

Exercise #3

Building a Thematic Error Matrix from a Stratified Random Sample

(an exercise to accompany the AmericaView Video Lessons on Assessing the Accuracy of Remotely Sensed Data)

This exercise accompanies the AmericaView Accuracy Assessment Lessons prepared by Dr. Russell G. Congalton, University of New Hampshire and Director of NewHampshireView.

Objective: To use a stratified random sampling approach to create a thematic error matrix and see the impact of the sampling scheme on the assessment. A comparison with Exercise #2 is enlightening.

This exercise is a follow-on from Exercise #2. At the end of this exercise, you will be asked to analyze and compare the results of Exercise #2 with the results achieved here.

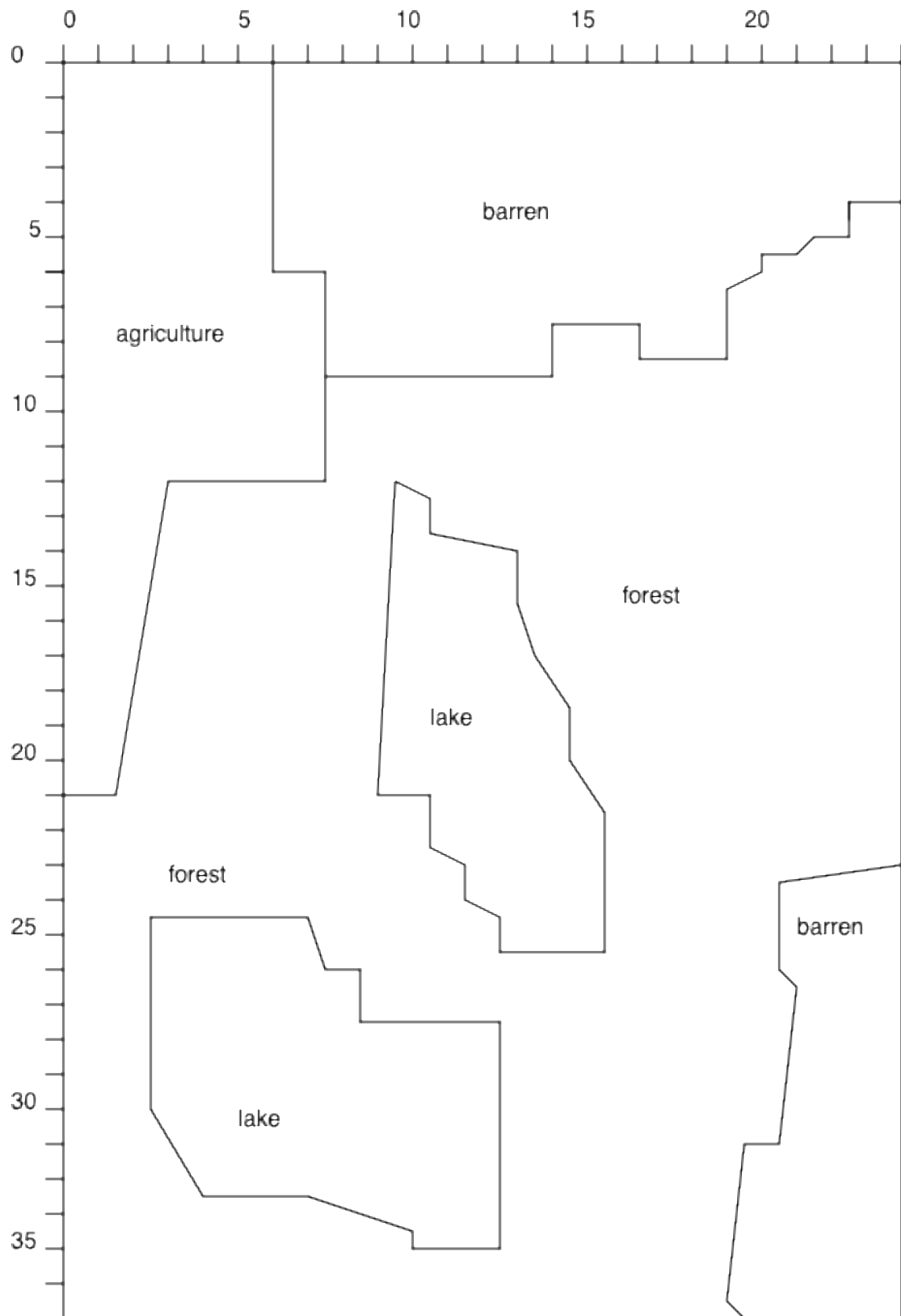
Instructions: You will select **80 stratified random samples (20 samples per reference data map class)** to compare the map to the reference data and create an error matrix as follows:

1. You have been provided with a map that was generated from remotely sensed data and a reference data set. (The goal here is to see the impact of stratified random sampling on your assessment and so we will not worry about the sample unit size here)
2. Use a random number table or better yet, a computer application, to generate a stratified random sample with **20 samples per reference data map class**. The random number generator needs to provide an x and a y coordinate. The x coordinate must range between 0 – 24 and the y coordinate between 0 – 37. Any duplicate pairs should be discarded.
3. Plot the locations of the x, y samples on both map and the reference data and then use this information to create a table like this:

Coordinate (x,y)	Map Label	Reference Label
5, 10	barren	barren
13, 5	forest	lake
30, 5	lake	lake
Etc.	Etc.	Etc.

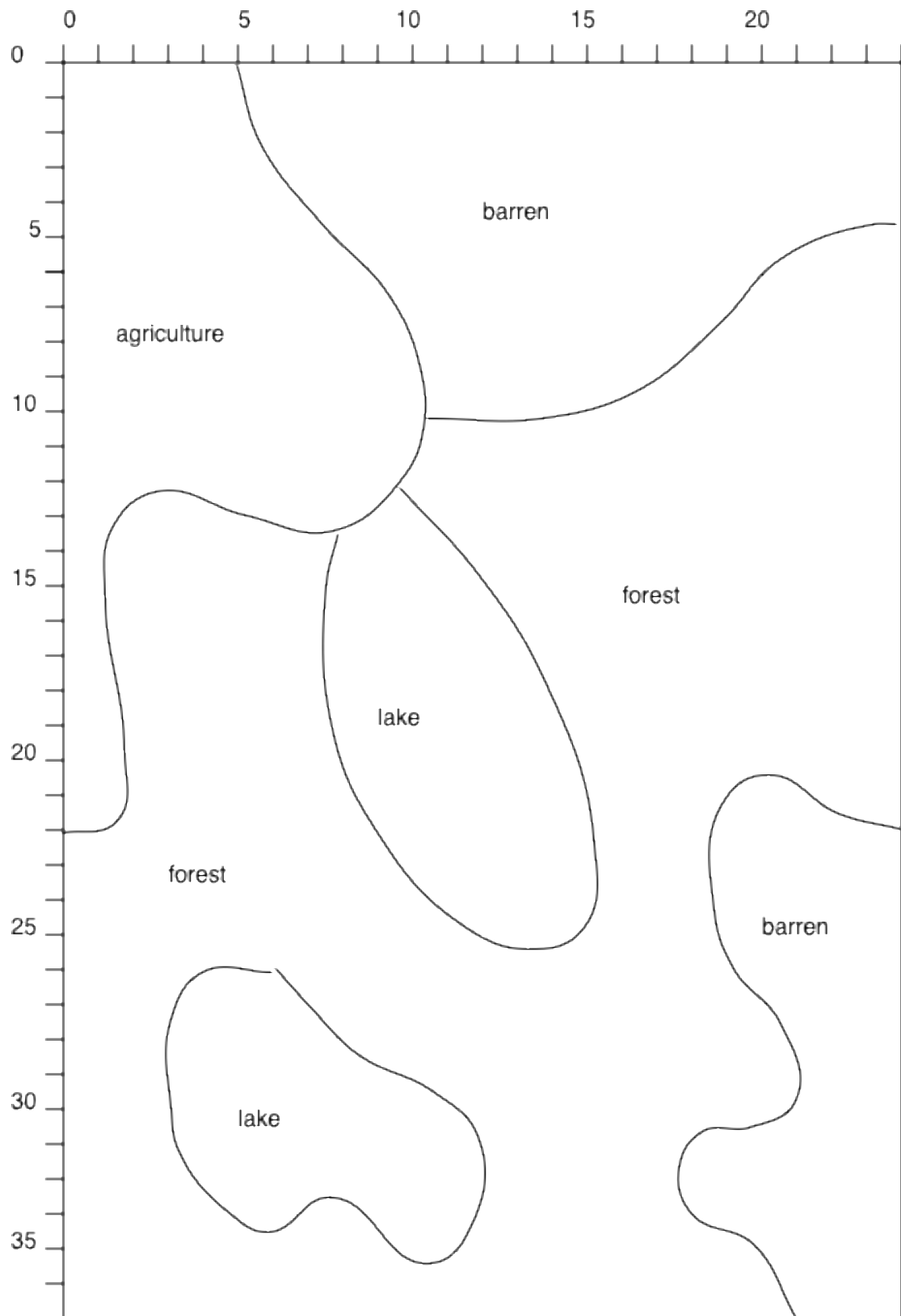
Remember: You should collect only 20 samples per reference data map class.

4. From this table, conduct the thematic accuracy assessment by compiling an error matrix for this 4 land cover class map (the matrix should be 4 x4 (Agriculture, Barren, Forest, and Lake) with the reference data on the x axis and the map on the y axis). The column totals in the error matrix should each be 20 since you have used a stratified random sampling approach with 20 reference data samples per map class.
5. Once the error matrix is compiled, compute the overall, producer's and user's accuracies.



CLASSIFIED DATA

Exercise #3 – Stratified Random Sampling.

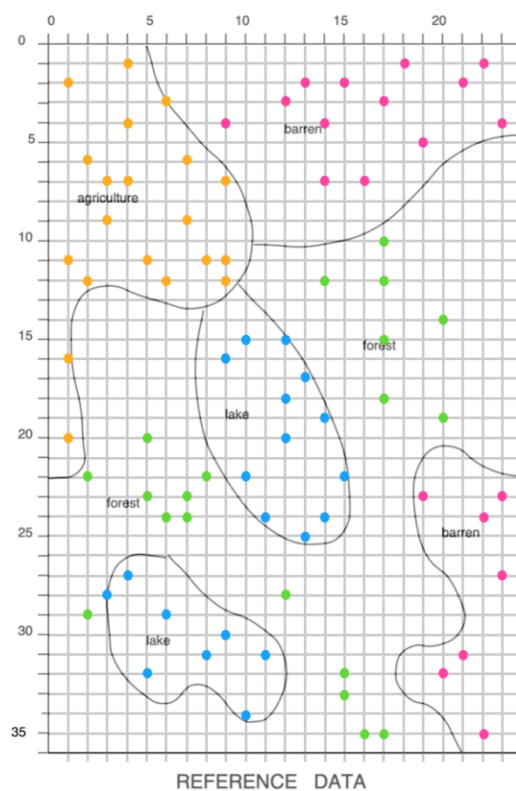
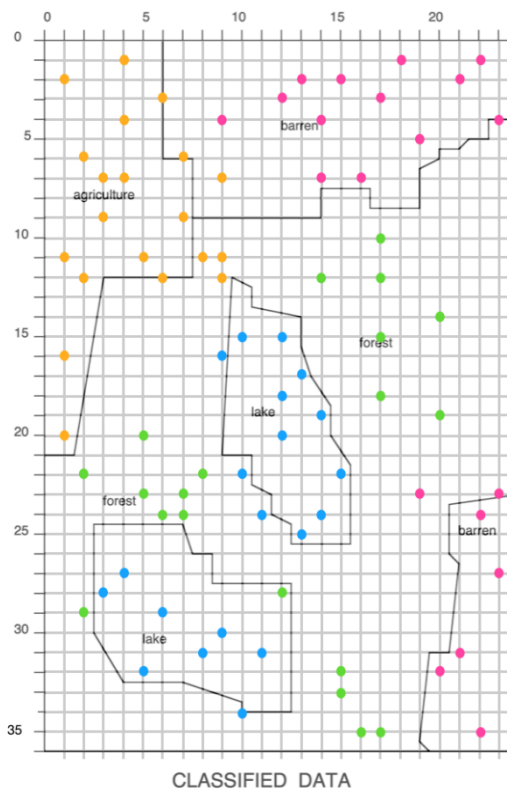


REFERENCE DATA

Exercise #3 – Stratified Random Sampling.

Observations to think about:

1. You now have the same number of reference data samples per map class in the error matrix regardless of the proportion of the map class in the map. How does having this same number of samples impact the accuracy?
2. Think about the advantages of using a stratified random approach over a simple random one. Are there situations when one approach is more appropriate than the other? Try to think of a number of scenarios where one approach would be better than the other.
3. Both these exercises (2&3) used only 80 total samples which is not sufficient for a valid assessment, but are enough to demonstrate some sampling issues. What would happen if you had 50 samples per map class or 200 samples?
4. Below are the results of one version of this exercise (it will be different than yours because the results are only a sample). Compare this example to your results and think why they may be different or similar.



Reference Data

Error Matrix	Forest	Lake	Barren	Agriculture	Total
Forest	19	4	3	4	30
Lake	1	16	0	0	17
Barren	0	0	17	2	19
Agriculture	0	0	0	14	14
Total	20	20	20	20	80

Map Data

Overall Accuracy = $66 / 80 = 82.5\%$

Producers Accuracy:

F = $19/20 = 0.95$ or 95%

L = $16/20 = 0.80$ or 80%

B = $17/20 = 0.85$ or 85%

A = $14/20 = .70$ or 70%

Users Acc

F = $19/30 = 0.63 = 63\%$

L = $16/17 = 0.94 = 94\%$

B = $17/19 = 0.89 = 89\%$

A = $14/14 = 1.0 = 100\%$