

Pattern Recognition in Road Accidents

ADWAIT SAHASRABHOJANEE

XUE YU

SREEJITH SREEKUMAR

XUEXIAN LI

Introduction

- Importance to evaluate road accident pattern
- Data complexity
 - Both categorical and numerical data
 - Temporal and spatial characteristics



Image courtesy: <https://www.colourbox.com/vector/car-crash-accident-with-two-damaged-auto-vector-29357577>

Approach

Data collection

- U.K. Department of Transportation
- 1 million accidents and 6 years of traffic flow

Unsupervised machine learning

- Exploratory Data Analysis
- Association analysis
- PCA
- Clustering

Association analysis

X = Condition of the accident

Y = 'Police_officer_attend = Yes'

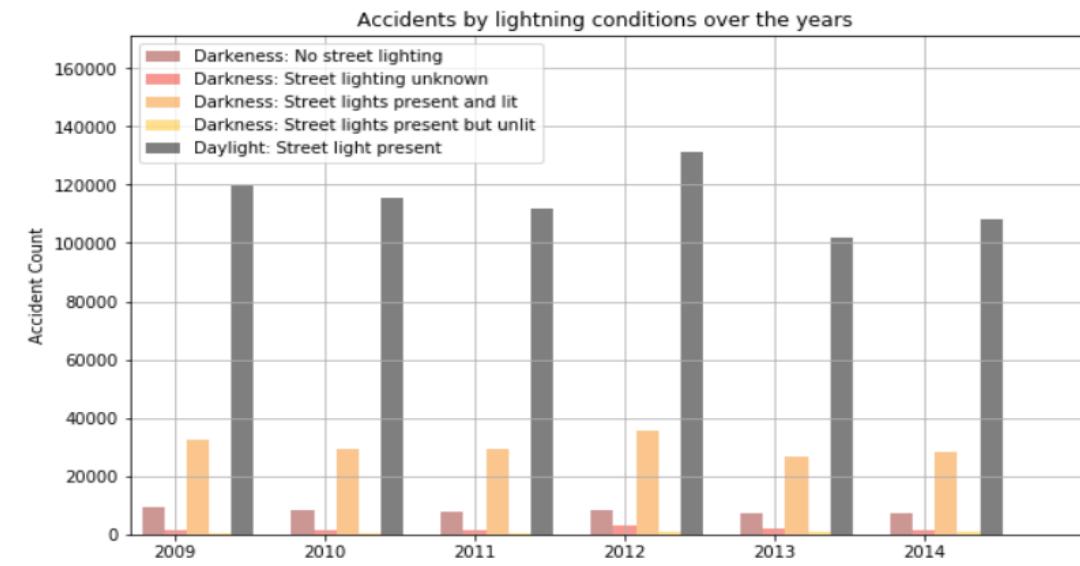
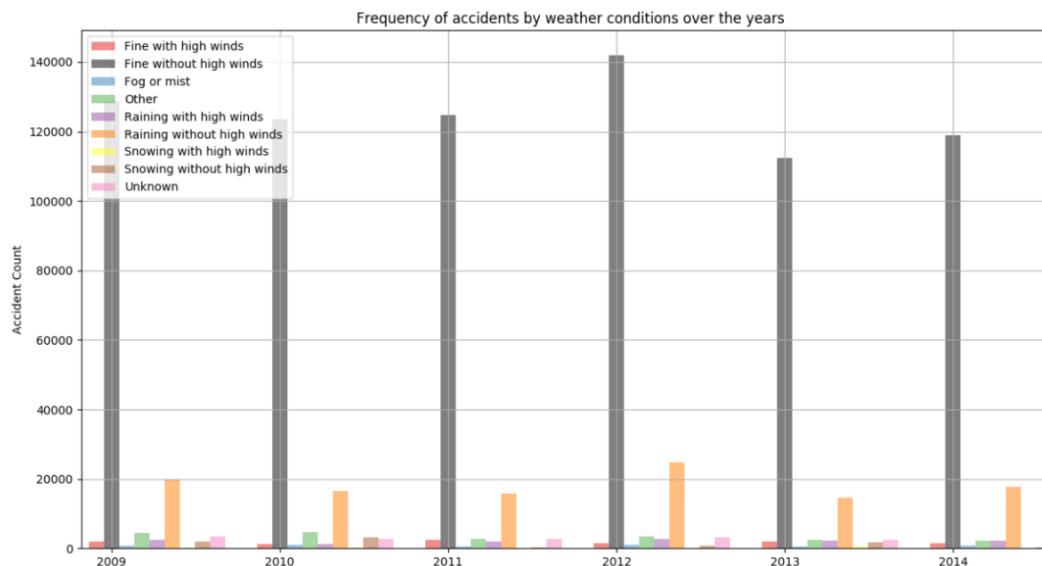
$$\text{Support}(X \rightarrow Y) = \frac{P(X \cup Y)}{P(I)} = \frac{\text{num}(X \cup Y)}{\text{num}(I)}$$

Interesting Findings

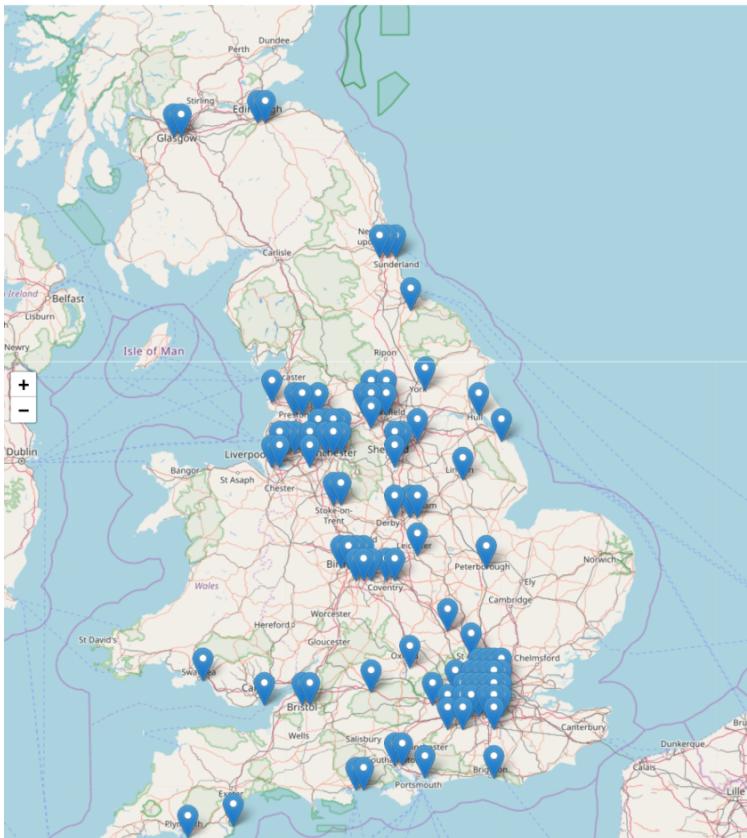
Number of Vehicles = 1, Police Officer Attend = Yes

Number of Vehicles ≥ 3 , Police Officer Attend = Yes

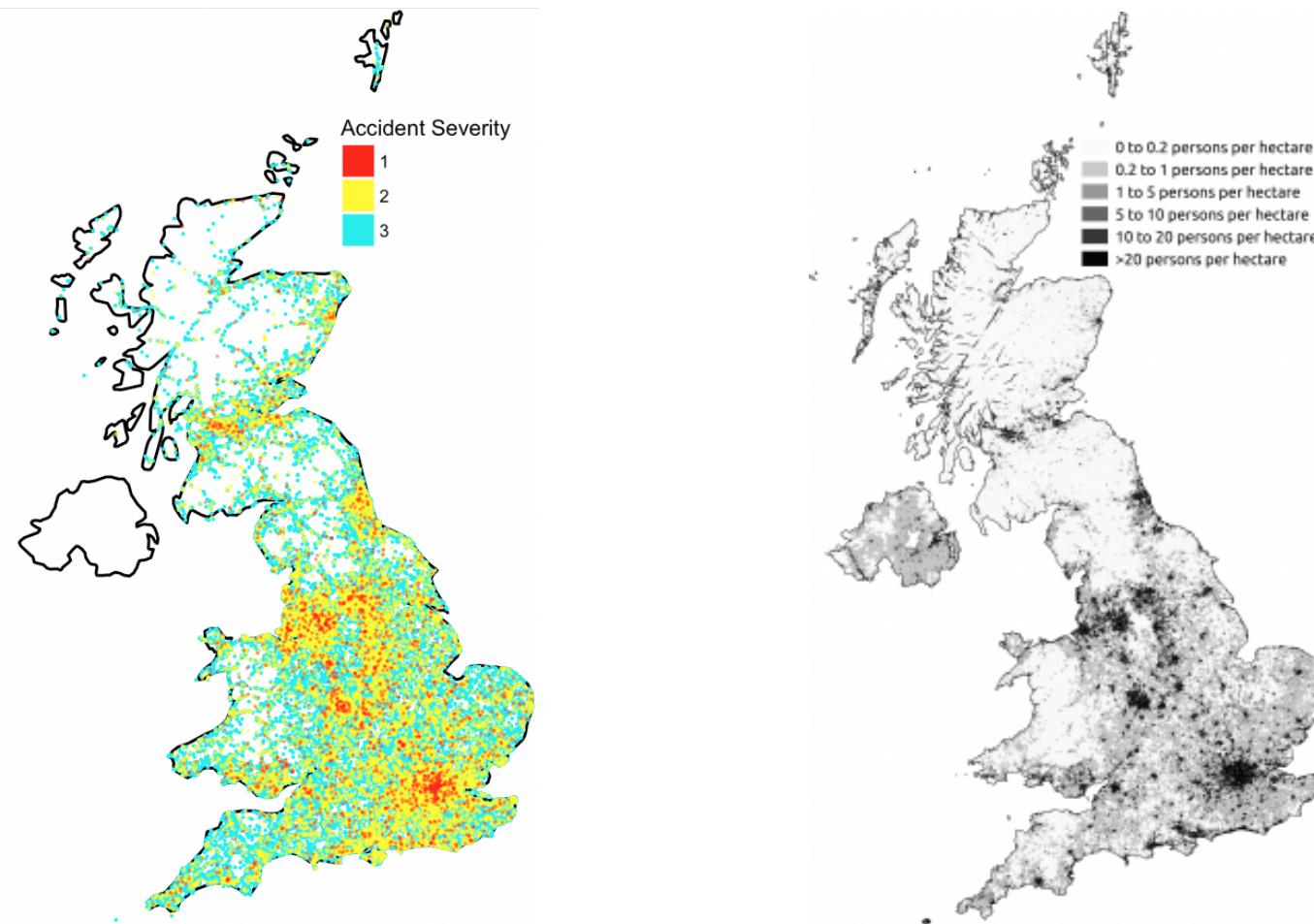
Exploratory Data Analysis



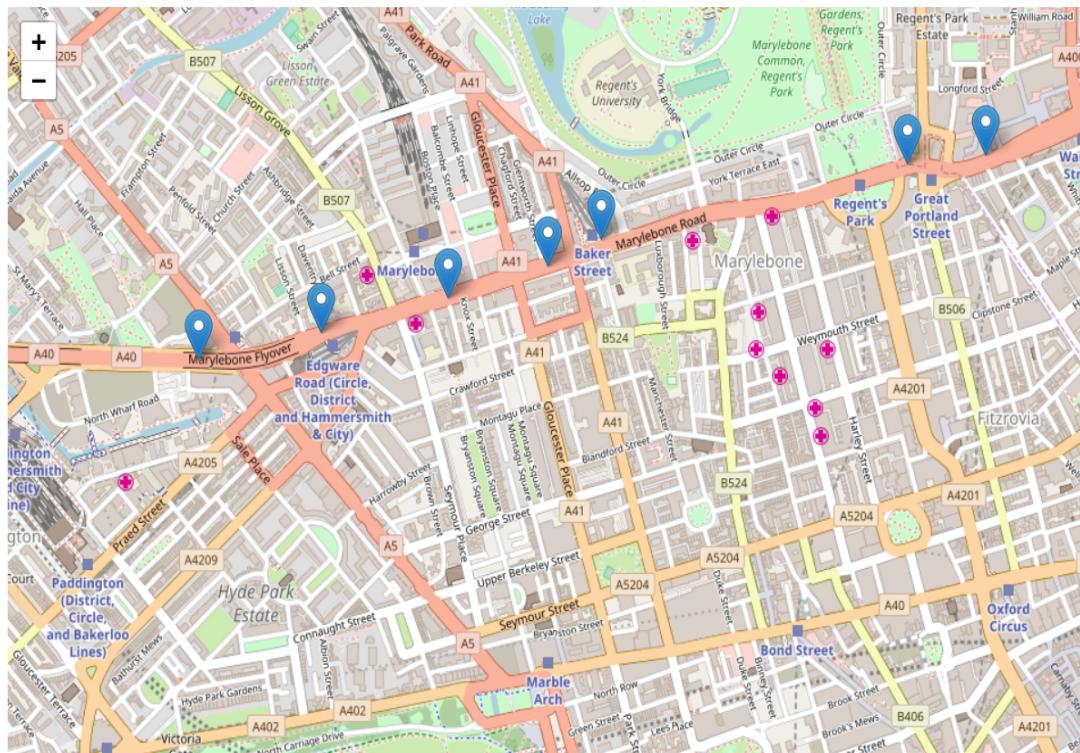
Spatial Distribution of Accidents



Scope of the Project

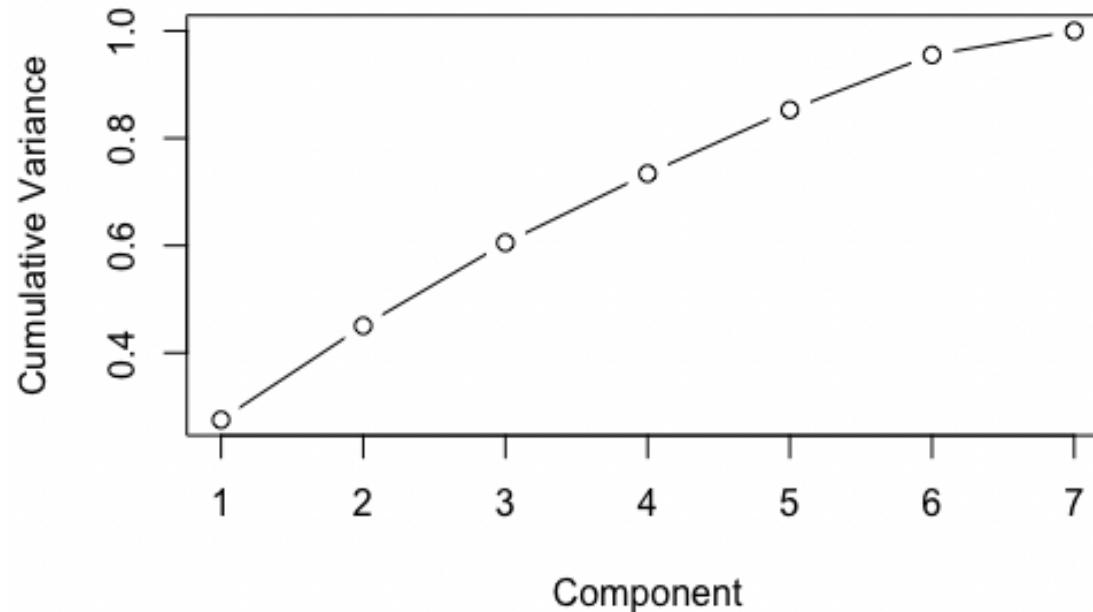


Studying Traffic Density in London



Segregation of accidents in London using numeric attributes

Principal Component Analysis



DBSCAN on first six components



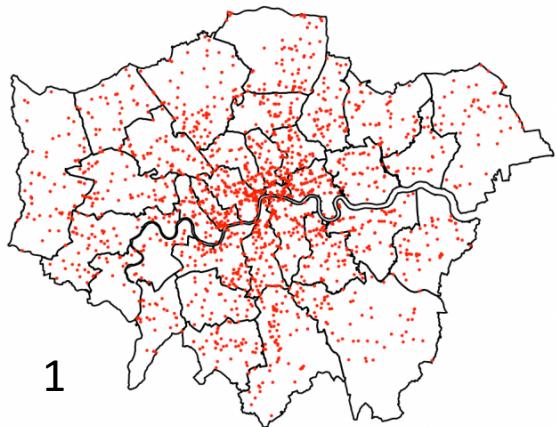
Clustering on Nominal Variables

1. Variables are one-hot encoded
2. Manhattan distances are calculated
3. The K-medoids algorithm is used to cluster the data

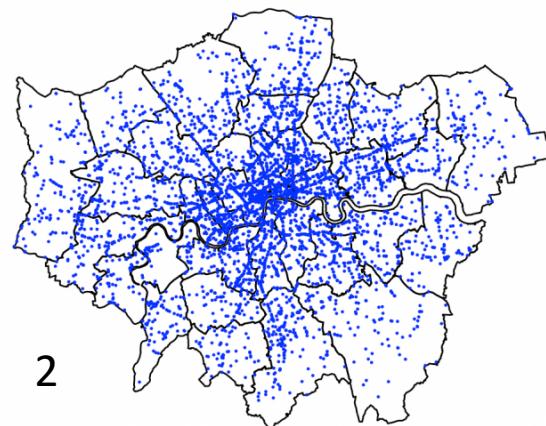
Why doesn't K-means work?

The centroids in K-means can be points that do not exist in the data.

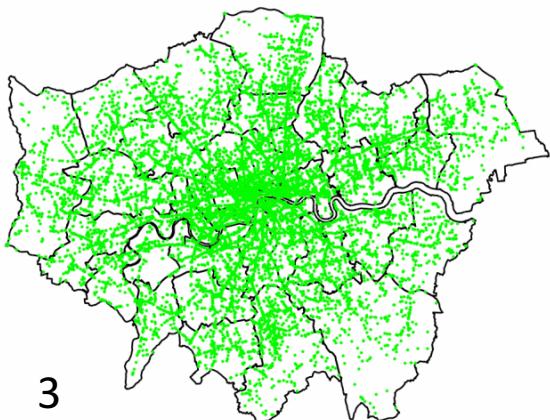
Results of clustering



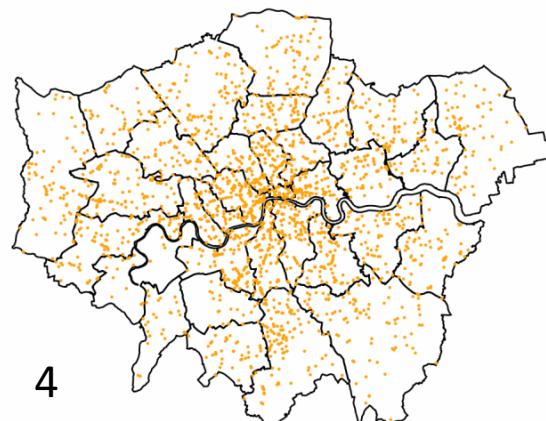
1



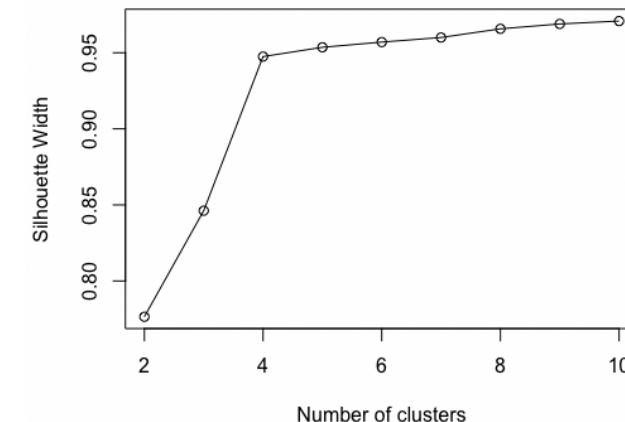
2



3



4

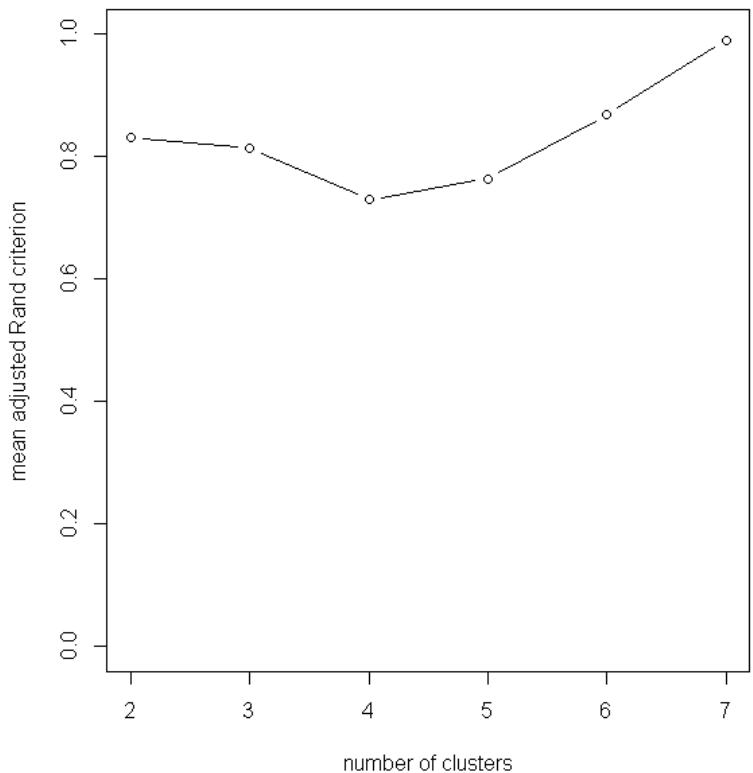


Cluster	Road_Surface_Conditions	Light_Conditions	Carriageway_Hazards	Special_Conditions_at_Site
1	Wet/Damp	Darkness: Street lights present and lit	None	None
2	Dry	Darkness: Street lights present and lit	None	None
3	Dry	Daylight: Street light present	None	None
4	Wet/Damp	Daylight: Street light present	None	None

Hierarchical clustering

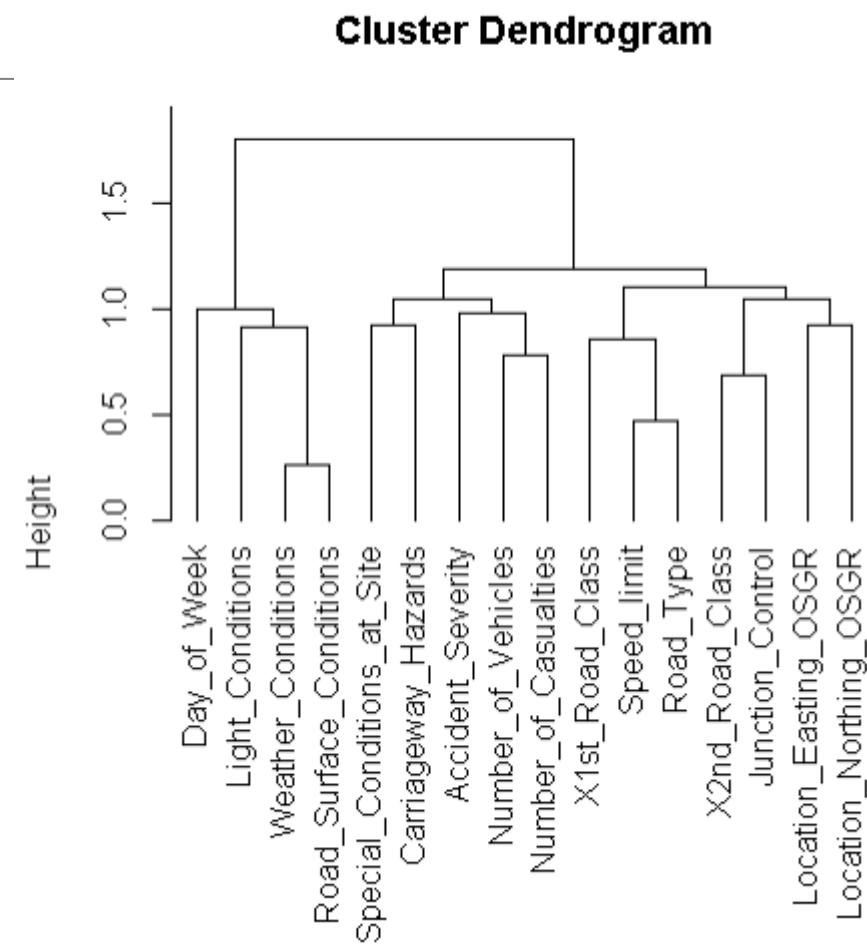
Algorithm

- Evaluate the variable contributions
- Separate quantitative and qualitative variables
- hierarchical/k-means clustering
 - PCAMIX
- Bootstrap to evaluate cluster numbers



Results for hierarchical clustering

- Three clusters
- generally grouped as:
 - Weather condition
 - Accident condition
 - Road condition



Conclusions

Pattern recognition results

- PCA
- Clustering analysis with k-means, DBSCAN, and k-medoids algorithms
- Hierarchical analysis of variables weight

Suggested future work

- Incorporate socio-demographic data into analysis
- Build interactive visualizations of pattern recognition