

CHIS ASSP Banding Sheet QA/QC

Amelia DuVall

2 April 2020

This document explains the QA/QC conducted on the Banding sheet of the CHIS ASSP Mistnetting Database. This work was completed in collaboration with Emily Kelsey, USGS.

Load libraries

```
library(tidyverse)
library(lubridate)
```

Set-up data for analysis

Read in Banding data sheet (copied/pasted from database version 3/24/2020). Use “na.strings” function to change “ND” values to be read as “NA” so that morphometric data is interpreted as integers rather than characters.

```
banding <- read.csv("bandingQAQC.csv", header = TRUE, na.strings = c("NA", "ND"))
```

Filter data to ASSP species and banded individuals only.

```
ASSP <- group_by(.data = banding) %>%
  filter(species == "ASSP") %>%
  filter(band_no != "notbanded") %>%
  ungroup()
```

Summarize band numbers and capture rates.

Summarize data by band number and determine capture rate for each band number (e.g., each individual).

```
summary <- group_by(.data = ASSP, band_no) %>%
  summarise(no_captures = n()) %>%
  ungroup()
```

There are 3644 unique band numbers.

Summarize capture rates.

```
summarycaptures <- group_by(summary, no_captures) %>%
  summarise(count = n()) %>%
  ungroup()

show(summarycaptures)
```

```
## # A tibble: 4 x 2
##   no_captures count
##       <int> <int>
## 1         1  3466
## 2         2   159
## 3         3    16
## 4         4     3
```

Summarize recapture rates.

```
recap <- group_by(.data = ASSP, recapture) %>%
  summarise(no_captures = n()) %>%
  ungroup()

show(recap)
```

```
## # A tibble: 3 x 2
##   recapture no_captures
##   <fct>      <int>
## 1 N          3607
## 2 SNR         41
## 3 Y          196
```

The recapture rates will not necessarily match the unique band numbers because we might have encountered individuals only as a recapture (i.e., we did not band them).

How else can we proof these data?

Capture time

```
#Change data type of time stamp and pull out hour as new field
ASSP$capture_time <- mdy_hm(ASSP$capture_time, tz="US/Pacific")
ASSP$cap_hour <- hour(ASSP$capture_time)

unique(ASSP$cap_hour)
```

```
## [1] 22 23 0 1 2 3 4 21 20 11 10 12
```

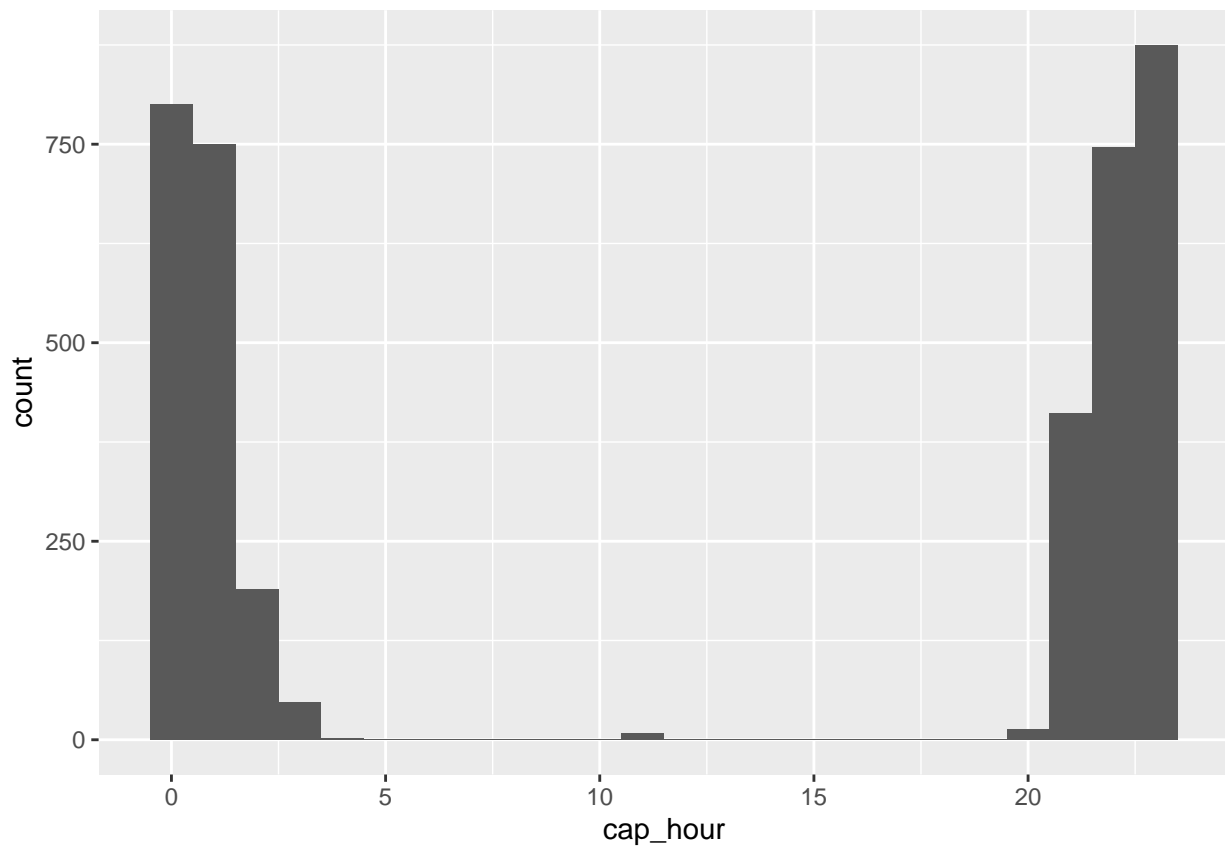
```
#Some problematic data, unique values at 11, 10, 12
#4 also seems like a really late morning capture time, but plausible, double-check.

summary(ASSP$cap_hour)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	1.00	21.00	12.18	22.00	23.00

```
#No NA's in capture time. Double-checked database, this is correct.
#They were filtered out (destroyed/lost bands).
```

```
ggplot(data = ASSP) +
  geom_histogram(mapping = aes(x = cap_hour), binwidth = 1)
```



```
#Appears that some time stamps were not entered in military time
```

```
captimesite <- arrange(ASSP, site_code, capture_time)
#Unsure how to check min/sec of capture time without preservation of data entry order
```

```
#Isolate questionable data
```

```
cap_hour_chk <- group_by(.data = ASSP, cap_hour) %>%
  filter(cap_hour %in% c(4, 10, 11, 12)) %>%
  ungroup()
```

```
write.csv(cap_hour_chk, "captureQAQC.csv")
```

```
#Verified capture times with raw data and changed any mistakes. Tracked changes in csv.
```

Release time

```
#Change data type of release time stamp and pull out hour as new field
ASSP$release_time <- mdy_hm(ASSP$release_time, tz="US/Pacific")
ASSP$rel_hour <- hour(ASSP$release_time)

unique(ASSP$rel_hour)
```

```
## [1] NA 23 0 1 2 21 22 3 4 12 20
```

```
#Problematic data: 12
#Double-check 3am and 4am also
```

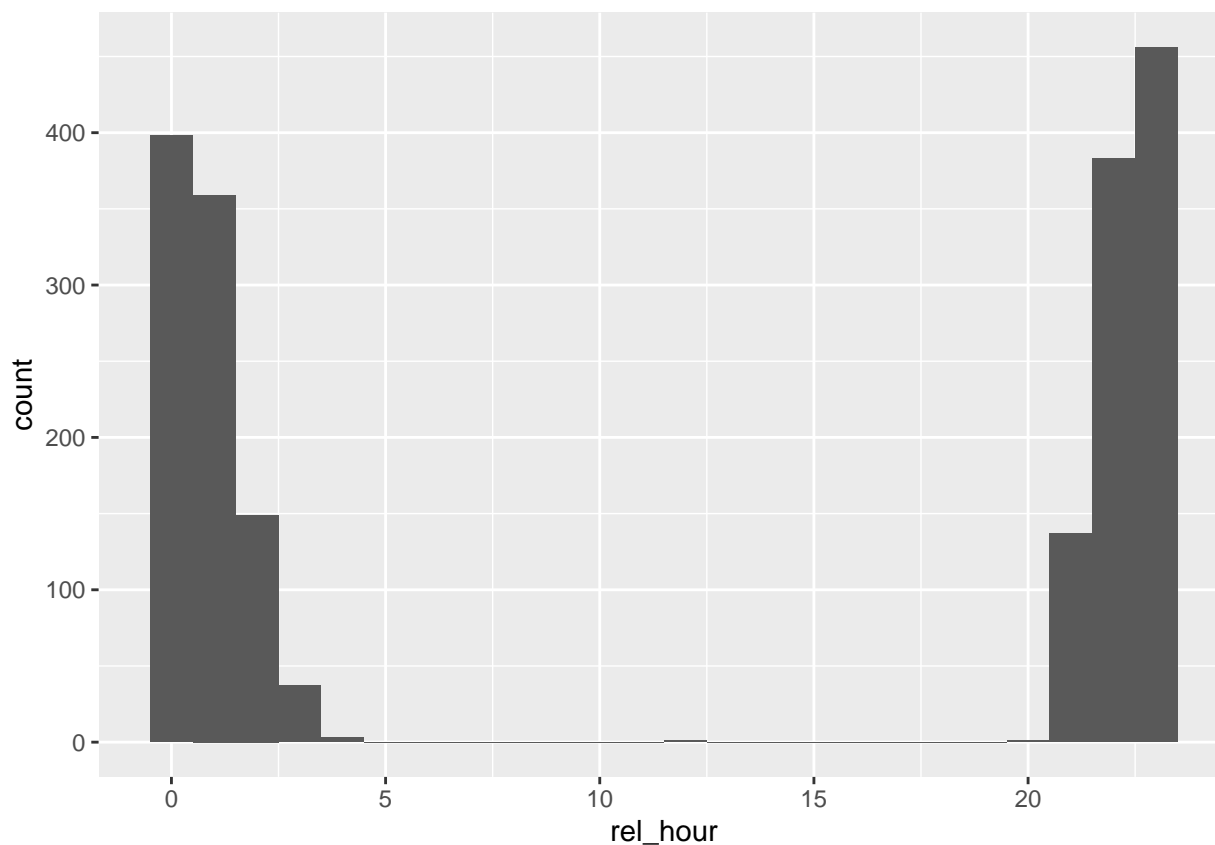
```
summary(ASSP$rel_hour)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.00   1.00   21.00   11.75   22.00   23.00   1920
```

```
#1920 NA's
#Min and max make sense
```

```
ggplot(data = ASSP) +
  geom_histogram(mapping = aes(x = rel_hour), binwidth = 1)
```

```
## Warning: Removed 1920 rows containing non-finite values (stat_bin).
```



```

#Also appears to be data not entered in military time

#Isolate questionable data
rel_hour_chk <- group_by(.data = ASSP, rel_hour) %>%
  filter(rel_hour %in% c(3, 4, 12)) %>%
  ungroup()

write.csv(rel_hour_chk, "reltimeQAQC.csv")
#Verified with raw data and changed any mistakes. Tracked changes in csv.

```

Brood patch

```
unique(ASSP$BP)
```

```
## [1] 4 1.5 0 1 3 5 <NA> 2 D PD B b pd d 4.5
## Levels: 0 1 1.5 2 3 4 4.5 5 b B d D pd PD
```

```

#Fix lowercase values
#All in range of values included in data dictionary

#Not sure how else to QAQC?

```

Mass

```

#=====
## Mass(uncorrected)
unique(ASSP$uncorr_mass)

```

```
## [1] 31 36 34 32 33 35 37 38 39 62 61 60 NA 58 40 41 26 44 68
## [20] 47 48 50 59 63 51 55 64 69 75 42 65 67 57 49 54 66 56 70
## [39] 30 53 72 71 52 46 45 74 73 43 29 28 157
```

```
summary(ASSP$uncorr_mass)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    26.00   35.00   52.00   48.14   60.00   157.00    493
```

```

#Large range between 26 - 157 but it's uncorrected data
#493 NAs

```

```

ggplot(data = ASSP) +
  geom_bar(mapping = aes(x = uncorr_mass), binwidth = 1)

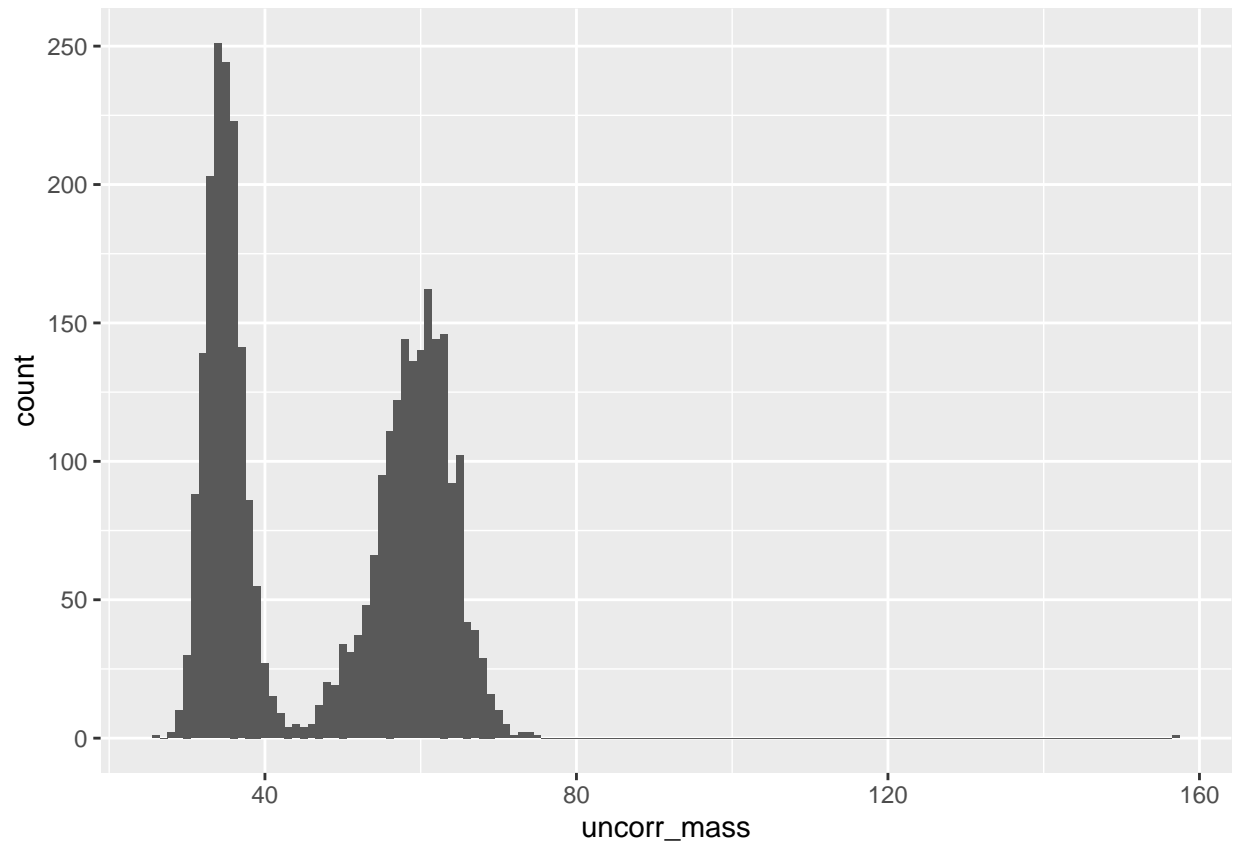
```

```

## Warning: `geom_bar()` no longer has a `binwidth` parameter. Please use
## `geom_histogram()` instead.

```

```
## Warning: Removed 493 rows containing non-finite values (stat_bin).
```



```
#Bimodal! Could be due to two different weighing containers used.
#Also really high outlier (157)
```

```
#Decided not to QAQC this field, but to check mass_corr values.
```

```
#=====
##Mass tare
unique(ASSP$mass_tare)
```

```
## [1] 0 NA 27 13 26 14 25 23 30 22 31 33 35 29 28 17 24 32 12 21 16 36 20 15 10
## [26] 18 19 9 11
```

```
summary(ASSP$mass_tare)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.0     0.0    16.0    12.7   25.0    36.0    494
```

```
#494 NAs
#1 more NA in mass tar than uncorr_mass. Why?
```

```
#Decided not to QAQC this field, but to check mass_corr values.
```

```
#=====
##Mass (corrected)
```

```
unique(ASSP$mass_corr)
```

```
## [1] 31 36 34 32 33 35 37 38 39 62 61 60 NA 58 40 41 26 44 43  
## [20] 42 29 30 23 17 28 22 48 45 6 7 135 20 27 24 46
```

```
summary(ASSP$mass_corr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      6.00   34.00   35.00   35.45   37.00   135.00     494
```

```
#494 NAs
```

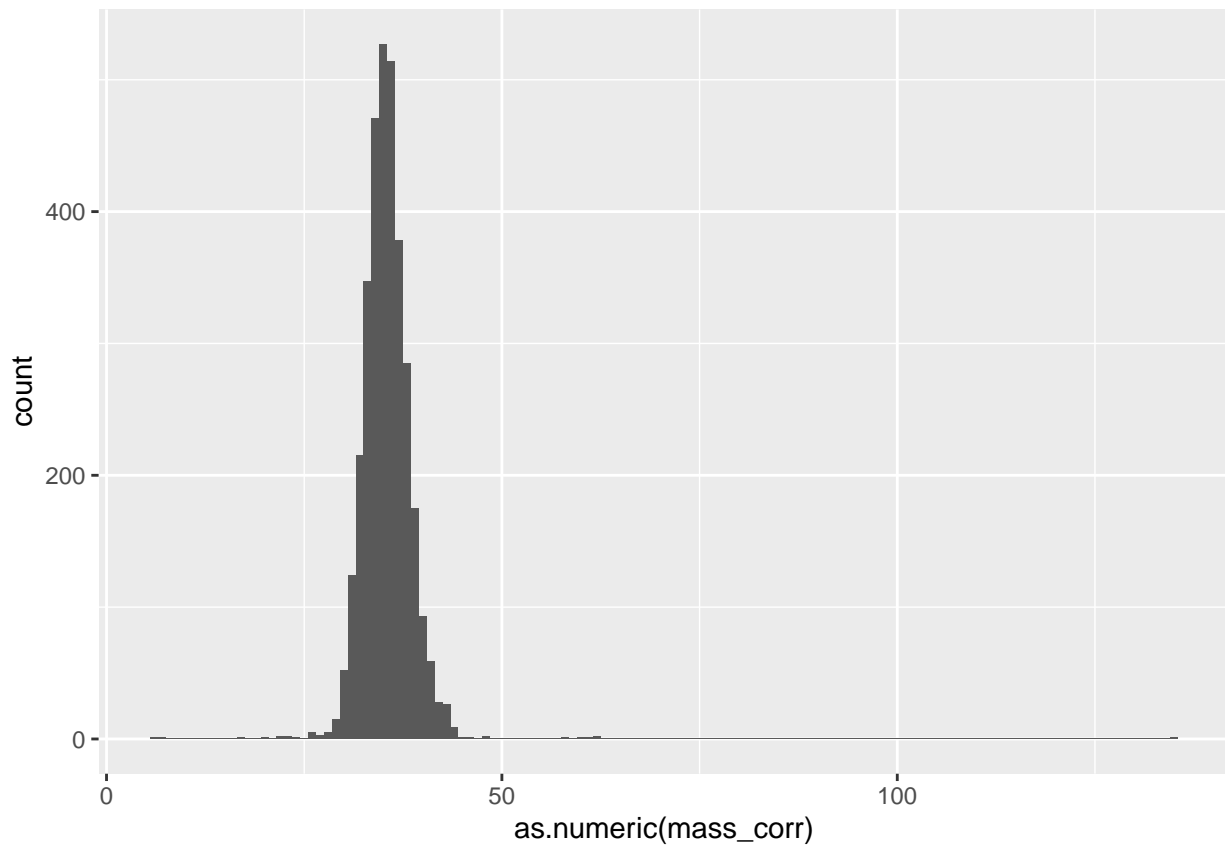
```
#large range between 6 - 135
```

```
#mean 35.45
```

```
#median 35
```

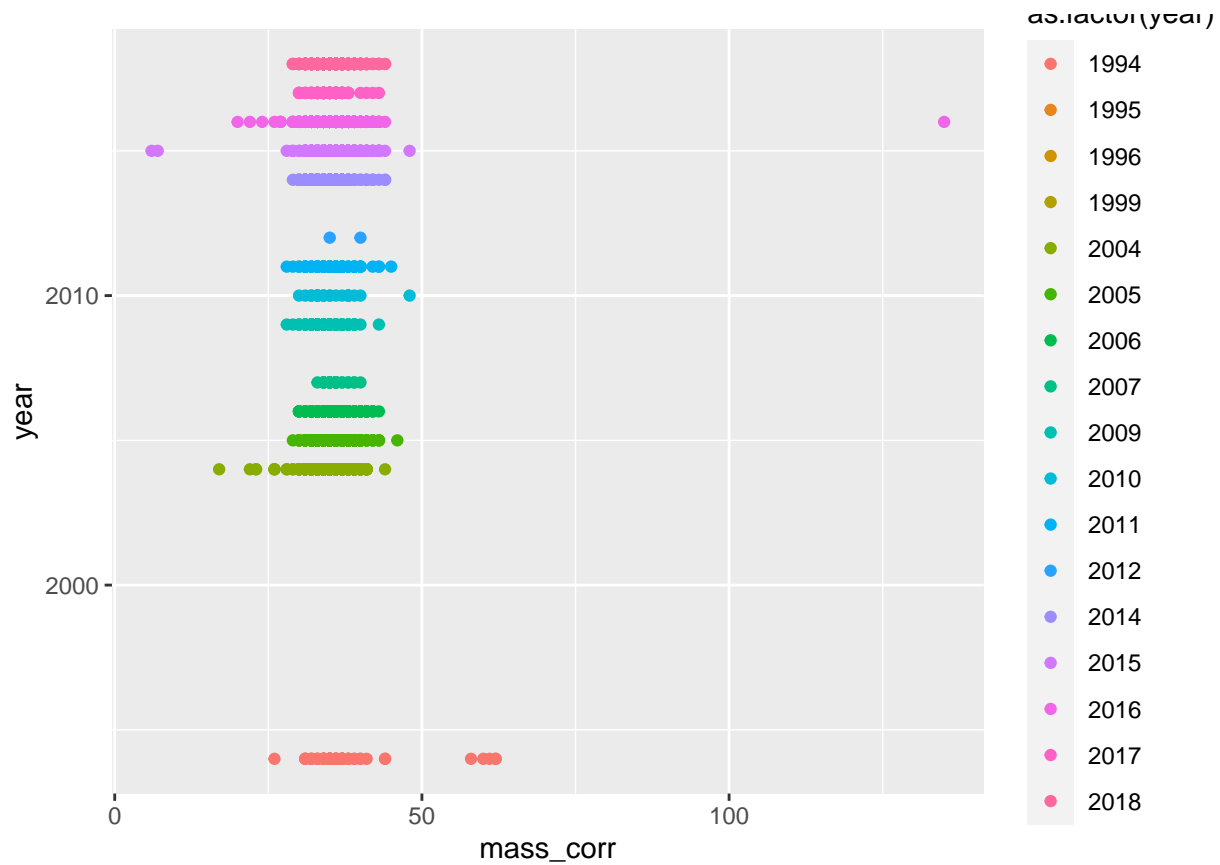
```
ggplot(data = ASSP) +  
  geom_histogram(mapping = aes(x = as.numeric(mass_corr)), binwidth = 1)
```

```
## Warning: Removed 494 rows containing non-finite values (stat_bin).
```



```
ggplot(data = ASSP) +  
  geom_point(mapping = aes(x = mass_corr, y = year, color = as.factor(year)))
```

```
## Warning: Removed 494 rows containing missing values (geom_point).
```



```
#Need to double-check outliers
```

```
#Re Adams 2016 paper, Mass (g) 36.1 +/- 2.8 (f) and 34.7 +/- 2.1 (m)
```

```
#Isolate questionable data
```

```
mass_corr_chk <- filter(ASSP, mass_corr < 25 | mass_corr > 50)
```

```
write.csv(mass_corr_chk, "masscorrQAQC.csv")
```

```
#Verified with raw data and changed any mistakes. Tracked changes in csv.
```

Culmen

```
summary(ASSP$culmen)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    10.50  14.10   14.50   14.51  14.99   24.80  1253
```

```
#Large range between 10.50 - 24.80
```

```
#Mean at 14.51
```

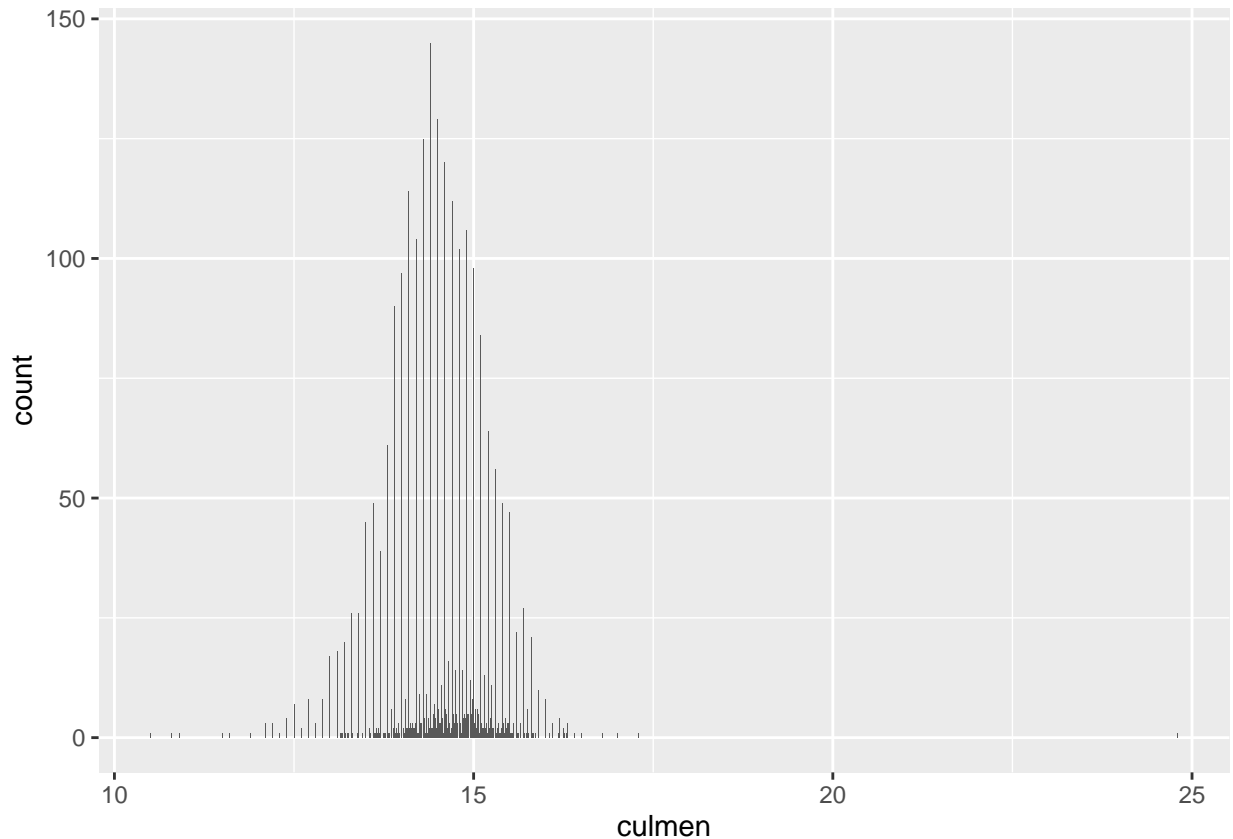
```
#Medium at 14.50
```



```
#Larger values seen incongruous
```

```
ggplot(data = ASSP) +  
  geom_bar(mapping = aes(x = culmen))
```

```
## Warning: Removed 1253 rows containing non-finite values (stat_count).
```



```
#Re Adams 2016 paper, Bill length (mm) 14.9 ? 0.5 (f) and 14.6 ? 0.8 (m)  
#Re Pyle guide, 13.1-15.2 (95% CI)
```

```
#Isolate questionable data
```

```
culmen_chk <- filter(ASSP, culmen < 13 | culmen > 16)
```

```
write.csv(culmen_chk, "culmenQAQC.csv")
```

```
#Verified with raw data and changed any mistakes. Tracked changes in csv.
```

```
#Seems like the range I picked (13-16) was too narrow. Most were not typos.
```

```
#Assumed values w/in (12-16.9) were not typos after reviewing data.
```

Skull length

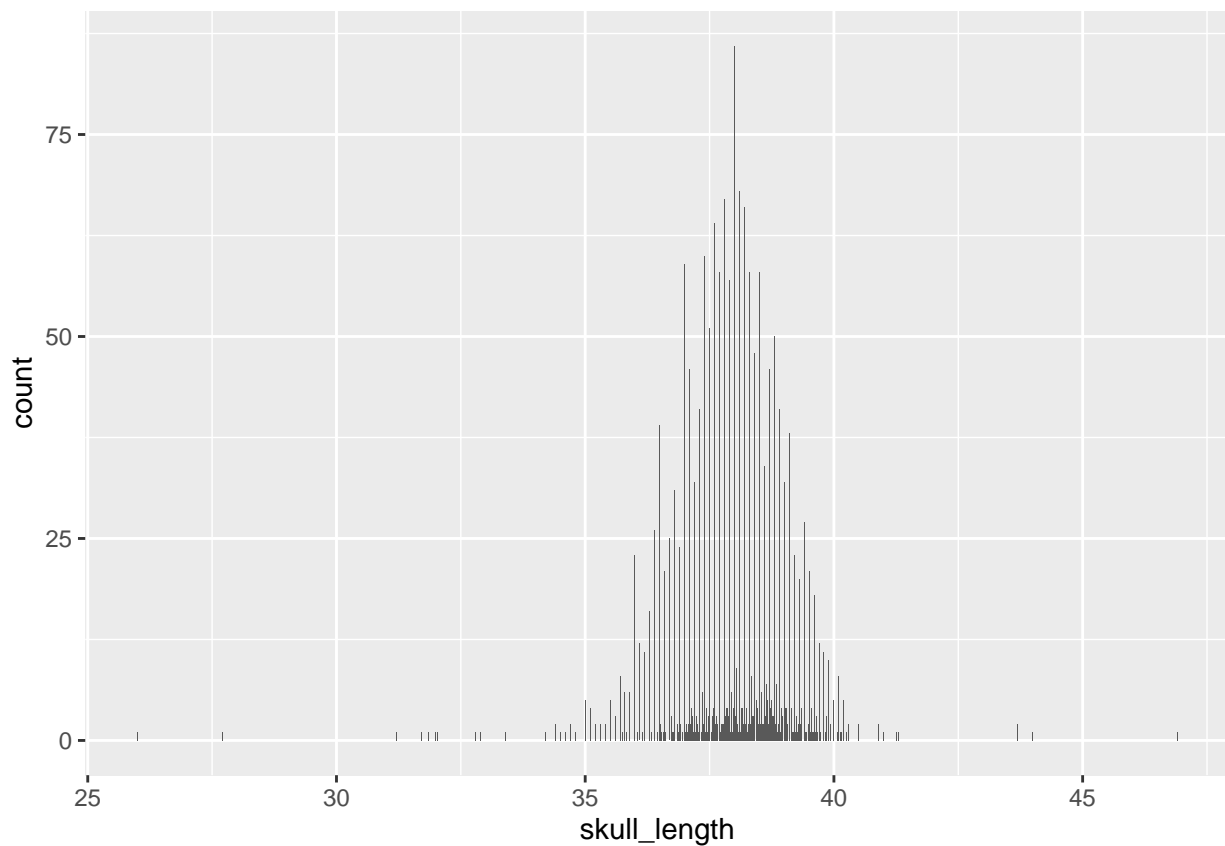
```
summary(ASSP$skull_length)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      26.00   37.30   38.00   37.93   38.65   46.90    1805
```

```
#Range between 26 - 46.90
#Mean at 37.93
#Median at 38
#Lower values seen incongruous
```

```
ggplot(data = ASSP) +
  geom_bar(mapping = aes(x = skull_length))
```

```
## Warning: Removed 1805 rows containing non-finite values (stat_count).
```



```
#Re Adams 2016 paper, skull length (mm) 38.1 ± 1.1 (f) and 37.9 ± 0.8 (m)
#No info in Pyle guide
```

```
#Isolate questionable data
skull_chk <- filter(ASSP, skull_length < 35 | skull_length > 41)
write.csv(skull_chk, "skullQAQC.csv")
```

Tarsus

```
summary(ASSP$tarsus)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##    12.60   22.80   23.45   23.55   24.10   235.00   1252
```

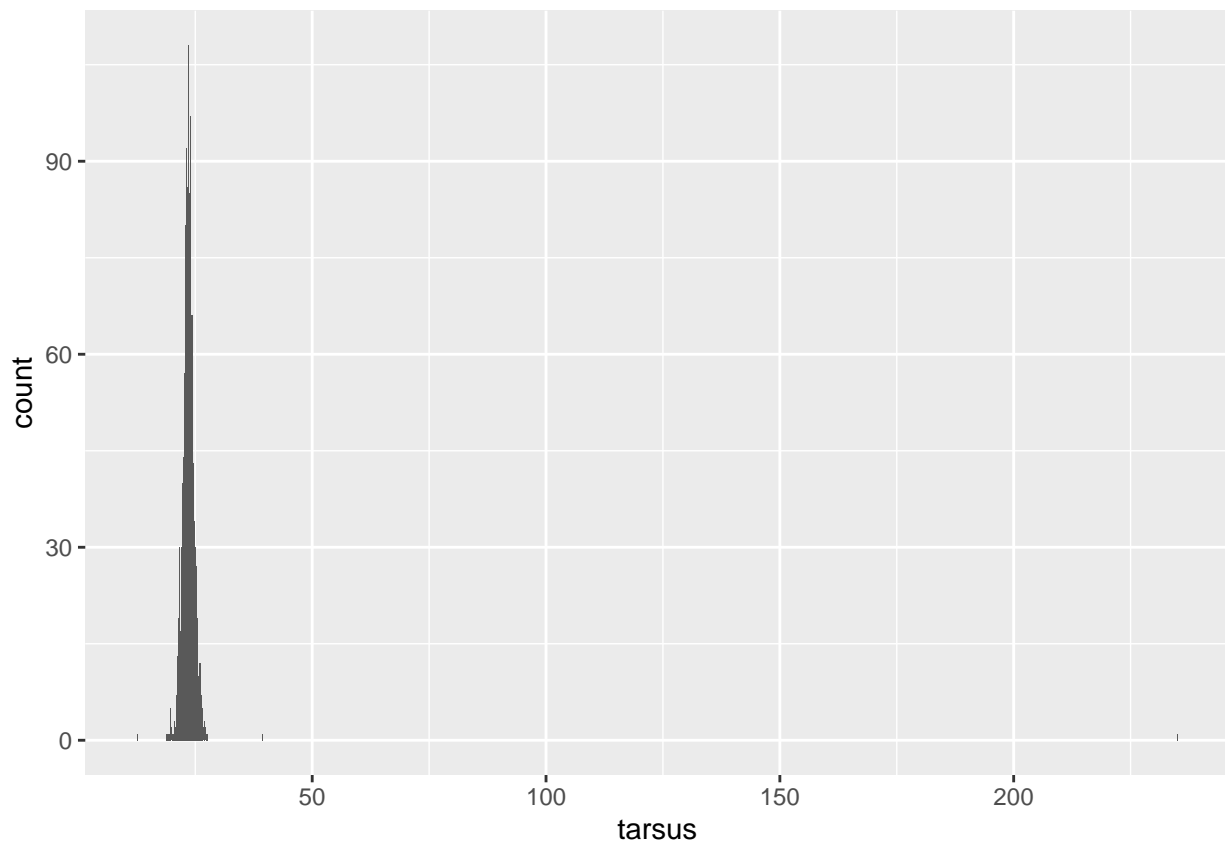
```
#Range between 12.60 - 235
```

```
#Mean at 23.55
```

```
#Median at 23.45
```

```
ggplot(data = ASSP) +  
  geom_bar(mapping = aes(x = tarsus))
```

```
## Warning: Removed 1252 rows containing non-finite values (stat_count).
```



```
#Re Adams 2016 paper, tarsus (mm) 23.2 ? 0.9 (f) and 23.1 ? 0.8 (m)
```

```
#Re Pyle guide, 21-25 (95% CI)
```

```
#Isolate questionable data
```

```
tarsus_chk <- filter(ASSP, tarsus < 19 | tarsus > 27)
```

```
write.csv(tarsus_chk, "tarsusQAQC.csv")
```

Wing

```
summary(ASSP$wing)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##      123     138     141     141     143    1425    161
```

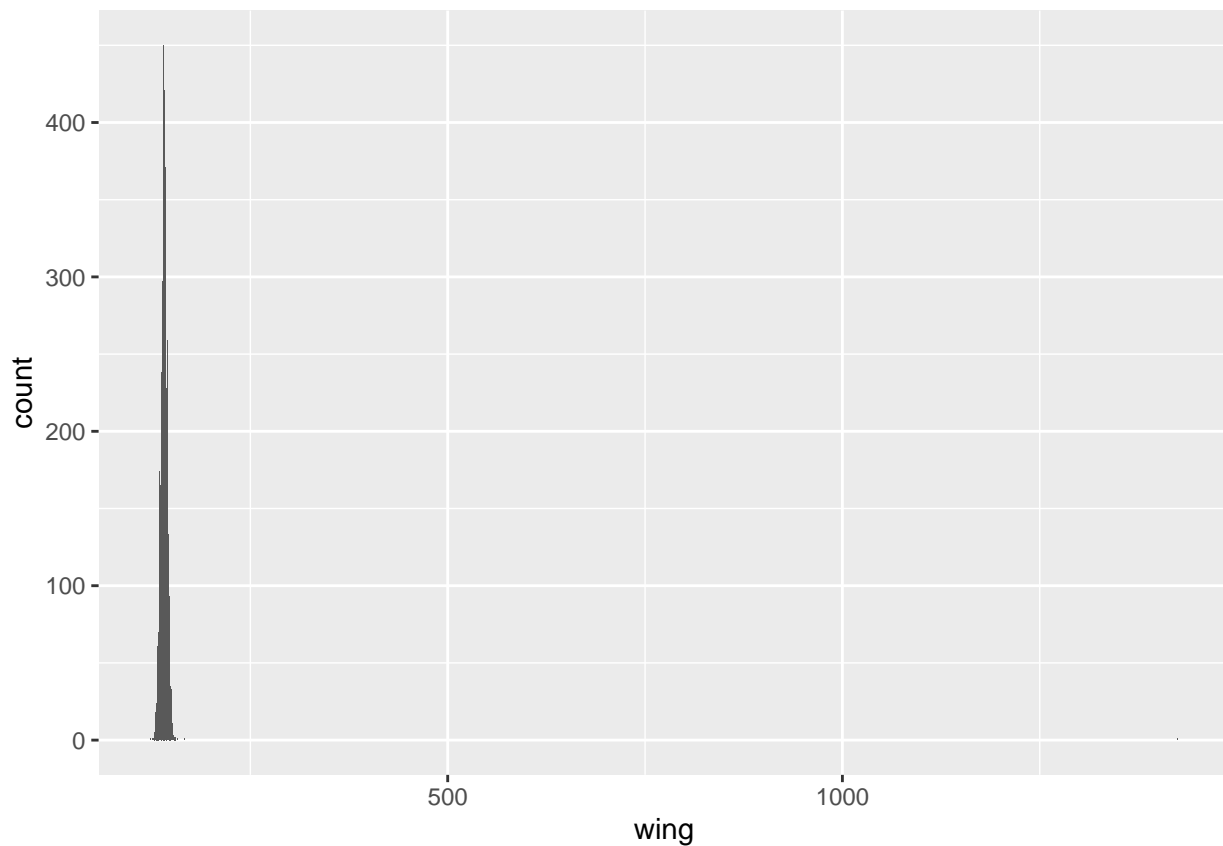
```
#Range between 123 - 1425
```

```
#Mean at 141
```

```
#Median at 141
```

```
ggplot(data = ASSP) +  
  geom_bar(mapping = aes(x = wing))
```

```
## Warning: Removed 161 rows containing non-finite values (stat_count).
```



```
#Re Adams 2016, max flat wing (mm) 142.7 ? 2.8 (f) and 140.4 ? 3.3 (m)
```

```
#Re Pyle guide, wing chord 132-148 (95% CI)
```

```
#Isolate questionable data
```

```
wing_chk <- filter(ASSP, wing < 130 | wing > 150)
```

```
write.csv(wing_chk, "wingQAQC.csv")
```

Tail

```
summary(ASSP$tail)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##      9.00   76.50   78.00   72.10   80.75   83.00   3834
```

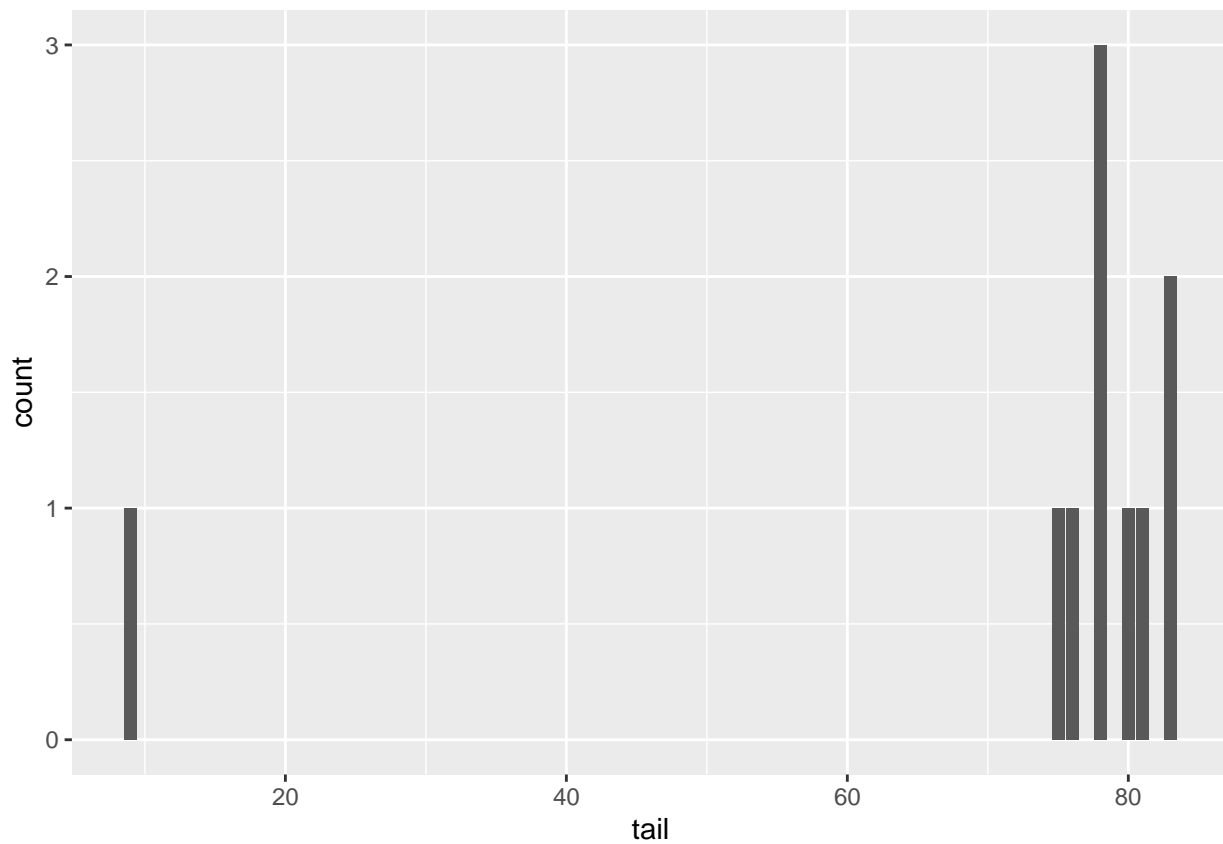
```
#Range between 9 - 83
```

```
#Mean 72.10
```

```
#Median 78
```

```
ggplot(data = ASSP) +  
  geom_bar(mapping = aes(x = tail))
```

```
## Warning: Removed 3834 rows containing non-finite values (stat_count).
```



```
#Re Pyle, 72-84 (95% CI)
```

```
#Isolate questionable data
```

```
tail_chk <- filter(ASSP, tail < 71 | tail > 85)
```

```
write.csv(tail_chk, "tailQAQC.csv")
```