Getting started with graphics: Exercises

Erasmus Q-Intelligence B.V.

1 Graphics

Exercise 1.1. Load the `patents` data from the provided `patents.Rds` 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`.