

Your project title

Info Innovators - Kevin Mao, Arnav Meduri, Ben Trokenheim, Ricardo Urena

2025-03-20

```
Rows: 55,367
Columns: 43
$ ...1      <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ~
$ `Unnamed: 0` <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ~
$ fire_name    <chr> NA, NA, NA, "WNA 1", NA, NA, NA, NA, NA, "CRAZY", NA-
$ fire_size     <dbl> 10.0, 3.0, 60.0, 1.0, 2.0, 1.0, 5.2, 1.0, 1.0, 1.0, 8-
$ fire_size_class <chr> "C", "B", "C", "B", "B", "B", "B", "B", "B", "B"-
$ stat_cause_descr <chr> "Missing/Undefined", "Arson", "Arson", "Debris Burnin-
$ latitude       <dbl> 18.10507, 35.03833, 34.94780, 39.64140, 30.70060, 32.-
$ longitude      <dbl> -66.75304, -87.61000, -88.72250, -119.30830, -90.5914-
$ state          <chr> "PR", "TN", "MS", "NV", "LA", "GA", "GA", "TX", "AR", ~
$ disc_clean_date <chr> "2/11/2007", "12/11/2006", "2/29/2004", "6/6/2005", "-
$ cont_clean_date <chr> NA, NA, NA, "6/6/2005", NA, NA, NA, NA, "4/19/200-
$ discovery_month <chr> "Feb", "Dec", "Feb", "Jun", "Sep", "Mar", "Jan", "Nov-
$ disc_date_final <chr> NA, NA, NA, "6/6/2005 2:14", NA, NA, NA, NA, "4/1-
$ cont_date_final <chr> NA, NA, NA, "6/6/2005 18:00", NA, NA, NA, NA, "4/-
$ putout_time     <chr> NA, NA, NA, "0 days 00:00:00.000000000", NA, NA, NA, ~
$ disc_date_pre   <chr> "1/12/2007", "11/11/2006", "1/30/2004", "5/7/2005", "-
$ disc_pre_year   <dbl> 2007, 2006, 2004, 2005, 1999, 1999, 2008, 2005, 2007, ~
$ disc_pre_month   <chr> "Jan", "Nov", "Jan", "May", "Aug", "Feb", "Dec", "Oct-
$ wstation_usaf    <chr> "785140", "723235", "723235", "724880", "722312", "72-
$ dstation_m       <dbl> 58917.70, 32665.34, 105122.15, 43433.33, 25940.57, 53-
$ wstation_wban    <dbl> 11603, 13896, 13896, 23185, 3908, 366, 99999, 99999, ~
$ wstation_byear   <dbl> 1945, 1978, 1978, 1942, 1987, 1942, 2006, 1984, 1973, ~
$ wstation_eyear   <dbl> 2018, 2020, 2020, 2020, 2016, 2020, 2020, 2020, 2020, ~
$ Vegetation       <dbl> 12, 15, 16, 0, 12, 12, 12, 15, 15, 12, 12, 16, 12-
$ fire_mag         <dbl> 1.0, 0.1, 1.0, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1-
$ weather_file     <chr> "785140-11603-2007.gz", "723235-13896-2006.gz", "7232-
$ Temp_pre_30      <dbl> 24.480974, 7.553433, 4.971930, 16.275967, -1.000000, ~
$ Temp_pre_15      <dbl> 24.716923, 7.010000, 5.782766, 18.996181, -1.000000, ~
$ Temp_pre_7       <dbl> 24.9025974, 0.3435294, 5.5587500, 18.1425641, -1.0000-
```

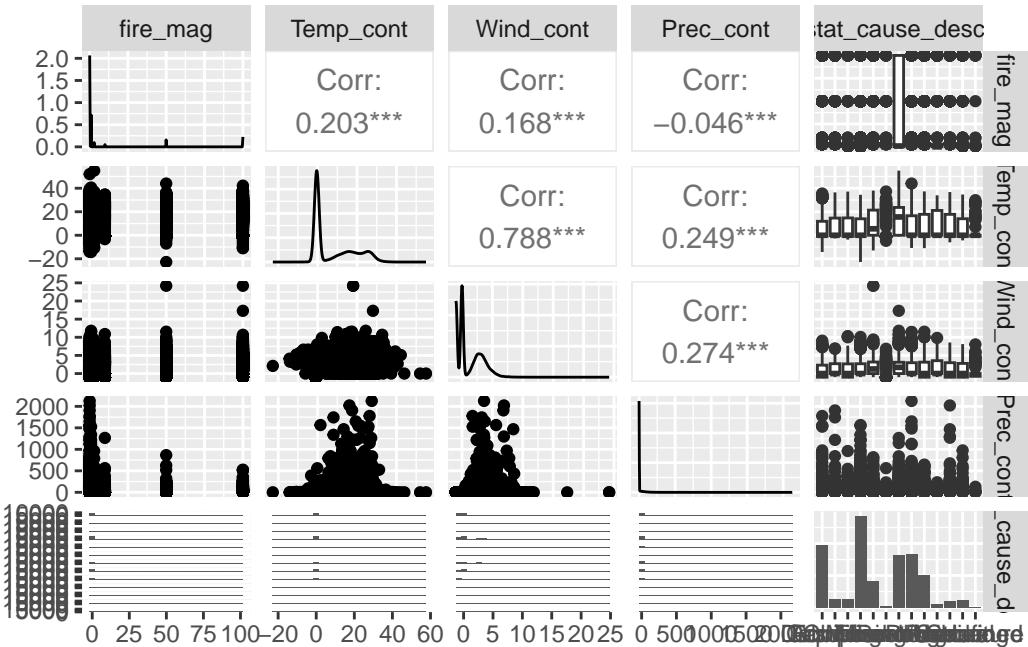
```

$ Temp_cont <dbl> 24.527961, 10.448298, 13.696600, 0.000000, -1.000000, ~
$ Wind_pre_30 <dbl> 4.341807, 2.709764, 3.364499, 4.054982, -1.000000, -1~
$ Wind_pre_15 <dbl> 3.492857, 2.881707, 2.923830, 3.398329, -1.000000, -1~
$ Wind_pre_7 <dbl> 3.262092, 1.976471, 2.695833, 3.671282, -1.000000, -1~
$ Wind_cont <dbl> 3.250413, 2.122320, 3.369050, 0.000000, -1.000000, -1~
$ Hum_pre_30 <dbl> 78.21659, 70.84000, 75.53163, 44.77843, -1.000000, -1.~ 
$ Hum_pre_15 <dbl> 76.79375, 65.85891, 75.86861, 37.14081, -1.000000, -1.~ 
$ Hum_pre_7 <dbl> 76.38158, 55.50588, 76.81283, 35.35385, -1.000000, -1.~ 
$ Hum_cont <dbl> 78.72437, 81.68268, 65.06380, 0.000000, -1.000000, -1.0~ 
$ Prec_pre_30 <dbl> 0.0, 59.8, 168.8, 10.4, -1.0, -1.0, 26.0, 28.4, 6.6, ~ 
$ Prec_pre_15 <dbl> 0.0, 8.4, 42.2, 7.2, -1.0, -1.0, 0.0, 27.5, 3.3, 1.8, ~ 
$ Prec_pre_7 <dbl> 0.0, 0.0, 18.1, 0.0, -1.0, -1.0, 0.0, 1.2, 0.0, 0.0, ~ 
$ Prec_cont <dbl> 0.0, 86.8, 124.5, 0.0, -1.0, -1.0, 0.0, 55.4, 46.4, 0~ 
$ remoteness <dbl> 0.01792339, 0.18435495, 0.19454351, 0.48744745, 0.214~ 

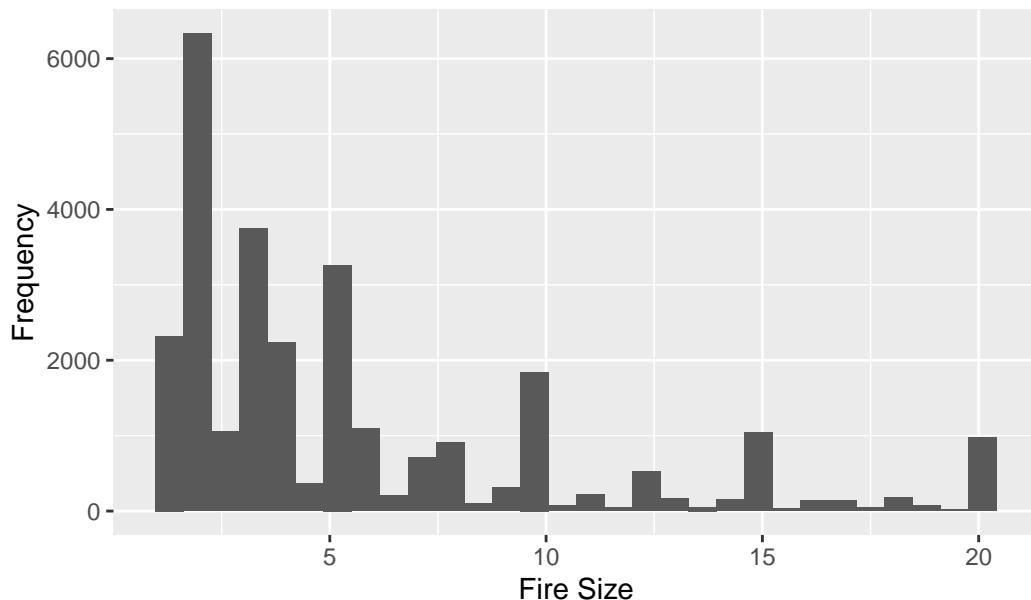
```

! Important

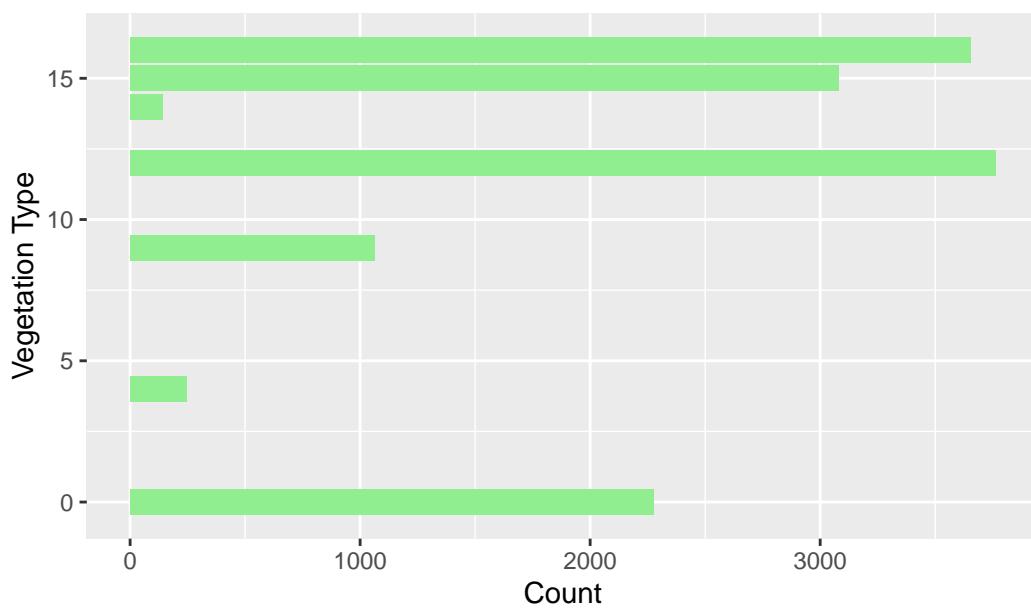
Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.



Distribution of Fire Sizes (Middle 50%)



Frequency of Vegetation Types

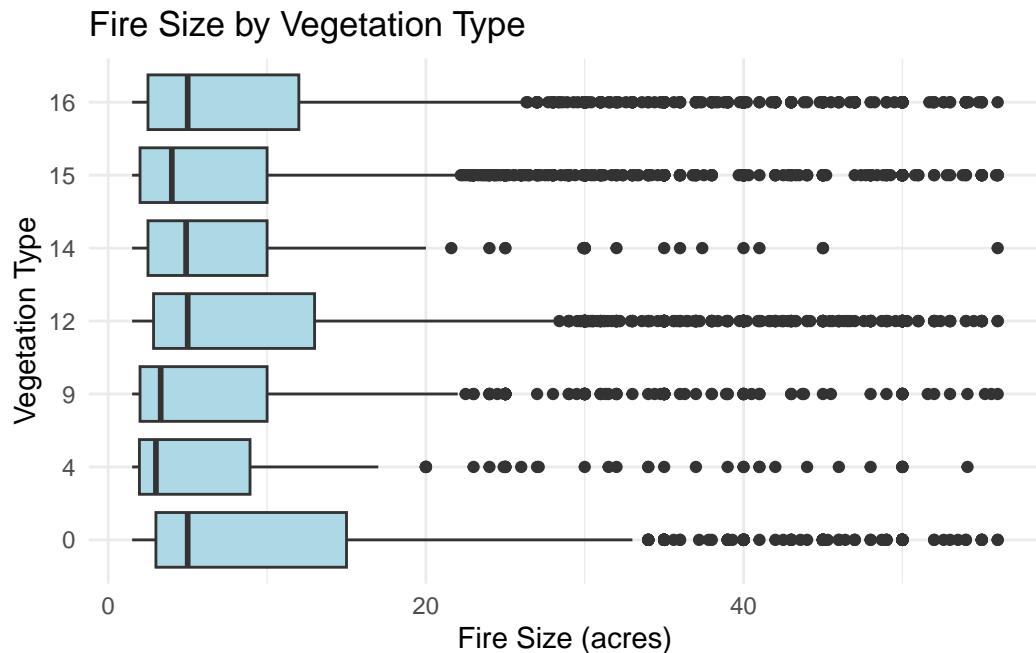


0	4	9	12	14	15	16
16.009558	1.728864	7.484714	26.445991	1.004990	21.652962	25.672921

The dataset includes 28 distinct vegetation types, each classified by a numerical code. Among these, a few vegetation types dominate the data. The most frequently occurring types are:

- Open Shrubland (code 12) with 3,763 observations (about 26% of the dataset)
- Secondary Tropical Evergreen Broadleaf Forest (code 16) with 3,653 observations (about 26%)
- Polar Desert/Rock/Ice (code 15) with 3,081 observations (about 22%)

Less common vegetation types represented in the data include desert and temperate evergreen needleleaf forests.



```
# A tibble: 7 x 4
  Vegetation count mean_fire_size median_fire_size
  <dbl> <int> <dbl> <dbl>
1 0 2278 11.0 5
2 12 3763 10.2 5
3 16 3653 9.67 5
4 14 143 8.98 4.9
```

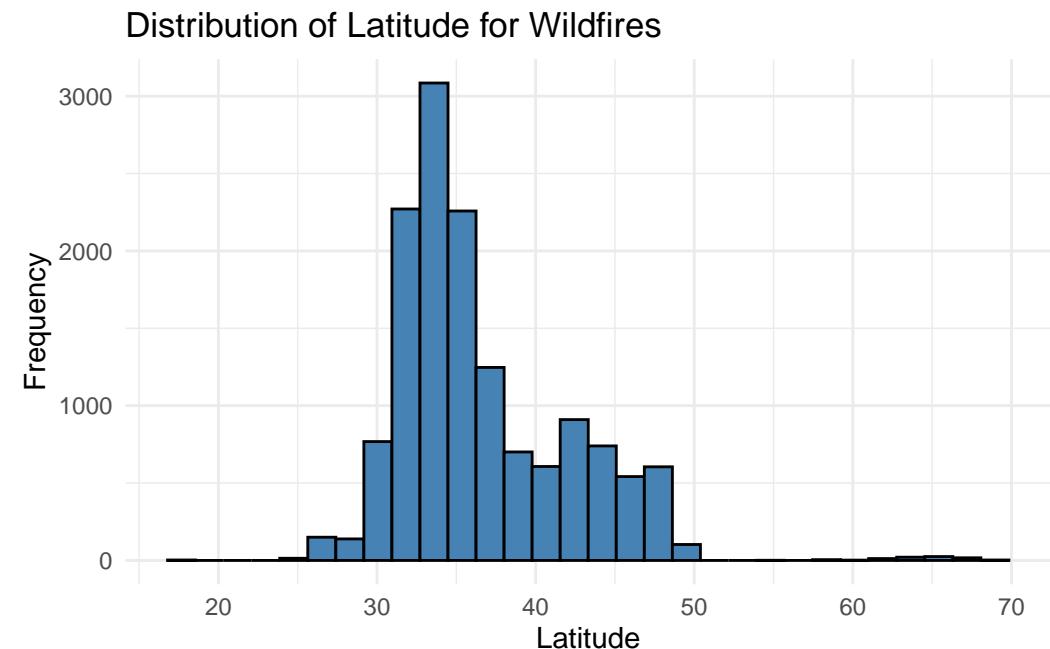
5	15	3081	8.91	4
6	4	246	8.39	3
7	9	1065	8.33	3.3

We explored the relationship between vegetation type and fire size using a boxplot and summary statistics. The boxplot shows variation in fire size distributions across vegetation types. Notably, vegetation type 0 has the largest spread and the highest median fire size at 5.0 acres. This type also has the highest mean fire size at 11.02 acres, suggesting a tendency for larger fires in this vegetation category.

Vegetation types 12 and 16 also exhibit relatively high median fire sizes (both around 5.0 acres) with mean fire sizes of 10.15 and 9.67 acres, respectively. Conversely, vegetation types 9 and 4 have lower median fire sizes (3.3 and 3.0 acres), along with the lowest mean fire sizes, 8.33 and 8.39 acres, respectively.

Overall, there is noticeable variation in fire size depending on vegetation type. Some vegetation types appear to be more prone to larger fires, which could be due to factors like fuel availability or vegetation density.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
17.98	32.85	34.84	36.62	39.93	69.26

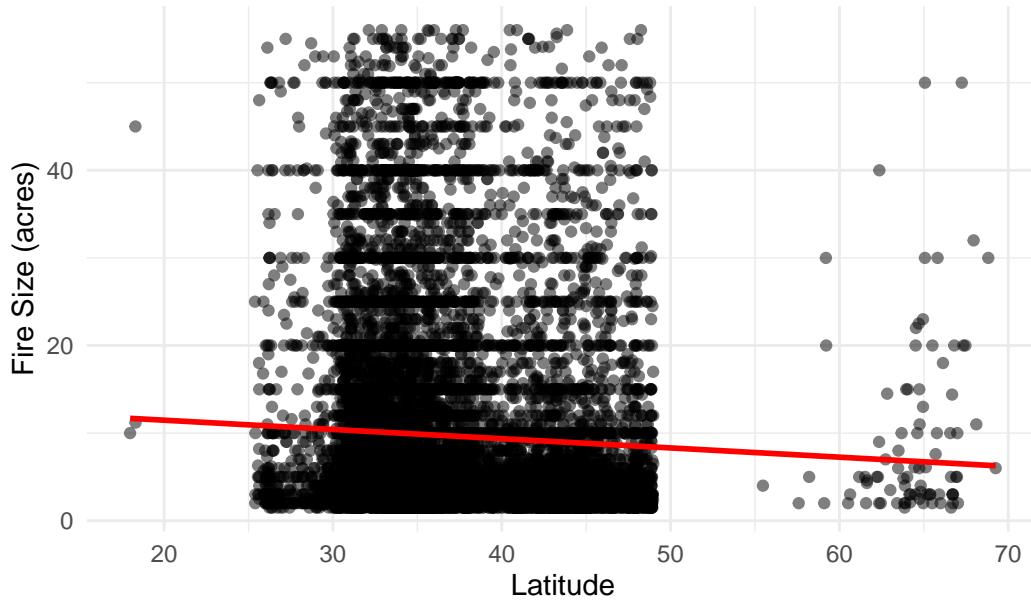


The latitude of wildfires in the dataset ranges from 17.98° to 69.26° , with a median of 34.84° and a mean of 36.62° . The middle 50% of the data falls between 32.85° and 39.93° , suggesting

that most wildfires occur in mid-latitude regions of the United States. This range corresponds to areas that commonly experience wildfires, such as parts of California and other western states.

The distribution of latitudes appears to be centered around the mid-30s to upper-30s, which may reflect the concentration of fire-prone areas in those geographic zones.

Fire Size vs. Latitude

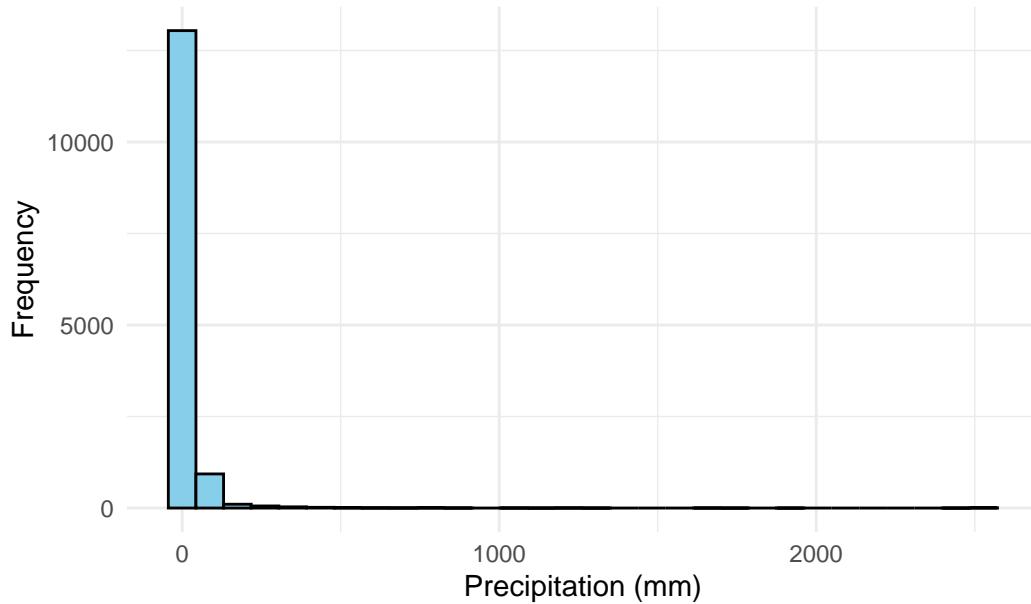


```
[1] -0.05139251
```

Not really any relationship between latitude and fire_size

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.00	-1.00	0.00	14.63	4.30	2527.00

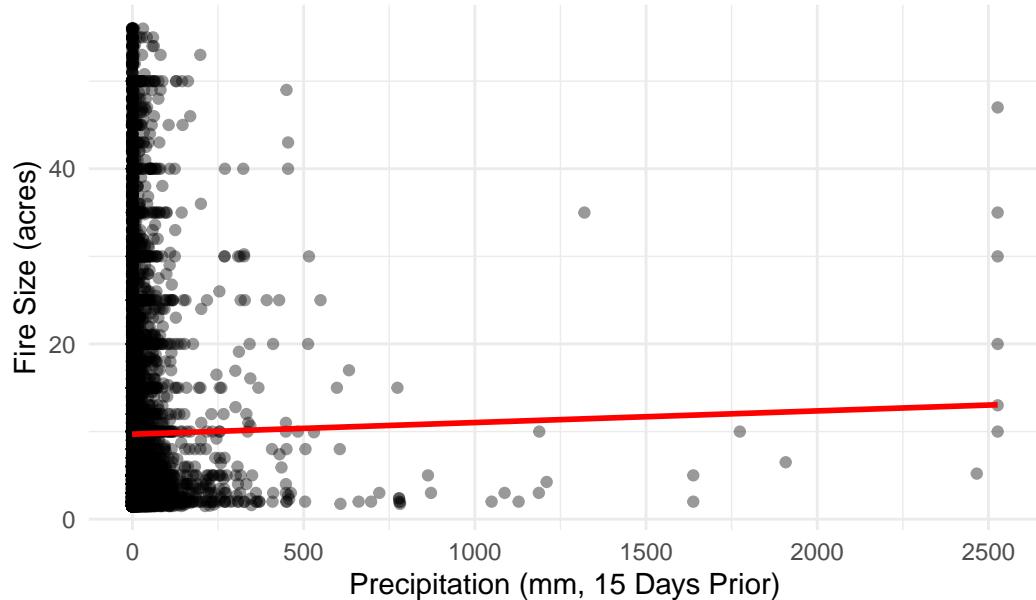
Distribution of Precipitation (15 Days Prior)



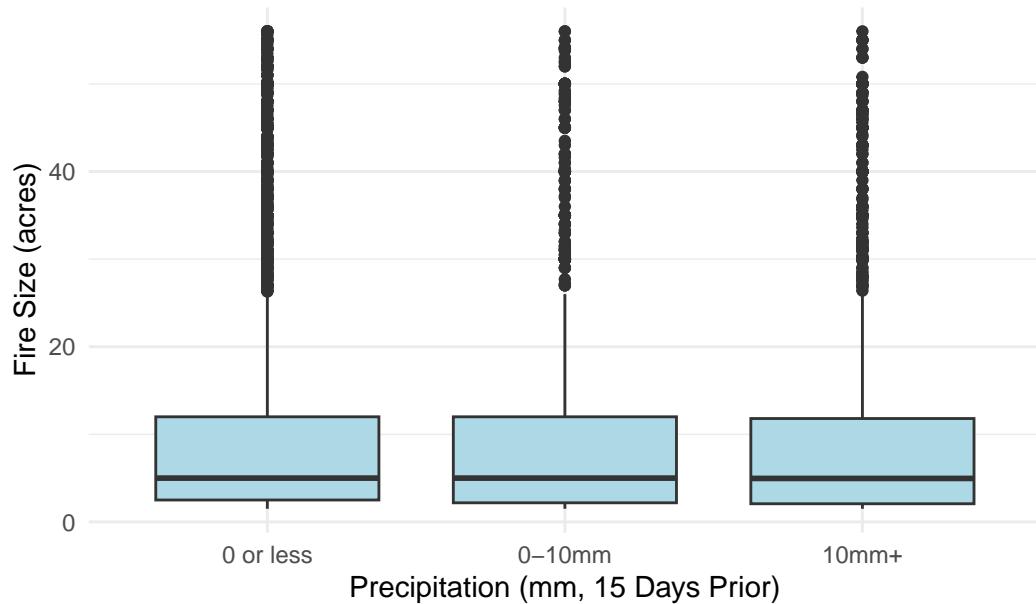
Most values are 0 mm or -1 mm, likely indicating missing data. The median is 0 mm, while the mean is 14.63 mm, skewed by extreme outliers (up to 2,527 mm).

The distribution is highly right-skewed, with most fires occurring after little to no precipitation, consistent with dry conditions increasing fire risk.

Fire Size vs. Precipitation (15 Days Prior)



Fire Size by Precipitation Bins



The scatterplot of `Prec_pre_15` and `fire_size` shows no strong relationship between prior precipitation and fire size. Most fires occurred with little to no precipitation, and there is substantial variability in fire size regardless of precipitation level. A few extreme precipitation values do not appear to have a significant effect on fire size.

The boxplot comparing precipitation bins (0 or less, 0-10mm, 10mm+) reveals similar distributions of fire size across all groups, with no meaningful differences in medians or spread. This suggests that precipitation up to 15 days prior to a fire may have limited impact on the size of the fire in this dataset.