Wildfires in California

Tali Motzkin (206079253)

Idan Megira (318995537)

Introduction

- Powerful and large-scale fires pose a danger to accelerating global warming (Oris et al., 2014) and to actual damage to property, human life and biological systems.

- In recent years there has been a danger of an increase in fires worldwide because of climate change factors (Kasischke et al., 1995). Due to anthropogenic activity, greenhouse gases in the earth's atmosphere have been rising steadily over the past 30-40 years and are expected to double within the next century, which is reflected in a warming of the atmosphere of 1-4°C (Stocks, 1993).

Introduction

- This type of warming increases the chance of stronger and more frequent fires, especially in hot and dry areas like the state of California which is characterized by mediterranean climate (Williames et al., 2021).

- This work we will examine: the fire's properties in California during years 1983-2021. While also examining the factors that affect the extent of fires, and their spatial and temporal distribution.

Literature Review

- Wildfires are one of the most frequent natural disasters in California. Thus, climate change and land cover changes, not only prolong the fire season, but also increase the severity of the damage and scale of the burned area (Li & Banerjee, 2021).

- According to Williams et al. (2019), California's annual wildfire extent increased fivefold between the years 1972-2018. The increasing in Summertime fires is mainly caused by cumulative drying effects of atmospheric aridity and precipitation deficits. While the increasing in Falltime fires is mainaly caused by strong dry wind.

Literature Review

- Keely & syphard (2017), shows that the correlation between burned area and climate is relatively weak in Central and south coast. This is likely partly because fire-climate relationships in these regions are strongly manipulated by humans via ignitions, suppression, and land cover change (Balch et al., 2017).

- In other study, Keeley and Syphard (2018) analysis shows that the frequency of wildfires in California declined greatly after 1980, but there has been no significant change in the total annual burned area. They also investigated the causes of fire, for example they showed that in recent decades man-made ignition sources have become less frequent, except of ignition made from power lines.

Literature Review

- Furthermore, Miller et al. (2012) showed that fire size, duration, number of fires, and total area burned per year were all significantly higher for lightning-ignited fires than for human-ignited fires.

- In Conclusion, Not all articles examined the entire territory of California. For example, Keely & Syphard (2017) - mainly examined South and Central California. Also, articles have shown different trends therefore we wanted to investigate by our own.

Hypothesis

- We assume that the effect of climatic factors on the extent of fires, in terms of burned area, will be weak. Because fires are affected by many factors, that are not necessarily climatic such as human factors. thus, they are very hard to measure.

- Also, we assume that we will see an overall increasing in number of fires and burned area during the examined years, especially in forested areas.

Data sources and R Libraries

- **Meteorological data** was downloaded from CIMIS - The California Irrigation Management Information System between the years 1980-2020.
- **CIMIS Station's location layer** - Was downloaded from ArcGIS online.
- **Fires Layer** - Was downloaded from ArcGIS online. It contains fire perimeters between the years 1878-2020.
- **Cause of fire** - the table was extracted from the Fire Layer Metadata.
- **Counties of California** - the shapefile was downloaded from "California Geo-Portal".
- **R Libraries**:

```r
library(sf)
library(tidyr)
library(dplyr)
library(ggplot2)
library(viridis)
library(RColorBrewer)
library(Hmisc)
library(ggcorrplot)
library(Hmisc)
library(moments)
```

Methodology

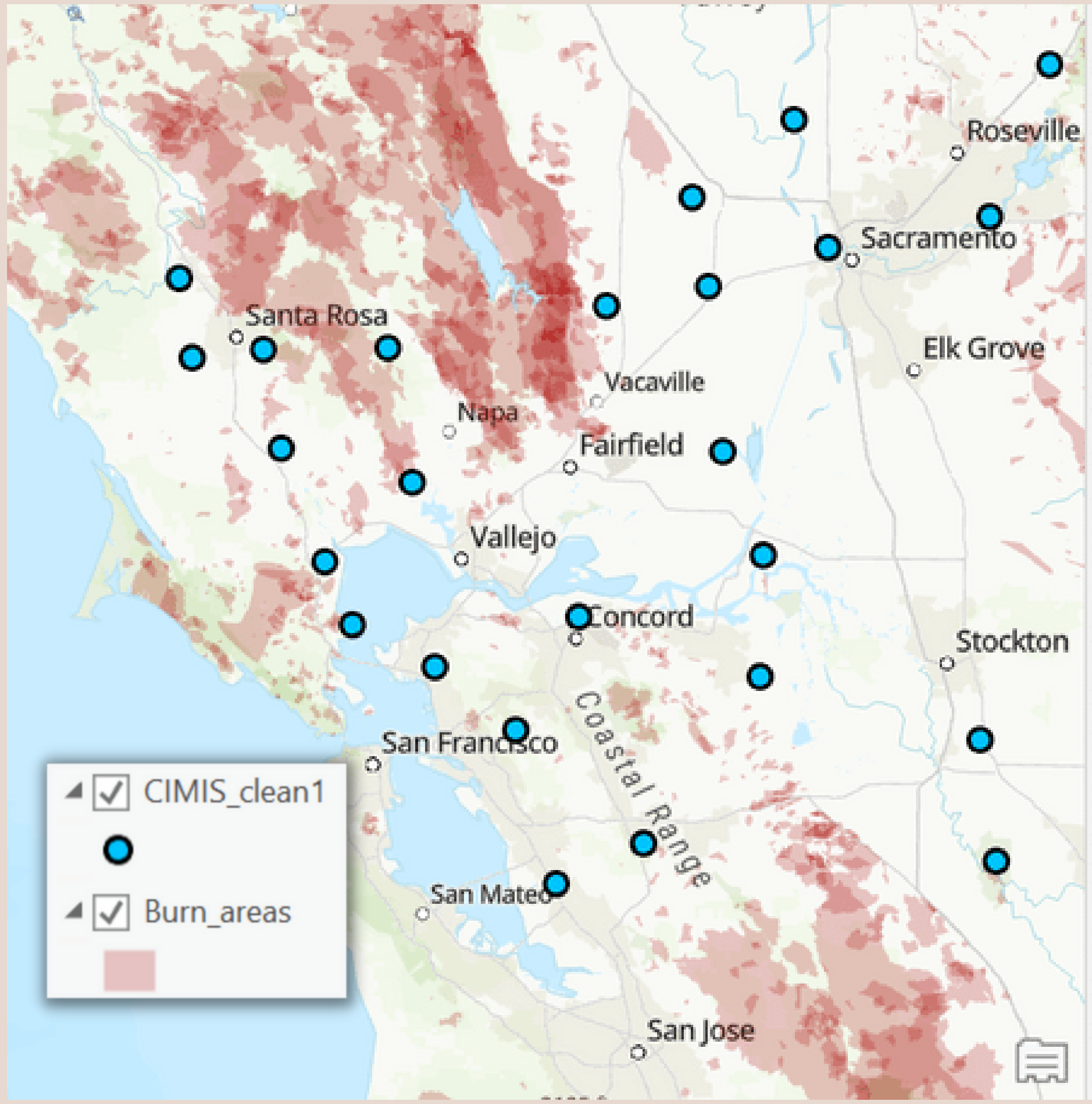
Data Loading

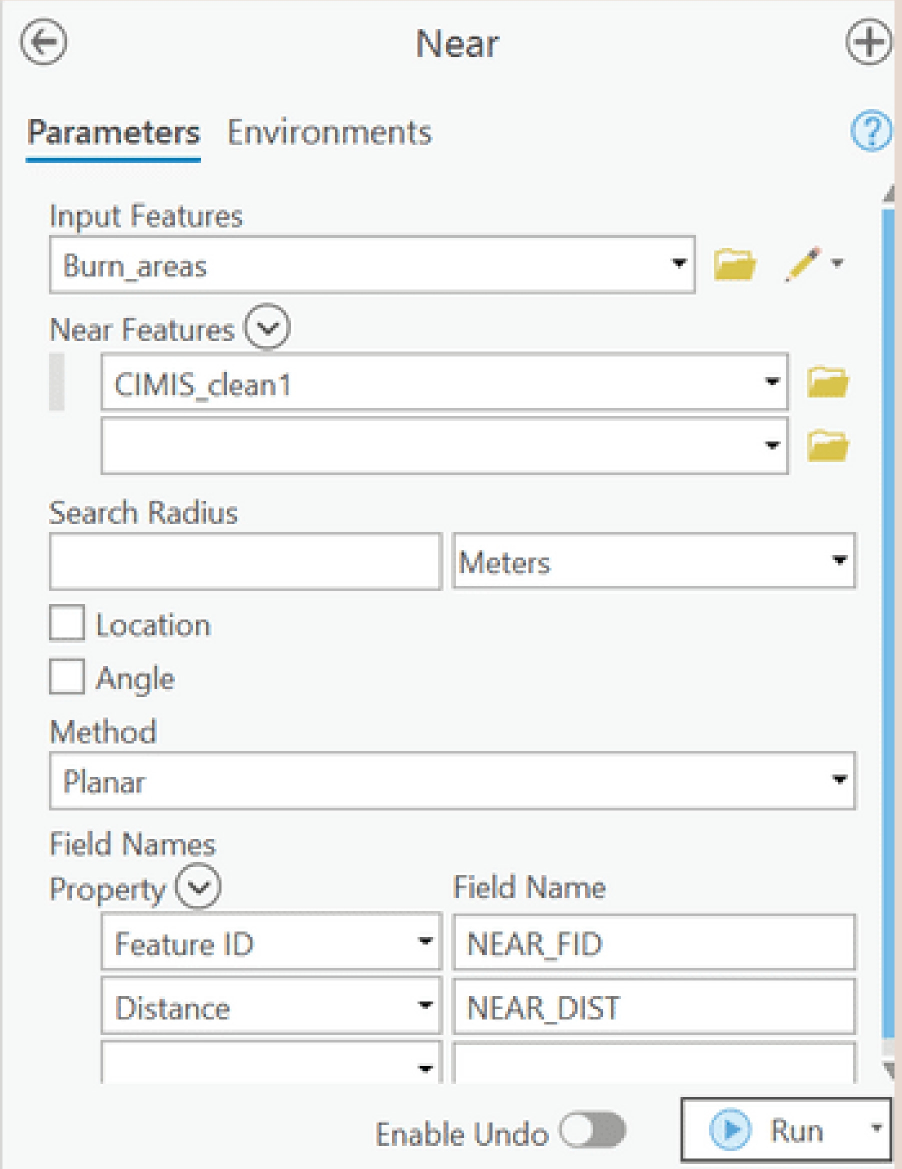
```
####Load Weather and Cause of fire tables, and CIMIS station data from the gdb project
Weather_data=read.csv("Raw_tables\\Weather_data.csv")
Cause_code = read.csv("Raw_tables\\Cause_code.csv")
gdb_project = st_read("Raw_Layers\\Project.gdb")
project_layers = st_layers(dsn = "Raw_Layers\\Project.gdb")
CIMIS_New = st_read("Raw_Layers\\Project.gdb", layer="CIMIS_new")
```

```
#### Comparing between the spatial station data (CIMIS_new) and the weather station
#### data (Weather_data) and making a new layer which contains only the spatial
#### stations that have weather data
CIMIS_clean1 = CIMIS_new[CIMIS_new$ID %in%  Weather_data$Stn.Id,]
st_write(CIMIS_clean1, "CIMIS_clean1.shp")
```

Methodology

Using the Near tool, we created a new field contains the nearest meteorological station for each fire






```
#### reading the gdb project of the two layers: burn areas and the clean meteorological
#### spatial station. Each burn area has the nearest station (determined by the FID column)
Burn_areas = st_read("Raw_Layers\\Project.gdb", layer="Burn_areas")
CIMIS_clean1 = st_read("Raw_Layers\\Project.gdb", layer = "CIMIS_clean")
CIMIS_clean1$FID = 0:(length(CIMIS_clean1$ID)-1)
#We couldn't load the FID column from the ARCGIS so we made one manually
```

Methodology

Fires Data: Cleaning and Tidying

```
#### cleaning and tiding the Data of the three tables
Burn_areas_clean = Burn_areas %>%
  select(AGENCY:FIRE_NAME, ALARM_DATE:CAUSE,
         GIS_ACRES, Shape_Length:Shape_Area,NEAR_FID:Shape) %>%
  mutate(treat_duration = difftime(CONT_DATE, ALARM_DATE, units="days")) %>%
  separate(ALARM_DATE, into=c("year", "month","day_time"), sep="-",
           convert = TRUE) %>%
  separate(day_time, into =c("day", "time"), sep =2,convert= TRUE) %>%
  separate(CONT_DATE, into=c("Cont_year", "Cont_month","Cont_day_time"), sep="-",
           convert = TRUE) %>% #seperate for the cont_date
  separate(Cont_day_time, into =c("Cont_day", "Cont_time"), sep =2,convert= TRUE) %>%
  select(-time, -Cont_time)
```

```
#1. we selected the relevant columns
#2. we calculated a new column which contains the treatment
#duration of the fire contamination(days)
#3. we separated the starting and the ending dates of the fires to
# three different columns: "day", "month" and "year" and de-select unnecessary columns
```

Methodology

Weather Data: Cleaning and Tidying

```
Weather_data_clean = Weather_data %>%
  select(Stn.Id:Total.ETo..mm., Total.Precip..mm.,Avg.Sol.Rad..W.sq.m.
         ,Avg.Vap.Pres..kPa.,Avg.Max.Air.Temp..C.,Avg.Min.Air.Temp..C.
         , Avg.Air.Temp..C., Avg.Max.Rel.Hum....,Avg.Min.Rel.Hum....,Avg.Rel.Hum....
         ,Avg.Dew.Point..C.,Avg.Wind.Speed..m.s.,Avg.Soil.Temp..C.) %>%
  rename(Total_ETo_mm = Total.ETo..mm., Tot_Precip_mm = Total.Precip..mm.,
         Avg_Sol_Rad_Wsqm = Avg.Sol.Rad..W.sq.m., Avg_Vap_pres_kPa=Avg.Vap.Pres..kPa.,
         Avg_Max_Air_T_C=Avg.Max.Air.Temp..C., Avg_Min_Air_Temp_C= Avg.Min.Air.Temp..C.,
        Avg_Air_T_C = Avg.Air.Temp..C., Avg_Max_Rel_Hum= Avg.Max.Rel.Hum....,
       Avg_Min_Rel_Hum = Avg.Min.Rel.Hum...., Avg_Rel_Hum= Avg.Rel.Hum....,
       Avg_DPoint_c=Avg.Dew.Point..C., Avg_Wind_S_ms=Avg.Wind.Speed..m.s.,
       Avg_Soil_Temp_C=Avg.Soil.Temp..C.) %>%
  separate(Month.Year , into=c("month", "year_time"), sep="/",
           convert = TRUE) %>%
  separate(year_time, into =c("year", "time"), sep =4,convert= TRUE) %>%
  select(-time)
#1. we selected the relevant columns
#2. we renamed some of the columns and separated the measuring date
# to 2 columns: "month" and "year" and de-select unnecessary columns
```

Methodology

Merging the data frame

```
#### join
CIMIS_data_clean = CIMIS_clean1 %>%
  st_drop_geometry() %>%
  select(ID, FID)
#1. we drop Geometry for the join
#2. we selected the relevant columns

Burn_Weather_join = Burn_areas_clean %>%
  inner_join(CIMIS_data_clean, by = c("NEAR_FID"="FID")) %>%
  inner_join(Weather_data_clean, by = c("ID"="Stn.Id", "month", "year")) %>%
  left_join(Cause_code ,by = c("CAUSE"= "Code")) %>%
  select(-NEAR_FID)
#we joined the 4 tables into one main table that includes the fires
#data, meteorological station data and the cause of the fire
```

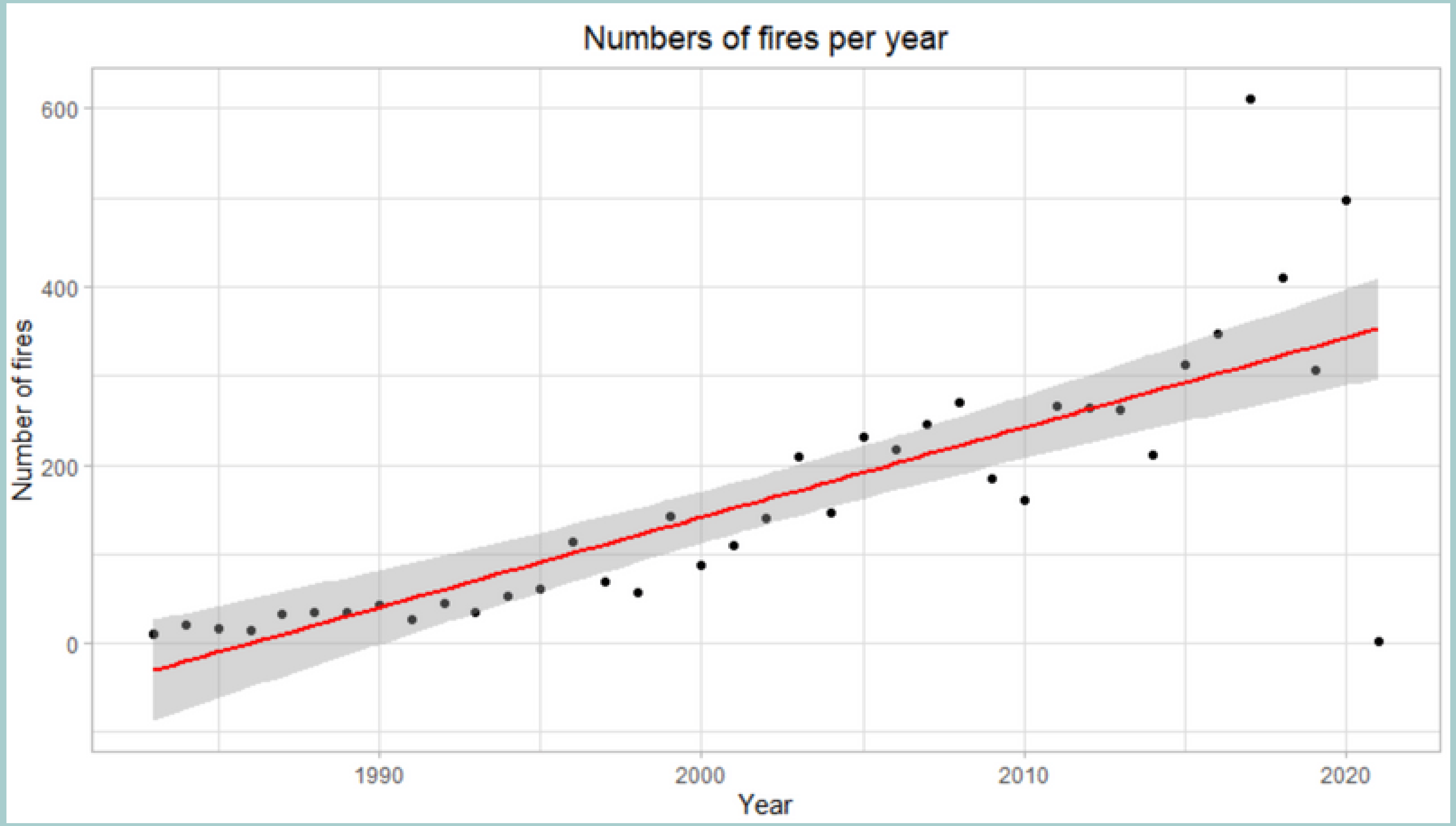
```
converted the sf data frame to tibble format for convenience while working with big
le_Burn_weather = as.tibble(Burn_Weather_join)
```

We decided to use inner join because climatic data and meteorological station data was only available for 1980-2020

General Trends

Number of fires - by year

- There has been an upward trend in the number of fires over the examined years.

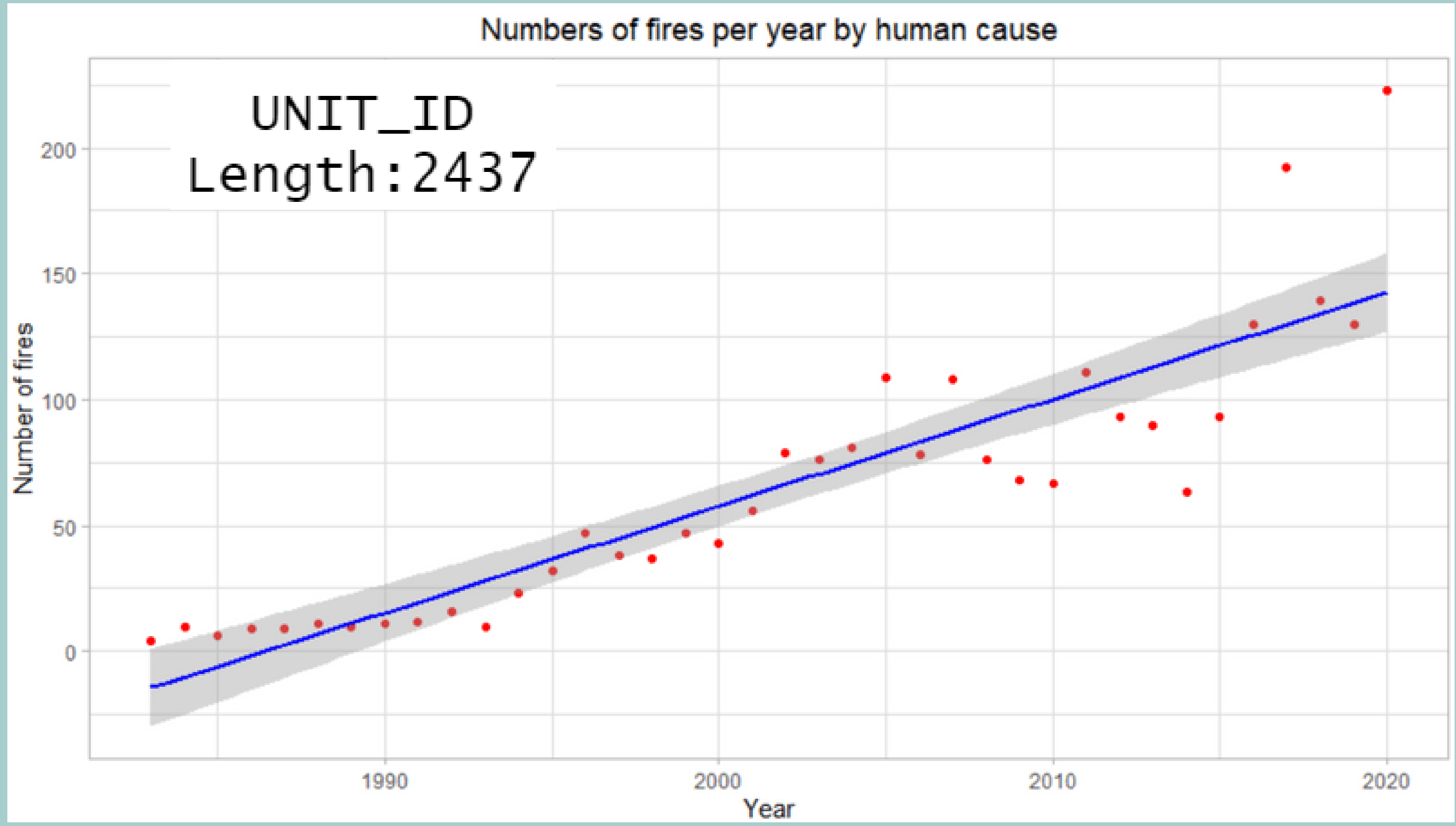


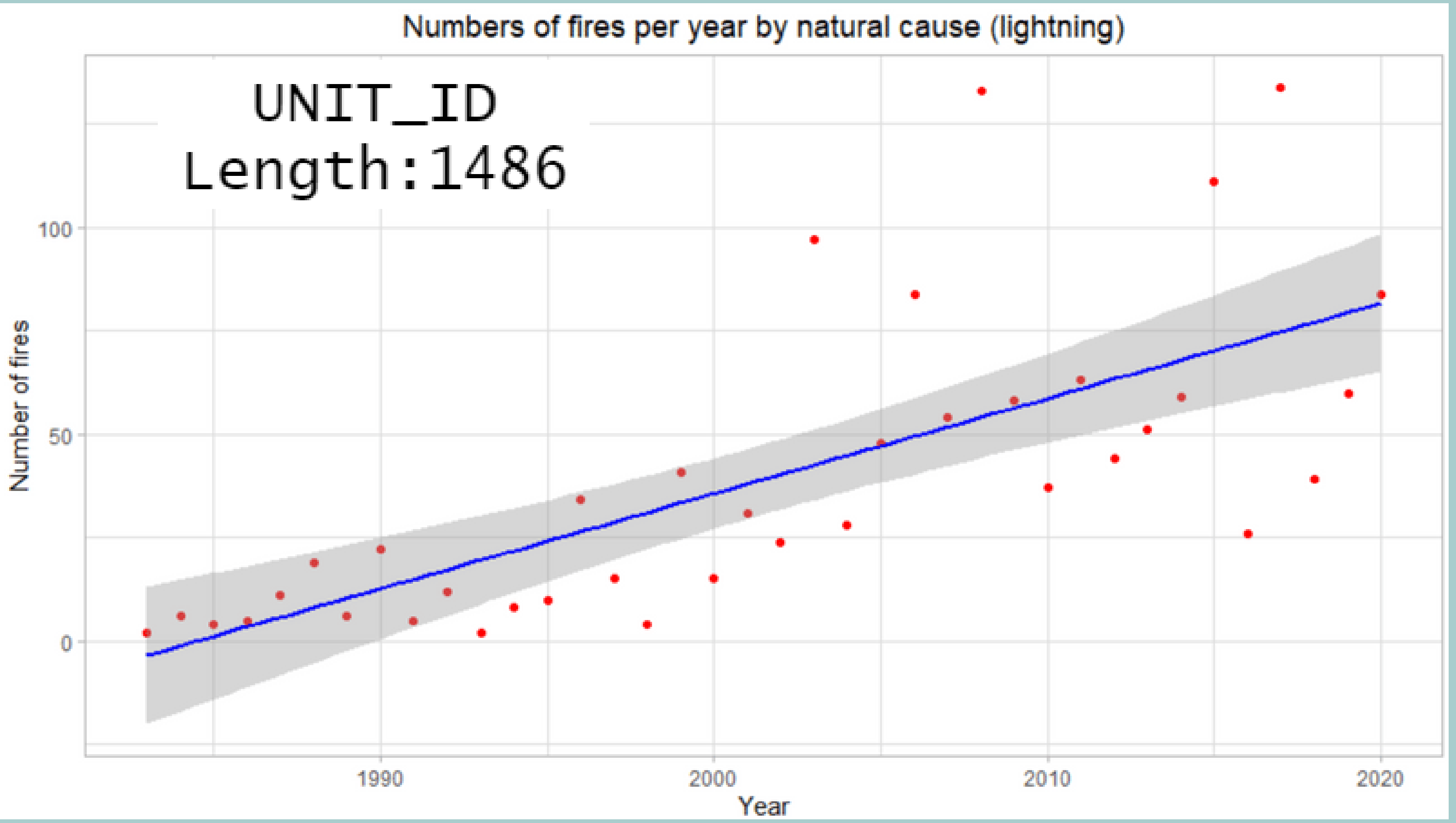

```
#show how much fires were per year since 1983 till 2020
tibble_Burn_weather %>%
  group_by(year)%>%
  count() %>%
  ggplot(aes(year,n))+
  geom_point(col = "black")+
  labs(title = "Numbers of fires per year",
       x="Year", y="Number of fires")+
  geom_smooth(method = lm, col="red")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

General Trends

Causes - Human vs Natural

- As can be seen, there is an increase in both man-made and natural-caused fires.





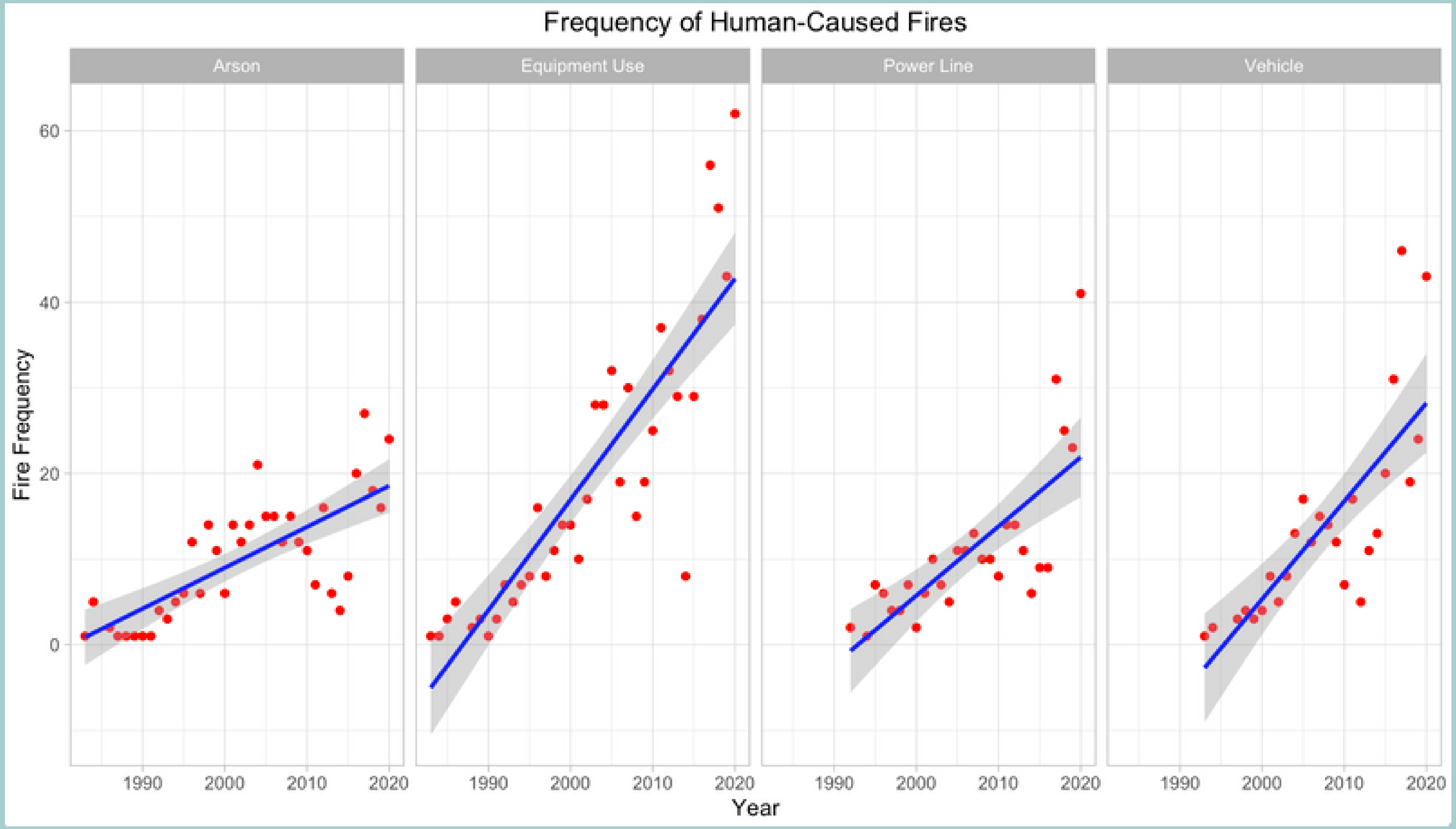

```
#Numbers of fires per year by human cause
tibble_Burn_weather %>%
  filter(Desc %in% c("Aircraft","Arson","Campfire", "Debris",
                     "Equipment Use", "Escaped Prescribed Burn",
                     "Firefighter Trainning","Illigal Alien Campfire",
                     "Non-Firefighter Training", "Playing with Fire",
                     "Power Line", "Railroad", "Smoking",
                     "Structure", "Vehicle")) %>%
  group_by(year)%>%
  count() %>%
  ggplot(aes(year,n))+
  geom_point(col = "red")+
  labs(title = "Numbers of fires per year by human cause",
       x="Year", y="Number of fires")+
  geom_smooth(method = lm, col="blue")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

```
#plot of lightning (natural cause)
tibble_Burn_weather %>%
group_by(year,Desc)%>%
  count() %>%
  filter(Desc == "Lightning") %>%
  ggplot(aes(year,n))+
  geom_point(col = "red")+
  labs(title = "Numbers of fires per year by natural
       cause (lightning)",
       x="Year", y="Number of fires")+
  geom_smooth(method = lm, col="blue")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

General Trends

Human Causes - Further trends

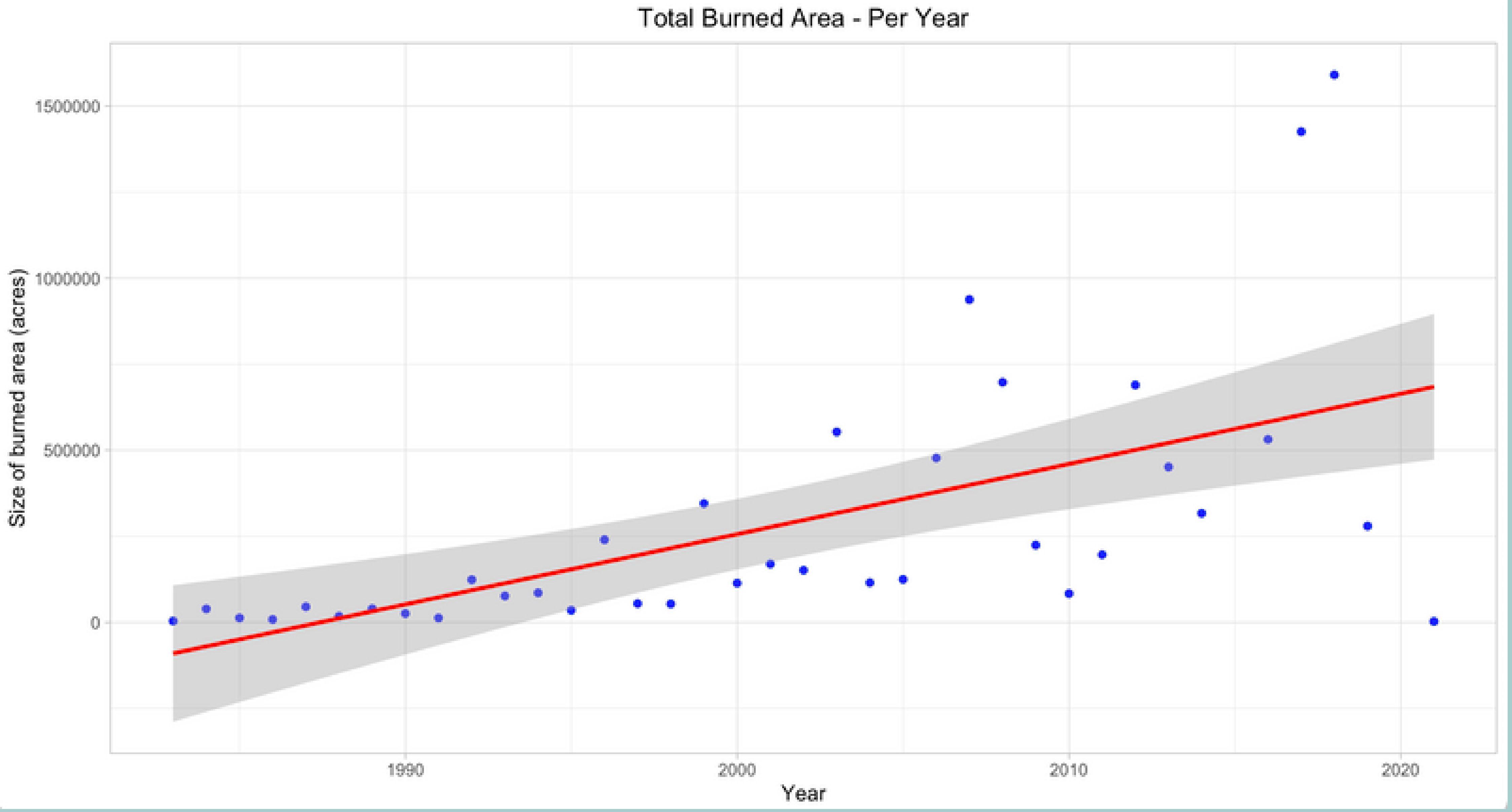
- A more in-depth examination of man-made fire, showed that there was mainly an increase in fires caused by Equipment use




```
##ignition from power lines, vehicle, equipment use and arson
tibble_Burn_weather %>%
  group_by(year,Desc)%>%
  count() %>%
  filter(Desc %in% c("Power Line","Arson","Vehicle", "Equipment Use")) %>%
  ggplot(aes(year,n))+
  geom_point(col = "red")+
  labs(title = "Numbers of fires per year by 4 human ignition causes",
       x="Year", y="Number of fires")+
  geom_smooth(method = lm, col="blue")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))+
  facet_wrap(~Desc, nrow=1)
```

Total Burned area (Per Year)

- There is an increase in the amount of areas burned each year



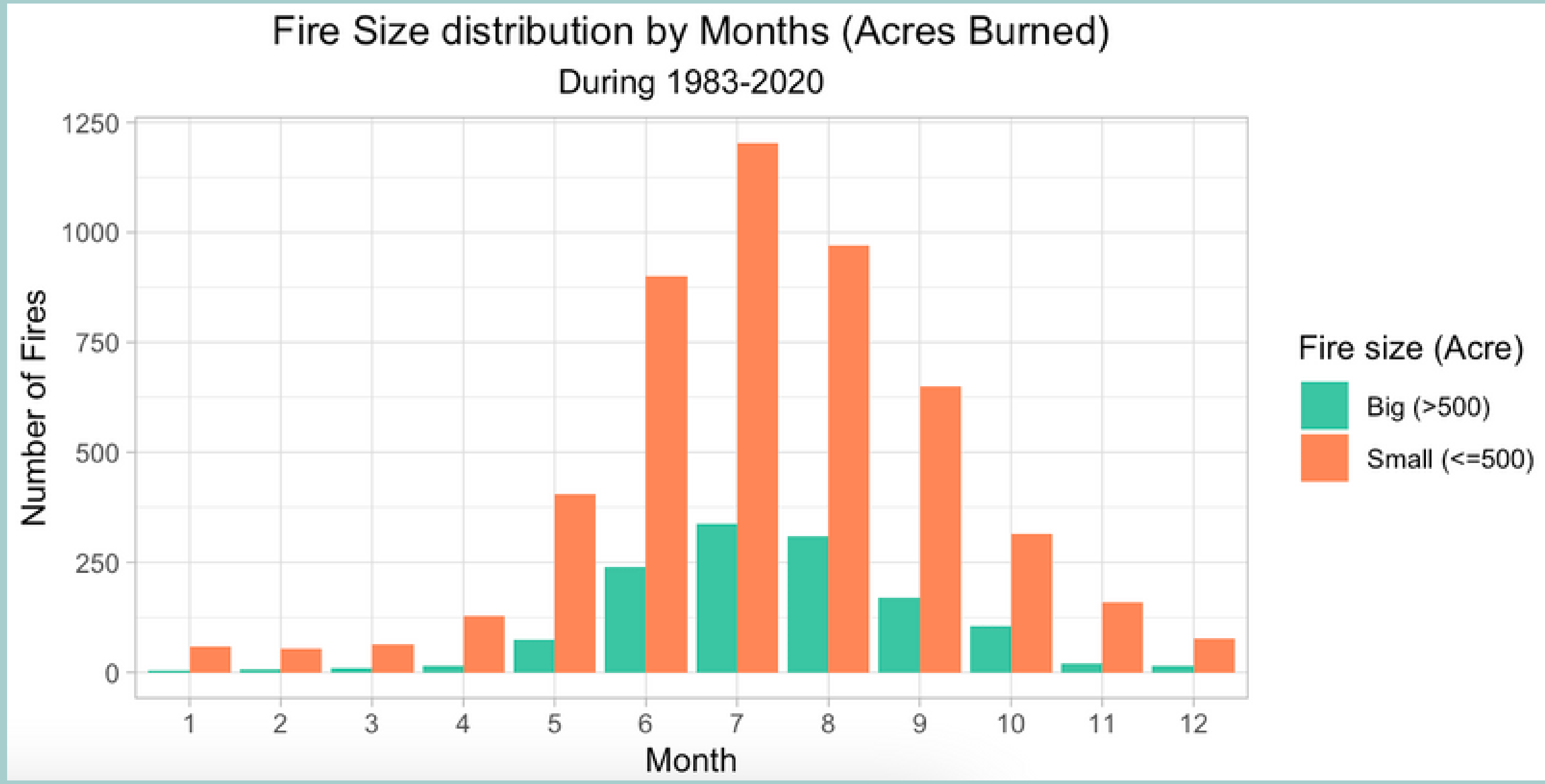

```
#sum the total burned area per year
tibble_Burn_weather %>%
  group_by(year)%>%
  summarise(total = sum(GIS_ACRES)) %>%
  ggplot(aes(year,total))+
  geom_point(col = "blue")+
  labs(title = "Total Burned Area - Per Year",
       x="Year", y="Size of burned area (acres)")+
  geom_smooth(method = lm, col="red")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

Fire Size - Big VS Small

```r
#decide which fires are small\big
Burn_size = tibble_Burn_weather %>%
  filter(!is.na(GIS_ACRES)) %>%
  mutate(godel = case_when(GIS_ACRES <= 500 ~ "Small (<=500)",
                           GIS_ACRES > 500 ~ "Big (>500)"))
```

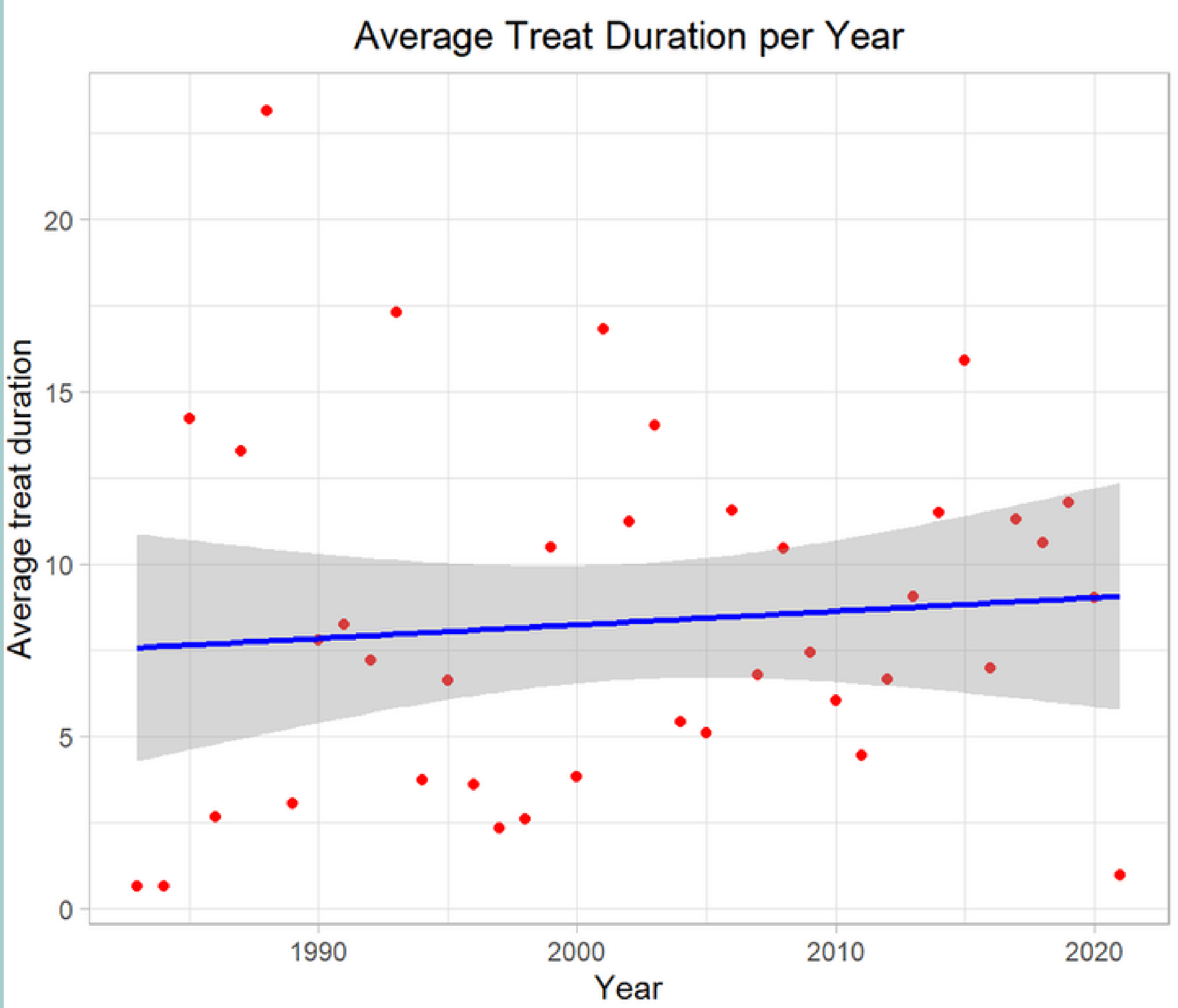
```r
#make a plot of big and small fires by month as accumulation
#of all years
Burn_size %>%
  group_by(month,godel)%>%
  count() %>%
  ggplot(aes(x= as.factor(month)))+
  geom_col(aes(fill = as.factor(godel), y=n), position = "dodge")+
  scale_fill_brewer(palette="Set2")+
  labs(title = "Fire Size distribution by Months (Acres Burned)",
       subtitle = "During 1983-2020",
       y="Number of Fires", x="Month",fill="Fire size (Acre)")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5),
        plot.subtitle = element_text(hjust=0.5))
```

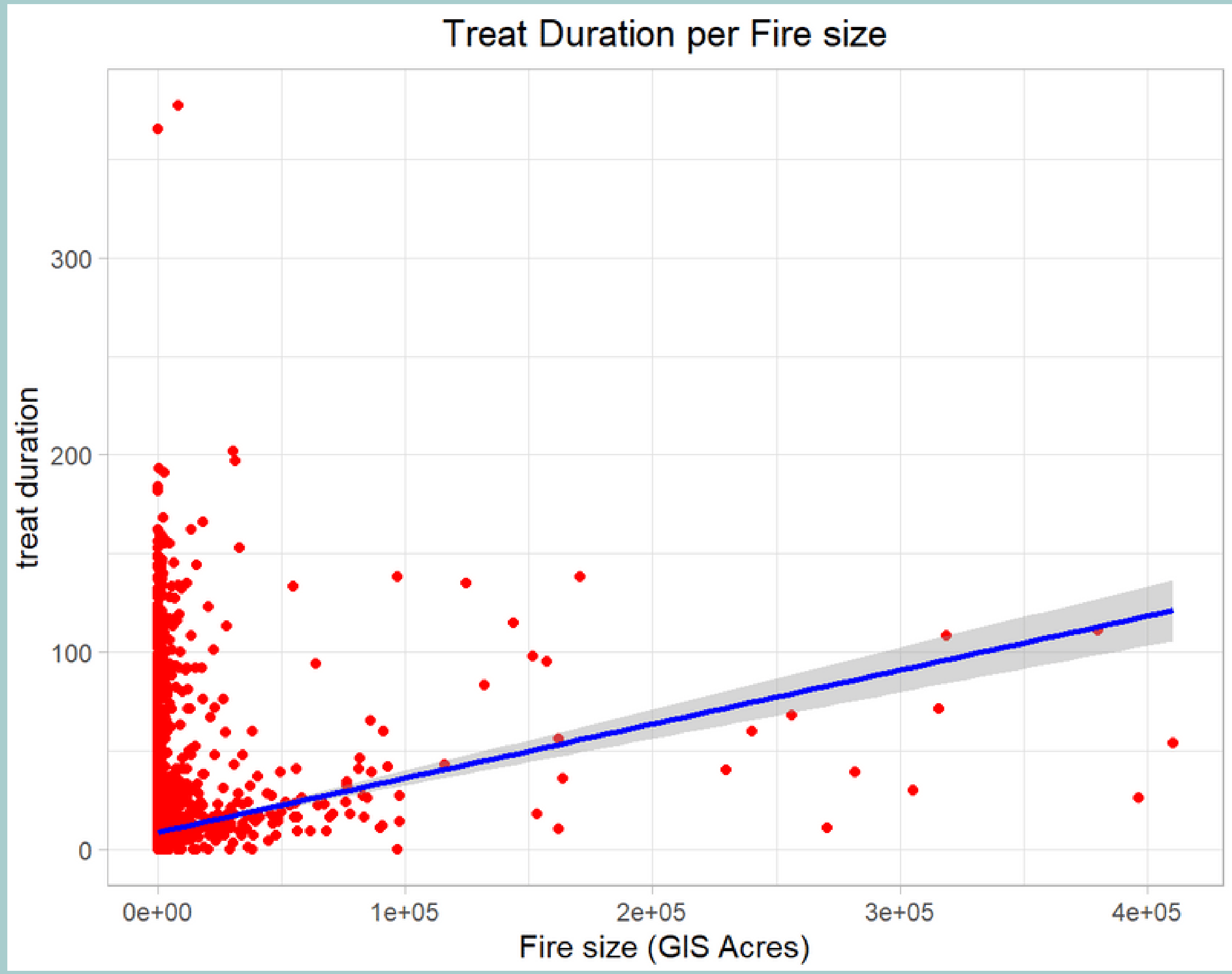
- Most major fires happen in the summer months.
- Size Classification was done according to Li & Banerjee (2021)



Treat Duration

- In addition we tried to examine the issue of the fire treatment duration.
- But we found no clear trend.






```
#Treat duration of fires per year
tibble_Burn_weather %>%
  filter(treat_duration >= 0) %>%
  group_by(year)%>%
  summarise(Avg_Treat_Duration =mean(treat_duration )) %>%
  ggplot(aes(year,Avg_Treat_Duration))+
  geom_point(col = "red")+
  labs(title = "Average Treat Duration per Year",
       x="Year", y="Average treat duration")+
  geom_smooth(method = lm, col="blue")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

```
#Treat duration of fires per Size of fire
tibble_Burn_weather %>%
  filter(treat_duration >= 0, GIS_ACRES < 500000) %>%
  ggplot(aes(GIS_ACRES,treat_duration))+
  geom_point(col = "red")+
  labs(title = "Treat Duration per Fire size",
       x="Fire size (GIS Acres)", y="treat duration")+
  geom_smooth(method = lm, col="blue")+
  theme_light()+
  theme(plot.title = element_text(hjust=0.5))
```

Checking if the Data is normal

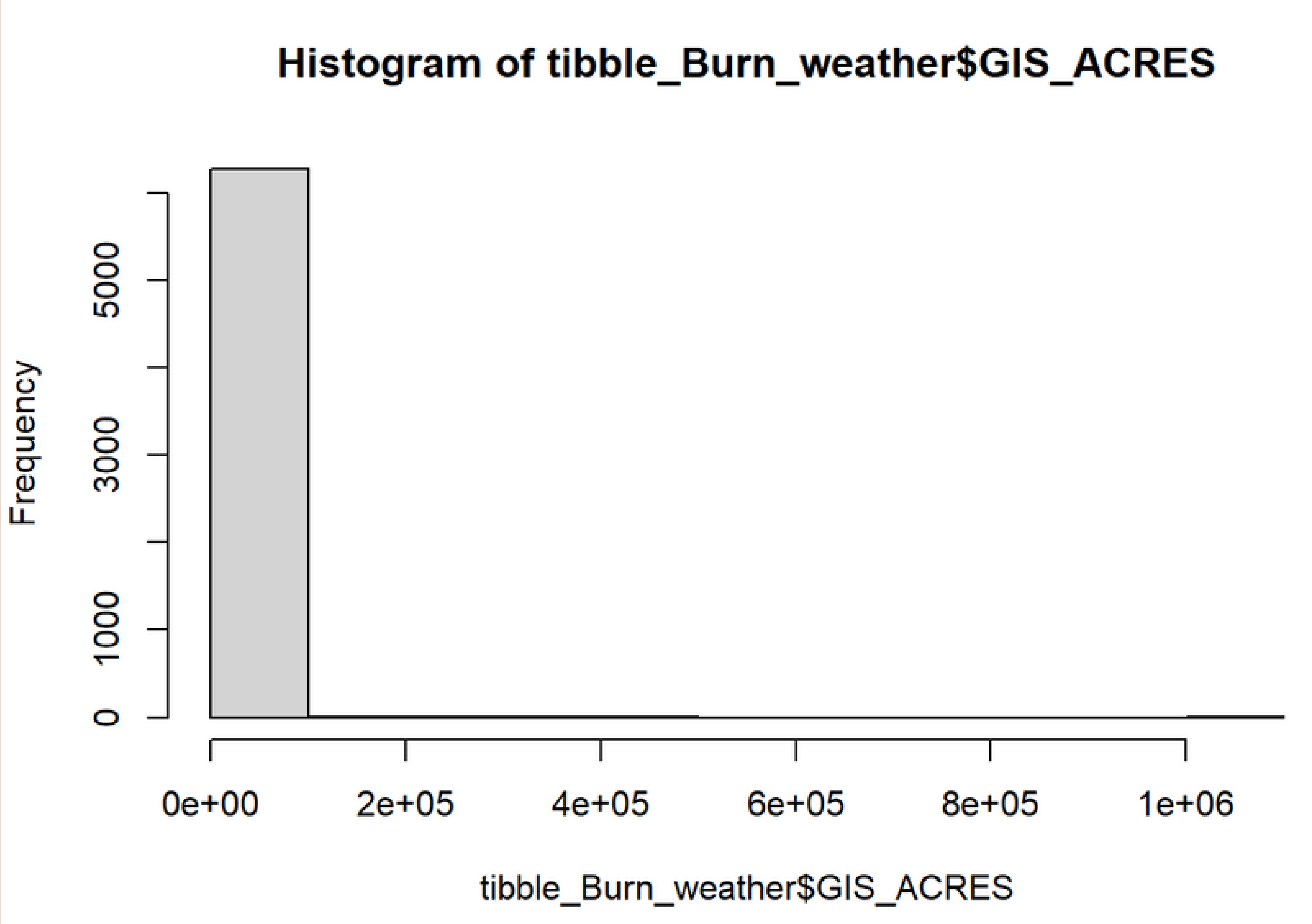
- We couldn't pass the shapiro test due to limitation for 5000 observations.
- We checked the distribution using histogram plot, skewness and kurtosis.

```
> shapiro.test(tibble_Burn_weather$Avg_Air_T_C)
Error in shapiro.test(tibble_Burn_weather$Avg_Air_T_C) :
  sample size must be between 3 and 5000
```

```
hist(tibble_Burn_weather$GIS_ACRES)
skewness(tibble_Burn_weather$GIS_ACRES, na.rm = TRUE)
kurtosis(tibble_Burn_weather$GIS_ACRES, na.rm = TRUE)
```

```
[1] 26.99809
```

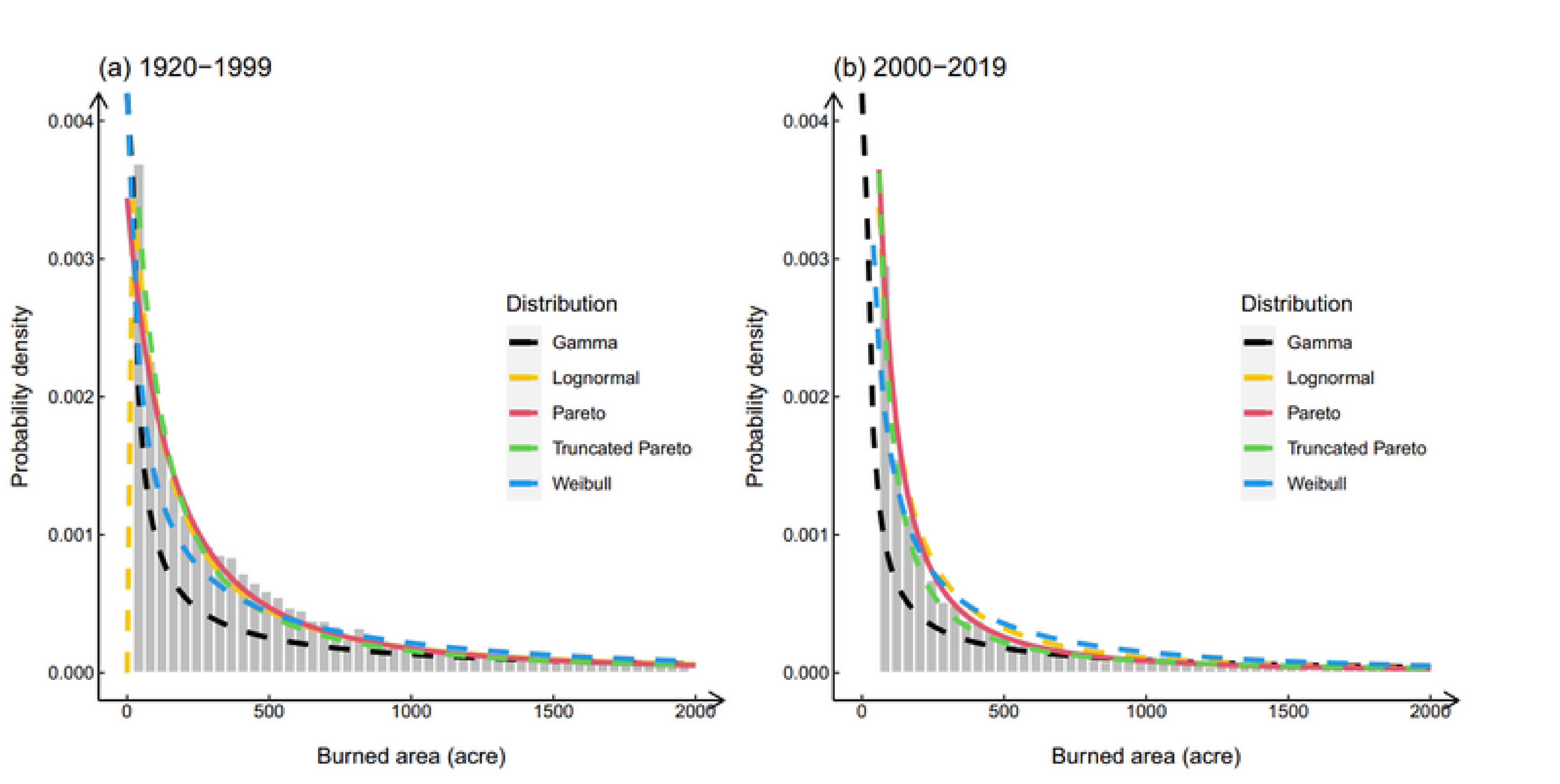
```
[1] 1089.892
```



- Kurtosis very high - means that the frequency is high and the data has a "peak" and the Skewness is also high meaning there is a right "tail".

Checking if the Data is normal

- We saw that our data does not have a normal distribution but we have enough observation to pass the linear regression tests.
- From the litruture we also saw that the Burned area is probably a "Pareto" distribution.



(Lee & Banerjee, 2021)

Multiple Linear Regression of Climatic factors (Overall)

- Checking The relation between the Total burned area and the Climatic Factors, by using a Multiple Linear Regression.
- The R^2 of the model is very low (0.004866), relation is almost non-existent
- P-value shows high significance level, meaning there is a <u>significantly not</u> linear relation between the Total burned area and: <u>Solar Radiation</u>, <u>Air Temp</u> and <u>Humidity</u>

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -2.574e+03  2.539e+03  -1.014  0.31065    
Tot_Precip_mm  -6.546e-02  5.794e-01  -0.113  0.91004    
Total_ETo_mm   -4.525e+00  9.518e+00  -0.475  0.63449    
Avg_Sol_Rad_Wsqm -1.976e+01  7.015e+00  -2.817  0.00486 ** 
Avg_Air_T_C     3.784e+02  7.989e+01   4.736 2.23e-06 ***
Avg_Rel_Hum     5.952e+01  2.436e+01   2.443  0.01458 *  
Avg_Wind_S_ms   1.562e+02  5.447e+02   0.287  0.77434    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20570 on 6041 degrees of freedom
  (250 observations deleted due to missingness)
Multiple R-squared:  0.004866,	Adjusted R-squared:  0.003877 
F-statistic: 4.923 on 6 and 6041 DF,  p-value: 4.944e-05
```

```
#linear regression of precipitation, evapotranspiration, temperature,
#wind speed, humidity, solar
climate_model = lm(data = tibble_Burn_weather, GIS_ACRES ~ Tot_Precip_mm+
                Total_ETo_mm + Avg_Sol_Rad_Wsqm + Avg_Air_T_C +
                Avg_Rel_Hum + Avg_Wind_S_ms)
summary(climate_model)
climate_model
```

Linear Regression of Climatic factors (Overall)

```
> Air_Temp=lm(data=tibble_Burn_weather, GIS_ACRES ~ Avg_Air_T_C)
> summary(Air_Temp)

Call:
lm(formula = GIS_ACRES ~ Avg_Air_T_C, data = tibble_Burn_weather)

Residuals:
    Min      1Q  Median      3Q     Max 
  -4352   -2676   -2193   -1616 1029676 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   -251.2     1120.2  -0.224   0.8226  
Avg_Air_T_C    129.4       52.4   2.470   0.0136 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20700 on 6276 degrees of freedom
  (20 observations deleted due to missingness)
Multiple R-squared:  0.0009708, Adjusted R-squared:  0.0008116 
F-statistic: 6.099 on 1 and 6276 DF,  p-value: 0.01355
```

```
> Sol_Rad=lm(data=tibble_Burn_weather, GIS_ACRES ~ Avg_Sol_Rad_Wsqm)
> summary(Sol_Rad)

Call:
lm(formula = GIS_ACRES ~ Avg_Sol_Rad_Wsqm, data = tibble_Burn_weather)

Residuals:
    Min      1Q  Median      3Q     Max 
  -4872   -2505   -2096   -1757 1030092 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      4840.747   1183.375   4.091 4.36e-05 ***
Avg_Sol_Rad_Wsqm   -8.725      4.190  -2.082   0.0374 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20700 on 6279 degrees of freedom
  (17 observations deleted due to missingness)
Multiple R-squared:  0.0006899, Adjusted R-squared:  0.0005307 
F-statistic: 4.335 on 1 and 6279 DF,  p-value: 0.03738
```

```
> Rel_Hum=lm(data=tibble_Burn_weather, GIS_ACRES ~Avg_Rel_Hum)
> summary(Rel_Hum)

Call:
lm(formula = GIS_ACRES ~ Avg_Rel_Hum, data = tibble_Burn_weather)

Residuals:
    Min      1Q  Median      3Q     Max 
  -3260   -2492   -2267   -1966 1030118 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1517.85    1013.26   1.498    0.134
Avg_Rel_Hum    18.33      19.45   0.943    0.346

Residual standard error: 20710 on 6271 degrees of freedom
  (25 observations deleted due to missingness)
Multiple R-squared:  0.0001417, Adjusted R-squared:  -1.774e-05 
F-statistic: 0.8887 on 1 and 6271 DF,  p-value: 0.3459
```

- Further checking the relation between the Total Burned Area and the 3 significant factors.
- Unsurprisingly, the R ^ 2 is still weak, the relation is almost non-existent.

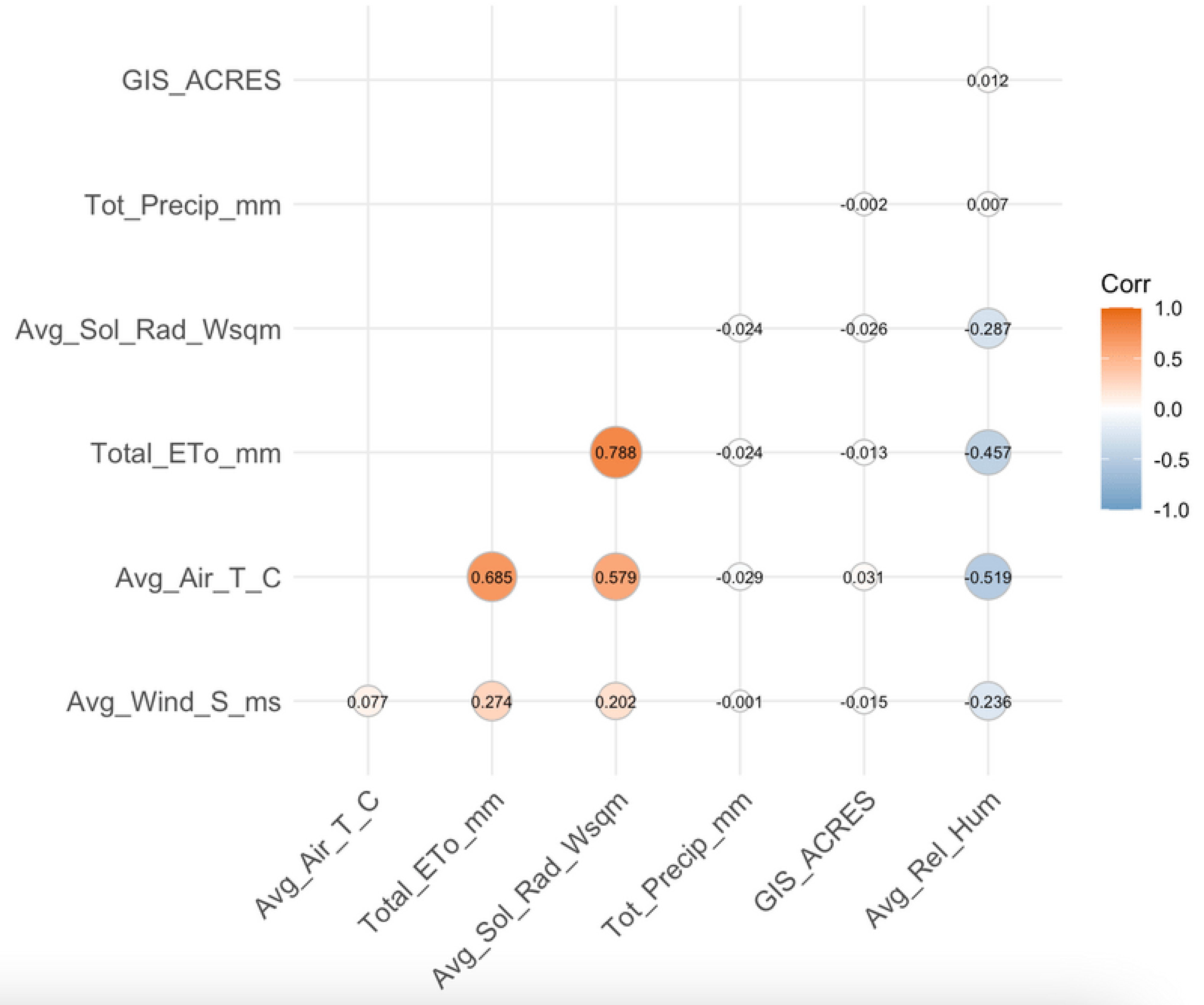
Pearson Correlation Test

Examining the effect of a specific climate parameter on the total burned areas, by using pearson test.

```
##checking the strong correlations:
cor.test(tibble_Burn_weather$GIS_ACRES, tibble_Burn_weather$Avg_Air_T_C,
    method = "pearson", use="na.or.complete")
cor.test(tibble_Burn_weather$GIS_ACRES, tibble_Burn_weather$Avg_Sol_Rad_Wsqm,
    method = "pearson", use="na.or.complete")
cor.test(tibble_Burn_weather$GIS_ACRES, tibble_Burn_weather$Avg_Rel_Hum,
    method = "pearson", use="na.or.complete")
```

```
[1] 0.031158
[1] -0.02626579
[1] 0.01190393
```

Correlation matrix -Climatic Factors




```
hello=tibble_Burn_weather %>%
  select(GIS_ACRES, Tot_Precip_mm, Total_ETo_mm,
         Avg_Sol_Rad_Wsqm, Avg_Air_T_C, Avg_Rel_Hum, Avg_Wind_S_ms)

cor1=rcorr(as.matrix(hello[,c(1:7)]))
cor1_df=as.data.frame(cor1$r)
cor1_df

ggcorrplot(cor1_df, hc.order = TRUE,
           type="lower",
           lab=TRUE,digits = 3,
           lab_size=2.5,
           method="circle",
           colors = c("blue", "white", "red"))
ggtheme=theme_gray()
dev.off()
```

Linear Regression of Climatic factors (Summertime)

```
#filter summer months and run regression
summer = filter(tibble_Burn_weather, month %in% c(5:10))
summer_model = lm(data = summer, GIS_ACRES ~ Tot_Precip_mm+
                    Total_ETo_mm + Avg_Sol_Rad_Wsqm + Avg_Air_T_C +
                    Avg_Rel_Hum + Avg_Wind_S_ms)
summary(summer_model)
```

```
Call:
lm(formula = GIS_ACRES ~ Tot_Precip_mm + Total_ETo_mm + Avg_Sol_Rad_Wsqm +
    Avg_Air_T_C + Avg_Rel_Hum + Avg_Wind_S_ms, data = summer)

Residuals:
    Min      1Q  Median      3Q     Max
  -9090   -3222   -2046    -834 1027548

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.987e+03  3.250e+03  -0.611  0.54099
Tot_Precip_mm    -7.207e-02  5.964e-01  -0.121  0.90383
Total_ETo_mm     -5.165e+00  9.897e+00  -0.522  0.60174
Avg_Sol_Rad_Wsqm -2.307e+01  7.931e+00  -2.909  0.00364 **
Avg_Air_T_C       3.722e+02  9.209e+01   4.042 5.37e-05 ***
Avg_Rel_Hum       6.896e+01  2.699e+01   2.555  0.01063 *
Avg_Wind_S_ms     2.700e+02  6.328e+02   0.427  0.66957
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21160 on 5450 degrees of freedom
  (226 observations deleted due to missingness)
Multiple R-squared:  0.005153,	Adjusted R-squared:  0.004058
F-statistic: 4.705 on 6 and 5450 DF,  p-value: 8.735e-05
```

```
# Summer - pearson test for the 3 climatic factors
cor(summer$GIS_ACRES, summer$Avg_Sol_Rad_Wsqm, method = "pearson", use="na.or.complete")
cor(summer$GIS_ACRES, summer$Avg_Air_T_C, method = "pearson", use="na.or.complete")
cor(summer$GIS_ACRES, summer$Avg_Rel_Hum, method = "pearson", use="na.or.complete")
```

- We tried to run the same test, but only for fires the occurred in summer months (May-October).
- The R^2 of the model is still very low, the relation is almost non-existent
- Next, we performed a Pearson test for the three significant factors. The results:

		r
GIS_Acres	Average Solar Radiation	-0.04430782
GIS_Acres	Average Air Temperature	0.02534028
GIS_Acres	Average Relative humidity	0.01893242

Linear Regression of Climatic factors (Falltime)

```
#filter fall month and run regression
fall = filter(tibble_Burn_weather, month %in% c(11,12,1,2))
fall_model = lm(data = fall, GIS_ACRES ~ Tot_Precip_mm+
                    Total_ETo_mm + Avg_Sol_Rad_Wsqm + Avg_Air_T_C +
                    Avg_Rel_Hum + Avg_Wind_S_ms)
summary(fall_model)
```

```
Call:
lm(formula = GIS_ACRES ~ Tot_Precip_mm + Total_ETo_mm + Avg_Sol_Rad_Wsqm +
    Avg_Air_T_C + Avg_Rel_Hum + Avg_Wind_S_ms, data = fall)

Residuals:
   Min     1Q Median     3Q    Max
-12909  -3062  -1306    228 274480

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       9068.245   9922.862   0.914   0.3614
Tot_Precip_mm       35.345     24.283   1.456   0.1463
Total_ETo_mm       365.719    167.362   2.185   0.0295 *
Avg_Sol_Rad_Wsqm  -149.800     64.586  -2.319   0.0209 *
Avg_Air_T_C       -406.151    397.272  -1.022   0.3073
Avg_Rel_Hum         -2.694    100.709  -0.027   0.9787
Avg_Wind_S_ms    -4192.693   2661.122  -1.576   0.1160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17140 on 376 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.02441,   Adjusted R-squared:  0.00884
F-statistic: 1.568 on 6 and 376 DF,  p-value: 0.1553
```

```
# Fall - pearson test for the 2 significant factors
cor(fall$GIS_ACRES, fall$Avg_Air_T_C, method = "pearson", use="na.or.complete")
cor(fall$GIS_ACRES, fall$Avg_Wind_S_ms, method = "pearson", use="na.or.complete")
```

- Running the same test, but now for fires occured in fall months (November-February)
- The R^2 of model is still low, but it is higher than the R^2 of the previous models
- Next, we performed a Pearson test for the 2 significant factors. The results:

		r
GIS_Acres	Total_ETo_	0.07132572
GIS_Acres	Average Solar Radiation	-0.01525809

Chi Square Test for Independence (All-times)

- We checked the relationships between the variable of burned Acres to all of the climate conditions.
- We also checked the relationship between the variable of burned Acres to the ignition cause. We aggregated all of the human ignition causes into one sign "0" in the new column "binary" and the lightning cause was marked as "1".

```
hum_nat = tibble_Burn_weather %>%
  filter(Desc %in% c("Aircraft","Arson","Campfire", "Debris",
                     "Equipment Use", "Escaped Prescribed Burn",
                     "Firefighter Trainning","Illigal Alien Campfire",
                     "Non-Firefighter Training", "Playing with Fire",
                     "Power Line", "Railroad", "Smoking",
                     "Structure", "Vehicle", "Lightning")) %>%
  mutate(binary = case_when(Desc == "Lightning"~ 1, TRUE ~ 0))
```

```
chisq.test(hum_nat$GIS_ACRES, hum_nat$binary)
```

```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$binary
X-squared = 3920.9, df = 3919, p-value = 0.4885
```

- For the ignition causes there was no relationship between them to the size of fire because the P value is bigger than 0.05

Chi Square Test for Independence (All-times)

```
chisq.test(hum_nat$GIS_ACRES, hum_nat$Avg_Air_T_C)
chisq.test(hum_nat$GIS_ACRES, hum_nat$Total_ETo_mm)
chisq.test(hum_nat$GIS_ACRES, hum_nat$Tot_Precip_mm)
chisq.test(hum_nat$GIS_ACRES, hum_nat$Avg_Wind_S_ms)
chisq.test(hum_nat$GIS_ACRES, hum_nat$Avg_Rel_Hum)
```

```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$Avg_Air_T_C
X-squared = 1105158, df = 1104030, p-value = 0.2238
```

```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$Total_ETo_mm
X-squared = 7911131, df = 7904834, p-value = 0.05666
```

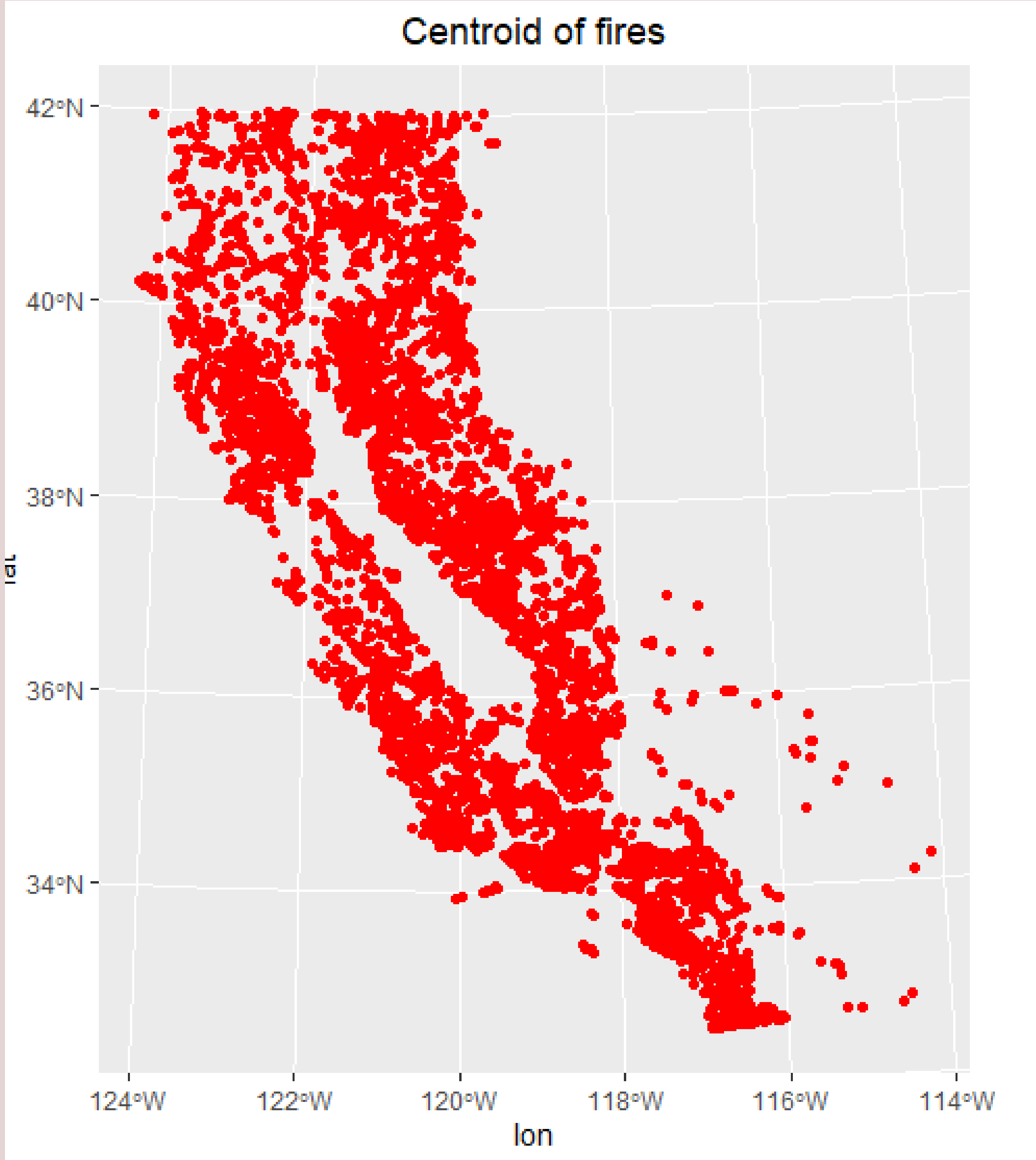
```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$Tot_Precip_mm
X-squared = 1740036, df = 1738260, p-value = 0.1704
```

```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$Avg_Wind_S_ms
X-squared = 172524, df = 172348, p-value = 0.3818
```

```
        Pearson's Chi-squared test

data:  hum_nat$GIS_ACRES and hum_nat$Avg_Rel_Hum
X-squared = 297768, df = 297464, p-value = 0.3465
```

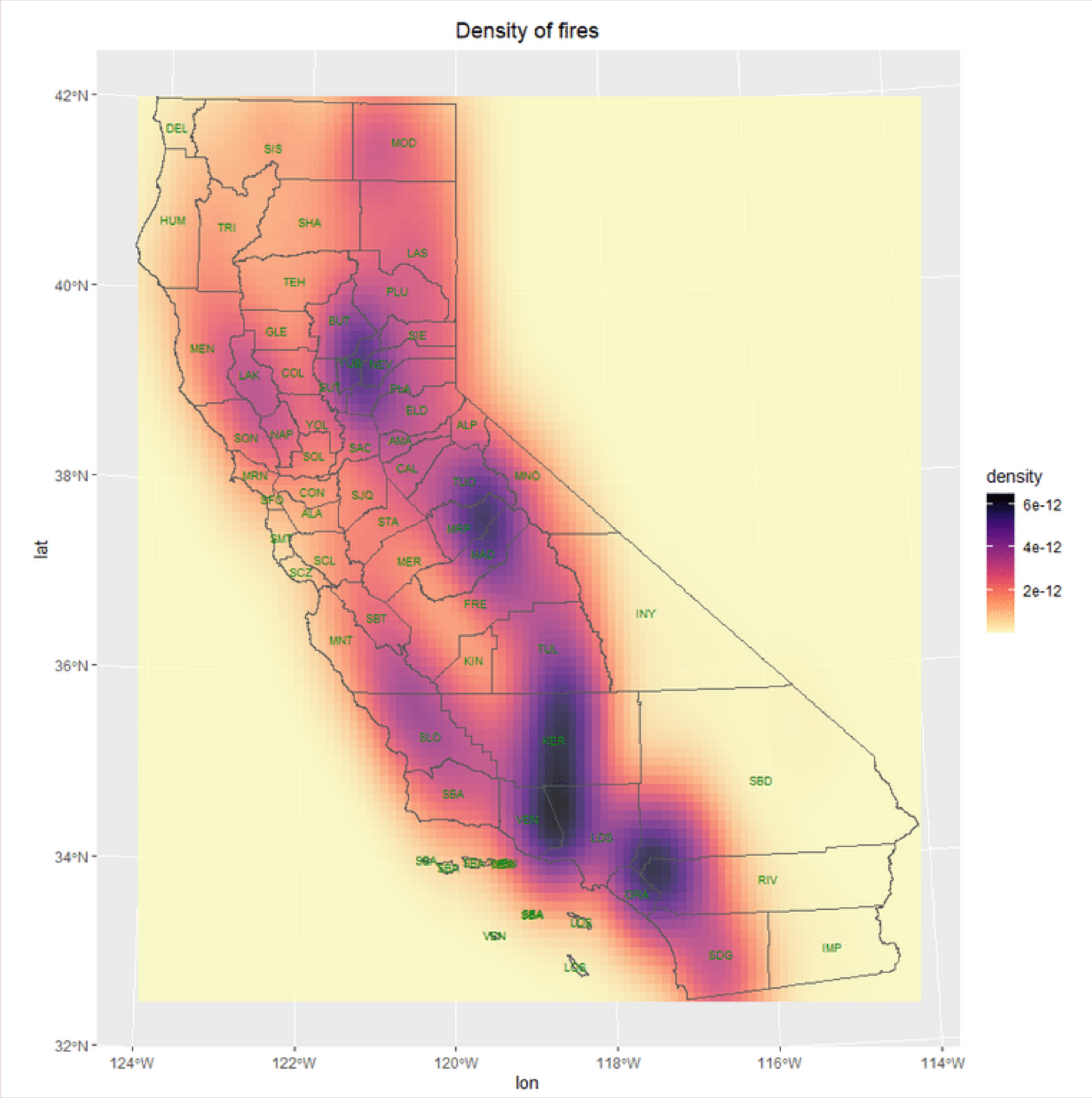
- For the climate factors there was no relationship between them to the size of fire because all of the P values were bigger than 0.05, Except the total evapotranspiration factor which it's P value is 0.056 so there is a little relation between this factor to the size of the burned area.

Fires Distribution




```r
#finding the fires centroid and plotting the results
Burn_centroid = st_centroid(Burn_Weather_join)
ggplot()+
  geom_sf(data =Burn_centroid, color = 'red')+
  labs(title = "Centroid of fires", x="lon", y="lat")+
  theme(plot.title = element_text(hjust=0.5))
```

Fires Distribution




```
#finding the lat and lon attributes of the centroid for the ggplot plotting
sep_coords <- Burn_centroid %>%
  mutate(lon = sf::st_coordinates(.)[,1],
              lat = sf::st_coordinates(.)[,2])

#plotting the density of fires
ggplot() +
  geom_sf(data =sep_coords, color = 'white', alpha = 0.2)+
  stat_density_2d(data = sep_coords,
                  mapping = aes(x = lon,
                                y = lat,
                                fill = stat(density)),
                  geom = 'tile',
                  contour = FALSE,
                  alpha = 0.8) +
  scale_fill_viridis(option = "A", direction = -1) +
  geom_sf(data = cal_crs, fill = NA) +
  geom_sf_text(data = cal_crs, aes(label= COUNTY_ABB), colour = 'green4', size = 2.5)+
  labs(title = "Density of fires")+
  theme(plot.title = element_text(hjust=0.5))
```

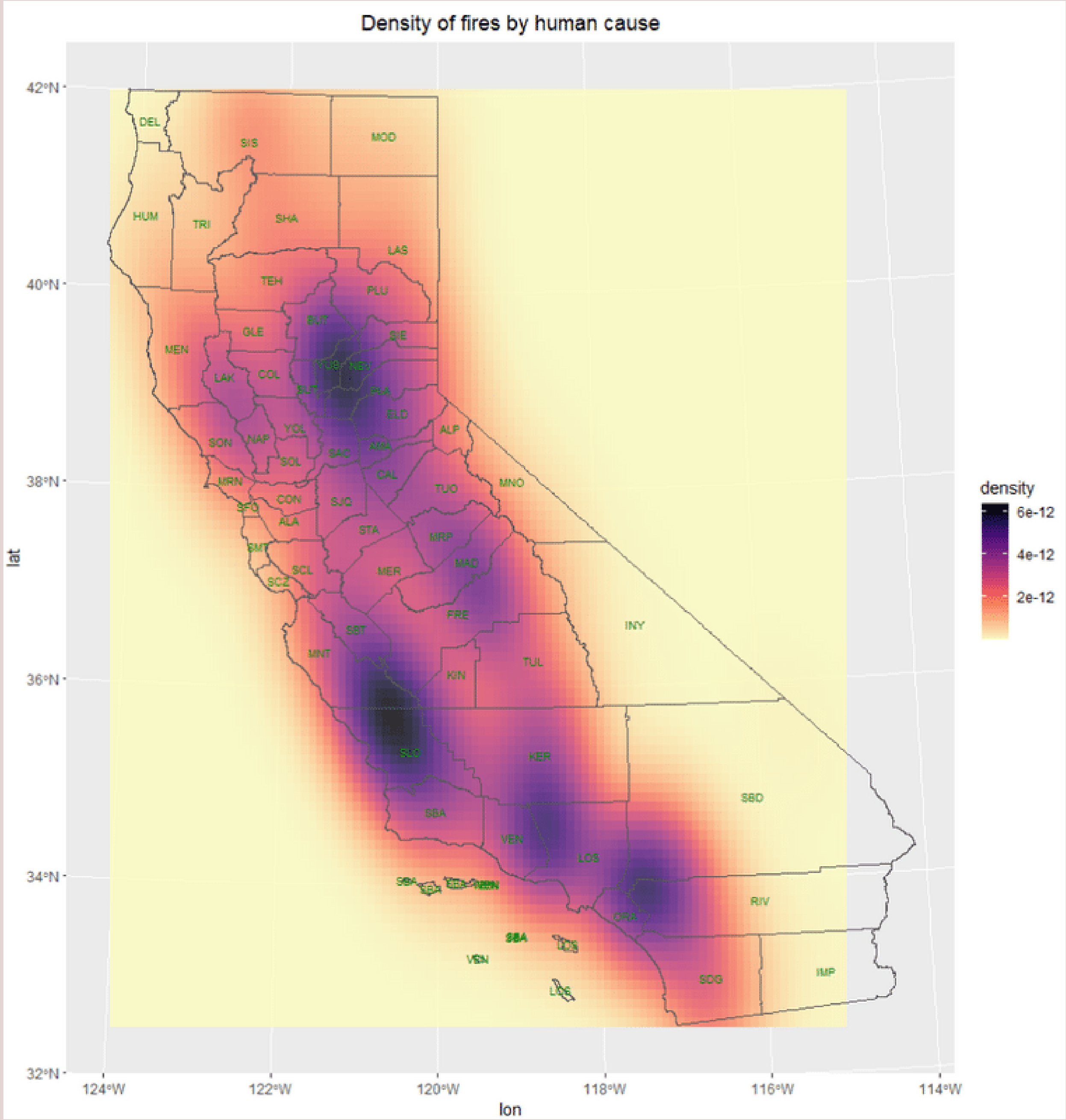
- High fire density especially in following counties:

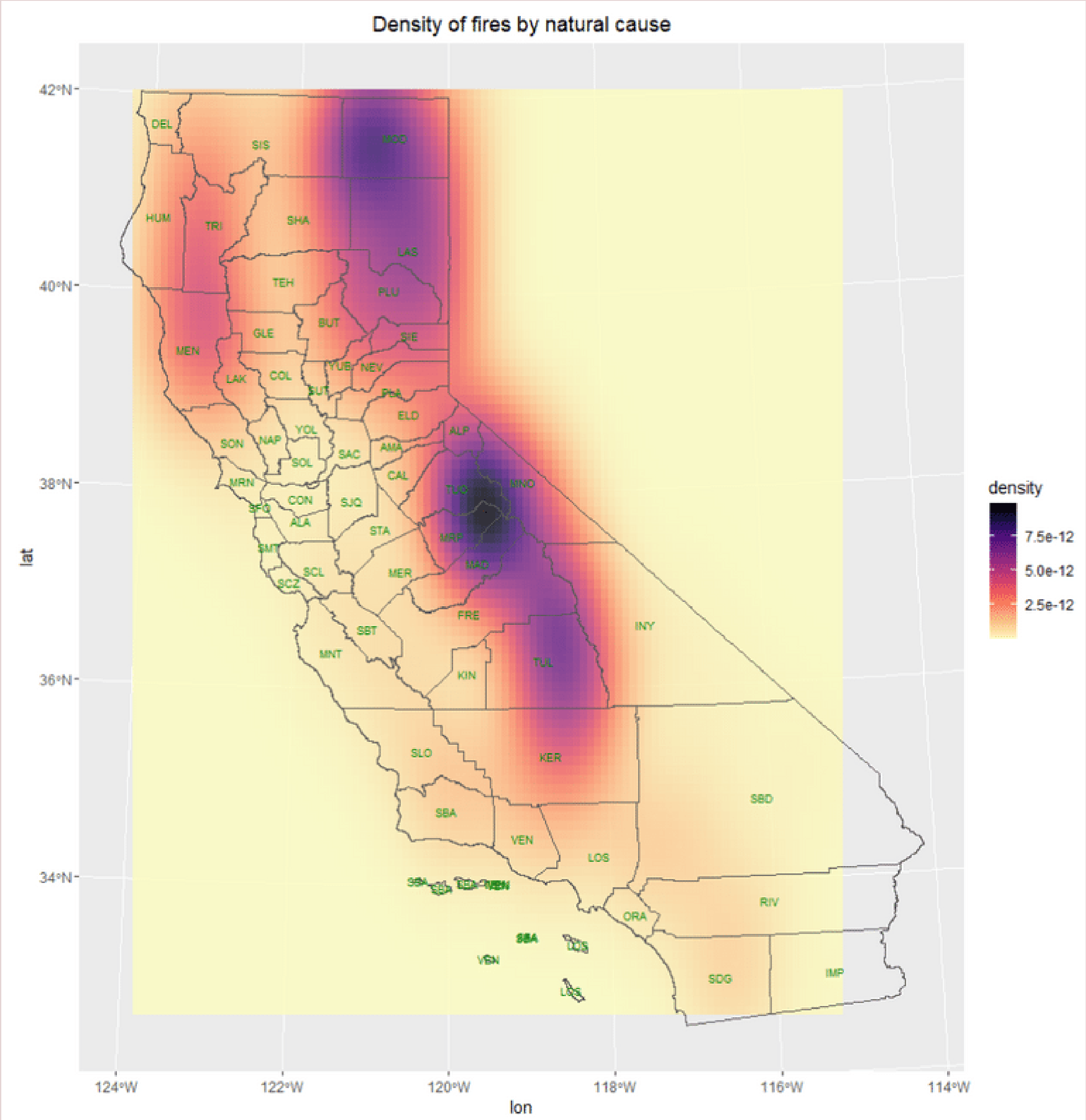
- Kern
- Ventura
- Orange
- Mariposa
- Yuba

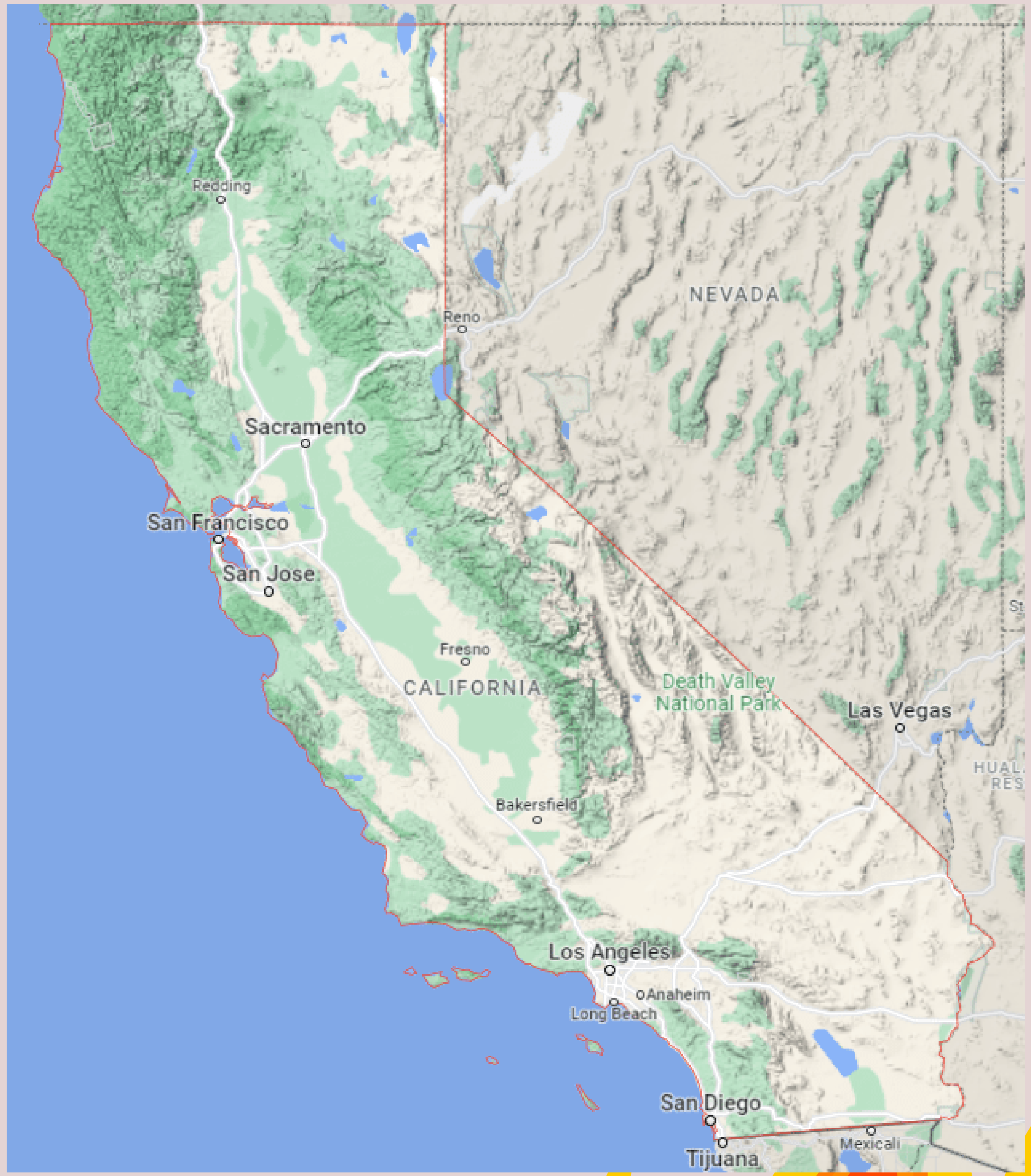
Fires Distribution

Causes: Human VS Natural

```
human = sep_coords %>%
  filter(Desc %in% c("Aircraft","Arson","Campfire", "Debris",
                   "Equipment Use", "Escaped Prescribed Burn",
                   "Firefighter Trainning","Illigal Alien Campfire",
                   "Non-Firefighter Training", "Playing with Fire",
                   "Power Line", "Railroad", "Smoking",
                   "Structure", "Vehicle"))

natural = sep_coords %>%
  filter(Desc == "Lightning" )
```

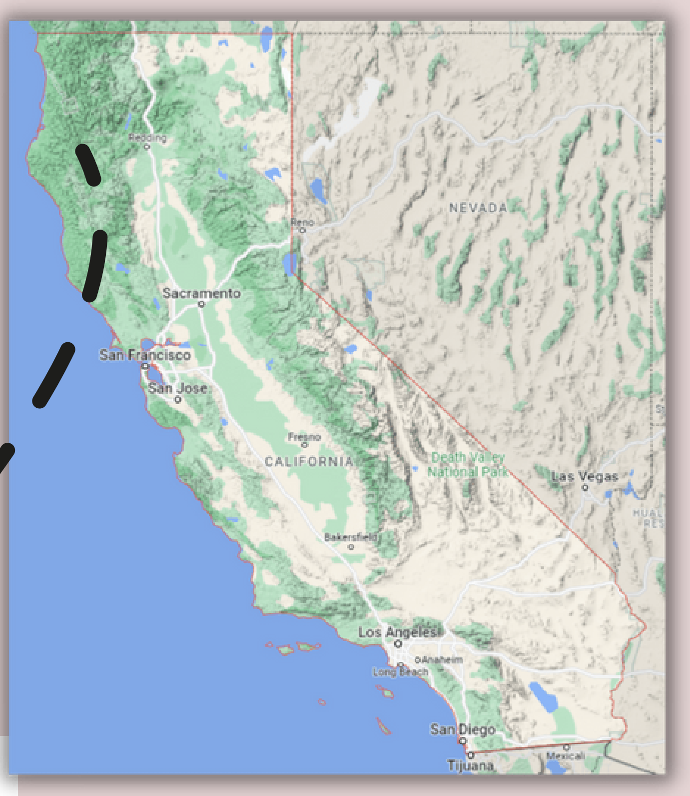






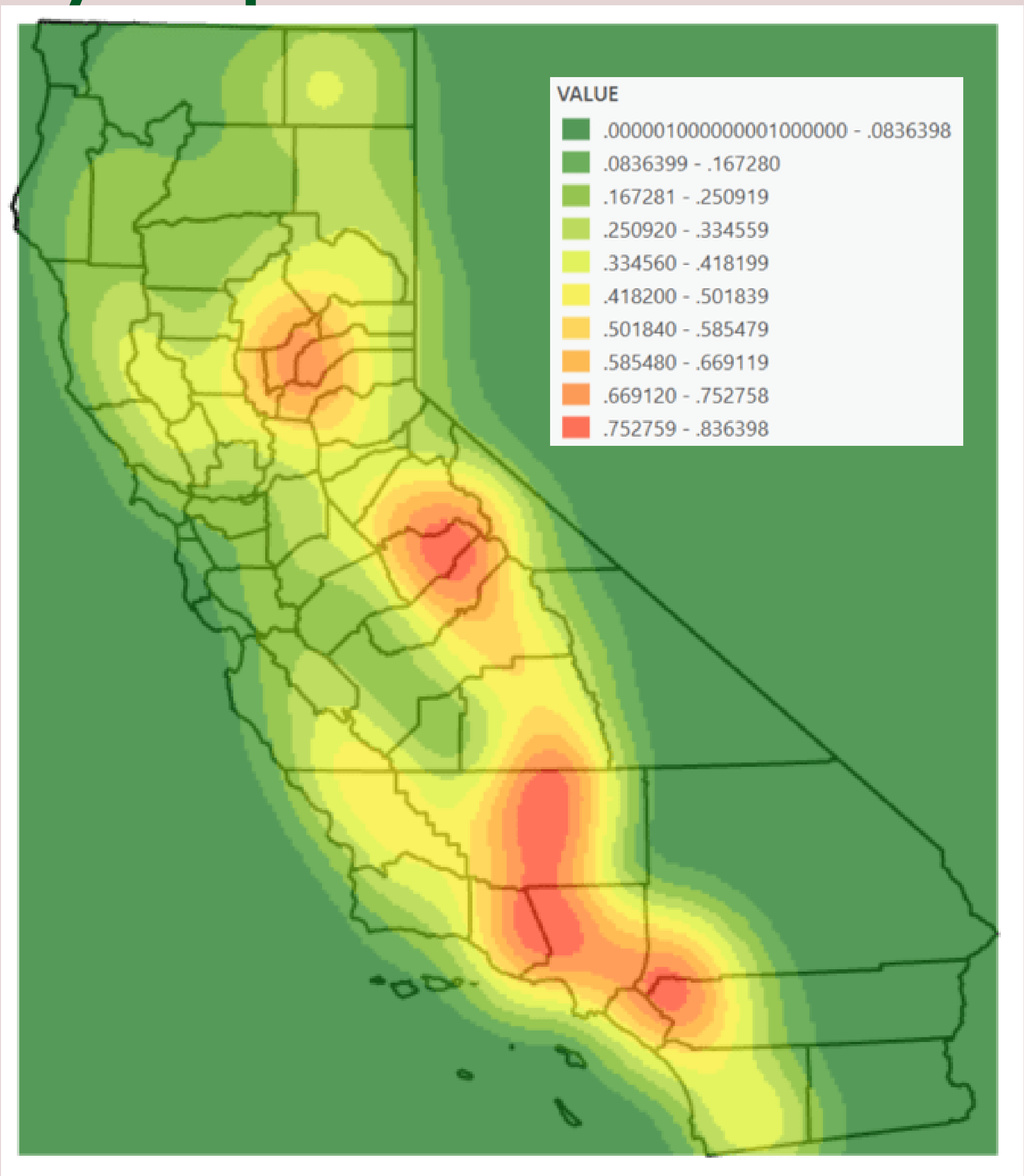
- Correlation between the human caused fire density to populates areas (More urban areas, especially along the coast)
- Correlation between the nature caused fire density to green areas

Kernel Density of fires
(ArcGIS PRO)

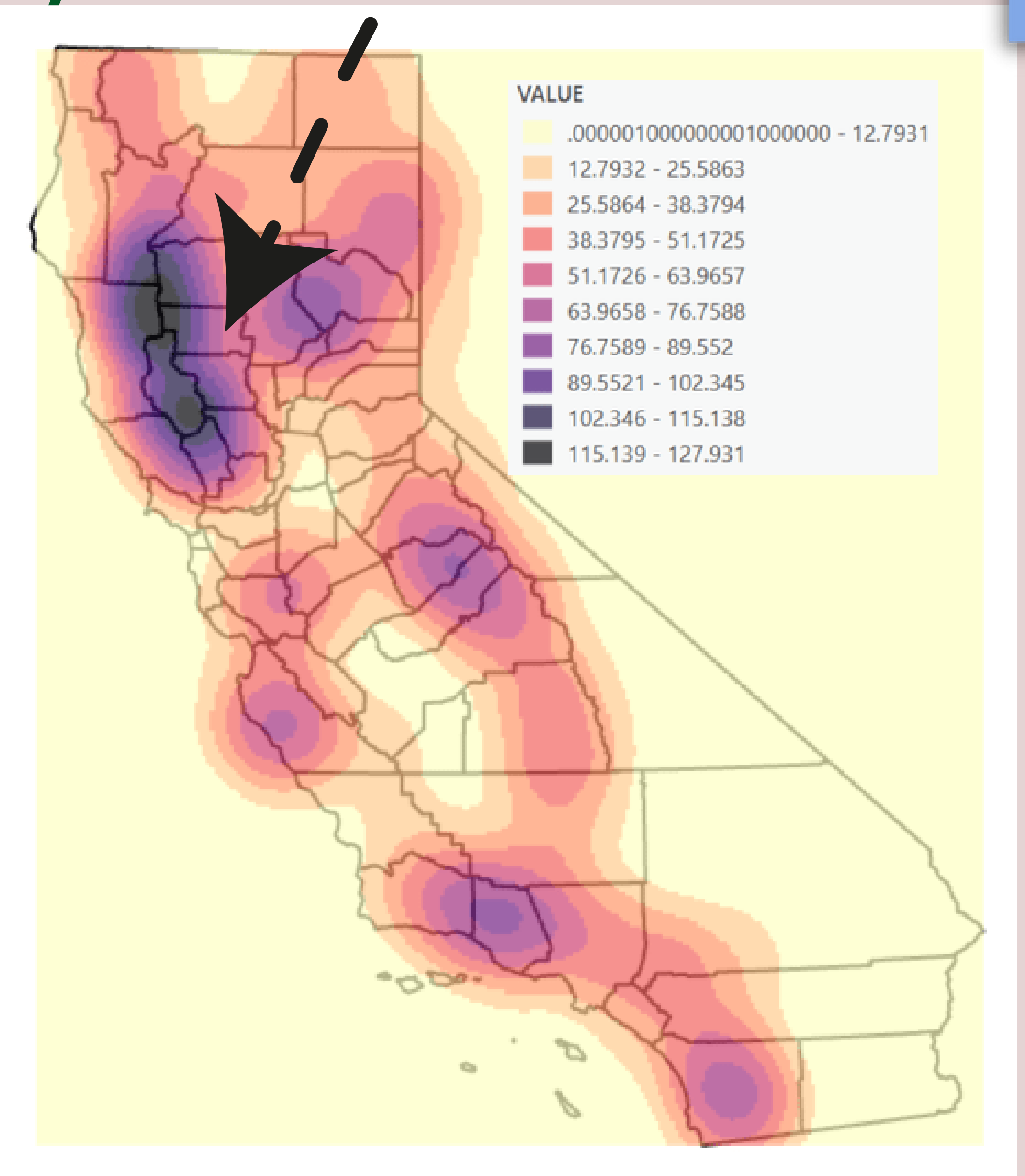


- Here we can see a correlation between the size of fire factor in kernel density to very wide forest areas.

By temprature



By fire size

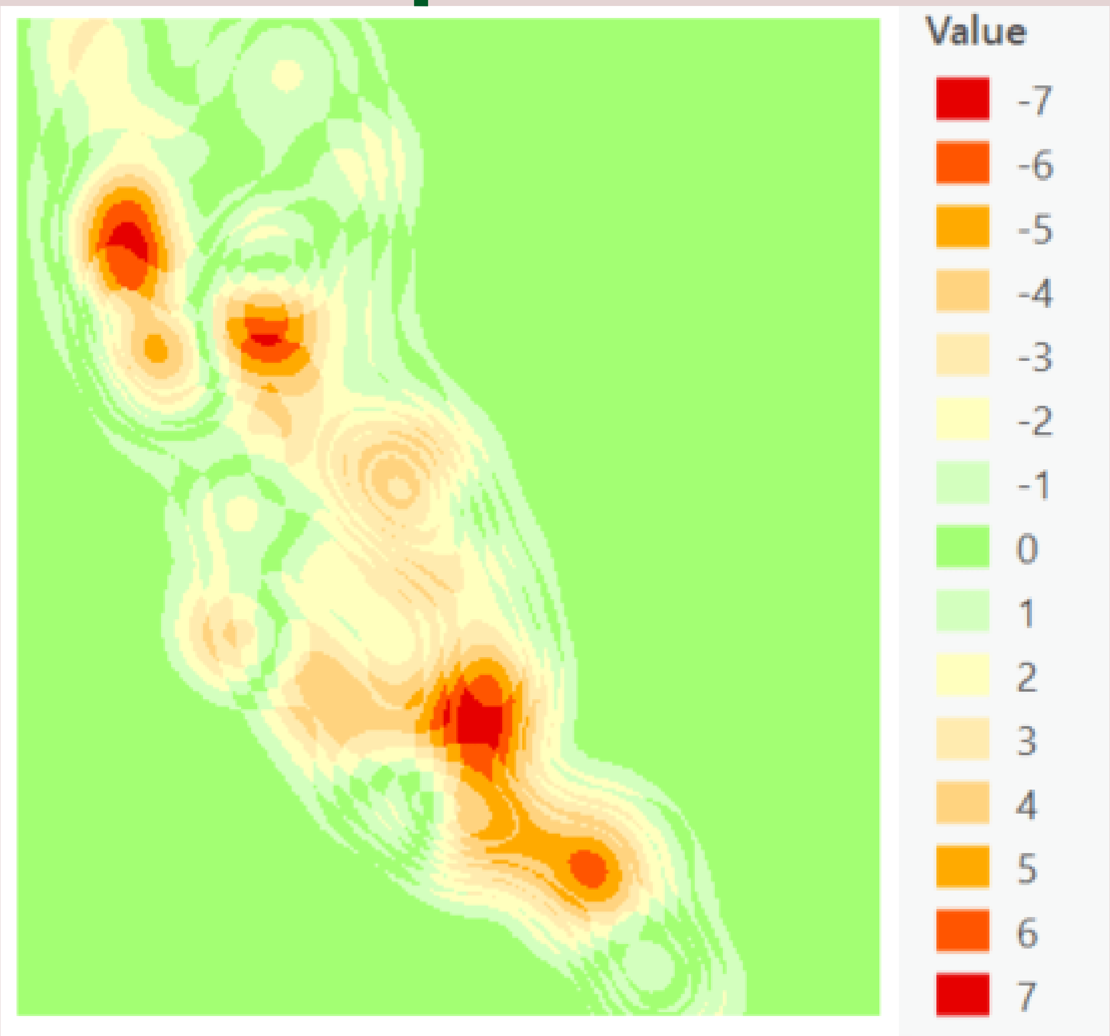


Reclassify -->

- We reclassified the rasters So that we can compare the two layers.

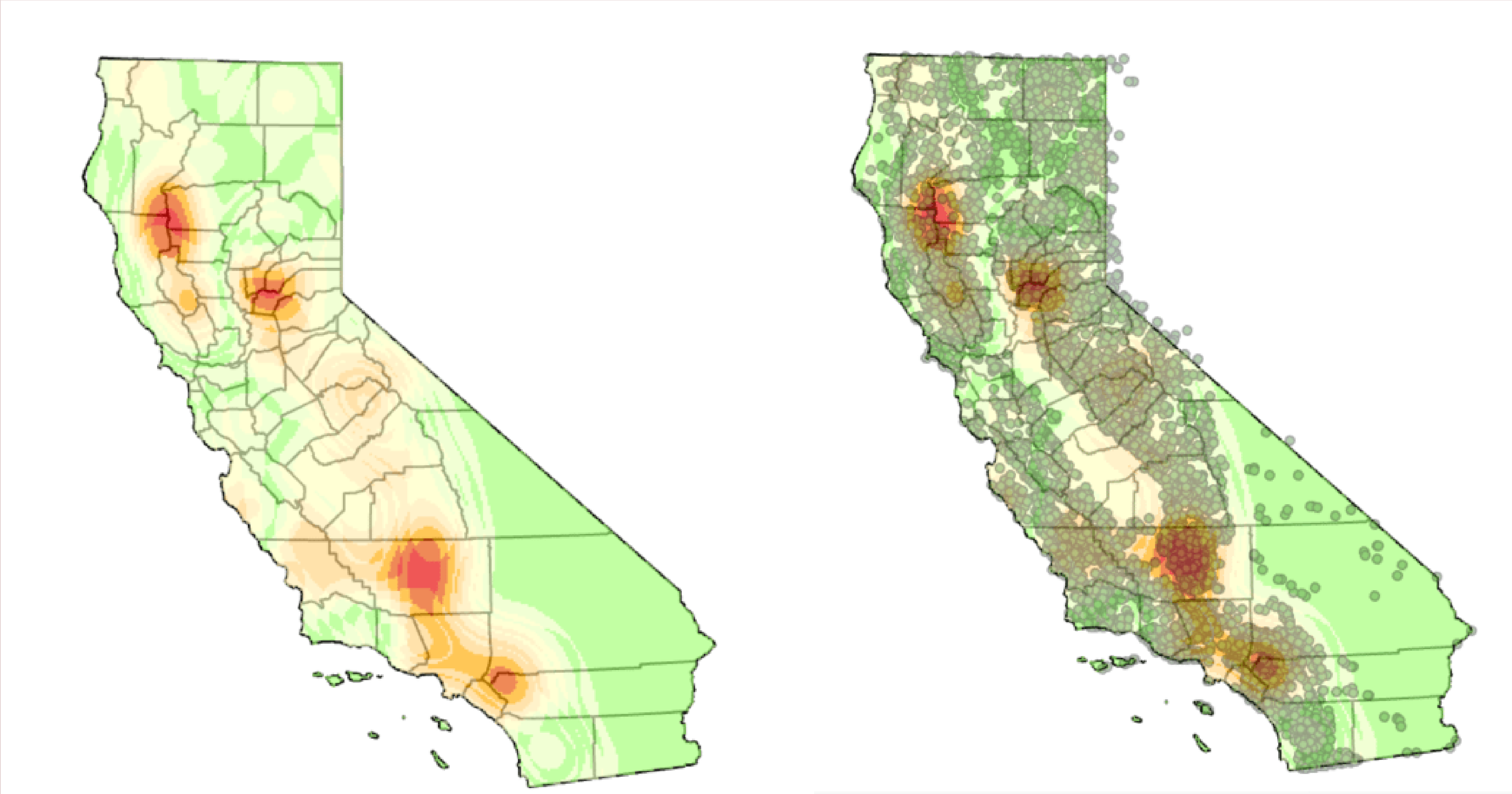
Kernel Density of fires
(ArcGIS PRO)

rater_temp - raster_size



- We subtracted between the two layers so that the pixels closest to 0 are the pixels with the best fit.

extract by mask (California counties)



- Most of the suitable areas are in areas where there have been almost no fires (near Navada Desert).

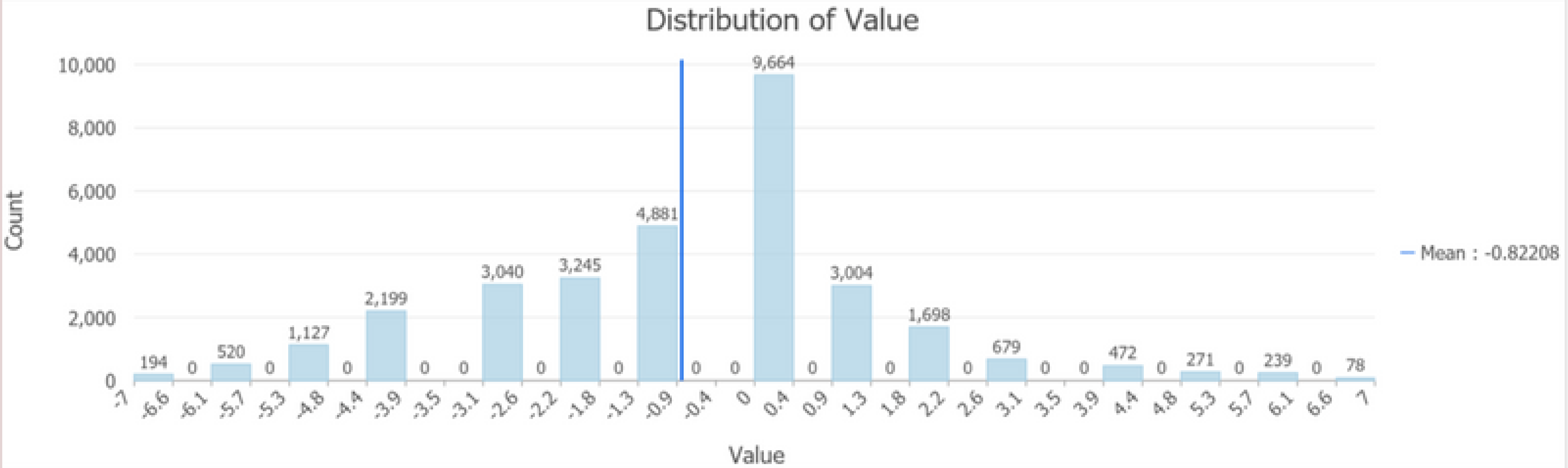
Kernel Density of fires
(ArcGIS PRO)

Statistics

Statistics

Adjust for cell count

	Dataset
Mean	-0.8
Median	0
Std. Dev.	2.25
Rows	72,790
Count	31,311
Nulls	41,479
Min	-7
Max	7
Sum	-25,740
Skewness	0.033
Kurtosis	3.7



30% of pixels (with 0 score) have a "good" match

- 30% is a low correlation, and as we said most of them are not within the fires area.

Conclusions

- Consistent with other studies, (Keely & syphard, 2017; williams et al., 2019), Our study has shown that the correlation between total burned area and climate factors is weak.

- This is largely because fire-climate relationship in California are strongly manipulated by humans via ignitions, land cover change, fire-preventing technologies and many-more. In other words, fires can be cause by a combination of various factors: climate factors, an non- climate factors, which make it complicated and hard to measure.

- We can see also a spatial correlation between human and natural caused fire to the land uses in California. Furthermore, large forests have the potential for expanding fires.

Conclusions

- While Miller et al (2012) showed that fire size, duration, number of fires, and total area burned per year were all significantly higher for lightning-ignited fires than for human-ignited fires.
- Our study shows that human-caused fires are getting more frequent during the examined years than lightning-caused fires.
- Also, as opposed to Keeley and Syphard (2018), the fires size and frequncy of our fire data has become more frequent and larger over the years.
- It could by explained by the fact that Miller (2012) analysed the fires which occurred in northwestern California during 1910-2008, And Keeley and Syphard (2018) analysed Central and South-coust California fire's during 1910-2010. While our study analysed the fires occurred in all California and during 1983-2020. And Also we used diffrent methodology and data sources
- Therefore, It is important that while we analyze data to look from where it came and when the phenomena has occurred.

Bibliography

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. (2017). Human-started wildfires expand the fire niche across the United States. *Proceeding of the National Academy of Scienced of the United States of America, 114*, 2946-2951.

Kasischke, E. S., Christensen Jr, N. L., & Stocks, B. J. (1995). Fire, global warming, and the carbon balance of boreal forests. *Ecological applications, 5*(2), 437-451.

Keely, E. J., & Syphard, D. A. (2018). Historical patterns of wildfire ignition sourced in California ecosystems. I*nternational Journal of Wildland Fire, 27*, 781-799.

Keely, E. J., & Syphard, D. A.. (2017) Different historical fire–climate patterns in California. *International Journal of Wildland Fire, 26*, 253-268.

Li, S., Banerjee, T. (2021). Spatial and temporal pattern of wildfires in California from 2000 to 2019. *Sci Rep, 11*, 8779 .

Bibliography

Miller, D. J., Skinner, N. C., Safford, D. H., Knapp. E. E., & Ramirez., C. M. (2012). Trends of causes of severity, size, and number of fires in northwestern California, USA. *Ecological Applications, 22*(1), 184-203.

Oris, F., Asselin, H., Ali, A. A., Finsinger, W., & Bergeron, Y. (2014). Effect of increased fire activity on global warming in the boreal forest. *Environmental Reviews, 22*(3), 206-219.

Stocks, B. J. (1993). Global warming and forest fires in Canada. *The Forestry Chronicle, 69*(3), 290-293.

Williams, P. A., Abatzoglou, T. J., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, K. J., & Lettenmaier, P, D. (2019). Observed Impacts of Anthropogenic climate change on wildfire in California. *Earth's Future, 7*, 892-910.