Florida 2023 Large Lake Sample Design Documentation

01/12/2023

Contacts:

Data Analyst

Vacant Position
Environmental Consultant
Florida Department of Environmental Protection (FDEP)
Division of Environmental Assessment and Restoration (DEAR)
Watershed Monitoring Section (WMS)

Section Administrator

Jay Silvanima
Environmental Administrator
Florida Department of Environmental Protection (FDEP)
Division of Environmental Assessment and Restoration (DEAR)
Watershed Monitoring Section (WMS)
2600 Blair Stone Road
Mail Station 3560
Tallahassee, Florida 32399-2400
James.Silvanima@FloridaDEP.gov (mailto:James.Silvanima@FloridaDEP.gov)
phone: 850-245-8507
fax: 850-245-8554

Analysis and Reporting Coordinator

Stephanie Sunderman-Barnes
Environmental Consultant
Florida Department of Environmental Protection (FDEP)
Division of Environmental Assessment and Restoration (DEAR)
Watershed Monitoring Section (WMS)
2600 Blair Stone Road
Mail Station 3560
Tallahassee, Florida 32399-2400
Stephanie.Sundermanbarnes@FloridaDEP.gov (mailto:Stephanie.Sundermanbarnes@FloridaDEP.gov)
850-245-8517

Description of Sample Design for Panel Design

Target population: All lakes in Florida that are greater than or equal to 10.0 hectares. Defined by FDEP based on the 2007 24K NHD water body GIS coverage. This identifies 1,699 large lakes.

Sample Frame:

The frame was provided by Florida Department of Environmental Protection, named *'Cycle17_LargeLakes_coverage_2023.shp'*. The GIS coverage is a large lake polygon coverage provided by FDEP based on the 2007 24K NHD water body data. GIS Data:

Attributes include: GNIS_NAME, REACHCODE, HECTARES, REPUNIT, RESOURCE, Comments, LakeCode, Shape_Leng, Shape_Area, Geometry

The state of Florida is divided into six regions (reporting units) for sampling design. The reporting unit name attribute in Cycle17_LargeLakes_2023_coverage.shp is REPUNIT and the six values are: ZONE 1, ZONE 2, ZONE 3, ZONE 4, ZONE 5, ZONE 6.

Survey Design for Cycle 17 Large Lake Site Selections:

Purpose of this R script is to select large lake sites throughout the 6 reporting units as part of an annual water quality sample survey of Florida large lakes.

Code developed using R version 4.1.2 and spsurvey version 5.4.0.

Load the libraries that will be called in this script.

```
library (spsurvey)
library(dplyr)
```

Check working directory and change location if needed.

```
getwd()
```

```
## [1] "C:/R/FL Large Lake Selections"
```

Create two simple features objects from shapefiles.
1) Polygon features representing the 2023 target population of large lakes (Cycle17_LargeLakes_coverage_2023).
2) Polygon features representing the Zones (Watershed_Monitoring_Section_(WMS)_Cycle_3_Reporting_Units).
Change all projections to Florida Albers HARN(CRS code 3087).

Calculate lake area per reporting Zone.

```
dsgn_ll <- st_read(dsn=".",layer="Cycle17_LargeLakes_coverage_2023")
```

```
## Reading layer `Cycle17_LargeLakes_coverage_2023' from data source
##   `C:\R\FL Large Lake Selections' using driver `ESRI Shapefile'
## Simple feature collection with 1699 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 85113.99 ymin: 269537.8 xmax: 790842.1 ymax: 778314.2
## Projected CRS: FDEP_Albers_HARN
```

```
wms_c3_reporting_units <- st_read(dsn=".",layer="Watershed_Monitoring_Section_(WMS)_Cycle_3_Repo
rting_Units")
```

```
## Reading layer `Watershed_Monitoring_Section_(WMS)_Cycle_3_Reporting_Units' from data source `
C:\R\FL Large Lake Selections' using driver `ESRI Shapefile'
## Simple feature collection with 6 features and 5 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.63542 ymin: 24.54522 xmax: -80.03095 ymax: 31.00084
## Geodetic CRS:  WGS 84
```

```
wms_c3_reporting_units <- st_transform(wms_c3_reporting_units, crs = 3087)
wms_c3_reporting_units
```

```
## Simple feature collection with 6 features and 5 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 52604.84 ymin: 61813.16 xmax: 793961.6 ymax: 781606.9
## Projected CRS: NAD83(HARN) / Florida GDL Albers
##   OBJECTID REPORTING_  ZONE_NAME   SHAPEAREA  SHAPELEN
## 1        1     ZONE 1     NWFWMD 28928958259 2240796.2
## 2        2     ZONE 2      SRWMD 19579475536  806595.1
## 3        3     ZONE 6 SFWMD EAST 20817253887 2706651.1
## 4        4     ZONE 4     SWFWMD 25325641885 1942336.2
## 5        5     ZONE 3     SJRWMD 30378451186 2491916.5
## 6        6     ZONE 5 SFWMD WEST 20914035527 1742555.9
##                          geometry
## 1 MULTIPOLYGON (((295836.8 63...
## 2 MULTIPOLYGON (((421510.4 73...
## 3 MULTIPOLYGON (((761186.7 39...
## 4 MULTIPOLYGON (((545245.4 60...
## 5 MULTIPOLYGON (((548234.1 73...
## 6 MULTIPOLYGON (((654296.6 50...
```

For large lakes target population: Convert all column names to lowercase. Inspect data.

```
names(dsgn_ll)<-tolower(names(dsgn_ll))
names(dsgn_ll)
```

```
##  [1] "gnis_name"  "reachcode"  "hectares"   "repunit"    "resource"
##  [6] "comments"   "lakecode"   "shape_leng" "shape_area" "geometry"
```

```
head(dsgn_ll)
```

```
## Simple feature collection with 6 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 230971 ymin: 456358.9 xmax: 590387.8 ymax: 725567.1
## Projected CRS: FDEP_Albers_HARN
##         gnis_name      reachcode hectares repunit   resource
## 1            <NA> 03100206000966 11.48807  Zone 4 Large Lake
## 2       Echo Lake 03100206001022 10.63352  Zone 4 Large Lake
## 3            <NA> 03100206018137 12.84469  Zone 4 Large Lake
## 4 Starvation Lake 03100206000954 18.55653  Zone 4 Large Lake
## 5            <NA> 03100208003947 16.63570  Zone 3 Large Lake
## 6     Cravy Lakes 03140203002076 11.34753  Zone 1 Large Lake
##                                                                          comments
## 1                                                                            <NA>
## 2                                                                            <NA>
## 3                                                                            <NA>
## 4                                                                            <NA>
## 5 Converted small lake to large lakes 2017; move to large lakes in 2017 NAW 08/9/2016
## 6                                                                            <NA>
##              lakecode shape_leng shape_area                       geometry
## 1  613-03100206000966   2050.146   114880.7 MULTIPOLYGON (((546091.8 45...
## 2  614-03100206001022   1391.761   106335.2 MULTIPOLYGON (((537024.1 45...
## 3  617-03100206018137   1825.429   128446.9 MULTIPOLYGON (((534077.1 45...
## 4  625-03100206000954   2506.460   185565.3 MULTIPOLYGON (((546448.4 45...
## 5 1009-03100208003947   2643.236   166357.0 MULTIPOLYGON (((590385.9 55...
## 6 2125-03140203002076   1742.685   113475.3 MULTIPOLYGON (((231291.4 72...
```

```
tail(dsgn_ll)
```

```
## Simple feature collection with 6 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 626068 ymin: 414157.1 xmax: 721956.2 ymax: 536314.9
## Projected CRS: FDEP_Albers_HARN
##                 gnis_name      reachcode   hectares repunit   resource
## 1694                 <NA> 03080101009581   11.20807  Zone 3 Large Lake
## 1695           Johns Lake 03080102005491  989.77148  Zone 3 Large Lake
## 1696    Blue Cypress Lake 03080101011546 2697.70312  Zone 3 Large Lake
## 1697 Little Sawgrass Lake 03080101045661   36.69663  Zone 3 Large Lake
## 1698   Lake Hellen Blazes 03080101011247  103.05471  Zone 3 Large Lake
## 1699        Sawgrass Lake 03080101028858  199.75477  Zone 3 Large Lake
##                                           comments            lakecode
## 1694 Moved from small lake group due to size > 10 ha 4644-03080101009581
## 1695                                           <NA> 4645-03080102005491
## 1696                                           <NA> 4647-03080101011546
## 1697                                           <NA> 4652-03080101045661
## 1698                                           <NA> 4653-03080101011247
## 1699                                           <NA> 4654-03080101028858
##      shape_leng shape_area                       geometry
## 1694   2547.443   112080.7 MULTIPOLYGON (((679161.6 53...
## 1695  50226.142  9897714.8 MULTIPOLYGON (((626774.3 50...
## 1696  21075.838 26977031.2 MULTIPOLYGON (((718637.8 41...
## 1697   3278.783   366966.3 MULTIPOLYGON (((715095.3 45...
## 1698   5281.432  1030547.1 MULTIPOLYGON (((714166.8 44...
## 1699   8205.094  1997547.7 MULTIPOLYGON (((717269.3 45...
```

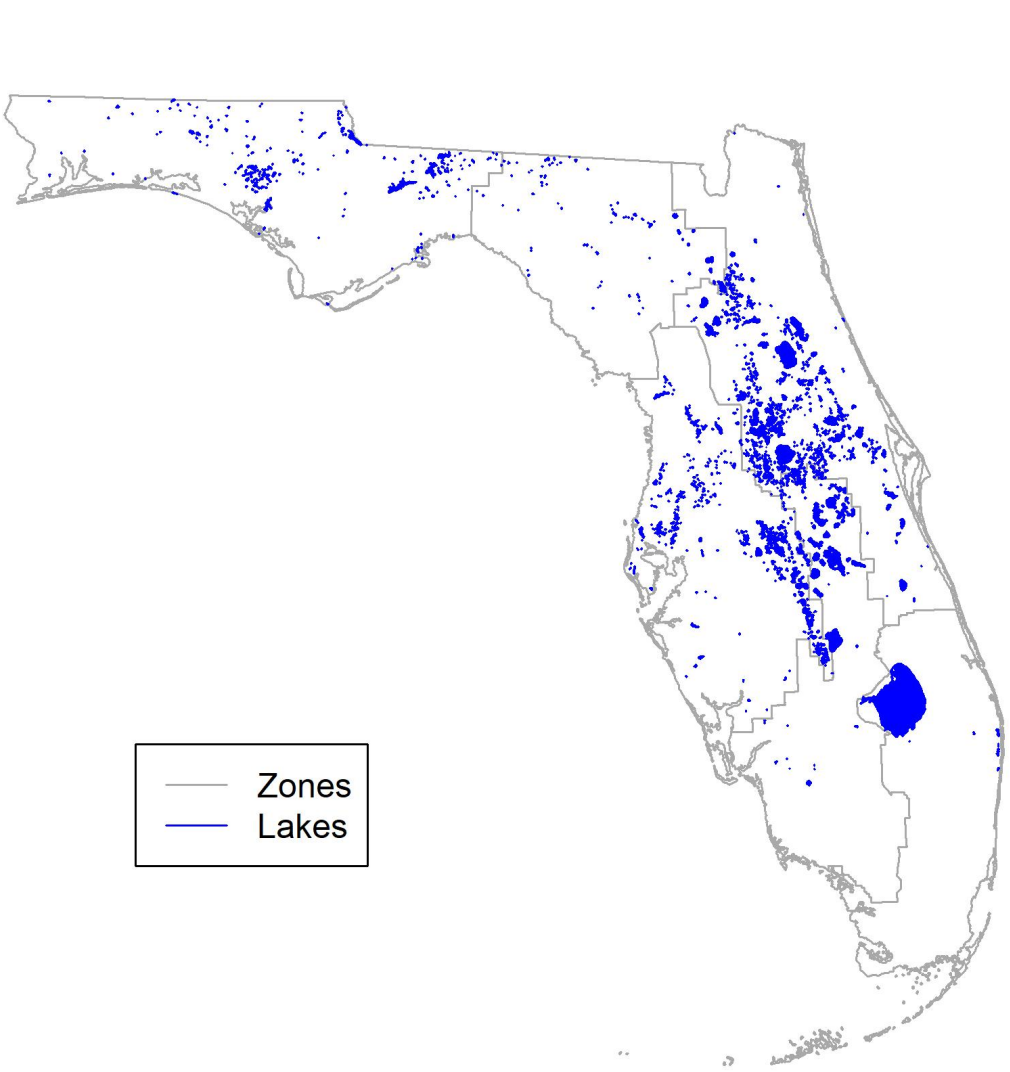
Calculate Lake polygon areas per c3_zone for all zones.

```
lakearea<-tapply(dsgn_ll$hectares,list(dsgn_ll$repunit), sum)
lakearea[is.na(lakearea)] <- 0
round(addmargins(lakearea),1)
```

```
##   Zone 1   Zone 2   Zone 3   Zone 4   Zone 5   Zone 6      Sum
##  18077.9   7581.9 121935.8  43095.0  60545.3 129999.5 381235.4
```

Plot the Zone polygons and 2023 target population of large lakes polygons.

```
jpeg('2023_LL_Population.jpg', units = 'in', width = 7, height = 7, res = 300)
plot(st_geometry(wms_c3_reporting_units), border='darkgray', main= '2023 Target Population of La
rge Lakes')
plot(st_geometry(dsgn_ll), border = 'blue', col = 'blue', add = TRUE)
legend(150000, 300000, legend=c('Zones','Lakes'), col=c('darkgray','blue'),lty=c(1,1))
dev.off()
```



Create a factor column to use as a stratum variable and check the components in the stratum column.

```
dsgn_ll$stratum<-factor(as.character(dsgn_ll$repunit))
levels(dsgn_ll$stratum)
```

```
## [1] "Zone 1" "Zone 2" "Zone 3" "Zone 4" "Zone 5" "Zone 6"
```

```
head(dsgn_ll)
```

```
## Simple feature collection with 6 features and 10 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 230971 ymin: 456358.9 xmax: 590387.8 ymax: 725567.1
## Projected CRS: FDEP_Albers_HARN
##         gnis_name      reachcode hectares repunit   resource
## 1            <NA> 03100206000966 11.48807  Zone 4 Large Lake
## 2       Echo Lake 03100206001022 10.63352  Zone 4 Large Lake
## 3            <NA> 03100206018137 12.84469  Zone 4 Large Lake
## 4 Starvation Lake 03100206000954 18.55653  Zone 4 Large Lake
## 5            <NA> 03100208003947 16.63570  Zone 3 Large Lake
## 6     Cravy Lakes 03140203002076 11.34753  Zone 1 Large Lake
##                                                                        comments
## 1                                                                          <NA>
## 2                                                                          <NA>
## 3                                                                          <NA>
## 4                                                                          <NA>
## 5 Converted small lake to large lakes 2017; move to large lakes in 2017 NAW 08/9/2016
## 6                                                                          <NA>
##              lakecode shape_leng shape_area                       geometry
## 1  613-03100206000966   2050.146   114880.7 MULTIPOLYGON (((546091.8 45...
## 2  614-03100206001022   1391.761   106335.2 MULTIPOLYGON (((537024.1 45...
## 3  617-03100206018137   1825.429   128446.9 MULTIPOLYGON (((534077.1 45...
## 4  625-03100206000954   2506.460   185565.3 MULTIPOLYGON (((546448.4 45...
## 5 1009-03100208003947   2643.236   166357.0 MULTIPOLYGON (((590385.9 55...
## 6 2125-03140203002076   1742.685   113475.3 MULTIPOLYGON (((231291.4 72...
##   stratum
## 1  Zone 4
## 2  Zone 4
## 3  Zone 4
## 4  Zone 4
## 5  Zone 3
## 6  Zone 1
```

Create lake area categories for each stratum (Zone).

First create column area_cat

```
dsgn_ll$area_cat <- rep(NA,nrow(dsgn_ll))
dsgn_ll$area_cat <- dsgn_ll$hectares
abbr <- c("Zone 1","Zone 2", "Zone 3", "Zone 4", "Zone 5", "Zone 6")
names(abbr) <- levels(dsgn_ll$stratum)
```

Unequal probability of selection within each strata based on lake area. Within each stratum lakes are ordered from smallest to largest area. Then the cumulative sum of lake area is computed, split into five equal area parts, and the lake identified at the boundary of the parts. The area of those five lakes is used to identify the lake area categories for the unequal probability sampling. For example using large lakes, the Suwannee region has 89 large lakes with a total lake area of 7632.871 ha. The five cumulative area breaks are 48 ha, 222 ha, 467 ha, 805 ha, and 1630+ ha. The area categories for the Suwannee are [10,49], (49,215], (215,467], (467,805], and (805,1.63e+03]. The number of lakes in each area category is 68, 14, 4, 2, and 1. The total lake area in each area category is 1484.778 ha, 1545.794 ha, 1406.269 ha, 1534.205 ha, and 1633.170 ha, respectively.

The below loop creates five lake categories based on size per reporting unit and populates the area_cat column with this information.

It also exports out a .csv file of the 30 area categories (five per zone).

```
for(i in levels(dsgn_ll$stratum) ) {
  tst <- dsgn_ll$stratum == i
  itmp <- order(dsgn_ll$hectares[tst])
  tmp <- cumsum(sort(dsgn_ll$hectares[tst]))
  n <- length(itmp)
  if(i != "Zone 6")   # Special case for Lake Okeechobee
  {ctmp <- cut(tmp,
               breaks=c(seq(10, tmp[n], length=6)),
               include.lowest = TRUE)
  }
  else
  {ctmp <- cut(tmp,
               breaks=c(seq(10, tmp[n-1], length=5), tmp[n]),
               include.lowest = TRUE)
  }
  icut <- cumsum(table(ctmp))
  acut <- unique(ceiling(dsgn_ll$hectares[tst][itmp][icut]))
  dsgn_ll$area_cat[tst] <-
    paste(abbr[i],(cut(dsgn_ll$hectares[tst], breaks=c(10,acut),
                       include.lowest=TRUE)), sep="_")
}
dsgn_ll$area_cat <- factor(dsgn_ll$area_cat)


area_cat <-levels(dsgn_ll$area_cat)
area_cat <-as.data.frame(area_cat)

write.csv(area_cat, "area_cat.csv")
```

Create dataframe of lake category, lake numbers, and size.

rename columns to ‘Zones’, ‘Number of Lakes, and ’hectares’ and remove row names.

Write out dataframe frame to .csv file.

```
Lakeareas <- data.frame(addmargins(table(dsgn_ll$repunit)),round(addmargins(lakearea),1))
names(Lakeareas) <-c("Zones", "Number_of_Lakes", "hectares")
row.names(Lakeareas)<- NULL
Lakeareas
```

```
##    Zones Number_of_Lakes hectares
## 1 Zone 1             239  18077.9
## 2 Zone 2              92   7581.9
## 3 Zone 3             733 121935.8
## 4 Zone 4             501  43095.0
## 5 Zone 5             125  60545.3
## 6 Zone 6               9 129999.5
## 7    Sum            1699 381235.4
```

```
write.csv(Lakeareas, "2023 Large Lake Framesize.csv", row.names = FALSE)
```

Create lake size category data frame, rename columns to ‘area_cat’ and ‘Number_of_Lakes’ and remove row names.
Write out data frame to .csv file.

```
area_cat_summary <- data.frame(table(dsgn_ll$area_cat))
names(area_cat_summary) <- c('area_cat','Number_of_Lakes')
area_cat_summary
```

```
##                      area_cat Number_of_Lakes
## 1  Zone 1_(1.02e+03,1.43e+03]              2
## 2  Zone 1_(1.43e+03,3.27e+03]              2
## 3       Zone 1_(144,1.02e+03]             11
## 4             Zone 1_(45,144]             48
## 5              Zone 1_[10,45]            176
## 6            Zone 2_(215,467]              5
## 7            Zone 2_(467,805]              2
## 8             Zone 2_(49,215]             14
## 9       Zone 2_(805,1.59e+03]              1
## 10             Zone 2_[10,49]             70
## 11 Zone 3_(1.78e+03,3.26e+03]              9
## 12      Zone 3_(182,1.78e+03]             38
## 13 Zone 3_(3.26e+03,7.44e+03]              4
## 14 Zone 3_(7.44e+03,1.76e+04]              2
## 15            Zone 3_[10,182]            680
## 16 Zone 4_(1.32e+03,1.86e+03]              6
## 17           Zone 4_(175,429]             32
## 18      Zone 4_(429,1.32e+03]             11
## 19             Zone 4_(55,175]             84
## 20              Zone 4_[10,55]            368
## 21 Zone 5_(1.89e+03,3.97e+03]              3
## 22 Zone 5_(3.97e+03,9.66e+03]              2
## 23      Zone 5_(730,1.89e+03]              8
## 24 Zone 5_(9.66e+03,1.25e+04]              1
## 25            Zone 5_[10,730]            111
## 26           Zone 6_(150,163]              1
## 27           Zone 6_(163,228]              1
## 28      Zone 6_(228,1.29e+05]              1
## 29             Zone 6_(59,150]              1
## 30              Zone 6_[10,59]              5
```

```
write.csv(area_cat_summary, "2023 Large Lake Size Categories.csv", row.names = FALSE)
```

A Generalized Random Tessellation Stratified (GRTS) survey design for an areal lake resource was used. The GRTS design includes reverse hierarchical ordering of the selected sites.

Create the stratification to be used.

Create the list of lake size categories to by used.

```
LL_base <-c("Zone 1"=15,"Zone 2"=15,"Zone 3"=15, "Zone 4"=15, "Zone 5"=15, "Zone 6"=15)
LL_over <-c("Zone 1"=135,"Zone 2"=135,"Zone 3"=135, "Zone 4"=135, "Zone 5"=135, "Zone 6"=135)
LL_select <-c("Zone 1"="unequal","Zone 2"="unequal","Zone 3"="unequal", "Zone 4"="unequal", "Zon
e 5"="unequal", "Zone 6"="unequal")

LL_catyn <-list("Zone 1" = c('Zone 1_(1.02e+03,1.43e+03]' = 3,
                          'Zone 1_(1.43e+03,3.27e+03]' = 3,
                          'Zone 1_(144,1.02e+03]' = 3,
                          'Zone 1_(45,144]' = 3,
                          'Zone 1_[10,45]' = 3),
                        "Zone 2" = c('Zone 2_(215,467]' = 3,
                          'Zone 2_(467,805]' = 3,
                          'Zone 2_(49,215]' = 3,
                          'Zone 2_(805,1.59e+03]' = 3,
                          'Zone 2_[10,49]' = 3),
                        "Zone 3" = c('Zone 3_(1.78e+03,3.26e+03]' = 3,
                          'Zone 3_(182,1.78e+03]' = 3,
                          'Zone 3_(3.26e+03,7.44e+03]' = 3,
                          'Zone 3_(7.44e+03,1.76e+04]' = 3,
                          'Zone 3_[10,182]' = 3),
                        "Zone 4" = c('Zone 4_(1.32e+03,1.86e+03]' = 3,
                          'Zone 4_(175,429]' = 3,
                          'Zone 4_(429,1.32e+03]' = 3,
                          'Zone 4_(55,175]' = 3,
                          'Zone 4_[10,55]' = 3),
                        "Zone 5" = c('Zone 5_(1.89e+03,3.97e+03]' = 3,
                          'Zone 5_(3.97e+03,9.66e+03]' = 3,
                          'Zone 5_(730,1.89e+03]' = 3,
                          'Zone 5_(9.66e+03,1.25e+04]' = 3,
                          'Zone 5_[10,730]' = 3),
                        "Zone 6" = c('Zone 6_(150,163]' = 3,
                          'Zone 6_(163,228]' = 3,
                          'Zone 6_(228,1.29e+05]' = 3,
                          'Zone 6_(59,150]' = 3,
                          'Zone 6_[10,59]' = 3))
```

Run random sample once to get random seed and put result into set.seed.
Reason is so that exactly the same sites can be reproduced if rerun. Don't change set.seed unless you want a different set of sites.

```
sample(1000000,1)
```

```
set.seed(598771)
```

Create variable to keep track of how long spsurvey takes to run grts function.

```
dsgntime <- proc.time()  # keep track of how long spsurvey takes
```

Create the GRTS survey design

Stratification:
Stratify by zones created for statewide coverage.
Expected sample size: 15 sites within six of the state's zones.

Oversample: 9x sample sites for each zone.

Site Use: The base design has 15 sites for each of the six zones in the stratum. Sites are listed in SiteID order and must be used in that order. All sites that occur prior to the last site used must have been evaluated for use and then either sampled or the reason documented as to why that site was not used.

Print the initial six lines of the survey design.
Print dsgntime to view run time for grts function in minutes.

```
sites<- grts(dsgn_ll,
             stratum_var="stratum",
             n_base = LL_base,
             n_over = LL_over,
             seltype = LL_select,
             caty_var = "area_cat",
             caty_n = LL_catyn,
             wgt_units = "ha",
             pt_density =  1000,
             DesignID="FLLL23001")

head(sites)
```

```
## $sites_legacy
## NULL
##
## $sites_base
## Simple feature collection with 90 features and 20 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: 238618.4 ymin: 281401 xmax: 790773.7 ymax: 761365.5
## Projected CRS: FDEP_Albers_HARN
## First 10 features:
##           siteID siteuse replsite lon_WGS84 lat_WGS84 stratum            wgt
## 1  FLLL23001-001    Base     None -84.18274  30.44043  Zone 1 1261.4117 [ha]
## 2  FLLL23001-002    Base     None -84.55497  30.43627  Zone 1 1619.2784 [ha]
## 3  FLLL23001-003    Base     None -84.94773  30.75563  Zone 1 1619.2784 [ha]
## 4  FLLL23001-004    Base     None -85.65047  30.47060  Zone 1 1194.2013 [ha]
## 5  FLLL23001-005    Base     None -84.25058  30.52126  Zone 1 1261.4117 [ha]
## 6  FLLL23001-006    Base     None -84.53780  30.46283  Zone 1 1619.2784 [ha]
## 7  FLLL23001-007    Base     None -85.04250  30.85660  Zone 1 1194.2013 [ha]
## 8  FLLL23001-008    Base     None -85.62800  30.50113  Zone 1 1261.4117 [ha]
## 9  FLLL23001-009    Base     None -84.32996  30.52692  Zone 1  875.6244 [ha]
## 10 FLLL23001-010    Base     None -84.20881  30.64885  Zone 1 1075.4480 [ha]
##                     ip                       caty     gnis_name      reachcode
## 1  0.0007927626 [1/ha]            Zone 1_(45,144]  Piney Z Lake 03120001008774
## 2  0.0006175590 [1/ha] Zone 1_(1.43e+03,3.27e+03]  Lake Talquin 03120003001574
## 3  0.0006175590 [1/ha] Zone 1_(1.43e+03,3.27e+03] Lake Seminole 03130004004753
## 4  0.0008373798 [1/ha]             Zone 1_[10,45]     Sand Lake 03140101000846
## 5  0.0007927626 [1/ha]            Zone 1_(45,144]     Lake Hall 03120003001525
## 6  0.0006175590 [1/ha] Zone 1_(1.43e+03,3.27e+03]  Lake Talquin 03120003001574
## 7  0.0008373798 [1/ha]             Zone 1_[10,45]    Sweet Pond 03130004001651
## 8  0.0007927626 [1/ha]            Zone 1_(45,144]    Gully Lake 03140101001084
## 9  0.0011420422 [1/ha] Zone 1_(1.02e+03,1.43e+03]  Lake Jackson 03120003001734
## 10 0.0009298450 [1/ha]      Zone 1_(144,1.02e+03]  Lake Iamonia 03120003001444
##      hectares repunit   resource
## 1    85.87792  Zone 1 Large Lake
## 2  3271.88680  Zone 1 Large Lake
## 3  1591.36654  Zone 1 Large Lake
## 4    33.08317  Zone 1 Large Lake
## 5    75.86576  Zone 1 Large Lake
## 6  3271.88680  Zone 1 Large Lake
## 7    15.35576  Zone 1 Large Lake
## 8   142.72525  Zone 1 Large Lake
## 9  1425.85241  Zone 1 Large Lake
## 10 1015.47034  Zone 1 Large Lake
##                                          comments            lakecode
## 1                                            <NA> 4291-03120001008774
## 2                                            <NA> 4035-03120003001574
## 3                                            <NA> 4065-03130004004753
## 4                                            <NA> 3978-03140101000846
## 5                                            <NA> 4075-03120003001525
## 6                                            <NA> 4035-03120003001574
## 7                                            <NA> 4283-03130004001651
## 8                                            <NA> 3971-03140101001084
```

```
## 9  Too shallow. Unable to get boat close to location. 4068-03120003001734
## 10                 Updated by Andy Roach on 8-23-18. 4459-03120003001444
##    shape_leng shape_area                   area_cat sframe_stratum
## 1    3855.003   858779.2            Zone 1_(45,144]         Zone 1
## 2  165804.747 32718868.0 Zone 1_(1.43e+03,3.27e+03]         Zone 1
## 3   57993.784 15913665.4 Zone 1_(1.43e+03,3.27e+03]         Zone 1
## 4    2829.936   330831.7             Zone 1_[10,45]         Zone 1
## 5    4590.483   758657.6            Zone 1_(45,144]         Zone 1
## 6  165804.747 32718868.0 Zone 1_(1.43e+03,3.27e+03]         Zone 1
## 7    2131.286   153557.6             Zone 1_[10,45]         Zone 1
## 8   10671.117  1427252.5            Zone 1_(45,144]         Zone 1
## 9   53764.645 14258524.1 Zone 1_(1.02e+03,1.43e+03]         Zone 1
## 10  51202.663 10154703.4      Zone 1_(144,1.02e+03]         Zone 1
##                     geometry
## 1  POINT (382465.2 714779.2)
## 2    POINT (346746 714423.8)
## 3  POINT (309329.4 750090.5)
## 4  POINT (241677.4 719172.9)
## 5  POINT (375974.2 723760.4)
## 6    POINT (348406 717363.7)
## 7  POINT (300357.6 761365.5)
## 8  POINT (243878.4 722532.8)
## 9    POINT (368364 724406.6)
## 10 POINT (380003.1 737910.8)
##
## $sites_over
## Simple feature collection with 810 features and 20 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: 129595.6 ymin: 271950.4 xmax: 790773.7 ymax: 774611.3
## Projected CRS: FDEP_Albers_HARN
## First 10 features:
##           siteID siteuse replsite lon_WGS84 lat_WGS84 stratum            wgt
## 1  FLLL23001-091    Over     Next -86.19852  30.96293  Zone 1 1194.2013 [ha]
## 2  FLLL23001-092    Over     Next -84.37170  30.61329  Zone 1 1194.2013 [ha]
## 3  FLLL23001-093    Over     Next -84.60311  30.42443  Zone 1 1619.2784 [ha]
## 4  FLLL23001-094    Over     Next -85.01674  30.79476  Zone 1 1261.4117 [ha]
## 5  FLLL23001-095    Over     Next -85.58465  30.32025  Zone 1  875.6244 [ha]
## 6  FLLL23001-096    Over     Next -84.30644  30.53134  Zone 1  875.6244 [ha]
## 7  FLLL23001-097    Over     Next -84.19399  30.53212  Zone 1 1194.2013 [ha]
## 8  FLLL23001-098    Over     Next -84.98605  30.95543  Zone 1 1194.2013 [ha]
## 9  FLLL23001-099    Over     Next -85.74066  30.50359  Zone 1 1261.4117 [ha]
## 10 FLLL23001-100    Over     Next -84.32283  30.54994  Zone 1  875.6244 [ha]
##                     ip                       caty       gnis_name
## 1  0.0008373798 [1/ha]             Zone 1_[10,45]    Manning Pond
## 2  0.0008373798 [1/ha]             Zone 1_[10,45]            <NA>
## 3  0.0006175590 [1/ha] Zone 1_(1.43e+03,3.27e+03]    Lake Talquin
## 4  0.0007927626 [1/ha]            Zone 1_(45,144]    Cow Pen Pond
## 5  0.0011420422 [1/ha] Zone 1_(1.02e+03,1.43e+03] Deer Point Lake
## 6  0.0011420422 [1/ha] Zone 1_(1.02e+03,1.43e+03]    Lake Jackson
## 7  0.0008373798 [1/ha]             Zone 1_[10,45]    Lake Kanturk
## 8  0.0008373798 [1/ha]             Zone 1_[10,45]            <NA>
```

```
## 9  0.0007927626 [1/ha]           Zone 1_(45,144]       Russ Pond
## 10 0.0011420422 [1/ha] Zone 1_(1.02e+03,1.43e+03]    Lake Jackson
##         reachcode   hectares repunit   resource
## 1  03140202002828   12.86853  Zone 1 Large Lake
## 2  03120003001449   29.78604  Zone 1 Large Lake
## 3  03120003001574 3271.88680  Zone 1 Large Lake
## 4  03130004005001   76.12225  Zone 1 Large Lake
## 5  03140101006086 1200.39693  Zone 1 Large Lake
## 6  03120003001734 1425.85241  Zone 1 Large Lake
## 7  03120001004349   30.12627  Zone 1 Large Lake
## 8  03130004001588   35.45931  Zone 1 Large Lake
## 9  03140203002094   75.08377  Zone 1 Large Lake
## 10 03120003001734 1425.85241  Zone 1 Large Lake
##
comments
## 1
<NA>
## 2
<NA>
## 3
<NA>
## 4
<NA>
## 5
<NA>
## 6                                                                    Too shallow. Unable to
get boat close to location.
## 7  Would require hand launch of small boat. Violation of District COVID-19 distancing require
ments. Water level very low.
## 8
<NA>
## 9                        Would require hand launch of small boat. Violation of District COVID
-19 distancing requirements.
## 10                                                                   Too shallow. Unable to
get boat close to location.
##               lakecode shape_leng shape_area                  area_cat
## 1  3917-03140202002828   1433.270   128685.3             Zone 1_[10,45]
## 2  4055-03120003001449   3718.831   297860.4             Zone 1_[10,45]
## 3  4035-03120003001574 165804.747 32718868.0 Zone 1_(1.43e+03,3.27e+03]
## 4  4285-03130004005001   5203.878   761222.5            Zone 1_(45,144]
## 5  4387-03140101006086  56499.930 12003969.3 Zone 1_(1.02e+03,1.43e+03]
## 6  4068-03120003001734  53764.645 14258524.1 Zone 1_(1.02e+03,1.43e+03]
## 7  4295-03120001004349   5292.853   301262.7             Zone 1_[10,45]
## 8  4049-03130004001588   6093.373   354593.1             Zone 1_[10,45]
## 9  3934-03140203002094   9659.811   750837.7            Zone 1_(45,144]
## 10 4068-03120003001734  53764.645 14258524.1 Zone 1_(1.02e+03,1.43e+03]
##    sframe_stratum                  geometry
## 1          Zone 1 POINT (190085.1 774611.3)
## 2          Zone 1 POINT (364391.1 734001.7)
## 3          Zone 1 POINT (342120.2 713131.5)
## 4          Zone 1 POINT (302763.8 754484.2)
## 5          Zone 1 POINT (247776.7 702405.7)
```

```
## 6          Zone 1 POINT (370620.9 724891.2)
## 7          Zone 1   POINT (381402 724955.8)
## 8          Zone 1 POINT (305841.4 772285.2)
## 9          Zone 1 POINT (233078.7 722952.8)
## 10         Zone 1 POINT (369054.1 726958.8)
##
## $sites_near
## NULL
##
## $design
## $design$call
## grts(sframe = dsgn_ll, n_base = LL_base, stratum_var = "stratum",
##     seltype = LL_select, caty_var = "area_cat", caty_n = LL_catyn,
##     n_over = LL_over, wgt_units = "ha", pt_density = 1000, DesignID = "FLLL23001")
##
## $design$stratum_var
## [1] "stratum"
##
## $design$stratum
## [1] "Zone 1" "Zone 2" "Zone 3" "Zone 4" "Zone 5" "Zone 6"
##
## $design$n_base
## Zone 1 Zone 2 Zone 3 Zone 4 Zone 5 Zone 6
##     15     15     15     15     15     15
##
## $design$seltype
##    Zone 1    Zone 2    Zone 3    Zone 4    Zone 5    Zone 6
## "unequal" "unequal" "unequal" "unequal" "unequal" "unequal"
##
## $design$caty_var
## [1] "area_cat"
##
## $design$caty_n
## $design$caty_n$`Zone 1`
## Zone 1_(1.02e+03,1.43e+03] Zone 1_(1.43e+03,3.27e+03]
##                          3                          3
##      Zone 1_(144,1.02e+03]            Zone 1_(45,144]
##                          3                          3
##             Zone 1_[10,45]
##                          3
##
## $design$caty_n$`Zone 2`
##      Zone 2_(215,467]      Zone 2_(467,805]       Zone 2_(49,215]
##                     3                     3                     3
## Zone 2_(805,1.59e+03]        Zone 2_[10,49]
##                     3                     3
##
## $design$caty_n$`Zone 3`
## Zone 3_(1.78e+03,3.26e+03]      Zone 3_(182,1.78e+03]
##                          3                          3
## Zone 3_(3.26e+03,7.44e+03] Zone 3_(7.44e+03,1.76e+04]
##                          3                          3
```

```
##            Zone 3_[10,182]
##                          3
##
## $design$caty_n$`Zone 4`
## Zone 4_(1.32e+03,1.86e+03]           Zone 4_(175,429]
##                          3                          3
##      Zone 4_(429,1.32e+03]            Zone 4_(55,175]
##                          3                          3
##             Zone 4_[10,55]
##                          3
##
## $design$caty_n$`Zone 5`
## Zone 5_(1.89e+03,3.97e+03] Zone 5_(3.97e+03,9.66e+03]
##                          3                          3
##      Zone 5_(730,1.89e+03] Zone 5_(9.66e+03,1.25e+04]
##                          3                          3
##            Zone 5_[10,730]
##                          3
##
## $design$caty_n$`Zone 6`
##      Zone 6_(150,163]      Zone 6_(163,228] Zone 6_(228,1.29e+05]
##                     3                     3                     3
##       Zone 6_(59,150]        Zone 6_[10,59]
##                     3                     3
##
##
## $design$aux_var
## NULL
##
## $design$legacy
## [1] FALSE
##
## $design$mindis
## NULL
##
## $design$n_over
## $design$n_over$`Zone 1`
## [1] 135
##
## $design$n_over$`Zone 2`
## [1] 135
##
## $design$n_over$`Zone 3`
## [1] 135
##
## $design$n_over$`Zone 4`
## [1] 135
##
## $design$n_over$`Zone 5`
## [1] 135
##
## $design$n_over$`Zone 6`
```

```
## [1] 135
##
##
## $design$n_near
## NULL
```

```
dsgntime <- (proc.time() - dsgntime)/60
dsgntime
```

```
##     user   system  elapsed
## 28.03883  2.52550 30.67633
```

Print the column names of the survey design

```
names(sites)
```

```
## [1] "sites_legacy" "sites_base"   "sites_over"   "sites_near"   "design"
```

Merge the data for the base and oversample sites.
Check column names and rename as needed to match the names from previous years' selections.

```
sites<-sp_rbind(sites)
names(sites)
```

```
##  [1] "siteID"         "siteuse"        "replsite"       "lon_WGS84"
##  [5] "lat_WGS84"      "stratum"        "wgt"            "ip"
##  [9] "caty"           "gnis_name"      "reachcode"      "hectares"
## [13] "repunit"        "resource"       "comments"       "lakecode"
## [17] "shape_leng"     "shape_area"     "area_cat"       "sframe_stratum"
## [21] "geometry"
```

```
names(sites)[names(sites) == 'caty'] <- 'mdcaty'
names(sites)[names(sites) == 'sframe_stratum'] <- 'sframe_st'
names(sites)
```

```
##  [1] "siteID"     "siteuse"    "replsite"   "lon_WGS84"  "lat_WGS84"
##  [6] "stratum"    "wgt"        "ip"         "mdcaty"     "gnis_name"
## [11] "reachcode"  "hectares"   "repunit"    "resource"   "comments"
## [16] "lakecode"   "shape_leng" "shape_area" "area_cat"   "sframe_st"
## [21] "geometry"
```

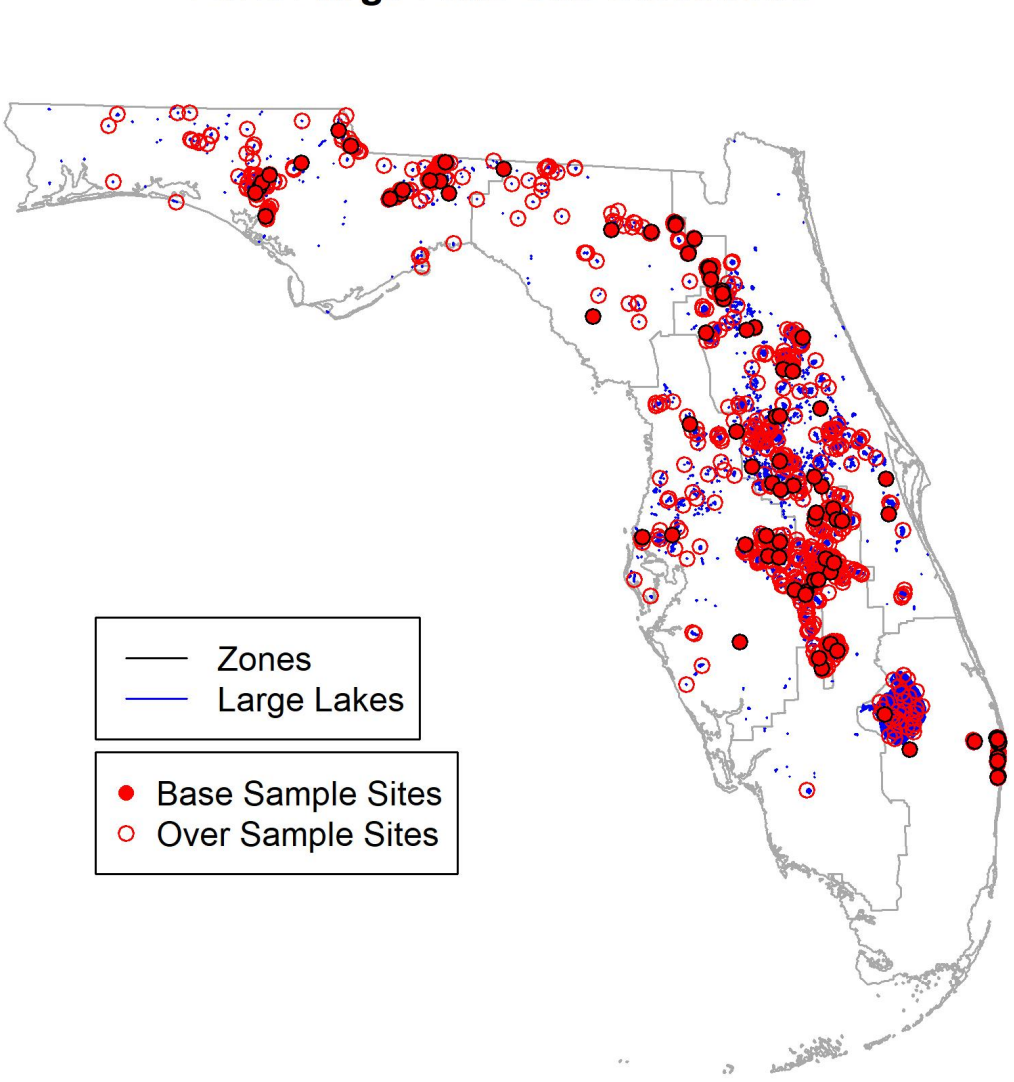
Check that the number of site selections for each Zone matches the requested design.

```
table(sites$stratum)
```

```
##
## Zone 1 Zone 2 Zone 3 Zone 4 Zone 5 Zone 6
##    150    150    150    150    150    150
```

Plot the Zone polygons, 2023 target population of large lakes, and 2023 site selections.

```
sites_base <-subset(sites, (sites$siteuse == 'Base'))
sites_over <-subset(sites, (sites$siteuse == 'Over'))
jpeg('2023_LL_Sites.jpg', units = 'in', width = 7, height = 7, res = 300)
plot(st_geometry(wms_c3_reporting_units), border='darkgray', main= '2023 Large Lake Site Selecti
ons')
plot(st_geometry(dsgn_ll), border = 'blue', col = 'blue', add = TRUE)
plot(st_geometry(sites_over), pch = 1, col = 'red', add = TRUE)
plot(st_geometry(sites_base), pch = 21, bg = 'red', add = TRUE)
legend(120000, 400000, legend=c('Zones','Large Lakes'), col=c('black','blue'),lty=c(1,1))
legend(120000, 300000, legend=c('Base Sample Sites','Over Sample Sites'), col=c('red','red'),pch
=c(16,1))
dev.off()
```



Check the selection summary and export the site selections in shapefile and CSV file formats.

```
sp_summary(sites, ~siteuse*stratum)
```

```
##    total     siteuse      stratum       siteuse:stratum
##  total:900   Base: 90   Zone 1:150   Over:Zone 1:135
##              Over:810   Zone 2:150   Over:Zone 2:135
##                         Zone 3:150   Over:Zone 3:135
##                         Zone 4:150   Over:Zone 4:135
##                         Zone 5:150   Over:Zone 5:135
##                         Zone 6:150   Over:Zone 6:135
##                                      Base:Zone 1: 15
##                                      Base:Zone 2: 15
##                                      Base:Zone 3: 15
##                                      (Other)    : 45
```

```
sf::st_write(sites,"2023_LL_sites.shp", append = FALSE)
```

```
## Deleting layer `2023_LL_sites' using driver `ESRI Shapefile'
## Writing layer `2023_LL_sites' to data source
##   `2023_LL_sites.shp' using driver `ESRI Shapefile'
## Writing 900 features with 20 fields and geometry type Point.
```

```
write.csv(sites, "2023_LL_site_selections.csv", row.names = FALSE)
```

Site Selection Summary

Description of Sample Design Output

The sites are provided as a shapefile that can be read directly by ArcGIS Pro.
The dbf file associated with the shapefile may be read by Excel.

The dbf file has the following variable definitions:
SiteID: Unique site identification (character)
siteUse: Whether the site is a legacy site (Legacy), base site (Base), reverse hierarchically ordered replacement site (Over), or nearest neighbor replacement site (Near).
replsite: The replacement site ordering. replsite is None if the site is not a replacement site, Next if it is the next reverse hierarchically ordered replacement site to use, or Near_, where the word following _ indicates the ordering of sites closest to the originally sampled site.
ln_WGS84: Longitude coordinates using the WGS84 coordinate system (EPSG:4326)
lt_WGS84: Latitude coordinates using the WGS84 coordinate system (EPSG:4326)
stratum: A stratum indicator. stratum is None if the sampling design was unstratified. If the sampling design was stratified, stratum indicates the stratum. *wgt*: Weight (in hectares), inverse of inclusion probability, to be used in statistical analyses.
ip: The site's original inclusion probability (the reciprocal of wgt).
mdcaty: An unequal probability grouping indicator. mdcaty is None if the sampling design did not use unequal inclusion probabilities. If the sampling design did use unequal inclusion probabilities, mdcaty indicates the unequal probability level.
gnis_nm: The name of the selected site's streams feature, derived from the USGS Geographic Names Information System database.
reachcd: The identification code for the selected site's corresponding streams flowline feature found in the National Hydrography Dataset for Florida.
hectares: The calculated area of the selected lake feature, expressed in hectares. *repunit*: The reporting unit (Zone) that the selected site is located within.

resourc: The water resource for the selected site's stream feature.
comments: Comments associated with the selected lake's feature.
lakecode: A unique feature ID containing the concatenation of the feature class OBJECTID and the REACHCODE for the selected site's river feature.
shape_leng: The length of the selected lake polygon feature, expressed in meters. *shape_area*: The area of the selected lake polygon feature, expressed in square meters. *sframe_st*: Strata used in the survey design. *geometry*: Geometry of the selected lake feature.

Projection Information
Projected Coordinate System: FDEP Albers HARN
Geographic Coordinate System: GCS_North_American_1983_HARN
Datum: D_North_American_1983_HARN
Spheroid: GRS_1980
Prime Meridian: Greenwich
Angular Unit: Degree
Projection: Albers
False_Easting: 400000.0
False_Northing: 0.0
Central_Meridian: -84.0
Standard_Parallel_1: 24.0
Standard_Parallel_2: 31.5
Latitude_Of_Origin: 24.0
Linear Unit: Meter

Evaluation Process

The survey design weights that are given in the design file assume that the survey design is implemented as designed. That is, all the base sites are used, and only the sites that are in the base sample (not in the over sample) are used.

As the base sites are evaluated, if a site is unable to be sampled for reasons described below, the base site is replaced with an over sample site to achieve the sample size planned. The site selection order must be followed when using over sample sites as replacements for base sites. When base sites are replaced, the survey design weights are no longer correct and must be adjusted before analyses can be performed. The weight adjustment requires knowing what happened to each site in the base design and any over sample sites that were evaluated as replacements.

If a site is unable to be sampled, it must be categorized into one of the following exclusion categories:
DRY
NO PERMISSION FROM OWNER
OTHERWISE UNSAMPLEABLE (e.g. safety concerns)
UNABLE TO ACCESS
WRONG RESOURCE/NOT PART OF TARGET POPULATION

Data analysts examine the site evaluation and exclusion information, and further reduce the sites into two categories for analysis.
1. *Non-Target* (TNTStatus = NT): Sites in the exclusion category *WRONG RESOURCE/NOT PART OF TARGET POPULATION*.
2. *Target* (TNTStatus = T): All other evaluated sites. Includes sites that were sampled and sites in exclusions categories *DRY*, *NO PERMISSION FROM OWNER*, *OTHERWISE UNSAMPLEABLE*, and *UNABLE TO ACCESS*.

Statistical Analysis

Any statistical analysis of data must incorporate information about the monitoring survey design. In particular, when estimates of characteristics for the entire target population are computed, the statistical analysis must account for any stratification or unequal probability selection in the design. Procedures for doing this are available from the resources listed in the bibliography. A statistical analysis library of functions is available to do common population estimates in the statistical software environment R.

Contacts for Questions and Inquiries about GRTS Sampling Program/Protocols

Anthony (Tony) R. Olsen
USEPA NHEERL
Western Ecology Division
200 S.W. 35th Street
Corvallis, OR 97333
Voice: (541) 754-4790
Fax: (541) 754-4716
email: Olsen.Tony@epa.gov (mailto:Olsen.Tony@epa.gov)

or

Michael Dumelle
USEPA
Freshwater Ecology Branch
Office of Research and Development
Corvalis, OR 97333
email: Dumelle.Michael@epa.gov (mailto:Dumelle.Michael@epa.gov)

Bibliography

Dumelle M., T.M. Kincaid, A.R. Olsen, and M.H. Weber. 2022. spsurvey: Spatial Sampling Design and Analysis. R package version 5.4.0.

Diaz-Ramos, S., D. L. Stevens, Jr, and A. R. Olsen. 1996. EMAP Statistical Methods Manual. EPA/620/R-96/002, U.S. Environmental Protection Agency, Office of Research and Development, NHEERL-Western Ecology Division, Corvallis, Oregon.

Stevens, D.L., Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics, 8:167-95.

Stevens, D.L., Jr. and Olsen, A.R. 1999. Spatially restricted surveys over time for aquatic resources. Journal of Agricultural, Biological, and Environmental Statistics, 4:415-428

Stevens, D. L., Jr., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14:593-610.

Stevens, D. L., Jr., and A. R. Olsen. 2004. Spatially-balanced sampling of natural resources in the presence of frame imperfections. Journal of American Statistical Association:99:262-278.