Add and Rename Variables

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

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

> 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")
> d <- expandCounts(d,~Number.of.Fish,~Length.or.Lower.Length.IN+Length.Upper.IN,new.name="Len")
> d1 <- filter(d,Species=="LAKE STURGEON",Waterbody.Name=="BARKER LAKE",!is.na(Weight.Pounds))
> d1 <- select(d1,Species,Survey.Year,Survey.Begin.Date,Len,Weight.Pounds)
> head(d1,n=4)
        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
3 LAKE STURGEON        2010        2010-05-04 59.7          55.6
4 LAKE STURGEON        2010        2010-05-04 62.5          66.5
```

Variable Additions

Simple Mutations

```
> tmp <- mutate(d1,loglen=log(Len),logwt=log(Weight.Pounds))
> head(tmp,n=4)
        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
3 LAKE STURGEON        2010        2010-05-04 59.7          55.6 4.089332 4.018183
4 LAKE STURGEON        2010        2010-05-04 62.5          66.5 4.135167 4.197202
```

Simple Special Purpose Mutations

```
> tmp <- mutate(d1,mon1=month(Survey.Begin.Date),
                   mon2=month(Survey.Begin.Date,label=TRUE))
> head(tmp,n=4)
        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
3 LAKE STURGEON        2010        2010-05-04 59.7          55.6    5  May
4 LAKE STURGEON        2010        2010-05-04 62.5          66.5    5  May

> tmp <- mutate(d1,Species1=capFirst(Species),
                   Species2=capFirst(Species,which="first"))
> head(tmp,n=4)
        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
3 LAKE STURGEON        2010        2010-05-04 59.7          55.6 Lake Sturgeon Lake sturgeon
4 LAKE STURGEON        2010        2010-05-04 62.5          66.5 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) )
> head(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
4 LAKE STURGEON        2010        2010-05-04 62.5          66.5    62     62    58  large
5 LAKE STURGEON        2010        2010-05-04 55.7          38.8    54     54    54 medium
6 LAKE STURGEON        2010        2010-05-04 56.4          45.7    56     56    56 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)
> head(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
4 LAKE STURGEON 2010        2010-05-04 62.5 66.5
5 LAKE STURGEON 2010        2010-05-04 55.7 38.8
6 LAKE STURGEON 2010        2010-05-04 56.4 45.7
```

```
> 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")
> 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!