

All the World's A Stage: Elements to visualize multivariate spatiotemporal data in R

R Conference New York
August 15, 2020

Emily Dodwell – Principal Inventive Scientist, AT&T Big Data and AI Research

in collaboration with

Di Cook – Professor of Business Analytics, Econometrics and Business Statistics

Weihao Li – Honours Research Student



Australia Bushfire Season 2019-2020

The New York Times

'It's an Atomic Bomb': Australia Deploys Military as Fires Spread

With more than a month still to go in the fire season, Australia has announced a large-scale use of military assets, a decision last seen since World War II.

BBC NEWS

Home | Video | World | US & Canada | UK | Business | Tech | Science | World | Africa | Asia | Australia | Europe | Latin America | Middle East

Australia fires: Storms wreak damage but bushfires 'far from over'

20 January 2020

Matthew Abbott for The New York Times

By Livia Albeck-Ripka, Isabella Kwai, Thomas Fuller and Jamie Tarabay

Published Jan. 4, 2020 Updated Jan. 5, 2020

f t b 480

A firefighter in a yellow uniform and helmet sprays water on a burning forest. The background shows a dense forest engulfed in flames.

ALJAZEERA News Documentaries Shows Opinion More Live

NEWS / AUSTRALIA

Melbourne chokes amid warning gigantic fires could become routine

Scientists say severity of fires in Australia and elsewhere will only be reduced with efforts to reduce global warming.

13 Jan 2020 f t

Support The Guardian Available for everyone, funded by readers Contribute → Subscribe →

Sign in The Guardian

News Opinion Sport Culture Lifestyle

Environment ▶ Climate change Wildlife Energy Pollution More

Bushfires

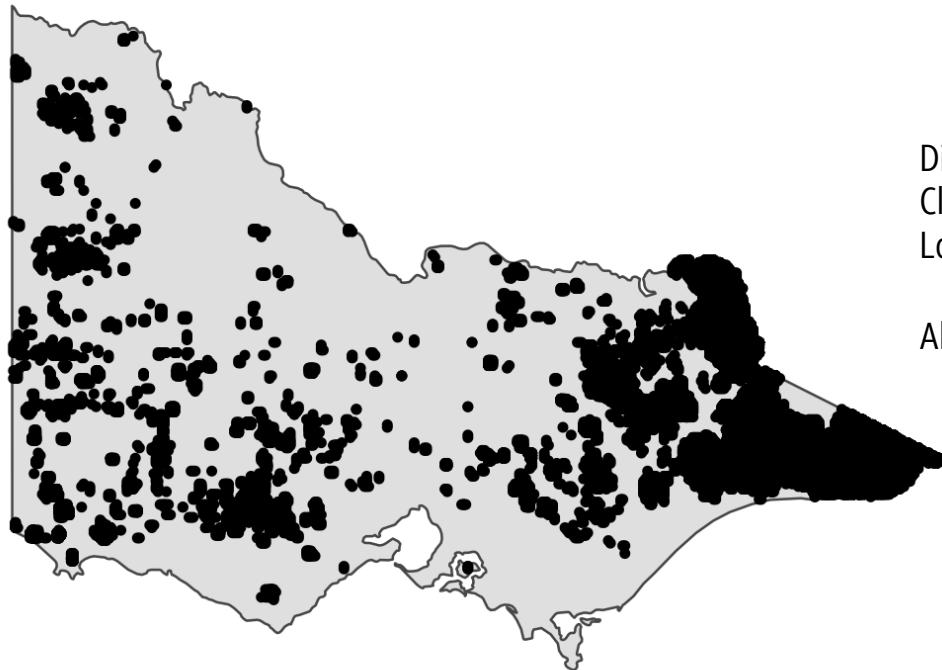
This article is more than 6 months old

A billion animals: some of the species most at risk from Australia's bushfire crisis

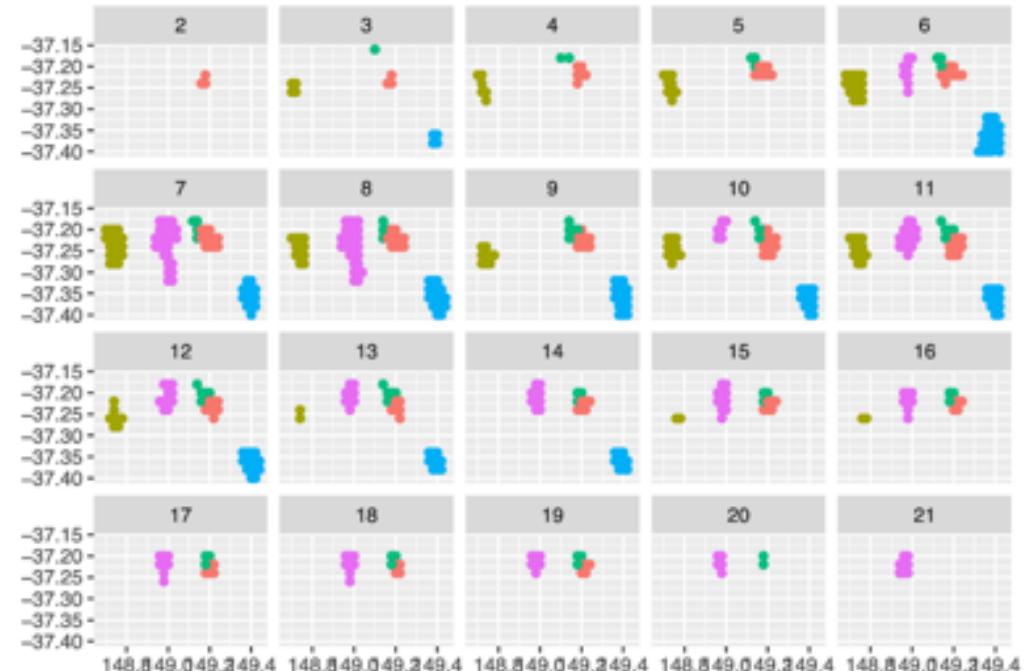
Fires take an enormous toll on wildlife, with huge numbers of mammals, birds, reptiles and insects killed

Himawari-8 Hotspot Data¹ Processing

October 2019-March 2020: 75,936 Hotspots Detected in Victoria (Firepower > 100)



Discretize time
Cluster hotspots = “fire ID”
Log centroid as fire location at t_n
Algorithm in forthcoming paper



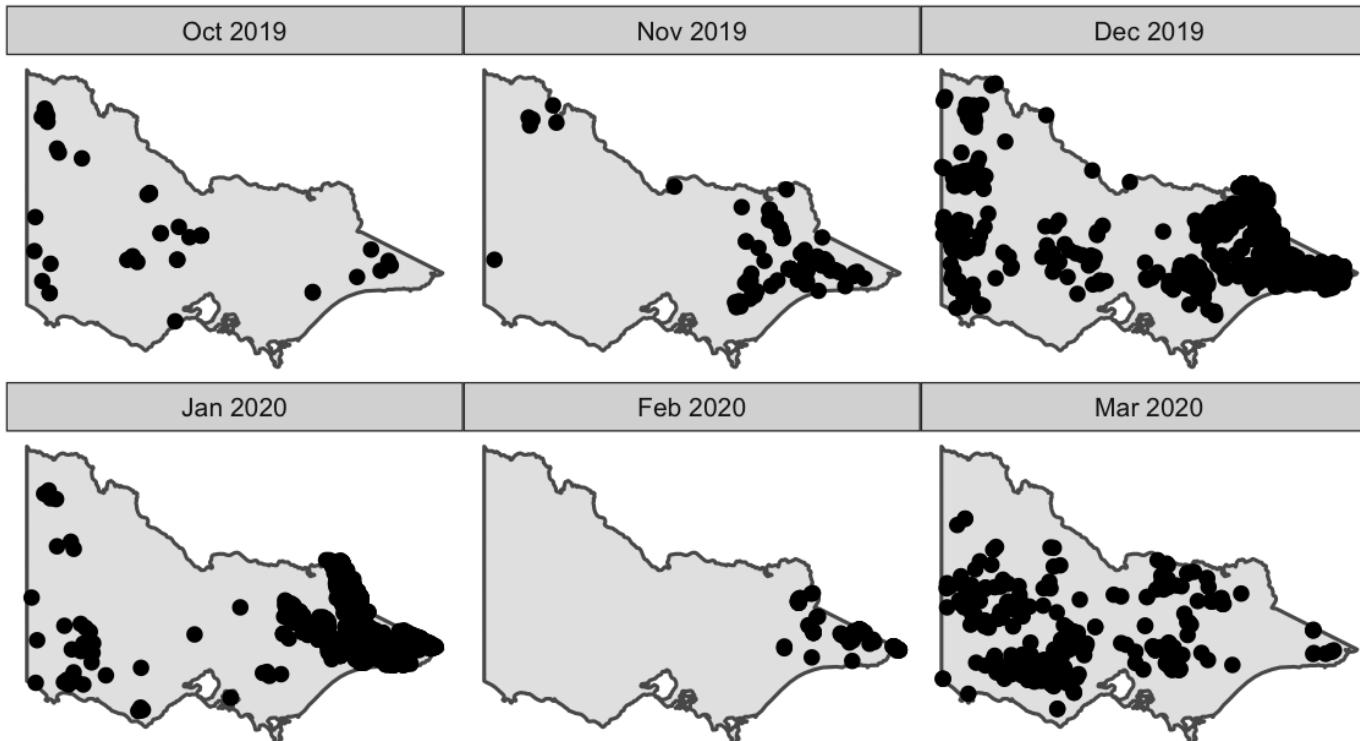
Note: This is a bad map

“The truth behind the ‘misleading’ fire maps that have gone viral during Australia’s bushfire crisis”
by Peta Fuller (ABC)

¹Data available for download from Japan Aerospace Exploration Agency (JAXA) Himawari Monitor:
https://www.eorc.jaxa.jp/ptree/registration_top.html

Motivation

2,063 Distinct Fires in Victoria (October 2019-March 2020)



How can EDA visualization support project goals?

- Predict ignition locations
- Satellite data to monitor and track movement of fires
- Understand patterns among and between variables that contribute to initial ignition and continued spread -- implications for resource allocation and future assistance

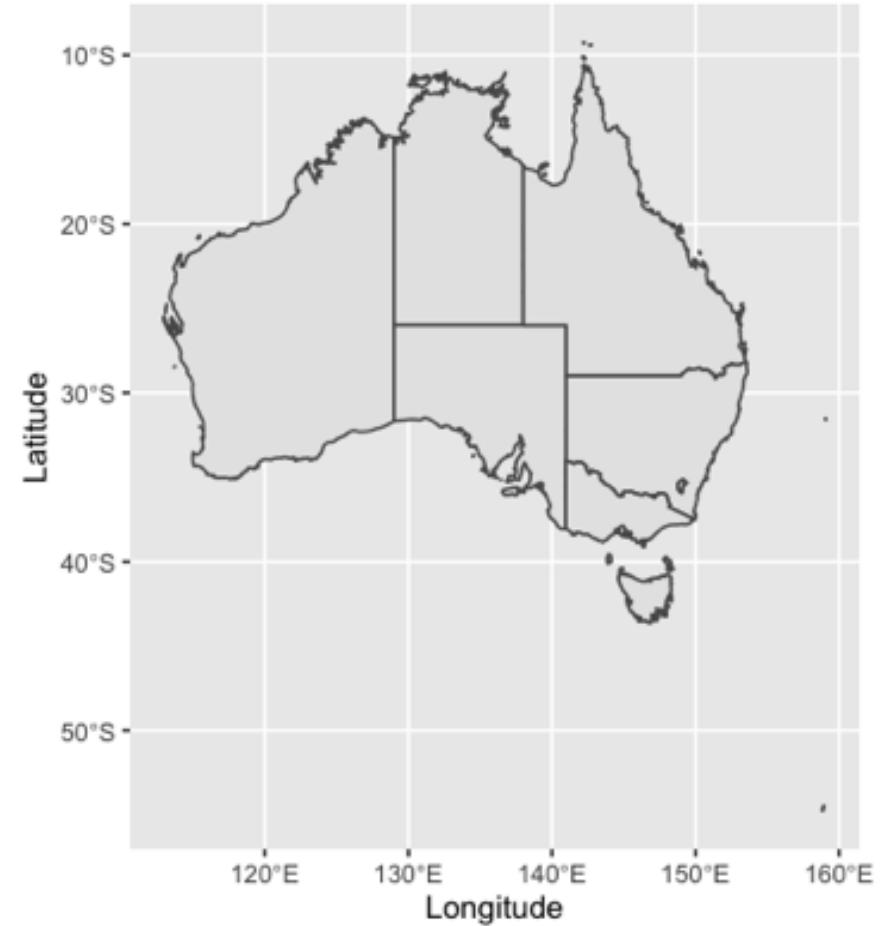
...back up. How did you get Victoria on a map?

```
library(tidyverse)
library(rnaturalearth)

australia <- ne_states(country = "Australia", returnclass = c("sf"))

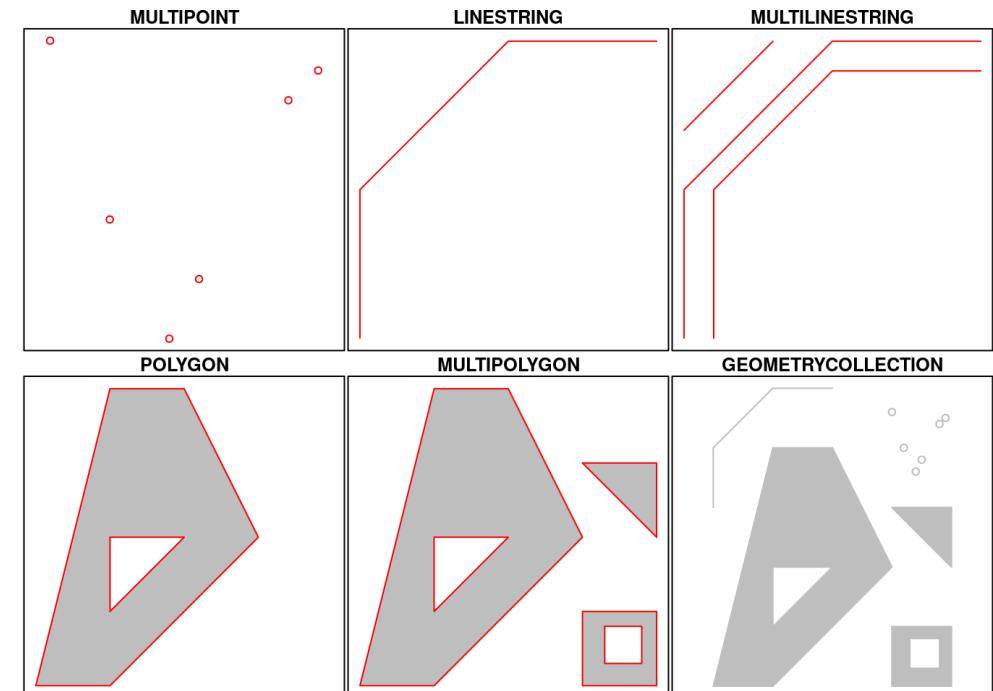
ggplot(australia) +
  geom_sf() +
  xlab("Longitude") + ylab("Latitude")
```

sf denotes that `australia` is a
“simple feature” object



...back up. How did you get Victoria on a map?

```
> australia %>% select(name, geometry)
Simple feature collection with 11 features and 1 field
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: 112.9194 ymin: -54.75042 xmax: 159.1065 ymax: -9.240167
CRS:           +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
First 10 features:
#> #>   name          geometry
#> #> 1744 Western Australia MULTIPOLYGON (((128.1589 -1...
#> #> 1745 Northern Territory MULTIPOLYGON (((129.6325 -1...
#> #> 1759 South Australia MULTIPOLYGON (((134.1956 -3...
#> #> 1760 Queensland MULTIPOLYGON (((138.0002 -1...
#> #> 1761 New South Wales MULTIPOLYGON (((153.5065 -2...
#> #> 1762 Jervis Bay Territory MULTIPOLYGON (((150.7038 -3...
#> #> 1763 Victoria MULTIPOLYGON (((149.9721 -3...
#> #> 2900 Tasmania MULTIPOLYGON (((147.9985 -4...
#> #> 2922 Macquarie Island MULTIPOLYGON (((158.8934 -5...
#> #> 2926 Lord Howe Island MULTIPOLYGON (((159.0444 -3...
```



Source: sf package vignette "1. Simple Features for R"

We got Victoria on a map!

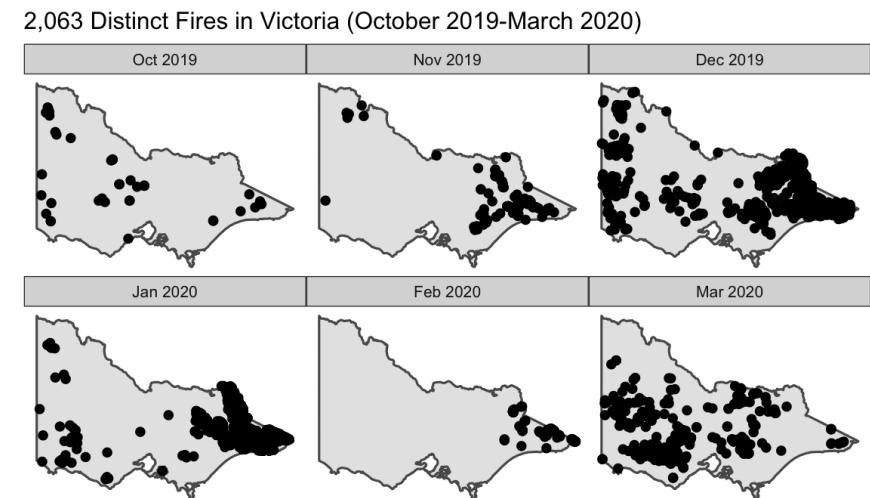
```
vic <- australia %>% filter(name == "Victoria")
```

```
ggplot(vic) +  
  geom_sf() +  
  # xlab("Longitude") + ylab("Latitude") +  
  ggthemes::theme_map()
```

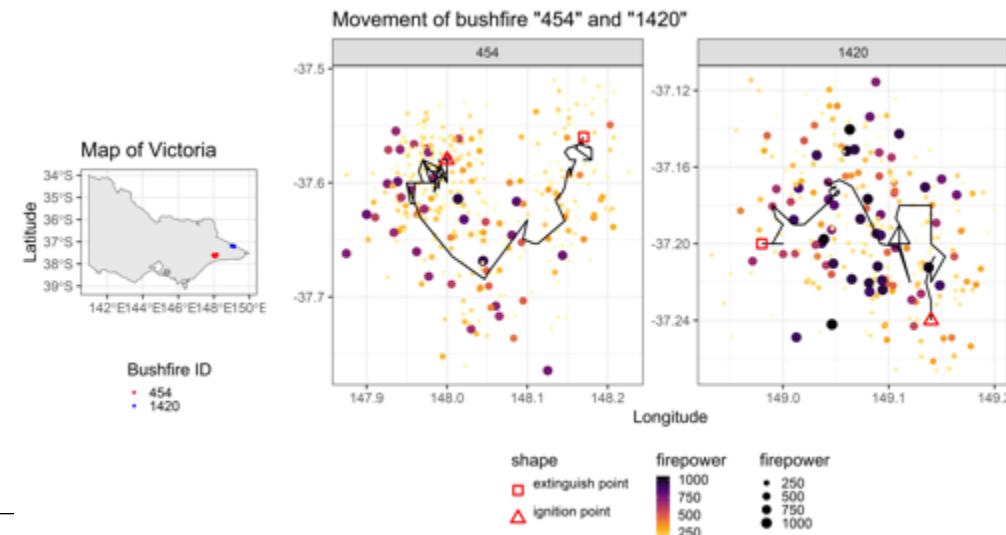


Questions of interest will inform granularity, e.g.:

Spatial distribution
of fires over time:



Trajectory of
individual fire(s):



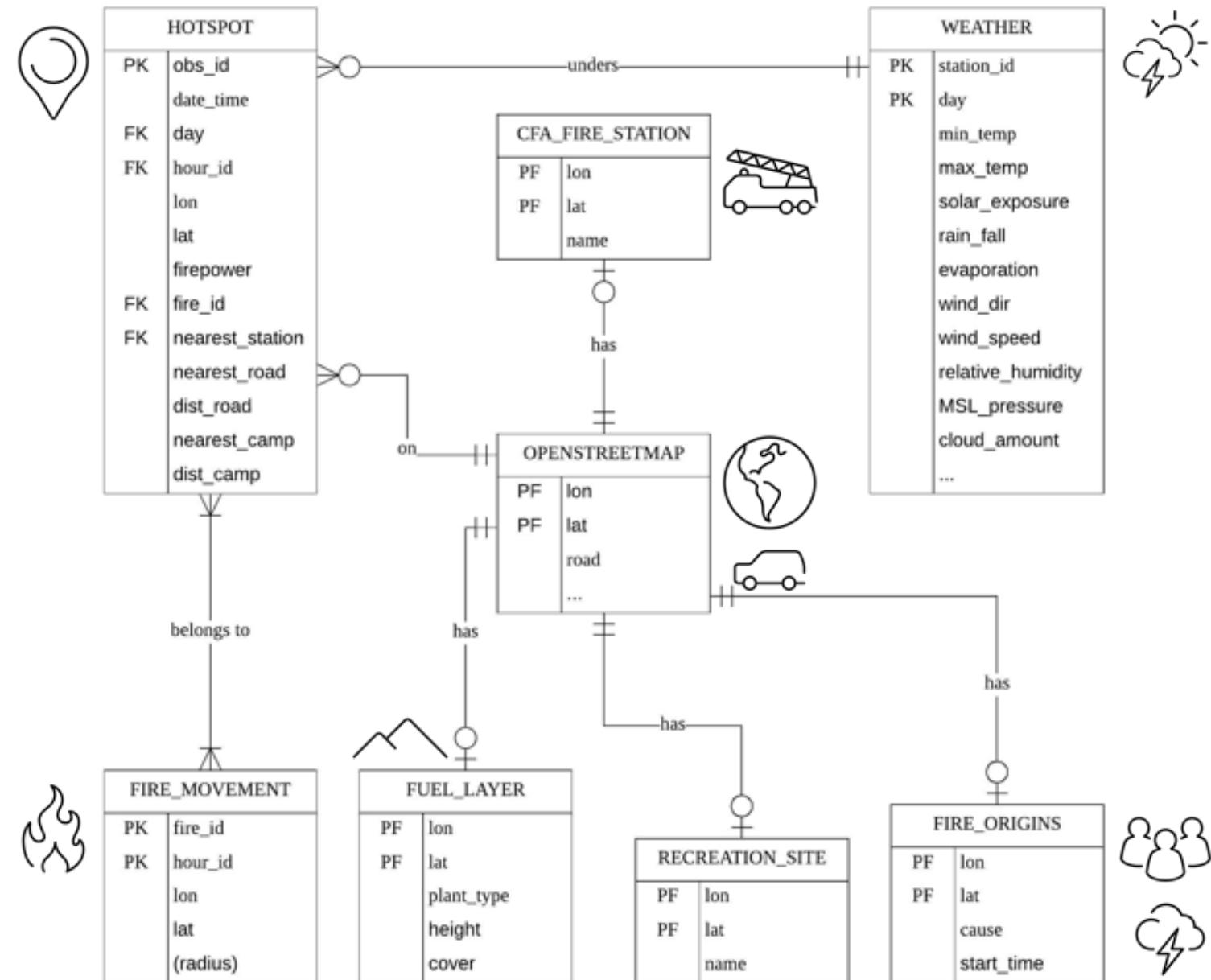
Data Schematic

Additional Layers:

What other useful information can be included to predict fire ignition?

Sources:

- Himawari-8 Hotspot Data
- Bureau of Meteorology & CSIRO
- OpenStreetMap
- Victoria Department of Environment, Land, Water & Planning
- Australian Bureau of Agricultural and Resource Economics and Sciences



Visualization Problem Statement

Data Structure(s)

Original data: location of fires (by hour)

fire_id	hour	lat	lon	firepower	dist_cfa	nearest_cfa	dist_road	...
1	1							...
1	2							...

Characterize each fire with summary statistics, e.g.:

fire_id	t ₀	lat _{t₀}	lon _{t₀}	burn_time	dist_traveled	nearest_cfa_t ₀	max_firepower	...
1	1							...
2	23							...

Summarize by associated geographic entity (cfa at t₀), e.g.:

cfa	week	lat	lon	total_fires	avg_burn_time	avg_dist_fire_t ₀	...
							...
							...

Longer you spend thinking about variables →

more questions you ask →

of potential plots → ∞

Keep in Mind:

- Who's the audience?
- What do you want them to know?
- Appropriate graphics?
- Helpful user interaction?

Visualization Problem Statement (Generalized)

Data Structure

ID	t	s_1	s_2	v_1	\dots	v_k
					\dots	
					\dots	

ID : key

t : time (consistent steps or variable)

(s_1, s_2) : location, i.e. longitude and latitude

(v_1, \dots, v_k) : variables associated with ID at time t and location (s_1, s_2) – discrete, continuous, categorical

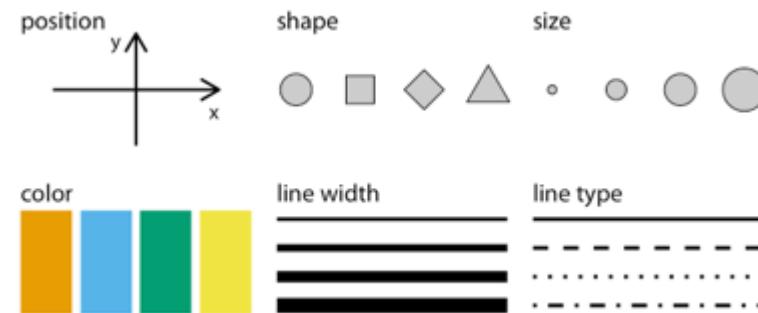
Analytical Tasks

- Where are events happening in space and time?
- Is there an association between $v_k, (s_1, s_2), t$?

Components of Tasks

- Map v_k to color of point/line on a map
- Map v_k to y-axis against time, separately by ID
- Summarize v_k by ID and t

Data Visualization Aesthetics



Source:
Fundamentals of Data Visualization
by Claus Wilke

[https://serialmentor.com/
dataviz/aesthetic-
mapping.html](https://serialmentor.com/dataviz/aesthetic-mapping.html)

Interacting with your data (Part 1): tmap package

```
library(tmap)
library(sf)

current.mode <- tmap_mode("view")

# Convert data.frame of initial ignition locations to sf object
f <- st_as_sf(ignitions, coords = c("lon", "lat"))

tm_basemap(leaflet::providers$OpenStreetMap.Mapnik) +
  tm_shape(vic, bbox = "Victoria") +
  tm_borders() +
  tm_shape(f) +
  tm_bubbles(size = "active_hrs", scale = .4,
             popup.vars = c("Start time" = "date", "Burn time"="active_hrs")) +
  tm_facets(by = "month", nrow = 2, free.coords = FALSE)
```

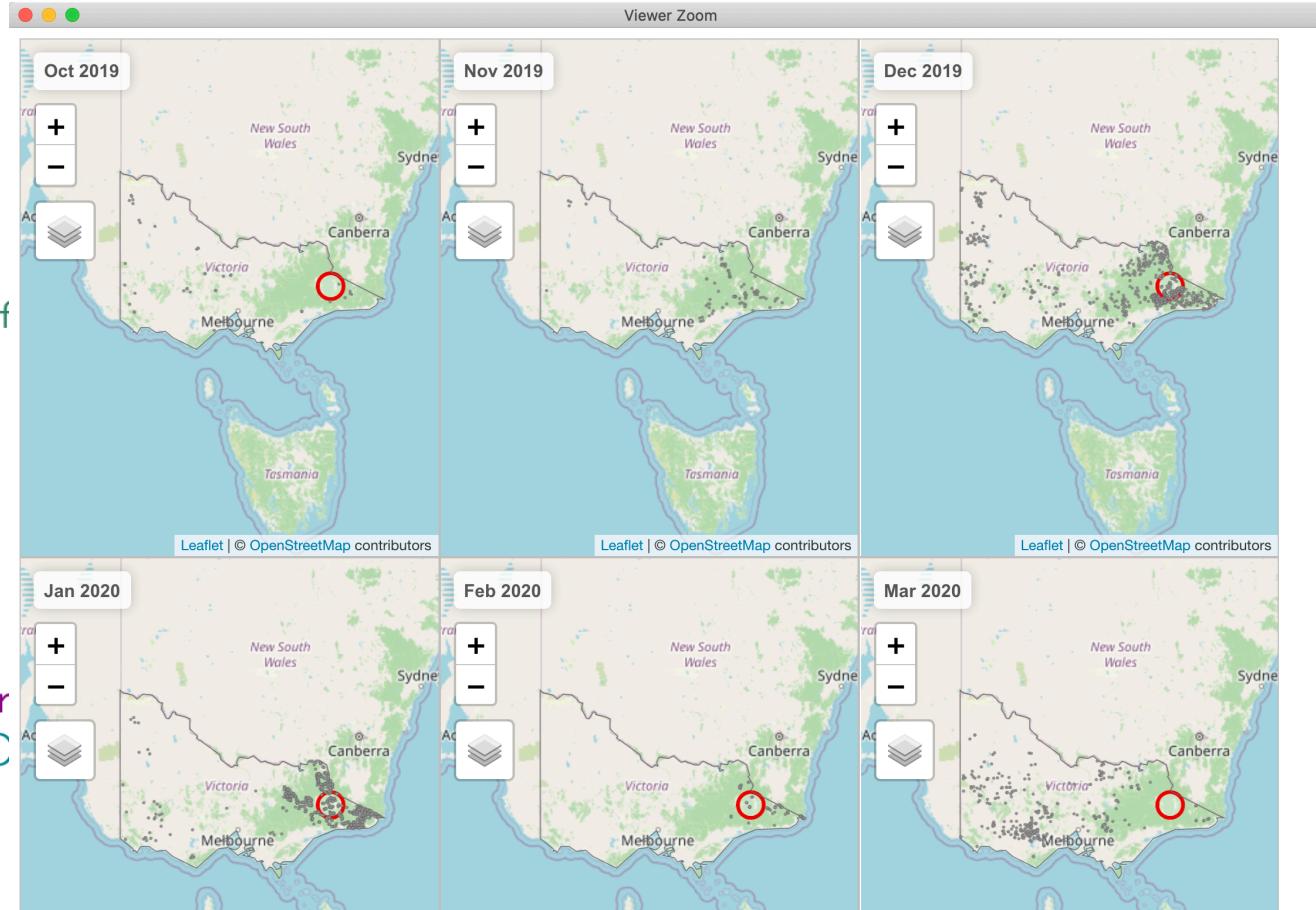
Interacting with your data (Part 1): tmap package

```
library(tmap)
library(sf)

current.mode <- tmap_mode("view")

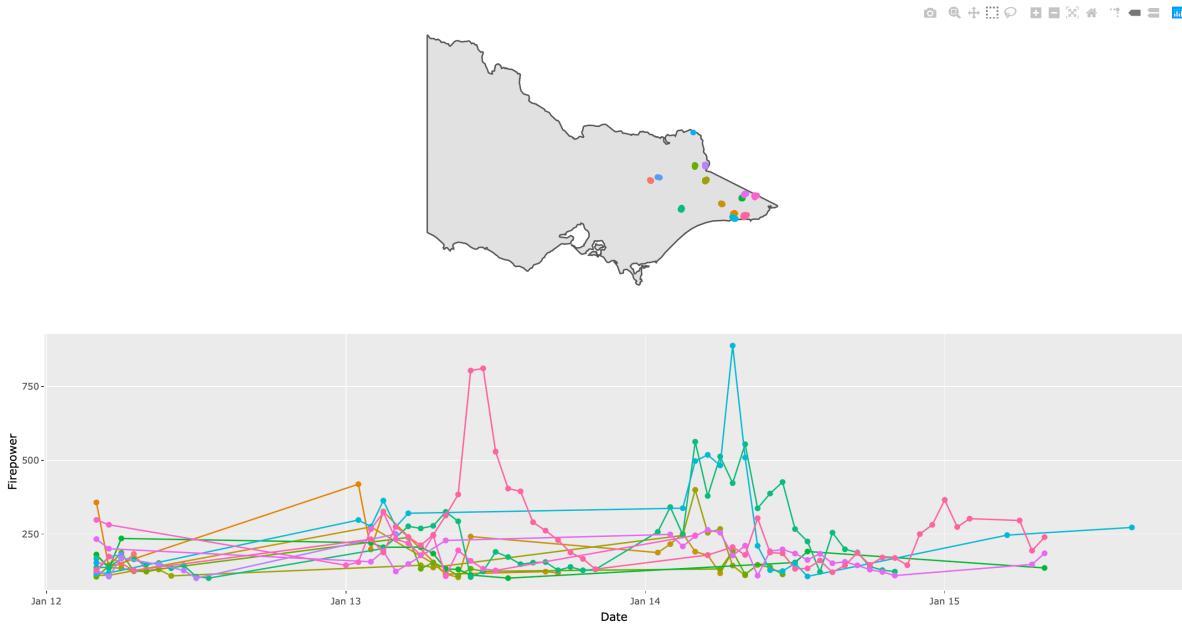
# Convert data.frame of initial ignition locations to sf
f <- st_as_sf(ignitions, coords = c("lon", "lat"))

tm_basemap(leaflet::providers$OpenStreetMap.Mapnik) +
  tm_shape(vic, bbox = "Victoria") +
  tm_borders() +
  tm_shape(f) +
  tm_bubbles(size = "active_hrs", scale = .4,
             popup.vars = c("Start time" = "date", "Burr"))
  tm_facets(by = "month", nrow = 2, free.coords = FALSE)
```



Interacting with your data (Part 2): plotly for linked plots

Victoria Bushfires (Started 2020-01-12)



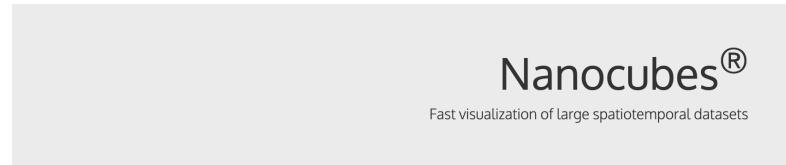
```
server <- function(input, output) {  
  sd <- highlight_key(fires, ~fire_id)  
  
  output$vicPlot <- renderPlotly({  
    p <- ggplot(vic) +  
      geom_sf() +  
      ggthemes::theme_map() +  
      geom_point(aes(x = lon, y = lat, color = fire_id), sd)  
  
    ggplotly(p) %>%  
      highlight("plotly_selected", off = "plotly_doubleclick") %>%  
      layout(showlegend = FALSE)  
  })  
  
  output$firepowerPlot <- renderPlotly({  
    q <- ggplot(sd, aes(x = date, y = med_firepower, group = fire_id)) +  
      geom_line(aes(color = fire_id)) + geom_point(aes(color = fire_id)) +  
      xlab("Date") +  
      ylab("Firepower")  
  
    ggplotly(q) %>%  
      highlight("plotly_selected") %>%  
      layout(showlegend = FALSE)  
  })  
}
```

Resources: Learning more and extending visualization

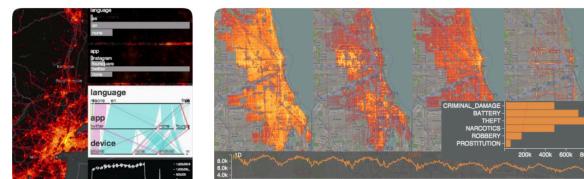
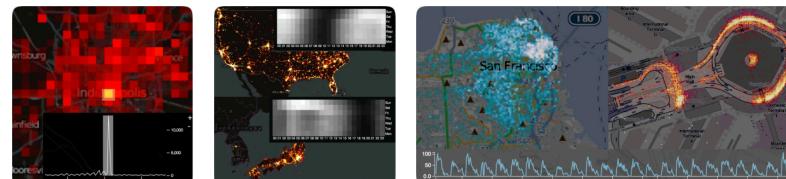
Talks & books I found helpful

- May 21, 2020: Mappy Hour with R-Ladies NYC
"Uncharted Territory: An Introduction to Making Clear and Accurate Maps" by Ayanthi Gunawardana
- "How to Be a Resilient Geospatial R useR: First Steps, Learning More, and the Spatial R-Ladies Community" by Angela Li
- *Geocomputation with R* by Robin Lovelace, Jakub Nowosad, Jannes Muenchow:
<https://geocompr.robinlovelace.net/>
- Documentation for packages `sf`, `tmap`, `plotly`

Solutions outside of R (scalability)



The Nanocubes® technology provides you with real-time visualization of large datasets. Slice and dice your data with respect to space, time, or some of your data attributes, and view the results in real-time on a web browser over heatmaps, bar charts, and histograms. We've used it for tens of billions of data points: maybe you can push it even farther!



A screenshot of the Kepler.gl website. The header includes links for "USER GUIDE", "DOCUMENTATION", "GITHUB", and "SUPPORT POLICY". The main content features the text "Make an impact with your location data" and "Kepler.gl is a powerful open source geospatial analysis tool for large-scale data sets." Below this is a "GET STARTED" button and a "GITHUB" button. The main area shows a map with a 3D point cloud visualization and a sidebar with various data selection and analysis tools.

Uber
kepler.gl

Many thanks to:



Di Cook, Monash University



Weihao Li, Monash University



emily@research.att.com



linkedin.com/emdodwell



@emdodwell



github.com/emdodwell