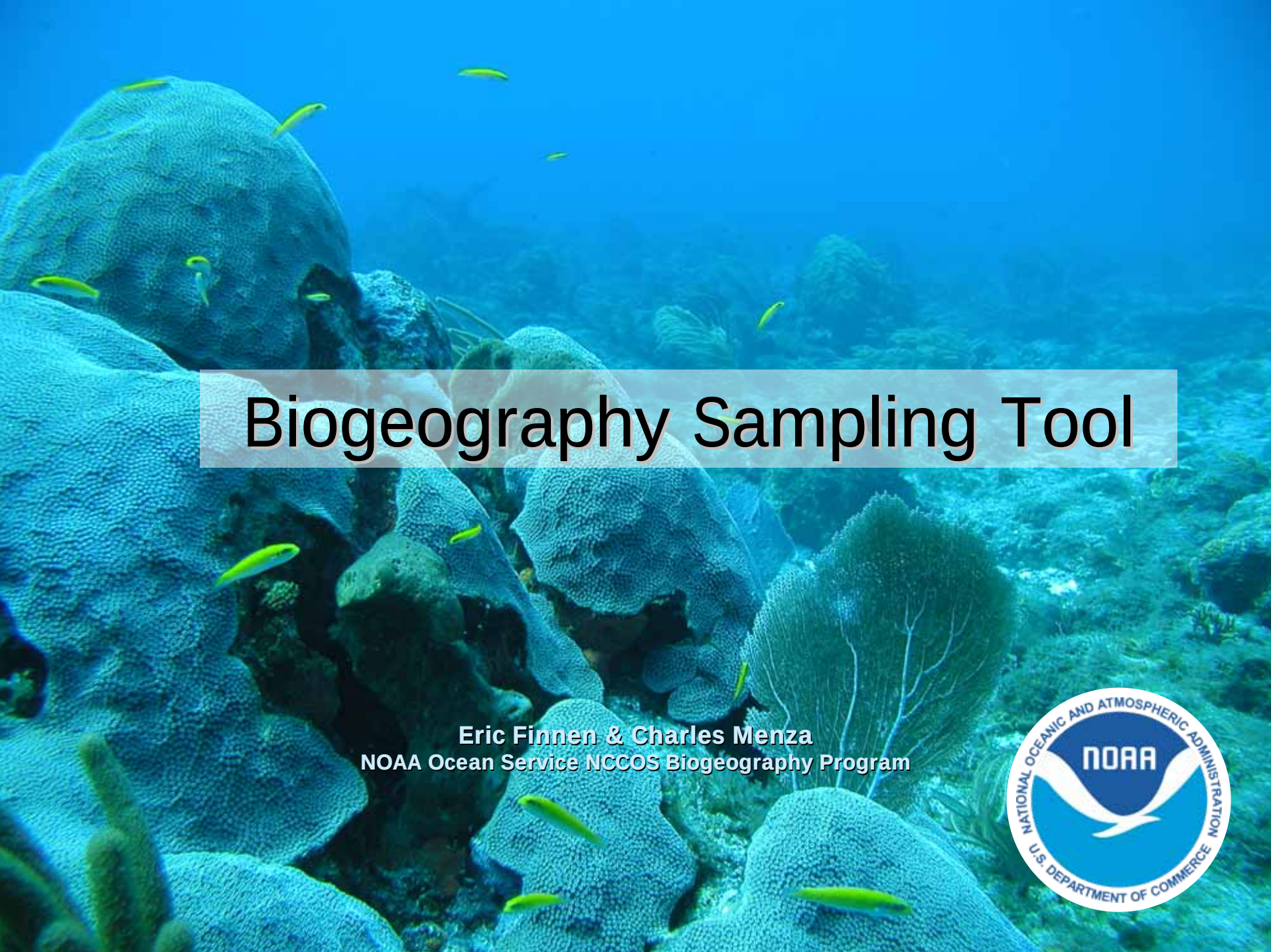


Biogeography Sampling Tool

Eric Finnen & Charles Menza
NOAA Ocean Service NCCOS Biogeography Program

Outline

- History of Project
  - Random Point generator which then evolved into a “smarter version” to incorporate statistical sampling techniques

Scope

- Simple tool to derive sampling locations for fieldwork using multiple approaches and analyzing the results.

But, what can it do for me??

- Can answer questions such as
  - Given a fixed budget, how should sample be allocated to get the most precision from a stratified sample?
  - Given a fixed sample size, what is the most precision that I can get from a stratified sample?
  - What is the smallest sample size that will provide a given level of survey precision?
  - Given a particular sample allocation plan, what level of precision can I expect?

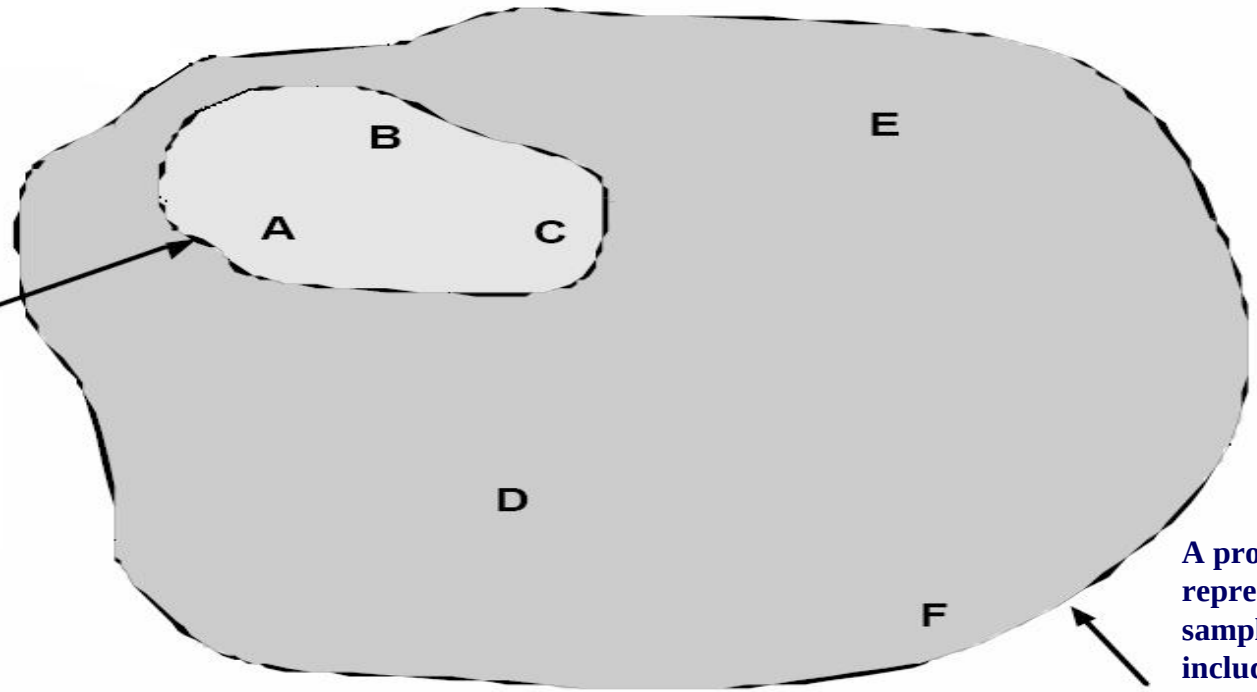
Mission

- Determine which Sampling Approach is best. Generally constraints are
  - Sampling Objectives
  - Cost
  - Expertise
  - Available Data

Mission (continued)

- Representative sampling may be considered as the measure of the degree to which data accurately and precisely represent a characteristic of a population, parameter variations at a sampling point, a process condition, or an environmental condition [American National Standards Institute/American Society for Quality Control (ANSI/ASQC) 1994].

Example : Sampling Bias



If only areas A, B & C were sampled they would miss the other strata that included D, E & F

A proper representative sample would include all appropriate strata

Why do We Care?

- We want to avoid sampling biased
  - A **biased sample** is one that is falsely taken to be typical of a population from which it is drawn.

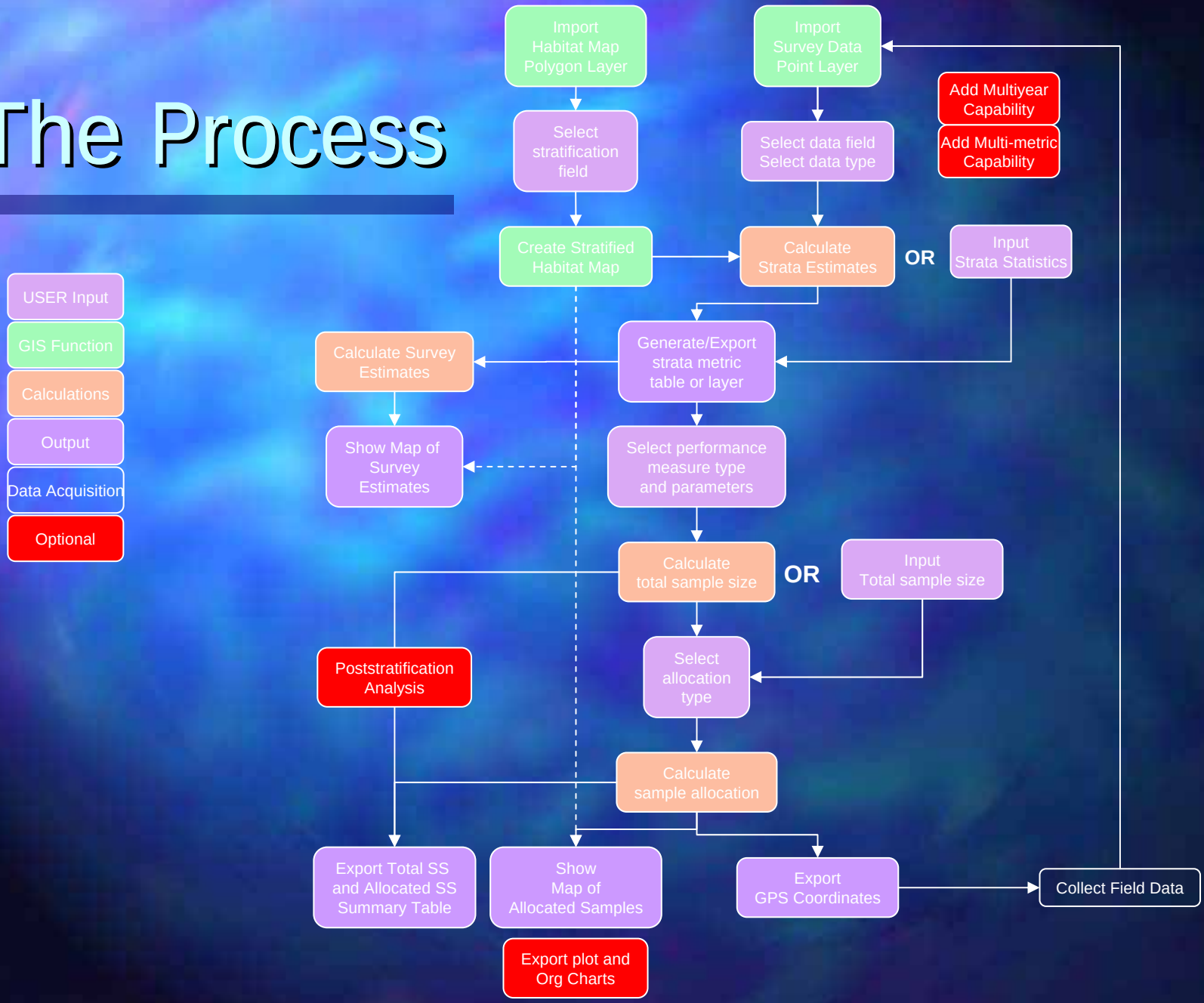
Basics

- Cochran Sampling Technique
  - With *simple random samples*, every possible sample has the same probability of being selected.
  - With *stratified random sampling*, the population is divided into strata and a simple random sample is selected from each stratum.

The Basics

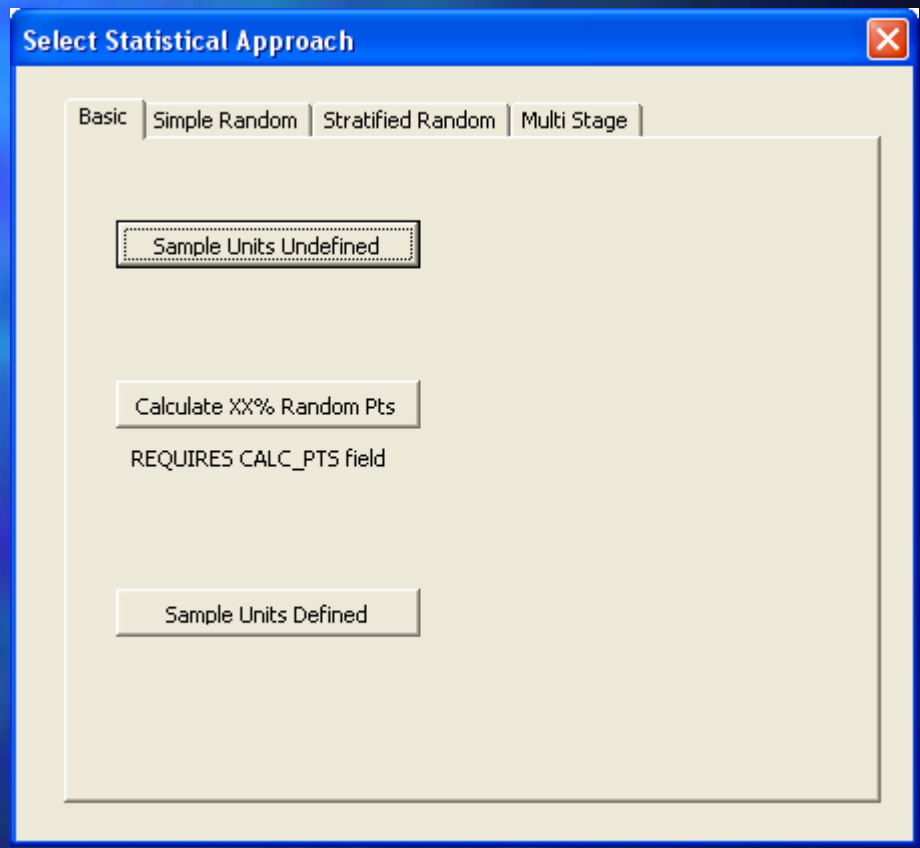
- We want to sample a population which is delimited by an area what do we do??
  - Establish Survey Objectives
  - Determine mean & variances (optional)
  - Determine what precision could be afforded

The Process


Import Habitat Map Polygon Layer
Import Survey Data Point Layer
Add Multiyear Capability
Add Multi-metric Capability
Select stratification field
Select data field
Select data type
Create Stratified Habitat Map
Calculate Strata Estimates
OR
Input Strata Statistics
USER Input
GIS Function
Calculations
Output
Data Acquisition
Optional
Calculate Survey Estimates
Generate/Export strata metric table or layer
Show Map of Survey Estimates
Select performance measure type and parameters
Calculate total sample size
OR
Input Total sample size
Poststratification Analysis
Select allocation type
Calculate sample allocation
Export Total SS and Allocated SS Summary Table
Show Map of Allocated Samples
Export GPS Coordinates
Collect Field Data
Export plot and Org Charts

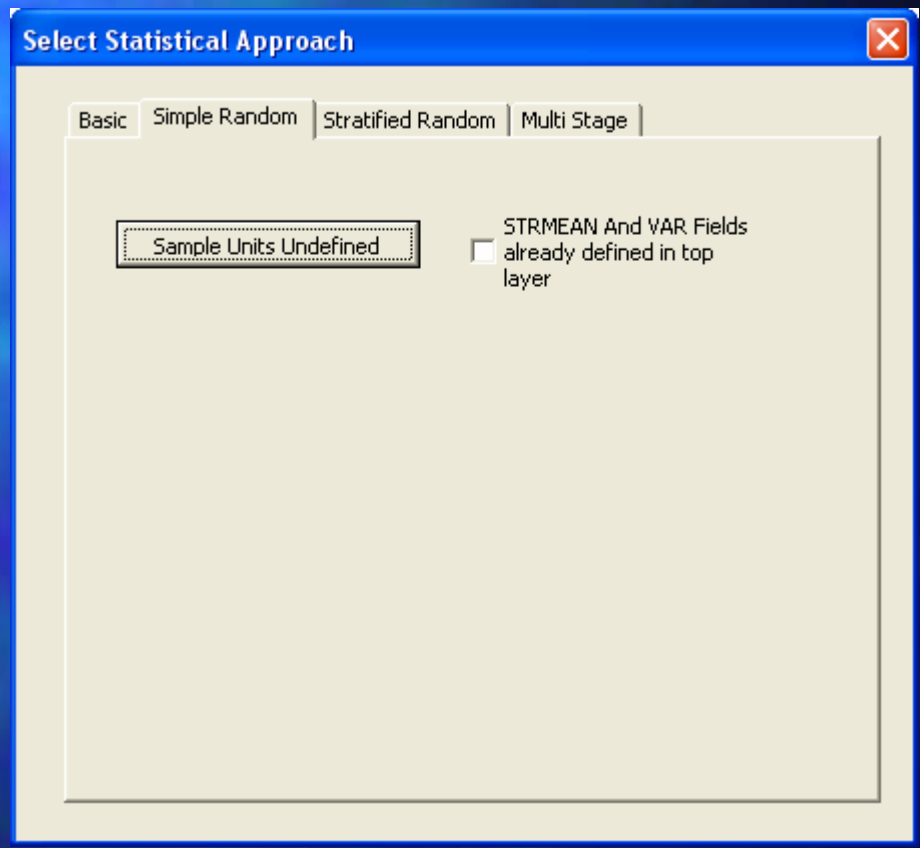
The Basics

- Basic – Sample Units Defined
- Basic – Sample Units Undefined
- Basic – Field pre-populated with points field



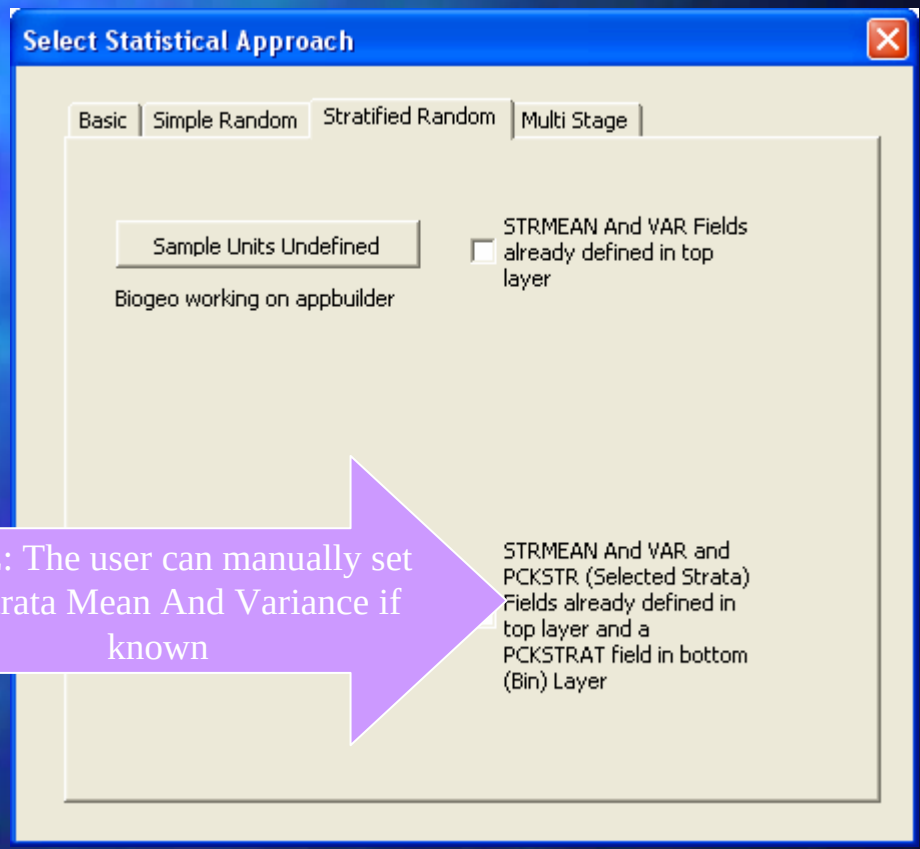
The Basics

- Simple Random – Sample Units Undefined



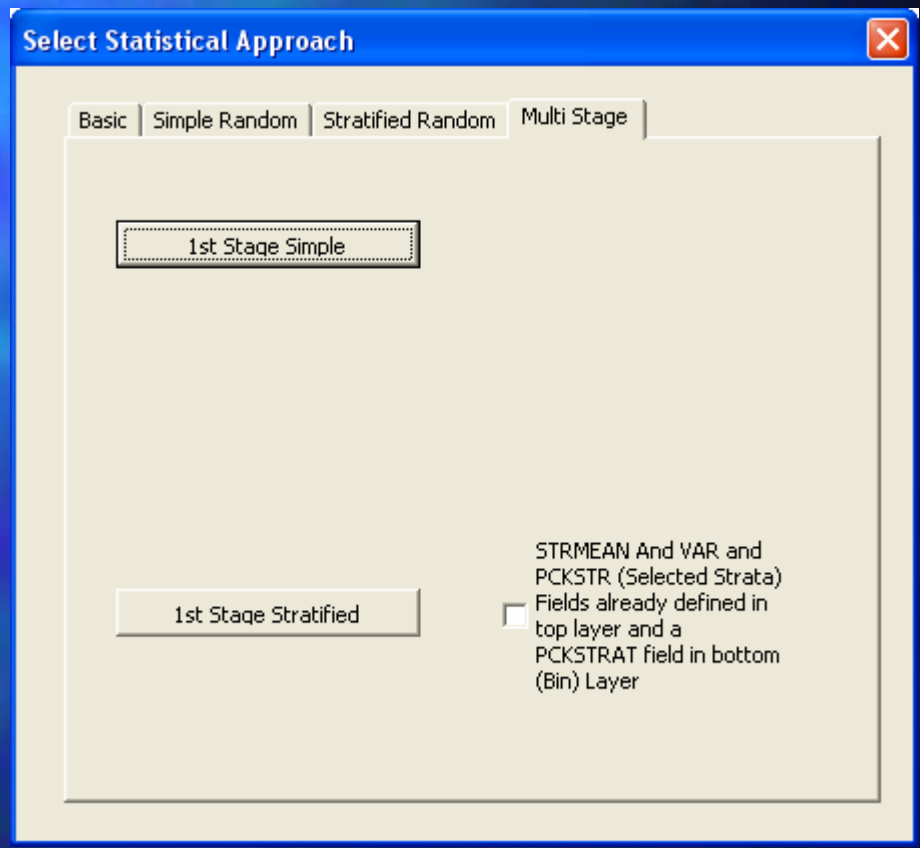
The Basics

- Stratified Random – Sample Units Defined
- Stratified Random – Sample Units Undefined



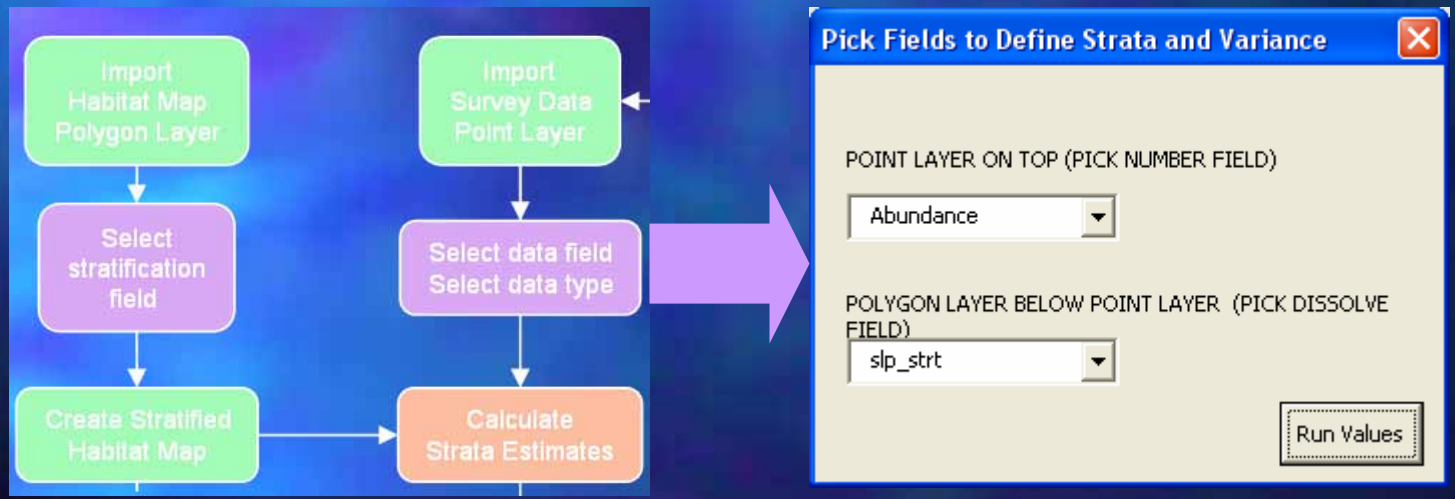
The Basics

- Multistage – Simple Random
- Simple Random – Sample Units Undefined



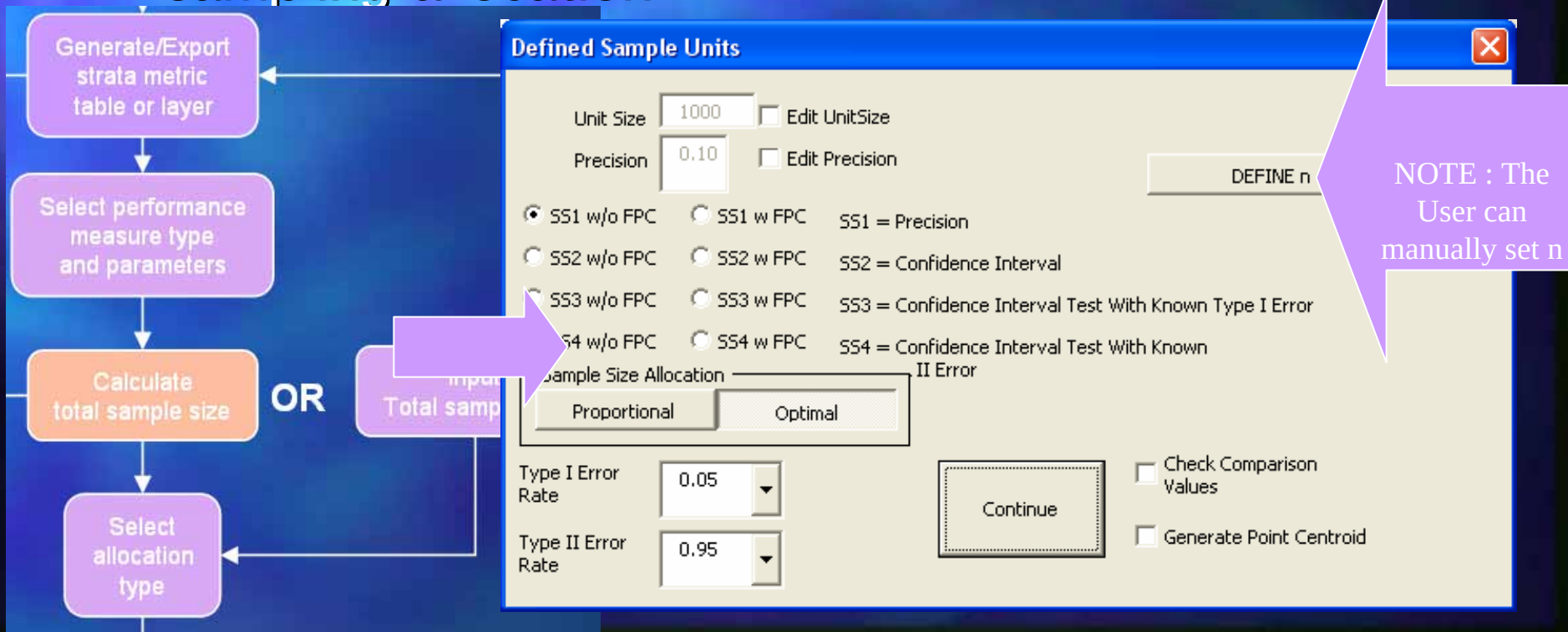
Where Do We Start?

- Step 1 Generate Strata Mean and Variance
  - Our example is Stratified Random – Sample Units Undefined
  - survey data (points) and appropriate strata layer (polygon)
  - Select point layer and polygon layer which we want to use to determine strata mean and variance
  - Usually computed using pilot data or data from precious surveys



Where Do We Start?

- Step 2A
- Determine appropriate sampling allocation



The Math

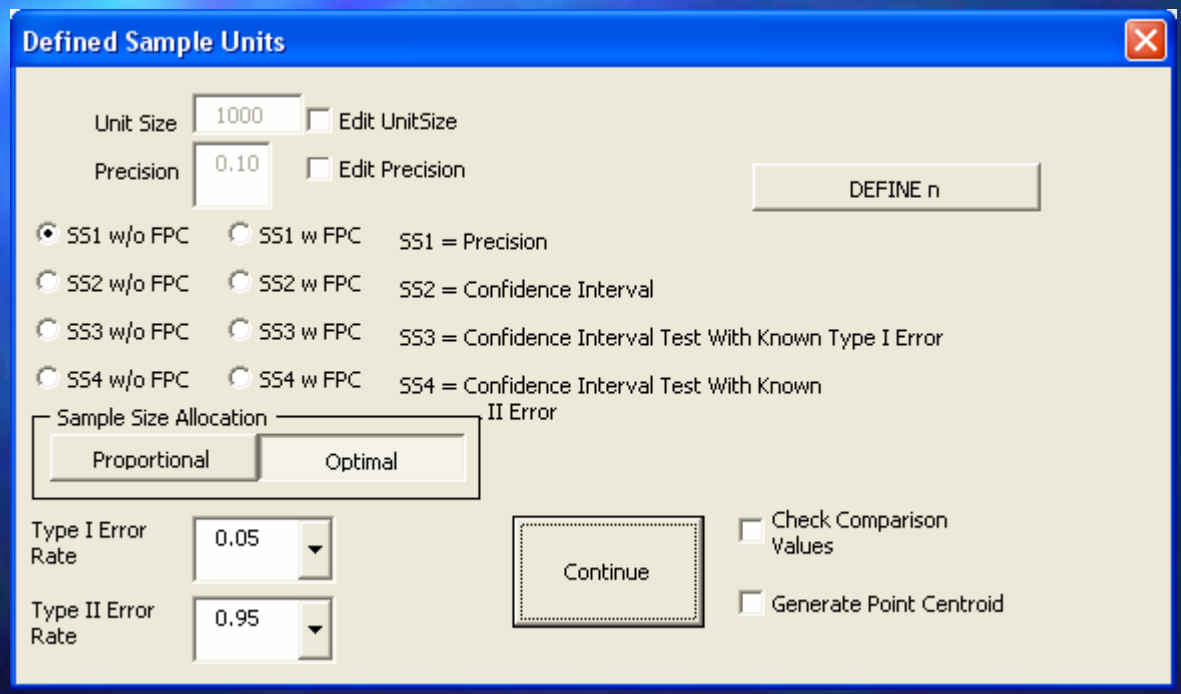
- Step 2B Decide on Allocation Methods: Proportional or Optimal
  - Proportional to area
    - Best if no variance data is available
    - Samples allocated by strata area
  - Optimal
    - Also called the Neyman Allocation Method
    - Samples allocated by area and metric variance

$$n_h = n \cdot \left(\frac{W_h}{\sum_h W_h} \right)$$

$$n_h = n \cdot \left(\frac{W_h s_h}{\sum_h W_h s_h} \right)$$

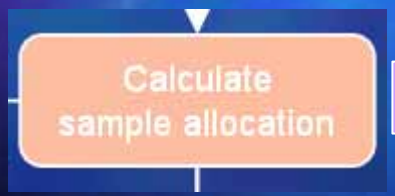
Where Do We Start?

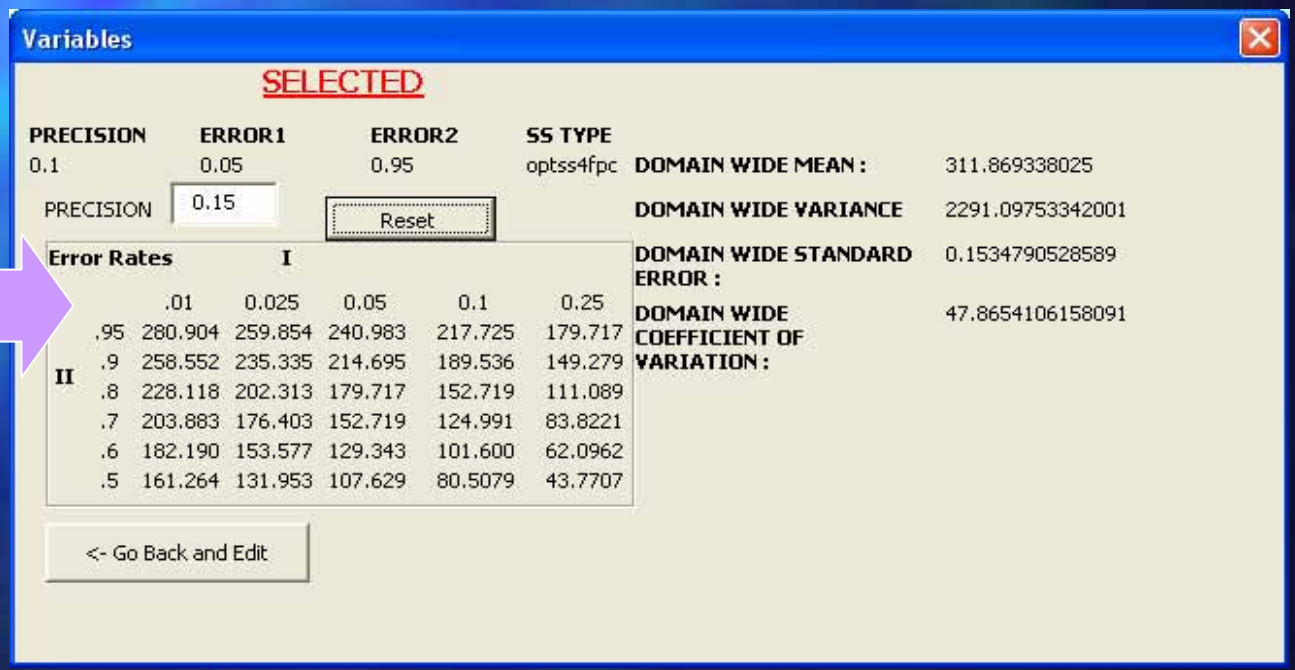
- Step 2C (optional)
- Check Comparison Values



Where Do We Start?

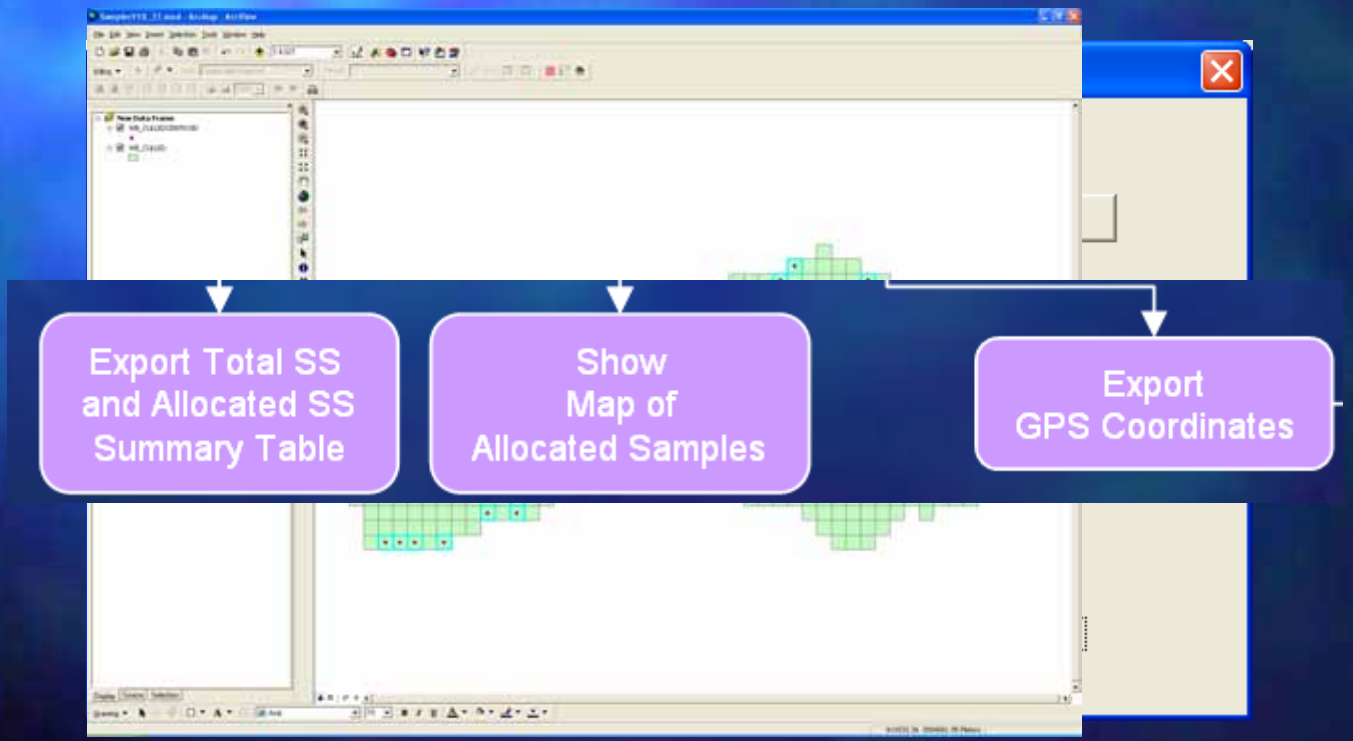
- Step 2C (optional)
- Enter Precision Values to see results





Where Do We Start?

- Step 2D (optional)
- Select point centroid option



Results Summary

- So for the statistically challenged
  - If all things spatially (area) are equal between two strata a higher calculated *nh* means that the areas are more heterogeneous
  - If *nh* is lower the area is more homogeneous

Conclusion

- So why is this application helpful/useful?
  - This tool provides a much needed spatial component to Cochran's Sampling methodology.
  - The user can quickly run numerous scenarios with varied sampling strategies, precision, unit size and error rates to arrive at the best sampling approach to meet his needs
  - Allows efficient planning of resources, prevents user from oversampling

Other Sample Uses

- Law Enforcement
- Environmental
- Defense
- Planning
- And many others..

Still In Concept

- Given a fixed sample size, how should sample be allocated to get the most precision from a stratified sample?
- What is the minimum cost to achieve a given level of survey precision?
- As Nh is maxed determine at what point you get the highest precision possible?
- Handle binomial data

Questions

Eric.finnen@noaa.gov

Charles.menza@noaa.gov