

期中報告

Bike Sharing Demand

報告者：黃琪婕



資料介紹



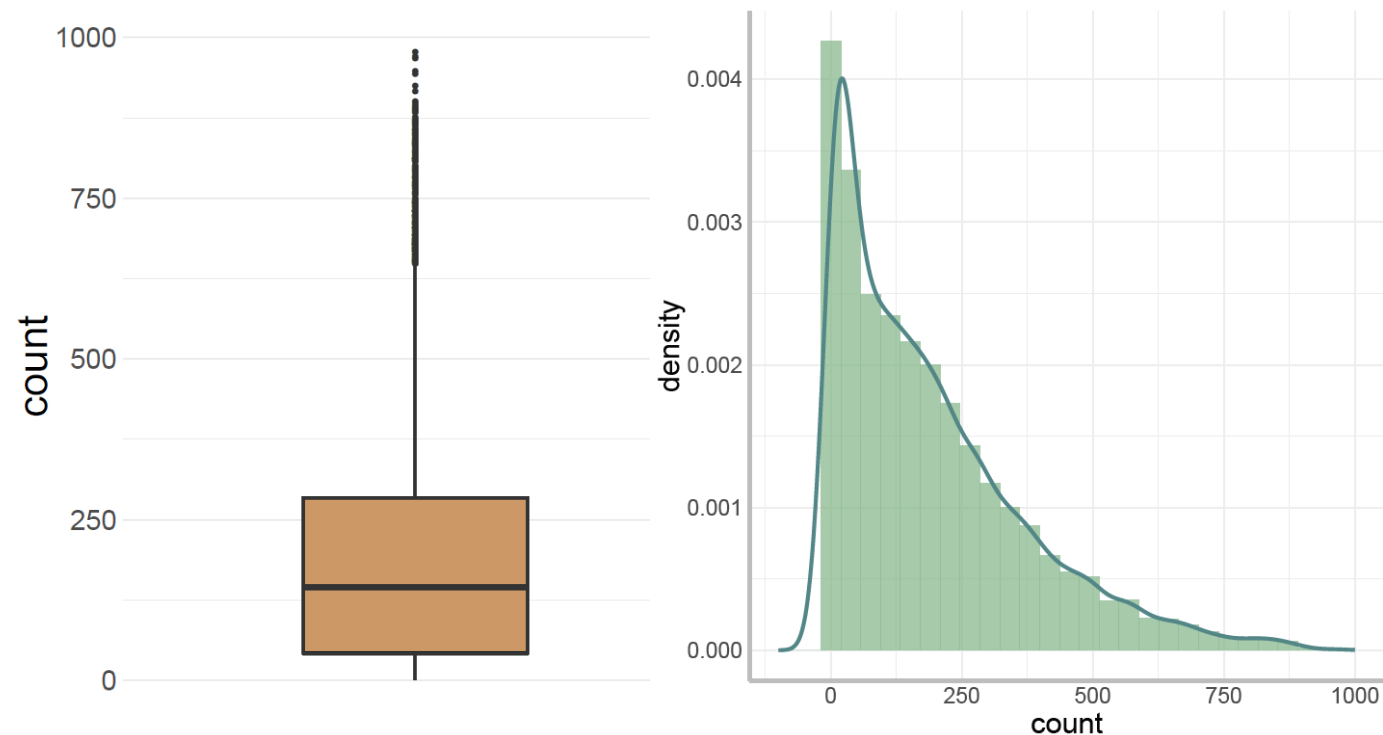
- 資料來源與名稱:Kaggle，Bike Sharing Demand
- 研究問題:
 - 將歷史使用模式與天氣數據結合，以預測華盛頓首都自行車租賃計劃中的自行車租賃需求
- 樣本數: 10886筆，無缺失值
- 12個欄位資料，以count作為應變數y。

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
1	2011-01-01 00:00:00	Spring	0	0	Clear	9.84	14.395	81	0.0000	3	13	16
2	2011-01-01 01:00:00	Spring	0	0	Clear	9.02	13.635	80	0.0000	8	32	40

Y - count

- 總租賃數量
- 連續變數
- 分布右偏
- 極端值處理

mean	sd	0%	25%	50%	75%	100%
191.57	181.14	1	42	145	284	977

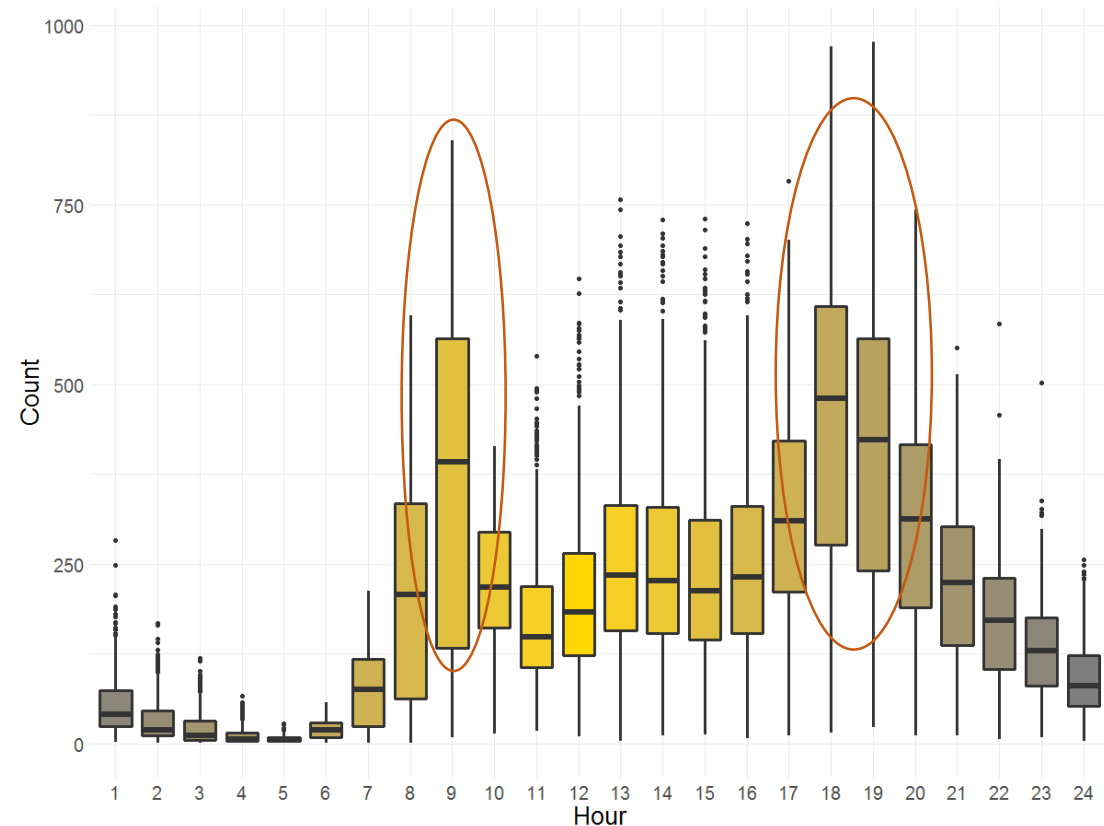


	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
1	2011-01-01 00:00:00	Spring	0	0	Clear	9.84	14.395	81	0.0000	3	13	16
2	2011-01-01 01:00:00	Spring	0	0	Clear	9.02	13.635	80	0.0000	8	32	40

X - datetime

hour & count

- 2011-01-01 00:00:00
 - 2012-12-19 23:00:00
 - 可變為時、日、月、年
 - 最小的預測時間為小時
-
- 早上9時有一個高峰，
 - 下午18~19時，有另一個高峰。

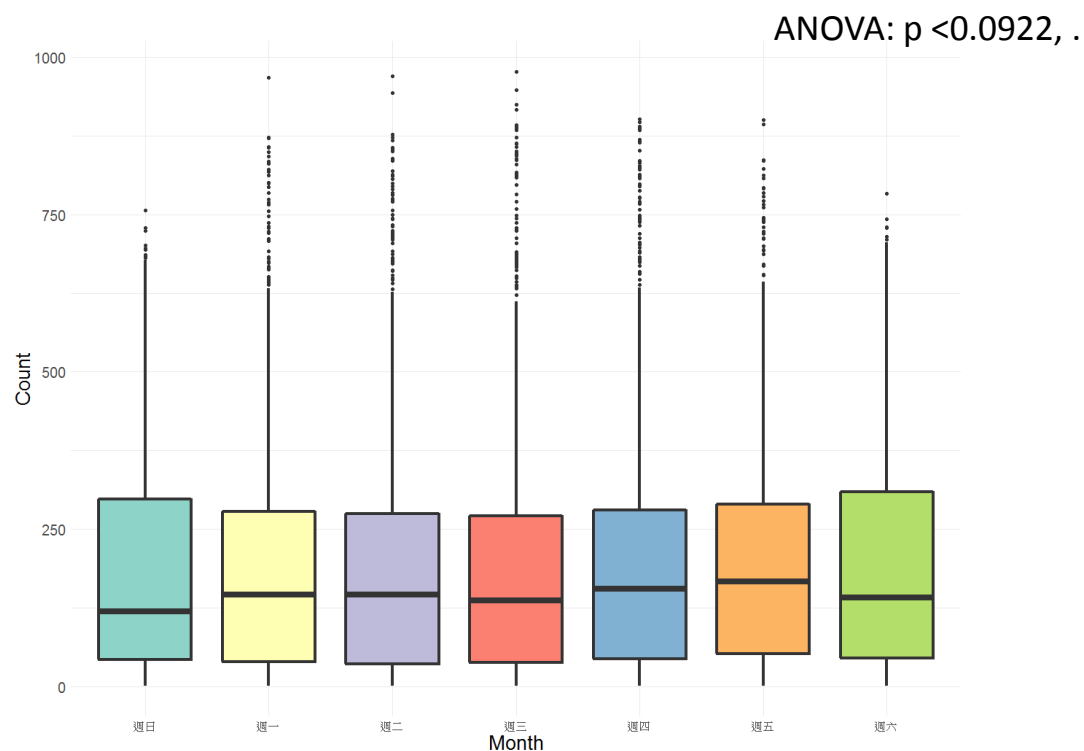


	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
1	2011-01-01 00:00:00	Spring	0	0	Clear	9.84	14.395	81	0.0000	3	13	16
2	2011-01-01 01:00:00	Spring	0	0	Clear	9.02	13.635	80	0.0000	8	32	40

X - datetime

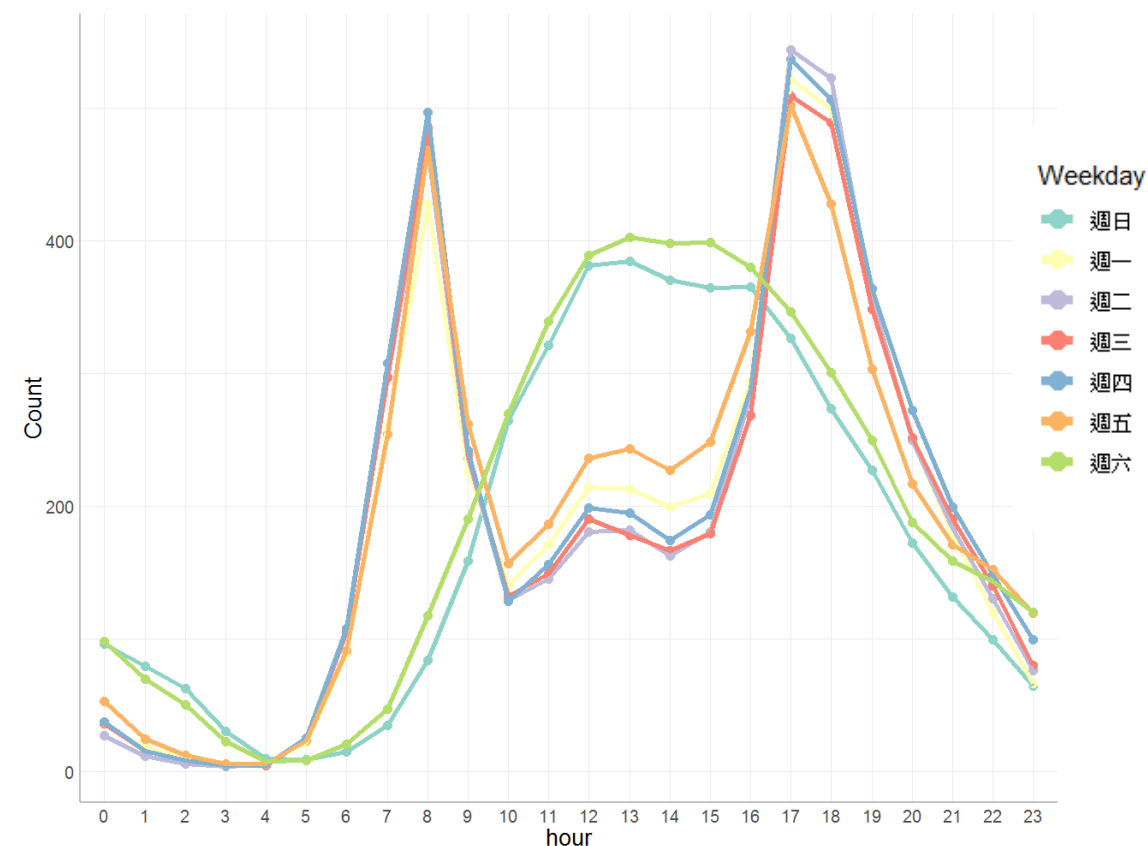
day& count

- 2011-01-01 00:00:00
- 2012-12-19 23:00:00
- 可變為時、日、月、年
- 最小的預測時間為小時
- 單看盒狀圖看不出day之間的差異，也就是週一到周日租賃數量都很平均。



day & hour & count

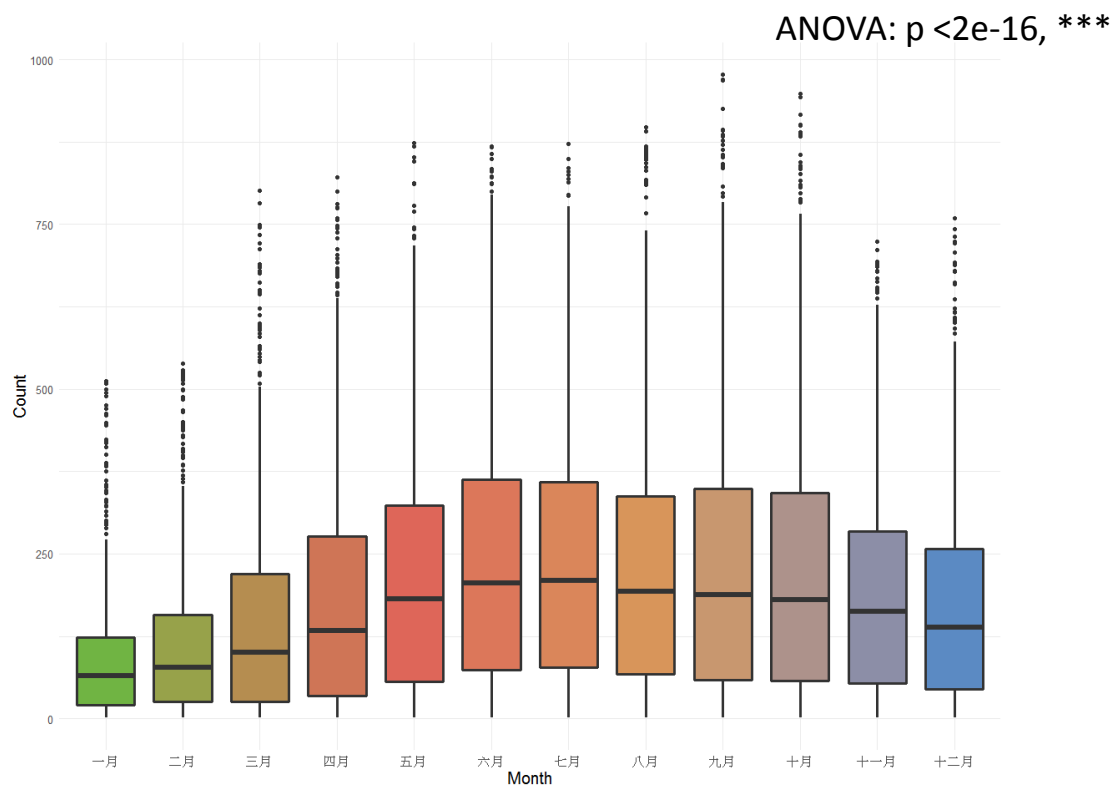
- 日+時，可以看出週六與周日租賃的趨勢與週一至週五不同。
- 週六與周日: 12~16時。
- 週一至週五: 7~8時(上班)與17~18時(下班)。



X - datetime

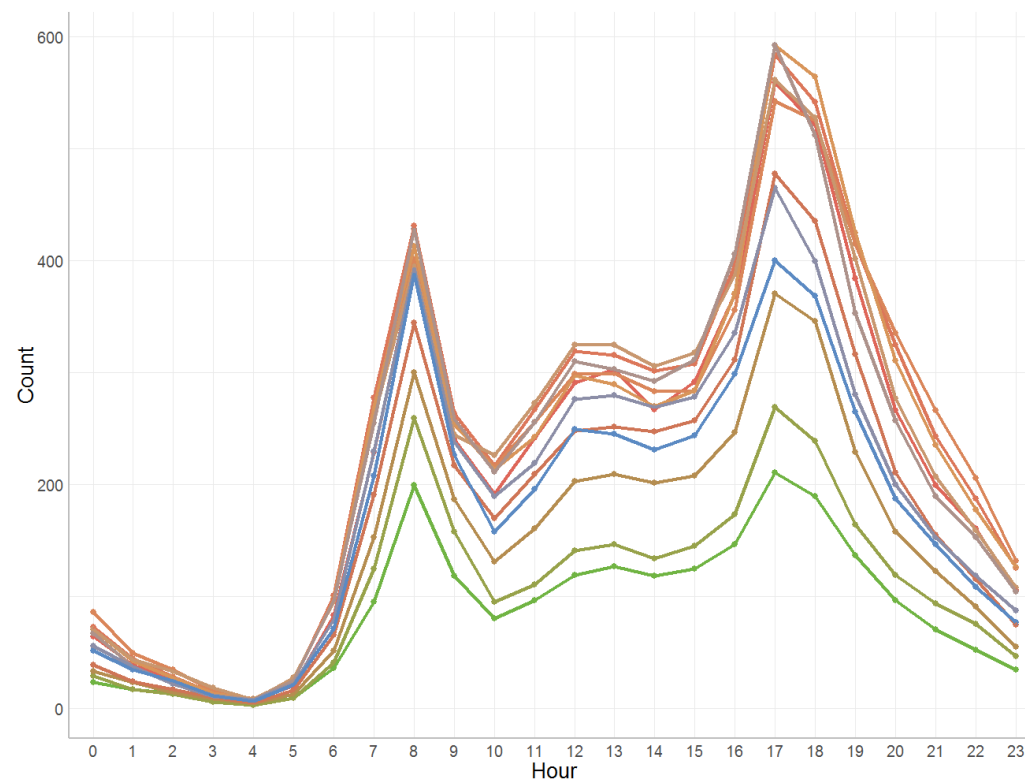
month& count

- 2011-01-01 00:00:00
- 2012-12-19 23:00:00
- 可變為時、日、月、年
- 最小的預測時間為小時
- 夏天的租賃數較多，冬天較少。



month & hour & count

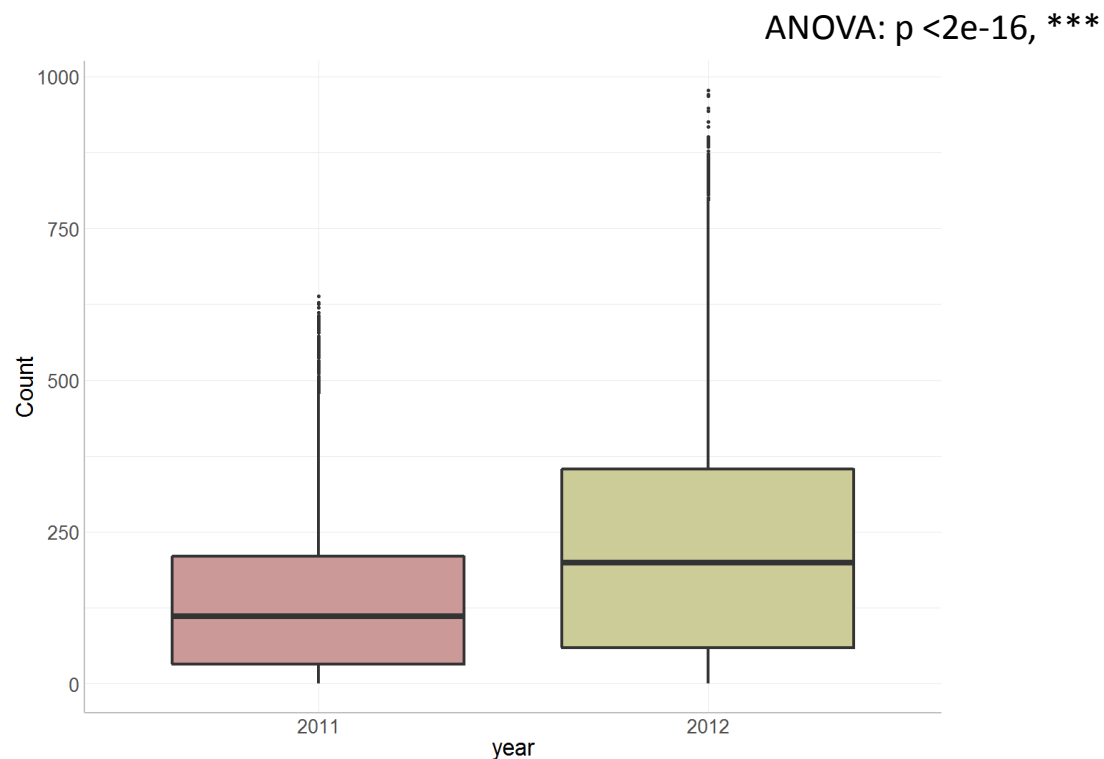
- 每個月的趨勢與hour & count一樣。
- 早上9時有一個高峰，
- 下午18~19時，有另一個高峰。



X - datetime

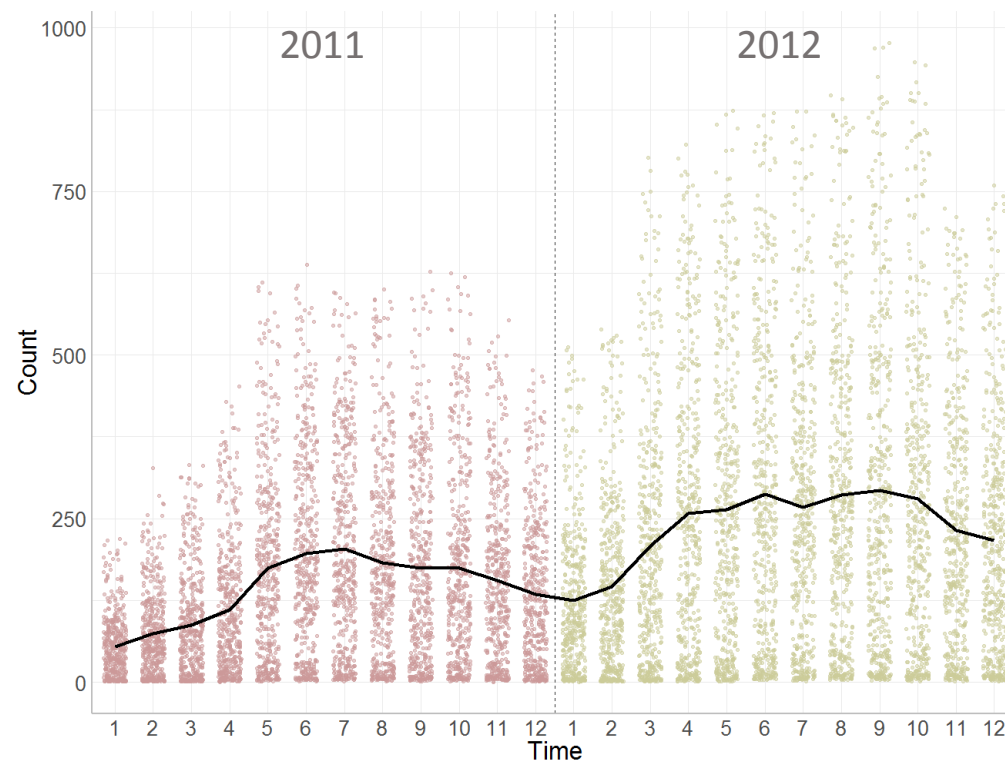
year & count

- 2011-01-01 00:00:00
- 2012-12-19 23:00:00
- 可變為時、日、月、年
- 最小的預測時間為小時
- 2012年租賃數顯著大於2011年。



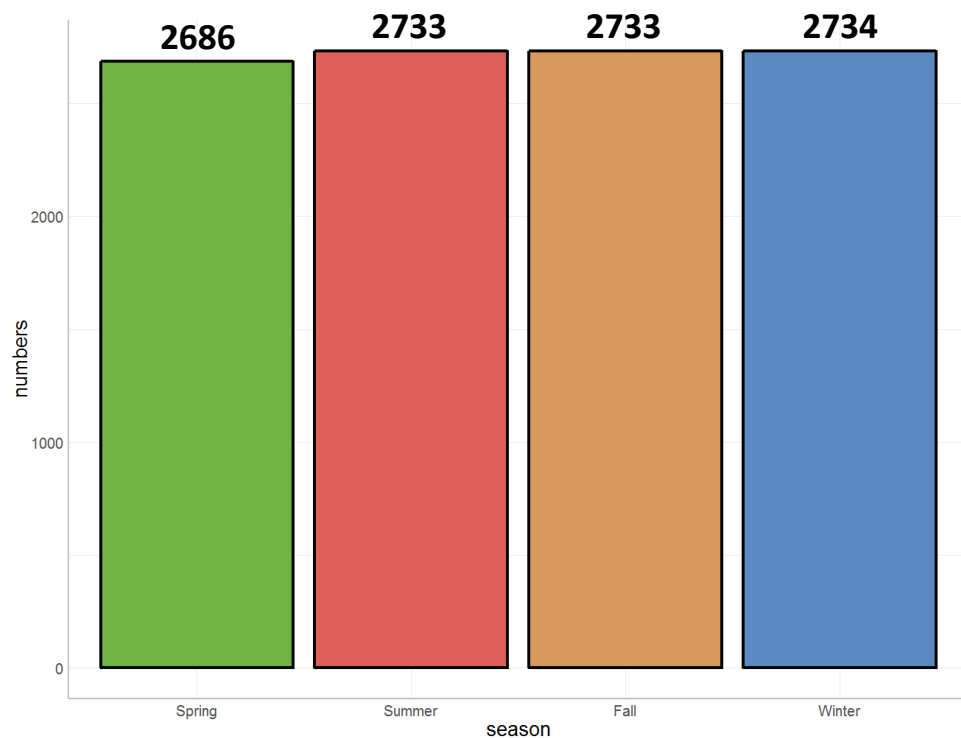
year & month & count

- 2年的月趨勢大致相同，與month& count的結果一樣。
- 皆是夏天的租賃數較多，冬天較少。
- 年趨勢並沒有太大的影響。



X - season

- 季節
- 類別資料
- 分為4類
 - 1: 春天 2: 夏天 3: 秋天 4: 冬天
- 共10886筆，季節分布平均。

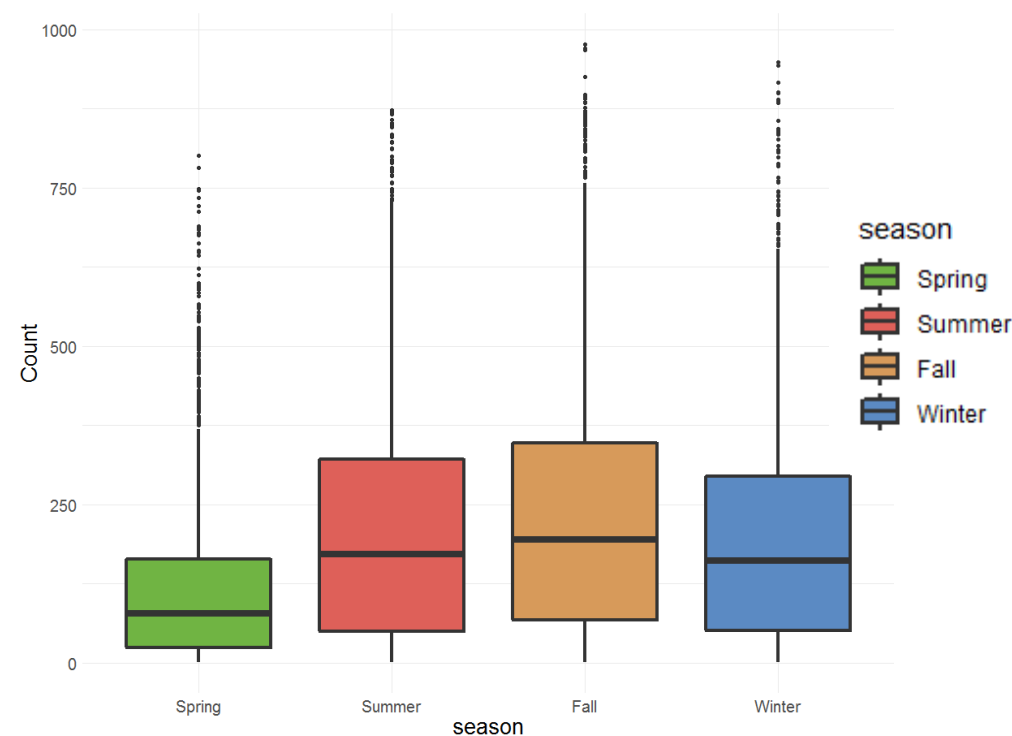


season & count

	mean	sd	0%	25%	50%	75%	100%
Spring	116.3433	125.2740	1	24	78	164	801
Summer	215.2514	192.0078	1	49	172	321	873
Fall	234.4171	197.1510	1	68	195	347	977
Winter	198.9883	177.6224	1	51	161	294	948

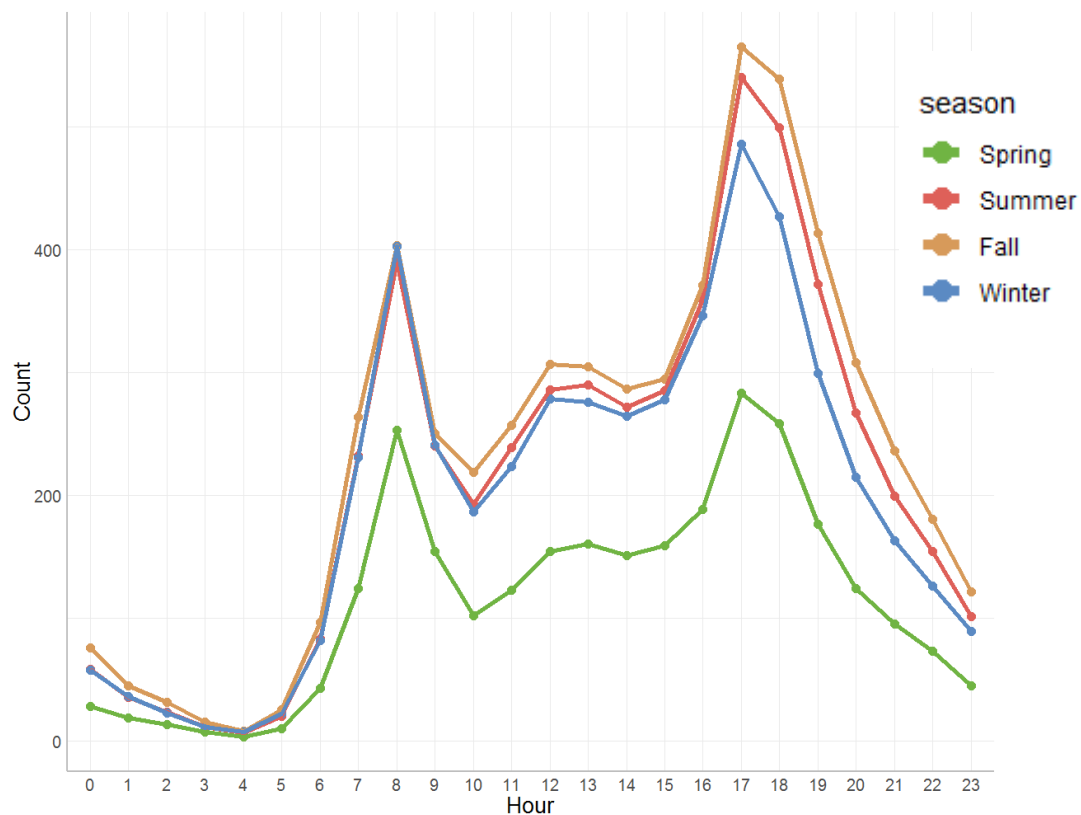
ANOVA: $p < 2e-16$, ***

- 將month & count 縮減為季節。
- 季節的平均租賃量有顯著差異。



season & hour & count

- 趨勢與 month & hour & count 一致。
- 4個季節趨勢一致。
- 春天較少，秋天最多。

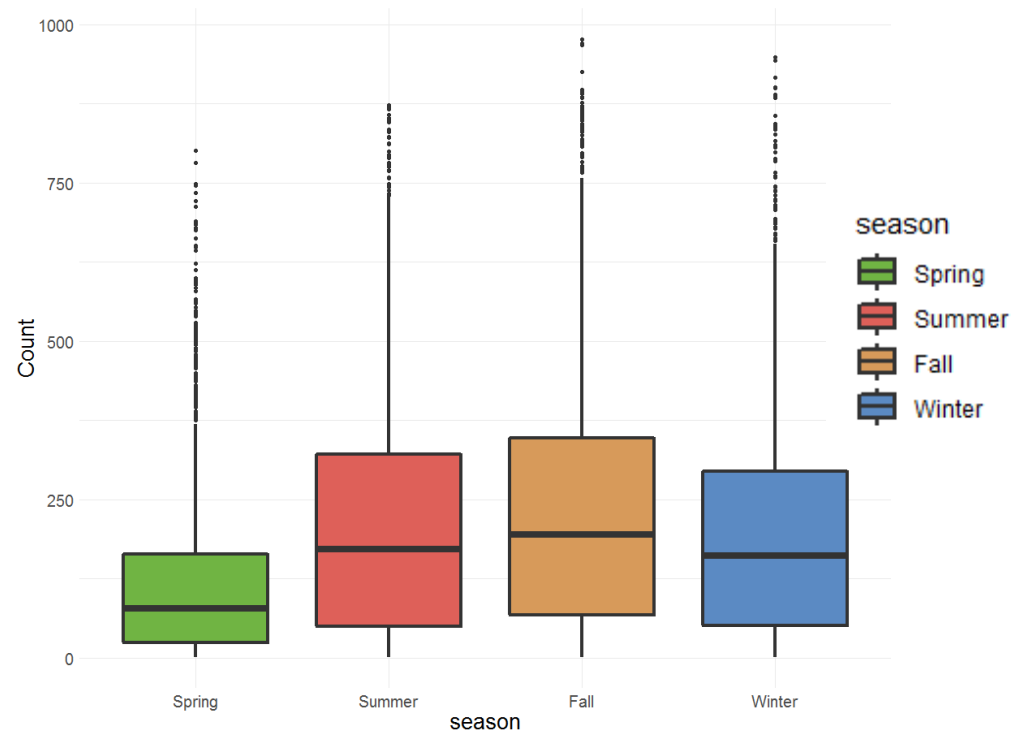


season & count

	mean	sd	0%	25%	50%	75%	100%
Spring	116.3433	125.2740	1	24	78	164	801
Summer	215.2514	192.0078	1	49	172	321	873
Fall	234.4171	197.1510	1	68	195	347	977
Winter	198.9883	177.6224	1	51	161	294	948

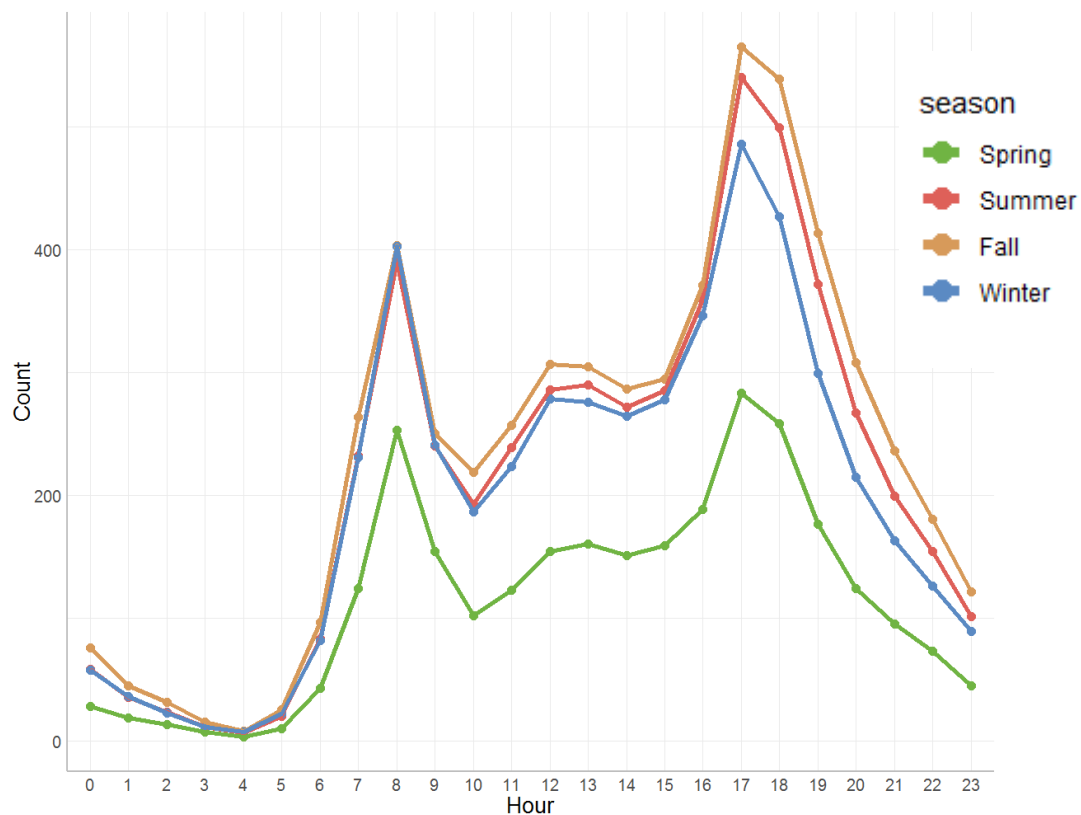
ANOVA: $p < 2e-16$, ***

- 將month& count縮減為季節。
- 季節的平均租賃量有顯著差異。



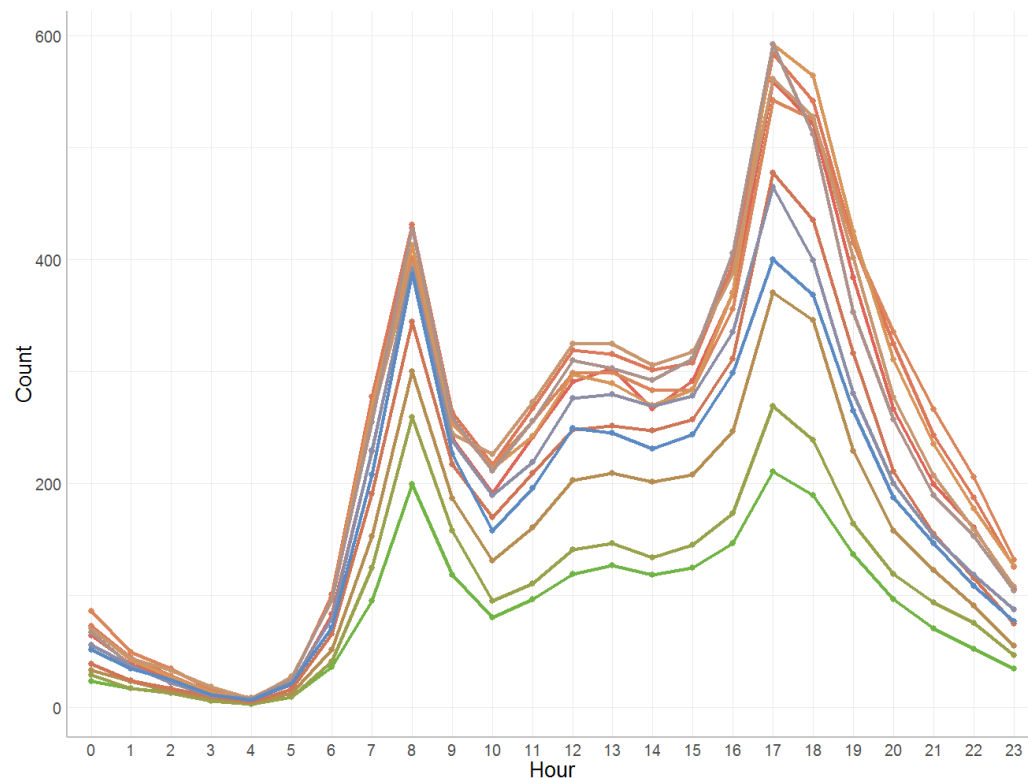
season & hour & count

- 趨勢與 month & hour & count 一致。
- 4個季節趨勢一致。
- 春天較少，秋天最多。



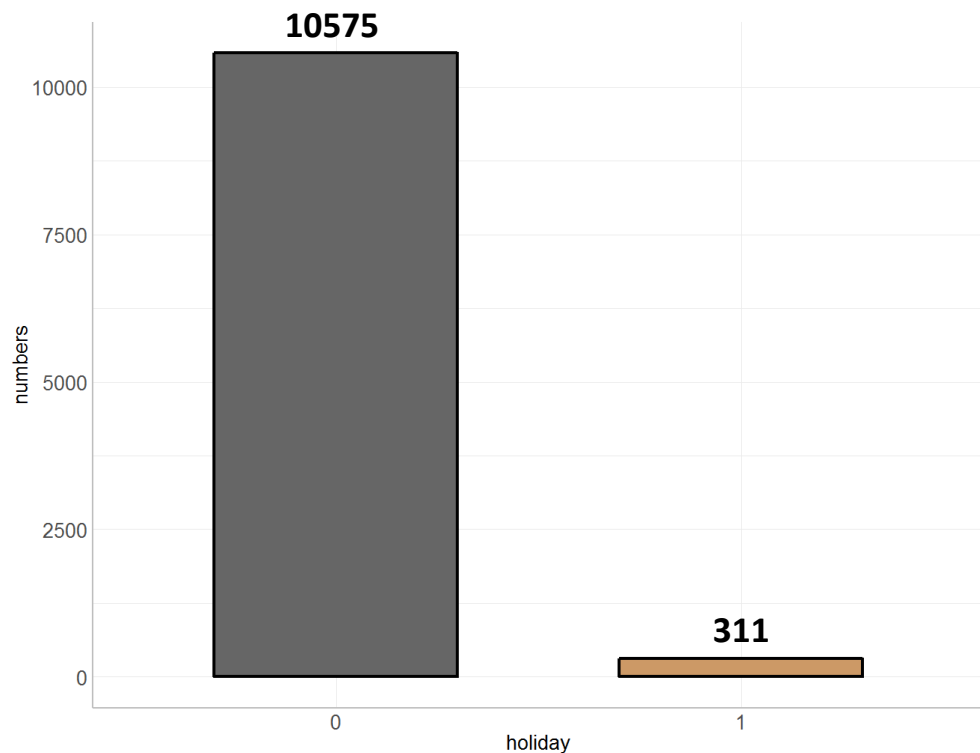
month & hour & count

- 每個月的趨勢與hour & count一樣。
- 早上9時有一個高峰，
- 下午18~19時，有另一個高峰。



X - holiday

- 當天是否是節假日
- 類別資料
- 分為2類
 - 0: 否 1: 是
- 大部分的時間都非節假日。

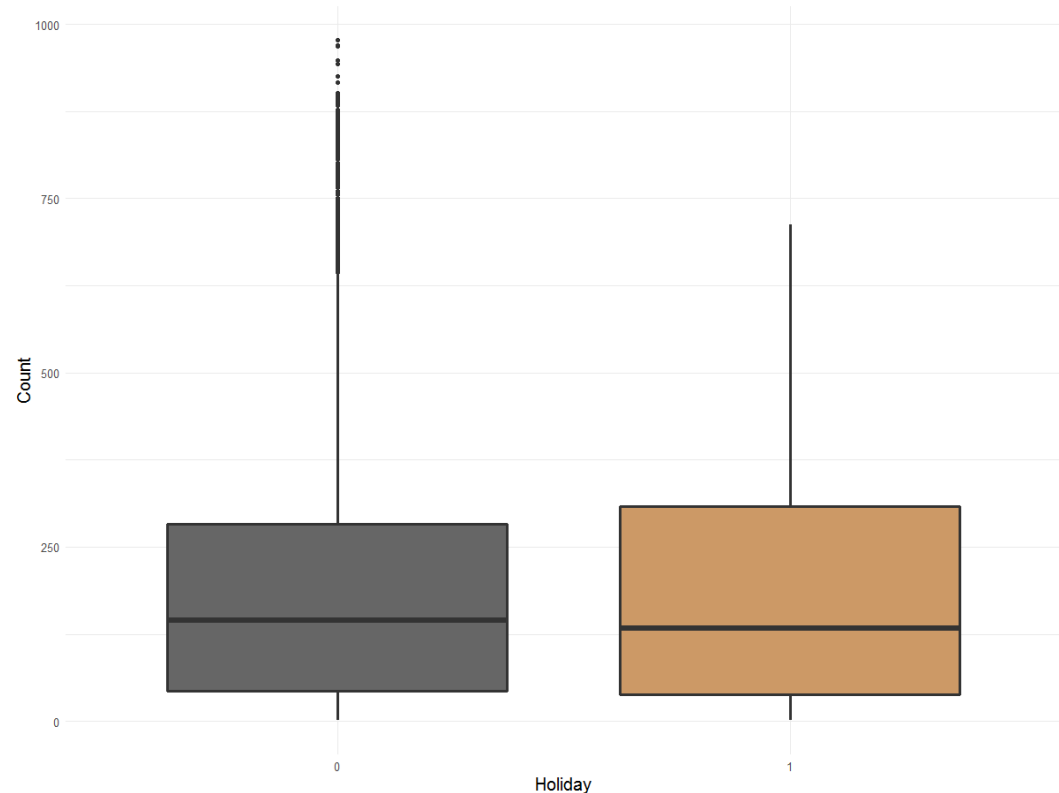


holiday & count

holiday	mean	sd	0%	25%	50%	75%	100%
0	191.74	181.51	1	43.0	145	283	977
1	185.88	168.30	1	38.5	133	308	712

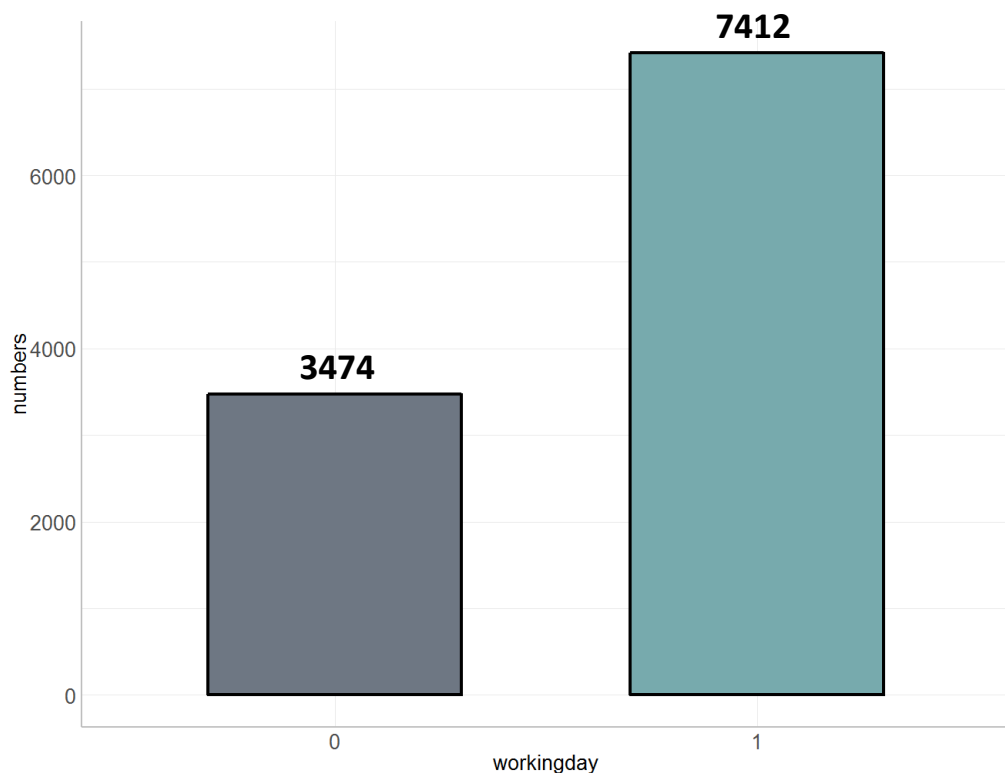
ANOVA: $p < 0.574$

- 是否節假日對平均租賃數量沒有影響。



X - workingday

- 當天是否是工作日
- 類別資料
- 分為2類
 - 0: 否 1: 是
- 工作的時間居多

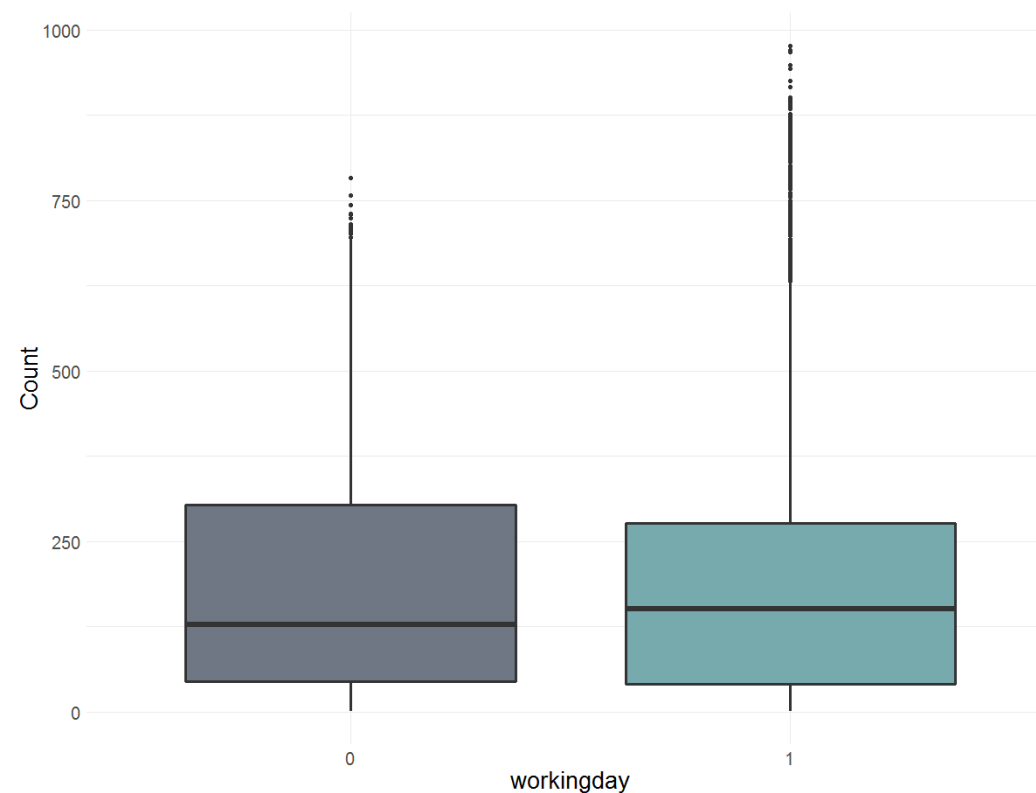


workingday & count

workingday	mean	sd	0%	25%	50%	75%	100%
0	188.51	173.72	1	44	128	304	783
1	193.01	184.51	1	41	151	277	977

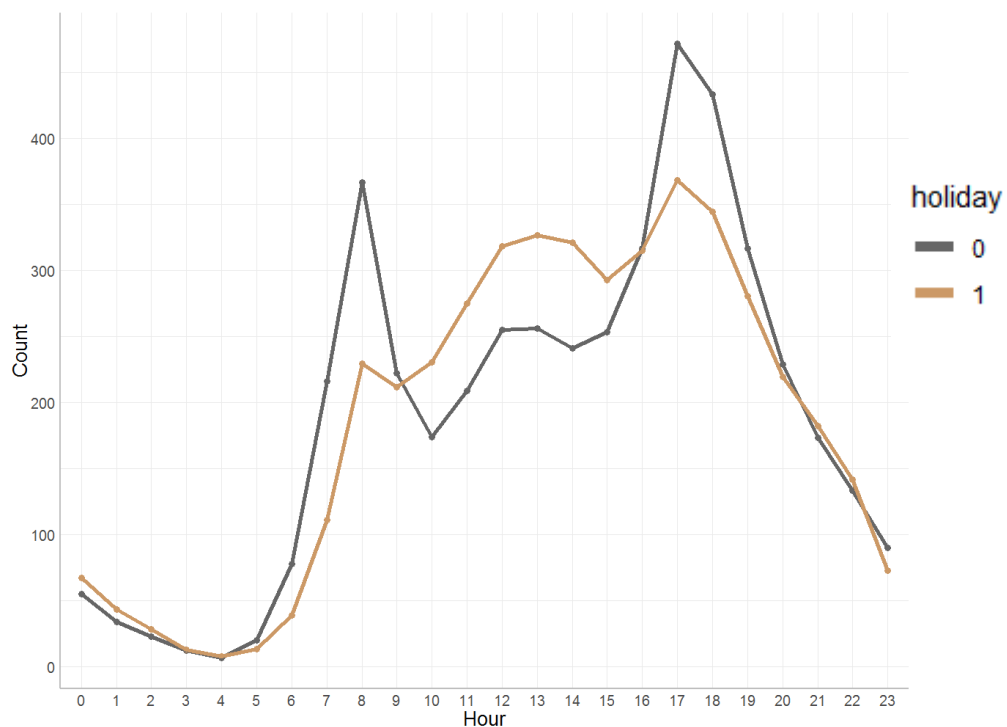
ANOVA: $p < 0.226$

- 是否工作日對平均租賃數量沒有影響。



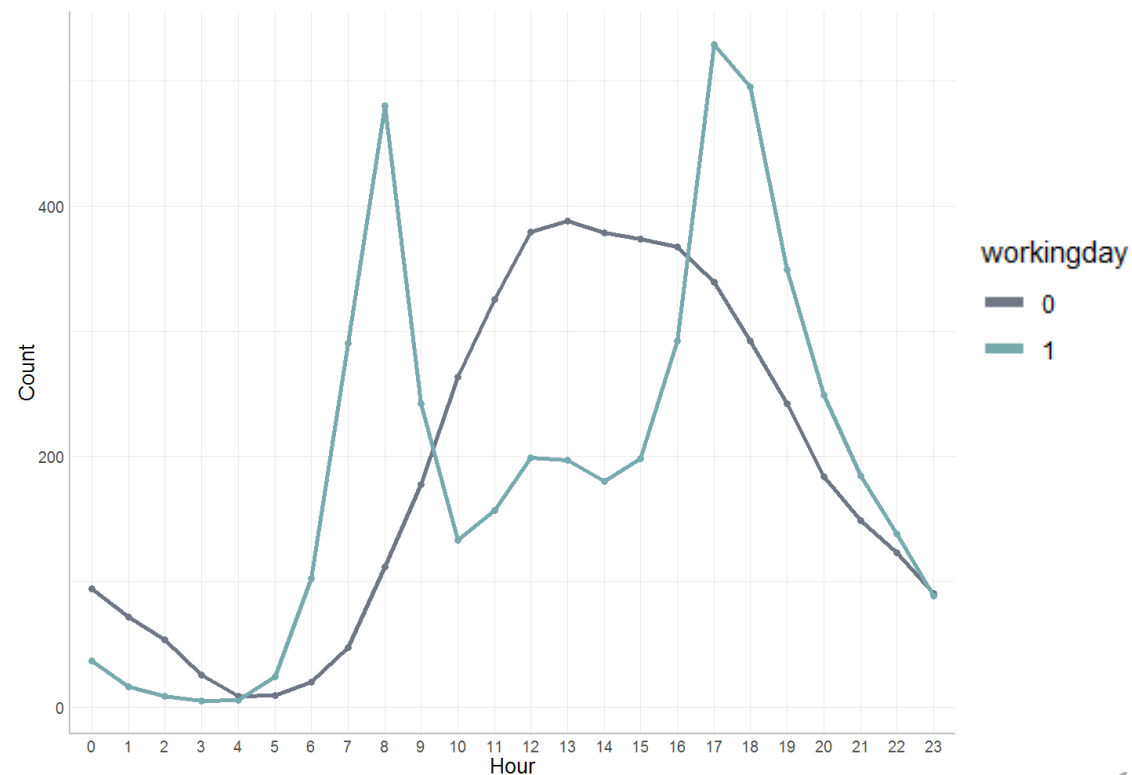
holiday & hour & count

- 是否節假日對於租賃數量好像沒有太大的影響。



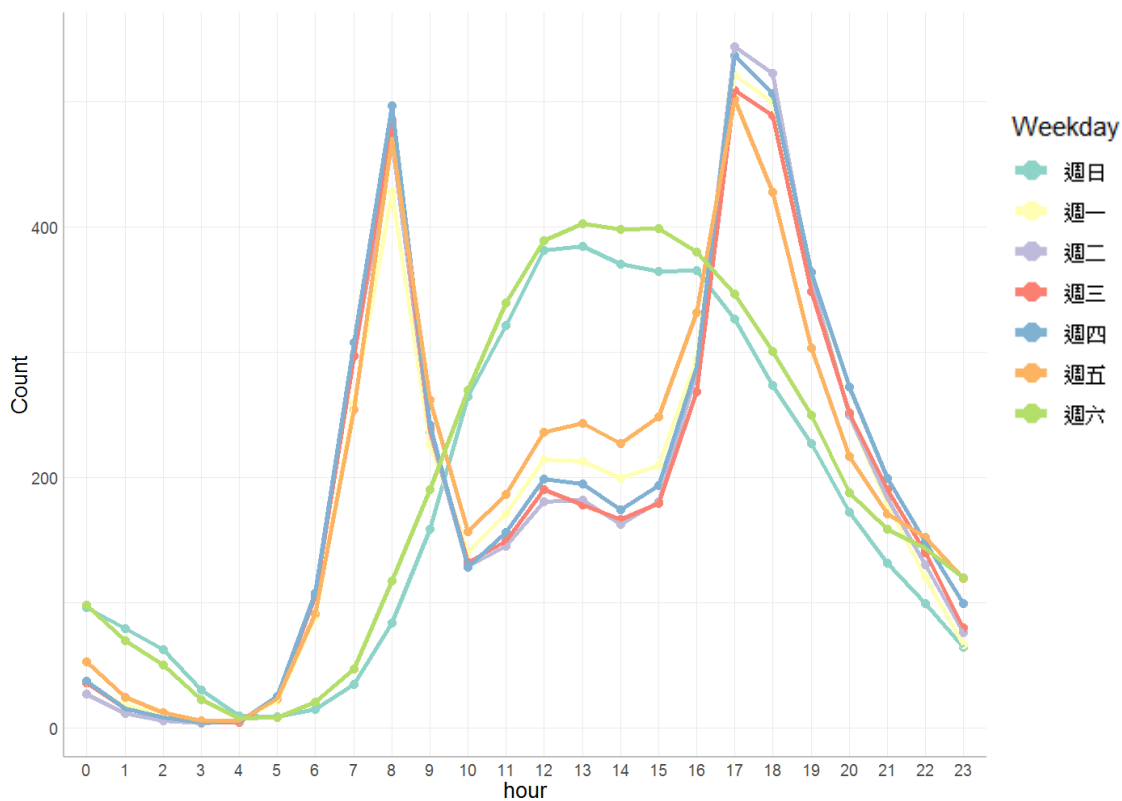
workingday & hour & count

- 趨勢與day & hour & count一致。
- 是否工作日，對於單車每小時的租賃數量趨勢有影響。



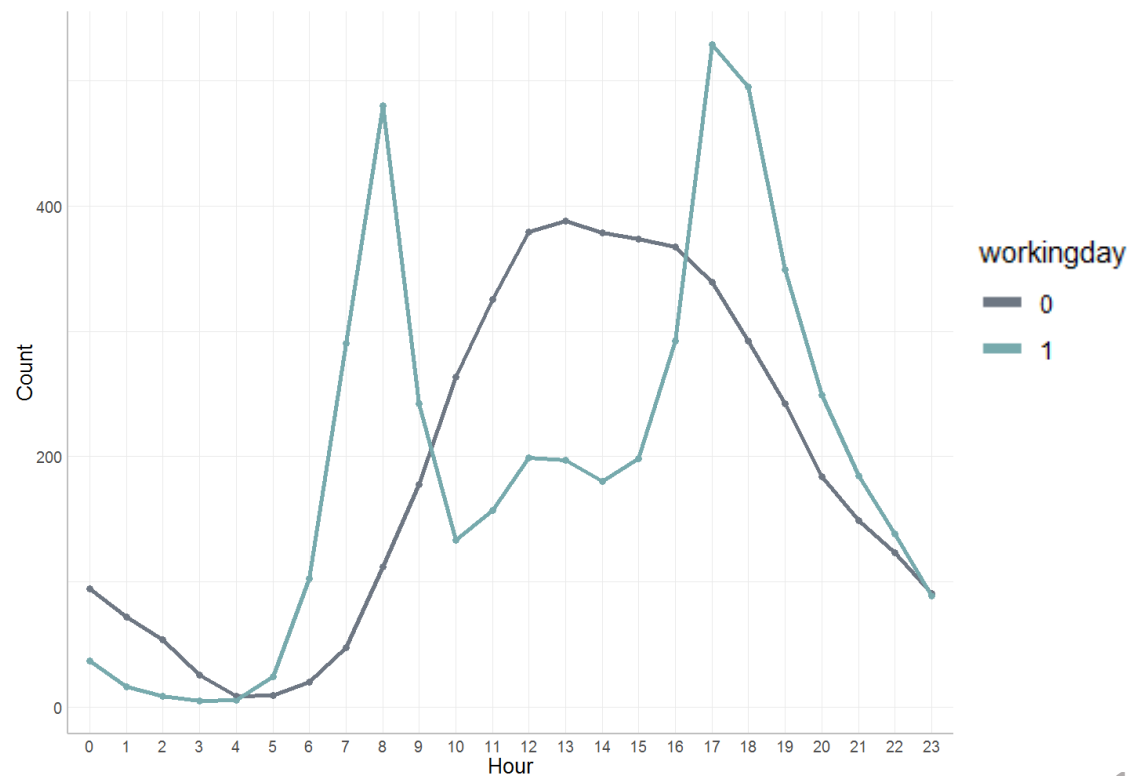
day & hour & count

- 日+時，可以看出週六與周日租賃的趨勢與週一至週五不同。
- 週六與周日: 12~16時。
- 週一至週五: 7~8時(上班)與17~18時(下班)。



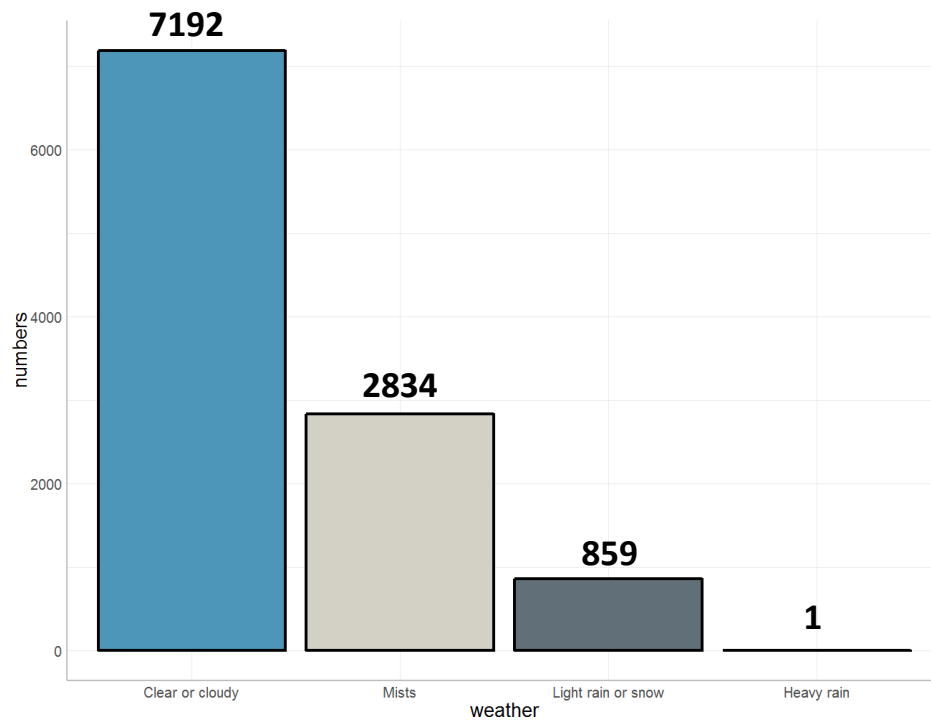
workingday & hour & count

- 趨勢與day & hour & count一致。
- 是否工作日，對於單車每小時的租賃數量趨勢有影響。
- 可取代一週變化。



X - weather

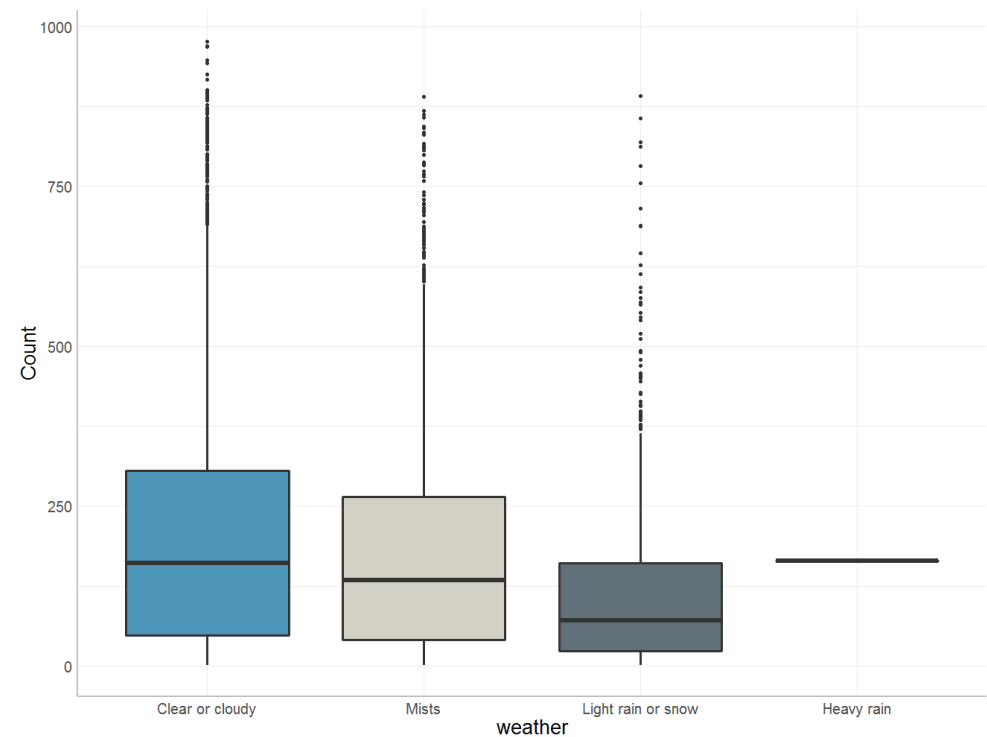
- 天氣狀況
- 類別資料
- 4種分類:
 - 1：晴，少雲，部分多雲，部分多雲
 - 2：薄霧+多雲，薄霧+破碎的雲，霧
 - 3：小雪，小雨+雷雨+散雲，小雨+散雲
 - 4：大雨+冰盤+雷雨+霧，雪+霧



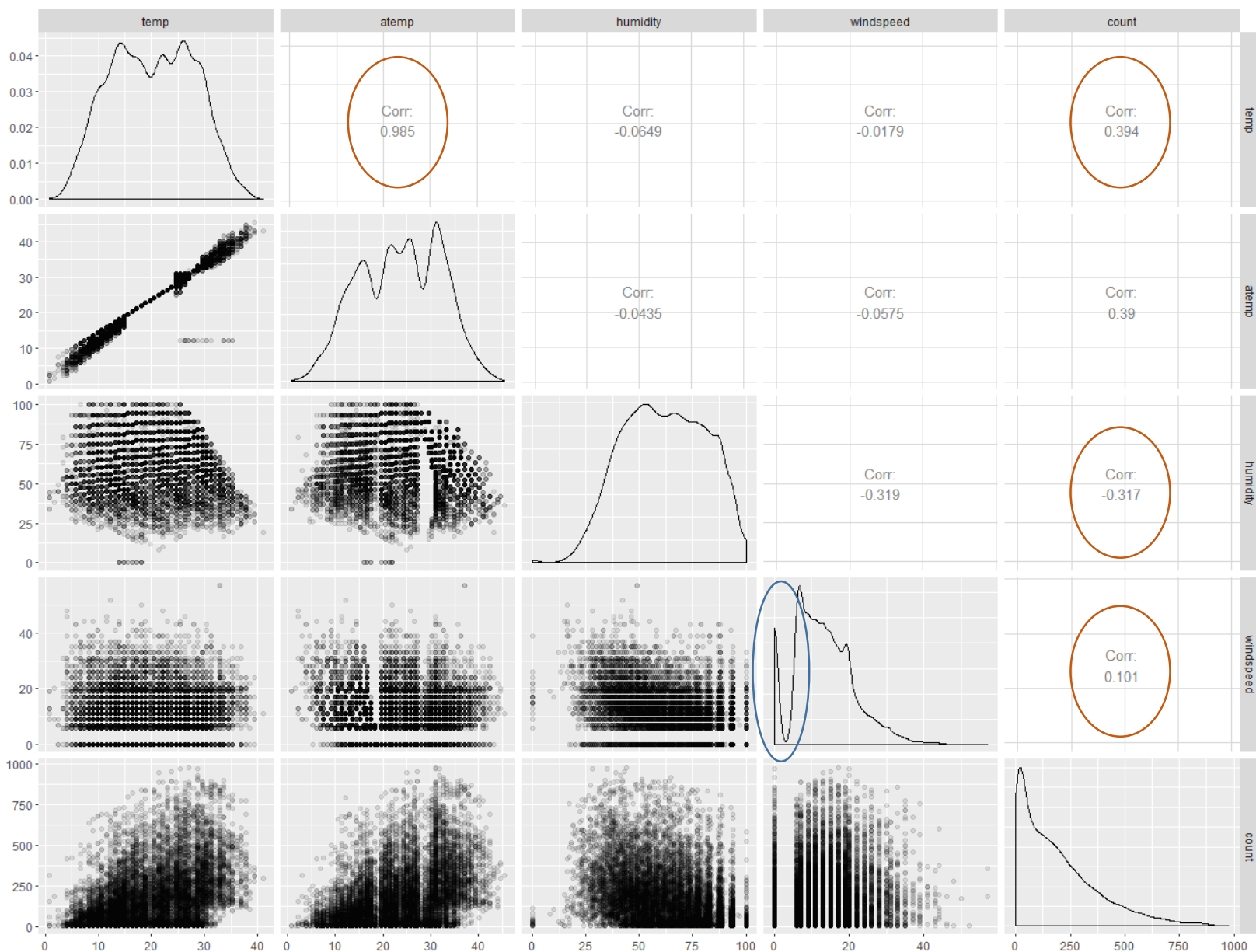
weather & count

weather	mean	sd	0%	25%	50%	75%	100%
Clear or cloudy	205.24	187.96	1	48	161	305	977
Mists	178.95	168.37	1	41	134	264	890
Light rain or snow	118.85	138.58	1	23	71	161	891
Heavy rain	164.00	NaN	164	164	164	164	164

ANOVA: $p < 2e-16$, ***



X - temp, atemp, humidity, windspeed



變數意義

- 皆為連續變數
- temp: 溫度
- atemp: 體感溫度
- humidity: 濕度
- windspeed: 風速

分布意義

- temp, atemp, humidity 分布近常態
- windspeed 的0值過多，推測 null填充。

相關係數

- temp & atemp 高相關，擇一
- temp & count : 正相關
- humidity & count: 負相關
- windspeed & count: 無相關

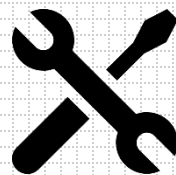
結論與未來工作



結論

- 時間的因素非常重要，尤其是小時為單位。
- 一週的租賃變化可以用是否工作日(workingday)取代。
- 月變化與季節解釋重疊
- Weather是一個很好的解釋變數
- 天氣數據中，選用溫度與濕度，體感溫度與風速刪除。

未來工作



- 目的: 預測每小時租賃數量
- 整理資料的極端值
- 使用Poisson Regression

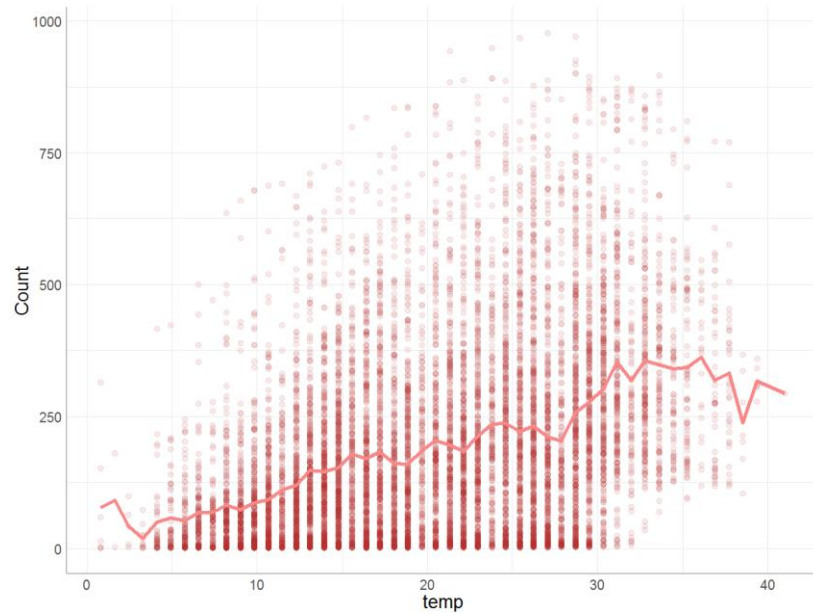
A cluster of three squares in the top-left corner: a dark blue square, a light blue square, and a dark grey square.A cluster of three squares in the bottom-right corner: a dark blue square, a light blue square, and a dark grey square.

THANK YOU!

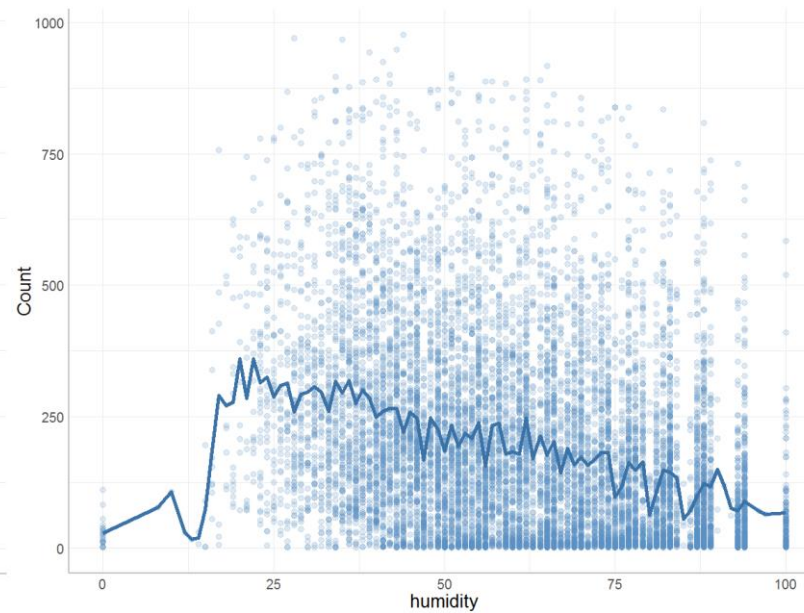
X – temp, humidity, windspeed

- 連續

相關係數: 0.39



相關係數: -0.31



相關係數: 0.10

