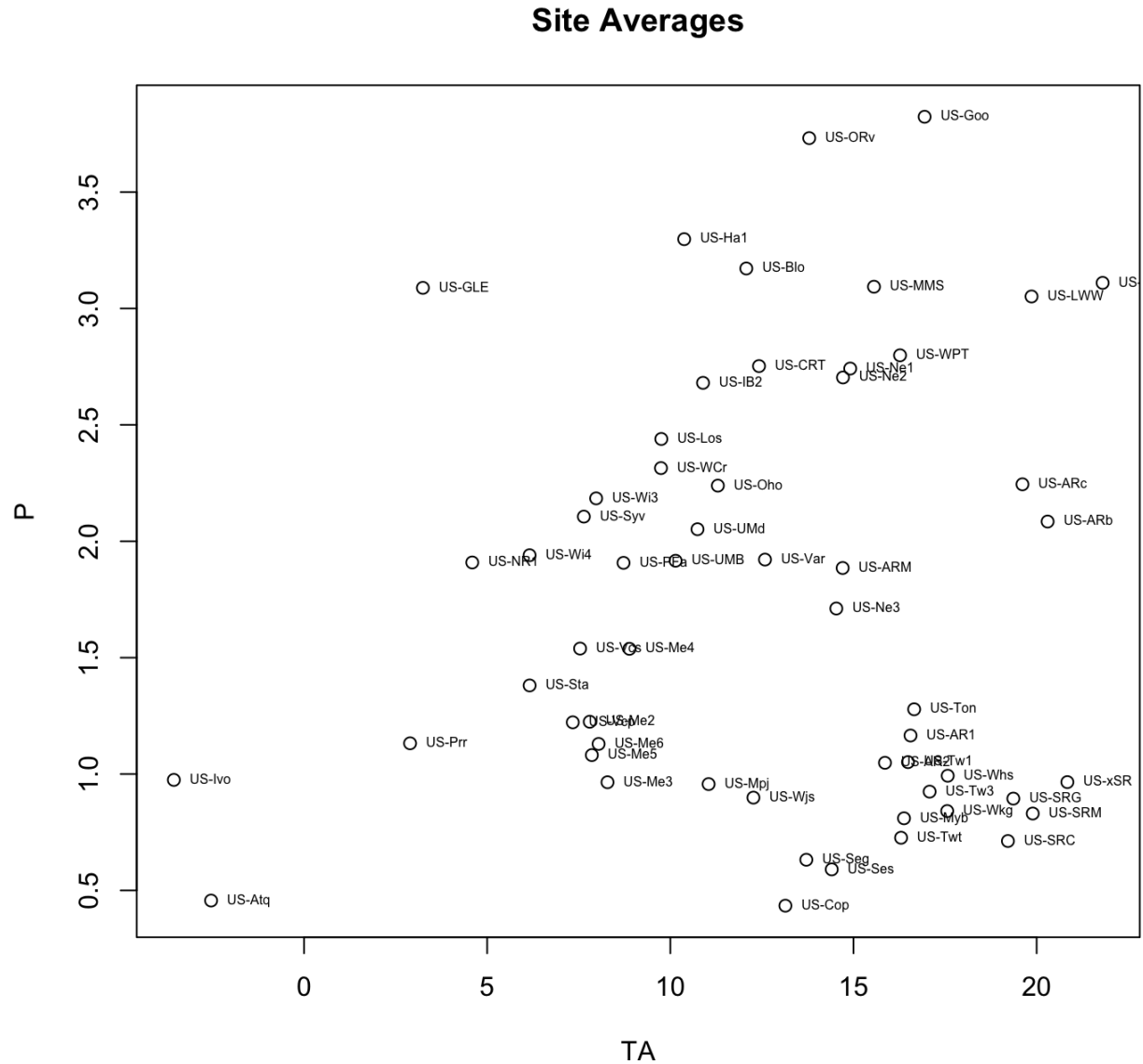
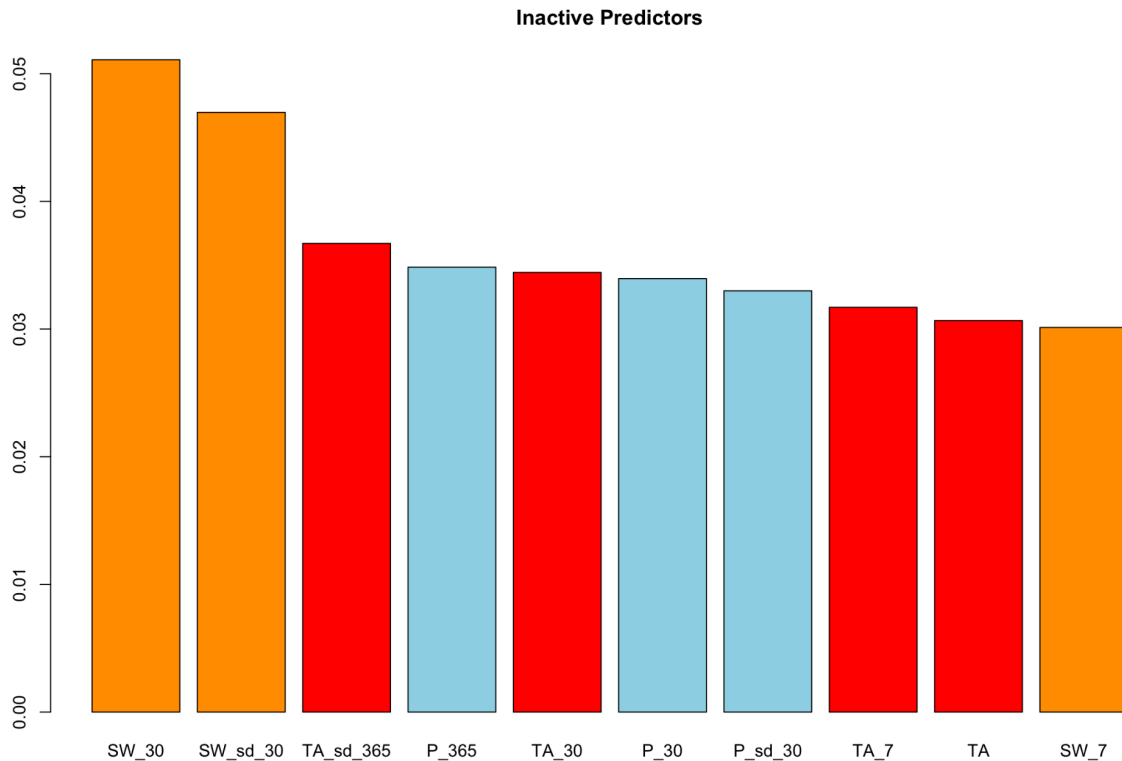


# Case12: Combined Inactive

Reminder of our site average information:



Starting off strong with our top 10 predictors:

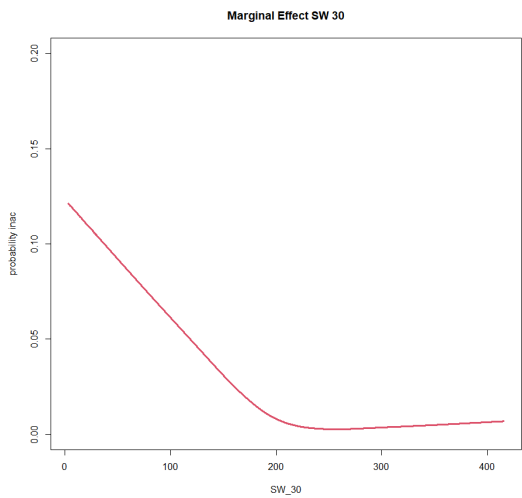
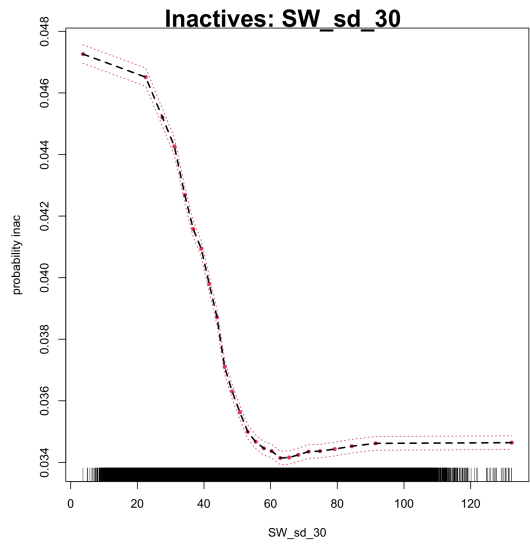
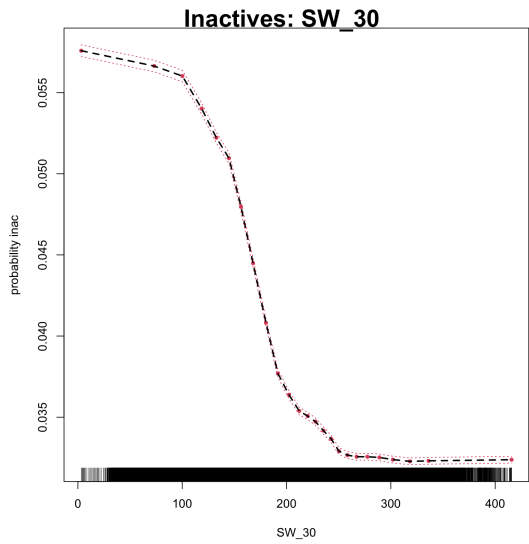


Things to note here are that the top predictors include a lot of standard deviation variables and they are almost all longer timescales. Sunlight, temperature, and precipitation come into play. One of my questions is if the longer timescales is tied to clumping of inactive identifications so that certain years that match an environmental condition make up the most of the inactive states.

On a monthly scale, these are the percentage of inactive classifications that belong to each month. The months of Nov-Dec make up .723 of the true inactive states, tying to its out-of season occurrence which is incorporated directly into its definition (near 0 GPP). Lets make some pdp plot of the top predictors now.

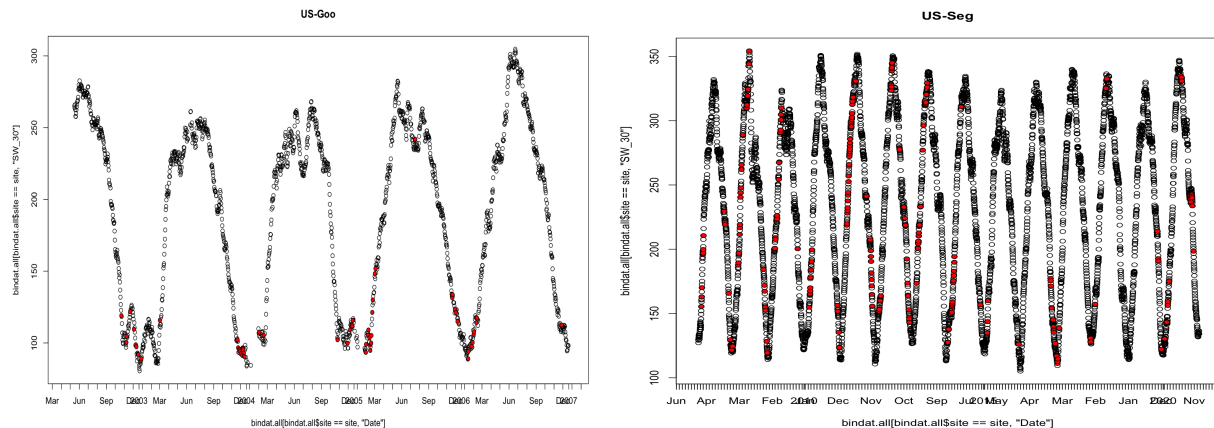
Month	Inac Percent
1	0.239
2	0.186
3	0.096

4	0.043
5	0.026
6	0.018
7	0.016
8	0.014
9	0.019
10	0.041
11	0.100
12	0.198



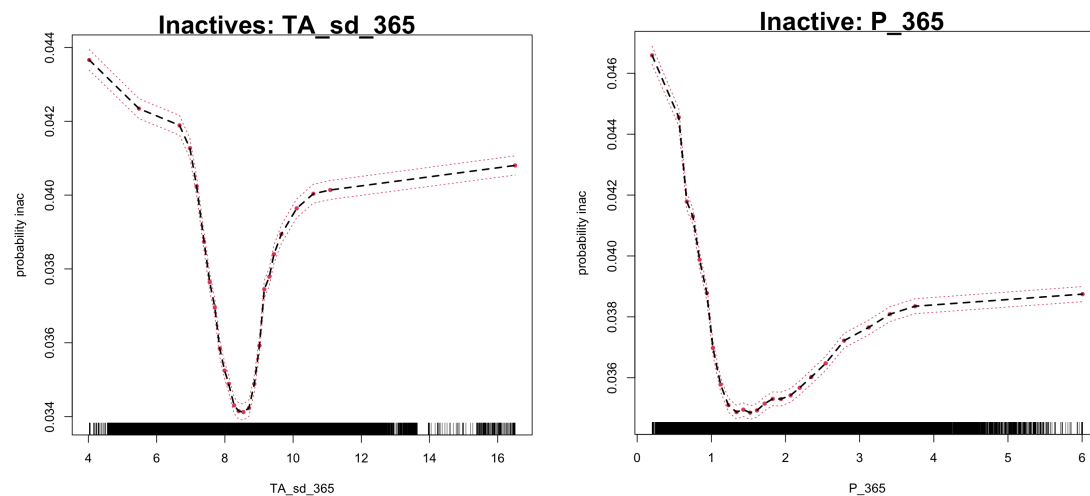
Im just now adding in marginal effects instead of partial dependency and I am so confused what the difference is.

This is a good example of what is most commonly seen with SW\_30, where the inactive occur in the months where inactive is switching from decreasing to increasing. In some sites, inactive occurs at the higher SW turning point as well, but occurrences are at the lower SW30. This is why low SW\_sd\_30 increases the probability of being inactive, because the sunlight is changing from low to high and thus the sd is lower. Some of the desert sites don't follow this pattern (Ton, SES, SEG), which I included an example of.



I am going to dive into some of the 365 variables now. What I am wondering is if these 365 variables are actually identifying years within a site that meet an elevated environmental criterion, or if it is just identifying the different sites. Here are some ways I can figure this out:

- (1) Comparing PDP plots to points we see
- (2) Looking at clustering in comparison to location of site environmentally



Looking at P\_365, it is clear that this is in connection to the sites rather than a year to year basis. These are the wettest sites NEE along P365. The inactive days are colored based on year to show the representation of inactivity in each of the years that data was collected from the site. Notice the wettest sites are occurring around 2-2.5 minimum P365, and thus the wet

sites are representative of that entire right side of the PDP plot. You can see that inactivity spreads along the range of the precipitation for these sites, indicating that the P365 variable isn't providing much information within the site, but rather pointing to the wetter sites in general. Precipitation in general is a more stricter variable I feel like, where sites have tighter ranges so there's no site truly covering across this entire range. Instead, it's broken up into climate types and stacked.

TA\_sd\_365 is weird on the other hand since its densities don't show that most of the inactives have a tasd365 of around 8. This could be tied to the dependency of temperature on other variables that are also important. A high sd is going to be more associated with average 365 temperatures since it can't have a high sd and be an extreme annual temperature for that site. At least that's what I am thinking.

