Statistics and Exploratory Data Analysis

Marcin Chlebus, Ewa Cukrowska-Torzewska

Laboratory 4: Graphical analysis of data (I)

Exercises for students

Exercise 1:

Use 'ess_data.csv'. This dataset contain selected variables from European Social Survey for 2014.

a) The variable 'cgtsmke' contains data on smoking behavior of individuals. Assign labels to data on smoking behavior as follows: 1 "I smoke daily", 2 "I smoke but not every day", 3 "I don't smoke now but I used to", 4 "I have only smoked a few times, 5 " I have never smoked ".
b) Summarize the distribution of the smoking behavior of Europeans (for all countries) using relevant charts.
c) Compare the distribution of smoking behavior of Polish people and other European nation (whichever you want) on one graph.
d) The variable 'fclcntr' contains answers to the question "How close do you feel to your country?". Assign labels to this variable as follows: 1 "Very close", 2 "Close", 3"Not very close", 4 "Not close at all".
e) Prepare a graph describing the variation across European countries in their attachment to the home country.
f) Plot the distribution of age (agea) for Poland and Germany. Use both histogram and density plots. Compare the distributions of the Polish and German populations by age – is there any noticeable pattern?
g) Summarize the distribution of the education level measured by the years of completed education (edulvlb) by country. In which European country study the longest?

Exercise 2:

Use 'ess_data.csv'. Using graphical visualization of data answer the following questions:

- What is the proportion of Europeans that is at most 160 cm high (height variable)
- What is the proportion of Poles that feels "very close" to their country?
- What is the proportion of Portuguese that declares that they "have never smoked"?
- What is the proportion of men in the Czech Republic that weights more than 100kg? What about women?

Exercise 3

Use 'ess_data.csv'. Create the boxplot for weight (for all countries). Based on your graph answer to the following questions:

- What are the range, the three quartiles and the interquartile range? Check your answers with calculating the relevant quartiles
- About how many outliers can you identify in the data?