**DataCraft\_Kurs08/24**

**< Modul: Datenvisualisierung in Python >**

**Tagesaufgaben 24.02.2025**

**Michael Matthiesen**

Aufgabe 1: Besuche unter folgendem Link das Unterforum „dataisbeautiful“ von Reddit und erkläre die folgende Grafik: <https://www.reddit.com/r/dataisbeautiful/comments/10uc6wo/the_most_common_place_name_in_each_european/>  
Schreibe für diese Grafik eine passende Zusammenfassung und beantworte dabei die Fragen:

* Was ist dargestellt?
  + - Die Grafik zeigt eine Übersicht der am häufigsten verwendeten Ortsnamen in Europa und zeigt, wie oft bestimmte Namen in verschiedenen Ländern vorkommen.
* Woher kommt die Grafik?
  + - Die Grafik stammt vom Online-Datenprojekt Landgeist, das sich auf Auswertungen von OpenStreetMap-Daten (Stand 2023) stützt.
* Welche Kritik gibt es daran?
  + - Kritisch wird vor allem angemerkt, dass die Datengrundlage je nach Region unterschiedlich präzise sein kann. Zudem fließen nur Städte, Gemeinden und Dörfer ein, während kleinere Siedlungen (z. B. Weiler) nicht berücksichtigt werden. Auch mögliche Übersetzungs- und Transkriptionsunterschiede sowie abweichende offizielle Namensschreibweisen können die Genauigkeit der Auswertung beeinträchtigen.

Aufgabe 2: „Die Wissenschaft macht keine Fehler!“ ist ein ausgedachter Satz, den Valentin hier auf diesen Aufgabenzettel geschrieben hat. Halte diesem Statement zwei Grafiken entgegen, die du der Sammlung „The top ten worst graphs“ entnommen hast: <https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/>  
Fasse dabei kurz zusammen, was dargestellt werden sollte (nur grob, bitte nicht inhaltlich) und was dabei womöglich schiefgelaufen ist. Unter „Discussion“ findest du hier mögliche Pro- und Kontraargumente.

* 1) Roeder K (1994) DNA fingerprinting: A review of the controversy (with discussion). *Statistical Science* 9:222-278, Figure 4
* Grafik 1 (Binned Frequency Data – D4S139 & D10S28)
* Was sollte gezeigt werden?  
  Die Verteilung bestimmter genetischer Marker in vier ostasiatischen Populationen (chinesisch, japanisch, koreanisch, vietnamesisch).
* Was lief schief?  
  Die überlagerte 3D-Darstellung mit mehreren Kurven erschwert den direkten Vergleich. Verschiedene Achsen- und Bin-Einstellungen (teils ungleiche Abstände) führen zu einer unübersichtlichen, schwer lesbaren Darstellung.
* 9) Cotter DJ, et al. (2004) Hematocrit was not validated as a surrogate endpoint for survival amoung epoetin-treated hemodialysis patients. *Journal of Clinical Epidemiology* 57:1086-1095, Figure 2
* Grafik 2 (3D-Balkendiagramm – Mortality Rate vs. Hematocrit Group)
* Was sollte gezeigt werden?  
  Der Zusammenhang zwischen Hämatokritwerten und Sterblichkeitsraten in verschiedenen Patientengruppen.
* Was lief schief?  
  Durch den 3D-Effekt und die eng beieinanderstehenden Balken sind die Werte nur schwer zu erkennen und zu vergleichen. Die Perspektive verdeckt Teile der Daten, was zu einer irreführenden Wahrnehmung führt.