

Add and Rename Variables

Derek H. Ogle, Northland College

2-Mar-2015

Preliminaries

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

> setwd("C:/aaaWork/Web/fishR/Courses/WIStatewide2015/CourseMaterial")
> d <- read.csv("FMDB_Sawyer.csv", stringsAsFactors=FALSE)
> 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))
> d2 <- select(d1, Species, Survey.Year, Survey.Begin.Date, Len, Weight.Pounds)
> head(d2)
```

	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
5	LAKE STURGEON	2010	2010-05-04	55.7	38.8
6	LAKE STURGEON	2010	2010-05-04	56.4	45.7

Variable Additions

Simple Mutations

```
> tmp <- mutate(d2, loglen=log(Len), logwt=log(Weight.Pounds))
> head(tmp)
```

	Species	Survey.Year	Survey.Begin.Date	Len	Weight.Pounds	loglen	logwt
1	LAKE STURGEON	2010	2010-05-04	58.0	43.9	4.060	3.782
2	LAKE STURGEON	2010	2010-05-04	61.5	70.5	4.119	4.256
3	LAKE STURGEON	2010	2010-05-04	59.7	55.6	4.089	4.018
4	LAKE STURGEON	2010	2010-05-04	62.5	66.5	4.135	4.197
5	LAKE STURGEON	2010	2010-05-04	55.7	38.8	4.020	3.658
6	LAKE STURGEON	2010	2010-05-04	56.4	45.7	4.032	3.822

Simple Special Purpose Mutations

```
> tmp <- mutate(d2, mon1=month(Survey.Begin.Date),
                mon2=month(Survey.Begin.Date, label=TRUE))
> head(tmp)
```

	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
5	LAKE STURGEON	2010	2010-05-04	55.7	38.8	5	May
6	LAKE STURGEON	2010	2010-05-04	56.4	45.7	5	May

```
> tmp <- mutate(d2,Species1=capFirst(Species),
                Species2=capFirst(Species,which="first"))
> head(tmp)
  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
5 LAKE STURGEON      2010      2010-05-04 55.7          38.8 Lake Sturgeon Lake sturgeon
6 LAKE STURGEON      2010      2010-05-04 56.4          45.7 Lake Sturgeon Lake sturgeon
```

Length Category Mutations

```
> tmp <- mutate(d2,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  3  4  2  2

> xtabs(~lcat2a,data=tmp)
lcat2a
46 48 50 52 54 56 58 60 62 64 66
 1  0  0  0  6  1  3  4  2  0  2

> xtabs(~lcatA,data=tmp)
lcatA
46 54 56 58
 1  6  1 11

> xtabs(~lcatB,data=tmp)
lcatB
      small      medium      large very_large
      1         10         8         0
```

Rename Variables

```
> tmp <- rename(d2,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] "county"          "waterbody"      "wbic"           "year"           "station"
 [6] "swimsID"         "site_seq"       "srvy_seq"       "srvy_begin"     "srvy_end"
[11] "srvy_status"     "dentry_name"    "dentry_date"    "vst_fish_seq"   "vst_type"
[16] "gear"            "date"           "substation"     "target"         "fish_data_seq"
[21] "net"             "species_code"   "species"        "len_in"         "len_up_in"
[26] "len_mm"          "len_up_mm"      "wt_lbs"         "wt_g"           "sex"
[31] "disease"         "injury"         "age"            "edge"           "age_strux"
[36] "mark_given"      "mark_found"     "mark2_found"    "tag_given"      "tag2_given"
[41] "tag_found"       "tag2_found"     "yoy"            "dentry_date2"   "dupdate_date"
[46] "dentry_name2"    "dupdate_name"   "inv_species"    "inv_bin"        "inv_length_unit"
[51] "inv_length"      "inv_count"      "status_code"    "Len"            "lennote"

> tmp <- changeDBNames(tmp,from="R",to="RDNR")
> names(tmp)
 [1] "County"          "Waterbody.Name" "WBIC"           "Survey.Year"    "Swims.Station.Id"
 [4] "Site.Seq.No"     "Srvy.Seq.No"    "Survey.Begin.Date"
 [7] "Survey.End.Date" "Survey.Status"  "Data.Entry.Name"
[10] "Entry.Date"      "Visit.Fish.Seq.No" "Visit.Type"
[13] "Gear"            "Sample.Date"    "Substation.Name"
[16] "Target.Species"  "Fish.Data.Seq.No" "Net.Number"
[19] "Species.Code"    "Species"        "Length.or.Lower.Length.IN"
[22] "Length.Upper.IN" "Length.or.Lower.Length.MM" "Length.Upper.MM"
[25] "Weight.Pounds"  "Weight.Grams"   "Gender"
[28] "Disease"         "Injury.Type"    "Age..observed.annuli."
[31] "Edge.Counted.Desc" "Age.Structure"  "Mark.Given"
[34] "Mark.Found"      "Second.Mark.Found" "Tag.Number.Given"
[37] "Second.Tag.Number.Given" "Tag.Number.Found" "Second.Tag.Number.Found"
[40] "YOY"            "Entry.Date.1"   "Last.Update.Date"
[43] "Data.Ent.Name"   "Last.Update.Name" "Invalid.Species"
[46] "Non.Standard.Bin" "Length.Unit.Error" "Length.Outside.Range"
[49] "Count.Outside.Range" "Status.Code"    "Len"
[52] "lennote"

> tmp <- changeDBNames(tmp,from="RDNR",to="DNR")
> write.csv(tmp,"LKS_Barker14.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 that have names that annoy you (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")`*).

Save your script!