Exercise

Classifying Data

Section 2 Exercise 3

04/2020

esri

THE SCIENCE OF WHERE™

Classifying Data

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 screen shots that you see in the course materials.

Time to complete

Approximately 20-30 minutes

Software requirements

ArcGIS Pro 2.5
ArcGIS Pro Standard license (or higher)

Note: The MOOC provides a separate ArcGIS account (user name and password) that you will need to use to license ArcGIS Pro and access other software applications used throughout the MOOC exercises. This account (user name ending with _cart) provides the appropriate ArcGIS Online role, ArcGIS Pro license, ArcGIS Pro extensions, and credits. We strongly recommend that you use the provided course ArcGIS account to ensure that you have the appropriate licensing to complete the exercises. Exercises may require credits. Using the provided course ArcGIS account ensures that you do not consume your organization's credits. Esri is not responsible for any credits consumed if you use a different account. Moreover, Esri will not provide technical support to students who use a different account.

Introduction

Now that you have explored the data that you will be working with, you can consider how to classify it. 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 underlined choropleth maps (https://bit.ly/33F7a3b) to illustrate how changing the classification changes the map.

Step 1: Create a map using a natural breaks classification

a   If necessary, start ArcGIS Pro and open the ExploringDataClassification_<your first and last name>.aprx project file that you saved in the previous exercise.

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

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

c Right-click the UK Election 2015 map and choose Paste.
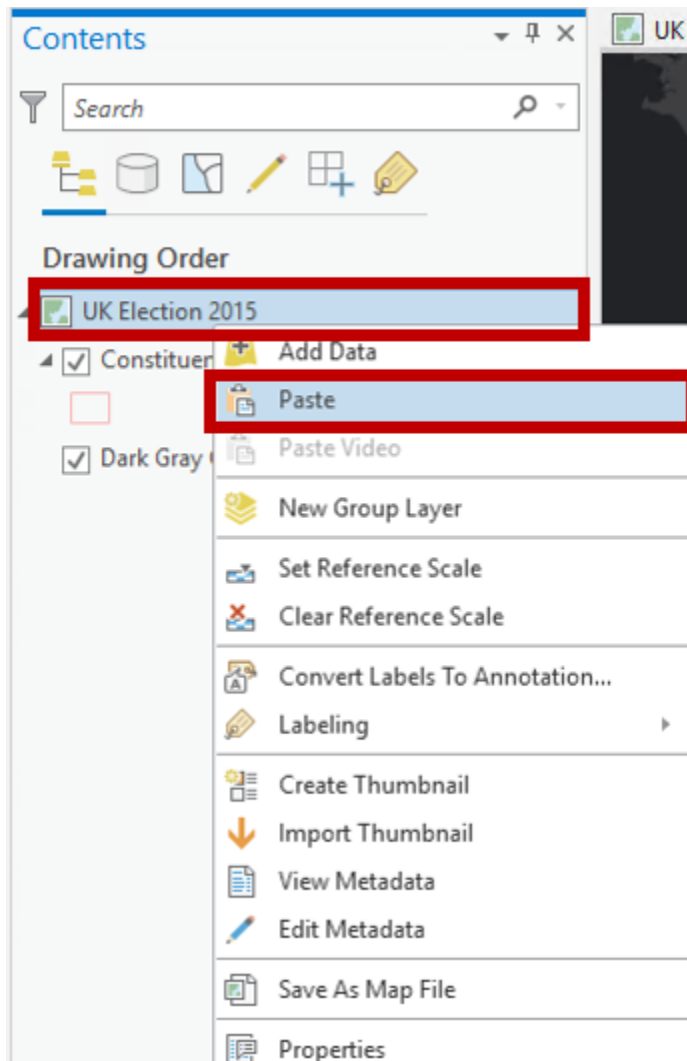
d Double-click the new layer name to open the Layer Properties dialog box.

e Click General, and then in the Name field, type **Constituencies_natural breaks** as the name for the new layer and click OK.

**f** In the Contents pane, turn off the Constituencies layer.

g Open the Symbology pane for the Constituencies_natural breaks layer.

Hint: With the new layer selected in the Contents pane, click the Appearance tab, and then in the Drawing group, click 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 choose Dock.

In the Symbology pane, the Primary 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 (https://bit.ly/2req3YA) that represents the range. For instance, if your classification scheme has five classes, five different symbol colors are assigned. Maps that vary in color in this way are called choropleth maps. Choropleth maps apply different shades of the same color so that lighter shades match lower data values and darker shades match higher data values.

**h** For Primary Symbology, choose Graduated Colors.
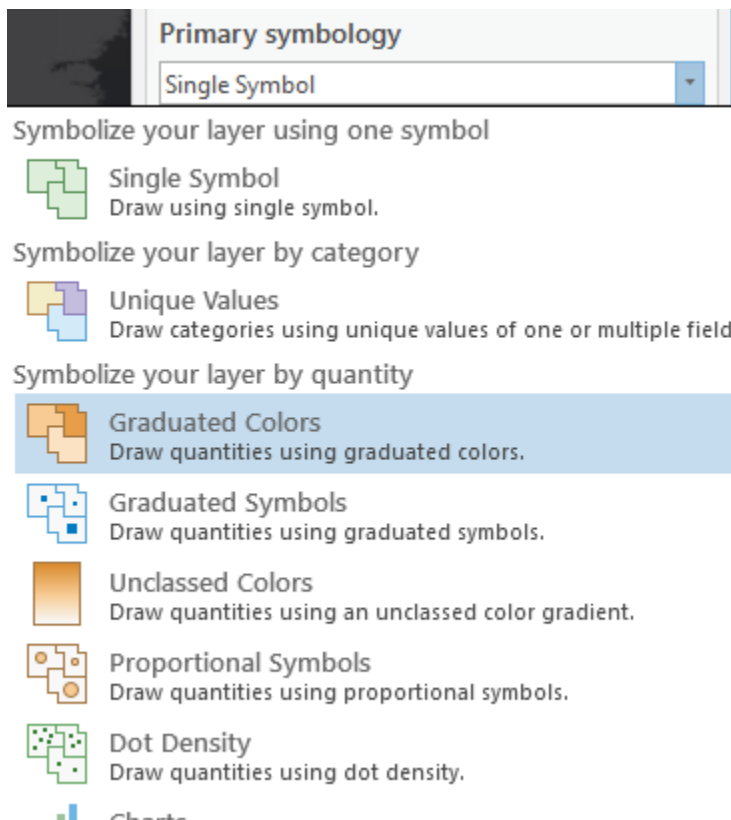
i For Field, choose Turnout.

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, which makes comparing areas across the map problematic. To create a final map that supports visual comparison like-for-like across the whole map, you must first account for underlying geographical and population differences. You can achieve this by normalizing the data, either by calculating a new field or by selecting a field to normalize against. Normalizing divides each data value by a common denominator, turning totals into rates. Mapping rates will result in a map that supports visual comparisons between areas.

**j** 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.
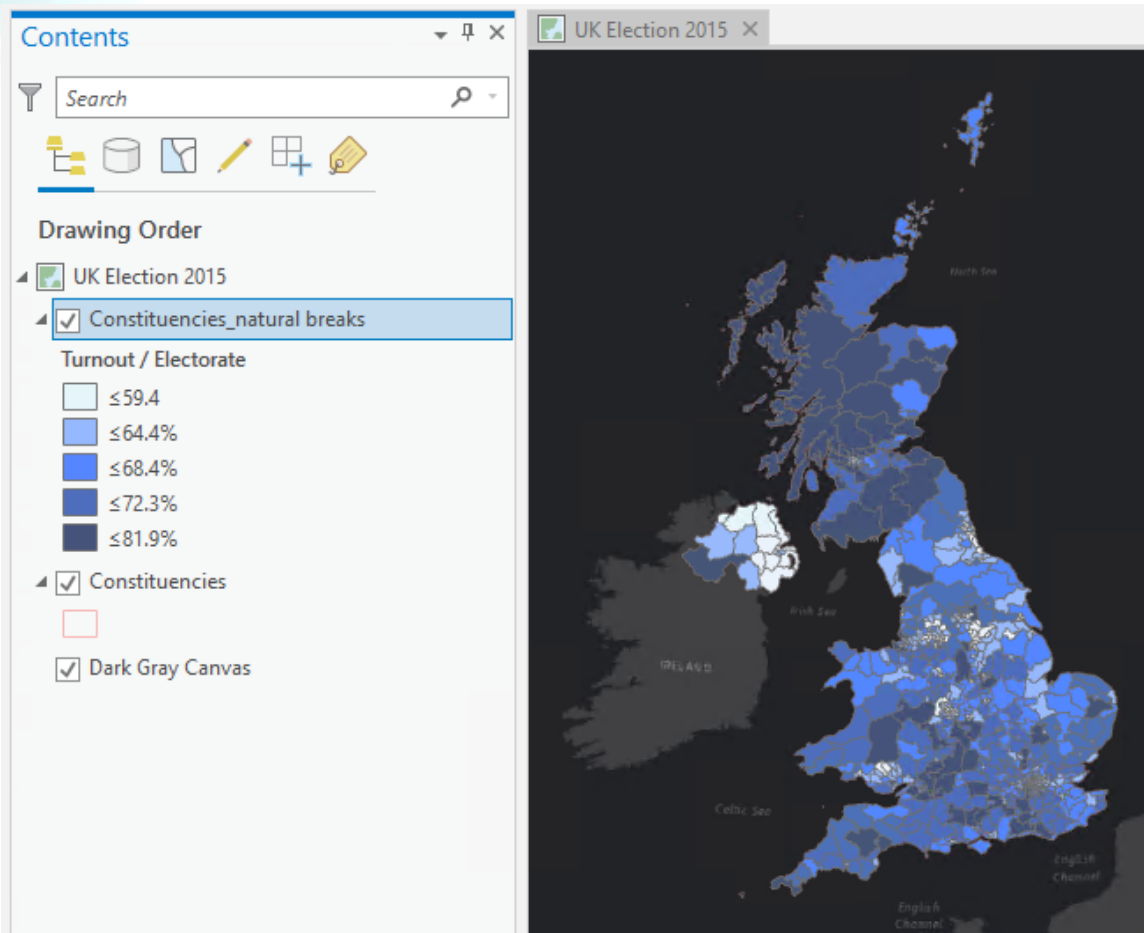
Choropleth map symbolized with graduated colors, using the natural breaks (Jenks) data classification method.

The default classification method, Natural Breaks (Jenks) (https://bit.ly/2UefL9U), 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 higher turnout is shown with darker symbols, as you can see in the legend.
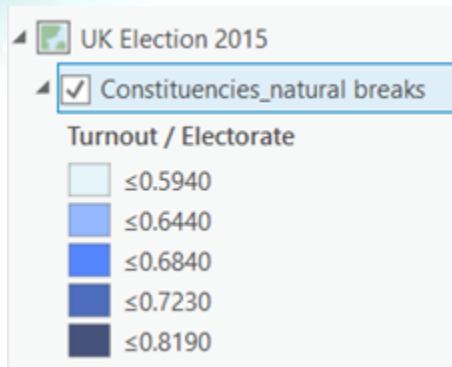
Looking at the map and the legend, what patterns do you see? What message does the classification communicate? With all the different classification methods, you will likely see areas of high rates relative to low rates. However, there are always subtle differences in the resulting maps. So, for instance, does a particular technique make it easy to see the highest and lowest areas, or does it cause these areas to become grouped with other areas? How is the data distributed across the whole map? Are there sharp changes between some areas that need to be examined further, 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 the natural groupings inherent in the data. This method identifies the class breaks that best group similar values and maximize the differences between classes. It divides the features into classes whose boundaries are set where there are relatively big differences in the data values.

k Near 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. However, a percentage would be more meaningful. You can change the labels to show a percentage by clicking each label and typing a new value and the percent symbol (for example, 0.5940

becomes 59.4%). You could label each class in many different ways, including showing class intervals (0–59.4, 50.5–64.4, and so on), but for now, you will just make a simple change to the default.

l Update the labels for the symbols to percentages using the following values:

Note: Change only the number to a percentage. Do not change the less-than/equal symbol.

- Change ≤ 0.5940 to ≤ **59.4%**
- Change ≤ 0.6440 to ≤ **64.4%**
- Change ≤ 0.6840 to ≤ **68.4%**
- Change ≤ 0.7230 to ≤ **72.3%**
- Change ≤ 0.8190 to ≤ **81.9%**

Classes	Histogram	Scales	

			More ▼

Symbol		Upper value	Label
☐	▼	≤ 0.594002	≤ 59.4%
☐	▼	≤ 0.643998	≤ 64.4%
◼	▼	≤ 0.684001	≤ 68.4%
◼	▼	≤ 0.723005	≤ 72.3%
◼	▼	≤ 0.818995	≤ 81.9%

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

▲ ☑ UK Election 2015

 ▲ ☑ Constituencies_natural breaks

Turnout / Electorate

☐ ≤ 59.4%
◼ ≤ 64.4%
◼ ≤ 68.4%
◼ ≤ 72.3%
◼ ≤ 81.9%
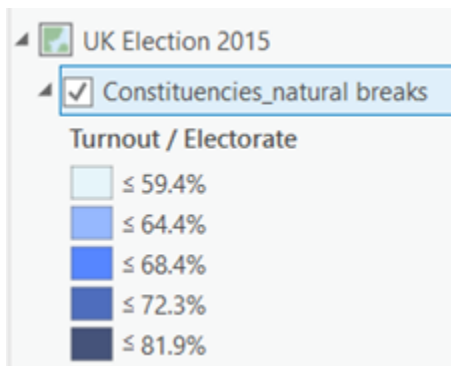
n Save your project.

When using graduated color symbology, you can use the different view modes in the Symbology pane to see your data classification in different ways.

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

o In the Symbology pane, click the Classes tab and the Histogram tab 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.

p Experiment with changing the number of classes by changing the Classes parameter 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 how they affect the way that the data appears on the map. Remember that you can pan and zoom the map as you do this. As you change settings, you will notice that the legend in the Contents pane updates and the changes 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 data's pattern and distribution. Ask yourself what each method shows and what the key aspect of the highlighted data is. Differences will sometimes be pronounced and, other times, subtle. All have consequences for how people read and interpret the pattern.

q Save your project.

Step 2: 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 in classes with too few or too many values.

Because features are grouped in equal numbers in each class, the resulting map will not necessarily reflect the underlying data distribution. Similar features might be placed in adjacent classes, and features with widely different values might be put in the same class. You can minimize this distortion by increasing the number of classes, although this 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 and choose Paste.

c Name the new layer **Constituencies_quantile**.

d Turn off the Constituencies_natural breaks layer and turn on the Constituencies_quantile layer, if necessary.

e Verify that the Constituencies_quantile layer is selected.

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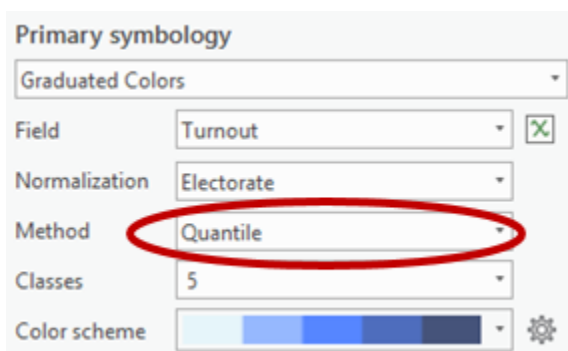
What differences or similarities do you notice with the change in classification method? What patterns are visible? What is the key aspect of the data that is highlighted with this classification method? Remember, you can vary the number of classes to see how that changes things.

Hint: In the Contents pane, turn the Constituencies_natural breaks and Constituencies_quantile layers on and off to see the differences between the maps.
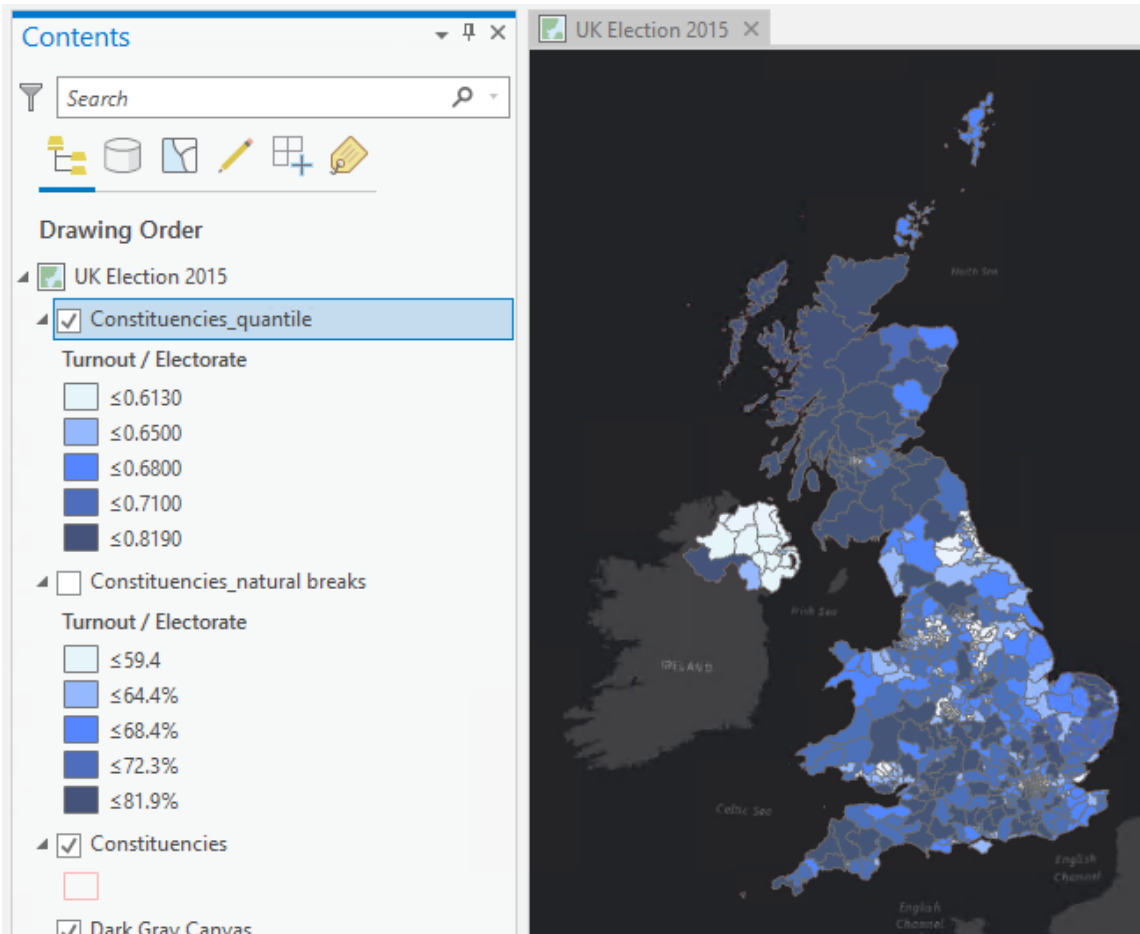
Choropleth map symbolized with graduated colors, using the quantile data classification method.

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

h Save your project.

Step 3: Create a map using an equal interval classification

Equal interval classification (https://bit.ly/2Jac2Ut) 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 paste it to the UK Election 2015 map.

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

c Turn off the Constituencies_quantile layer and turn on the Constituencies_equal interval layer, if necessary.

d Verify that the Constituencies_equal interval layer is selected.

e In the Symbology pane, change 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.
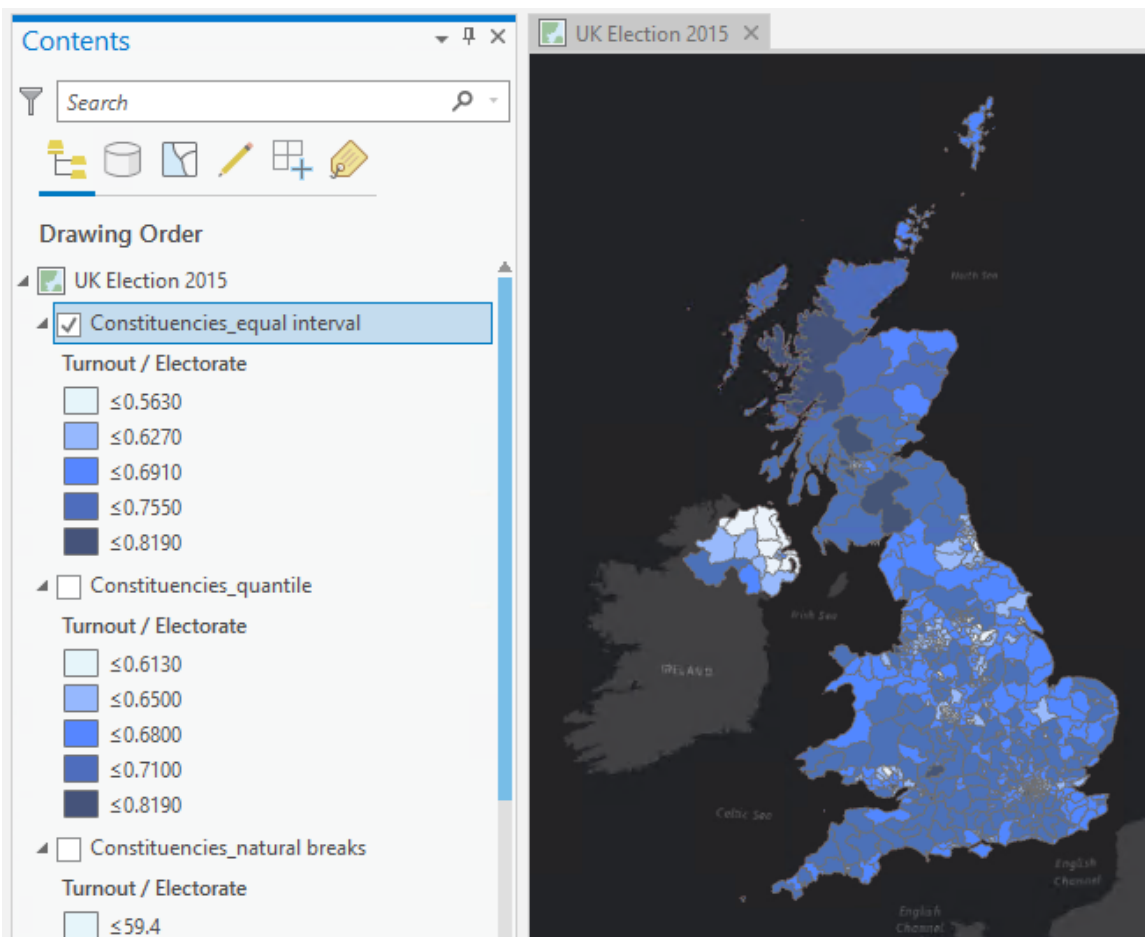
f Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

g Save your project.

You have explored three of the most common classification methods, but there are many more. We encourage you to explore other methods on your own.

Step 4: Create a map using a geometric interval classification

The geometric 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 paste it to the UK Election 2015 map.

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

c Turn off the Constituencies_equal interval layer and turn on the Constituencies_geometric interval layer, if necessary.

d Verify that the Constituencies_geometric interval layer is selected.

e In the Symbology pane, change the Method to Geometric Interval.
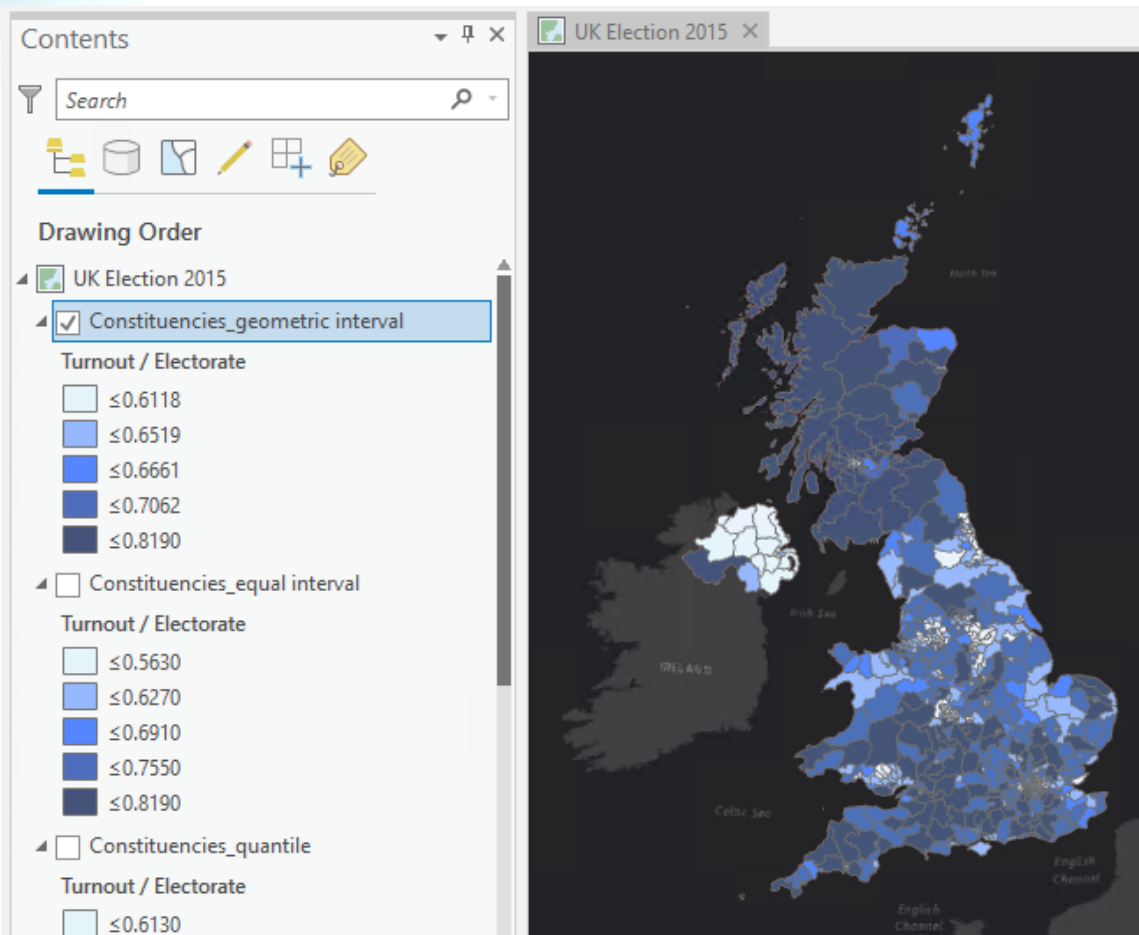
Choropleth map symbolized with graduated colors, using the geometric interval data classification method.

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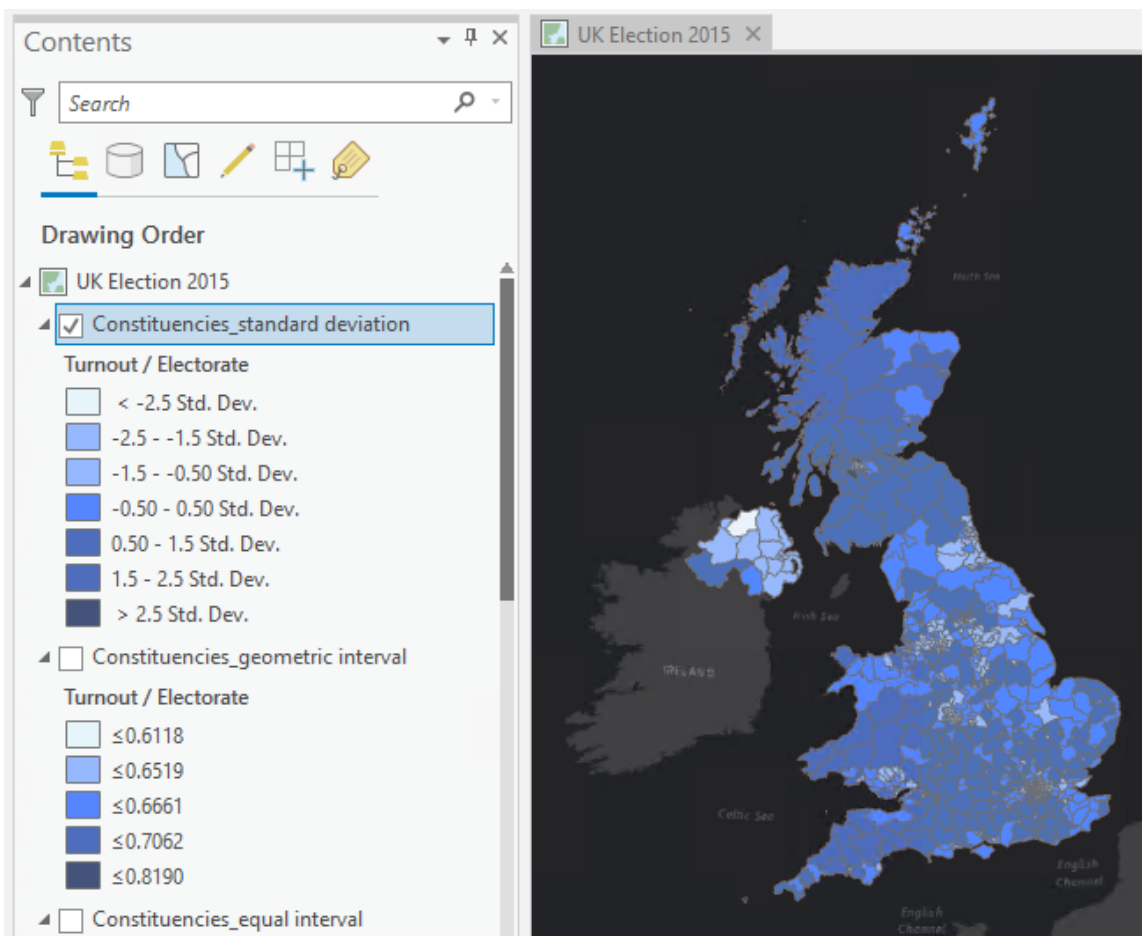
f Examine the histogram view for this map to see how the classification method has been applied across the data distribution.

g Save your project.

Step 5: 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 paste it to the UK Election 2015 map.

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

c. Turn off the Constituencies_geometric interval layer and turn on the Constituencies_standard deviation layer, if necessary.

d. Verify that the Constituencies_standard deviation layer is selected.

e. 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.

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

g Save your project.

Step 6: Create a map using 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 paste it to the UK Election 2015 map.

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

c Turn off the Constituencies_standard deviation layer and turn on the Constituencies_manual interval layer, if necessary.

d Verify that the Constituencies_manual interval layer is selected.

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

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

f In the Symbology pane, click the Histogram tab to switch to the 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 to you, based on the sort of characteristics in the data that you want to emphasize.
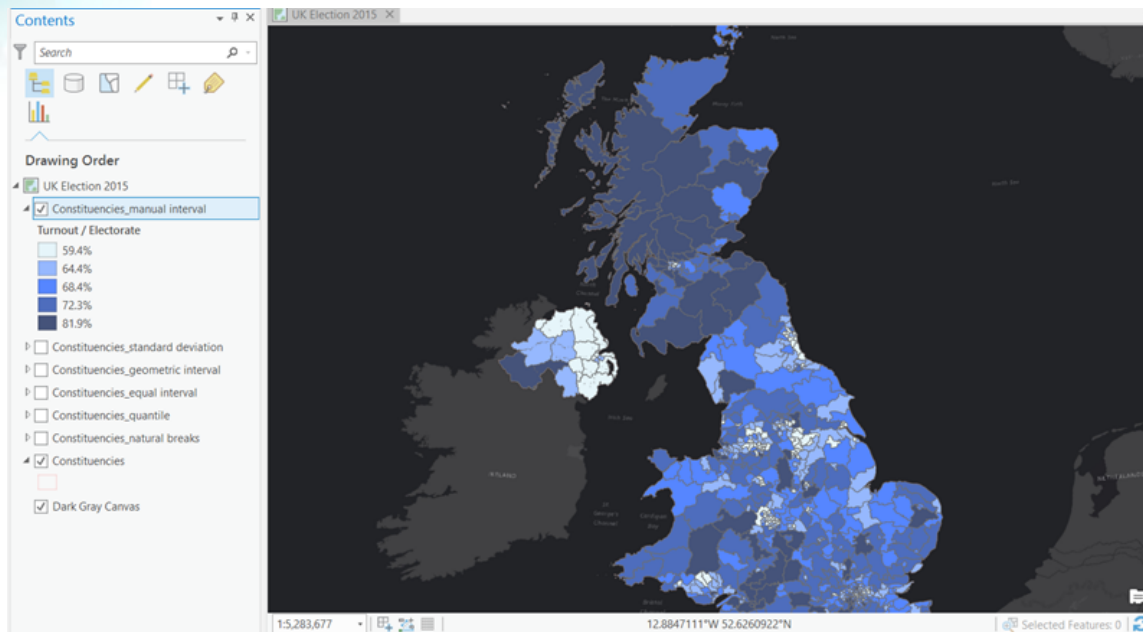
Choropleth map symbolized with graduated colors, using the manual interval data classification method.

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.

j Save your project.

## Step 7: 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 and turn on the Constituencies_unclassed layer, if necessary.

d In the Symbology pane, from the Primary 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: **50%**
- Lower Label: **82%**
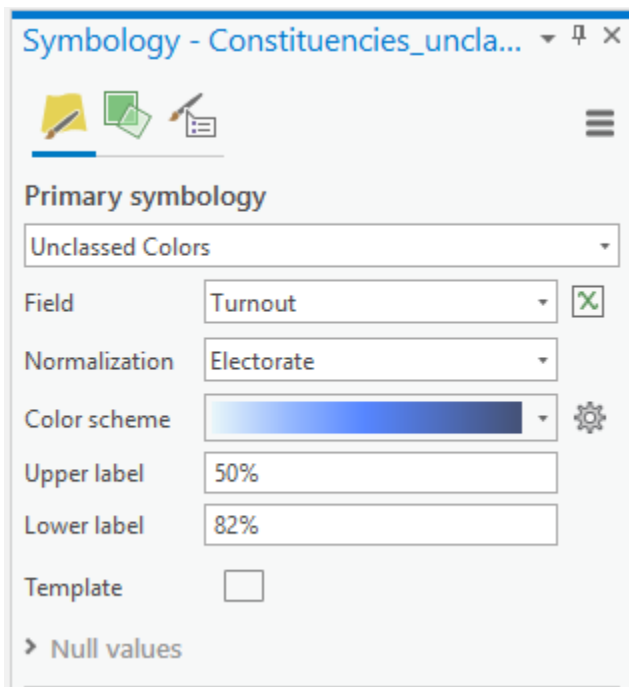
Dragging the handles in the histogram shifts the way that color is applied to the unclassified values.

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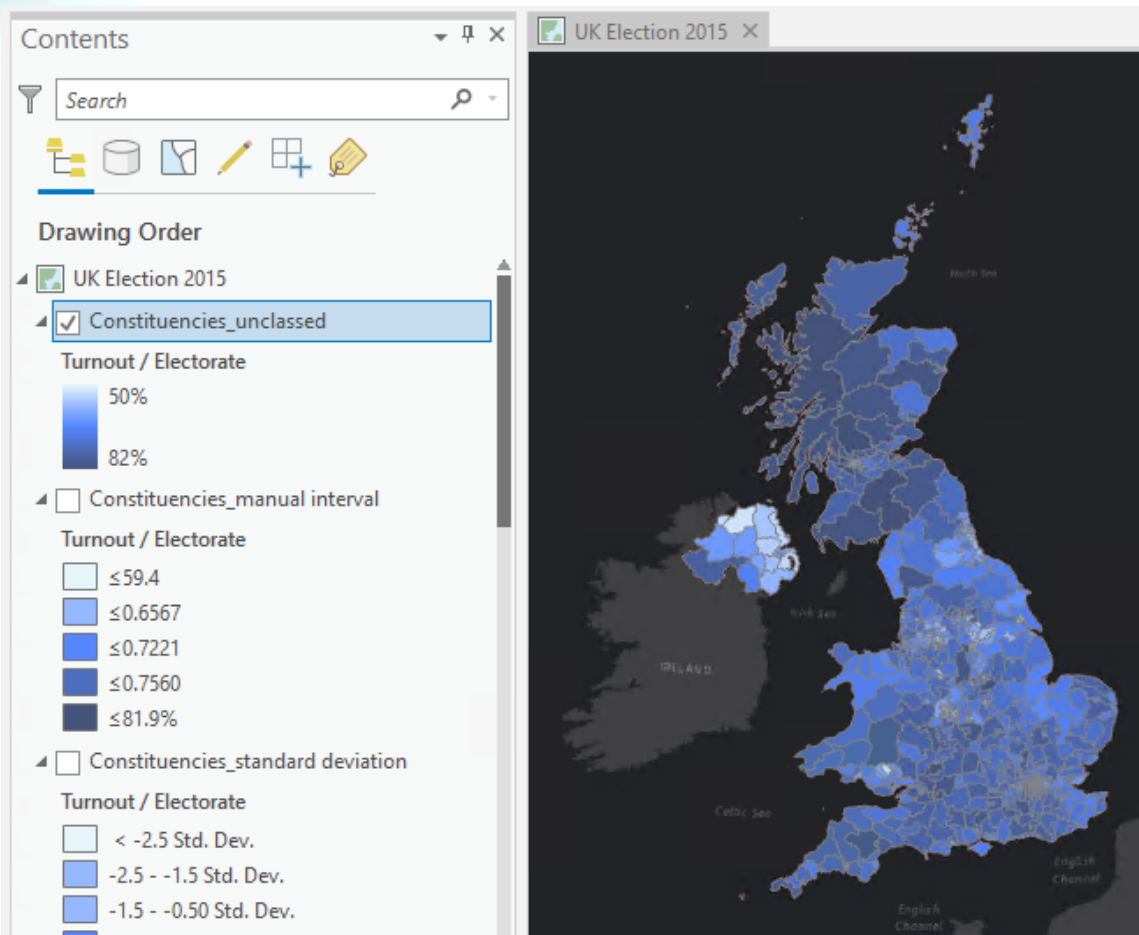
What is the key aspect of the data that is highlighted with this classification method?

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.

ⓘ Save your project and exit ArcGIS Pro.

Conclusion

You have explored a range of classification techniques and learned that there are numerous ways to classify a single dataset. Therefore, there is ample opportunity to represent the data well or poorly, objectively or persuasively.

For any dataset that you classify, first determine which method of classification will most effectively communicate your message without obfuscating the truth.

You will use this same dataset in a future exercise to create different thematic maps and study color and symbology in more detail. However, for now, the most important lesson to understand is that data can be manipulated to tell different stories.

Stretch Goal

Continue exploring the classification of numerical data with the following stretch goal:

- Classify some of the other data in the UK Election 2015 dataset. Observe how the 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.

Post your questions and observations about this stretch goal in the Lesson Forum. Be sure to include the **#stretch** hashtag in your post title. We would love to hear about any experimenting that you did beyond the steps of the original exercise!