# Flannelmouth Exploratory Graphs

*Jan K Boyer*

*2/18/2020*

## Load and format data

```
# exploratory figures
# see size structure and size at maturity so we can decide how to bin
# fish into size/age classes for meodelling

#Author: Jan Boyer, AGFD, jboyer@azgfd.gov
#Inputs: ./data/all_flannelmouth.csv
#Outputs: none
#Dependencies: none

require(tidyverse)
```

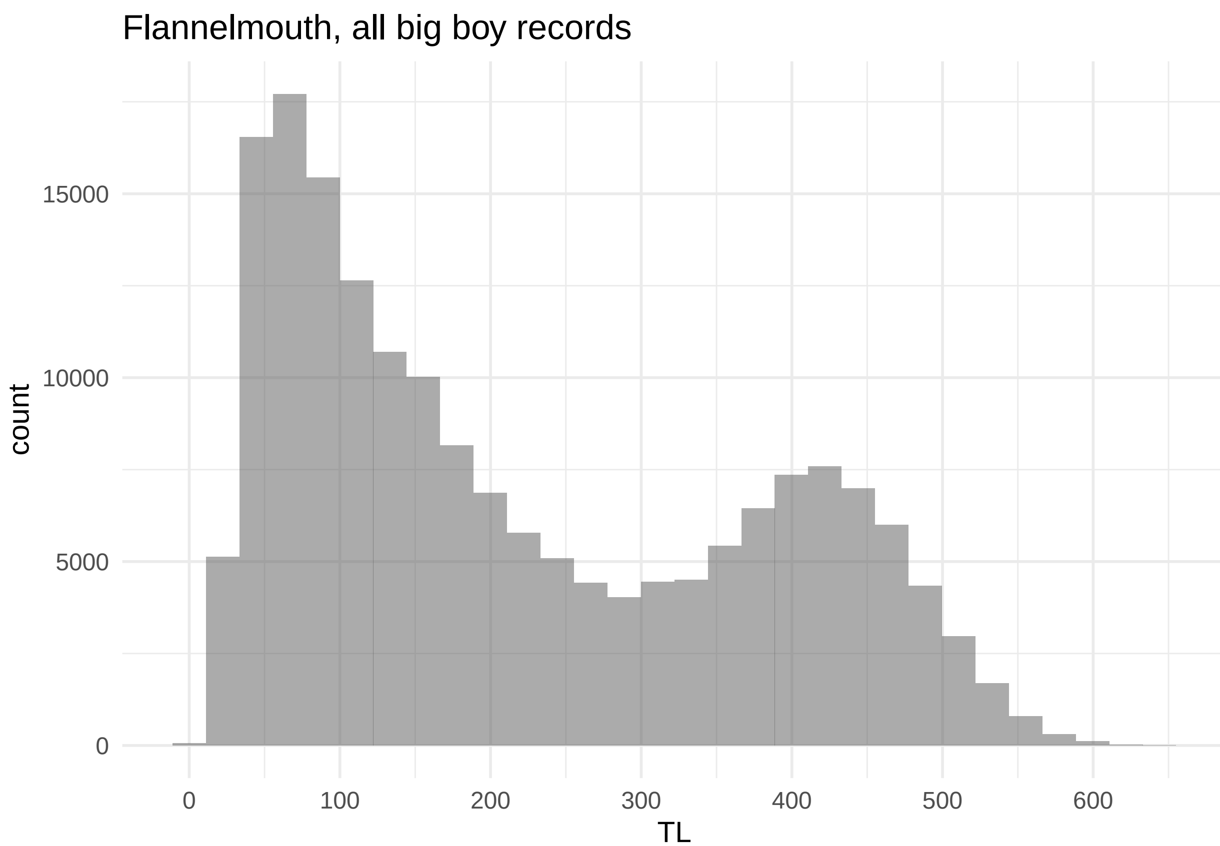
```
## Loading required package: tidyverse

## -- Attaching packages --------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.0     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
theme_set(theme_minimal()) #ggplot defaults are ugly, set a better theme


# load data - all flannelmouth records from big boy
fms <- read.csv("./data/all_flannelmouth.csv", stringsAsFactors = FALSE)

#format datetime as datetime
fms <- fms %>%
  mutate(START_DATETIME = as.POSIXct(START_DATETIME))

#subset data
fms <- fms %>%
  filter(!is.na(TL)) %>% #subset to fish with length data
  # remove very large fish. maybe these records are real, but I'm sceptical,
  # and we don't need to waste our graph space looking at long tails
  filter(TL <= 650)
```

## examine length frequency

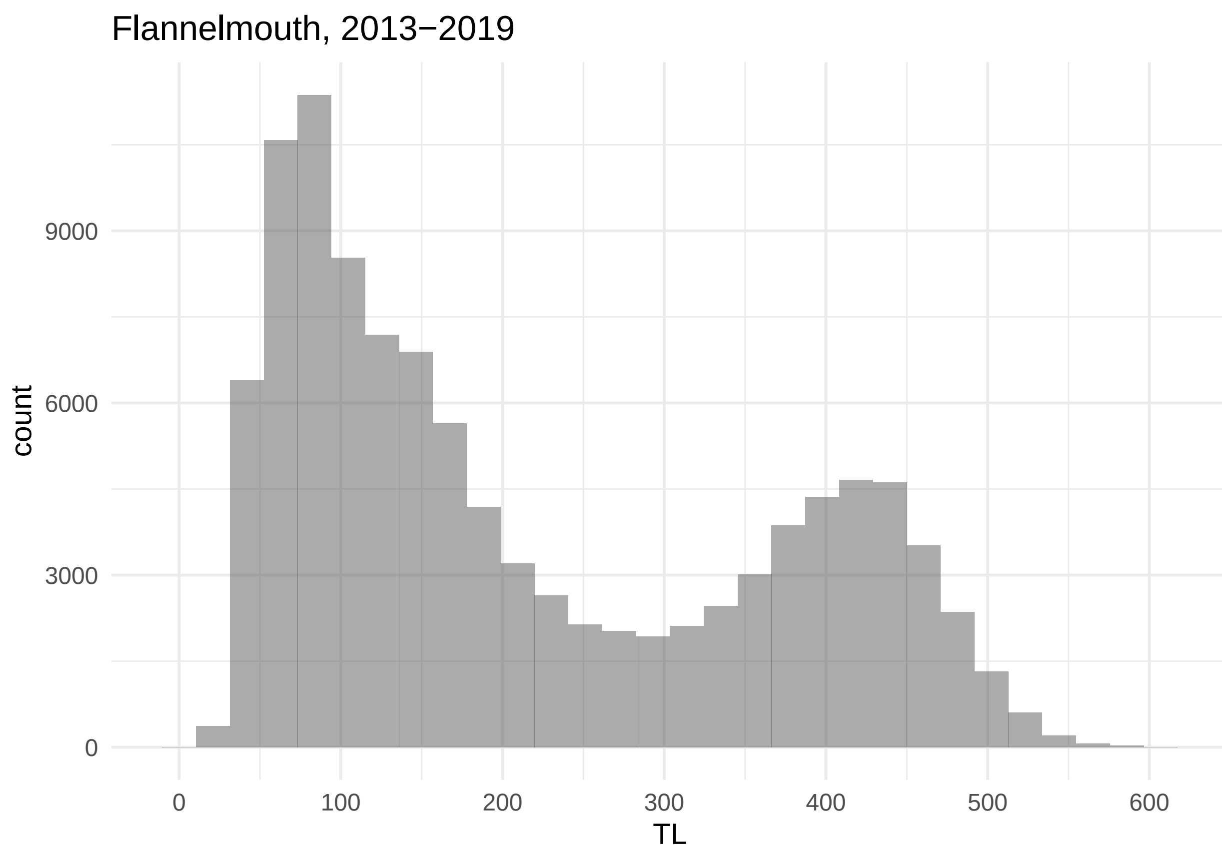
```
# length freq histogram
fms %>%
  ggplot(aes(TL)) +
  geom_histogram(alpha = 0.5) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, all big boy records")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



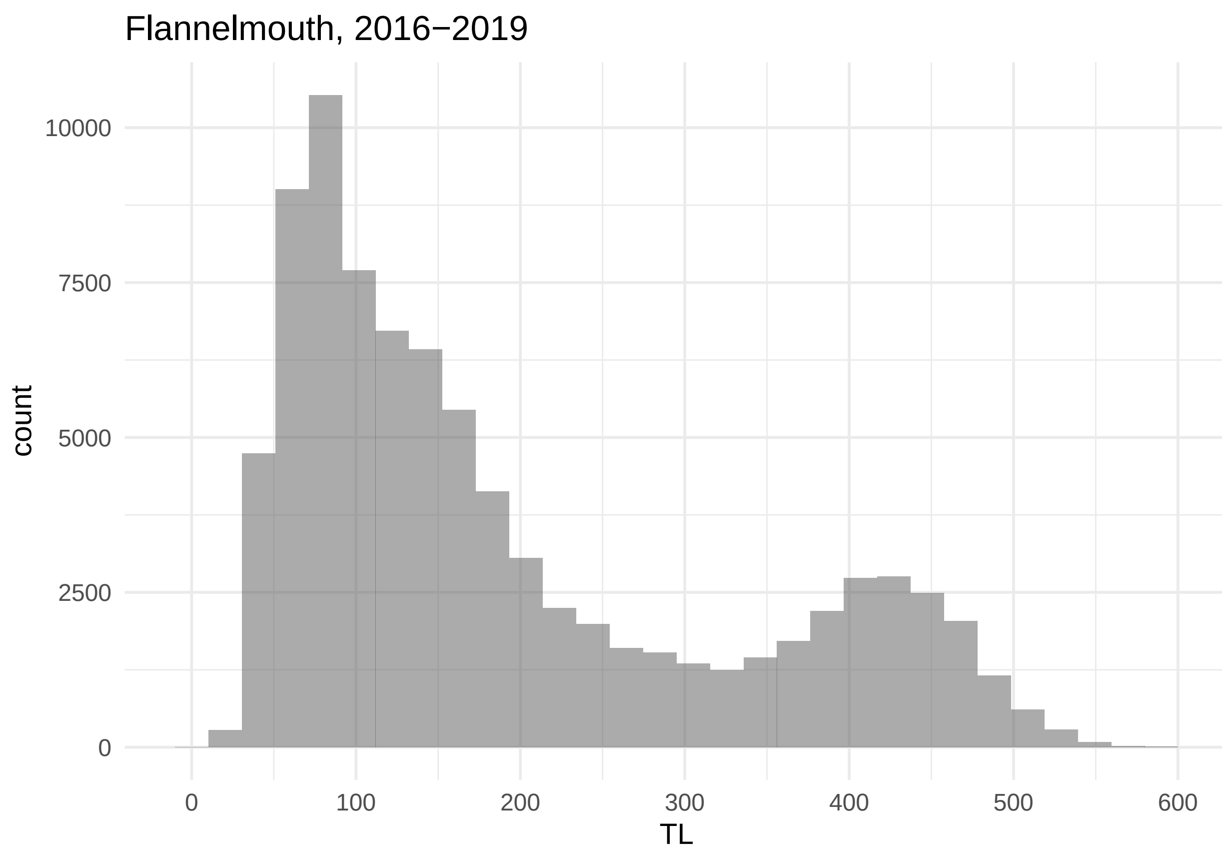

```
#natural break at about 275 mm

#is it different if we restrict to recent years?
fms %>%
  filter(year >= 2013) %>%
  ggplot(aes(TL)) +
  geom_histogram(alpha = 0.5) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



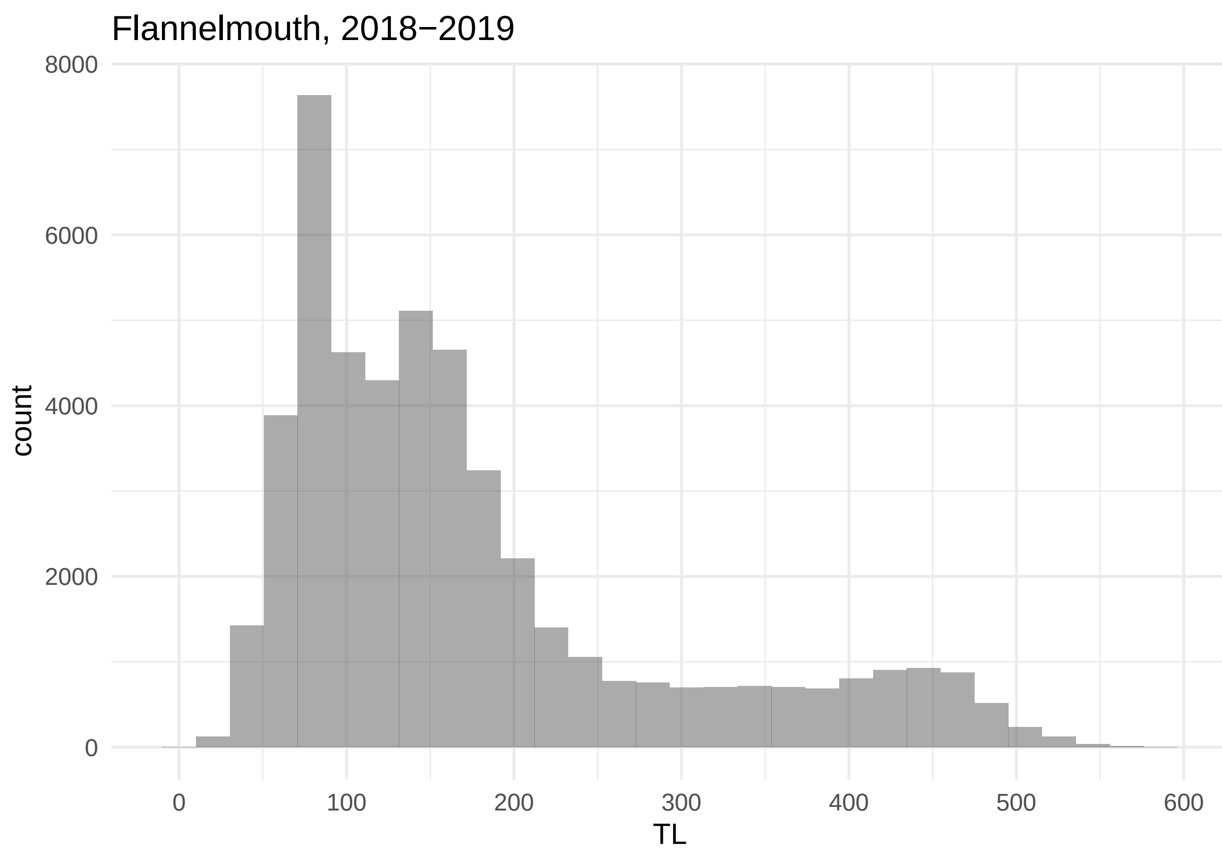

```r
#no difference here

fms %>%
  filter(year >= 2016) %>%
  ggplot(aes(TL)) +
  geom_histogram(alpha = 0.5) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2016-2019")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



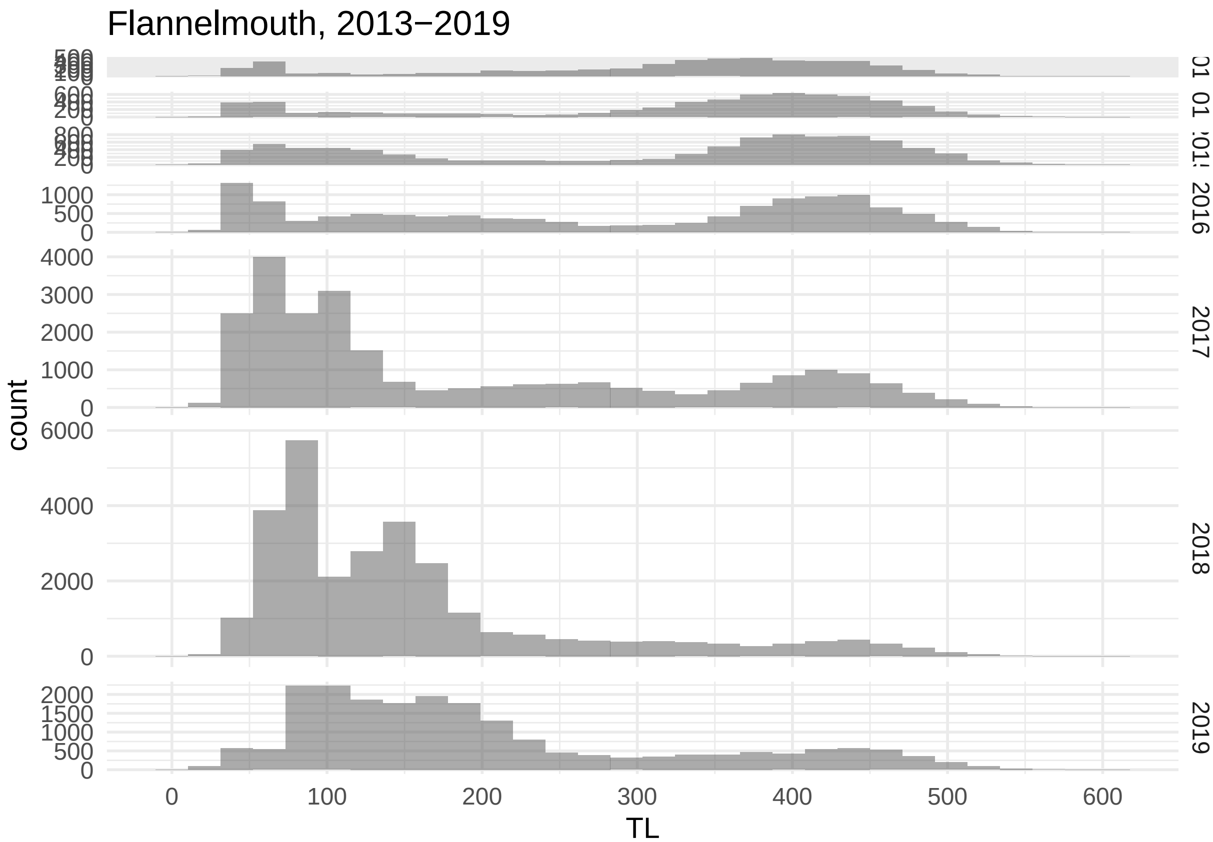

```r
#size break appears to be more like 325 for 2016 - present
#change may be due to increased number of juveniles

fms %>%
  filter(year >= 2018) %>%
  ggplot(aes(TL)) +
  geom_histogram(alpha = 0.5) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2018-2019")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

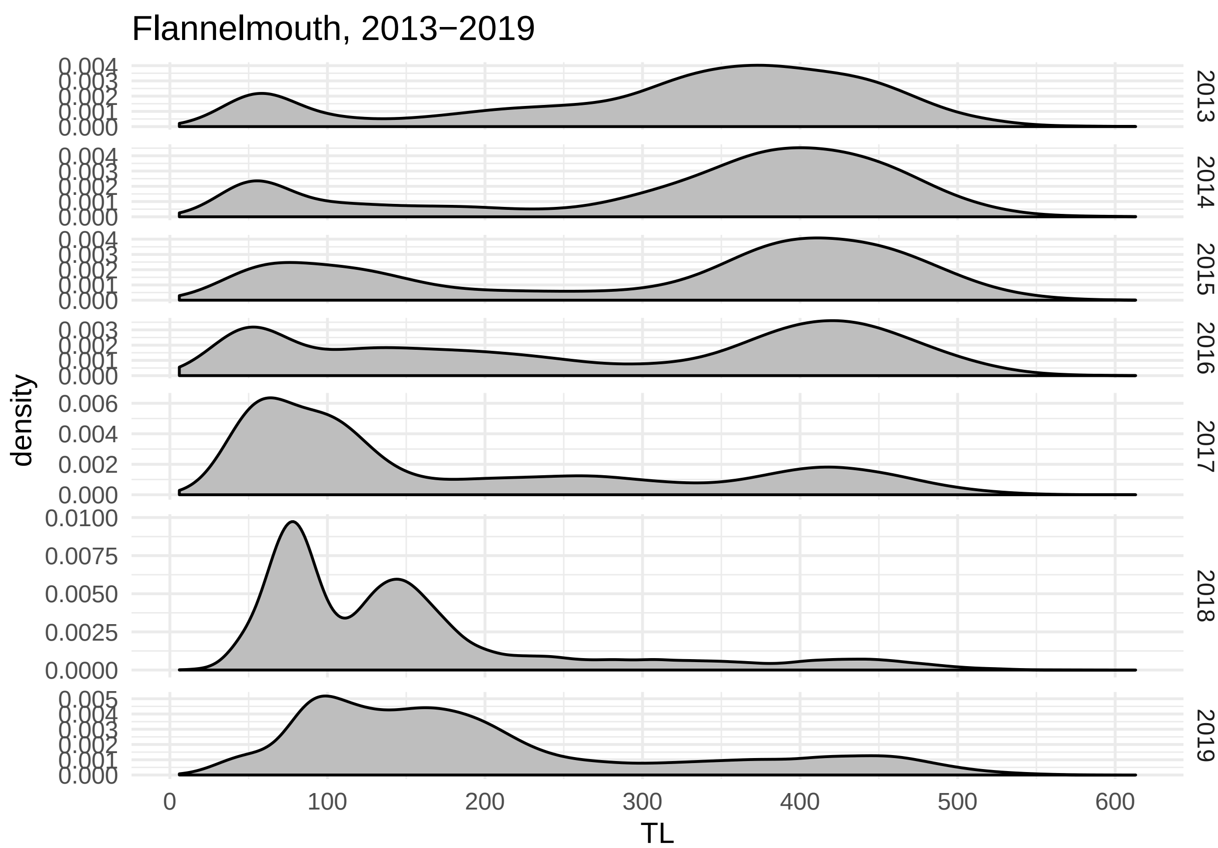



```r
#look at each year separately
fms %>%
  filter(year >= 2013) %>%
  ggplot(aes(TL)) +
  geom_histogram(alpha = 0.5) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019") +
  facet_grid(rows = vars(year), scales = "free", space = "free")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



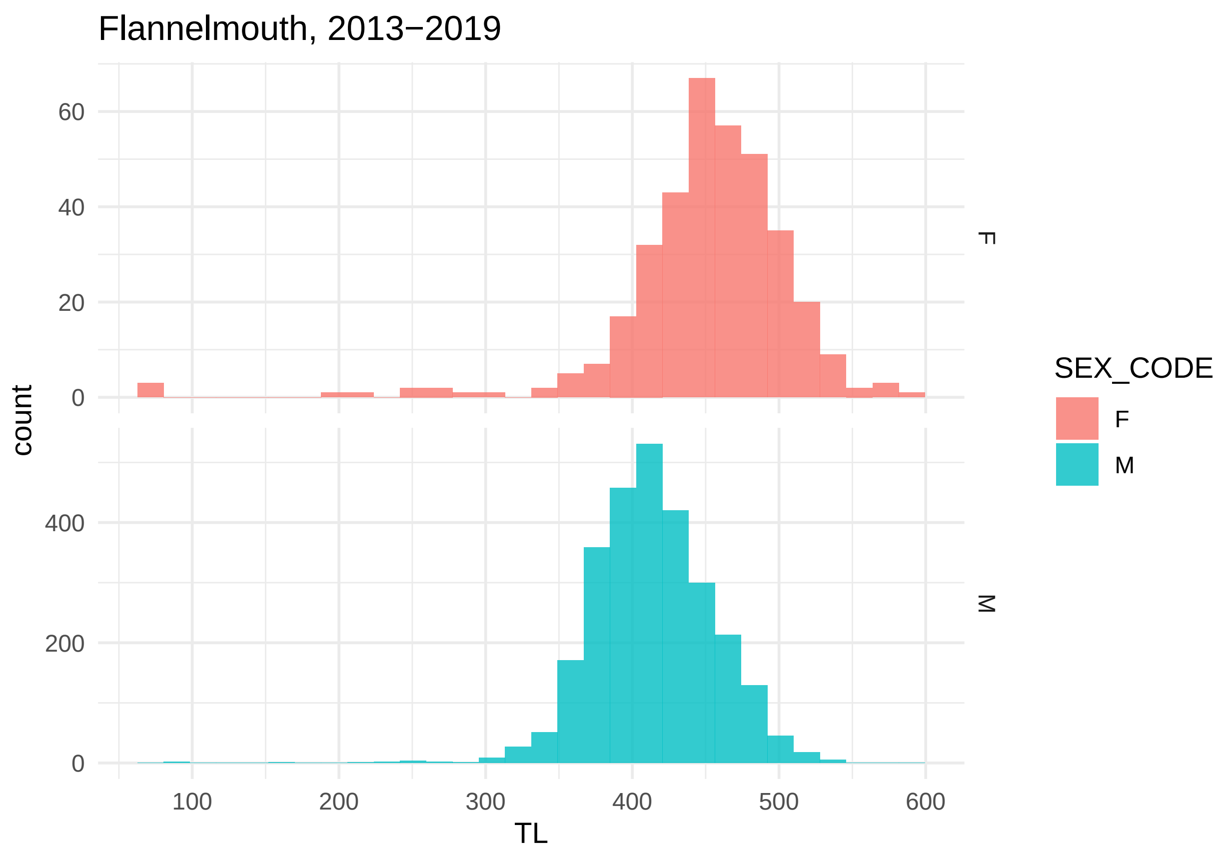

```r
# is huge increase in juveniles in 2017 due to start of JCM west?

#same data on a different graph type
fms %>%
  filter(year >= 2013) %>%
  ggplot(aes(TL)) +
  geom_density(fill = "gray") +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019") +
  facet_grid(rows = vars(year), scales = "free", space = "free")
```



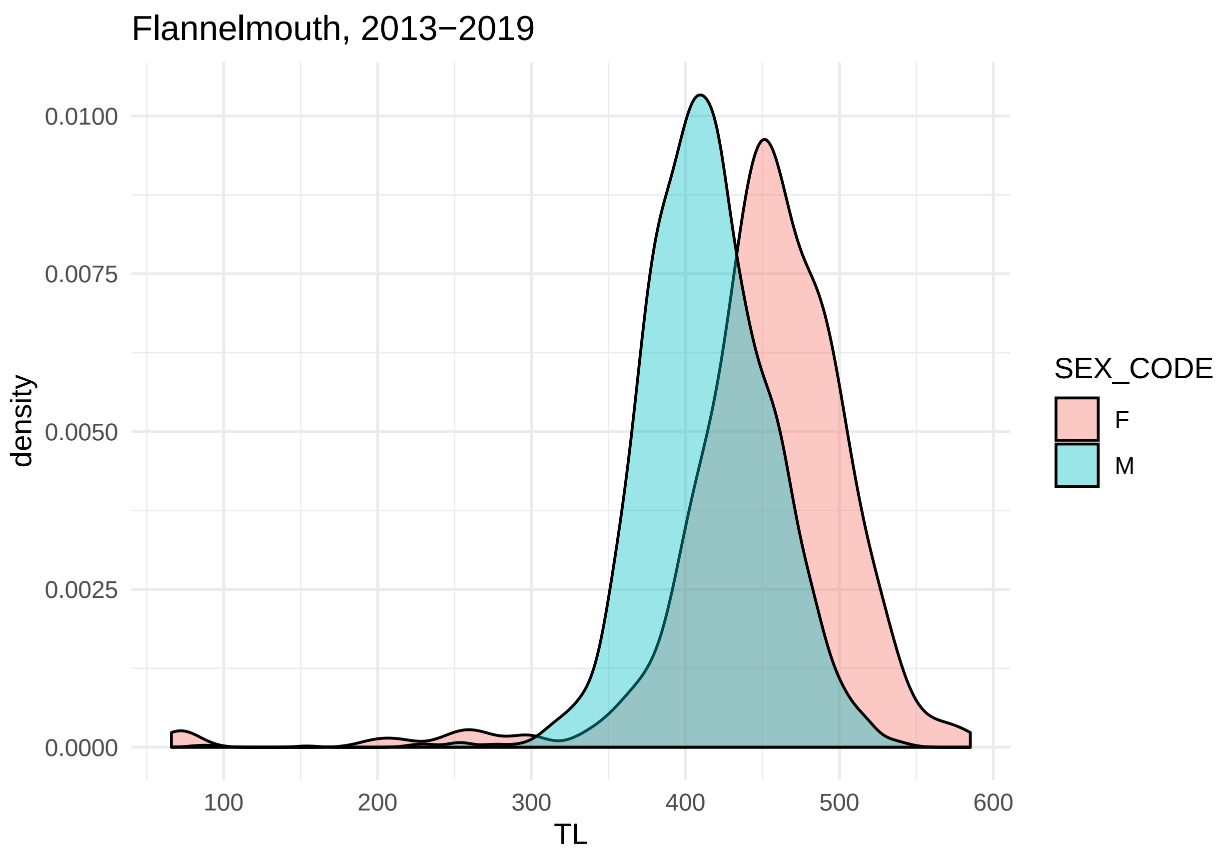

## maturity

```
#maximize sample size:
#include ripe fish and tuberculated or colored fish
fms %>%
  filter(year >= 2013) %>%
  filter(SEX_CODE %in% c("M", "F") &
         (SEX_COND_CODE == "R" |
           SEX_CHAR_CODE %in% c("C", "T", "B"))) %>%
  ggplot(aes(x = TL, fill = SEX_CODE)) +
  geom_histogram(alpha = 0.8) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019")  +
  facet_grid(rows = vars(SEX_CODE), scales = "free")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

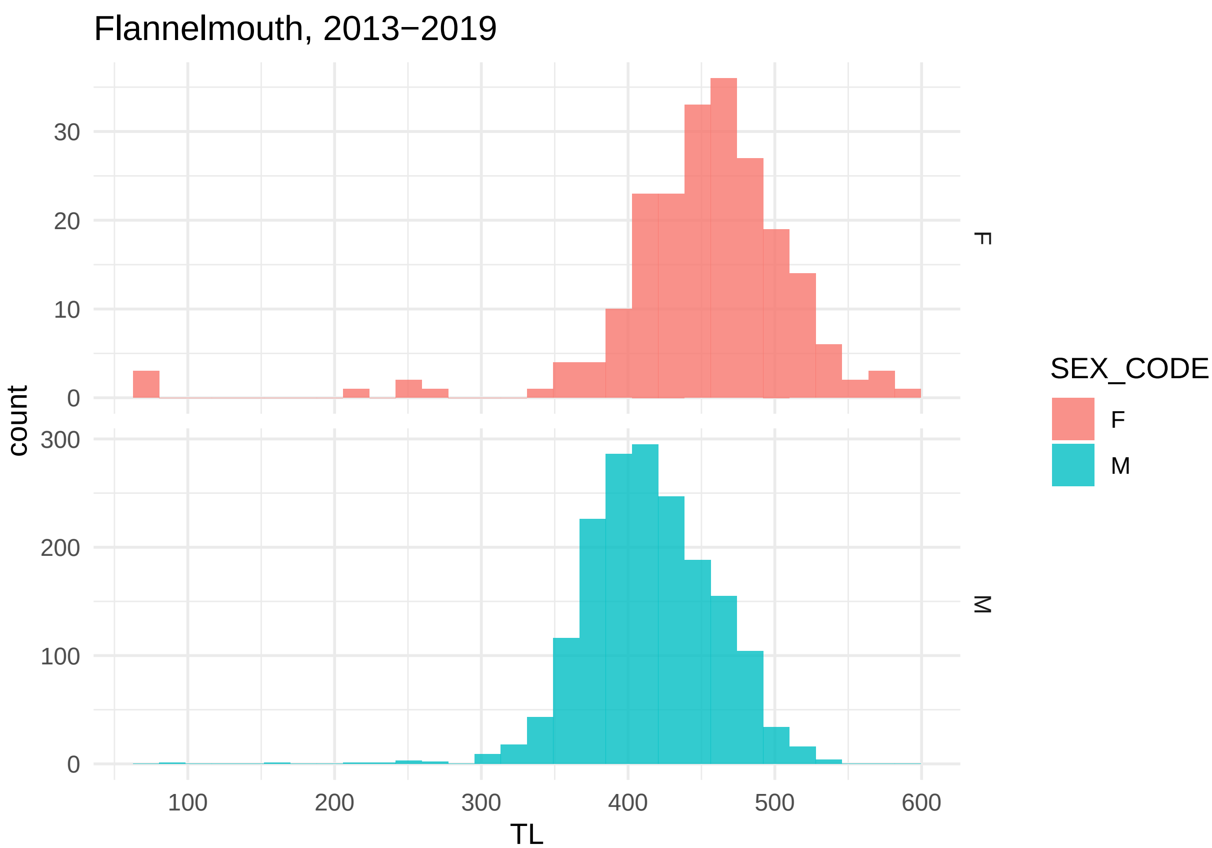



```r
#another way to look at same data
fms %>%
  filter(year >= 2013) %>%
  filter(SEX_CODE %in% c("M", "F") &
           (SEX_COND_CODE == "R" |
              SEX_CHAR_CODE %in% c("C", "T", "B"))) %>%
  ggplot(aes(x = TL, fill = SEX_CODE)) +
  geom_density(alpha = 0.4) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019")
```

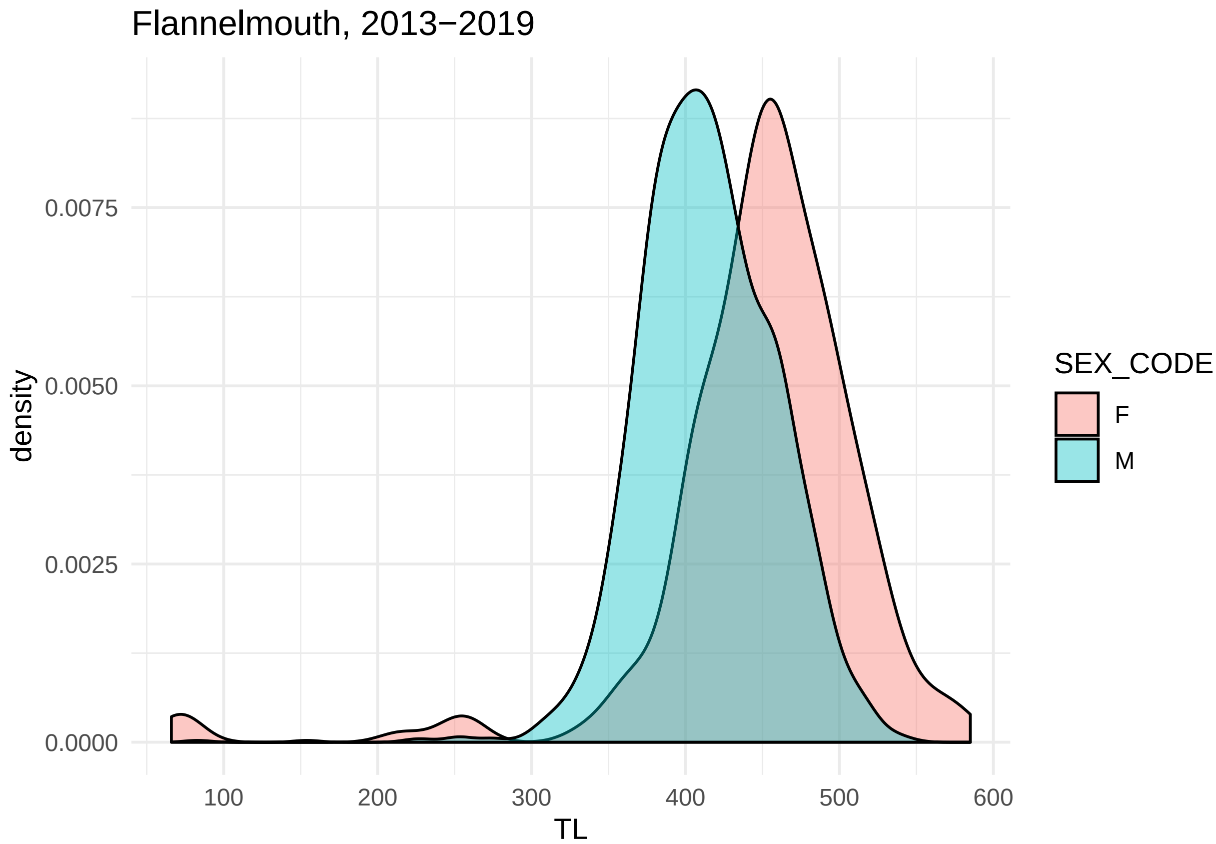



```
#only include ripe fish - will reduce sample size
fms %>%
  filter(year >= 2013) %>%
  filter(SEX_CODE %in% c("M", "F") &
           (SEX_COND_CODE == "R" )) %>%
  ggplot(aes(x = TL, fill = SEX_CODE)) +
  geom_histogram(alpha = 0.8) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019")  +
  facet_grid(rows = vars(SEX_CODE), scales = "free")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
#same data as above, but density plot
fms %>%
  filter(year >= 2013) %>%
  filter(SEX_CODE %in% c("M", "F") &
           (SEX_COND_CODE == "R" )) %>%
  ggplot(aes(x = TL, fill = SEX_CODE)) +
  geom_density(alpha = 0.4) +
  scale_x_continuous(breaks = seq(0, 700, by = 100)) +
  ggtitle("Flannelmouth, 2013-2019")
```




Based on maturity plots, somewhere around 300-350mm would be reasonable to set as size break based on size at reproductive maturity. As expected, males mature at smaller sizes (300mm, vs. 325-350 for females)

length freq histograms are messier, size distribution changes a lot each year. But, the ~325mm from maturity plots seems reasonable based on length freq histograms too. The only other natural break I see is around 125mm for the smallest size class, but these fish don't have PIT tags so we can't analyze a <125mm size class anyway