

UNIVERSITY OF KOBLENZ - AG SOFTLANG

DATA SCIENCE

Assignment 3: Visualization

TO BE SUBMITTED UNTIL THE 1.12.2021 (2 P.M)

November 17, 2021

Question 3.1: Find a Visualization

The supplementary material of this assignment contains three CSV files¹. We will describe the data sets in a second. You are supposed to find the best visualization for each data set just using a single plot (for each data set). If you see alternative options for a data set, you may also provide up to two of them, with a short clarification of the central difference between both options.

Please provide us with the plots as PDF (one PDF for each data set eventually including the alternatives) and the code for producing the plots from the CSV files. You may freely choose the library or technique you want to use for plotting (as long as you are not doing it by hand).

The following three data sets need to be visualized.

- The first data set contains a single categorical variable describing the religion of the people in Germany. There are in total 18084 persons contained in the data set and twelve levels, e.g.: *Unaffiliated*, *Catholic Church*, *Evangelical Church* or *Islam*.
- The second data set includes data on the salary of people in Germany. The data is accomplished by the age of each person and whether he or she works in the East or West of Germany.
- The third data set contains a single numeric variable measuring the height of *Kung San*² people.

¹Two of the data sets are made up (simulated using realistic parameters).

²https://en.wikipedia.org/wiki/%C7%83Kung_people