Cartography. MOOC

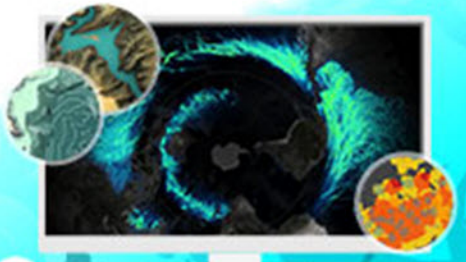
Exercise

Exploring Data Classification

Section 2 Exercise 2

04/2018

esri
THE SCIENCE OF WHERE™



Exploring Data Classification

Instructions
Use this guide and ArcGIS Pro to reproduce the results of the exercise on your own.

Note: The version of ArcGIS Pro that you are using for this course may produce slightly different results from the screenshots that you see in the course materials.

Time to complete
Approximately 30-40 minutes.

Software requirements
ArcGIS Pro 2.1

Introduction

All maps are made from data. Part of making a good map is to be numerate and to appreciate how your manipulation of the data plays a vital role in the message that your map communicates.

For a lot of topographic mapping, you're symbolizing data that has been surveyed, encoding meaning to the coordinates by symbolizing them as points, lines, and areas, often of different types.

For thematic mapping (https://bit.ly/2JCuwMA), you're often dealing with a dataset that represents a variable of interest. Your map will likely show certain trends, such as where the place that exhibits the highest or lowest value is or where certain areas share similar characteristics. The key is understanding that how you manipulate the data can tell different stories. It is important to ensure that you are not inadvertently telling a false story.

This exercise uses ArcGIS Pro to explore alternative methods of classifying numerical data for thematic mapping. Data classification is not unique to thematic mapping, but the techniques explored here can be used to understand and classify data more generally. You will create a range of choropleth maps (https://bit.ly/2qnGbpI) to illustrate how changing the classification changes the map.

Step 1: Download the exercise file

In this step, you will download and save the exercise file.

a. Open a new web browser tab or window.

b. Go to https://bit.ly/2qnpg6p and download the exercise file.

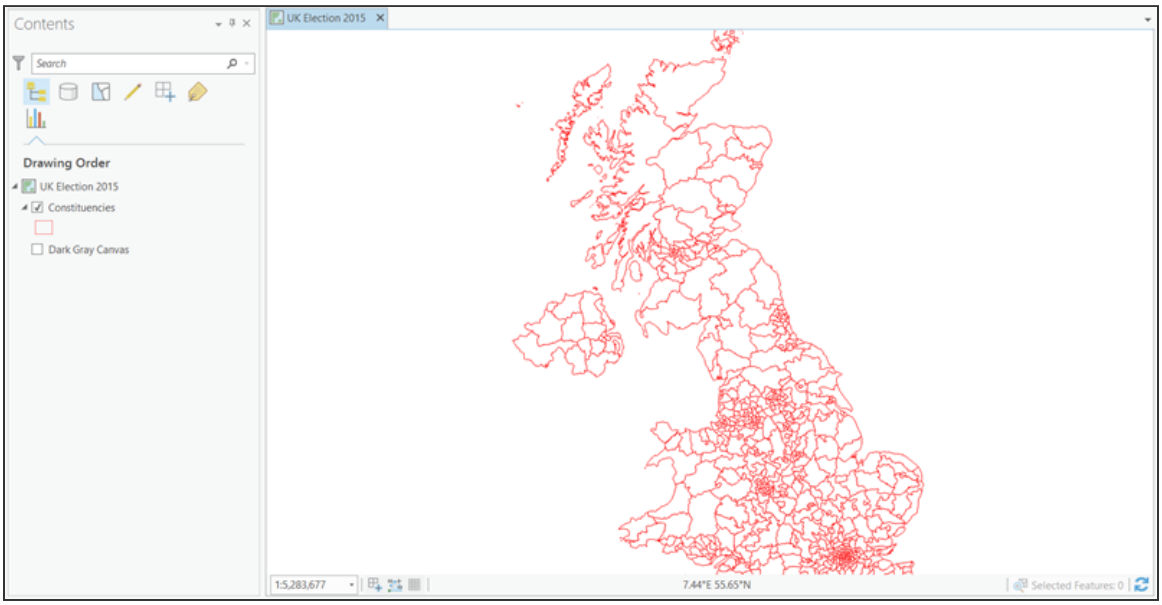
c. Save the file in a location that you will remember.

Step 2: Open an ArcGIS Pro project

You will open the exercise project in ArcGIS Pro to get started.

a. If you left ArcGIS Pro open at the end of the last exercise, browse to and open the Sec2Ex2_ExploringDataClassification.ppkx file.

Note: If you closed ArcGIS Pro at the end of the last exercise, start ArcGIS Pro again and, if necessary, sign in using the credentials provided at the start of this course (username includes _cart). Then browse to and open the Sec2Ex2_ExploringDataClassification.ppkx file.



The project opens to a United Kingdom Election 2015 election map, which includes a single layer showing the electoral constituencies for the United Kingdom of Great Britain and Northern Ireland, and a Dark Gray Canvas basemap layer for geographical context (which is not turned on).

The project also contains numerical election data for each of the constituencies. You will classify this data using a variety of methods to visualize the results of the 2015 UK election.

b From the Project tab, click Save As, and type a name for your map, such as **Sec2Ex2_ExploringDataClassification_<yourfirstandlastname>**.

Note: Saving your work regularly is important when working in ArcGIS Pro. Remember to save periodically as you work through this exercise.

Step 3: Determine the spatial reference of the data

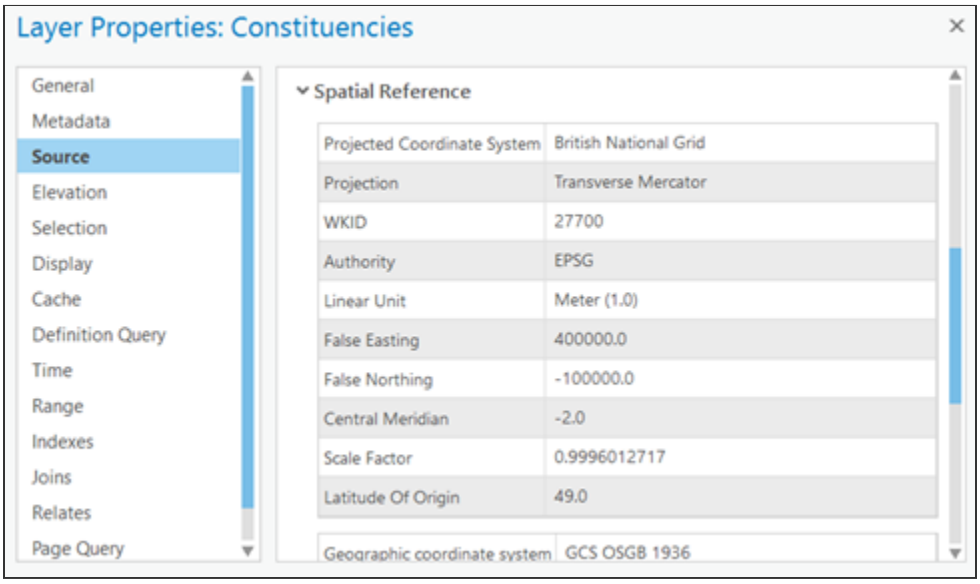
Before classifying the data, you should examine its spatial reference to gain an awareness of what coordinate system and projection that the data is specified in. This often helps in later steps when it can be an important part of analysis steps or cartography.

a In the Contents pane, double-click the Constituencies layer to open the Layer Properties dialog box.

Note: You can also right-click the layer name and choose Properties.

b From the Source tab, view the data source information for the layer.

c Scroll down, if necessary, and click Spatial Reference to expand that section.



The spatial reference uses a British National Grid projected coordinate system based on a Transverse Mercator projection. This option is the most common coordinate system and projection used for data of the UK. If you looked at a different part of the world, a different coordinate system and projection would likely be used that would be more relevant to that specific area.

d Close the Layer Properties dialog box.

Next, you will turn on the Dark Gray Canvas basemap layer to provide geographical context for the map.

e In the Contents pane, turn on the Dark Gray Canvas layer, and then zoom out to see all of the United Kingdom.

Now you can see how the projected coordinate system that the UK uses for its mapping warps the rest of the world map.

Step 4: Explore the data

To ensure that the message of your map is clear and accurate, you must understand the data before classifying it.

a Open the Constituencies layer attribute table.

Hint: Right-click the Constituencies layer > Attribute Table.

b Examine the field names in the attribute table, and use the field descriptions in the following table to learn more about the data.

Note: You can dock the attribute table pane in different parts of your window or make it larger or smaller by clicking and dragging the border.

All you're doing here is getting familiar with the data, which is important before working with it or making a map.

Field name	Description
ID	Alphanumeric code used to identify individual constituencies.
Constituency	The name of the electoral constituency.
Headline	The result of the election showing which political party won and by how much.
Winning_MP	The name of the winning member of Parliament who represents the winning political party.
First	Abbreviated name of the winning political party in that constituency.
First_Votes	The number of votes for the winning political party in that constituency.
First_Share	The overall share of votes for the winning political party in that constituency (largest share wins).
Second	Abbreviated name of the political party that came second in that constituency.
Second_Votes	The number of votes for the political party that came second in that constituency.
Second_Share	The overall share of votes for the party that came second in that constituency.
Third	Abbreviated name of the political party that came third in that constituency.
Third_Votes	The number of votes for the political party that came third in that constituency.
Third_Share	The overall share of votes for the party that came third in that constituency.
Other_Votes	The number of votes for all other political parties combined in that constituency.
Other_Share	The overall share of votes for all other political parties combined in that constituency.
Turnout	The number of people who actually voted in the constituency.
Electorate	The total number of eligible voters in the constituency.

c After you have examined the data, close the attribute table.

Now that you have an idea of the data available in the layer, you will create several choropleth maps to see how changing the data classification changes the message of the map.

Step 5: Create a map using natural breaks

You will make a simple choropleth (graduated colors ([https://bit.ly/2HrlJwi]) map to show the voter-turnout percentage as a proportion of the total electorate in each constituency.

a In the Contents pane, right-click the Constituencies layer and choose Copy.

b Right-click the UK Election 2015 map view item and choose Paste.

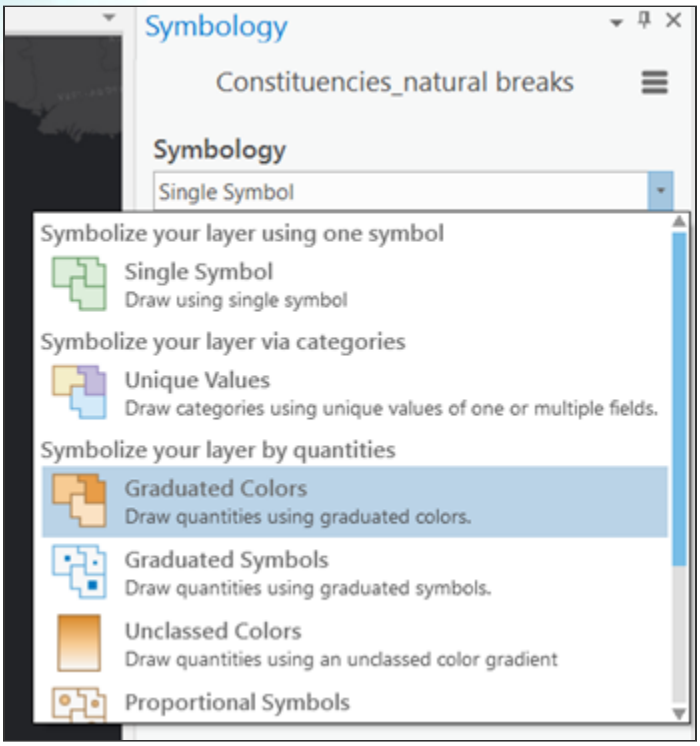
c Double-click the layer name to open the Layer Properties dialog box. On the General tab, in the Name field, type **Constituencies_natural breaks** as the name for the new layer.

d Open the Symbology pane for the Constituencies_natural breaks layer.

Hint: With the new layer selected, click the Appearance tab and choose Symbology. Alternatively, in the Contents pane, you can right-click the layer name and choose Symbology. If the Symbology pane is not docked, right-click the pane title and click Dock.

In the Symbology pane, the Symbology drop-down list provides several options for symbolizing the layer. You will first draw quantities using graduated colors, with darker colors indicating higher percentages.

Graduated-color symbology shows a quantitative difference between mapped features by varying the color of symbols. Data is classified into ranges that are each assigned a different color from a color scheme to represent the range. For instance, if your classification scheme has five classes, five different symbol colors are assigned. Maps that vary in color this way are called choropleth maps and work by applying different shades of the same color so that lighter shades match lower data values and darker shades match higher data values.



e. For Symbology, choose Graduated Colors.

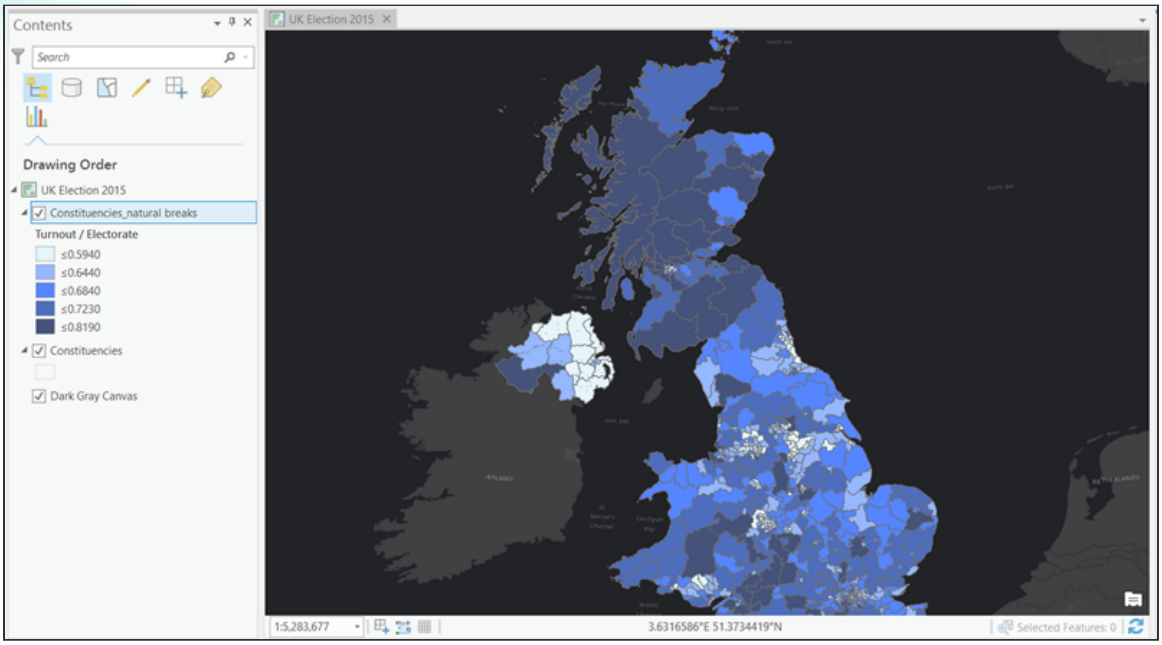
f. For Field, choose Turnout.

Recall that the Turnout field contains the number of people who actually voted in the constituency.

Choosing Turnout for the field will display the data totals. However, you should not map totals with a choropleth map because the areas and base population data differ between areas, creating problems when comparing areas across the map. Put simply, you have to mitigate for underlying geographical and population differences to create a final map that supports visual comparison like-for-like across the whole map. The data has to be normalized either by calculating a new field or by selecting a field to normalize against. Normalizing divides each data value by a common denominator so that you turn totals into rates. Mapping rates results in a map that supports visual comparisons between areas.

g. For Normalization, choose Electorate.

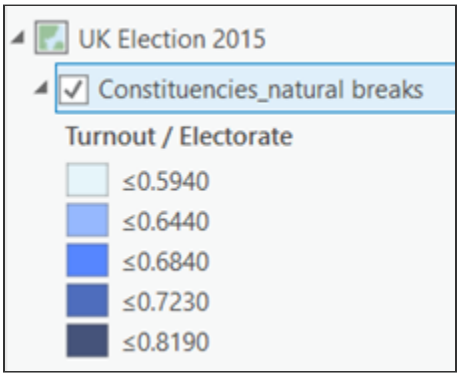
Because you want to show the percentage of voter turnout as a proportion of the total electorate, you normalize against the Electorate field, which contains the total number of eligible voters in the constituency.



Choropleth map symbolized with graduated colors, using the Natural Breaks (Jenks) data classification method.

The default classification method, Natural Breaks (Jenks) (https://bit.ly/2GQSiCL), and the number of classes are applied.

Now you have a default choropleth map showing the percentage turnout classified into five classes by the natural breaks method. Lower turnout is shown with lighter symbols, and larger turnout is shown with darker symbols, as you can see in the legend.

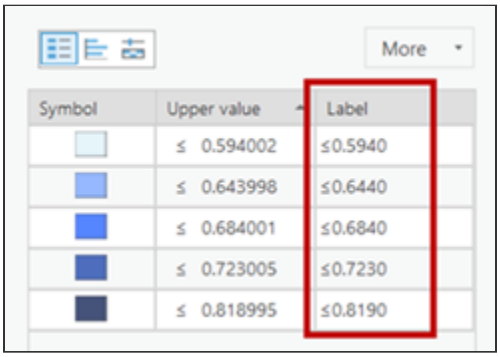


Looking at the map and the legend, what patterns do you see? What message does the classification provide? For all these different classification methods, you'll likely see the relative areas of high rates relative to low rates. However, there is always a subtle nuance. So, for instance, does the technique make it easy to see the highest and lowest areas, or do the areas tend to become grouped with other areas? How does the distribution of data appear across the whole map? Are there sharp changes between some areas that need picking out,

or are the changes more gradual across space? Keep these and other questions in mind as you change the classification methods because changing the method does change the map and the map's message.

The natural breaks (Jenks) classification method uses classes based on natural groupings inherent in the data. Class breaks that best group similar values and maximize the differences between classes are identified. The features are divided into classes whose boundaries are set where there are relatively big differences in the data values.

h. At the bottom of the Symbology pane, notice the labels associated with the symbols.

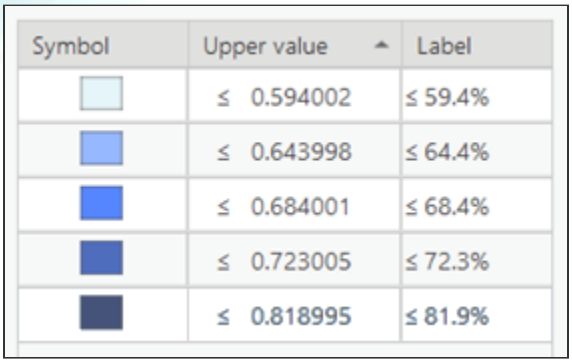


The labels show the proportion of turnout as a value between 0 and 1. A percentage would be more meaningful. You can change the labels to show a percentage by clicking the label and typing a new value and a percent symbol (for example, 0.5940 becomes 59.4%). There are different ways we might label each class such as showing a class interval (0 - 59.4; 50.5 - 64.4 and so on), but for now we'll just make a simple change to the default.

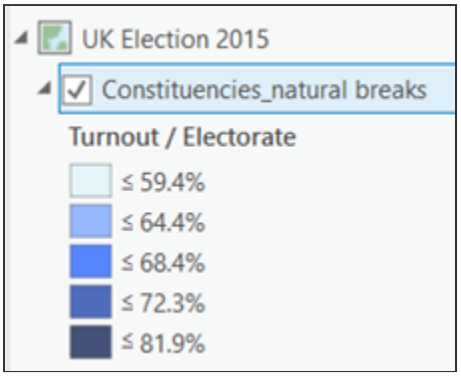
i. Update the labels for the symbols to percentages using the following values:

- Change 0.594002 to **59.4%**
- Change 0.643998 to **64.4%**
- Change 0.684001 to **68.4%**
- Change 0.723005 to **72.3%**
- Change 0.818995 to **81.9%**

Note: Change only the value.



j In the Contents pane, examine the map legend and notice how the percentages make the data and symbology more meaningful.



You can view your data classification in different ways by using the different view modes in the Symbology pane when drawing with graduated color symbology.

View mode	Description
Label View	Allows you to manage the symbol, values, descriptive labels, and grouping of the symbol classes.
Histogram View	Offers a visual tool for editing the classes and understanding how the data is represented by different classification methods.
Scale Range View	Allows you to specify the scale ranges in which each symbol class draws (though this tool is not particularly useful for this exercise).

k In the Symbology pane, use the Label View button and the Histogram View button to switch between a label view and a histogram view.

Ensure that you check the histogram view for this map and all subsequent classification maps to see how the data distribution relates to the different classification schemes.

l. Experiment with changing the number of classes by changing the classes drop-down in the Symbology pane.

Do fewer classes help simplify and clarify the message of the map? Do more classes give a different message? A good rule of thumb is to limit the number of classes to between four and seven. Too few and you lose a great deal of variation in how the data is presented. Too many and it becomes hard to see subtle differences between areas classified slightly differently.

Note: For more information about data classification, refer to The Basics Of Data Classification (https://bit.ly/2aMWAzN).

Now you will explore different classification techniques and see their effect on the way that the data appears on the map. Remember that you can pan and zoom the map as you do this. You can see how the legend updates in the Contents pane as you change settings and they are automatically applied to the map.

Note: Refer to ArcGIS Pro Help for more information about data classification methods (https: //bit.ly/2ql2WeF).

As you make changes, consider the visual impression of the pattern of data and its distribution. Ask yourself what each method shows and what the key aspect is of the highlighted data. Sometimes differences will be quite pronounced, sometimes subtle. All have consequences for how people read the pattern and, subsequently, what they interpret.

Step 6: Create a map using a quantile classification

In a quantile classification, each class contains an equal number of features. A quantile classification is well-suited to linearly distributed data. Quantile classification assigns the same number of data values to each class, resulting in no empty classes or classes with too few or too many values.

Because features are grouped in equal numbers in each class, the resulting map won't necessarily reflect the underlying data distribution. Similar features can be placed in adjacent classes, and features with widely different values can be put in the same class. You can minimize this distortion by increasing the number of classes though, of course, that introduces other trade-offs.

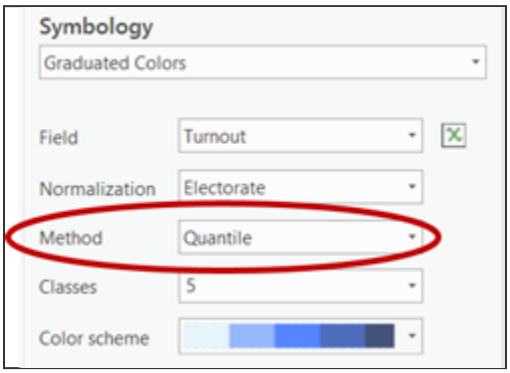
a. In the Contents pane, right-click the Constituencies_natural breaks layer and choose Copy.

b. Right-click the UK Election 2015 map view item and choose Paste.

c. Name the new layer **Constituencies_quantile**.

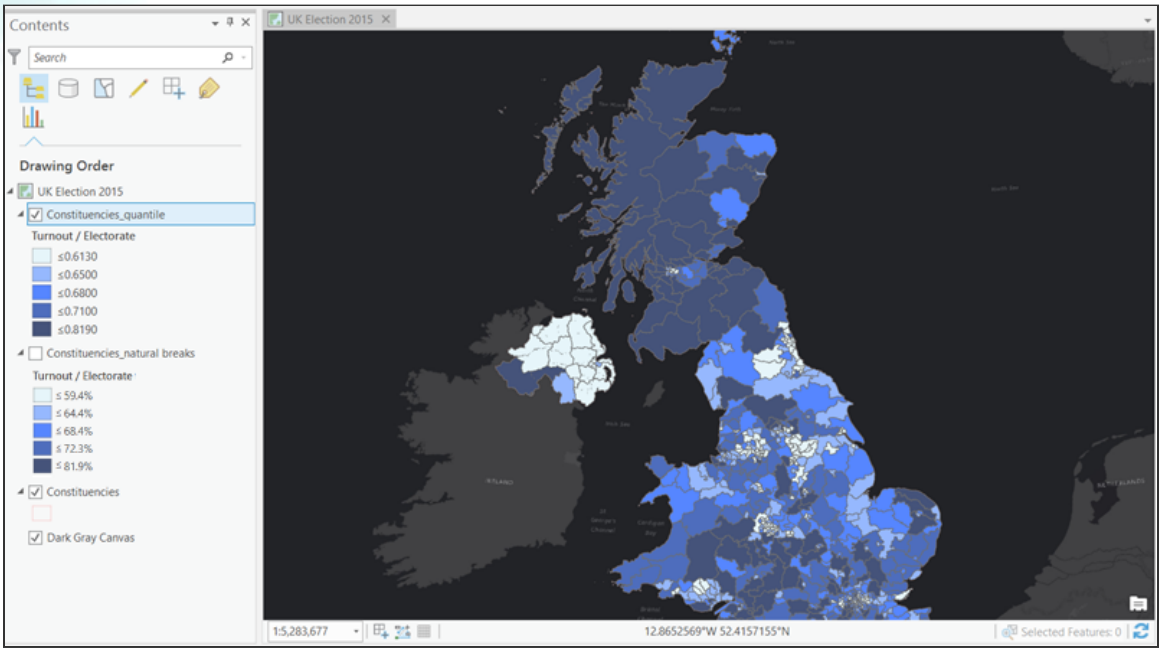
d. Turn off the Constituencies_natural breaks layer.

e. In the Symbology pane for the Constituencies_quantile layer, change the Method to Quantile.



What differences or similarities do you notice with the change in classification method? What patterns are visible? What is the key aspect of the data that is highlighted with this classification method? Remember, you can vary the number of classes to see how things change.

Hint: You can turn the natural breaks layer and this new quantile layer on and off in the Contents pane to see the difference between your first and second map.



Choropleth map symbolized with graduated colors, using the quantile data classification method.

f Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

Step 7: Create a map using an equal interval classification

The equal interval classification (https://bit.ly/2Hu3uq0) divides the range of data values into equal-sized subranges. ArcGIS Pro will automatically determine the class breaks based on the value range. For example, if you specify three classes for a field whose values range from 0 to 300, the application will create three classes with ranges of 0-100, 101-200, and 201-300.

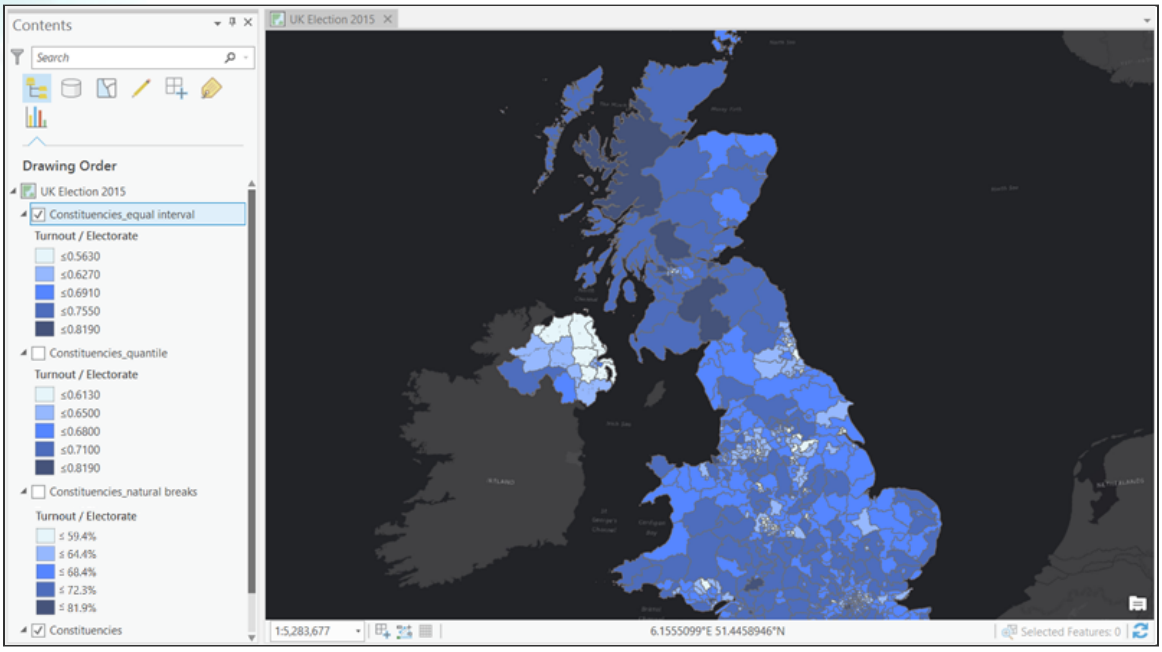
Equal interval is best applied to familiar data ranges, such as percentages and temperature. This method emphasizes the amount of an attribute value relative to other well-known values.

a In the Contents pane, copy the Constituencies_natural breaks layer and then paste another instance into the Contents pane.

b Name the new layer **Constituencies_equal interval**.

c Turn off the Constituencies_quantile layer.

d In the Symbology pane, change the Method to Equal Interval.



Choropleth map symbolized with graduated colors, using the equal interval data classification method.

What is the key aspect of the data that is highlighted with this classification method? Remember, you can vary the number of classes to see how things change.

e Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

Step 8: Create a map using a geometric interval classification

The geometrical interval classification scheme creates class breaks based on class intervals that have a geometric series. The geometric coefficient in this classifier can change once (to its inverse) to optimize the class ranges. The algorithm creates geometrical intervals by minimizing the sum of squares of the number of elements in each class. This ensures that each class range has approximately the same number of values in each class and that the change between intervals is fairly consistent.

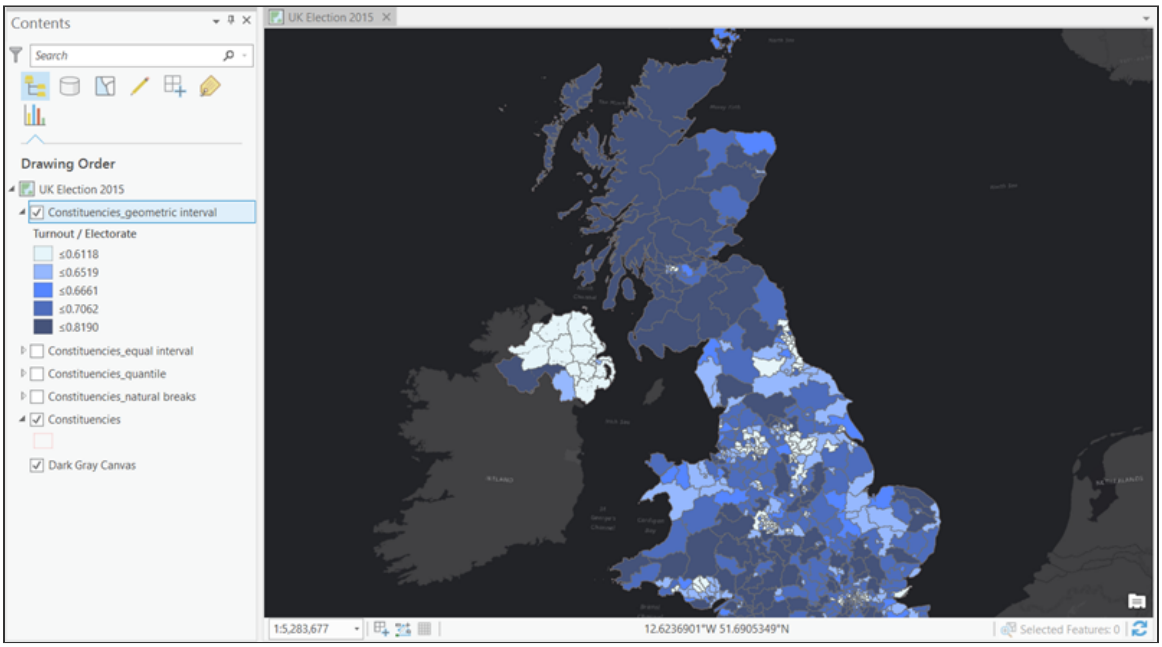
This algorithm was specifically designed to accommodate continuous data. It is a compromise method between equal interval, natural breaks (Jenks), and quantile. It creates a balance between highlighting changes in the middle values and the extreme values, producing a result that can be more relevant.

a In the Contents pane, copy the Constituencies_natural breaks layer and then paste another instance into the Contents pane.

b. Name the new layer **Constituencies_geometric interval**.

c. Turn off the Constituencies_equal interval layer.

d. In the Symbology pane, change the Method to Geometric Interval.



Choropleth map symbolized with graduated colors, using the geometric interval data classification method.

What is the key aspect of the data that is highlighted with this classification method? Remember, you can vary the number of classes to see how things change.

e. Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

Step 9: Create a map using a standard deviation classification

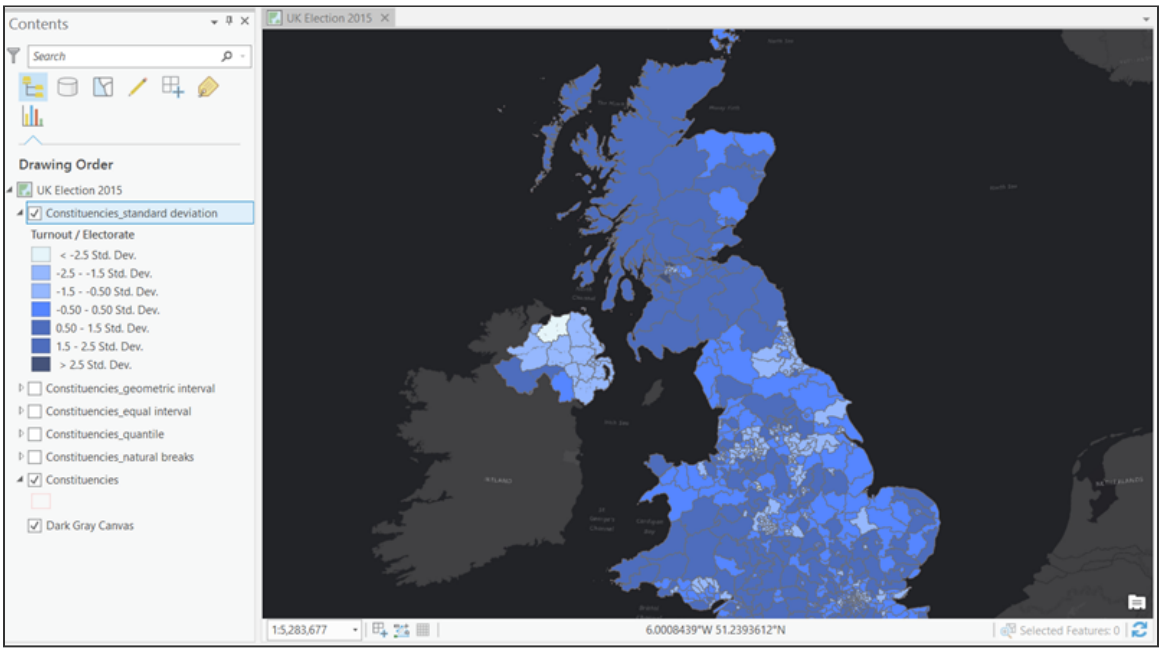
The standard deviation classification method shows you how much a feature's attribute value varies from the mean. ArcGIS Pro calculates the mean and standard deviation automatically. Class breaks are created with equal value ranges that are a proportion of the standard deviation, usually at intervals of one, one-half, one-third, or one-fourth standard deviations using mean values and the standard deviations from the mean.

a. In the Contents pane, copy the Constituencies_natural breaks layer, and then paste another instance into the Contents pane.

b. Name the new layer **Constituencies_standard deviation**.

c. Turn off the Constituencies_geometric interval layer.

d. In the Symbology pane, change the Method to Standard Deviation.



Choropleth map symbolized with graduated colors, using the standard deviation data classification method.

What is the key aspect of the data that is highlighted with this classification method? Remember, you can vary the number of classes to see how things change.

e. Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

Step 10: Create a map using a manual interval classification

Now that you have explored the various standard classification schemes available in ArcGIS Pro, you will make a manual classification scheme.

If you want to define your own classes, you can manually add class breaks and set class ranges that are appropriate for your data. You can also start with one of the standard classifications and make adjustments as needed.

First, you will make a copy of the Constituencies_natural breaks layer to use for your manual classification scheme.

a. In the Contents pane, copy the Constituencies_natural breaks layer and then paste another instance into the Contents pane.

b. Name the new layer **Constituencies_manual interval**.

c. Turn off the Constituencies_standard deviation layer.

d. Ensure that the Constituencies_manual interval layer is selected.

e. In the Symbology pane, update Method to Manual Interval.

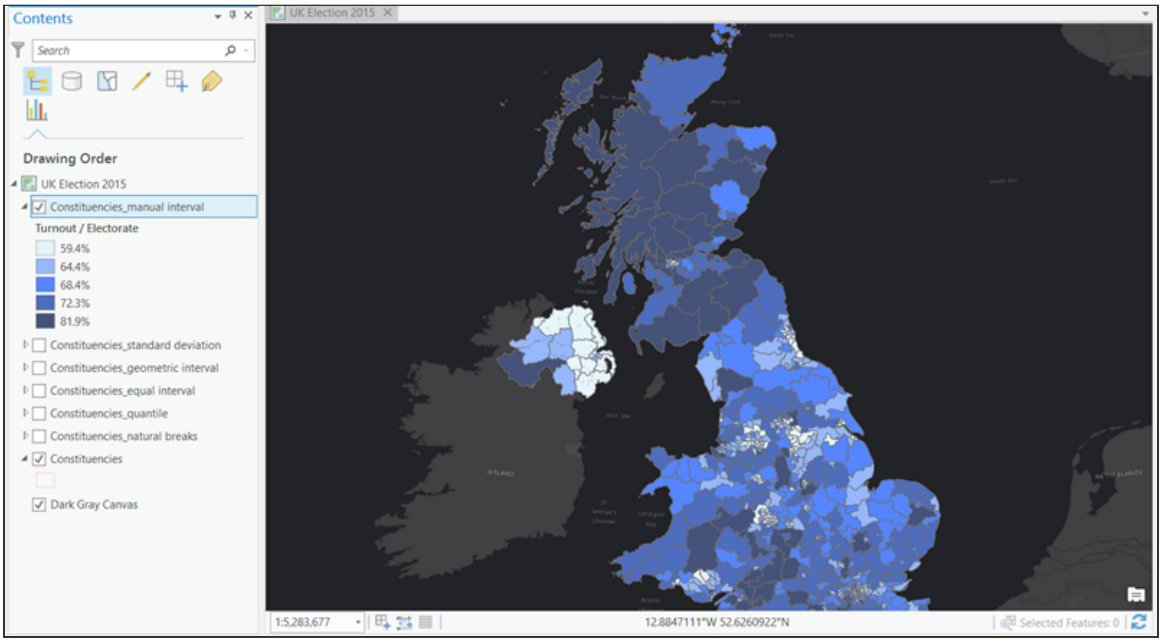
It's useful to use the histogram view of the data when creating a manual classification.

f. In the Symbology pane, use the Histogram View button to switch to histogram view.

g. Move the sliders up and down across the histogram so that you attempt to group together values of data that display similar characteristics.

You might also incorporate a specific value or values as part of your classification scheme.

h. Experiment until you achieve a classification scheme that makes sense and which makes the map make sense to you, based on the sort of characteristics in the data that you want to emphasize.



Choropleth map symbolized with graduated colors, using the manual interval data classification method.

What is the key aspect of the data that is highlighted with this classification method?

i. Examine your manual breaks method compared to the other methods to see how the different classification schemes change the way the map looks.

Step 11: Create a map using an unclassed representation

Now that you have seen how different standard classification methods affect the message of the map, you will learn another approach to represent the data. The alternative is to simply not classify the data at all and apply a unique shade of color to each individual data value in your array. This is called an unclassed technique.

a. In the Contents pane, paste another copy of the Constituencies_natural breaks layer.

b. Name the new layer **Constituencies_unclassed**.

c. Turn off the Constituencies_manual interval layer.

d. In the Symbology pane, from the Symbology drop-down list, choose Unclassed Colors.

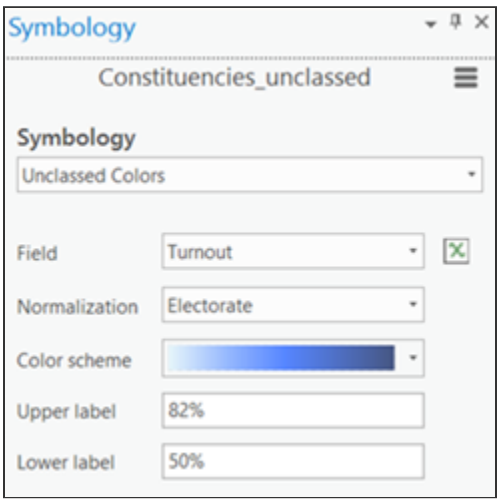
e. For Field, choose Turnout.

f. For Normalization, choose Electorate.

To control how the legend for this layer is labeled, you will edit the upper and lower labels. By default, they are the minimum and maximum values of the symbol field.

g. Update the following label fields with the provided percentages:

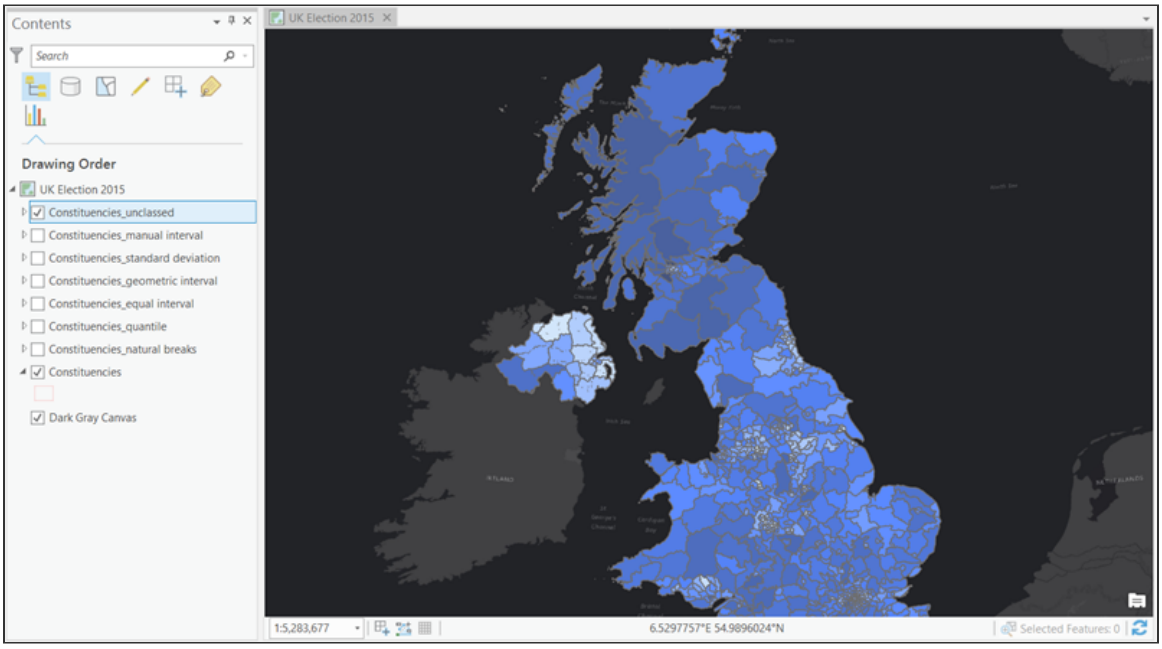
- Upper Label: **82%**
- Lower Label: **50%**



Dragging the handles in the histogram shifts the way that color is applied to the unclassified values.

h Experiment with these handles and see how the colors are applied to the map.

This is a quick way to change how all data values above or below a particular number can be made to appear the same way.



Choropleth map symbolized with unclassed colors.

What is the key aspect of the data that is highlighted with this classification method?

Unclassed techniques can be very useful in highlighting extremes in your data and picking out the highs and lows. They are, however, the inverse of classifying, so bear this in mind when you want to make a map that shows your reader how similar different places are.

i When you are finished, save your project and exit ArcGIS Pro.

Conclusion

You have explored a range of classification techniques and learned that there are an almost infinite number of alternatives to classify a single dataset. Therefore, there is ample opportunity to represent the data well or poorly, objectively or in a persuasive manner.

For any dataset that you might classify, determine which method of classification that you think is the most effective at getting your message across without obfuscating the truth.

You'll work with this same dataset in a future exercise to create a range of different thematic maps and begin to consider different colors and the symbology in more detail. However, for now, appreciating how data can be manipulated to show it differently is foundational.

Stretch Goal

You can continue your exploration of classifying numerical data with the following stretch goal.

- Try classifying some of the other data in the UK election dataset and gain some experience in seeing how different classification techniques can be used to tease out aspects of different data.

Note: Any fields with "_Share" in their names are already percentages and do not need to be normalized.

Use the Lesson Forum to post your questions, observations, and map examples.