Load Data from CSV

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

```
> library(fishWiDNR)   # for setDBClasses()
> library(FSA)         # for expandCounts()
```

Loading Data and Initial Examination

```
> 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")
> str(d)
```

```
'data.frame':   108872 obs. of  54 variables:
 $ County                   : Factor w/ 1 level "SAWYER": 1 1 1 1 1 1 1 1 1 1 ...
 $ Waterbody.Name           : Factor w/ 132 levels "ALDER CREEK",..: 39 39 39 39 39 39 39 39 39 39 ...
 $ WBIC                     : int  2391200 2391200 2391200 2391200 2391200 2391200 2391200 2391200 2391200 239
 $ Survey.Year              : int  2003 2003 2003 2003 2003 2003 2003 2003 2003 2003 ...
 $ Station.Name             : chr  "GRINDSTONE LAKE_GENERAL LAKE STATION" "GRINDSTONE LAKE_GENERAL LAKE STATIO
 $ Swims.Station.Id         : int  10005586 10005586 10005586 10005586 10005586 10005586 10005586 10005586 100
 $ Site.Seq.No              : int  110432 110432 110432 110432 110432 110432 110432 110432 110432 110432 ...
 $ Survey.Seq.No            : int  51723 51723 51723 51726 51726 51726 51737 51737 51737 51737 ...
 $ Survey.Begin.Date        : POSIXct, format: "2003-04-23" "2003-04-23" "2003-04-23" ...
 $ Survey.End.Date          : POSIXct, format: "2003-04-23" "2003-04-23" "2003-04-23" ...
 $ Survey.Status            : Factor w/ 3 levels "DATA ENTRY COMPLETE",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Data.Entry.Name          : chr  "warwir" "warwir" "warwir" "warwir" ...
 $ Entry.Date               : POSIXct, format: "2003-06-30" "2003-06-30" "2003-06-30" ...
 $ Visit.Fish.Seq.No        : int  485573 485573 485573 485616 485616 485616 485595 485595 485595 485595 ...
 $ Visit.Type               : Factor w/ 3 levels "ELECTROFISHING",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Gear                     : Factor w/ 14 levels "BACKPACK SHOCKER",..: 6 6 6 6 6 6 6 6 6 6 ...
 $ Sample.Date              : POSIXct, format: "2003-04-23" "2003-04-23" "2003-04-23" ...
 $ Substation.Name          : chr  NA NA NA NA ...
 $ Target.Species           : Factor w/ 18 levels "ALL SPECIES",..: 18 18 18 18 18 18 18 18 18 18 ...
 $ Fish.Data.Seq.No         : int  2104922 2104921 2104920 2106938 2106940 2106939 2106189 2106188 2106187 210
 $ Net.Number               : chr  "4" "4" "4" "H4" ...
 $ Species.Code             : Factor w/ 101 levels "A01","A01J","A02",..: 51 93 23 97 97 23 77 79 93 23 ...
 $ Species                  : Factor w/ 101 levels "ALL SPECIES",..: 32 101 64 95 95 64 69 10 101 64 ...
 $ Number.of.Fish           : int  1 12 19 4 1 1 1 2 60 2 ...
 $ Length.or.Lower.Length.IN: num  NA NA NA NA NA NA NA NA NA NA ...
 $ Length.Upper.IN          : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Length.or.Lower.Length.MM: num  NA NA NA NA NA NA NA NA NA NA ...
 $ Length.Upper.MM          : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Weight.Pounds            : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Weight.Grams             : num  NA NA NA NA NA NA NA NA NA NA ...
 $ Gender                   : Factor w/ 4 levels "F","M","NS","U": NA NA NA NA NA NA NA NA NA NA ...
 $ Disease                  : Factor w/ 0 levels: NA NA NA NA NA NA NA NA NA NA ...
 $ Injury.Type              : Factor w/ 1 level "DEAD": NA NA NA NA NA NA NA NA NA NA ...
 $ Age..observed.annuli.    : logi  NA NA NA NA NA NA ...
 $ Edge.Counted.Desc        : Factor w/ 0 levels: NA NA NA NA NA NA NA NA NA NA ...
 $ Age.Structure            : Factor w/ 2 levels "SCALE","SPINE": NA NA NA NA NA NA NA NA NA NA ...
 $ Mark.Given               : Factor w/ 9 levels "AN","BC","LP",..: NA NA NA NA NA NA NA NA NA NA ...
 $ Mark.Found               : Factor w/ 11 levels "AN","BC","LP",..: NA NA NA NA NA NA NA NA NA NA ...
```

```
 $ Second.Mark.Found        : Factor w/ 3 levels "LP","PIT","RV": NA NA NA NA NA NA NA NA NA NA ...
 $ Tag.Number.Given         : chr  NA NA NA NA ...
 $ Second.Tag.Number.Given  : chr  NA NA NA NA ...
 $ Tag.Number.Found         : chr  NA NA NA NA ...
 $ Second.Tag.Number.Found  : chr  NA NA NA NA ...
 $ YOY                      : Factor w/ 2 levels "N","Y": NA NA NA NA NA NA NA NA NA NA ...
 $ Entry.Date.1             : POSIXct, format: "2003-06-30" "2003-06-30" "2003-06-30" ...
 $ Last.Update.Date         : POSIXct, format: "2006-05-02" "2006-05-02" "2006-05-02" ...
 $ Data.Ent.Name            : chr  "prattf" "prattf" "prattf" "prattf" ...
 $ Last.Update.Name         : chr  "toberj" "toberj" "toberj" "toberj" ...
 $ Invalid.Species          : chr  NA NA NA NA ...
 $ Non.Standard.Bin         : chr  NA NA NA NA ...
 $ Length.Unit.Error        : chr  NA NA NA NA ...
 $ Length.Outside.Range     : chr  NA NA NA NA ...
 $ Count.Outside.Range      : chr  NA NA NA NA ...
 $ Status.Code              : chr  NA NA NA NA ...

> head(d)   # also can use tail(d) or headtail(d)

  County  Waterbody.Name    WBIC Survey.Year                          Station.Name Swims.Station.Id
1 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
2 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
3 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
4 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
5 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
6 SAWYER GRINDSTONE LAKE 2391200        2003 GRINDSTONE LAKE_GENERAL LAKE STATION         10005586
  Site.Seq.No Survey.Seq.No Survey.Begin.Date Survey.End.Date                  Survey.Status
1      110432         51723        2003-04-23      2003-04-23 DATA ENTRY COMPLETE AND PROOFED
2      110432         51723        2003-04-23      2003-04-23 DATA ENTRY COMPLETE AND PROOFED
3      110432         51723        2003-04-23      2003-04-23 DATA ENTRY COMPLETE AND PROOFED
4      110432         51726        2003-04-24      2003-04-24 DATA ENTRY COMPLETE AND PROOFED
5      110432         51726        2003-04-24      2003-04-24 DATA ENTRY COMPLETE AND PROOFED
6      110432         51726        2003-04-24      2003-04-24 DATA ENTRY COMPLETE AND PROOFED
  Data.Entry.Name Entry.Date Visit.Fish.Seq.No Visit.Type     Gear Sample.Date Substation.Name
1          warwir 2003-06-30            485573    NETTING FYKE NET  2003-04-23            <NA>
2          warwir 2003-06-30            485573    NETTING FYKE NET  2003-04-23            <NA>
3          warwir 2003-06-30            485573    NETTING FYKE NET  2003-04-23            <NA>
4          warwir 2003-06-30            485616    NETTING FYKE NET  2003-04-24            <NA>
5          warwir 2003-06-30            485616    NETTING FYKE NET  2003-04-24            <NA>
6          warwir 2003-06-30            485616    NETTING FYKE NET  2003-04-24            <NA>
  Target.Species Fish.Data.Seq.No Net.Number Species.Code       Species Number.of.Fish
1        WALLEYE          2104922          4          M50    CREEK CHUB             1
2        WALLEYE          2104921          4          X15  YELLOW PERCH            12
3        WALLEYE          2104920          4          L02 NORTHERN PIKE            19
4        WALLEYE          2106938         H4          X22       WALLEYE             4
5        WALLEYE          2106940         H7          X22       WALLEYE             1
6        WALLEYE          2106939         H7          L02 NORTHERN PIKE             1
  Length.or.Lower.Length.IN Length.Upper.IN Length.or.Lower.Length.MM Length.Upper.MM Weight.Pounds
1                        NA              NA                         NA              NA            NA
2                        NA              NA                         NA              NA            NA
3                        NA              NA                         NA              NA            NA
4                        NA              NA                         NA              NA            NA
5                        NA              NA                         NA              NA            NA
6                        NA              NA                         NA              NA            NA
  Weight.Grams Gender Disease Injury.Type Age..observed.annuli. Edge.Counted.Desc Age.Structure
1           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
2           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
3           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
4           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
5           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
```

```
6           NA   <NA>    <NA>        <NA>                    NA              <NA>          <NA>
  Mark.Given Mark.Found Second.Mark.Found Tag.Number.Given Second.Tag.Number.Given Tag.Number.Found
1       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
2       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
3       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
4       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
5       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
6       <NA>       <NA>              <NA>             <NA>                    <NA>             <NA>
  Second.Tag.Number.Found  YOY Entry.Date.1 Last.Update.Date Data.Ent.Name Last.Update.Name
1                    <NA> <NA>   2003-06-30       2006-05-02        prattf           toberj
2                    <NA> <NA>   2003-06-30       2006-05-02        prattf           toberj
3                    <NA> <NA>   2003-06-30       2006-05-02        prattf           toberj
4                    <NA> <NA>   2003-07-02       2006-05-02        prattf           toberj
5                    <NA> <NA>   2003-07-02       2006-05-02        prattf           toberj
6                    <NA> <NA>   2003-07-02       2006-05-02        prattf           toberj
  Invalid.Species Non.Standard.Bin Length.Unit.Error Length.Outside.Range Count.Outside.Range
1            <NA>             <NA>              <NA>                 <NA>                <NA>
2            <NA>             <NA>              <NA>                 <NA>                <NA>
3            <NA>             <NA>              <NA>                 <NA>                <NA>
4            <NA>             <NA>              <NA>                 <NA>                <NA>
5            <NA>             <NA>              <NA>                 <NA>                <NA>
6            <NA>             <NA>              <NA>                 <NA>                <NA>
  Status.Code
1        <NA>
2        <NA>
3        <NA>
4        <NA>
5        <NA>
6        <NA>

> nrow(d)

[1] 108872
```

Expanding Counts of Fish to Individual Measurements

Example portion of the Sawyer Co. FM database with "tallied" fish lengths.

Fish.Data.Seq.No	Species	Number.of.Fish	Length.or.Lower.Length.IN	Length.Upper.IN
8837567	WALLEYE	1	12.5	12.9
8837042	WALLEYE	1	15.5	15.9
8837041	WALLEYE	2	15.0	15.4
8837040	WALLEYE	1	14.5	14.9
8837637	WALLEYE	1	19.5	19.9
8837636	WALLEYE	2	15.0	15.4
8837635	WALLEYE	1	13.0	13.4

Example expansion but keeping the length bins.

Fish.Data.Seq.No	Species	Length.or.Lower.Length.IN	Length.Upper.IN
8837567	WALLEYE	12.5	12.9
8837042	WALLEYE	15.5	15.9
8837040	WALLEYE	14.5	14.9
8837637	WALLEYE	19.5	19.9
8837635	WALLEYE	13.0	13.4
8837041	WALLEYE	15.0	15.4
8837041	WALLEYE	15.0	15.4
8837636	WALLEYE	15.0	15.4
8837636	WALLEYE	15.0	15.4

Example expansion but adding random digit from within the length bin.

Fish.Data.Seq.No	Species	Length.or.Lower.Length.IN	Length.Upper.IN	Len	lennote
8837567	WALLEYE	12.5	12.9	12.8	Expanded length
8837042	WALLEYE	15.5	15.9	15.7	Expanded length
8837040	WALLEYE	14.5	14.9	14.8	Expanded length
8837637	WALLEYE	19.5	19.9	19.5	Expanded length
8837635	WALLEYE	13.0	13.4	13.4	Expanded length
8837041	WALLEYE	15.0	15.4	15.0	Expanded length
8837041	WALLEYE	15.0	15.4	15.2	Expanded length
8837636	WALLEYE	15.0	15.4	15.4	Expanded length
8837636	WALLEYE	15.0	15.4	15.0	Expanded length

```
> # without random digits
> d1 <- expandCounts(d,~Number.of.Fish)

33 rows had zero counts in Number.of.Fish.
73782 rows had an individual measurement.
35027 rows with multiple measurements were expanded to 374231 rows of individual measurements.

> # with random digits
> d1 <- expandCounts(d,~Number.of.Fish,~Length.or.Lower.Length.IN+Length.Upper.IN,new.name="Len")

33 rows had zero counts in Number.of.Fish.
73782 rows had an individual measurement.
35027 rows with multiple measurements were expanded to 374231 rows of individual measurements.

> nrow(d1)

[1] 448046

> # sum of Number.of.Fish variable (note from above the number of rows that had zero fish)
> sum(d$Number.of.Fish,na.rm=TRUE)

[1] 448013
```

Individual Variables

```
> d1$Length.or.Lower.Length.IN
```

```
  [1]   NA   NA   NA   NA   NA   NA   NA   NA  2.0  2.6  2.4  2.3   NA 14.5  2.9  4.5  2.5  2.2   NA
 [20]   NA   NA   NA 17.0 11.4  4.8  6.6  7.3  4.6  7.0  5.9  7.1  6.2  8.5  8.0  7.0  6.6  6.0  4.9
 [39]  4.8 10.0  9.7  9.3  9.0  8.8  8.2  7.6  7.2  6.7  6.6  5.3  5.1  5.9  5.5  5.4  7.2  8.2  7.5
 [58]  6.9  6.6  6.5  6.4  6.3  5.5  5.3  5.0  4.9  4.4  7.3  7.1  6.8 13.9 13.2 13.1 11.2 11.1  8.2
 [77]  7.9  4.0  5.0  3.2  5.8  3.7  3.5  3.0  3.6  4.2  5.5  2.2  2.2  5.2  4.0  3.9  2.5  3.5  3.4
 [96]  3.0  4.0  4.0  4.8  6.0  6.0  6.0  7.3  4.3  2.1  5.2  3.6  5.5  5.3  3.0  3.3  3.2  2.7  5.0
[115]  4.3  4.2  3.9  3.8  3.4  2.8  2.7  2.4  6.0  4.4  4.0  3.5  3.3  3.2  3.0  2.8  2.7  8.3  6.0
[134]  5.6  5.4  5.2  5.0  4.6  3.8  2.9  2.8  3.5  2.2  4.2  5.0  4.3  2.6  2.2  5.0  3.3  3.2  3.0
[153] 10.0  3.2  4.8  3.6  5.5  6.2  4.3  6.0  4.0  3.8  3.7  3.4  3.1  5.5  5.3  5.2  5.0  4.5  6.0
[172]  5.5  5.3  4.6  4.4  4.3  4.1  3.7  2.7  2.6  6.0  5.3  5.0  4.0  2.6  2.3  7.4  5.0  5.7  5.8
[191]  5.2  4.0  3.5  9.0  8.7  8.5  7.7  7.4  6.4  6.2  6.0  5.6  6.7  6.0  3.8  3.5  6.3  4.8  4.7
[210]  3.8  3.6  3.3  2.7  2.5 10.8  4.7  6.8  3.2  2.8  8.4  7.8  7.7  7.6  7.4  7.3  6.7  6.4  5.7
[229]  5.6  5.8  5.7  5.6  5.3  2.5  2.2  2.0  5.0  4.8  4.2  4.0  3.7  5.5  5.0  3.0  2.3  1.8  4.4
[248]  2.3  2.2  7.2  6.5  6.0  4.5  4.0  3.8  3.7  3.5  3.3  3.5  1.6  2.3  2.2 18.2  4.0  4.3  6.4
[267]  3.7  4.4  4.0 37.1  4.1 35.0 36.2 15.1  5.2  7.1  3.8  5.0  4.2  3.6  4.7  4.0  4.0  3.9  4.5
[286]  4.2  3.7  4.0  3.6  7.4  6.5  6.8  6.1 27.1  3.7  3.7  4.1  5.6  5.6  6.8 27.6  7.9  6.5  6.5
```

```
> d1$Species
```

```
  [1] CREEK CHUB       WALLEYE          NORTHERN PIKE    PUMPKINSEED      ROCK BASS
  [6] GOLDEN SHINER    BLACK CRAPPIE    MUSKELLUNGE      LARGEMOUTH BASS  BLUEGILL
 [11] BLUEGILL         BLUEGILL         LARGEMOUTH BASS  WALLEYE          BLACK BULLHEAD
 [16] BLACK CRAPPIE    BLACK CRAPPIE    YELLOW PERCH     NORTHERN PIKE    BLACK CRAPPIE
 [21] BLUEGILL         LARGEMOUTH BASS  BROOK TROUT     BROOK TROUT      PUMPKINSEED
 [26] YELLOW PERCH     YELLOW PERCH     PUMPKINSEED      BLACK BULLHEAD   YELLOW PERCH
 [31] BLACK CRAPPIE    ROCK BASS        BLACK CRAPPIE    YELLOW PERCH     YELLOW PERCH
 [36] YELLOW PERCH     YELLOW PERCH     YELLOW PERCH     YELLOW PERCH     BLACK CRAPPIE
 [41] BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE
 [46] BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE    BLACK CRAPPIE
 [51] BLACK CRAPPIE    GOLDEN SHINER    GOLDEN SHINER    GOLDEN SHINER    LARGEMOUTH BASS
 [56] YELLOW PERCH     ROCK BASS        ROCK BASS        ROCK BASS        ROCK BASS
 [61] ROCK BASS        ROCK BASS        ROCK BASS        ROCK BASS        ROCK BASS
 [66] ROCK BASS        ROCK BASS        PUMPKINSEED      PUMPKINSEED      PUMPKINSEED
 [71] LARGEMOUTH BASS  LARGEMOUTH BASS  LARGEMOUTH BASS  LARGEMOUTH BASS  LARGEMOUTH BASS
 [76] LARGEMOUTH BASS  LARGEMOUTH BASS  ROCK BASS        PUMPKINSEED      PUMPKINSEED
 [81] BLUEGILL         YELLOW PERCH     YELLOW PERCH     YELLOW PERCH     YELLOW PERCH
 [86] YELLOW PERCH     YELLOW PERCH     ROCK BASS        ROCK BASS        ROCK BASS
 [91] PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED
 [96] PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED
[101] BLUEGILL         BLUEGILL         BLUEGILL         BLUEGILL         ROCK BASS
[106] PUMPKINSEED      PUMPKINSEED      YELLOW PERCH     YELLOW PERCH     BLUNTNOSE MINNOW
[111] ROCK BASS        ROCK BASS        ROCK BASS        PUMPKINSEED      PUMPKINSEED
[116] PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED      PUMPKINSEED
101 Levels: ALL SPECIES AMERICAN BROOK LAMPREY ... YELLOW PERCH
```

Application Assignment

Create a script that performs the following tasks:

1. Load your FM data into R.
2. Set the classes on your data.frame.
3. How many rows are in your data.frame?
4. How many total fish are in your data.frame?
5. Expand the counts to individual fish (assign to a new data.frame).
6. How many variables (columns) and individual fish (rows) are in your new data.frame?
7. Display all expanded length measurements.
8. Show all "values" for two other variables of your choice.

Save your script!