

## Topic 2

# Mapped Data Visualization and Summary

### 2.1 Visceral Visions

This and the following topics will apply map analysis techniques in a case study format. The map-ematical procedures used to translate data into useful information is demonstrated using a subset of the data collected for a large multidisciplinary project of the Water Management Unit of ARS and the Experiment Station at Colorado State University. The MapCalc Learner software by Red Hen Systems is used throughout the case study for analysis and display of the data. The “hands-on” exercises accompanying this book are ideal for classroom and self-study.

The most fundamental concept in analyzing precision farming data is the recognition that all of the maps are “numbers first, pictures later.” While the human mind is good at a lot of things, noting subtle differences within large, complex sets of spatial data isn’t one of them. To visualize these data we first must aggregate

the detailed information into a few discrete categories then assign colors to each of the groupings.

For example, the yield maps shown in figure 2-1

appear dramatically different but were generated from the same harvest data. Both maps have five categories extending from low to high yield and use the same color scheme.

The differences in appearance arise from the methods used in determining the breakpoints—the one on the left uses “Equal Ranges” while the on the right uses “Equal Count.” **Equal Ranges** is similar to an elevation contour map since it divides the data range (2.33 to 295bu/ac) into five steps of about 60 bushels each. Note that the middle interval (119.3 to 177.8) has more than 100 acres and comprises over half of the field (54%).

**Equal Count**, on the other hand, divides the same data into five classes, but this time each category represents about the same area (20%) and contains a similar number of samples (about 660).

The differences between the two approaches are best visualized from a combined statistical/graphical perspective. The left

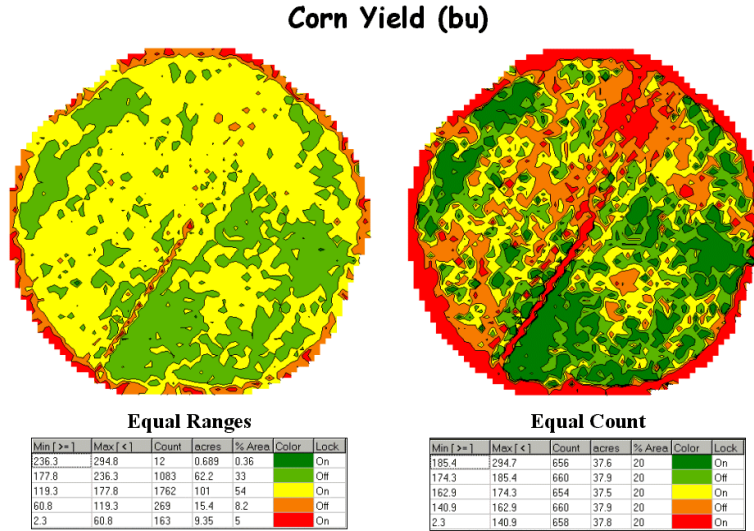


Figure 2-1. Two different renderings (categorizations) of corn yield data.

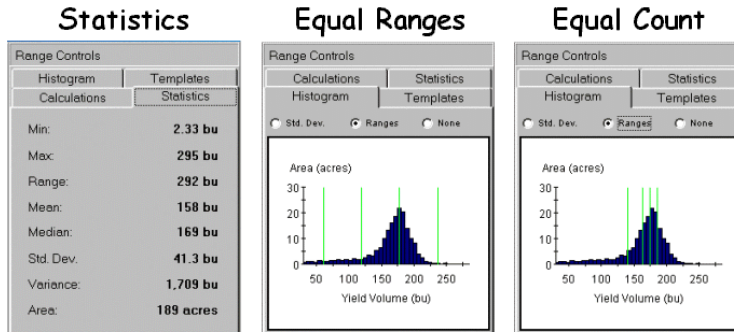


Figure 2-2. Statistical summary of yield data.

panel in figure 1-2 shows the basic *Descriptive Statistics* for the data set—*min, max, range, mean, median, standard deviation, variance* and *total area*. While these statistics report a fairly good harvest overall (an average yield of 158bu/ac), they fail to map the yield patterns (spatial variability) throughout the field.

The two histograms on the right side of figure 2-2 link the statistical perspective with the map displays. Recall that a histogram is simply a plot of the data range (X-axis) versus the number of samples (Y-axis). The “vertical bars” superimposed on the histograms identify the breakpoints for the categories—evenly spaced along the Y-axis for Equal Ranges while variably spaced for Equal Count.

Note the direct relationship between the placement of the bars and the category intervals shown in the map legends shown in figure 1-1. For example, the first interval begins at 2.3 bushels for both maps but breaks at 60.8 for Equal Ranges and 140.9 for Equal Count. The differences in the placement of the breakpoints (vertical bars) account for the differences in the appearance in the two yield maps.

Similarly, changes in the *number of intervals*, and to a lesser degree, changes in the *color scheme* will produce different looking maps. By its very nature, discrete contour mapping aggregates the spatial variability within field-collected data. The assignment into categories is inherently a subjective process and any changes in the contouring parameters result in different map renderings. Since decisions are based on the appearance of spatial patterns, different

actions could be taken depending on which visualization of the actual data was chosen.

The map on the left-side of figure 2-3 shows the Equal Count contour map of yield composed of irregular polygons that identify the discrete categories. The inset directly above the contour map characterizes the underlying *lattice data structure* for a portion of the field. A yield value is stored for each of the line intersections of the superimposed regular grid. The breakpoint values for the map categories are plotted along the lines connecting the actual data. The contours are formed by connecting the breakpoints, smoothing the boundary lines and filling the polygons with the appropriate color.

The *grid data structure* utilizes a superimposed reference grid as well. However, as depicted in the grid map and inset on the right side of figure 2-3, the centroid of each grid cell is assigned the actual value and the entire cell is filled with the appropriate color. Note similarities and differences in the yield patterns between the contour and grid displays—best yields generally in the southeast and northwest portions of the field.

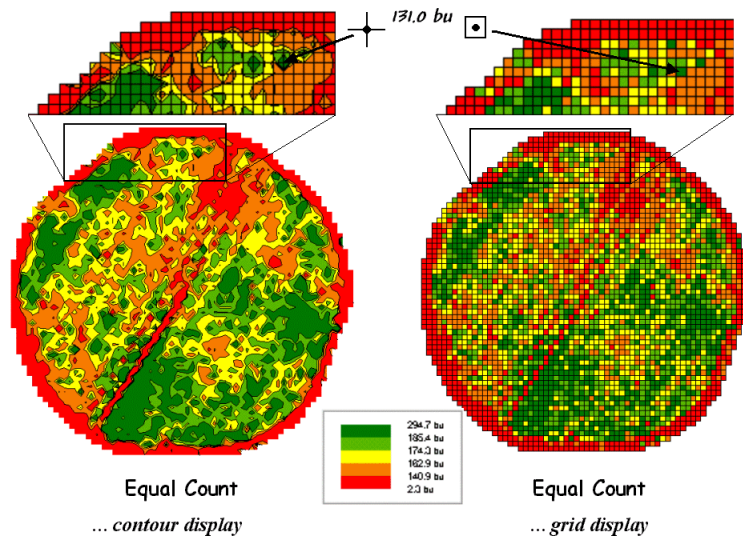


Figure 2-3. A reference grid is used to directly link the map display and the stored data.

The next section

will investigate the important considerations in representing precision farming data as continuous map surfaces, viewing these data as 3D displays and pre-processing procedures used to “normalize” maps for subsequent analysis.

## 2.2 Preprocessing and Normalization

Preprocessing and normalization are critical steps in map analysis that are often overlooked when the sole objective is to generate graphical output. In analysis, however, map values and their numerical relationships takes center stage.

**Preprocessing** involves conversion of raw data into consistent units that accurately represent field conditions. For example, a yield monitor’s *calibration* is used to translate electronic signals into measurements of crop production units, such as bushels per acre (measure of volume) or tons per hectare (measure of mass). The appropriate units (e.g., bushels vs. tons) and measurement system (English acres vs. metric hectares) is determined and applied to the raw data. While this point is important to software developers, it is somewhat moot in most practical applications after the initial installation of software preferences.

A bit more cantankerous preprocessing concern involves raw data adjustments and corrections. Like calibrations, *adjustments* involve “tweaking” of the values... sort of like a slight turn on that bathroom scale to alter the reading to what you know is your true weight.

*Corrections*, on the other hand, dramatically change measurement values, both numerically and spatially. For example, the time lag from a combine’s header to the yield monitor can require considerable repositioning of yield measurements.\* Also, a combine can pass over the same place more than once and the duplicate records must be corrected. Even more difficult is the corrections of a reading from a partial header width or the effects of emptying and filling the thrashing bin as a combine exits from one pass and enters another.

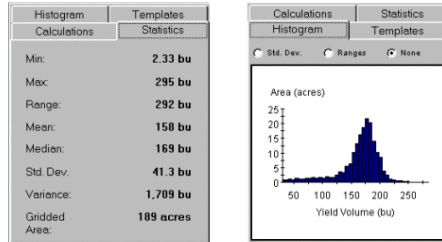
The old adage “garbage in, garbage out” holds true for any data collection endeavor. Procedures for calibration, adjustments and corrections are used to translate raw data

measurements into the most accurate map values possible. But even the best data often needs a further refinement before it is ready for analysis.

**Normalization** involves standardization of a data set, usually for comparison among different types of data. For example, crop rotation can provide different

views of the productivity of a field. While direct visual comparison of onion and corn yield maps might provide insight, some form of normalization of the data is needed prior to any quantitative analysis. In a sense, normalization techniques allow you to “compare apples and oranges” using a standard “mixed fruit scale of numbers.”

Insert A – Not Normalized



Insert B – Goal\_250 Normalized

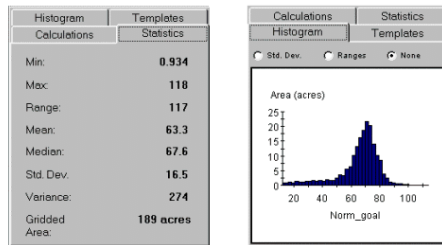
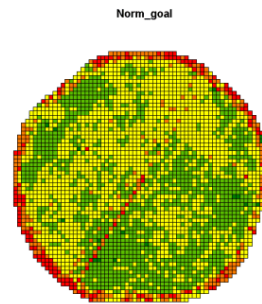
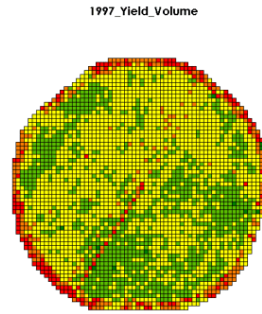


Figure 2-4. Comparison of original and “goal normalized” data.



The most basic normalization procedure uses a *goal* to adjust map values. For example, a goal of 250 bushels per acre might be used to normalize a yield map for corn. The equation,

$$\text{Norm\_GOAL} = (\text{mapValue} / 250) * 100$$

derives the percentage of the goal achieved by each location in a field. In evaluating the equation, the computer substitutes a map value for a field location, completes the calculation, stores the result, and then repeats the process for all of the other map locations.

Insert B in figure 2-4 shows the results of goal normalization. Note the differences in the descriptive statistics between the original and normalized data—a data range of 2.33 to 295 with an average of 158 bushels per acre for the original data versus .934 to 118 with an average of 63.3 percent of the goal.

Now note that the two histograms look nearly identical and that the same holds for the two maps. In theory, the histogram and map patterns are identical. The slight differences in the figure are

simply an artifact of rounding the display intervals for the rescaled graphics. While the descriptive statistics are different, the relationships (patterns) in the normalized histogram and map are identical to the original data.

That’s an important point— both the numeric and spatial relationships in the data are preserved during normalization. In effect, normalization simply “rescales” the values like changing from one set of units to another (e.g., switching from bushels per acre to cubic meters per hectare). The significance of the goal normalization is that the new scale allows comparison among different fields and even crop types— the “mixed fruit” expression of apples and oranges.

Other commonly used normalization expressions are 0-100 and SNV. **0-100** normalization forces a consistent range of values by spatially evaluating the equation—

$$\text{Norm}_{0-100} = ((\text{mapValue} - \text{min}) * 100) / (\text{max} - \text{min})$$

where max and min are the maximum and minimum values of the original data.

The result is a rescaling of the data to the range 0 to 100 while retaining the same relative numeric and spatial patterns of the original data. While goal normalization benchmarks a standard value without regard to the original data, the 0-100 procedure rescales the original data range to a fixed, standard range.

The third normalization procedure, standard normal variable (*SNV*), uses yet another approach. It rescales the data based on its central tendency by applying the equation—

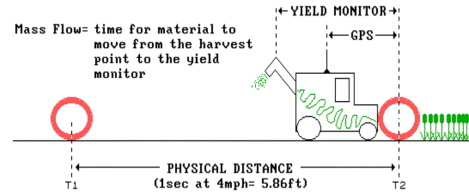
$$\text{Norm}_{SNV} = ((\text{mapValue} - \text{mean}) / \text{stdev}) * 100$$

where mean is the average and stdev is the standard deviation of the original data.

The result is rescaling of the data to the percent variation from the average. Mapped data expressed in this form enables you to easily identify “statistically unusual” areas— +100% locates areas that are one standard deviation above the typical yield; -100% locates areas that are one standard deviation below.

Map preprocessing and normalization are often forgotten steps in a rush to make a map, but they are critical pre-cursor to a host of subsequent

analyses. For precision farming to move beyond pretty maps, the map values themselves become the focus— both their accuracy and their appropriate scaling.



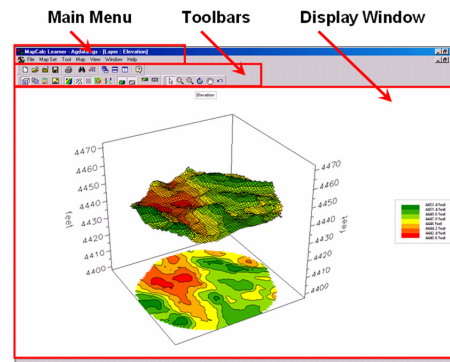
\* See [www.innovativegis.com/basis/pfprimer/](http://www.innovativegis.com/basis/pfprimer/), an online book entitled *Precision Farming Primer*, Topic 1 Continuous Data Logging, Section 2 “Putting Yield in Its Proper Place” for discussion on adjusting yield measurement position.


## 2.3 Exercises


Access MapCalc using the *Agdata.rgs* data set as you did in the previous exercise (Exercise 1.3.2) by selecting **Start → Programs → MapCalc Learner → MapCalc Learner → Open existing map set → PA\_AgData.rgs**. The following set of exercises utilizes this database.


### 2.3.1 Interacting with Map Displays


The MapCalc user interface consists of three components—*Main Menu* with drop-down lists, *Toolbars* with direct access buttons and *Display Window* for 2D and 3D plotting of mapped data. In addition to these basic components special dialog boxes will pop-up depending on the operation selected. The pop-up windows are used for specifying input or contain summary reports. The map displays can be easily modified and extensive tools for data summary and analysis are available. This exercise focuses on the basic display capabilities.





 In the toolbar section select the *Plus+ Magnifying Glass* button, then click and drag a rectangular area on the map to enlarge the area.


 Select the *Minus - Magnifying Glass* button, then click and slide the cursor vertically up and down to continuously enlarge and reduce the display.


 Select the *Hand* button then click and drag to move the map display in any direction.


 Select the *Rotate* button then click and drag to rotate the 3D plot in any direction.


 Select the *Reset* button to restore the default viewing specification of the map.

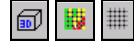
 Toggle on and off the *Contour Fill*, *Contour Lines* and *Grid Mesh* buttons. The grid mesh identifies the analysis frame used to store the data.

 Toggle on and off the *Floor* and *Ceiling* projections.


 Toggle on and off the *Use Cells* button to switch between lattice and grid display types. The *lattice* display type “pushes up” the intersections of the grid lines to the relative height of the elevation value stored for each location. The *grid* display type “pushes up” each cell centered on the data value to its relative height.

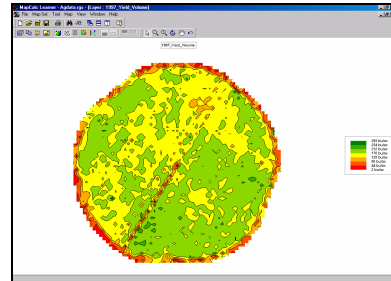
 Toggle on and off the *Data Type* button to switch between continuous and discrete data representation. This will take awhile as each elevation value is being represented as a separate symbol. *Continuous* data encompass a range of values including decimal fractions, such as elevation or soil nutrient maps. *Discrete* data contain integer values that represent distinct categories, such as land use or soil type maps.

 Toggle on and off the *3D View* button to switch between 2D and 3D display.

 Spend a few moments switching on and off the *3D View*, *Use Cells* and *Layer Mesh* buttons and note the changes in the map display.

### 2.3.2 Thematic Mapping

 Click on the *Binocular* button and select the **1997\_Yield\_Volume** map from the listing of map layers and press the *View* button.




Confirm that the display is a Continuous 2D Lattice with Layer Mesh off.



As you *Mouse-Over* a map display the map value at each location is reported as the cursor moves.

Column Name	Value	Units
Latitude	32.330753	
Longitude	-142.030587	
Vent_Shallow_Conductivity	18.2	me/m
Vent_Deep_Conductivity	25.8	me/m
Order_III_Soil_Survey	2 - BIB	
1997_Yield_Volume	153	bu/ac
1997_Yield_Mass	10.819	lb/ac
1997_Yield_Moisture	17.8	%
1996_Fall_P	7.07	pphm
1996_Fall_K	1.95	pphm
1996_Fall_3DM	1.08	%
1996_Fall_pH	7.58	


Double-click anywhere in a map display to pop-up the *Data Inspection* tool that reports the values for all map layers at the current cursor position (drill-down).

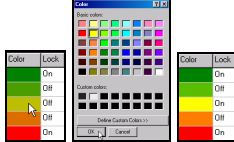

 Click on the *Shading Manager* button (or right-click on the display and select *Shading Manager*) to pop-up the map legend editing window.


Min	Max	Count	Area	% Gidded Area	Color	Lock
254	295	3	0.172	0.001	Green	On
212	254	64	3.67	1.9	Light Green	Off
170	212	1514	66.9	46	Yellow	Off
128	170	1204	68.1	37	Light Yellow	Off
86	128	239	13.7	7.3	Orange	Off
44	86	157	9.01	4.8	Red-Orange	Off
2	44	108	6.2	3.3	Red	On

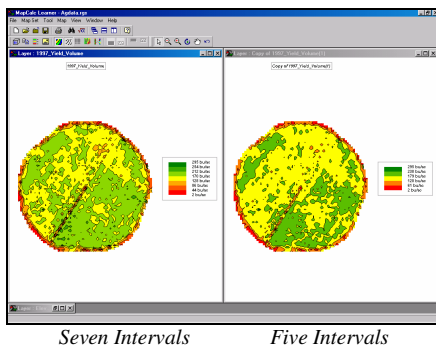
Note that there are 7 contour intervals from 2 to 295bu/ac and that each interval step is 42. Also note that the 170 to 212 interval covers most of the field (1514 cells for


86.9 acres or 46%). The interval is assigned a light green color.



 Click on the *Binocular* button and select the **1997\_Yield\_Volume** map from the listing of map layers and press the *Clone View* button.


  Change the *Number of Ranges* from 7 to 5 and note the change in the listing of the contour intervals. Change the *Color Ramp* so the yellow inflection is at the 3<sup>rd</sup> interval by 1) clicking **Off** the *Lock* by 4<sup>th</sup> interval, 2) clicking **On** the color block at the 3<sup>rd</sup> interval, 3) selecting **Yellow** from the color pallet and pressing **OK**. Press **OK** in the shading manager window to finalize the map display.

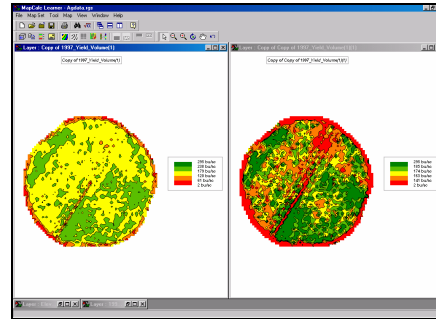
 Click on the *Tile Vertically* button to display all of the open map displays. Minimize the Elevation map so the two displays of the yield data are side-by-side. Note the differences between the two maps—spatial patterns appear to have changed yet the data is the same.



 Click on the *Binocular* button and select the **1997\_Yield\_Volume** map from the listing of map layers and press the *Clone View* button to create another copy of the same data.

  Access the *Shading Manager* and change the *Calculation Mode* from *Equal Ranges* to **Equal Count**. Note the changes in the contour intervals. In this instance the contour intervals are defined by equal percentages of the map area (approximately 37 acres each) with the min/max range for each interval varying widely.


 Place the *Equal Ranges* and *Equal Count* displays side-by-side—note the differences in the map appearance, yet the data are the same.




Min [y]	Max [x]	Count	acres	% Gridded Area	Color	Lock
236	295	11	0.631	0.33	On	Off
179	238	1007	57.8	31	Off	Off
120	179	1837	1.05	56	Off	Off
61	120	271	15.6	8.2	Off	Off
2	61	163	9.35	5	On	Off


  

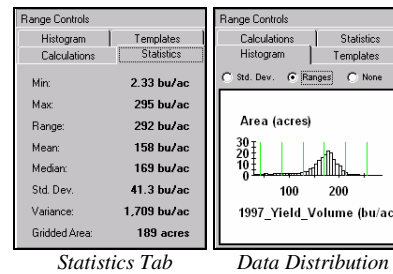
Min [y]	Max [x]	Count	acres	% Gridded Area	Color	Lock
185	295	676	38.8	21	Off	Off
174	185	664	38.1	20	Off	Off
163	174	636	35.9	19	Off	Off
141	163	664	38.1	20	Off	Off
2	141	659	37.8	20	On	Off



 Maximize the *Equal Count* map by clicking on the *Box* button in the upper-right corner of the display window. Note that each map display is contained in its own window and standard Microsoft Windows procedures are used.

### 2.3.3 Map Summary Statistics and Charts

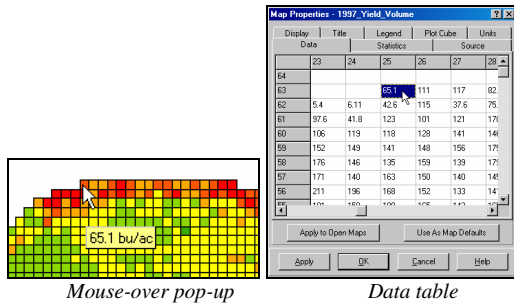
 Click on the *Binocular* button and select the **1997\_Yield\_Volume** map to restore the original display of the yield data.

 Click on the *Shading Manager* button then the *Statistics* tab for a summary of the map data comprising the yield map. Click on the *Histogram* tab and select the *Ranges* option for a plot of the data distribution indicating the contour interval breaks. Click *OK* to dismiss the shading manager.



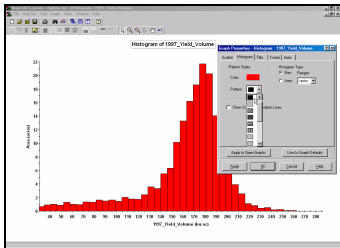
  Click on the *Layer Mesh* and the *Use Cells* buttons to generate a 2D grid display of the data.

Mouse-over the upper-left grid cell in field and note the map value stored at this location (65.1bu/ac).



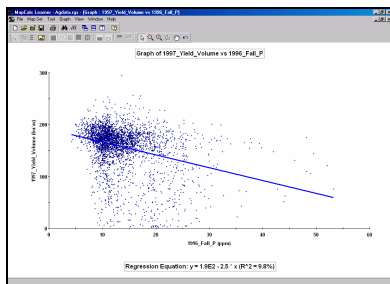
From the Main Menu, select **Map** → **Properties** then click on the *Data* tab to view the data table. Note that the top-right entry in the table (column= 25, row= 63, value= 65.1) corresponds to the upper-left cell in the map display.

From the Main menu, select **Map Set** → **New Graph** → **Histogram** → **1997\_Yield\_Volume** to generate a histogram plot of the yield data.



Right-click on the plot, then select *Properties* and set the color to *red* and the pattern to *solid*.

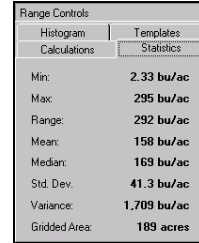
From the Main menu, select **Map Set** → **New Graph** → **Scatterplot** → specifying **1996\_Fall\_P** as the X-axis and **1997\_Yield\_Volume** as the Y-axis to generate a scatter plot of the yield data versus phosphorous.




Note that a linear regression line is fitted to the data to help visualize the overall numerical relationship between the two map variables (see Topic 6 for more information).

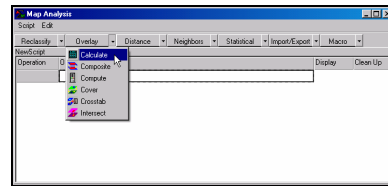
### 2.3.4 Map Normalization

Generate a statistical summary report of the yield map data by right-clicking on the map display and selecting **Shading Manager** → **Statistics** tab.

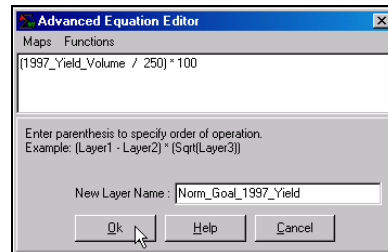


Note the *Min* (2.33), *Max* (295), *Mean* (158) and *Std. Dev.* (41.3) values for the yield map data.

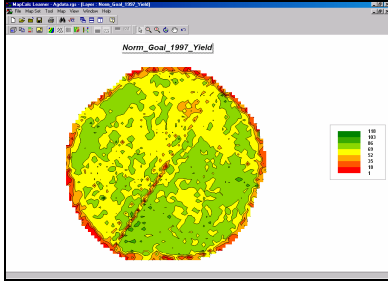
 Access the analytical side of MapCalc by clicking on the *Map Analysis* button.




Select the **Overlay** group of operators and choose the **Calculate** operation.

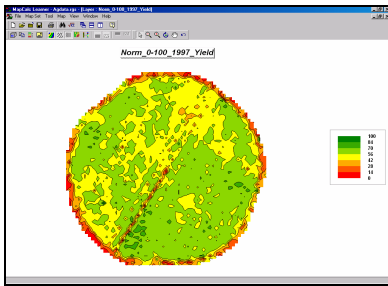
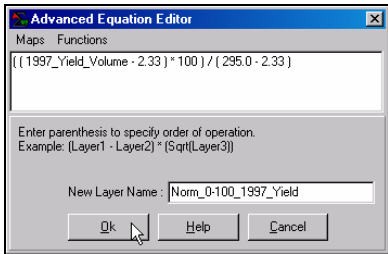


Enter the Goal Normalization equation (see above) by 1) selecting **Maps** → then choosing **1997\_Yield\_Volume**, 2) selecting **Functions** → **Operators** → “/” for divide, 3) enter **250** for the goal, 4) insert parenthesis around this portion of the equation, 5) enter “\*” for multiply, 6) enter **100**, 7) specify **Norm\_Goal\_1997\_Yield** as the *New Layer Name*, and press the *OK* button to submit the command and create a new map layer. Close the *Map Analysis* window to view the full map.




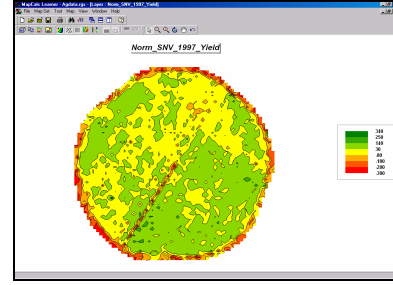
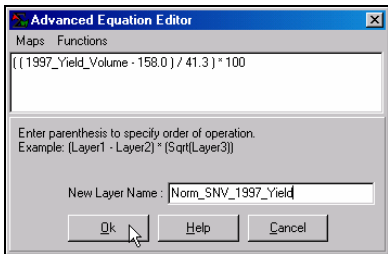
Note that the yield data has been normalized as a percent of the 250bu/ac goal and ranges from 1 to 118% with most of the low values concentrated around the edge of the field and along the access road in the southwest portion.

 Repeat the map analysis process to generate a 0-100 normalized map.




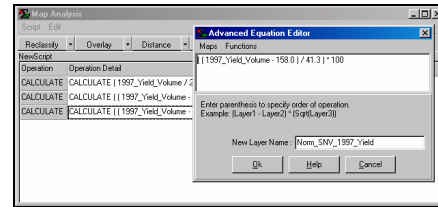
Note that the yield data has been normalized to the range 0 through 100.

 Repeat the map analysis process to generate a Standard Normal Variable (SNV) normalized map.



Note that the yield data has been normalized as percent deviation from the average yield value.

 Re-access the *Map Analysis* window and note the session log contains the three commands you entered. Double-click on any of the commands causes it to be recalled so you can edit it.



Additional editing capabilities include Copy, Paste, Delete, Reordering, Notes and Pauses to extend the session log into a full featured *Macro Builder* for creating your own processing sequences or spatial models that can be stored and reused in either batch or interactive modes by others.

You can exit the program by selecting **File** → **Exit** or by clicking on the "X" in the upper-right corner of the MapCalc program window. If you want to save your work, specify a new file name, such as *AgData\_Topic2\_exercises.rgs*. Each exercise set assumes you will start with the basic *AgData.rgs* data set and this database will become cluttered with exercise maps if you save your results to it each time.