## Data Dictionary

| **Column Name** | **Description** |
| --- | --- |
| wwtp\_jurisdiction | State, DC, US territory, or Freely Associated State jurisdiction name (2-letter abbreviation) in which the wastewater treatment plant provided in 'wwtp\_id' is located. |
| wwtp\_id | A unique identifier for wastewater treatment plants. This is an arbitrary integer used to provide a unique, but anonymous identifier for a wastewater treatment plant. This identifier is consistent over time, such that the same plant retains the same ID regardless of the addition or subtraction of other plants from the data set. |
| reporting\_jurisdiction | The CDC Epidemiology and Laboratory Capacity (ELC) jurisdiction, most frequently a state, reporting these data (2-letter abbreviation) |
| sample\_location | Sample collection location in the wastewater system, whether at a wastewater treatment plant (or other community level treatment infrastructure such as community-scale septic) or upstream in the wastewater system |
| sample\_location\_specify | A unique identifier for "upstream" sample locations. Specifically, when 'sample\_location' is "upstream", this field has a non-empty value, which provides a unique, but anonymous identifier for the upstream sample collection sites. This identifier is consistent over time, such that the same sample collection site retains the same ID regardless of the addition or subtraction of other sample collection sites from the data set. |
| key\_plot\_id | A unique identifier for the geographic area served by this sampling site, called a sewershed. This is an underscore-separated concatenation of the fields 'wwtp\_jurisdiction', 'wwtp\_id', and, if 'sample\_location' is "upstream", then also 'sample\_location\_specify', and sample\_matrix. |
| county\_names | The county and county-equivalent names corresponding to the FIPS codes in 'county\_fips' |
| county\_fips | 5-digit numeric FIPS codes of all counties and county equivalents served by this sampling site (i.e., served by this wastewater treatment plant or, if 'sample\_location' is "upstream", then by this upstream location). Note that multiple sampling sites or treatment plants may serve a single county, and that a single sampling site or treatment plant may serve multiple counties. Counties listed may be entirely or only partly served by this sampling site |
| population\_served | Estimated number of persons served by this sampling site (i.e., served by this wastewater treatment plant or, if 'sample\_location' is "upstream", then by this upstream location). |
| date\_start | The start date of the interval over which the metric is calculated. Intervals are inclusive of start and end dates. |
| date\_end | The end date of the interval over which metric is calculated. Intervals are inclusive of start and end dates. |
| ptc\_15d | The percent change in SARS-CoV-2 RNA levels over the 15-day interval defined by 'date\_start' and 'date\_end'. Percent change is calculated as the modeled change over the interval, based on linear regression of log-transformed SARS-CoV-2 levels. SARS-CoV-2 RNA levels are wastewater concentrations that have been normalized for wastewater composition |
| detect\_prop\_15d | The proportion of tests with SARS-CoV-2 detected, meaning a cycle threshold (Ct) value <40 for RT-qPCR or at least 3 positive droplets/partitions for RT-ddPCR, by sewershed over the 15-day window defined by 'date\_start' and "date\_end'. The detection proportion is the percent calculated by dividing the 15-day rolling sum of SARS-CoV-2 detections by the 15-day rolling sum of the number of tests for each sewershed and multiplying by 100. |
| percentile | This metric shows whether SARS-CoV-2 virus levels at a site are currently higher or lower than past historical levels at the same site. 0% means levels are the lowest they have been at the site; 100% means levels are the highest they have been at the site. Public health officials watch for increasing levels of the virus in wastewater over time and use this data to help make public health decisions. |
| sampling\_prior | Indicates whether the site was collecting wastewater samples before or on December 1, 2021. |
| first\_sample\_date | The first date samples were collected at a site. |