Exercise – Data Frames

Answer the following questions with R code by creating (*and editing if you make a mistake*) an R script and iteratively running the code in RStudio.

1. Load the data in the **RuffeBio.xlsx** file into a data frame in R.

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/Exercises")
> ruf <- read.csv("Data/RuffeBio.csv")
```

2. How many variables are in this data frame? How many individuals/observations?

```
> str(ruf)
'data.frame': 40 obs. of  10 variables:
 $ fishID  : int  60 61 62 63 64 65 66 67 68 69 ...
 $ locShort: Factor w/ 1 level "St. Louis R. (2007)": 1 1 1 1 1 1 1 1 1 ...
 $ year    : int  2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
 $ month   : int  9 9 9 9 9 9 9 9 9 9 ...
 $ day     : int  20 20 20 20 20 20 20 20 20 20 ...
 $ date    : Factor w/ 1 level "9/20/2007": 1 1 1 1 1 1 1 1 1 ...
 $ tl      : int  134 111 110 115 92 88 95 90 99 107 ...
 $ wt      : num  24.6 14.7 12.3 16 8.3 7.8 9.7 8.2 11.7 13 ...
 $ sex     : Factor w/ 3 levels "female","male",..: 1 1 1 1 1 1 1 1 1 ...
 $ maturity: Factor w/ 3 levels "","immature",..: 3 3 2 3 3 3 3 3 3 ...
```

   There are 10 variables and 40 individuals/observations in this data frame.

3. Specifically, what is the name of the first variable
   The name of the first variable is fishID.

4. Show all variables for the fifth individual.

```
> ruf[5,]
  fishID            locShort year month day      date tl  wt    sex maturity
5     64 St. Louis R. (2007) 2007     9  20 9/20/2007 92 8.3 female   mature
```

5. Show all variables for the fifth and seventh individuals.

```
> ruf[c(5,7),]
  fishID            locShort year month day      date tl  wt    sex maturity
5     64 St. Louis R. (2007) 2007     9  20 9/20/2007 92 8.3 female   mature
7     66 St. Louis R. (2007) 2007     9  20 9/20/2007 95 9.7 female   mature
```

6. Show the total lengths for all individuals.

```
> ruf$tl
 [1] 134 111 110 115  92  88  95  90  99 107  NA  99 102 105  90 102 114  NA  56  90
[21] 101 109 110 111 101  95  84 105 120 104 102  99  84  87  81  81  65  42  NA 115
```

7. Show ONLY the total length for the seventeenth individual.

```
> ruf$tl[17]
[1] 114
```

8. Show ONLY the total length for the fifth and seventeenth individuals.

```
> ruf$tl[c(15,17)]
[1]  90 114
```

9. For each situation below, create a new data frame (from the original) and record how many fish are in that data frame.

   (a) Just female ruffe.

```
> ruf1 <- Subset(ruf,sex=="female")
> nrow(ruf1)
[1] 31
```

   (b) Just ruffe greater than 110 mm.

```
> ruf2 <- Subset(ruf,tl>110)
> nrow(ruf2)
[1] 7
```

   (c) Just ruffe between 80 and 110 mm.

```
> ruf3 <- Subset(ruf,tl>80 & tl<110)
> nrow(ruf3)
[1] 25
```

   (d) Excluding all fish of an "unknown" sex.

```
> ruf5 <- Subset(ruf,sex!="unknown")
> nrow(ruf5)
[1] 39
```

10. Create new variables in the original data frame for the following situations.

   (a) Natural log of length and weight.

```
> ruf$logtl <- log(ruf$tl)
> ruf$logw <- log(ruf$w)
> view(ruf)
   fishID            locShort year month day      date  tl   wt    sex maturity
1      60 St. Louis R. (2007) 2007     9  20 9/20/2007 134 24.6 female   mature
3      62 St. Louis R. (2007) 2007     9  20 9/20/2007 110 12.3 female immature
12     71 St. Louis R. (2007) 2007     9  20 9/20/2007  99  8.4 female   mature
13     72 St. Louis R. (2007) 2007     9  20 9/20/2007 102 11.4 female   mature
20     79 St. Louis R. (2007) 2007     9  20 9/20/2007  90  7.6 female   mature
26     85 St. Louis R. (2007) 2007     9  20 9/20/2007  95 10.4 female   mature
   logtl  logw
1  4.898 3.203
3  4.700 2.510
12 4.595 2.128
13 4.625 2.434
20 4.500 2.028
26 4.554 2.342
```

   (b) Length categories that are 10-mm wide.

```
> Summarize(~tl,data=ruf)
       n     mean       sd      min       Q1   median       Q3      max percZero
   37.00    97.16    17.52    42.00    90.00   101.00   109.00   134.00     0.00
> ruf <- lencat(~tl,data=ruf,startcat=40,w=10)
> view(ruf)
   fishID            locShort year month day      date  tl   wt    sex maturity
4      63 St. Louis R. (2007) 2007     9  20 9/20/2007 115 16.0 female   mature
6      65 St. Louis R. (2007) 2007     9  20 9/20/2007  88  7.8 female   mature
7      66 St. Louis R. (2007) 2007     9  20 9/20/2007  95  9.7 female   mature
12     71 St. Louis R. (2007) 2007     9  20 9/20/2007  99  8.4 female   mature
```

```
28     87 St. Louis R. (2007) 2007     9  20 9/20/2007 105 11.7   male   mature
32     91 St. Louis R. (2007) 2007     9  20 9/20/2007  99  9.1   male   mature
   logtl  logw LCat
4  4.745 2.773  110
6  4.477 2.054   80
7  4.554 2.272   90
12 4.595 2.128   90
28 4.654 2.460  100
32 4.595 2.208   90
```

(c) Fulton's condition factor (The weight of the fish divided by the cubed length of the fish multiplied by 10000).

```
> ruf$fult <- ruf$wt/(ruf$tl^3)*10000
> view(ruf)
   fishID            locShort year month day      date  tl   wt    sex maturity
1      60 St. Louis R. (2007) 2007     9  20 9/20/2007 134 24.6 female   mature
3      62 St. Louis R. (2007) 2007     9  20 9/20/2007 110 12.3 female immature
11     70 St. Louis R. (2007) 2007     9  20 9/20/2007  NA  9.7 female   mature
13     72 St. Louis R. (2007) 2007     9  20 9/20/2007 102 11.4 female   mature
32     91 St. Louis R. (2007) 2007     9  20 9/20/2007  99  9.1   male   mature
39     98 St. Louis R. (2007) 2007     9  20 9/20/2007  NA 10.5   male   mature
   logtl  logw LCat    fult
1  4.898 3.203  130 0.10224
3  4.700 2.510  110 0.09241
11    NA 2.272 <NA>      NA
13 4.625 2.434  100 0.10742
32 4.595 2.208   90 0.09379
39    NA 2.351 <NA>      NA
```

11. If you have time ...

(a) Show the length frequency table by sex.

```
> table(ruf$sex,ruf$LCat)

         40 50 60 70 80 90 100 110 120 130
  female  0  1  1  0  2  8   9   7   1   1
  male    0  0  0  0  4  1   1   0   0   0
  unknown 1  0  0  0  0  0   0   0   0   0
```

(b) Create a length variable that is the total length in inches.

```
> ruf$tlin <- ruf$tl/25.4
```

(c) Create a subset of just male ruffe with a total length less than 100 mm.

```
> ruf4 <- Subset(ruf,sex=="male" & tl<100)
> nrow(ruf4)
[1] 5
```

(d) What is the *tl* for all but the 10th individual?

```
> ruf$tl[-10]
 [1] 134 111 110 115  92  88  95  90  99  NA  99 102 105  90 102 114  NA  56  90 101
[21] 109 110 111 101  95  84 105 120 104 102  99  84  87  81  81  65  42  NA 115
```

(e) Show all recorded information for the 11th individual.

```
> ruf[11,]
   fishID            locShort year month day      date tl  wt    sex maturity logtl
11     70 St. Louis R. (2007) 2007     9  20 9/20/2007 NA 9.7 female   mature    NA
    logw LCat fult tlin
11 2.272 <NA>   NA   NA
```