

36-315: Statistical Graphics and Visualization

Handout 25

Date: April 23, 2003

Data mining by projections and parallel-coordinate plots

Prediction is asymmetric—These are different questions:

- "Do women make more money?"
- "Are people with money likely to be women?"

The first is *descriptive*, asking what demographics women have in PA. The second is *discriminative* or *diagnostic*, asking what variables distinguish women from men.

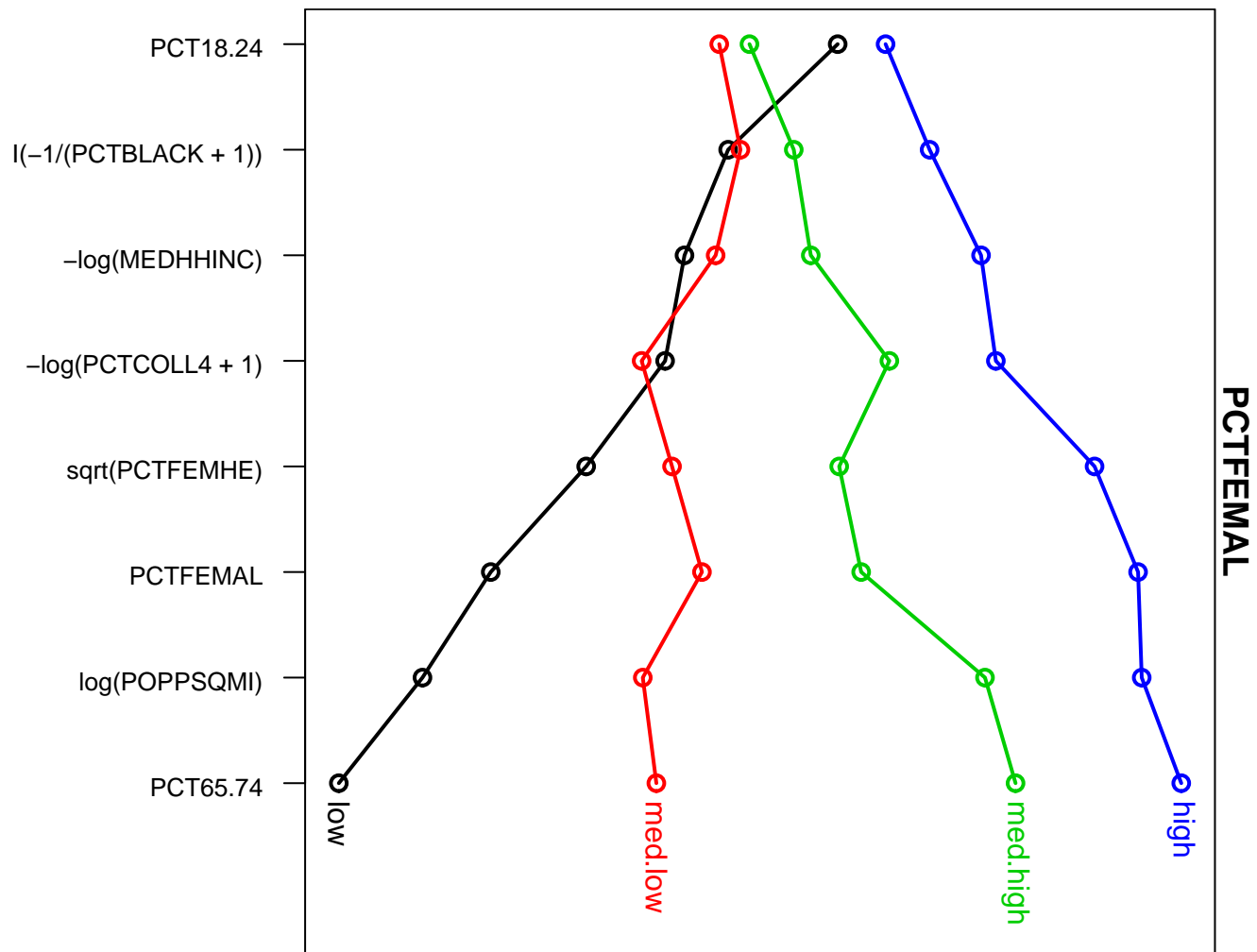
Descriptive approach—Divide data by response, show mean in each group.

Descriptive approach can sometimes work better than prediction curves—e.g. PCT18.24, PCTGQ.

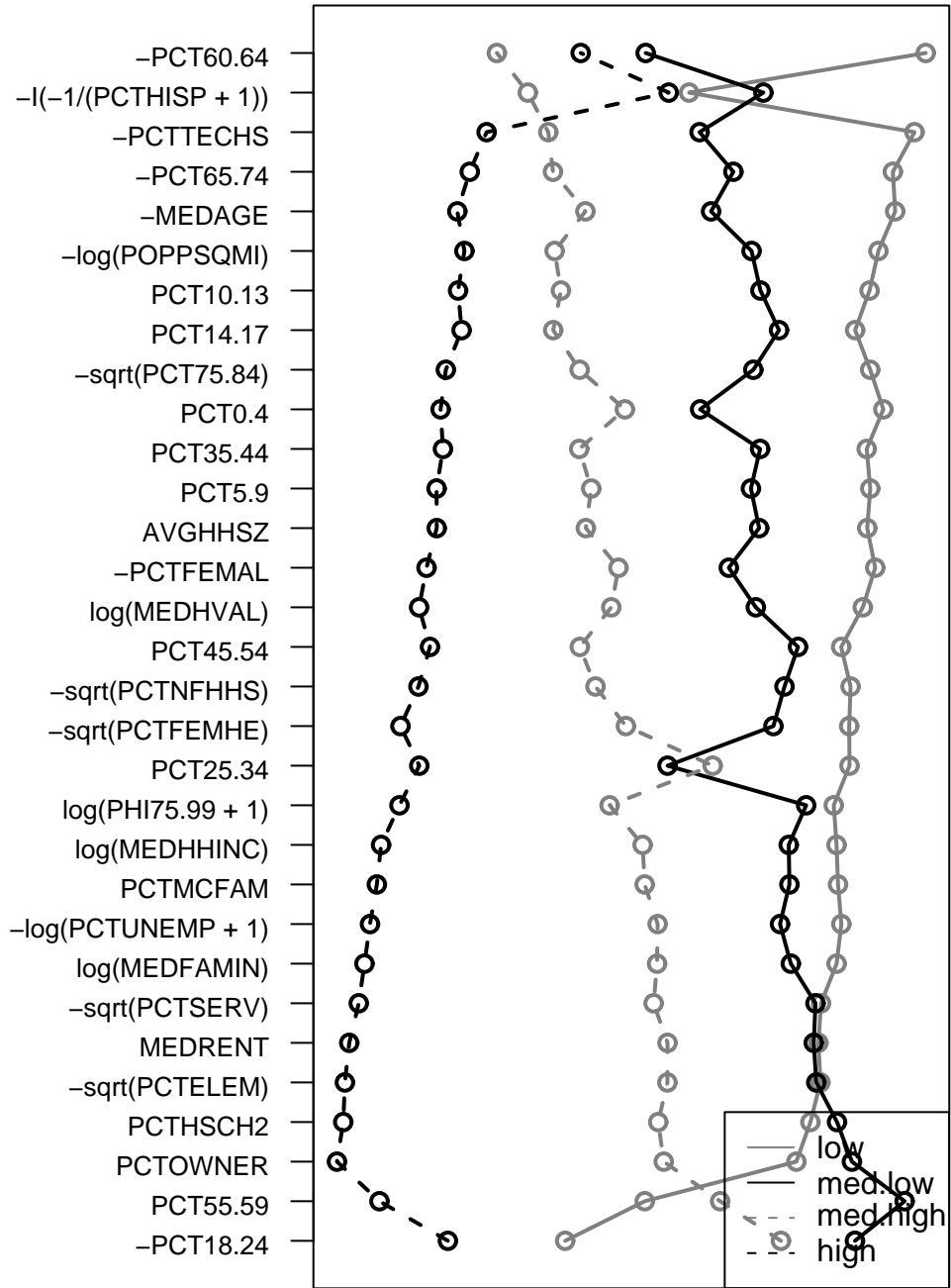
Descriptive approach can suggest a predictor ranking, other than using contour plots.

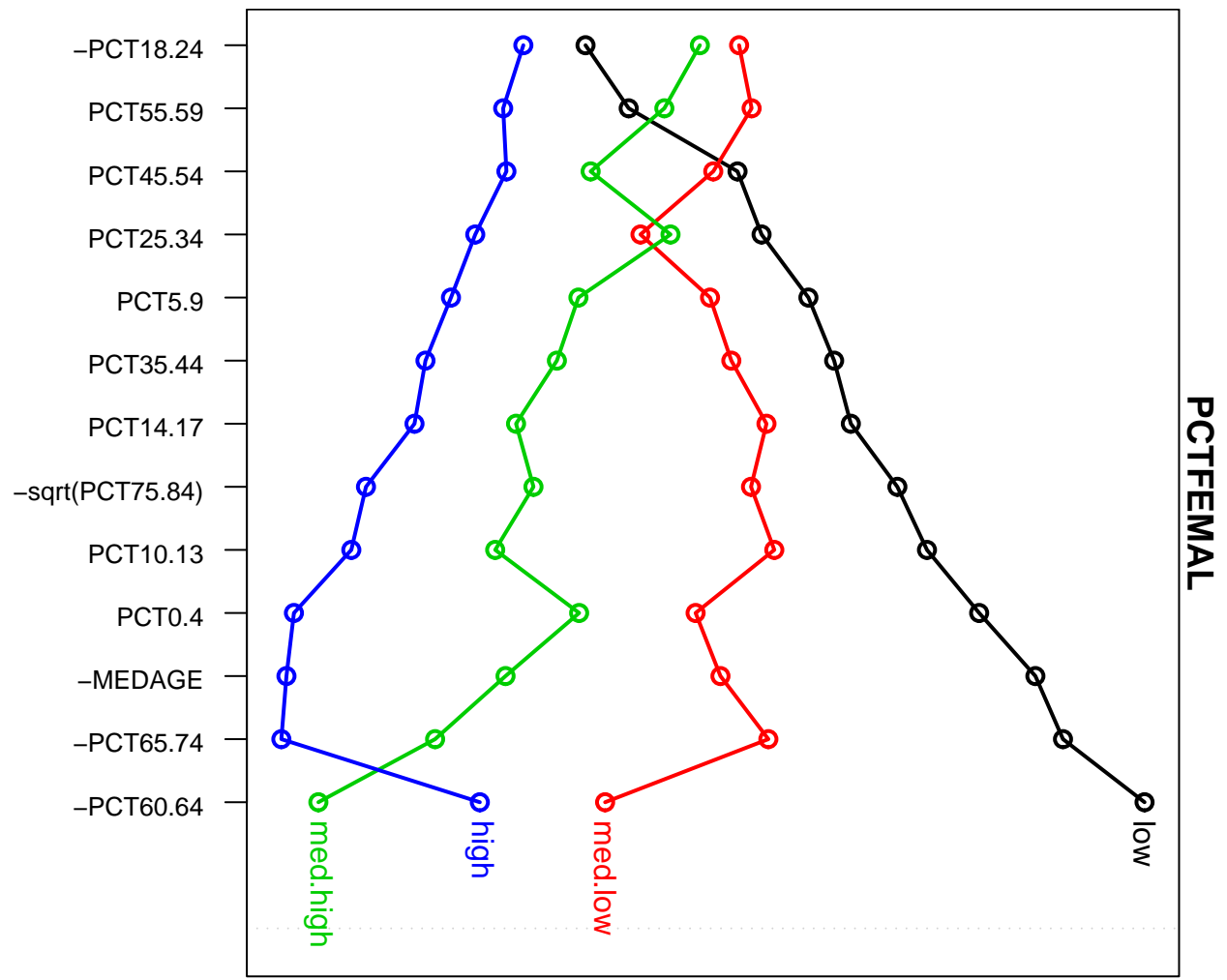
Projection can show variations with each response group, allowing further subdivision.

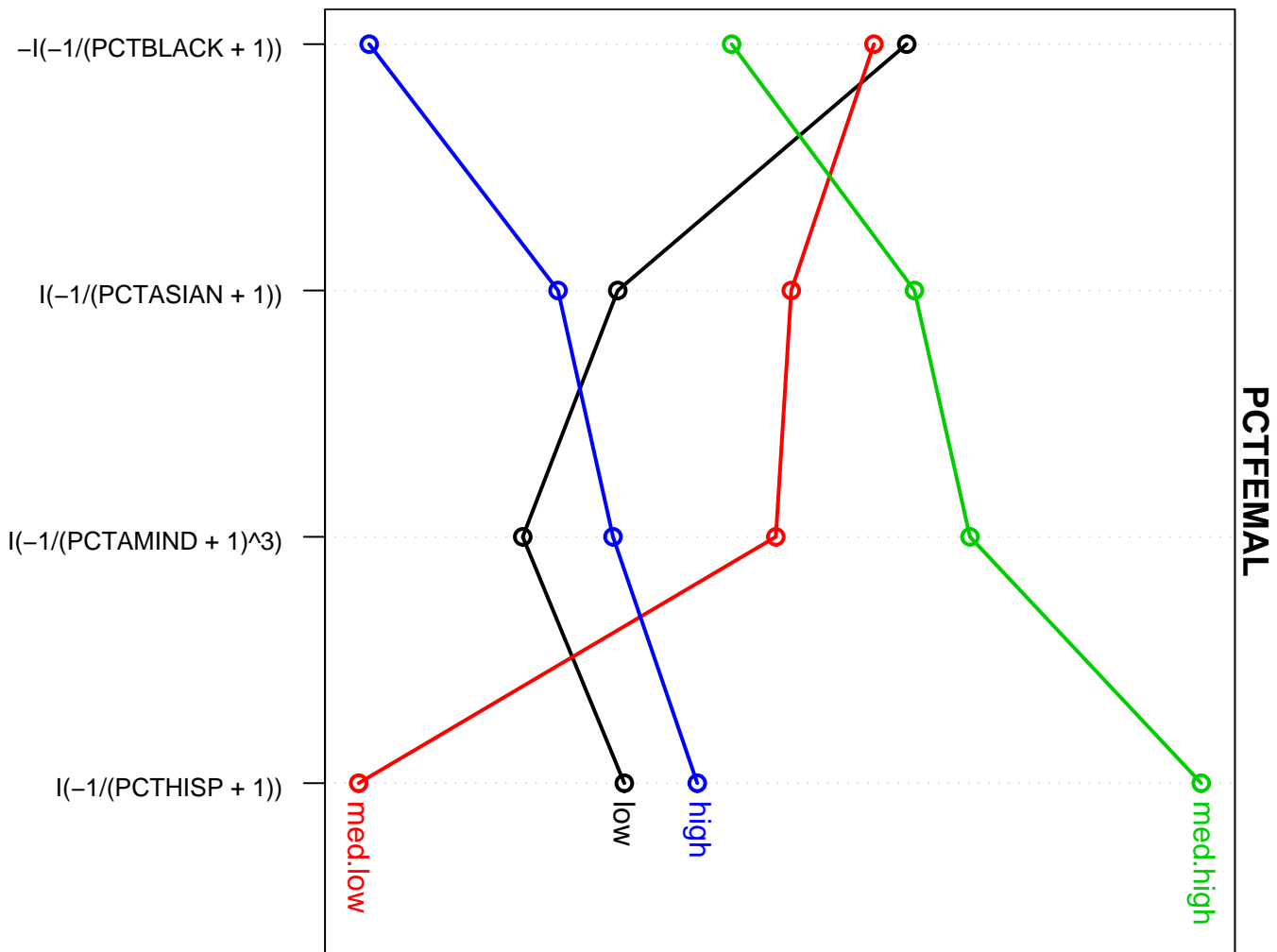
The diagnostic approach is much simplified if you rank predictors.

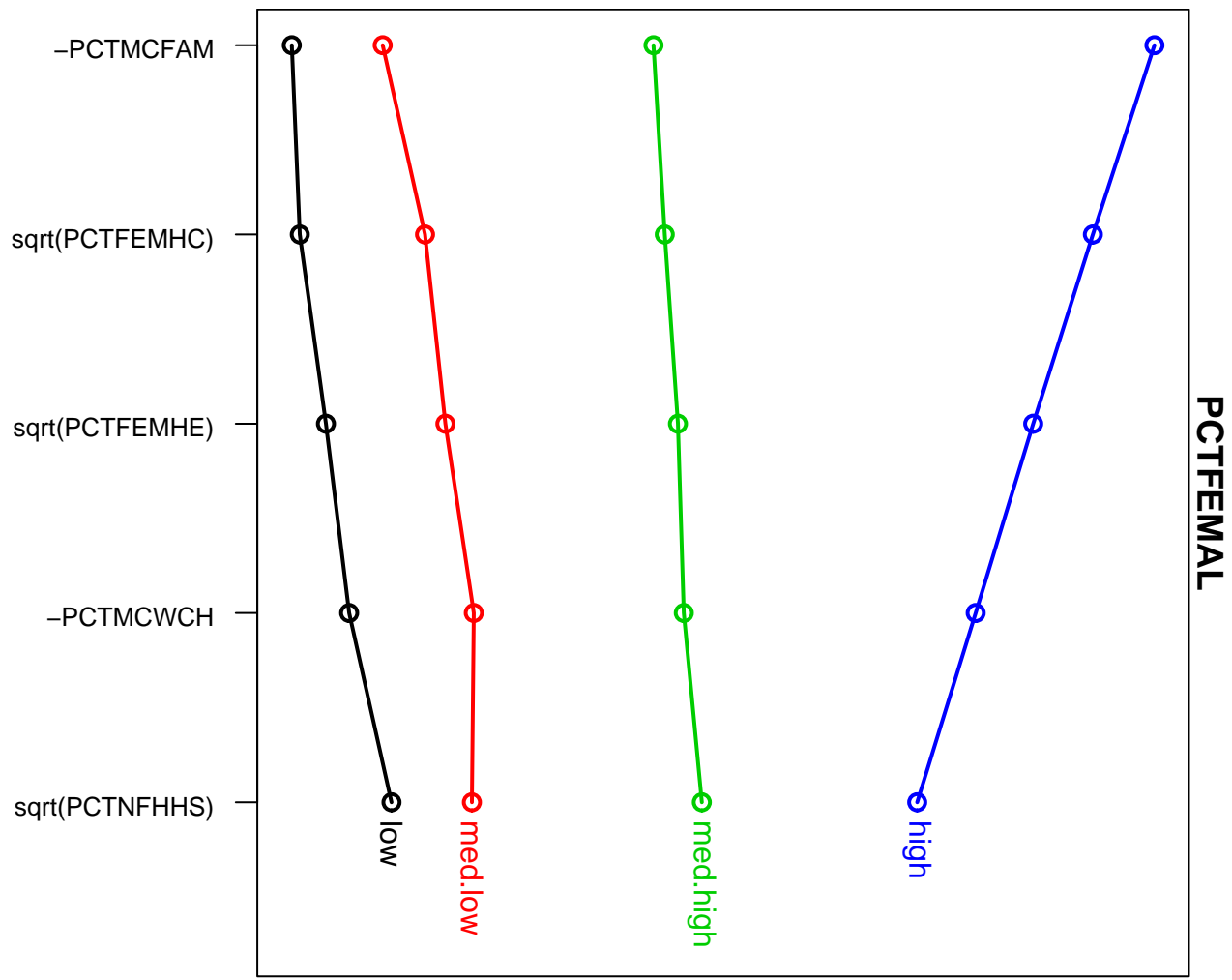


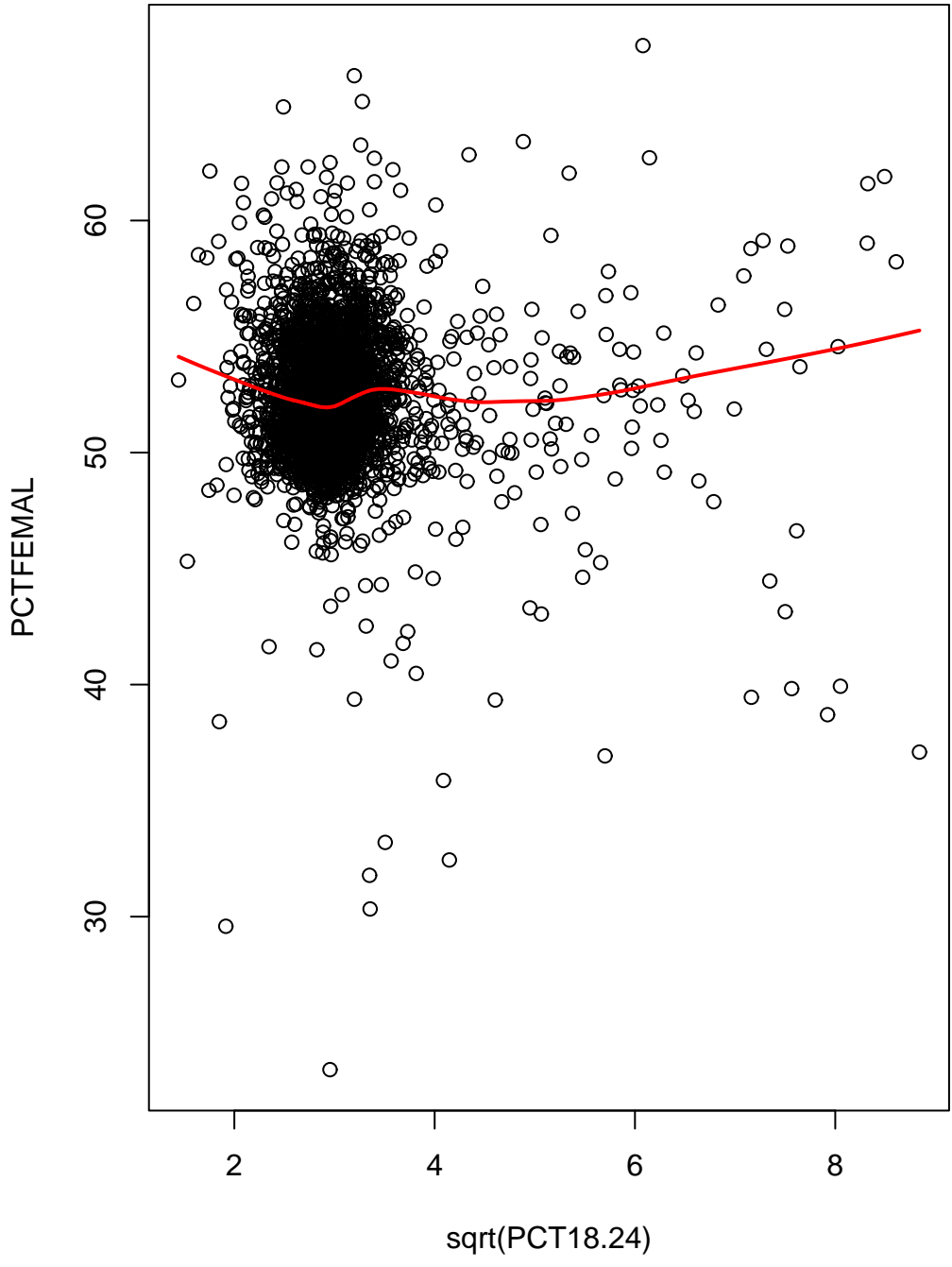
PCTFEMAL

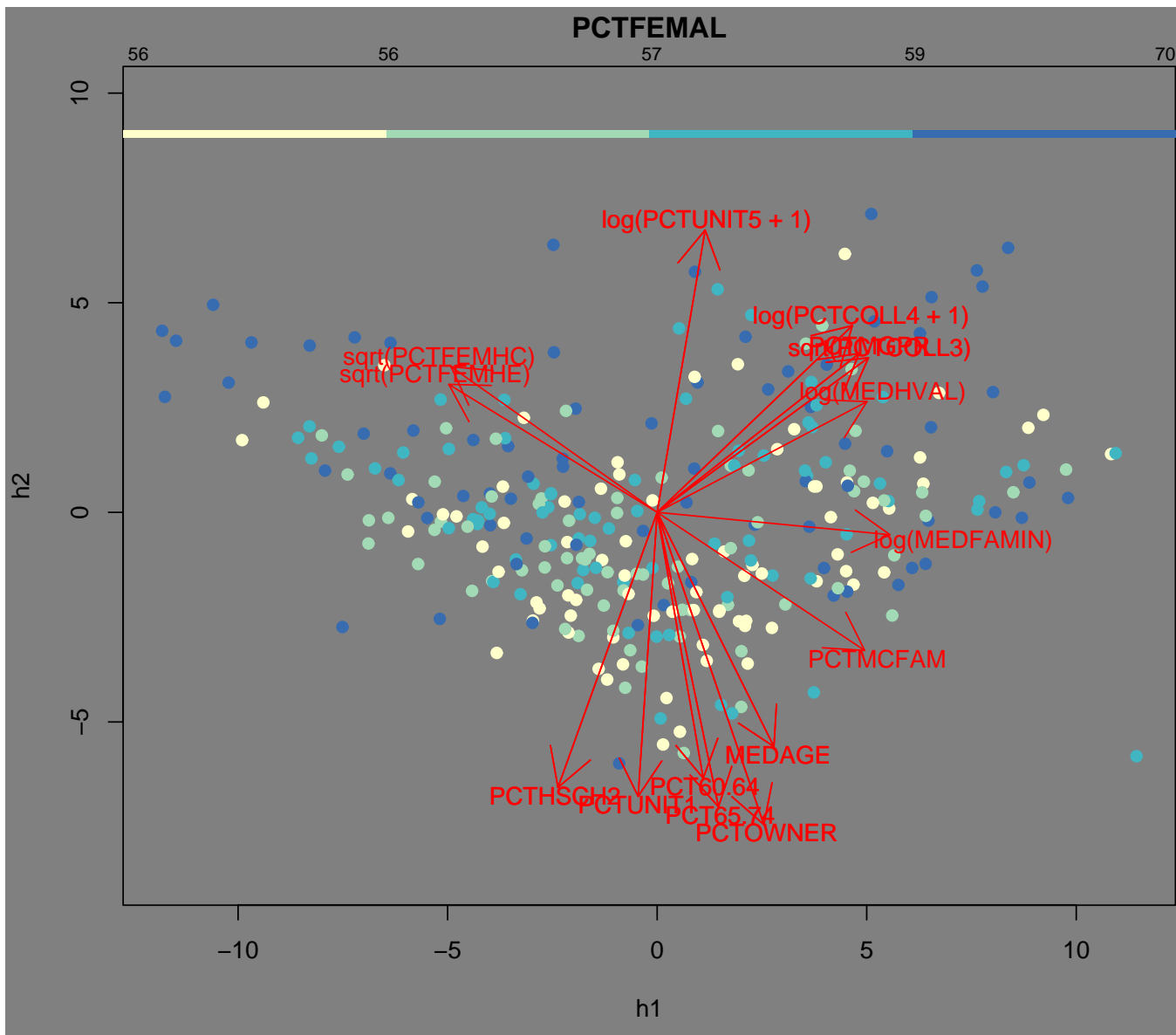




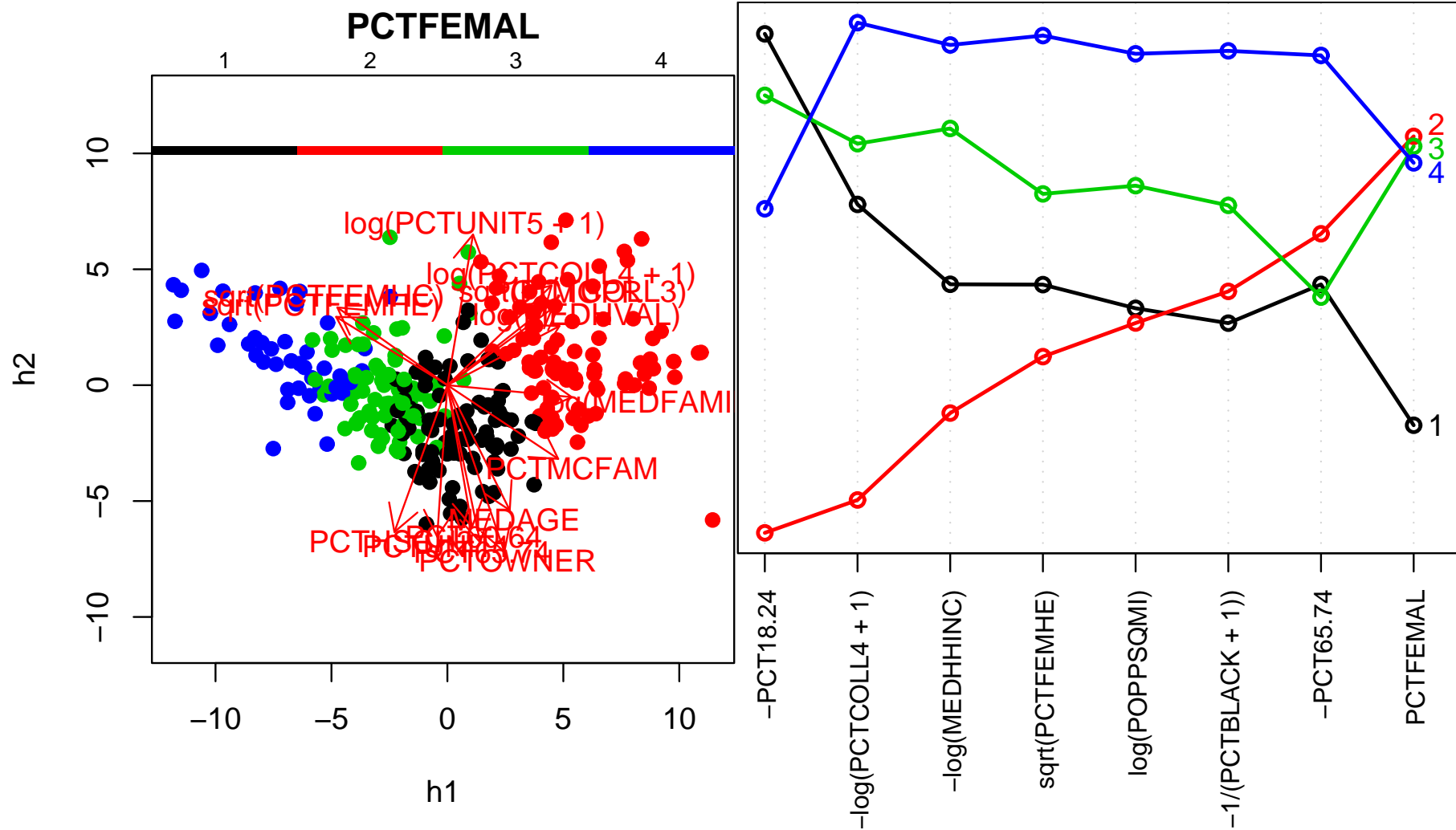




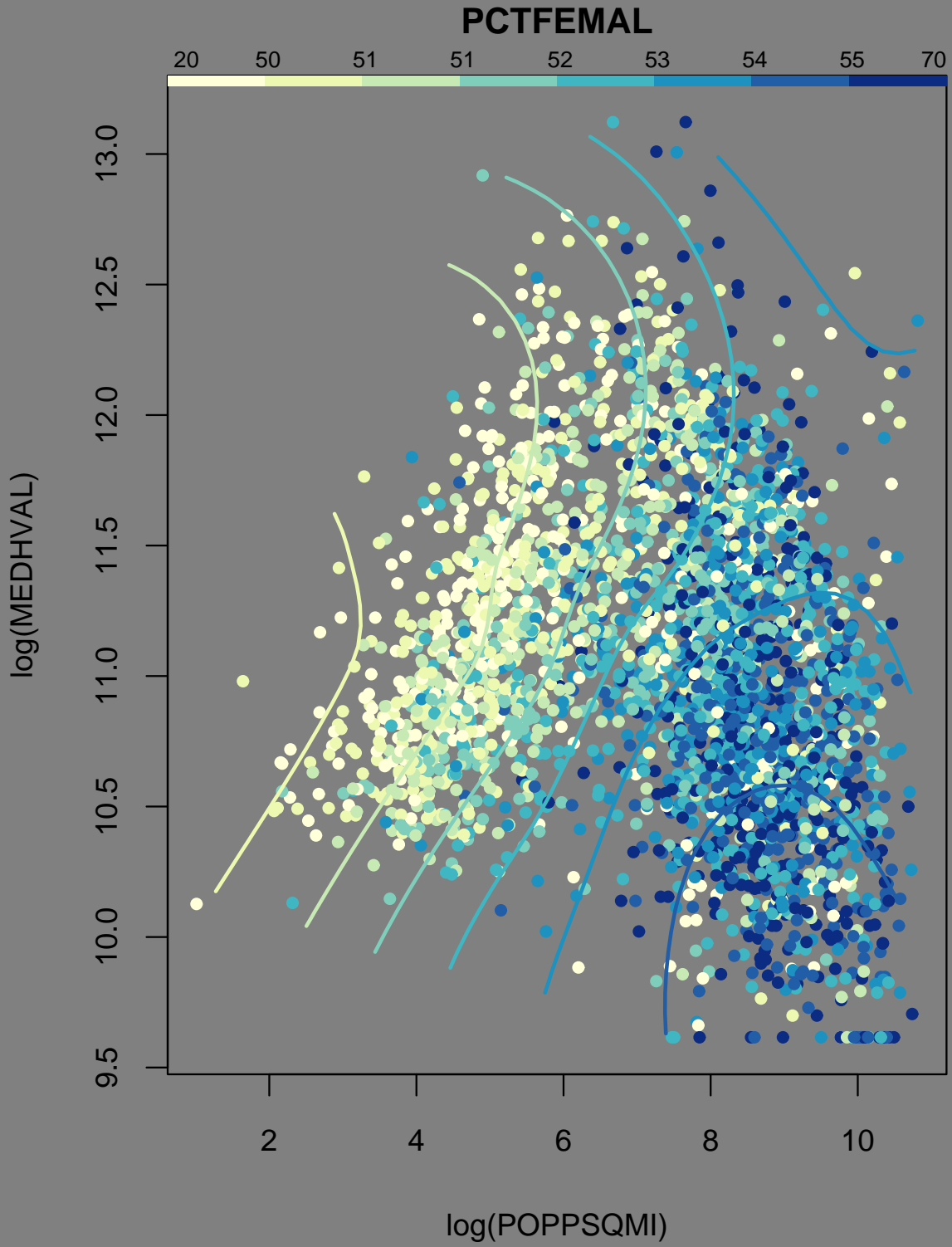


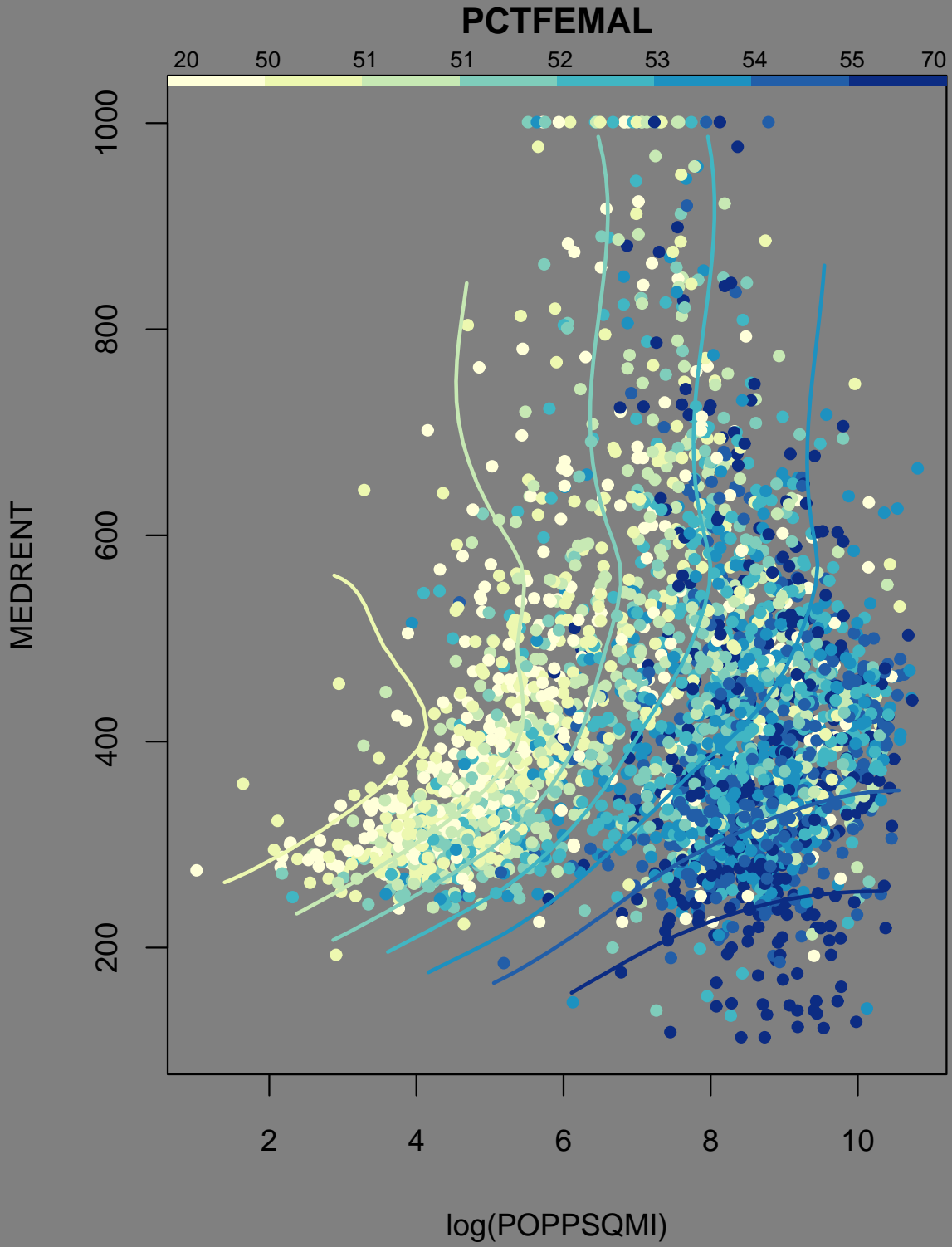


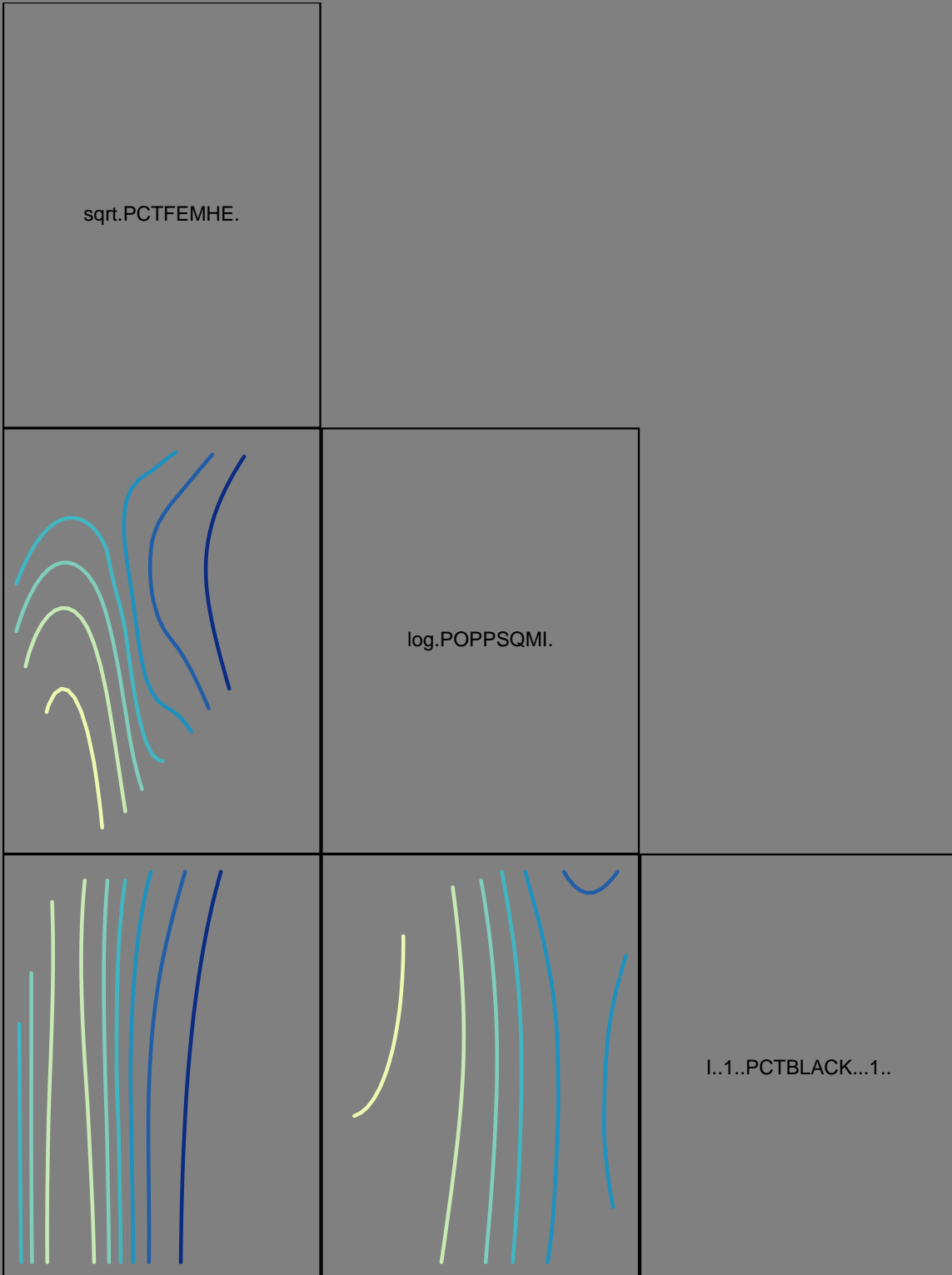
PCA projection of high response group (high PCTFEMAL)



Parallel-coordinates of high response group







Ranking predictors of PCTFEMAL