

Data Map for "Potential for mobility hub solutions in rural areas with low bus frequency".
Author: Scott Urban, College Lecturer for Visiting Students in Economics
scott.urban@hertford.ox.ac.uk

This data map is for re-creating the data in this paper. For those whose primary interest is in using the bus-frequency data produced by the paper, please refer to "data_notes.txt".

I. Introduction

The data analytics for this paper were carried out in Stata version 14.0 and QGIS version 3.32 ("Lima").

The routines for the "Gtfs_Sep2023_OA.do" file are data-intensive and require significant computing time. (I ran them on Intel Core i7 with 64GB RAM.)

You can skip the routines importing and processing the raw data by using any of the three "substantial" Stata datasets described in section III below.

This data map sets out the workflow of the Stata do-files and lists the underlying datasets that are called by the do-files -- some of which are included in Data.zip, others of which need to be downloaded as described in section V below.

II. Stata do-files

A. Gtfs_Sep2023_OA.do

1. Imports into Stata the GTFS data downloaded from UK DfT Bus Open Data Service (see "Underlying datasets required for Stata do-files" below).
2. Creates a CSV file (gtfs_stops.csv) of unique stops in stops_id dataset, with latitude and longitude, to import into QGIS
3. Merges this back in, with bus-stops geolocated in OAs (produced in QGIS -- see "QGIS" below).
4. Counts number of unique bus departures at UK ONS Census levels: OA, LSOA and MSOA.
5. Saves an OA-level Stata dataset: busfreq_England3OA.dta

B. Paper_Sep2023_OA.do

1. Imports and saves in Stata format a number of lookup tables and other datasets.
2. Imports OA-level public transport isochrone downloaded from ONS Geoportal.
3. Merge these with OA-level bus-frequency statistic developed in "Gtfs_Sep2023_OA.do" (saved in "busfreq_England3OA.dta")
4. Merge these with OA-level Propensity to Cycle data downloaded from the Propensity to Cycle Tool
5. Apply Bus Rule 1 and Bus Rule 2 (see paper)
6. Generate histogram and other figures for paper
7. Tabulate statistical tables for paper

8. Produce maps for paper (shapefiles must be downloaded from ONS -- see section V below).
9. Run an analysis comparing Transport Desert scores with my Frequency variable

III. Substantial Stata datasets produced

A. gtfs_england2OA.dta

The fullest picture of GTFS data -- contains all bus stops, routes and schedules; geolocated to ONS Census areas.

B. busfreq_England3OA.dta

An OA-level dataset reporting the raw bus frequency for localities in England as at September 2023. The identifying variable is oa21cd

C. rural_1.dta

An OA-level dataset with a variable reporting the raw and adjusted (i.e. rules 1- and 2-applied) bus frequency data, merged to ONS isochrone data and Propensity to Cycle data.

IV. QGIS (geolocating bus-stops in UK ONS Census Output Areas)

Here is how I joined the stops.txt data to OA boundaries (from ONS Geoportal).

Stata

- Create gtfs_stops.csv (as described in "Gtfs_Sep2023_OA.do" above) which has one observation per stop.

QGIS

- Load this point-data file. It is in CSV form, so choose "Delimited text" from the Data Source Manager in QGIS. Be sure to set the layer to WGS84 EPSG:4326 (if not already set that way).
- Reproject this layer to 27700 (British National Grid).
- Repair the ONS OA boundaries map, producing a "Fixed geometries" layer. Note: I used the full-realm OA boundaries, not the clipped ones.
- Note: Create a spatial index of the input file (Reprojected stops data) first. Do this by right-clicking on it --> Properties --> Source (from left-hand panel) --> Create spatial index.
- "Join Attributes by Location" tool. Input: Reprojected stops point-data. Tick "Intersect". Choose "One-to-many" relationship. "By comparing to:" is "Fixed geometries". --> "Run" the algorithm.
- Export this joined layer as a CSV, which will be used in Stata as above.

V. Underlying datasets required for Stata do-files

File	Comment
.txt files contained in "itm_all_gtfs" folder These need to be downloaded from DfT using link noted in the cell adjacent -->	Please download these from DfT Bus Open Data Service, https://data.bus-data.dft.gov.uk/timetable/download/
buapop.xls	Population at Built-up Area level (ONS)
custom-filtered-2023-09-08T15_14_00Z.csv	ONS: population at OA level
East_Midlands East_of_England London_East London_West North_East North_West_North North_West_South South_West West_Midlands Yorkshire_and_the_Humber South_East Each of the above appear as separate CSV files of the following naming convention (with `v` standing in for the above names): UK_Travel_Area_Isochrones_(Nov_Dec_2022)_by_Public_Transport_and_Walking_for_`v`_-_Generalised_to_10m.csv	Please download these from ONS Geoportal https://geoportal.statistics.gov.uk/search?tags=isochrones
Local_Authority_District_to_Combined_Authority_(December_2022)_Lookup_in_England.csv	A lookup table from ONS
Local_Authority_District_to_County_(April_2021)_Lookup_in_England.csv	A lookup table from ONS
Local_Authority_District_to_Region_(December_2022)_Lookup_in_England.csv	A lookup table from ONS
Lower_Layer_Super_Output_Area_(2021)_to_Ward_(2022)_to_LAD_(2022)_Lookup_in_England_and_Wales_v3.csv	A lookup table from ONS
Lower_Layer_Super_Output_Area_2021_EW_BFE_V8_7792113955784831350.csv	LSOA geometry from ONS
Lower_Tier_Local_Authority_to_Upper_Tier_Local_Authority_(December_2022)_Lookup_in_England_and_Wales.csv	A lookup table from ONS
LSOA_(2011)_to_LSOA_(2021)_to_Local_Authority_District_(2022)_Lookup_for_England_and_Wales_(Version_2) (exact fit).csv	A lookup table from ONS

File	Comment
LSOA_(2021)_to_Built_Up_Area_to_Local_Authority_District_to_Region_(December_2022)_Lookup_in_England_and_Wales_v2.csv	A lookup table from ONS
Middle_Layer_Super_Output_Areas_2021_EW_BFE_V6_698929597320552670.csv	MSOA geometry from ONS
Mobility Hub paper - Oct 2023.xlsx	A table of summary statistics reported by the "Paper_Sep2023_OA.do" file and copied manually to Excel.
oa_stops_joined4.csv	Bus stops geolocated in ONS Census Output Areas, using QGIS (see "QGIS" below).
Output_Area_to_Lower_layer_Super_Output_Area_to_Middle_layer_Super_Output_Area_to_Local_Authority_District_(December_2021)_Lookup_in_England_and_Wales_v3.csv	A lookup table from ONS
Output_Area_to_Lower_layer_Super_Output_Area_to_Middle_layer_Super_Output_Area_to_Local_Authority_District_(December_2021)_Lookup_in_England_and_Wales_v3.csv	A lookup table from ONS
Potential for 'Mobility Hub' Solutions.xlsx	For a list of local-authority names
Propensity to cycle CSV files, mine are: pct_avon pct_oxfordshire pct_london pct_set1 pct_set2 pct_set3	Download from PCT.bike https://www.pct.bike/m/?r=cambridgeshire#nattcommutezones
Rural_Urban_Classification_(2011)_of_Lower_Layer_Super_Output_Areas_in_England_and_Wales.csv	ONS rural-urban classification (OA-level)
Rural_Urban_Classification_2011_lookup_tables_for_local_authority_areas.xlsx	ONS rural-urban classification (LAD-level)
ukpopestimatesmid2021on2021geographyfinal.xls	Population estimates from ONS
Output_Areas_2021_EW_BFC_V8_-9054797207862162063/OA_2021_EW_BFC_V8.shp CTYUA_MAY_2023_UK_BFC.shp LAD_MAY_2023_UK_BFC.shp LSOA_2021_EW_BFE_V8.shp WD_DEC_22_GB_BFC.shp	These shapefiles are all available from ONS Geoportal: https://geoportal.statistics.gov.uk/ Select "Boundaries" menu from the top or simply search for the files in the search bar. NOTE: You will need the .dbf files that go with each of these .shp files.

VI. Underlying datasets required for QGIS

File	Comment
Shapefile for OA boundaries. These are too big to include here, but can be downloaded from ONS at this link -->	https://geoportal.statistics.gov.uk/datasets/ons::output-areas-2021-boundaries-ew-bfe/explore
gtfs_stops.csv	Bus-stop dataset (1 obs per row) with latitude and longitude. Created by "Gtfs_Sep2023_OA.do" described above.