

Sampling Design Tool for ArcGIS Pro

National Oceanic and Atmospheric Administration (NOAA)
National Centers for Coast Ocean Science (NCCOS)

TABLE OF CONTENTS

Overview.....	2
Requirements	2
Software	2
Data	2
Key Features	2
Installation and Setup	3
Create Points	4
Create Simple Random	4
Create Stratified Random	4
Create Two-Stage Random	4
Select Features.....	5
Select Simple Random	5
Select Stratified Random	5
Analyze Samples	5
Analyze Simple Random	5
Analyze Stratified Random	6
Sample Analysis window	7
Acknowledgements.....	9
Examples.....	10
Create Stratified Random points.....	10
Refining Sample Selection using previously collected data.....	14

Overview

The Sampling Design Tool for ArcGIS provides a means to effectively develop sampling strategies in a geographic information system (GIS). The tool was produced as part of an iterative process of sampling design development, whereby existing data informs new design decisions. The objective of this process, and hence a product of this tool, is an optimal sampling design which can be used to achieve accurate, high-precision estimates of population metrics at a minimum of cost. The tool can be used to sample any type of population defined in space, be it coral reefs or corn fields.

Requirements

Software

ArcGIS Pro 3.0 - SamplingDesign3.esriAddinX

Data

Feature classes in geodatabases can be used in all tools. Shapefiles can only be used in the Create Simple Random and Select Simple Random tools. Hosted feature layers can be used in the Create and Sample tools, but expect performance to be slower. Hosted feature layers cannot be used in the Analysis tools.

Key Features

Create new point samples - select random points within a polygon layer

Select samples from a GIS layer - select random samples from existing point, line or polygon sample frames

Sample in a GIS - sample spatial layers from within an existing GIS project, eliminating the need to transfer data between GIS and statistical programs, and easily incorporating inherently spatial data (e.g., benthic habitat maps, administrative boundaries)

Random sampling - eliminates sampling selection biases and corresponding criticisms encountered when samples are selected non-randomly

Multiple sampling designs – samples can be selected using simple, stratified, and two-stage sampling designs

Sample unit-based sampling – points or polygons are selected from a sample frame

Area-based sampling – random points are generated within a polygon

Analysis – previously collected data can be used to compute sample size requirements or efficiently allocate samples among strata

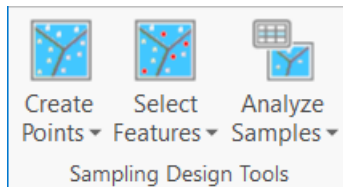
Computations –mean, standard error, confidence intervals for sample data and inferences of population parameters with known certainty

Installation and Setup

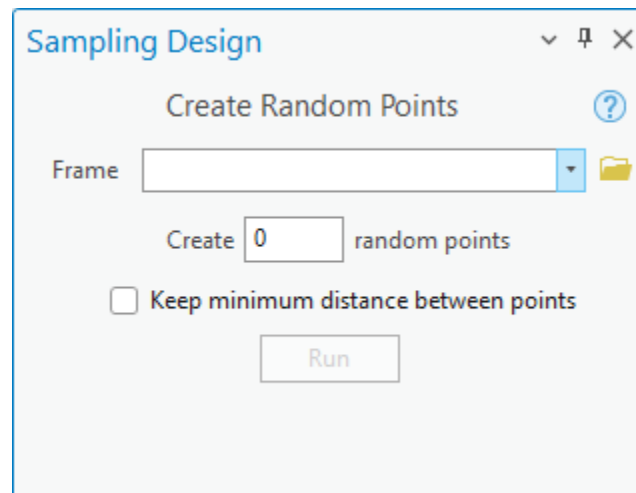
In a file manager, double click on the file SamplingDesign.esriAddInX to start the ESRI ArcGIS Add-In Installation Utility. Once installed on your computer, the Sampling Design Tool will appear as a tab on the [ribbon](#).

For more information about ArcGIS Pro Add-ins, please visit the [documentation page](#).

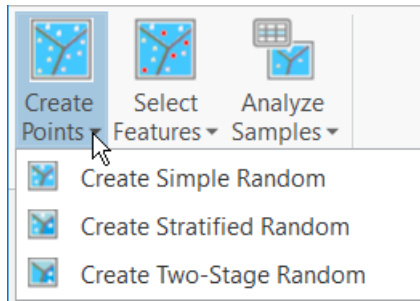
To start a process, click on the Sampling Design Tool tab and select a tool.



The selected tool will open a dockable pane.



Create Points



Random points are commonly used to sample a population defined by an area on a map (e.g., political jurisdiction, marine protected area, habitat). By taking measurements at points distributed randomly throughout the population, an unbiased survey of the population can be undertaken. This procedure is a form of area-based sampling and is aided by data in a GIS. The Sampling Design Tool offers 3 ways to generate point samples: simple random, stratified

random and two-stage random. The choice of which method to use will depend on survey objectives and types of available data.

Create Simple Random

This procedure generates randomly placed points within a population defined by a polygon dataset. This procedure is typically used when there is little information available for the population, covariates or spatial patterns of intended measurements. If the area for which random points will be assigned consists of multiple polygons, the distinct polygons are automatically [dissolved](#) together and treated as one.

Create Stratified Random

This procedure generates randomly placed points within mutually-exclusive subareas of a population, or strata. Strata are identified by choosing an attribute (column) in the sample frame's attribute table which distinguishes polygons amongst different strata. All polygons with the same value are considered part of the same stratum. Several methods can be used to allocate points among strata. The user can set the number of points to each stratum themselves, assigned proportional to the area of each stratum (i.e. larger strata get more points) or assigned equally among all strata. If either of the two latter methods is used, the user must identify the total number of sample units to allocate. A stratified sample is superior to a simple random sample if a polygon dataset separates a heterogeneous population into internally homogenous groups (e.g., benthic habitat map).

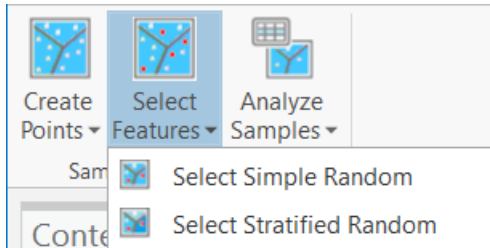
Create Two-Stage Random

This procedure samples in two simple random stages. First, a sample of primary sampling units (PSUs) is selected. PSUs must be defined by a polygon dataset. Then random points are placed in each selected PSU. The user selects the number or percentage of PSUs to be sampled and the number of points to be placed in each PSU. This procedure is generally used when the variance of a measured metric is highly variable at fine spatial scales.

The output of all create point sample procedures is a point dataset. The user must select where the dataset will be saved. The attribute table of all output datasets will contain their X and Y geographic coordinates as defined by the data frame's coordinate system.

The results of using the option “Keep a minimum distance between points” will be dependent on the scale of the map, the distance and units chosen, and the timeout period. If not all points can be created within the timeout period, a message box will report how many points were able to be randomly placed. Choosing a minimum distance can be used to mitigate spatial autocorrelation. It may also affect sampling bias.

Select Features



A sample can be selected from a sample frame (point, line or polygon dataset) using simple random or stratified random sampling. In these procedures, a selection set is created from the list of sampling units.

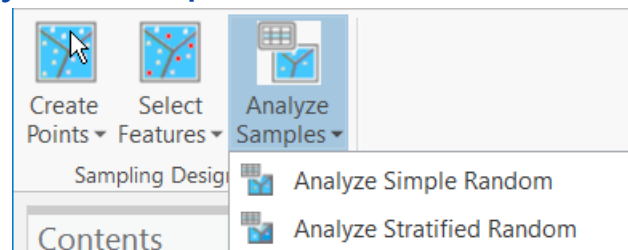
Select Simple Random

This procedure randomly selects a number or percentage of sample units.

Select Stratified Random

This procedure selects sample units from different strata. Strata are identified by choosing a numeric or text attribute (column) in the sample frame's attribute table. All sample units with the same value are considered part of the same stratum. Several methods can be used to allocate sample units among strata. The user can set the number of sample units to each stratum themselves, sample units can be assigned proportional to the area of each stratum (i.e. larger strata get more sample units) or sample units can be assigned equally among all strata. If either of the two latter methods are used the user must identify the number of sample units to allocate. A stratified sample is superior to a simple random sample if a polygon dataset separates a heterogeneous population into internally homogenous groups (e.g., benthic habitat map).

Analyze Samples



A common procedure when gathering data for a monitoring program is to analyze existing data to refine the sample selection procedure. This process produces more-efficient sampling designs. An analysis of past data can suggest new strata, a more efficient sample allocation among

strata, or the need for fewer (or more) samples to achieve a desired objective.

Analyze Simple Random

This procedure generates randomly placed points within a population defined by a polygon dataset. Select a sample frame (a polygon or point dataset), a sample

point layer, and a metric field (a numeric attribute of the sample layer). The metric field corresponds to a continuous numerical attribute and represents measurements taken from a population.

Analyze Stratified Random

This procedure generates randomly placed points within mutually-exclusive subareas of a population, or strata. Select a sample frame layer (a polygon or point dataset), a stratum (an attribute of the sample frame layer), a sample point layer, and a metric field (a numeric attribute of the sample layer). The metric field corresponds to a continuous numerical attribute and represents measurements taken from a population.

The Sampling Design Tool provides two main outputs: sample size requirements and sample data statistics. Sample statistics include the mean, standard error, coefficient of variation and upper and lower 95% confidence limits. The confidence interval is given by $\pm z_{\alpha} s(\bar{y}_{st})$, where z is the critical value of a Normal distribution with 95% confidence (i.e. $\alpha = 0.05$) and $s(\bar{y}_{st})$ is the square root of the standard error. Note: population estimates are only valid if the appropriate sampling design is used. Statistics are calculated using computational methods presented in Cochran (1977). Please read Cochran's (1977) [Sampling Techniques](#) for more information on analyses.

Sample Analysis window

The screenshot shows a software window titled "Sampling Analysis". It is divided into two main sections: "Sample Size Estimates" and "Statistics".

Sample Size Estimates:

- Method: t-Test (Type I/II Error) (selected)
- Finite Population Correction:
- Degree of Precision: 0.10
- Confidence Level: A table showing sample size requirements for various power and confidence level combinations.
- Highlight maximum sample size: (value: 100)
- Choose sample size: (value: 100)
- Buttons: Allocate Samples, Cancel

Statistics:

- Analysis Type: Area-Based Analysis (selected), Sampling Unit-Based Analysis
- Sampling Unit Size: 100
- Statistics Table:

Mean	52.74
Upper Confidence Limit	52.95
Lower Confidence Limit	52.53
Standard Error	1.74
Coefficient of Variation	3.29%
- Button: Export Statistics

Sample size requirements are based on desired objectives. The user must select one of four analytical approaches and a desired precision. The **Degree of Precision** sets the amount of desired agreement among repeated measurements. Precision is set as a proportion (e.g., 0.1, 0.2, or 0.5), which correspond with a percentage (e.g., 10%, 20%, or 50%). If the chosen sample frame is a polygon dataset the user must also select whether the analysis is sampling unit-based or area-based. Sampling unit-based means records in the sample frame dataset are used to define sampling probabilities (this is the

normal way). Area-based sampling indicates polygon areas and user-defined measurement plot sizes are used to estimate sampling probabilities.

Four distinct analytical approaches to examine sample variability are provided: Coefficient of variation, Confidence Interval (CI), *t*-test with various Type I errors, and *t*-test with various Type I and Type II errors (i.e. power analysis). The correct choice will depend on how the data will be analyzed.

The coefficient of variation (CV) is a standardized measure of the standard error and represents sample variability. If CV is chosen as the analytical approach, the output reflects the sample size needed for the CV to equal the user-defined precision. This is the simplest analytical approach and does not require data to fit any model assumptions.

A confidence interval is an interval estimate of a population parameter and is an alternative method of representing variability. By selecting precision and a confidence level (reciprocal of the Type I error rate), the user identifies the range of acceptable imprecision by a lower and upper bound. For instance, if precision is set to 0.10 and confidence level is set to 95%, the interval represented by +/-10% of the sample's mean would contain the population mean 95% of the time.

Two different analytical approaches are provided when samples will be compared using *t*-tests. The first approach only incorporates a confidence level whereas the second also incorporates statistical power (reciprocal of the Type II error rate). The results are presented as a matrix of sample size requirements for several conventional confidence and power rates. These tests have statistical assumptions which must be met in order for results to be valid.

To investigate the impact of the **finite population correction** (FPC), click the corresponding checkbox. The FPC provides a reduction in standard error for sampling a greater proportion of the sample frame and consequently lessens sample size requirements.

If there is a limit of how many samples can be allocated, click the **Highlight maximum sample size** checkbox. This will highlight the cells that are below this maximum threshold.

To select the sample size, click on a cell in the table Confidence Level table, then click the **Allocate Samples** button. If the Coefficient of variation approach was selected, just click the **Allocate Samples** button. The user can also manually set the sample size in the **Choose Sample Size** text box once the checkbox has been clicked.

A report of the analysis can be saved by clicking the "**Export Statistics**" button. This will save a text file for permanent records including, the data used, analytical selections (e.g., FPC, sample unit based), sample size requirements, population statistics and strata statistics.

Sample Allocation can be accomplished using proportional, optimal or user-defined settings. The Optimal or Neyman allocation method uses the proportion of sample units and estimated variance with each stratum to allocate samples. Strata with more sample units and variance (i.e. heterogeneous) will receive more samples. This option is not

available unless existing sample data is being analyzed, because it requires an estimate of stratum variance.

Acknowledgements

Eric Finnen completed the first version of the Sampling Design Tool and laid the foundation for subsequent versions. Charles Menza and Chris Caldow has provided guidance and support at every step of tool production.

Contact [Ken Buja](#) for questions on technical matters related to the tool.

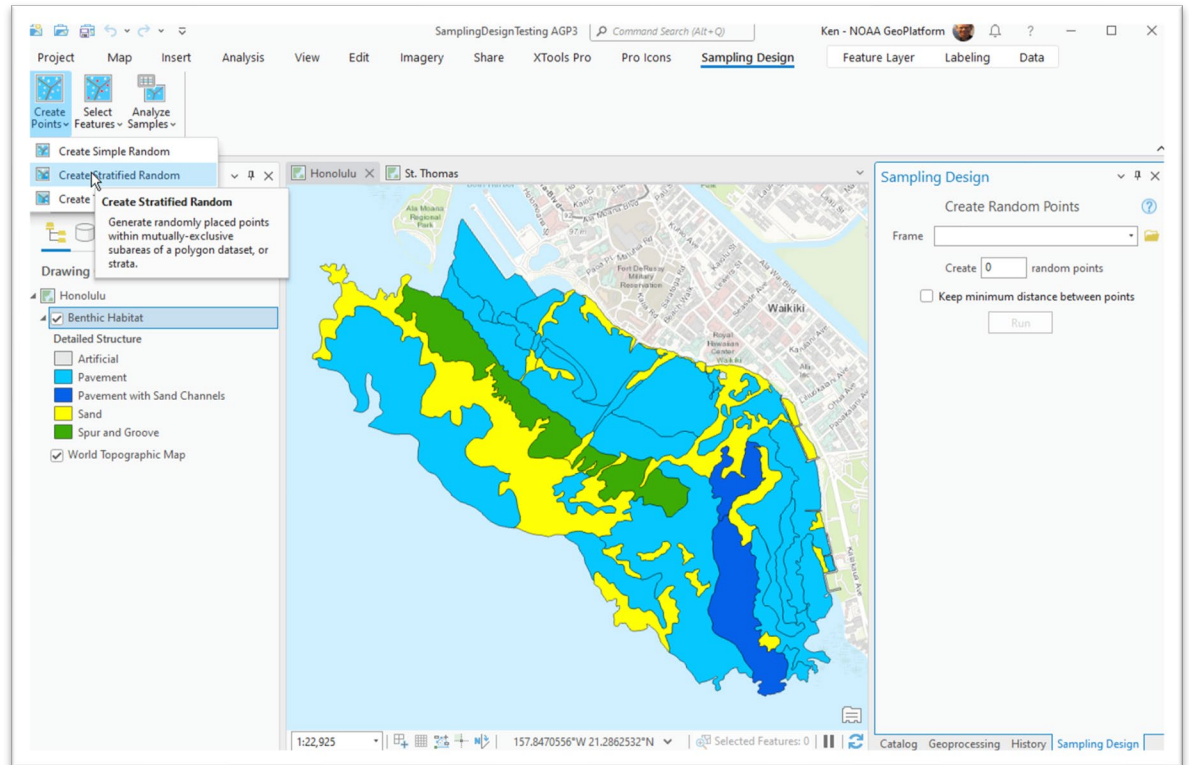


NOAA
National Ocean Service
National Centers for Coastal Ocean Science
[Marine Spatial Ecology](#)
1305 East West Highway
Silver Spring, MD 20910
Phone: (240) 533-0300

Examples

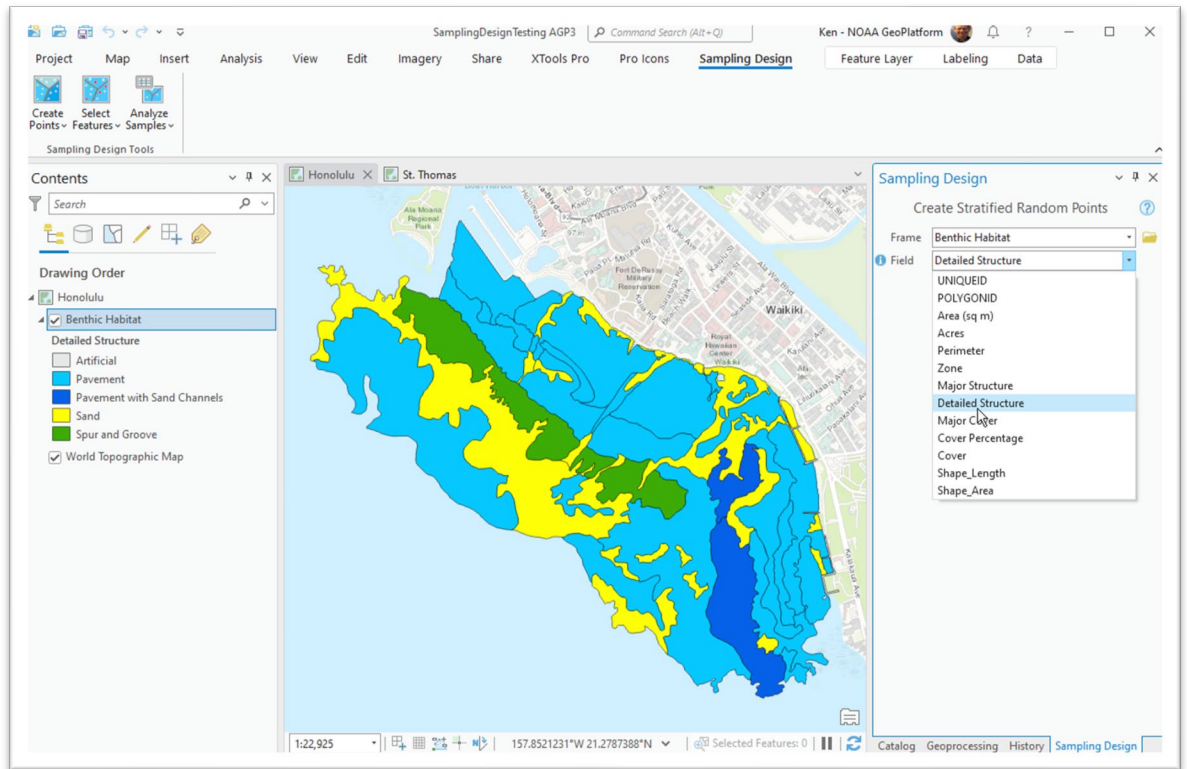
Create Stratified Random points

1. Click the Sampling Design tab, then click the **Create Points** button and select the **Create Stratified Random** tool.



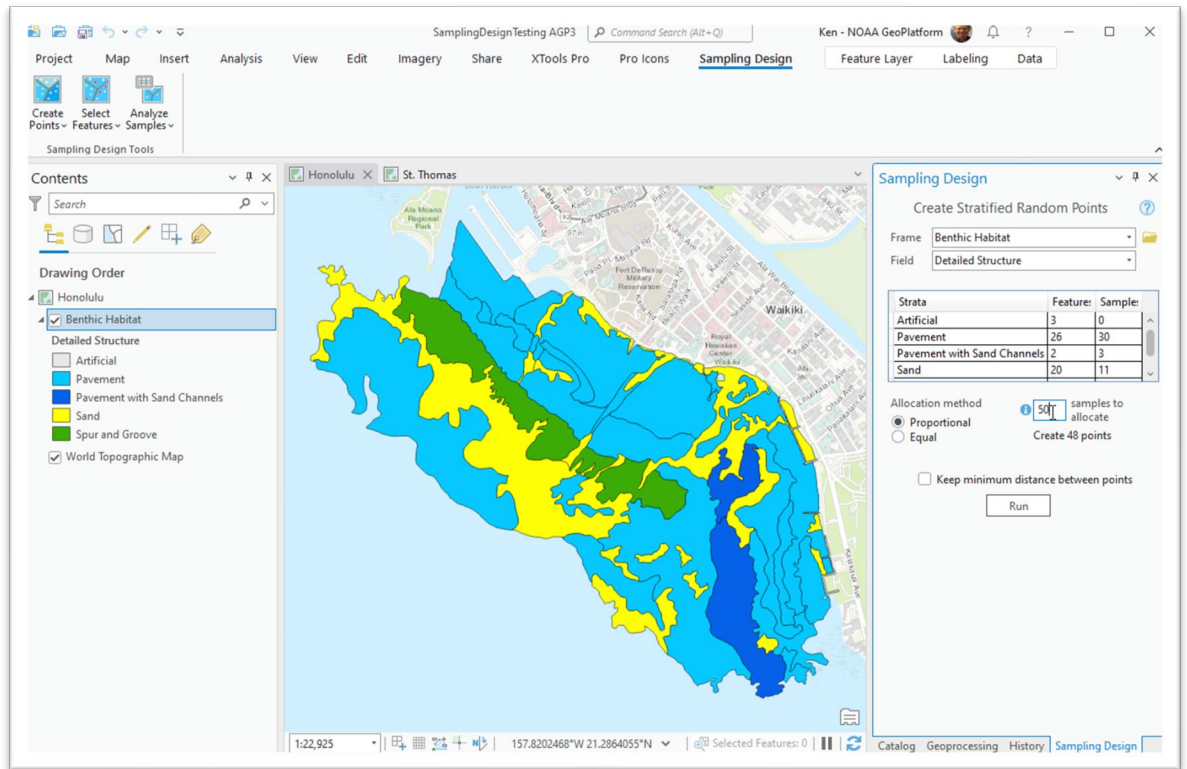
2. Select the Frame feature class polygon with the dropdown menu or the Browse button, then select the stratification field. Note that an information

icon will appear when hovering the mouse over an item.



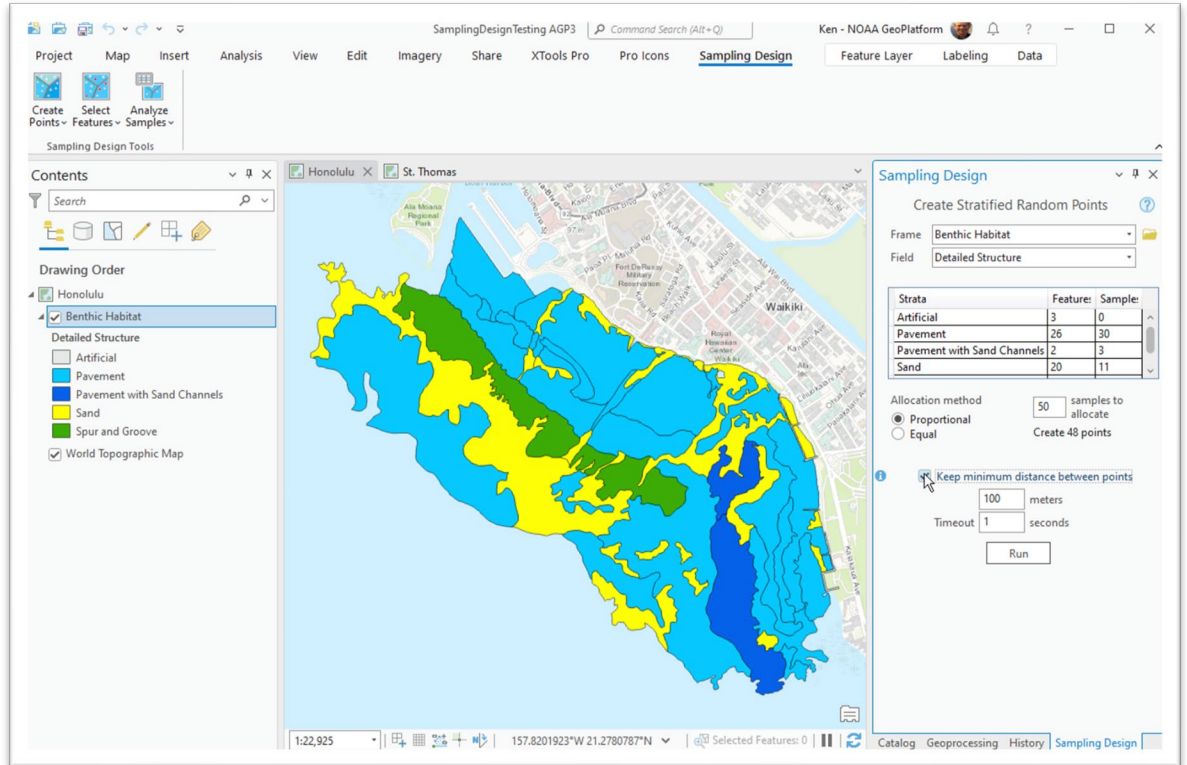
3. Click **Analyze** to list all strata.
4. Put the number of samples to allocate in the text box and the samples will be automatically allocated to each stratum according to the method selected. You can also change the samples allocated to a particular stratum by clicking

on the cell typing in a number.



5. Click the checkbox if you want to enable the Random Seed option.

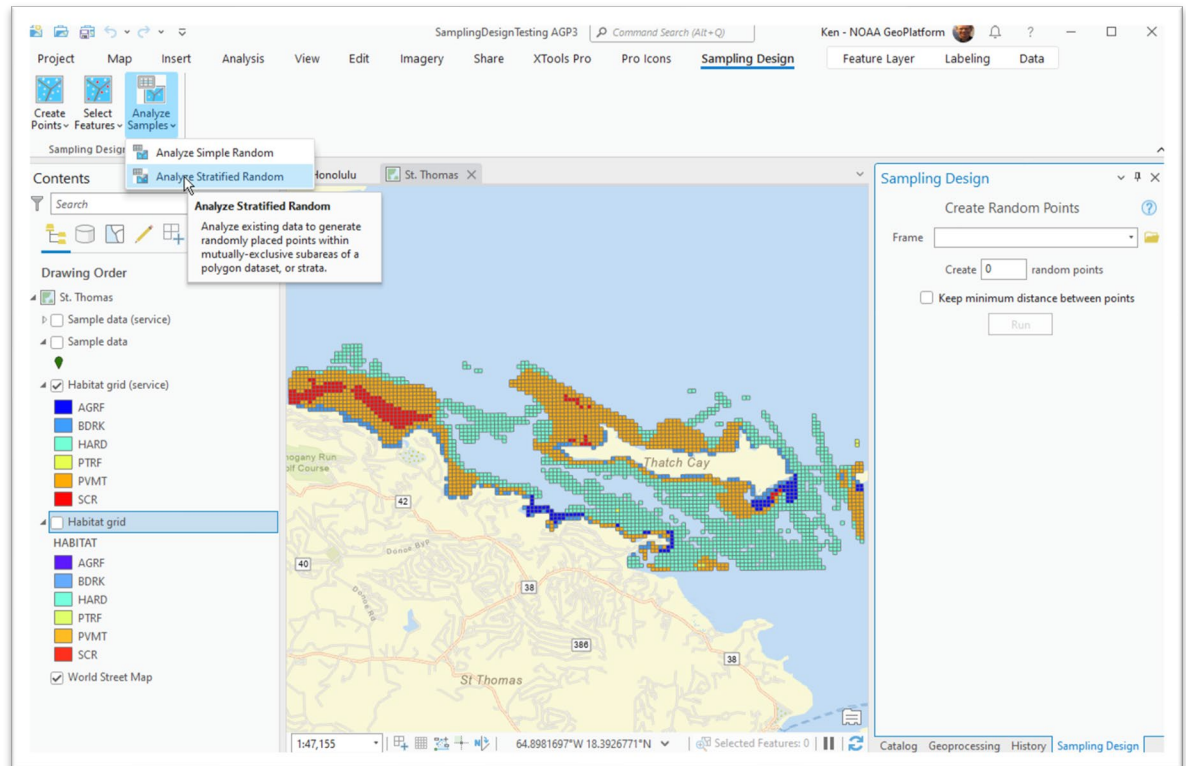
6. If you want to keep a minimum distance between, click the checkbox and specify the distance.



7. Click **Run** and use the Save dialog to create the dataset.

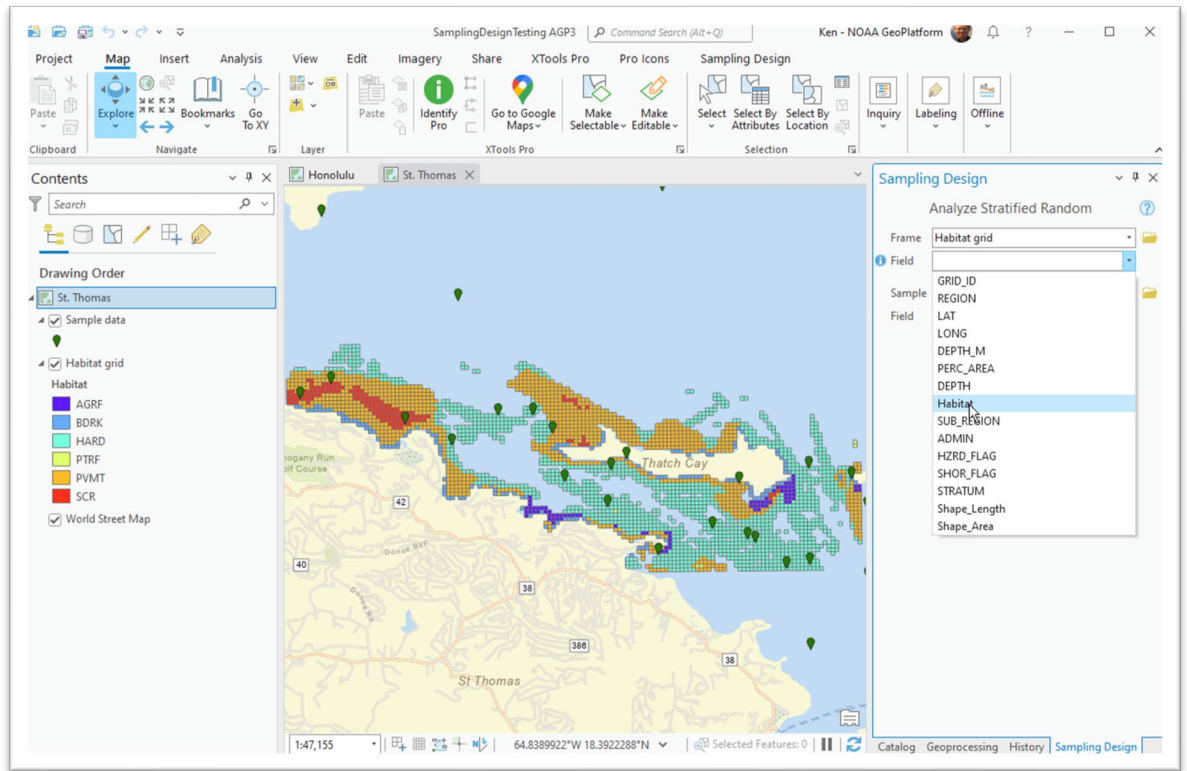
Refining Sample Selection using previously collected data

1. To refine sample selection, previous data (“Sample data”) and a sample frame (“Habitat grid”) are required.



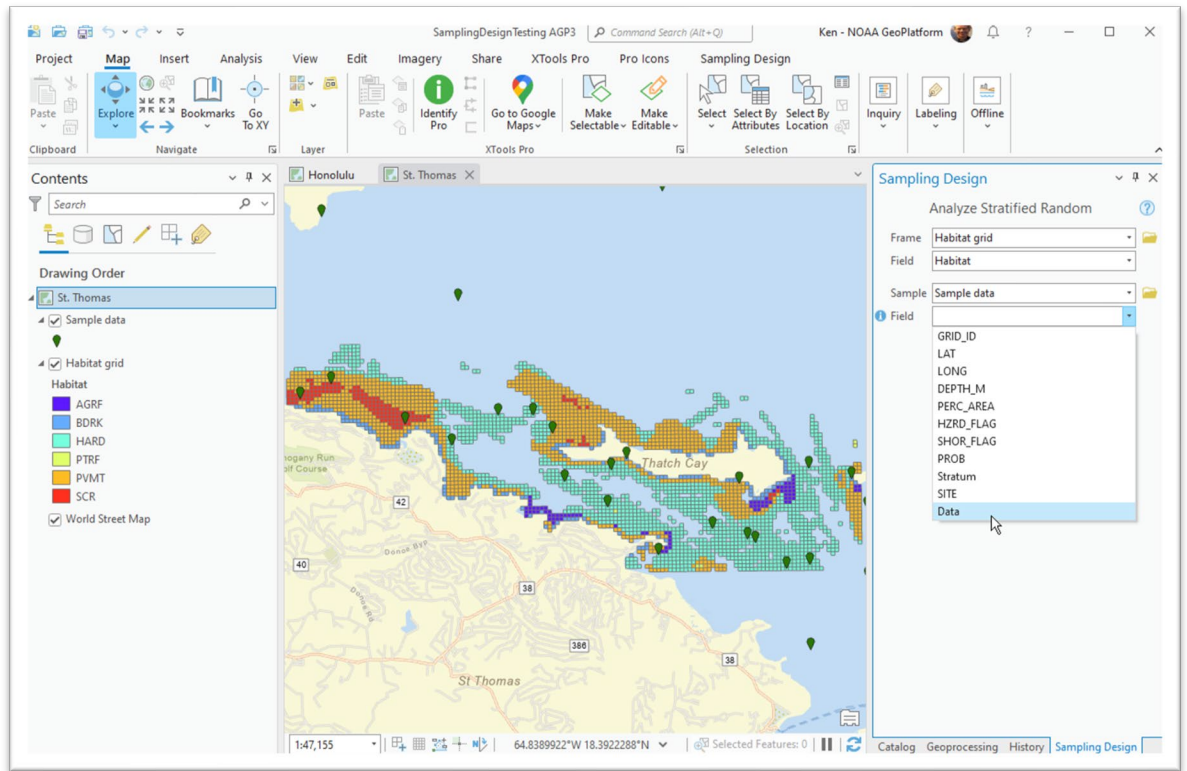
2. Click the Sampling Design tab, then click the **Analyze Samples** button and select the **Analyze Stratified Random** tool.
3. Select the frame data layer. Use the dropdown menu or folder browser to choose a point or polygon dataset. Select the field used to define strata in the frame using the dropdown menu, populated from the frame’s attribute table. Note that an information icon will appear when hovering the mouse over an

item.

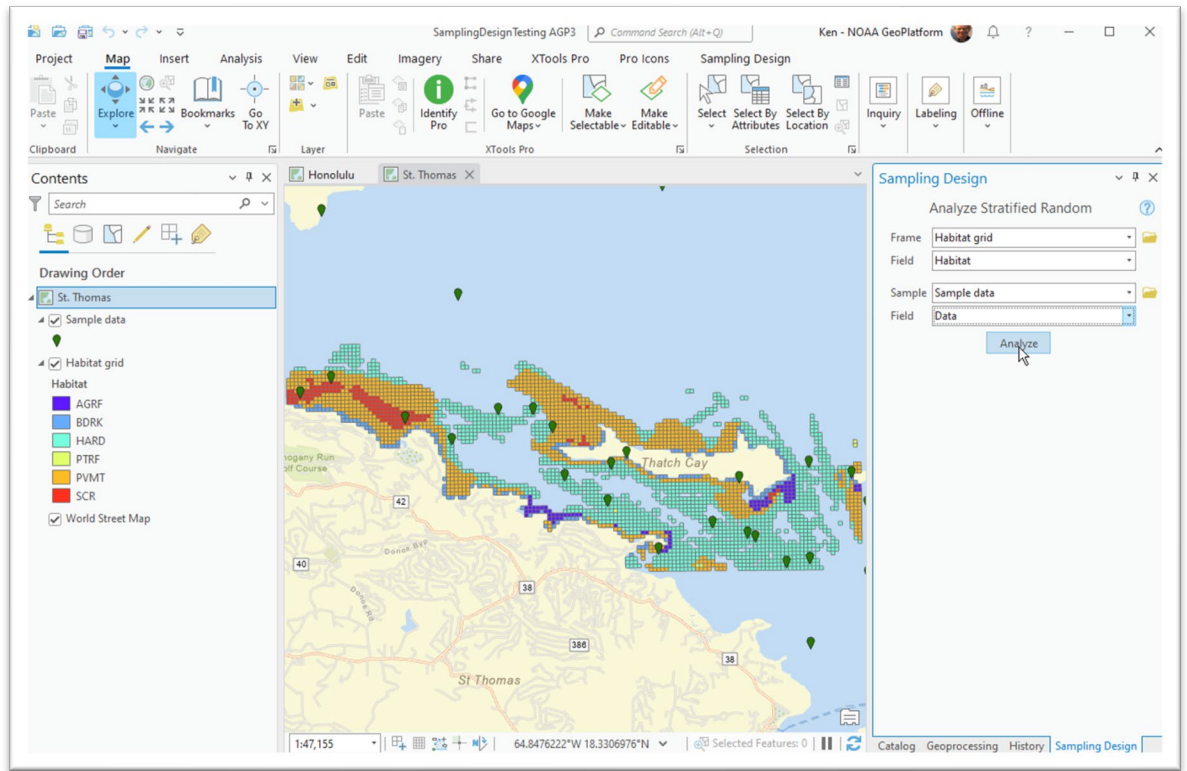


4. Select a sample layer with previously collected data and which will be examined. Select a metric field on which to perform analyses using the

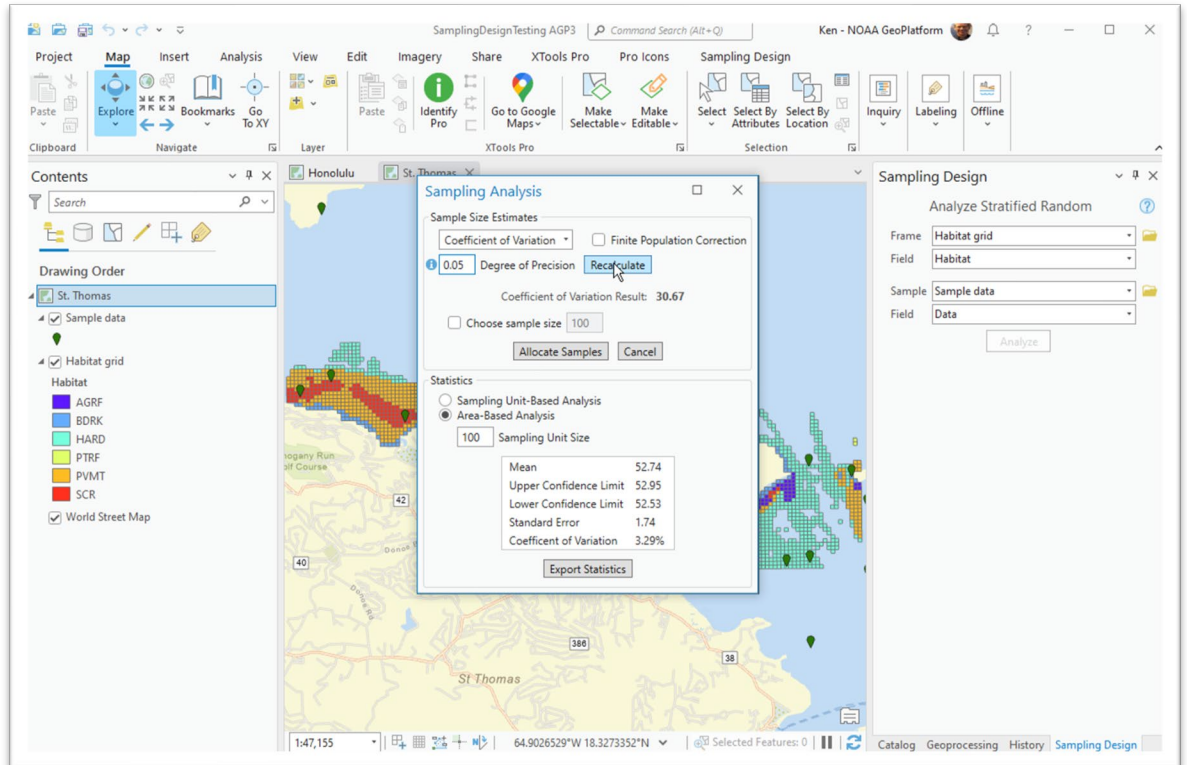
dropdown menu, populated from the sample layer's attribute table.



5. Click **Analyze**.

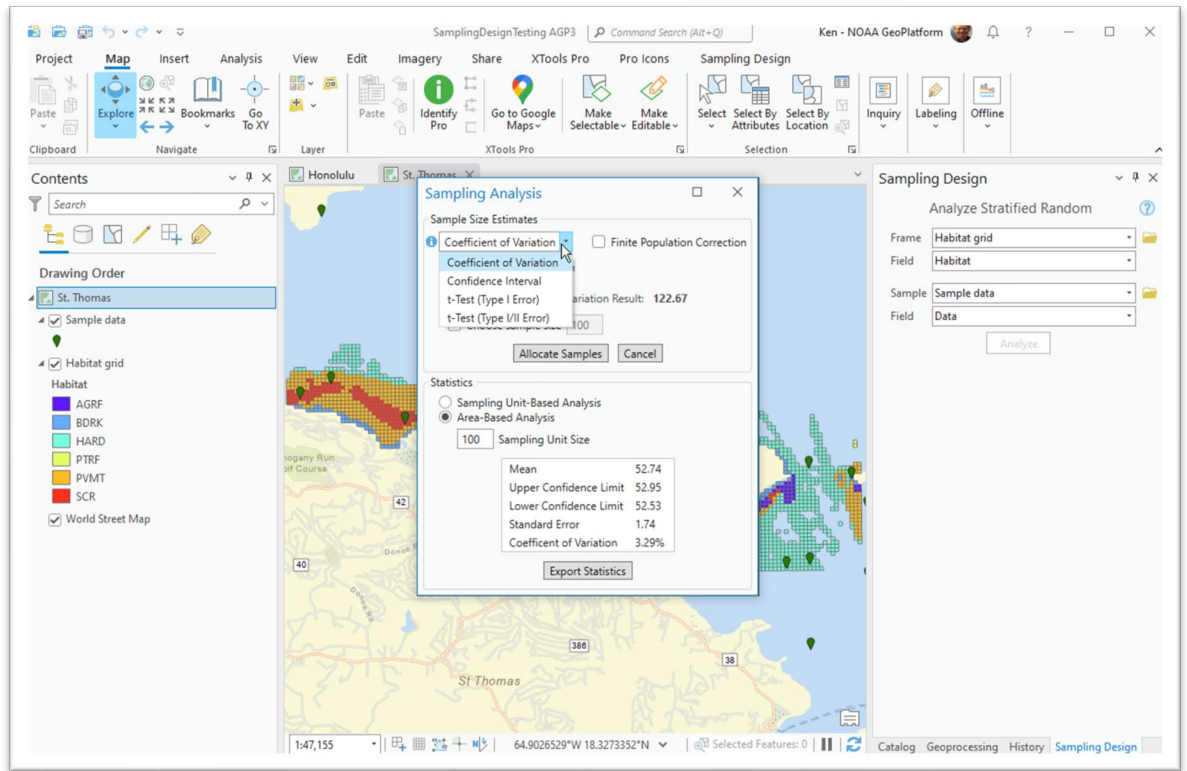


6. If needed, a message box will appear to describe any computational problems. Either fix these problems or continue.
7. Choose the survey objective, the degree of precision and whether or not the finite population correction will be used. In this example, we want to know what sample size is required to obtain a coefficient of variation with a 5% degree of precision. Click **Recalculate**.



8. Sample size estimates to achieve the desired survey objective are shown either as a single value (for Coefficient of variation) or in a table with numerous sample size requirements provided to help the user identify

variability.



9. Unfortunately, the project budget dictates that only 120 sites can be selected. The allocation can be amended to by checking the **Choose sample size** feature and entering a sample size.

10. Click **Allocate Samples**.

Sampling Analysis

Sample Size Estimates

t-Test (Type I/II Error) Finite Population Correction

0.05 Degree of Precision

	Confidence Level				
Power	99%	97.5%	95%	90%	75%
95%	4368.98	3693.34	3179.51	2639.39	1923.41
90%	3655.36	3039.77	2575.40	2091.80	1460.61
80%	2869.51	2327.32	1923.41	1508.89	981.33
70%	2357.65	1868.85	1508.89	1144.62	692.43
60%	1964.84	1521.09	1198.23	876.35	487.75
50%	1633.03	1230.98	942.47	659.85	330.12

Highlight maximum sample size 100

Choose sample size 120

Allocate Samples Cancel

Statistics

Sampling Unit-Based Analysis

Area-Based Analysis

100 Sampling Unit Size

Mean	52.74
Upper Confidence Limit	52.95
Lower Confidence Limit	52.53
Standard Error	1.74
Coefficient of Variation	3.29%

Export Statistics

11. Choose to allocate samples using the Optimal or Neyman allocation method.

Sampling Design

Analyze Stratified Random

Frame: Habitat grid

Field: Habitat

Sample: Sample data

Field: Data

Analyze

Strata	Feature	Sample
AGRF	84	3
BDRK	156	6
HARD	1455	59
PTRF	5	0

Allocation method

Proportional

Optimal (Neyman)

120 samples to allocate

Create 118 points

Keep minimum distance between points

Run

12. (Optional) Add 3 extra sample units to each stratum. This will add a small buffer in case a sampling unit allocated in Step 11 cannot be reached. Add extra samples by typing into the table.

The screenshot shows the Sampling Design software interface. The main map displays a habitat grid over a geographic area, with various strata (HARD, PTRF, PVMT, SCR) color-coded. The Sampling Design panel on the right is titled "Analyze Stratified Random" and includes the following settings:

- Frame: Habitat grid
- Field: Habitat
- Sample: Sample data
- Field: Data

An "Analyze" button is visible below these settings. Below the Analyze button is a table showing the current sampling design:

Strata	Features	Samples
HARD	1455	62
PTRF	5	3
PVMT	1085	47
SCR	152	8

Below the table, the Allocation method is set to "Optimal (Neyman)" with a "120 samples to allocate" input field. A "Run" button is highlighted with a mouse cursor. The "Keep minimum distance between points" checkbox is unchecked.

13. Click **Run** and use the Save dialog to create the dataset.