

Initial cross-system data comparison

2022-04-22

General summary

- 849 fish were tagged across the four groups (coastal Massachusetts and the Hudson, Delaware, and Potomac Rivers) from 2014-2019
- 2016 was the only year where fish were concurrently tagged by each group
- Types of data collected were tag date, transmitter ID, external ID, in-house ID, total length, fork length, wet weight, scale age, sex, general area of tagging
- All tagging groups collected the date of tagging, age, and total length
 - Hudson fish are in the process of being aged
 - 15% Potomac fish have scales collected for ageing, but these have not yet been processed
- Hudson, Delaware, and Potomac River systems have sex data; coastal Massachusetts does not
- Hudson and Potomac have weights; Delaware and Massachusetts do not

```
library(ggplot2); library(data.table)
```

```
all_info <- tar_read(combined_tags)
all_info[, yr := as.factor(year(tagdate))]
all_info[, system := as.factor(system)]
```

```
all_info
```

```
##      tagdate    tl    sex scales    age fishid exttag tagginglocationrkm
##      <Date> <num> <char> <char> <num> <char> <char>                <num>
##  1: 2016-04-15   580      M      x      5  SK005      5                118
##  2: 2016-04-15   656      M      x     NA  SK003      3                118
##  3: 2016-04-15   575      M      x      5  SK004      4                118
##  4: 2016-04-15   950      M      x     11  SK001      1                118
##  5: 2016-04-15   552      M      x      5  SK002      2                118
##  ---
## 845: 2018-04-12  1079      F <NA>    NA <NA> <NA>                NA
## 846: 2018-04-12  1003      F <NA>    NA <NA> <NA>                NA
## 847: 2018-04-12  1041      F <NA>    NA <NA> <NA>                NA
## 848: 2018-04-12  1067      F <NA>    NA <NA> <NA>                NA
## 849: 2018-04-12   978      F <NA>    NA <NA> <NA>                NA
##      taglocation    transmitter    system actagsn finclip    notes    lat    lon
##      <char>          <char>    <fctr>    <num>    <char> <char> <char> <char>
##  1: Lower Zone A69-1601-54015      DE      NA <NA> <NA> <NA> <NA>
##  2: Lower Zone A69-1601-54016      DE      NA <NA> <NA> <NA> <NA>
##  3: Lower Zone A69-1601-54017      DE      NA <NA> <NA> <NA> <NA>
##  4: Lower Zone A69-1601-54018      DE      NA <NA> <NA> <NA> <NA>
##  5: Lower Zone A69-1601-54019      DE      NA <NA> <NA> <NA> <NA>
##  ---
```

```
## 845: Pt Lookout A69-9002-6770 Potomac NA <NA> <NA> <NA> <NA>
## 846: Pt Lookout A69-9002-6769 Potomac NA <NA> <NA> <NA> <NA>
## 847: Pt Lookout A69-9002-6774 Potomac NA <NA> <NA> <NA> <NA>
## 848: Pt Lookout A69-9002-6773 Potomac NA <NA> <NA> <NA> <NA>
## 849: Pt Lookout A69-9002-6772 Potomac NA <NA> <NA> <NA> <NA>
##      wgt      yr
##      <num> <fctr>
## 1:      NA 2016
## 2:      NA 2016
## 3:      NA 2016
## 4:      NA 2016
## 5:      NA 2016
## ---
## 845: 17700 2018
## 846: 13200 2018
## 847: 16800 2018
## 848: 15600 2018
## 849: 13000 2018
```

Systems

```
addmargins(xtabs(~ system, addNA = T, data = all_info))
```

```
## system
##      DE Hudson MA Coast Potomac Sum
##      298      100      272      179 849
```

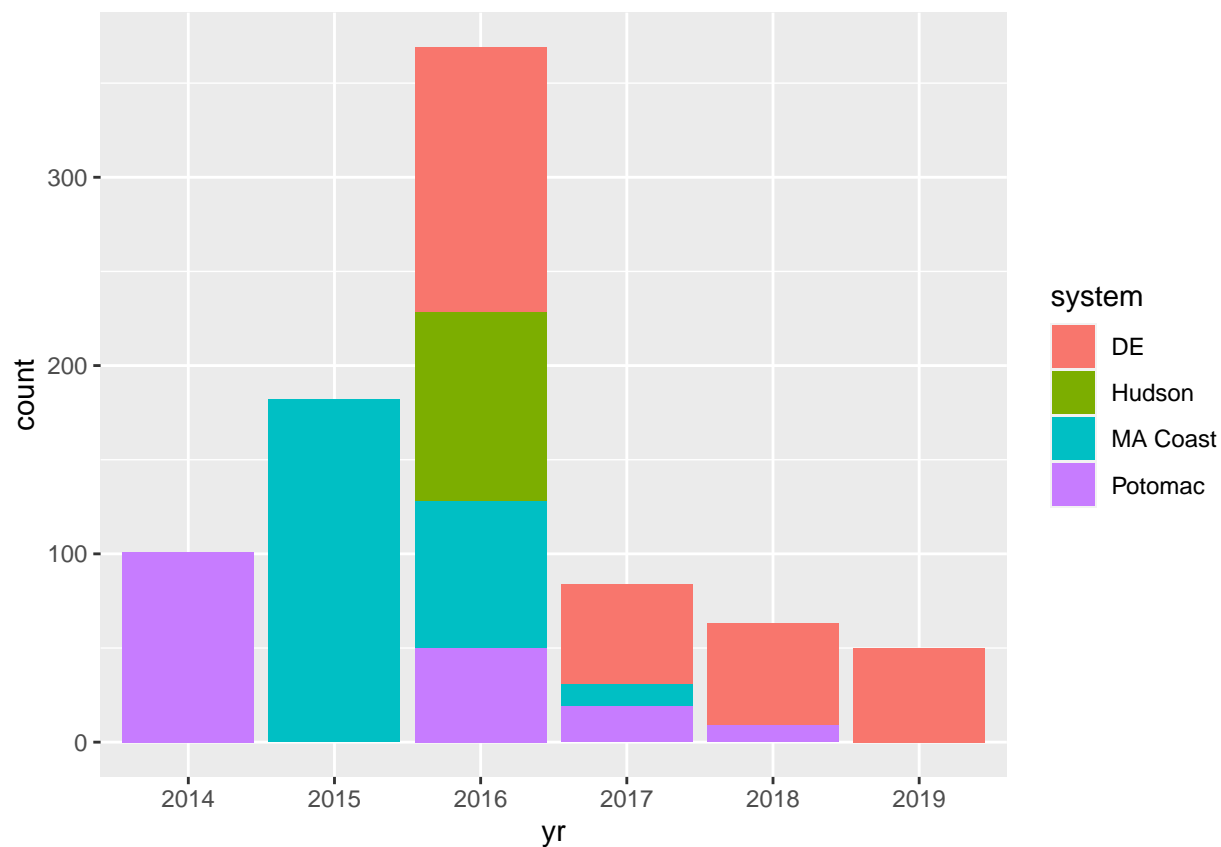
Year tagged

- Potomac tags skew older, Delaware tags skew younger, Massachusetts tags are in the middle, and Hudson were only tagged in 2016
- 2016 is the only year that had fish tagged in all systems

```
addmargins(xtabs(~ yr + system, addNA = T, data = all_info))
```

```
##      system
## yr      DE Hudson MA Coast Potomac Sum
## 2014    0      0      0      101 101
## 2015    0      0     182       0 182
## 2016  141     100      78      50 369
## 2017   53      0      12      19  84
## 2018   54      0       0       9  63
## 2019   50      0       0       0  50
## Sum  298     100     272     179 849
```

```
ggplot(data = all_info) +
  geom_bar(aes(x = yr, fill = system))
```



Age

- A subset of Potomac-tagged fish have not yet been, but are planning to be, aged. There will likely wind up being 1 Hudson River *NA* and 1 Potomac River *NA*.

```
# Cross-tabulation
addmargins(xtabs(~ age + system, addNA = T, data = all_info))
```

```
##      system
## age   DE Hudson MA Coast Potomac Sum
##  3      0      0     10      5   15
##  4      8      0     41     10   59
##  5     19      0     56      7   82
##  6     22      4     38     26   90
##  7     30     10     33     38  111
##  8     23     15     23     21   82
##  9     23     23     21     14   81
## 10     39     20      9     16   84
## 11     23     19      8      9   59
## 12     19      4      8      2   33
## 13     21      4      5      3   33
## 14     14      0      1      0   15
## 15      7      0      0      0    7
## 16      3      0      0      0    3
```

```
##   17     1     0     0     0   1
##   18     2     0     0     0   2
##   19     1     0     0     0   1
##   <NA> 43     1    19    28  91
##   Sum 298    100    272    179 849
```

```
# Range of ages
all_info[, .(min = min(age, na.rm = T),
               median = median(age, na.rm = T),
               max = max(age, na.rm = T)),
          by = system]
```

```
##      system   min median   max
##      <fctr> <num>  <num> <num>
## 1:      DE      4     10    19
## 2: MA Coast      3      6    14
## 3: Potomac      3      7    13
## 4:  Hudson      6      9    13
```

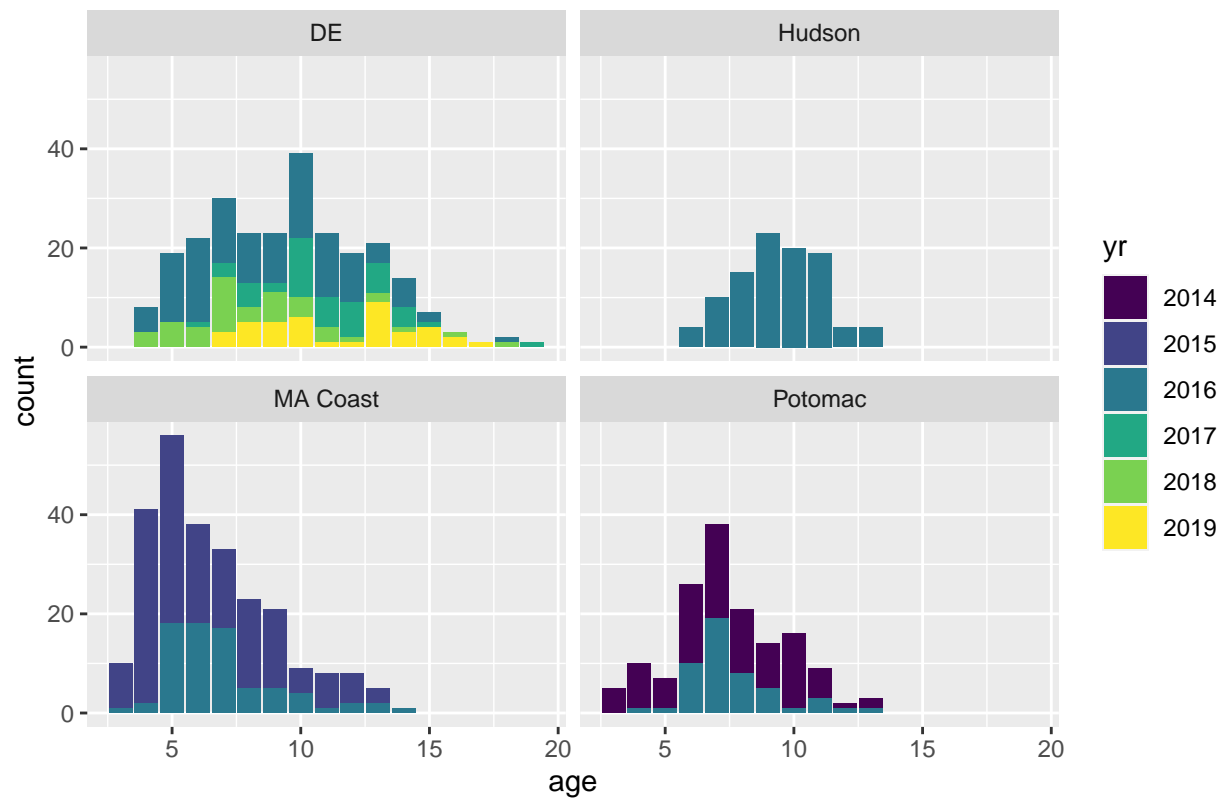
- Delaware fish are oldest, followed by Hudson, then Potomac, then Massachusetts.

```
TukeyHSD(aov(age ~ system, data = all_info))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = age ~ system, data = all_info)
##
## $system
##              diff          lwr          upr          p adj
## Hudson-DE      -0.1158645 -0.8981335  0.6664044 0.9811159
## MA Coast-DE     -2.9524917 -3.5386888 -2.3662946 0.0000000
## Potomac-DE      -2.0396052 -2.7179440 -1.3612665 0.0000000
## MA Coast-Hudson -2.8366271 -3.6197603 -2.0534940 0.0000000
## Potomac-Hudson  -1.9237407 -2.7780322 -1.0694492 0.0000001
## Potomac-MA Coast 0.9128864  0.2335512  1.5922216 0.0031946
```

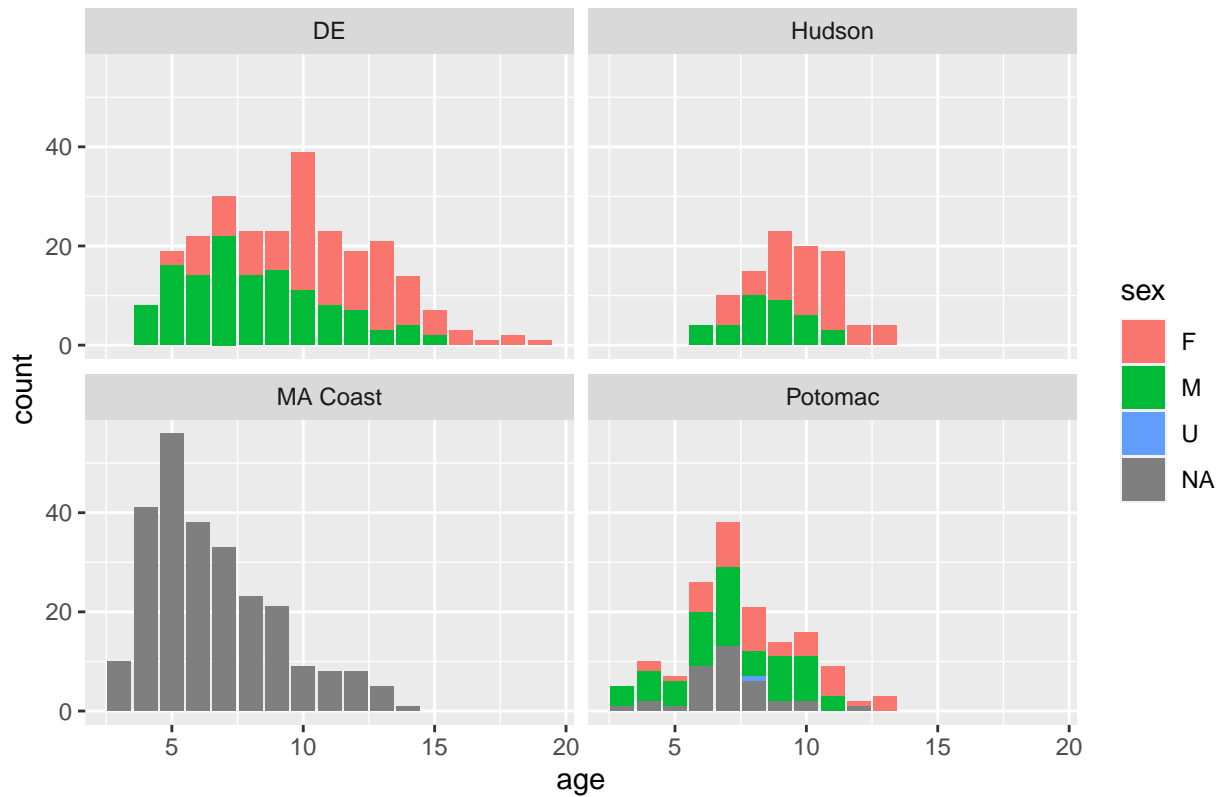
```
ggplot(data = all_info[!is.na(age)]) +
  geom_bar(aes(x = age, fill = yr)) +
  scale_fill_viridis_d() +
  facet_wrap(~ system) +
  labs(title = 'Age by year and system tagged')
```

Age by year and system tagged



```
ggplot(data = all_info[!is.na(age)]) +
  geom_bar(aes(x = age, fill = sex)) +
  facet_wrap(~ system) +
  labs(title = 'Age by system tagged and sex')
```

Age by system tagged and sex



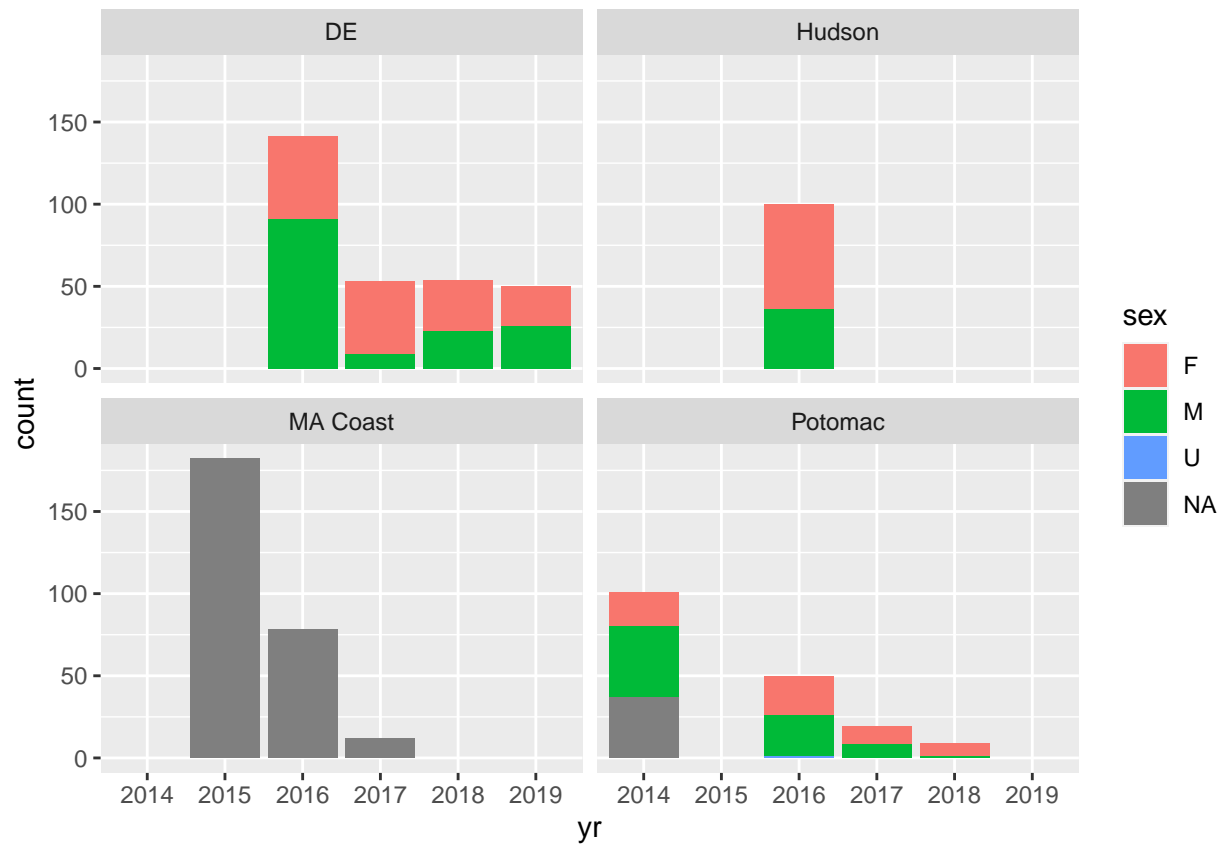
Sexes

Fish collected off the Massachusetts coast and 21% of Potomac fish were not sexed. “NA” means that sex was not recorded, while “U” means that the sex was not able to be determined.

```
addmargins(xtabs(~ sex + system, addNA = T, data = all_info))
```

```
##      system
## sex    DE Hudson MA Coast Potomac Sum
## F     149     64      0      64 277
## M     149     36      0      77 262
## U       0      0      0       1  1
## <NA>    0      0    272     37 309
## Sum   298    100    272    179 849
```

```
ggplot(data = all_info) +
  geom_bar(aes(x = yr, fill = sex)) +
  facet_wrap(~ system)
```

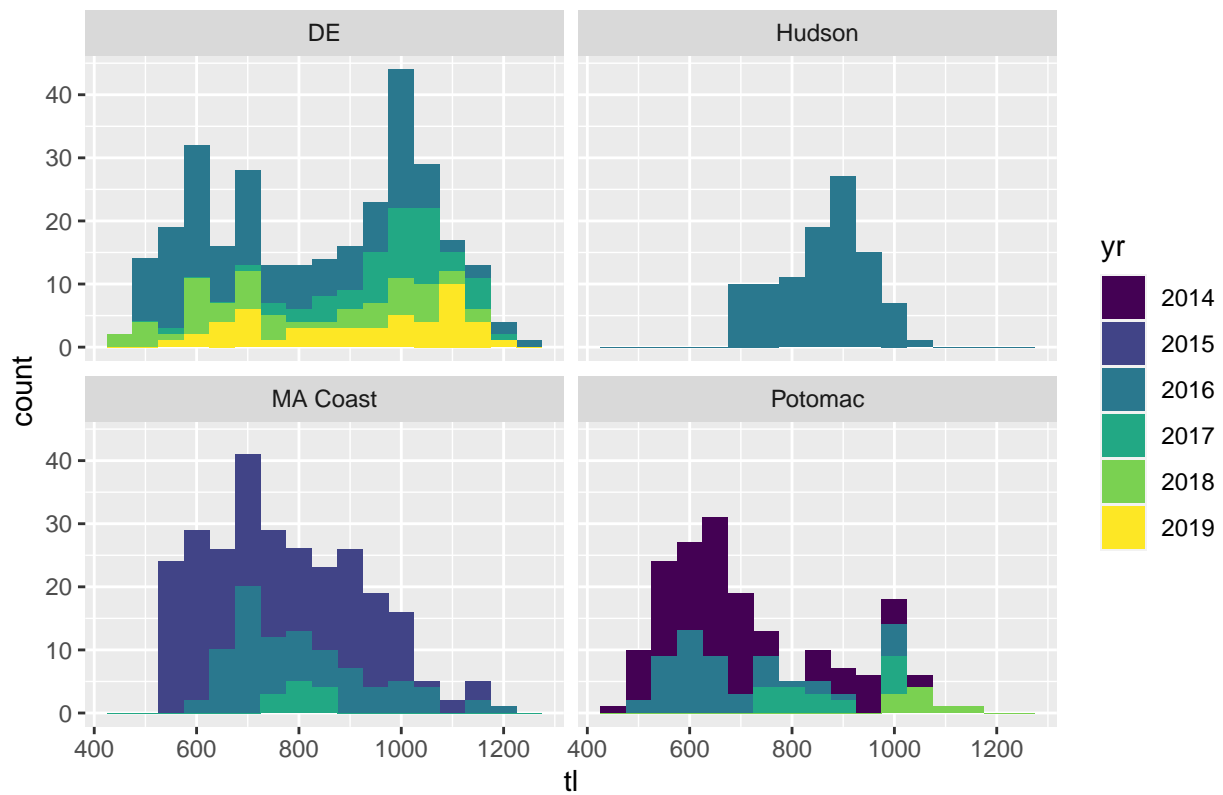


Lengths

All fish collected have total lengths, while only Hudson fish have fork lengths. Seems TL is the way to go here.

```
ggplot(data = all_info) +
  geom_histogram(aes(x = tl, fill = yr), binwidth = 50) +
  scale_fill_viridis_d() +
  facet_wrap(~ system) +
  labs(title = 'Total length by year and tagged')
```

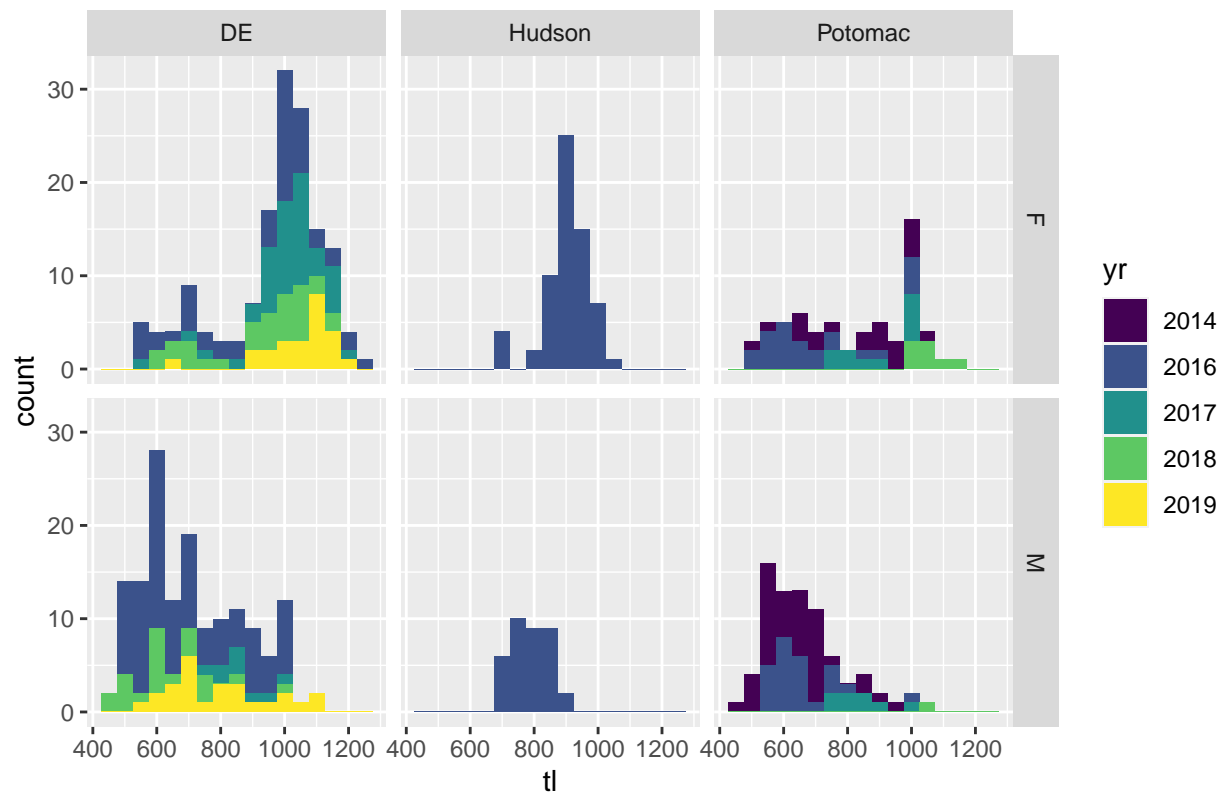
Total length by year and tagged



- Potomac and MA coast fish skew smaller, Hudson skews larger, and DE is bimodal (related in part to sex, see below)

```
ggplot(data = all_info[!is.na(sex) & sex != 'U']) +
  geom_histogram(aes(x = tl, fill = yr), binwidth = 50) +
  scale_fill_viridis_d() +
  facet_grid(sex ~ system) +
  labs(title = 'Total length by sex and year and system tagged')
```


Total length by sex and year and system tagged



Weights

Only Potomac and Hudson fish have recorded weights, with Potomac fish skewing much lighter.

```
all_info[!is.na(wgt), .N, by = system]
```

```
##      system      N
##      <fctr> <int>
## 1: Potomac   178
## 2: Hudson    97
```

```
ggplot(data = all_info[!is.na(wgt)]) +
  geom_histogram(aes(x = wgt, fill = yr), binwidth = 1000) +
  scale_fill_viridis_d() +
  facet_wrap(~ system)
```

