

Regional-level analysis

We'll also sometimes want to take a step back and group some of our observations together to do data analysis at a different level.

- Aggregate at the level of regions using the `aggregate()` function. (*Hint:* Pass in `FUN=median`.)
- Compute the correlations between the resulting columns.
- How do these compare with the correlations you calculated at the state level? What do you think explains the difference?

Adding interaction terms

We can add *interaction terms* to a linear regression very easily: in the list of predictors we pass in to the `lm()` function, we can include the interaction of `var1` with `var2` by including `var1:var2` or `var1*var2`.

- What's the difference between including `var1:var2` or `var1*var2`? (*Hint:* Try regressing against nothing aside from the interaction term.)
- How much additional predictive power can you get by including well-chosen interaction terms in your regression? Which interaction terms help the most?