Alberta Environment FSI Workshop

Carl Schwarz

2017-05-14

Objectives

- Create a draft Watershed Assessment Report
- How many sample locations do I need in the watershed to detect a trend?
- What is impact of stratification of samples in a watershed

Workflow

- use *R* and *RMarkdown* to do computations and create *MSWord* (or other format) draft documents that can be edited to create final report

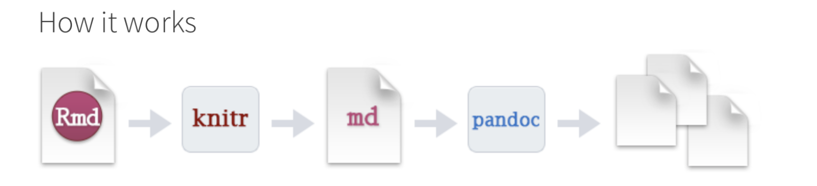


Figure 1:

Why this approach?

- Reproducible output.
- Final output and data are easy to syncronize.
- Reduce amount of copying and pasting output from *R* to *MSWord*.

But

- You can only get 90% of the way to the final report but need some customization unless you go to the next level of *RMarkdown* using *LaTeX* (a professional typesetting program).

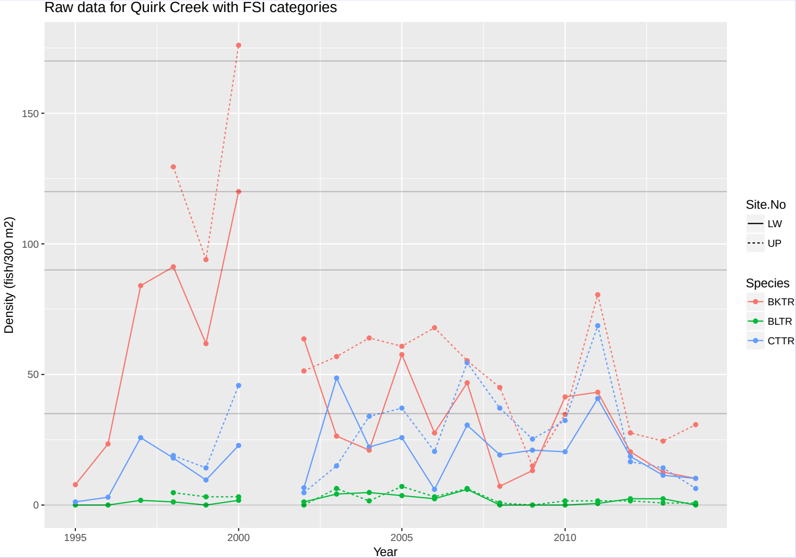
Computer Preparation

- Install *R* and associated packages.
  Caution with *ggmaps* as current version doesn't talk nicely with Google Maps. You will need to install from GitHub
- Install *RStudio* and associated *R* packages.
- Install *JAGS* and associated *R* packages.
- Get FWIS data for your watershed.
  We will use Quirk Creek as an example.
- Download material from GitHub
  Visit `https://github.com/cschwarz-stat-sfu-ca/ABgov-fish`
  Download zip file.
  Unzip zip file and discard zip file.

Watershed Assessment

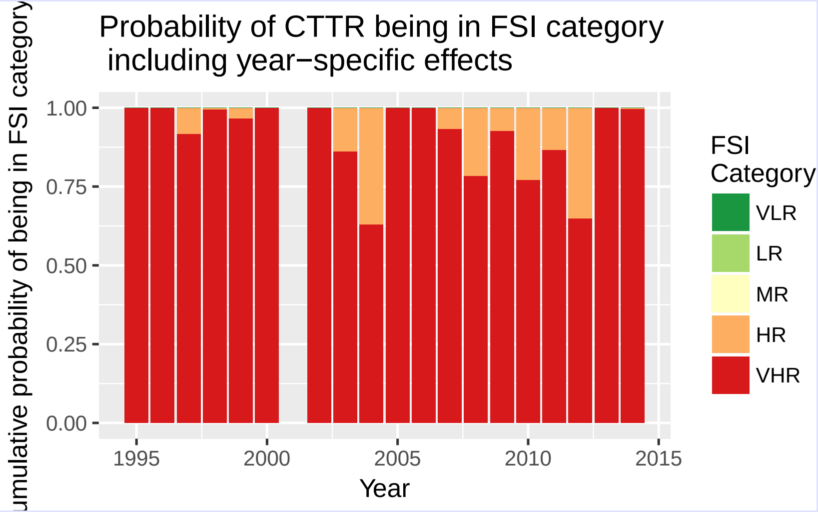
Watershed Assessment

Take this:



Watershed Assessment II

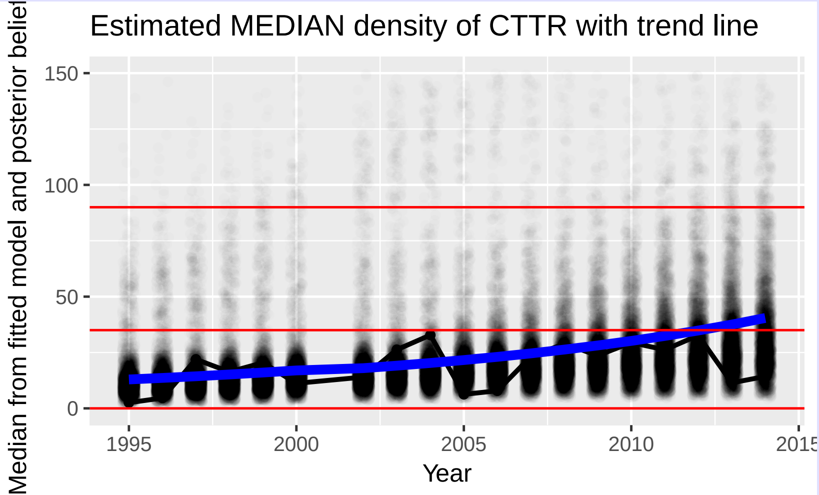
Objective: Take previous plot and classify watershed as:



But: - Year-specific effects add extra layer of variation

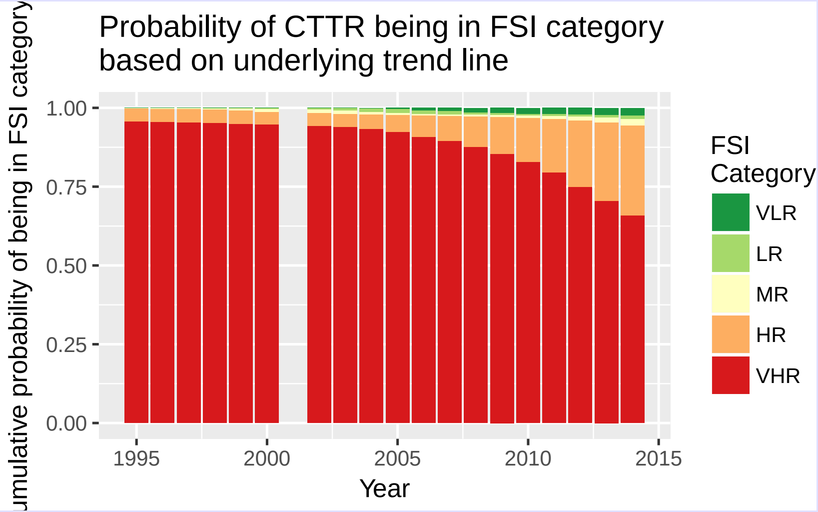
Watershed Assessment III

Estimate the underlying trend:

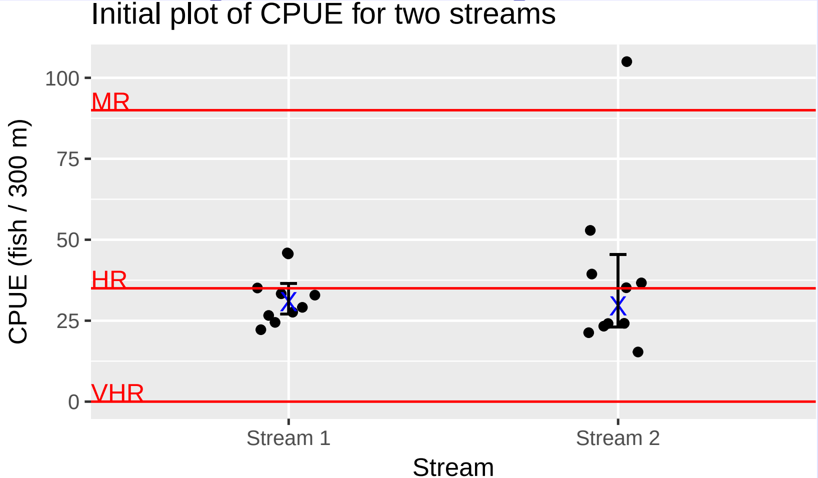


Watershed Assessment IV

Base FSI classification based on underlying trend:



Probabilistic assignment to FSI categories



What is the P(median (blue X)) lies in EACH FSI category?

- Bootstrap/simulate from each set of data.
- Compute the median from each bootstrap sample.
- What proportion of medians fall in each FSI category?

Probabilistic assignment to FSI categories

Generated values for median from Stream 1:

```
     log.median median
[1,]       3.61  37.03
[2,]       3.31  27.37
[3,]       3.38  29.23
[4,]       3.45  31.60
[5,]       3.43  30.87
```

Create a 0/1 indicator variable for each FSI category:

```
     prob.FSI.cat[1] prob.FSI.cat[2] prob.FSI.cat[3] prob.F
[1,]               0               1               0
[2,]               1               0               0
[3,]               1               0               0
[4,]               1               0               0
[5,]               1               0               0
```

Probabilistic assignment to FSI categories

Repeat for many thousands of simulations:

```
Probability of Stream 1 in each FSI category

                Probability
prob.FSI.cat[1]        0.89
prob.FSI.cat[2]        0.11
prob.FSI.cat[3]        0.00
prob.FSI.cat[4]        0.00
prob.FSI.cat[5]        0.00
```

Probabilistic assignment to FSI categories

Similarly for Stream 2:

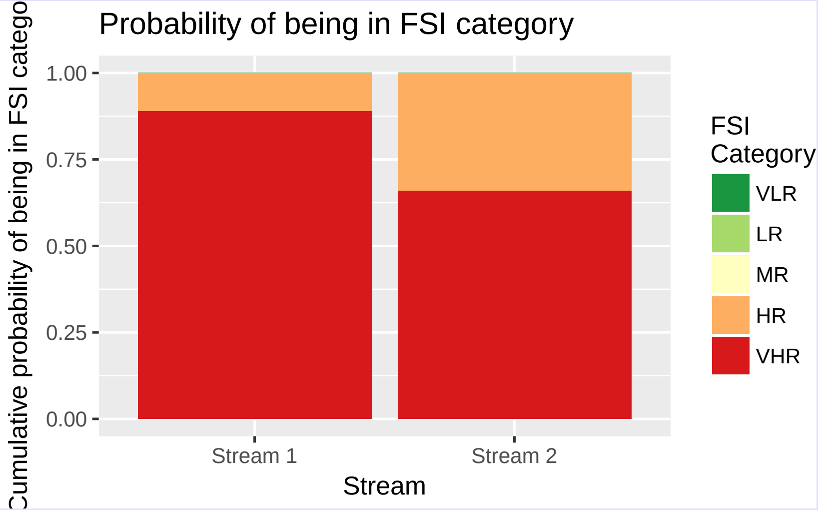
The same set of computation can be done for Stream 2 and we obtain the following results for Stream 2.

```
Probability of Stream 2 in each FSI category


                 Probability
prob.FSI.cat[1]         0.66
prob.FSI.cat[2]         0.34
prob.FSI.cat[3]         0.00
prob.FSI.cat[4]         0.00
prob.FSI.cat[5]         0.00
```

Probabilistic assignment to FSI categories

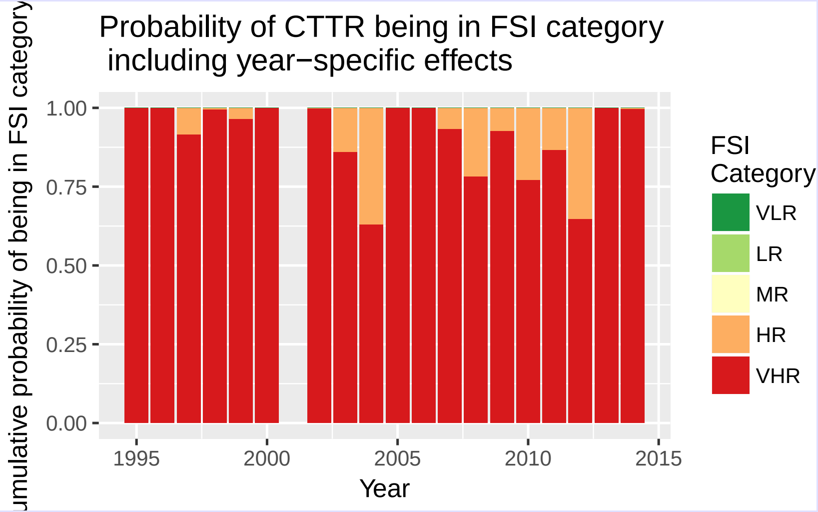
Create categories from the probabilties



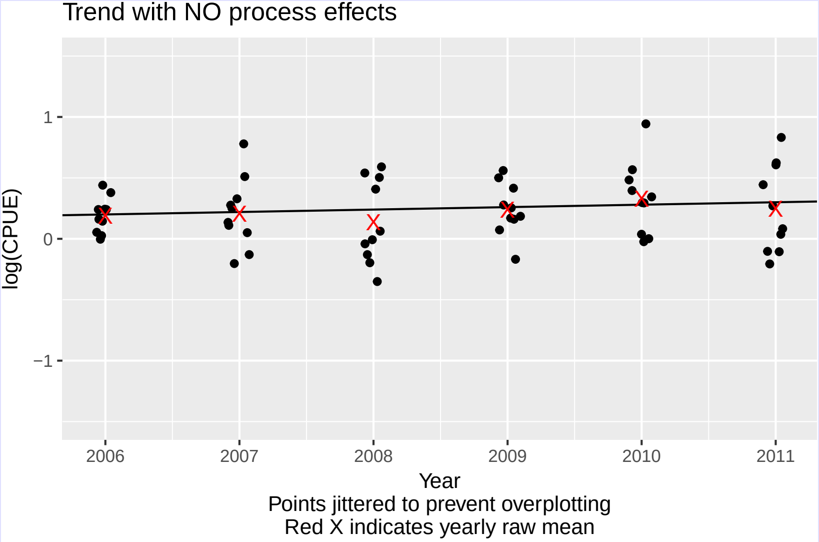
Year-Specific Effects (Process Error)

Role of process error (year-specific effects)

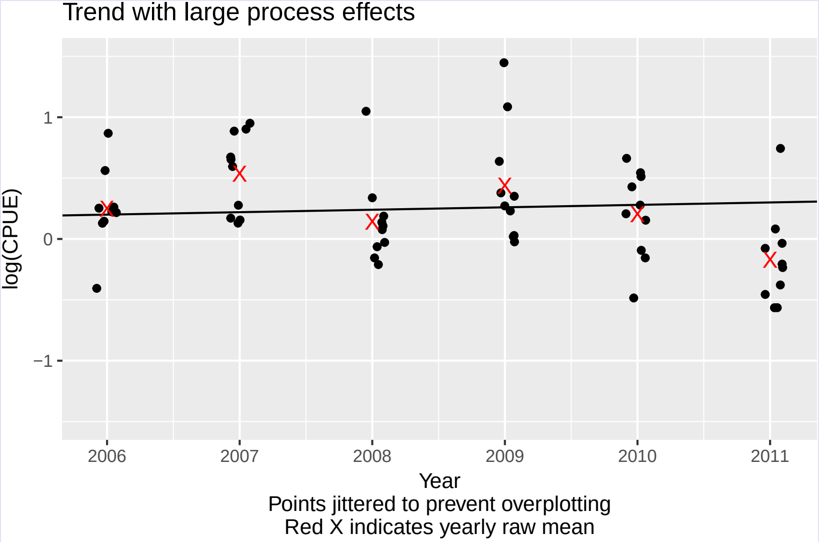
Year-specific effects add extra variaton to the assessment



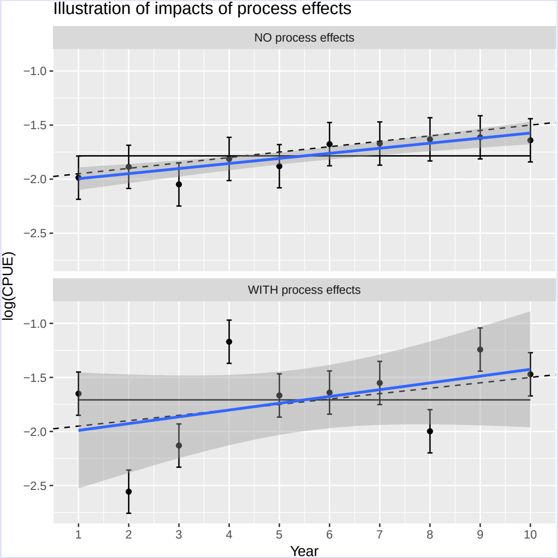
What are year-specific effects (process error)?



What are year-specific effects (process error)?



Impacts of year-specific effects (process error)



Fitting trend accounting for year-specific effects (process error)

Linear mixed model accounting for yearly trend and yearly process error:

```
plotdata.PE$YearC <- factor(plotdata.PE$Year)
mixed.fit <- lmerTest::lmer(logCPUE ~ Year + (1|YearC), dat
summary(mixed.fit)$coefficients
```

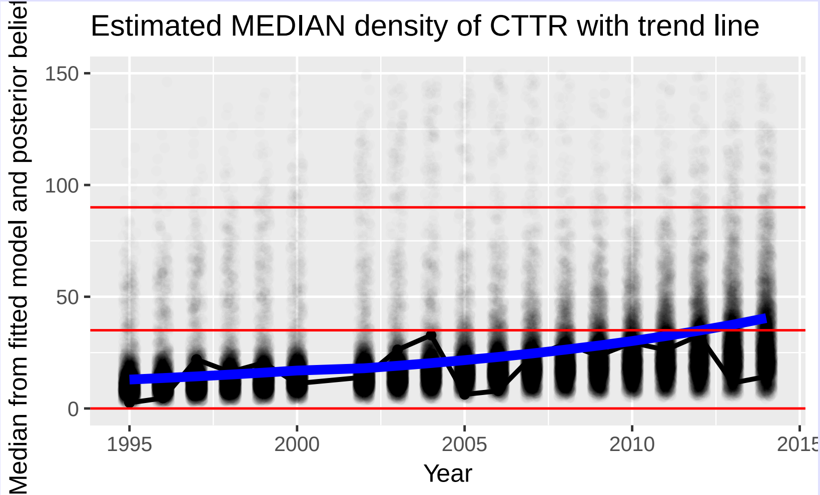
```
##                Estimate   Std. Error       df   t value
## (Intercept) 160.66702202 106.48797177 4.009495  1.50878
## Year         -0.07987644   0.05301864 4.009495 -1.50657
```

```
VarCorr(mixed.fit)
```

```
##  Groups   Name        Std.Dev.
##  YearC    (Intercept) 0.18682
##  Residual             0.37803
```

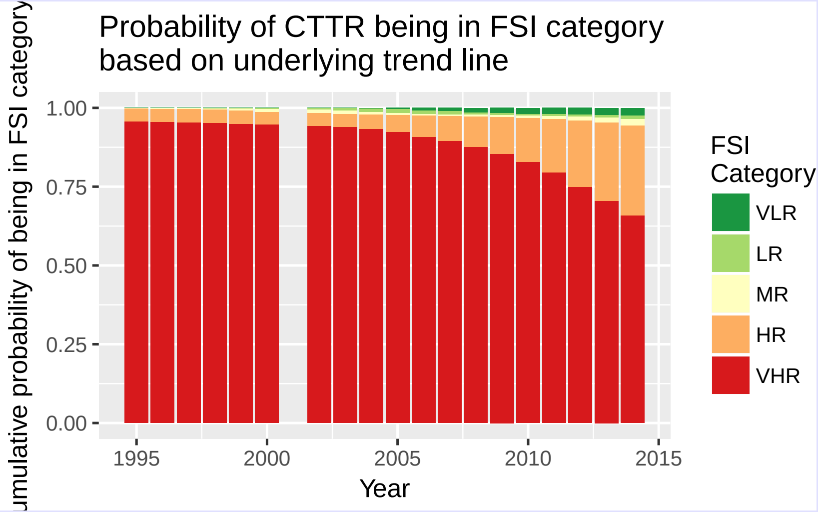
Fit trend on log(CPUE) scale

Similar LMM fit but also add SITE effects:



Fit trend on log(CPUE) scale

Assign FSI category based on variation of TREND line



Creating draft watershed assessment report

Draft watershed assessment report: Workflow

- Install *R*, *RStudio*, *JAGS* and associated packages
- Create working directory for your work.
- Download files from `https://github.com/cschwarz-stat-sfu-ca/ABgov-fish` into working directory and unzip
- Move FWIS data file to Data directory of WatershedReport directory.
- Duplicate and rename *WatershedTemplate.Rmd* and *SpeciesSubsection-BKTR.Rmd*
- Clone *SpeciesSubsection-BKTR.Rmd* file.
- Edit and Run *xx-WatershedTemplate.Rmd* file.
- View *xx-WatershedTemplate.docx* until satisfied.
- Edit *docx* file for finishing touches.

RMarkdown: Combining *R* and text into a single report.

Types of chunks:

- The headers (between the —'s)
- Code chunks. These are sections of *R* code of the form delimited by triple BACK quotes with *R* code in between.
- Text that includes results of *R* expressions

Using *RMarkdown*

- Run complete chunks using the *Run* menu item in *RStudio*.
- Run sections of code by highliting and pressing CTRL-R (Windows) or CMD-Return (Macintosh)
- Help availble from *RStudio* Help menu and Google
  `http://rmarkdown.rstudio.com`
  `https://yihui.name/knitr/options/`

Basic steps to creating draft watershed assessment report

- Carefully go through first *R* code chunk ensuring that data has been entered properly, and making sure that you select proper subset, clean up any coding errors, etc
- Run main document (excluding Species sections) until happy.
- Do ONE species at a time template revisions until done.
- Write your final summary message.
- Finally *knit* everything together into final *MSWord* document.

Exercise: Quirk Creek Watershed Assessment

Refer to detailed document in the WatershedReport directory.

How many years do I need to sample to detect a trend?

Detecting a trend

Power to detect a trend depends on

- Size of trend to detect. Larger trends are easier to detect than smaller trends.
- Sample size
  - Number of years of sampling.
  - Number of sites sampled each year.
- Variance components (noise)
  - Year-specific effect standard deviation (process error)
  - Site-to-Site effect standard deviation (cancels when the site is repeatedly measured over time)
  - Residual noise standard deviation.

The year-specific effects standard deviation is OFTEN THE LIMITING factor and CANNOT BE CONTROLLED!

- Not nessary to sample large number of sites in each year.

Detecting a trend - workflow

- Read in and edit FWIS data as before
- Estimate the variance components (standard deviations)
- Estimate power to detect various size trends under different scenarios
- Plot the data
- Be demoralized.

Exercise: Quirk Creek Power and Sample Size

Refer to detailed document in TrendPowerAnalysis directory.

Impact of stratification

Impact of different types of stratification

Uncertainty in estimating median in a year may be due to other variables such as

- HUC Classification
- Stream Order
- Other variables

If this uncertainty could be reduced, it may improve the FSI categorization

BUT

- Year-specific effect (process error) may render this moot.

Estimating uncertainty (theory)

- SRS (simple random sample)
  - $Y_i = log(CPUE)$.
  - s is the overall standard deviation.
  - $SE(\overline{Y}) = s/\sqrt{n}$
- Stratified SRS
  - $\overline{Y}_h$ mean in stratum h
  - s_h standard deviation in stratum h
  - n_h sample size in stratum h
  - $SE(\overline{Y}_h) = s_h/\sqrt{n_h}$ SE in stratum h
  - W_h population level stratum weight
  - $\overline{Y}_{overall} = W_1\overline{Y}_1 + W_2\overline{Y}_2 + ...$
  - $SE(\overline{Y}_{overall}) = \sqrt{W_1^2 SE_1^2 + W_2^2 SE_2^2 + ...}$

Sample allocation

- Equal allocation: $n_h = n_{total}/H$ where H is number of strata
- Proportional allocation: $n_h = n_{total} \times W_h$
- Optimal allocation $n_h = n_{total} \times \frac{W_h S_h}{W_1 S_1 + W_2 S_2 + ...}$.

Most gains in precision occur moving from Equal to Proportional Allocation; moving to optimal allocation typically leads to smaller improvments

Stratification - Workflow

- Read and summarize CPUE information. CAUTION. Current FWIS format does not include HUC10 or stream order
- Find POPLATION WEIGHTS from GIS for each stratification option
- Estimate s_h for each stratum
- Do trial allocaitons, say with $n_{tota} = 100$
- Estimate uncertainty in mean

Is there a benefit from stratification vs. no stratification?
Is there much difference between two types of stratification? Is there much gain moving from Equal -> Proportional -> Optimal?

Exercise: Nordegg River stratification