[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)  
  
Lest euch den Blog-Artikel durch und vollzieht die darin besprochenen Schritte mit diesem Datensatz hier nach:

* 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.

Dann importiere die .csv in GoogleSheets:

Deutsche Besonderheiten:

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

Hier die Schritte aus dem Blogpost in der Übersicht (vielleicht müsst ihr nicht alle anwenden)

Datei vorbereiten:

1. Daten ansehen und verstehen

2. Datei umbenennen

3. Roh-Daten duplizieren und nie wieder anfassen

4. Quelle(n) in einem Extrablatt speichern (mit Datum), ggf. hier auch Erklärung zu den Daten ablegen

Kopfzeile säubern:

1. Alles oberhalb der Kopfzeile löschen

2. Verbundene Zellen trennen und doppelzeilige Überschriften entfernen

3. Metriken in die Kopfzeile bringen

4. Prüfen, ob der Inhalt der Kopfzeile sinnvoll ist

5. Weil es praktisch ist: erste Zeile (und Spalte) fixieren

Daten bereinigen:

1. Nicht benötigte Spalten löschen

2. Nicht benötigte Zeilen löschen

3. Tausendertrennzeichen löschen

4. Text kürzen

5. Datumsangaben korrigieren

6. Variablen in eigene Spalte ziehen

**Übung zum Datensatz**   
  
Wir müssen auswählen, was wir als in der Grafik darstellen wollen:

* Zeige wie sich das Alter bei der Habilitation im Bereich Kunst- und Kunstgeschichte insgesamt verändert hat
* Zeige, wie sich das Durchschnittsalter der habilitierten Personen in Deutschland insgesamt verändert hat
* In welchem Bereich ist das Habilitationsalter am höchsten?

**\* Sternchen-Übung \***  
  
Aus welchem Gebiet stammen im Schnitt die jüngsten männlichen Eingebürgerten?

* Suche auf [<https://www-genesis.destatis.de/genesis/online>] nach 'Durchschnittsalter der Eingebürgerten: Deutschland, Jahre, Ländergruppierungen/Staatsangehörigkeit, Geschlecht,Familienstand
* Lade den Datensatz als Flat.csv herunter
* Importiere ihn in GoogleSheets
* Lösche nicht benötigte Spalten
* Filtere auf 'männlich' und 'ledig
* Sortiere den Datensatz nach dem Durchschnittsalter