

Size Structure I

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

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

> # load needed packages
> library(fishWiDNR) # for setDBClasses()
> library(dplyr)     # for filter(), select(), mutate(), group_by(), summarize()
> # options(dplyr.print_max=1e9)
> library(FSA)       # for expandCounts(), filterD(), Summarize(), hist(),
> library(lubridate) # for month()

> # load FM data, expand lengths, select pertinent variables ... mostly copied code from previous
> 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")
> d <- mutate(d, Mon=month(Survey.Begin.Date, label=TRUE))
> d <- select(d, Species, Waterbody.Name, Survey.Year, Gear, Survey.Begin.Date, Mon, Len)

> Spr <- filterD(d, Survey.Year==2013, Mon %in% c("Apr", "May", "Jun"))
> BGSpr <- filterD(Spr, Species=="BLUEGILL")
> BGSprLC <- filterD(BGSpr, Waterbody.Name=="LAKE CHETAC", Gear=="BOOM SHOCKER")
```

So ...

- Spr has all species sampled from all water bodies in the Spring of 2013.
- BGSpr has only Bluegill sampled from all water bodies in the Spring of 2013.
- BGSprLC has only Bluegill sampled with boom shockers from Lake Chetac in the Spring of 2013.

... and they all look similar to this ...

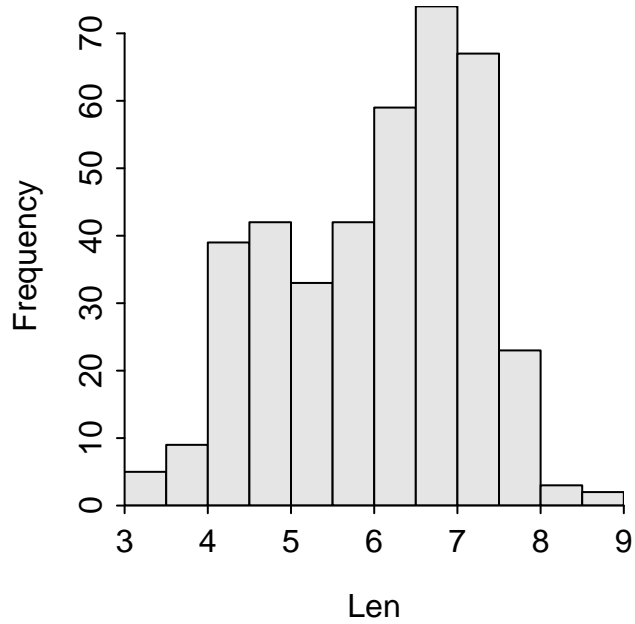
	Species	Waterbody.Name	Survey.Year	Gear	Survey.Begin.Date	Mon	Len
1	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	4.0
2	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	4.7
3	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	4.7
396	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	5.6
397	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	6.6
398	BLUEGILL	LAKE CHETAC	2013	BOOM SHOCKER	2013-05-09	May	6.6

Very Simple Summaries

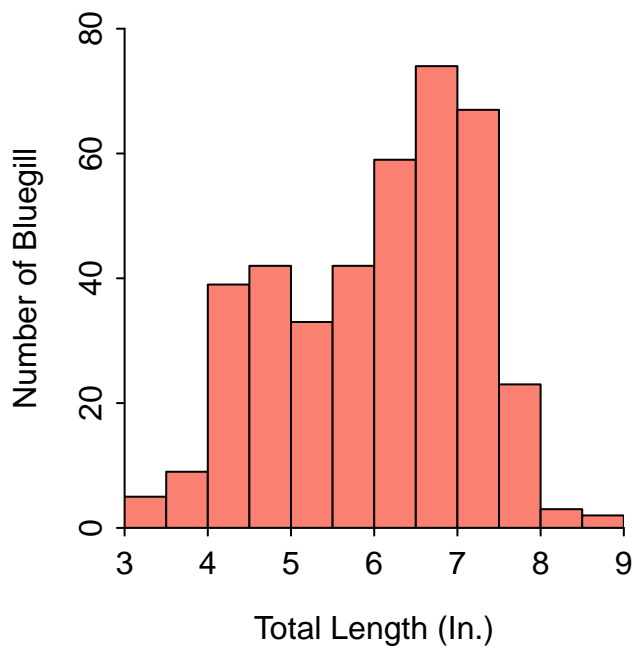
```
> Summarize(~Len, data=BGSprLC, digits=2)
      n      mean      sd      min      Q1      median      Q3      max percZero
398.00   5.98   1.16   3.00   5.00   6.20   6.90   8.90   0.00
```

Length Frequency Histograms

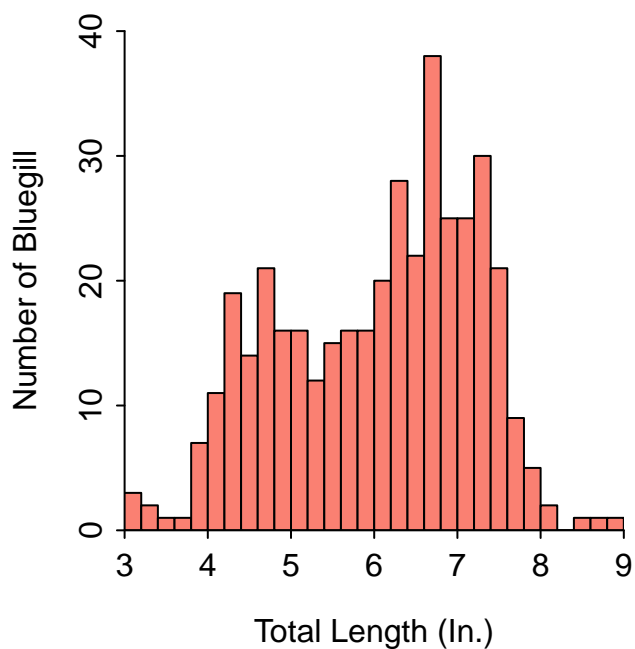
```
> hist(~Len,data=BGSprLC)
```



```
> hist(~Len,data=BGSprLC,xlab="Total Length (In.)",ylab="Number of Bluegill",  
      xlim=c(3,9),ylim=c(0,80),col="salmon")
```



```
> hist(~Len,data=BGSprLC,xlab="Total Length (In.)",ylab="Number of Bluegill",
      xlim=c(3,9),ylim=c(0,40),breaks=seq(3,9,0.2),col="#FA8072")
```



Multiple Summaries at Once

```
> BGSpr <- group_by(BGSpr,Waterbody.Name)
> summarize(BGSpr,n=n(),meanLen=mean(Len)) # see use of na.rm=TRUE below
Source: local data frame [11 x 3]
```

	Waterbody.Name	n	meanLen
1	BLACK DAN LAKE	599	NA
2	CONNORS LAKE	198	NA
3	DURPHEE LAKE	603	NA
4	GREEN LAKE	144	6.567361
5	LAKE CHETAC	589	NA
6	LAKE CHIPPEWA	746	NA
7	LAKE OF THE PINES	303	NA
8	LOWER CLAM LAKE	35	4.554286
9	MOOSE LAKE	1	NA
10	ROUND LAKE	414	NA
11	WHITEFISH LAKE	72	NA

```
> summarize(BGSpr,n=n(),valid_n=sum(!is.na(Len)),
             meanLen=mean(Len,na.rm=TRUE),sdLen=sd(Len,na.rm=TRUE),
             minLen=min(Len,na.rm=TRUE),maxLen=max(Len,na.rm=TRUE) )
Source: local data frame [11 x 7]
```

	Waterbody.Name	n	valid_n	meanLen	sdLen	minLen	maxLen
1	BLACK DAN LAKE	599	241	4.352697	0.9151520	2.1	7.0
2	CONNORS LAKE	198	108	5.155556	1.1018534	1.7	7.0
3	DURPHEE LAKE	603	574	6.603136	0.5071123	1.4	7.9
4	GREEN LAKE	144	144	6.567361	1.1392446	2.8	8.4
5	LAKE CHETAC	589	400	5.979250	1.1819420	2.0	8.9
6	LAKE CHIPPEWA	746	181	5.758011	1.1447001	3.7	8.0
7	LAKE OF THE PINES	303	90	5.000000	1.1646478	1.7	6.8
8	LOWER CLAM LAKE	35	35	4.554286	1.0042096	2.7	6.2
9	MOOSE LAKE	1	0	NaN	NaN	NA	NA
10	ROUND LAKE	414	309	5.070874	1.3018442	1.8	8.7
11	WHITEFISH LAKE	72	67	4.392537	1.3614067	2.1	7.4

```
> BGSpr <- filterD(BGSpr,Len>=3)
> summarize(BGSpr,n=n(),valid_n=sum(!is.na(Len)),
             meanLen=round(mean(Len,na.rm=TRUE),2),sdLen=round(sd(Len,na.rm=TRUE),2),
             minLen=min(Len,na.rm=TRUE),maxLen=max(Len,na.rm=TRUE),
             PSDQ=perc(Len,6,digits=0),PSD7=perc(Len,7,digits=0),PSDP=perc(Len,8,digits=0) )
Source: local data frame [10 x 10]
```

	Waterbody.Name	n	valid_n	meanLen	sdLen	minLen	maxLen	PSDQ	PSD7	PSDP
1	BLACK DAN LAKE	236	236	4.39	0.89	3.0	7.0	4	1	0
2	CONNORS LAKE	102	102	5.32	0.89	3.0	7.0	28	1	0
3	DURPHEE LAKE	573	573	6.61	0.46	4.5	7.9	94	21	0
4	GREEN LAKE	142	142	6.62	1.06	3.0	8.4	79	44	6
5	LAKE CHETAC	399	399	5.99	1.17	3.0	8.9	57	24	2
6	LAKE CHIPPEWA	181	181	5.76	1.14	3.7	8.0	44	20	1
7	LAKE OF THE PINES	83	83	5.23	0.87	3.0	6.8	20	0	0
8	LOWER CLAM LAKE	34	34	4.61	0.97	3.0	6.2	12	0	0
9	ROUND LAKE	296	296	5.18	1.21	3.0	8.7	25	9	2
10	WHITEFISH LAKE	59	59	4.65	1.25	3.0	7.4	15	8	0

```
> Spr <- group_by(Spr,Waterbody.Name,Species)
> tmp <- summarize(Spr,n=n(),valid_n=sum(!is.na(Len)),
                   meanLen=round(mean(Len,na.rm=TRUE),2),sdLen=round(sd(Len,na.rm=TRUE),2) )
> tmp
# only partial results shown
Source: local data frame [122 x 6]
Groups: Waterbody.Name
```

	Waterbody.Name	Species	n	valid_n	meanLen	sdLen
1	BLACK DAN LAKE	BLACK BULLHEAD	2	0	NA	NA
2	BLACK DAN LAKE	BLACK CRAPPIE	402	402	6.89	1.42
3	BLACK DAN LAKE	BLUEGILL	599	241	4.35	0.92
4	BLACK DAN LAKE	LARGEMOUTH BASS	76	76	11.01	3.15
5	BLACK DAN LAKE	MUSKELLUNGE	38	15	34.88	7.35
6	BLACK DAN LAKE	NORTHERN PIKE	8	8	22.91	5.82
7	BLACK DAN LAKE	PUMPKINSEED	43	31	4.61	1.12
8	BLACK DAN LAKE	PUMPKINSEED X BLUEGILL	13	9	5.36	1.01
9	BLACK DAN LAKE	ROCK BASS	4	4	4.40	2.23
10	BLACK DAN LAKE	WALLEYE	180	180	10.74	5.02
11	BLACK DAN LAKE	WARMOUTH	5	2	5.00	0.71
12	BLACK DAN LAKE	WHITE CRAPPIE	14	14	8.07	1.93
13	BLACK DAN LAKE	WHITE SUCKER	1	0	NaN	NaN
14	BLACK DAN LAKE	YELLOW PERCH	122	122	5.86	1.05

15	BLUEBERRY LAKE	LARGEMOUTH BASS	61	61	11.43	2.01
16	CONNORS LAKE	BLUEGILL	198	108	5.16	1.10
17	CONNORS LAKE	COMMON SHINER	4	0	NaN	NaN
18	CONNORS LAKE	GOLDEN SHINER	4	0	NaN	NaN
19	CONNORS LAKE	LARGEMOUTH BASS	3	2	15.25	0.07
20	CONNORS LAKE	MUSKELLUNGE	41	41	26.96	9.28
21	CONNORS LAKE	NORTHERN PIKE	8	8	22.05	3.50
22	CONNORS LAKE	PUMPKINSEED	13	9	5.24	1.07
23	CONNORS LAKE	ROCK BASS	24	10	5.86	0.78
24	CONNORS LAKE	SMALLMOUTH BASS	32	32	12.68	3.16
25	CONNORS LAKE	WALLEYE	501	501	12.83	4.67
26	CONNORS LAKE	WHITE SUCKER	27	9	10.78	2.08
27	CONNORS LAKE	YELLOW PERCH	1811	82	5.58	1.49
28	DURPHEE LAKE	BLACK CRAPPIE	100	100	7.46	0.74
29	DURPHEE LAKE	BLUEGILL	603	574	6.60	0.51
30	DURPHEE LAKE	LARGEMOUTH BASS	53	53	14.25	1.82
31	DURPHEE LAKE	NORTHERN PIKE	99	99	20.93	3.94
32	DURPHEE LAKE	ROCK BASS	17	15	6.77	0.91
33	DURPHEE LAKE	SMALLMOUTH BASS	3	3	12.07	2.16
34	DURPHEE LAKE	WALLEYE	24	24	19.10	6.12
35	DURPHEE LAKE	WHITE SUCKER	176	0	NaN	NaN
36	DURPHEE LAKE	YELLOW PERCH	4	4	3.70	1.07
37	GREEN LAKE	BLUEGILL	144	144	6.57	1.14
38	GREEN LAKE	LARGEMOUTH BASS	73	73	12.16	2.07
39	HAYWARD LAKE	LARGEMOUTH BASS	28	28	14.05	1.65
40	HAYWARD LAKE	MUSKELLUNGE	12	12	40.79	5.90
41	HAYWARD LAKE	NORTHERN PIKE	86	86	19.27	6.40
42	HAYWARD LAKE	SMALLMOUTH BASS	1	1	19.00	NaN
43	HAYWARD LAKE	WALLEYE	3	3	19.00	7.55
44	LAKE CHETAC	BLACK CRAPPIE	3619	548	6.15	1.35
45	LAKE CHETAC	BLUEGILL	589	400	5.98	1.18
46	LAKE CHETAC	BOWFIN	10	10	20.42	5.82
47	LAKE CHETAC	LARGEMOUTH BASS	274	274	12.68	2.76
48	LAKE CHETAC	NORTHERN PIKE	40	40	23.01	5.36
49	LAKE CHETAC	PUMPKINSEED	36	36	5.90	0.90
50	LAKE CHETAC	ROCK BASS	2	0	NaN	NaN
51	LAKE CHETAC	SMALLMOUTH BASS	10	10	11.68	2.95
52	LAKE CHETAC	WALLEYE	72	71	18.51	6.83
53	LAKE CHETAC	YELLOW PERCH	2222	385	7.01	1.36
54	LAKE CHIPPEWA	BLACK CRAPPIE	306	306	7.98	1.46
55	LAKE CHIPPEWA	BLUEGILL	746	181	5.76	1.14
56	LAKE CHIPPEWA	LARGEMOUTH BASS	147	147	13.34	2.78
57	LAKE CHIPPEWA	MUSKELLUNGE	7	7	35.41	14.77
58	LAKE CHIPPEWA	NORTHERN PIKE	185	184	20.87	3.62
59	LAKE CHIPPEWA	ROCK BASS	10	10	6.56	1.08
60	LAKE CHIPPEWA	SMALLMOUTH BASS	28	28	14.03	2.35
61	LAKE CHIPPEWA	WALLEYE	182	182	17.53	4.82
62	LAKE CHIPPEWA	WHITE SUCKER	3	0	NaN	NaN
63	LAKE CHIPPEWA	YELLOW PERCH	950	931	6.33	0.86
64	LAKE OF THE PINES	BLACK CRAPPIE	31	29	6.01	1.46
65	LAKE OF THE PINES	BLUEGILL	303	90	5.00	1.16
66	LAKE OF THE PINES	COMMON SHINER	1	0	NaN	NaN
67	LAKE OF THE PINES	GOLDEN SHINER	8	0	NaN	NaN
68	LAKE OF THE PINES	JOHNNY DARTER	1	0	NaN	NaN
69	LAKE OF THE PINES	LARGEMOUTH BASS	19	19	11.24	3.38
70	LAKE OF THE PINES	MUSKELLUNGE	54	54	30.55	6.39
71	LAKE OF THE PINES	NORTHERN PIKE	2	2	24.90	0.14
72	LAKE OF THE PINES	PUMPKINSEED	28	11	4.88	0.89
73	LAKE OF THE PINES	ROCK BASS	17	1	6.90	NaN

74	LAKE OF THE PINES	SMALLMOUTH BASS	3	3	12.93	2.16
75	LAKE OF THE PINES	WALLEYE	250	250	11.47	3.77
76	LAKE OF THE PINES	WHITE SUCKER	8	5	7.08	1.34
77	LAKE OF THE PINES	YELLOW PERCH	254	25	3.66	1.35
78	LOWER CLAM LAKE	BLACK CRAPPIE	4	4	7.78	1.27
79	LOWER CLAM LAKE	BLUEGILL	35	35	4.55	1.00
80	LOWER CLAM LAKE	LARGEMOUTH BASS	19	19	13.92	3.04
81	LOWER CLAM LAKE	MUSKELLUNGE	43	43	23.98	7.63
82	LOWER CLAM LAKE	NORTHERN PIKE	17	17	22.82	4.53
83	LOWER CLAM LAKE	PUMPKINSEED	8	8	5.04	0.72
84	LOWER CLAM LAKE	SHORTHEAD REDHORSE	1	1	14.20	NaN
85	LOWER CLAM LAKE	WALLEYE	95	95	15.23	4.76
86	LOWER CLAM LAKE	WHITE SUCKER	2	2	11.65	2.05
87	LOWER CLAM LAKE	YELLOW PERCH	8	8	4.03	1.00
88	MOOSE LAKE	BLACK CRAPPIE	103	103	9.58	1.67
89	MOOSE LAKE	BLUEGILL	1	0	NaN	NaN
90	MOOSE LAKE	GOLDEN REDHORSE	1	1	17.70	NaN
91	MOOSE LAKE	LARGEMOUTH BASS	1	1	17.50	NaN
92	MOOSE LAKE	MUSKELLUNGE	12	12	28.55	6.16
93	MOOSE LAKE	ROCK BASS	35	0	NaN	NaN
94	MOOSE LAKE	SHORTHEAD REDHORSE	21	9	11.96	2.17
95	MOOSE LAKE	SMALLMOUTH BASS	13	13	13.58	3.77
96	MOOSE LAKE	STRIPED SHINER	1	1	9.40	NaN
97	MOOSE LAKE	WALLEYE	115	115	10.04	3.77
98	MOOSE LAKE	WHITE SUCKER	3	3	11.77	2.87
99	MOOSE LAKE	YELLOW PERCH	31	31	6.92	0.66
100	OSPREY LAKE	LARGEMOUTH BASS	30	30	12.21	3.01
101	ROUND LAKE	BLACK CRAPPIE	60	60	9.84	1.34
102	ROUND LAKE	BLUEGILL	414	309	5.07	1.30
103	ROUND LAKE	LARGEMOUTH BASS	58	58	12.09	3.46
104	ROUND LAKE	MUSKELLUNGE	44	44	27.15	11.47
105	ROUND LAKE	NORTHERN PIKE	33	33	22.92	5.16
106	ROUND LAKE	PUMPKINSEED	22	22	5.13	1.13
107	ROUND LAKE	ROCK BASS	66	33	7.04	1.57
108	ROUND LAKE	SMALLMOUTH BASS	144	144	15.81	3.27
109	ROUND LAKE	WALLEYE	191	191	11.40	4.69
110	ROUND LAKE	WARMOUTH	1	0	NaN	NaN
111	ROUND LAKE	WHITE SUCKER	27	1	5.70	NaN
112	ROUND LAKE	YELLOW PERCH	616	616	6.07	1.10
113	WHITEFISH LAKE	BLACK CRAPPIE	45	45	6.76	2.38
114	WHITEFISH LAKE	BLUEGILL	72	67	4.39	1.36
115	WHITEFISH LAKE	LARGEMOUTH BASS	63	63	10.63	2.19
116	WHITEFISH LAKE	MUSKELLUNGE	8	8	23.26	11.92
117	WHITEFISH LAKE	NORTHERN PIKE	66	66	21.97	3.30
118	WHITEFISH LAKE	PUMPKINSEED	11	11	6.22	1.05
119	WHITEFISH LAKE	ROCK BASS	16	16	7.42	0.86
120	WHITEFISH LAKE	SMALLMOUTH BASS	34	34	14.08	3.42
121	WHITEFISH LAKE	WALLEYE	205	205	13.04	4.65
122	WHITEFISH LAKE	YELLOW PERCH	31	31	3.97	1.32

```
> write.csv(tmp, "LenSum_Sawyer_Spr13.csv", row.names=FALSE)
```

Application Assignment

Create a script that performs the following tasks:

1. Load and prepare your FM data in R (**HINT:** *use all or some of your scripts from previous application assignments*).
2. Reduce your `data.frame` to one year and several (4 or more) species of interest. Call this the *original data.frame*.
3. Reduce the *original data.frame* to one water body and one species of interest.
 - Compute summary statistics for the length variable.
 - Construct a length frequency histogram.
 - Does your description of the length frequency change dramatically with different bin widths?
4. Reduce the *original data.frame* to only one species (this will include several water bodies).
 - Efficiently construct summary statistics for the length variable for each water body. Include PSD values that are of interest to you (**HINT:** *use, for example, `psdVal("Largemouth Bass",units="in")` to find Gabelhouse lengths for a particular species*).
5. (*Time Permitting*) Re-create the summary statistics for one species in each water body but include calculations of the median and first and third quartiles (**HINT:** *use, for example, `quantile(x,0.50,na.rm=TRUE)` to compute the median (i.e., 50% quantile) of the data in x*).
6. (*Time Permitting*) Compute summary statistics of the length variable for each water body AND each of the several species of interest to you. Save the summary statistics to an object and write the results to a CSV file.

Save your script!