Model interpretability summary presentation

# Basic preparation

Split data out into Training, Holdout 1, Holdout 2

Train XGBoost model using all variables on Training

Apply model to Holdout 1 and plot PR curve and Gains chart

# Explore details of the model:

* Hyperparameters
* Plot of trees
* Variable importance and correlation
* Split points variable importance
* Interactions variable importance

# Train model using fewer variables:

* Vars with > 1% importance
* Top 50 vars, Top 10 vars
* Top 50 vars, used in top 10 interactions

Plot PR curves for all of these and move forward with the one with best performance

# Understanding what the model is doing with a single patient

* Simple patient profiles – just the list of variables and their values that weren’t missing for a patient
* Timeline of date difference variables
* LIME method

# other methods

* How model score changes as variables change
  + Relative risk curves
  + Partial dependence plots
* Characteristics of clusters of patients
  + Cluster patients by decile of risk score, run cluster descriptions
  + Bivariate stats comparing top 1% and bottom 1%
  + Dimension reduction, then plotting on a graph
  + Timeline of date difference variables for top 1% of patients
  + LIME variation using partial dependence plots