

# Exercise 2: Classification



## Introduction

In this exercise you learn how to symbolize quantitative values by classifying data using the layer's symbology settings.

## Table of Contents

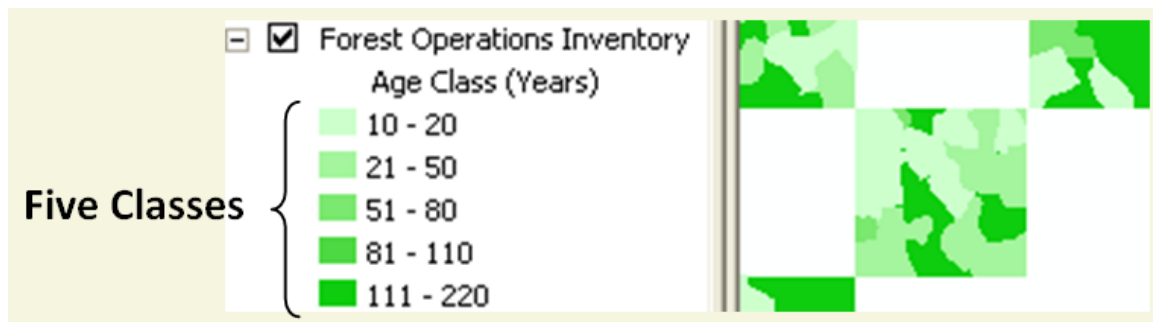
Introduction .....	1
Step 1 - Standard Classifications .....	1
Step 2 – Normalization.....	10

## Overview of major concepts

1. Standard Classifications
2. Normalization

## Part 1: Standard Classifications

When you classify data in ArcMap you are sorting an attribute's values into groups (or categories) called classes. The values in each class are represented by the same symbol and color. For example, the next screen capture illustrates how a Forest Operations Inventory layer is symbolized into five classes—each class is a range (in years) for the age of trees.



As you will experience in this exercise, ArcMap employs several classification methods. Data classification helps show data differentiation or trends. The result is an attribute-based legend that helps the user visually interpret the data.

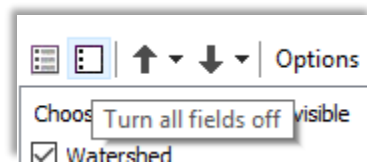
**Goal for this step:** Sample ArcMap's standard classifications methods to best represent the distribution of noxious weeds among 5th-field watersheds in the area of the Western Oregon Plan Review (WOPR).

1. Launch **ArcMap** (if needed), and open **Classification.mxd** in the \Data folder.

Currently, all watersheds have the same "Single Symbol" symbology. The "Weeds by 5th Field Watersheds" layer is a customized Hydrologic Unit Code (HUC) layer that includes an attribute on the weed acreage for each watershed. Also on the map is the name of each BLM District Office in Western Oregon.

Because classifying data in ArcMap relies on a layer's attribute, we will streamline the attribute names you will need to make a weed-distribution map. By default, all attributes are visible. As there are only four attributes we need to work with in this step, many of the attributes can become invisible.

2. Open the Properties for the **Weeds by 5th Field Watersheds** layer. *Hint: Right-click layer name in TOC.*
3. Activate the **Fields** tab.
4. **Turn off** all fields.



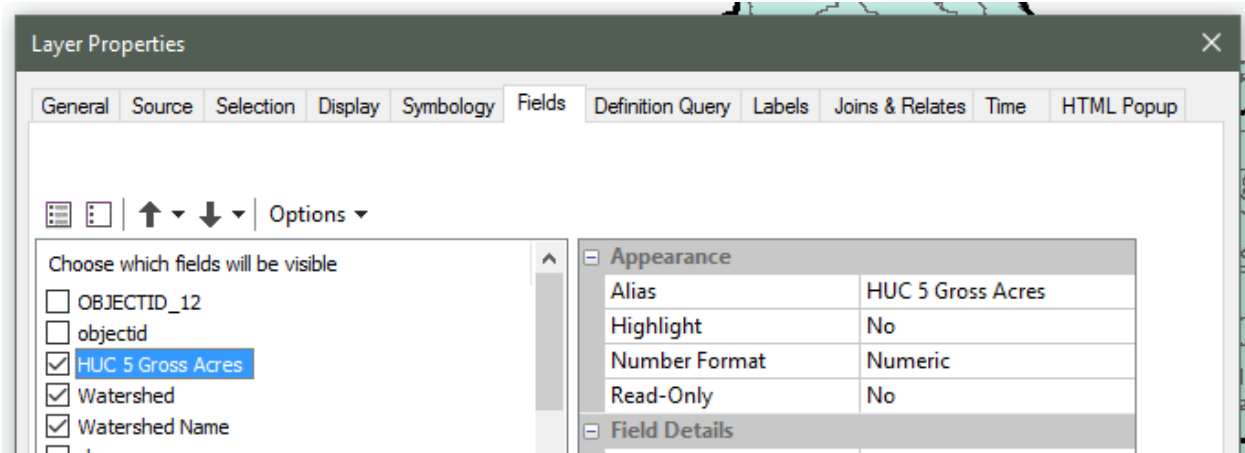
5. Add checkmarks for the following attributes:

- acres
- watershed
- watershed\_name
- huc5gridgross (Near the bottom of the list)

In the right panel of the field tab are the field properties. When a field is highlighted in the list of fields, the properties will be displayed. Here you can enter an Alias for a field. An Alias is a second name you can enter for a more intuitive field name that serves as the attribute's field header.

6. Under the Alias column, rename the following field names:

- acres → **HUC 5 Gross Acres** (*weed acres per watershed*)
- watershed → **Watershed**
- Watershed\_name → **Watershed Name**
- huc5gridgross → **HUC5 Weed Acres** (*watershed acres*)



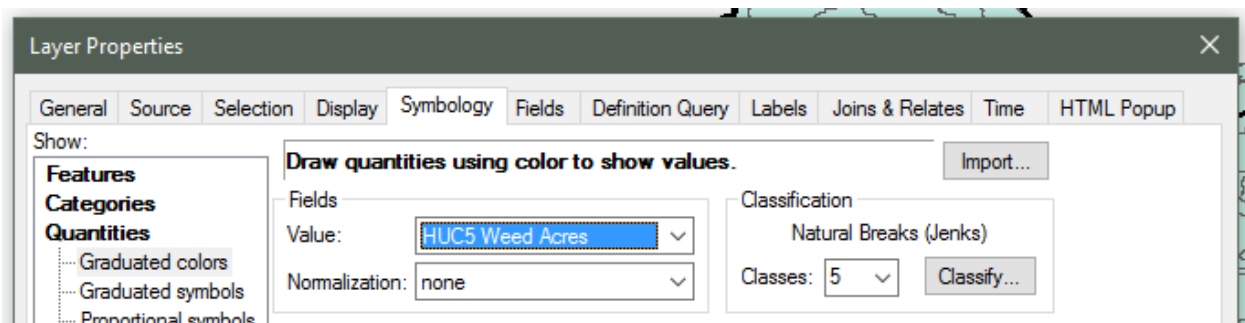
7. Click the **Apply** button; verify the aliases have changed.

You have just simplified the layer's attributes by limiting the number of fields and using readable aliases. The alias is used when you symbolize the data based on an attribute. Next, we will assign classifications. Our first classification method employs ArcMap's default classification method called Natural Breaks (Jenks1).

8. Activate the **Symbology** tab.

9. Under the Show list, single-click on **Quantities > Graduated colors**.

10. Set the Value field to **HUC5 Weed Acres**.

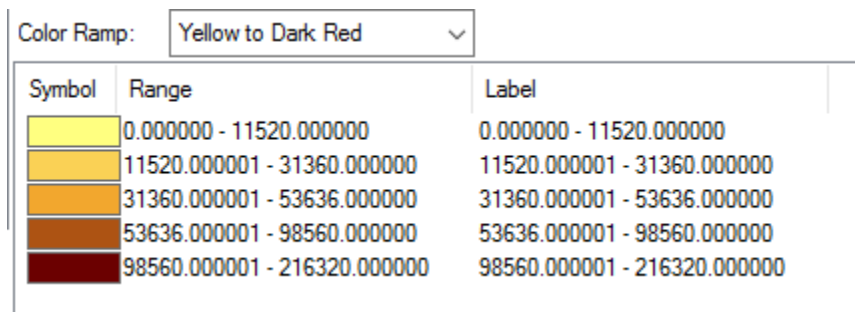


- i. By default, “Graduated colors” is highlighted. Once you specify an attribute, the attribute’s values are organized into classes—each class possessing a different, graduated color shade.

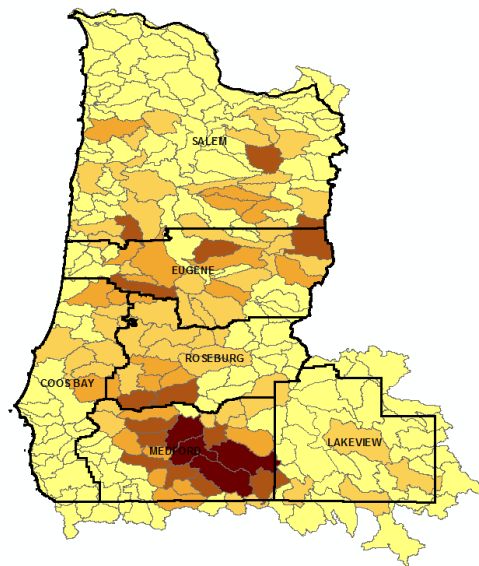
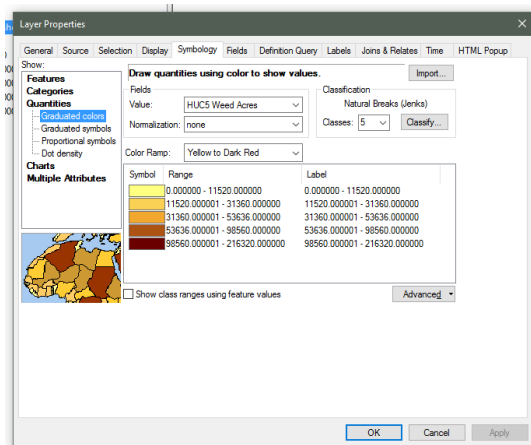
By selecting the attribute to classify, ArcMap automatically puts that attribute’s values into five classes using a Natural Breaks (Jenks) type classification. Each class is grouped by natural breaks (as the breaks occur in the total range of values), determined by a numerical algorithm developed by G.F. Jenks, Dept. of Geography, University of Kansas

11. Right-click the **Color Ramp graphic** → remove the checkmark for **Graphic View**.
12. Change the Color Ramp to **Yellow to Dark Red**.

Because the spread of noxious weeds is a concern, the classification colors should be indicative of their “threat” level. A yellow fill color will apply to watersheds with the lowest acreage of weeds. For watersheds with the highest acreage of weeds, dark red will be used.



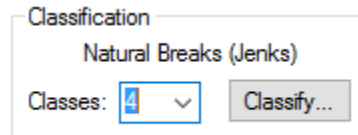
13. Click the **Apply** button.
14. As needed, move the **Layer Properties** window to see the **classification** results.



Based on the five classes of the Natural Breaks classification, watersheds in the BLM Medford District have some of the highest weed acreages.

Because of the various classification methods available in ArcMap, it is relatively easy to display misleading information. For any map you make (especially public maps) is critical that the map conveys the most accurate information possible. If you have questions regarding the appropriateness of your classification method, talk to your office's GIS Coordinator. Let's see how the map looks when we reduce the number of classes.

15. Set the number of **Natural Breaks (Jenks)** Classes to **4**.
16. Click **Apply**.

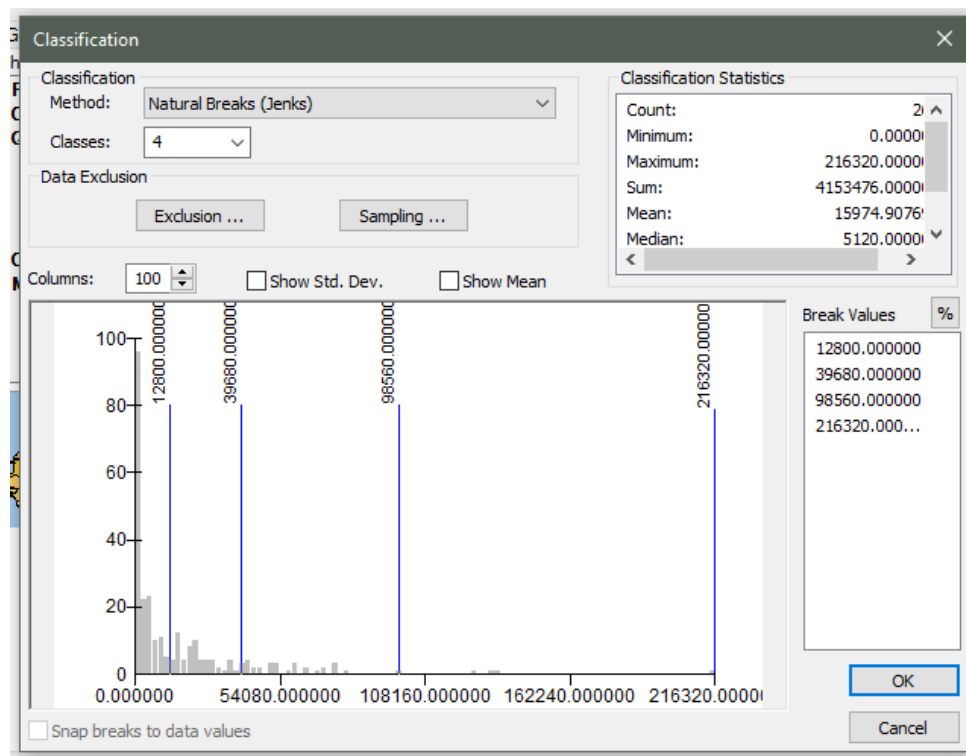


Although the change in the Data View may seem subtle, if you compare the next screen capture with the screen capture on the previous page, you should notice more watersheds are regrouped into the darker red class.

Which map is more correct, Natural Breaks with four classes or five classes? You should first ask whether or not the Natural Breaks classification is correctly portraying the data? If the answer is "yes," then you should ask whether or not the map is easier to interpret using four or five classes? The answer to the previous question is subject to the map reader's interpretation. Let's see how other classification methods affect the weeds-distribution map.

17. Click the Classify button.

The Classification dialog window opens. In the window is a histogram which shows the distribution of weed acreage for each 5th-field watershed. The horizontal X-axis represents a range of weed acreage values for each watershed, and starts at zero acres. The vertical Y-axis represents the number of times (i.e., the frequency) a weed-acreage value occurs. If you haven't already noticed, zero acres have the highest frequency.



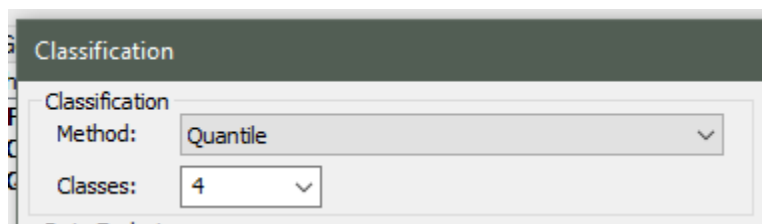
i. Y axis = Frequency

X axis = Weed Acres/Watershed

The vertical blue lines in the histogram represent the “Break Values” between classes. Values between the class breaks receive the same symbology and graduated color.

We have now seen the results from using the Natural Breaks classification. Let’s see how the map changes when we apply the following standard classification methods: Quantile, Standard Deviation, and Equal Interval.

18. From the Method drop-down list, choose **Quantile**.



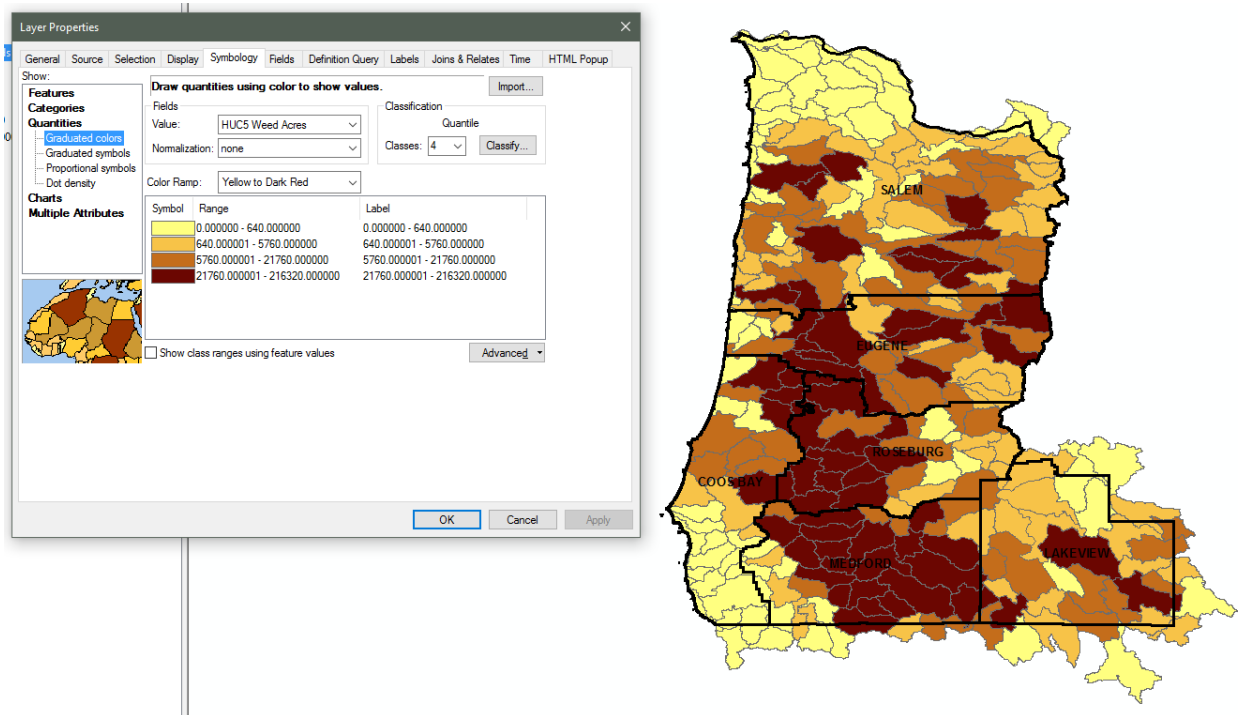
For each class, the Quantile classification method groups the same number of features. In our example, the watersheds are polygon features. As we are continuing with four classes, the “Weeds by 5th Field Watersheds” layer is symbolized with 65 watersheds per class (i.e.,  $260 \text{ polygons} \div 4 \text{ classes} = 65 \text{ polygons/class}$ ).

19. Click **OK** to close the Classification window.

20. Click **OK**.



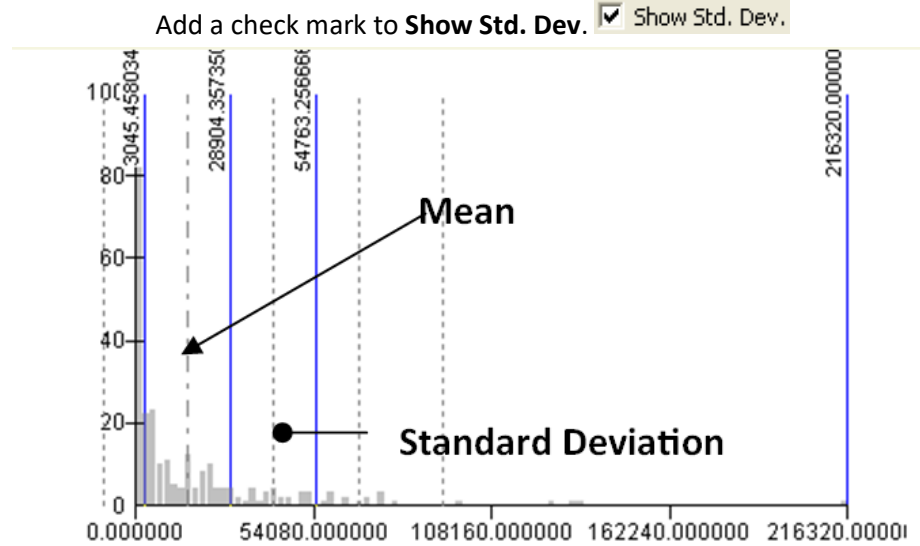
The Layer Properties window closes and the Data View redraws to the new classification.



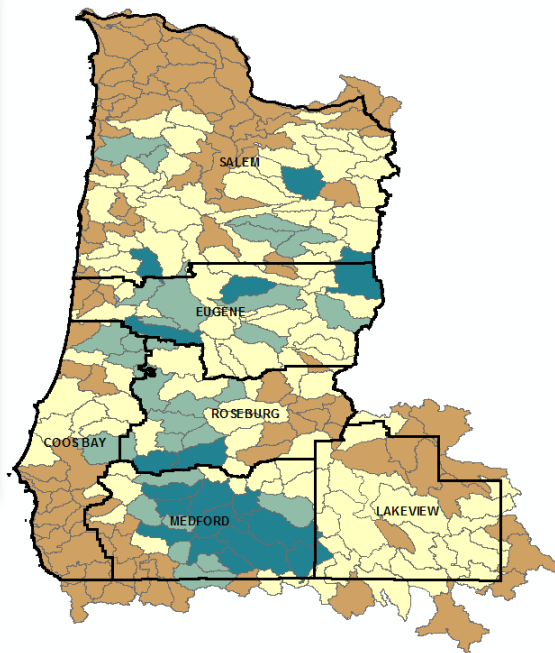
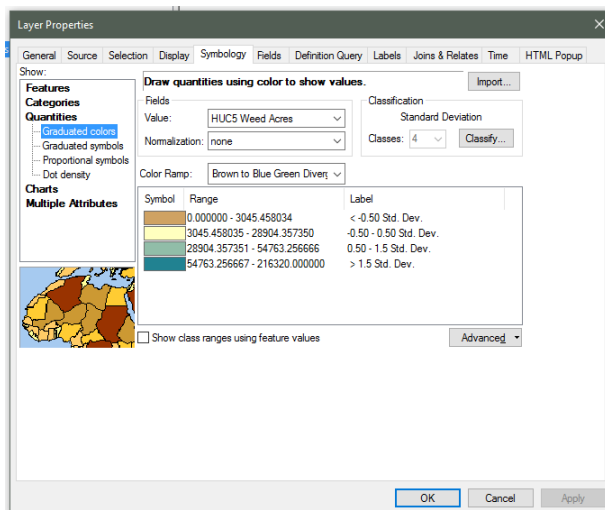
Even if you are unsure whether or not your map is correct, you should at least be able to see that the results from a Quantile classification differs from that of a Natural Breaks classification. Quantile classification works best with linearly distributed data such as population percentages (e.g., 0% to 100%). Quantile classification can be misleading because low values can be included in the same class as high values. Therefore, quantile classification is inappropriate for visualizing the distribution of noxious weeds. Next, let's apply the Standard Deviation classification.

21. Open the Properties for the **Weeds by 5th Field Watersheds** layer.
22. Under the **Symbology** tab, click the **Classify** button.
23. From the Method drop-down list, choose **Standard Deviation**.
  - i. The histogram changes to the new classification method. The Standard Deviation classification is used to visualize how values vary from the mean value.
24. Add a checkmark to '**Show Mean**'. ☒ Show Mean

On the histogram, a vertical dashed line indicates the location of the mean value. For the weeds data, the average weed acreage per watershed is 15,975 acres.



25. Click **OK**, then **Apply**. The Data View redraws the new classification.



In the Layer Properties window, you should notice the Color Ramp has changed. Watersheds with weed acreages below the mean are orange. Watersheds with weed acreage near the mean are yellow. Watersheds with weed acreages above the mean have various shades of green.

The Standard Deviation classification works best with “normally” distributed data; that is, data that is distributed symmetrically above and below the mean (Do you remember the “bell-shaped curve” from Statistics?). The weeds data is skewed below the mean—with a large number of values at 0 acres/watershed. Because of the asymmetry in values, the Standard



Deviation classification inaccurately represents the distribution of noxious weeds. Our final classification to apply is the Equal Interval classification. Because of a software inconsistency occurring when changing the Standard Deviation classification, you may need to reselect the “HUC5 Weed Acres” attribute.

From the Value drop-down list, reselect HUC5 Weed Acres.



The layer’s classification resets to Natural Breaks (Jenks), four classes.

26. In the **Layer Properties** window, click the **Classify** button.

i. The Classification window opens.

27. From the Method drop-down list, choose **Equal Interval**.

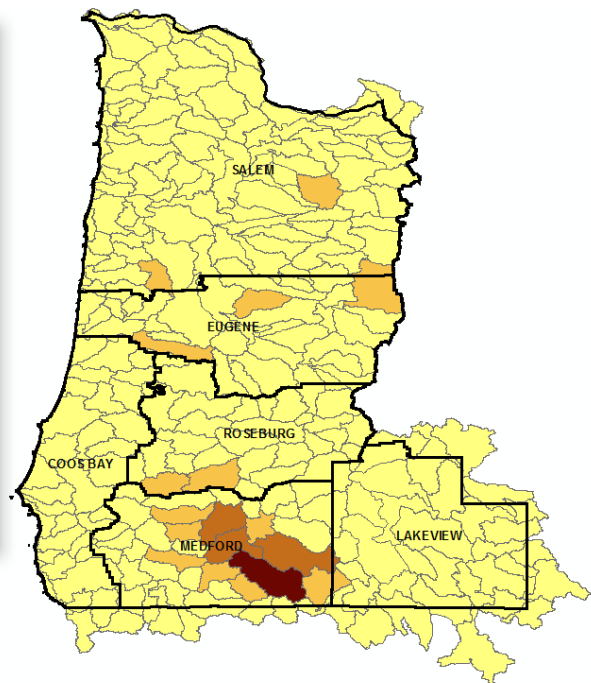
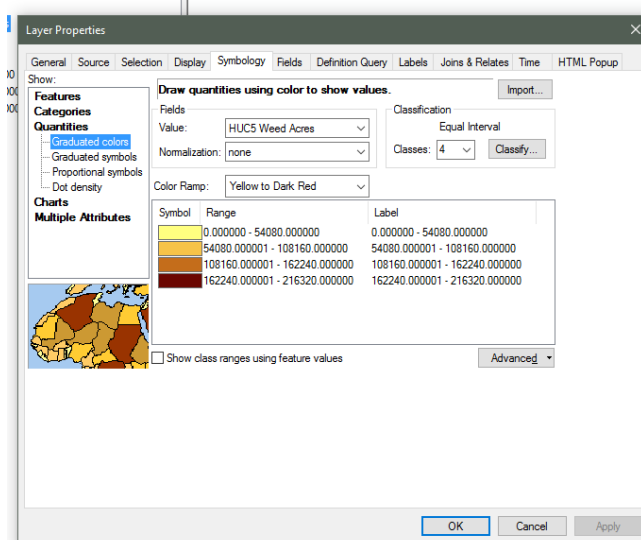
The Equal Interval classification divides the range of values into equal sized sub-ranges. For the weeds data, the acreage values range from 0 to 216,320 acres. Because we specified four classes, each sub-range equals 54,080 acres (i.e., 216,320 acres ÷ 4 classes = 54,080 acres/class).

28. Change the number of classes to **4**, if needed.

29. Click **OK**.

30. Verify the Color Ramp is **Yellow to Dark Red**. *If not, change the Color Ramp.*

31. Click **Apply**. *As in the next screen capture, the Data View redraws the new classification.*



Equal interval classification is useful in emphasizing relative amounts of an attribute. The classification works best with familiar data ranges such as percentages and temperatures.

However, when dealing with acreage per watershed, the map reader may have difficulty conceptualizing that each class represents a specific acreage (i.e., 54,080 acres). Therefore, Equal Interval classification is an inappropriate classification method for visualizing the distribution of noxious weeds for each watershed.

Which classification method should be used with the weeds data? Short of creating your own classification (see the next step), the distribution of noxious weeds for each watershed is best represented with a Natural Breaks classification using either four or five classes.

## Part 2: Normalization

By changing the class-break values, you can customize any of ArcMap's classification methods. You can either manually adjust the vertical blue lines in the classification histogram, or enter new break point values. Our goal for this step is to visualize the percentage of weed coverage for each watershed.

Why use percentages? A map reader can better conceptualize percentage values as opposed to large area values (such as weed acres per watershed). Based on the attribute values we have worked with so far, the "Weeds by 5th Field Watersheds" layer has no attribute information on the percent weed cover per watershed. The only attribute information available is two acreage fields:

- HUC5 Weed Acres (weed acreage per watershed)
- HUC5 Gross Acres (total watershed acreage)

*In the Layer Properties window, what attribute is referenced by the **Value Field**?*

1. From the Normalization drop-down list, choose HUC5 Gross Acres.
  - i. *Still within the layer **properties**→**symbolology** tab.*

Values from the HUC5 Weed Acres attribute have been subjected to various classification methods of this exercise. By dividing the HUC5 Weed Acres attribute values by the HUC5 Gross Acres attribute values, a percentage is derived:

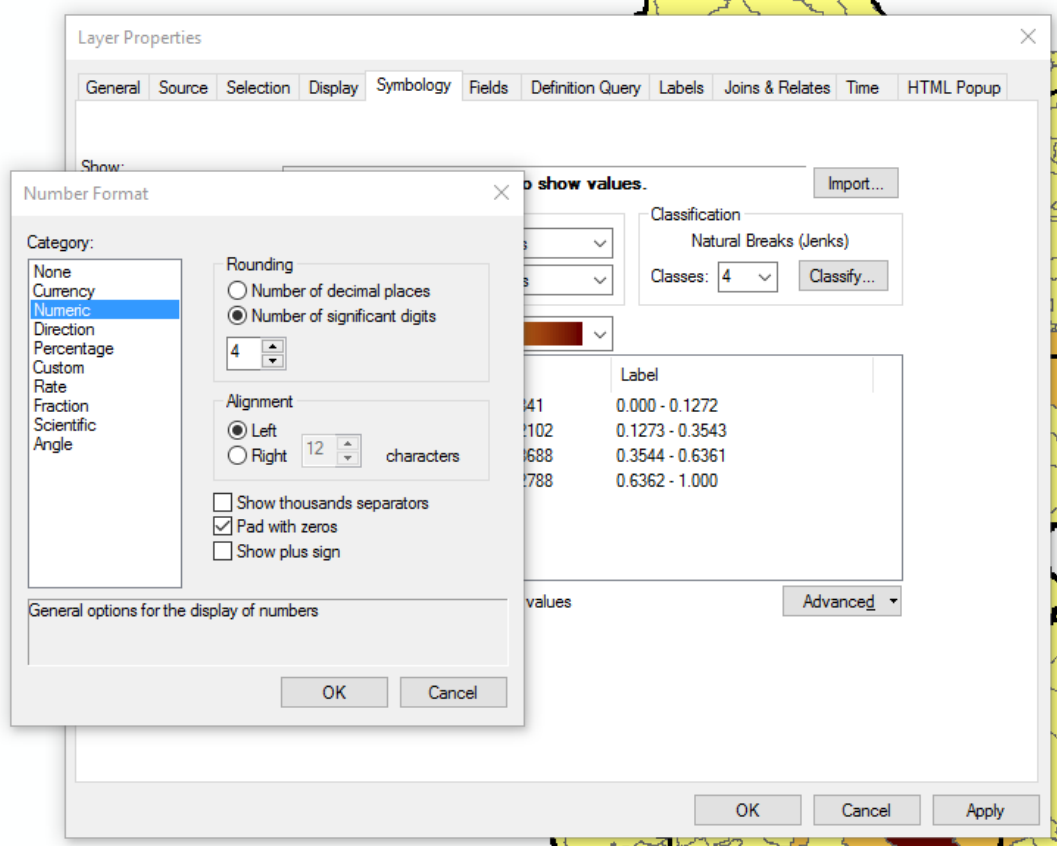
$$\text{HUC5 Weed Acres} \div \text{HUC5 Gross Acres} = \% \text{ Weed Cover}$$

Dividing one "area" attribute value by another "area" is called normalization. When should you normalize data? Use normalization when your data is represented by large numbers (e.g., watershed area, district area, county population, etc.). Normalization is done by ArcMap. You just need to specify the "Normalization" attribute.

Notice, nearly all the class values are less than one. There is also a change to the classification method. After applying the normalization, ArcMap reverts back to the Natural Break (Jenks) classification. However, the four classes you specified earlier are still retained. Shortly, we will customize the Natural Breaks classification.

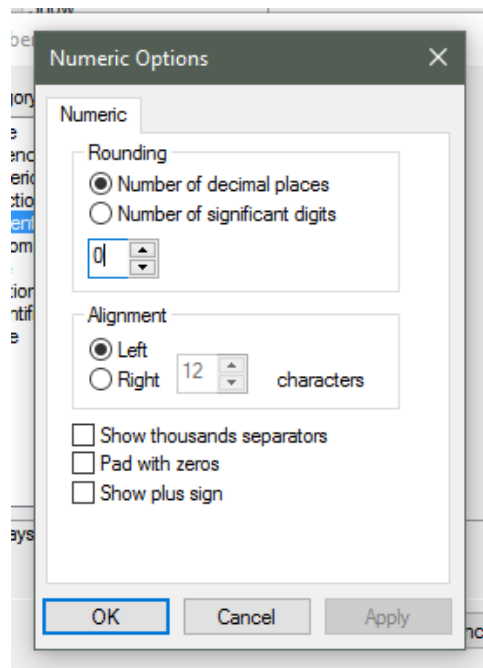
- ii. We can further improve symbol-label readability by setting the labels to display percentages.
2. Click on the header for the **Label column** → **Format Labels**.

The Number Format dialog window opens.



3. Select Percentage.
4. Enable the option for “The number represents a fraction...”
5. Click the Numeric Options button.





We can also set the number of decimal places used in the symbol labels. The default setting is to use six decimal places—even if zeroes are the only value to report.



6. Set the number of **decimal places** to zero.

7. Click the **OK** buttons until you are back to the **Layer Properties** dialog window.

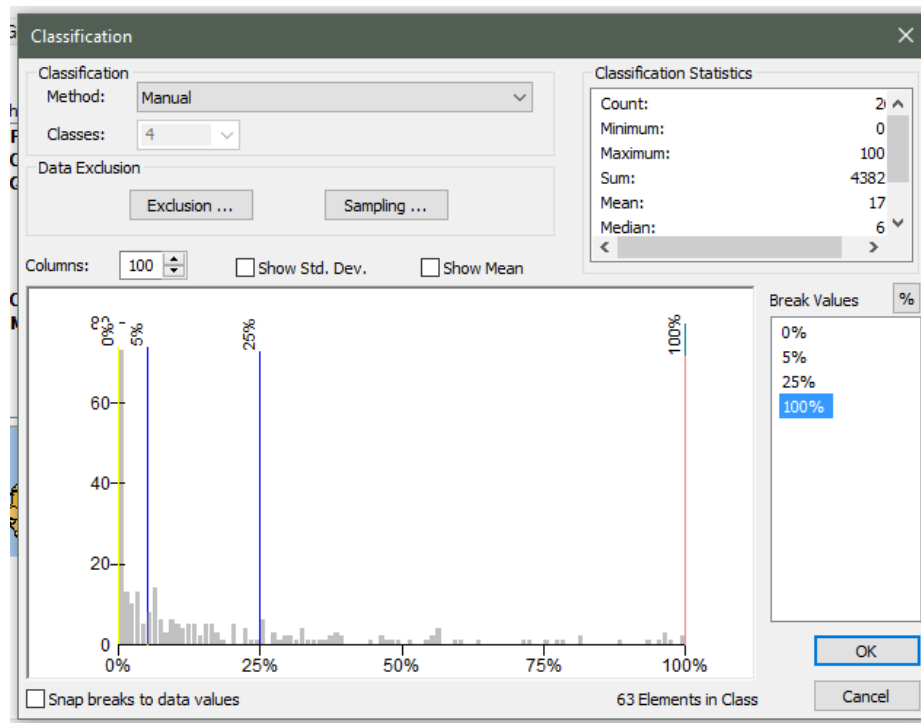
Under the Label column, each class is labeled with a percentage range.

Symbol	Range	Label
	0.00000000 - 0.127162841	0% - 13%
	0.127162842 - 0.354342102	14% - 35%
	0.354342103 - 0.636068688	36% - 64%
	0.636068689 - 0.999962788	65% - 100%

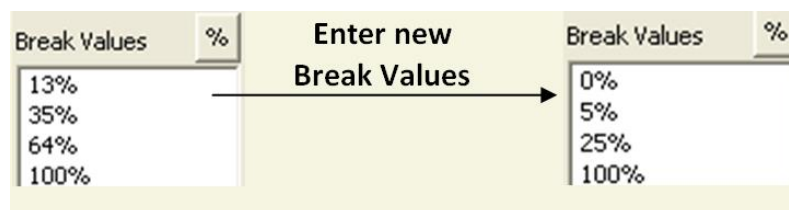
Why are we normalizing the weeds data? It is easier for the map reader to understand how serious noxious weeds have spread when the weeds cover over 65% of the watershed's area versus a large area number. Before we apply the normalized values, a "weeds" specialist wants the map to further emphasize those watersheds where weed coverage is greater than 25%. This allows the specialist to better allocate resources needed for noxious-weed management.

8. Click the **Classify** button.

As illustrated in the next screen capture, the classification histogram for the Natural Breaks (Jenks) classification is shown. At the window's lower right, a user can manually change any or all of the listed Break Values. Each Break Value is graphically represented as a blue vertical line in the Classification histogram.



To enter a Break Value, single click to highlight the existing value, enter the **new Break Value** (without the percent sign), and press the <Enter> key.



9. Change the following Break Values:

- 13% → 0%
- 35% → 5%
- 64% → 25%

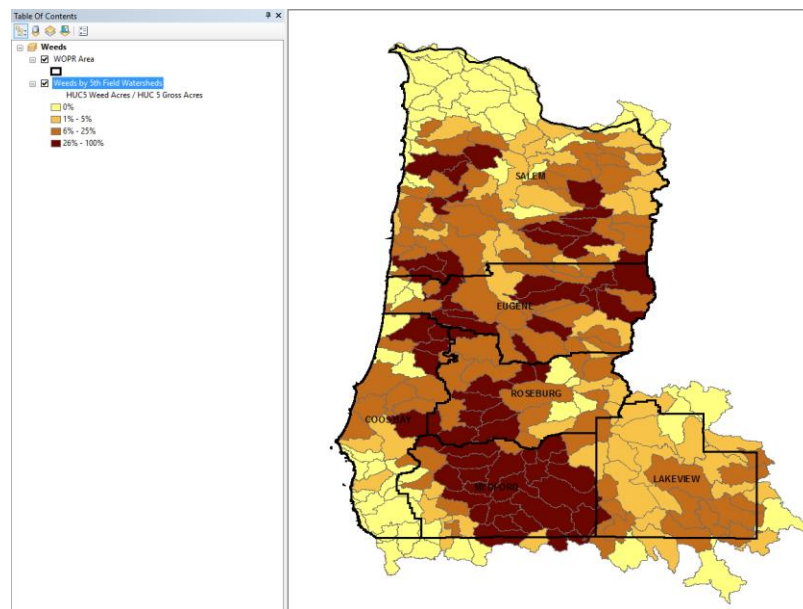
With the new Break Values, two changes occur in the Classification dialog window. First, the Classification method is now “Manual.” Second, in the Classification histogram, the positions of the blue vertical lines are repositioned to match your new Break Values. We are ready to apply your customized classification method.

With the new class breaks, the “weeds” specialist can ignore the watersheds in yellow, and focus on those watersheds with greater than 25% weed coverage.

Fields		Classification	
Value:	HUC5 Weed Acres	Manual	
Normalization:	HUC 5 Gross Acres	Classes: 4	<b>Classify...</b>
Color Ramp:			
Symbol	Range	Label	
	0.00000000	0%	
	0.000000001 - 0.050000000	1% - 5%	
	0.050000001 - 0.250000000	6% - 25%	
	0.250000001 - 1.00000000	26% - 100%	

10. Click **OK**.

11. Click **OK** to close the **Layer Properties** window.



12. Click the **Save** button.

13. Save your classifications as a layer file in the ...**\Data\Workspace folder**.

14. Exit **ArcMap**, unless you are moving on to the next exercise.

The primary purpose of this exercise is give you experience in changing a layer's symbol and color. Nearly all ArcMap users need to know this invaluable skill.

To learn more about classifying data in ArcMap; From ArcGIS Desktop Help, search for the topic called "Ways to map quantitative data."

**END EXERCISE**