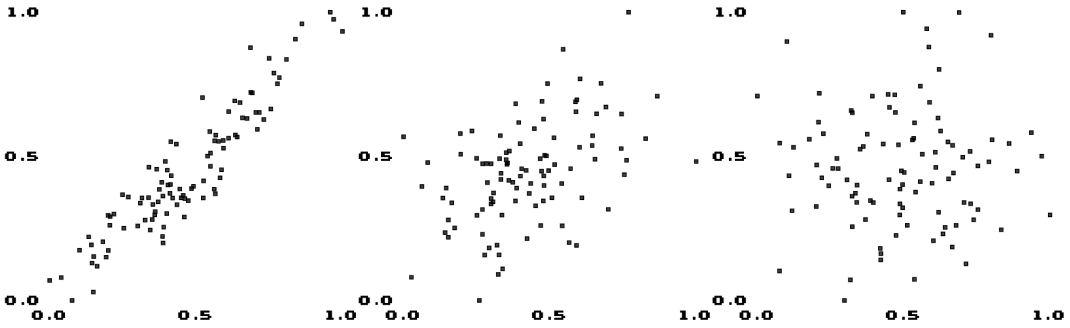
Statistics for Data Science Exercise 10
Wintersemester 2023/24
Please complete these problems before the exercise session on Tuesday 9 January, 2024, 8:30. Please be prepared to present your solutions to any problems that you completed successfully.

1. (a) Play the Guess the Correlation game:[†]
   `https://www.guessthecorrelation.com`
   (b) Now that you have developed a "feel" for estimating Pearson correlation coefficients, what do you estimate the correlation coefficients of the following scatter plots to be (based only on the visualization)?



2. Compute a p-value for the example on pg. 345 of the lecture notes using a Monte Carlo permutation test.
3. Download the file `shoeheight.txt` from the course webpage. The data contains the heights and shoe sizes of 50 students.
   (a) Plot a scatter plot of the data and based on the plot, estimate the sign and the magnitude of the correlation between the variables.
   (b) Calculate the Pearson correlation coefficient and the Spearman rank correlation coefficient of the data.
   (c) Use a Monte Carlo permutation test to test the statistical significance of the correlation coefficients.
4. A factory produces warm, woollen slippers. The number of weekly slipper orders y is believed to be linearly dependent on last week's median temperature x (in Celsius degrees). Consider the following sample of the variables:

x	-10	-17	-4	-7	-5	-6	-11
y	105	163	43	69	48	56	115

   (a) Fit a linear model into the sample.
   (b) Calculate the least squares estimates of the model parameters.
   (c) Plot a scatter plot of the sample and add an estimated regression line to the figure.
   (d) What is the coefficient of determination of the model?

[†]A version of this game which does not require logging in using a Google account is available at `https://www.geogebra.org/m/KE6JfuF9`