Getting started with graphics: Exercises

Andreas Alfons and Pieter Schoonees
Erasmus University Rotterdam

1 Getting started

Exercise 1.1. To familiarize yourself with the RStudio interface, perform the following tasks.

(a) Obtain a list of all data sets available in package `euR`. Choose three data sets and view their help files to get information on the variables.

(b) Create a new R script. Consider again the three data sets from before, and write the commands for loading them and producing summaries into the R script. Save the R script in the working directory.

(c) Close the R script and reopen it. Execute the commands on the R console using the appropriate keyboard shortcut (do not copy & paste) and examine the output.

(d) In the top right panel, switch to the *History* tab and browse through your history of commands. Double click on one of the commands. What happens?

Exercise 1.2. To familiarize yourself with the R help facilities, perform the following tasks.

(a) In the bottom right panel, switch to the *Packages* tab. Choose a package and load it. Click on the name of the package. What happens?

(b) Browse the list of available help topics in the package from before. Use links in a help file to navigate from one topic to another. Use the *back* and *forward* arrows in the menu bar of the help tab to browse through the last opened help topics. Use the text field that displays *"Find in Topic"* to search for text in a help file.

2 Graphics

Exercise 2.1. Load the `patents` data from the provided `patents.RData` file. The data set contains information on patents granted in 2012 in each of the 50 US federal states and the District of Columbia. It consists of the following variables:

total The total number of granted patents.

utility The number of granted utility patents.

design The number of granted design patents.

plant The number of granted plant patents.

population The number of inhabitants.

area Land area in km^2.

governor Party affiliation of the state governor.

area Land area divided into three categories: `"small"`, `"medium"` and `"large"`.

density The population density.

densitycat Population density divided into two categories: `"low"` and `"high"`.

logdensity Logarithm of population density.

logtotal Transformed total number of granted patents, i.e., `log(total+1)`.

logutility Transformed number of granted utility patents, i.e., `log(utility+1)`.

logdesign Transformed number of granted design patents, i.e., `log(design+1)`.

logplant Transformed number of granted plant patent, i.e., `log(plant+1)`.

The data were scraped from http://www.statsamerica.org/profiles/sip_index.html and http://en.wikipedia.org/wiki/List_of_current_United_States_governors.

(a) First produce a scatterplot of `total` vs `density`, then another scatterplot of `logtotal` vs `logdensity`. What do you observe about the distribution of those variables and the effect of the log-transformation?

(b) Produce histograms of `total` and `logtotal` to compare the distribution of the number of granted patents before and after the log-transformation. Play with the number of bins to get a more complete picture of the distributions.

(c) Produce density plots of `total` and `logtotal` and compare the estimated densities.

(d) Produce a boxplot of `logdensity`. Do you find any outliers?

(e) Produce conditional boxplots of `total` and `logtotal` with observations grouped by population density category (`densitycat`).

(f) Produce barplots of the factors `governor` and `areacat`.