Ex.2.1. The level of water in Wisła river was measured in March of 10 randomly chosen years. The following measurements were obtained:

$$m,\ l,\ l,\ h,\ h,\ l,\ h,\ m,\ h,\ l,$$

where m stands for *middle*, l for *low* and h for *high*.

(a) What kind of data does this data set represent?

(b) Save this data in R as (ordered!) `factor`. Set labels *low, middle* and *high* for the levels *l, m, h*, respectively.

(c) Create a table of the counts at each factor's level.

(d) Create a full frequency table for this data.

(e) Create a bar plot and pie plot for this data.

Ex.2.2. In `mtcars` built-in data set there is a variable *cyl* being the number of cylinders in the sample of 32 randomly chosen cars. Create a full frequency table and store it in a `data.frame` object. Name rows and columns accordingly.

Ex.2.3. Install and load **coin** package. Familiarize with `malformations` data set from a study on the relationship between maternal alcohol consumption and congenital malformations.

(a) Create contingency table for variables in this data set.

(b) Create frequency table for variables in this data set. Experiment with different marginal sums.

(c) Which marginal sum would you choose if you would like to assess whether alcohol consumption (at a given level) increases the risk of congenital malformations? What is the risk of getting congenital malformations if one consumes the alcohol at the level between 1-2?

(d) Create a new dichotomous (factor) variable *consumption_high*, that is equal to *No*, in case the level of alcohol consumption is 0 or 1-2, and is equal to *Yes*, otherwise.

(e) Repeat points (a) and (b) for variables *consumption_high* (created at point (d)) and *malformation*.

(f) Compute *risk ratio* using *consumption_high* variable (created at point (d)) and *malformation*. Is the risk of getting congenital malformations increased when alcohol consumption is high? If it so, how many times?

Ex.2.4. In the built-in package **datasets** there is a data set called `HairEyeColor`.

(a) Create a frequency table for all three variables using different values of marginal sums. Note the differences.

(b) Create a frequency table for *Eye*, subject to men with brown hair and represent the data with an adequate graphic.

(c) Repeat the above point for women with blue eyes.

Ex.2.5. File *data_tumor.csv* contains information on patients with tumor.

(a) Read this data file into `data.frame` object. Call the `summary()` function for this object.

(b) Perform a preliminary analysis on the variable *receptors.estrogen*:

i. Remove the `NA`s.
ii. Find the frequency table.
iii. Represent the data with an adequate plot.

(c) Perform a preliminary analysis on the variable *age*:
i. Compute mean, median, etc.
ii. Check on outliers and plot a boxplot.
iii. Plot a histogram with `hist()` function.
iv. Estimate density using the `density()` function.

(d) Perform a preliminary analysis on the variable *VGEF*.

Ex.2.6. In the package **e1071** there are build-in functions to compute sample skewness and kurtosis. Apply them to *Wind* and *Temp* variables in the `airquality` data set.