* Introduction
* Reducing crime is an important public policy objective. An important way to reduce it is appropriately allocate police resources (staff, funding).
* In order to allocate resources, police departments need to understand which counties most require support.
* An addition to previous prediction methods
  + Use of standard models
  + Use of twitter

* Data
* Describe source
* Describe variable types and why useful
* Show imbalance and why logged

1. Describe why logging dependent ensures we are less likely to get “biased” predictions
2. Describe potential trade-off (low-crime communities) and why less of a problem

* Methods; Models
* First method: Kernel Ridge Regression
* Why? Non-linear dynamics
* Steps to take
  + Write own cross-validation function for radial basis function, non-homogeneous polynomial, in parallel.
* **Look into**

1. Is the linear kernel the same as the “usual”?
2. Derive what the features are in non-homogeneous polynomial

* Second method: Boosting regression trees

Why? Powerful methods

* + Apply first to regression tree
  + Judge with out of bag error to prevent overfitting
  + Show how much iterations should be used
  + Check with cv how long the tree depth should be for each base learner
* Variable importance
  + Permutation feature importance on out of bag observations

After all this is done, try..

* + Combination of kernel ridge regression and trees in boosting?

Regression tree

* Why? Strong performance
* Consider looking into KTboost
* Methods; Evaluation
* Cross-validation for finding the parameters
* RMSE for prediction
* Permutation for the regression tree & Kernel Ridge regression

Method

1. Kernel ridge regression
   * Regression
     + Linear model
   * Ridge regression (why applicable in this case?)
     + L2 penalty
   * Kernels to fit non-linear relationships
     + Phi() for the different kernels
2. Regression trees and boosting
   * Regression tree
     + Illustration of tree and splitting criterion
   * Boosting to deal with stability issue of trees (Intuitive language here)
   * Gradient boosting (use slides)
3. Evaluation mechanisms
   * CV
   * RMSE
   * Variable importance

**Methods to add if done:**

* KTbooost
* Kernel Ridge Regression: boosted