911 Emergency Call Prediction

Presented by Ya-Ting (Tina) Chang

Agenda

- Data Preprocessing
- Exploratory Data Analysis
- Predictive Modeling
- Conclusion

Data Preprocessing

Selected Datasets

National Climate Data Center

- Climate records from stations across the globe
- From year 2008 to 2018
- Daily records (size: 1.7GB)

Emergency-911 call

- Provided by montcoalert.org
- Available on Kaggle
- Montgomery County, PA
- Dec 2015 to March 2018
- 326,425 daily records

Data Preprocessing - NCDC

- Focus on year 2017
- Select stations in US
- Remove missing value in columns: STATE, PRCP, TMAX, TMIN
- Transform TMAX and TMIN to Fahrenheit degree scale
- Finalized data format

id	date	PRCP	TMAX	TMIN	lat	long
USW00094732	2017-01-01	0	53.06	30.02	40.0819	-75.0111

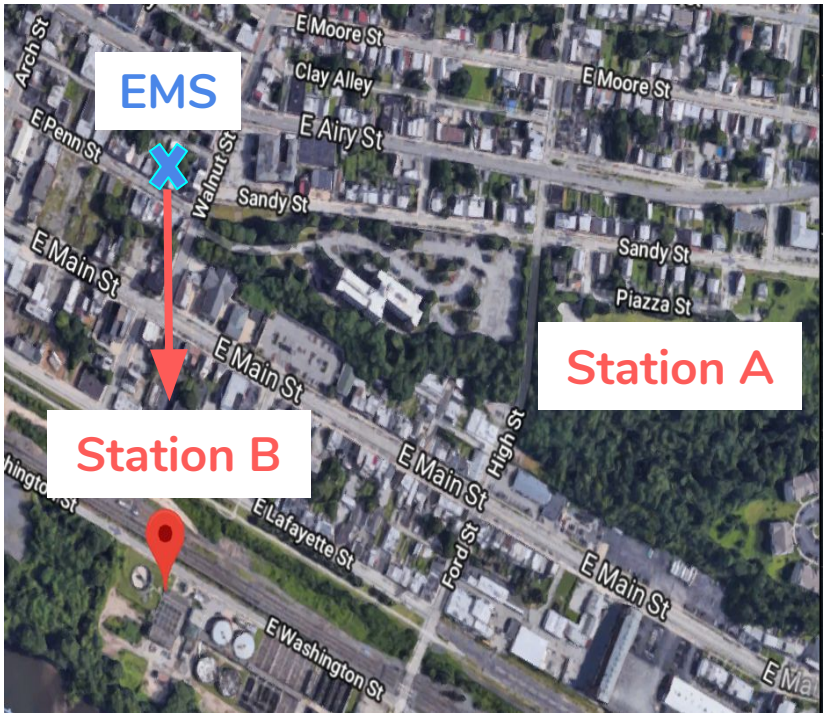
Data Preprocessing - 911 Call

- Focus on year 2017
- Three main categories: EMS, Fire, Traffic
- Too many categories in TOWNSHIP and ZIP CODE:
  - keep the most common 30 townships and aggregate the remaining into 'others'
  - extract the first 3 numbers in zip code, i.e. 19525 ⇒ 195
- Finalized data format

date	lat	long	Zip	township	month	week	weekday	type
2017-12-31	40.39682	-75.49774	180	PENNSBURG	12	53	1	EMS

Weather in location of emergency call

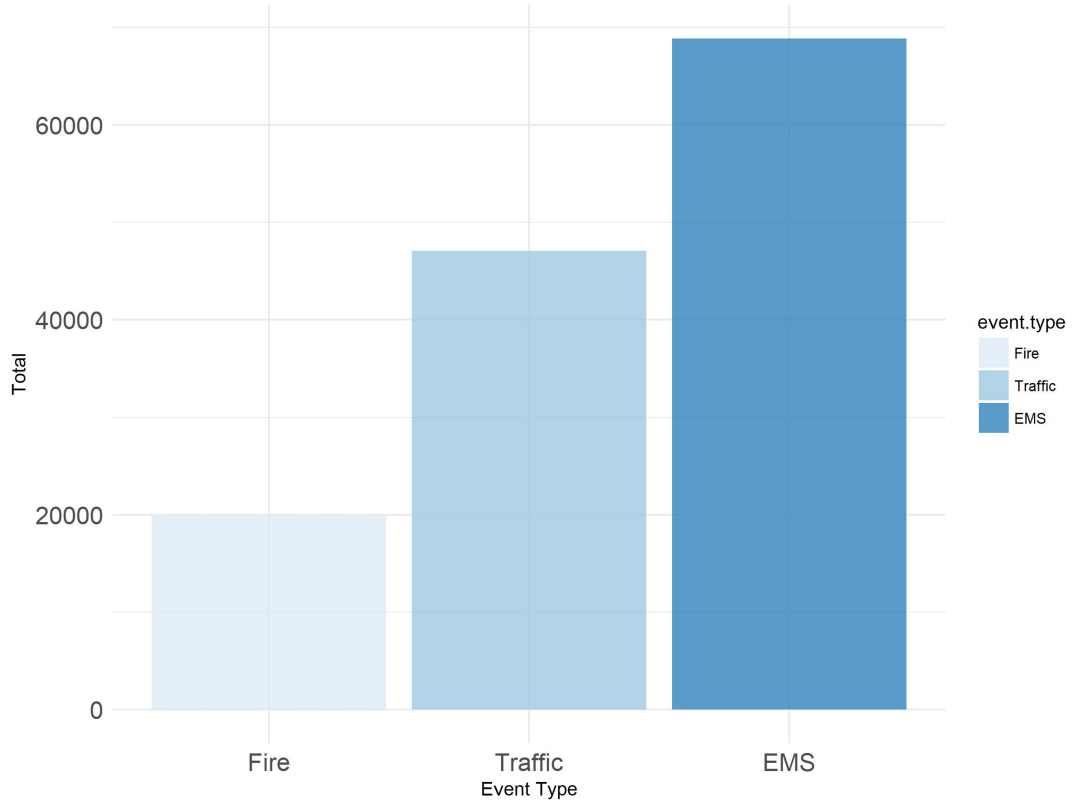
The weather in the emergency event's place is determined by the nearest station's record



Exploratory Data Analysis

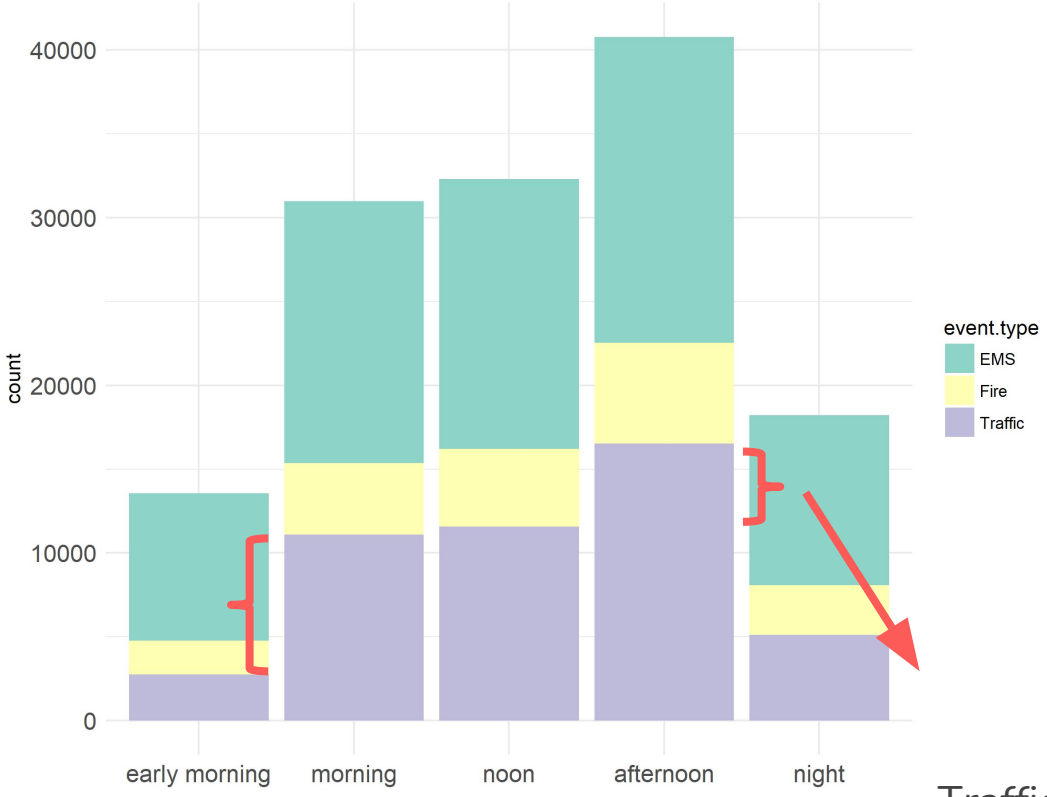
1. Time Factor
2. Location Factor
3. Temperature Factor

Total Number of Events



Fire	48,919 (15.0%)
Traffic	116,065 (35.6%)
EMS	161,441 (49.4%)

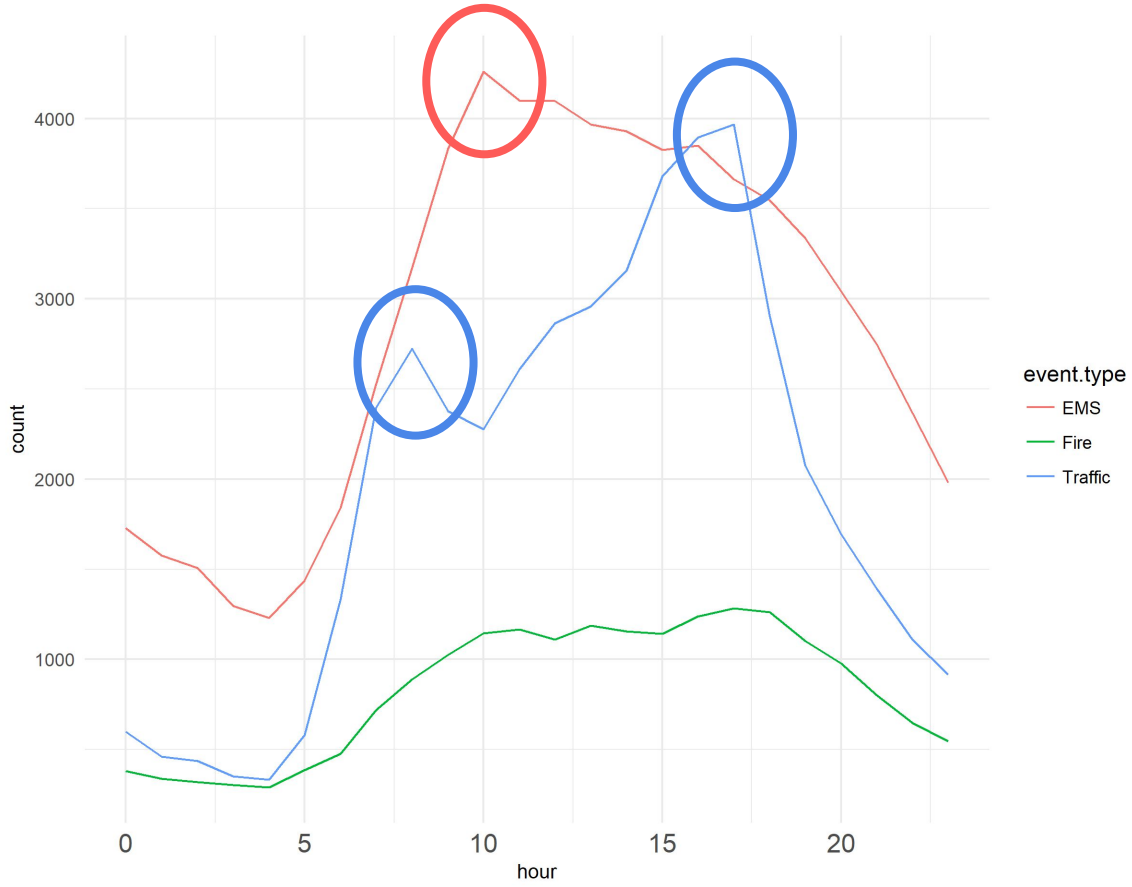
Occurrences in Time of Day



Early morning	0 am ~ 5 am
Morning	5 am ~ 10 am
Noon	10 am ~ 14 pm
Afternoon	14 pm ~ 19 pm
Night	19 pm ~ 24 pm

Traffic emergency increased greatly in morning and afternoon

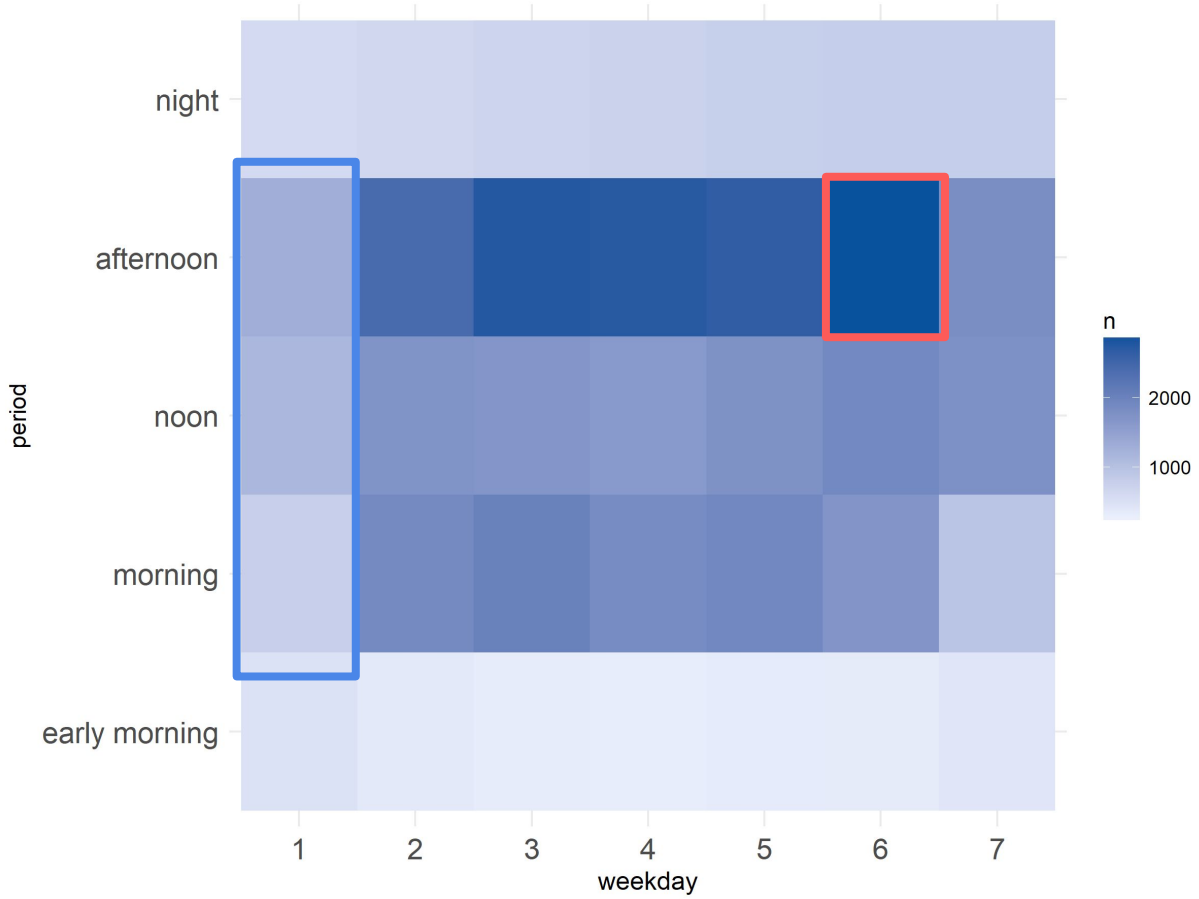
Trend in Hourly Occurrences



Two peaks of Traffic
1. go work (8 - 9)
2. back home (17 - 18)

An obvious peak of EMS at 10 am

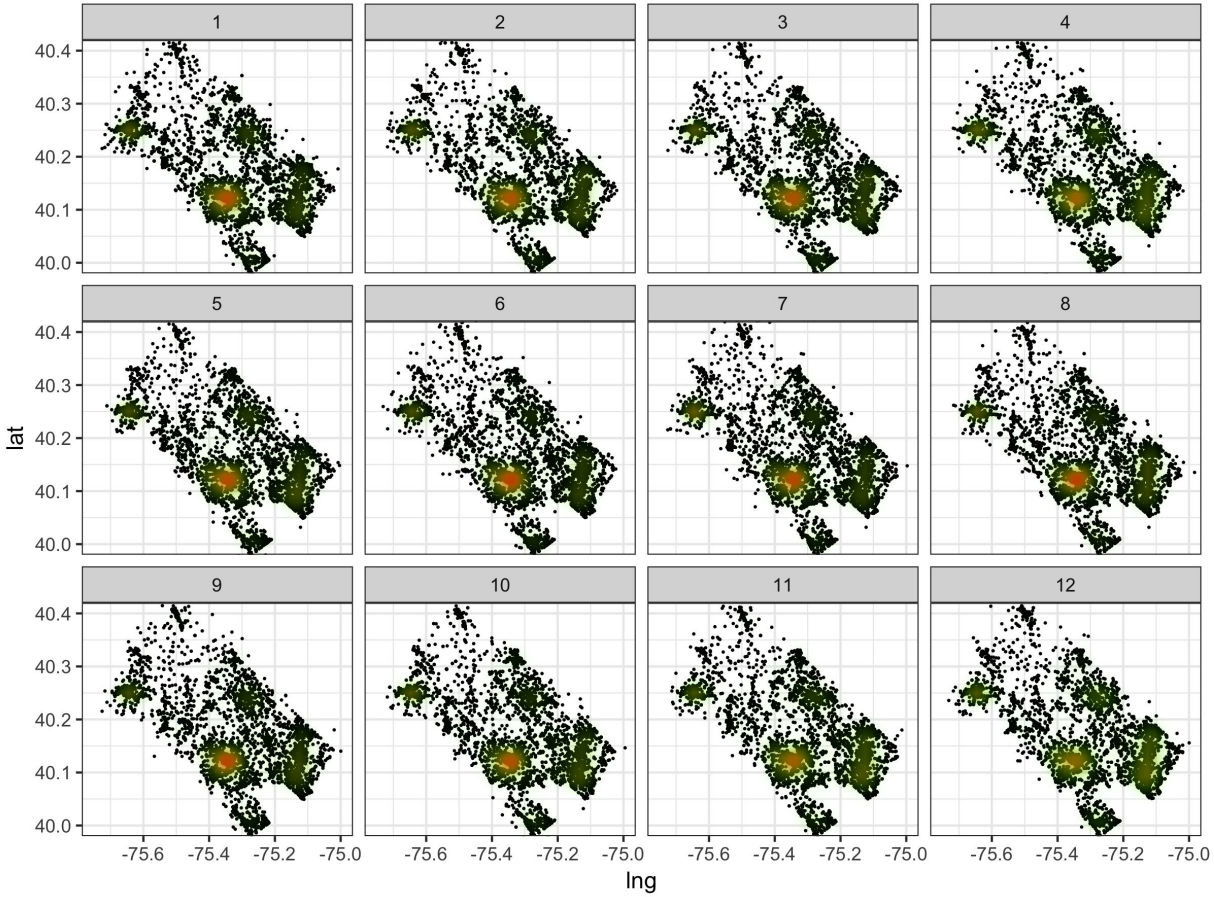
Heatmap of weekday and time of day

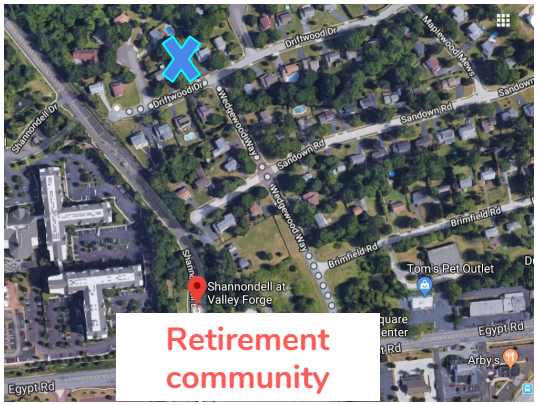


Least emergency call happened on Monday

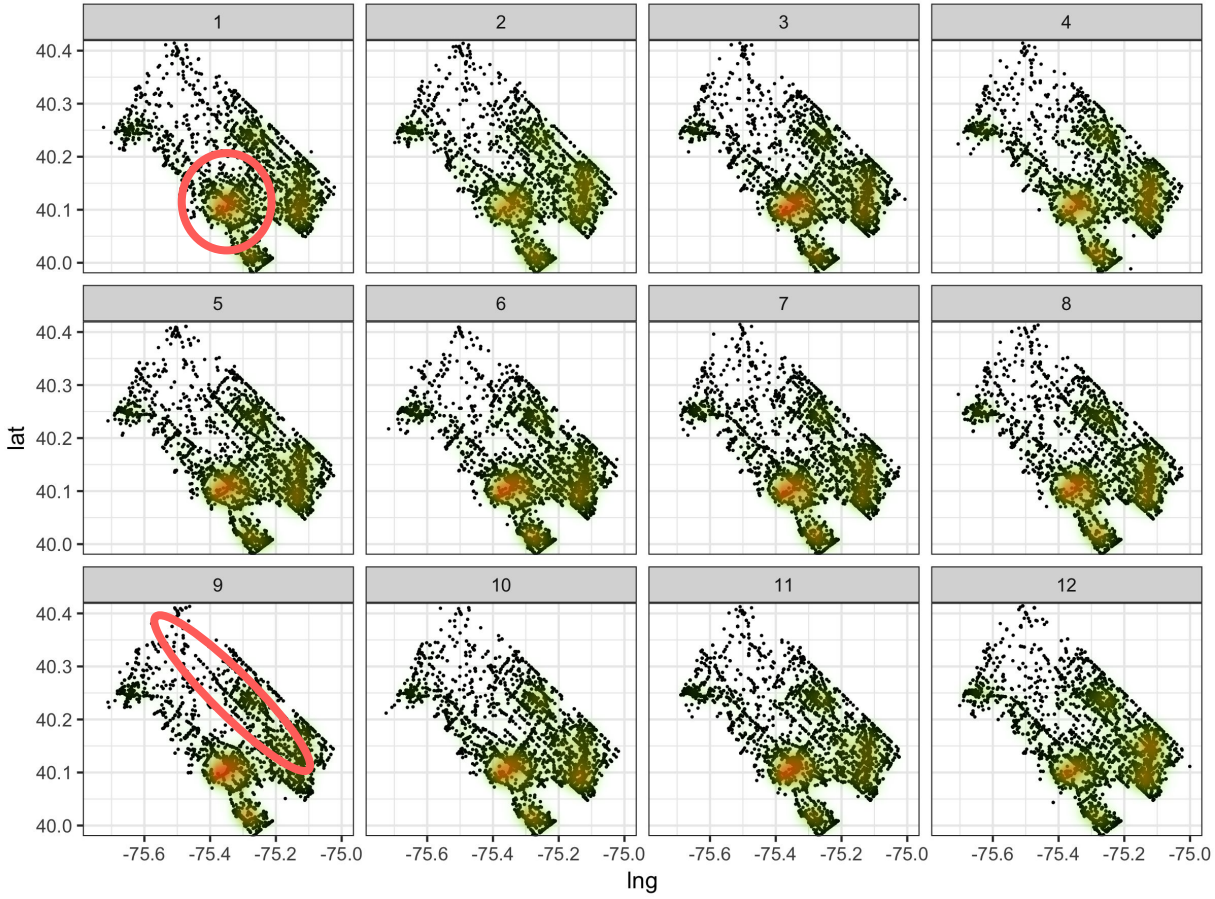
Most emergency call happened on Saturday night

EMS Density by Months

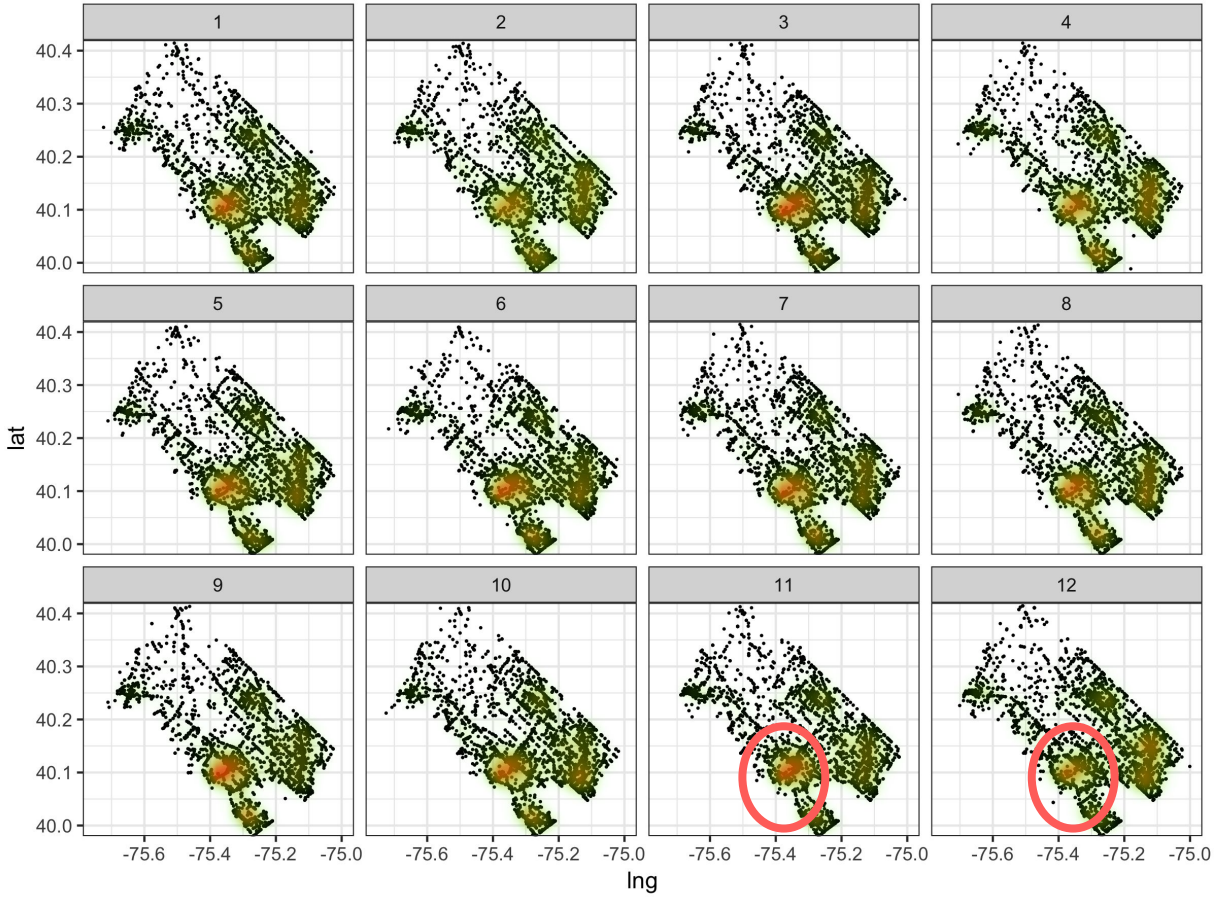




Traffic Density by Months

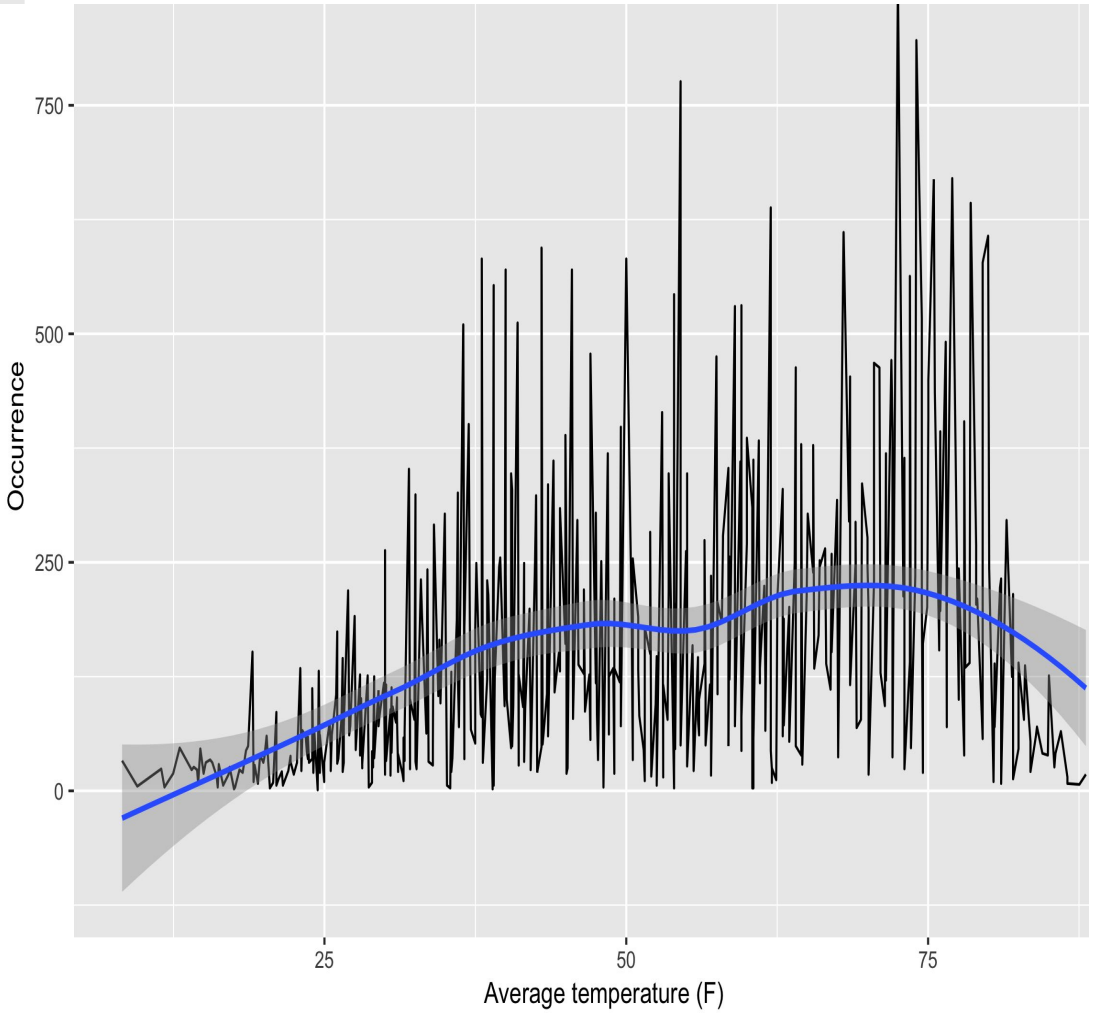


Fire Density by Months



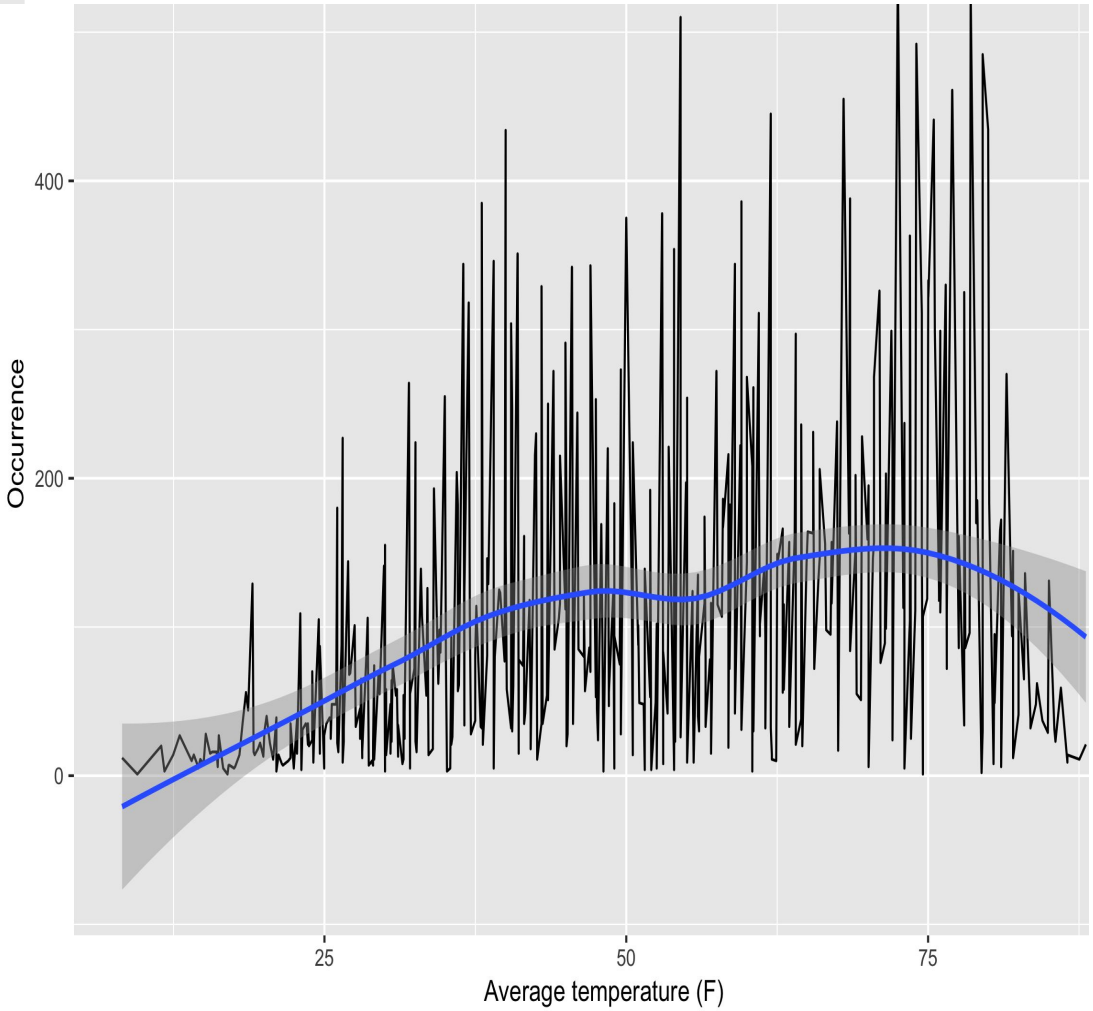
In December, the number of fire emergency at this region greatly decreased.

EMS Occurrence with Temperature



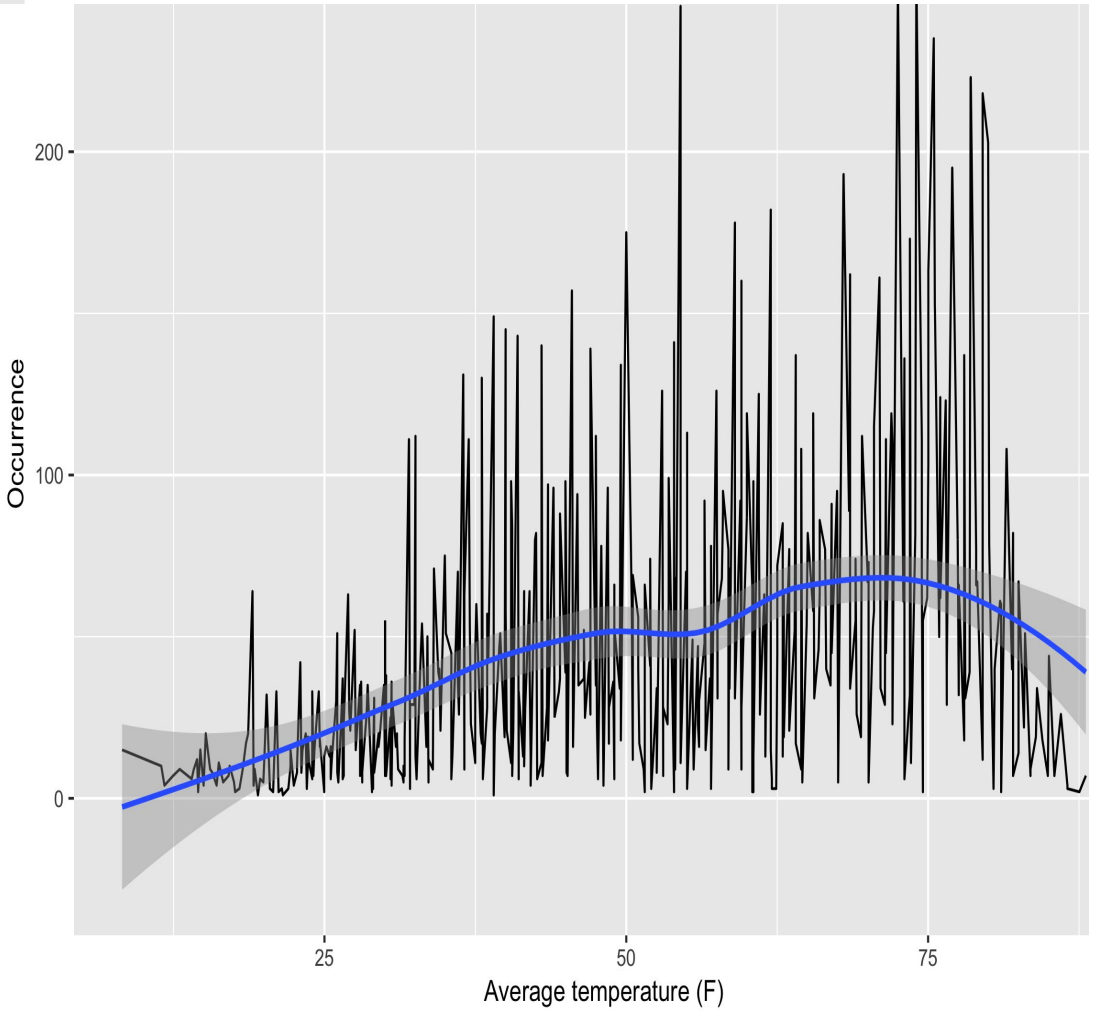
Even though there are dramatic fluctuations, the overall trend of occurrence is slightly **increasing** with temperature

Traffic Occurrence with Temperature



Even though there are dramatic fluctuations, the overall trend of occurrence is slightly **increasing** with temperature

Fire Occurrence with Temperature



Even though there are dramatic fluctuations, the overall trend of occurrence is slightly **increasing** with temperature

Predictive Modeling

X

Feature
Time/Location/Temperature

Y

Event Type
(EMS/Traffic/Fire)

Feature Generation

Weather-related

- PRCP
- TMAX
- TMIN
- Tdiff = TMAX - TMIN
- Tavg = (TMAX+TMIN)/2

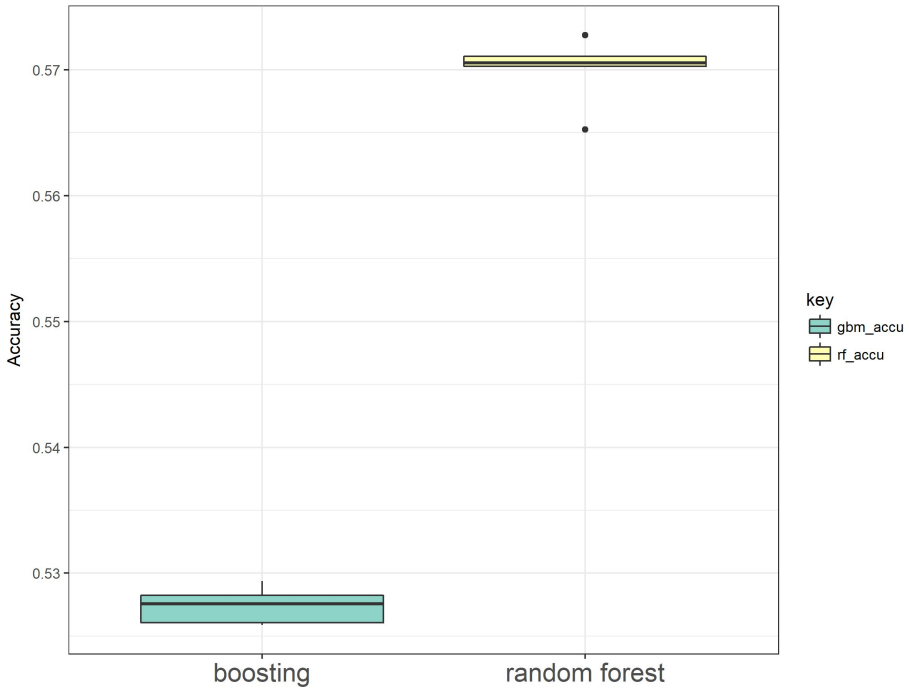
Time-related

- Month (1-12)
- Weekday (1-7)
- Weekend (bool)
- Hour (0-23)
- Season (spring, summer, fall, winter)

Location-related

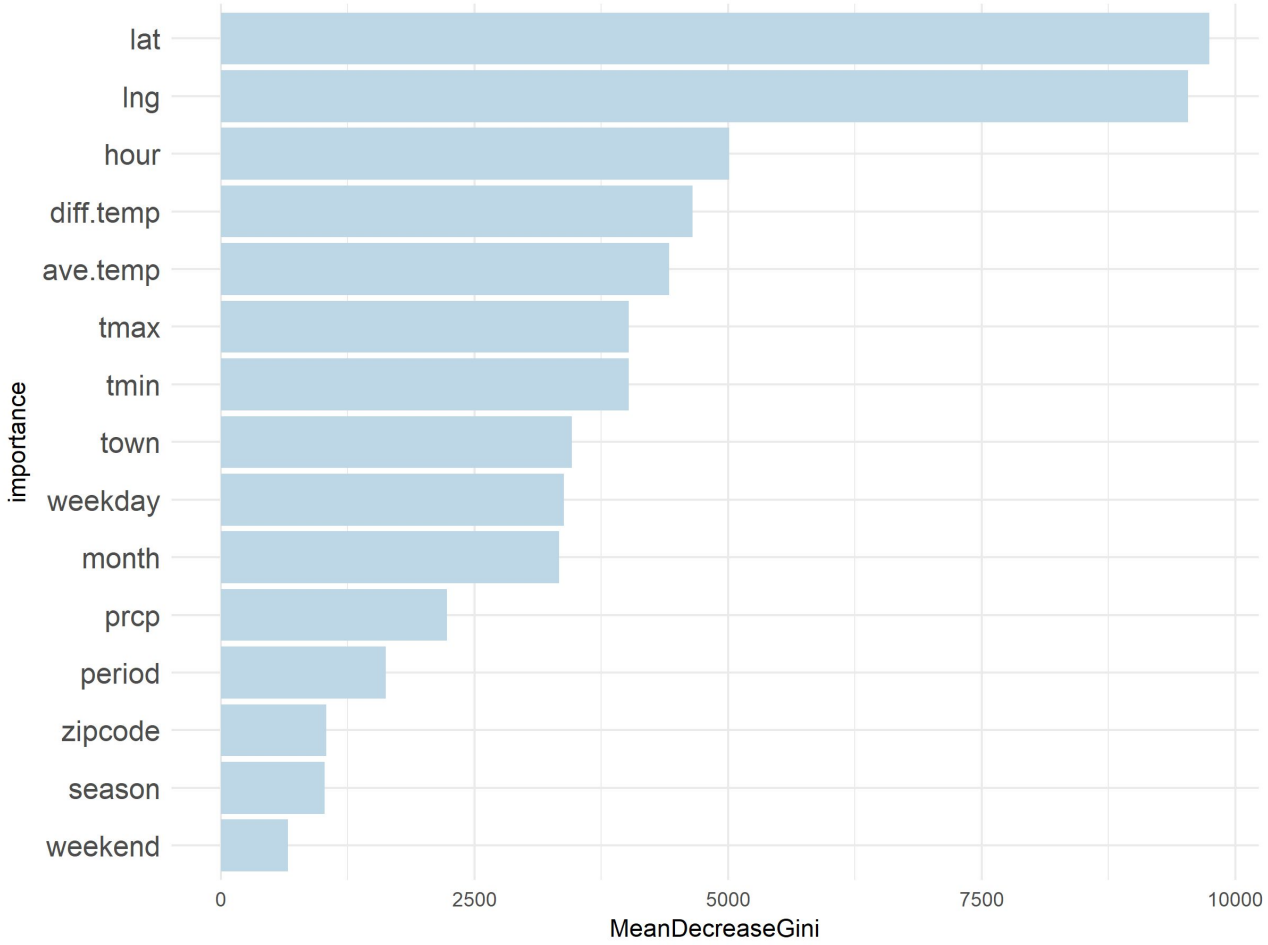
- Latitude (float)
- Longitude (float)
- Zipcode (10 class)
- Town (31 class)

Result



CV round	Random Forest	Gradient Boost Tree
1	0.5655	0.5239
2	0.5709	0.5187
3	0.5652	0.5228
4	0.5742	0.5238
5	0.5734	0.5213
Ave	0.5698	0.5221

Feature Importance - RF



Conclusion

- Finding:
  - Location is the most important feature predicting the type of emergency call
  - There are some peaks in time in the occurrence of EMS and Traffic
  - With temperature increasing, more 911 calls happen

- Future work:
  - Demographic information might improve the accuracy of models
  - Feature engineering might be a good way to optimize models