

Road Safety in UK 2016

Yang Dai, MPA Candidate, Evans School of Public Policy & Governance

March 2018

Introduction

There are more than 130,000 road accidents logged in UK in 2016. This study aims to understand where and when did these accidents happen, and what factors may be associated with certain types of accidents.

Dataset

Road safety data in UK in 2016

- There are 3 datasets: Accidents, Vehicle, Casualty. As it's easy to understand, an accident is mostly likely involve more than 1 vehicle, and it's likely to have multiple casualties from different vehicles. So Vehicle and Casualty are linked with Accidents by Accident_Index, Casualty and Vehicle are linked by Vehicle_Reference.

UK population data 2016 mid-year

UK shapefiles

Findings

1. There are more accidents happened in urban area than rural area.

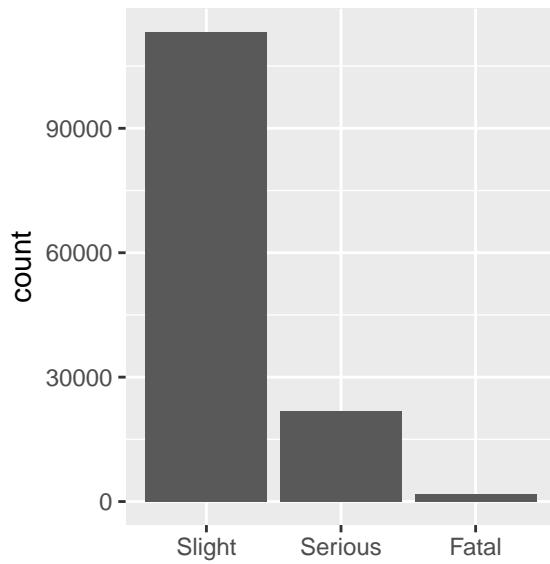
Area	Count
Urban	88468
Rural	48146

Area	Percent
Urban	0.6475764
Rural	0.3524236

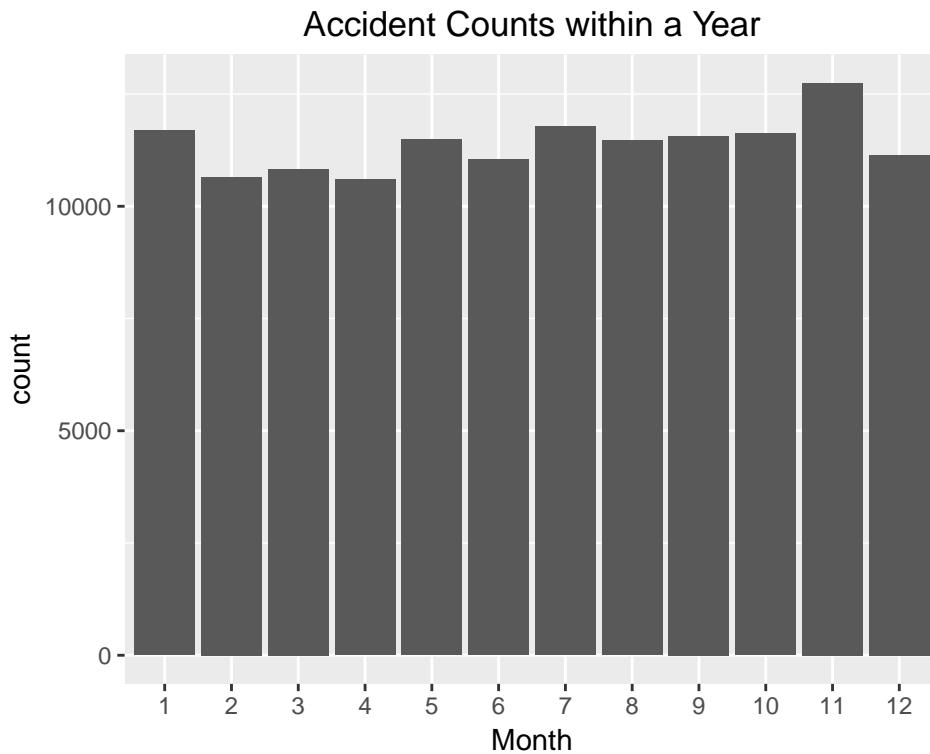
2. Most of the accidents are Slight, with few Serious accidents and even fewer Fatal ones.

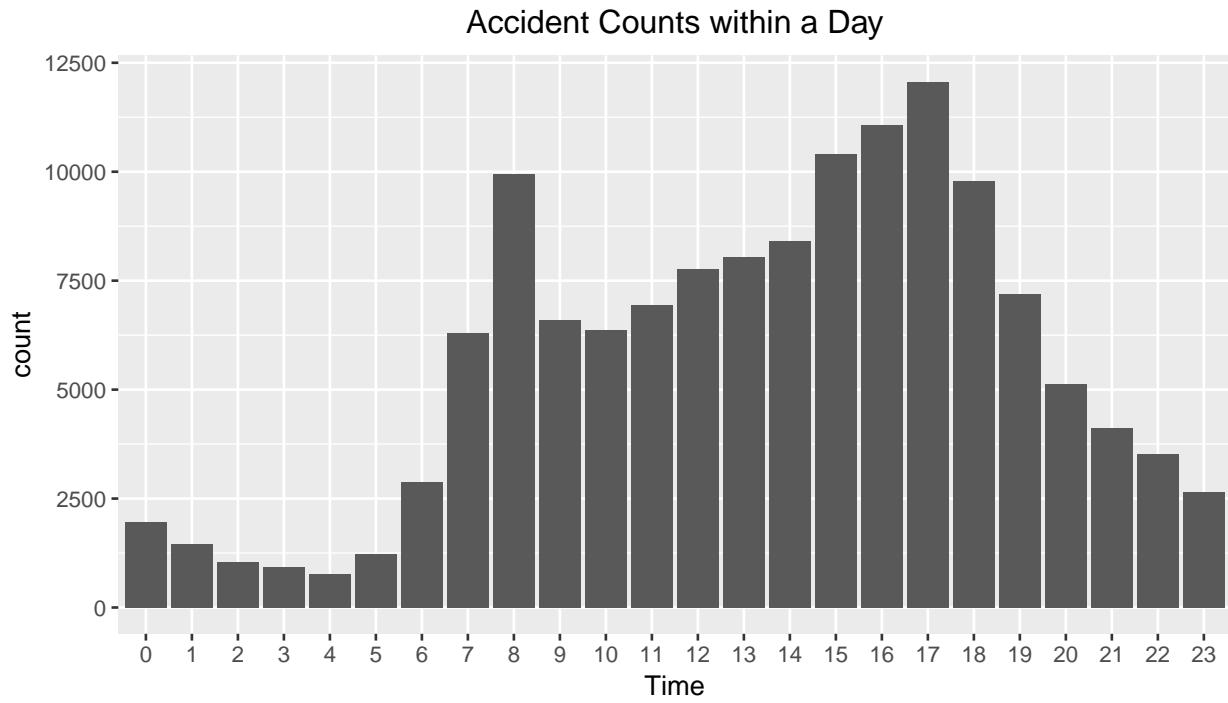
Accident Severity	Count
Slight	113201
Serious	21725
Fatal	1695

	Cum. Percent
Slight	0.8285769
Serious	0.9875934
Fatal	1.0000000

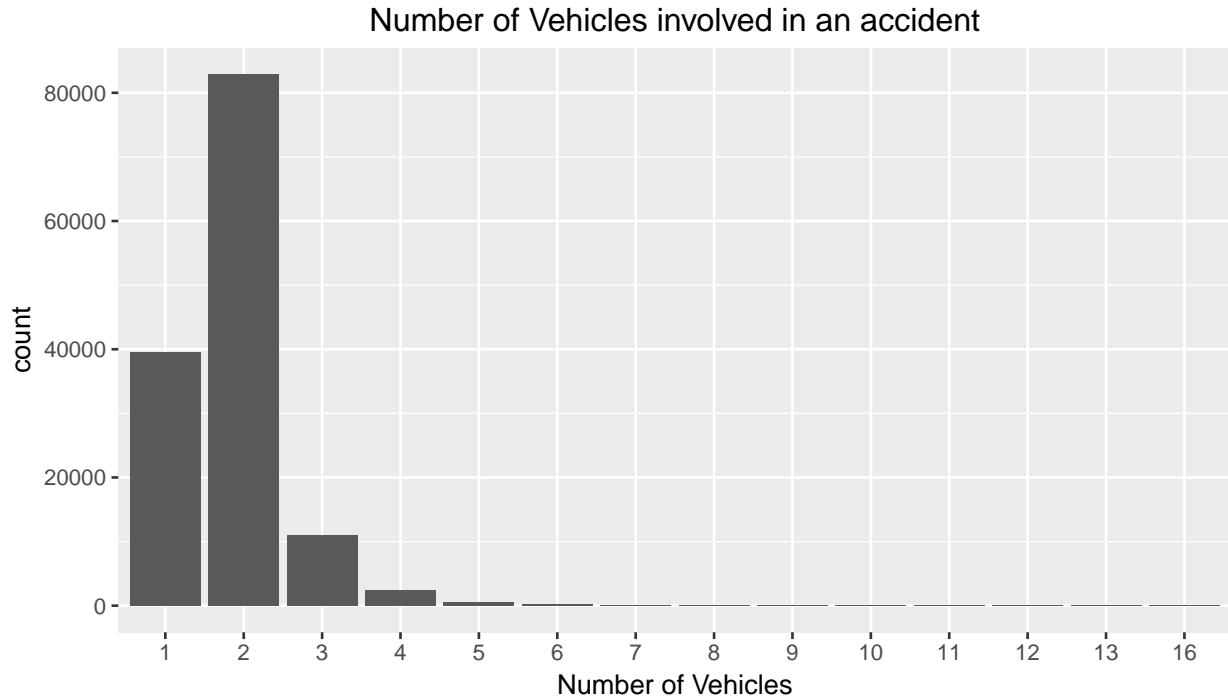


3. Most accidents happened during morning peak (8am) and afternoon peak (3pm-6pm). Accidents happened roughly equally (in amount) through 12 months.





- 99% of accidents involved 4 vehicles and below. 2 vehicles involved is most common.



- There are statistically significant yet small differences between the distribution of Accident Severity within each type of Road Surface Condition.

Cell Contents |—————|| N || N / Table Total |—————|

Total Observations in Table: 135852

| accidents_EDA\$Accident_Severity

accidents_EDA\$Road_Surface_Conditions	Slight	Serious	Fatal	Row Total
Dry 0.606	82336 0.116	15723 0.008	1147	99206
Wet or damp 0.207	28108 0.041	5592 0.004	520	34220
Snow 0.002	228 0.000	32 0.000	3	263
Frost or ice 0.012	1696 0.002	294 0.000	18	2008
Flood over 3cm. deep 0.001	131 0.000	22 0.000	2	155
Column Total	112499	21663	1690	135852

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 41.21899 d.f. = 8 p = 1.896968e-06

- t:

	Slight	Serious	Fatal
Dry	82336	15723	1147
Wet or damp	28108	5592	520
Snow	228	32	3
Frost or ice	1696	294	18
Flood over 3cm. deep	131	22	2

- prop.row:

	Slight	Serious	Fatal
Dry	0.8299	0.1585	0.01156
Wet or damp	0.8214	0.1634	0.0152
Snow	0.8669	0.1217	0.01141
Frost or ice	0.8446	0.1464	0.008964
Flood over 3cm. deep	0.8452	0.1419	0.0129

- prop.col:

	Slight	Serious	Fatal
Dry	0.7319	0.7258	0.6787

	Slight	Serious	Fatal
Wet or damp	0.2499	0.2581	0.3077
Snow	0.002027	0.001477	0.001775
Frost or ice	0.01508	0.01357	0.01065
Flood over 3cm. deep	0.001164	0.001016	0.001183

- prop.tbl:

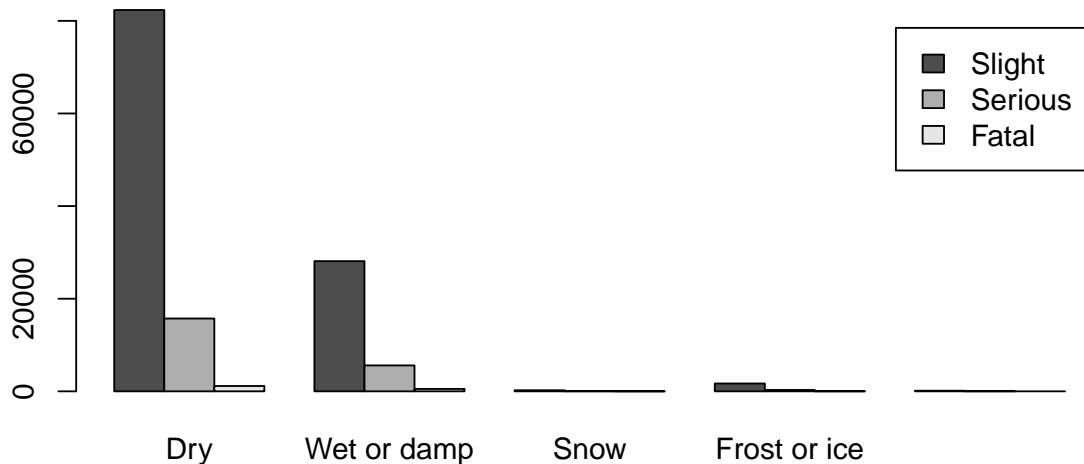
	Slight	Serious	Fatal
Dry	0.6061	0.1157	0.008443
Wet or damp	0.2069	0.04116	0.003828
Snow	0.001678	0.0002356	2.208e-05
Frost or ice	0.01248	0.002164	0.0001325
Flood over 3cm. deep	0.0009643	0.0001619	1.472e-05

- chisq:

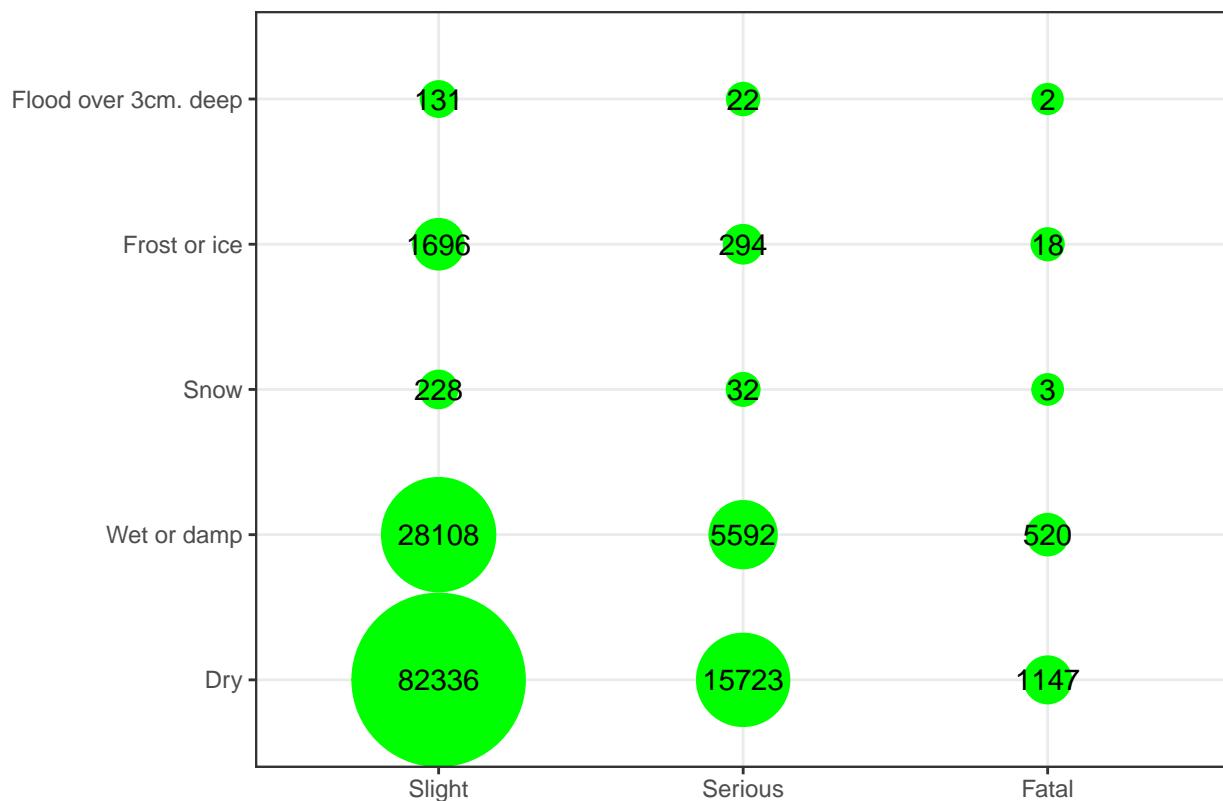
Table 10: Pearson's Chi-squared test: t

Test statistic	df	P value
41.22	8	1.897e-06 * * *

I think the data on "Snow", "Frost or ice", "Flood over 3cm. deep" is too few to draw some conclusion, but we can see there's slightly higher portion of Serious accidents on Wet or damp roads than on Dry road. Chi-squared test reject the null hypothesis that there is no difference in accident severity with different road surface conditions.



Accident Severity in different Road Surface Conditions



6. If there are more female driver involved in an accident, this accident tends to be a Slight one.

	CoefficientExp	Significant
(Intercept)	6.748	TRUE
Avg_Sex_of_Driver	1.547	TRUE
Avg_Age_of_Driver	0.9945	TRUE
Avg_Age_of_Vehicle	0.9824	TRUE

(I averaged Sex of Driver, Age of Driver, Age of Vehicle. It's easy to understand why I chose last 2 variables. The logic behind average Sex of Driver lay on how I code them. As you see above, I code Male as 0, Female as 1. Thus, if the average Sex of Driver is close to 1, that means there are more female drivers, and vice versa.)

7. Mapping the 'dangerous'(have higher total accident severity score) area In mapping, I want to code the accident severity differently from above. Because they coded 1 as Fatal, 3 as Slight, I want them to be reversed to reflect common sense (higher score means more dangerous).

I want to know which Local Authority Districts (LADs) are dangerous (have higher total accident severity score). So I aggregate the accident severity to the LAD level. Remember how I code the accident severity, I reversed the original code, so now 1 is Slight, 2 is Serious, and 3 is Fatal. And the aggregation of these value is a rough estimation (accident severity score) of how dangerous a LAD is. This contains a implicit (and not necessarily right) assumption that a fatal accident is 2 times more dangerous than a slight one. Accident Severity Rate is the ratio of severity to population.

Accident Severity Rate in UK 2016

