QAQC of the CINP ASSP 1994 - 2018 CPUE data

EKelsey

4/10/2020

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

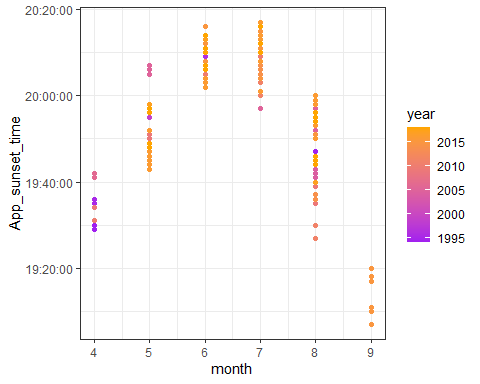
Load data and add columns neccessary for QAQC:

metadata <- read.csv('~/WERC-SC/ASSP\_share/ASSP\_4\_metadata\_CPUE\_20200325.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) %>%   
 filter(TRUE)

# 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()

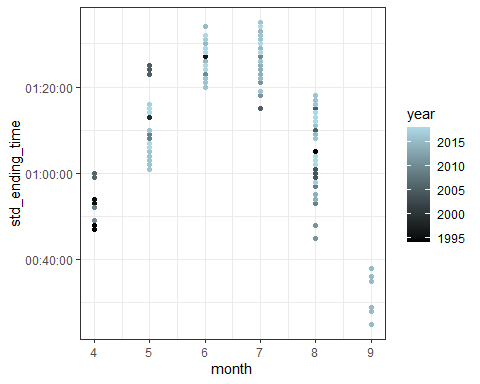
 .

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 for these 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))

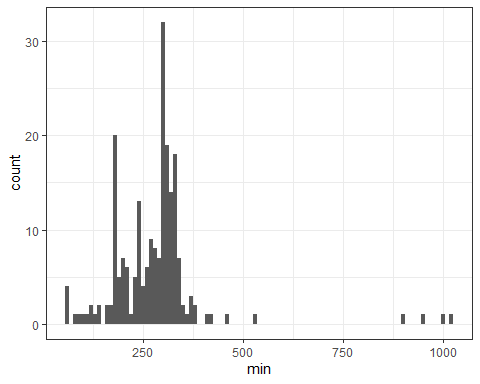
## Min. 1st Qu. Median   
## "1970-01-01 00:00:00" "1970-01-01 01:24:00" "1970-01-01 02:00:00"   
## Mean 3rd Qu. Max.   
## "1970-01-01 03:37:47" "1970-01-01 02:17:30" "1970-01-01 23:59:00"   
## NA's   
## "21"

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

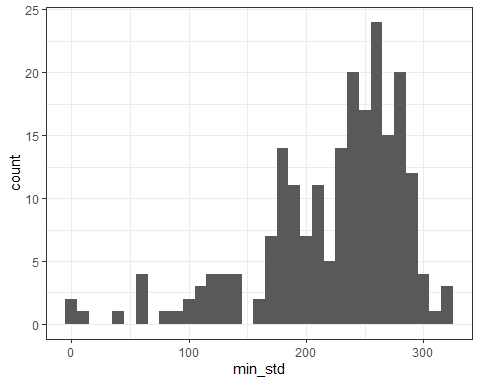
Here we visualize the total number of minutes calcuated for each netting session. We want to check that minutes were added correctly across multiple open/close sessions and also that minutes were added accurately across midnight. It looks like minutes were not added accurately across midnight on four occasions (the outliers)

## 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)

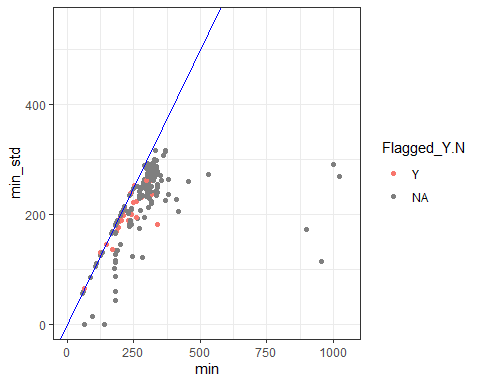
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0 189.0 239.6 223.0 267.3 317.1 22

Here we visualise the total number of mintues for the standardized session (from net open to net close or standard ending [5.3 hours after sunset] whichever came first). The standardized minutes cut out the erroneous minute calculations, as hoped. Also the max number of standardized minutes is 317, which makes sense as the maximum amount of time between sunset and 5.3 hours after.

## 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") +  
 xlim(0,1050) + ylim(0, 550) +  
 theme\_bw()

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

 .

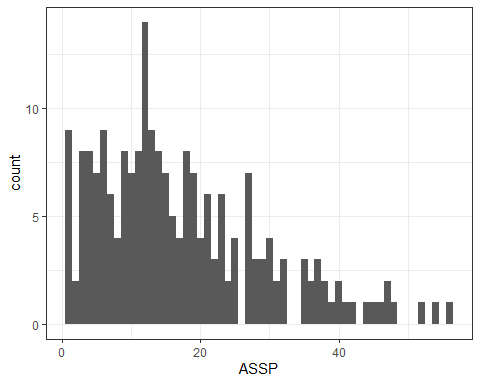
This plot of minutes vs. standardized mintues (before the 5.3 hour standardized ending). Blue line = slope of 1. Here we can make sure that minutes standardized is always equal or less than total minutes and total number of standardized minutes isn’t >317, indicating that the 5.3 hour cutoff was applied. Red points = data that has been flagged due to inconsistencies in data entry.

# 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)

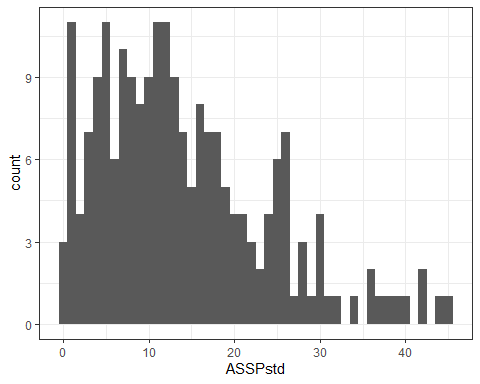
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.00 8.00 14.00 17.18 23.00 56.00 27

This graph and summary stats 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).



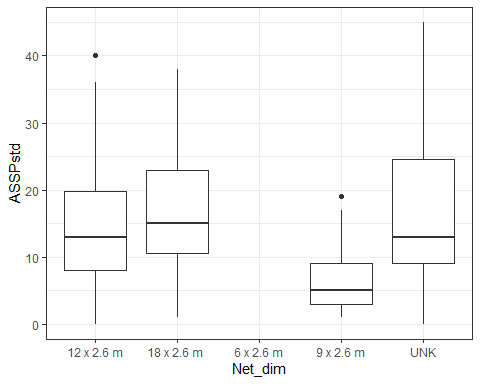
# summary of standardized ASSP catches  
summary(metadata$ASSPstd)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 7.00 12.00 14.08 19.00 45.00 27

This graph and summary stats 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

ggplot(metadata, aes(Net\_dim, ASSPstd)) +  
 geom\_boxplot() +  
 theme\_bw()

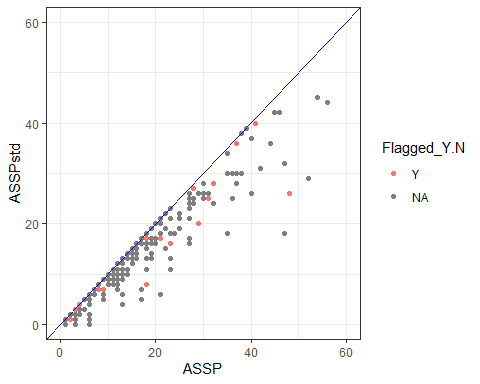
## Warning: Removed 27 rows containing non-finite values (stat\_boxplot).

 . > The size of the net doesn’t seem to effect the number of birds caught in a specific night

## 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).

 .

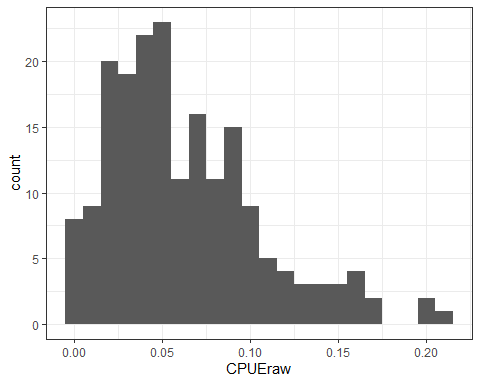
This plot of total number of ASSP vs. total number of ASSP before the standard ending. Blue line = slope of 1. Here we double check that the standardized number of ASSP is always equal to or less than the total number. Red points = data that has been flagged due to inconsistencies in data entry.

# 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)

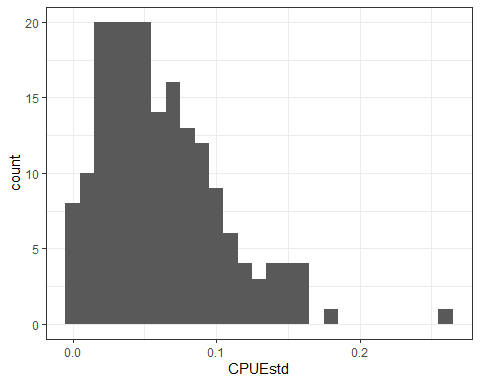
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00282 0.03245 0.05298 0.06270 0.08795 0.20779 46

This graph and summary stats 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)

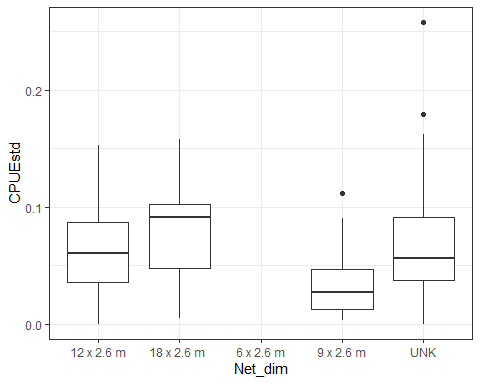
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00000 0.02858 0.05351 0.06166 0.08511 0.25810 47

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

## what other variables could influence CPUE?

ggplot(metadata, aes(Net\_dim, CPUEstd)) +  
 geom\_boxplot() +  
 theme\_bw()

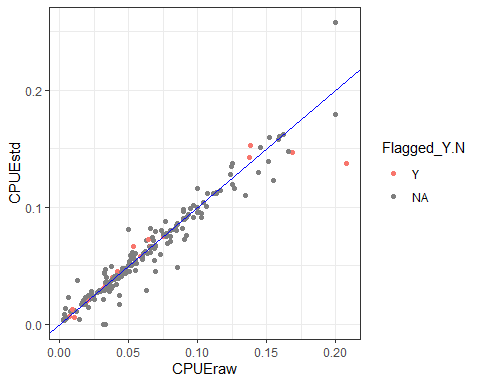
## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).

 > Here is the frequency of standardized CPUE values broken up by the dimensions of the net. Unknown net sizes make it hard to determine if the size of the net influenced catch rates

## 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).

 .

This 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. 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