



Einleitung

Ziel dieser Übung ist es, die im Modul „Graphische Datenexploration und Datenvisualisierung“ vermittelten Techniken anzuwenden. Grundlage dafür ist ein Datensatz mit dem regionalen Portrait der Schweizer Gemeinden des Jahres 2014, der über das Open-Data Portal des Bundes frei zugänglich ist (<https://opendata.swiss/de/dataset/regionalportraits-kennzahlen-aller-gemeinden>). Der auf Moodle abgelegte Datensatz ist bereits mit einzelnen Merkmalen zur räumlichen Gliederung der Schweiz ergänzt:

- Kanton
- Grossregionen der Schweiz
- Sprachgebiete
- Städtische/Ländliche Gebiete

Zusatzerläuterungen zu den Daten finden Sie in den Dokumenten:

- be-b-00.04-rgs-15.xls (Raumgliederung Schweiz)
- je-d-21.03.01.xls (Kennzahlen Regionalportrait)

Es handelt sich um einen Datensatz, der nicht explizit für Analysezwecke aufbereitet ist, vielmehr ist das Regionalportrait eine Sammlung unterschiedlichster Kennzahlen. Eine erste Übertragung in ein einheitliches Datenformat ist im File `gemeindedaten.csv` vorgenommen. Ansonsten sind die Daten aber frisch aus dem Netz. Was können wir daraus über die Gemeinden in der Schweiz in Erfahrung bringen?

1. Laden des Datensatzes und erste Begutachtung

1. Laden Sie den Datensatz `gemeindedaten.csv`
2. Verschaffen Sie sich einen ersten Überblick zu den Daten? Hat mit dem Einlesen alles wie erwartet geklappt? Sind die Daten so kodiert, wie Sie es erwarten? Stimmen die Datenformate der Variablen? Gibt es fehlende Werte und wie sind diese kodiert?
3. Falls nötig, definieren Sie im `dataframe` fehlende Werte so, dass R diese tatsächlich als fehlende Werte auffasst.
4. Beantworten Sie folgende Fragen: Wie viele Gemeinden gab es in der Schweiz im Jahr 2014? Was ist die mittlere Einwohnerzahl einer Schweizer Gemeinde? Wie viele Einwohner leben in der grössten Gemeinde? Wie viele in der kleinsten?

2. Graphische Datenexploration

Führen Sie visuelle Auswertungen durch, mit welchen Sie folgende Fragen beantworten lassen:

5. In welchem Kanton gibt es am meisten Gemeinden? In welchem am wenigsten?
6. Betrachten Sie die Einwohnerzahlen der Gemeinden gruppiert nach Sprachregionen. Wie heissen die jeweils grössten Gemeinden?
7. Betrachten Sie die Veränderung der Einwohnerzahl von 2010 bis 2014 nach Sprachregionen. In welcher Sprachregionen sind die Gemeinden am stärksten gewachsen? In welcher am wenigsten oder gibt es Sprachregionen, in welcher die Einwohnerentwicklung in der Tendenz sogar eher rückläufig ist? Analysieren sie zusätzlich graphisch, ob die Unterscheidung von städtischen und ländlichen Gemeinden dabei eine Rolle spielt?
8. Untersuchen Sie die Zusammenhangsstruktur folgender Variablen:
bev_dichte, bev_ausl, alter_0_19, alter_20_64, alter_65, bevbew_geburt, sozsich_sh, strafen_stgb
Gibt es Korrelationen? Falls ja, lassen Sie sich erklären oder sind sie eher unerwartet? Suchen Sie sich einen Ihnen interessant erscheinenden Zusammenhang und schauen Sie sich diesen in einem eigenen Scatterplot an
9. Visualisieren Sie eine Kontingenztafel mit den Variablen *Stadt_Land* und *Sprachregionen*. Welcher Gemeindetyp überwiegt bei deutschsprachigen Gemeinden, welcher bei italienischsprachigen Gemeinden. Gibt es in jeder Sprachregion isolierte Städte?
10. Erstellen Sie ein politisches Profil nach Sprachregionen mit der Hilfe der Variablen zu den Wähleranteilen.

3. Eine „hübsche“ Datenvisualisierung

11. Wählen Sie eine der bereits erstellten Graphiken aus und verfeinern Sie diese so, dass Sie von anderen gut verstanden wird (Titel, Fussnote etc.) und ästhetisch möglichst überzeugt.