

W06-1: Visualizing data

This worksheet will test your visualization competences. After completing this worksheet you should be able to plot simple visualizations and decide if a simple data transformation might enhance the visualization.

Things you need for this worksheet

- R — the interpreter can be installed on any operation system. For Linux, you should use the r-cran packages supplied for your Linux distribution. If you use Ubuntu, [this](#) is one of many starting points. If you use windows, you could install R from the official [CRAN](#) web page.
- R Studio — we recommend to use R Studio for (interactive) programming with R. You can download R Studio from the official [web page](#).
- your combined world bank data sets computed within [W05-1: Melting a data set](#)
- [this](#) csv data set which matches world bank country acronyms to regions according to this [overview](#).

Learning log assignments

😄 As always, please add these entries to your today's learning log at [teachwiki](#):

- Favorite aspect of the session (if any)
- Superfluous aspect of the session (if any)
- Eureka effect (if any)
- Links to what I've learned so far (if any)
- Questions (if any)

For more information see this short [howto](#).

As today's special, please complete the following assignment:

As part of [W05-1: Melting a data set](#) you have transformed and combined the world bank information on CO₂ emissions and GNI on a country level since 1960.

Today we want to visualize parts of this data but partly on a regional basis. Therefore, you have to merge your data set from [W05-1](#) with the country-to-region mapping data set provided above. Since we want to focus on visualization in today's worksheet, the necessary code to do this is provided at the end of the worksheet.

Let's do some visualization:

😄 Please extend the R script given below to create one visualization for each of CO₂ and GNI which allows us to get an estimate over the distribution function of the parameters.

😄 Based on the visualization results above, please extend your script to include an appropriate transformation of the CO₂ and GNI values to get their distribution a little bit closer to a normal distribution. Save the result as part of your data frame (i.e. new columns) and visualize the results

again (using the same visualization type as above).

After the data has been transformed, please continue with some more visualizations:

☹ Please extend the script to include two box whisker plots which respectively summarize the CO₂ and GNI value distribution for each of the world bank regions.

☹ Finally, please visualize the correlation between all CO₂ and GNI values in one plot and upload your final script to your learning log.

"Snippet 01"

```
rm(list = ls(all = T))

#### General settings
#####
inpath <- "<your input path>"
setwd(inpath)

library(lattice)

#### Load the merged world bank data set and merge it with the region
list #####
wb <- read.table("<your world bank data set>", header = TRUE,
                sep = ",")
cntry <- read.table("wb-db_country_classes.csv", header = TRUE,
                   sep = ",")
wbc <- merge(wb, cntry[,!(grepl("Economy|X", colnames(cntry)))],
             by.x = "Country.Code", by.y = "Code")
wbc$Region[wbc$Region == ".."] <- NA
wbc$Region <- droplevels(wbc$Region)

#### Make some plots
#####
```

From:
<http://moc.environmentalinformatics-marburg.de/> - Marburg Open Courseware

Permanent link:
<http://moc.environmentalinformatics-marburg.de/doku.php?id=courses:msc:data-management:worksheets:dm-ws-06-1>

Last update: 2014/11/25 11:37

