BayREN Bay Area Existing Buildings Dashboard Data FAQ

How do the data sources in the dashboard differ from nationally available data sources such as ComStock and ResStock?

ComStock[1] and ResStock[2], published by the National Renewable Energy Laboratory (NREL), are representative national datasets for existing buildings. Those datasets aim to create representative energy consumption for a city, state, or Census tract. The building quantities for any selected region in ComStock and ResStock are based on statistically sampled national datasets using the Residential Energy Consumption Survey and the Commercial Building Energy Consumption Survey among other sources. NREL translates these datasets to common building prototypes and then provides a statistically driven expectation of the number of each type of building by use, age, and size.

By contrast, the Bay Area Existing Buildings Dashboard uses parcel-level data based on tax assessor data from each county and the ReGrid[3] dataset. The use of each building is defined at a parcel level. In many cases area and building age are provided as well. While there are gaps in these data sources, they are expected to be more accurate than sampled data from NREL as the starting point more closely reflects the actual size and use of the buildings. Where gaps exist, the statistically representative relationships provided in NREL's ResStock and ComStock datasets have been used to fill those gaps.

Energy use data in the Bay Area Existing Buildings Dashboard uses the NREL ResStock and ComStock datasets by matching the building characteristics to the representative NREL prototype (or in some cases, blending results from a 2 or 3 NREL prototypes to represent certain types of Bay Area building types such as mixed-use buildings). Estimated energy is then assigned to Bay Area buildings as an energy use intensity normalized per floor area (in units of kBtu/SF and kWh/SF).

Next, for building types for which comprehensive benchmark data is available (representing actual metered energy trends for common building types), the energy use intensity is scaled so that the average of Bay Area buildings matches the average of benchmark energy data. Commercial energy use benchmark values were acquired from the 2022 California Commercial End-Use Survey (CEUS) and the San Francisco Existing Buildings Energy Performance Ordinance Report. To get values more closely aligned to BayREN territory, the California CEUS was further segmented to only include PG&E data, allowing the team to exclude values for LADWP, SMUD, and SDG&E. Finally, total energy consumption is derived by multiplying this energy intensity by known building floor area.

---

[1] https://comstock.nrel.gov/
[2] https://resstock.nrel.gov/datasets
[3] https://regrid.com/

How has the accuracy of the building counts been verified, and how accurate is the data believed to be?

Building count data is based on aggregation of the parcels based on use as defined in the County Assessor and ReGrid data. As parcel use designations vary and are sometimes misrecorded or incomplete in these datasets, numerous fields have been used to refine the categorization of buildings appropriately. To determine the accuracy of the data, statistical random sampling of the overall dataset and specific use types were performed and manually reviewed. These reviews resulted in corrections to the data to improve accuracy. Based on this approach, the final dataset is estimated to be 85-90% accurate, with the remaining inaccuracies identified being difficult to correct without manual review of all impacted parcels or compromising accuracy in another manner. The steps taken to improve the data quality are noted below:

- Based on review, ReGrid derived data fields overcount residential properties, particularly single-family residential. To account for this, County Assessor data was used where available to reclassify structures.
- Land Use data was unavailable for the Presidio and incorrectly categorized as almost entirely residential in ReGrid. Therefore, to correct, buildings were classified based on OpenStreetMap and basemap data.

- Specific measures to improve accuracy of education buildings include:
  - Used Owner field to identify properties including the terms:
    - Union S D
    - Unified S D
    - Regents of the University
  - Religious schools are included in Public Assembly instead of Education.
- Specific measures to improve accuracy of medical buildings include:
  - The Glen Ellen State Hospital was set to “No Energy Use” since it is currently abandoned.
- Specific measures to improve accuracy of multi-family buildings include:
  - Single family buildings over 15,000sq ft were switched to multi-family 5+ units.
  - Buildings with a Use Description that included Apartments were moved to multi-family 5+ units. Exceptions were those that specified less than 4 units.
- Specific measures to improve accuracy of single-family homes include:
  - Single family structures that share a wall with an adjacent single-family structure were moved to Single family attached.
  - Duplexes can be found in both single family attached and multi-family 2-4 units. Where clearly specified as duplex, the building is classified as multi-family

What limitations and known inaccuracies remain?

The known inaccuracies that have been identified are listed below. Unfortunately, corrective actions for these inaccuracies are challenging without manual review or compromising data accuracy in another manner. Assuming data quality in the underlying County Assessor or ReGrid data improves, future updates of the dashboard may improve data quality.

- Detached garages are often categorized as single-family homes as they are identified as structures in the ReGrid data and assigned a building footprint and provided with a single-family residential use based on the parcel designation. Filtering large garages from the dataset would also remove small homes. Currently, single-family homes under 600 SF are excluded from the dashboard counts. Increasing this size threshold jeopardizes excluding actual homes with smaller areas from the count.
- Standalone garages are occasionally being captured as Public Assembly based on the underlying dataset. No defining characteristics that could be systematically adjusted were identified in the available data.
- Some retirement (assisted living) communities show up in Medical – Other. A method to separate out apartment/condo style from those who offer more intensive care was not determined.
- Food Service is undercounted in existing buildings and is often classified under Retail or other categories. This arises from Food Service often being retrofit into retail spaces such as in strip malls or first floor retail spaces without a change to land use designation. OpenStreetMap data was investigated as a source of improvement for Food Service counts, but it could not be confirmed that this source would provide greater accuracy.
- RV Parks and Campgrounds were included under Entertainment & Culture and may include temporary structures such as a large RV that may be moved.