

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 occurrences 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

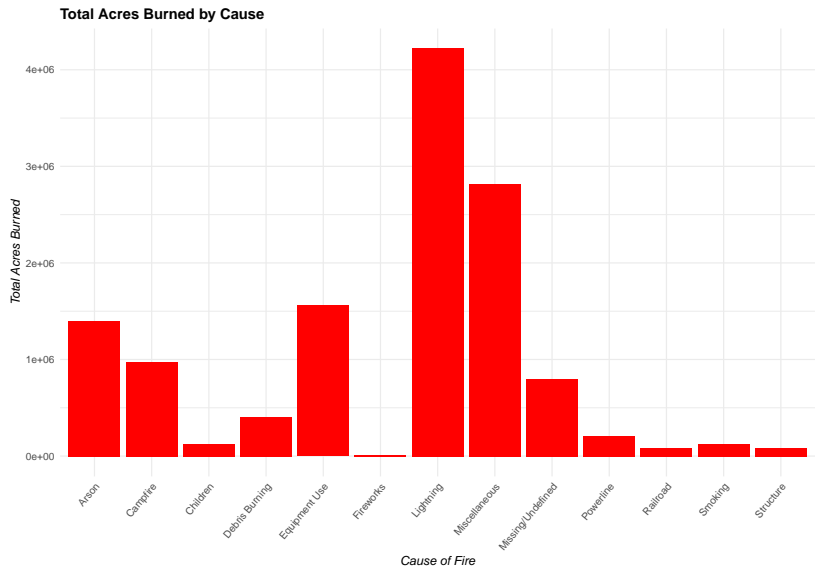
- ▶ Data was provided in the form of an SQLite database
- ▶ Using the library RSQLite I can send a query that will return a dataframe.
- ▶ From there I will group by counties in California and select occurrences that were caused by arson.

R Data Code

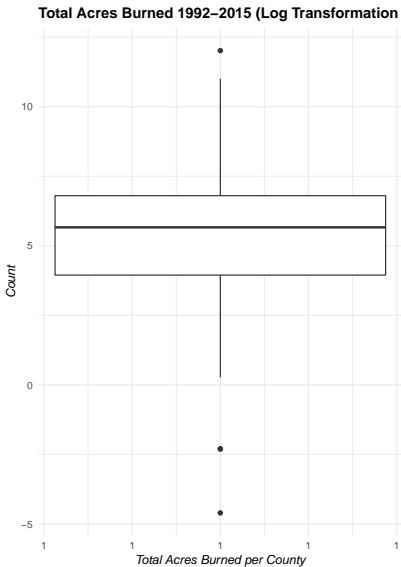
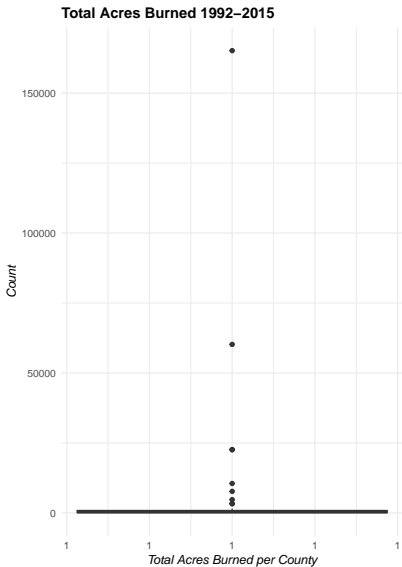
```
## Extract fire table from database
db <-dbConnect(SQLite(), "fire_database.sqlite")
res <- dbSendQuery(db, "SELECT * FROM Fires WHERE State == 'CA'")
cal_fires <- dbFetch(res)

## 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



Boxplot



Summary Statistics

```
pander(summarise(num_fires, Mean = mean(sum_size), Median = median(sum_size),  
                  Std_dev = sd(sum_size), IQR = IQR(sum_size)))
```

Mean	Median	Std_dev	IQR
5565	288.2	23296	846.5

With the mean being much higher than the median we can come to the conclusion there are several counties with a high value for total acres burned. With a high value for our standard deviation, we can also conclude that there is a large spread in our data (or some large outliers).

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.

Bootstrap Standard Deviation CI

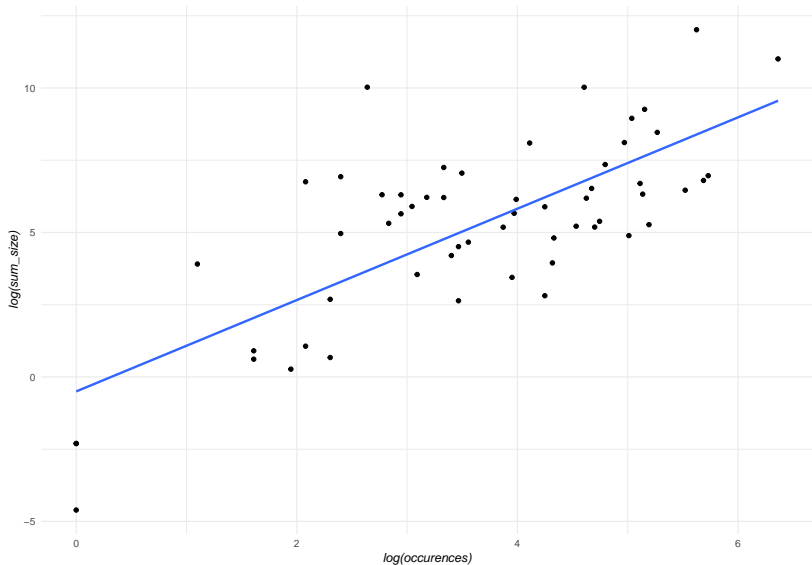
```
set.seed(385)
sd_sample_data <- function(data, idx) {
  sd(data[idx]) ## Mean of a vector
}

b <- boot(num_fires$sum_size, sd_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%      ( 3157, 38179 )
## Calculations and Intervals on Original Scale
```

We are 95% confident that the true standard deviation of acres burned in a California county is between 3,157 and 38,179 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 3: Fitting linear model: $\log(\text{sum_size}) \sim \log(\text{occurences})$

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.

Prediction

```
predict(fire_mod, data.frame(occurences=c(13, 350))) ## Take the exp of the result == total acres burned`
```

```
##          1          2  
## 3.554013 8.758995
```

If a county had 13 cases of arson the expected amount of acres burned would be 34.8133175.

If a county had 350 cases of arson the expected amount of acres burned would be 6310.6881081. Since 350 cases of arson cannot be found in the data set, this value should be considered with caution.

Creating Map of California Counties (Code)

```
## Map plot
counties <- map_data("county") %>% filter(region == "california")

## Fix county names
colnames(num_fires) <- c("subregion", "occurences", "mn_fire_size", "sum_fire_size")

## Fix subregion names
num_fires$subregion <- tolower(num_fires$subregion)

## Plot
cal_map <- ggplot(data=counties, aes(x=long, y=lat, group=group)) +
  coord_fixed(1.3) +
  geom_polygon(color="black", fill="gray") +
  theme_void()
```

Creating Map of California Counties (Map)



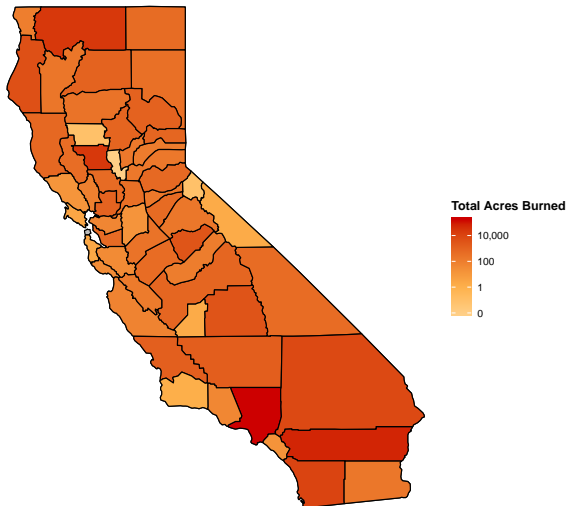
Wildfire Map of Total Acres Burned by Arson (Code)

```
## Join datasets
fire_map <- inner_join(counties, num_fires, by="subregion")

## Plot map with data
fire_base_map <- cal_map +
  geom_polygon(data=fire_map, aes(fill=sum_fire_size), color="black") +
  scale_fill_gradient2(trans="log10", low="#FFFFE0", mid="#FEB24C", high="#CD0000", labels=comma) +
  ggtitle("Total Acres Burned by County 1992-2015") +
  labs(fill="Total Acres Burned") +
  theme_void() +
  theme(title = element_text(face="bold"))
```


Wildfire Map of Total Acres Burned by Arson (Map)

Total Acres Burned by County 1992–2015



Wildfire Map With Biggest Fire per Year (Code)

```
## Create tibble of biggest fire per year
cal_fires %>%
  group_by(FIRE_YEAR) %>%
  summarise(max=which.max(FIRE_SIZE),
            subregion=cal_fires[which.max(FIRE_SIZE),]$FIPS_NAME,
            long=cal_fires[which.max(FIRE_SIZE),]$LONGITUDE,
            lat=cal_fires[which.max(FIRE_SIZE),]$LATITUDE) -> biggest_fire

## Plot with points
map_with_points <- fire_base_map +
  geom_point(data=biggest_fire, aes(x=long, y=lat), inherit.aes=FALSE, size=3) +
  geom_label_repel(data=biggest_fire, aes(x=long, y=lat, label = FIRE_YEAR),
                  box.padding = 0.35,
                  point.padding = 0.5,
                  segment.color = 'grey50', inherit.aes=FALSE)
```

Wildfire Map With Biggest Fire per Year (Map)

Total Acres Burned by County 1992–2015

