Prepare Data for Age-Length Key

Derek H. Ogle, Northland College

5-Mar-2015

Preliminaries

```
> # clears objects in R workspace
> rm(list = ls())


> # load needed packages
> library(fishWiDNR)   # for setDBClasses()
> library(FSA)         # for lencat(), filterD()
> library(dplyr)       # for select(), mutate(), arrange(), %>%
> library(lubridate)   # for month()
```

Loading Data and Initial Manipulations

```
> # Load and prepare the data
> setwd("C:/aaaWork/Web/fishR/Courses/WiDNR_Statewide_2015/Day1_IntroR_FMData")
> d <- read.csv("SAWYER_fish_raw_data_012915.csv",stringsAsFactors=FALSE,na.strings=c("-","NA","")) %>%
    setDBClasses(type="RDNR") %>%
    select(County,Waterbody.Name,Survey.Year,Sample.Date,Gear,Fish.Data.Seq.No,Species,
           Length.or.Lower.Length.IN,Gender,Age..observed.annuli.,Edge.Counted.Desc,Age.Structure) %>%
    mutate(mon=month(Sample.Date,label=TRUE)) %>%
    mutate(lcat=lencat(Length.or.Lower.Length.IN,w=0.5)) %>%
    arrange(Species,Length.or.Lower.Length.IN)
>
> wae <- filterD(d,Waterbody.Name=="NELSON LAKE",Survey.Year==2014,mon=="May",Species=="WALLEYE")
> wae.aged <- filterD(wae,!is.na(Age..observed.annuli.))
```

Critical Thinking Interlude

Answer the following questions from your understanding of the code above:

1. Are there any questions about the R code (i.e., which functions have you not seen previously)?
2. How many data.frames were created?
3. How many variables are in each data.frame?
4. Describe the individuals that are in each data.frame?

Some Exploration

```
> xtabs(~Gender+lcat,data=wae)
      lcat
Gender  6  7 7.5  8 8.5  9 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17.5 18 18.5
     F  0  0   0  0   0  0  0    0  0    0  0    1  1    1  2    4  6    6  6    6 10    3 12    9
     M  0  0   0  0   0  0  0    0  0    6  4    6  4   11 23   21 26   32 20   13 11    7  4    3
     U  1  2   3  1   3  2  2    1  1    3  3    1  5    5  2    4  0    0  0    0  4    1  1    1
      lcat
Gender 19 19.5 20 20.5 21 21.5 22 22.5 23 23.5 24 24.5 25 25.5 26 26.5 27 27.5 28
     F  8    6  5    1  1    6  7    3  3   11 10    6  4    8  7    3  2    1  3
     M  0    3  1    3  2    2  0    2  1    1  0    0  0    0  0    0  0    0  0
     U  0    0  0    0  1    0  0    0  0    0  0    0  0    0  0    0  0    0  0


> xtabs(~Gender+lcat,data=wae.aged)
      lcat
Gender 10 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17 17.5 18 18.5 19 19.5 20 20.5
     F  0  0    0  0    1  1    1  1    1  1    5  5    5  4    0  5    4  1    2  2    0
     M  0  0    2  2    2  3    2  5    5  5    4  5    5  2    3  4    3  0    1  0    1
     U  2  1    1  0    0  0    1  0    0  0    0  0    0  0    0  0    0  0    0


> clrs <- c("black","gray40","gray70")
> plot(Length.or.Lower.Length.IN~jitter(Age..observed.annuli.),data=wae.aged,pch=16,
         col=clrs[Gender],xlab="Age (yrs)",ylab="Total Length (in)")
> legend("bottomright",levels(wae$Gender),col=clrs,pch=16,cex=0.75,bty="n")
```
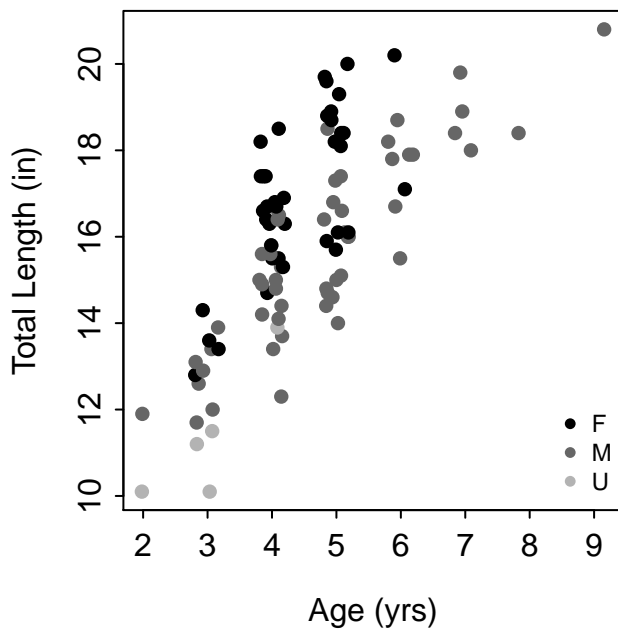
Critical Thinking Interlude

Answer the following questions from your understanding of the code above:

1. Are there any questions about the R code?
2. What decisions were made when sampling these fish for age?
3. How do those decisions affect how we can use these data?
4. What problems do you forsee with applying an age-length key developed from the aged fish in this sample to those fish that were not aged in the sample?