**Summary of Bootstrapping Results**

*Summary of methods:*

* Subdivided data into 24 unique combinations of year, region and salmon species
* Created collection numbers for each of the 24 data frames denoting unique date/site combinations
* Using a for loop and some dplyr functions, ran 10,000 simulations on each data frame where during each simulation the following operations were performed:
  + A collection was randomly sampled (with replacement)
  + A number of fish, n (where n is the number of fish in that collection), were randomly sampled with replacement
  + The mean number of lice per fish (for both species of lice) was calculated
  + This mean value was stored in a data frame
* This resulted in 10,000 means generated from each data frame
* To obtain an estimate and 95% confidence interval, these means were ordered, and the 5000th (estimate), 9750th (upper confidence limit), and 250th (lower confidence limit) values were taken

*Issues:*

* Since the data is very zero inflated, with the exception of one measurement (*C. clemensi*, Johnstone Strait, Pink Salmon, 2018 & *C. clemensi*, Johnstone Strait, Sockeye Salmon, 2017), every single value that was taken as the 250th ordered value for the lower confidence limit was a zero value.
  + I think this is because so many of the collections had a large number of fish that had no lice on them, so often a large number of the means (sometimes >70%) were zero
* The estimates (5000th values) for the *L. salmonis* measurements had the same problem as the lower confidence limit for both of the species. With the exception of two measurements (Discover Islands & Johnstone Strait, 2015, Pink Salmon), every single estimate was zero

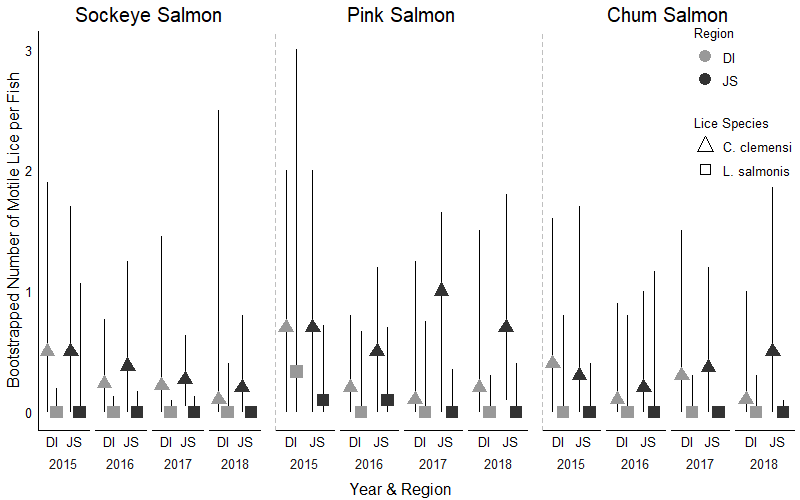
*Attempted Fixes:*

* To try and fix the L. salmonis estimate problem, I reproduced the graphs instead used the mean rather than the median for the point estimates

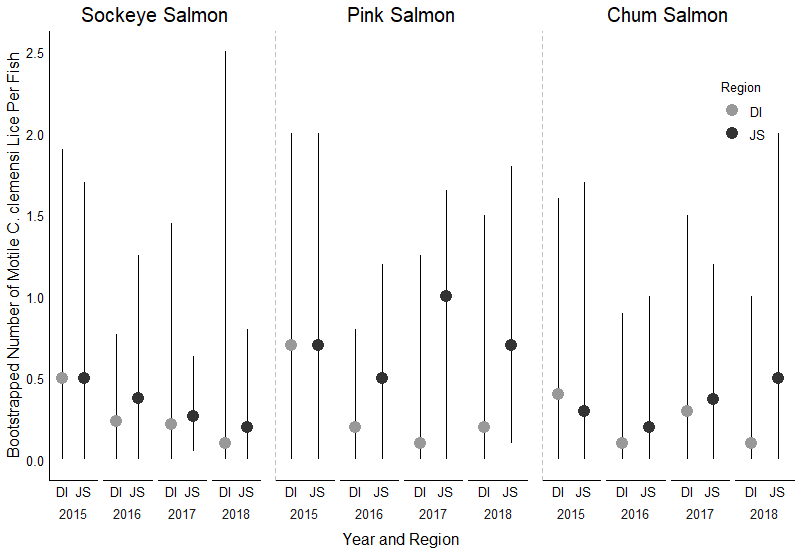
*Conclusion:*

I’m not completely sure what to make of these results, but hopefully some of this is useful. I wasn’t sure what to expect, but I didn’t anticipate having the problem of the zero values as the median estimates and the lower confidence limit estimates. I’m not sure if I’m correctly displaying the data in a way that would be most useful, but I’ve had multiple people look over my code, so I’m at least confident that it’s doing what I want it to do.

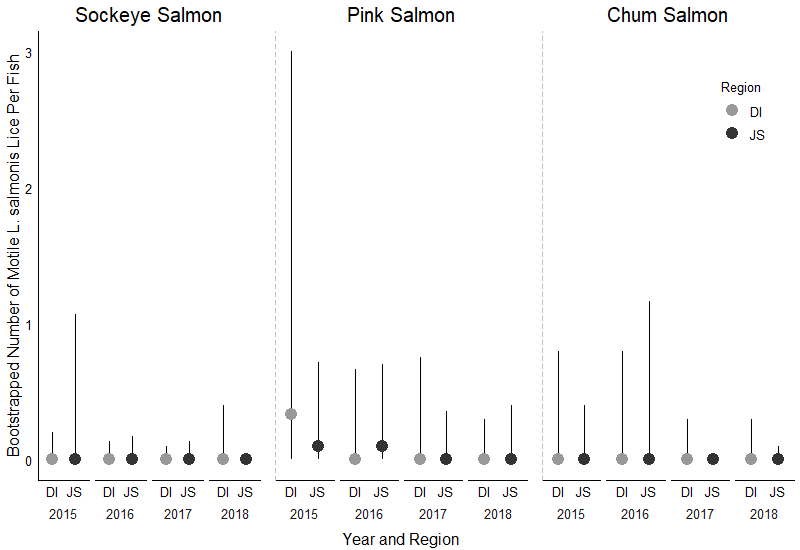
--- The relevant plots are below with figure captions ---



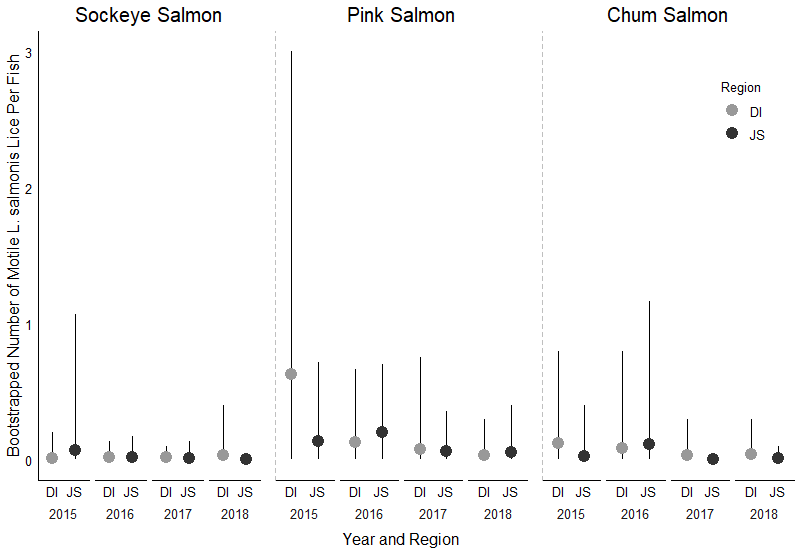
**Figure 1:** All combinations of lice, salmon, region and year. The estimates for these plots are the median (5000th) values from the ordered bootstrapped means.



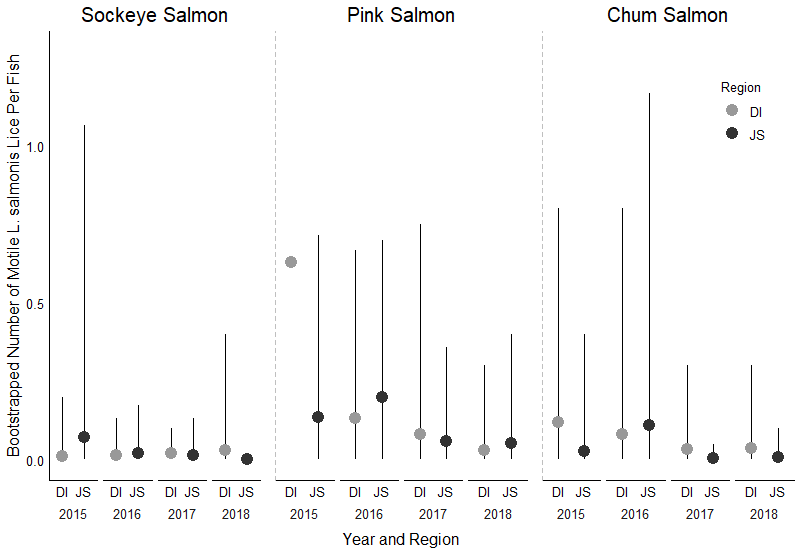
**Figure 2:** All combinations of salmon, region and year for  *C. clemensi*. The estimates for these plots are the median (5000th) values from the ordered bootstrapped means.



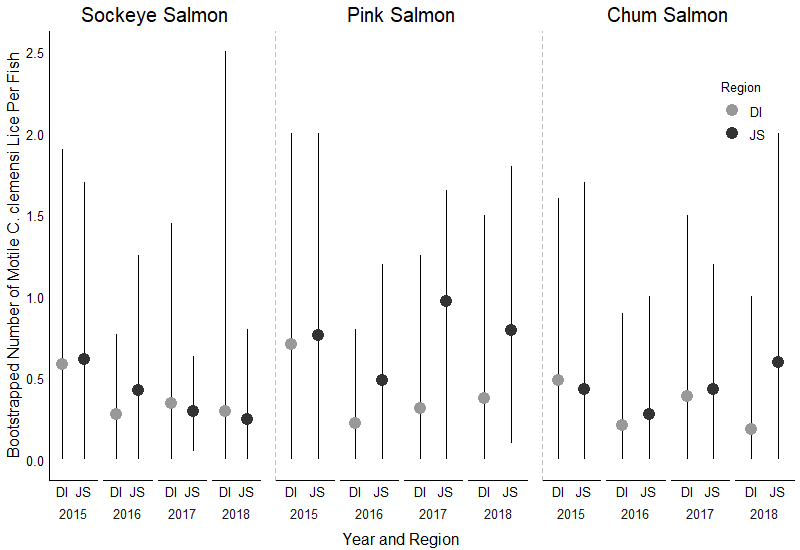
**Figure 3:** All combinations of salmon, region and year for  *L. salmonis*. The estimates for these plots are the median (5000th) values from the ordered bootstrapped means.



**Figure 4:** All combinations of salmon, region and year for  *L. salmonis*. The estimates for these plots are the means from the ordered bootstrapped means.



**Figure 5:** All combinations of salmon, region and year for  *L. salmonis*. The estimates for these plots are the means from the ordered bootstrapped means. This plot is identical to Fig. 4, with the only difference being that the y-axis scale was changed to eliminate the error bars from the Discovery Island, 2015, pink salmon value, so as to allow for easier viewing of the other values displayed here.



**Figure 4:** All combinations of salmon, region and year for  *C. clemensi*. The estimates for these plots are the means from the ordered bootstrapped means.