DS5230 Project Update 1 : Pattern Recognition in Accidents in the UK

Sahasrabhojanee, Adwait
sahasrabhojanee.a@husky.neu.edu

Sreekumar, Sreejith
sreekumar.s@husky.neu.edu

Yu, Xue
yu.xue1@husky.neu.edu

Li, Xuexian
li.xuex@husky.neu.edu

November 1, 2017

1 Introduction

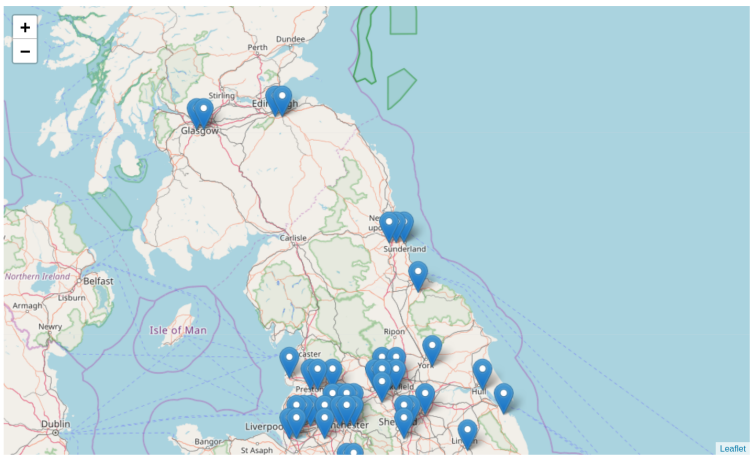
The Department of Transport in the United Kingdom has extensively recorded all the road accidents and hence provides an exclusively comprehensive dataset to explore the patterns and influencing factors regarding road casualties (Austin, 1994). According to the data, the number of road casualties is alarmingly high.

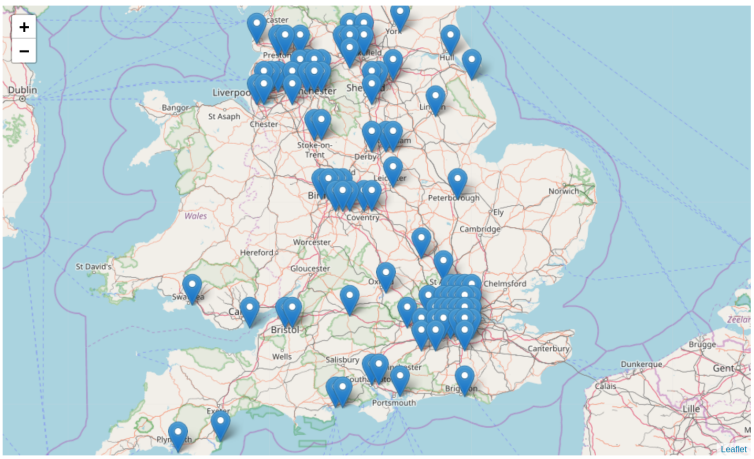
As an initial study of the dataset, we explore noticable patterns using statistical methods and exploratory data analysis.

2 Exploring accidents the UK region for trends and patterns

2.1 Spatial Study

The U.K. includes the island of Great Britain, the north-eastern part of the island of Ireland and many smaller islands. Locating the accident spots using geo co-ordinates in the records and isolate and pin point the top 200 accident spots, we have:





Over all the years from 2009-2014, the cities which have remained top 13 in casualties are London, Birmingham, Kirklees, Liverpool, Leeds, City of Edinburgh, Wakefield, Rushcliffe, Sunderland, North East Derbyshire, Elmbridge, East Lindsey, and Manchester. There is a high concentration of pins around these areas. As one might expect, these are the most populated cities in the UK[1].

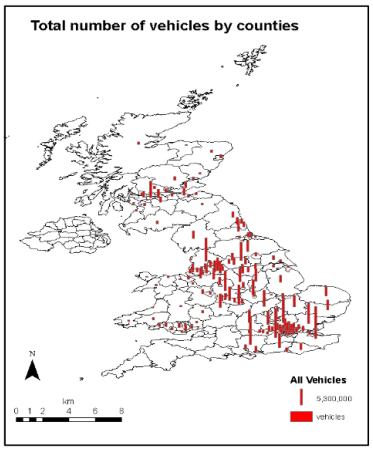


Figure 1: Total number of vehicles in various counties of U.K.(2014)

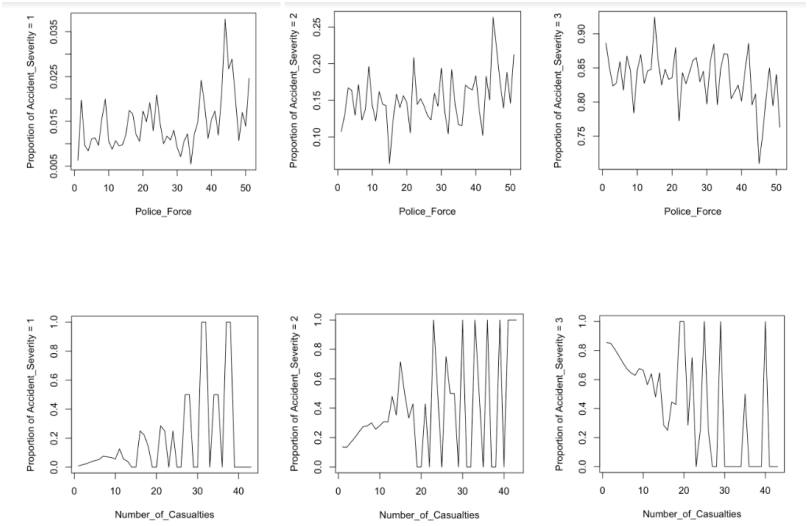
The heat map of the number of vehicles in the UK for different counties, more or less correlates with these locations as well.

2.2 Summary Statistics

Looking at the summary statistics of the records, we have:

Feature	Mean	Standard Deviation	Median	Mode
Police Force	29.59	25.49	23	1
Number of Vehicles	1.826	0.709	2	2
Number of Casualties	1.34	.82	1	1
Speed Limit	38.52	13.9	30	30

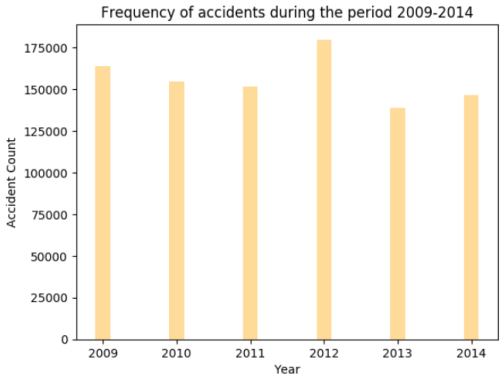
The dataset groups the severity of the accidents in levels 1-3. However, it's not mentioned which is more severe on this scale. Since this is critical to our analysis, we tried mapping the proportion of accidents in each severity level with the number of officers police force who attended the scene and the number of casualties.



As the number of casualties and the police force increases, the probability of the accident being assigned a severity of 1 increases, and the probability of being assigned a severity of 3 decreases. The probability of being assigned a severity of 2 increases sharply at first, but then slows down. This shows that a severity of 1 is the most severe, and a severity of 3 is the least severe.

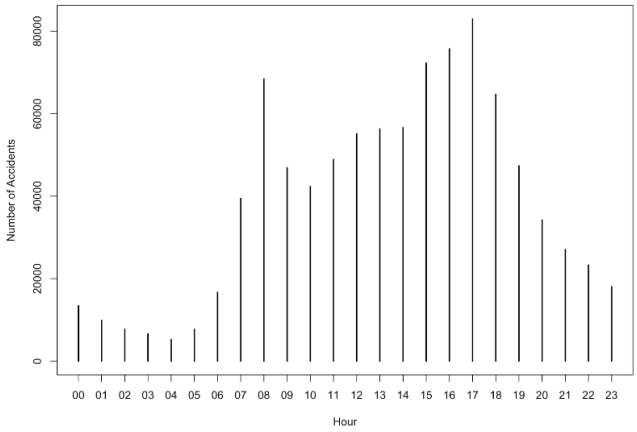
2.3 Trends

How has the number of accidents varied over the years?

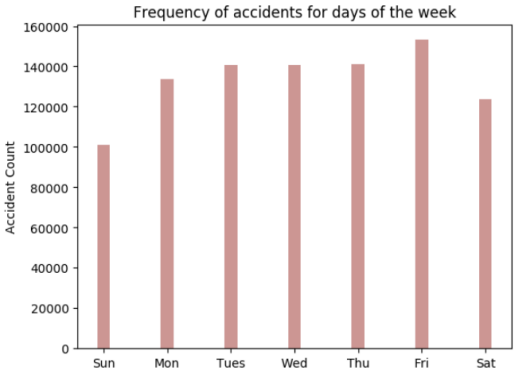


As seen in the bar plot, the accidents seem to have very slight variations, gradually decreasing trend over the years with an exception in 2012 when the count of casualties shot up over 175000. However, we don't know why this abrupt increase has happened yet.

The variation in the number of accidents over different hours of the day indicates that the most of the accident occur during the evening, with it's peak between 5.00pm - 6.00pm. This observation seems natural since more vehicles are expected after daily work hours, which usually ends around this time.



Moreover, according to the plot of the frequency of accidents over days in the week, friday leads in the number of casualties.

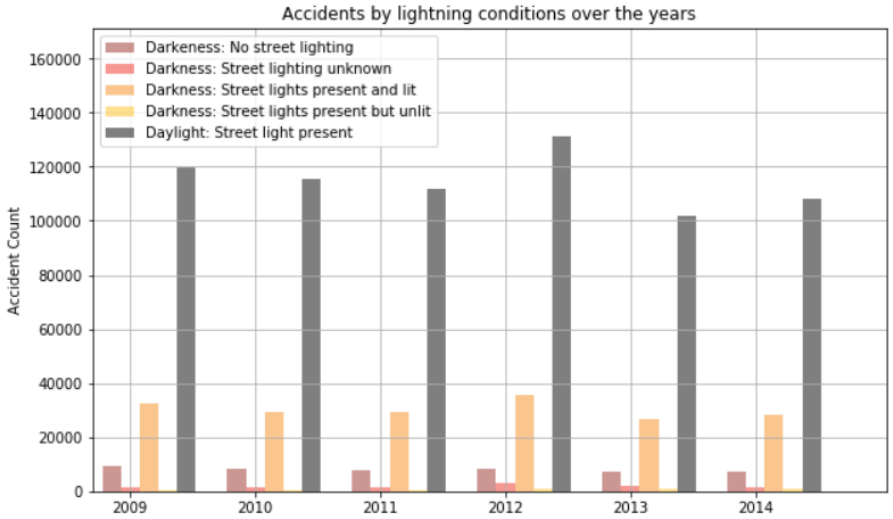


How much do variation in seasons affect road accidents?

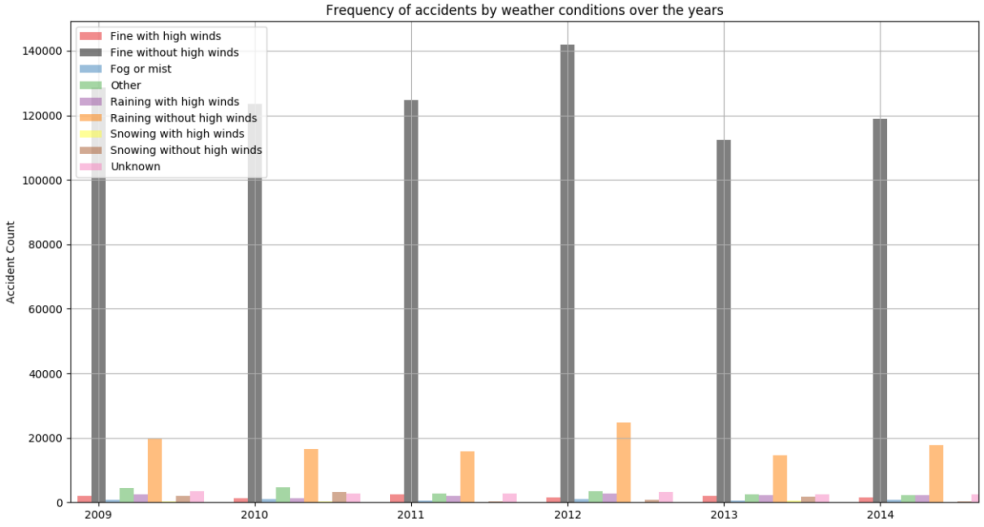
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Fine with high winds	11.09	9.33	9.51	5.07	8.20	4.13	2.88	3.17	8.95	9.56	16.97	11.13
Fine without high winds	7.09	6.95	8.92	8.38	9.20	8.98	9.21	8.63	9.23	8.92	8.04	6.46
Fog or mist	15.50	10.12	12.96	2.84	1.38	2.33	1.11	1.24	3.44	9.93	18.74	20.42
Other	19.34	12.36	4.65	3.21	3.16	3.05	3.99	3.84	3.78	7.32	12.64	22.68
Raining with high winds	10.00	4.75	4.46	4.83	3.95	5.05	3.53	3.67	8.44	9.81	24.58	16.95
Raining without high winds	7.16	6.33	4.83	6.08	6.05	7.95	9.46	8.23	7.79	11.94	13.47	10.72
Snowing with high winds	19.41	26.06	18.51	3.23	0.54	0.00	0.18	0.18	1.08	0.99	4.22	25.61
Snowing without high winds	26.84	28.70	4.77	0.90	0.23	0.25	0.12	0.08	0.20	0.56	7.34	30.01
Unknown	10.09	8.61	7.73	7.04	7.96	7.27	7.25	6.77	7.74	9.20	10.70	9.64

The analysis shows that 30.1%, 26.84% and 28.70% of all accidents in snowy weather occurs in December, January and February respectively. The numbers are quite high for environments with fog or mist, as well as snow with high winds during these months. May be taking more precautions and saftey measures for the weather can bring down accidents during these months.

How are the accidents dependent on lighting conditions. Do they happen mostly on roads with insufficient lighting at night ?



The yearly trend for accident correlation with lighting conditions hasn't shifted much over the years. As we can see, most of the accidents occur during the daylight. The number of accidents that happen during night time on roads with street light are very less compared to this. So probably, these accidents didn't occur because of insufficient lighting. In addition, chances that lights aren't lit if they're present are highly unlikely - This could be the reason why the fourth bar in every year is short.



After analyzing the trend in the frequency of accidents with weather conditions during the period 2009-2014, we realized that most of the accidents happen during fine weather without any wind or storm. Although some accidents are weather prone as disussed before, it definitely isn't the main cause of accidents.

3 Variable Correlations

We calculated the correlation matrix of the variables relating to the casualty variable. From the table, none of the variables seem to be highly correlated with any others.

	Longitude	Latitude	Police Force	Accident Severity	Number of vehicles	Day of Week	Time	_1st Road Class	Speed limit	_2nd Road_ lass	Urban or Rural Area	Number of Casualties
Longitude	1											
Latitude	-0.44	1										
Police_Force	-0.38	0.17	1									
Accident_Severity	0.05	-0.04	-0.06	1								
Number_of_Vehicles	0.01	-0.03	0	0.07	1							
Day_of_Week	0	0.01	0	0.01	0	1						
Time	-0.01	0.01	-0.02	0	0.01	0.04	1					
_1st_Road_Class	-0.05	0.05	0.04	0	-0.15	0	0.03	1				
Speed_limit	-0.07	0.06	0.21	-0.09	0.1	-0.01	-0.05	-0.38	1			
_2nd_Road_Class	0.06	-0.04	-0.11	0.07	0.06	0.01	0.03	0.1	-0.34	1		
Urban_or_Rural_Area	-0.11	0.06	0.25	-0.09	0.05	-0.02	-0.04	-0.22	0.68	-0.29	1	
Number_of_Casualties	-0.03	0.03	0.01	-0.06	0.23	0	0.02	-0.08	0.14	-0.03	0.11	1

4 Next Steps

5 References

[1] http://www.citymayors.com/gratis/uk_topcities.html