Ratemaking Capstone Instructions

You will get these data files

- clms.RData
- pol_dates.RData
- pol_rating.RData
- states.csv

This data is all made up. It is for a fictitious claims-made Architects and Engineers program.
You have data for policy years 2007 through 2016. The program is countrywide. You have some policy data, split into two sets, and claims data.

The `pol_dates` files have policy number, inception date, and expiration date.
The `pol_rating` files have some more policy attributes like revenue, employee count, etc. The `clms` file has claim info, on a ground-up basis, for the training set. The `states` file gives you state groupings. In real life you wouldn't have this. But this is not real life. :)

The claims valuation date is Dec 31, 2016.

We do not have limits on the policies or how much was actually charged.

Your task is to come up with a pricing model for this program. You will document your analysis of frequency and severity and why you chose the model you use. Then, you will calculate pure premium values for each of the "test" policies with the following per occurence limits (and unlimited aggregate):

- $1m
- $2m
- $5m

You will use the following skills in the capstone project

- general R skills, like indexing and arithmetic calculations
- loading data files
- visualization do to "exploratory data analysis"
- dpylr for data manipulation
- dpylr and base R for data aggregation
- decision trees to guide analysis
- glms for frequency model(s)
- RMarkdown for report writing
- curve fitting for severity distribution(s)

This capstone will also require you to do real actuarial work. Doing one-way analyses and figuring out whether data should be grouped will take some time. Deciding which model to use in ratemaking can be as much art as science. . . no one answer will be "correct".

I am really excited for you to be part of this capstone project. You will learn a lot of R and I hope you have a blast!

– Adam