|  |  |
| --- | --- |
| **Where you’re at** | **Prompts what to do** |
| I found a dataset in xlsx format | - Make sure to save the source where you found it  - Make a copy  - Clear all formatting, before starting to analyze |
| I found a dataset in csv format | Import it into Excel (either via the “Import” or the “Open” button under Menu > File)  In the import dialogue, make sure to select “delimited” (“getrennt” in German) and in the next step define by checking the adequate and uncheck the inadequate boxes what your delimiter is (Comma? Semicolon? Tab?) |
| I found a **table in a pdf** | You need to open it in Tabula and select the table area (make sure to catch the edges as well), then click extract.  If the extraction preview looks weird try the other extraction method (stream vs lattice)  When considering how to export the extracted data, your default thought should be “csv”, unless there are commas within any of your columns – then select “tsv” |
| I tried opening a **csv file**, but it is messy and doesn’t look right as usual csvs | Usually this happens with scraped data (e.g. from pdfs with Tabula) 🡪 open it in Atom text editor and see whether you can spot what is wrong; think about whether Regular Expression patterns (.\*) can help you clean it up |
| I have a large table with too detailed data that **I want to summarize** | You can summarize it with help of a Pivot table: Mark the area you want to summarize, find the Pivot Table button.  Next you want to select and drag and drop the fields you want to have in your Pivot table |
| I want to find the **largest/smallest/median/mean** value in my dataset | Use functions like  =MIN  =MAX  =MEDIAN  =MEAN  When you start typing them, Excel will show you a preview what it needs to complete the task  Another way for the smallest/largest value is to mark the area and add filters; these double as sorting buttons (descending/ascending) |
| I want to highlight all **values of a certain range** in my dataset | Look into Conditional formatting |
| I want to **merge two datasets** together without having to copy-paste single values around | Look into the function =VLOOKUP (German: =SVERWEIS) |
| I used a bunch of formulas, copy pasted things around, now I got a **#REF error** | That means Excel cannot establish any longer where the information came from. To avoid this, you should clear your data from formulas from time to time before you proceed. To do that, you copy your table and right-click in an empty area look for “Paste Special” (German: “Inhalte einfügen”) and select “Values” (German: “Werte”) |
| I would like to **change the orientation** of the table (what is in rows should be in columns and vice versa) | Copy your table, right click in an empty area, look for “Paste Special” (German: “Inhalte einfügen”) and select “Transpose” (German: “Transponieren”) |
| I have a datasets with different spellings for the same thing | Use Open Refine (Text facet, Cluster)  <https://gcgruen.github.io/teaching/open-refine/> |