# BAN502 - Predictive Analytics

## Goode, Stephen

### Course Project, PHASE 1

library(tidyverse, quietly = TRUE)

## Registered S3 methods overwritten by 'ggplot2':  
## method from   
## [.quosures rlang  
## c.quosures rlang  
## print.quosures rlang

## -- Attaching packages ----------------------------------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.1 v purrr 0.3.2  
## v tibble 2.1.1 v dplyr 0.8.1  
## v tidyr 0.8.3 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts -------------------------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2, quietly = TRUE)  
library(gapminder)  
library(dplyr)  
library(ggplot2)  
library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(leaps)  
library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(ROCR)

## Loading required package: gplots

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(VIM) #visualizing missingness

## Loading required package: colorspace

## Loading required package: grid

## Loading required package: data.table

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

## VIM is ready to use.   
## Since version 4.0.0 the GUI is in its own package VIMGUI.  
##   
## Please use the package to use the new (and old) GUI.

## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues

##   
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':  
##   
## sleep

library(visdat)  
library(mice)

## Registered S3 methods overwritten by 'lme4':  
## method from  
## cooks.distance.influence.merMod car   
## influence.merMod car   
## dfbeta.influence.merMod car   
## dfbetas.influence.merMod car

##   
## Attaching package: 'mice'

## The following object is masked from 'package:tidyr':  
##   
## complete

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

library(rpart)  
library(rpart.plot)  
library(RColorBrewer)  
library(rattle)

## Rattle: A free graphical interface for data science with R.  
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

rain <- read\_csv("rain.csv",   
 col\_types = cols(Date = col\_date(format = "%m/%d/%Y")))

str(rain)

## Classes 'spec\_tbl\_df', 'tbl\_df', 'tbl' and 'data.frame': 28003 obs. of 20 variables:  
## $ Date : Date, format: "2008-12-05" "2008-12-06" ...  
## $ MinTemp : num 17.5 14.6 9.8 14.1 20.5 20.1 9.6 14 12.5 17.4 ...  
## $ MaxTemp : num 32.3 29.7 27.7 20.9 31.8 32.7 23.9 28.3 28.4 43 ...  
## $ Rainfall : num 1 0.2 NA 0 0 0 0 0 0 0 ...  
## $ WindGustDir : chr "W" "WNW" "WNW" "ENE" ...  
## $ WindGustSpeed: num 41 56 50 22 41 48 41 48 37 39 ...  
## $ WindDir9am : chr "ENE" "W" NA "SSW" ...  
## $ WindDir3pm : chr "NW" "W" "WNW" "E" ...  
## $ WindSpeed9am : num 7 19 NA 11 19 13 19 17 20 7 ...  
## $ WindSpeed3pm : num 20 24 22 9 20 30 11 24 9 17 ...  
## $ Humidity9am : num 82 55 50 69 54 56 44 43 38 40 ...  
## $ Humidity3pm : num 33 23 28 82 24 15 22 15 16 8 ...  
## $ Pressure9am : num 1011 1009 1013 1012 1008 ...  
## $ Pressure3pm : num 1006 1005 1010 1010 1006 ...  
## $ Cloud9am : num 7 NA 0 8 NA NA NA NA NA NA ...  
## $ Cloud3pm : num 8 NA NA 1 NA NA NA NA NA NA ...  
## $ Temp9am : num 17.8 20.6 17.3 17.2 23.8 24.6 14.9 17.9 17.2 25.6 ...  
## $ Temp3pm : num 29.7 28.9 26.2 18.1 30.8 32.1 22.1 27.6 26.6 41.5 ...  
## $ RainToday : chr "No" "No" NA "No" ...  
## $ RainTomorrow : chr "No" "No" "No" "Yes" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Date = col\_date(format = "%m/%d/%Y"),  
## .. MinTemp = col\_double(),  
## .. MaxTemp = col\_double(),  
## .. Rainfall = col\_double(),  
## .. WindGustDir = col\_character(),  
## .. WindGustSpeed = col\_double(),  
## .. WindDir9am = col\_character(),  
## .. WindDir3pm = col\_character(),  
## .. WindSpeed9am = col\_double(),  
## .. WindSpeed3pm = col\_double(),  
## .. Humidity9am = col\_double(),  
## .. Humidity3pm = col\_double(),  
## .. Pressure9am = col\_double(),  
## .. Pressure3pm = col\_double(),  
## .. Cloud9am = col\_double(),  
## .. Cloud3pm = col\_double(),  
## .. Temp9am = col\_double(),  
## .. Temp3pm = col\_double(),  
## .. RainToday = col\_character(),  
## .. RainTomorrow = col\_character()  
## .. )

summary(rain)

## Date MinTemp MaxTemp Rainfall   
## Min. :2007-11-01 Min. :-8.50 Min. :-3.00 Min. : 0.000   
## 1st Qu.:2011-01-07 1st Qu.: 7.60 1st Qu.:17.90 1st Qu.: 0.000   
## Median :2013-06-10 Median :12.00 Median :22.60 Median : 0.000   
## Mean :2013-04-04 Mean :12.16 Mean :23.18 Mean : 2.265   
## 3rd Qu.:2015-06-13 3rd Qu.:16.80 3rd Qu.:28.20 3rd Qu.: 0.650   
## Max. :2017-06-25 Max. :30.50 Max. :47.00 Max. :268.600   
## NA's :132 NA's :64 NA's :295   
## WindGustDir WindGustSpeed WindDir9am WindDir3pm   
## Length:28003 Min. : 7.00 Length:28003 Length:28003   
## Class :character 1st Qu.: 31.00 Class :character Class :character   
## Mode :character Median : 39.00 Mode :character Mode :character   
## Mean : 40.02   
## 3rd Qu.: 48.00   
## Max. :135.00   
## NA's :1840   
## WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm   
## Min. : 0.00 Min. : 0.00 Min. : 1.00 Min. : 0.00   
## 1st Qu.: 7.00 1st Qu.:13.00 1st Qu.: 57.00 1st Qu.: 36.00   
## Median :13.00 Median :19.00 Median : 70.00 Median : 52.00   
## Mean :13.97 Mean :18.59 Mean : 68.86 Mean : 51.54   
## 3rd Qu.:19.00 3rd Qu.:24.00 3rd Qu.: 83.00 3rd Qu.: 66.00   
## Max. :87.00 Max. :83.00 Max. :100.00 Max. :100.00   
## NA's :308 NA's :526 NA's :366 NA's :694   
## Pressure9am Pressure3pm Cloud9am Cloud3pm   
## Min. : 980.5 Min. : 978.2 Min. :0.000 Min. :0.000   
## 1st Qu.:1013.0 1st Qu.:1010.5 1st Qu.:1.000 1st Qu.:2.000   
## Median :1017.7 Median :1015.3 Median :5.000 Median :5.000   
## Mean :1017.7 Mean :1015.3 Mean :4.459 Mean :4.513   
## 3rd Qu.:1022.4 3rd Qu.:1020.0 3rd Qu.:7.000 3rd Qu.:7.000   
## Max. :1041.0 Max. :1037.0 Max. :8.000 Max. :8.000   
## NA's :2837 NA's :2817 NA's :10673 NA's :11341   
## Temp9am Temp3pm RainToday RainTomorrow   
## Min. :-5.60 Min. :-4.20 Length:28003 Length:28003   
## 1st Qu.:12.30 1st Qu.:16.60 Class :character Class :character   
## Median :16.70 Median :21.10 Mode :character Mode :character   
## Mean :16.96 Mean :21.63   
## 3rd Qu.:21.50 3rd Qu.:26.40   
## Max. :38.60 Max. :45.20   
## NA's :196 NA's :532

rain = rain %>%  
 separate(Date, sep="-", into = c("Year", "Month", "Day"))

rain = rain %>% mutate(RainTomorrow = as\_factor(RainTomorrow)) %>% mutate(RainTomorrow = fct\_recode(RainTomorrow,  
 "Yes" = "1",  
 "No" = "0"))

## Warning: Unknown levels in `f`: 1, 0

rain = rain %>% mutate(RainToday= as\_factor(RainToday)) %>% mutate(RainToday = fct\_recode(RainToday,  
 "Yes" = "1",  
 "No" = "0"))

## Warning: Unknown levels in `f`: 1, 0

rain = rain %>% mutate(WindDir9am = as\_factor(WindDir9am)) %>% mutate(WindDir9am = fct\_recode(WindDir9am,  
 "E" = "1",  
 "ENE" = "2",  
 "ESE" = "3",  
 "N" = "4",  
 "NE" = "5",  
 "NNE" = "6",  
 "NNW" = "7",  
 "NW" = "8",  
 "S" = "9",  
 "SE" = "10",  
 "SSE" = "11",  
 "SSW" = "12",  
 "SW" = "13",  
 "W" = "14",  
 "WNW" = "15",  
 "WSW" = "16"))

## Warning: Unknown levels in `f`: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,  
## 14, 15, 16

rain = rain %>% mutate(WindDir3pm = as\_factor(WindDir3pm)) %>% mutate(WindDir3pm = fct\_recode(WindDir3pm,  
 "E" = "1",  
 "ENE" = "2",  
 "ESE" = "3",  
 "N" = "4",  
 "NE" = "5",  
 "NNE" = "6",  
 "NNW" = "7",  
 "NW" = "8",  
 "S" = "9",  
 "SE" = "10",  
 "SSE" = "11",  
 "SSW" = "12",  
 "SW" = "13",  
 "W" = "14",  
 "WNW" = "15",  
 "WSW" = "16"))

## Warning: Unknown levels in `f`: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,  
## 14, 15, 16

rain = rain %>% mutate(WindGustDir = as\_factor(WindGustDir)) %>% mutate(WindGustDir = fct\_recode(WindGustDir,  
 "E" = "1",  
 "ENE" = "2",  
 "ESE" = "3",  
 "N" = "4",  
 "NE" = "5",  
 "NNE" = "6",  
 "NNW" = "7",  
 "NW" = "8",  
 "S" = "9",  
 "SE" = "10",  
 "SSE" = "11",  
 "SSW" = "12",  
 "SW" = "13",  
 "W" = "14",  
 "WNW" = "15",  
 "WSW" = "16"))

## Warning: Unknown levels in `f`: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,  
## 14, 15, 16

rain = rain %>% mutate(RainToday = as\_factor(as.character(RainToday))) %>% mutate(RainToday = fct\_recode(RainToday,  
 "No" = "0",  
 "Yes" = "1"))

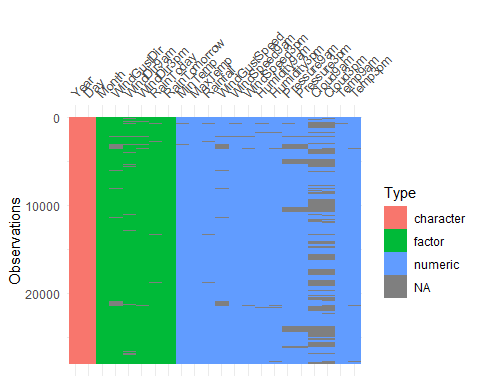
## Warning: Unknown levels in `f`: 0, 1

rain = rain %>% mutate(RainTomorrow = as\_factor(as.character(RainTomorrow))) %>% mutate(RainTomorrow = fct\_recode(RainTomorrow,  
 "No" = "0",  
 "Yes" = "1"))

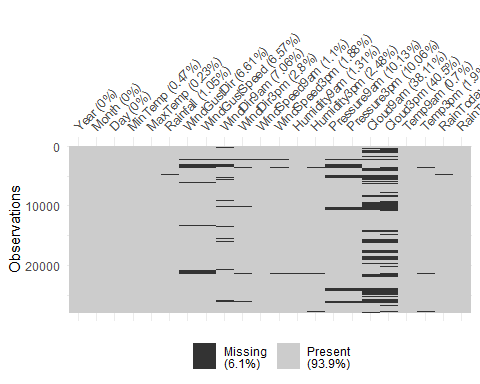
## Warning: Unknown levels in `f`: 0, 1

rain = rain %>% mutate(Month = as\_factor(Month)) %>% mutate(Month = fct\_recode(Month,  
 "January" = "01",  
 "February" = "02",  
 "March" = "03",  
 "April" = "04",  
 "May" = "05",  
 "June" = "06",  
 "July" = "07",  
 "August" = "08",  
 "September" = "09",  
 "October" = "10",  
 "November" = "11",  
 "December" = "12"))

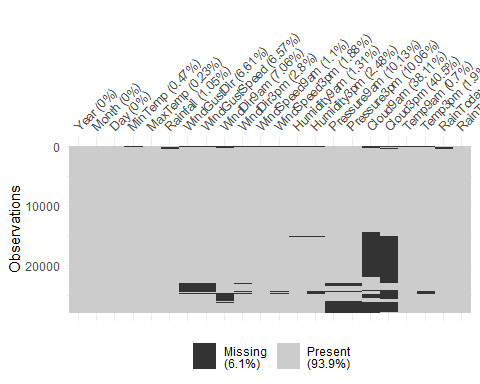
start\_time = Sys.time() #for timing  
  
vis\_dat(rain)



vis\_miss(rain)

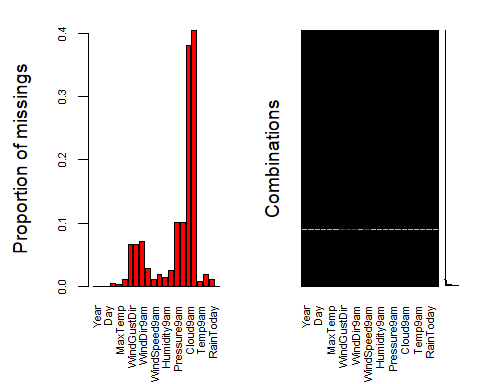


vis\_miss(rain, cluster = TRUE)



vim\_plot = aggr(rain, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)

## Warning in plot.aggr(res, ...): not enough vertical space to display  
## frequencies (too many combinations)



end\_time = Sys.time()  
end\_time-start\_time

## Time difference of 3.274876 mins

str(rain)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 28003 obs. of 22 variables:  
## $ Year : chr "2008" "2008" "2008" "2008" ...  
## $ Month : Factor w/ 12 levels "December","January",..: 1 1 1 1 1 1 2 2 2 2 ...  
## $ Day : chr "05" "06" "16" "17" ...  
## $ MinTemp : num 17.5 14.6 9.8 14.1 20.5 20.1 9.6 14 12.5 17.4 ...  
## $ MaxTemp : num 32.3 29.7 27.7 20.9 31.8 32.7 23.9 28.3 28.4 43 ...  
## $ Rainfall : num 1 0.2 NA 0 0 0 0 0 0 0 ...  
## $ WindGustDir : Factor w/ 16 levels "W","WNW","ENE",..: 1 2 2 3 2 2 1 1 4 5 ...  
## $ WindGustSpeed: num 41 56 50 22 41 48 41 48 37 39 ...  
## $ WindDir9am : Factor w/ 16 levels "ENE","W","SSW",..: 1 2 NA 3 2 4 5 2 6 6 ...  
## $ WindDir3pm : Factor w/ 16 levels "NW","W","WNW",..: 1 2 3 4 2 3 5 6 7 5 ...  
## $ WindSpeed9am : num 7 19 NA 11 19 13 19 17 20 7 ...  
## $ WindSpeed3pm : num 20 24 22 9 20 30 11 24 9 17 ...  
## $ Humidity9am : num 82 55 50 69 54 56 44 43 38 40 ...  
## $ Humidity3pm : num 33 23 28 82 24 15 22 15 16 8 ...  
## $ Pressure9am : num 1011 1009 1013 1012 1008 ...  
## $ Pressure3pm : num 1006 1005 1010 1010 1006 ...  
## $ Cloud9am : num 7 NA 0 8 NA NA NA NA NA NA ...  
## $ Cloud3pm : num 8 NA NA 1 NA NA NA NA NA NA ...  
## $ Temp9am : num 17.8 20.6 17.3 17.2 23.8 24.6 14.9 17.9 17.2 25.6 ...  
## $ Temp3pm : num 29.7 28.9 26.2 18.1 30.8 32.1 22.1 27.6 26.6 41.5 ...  
## $ RainToday : Factor w/ 2 levels "No","Yes": 1 1 NA 1 1 1 1 1 1 1 ...  
## $ RainTomorrow : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...

summary(rain)

## Year Month Day MinTemp   
## Length:28003 May : 2622 Length:28003 Min. :-8.50   
## Class :character March : 2522 Class :character 1st Qu.: 7.60   
## Mode :character January: 2505 Mode :character Median :12.00   
## July : 2417 Mean :12.16   
## June : 2400 3rd Qu.:16.80   
## October: 2320 Max. :30.50   
## (Other):13217 NA's :132   
## MaxTemp Rainfall WindGustDir WindGustSpeed   
## Min. :-3.00 Min. : 0.000 W : 1913 Min. : 7.00   
## 1st Qu.:17.90 1st Qu.: 0.000 SE : 1863 1st Qu.: 31.00   
## Median :22.60 Median : 0.000 E : 1810 Median : 39.00   
## Mean :23.18 Mean : 2.265 S : 1779 Mean : 40.02   
## 3rd Qu.:28.20 3rd Qu.: 0.650 N : 1778 3rd Qu.: 48.00   
## Max. :47.00 Max. :268.600 (Other):17009 Max. :135.00   
## NA's :64 NA's :295 NA's : 1851 NA's :1840   
## WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm   
## N : 2275 SE : 2036 Min. : 0.00 Min. : 0.00   
## SSE : 1781 W : 1924 1st Qu.: 7.00 1st Qu.:13.00   
## SE : 1764 S : 1877 Median :13.00 Median :19.00   
## E : 1739 SSE : 1862 Mean :13.97 Mean :18.59   
## NW : 1711 WSW : 1850 3rd Qu.:19.00 3rd Qu.:24.00   
## (Other):16756 (Other):17669 Max. :87.00 Max. :83.00   
## NA's : 1977 NA's : 785 NA's :308 NA's :526   
## Humidity9am Humidity3pm Pressure9am Pressure3pm   
## Min. : 1.00 Min. : 0.00 Min. : 980.5 Min. : 978.2   
## 1st Qu.: 57.00 1st Qu.: 36.00 1st Qu.:1013.0 1st Qu.:1010.5   
## Median : 70.00 Median : 52.00 Median :1017.7 Median :1015.3   
## Mean : 68.86 Mean : 51.54 Mean :1017.7 Mean :1015.3   
## 3rd Qu.: 83.00 3rd Qu.: 66.00 3rd Qu.:1022.4 3rd Qu.:1020.0   
## Max. :100.00 Max. :100.00 Max. :1041.0 Max. :1037.0   
## NA's :366 NA's :694 NA's :2837 NA's :2817   
## Cloud9am Cloud3pm Temp9am Temp3pm   
## Min. :0.000 Min. :0.000 Min. :-5.60 Min. :-4.20   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:12.30 1st Qu.:16.60   
## Median :5.000 Median :5.000 Median :16.70 Median :21.10   
## Mean :4.459 Mean :4.513 Mean :16.96 Mean :21.63   
## 3rd Qu.:7.000 3rd Qu.:7.000 3rd Qu.:21.50 3rd Qu.:26.40   
## Max. :8.000 Max. :8.000 Max. :38.60 Max. :45.20   
## NA's :10673 NA's :11341 NA's :196 NA's :532   
## RainToday RainTomorrow  
## No :21525 No :21713   
## Yes : 6183 Yes: 6290   
## NA's: 295   
##   
##   
##   
##

start\_time = Sys.time() #for timing  
rain\_imp = mice(rain, m=1, method = "pmm", seed = 12345)

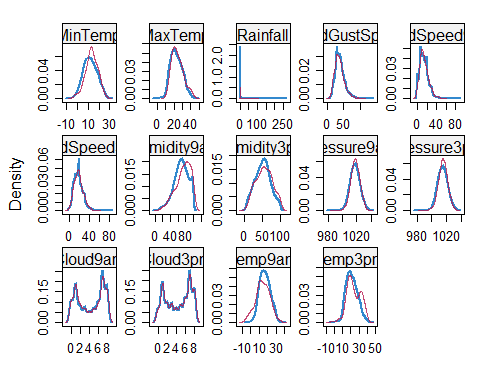
##   
## iter imp variable  
## 1 1 MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday  
## 2 1 MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday  
## 3 1 MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday  
## 4 1 MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday  
## 5 1 MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday

## Warning: Number of logged events: 2

#in line above: m=1 -> runs one imputation, seed sets the random number seed to get repeatable results  
summary(rain\_imp)

## Class: mids  
## Number of multiple imputations: 1   
## Imputation methods:  
## Year Month Day MinTemp MaxTemp   
## "" "" "" "pmm" "pmm"   
## Rainfall WindGustDir WindGustSpeed WindDir9am WindDir3pm   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm   
## "pmm" "pmm" "pmm" "pmm" "pmm"   
## RainToday RainTomorrow   
## "pmm" ""   
## PredictorMatrix:  
## Year Month Day MinTemp MaxTemp Rainfall WindGustDir WindGustSpeed  
## Year 0 1 0 1 1 1 1 1  
## Month 0 0 0 1 1 1 1 1  
## Day 0 1 0 1 1 1 1 1  
## MinTemp 0 1 0 0 1 1 1 1  
## MaxTemp 0 1 0 1 0 1 1 1  
## Rainfall 0 1 0 1 1 0 1 1  
## WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am  
## Year 1 1 1 1 1  
## Month 1 1 1 1 1  
## Day 1 1 1 1 1  
## MinTemp 1 1 1 1 1  
## MaxTemp 1 1 1 1 1  
## Rainfall 1 1 1 1 1  
## Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am  
## Year 1 1 1 1 1 1  
## Month 1 1 1 1 1 1  
## Day 1 1 1 1 1 1  
## MinTemp 1 1 1 1 1 1  
## MaxTemp 1 1 1 1 1 1  
## Rainfall 1 1 1 1 1 1  
## Temp3pm RainToday RainTomorrow  
## Year 1 1 1  
## Month 1 1 1  
## Day 1 1 1  
## MinTemp 1 1 1  
## MaxTemp 1 1 1  
## Rainfall 1 1 1  
## Number of logged events: 2   
## it im dep meth out  
## 1 0 0 constant Year  
## 2 0 0 constant Day

densityplot(rain\_imp)



#red imputed, blue original, only shows density plots when more than 1 value the variable was imputed  
#note that the density plots are fairly uninteresting given the small amount of missing data  
rain\_complete = complete(rain\_imp)  
summary(rain\_complete)

## Year Month Day MinTemp   
## Length:28003 May : 2622 Length:28003 Min. :-8.50   
## Class :character March : 2522 Class :character 1st Qu.: 7.60   
## Mode :character January: 2505 Mode :character Median :12.00   
## July : 2417 Mean :12.17   
## June : 2400 3rd Qu.:16.80   
## October: 2320 Max. :30.50   
## (Other):13217   
## MaxTemp Rainfall WindGustDir WindGustSpeed   
## Min. :-3.00 Min. : 0.00 W : 2064 Min. : 7.00   
## 1st Qu.:17.90 1st Qu.: 0.00 SE : 2007 1st Qu.: 30.00   
## Median :22.60 Median : 0.00 E : 1937 Median : 39.00   
## Mean :23.18 Mean : 2.28 N : 1896 Mean : 39.85   
## 3rd Qu.:28.20 3rd Qu.: 0.80 S : 1896 3rd Qu.: 48.00   
## Max. :47.00 Max. :268.60 SSE : 1890 Max. :135.00   
## (Other):16313   
## WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm   
## N : 2476 SE : 2098 Min. : 0.00 Min. : 0.00   
## SSE : 1896 W : 1970 1st Qu.: 7.00 1st Qu.:13.00   
## SE : 1889 S : 1953 Median :13.00 Median :19.00   
## E : 1878 SSE : 1923 Mean :13.97 Mean :18.58   
## NW : 1856 WSW : 1891 3rd Qu.:19.00 3rd Qu.:24.00   
## S : 1758 SW : 1879 Max. :87.00 Max. :83.00   
## (Other):16250 (Other):16289   
## Humidity9am Humidity3pm Pressure9am Pressure3pm   
## Min. : 1.00 Min. : 0.00 Min. : 980.5 Min. : 978.2   
## 1st Qu.: 57.00 1st Qu.: 37.00 1st Qu.:1013.1 1st Qu.:1010.6   
## Median : 70.00 Median : 52.00 Median :1017.7 Median :1015.4   
## Mean : 68.91 Mean : 51.62 Mean :1017.7 Mean :1015.3   
## 3rd Qu.: 83.00 3rd Qu.: 66.00 3rd Qu.:1022.4 3rd Qu.:1020.0   
## Max. :100.00 Max. :100.00 Max. :1041.0 Max. :1037.0   
##   
## Cloud9am Cloud3pm Temp9am Temp3pm   
## Min. :0.000 Min. :0.00 Min. :-5.60 Min. :-4.20   
## 1st Qu.:1.000 1st Qu.:2.00 1st Qu.:12.30 1st Qu.:16.60   
## Median :5.000 Median :5.00 Median :16.70 Median :21.10   
## Mean :4.404 Mean :4.51 Mean :16.93 Mean :21.67   
## 3rd Qu.:7.000 3rd Qu.:7.00 3rd Qu.:21.50 3rd Qu.:26.40   
## Max. :8.000 Max. :8.00 Max. :38.60 Max. :45.20   
##   
## RainToday RainTomorrow  
## No :21726 No :21713   
## Yes: 6277 Yes: 6290   
##   
##   
##   
##   
##

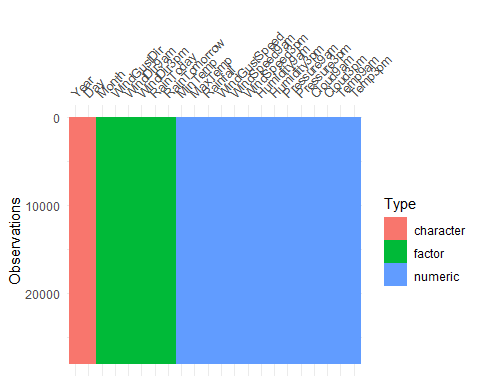
str(rain\_complete)

## 'data.frame': 28003 obs. of 22 variables:  
## $ Year : chr "2008" "2008" "2008" "2008" ...  
## $ Month : Factor w/ 12 levels "December","January",..: 1 1 1 1 1 1 2 2 2 2 ...  
## $ Day : chr "05" "06" "16" "17" ...  
## $ MinTemp : num 17.5 14.6 9.8 14.1 20.5 20.1 9.6 14 12.5 17.4 ...  
## $ MaxTemp : num 32.3 29.7 27.7 20.9 31.8 32.7 23.9 28.3 28.4 43 ...  
## $ Rainfall : num 1 0.2 0 0 0 0 0 0 0 0 ...  
## $ WindGustDir : Factor w/ 16 levels "W","WNW","ENE",..: 1 2 2 3 2 2 1 1 4 5 ...  
## $ WindGustSpeed: num 41 56 50 22 41 48 41 48 37 39 ...  
## $ WindDir9am : Factor w/ 16 levels "ENE","W","SSW",..: 1 2 16 3 2 4 5 2 6 6 ...  
## $ WindDir3pm : Factor w/ 16 levels "NW","W","WNW",..: 1 2 3 4 2 3 5 6 7 5 ...  
## $ WindSpeed9am : num 7 19 31 11 19 13 19 17 20 7 ...  
## $ WindSpeed3pm : num 20 24 22 9 20 30 11 24 9 17 ...  
## $ Humidity9am : num 82 55 50 69 54 56 44 43 38 40 ...  
## $ Humidity3pm : num 33 23 28 82 24 15 22 15 16 8 ...  
## $ Pressure9am : num 1011 1009 1013 1012 1008 ...  
## $ Pressure3pm : num 1006 1005 1010 1010 1006 ...  
## $ Cloud9am : num 7 1 0 8 1 0 0 1 8 0 ...  
## $ Cloud3pm : num 8 1 1 1 4 1 1 1 3 0 ...  
## $ Temp9am : num 17.8 20.6 17.3 17.2 23.8 24.6 14.9 17.9 17.2 25.6 ...  
## $ Temp3pm : num 29.7 28.9 26.2 18.1 30.8 32.1 22.1 27.6 26.6 41.5 ...  
## $ RainToday : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ RainTomorrow : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...

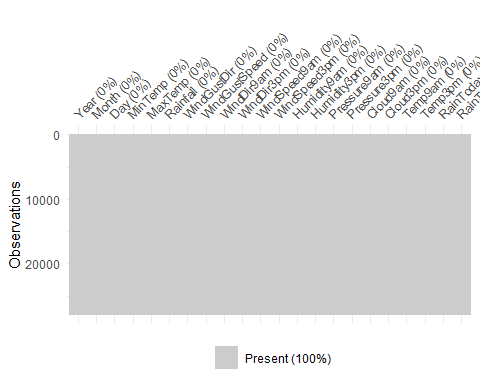
end\_time = Sys.time()  
end\_time-start\_time

## Time difference of 2.456269 mins

start\_time = Sys.time() #for timing  
  
vis\_dat(rain\_complete)

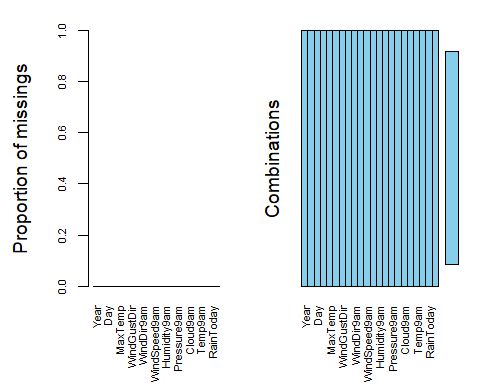


vis\_miss(rain\_complete)



vim\_plot = aggr(rain\_complete, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)

## Warning in plot.aggr(res, ...): not enough horizontal space to display  
## frequencies

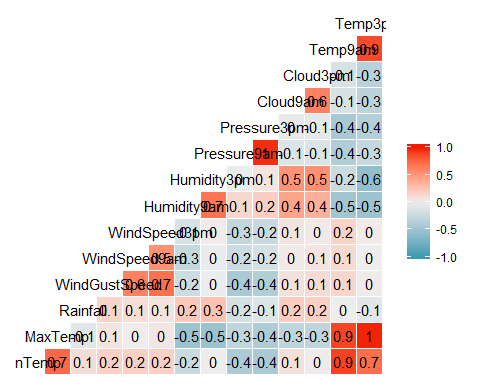


end\_time = Sys.time()  
end\_time-start\_time

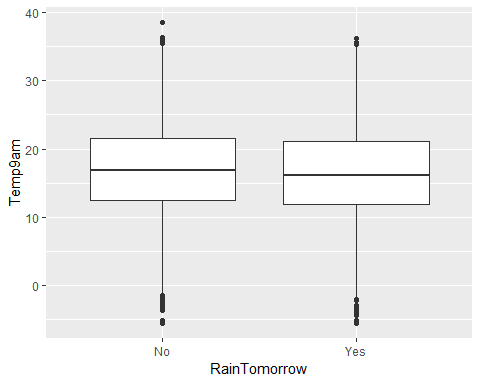
## Time difference of 23.74614 secs

ggcorr(rain\_complete, label = TRUE)

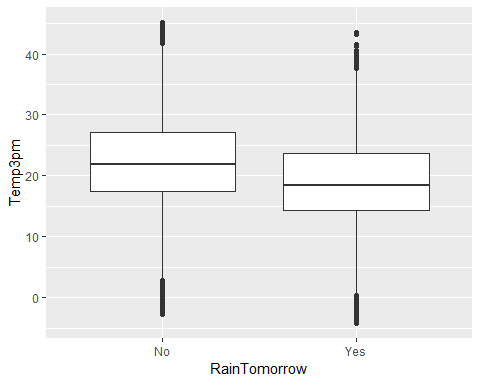
## Warning in ggcorr(rain\_complete, label = TRUE): data in column(s) 'Year',  
## 'Month', 'Day', 'WindGustDir', 'WindDir9am', 'WindDir3pm', 'RainToday',  
## 'RainTomorrow' are not numeric and were ignored



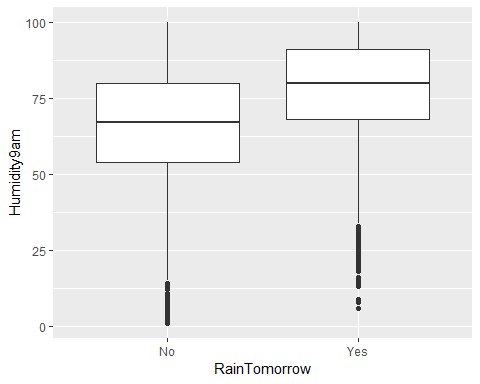
ggplot(rain\_complete,aes(x=RainTomorrow, y=Temp9am)) + geom\_boxplot()



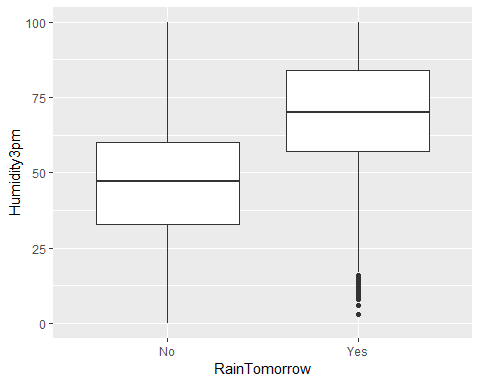
ggplot(rain\_complete,aes(x=RainTomorrow, y=Temp3pm)) + geom\_boxplot()



ggplot(rain\_complete,aes(x=RainTomorrow, y=Humidity9am)) + geom\_boxplot()



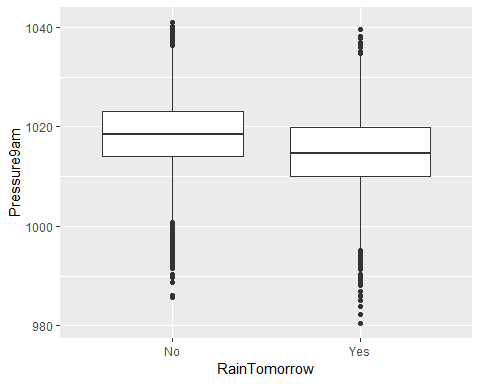
ggplot(rain\_complete,aes(x=RainTomorrow, y=Humidity3pm)) + geom\_boxplot()



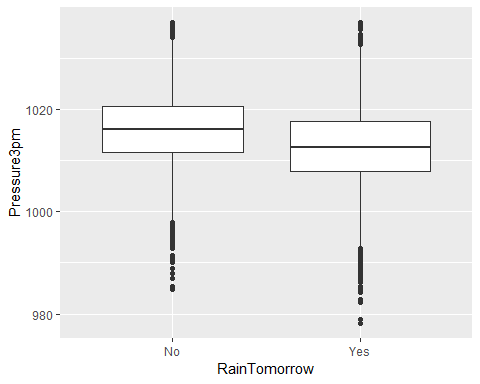
ggsave("Figure 1.png")

## Saving 5 x 4 in image

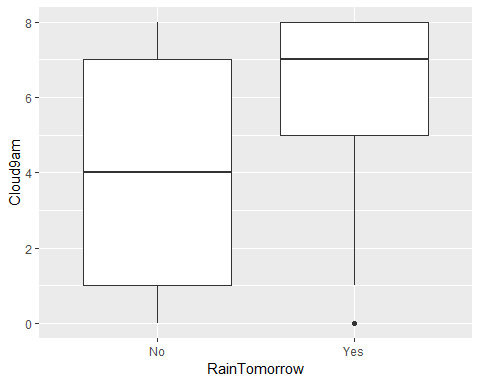
ggplot(rain\_complete,aes(x=RainTomorrow, y=Pressure9am)) + geom\_boxplot()



ggplot(rain\_complete,aes(x=RainTomorrow, y=Pressure3pm)) + geom\_boxplot()



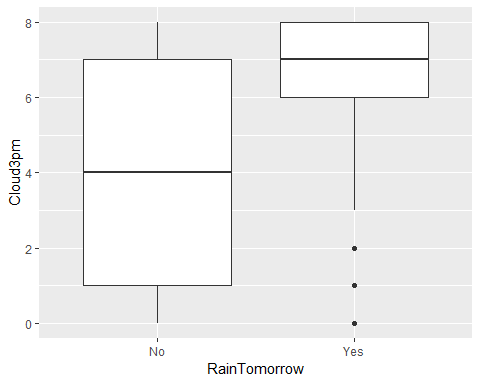
ggplot(rain\_complete,aes(x=RainTomorrow, y=Cloud9am)) + geom\_boxplot()



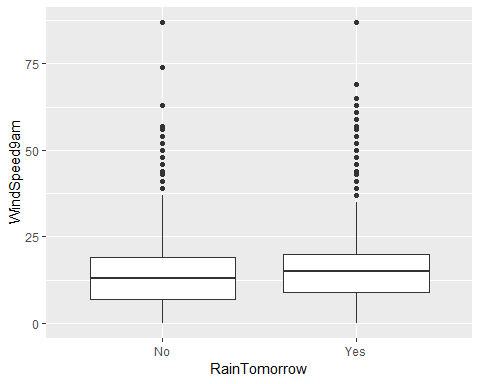
ggsave("Figure 2.png")

## Saving 5 x 4 in image

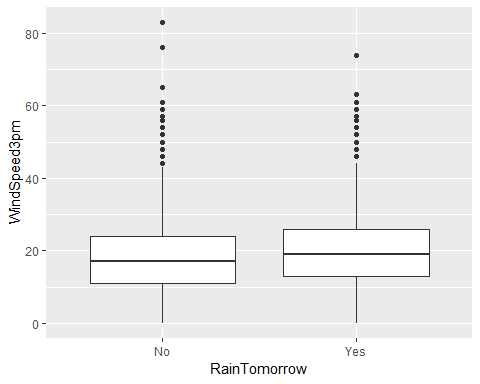
ggplot(rain\_complete,aes(x=RainTomorrow, y=Cloud3pm)) + geom\_boxplot()



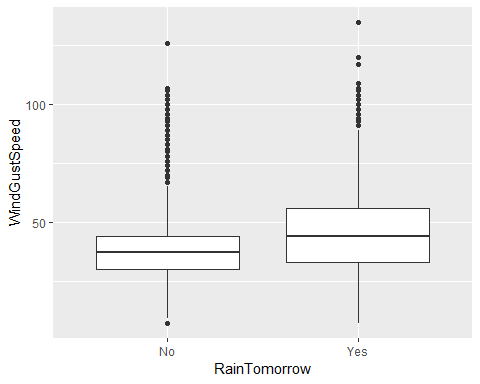
ggplot(rain\_complete,aes(x=RainTomorrow, y=WindSpeed9am)) + geom\_boxplot()



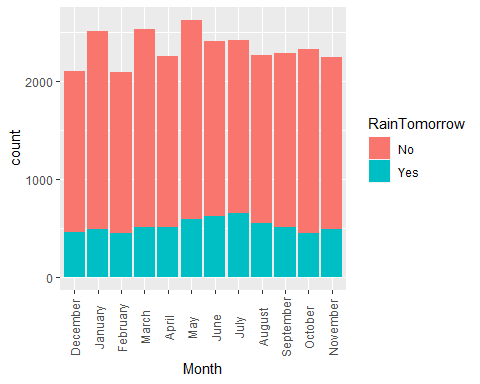
ggplot(rain\_complete,aes(x=RainTomorrow, y=WindSpeed3pm)) + geom\_boxplot()



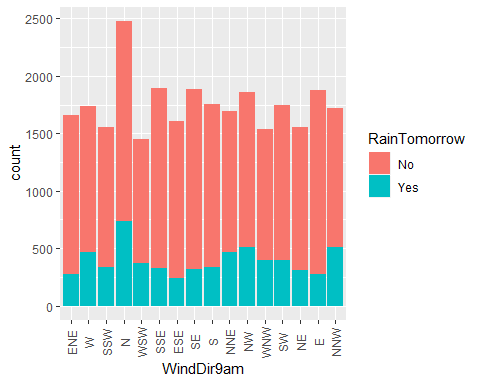
ggplot(rain\_complete,aes(x=RainTomorrow, y=WindGustSpeed)) + geom\_boxplot()



ggplot(rain\_complete,aes(x=Month, fill = RainTomorrow)) + geom\_bar() +  
 theme(axis.text.x = element\_text(angle=90, vjust=0.5))



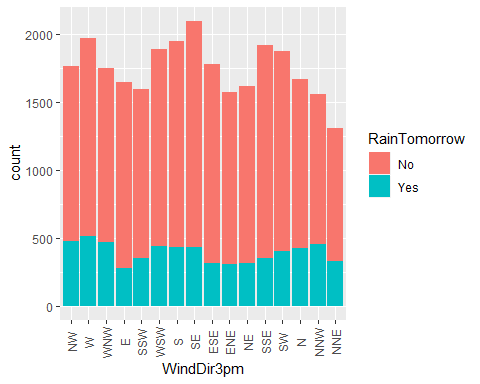
ggplot(rain\_complete,aes(x=WindDir9am, fill = RainTomorrow)) + geom\_bar() +  
 theme(axis.text.x = element\_text(angle=90, vjust = 0.5))



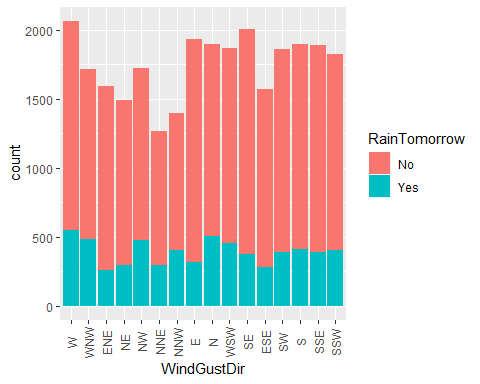
ggsave("Figure 3.png")

## Saving 5 x 4 in image

ggplot(rain\_complete,aes(x=WindDir3pm, fill = RainTomorrow)) + geom\_bar() +  
 theme(axis.text.x = element\_text(angle=90, vjust=0.5))



ggplot(rain\_complete,aes(x=WindGustDir, fill = RainTomorrow)) + geom\_bar() +  
 theme(axis.text.x = element\_text(angle=90, vjust=0.5))



# Let's split the data  
set.seed(12345)  
train.rows = createDataPartition(y = rain\_complete$RainTomorrow, p=0.7, list = FALSE) #70% in training  
train = rain\_complete[train.rows,]   
test = rain\_complete[-train.rows,]

options(scipen = 999) #Disable exponential notation  
  
allmod = glm(RainTomorrow ~., train, family = "binomial")   
summary(allmod)

##   
## Call:  
## glm(formula = RainTomorrow ~ ., family = "binomial", data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.4039 -0.5468 -0.2840 -0.1103 3.1808   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 40.523296 4.199483 9.650 < 0.0000000000000002 \*\*\*  
## Year2008 -1.446082 1.000491 -1.445 0.148353   
## Year2009 -1.439775 0.987648 -1.458 0.144901   
## Year2010 -1.429435 0.987481 -1.448 0.147741   
## Year2011 -1.367443 0.987472 -1.385 0.166116   
## Year2012 -1.411081 0.988270 -1.428 0.153341   
## Year2013 -1.498683 0.987732 -1.517 0.129192   
## Year2014 -1.574003 0.987418 -1.594 0.110923   
## Year2015 -1.351221 0.987513 -1.368 0.171216   
## Year2016 -1.492361 0.987396 -1.511 0.130684   
## Year2017 -1.611088 0.990545 -1.626 0.103851   
## MonthJanuary -0.216621 0.115759 -1.871 0.061303 .   
## MonthFebruary 0.067620 0.118730 0.570 0.568999   
## MonthMarch 0.127532 0.114155 1.117 0.263915   
## MonthApril 0.409403 0.116182 3.524 0.000425 \*\*\*  
## MonthMay 0.430363 0.117966 3.648 0.000264 \*\*\*  
## MonthJune 0.426380 0.121545 3.508 0.000451 \*\*\*  
## MonthJuly 0.529846 0.122553 4.323 0.0000154 \*\*\*  
## MonthAugust 0.507292 0.123972 4.092 0.0000428 \*\*\*  
## MonthSeptember 0.385954 0.120631 3.199 0.001377 \*\*   
## MonthOctober 0.345612 0.116028 2.979 0.002895 \*\*   
## MonthNovember 0.217231 0.112982 1.923 0.054516 .   
## Day02 -0.158480 0.169111 -0.937 0.348688   
## Day03 -0.083616 0.170642 -0.490 0.624128   
## Day04 -0.132151 0.169163 -0.781 0.434680   
## Day05 -0.268438 0.167872 -1.599 0.109806   
## Day06 -0.144031 0.163689 -0.880 0.378908   
## Day07 -0.371484 0.174493 -2.129 0.033260 \*   
## Day08 -0.200879 0.170320 -1.179 0.238231   
## Day09 -0.103224 0.162395 -0.636 0.525013   
## Day10 -0.163396 0.167108 -0.978 0.328179   
## Day11 0.185540 0.161842 1.146 0.251619   
## Day12 -0.219382 0.169600 -1.294 0.195830   
## Day13 -0.074257 0.164415 -0.452 0.651526   
## Day14 -0.163461 0.170715 -0.958 0.338309   
## Day15 -0.251802 0.171067 -1.472 0.141034   
## Day16 -0.116813 0.165763 -0.705 0.480998   
## Day17 -0.054343 0.167850 -0.324 0.746122   
## Day18 -0.216806 0.169000 -1.283 0.199536   
## Day19 0.103832 0.167311 0.621 0.534867   
## Day20 0.001543 0.168537 0.009 0.992693   
## Day21 -0.406060 0.170326 -2.384 0.017125 \*   
## Day22 -0.056217 0.165954 -0.339 0.734799   
## Day23 -0.225829 0.170136 -1.327 0.184396   
## Day24 0.039280 0.165507 0.237 0.812397   
## Day25 -0.049519 0.166791 -0.297 0.766548   
## Day26 -0.173447 0.166316 -1.043 0.297006   
## Day27 -0.360841 0.169999 -2.123 0.033787 \*   
## Day28 0.020289 0.165188 0.123 0.902245   
## Day29 -0.129517 0.170071 -0.762 0.446330   
## Day30 -0.136010 0.171244 -0.794 0.427050   
## Day31 -0.008080 0.201312 -0.040 0.967985   
## MinTemp 0.031476 0.011966 2.630 0.008527 \*\*   
## MaxTemp -0.035455 0.018787 -1.887 0.059133 .   
## Rainfall 0.006818 0.003114 2.189 0.028593 \*   
## WindGustDirWNW -0.134899 0.115895 -1.164 0.244431   
## WindGustDirENE -0.140824 0.141792 -0.993 0.320624   
## WindGustDirNE -0.124970 0.138220 -0.904 0.365921   
## WindGustDirNW -0.075859 0.119465 -0.635 0.525432   
## WindGustDirNNE -0.323969 0.142716 -2.270 0.023206 \*   
## WindGustDirNNW -0.124438 0.129054 -0.964 0.334926   
## WindGustDirE -0.059462 0.136262 -0.436 0.662564   
## WindGustDirN -0.090012 0.125102 -0.720 0.471828   
## WindGustDirWSW -0.010698 0.114764 -0.093 0.925734   
## WindGustDirSE -0.033419 0.134708 -0.248 0.804070   
## WindGustDirESE 0.018372 0.139826 0.131 0.895468   
## WindGustDirSW -0.022023 0.121418 -0.181 0.856068   
## WindGustDirS -0.048005 0.129326 -0.371 0.710496   
## WindGustDirSSE 0.189876 0.130856 1.451 0.146772   
## WindGustDirSSW -0.093345 0.127798 -0.730 0.465140   
## WindGustSpeed 0.053522 0.002509 21.336 < 0.0000000000000002 \*\*\*  
## WindDir9amW 0.007280 0.134887 0.054 0.956956   
## WindDir9amSSW -0.084535 0.141391 -0.598 0.549918   
## WindDir9amN 0.274941 0.124807 2.203 0.027600 \*   
## WindDir9amWSW 0.018648 0.139048 0.134 0.893311   
## WindDir9amSSE -0.289545 0.138807 -2.086 0.036983 \*   
## WindDir9amESE -0.145755 0.146176 -0.997 0.318707   
## WindDir9amSE -0.162948 0.138613 -1.176 0.239772   
## WindDir9amS -0.307493 0.140937 -2.182 0.029126 \*   
## WindDir9amNNE 0.446398 0.132938 3.358 0.000785 \*\*\*  
## WindDir9amNW 0.043224 0.131649 0.328 0.742664   
## WindDir9amWNW -0.014128 0.137506 -0.103 0.918163   
## WindDir9amSW 0.018622 0.135313 0.138 0.890538   
## WindDir9amNE 0.109425 0.139602 0.784 0.433138   
## WindDir9amE -0.042250 0.139831 -0.302 0.762539   
## WindDir9amNNW -0.012872 0.134980 -0.095 0.924029   
## WindDir3pmW 0.005504 0.122984 0.045 0.964303   
## WindDir3pmWNW -0.030477 0.121631 -0.251 0.802148   
## WindDir3pmE -0.336717 0.142037 -2.371 0.017758 \*   
## WindDir3pmSSW -0.098519 0.140946 -0.699 0.484562   
## WindDir3pmWSW -0.050129 0.129297 -0.388 0.698238   
## WindDir3pmS -0.183267 0.137356 -1.334 0.182121   
## WindDir3pmSE -0.070031 0.133866 -0.523 0.600876   
## WindDir3pmESE -0.431885 0.142024 -3.041 0.002359 \*\*   
## WindDir3pmENE -0.257903 0.140464 -1.836 0.066345 .   
## WindDir3pmNE -0.457638 0.137496 -3.328 0.000874 \*\*\*  
## WindDir3pmSSE -0.355456 0.139671 -2.545 0.010929 \*   
## WindDir3pmSW -0.013615 0.134512 -0.101 0.919377   
## WindDir3pmN -0.230455 0.129459 -1.780 0.075053 .   
## WindDir3pmNNW 0.168117 0.123132 1.365 0.172147   
## WindDir3pmNNE -0.114857 0.138109 -0.832 0.405612   
## WindSpeed9am -0.009697 0.003348 -2.896 0.003779 \*\*   
## WindSpeed3pm -0.031631 0.003420 -9.250 < 0.0000000000000002 \*\*\*  
## Humidity9am 0.004408 0.002457 1.794 0.072793 .   
## Humidity3pm 0.053608 0.002642 20.288 < 0.0000000000000002 \*\*\*  
## Pressure9am 0.155921 0.015084 10.337 < 0.0000000000000002 \*\*\*  
## Pressure3pm -0.202641 0.015090 -13.429 < 0.0000000000000002 \*\*\*  
## Cloud9am 0.046136 0.011099 4.157 0.0000323 \*\*\*  
## Cloud3pm 0.184612 0.011719 15.753 < 0.0000000000000002 \*\*\*  
## Temp9am 0.033266 0.018081 1.840 0.065784 .   
## Temp3pm 0.017248 0.021090 0.818 0.413446   
## RainTodayYes 0.532746 0.058751 9.068 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20884 on 19602 degrees of freedom  
## Residual deviance: 13656 on 19491 degrees of freedom  
## AIC: 13880  
##   
## Number of Fisher Scoring iterations: 6

emptymod = glm(RainTomorrow ~1, train, family = "binomial")   
summary(emptymod)

##   
## Call:  
## glm(formula = RainTomorrow ~ 1, family = "binomial", data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7133 -0.7133 -0.7133 -0.7133 1.7282   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.23901 0.01711 -72.39 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20884 on 19602 degrees of freedom  
## Residual deviance: 20884 on 19602 degrees of freedom  
## AIC: 20886  
##   
## Number of Fisher Scoring iterations: 4

#forward  
forwardmod = stepAIC(emptymod, direction = "forward", scope=list(upper=allmod,lower=emptymod), trace = TRUE)

## Start: AIC=20886.22  
## RainTomorrow ~ 1  
##   
## Df Deviance AIC  
## + Humidity3pm 1 16413 16417  
## + Cloud3pm 1 17794 17798  
## + Cloud9am 1 18780 18784  
## + RainToday 1 19155 19159  
## + Humidity9am 1 19368 19372  
## + Rainfall 1 19799 19803  
## + Pressure9am 1 19883 19887  
## + WindGustSpeed 1 19928 19932  
## + Pressure3pm 1 20067 20071  
## + Temp3pm 1 20160 20164  
## + MaxTemp 1 20404 20408  
## + WindDir9am 15 20544 20576  
## + MinTemp 1 20705 20709  
## + WindGustDir 15 20690 20722  
## + WindSpeed9am 1 20724 20728  
## + WindDir3pm 15 20710 20742  
## + WindSpeed3pm 1 20753 20757  
## + Month 11 20821 20845  
## + Year 10 20840 20862  
## + Temp9am 1 20877 20881  
## <none> 20884 20886  
## + Day 30 20847 20909  
##   
## Step: AIC=16416.84  
## RainTomorrow ~ Humidity3pm  
##   
## Df Deviance AIC  
## + WindGustSpeed 1 15353 15359  
## + Pressure3pm 1 15436 15442  
## + Pressure9am 1 15570 15576  
## + Cloud3pm 1 15771 15777  
## + RainToday 1 16054 16060  
## + WindDir3pm 15 16080 16114  
## + Cloud9am 1 16158 16164  
## + Rainfall 1 16176 16182  
## + WindDir9am 15 16170 16204  
## + WindSpeed9am 1 16226 16232  
## + WindGustDir 15 16212 16246  
## + MinTemp 1 16266 16272  
## + WindSpeed3pm 1 16267 16273  
## + Temp9am 1 16324 16330  
## + MaxTemp 1 16339 16345  
## + Temp3pm 1 16356 16362  
## + Month 11 16372 16398  
## + Year 10 16388 16412  
## + Humidity9am 1 16407 16413  
## <none> 16413 16417  
## + Day 30 16378 16442  
##   
## Step: AIC=15359.25  
## RainTomorrow ~ Humidity3pm + WindGustSpeed  
##   
## Df Deviance AIC  
## + Cloud3pm 1 14874 14882  
## + Pressure3pm 1 14891 14899  
## + Pressure9am 1 15052 15060  
## + WindDir3pm 15 15116 15152  
## + Cloud9am 1 15159 15167  
## + WindSpeed3pm 1 15168 15176  
## + RainToday 1 15170 15178  
## + WindDir9am 15 15180 15216  
## + Rainfall 1 15244 15252  
## + WindGustDir 15 15232 15268  
## + Temp3pm 1 15268 15276  
## + WindSpeed9am 1 15274 15282  
## + MinTemp 1 15275 15283  
## + MaxTemp 1 15280 15288  
## + Temp9am 1 15305 15313  
## + Humidity9am 1 15316 15324  
## <none> 15353 15359  
## + Year 10 15337 15363  
## + Month 11 15338 15366  
## + Day 30 15319 15385  
##   
## Step: AIC=14881.67  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm  
##   
## Df Deviance AIC  
## + Pressure3pm 1 14446 14456  
## + Pressure9am 1 14614 14624  
## + WindDir3pm 15 14684 14722  
## + RainToday 1 14717 14727  
## + Temp3pm 1 14735 14745  
## + WindSpeed3pm 1 14744 14754  
## + WindDir9am 15 14724 14762  
## + MaxTemp 1 14759 14769  
## + MinTemp 1 14769 14779  
## + Rainfall 1 14780 14790  
## + Temp9am 1 14789 14799  
## + WindSpeed9am 1 14791 14801  
## + WindGustDir 15 14778 14816  
## + Cloud9am 1 14844 14854  
## + Humidity9am 1 14850 14860  
## <none> 14874 14882  
## + Year 10 14859 14887  
## + Month 11 14862 14892  
## + Day 30 14838 14906  
##   
## Step: AIC=14455.71  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm  
##   
## Df Deviance AIC  
## + Pressure9am 1 14297 14309  
## + RainToday 1 14311 14323  
## + WindSpeed3pm 1 14329 14341  
## + WindDir3pm 15 14351 14391  
## + WindDir9am 15 14355 14395  
## + Rainfall 1 14384 14396  
## + Month 11 14367 14399  
## + WindSpeed9am 1 14394 14406  
## + Humidity9am 1 14402 14414  
## + Cloud9am 1 14420 14432  
## + Temp3pm 1 14428 14440  
## + WindGustDir 15 14407 14447  
## + MinTemp 1 14436 14448  
## + MaxTemp 1 14438 14450  
## + Temp9am 1 14443 14455  
## <none> 14446 14456  
## + Year 10 14433 14463  
## + Day 30 14410 14480  
##   
## Step: AIC=14308.59  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am  
##   
## Df Deviance AIC  
## + RainToday 1 14081 14095  
## + WindSpeed3pm 1 14177 14191  
## + Rainfall 1 14208 14222  
## + WindDir3pm 15 14218 14260  
## + Humidity9am 1 14255 14269  
## + Cloud9am 1 14259 14273  
## + Month 11 14241 14275  
## + WindDir9am 15 14236 14278  
## + WindSpeed9am 1 14266 14280  
## + WindGustDir 15 14262 14304  
## + MinTemp 1 14290 14304  
## <none> 14297 14309  
## + Temp3pm 1 14295 14309  
## + Temp9am 1 14296 14310  
## + MaxTemp 1 14296 14310  
## + Year 10 14284 14316  
## + Day 30 14261 14333  
##   
## Step: AIC=14094.56  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday  
##   
## Df Deviance AIC  
## + WindSpeed3pm 1 13973 13989  
## + WindDir3pm 15 14014 14058  
## + WindDir9am 15 14016 14060  
## + WindSpeed9am 1 14051 14067  
## + Cloud9am 1 14060 14076  
## + Month 11 14040 14076  
## + Rainfall 1 14074 14090  
## + MinTemp 1 14074 14090  
## + Humidity9am 1 14075 14091  
## + WindGustDir 15 14050 14094  
## + Temp3pm 1 14078 14094  
## <none> 14081 14095  
## + Temp9am 1 14079 14095  
## + MaxTemp 1 14079 14095  
## + Year 10 14070 14104  
## + Day 30 14045 14119  
##   
## Step: AIC=13989.06  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm  
##   
## Df Deviance AIC  
## + WindDir9am 15 13914 13960  
## + WindDir3pm 15 13916 13962  
## + Cloud9am 1 13952 13970  
## + MinTemp 1 13958 13976  
## + Month 11 13940 13978  
## + WindSpeed9am 1 13964 13982  
## + Temp9am 1 13964 13982  
## + Temp3pm 1 13966 13984  
## + Rainfall 1 13968 13986  
## + MaxTemp 1 13969 13987  
## + Humidity9am 1 13969 13987  
## <none> 13973 13989  
## + WindGustDir 15 13950 13996  
## + Year 10 13963 13999  
## + Day 30 13938 14014  
##   
## Step: AIC=13960.46  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am  
##   
## Df Deviance AIC  
## + MinTemp 1 13883 13931  
## + Cloud9am 1 13892 13940  
## + Temp9am 1 13895 13943  
## + Temp3pm 1 13898 13946  
## + WindDir3pm 15 13874 13950  
## + MaxTemp 1 13904 13952  
## + Rainfall 1 13907 13955  
## + WindSpeed9am 1 13908 13956  
## + Month 11 13889 13957  
## + Humidity9am 1 13912 13960  
## <none> 13914 13960  
## + Year 10 13904 13970  
## + WindGustDir 15 13896 13972  
## + Day 30 13878 13984  
##   
## Step: AIC=13931.33  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp  
##   
## Df Deviance AIC  
## + Month 11 13820 13890  
## + WindDir3pm 15 13825 13903  
## + Cloud9am 1 13863 13913  
## + MaxTemp 1 13873 13923  
## + Humidity9am 1 13874 13924  
## + WindSpeed9am 1 13875 13925  
## + Rainfall 1 13878 13928  
## + Temp9am 1 13879 13929  
## + Temp3pm 1 13879 13929  
## <none> 13883 13931  
## + WindGustDir 15 13856 13934  
## + Year 10 13872 13940  
## + Day 30 13848 13956  
##   
## Step: AIC=13890.22  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp +   
## Month  
##   
## Df Deviance AIC  
## + WindDir3pm 15 13768 13868  
## + Cloud9am 1 13797 13869  
## + Humidity9am 1 13807 13879  
## + WindSpeed9am 1 13808 13880  
## + Rainfall 1 13814 13886  
## <none> 13820 13890  
## + Temp9am 1 13819 13891  
## + MaxTemp 1 13819 13891  
## + Temp3pm 1 13820 13892  
## + WindGustDir 15 13795 13895  
## + Year 10 13808 13898  
## + Day 30 13784 13914  
##   
## Step: AIC=13867.77  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp +   
## Month + WindDir3pm  
##   
## Df Deviance AIC  
## + Cloud9am 1 13745 13847  
## + WindSpeed9am 1 13757 13859  
## + Humidity9am 1 13758 13860  
## + Rainfall 1 13761 13863  
## <none> 13768 13868  
## + MaxTemp 1 13766 13868  
## + Temp9am 1 13767 13869  
## + Temp3pm 1 13768 13870  
## + Year 10 13755 13875  
## + WindGustDir 15 13752 13882  
## + Day 30 13732 13892  
##   
## Step: AIC=13847.15  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp +   
## Month + WindDir3pm + Cloud9am  
##   
## Df Deviance AIC  
## + WindSpeed9am 1 13735 13839  
## + Rainfall 1 13739 13843  
## + Humidity9am 1 13741 13845  
## <none> 13745 13847  
## + Temp3pm 1 13745 13849  
## + Temp9am 1 13745 13849  
## + MaxTemp 1 13745 13849  
## + Year 10 13732 13854  
## + WindGustDir 15 13730 13862  
## + Day 30 13710 13872  
##   
## Step: AIC=13838.92  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp +   
## Month + WindDir3pm + Cloud9am + WindSpeed9am  
##   
## Df Deviance AIC  
## + Rainfall 1 13729 13835  
## + Humidity9am 1 13733 13839  
## <none> 13735 13839  
## + MaxTemp 1 13734 13840  
## + Temp9am 1 13735 13841  
## + Temp3pm 1 13735 13841  
## + Year 10 13722 13846  
## + WindGustDir 15 13720 13854  
## + Day 30 13699 13863  
##   
## Step: AIC=13834.58  
## RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm + Pressure3pm +   
## Pressure9am + RainToday + WindSpeed3pm + WindDir9am + MinTemp +   
## Month + WindDir3pm + Cloud9am + WindSpeed9am + Rainfall  
##   
## Df Deviance AIC  
## <none> 13729 13835  
## + Humidity9am 1 13727 13835  
## + MaxTemp 1 13728 13836  
## + Temp9am 1 13728 13836  
## + Temp3pm 1 13728 13836  
## + Year 10 13715 13841  
## + WindGustDir 15 13714 13850  
## + Day 30 13693 13859

summary(forwardmod)

##   
## Call:  
## glm(formula = RainTomorrow ~ Humidity3pm + WindGustSpeed + Cloud3pm +   
## Pressure3pm + Pressure9am + RainToday + WindSpeed3pm + WindDir9am +   
## MinTemp + Month + WindDir3pm + Cloud9am + WindSpeed9am +   
## Rainfall, family = "binomial", data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.4370 -0.5478 -0.2883 -0.1156 3.1848   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 37.732667 3.974035 9.495 < 0.0000000000000002 \*\*\*  
## Humidity3pm 0.055631 0.001505 36.963 < 0.0000000000000002 \*\*\*  
## WindGustSpeed 0.052479 0.002423 21.662 < 0.0000000000000002 \*\*\*  
## Cloud3pm 0.181985 0.011533 15.779 < 0.0000000000000002 \*\*\*  
## Pressure3pm -0.198572 0.013698 -14.497 < 0.0000000000000002 \*\*\*  
## Pressure9am 0.153179 0.013670 11.206 < 0.0000000000000002 \*\*\*  
## RainTodayYes 0.556491 0.056688 9.817 < 0.0000000000000002 \*\*\*  
## WindSpeed3pm -0.030682 0.003321 -9.238 < 0.0000000000000002 \*\*\*  
## WindDir9amW 0.019538 0.132124 0.148 0.882440   
## WindDir9amSSW -0.063186 0.138326 -0.457 0.647823   
## WindDir9amN 0.257517 0.121651 2.117 0.034273 \*   
## WindDir9amWSW 0.032057 0.136547 0.235 0.814389   
## WindDir9amSSE -0.223116 0.136080 -1.640 0.101089   
## WindDir9amESE -0.132224 0.144561 -0.915 0.360373   
## WindDir9amSE -0.136997 0.136373 -1.005 0.315101   
## WindDir9amS -0.257518 0.137768 -1.869 0.061593 .   
## WindDir9amNNE 0.412057 0.130628 3.154 0.001608 \*\*   
## WindDir9amNW 0.041994 0.129019 0.325 0.744814   
## WindDir9amWNW -0.018687 0.134812 -0.139 0.889756   
## WindDir9amSW 0.027903 0.132883 0.210 0.833684   
## WindDir9amNE 0.094334 0.138458 0.681 0.495669   
## WindDir9amE -0.041809 0.138479 -0.302 0.762717   
## WindDir9amNNW -0.024895 0.131472 -0.189 0.849815   
## MinTemp 0.043551 0.004710 9.248 < 0.0000000000000002 \*\*\*  
## MonthJanuary -0.211094 0.112536 -1.876 0.060685 .   
## MonthFebruary 0.080081 0.114678 0.698 0.484982   
## MonthMarch 0.120891 0.110439 1.095 0.273675   
## MonthApril 0.415425 0.112460 3.694 0.000221 \*\*\*  
## MonthMay 0.428838 0.113509 3.778 0.000158 \*\*\*  
## MonthJune 0.431928 0.116561 3.706 0.000211 \*\*\*  
## MonthJuly 0.545005 0.117471 4.640 0.00000349 \*\*\*  
## MonthAugust 0.522827 0.119221 4.385 0.00001158 \*\*\*  
## MonthSeptember 0.388165 0.116973 3.318 0.000905 \*\*\*  
## MonthOctober 0.348486 0.114232 3.051 0.002283 \*\*   
## MonthNovember 0.217718 0.111315 1.956 0.050480 .   
## WindDir3pmW 0.027108 0.117650 0.230 0.817771   
## WindDir3pmWNW -0.024974 0.119449 -0.209 0.834386   
## WindDir3pmE -0.313432 0.133083 -2.355 0.018514 \*   
## WindDir3pmSSW -0.100370 0.130983 -0.766 0.443511   
## WindDir3pmWSW -0.025845 0.122715 -0.211 0.833189   
## WindDir3pmS -0.166117 0.125523 -1.323 0.185702   
## WindDir3pmSE -0.014343 0.122818 -0.117 0.907033   
## WindDir3pmESE -0.396296 0.132549 -2.990 0.002791 \*\*   
## WindDir3pmENE -0.279491 0.130329 -2.144 0.031993 \*   
## WindDir3pmNE -0.501907 0.128244 -3.914 0.00009089 \*\*\*  
## WindDir3pmSSE -0.292520 0.129310 -2.262 0.023687 \*   
## WindDir3pmSW 0.000205 0.126140 0.002 0.998704   
## WindDir3pmN -0.254959 0.122786 -2.076 0.037852 \*   
## WindDir3pmNNW 0.143076 0.120332 1.189 0.234433   
## WindDir3pmNNE -0.187786 0.129761 -1.447 0.147851   
## Cloud9am 0.047964 0.010326 4.645 0.00000340 \*\*\*  
## WindSpeed9am -0.010217 0.003179 -3.214 0.001310 \*\*   
## Rainfall 0.007603 0.003087 2.463 0.013787 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20884 on 19602 degrees of freedom  
## Residual deviance: 13729 on 19550 degrees of freedom  
## AIC: 13835  
##   
## Number of Fisher Scoring iterations: 6

#backstep  
backmod = stepAIC(allmod, direction = "backward", trace = TRUE)

## Start: AIC=13880.44  
## RainTomorrow ~ Year + Month + Day + MinTemp + MaxTemp + Rainfall +   
## WindGustDir + WindGustSpeed + WindDir9am + WindDir3pm + WindSpeed9am +   
## WindSpeed3pm + Humidity9am + Humidity3pm + Pressure9am +   
## Pressure3pm + Cloud9am + Cloud3pm + Temp9am + Temp3pm + RainToday  
##   
## Df Deviance AIC  
## - Day 30 13693 13857  
## - WindGustDir 15 13672 13866  
## - Year 10 13671 13875  
## - Temp3pm 1 13657 13879  
## <none> 13656 13880  
## - Humidity9am 1 13660 13882  
## - Temp9am 1 13660 13882  
## - MaxTemp 1 13660 13882  
## - Rainfall 1 13661 13883  
## - MinTemp 1 13663 13885  
## - WindSpeed9am 1 13665 13887  
## - WindDir3pm 15 13698 13892  
## - Cloud9am 1 13674 13896  
## - WindDir9am 15 13711 13905  
## - Month 11 13711 13913  
## - RainToday 1 13737 13959  
## - WindSpeed3pm 1 13743 13965  
## - Pressure9am 1 13765 13987  
## - Pressure3pm 1 13842 14064  
## - Cloud3pm 1 13916 14138  
## - Humidity3pm 1 14088 14310  
## - WindGustSpeed 1 14134 14356  
##   
## Step: AIC=13857.17  
## RainTomorrow ~ Year + Month + MinTemp + MaxTemp + Rainfall +   
## WindGustDir + WindGustSpeed + WindDir9am + WindDir3pm + WindSpeed9am +   
## WindSpeed3pm + Humidity9am + Humidity3pm + Pressure9am +   
## Pressure3pm + Cloud9am + Cloud3pm + Temp9am + Temp3pm + RainToday  
##   
## Df Deviance AIC  
## - WindGustDir 15 13708 13842  
## - Year 10 13707 13851  
## - Temp3pm 1 13694 13856  
## <none> 13693 13857  
## - Humidity9am 1 13696 13858  
## - MaxTemp 1 13697 13859  
## - Temp9am 1 13697 13859  
## - Rainfall 1 13698 13860  
## - MinTemp 1 13700 13862  
## - WindSpeed9am 1 13702 13864  
## - WindDir3pm 15 13735 13869  
## - Cloud9am 1 13711 13873  
## - WindDir9am 15 13745 13879  
## - Month 11 13748 13890  
## - RainToday 1 13774 13936  
## - WindSpeed3pm 1 13781 13943  
## - Pressure9am 1 13803 13965  
## - Pressure3pm 1 13880 14042  
## - Cloud3pm 1 13952 14114  
## - Humidity3pm 1 14124 14286  
## - WindGustSpeed 1 14173 14335  
##   
## Step: AIC=13841.79  
## RainTomorrow ~ Year + Month + MinTemp + MaxTemp + Rainfall +   
## WindGustSpeed + WindDir9am + WindDir3pm + WindSpeed9am +   
## WindSpeed3pm + Humidity9am + Humidity3pm + Pressure9am +   
## Pressure3pm + Cloud9am + Cloud3pm + Temp9am + Temp3pm + RainToday  
##   
## Df Deviance AIC  
## - Year 10 13721 13835  
## - Temp3pm 1 13708 13840  
## <none> 13708 13842  
## - Humidity9am 1 13711 13843  
## - MaxTemp 1 13711 13843  
## - Temp9am 1 13712 13844  
## - Rainfall 1 13713 13845  
## - MinTemp 1 13714 13846  
## - WindSpeed9am 1 13716 13848  
## - Cloud9am 1 13726 13858  
## - WindDir9am 15 13756 13860  
## - WindDir3pm 15 13758 13862  
## - Month 11 13762 13874  
## - RainToday 1 13789 13921  
## - WindSpeed3pm 1 13798 13930  
## - Pressure9am 1 13816 13948  
## - Pressure3pm 1 13892 14024  
## - Cloud3pm 1 13968 14100  
## - Humidity3pm 1 14137 14269  
## - WindGustSpeed 1 14200 14332  
##   
## Step: AIC=13835.12  
## RainTomorrow ~ Month + MinTemp + MaxTemp + Rainfall + WindGustSpeed +   
## WindDir9am + WindDir3pm + WindSpeed9am + WindSpeed3pm + Humidity9am +   
## Humidity3pm + Pressure9am + Pressure3pm + Cloud9am + Cloud3pm +   
## Temp9am + Temp3pm + RainToday  
##   
## Df Deviance AIC  
## - Temp3pm 1 13722 13834  
## <none> 13721 13835  
## - Humidity9am 1 13724 13836  
## - MaxTemp 1 13724 13836  
## - Temp9am 1 13725 13837  
## - Rainfall 1 13727 13839  
## - MinTemp 1 13728 13840  
## - WindSpeed9am 1 13730 13842  
## - Cloud9am 1 13738 13850  
## - WindDir9am 15 13769 13853  
## - WindDir3pm 15 13771 13855  
## - Month 11 13774 13866  
## - RainToday 1 13803 13915  
## - WindSpeed3pm 1 13812 13924  
## - Pressure9am 1 13831 13943  
## - Pressure3pm 1 13907 14019  
## - Cloud3pm 1 13982 14094  
## - Humidity3pm 1 14149 14261  
## - WindGustSpeed 1 14213 14325  
##   
## Step: AIC=13833.62  
## RainTomorrow ~ Month + MinTemp + MaxTemp + Rainfall + WindGustSpeed +   
## WindDir9am + WindDir3pm + WindSpeed9am + WindSpeed3pm + Humidity9am +   
## Humidity3pm + Pressure9am + Pressure3pm + Cloud9am + Cloud3pm +   
## Temp9am + RainToday  
##   
## Df Deviance AIC  
## <none> 13722 13834  
## - MaxTemp 1 13726 13836  
## - Temp9am 1 13726 13836  
## - Humidity9am 1 13727 13837  
## - Rainfall 1 13727 13837  
## - MinTemp 1 13728 13838  
## - WindSpeed9am 1 13730 13840  
## - Cloud9am 1 13739 13849  
## - WindDir9am 15 13770 13852  
## - WindDir3pm 15 13772 13854  
## - Month 11 13775 13865  
## - RainToday 1 13804 13914  
## - WindSpeed3pm 1 13814 13924  
## - Pressure9am 1 13842 13952  
## - Pressure3pm 1 13922 14032  
## - Cloud3pm 1 13984 14094  
## - WindGustSpeed 1 14217 14327  
## - Humidity3pm 1 14352 14462

summary(backmod)

##   
## Call:  
## glm(formula = RainTomorrow ~ Month + MinTemp + MaxTemp + Rainfall +   
## WindGustSpeed + WindDir9am + WindDir3pm + WindSpeed9am +   
## WindSpeed3pm + Humidity9am + Humidity3pm + Pressure9am +   
## Pressure3pm + Cloud9am + Cloud3pm + Temp9am + RainToday,   
## family = "binomial", data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.4299 -0.5499 -0.2871 -0.1135 3.2015   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 37.628563 3.994474 9.420 < 0.0000000000000002 \*\*\*  
## MonthJanuary -0.204161 0.112828 -1.809 0.070375 .   
## MonthFebruary 0.087496 0.115433 0.758 0.448465   
## MonthMarch 0.123342 0.111458 1.107 0.268458   
## MonthApril 0.397385 0.113145 3.512 0.000444 \*\*\*  
## MonthMay 0.409525 0.115335 3.551 0.000384 \*\*\*  
## MonthJune 0.417349 0.119188 3.502 0.000462 \*\*\*  
## MonthJuly 0.529736 0.120902 4.382 0.0000118 \*\*\*  
## MonthAugust 0.504967 0.122155 4.134 0.0000357 \*\*\*  
## MonthSeptember 0.368343 0.118993 3.095 0.001965 \*\*   
## MonthOctober 0.346875 0.114643 3.026 0.002481 \*\*   
## MonthNovember 0.218038 0.111474 1.956 0.050471 .   
## MinTemp 0.031025 0.011867 2.614 0.008939 \*\*   
## MaxTemp -0.024747 0.012469 -1.985 0.047172 \*   
## Rainfall 0.007232 0.003098 2.335 0.019564 \*   
## WindGustSpeed 0.053670 0.002471 21.719 < 0.0000000000000002 \*\*\*  
## WindDir9amW 0.015709 0.132316 0.119 0.905492   
## WindDir9amSSW -0.068764 0.138631 -0.496 0.619882   
## WindDir9amN 0.257064 0.121745 2.112 0.034729 \*   
## WindDir9amWSW 0.034297 0.136726 0.251 0.801932   
## WindDir9amSSE -0.231354 0.136188 -1.699 0.089360 .   
## WindDir9amESE -0.133908 0.144720 -0.925 0.354817   
## WindDir9amSE -0.136617 0.136548 -1.001 0.317066   
## WindDir9amS -0.267424 0.137943 -1.939 0.052544 .   
## WindDir9amNNE 0.416981 0.130759 3.189 0.001428 \*\*   
## WindDir9amNW 0.037789 0.129139 0.293 0.769811   
## WindDir9amWNW -0.016503 0.134958 -0.122 0.902675   
## WindDir9amSW 0.024881 0.133065 0.187 0.851671   
## WindDir9amNE 0.094230 0.138561 0.680 0.496463   
## WindDir9amE -0.039525 0.138620 -0.285 0.775544   
## WindDir9amNNW -0.027250 0.131607 -0.207 0.835968   
## WindDir3pmW 0.025071 0.117641 0.213 0.831240   
## WindDir3pmWNW -0.029020 0.119548 -0.243 0.808202   
## WindDir3pmE -0.316376 0.133366 -2.372 0.017681 \*   
## WindDir3pmSSW -0.089568 0.131046 -0.683 0.494299   
## WindDir3pmWSW -0.024031 0.122672 -0.196 0.844692   
## WindDir3pmS -0.155443 0.125673 -1.237 0.216129   
## WindDir3pmSE -0.013708 0.123089 -0.111 0.911327   
## WindDir3pmESE -0.393853 0.132811 -2.966 0.003022 \*\*   
## WindDir3pmENE -0.278525 0.130560 -2.133 0.032899 \*   
## WindDir3pmNE -0.499564 0.128582 -3.885 0.000102 \*\*\*  
## WindDir3pmSSE -0.286073 0.129521 -2.209 0.027195 \*   
## WindDir3pmSW 0.005013 0.126199 0.040 0.968315   
## WindDir3pmN -0.256826 0.122950 -2.089 0.036720 \*   
## WindDir3pmNNW 0.151494 0.120517 1.257 0.208742   
## WindDir3pmNNE -0.186301 0.130030 -1.433 0.151928   
## WindSpeed9am -0.009516 0.003299 -2.885 0.003920 \*\*   
## WindSpeed3pm -0.032305 0.003389 -9.533 < 0.0000000000000002 \*\*\*  
## Humidity9am 0.005044 0.002266 2.226 0.026030 \*   
## Humidity3pm 0.051958 0.002137 24.312 < 0.0000000000000002 \*\*\*  
## Pressure9am 0.157845 0.014524 10.868 < 0.0000000000000002 \*\*\*  
## Pressure3pm -0.203240 0.014587 -13.933 < 0.0000000000000002 \*\*\*  
## Cloud9am 0.046399 0.011041 4.202 0.0000264 \*\*\*  
## Cloud3pm 0.183262 0.011579 15.827 < 0.0000000000000002 \*\*\*  
## Temp9am 0.037969 0.017404 2.182 0.029134 \*   
## RainTodayYes 0.533482 0.058399 9.135 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20884 on 19602 degrees of freedom  
## Residual deviance: 13722 on 19547 degrees of freedom  
## AIC: 13834  
##   
## Number of Fisher Scoring iterations: 6