Milliman

# Easy Tree-sy, continued

CAS RPM Seminar

San Diego, California

Peggy Brinkmann, FCAS, MAAA

## Advantages of Trees

- Easy to Interpret
- Automatic Variable Selection
- Automatic Interactions and Local Effects
- Handles Missing Values
- Handles Outliers
- Handles Monotonic Transformation

Milliman

## Applications of Trees

**Enhancing GLMs.** Decision trees do not require a lot of pre-processing of predictor variables to handle missing values and non-linear relationships – making them ideal for quickly screening a large number of potential predictor variables. Analyzing the "residuals" from a GLMs with a decision tree can help you identify additional transformations and/or interactions to improve the fit of your model, and as a check to make sure that no "signal" has been missed.

**Portfolio diagnostics.** A decision tree run with loss ratio as the target variable can help you identify segments with good profitability (to target marketing efforts) and poor profitability (for pricing revisions and/or underwriting action).

**Checking/Quality Control.** Ever tried to figure out why your complex calculated value (e.g. rerated premium, credit score) doesn't match to another source (e.g. company inforce premium, vendor calculated score)? Use the input variables as the predictors and the difference in values as the target variable, and you'll quickly find the sources of the discrepancies.



## Demo

- Commercial lines automobile carrier
- Eight years of loss and premium data
- Want better segment underwriting and pricing
- Have looked at variables in one-way analysis
- Now what?

## Data, part 1

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | reccount | polno | state | effdate | effyr | expdate | cancel | lapse | poltype | vehtype | deluxmulti | color | V96 |
| 2 | 1 | FLHP0000001 | OH | 20140415 | 2014 | 20150415 | NA | NA | delux | full | No | White | G |
| 3 | 2 | FLHP0000002 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | full | No | White | G |
| 4 | 3 | FLHP0000006 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | full | No | White | G |
| 5 | 4 | FLHP0000009 | OH | 20140407 | 2014 | 20150407 | NA | NA | delux | stand | No | White | H |
| 6 | 5 | FLHP0000010 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | full | No | White | G |
| 7 | 6 | FLHP0000011 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | full | No | White | G |
| 8 | 7 | FLHP0000012 | OH | 20140428 | 2014 | 20150428 | NA | NA | delux | econ | No | White | G |
| 9 | 8 | FLHP0000017 | OH | 20140421 | 2014 | 20150421 | NA | NA | delux | full | No | Blue | H |
| 10 | 9 | FLHP0000018 | OH | 20140425 | 2014 | 20150425 | NA | NA | delux | full | No | White | H |
| 11 | 10 | FLHP0000019 | OH | 20140414 | 2014 | 20150414 | NA | NA | delux | full | No | White | G |
| 12 | 11 | FLHP0000020 | OH | 20140407 | 2014 | 20150407 | NA | NA | delux | econ | No | White | H |
| 13 | 12 | FLHP0000022 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | full | No | White | G |
| 14 | 13 | FLHP0000023 | OH | 20140417 | 2014 | 20150417 | NA | NA | delux | full | No | White | G |
| 15 | 14 | FLHP0000025 | OH | 20140411 | 2014 | 20150411 | NA | NA | delux | econ | No | White | G |
| 16 | 15 | FLHP0000027 | OH | 20140408 | 2014 | 20150408 | NA | NA | delux | stand | No | White | H |
| 17 | 16 | FLHP0000028 | OH | 20140408 | 2014 | 20150408 | NA | NA | delux | stand | No | White | G |
| 18 | 17 | FLHP0000029 | OH | 20140408 | 2014 | 20150408 | NA | NA | delux | stand | No | Blue | G |

Milliman

5

## Data, part 2

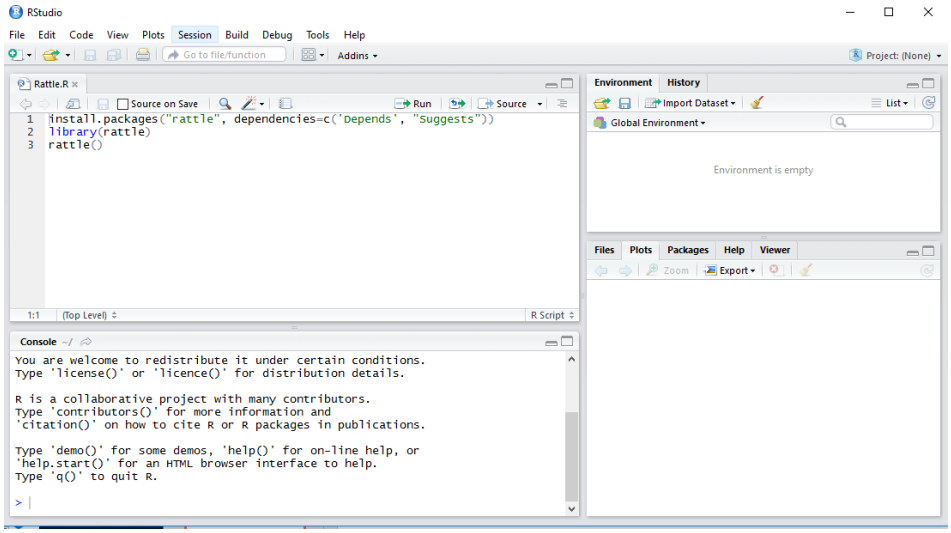
| N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| limit | prem | incloss | claimct | exposure | modelyea | latepay | mileage_g | countynar | size_num | deductibl | power_nu | policyage f |
| 100 | 573 | 0 | 0 | 1 | 1996 | NA | 4 | F | 1500 | | 1 | 0 |
| 100 | 1246 | 0 | 0 | 1 | 1978 | NA | 3 | G | 1500 | | 1 | 0 |
| 100 | 991 | 0 | 0 | 1 | 1986 | NA | 2 | H | 1500 | | 1 | 0 |
| 25 | 739 | 0 | 0 | 1 | 2004 | NA | 5 | B | 1500 | | 1 | 0 |
| 100 | 732 | 0 | 0 | 1 | 1998 | NA | 4 | F | 2000 | | 1 | 0 |
| 25 | 742 | 0 | 0 | 1 | 2000 | NA | 4 | F | 1500 | | 1 | 0 |
| 100 | 1448 | 0 | 0 | 1 | 1990 | NA | 5 | B | 2000 | | 1 | 0 |
| 100 | 2646 | 0 | 0 | 1 | 1985 | NA | 4 | A | 2000 | | 1 | 0 |
| 500 | 1138 | 0 | 0 | 1 | 2006 | NA | 3 | D | 4000 | | 1 | 0 |
| 100 | 1773 | 0 | 0 | 1 | 1988 | NA | 4 | D | 2000 | | 1 | 0 |
| 500 | 895 | 0 | 0 | 1 | 2001 | NA | 6 | F | 3000 | | 1 | 0 |
| 100 | 1063 | 0 | 0 | 1 | 2005 | NA | | E | 2000 | | 1 | 0 |
| 100 | 1219 | 0 | 0 | 1 | 1960 | NA | 3 | B | 2000 | | 1 | 0 |
| 100 | 1537 | 0 | 0 | 1 | 1983 | NA | 4 | F | 1500 | | 1 | 0 |
| 500 | 1585 | 0 | 0 | 1 | 1990 | NA | 3 | B | 2500 | | 1 | 0 |
| 100 | 717 | 0 | 0 | 1 | 1987 | NA | 3 | B | 2000 | | 1 | 0 |
| 100 | 2585 | 0 | 0 | 1 | 1974 | NA | 5 | A | 2000 | | 1 | 0 |

Milliman

6

# Screenshots

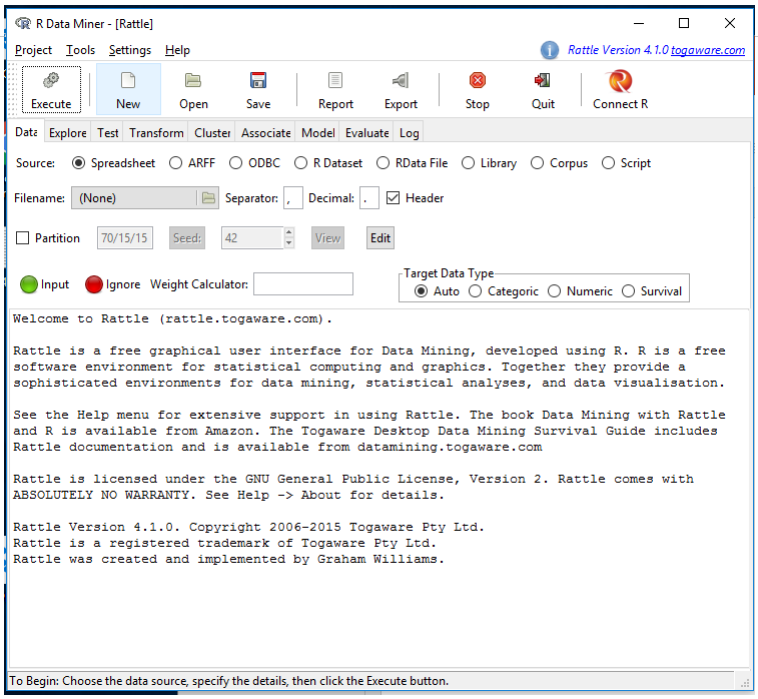





This is all the R code you need to write to run Rattle!

Click Run to submit

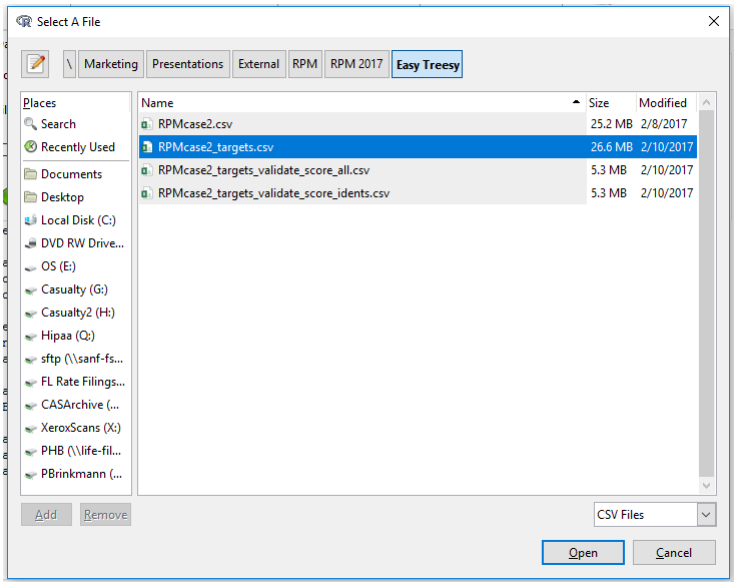
Milliman



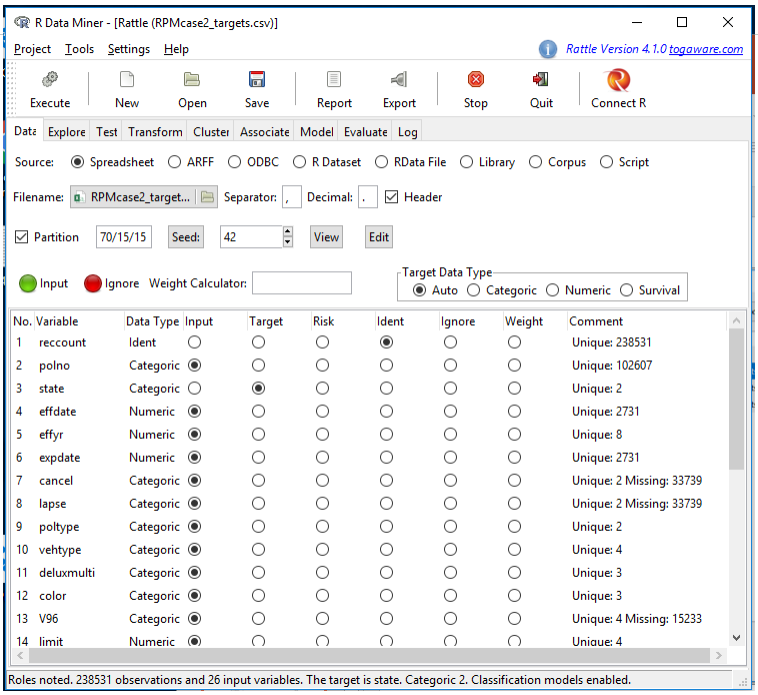

Select your data format, filename, delimiter (if applicable), and if it has a header with variable names

Then click "Execute"

Milliman

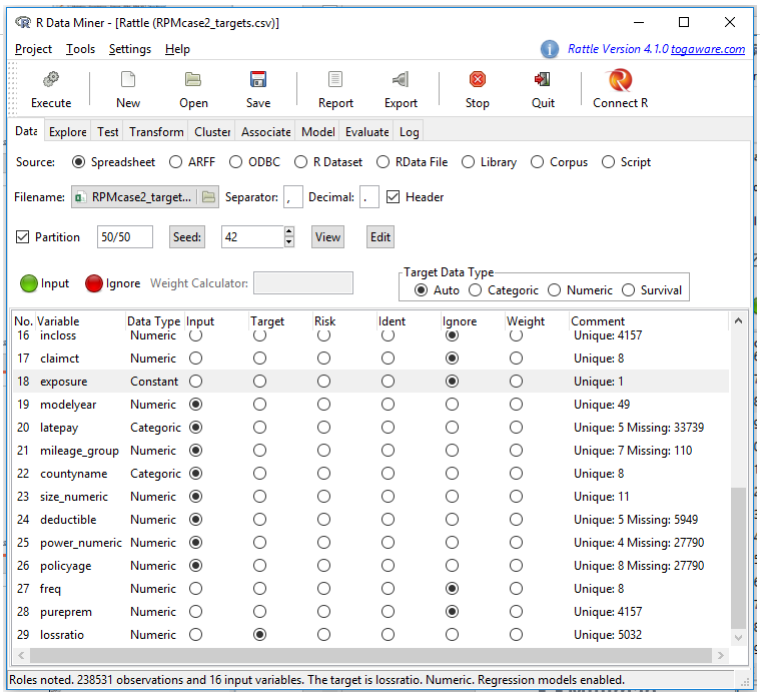



Milliman



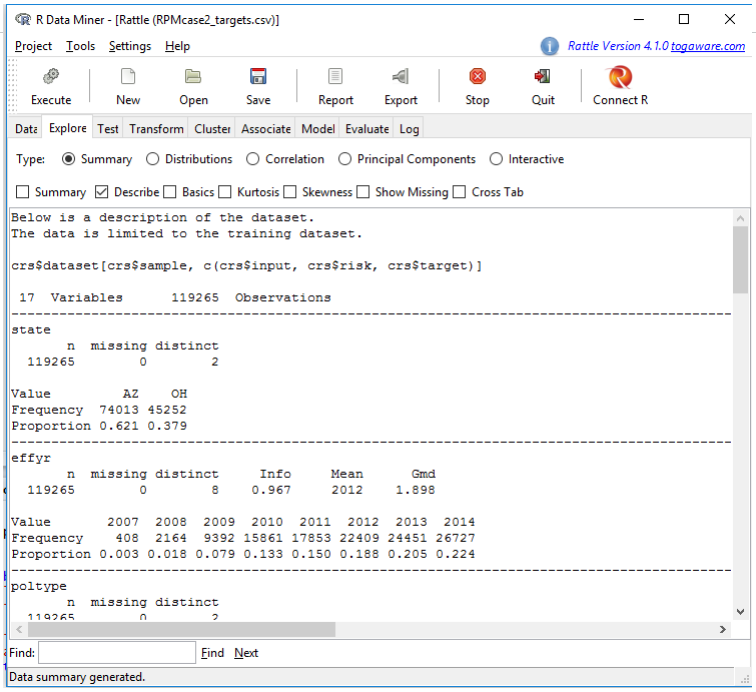

Rattle will guess how to use each variable; here reset the target to loss ratio, specify the weight (prem), and ignore variables like frequency, exposure, and cancellations

Milliman



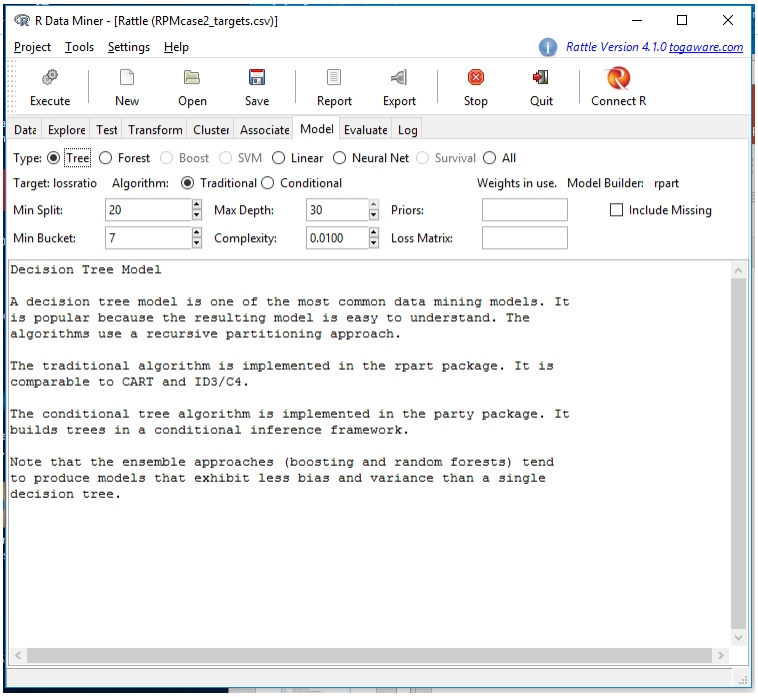

After making selections, hit Execute to refresh the data – see note at the bottom of the window

Milliman




Basic statistics for each variable created from Summary - Describe

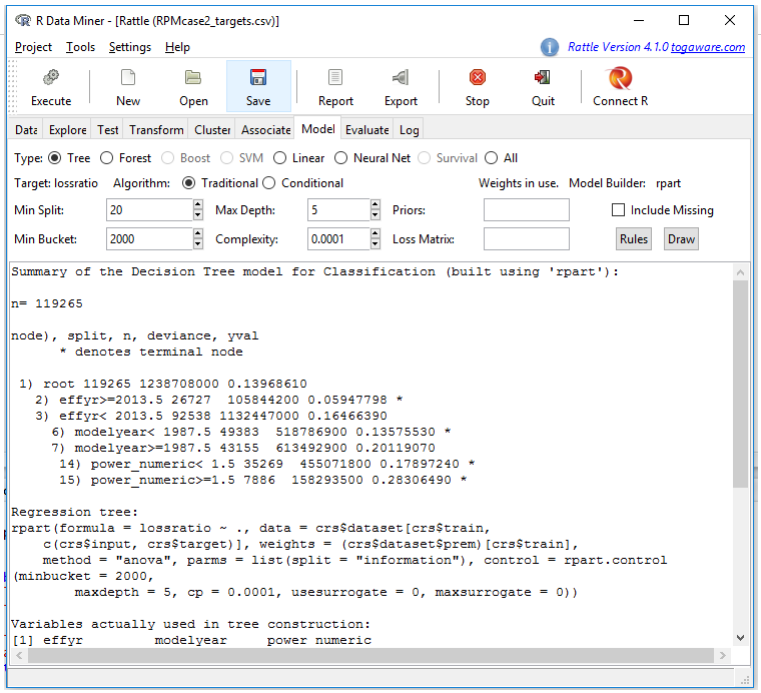
Milliman




Rattle can fit other types of models besides trees, as well as k-means clusters and association analysis

Note the default parameters for trees – not appropriate for loss ratio analysis!

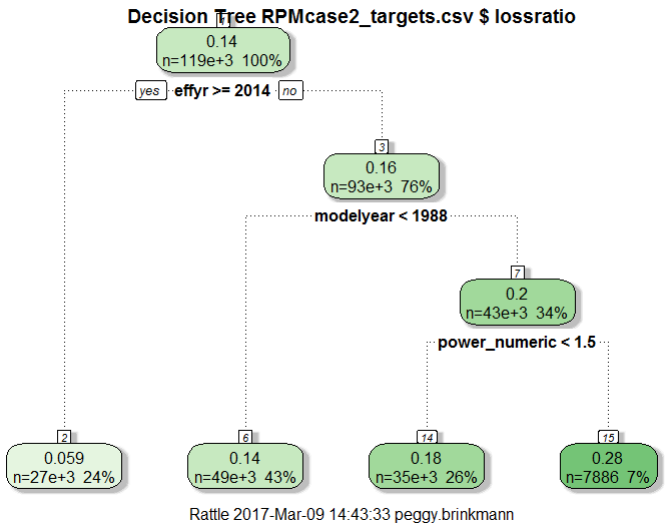
Milliman




Reset parameters to require at least 1K policies per node, and limit tree to 5 levels of splits

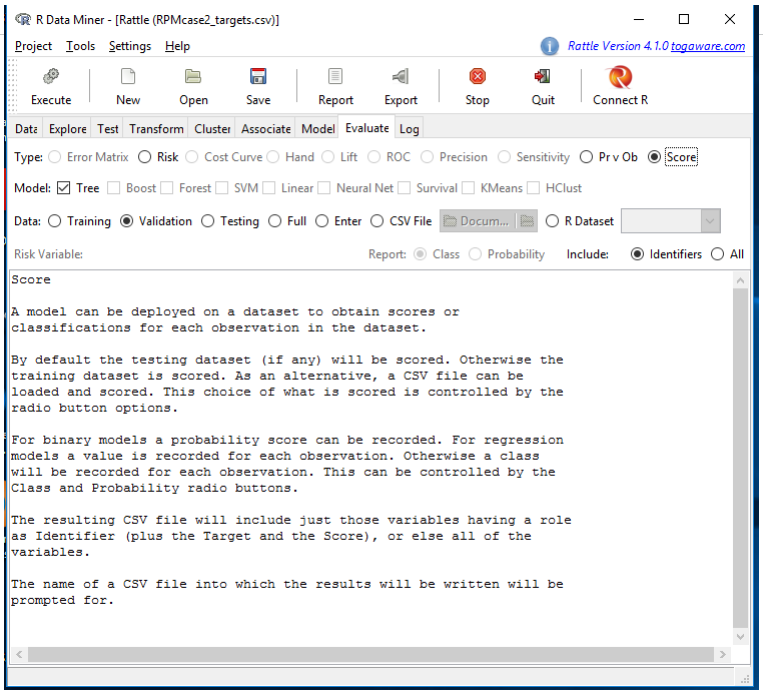
For regression tree, complexity parameter is the minimum change in R squared to make a split

Milliman




Output from the Draw function is sent to R Studio plot window

Milliman




Evaluate tab has variety of measures that can be applied to training, validation, or full dataset

For regression tree, the most useful is Score, which applies the tree nodes to a dataset

Milliman

| | A | B | C | D | AD |
|---|---|---|---|---|---|
| 1 | reccount | polno | state | effdate | rpart |
| 2 | 1 | FLHP0000001 | OH | 20140415 | 0.059478 |
| 3 | 3 | FLHP0000006 | OH | 20140411 | 0.059478 |
| 4 | 4 | FLHP0000009 | OH | 20140407 | 0.059478 |
| 5 | 5 | FLHP0000010 | OH | 20140411 | 0.059478 |
| 6 | 8 | FLHP0000017 | OH | 20140421 | 0.059478 |
| 7 | 9 | FLHP0000018 | OH | 20140425 | 0.059478 |
| 8 | 10 | FLHP0000019 | OH | 20140414 | 0.059478 |
| 9 | 11 | FLHP0000020 | OH | 20140407 | 0.059478 |
| 10 | 13 | FLHP0000023 | OH | 20140417 | 0.059478 |
| 11 | 14 | FLHP0000025 | OH | 20140411 | 0.059478 |
| 12 | 15 | FLHP0000027 | OH | 20140408 | 0.059478 |
| 13 | 21 | FLHP0000035 | OH | 20140409 | 0.059478 |
| 14 | 22 | FLHP0000036 | OH | 20140409 | 0.059478 |
| 15 | 24 | FLHP0000039 | OH | 20140428 | 0.059478 |
| 16 | 25 | FLHP0000041 | OH | 20140415 | 0.059478 |
| 17 | 28 | FLHP0000045 | OH | 20140415 | 0.059478 |
| 18 | 31 | FLHP0000051 | OH | 20140414 | 0.059478 |
| 19 | 34 | FLHP0000057 | OH | 20140411 | 0.059478 |

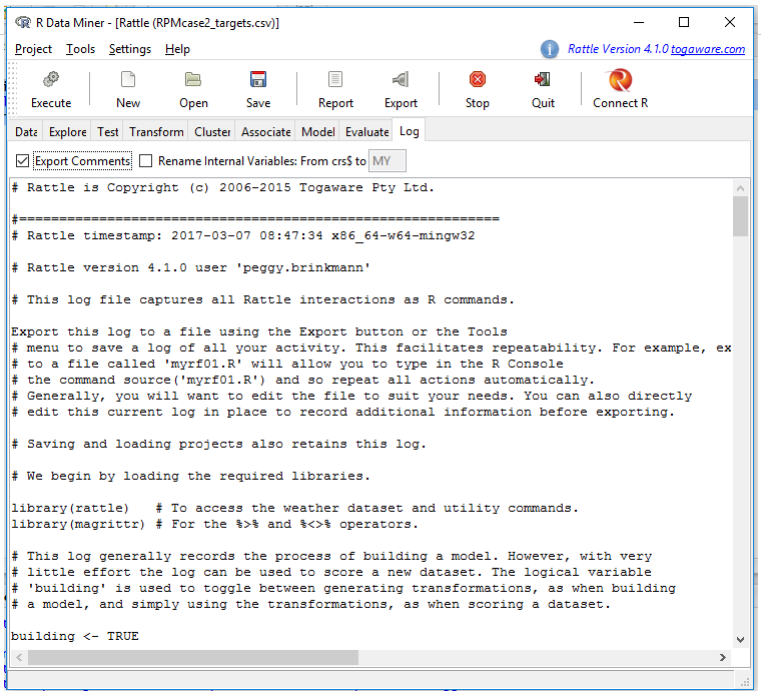
The column “rpart” shows the expected loss ratio for the node based on the training data

Milliman

| Row Labels | Sum of exposure | Sum of prem | Sum of incloss | Sum of Loss ratio |
|---|---|---|---|---|
| 0.059477976 | 26509 | 49133517 | 3088809 | 0.06286562 |
| 0.135755252 | 49354 | 88307588 | 12283788 | 0.139102293 |
| 0.178972424 | 35439 | 55420814 | 10251548 | 0.184976496 |
| 0.283064949 | 7964 | 15219797 | 3560380 | 0.233930847 |
| **Grand Total** | **119266** | **208081716** | **29184525** | **0.140255115** |

Summarize the data by node in Excel to see how the tree performs on new data

Milliman




We can get the R code to perform the analysis from the Log, like recording a macro in Excel

This can be a way to get a R program started, and you can modify and customize the code in R Studio

Milliman

Milliman

# Thank You

peggy.brinkmann@milliman.com
415-394-3726