Summaries and visualization of distributions

Reflection on the last week

Objectives

Organizing your work

Descriptive Statistics

Characterizing centrality

Mean (*průměr*)

`mean(x)`

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}(\sum_{i=1}^{n} x_i)$$

Median (*medián*)

`median(x)`
- **Robust**, minimizes influence of outliers.

What are outliers? (*odlehlé hodnoty*)

- **Outliers** are data points that significantly differ from other observations.
- May indicate a measurement error, an exceptional observation, etc.

Characterizing centrality

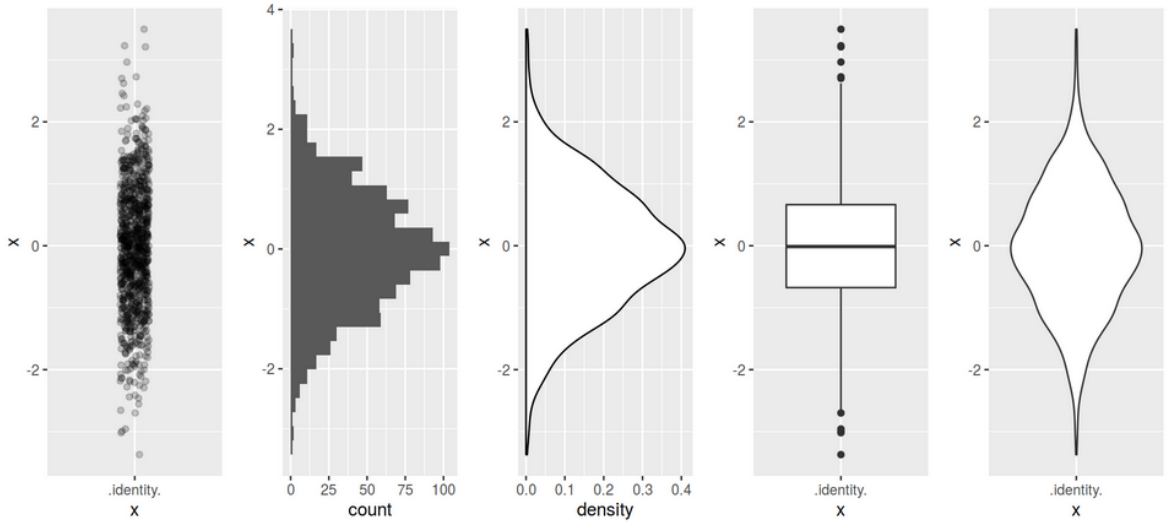

Figure 1: Various plots of a normal distribution

Characterizing dispersion and/or spread

Range (*rozpětí*)

`max(x) - min(x)` or `range(x)`

Variance and Standard deviation (*rozptyl* a *směrodatná odchylka*)

`sd(x)`

$$\sigma = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$

Interquartile range (midspread, IQR, *kvantil*, *mezikvartilové rozpětí*)

`IQR(x)`
- **Robust**, minimizes influence of outliers.

Characterizing dispersion and/or spread

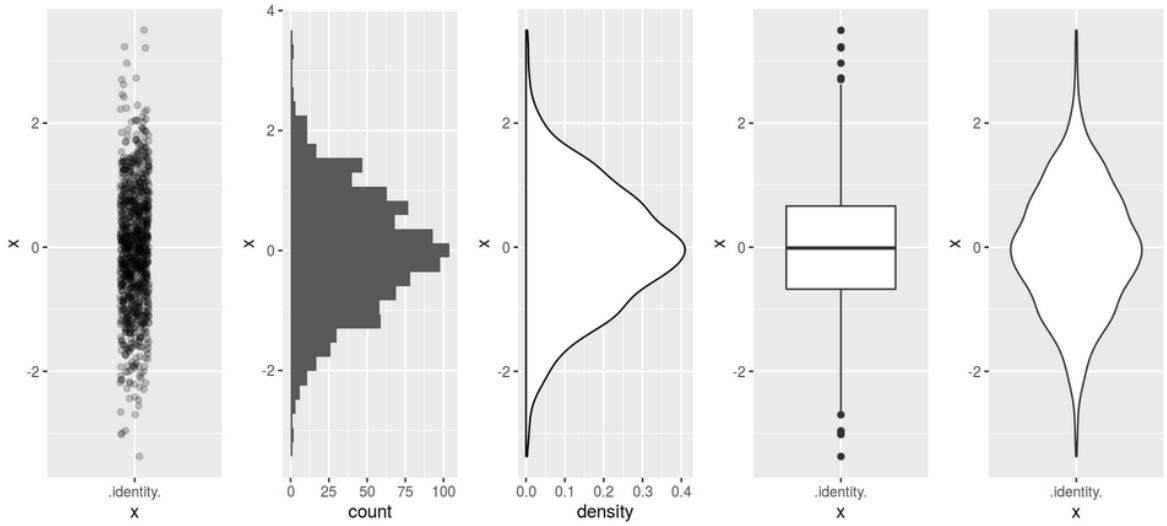

Figure 2: Various plots of a normal distribution

Brainstorming

- Why do we visualize data?
- What elements does a *good* graph contain?
- How are these elements called?

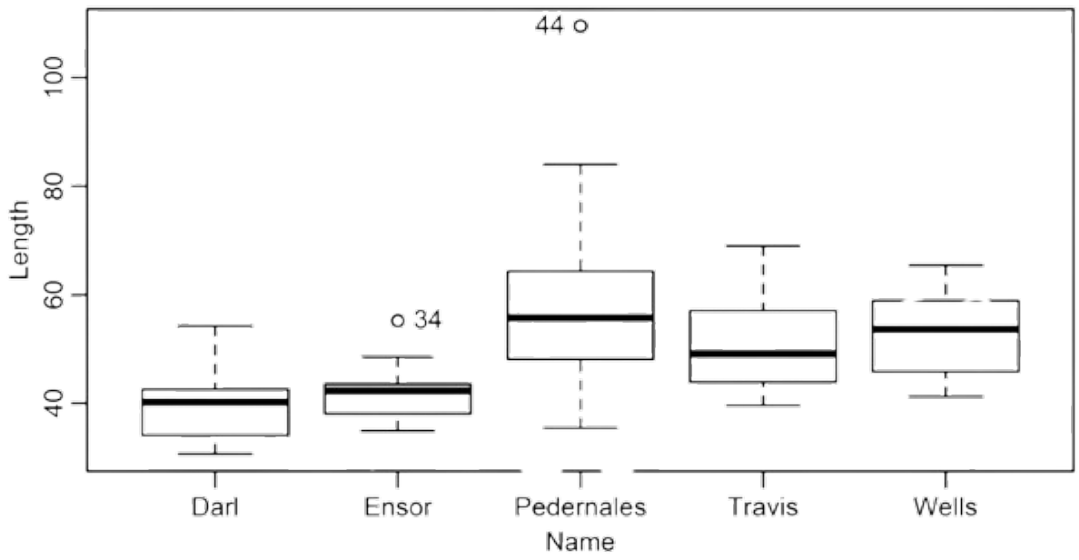

FIGURE 15 Box-and-whiskers plots for dart point lengths.

Figure 3: Boxplots from Carlson 2017

Plots for one variable

Histogram

Density plot

Exercises

Assignments

- Read Make a plot chapter in *Data Visualization* book by K. J. Healy.

Optional

- Go through *Visualize data* tutorials here.