Wildfire Incidence in Arizona

Group 3 | STAT574E Final Project

Raymond Owino, Alex Salce, Matthew Wallace

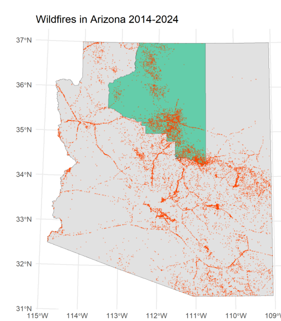
2024-12-11

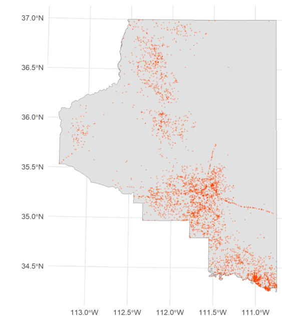
Wildfire Incidence in Arizona

Our project investigates different approaches to spatial statistical modeling of wildfire incidence data in Arizona, with a focus on Coconino County (Northern AZ).

The “Wildfire Incidence” data we will be studying specifies coordinates for a fire’s origin, and its resulting size in acres.

Overall practical goal: modeling wildfire incidence to aid in prediction/assessment of wildfire risks in AZ based on relevant information.





Wildfire Incidence Dataset

Our incidence data uses the **National Interagency Fire Center** Wildland Fire Incident Locations dataset.

This dataset includes point locations and corresponding data for all wildland fires in the United States reported to the IRWIN system, which aggregates wildfire incidence data from around the country.

The dataset has all IRWIN data entries since 2014 (it is still updated daily).

There are over 300K records in this dataset, and just over 18K in Arizona.

Wildfire Incidence Dataset

Refinement of this dataset utilized the following data attributes (of 96 columns) for incidence filtering & covariates.

- `IncidentSize` | Size of the resulting wildfire in acres
- `FireCause` | Human, Natural, Unknown, Undetermined
- `FireDiscoveryDateTime` | Date & time of incident reporting
- `IncidentTypeCategory` | WF (wildfire) or RX (prescribed burn)
- `x` and `y` | Spatial coordinates in lat/lon

Wildfire Incidence Data - Continuous, Fixed Spatial Index

This dataset offered us a unique opportunity to take multiple approaches to spatial statistical modeling. Using the available data attributes, we can treat our response data in two distinct ways

- **Continuous, fixed** - Using `x` and `y` as coordinates and `IncidentSize` as response, we can treat wildfire incidence as fixed, continuous data.

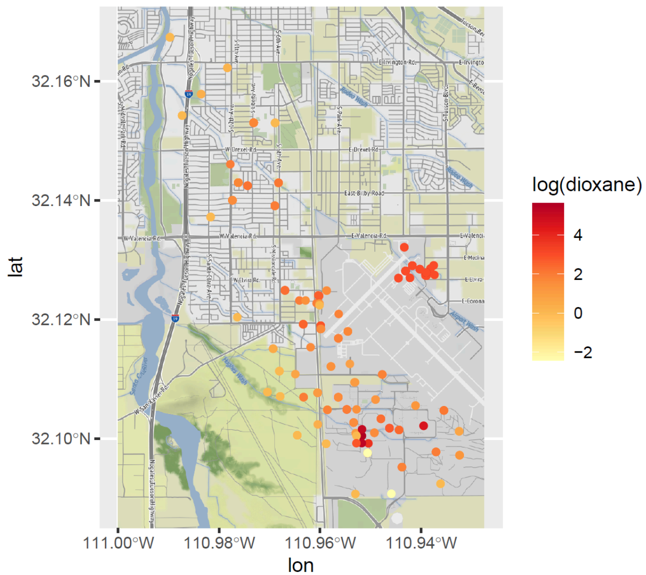
Modeling approach

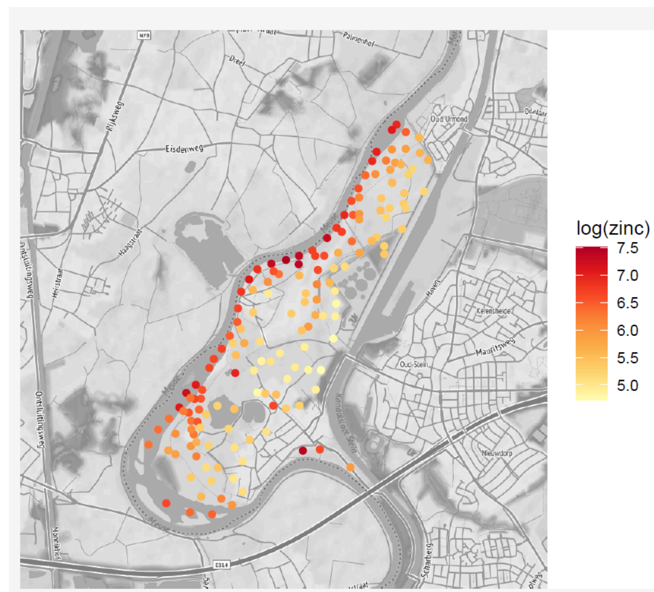
- **Spatial Linear Model** - Module 2, similar to Holland example in *Continuously indexed spatial data (geostatistics)*, or the `dioxane` analysis in HW2.

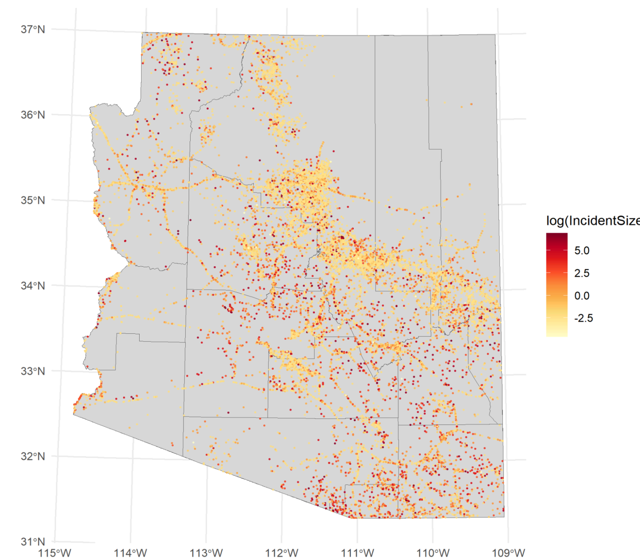
$$\mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{e} \quad \boldsymbol{e} \sim \mathrm{N}(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$$

Wildfire Incidence Data - Continuous, Fixed Spatial Index

Spatial Linear Model Data







Wildfire Incidence Data - Point Process 1

This dataset offered us a unique opportunity to take multiple approaches to spatial statistical modeling. Using the available data attributes, we can treat our response data in two distinct ways

- **Point process** - Using `x` and `y` as coordinates and `IncidentSize` as a *threshold*, we can study "large wildfires" (`IncidentSize` ≥ 1000 acres) as point process data. We can also filter by `FireCause` (Human or Natural).

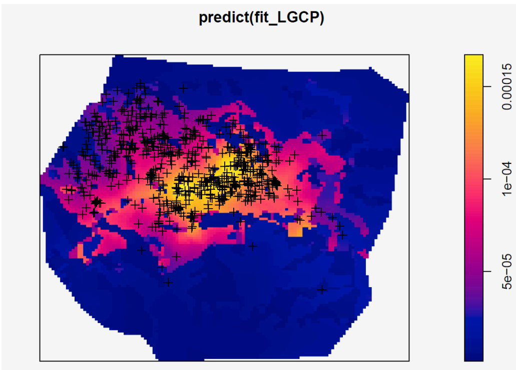
Modeling approach

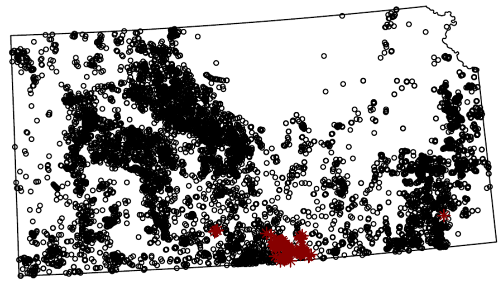
- **Log-Gaussian Cox Process** Module 4, similar to Gorillas LGCP example in *Random spatial index (point pattern)*, or the `earthquakes_um` analysis of Kansas in HW4.

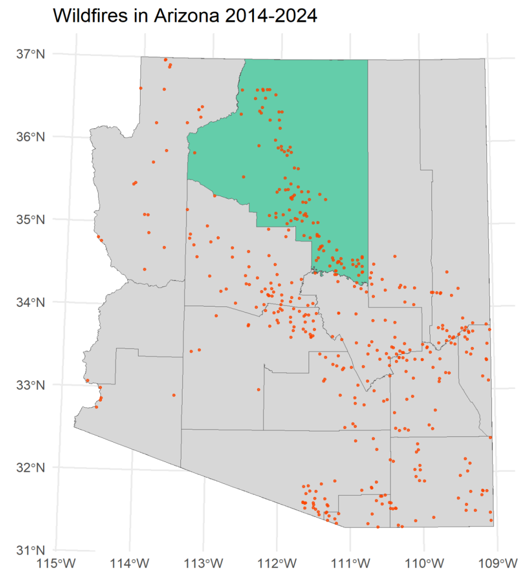
$$\log(\lambda(u)) = \boldsymbol{Z}(u)\boldsymbol{\beta} + e(u), \quad e(u) \sim \mathrm{N}(0, C(\boldsymbol{\theta})), \quad C(u, u') = \sigma^2 e^{-||u-u'||/h}$$

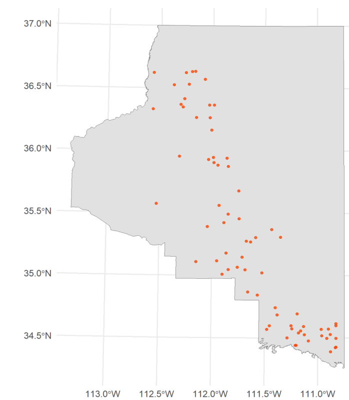
Wildfire Incidence Data - Point Process 1

Log-Gaussian Point Process Model Data









Wildfire Incidence Data - Point Process 2

This dataset offered us a unique opportunity to take multiple approaches to spatial statistical modeling. Using the available data attributes, we can treat our response data in two distinct ways

- **Point process** - Using `x` and `y` as coordinates and `IncidentSize` as a *threshold*, we can study "large wildfires" (`IncidentSize` ≥ 1000 acres) as point process data. We can also filter by `FireCause` (Human or Natural).

Modeling approach

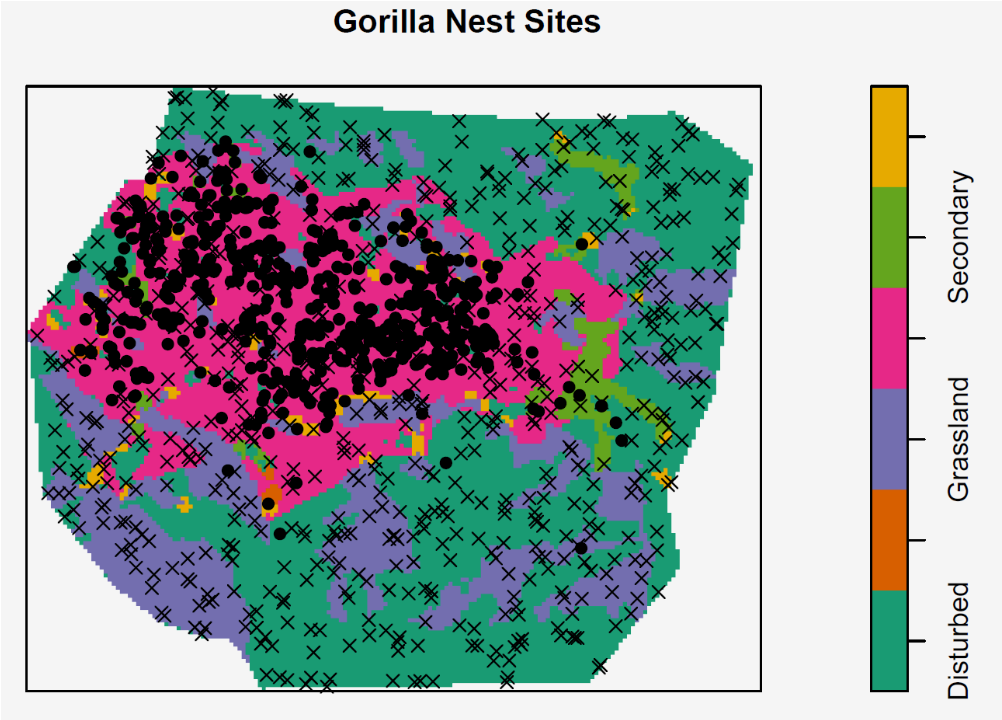
- **Binary GLM Spatial Logistic Regression** - Module 5, similar to Gorillas Logistic Regression GLM example in *Non-Gaussian spatial data*.

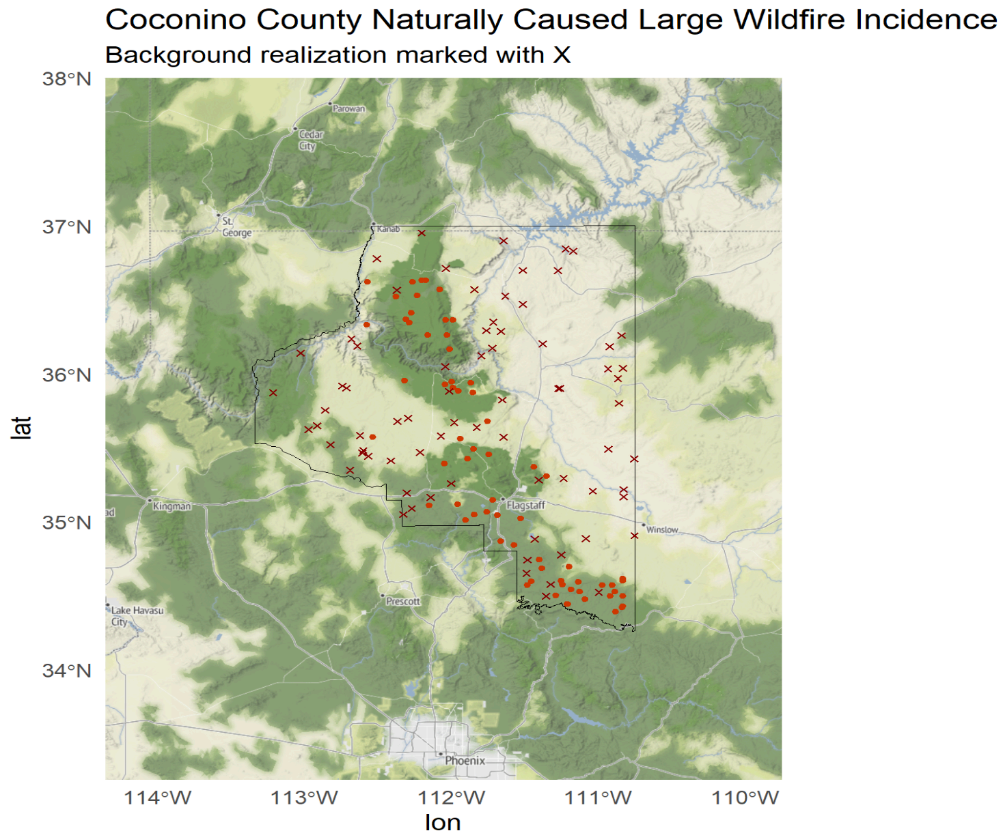
$$\mathrm{logit}(\lambda_1(\boldsymbol{s})) = \mathbf{x}(\boldsymbol{s})^T\boldsymbol{\beta} + e(\boldsymbol{s}) + \log(\lambda_0),$$

$$Y(\boldsymbol{s}) \sim \mathrm{Bern}(p(\boldsymbol{s})), \quad \mathrm{E}[Y(\boldsymbol{s})] = p(\boldsymbol{s}) = \frac{\lambda_1(\boldsymbol{s})}{\lambda_0(\boldsymbol{s}) + \lambda_1(\boldsymbol{s})}$$

Wildfire Incidence Data - Point Process 2

Binary GLM Spatial Logistic Regression Model Data





Research Questions (1)

Does the proximity of a wildfire's origin to a road influence its resulting size?

(a) All wildfires *(b) Proximity to major roads*

Figure 1: Reasearch question 1 framework

Data explorations showed some clear wildfire patterns near roads (the outlines are visible). We wanted to include the distance in meters to the nearest major roads (and "remote" roads) as usable predictors. For each point, the `roads()` function in the `tigris` package to generate `sf` objects for AZ roads, and `st_distance()` function in the `sf` package helped us generate this data to be used as a covariate.

Research Questions (2)

Can environmental or human factors be useful predictors in modeling resulting wildfire size?

Add images like screenshots of GIS here

Figure 2: Research question 2 framework

Raymond to update this slide with environmental covariates, quick summary of what the data you got was and where it came from/units

Research Questions (2)

Can environmental or human factors be useful predictors in modeling resulting wildfire size?

Add images like population density maps
from exploration

Figure 3: Research question 2 framework

Matt to update this slide with population density data details (packages, units, etc)

Research Questions (3)

Are the patterns of human or non human caused fires spatially CSR, or do they exhibit an inhomogeneous spatial intensity?

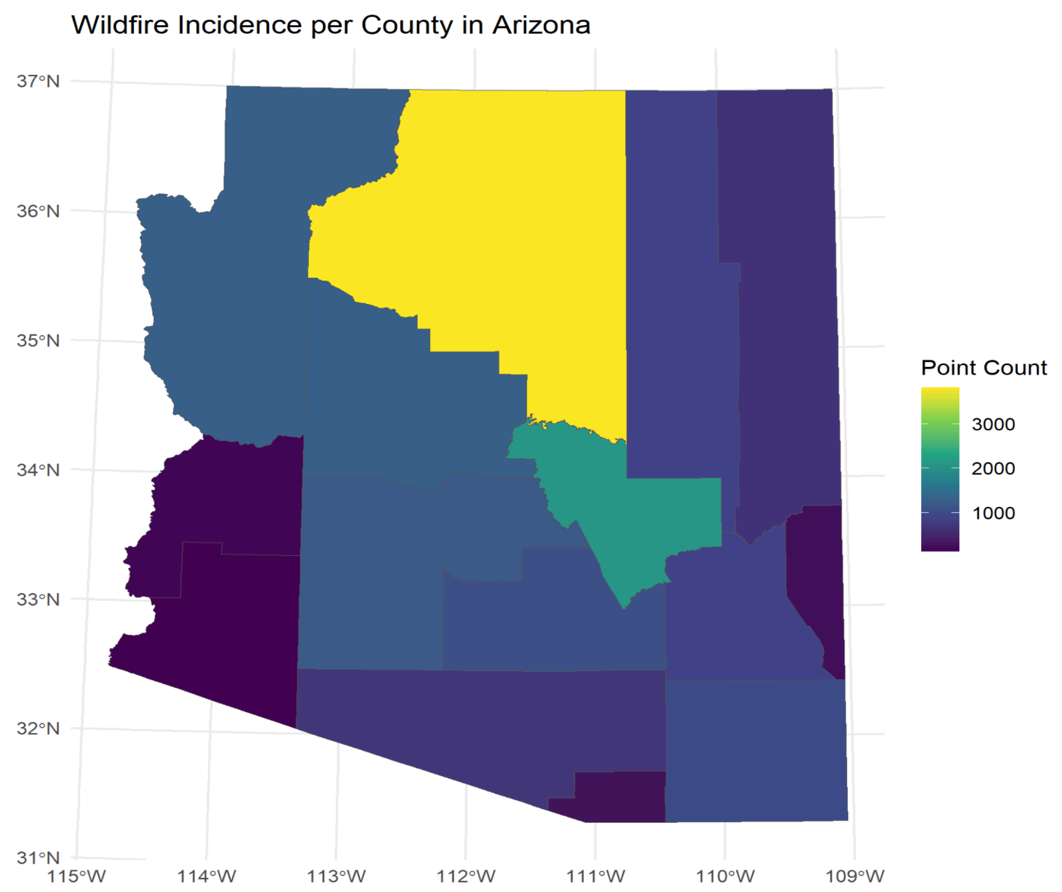
Matt to update this slide

*Add screenshots, like F(r), G(r), K(r) curve just as examples, maybe an intensity map. Only for high leve visual

Modeling - Coconino County

We opted to focus our modeling efforts on Coconino County based on some inital model fits, and overall size of the dataset.

Coconino County is home to national forests like Coconino, Kaibab, and Apache-Sitgreaves, and has the most wildfire incidence in the state by a good margin.



Spatial Linear Model approach

- **Spatial Linear Model** - use the continuous, fixed response data to fit a spatial linear model to predict resulting model size for input points and covariate data.

$$\mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{e} \quad \boldsymbol{e} \sim \mathrm{N}(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$$

- How you landed on the model you did (AIC, other analysis)
- Any data filtering you did, variable transformations
- Point estimates and Confidence intervals on fixed effects
- Fit parameters for spatial variance (nugget, range, sill)
  - de (sill), ie (nugget), range (range)
- `loocv`, diagnostic plots?
- Visuals (ESV, relevant exploration images, etc)
- What might be the practical applications of this model?

LGCP Modeling Approach

Make sure to include CSR analysis to address RQ3

- **Log-Gaussian Cox Process** Treat large wildfire (≥ 1000 acres) incidence locations as arising from a random intensity surface and fit LGCP model accounting for fixed effects and spatial dependence

$$\log(\lambda(u)) = \boldsymbol{Z}(u)\boldsymbol{\beta} + e(u), \quad e(u) \sim \mathrm{N}(0, C(\boldsymbol{\theta})), \quad C(u, u') = \sigma^2 e^{-||u-u'||/h}$$

- Visuals of the data you're investigating (plots of points on coco county) - F, G, K plots with model and CIs plotted for most interesting plots - is data CSR? For what filters (human cause/natural cause, >=1000 acres, etc) - How you landed on the model you did (AIC, other analysis) - Any data filtering you did, variable transformations - Point estimates and Confidence intervals on fixed effects - Fit parameters for spatial variance (nugget, range, sill) - Prediction intensity surface

Binary GLM Spatial Logistic Regression approach

- **Binary GLM Spatial Logistic Regression** - Treat large wildfire incidence as realization of a point process λ_1 (response value$=1$) and generate constant background process λ_0 (response value$=0$). Perform binary logistic regression to model probability of large wildfire incidence for a point and chosen predctors

$$\operatorname{logit}(\lambda_1(\boldsymbol{s})) = \mathbf{x}(\boldsymbol{s})^T\boldsymbol{\beta} + e(\boldsymbol{s}) + \log(\lambda_0)$$