

# **Requirement List**

## **Clean and visualize data: Corona 7-day incidence and deaths cases**

1. Read and clean the relevant data (cases, deaths, sex, age, county)
  - a. Note: Data is partly in German
  - b. The actual data:  
[https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6\\_0/explore?showTable=true](https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0/explore?showTable=true)
  - c. Explanation of this data:  
<https://www.arcgis.com/home/item.html?id=dd4580c810204019a7b8eb3e0b329dd6>
  - d. The counties:  
[https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/917fc37a709542548cc3be077a786c17\\_0/explore](https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/917fc37a709542548cc3be077a786c17_0/explore)
2. ~~If there is missing data, use expectation maximization (Note: We don't expect there to be missing data, it seems like the data is complete, but in case we find missing data throughout working on the project, this is what we'll do)~~
3. Group the data into meaningful units (sex, age, county) and calculate number of cases and number of deaths per part of each group (e.g. male, female)
4. Calculate the 7-day window (=incidence) for cases and deaths, in total and per part of each group
5. **Plotting:**
  - a. Bar-plot total 7-day incidence values on each day for the dates 02.01.2020 until 07.07.2021
  - b. Bar-plot 7-day incidence values depending on the variables
    - i. Age group
    - ii. Sex
    - iii. County
    - iv. Such that different age groups, sexes and counties can, respectively, be compared with each other directly
  - c. Use stacked bars to also put the #deaths in the last 7 days into the plots