Combining all MRIP Data

MRIP data consists of 3 datasets, some of them overlap. They are catch\_2017\_cls, trip\_2017\_cls, size\_2017\_cls. Don’t be confused with the “cls” wording in their names, these are in fact MRIP datasets.

There are also the “I” datasets, but they contain info on same boats and for record linkage, the 3 main datasets will work just fine.

The format of the data from MRIP is one species per interview per row. So in order to tidy the data, I re-format the data such that each row corresponds to an interview, and the columns are the variables associated with each species and variables associated with the trip of interest. i.e. each row has variables like: “red\_snapper\_release”, “red\_snapper\_claim”, “grouper\_release”, “grouper\_claim”

I started with the catch\_2017\_cls data set (call it mrip\_catch from now on). I dropped the variables:

1. “TOT\_LEN
2. “TOT\_LEN\_B1
3. “tot\_len\_a”
4. “WGT\_AB1”
5. “WGT\_B1
6. “WGT\_A”
7. “MODE\_FX”
8. “CLAIM”
9. “RELEASE” since I will look at “RELEASE\_UNADJ”
10. “HARVEST
11. “HARVEST\_UNADJ”
12. “tot\_cat”
13. “MODE\_FX”
14. “AREA\_X”
15. Month
16. FL\_REG
17. Alt\_flag
18. Arx\_method
19. Kod
20. SUB\_REG
21. Region

I dropped the variables because the combination of all the variables and all the species is very large, so I drop variables to reduce the number of variables in the final dataset.

Then I go to the trip\_2017\_cls dataset (called mrip\_trip from now on). There is already one interview per row. I only use the variables: ID\_CODE, PRIM1\_COMMON, and PARTY from mrip\_trip.

Then look at size\_2017\_cls (mrip\_size). For now, I will not use it because it has many of same variables, with some different variables for measuring fish. I think this is too complicated for right now.

Then, take the updated mrip\_catch (I call it m\_catch) and then left join on the updated mrip\_trip. The result is all the mrip data, it has one row per interview with 1257 variables.

Note: I pulled date out of the ID\_CODE number