Stat 610 Homework 3

Thursday, September 29, 11:59pm.

Background

We are going to be interested in the relationship between tornado magnitude and tornado damage, and whether the relationships is different from state to state.

Assignment

Your assignment is as follows:

1. Get the data using `readRDS('tornadoes.RDS')`
2. Use `lapply` or `sapply` and the `typeof` function to get the classes of all of the columns of `td`. What happens when you use `lapply` or `sapply` on a data frame? Why?
3. Using the split/apply/combine strategy discussed in class, compute the fraction of tornadoes with magnitude (`mag`) greater than or equal to 3 for each state. Remember that the previous part showed that magnitude is of type character and you will need to make it into a number.
4. Write a function that takes a data frame as its argument and returns the slope coefficient in a linear model with loss (`loss`) as a linear function of magnitude (`mag`, coded as a number). (Hint: the `coef` function will extract the coefficients from the output of the `lm` function.)
5. Using the split/apply/combine strategy discussed in class and the function you wrote in the previous section, compute the slope in a linear model of loss as a function of magnitude for each state.
6. Notice that the coding for loss changes after 1996 (https://www.spc.noaa.gov/wcm/data/SPC_severe_database_description.pdf), that loss = 0 means unknown, and that mag = -9 also means unknown. Split the analysis by pre/post-1996 in addition to state (so now there should be two values for each state, one for pre-1996 and one for post-1996) and modify either the dataset or your function so that it does not include loss = 0 or mag = -9 in the calculations.
7. One might expect states that see strong tornadoes more often to build more defensively against hurricanes. To check this, merge the results in parts (3) and (6), and see if there is a correlation between the fraction of hurricanes with magnitude at least 3 and the increase in loss due to stronger hurricanes.

Extra credit:

1. (1 point for a good answer): Non-zero log losses post 1996 are strongly bi-modal, even after accounting for magnitude. Extra credit for figuring out why this is (fair warning: I don't know why either and am asking because I would like to know).

2. (.5 points for a good answer): Extra credit for noticing any coding problems that I missed.

Submission parameters

– Submit an `.Rmd` file with your answers to the questions, your code, and a description of what it is doing.