R Handout - R Summarization — Dr. Derek Ogle

Northland College, Midwest (Wichita) — Dec 2013

```
> library(FSA)        # view, Subset, Summarize, hist
> library(plotrix)    # plotH
```

0.1 Herman Lake Data

```
> ## setwd("C:/aaaWork/Web/fishR/courses/Midwest2012/CourseMaterial/")
> Herman <- read.table("Herman.txt",header=TRUE,sep="\t")
> str(Herman)

'data.frame': 5931 obs. of  3 variables:
 $ tl  : num  16.6 19.2 20.1 21.1 27.7 27.5 28 29.7 30.3 31.4 ...
 $ spec: Factor w/ 4 levels "bbh","bkc","wae",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ yr  : int  1999 1999 1999 1999 1999 1999 1999 1999 1999 1999 ...

> levels(Herman$spec)

[1] "bbh" "bkc" "wae" "yep"

> Herman$fyr <- factor(Herman$yr)
> str(Herman)

'data.frame': 5931 obs. of  4 variables:
 $ tl  : num  16.6 19.2 20.1 21.1 27.7 27.5 28 29.7 30.3 31.4 ...
 $ spec: Factor w/ 4 levels "bbh","bkc","wae",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ yr  : int  1999 1999 1999 1999 1999 1999 1999 1999 1999 1999 ...
 $ fyr : Factor w/ 5 levels "1997","1999",..: 2 2 2 2 2 2 2 2 2 2 ...

> levels(Herman$fyr)

[1] "1997" "1999" "2001" "2003" "2005"
```

0.2 Simple Tabular or Numerical Summaries

Frequency and Percentage Tables

```
> ( tSpec <- table(Herman$spec) )


 bbh  bkc  wae  yep
5202  268  191  270
> ( ptSpec <- prop.table(tSpec)*100 )


   bbh    bkc    wae    yep
87.709  4.519  3.220  4.552
> round(ptSpec,1)


 bbh  bkc  wae  yep
87.7  4.5  3.2  4.6
> ( tSpecYr <- table(Herman$spec,Herman$fyr) )
```

```
      1997 1999 2001 2003 2005
  bbh    0   99  100 4811  192
  bkc   66    0  100  100    2
  wae    0   51   14   80   46
  yep    0   50   55  113   52
> addmargins(tSpecYr)


      1997 1999 2001 2003 2005  Sum
  bbh    0   99  100 4811  192 5202
  bkc   66    0  100  100    2  268
  wae    0   51   14   80   46  191
  yep    0   50   55  113   52  270
  Sum   66  200  269 5104  292 5931
> round(prop.table(tSpecYr,margin=2)*100,1)


       1997  1999  2001  2003  2005
  bbh   0.0  49.5  37.2  94.3  65.8
  bkc 100.0   0.0  37.2   2.0   0.7
  wae   0.0  25.5   5.2   1.6  15.8
  yep   0.0  25.0  20.4   2.2  17.8
```

Summary Statistics

```
> Hbbh <- Subset(Herman,spec=="bbh")
> Summarize(Hbbh$tl,digits=1)
       n      mean        sd      min       Q1   median       Q3      max percZero
  5202.0      25.3       2.6     16.3     23.8     24.8     25.9     35.4      0.0
> Summarize(tl~spec,data=Herman,digits=1)
  spec    n mean  sd  min   Q1 median   Q3  max percZero
1  bbh 5202 25.3 2.6 16.3 23.8   24.8 25.9 35.4        0
2  bkc  268 23.3 2.5 17.3 21.5   23.1 24.6 31.1        0
3  wae  191 33.8 7.7 16.6 27.2   34.3 36.7 71.7        0
4  yep  270 23.3 3.8 11.6 21.8   23.3 25.6 31.9        0
> Summarize(tl~fyr*spec,data=Herman,digits=1)
    fyr spec    n mean  sd  min   Q1 median   Q3  max percZero
1  1999  bbh   99 28.6 3.3 16.3 28.2   29.4 30.5 35.4        0
2  2001  bbh  100 29.0 3.1 16.6 27.8   29.2 31.1 33.8        0
3  2003  bbh 4811 25.0 2.2 19.0 23.7   24.7 25.7 32.9        0
4  2005  bbh  192 29.9 2.7 21.7 28.7   30.5 31.7 34.8        0
5  1997  bkc   66 24.8 2.1 19.1 23.8   25.2 26.0 30.7        0
6  2001  bkc  100 21.5 1.5 17.4 20.6   21.6 22.6 25.5        0
7  2003  bkc  100 24.0 2.4 19.6 22.6   23.6 24.6 31.1        0
8  2005  bkc    2 18.4 1.6 17.3 17.9   18.5 19.0 19.6        0
9  1999  wae   51 33.2 6.7 16.6 31.4   33.3 35.2 59.2        0
10 2001  wae   14 29.2 7.3 19.3 22.4   31.5 34.7 40.7        0
11 2003  wae   80 38.6 4.8 32.4 35.4   36.6 40.6 54.1        0
12 2005  wae   46 27.4 7.5 21.4 25.2   26.4 26.8 71.7        0
13 1999  yep   50 24.0 3.7 14.5 20.9   25.1 26.8 29.8        0
14 2001  yep   55 21.2 3.3 13.6 20.4   22.0 22.9 29.2        0
15 2003  yep  113 23.2 1.8 18.4 22.2   23.3 24.0 29.4        0
16 2005  yep   52 25.3 5.9 11.6 24.4   28.0 29.1 31.9        0
```
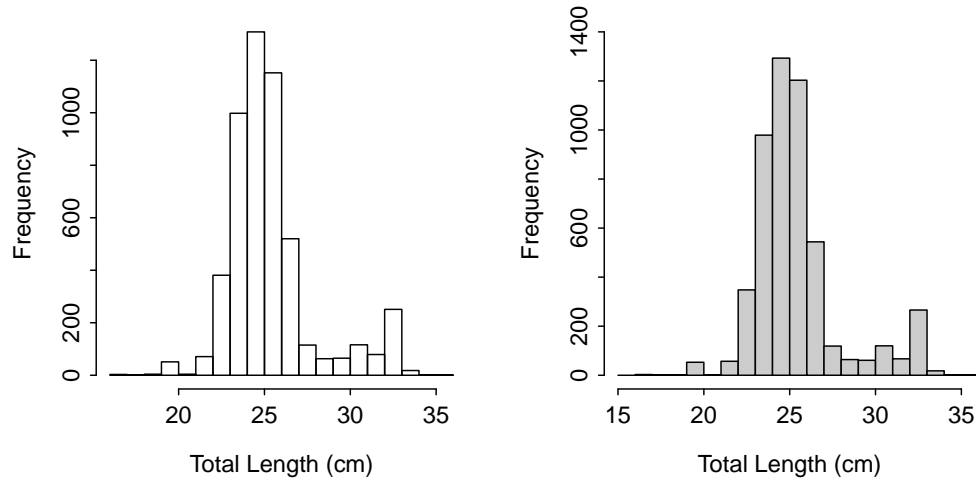
0.3 Simple Graphics

Single Histograms

```
> hist(Hbbh$tl,xlab="Total Length (cm)",main="")
> hist(Hbbh$tl,xlab="Total Length (cm)",main="",xlim=c(15,35),ylim=c(0,1400),right=FALSE,
      col="gray80",breaks=seq(15,36,1))
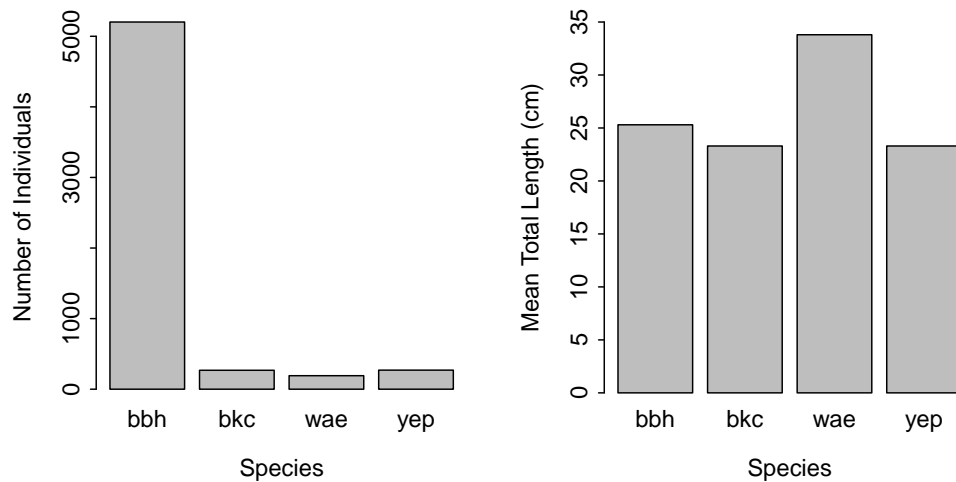```



**"Bar" Charts and Plots**

```
> tSpec

 bbh bkc wae yep
5202 268 191 270
> barplot(tSpec,xlab="Species",ylab="Number of Individuals")
> (sumSpec <- Summarize(tl~spec,data=Herman,digits=1))
  spec   n mean sd min   Q1 median   Q3 max percZero
1 bbh 5202 25.3 2.6 16.3 23.8   24.8 25.9 35.4   0
2 bkc 268 23.3 2.5 17.3 21.5   23.1 24.6 31.1   0
3 wae 191 33.8 7.7 16.6 27.2   34.3 36.7 71.7   0
4 yep 270 23.3 3.8 11.6 21.8   23.3 25.6 31.9   0
> plotH(mean~spec,data=sumSpec,xlab="Species",ylab="Mean Total Length (cm)",ylim=c(0,35))
```

Number of Individuals

5000
3000
1000
0

bbh    bkc    wae    yep

Species

Mean Total Length (cm)

35
30
25
20
15
10
5
0

bbh    bkc    wae    yep

Species

**Single Scatterplots**

```
> BullTroutRML1 <- read.table("BullTroutRML1.txt",header=TRUE,sep="\t")
> str(BullTroutRML1)

'data.frame': 137 obs. of  3 variables:
 $ fl  : int  90 180 201 346 359 362 373 380 375 396 ...
 $ mass: int  11 107 119 587 539 659 719 779 839 755 ...
 $ era : Factor w/ 2 levels "1977-79","2001": 1 1 1 1 1 1 1 1 1 1 ...

> BullTroutRML1$logFL <- log(BullTroutRML1$fl)
> BullTroutRML1$logW <- log(BullTroutRML1$mass)
> str(BullTroutRML1)

'data.frame': 137 obs. of  5 variables:
 $ fl   : int  90 180 201 346 359 362 373 380 375 396 ...
 $ mass : int  11 107 119 587 539 659 719 779 839 755 ...
 $ era  : Factor w/ 2 levels "1977-79","2001": 1 1 1 1 1 1 1 1 1 1 ...
 $ logFL: num  4.5 5.19 5.3 5.85 5.88 ...
 $ logW : num  2.4 4.67 4.78 6.38 6.29 ...

> view(BullTroutRML1)

     fl mass   era logFL  logW
29  439  988 2001 6.084 6.896
111 190   95 2001 5.247 4.554
114 184   83 2001 5.215 4.419
130 130   36 2001 4.868 3.584
133 102   12 2001 4.625 2.485
134  93   12 2001 4.533 2.485
```

```
> plot(mass~fl,data=BullTroutRML1,xlab="Fork Length (mm)",ylab="Weight (g)")
> plot(logW~logFL,data=BullTroutRML1,xlab="log Fork Length (mm)",ylab="log Weight (g)",
    pch=16)
> cor(BullTroutRML1$logW,BullTroutRML1$logFL)

[1] 0.9925
```
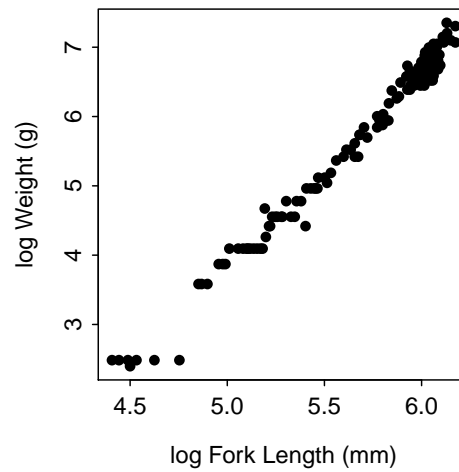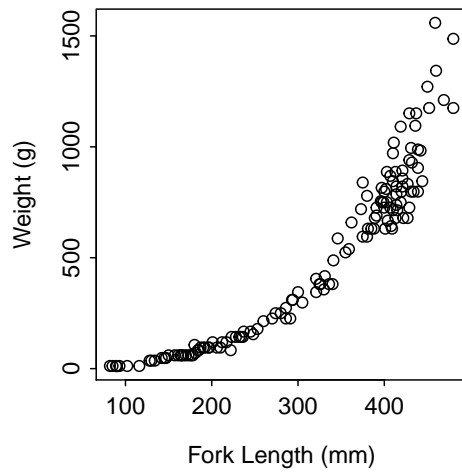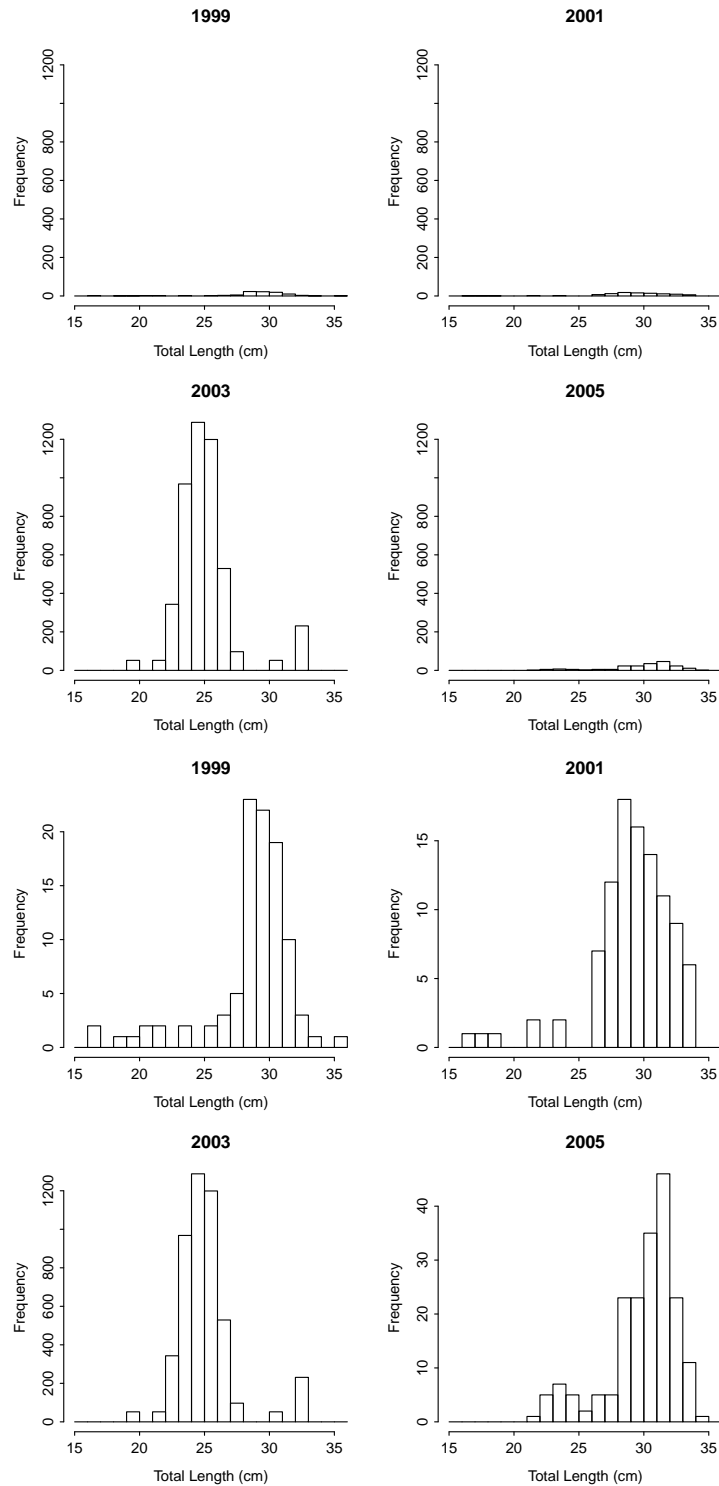
0.4 Some "Extras"

Multiple Histograms

```
> hist(tl~fyr,data=Hbbh,xlab="Total Length (cm)",main="",right=FALSE,breaks=seq(15,36,1))
> hist(tl~fyr,data=Hbbh,xlab="Total Length (cm)",main="",right=FALSE,
      breaks=seq(15,36,1),same.ylim=FALSE)
```

Scatterplots with Different Points

```
> pts <- c(16,3)
> clrs <- c("black","red")
> plot(logW~logFL,data=BullTroutRML1,xlab="log Fork Length (mm)",ylab="log Weight (g)",
 pch=pts[era],col=clrs[era])
> legend("topleft",legend=levels(BullTroutRML1$era),pch=pts,col=clrs)
```