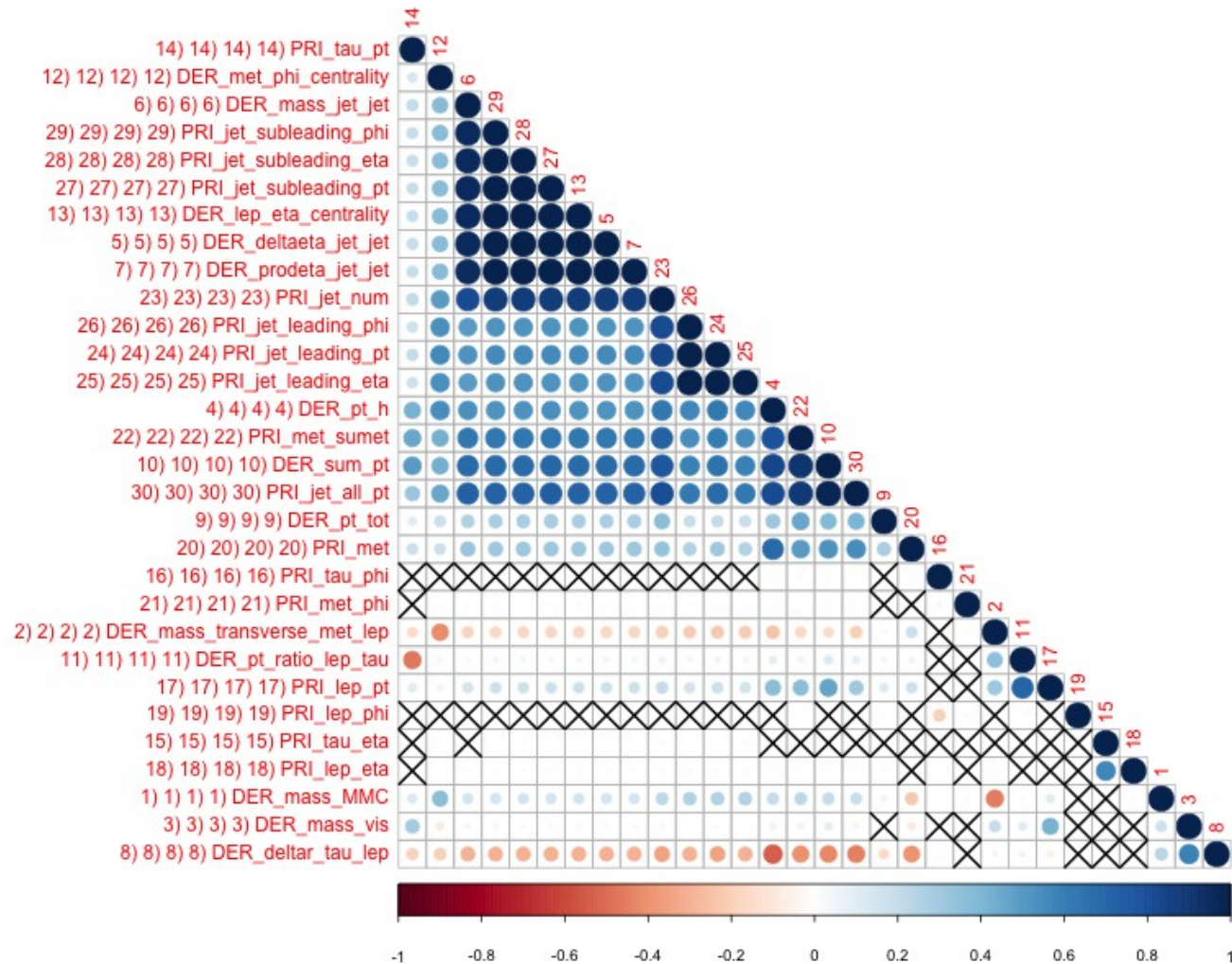


EDA

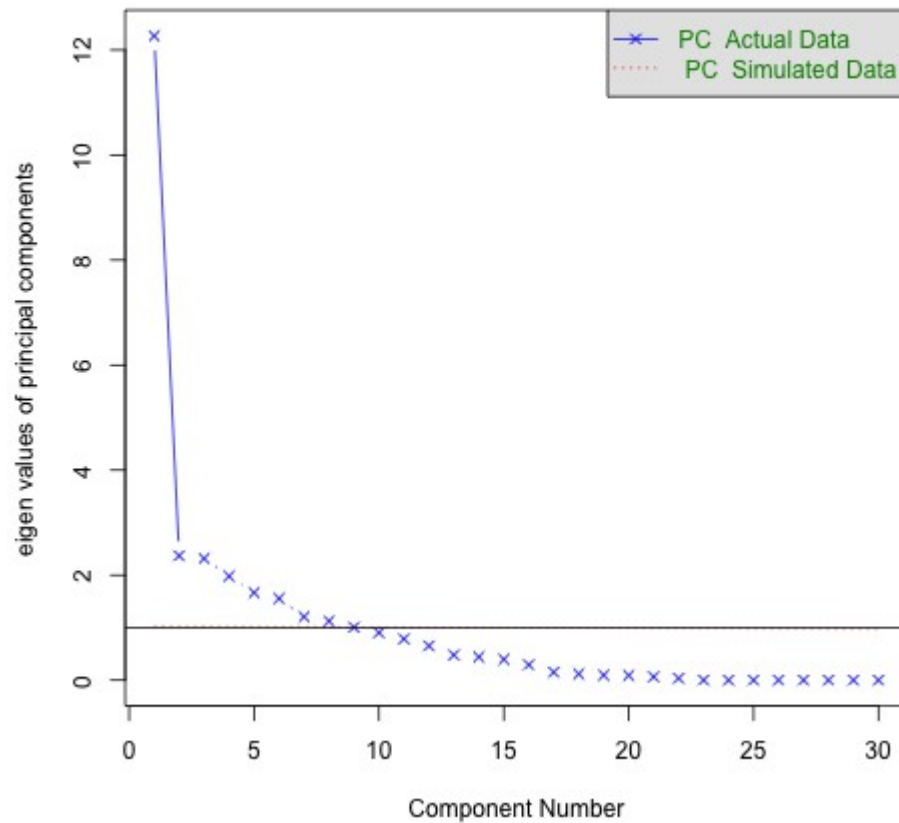


Statistically Significant Correlation Among Features

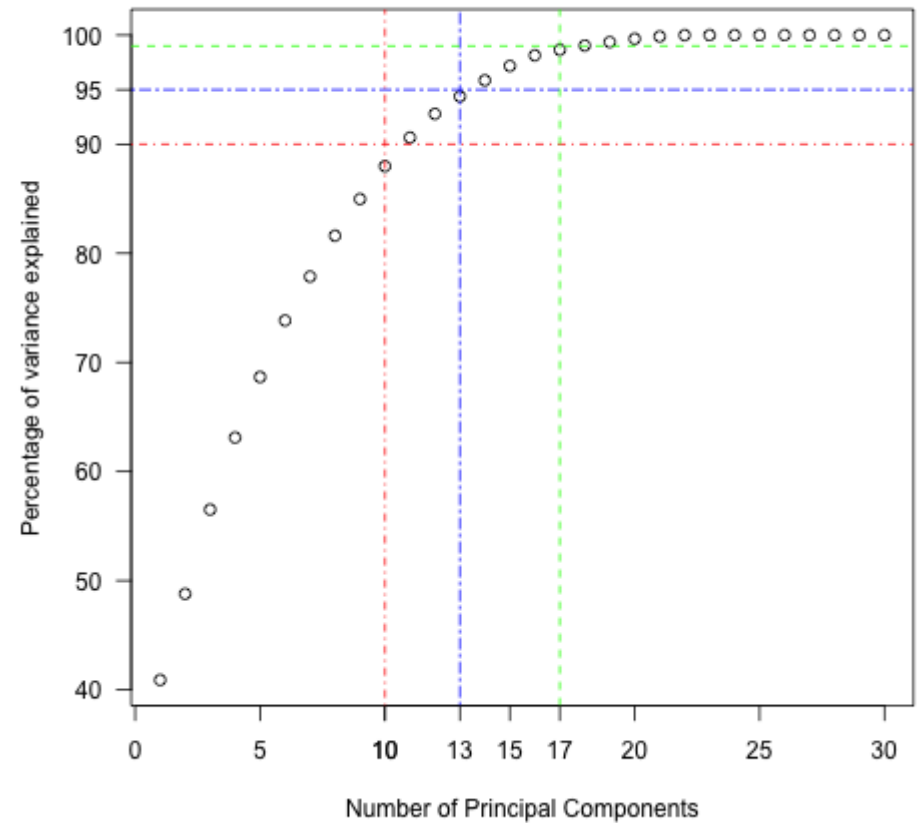


10 PCA Components Account for 90 % of Variance

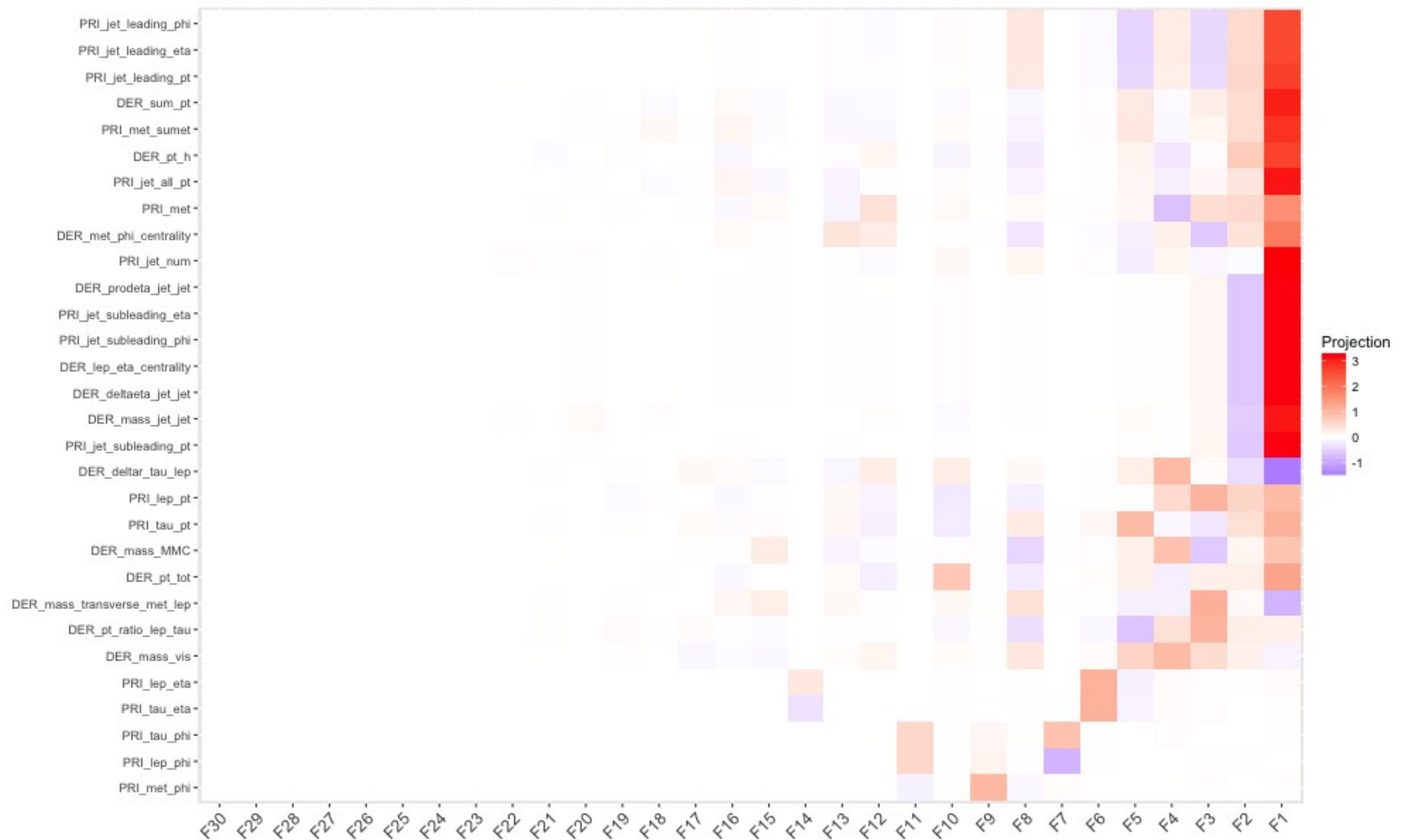
Parallel Analysis Scree Plots



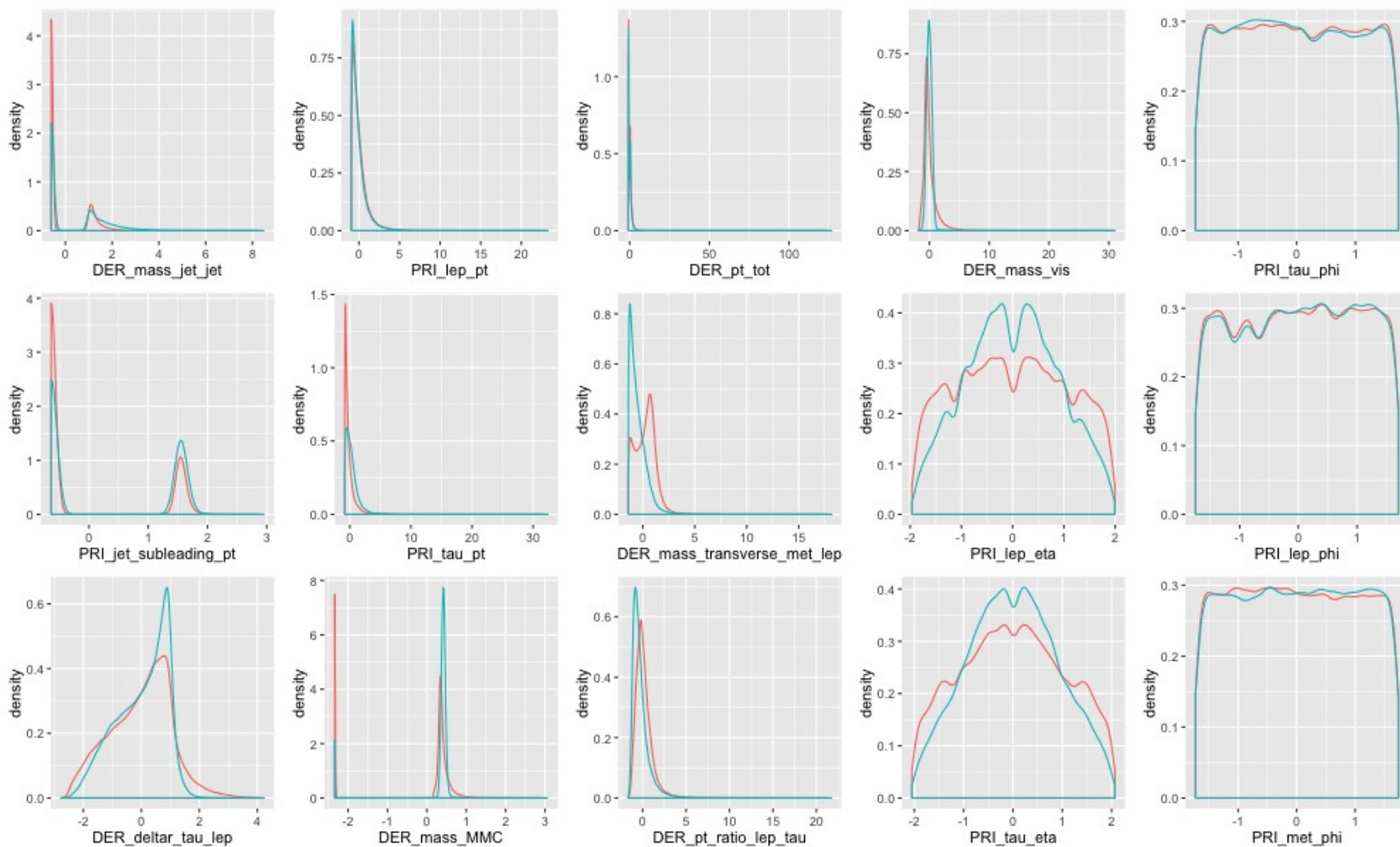
PCA



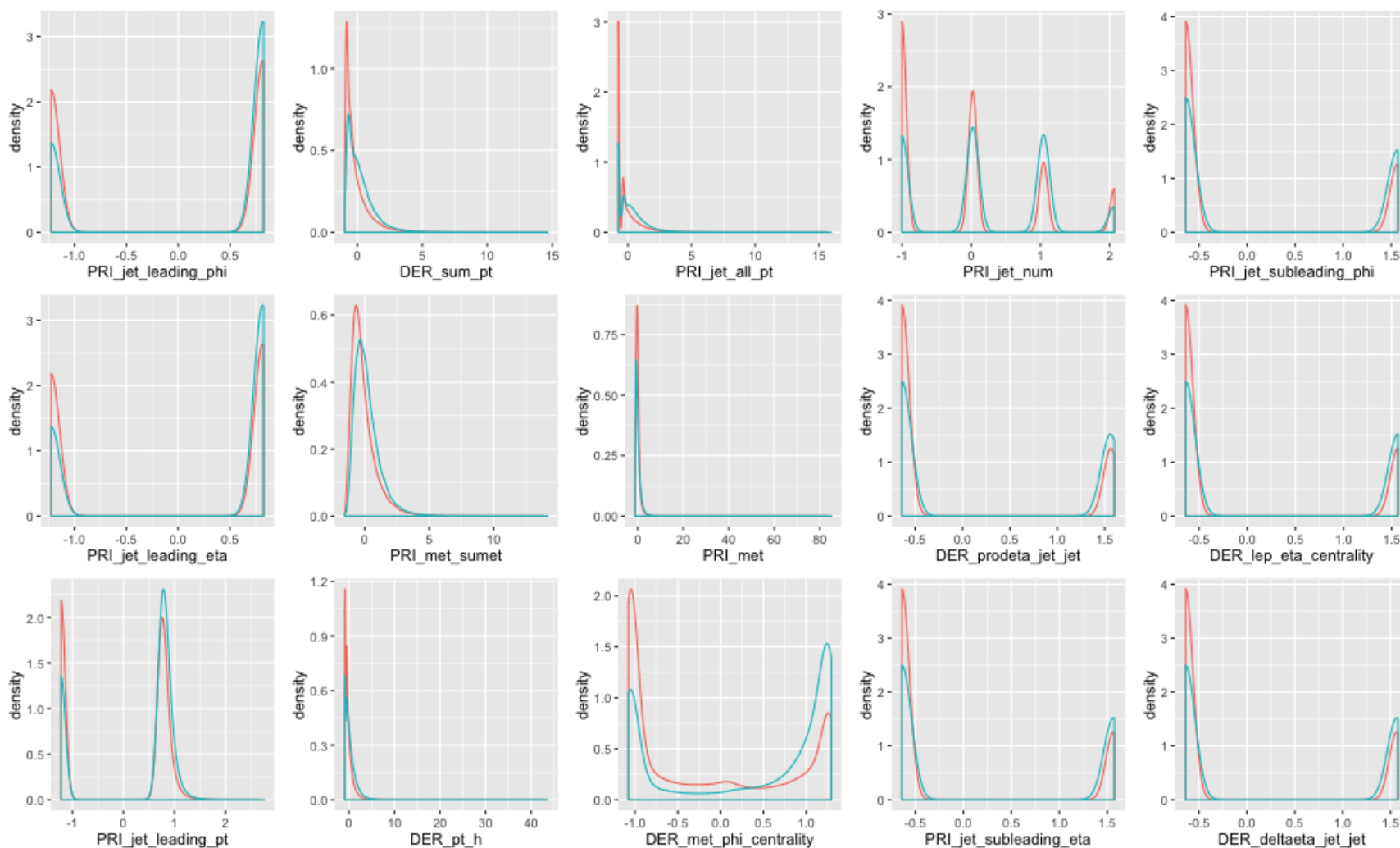
PCA Feature Vectors used to reduce number of features



Least Influential Features cannot distinguish “s” from “b”



Higher weighted Variance can distinguish “s” and “b”



Undefined Variables by Jet No.

- All Jets (0,1,2,3)

1 15.2456 1 DER_mass_MMC

- No Jets(0)

2 39.9652 24 PRI_jet_leading_pt

3 39.9652 25 PRI_jet_leading_eta

4 39.9652 26 PRI_jet_leading_phi

- No Jet and 1 Jet (0,1)

5 70.9828 5 DER_deltaeta_jet_jet

6 70.9828 6 DER_mass_jet_jet

7 70.9828 7 DER_prodeteta_jet_jet

8 70.9828 13 DER_lep_eta_centrality

9 70.9828 27 PRI_jet_subleading_pt

10 70.9828 28 PRI_jet_subleading_eta

11 70.9828 29 PRI_jet_subleading_phi

EDA Insights

- Last 5 variables in PCA feature plots have least influence in their current form
- Remove all derived features (13) + above features (5) and use SVM and Neural Networks
- Split training and test data into six categories according to level of undefined features