Data Exploration

Alex Salce

November 8th, 2024

Description of Dataset

For my data exploration, I wanted to find a data set that would serve as a foundation to allow for investigations of questions different analysis techniques that we have learned in the course. Based on what we have learned so far, I wanted to find a geostatistical data set that I had questions about, but also something that I had at least some kind of intuitive grasp on.

I grew up in Flagstaff, AZ, a densely forested part of the state. In general, life in Flagstaff is fairly worry-free, aside from the ever-present threat of wildfires in the area. Being curious about what kind of wildfire data was available turned out to be fruitful for geostatistical data. In my search, I managed to find a Wildland Fire Incident Location data set, hosted via ArcGIS Open data, through the National Interagency Fire Center website.

https://data-nifc.opendata.arcgis.com/datasets/b4402f7887ca4ea9a6189443f220ef28_0/explore?location=-0.000000%2C0.000000%2C1.38

This data set contains all wildfire data in the continental US reported to the IRWIN system. It contains a wealth of geostatistical data including spatial coordinates of fire origin, dates and times of incidence, total acreage of the fire, among many other factors that may be of interest.

Wildfire incidence data fits the mold of geostatistical data, and on its own can be cut in many interesting ways. But, I also believe it has potential to be combined with other data sets to approach some questions that I have some intuition about.

For the exploration I also used the `rFIA` package (thanks for the inspiration, Dani!) to capture abundance data that may be useful for the project. Additionally, data like roads may help to analyze how certain fires end up getting out of control/becoming large.

In general, this is a data set that I think leaves the door open for lots of different questions and analyses.

Exploratory Data Analysis

In this section, I put together some visualizations that were interesting to me and to serve as proof-of-concept for combining some data.

```
library(dplyr)
library(readr)
library(tigris)
library(sf)
library(rFIA)
library(openmeteo)
library(gganimate)
library(ggplot2)
library(mapview)
library(tidyverse)
library(leaflet)
```

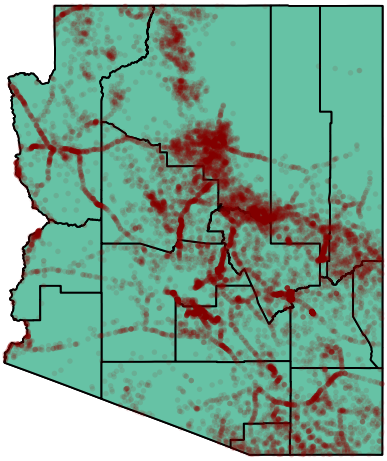
```r
library(htmltools)

# load datasets
load("./R_data/az_fia.RData")
load("./R_data/az_fia_density.RData")
load("./R_data/az_roads_cc_sf.RData")
```

```r
path <- "./wildland_fire_incident_locations/WFIGS_Incident_Locations_-265102656568979733.csv"
wfigs_incident_location_data <- read_csv(path, show_col_types = FALSE)

#convert date data to date objects

wfigs_incident_location_data <- wfigs_incident_location_data %>%
  mutate_at(vars(FireDiscoveryDateTime, ContainmentDateTime, ControlDateTime, CreatedOnDateTime_dt, Modi

wfigs_incident_location_data <- wfigs_incident_location_data %>%
  filter(!is.na(InitialLatitude) & !is.na(InitialLongitude))

wfigs_az <- wfigs_incident_location_data %>% filter(POOState=="US-AZ")
# wfigs_ca <- wfigs_incident_location_data %>% filter(POOState=="US-CA")


rm(wfigs_incident_location_data)
```

```r
wfigs_az <- wfigs_az %>% filter(!is.na(IncidentSize))

wfigs_az_sf <- st_as_sf(wfigs_az, coords = c("InitialLongitude", "InitialLatitude"))

arizona_sf <- states() %>% filter_state("arizona")

az_counties_sf <- counties(state = "AZ", cb = TRUE)

# data saved, long load
# az_roads_cc_sf <- roads(state = "AZ", county = "coconino")

st_crs(wfigs_az_sf) <- st_crs(arizona_sf)
st_crs(az_counties_sf) <- st_crs(arizona_sf)
st_crs(az_roads_cc_sf) <- st_crs(arizona_sf)

#there were some coordinates that were reported outside of the state
wfigs_az_sf <- st_intersection(wfigs_az_sf, arizona_sf)
```

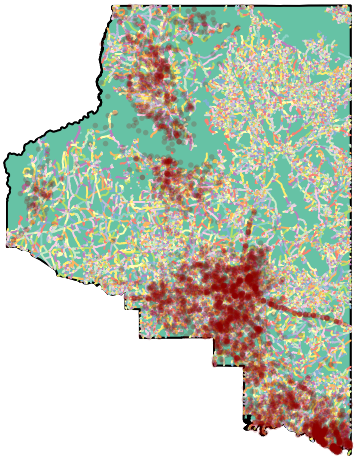
```r
plot(arizona_sf$geometry, main = "Wildfires in Arizona 2014-2024")
plot(az_counties_sf, add = T)
plot(wfigs_az_sf$geometry, pch = 16, cex = 0.4, col = scales::alpha("darkred",
    0.1), add = T)
```

Wildfires in Arizona 2014−2024



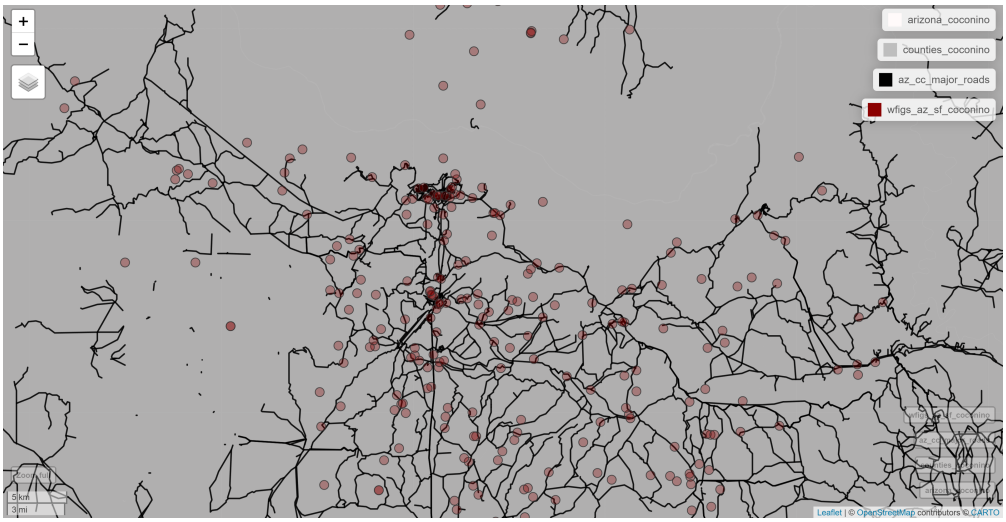
```r
# Extract Coconino County
coconino <- az_counties_sf %>%
    filter(NAME == "Coconino")

# Crop Arizona and counties to the buffered area
arizona_coconino <- st_intersection(arizona_sf, coconino)
counties_coconino <- st_intersection(az_counties_sf, coconino)
wfigs_az_sf_coconino <- st_intersection(wfigs_az_sf, coconino)

az_cc_major_roads <- az_roads_cc_sf %>%
    filter(MTFCC %in% c("S1100", "S1200"))  # Primary and secondary roads

az_cc_minor_roads <- az_roads_cc_sf %>%
    filter(MTFCC %in% c("S1400", "S1500"))  # Minor roads



plot(arizona_coconino$geometry, main = "Wildfires in Coconino County 2014-2024")
plot(counties_coconino, add = T)
plot(az_cc_major_roads, add = T)
plot(az_cc_minor_roads, alpha = 0.2, add = T)
plot(wfigs_az_sf_coconino$geometry, pch = 16, cex = 0.4, col = scales::alpha("darkred",
    0.2), add = T)
```

Wildfires in Coconino County 2014−2024

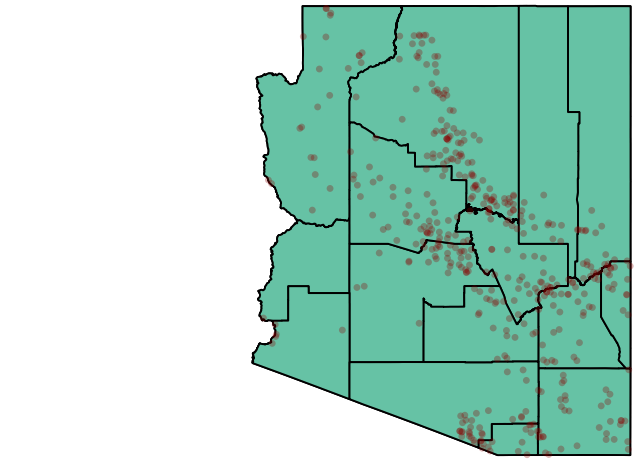


```
# fun interactive version of the map:

# mapview(arizona_coconino, col.regions = 'snow') +
# mapview(counties_coconino, col.regions = 'gray')+
# mapview(az_cc_major_roads, color = 'black', alpha = 0.8) +
# mapview(az_cc_minor_roads, color = 'darkgray', alpha = 0.3) +
# mapview(wfigs_az_sf_coconino, col. = 'darkred', alpha.regions =
# 0.3)
```

```
knitr::include_graphics("mapview_roads_az_cc.png")
```

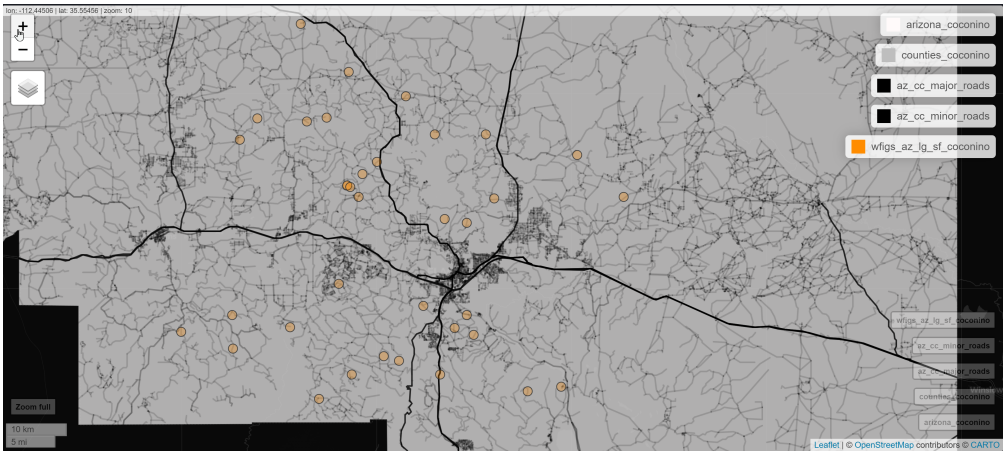


```r
wfigs_az_lg_sf <- wfigs_az_sf %>%
    filter(IncidentSize >= 1000)

plot(arizona_sf$geometry, main = "Large Wildfires (>1000 acres) in Arizona 2014-2024")
plot(az_counties_sf, add = T)
plot(wfigs_az_lg_sf$geometry, pch = 16, cex = 0.5, col = scales::alpha("darkred",
    0.3), add = T)
```




```r
wfigs_az_lg_sf_coconino <- st_intersection(wfigs_az_lg_sf, coconino)

# fun interactive version of the map:

# mapview(arizona_coconino, col.regions = 'snow') +
# mapview(counties_coconino, col.regions = 'gray')+
# mapview(az_cc_major_roads, color = 'black', alpha = 0.8) +
# mapview(az_cc_minor_roads, color = 'black', alpha = 0.15) +
# mapview(wfigs_az_lg_sf_coconino, col. = 'darkorange', alpha.regions
# = 0.3)
```

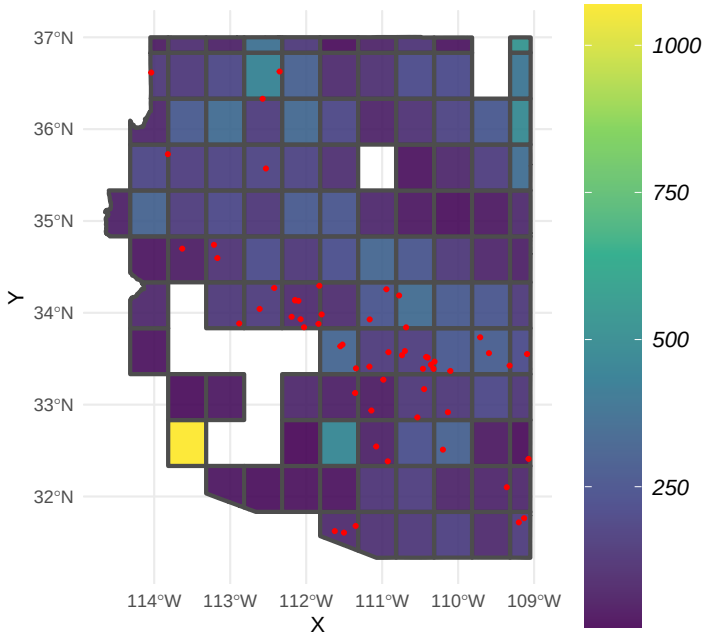
```r
knitr::include_graphics("mapview_roads_az_cc_lgfires.png")
```




```
# this chunk is used to generate az_fia_density. it takes a while to load, so
# I just saved the environment object and will load it for the exploration.

# # Create square grid
# az_grid <- st_make_grid(arizona_sf, cellsize = 0.5, square = TRUE)
#
# # Convert to sf object and add grid ID
# az_grid_sf <- st_sf(az_grid) %>% mutate(grid_id = 1:length(lengths(az_grid)))
#
# # Clip grid to Arizona boundary
# az_grid_clipped <- st_intersection(az_grid_sf, arizona_sf)
# az_grid_clipped <- az_grid_clipped %>% rename(geometry = az_grid)
#
# az_fia_density  <- tpa(az_fia, polys = az_grid_clipped, returnSpatial = TRUE)
#
#
#
# az_fia_density_plot <- plotFIA(az_fia_density, TPA, min.year = 2024)
```

```
year_filter <- 2020

# year filtering
az_fia_density_yr <- az_fia_density %>% filter(YEAR == year_filter)
wfigs_az_yr <- wfigs_az_lg_sf %>% filter(year(CreatedOnDateTime_dt)==year_filter)

# take the intersection of the wildfire data and available density data
wfigs_az_yr <- st_intersection(wfigs_az_yr, az_fia_density_yr)


az_fia_density_plot_yr <- plotFIA(az_fia_density_yr, TPA)



coords <- st_coordinates(wfigs_az_yr$geometry)
coords <- as.data.frame(coords)


wf_density_plot <- az_fia_density_plot_yr +
  geom_point(coords, mapping = aes(x = X, y = Y), color = "red", size=0.75)
```

```
wf_density_plot
```



Overview of the Dataset

For this exploration, I am only looking at Arizona, and more specifically Coconino County in some cases, since that is where home is and I have some familiarity with it.

There are some interesting patterns that seem to generally track with my intuition. One question that was interesting to me before ever looking at the data was whether there was a connection with a fire getting out of control and whether it started somewhere accessible, i.e. near a road of some kind. Having a look at fires that ended up over 1000 acres, it does appear that most start somewhat "off the beaten path", away from major roads and maybe only off of smaller service roads. You can also see some pretty obvious patterns of wildfires that start right next to major roads, particularly some of the interstates, although they seem to get put out before becoming large.

I also imported tree abundance data from the `rFIA` package to look at a possible connection between size, point pattern intensity, or other insights based on tree abundance for certain regions. The data plotted from 2020 shows a proof-of-concept that I am able to get at least some kind of estimate in a grid for approximate abundance. There are further breakdowns of tree types, etc. in the data that may be interesting to study in terms of wildfire prevalence and chance of becoming a large fire.

There are other datasets that I am seeking that I decided not to include here yet since they were more ambitious, and I wanted for our group to come to a consensus before we delve any deeper, but I believe there are lots of interesting approaches to this data other than my initial thoughts that we could explore, especially combining other data sets.

Challenges

There will be some challenges dealing with this data that I have already encountered. First, getting an understanding of all of the available fields in the wildfire incidence data set will be time consuming, since there are 97 variables that are not all populated. Deciding what is useable/important may be tricky. Additionally, as is often the case with data sets from the "wild", there is a fair amount of data cleanup that has to happen, which was definitely the case here.

The rFIA data is incomplete, and there are certain regions of Arizona that did not have data available. You'll notice, for one, that I had to make a grid to get any kind of resolution on the abundance data to give any better approximations for regions where there were fires. There were also fire data points that were in the no-data regions, for which I had to remove some wildfire data using `st_intersection`. There may be some challenges getting some reasonable data fits using this data.

I was having a hard time getting historical climate data spatially. For example, if it was particularly windy the day the fire started, it would be worth including in any models we might be interested in fitting. One of the factors I would like to include is previous annual total precipitation for a region, which I found to be more difficult than anticipated, so I will save for later if we go this route.

Lastly, this is a TON of data, and it's tricky to negotiate how big the files are and how long it takes to load some of this. I managed to save a few environment objects that took a long time to get created as `*.RData` files, but overall there is a lot of data to filter and work with.

Depending on the questions we come up with if we pursue this data, there will certainly be more challenges formatting data, acquiring data, and getting an analysis set up.