# Processing survey length/age composition data
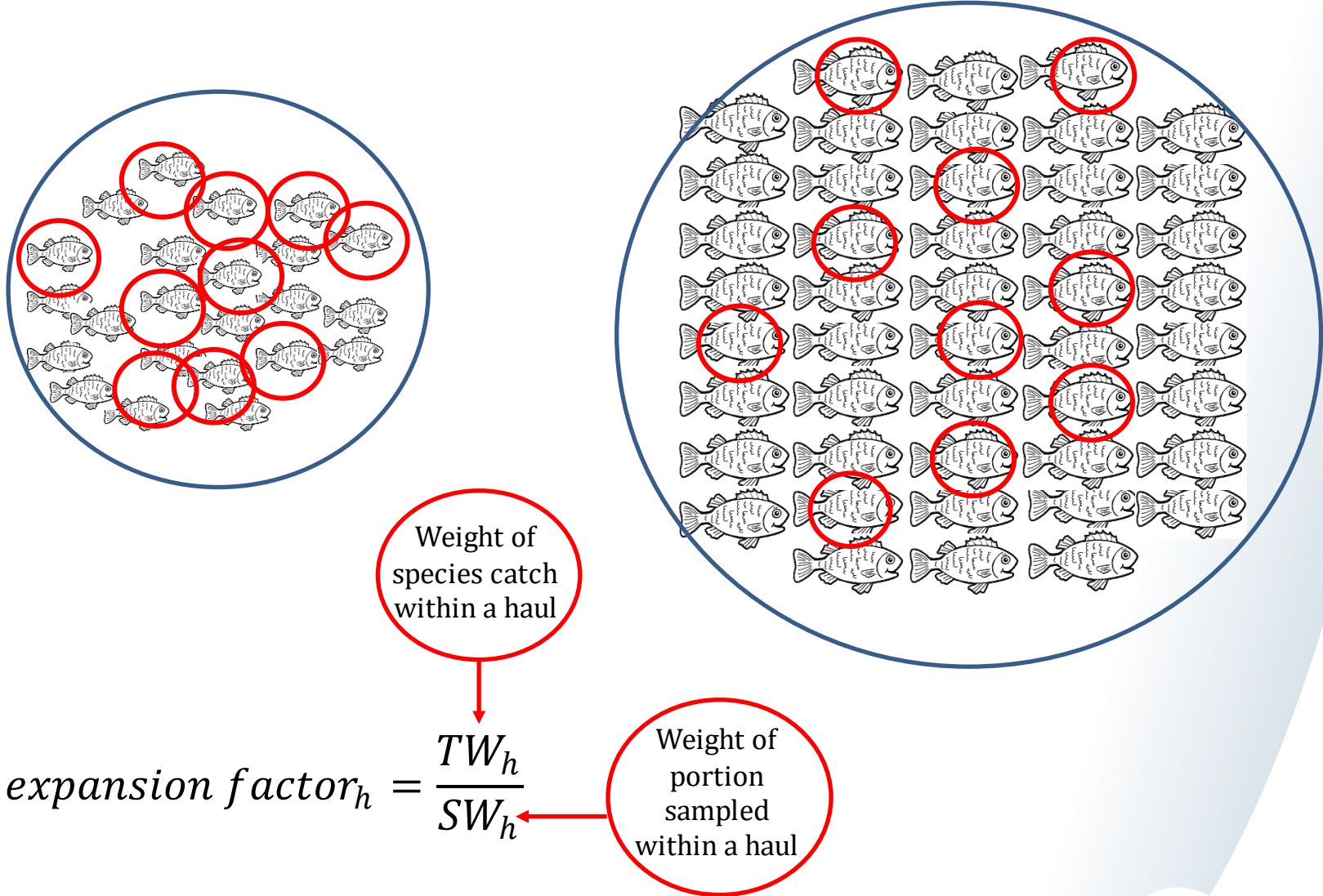
FISH576, Week 2

# What's length and age composition data used for in the model?

- To estimate growth parameters
- To estimate natural mortality
- To estimate recruitment and recruitment deviations
- To describe survey selectivity curves
- To describe fishery selectivity and retention curves

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service

**NOAA FISHERIES**
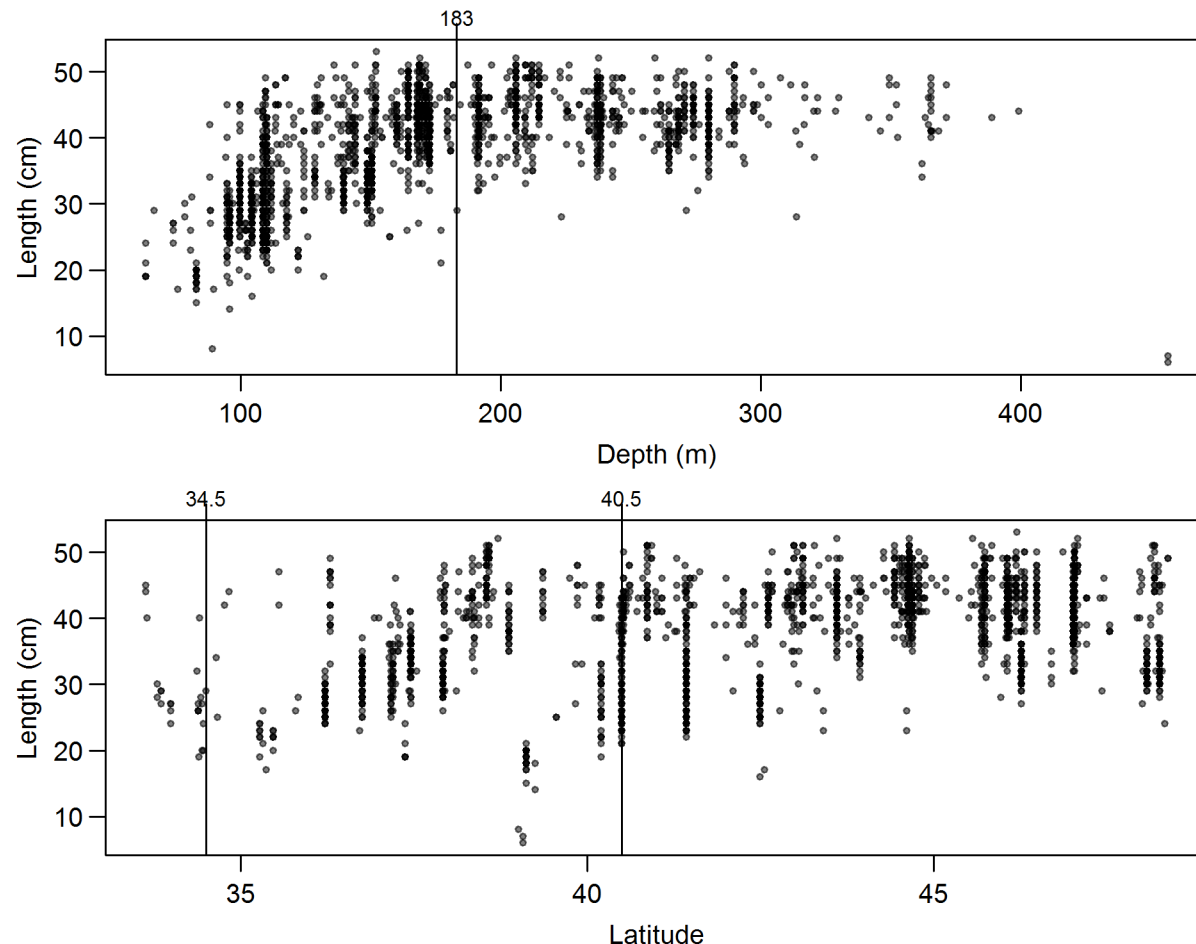
# Length composition data

- Set # lengths for each species randomly sampled each haul, no matter of the amount of catch.

- The numbers (and weight) of fish sampled each survey haul are not proportional to amount caught.

- Therefore, we need to account for differences in catch among hauls.

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service

# Length composition expansion: stage one



Weight of species catch within a haul

$$expansion\ factor_h = \frac{TW_h}{SW_h}$$

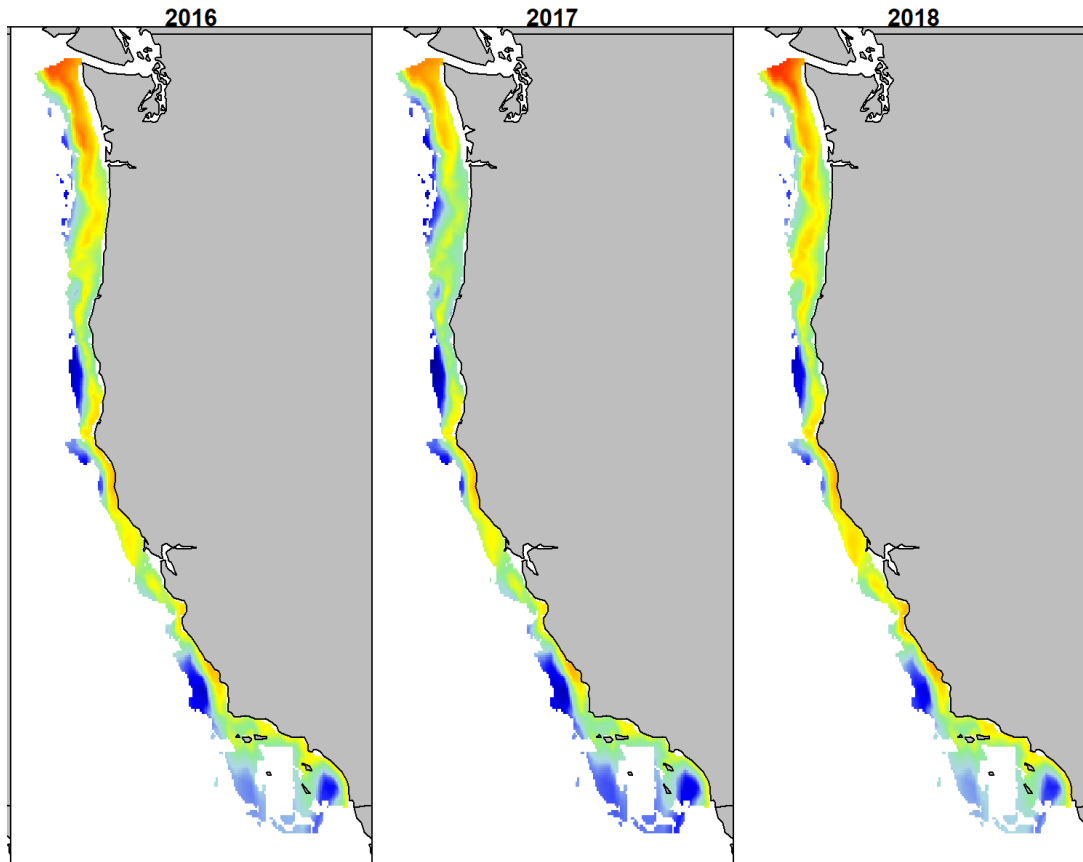Weight of portion sampled within a haul

**NOAA FISHERIES**

# Why composition data expansion matter

- Fish tend of aggregate by size.

- In groundfish we often observe ontogenetic movements.



Widow rockfish example: smaller fish are found in shallower waters and more often in southern areas

# Length composition expansion: stage two



2016   2017   2018

Spiny dogfish shark example: fish are more abundant in the north, around U.S.-Canada border
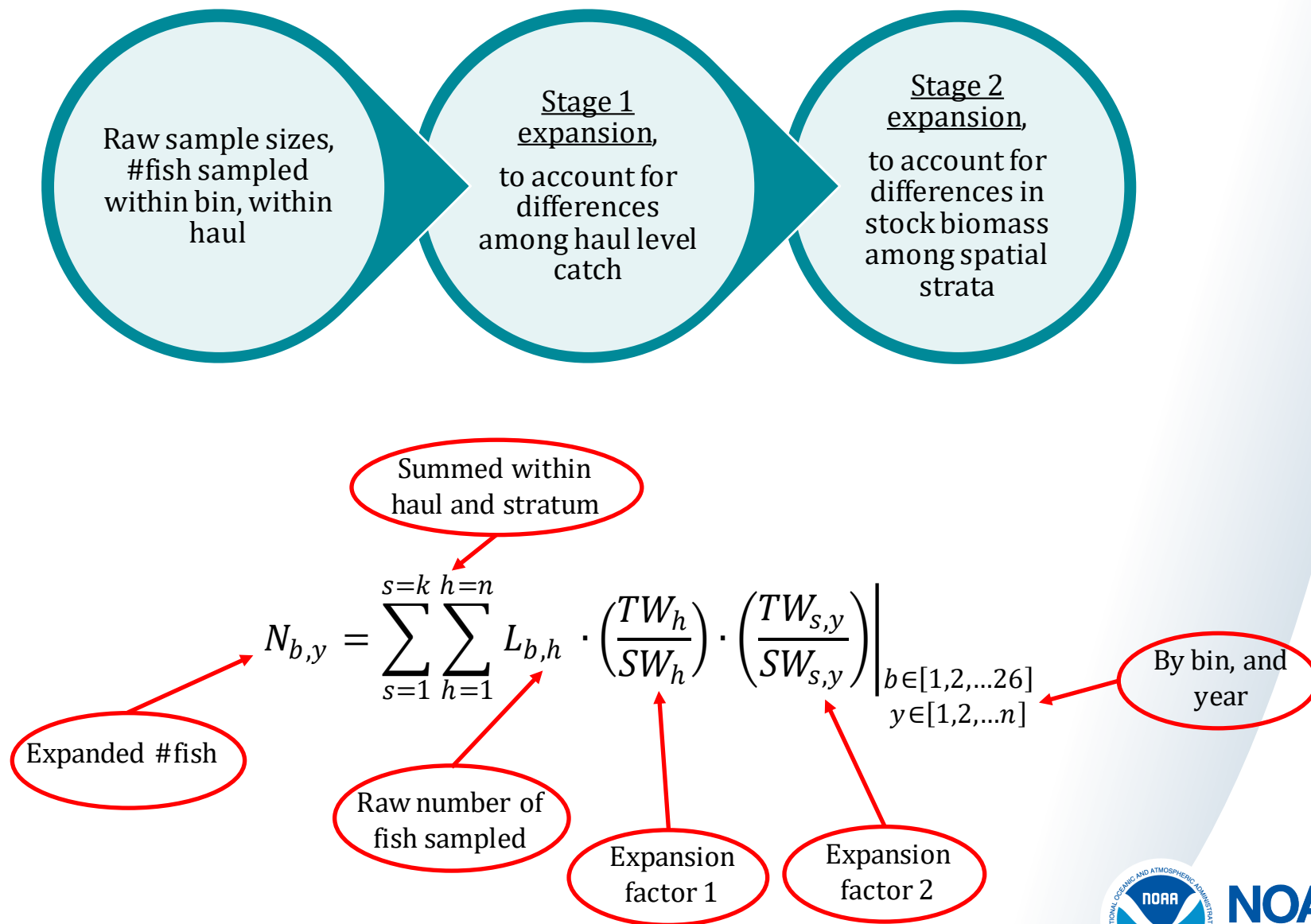
Stock biomass within stratum

$$expansion\ factor_y = \frac{TW_s}{SW_s}$$

Weight of sampled fish within a stratum

**NOAA FISHERIES**

# Putting everything together:

**Raw sample sizes, #fish sampled within bin, within haul**

**Stage 1 expansion,** to account for differences among haul level catch

**Stage 2 expansion,** to account for differences in stock biomass among spatial strata

Summed within haul and stratum

$$N_{b,y} = \sum_{s=1}^{s=k} \sum_{h=1}^{h=n} L_{b,h} \cdot \left(\frac{TW_h}{SW_h}\right) \cdot \left(\frac{TW_{s,y}}{SW_{s,y}}\right)\Bigg|_{\substack{b \in [1,2,\ldots 26] \\ y \in [1,2,\ldots n]}}$$

Expanded #fish

Raw number of fish sampled

Expansion factor 1

Expansion factor 2

By bin, and year

**NOAA FISHERIES**

# Luckily!

- We have a generalized R code, to expand the data and put it in the format needed for ss3 input file.

- It is publicly available on GitHub at: https://github.com/pfmc-assessments/nwfscSurvey

- It is important to explore the data, there are a lot of details, specific to individual species that are not be accounted for in the generalized code.

- Generalized code require species specific inputs, that include length and age bins, spatial strata you want to use in your analysis.

NOAA FISHERIES

# Length and age bins

- Length data are usually split by 2 cm bins (for large organisms can be 5cm).

- Rule of thumb – the number of length bins in many assessment is about 30

- Lowest bin includes all fish in that bin and smaller.

- Highest bin includes all fish in that bin and larger

- Rule of thumb –highest bin should not have more than 5% of data.

- Ages are usually split by 1 year bins, with highest bin accumulating fish older than that bin.

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service
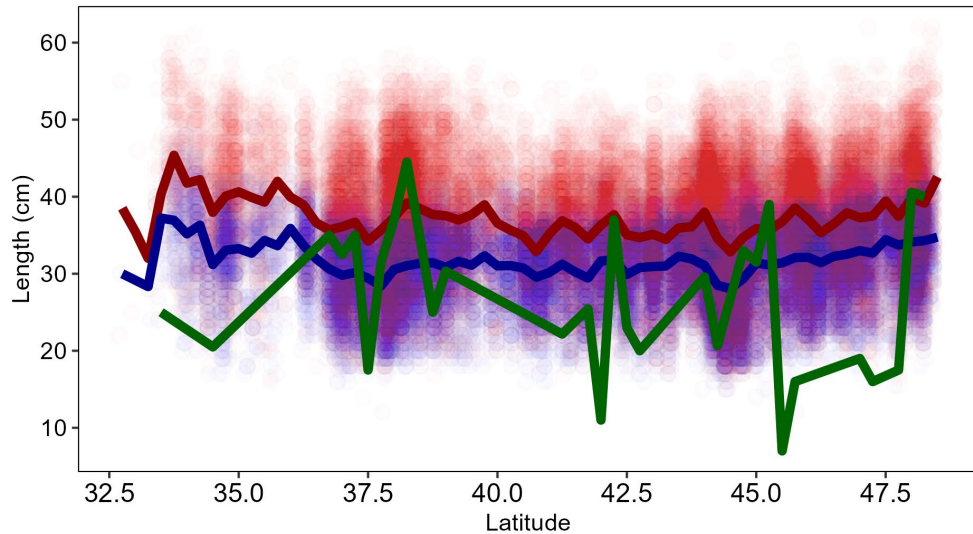
# Spatial strata

- Needed to calculate design-based indices

- Also used in length composition data expansion.

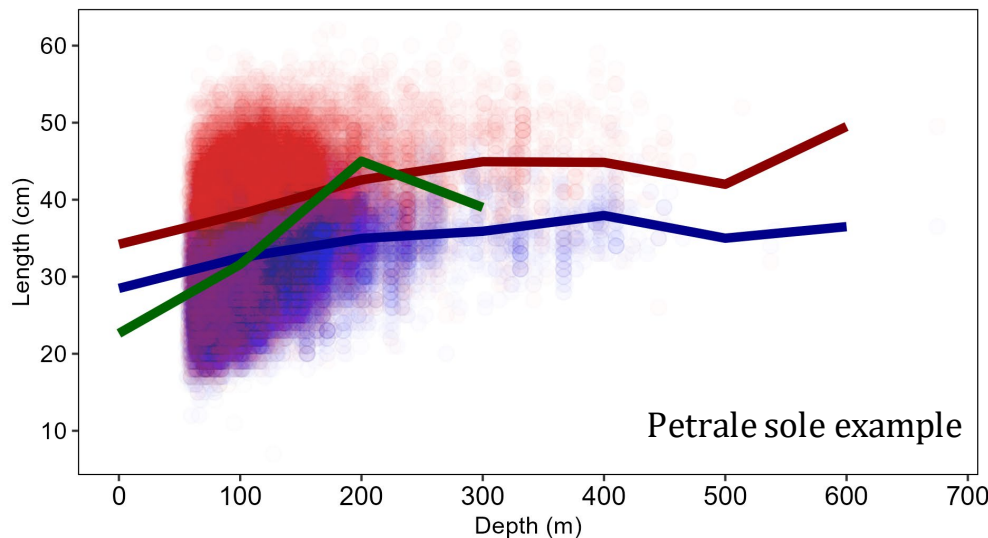- Defined based on species depth range and spatial distribution.



Petrale sole example
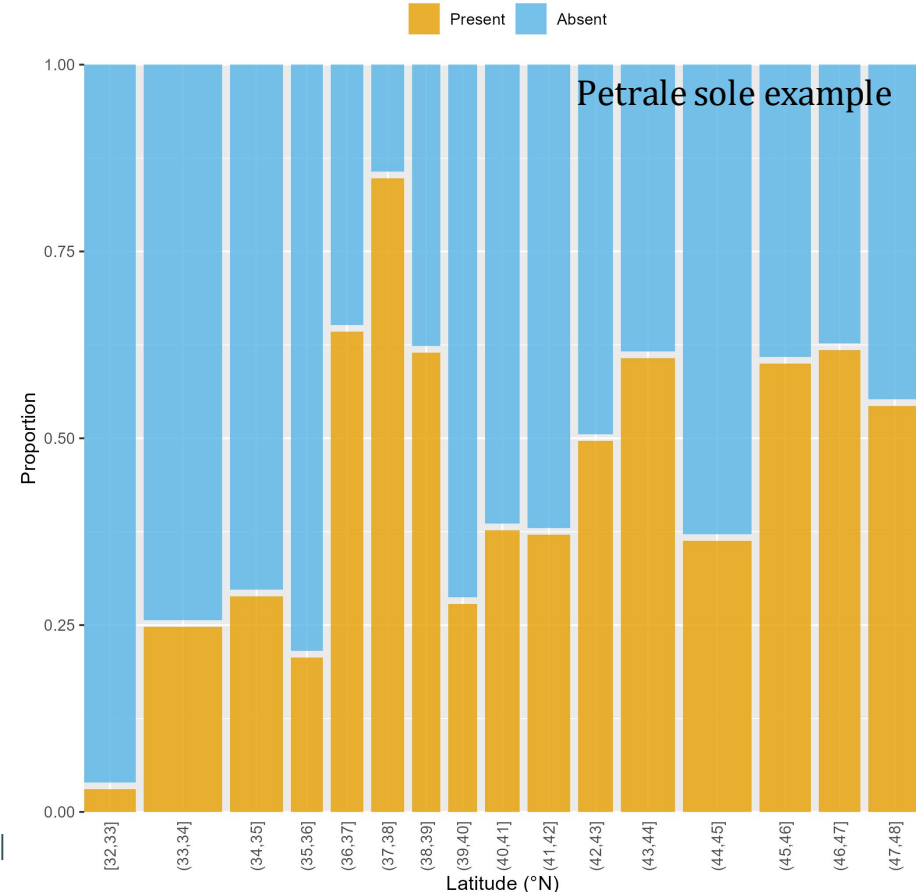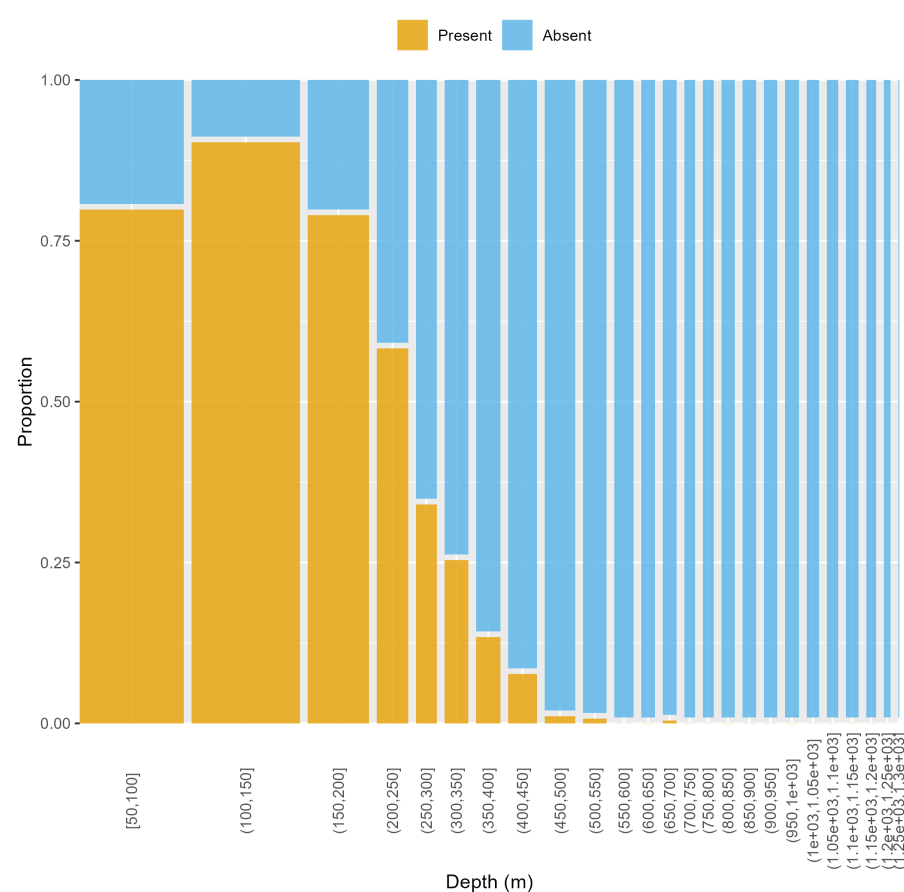
# Spatial strata



Petrale sole example

- Latitudinal bins are often alight with state borders,

- Depth bins more species specific.

- Need to locate strata used in the last full assessment.

# Age composition data

- From systematic sampling of survey catch
- Set # ages for each species randomly sampled each haul
- Age samples are usually is a set subset of fish sampled for length
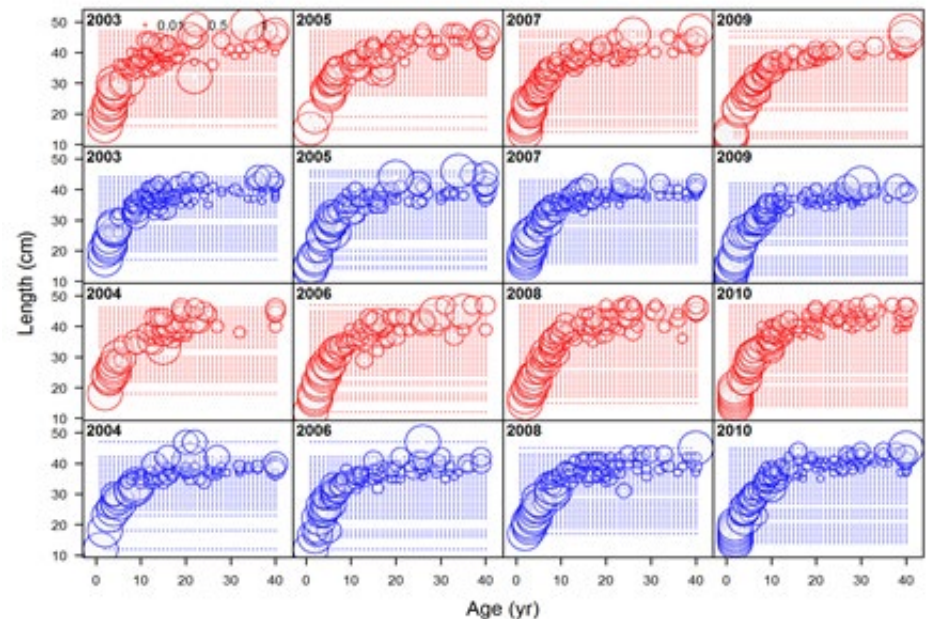- Otoliths are used for ageing of rockfish

NOAA FISHERIES

# Age composition data

- Multiple ways to enter age data – marginal length compositions and conditional ages-at-length compositions (CAAL).

- Survey age data are inputted as CAAL in most ss3 models.

- Facilitates the estimation of growth within the model.

- Goes around the issue of double counting the same fish (since age samples are usually a subset of length samples).

- But, model estimation slows considerably using CAAL data.



Age comp data, retained, Fishery
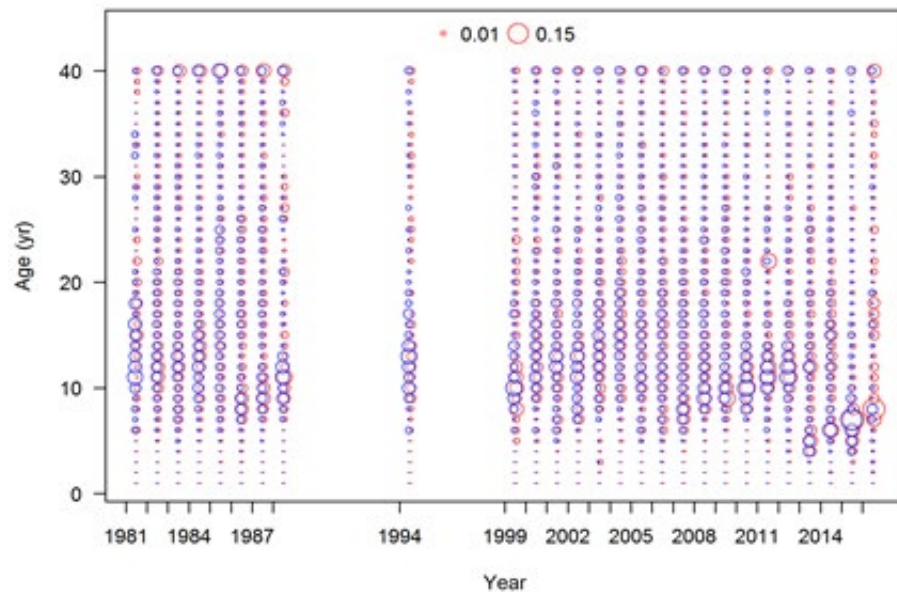


Conditional age-at-length data, whole catch, NWFSC shelf-slope survey

# Lets switch to nwfscSurvey

- [https://github.com/pfmc-assessments/nwfscSurvey](https://github.com/pfmc-assessments/nwfscSurvey)

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service

# nwfscSurvey

```
length_comps <- get_expanded_comps(
  bio_data = bio,
  catch_data = catch,
  comp_bins = seq(10, 80, 2),
  strata = strata,
  dir = getwd(),
  comp_column_name = "length_cm",
  output = "full_expansion_ss3_format",
  two_sex_comps = TRUE,
  month = "7",
  fleet = "7",
  input_n_method = "stewart_hamel")
```

Stewart & Hamel (2014),
Input N is a function of both
#fish and #haul

NOAA
FISHERIES

# Lets look at nwfscSurvey outputs

U.S. Department of Commerce | National Oceanic and Atmospheric Administration | National Marine Fisheries Service

# Exercise

- Extract the catch and bio data,
- Develop length compositions for WCGBTS,
- Create plots to evaluate data by depth and latitude.
- Evaluate current data bins. Do they make sense?
- Repeat the same steps for triennial survey.