The OECD *Metropolitan Database*

Dataset access	http://stats.oecd.org/Index.aspx?DataSetCode=CITIES
Contact	RegionStat@oecd.org
Last dataset update	August, 2022

Database metadata and release notes
Updated: August, 2022

Version: August, 2022

TABLE OF CONTENTS

Introduction

The OECD, in cooperation with the EU, has developed a harmonised definition of functional urban areas (FUAs). Being composed of a city and its commuting zone, FUAs encompass the economic and functional extent of cities based on daily people's movements (OECD, 2012); (Dijkstra, Poelman, & Veneri, 2019). The definition of FUA aims at providing a functional/economic definition of cities and their area of influence, by maximising international comparability and overcoming the limitation of using purely administrative approaches. At the same time, the concept of FUA, unlike other approaches, ensures a minimum link to the government level of the city or metropolitan area.

FUAs are defined in several steps. First, a population grid makes it possible to define 'urban centres' independently from administrative or statistical boundaries. An urban centre is a pure grid-based concept, a cluster of contiguous cells of high density and with more than 50,000 inhabitants. Subsequently, this dense urban centre is adapted to the closest local units to define a city (or 'core'). Next, commuting flows are used to identify which of the surrounding, less densely populated local units were part of the city's labour market (commuting zone). Commuting zones are defined as all municipalities with at least 15% of their employed residents working in a certain city core. Municipalities surrounded by a single functional urban area are included and non-contiguous municipalities are dropped.

Local units used as building blocks to define FUAs are different across countries. In the case of European countries they are Local Administrative Units (LAU) according to the terminology adopted by Eurostat. In most cases local units are municipalities. In the case of the United States, census tracts are used for the method, and the final boundaries of the city and the commuting zones are adapted to the county boundaries.

This methodology makes it possible to compare functional urban areas of similar size across countries. A classification of functional urban areas into four types according to population size is proposed:

- Small urban areas, with a population below 100 000 people;
- Medium-sized urban areas, with a population between 100 000 and 250 000;
- Metropolitan areas, with a population between 250 000 and 1.5 million;
- Large metropolitan areas, with a population of 1.5 million or more.

The Metropolitan database provides socio-economic and environmental indicators of 691 OECD functional urban areas over 250,000 inhabitants in 36 OECD countries. Most of the indicators presented in the database are modelled based on aggregation of data at smaller geographic scale (e.g. population by age). Other indicators are modelled using geo-spatial data sources (e.g. air quality) or by downscaling/adjusting indicators available at slightly different geographic scale through the use of population grid (e.g. GDP). The database contains annual data from 2000 to the most recent available year. Additional information, including the geographic files, list of FUAs by country, and correspondence between local units and FUAs can be found in the following link:

http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm

The OECD Metropolitan Database provides a major source for the analysis on socio-economic trends of cities and functional urban areas regularly published in the OECD Regions and Cities at a Glance publication series.

The rest of the document is divided as follows: the next section presents the different datasets available in the database. The third section describes the methodology and the final section presents the data sources.

Datasets

The OECD Metropolitan database is composed by the following datasets:

- **Demography** – Provides indicators relating to demographic composition, area, and density;
- **Economy** – Provides indicators relating to economic activity such as GDP, productivity, and employment;
- **Labour –** Provides indicators relating to labour market conditions such as employment, unemployment, and labour force;
- **Social –** Provides indicators relating to the environment, income distribution and crime;
- **Digitalization** – Provides indicators relating to the broadband coverage;
- **Territorial organization** – Provides indicators relating to the territorial fragmentation and polycentricity of metropolitan areas.

There are four main methods used to calculate the different indicators presented in the database (detailed information presented in the methodology section):

- **Aggregation of local administrative data:** The process of collecting data at the scale of local units (usually communes or municipalities) and aggregating it up to the FUA level;
- **Geographic processing of geo-localised data:** The processing of geo-localised data to estimate the indicator at the FUA level, including processing raster files and using geographic files to calculate areas;
- **Downscaling from larger regions:** In this method, the indicators are created by using the values in TL3 regions as data inputs and deriving FUA estimates based on the distribution of population. The distribution of population is measured using grid level data;
- **Derived indicators:** Indicators that are derived from other indicators: such as density (population/area), growth (changes per year), or national shares.

The following tables present all indicators available in the database, the variables names, methodology used to construct them, and unit of measurement.

Demography, Area, and Population Density

Table 1 Demographic indicators

Category	Description	Variable Name	Method	Unit	Notes
Demographic Composition and Evolution	Population, all ages, administrative data	T_T	Aggregation of local administrative data.	Persons	1;2
	Population of the core area	T_T_CORE	Aggregation of local administrative data.	Persons	1;2
	Population of the hinterland area	T_T_HINTER	Aggregation of local administrative data.	Persons	1;2
	Growth/shrinking index of the total population (2001=100)	POP_TOT_GI	Derived	Index	
	Youth population group (0-14)	T_Y0_14	Aggregation of local administrative data	Persons	1;2
	Working age population group (15-64)	T_Y15_64	Aggregation of local administrative data	Persons	1;2
	Elderly population group (65+)	T_Y65_MAX	Aggregation of local administrative data	Persons	1;2
	Population of the metropolitan area as a share of the national population	T_T_SH_NAT	Derived	Ratio	3
	Youth dependency ratio (-15 over population 15-64)	YOU_DEP_RA	Derived	Ratio	
	Elderly dependency ratio (65+ over population 15-64)	ELD_DEP_RA	Derived	Ratio	
Area and Population Density	Metropolitan area total land area	SURF	Geographic Processing	Square Km.	4
	Metropolitan area core land area	SURF_CORE	Geographic Processing	Square Km.	4
	Metropolitan area hinterland land area	SURF_HINTER	Geographic Processing	Square Km.	4
	Core land area as a share of the total metropolitan land area	SURF_CORE_SH	Derived	Ratio	
	Hinterland land area as a share of the total metropolitan land area	SURF_HINTER_SH	Derived	Ratio	
	Urbanised area (Built-up area or land for urban use in a metropolitan area in km2)	URB_AREA	Geographic Processing	Square Km.	5
	Urbanised area per capita (m2 per capita)	URB_AREA_PC	Derived	Ratio	5;6
	Urbanised area of the hinterland area (km2)	URB_AREA_HINTER	Geographic Processing	Square Km.	5

	Urbanised area of the core area (km2)	URB_AREA_CORE	Geographic Processing	Square Km.	5
	Population density (inhabitants per km2)	POP_DEN	Derived	Persons per km^2	7
	Population density of the core area (inhabitants per km2)	POP_DEN_CORE	Derived	Persons per km^2	7
	Population density of the hinterland area (inhabitants per km2)	POP_DEN_HINTER	Derived	Persons per km^2	7

Notes

1) In cases where the data was not available at local unit level, indicators were computed using TL3 and TL2 regions. In cases where the boundaries of the TL3 or TL2 regions did not match with those of the FUAs, indicators were adjusted based on the GHS population grid.

2) For years with missing values, the data was interpolated using linear interpolation.

3) The national values for each country come from the OECD regional dataset https://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR

4) The geographic files can be found in the following link: http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm

5) Values of urbanised area were estimated using the Global Human Settlement (GHS) built-up grid at a 1km resolution. https://ghsl.jrc.ec.europa.eu/ghs_bu.php

6) This measure was constructed using the T_T variable.

7) Population density was estimated using the variable T_T.

Economy

Table 2 Economic variables

Category	Description	Variable Name	Method	Unit	Notes
Metropolitan area Economy	GDP	GDP_REAL_PPP	Downscaling, Aggregation of local administrative data	Millions USD, constant prices, constant PPP, base year 2015	1
	GDP per capita	GDP_PC_REAL_PPP	Derived	USD, constant prices, constant PPP, base year 2015	2
	GDP of the metropolitan area as a share of the national GDP	GDP_SH_NAT	Derived	Ratio	3
	Employment at place of work	EMP_IND_TOTAL	Aggregation of local administrative data, Downscaling	Persons	1
	Labour productivity	GDP_PW_REAL_PPP	Derived	GDP per worker in USD, constant prices, constant PPP, base year 2015	4

Notes

1) Estimates of GDP and employment at place of work at the metropolitan area level are derived from the values of TL3 regions. For Chile, Mexico, Australia, and Colombia the values are derived from TL2 regions. Regional values come from http://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM. The values for United States are compiled by aggregating county-level GDP/employment at place of work estimates.

2) The variable for population is T_T.

3) The national values for each country come from the OECD regional dataset http://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM

4) The variable is created by dividing the variable GDP_REAL_PPP by EMP_IND_TOTAL. For comparability across countries, and given the high number of part-time jobs in the US, productivity of US metropolitan areas is calculated using the employment at place of residence (variable EMP_Y15_MAX).

5) Estimates of employment at place of work are derived from the values of TL3 regions. For Chile, Mexico, Australia, France and Canada the values are derived from TL2 regions. Regional values come from http://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM.

Labour

Table 3: Labour variables

Category	Description	Variable Name	Method	Unit	Notes
Labour	Labour force (15-64 years old)	LF_Y15_64	Downscaling or aggregation	Persons	1
	Labour force (15 years old and over)	LF_Y15_MAX	Downscaling or aggregation	Persons	1
	Employment (15-64 years old)	EMP_Y15_64	Downscaling or aggregation	Persons	1
	Employment (15 years old and over)	EMP_Y15_MAX	Downscaling or aggregation	Persons	1
	Unemployment (15-64 years old)	UNEM_Y15_64	Downscaling or aggregation	Persons	1
	Unemployment (15 years old and over)	UNEM_Y15_MAX	Downscaling or aggregation	Persons	1
	Participation rate (labour force 15-64 over population 15-64)	PARTIC_RA_15_64	Derived	Ratio	
	Participation rate (labour force 15+ over population 15+)	PARTIC_RA_15_MAX	Derived	Ratio	
	Employment rate (employment 15-64 over population 15-64)	EMP_RA_15_64	Derived	Ratio	
	Employment rate (employment 15+ over population 15+)	EMP_RA_15_MAX	Derived	Ratio	
	Unemployment rate (unemployment 15-64 over labour force 15-64)	UNEM_RA_15_64	Derived	Ratio	
	Unemployment rate (unemployment 15+ over labour force 15+)	UNEM_RA_15_MAX	Derived	Ratio	
	Participation rate growth index (2001=100), 15-64 years old	PARTIC_RA_15_64_GR_2001	Derived	Index	
	Participation rate growth index (2007=100), 15-64 years old	PARTIC_RA_15_64_GR_2007	Derived	Index	

	Participation rate growth index (2001=100), 15 years old and over	PARTIC_RA_15_MAX_GR_2001	Derived	Index	
	Participation rate growth index (2007=100), 15 years old and over	PARTIC_RA_15_MAX_GR_2007	Derived	Index	
	Employment rate growth index (2001=100), 15-64 years old	EMP_RA_15_64_GR_2001	Derived	Index	
	Employment rate growth index (2007=100), 15-64 years old	EMP_RA_15_64_GR_2007	Derived	Index	
	Employment rate growth index (2001=100), 15 years old and over	EMP_RA_15_MAX_GR_2001	Derived	Index	
	Employment rate growth index (2007=100), 15 years old and over	EMP_RA_15_MAX_GR_2007	Derived	Index	
	Unemployment rate growth index (2007=100), 15-64 years old	UNEM_RA_15_64_GR_2007	Derived	Index	
	Unemployment rate growth index (2007=100), 15 years old and over	UNEM_RA_15_MAX_GR_2007	Derived	Index	

Notes

1) Estimates of labour indicators for metropolitan areas are derived from the values of TL3 regions. For Chile and Mexico the values are derived from TL2 regions. For the US, county-level data is aggregated to FUAs.

Regional values come from http://stats.oecd.org/Index.aspx?DataSetCode=REGION_LABOUR

Social

Table 4 Social variables

Category	Description	Variable Name	Method	Unit	Notes
Environment	Mean population exposure to PM2.5 air pollution	PWM_EX	Geographic Processing	µg/m^3	1
	Mean population exposure to PM2.5 air pollution in the city	PWM_EX_CORE	Geographic Processing	µg/m^3	1
	Share of population exposed to a level of PM2.5 above 5 µg/m^3	SPEX_MORE_THAN_5	Geographic Processing	Ratio	1
	Share of population in the city exposed to a level of PM2.5 above 5 µg/m^3	SPEX_MORE_THAN_5_CORE	Geographic Processing	Ratio	1
	Share of population exposed to a level of PM2.5 above 10 µg/m^3	SPEX_MORE_THAN_10	Geographic Processing	Ratio	1
	Share of population in the city exposed to a level of PM2.5 above 10 µg/m^3	SPEX_MORE_THAN_10_CORE	Geographic Processing	Ratio	1
	Share of population exposed to a level of PM2.5 above 15 µg/m^3	SPEX_MORE_THAN_15	Geographic Processing	Ratio	1
	Share of population in the city exposed to a level of PM2.5 above 15 µg/m^3	SPEX_MORE_THAN_15_CORE	Geographic Processing	Ratio	1
	Share of population exposed to a level of PM2.5 above 25 µg/m^3	SPEX_MORE_THAN_25	Geographic Processing	Ratio	1
	Share of population in the city exposed to a level of PM2.5 above 25 µg/m^3	SPEX_MORE_THAN_25_CORE	Geographic Processing	Ratio	1
	Share of population exposed to a level of PM2.5 above 35 µg/m^3	SPEX_MORE_THAN_35	Geographic Processing	Ratio	1
	Share of population in the city exposed to a level of PM2.5 above 35 µg/m^3	SPEX_MORE_THAN_35_CORE	Geographic Processing	Ratio	1
	Tree cover, as a share of the metropolitan core area	TREECOVER_SHARE_CORE	Geographic Processing	Ratio	1

	Total protected area, including all IUCN management categories	PA_AREA	Geographic Processing	km^2	
	Total coastal protected area (protected area located within 50km from the coast)	COAST_PA_AREA	Geographic Processing	km^2	
	Protected area, as a share of the metropolitan land area	PA_SHARE	Derived	Ratio	
	Coastal protected area, as a share of the metropolitan land area	COAST_PA_SHARE	Derived	Ratio	
	Annual Cooling Degree Days (Sum over a year of the differences between the daily mean outdoor air temperature and the threshold temperature (22°C) when the daily mean outdoor temperature is above 22°C)	CDD	Geographic Processing	CDD	
	Annual Heating Degree Days (Sum over a year of the differences between the threshold temperature (15.5°C) and the daily mean outdoor temperature when the daily mean outdoor temperature is below 15.5°C)	HDD	Geographic Processing	HDD	
	Change in Annual Cooling Degree Days from 1970-1979 to 2009-2018 (in degree days)	CDD_CHANGE	Derived	CDD	
	Change in Annual Heating Degree Days from 1970-1979 to 2009-2018 (in degree days)	HDD_CHANGE	Derived	HDD	
Income distribution	Disposable Income per equivalised household (in USD constant prices, constant PPP, base year 2010)	INCOME_DISP_HH_REAL_PPP	See methodology section		
	Gini (at disposable income, after taxes and transfers)	GINI			
	Poverty rate after taxes and transfers, poverty line 60%	PVT6A			
Safety	Annual reported homicides	HOMIC	Aggregation of local administrative data, or Geographic Processing, or Downscaling	Count	2
	Homicide rate (reported homicides per 100,000 inhabitants)	HOMIC_RA		Homicides per 100,000 inhabitants	2;3

Notes

1) Air pollution and land cover indicators are provided by the OECD Environment Directorate. See (Mackie A. I., 2016) for information on the methodology and data sources for air pollution indicators and (Haščič, 2018) for the land cover indicators.

2) For Canadian metropolitan areas homicide indicators are estimated from Census Metropolitan Area data. In cases where the CMA does not exactly coincide with the FUA, a correction is applied by adjusting the estimate based on the share of population living in the FUA using grid level data. Estimates of homicides data for metropolitan areas in Korea and Japan are derived from the values of TL3 regions. Regional values come from https://stats.oecd.org/index.aspx?queryid=67085.

3) The variable used for population is T_T.

Digitalization

Table 5 :Digitalization variables

Category	Description	Variable Name	Method	Unit	Notes
Digitalization	Share of households having a downloading speed of at least 30 Mbits/s	D_SPEED_30MB	Aggregation of local administrative data	Ratio	See methodology section
	Share of households having a downloading/uploading speed of at least 25/3 Mbits/s	DU_SPEED_25_3MB			
	Share of households having a downloading/uploading speed of at least 100/10 Mbits/s	DU_SPEED_100_10MB			
	Share of households having a downloading/uploading speed of at least 250/25 Mbits/s	DU_SPEED_250_25MB			
	Share of households having a downloading/uploading speed of at least 1000/100 Mbits/s	DU_SPEED_1000_100MB			
	Share of households having access to ADSL, SDSL,or HDSL	XDSL			
	Share of households having access to cable (CATV)	CABLE			
	Share of households having access to fiber (FTTH, FTTB)	FIBER			
	Share of households having access to Wi-MAX	WI_MAX			

Territorial organisation

Table 6 : Territorial organisation variables

Category	Description	Variable Name	Method	Unit	Notes
Territory	Polycentricity (1=yes, 0=no)	POLY	Geographic Processing	Boolean	1
	Local governments (count)	LOC_GOV		Count	1

	Territorial fragmentation (local governments per 100 000 inhabitants)	FRAGMENTATION	Derived		2
	Average population size of local governments	AVG_POP_SIZE_GOV		Persons	1

Notes

1) See methodology section for more details.

2) The territorial fragmentation is defined as the number of local governments per 100,000 inhabitants of the metropolitan area.

Methodology

This section presents three commonly used methods used to create the statistics for metropolitan areas: aggregation of data, geo-spatial processing, and downscaling indicators from regional administrative data.

Aggregation of data

The method of aggregation of data consists in obtaining data on the geographic units (i.e. local units) that make-up the functional urban areas to provide summary statistics of the variable of interest. The data required for this part are:

1) Correspondence between local units and FUAs;

2) The variables of interest for the local units (for example population);

To illustrate, suppose FUA *A1* contains three local units: *X, Y,* and *Z* with a population of 10, 15, and 20, respectively. Therefore, the total population of FUA *A1* would be 45.

The correspondence between local units and FUAs can be found at the following link:

http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm

The source for each variable are presented in the data sources section.

Digitalization

Digitalization indicators were compiled by aggregating local level data. The table below describes for each country the input data that was used to compile FUA-level indicators.

Country	Type of data	Minimum aggregation	Technologies	Coverage definition	Speeds
Belgium	Webscraping from web map	Sections	Fixed network	Households	Download >1, >30, >60, >100 Mbit/s
Denmark	Table	Municipalities	Fixed wireless, Fiber, Cable TV, xDSL	Households, Businesses, & Households	Download 2, 10, 30, 100, 500, 1000 ; Upload 2, 10, 30, 100 ; Download/Upload 2/0.5, 5/2, 10/1, 10/2, 30/5, 100/30

				businesses, summer houses	
Estonia	Tables	Settlements, counties	Cable, 3G, 4G		Download 1/1, 30/1, 100/1, 500/1, 30/30, 100/30, 100/100 , 500/<500, 500/500 MB/s
Finland	Table	Municipalities	Fibre (2015)	Households	Download 30, 100 Mbit/s
France	Table	Municipalities	DSL, Cable, Ftth	Buildings	Download >3, >8, > 30, >1000 MNit/s
Germany	Webscraping from web map	Municipalities	DSL, HSDPA, LTE, FTTH/B, CATV (Cable), WLAN/WiFi, Wimax, PLC	Households	Download >1, >2, >6, >16, >30, >50 Mbit/s
Hungary	Table	Settlements		Subscriptions	Download 30, 100 Mbit/s
Italy	Table, shapefiles	Municipalities	ADSL, VDSL, EVDSL, FTTC, FTTH	Households, People, served addresses.	
Luxembourg	Table	TL2		Households	
Mexico	Table	Municipalities	Cable coaxial, Optical Fiber, DSL, Satellite, Fixed wireless	Accesses	Not available
Netherlands	Webscraping from interactive map	Municipalities	FTTH	Homes	Not available
Norway	Table	Municipalities, counties	DSL, VDSK, HFC-D3, FTTH, SAT, LTE, Wi-fi	Households	Donwload/Upload ≥4/0,5, ≥10/0,8, ≥25/5, ≥30/5, ≥50/10, ≥100/10, ≥50/50, ≥100/100
Poland	Table	Municipalities, counties, towns	Not provided	Buildings, housing units	Building and housing units penetration with at least 30MB/s and 100 MB/s
Sweden	Table	Municipalities, counties	ADSL, VDSL, HFC, Ftth, Wireless, UMTS, LTE, fixed networks	Households, businesses	>2, >10, > 30, >1000 M (fixed networks); >2, >30 M (Wireless)
United Kingdom	Table	Fixed local and unitary authorities	Fibre	Premises.	>2, >5, >10 >30, >300 Mbit/s
United States	Table	Counties, States	ADSL, Cable, fiber, other, satellite, wireless	Population	Download/Upload 0.2/0.2; 4/1; 10/1; 25/3; 100/10; 250/25; 1000/100 Mbit/s

Geographic processing

Geo-spatial data processing refers to the use of data that have an explicit or implicit information on geographical location. Geo-spatial data processing is used in the context of this note to model indicators for geographic areas (i.e. metropolitan). For this purpose, it is necessary to have data that has geographic information associated with it (such as latitude and longitude) and a geographic file (e.g. a shapefile with boundaries). For example, the population variable T_T is estimated by using a population grid (i.e. raster file) and the FUA geographic file (i.e. Shapefile).

There are several GIS programs that can be used to provide summary statistics, including: ARCGIS, R, Python, QGIS, and PostGIS.

Territory, areas and population density

Territorial statistics are calculated using geographic techniques. The OECD uses PostgreSQL for these calculations.

Metropolitan area total land area (AREA): calculated using the ST_Area function that returns the area of the polygon corresponding to the metropolitan area.

Metropolitan area core land area (AREA_FUA_CORE): calculated using the ST_Area function that returns the area of the polygon corresponding to the core of the metropolitan area.

Metropolitan area hinterland land area (AREA_FUA_HINTER): calculated using the ST_Area function that returns the area of the polygon corresponding to the hinterland of the metropolitan area.

Urbanised area (Built-up area or land for urban use in a metropolitan area in km2): calculated from the GHS built up area grids (2000 and 2014) and the polygon corresponding to the metropolitan area using the ST_SummaryStats function.

Urbanised area of the core area (km²): calculated from the GHS built up area grids (2000 and 2014) and the polygon corresponding to the metropolitan area core using the ST_SummaryStats function.

Urbanised area of the hinterland area (km²): calculated from the GHS built up area grids (2000 and 2014) and the polygon corresponding to the metropolitan area hinterland using the ST_SummaryStats function.

Polycentricity: a polycentric metropolitan area is defined as a metropolitan area having several distinct urban cores. The function ST_NumGeometries was used to count the number of distinct polygons corresponding to urban cores.

Local governments (count): defined as the number of local governments within the metropolitan area.

Local governments in the core (count): defined as the number of local governments within the core of the metropolitan area.

Protected areas

The main data source used to compute indicators on protected areas is the World Database on Protected Areas.

The World Database on Protected Areas (UNEP-WCMC, 2017) is a worldwide record of marine and terrestrial protected areas. Launched by the International Union for Conservation of Nature (IUCN) and UN Environment, the geospatial database has been compiled and is updated monthly by the UN Environment World Conservation Monitoring Centre (UNEP-WCMC).

The database is made up of about 242,000 records of protected areas, split into two shape files. Each protected area is recorded either as a polygon, delimiting the boundaries of the area or as a point with a reported area providing information on the extent of the protected area. One shape file contains all the protected areas recorded as polygons and the other one is for protected areas recorded as points.

Non-geospatial information is also available for each record, giving more details on the protected areas. Among the 28 fields accessible through the attributes table, five are useful for the analysis described in this document:

- IUCN management categories (IUCN_CAT): the different categories of protected areas made by the International Union for Conservation of Nature (IUCN) correspond to the management objectives within the areas. Seven different categories can be distinguished , going from the most restrictive natural zone management to a zone with sustainable use of natural resources (Ia: Strict Nature Reserve, Ib: Wilderness Area, II: National Park, III: Natural Monument or Feature, IV: Habitat/Species Management Area, V: Protected

Landscape/Seascape, VI: Protected area with sustainable use of natural resources). This variable can also take the following values: not applicable, not assigned or not reported.
- Status (STATUS): refers to the administrative status of the protected areas: 'Designated', 'Inscribed', Adopted', 'Proposed' or 'Established'
- Status year (STATUS_YR): year corresponding to the entry into force of the current status of the protected area
- Designation (DESIG): corresponds to the subnational, national or international framework or agreement the protected area is part of
- Reported area (REP_AREA): Protected area extent (useful for protected areas recorded as points)

Following the methodology developed for country-level indicators (Mackie, Sentier, Haščič, & Linster, 2017), protected areas with 'not reported' or 'proposed' status, and UNESCO Man and Biosphere Reserves are excluded for the analysis as well as protected areas recorded as points without a reported area.

The shape file containing protected areas recoded as polygons was dissolved to avoid overlaps between protected areas and converted afterwards into a 300 meter-resolution raster file. The raster does not take into account IUCN management categories.

Indicators on protected areas and on coastal protected areas are computed from this raster file, following the steps below:

1. Protected area

- The metropolitan land area (RA) is calculated from the metropolitan areas' shape file.
- The protected area extent (PA) is calculated from the protected areas raster, the protected areas recorded as points shape file and the metropolitan boundaries' shape file. The first part of the protected area extent (PA1) is calculated as the sum of the reported areas of all the points located within the metropolitan area. The second part (PA2) is calculated as the protected zones extent within the metropolitan boundaries measured from the raster. The total protected area extent (PA) is thus calculated as PA1 + PA2.
- The share of protected area within the metropolitan area (%) is calculated as 100*PA/RA.

2. Coastal protected area

- A 50km-buffer is created around the coastlines.
- The metropolitan coastal area (CA) is calculated for each metropolitan area as the area of the intersection between the 50km-buffer and the metropolitan areas' shape file.
- The coastal protected area extent (CPA) is calculated from the protected areas raster, the protected areas recorded as points shape file, the 50km-buffer and the metropolitan boundaries' shape file. The first part of the coastal protected area extent (CPA1) is calculated as the sum of the reported areas of all the points located within the intersection between the buffer and the metropolitan area. The second part (CPA2) is calculated as the protected zones extent within the intersection between the buffer and the metropolitan area measured from the raster. The coastal protected area extent (CPA) is thus calculated as CPA1 + CPA2.
- The share of coastal protected area within the metropolitan area (%) is calculated as 100*CPA/CA.

Cooling degree days

The data used to compute Cooling Degree Days indicators at the functional urban area level comes from the historical global gridded degree-days database of Cooling Degree Days (CDD) and Heating Degree Days (HDD) (Mistry, 2019). The database includes three types of indicators corresponding to CDD, HDD, and CDD computed using wet-bulb

temperature (CDDwb). Each indicator is available at six different threshold temperatures: 18, 18.3, 22, 23, 24 and 25°C for CDD and CDDwb and 10, 15, 15.5, 16, 17 and 18°C for HDD. The database provides these three indicators both by year and by month over the period 1970-2018.

The dataset used to compute indicators at the functional urban area level is the Cooling Degree Days raster corresponding to a threshold temperature of 22 degrees Celsius. The 49 bands of the raster correspond to the annual Cooling Degree Days values from 1970 to 2018 included.

Indicators were computed using the geopandas and rasterstats python libraries and by intersecting the raster file with the shape file corresponding to the functional urban areas' boundaries. For each functional urban area, the average cell value is calculated. All cells having an intersection with the functional urban area are included in the mean value calculation. The cells with missing values are ignored.

Downscaling indicators from regional data

Social, economic and labour statistics at sub-national level (such as GDP and Labour) which are comparable across countries are generally available for administrative regions (TL2 and TL3 regions of the OECD Regional database). GDP and Labour indicators for metropolitan areas are currently modelled by adjusting the regional values officially available from the countries NSOs to the boundaries of metropolitan areas, based on the population living in each region.

Geographic Information System (GIS) techniques are increasingly adopted in the literature, especially in the field of environmental indicators and other issues that are particularly attached to the geography of the territory, rather than their functional or political organisation (Nordhaus, 2006; Doll, 2000).

The methodology is similar to the one applied by (Milego, 2006) and uses the socio-economic values (GDP, employment and unemployment) in TL3 regions as data inputs and combines it with the distribution of population. The distribution of population is obtained from grid level data. The suggested methodology is composed of three main steps:

1) Intersect the FUA boundaries with the TL3 boundaries by the use of GIS techniques;

2) Calculate the share of population living in the intersection of the TL3 boundary and the FUA;

3) Derive the variable of interest based on the share of population living in the area calculated in the previous step.

It has to be noted that this technique assumes that the variable of interest has the same spatial distribution as population. As a consequence, the results of the GDP and labour indicators modelled with above mentioned approach should be interpreted with caution and for illustrative purposes only.

Example 1. Downscaling indicators from regional data

The following example illustrates the steps in the process.

The Figure below shows the boundaries of the Austrian FUA of Linz highlighted in red. There are three TL3 regions (Mostviertel-Eisenwurzen, Linz-Wels, and Mühlviertel) that intersect the FUA.

Figure 1 FUA and TL3 map

The first step is to calculate the intersection area between each TL3 region and the FUA. The next step requires to calculate the share of population living in the intersection. For this, it is necessary to obtain the population of the TL3 region and the population of the intersection. This step is done using GIS-processing. In the example above, the results are:

Table 7 Share of population between FUA and TL3

FUA Name	TL3 Name	Share of TL3 population living in FUA
Linz	Mostviertel-Eisenwurzen	7.25%
Linz	Linz-Wels	75.30%
Linz	Mühlviertel	90.02%

The next step is to take the variable of interest, in this case GDP to have an estimate for the FUA.

Table 8 Allocated GDP

FUA Name	TL3 Name	Share of TL3 population living in FUA	Total TL3 2016 GDP (Millions USD, current prices)	FUA share of 2016 GDP (Millions USD, current prices)
Linz	Mostviertel-Eisenwurzen	7.25%	7 436	539
Linz	Linz-Wels	75.30%	28 865	21 735
Linz	Mühlviertel	90.02%	5 279	4 752

Estimated GDP (USD in current prices) of Linz	27 026

Methodology to calculate income levels and inequalities within metropolitan areas

The income is defined as the household disposable income per equivalized household income, which represents the amount of income available to households after payment of taxes, social security contributions paid by workers and private current transfers. This value is equivalized by dividing by the square of the number of people in the household.

To estimate both the mean income and the income inequalities for each metropolitan area, tax records or household income surveys at the local level (municipalities, counties, small areas…) were gathered from National Statistics or other public agencies (i.e. tax agencies, social security, etc.). The method described below is detailed also in (Boulant, Brezzi, & Veneri, 2016). Table 7 provides the source and type of income data available in each country at the scale of local unit, the number of local units considered (municipalities, counties, etc.) and the available years. These data have been used to model income levels in the OECD metropolitan areas and, for a subset of countries, income inequality (Gini coefficients).

Table 9. Income Data provided by each country

Country	Years	Type of local unit	Number of local units	Income Distribution	Number of metropolitan areas
Australia	2011-2015	Statistical Area 2	2310	No	6
Austria	2013-2015	Municipality	2376 (2014), 2122 (2015)	Yes, deciles	3
Belgium	2013-2015	Municipality	647	Yes, intervals	4
Canada	2015	Metropolitan area	34	Yes, quantiles	34
Chile	2015	Taxpayers	266968	Yes, microdata provided	3
Denmark	2014-2016	Municipality	99	Yes, deciles	1
Estonia	2015-2017	Municipality	268	No	1
Finland	2015-2016	Municipality	312	No	1
France	2012-2015	Municipality	32252 (2015), 32974 (2014), 32941 (2013), 32950 (2012)	Yes, deciles	15
Germany	2013	Municipality	14575	No	26
Hungary	2014-2016	Municipality	3154	No	1
Italy	2014-2016	Municipality	7978 (2016), 7999 (2015), 8048 (2014)	Yes, intervals	13
Mexico	2015	Municipality	2446	No	38
Norway	2014-2016	Municipality	476	Yes, intervals	1
Portugal	2015	Municipality	308	Yes, deciles	2
Sweden	2014-2016	Municipality	290	Yes, intervals	1
Netherlands	2014-2015	Municipality	4057 (2014), 4064 (2015)	-No	6
United States	2014, 2016	County	3142	Yes, intervals	93
United Kingdom	2015-2016	Middle layer super output area	7201	No	21

Method to estimate total household disposable income at the metropolitan level

Available information on household income at local unit scale generally refers to taxable income, gross income or other definitions that are not comparable across countries. In absence of detailed information on public transfers and taxes, which would be required to get the disposable income from the taxable income, we chose a simple method based on coefficients estimated based on the mean equivalized household disposable income available for larger subnational regions, via the OECD Regional Database, or for the whole country, via the OECD Income Distribution Database. For most countries, the same coefficients as the last analysis on income in metropolitan areas were kept, with the year of reference 2010 and coefficients for each TL2 region. For the other countries, coefficients were computed at the national level with more recent years of reference. For some countries, the number of households had to be estimated as well, based similarly on coefficients computed at the regional or national level.

Table 10. Coefficient references

Country	Year of reference	Level	Source	Households estimation
Australia	2010	Regional	OECD Regional Database	Estimated from official population
Austria	2010	Regional	OECD Regional Database	Estimated from official population (2013,2014), Official (2015)
Belgium	2010	Regional	OECD Regional Database	Official
Canada	-	-	Stats Canada	Official
Chile	2015	National	OECD Income Distribution Database	Estimated from the microdata
Denmark	2010	Regional	OECD Regional Database	Official
Estonia	2010	Regional	OECD Regional Database	Estimated from official population
Finland	2010	Regional	OECD Regional Database	Official
France	2013	National	OECD Income Distribution Database	Official
Germany	2013	National	OECD Income Distribution Database	Estimated from official population
Hungary	2010	Regional	OECD Regional Database	Estimated from official population
Italy	2010	Regional	OECD Regional Database	Estimated from official population
Mexico	2010	Regional	OECD Regional Database	Estimated from official population
Norway	2010	Regional	OECD Regional Database	Official
Portugal	2015	National	OECD Income Distribution Database	Official
Sweden	2010	Regional	OECD Regional Database	Estimated from official population
Netherlands	-	-	Stats Netherlands	Official
USA	2010	Regional	OECD Regional Database	Official
UK	2014	National	OECD Income Distribution Database	Estimated from official population

The method adopted to compute **total household disposable income** for metropolitan areas consists in the following steps:

- First identify the correspondence between the municipalities and the metropolitan areas, using the OECD metropolitan area shapefiles and official shapefiles of the local units borders at the year, for which the income data were published by the national statistics organizations.
- Second, regional or national household incomes from tax records are computed by aggregating values of all local units up to the TL2 region or the whole country. The resulting total income measures are compared with the regional or national values of household equivalized disposable income provided by the OECD Regional Database or the OECD Income Distribution Database.
- Third, a coefficient — computed as the ratio between the two income values at regional level — is applied to re-scale the income value of each local unit to a disposable income definition. The same is done with the

population data to estimate the number of households. The income value is also deflated and converted into Purchase Power Parties (PPP).

- Fourth, the total disposable income, the number of households and the population of each local unit are aggregated according to the metropolitan areas to eventually estimate the household disposable income per equivalized household for each metropolitan area.

The same method was applied to estimate the income levels for the city and the commuting zone of each metropolitan area.

Method to estimate the distribution of household disposable income within areas (income inequalities)

Estimates of the Gini coefficient – or other indicators of inequality – for household disposable income are not available directly for most metropolitan areas in OECD countries. For a subset of countries, some information is available on the distribution of income within local units. While the form of this information differ somewhat across countries (e.g. average or total income by quantiles of households), this allows producing "proxies" of income distribution within local units and metropolitan areas. Proxy measures were computed by first "simulating" the entire income distribution in each metropolitan area, based on the assumption that this income distribution has the lognormal functional form (Balintfy & Goodman, 1973; Lopez & Servén, 2006) and by fitting this model through the upper-bound of each population quintiles. This method has the advantage of avoiding the dependence of the Gini coefficient on the granularity of measurement (using income quintiles, for example, would result in a lower Gini coefficient than using income deciles taken from the same distribution) and on the number of local units included in a metropolitan area.

Coefficients for respectively quantiles, deciles or intervals had to be computed at the regional or national level considering the same OECD databases, as for the average disposable income.

The distribution of household disposable income in each metropolitan area is modelled as a mixture of various log-normal distributions. Once a sample of households, whose incomes fit the income quintiles, is generated, the Gini coefficient can be computed for each metropolitan area.

Income values were also generated at the national level by sampling only a certain amount of municipalities to limit the computation time. The number of municipalities to sample per country was chosen based on the average share of national population associated. This enabled to compute the Gini coefficients at the national level.

Table 11. Municipality sampling at the national level

Country	Number of municipalities sampled	Share of the national population
Austria	1000	47%
Belgium	300	85%
Denmark	98	100%
France	1500	28%
Italy	2000	25%
Norway	200	47%
Portugal	150	49%
Sweden	150	52%
USA	500	16%

The poverty rates for each metropolitan area were computed using the generated incomes. The generated incomes at the national level enabled to compute an estimate of the median. The poverty line was defined as half of this median.

The poverty rates were then computed for each metropolitan area based on this poverty line, by taking the share of people whose income is below this threshold.

Data sources and availability

Table 12 Overview of data sources

Indicator	Data sources
Geographic Files	The geographic files include the files for the local administrative units (LAU) and the functional urban areas. The LAU sources are detailed in the next table. The FUA geographic files are available in the following link: http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm
Social	The underlying PM2.5 concentration estimates are taken from the Global Burden of Disease (GBD) 2017 project. They are derived by integrating satellite observations. See (Mackie A. I., 2016). Crime data at the metropolitan area level was computed using municipality level homicides records. Data was retrieved from the Eurostat Functional Urban Areas Database for European countries. For Canada, crime data was available at the Census metropolitan Area level. Homicides estimates in Japan and Korea are based on TL3 level data taken from the OECD Regional Database. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb_llivcon&lang=en https://stats-3.oecd.org/Index.aspx?DataSetCode=REGION_SOCIAL

Demography

Population

Table 13 Sources for gridded population

Country	Years available	Source
All countries	2000, 2015	https://ghsl.jrc.ec.europa.eu/ghs_pop.php

Population by age

Table 14 Sources for population by age

Country	Source	Link	Years	Definition
AUS	ERP by SA2 (ASGS 2016), Age and Sex, 2001 Onwards, Australian Bureau of Statistics	https://stat.data.abs.gov.au/	2001-2020	Estimated Resident Population (ERP), on 30th of June.
AUT	Population status, Statistics Austria	https://data.statistik.gv.at/web/meta.jsp?dataset=OGD_bevstandjbab2002_BevStand_2021	2002-2021	Population at the beginning of the year
BEL	Population by place of residence, nationality (Belgian/non-Belgian), marital status, age and gender, Directorate-general Statistics - Statistics Belgium	https://bestat.statbel.fgov.be/bestat/crosstable.xhtml?datasource=65ee413b-3859-4c6f-a847-09b631766fa7	2000-2008	Population on the 1st of January
BEL	Population par lieu de résidence, nationalité, état civil, âge et sexe, Statbel (Statistics Belgium)	https://statbel.fgov.be/fr/open-data/population-par-lieu-de-residence-nationalite-etat-civil-age-et-sexe-6	2009-2021	Resident population on the 1st of January
CAN	Statistics Canada, Centre for Demography, customized data	Data was sent by Statistics Canada.	2001-2020	Annual population estimates on July 1
CHE	Population living in the economic domicile by municipality, sex and age, 2000, Permanent resident population by age, by canton, district and municipality, from 2010 to 2020, Federal Statistical Office	https://www.bfs.admin.ch/bfs/fr/home/statistiques/catalogues-banques-donnees/donnees.assetdetail.286146.html, https://www.bfs.admin.ch/bfs/fr/home/statistiques/catalogues-banques-donnees/tableaux.assetdetail.18344317.html	2001, 2011-2021	Permanent resident population on the 31st of December. Population counts are allocated to the following year for consistency across countries. Values for the period 2002-2010 are interpolated (linear interpolation).
CHL	Population projections, base 2017, estimates and projections 2002-2035, communes, INE	https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion	2002-2022	Population projections
COL	Projections and retro-projections of the municipal population for the period 1985-2017 and 2018-2035 based on the CNPV 2018, DANE	https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	2000-2022	Municipal population projections
CZE	Distribution of the population - basic age groups (time series) by municipality, Czech Statistical Office	https://vdb.czso.cz/vdbvo2/faces/en/index.jsf?page=vystup-objekt-parametry&pvo=DEM02&sp=A&pvokc=&katalog=33156&z=T (the dataset for all municipalities was sent by the Czech Statistical Office by email)	2011-2022	Population count on December 31st. Population counts are allocated to the following year for consistency across countries.

DEU	Population by 5-year age groups, small regions TL3, OECD Regional Database	https://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR	2000-2021	Population on 31st of December restated to 1st of January the following year by the OECD Secretariat.
DNK	Population 1. January by municipality, size of the city, age and sex (Table `BY2`), Population 1. January by municipality, sex, age and marital status (Table `BEF1A07` DISCONTINUED), Statistics Denmark	https://www.statbank.dk/statbank5a/SelectVarVal/Define.asp?MainTable=BY2&PLanguage=1&PXSId=0&wsid=cftree, https://www.statbank.dk/statbank5a/SelectVarVal/Define.asp?MainTable=BEF1A07&PLanguage=1&PXSId=0&wsid=cftree	2005-2021	Population on 1st of January
ESP	Population by sex, municipalities and age (five-year groups), INE	https://www.ine.es/dynt3/inebase/en/index.htm?padre=6232&capsel=6233	2003-2021	Population on the 1st of January
FIN	Population according to age (1-year) and sex by area, 1972-2020, Table `statfin_vaerak_pxt_11re.px`, Statistics Finland	Fetched from Statistics Finland API https://www.stat.fi/org/avoindata/pxweb_en.html	2000-2022	Population on the 31st of December. Population counts are allocated to the following year for consistency across countries.
FRA	Population Census, Population by sex and age (table POP1B), INSEE	2006: https://www.insee.fr/fr/statistiques/2120422?sommaire=2402741#consulter, 2007: https://www.insee.fr/fr/statistiques/2120360?sommaire=2402748#consulter, 2008: https://www.insee.fr/fr/statistiques/2119934?sommaire=2403563#consulter, 2009: https://www.insee.fr/fr/statistiques/2116765?sommaire=2403625#consulter, 2010: https://www.insee.fr/fr/statistiques/2053581?sommaire=2118618#consulter, 2011: https://www.insee.fr/fr/statistiques/2050369?sommaire=2404802#consulter, 2012: https://www.insee.fr/fr/statistiques/2046655?sommaire=2118088#consulter, 2013: https://www.insee.fr/fr/statistiques/2045005?sommaire=2117002#consulter, 2014: https://www.insee.fr/fr/statistiques/2863610?sommaire=2867849#consulter, 2015: https://www.insee.fr/fr/statistiques/3561090?sommaire=3561107#consulter, 2016: https://www.insee.fr/fr/statistiques/4171341?sommaire=4171351#consulter, 2017: https://www.insee.fr/fr/statistiques/4515539?sommaire=4516122#consulter, 2018: https://www.insee.fr/fr/statistiques/5395878?sommaire=5395927#consulter	2006-2018	Population on the 1st of January
GBR	Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland, Mid-2001 to mid-2020 detailed time series, Office for National Statistics	https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland	2001-2020	Mid-year population estimates
ITA	Reconstructed resident population - Years 2002-2019 (dataflow `164_164`), Resident population on 1 January (dataflow `22_289`), Istat	Fetched from Istat API https://www.istat.it/it/metodi-e-strumenti/web-service-sdmx	2002-2022	Resident population on 1st of January
KOR	Dataset `101_DT_1B040M5_Y`: Resident registration year-over-year population by city/gun/gu/sex/age (5 years old) Years 1993-2021, KOSIS	https://kosis.kr/statisticsList/statisticsListIndex.do?vwcd=MT_ZTITLE&menuId=M_01_01&outLink=Y&entrType=	2010-2021	Resident population
LUX	Population by 5-year age groups, small regions TL3, OECD Regional Database	https://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR	2000-2021	Population on 1st of January
MEX	2020, 2010, and 2000 Census of Population and Housing and Population and Housing Count 2005, INEGI	http://en.www.inegi.org.mx/programas/ccpv/2020/, http://en.www.inegi.org.mx/programas/ccpv/2010/, http://en.www.inegi.org.mx/programas/ccpv/2005/, http://en.www.inegi.org.mx/programas/ccpv/2000/	2000, 2005, 2010, 2020	Census of population. Values for intercensal years are interpolated (linear interpolation).
NOR	Table `07459`: Population, by region, age, contents, year and sex, Statistics Norway	Fetched from Statistics Norway API https://www.ssb.no/en/statbank/table/07459/	2000-2022	Population on the 1st of January
PRT	Table `0008273`: Resident population (No.) by Place of residence (NUTS - 2013), Sex and Age group; Annual - Statistics Portugal,	Fetched from Statistics Portugal API https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003182&contexto=bd&selTab=tab2,	2000-2020	Resident population

	Annual estimates of resident population. Table `0003182`: Resident population (No.) by Place of residence (NUTS - 2002), Sex and Age group; Annual - Statistics Portugal, Annual estimates of resident population	https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008273&contexto=bd&selTab=tab2		
SVN	Table `05C4003S`: Population by age and sex, municipalities, Slovenia, half-yearly, Statistical Office of the Republic of Slovenia	Fetched from the Statistical Office of the Republic of Slovenia API https://pxweb.stat.si/SiStatData/pxweb/en/Data/-/05C4003S.px	2008-2021	Population on 1st of January.
SWE	Population by region, marital status, age and sex. Year 1968 - 2020, Statistics Sweden	Fetched from Statistics Sweden API https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101A/BefolkningNy/#	2000-2022	Population on December 31st. Population counts are allocated to the following year for consistency across countries.
USA	Intercensal Estimates of the Resident Population by Age, Sex, U.S. Census Bureau, Population Division	https://www2.census.gov/programs-surveys/popest/datasets/2000-2010/intercensal/county/co-est00int-agesex-5yr.csv	2000-2010	Resident population estimate on the 1st of July, metadata [here](https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2000-2010/intercensal/county/co-est00int-agesex-5yr.pdf)
USA	Intercensal Estimates of the Resident Population by Age, Sex, U.S. Census Bureau, Population Division	https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/asrh/CC-EST2020-AGESEX-XX.csv	2010-2020	Population estimate on the 1st of July, metadata [here](https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010-2020/cc-est2020-agesex.pdf)
Other countries	Population by age groups in TL3 regions, OECD regional database	http://stats.oecd.org/Index.aspx?DataSetCode=REGION_DEMOGR	2000-2021	Regional population values are downscaled to FUAs.

Economy

Table 15 Sources for economic variables

Country	Source and link	Definition	Years Available
OECD (except CHL, MEX, AUS, COL, USA, CAN)	http://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM	Estimated based on GDP data at TL3 level from OECD Regional Database	2001-2020
CHL, MEX, AUS, COL, CAN	http://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM	Estimated based on GDP data at TL2 level from OECD Regional Database	2001-2020
USA	https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1	Estimated based on GDP data at the county level	2001-2019

Labour

Table 16 Sources for labour variables

Country	Source and link	Definition	Years available
OECD except (FRA, GRC, CHL, NLD, LUX, CAN, FIN, ISL, PRT, JPN, COL, POL, TUR)	https://stats.oecd.org/Index.aspx?DataSetCode=REGION_LABOUR	Estimated based on labour data at TL3 level from the OECD Regional Database	2000-2020

FRA, GRC, CHL, NLD, LUX, CAN, FIN, ISL, PRT, JPN, COL, POL, TUR, ROU, BGR, HRV, CYP, MLT	https://stats.oecd.org/Index.aspx?DataSetCode=REGION_LABOUR	Estimated based on labour data at TL2 level from the OECD Regional Database	2000-2020
USA	Local Area Unemployment Statistics, US Bureau of Labour Statistics, https://www.bls.gov/lau/#cntyaa	Annual averages of employment, unemployment, and labour force data by place of residence	2000-2020

Income

Table 17 Income data sources and original definition

Country	Source and link	Definition
Australia	Australian Bureau of Statistics, Estimates of Personal Income for Small Areas: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6524.0.55.0022011-2015?OpenDocument	Total income from all sources: Wages and salaries, Own incorporated business, Superannuation and annuities, Investments and Other income (excl. Government pensions & allowances)
Austria	Statistics Austria, Integrated Wage and Income Tax Statistics (sent by Statistics Austria)	Net income = total income including transfer payments – tax paid
Belgium	Statistics Belgium, Revenu total net imposable et impôts par déclaration et par commune : https://statbel.fgov.be/fr/themes/menages/revenus-fiscaux/plus	Taxable net total income (all net income after subtracting deductible expenses) - Total income tax (amount of state taxes, local taxes and agglomeration taxes)
Canada	Statistics Canada, sent by Statistics Canada	Household disposable income, defined as employment income + semi-employment income + investment income + transfers received – transfers paid.
Chile	Ministerio de Desarrollo Social, Gobierno de Chile, Encuesta CASEN : http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/basedatos.php	Current household income
Denmark	Statistics Denmark, Disposable family income by municipality; Avg. equivalised disposable Income in decile groups, by decile average, municipality and time: http://www.statistikbanken.dk/statbank5a/SelectTable/Omrade0.asp?SubjectCode=04&ShowNews=OFF&PLanguage=1	Disposable income per fiscal household excluding imputed rent
Estonia	Statistics Estonia , Number of recipients and Average monthly gross income per employee by regions and administrative units: http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=IM005&lang=1	Gross income – remuneration subject to social tax, paid to the employee or public servant; scholarship, allowance and pension paid in relation to the employment or service relationship; remuneration paid for the performance of work paid pursuant to a legal act or other legislation; remuneration paid to a person after the end of employment or service relationship (excl. benefit paid to the employee or public servant upon the termination of contract or upon removal from post) according to the Estonian Tax and Customs Board declaration of income and social tax, unemployment insurance premiums and contributions to mandatory funded pension
Finland	Statistics Finland , Numbers and income of dwelling population and household-dwelling units by Municipality, Year and Data : http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tjkt/statfin_tjkt_pxt_001.px/?rxid=97f42fff-900d-4ede-bb39-1dab702e3f82	Households' disposable money income includes monetary income items and benefits in kind connected to employment relationships. Money income does not include imputed income items, of which the main one is imputed rent. The formation of disposable money income can be described as follows: wages and salaries + entrepreneurial income + property income (without imputed rent) + current transfers received (without imputed rent) – current transfers paid.
France	INSEE, Dispositif Fichier Localisé social et fiscal, Distribution des niveaux de vie et composition du revenu disponible: https://www.insee.fr/fr/statistiques/3560118	Standard of living, defined as the household disposable income divided by the number of consumption units in the household. The consumption units are defined based on the OECD equivalence scale.
Germany	Statistische Ämte des Bundes und der Länder, Lohn und Einkommenssteuer : https://www.govdata.de/sr_RS_latin/web/guest/daten/-/details/stlae-service-73111-01-01-5	Net income = total amount of income – tax on wage and income
Hungary	Hungarian Ministry of Finance. Net personal income per municipality (sent by Hungarian Ministry of Finance)	Net personal income. The net personal income data is the income after tax per capita (for one year). The net income is equivalent the domestic income minus tax per population.
Italy	Ministry of Economy and Finance, Dichiarazioni fiscali: http://www1.finanze.gov.it/finanze3/pagina_dichiarazioni/dichiarazioni.php	Total taxable income from fiscal declarations
Mexico	CONEVAL (Consejo Nacional de Evaluacion de la Politica de Desarrollo Social), Ingreso corriente total per cápita (ICTPC) mensual promedio, por municipio (sent by CONEVAL)	Household total income is equal to monetary income and non-monetary income: work-related income (remuneration for subordinate work and independent work income), property rental income and transfers (including in kind transfers).
Netherlands	Statistics Netherlands, Huishoudensinkomen naar postcode: https://www.cbs.nl/nl-nl/maatwerk/2018/15/huishoudensinkomen-naar-postcode4-2014-2015	Disposable income is gross income minus current transfers paid as alimony of the ex-spouse(s), income insurance premium such as premiums paid for social/national/private insurance in related to unemployment/disability/old-age/next-of-kin, health insurance premiums.

Norway	Statistics Norway, Tax statistics for personal tax payers: https://www.ssb.no/en/statbank/table/05854?rxid=33b51e87-8ceb-4b4e-943b-98942afa6081 Income intervals, by sex. Number of residents 17 years and older: https://www.ssb.no/en/statbank/table/08411/?rxid=4f133c74-017f-4b49-bb77-a2867e8b2f12	Ordinary income after special deductions is the equivalent of net income. Ordinary income after special deductions is the basis for municipal income tax, county income tax and community tax. Special deductions are given due to age, disabilities or reduced ability to earn an income, unusual high expenses due to illness, and parents' deductions.
Portugal	Sent by the Portuguese Tax Authority	Gross income
Sweden	Statistics Sweden, Income and Tax Statistics: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__HE__HE0110__HE0110A/SamForvInk1/?rxid=660355cb-8963-4b2b-a355-c531348d6192	Income from employment and business. It also includes income from pensions, sick pay, and unemployment benefits.
UK	Office for National Statistics, Small area income estimates for middle layer super output areas, England & Wales: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/smallareaincomeestimatesformiddlelayersuperoutputareasenglandandwales	Weekly net household income in 2014 and annual net household income in 2016, which is the sum of the gross income of every member of the household plus any income from taxes/benefits such as Working Families Tax Credit
USA	United States Census Bureau, Aggregate household income in the past 12 months, 5-year estimates: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_B19025&prodType=table Shares of aggregate income by quintile, 5-year estimates: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_B19082&prodType=table	Total income is the sum of the amounts reported separately for wage or salary income; net selfemployment income; interest, dividends, or net rental or royalty income or income from estates and trusts; Social Security or Railroad Retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income.

Crime

Table 18 : Crime data sources and original definition

Country	Source and link	Definition	Years available
Canada	Statistics Canada: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510007101	Number of homicide victims	2010-2018
Chile	National Public Security Fund: http://www.fnsp.gob.cl/estadisticas-delictuales-por-comuna/	Police Cases for Crimes of Greater Social Connotation. Homicides.	2012 - 2016
Colombia	Colombia National Police: https://www.policia.gov.co/grupo-informaci%C3%B3n-criminalidad/estadistica-delictiva	Common homicides	2013 - 2018
European countries	Eurostat, living conditions, functional urban areas: https://ec.europa.eu/eurostat/web/cities/data/database	Number of murders and violent deaths	2009 - 2018
Japan	OECD Regional Questionnaire; information provided by the delegate of the Working Party of Territorial Indicators (WPTI)	Intentional homicides	2013 - 2017
	National Police Agency	Number of homicide cases	2018
Korea	OECD Regional Questionnaire; information provided by the delegate of the Working Party of Territorial Indicators (WPTI)	Intentional homicides	2013 - 2017
Mexico	Government of Mexico: https://www.gob.mx/sesnsp/acciones-y-programas/incidencia-delictiva-del-fuero-comun-nueva-metodologia?state=published	Intentional homicides and femicides	2015 - 2018
United States	National Archive of Criminal Justice Data: https://www.icpsr.umich.edu/icpsrweb/NACJD/series/57	Murder offenses and arrests counts	2014, 2016

Table 19 :Environment data sources

Indicators	Source and link	Years available
Cooling degree days	Mistry, Malcolm Noshir (2019): A high-resolution (0.25 degree) historical global gridded dataset of monthly and annual cooling and heating degree-days (1970-2018) based on GLDAS data. PANGAEA, https://doi.org/10.1594/PANGAEA.903123, Supplement to: Mistry, MN (2019): Historical global gridded degree-days: A high-spatial resolution database of CDD and HDD. Geoscience Data Journal, https://doi.org/10.1002/gdj3.83	1970 - 2018
Air pollution - PM2.5 particles concentration	Air quality and health: Exposure to PM2.5 fine particles - countries and regions", OECD Environment Statistics (database), using IHME GBD 2019 concentration estimates https://doi.org/10.1787/96171c76-en	2000 - 2019
Protected areas	IUCN and UNEP-WCMC (2017), The World Database on Protected Areas (WDPA), On-line, January 2017, Cambridge, UK: UNEP-WCMC, www.protectedplanet.net	2017
Tree cover	"Land resources: Land cover change in countries and regions", OECD Environment Statistics (database), https://doi.org/10.1787/3bce4397-en	2000 - 2018

Digitalization

Table 20 : Digitalization data sources

Country	Source and link	Years available
Belgium	Belgian Institute for Postal services and Telecommunications	March 2017
Denmark	https://tjekditnet.dk/file/2017baggrundsdatakommunerxlsx	2017
Estonia	https://www.netikaart.ee/tsaApp	2017 Q3
Finland	https://www.traficom.fi/fi/	2015, 2016
France	Observatoire – France très haut débit	2017 T2
Germany	https://www.bmvi.de/DE/Themen/Digitales/Breitbandausbau/Breitbandatlas-Karte/start.html	March 2018
Hungary	Sent by the Hungarian Ministry for Innovation and Technology	2017
Italy	https://maps.agcom.it/	October 2017

Luxembourg	https://ec.europa.eu/digital-single-market/en/connectivity	2017
Mexico	https://bit.ift.org.mx/BitWebApp/	2017
Netherlands	https://www.stratix.nl/glaskaart/	2018
United Kingdom	https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-update-spring-2018	January 2018
United States	https://opendata.fcc.gov/Wireline/Area-Table-Dec2016/xv2f-wqqz	January 2018

Bibliography

Balintfy, L., & Goodman, S. (1973). Socio-Economic Factors in Income Inequality: A Log-Normal Hypothesis. *Zeitschrift fiir Nationalokonomie, Vol. 33*, 389-402.

Boulant, J., Brezzi, M., & Veneri, P. (2016). *Income Levels and Inequality in Metropolitan Areas: A Comparative Approach in OECD Countries.* Paris: OECD Publishing.

Dijkstra, L., Poelman, H., & Veneri, P. (2019). The EU-OECD definition of a functional urban area. *OECD Regional Development Working Papers, No. 2019/11*. doi:https://doi.org/10.1787/d58cb34d-en

Doll, C. J. (2000). *Night-time imagery as a tool for global mapping of social-economic parameters and greenhouse gas emissions.*

European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network - CIESIN. (2015). *GHS population grid, derived from GPW4, multitemporal (1990, 2000, 2015).*

Haščič, I. a. (2018). "Land Cover Change and Conversions: Methodology and Results for OECD and G20 Countries". In *OECD Green Growth Papers,* (Vol. No. 2018/04). Paris: OECD Publishing. doi:https://doi.org/10.1787/72a9e331-en

Hijmans, R. J. (2019). *Introduction to the 'raster' package (version 2.9-5).*

Lopez, J., & Servén, L. (2006). A Normal Relationship? Poverty, Growth, and Inequality. *World Bank Policy Research Working Paper*.

Mackie, A. I. (2016). Population Exposure to Fine Particles: Methodology and Results for OECD and G20 Countries. *OECD Green Growth Papers, No. 2016/02*.

Mackie, A., Sentier, S., Haščič, I., & Linster, M. (2017). "Indicators on Terrestrial and Marine Protected Areas: Methodology and Results for OECD and G20 countries". In *OECD Environment Working Papers.* Paris: OECD Publishing. doi:https://doi.org/10.1787/e0796071-en

Maraut, S. D. (2008). The OECD REGPAT Database: A Presentation. *OECD Science, Technology and Industry Working Papers, No. 2008/02*.

Milego, R. a. (2006). *Espon Database.* Espon Publishing. Retrieved from Espo 2013 Database.

Mistry, M. N. (2019). A high-resolution (0.25 degree) historical global gridded dataset of monthly and annual cooling and heating degree-days (1970-2018) based on GLDAS data. *Geoscience Data Journal*. doi:https://doi.org/10.1002/gdj3.83

Nordhaus, W. Q. (2006). *The G-Econ Database on Gridded Output: methods and data*. Retrieved from gecon.yale.edu

OECD. (2012). *Redefining "urban". A new way to measure metropolitan areas.* Paris: OECD Publishing.

UNEP-WCMC, I. a. (2017). *The World Database on Protected Areas (WDPA).* Retrieved from www.protectedplanet.net