



On the Correlation of Traffic Severity and Weather in California



Ching Wang*

Kuan-Hsun Wu**

Department of Statistics, National Chengchi University

Introduction

Motivation: Provide guidance to local governments for enhancing public facilities and overall safety.

- Building on previous findings [1] linking weather and traffic accidents in snowy mountain areas.
- Identify the impact of weather on the severity of traffic accidents.
- Determine suitable model, considering geographic features.

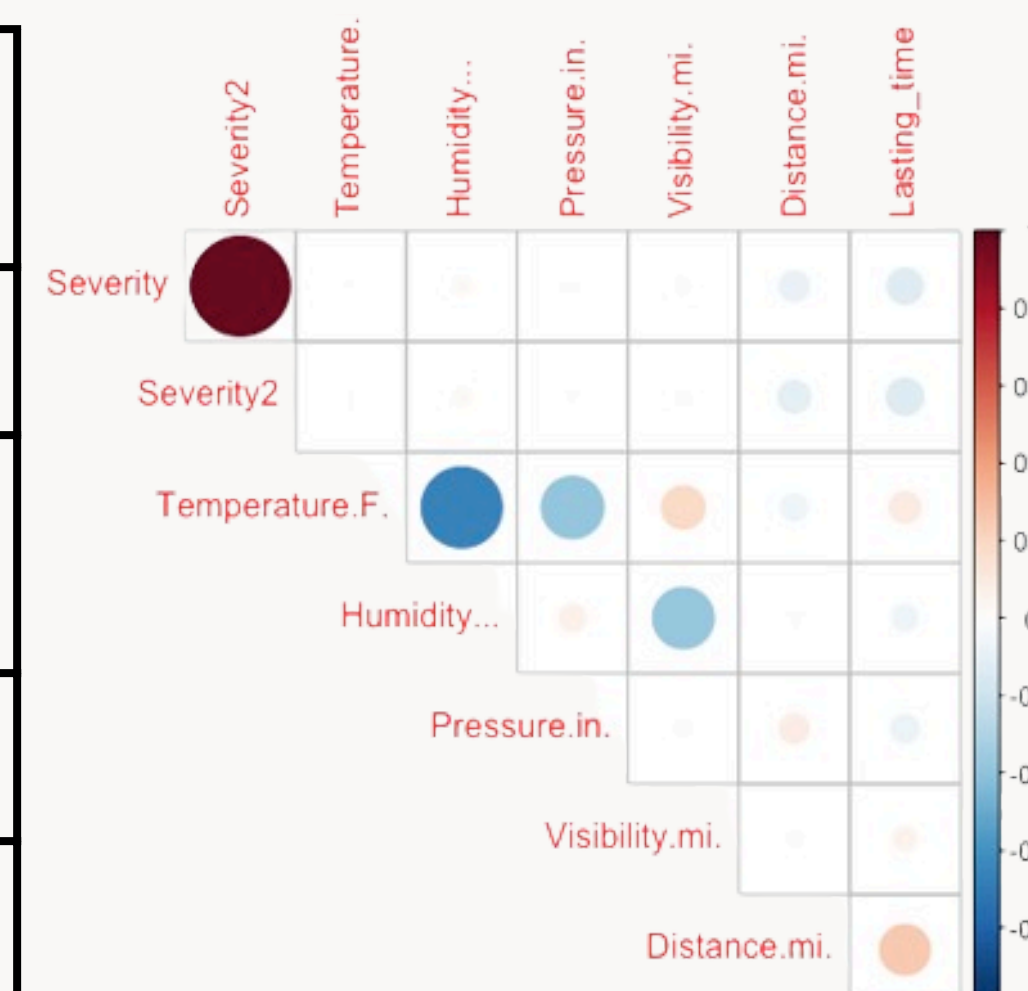
Data Introduction

- **Data:** The dataset, with 7.7 million entries, is derived from the "A Countrywide Traffic Accident Dataset" on Kaggle.
- The dataset comprises 45 attributes, categorised into 5 distinct groups.

Important Attribute Table

Traffic Attributes (10)	severity, start_time, start_point, distance
Address Attributes (8)	city, county, state
Weather Attributes (10)	temperature, humidity, pressure, visibility, wind_speed, precipitation
POI Attributes (13)	normal traffic facilities
Period-of-Day (4)	Civil Twilight

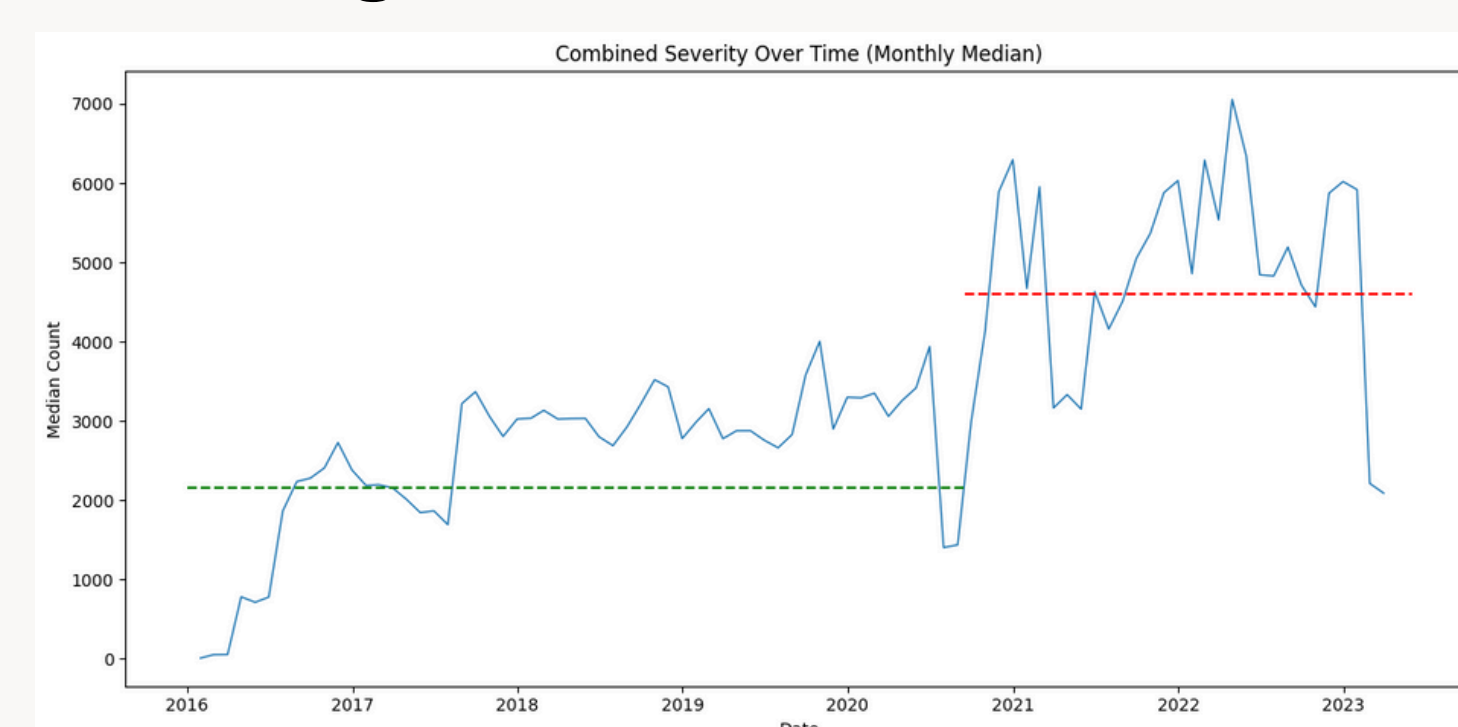
Variable Correlation Plot



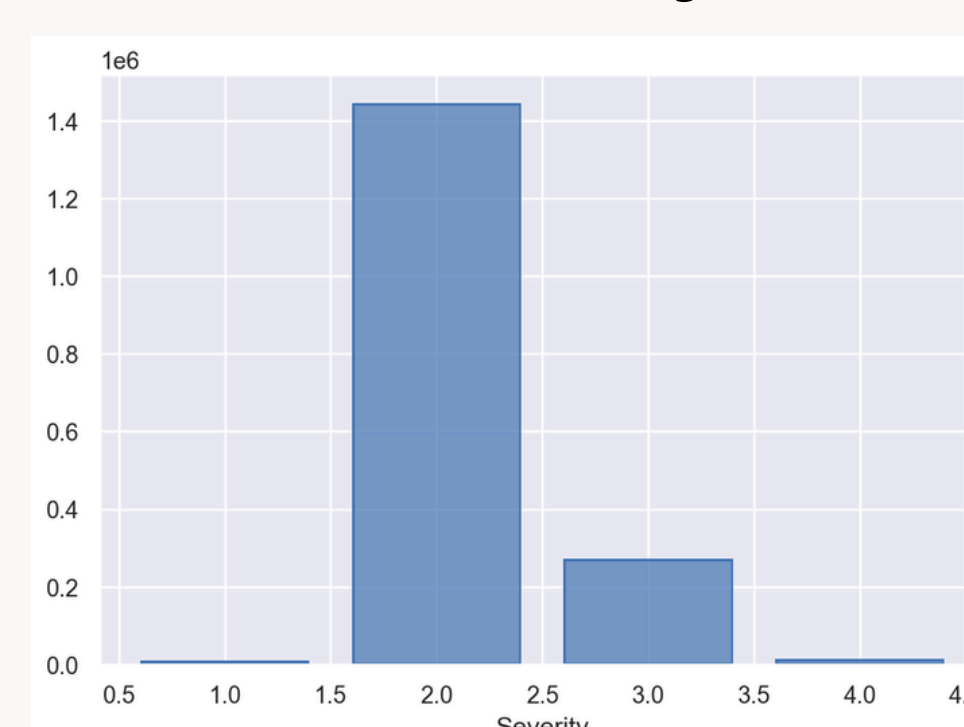
Data preprocessing

- **Geographic:** Obtain 23% of entries from California, and zoom in Tehama.
- **Time:** A change point in accident counts was found on September 15, 2020, using the Pettitt test. Data before this point was chosen.
- **Severity:** To mitigate the effect of imbalance in Severity, levels 1 and 2 of the severity attribute were merged, as were levels 3 and 4.

Change Point of Accident counts



Unbalanced Severity Attribute



Model

Ordinal Logistic Regression:

$$\text{logit}(P(Y_i \leq j)) = \beta_0 + \sum_{i=1}^p \beta_i X_i$$

- Six variables were chosen for the model based on the correlation plot.
- Adjusted R-square(McFadden): 0.0242

Confusion Matrix

predictions\actual	2	3	4
2	52	28	0
3	0	0	2
4	1	0	0

Deviance Table

	Pr(>Chisq)
Distance	0.0009
Temperature	0.0862
Humidity	0.1853
Pressure	0.7171
Visibility	0.0563
Lasting_time	0.2514

Geographically Weighted Regression (GWR):

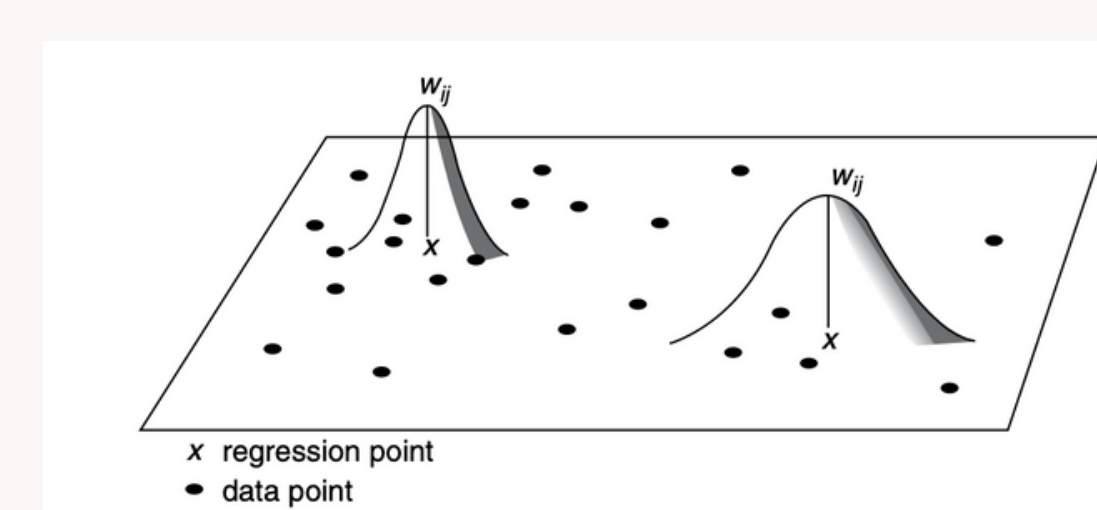
- Geographical Factor was not considered in above regression models.

$$Y_i = \beta_0 + \sum_{i=1}^p \beta_i X_i \quad (\text{global regression})$$

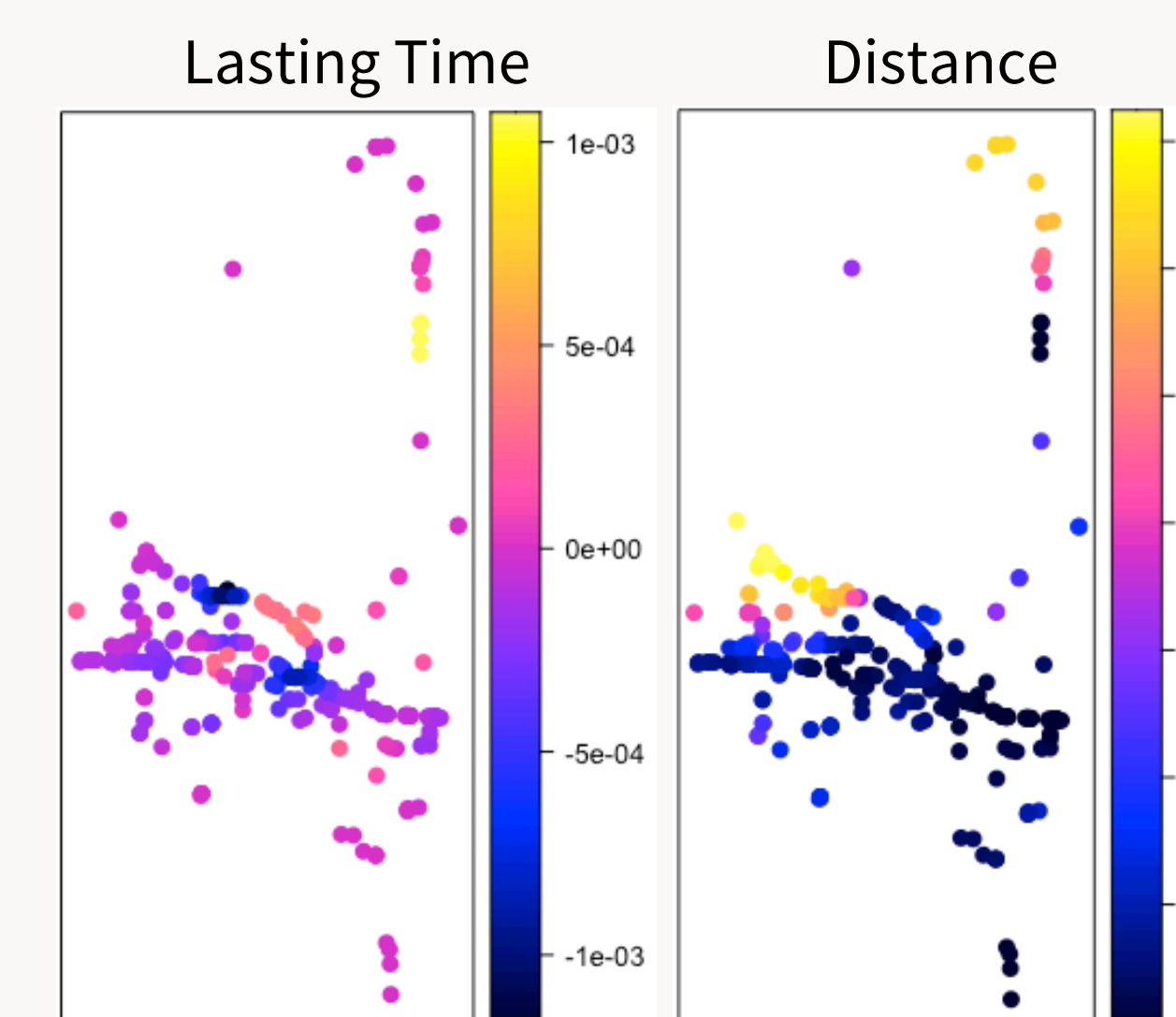
$$Y_i = \beta_{0,(u_i,v_i)} + \sum_{j=1}^p \beta_{(u_i,v_i)} X_{ij} \quad (\text{Local regression, spatial distance considered})$$

- GWR considers local variation and parameter estimate differs from each other.

Geographic Weight Illustration



Parameter Estimate



GWR Evaluation

	Lasting_time	Distance	Lasting_time+Distance
pseudo_r_square	0.0001	0.0121	0.01279

Conclusion

Ordinal Logistic Regression Issues

- Omission of Crucial Variables:
 - Human factors in car accidents may be overlooked.
- Imbalanced Response Variable:
 - Severity variable imbalance, with a majority labeled as severity 2.

Geographically Weighted Regression Challenges

- Limited Geographic Scope
 - Analysis constrained to Tehama due to computational limits.
- Spatial Variation in Parameter Estimation
 - Model captures spatial variation but doesn't help explanation.

Future prospective

Geographical Models

- The correlation of each area does exhibit variation.
- Perhaps Tehama is too small to show spatial variation, we can consider larger data scope in the future.

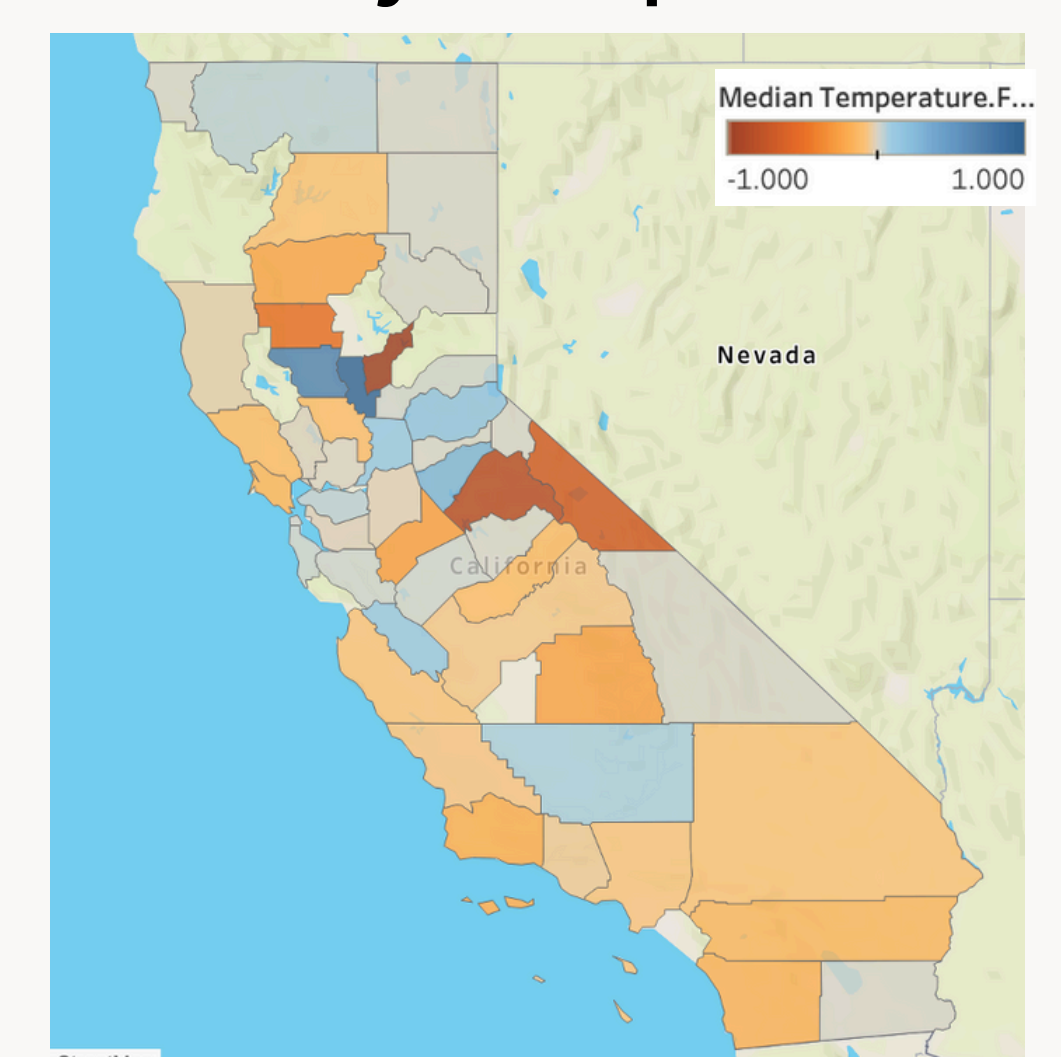
Computational Resource

- Search of optimal bandwidth requires large resource.
- R does not completely support GWR. Software suite such as ArcGIS is needed.

Human-Related Factors

- Include more human related factors might help increasing the explanation in response variable.

CA Partial Correlation Plot Severity - Temperature



Cited References

- Pang, J., Krathaus, A., Benedyk, I., Ahmed, S. S., & Anastasopoulos, P. C. (2022). A temporal instability analysis of environmental factors affecting accident occurrences during snow events: The random parameters hazard-based duration model with means and variances heterogeneity. Analytic methods in accident research, 34, 100215.
- Moosavi, S., Samavatian, M. H., Parthasarathy, S., & Ramnath, R. (2019). A countrywide traffic accident dataset. arXiv preprint arXiv:1906.05409.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2003). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.