Raindrops don't keep falling on my head

REGINA

1REGINACHEONG@GMAIL.COM

Agenda

1. Background
2. Data Source
3. Tools
4. Findings
5. Conclusion

Background

When should I bring an umbrella to fetch my kid at Newton?

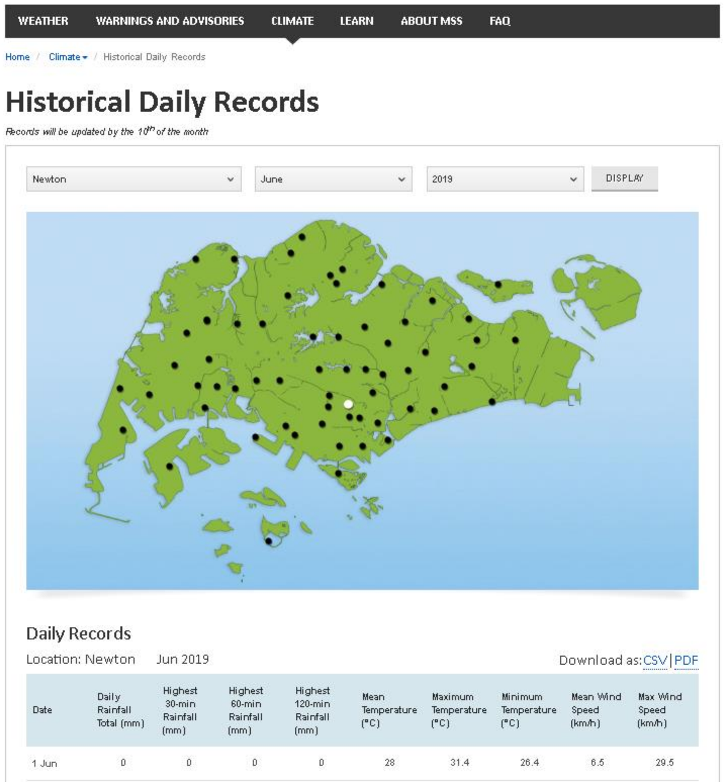
How much transport allowance should I budget for grab rides for my kid?

Data Source

National Environment Agency

Safeguard · Nurture · Cherish

Newton – 2014 to 2019



Historical Daily Records

Records will be updated by the 10th of the month

Newton | June | 2019

Daily Records

Location: Newton Jun 2019

Download as: CSV | PDF

Date	Daily Rainfall Total (mm)	Highest 30-min Rainfall (mm)	Highest 60-min Rainfall (mm)	Highest 120-min Rainfall (mm)	Mean Temperature (°C)	Maximum Temperature (°C)	Minimum Temperature (°C)	Mean Wind Speed (km/h)	Max Wind Speed (km/h)
1 Jun	0	0	0	0	28	31.4	26.4	6.5	29.5

Tools

1. Web scrapping
   - Beautifulsoup4
   - Selenium

2. EDA and Data Cleaning
   - Numpy
   - Pandas

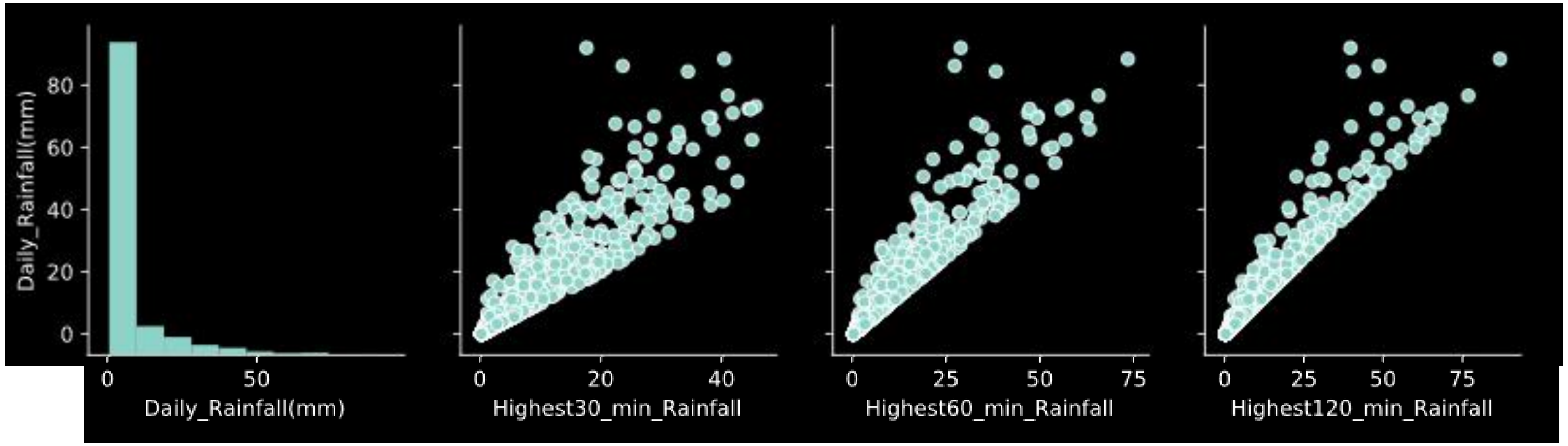
3. Model Building and Selection
   - Scikit
   - Scipy.stats
   - Statsmodel
   - Seaborn and Matplotlib

From 3600 data rows -> 1800 rows

Findings – Remove “noisy” parameters

PARAMETERS

1. Date
2. Daily Rainfall Total (mm)
3. Highest 30-min Rainfall (mm)
4. Highest 60-min Rainfall (mm)
5. Highest 120-min Rainfall (mm)
6. Mean Temperature (°C)
7. Maximum Temperature (°C)
8. Minimum Temperature (°C)
9. Mean Wind Speed (km/h)
10. Max Wind Speed (km/h)

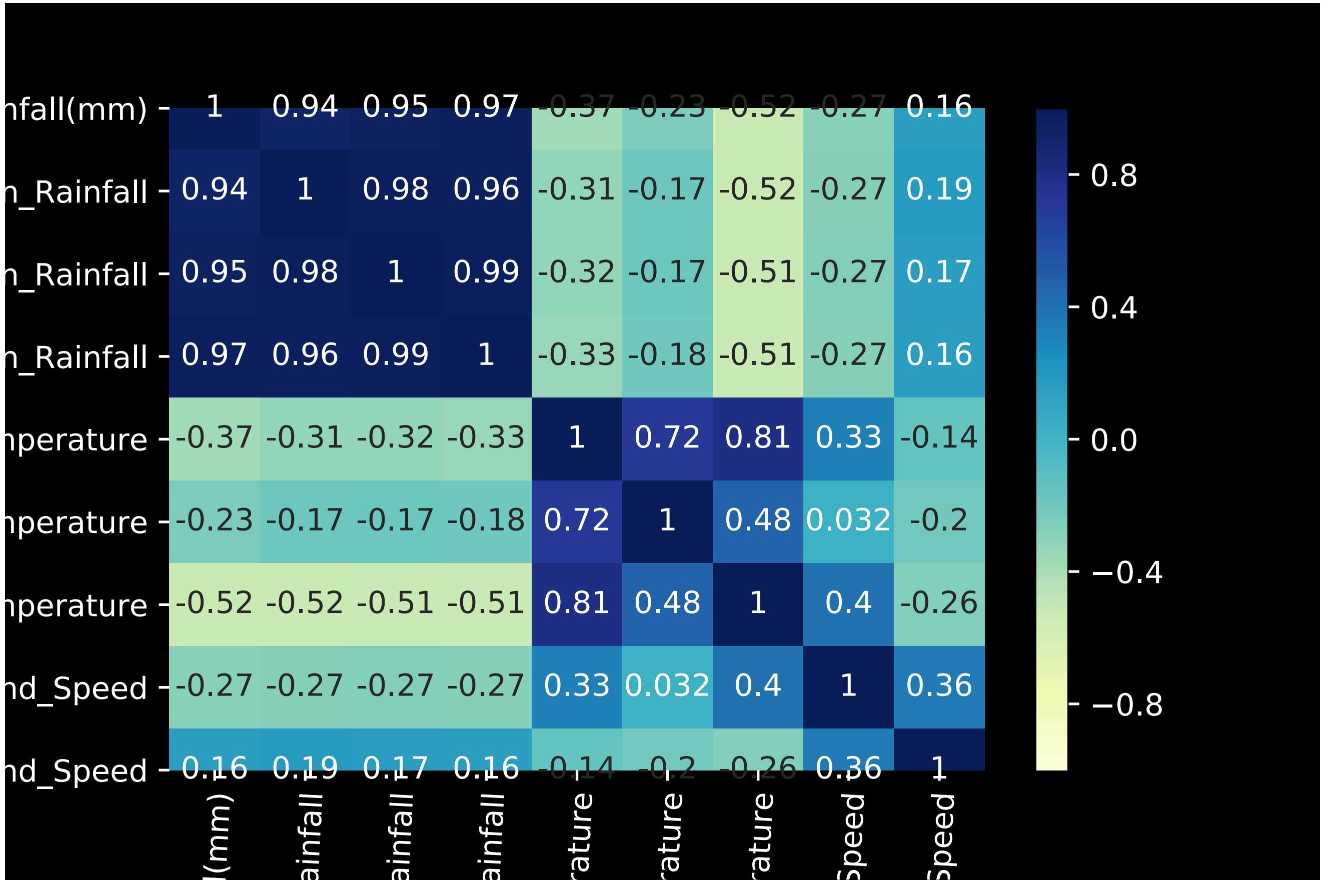


Findings

PARAMETERS

1. Date
2. Daily Rainfall Total (mm)
3. Highest 30-min Rainfall (mm)
4. Highest 60-min Rainfall (mm)
5. Highest 120-min Rainfall (mm)
6. Mean Temperature (°C)
7. Maximum Temperature (°C)
8. Minimum Temperature (°C)
9. Mean Wind Speed (km/h)
10. Max Wind Speed (km/h)

COLLINEARITY

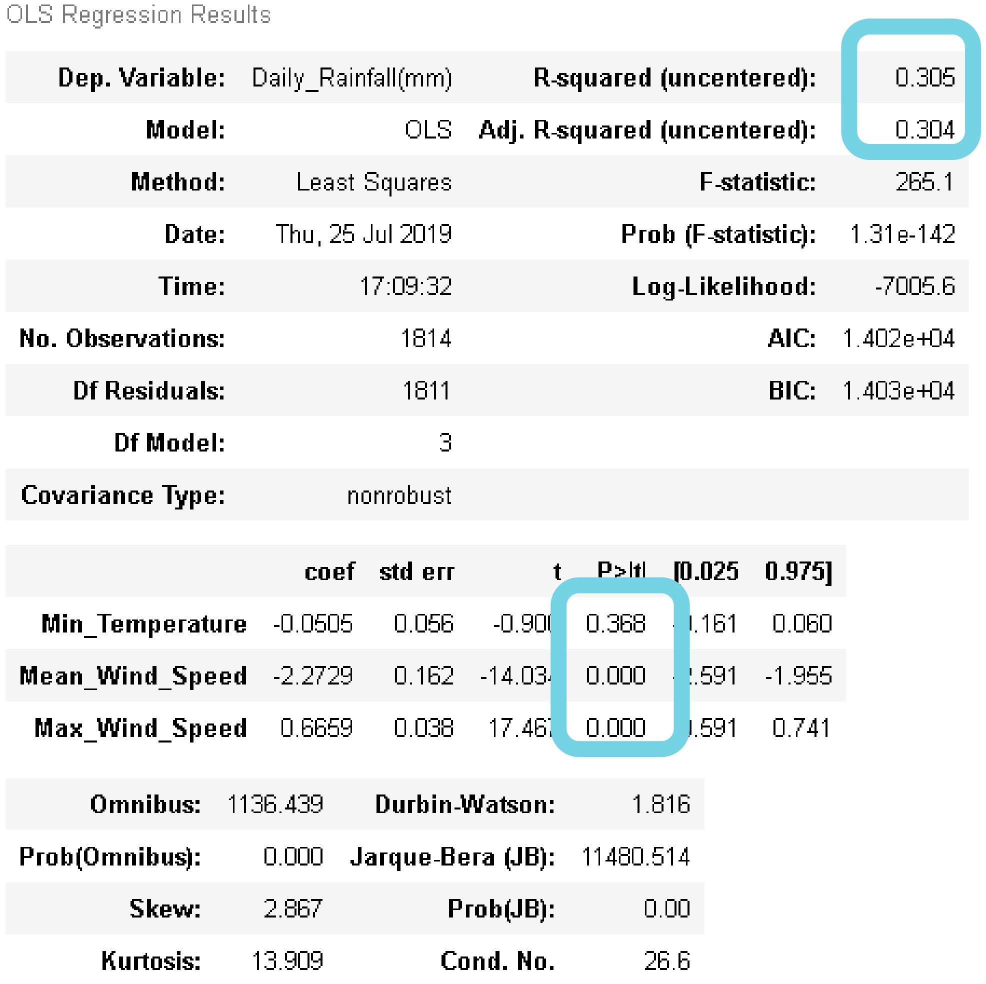


	l(mm)	ainfall	ainfall	ainfall	rature	rature	rature	Speed	Speed
nfall(mm)	1	0.94	0.95	0.97	-0.37	-0.23	-0.52	-0.27	0.16
n_Rainfall	0.94	1	0.98	0.96	-0.31	-0.17	-0.52	-0.27	0.19
n_Rainfall	0.95	0.98	1	0.99	-0.32	-0.17	-0.51	-0.27	0.17
n_Rainfall	0.97	0.96	0.99	1	-0.33	-0.18	-0.51	-0.27	0.16
nperature	-0.37	-0.31	-0.32	-0.33	1	0.72	0.81	0.33	-0.14
nperature	-0.23	-0.17	-0.17	-0.18	0.72	1	0.48	0.032	-0.2
nperature	-0.52	-0.52	-0.51	-0.51	0.81	0.48	1	0.4	-0.26
nd_Speed	-0.27	-0.27	-0.27	-0.27	0.33	0.032	0.4	1	0.36
nd_Speed	0.16	0.19	0.17	0.16	-0.14	-0.2	-0.26	0.36	1

Findings – Model 1

PARAMETERS

1. Date
2. Daily Rainfall Total (mm)
3. Highest 30-min Rainfall (mm)
4. Highest 60-min Rainfall (mm)
5. Highest 120-min Rainfall (mm)
6. Mean Temperature (°C)
7. Maximum Temperature (°C)
8. Minimum Temperature (°C)
9. Mean Wind Speed (km/h)
10. Max Wind Speed (km/h)



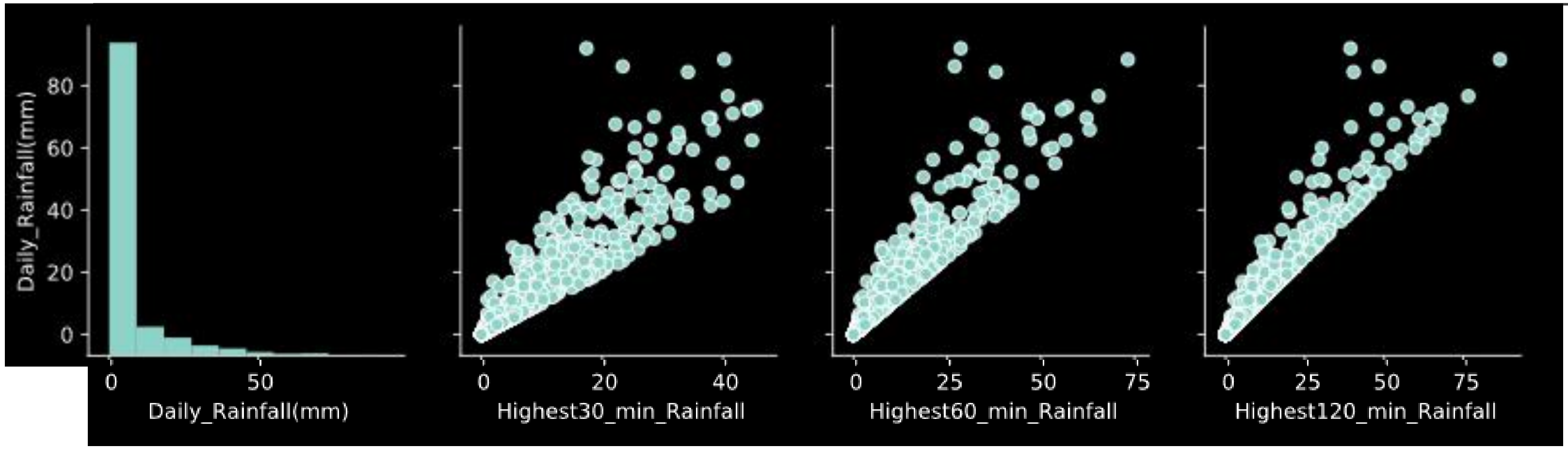
OLS Regression Results

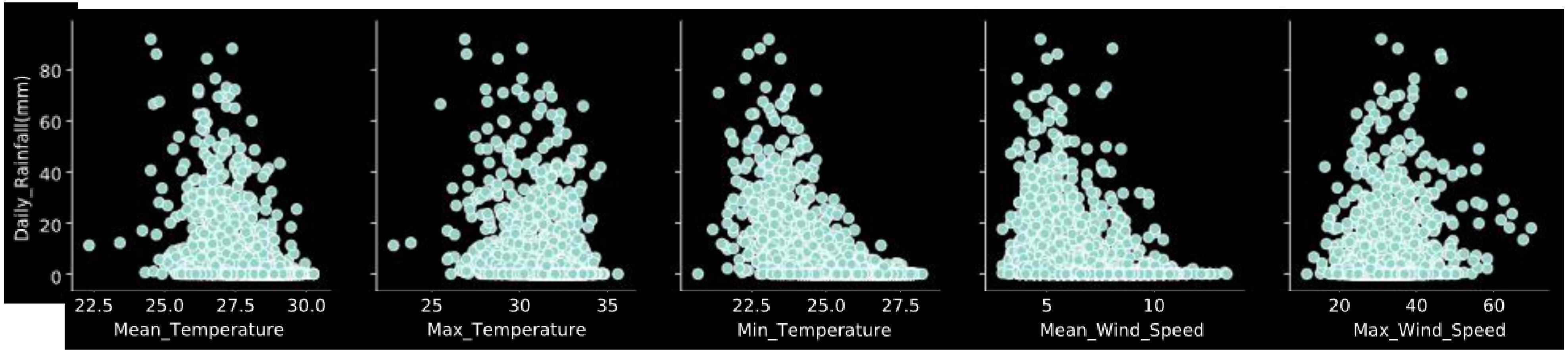
Dep. Variable:	Daily_Rainfall(mm)	**R-squared (uncentered):**	0.305
Model:	OLS	**Adj. R-squared (uncentered):**	0.304
Method:	Least Squares	**F-statistic:**	265.1
Date:	Thu, 25 Jul 2019	**Prob (F-statistic):**	1.31e-142
Time:	17:09:32	**Log-Likelihood:**	-7005.6
No. Observations:	1814	**AIC:**	1.402e+04
Df Residuals:	1811	**BIC:**	1.403e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Min_Temperature	-0.0505	0.056	-0.900	0.368	-0.161	0.060
Mean_Wind_Speed	-2.2729	0.162	-14.034	0.000	-2.591	-1.955
Max_Wind_Speed	0.6659	0.038	17.467	0.000	0.591	0.741

Omnibus:	1136.439	**Durbin-Watson:**	1.816
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	11480.514
Skew:	2.867	**Prob(JB):**	0.00
Kurtosis:	13.909	**Cond. No.**	26.6

Findings – Collinearity

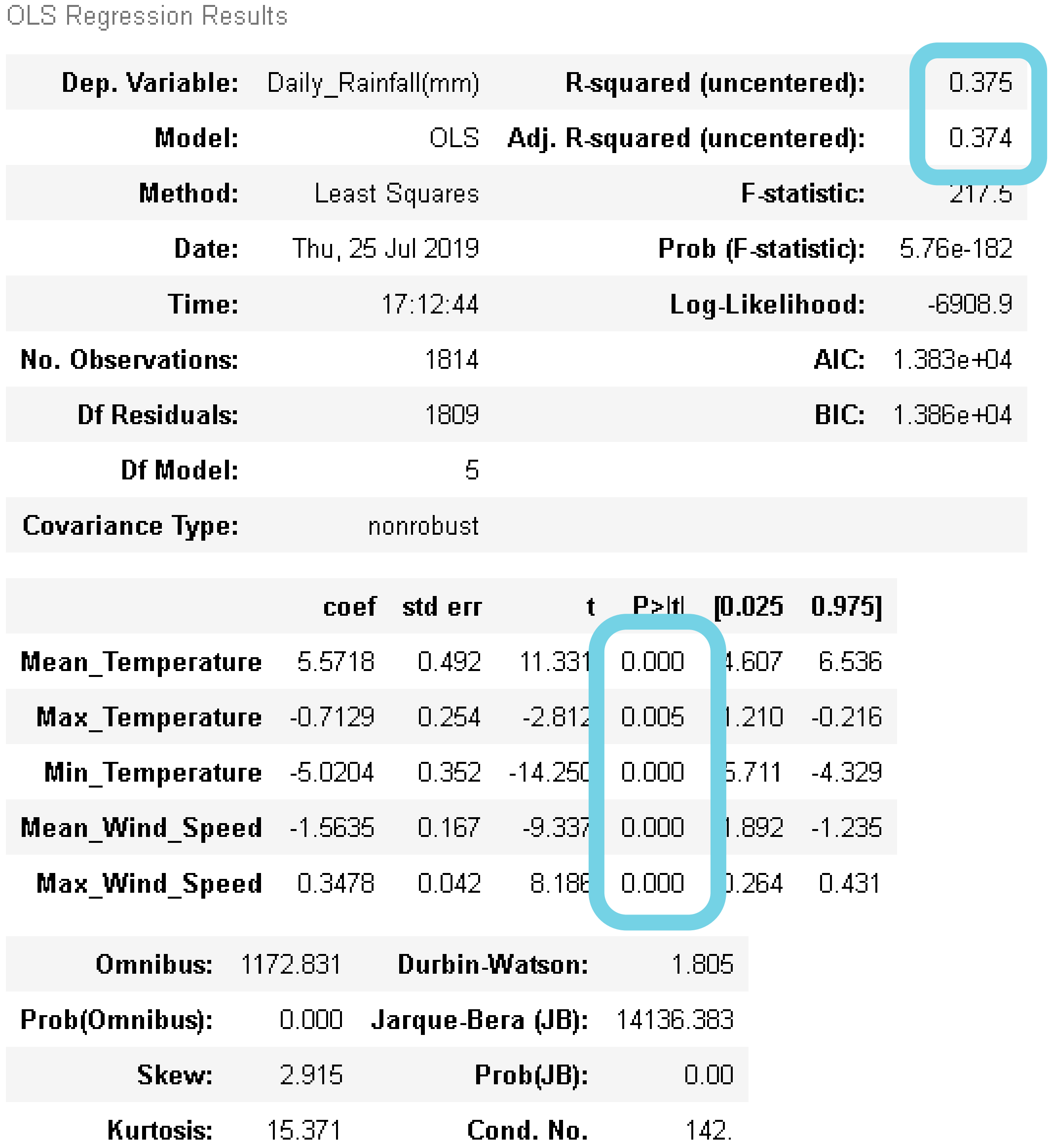




Findings – Model 2

PARAMETERS

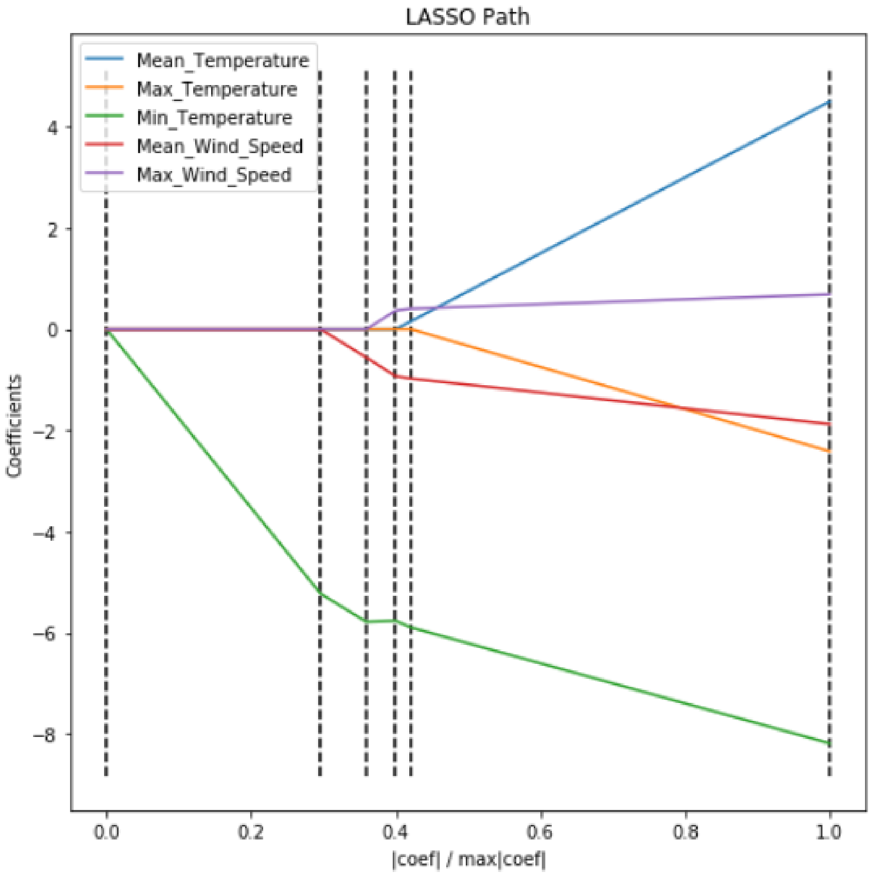
1. Date
2. Daily Rainfall Total (mm)
3. Highest 30-min Rainfall (mm)
4. Highest 60-min Rainfall (mm)
5. Highest 120-min Rainfall (mm)
6. Mean Temperature (°C)
7. Maximum Temperature (°C)
8. Minimum Temperature (°C)
9. Mean Wind Speed (km/h)
10. Max Wind Speed (km/h)



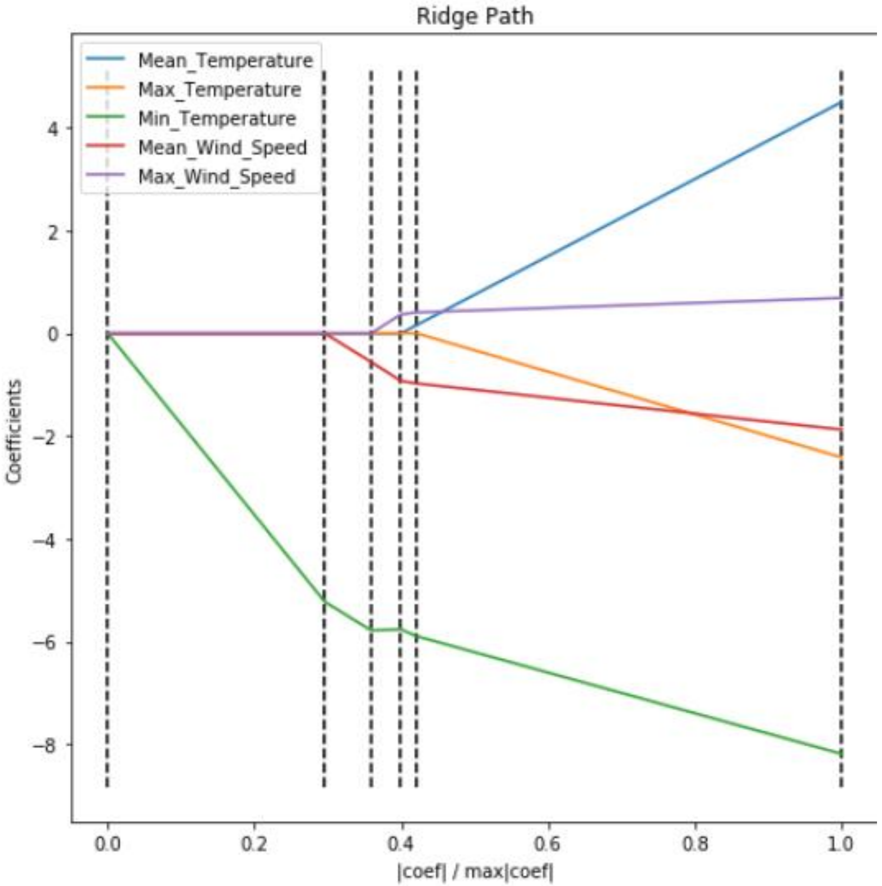
```
OLS Regression Results

     Dep. Variable:  Daily_Rainfall(mm)        R-squared (uncentered):       0.375
            Model:                 OLS   Adj. R-squared (uncentered):       0.374
           Method:       Least Squares                   F-statistic:       217.5
             Date:    Thu, 25 Jul 2019             Prob (F-statistic):  5.76e-182
             Time:            17:12:44                Log-Likelihood:     -6908.9
 No. Observations:                1814                           AIC:   1.383e+04
     Df Residuals:                1809                           BIC:   1.386e+04
         Df Model:                   5
  Covariance Type:           nonrobust

                     coef   std err        t   P>|t|   [0.025   0.975]
Mean_Temperature   5.5718     0.492   11.331   0.000    4.607    6.536
 Max_Temperature  -0.7129     0.254   -2.812   0.005    1.210   -0.216
 Min_Temperature  -5.0204     0.352  -14.250   0.000    5.711   -4.329
 Mean_Wind_Speed  -1.5635     0.167   -9.337   0.000    1.892   -1.235
  Max_Wind_Speed   0.3478     0.042    8.186   0.000    0.264    0.431

      Omnibus:  1172.831     Durbin-Watson:      1.805
Prob(Omnibus):     0.000  Jarque-Bera (JB):  14136.383
         Skew:     2.915          Prob(JB):       0.00
     Kurtosis:    15.371         Cond. No.       142.
```

Findings – LASSO vs Ridge Models

LASSO

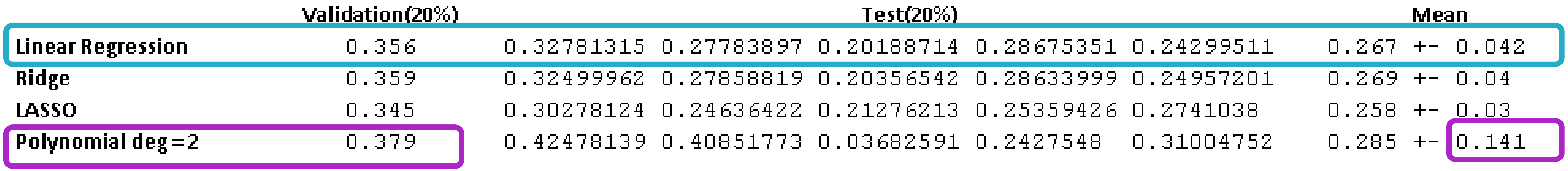


RIDGE



Findings – Model Selection

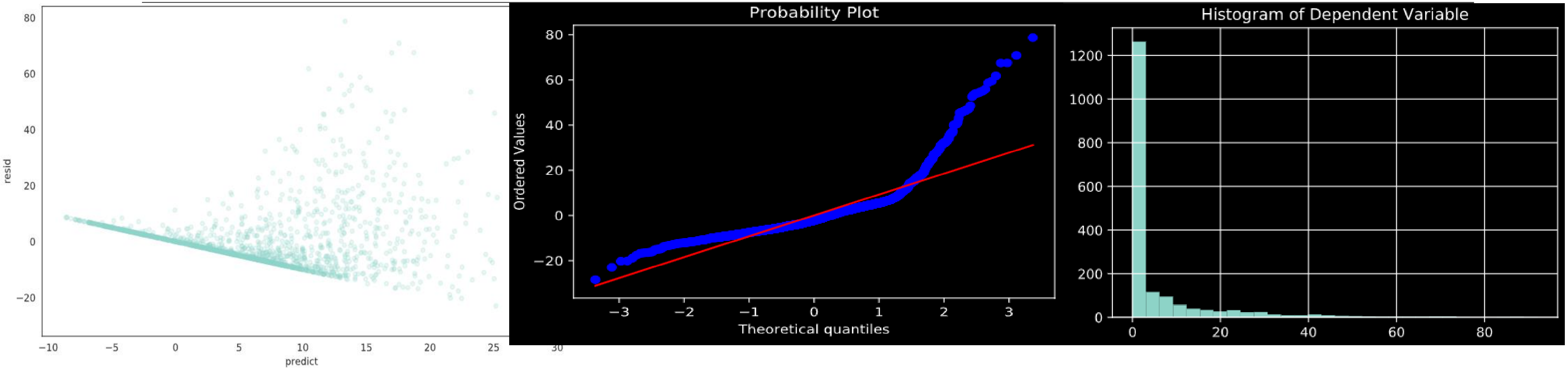
APPROACH: TRAINING (60%) VS VALIDATION (20%) VS TEST (20%)



	Validation(20%)	Test(20%)					Mean
Linear Regression	0.356	0.32781315	0.27783897	0.20188714	0.28675351	0.24299511	0.267 +- 0.042
Ridge	0.359	0.32499962	0.27858819	0.20356542	0.28633999	0.24957201	0.269 +- 0.04
LASSO	0.345	0.30278124	0.24636422	0.21276213	0.25359426	0.2741038	0.258 +- 0.03
Polynomial deg=2	0.379	0.42478139	0.40851773	0.03682591	0.2427548	0.31004752	0.285 +- 0.141

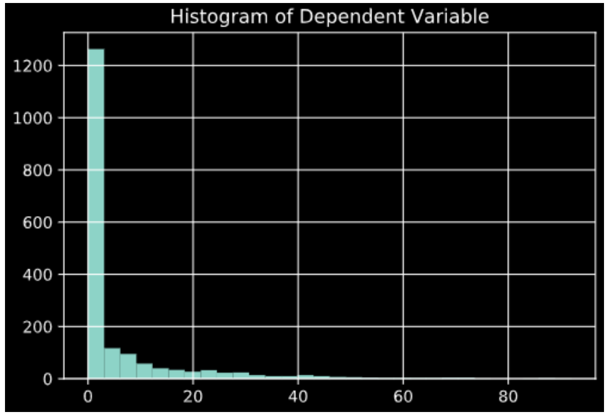
Select Model: Linear Regression

Findings – Linear Regression Model Evaluation

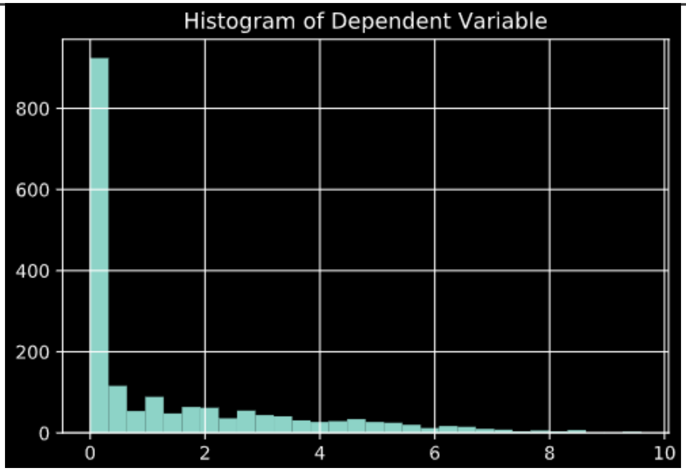


Evaluation: The raw Model cant predict extreme target values

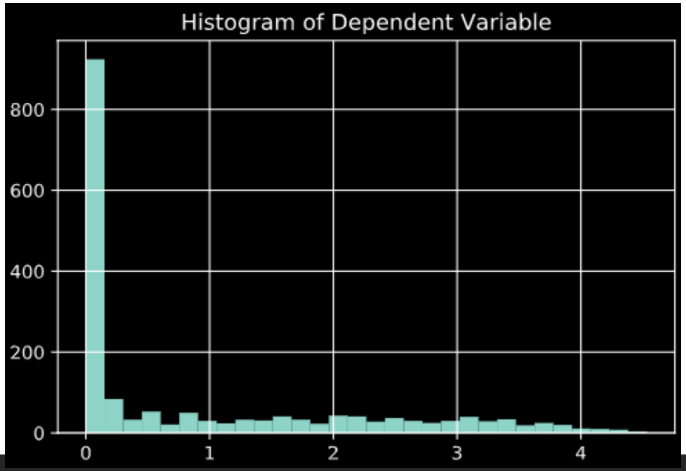
Findings – Model Evaluation



Square root of y



Log y

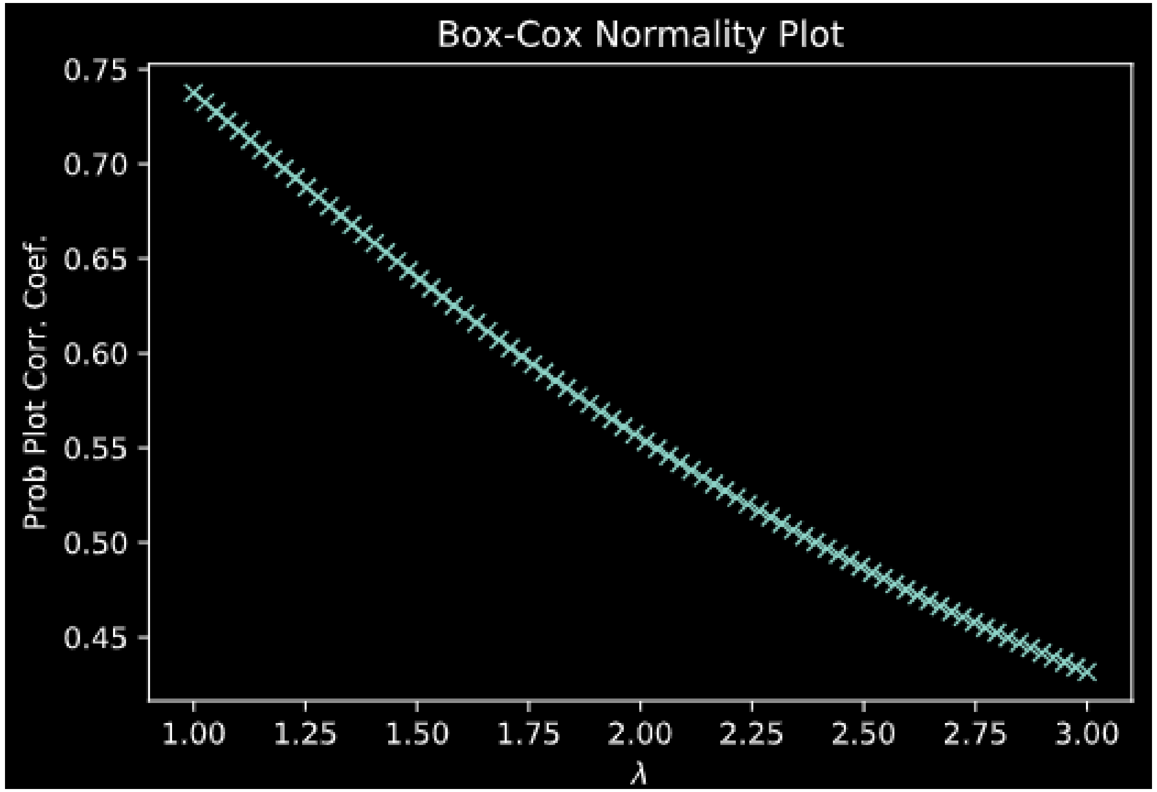


Conclusion

Historical Temperature and Wind Speed alone are *not* good parameters to predict Rainfall at Newton.

Probable reasons:

1. Cloud movements/weather in surrounding areas.
2. There is a seasonal/time dimension that is not accounted.



Thank you

REGINA