Using R to Assist with Invasive Species Monitoring and Challenge Biology Students

http://wdfw.wa.gov/ais/esox_lucius/

Stephen Hayes
Biology Lecturer
Gonzaga University

Introductions

Wildlife Biologist / Previous SAS User / R Convert

```
*/
libname sturgeon 'C:\Consulting\WDOW\Sturgeon\Data\Detections';
libname sturgloc 'C:\Consulting\WDOW\Sturgeon\Analysis\Location Analysis';
libname sturgmrg 'C:\Consulting\WDOW\Sturgeon\Analysis\Merged Analysis';
libname sturgdet 'C:\Consulting\WDOW\Sturgeon\Analysis\Models';

/*
*****     Run this section of the code once per season. Run the macro
portion, before the mallard data step, after isolating the appropriate
season;


data detect (keep=Date_and_Time__UTC_ Transmitter tag year month day date
date1 hour min sec time1 time_detect datetime loc);

set sturgeon.winter_detect;

loc = 0;

informat Date_and_Time__UTC_ $19.;
year       = substr(Date_and_Time__UTC_,1,4);
month      = substr(Date_and_Time__UTC_,6,2);
day        = substr(Date_and_Time__UTC_,9,2);
hour       = substr(Date_and_Time__UTC_,12,2);
min        = substr(Date_and_Time__UTC_,15,2);
sec        = substr(Date_and_Time__UTC_,18,2);
```

Northern Pike (*Esox lucius*)

"Pitiless Water-Wolf"

http://kalispeltribe.com/kalispel-natural-resources-department/northern-pike

Large: 2-4 feet, 10-15 lbs
Voracious, apex predator
Invasive Species in WA

Northern Pike Invasion in Pend Oreille River

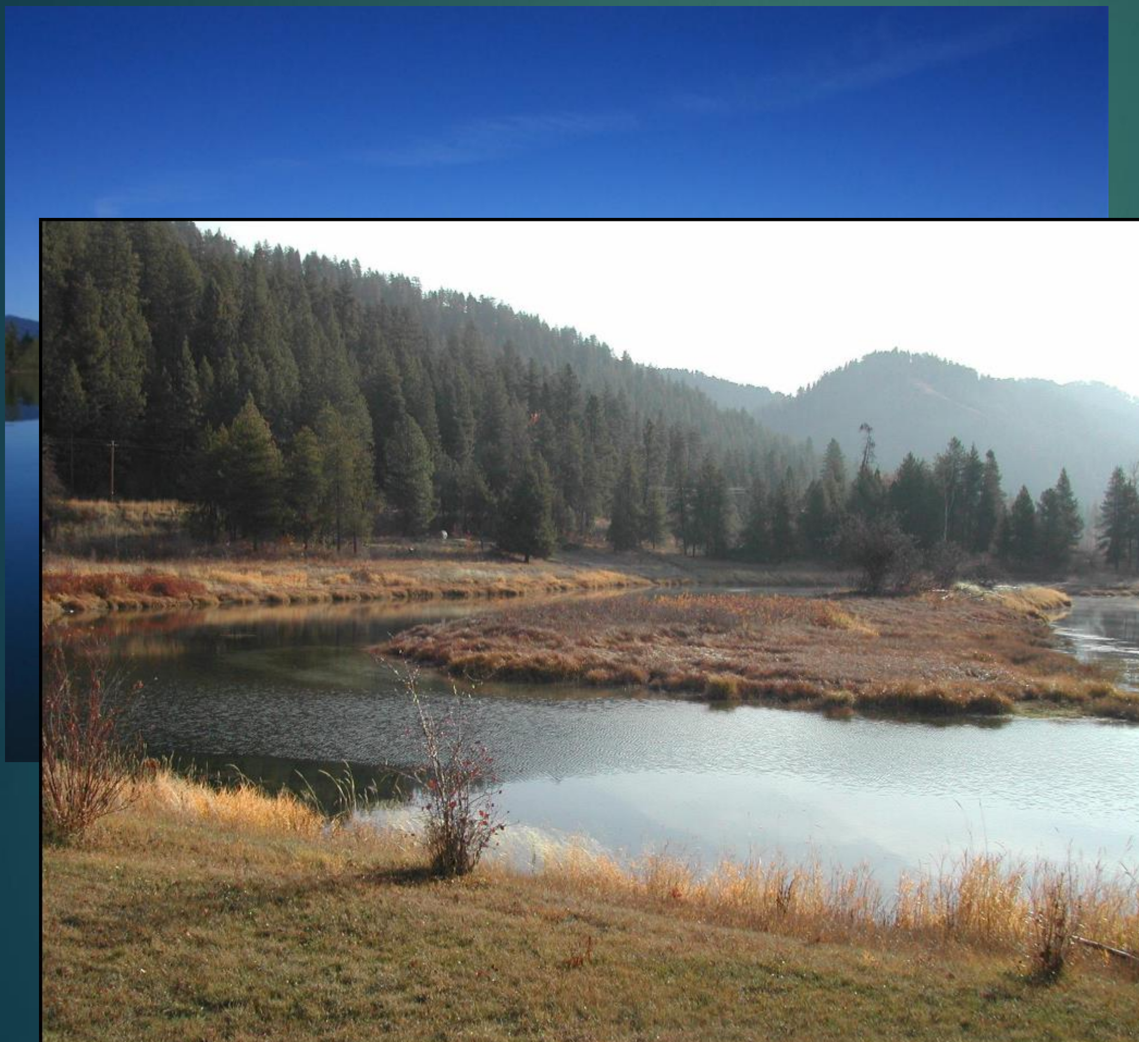
2004: Appeared in Box Canyon

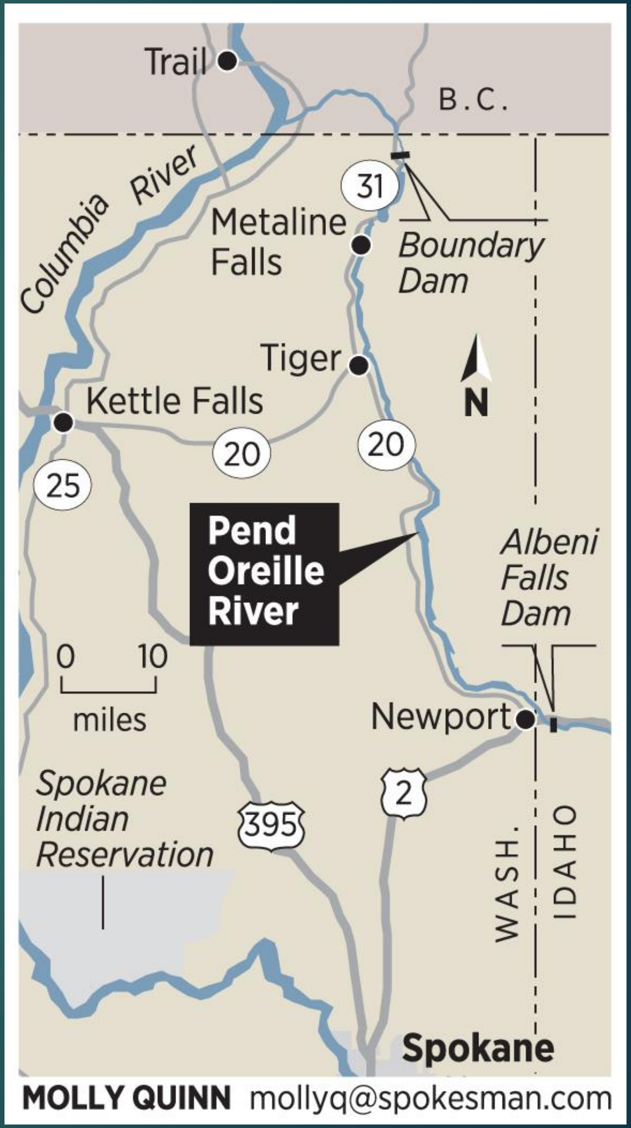
2006-2010: Exponential growth

2010 Population Estimate:
- At least 5,500 adults and 10,000 throughout river

Threats:
- Native fish
- Salmon/steelhead/trout recovery
- Downstream movement to Columbia River

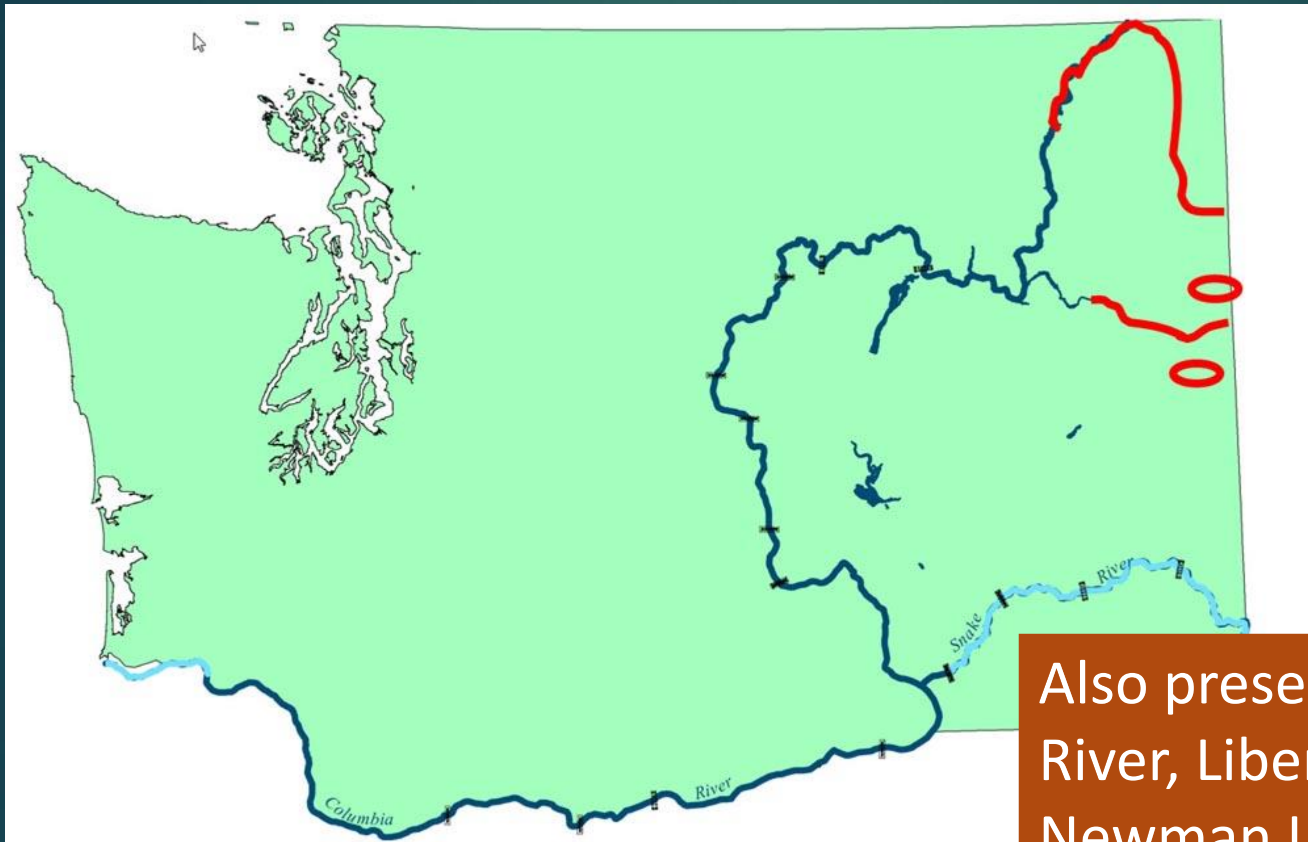
Pend Oreille River, WA





http://wdfw.wa.gov/publications/01169/

Northern Pike Invasion



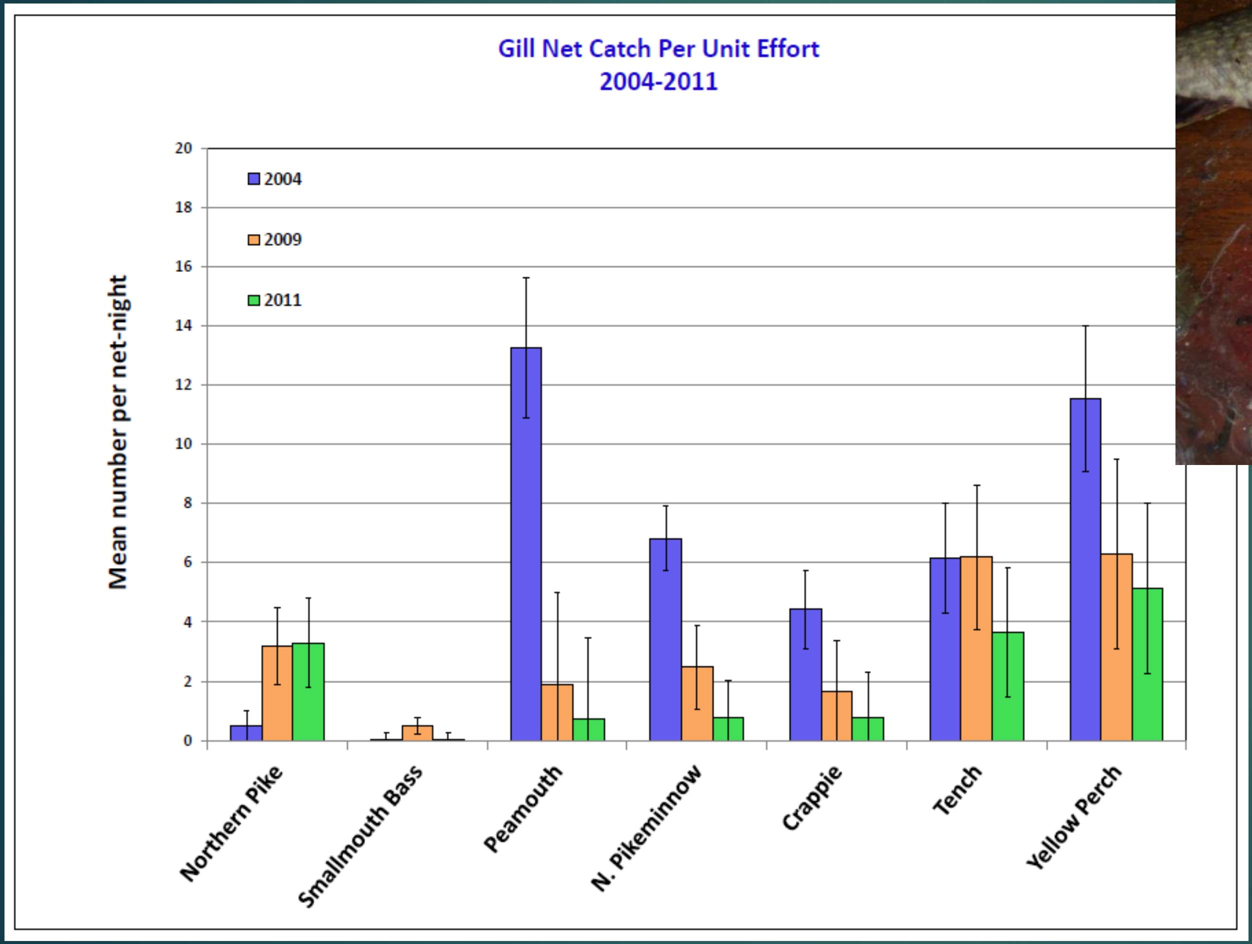
http://wdfw.wa.gov/ais/esox_lucius/

Also present in Spokane River, Liberty Lake, and Newman Lake

Where did they come from?

Illegal introduction…

Ecological Effects of Northern Pike Invasion



Native fish CPUE declined while Northern Pike CPUE increased

http://wdfw.wa.gov/publications/01169/

Northern Pike Removal

Since spring 2012, pike have been targeted for removal by:

- Angler Harvest
- Derbies
- Gill Nets

Target spawning pike in sloughs w/nets

http://wdfw.wa.gov/publications/01465/

Northern Pike Monitoring

http://wdfw.wa.gov/publications/01465/

What is the status of the invasion?

What is the removal goal?

Reduce pike CPUE to 2004-2006 levels

South End of Reservoir: ≤ 1.73 pike/net

North End of Reservoir: ≤ 0.5 pike/net

Monitoring Sampling Plan

Biologist Needs:

- Point and interval estimates of CPUE
- Estimates for North, South, and Sloughs
- Place several nets within sloughs
- Geographically "balanced" distribution of nets

http://wdfw.wa.gov/publications/01465/

Monitoring Sampling Plan

Statistical Constraints:

Population: Box Canyon Reservoir

What is the sample unit? (independence)

What is the sample frame?

How can a random sample be geographically “balanced”?

http://wdfw.wa.gov/publications/01465/

Monitoring Sampling Plan

Sample Unit: Net Location

- Independent if separated by 500 m
- Sloughs create a challenge…

http://wdfw.wa.gov/publications/01465/

Monitoring Sampling Plan

Sloughs are treated as clusters in a 2-stage sampling design…

World Imagery - Source: Esri and others

Why a geographically “balanced” sample?

Random samples tend to be clumped

Systematic samples don’t have design-based variance estimators

Solution?

Spatially Balanced Sampling of Natural Resources

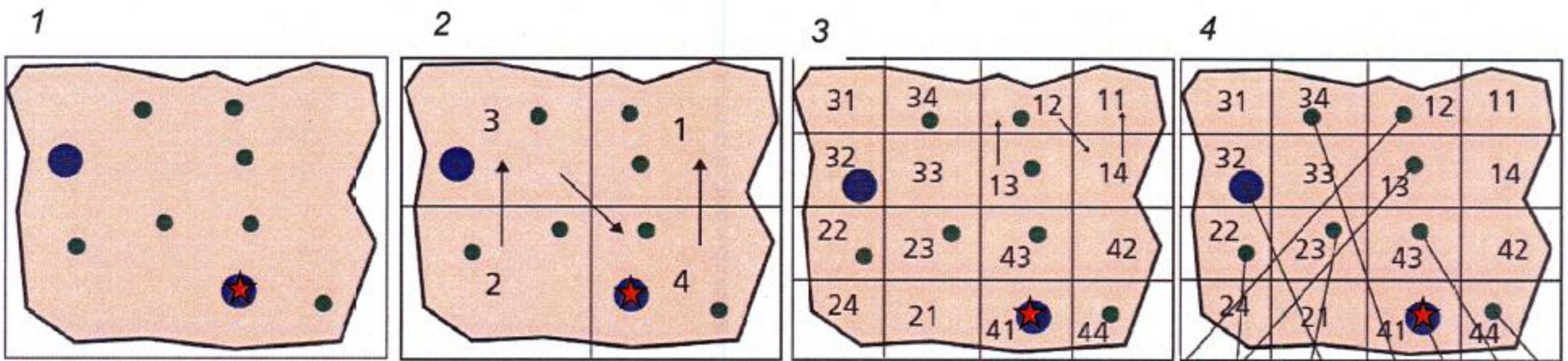
Don L. STEVENS Jr. and Anthony R. OLSEN

The spatial distribution of a natural resource is an important consideration in designing an efficient survey or monitoring program for the resource. Generally, sample sites that are spatially balanced, that is, more or less evenly dispersed over the extent of the resource, are more efficient than simple random sampling. We review a unified strategy for selecting spatially balanced probability samples of natural resources. The technique is based on creating a function that maps two-dimensional space into one-dimensional space, thereby defining an ordered spatial address. We use a restricted randomization to randomly order the addresses, so that systematic sampli
linear structure results in a spatially well-balanced random sample. Variable inclusion probability, propo
ancillary variable, is easily accommodated. The basic technique selects points in a two-dimensional conti
sampling finite populations or one-dimensional continua embedded in two-dimensional space. An extensi
a way to order the sample points so that any set of consecutively numbered points is in itself a spatially we
property is extremely useful in adjusting the sample for the frame imperfections common in environmental sampling.

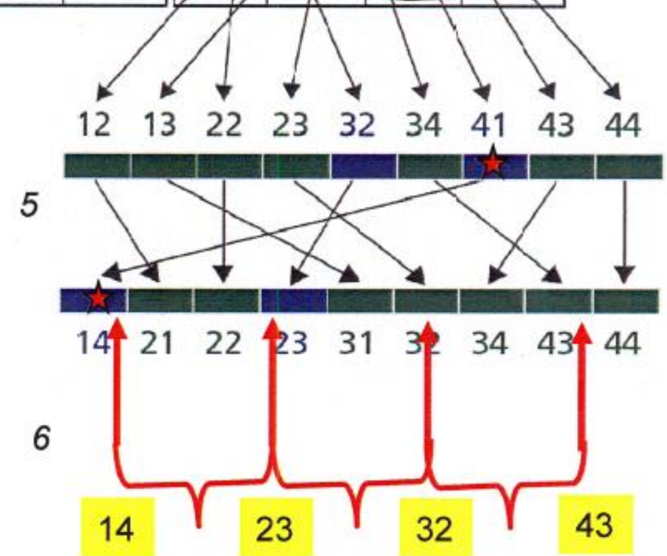
KEY WORDS: Environmental sampling; Imperfect sampling frame; Monitoring; Non-response; Spatial sampling; Survey design; Systematic sampling.

Journal of the American Statistical Association
March 2004, Vol. 99, No. 465, Theory and Methods
DOI 10.1198/016214504000000250

GRTS Sample Selection Process



- *Step 1*: Create sample frame
- *Step 2*: Randomly place grid over the region, sub-divide region and randomly assign recursive addresses to sub-regions
- *Step 3*: Sub-divide sub-regions; randomly assign numbers independently to each new sub-region creating hierarchical addresses. Continue sub-dividing until only one 'lake' per cell.
- *Step 4*: Identify each lake with cell address; assign each 'lake' length 1 (areal resource treated as points), or with length proportional to actual area
- *Step 5*: Place 'lakes' on line in numerical cell address order. *Reverse order of address digits and resort.*
- *Step 6*: Take a systematic sample with a random start on the line to generate spatially balanced (GRTS) sample



Final GRTS sample

GRTS can be implemented in R!

Package 'spsurvey'

October 23, 2015

Version 3.1

Date 2015-10-23

Title Spatial Survey Design and Analysis

Depends R (>= 2.10), sp

Imports methods, deldir, foreign, graphics, grDevices, MASS, rgeos, stats

Description This group of functions implements algorithms for design and analysis of probability surveys. The functions are tailored for Generalized Random Tessellation Stratified survey designs.

License GPL (>= 2)

URL `http://www.epa.gov/nheerl/arm/`

NeedsCompilation yes

Author Tom Kincaid [aut, cre],
Tony Olsen [aut],
Don Stevens [ctb],
Christian Platt [ctb],
Denis White [ctb],
Richard Remington [ctb]

Maintainer Tom Kincaid <Kincaid.Tom@epa.gov>

Monitoring Plan Implementation

A. Choose net locations

B. Sample (set nets at locations)

C. Point and interval estimate of CPUE

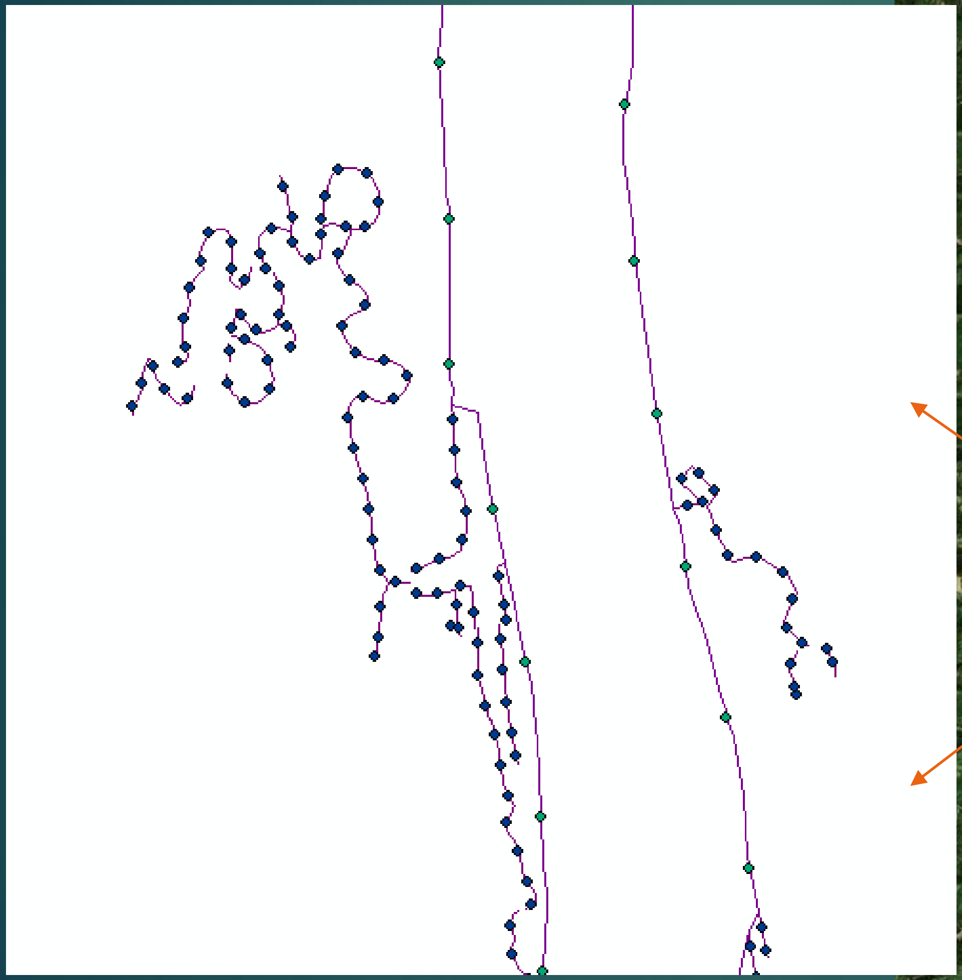
Choosing Net Locations

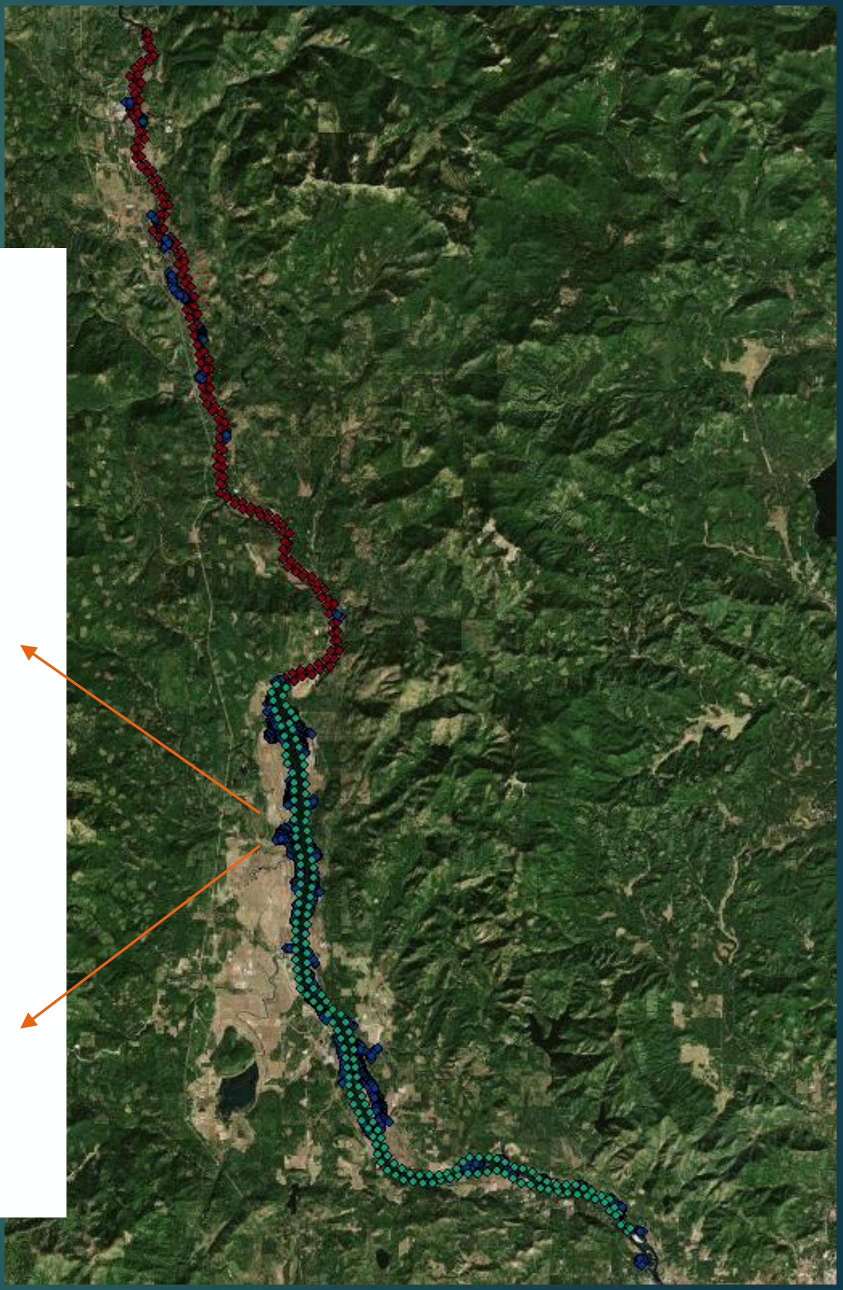
1. Create GIS file of Sample Frame
   a. North, South net locations
   b. Sloughs and net locations

2. Cluster sampling in sloughs requires 2 steps:
   a. Select PSU (slough)
   b. Select SSU within PSU (nets within sloughs)

3. Use spsurvey to select sample points

Sample Frame in GIS





Allocating Nets to Sloughs

a. Select sloughs (PSU's) with probability proportional to size (pps)

b. Select net locations (SSU's) to achieve geographical balance (spsurvey)

   i. Net allocation to sloughs based on area:

PSU size class	PSU size range	Number of ssu's sampled
Small	1 – 4	2
Small	5 – 6	3
Medium	7 – 12	4
Medium	13 – 20	5
Large	21 – 30	8
Large	> 30	10

Allocating Nets to Sloughs

Excel used to choose PSU’s

R used to select net locations

Not ideal, user friendly and met time constraint…

See Excel example file: “Slough Auto Allocation.xlsx”

See R example file: “Kalispel_grts_.r”

Monitoring Plan Implementation

A. Choose net locations ✓

B. Sample (set nets at locations) ✓

C. Point and interval estimate of CPUE
   i. 2-stage cluster sampling makes this interesting …

$$\hat{\mu} = W_N \overline{Y}_N + W_S \overline{Y}_S + W_{SL} \overline{Y}_{SL},$$

Point and Interval Estimation

Variance Estimation is “involved”:

The estimated variance of $\hat{\mu}$ is

$$s_{\hat{\mu}}^2 = W_N^2 s_{\bar{\bar{Y}}_N}^2 + W_S^2 s_{\bar{\bar{Y}}_S}^2 + W_{SL}^2 s_{\bar{\bar{Y}}_{SL}}^2 . \qquad (2.2)$$

The variance of $\bar{Y}_{Sl}$ is estimated as

$$s_{\bar{\bar{Y}}_{SL}}^2 = s_{PSU}^2 + s_{SSU}^2 ,$$

where,

$$s_{SSU}^2 = \sum_{i=1}^{n} \frac{M_i^2}{\pi_i M^2}\left(1 - \frac{m_i}{M_i}\right) s_i^2 / m_i \qquad (2.4)$$

Point and Interval Estimation

PSU Variance estimated with:

$$v\left(\widehat{Y}_{\mathrm{NHT}}\right) = \sum_{i \in s} c_i e_i^2 ,$$

where $e_i = \frac{y_i}{\pi_i} - \widehat{B}$ with

$$\widehat{B} = \frac{\sum_{i \in s} a_i \left(y_i / \pi_i\right)}{\sum_{i \in s} a_i}$$

Haziza et al. (2008)

Brewer 4	v_{B4}	$\frac{n}{n-1}\left(1 - \pi_i - \frac{\pi_i}{n-1} + n^{-1}(n-1)^{-1}\sum_{k \in U} \pi_k^2\right)$

See R example file: “SPIN_summary_example.r”

Results

Removal Efforts Are Working:

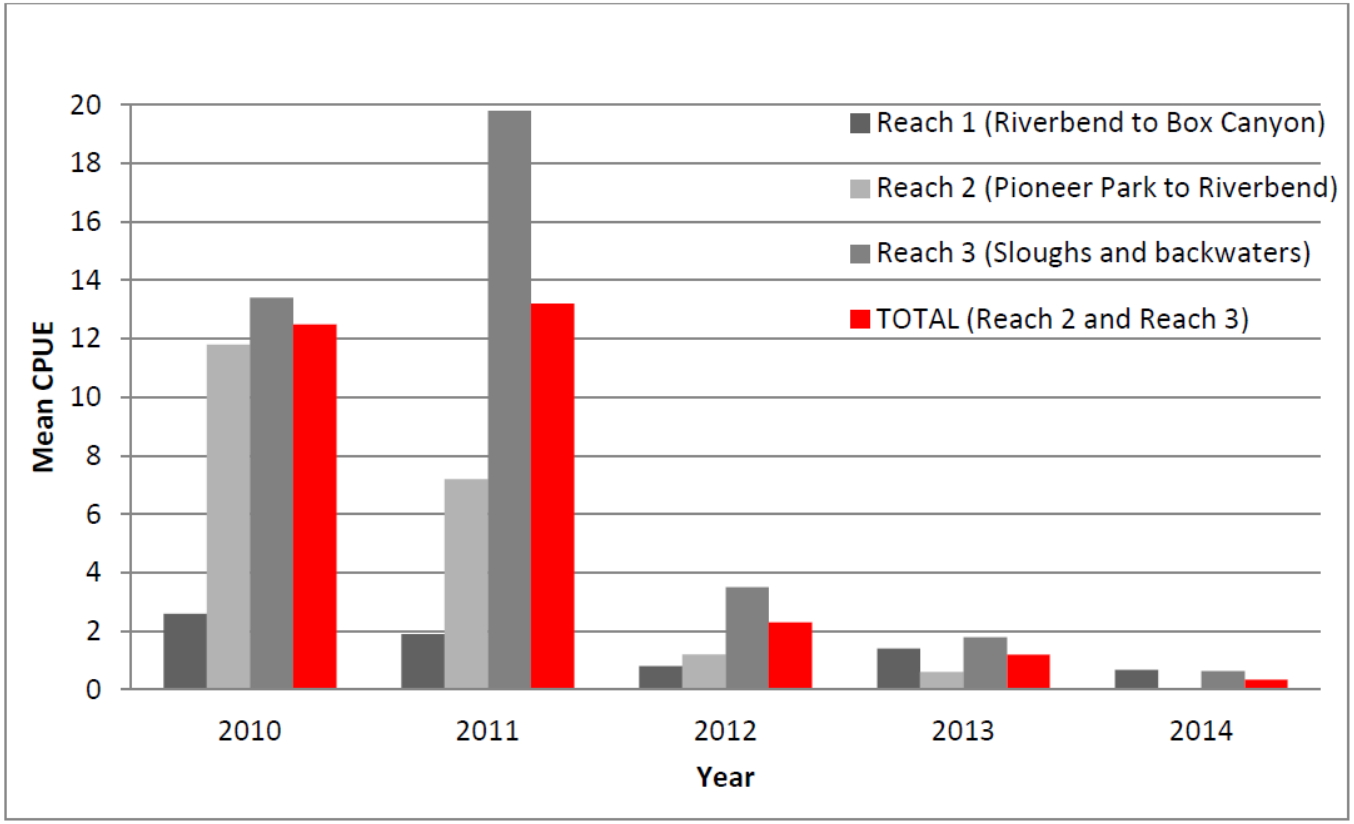


Figure 1. Spring Pike Index Netting Survey results for 2010-2014. Note that northern pike suppression was initiated in 2012 prior to the SPIN Survey that year.

What's missing here?

http://wdfw.wa.gov/publications/01636/wdfw01636.pdf

What’s Next?

Use R to allocate sample (not Excel)

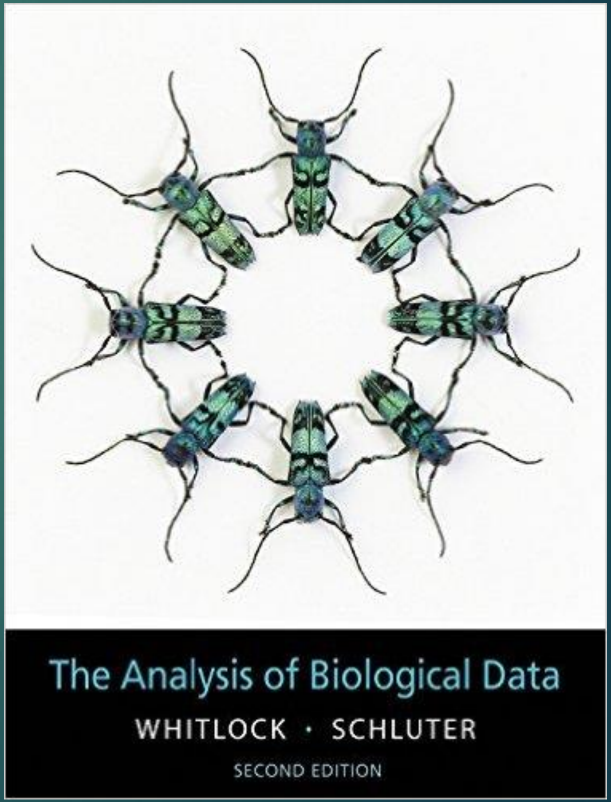
Simulation to estimate sample size for desired bound

“Nice” Interface for Users …. Ideas?

Undergraduate Biology Students and R

Biology 305
Upper Division Biology Elective
Fulfills Research Option Requirement

See:
http://whitlockschluter.zoology.ubc.ca/



Undergraduate Biology Students and R

Example Exercise:
R Simulation of the Sampling Distribution

{Inspired by “Teaching Statistics” (Gelman and Nolan 2002)}

The class records enters their height and their parents height in a vector.

Generate a sample from the population to estimate mean.

Write a function to generate sample, then a *for loop* to simulate the sampling distribution … takes 2 days and lots of help!