* Reminder of project purpose (Aaron 1:30)
  + British Auto data
  + Predict Severity Level (Fatal, Severe, Slight)
  + Multinomial
  + Business use: set initial insurance reserves or fraud
* Data Sources (Donald 1:30)
  + Description of files
    - Accident file – one per accident
    - Vehicle file – one record for every vehicle involved
    - Casualty file – one for every person hurt
  + One to many relationship
  + Predictors all categorical and listed as numbers
    - Created factors with descriptions from crosswalk
    - Created dummy variables for each level
* Data Cleaning and Variable exploration (Richard 3:00)
  + Missing values?
  + Near Zero values?
  + Show several variables with the Shiny app and graph
  + Describe what the graph means
    - Bars are distribution
    - Lines are average for severity
    - Want separation between lines
  + Show:
    - Urban vs Rural
    - Show Weather combined
      * Difficult to see what is going on
    - Switch to Dummy Variable for Snow
      * Explain how this seemed counter intuitive until we thought about it
    - Switch to Fog
      * Very predictive
* Feature Creation (Aaron 1:00)
  + Use accident level as base
    - Use others for feature generation at the accident level
    - Examples of new features:
      * Indicator of a motorcycle in the accident
      * Count of kids hurt
      * Etc
    - Rolled up to accident level and merged
    - Chosen by brainstorm/common sense
  + Description of Train/Test/Holdout and their use
* Generation of multiple models
  + Random Forest (Aaron 0:30)
    - 5 fold cross validation
    - Tuning parameters selected
    - Final model:
      * Accuracy, Kappa, Confusion Matrix
  + Multinom – neural net (Aaron 0:30)
    - Tuning parameters selected
    - Final model:
      * Accuracy, Kappa, Confusion Matrix
  + One vs Many classification approach (Donald 3:00)
    - Spend most of time talking about this one
    - Talk about how it was done
    - Decisions that had to be made
    - AUC optimization and order of models
    - Business rules to put them together
    - Final model:
      * Accuracy, Confusion Matrix
* Model Selection (Richard 2:00)
  + Show all three Accuracy and Confusion Matrix
    - Based on Test data
  + Which one we select and why
  + Accuracy of the selected model on Holdout
  + Accuracy and Confusion Matrix
* Summary of what we learned? (Aaron 0:30)