Statistics and Exploratory Data Analysis

Marcin Chlebus, Ewa Cukrowska-Torzewska

Laboratory 5: Graphical analysis of data (II)

Exercises for students

Exercise 1

Use data on German soccer league that is available in R (to load the data type: data("Bundesliga")).

1) Plot the goals per game over time
2) Plot the goals per game for each year

Exercise 2

Use data on Titanic surviving that are available in R ("Titanic" dataset from the package "datasets") create mosaic plots showing:

3) The relation between surviving the Titanic sinking and sex
4) The relation between surviving the Titanic sinking and class of the ticket (1^{st} class, 2^{nd} class, 3^{rd} class, and crew)
5) The relation between surviving the Titanic sinking, sex and class of the ticket (1^{st} class, 2^{nd} class, 3^{rd} class, and crew)

Exercise 3

Use the data on movies that are available in R ("movies" dataset from the package "ggplot2movies"). Using the data and relevant plots answer the following questions:

1) Did the length of the movies increased over time?
2) Did the budget of the movies increased over time? (hint: there are many missing observations for budget data, so you need to ignore NA)
3) Is there any relation between the budget and the rating?
4) Is there any relation between the rating and the number of votes?
5) Is the variable "votes" normally distributed? Is the variable "rating" normally distributed?
6) Does the distribution and density of the rating differ by the type of the movie (action, comedy, drama, etc.)?

Hint:

To load the data that are available in R packages you need to install the package, load it and load the data. For example for exercise 3:

```
install.packages("ggplot2movies")
library(ggplot2movies)
data("movies")
movies
```