R Wizardry Final Project

Teresa Tufts

April 2017

In my job with the Department of Environment for the Government of Nunavut, we help collect and analyze fisheries data to monitor and promote sustainable fisheries. Our division is tasked with developing the fishery industry in Nunavut in hopes of creating a source of income and food security for remote communities. To ensure that a waterbody's fish populations are healthy and able to support a new commercial fishery or continue to support an existing one, we need to know a bit about the fish populations themselves. This document demonstrates various analyses in R that can be used to explore fisheries data in this context. I use a dataset from two lakes found on Baffin Island, Nunavut, called PG008 and PG027.

This 'technique tutorial' is by no means revolutionary nor is it terribly sophisticated. These are basic techniques in R that can be used by any budding fisheries scientist to visualize and better understand their data.

Contents:

1. Summarizing the number of fish caught per year

2. The maturity stages

3. The length-weight relationship

4. The weight-time relationship

5. The sex-size relationship

6. Revisiting the length-weight relationship with regression

7. The size-age relationship

8. Mortality rates

```r
library(dplyr) #for mutate

# You will need to install the FSA package before you begin

library(FSA) #for mapvalues, catchCurve, fact2num, Summarize, vbFuns, vbStarts
, residPlot
```

```
library(ggplot2) #for ggplot, qplot
```

First, make sure the R environment is cleared then load the dataset

```
rm(list=ls(all=T))
char = read.csv("~/Desktop/R/Final project/CharData.csv")
```

## Next, organize the dataset a bit to make it more managable

```
char = mutate(char,fyear=factor(year)) # year was an integer but it should be
made into a factor variable
char = as.data.frame(char)
char = select(char,locWB, water.type, fl, wt, sex, mat, age, fyear) # get rid
of columns not needed or not relevent to the analyses you'll be doing


# rename some variables
char = mutate(char,new_sex=mapvalues(sex,from=c("M","F","U"),to=c("male","fema
le","unknown")))
char = mutate(char, new_mat = mapvalues(mat, from = c(" M","I","M","na","R","R
R","S", "U"), to = c("mature", "immature", "mature", "na", "ripe", "running ri
pe", "spent", "unknown")))
# change some of the variables you'll be using into numeric variables, as nece
ssary.
char = transform(char, fl = as.numeric(fl), wt = as.numeric(wt), age = as.nume
ric(age))
```

Now that the data is a little bit more tidy than it was, we will look at some summaries of the data.

1. Summarizing the number of fish caught per year

Here, we'll look at the percentage of fish caught from each waterbody, by year.

```
by.year = xtabs(~fyear + locWB,data = char)
by.year = as.data.frame(by.year)
# make a plot to visualize the data
g = ggplot(data = by.year) + geom_point(aes(x=fyear, y=locWB, size = Freq), co
lor = 2) + theme_bw() +xlab("Year") + ylab("Water body")
g + scale_size_continuous(range = c(1,15))
```
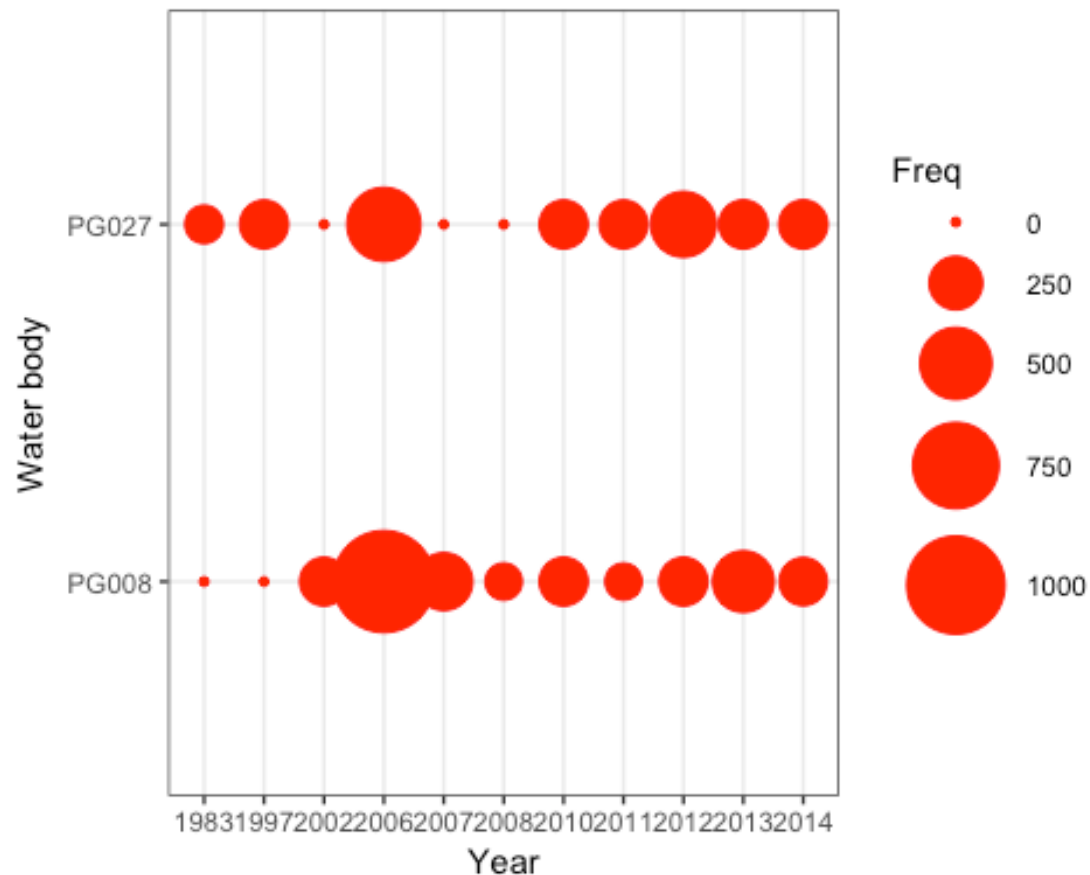
2. The maturity stages

Now, let's summarize the maturity stages of fish caught by the fish's sex, by year and by waterbody

```
# first, get rid of na and unknown sex and maturity values
char.2 = subset(char, new_mat != "na")
char.2 = subset(char.2, new_mat != "unknown")
char.2 = subset(char.2, new_sex != "na")
char.2 = subset(char.2, new_sex != "unknown")
char.2 = droplevels(char.2)
levels(char.2$new_sex) # you can see the unwanted levels are dropped

## [1] "female" "male"

# now make a table
ftable(xtabs(~locWB+fyear+new_sex+new_mat,data = char.2))
```

```
##                    new_mat mature immature ripe running ripe spent
## locWB fyear new_sex
## PG008 2002  female               0        2   20            0     0
##             male                 0        2   24            0     0
##       2007  female               0        1   25            0     0
##             male                 0        1   23            0     0
##       2010  female              62       18    5            0     0
##             male                22       33   59            0     0
##       2011  female               0        0    0            0     0
##             male                 0        0    0            0     0
##       2012  female               0        0    0            0     0
##             male                 0        0    0            0     0
##       2013  female              64       21   10            0     0
##             male                33       29   42            0     0
##       2014  female               0       52   47            0     2
##             male                 0       36   52            0     3
## PG027 2002  female               0        0    0            0     0
##             male                 0        0    0            0     0
##       2007  female               0        0    0            0     0
##             male                 0        0    0            0     0
##       2010  female               1        9   51            0     0
##             male                65       27   40            0     0
##       2011  female              25       13    2            0    30
##             male                71       30   15            0     5
##       2012  female               0       17   33            1    31
##             male                 9       20   48            0    35
##       2013  female               0       22   42            0     0
##             male                 0       34  100            1     0
##       2014  female              32        3   13            0    25
##             male                60       26   35            0     5
```

```
# Use the table to look at the data and see if there is anything that interests you about certain maturity stages. If so, you can do further statistical analyses
```

## 3. The length-weight relationship

Next, let's summarize the length-weight relationship of fish. Let's use the year 2014 as an example

```r
lw = filterD(char, fyear=="2014") # filter out only the year you're interested in
lw = select(lw,fl, wt, locWB, new_sex) # select the variables you want to work with
lw = lw[lw$fl != 0,] # omit observations with 0 length
```

```
# first, make a dataframe for the length data
len = data.frame(x = lw$fl, y = lw$locWB, na.rm = TRUE)
# plot it
ggplot(len, aes(x=x, fill=y)) + geom_histogram(alpha=0.2, position="identity",
binwidth = 10)
```
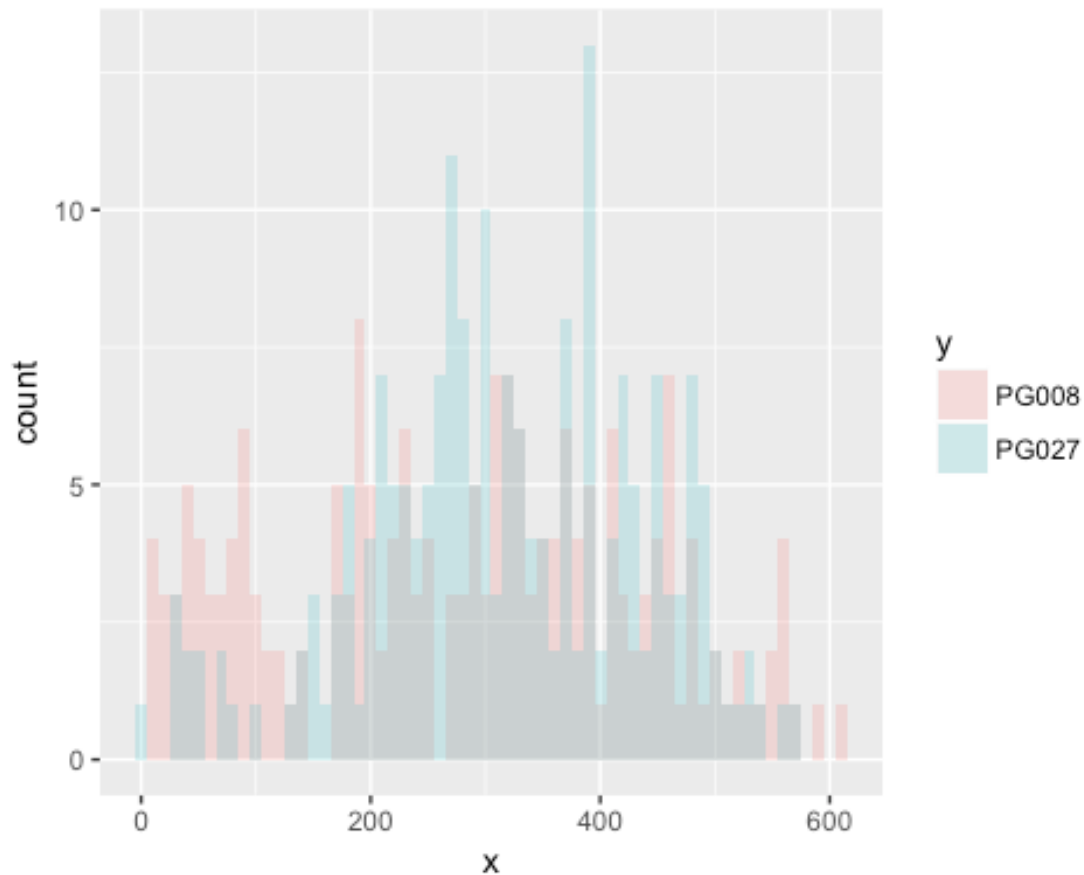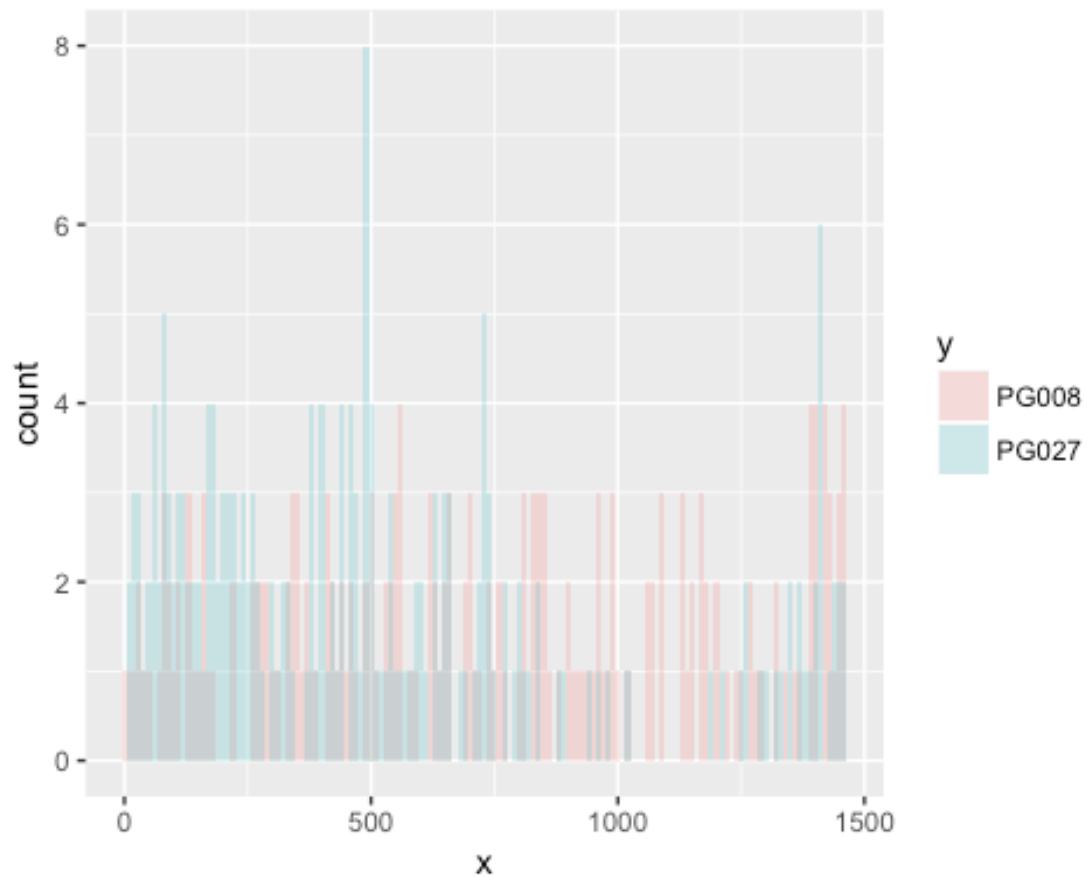


```
# notice that there seems to be more medium length fish in PG027 compared to P
G008 which has a more even distribution of lengths

# next, make a dataframe for weight data
weight = data.frame(x = lw$wt, y = lw$locWB, na.rm = TRUE)
# plot it
ggplot(weight, aes(x=x, fill=y)) + geom_histogram(alpha=0.2, position="identit
y", binwidth = 10)
```

```
# notice that there seems to be fewer heavy fish in PG027 compared to PG008

# Now compare sizes between the two lakes
qplot(fl, wt,data = lw, colour = locWB, xlab = "fork length (cm)", ylab = "wei
ght (g)", main = "Size comparison")
```

Size comparison
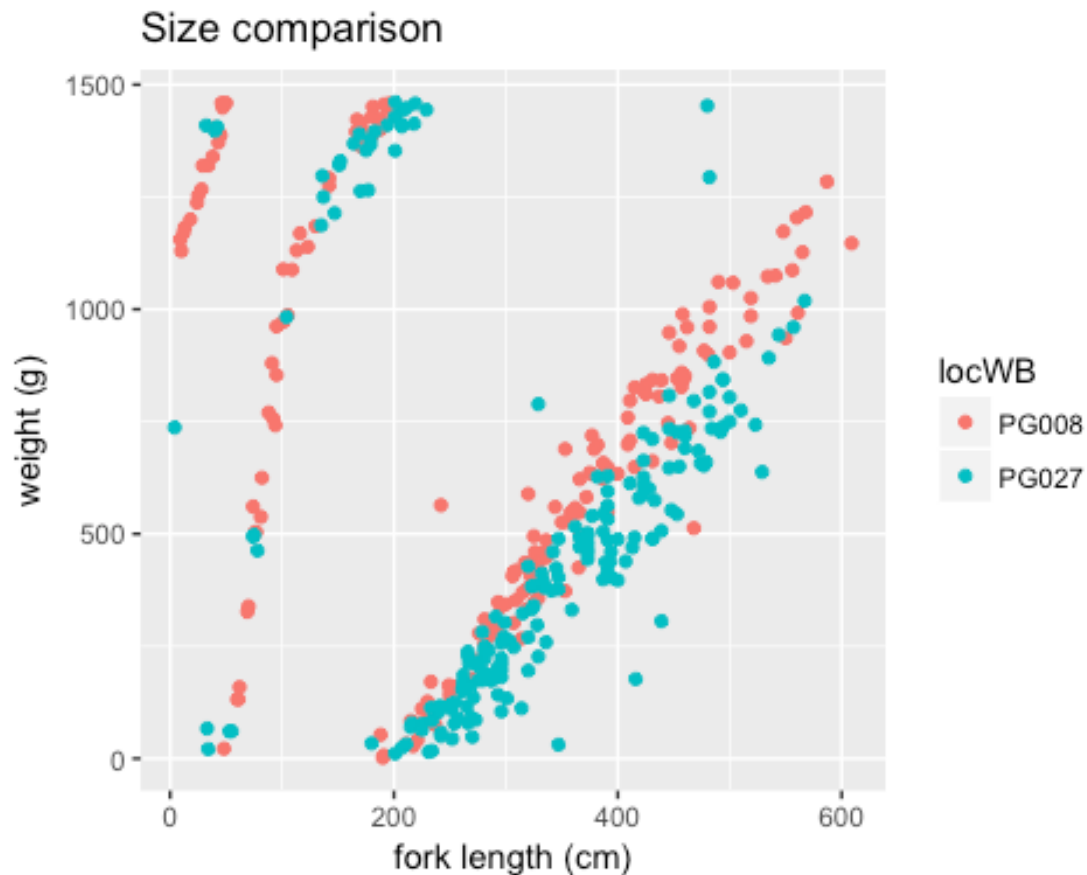
can see what looks like 3 different size classes in each lake. These either
represent juveniles, 'teens', and adults in the population or, alternatively,
a combination of juveniles and 'dwarf' adults.

Let's view data for the 'adult' size class only. Since there's no age data f
or 2014 we'll have to 'eyeball' it to determine which size likely counts as 'r
egular' adults
```
adults = lw[lw$fl > 220,]
qplot(fl, wt,data = adults, colour = locWB, xlab = "fork length (cm)", ylab =
"weight (g)", main = "Size comparison of adults")
```

Size comparison of adults
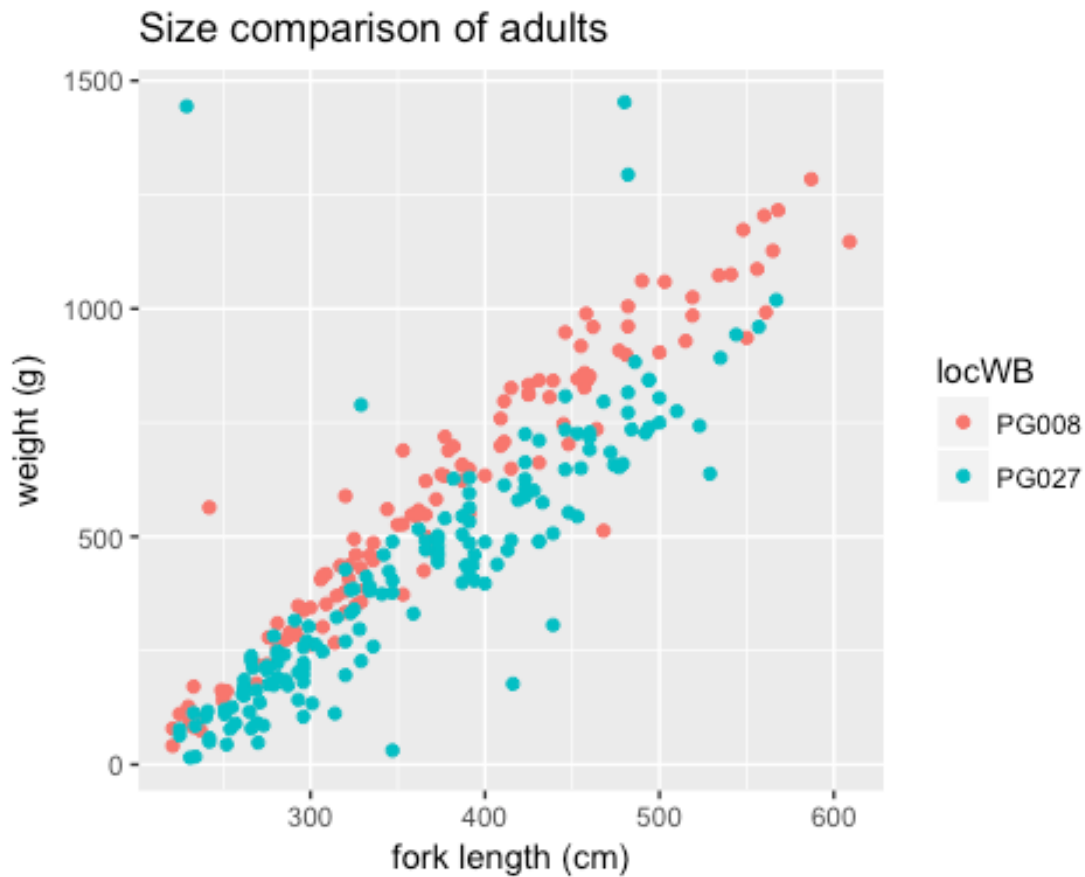
```r
# see how well length and weight are correlated in each lake
pg08 = filterD(adults, locWB=="PG008")
pg027 = filterD(adults, locWB=="PG027")
with(pg08,cor(fl,wt))

## [1] 0.962219

# In PG008, the correlation between length and weight in the adults size class
is 0.962219
with(pg027,cor(fl,wt))

## [1] 0.8213275

# In PG027, the correlation between length and weight is 0.8213275
# we can conclude that the weight-length relationship is stronger in PG008 tha
n in PG027
```

4. The weight-time relationship

Let's test to see if the mean weight of the fish differed between locations over the years

```
Summarize(wt~locWB,data=char)
```

```
##    locWB    n nvalid      mean       sd min  Q1 median   Q3  max percZero
## 1 PG008 2724   2724 1147.9394 406.7917   1 868   1464 1464 1464        0
## 2 PG027 2045   2045  970.9814 528.5456   1 460   1245 1464 1464        0
```

```
# from the table we can see that the mean weights differ between lakes but we
can't tell if the difference is significant
```

```
ggplot(char, aes(x= fyear, y = wt, fill = locWB)) +
  geom_boxplot( ) +
  ggtitle("Fish weights")+
  xlab("Year")+
  ylab("Weight (g)")
```
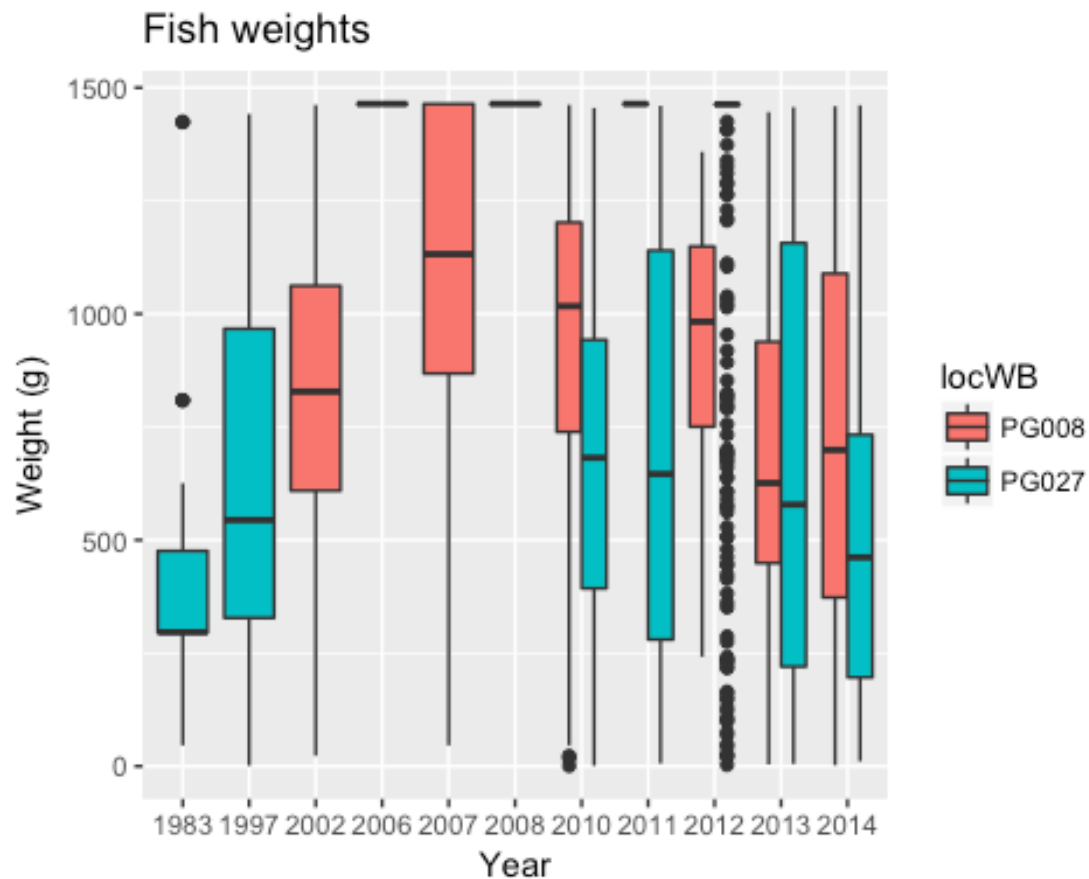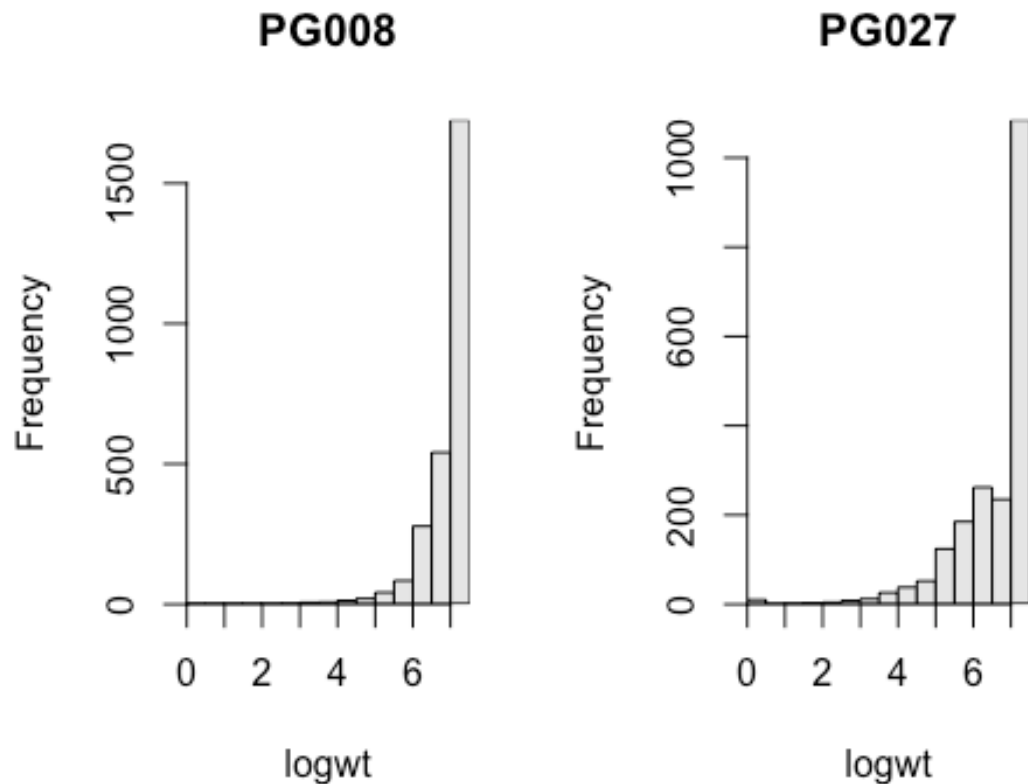


```
# The plot also suggests that PG008 has heavier fish
```

```
# Before moving ahaed, though, check to see that the data is normally distribu
ted
hist(wt~locWB,data=char,same.ylim=FALSE)
```
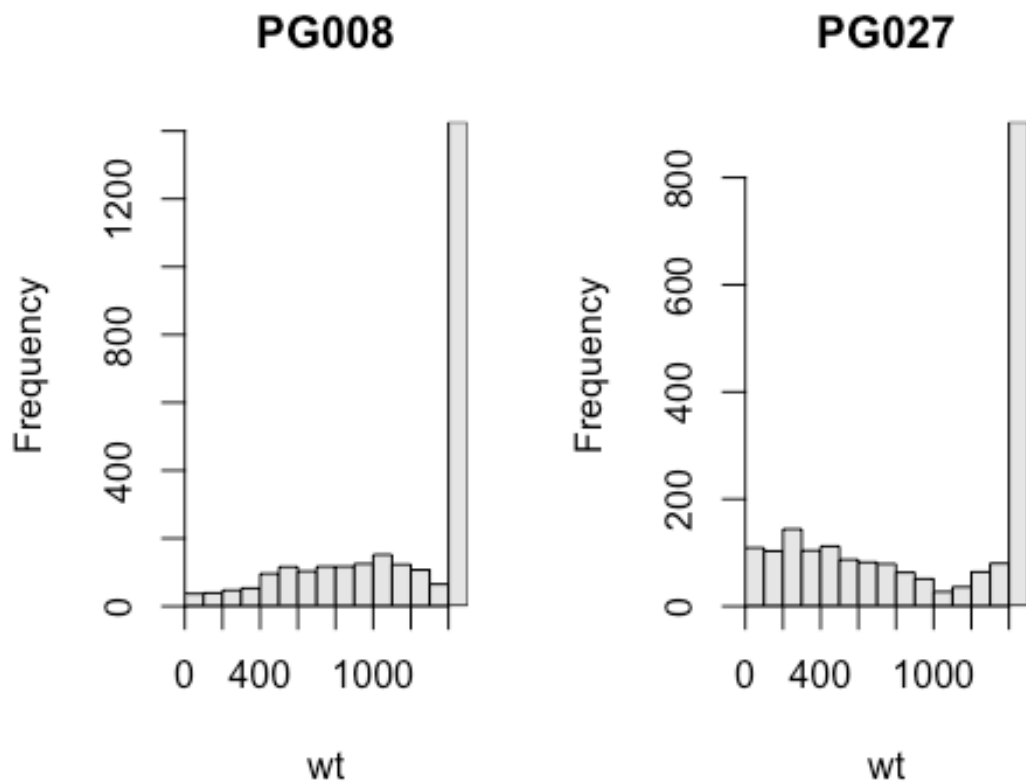


```
# data does not look normal. We should log-transform it.

char.log = mutate(char,logwt=log(wt), logfl=log(fl))
hist(logwt~locWB,data=char.log,same.ylim=FALSE)
```

PG008

PG027

```
# this looks a bit better

# Use anova to see which variables influence the weight of the fish
anova(lm(formula = logwt ~ locWB * logfl, data = char.log))

## Analysis of Variance Table
##
## Response: logwt
##                Df Sum Sq Mean Sq  F value   Pr(>F)
## locWB           1  153.5 153.506 227.8850 < 2e-16 ***
## logfl           1  204.5 204.514 303.6086 < 2e-16 ***
## locWB:logfl     1    3.8   3.789   5.6252 0.01774 *
## Residuals    4765 3209.8   0.674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# The location and length have a significant effect on the weight of the fish,
as well as their interaction.
```

5. The sex-size relationship
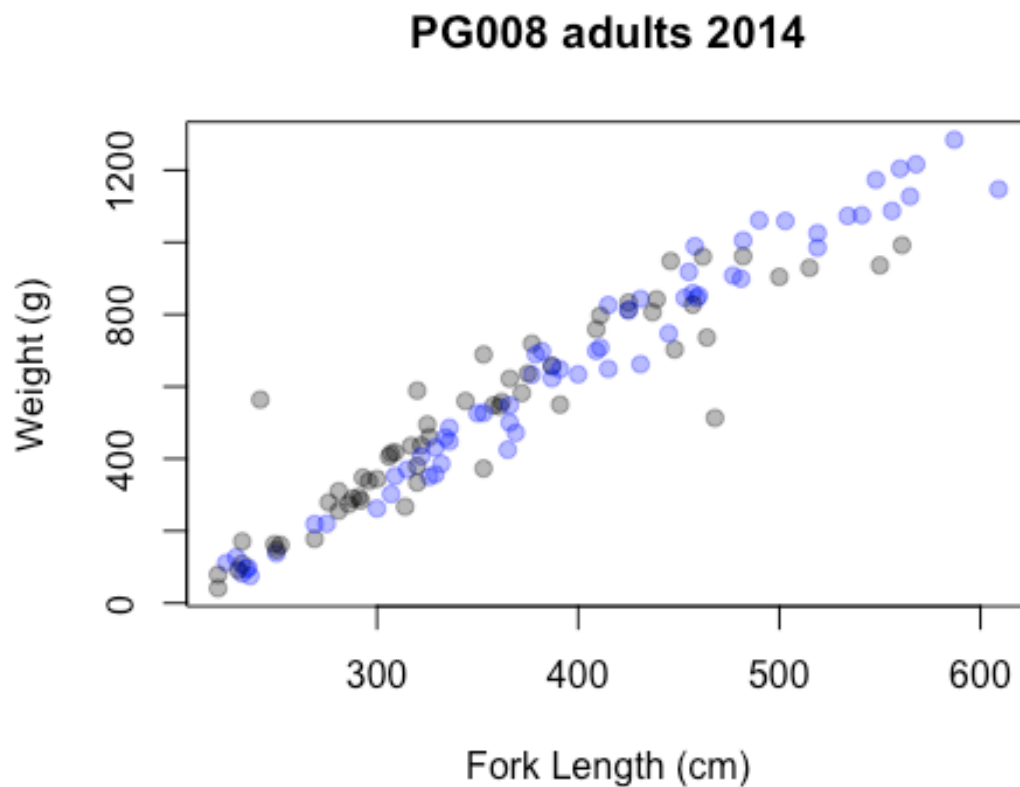
Let's see if sex has an effect on the size of the fish

```
#using the 'adults' dataset for 2014 (pg08 and pg027)

# visualize the relationship with a plot
clr1 <- c("black","blue")
clr2 <- col2rgbt(clr1,1/3)
plot(wt~fl,data=pg08,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",ylab="We
ight (g)", main = "PG008 adults 2014")
```
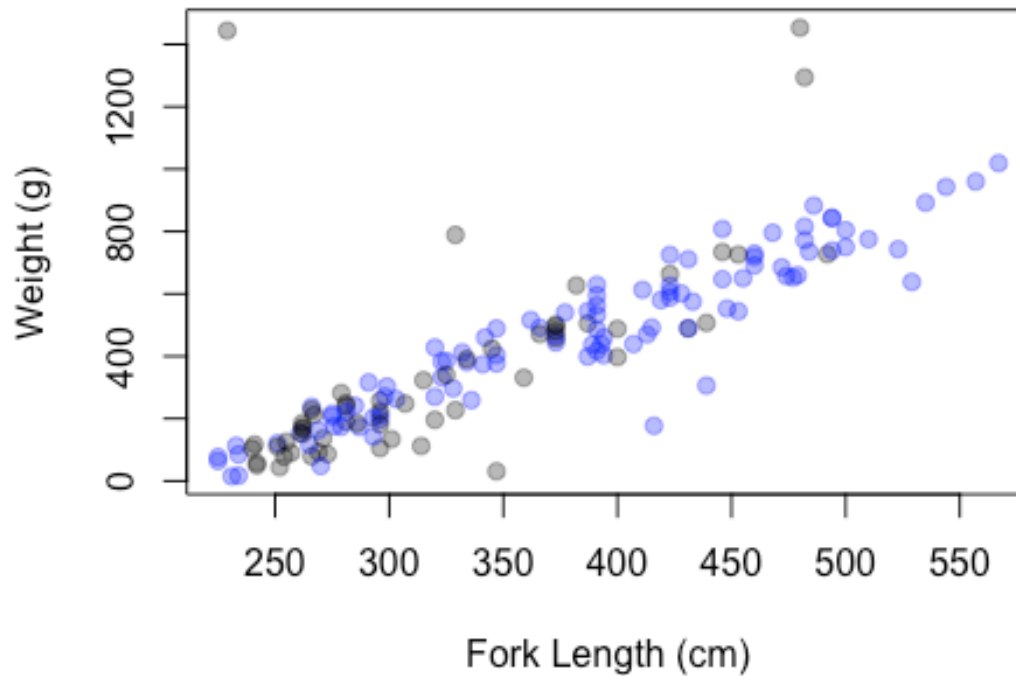

PG008 adults 2014

```
plot(wt~fl,data=pg027,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",ylab="W
eight (g)", main = "PG027 adults 2014")
```
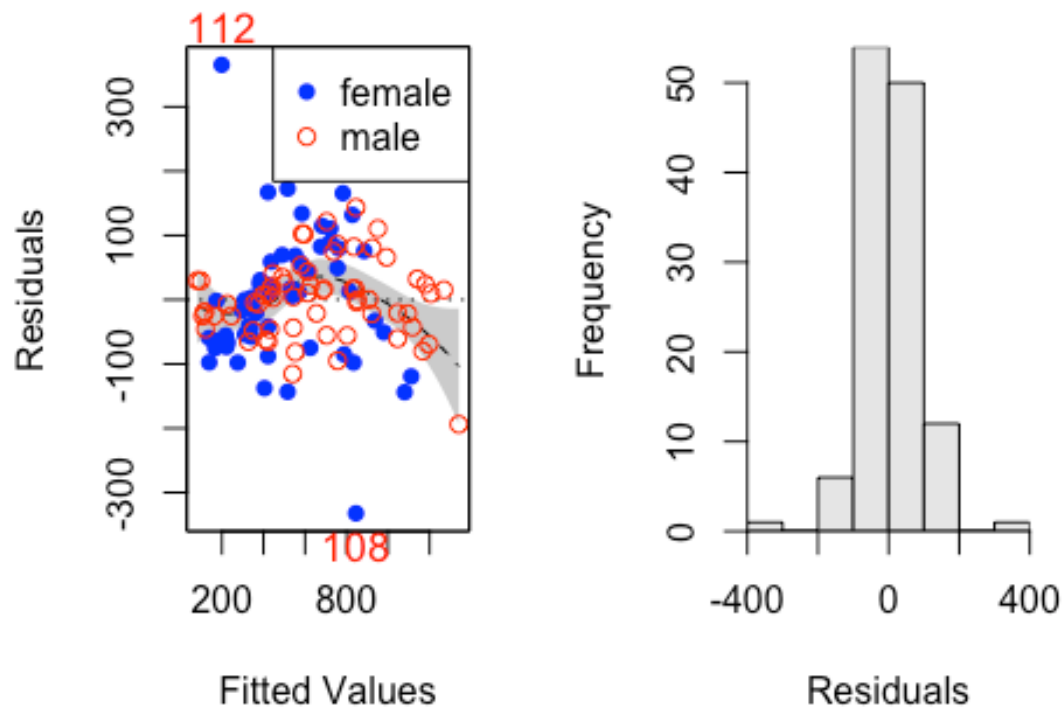
PG027 adults 2014


```
# based on visual inspection, there doesn't seem to be a large effect of sex on the size of the fish from either lake in the year 2014

# Find the residuals for the weight-length relationship by sex for each lake then perform an analysis of variance to see if the differences are significant

#For lake PG008
reg1 = lm(wt~fl*new_sex,data=pg08)
residPlot(reg1)
```
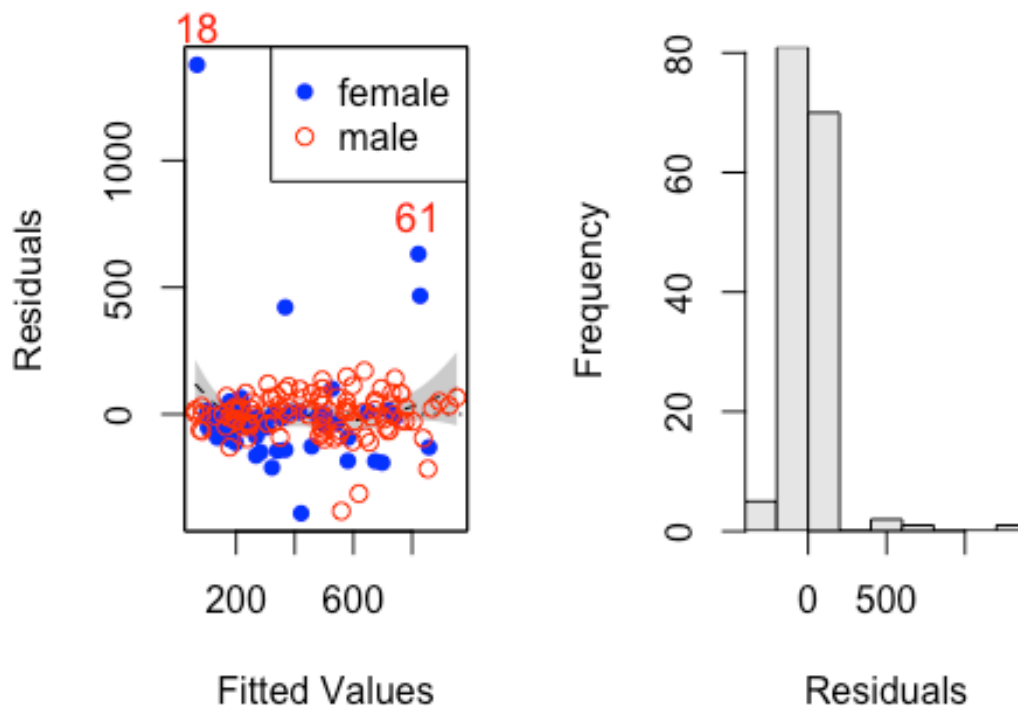
```
anova(reg1)

## Analysis of Variance Table
##
## Response: wt
##             Df    Sum Sq   Mean Sq   F value   Pr(>F)
## fl           1 11064828  11064828 1583.1286 < 2e-16 ***
## new_sex      1     1718      1718    0.2458 0.62096
## fl:new_sex   1    45545     45545    6.5164 0.01194 *
## Residuals  120   838706      6989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


#For Lake PG027

reg2 = lm(wt~fl*new_sex,data=pg027)
residPlot(reg2)
```

```
anova(reg2)

## Analysis of Variance Table
##
## Response: wt
##              Df  Sum Sq Mean Sq  F value Pr(>F)
## fl            1 8345574 8345574 330.2655 <2e-16 ***
## new_sex       1   50859   50859   2.0127 0.1580
## fl:new_sex    1   33092   33092   1.3096 0.2542
## Residuals   156 3942008   25269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# From the anova, we can see that fork length is a significant contributer to
the weight of the fish in both lakes. Sex does not have a significant effect o
n the size of adults in lake PG027 in the year 2014 but it did have a signific
ant affect on weight in lake PG008 when interacting with the fork length of th
e fish.

# Maybe 2014 was just a weird year for char. Let's scale it up to the whole da
taset to look at the effect of sex on size across years

char.adults = select(char,fl, wt, locWB, fyear, new_sex)
```
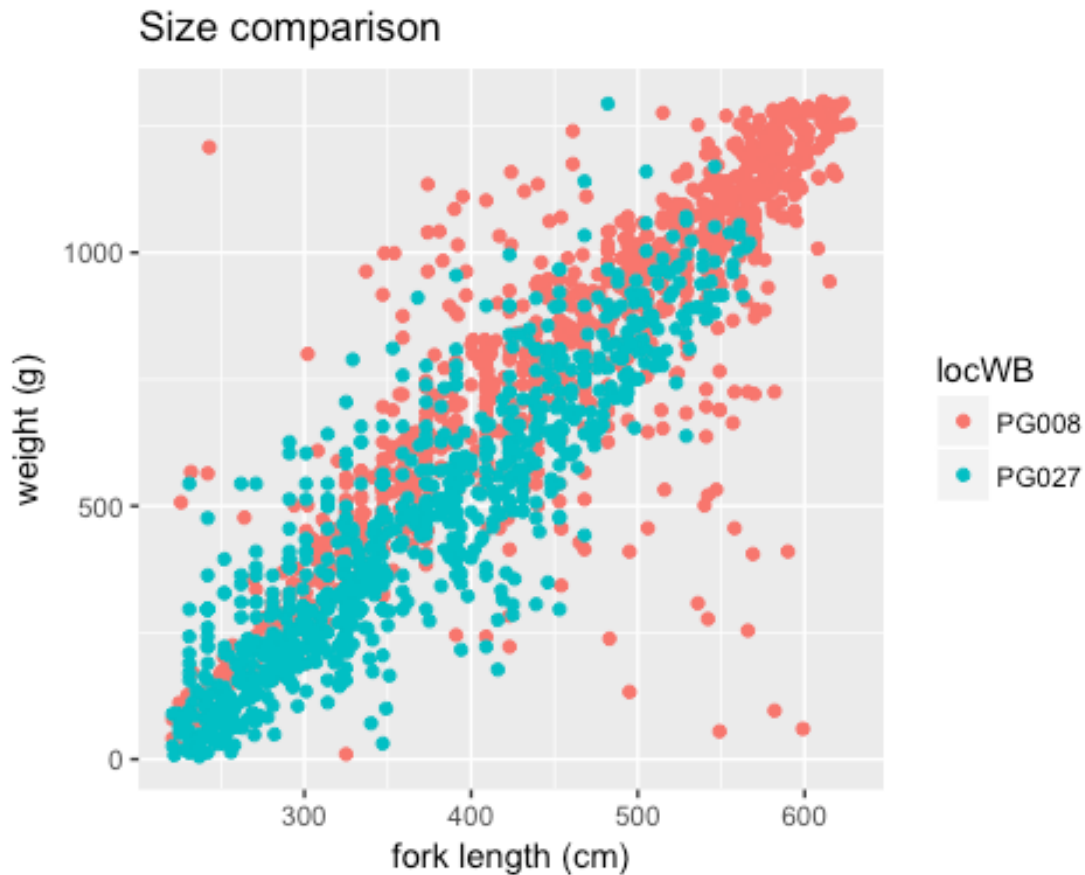
```
# restrict the dataset slightly to look at the 'normal' adults
char.adults = char.adults[char.adults$fl > 220,]
char.adults = char.adults[char.adults$wt < 1300,]

qplot(fl, wt,data = char.adults, colour = locWB, xlab = "fork length (cm)", yl
ab = "weight (g)", main = "Size comparison")
```
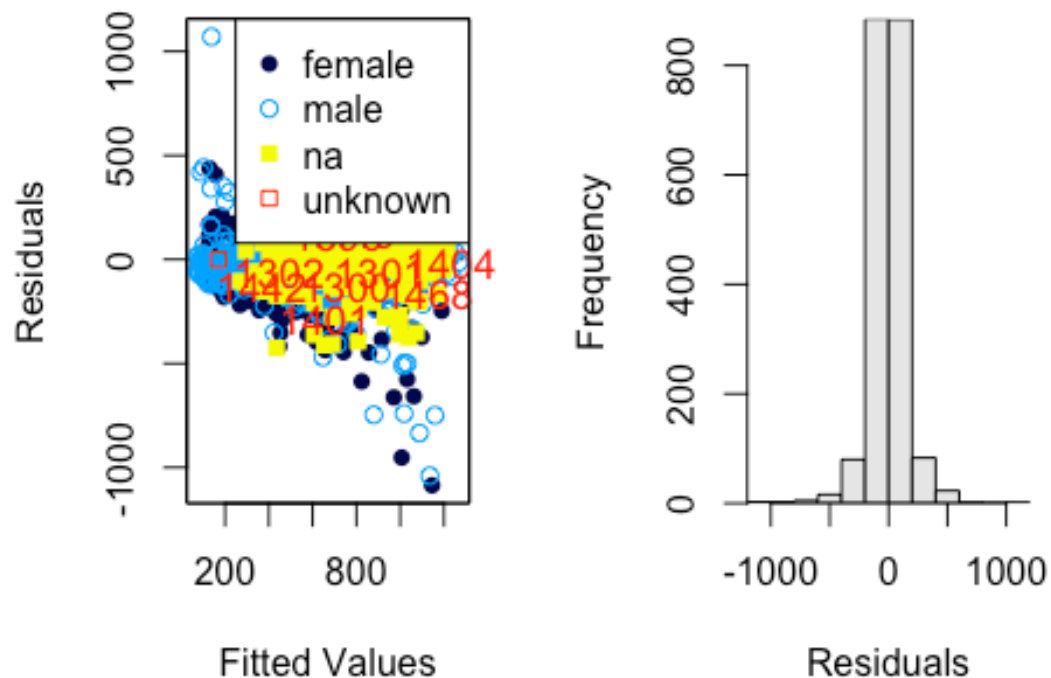


```
# Do the regression and analysis of variance
reg3 = lm(wt~fl*new_sex,data=char.adults)
residPlot(reg3)
```

```
anova(reg3)

## Analysis of Variance Table
##
## Response: wt
##                Df     Sum Sq    Mean Sq   F value      Pr(>F)
## fl              1 178552599  178552599 8231.1287 < 2.2e-16 ***
## new_sex         3      95248      31749    1.4636 0.2225870
## fl:new_sex      2     321864     160932    7.4188 0.0006167 ***
## Residuals    1975   42842409      21692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# The sex of the fish only has an effect on the weight of the fish when it is
interacting with the fork length. Like in the analysis for 2014, the fork leng
th has a significant effect on the fish's weight.

#make a plot to visualize the distribution of the sexes througout the differen
t sizes
plot(wt~fl,data=char.adults,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",y
lab="Weight (g)")
```
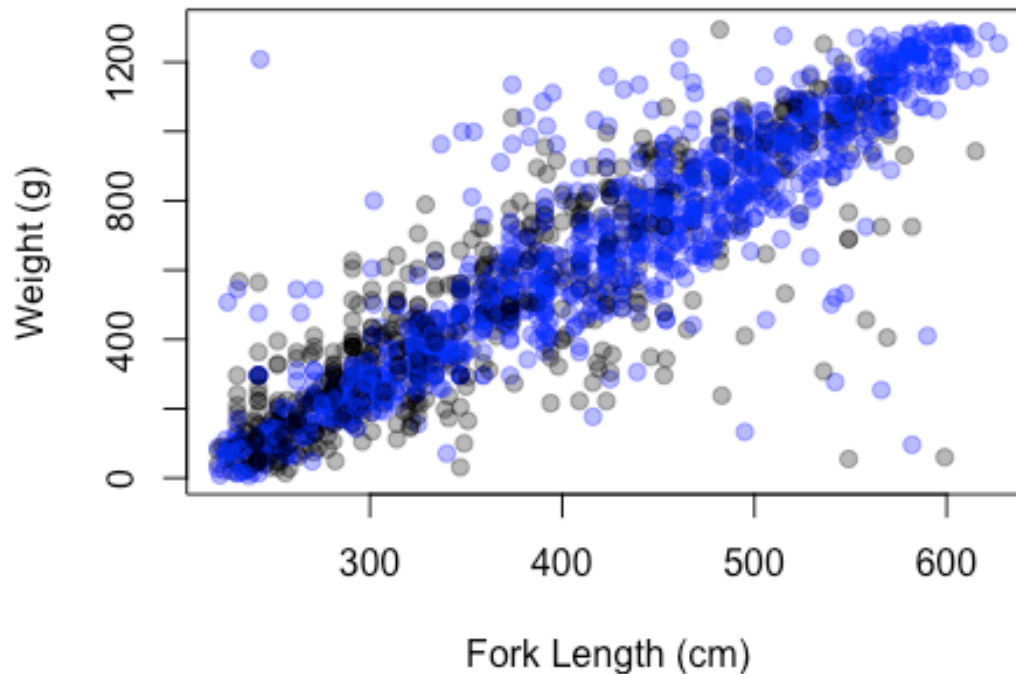
6. Revisiting the length-weight relationship with regression

Above (in #2) we performed a correlation test for the relationship between length and weight. We can also use a regression to model the linear relationship and to find out the R^2 value is for the weight-length relationship. Unlike a correlation, you can use a regression to make predictions about the data

```
# using the 'adults' dataset for 2014
# do the linear regression
lr1 = lm(wt~fl,data=pg08)
residPlot(lr1)
```
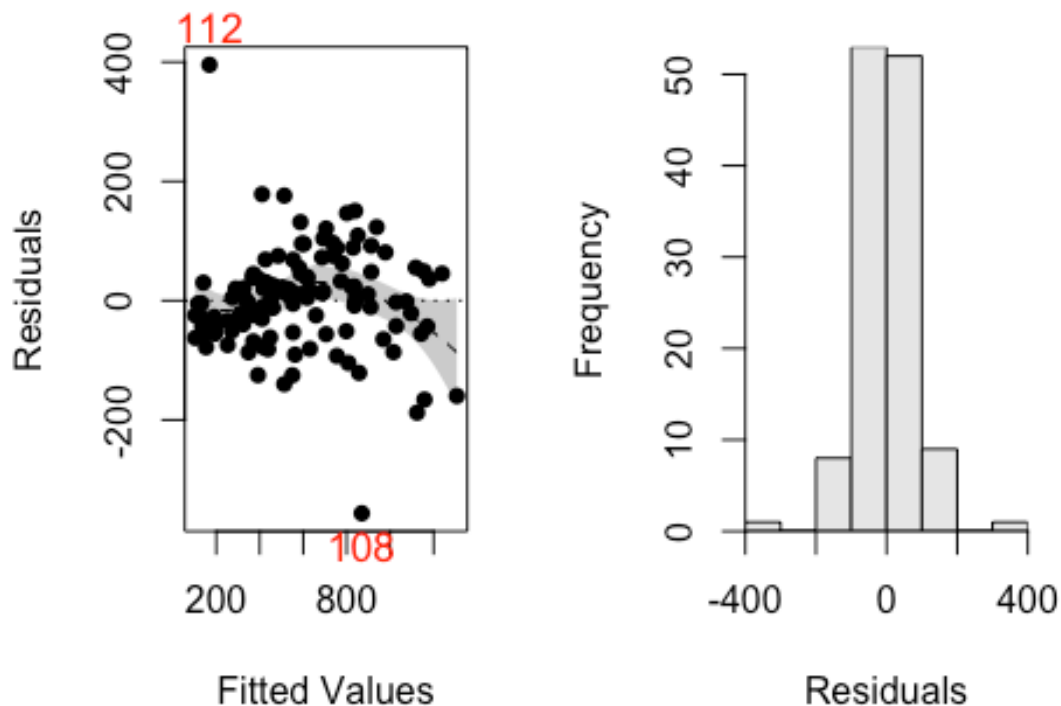
```
summary(lr1)

##
## Call:
## lm(formula = wt ~ fl, data = pg08)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -356.61  -51.15    0.25   42.26  395.24
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -581.69923   30.86426  -18.85   <2e-16 ***
## fl             3.10109    0.07945   39.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 85.22 on 122 degrees of freedom
## Multiple R-squared:  0.9259, Adjusted R-squared:  0.9253
## F-statistic:  1524 on 1 and 122 DF,  p-value: < 2.2e-16

# R^2 for PG008 is 0.9259. This suggests that length largely explains the vari
ation in weight in these fish. This value is similar to the value we found for
the correlation (0.962)
```
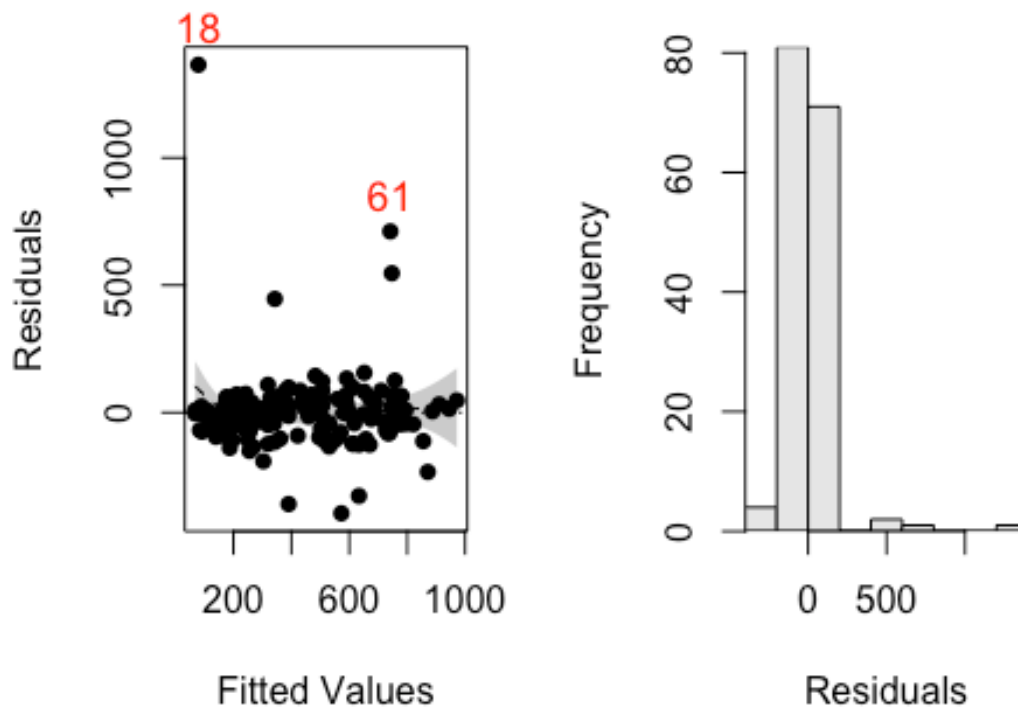
```r
# make a table of parameter estimates for the length- weight relationship
cbind(ests=coef(lr1),confint(lr1))
```

```
##                    ests        2.5 %      97.5 %
## (Intercept) -581.699227 -642.798104 -520.60035
## fl             3.101089    2.943818    3.25836
```

```r
# Repeat with adults from PG027
lr2 = lm(wt~fl,data=pg027)
residPlot(lr2)
```



```r
summary(lr2)
```

```
##
## Call:
## lm(formula = wt ~ fl, data = pg027)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.68  -62.75   -3.56   35.43 1365.31
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -526.258       54.153   -9.718     <2e-16 ***
## fl               2.642        0.146   18.098     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159.6 on 158 degrees of freedom
## Multiple R-squared:  0.6746, Adjusted R-squared:  0.6725
## F-statistic: 327.5 on 1 and 158 DF,  p-value: < 2.2e-16

# R^2 for PG027 is 0.6746. This is much lower than the fish from lake PG008. T
his value is quite different from the one we found for the correlation (0.821)
.
# In this example we can see how correlation and regression are quite differen
t and careful consideration should be taken when deciding which one to use. Se
e http://www.graphpad.com/support/faqid/1141/ for more info

# make a table of model parameters for the length-weight relationship. This ta
ble shows the 95% confidence intervals
cbind(ests=coef(lr2),confint(lr2))

##                      ests        2.5 %       97.5 %
## (Intercept) -526.257797 -633.214154 -419.301441
## fl             2.641678     2.353378    2.929979
```

7. The size-age relationship

Determine the body size of the fish as a function of their age

```
qplot(age, fl,data = char, colour = locWB, xlab = "age", ylab = "fork length (
cm)", main = "size at age")
```

size at age

```
# from this plot we can see how the age-length relationship is quite non-linea
r. To simplify things, let's restrict the age to under 20 because above that a
ge the relationship gets pretty strange and might be due to an ecologically un
ique morphotype of char

char.norm = filterD(char,!is.na(fl),!is.na(age), age < 20)
sum <- Summarize(fl~age,data=char.norm,digits=1) #getting the mean length per
each age

## Warning: RHS variable was converted to a factor.

plot(fl~age,data=char.norm,pch=19,col=col2rgbt(1,1/5),xlab="Age (yrs)",ylab="F
ork Length (cm)") # plot of the lengths at each age
lines(mean~fact2num(age),data=sum,lwd=2,col="red") #plot the line that shows t
he mean length at each age
```
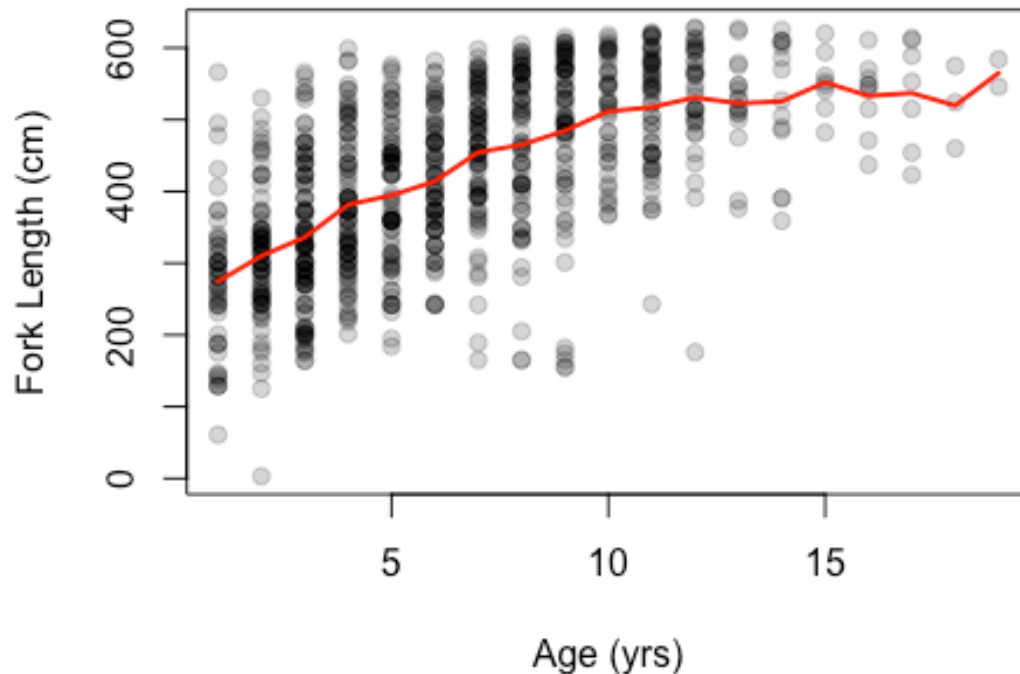
```
#because the relationship isn't linear, we will use a non-linear regression te
chnique

#Set up a function using the von Bertalanffy growth equation to evaluate the b
ody size of the fish as a function of age
vonb = vbFuns("Typical", msg = TRUE) # msg=true let's you see what the test is
doing and what the tems mean

## You have chosen the 'Typical', 'Traditional', or 'BevertonHolt' parameteriz
ation.
##
##    E[L|t] = Linf*(1-exp(-K*(t-t0)))
##
##    where Linf = asymptotic mean length
##          K = exponential rate of approach to Linf
##          t0 = the theoretical age when length = 0 (a modeling artifact)

# find the starting values for the test using a built in function in the FSA p
ackage
start.vb = vbStarts(fl~age,data=char.norm,type="Typical",plot=TRUE)
```
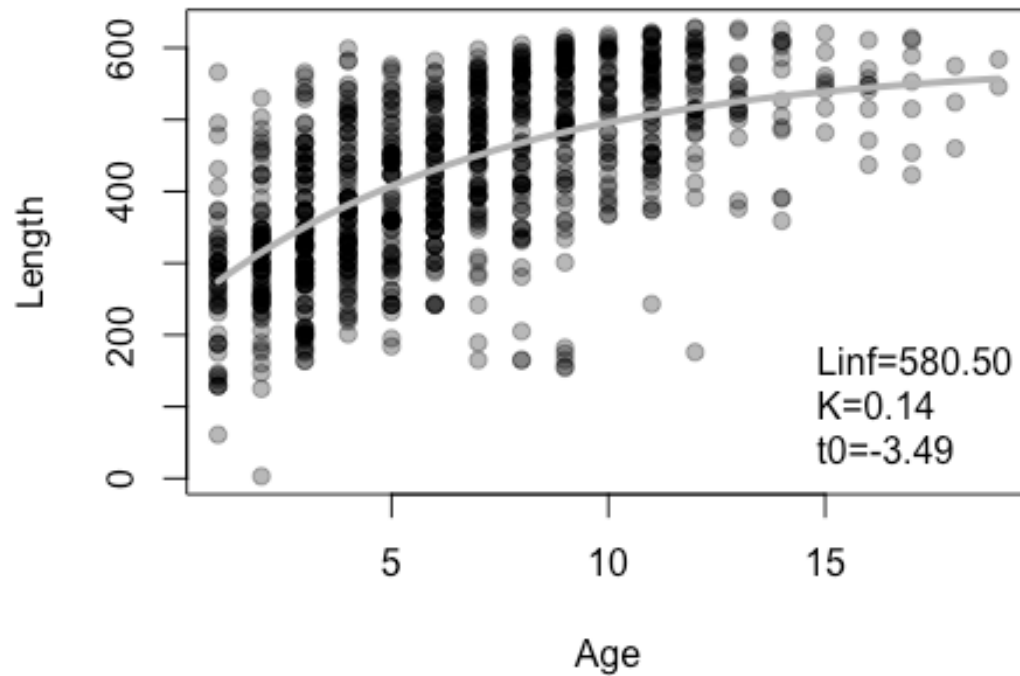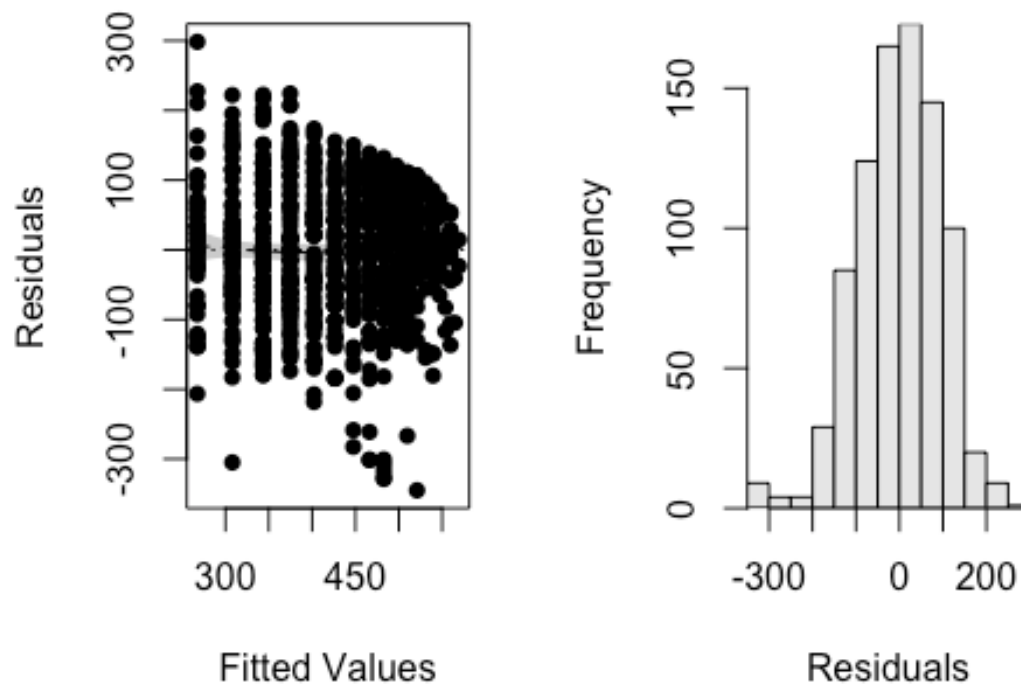
von B (Typical) STARTING VALUES

Linf=580.50
K=0.14
t0=-3.49

```r
# check that the residuals are normally distributed
fit1 = nls(fl~vonb(age,Linf,K,t0),data=char.norm,start=start.vb) # Taking the
relationship between the function just created and the length of the fish then
finding their nonlinear least squares.
residPlot(fit1)
```

```r
# get a summary
summary(fit1,correlation=TRUE)

##
## Formula: fl ~ vonb(age, Linf, K, t0)
##
## Parameters:
##        Estimate Std. Error t value Pr(>|t|)
## Linf 602.38092   29.83301  20.192  < 2e-16 ***
## K      0.12858    0.02281   5.638 2.33e-08 ***
## t0    -3.56713    0.69834  -5.108 4.01e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 94.75 on 865 degrees of freedom
##
## Correlation of Parameter Estimates:
##     Linf  K
## K   -0.96
## t0  -0.84  0.95
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 5.911e-06
```

```
# The correlation is high (0.96), meaning the age and length are strongly rela
ted for fish younger than 20 years old

cf = coef(fit1) # save the coefficients to add to the plot
plot(fl~age,data=char.norm,pch=19,col=col2rgbt(1,1/5),xlab="Age (yrs)",ylab="F
ork Length (cm)")
curve(vonb(x,cf), add = TRUE) #add = true adds the line to the existing plot
```
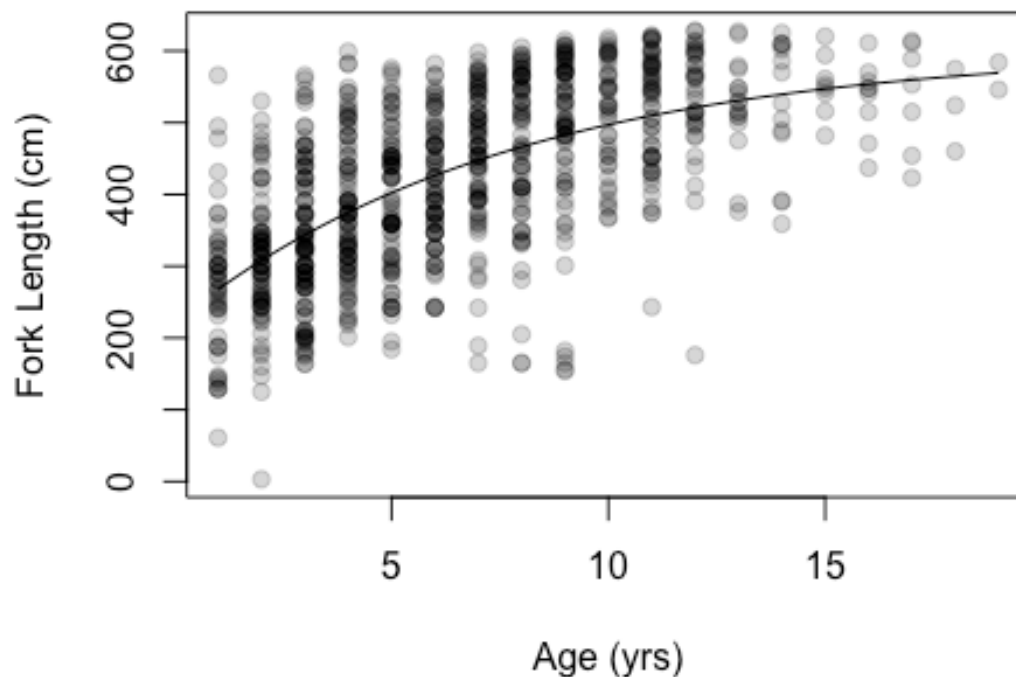

8. Mortality rates

Often with fisheries data, you want to see how quickly fish are dissapearing from a population to evaluate if a fishery is sustainable or not. Here we will be estimating annual mortality rates

```
# We will compare mortality rates in lake PG008 for the years 2007 and 2010 as
an example.
mort2007 = filterD(char,water.type == "freshwater",fyear=="2007", locWB == "PG
```
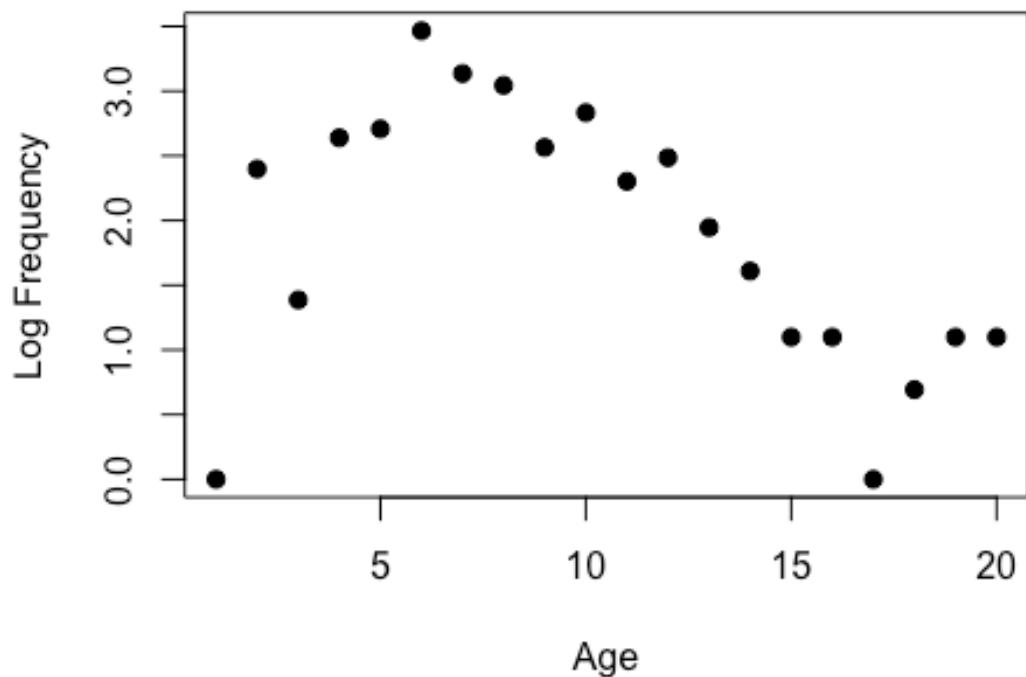
```
008", !is.na(fl),!is.na(age))

mort2007 = data.frame(xtabs(~age,data=mort2007)) # finding the frequency of ea
ch age
mort2007 = mutate(mort2007,age=as.numeric(age)) # make sure age is a numeric v
ariable

mort2007 = mutate(mort2007,logfreq=log(Freq)) # log transform the data to make
it more normal
plot(logfreq ~ age, data = mort2007, xlab="Age",ylab="Log Frequency",pch=19)
```



```
# visually inspect for the 'descending limb' to determin when the frequency of
ages strats to decrease. It looks to be around age 6 for lake PG008 in the yea
r 2007

cc1 = catchCurve(Freq~age, data = mort2007, ages2use = 6:20) # using ages 6-20
to determine the mortality rate
summary(cc1)

##    Estimate Std. Error  t value     Pr(>|t|)
## Z  0.211768 0.02497372 8.479635 1.175436e-06
## A 19.084762         NA       NA           NA
```
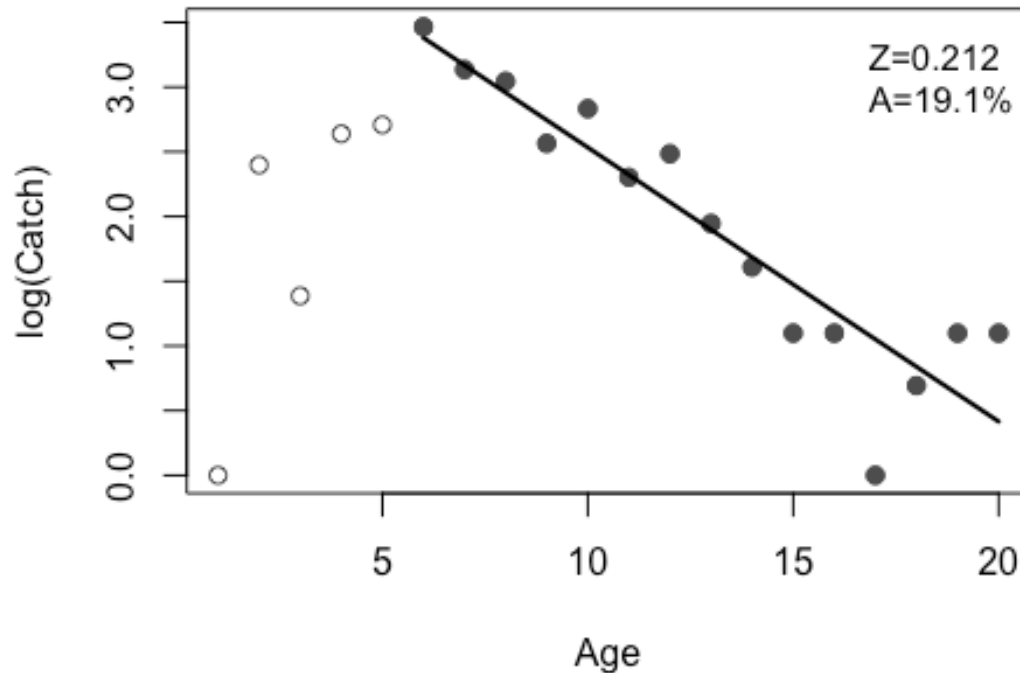
```
# summary shows the Z estimate which is the instantenous mortality rate
# It also shows the A estimate which is the annual mortality rate, which for 2
007 is 19%
plot(cc1)
```



```
# Do the same for lake PG008 in 2010
mort2010 = filterD(char,water.type == "freshwater",fyear=="2010", locWB == "PG
008", !is.na(fl),!is.na(age))
mort2010 = data.frame(xtabs(~age,data=mort2010))
mort2010 = mutate(mort2010,age=as.numeric(age))

mort2010 = mutate(mort2010,logfreq=log(Freq))
plot(logfreq ~ age, data = mort2010, xlab="Age",ylab="Log Frequency",pch=19)
```
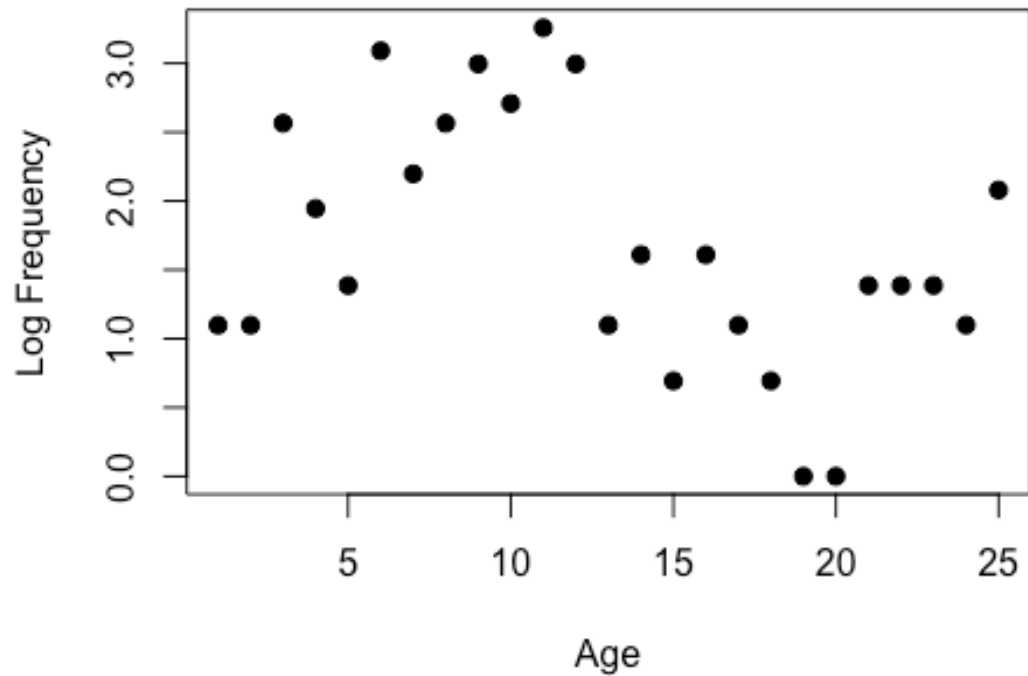
```
# visually inspect for the 'descending limb', looks to be around age 12

cc2 = catchCurve(Freq~age, data = mort2010, ages2use = 12:25)
summary(cc2)

##     Estimate Std. Error  t value  Pr(>|t|)
## Z 0.03114889 0.05287816 0.589069 0.5667394
## A 3.06687573         NA       NA        NA

# the instantenous mortality rate is 0.03
# the annual mortality rate is 3%
plot(cc2)
```
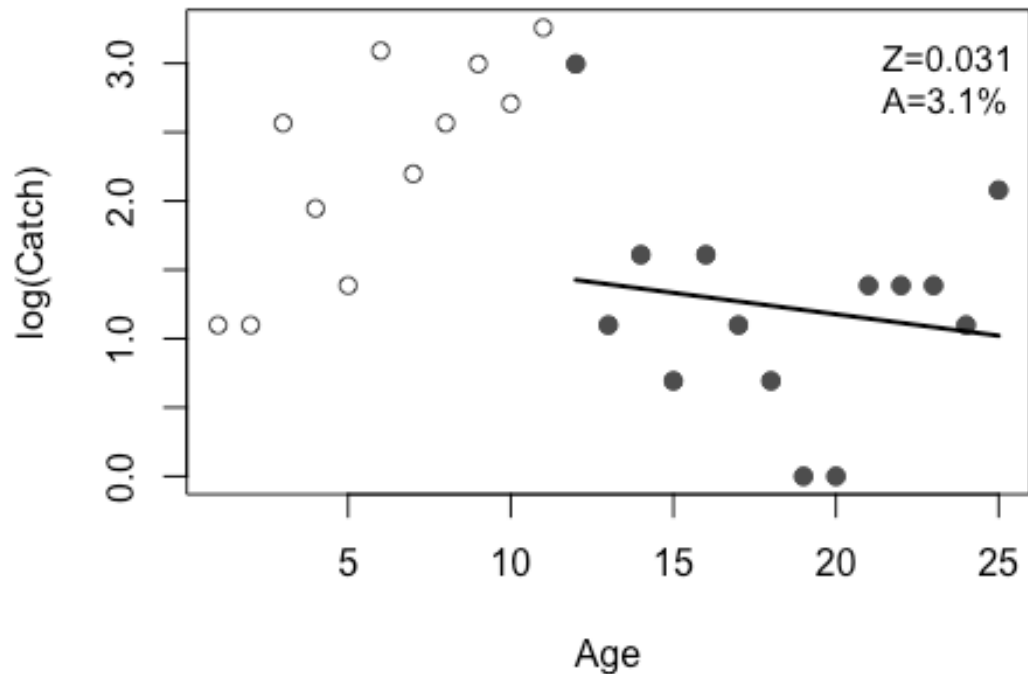
R Code

```r
library(dplyr)
library(FSA)
library(ggplot2)
```

```r
rm(list=ls(all=T))
char = read.csv("~/Desktop/R/Final project/CharData.csv")
```

```r
char = mutate(char,fyear=factor(year))
char = as.data.frame(char)
char = select(char,locWB, water.type, fl, wt, sex, mat, age, fyear)
char = mutate(char,new_sex=mapvalues(sex,from=c("M","F","U"),to=c("male","female","unknown")))
char = mutate(char, new_mat = mapvalues(mat, from = c(" M","I","M","na","R","RR","S", "U"), to = c("mature", "immature",
"mature", "na", "ripe", "running ripe", "spent", "unknown")))
char = transform(char, fl = as.numeric(fl), wt = as.numeric(wt), age = as.numeric(age))
```

```r
by.year = xtabs(~fyear + locWB,data = char)
by.year = as.data.frame(by.year)
g = ggplot(data = by.year) + geom_point(aes(x=fyear, y=locWB, size = Freq), color = 2) + theme_bw() +xlab("Year") +
ylab("Water body")
g + scale_size_continuous(range = c(1,15))
```

```
```
```{r}
char.2 = subset(char, new_mat != "na")
char.2 = subset(char.2, new_mat != "unknown")
char.2 = subset(char.2, new_sex != "na")
char.2 = subset(char.2, new_sex != "unknown")
char.2 = droplevels(char.2)
levels(char.2$new_sex) # you can see the unwanted levels are dropped
ftable(xtabs(~locWB+fyear+new_sex+new_mat,data = char.2))
```
```{r}
lw = filterD(char, fyear=="2014") # filter out only the year you're interested in
lw = select(lw,fl, wt, locWB, new_sex) # select the variables you want to work with
lw = lw[lw$fl != 0,] # omit observations with 0 length
len = data.frame(x = lw$fl, y = lw$locWB, na.rm = TRUE)
ggplot(len, aes(x=x, fill=y)) + geom_histogram(alpha=0.2, position="identity", binwidth = 10)
weight = data.frame(x = lw$wt, y = lw$locWB, na.rm = TRUE)
ggplot(weight, aes(x=x, fill=y)) + geom_histogram(alpha=0.2, position="identity", binwidth = 10)
qplot(fl, wt,data = lw, colour = locWB, xlab = "fork length (cm)", ylab = "weight (g)", main = "Size comparison")
adults = lw[lw$fl > 220,]
qplot(fl, wt,data = adults, colour = locWB, xlab = "fork length (cm)", ylab = "weight (g)", main = "Size comparison of adults")
pg08 = filterD(adults, locWB=="PG008")
pg027 = filterD(adults, locWB=="PG027")
with(pg08,cor(fl,wt))
with(pg027,cor(fl,wt))
```
```{r}
Summarize(wt~locWB,data=char)
ggplot(char, aes(x= fyear, y = wt, fill = locWB)) +
  geom_boxplot( ) +
  ggtitle("Fish weights")+
  xlab("Year")+
  ylab("Weight (g)")
hist(wt~locWB,data=char,same.ylim=FALSE)
char.log = mutate(char,logwt=log(wt), logfl=log(fl))
hist(logwt~locWB,data=char.log,same.ylim=FALSE)
anova(lm(formula = logwt ~ locWB * logfl, data = char.log))
```
```{r}
clr1 <- c("black","blue")
clr2 <- col2rgbt(clr1,1/3)
plot(wt~fl,data=pg08,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",ylab="Weight (g)", main = "PG008 adults 2014")
plot(wt~fl,data=pg027,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",ylab="Weight (g)", main = "PG027 adults 2014")
reg1 = lm(wt~fl*new_sex,data=pg08)
residPlot(reg1)
anova(reg1)
reg2 = lm(wt~fl*new_sex,data=pg027)
residPlot(reg2)
anova(reg2)
char.adults = select(char,fl, wt, locWB, fyear, new_sex)
char.adults = char.adults[char.adults$fl > 220,]
char.adults = char.adults[char.adults$wt < 1300,]
qplot(fl, wt,data = char.adults, colour = locWB, xlab = "fork length (cm)", ylab = "weight (g)", main = "Size comparison")
reg3 = lm(wt~fl*new_sex,data=char.adults)
residPlot(reg3)
anova(reg3)
plot(wt~fl,data=char.adults,pch=19,col=clr2[new_sex],xlab="Fork Length (cm)",ylab="Weight (g)")
```
```{r}
lr1 = lm(wt~fl,data=pg08)
residPlot(lr1)
```

```r
summary(lr1)
cbind(ests=coef(lr1),confint(lr1))
lr2 = lm(wt~fl,data=pg027)
residPlot(lr2)
summary(lr2)
cbind(ests=coef(lr2),confint(lr2))
```

```{r}
qplot(age, fl,data = char, colour = locWB, xlab = "age", ylab = "fork length (cm)", main = "size at age")
char.norm = filterD(char,!is.na(fl),!is.na(age), age < 20)
sum <- Summarize(fl~age,data=char.norm,digits=1)
plot(fl~age,data=char.norm,pch=19,col=col2rgbt(1,1/5),xlab="Age (yrs)",ylab="Fork Length (cm)")
lines(mean~fact2num(age),data=sum,lwd=2,col="red")
vonb = vbFuns("Typical", msg = TRUE)
start.vb = vbStarts(fl~age,data=char.norm,type="Typical",plot=TRUE)
fit1 = nls(fl~vonb(age,Linf,K,t0),data=char.norm,start=start.vb)
residPlot(fit1)
summary(fit1,correlation=TRUE)
cf = coef(fit1)
plot(fl~age,data=char.norm,pch=19,col=col2rgbt(1,1/5),xlab="Age (yrs)",ylab="Fork Length (cm)")
curve(vonb(x,cf), add = TRUE)
```

```{r}
mort2007 = filterD(char,water.type == "freshwater",fyear=="2007", locWB == "PG008", !is.na(fl),!is.na(age))
mort2007 = data.frame(xtabs(~age,data=mort2007))
mort2007 = mutate(mort2007,age=as.numeric(age))
mort2007 = mutate(mort2007,logfreq=log(Freq))
plot(logfreq ~ age, data = mort2007, xlab="Age",ylab="Log Frequency",pch=19)
cc1 = catchCurve(Freq~age, data = mort2007, ages2use = 6:20) #using ages 6-20 to determine the mortality rate
summary(cc1)
plot(cc1)
mort2010 = filterD(char,water.type == "freshwater",fyear=="2010", locWB == "PG008", !is.na(fl),!is.na(age))
mort2010 = data.frame(xtabs(~age,data=mort2010))
mort2010 = mutate(mort2010,age=as.numeric(age))
mort2010 = mutate(mort2010,logfreq=log(Freq))
plot(logfreq ~ age, data = mort2010, xlab="Age",ylab="Log Frequency",pch=19)
cc2 = catchCurve(Freq~age, data = mort2010, ages2use = 12:25)
summary(cc2)
plot(cc2)
```

```{r}
sessionInfo()
#R version 3.3.1 (2016-06-21)
#Platform: x86_64-apple-darwin13.4.0 (64-bit)
#Running under: OS X 10.9.5 (Mavericks)
#[1] en_CA.UTF-8/en_CA.UTF-8/en_CA.UTF-8/C/en_CA.UTF-8/en_CA.UTF-8
#attached base packages:
#[1] stats    graphics grDevices utils    datasets methods  base
#other attached packages:
#[1] ggplot2_2.2.1 FSA_0.8.8    dplyr_0.5.0
#loaded via a namespace (and not attached):
# [1] Rcpp_0.12.6    knitr_1.15.1    magrittr_1.5    splines_3.3.1    MASS_7.3-45
# [6] munsell_0.4.3    lattice_0.20-33    colorspace_1.2-6 R6_2.1.2    minqa_1.2.4
#[11] car_2.1-2    plyr_1.8.4    caTools_1.17.1    tools_3.3.1    parallel_3.3.1
#[16] nnet_7.3-12    pbkrtest_0.4-6    grid_3.3.1    nlme_3.1-128    gtable_0.2.0
#[21] mgcv_1.8-12    quantreg_5.26    plotrix_3.6-4    KernSmooth_2.23-15 DBI_0.4-1
#[26] MatrixModels_0.4-1 gtools_3.5.0    lme4_1.1-12    lazyeval_0.2.0    assertthat_0.1
#[31] digest_0.6.9    tibble_1.1    Matrix_1.2-6    nloptr_1.0.4    bitops_1.0-6
#[36] labeling_0.3    gdata_2.17.0    gplots_3.0.1    scales_0.4.1    SparseM_1.7
```