**Zusammenfassende Übungen**

**Bitte für jeden Datensatz in jedem Fall diese Schritte ausführen:**

Daten dokumentieren

* **Datei sinnvoll benennen**
* **Roh-Daten duplizieren und nie wieder anfassen**
* **Datenquelle in eigenem Sheet notieren: Wo? Was? Wann? Ggf. Hinweise notieren**
  + **Fragen an die Daten in eigenem Sheet aufschreiben**

**Erinnerung:** Deutsche Besonderheiten:

* In deutschen Tabellen wird ';' statt ',' als Trenner für .csv verwendet
* Dafür schreiben wir in Deutschland '42,3' statt '42.3'
* In Tabelle gibt es oft fehlende Werte, die zum Beispiel mit '-' angegeben werden
* Oft wird auch nicht das weltweit gebräuchliche 'uft-8' als Zeichenformat verwendet, sondern 'latin-1' oder 'iso-8859-1'

**Übung 1**

Aus welchem Gebiet stammen im Schnitt die jüngsten männlichen Eingebürgerten in NRW 2020?

* Suche auf [<https://www-genesis.destatis.de/genesis/online>] nach 'Durchschnittsalter der Eingebürgerten: Bundesländer und

##### Ausland, Jahre, Ländergruppierungen/Staatsangehörigkeit, Geschlecht, Familienstand’ - Nummer der Tabelle: 12511-0017

* Lade den Datensatz für das neueste verfügbare Jahr als **Flat.csv** herunter
* Importiere ihn in GoogleSheets (Obacht: Trennzeichen ist: ;)
* Lösche nicht benötigte Spalten
* Filtere auf 'männlich' und 'ledig
* Filtere auf NRW (AGS=5 - Spalte H - 1\_Auspraegung\_Code)
* Sortiere den Datensatz nach dem Durchschnittsalter

Bereite die Daten für eine Choroplethenkarte im Datawrapper vor:

* Zeige auf einer Karte, wo die jüngsten, ledigen, weiblichen Personen eingebürgert werden
* Filtere so, dass alle Bundesländer angezeigt werden (aber auch nur die) (Spalte H “1\_Auspraegung\_Code”)
* Filtere so, dass für jedes Bundesland das Durchschnittsalter aller Menschen angezeigt wird, die aus dem vereinigten Königreich (“Vereinigtes Königreich einschl.brit.Überseegebiete”) einwanderten (Spalte U- 4\_Auspraegung\_Label)
* Falls Du Zeit hast, lade die Daten gerne in den Datawrapper und erstelle eine Choroplethenkarte für Deutschland

**Übung 2 \***

* Gehe auf [<https://www-genesis.destatis.de/>]
* suche nach 'Durchschnittsalter der habilitierten Personen'.
* Dann auf 'Werteabruf' und die Tabelle bitte als '**csv**' (nicht flat\_csv) in ./data abspeichern (Es sollten Werte von 2011 bis 2020 enthalten sein)
* Dann importiere die .csv in GoogleSheets (Obacht: Trennzeichen ist: ;)
* Bereinige Kopf und Fußzeile
* Bereinige die Spalten / Spaltennamen

**Beantworte diese Fragen zum Datensatz**   
 - Wie hat sich das Alter bei der Habilitation im Bereich Kunst- und Kunstgeschichte insgesamt verändert?  
 - Wie hat sich das Durchschnittsalter der habilitierten Personen in Deutschland insgesamt verändert?  
 - In welchem Bereich war das Habilitationsalter 2019 bei Frauen am niedrigsten?

Blogpost zur Datenbereinigung:  
[Lisa Charlotte Rost 2019: How to prepare your data for analysis and charting in Excel & Google Sheets](https://blog.datawrapper.de/prepare-and-clean-up-data-for-data-visualization)