Rapid Classification of (Astronomical) Time-Series Data

Challenge to uncover & understand the dynamic universe as captured by new sky surveys where:

- data plentiful but uncertain, ratty, disparate
- science of transient events requires rapid response w/ sparse resources
- humans cannot possibly be in real-time loop
- Rumsfeldian landscape of critters → non-trivial anomaly detection
Rapid Classification of (Astronomical) Time-Series Data

Our approach:
Make probabilistic statements about everything at each update

Draw from rich **machine learning** field & combine with traditional fitting

Abstract out the role of real-time human in taking raw data to organized science classes of variables & transients
Transients Classification Project

DataMarshall
- `psycopg2`, `mysqldb`, `ipython`, `xmlrpclib`

Group-Think
- crowdsourcing platform
  labelling 2D images

PyMPChecker
- asteroid webservice
  - `mysqldb`, `pyephem`, `ipython (parallel)`, `Django`

FeatureExtraction/Classification
- `numpy`, `scipy`, `Rpy`, `ipython`, `unittest`, `jpytype`

Telescope Response
- `socket`, `libxml`, `pyserial`, `pyparallel`

VizTools
- `sqlite3`, `psycopg2`, `imap`
Transients Classification Project

NSF Cyber-enabled Discovery & Innovation (CDI) Project (2009-2011)

Goal: build this framework (for real surveys) & algorithmic/computational advances in parallel machine learning

e.g. Hadoopification of ML frameworks, trying ipython on EC2