

MATH 324 Final Project

Nathan Howland, Madilyn Webb, Ellise Putnam, Addison Hart

1. Project Description

This experiment is using a data set of the weather patterns across various cities in Australia, which contains variables such as the humidity, pressure, cloud coverage, temperature, and much more. This experiment was conducted as an observational study, where the goal of the study was to predict the possibility of rain based on these variables. We wanted to determine: 1) Whether additional weather variables led to a sufficiently improved model for predicting rainfall in Australia 2) Whether there are different variables that conclude better predictions based on the city within Australia

Should be in paragraph form (not bullets). Written formally for a 3rd party such as public policymakers, or an external review board. As part of your description, consider the following:

- *Type of study [Designed experiment, Observational study]*
- *Description of the dataset*
- *Objective of study*

1.1 Research Questions

What are the overarching research questions that the study is targeting? (not just a restatement of the original research questions)

Research Question 1: What model should we use to predict rainfall across Australia? We want to discover which variables are the most reliable predictors of the rainfall for the “Next day”.

Research Question 2: Should we use different models for different locations? Are there models that are better suited for different locations across Australia, and what variables are the best for these locations?

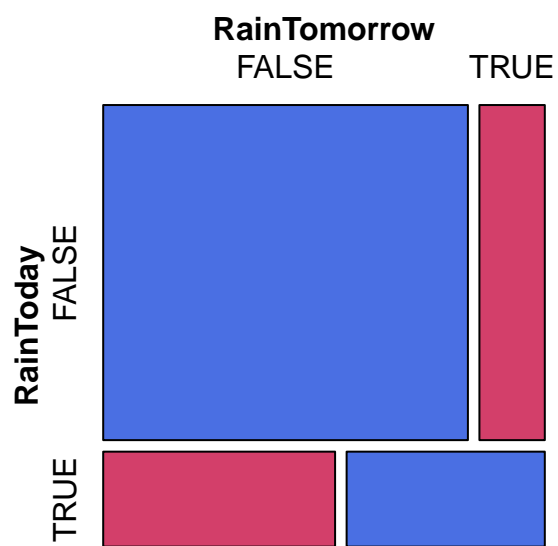
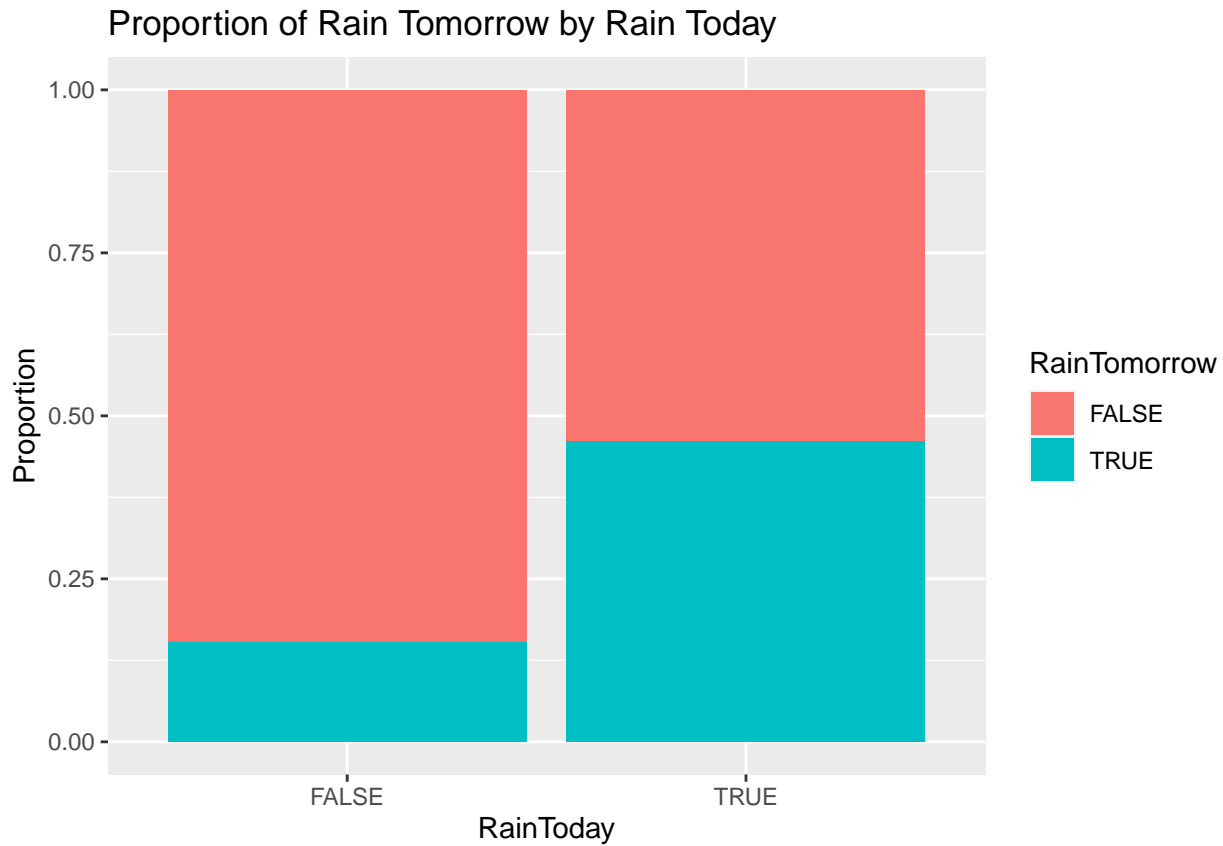
1.2 Variables

What is (are) possible explanatory and response variables?

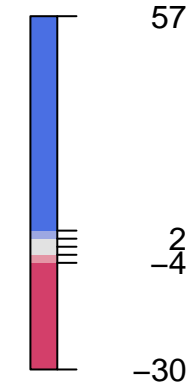
Create a table here - it should include variable names and types

2. Detailed Exploratory Data Analysis (EDA)

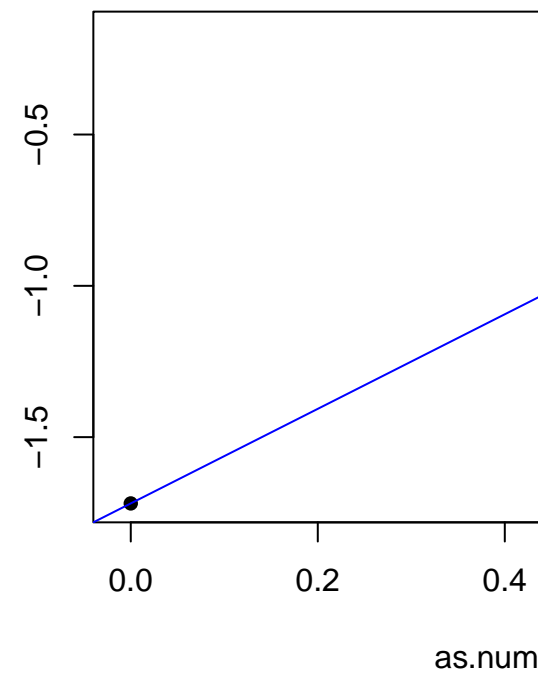
Question 1: Simple logistic regression model using RainToday as a predictor



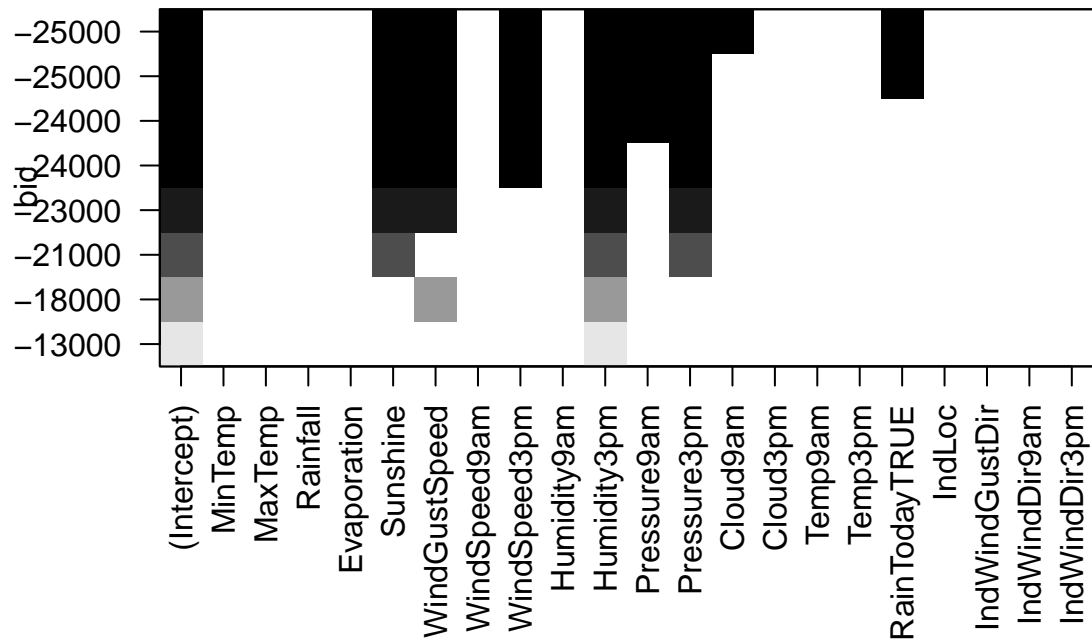
Pearson residuals:



p-value = $< 2.22e-16$



Question 2: Best model (using BIC) of all locations



```
##               model p   rsq  rss adjr2    cp    bic stderr
## 1                H3 2 0.207 7681 0.207 14165 -13089  0.369
## 2              WG-H3 3 0.271 7061 0.271  8476 -17821  0.354
## 3            S-H3-P3 4 0.312 6662 0.312  4808 -21095  0.344
## 4          S-WG-H3-P3 5 0.335 6444 0.335  2807 -22961  0.338
## 5        S-WG-WS3-H3-P3 6 0.345 6348 0.345  1928 -23795  0.335
## 6      S-WG-WS3-H3-P9-P3 7 0.351 6293 0.350  1422 -24278  0.334
## 7    S-WG-WS3-H3-P9-P3-RT 8 0.358 6225 0.358   797 -24882  0.332
## 8 S-WG-WS3-H3-P9-P3-C9-RT 9 0.359 6212 0.359   686 -24982  0.332
```

##

Model variables with abbreviations

##

H3

WG-H3

S-H3-P3

S-WG-H3-P3

S-WG-WS3-H3-P3

S-WG-WS3-H3-P9-P3

S-WG-WS3-H3-P9-P3-RT

S-WG-WS3-H3-P9-P3-C9-RT Sunshine-WindGustSpeed-WindSpeed3pm-Humidity3pm-Pressure9am-Pressure3pm-Cloud

##

model with largest adjr2

8

##

Number of observations

56420

##

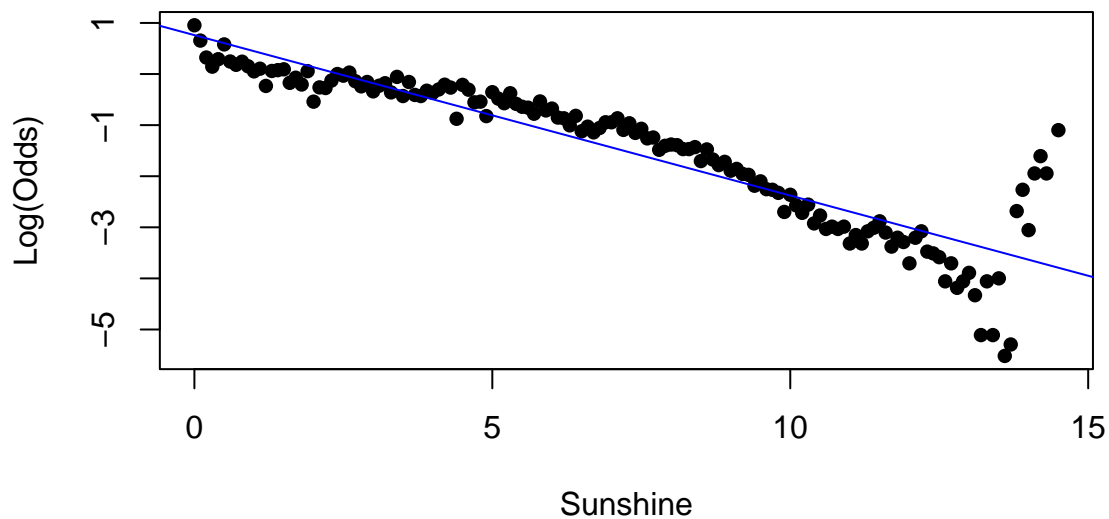
Call:

glm(formula = RainTomorrow ~ Sunshine + WindGustSpeed + WindSpeed3pm +

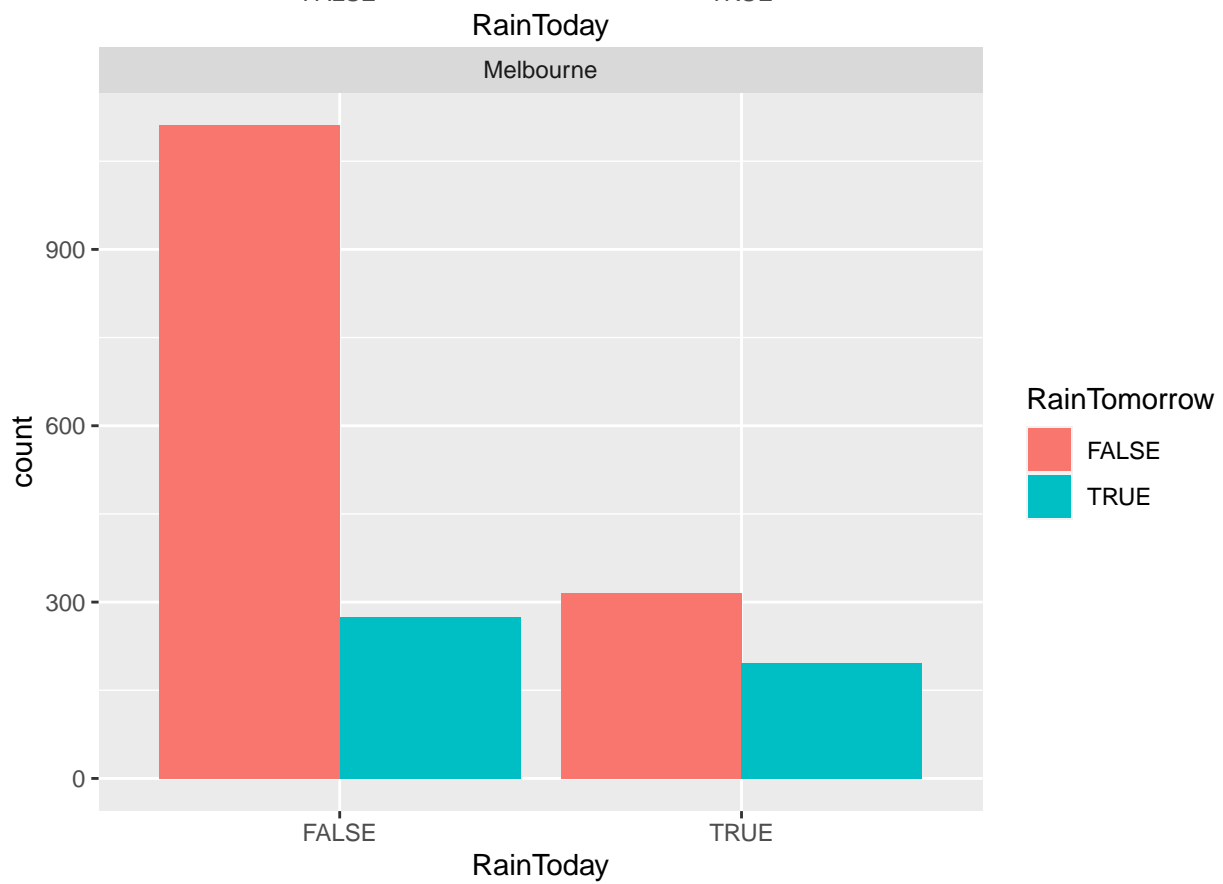
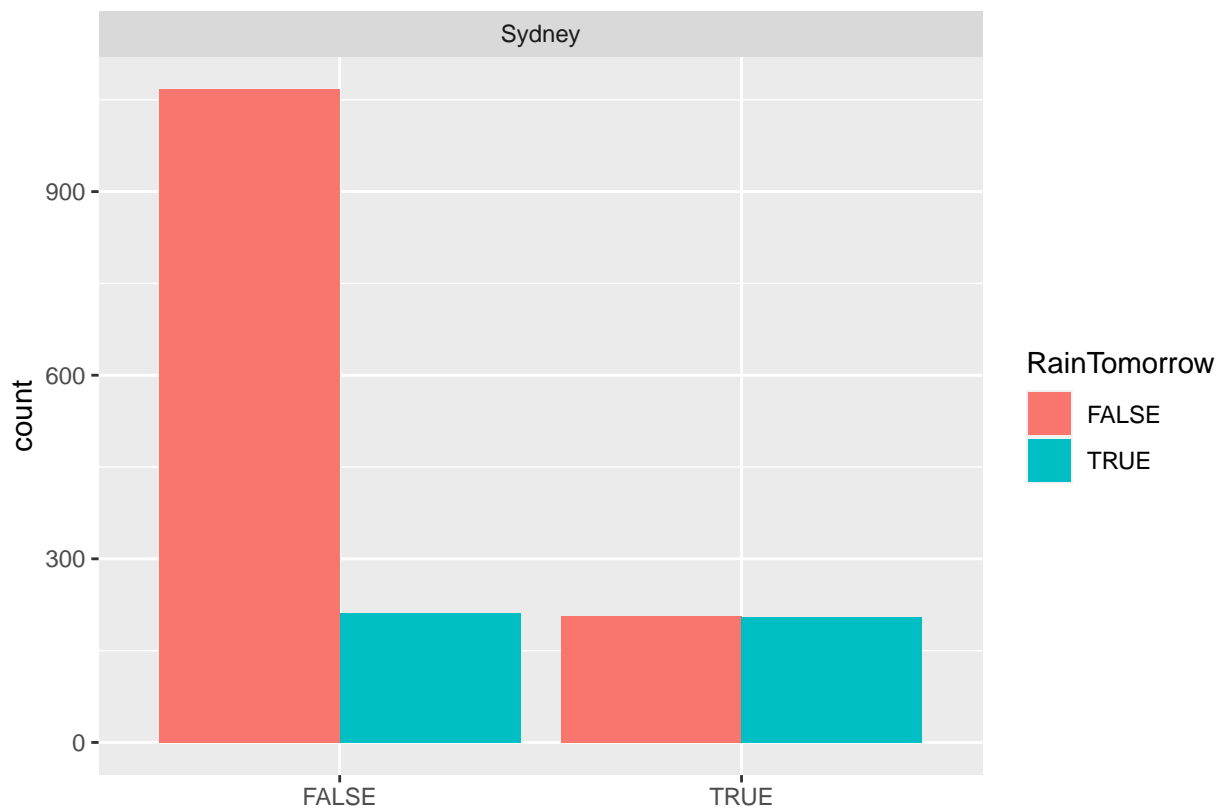
Humidity3pm + Pressure3pm + RainToday, family = "binomial",

data = clean_weather)

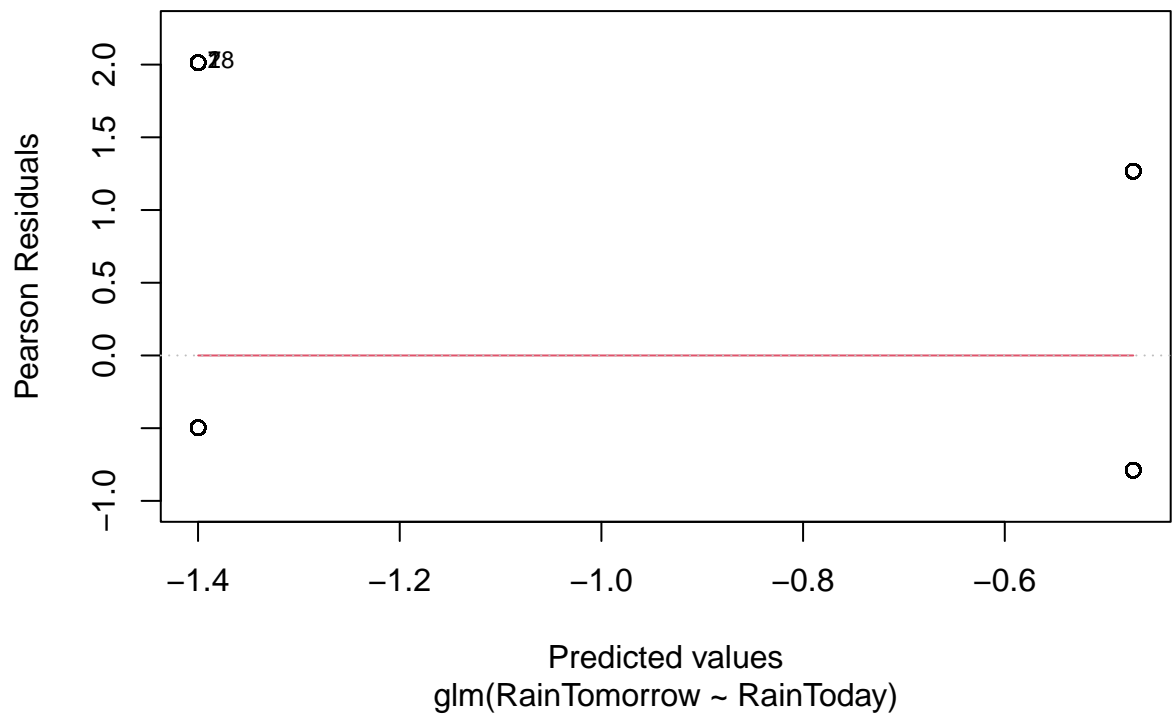
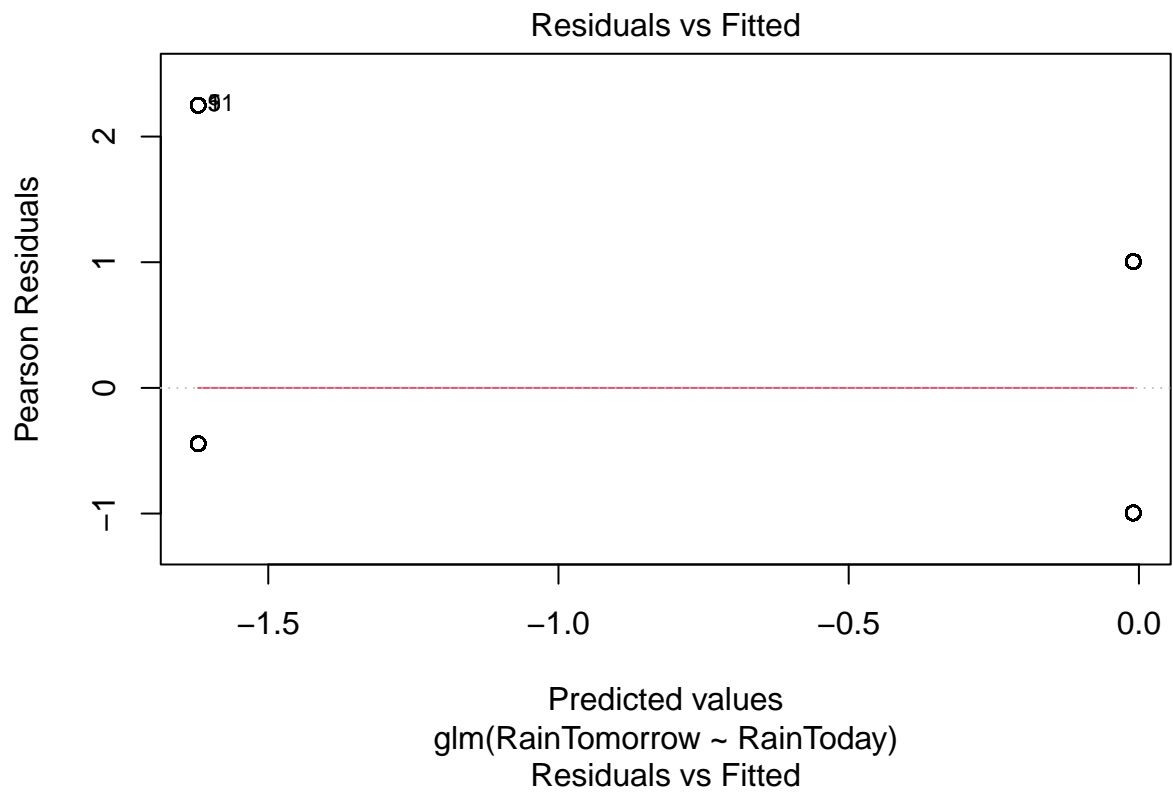
```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  74.4980470   2.0850771   35.73  <2e-16 ***
## Sunshine    -0.1679587   0.0041962  -40.03  <2e-16 ***
## WindGustSpeed 0.0508944   0.0013605   37.41  <2e-16 ***
## WindSpeed3pm -0.0330498   0.0020700  -15.97  <2e-16 ***
## Humidity3pm   0.0533006   0.0009395   56.73  <2e-16 ***
## Pressure3pm  -0.0781103   0.0020437  -38.22  <2e-16 ***
## RainTodayTRUE 0.3888131   0.0286052   13.59  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 59493  on 56419  degrees of freedom
## Residual deviance: 38526  on 56413  degrees of freedom
## AIC: 38540
##
## Number of Fisher Scoring iterations: 5
```

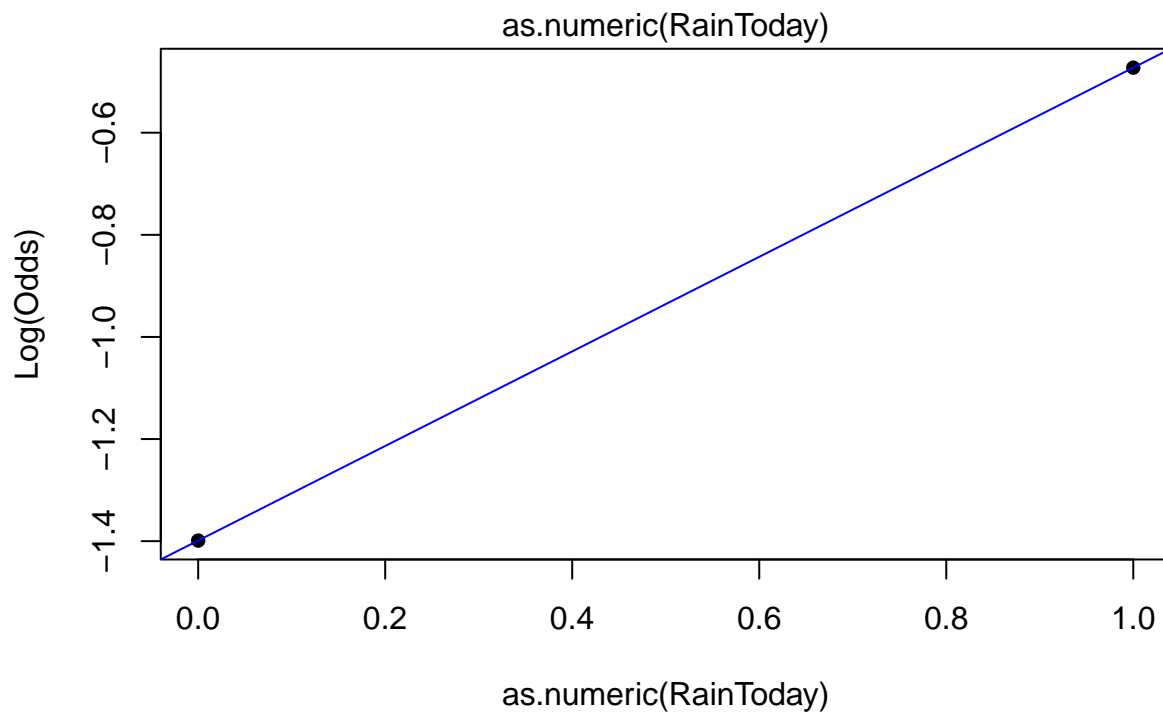
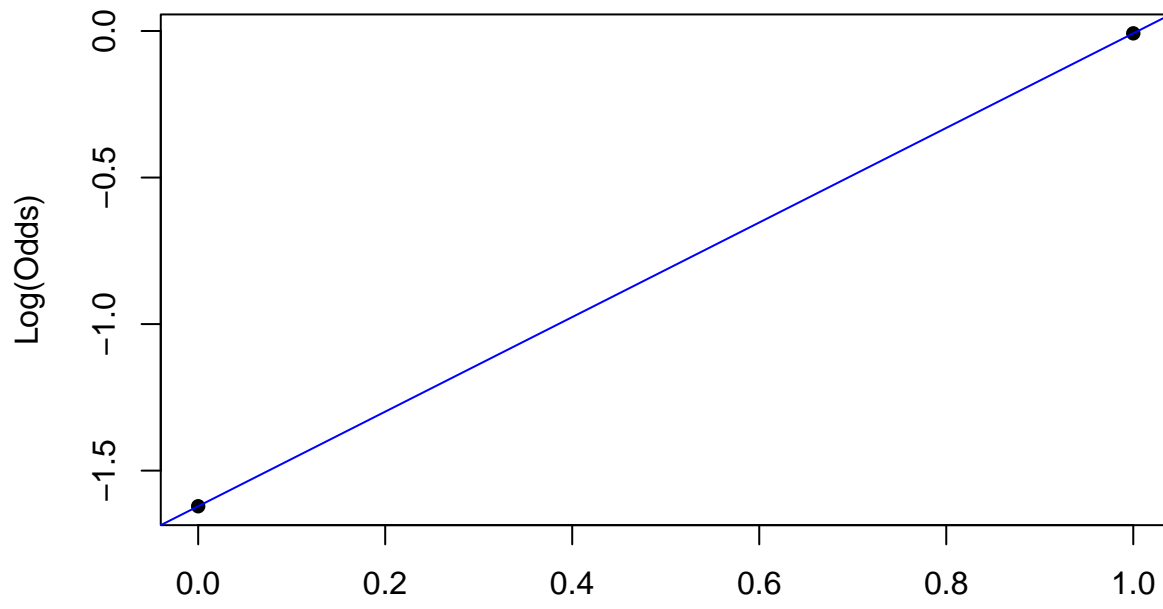


Question 3: Simple Logistic regression model dependent on the location



```
##
## Call:
## glm(formula = RainTomorrow ~ RainToday, family = binomial(link = "logit"),
##      data = SydneyWeather)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.62075    0.07534  -21.51  <2e-16 ***
## RainTodayTRUE  1.61104    0.12404   12.99  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1886.3  on 1689  degrees of freedom
## Residual deviance: 1716.3  on 1688  degrees of freedom
## AIC: 1720.3
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = RainTomorrow ~ RainToday, family = binomial(link = "logit"),
##      data = MelbourneWeather)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.39989    0.06745  -20.75  < 2e-16 ***
## RainTodayTRUE  0.92735    0.11309    8.20 2.41e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2126.9  on 1897  degrees of freedom
## Residual deviance: 2061.1  on 1896  degrees of freedom
## AIC: 2065.1
##
## Number of Fisher Scoring iterations: 4
```

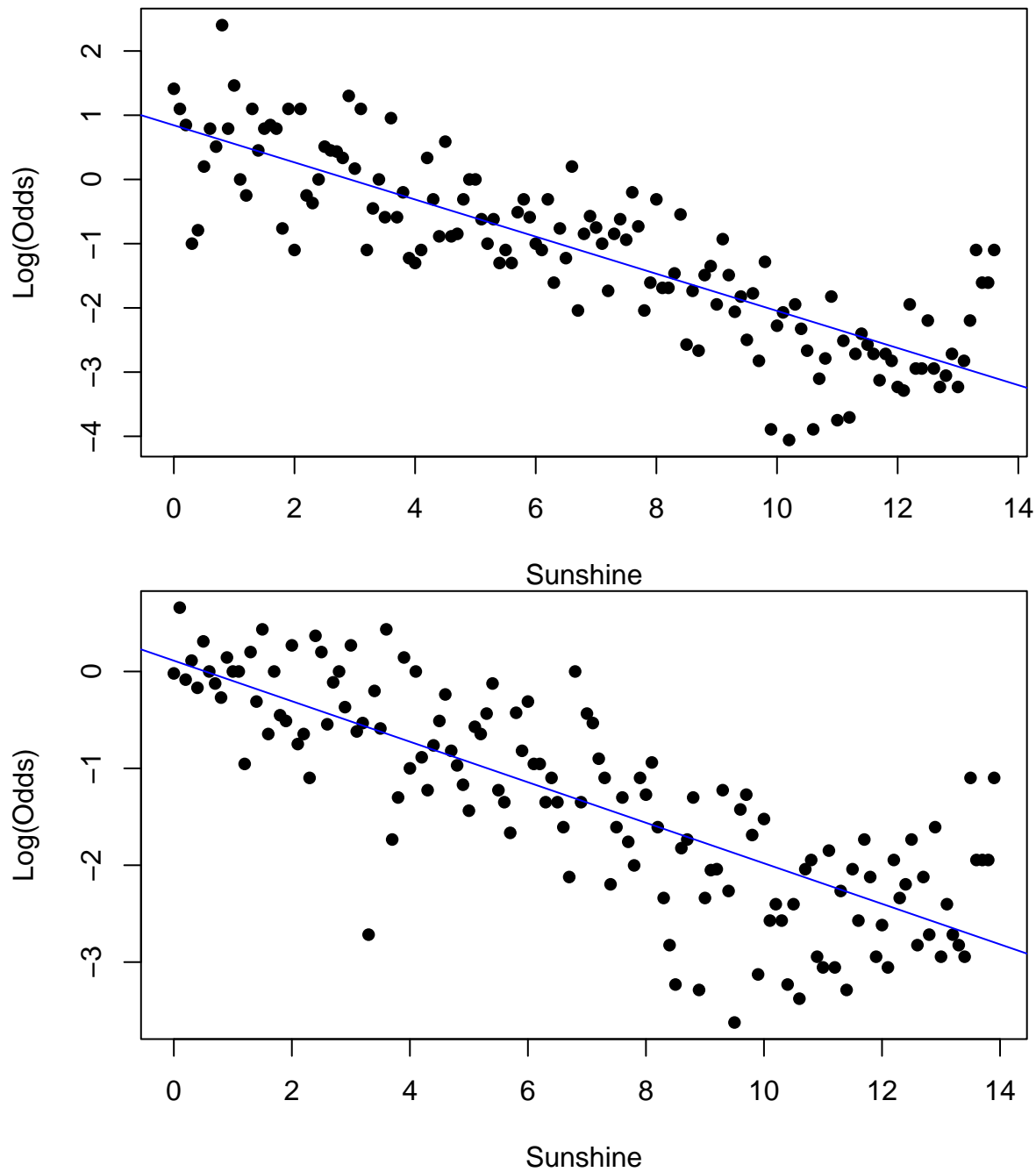




Question 4: Best model (using BIC) of two locations

##	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am
## 1	12.9	20.3	0.2	3.0	10.9	37	11
## 2	13.3	21.5	0.0	6.6	11.0	41	11
## 3	15.3	23.0	0.0	5.6	11.0	41	6
## 4	12.9	26.7	0.2	3.8	12.1	33	13
## 5	14.8	23.8	0.0	6.8	9.6	54	13
## 6	14.5	22.1	0.2	6.2	1.8	31	11
## 7	18.1	20.4	5.2	6.6	0.1	48	22
## 8	12.0	23.9	0.2	4.0	11.9	44	13

## 9	14.4	18.1	0.4	5.4	2.8	48	15
## 10	13.4	19.2	8.4	8.4	2.7	46	17
##	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	
## 1	26	70	57	1028.8	1025.6	3	
## 2	28	75	58	1025.9	1022.4	2	
## 3	19	70	63	1021.4	1017.8	1	
## 4	24	73	56	1018.0	1015.0	1	
## 5	26	76	69	1016.0	1014.7	2	
## 6	15	82	59	1020.7	1017.5	7	
## 7	28	78	77	1014.7	1012.1	8	
## 8	24	65	54	1018.1	1016.3	1	
## 9	20	83	68	1022.3	1020.5	7	
## 10	30	67	53	1023.9	1022.1	7	
##	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow	WindGustDir	WindDir9am
## 1	1	16.9	19.8	FALSE	FALSE	ENE	W
## 2	5	17.6	21.3	FALSE	FALSE	ENE	W
## 3	4	19.0	22.2	FALSE	FALSE	NNE	W
## 4	5	17.8	22.5	FALSE	FALSE	NE	W
## 5	7	20.2	20.6	FALSE	TRUE	SSE	SSE
## 6	7	16.2	20.5	FALSE	FALSE	ENE	WNW
## 7	8	18.7	18.6	TRUE	TRUE	SSE	S
## 8	3	17.6	20.7	FALSE	FALSE	SE	W
## 9	7	14.8	17.6	FALSE	TRUE	ESE	S
## 10	7	15.2	17.8	TRUE	TRUE	SE	SE
##	WindDir3pm	Location					
## 1	E	Sydney					
## 2	ENE	Sydney					
## 3	ENE	Sydney					
## 4	ENE	Sydney					
## 5	SE	Sydney					
## 6	ENE	Sydney					
## 7	SSE	Sydney					
## 8	ESE	Sydney					
## 9	SE	Sydney					
## 10	SSE	Sydney					



A BIC plot is made of both data from Sydney and data from Melbourne to derive the explanatory variables that maximize the BIC score, which we then use to create a simple logistic model. A linearity condition assesment using the logarithm of the odds is also employed.

Spend time investigating the data thoroughly and use this section to present figures that allow you to visualize the relationship of the response to different predictors. Include descriptions of the figures that illustrate what you can learn from them.

3. Statistical Analysis

Research Question 1

Describe the statistical analysis that you used to answer both research questions. Each analysis should be labeled for the appropriate research question. This section should summarize all relevant analyses that lead to your final conclusions/decisions/recommendations. Be sure to include:

- *Model Assumptions & how checked/verified*
- *Interpretation of relevant estimates/statistics/p-values IN CONTEXT*

Question 1: Simple logistic regression model using RainToday as a predictor

Question 2: Best model (using BIC) of all locations

Question 3: Simple Logistic regression model dependent on the location

Question 4: Best model (using BIC) of two locations

```
## [1] "Brier score for Sydney: 0.117757762827891"
```

```
## [1] "Brier score for Melbourne: 0.138845057528619"
```

You should not include your code. However, I would like to see the code for final models that you are using to answer the research question in the appendix.

4. Conclusions

Succinct response to each question laid out in 1.1. This a much shorter version of section 3, and focuses on conclusions rather than the analyses.

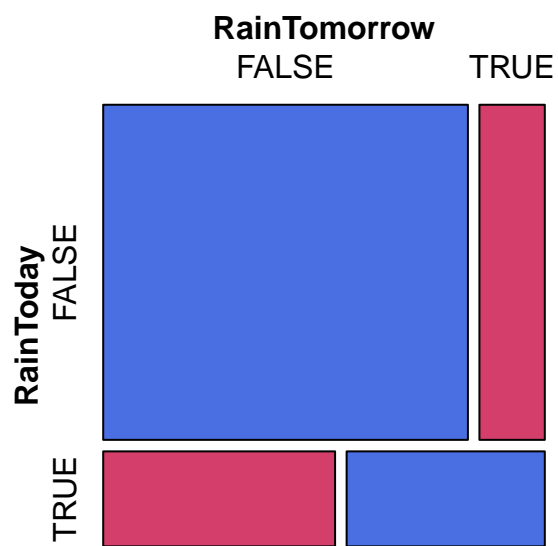
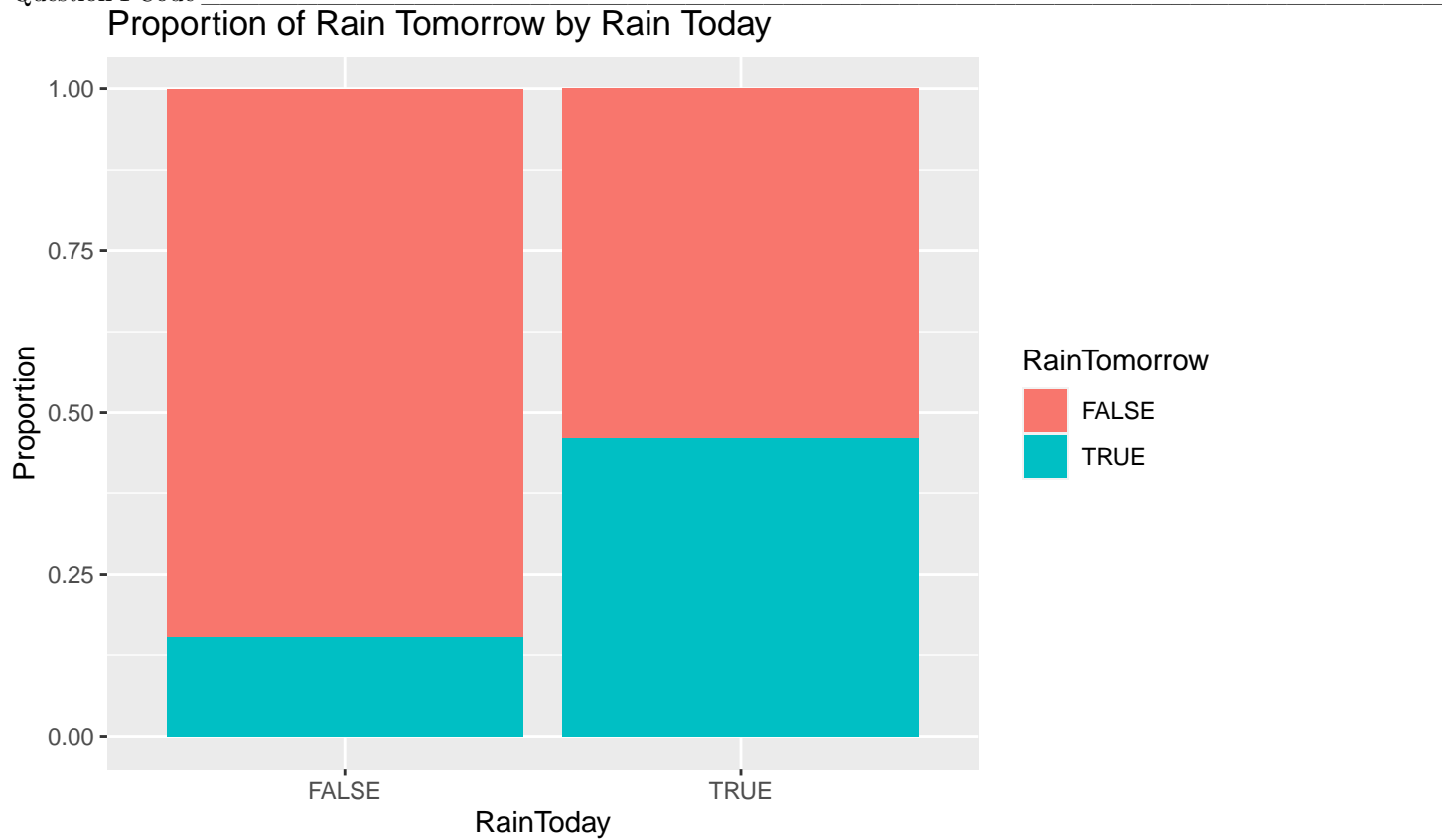
Research Question 1:

Research Question 2:

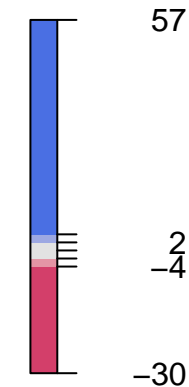
(all research questions in 1.1 should be addressed, in the same order)

5. Appendix

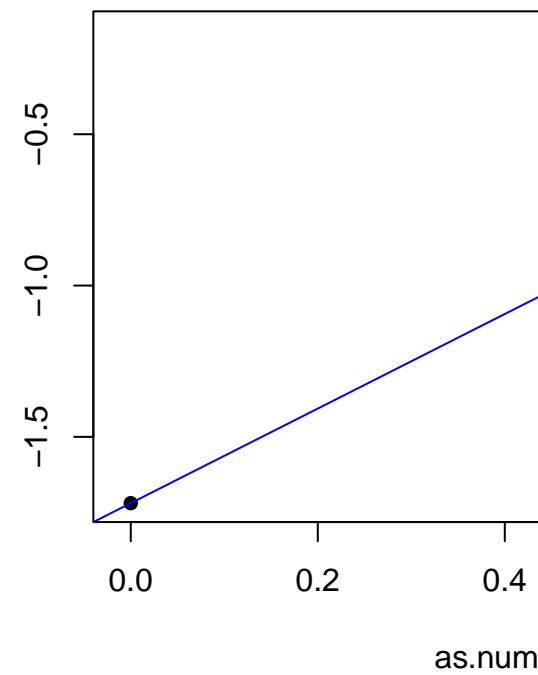
Question 1 Code



Pearson residuals:



p-value = $< 2.22e-16$



Question 2 Code

```
## Start: AIC=-99395.67
## RainTomorrow ~ 1
```

```

##
##          Df Sum of Sq    RSS    AIC
## + Humidity3pm      1  2009.19 7680.6 -112504
## + Sunshine         1  1992.02 7697.8 -112378
## + Cloud3pm         1  1463.07 8226.8 -108629
## + Cloud9am         1  1017.03 8672.8 -105650
## + RainToday        1   925.78 8764.1 -105059
## + Humidity9am       1   711.80 8978.0 -103698
## + Pressure9am      1   629.17 9060.7 -103181
## + Rainfall         1   626.83 9063.0 -103167
## + WindGustSpeed     1   526.77 9163.1 -102547
## + Pressure3pm      1   514.46 9175.4 -102472
## + Temp3pm          1   326.59 9363.3 -101328
## + Location         25   283.03 9406.8 -101018
## + MaxTemp          1   210.72 9479.1 -100634
## + WindDir9am       15   169.67 9520.2 -100362
## + Evaporation       1   163.76 9526.1 -100355
## + WindGustDir      15   128.04 9561.8 -100116
## + WindDir3pm       15   117.98 9571.9 -100057
## + WindSpeed3pm     1    76.51 9613.3  -99841
## + MinTemp          1    74.07 9615.8  -99827
## + WindSpeed9am     1    68.22 9621.6  -99792
## + IndWindDir3pm    1    26.75 9663.1  -99550
## + IndWindGustDir   1    15.26 9674.6  -99483
## + IndWindDir9am    1     9.94 9679.9  -99452
## + Temp9am          1     3.20 9686.6  -99412
## + IndLoc           1     1.17 9688.7  -99400
## <none>              9689.8  -99396
##
## Step:  AIC=-112504.2
## RainTomorrow ~ Humidity3pm
##
##          Df Sum of Sq    RSS    AIC
## + WindGustSpeed     1    619.31 7061.3 -117245
## + Pressure3pm       1    564.97 7115.7 -116813
## + Pressure9am       1    496.57 7184.1 -116273
## + Sunshine          1    446.63 7234.0 -115882
## + Cloud3pm          1    318.69 7362.0 -114893
## + RainToday         1    203.09 7477.6 -114014
## + WindDir3pm        15    206.58 7474.1 -114012
## + Rainfall          1    171.81 7508.8 -113779
## + WindDir9am        15    138.79 7541.9 -113503
## + Location          25    139.76 7540.9 -113490
## + WindSpeed9am      1    118.76 7561.9 -113381
## + WindGustDir       15    122.32 7558.3 -113380
## + Cloud9am          1    110.95 7569.7 -113323
## + WindSpeed3pm      1     53.64 7627.0 -112898
## + Evaporation       1     45.86 7634.8 -112840
## + MaxTemp           1     38.82 7641.8 -112788
## + IndWindDir3pm     1     32.51 7648.1 -112742
## + Humidity9am       1     31.05 7649.6 -112731
## + MinTemp           1     29.65 7651.0 -112720
## + Temp9am           1     25.65 7655.0 -112691
## + Temp3pm           1     23.62 7657.0 -112676

```

```

## + IndWindGustDir 1      7.60 7673.1 -112558
## <none>              7680.6 -112504
## + IndWindDir9am 1      0.13 7680.5 -112503
## + IndLoc          1      0.06 7680.6 -112503
##
## Step: AIC=-117245.4
## RainTomorrow ~ Humidity3pm + WindGustSpeed
##
##           Df Sum of Sq  RSS    AIC
## + Sunshine      1    349.35 6712.0 -120106
## + Pressure3pm    1    237.50 6823.8 -119174
## + Location      25    214.49 6846.8 -118936
## + Cloud3pm       1    185.67 6875.7 -118747
## + WindSpeed3pm   1    181.67 6879.7 -118714
## + Pressure9am    1    162.29 6899.0 -118555
## + Rainfall       1    102.44 6958.9 -118068
## + WindDir3pm     15    104.87 6956.5 -118060
## + RainToday      1    100.31 6961.0 -118051
## + WindDir9am     15     89.78 6971.6 -117937
## + Cloud9am       1     54.98 7006.4 -117684
## + WindGustDir    15     46.15 7015.2 -117585
## + MaxTemp        1     33.34 7028.0 -117510
## + Temp3pm        1     30.05 7031.3 -117484
## + WindSpeed9am   1     28.40 7032.9 -117471
## + IndWindDir3pm  1     14.38 7047.0 -117358
## + Temp9am        1      9.50 7051.8 -117319
## + MinTemp        1      5.95 7055.4 -117291
## + Evaporation    1      2.38 7059.0 -117262
## + IndWindDir9am  1      1.17 7060.2 -117253
## + IndWindGustDir 1      0.33 7061.0 -117246
## + IndLoc         1      0.27 7061.1 -117246
## <none>              7061.3 -117245
## + Humidity9am    1      0.00 7061.3 -117243
##
## Step: AIC=-120106.1
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine
##
##           Df Sum of Sq  RSS    AIC
## + Pressure3pm    1    267.977 6444.0 -122403
## + Pressure9am    1    184.590 6527.4 -121677
## + Location      25    180.222 6531.8 -121592
## + MaxTemp        1    123.840 6588.2 -121155
## + Temp3pm        1    112.686 6599.3 -121059
## + WindSpeed3pm   1    104.001 6608.0 -120985
## + Rainfall       1     73.485 6638.5 -120725
## + Temp9am        1     68.528 6643.5 -120683
## + RainToday      1     65.701 6646.3 -120659
## + WindDir3pm     15     68.688 6643.3 -120656
## + WindDir9am     15     67.498 6644.5 -120646
## + MinTemp        1     32.535 6679.5 -120378
## + WindSpeed9am   1     29.215 6682.8 -120350
## + Evaporation    1     23.559 6688.4 -120302
## + WindGustDir    15     26.711 6685.3 -120301
## + Cloud3pm       1     14.066 6697.9 -120222

```

```

## + IndWindDir3pm 1 11.564 6700.4 -120201
## + Cloud9am 1 8.419 6703.6 -120175
## + Humidity9am 1 7.705 6704.3 -120169
## + IndWindDir9am 1 1.411 6710.6 -120116
## <none> 6712.0 -120106
## + IndLoc 1 0.178 6711.8 -120106
## + IndWindGustDir 1 0.001 6712.0 -120104
##
## Step: AIC=-122402.9
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm
##
## Df Sum of Sq RSS AIC
## + Location 25 135.482 6308.5 -123552
## + WindSpeed3pm 1 95.822 6348.2 -123246
## + Pressure9am 1 58.049 6386.0 -122911
## + RainToday 1 49.451 6394.6 -122836
## + Rainfall 1 44.948 6399.1 -122796
## + WindDir9am 15 46.455 6397.6 -122781
## + Cloud9am 1 22.228 6421.8 -122596
## + WindDir3pm 15 24.552 6419.5 -122588
## + WindSpeed9am 1 13.501 6430.5 -122519
## + MaxTemp 1 8.072 6435.9 -122472
## + MinTemp 1 7.330 6436.7 -122465
## + Temp3pm 1 5.497 6438.5 -122449
## + Cloud3pm 1 3.108 6440.9 -122428
## + WindGustDir 15 6.267 6437.7 -122428
## + IndWindDir3pm 1 2.845 6441.2 -122426
## + IndWindDir9am 1 0.496 6443.5 -122405
## + Humidity9am 1 0.289 6443.7 -122403
## <none> 6444.0 -122403
## + IndWindGustDir 1 0.112 6443.9 -122402
## + IndLoc 1 0.061 6444.0 -122401
## + Temp9am 1 0.059 6444.0 -122401
## + Evaporation 1 0.005 6444.0 -122401
##
## Step: AIC=-123551.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
## Location
##
## Df Sum of Sq RSS AIC
## + Pressure9am 1 66.620 6241.9 -124149
## + WindSpeed3pm 1 60.818 6247.7 -124096
## + RainToday 1 35.876 6272.7 -123871
## + WindDir9am 15 38.657 6269.9 -123869
## + WindDir3pm 15 38.299 6270.2 -123865
## + Rainfall 1 32.619 6275.9 -123842
## + Cloud9am 1 15.692 6292.8 -123690
## + WindGustDir 15 13.781 6294.8 -123645
## + MaxTemp 1 8.684 6299.9 -123627
## + Temp3pm 1 7.377 6301.2 -123616
## + WindSpeed9am 1 6.692 6301.8 -123610
## + MinTemp 1 6.508 6302.0 -123608
## + Cloud3pm 1 2.905 6305.6 -123576
## + IndWindDir3pm 1 1.770 6306.8 -123566

```

```

## + IndWindGustDir 1      0.815 6307.7 -123557
## + Evaporation    1      0.471 6308.1 -123554
## + IndWindDir9am  1      0.421 6308.1 -123553
## <none>           6308.5 -123552
## + Temp9am        1      0.029 6308.5 -123550
## + Humidity9am     1      0.007 6308.5 -123550
##
## Step: AIC=-124148.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am
##
##              Df Sum of Sq    RSS    AIC
## + RainToday      1    60.331 6181.6 -124695
## + WindSpeed3pm    1    53.119 6188.8 -124629
## + Rainfall        1    43.801 6198.1 -124544
## + WindDir9am      15    16.836 6225.1 -124271
## + WindDir3pm      15    15.905 6226.0 -124263
## + Cloud9am        1    10.234 6231.7 -124239
## + MaxTemp         1     5.530 6236.4 -124197
## + Cloud3pm        1     4.452 6237.5 -124187
## + Temp3pm         1     3.526 6238.4 -124179
## + WindSpeed9am    1     3.131 6238.8 -124175
## + WindGustDir     15     5.398 6236.5 -124168
## + IndWindDir3pm   1     2.009 6239.9 -124165
## + MinTemp         1     1.524 6240.4 -124160
## + Temp9am         1     0.297 6241.6 -124149
## + Humidity9am     1     0.251 6241.7 -124149
## <none>           6241.9 -124149
## + IndWindDir9am   1     0.195 6241.7 -124148
## + Evaporation     1     0.130 6241.8 -124148
## + IndWindGustDir  1     0.001 6241.9 -124147
##
## Step: AIC=-124694.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday
##
##              Df Sum of Sq    RSS    AIC
## + WindSpeed3pm    1    49.361 6132.2 -125145
## + WindDir9am      15    19.512 6162.1 -124843
## + Cloud9am        1    12.441 6169.1 -124806
## + WindDir3pm      15    15.156 6166.4 -124803
## + Rainfall        1    11.261 6170.3 -124796
## + MaxTemp         1     9.326 6172.3 -124778
## + Humidity9am     1     6.576 6175.0 -124753
## + Temp3pm         1     5.203 6176.4 -124740
## + Cloud3pm        1     4.943 6176.6 -124738
## + WindSpeed9am    1     3.304 6178.3 -124723
## + Evaporation     1     2.468 6179.1 -124715
## + Temp9am         1     2.086 6179.5 -124712
## + IndWindDir3pm   1     1.699 6179.9 -124708
## + WindGustDir     15     4.663 6176.9 -124707
## + MinTemp         1     0.815 6180.8 -124700
## <none>           6181.6 -124695
## + IndWindGustDir  1     0.051 6181.5 -124693

```



```

## + IndWindDir9am    1      0.046 6181.5 -124693
##
## Step: AIC=-125145
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm
##
##              Df Sum of Sq    RSS    AIC
## + WindDir9am    15   16.6619 6115.6 -125269
## + WindDir3pm    15   13.8421 6118.4 -125243
## + Rainfall       1    9.5600 6122.7 -125231
## + Cloud9am       1    9.2872 6122.9 -125229
## + Humidity9am     1    5.3981 6126.8 -125193
## + MaxTemp        1    5.2164 6127.0 -125191
## + Cloud3pm       1    4.4558 6127.8 -125184
## + Temp3pm        1    2.3985 6129.8 -125165
## + IndWindDir3pm   1    2.1731 6130.0 -125163
## + WindGustDir    15    5.1208 6127.1 -125162
## + Evaporation     1    1.5629 6130.7 -125157
## + WindSpeed9am    1    1.4958 6130.7 -125157
## + MinTemp        1    1.4650 6130.8 -125156
## + Temp9am        1    1.2892 6130.9 -125155
## <none>                                6132.2 -125145
## + IndWindDir9am   1    0.0895 6132.1 -125144
## + IndWindGustDir  1    0.0168 6132.2 -125143
##
## Step: AIC=-125268.5
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am
##
##              Df Sum of Sq    RSS    AIC
## + Rainfall       1   10.1685 6105.4 -125360
## + WindDir3pm     15   12.8997 6102.7 -125358
## + Cloud9am       1    7.6503 6107.9 -125337
## + WindGustDir    15   10.6100 6105.0 -125336
## + Humidity9am     1    4.6892 6110.9 -125310
## + MaxTemp        1    4.2245 6111.3 -125306
## + Cloud3pm       1    3.7309 6111.8 -125301
## + Evaporation     1    2.3403 6113.2 -125288
## + Temp3pm        1    1.7246 6113.8 -125282
## + IndWindDir3pm   1    1.7220 6113.8 -125282
## + MinTemp        1    1.4533 6114.1 -125280
## + WindSpeed9am    1    1.4383 6114.1 -125280
## + Temp9am        1    0.7472 6114.8 -125273
## <none>                                6115.6 -125269
## + IndWindGustDir  1    0.0835 6115.5 -125267
##
## Step: AIC=-125360.4
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall
##
##              Df Sum of Sq    RSS    AIC
## + WindDir3pm     15   12.6525 6092.7 -125447
## + Cloud9am       1    7.8906 6097.5 -125431

```

```

## + WindGustDir      15    10.2931 6095.1 -125426
## + Humidity9am       1     5.7841 6099.6 -125412
## + Cloud3pm          1     3.9624 6101.4 -125395
## + MaxTemp           1     3.8364 6101.6 -125394
## + Evaporation        1     1.9763 6103.4 -125377
## + IndWindDir3pm     1     1.8152 6103.6 -125375
## + MinTemp           1     1.4575 6103.9 -125372
## + WindSpeed9am      1     1.4381 6104.0 -125372
## + Temp3pm           1     1.3807 6104.0 -125371
## + Temp9am           1     0.7539 6104.6 -125365
## <none>                                6105.4 -125360
## + IndWindGustDir    1     0.0744 6105.3 -125359
##
## Step: AIC=-125447.4
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm
##
##              Df Sum of Sq  RSS    AIC
## + Cloud9am      1    7.6139 6085.1 -125516
## + Humidity9am    1    6.2602 6086.5 -125503
## + MaxTemp        1    4.3936 6088.3 -125486
## + Cloud3pm       1    3.7942 6088.9 -125481
## + WindGustDir    15    6.1975 6086.5 -125475
## + Evaporation    1    2.5420 6090.2 -125469
## + Temp3pm        1    1.6645 6091.1 -125461
## + WindSpeed9am   1    1.6270 6091.1 -125461
## + MinTemp        1    1.1721 6091.6 -125456
## + Temp9am        1    1.1289 6091.6 -125456
## <none>                                6092.7 -125447
## + IndWindGustDir 1    0.1379 6092.6 -125447
##
## Step: AIC=-125516
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am
##
##              Df Sum of Sq  RSS    AIC
## + Cloud3pm       1    6.7462 6078.4 -125577
## + MaxTemp         1    6.1858 6078.9 -125571
## + Humidity9am     1    4.8721 6080.3 -125559
## + Evaporation     1    3.4365 6081.7 -125546
## + WindGustDir     15    6.4358 6078.7 -125546
## + Temp3pm         1    3.0017 6082.1 -125542
## + Temp9am         1    2.0987 6083.0 -125533
## + WindSpeed9am    1    1.7349 6083.4 -125530
## <none>                                6085.1 -125516
## + IndWindGustDir 1    0.1999 6084.9 -125516
## + MinTemp         1    0.1969 6084.9 -125516
##
## Step: AIC=-125576.6
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm

```

```

##
##          Df Sum of Sq    RSS    AIC
## + MaxTemp      1    4.6530 6073.7 -125618
## + Humidity9am   1    3.7592 6074.6 -125609
## + WindGustDir   15    6.3038 6072.1 -125605
## + Evaporation   1    2.8164 6075.6 -125601
## + Temp3pm       1    2.1584 6076.2 -125595
## + WindSpeed9am  1    1.8292 6076.6 -125592
## + Temp9am       1    1.1622 6077.2 -125585
## + MinTemp       1    0.4569 6077.9 -125579
## <none>                                6078.4 -125577
## + IndWindGustDir 1    0.1736 6078.2 -125576
##
## Step: AIC=-125617.8
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp
##
##          Df Sum of Sq    RSS    AIC
## + MinTemp      1   17.7256 6056.0 -125781
## + Temp3pm       1    5.4918 6068.2 -125667
## + Temp9am       1    3.3477 6070.4 -125647
## + WindGustDir   15    6.3138 6067.4 -125646
## + Humidity9am   1    2.6059 6071.1 -125640
## + WindSpeed9am  1    1.4489 6072.3 -125629
## + Evaporation   1    0.5762 6073.2 -125621
## <none>                                6073.7 -125618
## + IndWindGustDir 1    0.0837 6073.6 -125617
##
## Step: AIC=-125780.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp
##
##          Df Sum of Sq    RSS    AIC
## + Humidity9am   1    3.1132 6052.9 -125808
## + WindGustDir   15    5.8149 6050.2 -125805
## + Evaporation   1    2.7729 6053.2 -125805
## + Temp3pm       1    1.5158 6054.5 -125793
## + Temp9am       1    0.3217 6055.7 -125782
## <none>                                6056.0 -125781
## + WindSpeed9am  1    0.2078 6055.8 -125781
## + IndWindGustDir 1    0.0881 6055.9 -125780
##
## Step: AIC=-125807.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp +
##      Humidity9am
##
##          Df Sum of Sq    RSS    AIC
## + WindGustDir   15    6.1778 6046.7 -125835
## + Evaporation   1    1.6168 6051.3 -125821
## + WindSpeed9am  1    0.7187 6052.2 -125812

```

```

## + Temp3pm          1      0.6317 6052.3 -125812
## <none>              6052.9 -125808
## + Temp9am          1      0.1993 6052.7 -125808
## + IndWindGustDir   1      0.0842 6052.8 -125806
##
## Step: AIC=-125835.3
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp +
##      Humidity9am + WindGustDir
##
##              Df Sum of Sq    RSS      AIC
## + Evaporation   1    1.53806 6045.2 -125848
## + Temp3pm        1     0.61583 6046.1 -125839
## + WindSpeed9am   1     0.47053 6046.2 -125838
## <none>           6046.7 -125835
## + Temp9am        1     0.15360 6046.6 -125835
##
## Step: AIC=-125847.7
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp +
##      Humidity9am + WindGustDir + Evaporation
##
##              Df Sum of Sq    RSS      AIC
## + WindSpeed9am   1     0.55665 6044.6 -125851
## + Temp3pm        1     0.52771 6044.6 -125851
## <none>           6045.2 -125848
## + Temp9am        1     0.10969 6045.1 -125847
##
## Step: AIC=-125850.9
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp +
##      Humidity9am + WindGustDir + Evaporation + WindSpeed9am
##
##              Df Sum of Sq    RSS      AIC
## + Temp3pm        1     0.49027 6044.1 -125853
## <none>           6044.6 -125851
## + Temp9am        1     0.20045 6044.4 -125851
##
## Step: AIC=-125853.4
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine + Pressure3pm +
##      Location + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +
##      Rainfall + WindDir3pm + Cloud9am + Cloud3pm + MaxTemp + MinTemp +
##      Humidity9am + WindGustDir + Evaporation + WindSpeed9am +
##      Temp3pm
##
##              Df Sum of Sq    RSS      AIC
## <none>           6044.1 -125853
## + Temp9am        1     0.09461 6044.0 -125852
##
## Call:

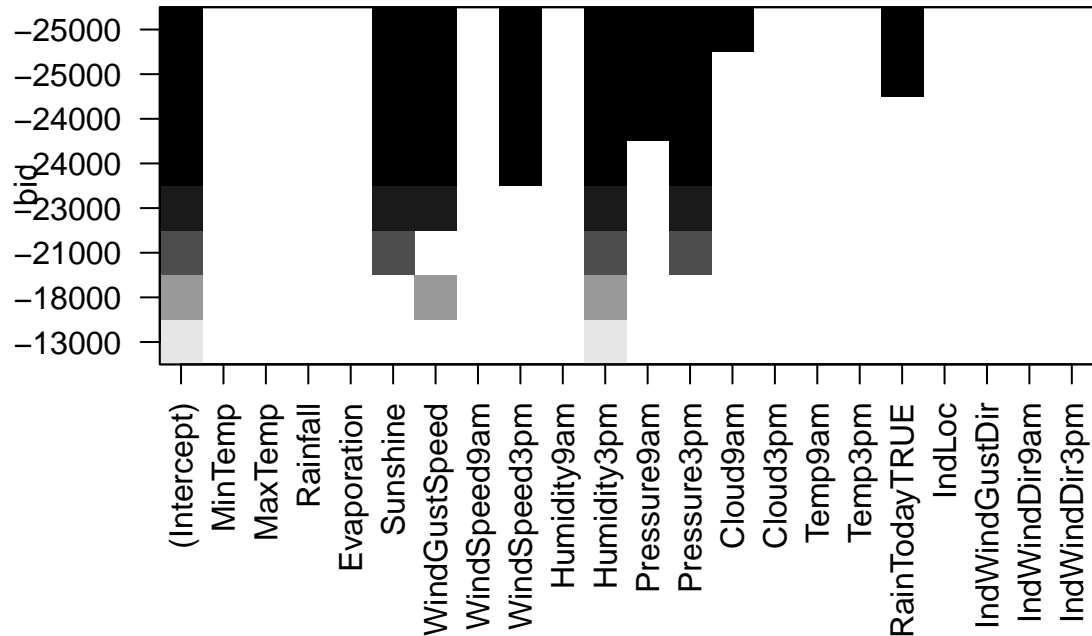
```

```

## lm(formula = RainTomorrow ~ Humidity3pm + WindGustSpeed + Sunshine +
##     Pressure3pm + Location + Pressure9am + RainToday + WindSpeed3pm +
##     WindDir9am + Rainfall + WindDir3pm + Cloud9am + Cloud3pm +
##     MaxTemp + MinTemp + Humidity9am + WindGustDir + Evaporation +
##     WindSpeed9am + Temp3pm, data = clean_weather)
##
## Coefficients:
##             (Intercept)                Humidity3pm                WindGustSpeed
##             8.2008485                0.0086312                0.0080625
##             Sunshine                Pressure3pm                LocationBrisbane
##             -0.0259849                -0.0264062                0.0149567
##             LocationCairns                LocationCanberra                LocationCobar
##             -0.0234509                -0.0398485                0.0212171
##             LocationCoffsHarbour                LocationDarwin                LocationHobart
##             -0.0083016                -0.0446166                -0.1074446
##             LocationMelbourne                LocationMelbourneAirport                LocationMildura
##             -0.0483958                -0.0829032                0.0141485
##             LocationMoree                LocationMountGambier                LocationNorfolkIsland
##             -0.0080463                -0.0186528                -0.0531244
##             LocationNuriootpa                LocationPerth                LocationPerthAirport
##             -0.0121691                0.0306939                0.0037245
##             LocationPortland                LocationSale                LocationSydney
##             0.0029331                -0.0955003                -0.0177373
##             LocationSydneyAirport                LocationTownsville                LocationWaggaWagga
##             -0.0387042                -0.0866687                0.0109036
##             LocationWatsonia                LocationWilliamtown                LocationWoomera
##             -0.0499719                -0.0339462                0.0105455
##             Pressure9am                RainTodayTRUE                WindSpeed3pm
##             0.0179909                0.0819036                -0.0043065
##             WindDir9amENE                WindDir9amESE                WindDir9amN
##             0.0215685                -0.0113344                0.0162846
##             WindDir9amNE                WindDir9amNNE                WindDir9amNNW
##             0.0262873                0.0389413                -0.0036627
##             WindDir9amNW                WindDir9amS                WindDir9amSE
##             -0.0012844                -0.0332563                -0.0213697
##             WindDir9amSSE                WindDir9amSSW                WindDir9amSW
##             -0.0274714                -0.0218372                -0.0042812
##             WindDir9amW                WindDir9amWNW                WindDir9amWSW
##             0.0096174                -0.0028188                0.0039351
##             Rainfall                WindDir3pmENE                WindDir3pmESE
##             0.0022702                -0.0043172                0.0035435
##             WindDir3pmN                WindDir3pmNE                WindDir3pmNNE
##             0.0201213                -0.0105967                0.0136750
##             WindDir3pmNNW                WindDir3pmNW                WindDir3pmS
##             0.0297067                0.0346413                0.0010783
##             WindDir3pmSE                WindDir3pmSSE                WindDir3pmSSW
##             0.0182143                0.0055815                -0.0059259
##             WindDir3pmSW                WindDir3pmW                WindDir3pmWNW
##             -0.0237499                0.0043942                0.0329164
##             WindDir3pmWSW                Cloud9am                Cloud3pm
##             -0.0144192                -0.0050430                0.0046128
##             MaxTemp                MinTemp                Humidity9am
##             0.0101766                -0.0076491                -0.0005756
##             WindGustDirENE                WindGustDirESE                WindGustDirN

```

```
##          -0.0096221          0.0185454          -0.0200001
##          WindGustDirNE          WindGustDirNNE          WindGustDirNNW
##          -0.0109035          -0.0152354          0.0013528
##          WindGustDirNW          WindGustDirS          WindGustDirSE
##          0.0021177          0.0215562          0.0228459
##          WindGustDirSSE          WindGustDirSSW          WindGustDirSW
##          0.0317376          0.0179158          0.0117916
##          WindGustDirW          WindGustDirWNW          WindGustDirWSW
##          0.0030004          0.0071477          0.0140765
##          Evaporation          WindSpeed9am          Temp3pm
##          0.0022531          -0.0005254          -0.0032099
```

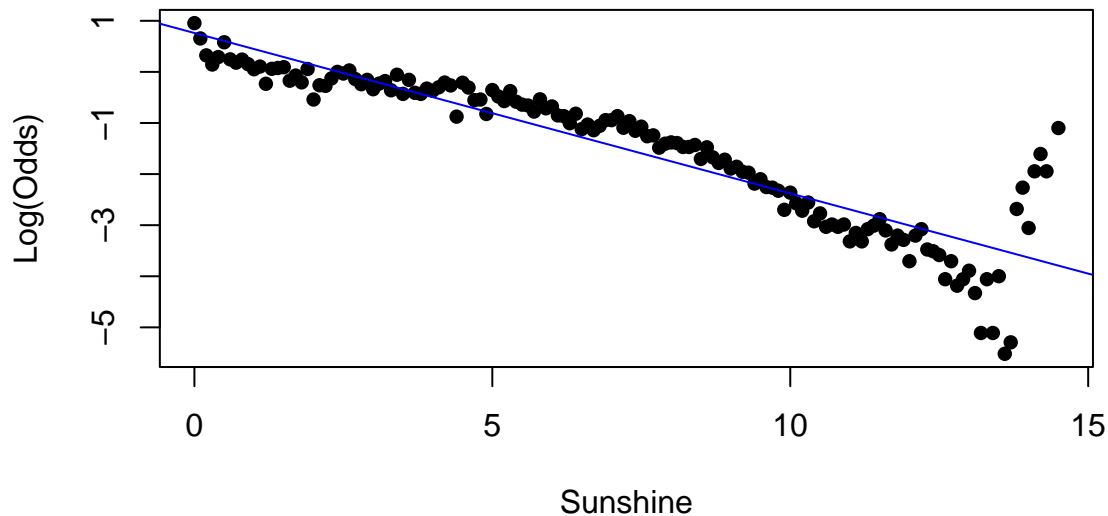


```
##          model p    rsq  rss  adjr2    cp    bic  stderr
## 1          H3  2  0.207 7681  0.207 14165 -13089  0.369
## 2         WG-H3  3  0.271 7061  0.271  8476 -17821  0.354
## 3        S-H3-P3  4  0.312 6662  0.312  4808 -21095  0.344
## 4       S-WG-H3-P3  5  0.335 6444  0.335  2807 -22961  0.338
## 5      S-WG-WS3-H3-P3  6  0.345 6348  0.345  1928 -23795  0.335
## 6     S-WG-WS3-H3-P9-P3  7  0.351 6293  0.350  1422 -24278  0.334
## 7    S-WG-WS3-H3-P9-P3-RT  8  0.358 6225  0.358   797 -24882  0.332
## 8 S-WG-WS3-H3-P9-P3-C9-RT  9  0.359 6212  0.359   686 -24982  0.332
```

```
## Model variables with abbreviations
```

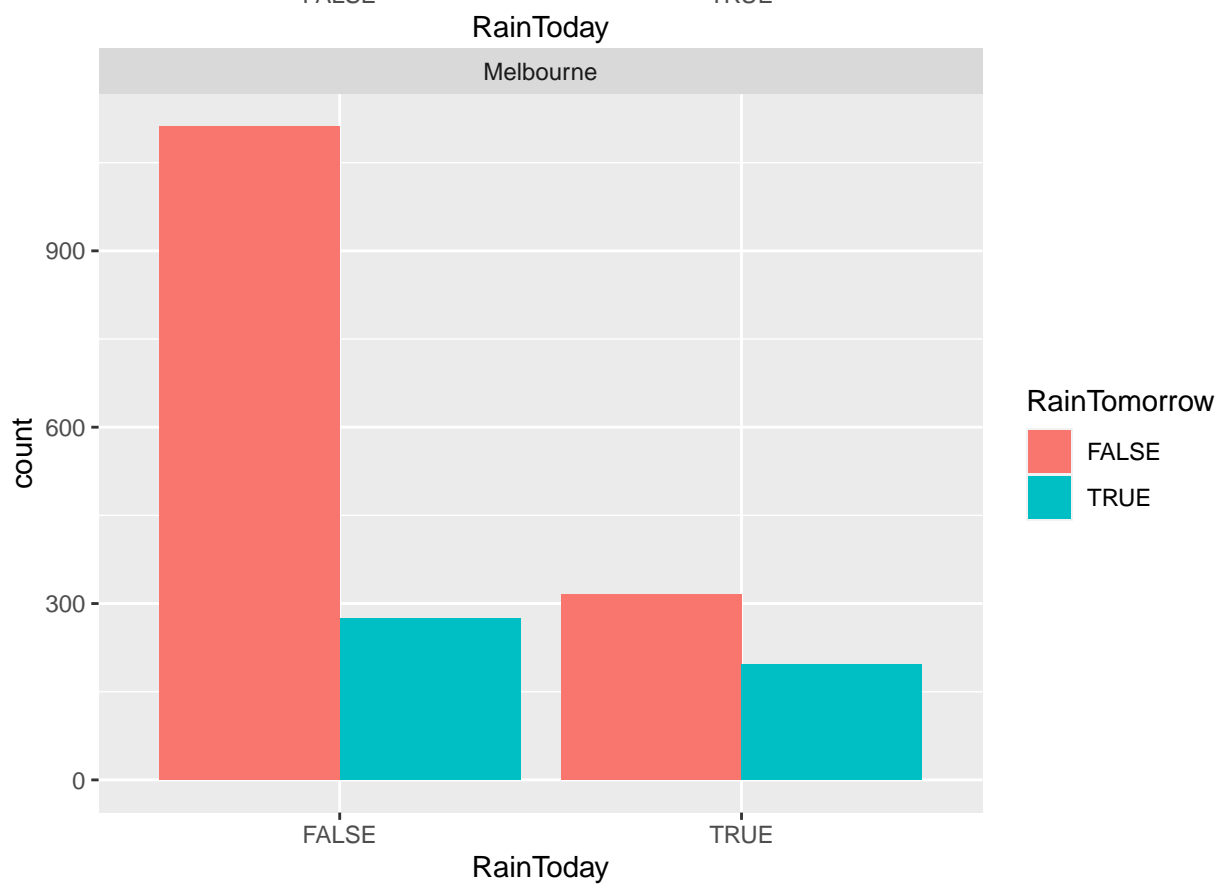
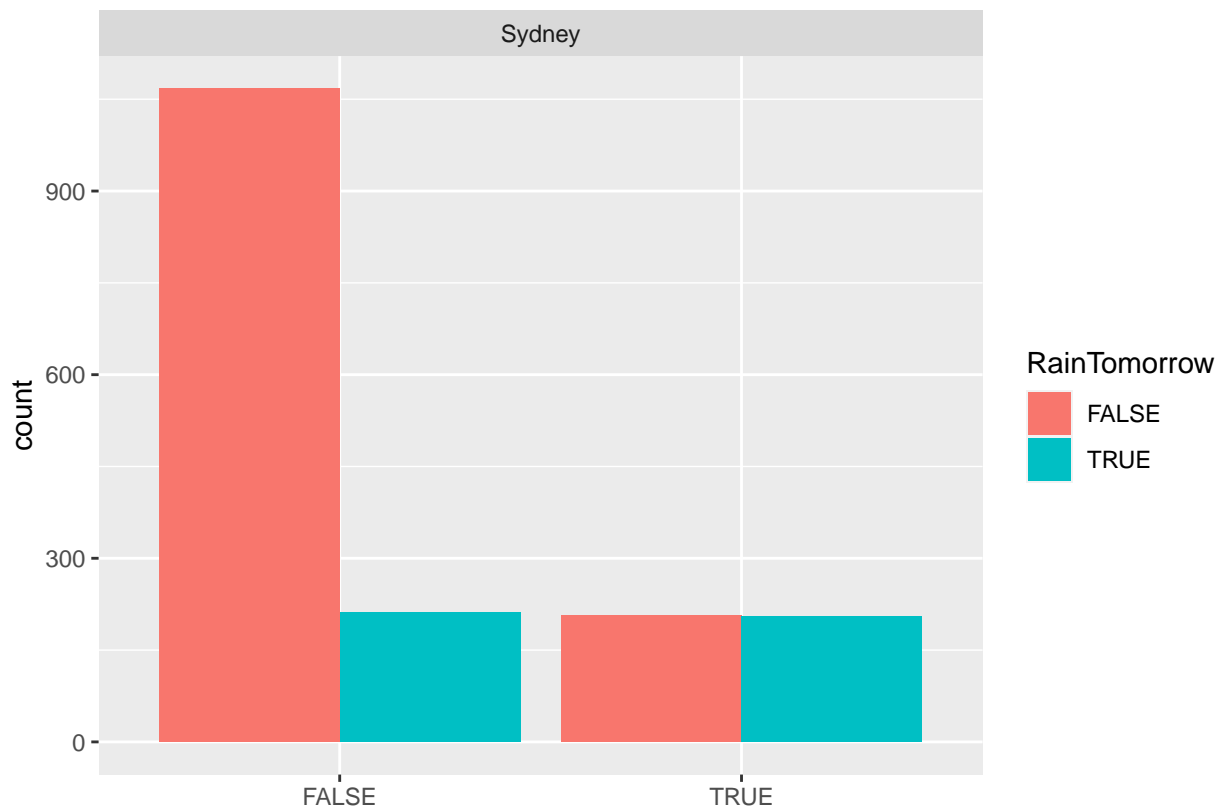
```
## H3
## WG-H3
## S-H3-P3
## S-WG-H3-P3
## S-WG-WS3-H3-P3
## S-WG-WS3-H3-P9-P3
## S-WG-WS3-H3-P9-P3-RT
## S-WG-WS3-H3-P9-P3-C9-RT
## model with largest adjr2
```

```
## 8
##
## Number of observations
## 56420
##
## Call:
## glm(formula = RainTomorrow ~ Sunshine + WindGustSpeed + WindSpeed3pm +
##      Humidity3pm + Pressure3pm + RainToday, family = "binomial",
##      data = clean_weather)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  74.4980470   2.0850771   35.73  <2e-16 ***
## Sunshine    -0.1679587   0.0041962  -40.03  <2e-16 ***
## WindGustSpeed  0.0508944   0.0013605   37.41  <2e-16 ***
## WindSpeed3pm -0.0330498   0.0020700  -15.97  <2e-16 ***
## Humidity3pm   0.0533006   0.0009395   56.73  <2e-16 ***
## Pressure3pm  -0.0781103   0.0020437  -38.22  <2e-16 ***
## RainTodayTRUE  0.3888131   0.0286052   13.59  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 59493  on 56419  degrees of freedom
## Residual deviance: 38526  on 56413  degrees of freedom
## AIC: 38540
##
## Number of Fisher Scoring iterations: 5
```



```
## [1] 0.1066608
```

Question 3 Code

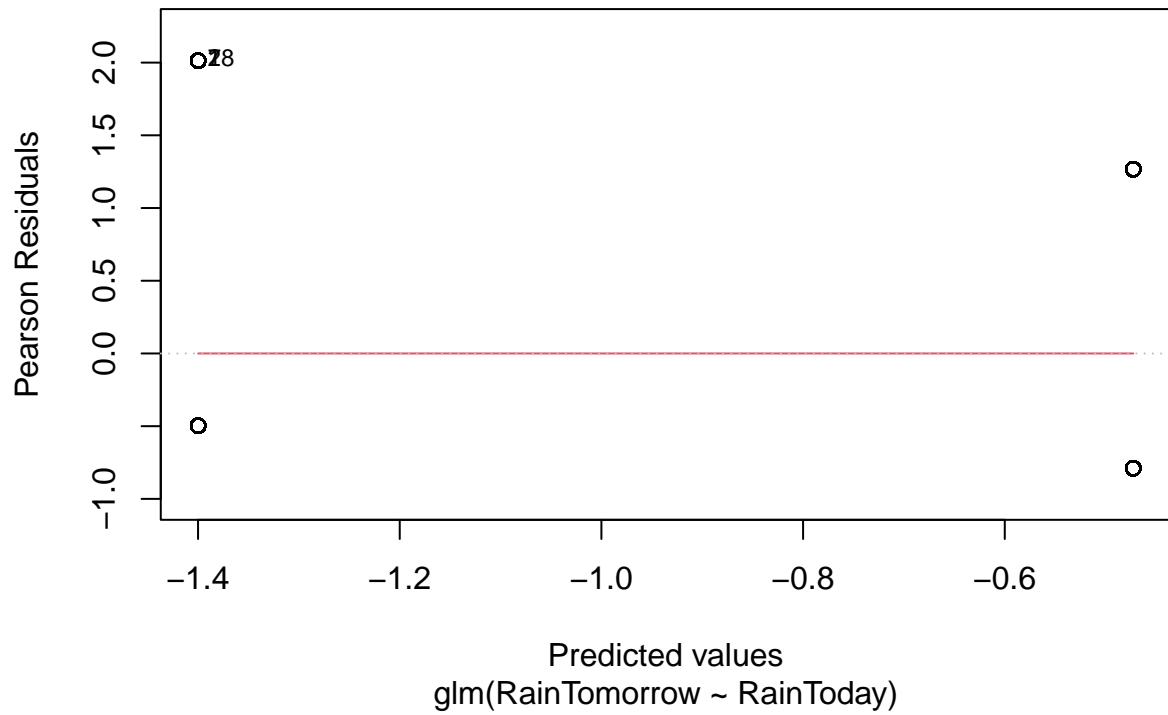
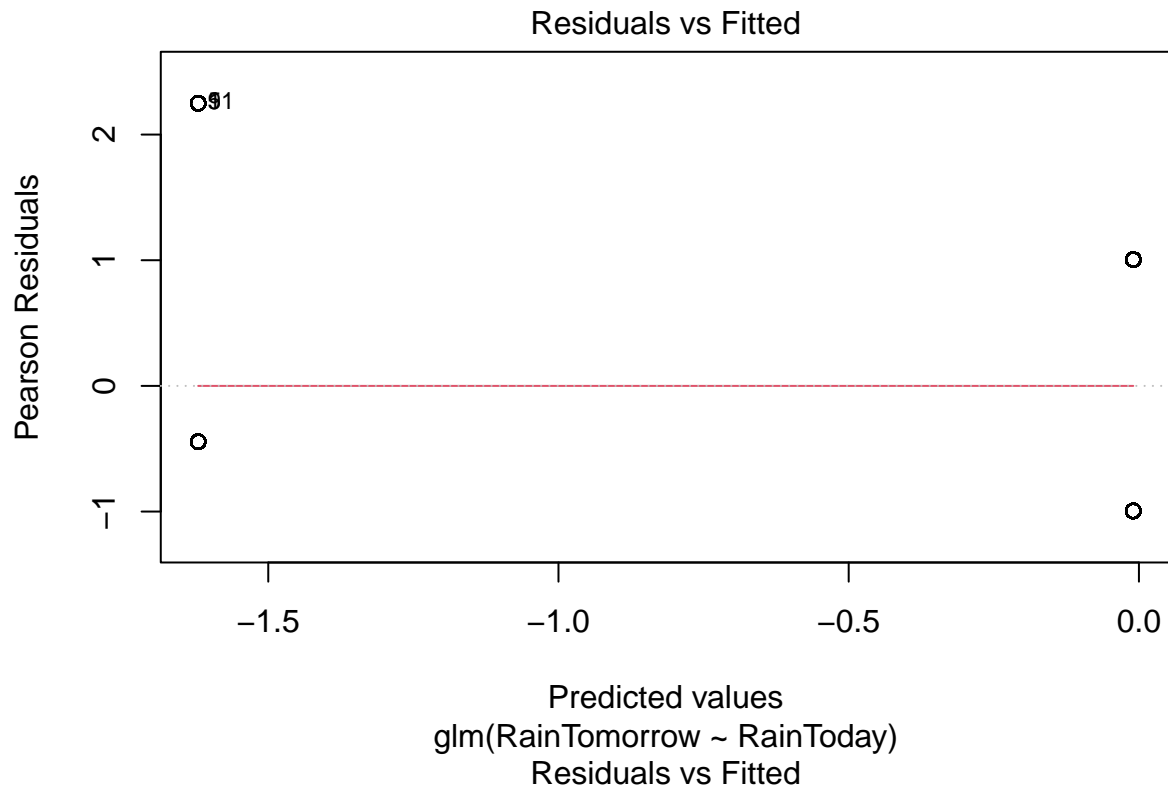


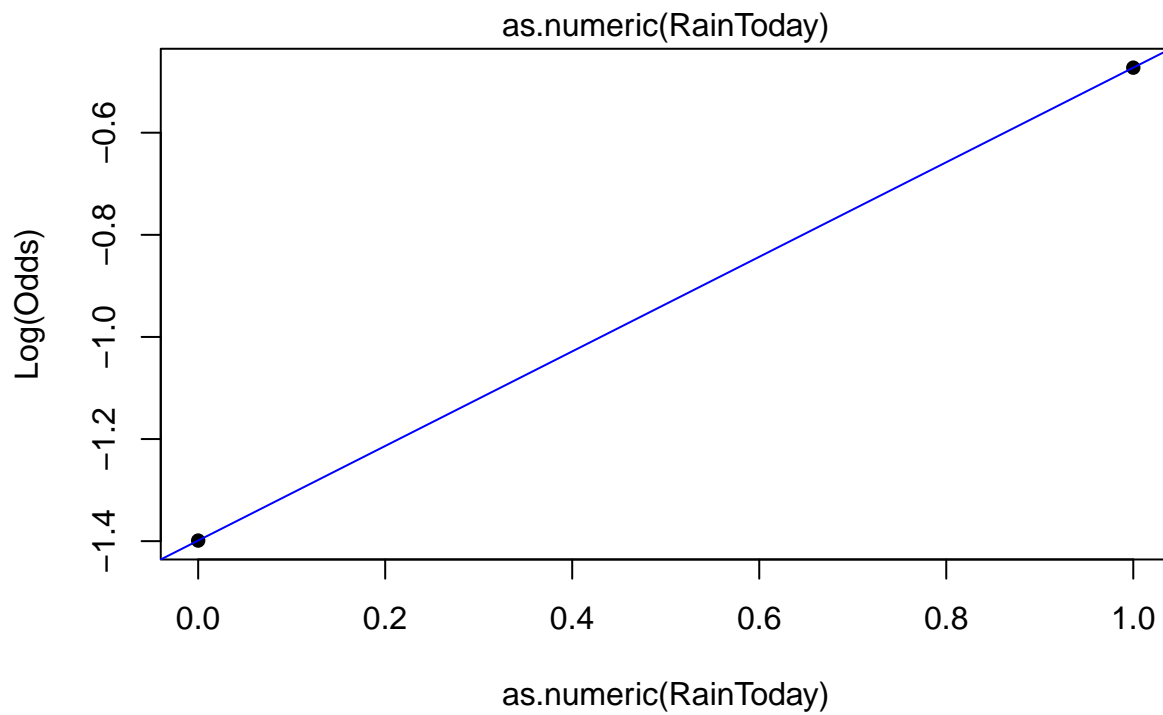
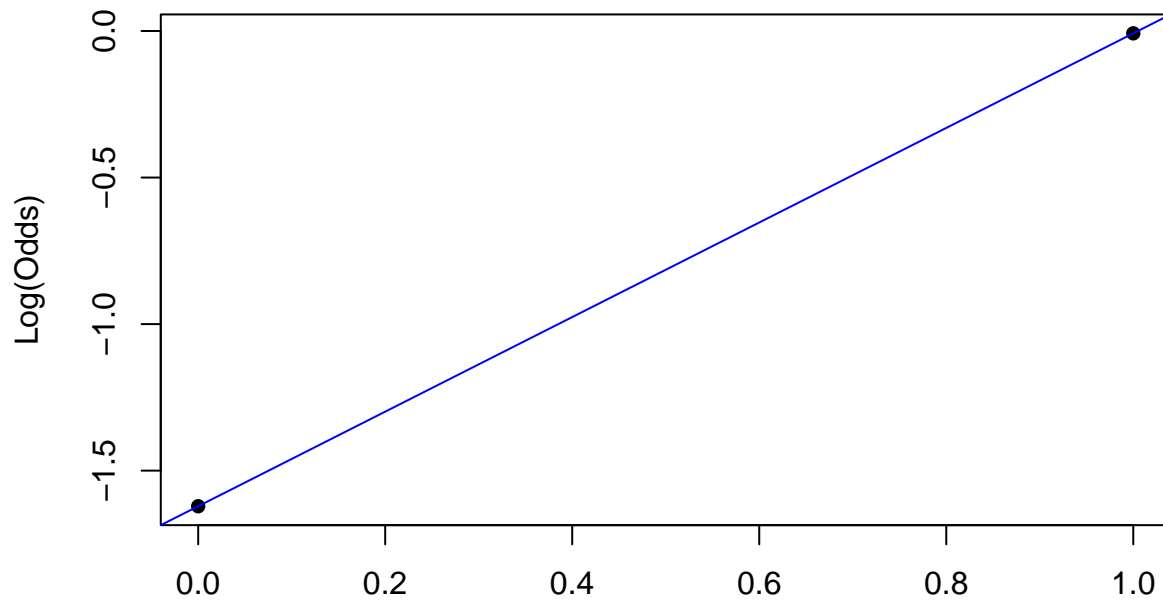
##


```

## Call:
## glm(formula = RainTomorrow ~ RainToday, family = binomial(link = "logit"),
##      data = SydneyWeather)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.62075    0.07534  -21.51  <2e-16 ***
## RainTodayTRUE  1.61104    0.12404   12.99  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1886.3  on 1689  degrees of freedom
## Residual deviance: 1716.3  on 1688  degrees of freedom
## AIC: 1720.3
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = RainTomorrow ~ RainToday, family = binomial(link = "logit"),
##      data = MelbourneWeather)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.39989    0.06745  -20.75  < 2e-16 ***
## RainTodayTRUE  0.92735    0.11309    8.20 2.41e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2126.9  on 1897  degrees of freedom
## Residual deviance: 2061.1  on 1896  degrees of freedom
## AIC: 2065.1
##
## Number of Fisher Scoring iterations: 4

```





```
## [1] 0.1651841
## [1] 0.1797378
## [1] 1.047284
## [1] 0.6155987
```

Question 4 Code

This section should include the code for the linear models used to answer the research questions, as well as code for any comparisons you are making.