Exercise 1

Data Analysis

March 13, 2024

Install the R package `nycflights13` with
`library("nycflights13")`
and then load the dataset `flights` with
`data(flights)`
and use `attach(flights)` to get direct access to the variables. See also help page `?flights` and `str(flights)` for more detailed information.

In this exercise we will use simple plots to learn something about the data structure.

1. Show the variable `air_time` in a univariate scatterplot. What do you see?

2. Show the variable `distance` in a univariate scatterplot. What do you see?

3. Show `air_time` as well as `distance` in histograms, and adjust the histogram presentatation to get a good idea about the data structure. What do you see?

4. Does the gap in the distribution of `distance` (shorter versus long distances) correspond to the gap in the distribution of `air_time` (shorter versus long)? Find an answer by using different color information in the plot(s).

5. Investigate graphically if the distribution of arrival delay differs for short versus long flight distance. You could try to show two histograms in one plot using the same scale. A second histogram can be added to a plot by using the argument `add=TRUE`. What do you conclude?

6. Do the same as in 5. based on the variable `dep_delay`.

7. At which hour of the day can we expect the biggest aarival delays? When do we have the smallest arrival delays? How could you answer this question using simple plots?

8. Is there is a difference of the distribution of arrival delays in the summer and winter months?

9. Install the R package `epade` and load the package. Use for the following task the function `histogram.ade()`, see also help file, to create superimposed histograms. Compare the airlines (`carrier`) AA (American Airlines) and WN (Southwest Airlines) concerning their distribution of arrival delays. You first need to create a data frame including the arrival delays of these carriers and a factor variable with the carriers. Use appropriate histogram parameters to make possible differences of the carriers clearly visible.

10. Compare visually the average arrival delays of the carriers AA and WN across the months. What do you conclude?