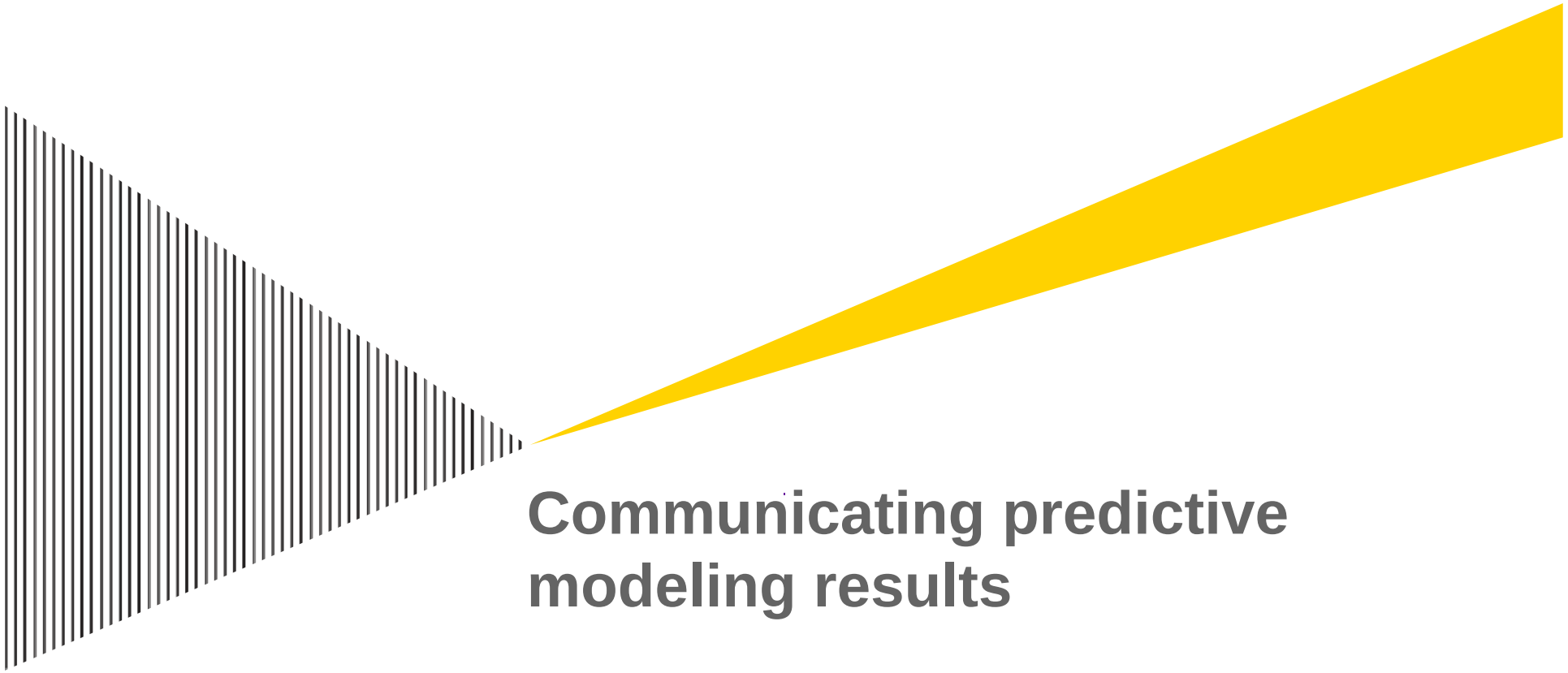


Communicating predictive modeling results

Chuck Boucek

CAS Seminar on Predictive Modeling – San Diego, California
6-7 October 2008

ERNST & YOUNG
Quality In Everything We Do

Agenda

- Data validation
- Hypothesis building
- Model building
- Model testing
- Monitoring model results

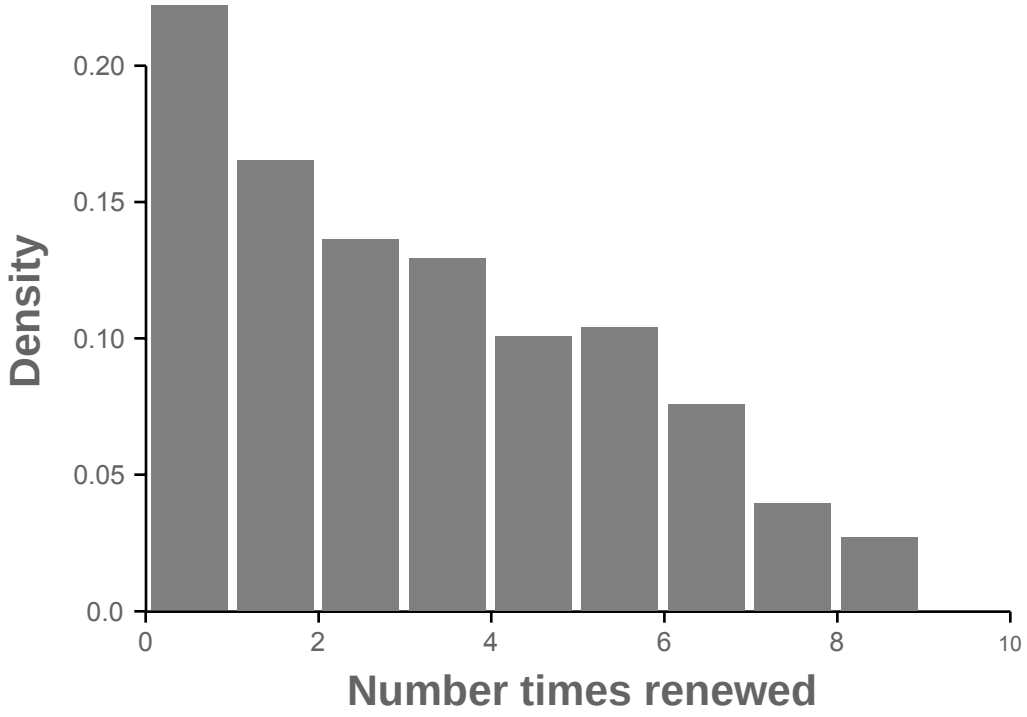
Data validation

Goals

- Validate reasonableness of data
- Understand key patterns in data
- Understand changes in data and underlying business through time

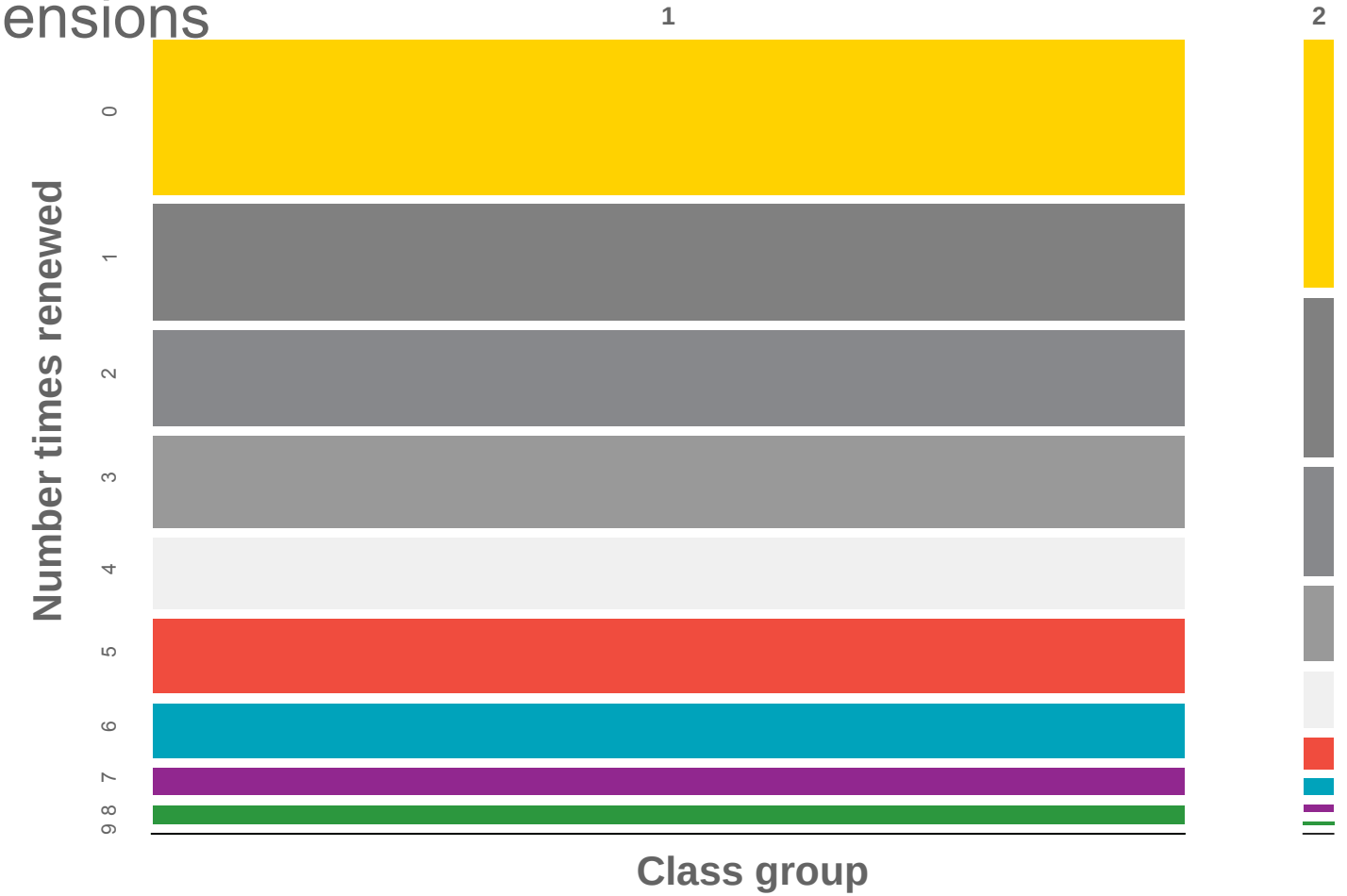
Data validation

Histogram is a simple tool for reasonability testing of modeling database



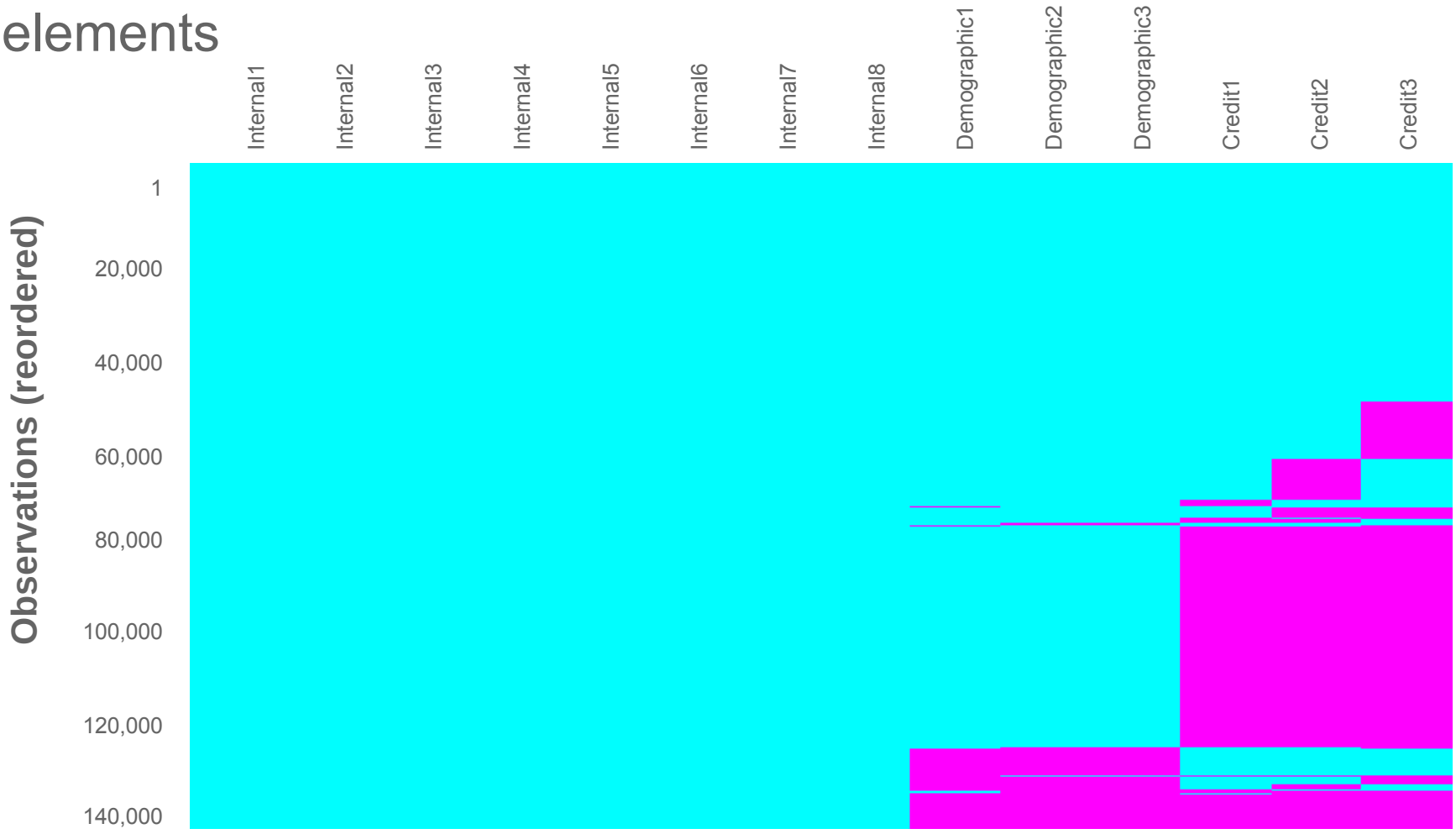
Data validation

Mosaic Plot shows the distribution of predictors in two dimensions



Data validation

Missing data plot shows the relationship of missing data elements

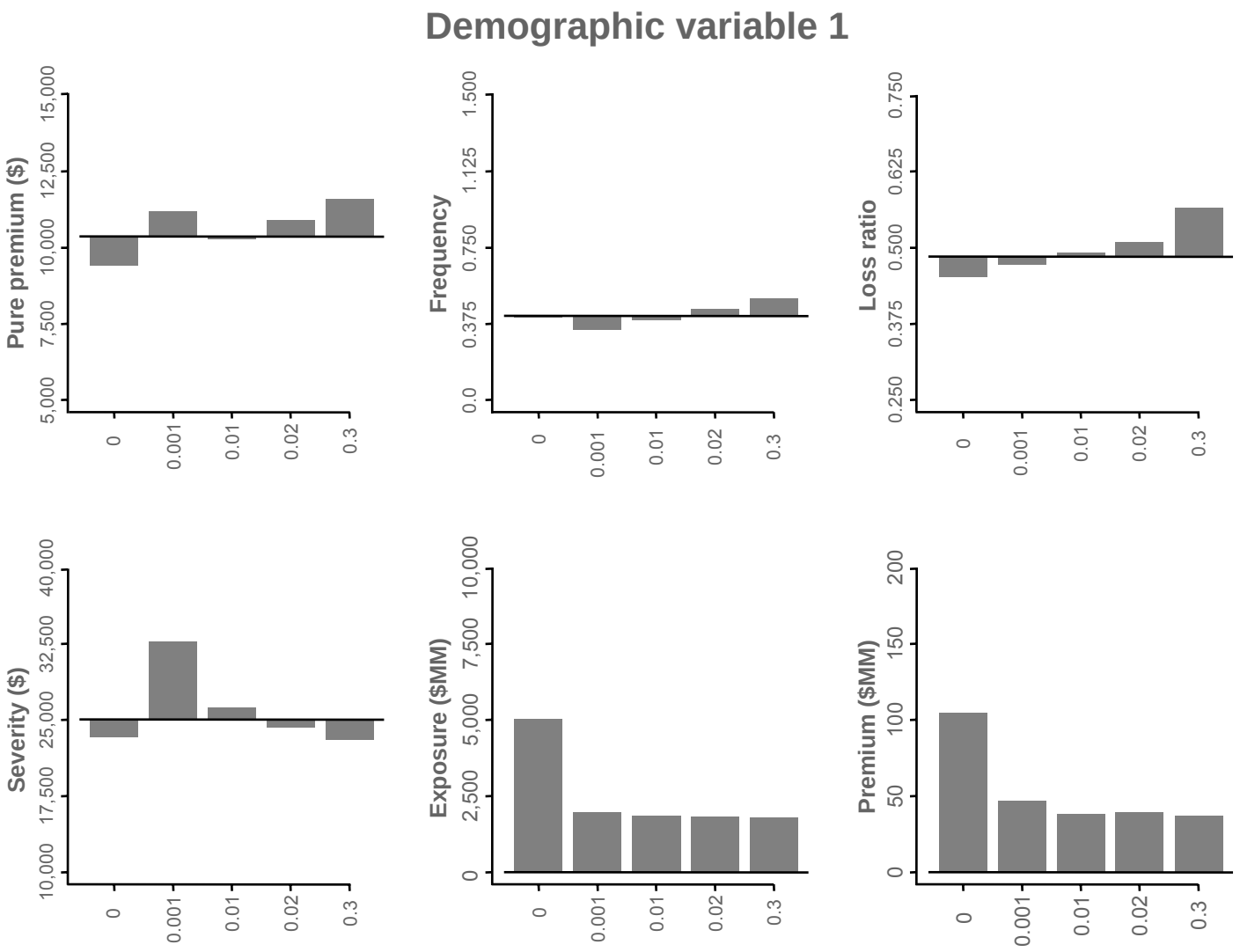


Hypothesis building

Goals

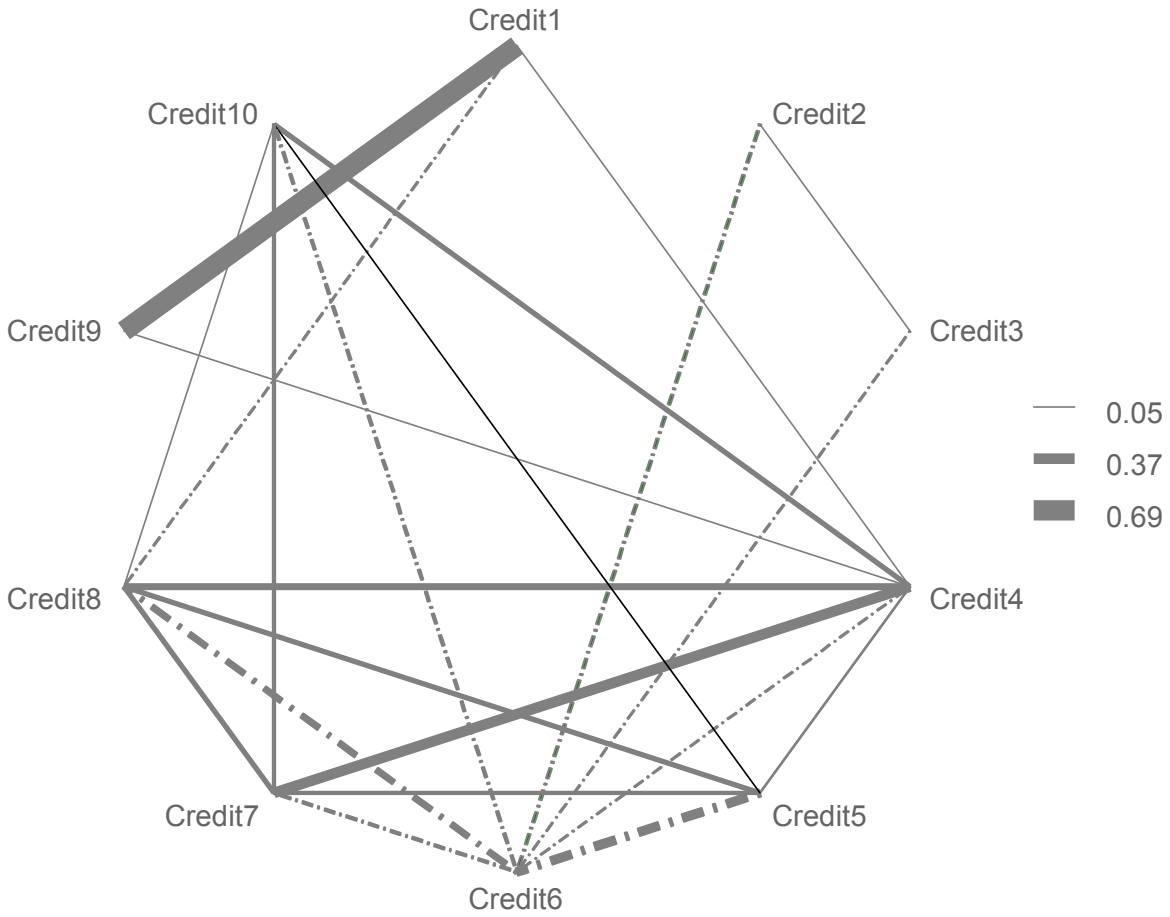
- Perform initial analysis of potential predictor variables
- Limit the list of predictor variables to be employed in subsequent phases of model building
- Further reasonability testing of data

Hypothesis building



Hypothesis building

Correlation web concisely summarizes a correlation matrix



Model building

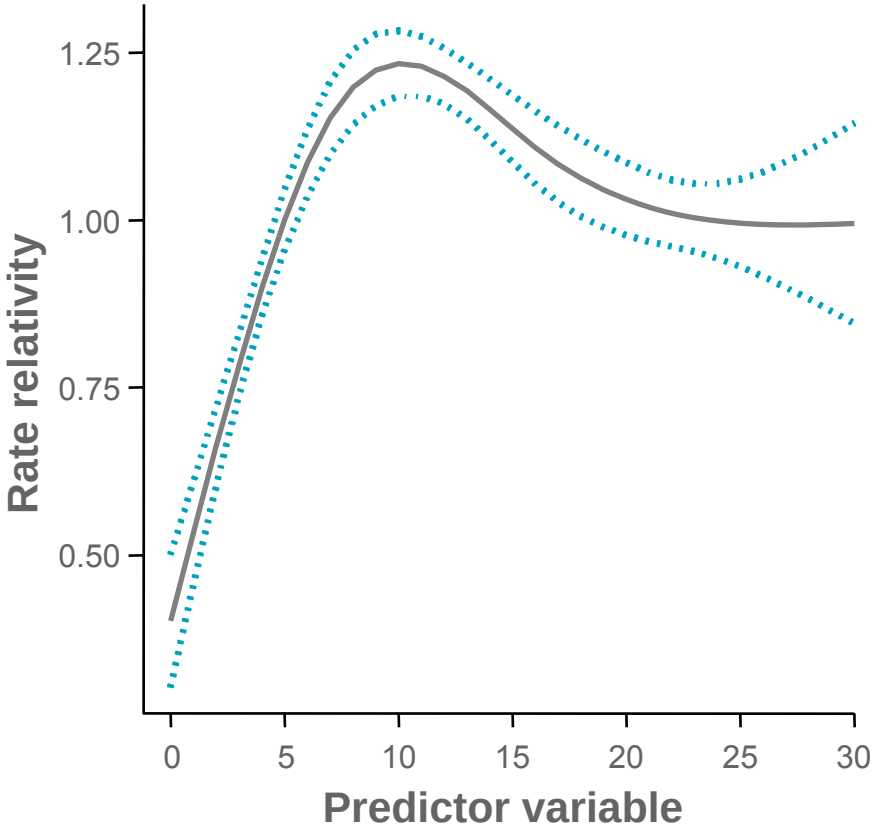
- Model building is an iterative process
- Understanding patterns and relationships throughout this process is critical

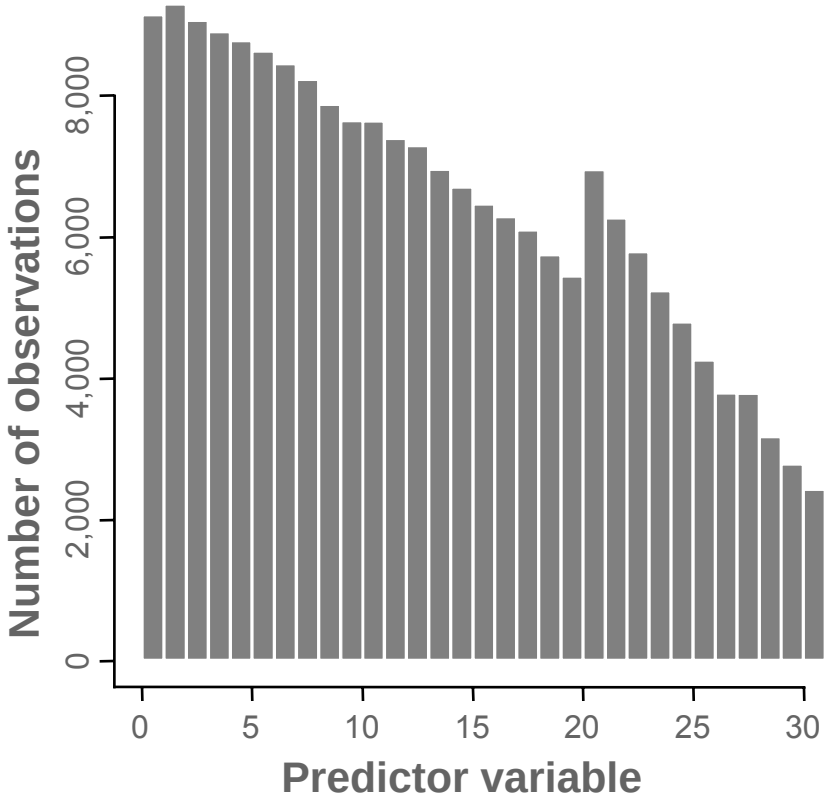
Model building

- Partial plots are a key tool to visualize predictor variables throughout the model building process
- What is a partial plot?
  - Linear predictor = $k + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
  - Predicted value = $(e^{k}) \times (e^{\beta_1 X_1}) \times (e^{\beta_2 X_2}) \times (e^{\beta_3 X_3}) \times (e^{\beta_4 X_4})$
- Partial plot demonstrates an individual predictor variable's contribution to final prediction

Model building

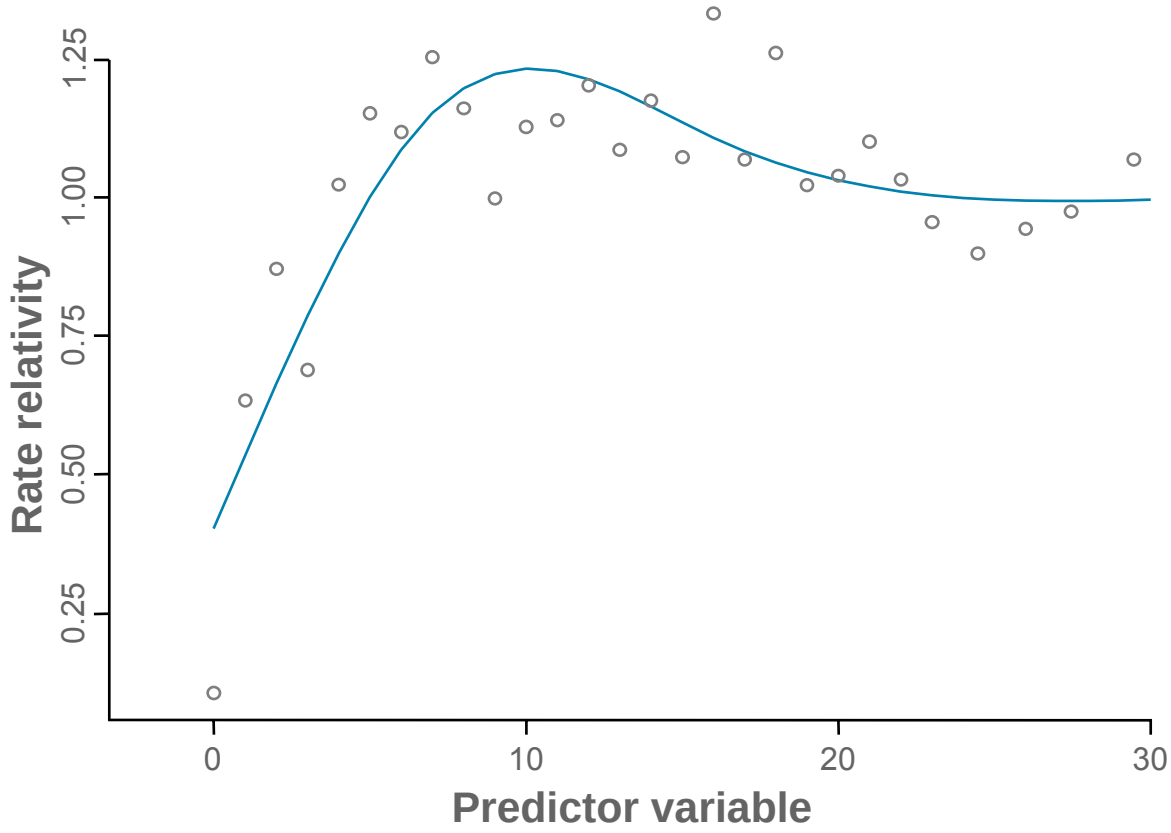
Partial plot demonstrates an individual predictor variable's contribution to final prediction





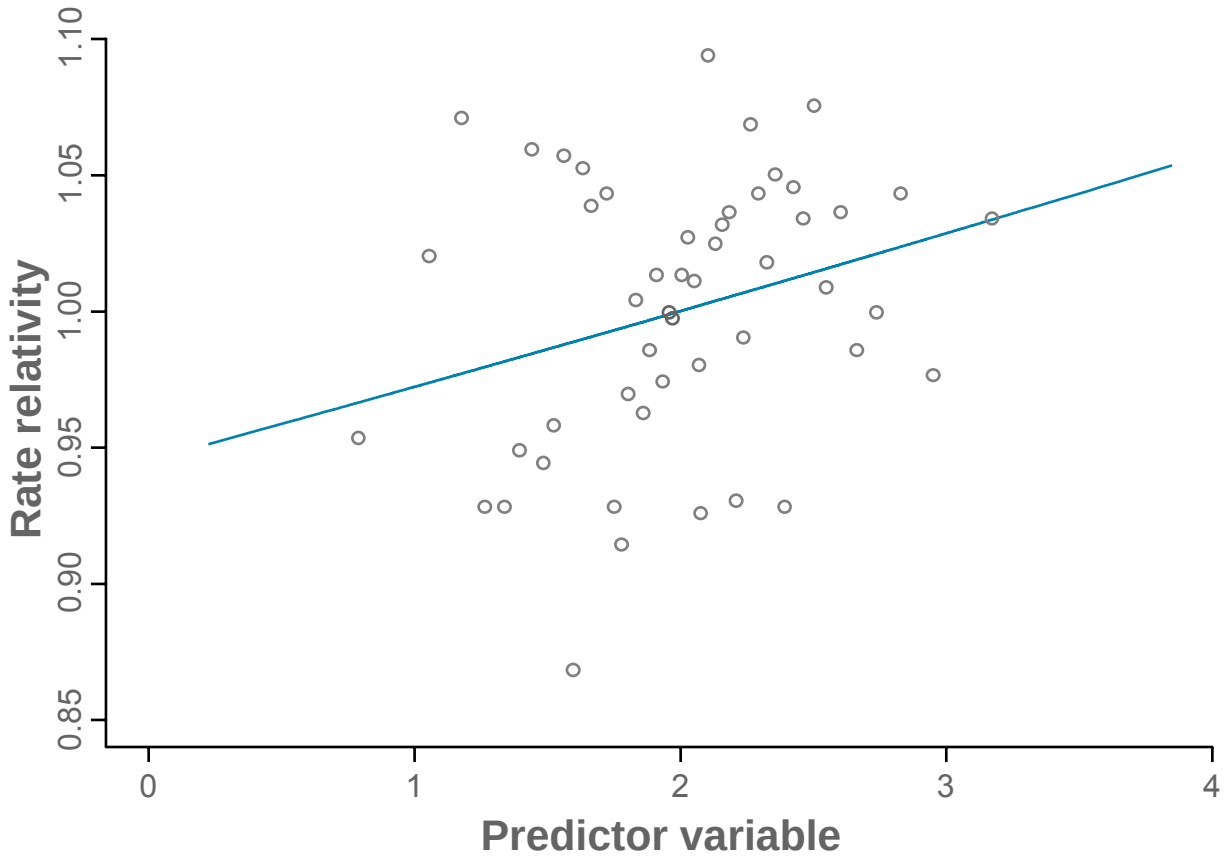
Model building

Partial plot with modified scatter plot of variable



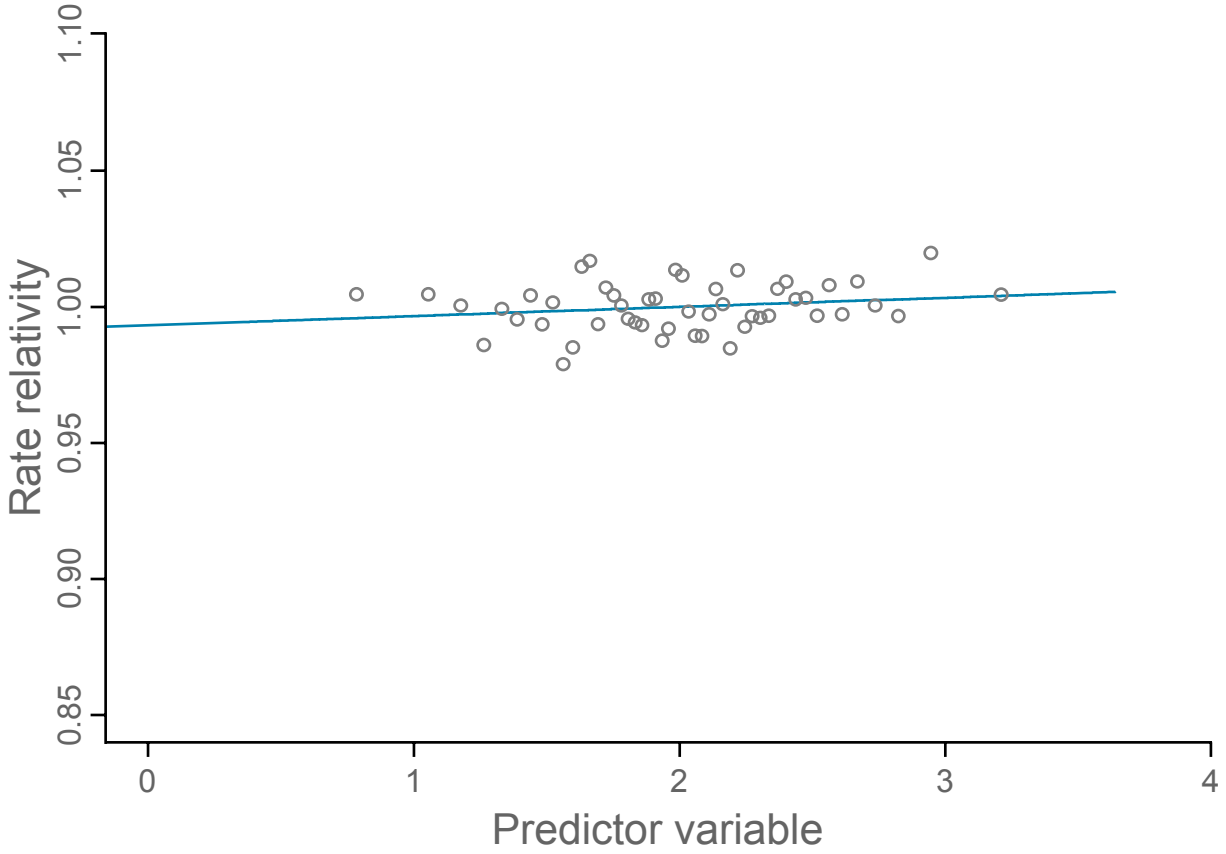
Model building

Low P-value due to variability



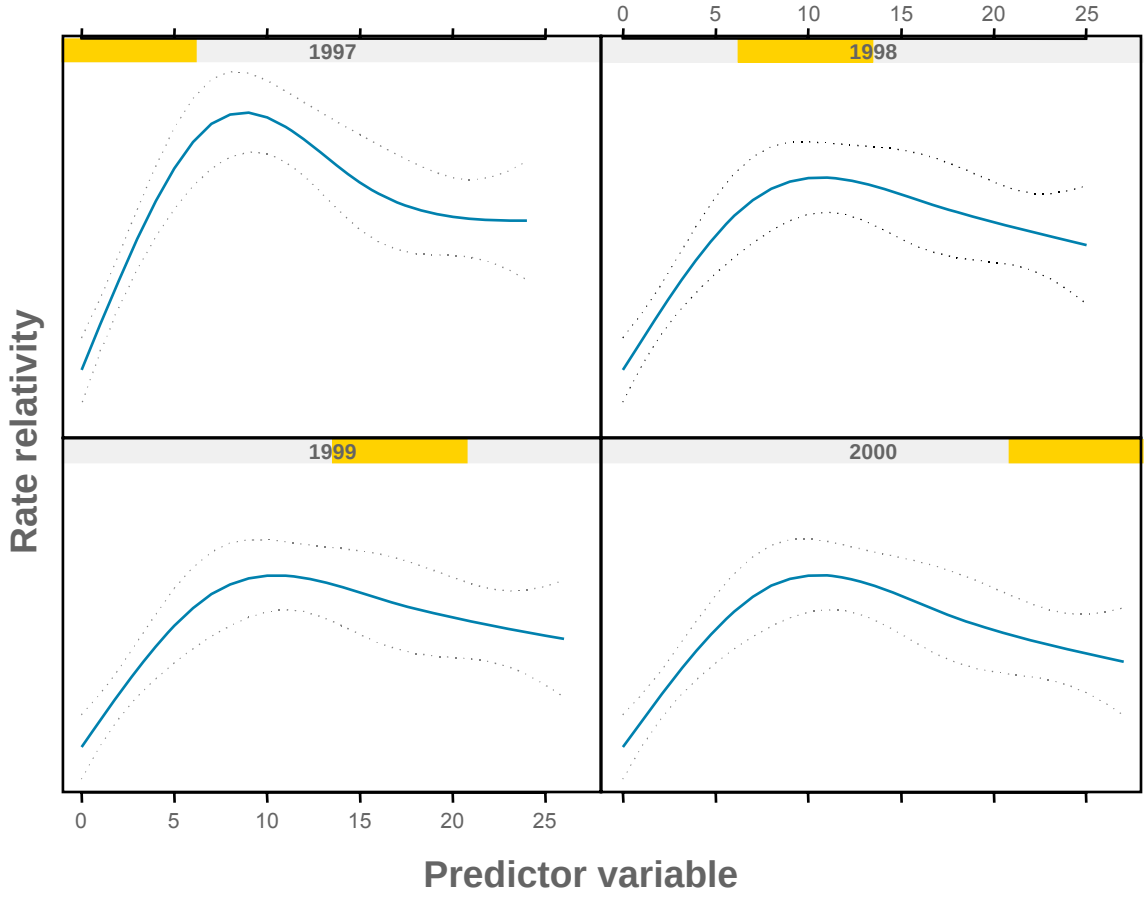
Model building

Low P-value due to little differentiation



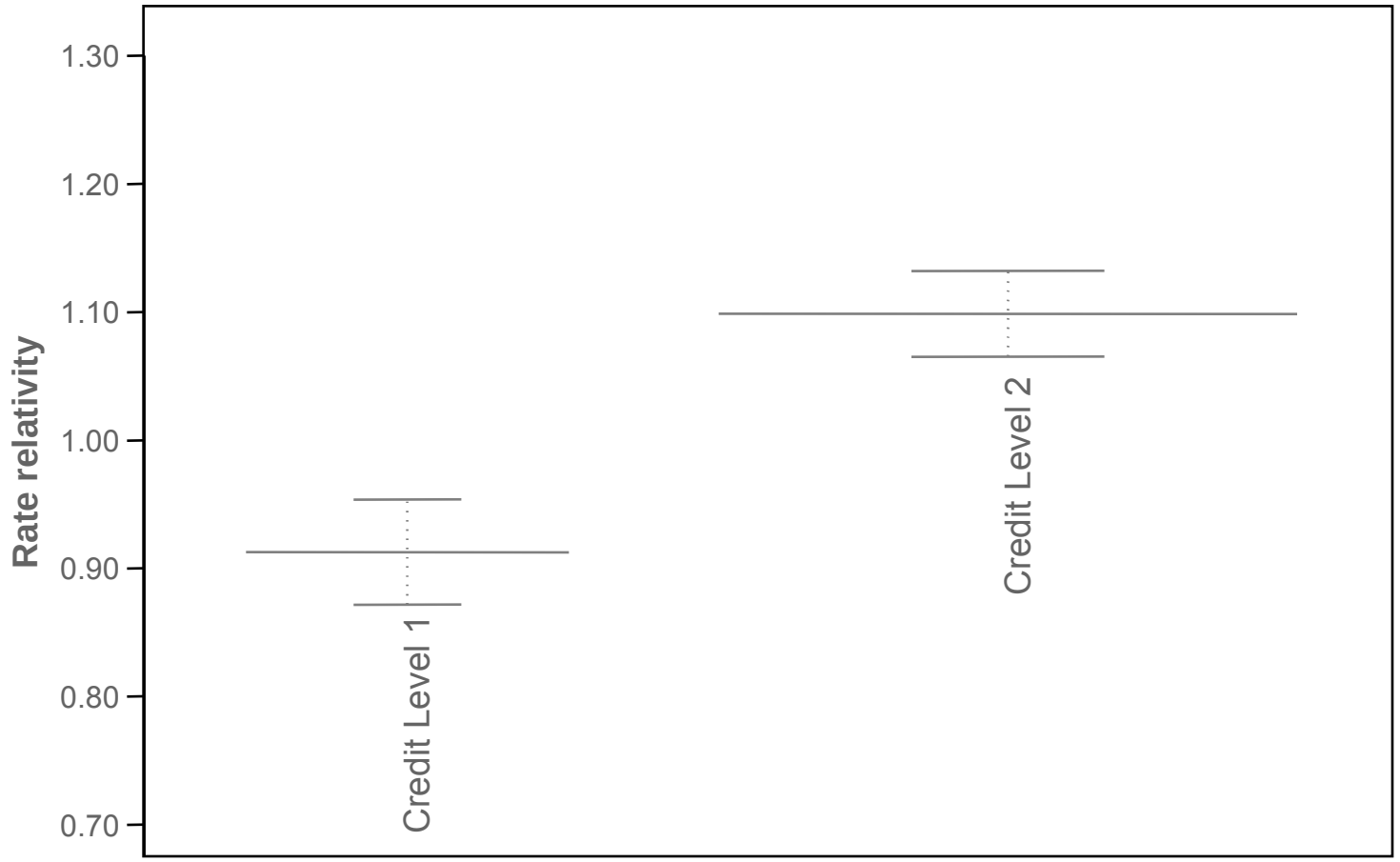
Model building

Time consistency plot is a critical tool for numeric predictors



Model building

Partial plot for a factor variable

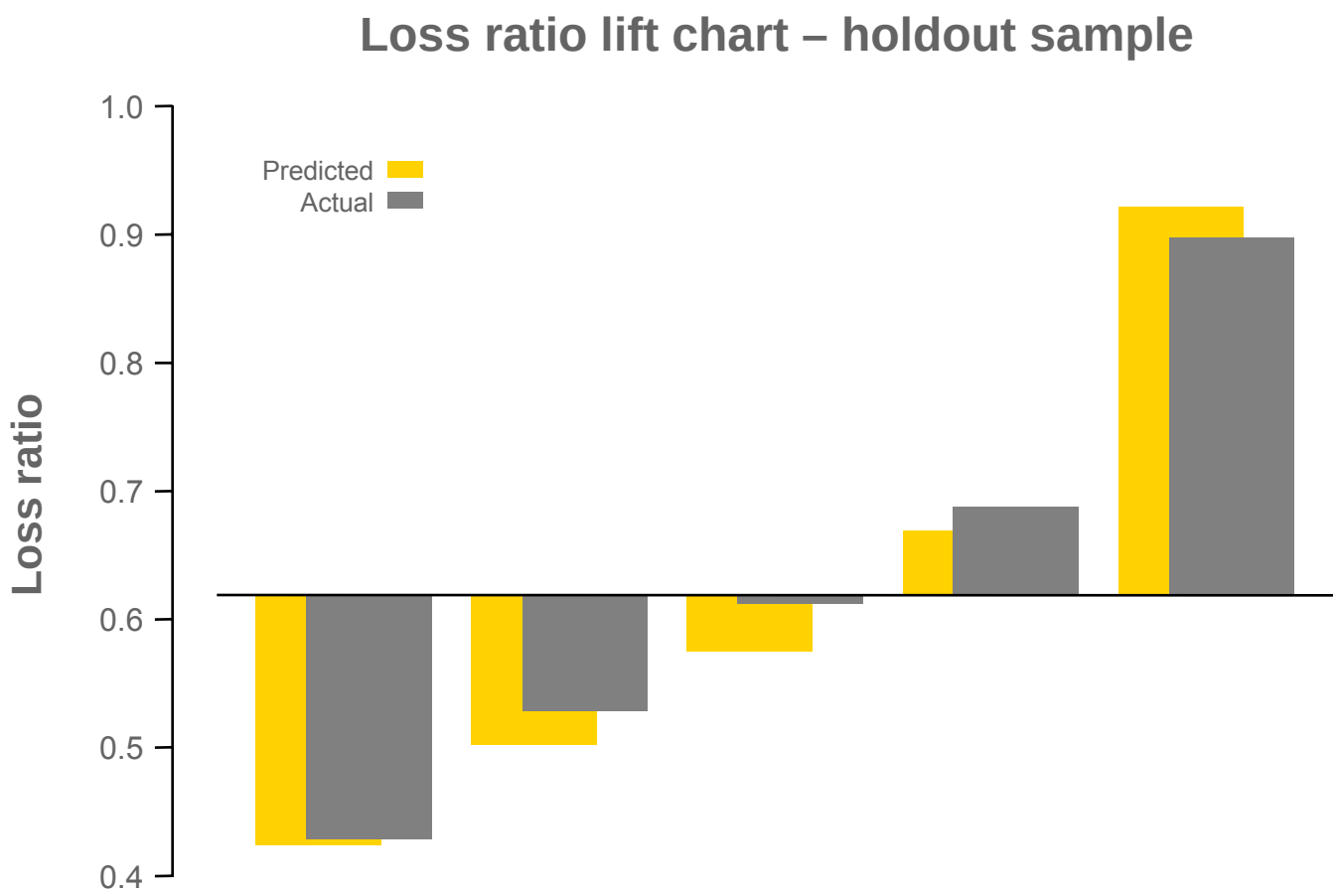


Model testing

- Likely the most critical visualizations in predictive modeling work
  - Management's perception of a project's success will likely depend on these visualizations
- Holdout tests
- Cross-validation tests

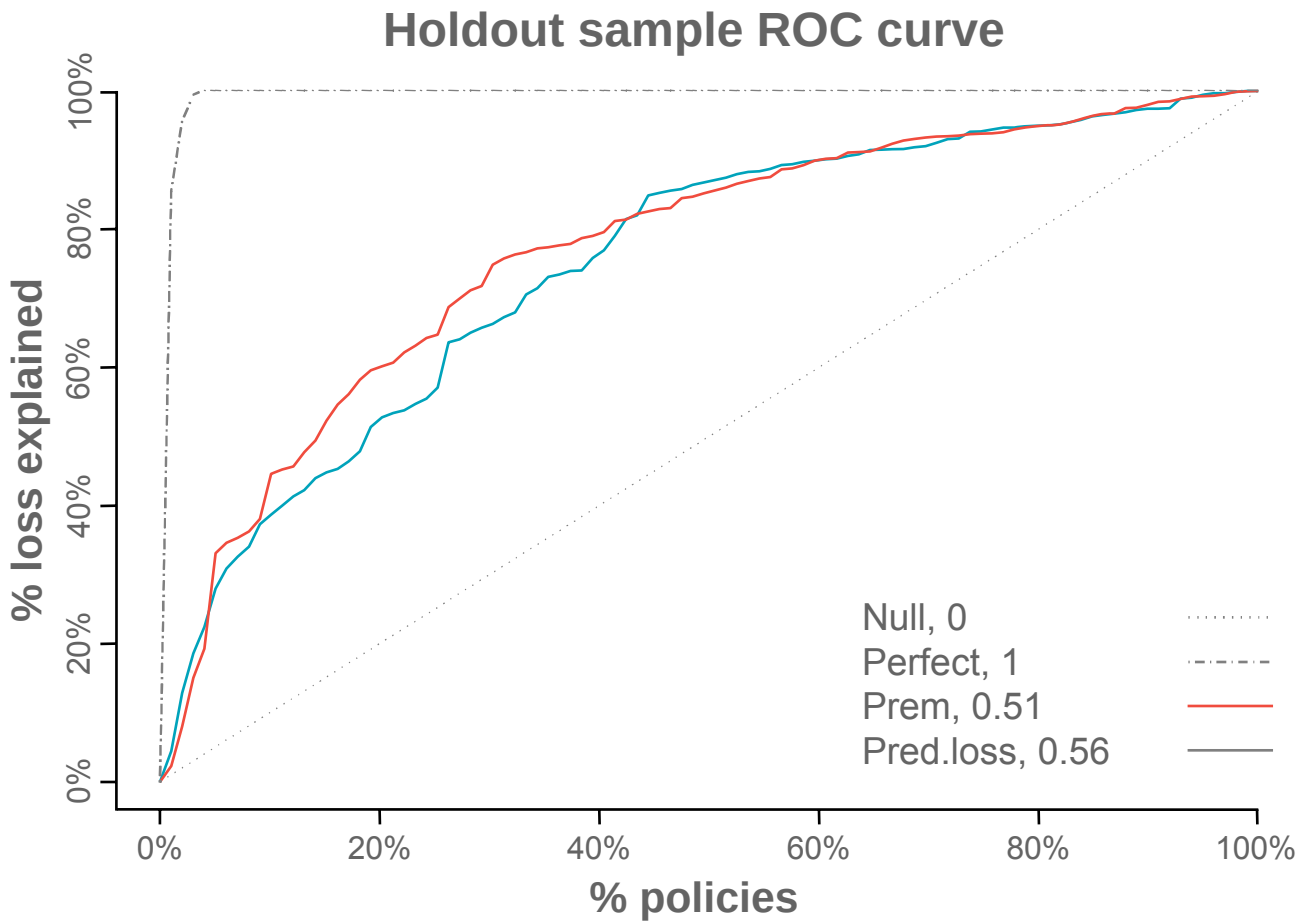
Model testing

Lift chart shows overall model performance



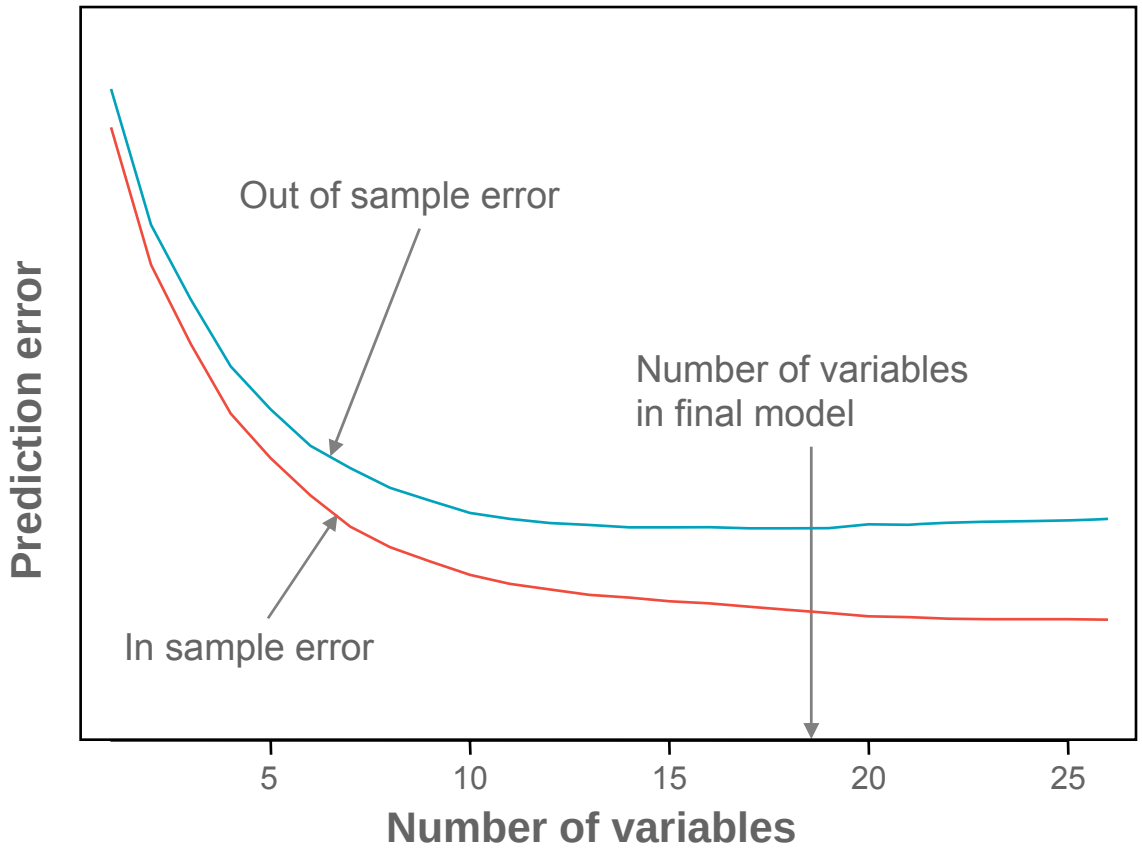
Model testing

ROC curve shows overall model performance



Model testing

Classical cross-validation exhibit



Monitoring model results

The work does not end when the lift chart looks good

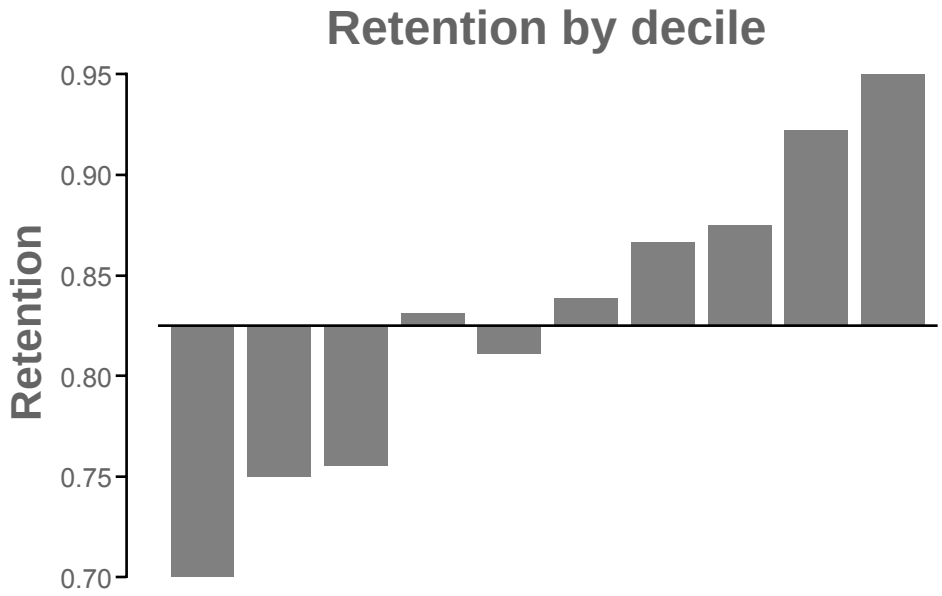
Monitoring tools

- Decile management
- Exception analysis
- Model versus actual results

Monitoring model results

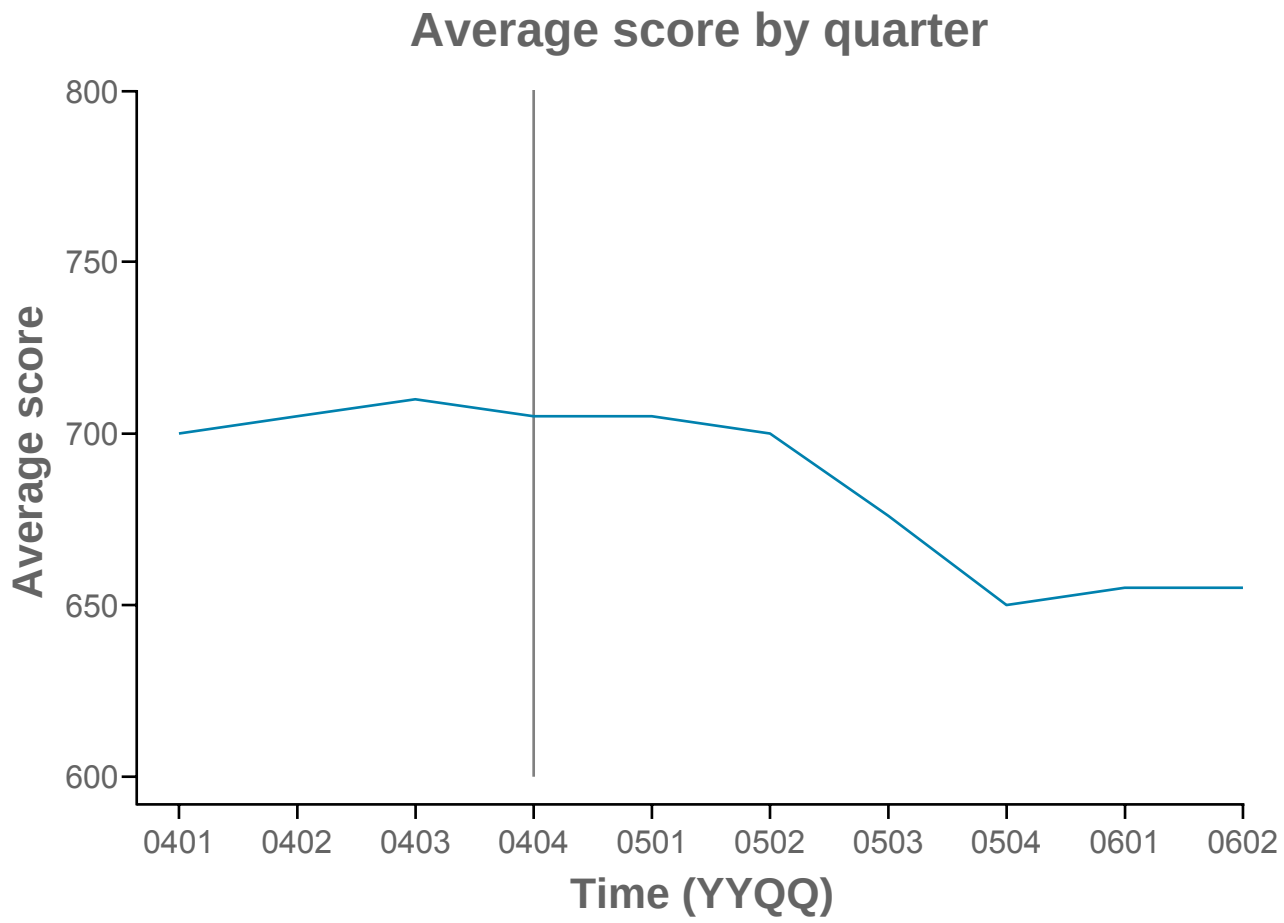
Decile management

- Retention
- Loss ratio
- Rate action
- Tier/schedule mod



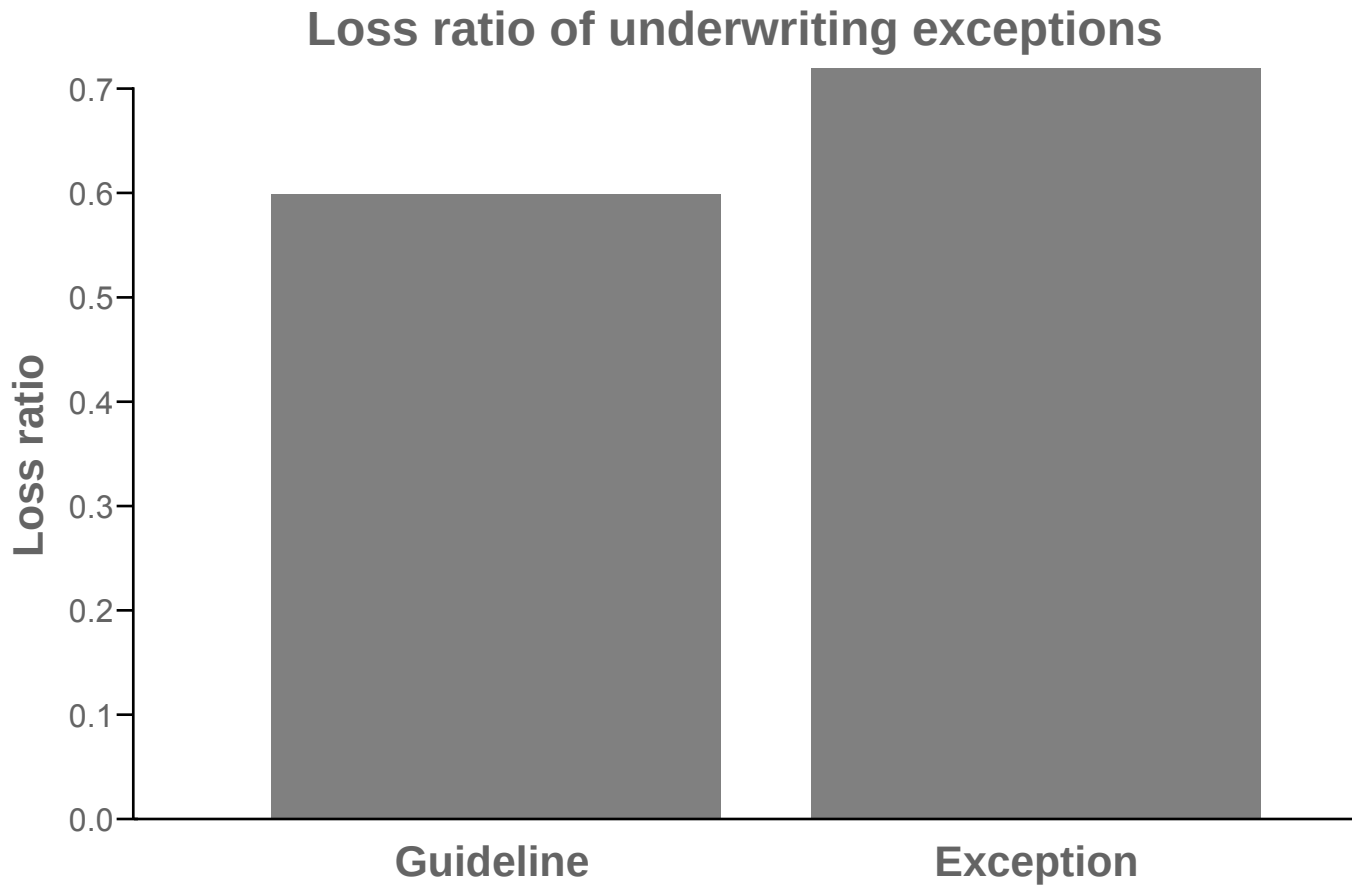
Monitoring model results

Average score over time



Monitoring model results

Loss ratio of model exceptions



Summary

- Capture the statistical concept in a graphical image
- Limit the number of concepts presented on a single slide
- Listen to your audience
- Present results throughout the life of the project