fire-color: matching/casual inference methods

identifying post-fire color changes in lakes of the western U.S.

Bella Oleksy et al.

Last compiled on 10 January, 2022

Contents

1 Overview

In this document I'm going to be assessing some different methods for causual inference, such as propensity scoring, regression discontinuity, and difference-in-differences.

Big picture: we want to assess responses of lake color to fire and also compare those lakes to similar/matching lakes that did not burn. The reason we want to incorporate an apporach like this is because it allows us to "design" an observational study that mimics some of the particular characteristics of a randomized control trial. The "causal effect" we are trying to measure ist he impact of fire on lake color. So in our case the **control** is lakes in unburned watersheds and the **treatment** is lakes that experienced burns. Because fires are not randomly distributed across the landscape, we need to be careful in our design and how we interpret our results. We achieve the goal of matching by comparing populations of lakes that are similar in a number of covariates.

Causal inference definition in Stuart 2010 *Statistical Science*: "Any method that aims to equate (or"balance") the distribution of covariates int he treatment and control group. This may involve 1:1 matching, weighting, or subclassification."

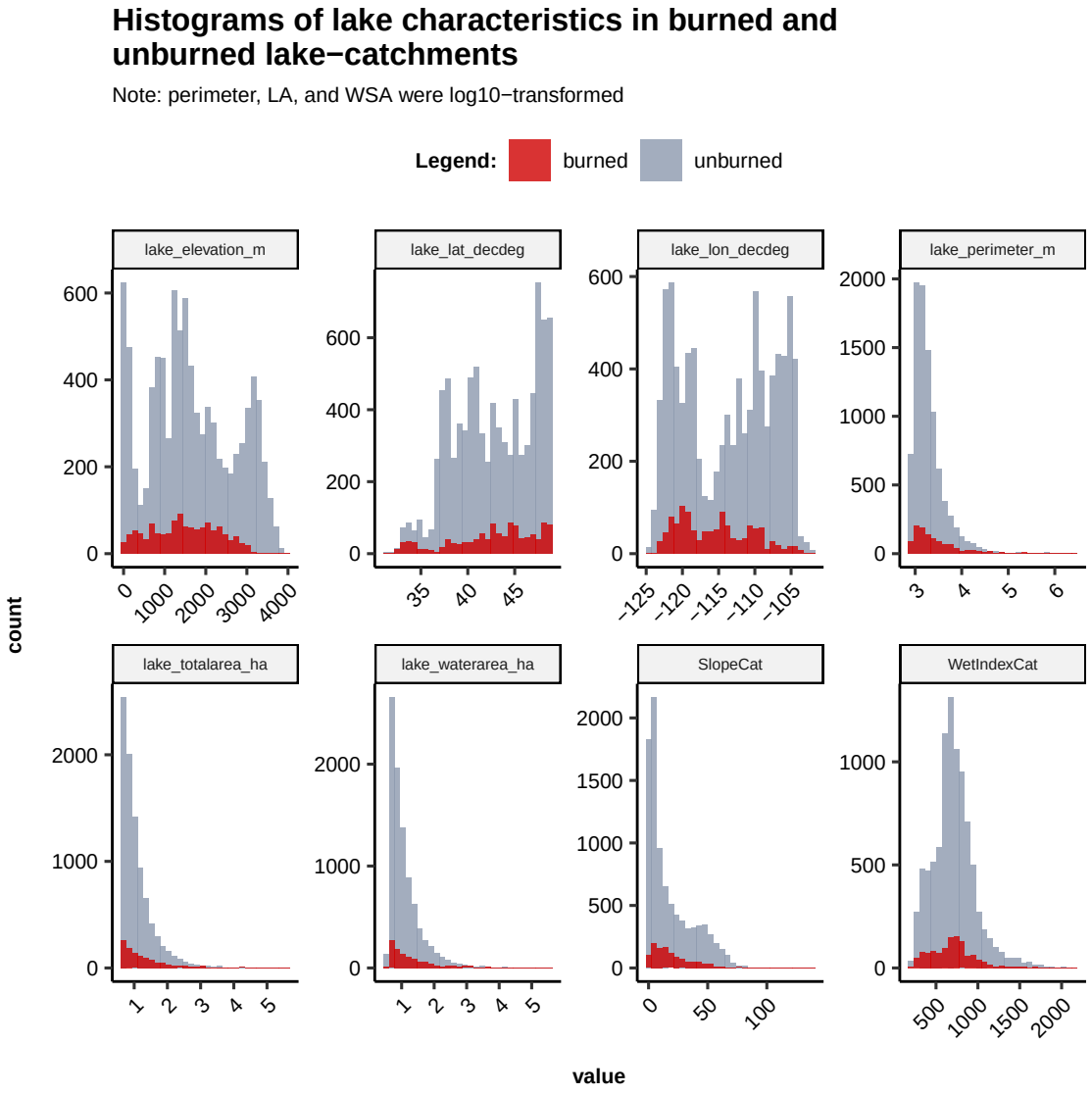
1.1 Covariates

Of all the lakes in the LakeCat dataset > 4 ha, how many burned at all (>1%) versus not at all?

Table 1

burn_YN	n
burned	1157
unburned	9074

And are there obvious differences among the lakes?



Nothing too different pops out at me with the histograms, but let's see what we can learn beyond just eye-balling the data.

1.1.1 Propensity score example via MatchIt

```
# Create binary variable for fire
library(recipes)
AllLakes <- AllLakes %>%
    recipe(lake_lat_decdeg ~ .) %>%
    step_dummy(burn_YN, one_hot = TRUE) %>%
    prep() %>%
    bake(AllLakes) %>%
    select(-burn_YN_unburned) %>%
    rename(fire = burn_YN_burned)  ## fire = 0 = no fire; fire = 1 = yes fire.

require(MatchIt)
matching <- matchit(fire ~ lake_totalarea_ha + lake_waterarea_ha + lake_elevation_m +
    WetIndexCat + SlopeCat, method = "nearest", data = AllLakes, distance = "glm",
    link = "probit", discard = "both")
```

```
summary(matching$model)
```

Calling summary(matching$model) gives the result of a binary probit model which estimates the propensity score, which is the probability of a lake having burned. As expected, slope and elevation are good predictors of fire. So is WetIndex which stands for Mean Composite Topographic Index (fairly highly correlated with slope, r= -0.4659471).