Big Data Analysis Final Report – Solar Radiation Prediction

隊名：茶水隊

組員：黃冠瑋、胡恩愷、周擎

Outline

- Introduction
- Literature Reviews and Questions
- Exploratory Data Analysis
- Modeling
- Conclusion

Introduction

- 再生能源的重要性

- 太陽能
  - 取之不盡
  - 無汙染的特性
- 太陽能發電系統
  - 主要受太陽輻射照度的影響
  - 能量分散、密度低、易受環境與天文因素影響

Motivation

- 台灣目前最主要是以風力及火力發電，風力發電充滿不穩定性，而火力發電則是對地球生態造成嚴重汙染

- 希望藉由找到變數與熱輻射的關係，預測未來在哪些地點可以大力推廣太陽能發電，而有效的設立電廠，一方面可以吸收最多的太陽能量，另一方面可以節省成本，達到雙贏的局面。

Literature Reviews

- 行政院原子能委員會《太陽能預測模式與分析技術研究》
  - 分析WRF模式預測未來6至48 小時各種天氣要素
  - 包含溫度、濕度、氣壓、風速、風向以及短波輻射量等

- 結論出幾項結果
  1. 不同地區與日期有顯著的不同
  2. 不同地區與時間有顯著的不同
  3. 在直接日照的數據分析分面，以南部地區在八月份的中午時刻最強。

Question

- 台灣哪些地方會比較適合興建太陽能發電廠？
- 溫度、壓力、濕度、風速如何影響輻射？
- 在這些變數中，我們考慮的先後順序為何？

Question

- 解決方法
  - 從資料選取幾項我們認為影響較深的變數
  - 線性迴歸模型去配適夏威夷的變數以及輻射結果
  - 檢定
  - 台灣的資料做適度的等比例轉換
  - 套入模型

Data

- 夏威夷太陽輻射資料　　資料來源:Kaggle
  - 時間：2016/9/1~2016/12/31
  - 變數：輻射、溫度、壓力、濕度、風向、風速、日出、日落

```
> head(solar)
    UNIXTime                  Data     Time Radiation Temperature Pressure Humidity
1 1475229326 9/29/2016 12:00:00 AM 23:55:26      1.21          48    30.46       59
2 1475229023 9/29/2016 12:00:00 AM 23:50:23      1.21          48    30.46       58
3 1475228726 9/29/2016 12:00:00 AM 23:45:26      1.23          48    30.46       57
4 1475228421 9/29/2016 12:00:00 AM 23:40:21      1.21          48    30.46       60
5 1475228124 9/29/2016 12:00:00 AM 23:35:24      1.17          48    30.46       62
6 1475227824 9/29/2016 12:00:00 AM 23:30:24      1.21          48    30.46       64
  WindDirection.Degrees. Speed TimeSunRise TimeSunSet
1                 177.39  5.62    06:13:00   18:13:00
2                 176.78  3.37    06:13:00   18:13:00
3                 158.75  3.37    06:13:00   18:13:00
4                 137.71  3.37    06:13:00   18:13:00
5                 104.95  5.62    06:13:00   18:13:00
6                 120.20  5.62    06:13:00   18:13:00
```

Data Variable

- 原始資料

	A	B	C	D	E	F	G	H	I	J	K
1	UNIXTime	Data	Time	Radiation	Temperature	Pressure	Humidity	WindDirection	Speed	TimeSunRise	TimeSunSet
2	1475229326	9/29/2016	23:55:26	1.21	48	30.46	59	177.39	5.62	06:13:00	18:13:00
3	1475229023	9/29/2016	23:50:23	1.21	48	30.46	58	176.78	3.37	06:13:00	18:13:00
4	1475228726	9/29/2016	23:45:26	1.23	48	30.46	57	158.75	3.37	06:13:00	18:13:00
5	1475228421	9/29/2016	23:40:21	1.21	48	30.46	60	137.71	3.37	06:13:00	18:13:00

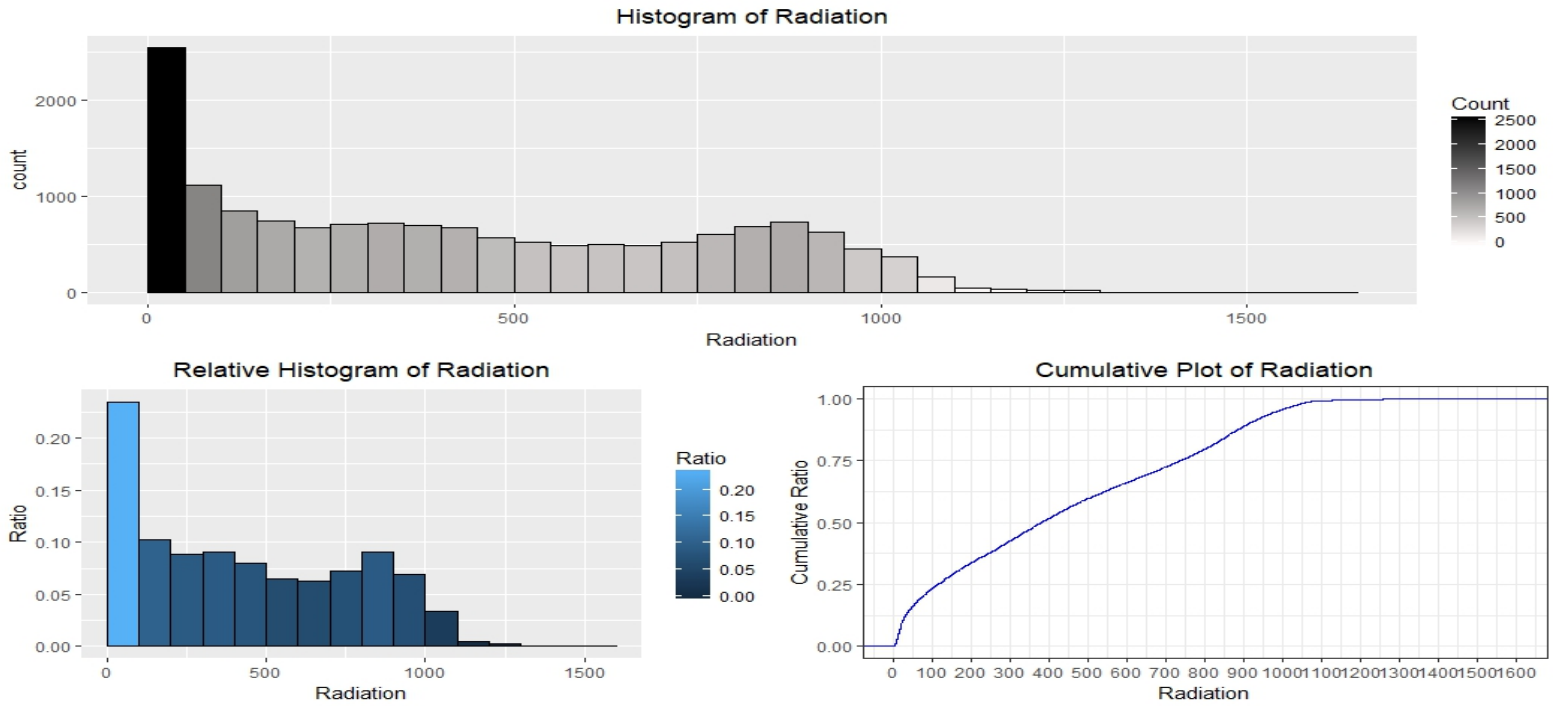
- 處理過後的資料

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	UNIXTime	DayOrNight	Time	Radiation	Temperature	Pressure	Humidity	WindDirection	Speed	TimeSunRise	TimeSunSet	SunshineTime	Month	Day	Year
2	1475229326	0	2355	1.21	48	30.46	59	177.39	5.62	613	1813	12	9	29	2016
3	1475229023	0	2350	1.21	48	30.46	58	176.78	3.37	613	1813	12	9	29	2016
4	1475228726	0	2345	1.23	48	30.46	57	158.75	3.37	613	1813	12	9	29	2016
5	1475228421	0	2340	1.21	48	30.46	60	137.71	3.37	613	1813	12	9	29	2016

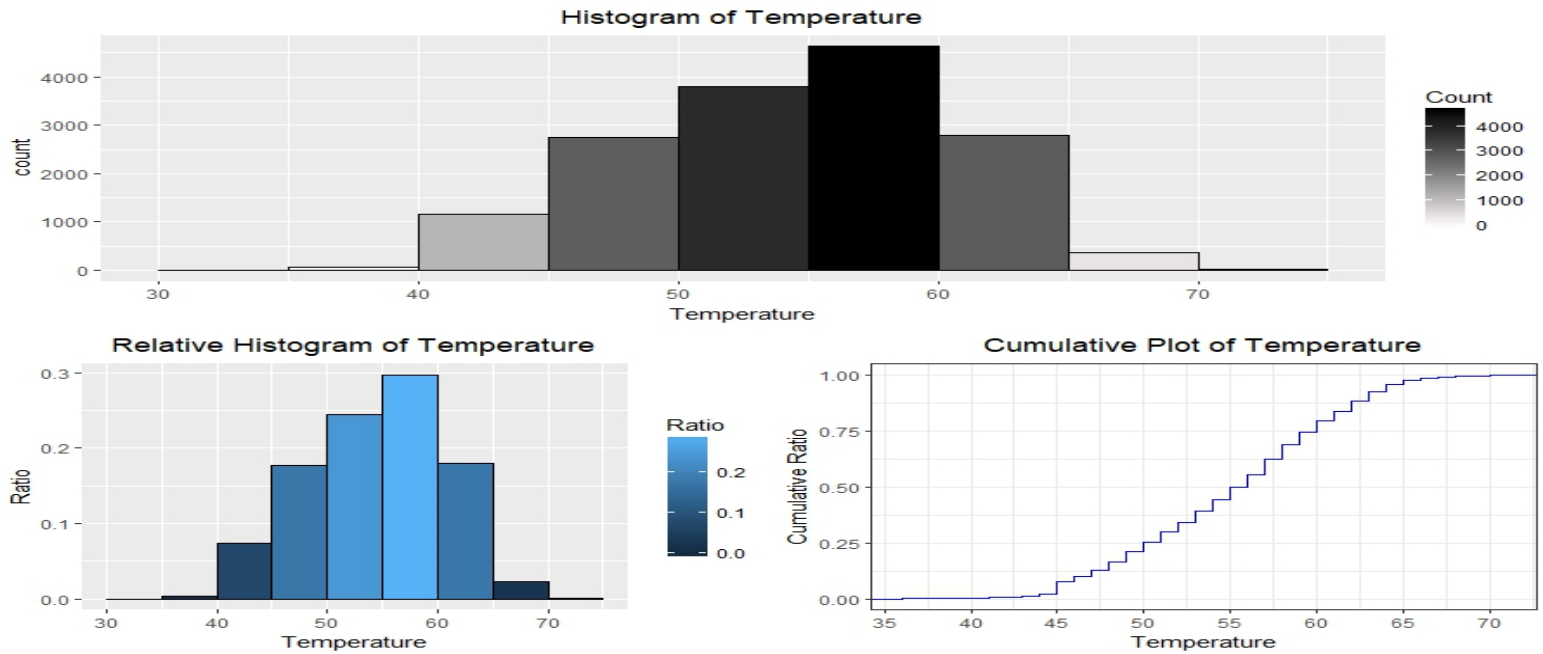
Summary

SUMMARY						
統計量\變數	Radiation	Temperature	Pressure	Humidity	Speed	SunshineTime
Min.	1.11	34	30.19	8	0	10.93
1st Qu.	1.23	46	30.4	56	3.37	11.03
Median	2.66	50	30.43	85	5.62	11.4
Mean	207.12	51.1	30.42	75.02	6.244	11.51
3rd Qu.	354.24	55	30.46	97	7.87	11.92
Max.	1601.26	71	30.56	103	40.5	12.52

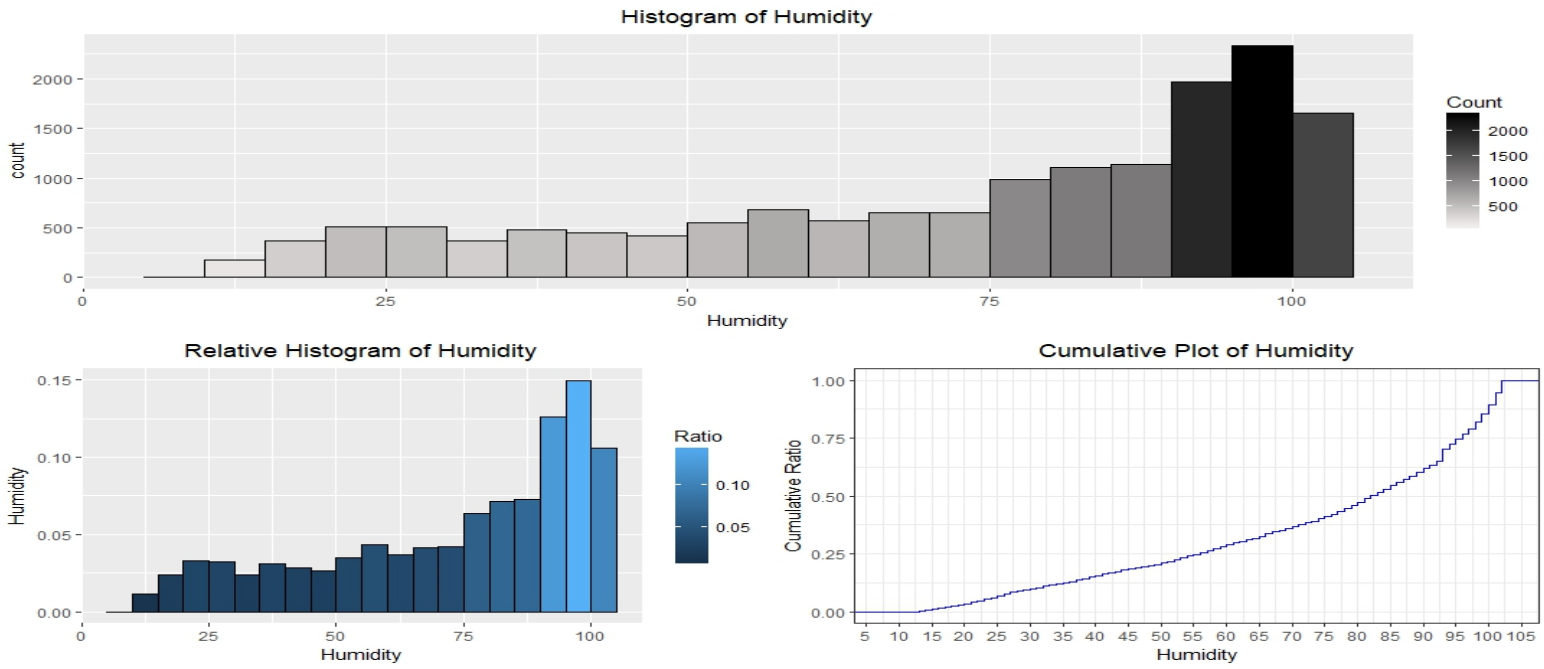
太陽輻射量(Radiation)



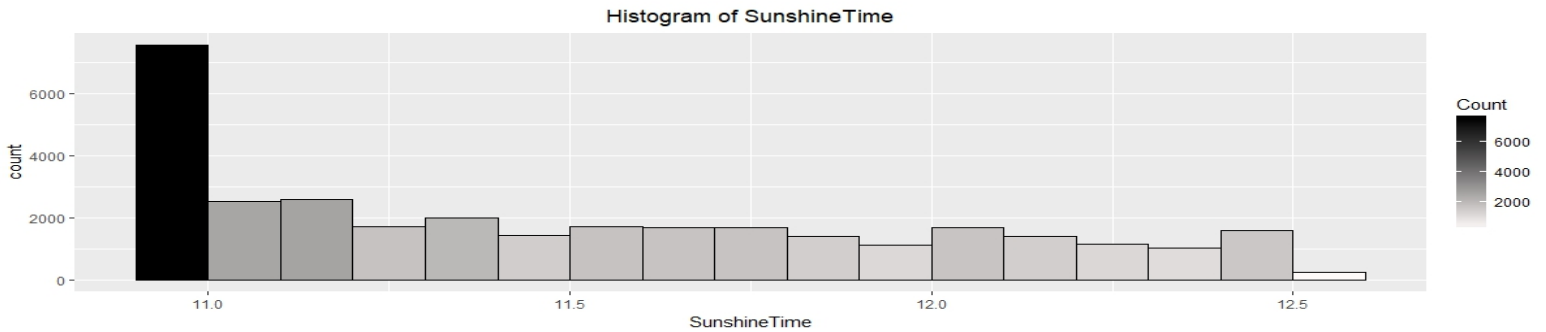
溫度(Temperature)

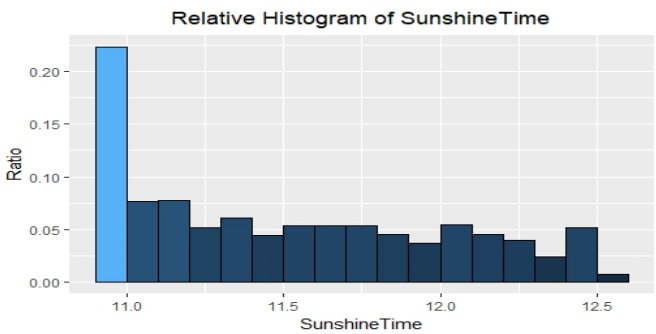


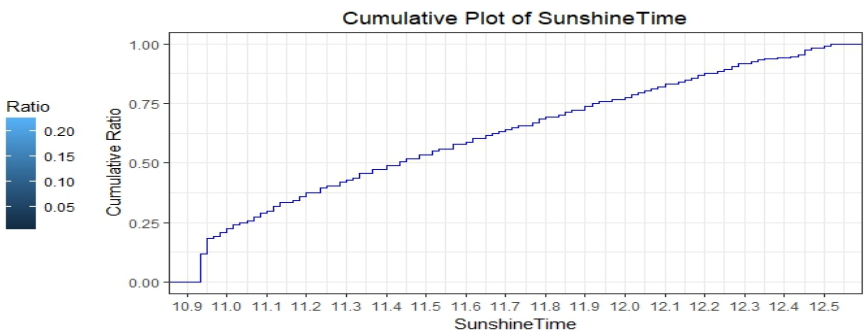
相對濕度(Humidity)



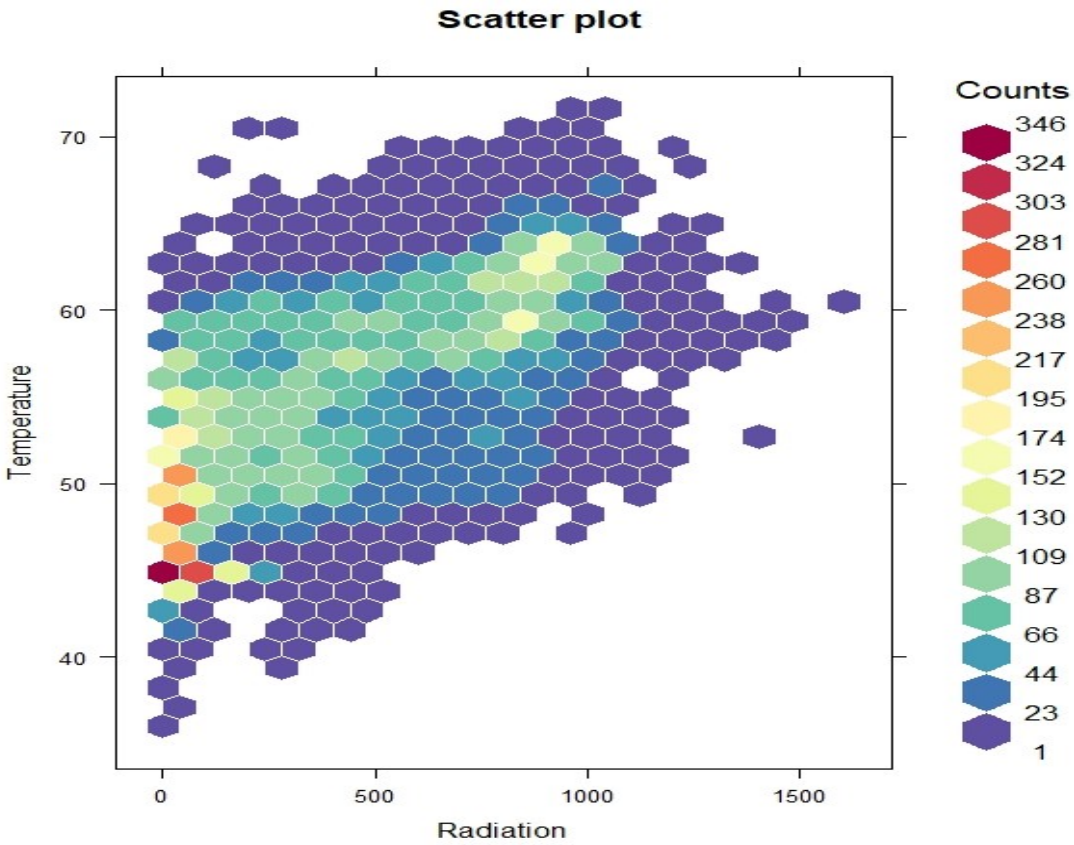
日照時間(SunShineTime)



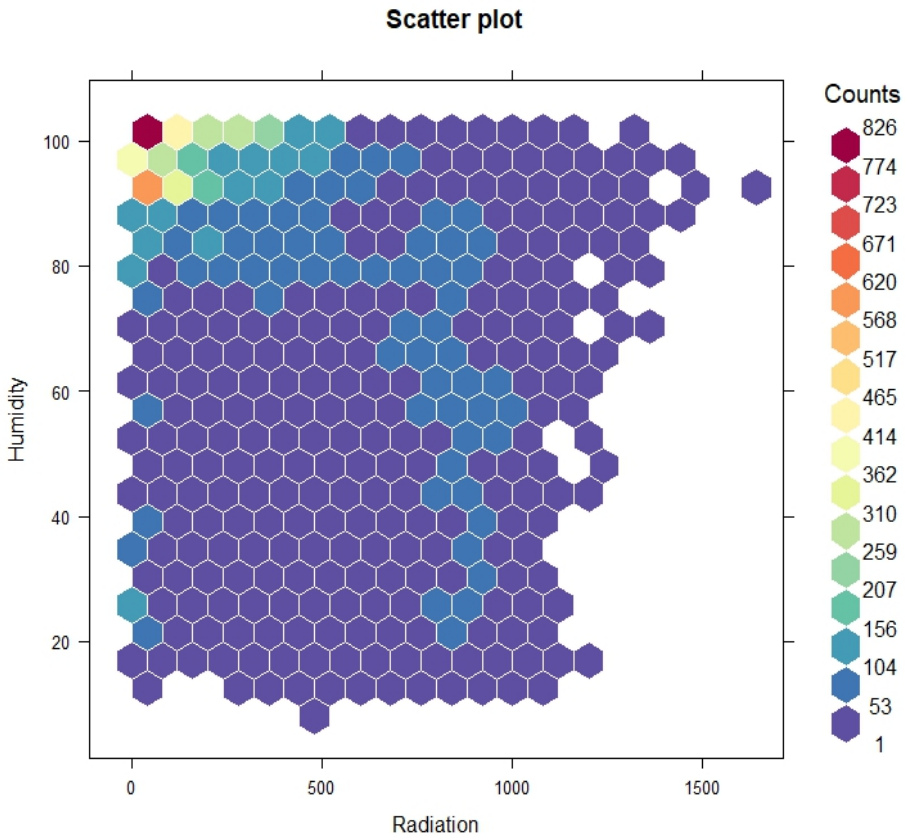




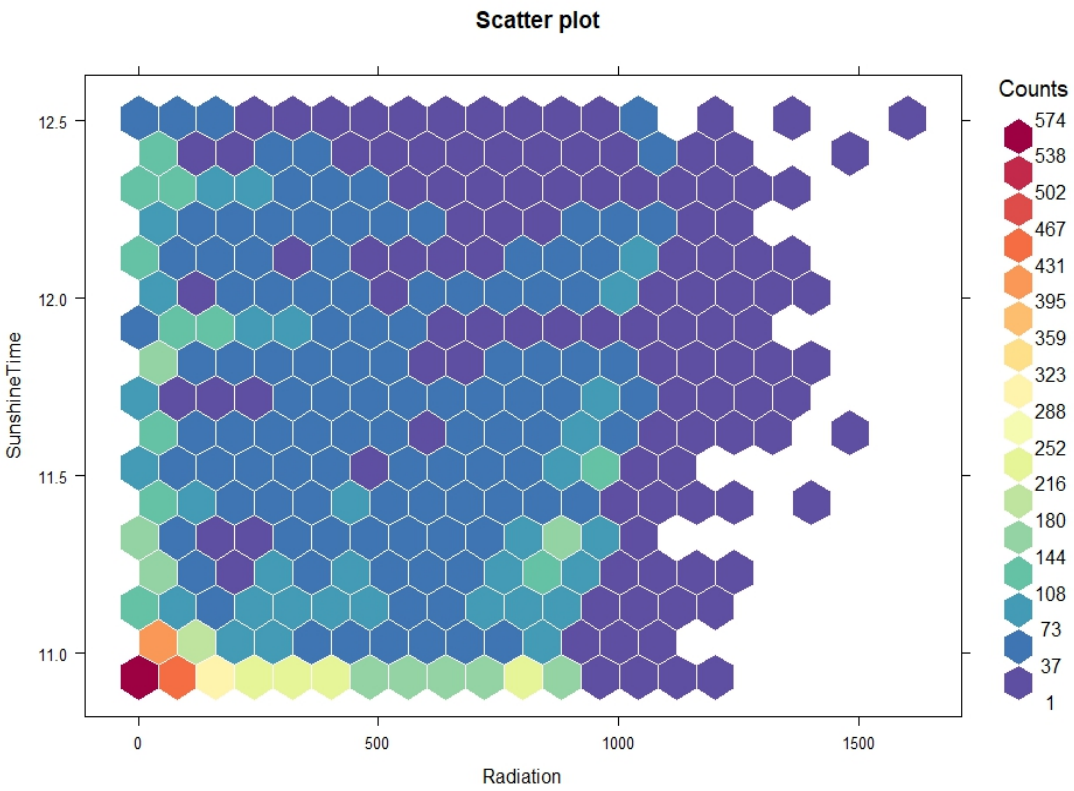
Radiation VS Temperature



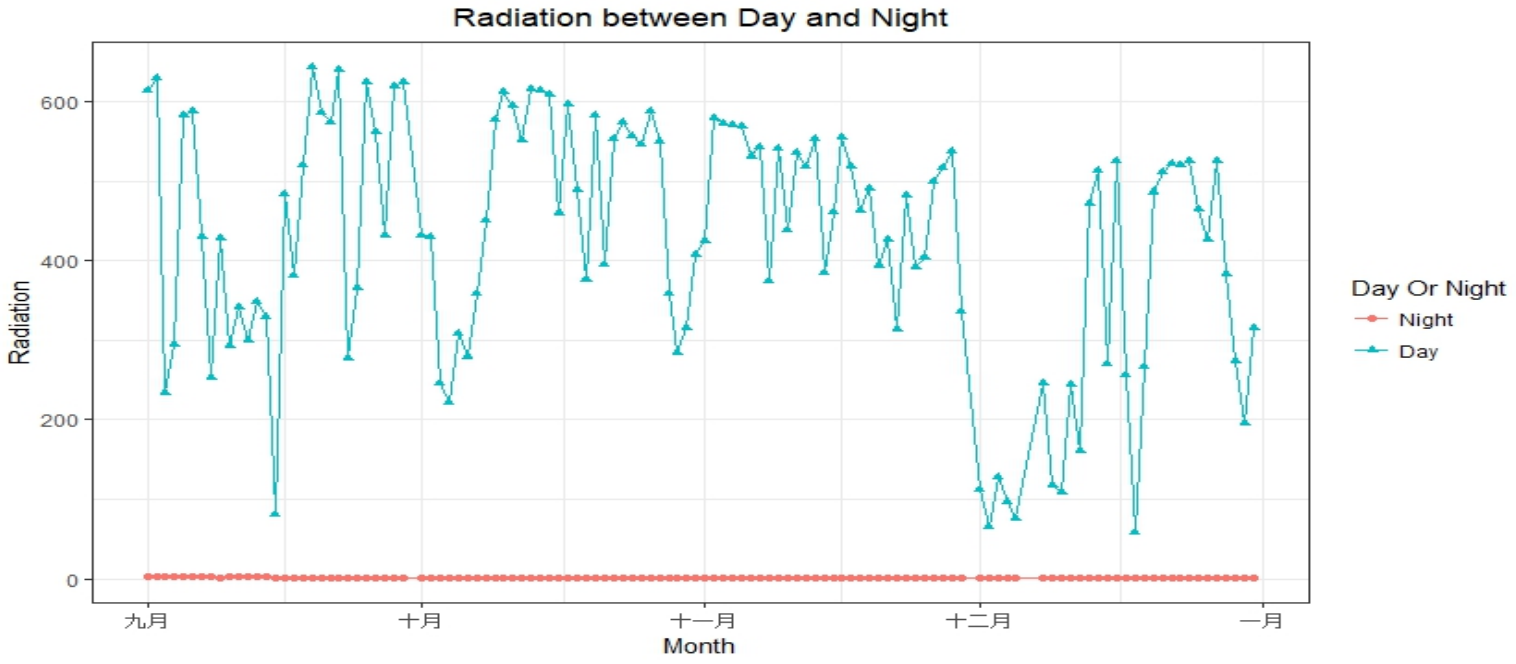
Radiation VS Humidity



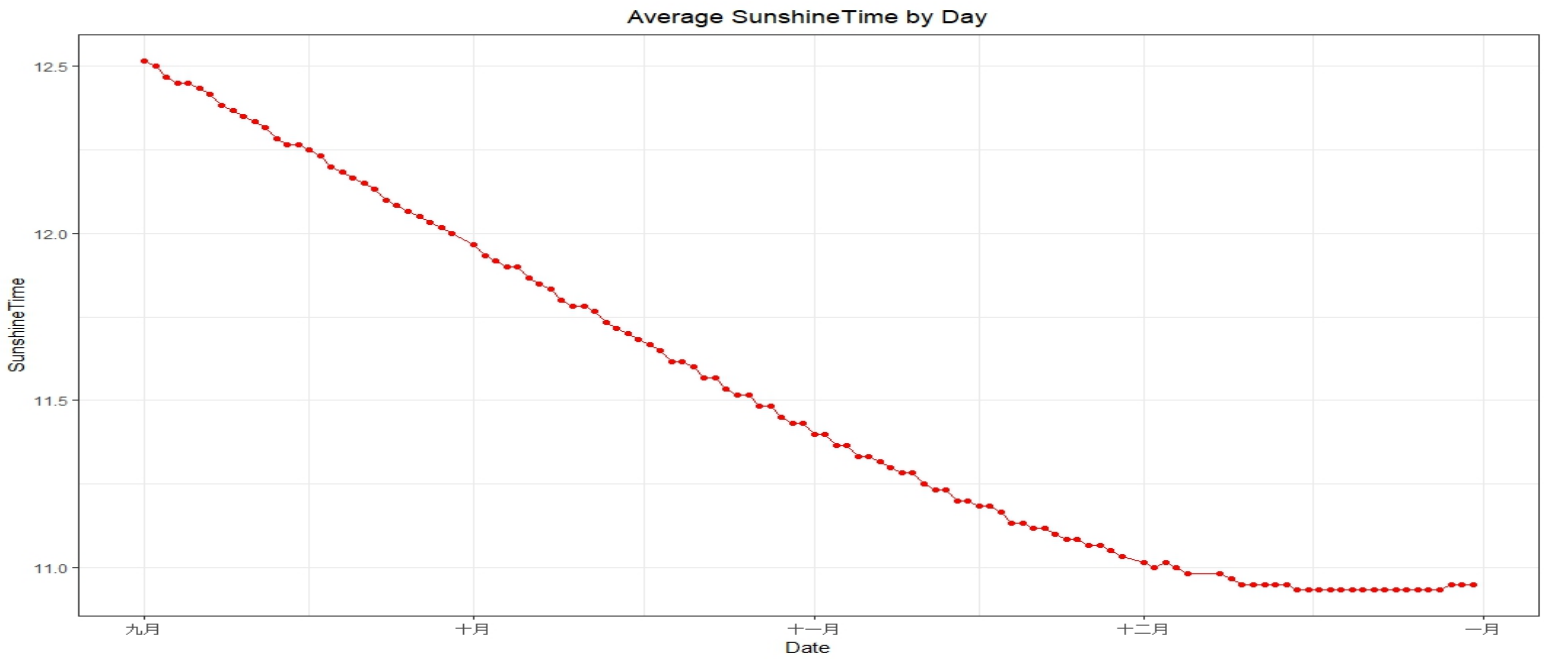
Radiation VS SunshineTime



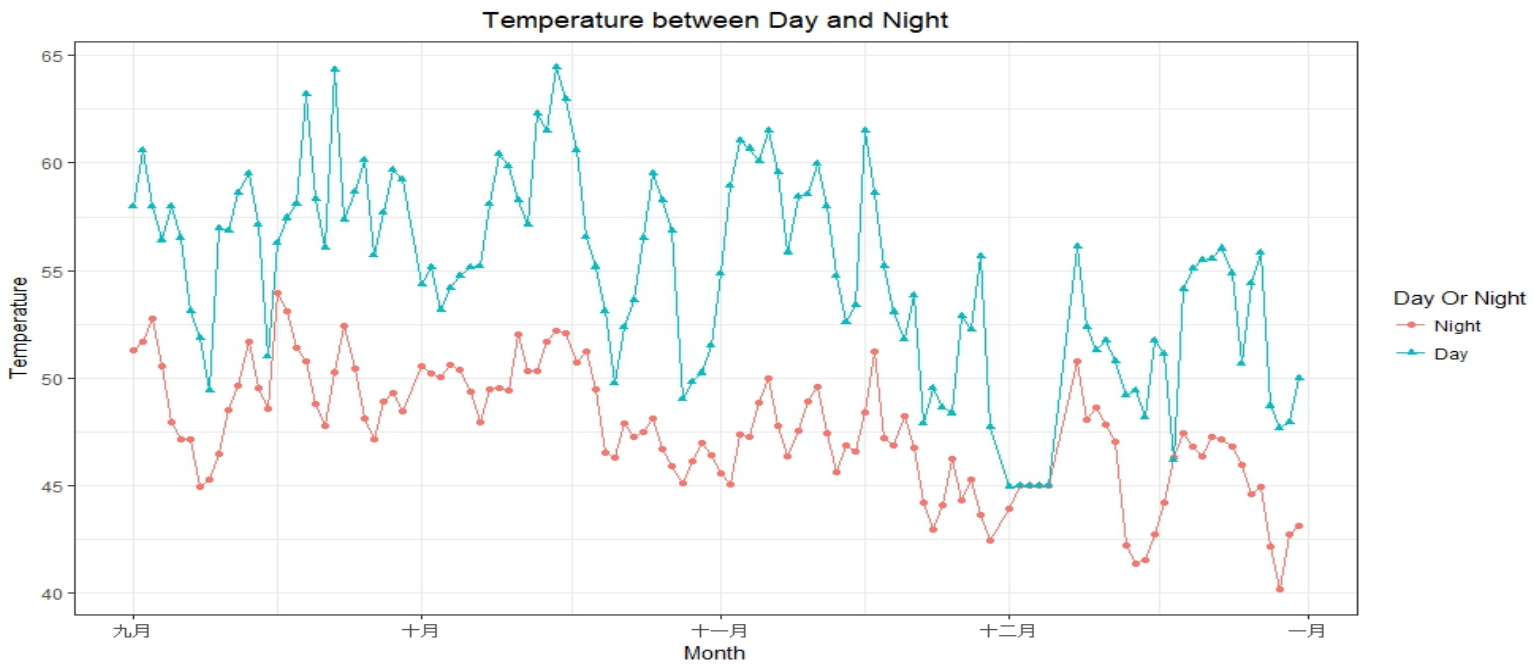
Radiation between Day and Night



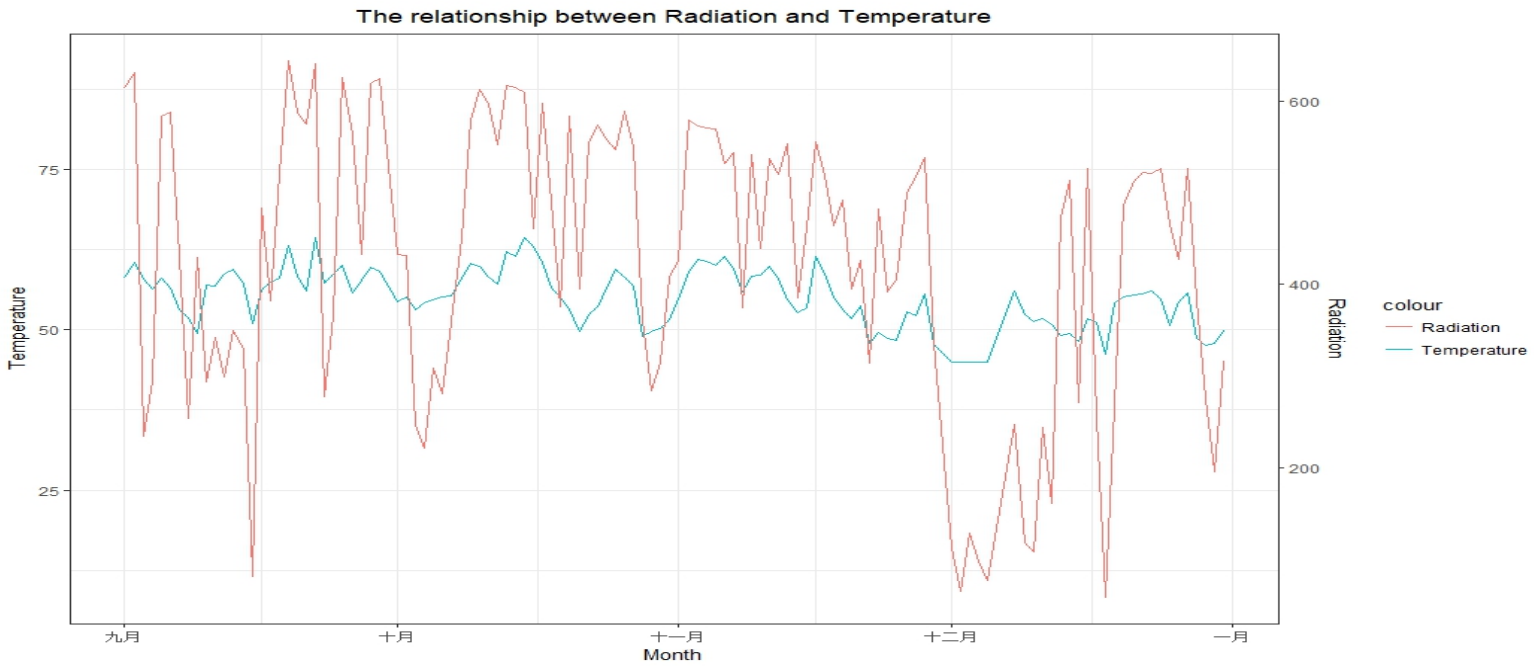
Winter is Coming



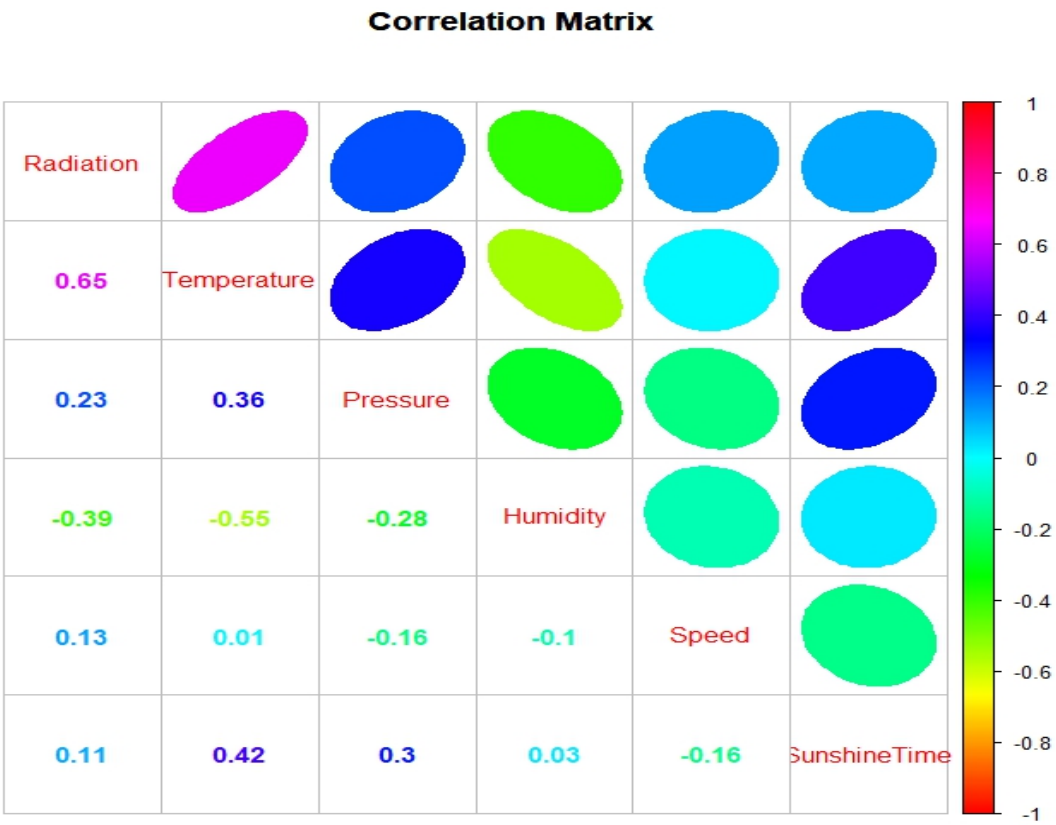
Temperature between Day and Night



The relationship between Radiation and Temperature



Correlation Matrix



Regression Model

- $\boldsymbol{y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \epsilon \cdot \epsilon \sim N(0, \sigma^2)}$

y : Radiation

x_1 : Temperature

x_2 : Pressure

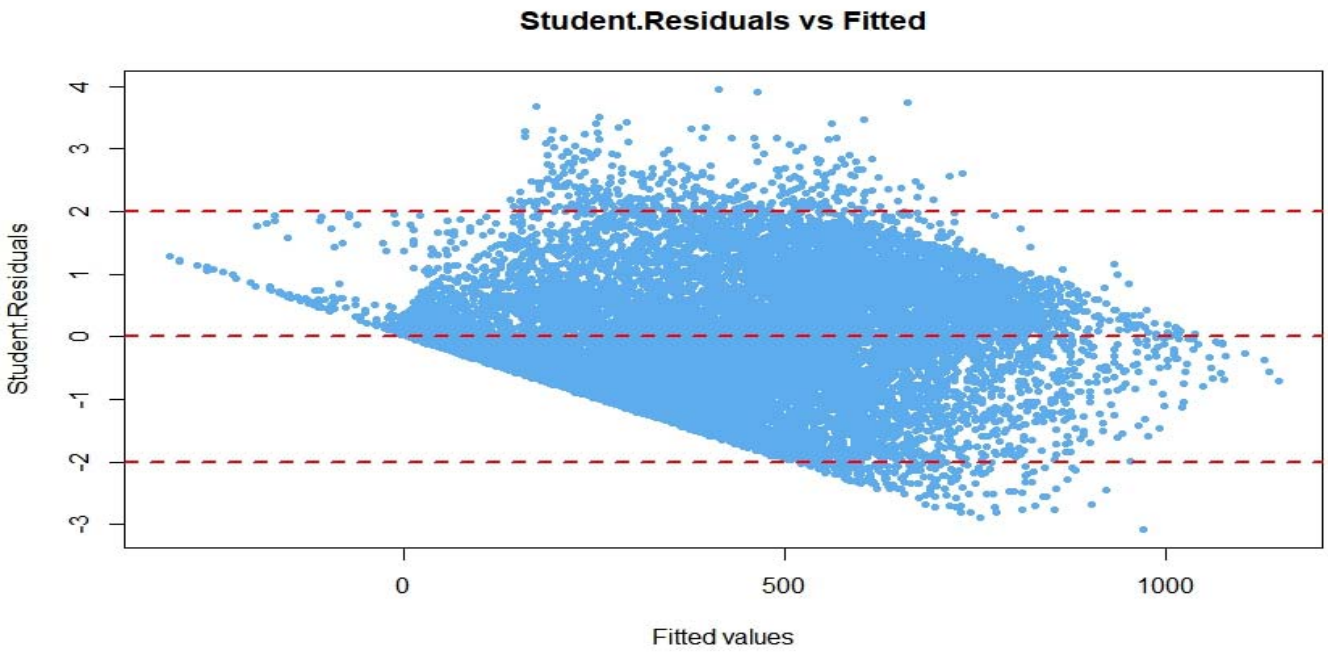
x_3 : Humidity

x_4 : Speed

Regression Model

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.23e+03	1.24e+03	-4.217	2.49e-05
Temperature	3.41e+01	4.14e-01	82.428	< 2e-16
Pressure	1.23e+02	4.09e+01	3.017	0.00256
Humidity	-5.19e-01	9.48e-02	-5.475	4.45e-08
Speed	1.10e+01	5.72e-01	19.151	< 2e-16
Multiple R-squared:	0.4322			
Adjusted R-squared:	0.432			

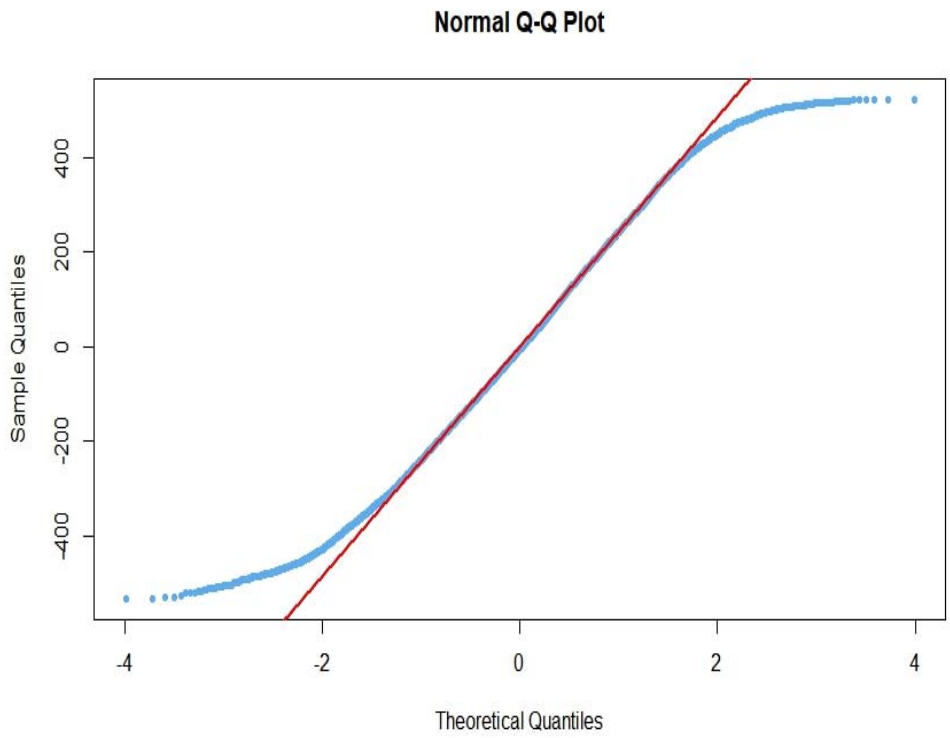
Residual Plot



Regression Model(fitting)

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.19e+03	1.14e+03	-3.686	0.000229
Temperature	3.47e+01	3.77e-01	92.151	< 2e-16
Pressure	8.89e+01	3.75e+01	2.373	0.017644
Humidity	-1.07e+00	8.69e-02	-12.338	< 2e-16
Speed	1.19e+01	5.28e-01	22.594	< 2e-16
Multiple R-squared:	0.5154			
Adjusted R-squared:	0.5152			

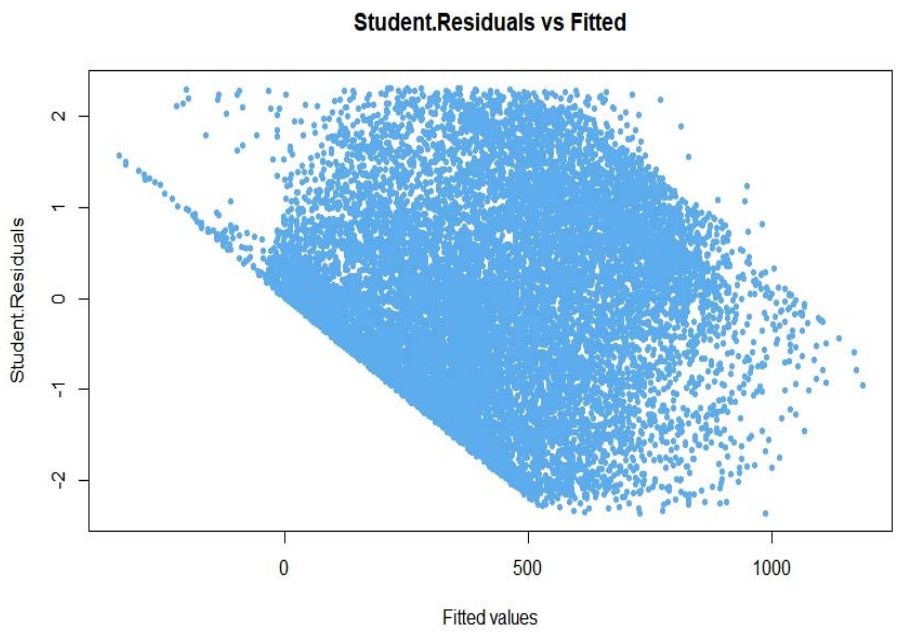
Normal Q-Q Plot and Testing Result



- **Anderson-Darling normality test**

  A = 19.475, p-value < 2.2e-16

Residual Plot and Testing Result

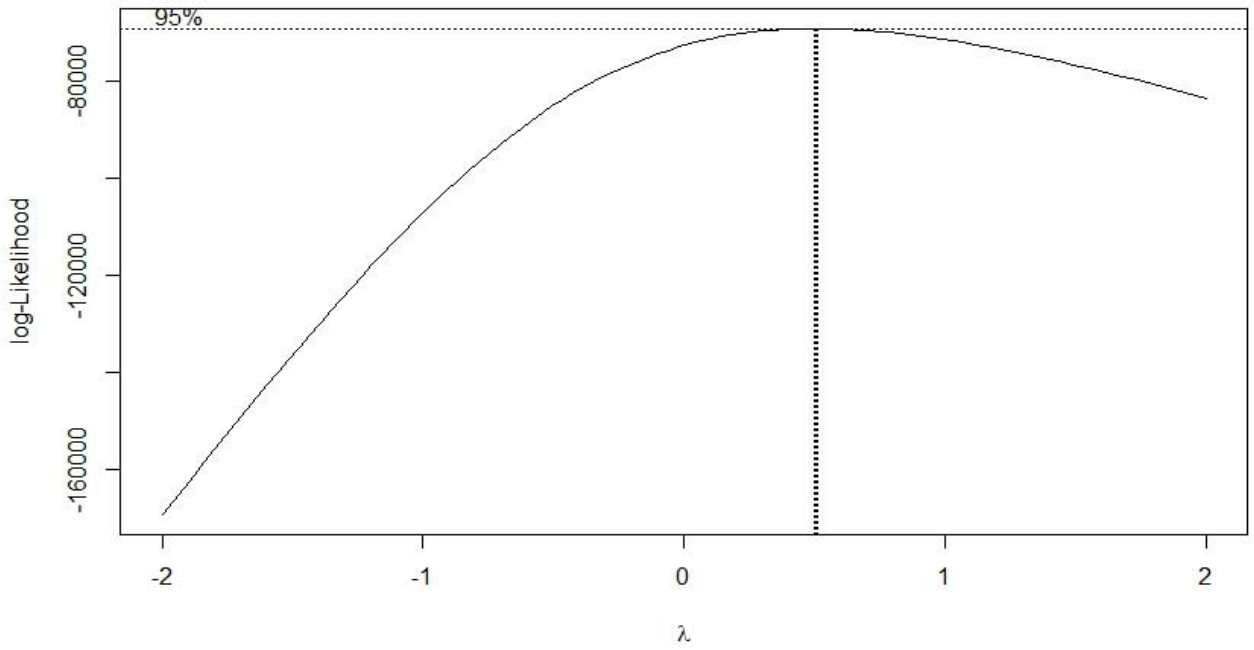


- **studentized Breusch-Pagan test**

  BP = 662.31, df = 4, p-value < 2.2e-16

Box-Cox Transformation

$\lambda = 0.5$



Refitting Model

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.073e+02	3.195e+01	-3.358	0.000787
Temperature	1.011e+00	1.059e-02	95.470	< 2e-16
Pressure	2.272e+00	1.053e+00	2.157	0.031030
Humidity	-1.381e-02	2.443e-03	-5.653	1.61e-08
Speed	3.155e-01	1.486e-02	21.235	< 2e-16

Multiple R-squared: 0.51

Adjusted R-squared: 0.5099

- **Anderson-Darling normality test**

  A = 34.871, p-value < 2.2e-16

- **studentized Breusch-Pagan test**

  BP = 366.47, df = 4, p-value < 2.2e-16

Conclusion

- 不好的模型，不能將台灣資料帶入求出我們想要的結果
  - 做出的迴歸模型，雖然在解釋力上有5成多的強度，但在檢驗殘差為常態時，結果是顯著拒絕假設
  - 在針對殘差的變異數為常數時，得到的結果也是否定的
- BOX-COX轉換
  - 經過我們的檢定後發現轉換後的模型並沒有比原本模型好太多，且原本違反假設的事實也並沒有因為我們的轉換而改善。

Conclusion

- 配合我們找到的文獻，我們認為最主要影響是變數的不足
  - 行政院的報告中，他們利用了更多的變數
  - 從原始資料中我們僅用了4種
- 其他因素
  - 空氣中的水和二氧化碳會吸收紅外線輻射
  - 臭氧會吸收紫外線輻射
  - 空氣分子、塵埃、雲滴會改變輻射方向
  - 雲層和塵埃對太陽輻射則是有反射作用，使部份太陽輻射來不到地面

Conclusion

- 雖然我們無法做出有效的模型來預測，但我們卻可以從一開始對資料的分析得到一些相關訊息
  - 輻射跟溫度、壓力、風速及日照時間是呈現正相關的
  - 濕度是呈現負相關
- 在地點的選擇上，我們可以選擇在這些正相關變數中表現較好，而濕度表現較差的地點來建設發電廠。