Exercise 2

Riccardo De Bin and Vindi Jurinovic

October 26, 2015

Writing your R-code, you can use the character # to insert a comment, so you can describe your commands (and remind yourself what you did weeks or years ago). Example:

```
> x <- 2 + 4 + 1 # + 4 + 2
```

Here, only the expression `x <- 2 + 4 + 1` will be executed and `+ 4 + 2` will be ignored.

If you want to save a plot into a file named, say, `'myPlot.pdf'`, use the following code:

```
> pdf('myPlot.pdf')
```

This command specifies the name and the file format. You can also choose `jpeg(), png()`,...

```
> plot(variable1, variable2, col=..., pch=..., ...)
```

Here you are creating your plot with the chosen variables and arguments.

```
> dev.off()
```

This command closes the graphics device. Without it, the file is not created properly.

1 R Project Part I

Here, we will answer some questions for your data project. The question numbers are the same as on the question sheet.

2 Describe the US population with regards to:

a) demographic characteristics (age, gender, ethnicity...). Recode the age variable into following categories: 20-34, 35-49, 50-64, 65-79, 80 or higher. Add this new variable (a factor!) to your data set.
b) self-rated health.

3 Lifetime prevalence of cancer in the population

a) Estimate the lifetime prevalence of cancer. Can you also give an interval estimate?
b) What are the prevalences estimates in those who were exposed to pollutants at work for a longer time period, and in those who weren't? Is there a significant difference in prevalence between these two subgroups?

4 HDL cholesterol and gender

a) Look at the distribution of high-density lipoprotein (HDL) cholesterol levels. What shape does it have? Apply an appropriate transformation to normalize HDL and save it as a new variable. (We already did this last week.)

b) Is there a significant difference between men and women in HDL cholesterol levels (using normalized variable)?