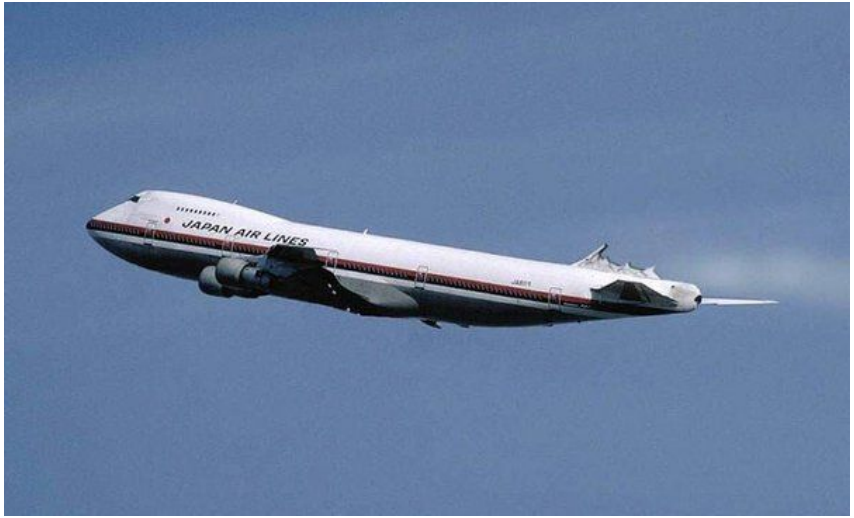
Plane Accidents and Fatalities

Huy Nguyen



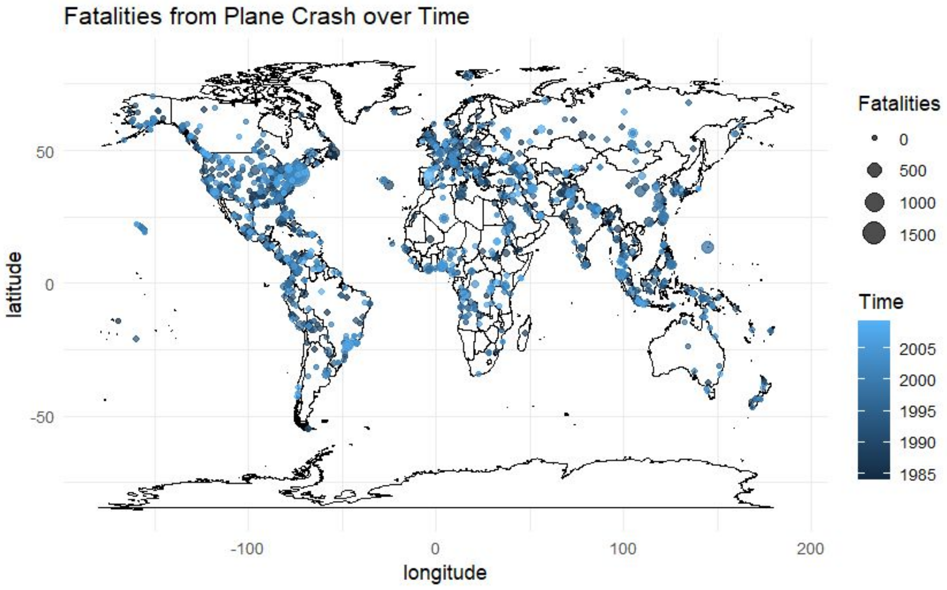
Introduction

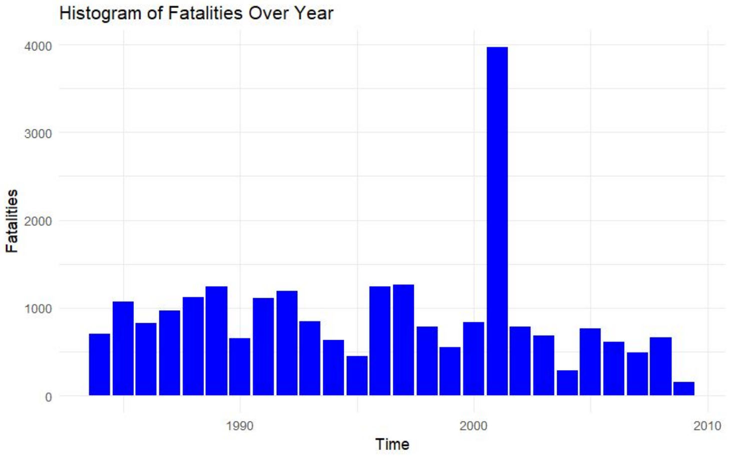
- The airline industry has remained strong despite global events
- Despite being one of the safest modes of transportation, accidents bring huge loss of life on the plane and ground
- While technology plays a role, regulations in each country can’t be ignored. Examples include maintenance, training (B737 Max), emergency protocols
- In our project, we want to: see patterns of fatalities globally, benchmark different models, and provide some safety recommendations

Data and Biases

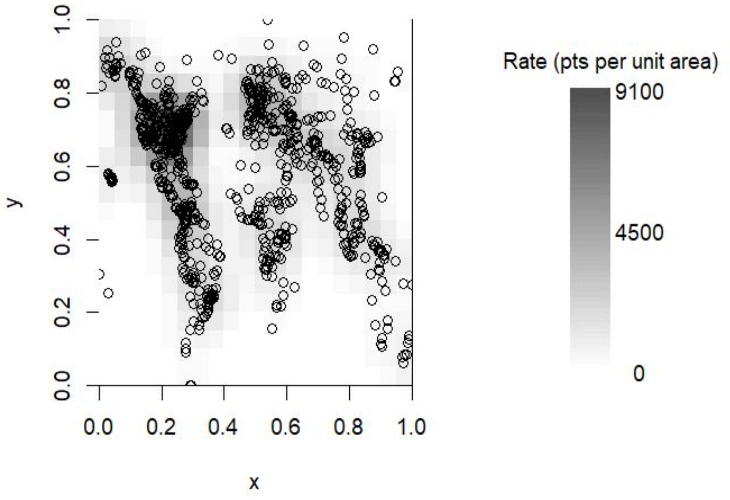
- The data started from 1908 to 2016. We want to focus on modern planes, so 1983 is our starting point (when GPS was introduced)
- The data did not have coordinates but rather location descriptions. We will use geocode() to find keywords and assign coordinates
- Relevant attributes: Longitude, Latitude, Total fatalities, Date
- Marks: plane + ground fatalities
- Some drawbacks and biases:
  - There will be no data on ocean crashes
  - Coordinates will not always be accurate (Ex: Off the coast of Los Angeles)
  - Some crashes happen at the same time or location, so the point process is not purely simple
  - Flights with missing values for the fatalities are excluded

Exploratory Data Analysis

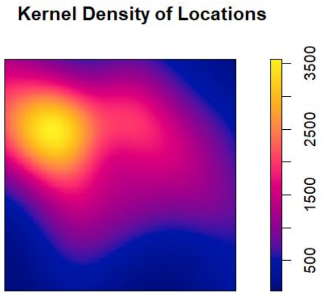




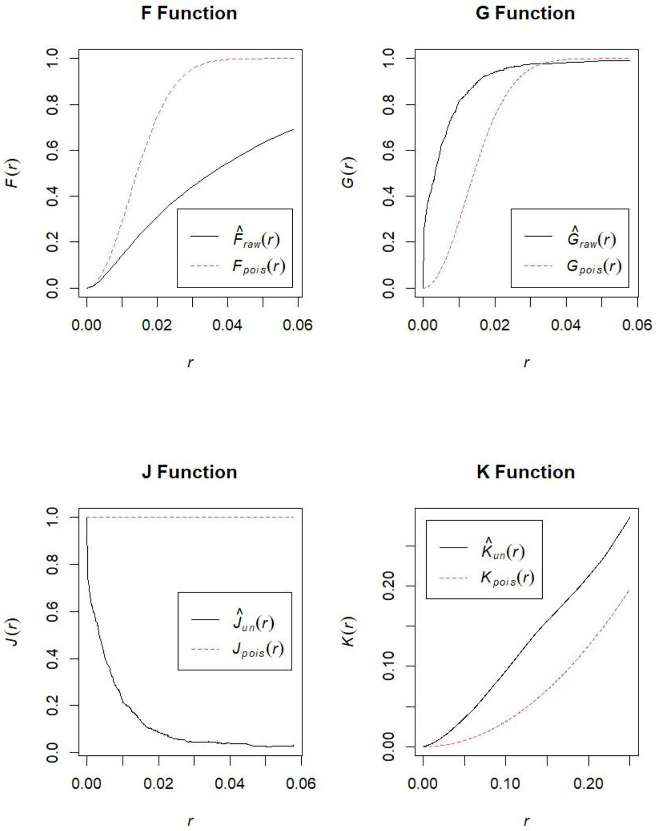
Kernel Density



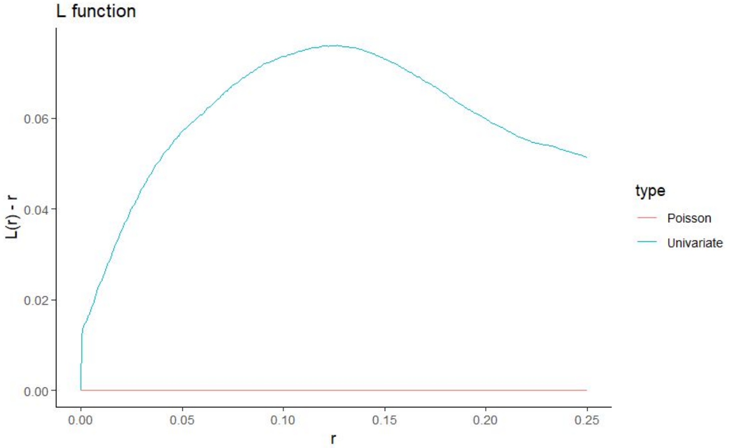
- Highest rates in contiguous US, then Europe
- Surprisingly high rates in Central America and Alaska
- One possible hypothesis is that small plane accidents dominate these remote/mountainous regions



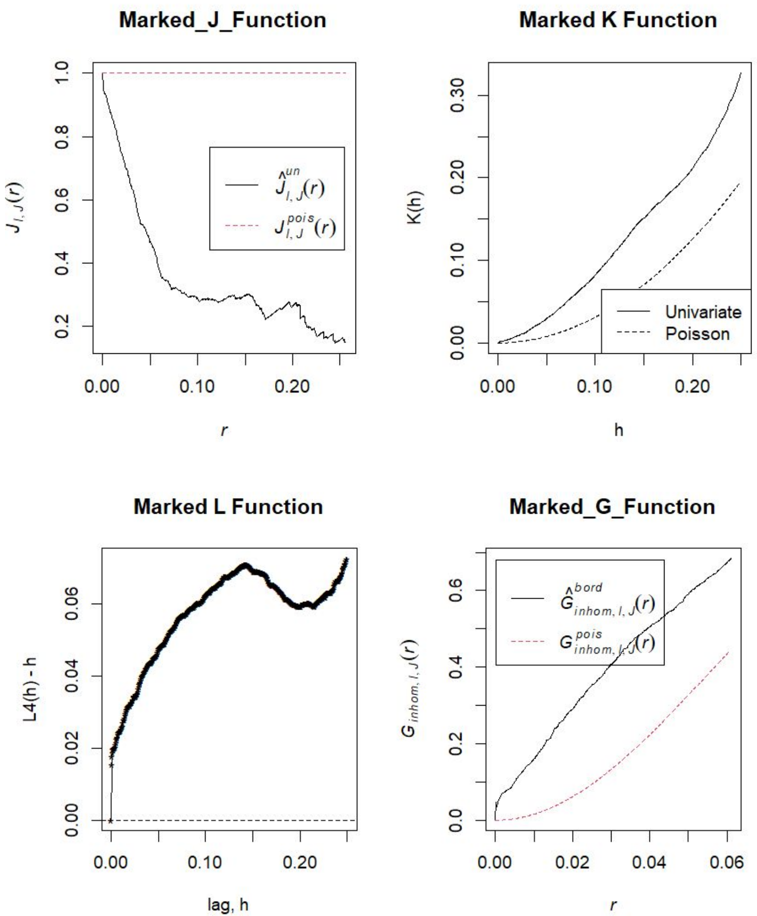
Unmarked F, G, J, K, L



- All functions indicate clustering
- Quite makes sense since most accidents happen during landing/takeoff



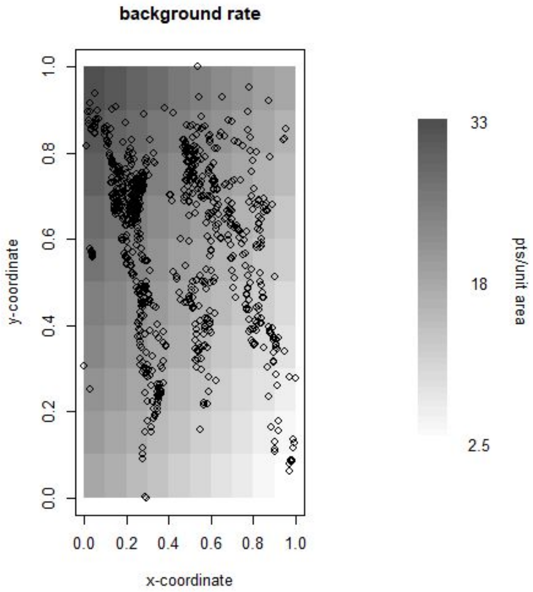
Marked J, K, L, G

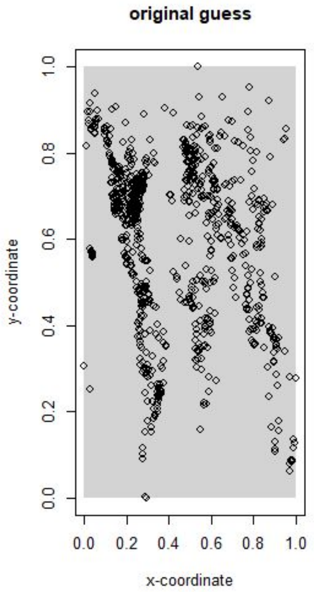


- Two Marks
  - Minor: Accidents with < 10 fatalities
  - Major: Accidents with > 100 fatalities
- The plots show clustering patterns
- Ensuring flight safety over certain regions may benefit both large and small planes.

Pseudo Log Likelihood

	μ	α	β	γ	a1
Estimate	17.7	-16.9	16.6	0.98	88.2
SE	14.9	17.7	15.8	0.03	3.92





Hawkes (Basic)

	μ	κ	α	β
Estimate	0.13	0.016	2.98	1430
SE	0.028	4.9*10^(-4)	0.095	7.14

- Better performance than Pseudo Log Likelihood

Poisson-Stoyan

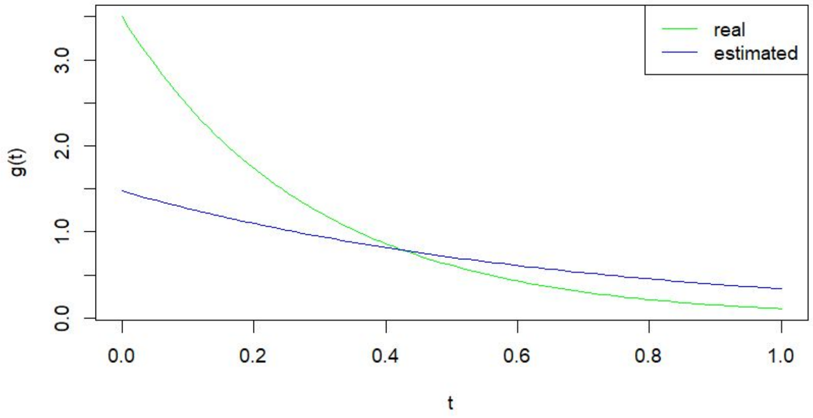
	μ	κ	α	β
Estimate	0.2	0.39	131	0.016
SE	0.005	0.074	68.96	0.003

- Much worse α estimation than Hawkes

Hawkes-Stoyan

	μ	κ	α	β
Estimate	0.809	0.973	1.327	1.473
SE	0.26	0.37	0.55	0.58

- Best model we have so far!
- Still many rooms for improvements however



Conclusions and Future Directions

- The project shows clustering trends for fatalities
- Thus, regions such as US or EU should enhance regulations and training especially for smaller/private planes
- Probably focus on main airports within areas of high accident rates so evacuation process is more efficient
- Add ocean and more accurate coordinates from Wikipedia
- Include years before GPS and Autopilot for comparison
- Focus only on the US or Europe for more insights
- Break out by classes ie First Class, Business, Economy

Thank you for your time!

Questions?

	SVM	Hierarchical
Accuracy	92.6%	71.7%