Summarization I

Preliminaries

Load Necessary Packages

```
> library(FSA)      # for filterD(), rcumsum(), hist(), Summarize()
> library(dplyr)    # for mutate()
```

Load Data

```
> # Set your working directory to where your external data files (and scripts) are located.
> setwd("C:/aaaWork/Web/GitHub/RcourseNunavut2016/Handouts")
> dSC <- read.csv("SawyerCo_reduced.csv")
> names(dSC)
 [1] "waterbody" "year"      "mon"       "gear"      "species"   "len"       "weight"    "sex"
 [9] "age"       "age_strux" "lennote"
```

```
> levels(dSC$sex)
[1] ""  "F" "M" "U"
```

```
> dSC <- mutate(dSC,sex=mapvalues(sex,from="",to="ND"),fyear=factor(year),lcat25=lencat(len,w=25))
> levels(dSC$waterbody)
 [1] "BLACK DAN LAKE"  "CHIPPEWA RIVER"  "CONNORS LAKE"    "GRINDSTONE LAKE" "HUNTER LAKE"
 [6] "LAKE CHETAC"     "LAKE CHIPPEWA"   "MOSQUITO BROOK"  "NAMEKAGON RIVER" "NELSON LAKE"
[11] "SAND LAKE"
```

```
> levels(dSC$species)
 [1] "Black Crappie"      "Bluegill"           "Brook Trout"        "Brown Trout"
 [5] "Lake Sturgeon"      "Largemouth Bass"    "Muskellunge"        "Northern Pike"
 [9] "Pumpkinseed"        "Rock Bass"          "Shorthead Redhorse" "Smallmouth Bass"
[13] "Walleye"            "White Sucker"       "Yellow Perch"
```

```
> LChip_WAE <- filterD(dSC,waterbody=="LAKE CHIPPEWA",species=="Walleye")
> LChip_WAE11 <- filterD(LChip_WAE,year==2011)
```

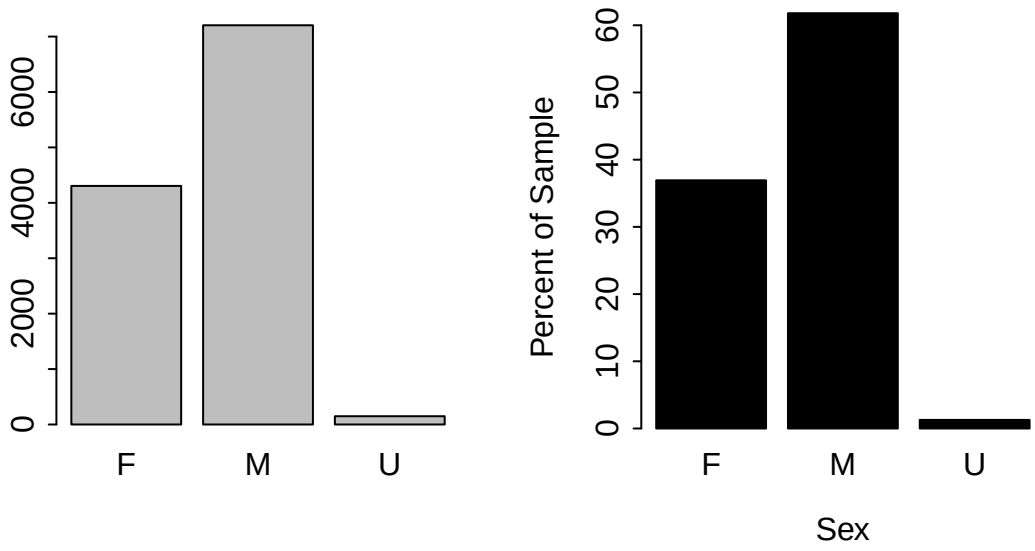
Univariate Summaries – Categorical Variables

```
> ( t_sex <- xtabs(~sex,data=LChip_WAE11) )
sex
  ND    F    M    U
1750 4305 7204  149
```

```
> ( t_sex1 <- t_sex[-1] )
sex
   F    M    U
4305 7204  149
```

```
> ( tp_sex1 <- prop.table(t_sex1)*100 )
sex
        F         M         U
36.927432 61.794476  1.278092
```

```
> barplot(t_sex1)  # Left
> barplot(tp_sex1,xlab="Sex",ylab="Percent of Sample",ylim=c(0,60),col="black")  # Right
```




```
> ( t_len25 <- xtabs(~lcat25,data=LChip_WAE11) )
lcat25
 100  125  150  175  200  225  250  275  300  325  350  375  400  425  450  475  500  525  550  575
   5   32  544  829   32   34  122  283  842 1857 1986 2020 1714  973  679  356  273  184  176  132
 600  625  650  675  700  725  750
 111  101   52   30   30   10    1
```

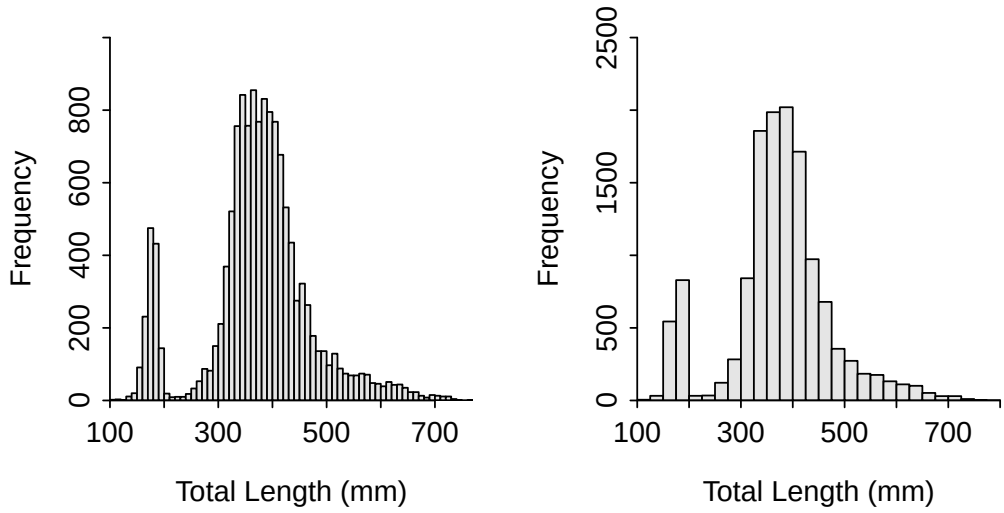
```
> tp_len25 <- prop.table(t_len25)*100
> round(tp_len25,1)
lcat25
 100  125  150  175  200  225  250  275  300  325  350  375  400  425  450  475  500  525  550  575
 0.0  0.2  4.1  6.2  0.2  0.3  0.9  2.1  6.3 13.8 14.8 15.1 12.8  7.3  5.1  2.7  2.0  1.4  1.3  1.0
 600  625  650  675  700  725  750
 0.8  0.8  0.4  0.2  0.2  0.1  0.0
```

```
> round(cumsum(tp_len25),1)
  100   125   150   175   200   225   250   275   300   325   350   375   400   425   450   475
  0.0   0.3   4.3  10.5  10.8  11.0  11.9  14.0  20.3  34.2  49.0  64.0  76.8  84.1  89.1  91.8
  500   525   550   575   600   625   650   675   700   725   750
 93.8  95.2  96.5  97.5  98.3  99.1  99.5  99.7  99.9 100.0 100.0
```

```
> round(rcumsum(tp_len25),1)
  100   125   150   175   200   225   250   275   300   325   350   375   400   425   450   475
100.0 100.0  99.7  95.7  89.5  89.2  89.0  88.1  86.0  79.7  65.8  51.0  36.0  23.2  15.9  10.9
  500   525   550   575   600   625   650   675   700   725   750
  8.2   6.2   4.8   3.5   2.5   1.7   0.9   0.5   0.3   0.1   0.0
```

Univariate Summaries – Quantitative Variables

```
> hist(~len,data=LChip_WAE11,xlab="Total Length (mm)",ylim=c(0,1000),w=10)  # Left
> hist(~len,data=LChip_WAE11,xlab="Total Length (mm)",ylim=c(0,2500),breaks=seq(100,800,25))  # Right
```




```
> Summarize(~len,data=LChip_WAE11,digits=1)
       n   nvalid     mean       sd      min       Q1   median       Q3      max percZero
 13408.0  13408.0    374.0     98.8    104.0    333.0    376.0    422.0    767.0      0.0
```

Bivariate Summaries – Categorical Variables

```
> ( t_seas <- xtabs(~mon+fyear,data=LChip_WAE) )
     fyear
mon    2010  2011  2012  2013  2014
  Apr   205 11658    48     0     0
  May   175     0   109   182   327
  Sep     0  1750    29    85   418
```

```
> round(prop.table(t_seas)*100,1)
     fyear
mon   2010 2011 2012 2013 2014
  Apr  1.4 77.8  0.3  0.0  0.0
  May  1.2  0.0  0.7  1.2  2.2
  Sep  0.0 11.7  0.2  0.6  2.8
```

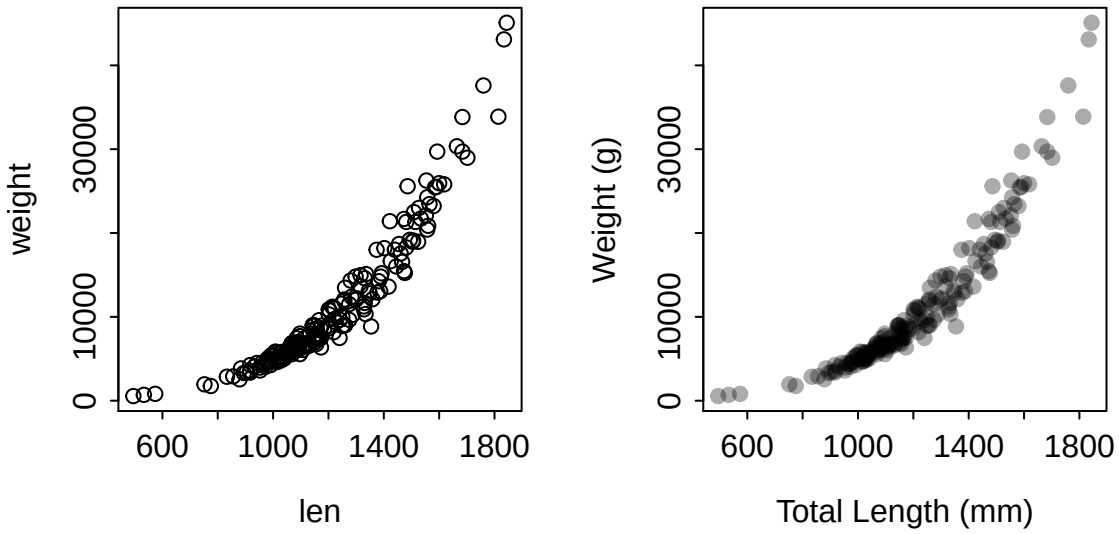
```
> round(prop.table(t_seas,margin=2)*100,1)
     fyear
mon   2010 2011 2012 2013 2014
  Apr 53.9 86.9 25.8  0.0  0.0
  May 46.1  0.0 58.6 68.2 43.9
  Sep  0.0 13.1 15.6 31.8 56.1
```

```
> round(prop.table(t_seas,margin=1)*100,1)
     fyear
mon   2010 2011 2012 2013 2014
  Apr  1.7 97.9  0.4  0.0  0.0
  May 22.1  0.0 13.7 23.0 41.2
  Sep  0.0 76.7  1.3  3.7 18.3
```

Bivariate Summaries – Quantitative Variables

```
> Sturg <- filterD(dSC,species=="Lake Sturgeon",waterbody %in% c("CHIPPEWA RIVER","HUNTER LAKE"))
```

```
> plot(weight~len,data=Sturg) # Left
> plot(weight~len,data=Sturg,pch=19,col=col2rgbt("black",1/3),
      ylab="Weight (g)",xlab="Total Length (mm)")  # Right
```




```
> with(Sturg,cor(weight,len))
[1] NA
```

```
> with(Sturg,cor(weight,len,use="pairwise.complete.obs"))
[1] 0.9303562
```

```
> with(Sturg,cor(weight,len,use="pairwise.complete.obs",method="spearman"))
[1] 0.981568
```