Pesticide Project

**By: Nick Kory,
Abenezer Yitagesu,
Zoe Shang,
Joel Borja**

Outline

- Introduction
- Application Goals
- Technologies Used
- Frontend
- Login
- Backend
- Conclusion/Future work

Introduction

Purpose

Responsive web design for mobile devices/desktop dashboard where pesticide products have been and will be applied at UC Riverside; document pesticide product applications; reports product use.

Users

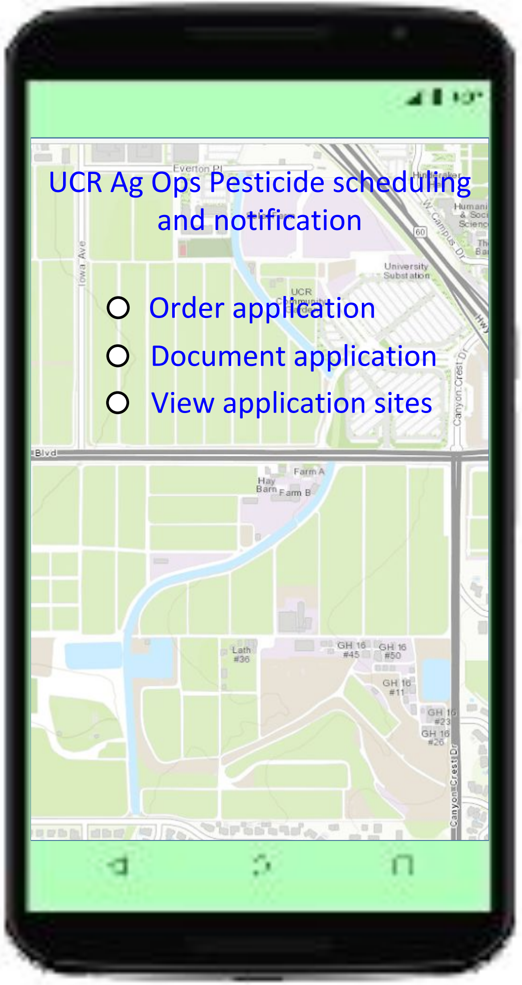
Applicators, pest control supervisor, Ag Ops personnel, PIs, and other land users.

Application Goals

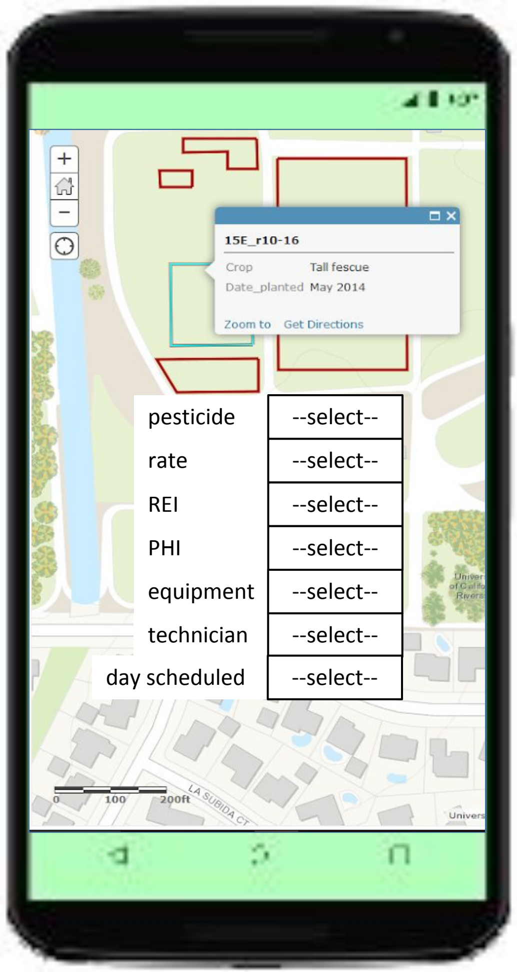
App will display on web map:

- Locations that have been treated with product.
- Locations currently under REI.
- Locations scheduled for product application.
- Date/time of scheduled applications

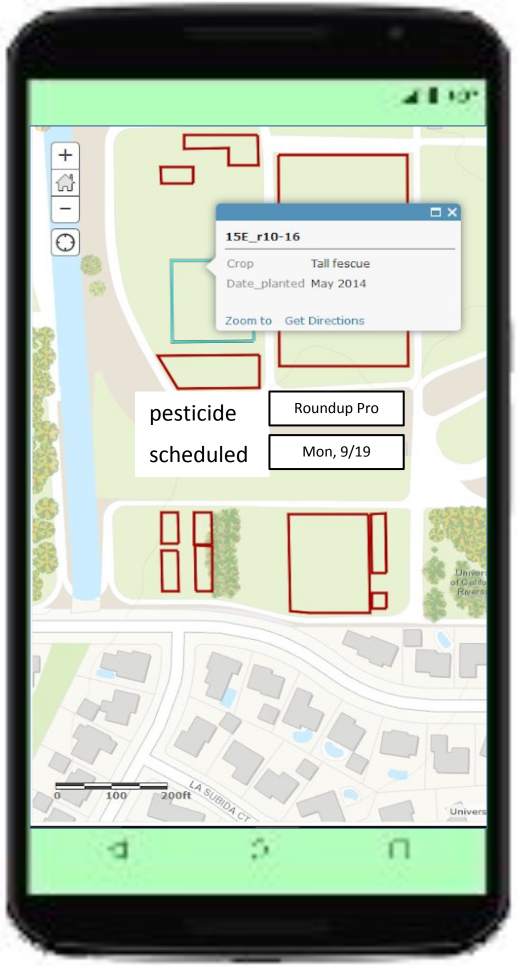
Record pesticide applications by treatment block, include pesticide name and treatment end date/time.



Pesticide app scheduling and notification mobile app

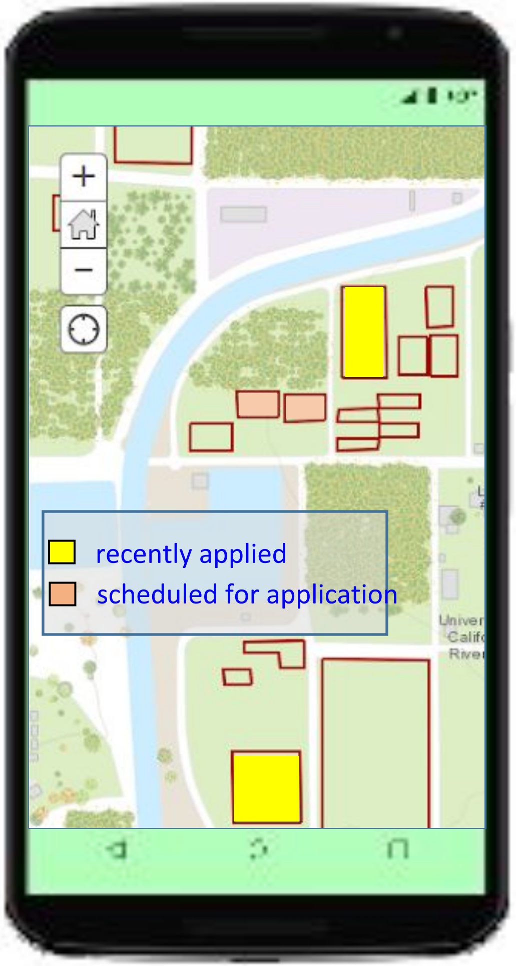


Application order information screen



Application sites screen

- **User** views scheduled application information



Application sites screen

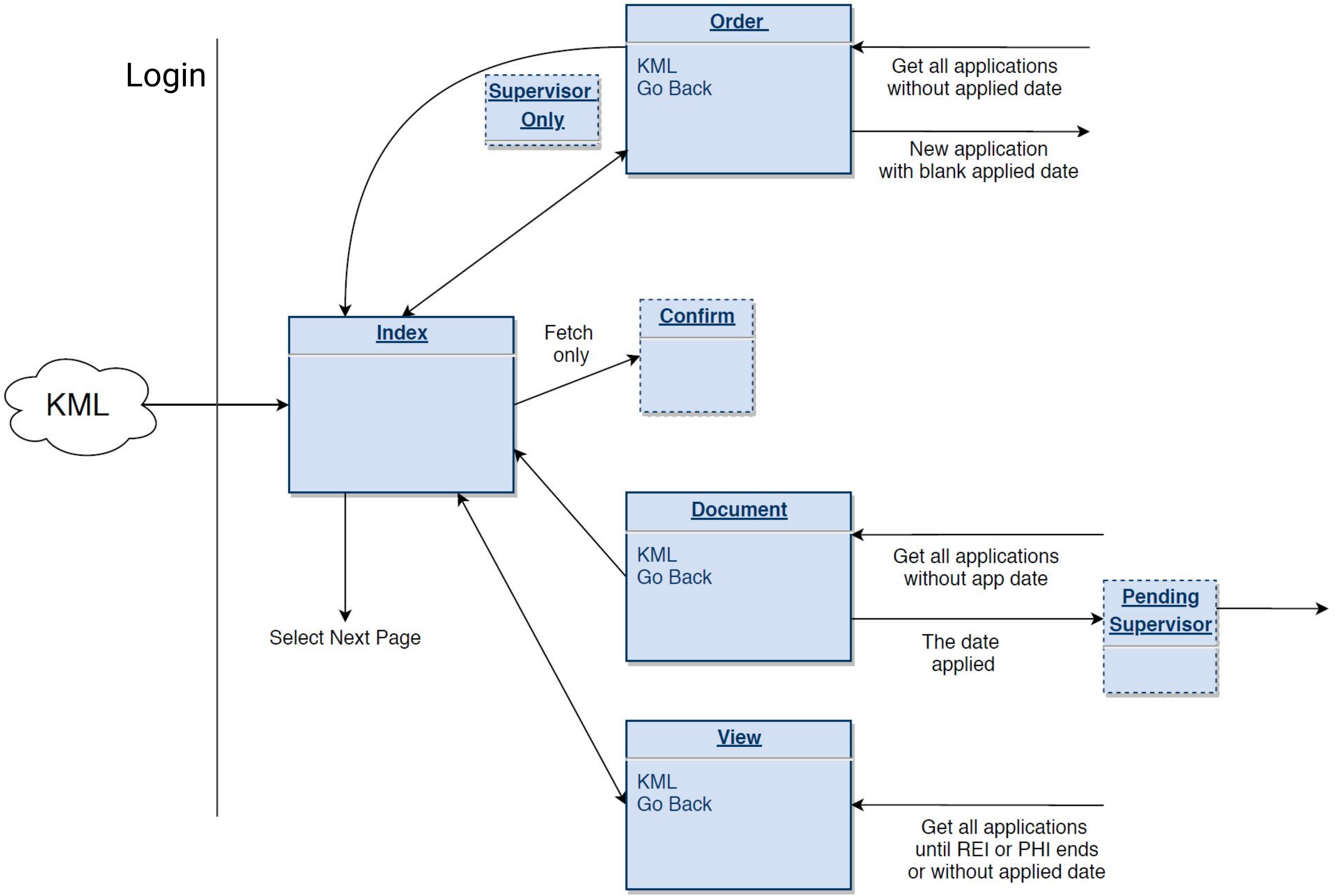
Technologies Used

Back End

- MySQL Server

Front End

- HTML, CSS, PHP, JavaScript
- Leaflet API > Google Maps API
- Openstreetmap
- Bootstrap


Login
KML
Order
KML
Go Back
Supervisor Only
Get all applications without applied date
New application with blank applied date
Index
Fetch only
Confirm
Select Next Page
Document
KML
Go Back
Get all applications without app date
The date applied
Pending Supervisor
View
KML
Go Back
Get all applications until REI or PHI ends or without applied date

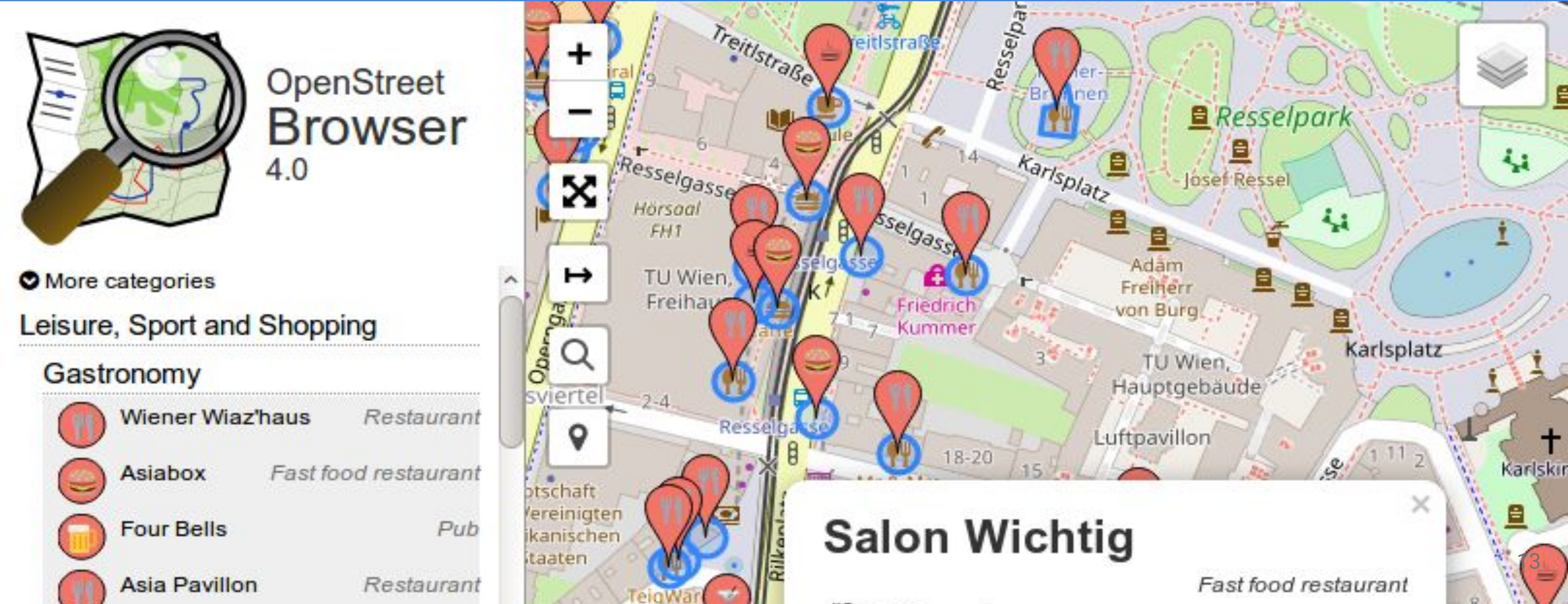
	API Function	API Function Description	API Name
Read	Get all applications without applies date	Returns .json objects where application date blank	Order → select
Read	Get all applications until REI ends or until PHI ends	Returns .json objects where REI date ≥ today or PHI ≥ today	View → select
Write	Insert new application record with blank applied date	Params: All columns from app_record table except applied date returns: 1 if success 0 if fail	Document → select for update insert
Write	Update application record with applied date	Params: Primary key from .json objects & date returns: 1 if success 0 if fail	Update → select for update

Frontend

Leaflet

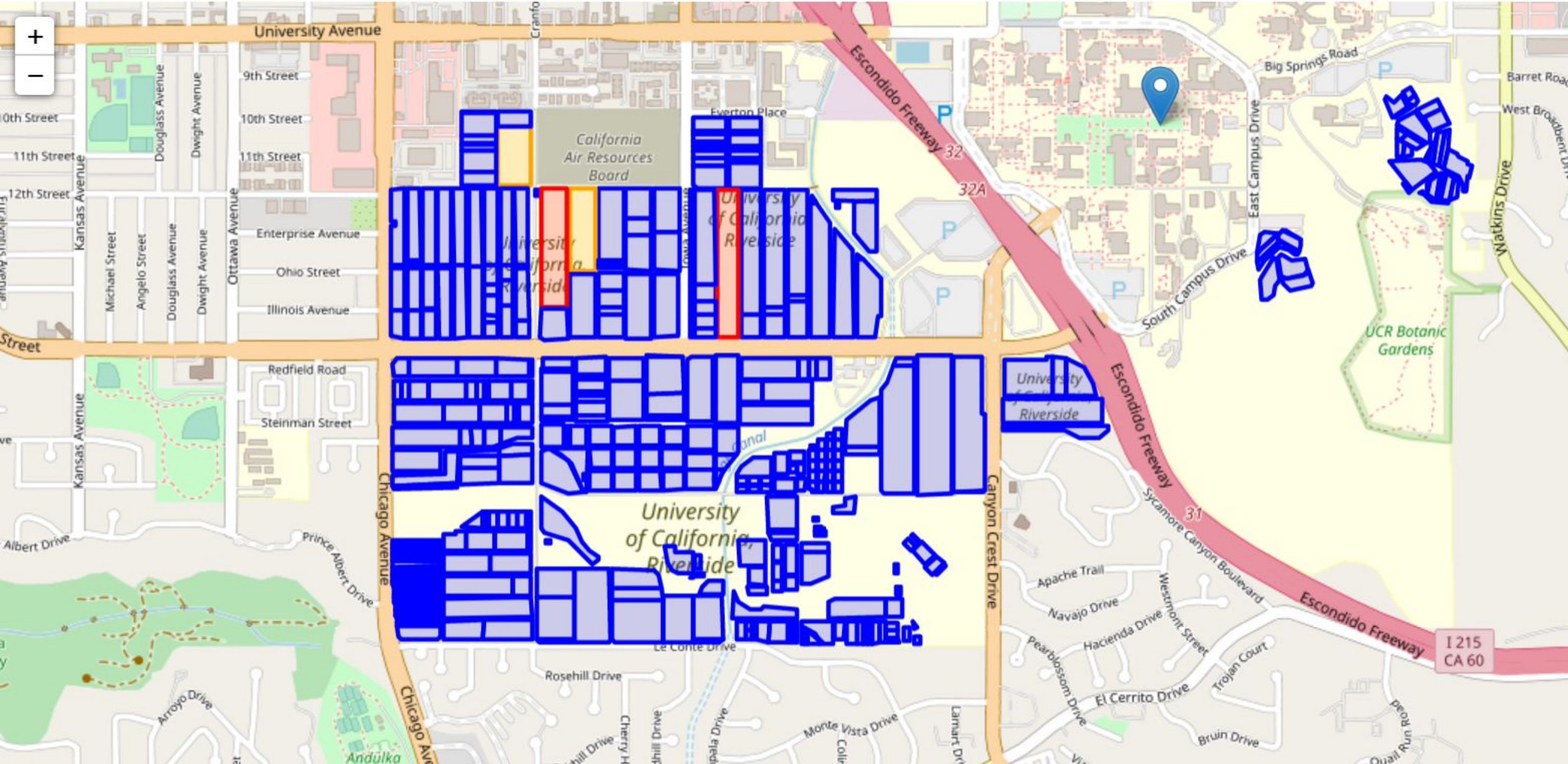
an open-source JavaScript library
for mobile-friendly interactive maps

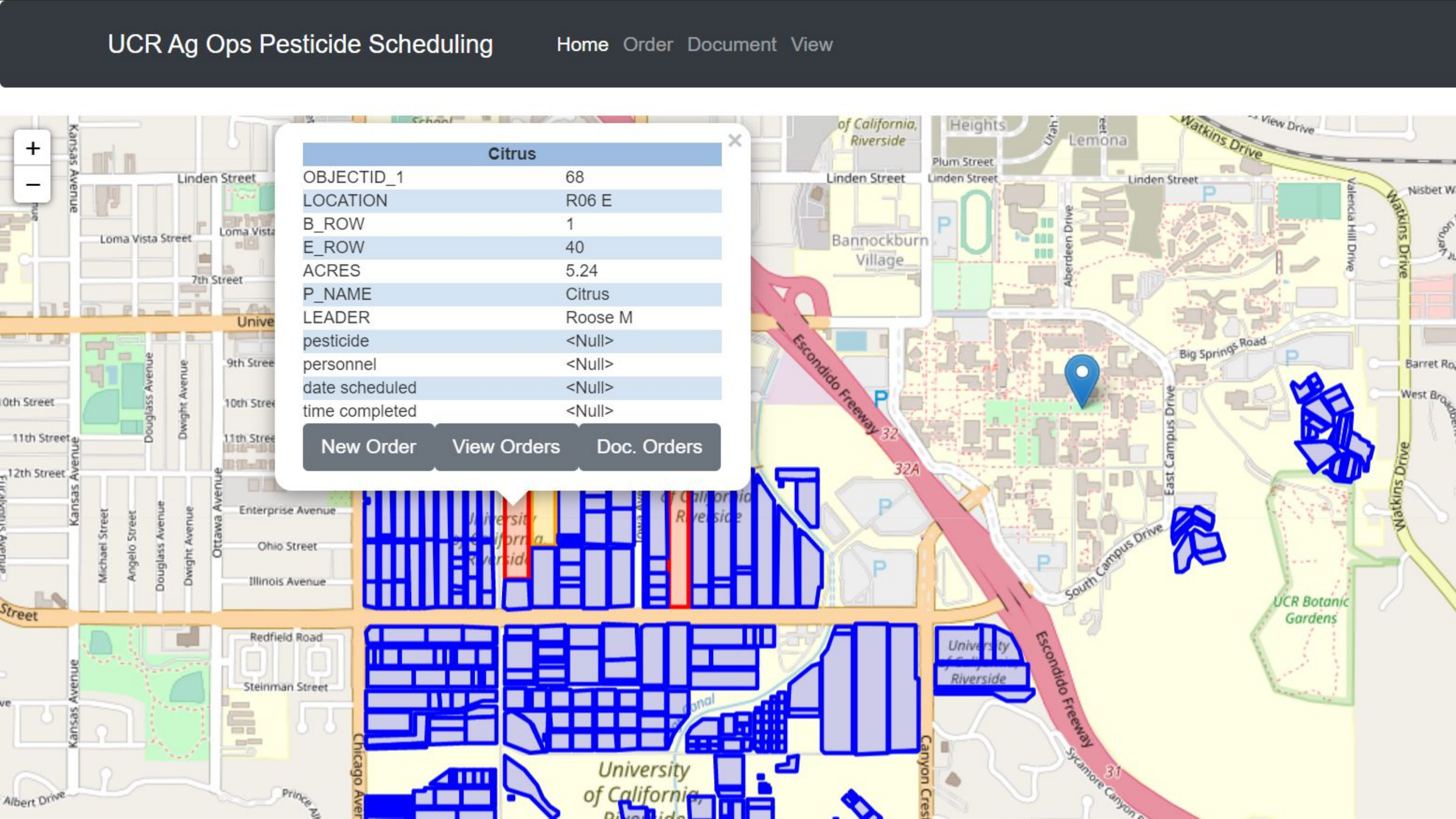
Frontend


OpenStreet
Browser
4.0
More categories
Leisure, Sport and Shopping
Gastronomy
Wiener Wiaz'haus Restaurant
Asiabox Fast food restaurant
Four Bells Pub
Asia Pavillon Restaurant
Treitlstraße
Resselgasse
Karlsplatz
Resselpark
Josef Ressel
Adam Freiherr von Burg
TU Wien, Hauptgebäude
Luftpavillon
Friedrich Kummer
Salon Wichtig
Fast food restaurant

UCR Ag Ops Pesticide Scheduling

Home Order Document View




UCR Ag Ops Pesticide Scheduling
Home
Order
Document
View
Citrus
OBJECTID_1 68
LOCATION R06 E
B_ROW 1
E_ROW 40
ACRES 5.24
P_NAME Citrus
LEADER Roose M
pesticide <Null>
personnel <Null>
date scheduled <Null>
time completed <Null>
New Order
View Orders
Doc. Orders
Linden Street
Loma Vista Street
7th Street
Bannockburn Village
Escondido Freeway
Big Springs Road
East Campus Drive
South Campus Drive
Watkins Drive
UCR Botanic Gardens
University of California, Riverside
Chicago Avenue
Canyon Crest
Kansas Avenue
Enterprise Avenue
Ohio Street
Illinois Avenue
Redfield Road
Steinman Street

UCR Ag Ops Pesticide Scheduling Home Order Document View

Order

Feild:

324

Pesticide:

15-15-15,

Rate:

Rate Information

REI:

REI Information

Login

- Used: HTML, CSS, and PHP
- Attempted registering users
- Attempted checking if users are registered
- Attempted to implement a recover lost password feature
- Attempted returned to home

LOGIN MODES

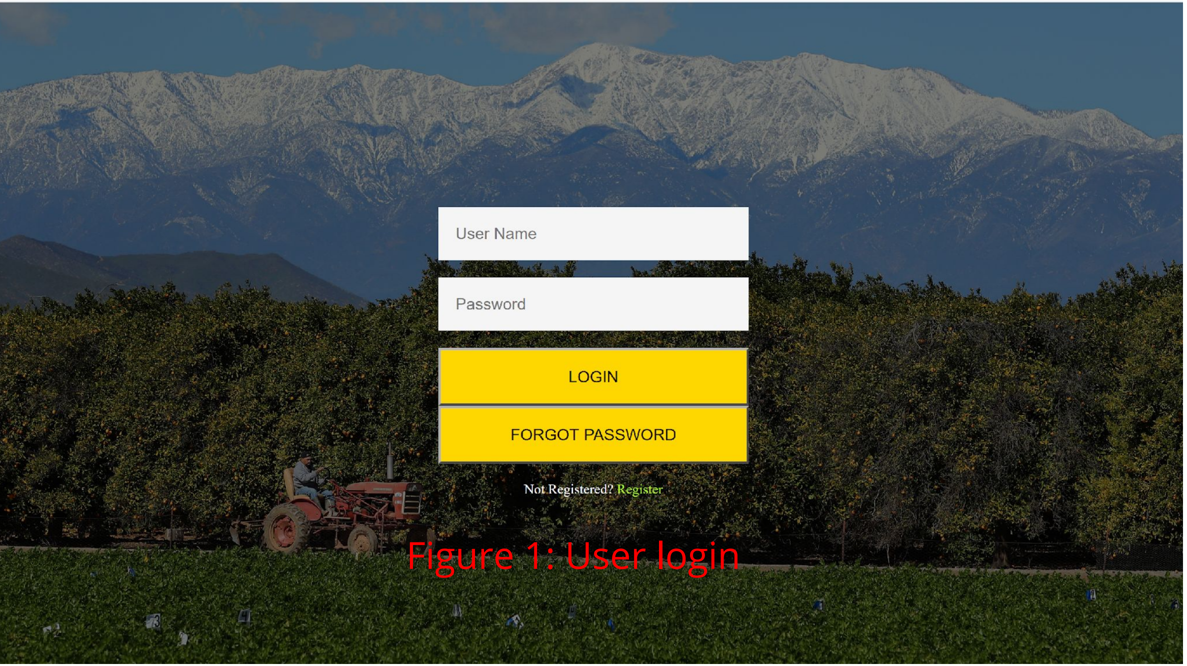


Figure 1: User login

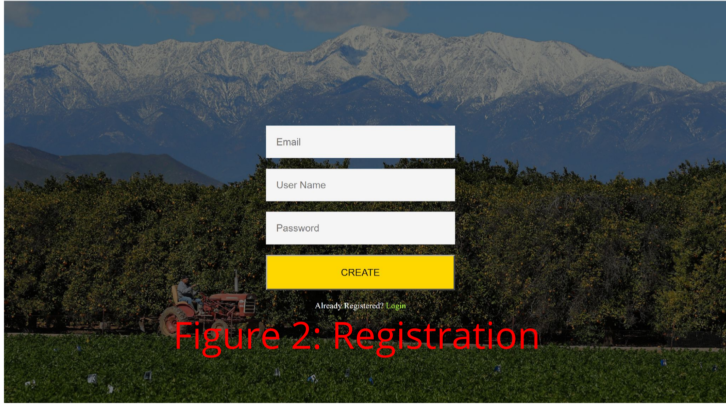


Figure 2: Registration

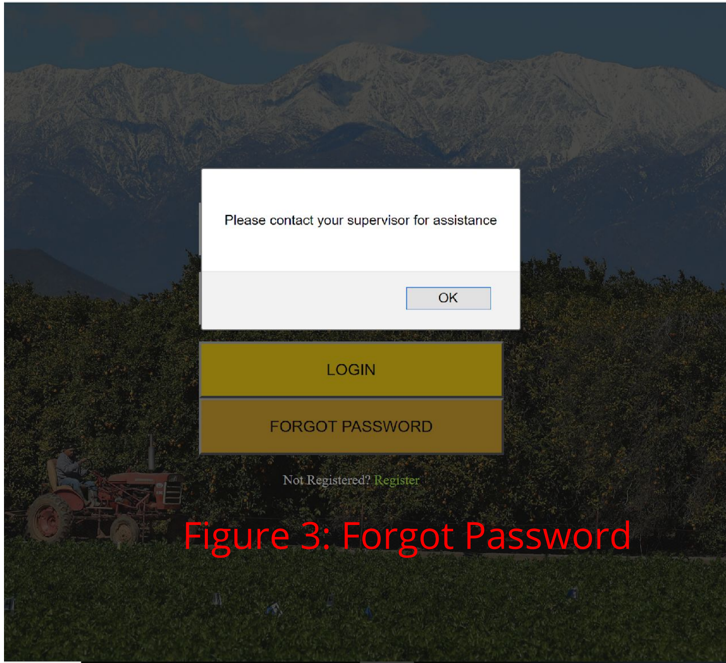


Figure 3: Forgot Password

Back End

- Database: MySQL
- Spring boot framework
- API

  Submit Order

  Fetch Products

  Fetch Technicians

Order

Feild:

173

Pesticide:

89, 27-3-4

Rate:

Rate Information

REI:

REI Information

PHI:

PHI Information

Equipment:

PHI Information

Technician:

Micky

Day Scheduled:

Submit Order

Order

Feild:

173

Pesticide:

✓ 89, 27-3-4
255, 15-5-8 Microgreen
91, 31-0-0 PAR EX

Rate Information

REI:

REI Information

PHI:

PHI Information

Equipment:

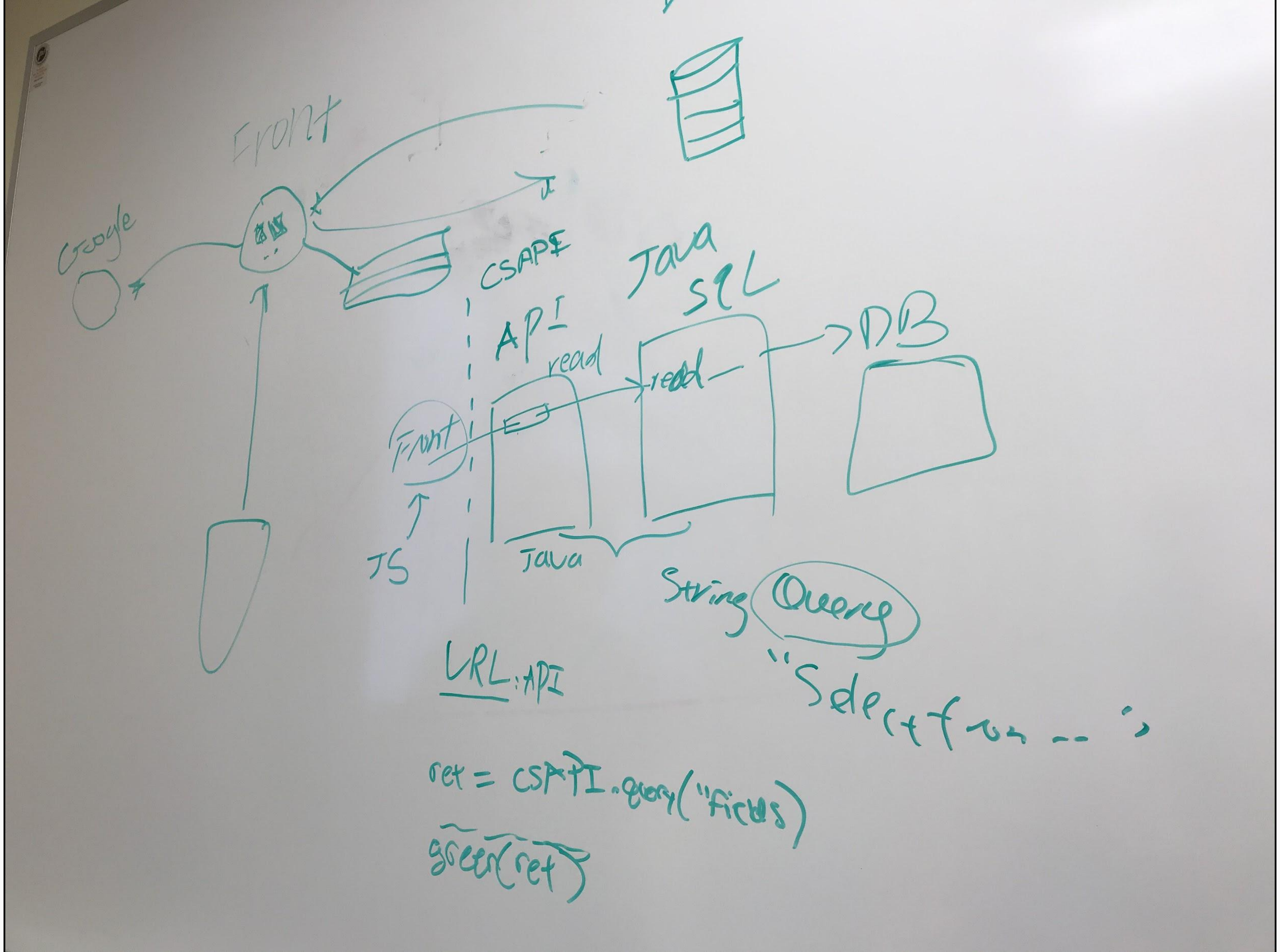
PHI Information

Technician:

Micky

Day Scheduled:

Submit Order



Front
Google
CSAPI
Java
SQL
API
read
read
DB
Front
JS
Java
String
Query
"Select from ... ";
URL: API
ret = CSAPI.query("fields")
green(ret)

Future Works

- Integrating UC Riverside identity management.
- Notifications to supervisors when an order is created (via the app or via email).
- Admin view.
- “Tunnel-proof” code that saves orders and posts when a WiFi or data connection is lost then re-established.
- Integrating dynamic field loads when ArcGIS layer is updated.
- Protection against SQL injections.
- Feature for supervisors to add/change permissions of users.
- Read receipts when a field worker sees a new order.
- Package deployment to UC Riverside hosted web server.
- Create test / dev / prod implementations.

Questions?

By: Nick Kory,
Abenezer Yitagesu,
Zoe Shang,
Joel Borja

Spatial Computing(CS225) Project Presentation

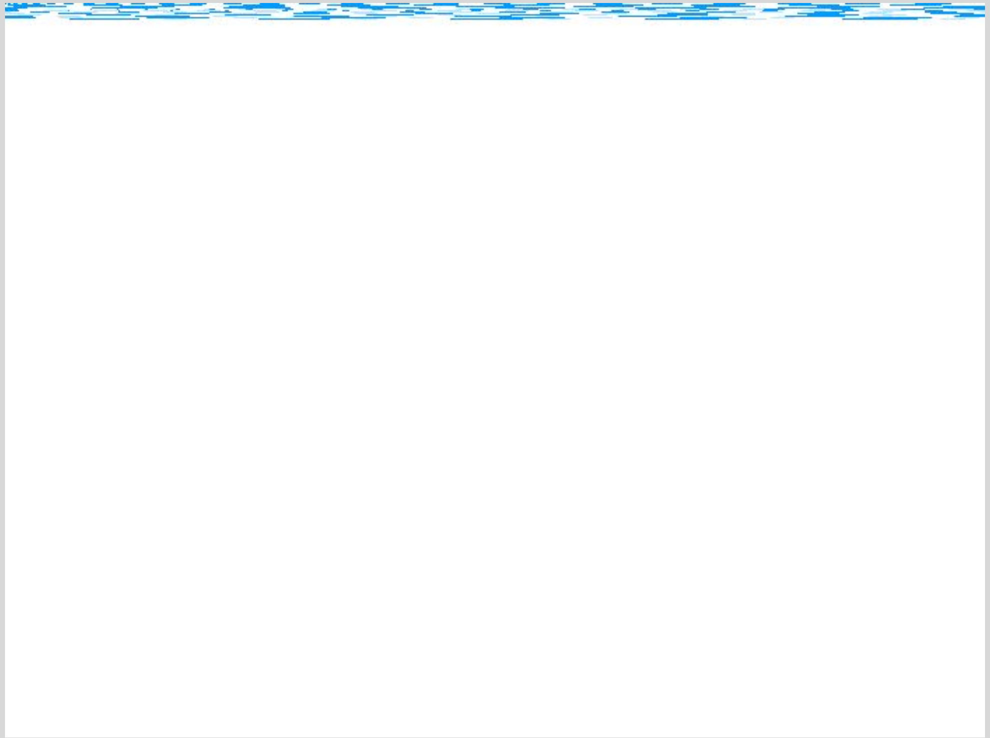
Group 8

Siddharth Shenoy, Devansh Sheth, Ganesh Krishnan Sivaram, Mahip Shah

Finding correlation between vegetation, temperature, humidity and wildfires using satellite data for counties of California

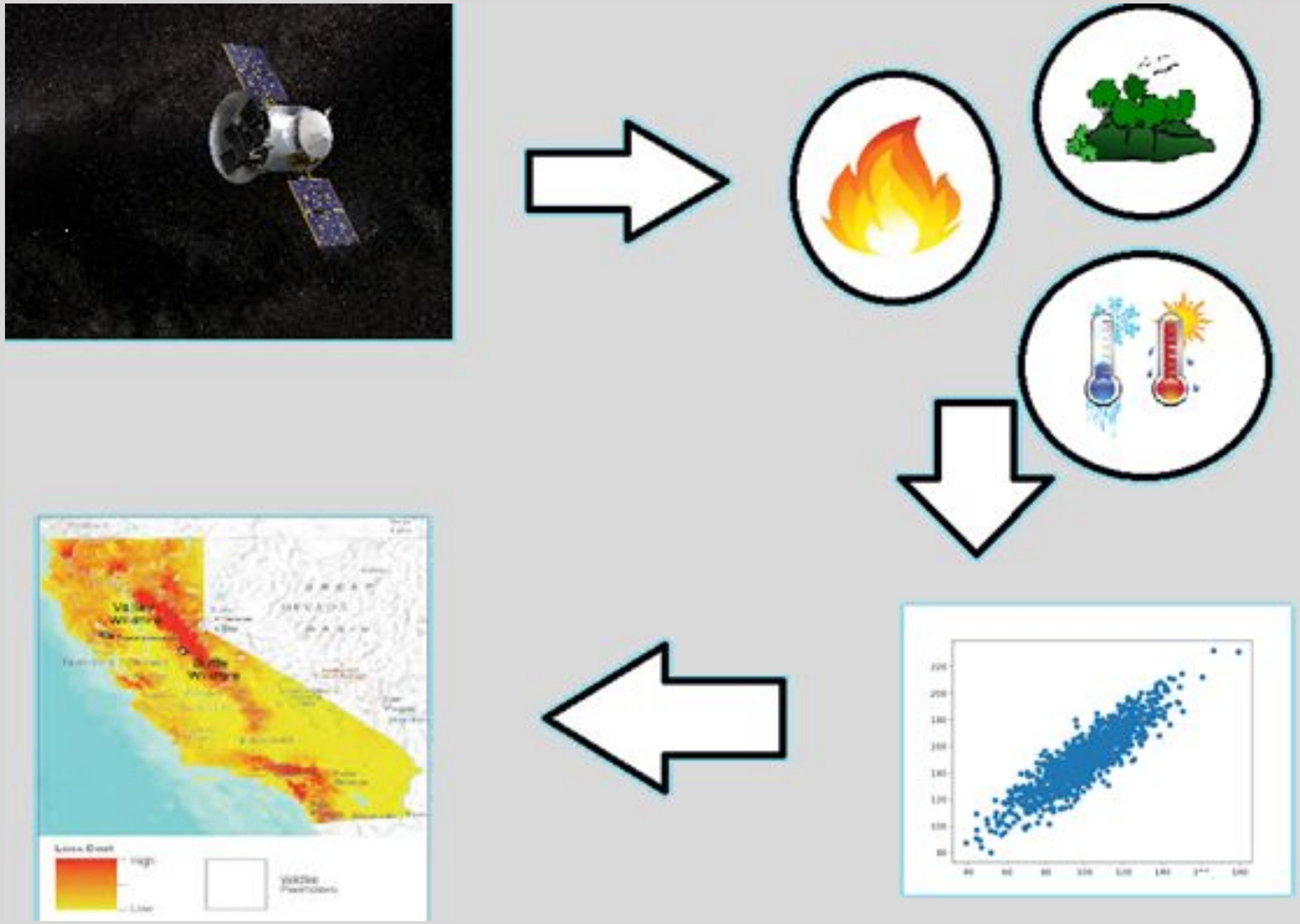
- BUT WHY ARE WE DOING IT? AND WHAT ARE WE ACTUALLY DOING?

WHY ?

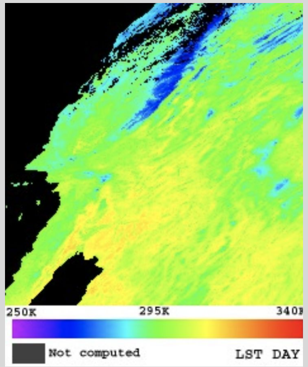


- California fires are getting bigger.
- Large wildfires in the US now burn more than twice the area they did in 1970.
- A recent study found that the portion of California that burns from wildfires every year has increased more than five-fold since 1972.

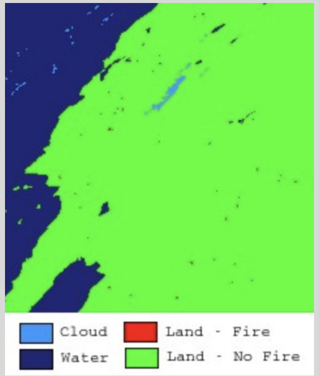
HOW ?



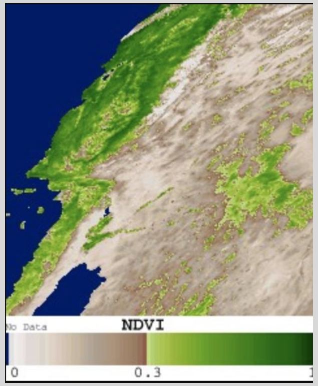
The Satellite Dataset



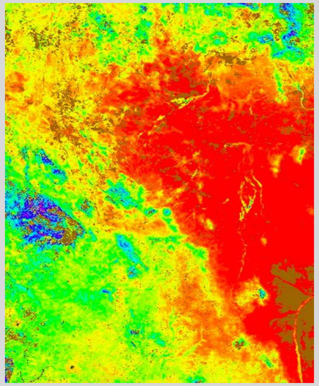
Temperature Data
MODIS MOD11A2



Fire Data
MODIS MOD14A2



Vegetation Data
MODIS MOD13A2



Humidity Data
MODIS MOD16A2

Reality

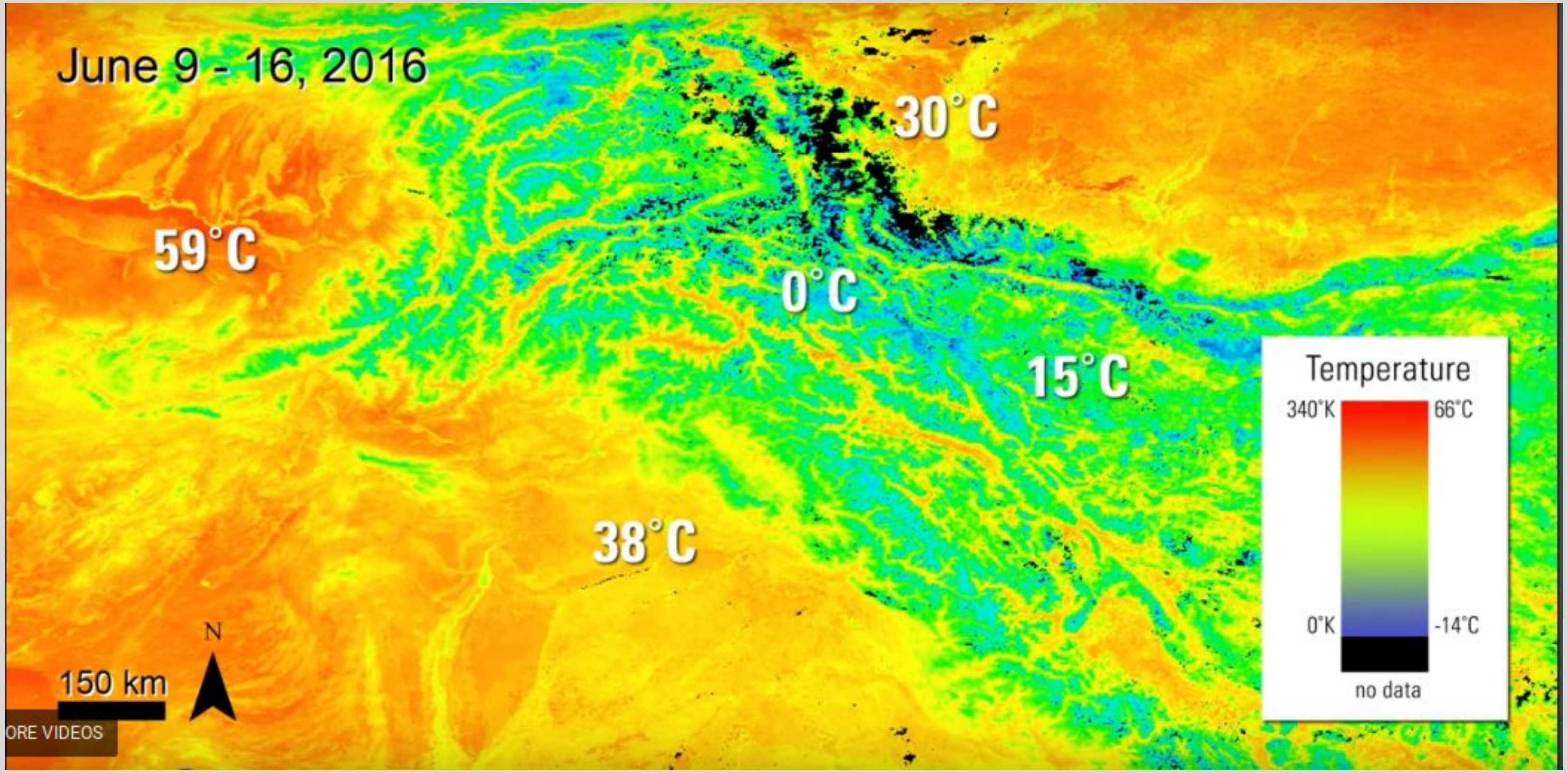
Challenges faced

- Data Format Conversion
- Normalizing data
- Data combining
- Correlation between four features
- Verification of actual occurrence of wildfire with our data

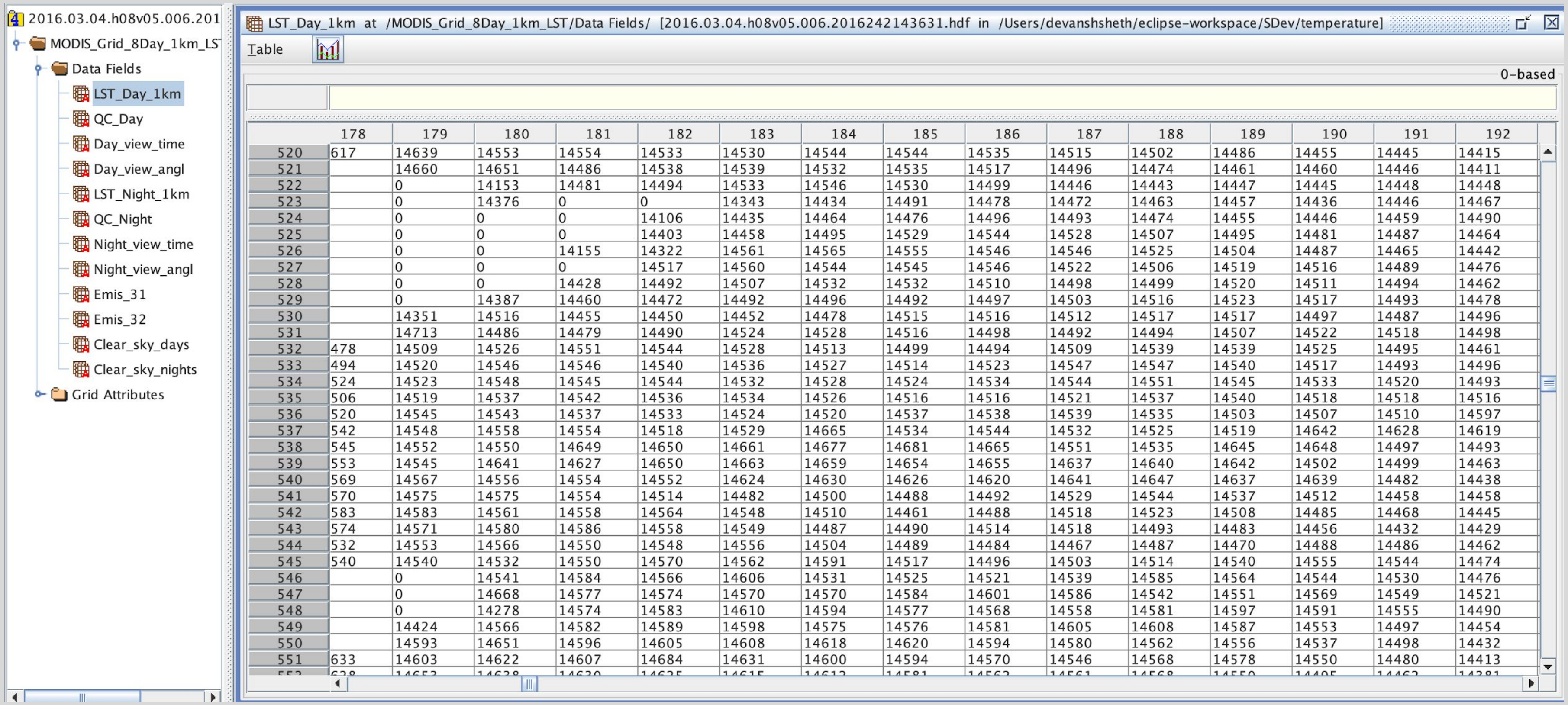
Data processing

- Extracting values from HDF file
- Aggregate data for various counties
- Normalize the data for a common resolution
- Merge the four factors into a single dataset

Actual Satellite Data



HDF VIEW



	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192
520	617	14639	14553	14554	14533	14530	14544	14544	14535	14515	14502	14486	14455	14445	14415
521		14660	14651	14486	14538	14539	14532	14535	14517	14496	14474	14461	14460	14446	14411
522		0	14153	14481	14494	14533	14546	14530	14499	14446	14443	14447	14445	14448	14448
523		0	14376	0	0	14343	14434	14491	14478	14472	14463	14457	14436	14446	14467
524		0	0	0	14106	14435	14464	14476	14496	14493	14474	14455	14446	14459	14490
525		0	0	0	14403	14458	14495	14529	14544	14528	14507	14495	14481	14487	14464
526		0	0	14155	14322	14561	14565	14555	14546	14546	14525	14504	14487	14465	14442
527		0	0	0	14517	14560	14544	14545	14546	14522	14506	14519	14516	14489	14476
528		0	0	14428	14492	14507	14532	14532	14510	14498	14499	14520	14511	14494	14462
529		0	14387	14460	14472	14492	14496	14492	14497	14503	14516	14523	14517	14493	14478
530		14351	14516	14455	14450	14452	14478	14515	14516	14512	14517	14517	14497	14487	14496
531		14713	14486	14479	14490	14524	14528	14516	14498	14492	14494	14507	14522	14518	14498
532	478	14509	14526	14551	14544	14528	14513	14499	14494	14509	14539	14539	14525	14495	14461
533	494	14520	14546	14546	14540	14536	14527	14514	14523	14547	14547	14540	14517	14493	14496
534	524	14523	14548	14545	14544	14532	14528	14524	14534	14544	14551	14545	14533	14520	14493
535	506	14519	14537	14542	14536	14534	14526	14516	14516	14521	14537	14540	14518	14518	14516
536	520	14545	14543	14537	14533	14524	14520	14537	14538	14539	14535	14503	14507	14510	14597
537	542	14548	14558	14554	14518	14529	14665	14534	14544	14532	14525	14519	14642	14628	14619
538	545	14552	14550	14649	14650	14661	14677	14681	14665	14551	14535	14645	14648	14497	14493
539	553	14545	14641	14627	14650	14663	14659	14654	14655	14637	14640	14642	14502	14499	14463
540	569	14567	14556	14554	14552	14624	14630	14626	14620	14641	14647	14637	14639	14482	14438
541	570	14575	14575	14554	14514	14482	14500	14488	14492	14529	14544	14537	14512	14458	14458
542	583	14583	14561	14558	14564	14548	14510	14461	14488	14518	14523	14508	14485	14468	14445
543	574	14571	14580	14586	14558	14549	14487	14490	14514	14518	14493	14483	14456	14432	14429
544	532	14553	14566	14550	14548	14556	14504	14489	14484	14467	14487	14470	14488	14486	14462
545	540	14540	14532	14550	14570	14562	14591	14517	14496	14503	14514	14540	14555	14544	14474
546		0	14541	14584	14566	14606	14531	14525	14521	14539	14585	14564	14544	14530	14476
547		0	14668	14577	14574	14570	14570	14584	14601	14586	14542	14551	14569	14549	14521
548		0	14278	14574	14583	14610	14594	14577	14568	14558	14581	14597	14591	14555	14490
549		14424	14566	14582	14589	14598	14575	14576	14581	14605	14608	14587	14553	14497	14454
550		14593	14651	14596	14605	14608	14618	14620	14594	14580	14562	14556	14537	14498	14432
551	633	14603	14622	14607	14684	14631	14600	14594	14570	14546	14568	14578	14550	14480	14413

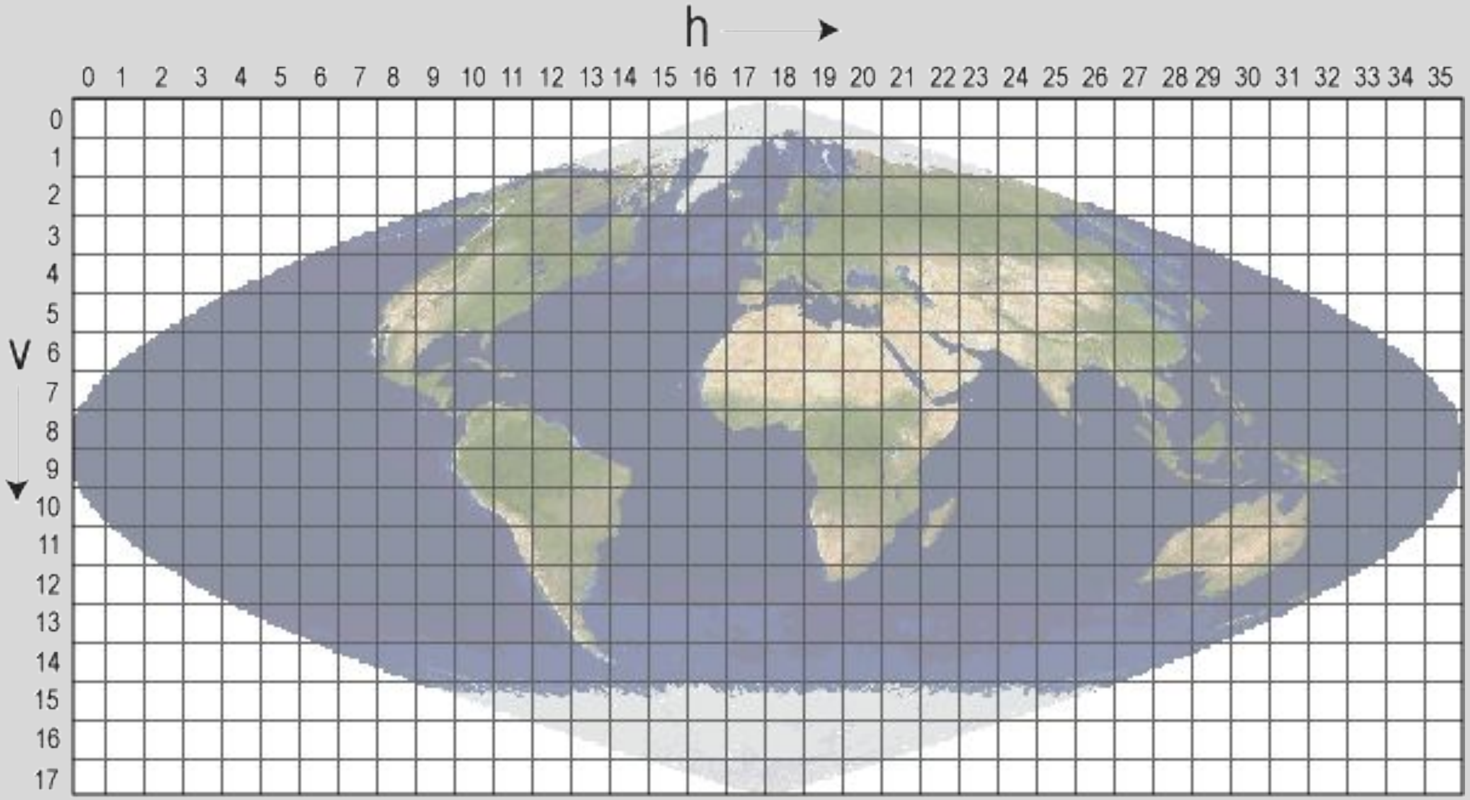
Conversion of data points

Data Layer Characteristics

SDS Layer Name	Description	Units	Data Type	Fill Value	Valid Range	Scaling Factor	Additional Offset
LST_Day_1km	Day Land Surface Temperature	Kelvin	16-bit unisnged integer	0	7500 to 65535	0.02	N/A
QC_Day	Daytime LST Quality Indicators	Bit Field	8-bit unsigned integer	N/A	0 to 255	N/A	N/A
[illegible]	[illegible]	Field	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Night_view_time	Local time of night observation	Hours	8-bit unsigned integer	255	0 to 240	0.1	N/A
Night_view_angle	View zenith angle of night observation	Degree	8-bit unsigned integer	255	0 to 130	1.0	-65
Emis_31	Band 31 emissivity	None	8-bit unsinged integer	0	1 to 255	0.002	0.49
Emis_32	Band 32 emissivity	None	8-bit unsigned integer	0	1 to 255	0.002	0.49
Clear_day_cov	N/A	None	16-bit unsigned integer	0	1 to 65535	0.0005	N/A
Clear_night_cov	N/A	None	16-bit unsigned integer	0	1 to 65535	0.0005	N/A

$$15592 \times .02 = 311.84$$

Sinusoidal Projection



County boundaries

- For each and every county in California, we approximate the boundaries in form of a polygon whose co ordinates are used to get the data relevant to that county

Output post processing

Vegetation data for each day for each county

2019.01.01	Alameda	0.2267418474
2019.01.01	Alpine	0.1088802544
2019.01.01	Amador	0.2672708952
2019.01.01	Butte	0.2119804167
2019.01.01	Calaveras	0.2694989798
2019.01.01	Colusa	0.1662770325
2019.01.01	Contra Costa	0.2484854556
2019.01.01	Del Norte	0.3211448463
2019.01.01	El Dorado	0.2282522958
2019.01.01	Fresno	0.1845172419
2019.01.01	Glenn	0.2183001748
2019.01.01	Humboldt	0.3335566537
2019.01.01	Imperial	0.1104355023
2019.01.01	Inyo	0.06437238688
2019.01.01	Kern	0.1920874498
2019.01.01	Kings	0.2378501078
2019.01.01	Lake	0.1480543058
2019.01.01	Lassen	0.1010948126
2019.01.01	Los Angeles	0.167926649
2019.01.01	Madera	0.2307867279
2019.01.01	Marin	0.3060344877
2019.01.01	Mariposa	0.2152502721
2019.01.01	Mendocino	0.3008924707
2019.01.01	Merced	0.2792191939
2019.01.01	Modoc	0.1195551398
2019.01.01	Mono	0.05804772432
2019.01.01	Monterey	0.2393006009
2019.01.01	Napa	0.2437045166
2019.01.01	Nevada	0.216698655
2019.01.01	Orange	0.1933352243
2019.01.01	Placer	0.2050904407

Temperature data for each day for each county

2019.01.01	Alameda	285.6409265
2019.01.01	Alpine	270.367217
2019.01.01	Amador	282.517969
2019.01.01	Butte	282.583559
2019.01.01	Calaveras	283.2362063
2019.01.01	Colusa	284.1931687
2019.01.01	Contra Costa	284.9329969
2019.01.01	Del Norte	280.9919602
2019.01.01	El Dorado	278.3651384
2019.01.01	Fresno	281.2031551
2019.01.01	Glenn	283.923294
2019.01.01	Humboldt	281.8464864
2019.01.01	Imperial	290.7841403
2019.01.01	Inyo	282.9137591
2019.01.01	Kern	286.1741523
2019.01.01	Kings	288.017766
2019.01.01	Lake	282.9989065
2019.01.01	Lassen	273.9235209
2019.01.01	Los Angeles	287.6298483
2019.01.01	Madera	280.2969498
2019.01.01	Marin	284.3173414
2019.01.01	Mariposa	280.8140821
2019.01.01	Mendocino	282.5648755
2019.01.01	Merced	285.3748298
2019.01.01	Modoc	274.5988018
2019.01.01	Mono	271.7146342
2019.01.01	Monterey	285.9114802
2019.01.01	Napa	283.9690859
2019.01.01	Nevada	277.7528152
2019.01.01	Orange	290.9759709
2019.01.01	Placer	278.5222309

Output post processing

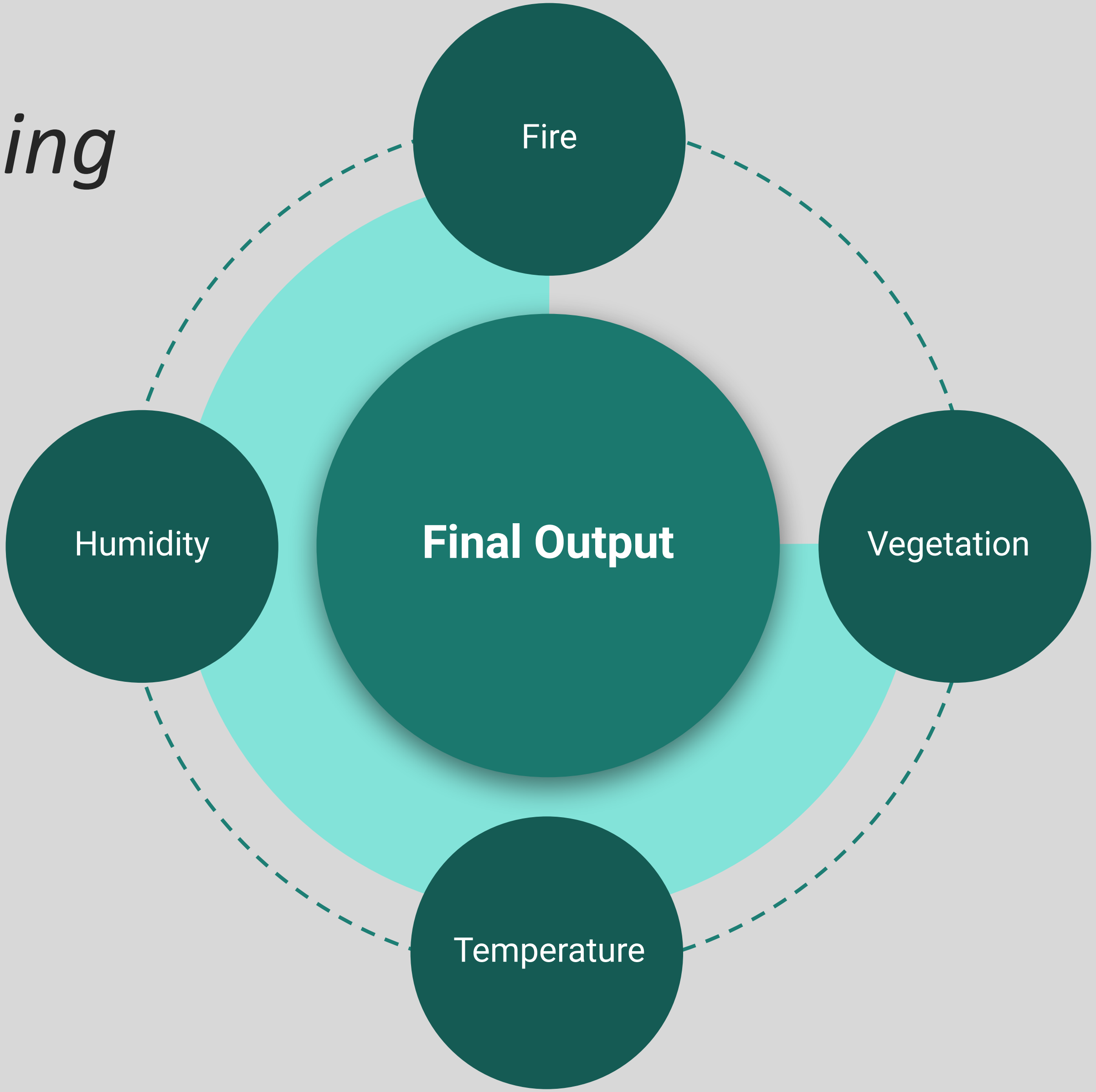
Fire data for each day for each county

2019.01.01	Calaveras	1
2019.01.01	El Dorado	2
2019.01.01	Fresno	19
2019.01.01	Inyo	2
2019.01.01	Kern	9
2019.01.01	Kings	2
2019.01.01	Lake	1
2019.01.01	Los Angeles	2
2019.01.01	Madera	2
2019.01.01	Monterey	2
2019.01.01	Napa	2
2019.01.01	Riverside	2
2019.01.01	San Bernardino	5
2019.01.01	San Diego	5
2019.01.01	San Joaquin	3
2019.01.01	San Luis Obispo	11
2019.01.01	Santa Barbara	5
2019.01.01	Sonoma	2
2019.01.01	Stanislaus	2
2019.01.01	Tehama	1
2019.01.01	Tulare	9
2019.01.01	Yolo	1
2019.01.09	Alameda	5
2019.01.09	Amador	11
2019.01.09	Calaveras	1
2019.01.09	Colusa	2
2019.01.09	El Dorado	10
2019.01.09	Fresno	19
2019.01.09	Humboldt	2
2019.01.09	Kern	8
2019.01.09	Kings	2

Humidity data for each day for each county

2019.01.01	Alameda	1477.133989
2019.01.01	Alpine	2318.423261
2019.01.01	Amador	1041.737199
2019.01.01	Butte	912.601891
2019.01.01	Calaveras	797.3416382
2019.01.01	Colusa	1229.386295
2019.01.01	Contra Costa	1562.712314
2019.01.01	Del Norte	914.2515258
2019.01.01	El Dorado	1003.121748
2019.01.01	Fresno	1636.582925
2019.01.01	Glenn	939.9926974
2019.01.01	Humboldt	1336.131513
2019.01.01	Imperial	2565.591397
2019.01.01	Inyo	2434.225649
2019.01.01	Kern	1673.688516
2019.01.01	Kings	1764.723261
2019.01.01	Lake	1159.707826
2019.01.01	Lassen	2279.86809
2019.01.01	Los Angeles	2016.074389
2019.01.01	Madera	1376.798862
2019.01.01	Marin	867.845135
2019.01.01	Mariposa	1208.625829
2019.01.01	Mendocino	1148.409436
2019.01.01	Merced	1462.684713
2019.01.01	Modoc	2106.22431
2019.01.01	Mono	2554.168936
2019.01.01	Monterey	806.0005959
2019.01.01	Napa	808.741114
2019.01.01	Nevada	1025.226868

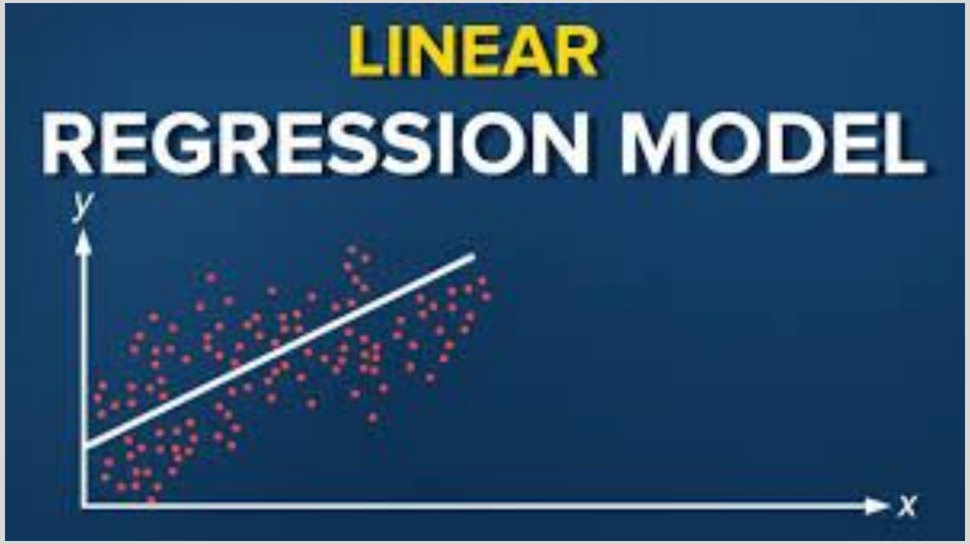
Data merging



First normalize the resolution. How ?

Complex calculations?

No



How do we do merging?

- First we convert the resolution of all the data with different resolution to daily resolution by using linear regression over the period
- Then we merge all the data from 4 files into one single file with each record containing vegetation , temperature, humidity and fire data for the county on a particular date
- Then we filter the data based on counties to get data for each county and sort them date-wise for the whole year

What do we get?

- Single CSV file with data points for each county for every day

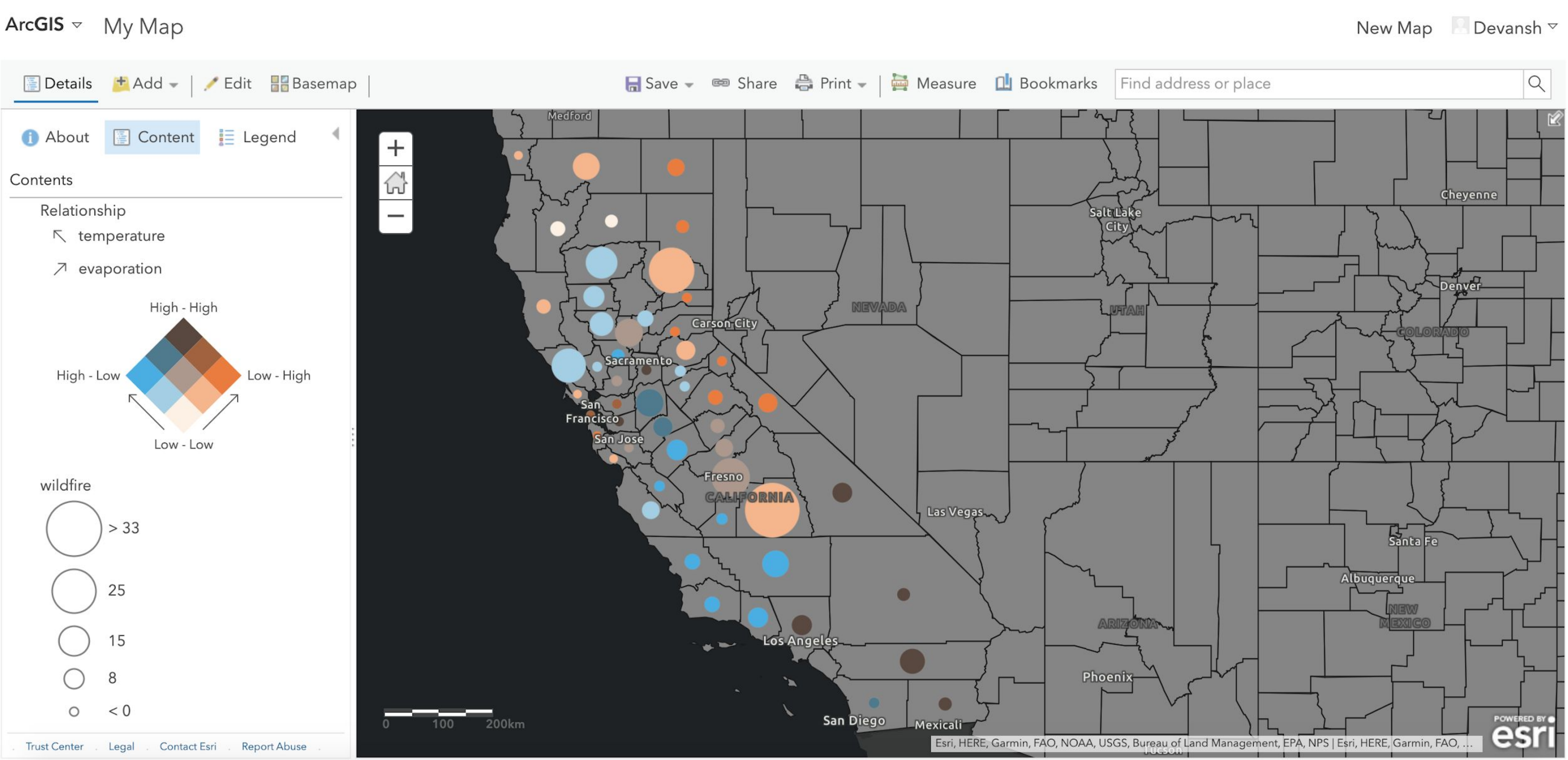
county	evaporation	vegetation	wildfire	temperature
Alameda	1540.61745	0.2444800521	-2.457494657	285.6409265
Alpine	1653.707409	0.1826841739	-2.522712048	270.367217
Amador	939.0322917	0.2970927166	-1.544451179	282.517969
Butte	980.9912948	0.2946972119	6.738157517	282.583559
Calaveras	823.1842422	0.2984718268	-2.218364222	283.2362063
Colusa	895.8056976	0.2543905143	7.520766213	284.1931687
Contra Costa	1450.308913	0.276925376	-2.522712048	284.9329969
Del Norte	1171.851784	0.373021073	-3.240103353	280.9919602
El Dorado	1113.793531	0.2793030768	4.151200995	278.3651384
Fresno	1184.618777	0.2250945681	17.8685923	281.2031551
Glenn	822.0136035	0.2811806412	5.629461865	283.923294
Humboldt	1282.209065	0.4023410539	-2.848799005	281.8464864
Imperial	2294.984942	0.1214066273	0.08598360394	290.7841403
Inyo	1488.087693	0.09792610717	4.825114039	282.9137591
Kern	920.6759412	0.217077208	9.520766213	286.1741523
Kings	882.6821281	0.2573915058	-1.58792944	288.017766
Lake	901.9449365	0.226880316	-0.6531468308	282.9989065
Lassen	1439.445816	0.1718064405	-0.3705381352	273.9235209
Los Angeles	1492.457644	0.2077347433	5.107722734	287.6298483
Madera	1119.806661	0.2655455157	3.194679256	280.2969498
Marin	1000.425629	0.333965632	-3.370538135	284.3173414
Mariposa	1019.732511	0.2532437272	0.4772879518	280.8140821
Mendocino	1081.74973	0.3676860185	0.738157517	282.5648755
Merced	910.0607244	0.2909325323	5.172940126	285.3748298
Modoc	1451.691769	0.1706357713	2.759896647	274.5988018
Mono	1524.108964	0.1259336784	4.064244474	271.7146342
Monterey	652.8207787	0.2792733363	3.194679256	285.9114802
Napa	725.2980304	0.2931885699	-2.414016396	283.9690859
Nevada	1072.512885	0.2936267091	-2.718364222	277.7528152
Orange	2185.948845	0.229680026	-3.305320744	290.9759709

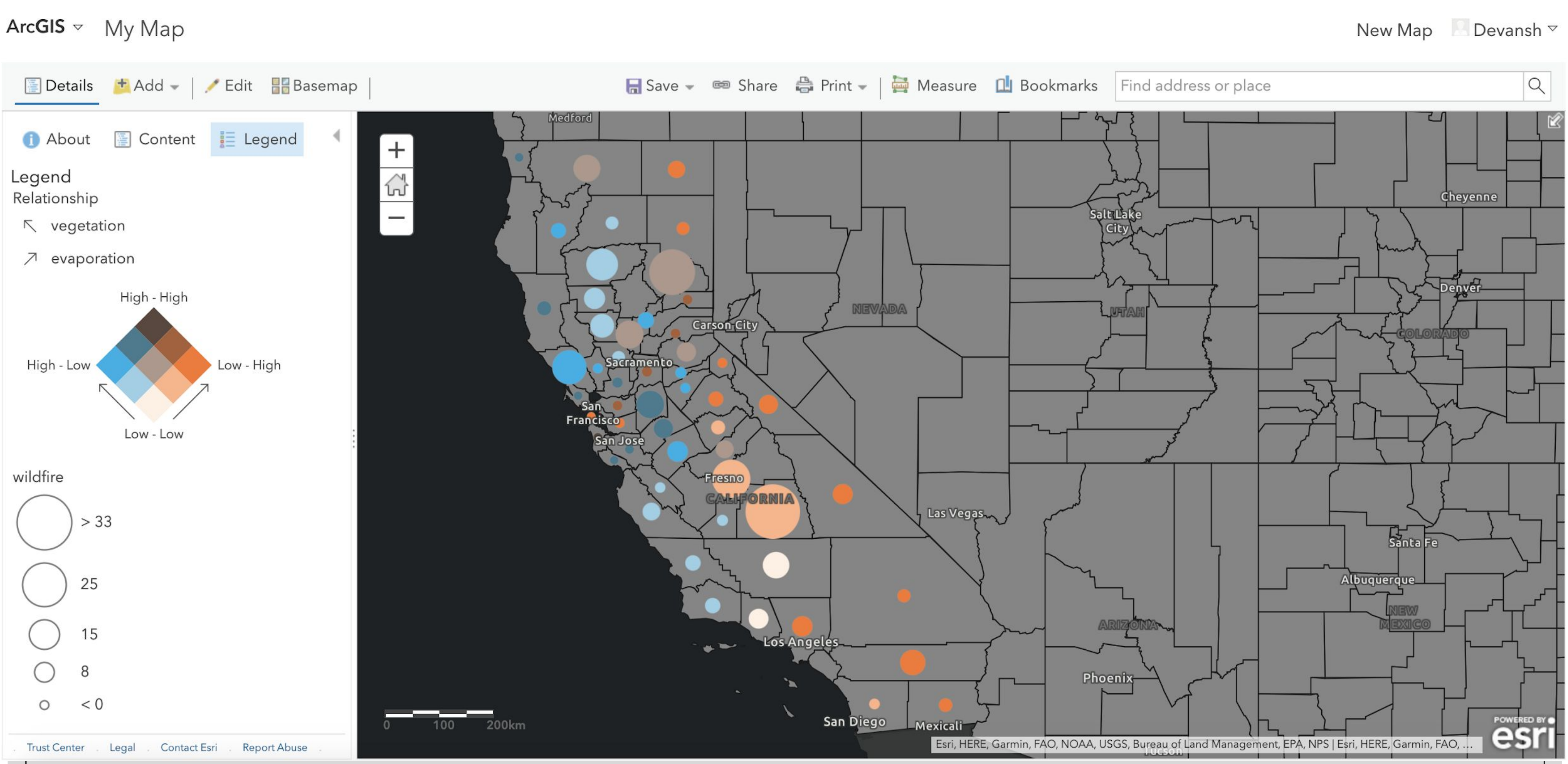
Finding correlation between parameters

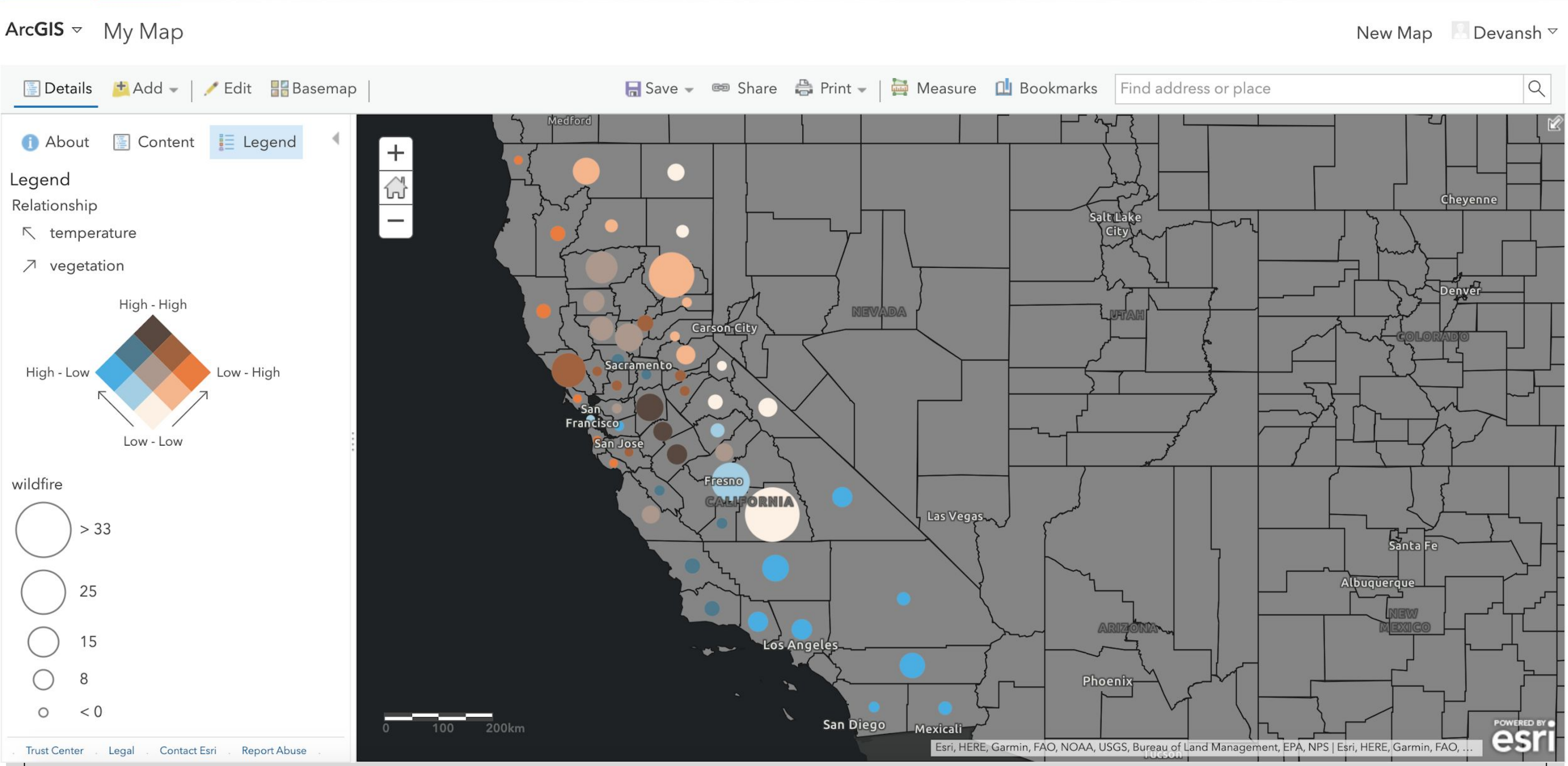
- Once we get the data in the expected format, we find the correlation between fire - vegetation, fire - humidity and fire - temperature using scatter plots.
- We have assumed there is positive correlation between the two attributes, and we want to prove it using the outputs.

Correlation with 2 variables at a time for the counties of California

For this we used ArcGIS to show a 3 way correlation between the parameters.


ArcGIS
My Map
New Map
Devansh
Details
Add
Edit
Basemap
Save
Share
Print
Measure
Bookmarks
Find address or place
About
Content
Legend
Contents
Relationship
temperature
evaporation
High - High
High - Low
Low - High
Low - Low
wildfire
> 33
25
15
8
< 0
Trust Center
Legal
Contact Esri
Report Abuse
Medford
Salt Lake City
Cheyenne
Denver
NEVADA
UTAH
COLORADO
Carson City
Sacramento
San Francisco
San Jose
Fresno
CALIFORNIA
Las Vegas
Santa Fe
Albuquerque
NEW MEXICO
ARIZONA
Los Angeles
Phoenix
San Diego
Mexicali
0
100
200km
Esri, HERE, Garmin, FAO, NOAA, USGS, Bureau of Land Management, EPA, NPS | Esri, HERE, Garmin, FAO, ...
POWERED BY esri


ArcGIS
My Map
New Map
Devansh
Details
Add
Edit
Basemap
Save
Share
Print
Measure
Bookmarks
Find address or place
About
Content
Legend
Legend
Relationship
vegetation
evaporation
High - High
High - Low
Low - High
Low - Low
wildfire
> 33
25
15
8
< 0
Trust Center
Legal
Contact Esri
Report Abuse
Medford
Salt Lake City
Cheyenne
Denver
NEVADA
UTAH
COLORADO
Carson City
Sacramento
San Francisco
San Jose
Fresno
CALIFORNIA
Las Vegas
Santa Fe
Albuquerque
NEW MEXICO
ARIZONA
Los Angeles
Phoenix
San Diego
Mexicali
0
100
200km
Esri, HERE, Garmin, FAO, NOAA, USGS, Bureau of Land Management, EPA, NPS | Esri, HERE, Garmin, FAO, ...
POWERED BY esri


ArcGIS
My Map
New Map
Devansh
Details
Add
Edit
Basemap
Save
Share
Print
Measure
Bookmarks
Find address or place
About
Content
Legend
Legend
Relationship
temperature
vegetation
High - High
High - Low
Low - High
Low - Low
wildfire
> 33
25
15
8
< 0
Trust Center
Legal
Contact Esri
Report Abuse
Medford
Salt Lake City
Cheyenne
Denver
NEVADA
UTAH
COLORADO
Carson City
Sacramento
San Francisco
San Jose
Fresno
CALIFORNIA
Las Vegas
Santa Fe
Albuquerque
NEW MEXICO
ARIZONA
Phoenix
Los Angeles
San Diego
Mexicali
0
100
200km
Esri, HERE, Garmin, FAO, NOAA, USGS, Bureau of Land Management, EPA, NPS | Esri, HERE, Garmin, FAO, ...
POWERED BY esri

Prescribed fires

A controlled or prescribed burn, is a wildfire set intentionally for purposes of forest management.

Detected by the data and is termed as wildfire but actually it is started by humans to control the occurrence of wildfire in that region in future and hence has no relation with vegetation and temperature on that day

How do we verify?

Basically check real-time occurrences

Name	County	Acres	Start date	Containment date	Notes
W-1 McDonald	Lassen	1,020	August 8	August 11	Caused by lightning strike
Gaines	Mariposa	1,300	August 16	August 20	
Mountain	Shasta	600	August 22	August 26	14 buildings destroyed, 7 damaged and 3 people injured
Long Valley	Lassen	2,438	August 24	August 27	
R-1 Ranch	Lassen	3,380	August 28	September 5	Caused by lightning strike
Cow	Inyo, Tulare	1,975	July 25	October 11	Caused by lightning strike
Springs	Mono	4,840	July 26	October 7	Caused by lightning strike
Tucker	Modoc	14,150	July 28	August 15	Unintentionally caused by vehicular traffic along California State Route 139[20][21]
Boulder	San Luis Obispo	1,127	June 5	June 11	
Sand	Yolo	2,512	June 8	June 17	7 structures destroyed, 2 injuries
West Butte	Sutter	1,300	June 8	June 10	

Maria Fire

It was first reported on October 31 in Ventura County

Size: 8,730 acres

Containment: 0%

Getty Fire

It started October 28 in Los Angeles

Size: 745 acres

Containment: 66%

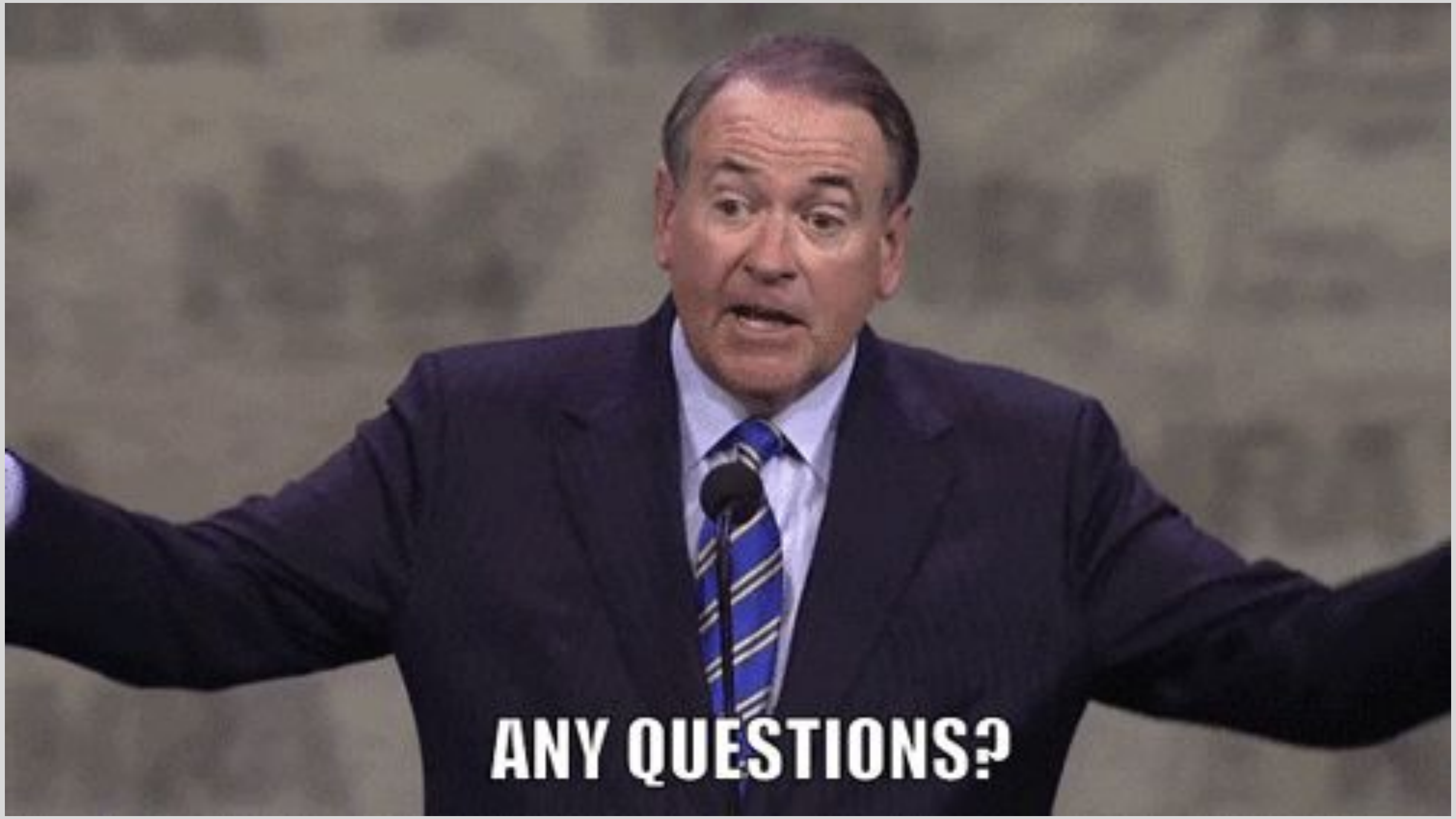
Cause: A eucalyptus tree limb fell into power lines.

Limitations

- Contribution of other factors (like sudden climate changes) to wildfires is ignored which can be significant
- Incorrect data points due to different landscape (specially water bodies)
- Noise in data due to cloud cover
- There will be areas where temperature and vegetation is high and humidity is low but there is no wildfire occurring which currently cannot be verified

Future Work

- Expansion of the idea to the whole world
- Addition of more factors to increase accuracy of the correlation
- Ignoring the prescribed fires to further analyse the occurrence of natural wildfire
- Using the analysis to suggest preventive measures for wildfires


ANY QUESTIONS?


nk you

Max-p-regions Problem

Group 9

Group Member:

Yongyi Liu

Shiyi Zhang

Tong Jia

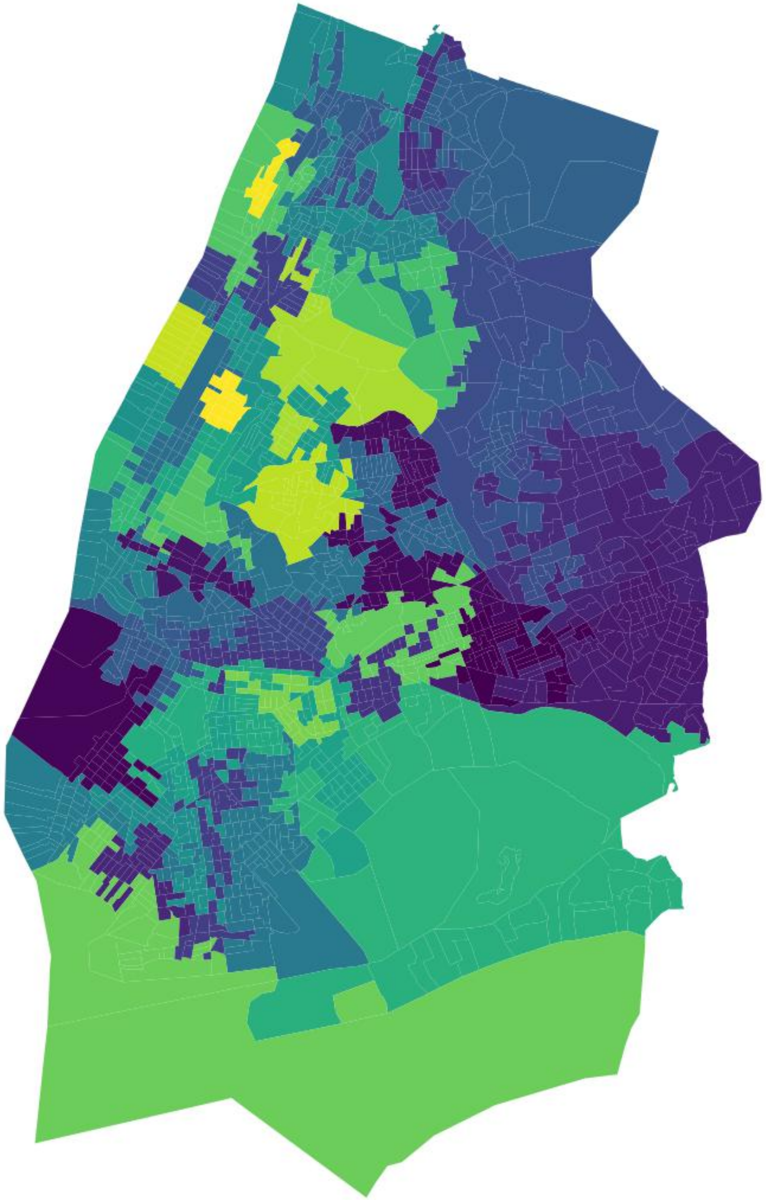
Xiangyu Li

Outline

- Overall Introduction
- Previous Work
- State-of-the-Art
- Our Improvement and Thoughs

Problem Definition

- Extensive attribute: population
- Inner attribute: Income
- Geographical Relationship



Problem Definition

Areas

Let $A = \{A_1, A_2, \ldots, A_n\}$ denote a set $n = |A|$ areas.

Attributes

Let A_{iy} denote the attribute y of area A_i, where $y \in Y = \{1, 2, \ldots, m\}$ with $m \geq 1$; and l_i denote a spatially extensive attribute of area A_i.

Relationship

Let $d : A \times A \to \mathbb{R}^+ \cup \{0\}$ be the dissimilarity between areas based on the set of attributes Y such that $d_{ij} \equiv d(A_i, A_j)$ satisfies the conditions $d_{ij} \geq 0, d_{ij} = d_{ij}$ and $d_{ij} = 0$ for $i, j = 1, 2, \ldots, n$. Distance funtions can also be utilized; i.e., d_{ij} can also satify the subadditivity, or triangle inequality, condition: $d_{ij} \leq d_{ik} + d_{kj}$ for $i, j, k = 1, 2, \ldots, n$.

Let $W = (V, E)$ denote the contiguity graph associated with A such that vertices $v_i \in V$ correspond to areas $A_i \in A$ and edges $\{v_i, v_j\} \in E$ if and only if areas A_i and A_j share a common border. For the max-p-regions model W must be a connected graph.

Feasible Partitions of A

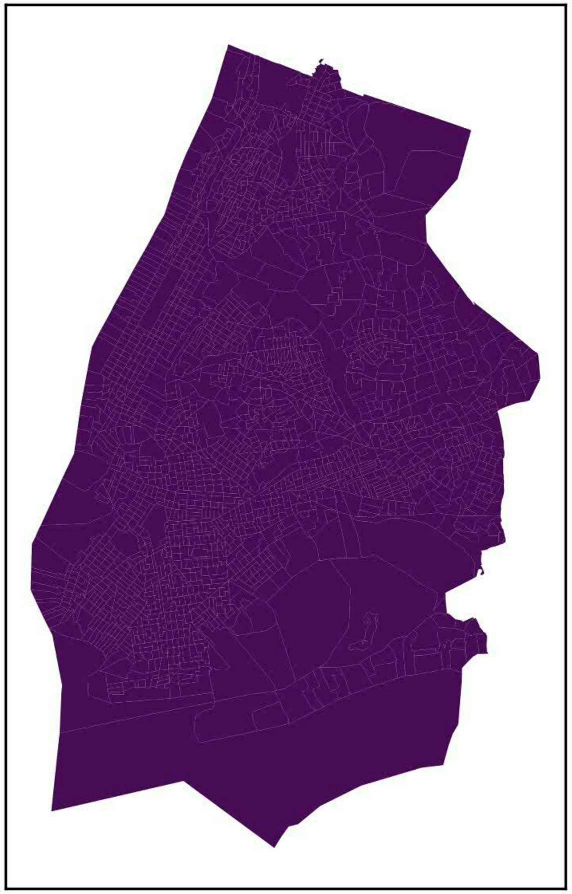
Let $P_p = \{R_1, R_2, \ldots, R_p\}$ denote a partition of areas A into p regions with $1 \leq p \leq n$ such that:

$$|R_k| > 0 \quad \text{for} \quad k = 1, 2, \ldots, p;$$

$$R_k \cap R_{k'} = \emptyset \quad \text{for} \quad k, k' = 1, 2, \ldots, p \ \wedge \ k \neq k';$$

$$\bigcup_{k=1}^{p} R_k = A;$$

$$\sum_{A_i \in R_k} l_i \geq \mathit{threshold} \begin{cases} \text{for } k = 1, 2, \ldots, p, \text{ and} \\ \mathit{threshold} \in \mathbb{R}^+ \cup \{0\} | 0 \leq \mathit{threshold} \leq \sum_{A_i \in A} l_i; \end{cases}$$



Objective Function

(1) $$Z = \left(-\sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k0} \right) * 10^h + \sum_{i} \sum_{j|j>i} d_{ij} t_{ij}.$$

Subject to:

(2) $$\sum_{i=1}^{n} x_i^{k0} \leq 1 \qquad \forall k = 1, \ldots, n;$$

(3) $$\sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{kc} = 1 \qquad \forall i = 1, \ldots, n;$$

(4) $$x_i^{kc} \leq \sum_{j \in N_i} x_j^{k(c-1)} \qquad \forall i = 1, \ldots, n;\ \forall k = 1, \ldots, n;\ \forall c = 1, \ldots, q;$$

(5) $$\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{kc}\, l_i \geq \text{threshold} * \sum_{i=1}^{n} x_i^{k0} \qquad \forall k = 1, \ldots, n;$$

(6) $$t_{ij} \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} - 1 \qquad \forall i, j = 1, \ldots, n | i < j;\ \forall k = 1, \ldots, n;$$

(7) $$x_i^{kc} \in \{0, 1\} \qquad \forall i = 1, \ldots, n;\ \forall k = 1, \ldots, n;\ \forall c = 0, \ldots, q;$$

(8) $$t_{ij} \in \{0, 1\} \qquad \forall i, j = 1, \ldots, n | i < j.$$

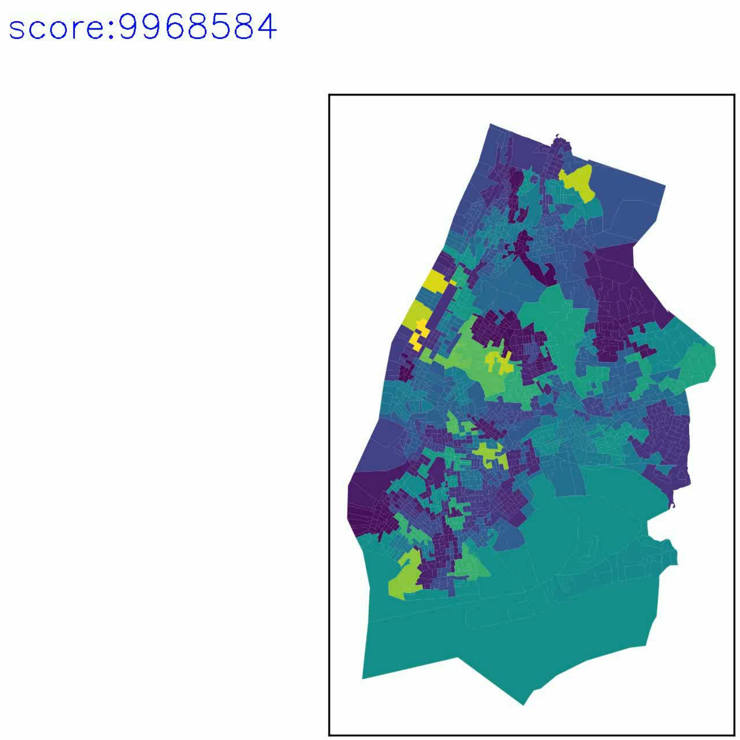
Previous Work

- a. apply the conventional clustering algorithm without considering the geographical location and define regions as the subset of spatially contiguous areas assigned to the same cluster
- b. Constructing homogenous regions by including x and y coordinates of the centroids of the areas as two additional attributes in the clustering algorithm.
- Problem in the previous work: the spatial contiguity of different region is not guaranteed

State-of-the-Art

- 1. Construction
- -> growing regions + assigning enclaves

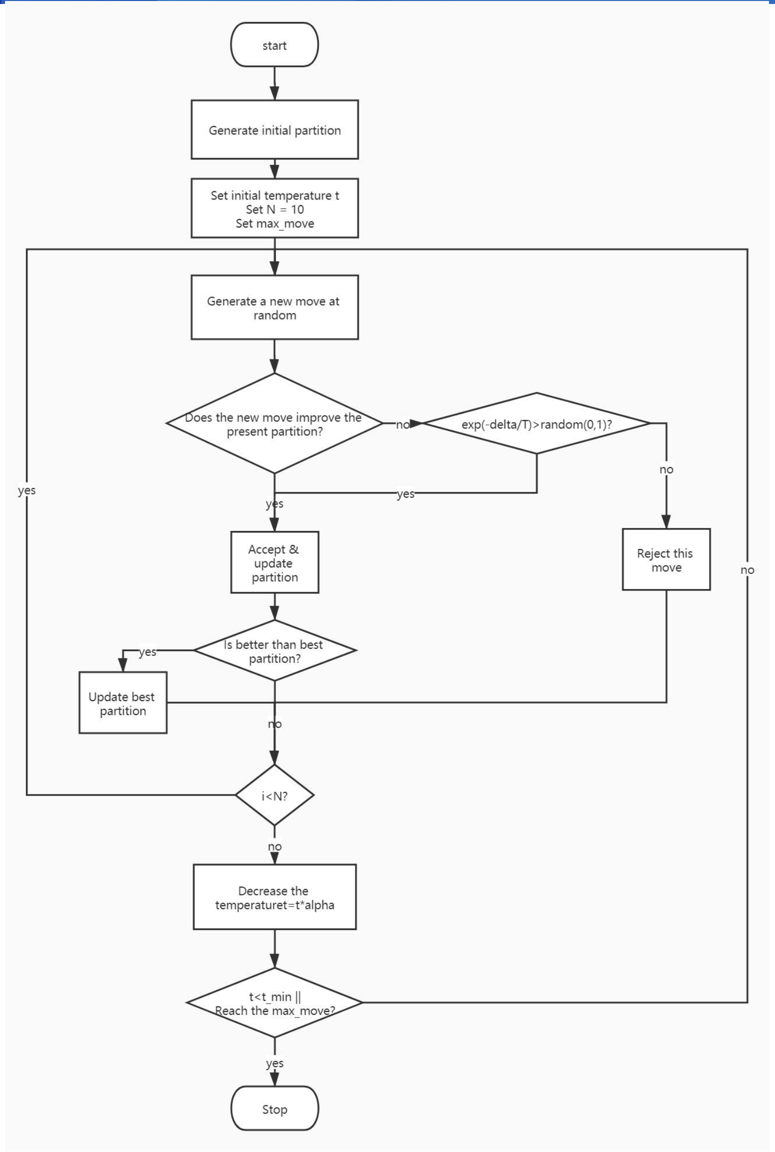
- 2. Local Search Phase
- -> a. Simulated Annealing
- -> b. Tabu Search



Simulated Annealing

- Pros:
- Escaping from the local minimum

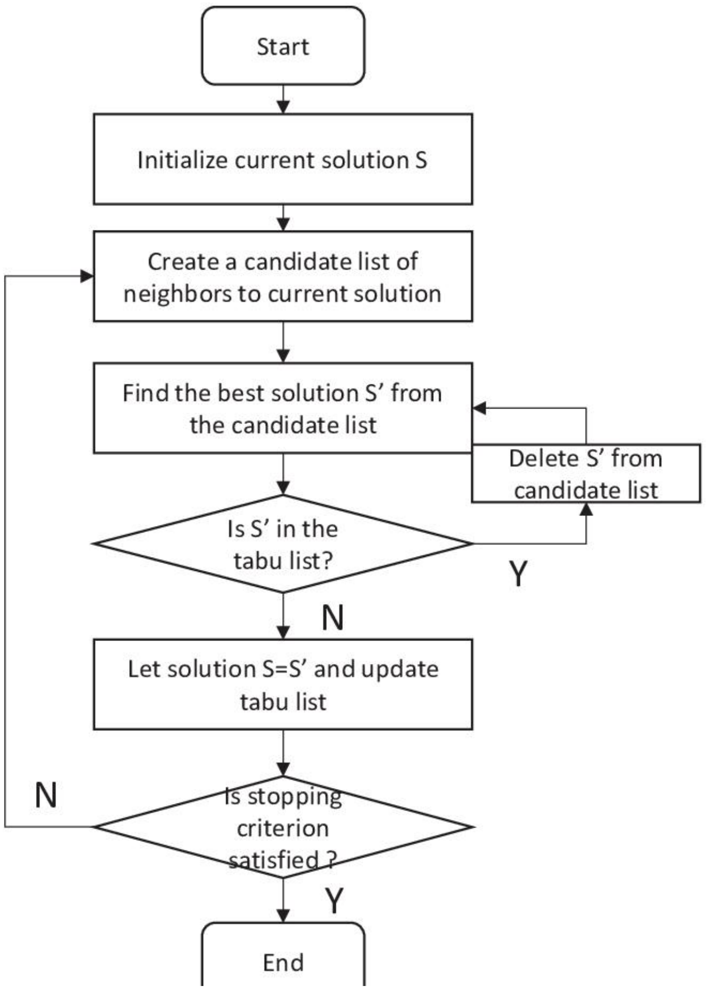
- Cons:
- Randomness (The annealing phase to hard to control.)
- Time-consuming (based on the parameter).



Tabu Search

- Pros:
- Escaping from the local minimum

- Cons:
- Relate on Initial Solution



Our Contribution

- Improvement in the construction phase

  -test how threshold affects the performance

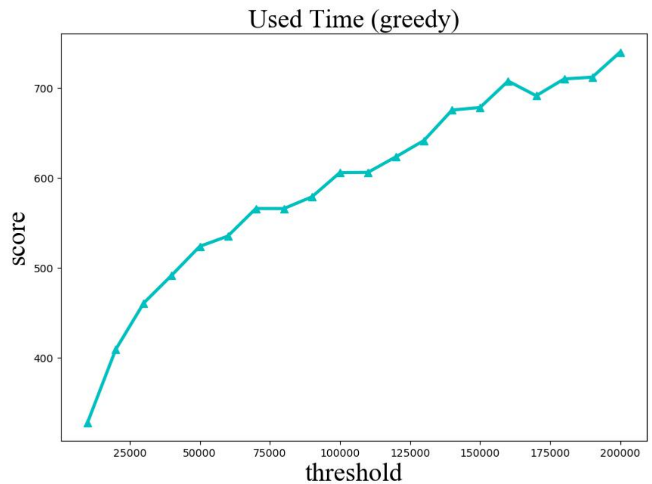
  -change the O(n^2) greedy method in the growing phase to a O(n) method that finds out the area with minimum extensive attribute

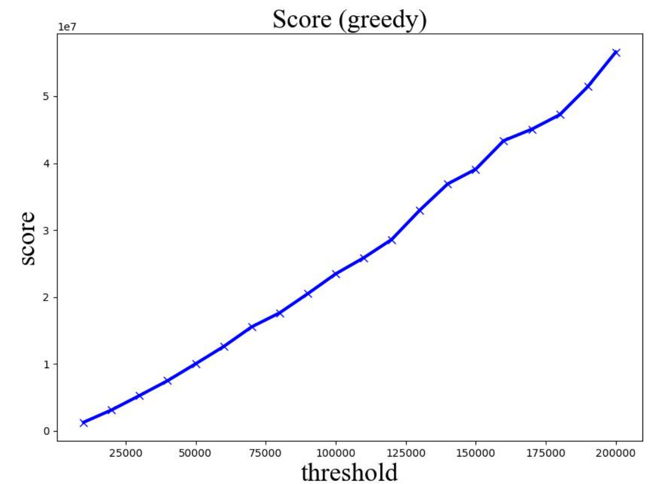
- Improvement in the local search phase

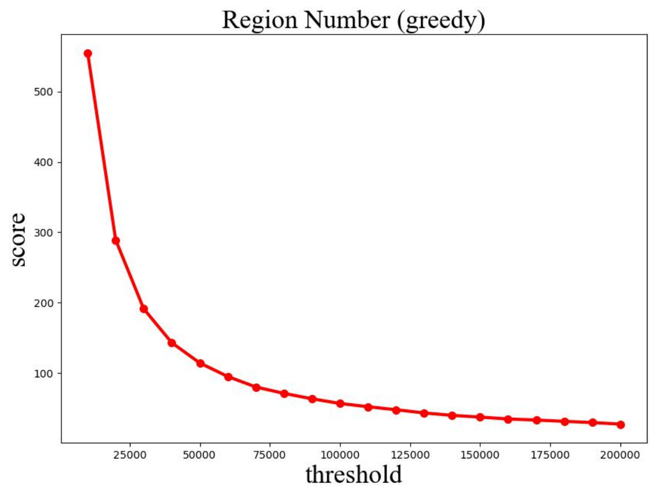
  -modified the simulated annealing search

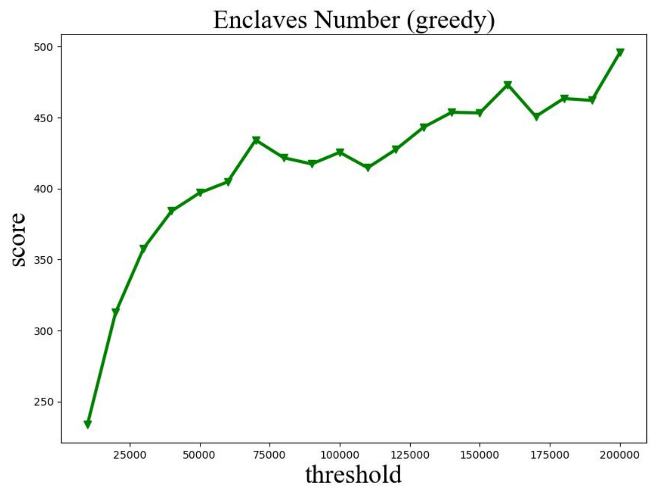
  -proposed a hybrid heuristic method that involves the advantage of SA, Tabu search and greedy method

How threshold affects performance

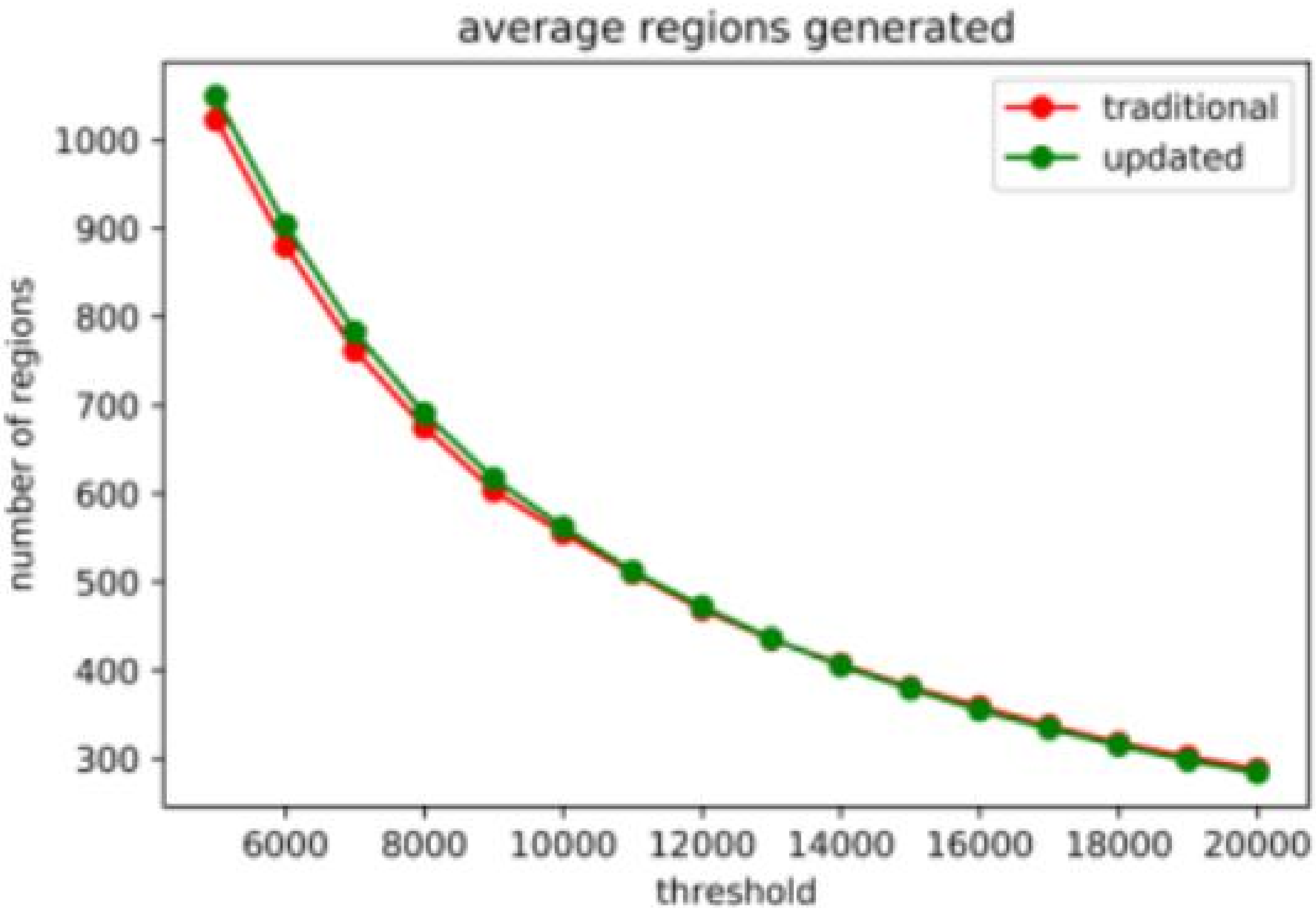








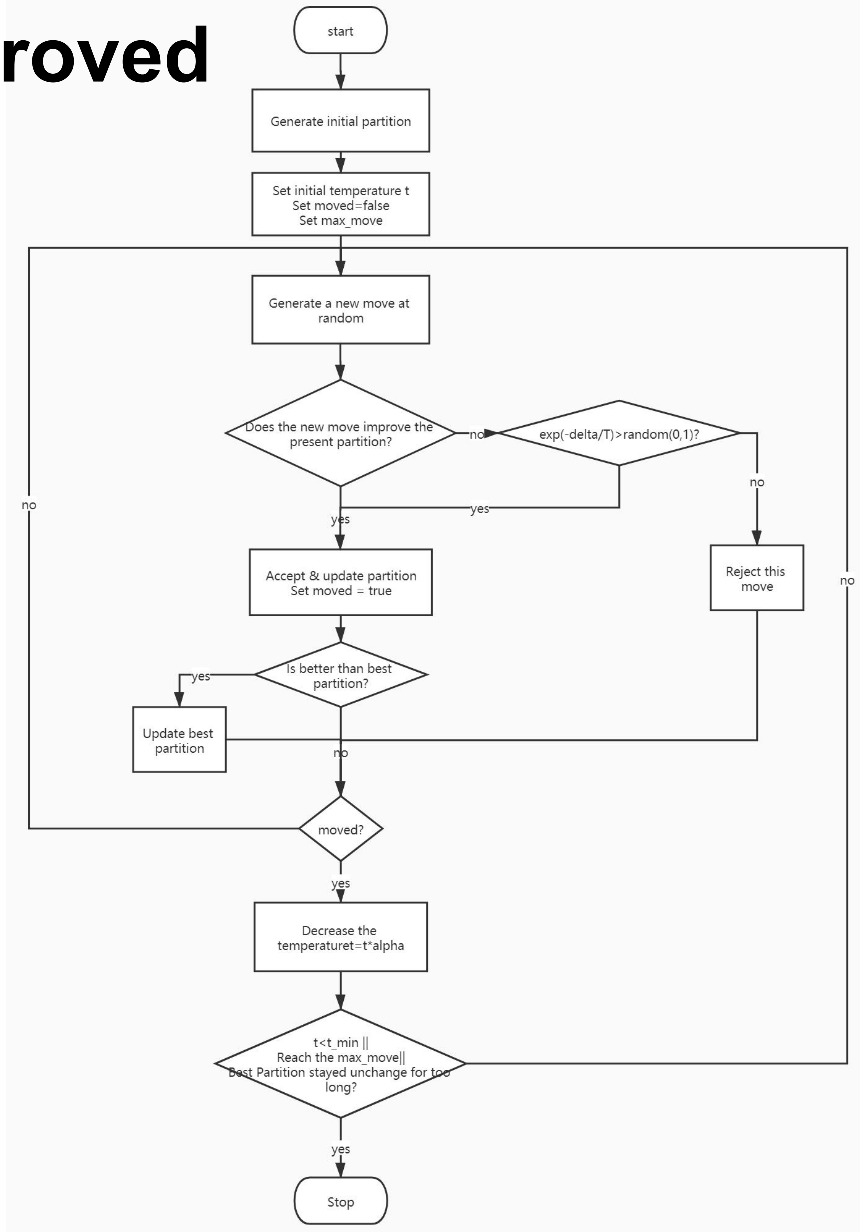
Changing O(n^2) to O(n)



Simulated Annealing Improved

- Improvements:
- Replace constant loop number with flexible loop number
- Stop the loop in advance to save time

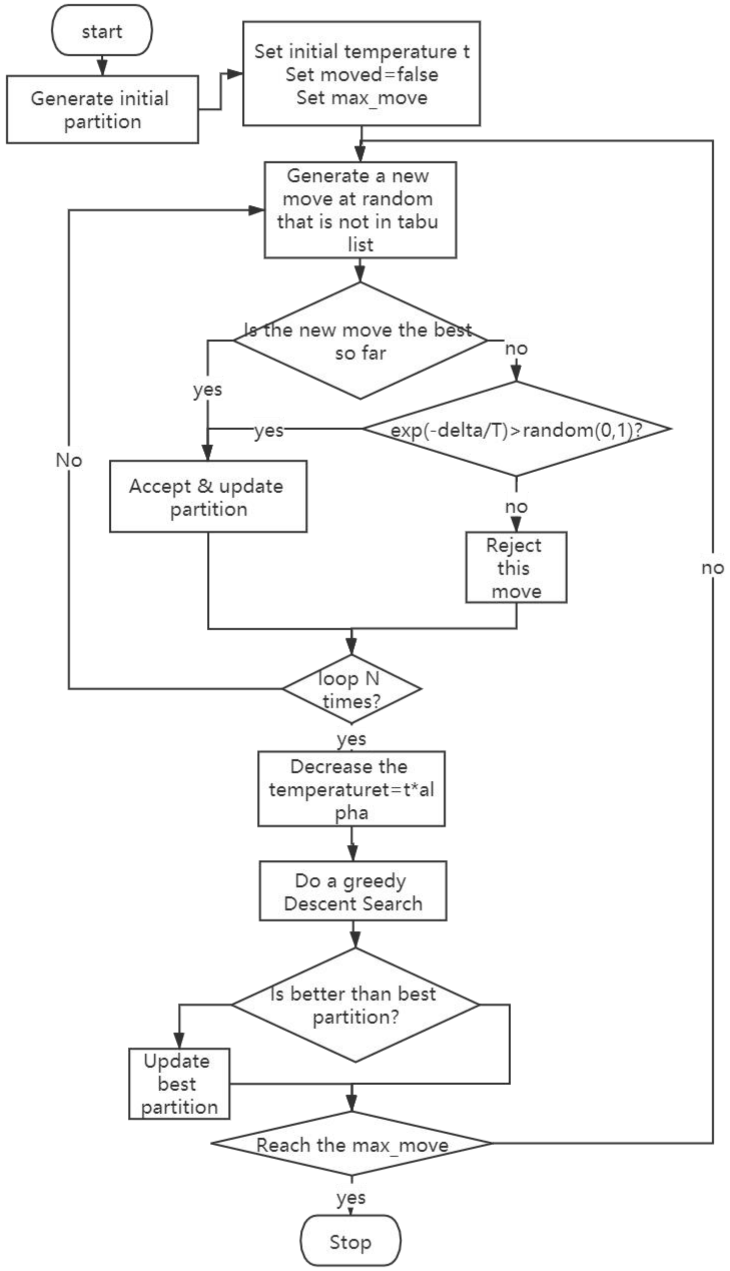
- Pros:
- Use less time



Mixed Search

- Improvements:
- Combine the Simulated Annealing Search with Tabu Search
- Add a Greedy descent to make sure we won't miss a good result

- Pros:
- 1. Escape local minimum
- 2. Avoid repeat search
- 2. Don't relay on Initial solution
- 3. Won't miss good solutions



Experimental Result

METHOD	TIME CONSUMED	WRD (BEFORE)	WRD (AFTER)	DIFF	IMPROVEMTN PERCENTAGE
TABU	25189.7	10130302.1	9742281.3	388020.8	3.83%
SA	106700.8	10130302.1	9825322.2	304979.9	3.01%
SA+	16068.8	10130302.1	9908254.6	222047.5	2.19%
MIXED	95968.8	10130302.1	9737182.8	393119.3	3.88%

Thanks

Scalable Max-P Regions

Faisal Almaarik
Hussah Alrashid

Outline

- Introduction
    - Problem Statement
- Related Work
- Scalable Max-P Regions
    - Data Preprocessing
    - General Approach
- Results/Future Work

Problem Statement

- What is the problem?
- Why Max-P regions is a unique Model?
- What are the challenges ?

What Is The Problem

- Homogenous regions
  - set of spatially contiguous areas that provide a high degree of similarity in the attributes
- Aggregation multiple areas into regions are not easy
  - shape of regions
  - the equality of an attribute value across the regions
  - membership of constraint
- Max-P regions model.

Why Max-P Regions Is A Unique Model?

- The number of regions is modeled as an endogenous parameter.
- The data form the final shape of each regions.
- Analytical study for regions.
  - Crime rate
  - Unemployment rate

What Are The Challenges?

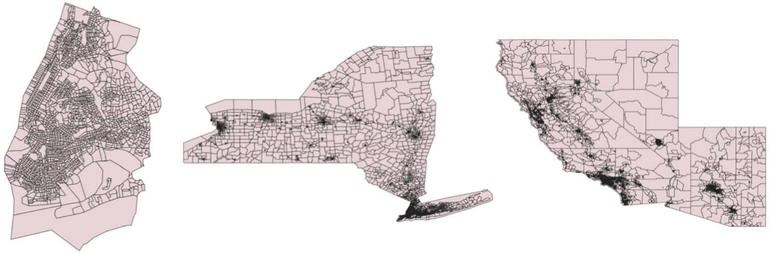
- Trapped in the local search algorithm looking for optimal solution.
- Finding feasible neighboring solutions efficiently.
- Computational time

Related Work

Heuristic Search Algorithm				Region Compactness		Feasible Solution Technique		
Simulated Annealing	Tabu	Greedy	Parallel	NMI	Corridor Region Formula	Move	Swap	Merge

Scalable Max-P Regions: Data Preprocessing

- Data Preprocessing:
  - Removing islands.
  - Merging different datasets.
- Datasets:
  - New York City - 2056 Areas
  - New York State - 4918 Areas
  - California + Arizona + Nevada States - 10263 Areas



(a) NY City (2056 Areas) (b) NY State (4096 Areas) (c) CA, NV, AZ States (10263 Areas)

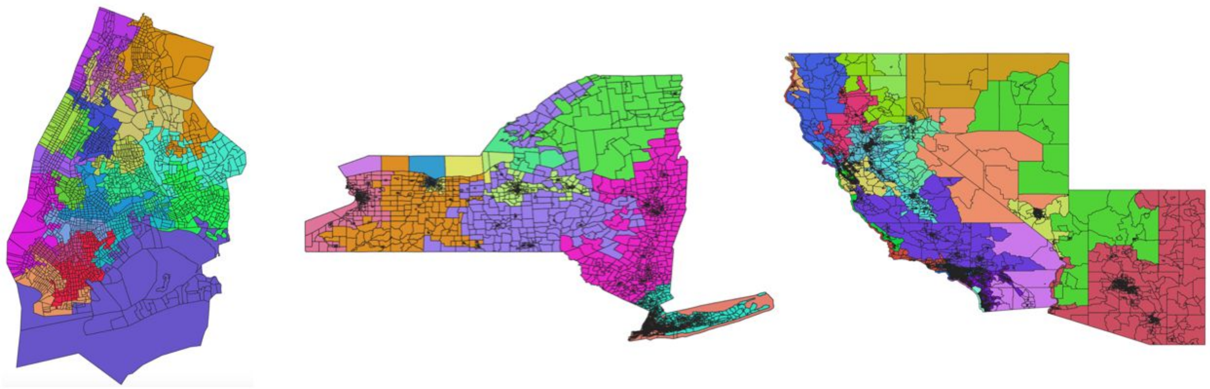
Scalable Max-P Regions: General Approach

- Construction Phase:
  - Grow Regions
    - Select seed area.
    - Add areas that reduces region dissimilarity.
    - Stops when threshold reaches a predefined value.
  - Enclaves Assignment
    - Select an enclave that has at least one assigned area as a neighbor.
    - Find neighboring regions.
    - Calculate neighboring regions dissimilarity when the enclave is added.
    - Enclave is added to the most similar region.

Scalable Max-P Regions: General Approach

- Local Search Phase (Tabu):
  - Find feasible neighbors by moving one area to another region
    - Create a list of border areas and their neighboring regions.
    - Move each area in the list to all the neighboring regions if it satisfies the following conditions:
      - Region size > 1
      - New threshold > predefined threshold
      - Spatial contiguity is preserved
  - Moves are taken until maximum number of iterations is made without improving the dissimilarity measure.

Results

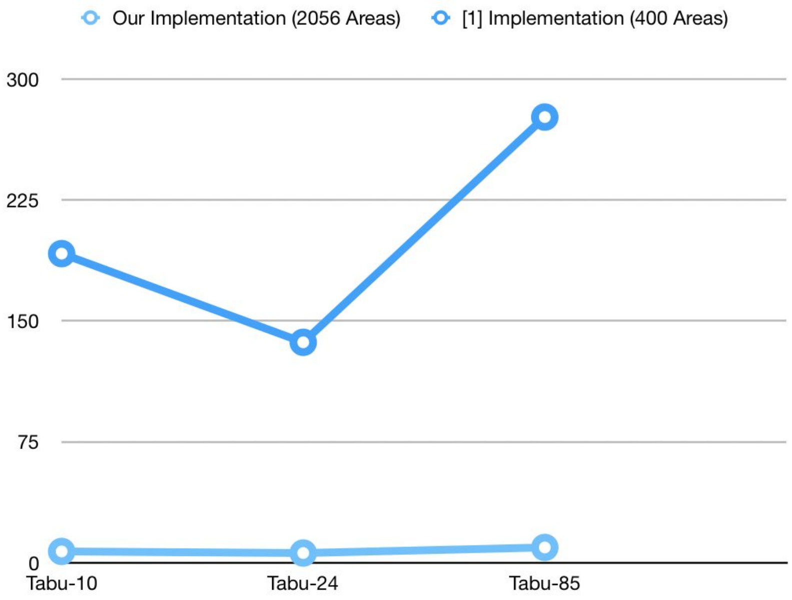


(a) NY City (2056 Areas) (b) NY State (4096 Areas) (c) CA, NV, AZ States (10263 Areas)

Figure 6: Dataset After Applying Max-P Regions

Results

Our Implementation when compared to the smallest dataset of [1]'s Implementation: 2056 vs 400 areas.



Future Work

- Partition the data and parallelize the grow regions phase.
- Test the implementation against larger datasets.

SCALABLE RANGE QUERY FOR POLYGON LAYERS

Laila Abdelhafeez
Carter Slocum

Outline

1. Problem definition
2. Limitations of the related work
3. Challenges
4. Contribution
5. Evaluation
6. Front End
7. Demonstration

SpatialSpark

GeoSpark

Hadoop-GIS

Spatial Big Data Solutions

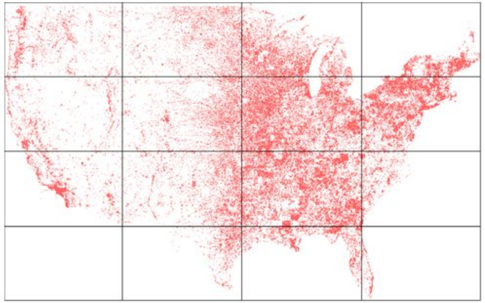
Simba

Magellan

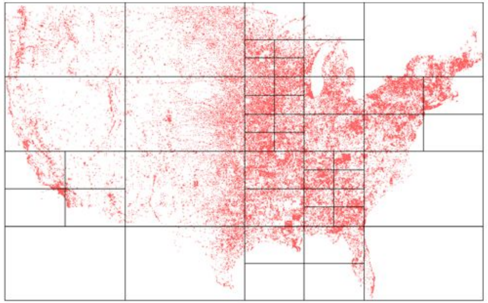
LocationSpark

Big spatial data systems

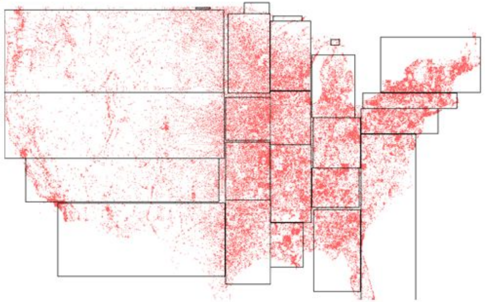
Data Partitioning — Global Index



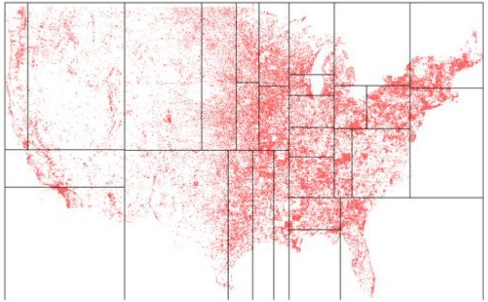
a) SRDD partitioned by uniform grids



b) SRDD partitioned by Quad-Tree



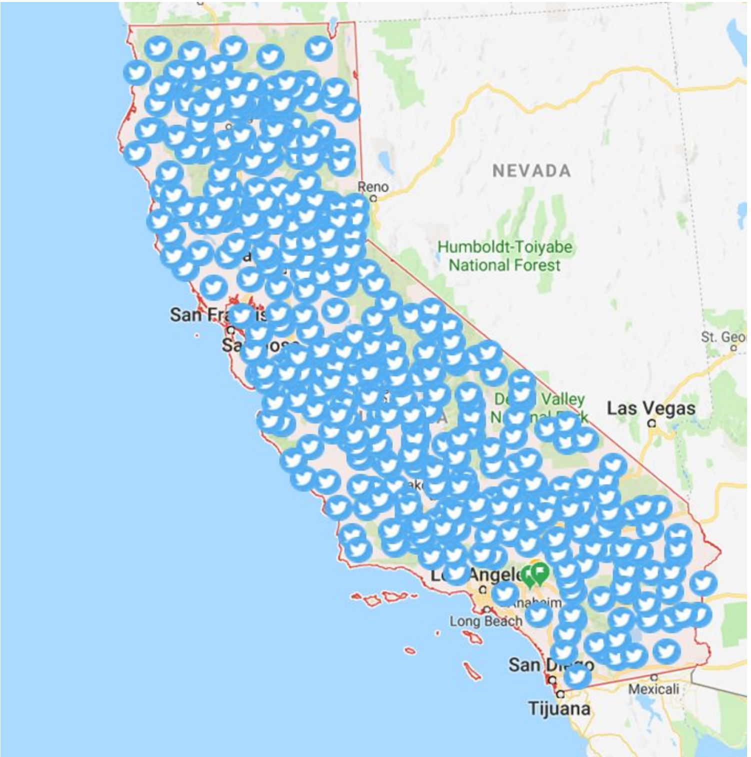
c) SRDD partitioned by R-Tree



d) SRDD partitioned by KDB-Tree

Problem definition

Given arbitrary boundary points for a polygon *"California"*, retrieve all data points *"tweets"* that are located inside the query polygon boundary.

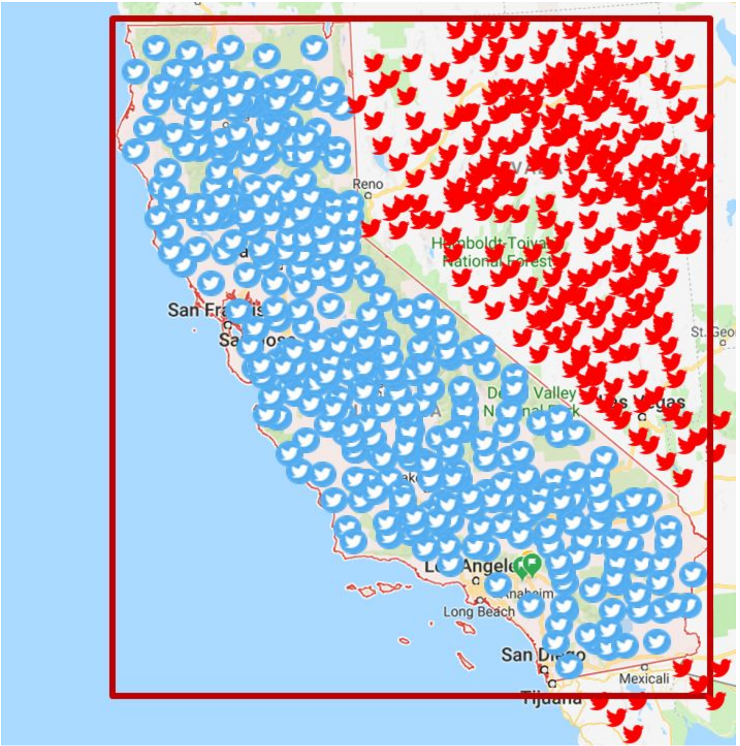


Filter-refine approach

Filter-refine approach employs two stages:

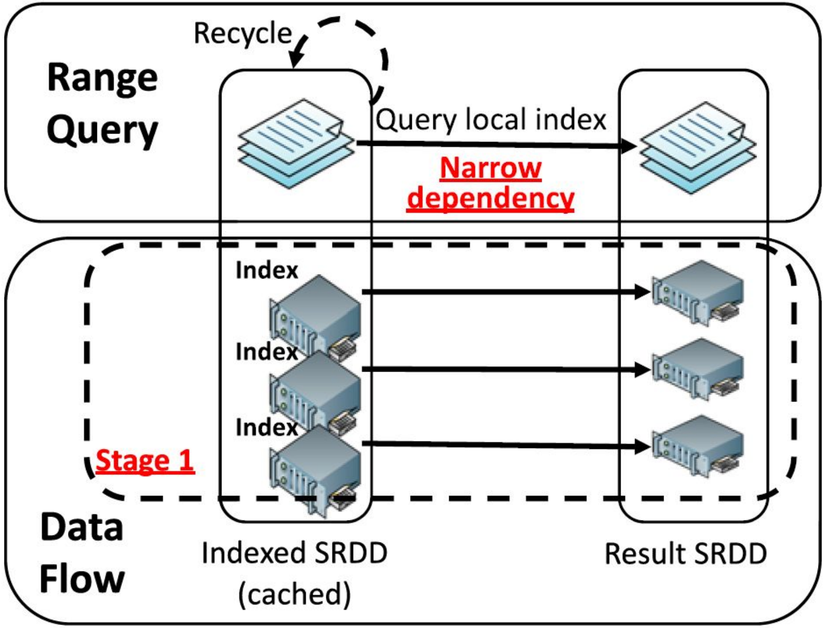
1. Filter the data using MBR — minimum bounding rectangle. Only the data within the MBR, **candidate set**, can be within the polygon.
2. Refine the candidate set using the exact geometry using the point in polygon operation.

   ***Blue tweets are within California but red tweets are not.*



Filter-refine in GeoSpark

1. Broadcast query polygon to all partitions.
2. Perform filter-refine on each partition of the data.



Challenges

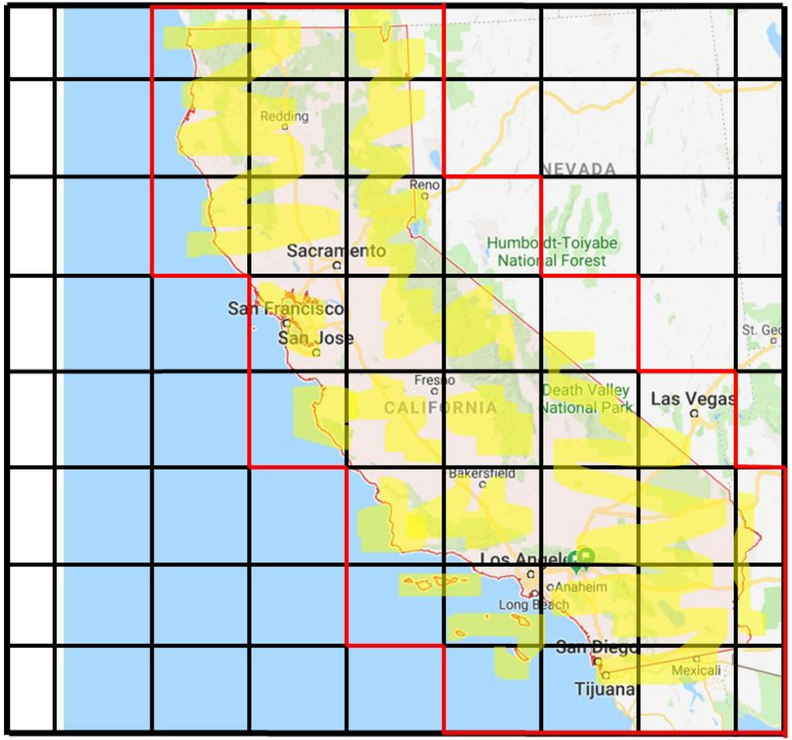
1. Point-in-polygon operation is **computationally expensive** for complex polygons.
2. BIG DATA leads to a huge number of point-in-polygon operation for just one polygon.
3. Processing polygon layers will result intolerable processing time.

Proposed solution

Make use of the local index built on each partition.

In analogy of divide and conquer approach:

1. Decompose the complex polygon into set of smaller polygons by clipping the query polygon with the underlying index.
2. Clipping the polygon will result three possible states:
   a. Grid cell is wholly contained in query polygon. (Add all cell data to result set)
   b. Grid cell is outside the query polygon. (Ignore all data in this cell)
   c. Grid cell intersects the query polygon resulting a smaller polygon. (perform filter-refine on the resulting polygon)

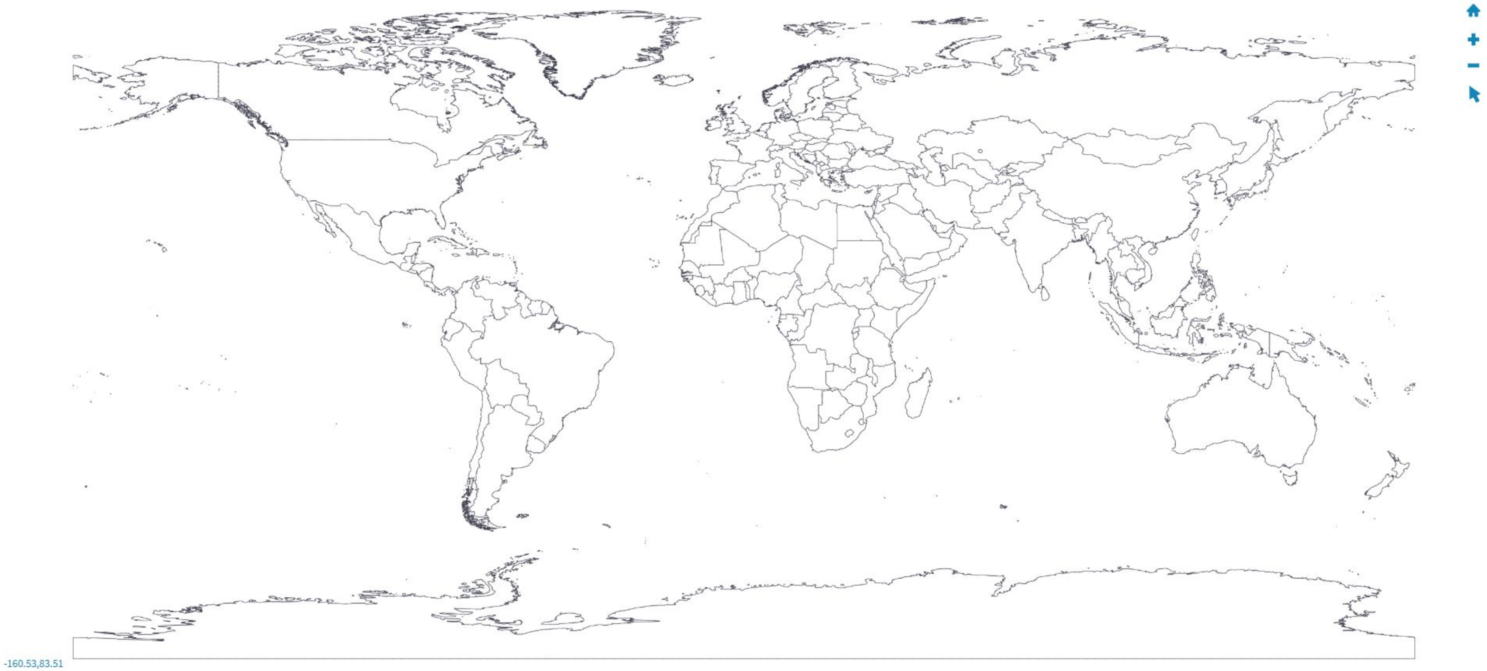


Missing problem?

- The previously mentioned solution, reduces the computation time of a single polygon query.
- An executor works on a partition of the data.
- A single polygon will appear in a small number of partitions, leaving other executors idle.
- Leverage these idle executors to process other polygons that are in different locations.

Polygon layer query

Given a polygon layer *"Countries of the world"*, retrieve all data points that is located inside each polygon separately.



Proposed solution cont.

1. Partition the polygons into the partitions that include them.
2. Clip the polygon with the partition boundary, to get a smaller polygon that only intersects with the current partition.
3. Process the polygons of each partition.
4. Group the results.

Polygon layer results

	Continents	Countries	States	Counties	Postal codes
File size	4 MB	12 MB	27 MB	1.1 GB	1.2 GB
Records	8	255	4,489	45,961	152,908
Source	ARCGIS	NE	NE	GADM	UCR STAR
Total time	7.6 minutes	7 minutes	8 minutes	14.6 minutes	9.7 minutes
Average size per record	0.5 MB	0.05 MB	0.006	0.025	0.008
Polygon partitioning time	1.9 minutes	50 seconds	9 seconds	36 seconds	27 seconds
Range query time	5.7 minutes	6.2 minutes	7.8 minutes	14 minutes	9.2 minutes

Displaying Results

1. Polygon Data are stored as large lists of Lat/Lon coordinates. (NA is over 60k)
2. Users may not wish to store multiple Gigabytes of polygon data for quick entry.
3. Need to quickly and Intuitively display Polygons and query results

Spatial Polygon Front-End

- Browser Based front end to display Map + polygons.
- Intuitive query on click.
- Far less data in memory and over Network.

Front End Challenges

- Cannot Load 6 GB of polygon data every page.
- Cannot send multiple MB of query data to back-end over Network
- Must reserve shape to overlay on Google Maps.
- Need to Map from Lat/Lon to Pixel coords

Spatial Polygon Front-End

- Built Using Google Maps Javascript API for Background map and Polygon Scaling
- Mapshaper to prune vertices while maintaining shape and polygons (Visvalingam)

Front End Results

- Data loaded reduced by 99.5%
- 3610 polygons converted and drawn in roughly one second
- Fully functional Interface for issuing Queries.