

JPL/Caltech Virtual Summer School – Big Data Analytics

The exercises related to the R module are described here. Several later modules will use more advanced R material and some type of plots we have seen will also be appearing in the visualization module.

Exercises:

(1) Apply to the astronomy dataset various graphing techniques learnt

<http://goo.gl/Zjcas6>

(2) Determine the most discriminating columns for the six different classes viz. AGN, blazars, CV, Flares, RR Lyrae, SN. The columns are: amplitude, beyond1std, fpr_mid20, fpr_mid35, fpr_mid50, fpr_mid65, fpr_mid80, linear_trend, max_slope, med_abs_dev, med_buf_range_per, pair_slope_trend, percent_amplitude, pdfp, skew, kurtosis, std, ls, rcorbor, magratio

(3) Use plot(), qplot() and ggplot2() methods to demonstrate the above