1. **Regions**

* USA

- Region 1: *Northeast*: Maine ; New Hampshire ; Vermont ; Massachusetts ; Rhode Islands ; Connecticut ; New York ; New Jersey ; Pennsylvania

- Region 2: *Midwest*: Ohio ; Indiana ; Illinois ; Michigan ; Wisconsin ; Minnesota ; Iowa ; Missouri ; North Dakota ; South Dakota ;Nebraska ; Kansas

- Region 3: *West*: Montana ; Idaho ; Wyoming ; Colorado ; New Mexico ; Arizona ; Utah ; Nevada ; Washington ; Oregon ; California ; Alaska ; Hawaii

- Region 4: *South*: Delaware ; Maryland ; District of Columbia ; Virginia ; West Virginia ; Noth Carolina ; South Carolina ; Georgia ; Florida ; Kentucky ; Tennessee ; Alabama ; Mississippi ; Arkansas ; Louisiana ; Oklahoma ; Texas

determined with: zipcode

sources : https://en.wikipedia.org/wiki/List\_of\_ZIP\_Code\_prefixes

<https://benfry.com/zipdecode/>

2019 https://www.census.gov/popclock/print.php?component=growth&image=//www.census.gov/popclock/share/images/growth\_1561939200.png

* Denmark

- Region 1: Hovedstaden

- Region 2: *Midtjylland*

- Region 3: *Nordjylland*

- Region 4: *Sjælland*

- Region 5: *Syddanmark*

baseline units: Region (NUTS 2)

source: 2020 https://www.statbank.dk/BY2

determined with: Declarative question

* France

- Region 1: *Île-de-France*

- Region 2: *Nord-Ouest*: Centre-Val de Loire ; Pays de la Loire ; Poitou-Charentes ; Bretagne ; Normandie

- Region 3: *Nord-Est*: Grand Est ; Hauts-de-France ; Bourgogne-Franche-Comté

- Region 4: *Sud-Ouest*: Aquitaine ; Limousin ; Languedoc-Roussillon ; Midi-Pyrénées

- Region 5: *Sud-Est*: Auvergne-Rhône-Alpes ; PACA

source: 2020 https://www.insee.fr/fr/statistiques/4277596?sommaire=4318291&q=population+par+r%C3%A9gion

determined with: zipcode; excludes Corsica and overseas territories

* Australia

- Region 1: (West Australia) *Northern Territory; Western Australia*

- Region 2: *Queensland*

- Region 3: (Broad New South Wales) *Australian Capital Territory; New South Wales*

- Region 4: *South Australia*

- Region 5: (Victoria-Tasmania) *Tasmania; Victoria; Rest*

baseline unit: States and Territories

source: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release

determined with: zipcode

* Brazil

- Region 1: *North*

- Region 2: *North-East*

- Region 3: *South-East*

- Region 4: *South*

- Region 5: *Central-West*

baseline unit: Regiões

determined with: TBD

source: https://ftp.ibge.gov.br/Estimativas\_de\_Populacao/Estimativas\_2020/estimativa\_dou\_2020.pdf

* Canada

- Region 1: (North-West) *British Columbia; Alberta; Yukon; Northwest Territories; Nunavut*

- Region 2: (Central) *Manitoba; Saskatchewan*

- Region 3: *Ontario*

- Region 4: *Quebec*

- Region 5: (East) *Nova Scotia; New Brunswick; Prince Edward Island; Newfoundland and Labrador*

baseline unit: Provinces and Territories

determined with: zip code (FSA first characters)

source: https://www.canadapost-postescanada.ca/cpc/en/support/articles/addressing-guidelines/postal-codes.page?

* China

- Region 1: *North China*

- Region 2: *Northeast China*

- Region 3: *East China*

- Region 4: *South Central China*

- Region 5: (West China) *Northwest China; Southwest China*

baseline unit: Greater administrative area (1956)

determined with: zip code

source: http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm

* Germany

- Region 1: (Northern*) Bremen, Hamburg, Lower Saxony, Mecklenburg-Western Pomerania, Schleswig-Holstein*

- Region 2: (Western*) North Rhine-Westphalia, Rhineland-Palatinate, Saarland*

- Region 3: (Central*) Hesse, Thuringia*

- Region 4: (Eastern) *Berlin, Brandenburg, Saxony, Saxony-Anhalt*

- Region 5: (Southern) *Baden-Württemberg, Bavaria*

baseline unit: Länd

determined with: zip code

source: https://www.destatis.de/EN/Themes/Society-Environment/Population/Current-Population/Tables/population-by-laender.html

* India

- Region 1: *Northern Zonal Council*

- Region 2: *Southern Zonal Council, Andaman and Nicobar Islands, Lakshadweep*

- Region 3: *Central Zonal Council*

- Region 4: *Eastern Zonal Council, North Eastern*

- Region 5: *Western Zonal Council*

baseline unit: Zonal Council

determined with: zip code

source: https://statisticstimes.com/demographics/india/indian-states-population.php

* Indonesia

- Region 1: (Western Java) *Banten; Jakarta; West Java*

- Region 2: (Eastern Java) *Central Java; Yogyakarta; East Java*

- Region 3: (Northern Islands) *West Kalimantan; Central Kalimantan; North Kalimantan; East Kalimantan; South Kalimantan; North Sulawesi; Gorontalo; Central Sulawesi; West Sulawesi; South Sulawesi; Southeast Sulawesi*

- Region 4: (Eastern Islands) *North Maluku; Maluku; Bali; West Nusa Tenggara; East Nusa Tenggara; West Papua; Papua*

- Region 5: (Sumatra) *Aceh; North Sumatra; West Sumatra; Riau; Riau Islands; Jambi; Bengkulu; South Sumatra; Bangka Belitung Islands; Lampung*

baseline unit: Provinces

determined with: zip code

source: https://web.archive.org/web/20171123162558/http://www.bps.go.id/linkTabelStatis/view/id/1267

* Italy

- Region 1: North-West

- Region 2: North-East

- Region 3: Centre

- Region 4: South

- Region 5: Islands

baseline unit: Gruppi di regioni (NUTS 1)

determined with: zip code

source: http://demo.istat.it/bilmens/index.php?anno=2019&lingua=ita

* Japan

- Region 1: (Kanto) *Chiba; Gunma; Ibaraki; Kanagawa; Saitama; Tochigi; Tōkyō*

- Region 2: (Kansai) *Hyōgo; Kyōto; Mie; Nara; Ōsaka; Shiga; Wakayama*

- Region 3: (North) *Akita; Aomori; Fukushima; Hokkaido; Iwate; Miyagi; Yamagata*

- Region 4: (Chubu) *Aichi; Fukui; Gifu; Ishikawa; Nagano; Niigata; Shizuoka; Toyama; Yamanashi*

- Region 5: (South) *Ehime; Fukuoka; Hiroshima; Kagawa; Kagoshima; Kōchi; Kumamoto; Miyazaki; Nagasaki; Ōita; Okayama; Okinawa; Saga; Shimane; Tokushima; Tottori; Yamaguchi*

baseline unit: Prectures

determined with: zip code

source: https://www.citypopulation.de/en/japan/

* Mexico

- Region 1: (Central-Western) *Aguascalientes; Colima; Jalisco; Guanajuato; Michoacan; Nayarit; San Luis Potosi; Zacatecas*

- Region 2: (Central-Eastern) *Federal District; Hidalgo; Mexico; Morelos; Puebla; Queretaro; Tlaxcala,*

- Region 3: (North-East) *Coahuila; Nuevo Leon; Tamaulipas*

- Region 4: (North-West) *Baja California; Baja California Sur; Chihuahua; Durango; Sinaloa; Sonora*

- Region 5: (South) *Campeche; Chiapas; Guerrero; Oaxaca; Quintana Roo; Tabasco; Varacruz; Yucatan*

baseline unit: States

determined with: zip code

source: https://www.inegi.org.mx/app/scitel/Default?ev=9

* Poland

- Region 1: (North) Pomeranian, West Pomeranian, Kuyavian-Pomeranian, Warman-Masurian, Podlaskie

- Region 2: (Central) Lubusz, Greater Poland

- Region 3: (South-West) Lower Silesian, Opole, Silesia

- Region 4: (Central-East) Lesser Poland, Subcarpathian

- Region 5: (South-East) Lodz, Masovian, Lubin, Holy Cross

baseline unit: Voivodeships

determined with: zip code (first number)

source: https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-stan-i-struktura-ludnosci-oraz-ruch-naturalny-w-przekroju-terytorialnym-stan-w-dniu-30-06-2020,6,28.html

* South Africa

- Region 1: *Gauteng*

- Region 2: (West) *Western Cape; Northern Cape*

- Region 3: (Center) *North West; Free State*

- Region 4: (North-East) *Limpopo; Mpumalanga*

- Region 5: (South-East) *KwaZulu-Natal; Eastern Cape*

baseline unit: Provinces

determined with: zip code

source: http://www.statssa.gov.za/census/census\_2011/census\_products/Census\_2011\_Census\_in\_brief.pdf

* South Korea

- Region 1: *Seoul*

- Region 2: (North) *Gyeonggi; Incheon; Gangwon*

- Region 3: (Central-West) *North Chungcheong; South Chungcheong; Daejeon; Sejong*

- Region 4: (South-West) *North Jeolla; South Jeolla; Gwanggju; Jeju*

- Region 5: (East) *South Gyeongsang; North Gyeongsang; Daegu; Busan; Ulsan*

baseline unit: Provinces

determined with: zip code

source: https://kosis.kr/eng/statisticsList/statisticsListIndex.do?menuId=M\_01\_01&vwcd=MT\_ETITLE&parmTabId=M\_01\_01&statId=1962001&themaId=#SelectStatsBoxDiv

* Spain

- Region 1: (East) *Cataluña; Comunidad Valenciana; Islas Baleares*

- Region 2: (Center) *Castilla-La Mancha; Comunidad de Madrid*

- Region 3: (South) *Andalucía; Canarias; Ceuta (Ciudad Autónoma); Extremadura; Melilla (Ciudad Autónoma); Región de Murcia*

- Region 4: (North) *Aragón; Cantabria; La Rioja; Navarra; País Vasco*

- Region 5: (North-West) *Castilla y León; Galicia; Principado de Asturias*

baseline unit: Autonomous communities and Cities

determined with: zip code

source: https://www.ine.es/jaxiT3/Tabla.htm?t=2915

* Turkey

- Region 1: (Anatolia) *Central Anatolia; Southeastern Anatolia; Eastern Anatolia*

- Region 2: *Marmara*

- Region 3: *Mediterranean*

- Region 4: *Aegean*

- Region 5: *Black Sea*

baseline unit: Regions

determined with: zip codes

source: https://www.icisleri.gov.tr/turkiyenin-nufus-haritasi

* Ukraine

- Region 1: (Center-East) *Chernihiv; Poltava; Sumy*

- Region 2: (Center-West) *Cherkasy; Kirovohrad; Kyiv; Vinnytsya; Zhytomyr*

- Region 3: (East) *Donetsk; Kharkiv; Luhansk*

- Region 4: (South) *Dnipropetrovsk; Kherson; Mykolayiv; Odesa; Zaporizhzhya*

- Region 5: (West) *Chernivtsi; Ivano-Frankivsk; Khmelnytski; Lviv; Rivne; Ternopil; Volyn; Zakarpattya*

baseline unit: Oblast

determined with: TBD

source: https://ukrstat.org/en/druk/publicat/kat\_e/publnasel\_e.htm

* UK

- Region 1: (London)

- Region 2: (Southern England) *South West; South East; East of England*

- Region 3: (Central UK) *West Midlands; East Midlands; Wales*

- Region 4: (Northern England) *North West; Yorkshire and The Humber; North East*

- Region 5: (Northern UK) *Scotland; Northern Ireland*

baseline unit: Regions of the UK (NUTS 1)

determined with: zip code

source: https://www.doogal.co.uk/postcodedownloads.php

1. **Urban-Rural Categories**

* USA

- Urban category 1: *Urban*

- Urban category 2: *Rural*

Determined by : zipcode

description: Urban = RUCA code 1 (core metropolitan)

sources : [https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes](https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/)

data/zipcodes/US\_zipcode.xlsx from https://www.arcgis.com/home/item.html?id=8d2012a2016e484dafaac0451f9aea24

* Denmark

- Urban category 1: Live in town with less than 20,000 inhabitants

- Urban category 2: *Live in town with more than 20,000 inhabitants*

source: 2020 https://www.statbank.dk/BEV22

determined with: Declarative question

* France

- Urban category 1: *Grand-Pôle*

- Urban category 2: *Couronne de Grand-Pôle*

- Urban category 3: *Nord-Est*

description: Match postal codes with statistical identifier and retrieve urban category associated with the latter.

source 1: https://public.opendatasoft.com/explore/dataset/correspondance-code-insee-code-postal/table/

source 2: https://www.insee.fr/fr/information/2115011

source 3: https://www.insee.fr/fr/information/2115011

determined with: zipcode, cf. code/FR\_commune.R

* Australia

- Urban category 1: (Rural) *Inner Regional Australia; Outer Regional Australia; Remote Australia; Very Remote Australia*

- Urban category 2: (Urban) *Major Cities of Australia*

description: From source 1, we obtain the Remoteness Area of each mesh blocks, from source 2 we obtain the Postal Area, and from source 3 the population. We then merge the three sources with the mesh blocks information. We sum the population of each mesh block contained in a Postal Area. If a Postal Area is linked to several urban categories or several region, we attribute the option tied to the largest share of population. Finally, we add additional postal codes from source 4. Those postal codes already have a region and a urban category.

determined with: zip code

source 1: Remoteness Area (RA)ASGS Ed 2016 in .csv format) https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.005July%202016?OpenDocument

source 2: POAs (Postal Areas ASGS Edition 2016 in .csv Format ): https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.003July%202016?OpenDocument

source 3: census: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2074.02016?OpenDocument

source 4: https://github.com/matthewproctor/australianpostcodes

* Canada

- Urban category 1: (Rural) Forward Sortation Area second character is 0

- Urban category 2: (Urban) Forward Sortation Area second character is different from 0

description: From the source, we compute share of population based on second character of the Forward Sortation Area (if 0 it corresponds to a rural area according to Canada Post)

determined: zip code (3 first characters, i.e. FSA)

source: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&PR=0

* China

- Rural : Live in Xiāng, or in Jiēdào, or a Zhèn of less than 50,000 inhabitants

- Urban: Live in a Zhèn of more than 50,000 inhabitants

description: We perform web-scraping on the data from the source and compute the share of population using the Status of the municipalities.

determined with: two declarative questions: *i)* In which kind of municipality do you live in? *Xiāng*; *Jiēdào*; *Zhèn*, *ii*) What is the size of the agglomeration you live in? *Less than 50,000 inhabitants; More than 50,000 inhabitants*. (TODO: check with translators if those categories make sense and if we're not missing anything else)

source: http://www.citypopulation.de/en/china/townships/

* Germany

- Urban category 1: *Cities*

- Urban category 2: *Towns and Suburbs*

- Urban category 3: *Rural areas*

description: From source 1, we obtain the list of zipcodes with the area type (defined according to https://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP\_DEGURBA) as well as the population of each city. However, source 1 does not list all the zipcodes for each city. Source 2 contains the additional zipcodes. We first merge the existing zipcodes in source 1 with the data from source 2. We then uniquely identify a city with the city's name and Vorwahl, and attribute to a zipcode the most rural category (thus by order of priority Rual area; Towns and Suburbs; Cities) of its city. We finally remove zipcodes duplicates by keeping the most rural category for each zipcode.

determined with: zip code

source 1: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugJ/31122019\_Auszug\_GV.html

source 2: https://gist.github.com/jbspeakr/4565964#file-german-zip-codes-csv

* India

- Urban category 1: *Less than 5,000 inhabitants*

- Urban category 2: *5,000–20,000 inhabitants*

- Urban category 3: *20,000-50,000 inhabitants*

- Urban category 4: *50,000-250,000 inhabitants*

- Urban category 5: *250,000-3 millions inhabitants*

- Urban category 6: *More than 3 millions inhabitants*

description: We use data from the source and remove duplicates by name.

determined with: Declarative question: What is the size of the agglomeration you live in? *Less than 5,000 inhabitants; 5,000–20,000 inhabitants; 20,000-50,000 inhabitants; 50,000–250,000 inhabitants; 250,000–3 millions inhabitants; More than 3 millions inhabitants*

source: https://censusindia.gov.in/pca/pcadata/DDW\_PCA0000\_2011\_wardlevel.rar

* Indonesia

- Urban category 1: (Rural) *In a Kabupaten outside of the Capital town*

- Urban category 2: (Urban) *Kota; Capital town of a Kabupaten*

determined with: declarative question: In which of the following areas are you currently living? *Kota; Capital town of a Kabupaten; In a Kabupaten outside of the Capital town* (TODO: check with translator if those catagories make sense)

source: https://unstats.un.org/unsd/demographic-social/meetings/2019/newyork-egm-statmeth/docs/s08-01-IDN.pptx

* Italy

- Urban category 1: *Cities*

- Urban category 2: *Towns and Suburbs*

- Urban category 3: *Rural areas*

description: From source 1, we obtain the population for each municipality, its area type (defined according to https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-008), as well as a statistical identifier. From source 2 we obtain the zip codes and statistical identifier. For some zip codes we obtain different area types, we address this issue by determining the percentage of population living in each area types for a given zip code and attribute the area type with the greatest share of population. There also are a few zipcodes for which we do not have an area type, we assign them to rural if their population is under the biggest rural area (23,000 inhabitants) and to towns and suburbs otherwise.

determined with: zip code

source 1: https://www.istat.it/storage/codici-unita-amministrative/Classificazioni-statistiche-Anni\_2017-2021.zip

source 2: http://www.comuni-italiani.it/

* Japan

- Urban category 1: *Living in a town of less than 100,000 inhabitants.*

- Urban category 2: *Living in a town of more than 100,000 inhabitants.*

description: From source 1 we obtain postal information on municipalities and town in Japan, and obtain population from source 2. We then merge the two datasets using the locality name and prefecture (cleaning for some discrepancies). In a second step, we regroup municipalities by zipcodes prefix (the first five numbers of the zipcodes). If the prefix is linked to both towns with less than 100,000 inhabitants and with more than 100,000 habitants, we link it the latter category.

determined with: zip code

source 1: https://www.post.japanpost.jp/zipcode/dl/roman/ken\_all\_rome.zip?210622

source 2: https://www.e-stat.go.jp/en/regional-statistics/ssdsview/municipality

* Mexico

- Urban category 1: *Rural*

- Urban category 2: *Semiurbano*

- Urban category 3: *Urbano*

description: From source 1 we obtain the postal code and the different categories of locality (among *Rural*; *Semiurbano*; *Urbano*) associated with it. For postal codes associated with several categories, we assign the most urban category.

determined with: zip code. From source 2 (slide 19), we retrieve the population distribution by categories (where we apply the definition from CS02a.2-1 in source 3.

source 1: https://www.correosdemexico.gob.mx/SSLServicios/ConsultaCP/CodigoPostal\_Exportar.aspx

source 2: https://www.issea.gob.mx/Docs/Censo%20INEGI%202021/Censo2020\_Principales\_resultados\_EUM.pdf

source 3: https://www.inee.edu.mx/wp-content/uploads/2019/03/CS02-2011.pdf

* Poland

- Urban category 1: *Living in a town of less than 20,000 inhabitants.*

- Urban category 2: *Living in a town of more than 20,000 inhabitants.*

description: From the official statistics we calculate the distribution of population by towns' sizes. (NB: we were not able to link it with a zip code as zip code and municipalities each overlap several elements of the other category for an important share of the population).

determined with: declarative question: What is the size of the agglomeration you currently live in? *Less than 20,000 inhabitants; More than 20,000 inhabitants.*

source: https://stat.gov.pl/en/topics/population/population/population-size-and-structure-and-vital-statistics-in-poland-by-territorial-division-as-of-december-31-2020,3,29.html

* South Africa

- Urban: *Live in a metropolitan municipality or in a capital of a District municipality*

- Rural: *Live in a District municipality other than the District capital.*

description: We compute the share of the population living in a metropolitan municipality or in the capital of a district municipality (as a proxy for living in a big urban area).

determined with: Declarative question: In which type of municipality do you currently live in? *A metropolitan municipality; In a capital of a District municipality; In a District municipality other than the District capital.*

source: http://interactive2.statssa.gov.za/webapi/

* South Korea

- Urban category 1: *Live in a District (i.e., “Gum”).*

- Urban category 2: *Live in a Town (i.e., “Si”) .*

- Urban category 3: *Live in a City (i.e., “Gu”).*

description: From source 1, we collect the share of population living in each Municipal administrative level. From source 2, we collect the Municipal administrative level assigned to each zip code.

determined with: zip code*.*

source 1: https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx\_cd=1200 [file "2020년 도시지역 인구현황(시군구별).xls"]

source 2: https://www.epost.go.kr/search/zipcode/areacdAddressDown.jsp [file "시스템에 사용할 경우 지역별 주소 DB 사용을 권장 (2021.09.03)"]

* Spain

- Urban category 1: *Living in a town of less than 20,000 inhabitants.*

- Urban category 2: *Living in a town of more than 20,000 inhabitants.*

description: From the official statistics we calculate the distribution of population by towns' sizes.

determined with: Declarative question: What is the size of the agglomeration you currently live in? *Less than 20,000 inhabitants; More than 20,000 inhabitants.*

source: https://www.ine.es/dynt3/inebase/en/index.htm?padre=517&capsel=525

* Turkey

- Urban category 1: *Living in a district with a share of rural population greater than the national average for districts.*

- Urban category 2: *Living in a district with a share of rural population smaller than the national average for districts.*

determined with: zip code

description: From source 1 (which is official data but with a better formatting for our matching) we obtain the District associated to each zip code. From source 2, we obtain the share of rural population and the total population (defined by national statistics) of each district. We then assign districts with a share of rural population greater than the national average (around 22%) as rural. We finally merge the two datasets.

source 1: https://github.com/muratgozel/turkey-neighbourhoods

source 2: https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr (ADNKS-GK137474-O29001 İbbs-Düzey1, İbbs-Düzey2, İl Ve İlçe Nüfusları / 2020 / İlçe Düzeyi / HEPSI)

* UK

- Urban category 1: (Rural and other) *Rural village; Rural hamlet and isolated dwellings; Rural town and fringe; Rural town and fringe in a sparse setting; Rural hamlet and isolated dwellings in a sparse setting; Rural village in a sparse setting; Accessible rural area; Remote rural area; Very remote rural area; Very remote small town; Accessible small town; Remote small town*

- Urban category 2: (City or Town) *Urban city and town; Urban city and town in a sparse setting*

- Urban category 3: (Large urban areas) *Urban major conurbation; Urban minor conurbation; Large urban area; Other urban area*

description: From the source we obtain the area type (defined according to the following categories https://www.ons.gov.uk/file?uri=/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification/rucladleafletmay2015tcm77406355.pdf ) for each zip code. We then compute the area type for each outcode (i.e. the first characters of a UK zip code) by attributing the area type with the greatest share of population within the outcode area. For outcode with no population in the data we assign them to the rural category, except for three outcodes that do not include any zip code of a rural category. For the latter, we assign them the City or Town category.

determined with: zip code / outcode

source: https://www.doogal.co.uk/postcodedownloads.php