

資料探勘導論

組員

01

劉灼庭 04170108

02

蕭少平 04170141

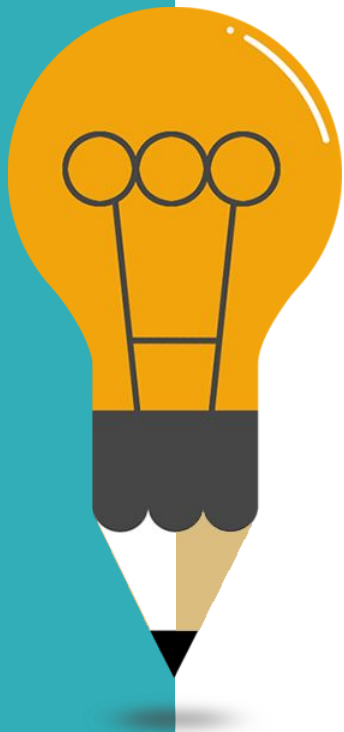
03

林茵婷 04170143

04

張雁婷 04170144

目錄



01

組員分工

02

選擇站點

03

預測結果

04

北上小客車

05

南下小客車

06

北上大客車

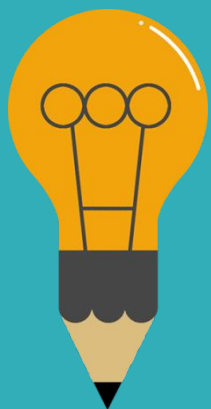
07

南下大客車

08

結論

組員 分工



劉灼庭同學：**北上小客車**模型建立與分析預測

張雁婷同學：**北上大客車**模型建立與分析預測

林茵婷同學：**南下小客車**模型建立與分析預測

蕭少平同學：**南下大客車**模型建立與分析預測

資料前處理、討論與各式各樣報告必須作業為大家平分謝謝！

選擇 站點 原因

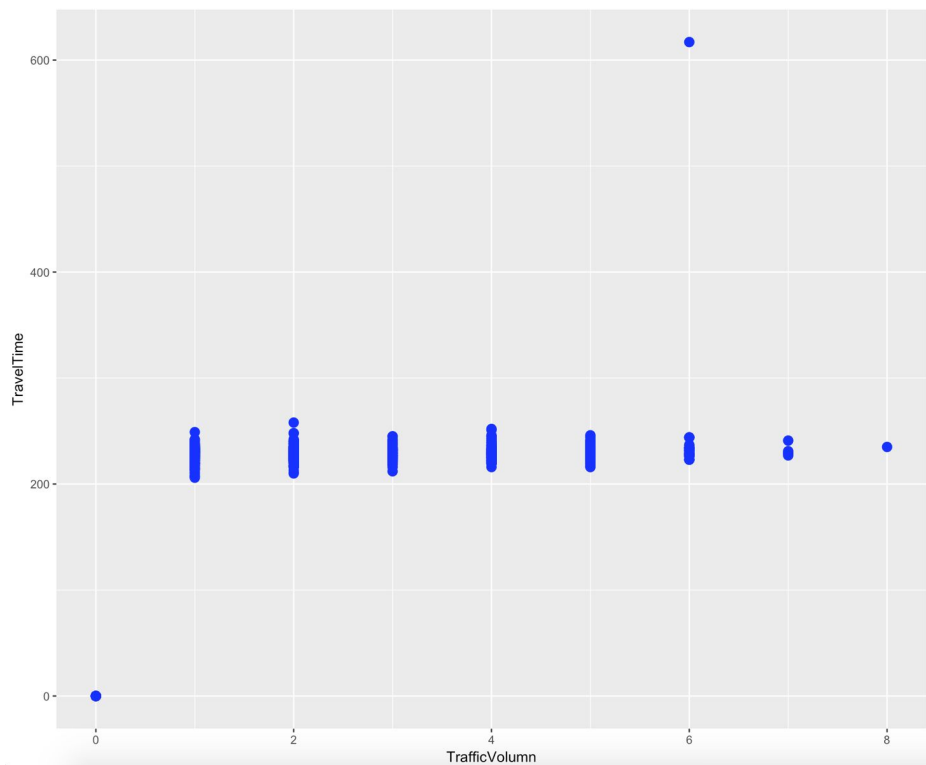
南港-石碇

- 1 選擇車禍發生比較少的路段
- 2 選擇有大客車通道的路段
- 3 考慮的因素比較全面性
- 4 往來台北宜蘭一定會經過

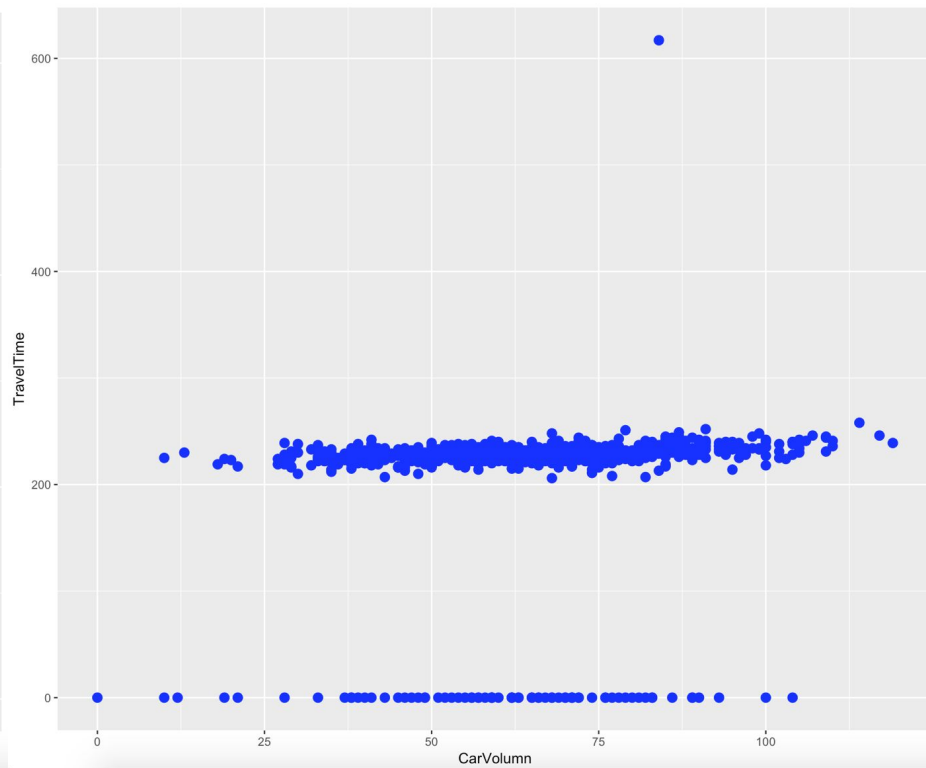


北上。大客車。8點

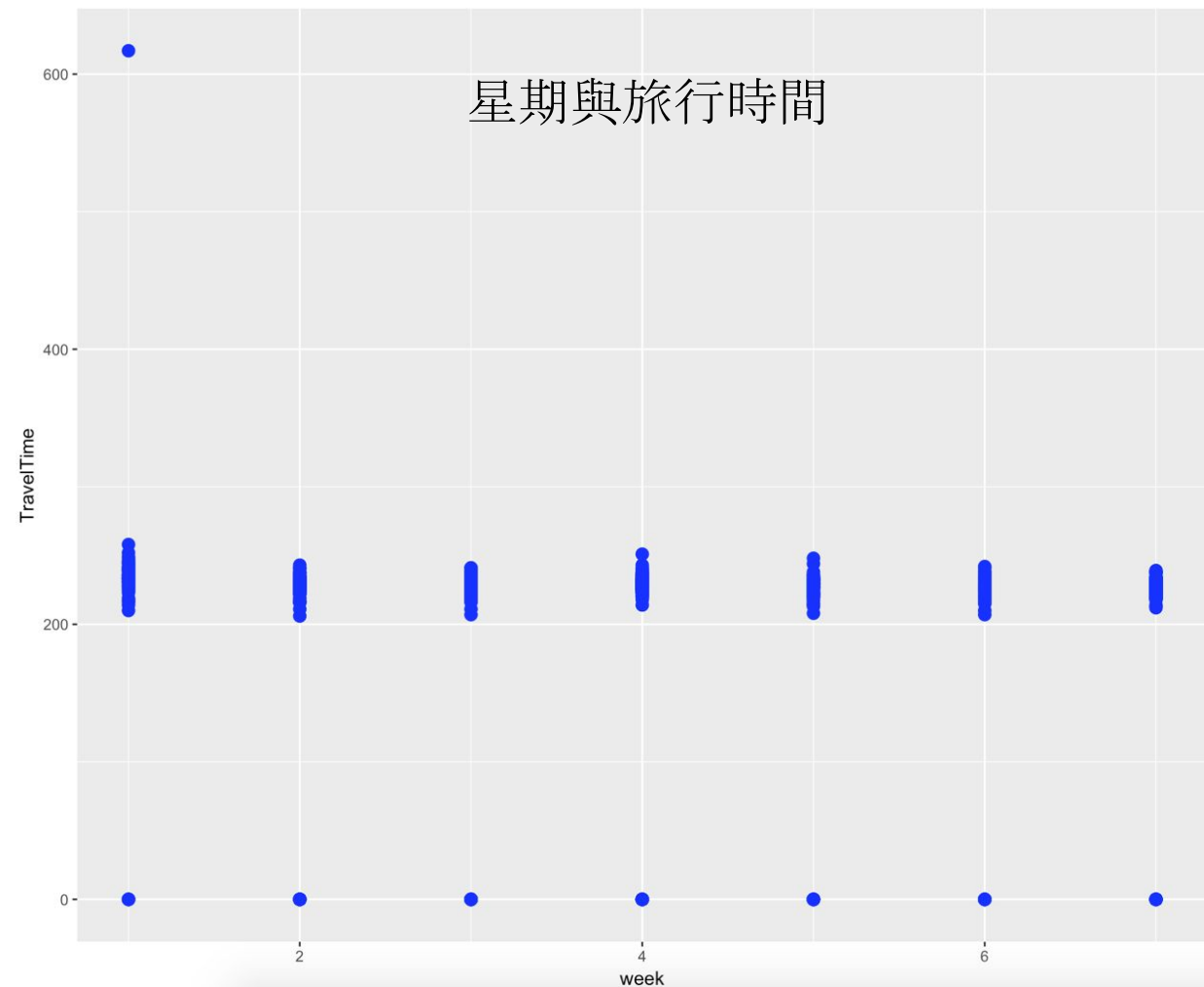
大客車流量與旅行時間



小客車流量與旅行時間



星期與旅行時間



✓ 星期一的旅行時間稍多一點

TrafficVolumn+CarVolumn+
mon(是否為星期一)

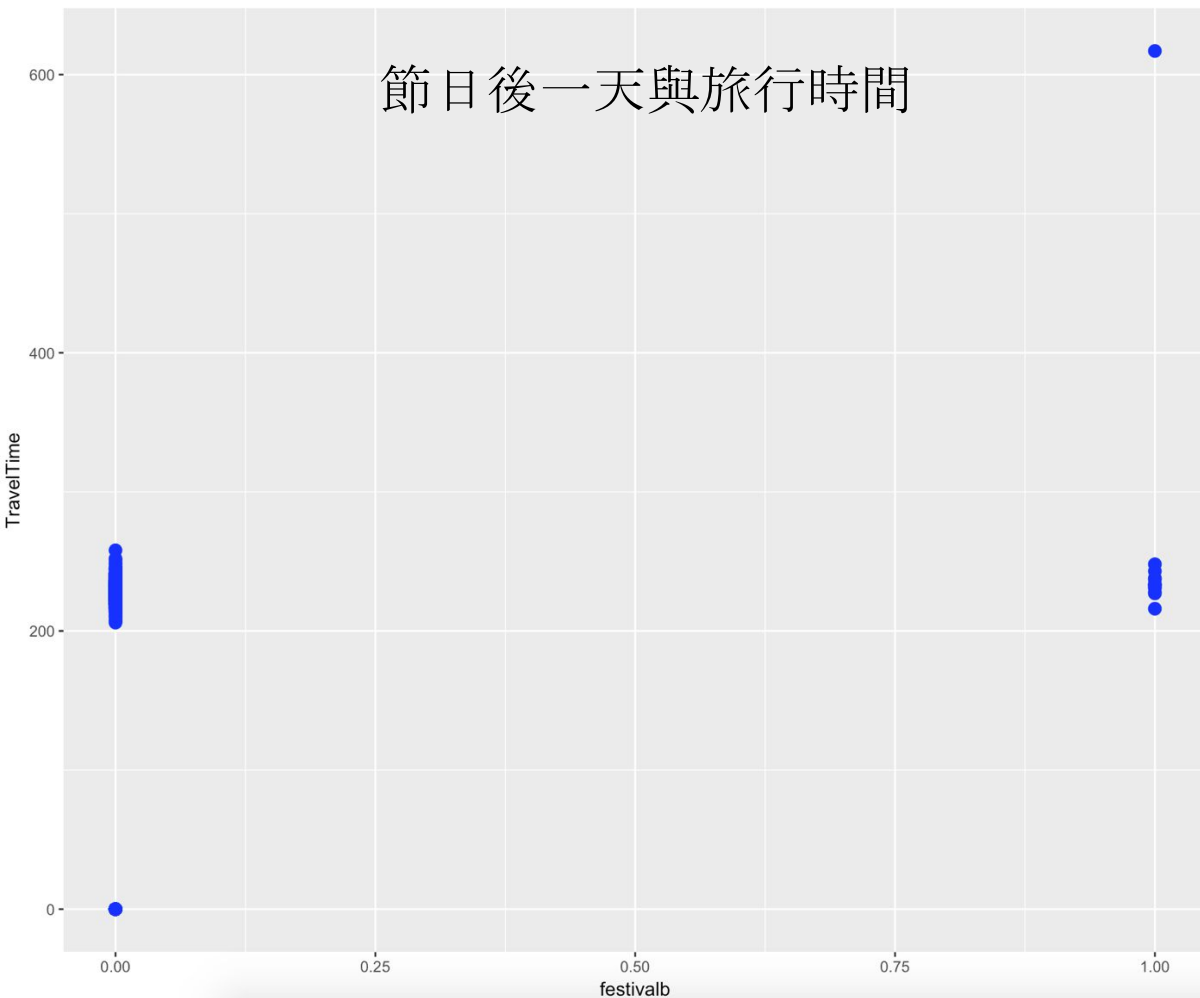
$R^2=0.2649$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	144.3594	8.8824	16.252	<2e-16 ***
mon	8.0742	7.4980	1.077	0.282
TrafficVolumn	22.5375	1.4403	15.648	<2e-16 ***
CarVolumn	0.1320	0.1345	0.981	0.327

TimeInterval	GantryFrom	GantryTo	VehicleType	TravelTime	CarVolumn	TrafficVolumn
2016/6/13 8:00	05F0055N	05F0001N	41	617	84	6
2016/6/14 8:00	05F0055N	05F0001N	41	235	66	3
2016/6/15 8:00	05F0055N	05F0001N	41	229	51	1
2016/6/16 8:00	05F0055N	05F0001N	41	232	66	3
2016/6/17 8:00	05F0055N	05F0001N	41	228	60	5
2016/6/18 8:00	05F0055N	05F0001N	41	241	59	5
2016/6/19 8:00	05F0055N	05F0001N	41	0	48	0
2016/6/20 8:00	05F0055N	05F0001N	41	231	102	2

節日後一天與旅行時間



**TrafficVolumn+CarVolumn+
mon+festivalb**
R2=0.2743

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	148.09026	8.91445	16.612	< 2e-16	***
mon	9.74220	7.47463	1.303	0.19286	
TrafficVolumn	22.54948	1.43202	15.747	< 2e-16	***
CarVolumn	0.05431	0.13609	0.399	0.68998	
festivalb	50.90756	16.57828	3.071	0.00222	**

✓關係明顯較小客車流量顯著, 將小客車流量、mon變量刪除

TrafficVolumn+festivalb

R2=0.2711

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	152.177	4.001	38.032	< 2e-16	***
TrafficVolumn	22.873	1.422	16.089	< 2e-16	***
festivalb	52.916	16.297	3.247	0.00122	**

調整空值與離群值

✓ 使用三期平均補空值及調整離群值

TrafficVolumn+festivalb

R2=0.03657

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	226.6187	0.5126	442.127	< 2e-16	***
festivalb	4.7999	1.7677	2.715	0.00678	**
TrafficVolumn	0.8047	0.1745	4.611	4.74e-06	***

預測6/14

$$y = 152.177 + 52.916[\text{是否為節日後一天}] + 22.873[\text{大客車流量}]$$

$x_1 = 0$, $x_2 = 3$, $y = 221$

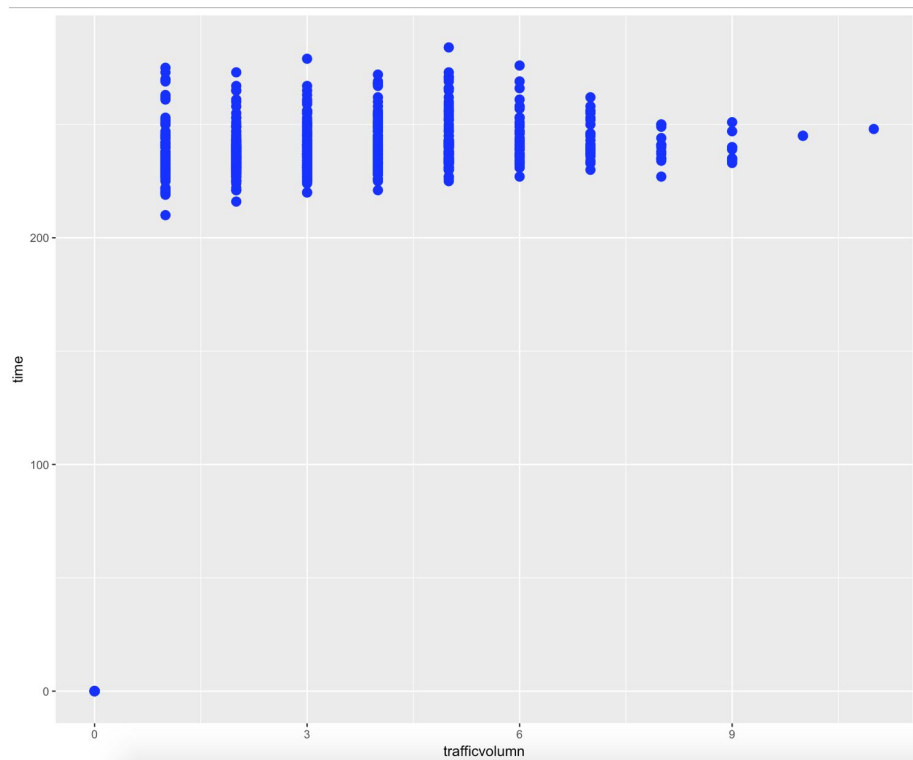
```
> vif(regN841.lm)
festivalb TrafficVolumn
1.001875      1.001875
```

2018/6/14 8:00	01F0017N	01F0005N	31	45	41
2018/6/14 8:00	05F0055N	05F0001N	31	210	68
2018/6/14 8:00	05F0055N	05F0001N	32	213	15
2018/6/14 8:00	05F0055N	05F0001N	41	222	1
2018/6/14 8:00	05F0055N	05F0001N	42	0	0
2018/6/14 8:00	05F0055N	05F0001N	5	0	0

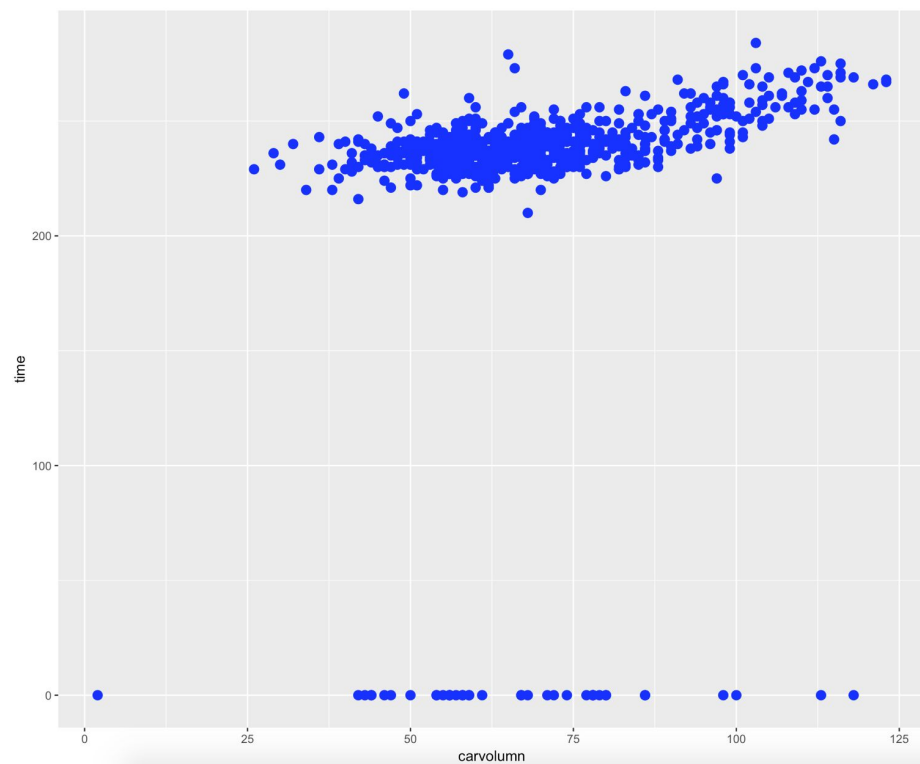
北上。大客車。13點

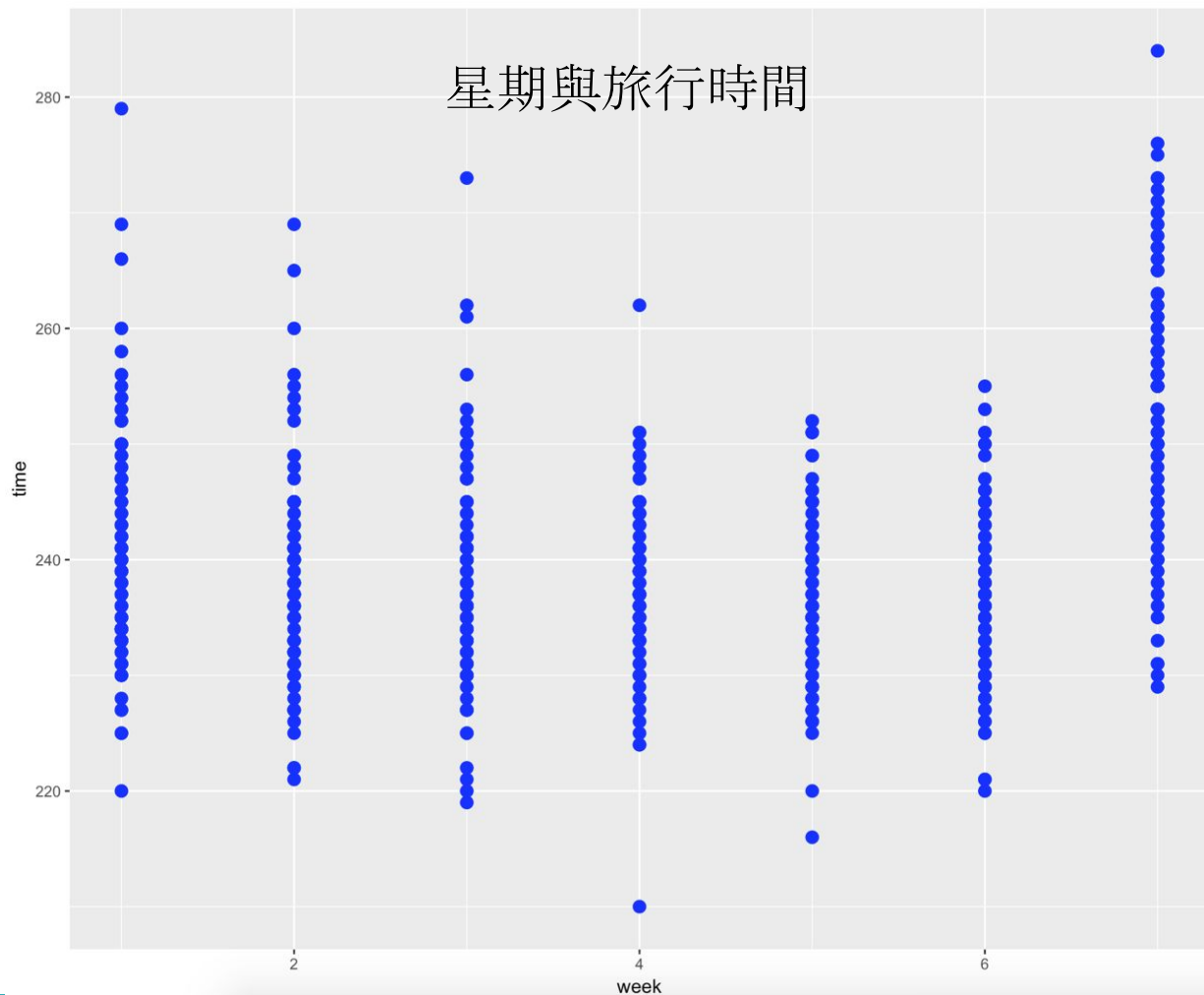
timeinterval	from	to	vehicletype	time	carvolumn	trafficvolumn
2016/6/13 13:00	05F0055N	05F0001N	41	250	58	8
2016/6/14 13:00	05F0055N	05F0001N	41	256	60	3
2016/6/15 13:00	05F0055N	05F0001N	41	241	67	7
2016/6/16 13:00	05F0055N	05F0001N	41	247	48	6
2016/6/17 13:00	05F0055N	05F0001N	41	239	65	7
2016/6/18 13:00	05F0055N	05F0001N	41	250	73	3
2016/6/19 13:00	05F0055N	05F0001N	41	261	107	2

大客車流量與旅行時間



小客車流量與旅行時間





✓ 星期一與星期日的旅行時間較長

TrafficVolumn+CarVolumn+
w7(是否為星期日)

$R^2=0.1843$

調整空值與離群值

✓ 使用三期平均補空值

TrafficVolumn+CarVolumn+w7

R2=0.4006

```
> vif(regN1341.lm)
```

w7	carvolumn	trafficvolumn
1.726783	1.723703	1.018933

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	220.7747	1.5259	144.680	< 2e-16	***
w7	8.0687	1.1113	7.260	9.96e-13	***
carvolumn	0.2297	0.0220	10.440	< 2e-16	***
trafficvolumn	0.5981	0.1559	3.835	0.000136	***

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

Residual standard error: 7.987 on 726 degrees of freedom
Multiple R-squared: 0.4006, Adjusted R-squared: 0.3981
F-statistic: 161.7 on 3 and 726 DF, p-value: < 2.2e-16

預測6/14

$$y = 220.7747 + 8.0687[\text{是否為星期日}] + 0.2297[\text{小客車流量}] + 0.5981[\text{大客車流量}]$$

$x_1=0$, $x_2=68$, $x_3=2$, $y=238$

2018/6/14 13:00	01F0017N	01F0005N	31	45	45
2018/6/14 13:00	05F0055N	05F0001N	31	220	66
2018/6/14 13:00	05F0055N	05F0001N	32	221	25
2018/6/14 13:00	05F0055N	05F0001N	41	237	1
2018/6/14 13:00	05F0055N	05F0001N	42	226	1
2018/6/14 13:00	05F0055N	05F0001N	5	0	0

結果探討

時間	車種	起點	終點	預測值	實際值	殘差	百分比
8	41	05F0055N	05F0001N	221	219	2	0.009%
13	41	05F0055N	05F0001N	235	231	4	0.02%

- ✓ 在第一個模型中，只直接比較調整前與調整後的 R square值，就判斷了沒有調整的資料比較適合預測，並沒有很深入的去了解影響離群 值的變數。