

# Initial cross-system data comparison

2022-01-20

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

```
all_info <- fread('embargo/derived/combined_tag_info.csv',  
                 na.strings = '')  
all_info[, yr := as.factor(year(tagdate))]
```

```
all_info
```

```
##      tagdate    tl    sex scales    age fishid exttag tagginglocationrkm  
##      <IDat> <int> <char> <char> <int> <char> <char>                <int>  
##  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>    <char>    <int>    <char> <char> <num> <num>  
##  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  
##      <int> <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

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

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

## Year tagged

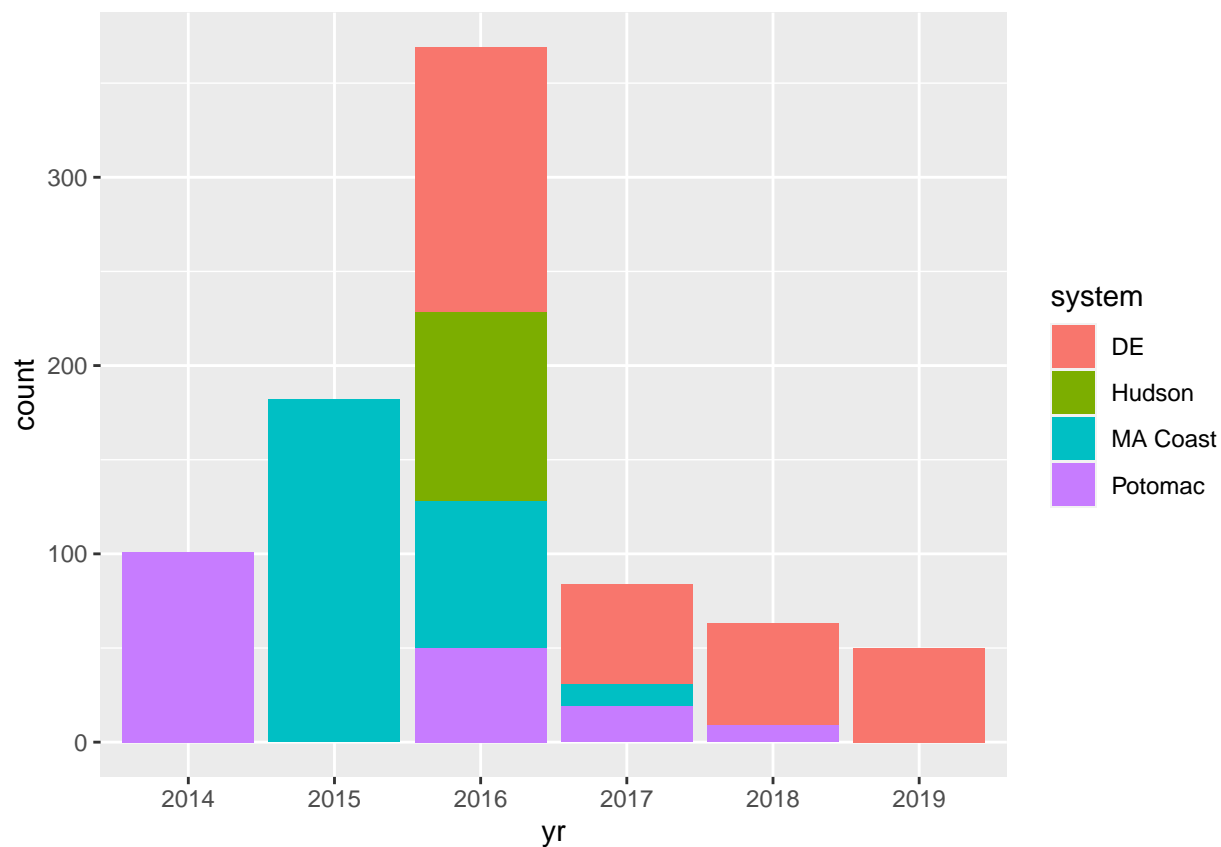
```
# Cross-tabulation
addmargins(xtabs(~ yr + system, addNA = T, data = all_info))
```

```
##      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
```

```
# Dirty chi-squared test
summary(xtabs(~ yr + system, addNA = T, data = all_info))
```

```
## Call: xtabs(formula = ~yr + system, data = all_info, addNA = T)
## Number of cases in table: 849
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 1066.7, df = 15, p-value = 6.645e-218
```

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



## Age

Hudson River and a subset of Potomac origin fish have not yet been, but are planning to be, aged.

```
# 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      0    38     26   86
## 7    30      0    33     38  101
## 8    23      0    23     21   67
## 9    23      0    21     14   58
## 10   39      0     9     16   64
## 11   23      0     8      9   40
## 12   19      0     8      2   29
## 13   21      0     5      3   29
## 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    100    19     28  190
```

```
## Sum 298 100 272 179 849
```

```
# Dirty chi-squared test
```

```
summary(xtabs(~ age + system, addNA = T, data = all_info))
```

```
## Call: xtabs(formula = ~age + system, data = all_info, addNA = T)
```

```
## Number of cases in table: 849
```

```
## Number of factors: 2
```

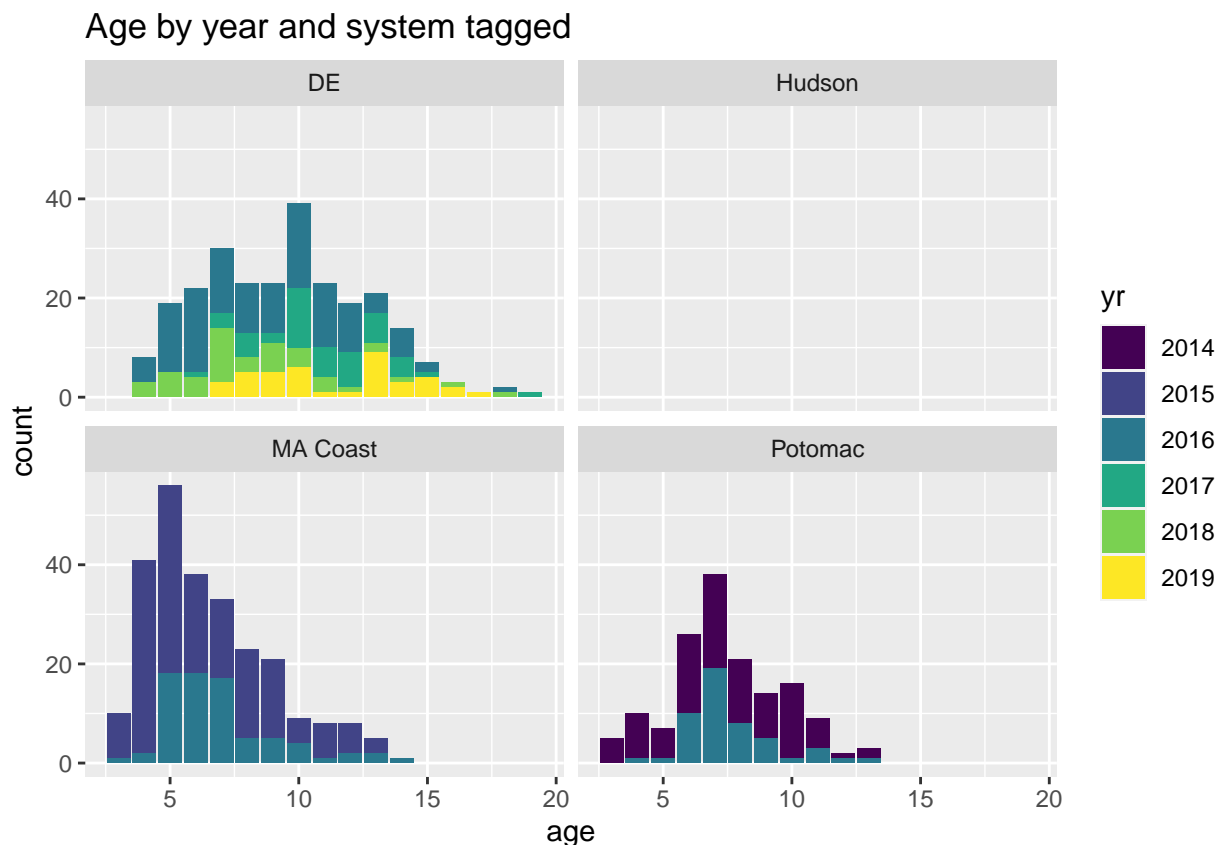
```
## Test for independence of all factors:
```

```
## Chisq = 602, df = 51, p-value = 3.55e-95
```

```
## Chi-squared approximation may be incorrect
```

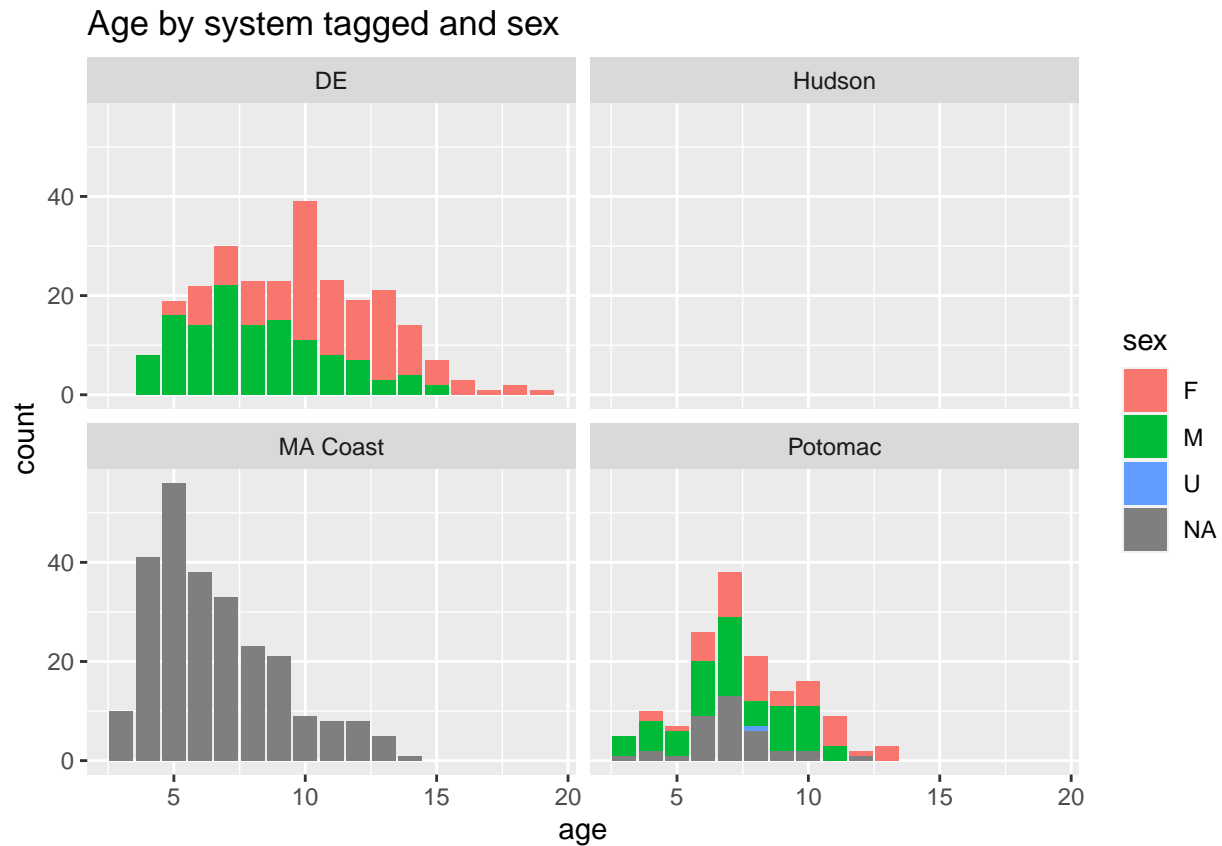
```
ggplot(data = all_info) +  
  geom_bar(aes(x = age, fill = yr)) +  
  scale_fill_viridis_d() +  
  facet_wrap(~ system) +  
  labs(title = 'Age by year and system tagged')
```

```
## Warning: Removed 190 rows containing non-finite values (stat_count).
```



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

```
## Warning: Removed 190 rows containing non-finite values (stat_count).
```



## Sexes

Fish collected off the Massachusetts coast and 21% of Potomac fish were not sexed.

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

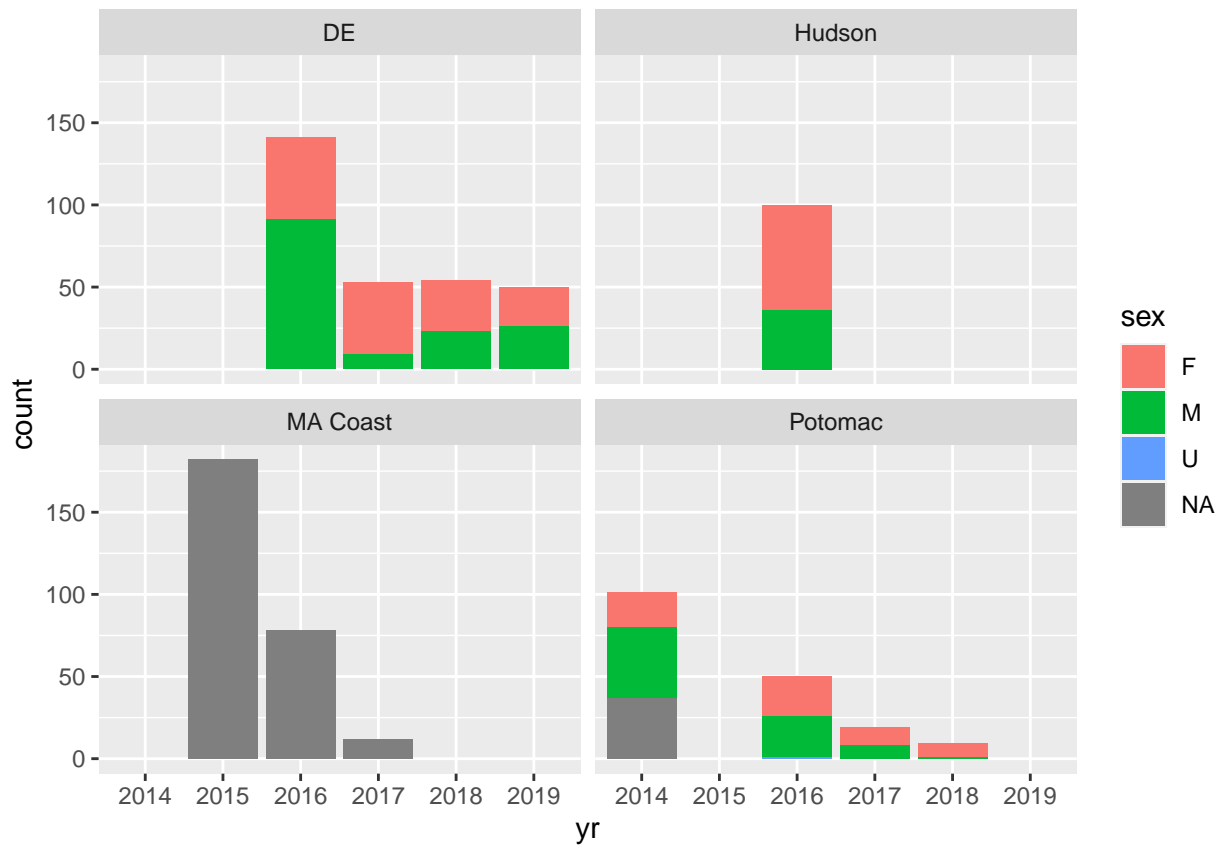
```
##      system
## sex    DE Hudson MA Coast Potomac Sum
##  F     149     64      0      64 277
##  M     149     36      0      77 262
##  U       0      0      0       1   1
##  <NA>    0      0    272     37 309
##  Sum   298    100    272    179 849
```

```
# Dirty chi-squared test
summary(xtabs(~ sex + system, addNA = T, data = all_info))
```

```
## Call: xtabs(formula = ~sex + system, data = all_info, addNA = T)
## Number of cases in table: 849
## Number of factors: 2
```

```
## Test for independence of all factors:
## Chisq = 738.9, df = 9, p-value = 3.047e-153
## Chi-squared approximation may be incorrect
```

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

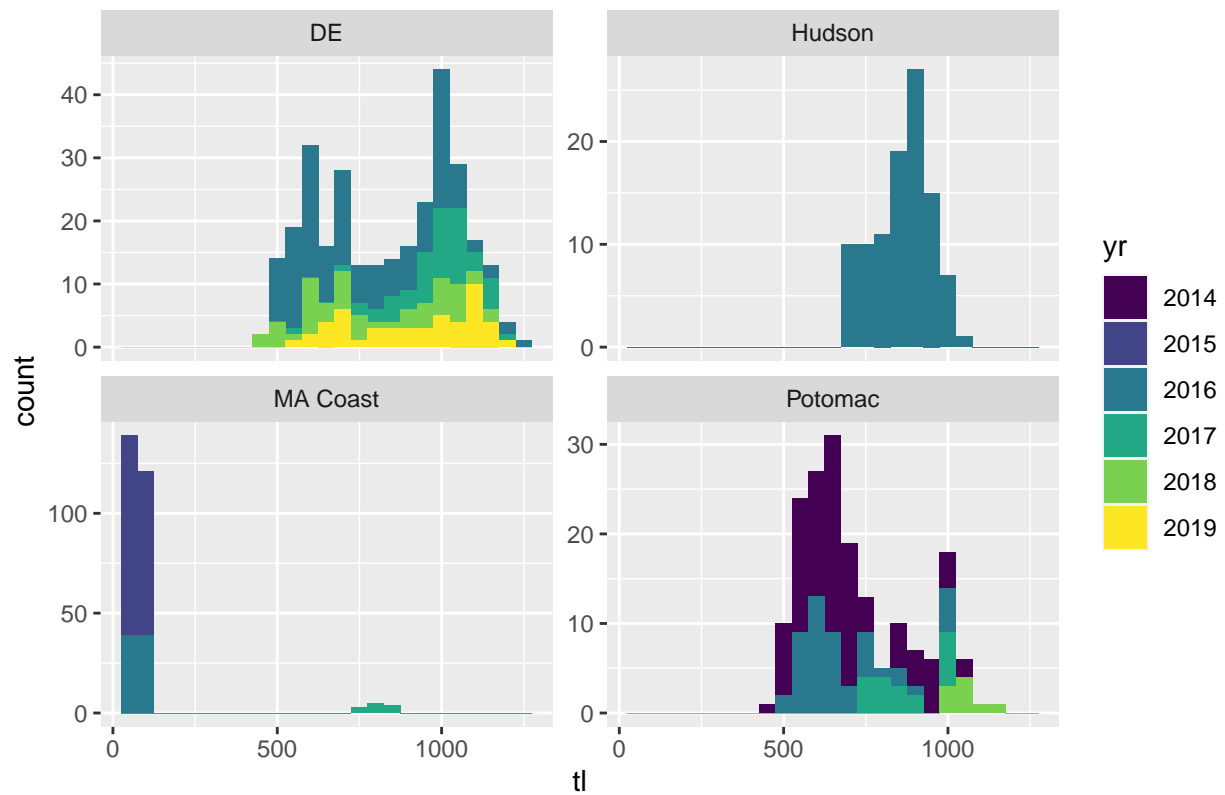


## Lengths

All fish collected have total lengths. Only Hudson fish have fork lengths.

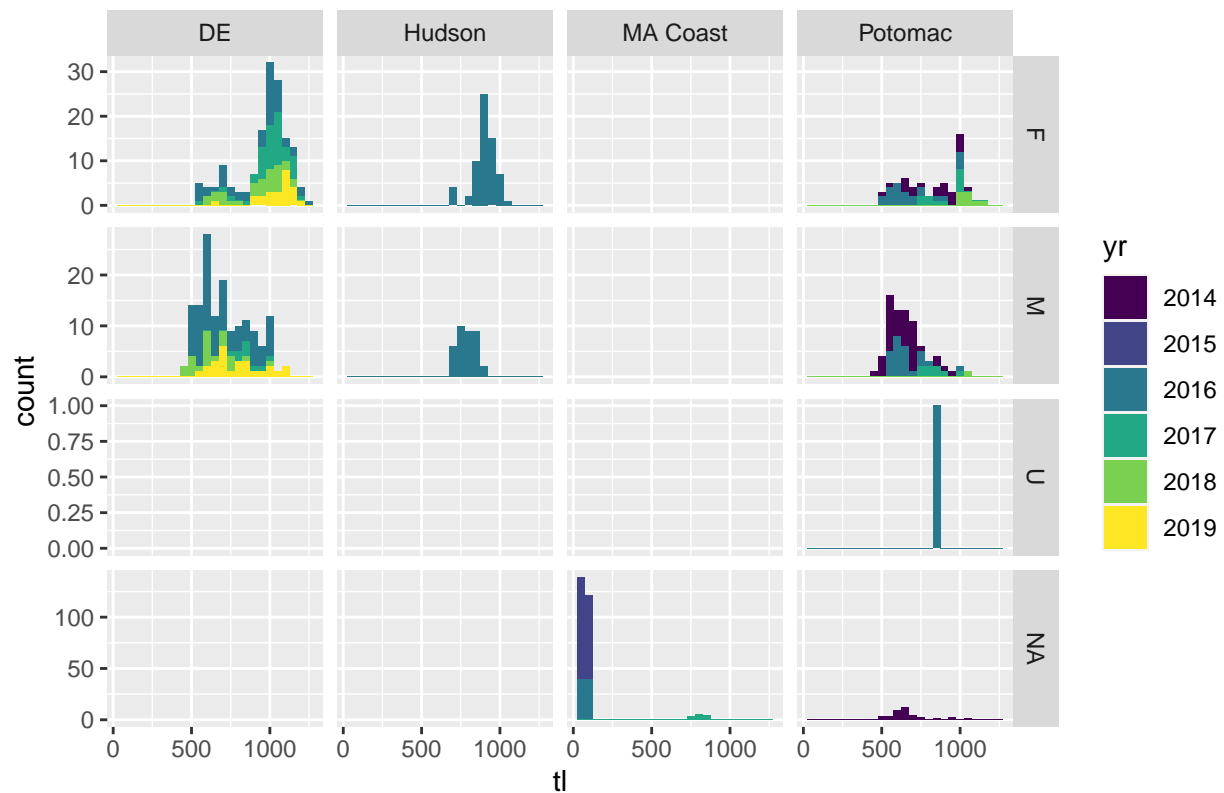
```
ggplot(data = all_info) +
  geom_histogram(aes(x = tl, fill = yr), binwidth = 50) +
  scale_fill_viridis_d() +
  facet_wrap(~ system, scales = 'free_y') +
  labs(title = 'Total length by year and tagged')
```

Total length by year and tagged



```
ggplot(data = all_info) +
  geom_histogram(aes(x = tl, fill = yr), binwidth = 50) +
  scale_fill_viridis_d() +
  facet_grid(sex ~ system, scales = 'free_y') +
  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.

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

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

```
ggplot(data = all_info) +
  geom_histogram(aes(x = wgt, fill = yr), binwidth = 1000) +
  scale_fill_viridis_d() +
  facet_wrap(~ system, scales = 'free_y')
```

```
## Warning: Removed 574 rows containing non-finite values (stat_bin).
```



