

Tutorial on propensity score matching and inverse probability of treatment weighting

Katherine Siegel and Laura Dee

2024-06-17

Description

This tutorial demonstrates how to use propensity score matching and inverse probability of treatment weighting using a real dataset from Siegel et al. 2022. The dataset for the entire western US is very large and unwieldy, so you'll work with a subset of data for a single year in Colorado.

Set up

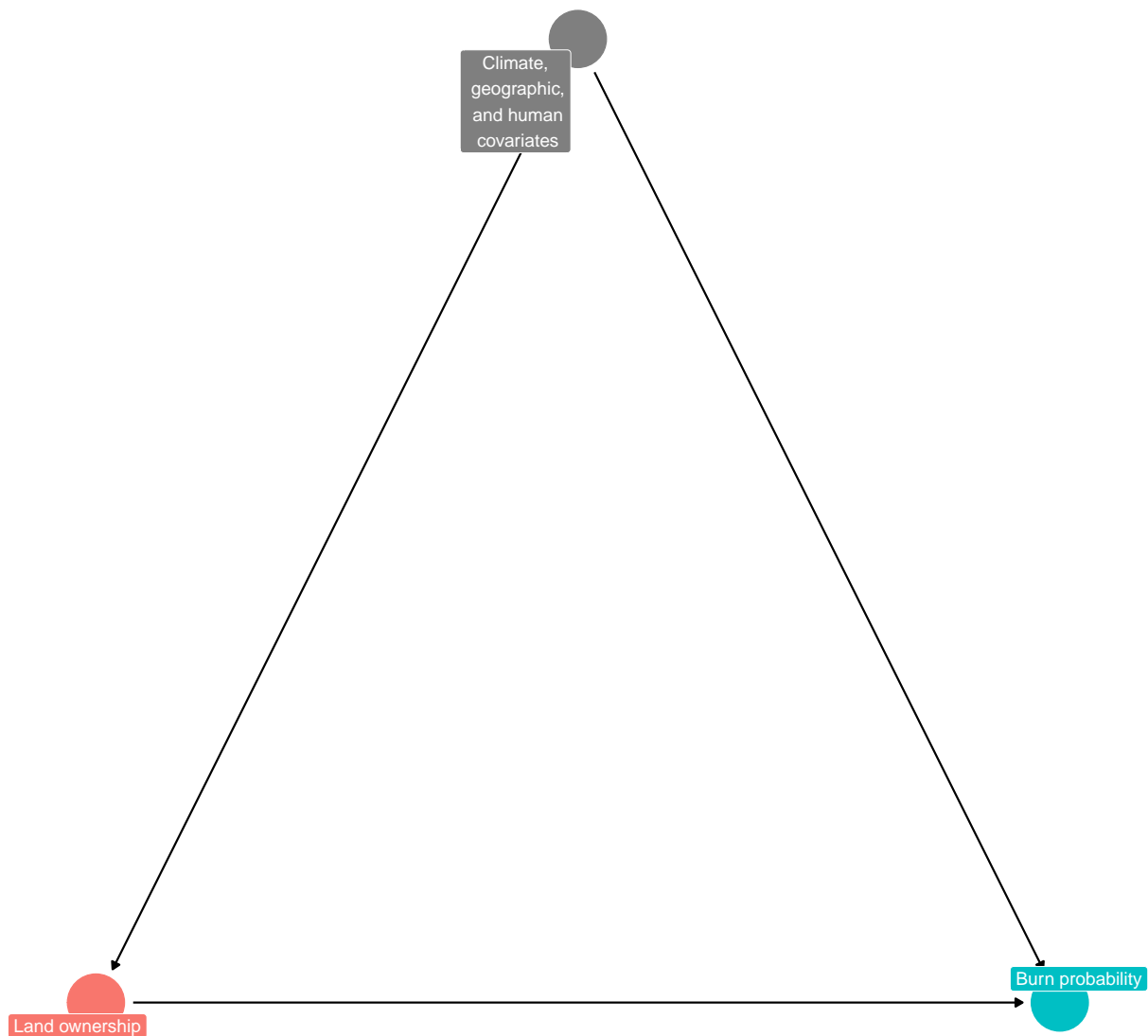
Load the packages used for data manipulation (tidyverse, sf), making a directed acyclic graph (ggdag), matching (MatchIt), weighting (ipw), and regression models (lme4).

The context

The Siegel et al. 2022 study examines the effect of forest management (through the proxy of land ownership) on annual burn probability in forests of the western US. Specifically, it looks at the effect of federal (treatment) vs. private (control) ownership on wildfire occurrence in sample units.

Directed acyclic graph

Here's a DAG for the research question:



The data

The data are in the file *matching_ipw_data_full.csv*.

Variable names

- state: the state the sample unit is from (Colorado)
- UID: a unique identifier for each sample unit
- year: the year that the fire and climate data is from (2002)

- burned: whether or not the unit burned in 2002 (0 = unburned, 1 = burned)
- prot_cat_recl: the ownership class. 0 = private, 1 = federal
- dist_rds_km: distance to the nearest road, in kilometers
- slope: slope, in degrees
- aspect_srai: solar radiation aspect index
- elev_km: elevation, in 1000 m
- lon: longitude
- lat: latitude
- pdsi_avg_season: seasonal average Palmer Drought Severity Index value (fall, spring, summer, winter)
- soil_avg_season: seasonal average soil moisture (fall, spring, summer, winter)
- tmmn_avg_season: seasonal average minimum temperature (fall, spring, summer, winter)
- tmmx_avg_season: seasonal average maximum temperature (fall, spring, summer, winter)
- vs_max_season: seasonal average maximum wind speed (fall, spring, summer, winter)
- total_precip_season: total seasonal precipitation (fall, spring, summer, winter)
- prev_yr_precip: total precipitation in the previous year

Data exploration

What's the breakdown of private (value = 0) vs. federal (value = 1) units?

Table 1: Units on federal (=1) and private (=0) land

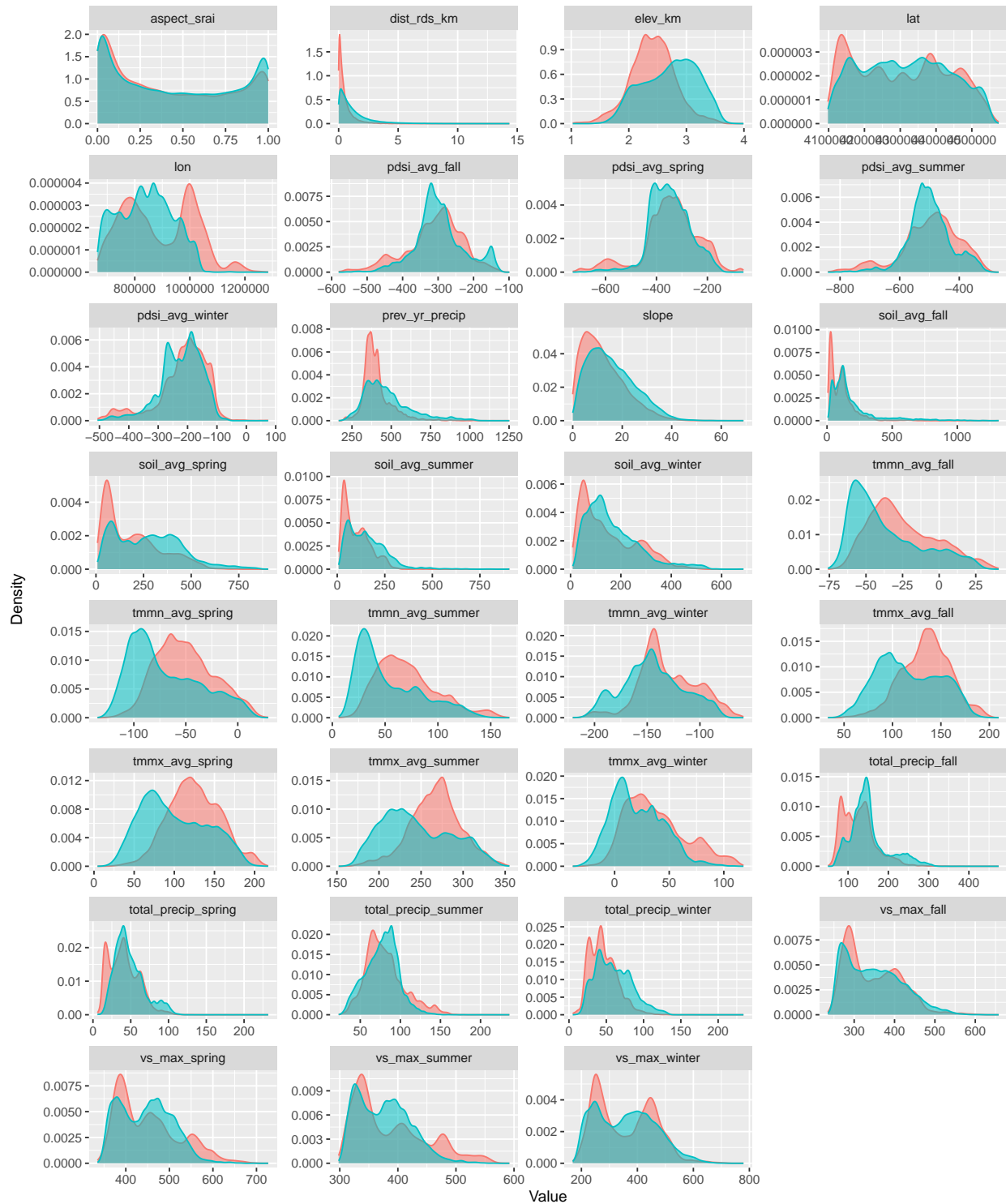
Var1	Freq
0	22877
1	60654

What's the breakdown of units that burned (value = 1) in 2002 vs. units that did not burn (value = 0)?

Table 2: Units that burned (=1) or did not burn (=0) in 2002

Var1	Freq
0	82189
1	1342

How do the private (in red) vs federal (in blue) units differ in terms of potential confounders?



Run naive regression

We could just run a naive regression, ignoring the potential impact of confounders. There are some highly correlated covariates in the model, but let's ignore them for now. Let's see what that would yield:

Table 3: Coefficient estimates for naive model

Variable	Estimate	Std. Error	p value
(Intercept)	-14.088	3.606	0.000
prot_cat_recl1	1.229	0.095	0.000
dist_rds_km	-0.079	0.029	0.007
slope	0.019	0.003	0.000
aspect_srai	0.149	0.084	0.077
elev_km	1.627	0.215	0.000
pdsi_avg_winter	-0.014	0.002	0.000
pdsi_avg_spring	-0.021	0.003	0.000
pdsi_avg_summer	0.036	0.004	0.000
pdsi_avg_fall	-0.004	0.003	0.143
soil_avg_winter	0.000	0.001	0.597
soil_avg_spring	0.006	0.001	0.000
soil_avg_summer	-0.002	0.001	0.200
soil_avg_fall	-0.002	0.001	0.003
tmmn_avg_winter	-0.112	0.011	0.000
tmmn_avg_spring	0.054	0.026	0.035
tmmn_avg_summer	-0.095	0.022	0.000
tmmn_avg_fall	0.131	0.026	0.000
tmmx_avg_winter	0.133	0.012	0.000
tmmx_avg_spring	-0.017	0.021	0.431
tmmx_avg_summer	0.265	0.021	0.000
tmmx_avg_fall	-0.395	0.029	0.000
vs_max_winter	0.044	0.004	0.000
vs_max_spring	-0.038	0.005	0.000
vs_max_summer	-0.007	0.005	0.157
vs_max_fall	-0.045	0.006	0.000
total_precip_winter	-0.152	0.008	0.000
total_precip_spring	-0.034	0.007	0.000
total_precip_summer	0.130	0.007	0.000
total_precip_fall	0.044	0.005	0.000
prev_yr_precip	-0.003	0.003	0.196

Use matching to overcome issues with observed confounding variables

Match the data

Match the data on the observable covariates, using the MatchIt package. You can play around with the settings to see how it affects the matched data you end up with.

Assess match quality Take a look at the quality of the matches: how many units were matched? Control = private units, Treated = federal units

Table 4: Breakdown of matched and unmatched units

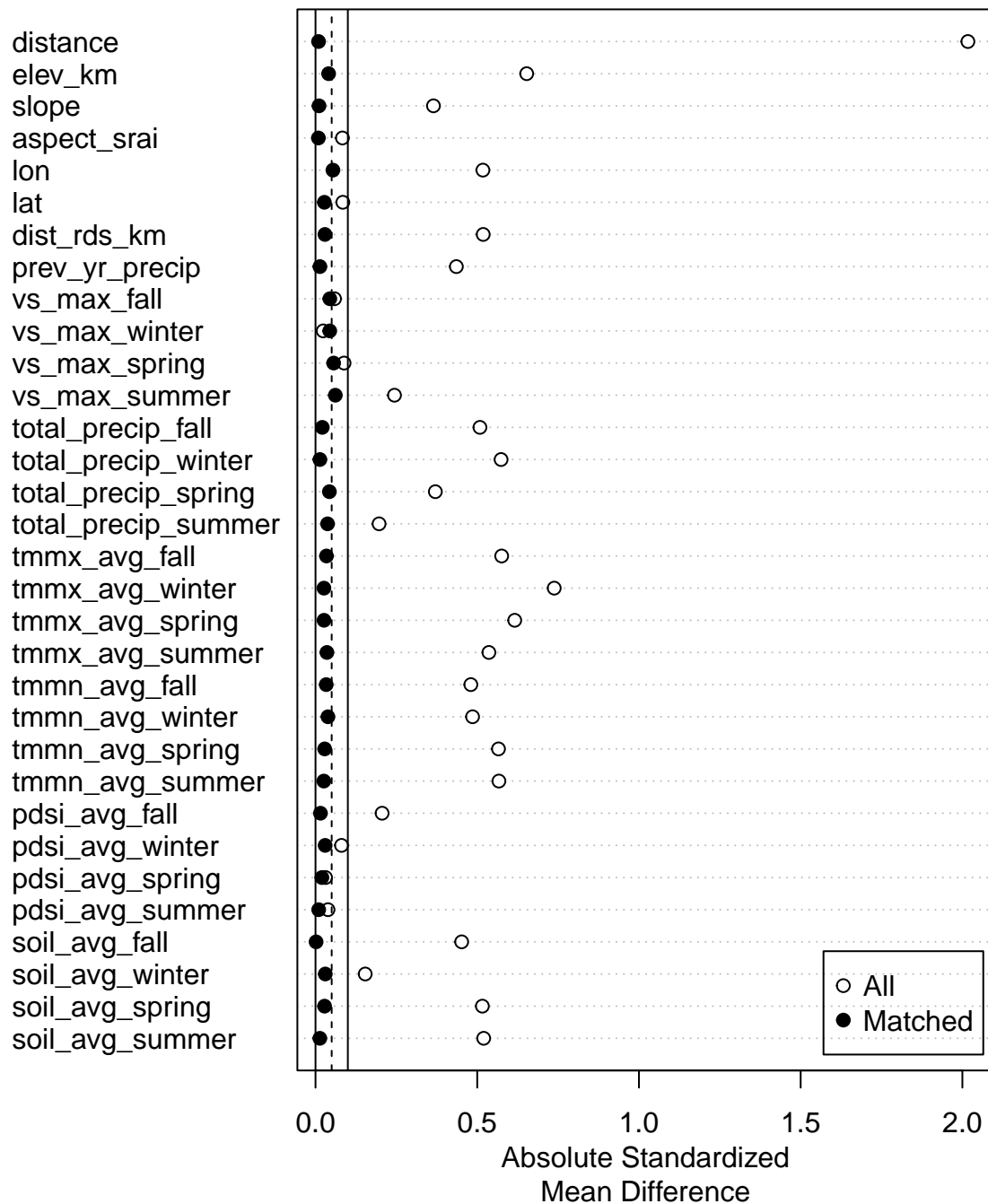
	Control	Treated
All (ESS)	22877	60654
All	22877	60654
Matched (ESS)	14941	14941
Matched	14941	14941
Unmatched	7936	45713
Discarded	0	0

What are the covariate means in the matched dataset for the treated (federal) and control (private) units?
What was the covariate balance after matching?

Table 5: Covariate balance of matched dataset

	Means Treated	Means Control	Standardized Mean Difference
distance	0.6175965	0.6158981	0.0094462
elev_km	2.5053351	2.4864021	0.0405370
slope	13.2518894	13.1504751	0.0108233
aspect_srai	0.4449302	0.4418271	0.0089396
lon	850134.9720249	845005.0610417	0.0538453
lat	4318592.0954453	4321889.5989592	-0.0274530
dist_rds_km	0.5686411	0.5290214	0.0291697
prev_yr_precip	415.8833411	413.8367579	0.0136069
vs_max_fall	333.2350579	330.1453718	0.0442657
vs_max_winter	337.1712067	332.5381835	0.0440739
vs_max_spring	433.4193829	430.1153203	0.0560467
vs_max_summer	369.8738371	367.1268322	0.0612961
total_precip_fall	135.7814738	136.7715012	-0.0209289
total_precip_winter	49.3297637	49.0044174	0.0135343
total_precip_spring	43.3679807	44.2090222	-0.0429109
total_precip_summer	74.1364701	73.3847801	0.0373767
tmmx_avg_fall	129.6127658	130.7068916	-0.0339943
tmmx_avg_winter	30.8183968	31.4282400	-0.0260555
tmmx_avg_spring	117.6968074	118.7651875	-0.0262852
tmmx_avg_summer	262.3405395	263.7941236	-0.0352977
tmmn_avg_fall	-29.4074471	-28.6041764	-0.0331496
tmmn_avg_winter	-136.7220623	-135.6087054	-0.0384818
tmmn_avg_spring	-56.4995203	-55.5338331	-0.0283374
tmmn_avg_summer	64.8160989	65.6082815	-0.0256975
pdsi_avg_fall	-307.1160119	-306.0554403	-0.0153424
pdsi_avg_winter	-221.0996587	-223.2403677	0.0298376
pdsi_avg_spring	-350.8192892	-352.3800950	0.0197357
pdsi_avg_summer	-503.8968610	-504.6041764	0.0098540
soil_avg_fall	124.6548870	124.9799433	-0.0017864
soil_avg_winter	162.4721460	165.6927247	-0.0302773
soil_avg_spring	210.6021016	215.4341521	-0.0280174
soil_avg_summer	113.0530531	114.3978984	-0.0133958

Comparison of standardized mean differences in the covariate values in the full vs. matched dataset



Analyze the matched dataset

Extract the matched data First, you'll need to extract the matched data and use the UIDs from the matched data to subset the full dataset for analysis.

Model the effect of ownership/management on wildfire probability Again, there are correlated covariates, but let's just ignore them

Table 6: Coefficient estimates for model with matching

Variable	Estimate	Std. Error	p value
(Intercept)	-41.235	7.821	0.000
prot_cat_recl1	0.945	0.116	0.000
dist_rds_km	-0.137	0.102	0.178
slope	0.025	0.006	0.000
aspect_srai	0.097	0.142	0.492
elev_km	2.268	0.419	0.000
pdsi_avg_winter	-0.020	0.004	0.000
pdsi_avg_spring	-0.021	0.006	0.001
pdsi_avg_summer	0.030	0.009	0.001
pdsi_avg_fall	-0.003	0.007	0.678
soil_avg_winter	0.001	0.002	0.469
soil_avg_spring	0.009	0.002	0.000
soil_avg_summer	-0.007	0.003	0.033
soil_avg_fall	0.000	0.001	0.807
tmmn_avg_winter	-0.067	0.020	0.001
tmmn_avg_spring	-0.100	0.045	0.026
tmmn_avg_summer	-0.147	0.043	0.001
tmmn_avg_fall	0.235	0.051	0.000
tmmx_avg_winter	0.170	0.024	0.000
tmmx_avg_spring	-0.020	0.039	0.613
tmmx_avg_summer	0.525	0.043	0.000
tmmx_avg_fall	-0.644	0.059	0.000
vs_max_winter	0.045	0.008	0.000
vs_max_spring	-0.083	0.012	0.000
vs_max_summer	0.007	0.010	0.474
vs_max_fall	-0.034	0.013	0.010
total_precip_winter	-0.280	0.018	0.000
total_precip_spring	-0.139	0.015	0.000
total_precip_summer	-0.002	0.016	0.924
total_precip_fall	-0.004	0.010	0.669
prev_yr_precip	0.059	0.007	0.000

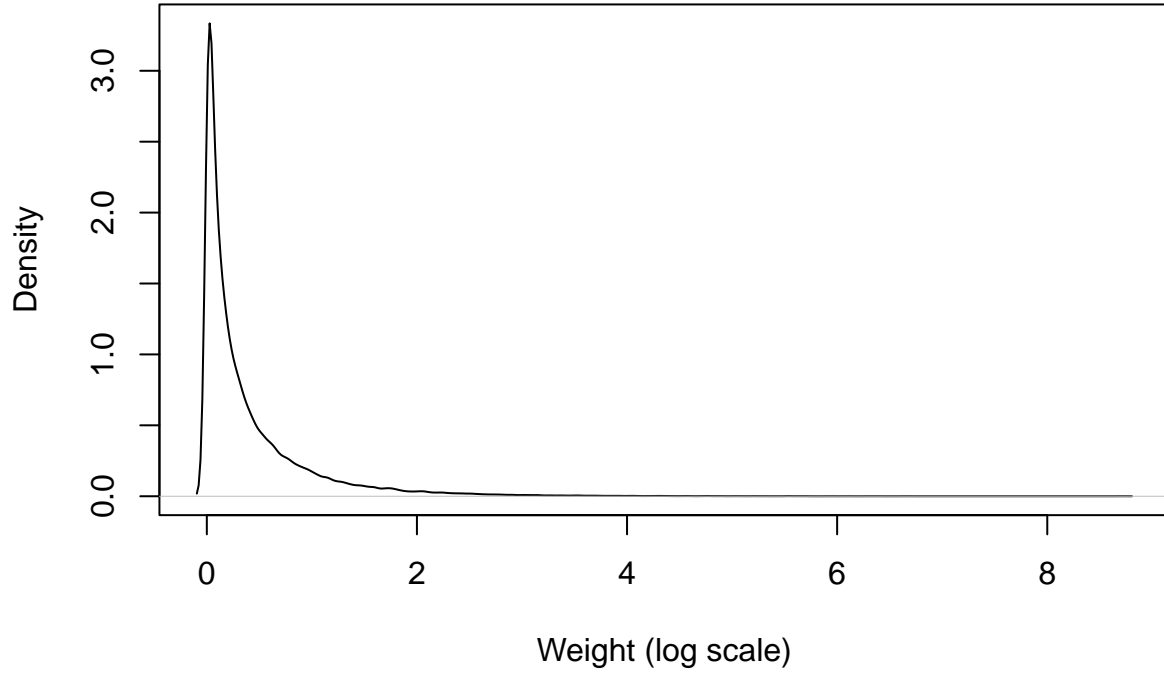
Use weighting to overcome issues with observed confounding variables

Weight the data

Use the package ipw

What's the range of weights?

```
## [1] 1.00 6069.64
```

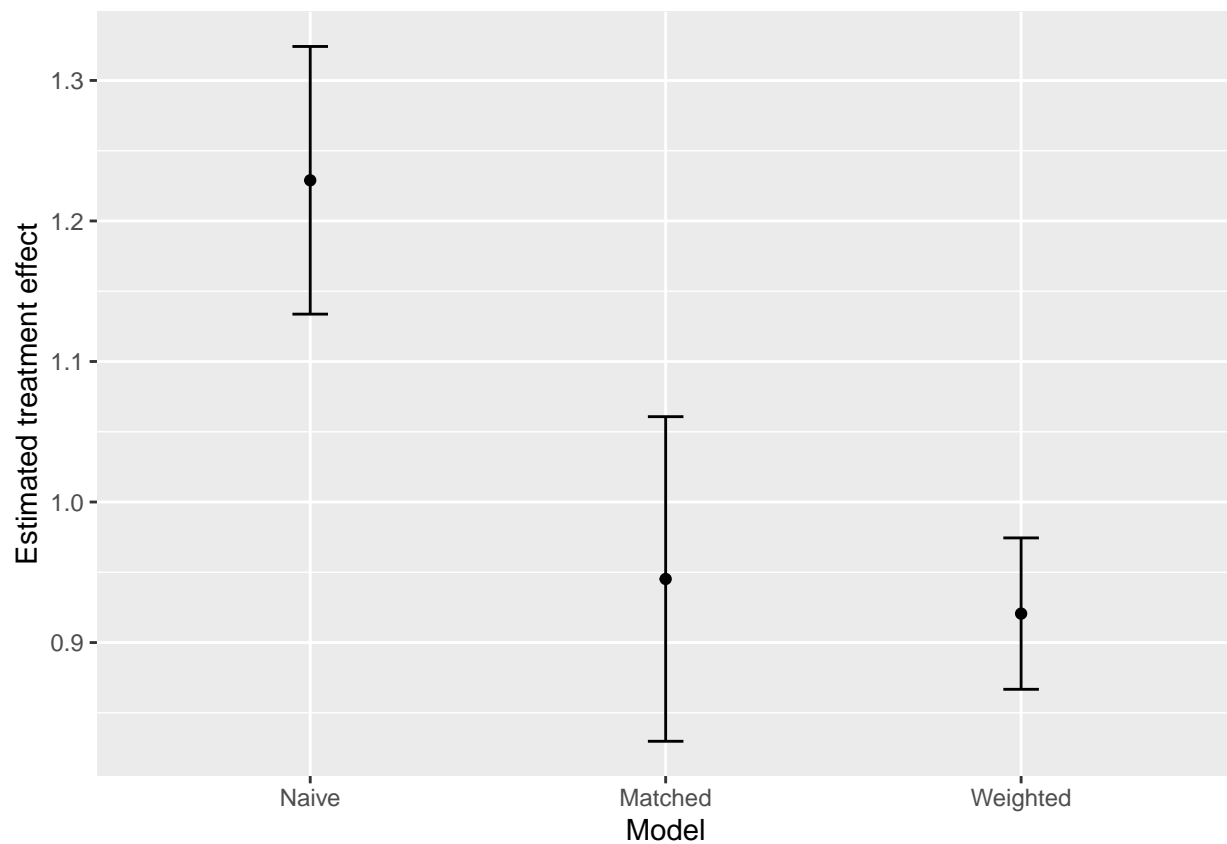
Model the effect of ownership

Table 7: Coefficient estimates for model with weighting

Variable	Estimate	Std. Error	p value
(Intercept)	-11.151	2.821	0.000
prot_cat_recl1	0.921	0.054	0.000
dist_rds_km	-0.014	0.026	0.581
slope	0.011	0.003	0.000
aspect_srai	0.077	0.066	0.245
elev_km	1.890	0.174	0.000
pdsi_avg_winter	-0.018	0.002	0.000
pdsi_avg_spring	-0.020	0.002	0.000
pdsi_avg_summer	0.036	0.003	0.000
pdsi_avg_fall	-0.003	0.002	0.130
soil_avg_winter	-0.001	0.001	0.331
soil_avg_spring	0.008	0.001	0.000
soil_avg_summer	-0.003	0.001	0.005
soil_avg_fall	-0.002	0.001	0.000
tmmn_avg_winter	-0.118	0.008	0.000
tmmn_avg_spring	0.059	0.019	0.002
tmmn_avg_summer	-0.127	0.017	0.000
tmmn_avg_fall	0.158	0.021	0.000
tmmx_avg_winter	0.152	0.009	0.000

Variable	Estimate	Std. Error	p value
tmmx_avg_spring	-0.026	0.016	0.108
tmmx_avg_summer	0.299	0.016	0.000
tmmx_avg_fall	-0.431	0.023	0.000
vs_max_winter	0.054	0.003	0.000
vs_max_spring	-0.060	0.004	0.000
vs_max_summer	0.001	0.004	0.863
vs_max_fall	-0.042	0.005	0.000
total_precip_winter	-0.165	0.006	0.000
total_precip_spring	-0.061	0.005	0.000
total_precip_summer	0.089	0.006	0.000
total_precip_fall	0.039	0.004	0.000
prev_yr_precip	0.004	0.002	0.060

Compare outputs from the naive, matched, and weighted regressions



```
## List of 136
## $ line :List of 6
## ..$ colour      : chr "black"
## ..$ linewidth    : num 0.5
## ..$ linetype     : num 1
## ..$ lineend      : chr "butt"
## ..$ arrow        : logi FALSE
## ..$ inherit.blank: logi TRUE
```

```

##   .- attr(*, "class")= chr [1:2] "element_line" "element"
## $ rect                                     :List of 5
##   ..$ fill           : chr "white"
##   ..$ colour         : chr "black"
##   ..$ linewidth      : num 0.5
##   ..$ linetype       : num 1
##   ..$ inherit.blank: logi TRUE
##   .- attr(*, "class")= chr [1:2] "element_rect" "element"
## $ text                                     :List of 11
##   ..$ family         : chr ""
##   ..$ face           : chr "plain"
##   ..$ colour         : chr "black"
##   ..$ size           : num 11
##   ..$ hjust          : num 0.5
##   ..$ vjust          : num 0.5
##   ..$ angle          : num 0
##   ..$ lineheight     : num 0.9
##   ..$ margin         : 'margin' num [1:4] 0points 0points 0points 0points
##   .. .- attr(*, "unit")= int 8
##   ..$ debug          : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   .- attr(*, "class")= chr [1:2] "element_text" "element"
## $ title                                     : NULL
## $ aspect.ratio                             : NULL
## $ axis.title                             : NULL
## $ axis.title.x                             :List of 11
##   ..$ family         : NULL
##   ..$ face           : NULL
##   ..$ colour         : NULL
##   ..$ size           : NULL
##   ..$ hjust          : NULL
##   ..$ vjust          : num 1
##   ..$ angle          : NULL
##   ..$ lineheight     : NULL
##   ..$ margin         : 'margin' num [1:4] 2.75points 0points 0points 0points
##   .. .- attr(*, "unit")= int 8
##   ..$ debug          : NULL
##   ..$ inherit.blank: logi TRUE
##   .- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.x.top                         :List of 11
##   ..$ family         : NULL
##   ..$ face           : NULL
##   ..$ colour         : NULL
##   ..$ size           : NULL
##   ..$ hjust          : NULL
##   ..$ vjust          : num 0
##   ..$ angle          : NULL
##   ..$ lineheight     : NULL
##   ..$ margin         : 'margin' num [1:4] 0points 0points 2.75points 0points
##   .. .- attr(*, "unit")= int 8
##   ..$ debug          : NULL
##   ..$ inherit.blank: logi TRUE
##   .- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.x.bottom                     : NULL

```

```

## $ axis.title.y                                :List of 11
## ..$ family      : NULL
## ..$ face         : NULL
## ..$ colour       : NULL
## ..$ size         : NULL
## ..$ hjust        : NULL
## ..$ vjust        : num 1
## ..$ angle        : num 90
## ..$ lineheight   : NULL
## ..$ margin       : 'margin' num [1:4] 0points 2.75points 0points 0points
## .. ..- attr(*, "unit")= int 8
## ..$ debug        : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.title.y.left                          : NULL
## $ axis.title.y.right                        :List of 11
## ..$ family      : NULL
## ..$ face         : NULL
## ..$ colour       : NULL
## ..$ size         : NULL
## ..$ hjust        : NULL
## ..$ vjust        : num 1
## ..$ angle        : num -90
## ..$ lineheight   : NULL
## ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.75points
## .. ..- attr(*, "unit")= int 8
## ..$ debug        : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text                                    :List of 11
## ..$ family      : NULL
## ..$ face         : NULL
## ..$ colour       : chr "grey30"
## ..$ size         : 'rel' num 0.8
## ..$ hjust        : NULL
## ..$ vjust        : NULL
## ..$ angle        : NULL
## ..$ lineheight   : NULL
## ..$ margin       : NULL
## ..$ debug        : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.x                                :List of 11
## ..$ family      : NULL
## ..$ face         : NULL
## ..$ colour       : NULL
## ..$ size         : NULL
## ..$ hjust        : NULL
## ..$ vjust        : num 1
## ..$ angle        : NULL
## ..$ lineheight   : NULL
## ..$ margin       : 'margin' num [1:4] 2.2points 0points 0points 0points
## .. ..- attr(*, "unit")= int 8
## ..$ debug        : NULL

```

```

## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.x.top :List of 11
## ..$ family : NULL
## ..$ face : NULL
## ..$ colour : NULL
## ..$ size : NULL
## ..$ hjust : NULL
## ..$ vjust : num 0
## ..$ angle : NULL
## ..$ lineheight : NULL
## ..$ margin : 'margin' num [1:4] 0points 0points 2.2points 0points
## ..- attr(*, "unit")= int 8
## ..$ debug : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.x.bottom : NULL
## $ axis.text.y :List of 11
## ..$ family : NULL
## ..$ face : NULL
## ..$ colour : NULL
## ..$ size : NULL
## ..$ hjust : num 1
## ..$ vjust : NULL
## ..$ angle : NULL
## ..$ lineheight : NULL
## ..$ margin : 'margin' num [1:4] 0points 2.2points 0points 0points
## ..- attr(*, "unit")= int 8
## ..$ debug : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.y.left : NULL
## $ axis.text.y.right :List of 11
## ..$ family : NULL
## ..$ face : NULL
## ..$ colour : NULL
## ..$ size : NULL
## ..$ hjust : num 0
## ..$ vjust : NULL
## ..$ angle : NULL
## ..$ lineheight : NULL
## ..$ margin : 'margin' num [1:4] 0points 0points 0points 2.2points
## ..- attr(*, "unit")= int 8
## ..$ debug : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text.theta : NULL
## $ axis.text.r :List of 11
## ..$ family : NULL
## ..$ face : NULL
## ..$ colour : NULL
## ..$ size : NULL
## ..$ hjust : num 0.5
## ..$ vjust : NULL

```

```

## ..$ angle          : NULL
## ..$ lineheight     : NULL
## ..$ margin         : 'margin' num [1:4] 0points 2.2points 0points 2.2points
## ..- attr(*, "unit")= int 8
## ..$ debug          : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.ticks        :List of 6
## ..$ colour          : chr "grey20"
## ..$ linewidth       : NULL
## ..$ linetype         : NULL
## ..$ lineend         : NULL
## ..$ arrow           : logi FALSE
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_line" "element"
## $ axis.ticks.x      : NULL
## $ axis.ticks.x.top   : NULL
## $ axis.ticks.x.bottom : NULL
## $ axis.ticks.y       : NULL
## $ axis.ticks.y.left  : NULL
## $ axis.ticks.y.right : NULL
## $ axis.ticks.theta   : NULL
## $ axis.ticks.r        : NULL
## $ axis.minor.ticks.x.top : NULL
## $ axis.minor.ticks.x.bottom : NULL
## $ axis.minor.ticks.y.left : NULL
## $ axis.minor.ticks.y.right : NULL
## $ axis.minor.ticks.theta : NULL
## $ axis.minor.ticks.r    : NULL
## $ axis.ticks.length   : 'simpleUnit' num 2.75points
## ..- attr(*, "unit")= int 8
## $ axis.ticks.length.x : NULL
## $ axis.ticks.length.x.top : NULL
## $ axis.ticks.length.x.bottom : NULL
## $ axis.ticks.length.y : NULL
## $ axis.ticks.length.y.left : NULL
## $ axis.ticks.length.y.right : NULL
## $ axis.ticks.length.theta : NULL
## $ axis.ticks.length.r    : NULL
## $ axis.minor.ticks.length : 'rel' num 0.75
## $ axis.minor.ticks.length.x : NULL
## $ axis.minor.ticks.length.x.top : NULL
## $ axis.minor.ticks.length.x.bottom : NULL
## $ axis.minor.ticks.length.y : NULL
## $ axis.minor.ticks.length.y.left : NULL
## $ axis.minor.ticks.length.y.right : NULL
## $ axis.minor.ticks.length.theta : NULL
## $ axis.minor.ticks.length.r    : NULL
## $ axis.line                   : list()
## ..- attr(*, "class")= chr [1:2] "element_blank" "element"
## $ axis.line.x                 : NULL
## $ axis.line.x.top              : NULL
## $ axis.line.x.bottom           : NULL
## $ axis.line.y                  : NULL

```

```

## $ axis.line.y.left           : NULL
## $ axis.line.y.right         : NULL
## $ axis.line.theta           : NULL
## $ axis.line.r               : NULL
## $ legend.background          :List of 5
##   ..$ fill                   : NULL
##   ..$ colour                 : logi NA
##   ..$ linewidth              : NULL
##   ..$ linetype               : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
## $ legend.margin              : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##   ..- attr(*, "unit")= int 8
## $ legend.spacing             : 'simpleUnit' num 11points
##   ..- attr(*, "unit")= int 8
## $ legend.spacing.x           : NULL
## $ legend.spacing.y           : NULL
## $ legend.key                 : NULL
## $ legend.key.size            : 'simpleUnit' num 1.2lines
##   ..- attr(*, "unit")= int 3
## $ legend.key.height          : NULL
## $ legend.key.width           : NULL
## $ legend.key.spacing         : 'simpleUnit' num 5.5points
##   ..- attr(*, "unit")= int 8
## $ legend.key.spacing.x       : NULL
## $ legend.key.spacing.y       : NULL
## $ legend.frame               : NULL
## $ legend.ticks               : NULL
## $ legend.ticks.length        : 'rel' num 0.2
## $ legend.axis.line           : NULL
## $ legend.text                :List of 11
##   ..$ family                 : NULL
##   ..$ face                   : NULL
##   ..$ colour                 : NULL
##   ..$ size                   : 'rel' num 0.8
##   ..$ hjust                  : NULL
##   ..$ vjust                  : NULL
##   ..$ angle                  : NULL
##   ..$ lineheight             : NULL
##   ..$ margin                 : NULL
##   ..$ debug                  : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ legend.text.position        : NULL
## $ legend.title               :List of 11
##   ..$ family                 : NULL
##   ..$ face                   : NULL
##   ..$ colour                 : NULL
##   ..$ size                   : NULL
##   ..$ hjust                  : num 0
##   ..$ vjust                  : NULL
##   ..$ angle                  : NULL
##   ..$ lineheight             : NULL
##   ..$ margin                 : NULL

```

```

## ..$ debug          : NULL
## ..$ inherit.blank: logi TRUE
## ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ legend.title.position      : NULL
## $ legend.position            : chr "right"
## $ legend.position.inside     : NULL
## $ legend.direction           : NULL
## $ legend.byrow               : NULL
## $ legend.justification       : chr "center"
## $ legend.justification.top   : NULL
## $ legend.justification.bottom : NULL
## $ legend.justification.left  : NULL
## $ legend.justification.right : NULL
## $ legend.justification.inside : NULL
## $ legend.location           : NULL
## $ legend.box                 : NULL
## $ legend.box.just            : NULL
## $ legend.box.margin          : 'margin' num [1:4] 0cm 0cm 0cm 0cm
## ..- attr(*, "unit")= int 1
## $ legend.box.background      : list()
## ..- attr(*, "class")= chr [1:2] "element_blank" "element"
## $ legend.box.spacing         : 'simpleUnit' num 11points
## ..- attr(*, "unit")= int 8
## [list output truncated]
## - attr(*, "class")= chr [1:2] "theme" "gg"
## - attr(*, "complete")= logi TRUE
## - attr(*, "validate")= logi TRUE

```