

# 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_FMDData")
> 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 2393500 ...
 $ 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 STATION" ...
 $ Swims.Station.Id: int   10005590 10005590 10005590 10005590 10005590 10005590 10005590 10005590 10005590 10005590 ...
 $ 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 39508941 39508941 ...
 $ 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 7529181 ...
 $ 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 ...

```

```
> head(d) # also can use tail(d) or headtail(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
4	SAWYER	SISSABAGAMA LAKE	2393500	2010	SISSABAGAMA LAKE_GENERAL	LAKE STATION
5	SAWYER	SISSABAGAMA LAKE	2393500	2010	SISSABAGAMA LAKE_GENERAL	LAKE STATION
6	SAWYER	SISSABAGAMA LAKE	2393500	2010	SISSABAGAMA 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
4	10005590	113071	39508941	2010-03-31	2010-04-13
5	10005590	113071	39508941	2010-03-31	2010-04-13
6	10005590	113071	39508941	2010-03-31	2010-04-13

  

	Survey.Status	Data.Entry.Name	Visit.Fish.Seq.No	Visit.Type	Gear	Sample.Date
1	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01
2	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01
3	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01
4	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01
5	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01
6	DATA ENTRY COMPLETE AND PROOFED	spooner_treaty	624568	NETTING	FYKE NET	2010-04-01

  

	Substation.Name	Target.Species	Fish.Data.Seq.No	Net.Number	Species.Code	Species	Number.of.Fish
1	<NA>	WALLEYE	7529172	ALL DAILY NET	X22	WALLEYE	1
2	<NA>	WALLEYE	7529173	ALL DAILY NET	X22	WALLEYE	2
3	<NA>	WALLEYE	7529174	ALL DAILY NET	X22	WALLEYE	1
4	<NA>	WALLEYE	7529175	ALL DAILY NET	X22	WALLEYE	5
5	<NA>	WALLEYE	7529176	ALL DAILY NET	X22	WALLEYE	11
6	<NA>	WALLEYE	7529177	ALL DAILY NET	X22	WALLEYE	18

  

	Length.or.Lower.Length.IN	Length.Upper.IN	Length.or.Lower.Length.MM	Length.Upper.MM	Weight.Pounds
1	11.0	11.4	279.4	289.56	NA
2	11.5	11.9	292.1	302.26	NA
3	12.0	12.4	304.8	314.96	NA
4	12.5	12.9	317.5	327.66	NA
5	13.0	13.4	330.2	340.36	NA
6	13.5	13.9	342.9	353.06	NA

  

	Weight.Grams	Gender	Disease.	Injury.Type	Age..observed.annuli.	Edge.Counted.Desc	Age.Structure
1	NA	M	<NA>	<NA>	NA	<NA>	<NA>
2	NA	M	<NA>	<NA>	NA	<NA>	<NA>
3	NA	M	<NA>	<NA>	NA	<NA>	<NA>
4	NA	M	<NA>	<NA>	NA	<NA>	<NA>
5	NA	M	<NA>	<NA>	NA	<NA>	<NA>
6	NA	M	<NA>	<NA>	NA	<NA>	<NA>

  

	Mark.Given	Mark.Found	Second.Mark.Found	Tag.Number.Given	Second.Tag.Number.Given	Tag.Number.Found
1	LP	<NA>	<NA>	<NA>	<NA>	<NA>

2	LP	<NA>	<NA>	<NA>	<NA>	<NA>
3	LP	<NA>	<NA>	<NA>	<NA>	<NA>
4	LP	<NA>	<NA>	<NA>	<NA>	<NA>
5	LP	<NA>	<NA>	<NA>	<NA>	<NA>
6	LP	<NA>	<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>
4	<NA>	<NA>	2010-04-06	<NA>	spooner_treaty	<NA>
5	<NA>	<NA>	2010-04-06	<NA>	spooner_treaty	<NA>
6	<NA>	<NA>	2010-04-06	<NA>	spooner_treaty	<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>
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] 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
8837042	WALLEYE	1	15.5	15.9
8837043	WALLEYE	1	16.0	16.4
8837044	WALLEYE	1	18.5	18.9
8837601	WALLEYE	1	15.5	15.9
8837602	WALLEYE	1	19.0	19.4
8837608	WALLEYE	1	12.5	12.9

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
8837042	WALLEYE	15.5	15.9
8837043	WALLEYE	16.0	16.4
8837044	WALLEYE	18.5	18.9
8837601	WALLEYE	15.5	15.9
8837602	WALLEYE	19.0	19.4
8837608	WALLEYE	12.5	12.9

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	20.4	Expanded length
8837042	WALLEYE	15.5	15.9	15.9	Expanded length
8837043	WALLEYE	16.0	16.4	16.4	Expanded length
8837044	WALLEYE	18.5	18.9	18.5	Expanded length
8837601	WALLEYE	15.5	15.9	15.8	Expanded length
8837602	WALLEYE	19.0	19.4	19.0	Expanded length
8837608	WALLEYE	12.5	12.9	12.8	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	BLUNTNOSSE 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!**