

Exploration for Data Filtering HPE

Anna Steel

August 10, 2016

Read in Data

- Open script from Fremont16.Rproj in GitHub to ensure directories are correct

```
## [1] "Prior to filtering, 130665 total positions in dataset"  
## [1] "Prior to filtering, 438 individual fish positioned"
```

Initial Data Cleaning

- remove excessively high HPEs (<1000) for ease later
- reduce dataset to only positions recorded after full VPS array was in place (don't yet have removal dates, so not incorporated here)
- Create points in a spatial dataframe as well (both lat-long WGS84 & UTM)
- other steps to include here? – perhaps the tags with ID > 65000 are not fish tags for our study; the VEMCO report only considered those with tagIDs in the 36xxx, 38xxx, and 39xxx series; when we remove the 5 tags in the 65xxx series our initial values for n fish tags and n positions match those in the vemco report

```
## [1] "Removing 0 positions with HPEs > 1000"  
## [1] "0 positions removed due to incomplete VPS array"
```

Explore HPEs relationship with measured error for sync and reference tags

- Only positions with Hpes<1000
- Uses XY coordinates from VEMCO (Azimuthal Equal Area Projection)

```
## [1] 2.68366822228912  
  
## pdf  
## 2  
  
## [1] "At threshold of HPEs<1, mean measured error (m) of sync/ref positions = 3.7"  
## [1] "At threshold of HPEs<1, median measured error (m) of sync/ref positions = 2.44"  
## [1] "At threshold of HPEs<1, 90%ile of measured error (m) of sync/ref positions = 7.78"  
## [1] "At threshold of HPEs<1, 83.72% of sync/ref positions retained"
```

Based on these graphics I will use a HPE filtering threshold of HPE=1; this seems to be a nice tradeoff between precision (mean ~ 5m in HPE bin 0.75-1) and quantity of data (~80% of data retained).

Spatial distribution of HPEs in fish positions (Similiar to Fig 9 in 2015 report)

- Eventually will add river line to these plots but needs more research to do in ggplot

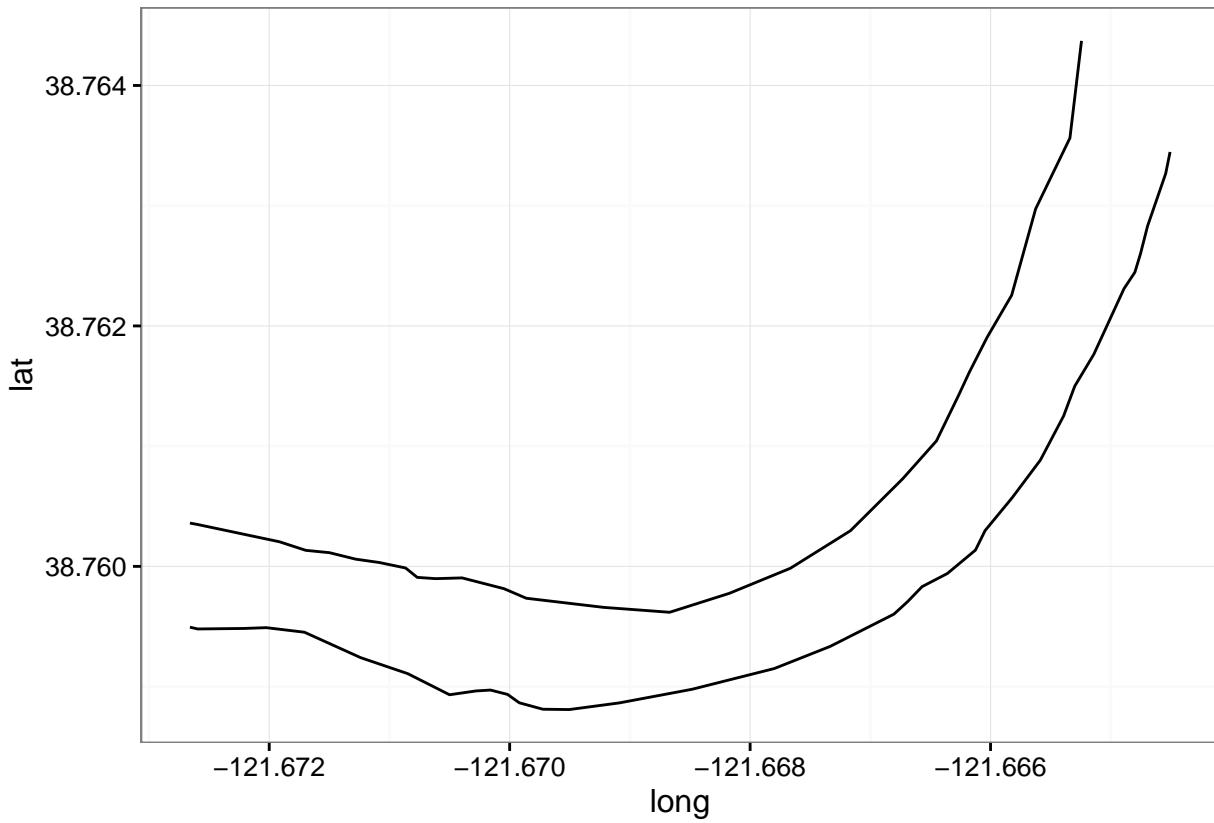
```
## OGR data source with driver: ESRI Shapefile
## Source: "GIS/2004_channel", layer: "2004_channel_frtightclipWGS84"
## with 2 features
## It has 1 fields

## Warning: Non Lab interpolation is deprecated
## Warning: Non Lab interpolation is deprecated
## Warning: Non Lab interpolation is deprecated

## pdf
## 2
```

**code pulled from 2015 to plot density contours of removed positions
(based on HPE>0.5)**

```
ggplot(ggriverbank) + aes(long,lat,group=group, fill = id) +
  geom_path(color="black") +
  coord_equal() + theme_bw()
```



```

# create the constant theme and plots
roundfunc <- function(l) { (l <- round(l, 3)) }
commonTheme = list(labs(color="Density",fill="Density",
    x="Longitude", y="Latitude"),
    theme_bw(),
    theme(legend.position=c(0,1),
        legend.justification=c(0,1)))

ptdens.keep = ggplot(data=alldf[alldf$Hpes<0.5,],
    aes(x=Longitude,y=Latitude)) +
    geom_point(alpha=.1, color="grey80") +
    stat_density2d(aes(fill=..level..,alpha=..level..),
        geom='polygon',colour='black') +
    scale_fill_continuous(low="white",high="blue") +
    guides(alpha="none") + coord_fixed(ratio=1) +
    geom_path(data = ggriverbank, aes(long, lat, group=group)) +
    scale_y_continuous(label=roundfunc) +
    scale_x_continuous(label=roundfunc) +
    commonTheme

ptdens.remove = ggplot(data=alldf[alldf$Hpes>=0.5 & alldf$Hpes<500,],
    aes(Longitude, Latitude)) +
    geom_point(alpha=.1, color="grey80") +
    stat_density2d(aes(fill=..level..,alpha=..level..),
        geom='polygon',colour='black') +

```

```

    scale_fill_continuous(low="white",high="red") +
    guides(alpha="none") + coord_fixed(ratio=1) +
    geom_path(data = ggriverbank, aes(long, lat, group=group)) +
    scale_y_continuous(label=roundfunc,
                        limits=range(ggriverbank$lat)) +
    scale_x_continuous(label=roundfunc,
                        limits=range(ggriverbank$long)) +
    commonTheme

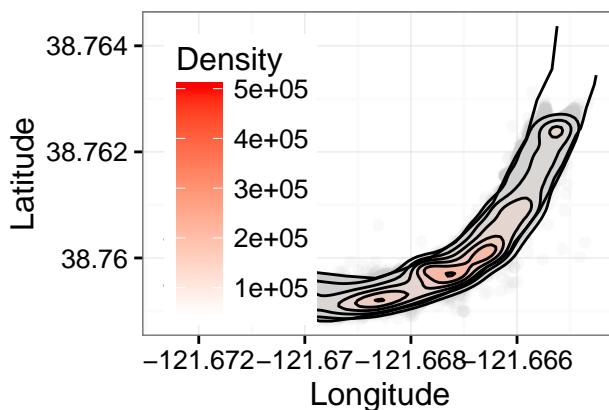
# print plots in proper grid
windows()
plot_grid(ptdens.remove, ptdens.keep, labels = c("A", "B"), ncol=1, align="v")

```

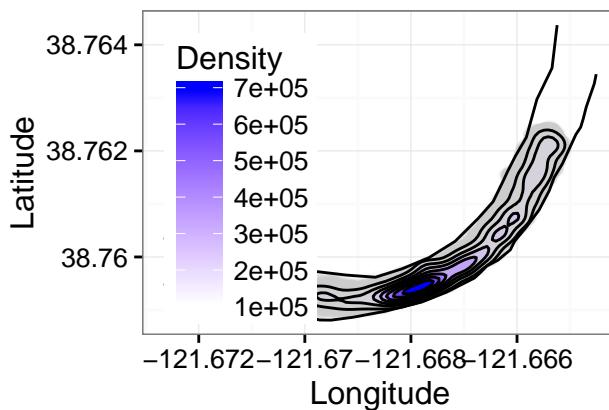
Warning: Removed 13 rows containing non-finite values (stat_density2d).

Warning: Removed 13 rows containing missing values (geom_point).

A



B



Implement HPEs threshold in filtering

- This DOES NOT match the report from VEMCO (they report 106,529 or 87% after filtering to HPE<1)
- I need to get the tagID metadata from someone; suspect there are tags here that aren't fish tags

[1] "After filtering @ HPEs<1, dataset retained:"

```
## [1] " 106529 positions,"  
  
## [1] " 86.63% of fish tag positions"  
  
## [1] " 433 individual fish"
```

Here is an excerpt from VEMCO's results document that is returned with the post-processed VPS dataset, regarding their estimate of the precision of the fish positions:

Fish tag precision is estimated here using a simple method of comparing pairs of calculated positions that are very close in time, based on the knowledge that over a short period of time a fish will travel a relatively short distance. The time durations between successive transmissions of a fish tag in this study follow a pseudorandom sequence, with a minimum of 1.0 seconds, a maximum of 2.0 seconds, and an average of 1.5 seconds. For each of the HPEs filters, the retained positions are analyzed as follows:

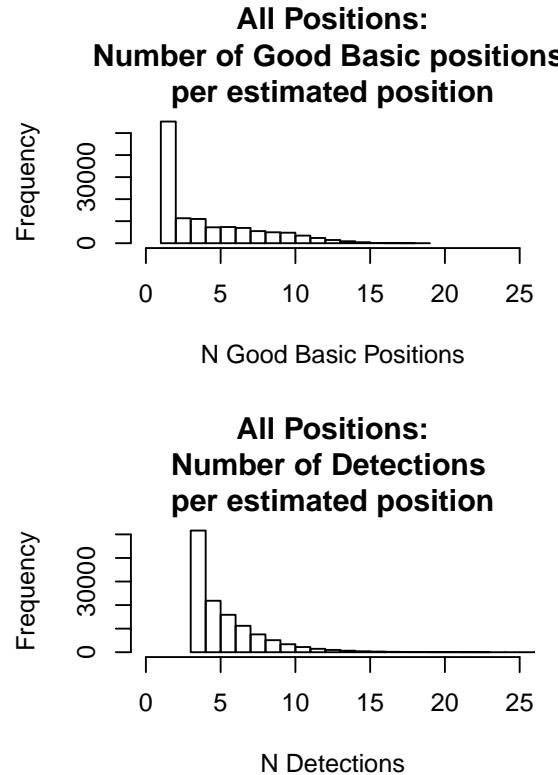
- Pairs of positions from the same transmitter that are within 2.0 seconds of each other are identified, and the horizontal distances between these pairs of positions are calculated
- Selected percentiles of these horizontal distances are calculated

The following is a summary of these analyses:

HPEs	#	Pos %	pos 50%ile	80%ile	90%ile	95%ile
0.2	25792	21%	1.34	2.04	2.56	3.16
0.5	83658	68%	1.57	2.54	3.45	4.62
1.0	106529	87%	1.68	2.93	4.21	5.86

For example, using an HPEs filter of 1.0, half of the pairs of calculated positions that are within 2.0 seconds of each other are less than 1.68 metres apart, and 80% are within 2.93m. Note that these statistics cannot be compared directly with the corresponding statistics for sync and reference tags because they include a component based on an average of 1.5 seconds of fish motion.

Histograms of individual detections/GoodBasic positions for each calculated fish location:



- not of much interest but code was already written

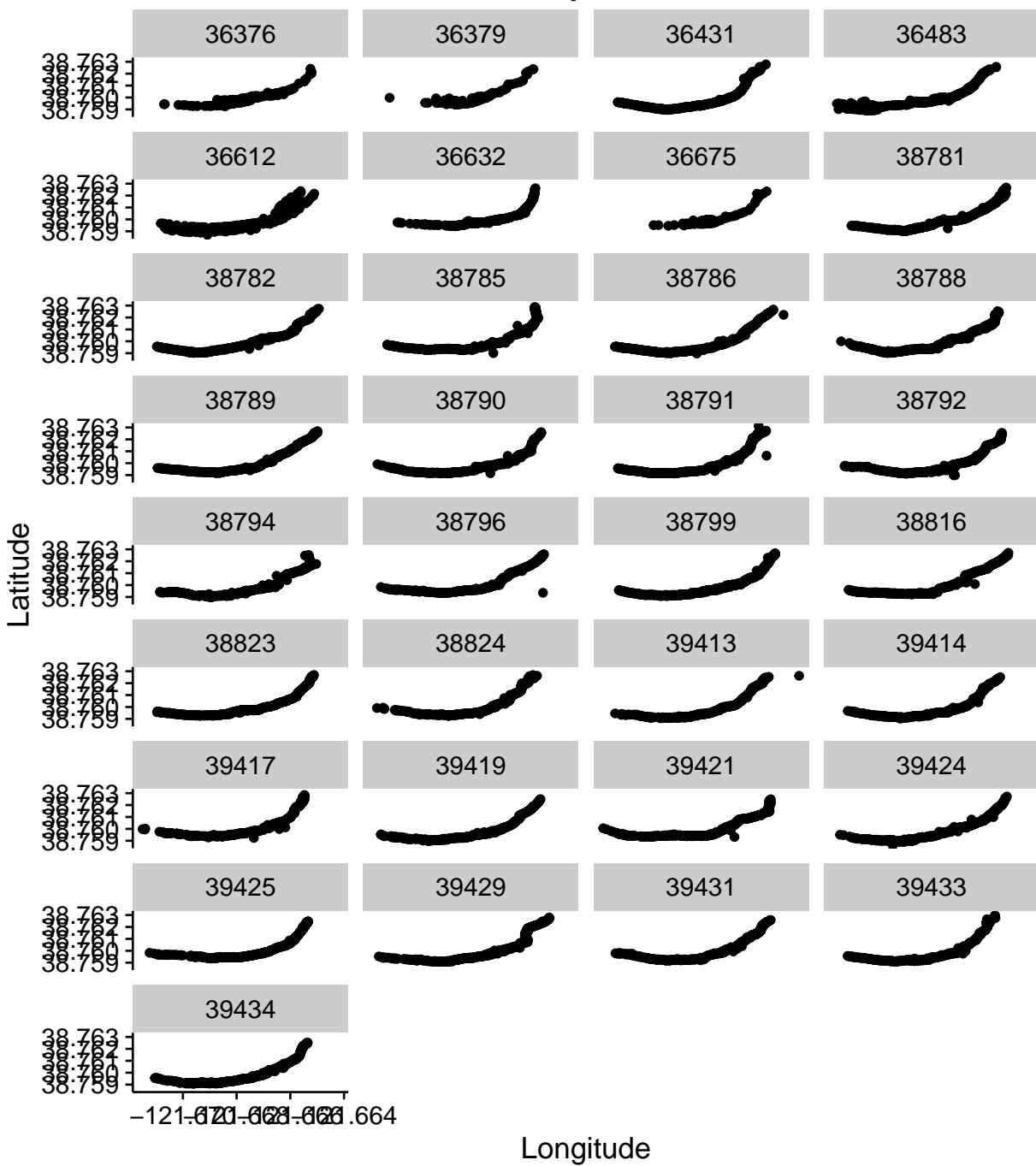
Positions remaining per fish (Figs 8 and 10 in 2015 report)

- considered positions per fish before filtering by HPE and after
- plots tracks of outliers to examine for predator-like behavior patterns
- ‘outliers’ defined according to standard boxplot metrics: all those points greater than $1.5 \times \text{IQR} + 75\text{th}\text{ percentile}$ of the data.
- plots percent of positions remaining for individual fish after filtering at $\text{HPE} < 1$ (Fig 8)

```
##           Min.        1st Qu.        Median        Mean        3rd Qu.
## 37.000000000 185.000000000 238.000000000 283.986143187 317.000000000
##             Max.
## 5577.000000000
```

```
## pdf
## 2
```

Tracks of outliers, positions not filtered



```

##              Min.       1st Qu.        Median        Mean       3rd Qu.
## 28.0000000000 165.0000000000 212.0000000000 246.025404157 283.0000000000
##             Max.
## 2015.0000000000

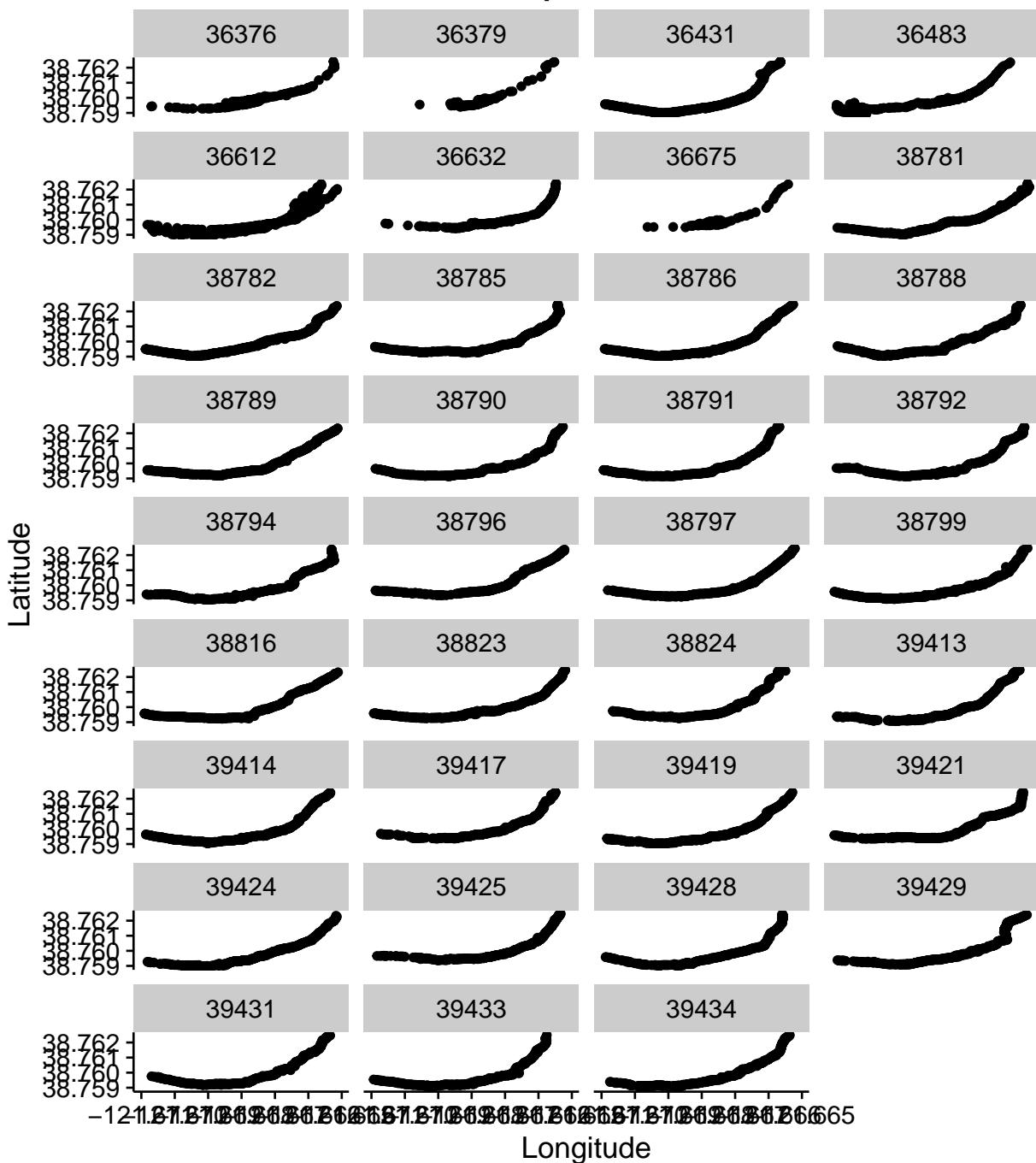
```

```

## pdf
## 2

```

Tracks for outliers, positions filtered HPEs<1



```
## Warning: Removed 529 rows containing non-finite values (stat_bin).
```

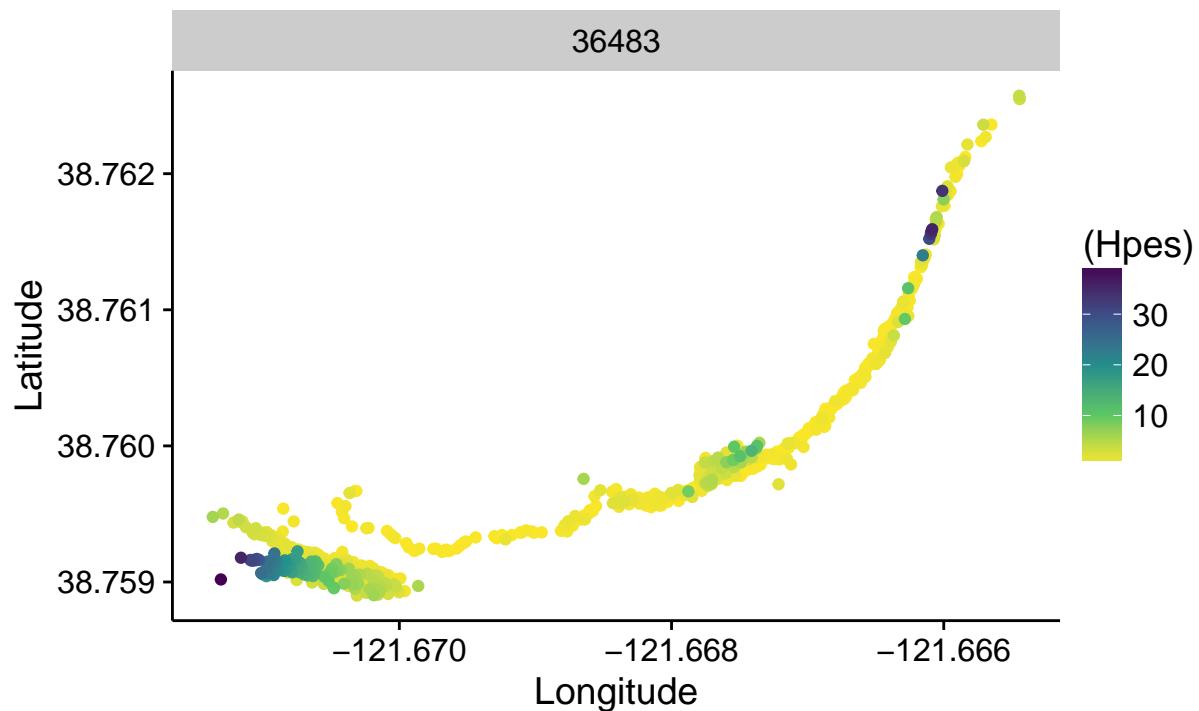
```
## Warning: Removed 23 rows containing non-finite values (stat_bin).
```

```
## pdf  
## 2
```

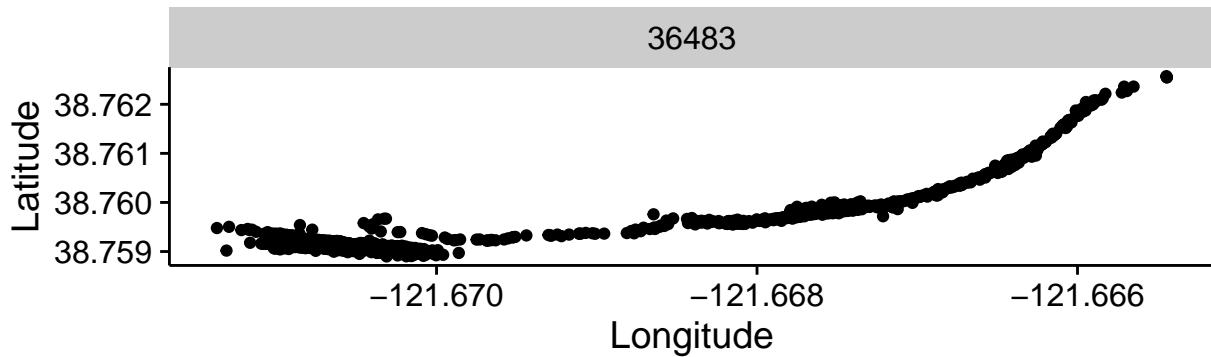
Look at fish which >40% positions removed

- how many are there?
- are the tracks indicative of predator behavior?
- where are the errors?

```
## [1] "Only 1 fish with >40% of positions removed"
```



'Poor' tracks with no filtering



'Poor' tracks filtered at HPEs < 1

