Code organization:

* Functions
  + If a task will be performed in other applications then it should be a distinct function
* Script files
  + Load, Clean, Func, Do method
    - Eventually will find that x of the functions are being used most often; can split up funcs into two scripts at that time
    - File.exists() and load() to import data after previous load and clean has been run

Computing

* Amazon ec2
  + Full price
    - Expensive but guaranteed, uninterrupted service
  + Auction
    - Cheapest option, not guaranteed
      * If time is claimed for full price then service is ended immediately
* Phoenix nap
  + More affordable than full price ec2, same service

Missing data; features not available in earlier years

* So what. Only matters when level 1 validation window extends into a region with and without a given feature, and even then many prediction rules are robust to missing data.
* For now - for each level 1 valid window if a feature is partially missing I will exclude the feature from the pre-PCA model matrix.
  + Later I will look into implementing imputation methods

Have closing lines, not early/mid-week lines

* Train with closing lines
  + Can incorporate as a feature of the website as an interactive recommendation system given a user input closing point spread – how good of a bet is this if the closing line is x?
* Really really really really want time series of line movement. Like, really.

Feature generation

* K-week average for k greater than number of previous games available
  + Instead use the average of greatest number of available games

PCA

* Adds to, does not replace the model matrix
* Can perform level 1 validation over choices of the number of and which features to include in the MM for model fitting

Stratified machines

* From most complex (one rule per week of season) to least complex (one rule for entire season)
* Hierarchical model as a middle path
  + Factor for month/week of season
* For now, use complete pooling/least complex model

Earliest data to use

* Use it all, decide later how much to pay attention to

Should I use games in weeks that are not used for validation?