# Procedure followed in splitting the urban sample

Procedures were carried out in Excel unless otherwise mentioned.

1. Beginning from the file dwellings-appliances-Filtered-UR-20150203.csv:
   1. Removed the rows of rural households, leaving the 197 urban ones, and

assigned random numbers between 0 and 1, using the RAND() function, to each row, saving the file as dwellings-appliances-Filtered-Urban-20150203-random.csv.

* 1. Sorted the rows in increasing order of these numbers, saving the file as dwellings-appliances-Filtered-Urban-20150203-random-sorted.csv.
  2. Saved the headings row, plus the first 98 household rows, without the random numbers column, as dwellings-appliances-Filtered-Urban1-20170405.csv.
  3. Saved the headings row, and the 99th through 196th household rows (rows 100 to 197), without the random numbers column, as dwellings-appliances-Filtered-Urban2-20170405.csv. (The 197th household row is thus not included in either Urban1 or Urban2).

1. Beginning from the file script-dwellings-Urban-20141126.csv:
   1. Using emacs, inserted a column (column C) headed “num”, and containing integers between 1 and 16, each integer corresponding to a type of dwelling, as follows:
2. Bungalow, detached, cavity wall
3. Bungalow, detached, solid wall
4. Bungalow, semi-detached, cavity wall
5. Bungalow, semi-detached, solid wall
6. Flat, non-top-floor, cavity wall
7. Flat, non-top-floor, solid wall
8. Flat, top-floor, cavity wall
9. Flat, top-floor, solid wall
10. House, detached, cavity wall
11. House, detached, solid wall
12. House, end-of-terrace, cavity wall
13. House, end-of-terrace, solid wall
14. House, mid-terrace, cavity wall
15. House, mid-terrace, solid wall
16. House, semi-detached, cavity wall
17. House, semi-detached, solid wall

This is the order in which dwellings were added to the CEDSS street network for all previous urban runs (I don’t know why semi-detached houses don’t directly follow detached ones, but this doesn’t seem important).

* 1. Using emacs, inserted a column (column E) headed “bedrooms”, and containing integers between 1 and 6, each corresponding to the number of bedrooms in the dwelling (a 6 indicating 6 or more bedrooms), saving the file as script-dwellings-Urban-20141126-withnums.csv.
  2. Copied the column of random numbers from dwellings-appliances-Filtered-Urban-20150203-random.csv as column F, and sorted the file on this column (smallest to largest), saving the results as script-dwellings-Urban-20141126-withnums-random-sorted.csv.
  3. Deleted column F, and the last row. Cut rows 100-197 and pasted them as rows 2-99 in a new file. Saved the rest of the file (now containing rows 1-99) as script-dwellings-Urban1-20170405.csv. Copied row 1 to the new file, and saved that as script-dwellings-Urban2-20170405.csv.
  4. Sorted script-dwellings-Urban1-20170405.csv and script-dwellings-Urban2-20170405.csv by columns C and E, saving the files as script-dwellings-Urban1-20170406.csv and script-dwellings-Urban2-20170406.csv.
  5. Copied colums C-E from script-dwellings-Urban1-20170405.csv and script-dwellings-Urban2-20170405.csv to dwellings-appliances-Filtered-Urban1-20170405.csv and dwellings-appliances-Filtered-Urban2-20170405.csv respectively, as columns D-F, sorted these files on column D then column F, and saved temporarily as dwellings-appliances-Filtered-Urban1-20170406x.csv and dwellings-appliances-Filtered-Urban2-20170406x.csv (see steps 6a and 6b for the purpose of this).
  6. Removed the columns C and E from script-dwellings-Urban1-20170406.csv and script-dwellings-Urban2-20170406.csv, and re-saved the files.

1. Starting from script-patch-20141126.csv, used the ordering of rows in script-dwellings-Urban1-20170406.csv and script-dwellings-Urban2-20170406.csv to produce script-patch-Urban1-20170406.csv and script-patch-Urban2-20170406.csv. Each dwelling name (e.g. dwu-940) was placed in the appropriate script-patch file twice in succession (using the ordering of dwelling-patches defined in the file Street network construction procedure.pdf), with the second occurrence having an “00” added (e.g. dwu-94000). This ensures that all the duplicates have higher numerical indices than the originals.
2. Starting from dwellings-appliances-Filtered-Urban1-20170405.csv and dwellings-appliances-Filtered-Urban2-20170405.csv:
   1. Resort the files by column B (respnr).
   2. In each file, duplicate all but the first row.
   3. Save as dwellings-appliances-Filtered-Urban1-20170406.csv and dwellings-appliances-Filtered-Urban2-20170406.csv
3. Starting from dwellings-appliances-Filtered-Urban1-20170406.csv and dwellings-appliances-Filtered-Urban2-20170406.csv:
   1. Multiply each entry in column B from row 100 onward by 100.
   2. Save the files as dwellings-appliances-Filtered-Urban1-20170407.csv and dwellings-appliances-Filtered-Urban2-20170407.csv.
4. Starting from script-dwellings-Urban1-20170406.csv and script-dwellings-Urban2-20170406.csv .
   1. Copy column B (respnr) from dwellings-appliances-Filtered-Urban1-20170406x.csv to column D in script-dwellings-Urban1-20170406.csv, and column B from dwellings-appliances-Filtered-Urban2-20170406x.csv to column D in script-dwellings-Urban2-20170406.csv.
   2. Resort the rows of script-dwellings-Urban1-20170406.csv and script-dwellings-Urban2-20170406.csv on column D (the point here is that a numerical sort on this column gets column B in the desired order, while sorting on column B does not). Remove column D.
   3. Duplicate rows 2-99 as rows 100-197. Save the files.
   4. Using emacs, add “00” to the end of each entry in column B from row 100 onward. Save the files as script-dwellings-Urban1-20170407.csv and script-dwellings-Urban2-20170406.csv.