. In any case, if needed, you will always be able to easily access all code used to generate the output presented in your report. This goes in the direction of literate programming:

I believe that the time is ripe for significantly better documentation of programs,  
and that we can best achieve this by considering programs to be works of literature.  
– Knuth (1992)

In this sample report, we just want to check that your system is set up properly to participate in the workshop. Please make sure you manage to compile this report and that the resulting PDF looks like the provided reference PDF.

## 2 Generating some data and making plots

We will simulate some meaningless data to have some values to plot and represent in tables. The code to simulate the data is shown below:

```
set.seed(42) # for reproducibility
## simulated values of two genes for samples from different batches
gene1_b1 <- sort(abs(rnorm(30, mean = 2, sd = 1)))
gene1_b1_groups <- rep(c("control", "case"), times = c(11, 19))
gene1_b2 <- sort(abs(rnorm(20, mean = 4, sd = 0.5)))
gene1_b2_groups <- rep(c("control", "case"), times = c(12, 8))

## put the simulated data in a data frame
orig_df <-
  data.frame(by = as.factor(rep(c("batch 1", "batch 2"), times = c(30, 20))),
            vals = c(gene1_b1, gene1_b2),
            cat = as.factor(c(gene1_b1_groups, gene1_b2_groups)))
```

Assume these values are measured intensities for a specific gene extracted from a larger experiment. The samples were processed in two batches. We call a customized plotting function to get a first idea of what the data look like (code not shown in the report).

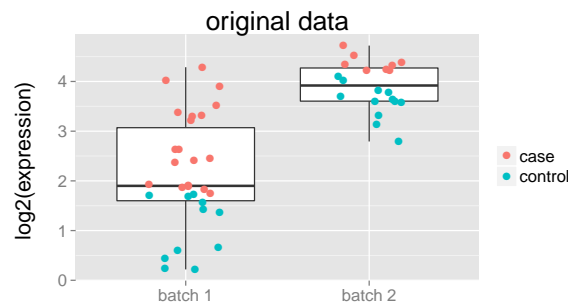


Figure 1: A first figure

There seems to be a batch effect which we would like to remove with a simple  $z$ -transform:

```
norm_df <-  
  data.frame(by = as.factor(rep(c("batch 1", "batch 2"), times = c(30, 20))),  
            vals = c(scale(gene1_b1), scale(gene1_b2)),  
            cat = as.factor(c(gene1_b1_groups, gene1_b2_groups)))
```

To see the effect of our batch effect removal step, we want to plot both, the original and the corrected data next to each other (code not shown in the report):

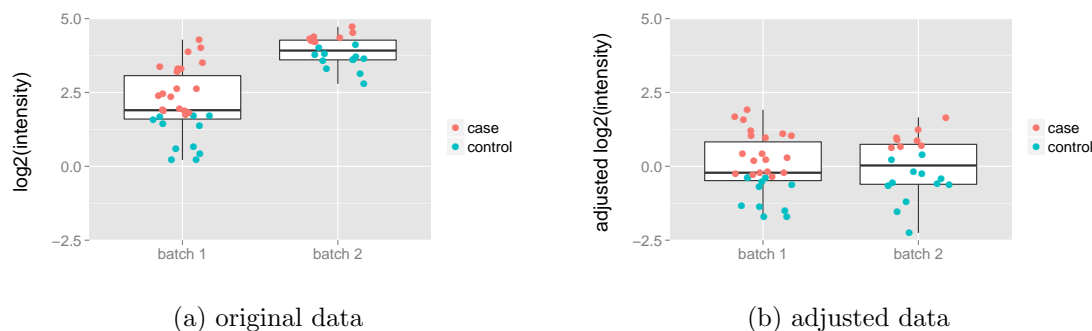


Figure 2: A second figure

We can use expressions of the form `\Sexpr{}` to use R code inline. For example, in Figure 2 we see that the  $z$ -transformation brings the 30 samples from batch 1 on the same scale as the 20 samples from batch 2. The sample counts in the previous sentence are obtained with inline R statements such as `\Sexpr{sum(orig_df[, "by"] == "batch 1")}`.

## 2.1 Let's add a table of values

The R package `xtable` is of great help to nicely display data frames as tables in a report. As an example, we present the sample counts per batch and per category in a table using the following code:

```
print(xtable(table(orig_df[, c("by", "cat")])), caption = "A table."))
```

	case	control
batch 1	19	11
batch 2	8	12

Table 1: A table.

### 3 Information about the R session

It is always a good idea to include information about the R session (version of R and used packages) in which the current report was generated:

```
print(toLatex(sessionInfo(), locale = FALSE))
```

- R version 3.0.2 (2013-09-25), x86\_64-pc-linux-gnu
- Base packages: base, datasets, grDevices, graphics, grid, methods, stats, utils
- Other packages: ggplot2 0.9.3.1, gridExtra 0.9.1, knitr 1.5, xtable 1.7-1
- Loaded via a namespace (and not attached): MASS 7.3-29, RColorBrewer 1.0-5, colorspace 1.2-4, compiler 3.0.2, dichromat 2.0-0, digest 0.6.3, evaluate 0.5.1, formatR 0.10, gtable 0.1.2, highr 0.3, labeling 0.2, markdown 0.6.3, munsell 0.4.2, plyr 1.8, proto 0.3-10, reshape2 1.2.2, scales 0.2.3, stringr 0.6.2, tools 3.0.2

### 4 End of sample file

That was it. If you produce the PDF of this report on your laptop, you should get a document that looks very similar to the file `final_report_latex.pdf` in the GitHub repository. If this is not the case, you have to fix the problems (e.g. install missing software) before the workshop in order to be able to fully participate.