Fire Data

Load Packages library(“rworldmap”) library(“readxl”) library(“purrr”) library(“maps”) library(“mapdata”) require(“devtools”) library(“ggpubr”)

```
suppressPackageStartupMessages({
library("RSQLite")#to help us subset our SQL dataset
library("tidyr")#used to drop NA values
library("rspatial") #used for map of hawaii
library("dbplyr") #used in SQL databse connection set-up
library("dplyr") #used in SQL databse connection set-up
library("ggplot2") #used for visualization
library("usmap")#used for state-by-state comparisons
library("rnaturalearth")#used for map of alaska
library("rnaturalearthdata")#used for map of alaska
library(patchwork) #to display multiple kernel density maps
library("rgdal")
})
```

Data Source:

```
#maps we will use in our analysis
#hawaii map
hawaii <- subset(getData("GADM", country = "usa", level = 1), NAME_1 == "Hawaii")
#hawaii <- subset(usa, NAME_1 == "Hawaii")

#will be re positioned alaska map
us_ak<- subset(ne_countries(scale='medium',returnclass = 'sf'), admin == "United States of America")
```

Acess Fire data with SQL

```
fires <- DBI::dbConnect(RSQLite::SQLite(), "FPA_FOD_20170508.sqlite")
src_dbi(fires)
```

```
## src:  sqlite 3.36.0 [/Users/juliajones/Desktop/Stuff/fire data/FPA_FOD_20170508.sqlite]
## tbls: ElementaryGeometries, Fires, geom_cols_ref_sys, geometry_columns,
##   geometry_columns_auth, geometry_columns_field_infos,
##   geometry_columns_statistics, geometry_columns_time, idx_Fires_Shape,
##   idx_Fires_Shape_node, idx_Fires_Shape_parent, idx_Fires_Shape_rowid, KNN,
##   NWCG_UnitIDActive_20170109, spatial_ref_sys, spatial_ref_sys_all,
##   spatial_ref_sys_aux, SpatialIndex, spatialite_history, sql_statements_log,
##   sqlite_sequence, vector_layers, vector_layers_auth,
##   vector_layers_field_infos, vector_layers_statistics, views_geometry_columns,
##   views_geometry_columns_auth, views_geometry_columns_field_infos,
##   views_geometry_columns_statistics, virts_geometry_columns,
##   virts_geometry_columns_auth, virts_geometry_columns_field_infos,
##   virts_geometry_columns_statistics
```

get column names for our dataset

```
dbListFields(fires, "fires")
```

```
##  [1] "OBJECTID"                   "FOD_ID"
##  [3] "FPA_ID"                     "SOURCE_SYSTEM_TYPE"
##  [5] "SOURCE_SYSTEM"              "NWCG_REPORTING_AGENCY"
##  [7] "NWCG_REPORTING_UNIT_ID"     "NWCG_REPORTING_UNIT_NAME"
##  [9] "SOURCE_REPORTING_UNIT"      "SOURCE_REPORTING_UNIT_NAME"
## [11] "LOCAL_FIRE_REPORT_ID"       "LOCAL_INCIDENT_ID"
## [13] "FIRE_CODE"                  "FIRE_NAME"
## [15] "ICS_209_INCIDENT_NUMBER"    "ICS_209_NAME"
## [17] "MTBS_ID"                    "MTBS_FIRE_NAME"
## [19] "COMPLEX_NAME"               "FIRE_YEAR"
## [21] "DISCOVERY_DATE"             "DISCOVERY_DOY"
## [23] "DISCOVERY_TIME"             "STAT_CAUSE_CODE"
## [25] "STAT_CAUSE_DESCR"           "CONT_DATE"
## [27] "CONT_DOY"                   "CONT_TIME"
## [29] "FIRE_SIZE"                  "FIRE_SIZE_CLASS"
## [31] "LATITUDE"                   "LONGITUDE"
## [33] "OWNER_CODE"                 "OWNER_DESCR"
## [35] "STATE"                      "COUNTY"
## [37] "FIPS_CODE"                  "FIPS_NAME"
## [39] "Shape"
```

Pare down fire data to what will be used in our analysis

```
fire_info<-tbl(fires, sql("SELECT OBJECTID, SOURCE_SYSTEM, SOURCE_REPORTING_UNIT_NAME, FIRE_NAME, FIRE_
#head(fire_info, n = 10)
causes<-data.frame(fire_info %>%
    dplyr::select(STAT_CAUSE_DESCR,LATITUDE,LONGITUDE,STATE,FIRE_SIZE,FIRE_SIZE_CLASS))
head(causes)
```

```
##   STAT_CAUSE_DESCR LATITUDE LONGITUDE STATE FIRE_SIZE FIRE_SIZE_CLASS
## 1    Miscellaneous 40.03694 -121.0058    CA      0.10               A
## 2        Lightning 38.93306 -120.4044    CA      0.25               A
## 3   Debris Burning 38.98417 -120.7356    CA      0.10               A
## 4        Lightning 38.55917 -119.9133    CA      0.10               A
## 5        Lightning 38.55917 -119.9331    CA      0.10               A
## 6        Lightning 38.63528 -120.1036    CA      0.10               A
```

```
fire_info=NA
```

Define graphical functions for presentation

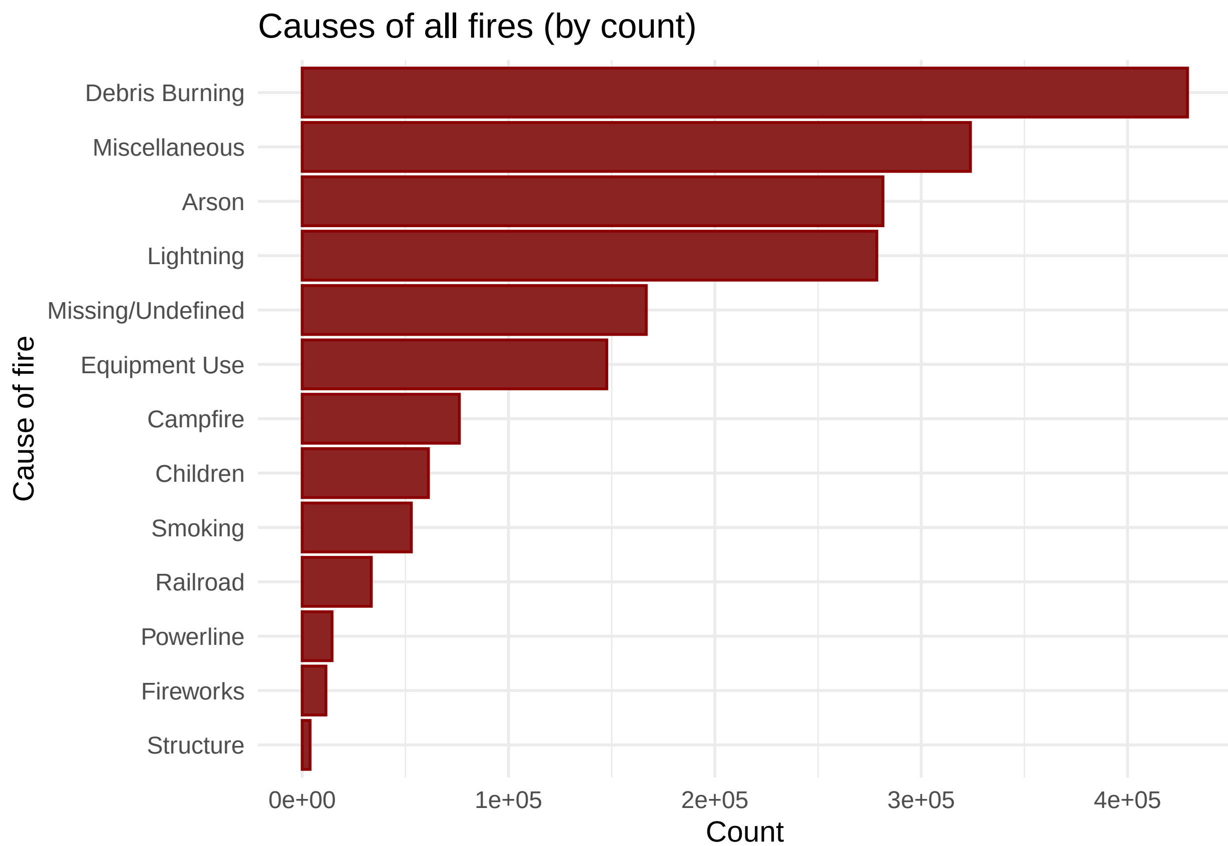
```
#function kernel_density_map creates: A 2d density map based on inputted latitude and longitude coordin
#df: data frame containing the latitude and longitude coordinates for each fire
#map: map we will overlay our kernel density on
#title: figure title
#hawaii: Boolean that sets limits on the dimensions of the kernel density map. When hawaii=true sets th
kernel_density_map<-function(df,map,title,hawaii=FALSE){
  us<-map
  #sets map dimensions to HI
  if(hawaii==TRUE){
     base_plot<-ggplot(causes, aes(x = LONGITUDE, y = LATITUDE))+theme_void()+xlim(-162,-150)+ylim(15,23
  }
```

```
  #sets dimensions to the lower 48
  else{
   base_plot<-ggplot(causes, aes(x = LONGITUDE, y = LATITUDE))+theme_void()+xlim(-125,-67)+ylim(25,50)
  }


  #overlay the inputted map and add labels
  base_plot<-base_plot+geom_polygon(data=us,aes(x=long,y=lat,group=group))+
  xlab('Longitude') + ylab('Latitude')
  #fine tune presentation of density plot
  base_plot<-base_plot+stat_density2d(aes(fill = ..level..), alpha = .5,
                         geom = "polygon", data = df) +
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma") +
  theme(legend.position = 'none')+ggtitle(title)

  return(base_plot)
}
#remove dual grouping (i think in geom polygon)?
scatter_map<-function(df,xx,yy,title){
  us<-map_data("state")
  #plot points that represent location of fire
  base_plot<-ggplot(df,aes_string(x=xx,y=yy))+
    geom_polygon(data=us,aes(x=long,y=lat,group=group),color='gray',fill=NA,alpha=.1)+
    geom_point(size=.05,alpha=.05,color="darkred")+xlim(-125,-65)+ylim(20,50)+
    ggtitle(title)+theme_minimal()+xlab("Longitude")+ylab("Latitude")

  #overlay map
 base_plot+xlim(-125,-65)+ylim(20,50)
}
#function bar_map creates a vertically aligned countplot from the passed in dataframe at column cat, se
#df: data frame containing the passed in column cat
#cat: column we will pull our information from
#title: figure title
bar_map<-function(df,cat,title){
ggplot(df, aes(x= reorder(cat,cat,function(x)+length(x)))) +
geom_bar(fill="brown4", color="darkred")+coord_flip()+ggtitle(title)

}

arrange_states<-function(AK, HI, l48){
#place alaska above hawaii in one pane
lb<-(AK+HI)+plot_layout(nrow=2)

#used to control figure sizing A=lower 48, B= alaska and hawaii figure
layout<- "
AAB
"
#put the lower 48 states on the left and the lb pane on the right
fig<-(l48/lb)+plot_layout(design = layout)
return(fig)
}
```

Examination of Fire Causes

Lets examine the most common causes for fires

```
# Show the causes of fire with counts in descending order
ggplot(dbGetQuery(fires, "
           SELECT
           STAT_CAUSE_DESCR,
           count(*) as [count]
           FROM Fires
           GROUP BY STAT_CAUSE_DESCR
           ORDER BY count DESC;"),aes(y=count,x=reorder(STAT_CAUSE_DESCR,count)))+geom_bar(stat="identi
```



Lets simplify the causes listed above into three more palatable categories. Natural fires (N) are those caused by lightning. Unknown fires (UK) are the miscellaneous fires or the missing/undefined fires. Man-made fires (M) are all of the other fire categories in our dataset.

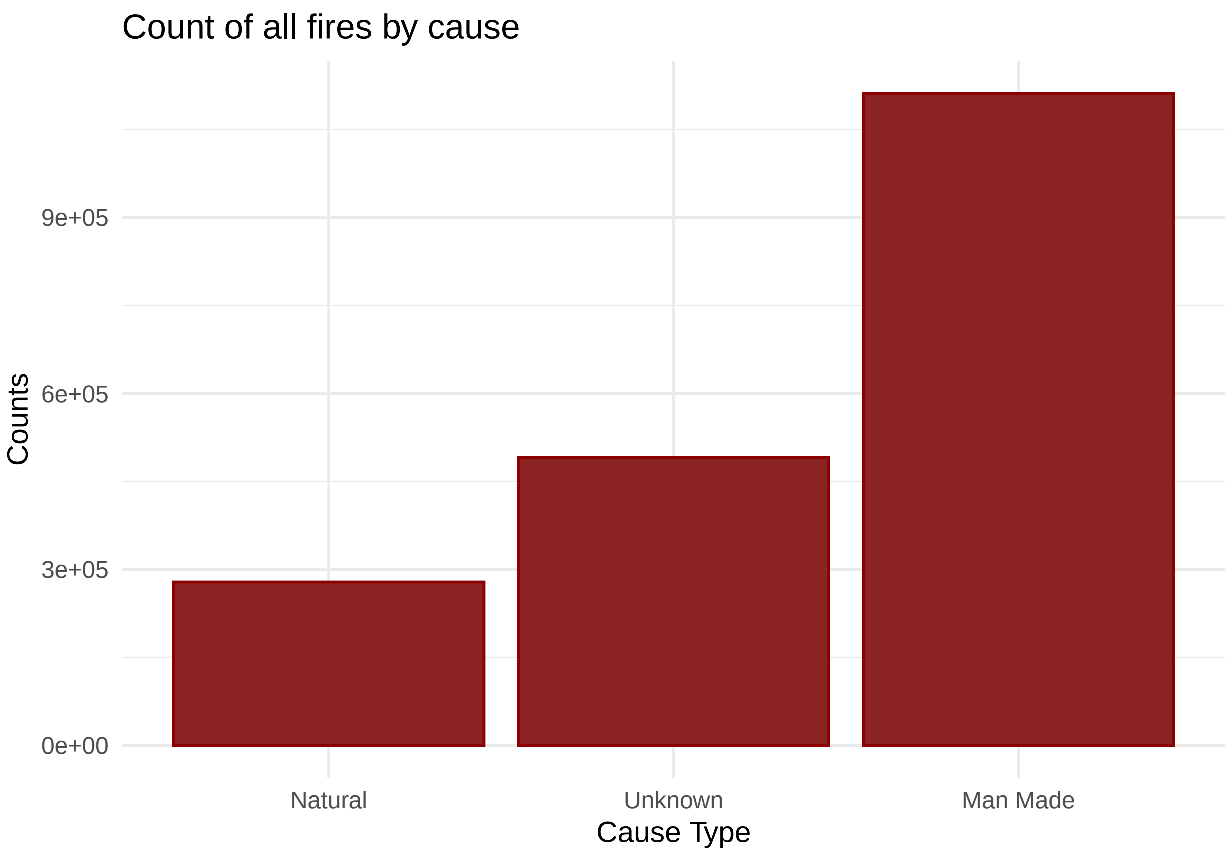
```
#sql query that separates our STAT_CAUSE_DESCR data into the three groups
# outlined above. This column is names CAUSE_GROUP and is what our df is grouped
# on. We then sum the count of each of these categories with the
# column count and the acres burned by each category with total_areas
fires_by_cause<-dbGetQuery(fires, "
           SELECT
           CASE
           WHEN STAT_CAUSE_DESCR == 'Lightning' THEN 'N'
           WHEN TRIM(STAT_CAUSE_DESCR) == 'Miscellaneous' OR
           STAT_CAUSE_DESCR == 'Missing/Undefined' THEN 'UK'
           ELSE 'M'
```

```
           END as 'CAUSE_GROUP',
           count(*) as [count],
           ROUND(SUM(FIRE_SIZE), 0) as total_area
           FROM Fires
           GROUP BY CAUSE_GROUP
           ")
#peek at results
fires_by_cause
```

```
##   CAUSE_GROUP   count total_area
## 1           M 1111469   29953119
## 2           N  278468   87033501
## 3          UK  490528   23145930
```

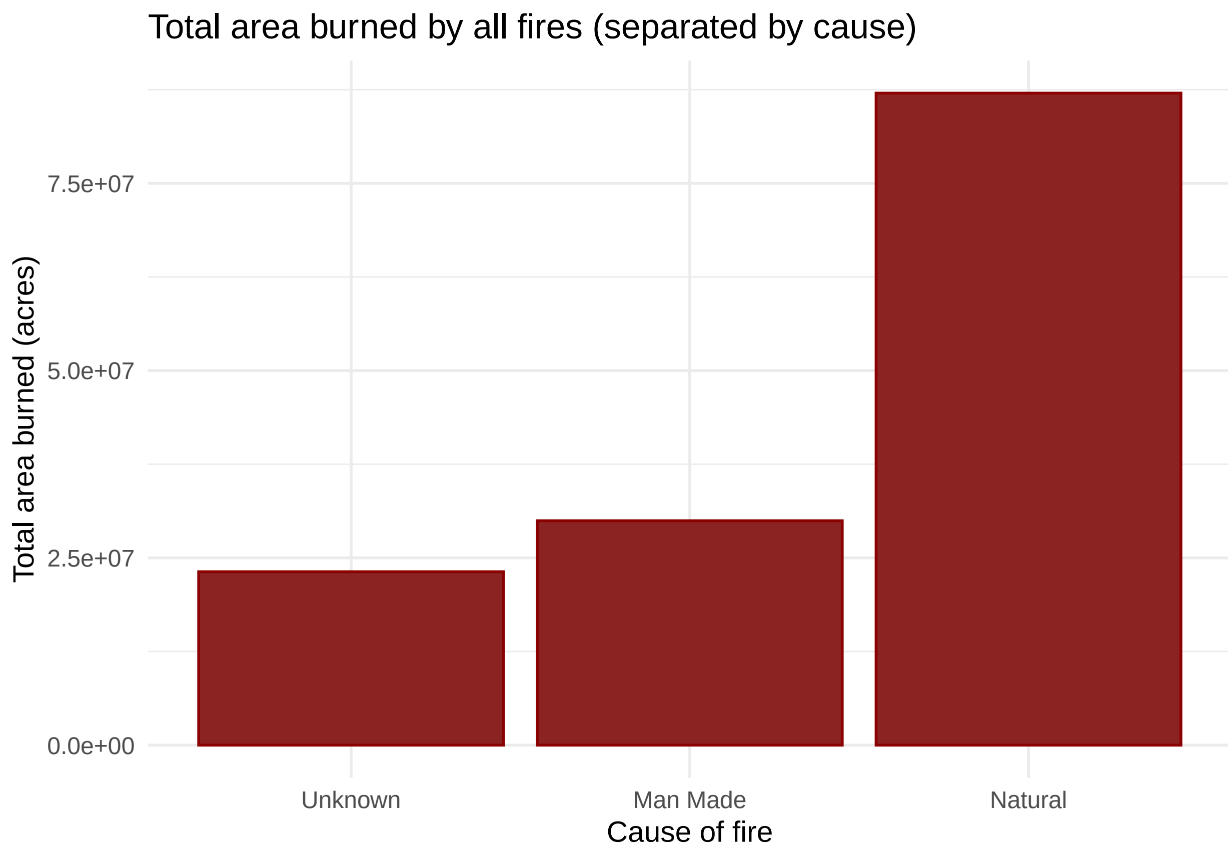
Now with one SQL call the differences between these categories both in raw counts and the area burned is clear. Lets represent these differences in a count plot

```
ggplot(fires_by_cause,aes(y=count,x=reorder(CAUSE_GROUP,count)))+
  geom_bar(stat="identity",fill="brown4",color="darkred")+
  theme_minimal()+scale_x_discrete(labels=c("Natural", "Unknown",
  "Man Made"))+labs(x="Cause Type",y="Counts",title=
                    "Count of all fires by cause")
```



Man made fires are by far the most common. Man-made fires are 3x more common than natural fires and 2x more common as unknown fires. But not every fires is the same in severity or magnitude. Now lets examine the area burned by each category of fire.

```
ggplot(fires_by_cause,aes(y=total_area,x=reorder(CAUSE_GROUP,total_area)))+
  geom_bar(stat="identity",fill="brown4", color="darkred")+
  theme_minimal()+scale_x_discrete(labels=
  c("Unknown", "Man Made","Natural"))+labs(x="Cause of fire",y="Total area burned (acres)",title="Total
```



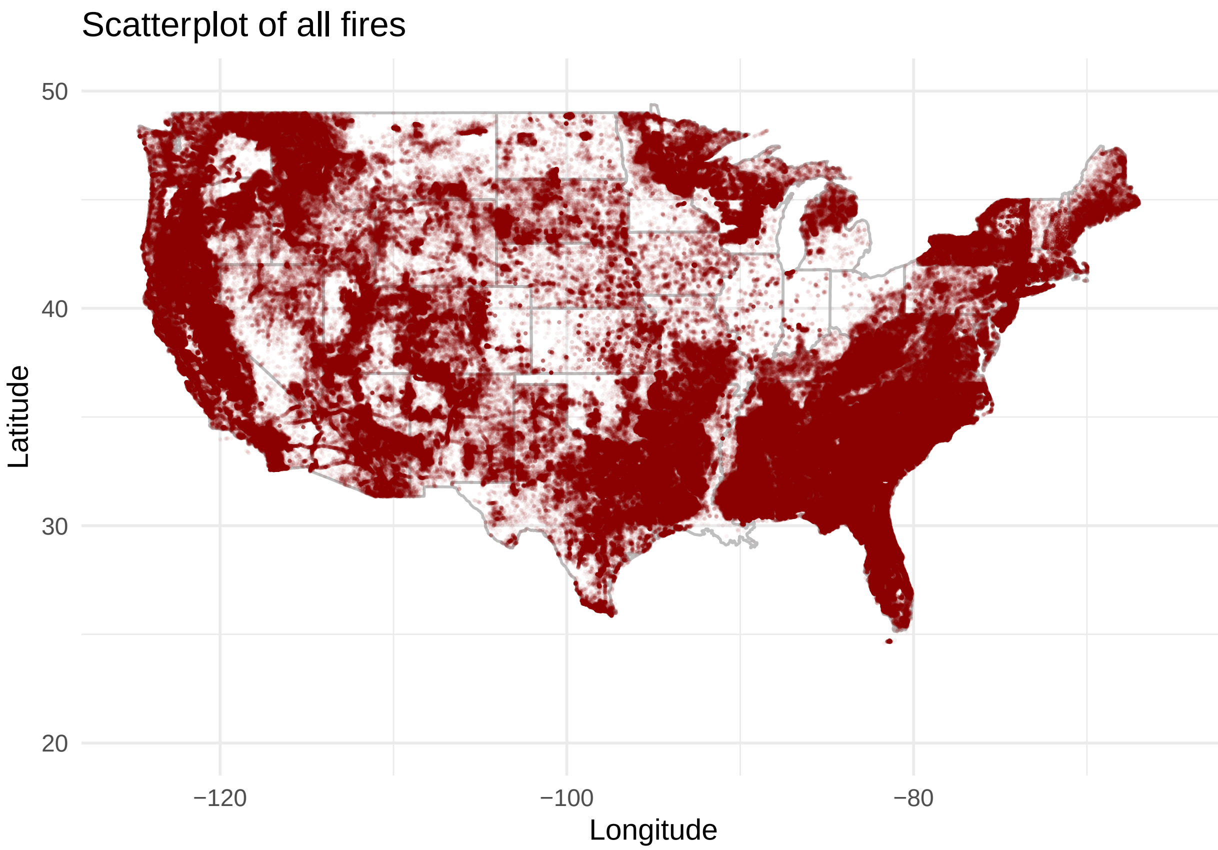
Even though the natural fires have the lowest count they have burned the most acreage. Perhaps this is because these natural fires tend to be further from population centers so they take longer to be noticed?

Map based analysis

Now we will examine every fire occurrence on the map of the United States to see if any trends are visible.

```
scatter_map(dbGetQuery(fires, "
SELECT
LATITUDE,
LONGITUDE
FROM fires"),"LONGITUDE","LATITUDE","Scatterplot of all fires")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.

## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.

## Warning: Removed 44819 rows containing missing values (geom_point).
```



With so many fires in our dataset this map is essentially worthless. So many points are overlapping the trends of the data are largely inscrutable. A more valuable metric would be the kernel density map of all of our fires. Lets see if this better elucidates the undelying trends in the data.

Due to inconsistencies with the map of Alaska in the mapping package I have been utilizing for the lower 48 states and Hawaii, I have chosen to use a map from a different package. This changes our kernel density code from the function kernel_density_map. Alaska still is using the same kernel density formula under the hood so there shouldn't be inconsistencies in our results.

```
#alaska map
ak_poly<-ggplot(data=dbGetQuery(fires, "
           SELECT
           LATITUDE, LONGITUDE
           FROM fires
           WHERE STATE in ('AK')
           "),aes(y=LATITUDE, x=LONGITUDE))+
  geom_sf(data=us_ak$geometry,inherit.aes = FALSE,color="black",fill="black")+
  stat_density2d(aes(fill = ..level..), alpha = .5,geom = "polygon")+
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma")+
  xlim(-180,-130)+ylim(51,71)+theme_void()+theme(legend.position = 'none')
```

Lets calculate our kernel density for Hawaii and the lower 48 states, which are located in different maps.

```
hi<-kernel_density_map(dbGetQuery(fires, "
           SELECT
           FIRE_YEAR, LATITUDE, LONGITUDE, FIRE_SIZE
           FROM fires
           WHERE STATE in ('HI')
           "),hawaii, "",TRUE)
```

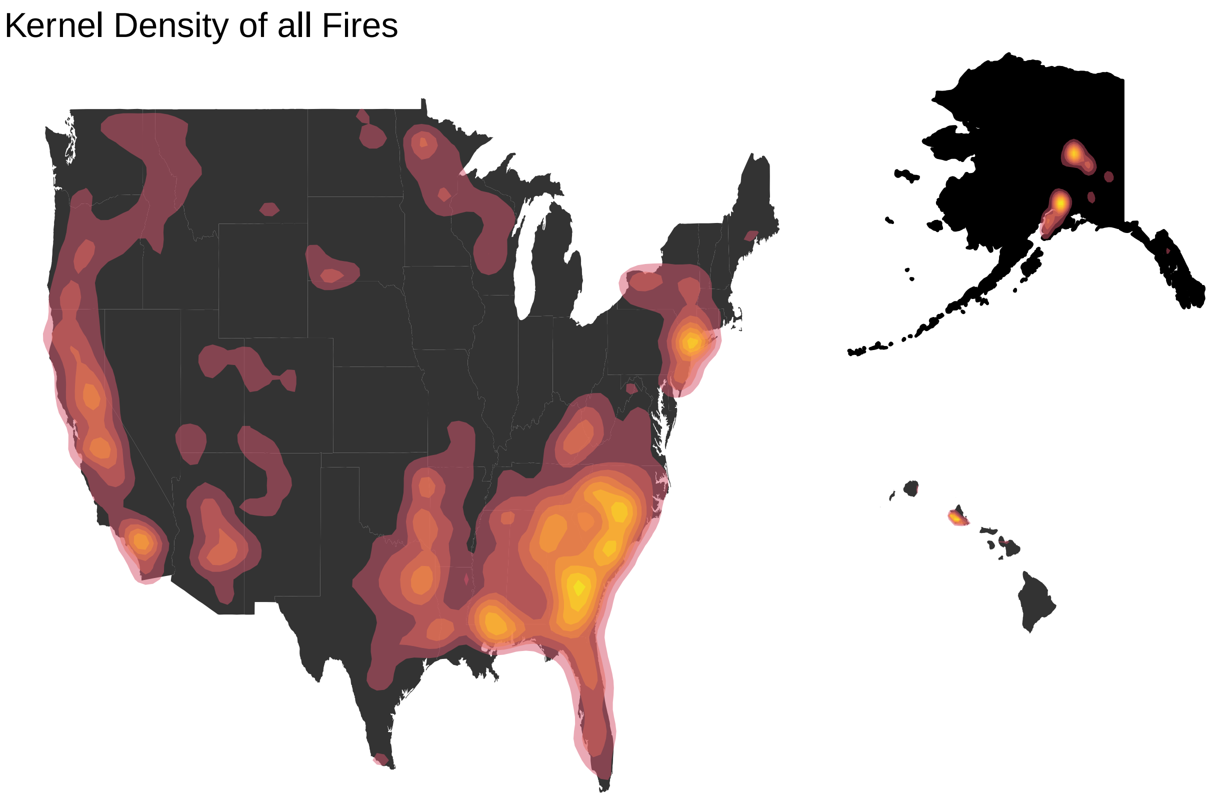
```
## Regions defined for each Polygons
```

```
#lower 48 map
l48<-kernel_density_map(causes,map_data("state"),"Kernel Density of all Fires",FALSE)
```

Now we will display all three of our maps with the patchwork library.

```
#place alaska above hawaii in one pane
lb<-(ak_poly+hi)+plot_layout(nrow=2)

#used to control figure sizing A=lower 48, B= alaska and hawaii figure
layout<- "
AAB
"
#put the lower 48 states on the left and the lb pane on the right
(l48/lb)+plot_layout(design = layout)
```

```
## Warning: Removed 45116 rows containing non-finite values (stat_density2d).
```



Fires appear to be densest in the south, with smaller clusters around California and New York. In Alaska the fires are centered around Alaska's two largest urban centers: Anchorage and Fairbanks In Hawaii the fires are centered on Oahu. Now lets see if fires follow different trends causally.

```
comb_causes<-dbGetQuery(fires, "
           SELECT
           CASE
           WHEN STAT_CAUSE_DESCR == 'Lightning' THEN 'N'
           WHEN TRIM(STAT_CAUSE_DESCR) == 'Miscellaneous' OR
           STAT_CAUSE_DESCR == 'Missing/Undefined' THEN 'UK'
           ELSE 'M'
           END as 'CAUSE_GROUP',
```

```
           FIRE_SIZE,
           LATITUDE,
           LONGITUDE,
           STATE
           FROM Fires
           ")

#unkown is at c[[3]] but wont examine those
c<-split(comb_causes, comb_causes$CAUSE_GROUP)
man<-drop_na(c[[1]])
natural<-drop_na(c[[2]])
```

Now that we have separated out fires by cause, lets see what these density maps elucidate. We will first examine the fires with natural causes (i.e. lightning strikes that led to a subsequent fire).

```
#nothing changes about our lower 48 call, except the dataframe we are passing in
natural_48<-kernel_density_map(natural,map_data("state"), "Kernel Density of natural fires")
```

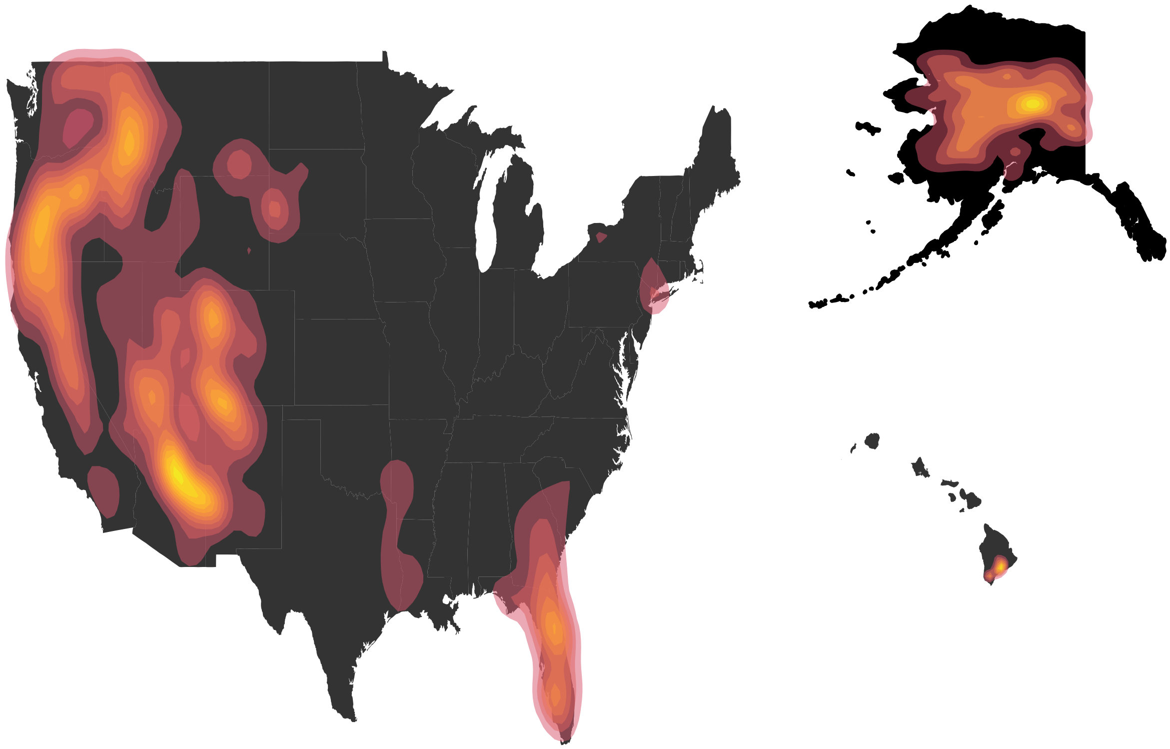
```
#for Hawaii and Alaska we will pull from this dataframe instead of making a SQL
#call for the data.
HI_natural<-natural[natural$STATE=="HI",]
AK_natural<-natural[natural$STATE=="AK",]

ak_poly_natural<-ggplot(AK_natural,aes(y=LATITUDE, x=LONGITUDE))+
  geom_sf(data=us_ak$geometry,inherit.aes = FALSE,color="black",fill="black")+
  stat_density2d(aes(fill = ..level..), alpha = .5,geom = "polygon")+
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma")+
  xlim(-180,-130)+ylim(51,71)+theme_void()+theme(legend.position = 'none')

hi_poly_natural<-kernel_density_map(HI_natural,hawaii, "",TRUE)
```

```
## Regions defined for each Polygons
```

```
lb<-(ak_poly_natural+hi_poly_natural)+plot_layout(nrow=2)
layout<- "
AAB
"
(natural_48/lb)+plot_layout(design = layout)
```

```
## Warning: Removed 4125 rows containing non-finite values (stat_density2d).
```

Kernel Density of natural fires



When looking at natural fires, their density is dramatically shifted west. Only the natural fires in Florida are notable in the south. In Alaska, the density of the fires have shifted from Alaska's urban centers (Anchorage and Fairbanks) to the center of the middle of the state, including Fairbanks. In Hawaii the fires are centered on the big island around two national parks encompassing the bottom of the big island.

Now we will examine the density maps of fires caused by humans.

```r
#lower 48 map
man_48<-kernel_density_map(man,map_data("state"),"Density of man-made fires")
```

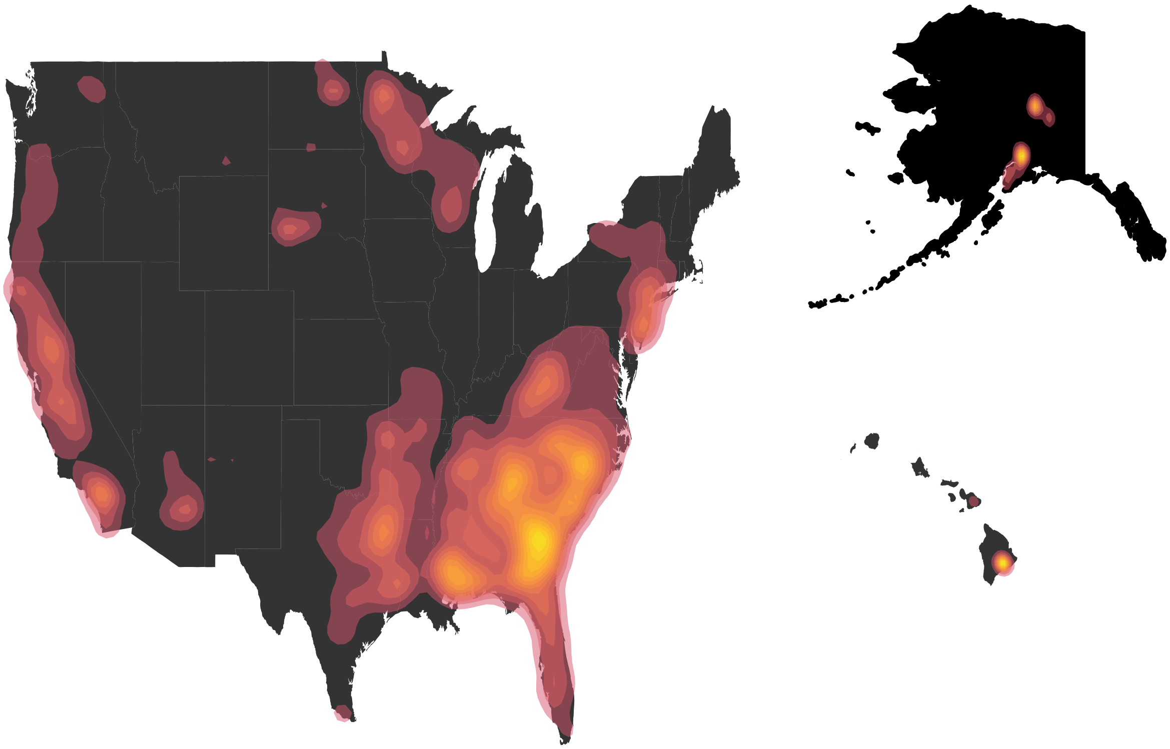
```r
#Hawaii and Alaska maps
HI_man<-man[man$STATE=="HI",]
AK_man<-man[man$STATE=="AK",]

ak_poly_man<-ggplot(AK_man,aes(y=LATITUDE, x=LONGITUDE))+
  geom_sf(data=us_ak$geometry,inherit.aes = FALSE,color="black",fill="black")+
  stat_density2d(aes(fill = ..level..), alpha = .5,geom = "polygon")+
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma")+
  xlim(-180,-130)+ylim(51,71)+theme_void()+theme(legend.position = 'none')

hi_poly_man<-kernel_density_map(HI_man,hawaii, "",TRUE)
```

```
## Regions defined for each Polygons
```

```r
lb<-(ak_poly_man+hi_poly_man)+plot_layout(nrow=2)
layout<- "
AAB
"
(man_48/lb)+plot_layout(design = layout)
```

```
## Warning: Removed 6933 rows containing non-finite values (stat_density2d).
```

Density of man−made fires



For the lower 48 states and Alaska this map of man-made fires closely matches that of the map of all fires. Hawaii's map hardly shifts, considering the big island is also the largest urban center in the state.

The above density maps count each fire equally, no matter the size. To get rid of this bias, lets examine the acres burned by state.

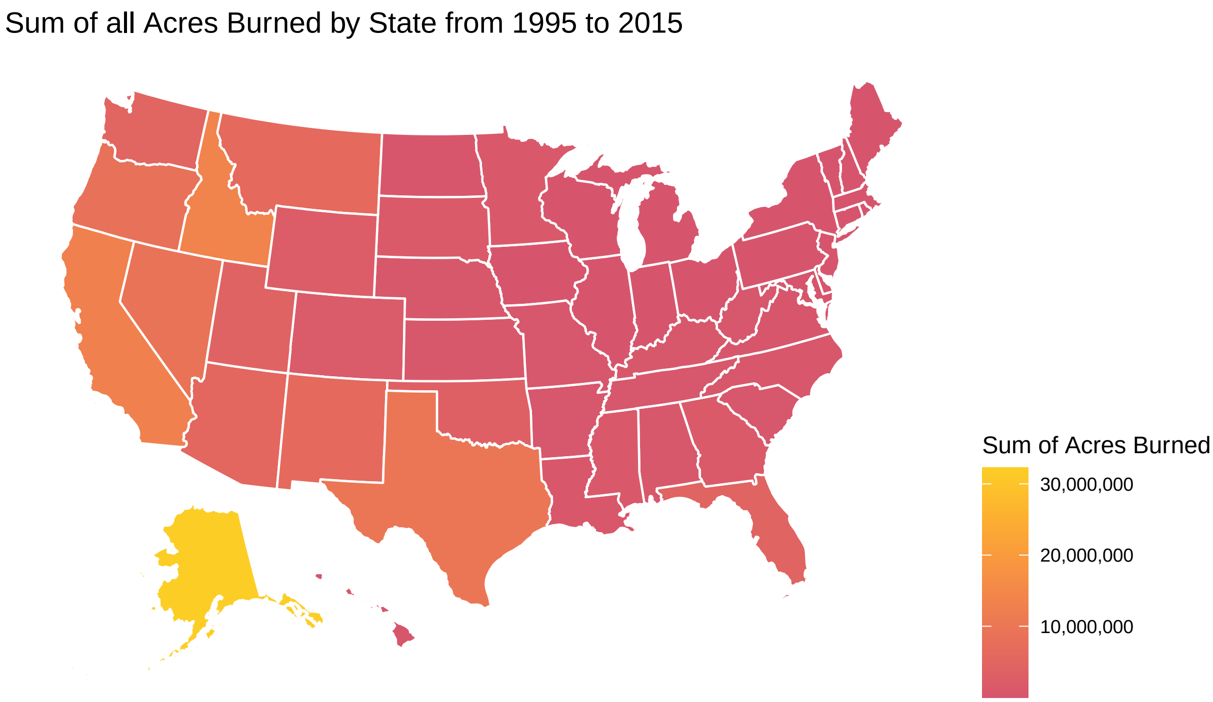
```
state_causes<-dbGetQuery(fires, "
           SELECT
           STATE AS state,
           ROUND(SUM(FIRE_SIZE), 0) as total_area
           FROM Fires
           Where STATE not in ('PR','DC')
           GROUP BY STATE
           ")

head(state_causes)
```

```
##   state total_area
## 1    AK   32233094
## 2    AL     920545
## 3    AR     508912
## 4    AZ    5576681
## 5    CA   12745859
## 6    CO    1842309
```

Examination of fires by state

```
plot_usmap(data = state_causes,values="total_area",color="white") +
  scale_fill_viridis_c(alpha=1,begin=.55, end=0.9,option="plasma",
```

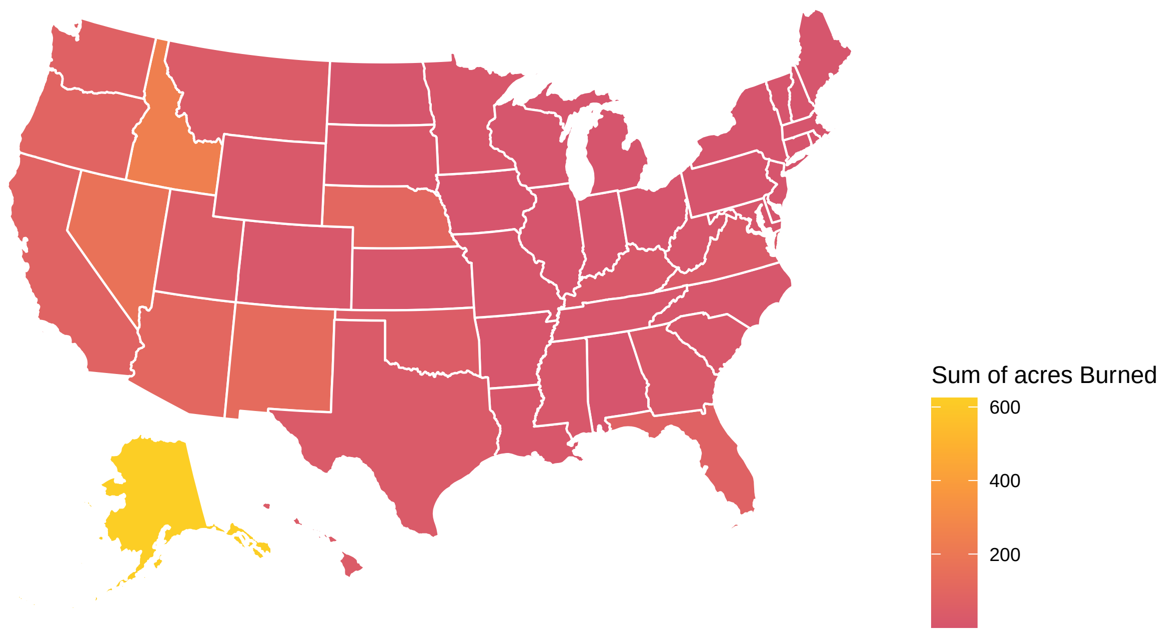
```
                       name = "Sum of Acres Burned", label = scales::comma) +
  labs(title = "Sum of all Acres Burned by State from 1995 to 2015") +
  theme(legend.position = "right")
```



At this scale acres burned is clearly connected to the size of the state with larger states having more total acres burned by fires. This makes sense because they have a larger burn potential then the smaller states.Lets see if this trend persists when we account for the area of the states.

```
#get area of states from built in R data
data("state")
df<-data.frame("area_km2"=state.area)
#see which state has had the most burn relative to their size
#(area burned/state size)
state_causes$burned_by_area<-state_causes$total_area/df$area_km2
plot_usmap(data = state_causes,values="burned_by_area",color="white")+
  scale_fill_viridis_c(alpha=1,begin=.55, end=0.9,option="plasma",
                       name = "Sum of acres Burned", label = scales::comma)+
  labs(title = "Sum of all Acres Burned by State from 1995 to 2015 (Controlled by Area)")+
  theme(legend.position = "right")
```

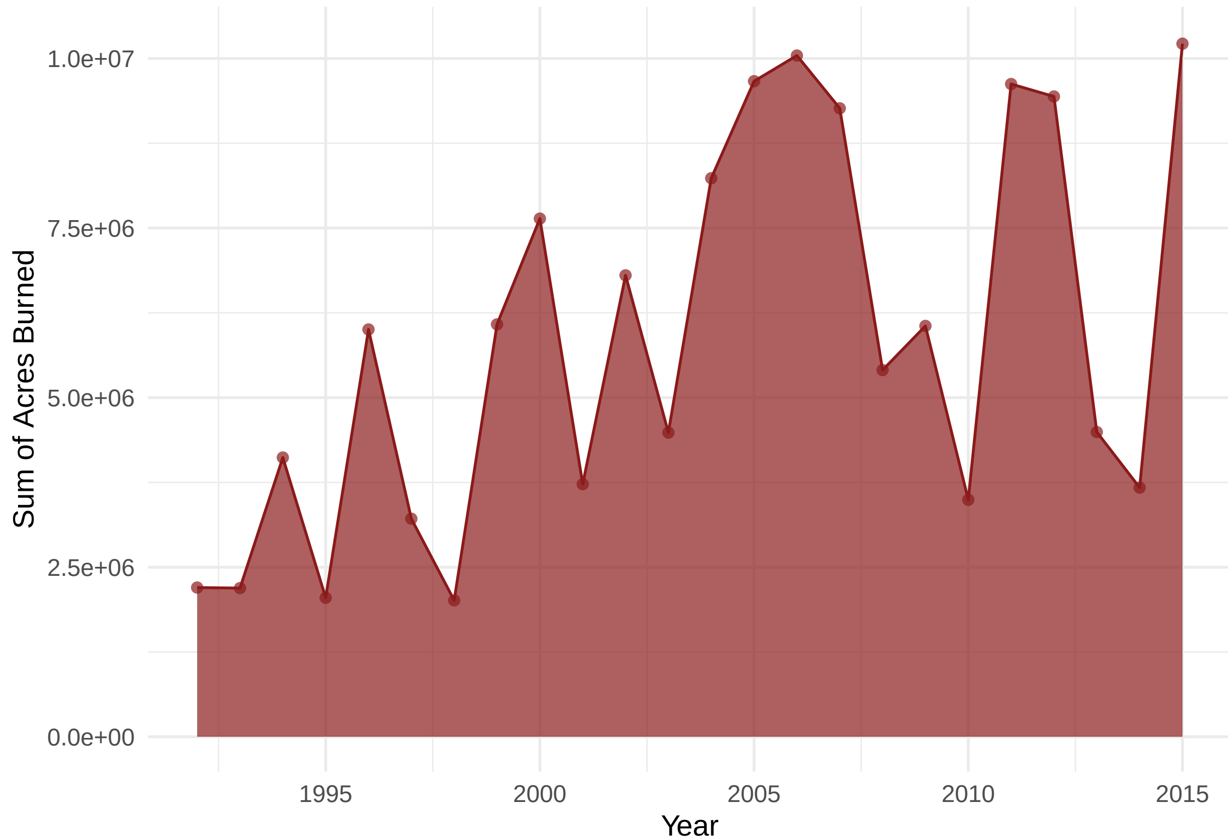
Sum of all Acres Burned by State from 1995 to 2015 (Controlled by Area)



When accounting for the size of the state, most of the states perform the same. The most obvious outlier is Alaska which lacks the population and in turn the resources (or need) to put out as many fires. Other larger states with lower population such as Idaho, Nevada, New Mexico and Nebraska follow this trend to a lesser degree.

Examanation of fires by year

```
ggplot(data=dbGetQuery(fires, "
           SELECT
           FIRE_YEAR as year,
           ROUND(SUM(FIRE_SIZE), 0) as total_area
           FROM Fires
           GROUP BY FIRE_YEAR
           "), aes(x=year, y=total_area, group=1)) +
  geom_area(alpha=0.7,fill="firebrick4")+
  geom_line(color="firebrick4")+
  geom_point(alpha=0.7,colour="firebrick4")+
  ylim(0,10250000)+theme_minimal()+xlab("Year")+ylab("Sum of Acres Burned")
```



Examining this plot, we can see the smallest fire year was in 1998 and the largest fire year was in 2015. 8,207,166, or 5x more acres burned in 2015 than 1998. Is there a functional reason for this difference (e.g., more man-made fires, more fires in an ill-equipped area) or did these two years follow the classical trends just on differing scales?

Start of unfinished work: First we will query our dataset for fires from 2015 and 1998

```
year_2015<-dbGetQuery(fires, "
           SELECT
           LATITUDE, LONGITUDE, FIRE_SIZE
           FROM fires
           WHERE FIRE_YEAR in (2015)
           ")

year_1998<-dbGetQuery(fires, "
           SELECT
           LATITUDE, LONGITUDE, FIRE_SIZE
           FROM fires
           WHERE FIRE_YEAR in (1998)
           ")
```

First we will calculate the 1998 data

```
#hawaii
hi_98<-kernel_density_map(year_1998,hawaii, "",TRUE)
```

```
## Regions defined for each Polygons
```

```
#alaska
ak_98<-ggplot(year_1998,aes(y=LATITUDE, x=LONGITUDE))+
```

```
  geom_sf(data=us_ak$geometry,inherit.aes = FALSE,color="black",fill="black")+
  stat_density2d(aes(fill = ..level..), alpha = .5,geom = "polygon")+
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma")+
  xlim(-180,-130)+ylim(51,71)+theme_void()+theme(legend.position = 'none')
#lower 48
l48_98<-kernel_density_map(year_1998,map_data("state"),"Density of fires in 1998")

panel_1998<-arrange_states(ak_98,hi_98,l48_98)
```

Then the 2015 data

```
#hawaii
hi_15<-kernel_density_map(year_2015,hawaii, "",TRUE)
```

```
## Regions defined for each Polygons
```

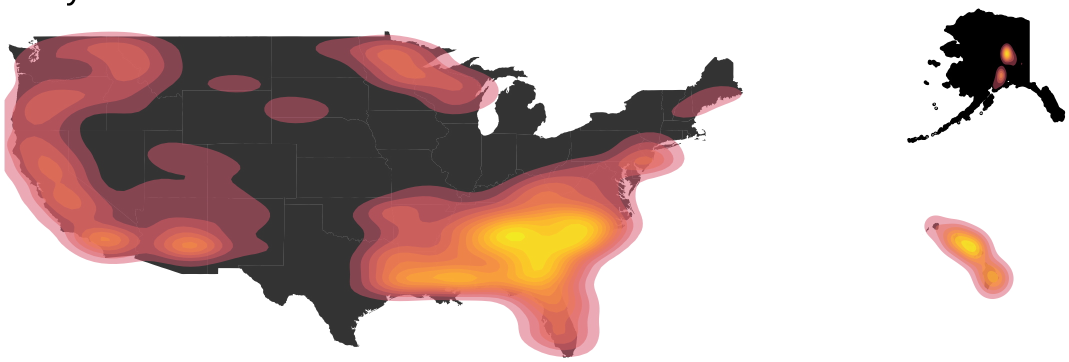
```
#alaska
ak_15<-ggplot(year_2015,aes(y=LATITUDE, x=LONGITUDE))+
  geom_sf(data=us_ak$geometry,inherit.aes = FALSE,color="black",fill="black")+
  stat_density2d(aes(fill = ..level..), alpha = .5,geom = "polygon")+
  scale_fill_viridis_c(alpha=0.3,begin=.55,option="plasma")+
  xlim(-180,-130)+ylim(51,71)+theme_void()+theme(legend.position = 'none')
#lower 48
l48_15<-kernel_density_map(year_2015,map_data("state"),"Density of fires in 2015")
#arrange
panel_2015<-arrange_states(ak_15,hi_15,l48_15)
```

Now we will present the graphs on top of one another to allow for direct comparison

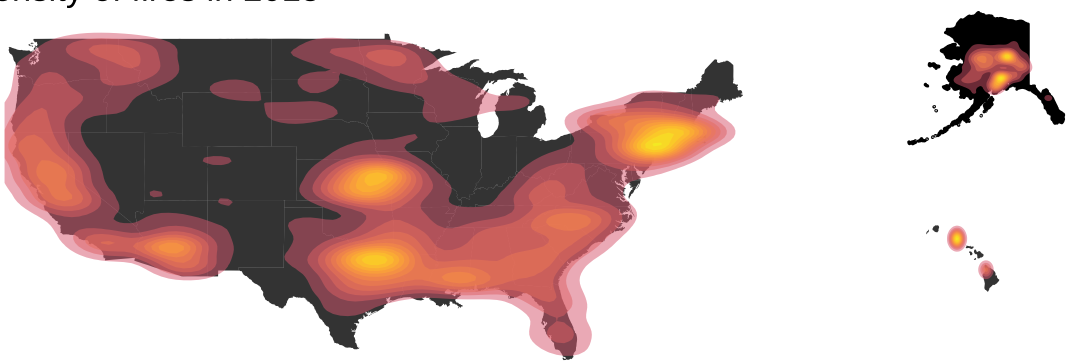
```
(panel_1998|panel_2015)+plot_layout(nrow=2)
```

```
## Warning: Removed 468 rows containing non-finite values (stat_density2d).

## Warning: Removed 67962 rows containing non-finite values (stat_density2d).

## Warning: Removed 68344 rows containing non-finite values (stat_density2d).

## Warning: Removed 808 rows containing non-finite values (stat_density2d).

## Warning: Removed 73720 rows containing non-finite values (stat_density2d).

## Warning: Removed 74487 rows containing non-finite values (stat_density2d).
```

Density of fires in 1998



Density of fires in 2015



Lower 48 states:

1998: The lower 48 states have their densest area in the south and largely match the shape of the map of all years of fires.

2015: The fires have more density centers and cover a larger portion of the map, especially in the Midwest and New England.

Alaska:

1998: The fires are centered around the urban centers of Alaska, just as they are in the map of all fire years.

2015: The fires are just as heavily centered in the urban centers, but have also spread to much of the area surrounding them in the middle of the state.

Hawaii:

1998: The fires seem to encompass the entire state. This either could be due to an unseemly amount of fires in Hawaii that year, or that there was not enough data for the density map to find its center. Either way Hawaii is too small to significantly effect the fire trends of the whole country.

2015: The fires are centered on the same island as the map with all of the fire years, with additional fires on the big island.