Filter Data

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

```
> library(fishWiDNR)   # for setDBClasses()
> library(dplyr)       # for select(), filter()
> library(FSA)         # for Summarize(), expandCounts()
```

```
> setwd("C:/aaaWork/Web/fishR/Courses/WiDNR_Statewide_2015/Day1_IntroR_FMData")
> d <- read.csv("FMDB_Sawyer_MultiYr_APEX.csv",stringsAsFactors=FALSE,na.strings=c("-","NA",""))
> d <- setDBClasses(d,type="RDNR")
> d <- expandCounts(d,~Number.of.Fish,~Length.or.Lower.Length.IN+Length.Upper.IN,new.name="Len")
> names(d)
```

```
 [1] "County"                    "Waterbody.Name"            "WBIC"
 [4] "Survey.Year"               "Station.Name"              "Swims.Station.Id"
 [7] "Site.Seq.No"               "Survey.Seq.No"             "Survey.Begin.Date"
[10] "Survey.End.Date"           "Survey.Status"             "Data.Entry.Name"
[13] "Entry.Date"                "Visit.Fish.Seq.No"         "Visit.Type"
[16] "Gear"                      "Sample.Date"               "Substation.Name"
[19] "Target.Species"            "Fish.Data.Seq.No"          "Net.Number"
[22] "Species.Code"              "Species"                   "Length.or.Lower.Length.IN"
[25] "Length.Upper.IN"           "Length.or.Lower.Length.MM" "Length.Upper.MM"
[28] "Weight.Pounds"             "Weight.Grams"              "Gender"
[31] "Disease"                   "Injury.Type"               "Age..observed.annuli."
[34] "Edge.Counted.Desc"         "Age.Structure"             "Mark.Given"
[37] "Mark.Found"                "Second.Mark.Found"         "Tag.Number.Given"
[40] "Second.Tag.Number.Given"   "Tag.Number.Found"          "Second.Tag.Number.Found"
[43] "YOY"                       "Entry.Date.1"              "Last.Update.Date"
[46] "Data.Ent.Name"             "Last.Update.Name"          "Invalid.Species"
[49] "Non.Standard.Bin"          "Length.Unit.Error"         "Length.Outside.Range"
[52] "Count.Outside.Range"       "Status.Code"               "Len"
[55] "lennote"
```

Selecting Variables – select()

```
> d1 <- select(d,Waterbody.Name,Gear,Survey.Year,Species,Len,Weight.Pounds,Gender,Mark.Given)
> headtail(d1)

        Waterbody.Name         Gear Survey.Year         Species  Len Weight.Pounds Gender
1      GRINDSTONE LAKE     FYKE NET        2003      CREEK CHUB   NA            NA   <NA>
2      GRINDSTONE LAKE     FYKE NET        2003         WALLEYE   NA            NA   <NA>
3      GRINDSTONE LAKE     FYKE NET        2003   NORTHERN PIKE   NA            NA   <NA>
448038   LAKE CHIPPEWA BOOM SHOCKER        2002         WALLEYE 13.0            NA   <NA>
448039     ISLAND LAKE BOOM SHOCKER        2007 LARGEMOUTH BASS  7.3            NA   <NA>
448043  BLAISDELL LAKE BOOM SHOCKER        2006     MUSKELLUNGE 13.7            NA   <NA>
       Mark.Given
1            <NA>
2            <NA>
3            <NA>
448038       <NA>
448039       <NA>
448043       <NA>

> tmp <- select(d,County:Swims.Station.Id)
> headtail(tmp)

       County  Waterbody.Name    WBIC Survey.Year                         Station.Name
1      SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION
2      SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION
3      SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION
448038 SAWYER   LAKE CHIPPEWA 2399700        2002   LAKE CHIPPEWA_GENERAL LAKE STATION
448039 SAWYER     ISLAND LAKE 2381800        2007     ISLAND LAKE_GENERAL LAKE STATION
448043 SAWYER  BLAISDELL LAKE 2402200        2006  BLAISDELL LAKE_GENERAL LAKE STATION
       Swims.Station.Id
1              10005586
2              10005586
3              10005586
448038         10005605
448039         10005570
448043         10005611

> tmp <- select(d,-(Station.Name:Status.Code))
> headtail(tmp)

       County  Waterbody.Name    WBIC Survey.Year  Len         lennote
1      SAWYER GRINDSTONE LAKE 2391200        2003   NA Observed length
2      SAWYER GRINDSTONE LAKE 2391200        2003   NA Observed length
3      SAWYER GRINDSTONE LAKE 2391200        2003   NA Observed length
448038 SAWYER   LAKE CHIPPEWA 2399700        2002 13.0 Expanded length
448039 SAWYER     ISLAND LAKE 2381800        2007  7.3 Expanded length
448043 SAWYER  BLAISDELL LAKE 2402200        2006 13.7 Expanded length

> tmp <- select(d,starts_with("Length"))                        # there is also an ends_with
> names(tmp)

[1] "Length.or.Lower.Length.IN" "Length.Upper.IN"           "Length.or.Lower.Length.MM"
[4] "Length.Upper.MM"           "Length.Unit.Error"         "Length.Outside.Range"
```

```
> tmp <- select(d,Survey.Seq.No,Species,Len,contains("Mark"))
> headtail(tmp)
```

```
       Survey.Seq.No         Species  Len Mark.Given Mark.Found Second.Mark.Found
1              51723      CREEK CHUB   NA       <NA>       <NA>              <NA>
2              51726         WALLEYE   NA       <NA>       <NA>              <NA>
3              51726   NORTHERN PIKE   NA       <NA>       <NA>              <NA>
448038         51356         WALLEYE 13.0       <NA>       <NA>              <NA>
448039         97739 LARGEMOUTH BASS  7.3       <NA>       <NA>              <NA>
448043         94228     MUSKELLUNGE 13.7       <NA>       <NA>              <NA>
```

Selecting Individuals – filter()

```
> levels(d1$Gear)
```

```
 [1] "BACKPACK SHOCKER"                       "BOOM SHOCKER"
 [3] "BOTTOM GILL NET"                        "DIP NET"
 [5] "FLOATING GILL NET"                      "FYKE NET"
 [7] "HOOK AND LINE"                          "LONG LINE SHOCKER"
 [9] "MINI BOOM SHOCKER"                      "MINI FYKE NET"
[11] "MINI FYKE NET WITH TURTLE EXCLUSION"    "MINI FYKE NET WITHOUT TURTLE EXCLUSION"
[13] "SEINE"                                  "STREAM SHOCKER"
```

```
> xtabs(~Gear,data=d1)
```

```
Gear
                      BACKPACK SHOCKER                           BOOM SHOCKER
                                  9467                                 131432
                       BOTTOM GILL NET                                DIP NET
                                   342                                    189
                     FLOATING GILL NET                               FYKE NET
                                  2883                                 193217
                         HOOK AND LINE                      LONG LINE SHOCKER
                                  1688                                     72
                     MINI BOOM SHOCKER                          MINI FYKE NET
                                  4479                                  15525
   MINI FYKE NET WITH TURTLE EXCLUSION MINI FYKE NET WITHOUT TURTLE EXCLUSION
                                 13873                                  24856
                                 SEINE                         STREAM SHOCKER
                                  2458                                  47565
```

```
> xtabs(~Waterbody.Name+Gear,data=d1)                           # only partial results shown
```

```
                   Gear
Waterbody.Name      BACKPACK SHOCKER BOOM SHOCKER BOTTOM GILL NET DIP NET
  ALDER CREEK                    182            0               0       0
  ASHEGON LAKE                     0           58               0       0
  BADGER CREEK                   105            0               0       0
  BARBER CREEK                    90            0               0       0
  BARBER LAKE                      0          979               0       0
  BARKER LAKE                      0          381              25       0
  BEAVER CREEK                     0            0               0       0
  BENSON CREEK                    74            0               0       0
  BILLY BOY FLOWAGE                0           92               0       0
  BLACK DAN LAKE                   0         1732               0       0
  BLACK LAKE                       0          213               0       0
  BLAISDELL LAKE                   0          404                41       0
  BLUEBERRY CREEK                 52            0                 0       0
  BLUEBERRY LAKE                   0          979                 0       0
  BRUNET RIVER                   133            0                 0       0

> tmp <- filter(d1,Waterbody.Name=="BARBER LAKE")
> xtabs(~Waterbody.Name,data=tmp)                               # only partial results shown

Waterbody.Name
      ALDER CREEK      ASHEGON LAKE      BADGER CREEK      BARBER CREEK       BARBER LAKE
                0                 0                 0                 0              3727
      BARKER LAKE      BEAVER CREEK      BENSON CREEK BILLY BOY FLOWAGE    BLACK DAN LAKE
                0                 0                 0                 0                 0
       BLACK LAKE    BLAISDELL LAKE   BLUEBERRY CREEK    BLUEBERRY LAKE      BRUNET RIVER
                0                 0                 0                 0                 0
    CALLAHAN LAKE
                0

> tmp <- droplevels(tmp)
> xtabs(~Waterbody.Name,data=tmp)

Waterbody.Name
BARBER LAKE
       3727

> tmp <- filter(d1,Waterbody.Name %in% c("BARBER LAKE","LAKE CHETAC"))
> tmp <- droplevels(tmp)
> xtabs(~Waterbody.Name,data=tmp)

Waterbody.Name
BARBER LAKE LAKE CHETAC
       3727       14827

> LCblg <- filter(d1,Waterbody.Name=="LAKE CHETAC",Species=="BLUEGILL")
> LCblg <- droplevels(LCblg)
> xtabs(~Gear,data=LCblg)

Gear
 BOOM SHOCKER      FYKE NET MINI FYKE NET
         1005           191           327

> LCblg <- filter(LCblg,Gear=="BOOM SHOCKER")
> Summarize(~Len,data=LCblg,digits=2)

       n     mean       sd      min       Q1   median       Q3      max percZero
 1005.00     6.16     1.08     2.60     5.50     6.20     7.00     9.80     0.00

> LCblgPREF <- filter(LCblg,Len>=7)
> Summarize(~Len,data=LCblgPREF,digits=2)

       n     mean       sd      min       Q1   median       Q3      max percZero
  259.00     7.45     0.43     7.00     7.20     7.30     7.65     9.80     0.00

> sturgWts <- filter(d1,Species=="LAKE STURGEON",!is.na(Weight.Pounds))
> headtail(sturgWts)
```

```
    Waterbody.Name            Gear Survey.Year       Species  Len Weight.Pounds Gender Mark.Given
1   CHIPPEWA RIVER         DIP NET        2006 LAKE STURGEON 54.3          32.0      M        PIT
2   CHIPPEWA RIVER         DIP NET        2006 LAKE STURGEON 59.7          47.0      F        PIT
3   CHIPPEWA RIVER         DIP NET        2006 LAKE STURGEON 54.8          37.0      M        PIT
415    BARKER LAKE BOTTOM GILL NET        2012 LAKE STURGEON 58.3          34.2   <NA>        PIT
416    BARKER LAKE BOTTOM GILL NET        2012 LAKE STURGEON 60.9          50.6   <NA>        PIT
417    BARKER LAKE BOTTOM GILL NET        2012 LAKE STURGEON 60.9          50.6   <NA>        PIT
```

Application Assignment

Create a script that performs the following tasks:

1. Load and prepare (set classes, expand counts, examine structure) your FM data in R (**HINT:** *use all or some of your script from the first application assignment*). Call this the *original data.frame*.
2. Create a data.frame that removes all variables related to the database (e.g., when datum was entered, who entered it, error flags, etc.).
3. Examine the sample size per water body and gear combination in the original data.frame.
4. Isolate (from the original data.frame) a water body of your choice and show the number of each species captured (in all gears).
5. Isolate (from the original data.frame) three water bodies of your choice and make one table that shows the number of each species captured in each water body (regardless of gear).
6. Isolate (from the original data.frame) one species of fish from one gear used in one waterbody.
   - Construct a table of frequency of each sex.
   - Summarize the length variable.
7. (*Time Permitting*) Suppose the waterbody and species you chose above has a minimum length limit (make up the minimum length). Isolate those fish that would be legal. Show that your filtering was successful.
8. (*Time Permitting*) Repeat the previous question but for a protected slot.
9. (*Time Permitting*) Repeat the previous question but for a harvest slot.
10. (*Time Permitting*) List all water bodies and species for which a weight in pounds was recorded (begin with the original data.frame).

Save your script!