Synthetic Populations for Regions of the World

Version 0.9

Data Description

Organization:

NSSAC, Biocomplexity Institute and Initiative, University of Virginia

Date:

30 April 2022

Summary:

Data description of fields in CSV files for deposited synthetic populations.

Version:

0.9

Overview:

A synthetic population as deposited will consist of the list of files specified below. Here `region` denotes the country or state, e.g., France or Virginia (US). The parameters `major` and `minor` encode the major and minor version of the population data.

Files:

- Person data   `{region}_person_v_{major}_{minor}.csv`
- Household data   `{region}_household_v_{major}_{minor}.csv`
- Residence location data   `{region}_residence_locations_v_{major}_{minor}.csv)`
- Activity location data   `{region}_activity_locations_v_{major}_{minor}.csv`
- Activity location assignment data `{region}_activity_location_assignment_v_{major}_{minor}.csv`
- Contact matrix data   `{region}_contact_matrix_v_{major}_{minor}.csv`

Construction parameters and notes

Activity encoding:

Enumerations used for encoding activity types:

```
Home:        1

Work:        2

Shopping:    3

Other:       4
```

```
School:          5
College:         6
Religion:        7
```

Note regarding IDs:

IDs are unique within a population. Populations from different regions, or from different versions of populations from the same region, will generally have incompatible IDs in the sense that they may clash. In particular, be aware of the following:

- `pid, hid, alid, rlid`: not unique across instances;
- `alid, rlid`: these are unified as `lid` (location IDs) at the location assignment step.

Person file (CSV):

`hid,pid,person_number,age,sex`

`0,0,1,55,1`

`1,1,177,0,1`

- **hid:** Household ID
- **pid:** Person ID
- **person_number:** numbering of persons within household (SPORDER in PUMS)
- **age:** age of person (AGEP in PUMS)
- **sex:** {1: MALE; 2: FEMALE} (SEX in PUMS)

Household file (CSV):

`hid,admin1,admin2,hh_size`

`1,140003969,160031729,1`

`2,140003973,160031700,1`

- **hid:** Household ID; same as in the person file;
- **admin1:** ADCW ID for admin1 region;
- **admin2:** ADCW ID for admin2 region; same as admin1 if a region does not have and admin2 division as per ADC Worldmap.
- **hh_size:** Number of persons in family

Residence locations file (CSV):

`rlid,longitude,latitude,admin1,admin2`

`1052998,-82.3277705,29.6538605,12,001`

```
1114821,-82.332464,29.6548805,12,001
```

- **rlid**: residence location ID
- **longitude**: the longitude of the location
- **latitude**: the latitude of the location
- **admin1**: see household file description
- **admin2**: see household file description

Activity locations file (CSV):

```
alid,longitude,latitude,admin1,admin2,work,shopping,school,\
other,college,religion
1,-99.45144,27.53944,48,479,1,0,0,1,0,0
2,-98.75611,33.3707,-1,48,503,1,218,0,1,0,0
```

- **alid**: the activity location ID
- **longitude**: the longitude of the location
- **latitude**: the latitude of the location
- **admin1**: see household file description
- **admin2**: see household file description
- **work**: attractor weight for work activity
- **shopping**: attractor weight for shopping activity
- **school**: attractor weight for school activity
- **other**: attractor weight for other activity
- **college**: attractor weight for college activity
- **religion**: attractor weight for religion activity

Location assignment file (CSV)

```
hid,pid,activity_number,activity_type,start_time,duration,lid
812952,74,0,1,0,25200,74
812952,74,2,5,28800,32400,1774067
```

- **hid:** the household ID of the person
- **pid:** the person ID of the person
- **activity_number:** the number of the activity in the activity sequence to which it belongs
- **activity_type:** the activity type (see above)
- **start_time:** the start time of activity in seconds since midnight Sunday/Monday
- **duration:** the duration of the activity in seconds
- **lid:** the location ID of the location where the activity takes place. (rlid or alid)

Contact matrix (CSV)

```
src_age,dst_age,num_contacts,total_duration,num_people
age_00_04,age_00_04,8624048,3.22569e+06,1680684
```

- **src_age:** age range of source group (as formatted string)
- **dst_age:** age range of target group
- **src_age:** age group of source
- **num_contacts:** number of contacts recorded in network
- **total_duration:** total contact duration in seconds
- **num_people:** number of people in source age group