**Information related to the Dataset**

There are 48 features with values for each of the 876 monitors (observations). The data comes from the US [Environmental Protection Agency (EPA)](https://www.epa.gov/), the [National Aeronautics and Space Administration (NASA)](https://www.nasa.gov/), the US [Census](https://www.census.gov/about/what/census-at-a-glance.html), and the [National Center for Health Statistics (NCHS)](https://www.cdc.gov/nchs/about/index.htm).

| **Variable** | **Details** |
| --- | --- |
| **id** | Monitor number – the county number is indicated before the decimal – the monitor number is indicated after the decimal **Example**: 1073.0023 is Jefferson county (1073) and .0023 one of 8 monitors |
| **fips** | Federal information processing standard number for the county where the monitor is located – 5 digit id code for counties (zero is often the first value and sometimes is not shown) – the first 2 numbers indicate the state – the last three numbers indicate the county **Example**: Alabama’s state code is 01 because it is first alphabetically (note: Alaska and Hawaii are not included because they are not part of the contiguous US) |
| **Lat** | Latitude of the monitor in degrees |
| **Lon** | Longitude of the monitor in degrees |
| **state** | State where the monitor is located |
| **county** | County where the monitor is located |
| **city** | City where the monitor is located |
| **CMAQ** | Estimated values of air pollution from a computational model called [**Community Multiscale Air Quality (CMAQ)**](https://www.epa.gov/cmaq) – A monitoring system that simulates the physics of the atmosphere using chemistry and weather data to predict the air pollution – ***Does not use any of the PM2.5 gravimetric monitoring data.*** (There is a version that does use the gravimetric monitoring data, but not this one!) – Data from the EPA |
| **zcta** | [Zip Code Tabulation Area](https://www2.census.gov/geo/pdfs/education/brochures/ZCTAs.pdf) where the monitor is located – Postal Zip codes are converted into “generalized areal representations” that are non-overlapping – Data from the 2010 Census |
| **zcta\_area** | Land area of the zip code area in meters squared – Data from the 2010 Census |
| **zcta\_pop** | Population in the zip code area – Data from the 2010 Census |
| **imp\_a500** | Impervious surface measure – Within a circle with a radius of 500 meters around the monitor – Impervious surface are roads, concrete, parking lots, buildings – This is a measure of development |
| **imp\_a1000** | Impervious surface measure – Within a circle with a radius of 1000 meters around the monitor |
| **imp\_a5000** | Impervious surface measure – Within a circle with a radius of 5000 meters around the monitor |
| **imp\_a10000** | Impervious surface measure – Within a circle with a radius of 10000 meters around the monitor |
| **imp\_a15000** | Impervious surface measure – Within a circle with a radius of 15000 meters around the monitor |
| **county\_area** | Land area of the county of the monitor in meters squared |
| **county\_pop** | Population of the county of the monitor |
| **Log\_dist\_to\_prisec** | Log (Natural log) distance to a primary or secondary road from the monitor – Highway or major road |
| **log\_pri\_length\_5000** | Count of primary road length in meters in a circle with a radius of 5000 meters around the monitor (Natural log) – Highways only |
| **log\_pri\_length\_10000** | Count of primary road length in meters in a circle with a radius of 10000 meters around the monitor (Natural log) – Highways only |
| **log\_pri\_length\_15000** | Count of primary road length in meters in a circle with a radius of 15000 meters around the monitor (Natural log) – Highways only |
| **log\_pri\_length\_25000** | Count of primary road length in meters in a circle with a radius of 25000 meters around the monitor (Natural log) – Highways only |
| **log\_prisec\_length\_500** | Count of primary and secondary road length in meters in a circle with a radius of 500 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_prisec\_length\_1000** | Count of primary and secondary road length in meters in a circle with a radius of 1000 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_prisec\_length\_5000** | Count of primary and secondary road length in meters in a circle with a radius of 5000 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_prisec\_length\_10000** | Count of primary and secondary road length in meters in a circle with a radius of 10000 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_prisec\_length\_15000** | Count of primary and secondary road length in meters in a circle with a radius of 15000 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_prisec\_length\_25000** | Count of primary and secondary road length in meters in a circle with a radius of 25000 meters around the monitor (Natural log) – Highway and secondary roads |
| **log\_nei\_2008\_pm25\_sum\_10000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 10000 meters of distance around the monitor (Natural log) |
| **log\_nei\_2008\_pm25\_sum\_15000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 15000 meters of distance around the monitor (Natural log) |
| **log\_nei\_2008\_pm25\_sum\_25000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 25000 meters of distance around the monitor (Natural log) |
| **log\_nei\_2008\_pm10\_sum\_10000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 10000 meters of distance around the monitor (Natural log) |
| **log\_nei\_2008\_pm10\_sum\_15000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 15000 meters of distance around the monitor (Natural log) |
| **log\_nei\_2008\_pm10\_sum\_25000** | Tons of emissions from major sources data base (annual data) sum of all sources within a circle with a radius of 25000 meters of distance around the monitor (Natural log) |
| **popdens\_county** | Population density (number of people per kilometer squared area of the county) |
| **popdens\_zcta** | Population density (number of people per kilometer squared area of zcta) |
| **nohs** | Percentage of people in zcta area where the monitor is that **do not have a high school degree** – Data from the Census |
| **somehs** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was **some high school education** – Data from the Census |
| **hs** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was completing a **high school degree** – Data from the Census |
| **somecollege** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was completing **some college education** – Data from the Census |
| **associate** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was completing an **associate degree** – Data from the Census |
| **bachelor** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was a **bachelor’s degree** – Data from the Census |
| **grad** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was a **graduate degree** – Data from the Census |
| **pov** | Percentage of people in zcta area where the monitor is that lived in [**poverty**](https://aspe.hhs.gov/2008-hhs-poverty-guidelines) in 2008 - or would it have been 2007 guidelines??<https://aspe.hhs.gov/2007-hhs-poverty-guidelines> – Data from the Census |
| **hs\_orless** | Percentage of people in zcta area where the monitor whose highest formal educational attainment was a **high school degree or less** (sum of nohs, somehs, and hs) |
| **urc2013** | [2013 Urban-rural classification](https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf) of the county where the monitor is located – 6 category variable - 1 is totally urban 6 is completely rural – Data from the [National Center for Health Statistics](https://www.cdc.gov/nchs/index.htm) |
| **urc2006** | [2006 Urban-rural classification](https://www.cdc.gov/nchs/data/series/sr_02/sr02_154.pdf) of the county where the monitor is located – 6 category variable - 1 is totally urban 6 is completely rural – Data from the [National Center for Health Statistics](https://www.cdc.gov/nchs/index.htm) |
| **aod** | Aerosol Optical Depth measurement from a NASA satellite – based on the diffraction of a laser – used as a proxy of particulate pollution – unit-less - higher value indicates more pollution – Data from NASA |