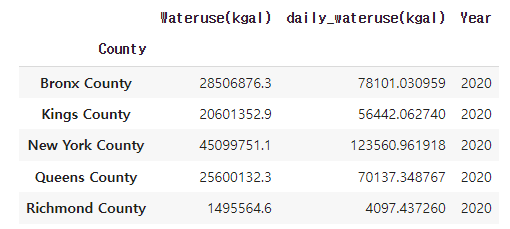
**Water Use Data**

**Abbreviation and Summary of Input Data sets**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CSV Name** | **Abbreviation** | **Description** | **Sector** | **Spatial** |
| **Energy\_and\_Water\_Data\_Disclosure\_for\_Local\_Law\_84\_2021\_\_Data\_for\_Calendar\_Year\_2020\_.csv** | **DF\_ZIP** | **Data and metrics on water and energy consumption in privately owned buildings over 25,000 ft2 and in City-owned buildings over 10,000 ft2. (2020)** | **PrimaryProperty Type (Residential:8, Industry:4, Commercial: 63)** | **Borough,**  **NYCBuilding Identification Number(BIN),City,Postcode, Longitude, Latitude** |
| **usco2015v2.0.csv** | **DF\_County** | **Water-use estimates for 2015 that are aggregated to the county level in the United States.** | **Irrigation, Livestock, Aquaculture, Mining, Thermoelectric, Public Supply, Total Withdrawal** | **County, FIPS** |

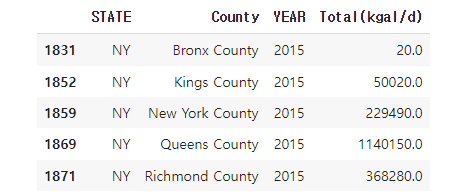
DF\_ZIP is the 2020 annual NYC water consumption data for privately owned buildings over 25,000 ft2 and City-owned buildings over 10,000 ft2. The data set includes 75 types of primary properties that can be categorized into groups (Residential:8, Industrial:4, Commercial: 63). Spatial granularity for the data is available in Borough, NYC Building Identification Number, specific address line, City, Borough, Postcode, and Longitude and Latitude. The data amount of water used in this data is in kgal per year and the year is 2020.

**DF\_County** contains the county-level water-use data that support the state-level estimates in Dieter. This dataset contains data for public supply, domestic, irrigation, thermoelectric power, industrial, mining, livestock, and aquaculture water-use categories for the whole US. The dataset contains total population data and water-use estimates for 2015 for the following categories: Public supply, domestic, irrigation, thermoelectric power, industrial, mining, livestock, and aquaculture. Data is aggregated to the county level.



**DF\_ZIP** when aggregated by countyand divided by 365 for daily estimated values of daily water use looks like the above graph.

**DF\_County** aggregated by county level and filtered for only NYC counties looks like the below plot.



When aggregated at the county level , there are differences in daily water use estimates. Calculation of standardized difference is done using formula

**Standardized\_Difference=**

**(DF\_County[**Total(kgal/d)**]-DF\_ZIP[**daily\_wateruse(kgal)]**)/ DF\_ZIP[**daily\_wateruse(kgal)**]**

|  |  |
| --- | --- |
| County | Standardized\_ Difference |
| Bronx | -0.999744 |
| Kings | -0.113782 |
| New York | 0.857302 |
| Queens | 15.255961 |
| Richmond | 88.880571 |

**Number of unique zip codes in DF\_ZIP in each county**

Total 243 unique zip codes

New York County 86

Queens County 71

Kings County 44

Richmond County 14

Bronx County 28

**Data Processing**

Each steps explained in detail on [Waterusage\_data.ipynb](https://github.com/vaishu1396/Hardening_NYC_Infrastructure/blob/main/Waterusage_data.ipynb)

**Limitation**

Among public supply, domestic, irrigation, thermoelectric power, industrial, mining, livestock, and aquaculture water-use categories for the whole US, we only used elements that goes under category of residential, commercial and industrial. Also, water use percentage calculated for each sectors(type of industry) is based on privately owned buildings over 25,000 ft2 and City-owned buildings over 10,000 ft2 at New York City.

Sectors(type of properties) in each zip code from DF\_ZIP and applied to DF\_County. As a result, final output data can cover missing values from DF\_ZIP. However, DF\_County is an estimated water use data therefore, the final output data is also an estimate (from the county-level input database) and it can be way off from the input data. Therefore, more effort to collect actual data to inform decision-making is needed. The final output data can be used in calculation, modeling, or visualization, but comparing the result or output with real standardized data from DF\_ZIP is recommended for better quality.