Exploratory Analysis of California Wildfires (1992-2015)

Grant Esparza

September 18, 2018

About the Data

- Obtained from Kaggle.
- Subset of data from Fire Program Analysis fire-occurrence database which contains 1.88 million wildfire records.
- This analysis focuses on occurences in California, which has 189,550 Wildfires over the twenty-four year period.
- Even more specifically I'll be looking at wildfires caused by arson.

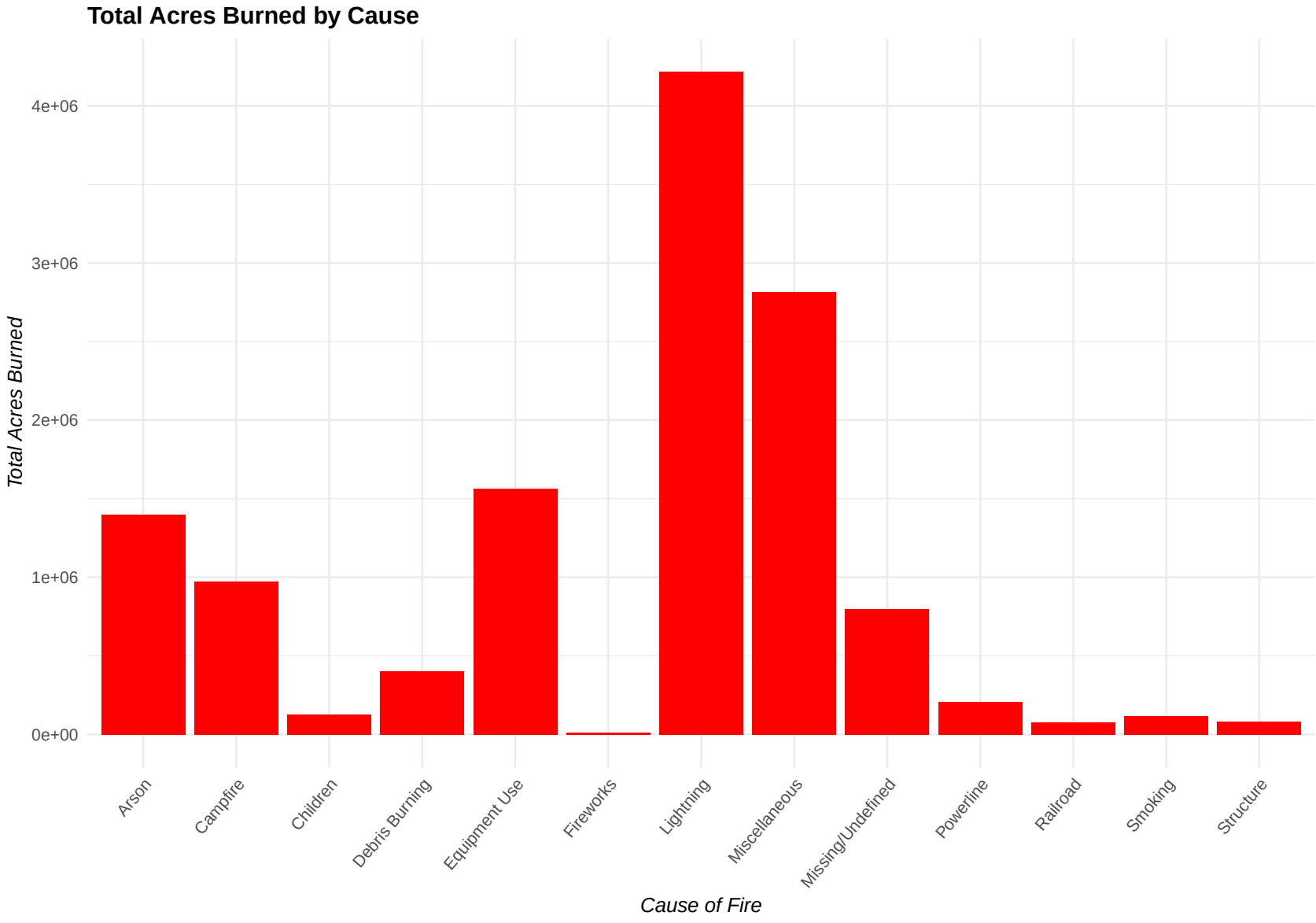
Data Cleaning

- Data was provided in the form of an SQLite database
- Using the library `RSQLite` I can send a query that will return a dataframe.
- From there I will group by counties in California and select occurences that were caused by arson.

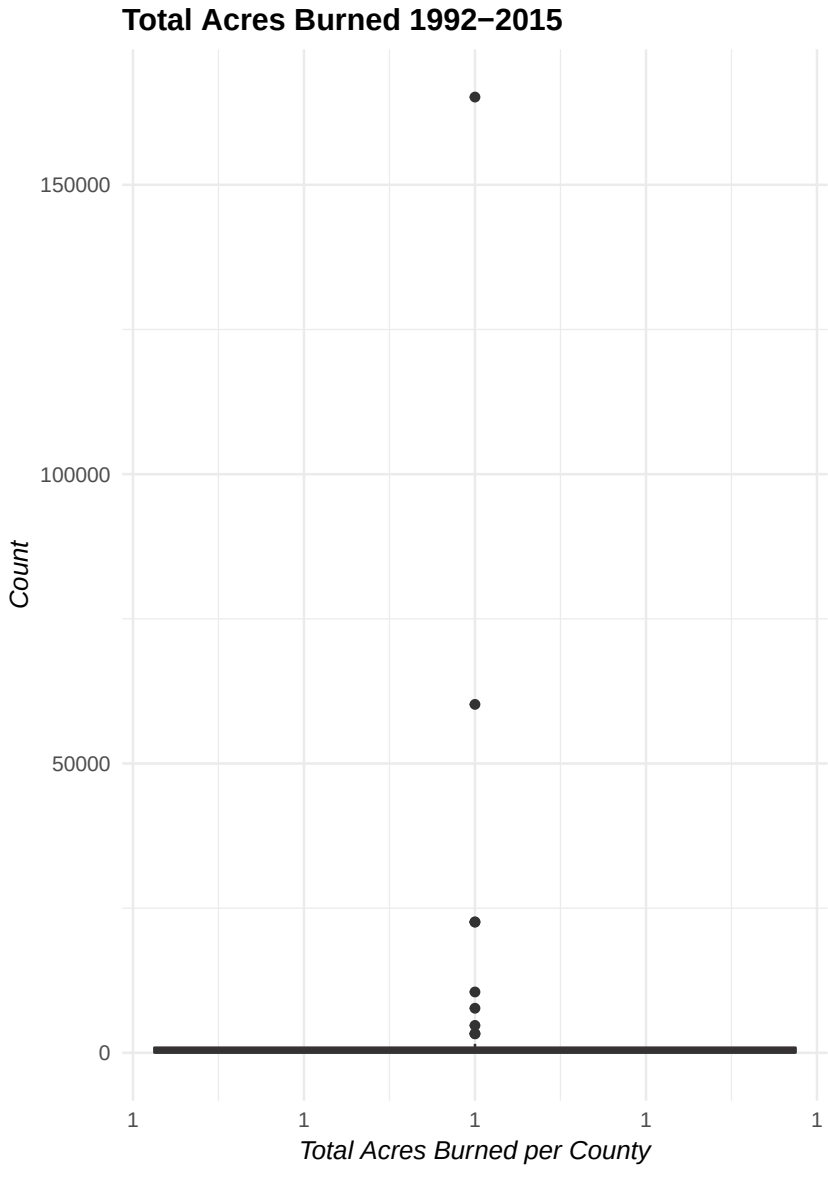
R Data Code

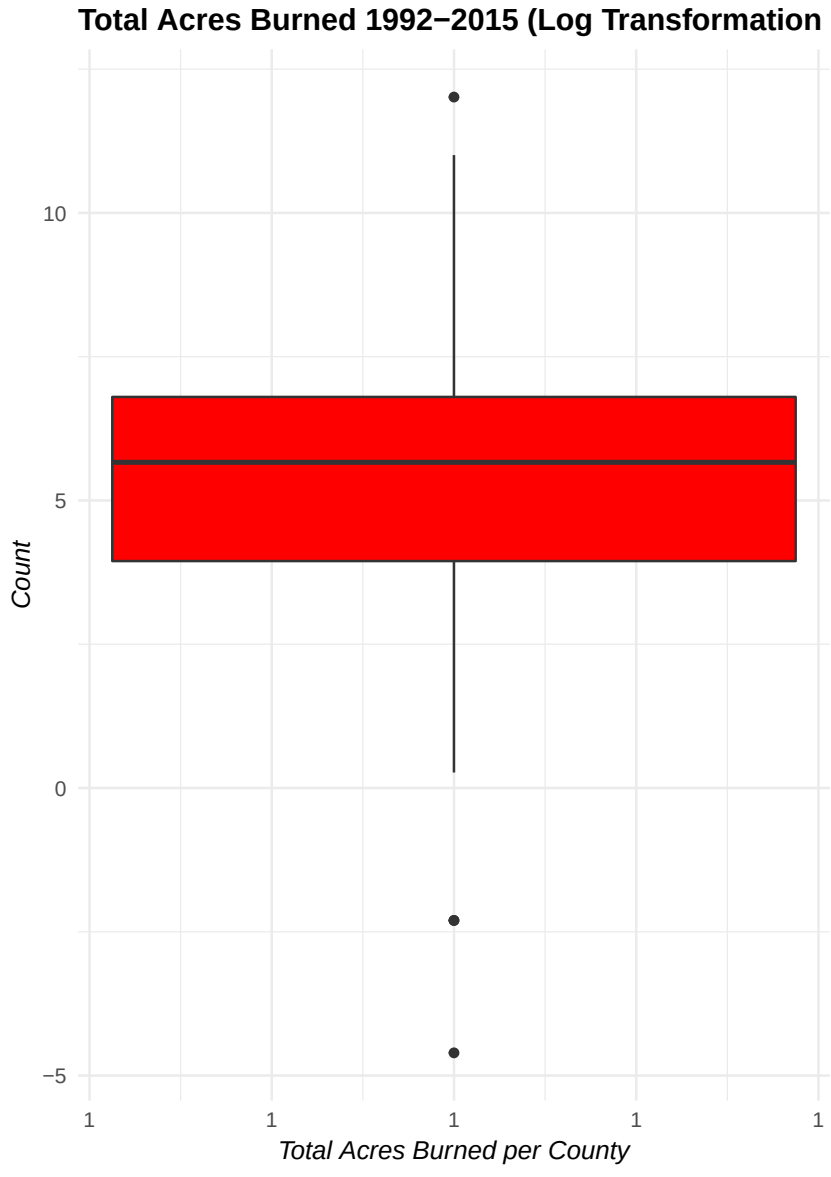
```
## Extract fire table from database
db <-dbConnect(SQLite(), "fire_database.sqlite")
res <- dbSendQuery(db, "SELECT * FROM Fires WHERE State == 'CA'")
cal_fires <- dbFetch(res)

## Get a tibble
cal_fires %>%
    filter(STAT_CAUSE_DESCR == "Arson") %>%
    group_by(FIPS_NAME) %>%
    summarise(occurences=n(), mn_size=mean(FIRE_SIZE), sum_size=sum(FIRE_SIZE)) -> num_fires
num_fires <- num_fires[1:57,]
```

Bar Chart


Total Acres Burned by Cause
Total Acres Burned
4e+06
3e+06
2e+06
1e+06
0e+00
Arson
Campfire
Children
Debris Burning
Equipment Use
Fireworks
Lightning
Miscellaneous
Missing/Undefined
Powerline
Railroad
Smoking
Structure
Cause of Fire

Boxplot


Total Acres Burned 1992−2015
150000
100000
50000
0
Count
1
1
1
1
1
Total Acres Burned per County


Total Acres Burned 1992−2015 (Log Transformation
10
5
0
−5
Count
1
1
1
1
1
Total Acres Burned per County

Bootstrap Mean CI

```
set.seed(385)
mean_sample_data <- function(data, idx) {
   mean(data[idx]) ## Mean of a vector
}

b <- boot(num_fires$sum_size, mean_sample_data, R=999)
boot.ci(b, type="perc")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = b, type = "perc")
##
## Intervals :
## Level     Percentile
## 95%   ( 1058, 12081 )
## Calculations and Intervals on Original Scale
```

We are 95% confident that the true mean number of acres burned in a California county is between 1058 and 12,081 acres.

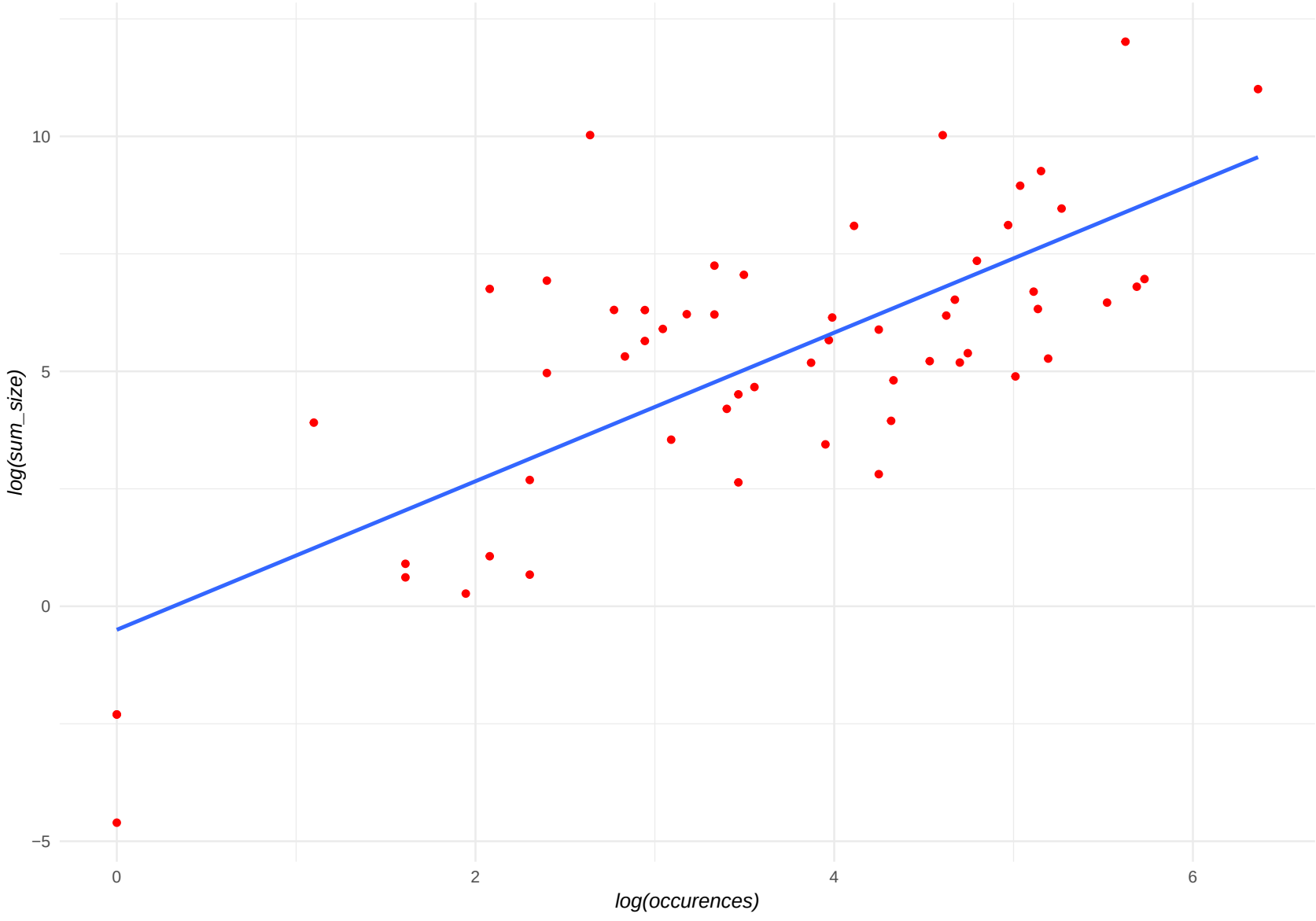
Bootstrap Median CI

```
set.seed(385)
median_sample_data <- function(data, idx) {
   median(data[idx]) ## Mean of a vector
}

b <- boot(num_fires$sum_size, median_sample_data, R=999)
boot.ci(b, type="perc")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = b, type = "perc")
##
## Intervals :
## Level     Percentile
## 95%   (178.2, 546.4 )
## Calculations and Intervals on Original Scale
```

We are 95% confident that the true median number of acres burned in a California county is between 178.2 and 546.4 acres.

Scatter Plot



Linear Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.50	0.75	-0.66	0.51
log(occurences)	1.58	0.19	8.21	0.00

Table 2: Fitting linear model: log(sum_size) ~ log(occurences)

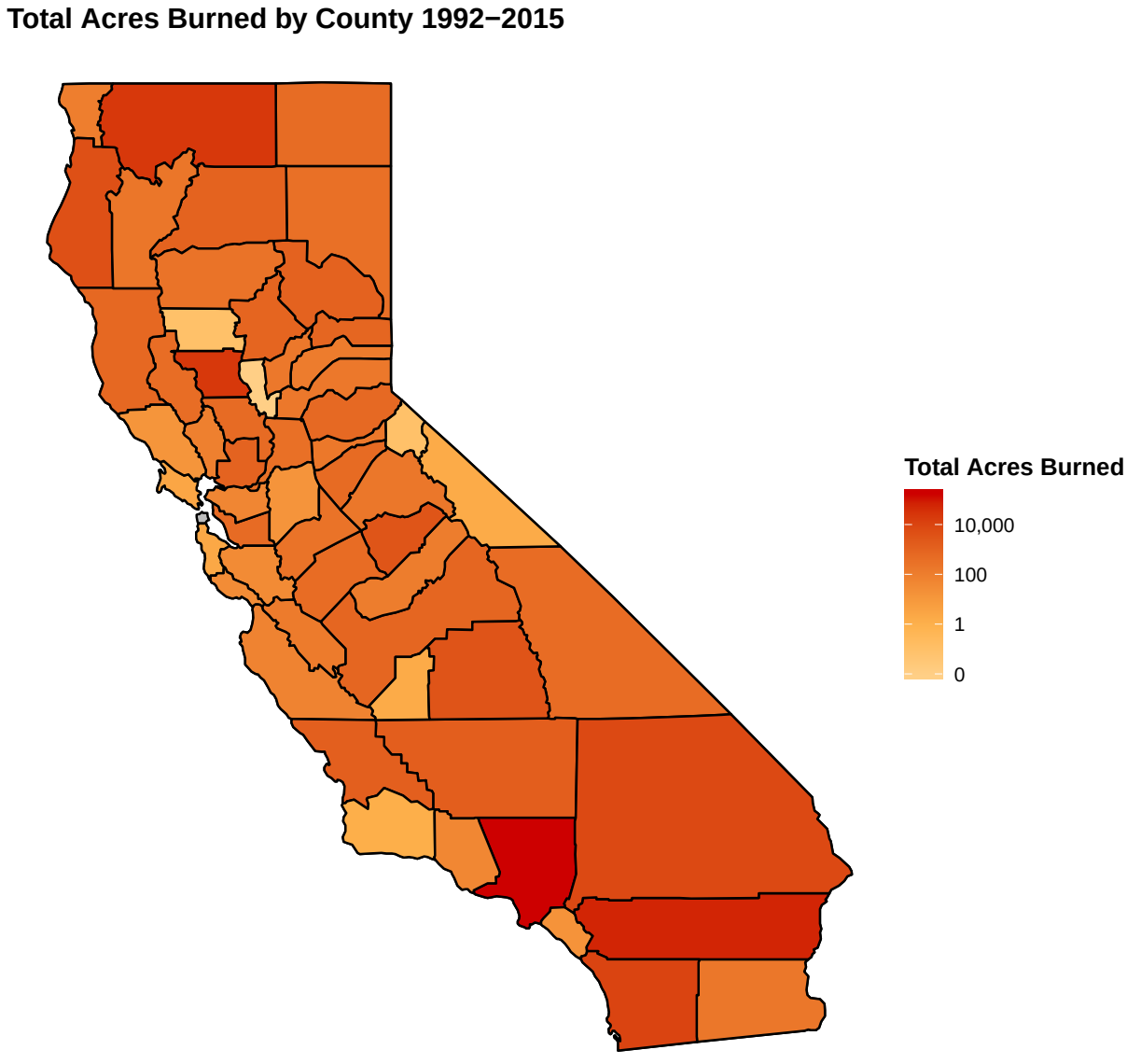
Observations	Residual Std. Error	R^2	Adjusted R^2
57	2.14	0.55	0.54

For every one percent increase in California wildfires in a county caused by arson, there is a 1.58% increase in total acres burned.

When the log of fires caused by arson is zero, the log of total acres burned will be -0.50

Creating Map of California Counties

Wildfire Map of Total Acres Burned by Arson



Wildfire Map With Biggest Fire per Year