***Jasmine Data Wrangling Notes***

* OSS detection files
  + Seven plots per stand number, but not all of these plots have surveys
    - But when I drop all the plots that have surveys, we end up a little short on nall (3217 v 3379)…
  + V1 through V3 are repeated surveys
    - Stopped if they found something
* Ugh we’re missing the covariates for these; seemingly
  + Date, air temp, DWD
  + BUT these are likely just three variables, with one value for each row
    - I can get these in here
    - If I want to run his original code, I can likely just simulate these
* Stand ID turns into stand number with a as.numeric(as.factor())
* He did a lot of stuff with poor replication form that is gonna make it a bit more of a pain to replicate
  + Doing it with specific years rather than just vectorizing all this
* Once I reworked some stuff to recreate original data; here’s what’s messed up
  + Level 1
    - nall too small
  + Level 2
    - Year2024 – Year2019 have 87 columns instead of 88
    - Same with PreTrt and PostTrt
    - Stand2 showing stand IDs rather than factor numbers
  + Level 3
    - Everything got the wrong nall
    - ye has a fully different dimension
      * This was a problem with my filtering
* Okay I think I fixed this all other than the nall, which is likely just due to the different source data files?
  + Okay I had my NA filter wrong, but now I think we’re good to go!
    - Have some extra rows because I don’t know which ones are missing covariates
* Stuff to figure out in translating Jas’ data
  + DWD
    - Seems like Jay’s data is just a count? (0-18)
    - Where’s this at?
      * Could summarize from sp\_year\_dwd
        + Logs AND stumps?

Jas says do just logs

* + - * + Actually use dwd.complete
  + Counts in series
    - Not in occupancy CSVs, which just show each subplot (not each survey)
    - Oss\_year\_sals – summarise from pass? Includes both species
      * + Probably combined in sals\_complete?

What is sal\_rep ?

* Where to draw from Jasmine’s data
  + Sp\_year\_sitesubplot
    - Air temp, survey date, treatment, landowner, tree farm, stand number
  + Site.complete
    - Has all the site data; easy enough
* Outstanding questions for Jasmine
  + Are there any new owners?
    - No
  + Are there any new tree farms?
    - Yes, there are now three
      * IN site\_level\_matrix.csv
    - Only two in the original
  + How do you want treatment handled?
* STAND != SITE because sometimes there were two sites in stands
* Notes for Jasmine
  + Jay has both a “Trt” and a “TrtGrp” variable
    - Trt refers to whether it has been cut at all, doesn’t separate between “Control” and “Pre-Treatment”
    - Dropping the “Trt” thing