The common matrix throughout all the document was year and country and the datasets shown above fulfilled the conditions the most.

The data sets that we decided to choose:

* annual-freshwater-withdrawals: Limited missing data and has data on most countries which are common across documents.
* Premature\_deaths\_due\_to\_UNSAFE\_WASH: Has limited missing data and the data provided is available for most countries.
* Proportion of bodies of water with good ambient water quality: Has limited data. There is only data for one year however the years and the countries the data is provided for is sufficient.
* Wastewater Discharges Per Year Per Country: Has limited missing data and the data provided is very informative. Considering this dataset to be the main we decided to choose common countries according to this document.

The datasets that we did not choose:

* Generation and discharge of wastewater in volume: It was quite informative and this dataset was a detailed version of the ‘Wastewater Discharges Per Year Per Country’ dataset. However, the main issue was that the countries which were most common across the document were not available in this dataset.
* Mortality rate attributed to unsafe water, unsafe sanitation, and lack of hygiene (per 100,000 population): This dataset has limited quality data as the data provided was only for one country. However, the data under the country had many missing data values.