Load Data from CSV

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

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

Loading Data and Initial Examination

```
> 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",""))
> d <- setDBClasses(d,type="RDNR")
> str(d)
'data.frame':   48683 obs. of  53 variables:
 $ County                    : Factor w/ 1 level "SAWYER": 1 1 1 1 1 1 1 1 1 1 ...
 $ Waterbody.Name            : Factor w/ 86 levels "ALDER CREEK",..: 64 64 64 64 64 64 64 64 64 64 ...
 $ WBIC                      : int  2393500 2393500 2393500 2393500 2393500 2393500 2393500 2393500 2393500 239
 $ Survey.Year               : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
 $ Station.Name              : chr  "SISSABAGAMA LAKE_GENERAL LAKE STATION" "SISSABAGAMA LAKE_GENERAL LAKE STAT
 $ Swims.Station.Id          : int  10005590 10005590 10005590 10005590 10005590 10005590 10005590 10005590 100
 $ Site.Seq.No               : int  113071 113071 113071 113071 113071 113071 113071 113071 113071 113071 ...
 $ Survey.Seq.No             : int  39508941 39508941 39508941 39508941 39508941 39508941 39508941 39508941 395
 $ Survey.Begin.Date         : POSIXct, format: "2010-03-31" "2010-03-31" "2010-03-31" ...
 $ Survey.End.Date           : POSIXct, format: "2010-04-13" "2010-04-13" "2010-04-13" ...
 $ Survey.Status             : Factor w/ 3 levels "DATA ENTRY COMPLETE",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Data.Entry.Name           : chr  "spooner_treaty" "spooner_treaty" "spooner_treaty" "spooner_treaty" ...
 $ Visit.Fish.Seq.No         : int  624568 624568 624568 624568 624568 624568 624568 624568 624568 624568 ...
 $ Visit.Type                : Factor w/ 2 levels "ELECTROFISHING",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Gear                      : Factor w/ 7 levels "BACKPACK SHOCKER",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Sample.Date               : POSIXct, format: "2010-04-01" "2010-04-01" "2010-04-01" ...
 $ Substation.Name           : chr  NA NA NA NA ...
 $ Target.Species            : Factor w/ 14 levels "ALL SPECIES",..: 14 14 14 14 14 14 14 14 14 14 ...
 $ Fish.Data.Seq.No          : int  7529172 7529173 7529174 7529175 7529176 7529177 7529178 7529179 7529180 752
 $ Net.Number                : chr  "ALL DAILY NET" "ALL DAILY NET" "ALL DAILY NET" "ALL DAILY NET" ...
 $ Species.Code              : Factor w/ 73 levels "A01J","A02","A03",..: 70 70 70 70 70 70 70 70 70 70 ...
 $ Species                   : Factor w/ 73 levels "AMERICAN BROOK LAMPREY",..: 67 67 67 67 67 67 67 67 67 67 .
 $ Number.of.Fish            : int  1 2 1 5 11 18 27 26 23 15 ...
 $ Length.or.Lower.Length.IN: num  11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 ...
 $ Length.Upper.IN           : num  11.4 11.9 12.4 12.9 13.4 13.9 14.4 14.9 15.4 15.9 ...
 $ Length.or.Lower.Length.MM: num  279 292 305 318 330 ...
 $ Length.Upper.MM           : num  290 302 315 328 340 ...
 $ 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/ 3 levels "F","M","U": 2 2 2 2 2 2 2 2 2 2 ...
 $ 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.     : int  NA NA NA NA NA NA NA NA NA NA ...
 $ Edge.Counted.Desc         : Factor w/ 1 level "Yes": NA NA NA NA NA NA NA NA NA NA ...
 $ Age.Structure             : Factor w/ 3 levels "OTOLITH","SCALE",..: NA NA NA NA NA NA NA NA NA NA ...
 $ Mark.Given                : Factor w/ 7 levels "AN","LP","LV",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Mark.Found                : Factor w/ 9 levels "AN","BC","LP",..: NA NA NA NA NA NA NA NA NA NA ...
 $ Second.Mark.Found         : Factor w/ 1 level "PIT": 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               : POSIXct, format: "2010-04-06" "2010-04-06" "2010-04-06" ...
 $ Last.Update.Date         : POSIXct, format: NA NA NA ...
 $ Data.Ent.Name            : chr  "spooner_treaty" "spooner_treaty" "spooner_treaty" "spooner_treaty" ...
 $ Last.Update.Name         : chr  NA NA NA NA ...
 $ 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 ...
```

```
> headtail(d)   # also can use tail(d) or head(d)
      County   Waterbody.Name    WBIC Survey.Year                          Station.Name
1     SAWYER SISSABAGAMA LAKE 2393500        2010 SISSABAGAMA LAKE_GENERAL LAKE STATION
2     SAWYER SISSABAGAMA LAKE 2393500        2010 SISSABAGAMA LAKE_GENERAL LAKE STATION
3     SAWYER SISSABAGAMA LAKE 2393500        2010 SISSABAGAMA LAKE_GENERAL LAKE STATION
48681 SAWYER     WINDIGO LAKE 2046600        2014     WINDIGO LAKE_GENERAL LAKE STATION
48682 SAWYER     WINDIGO LAKE 2046600        2014     WINDIGO LAKE_GENERAL LAKE STATION
48683 SAWYER   BLAISDELL LAKE 2402200        2014   BLAISDELL LAKE_GENERAL LAKE STATION
      Swims.Station.Id Site.Seq.No Survey.Seq.No Survey.Begin.Date Survey.End.Date
1             10005590      113071      39508941        2010-03-31      2010-04-13
2             10005590      113071      39508941        2010-03-31      2010-04-13
3             10005590      113071      39508941        2010-03-31      2010-04-13
48681         10005544      121911     515077184        2014-10-16      2014-10-17
48682         10005544      121911     515077184        2014-10-16      2014-10-17
48683         10005611      109201     515076773        2014-10-21      2014-10-21
                        Survey.Status Data.Entry.Name Visit.Fish.Seq.No     Visit.Type
1     DATA ENTRY COMPLETE AND PROOFED  spooner_treaty            624568        NETTING
2     DATA ENTRY COMPLETE AND PROOFED  spooner_treaty            624568        NETTING
3     DATA ENTRY COMPLETE AND PROOFED  spooner_treaty            624568        NETTING
48681 DATA ENTRY COMPLETE AND PROOFED  spooner_treaty            723742 ELECTROFISHING
48682 DATA ENTRY COMPLETE AND PROOFED  spooner_treaty            723742 ELECTROFISHING
48683 DATA ENTRY COMPLETE AND PROOFED          scheij            722558        NETTING
                Gear Sample.Date Substation.Name Target.Species Fish.Data.Seq.No    Net.Number
1          FYKE NET  2010-04-01            <NA>        WALLEYE          7529172 ALL DAILY NET
2          FYKE NET  2010-04-01            <NA>        WALLEYE          7529173 ALL DAILY NET
3          FYKE NET  2010-04-01            <NA>        WALLEYE          7529174 ALL DAILY NET
48681   BOOM SHOCKER  2014-10-17          GLIFWC        WALLEYE         10711298          <NA>
48682   BOOM SHOCKER  2014-10-17          GLIFWC        WALLEYE         10711299          <NA>
48683 BOTTOM GILL NET  2014-10-21            <NA>  LAKE STURGEON         10633383      ALL NETS
      Species.Code          Species Number.of.Fish Length.or.Lower.Length.IN Length.Upper.IN
1              X22          WALLEYE              1                      11.0            11.4
2              X22          WALLEYE              2                      11.5            11.9
3              X22          WALLEYE              1                      12.0            12.4
48681          W11  SMALLMOUTH BASS              1                      10.5            10.9
48682          W12  LARGEMOUTH BASS              1                       7.0             7.4
48683          Z98 NO FISH CAPTURED              0                        NA              NA
      Length.or.Lower.Length.MM Length.Upper.MM Weight.Pounds Weight.Grams Gender Disease.
1                         279.4          289.56            NA           NA      M     <NA>
2                         292.1          302.26            NA           NA      M     <NA>
3                         304.8          314.96            NA           NA      M     <NA>
48681                     266.7          276.86            NA           NA   <NA>     <NA>
48682                     177.8          187.96            NA           NA   <NA>     <NA>
48683                        NA              NA            NA           NA   <NA>     <NA>
      Injury.Type Age..observed.annuli. Edge.Counted.Desc Age.Structure Mark.Given Mark.Found
1            <NA>                   NA              <NA>          <NA>         LP       <NA>
```

```
2           <NA>                    NA              <NA>          <NA>         LP       <NA>
3           <NA>                    NA              <NA>          <NA>         LP       <NA>
48681       <NA>                    NA              <NA>          <NA>       <NA>       <NA>
48682       <NA>                    NA              <NA>          <NA>       <NA>       <NA>
48683       <NA>                    NA              <NA>          <NA>       <NA>       <NA>
      Second.Mark.Found Tag.Number.Given Second.Tag.Number.Given Tag.Number.Found
1                  <NA>             <NA>                    <NA>             <NA>
2                  <NA>             <NA>                    <NA>             <NA>
3                  <NA>             <NA>                    <NA>             <NA>
48681              <NA>             <NA>                    <NA>             <NA>
48682              <NA>             <NA>                    <NA>             <NA>
48683              <NA>             <NA>                    <NA>             <NA>
      Second.Tag.Number.Found  YOY Entry.Date Last.Update.Date  Data.Ent.Name Last.Update.Name
1                        <NA> <NA> 2010-04-06             <NA> spooner_treaty             <NA>
2                        <NA> <NA> 2010-04-06             <NA> spooner_treaty             <NA>
3                        <NA> <NA> 2010-04-06             <NA> spooner_treaty             <NA>
48681                    <NA> <NA> 2015-01-21             <NA> spooner_treaty             <NA>
48682                    <NA> <NA> 2015-01-21             <NA> spooner_treaty             <NA>
48683                    <NA> <NA> 2014-12-10             <NA>         scheij             <NA>
      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>
48681            <NA>             <NA>              <NA>                 <NA>                <NA>
48682            <NA>             <NA>              <NA>                 <NA>                <NA>
48683            <NA>             <NA>              <NA>                 <NA>                <NA>
      Status.Code
1            <NA>
2            <NA>
3            <NA>
48681        <NA>
48682        <NA>
48683        SC/H

> nrow(d)
[1] 48683
```

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
8837020	WALLEYE	1	20.0	20.4
8837525	WALLEYE	1	18.0	18.4
8837562	WALLEYE	2	14.5	14.9
8837563	WALLEYE	1	15.5	15.9
8837635	WALLEYE	1	13.0	13.4
8837636	WALLEYE	2	15.0	15.4
8837637	WALLEYE	1	19.5	19.9

Example expansion but keeping the length bins.

Fish.Data.Seq.No	Species	Length.or.Lower.Length.IN	Length.Upper.IN
8837020	WALLEYE	20.0	20.4
8837525	WALLEYE	18.0	18.4
8837562	WALLEYE	14.5	14.9
8837562	WALLEYE	14.5	14.9
8837563	WALLEYE	15.5	15.9
8837635	WALLEYE	13.0	13.4
8837636	WALLEYE	15.0	15.4
8837636	WALLEYE	15.0	15.4
8837637	WALLEYE	19.5	19.9

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
8837020	WALLEYE	20.0	20.4	20.2	Expanded length
8837525	WALLEYE	18.0	18.4	18.2	Expanded length
8837562	WALLEYE	14.5	14.9	14.8	Expanded length
8837562	WALLEYE	14.5	14.9	14.6	Expanded length
8837563	WALLEYE	15.5	15.9	15.8	Expanded length
8837635	WALLEYE	13.0	13.4	13.2	Expanded length
8837636	WALLEYE	15.0	15.4	15.3	Expanded length
8837636	WALLEYE	15.0	15.4	15.0	Expanded length
8837637	WALLEYE	19.5	19.9	19.5	Expanded length

```
> # without random digits
> d1 <- expandCounts(d,~Number.of.Fish)
13 rows had zero counts in Number.of.Fish.
38448 rows had an individual measurement.
10222 rows with multiple measurements were expanded to 94790 rows of individual measurements.

> # with random digits
> d1 <- expandCounts(d,~Number.of.Fish,~Length.or.Lower.Length.IN+Length.Upper.IN,new.name="Len")
13 rows had zero counts in Number.of.Fish.
38448 rows had an individual measurement.
10222 rows with multiple measurements were expanded to 94790 rows of individual measurements.

> nrow(d1)
[1] 133251

> # 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] 133238
```

Isolate Individual Variables

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

```
  [1]  8.2  8.1  8.6  9.3  8.9  9.7  6.3  7.6  9.1  5.9   NA   NA   NA   NA   NA   NA   NA   NA   NA
 [20]  7.9  8.0  8.7  8.8  9.0  9.2  7.4  7.6  7.7  8.3  8.4  8.8  5.2  6.0  7.5  7.6 11.1  5.7  7.5
 [39]  8.0  8.5  8.8  9.0  9.3 19.0  8.5  8.7  7.3  8.9 10.1  7.4  6.9 11.4  6.7  5.9  6.2   NA   NA
 [58]   NA  5.3  5.7  6.3  6.4  6.6  6.9  7.2  8.6  8.7  9.2  7.9  8.1  8.2  8.3  8.5  8.6  6.0  7.0
 [77]  7.2  7.6  7.8  8.1  8.6  9.1  4.9  5.0  5.1  6.3  5.4  6.5  6.6  6.8  7.2  7.6  8.9  9.1  9.2
 [96]  9.5  9.6 10.0  4.9  6.2  5.3  5.6  6.9  7.7  8.0  8.1  8.2  8.3  8.4  8.5  8.7  9.1  5.6 17.5
[115] 21.0  9.3  9.5 10.5 11.0 11.2 11.5  9.5  7.7   NA  5.0  9.8 10.2 10.3 12.0 12.2 12.5  8.6  8.1
[134]  5.7  7.2  8.0  9.1 11.4 11.8 12.2   NA   NA   NA   NA   NA  4.1  4.4  4.6  5.0  5.3  6.2  8.4
[153]  9.2 11.8 12.2 12.3  9.0 11.7   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
[172]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
[191]  4.5  6.4  6.9  7.8  8.0  8.1  8.4  8.5  9.1  5.6  5.7  6.1  6.4  6.8  7.1  7.2  7.3  7.5  7.6
[210]  7.7  9.0  9.1  9.3  9.5  5.0  5.2  5.3  5.4  5.8  6.1  6.3  6.8  7.1  6.5  6.9  7.3  5.3  5.8
[229]  7.1  7.8  8.3  8.9  5.6  6.9  6.5  6.7  6.9  7.0  6.2  6.5  6.9  7.2  7.5  7.7  7.9  5.2  5.5
[248]  5.9  6.7  6.8  6.9  7.2  7.4  7.5  7.7  7.9  8.3  7.4  7.7  8.0  8.2  8.3  8.7  9.0  9.2  9.5
[267] 10.5 11.5 11.6  5.5  6.5  7.0  7.5  4.9  5.4  5.6  6.0  6.3  6.6  7.5  7.7  7.9  8.1  8.2  8.4
[286]  9.1  9.2 12.5  5.7  5.8  6.5  6.6  7.5  7.6  7.8  8.2  8.3  8.6  5.6  5.8  6.2  6.3  6.4  6.9
```

```
> d1$Species
```

```
  [1] YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH
  [6] YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH
 [11] WHITE SUCKER    PUMPKINSEED     GOLDEN SHINER   GOLDEN SHINER   ROCK BASS
 [16] PUMPKINSEED     YELLOW PERCH    YELLOW PERCH    GOLDEN SHINER   BLACK CRAPPIE
 [21] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [26] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [31] BLACK CRAPPIE   YELLOW PERCH    BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [36] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [41] BLACK CRAPPIE   BLACK CRAPPIE   YELLOW BULLHEAD WHITE SUCKER    YELLOW BULLHEAD
 [46] YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH
 [51] YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH    YELLOW PERCH
 [56] YELLOW PERCH    BLUEGILL        BLUEGILL        BLACK CRAPPIE   BLACK CRAPPIE
 [61] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [66] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [71] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [76] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [81] BLACK CRAPPIE   BLACK CRAPPIE   YELLOW PERCH    YELLOW PERCH    YELLOW PERCH
 [86] YELLOW PERCH    BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [91] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
 [96] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   YELLOW PERCH    YELLOW PERCH
[101] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
[106] BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE   BLACK CRAPPIE
[111] BLACK CRAPPIE   BLACK CRAPPIE   GOLDEN SHINER   WHITE SUCKER    WHITE SUCKER
[116] YELLOW BULLHEAD YELLOW BULLHEAD YELLOW BULLHEAD YELLOW BULLHEAD YELLOW BULLHEAD
73 Levels: AMERICAN BROOK LAMPREY AMERICAN BROOK LAMPREY (AMMOCOETE) ... YELLOW PERCH
```

Application Assignment

Create a script that performs the following tasks:

1. Load your FM data into a data.frame in R.
2. Set the classes on the data.frame.
3. How many rows are in the data.frame?
4. How many total fish are in the data.frame?

5. Expand the counts of fish to individual fish with random digits for length (assign to a new data.frame).
6. How many variables (columns) and individual fish (rows) are in the new data.frame?
7. Display all expanded length measurements.
8. Show all "values" for two other variables of your choice.

Save your script!