

Windfire Analysis in U.S.

Data Science as a Field - CU Boulder

Xingyu Chen
Kaitlyn McGrew
Tittiwat Tonburinthip
Kexin Yu
Thejas Kiran

September 06, 2021

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Related Applications	3
1.3	Observations and Questions	3
2	Design of Data and Methodology	4
2.1	Data Resource and Explanation of Variables	4
2.2	Preparing the Data	5
3	Exploration	14
3.1	Title 1	14
3.2	Title 2	15
3.3	Title 3	16
4	Problems Tackled and Conclusion	17
5	References	18

List of Figures

1	Average Number of US Wildfires by Day of Year. Observe the trend of wildfires count over Time	6
2	Number of US Wildfires by Size Class per year.	7
3	Number of US Wildfires by cause type.	8
4	Number of US Wildfires by cause type.	9
5	average wildfire size by cause.	10
6	US Wildfires, 1992-2018. The spectrum from white to darkred indicates more wildfires in that state	11
7	US Wildfires in CA, 1992-2018. The spectrum from white to darkred indicates more wildfires in that state	13

1 Introduction

1.1 Motivation

1.2 Related Applications

1.3 Observations and Questions

2 Design of Data and Methodology

```
library(RSQLite)
library(tidyverse)
library(dbplyr)
library(lubridate)
library(data.table)
library(scales)
library(usmap)
#library(leaflet)
library(kableExtra)
```

2.1 Data Resource and Explanation of Variables

```
# create db connection
conn <- dbConnect(SQLite(), 'FPA_FOD_20210617.sqlite')

# pull the fires table into RAM
fires <- tbl(conn, "Fires") %>% collect()

# disconnect from db
dbDisconnect(conn)

# select the column we need for this project
fires <- fires[,c('FIRE_NAME', 'DISCOVERY_DATE', 'NWCG_CAUSE_CLASSIFICATION', 'NWCG_GENRE')]

# kable related variable
kbl(text_tbl, booktabs = T, longtable = T, caption = "The related-variables in our data") %>%
  kable_styling(full_width = T) %>%
  column_spec(1, color = "red") %>%
  column_spec(2, width = "25em")
```

Table 1: The related-variables in our data set.

Related-Variable	Description
FIRE_NAME	Name of the incident from the fire report
FIRE_YEAR	Calendar year in which the fire was discovered or confirmed to exist.
DISCOVERY_DATE	Date on which the fire was discovered or confirmed to exist.
NWCG_CAUSE_CLASSIFICATION	Description of the (statistical) cause of the fire.
CONT_DATE	Date on which the fire was declared contained or otherwise controlled (mm/dd/yyyy where mm=month, dd=day, and yyyy=year).

FIRE_SIZE	Estimate of acres within the final perimeter of the fire.
FIRE_SIZE_CLASS	Code for fire size based on the number of acres within the final fire perimeter expenditures (A=greater than 0 but less than or equal to 0.25 acres, B=0.26-9.9 acres, C=10.0-99.9 acres, D=100-299 acres, E=300 to 999 acres, F=1000 to 4999 acres, and G=5000+ acres).
LATITUDE	Latitude (NAD83) for point location of the fire (decimal degrees).
LONGITUDE	Longitude (NAD83) for point location of the fire (decimal degrees).
STATE	Two-letter alphabetic code for the state in which the fire burned (or originated), based on the nominal designation in the fire report.
COUNTY	County, or equivalent, in which the fire burned (or originated), based on nominal designation in the fire report.

2.2 Preparing the Data

```
fires_1 <- as.data.frame(fires)

fires_1$DISCOVERY_DATE<-as.Date(fires_1$DISCOVERY_DATE, format = "%m/%d/%Y")

fires_1 <- fires_1 %>%
  mutate(day = format(DISCOVERY_DATE, "%d"),
         month = format(DISCOVERY_DATE, "%m"),
         year = format(DISCOVERY_DATE, "%Y")) %>%
  group_by(month, day) %>%
  summarise(total = n()) %>%
  mutate(date = make_date(month = month, day = day))

ggplot() +
  geom_line(aes(x = date, y = total/27), fires_1, color = 'orange') +
  scale_x_date(date_breaks= "1 month", date_labels = "%b") +
  xlab("Day of Year") + ylab("Number of wildfire") +
  theme(plot.background = element_rect(fill = "#BFD5E3"))

fires_2 <- as.data.frame(fires)
size_classes <- c('A' = '0-0.25',
                  'B' = '0.26-9.9',
                  'C' = '10.0-99.9',
                  'D' = '100-299',
                  'E' = '300-999',
                  'F' = '1000-4999',
```

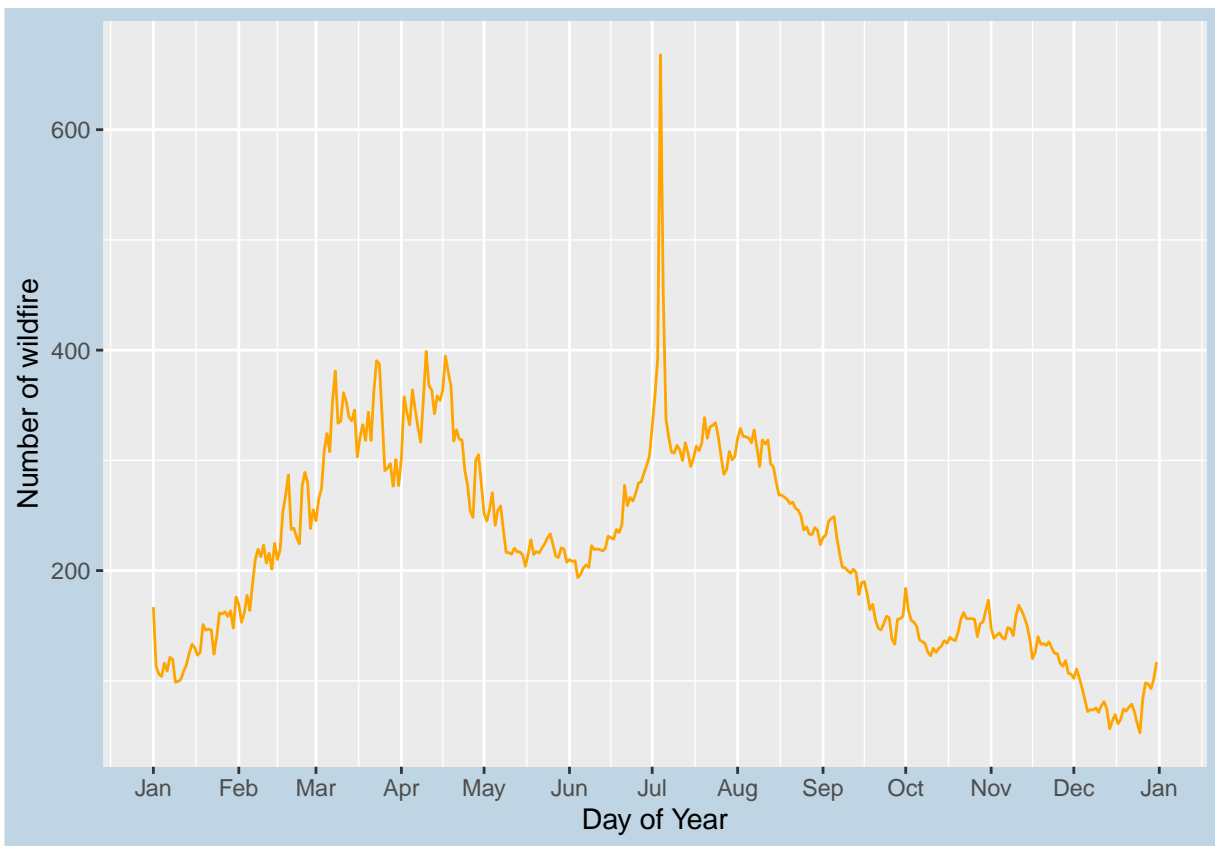


Figure 1: Average Number of US Wildfires by Day of Year. Observe the trend of wildfires count over Time

```

    'G' = '5000+')
fires_2 <- fires_2 %>%
  group_by(FIRE_SIZE_CLASS) %>%
  summarize(total = n()) %>%
  mutate(FIRE_SIZE_CLASS = size_classes[FIRE_SIZE_CLASS])

ggplot(data = fires_2, aes(x=FIRE_SIZE_CLASS, y = total/27, fill =FIRE_SIZE_CLASS)) +
  geom_bar(stat = "identity") +
  scale_fill_brewer(palette = "Reds") +
  xlab("Fire size (acres)") + ylab("Number of wildfires") +
  geom_text(label = paste0(round(fires_2$total/sum(fires_2$total)*100, 1), "%")) +
  theme(plot.background = element_rect(fill = "#BFD5E3"))

```

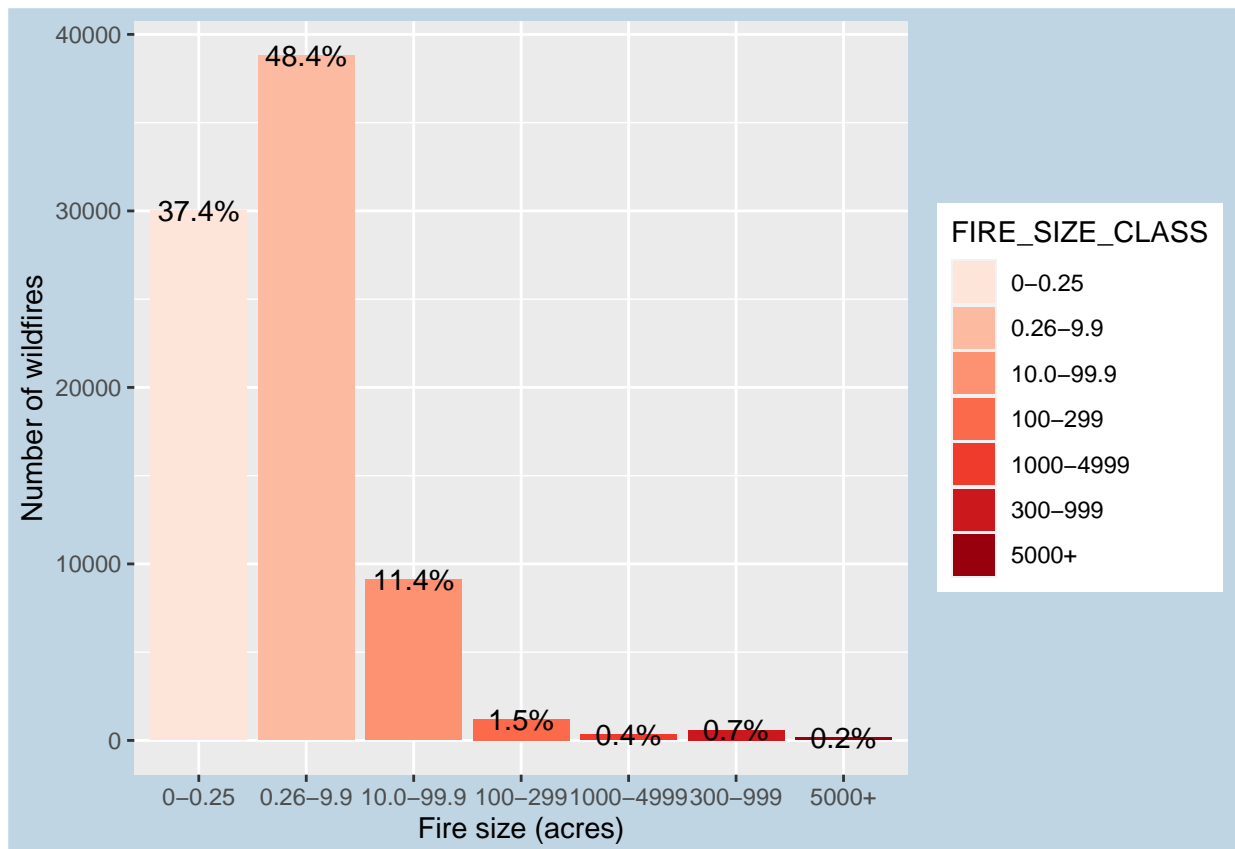


Figure 2: Number of US Wildfires by Size Class per year.

```

fires_3 <- as.data.frame(fires)
fires_3 <- fires_3 %>%
  group_by(NWCG_CAUSE_CLASSIFICATION) %>%
  summarize(total = n()) %>%
  na.omit() %>%
  arrange(desc(total))

```

```
ggplot(data = fires_3) +
  geom_bar(aes(x = "", y = total, fill = NWCG_CAUSE_CLASSIFICATION), stat = "identity")
  geom_text(aes(x = "", y = total, label = paste0(round(fires_3$total / sum(fires_3$total), 1), "%")))
  coord_polar(theta = "y") +
  theme_void()
```

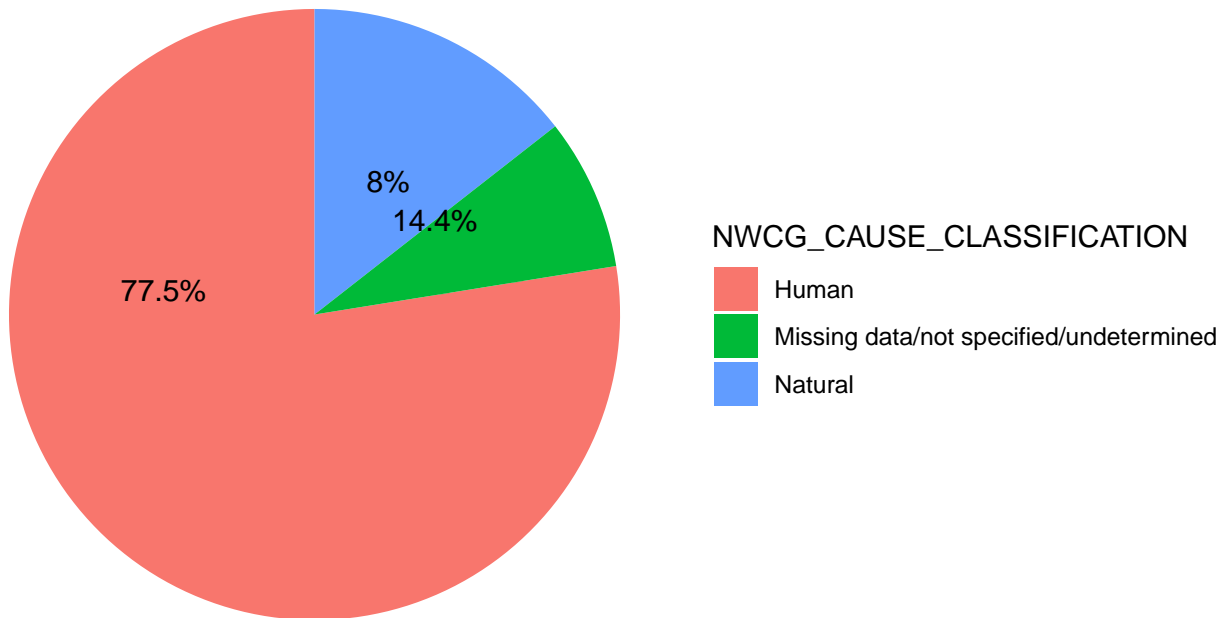


Figure 3: Number of US Wildfires by cause type.

```
fires_4 <- as.data.frame(fires)
fires_4 <- fires_4 %>%
  group_by(NWCG_GENERAL_CAUSE) %>%
  summarize(total = n()/1000) %>%
  na.omit() %>%
  arrange(desc(total))

ggplot(data = fires_4) +
  geom_bar(aes(x = reorder(NWCG_GENERAL_CAUSE, total), y = total), stat = "identity", fill = "#BFD5E3")
  coord_flip() +
  xlab("NWCG_GENERAL_CAUSE") + ylab("Number of wildfires (thousands)") +
  theme(plot.background = element_rect(fill = "#BFD5E3"))
```

```
fires_5 <- as.data.frame(fires)
fires_5 <- fires_5 %>%
  group_by(NWCG_GENERAL_CAUSE) %>%
  summarize(mean_size = mean(FIRE_SIZE, na.rm = TRUE)) %>%
  na.omit() %>%
  arrange(desc(mean_size))
```

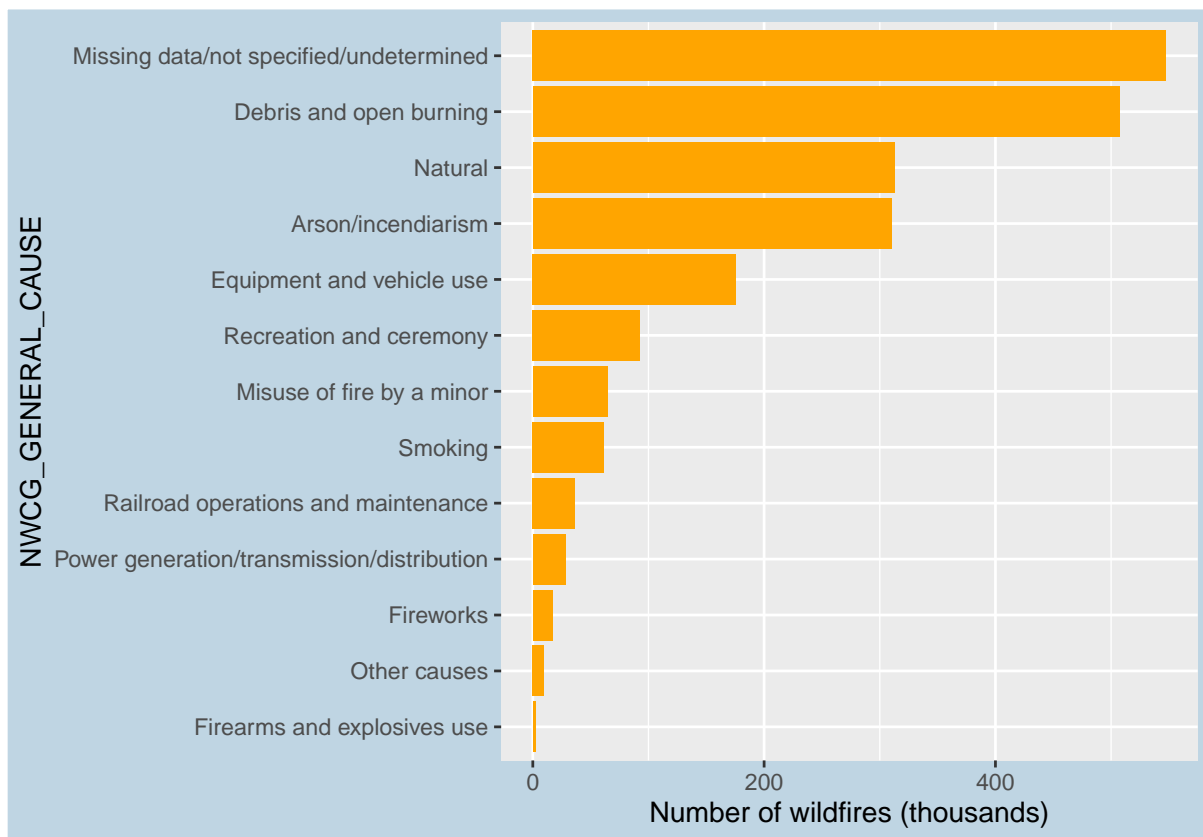



Figure 4: Number of US Wildfires by cause type.

```
ggplot(data = fires_5) +
  geom_bar(aes(x = reorder(NWCG_GENERAL_CAUSE, mean_size), y = mean_size), stat = "identity") +
  coord_flip() +
  xlab("NWCG_GENERAL_CAUSE") + ylab("Acres") +
  theme(plot.background = element_rect(fill = "#BFD5E3"))
```

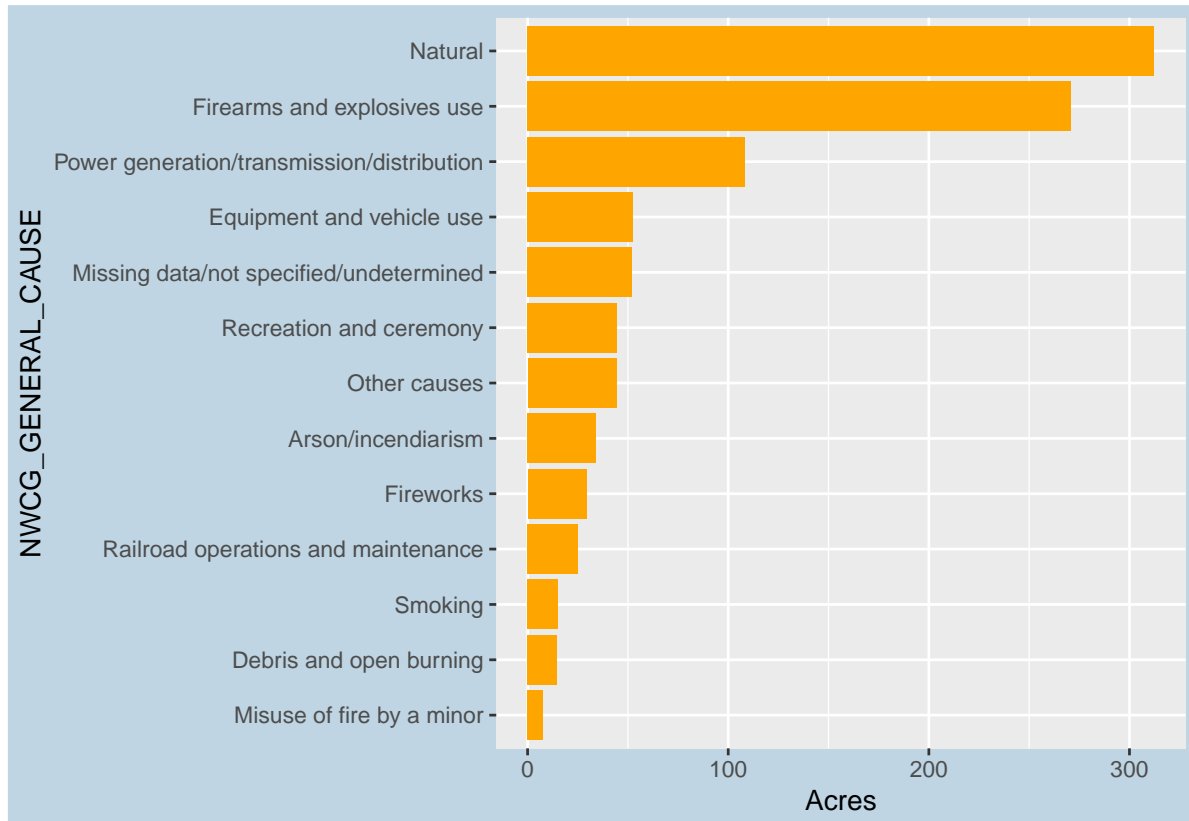


Figure 5: average wildfire size by cause.

```
fires_6 <- as.data.frame(fires)
fires_6 <- fires_6 %>%
  group_by(STATE) %>%
  summarize(total = n()) %>%
  na.omit()
fires_6 <- as.data.frame(fires_6)
colnames(fires_6)[1] = "state"

plot_usmap(data = fires_6, values = "total", color = "red", labels = TRUE) +
  scale_fill_continuous(low = "white", high = "darkred",
                        name = "Number of fires", label = scales::comma) +
  theme(legend.position = "right",
        legend.title = element_text(size=16),
```

```

legend.text = element_text(size=18),
plot.title = element_text(size=24),
plot.caption = element_text(size=20),
panel.background = element_rect(colour = "black"))

```

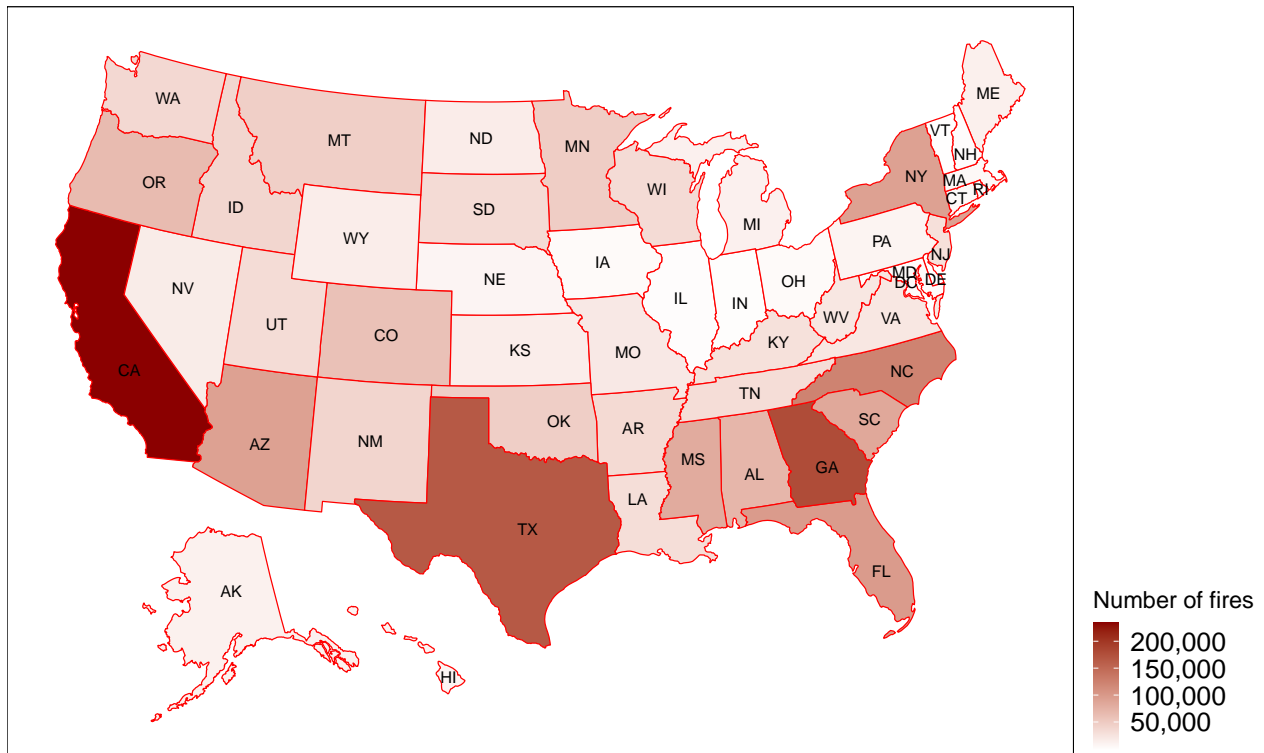


Figure 6: US Wildfires, 1992-2018. The spectrum from white to darkred indicates more wildfires in that state

```

fires_7 <- as.data.frame(fires)
fires_7 <- fires_7 %>%
  filter(STATE == 'CA') %>%
  group_by(FIPS_CODE) %>%
  summarize(total = n()) %>%
  na.omit()
fires_7 <- as.data.frame(fires_7)
colnames(fires_7)[1] = "fips"
#fires_7 <- fires_7[c("LONGITUDE", "LATITUDE", "total")]
#fire_transformed <- usmap::usmap_transform(fires_7)

plot_usmap(data = fires_7, values = "total", "counties", include = c("CA"), labels = TRUE,
# geom_point(data = fires_7, aes(x = LONGITUDE.1, y = LATITUDE.1, size = total), color = "darkred"),
  scale_fill_continuous(low = "white", high = "darkred", name = "Number of fires", label = FALSE),
  theme(legend.position = "right",

```

```
legend.title = element_text(size=16),  
legend.text = element_text(size=10),  
plot.title = element_text(size=24),  
plot.caption = element_text(size=20),  
panel.background = element_rect(colour = "black"))
```

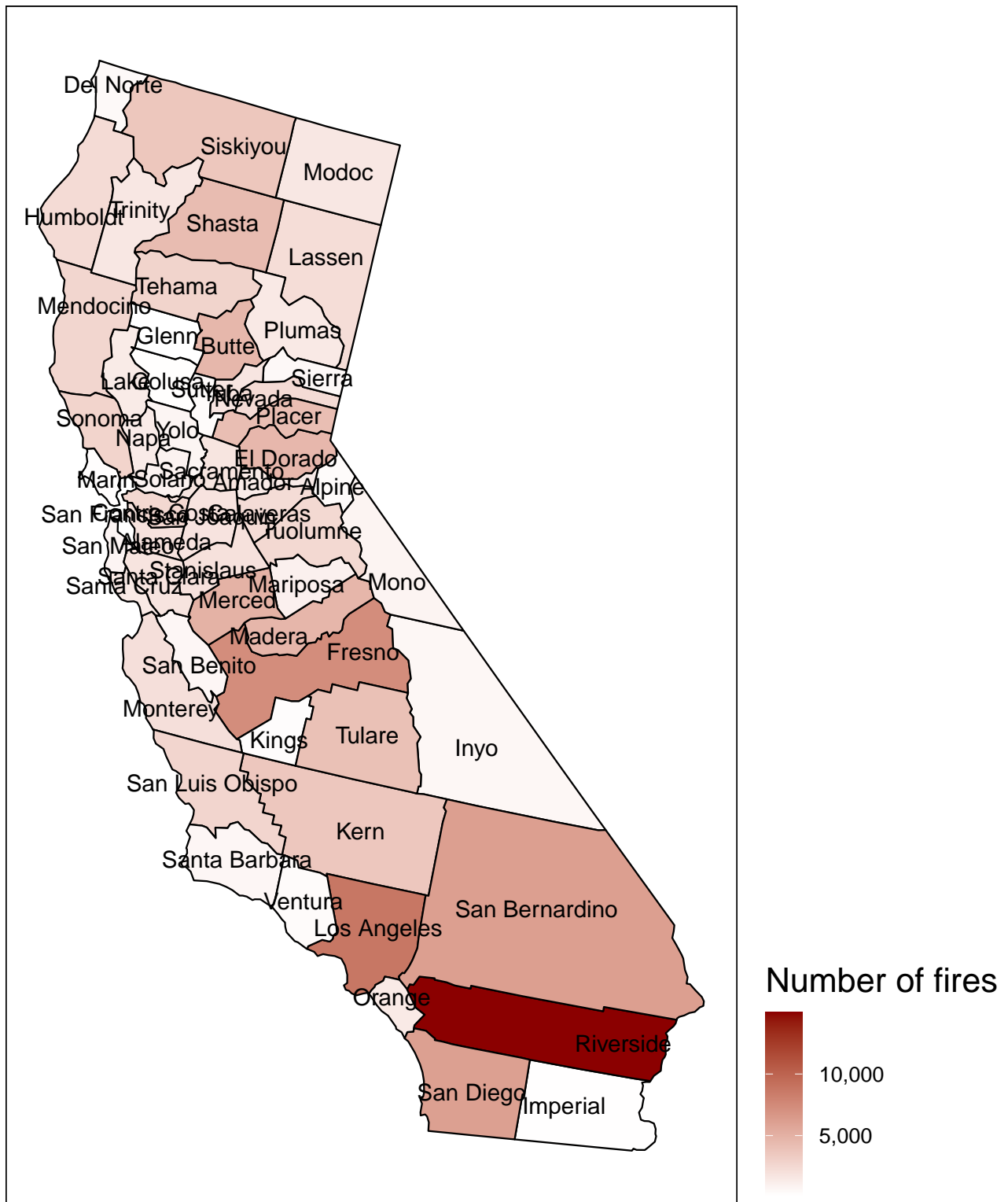


Figure 7: US Wildfires in CA, 1992-2018. The spectrum from white to darkred indicates more wildfires in that state

3 Exploration

3.1 Title 1

3.2 Title 2

3.3 Title 3

4 Problems Tackled and Conclusion

5 References