CPUE_modeling

Markus Min

March 04 2021

Description

This R markdown file will go through fitting a model to our CPUE data to isolate the effect of survey (year) on CPUE of ESA-listed rockfishes.

```
library(here)
library(tidyverse)
library(pscl)
```

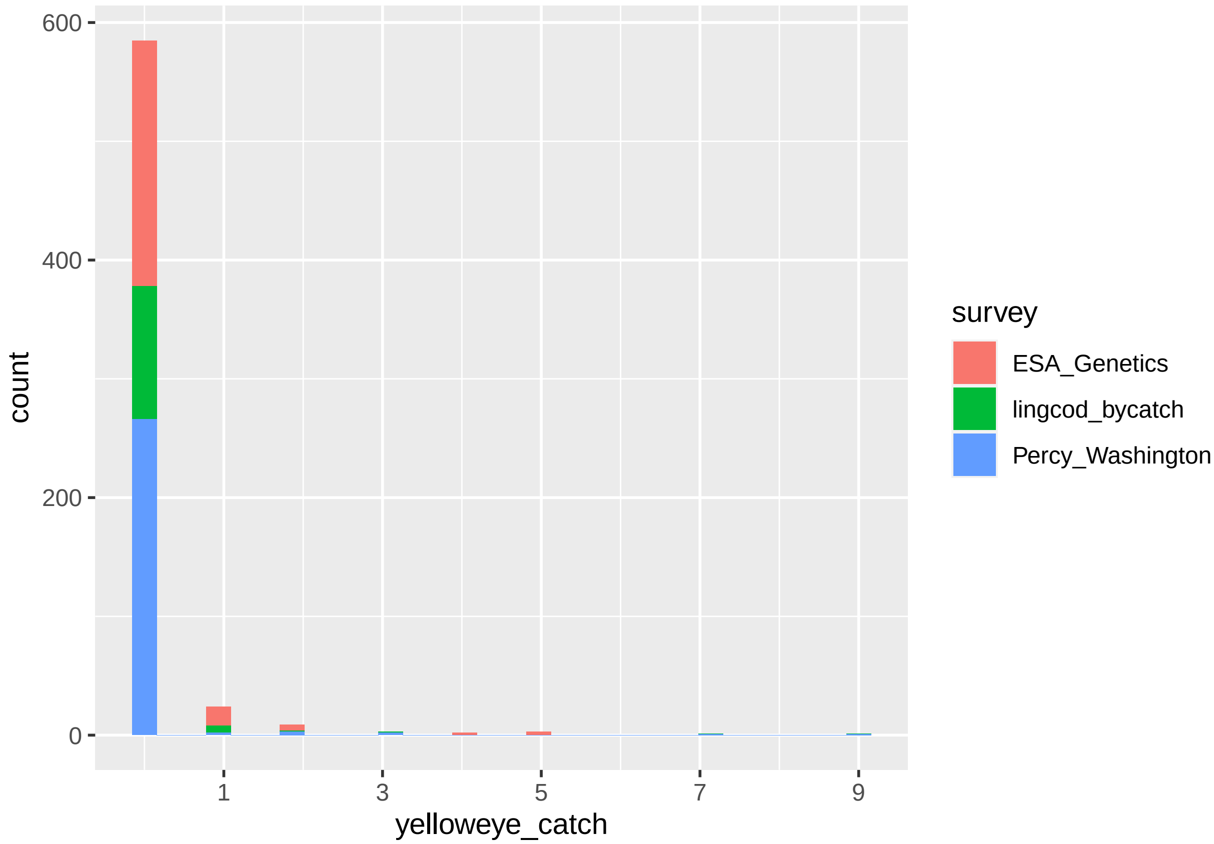
Load libraries

Import data

```
dat <- read.csv(here("hook_and_line_data", "CPUE_data_for_model.csv"), row.names = 1)
dat$month <- as.character(dat$month)
```

Plot data distributions

```
YE_depth_histogram <- ggplot(dat, aes(x = yelloweye_catch, fill = survey)) +
  geom_histogram()+
  scale_x_continuous(breaks = seq(1,10,2))

YE_depth_histogram
```

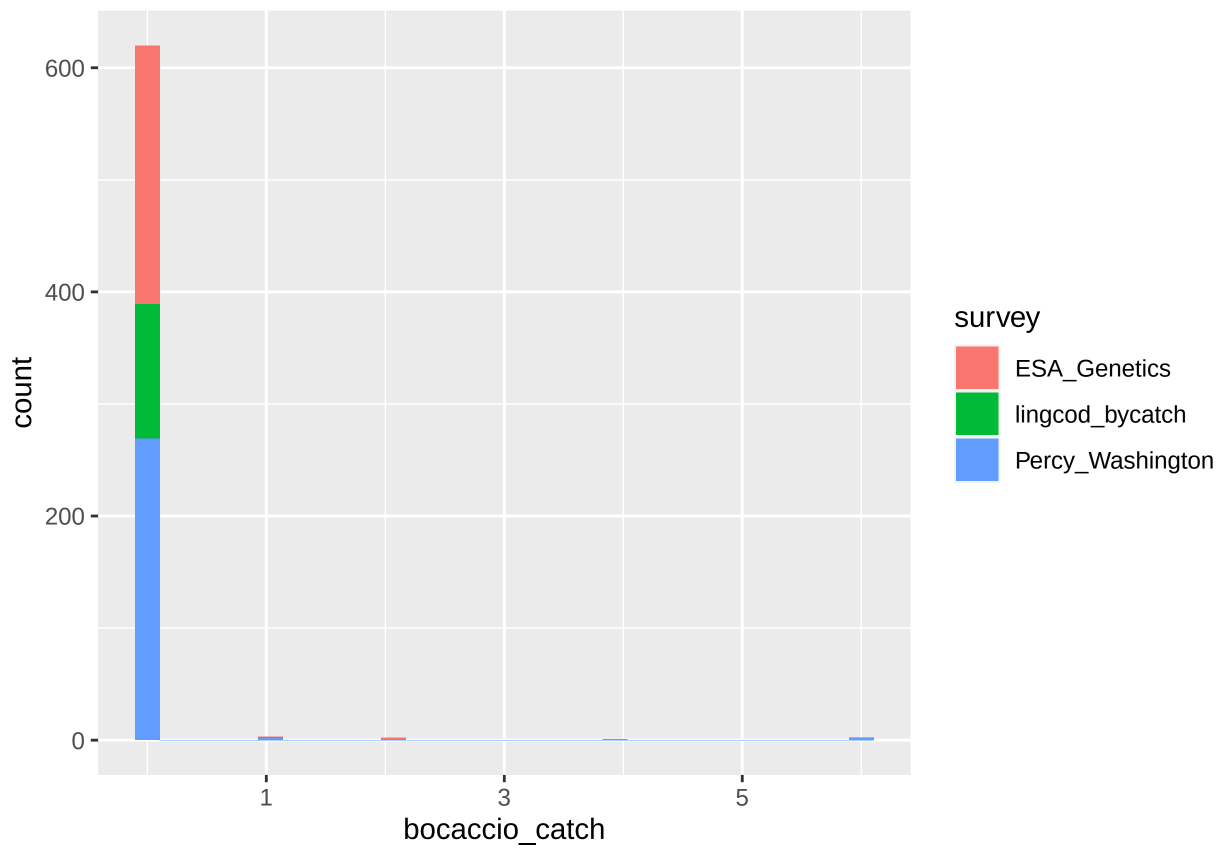


Catch

```
table(dat$yelloweye_catch)
```

```
##
##   0   1   2   3   4   5   7   9
## 585  24   9   3   2   3   1   1
```

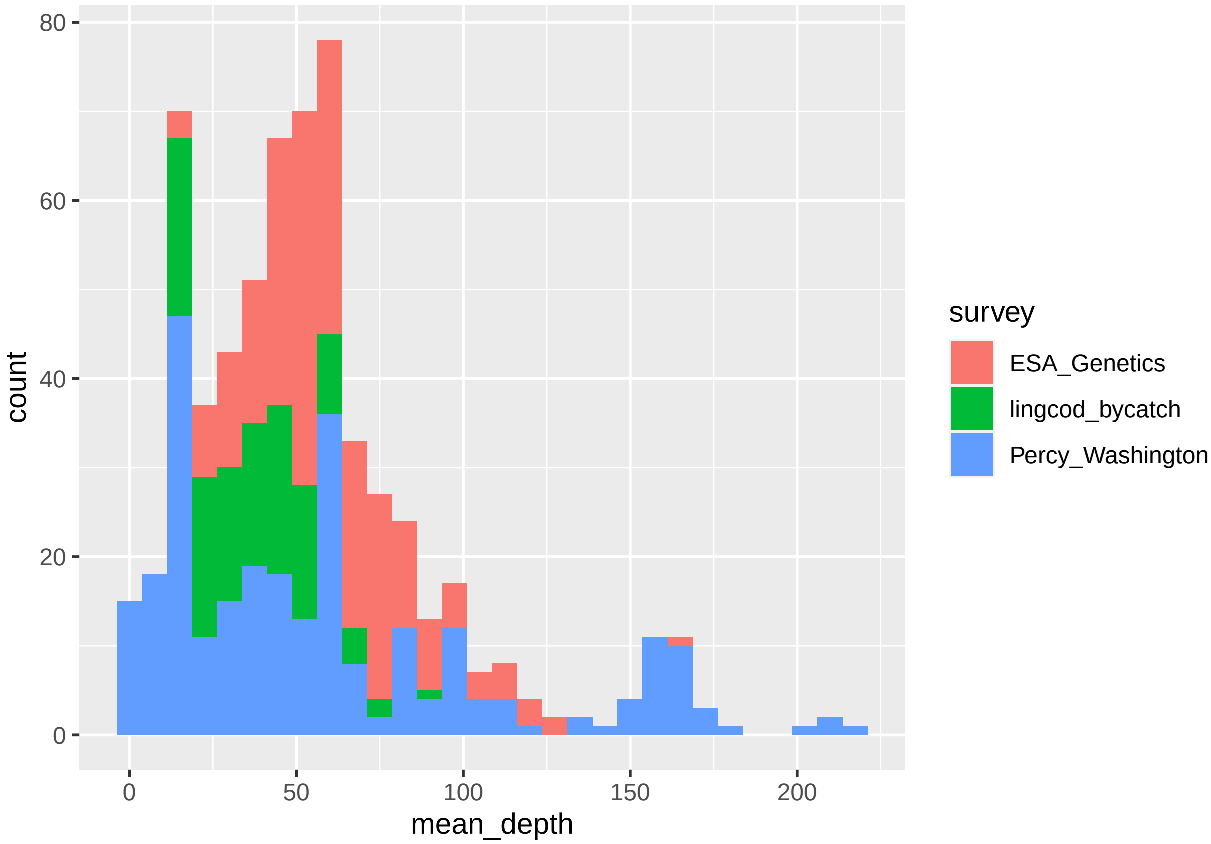
```
boc_depth_histogram <- ggplot(dat, aes(x = bocaccio_catch, fill = survey)) +
  geom_histogram()+
  scale_x_continuous(breaks = seq(1,10,2))

boc_depth_histogram
```




```r
table(dat$bocaccio_catch)
```

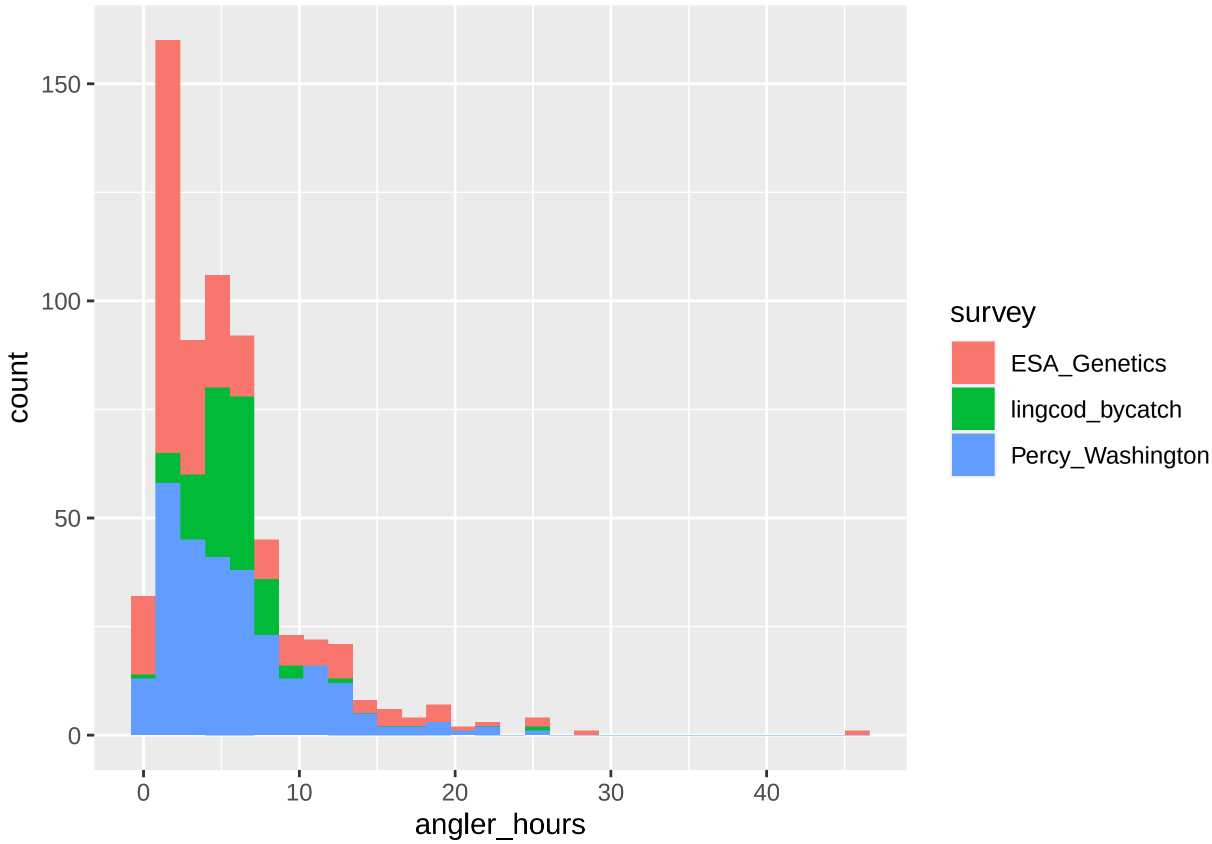
```
##
##   0   1   2   4   6
## 620   3   2   1   2
```

```r
depth_histogram <- ggplot(dat, aes(x = mean_depth, fill = survey)) +
  geom_histogram()

depth_histogram
```



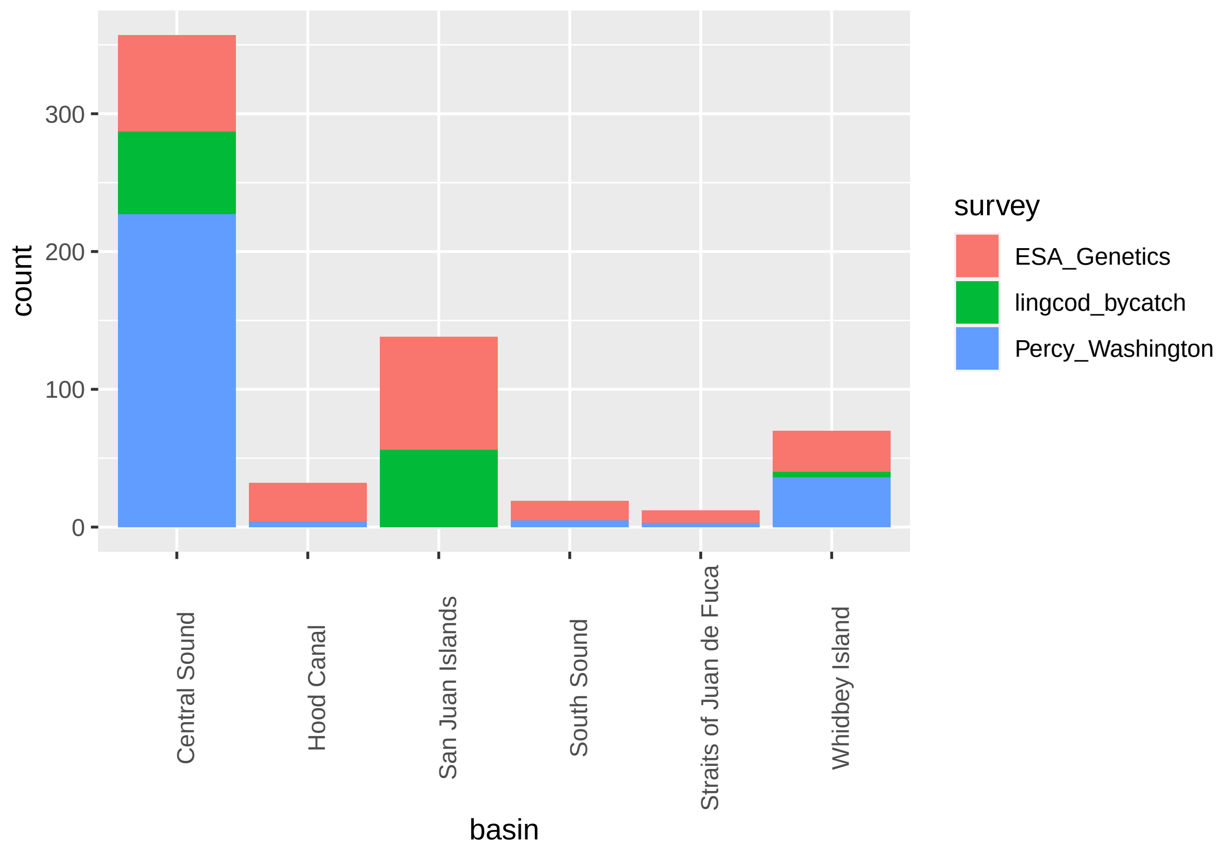
Depth

```
effort_histogram <- ggplot(dat, aes(x = angler_hours, fill = survey)) +
  geom_histogram()

effort_histogram
```



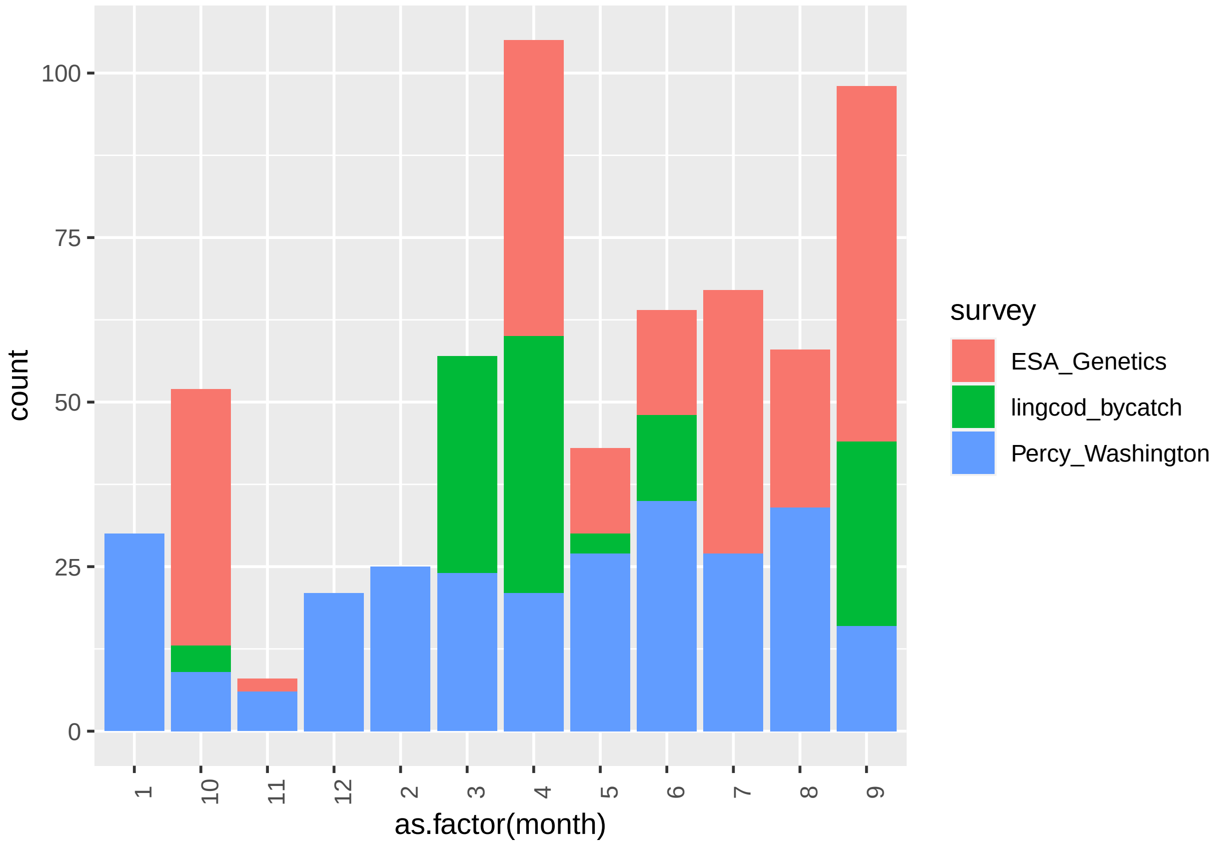
Effort

```
basin_histogram <- ggplot(dat, aes(x = basin, fill = survey)) +
  geom_histogram(stat = "count")+
  theme(axis.text.x = element_text(angle = 90))

basin_histogram
```



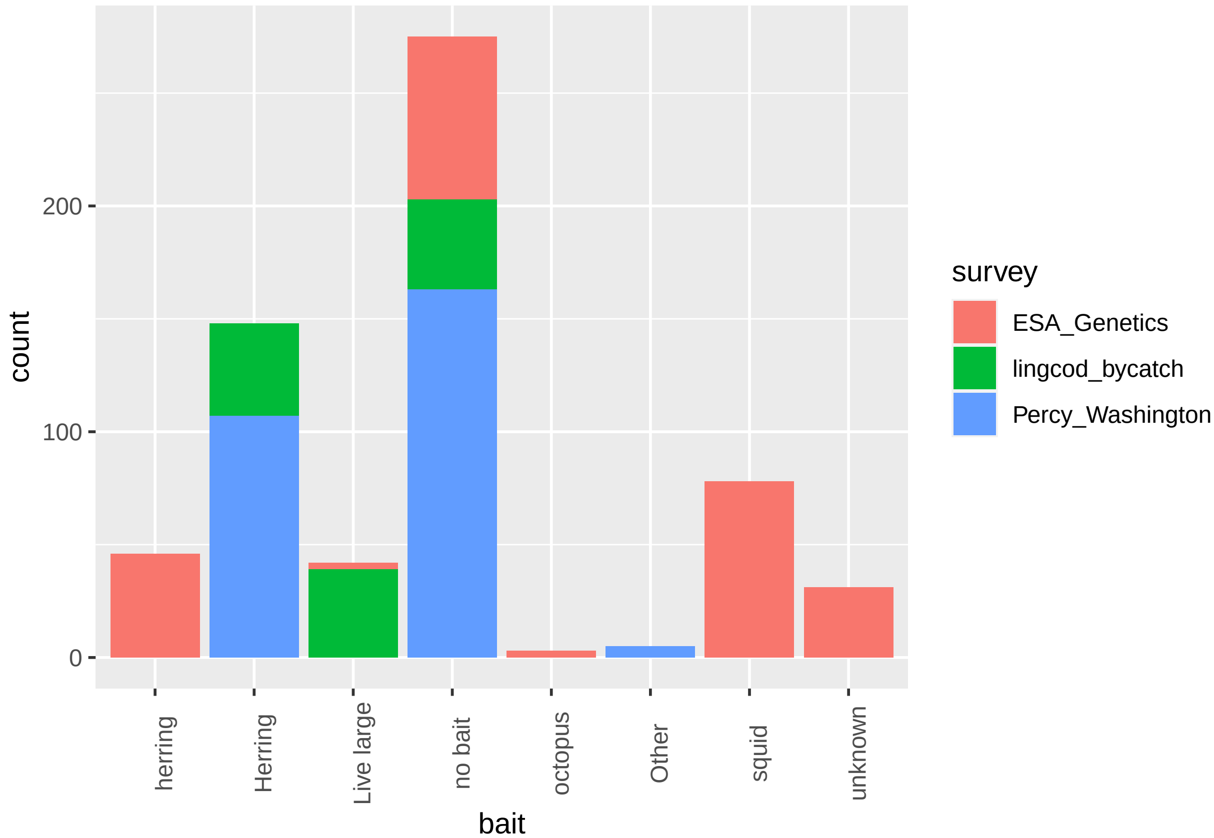
Basin

```
month_histogram <- ggplot(dat, aes(x = as.factor(month), fill = survey)) +
  geom_histogram(stat = "count")+
  theme(axis.text.x = element_text(angle = 90))

month_histogram
```



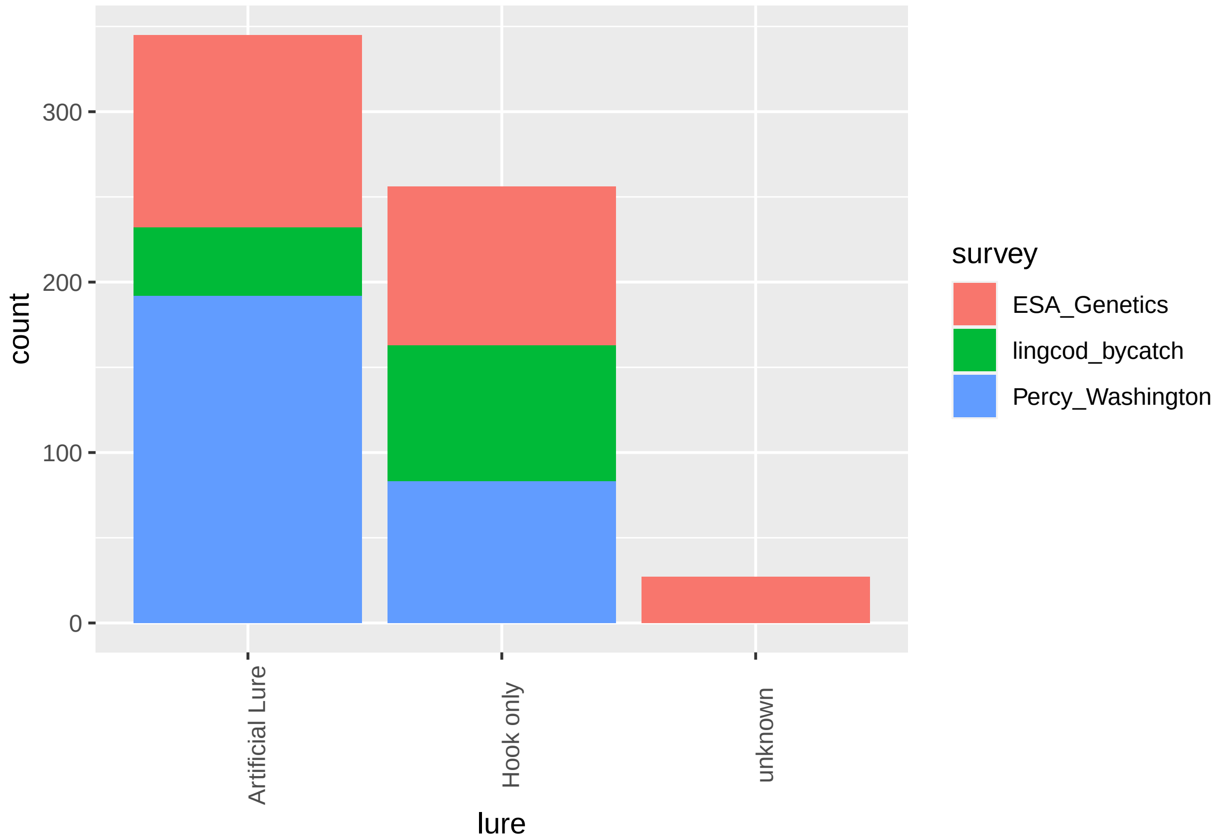
Month

```
bait_histogram <- ggplot(dat, aes(x = bait, fill = survey)) +
  geom_histogram(stat = "count")+
  theme(axis.text.x = element_text(angle = 90))

bait_histogram
```



Bait

```
lure_histogram <- ggplot(dat, aes(x = lure, fill = survey)) +
  geom_histogram(stat = "count")+
  theme(axis.text.x = element_text(angle = 90))

lure_histogram
```



Lure

Fit models for yelloweye

Fit model for yelloweye, all data

We will fit a GLM with ZIP distribution of catch, with fixed effects for depth, bait, lure, effort (angler_hours), month (seasonality), basin, and SURVEY.

For all models, bait, lure, and month don't work (`fitted probabilities numerically 0 or 1 occurredsystem is computationally singular`)

```
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + bait + lure + month + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + bait + lure + month + basin + survey + angler_hours, data = da
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning in value[[3L]](cond): system is computationally singular: reciprocal
## condition number = 3.10343e-26FALSE
```

```r
summary(yelloweye_zip)
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + bait + lure + month +
##     basin + survey + angler_hours | mean_depth + bait + lure + month +
##     basin + survey + angler_hours, data = dat)
##
## Pearson residuals:
##        Min         1Q     Median         3Q        Max
## -1.4067998 -0.2065385 -0.0888350 -0.0000884 10.8099225
##
## Count model coefficients (poisson with log link):
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -3.9707278         NA      NA       NA
## mean_depth                   0.0357615         NA      NA       NA
## baitHerring                 -0.7253178         NA      NA       NA
## baitLive large              -2.9663477         NA      NA       NA
## baitno bait                 -0.0218462         NA      NA       NA
## baitoctopus                 -0.0015955         NA      NA       NA
## baitOther                   -0.0019120         NA      NA       NA
## baitsquid                   -2.1462662         NA      NA       NA
## baitunknown                 -9.5333578         NA      NA       NA
## lureHook only               -0.0282247         NA      NA       NA
## lureunknown                 -9.5333578         NA      NA       NA
## month10                      0.5597921         NA      NA       NA
## month11                     -0.0020753         NA      NA       NA
## month12                     -0.3247762         NA      NA       NA
## month2                      -0.0114366         NA      NA       NA
## month3                       3.4467094         NA      NA       NA
## month4                       2.2610444         NA      NA       NA
## month5                      -1.5042564         NA      NA       NA
## month6                      -0.9449834         NA      NA       NA
## month7                       1.6842292         NA      NA       NA
## month8                      -0.4835493         NA      NA       NA
## month9                      -0.4258008         NA      NA       NA
## basinHood Canal              2.9216419         NA      NA       NA
## basinSan Juan Islands        2.0238117         NA      NA       NA
## basinSouth Sound            -0.0022871         NA      NA       NA
## basinStraits of Juan de Fuca -0.0001729         NA      NA       NA
## basinWhidbey Island          1.3247577         NA      NA       NA
## surveylingcod_bycatch       -2.1557684         NA      NA       NA
## surveyPercy_Washington      -0.6704206         NA      NA       NA
## angler_hours                 0.0655305         NA      NA       NA
```

```
##
## Zero-inflation model coefficients (binomial with logit link):
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -1.5035         NA      NA       NA
## mean_depth                      0.0154         NA      NA       NA
## baitHerring                    -0.9046         NA      NA       NA
## baitLive large                -17.5509         NA      NA       NA
## baitno bait                     2.9151         NA      NA       NA
## baitoctopus                    17.8110         NA      NA       NA
## baitOther                      15.2272         NA      NA       NA
## baitsquid                      -4.6907         NA      NA       NA
## baitunknown                    26.0748         NA      NA       NA
## lureHook only                   3.5003         NA      NA       NA
## lureunknown                    11.3032         NA      NA       NA
## month10                         2.5128         NA      NA       NA
## month11                        16.2806         NA      NA       NA
## month12                        -0.5708         NA      NA       NA
## month2                         17.7902         NA      NA       NA
## month3                          3.3936         NA      NA       NA
## month4                          4.7681         NA      NA       NA
## month5                         -1.8921         NA      NA       NA
## month6                        -38.3085         NA      NA       NA
## month7                          5.9332         NA      NA       NA
## month8                          0.7641         NA      NA       NA
## month9                          2.6517         NA      NA       NA
## basinHood Canal                -3.8142         NA      NA       NA
## basinSan Juan Islands          -2.4105         NA      NA       NA
## basinSouth Sound               15.6236         NA      NA       NA
## basinStraits of Juan de Fuca   15.9066         NA      NA       NA
## basinWhidbey Island            40.9559         NA      NA       NA
## surveylingcod_bycatch          -3.4208         NA      NA       NA
## surveyPercy_Washington         -1.3933         NA      NA       NA
## angler_hours                   -0.1823         NA      NA       NA
##
## Number of iterations in BFGS optimization: 116
## Log-likelihood: -147.3 on 60 Df
```

```r
AIC(yelloweye_zip)
```

```
## [1] 414.6191
```

```r
# Remove bait/lure
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + month + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + month + basin + survey + angler_hours, data = dat)
```

```
## Warning in value[[3L]](cond): system is computationally singular: reciprocal
## condition number = 6.99648e-23FALSE
```

```r
summary(yelloweye_zip)
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + month + basin + survey +
##     angler_hours | mean_depth + month + basin + survey + angler_hours,
##     data = dat)
##
## Pearson residuals:
##       Min        1Q    Median        3Q       Max
## -1.049713 -0.194151 -0.065523 -0.001742 10.527410
##
## Count model coefficients (poisson with log link):
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -3.5225553         NA      NA       NA
## mean_depth                   0.0281582         NA      NA       NA
## month10                      0.5466703         NA      NA       NA
## month11                     -0.0051841         NA      NA       NA
## month12                     -0.5237367         NA      NA       NA
## month2                      -0.0002556         NA      NA       NA
## month3                       3.1830175         NA      NA       NA
## month4                       0.8773874         NA      NA       NA
## month5                       0.4022520         NA      NA       NA
## month6                      -2.2501889         NA      NA       NA
## month7                       0.5205359         NA      NA       NA
## month8                      -0.5736331         NA      NA       NA
## month9                      -0.8313129         NA      NA       NA
## basinHood Canal              3.7399594         NA      NA       NA
## basinSan Juan Islands        1.5335437         NA      NA       NA
## basinSouth Sound            -0.0010838         NA      NA       NA
## basinStraits of Juan de Fuca -6.2353183         NA      NA       NA
## basinWhidbey Island          1.5533854         NA      NA       NA
## surveylingcod_bycatch       -1.1195005         NA      NA       NA
## surveyPercy_Washington       0.5788491         NA      NA       NA
## angler_hours                -0.0014098         NA      NA       NA
##
## Zero-inflation model coefficients (binomial with logit link):
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.776154         NA      NA       NA
## mean_depth                   -0.003741         NA      NA       NA
## month10                       3.808211         NA      NA       NA
## month11                      15.747847         NA      NA       NA
## month12                      -1.539098         NA      NA       NA
## month2                       15.989394         NA      NA       NA
```

```
## month3                        3.920630         NA       NA        NA
## month4                        2.249187         NA       NA        NA
## month5                        5.593413         NA       NA        NA
## month6                      -43.155566         NA       NA        NA
## month7                        4.797606         NA       NA        NA
## month8                        1.168228         NA       NA        NA
## month9                        2.740570         NA       NA        NA
## basinHood Canal              -0.825466         NA       NA        NA
## basinSan Juan Islands        -2.206585         NA       NA        NA
## basinSouth Sound             14.896836         NA       NA        NA
## basinStraits of Juan de Fuca  23.298445        NA       NA        NA
## basinWhidbey Island          42.413452         NA       NA        NA
## surveylingcod_bycatch         0.695372         NA       NA        NA
## surveyPercy_Washington        4.012607         NA       NA        NA
## angler_hours                 -0.593134         NA       NA        NA
##
## Number of iterations in BFGS optimization: 77
## Log-likelihood:  -151 on 42 Df
```

```r
AIC(yelloweye_zip)
```

```
## [1] 385.9627
```

```r
# Keep only survey, depth, and angler hours
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + survey + angler_hours, data = dat)

summary(yelloweye_zip)
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + survey + angler_hours |
##     mean_depth + survey + angler_hours, data = dat)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -0.7351 -0.2706 -0.2108 -0.1114  9.1357
##
## Count model coefficients (poisson with log link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.1394869  0.3964022  -0.352   0.7249
## mean_depth              0.0006386  0.0038152   0.167   0.8671
## surveylingcod_bycatch  -0.5174032  0.5618301  -0.921   0.3571
## surveyPercy_Washington  0.6427322  0.3375424   1.904   0.0569 .
```

```
## angler_hours            0.0575530  0.0276866   2.079   0.0376 *
##
## Zero-inflation model coefficients (binomial with logit link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              2.835624   0.468938   6.047 1.48e-09 ***
## mean_depth              -0.016152   0.005086  -3.176  0.00149 **
## surveylingcod_bycatch  -0.115882   0.624207  -0.186  0.85272
## surveyPercy_Washington  1.956934   0.484724   4.037 5.41e-05 ***
## angler_hours            -0.029884   0.028126  -1.062  0.28801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 17
## Log-likelihood: -203.5 on 10 Df
```

```r
AIC(yelloweye_zip)
```

```
## [1] 426.9106
```

```r
# Keep only survey, depth, basin, and angler hours
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth  + basin + survey + angler_hours, data = dat)

summary(yelloweye_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + basin + survey + angler_hours |
##     mean_depth + basin + survey + angler_hours, data = dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -0.93903 -0.28604 -0.12342 -0.08084  9.39493
##
## Count model coefficients (poisson with log link):
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -7.786e-01  7.008e-01  -1.111   0.2665
## mean_depth                     3.886e-03  4.896e-03   0.794   0.4274
## basinHood Canal                8.791e-01  4.599e-01   1.911   0.0559 .
## basinSan Juan Islands          2.109e-01  4.988e-01   0.423   0.6725
## basinSouth Sound              -1.165e-05         NA      NA       NA
```

```
## basinStraits of Juan de Fuca -4.547e-06         NA      NA       NA
## basinWhidbey Island          -1.203e+00  1.803e+00  -0.667   0.5046
## surveylingcod_bycatch        -3.658e-01  6.295e-01  -0.581   0.5611
## surveyPercy_Washington        6.268e-01  5.104e-01   1.228   0.2195
## angler_hours                  7.277e-02  3.047e-02   2.388   0.0169 *
##
## Zero-inflation model coefficients (binomial with logit link):
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   4.181e+00  8.246e-01   5.070 3.98e-07 ***
## mean_depth                   -1.914e-02  6.062e-03  -3.157  0.00160 **
## basinHood Canal              -1.847e+00  7.377e-01  -2.503  0.01230 *
## basinSan Juan Islands        -1.972e+00  6.439e-01  -3.063  0.00219 **
## basinSouth Sound              1.437e+01  8.173e+02   0.018  0.98597
## basinStraits of Juan de Fuca  1.480e+01  1.507e+03   0.010  0.99216
## basinWhidbey Island           6.905e-02  1.496e+00   0.046  0.96318
## surveylingcod_bycatch         2.751e-02  7.739e-01   0.036  0.97164
## surveyPercy_Washington        1.009e+00  6.703e-01   1.505  0.13222
## angler_hours                 -6.118e-02  3.247e-02  -1.884  0.05951 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 32
## Log-likelihood: -184.1 on 20 Df
```

```r
AIC(yelloweye_zip)
```

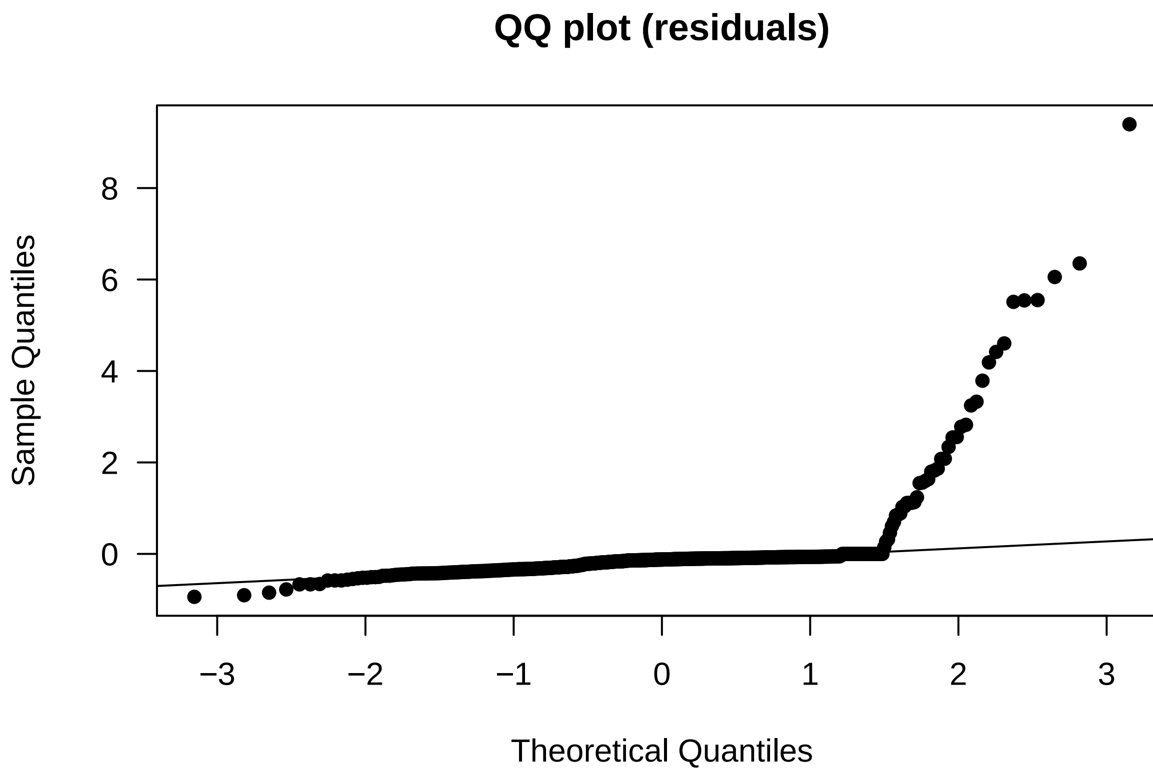
```
## [1] 408.2275
```

So with ZIP models (or other zero-inflated models), there are two processes being modeled: the probability of detection (binomial) and the count process (Poisson). As we see from the model outputs, each of these has its own significance values and estimates of effect size etc.

For the best model (`zeroinfl(formula = yelloweye_catch ~ mean_depth + basin + survey + angler_hours | mean_depth + basin + survey + angler_hours, data = dat)`), angler_hours is the only significant variable for counts, whereas mean depth and basin (Hood Canal or San Juan Islands) are significant in the detection process.

Diagnostic plots for best model

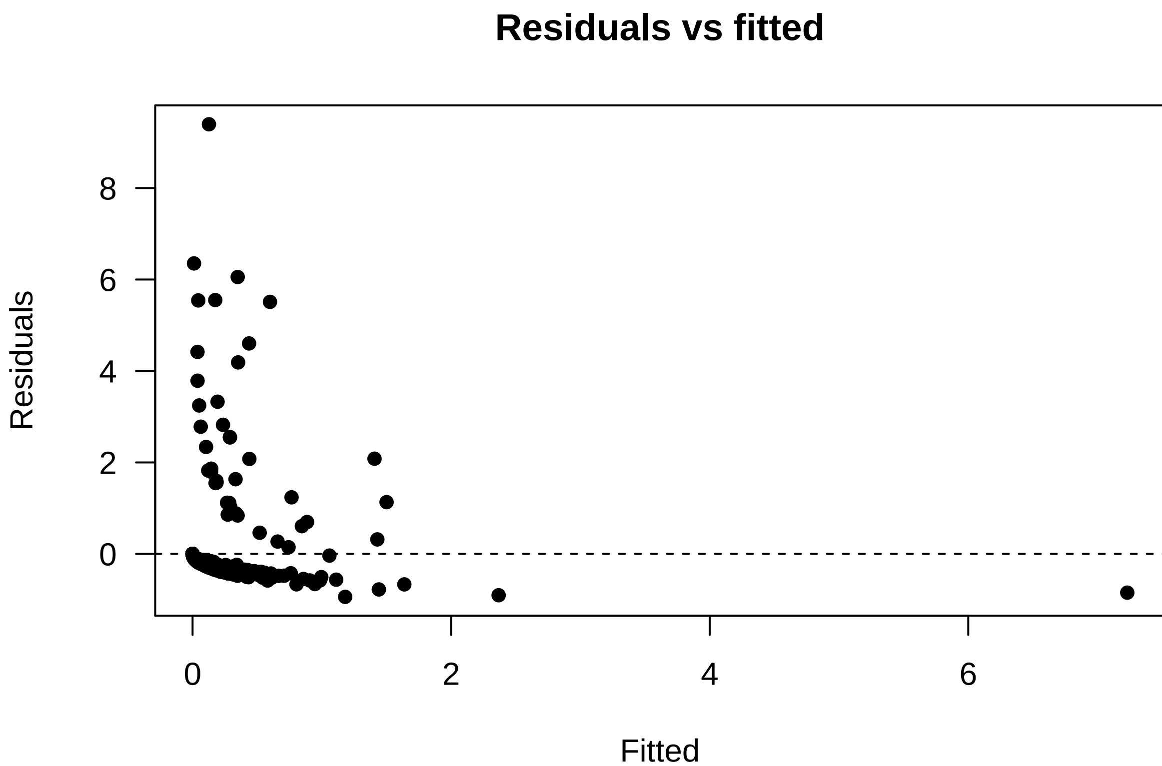
```r
## qq resids
qqnorm(residuals(yelloweye_zip), main = "QQ plot (residuals)", las = 1, pch = 16)
qqline(residuals(yelloweye_zip))
```



QQ plot of residuals

Not great!

```
plot(fitted(yelloweye_zip), residuals(yelloweye_zip), las = 1, pch = 16,
     xlab = "Fitted", ylab = "Residuals",
     main = "Residuals vs fitted")
abline(h=0, lty = "dashed")
```



Residuals vs. fitted

Again, not great!

```
## Type-I error
alpha_crit <- 0.05

nsamples <- length(residuals(yelloweye_zip))
nvariables <- 3
## threshold value for rho (correlation)
(rho_crit <- qnorm(1 - alpha_crit/2) / sqrt(nsamples))
```

ACF

```
## [1] 0.07865065
```

```
## rearrange residuals into matrix
rr <- matrix(residuals(yelloweye_zip), nsamples, nvariables)
```

```r
## get ACF
ACF <- apply(rr, 2, acf, lag.max = 5, plot = FALSE)
ACF <- lapply(ACF, function(x) x$acf)
## convert list to matrix; don't need row 1 b/c rho_0 = 1
ACF <- do.call(cbind, ACF)[-1,]

## check if any |values| > rho_crit
any(abs(ACF) > rho_crit)
```

```
## [1] FALSE
```

I'm not sure I did this right, but looks like no autocorrelation.

Fit model for yelloweye, only PSP

We will fit a GLM with ZIP distribution of catch, with fixed effects for depth, bait, lure, effort (angler__hours), month (seasonality), basin, and SURVEY.

```r
unique(dat$basin)
```

```
## [1] "Central Sound"           "Whidbey Island"
## [3] "South Sound"             "Straits of Juan de Fuca"
## [5] "Hood Canal"              "San Juan Islands"
```

```r
PSP_dat <- subset(dat, basin %in% c("Central Sound",  "Whidbey Island", "South Sound"))
# yelloweye_zip <- zeroinfl(yelloweye_catch ~
#                              # Predictors of counts
#                              mean_depth + bait + lure + month + basin + survey + angler_hours |
#                              # Predictors of detection
#                              mean_depth + bait + lure + month + basin + survey + angler_hours, data = 
#
#
#
# summary(yelloweye_zip)
# AIC(yelloweye_zip)

# Remove bait/lure
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + month + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + month + basin + survey + angler_hours, data = PSP_dat)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning in value[[3L]](cond): Lapack routine dgesv: system is exactly singular:
## U[31,31] = 0FALSE
```

```r
summary(yelloweye_zip)
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + month + basin + survey +
##     angler_hours | mean_depth + month + basin + survey + angler_hours,
##     data = PSP_dat)
##
## Pearson residuals:
##        Min         1Q     Median         3Q        Max
## -7.843e-01 -1.489e-08 -1.192e-08 -8.975e-14  1.575e+01
##
## Count model coefficients (poisson with log link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -6.27356         NA      NA       NA
## mean_depth                0.02656         NA      NA       NA
## month10                   0.93402         NA      NA       NA
## month11                 -21.81314         NA      NA       NA
## month12                  -0.40886         NA      NA       NA
## month2                  -32.15774         NA      NA       NA
## month3                    3.54295         NA      NA       NA
## month4                    3.98949         NA      NA       NA
## month5                    4.81324         NA      NA       NA
## month6                   -0.95739         NA      NA       NA
## month7                 -183.24710         NA      NA       NA
## month8                   -1.33595         NA      NA       NA
## month9                 -109.26451         NA      NA       NA
## basinSouth Sound        -39.98127         NA      NA       NA
## basinWhidbey Island       3.70035         NA      NA       NA
## surveylingcod_bycatch     0.50413         NA      NA       NA
## surveyPercy_Washington    3.78063         NA      NA       NA
## angler_hours             -0.03370         NA      NA       NA
##
## Zero-inflation model coefficients (binomial with logit link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -355.1127         NA      NA       NA
## mean_depth                0.7176         NA      NA       NA
## month10                  67.8434         NA      NA       NA
## month11                  22.4584         NA      NA       NA
## month12                -140.9277         NA      NA       NA
## month2                   27.3928         NA      NA       NA
## month3                  533.5563         NA      NA       NA
## month4                  603.0181         NA      NA       NA
## month5                 1218.6737         NA      NA       NA
```

```
## month6                  -688.3161         NA      NA       NA
## month7                    56.9670         NA      NA       NA
## month8                    42.7446         NA      NA       NA
## month9                    62.8064         NA      NA       NA
## basinSouth Sound          20.1751         NA      NA       NA
## basinWhidbey Island     1153.2434         NA      NA       NA
## surveylingcod_bycatch    242.6995         NA      NA       NA
## surveyPercy_Washington   767.9516         NA      NA       NA
## angler_hours             -62.6257         NA      NA       NA
##
## Number of iterations in BFGS optimization: 110
## Log-likelihood: -35.7 on 36 Df
```

```r
AIC(yelloweye_zip)
```

```
## [1] 143.4051
```

```r
# Keep only survey, depth, basin, and angler hours
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth  + basin + survey + angler_hours, data = PSP_dat)

summary(yelloweye_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + basin + survey + angler_hours |
##     mean_depth + basin + survey + angler_hours, data = PSP_dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -0.80278 -0.17081 -0.10557 -0.05099  6.81244
##
## Count model coefficients (poisson with log link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -3.90477    0.78002  -5.006 5.56e-07 ***
## mean_depth              0.00249    0.00401   0.621    0.535
## basinSouth Sound       -3.60485         NA      NA       NA
## basinWhidbey Island     1.23238    1.15457   1.067    0.286
## surveylingcod_bycatch   1.14796    1.32986   0.863    0.388
## surveyPercy_Washington  4.17012    0.65711   6.346 2.21e-10 ***
## angler_hours            0.06334    0.03188   1.987    0.047 *
```

```
##
## Zero-inflation model coefficients (binomial with logit link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -6.705429  43.088722  -0.156  0.87633
## mean_depth              -0.018962   0.006342  -2.990  0.00279 **
## basinSouth Sound        18.996934         NA      NA       NA
## basinWhidbey Island      9.439691  43.003151   0.220  0.82625
## surveylingcod_bycatch    9.637957  43.128113   0.223  0.82317
## surveyPercy_Washington 13.148521  43.106577   0.305  0.76035
## angler_hours            -0.216516   0.078749  -2.749  0.00597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 79
## Log-likelihood: -70.61 on 14 Df
```

```r
AIC(yelloweye_zip)
```

```
## [1] 169.2284
```

```r
# Keep only survey, depth, and angler hours
yelloweye_zip <- zeroinfl(yelloweye_catch ~
                            # Predictors of counts
                            mean_depth + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + survey + angler_hours, data = PSP_dat)

summary(yelloweye_zip)
```

```
##
## Call:
## zeroinfl(formula = yelloweye_catch ~ mean_depth + survey + angler_hours |
##     mean_depth + survey + angler_hours, data = PSP_dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -0.57660 -0.16496 -0.11898 -0.08856  8.65663
##
## Count model coefficients (poisson with log link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -3.020721   1.230687  -2.455   0.0141 *
## mean_depth               0.002946   0.004627   0.637   0.5243
## surveylingcod_bycatch   -0.132380   1.545764  -0.086   0.9318
## surveyPercy_Washington   1.759598   0.889473   1.978   0.0479 *
## angler_hours             0.182919   0.078422   2.332   0.0197 *
##
## Zero-inflation model coefficients (binomial with logit link):
```

```
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              2.748316   1.341070   2.049  0.04043 *
## mean_depth              -0.017700   0.006779  -2.611  0.00902 **
## surveylingcod_bycatch  -0.814094   1.740295  -0.468  0.63993
## surveyPercy_Washington  1.729796   0.978502   1.768  0.07709 .
## angler_hours             0.002482   0.058463   0.042  0.96613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 42
## Log-likelihood: -74.47 on 10 Df
```

```r
AIC(yelloweye_zip)
```

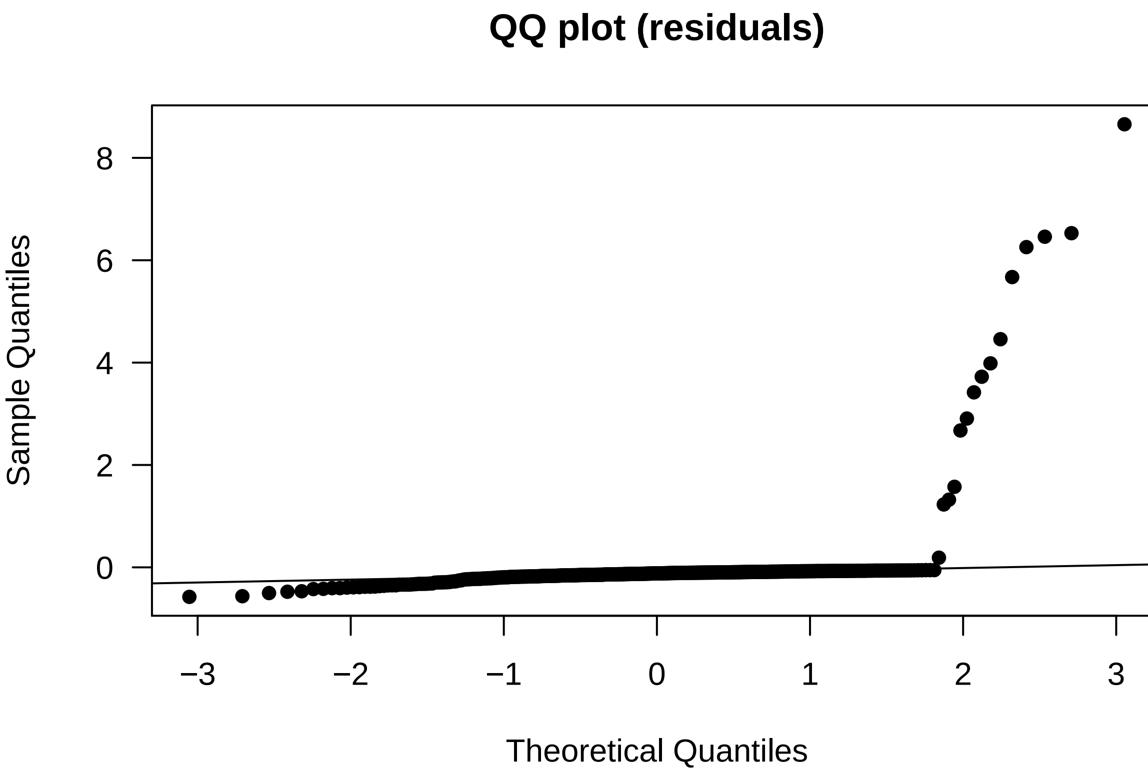
```
## [1] 168.9327
```

Here, the best model by AIC is `zeroinfl(formula = yelloweye_catch ~ mean_depth + survey + angler_hours | mean_depth + survey + angler_hours, data = PSP_dat)`.

In this model, mean depth is the only significant predictor of presence, whereas angler hours and survey (Percy Washington) are significant for counts.

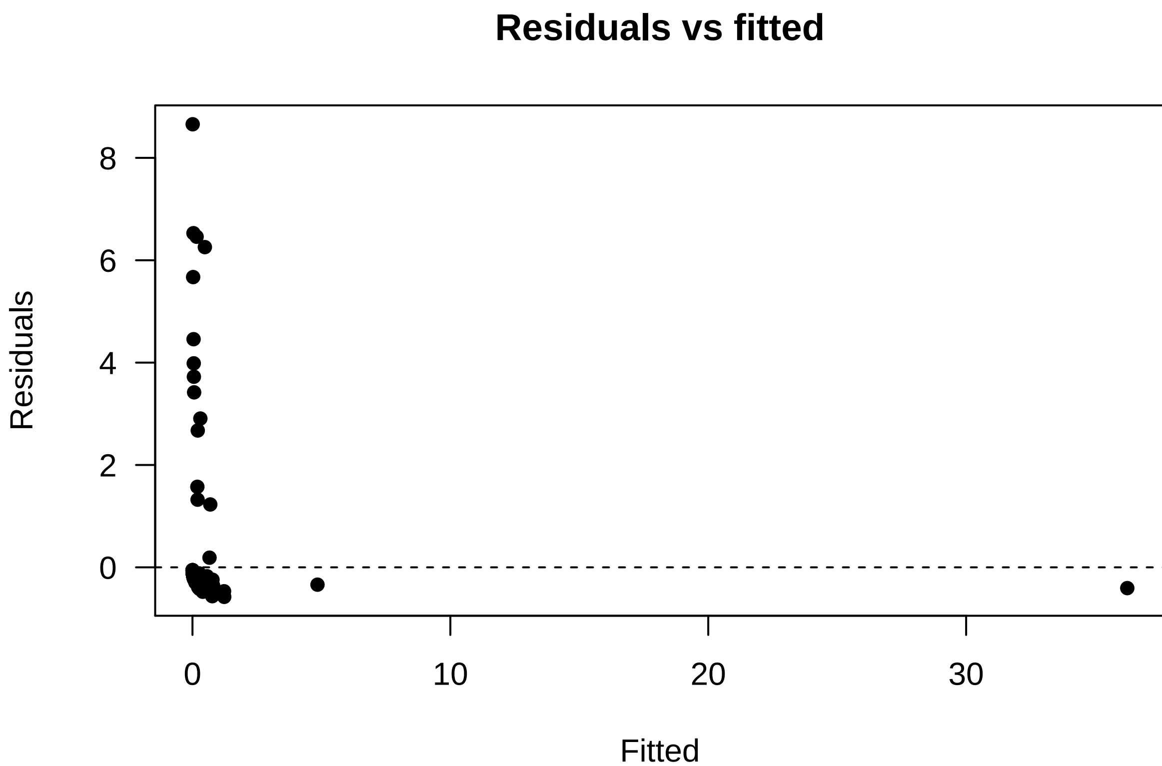
Diagnostic plots for best model

```r
## qq resids
qqnorm(residuals(yelloweye_zip), main = "QQ plot (residuals)", las = 1, pch = 16)
qqline(residuals(yelloweye_zip))
```



QQ plot of residuals
Not great!

```
plot(fitted(yelloweye_zip), residuals(yelloweye_zip), las = 1, pch = 16,
     xlab = "Fitted", ylab = "Residuals",
     main = "Residuals vs fitted")
abline(h=0, lty = "dashed")
```



Residuals vs. fitted
Again, not great!

```
## Type-I error
alpha_crit <- 0.05

nsamples <- length(residuals(yelloweye_zip))
nvariables <- 3
## threshold value for rho (correlation)
(rho_crit <- qnorm(1 - alpha_crit/2) / sqrt(nsamples))
```

ACF

```
## [1] 0.0931207
```

```
## rearrange residuals into matrix
rr <- matrix(residuals(yelloweye_zip), nsamples, nvariables)
```

```
## get ACF
ACF <- apply(rr, 2, acf, lag.max = 5, plot = FALSE)
ACF <- lapply(ACF, function(x) x$acf)
## convert list to matrix; don't need row 1 b/c rho_0 = 1
ACF <- do.call(cbind, ACF)[-1,]

## check if any |values| > rho_crit
any(abs(ACF) > rho_crit)
```

```
## [1] FALSE
```

Plot estimated probability of non-detection and expected count as a function of explanatory variables How do we modify this code for multiple explanatory variables? I guess we could model each separately?

As function of mean depth Well, I couldn't quite get this to run. Complicated to select only the relevant elements for a single explanatory variable...

```
# Remove NAs from dat
dat_noNA <- subset(dat, !is.na(mean_depth) & !is.na(basin))

# Sample size
nn <- dim(dat_noNA)[1]

## fitted for detection prob (pi)
mean_depth <- sort(dat$mean_depth)
gamma_hat_0 <- coef(yelloweye_zip)[2]
gamma_hat_1 <- coef(yelloweye_zip)[12]
pi_hat <- 1/(1+exp(-(gamma_hat_0 + gamma_hat_1 * mean_depth)))

## matrix of derivatives
derivs <- matrix(NA, nrow = nn, ncol = 4)
derivs[,1] <- derivs[,2] <- 0
derivs[,3] <- (exp(gamma_hat_0 + gamma_hat_1*mean_depth))/((exp(gamma_hat_0 + gamma_hat_1*mean_depth)+1
derivs[,4] <- (mean_depth*exp(gamma_hat_0 + gamma_hat_1*mean_depth))/((exp(gamma_hat_0 + gamma_hat_1*mea
se <- sqrt( diag ( derivs %*% vcov(yelloweye_zip) %*% t(derivs) ))
lower <- pi_hat - se * qt(0.025, nn-2, lower.tail = FALSE)
upper <- pi_hat + se * qt(0.025, nn-2, lower.tail = FALSE)

## fitted for mean & var (lambda)
beta_hat_0 <- coef(yelloweye_zip)[1]
beta_hat_1 <- coef(yelloweye_zip)[2]
lambda_hat <- exp(beta_hat_0 + beta_hat_1*mean_depth)

## matrix of derivatives
derivs_2 <- matrix(NA, nrow = nn, ncol = 4)
```

```r
derivs_2[,1] <- exp(beta_hat_0+beta_hat_1*mean_depth)
derivs_2[,2] <- mean_depth*exp(beta_hat_0+beta_hat_1*mean_depth)
derivs_2[,3] <- derivs_2[,4] <- 0
se_2 <- sqrt( diag ( derivs_2 %*% vcov(yelloweye_zip) %*% t(derivs_2) ))
lower_2 <- lambda_hat - se_2 * qt(0.025, nn-2, lower.tail = FALSE)
upper_2 <- lambda_hat + se_2 * qt(0.025, nn-2, lower.tail = FALSE)

## set plot area
par(mfrow = c(1, 2),
    mai = c(0.9, 0.9, 0.6, 0.1),
    omi = c(0, 0, 0, 0), bg = NA,
    cex.main = 1.2, cex.lab = 1.2)
## detections
plot(mean_depth, pi_hat, type = "l", las = 1, ylim = c(0, 1), lwd = 2, col = "darkgreen",
     xlab = "Tree density", ylab = expression(pi), main = "Missed detection")
lines(mean_depth, lower, lty = 2,  col = "darkgreen", lwd = 2)
lines(mean_depth, upper, lty = 2,  col = "darkgreen", lwd = 2)
## counts
plot(mean_depth, lambda_hat, type = "l", las = 1, ylim = c(0, 20), lwd = 2, col = "darkgreen",
     xlab = "Tree density", ylab = expression(lambda), main = "Counts")
lines(mean_depth, lower_2, lty = 2, col = "darkgreen", lwd = 2)
lines(mean_depth, upper_2, lty = 2, col = "darkgreen", lwd = 2)
```

Fit models for bocaccio

I think the bocaccio models are having fits with the fact that there were no bocaccio caught during the lingcod bycatch survey… maybe?

Fit model for bocaccio, all data

We will fit a GLM with ZIP distribution of catch, with fixed effects for depth, bait, lure, effort (angler_hours), month (seasonality), basin, and SURVEY.

```r
# Can't include bait/lure
# bocaccio_zip <- zeroinfl(bocaccio_catch ~
#                              # Predictors of counts
#                              mean_depth + bait + lure + month + basin + survey + angler_hours |
#                              # Predictors of detection
#                              mean_depth + bait + lure + month + basin + survey + angler_hours, data =
#
#
#
# summary(bocaccio_zip)
# AIC(bocaccio_zip)
```

```r
# Remove bait/lure
# bocaccio_zip <- zeroinfl(bocaccio_catch ~
#                               # Predictors of counts
#                               mean_depth + month + basin + survey + angler_hours |
#                               # Predictors of detection
#                               mean_depth + month + basin + survey + angler_hours, data = dat)
#
# summary(bocaccio_zip)
# AIC(bocaccio_zip)

# Keep only survey, depth, and angler hours
bocaccio_zip <- zeroinfl(bocaccio_catch ~
                             # Predictors of counts
                             mean_depth + survey + angler_hours |
                             # Predictors of detection
                             mean_depth + survey + angler_hours, data = dat)

summary(bocaccio_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = bocaccio_catch ~ mean_depth + survey + angler_hours |
##     mean_depth + survey + angler_hours, data = dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -0.71755 -0.09825 -0.06792 -0.04443 12.26150
##
## Count model coefficients (poisson with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.387e+00  2.201e+00  -0.630    0.528
## mean_depth              1.039e-02  9.606e-03   1.081    0.280
## surveylingcod_bycatch  -1.920e-06         NA      NA       NA
## surveyPercy_Washington  1.790e+00  1.289e+00   1.388    0.165
## angler_hours           -1.532e-02  5.086e-02  -0.301    0.763
##
## Zero-inflation model coefficients (binomial with logit link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             5.382983   1.724622   3.121  0.00180 **
## mean_depth             -0.006335   0.007853  -0.807  0.41988
## surveylingcod_bycatch  15.234235         NA      NA       NA
## surveyPercy_Washington  0.103437   1.423206   0.073  0.94206
## angler_hours           -0.168108   0.063601  -2.643  0.00821 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Number of iterations in BFGS optimization: 15
## Log-likelihood: -48.26 on 10 Df
```

```
AIC(bocaccio_zip)
```

```
## [1] 116.5287
```

```
# Keep only survey, depth, basin, and angler hours
bocaccio_zip <- zeroinfl(bocaccio_catch ~
                            # Predictors of counts
                            mean_depth + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth  + basin + survey + angler_hours, data = dat)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(bocaccio_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = bocaccio_catch ~ mean_depth + basin + survey + angler_hours |
##     mean_depth + basin + survey + angler_hours, data = dat)
##
## Pearson residuals:
##        Min         1Q     Median         3Q        Max
## -1.049e+00 -1.798e-02 -8.322e-04 -3.872e-06  4.359e+00
##
## Count model coefficients (poisson with log link):
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -3.533e+00  1.979e+00  -1.785 0.074234 .
## mean_depth                      7.664e-02  2.399e-02   3.194 0.001404 **
## basinHood Canal                -1.022e-06         NA      NA       NA
## basinSan Juan Islands          -2.142e-05         NA      NA       NA
## basinSouth Sound               -9.749e+00  3.488e+00  -2.795 0.005194 **
## basinStraits of Juan de Fuca -2.692e-06         NA      NA       NA
## basinWhidbey Island            -3.443e+00  1.676e+00  -2.055 0.039924 *
## surveylingcod_bycatch          -5.550e-04         NA      NA       NA
## surveyPercy_Washington          3.765e+00  1.083e+00   3.476 0.000509 ***
## angler_hours                   -1.209e-01  5.457e-02  -2.215 0.026762 *
##
## Zero-inflation model coefficients (binomial with logit link):
##                                Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                    2.2457     3.7826   0.594   0.5527
## mean_depth                     0.1559     0.1016   1.534   0.1251
## basinHood Canal               15.5661         NA      NA       NA
## basinSan Juan Islands         14.9871         NA      NA       NA
## basinSouth Sound             -28.4257    15.5192  -1.832   0.0670 .
## basinStraits of Juan de Fuca  16.9195         NA      NA       NA
## basinWhidbey Island          -12.3708     6.5703  -1.883   0.0597 .
## surveylingcod_bycatch         16.2864         NA      NA       NA
## surveyPercy_Washington         4.5271     4.1365   1.094   0.2738
## angler_hours                  -0.5133     0.2723  -1.885   0.0595 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 35
## Log-likelihood: -34.7 on 20 Df
```

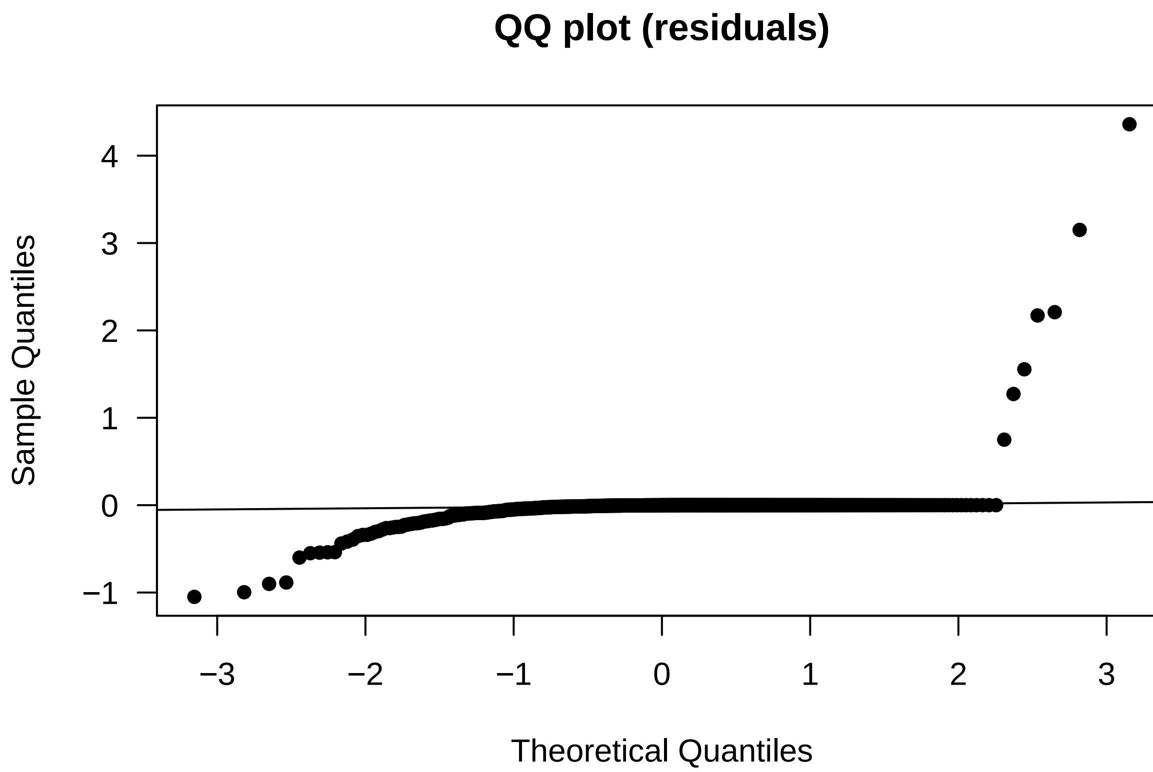
```
AIC(bocaccio_zip)
```

```
## [1] 109.4022
```

For the best model (`zeroinfl(formula = bocaccio_catch ~ mean_depth + basin + survey + angler_hours | mean_depth + basin + survey + angler_hours, data = dat)`), depth, South Sound, Whidbey Island, Percy Washington Survey, and angler hours are all significant for counts. Interestingly, the effect is only positive for depth and PW survey, and negative for all others. There are no significant variables for the detection process.

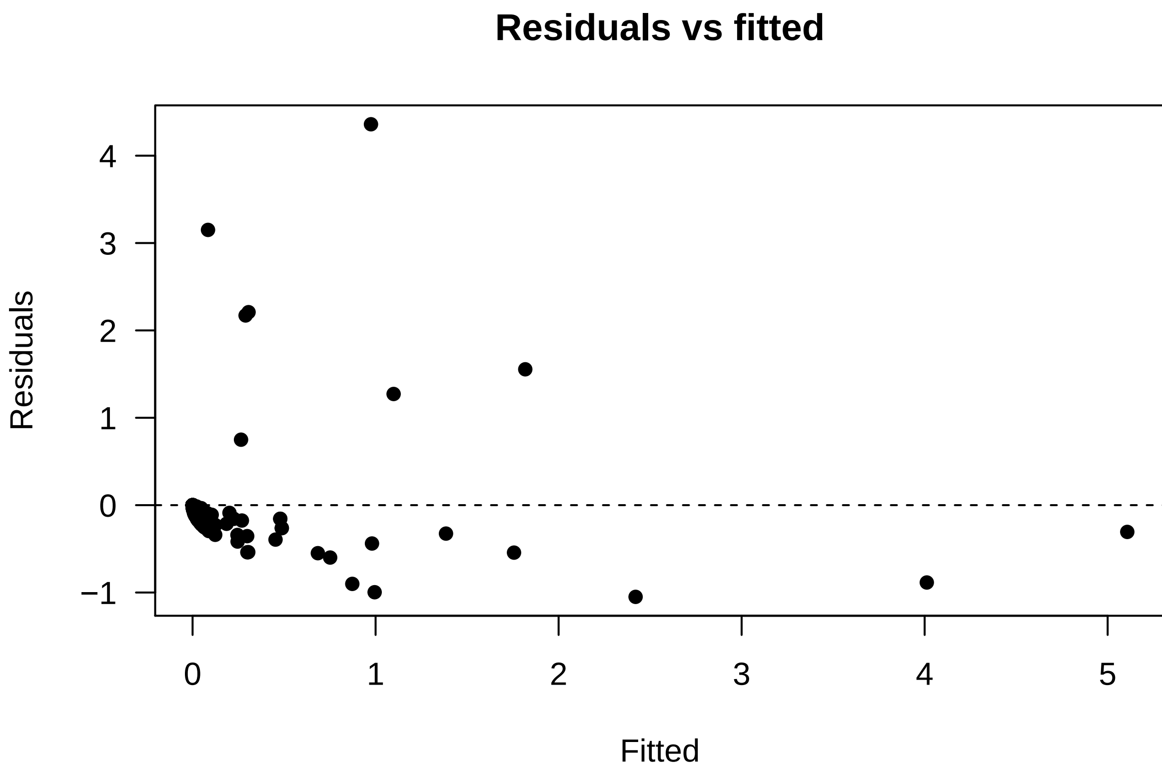
Diagnostic plots for best model

```
## qq resids
qqnorm(residuals(bocaccio_zip), main = "QQ plot (residuals)", las = 1, pch = 16)
qqline(residuals(bocaccio_zip))
```



QQ plot of residuals
Not great!

```
plot(fitted(bocaccio_zip), residuals(bocaccio_zip), las = 1, pch = 16,
     xlab = "Fitted", ylab = "Residuals",
     main = "Residuals vs fitted")
abline(h=0, lty = "dashed")
```



Residuals vs. fitted
Much better than the same plots for yelloweye... but still not great?

```
## Type-I error
alpha_crit <- 0.05

nsamples <- length(residuals(bocaccio_zip))
nvariables <- 3
## threshold value for rho (correlation)
(rho_crit <- qnorm(1 - alpha_crit/2) / sqrt(nsamples))
```

ACF

```
## [1] 0.07865065
```

```
## rearrange residuals into matrix
rr <- matrix(residuals(bocaccio_zip), nsamples, nvariables)
```

```r
## get ACF
ACF <- apply(rr, 2, acf, lag.max = 5, plot = FALSE)
ACF <- lapply(ACF, function(x) x$acf)
## convert list to matrix; don't need row 1 b/c rho_0 = 1
ACF <- do.call(cbind, ACF)[-1,]

## check if any |values| > rho_crit
any(abs(ACF) > rho_crit)
```

```
## [1] TRUE
```

I'm not sure I did this right, but looks like no autocorrelation.

Fit model for bocaccio, only PSP

We will fit a GLM with ZIP distribution of catch, with fixed effects for depth, bait, lure, effort (angler_hours), month (seasonality), basin, and SURVEY.

```r
unique(dat$basin)
```

```
## [1] "Central Sound"           "Whidbey Island"
## [3] "South Sound"             "Straits of Juan de Fuca"
## [5] "Hood Canal"              "San Juan Islands"
```

```r
PSP_dat <- subset(dat, basin %in% c("Central Sound",  "Whidbey Island", "South Sound"))

# Bait and lure break the model
# bocaccio_zip <- zeroinfl(bocaccio_catch ~
#                              # Predictors of counts
#                              mean_depth + bait + lure + month + basin + survey + angler_hours |
#                              # Predictors of detection
#                              mean_depth + bait + lure + month + basin + survey + angler_hours, data = 
#
#
#
# summary(bocaccio_zip)
# AIC(bocaccio_zip)

# Remove bait/lure
# this doesn't work though
# bocaccio_zip <- zeroinfl(bocaccio_catch ~
#                              # Predictors of counts
#                              mean_depth + month + basin + survey + angler_hours |
#                              # Predictors of detection
#                              mean_depth + month + basin + survey + angler_hours, data = PSP_dat)
```

```r
#
# summary(bocaccio_zip)
# AIC(bocaccio_zip)

# Keep only survey, depth, and angler hours
bocaccio_zip <- zeroinfl(bocaccio_catch ~
                            # Predictors of counts
                            mean_depth + survey + angler_hours |
                            # Predictors of detection
                            mean_depth + survey + angler_hours, data = PSP_dat)

summary(bocaccio_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = bocaccio_catch ~ mean_depth + survey + angler_hours |
##     mean_depth + survey + angler_hours, data = PSP_dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -0.73826 -0.11798 -0.08508 -0.06425 11.72780
##
## Count model coefficients (poisson with log link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.169e+00  2.242e+00  -0.521    0.602
## mean_depth              9.305e-03  9.871e-03   0.943    0.346
## surveylingcod_bycatch  -1.277e-06         NA      NA       NA
## surveyPercy_Washington  1.713e+00  1.287e+00   1.331    0.183
## angler_hours           -2.035e-02  5.293e-02  -0.385    0.701
##
## Zero-inflation model coefficients (binomial with logit link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             4.899279   1.731207   2.830  0.00466 **
## mean_depth             -0.005668   0.007875  -0.720  0.47165
## surveylingcod_bycatch  15.874865         NA      NA       NA
## surveyPercy_Washington  0.578502   1.417415   0.408  0.68317
## angler_hours           -0.175773   0.065650  -2.677  0.00742 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 16
## Log-likelihood: -46.96 on 10 Df
```

```r
AIC(bocaccio_zip)
```

```
## [1] 113.9243
```

```r
# Keep only survey, depth, basin, and angler hours
bocaccio_zip <- zeroinfl(bocaccio_catch ~
                            # Predictors of counts
                            mean_depth + basin + survey + angler_hours |
                            # Predictors of detection
                            mean_depth  + basin + survey + angler_hours, data = PSP_dat)

summary(bocaccio_zip)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## zeroinfl(formula = bocaccio_catch ~ mean_depth + basin + survey + angler_hours |
##     mean_depth + basin + survey + angler_hours, data = PSP_dat)
##
## Pearson residuals:
##        Min         1Q     Median         3Q        Max
## -1.0486192 -0.0398323 -0.0062379 -0.0002899  4.3600553
##
## Count model coefficients (poisson with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -3.52818    1.98251  -1.780 0.075133 .
## mean_depth               0.07659    0.02400   3.191 0.001419 **
## basinSouth Sound        -9.74462    3.48813  -2.794 0.005212 **
## basinWhidbey Island     -3.44218    1.67562  -2.054 0.039949 *
## surveylingcod_bycatch   -0.18326         NA      NA       NA
## surveyPercy_Washington  3.76224    1.08463   3.469 0.000523 ***
## angler_hours            -0.12086    0.05457  -2.215 0.026760 *
##
## Zero-inflation model coefficients (binomial with logit link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)               2.2588     3.7829   0.597   0.5504
## mean_depth                0.1554     0.1015   1.531   0.1258
## basinSouth Sound        -28.3494    15.4970  -1.829   0.0673 .
## basinWhidbey Island     -12.3391     6.5597  -1.881   0.0600 .
## surveylingcod_bycatch    15.1535         NA      NA       NA
## surveyPercy_Washington    4.5085     4.1342   1.091   0.2755
## angler_hours             -0.5120     0.2721  -1.882   0.0598 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Number of iterations in BFGS optimization: 61
## Log-likelihood: -34.7 on 14 Df
```

```r
AIC(bocaccio_zip)
```

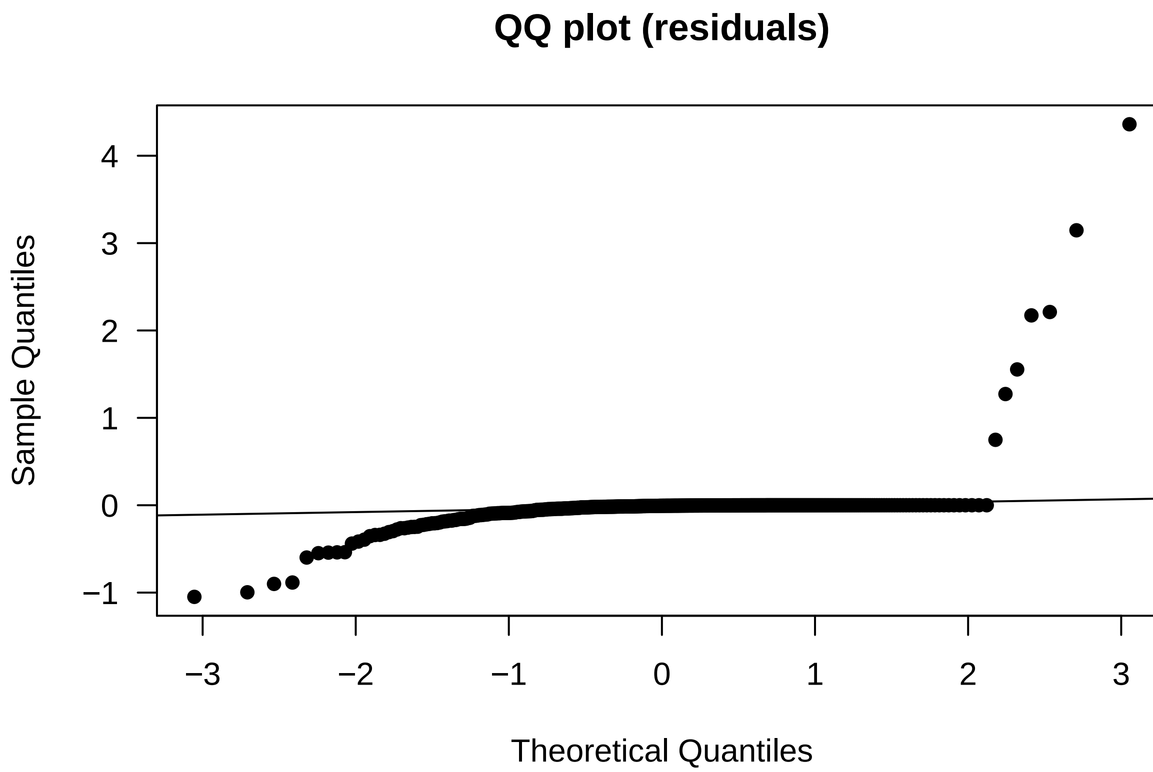
```
## [1] 97.40229
```

Here, the best model by AIC is `zeroinfl(formula = bocaccio_catch ~ mean_depth + basin + survey + angler_hours | mean_depth + basin + survey + angler_hours, data = PSP_dat)`.

In this model, there are no significant predictors of presence. However, PW survey and depth are significant positive predictors of counts, whereas South Sound, Whidbey Island, and angler hours (???) are negative predictors of counts.

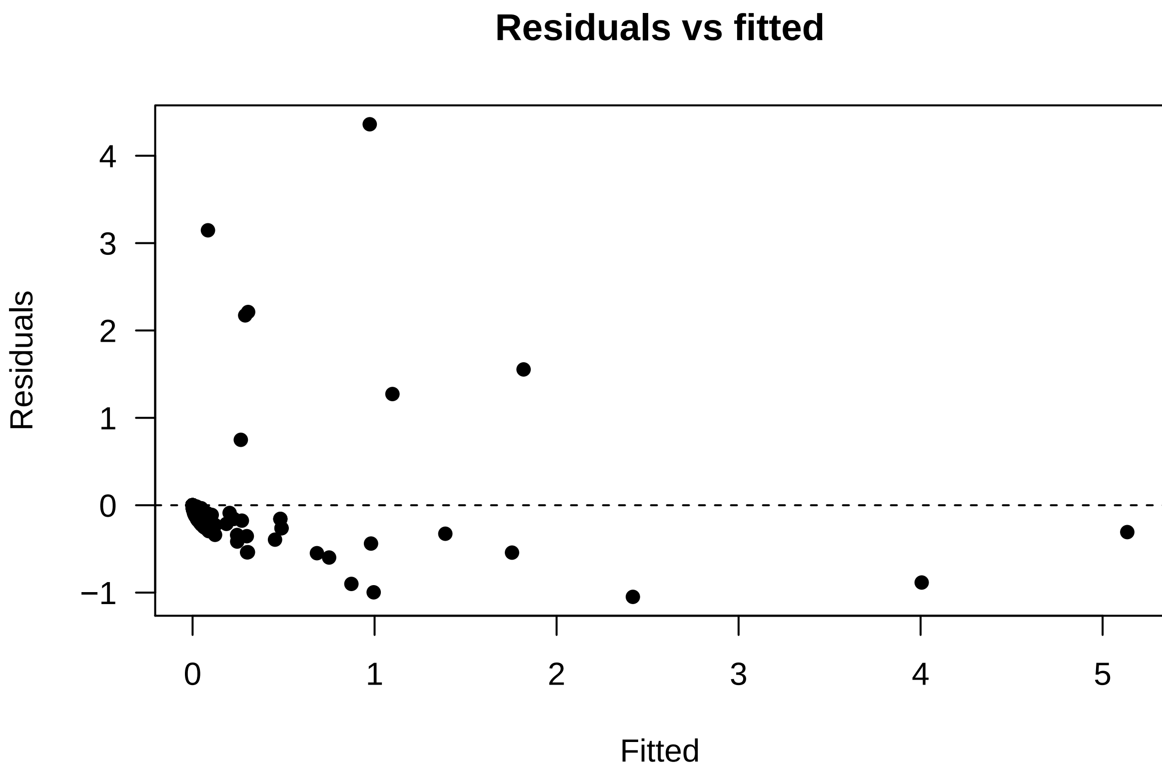
Diagnostic plots for best model

```r
## qq resids
qqnorm(residuals(bocaccio_zip), main = "QQ plot (residuals)", las = 1, pch = 16)
qqline(residuals(bocaccio_zip))
```



QQ plot of residuals
Not great!

```
plot(fitted(bocaccio_zip), residuals(bocaccio_zip), las = 1, pch = 16,
     xlab = "Fitted", ylab = "Residuals",
     main = "Residuals vs fitted")
abline(h=0, lty = "dashed")
```



Residuals vs. fitted
Again, not great but better than yelloweye.

```
## Type-I error
alpha_crit <- 0.05

nsamples <- length(residuals(bocaccio_zip))
nvariables <- 3
## threshold value for rho (correlation)
(rho_crit <- qnorm(1 - alpha_crit/2) / sqrt(nsamples))
```

ACF

```
## [1] 0.0931207
```

```
## rearrange residuals into matrix
rr <- matrix(residuals(bocaccio_zip), nsamples, nvariables)
```

```r
## get ACF
ACF <- apply(rr, 2, acf, lag.max = 5, plot = FALSE)
ACF <- lapply(ACF, function(x) x$acf)
## convert list to matrix; don't need row 1 b/c rho_0 = 1
ACF <- do.call(cbind, ACF)[-1,]

## check if any |values| > rho_crit
any(abs(ACF) > rho_crit)
```

```
## [1] TRUE
```