Planning Primary Care Locations using ArcGIS Online

Overview:

In this practical, we will explore how to use ArcGIS Online to implement location-allocation modelling. We will take the example of primary care delivery in the Southampton Clinical Commissioning Group (CCG), UK, and assume that several surgeries (primary care facilities) are having to close due to a budget rationalisation.

Data:

Data on provision:

Point your browser at:

https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data

…where you will find a primary care database. Download the following file:

epraccur – current GP practices, including their addresses.

[Additional relevant files, which we will not use here:

ebranchs – although it is tempting to use this file, unfortunately, the addresses and postcodes are for the main surgery, not the branch surgery. Our analysis will therefore not take into account the presence of branch surgeries.

Epracmem – contains details of the doctors working at each surgery]

Make sure you take a look at the .pdf of metadata that explains what the **epraccur** file contains.

After downloading **epraccur**, unzip the .csv and open it in Excel. Add titles for columns A (practice ID), B (practice name), M (status), and J (postcode). Delete all other columns. Add a new column called pcdistrict and enter this formula in the first cell to extract the postcode district (we will focus only on general practices in Southampton):

=LEFT(D2, 2)

Drag the formula down so it is copied to all cells in this column:

	A	B	C	D	E
1	pracid	name	postcode	status	pcdistrict
2	A81001	THE DENSI	TS18 1HU	A	TS
3	A81002	QUEENS P	TS18 2AW	A	TS
4	A81003	VICTORIA	TS26 8DB	C	TS
5	A81004	BLUEBELL	TS5 8SB	A	TS
6	A81005	SPRINGWI	TS14 7DJ	A	TS
7	A81006	TENNANT	TS18 2AT	A	TS
8	A81007	BANKHOU	TS24 7PW	A	TS
9	A81008	ALBERT HO	TS6 6TD	C	TS
10	A81009	VILLAGE M	TS5 6HF	A	TS
11	A81011	CHADWIC	TS24 7PW	A	TS
12	A81012	WESTBOU	TS3 6AL	A	TS
13	A81013	BROTTON	TS12 2FF	A	TS
14	A81014	QUEENSTF	TS23 2LA	A	TS
15	A81015	LAGAN SU	TS10 1TZ	A	TS

Sort on your new column, then delete all rows containing postcodes that do not begin ‘SO’.

Now save your revised .csv file with a new name and close down Excel.

Data on demand:

As a measure of demand, we will use census output areas, so head for:

https://borders.ukdataservice.ac.uk/

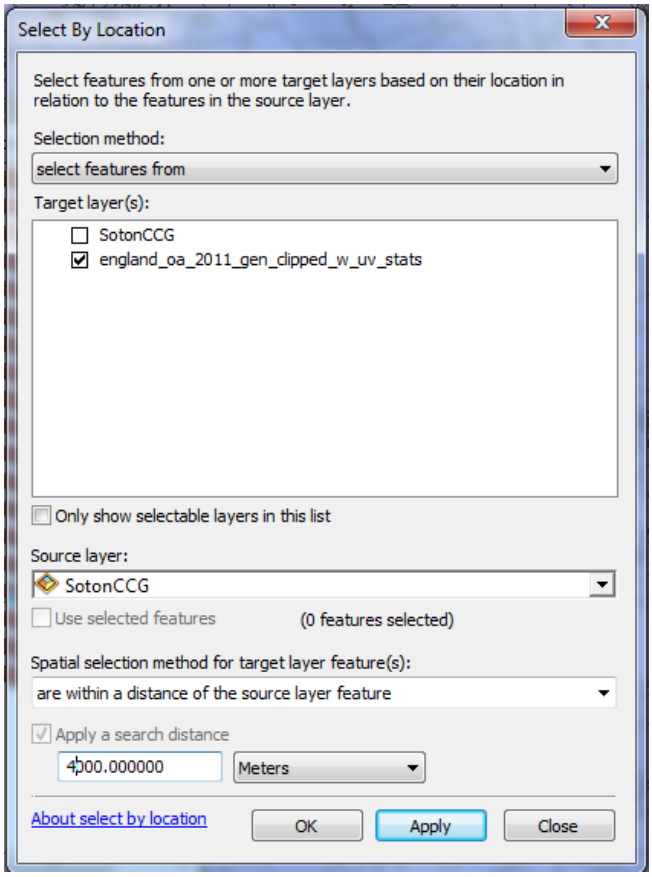
Click on easy download, select 2011 English census output areas, then select ‘English Census Output Areas, 2011, Clipped and Generalised with Univariate Census 2011 data’, choosing shape file format.

Click easy download again, but this time select English NHS Clinical Commissioning Groups, then ‘English NHS Clinical Commissioning Groups, 2013, Clipped and Generalised (simplified polygon geometry)’ and again choose the shape file version. Unzip both files, and be sure to check out the meta-data and terms and conditions that comes with each one.

Pre-processing in ArcGIS

Open up both sets of administrative boundaries that you just downloaded in ArcGIS. Use ‘Select by attributes’ to select ‘NHS Southampton’ from your CCG layer. Right-click on the layer, then select data / export data, to create a new output file called **sotonccg.**

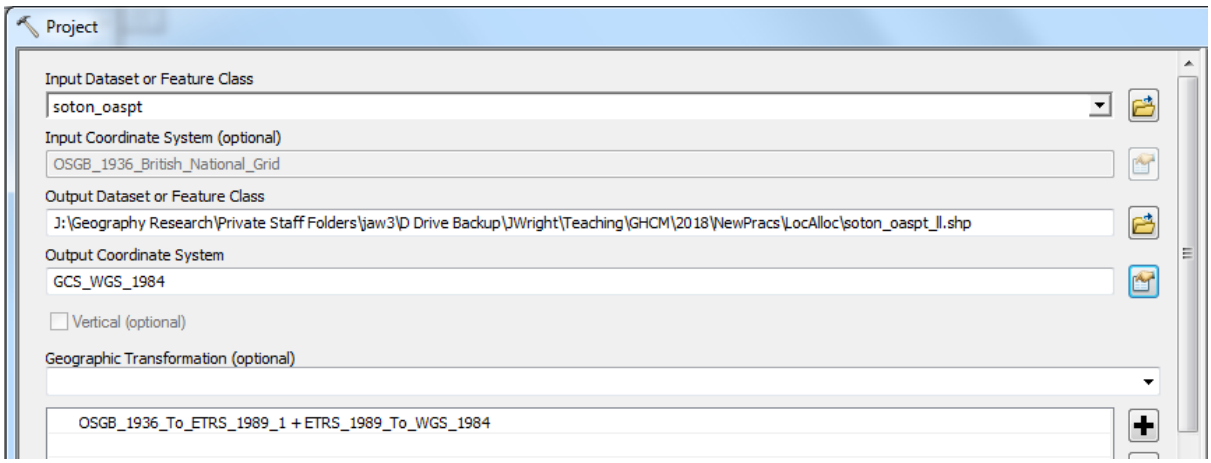
Use selection / Select by location, to select only those output areas within 4km of the Southampton CCG (we use 4km here, so that we can account for demand and facilities around the margins of the CCG):



Having selected those output areas that are close to the Southampton CCG, right-click on the output areas layer and select Data / export data, saving this subset of output areas as **soton_oas.**

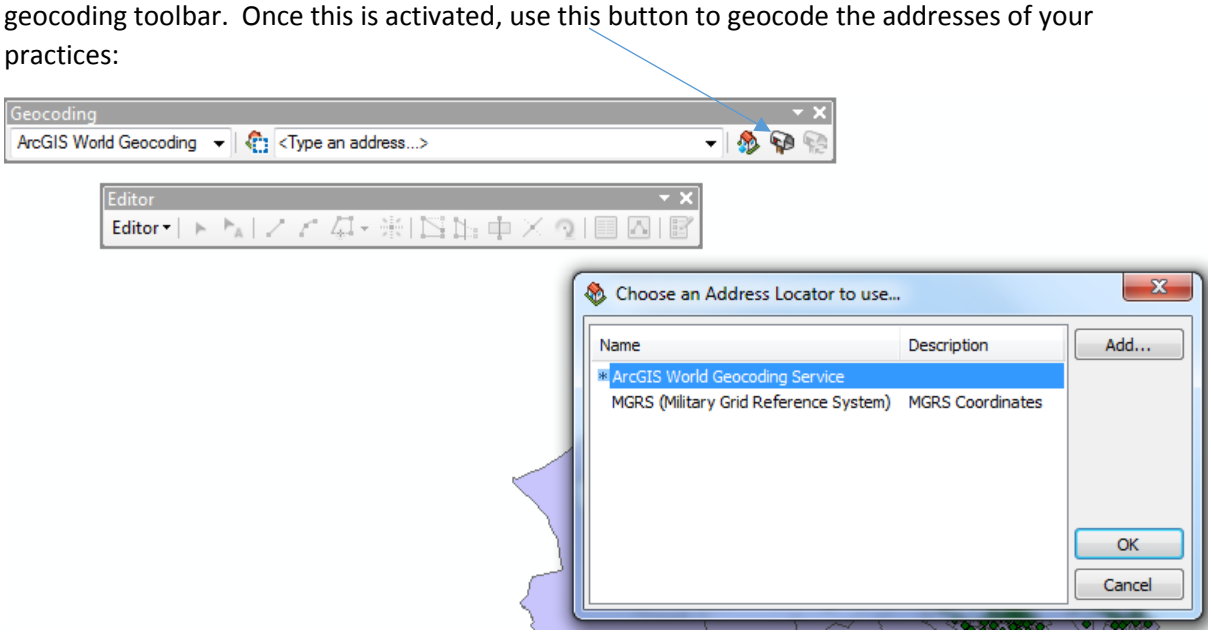
ArcGIS Online will be expecting demand to be represented as points, so we now need to convert these output area polygons to points. To do this, head for the Arctoolbox, and choose data management tools / features / feature to point. Use your Southampton output areas as the input and call the output **soton_oapts.**

Finally, we need to convert any layers we upload into ArcGIS Online into latitudes and longitudes, referenced to the WGS 1984 datum. To do this head for the Arctoolbox, then choose data management / projections and transformations / project, being sure to save the output in shape file format (because this is what ArcGIS Online expects):



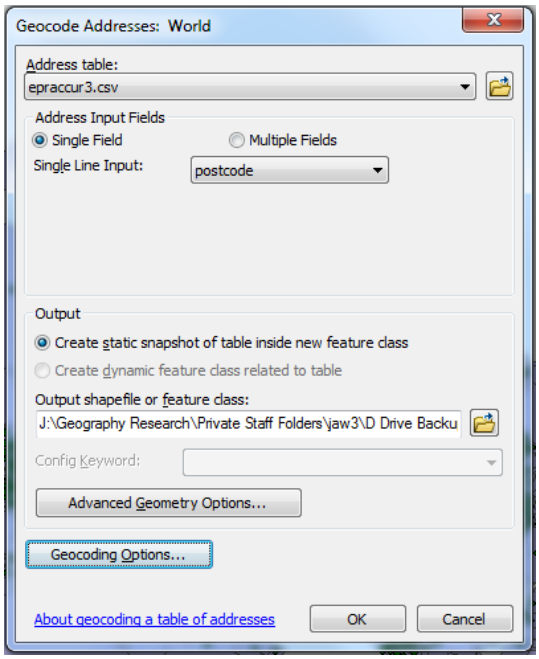
In a moment, we will need to make a distance calculation using the CCG boundaries, so take your Southampton CCG boundary file (**sotonCCG)** and convert this to latitude and longitude WGS 1984 as well.

Now open up your .csv of Southampton general practices that you created earlier. Right-click somewhere over the grey toolbar area at the top and choose *geocoding* to activate the ArcGIS geocoding toolbar. Once this is activated, use this button to geocode the addresses of your practices:



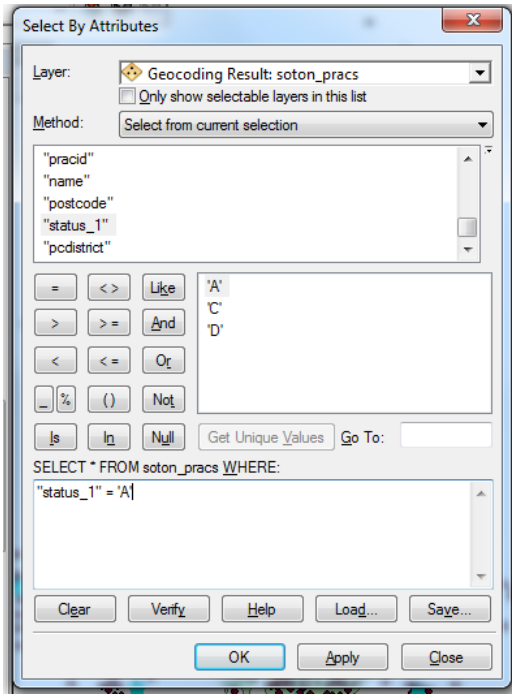
Select the 'Arcgis World geocoding service' as the address locator (which contains names plus latitudes and longitudes to be used during geocoding) and press OK. Next you will need to sign In to Arcgis Online with your university username and password.

On the next screen, choose your .csv file of practice characteristics and select *single field* geocoding based on the **postcode** field:



Now select only those practices that are within 4,000 metres of the Southampton CCG, just as we did before with the output areas. Note that you should use the version of the CCG map layer that is in geographic (latitude and longitude) coordinates.

Finally, we need to make sure that we only focus on practices that are open. To do this, head for the Selection menu / select by attributes, and select 'status_1' field codes of 'A' (if you refer back to the accompanying pdf with the practices data, you will see that this code means 'practice open'):



Note that our original 'status' field has been renamed 'status_1' because ArcGIS creates another field called 'status' during geocoding. Note also that under *Method* we need to choose *select from current selection*, so that we obtain practices that are both within 4km of the CCG and open.

Right-click on the layer, choose data and then export data again, to make a new layer with only this subset of the practices in shape file format. Have a quick look at the attribute table for this new layer and make a note of the number of records it contains.

Close down ArcGIS desktop and create two zip files containing all the constituent files of each of the two shape files you created – one for your output area centroids and one for the practices you just created.

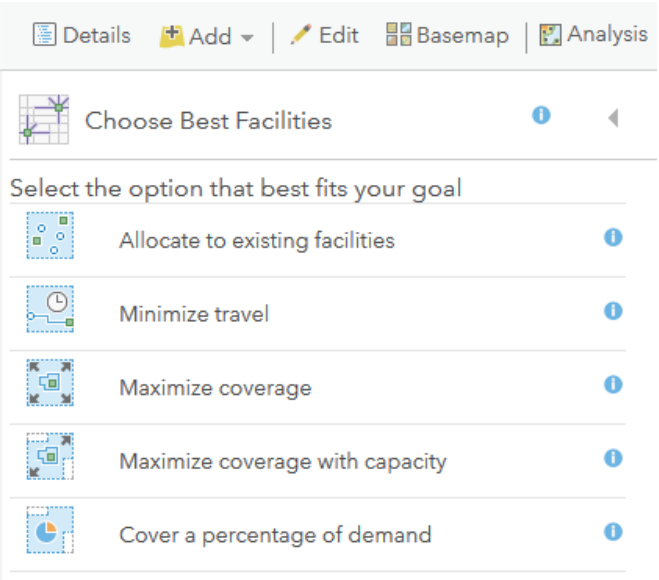
ArcGIS Online

Log in to the university's ArcGIS Online account using your username and log-in:

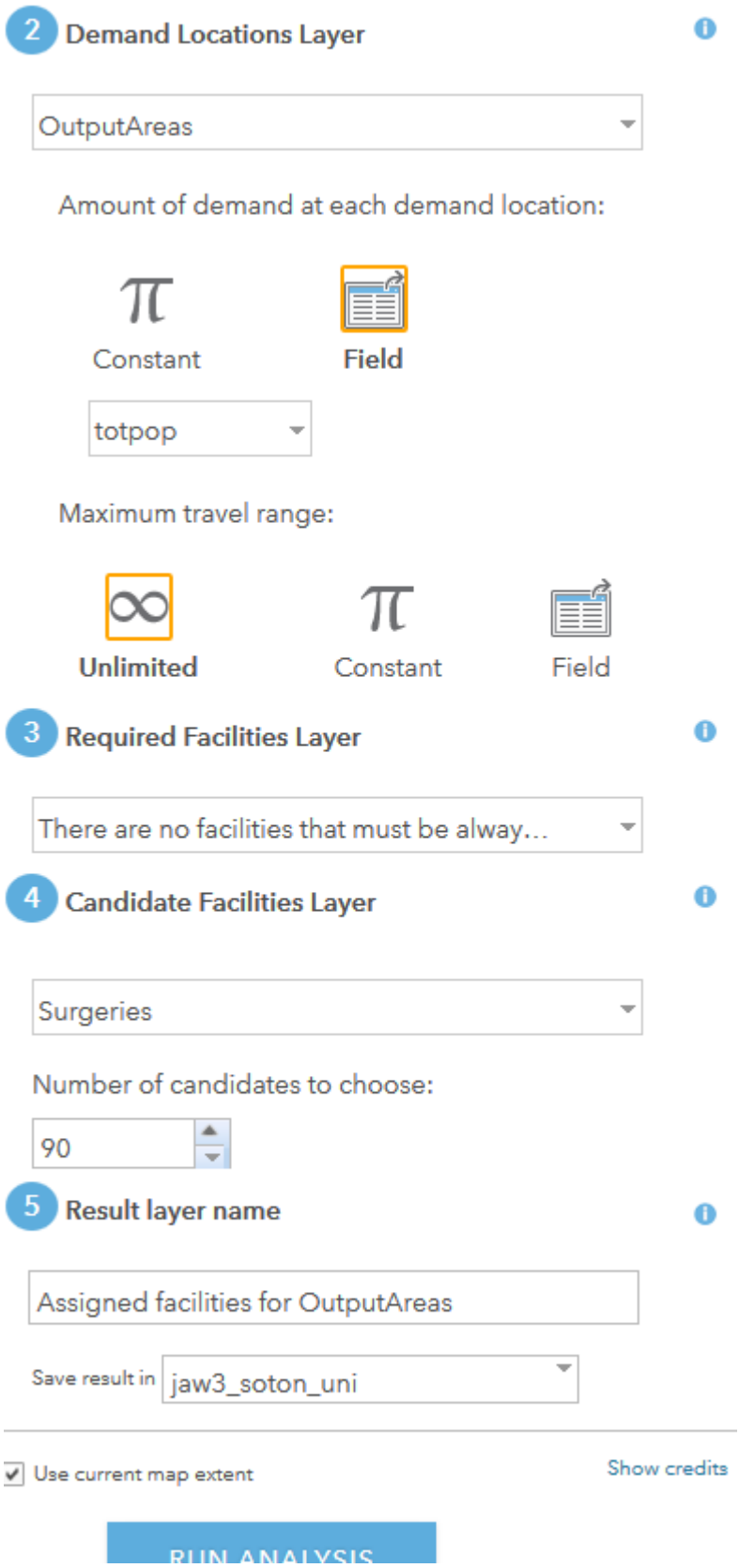
https://soton-uni.maps.arcgis.com/home/signin.html

Use the 'add' button / *add layer from file* to upload the two zip files (of output area points and general practices), following the method we used in an earlier practical.

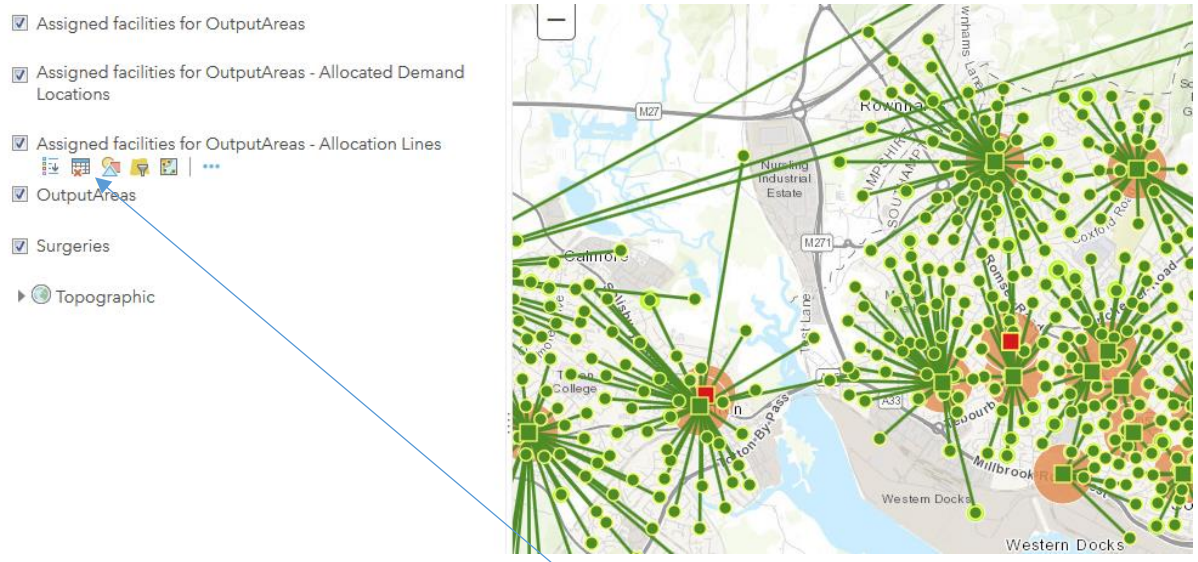
Next click on *Analysis* and then select *Find Locations* and *Choose Best Facilities*:



Select *minimize travel,* then using drive-times without traffic, select the output areas as your *demand layer,* choosing **totpop** (total population in 2011) as the *amount of demand at each demand location*. Use your general practices as the *candidate facility layer,* setting the *number of candidates to choose* to be slightly less than the total number of records in your general practices file (this will mean that several practices will be selected for closure) :



You can use *show credits* to check your credit balance with ESRI and how many credits the operation will cost. Otherwise, if you are happy, press *run analysis.* This analysis may take a little while to run, but you should see something like this:



Three new layers have been created, which are an updated demand layer (the circles - allocating the demand to facilities), an updated facilities layer (the squares - indicating whether or not they have been chosen and therefore remain open), and an updated allocation layer, shown as lines linking demand points to the facilities to which they have been allocated.

Explore the attributes for each of these in turn, using the attribute table button:

Assigned facilities for OutputAreas - Allocation Lines (Features: 1377, Selected: 1)

Allocated Demand	Total Weighted Travel Time (Minutes)	Total Travel Time (Minutes)	Assigned Facility ID	Allocated Demand ID	Total Travel Distance (Miles)
337.00	1,957.07	5.81	1	19	1.99
361.00	1,544.32	4.28	1	31	1.26
236.00	480.28	2.04	1	69	0.79
241.00	803.72	3.33	1	192	1.09

The allocation layer includes fields for the allocated demand (total population in 2011), the total weighted travel time (population * drive time in minutes), the total drive time and distance, and the ID of the facility that is the destination and the output area (demand point) that is the origin.

Assigned facilities for OutputAreas - Allocated Demand Locations (Features: 1377, Selected: 0)

Demand	Allocated Demand	Demand ID	Assigned Facility ID	Status
314.00	314.00	1	29	Ok
335.00	335.00	2	87	Ok
307.00	307.00	3	31	Ok

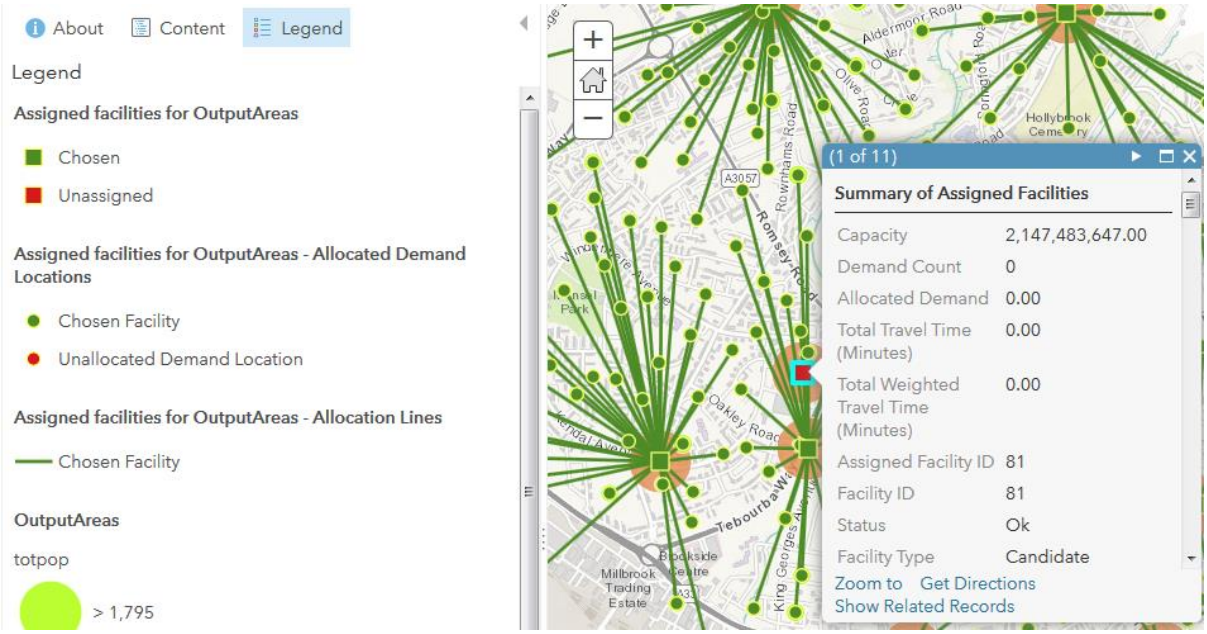
If you look at the demand locations (the output areas), new attributes include the demand (i.e. the total 2011 population at the output area), how much of this demand has been allocated, the facility to which the demand has been allocated, and whether or not the demand was successfully allocated to a facility (status).

Assigned facilities for OutputAreas (Features: 97, Selected: 0)

Capacity	Demand Count	Allocated Demand	Total Travel Time (Minutes)	Total Weighted Travel Time (Minutes)	Assigned Facility ID	Facility ID	Status	Facility Type
2,147,483,647.00	31	9,376.00	87.29	27,189.72	1	1	Ok	Chosen
2,147,483,647.00	0	0.00	0.00	0.00	2	2	Ok	Candidate

Finally, the facilities layer includes a very large number for capacity (here, we have assumed there is no limit to the population that can use each facility, so there is no capacity constraint or maximum population that can be served in this field). It includes a 'demand count', a count of the number of output areas assigned to it, and an 'allocated demand', a count of the population assigned to it. The 'total weighted travel time' is the sum of population times by drive-time in minutes for all the output areas assigned to the facility. Again, the ID is shown and a 'status' field indicates whether the software had any difficulty mapping the facility onto the road network. Crucially, the final field 'facility type' indicates whether or not the facility was 'chosen' (i.e. retained as a viable site) or remained a 'candidate' (i.e. not selected to remain open and scheduled for closure in our scenario). Such 'candidates' should not therefore have any demand allocated to them.

Displaying the legend makes it a little clearer what is happening, and you can click on an individual facility or point to view its attributes:



Following the instructions in our previous ArcGIS Online practical, if you wished, you could now download these three output layers into ArcGIS Desktop and work with them further.

Possible extensions to explore

This practical has provided an introduction to location-allocation modelling, but there is plenty of scope to do more, such as:

- Splitting the facilities out into two map layers – required facilities that must be retained (e.g. those lying within neighbouring CCGs), and candidate facilities, of which some are to be closed.
- Using a more sophisticated demand measure, e.g. weighting demand by different population groups
- Trying out some of the other location-allocation problems implemented in ArcGIS Online, such as *maximise coverage with capacity*