

URBAN FOOTPRINT

UrbanFootprint Technical Documentation NATIONWIDE PARCEL CANVAS CREATION METHODOLOGY

Overview	3
Methodology	3
Standardizing Geometries	3
Duplicate IDs	3
Exact and near duplicate geometries	4
Nested geometries	4
Overlapping geometries	5
Creating Crosswalk to UrbanFootprint Land Use Codes	6
UrbanFootprint’s Generalized Land Use Classification (GLUC) system	6
Dwelling Units	7
Unit Counts from CoreLogic Attributes	7
Dealing with Condominiums	8
Imputing Missing Data	8
Assigning Units from Census Data	10
Building Area by Type	10
Population and Households	12
Employment by Type	12
Disaggregation Logic	12
Parcel Area by Type	13
Built Form	13
Density-Based Classification	13
Non-Density Based Classification	14
Intersection Density and Land Development Category	16
Irrigated Area	16
Appendix	16
Appendix A: UrbanFootprint - Corelogic Crosswalk	16
Appendix B: Source-Specific Modifications	16

Overview

UrbanFootprint scenarios are built on a base-year geospatial dataset that describes the existing environment. This detailed “canvas” of data constitutes a baseline assessment of land use, demographic characteristics, and other conditions, providing the context for scenario painting and the foundation for analysis by the modules. Base canvas creation processes involve cleaning, compiling, and processing geographic and tabular data that come from a wide variety of sources. The UrbanFootprint (UF) canvas encompasses data from the following general categories:

- Residential fields
- Employment fields
- Parcel area fields
- Building area fields
- Irrigated area fields
- Roadway intersection density
- Built form type and land development category

This documentation covers the process used to generate the **parcel-scale** base canvas. For documentation on the process to build the block-scale base canvas, follow this link [here](#). The parcel canvas builds off of much of the logic used to generate the block canvas, but substitutes more refined parcel scale data where available. The default source for parcel information is CoreLogic, which is a leading provider of real estate data in the United States.

Methodology

The following section gives a detailed overview of the steps required to transform raw parcel data into an UrbanFootprint-ready base canvas. The section covers geospatial transformations required to create unique, non-overlapping parcel areas; assumptions and fixes made to incorporate residential unit and building area fields into the base canvas fields; the process to distribute employment data from the block scale to underlying parcels; and our typing process to give every parcel a built form key.

Standardizing Geometries

Raw parcel data can have a few issues with the geometries that make it hard for sketch planning tools like UrbanFootprint to work with. Common issues include duplicate IDs, stacked geometries, nested geometries and overlapping geometries. As a post process, very tiny geometries (less than 1 sqm) are dropped.

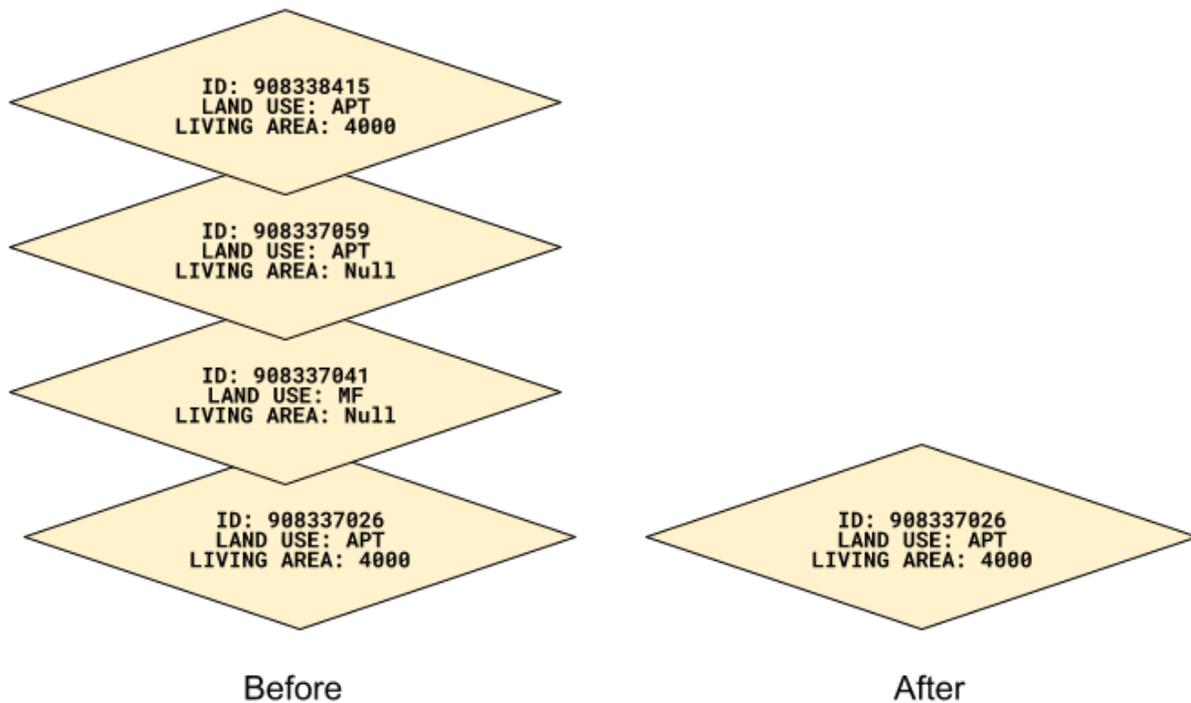
Duplicate IDs

In the rare cases where there were duplicate parcel IDs, the first parcel is retained and the duplicates are dropped. A verification step also ensures that duplicate parcel IDs don't exist in the dataset.

Exact and near duplicate geometries

In some cases such as parcels for condominiums, multiple geometries that are exactly the same or nearly the same are stacked on top of each other. These geometries are dissolved into a single flat geometry. For numeric data, the maximum across all parcel attributes was retained. For non-numeric data, the attributes of the parcel with the lowest parcel ID are retained.

Figure 1: Example of Stacked Geometries

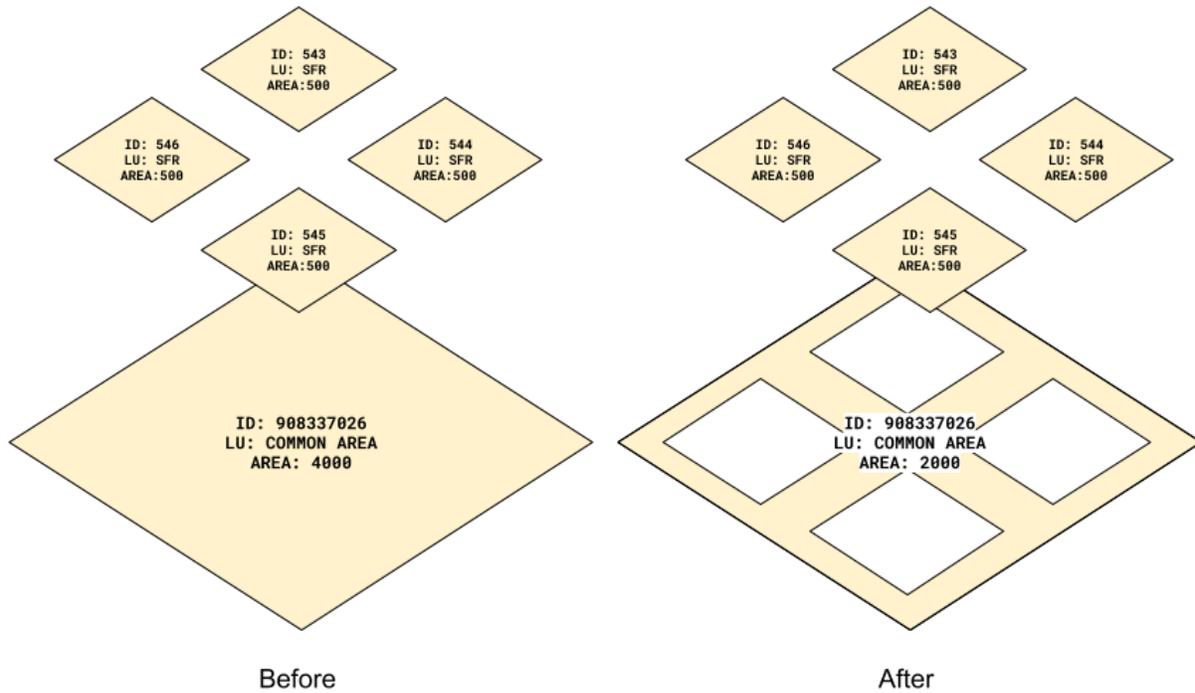


Nested geometries

Nested geometries sometimes occur in master planned developments or subdivided residential neighborhoods where the outdoor area is owned by one entity, but each unit is privately owned. In these cases, a parent geometry (eg: a planned unit development parcel) can have several children geometries (eg: the planned units) as individual parcels completely contained by the larger parent geometry. For scenario development to correctly account for the total land area in these cases, it is essential that no land is double counted, meaning that the parent geometry does not include the area of the children parcels. To fix this, the children geometries are cut out of the parent geometry.

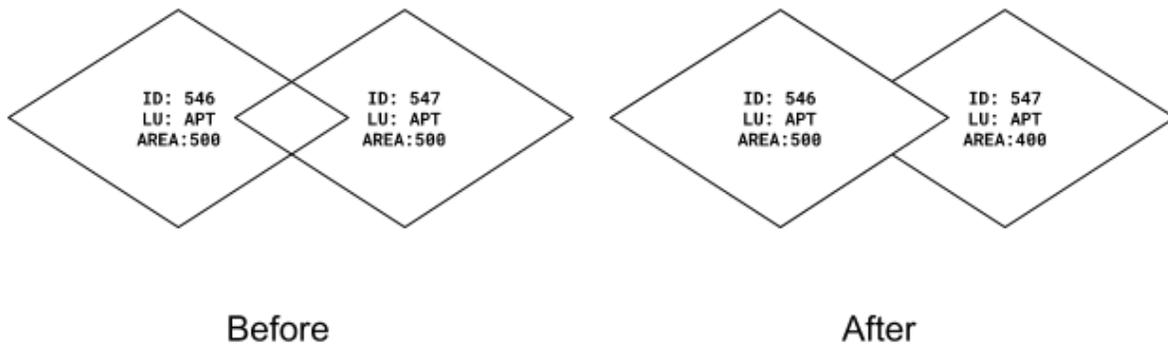
The result is that the parent geometry encompasses the shared common area, while each child parcel contains the information for the particular unit.

Figure 2: Nested Geometries Example



Overlapping geometries

Finally, there are parcels that overlap but are not entirely contained in one. This most often happens in condominiums where the floor plans of each unit differ across floors. For this type of parcelization to work in UF, it is important that no geometries are overlapping, even if they represent different floors in a structure. For these cases, the lower parcel ID has its geometry cut out and the overlapping portion is retained for the parcel with the higher parcel ID. The parcel area is recalculated for each new geometry, but each geometry retains its original attribute data, meaning that the unit number and building area are not impacted by the change in geometry.



Creating Crosswalk to UrbanFootprint Land Use Codes

The next step in the parcel canvas creation process is to crosswalk original land use code designations into standardized codes that facilitate the process to type parcels into UF built form types (Refer [Built Form](#) Section).

UrbanFootprint's Generalized Land Use Classification (GLUC) system

Most jurisdictions represent land use at the parcel scale, using land use codes that reflect specific uses. UF “translates” local data and land use information to represent development in terms of UF built forms, which are foundational to scenario development and analysis. While land use and urban form are the subject of both local/regional land use classification systems and UF built forms, they are distinct languages. Land use codes are predominantly use-based and static. UF built form types are more form-based, and are dynamic and expansive, designed to capture the universe of built form and land use combinations. In order for UF built forms and the variety of local land use classification systems to relate to each other, we developed the Generalized Land Use Classification (GLUC) system. It acts as a translator between the various land use dialects and UF's hybrid built form/land use language.

To translate CoreLogic data into UF, it is necessary to associate each land use code with a UF land use designation, creating a “crosswalk” to guide the translation process. Every UF built form is assigned one or more generalized land use designations, each of which is associated with one or more built form types. Using the crosswalk to narrow the range of built form types that each land use code could be translated to, the UF translation algorithm matches the demographic data of each parcel to the closest built form type.

For example, CoreLogic's 'APARTMENT' land use code (#106) is crosswalked to UF Land Use Designation 'residential multifamily - all'. The crosswalk informs the base canvas creation that these parcels are of multifamily uses, but could vary in the density of the actual structure. In contrast, CoreLogic's 'HIGH RISE CONDO' code (#117) is crosswalked to 'residential multifamily - high', which is more restrictive in the types of built forms that it can translate into.

The crosswalk between CoreLogic land use codes and UF land use designations can be found in [Appendix A](#).

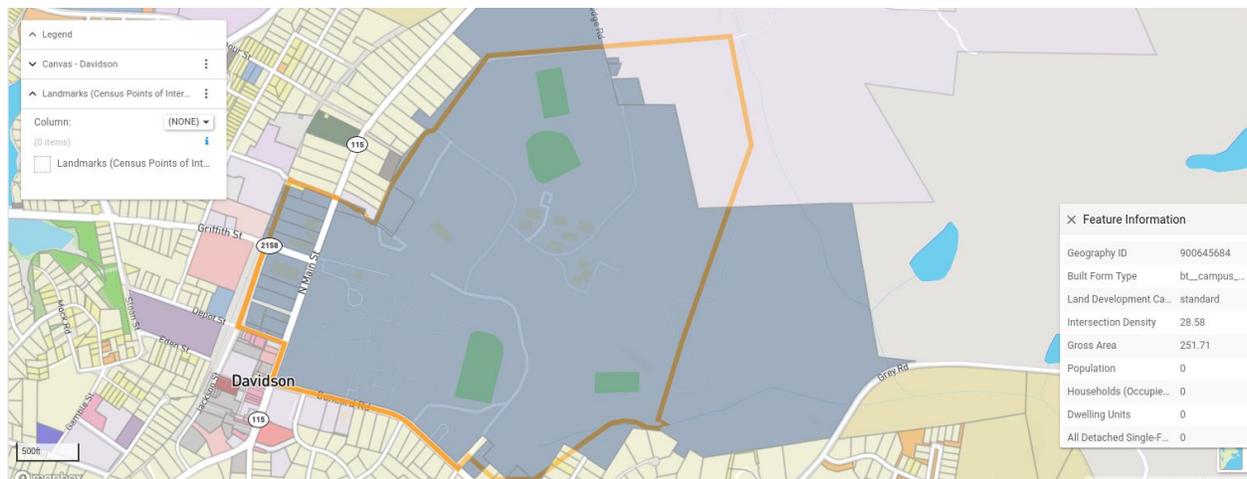
Supplemental Datasets

In addition to CoreLogic's Land Use data, supplemental datasets provide more specificity where Land Use data is lacking. Examples of supplemental datasets include OpenStreetMap (OSM), Schools, Parks, Military, Prison Facilities etc. A parcel is tagged with a land use from corelogic and land uses from these supplemental datasets. The land use for a parcel is then "resolved" by picking **one** land use from this set of options. The land use that is more specific is given priority, unless there is an exception-based rule. For example, between a choice of Commercial - All vs. Commercial - Restaurant, the Restaurant Land Use is chosen.

Supplemental datasets can be either polygons or points. For polygon based datasets, at least 25% of the parcel must be covered by the dataset to receive a tag from the dataset. For point datasets, the parcel which the point intersects is given a direct tag. The point dataset is also buffered by 50m and parcels that touch this buffer are also tagged. These buffered tags are only used in the "resolver" if the parcel's competing Land Use code was vacant or null.

The process of typing parcels using a polygon dataset can be seen below in Figure 1. The campus geometry defined by the Census TIGER dataset is the yellow boundary, while parcels typed as Campus College - Low show up in blue. All parcels where more than 25% of the polygon intersect with the landmark geometry are retyped accordingly. The large parcel in the upper right side of the image is not typed as campus, as less than 25% of its area intersects with the landmark polygon.

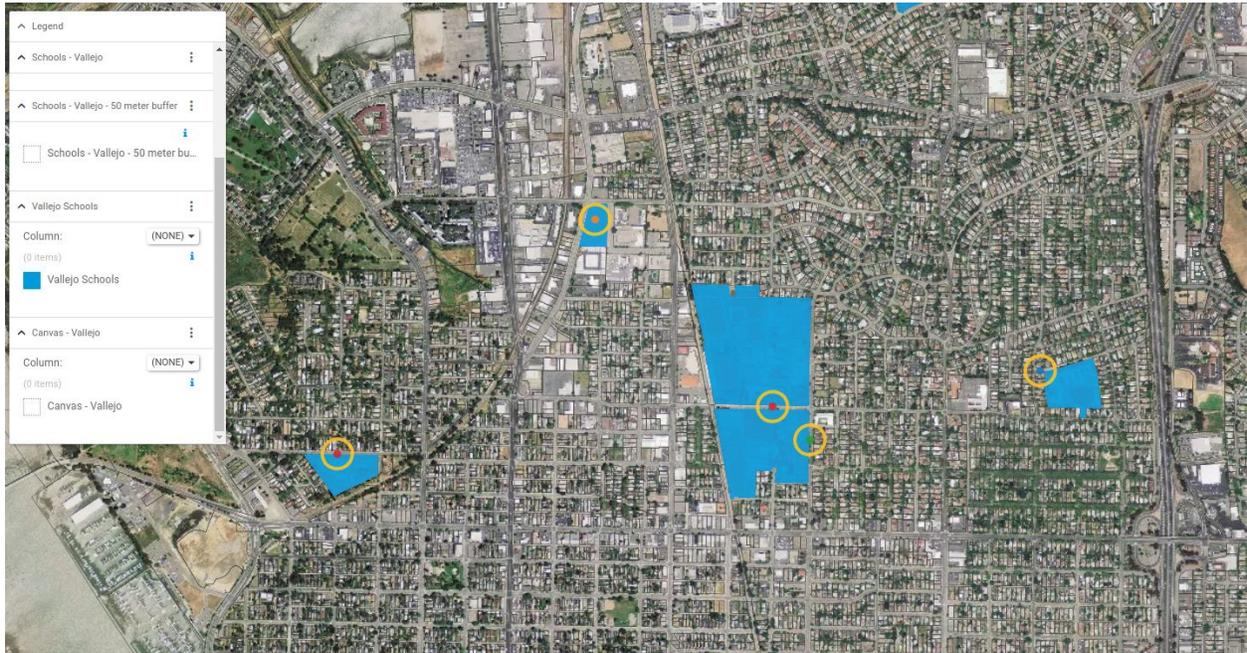
Figure 1: Polygon-based typing using Census Landmarks Dataset



An example of a point-based dataset is the SABINS schools dataset. This shapefile provides point locations of K-12 schools in the National Center for Education Statistics'

Common Core of Data (CCD). Figure 2 below demonstrates both cases of direct typing and buffered typing for point-based datasets. The buffered points picked up the null land use parcels and typed them correctly as a school, and the parcels where the point fell directly on the parcel was tagged as a school.

Figure 2: Point-based Typing of Schools using SABINS Dataset



Finally, some sources also directly indicate the built form to be typed with instead of using the UF Land Use Codes. For example, golf courses are directly typed as the built form and the UF Land Use Code is only used for disaggregating data. (Refer the [Disaggregate Block Data to Parcels](#) Section for more details).

A full list of the supplemental datasets used is shown in Table 1.

Table 1: Supplemental Datasets for Built Form Typing and Disaggregation

Dataset	Source	Geometry Type	UF Land Use/Built Form
OSM	Overpass API	Both Point and Polygon	See Table 2
Parks	Esri, TomTom North America, Inc. California Protected Areas Database (CPAD)	Polygon	Parks and recreation

NATIONWIDE PARCEL CANVAS CREATION

Landmarks	Census TIGER Landmarks	Polygon	Uses the MAF/TIGER Feature Class Code (MTFCC) ¹ . See Table 3
Military	Census TIGER Military Installations	Polygon	Military
Prison Facilities	Homeland Infrastructure Foundation-Level Data (HIFLD)	Polygon	Correctional Facilities
Places of Worship	HIFLD	Point	Religious Centers
Major Sporting Venues	HIFLD	Point	Commercial Recreation except for Golf Courses which have their own Built Form
Fire Stations	HIFLD	Point	Fire/EMS
Schools	School Attendance Boundary Information System (SABINS)	Point	Primary and Secondary Education ²
Hospitals	HIFLD	Point	Hospitals
Airports	HIFLD	Point	Air Transportation
Colleges and Universities	HIFLD	Point	Higher Education

Table 2: OSM Tags Crosswalk

UF Land Use	OSM Land Use/Amenity Tags
Cropland	farmland, farm, farmyard, orchard
Commercial - All	commercial
Industrial - All	industrial
Forest - All	forest
Cemeteries	cemetery
Retail	retail
Detention/Utilities	reservoir, basin
Greenbelt	conservation

¹ The [MAF/TIGER Feature Class Code \(MTFCC\)](#) is a 5-digit code assigned by the Census Bureau intended to classify and describe geographic objects or features. These codes can be found in the TIGER/Line products

² The schools are classified as elementary, middle or high based on the highest grade offered at the school. Schools are typed as 'urban' if the intersection density of the surrounding census block is greater than 150 intersections per square mile, and 'non-urban' if the intersection density falls below this threshold.

NATIONWIDE PARCEL CANVAS CREATION

Vineyard	vineyard
Religious Centers	places_of_worship
Primary and Secondary Education	school
Cemeteries	grave_yard, cemetery
Fire/EMS	fire_station, police

Table 3: Census MAF/TIGER Feature Class Code Crosswalk

Built Form	MTFCC Code	MTFCC Description
Campus - College (High)	K2540	University or College,
Airport	K2456	Airport–Intermodal, Transportation, Hub/Terminal
	K2457	Airport–Statistical Representation
	K2451	Airport or Airfield
Park	K2180	Park
	K2181	National Park Service Land
	K2182	National Forest or Other Federal Land
	K2183	Tribal Park, Forest, or Recreation Area
	K2184	State Park, Forest, or Recreation Area
	K2185	Regional Park, Forest, or Recreation Area
	K2186	County Park, Forest, or Recreation Area
	K2187	County Subdivision Park, Forest, or Recreation Area
	K2188	Incorporated Place Park, Forest, or Recreation Area
	K2189	Private Park, Forest, or Recreation Area
	K2190	Other Park, Forest, or Recreation Area
Golf Course	K2561	Golf Course
Hospital	K1231	Hospital/Hospice/Urgent Care Facility
Urban Civic	K2165	Government Center
Correctional Facility	K1235	Juvenile Institution
	K1236	Local Jail or Detention Center
	K1237	Federal Penitentiary, State Prison, or Prison Farm
Cemetery	K2582	Cemetery

Dwelling Units

Where available, the UrbanFootprint parcel canvas uses CoreLogic data to populate the dwelling units field. To make the CoreLogic data usable, it first needs to go through a series of standardization steps to address abnormalities present in the raw data. The process is as follows:

1. Assign unit counts from raw CoreLogic data
2. Curation to address condominium overcounting
3. Imputing missing data from land use data
4. Assigning units from census data where necessary

Unit Counts from CoreLogic Attributes

CoreLogic provides two attribute columns that help us identify the number of dwelling units present on a parcel: 'Building Units' and Units Number'. The definition of each column can be seen in Table 4 below.

Table 4: CoreLogic Dwelling Unit Attributes

CoreLogic Column	Column Description
BUILDING UNITS	Total Number of Buildings on the Parcel
UNITS NUMBER	Number of Residential, Apartment, or Business Units

The first field, Building Units, is useful for identifying single-family homes. However, it is insufficient for calculating the total number of units for multifamily units as it would only count the building that houses all of the units. For that reason, we use the second field, Units Number, to count the number of units for multifamily parcels.

As the Units Number field contains information on business units as well as residential units, the first step of our curation is to differentiate between the two. We do this by using the CoreLogic land use codes to classify parcels as residential, commercial/employment, or mixed use. Residential and mixed use parcels are assigned their Units Number value as the count of dwelling units, while parcels classified as commercial/employment are ignored.

Dealing with Condominiums

The raw parcel data has inconsistent patterns in how units are counted in developments such as condominiums or master planned subdivisions. In some cases, the Units Number field actually represents the number of units present in the whole development rather than counting the units present on a single parcel polygon. For example, a master planned area outside of Phoenix might be subdivided into 164 plots, each housing a single family home. The raw data reports the Units Number as 164 for every polygon. If the count is applied directly to the parcel, the number of units would be drastically

overcounted. In these cases, parcels that are a part of each development are grouped, and the units are evenly assigned across all residential parcels in the group.

Imputing Missing Data

There are many cases nationwide where dwelling unit information was not directly supplied in the raw data. In these cases, we use a variety of methods to impute the dwelling unit count instead.

The first imputation process we use is to directly assign one dwelling unit to all parcels that are typed as a single-family land use. Land use codes that fall in this category can be seen in Table 5 below:

Table 5: Single Family Residential Land Use Codes for Imputation

Code	Land Use Name
102	TOWNHOUSE/ROWHOUSE
103	APARTMENT/HOTEL
106	APARTMENT
109	CABIN
112	CONDOMINIUM
115	DUPLEX
135	MOBILE HOME LOT
136	MOBILE HOME PARK
137	MOBILE HOME

Beyond the previous list, Corelogic has other land use codes (see Table 6 below) that are applied to both residential and non-residential uses, usually denoting undeveloped parcels. In these cases, the Improvement Value field for the parcel is used to determine if the parcel should actually be assigned a dwelling unit. A dwelling unit is assigned if the parcel surpasses a bare-minimum threshold for the value of a residential building³.

Table 6: Generic Residential Land Use Codes

Code	Land Use Name
100	RESIDENTIAL (NEC)
148	PUD

Next, a conservative estimate of 10 dwelling units is assigned to parcels that are missing unit data, but were assigned the 'Multifamily 10 Units Plus' land use field.

³ By default, the minimum improvement value threshold for identifying dwelling units by land use designation is \$5,000.

For parcels that are missing dwelling unit data but carry a multifamily land use, a conservative density of 12 DU/acre is applied. The land use codes that receive this imputation are in Table 7 below.

Table 7: Multifamily Land Use Codes for Imputation

Code	Land Use Name
106	APARTMENT
132	MULTI FAMILY 10 UNITS LESS
133	MULTI FAMILY DWELLING

The next step is to remove clear outliers, or cases where the resulting dwelling unit density is far beyond what could be considered reasonable. For parcels classified as a single-family land use (see Table 2), the following corrections are applied:

- Very small parcels assigned more than one detached single-family DU → reassigned one DU⁴
- Assigned detached single-family DU density far exceeding viable density → reassigned density of large lot single family building type⁵

For parcels classified as a multifamily land use (see Table 7):

- Assigned multifamily DU density far exceeding viable density → reassigned density of lowest intensity multifamily type in UF built form library⁶

Assigning Units from Census Data

Even after all of these transformations, there are still instances in the raw data that simply do not have enough information available to impute dwelling units. Data from the UrbanFootprint block level canvas is used in these cases to both identify and fill in gaps. Parcel level data is first aggregated to the block scale to allow for a direct comparison of dwelling unit totals against our block canvas totals, which come from the 2010 US Decennial Census. Blocks satisfy the following conditions are flagged as cases where raw parcel data should be substituted with census dwelling unit totals:

- Blocks that have at least 10 dwelling units in the 2010 Decennial Census
- Blocks where our aggregate parcel dwelling unit total is more than 30% under the 2010 Decennial Census block total

For parcels that match these cases, block-level totals are disaggregated down to parcels using land use codes to identify parcels that can accommodate residential data (residential or mixed use parcels). This process is detailed further in the [Disaggregate Block Data to Parcels](#) section.

⁴ Parcel area less than 0.15 acres (6,500 square feet)

⁵ Dwelling unit density greater than 50 DU/acre reassigned 5 DU/acre. This density-based fix is necessary for cases where there are multiple single-family houses included in the same parcel. This occasionally occurs in master planned developments.

⁶ Dwelling unit density greater than 1000/acre re-assigned 10 DU/acre. For reference, the dwelling unit density of the skyscraper residential building type in UrbanFootprint is 608 DU/acre.

Source-Specific Modifications

Beyond the process described above, there are cases across the country where the raw data does not match the definition found in attribute columns or where the data does not follow the pattern of the rest of the dataset. In these cases, modifications were applied to make the data work within the framework specified above. The data inconsistencies fit into four categories:

1. Unreliable Building Units Data
2. Overtyped Apartment Land Use Code
3. Literal Application of Attached Single-Family Types
4. Overtyped Single-Family Land Use

Data sources where these modifications are applied can be found in [Appendix B](#). Each modification is described in more detail below.

Unreliable Building Units Data

Some data sources use the building units field to include structures such as small sheds or storage areas. In other cases, the raw building units data is unreliable when compared to satellite imagery, overcounting the actual number of buildings on detached single family parcels. For counties where these patterns are identified, the building units field is ignored and instead dwelling units counts are typed solely by the land use code imputation process.

Overtyped Apartment Land Use Code

In other data sources, the corelogic apartment land use code is used to denote that the unit is rented, rather than providing information as to whether the unit is in a multifamily structure or not. To improve the accuracy of the UF built form type assigned to these parcels, they are reclassified as detached single-family land use before typing.

Literal Application of Attached Single-Family Types

Throughout most of the country, duplex and triplex units are parcelized in such a way that each unit has its own geometry. Where building unit data is missing, these parcels are imputed such that each geometry receives one dwelling unit. There are certain data sources where Duplex, Triplex, and Quadruplex land use designations instead are applied to a single parcel that contains more than one unit. For these data sources where building unit data is missing, dwelling units are imputed as follows using a literal application of the unit type (i.e. - Duplexes receive two units, Triplexes receive three units).

Overtyped Single-Family Land Use

The Single Family Residential (SFR) land use code (#163) is most commonly applied to parcels where there is a dwelling unit on the structure, while vacant residential land uses are denoted with a vacant land use (#465). That said, there are counties where the SFR type is liberally applied to any residential use. In these cases, the dwelling unit

imputation process for SFR is skipped, relying on the 'Building Units' field to assign dwelling units instead.

Building Area by Type

Building area fields in the UF parcel canvas follow a similar pattern as the logic used to populate dwelling units. Where the raw CoreLogic parcels contain information on the building area of a parcel, this data is incorporated directly into the parcel canvas. Where the data is missing, building area is imputed based on default square foot per unit values that are then multiplied by the total number of dwelling units and employees of each subtype.

Building area in the UF canvas is defined as the total living area, meaning the area of a building that would be heated or cooled. This typically excludes garages, unfinished basements, and patios. The CoreLogic field used to populate the building area field is **Living Square Feet**, which is defined below in Table 8 alongside other building area fields present in the raw data.

Table 8: CoreLogic Building Area Attributes

Variable Name	Description
UNIVERSAL BUILDING SQUARE FEET	The Building Square Footage that can most accurately be used for assessments or comparables (e.g., Living, Adjusted, Gross).
BUILDING SQUARE FEET IND	The codes appearing in this field indicates the source used to populate the UNIVERSAL BUILDING SQUARE FEET field (e.g., Living, Adjusted, Gross). Please see "BLDSF" table for code descriptions.
BUILDING SQUARE FEET	The size of the building in Square Feet. This field is most commonly populated as a cumulative total when a county does not differentiate between Living and Non-living areas.
LIVING SQUARE FEET	This is the area of a building that is used for general living. This is typically the area of a building that is heated or air conditioned and generally does not include Garage, Porch or Unfinished Basement Square Footage values.
GROUND FLOOR SQUARE FEET	Square footage of the part of the building which is level with the ground (typically the front of the building). This is generally above the basement(s) and below the second floor.
GROSS SQUARE FEET	This is the square footage for the entire building. Typically this represents all square feet under the roof.
ADJUSTED GROSS SQUARE FEET	This is the square footage used by the county or local taxing / assessment authority to determine Improvement Value. This figure is typically 100% of the living area, plus lower percentage of non-living area.
BASEMENT SQUARE FEET	This is total square footage associated with Basement portion of a building. This would include both finished and unfinished areas.
GARAGE/PARKING SQUARE FEET	This is the total square footage of the primary garage or parking area (i.e., commercial). This includes both finished and unfinished areas.

To match the format of the UF canvas, the living square feet data is split into the various dwelling unit and employment subtypes present on the parcel. The logic used to distribute the square footage is defined below in Table 9.

Table 9: Building Area Distribution Logic

Case	Assignment Logic
Dwelling Units > 0 AND Employment = 0	Assign Living Square Feet data to dwelling unit type present on the parcel
Employment > 0 AND Dwelling Units = 0	Proportionally distribute building area based on the number of employees in each subcategory
Dwelling Units > 0 AND Employment > 0	Living Square Feet first split into residential and employment uses based on dwelling unit vs. employee proportions. Then assigned using the methodology for each case described above.

For cases where Living Square Feet data is missing for the parcel, building area is imputed using default square foot per unit values that are then multiplied by dwelling unit by type and employment by type values, respectively. A more in-depth description of this process, as well as the default values used, can be found in the Building Square Footage section of the UrbanFootprint Base Canvas Documentation.

Population and Households

The population and households fields are derived by multiplying the number of dwelling units present on each parcel by census (ACS 2016 5 Year Estimates) averages for occupancy and population per household at the tract level. The population and household figures are then rounded to the nearest integer, with a control process in place to ensure that the county-wide population or household total is not changed through the rounding process.

Employment by Type

Employment by Type is calculated at the census block level to create the UF block level canvas. The block level data is then disaggregated down to the parcels using the process described in below.

Disaggregate Block Data to Parcels

When data is not available at the parcel level, census-block level data is disaggregated down to the parcels. Disaggregation is always area-dependent which means that all else equal, larger parcels will receive more data while smaller parcels will receive less. However, the disaggregation will take into account land use information when available to better allocate data. To build out the nationwide parcel canvas, each parcel is first assigned a UF land use code using the crosswalk described in the above section. UF canvas columns for dwelling units and employment by type are set as TRUE or FALSE for each UF land use code. TRUE indicates that the particular UF land use code is capable of taking on the data from the block for that column. For example, a parcel given a residential UF land use code will take on some residential units but will not take on employment from its parent block. Similarly, a parcel given a retail land use code will take on some retail employees but will not take on industrial employees.

By default, parcels that are classified as ‘vacant’, ‘open space’ or ‘other’ are excluded from disaggregation. That said, there are exceptions to the rule. When the block has data to disaggregate and **all** the parcels under the block are typed as one of ‘vacant’, ‘open space’ or ‘other’, the data gets assigned to only the ‘open space’ or ‘other’ parcels. If all the parcels are vacant and the block still asserts data, the disaggregation logic falls back to the default area-based allocation.

Data is also not disaggregated to parcels below 100 square feet.

Parcel Area by Type

Beyond the general area of the parcel, data on area utilized by different land uses is captured by the parcel area by type fields. The highest level of these uses are four mutually-exclusive categories, described in Table 10 below.

Table 10: Mutually Exclusive Parcel Area Categories

Parcel Area Category	Description
Residential	Parcels that have dwelling units and no employment
Employment	Parcels that have employment and no dwelling units
Mixed Use	Parcels that have both dwelling units and employment
No Use	Parcels that have neither dwelling units or employment

Beneath these four mutually exclusive categories are a series of subtype fields which are **not** mutually exclusive. Each field is assigned the total parcel area if the subtype that they track is present on the parcel. For example, if a parcel has both retail

employment and office employment, both the retail parcel area and office parcel area will be the same value, which is the area of the parcel. A full list of the parcel area by type fields can be found in the Base Canvas Documentation.

Built Form

UF includes a default library of more than 50 building types. These building types are used to populate the 'built form' column of the parcel canvas by attempting to represent the ground truth into these distinct building types.

Density-Based Classification

The first step of assigning built forms is to divide all the parcels and building types into distinct sets based on UF Land Use Codes. Within each set of building types for a specific UF Land Use Code, the parcel gets a building type that "most closely" matches the attributes of the parcel. "Closeness" is defined as the *least standardized absolute difference* between the building type and parcel data attributes. To do this, the parcel and building type attributes are first standardized by subtracting the mean and dividing by the standard deviation of the corresponding building types set. Then, each parcel's standardized attribute score is compared to the standardized score of the attribute in the corresponding set of building types. The differences are then squared and summed. The building type which corresponds to the least sum-of-squares is assigned to that parcel as its built form. If the dwelling unit density and employment density were the axes of a two-dimensional graph, this sum-of-squares would represent the distance between the parcel's data point and the building type data point. Therefore, the least sum-of-squares would represent the building type that is "closest" to the parcel's data point. Similarly, multiple attributes could be represented on an n-dimensional graph.

$$\text{"Distance" to built form} = \sum_{\text{attributes}} (Z_{\text{parcel}} - Z_{\text{building type}})^2$$

where Z = standardized attribute density

Data attributes currently default to dwelling unit density for residential based land use codes, employment density for employment based land use codes and use both for mixed land use codes.

Non-Density Based Classification

Relying purely on this density-based approach won't capture the nuances of some building types like parks, open water, schools, or cemeteries. Built forms such as these are considered *non-density based*. Most of these classifications come from the supplemental datasets (Refer the [supplemental datasets](#) section for more details). A few exceptions are detailed below.

For assigning water to a parcel, the UF block canvas is used. UF blocks (which are based off of census blocks) are assigned water if the census asserts that the block's

area is entirely covered by water⁷. Parcels that intersect with these blocks are assigned water if they don't belong to residential-based building types. This is to account for residential parcels on the edge of water bodies.

Institutional types, such as courthouses, libraries, or city halls, are another example of where the density based approach falls short. From our raw source of parcel data, these parcels are sporadically categorized as 'public', 'tax exempt', or 'state property' which are hard to parse into more specific building types. For simplicity's sake and if those parcels haven't been typed by more specific supplemental datasets, these parcels are categorized as open space (usually state or regional parks) if they are rural, non-urban civic if they are in developed spaces but not in an urban setting and urban civic if they are in an urban setting.

Intersection Density and Land Development Category

Intersection density and land development category are attributes of a parcel that exist only in the context of where the parcel sits in space. In most cases, roadway is excluded from parcel geometries meaning that the actual intersection density of any parcel on its own would be zero. To address this discrepancy, these attributes are applied directly from the UF Block canvas. Each parcel is assigned the intersection density and land development category of the block that shares the most area with it. Intersection density is derived from OpenStreetMap road network data.

The methodology for calculating both columns is described in greater detail in the UrbanFootprint Base Canvas Documentation.

Irrigated Area

Irrigated area values in the parcel base canvas are modelled based on the building type assigned to each parcel. This methodology is described in greater detail in the UrbanFootprint Base Canvas Documentation.

⁷ Census blocks where AWATER > 0 and ALAND = 0

Appendix

Appendix A: UrbanFootprint - Corelogic Crosswalk

See link [here](#).

Appendix B: Source-Specific Modifications

State	County	Unreliable Building Units	Overtyped Apts	Overtyped Single Family	Literal Attached SF
Alabama	Morgan County	X			
Alaska	Kenai Peninsula		X		
	Matanuska-Susitna Borough		X		
California	Madera County		X		
	Nevada County		X		
Florida	Charlotte County	X			
	Walton County		X		
Georgia	Bartow County		X	X	
	Floyd County	X			
Michigan	Livingston County		X		
Minnesota	Olmsted County		X	X	
	Otter Tail County		X		
	Sherburne County		X	X	
	Winona County		X		
	Wright County		X		
Mississippi	Pearl River County			X	
New Hampshire	Belknap County		X		
New York	Columbia County		X		
	Dutchess County		X		
	King County				X
	Putnam County		X		
	Sullivan County		X		
	Tioga County		X		
	Ulster County		X		

NATIONWIDE PARCEL CANVAS CREATION

North Carolina	Buncombe County		X		
	Caldwell County		X	X	
	Craven County	X			
	Davidson County		X	X	
	Iredell County		X	X	
	Pender County		X		
	Wayne County		X	X	
	Wilson County		X	X	
Oklahoma	Canadian County		X	X	
	Wagoner County		X	X	
Oregon	Umatilla County	X			
Texas	Angelina County	X			
	Bastrop County	X			
	Brazoria County	X			
	Grayson County	X			
	Hardin County	X			
	Hunt County	X			
Utah	Washington County	X			
Virginia	Campbell County		X	X	
	Fauquier County		X	X	
	Franklin County		X	X	
	Henry County		X	X	
	Rockingham County		X	X	
	Spotsylvania County		X	X	
Washington	Grant County	X			
	Skagit County		X	X	