Choropleth Mapping

Overview:

In this exercise, we will look at how geospatial health data (on cancer mortality) can be visualised using ArcGIS, both using choropleths and with dot density maps.

Data set:

The data depicts US cancer mortality from 1970 to 1994 by county and was originally downloaded from here: http://nationalatlas.gov/atlasftp.html#cancerp. We wish to acknowledge both the National Atlas of the United States of America and the U.S. Geological Survey as the originators of the data set.

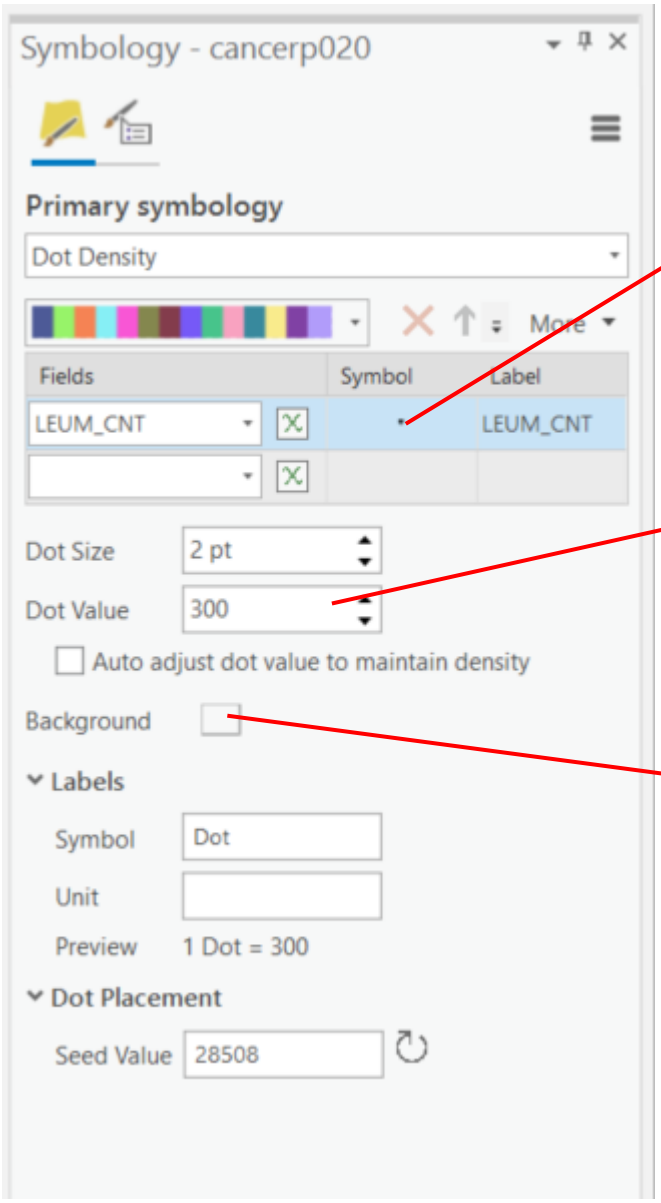
Exercise:

Briefly open up (e.g. with NotePad or a text editor) the 'cancerp020.txt' file, which contains the meta-data for this map layer, and have a quick look at the meta-data it contains. Next, open up the file in ArcGIS PRO. Because the USA straddles the international dateline and this data set is in latitude and longitude (and therefore there are a few parts of Alaska that appear to the right of your map display), you may need to zoom in on the North American continent to see your map properly.

Right-click on your map layer in the left-hand table of contents and select 'Attribute Table'. Within the attribute table, if you scroll to the right, you will find data on mortality from a great many different types of cancer (the first seven or so attribute fields contain locational information, such as area and perimeter). The first three letters of each field indicate the type of cancer ('acc' means all cancers – the other codes are contained in the meta-data), the next letter indicates whether the data relates to males ('m') or females ('f), and the final part of the field name indicates the type of data it contains. '…_rate' is a rate, adjusted to take account of the differing age structures in each county rather as we did in our earlier exercise. '.._cnt' is a count of the number of deaths recorded of that particular type of cancer. For example, 'blaf_cnt' contains a count of the number of females who died from bladder cancer. Now let us try generating maps of both the count of deaths and the rate.

Mapping counts:

Beginning with the counts, pick a type of cancer that is of interest to you (e.g. 'leum_cnt' = counts of leukaemia in males). Next, right-click on the map layer, select *symbology* to open the symbology pane. Select *Dot Density* as your Primary symbology. Under *fields*, select the field for your chosen type of cancer. ArcGIS will now create a dot density map, in which each death is represented by a dot randomly placed within each of the counties where the deaths were recorded. In the example below, each dot represents 300 deaths. Further ways of controlling what the output looks like are shown below.



You can double-click here to change the size and colour of your dots.

These settings control the density of dots, such as how many deaths each dot represents

These settings control the background polygon colours and boundaries.

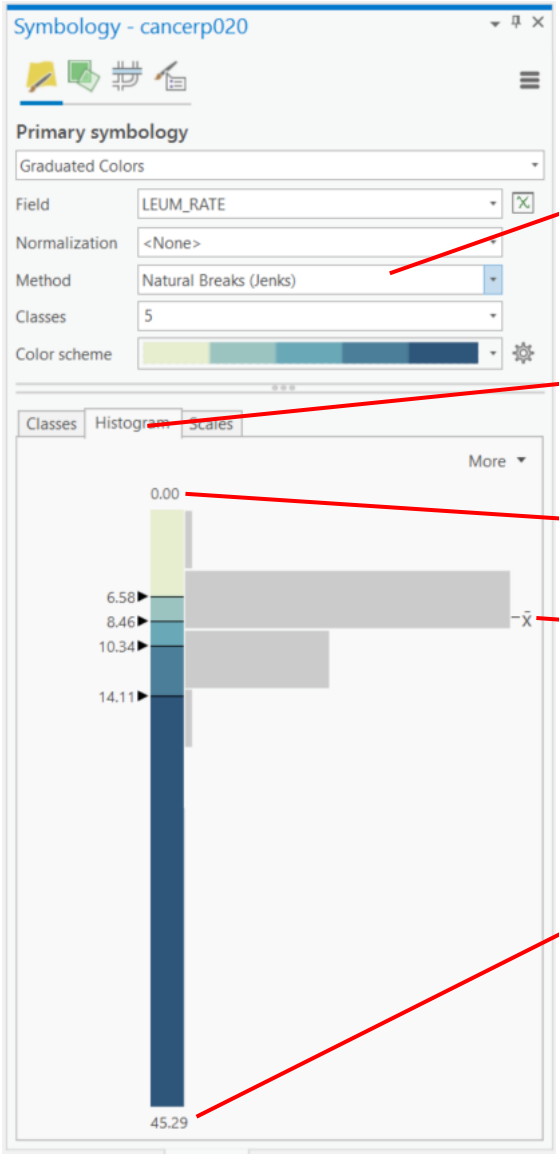
The effect of major settlements and cities is likely to be particularly striking in the output you produce. Were we to be planning healthcare on the basis of this map (for example, we might consider provision of staff for palliative or end-of-life care in this context), then such an illustration would be particularly helpful in understanding the counts of cases / deaths in each area as a basis for planning.

Mapping rates

Let us now try and map out the rates of cancer in each county. Let's head back to the *symbology* pane once more. This time, select *graduated* colors as your primary symbology to activate the choropleth mapping screen in ArcGIS PRO. Under *field* select the rate field for your chosen cancer

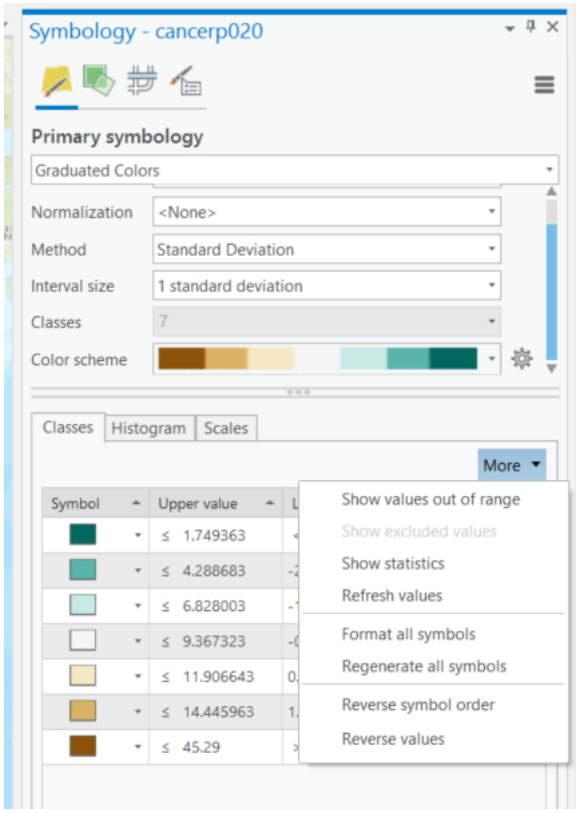
(e.g. 'leum_rate'). There is also a *normalisation* option here – we can ignore this here because we are already dealing with rates. Had we selected a count of deaths instead of a rate, we could have selected a field with total population under *normalisation* and ArcGIS would have divided the death count by total population for us to 'normalise' it (i.e. remove the effect of varying population sizes between areas). You should now see that ArcGIS PRO will attempt to apply a set of colours to the cancer rates you have selected. Typically, a small number of colour classes (usually five) will be selected and typically, a method known as 'natural breaks' will be used to decide at which rate to switch from using one colour to using a different one to display the data.

Click on the *Histogram* tab, ArcGIS PRO now shows us a histogram of the cancer rates by district. The gray bars show you the number of counties, whilst the black pointers show the class boundaries or break values where the map display switches from one colour to another. The black pointers can be dragged to change the ranges. By default, there are five different colour classes and under a 'natural breaks' scheme, the idea is that the break values between different colours fall where there is a 'break' in the histogram (in other words, there is an apparent dip in the grey bars). Try selecting different methods (e.g quantile, equal interval) and see how this changes the histogram.

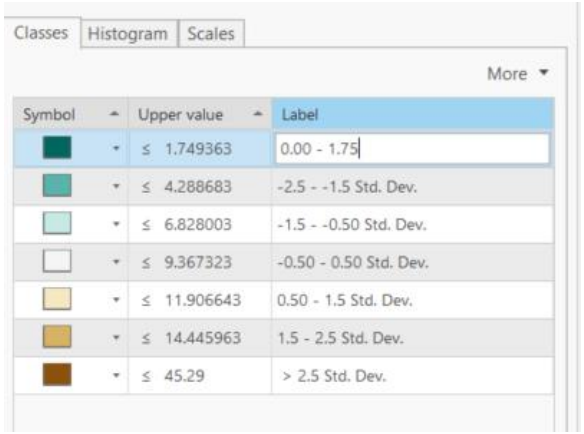


We could try and redesign our initial colour classification, so that areas with rates above the national average were assigned one shade, and areas below the national average were assigned a different shade. To do this, under *method* try selecting *standard deviation*. What this will do is to vary the colours, depending on whether they are above or below the national average. The rates will be expressed in terms of the degree of spread about the national average. The amount of spread around the average national rate can be measured in terms of what are known as standard deviations (this is a measure of how much spread about the mean there is in a data set. Roughly speaking, if data are normally distributed, 95% of data will fall within about 2 standard deviations of the mean value). Negative standard deviations indicate rates below the national average, whilst positive standard deviation values indicate rates above the national average – larger values are further away from the national average.

Depending on what colour scheme is selected by default, you may find it useful to flip the colours round, so that blue colours indicate low rates and red colours indicate high rates. You can do this by clicking on More and and selecting *Reverse symbol order.*



It is possible that you (and/or your readers) may find the use of labels that are in standard deviations confusing. If so, then you can click under the *label* column and overtype these with the original rates, as shown below:



Unlike the map of counts, which you could imagine using for healthcare planning, the map of rates (particularly when standardised to take account of age), can help us to understand patterns of disease risk and/or unusual artefacts in our data, such as possible misreporting. These two styles of map provide us with different perspectives on the same disease pattern.