

BinForDecisionTree

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
hour_2013 <- read.csv(file="D:/Capston/2013_Hour_By_Tract.csv", header=TRUE,
sep=",",
                      stringsAsFactors = FALSE)

hour_2013$dteday <- as.Date(hour_2013$dteday)
hour_2013$mday <- as.integer(format(hour_2013$dteday, "%d"))

hour_2013 <- subset(hour_2013, select = -c(yr, atemp))

hour_2013$season <- as.factor(hour_2013$season)
hour_2013$weather <- as.factor(hour_2013$weather)
hour_2013$holiday <- as.factor(hour_2013$holiday)
hour_2013$weekday <- as.factor(hour_2013$weekday)
hour_2013$workingday <- as.factor(hour_2013$workingday)
hour_2013$mth <- as.factor(hour_2013$mth)
hour_2013$tractID <- as.factor(hour_2013$tractID)

hour_2013$casual <- log1p(hour_2013$casual)
hour_2013$registered <- log1p(hour_2013$registered)
hour_2013$cnt <- log1p(hour_2013$cnt)

str(hour_2013)

## 'data.frame': 399458 obs. of 17 variables:
## $ dteday : Date, format: "2013-01-01" "2013-01-01" ...
## $ tractID : Factor w/ 113 levels "1","2","3","4",...: 2 8 16 17 19 21 29
30 32 36 ...
## $ jday : int 1 1 1 1 1 1 1 1 1 1 ...
## $ season : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
## $ mth : Factor w/ 12 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1
...
## $ hr : int 0 0 0 0 0 0 0 0 0 0 ...
## $ holiday : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ weekday : Factor w/ 7 levels "0","1","2","3",...: 3 3 3 3 3 3 3 3 3 3
...
## $ workingday: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ weather : Factor w/ 4 levels "1","2","3","4": 2 2 2 2 2 2 2 2 2 2 ...
```

```
## $ temp      : num  3.3 3.3 3.3 3.3 3.3 3.3 3.3 3.3 3.3 3.3 ...
## $ humidity   : int  68 68 68 68 68 68 68 68 68 68 ...
## $ windspeed  : num  11.1 11.1 11.1 11.1 11.1 11.1 11.1 11.1 11.1 11.1 ...
## $ casual     : num   0 0 0 0 0 ...
## $ registered: num  0.693 0.693 0.693 0.693 1.099 ...
## $ cnt        : num  0.693 0.693 0.693 0.693 1.099 ...
## $ mday       : int   1 1 1 1 1 1 1 1 1 1 ...
```

`summary(hour_2013)`

```
##      dteday      tractID      jday      season
## Min.   :2013-01-01    36      : 7888    Min.   : 1.0    1: 80222
## 1st Qu.:2013-04-19    72      : 7733    1st Qu.:109.0    2:103493
## Median :2013-07-14    38      : 7701    Median :195.0    3:113245
## Mean   :2013-07-09    71      : 7571    Mean   :190.7    4:102498
## 3rd Qu.:2013-10-02    27      : 7480    3rd Qu.:275.0
## Max.   :2013-12-31    39      : 7477    Max.   :365.0
##                               (Other):353608
##      mth      hr      holiday      weekday      workingday
## 8      : 38405    Min.   : 0.00    0:388645    0:55954    0:125945
## 10     : 37866    1st Qu.: 9.00    1: 10813    1:55069    1:273513
## 7      : 37477    Median :14.00
## 9      : 37363    Mean   :13.22
## 6      : 35690    3rd Qu.:18.00
## 5      : 35577    Max.   :23.00
## (Other):177080
## weather      temp      humidity      windspeed
## 1:148141    Min.   :-8.90    Min.   : 12.0    Min.   : 1.00
## 2:223743    1st Qu.: 8.30    1st Qu.: 44.0    1st Qu.: 9.30
## 3: 26198    Median :17.20    Median : 58.0    Median :14.80
## 4: 1376    Mean   :16.52    Mean   : 58.8    Mean   :15.36
##           3rd Qu.:24.40    3rd Qu.: 74.0    3rd Qu.:19.91
##           Max.   :35.00    Max.   :100.0    Max.   :53.70
##
##      casual      registered      cnt      mday
## Min.   :0.0000    Min.   :0.0000    Min.   :0.6931    Min.   : 1.00
## 1st Qu.:0.0000    1st Qu.:0.6931    1st Qu.:0.6931    1st Qu.: 8.00
## Median :0.0000    Median :1.3863    Median :1.3863    Median :16.00
## Mean   :0.3943    Mean   :1.4314    Mean   :1.5676    Mean   :15.67
## 3rd Qu.:0.6931    3rd Qu.:1.9459    3rd Qu.:2.0794    3rd Qu.:23.00
## Max.   :5.4889    Max.   :5.1358    Max.   :5.6699    Max.   :31.00
##
```

```
library(rpart)
library(rattle)
```

```
## Warning: package 'rattle' was built under R version 3.4.2
```

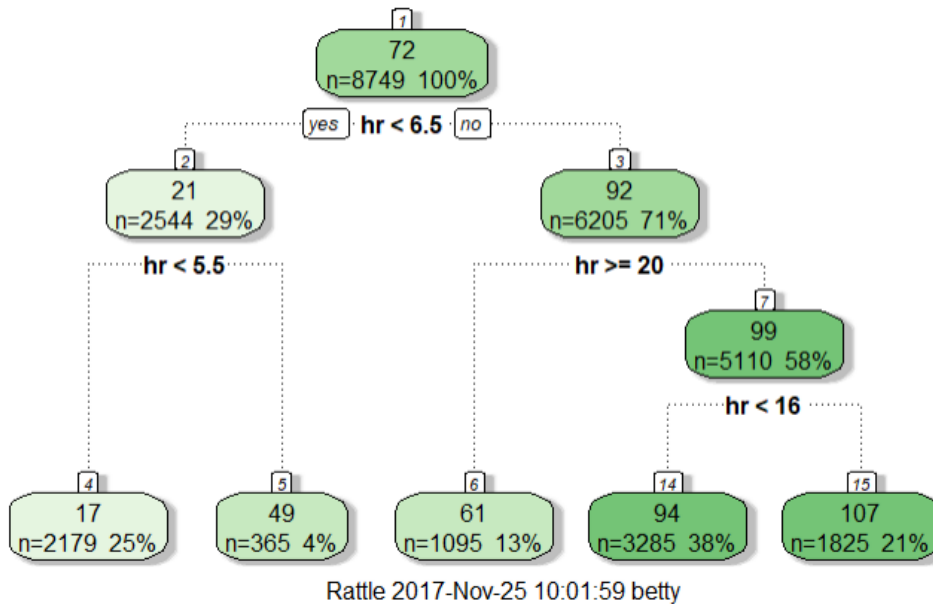
```
## Rattle: A free graphical interface for data science with R.
## Version 5.1.0 Copyright (c) 2006-2017 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
library(rpart.plot)

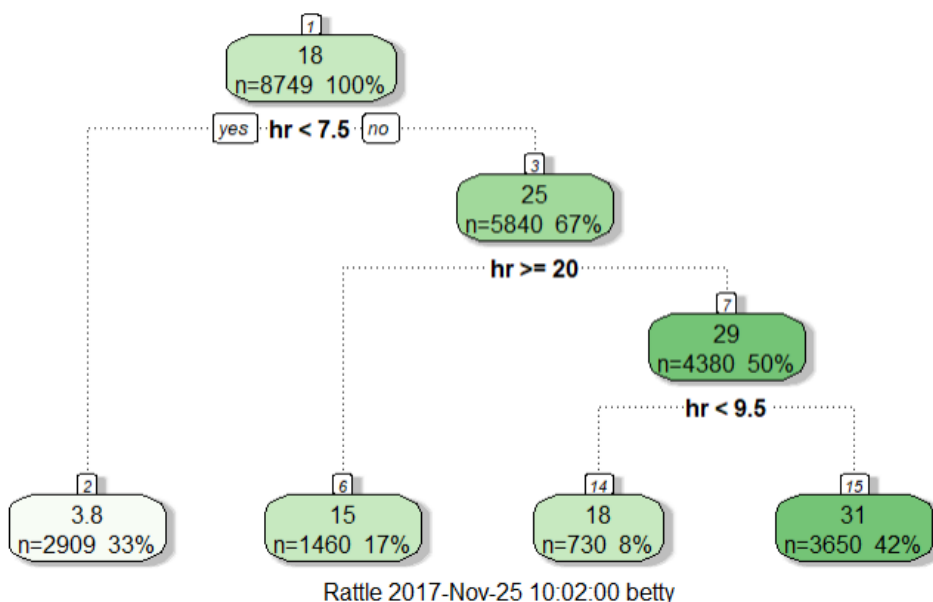
## Warning: package 'rpart.plot' was built under R version 3.4.2

library(RColorBrewer)

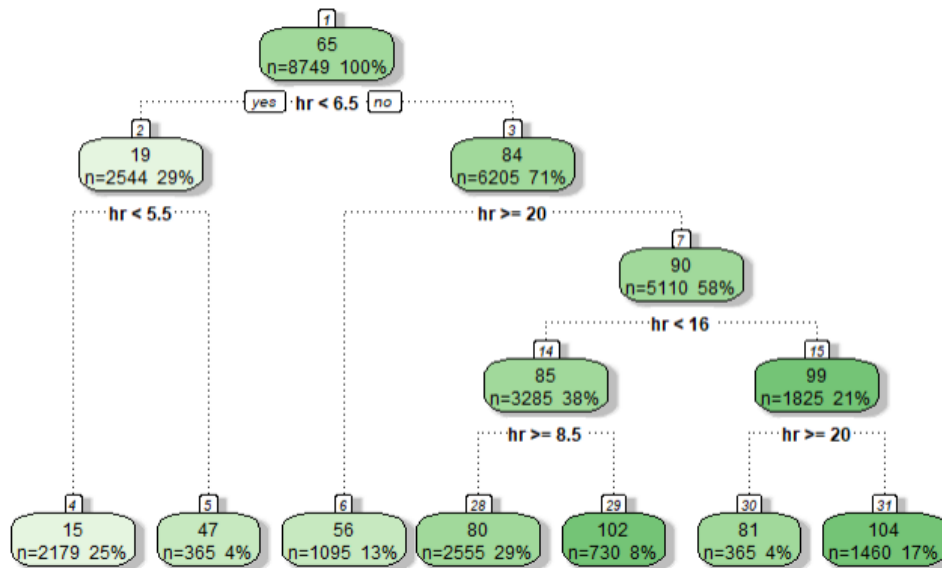
cnt <- aggregate(cnt~jday+hr, data=hour_2013, FUN=sum)
model <- rpart(cnt~hr, data = cnt)
fancyRpartPlot(model)
```



```
casual <- aggregate(casual~jday+hr, data=hour_2013, FUN=sum)
model <- rpart(casual~hr, data = casual)
fancyRpartPlot(model)
```



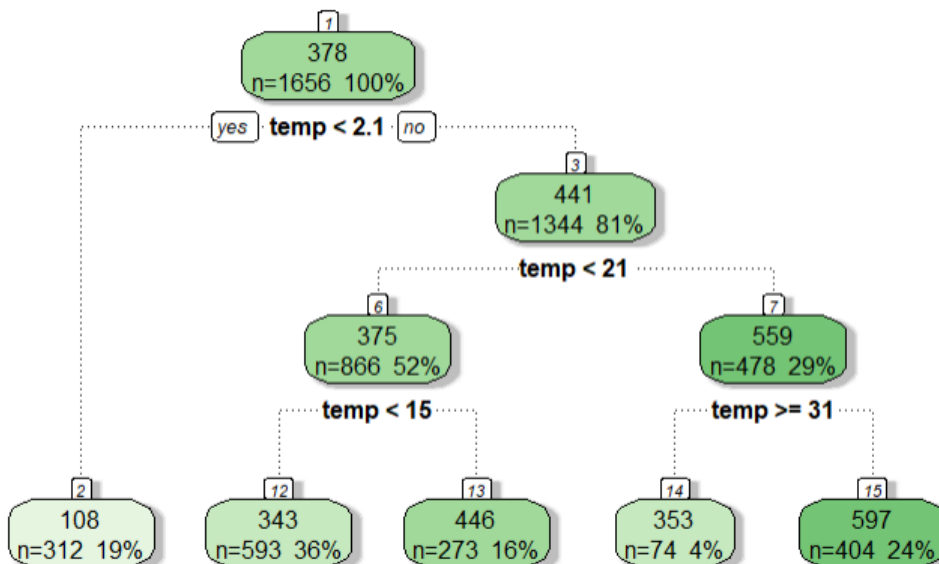
```
registered <- aggregate(registered~jday+hr, data=hour_2013, FUN=sum)
model <- rpart(registered~hr, data = registered)
fancyRpartPlot(model)
```



Rattle 2017-Nov-25 10:02:01 betty

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

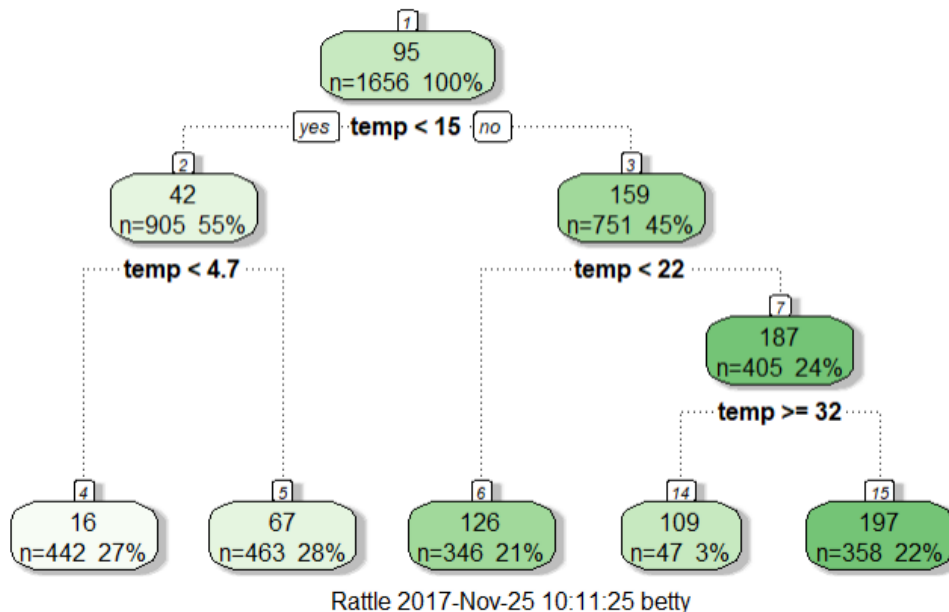
```
cnt <- aggregate(cnt~hr+temp, data=hour_2013, FUN=sum)
model <- rpart(cnt~temp, data = cnt)
fancyRpartPlot(model)
```



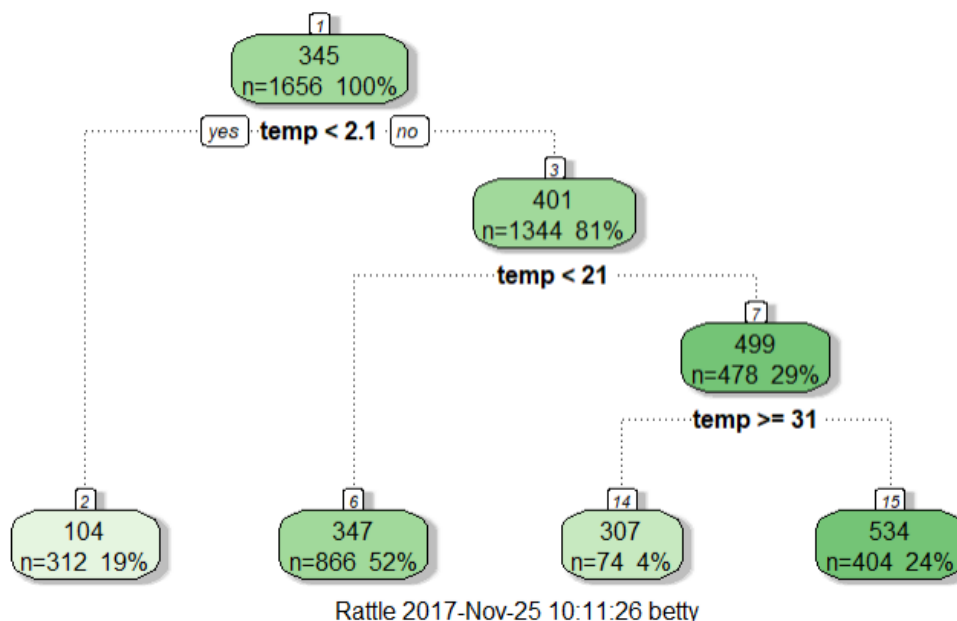
Rattle 2017-Nov-25 10:11:24 betty

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

```
casual <- aggregate(casual~hr+temp, data=hour_2013, FUN=sum)
model <- rpart(casual~temp, data = casual)
fancyRpartPlot(model)
```

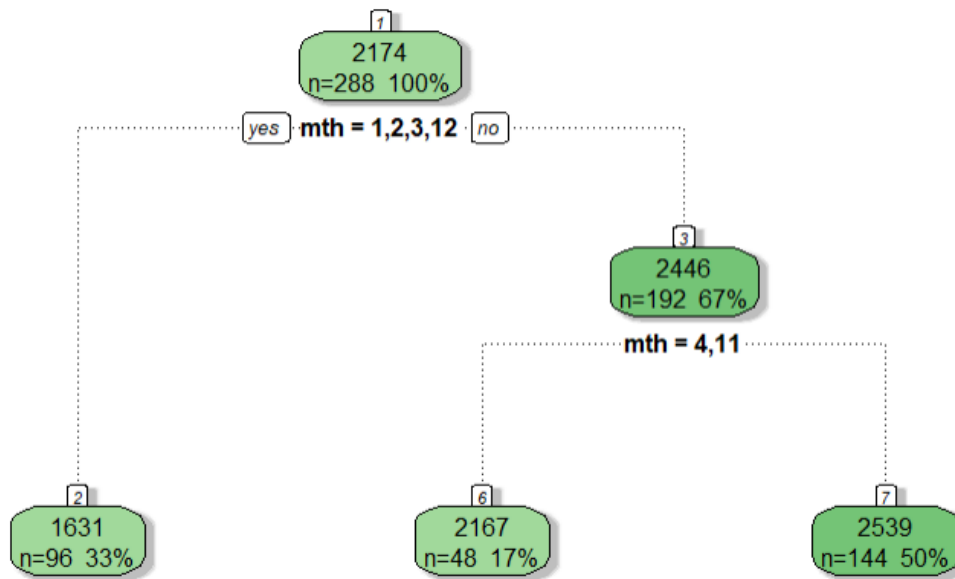


```
registered <- aggregate(registered~hr+temp, data=hour_2013, FUN=sum)
model <- rpart(registered~temp, data = registered)
fancyRpartPlot(model)
```



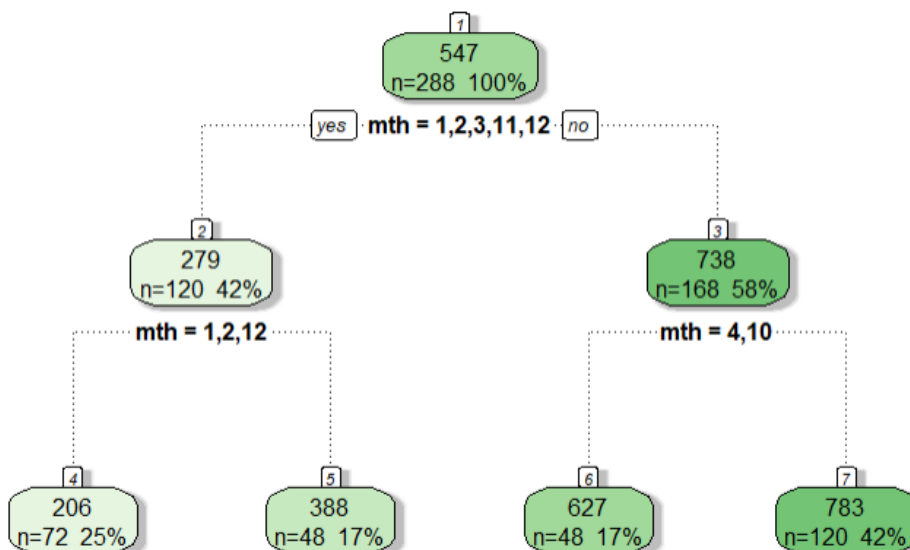
```
library(rpart)
library(rattle)
library(rpart.plot)
library(RColorBrewer)

cnt <- aggregate(cnt~hr+mth, data=hour_2013, FUN=sum)
model <- rpart(cnt~mth, data = cnt)
fancyRpartPlot(model)
```



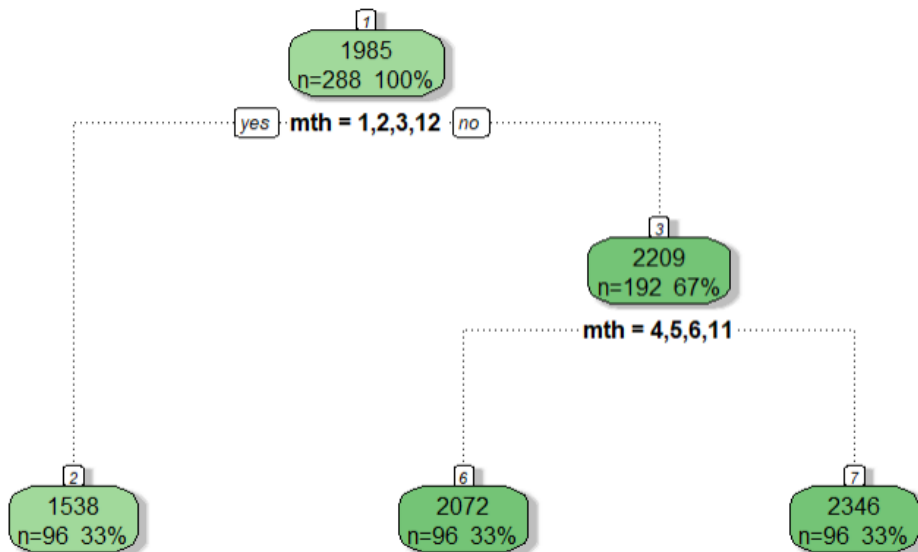
Rattle 2017-Nov-25 10:03:55 betty

```
casual <- aggregate(casual~hr+mth, data=hour_2013, FUN=sum)
model <- rpart(casual~mth, data = casual)
fancyRpartPlot(model)
```



Rattle 2017-Nov-25 10:03:55 betty

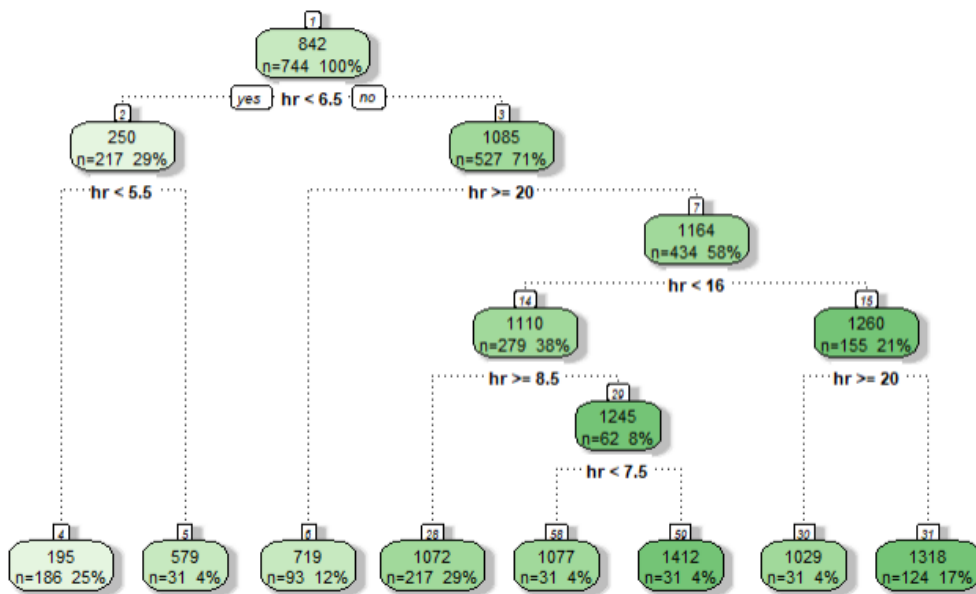
```
registered <- aggregate(registered~hr+mth, data=hour_2013, FUN=sum)
model <- rpart(registered~mth, data = registered)
fancyRpartPlot(model)
```



Rattle 2017-Nov-25 10:03:56 betty

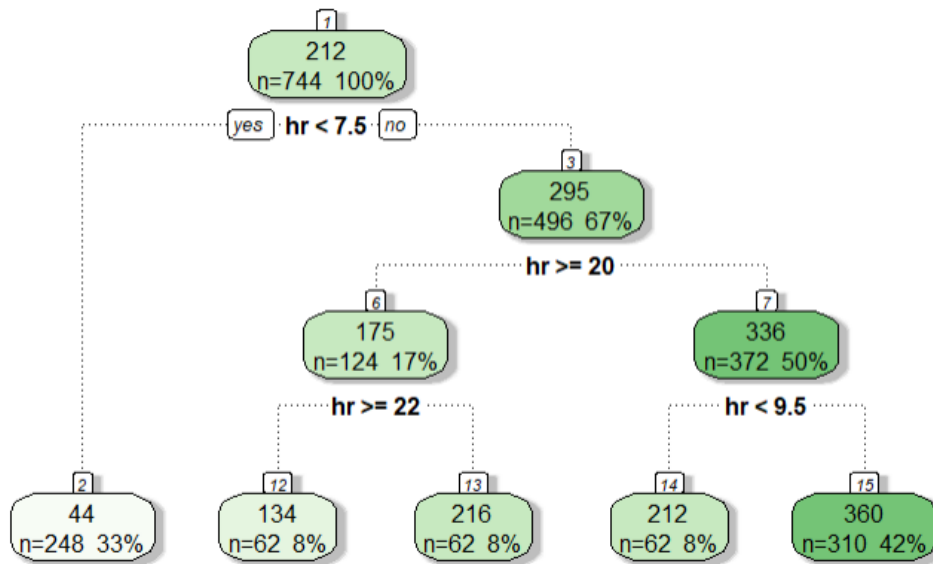
```
library(rpart)
library(rattle)
library(rpart.plot)
library(RColorBrewer)

cnt <- aggregate(cnt~mday+hr, data=hour_2013, FUN=sum)
model <- rpart(cnt~hr, data = cnt)
fancyRpartPlot(model)
```



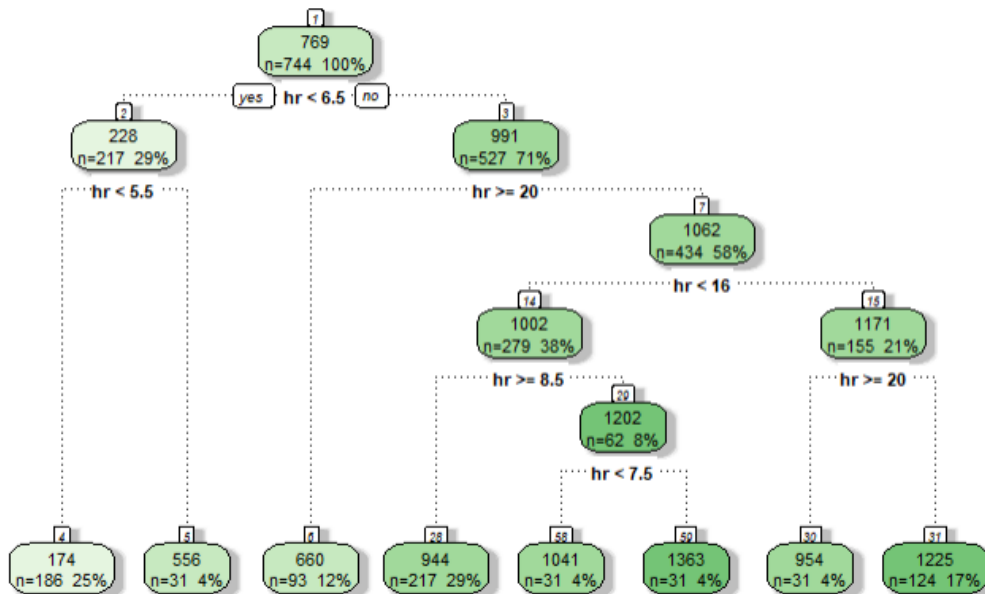
Rattle 2017-Nov-25 10:05:12 betty

```
casual <- aggregate(casual~mday+hr, data=hour_2013, FUN=sum)
model <- rpart(casual~hr, data = casual)
fancyRpartPlot(model)
```



Rattle 2017-Nov-25 10:05:13 betty

```
registered <- aggregate(registered~mday+hr, data=hour_2013, FUN=sum)
model <- rpart(registered~hr, data = registered)
fancyRpartPlot(model)
```



Rattle 2017-Nov-25 10:05:14 betty