Australia Next Day Rain Prediction

Allister Mounsey

11 February 2019

Contents

1 Introduction

2 Methods and Analysis

Figure 1 shows

Figure 2 shows seasonal

Separation. Figure 3 example

Not well separated Figure 4

Table 1: Results of Kolmogorov-Smirnov test on Conditional Distributions (RainTomorrow = 0 vs. RainTomorrow = 0)

NumericVariable	D
MinTemp	0.09
MaxTemp	0.14
Rainfall	0.34
Evaporation	0.14
Sunshine	0.47
WindGustSpeed	0.23
WindSpeed9am	0.08
WindSpeed3pm	0.08
Humidity9am	0.29
Humidity3pm	0.46
Pressure9am	0.26
Pressure3pm	0.23
Cloud9am	0.34
Cloud3pm	0.41
Temp9am	0.04
Temp3pm	0.17

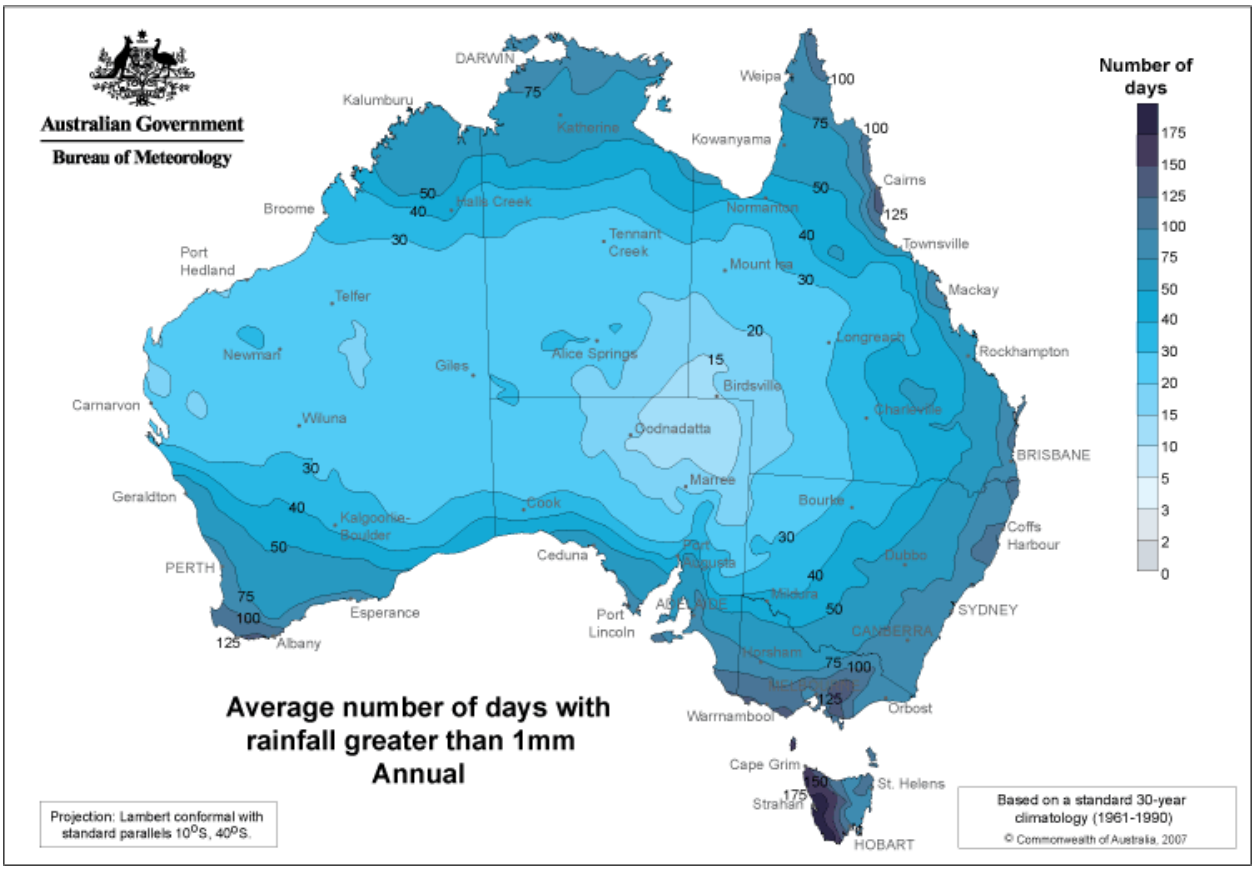

Figure 1: Rain Days across Australia Source: http://www.bom.gov.au/jsp/ncc/climate_averages/raindays/index.jsp

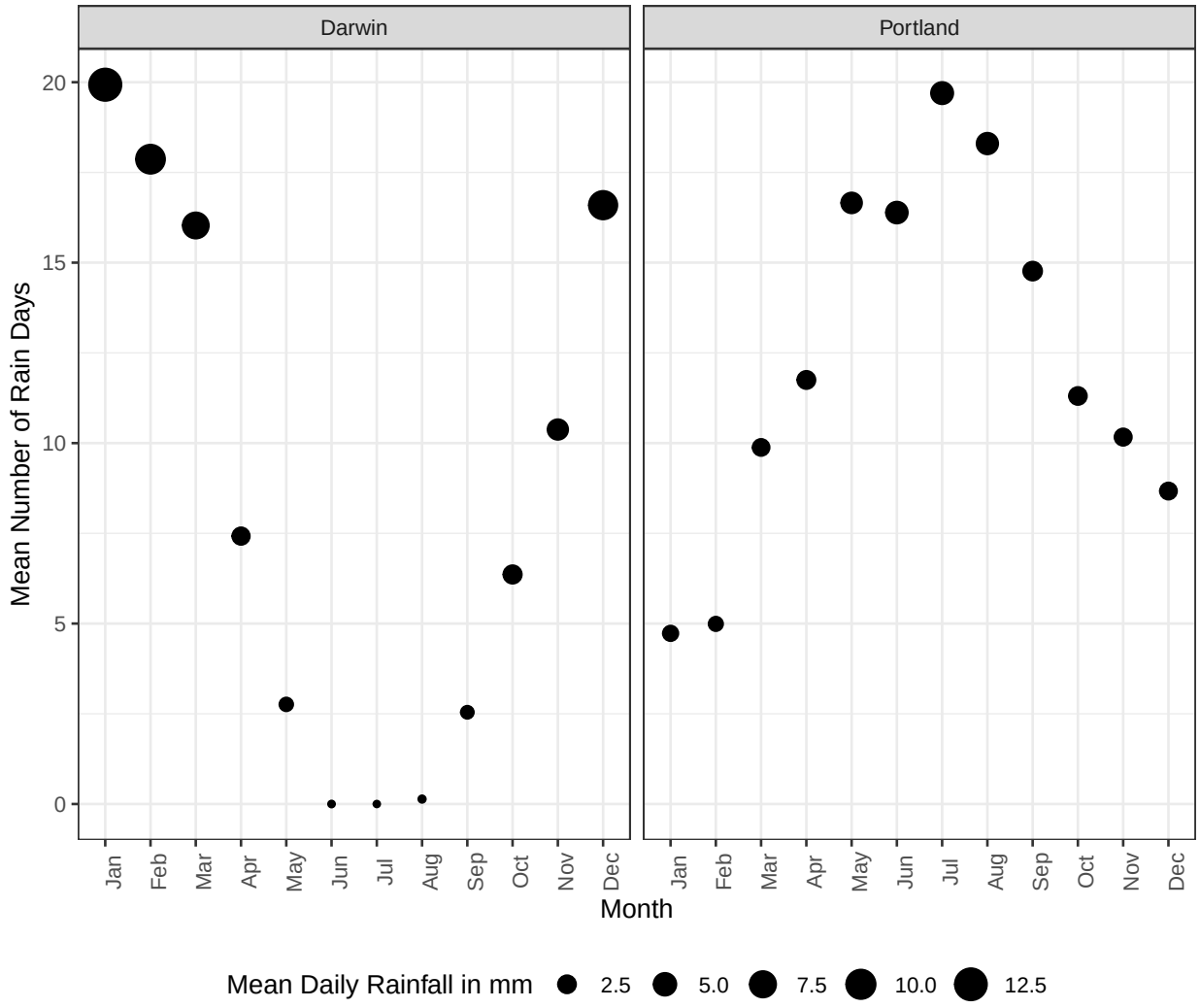

Figure 2: Monthly Rainfall Patterns by Location

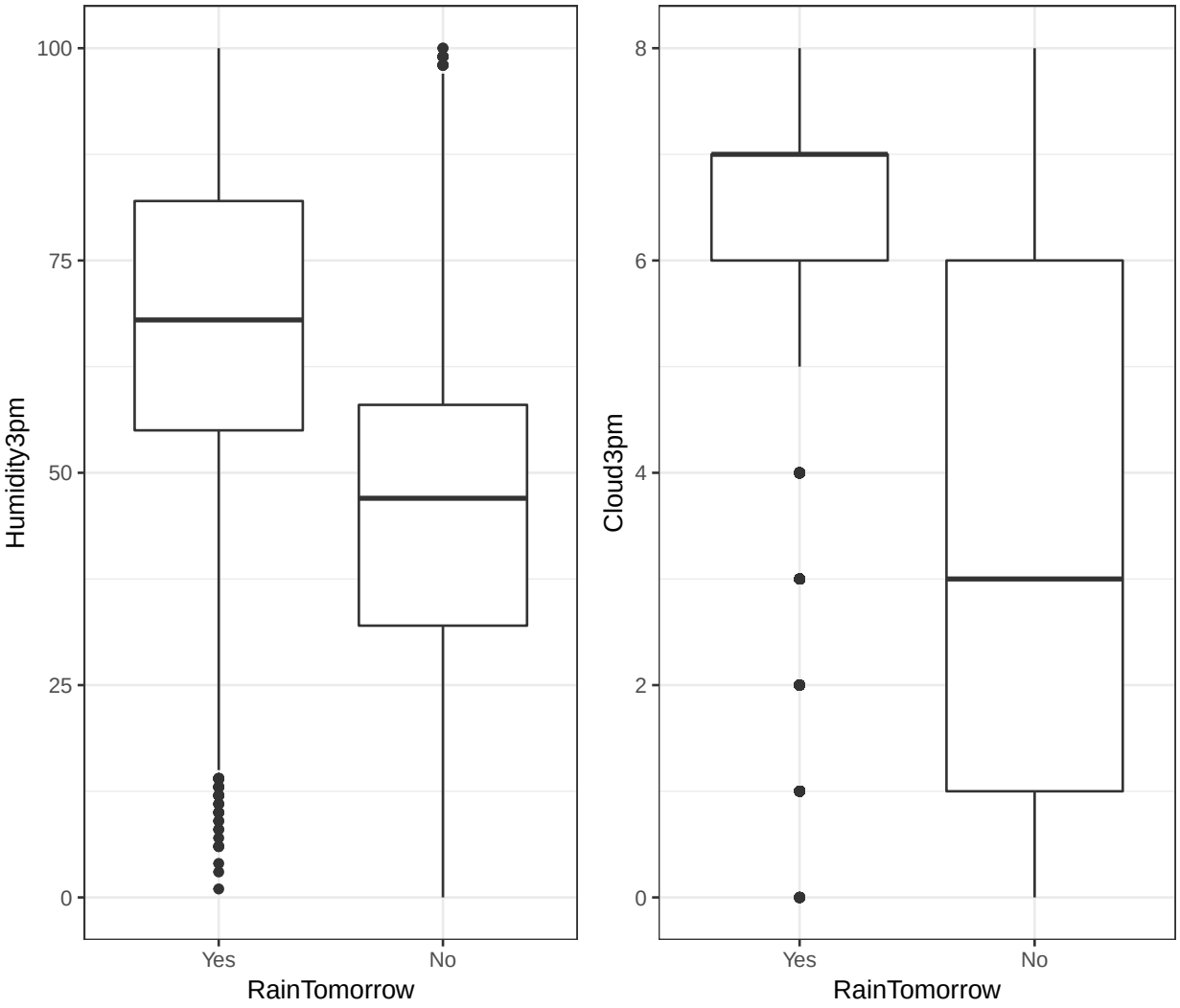

Figure 3: Examples of Some Well Separated Cases

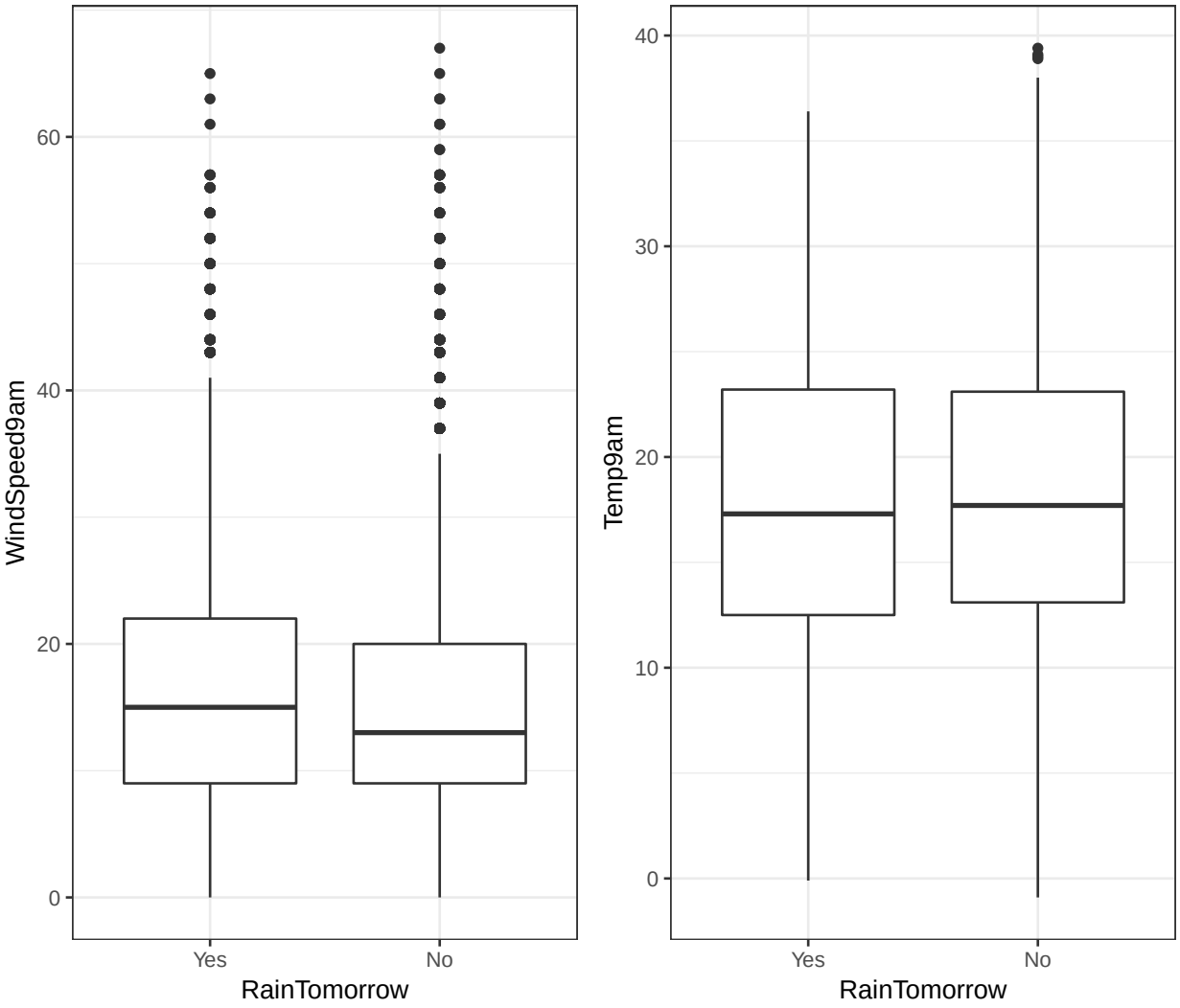

Figure 4: Examples of Some **NOT** Well Separated Cases

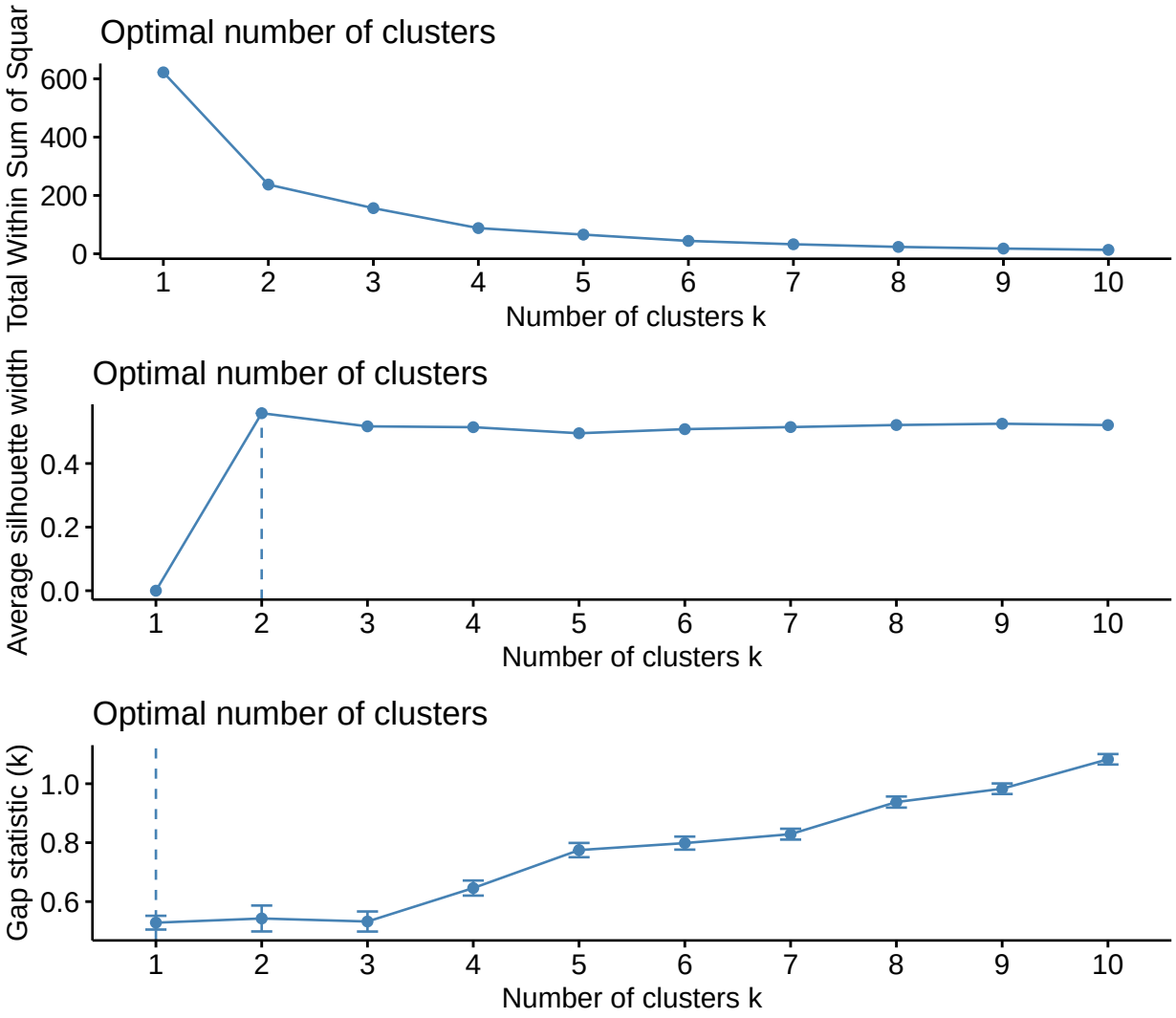

Figure 5: Finding Optimal Number for Location - Month Clustering

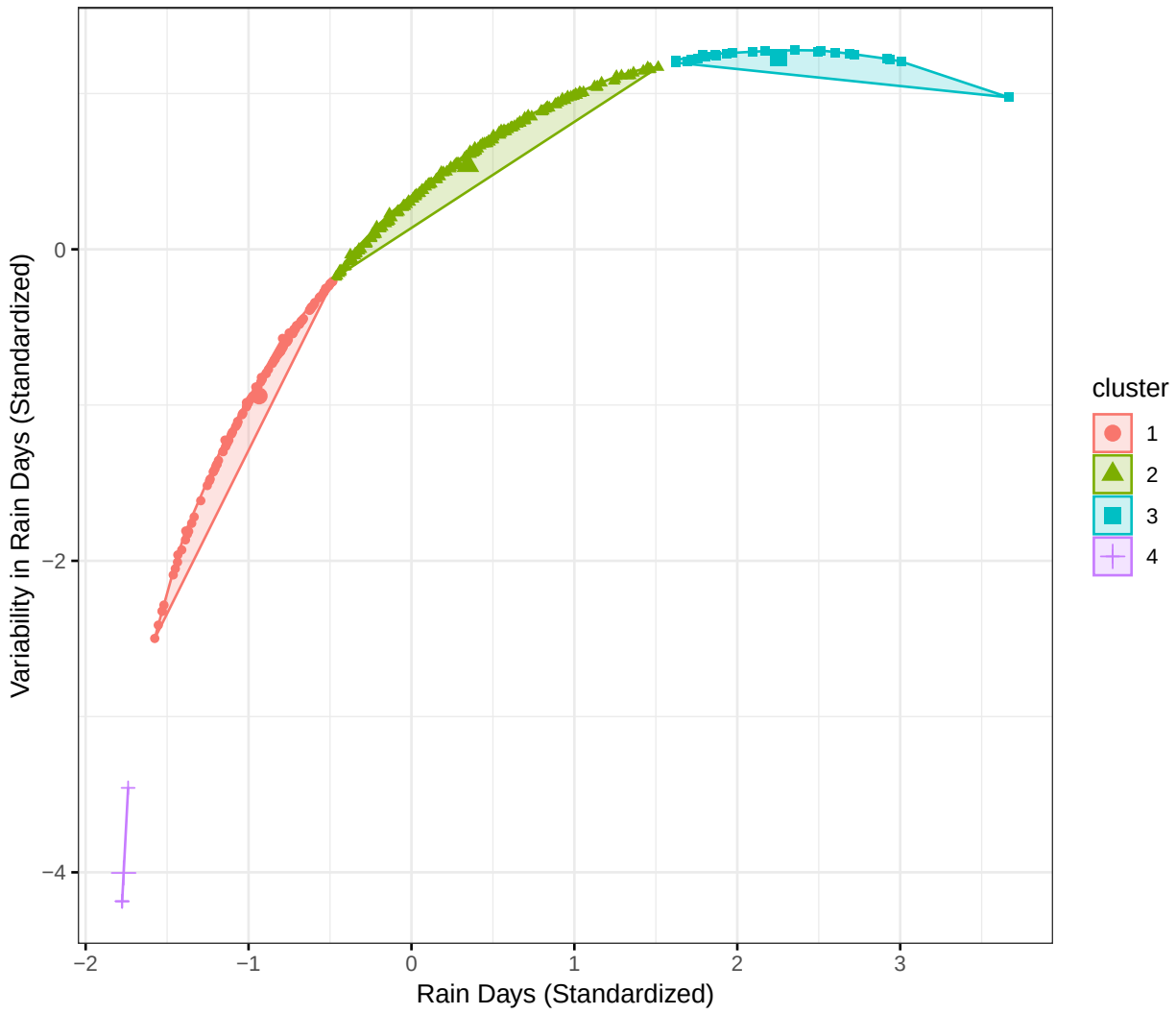

Figure 6: Number of Rain Days and Variability in Rain Days accross Clusters

Table 2: Location-Month Clusters

Location	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
AliceSprings	1	1	1	1	1	1	1	1	1	1	1	2
Brisbane	2	2	2	2	2	2	1	1	1	1	2	2
Cairns	3	3	3	2	2	2	2	1	1	2	2	2
Canberra	1	2	1	2	1	2	1	2	2	2	2	2
Cobar	1	2	1	1	2	2	1	4	1	1	1	1
CoffsHarbour	2	2	3	2	2	2	2	1	2	2	2	3
Darwin	3	3	3	2	1	4	4	4	1	2	2	3
Hobart	2	1	2	2	2	2	2	2	2	2	2	2
Melbourne	1	1	2	2	2	2	2	2	2	2	2	2
MelbourneAirport	1	1	2	2	2	2	2	2	2	2	2	2
Mildura	1	1	1	1	1	1	1	1	1	1	1	1
Moree	2	1	1	1	1	1	1	1	1	1	1	2
MountGambier	1	1	2	2	3	3	3	3	2	2	2	2
NorfolkIsland	2	2	2	2	2	3	3	2	2	2	2	2
Nuriootpa	1	1	1	2	2	2	2	2	2	1	1	2
Perth	1	1	1	1	2	2	3	2	2	1	1	1
PerthAirport	1	1	1	1	2	2	2	2	2	1	1	1
Portland	1	2	2	2	3	3	3	3	3	2	2	2
Sale	1	2	2	2	2	2	2	2	2	2	2	2
Sydney	2	2	2	2	1	2	2	2	2	1	2	2
SydneyAirport	2	2	2	2	2	2	2	2	2	2	2	2
Townsville	2	3	2	1	1	1	1	1	1	1	2	2
WaggaWagga	1	2	1	1	2	2	2	2	1	1	1	2
Watsonia	1	1	2	2	2	2	2	2	2	2	2	2
Williamtown	2	2	2	2	2	3	2	1	2	2	2	2
Woomera	1	1	1	1	1	1	1	1	1	1	1	1

Table 3: Conditional Probabilities No Rain Tomorrow vs. Rain Tomorrow Given WindGust Direction (Location = Cobar)

	E	ENE	ESE	N	NE	NNE	NNW	NW	S	SE	SSE	SSW	SW	W	WNW	WSW	Total
Yes	0.14	0.1	0.15	0.05	0.41	0.33	0.14	0.12	0.06	0.07	0.08	0.02	0.07	0	0.11	0.08	0.12
No	0.86	0.9	0.85	0.95	0.59	0.67	0.86	0.88	0.94	0.93	0.92	0.98	0.93	1	0.89	0.92	0.88
Total	1.00	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	1.00	1.00	1.00

Table 4: Probabilites of Rain Tomorrow Conditioned on rainDir9am

	RainTomorrow=0	RainTomorrow=1	Total
rainDirGust=0	0.13	0.87	1
rainDirGust=1	0.32	0.68	1
Total	0.22	0.78	1

3 Results

result from Logistic hyper parameter tuning

Table 5: Probabilites of Rain Tomorrow Conditioned on rainDir3pm

	RainTomorrow=0	RainTomorrow=1	Total
rainDir9am=0	0.14	0.86	1
rainDir9am=1	0.31	0.69	1
Total	0.22	0.78	1

Table 6: Probabilites of Rain Tomorrow Conditioned on rainDirGust

	RainTomorrow=0	RainTomorrow=1	Total
rainDir3pm=0	0.14	0.86	1
rainDir3pm=1	0.33	0.67	1
Total	0.22	0.78	1

Table 7: Optimal Hyperparameter Selection Using 10-fold nested Cross Validation

alpha	lambda	ROC
0.10	0.0003750	0.8912614
0.10	0.0037499	0.8899216
0.10	0.0374991	0.8863516
0.55	0.0003750	0.8912378
0.55	0.0037499	0.8889378
0.55	0.0374991	0.8854984
1.00	0.0003750	0.8912182
1.00	0.0037499	0.8875528
1.00	0.0374991	0.8806607

ROC	Sens	Spec	alpha	lambda	Resample
0.892	0.567	0.947	0.1	0.00037499130337396	Fold01
0.895	0.560	0.935	0.1	0.00037499130337396	Fold02
0.896	0.557	0.939	0.1	0.00037499130337396	Fold03
0.886	0.536	0.942	0.1	0.00037499130337396	Fold04
0.880	0.517	0.940	0.1	0.00037499130337396	Fold05
0.902	0.555	0.938	0.1	0.00037499130337396	Fold06
0.885	0.520	0.937	0.1	0.00037499130337396	Fold07
0.885	0.546	0.939	0.1	0.00037499130337396	Fold08
0.897	0.529	0.943	0.1	0.00037499130337396	Fold09
0.898	0.551	0.946	0.1	0.00037499130337396	Fold10
0.892	0.544	0.941	0.1	0.00037499130337396	Fold01
0.007	0.018	0.004	0.1	0.00037499130337396	Fold01

4 Conclusion