UMN Accessibility to Jobs by Transit – 2019 Update

2022-11-02

Abstract: The previous version of this dataset (2014) found on MAPC Internal Data Browser has two variables for each blockid. The 2019 dataset update has ninety-two variables for two observation years (2016 and 2017). The published datasets available for download are now in geopackage format that contain both the spatial data (blocks as polygons) and separate tabular data layers of access to jobs broken down at five minute intervals, from 5 minutes to 60 minutes. Therefore, processing the datasets requires a new approach to connect to geopkg format, and publication on Data Browser or DataCommon will require a new table definition and associated metadata.

Data Source:

https://conservancy.umn.edu/handle/11299/218072

Documentation:

K://DataServices/Datasets/Transportation/UMN_accessibility_observatory/2019/14460_Boston-Cambridge-Quincy_MA-NH/2019DataDoc_all_LEHD.pdf

note: the Field Descriptions provided with the 2019 data refer to the years 2015 and 2016, but the data layers in the geopkg have the suffixes _16 and _17, therefore it's assume that the descriptions were not updated and that the variable names are correct.

File Format Notes: The official ESRI documentation on handling Geopkg files notes that editing would require third party software, since ArcGIS cannot do this. A basic example on opening Geopkg layers in ArcGIS was successful, but the layers added to a map needed to be converted to feature classes in a File Geodatabase for any joins or other operations to be done. (Note, ArcGIS Pro also crashed when attempting the conversion step.)

See: https://www.esri.com/arcgis-blog/products/product/data-management/how-to-use-ogc-geopackages-in-arcgis-pro/

How to open in ArcGIS: K:/DataServices/Datasets/Transportation/UMN_accessibility_observatory/2019/Using%20Geopackages%20in%20ArcGIS.pdf

Using Geopkg in R:

Because the MA blocks in the UMN publications are found in two separate regions (14460 Boston and 39300 Providence), opening and joining layers in R was the easiest way to obtain the tabular data for MA.

The process involves importing the tabular data for a particular time interval (such as 45 minutes commuting time by transit) and binding them into one MA data frame, then joining those to the spatial data (blocks as polygons). There were approximately 84,000 blocks in the 14460 layer and 35,000

blocks in the 39300 layer, so a total of 119,000 blocks for the complete set of both regions. Filtering by MA blockids resulted in a final dataset of 86,000 blocks.

The process of connecting to the geopkg in R was quite straitght-forward and easy to reproduce. In the process of selecting out the datasets for the commute times of 15min, 30min, 45min, and 60min, a video screen-capture was made and posted to youtube.

See: Use Geopackage data in R [geopkg in R] - YouTube

Columns and variables in the datasets.

The current dataset is on the MAPC internal data browser:
Transporation | Accessibility to Jobs by Transit | Accessibility to Jobs by Transit within 30 minutes (blocks, Boston MSA)

The two variables for each blockid are:
threshold time threshold (in minutes) for job accessibility
jobs employment accessible within a time threshold

because the dataset itself is limited to the 30 minute transit window, the threshold value is = 30. in effect, the only variable for this dataset is the job count.

For the 2019 dataset, there are two years of data included for each 5 minute interval (2016 and 2017).

Each tabular data layer contains 92 variables for each year. There are two main types of LEHD vars with prefixes w_ (WAC worker area characteristics) and r_ (RAC residence area characteristics).

These can be broken down into the following groups:

WAC:
workers: jobs by three age brackets (<= 29, 30-54, >=55)
workers: jobs by three earnings brackets (<=$1250, $1251-$3333, >=$3334 /month)
workers: jobs by twenty NAICS sectors
workers: jobs by six race categories
workers: jobs by two ethnicities (Not Hispanic or Latino, Hispanic or Latino)
workers: jobs by four educ attainment categories (< high school, high sch no col, some col, >= bach)
workers: jobs by gender (m/f)
workers: jobs at firms with firm age in five brackets
workers: jobs at firms with firm size in five brackets

RAC:
same as above without the *firm age* and *firm size*.

Extracted layers as Shapefiles (for example **mass_blocks_45min.shp**) :
K:\DataServices\Datasets\Transportation\UMN_accessibility_observatory\2019\merged_shp

sample map of r_c000_16 (residential area characteristics, total workers, 2016)