QAQC of the CINP ASSP 1994 - 2018 CPUE data

EKelsey

4/14/2020

Load libraries

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

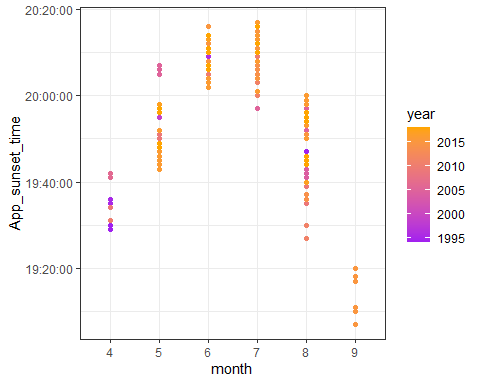
Load data and add columns neccessary for QAQC:

metadata <- read.csv('~/WERC-SC/ASSP\_share/ASSP\_4\_metadata\_CPUE\_20200413.csv') %>%   
 mutate\_at(c("App\_sunset", "std\_ending"), .funs = ~as.POSIXct(., format="%m/%d/%Y %H:%M")) %>%   
 mutate\_at(c("net\_open\_1", "net\_close\_1", "net\_open\_2", "net\_close\_2", "net\_open\_3",  
 "net\_close\_3", "net\_open\_4", "net\_close\_4", "net\_open\_5", "net\_close\_5"),  
 .funs = ~as.POSIXct(., format="%Y-%m-%d %H:%M:%S")) %>%  
 mutate\_at(c("App\_sunset", "std\_ending", "net\_open\_1", "net\_close\_1"),   
 .funs = list(time = ~ hms::as\_hms(.))) %>%   
 mutate(CPUE\_ratio = CPUEstd/CPUEraw,  
 month = as.character(month)) %>%   
 filter(TRUE)  
  
catches <- read.csv('~/WERC-SC/ASSP\_share/ASSP\_4\_catches\_BANDING\_20200414.csv') %>%   
 mutate\_at(c("App\_sunset", "std\_ending", "capture\_time", "release\_time"),  
 .funs = ~as.POSIXct(., format="%Y-%m-%d %H:%M:%S")) %>%   
 filter(species == "ASSP",  
 catchPastSS > 0, catchPastSS < 600)

# Time and Mistnetting Effort

## Graphical check of App\_sunset

ggplot(metadata, aes(month, App\_sunset\_time)) +  
 geom\_point(aes(color = year)) +  
 scale\_color\_gradient(low="purple", high="orange") +  
 theme\_bw()

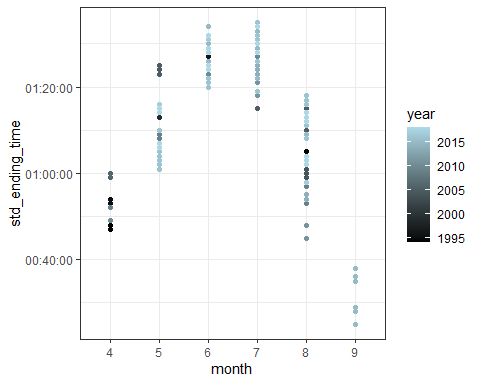
 .

This graph shows the time of apparent sunset for netting sessions each month. The range and timing for that time of year is as we would expect. Thus we conclude that the suncalc function was used effectively to get the sunset times associated with each mistnetting session.

## Graphical check of Std\_ending

### Plotted by month

ggplot(metadata, aes(month, std\_ending\_time)) +  
 geom\_point(aes(color = year)) +  
 scale\_color\_gradient(low="black", high="light blue") +  
 theme\_bw()

 .

This graph shows the time of standard ending (5.3 hours after sunset) for netting sessions each month. The range and timing of the standard ending track with sunset time as we would expect.

# Summarize net\_open and net\_close

# first (and usually only) net open time  
summary(as.POSIXct(metadata$net\_open\_1\_time))

## Min. 1st Qu. Median   
## "1970-01-01 00:00:00" "1970-01-01 20:45:00" "1970-01-01 21:02:30"   
## Mean 3rd Qu. Max.   
## "1970-01-01 20:33:23" "1970-01-01 21:36:00" "1970-01-01 23:35:00"   
## NA's   
## "22"

# first (and usually only) net close time  
summary(as.POSIXct(metadata$net\_close\_1\_time))

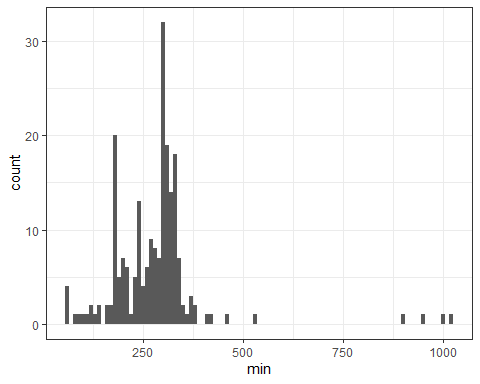
## Min. 1st Qu. Median   
## "1970-01-01 00:00:00" "1970-01-01 01:24:00" "1970-01-01 02:00:00"   
## Mean 3rd Qu. Max.   
## "1970-01-01 03:37:47" "1970-01-01 02:17:30" "1970-01-01 23:59:00"   
## NA's   
## "21"

This is not a perfect way to summarize net open and close times because the “summarize” function doesn’t recognize times across midnight here. But, by looking at the median and mean, we can tell that net open and close times are usually what we would expect, with a few late/early nights thrown in.

## Total mistnetting minutes per session

library(ggplot2)  
ggplot(metadata, aes(min)) +  
 geom\_histogram(binwidth = 10) +  
 theme\_bw()

## Warning: Removed 22 rows containing non-finite values (stat\_bin).



# summary of total mistnetting minutes  
summary(metadata$min)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 56.0 214.0 293.0 280.3 316.0 1022.0 22

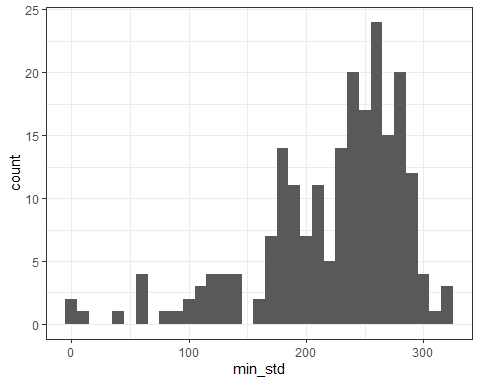
The graph and summary stats above show minutes calcuated for each netting session. We want to check that minutes were added correctly across multiple open/close sessions and also that minutes were added accurately across midnight. It looks like minutes were not added accurately across midnight on four occasions (the outliers to the far right)

## Total mistnetting standard minutes per session

### from start until end or standard ending, whichever came first

ggplot(metadata, aes(min\_std)) +  
 geom\_histogram(binwidth = 10) +  
 theme\_bw()

## Warning: Removed 22 rows containing non-finite values (stat\_bin).



# summary of mistnetting minutes cut to standard ending time  
summary(metadata$min\_std)

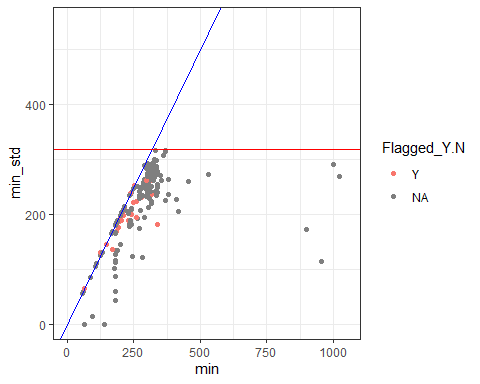
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0 189.0 239.6 223.0 267.3 317.1 22

The above graph and summary stats show the total number of mintues for the standardized session (from net open to net close or standard ending whichever came first). The standard ending is 5.3 hours (318 minutes) after sunset. Nets opened sometime after sunset.

## Compare min vs. min\_std for each session

library(ggplot2)  
ggplot(metadata, aes(min, min\_std)) +  
 geom\_point(aes(color = Flagged\_Y.N)) +  
 geom\_abline(intercept = 0, slope = 1, color = "blue") +  
 geom\_hline(yintercept = 318, color = "red") +  
 xlim(0,1050) + ylim(0, 550) +  
 theme\_bw()

## Warning: Removed 22 rows containing missing values (geom\_point).

 .

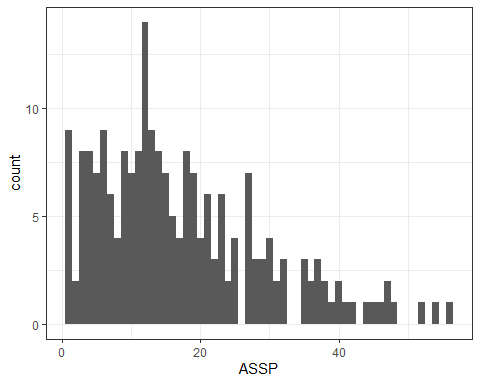
This plot of minutes vs. standardized mintues (before the 5.3 hour standardized ending). Blue line = slope of 1. Red line = standard ending. Here we can make sure standardized net open minutes is equal or less the total number of minutes and sunset to standard ending of 5.3 hours (318 minutes). Red points = data that has been flagged due to inconsistencies in data entry. It does not appear that the reason these entries were flagged effects net minutes.

# ASSP

## Histogram of total ASSP caught per session

ggplot(metadata, aes(ASSP)) +  
 geom\_histogram(binwidth = 1) +  
 theme\_bw()

## Warning: Removed 27 rows containing non-finite values (stat\_bin).



# summary of ASSP catches  
summary(metadata$ASSP)

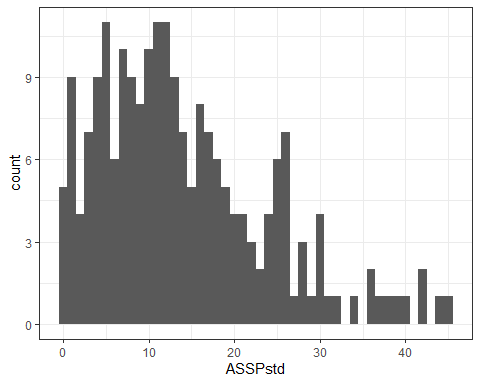
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.00 8.00 14.00 17.18 23.00 56.00 27

The graph and summary stats above show the distribution of total numbers of ASSP caught per session.

## Histogram of total ASSP caught per standardized session

ggplot(metadata, aes(ASSPstd)) +  
 geom\_histogram(binwidth = 1) +  
 theme\_bw()

## Warning: Removed 27 rows containing non-finite values (stat\_bin).



# summary of standardized ASSP catches  
summary(metadata$ASSPstd)

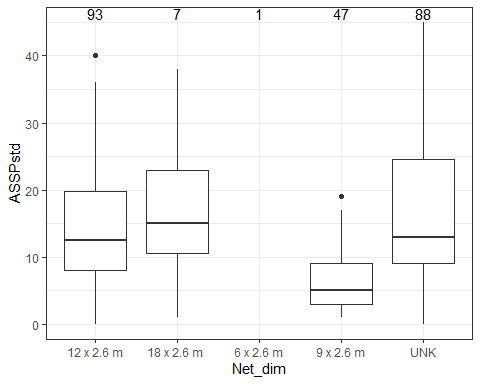
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 7.00 12.00 14.03 19.00 45.00 27

The graph and summary stats above show the distribution of total numbers of ASSP caught before standard ending or net close, whichever came first. This distribution is more constrained than the one above, which is what we would expect with the standard ending cutoff. Next we will look into what else could effect the number of birds caught

## number of ASSP caught in relation to net size

netUse <- metadata %>%   
 group\_by(Net\_dim) %>%   
 tally()  
  
ggplot(metadata, aes(Net\_dim, ASSPstd)) +  
 geom\_boxplot() +  
 geom\_text(data = netUse,  
 aes(Net\_dim, Inf, label = n), vjust = 1) +  
 theme\_bw()

## Warning: Removed 27 rows containing non-finite values (stat\_boxplot).

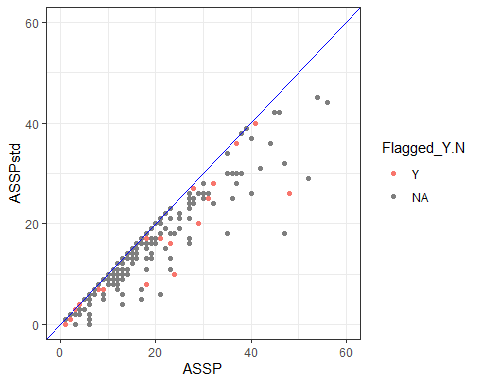
 .

The graph above shows the number of ASSP caught within the standardized period in relation to net dimensions. Number at top of each box plot = sample size. The size of the net doesn’t seem to effect the number of birds caught in a specific night. Below we will explore this effect with number of catches standardized to effort.

## comparison of ASSP vs ASSPstd

ggplot(metadata, aes(ASSP, ASSPstd)) +  
 geom\_point(aes(color = Flagged\_Y.N)) +  
 geom\_abline(intercept = 0, slope = 1, color = "blue") +  
 xlim(0,60) + ylim(0, 60) +  
 theme\_bw()

## Warning: Removed 27 rows containing missing values (geom\_point).

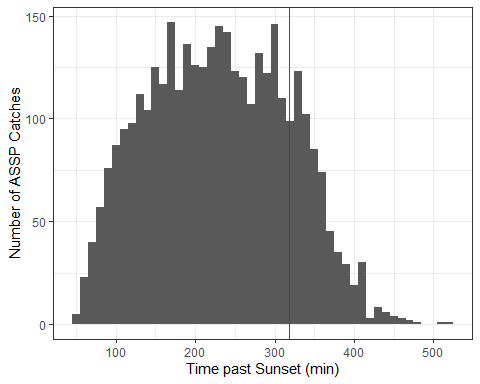
 .

The plot above shows the total number of ASSP vs. total number of ASSP before the standard ending. Blue line = slope of 1. Here we double check that the standardized number of ASSP is always equal to or less than the total number. Red points = data that has been flagged due to inconsistencies in data entry. It does not appear that the reason these entries were flagged effects the number of birds caught.

## Timing of ASSP catches

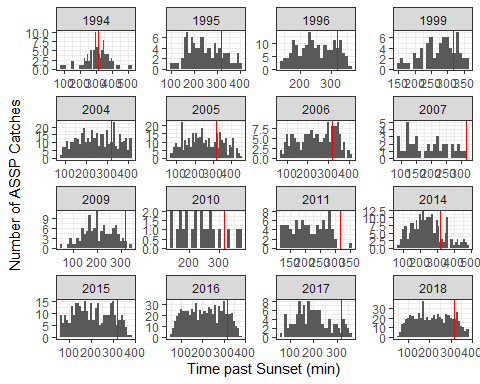
Next lets explore the frequency of catches in relation to the standard ending. Do catches start dropping off before 5.3 hours after sunset? After? NOTE - the timing of net closures will effect the number of late night captures

ggplot(catches, aes(catchPastSS)) +  
 geom\_histogram(binwidth = 10) +  
 geom\_vline(xintercept = 318, color = "red") +  
 xlab("Time past Sunset (min)") + ylab("Number of ASSP Catches") +  
 theme\_bw()



5.3 hour cutoff occurs when number of catches are still relatively high. How does that differ across years? In recent years (2014 - 2018) nets were usually closed at 2am. In earlier years nets were sometimes left open past 2am.

ggplot(catches, aes(catchPastSS)) +  
 geom\_histogram(binwidth = 10) +  
 geom\_vline(xintercept = 318, color = "red") +  
 xlab("Time past Sunset (min)") + ylab("Number of ASSP Catches") +  
 facet\_wrap(.~ year, scales = "free") +  
 theme\_bw()



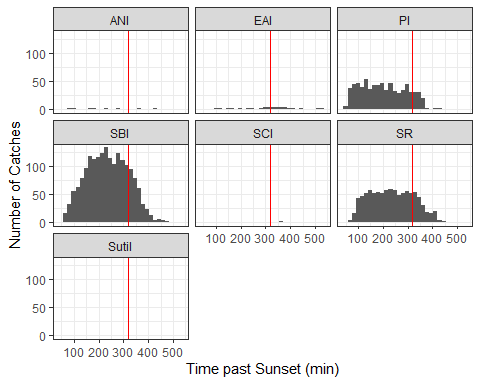
# NOTE: scales on x- and y- axis differ between panes

No clear pattern of catch timing can be seen across years.

How does catch time post sunset differ between netting locations?

ggplot(catches, aes(catchPastSS)) +  
 geom\_histogram() +  
 geom\_vline(xintercept = 318, color = "red") +  
 xlab("Time past Sunset (min)") + ylab("Number of Catches") +  
 facet\_wrap(.~ Location) +  
 theme\_bw()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



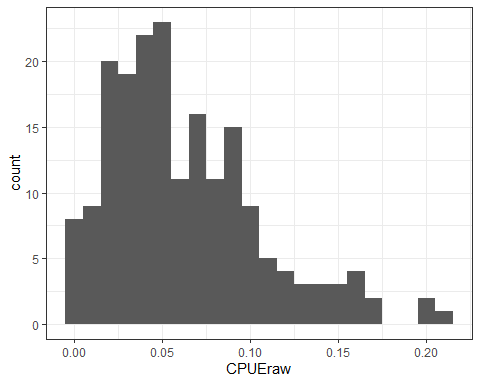
There doesn’t appear to be any distinct pattern in timing of catches between locations, especially given the variation in sample sizes between locations.

# CPUE

## visualization of CPUE per session

ggplot(metadata, aes(CPUEraw)) +  
 geom\_histogram(binwidth = 0.01) +  
 theme\_bw()

## Warning: Removed 46 rows containing non-finite values (stat\_bin).



# summary of CPUE per session  
summary(metadata$CPUEraw)

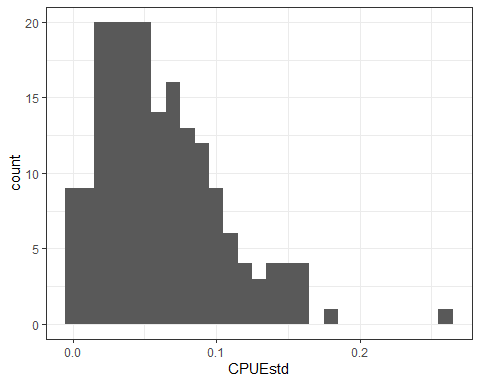
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00282 0.03245 0.05298 0.06270 0.08795 0.20779 46

The graph and summary stats above show the distribution of catch-per-unit-effort (ASSP/min).

## visualization of CPUE per standardized session

ggplot(metadata, aes(CPUEstd)) +  
 geom\_histogram(binwidth = 0.01) +  
 theme\_bw()

## Warning: Removed 47 rows containing non-finite values (stat\_bin).



# summary of CPUE per standardized session  
summary(metadata$CPUEstd)

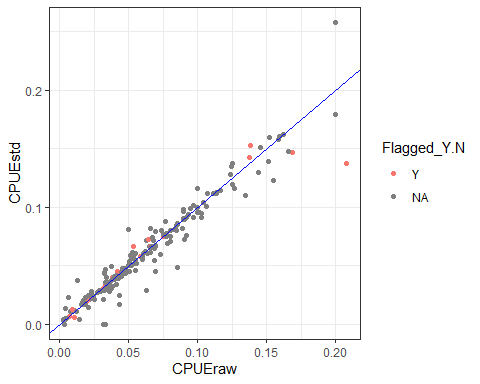
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00000 0.02858 0.05351 0.06161 0.08511 0.25810 47

The graph and summary stats above show the distribution of standardized catch-per-unit-effort (ASSPstd/min\_std). This distribution is more constrained than the previous one, which is what we would expect with the standard ending cutoff.

## comparision of CPUE vs CPUEstd

ggplot(metadata, aes(CPUEraw, CPUEstd)) +  
 geom\_point(aes(color = Flagged\_Y.N)) +  
 geom\_abline(intercept = 0, slope = 1, color = "blue") +  
 theme\_bw()

## Warning: Removed 47 rows containing missing values (geom\_point).

 .

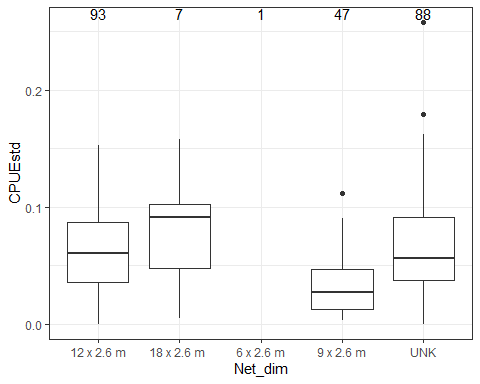
The above graph explores the correlation between CPUE and CPUE std. Blue line = slope of 1. As expected, the correlation is often 1:1, but with variation as the number of ASSP caught and number of mistnetting minutes were both effected by the standard ending cutoff but not always in a proportional way. Red points = data that has been flagged due to inconsistencies in data entry. It does not appear that the reason these entries were flagged effects the overall CPUE. The three outliers on the upper righthand side of the graph were checked to make sure the data was accurate. Sure enough, these were nights with high numbers of ASSP caught, but no errors in the data

## What other variables could influence CPUE?

### net dimensions

ggplot(metadata, aes(Net\_dim, CPUEstd)) +  
 geom\_boxplot() +  
 geom\_text(data = netUse,  
 aes(Net\_dim, Inf, label = n), vjust = 1) +  
 theme\_bw()

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).

 .

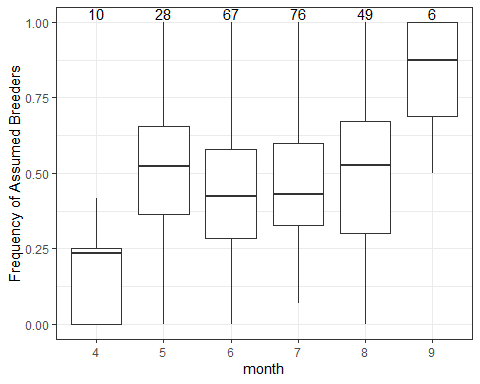
Here is the frequency of standardized CPUE values broken up by the dimensions of the net. The number above each box plot = the sample size. Unknown net sizes make it hard to determine if the size of the net influenced catch rates.

# Brood Patch and Assumed Breeders

Below we explore the relationship between ASSP assumed to be breeding and the time of year. ASSP were assumed to be breeding if they had a bare broodpatch (brood patch score of 2-4; Ainley et al. 1990)

monthCatches <- metadata %>%   
 group\_by(month) %>%   
 tally()  
  
ggplot(metadata, aes(month, BPfreq\_Y)) +  
 geom\_boxplot() +  
 ylab("Frequency of Assumed Breeders") +  
 geom\_text(data = monthCatches,  
 aes(month, Inf, label = n), vjust = 1) +  
 theme\_bw()

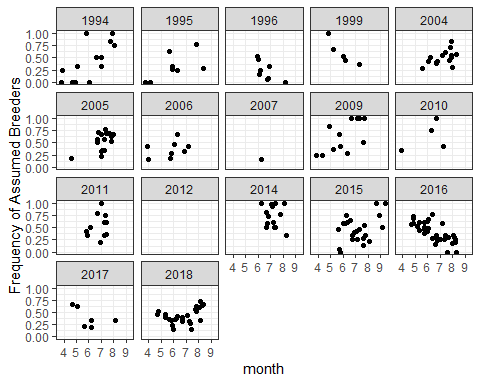
## Warning: Removed 27 rows containing non-finite values (stat\_boxplot).

 .

The above graph shows frequency of assumed breeders (brood patch 2-4) in relation to the month of netting effort. The number above each box plot = sample size. The number of assumed breeders appears to be higher later in the breeding season.

ggplot(metadata, aes(month, BPfreq\_Y)) +  
 geom\_point(position = "jitter") +  
 ylab("Frequency of Assumed Breeders") +  
 # scale\_color\_gradient(low="blue", high="yellow") +  
 facet\_wrap(.~year) +  
 theme\_bw()

## Warning: Removed 27 rows containing missing values (geom\_point).

 .

The above graph shows the frequency of assumed breeders caught in mistnetting sessions across months, broken up by years. No distinct patterns appear here.