

Add and Rename Variables

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

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

> # load needed packages
> library(fishWiDNR) # for setDBClasses(), changeDBNames()
> library(dplyr)     # for filter(), select(), mutate(), rename()
> library(lubridate) # for month()
> library(FSA)       # for expandCounts(), capFirst(), filterD()

> # load FM data and expand lengths ... mostly copied code from first and second handouts
> 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")
> d <- expandCounts(d, ~Number.of.Fish, ~Length.or.Lower.Length.IN+Length.Upper.IN, new.name="Len")

> d1 <- filterD(d, Species=="LAKE STURGEON", Waterbody.Name=="BARKER LAKE", !is.na(Weight.Pounds))
> d1 <- select(d1, Species, Survey.Year, Survey.Begin.Date, Len, Weight.Pounds)
> headtail(d1, n=2)
      Species Survey.Year Survey.Begin.Date  Len Weight.Pounds
1 LAKE STURGEON      2010      2010-05-04  58.0          43.9
2 LAKE STURGEON      2010      2010-05-04  61.5          70.5
24 LAKE STURGEON      2012      2012-08-02  58.3          34.2
25 LAKE STURGEON      2012      2012-08-02  58.3          34.2
```

Variable Additions

Simple Mutations

```
> tmp <- mutate(d1, loglen=log(Len), logwt=log(Weight.Pounds))
> headtail(tmp, n=2)
      Species Survey.Year Survey.Begin.Date  Len Weight.Pounds  loglen  logwt
1 LAKE STURGEON      2010      2010-05-04  58.0          43.9  4.060443  3.781914
2 LAKE STURGEON      2010      2010-05-04  61.5          70.5  4.119037  4.255613
24 LAKE STURGEON      2012      2012-08-02  58.3          34.2  4.065602  3.532226
25 LAKE STURGEON      2012      2012-08-02  58.3          34.2  4.065602  3.532226
```

Simple Special Purpose Mutations

```
> tmp <- mutate(d1, mon1=month(Survey.Begin.Date),
                mon2=month(Survey.Begin.Date, label=TRUE))
> headtail(tmp, n=2)
      Species Survey.Year Survey.Begin.Date  Len Weight.Pounds mon1 mon2
1 LAKE STURGEON      2010      2010-05-04  58.0          43.9    5  May
2 LAKE STURGEON      2010      2010-05-04  61.5          70.5    5  May
24 LAKE STURGEON      2012      2012-08-02  58.3          34.2    8  Aug
25 LAKE STURGEON      2012      2012-08-02  58.3          34.2    8  Aug
```

```
> tmp <- mutate(d1,Species1=capFirst(Species),
                Species2=capFirst(Species,which="first"))
> headtail(tmp,n=2)
```

	Species	Survey.Year	Survey.Begin.Date	Len	Weight.Pounds	Species1	Species2
1	LAKE STURGEON	2010	2010-05-04	58.0	43.9	Lake Sturgeon	Lake sturgeon
2	LAKE STURGEON	2010	2010-05-04	61.5	70.5	Lake Sturgeon	Lake sturgeon
24	LAKE STURGEON	2012	2012-08-02	58.3	34.2	Lake Sturgeon	Lake sturgeon
25	LAKE STURGEON	2012	2012-08-02	58.3	34.2	Lake Sturgeon	Lake sturgeon

Length Category Mutations

```
> tmp <- mutate(d1,lcat2=lencat(Len,w=2),
                lcat2a=lencat(Len,w=2,as.fact=TRUE),
                lcatA=lencat(Len,breaks=c(46,54,56,58,70)),
                lcatB=lencat(Len,breaks=c(small=0,medium=50,large=60,very_large=70),use.names=TRUE) )
> headtail(tmp)
```

	Species	Survey.Year	Survey.Begin.Date	Len	Weight.Pounds	lcat2	lcat2a	lcatA	lcatB
1	LAKE STURGEON	2010	2010-05-04	58.0	43.9	58	58	58	medium
2	LAKE STURGEON	2010	2010-05-04	61.5	70.5	60	60	58	large
3	LAKE STURGEON	2010	2010-05-04	59.7	55.6	58	58	58	medium
23	LAKE STURGEON	2012	2012-08-02	60.9	50.6	60	60	58	large
24	LAKE STURGEON	2012	2012-08-02	58.3	34.2	58	58	58	medium
25	LAKE STURGEON	2012	2012-08-02	58.3	34.2	58	58	58	medium

```
> xtabs(~lcat2,data=tmp)
lcat2
46 54 56 58 60 62 66
1 6 1 6 7 2 2

> xtabs(~lcat2a,data=tmp)
lcat2a
46 48 50 52 54 56 58 60 62 64 66
1 0 0 0 6 1 6 7 2 0 2

> xtabs(~lcatA,data=tmp)
lcatA
46 54 56 58
1 6 1 17

> xtabs(~lcatB,data=tmp)
lcatB
      small      medium      large very_large
      1         13         11         0
```

Rename Variables

```
> tmp <- rename(d1,year=Survey.Year,wt=Weight.Pounds)
> headtail(tmp)
      Species year Survey.Begin.Date  Len  wt
1 LAKE STURGEON 2010      2010-05-04 58.0 43.9
2 LAKE STURGEON 2010      2010-05-04 61.5 70.5
3 LAKE STURGEON 2010      2010-05-04 59.7 55.6
23 LAKE STURGEON 2012      2012-08-02 60.9 50.6
24 LAKE STURGEON 2012      2012-08-02 58.3 34.2
25 LAKE STURGEON 2012      2012-08-02 58.3 34.2

> tmp <- changeDBNames(d1)
> names(tmp)
[1] "srvy_begin" "species"      "year"          "Len"          "wt_lbs"

> tmp <- changeDBNames(tmp,from="R",to="RDNR")
> names(tmp)
[1] "Species"          "Survey.Year"      "Survey.Begin.Date" "Len"
[5] "Weight.Pounds"

> tmp <- changeDBNames(tmp,from="RDNR",to="DNR")
> names(tmp)
[1] "Survey Begin Date" "Species"          "Survey Year"      "Len"
[5] "Weight Pounds"

> tmp$"Weight Pounds"
 [1] 43.90 70.50 55.60 66.50 38.80 45.70 44.60 41.90 60.00 48.00 48.70 24.25 67.24 62.83 44.31 38.47
[17] 56.88 41.01 51.59 53.79 50.70 50.60 50.60 34.20 34.20

> write.csv(tmp,"LKS_Barker.csv",row.names=FALSE)
```

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 scripts from previous application assignments*).
2. Rename two or more variables to names that better fit your usage (or change all names according to the definitions in `changeDBNames()`).
3. Create a new variable that has the species names with only the first letters capitalized.
4. Create a new variable that has the water body names with only the first letters capitalized.
5. Create a new variable that is the length in mm computed from the length in inches (even though this is already in the FM database).
6. Isolate a game species from a waterbody (and possibly a gear) of interest to you.
7. Create a new variable that contains evenly-spaced length categories that are appropriate for your species. Construct a frequency table of that variable.
8. Create a new variable that contains length categories that could be defined as “not of interest”, “marginally interesting”, “preferred”, and “very interesting” to anglers for your species. Construct a frequency table of that variable.
9. (*Time Permitting*) Create a new variable that contains the Gabelhouse length categories (“stock”, “quality”, etc.) for your species (**HINT:** *use, for example, `psdVal("Largemouth Bass",units="in")` to find Gablehouse lengths for a particular species*).

Save your script!