Capital Cities Street Network Indices: Dataset Description

Dataset field descriptions

This dataset is composed by 176 rows, representing 176 capital or important cities from most countries of the world, and by 51 columns, that contain the values of street networks indices and socioeconomic information for each city.

The descriptions of the 51 columns are presented in the table 1.

Table 1. Description of dataset columns.

Column Name	Description
id	Unique identifyer of the urban area.
city	The name of the main city inside the urban area.
area	Total urban area in squared kilometers as reported in the GHS-UCDB[1].
geoarea	Geographical area to which the urban area belongs as reported in the GHS-UCDB. Encodings are reported in Table 2.
population	Total estimated population count in the urban area as reported by the GHS-UCDB.
climate_class	Köppen–Geiger climate class as reported by the GHS-UCDB. The climate class is reduced to a numerical code reported in Table 3.
builtup_area	Built-up area of the city as reported by the GHS-UCDB. The built-up area represents the total building area in the urban area.
income_group	Income group according to the classification of the World Bank, as reported by the GHS-UCDB.
orientation_entropy	This index represent the Shannon index of the street segment bearings (i.e., the angles of street segments) for the entire street network. It measures how ordered the network is based on its geometry and design. This index was calculated over the road network.

---

[1] Global Human Settlement - Urban Centres Database.

walk_prox_poi_mean	The walking proximity to points of interest (POI) measures the amount of different types of amenities that can be reached within a 15-minute walk from each node of the of pedestrian network.
walk_prox_poi_median	
walk_prox_poi_range	
walk_prox_poi_std	
walk_prox_poi_iqr	The city-scale mean, median, range, standard deviation, and interquartile range are reported.
walk_prox_pub_transport_mean	The walking proximity to public transport measures the amount of public transport stops or stations that can be reached within 400 meters from each node of the pedestrian network. 400 meters corresponds to the pedestrian shed and is regarded as a walkable distance for most people.
walk_prox_pub_transport_median	
walk_prox_pub_transport_range	
walk_prox_pub_transport_std	
walk_prox_pub_transport_iqr	The city-scale mean, median, range, standard deviation, and interquartile range are reported.
walk_circuity	Circuity measures the sinuosity of the street network as the ratio of network distance with respect to its straight-line length. This index is calculated over the pedestrian street network.
walk_street_len_mean	The street length or block length measures the length of street network segments. This index is calculated over the pedestrian street network. The city-scale mean, median, range, standard deviation, and interquartile range are reported.
walk_street_len_median	
walk_street_len_range	
walk_street_len_std	
walk_street_len_iqr	
walk_street_density	This index is calculated as the total length of the streets of the urban area divided by its area. It is calculated over the pedestrian street network.
walk_street_len_total	This index is the sum of all the lengths of the street segments of the urban area. It is calculated over the pedestrian street network.
walk_link_node_ratio	The link-node ratio is calculated as the number of edges divided by the number of nodes of an undirected graph. It is calculated over the pedestrian network.

walk_intersection_density	Intersections are defined as junctions that have at least two different exits. The intersection density is calculated as the total number of intersections of a network divided by its total urban area. This index is calculated over the pedestrian network.
walk_intersection_count	Intersections are defined as junctions that have at least two different exits. The intersection count is calculated as the total number of intersections of a network. This index is calculated over the pedestrian network.
bike_prox_poi_mean	The cycling proximity to points of interest (POI) measures the amount of different types of amenities that can be reached within a 15-minute bicycle ride from each node of the of cycling network. The city-scale mean, median, range, standard deviation, and interquartile range are reported.
bike_prox_poi_median	
bike_prox_poi_range	
bike_prox_poi_std	
bike_prox_poi_iqr	
bike_slope_mean	The slope, also called grade, gradient, incline or rise, is the measurement of the degree of elevation that a street segment has with respect to a horizontal plane. This index is calculated over the cycling network. The city-scale mean, median, range, standard deviation, and interquartile range are reported.
bike_slope_median	
bike_slope_range	
bike_slope_std	
bike_slope_iqr	
bike_circuity	Circuity measures the sinuosity of the street network as the ratio of network distance with respect to its straight-line length. This index is calculated over the cycling street network.
bike_street_len_mean	The street length or block length measures the length of street network segments. This index is calculated over the cycling street network. The city-scale mean, median, range, standard deviation, and interquartile range are reported.
bike_street_len_median	
bike_street_len_range	
bike_street_len_std	
bike_street_len_iqr	
bike_street_density	This index is calculated as the total length of the streets of the urban area divided by its

	area. It is calculated over the cycling street network.
bike_street_len_total	This index is the sum of all the lengths of the street segments of the urban area. It is calculated over the cycling street network.
bike_link_node_ratio	The link-node ratio is calculated as the number of edges divided by the number of nodes of an undirected graph. It is calculated over the cycling network.
bike_intersection_density	Intersections are defined as junctions that have at least two different exits. The intersection density is calculated as the total number of intersections of a network divided by its total urban area. This index is calculated over the cycling network.
bike_intersection_count	Intersections are defined as junctions that have at least two different exits. The intersection count is calculated as the total number of intersections of a network. This index is calculated over the cycling network.

Variable encodings

Climate Class

The climate class in this dataset is determined by the Köppen–Geiger classification. To simplify the classification, for each row, only the macro-class is maintained, ending up with a simplified 4-class classification. Although the Köppen–Geiger classificacion comprises more classes, table 2 reports the encodings of the climate classes of the values that appear in the dataset.

Table 2. Encodings for climate classes.

Köppen–Geiger climate class	Code	Group
Af Tropical, rainforest	1	
Am Tropical, monsoon	2	Tropical
Aw Tropical, savannah	3	
BWh Arid, desert, hot	4	
BWk Arid, desert, cold	5	
BSh Arid, steppe, hot	6	Arid
BSk Arid, steppe, cold	7	
Csa Temperate, dry summer, hot summer	8	
Csb Temperate, dry summer, warm summer	9	
Cwa Temperate, dry winter, hot summer	11	
Cwb Temperate, dry winter, warm summer	12	Temperate
Cfa Temperate, no dry season, hot summer	14	
Cfb Temperate, no dry season, warm summer	15	
Cfc Temperate, no dry season, cold summer	16	
Dwa Cold, dry winter, hot summer	21	
Dfa Cold, no dry season, hot summer	25	Cold
Dfb Cold, no dry season, warm summer	26	

Geographical Area

In the Global Human Settlement - Urban Centres Database (GHS-UCDB), all urban centres are are divided into nine geographical areas. Table 3 shows the code associated to each region and combines two regions that had low number of records in the unprocessed data. The region of Australia and New Zealand was merged with Oceania, and the region of Northern America was merged with Latin America and the Caribbean.

Table 3. Encodings for geographical areas.

Geographical Area	Code
Sub-Saharan Africa	SA
Oceania	OC
Australia and New Zealand	
Northern America	AM
Latin America and the Caribbean	
Northern Africa and Western Asia	NW
Europe	EU
Eastern and South-Eastern Asia	EA
Central and Southern Asia	CA