


101-106年台北市交通事故 資料分析

Business Analytics 107-1 Final Project | Group 4

江菩敏、李埏歆、李旻格、林南欣、陳廷語



Outline

- ◎ Data Cleaning
 - ◎ Exploratory Data Analysis
 - ◎ Building Models
 - ◎ Applications & Future Prospects
- 



Data Cleaning

資料維度



時間

年
月
日
小時
分鐘



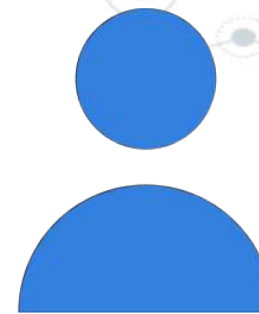
車禍嚴重程度

處理別
死亡人數
受傷人數
受傷程度



環境狀況

行政區序
地址
天氣
速限
道路型態
事故位置



個人資訊

當事人序
駕駛車種
性別
年齡

資料維度

◎ 當事人序：事故中肇事責任大小之順序

◎ 處理別：

1 → 造成人員當場或二十四小時內死亡之交通事故

2 → 造成人員受傷或超過二十四小時死亡之交通事故

發生年	發生月	發生日	發生時	發生分	處理別	區序	肇事地點	死亡人數	受傷人數	當事人序	車種	性別	年齡	受傷程度	4天候	7速限	8道路型態	9事故位置
106	2	8	19	5	201	大同區	大同區鄭	0	3	1	C03	1	44	2	6	40	5	2
106	2	8	19	5	201	大同區	大同區鄭	0	3	2	C03	1	34	2	6	40	5	2
106	2	8	19	5	201	大同區	大同區鄭	0	3	3	B03	1	50	3	6	40	5	2
106	2	8	19	5	201	大同區	大同區鄭	0	3	4	C03	1	27	2	6	40	5	2

調整資料品質

◎ 移除資料

- 性別：第3類（無、動物、堆置物）、第4類（肇事逃逸尚未查獲）
- 受傷程度：不明

◎ 更改資料

- 速限：3 → 30、440 → 40、0 → 200

調整資料型態

`as.ordered()`

◎ 速限

`as.factor()`

◎ 月份

◎ 季節

◎ 地區

◎ 24小時制

◎ 性別

◎ 駕駛車種

◎ 天氣

◎ 道路型態

◎ 事故位置

資料分組

季節

春、夏、秋、冬

24小時時段

夜間：0 - 6

上班尖峰時刻：7 - 9

中午：10 - 14

下午：15 - 16

下班尖峰時刻：17 - 19

晚間：20 - 23

駕駛車種

大卡/貨車

小型客貨車

機車

軍事用車

特種車

慢車

其他車種

人

天氣

暴雨、強風、風沙、霧或煙、雪

雨天

陰天

晴天

道路型態

交叉路

單路

圓環廣場

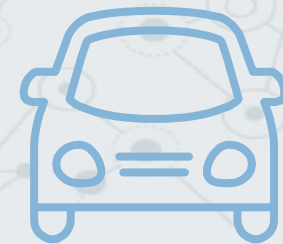
事故位置

交叉路口

路段

交流道

其他



Exploratory Data Analysis

反應變數

◎ 車禍總次數

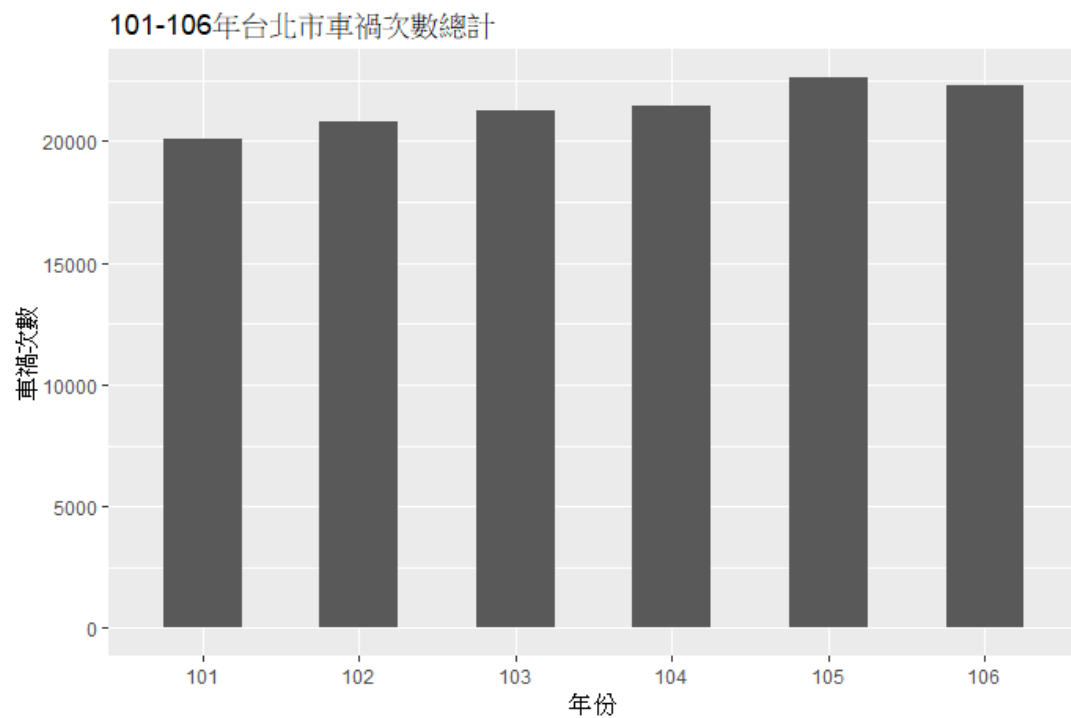
- 擷取當事人序為1的資料，避免同一場車禍有重複的資料
- 解釋變數：發生年、發生月、發生時、處理別、區序、肇事地點、天候、速限、道路型態和事故位置
- 適用一般線性迴歸模型或廣義線性模型中的Poisson

◎ 車禍受傷程度

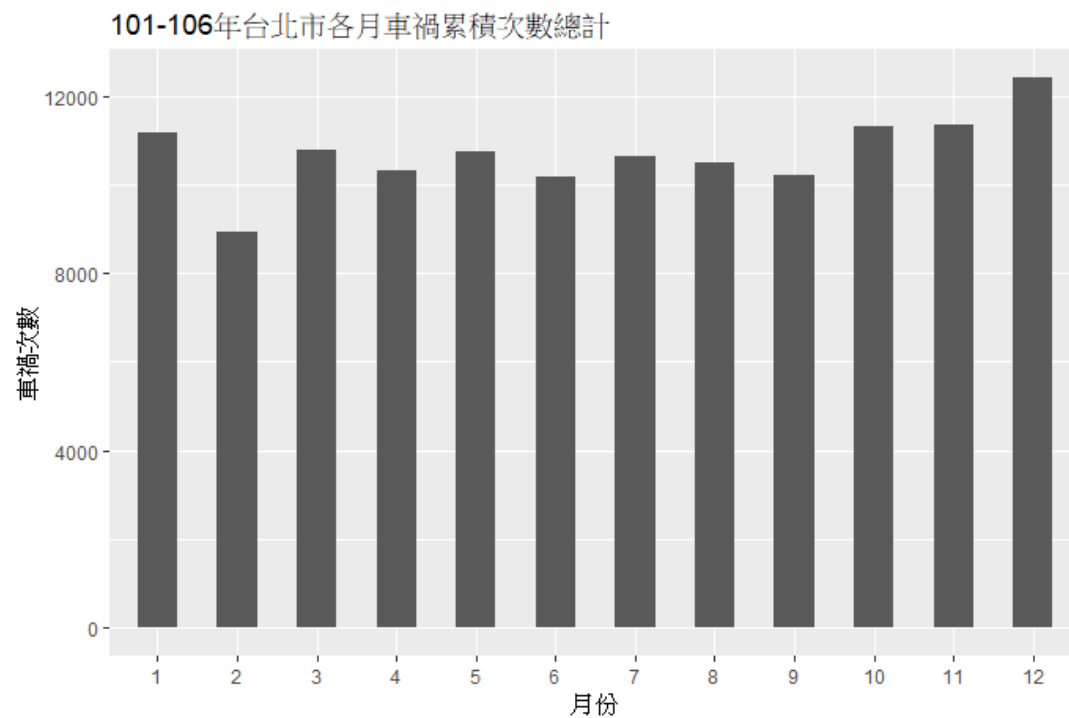
- 採用所有資料，以人為單位
- 解釋變數：年齡、性別和車種
- 適用廣義線性模型中的Binomial

車禍總次數 | 發生年&月

◎ 車禍總次數逐年增加

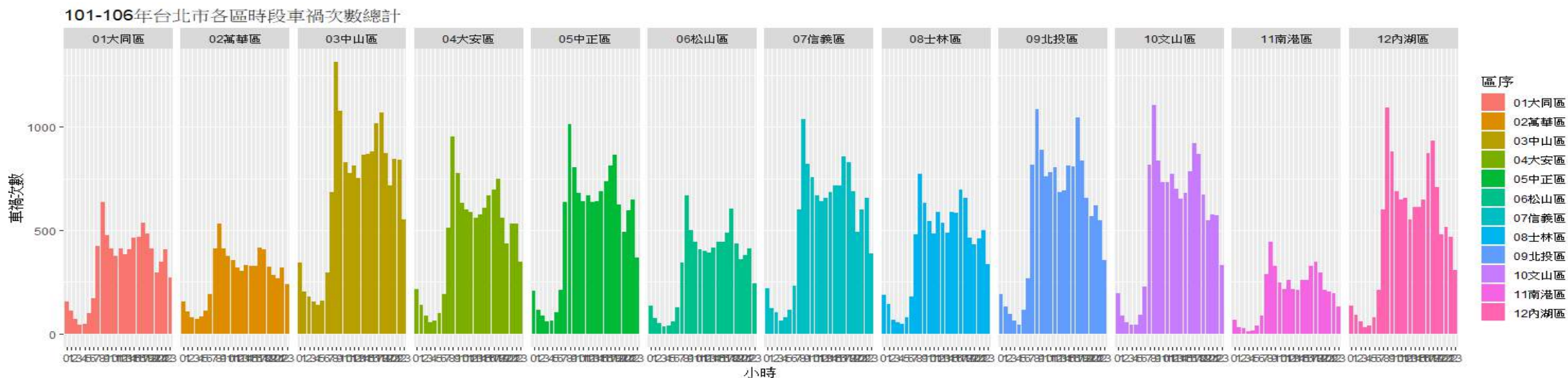


◎ 2月次數較少，12月為高峰。



車禍總次數 | 發生時段 & 區域

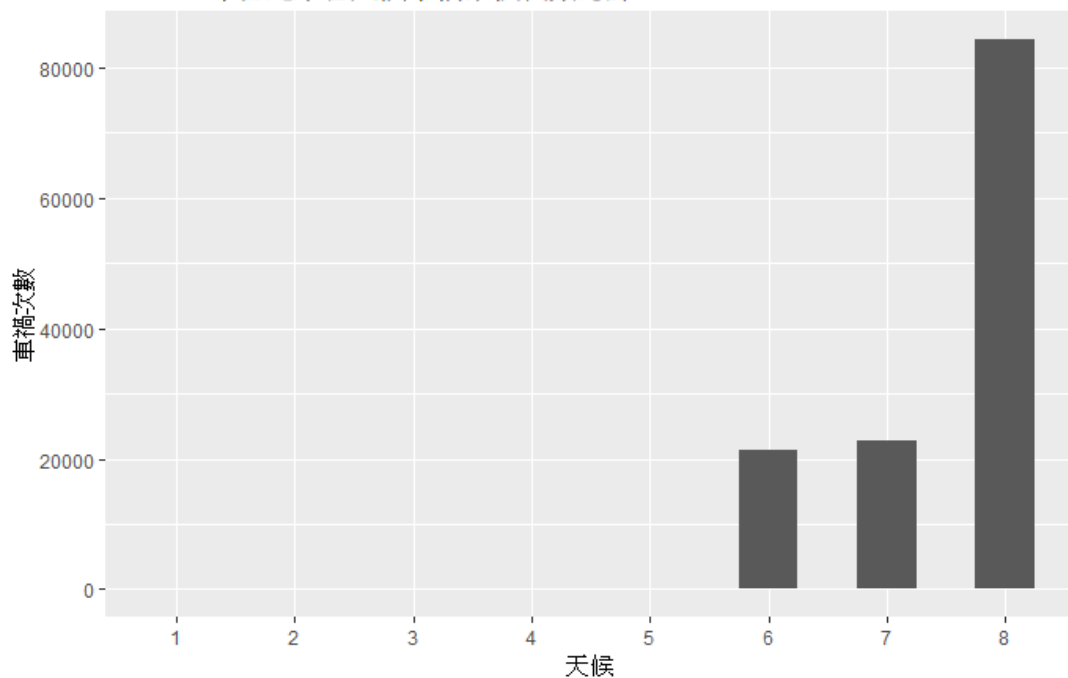
- ◎ 時段：雙峰分布，上、下班交通尖峰時段最為頻繁
- ◎ 中山區、大安區、中正區、信義區：21:00 - 22:00車禍次數較多
 - 公司行號較多、夜生活較為豐富、人口數較多
- ◎ 中山區為首，主要事故地點：民權西路路段
 - 新北市通勤族必經路段



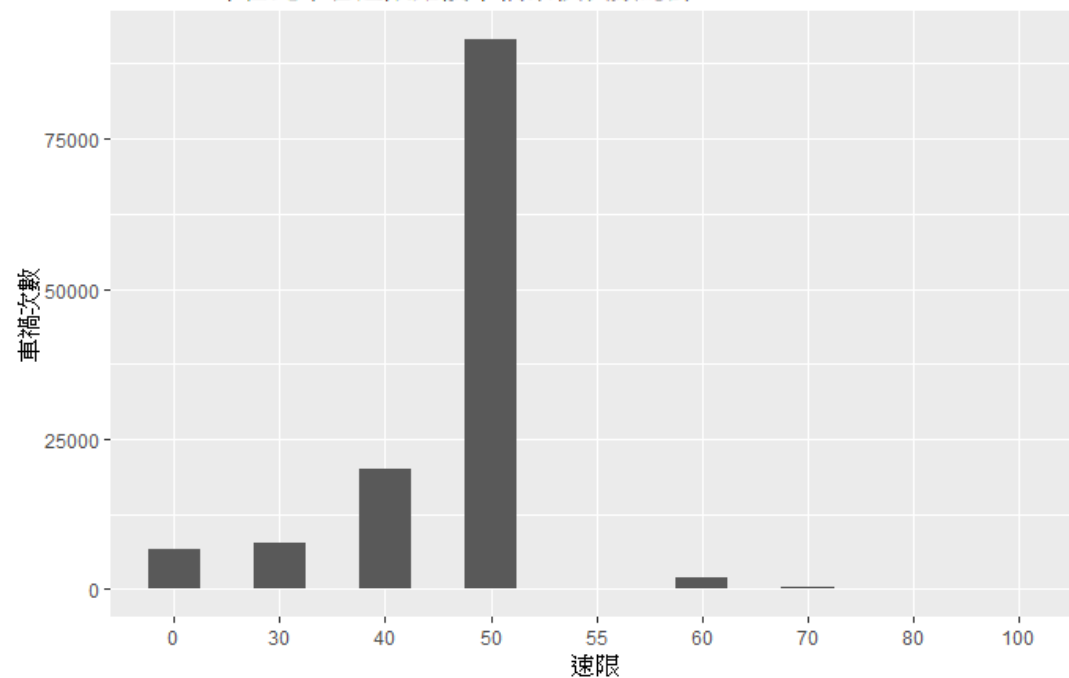
車禍總次數 | 天候 & 速限

- ◎ 天候：晴天、雨天、陰天次數多
- ◎ 速限：0（無速限）、30 - 50公里次數多

101-106年台北市各天候車禍累積次數總計



101-106年台北市各速限路段車禍累積次數總計



車禍總次數 | 道路型態 & 事故位置

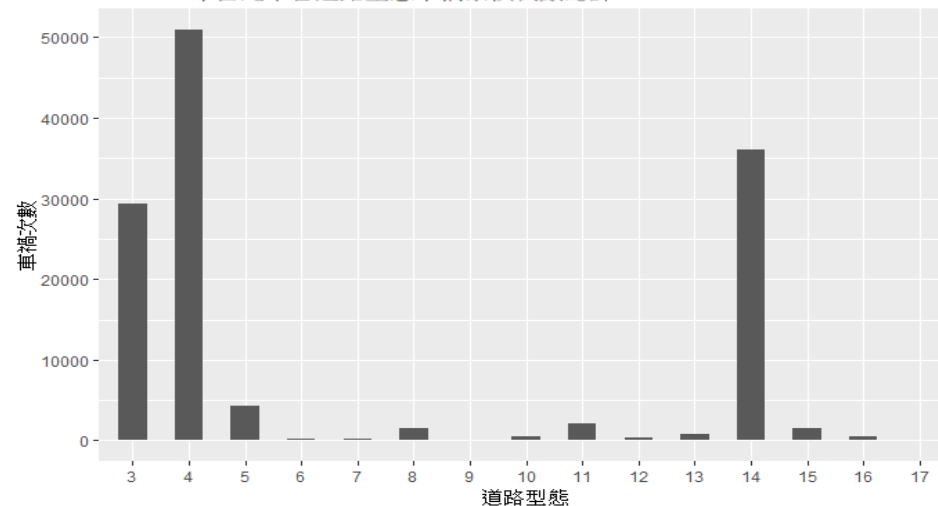
◎ 道路型態與事故位置相關

- 事故位置：交叉路口、一般車道
- 道路型態：三叉路、四叉路、直路

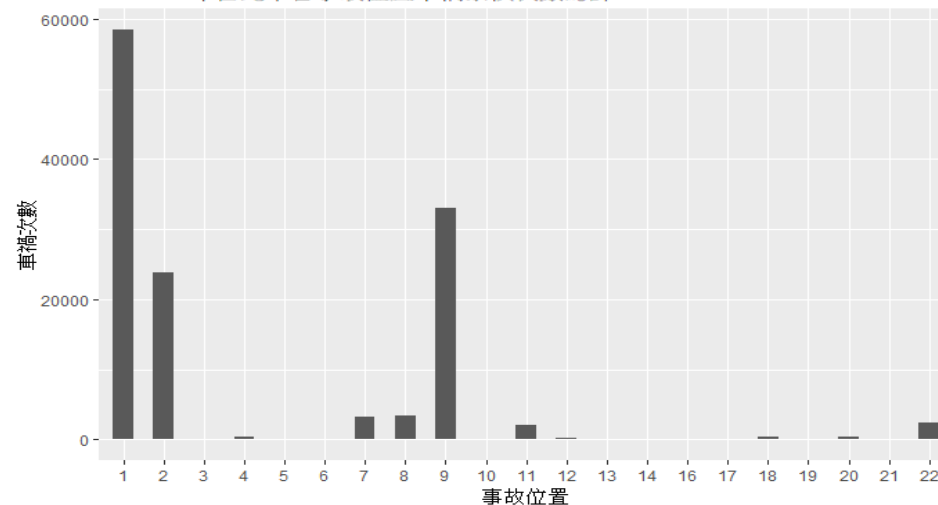
◎ 受各區域道路設計影響

- 北投區：彎曲路段、機車專用道

101-106年台北市各道路型態車禍累積次數總計

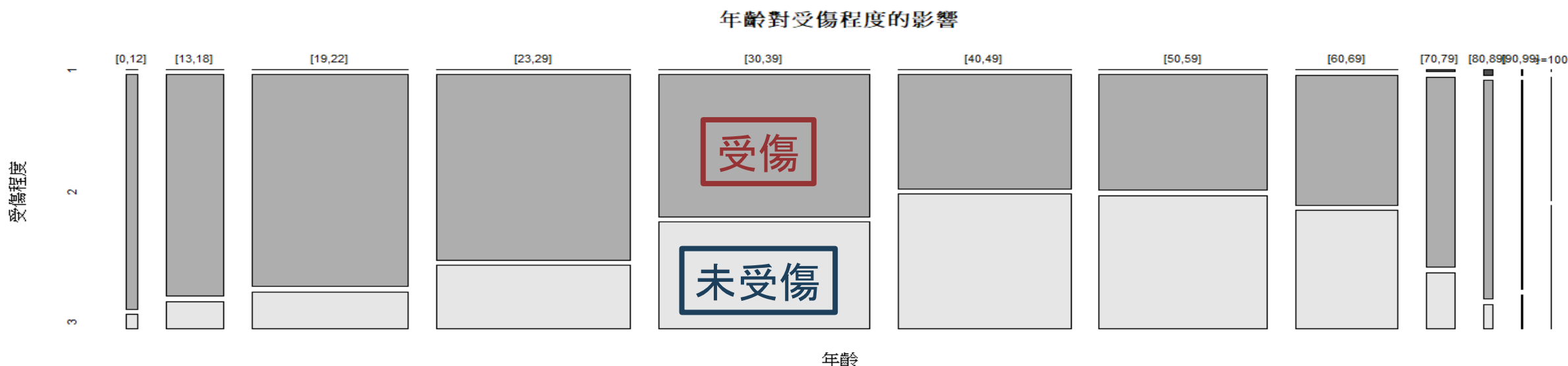


101-106年台北市各事故位置車禍累積次數總計



車禍受傷程度 | 年齡

- ◎ 年齡分佈：青少年、中壯年居多
- ◎ 隨著年齡增加，發生車禍於24小時內死亡的比例些微增加（70歲以上明顯）
- ◎ 未受傷的比例：隨年齡的增加而增加，50歲左右逐漸下降，呈非線性關係



車禍受傷程度 | 性別 & 車種

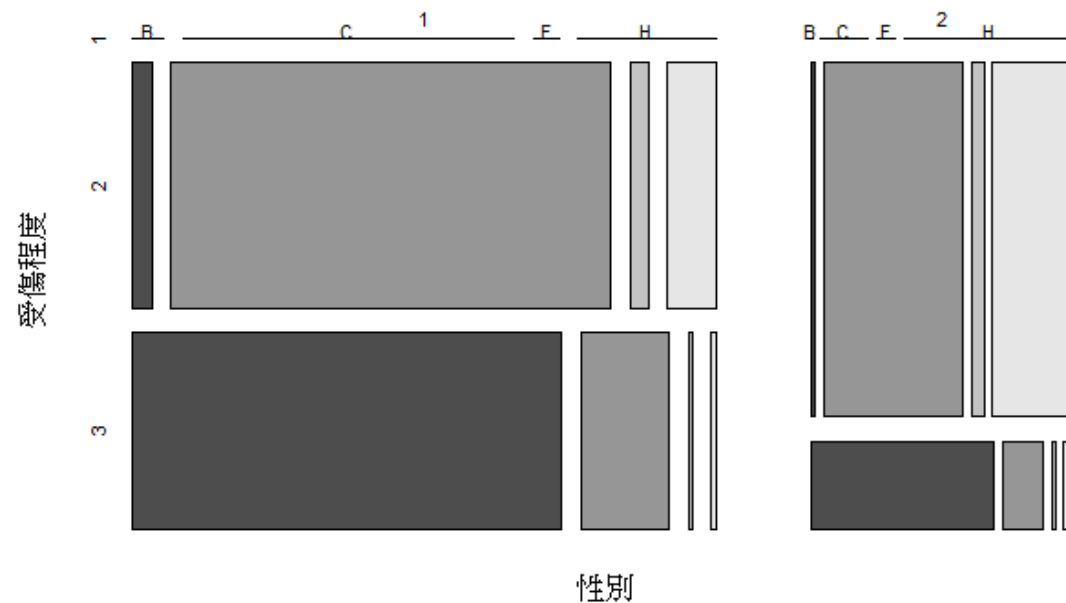
◎ 受傷比例：行人或乘客 > 自行車 > 機車 > 汽車

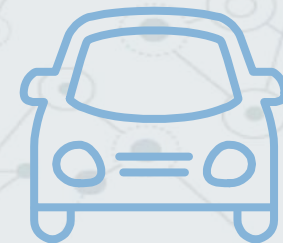
◎ 遭遇車禍次數：

- 機車 > 汽車 > 行人或乘客 > 自行車
- 男性 > 女性

◎ 未受傷比例：男性 > 女性

性別和車種對受傷程度的影響





Building Models

Model 1：車禍嚴重程度 | 問題定義

	Model 1	Model 2	Model 3
問題定義	車禍嚴重程度 和哪些因素相關	車禍當事人傷亡狀況 和哪些因素相關	肇事者傷亡狀況 和哪些因素相關
分析單位	車禍	車禍當事人	肇事者
應變數	處理別	受傷程度	受傷程度
自變數	時間、環境狀況	個人資訊	時間、環境狀況、 個人資訊

Model 1 : 車禍嚴重程度

最終模型自變數：

1. 事故位置
2. 時段

```
Call:
glm(formula = adminstration ~ accident_place + hr, family = binomial,
     data = tp1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2446  -0.0842  -0.0741  -0.0709   3.5332

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -6.23989    0.18767  -33.250  < 2e-16 ***
accident_placeOnTheRoad  0.34458    0.09753   3.533 0.000411 ***
accident_placeOthers    0.79698    0.22808   3.494 0.000475 ***
accident_placeSlipRoad -11.03431   234.04814  -0.047 0.962397
hrAroundNoon         0.24871    0.21109   1.178 0.238714
hrEvening            0.56108    0.21595   2.598 0.009373 **
hrEveningRushHR      0.25537    0.22155   1.153 0.249066
hrMorningRushHR      0.30941    0.21760   1.422 0.155040
hrNight              1.94854    0.20298   9.600  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6049.4  on 117304  degrees of freedom
Residual deviance: 5834.1  on 117296  degrees of freedom
AIC: 5852.1

Number of Fisher Scoring iterations: 15

> #whether it's a good model?
> deviance(g3)
[1] 5834.099
> 1-pchisq(deviance(g3),g3$df.residual) # p-value = 1 ???
[1] 1
```

Model 1 : 車禍嚴重程度 | 模型解釋

◎ 夜間時段 (0-6)

- 可見度較低
 - 注意力較難集中
 - 車流量小
- 夜間車速普遍較快



時間

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-6.23989	0.18767	-33.250	< 2e-16	***
accident_placeOnTheRoad	0.34458	0.09753	3.533	0.000411	***
accident_placeOthers	0.79698	0.22808	3.494	0.000475	***
accident_placeSlipRoad	-11.03431	234.04814	-0.047	0.962397	
hrAroundNoon	0.24871	0.21109	1.178	0.238714	
hrEvening	0.56108	0.21595	2.598	0.009373	**
hrEveningRushHR	0.25537	0.22155	1.153	0.249066	
hrMorningRushHR	0.30941	0.21760	1.422	0.155040	
hrNight	1.94854	0.20298	9.600	< 2e-16	***

Model 2：車禍當事人傷亡狀況 | 問題定義

	Model 1	Model 2	Model 3
問題定義	車禍嚴重程度 和哪些因素相關	車禍當事人傷亡狀況 和哪些因素相關	肇事者傷亡狀況 和哪些因素相關
分析單位	車禍	車禍當事人	肇事者
應變數	處理別	受傷程度	受傷程度
自變數	時間、環境狀況	個人資訊	時間、環境狀況、 個人資訊

Model 2 : 車禍當事人傷亡狀況

最終模型自變數：

1. 駕駛車種
2. 性別
3. 年齡
4. 年齡平方

Call:

```
glm(formula = how_bad_hurt ~ car_type + gender + age + agesq,  
     family = binomial, data = t)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6079	-0.3176	0.3817	0.5409	2.7465

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.221e+00	4.842e-02	45.876	< 2e-16	***
car_typeBus/Trucks	-5.869e+00	1.115e-01	-52.627	< 2e-16	***
car_typeCar	-5.107e+00	4.546e-02	-112.340	< 2e-16	***
car_typeHuman	4.143e-01	5.043e-02	8.214	< 2e-16	***
car_typeMilitary	-4.653e+00	1.048e+00	-4.439	9.05e-06	***
car_typeMotorcycle	-3.364e-01	4.420e-02	-7.610	2.75e-14	***
car_typeOthers	-2.981e+00	1.616e-01	-18.447	< 2e-16	***
car_typeSpecial Use	-3.449e+00	1.619e-01	-21.308	< 2e-16	***
gender2	7.312e-01	1.651e-02	44.296	< 2e-16	***
age	-1.849e-03	5.183e-04	-3.568	0.00036	***
agesq	5.420e-06	2.327e-06	2.329	0.01988	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 361783 on 272728 degrees of freedom
Residual deviance: 156714 on 272718 degrees of freedom
AIC: 156736

Number of Fisher Scoring iterations: 6

```
> #whether it's a good model?
```

```
> deviance(g2_2)
```

```
[1] 156713.8
```

```
> 1-pchisq(deviance(g2_2),g2_2$df.residual) # p-value = 1 ???
```

```
[1] 1
```

Model 2：車禍當事人傷亡狀況 | 模型解釋

◎ 交通工具的安全防護設計

大多以男性為保護目標

◎ 哈佛公共衛生學院研究：

在安全氣囊正常運作下，

女性傷亡比高於男性



性別

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.221e+00	4.842e-02	45.876	< 2e-16	***
car_typeBus/Trucks	-5.869e+00	1.115e-01	-52.627	< 2e-16	***
car_typeCar	-5.107e+00	4.546e-02	-112.340	< 2e-16	***
car_typeHuman	4.143e-01	5.043e-02	8.214	< 2e-16	***
car_typeMilitary	-4.653e+00	1.048e+00	-4.439	9.05e-06	***
car_typeMotorcycle	-3.364e-01	4.420e-02	-7.610	2.75e-14	***
car_typeOthers	-2.981e+00	1.616e-01	-18.447	< 2e-16	***
car_typeSpecial Use	-3.449e+00	1.619e-01	-21.308	< 2e-16	***
gender2	7.312e-01	1.651e-02	44.296	< 2e-16	***
age	-1.849e-03	5.183e-04	-3.568	0.00036	***
agesq	5.420e-06	2.327e-06	2.329	0.01988	*

Model 3：肇事者傷亡狀況 | 問題定義

	Model 1	Model 2	Model 3
問題定義	車禍嚴重程度 和哪些因素相關	車禍當事人傷亡狀況 和哪些因素相關	肇事者傷亡狀況 和哪些因素相關
分析單位	車禍	車禍當事人	肇事者
應變數	處理別	受傷程度	受傷程度
自變數	時間、環境狀況	個人資訊	時間、環境狀況、 個人資訊

Model 3 : 肇事者傷亡狀況

最終模型自變數：

1. 駕駛車種
2. 性別
3. 天氣
4. 事故位置
5. 時段

```
Call:
glm(formula = how_bad_hurt ~ car_type + gender + weather + accident_place +
    hr, family = binomial, data = tp1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6316	-0.2832	-0.2301	0.6083	2.8978

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.309741	0.066362	19.736	< 2e-16	***
car_typeBus/Trucks	-5.264737	0.148699	-35.405	< 2e-16	***
car_typeCar	-4.768837	0.059529	-80.109	< 2e-16	***
car_typeHuman	-0.306850	0.086661	-3.541	0.000399	***
car_typeMilitary	-12.175906	48.159656	-0.253	0.800404	
car_typeMotorcycle	-0.006985	0.056374	-0.124	0.901386	
car_typeOthers	-2.966728	0.257818	-11.507	< 2e-16	***
car_typeSpecial Use	-3.772833	0.338635	-11.141	< 2e-16	***
gender2	0.679618	0.023950	28.376	< 2e-16	***
weatherExtreme	0.746120	0.362778	2.057	0.039716	*
weatherRainy	0.046579	0.034286	1.359	0.174295	
weatherSunny	-0.023873	0.026448	-0.903	0.366729	
accident_placeOnTheRoad	0.290713	0.021139	13.752	< 2e-16	***
accident_placeOthers	-0.228542	0.059911	-3.815	0.000136	***
accident_placeSlipRoad	2.287122	0.251365	9.099	< 2e-16	***
hrAroundNoon	-0.004102	0.035423	-0.116	0.907819	
hrEvening	0.038683	0.038173	1.013	0.310888	
hrEveningRushHR	-0.195775	0.036893	-5.307	1.12e-07	***
hrMorningRushHR	-0.135220	0.036406	-3.714	0.000204	***
hrNight	0.894604	0.049950	17.910	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 159735 on 117304 degrees of freedom
Residual deviance: 71730 on 117285 degrees of freedom
AIC: 71770

Number of Fisher Scoring iterations: 9

```
> #whether it's a good model?
> deviance(g2)
[1] 71729.84
> 1-pchisq(deviance(g2),g2$df.residual) # p-value = 1 ???
[1] 1
```

Model 3 : 肇事者傷亡狀況 | 模型解釋

◎ 最容易發生嚴重傷害

1. 自行車
2. 行人

車種



Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.309741	0.066362	19.736	< 2e-16	***
car_typeBus/Trucks	-5.264737	0.148699	-35.405	< 2e-16	***
car_typeCar	-4.768837	0.059529	-80.109	< 2e-16	***
car_typeHuman	-0.306850	0.086661	-3.541	0.000399	***
car_typeMilitary	-12.175906	48.159656	-0.253	0.800404	
car_typeMotorcycle	-0.006985	0.056374	-0.124	0.901386	
car_typeOthers	-2.966728	0.257818	-11.507	< 2e-16	***
car_typeSpecial Use	-3.772833	0.338635	-11.141	< 2e-16	***
gender2	0.679618	0.023950	28.376	< 2e-16	***
weatherExtreme	0.746120	0.362778	2.057	0.039716	*
weatherRainy	0.046579	0.034286	1.359	0.174295	
weatherSunny	-0.023873	0.026448	-0.903	0.366729	
accident_placeOnTheRoad	0.290713	0.021139	13.752	< 2e-16	***
accident_placeOthers	-0.228542	0.059911	-3.815	0.000136	***
accident_placeSlipRoad	2.287122	0.251365	9.099	< 2e-16	***
hrAroundNoon	-0.004102	0.035423	-0.116	0.907819	



Applications & Future Prospects

實際應用

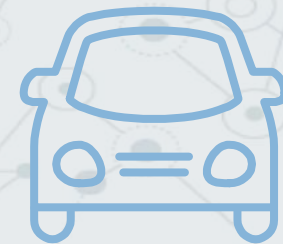


未來展望

資料
視覺化

資料收集
項目增加





Q & A

Goodness-of-fit Test : $p\text{-value} = 1$

車種對受傷程度的影響

