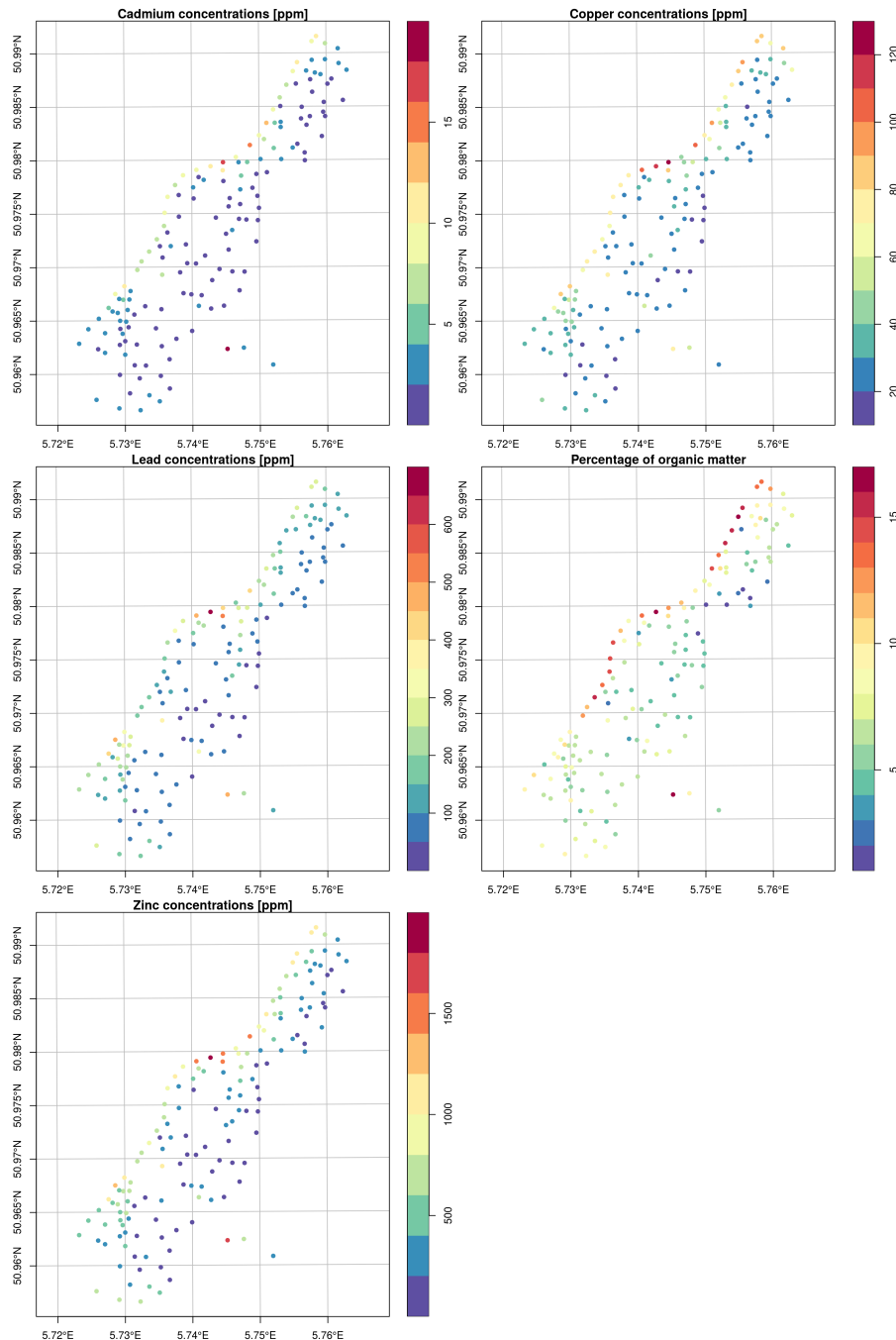## Exercise 1 - Basic Statistics - Solutions

Office hours: Friday 09:00-12:00

A. Berne   alexis.berne@epfl.ch
G. Ghiggi   gionata.ghiggi@epfl.ch
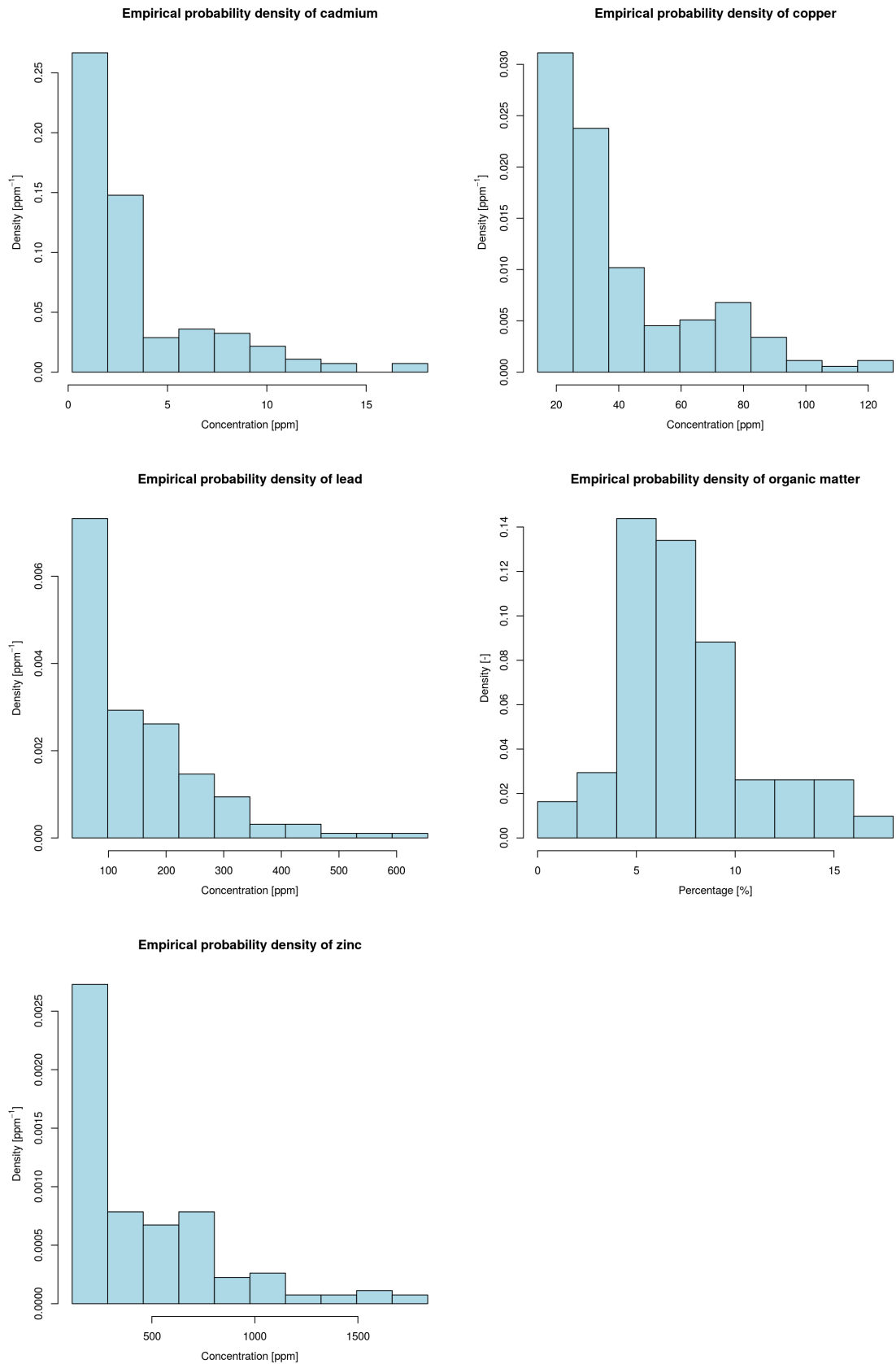M. Guidicelli   matteo.guidicelli@epfl.ch

1. *Plot and show the map of cadmium, copper, lead, zinc and organic matter concentrations for the studied domain. What can you say about the spatial distribution of these variables?*

The concentrations are larger along the Northwest border or the domain (i.e., along the river Meuse). The samples are distributed homogeneously over the domain, except for 3

locations in the Southeast.

2. *Plot and show the empirical probability density function (pdf) estimated from the values of cadmium, copper, lead, zinc and organic matter. Is normality a reasonable hypothesis?*

Empirical probability density of cadmium

Empirical probability density of copper

Empirical probability density of lead

Empirical probability density of organic matter

Empirical probability density of zinc

Normality is not a reasonable hypothesis. The distributions are all positively skewed (i.e., they have a tail on the right). The mean is always larger than the median (see the table

below). The distribution closest to normality is the one of the organic matter.

3. *Compute and give the values of the mean, the standard deviation, the median and the 90% percentile of the cadmium, copper, lead, zinc and organic matter concentrations. Which variable exhibits the minimum (relative) variability? And the maximum?*

	cadmium [ppm]	copper [ppm]	lead [ppm]	zinc [ppm]	organic matter [%]
mean	3.246	40.316	153.361	469.716	7.478
sd	3.524	23.680	111.320	367.074	3.433
median	2.1	31	123	326	6.9
90%	8.26	77.0	290.4	986.4	13.24
cv	1.086	0.587	0.726	0.781	0.459

The relative variability is determined by the coefficient of variation (i.e., the standard deviation divided by the mean). It is minimum for the organic matter (0.459) and maximum for the cadmium (1.086).

4. *Compute and give the minimum/maximum euclidean distance between each pair of data in the studied domain. How many pairs are separated by 200 m or less?*

The minimum distance is 43.98 m. The maximum is 4440.76 m. There are 314 pairs separated by 200 m or less.

5. *Demonstrate the following equality:* $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y].$

$$
\begin{aligned}
\mathrm{Var}[X + Y] &= \mathrm{E}[((X + Y) - \mathrm{E}[(X + Y)])^2] \\
&= \mathrm{E}[((X - \mathrm{E}[X]) + (Y - \mathrm{E}[Y]))^2] \\
&= \mathrm{E}[(X - \mathrm{E}[X])^2 + (Y - \mathrm{E}[Y])^2 + 2(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]
\end{aligned}
$$