

36-315: Statistical Graphics and Visualization

Handout 14

Date: March 5, 2003

How to explain outliers in a 2D scatterplot:

- Color or bubble by a third variable. If the outlier is not unusual among data with the same color or size, then you have explained the outlier.
- Rotate in the direction of a third variable. If the outlier lies in a special spot on the third axes (away from its 2D neighbors), then you have explained the outlier.

The equal-count rule for colors and bubble sizes:

- Different from equal spacing
- Each range covers the same amount of data. Ensures balanced use of color/area.
- Plot is identical under any monotonic transformation of z .

Slice plot—Slice z axis by equal-count rule, and display each slice separately. Slices can also be superposed using color (which is the same as a color plot).

Prediction surface—Depicts the mean of the z variable at a given (x, y) . Computed by averaging z values in a small neighborhood around (x, y) . `span` controls the size of the neighborhood.

Prediction surfaces are unreliable in areas with little data

- Isolated points cause flaring at edges

Choosing the viewpoint:

- Want to be high enough to see everything, low enough to see small hills.
- Better to look up a hill than down a chasm.

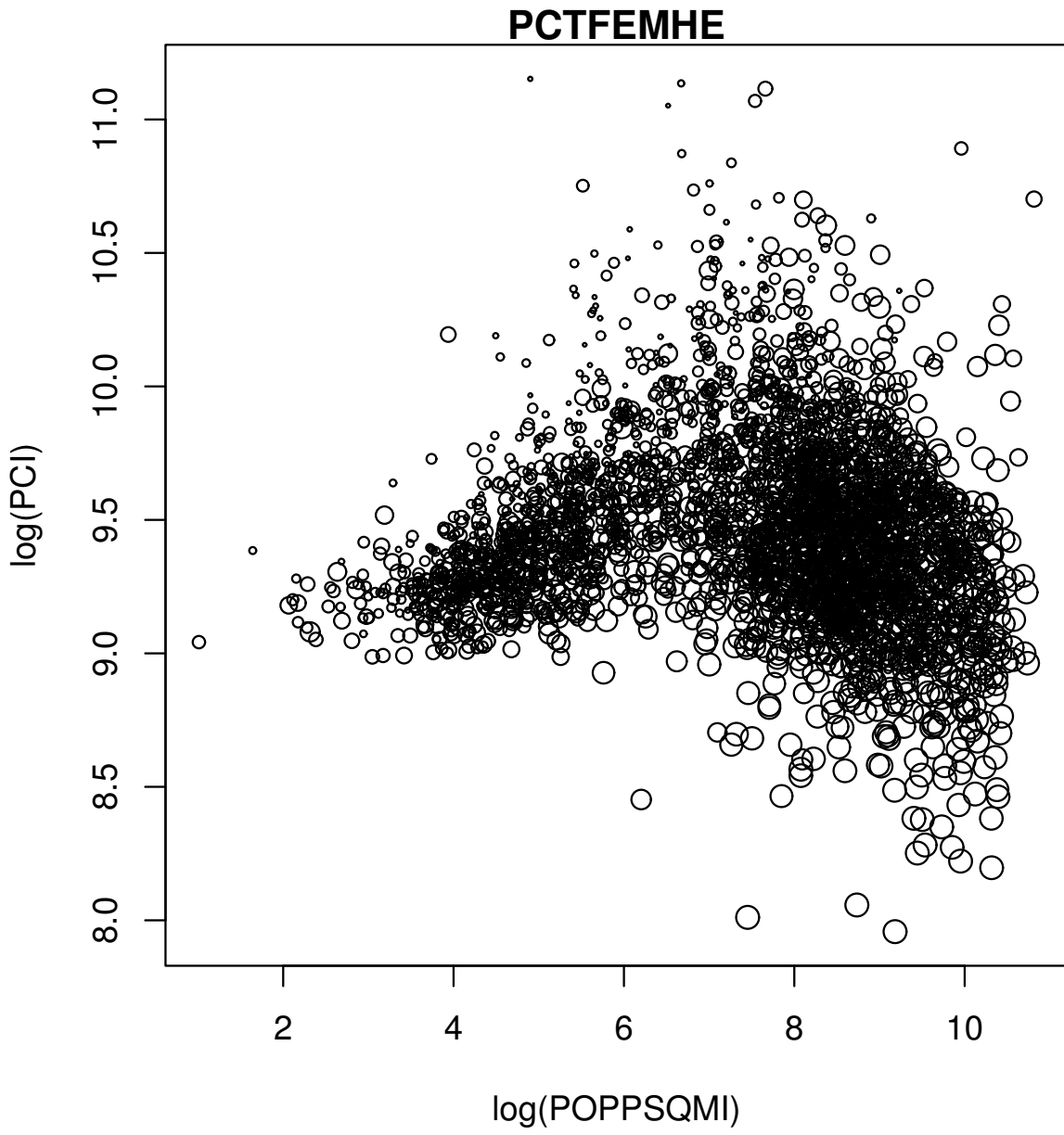
Advantages of three-dimensional plots:

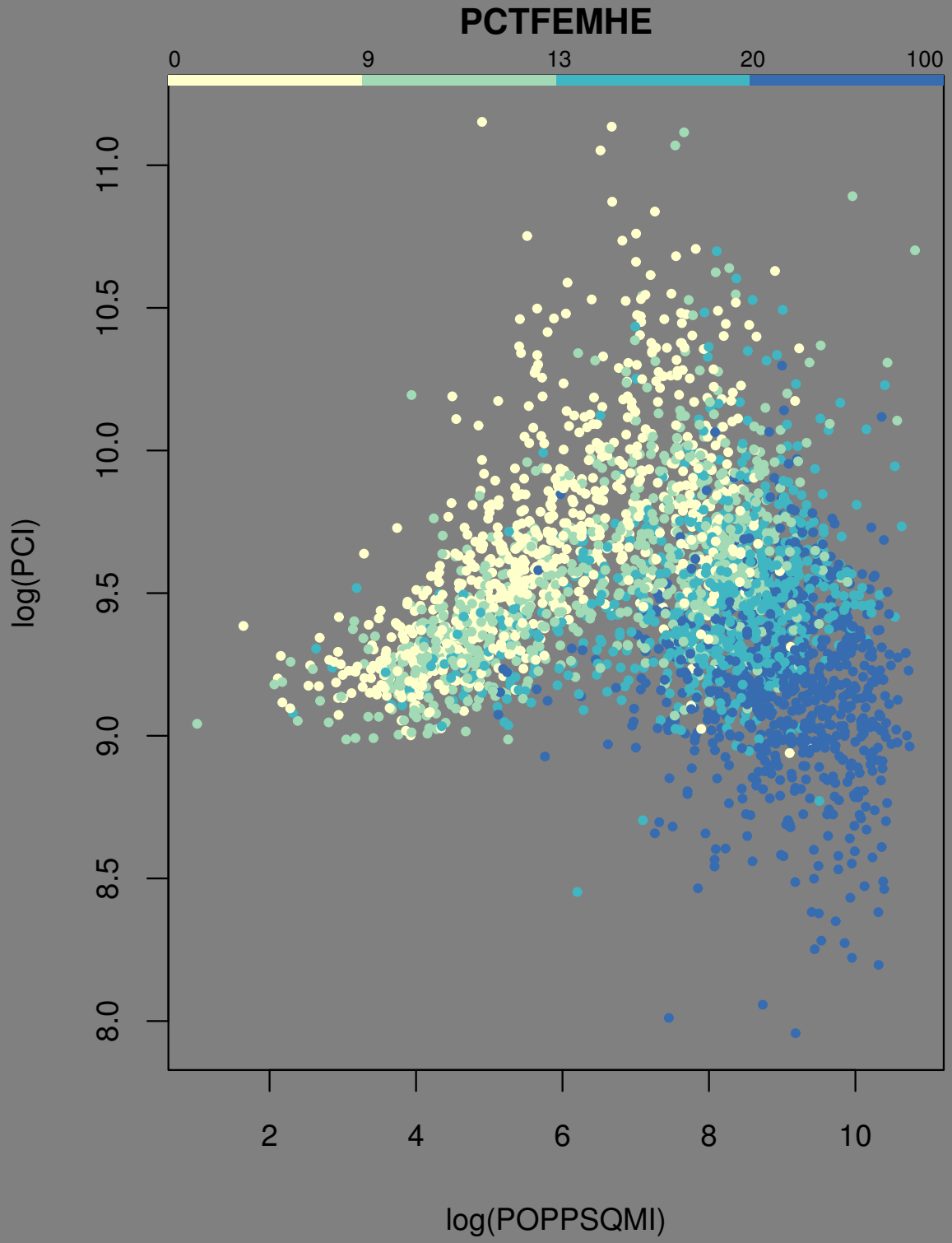
- See the big picture quickly

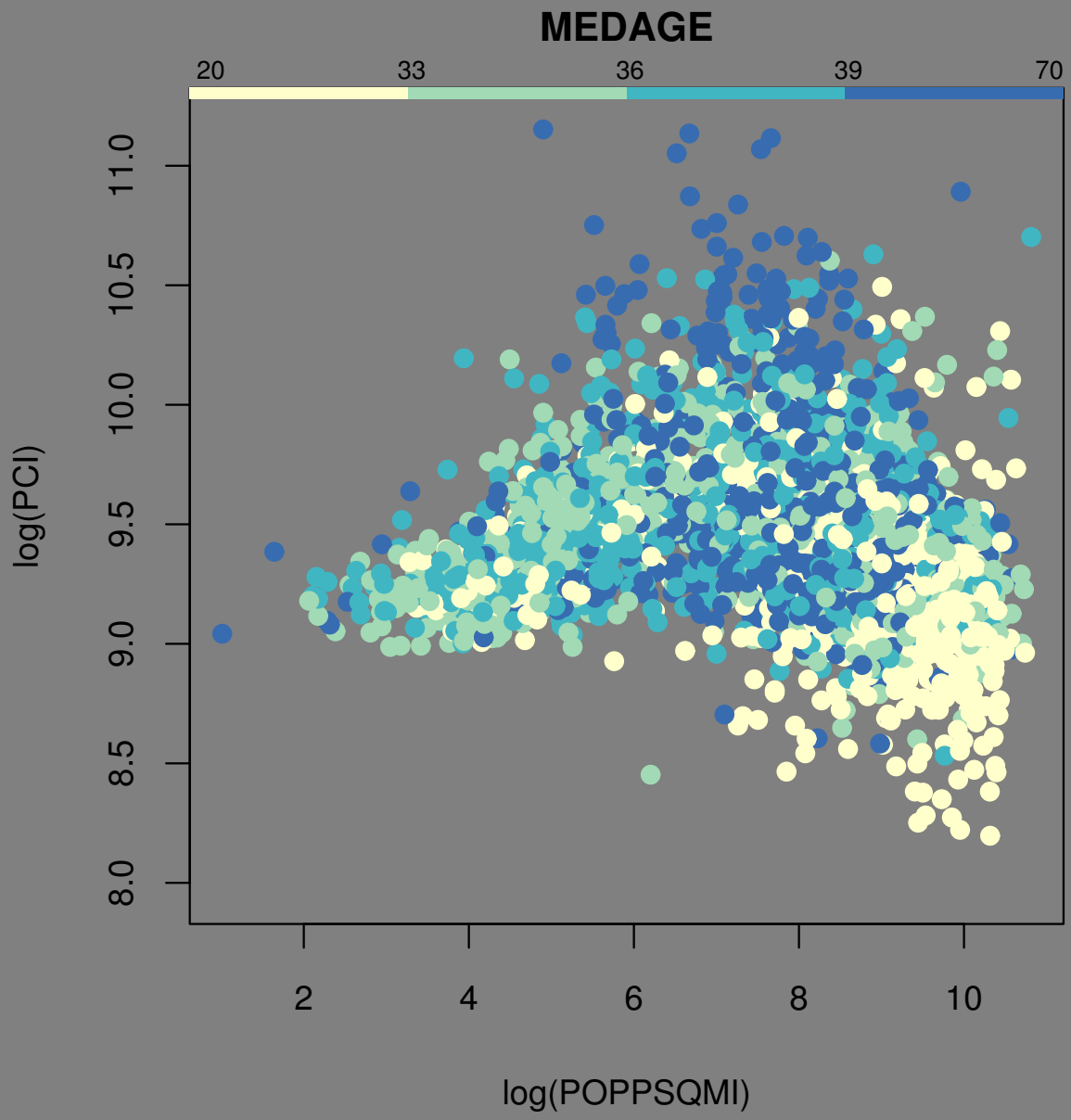
Disadvantages of three-dimensional plots:

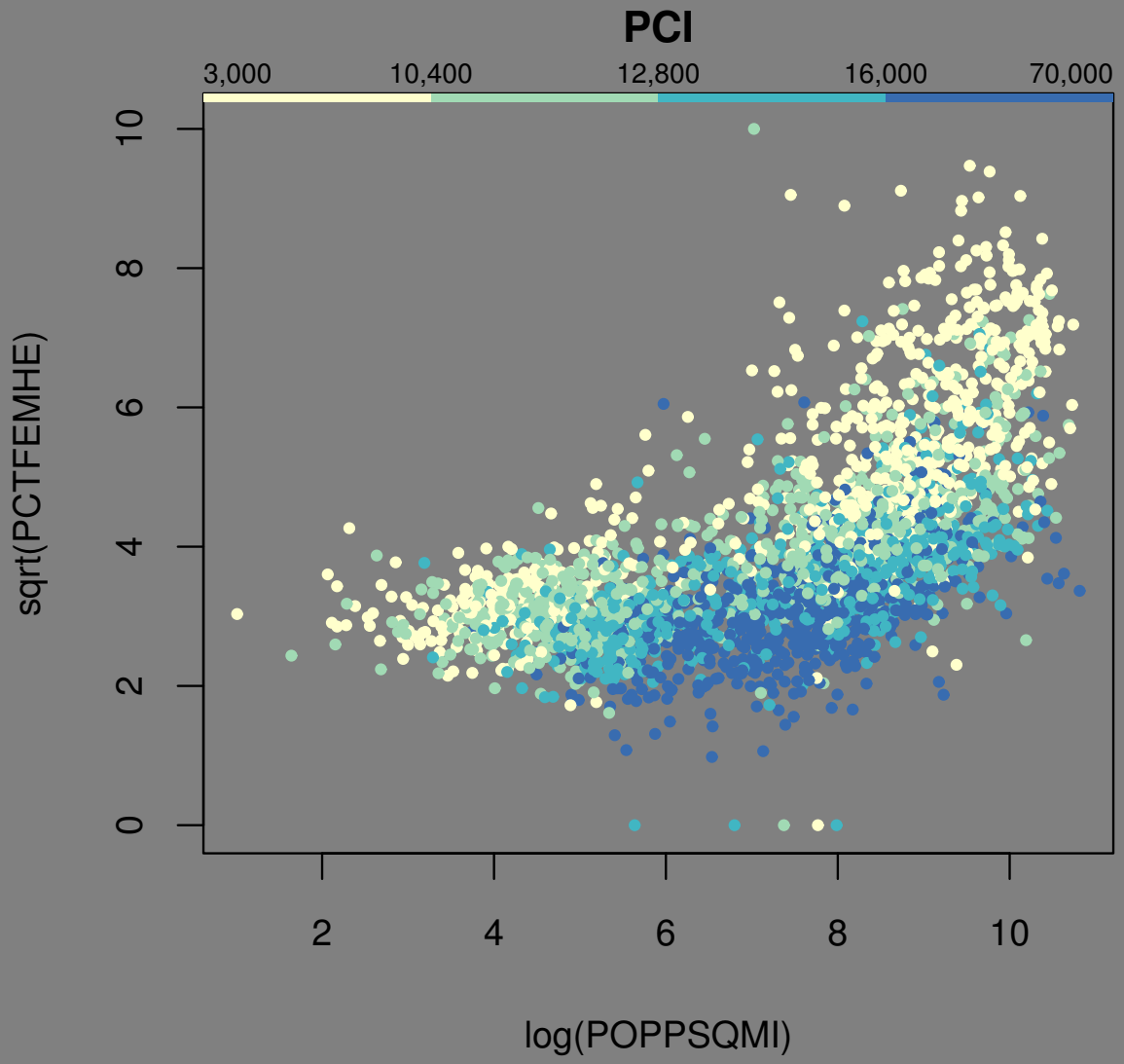
- Can't look up position on the axes
- Have to choose good viewpoint
- Have to choose good transformation for z (as well as x and y)

Explaining outliers:

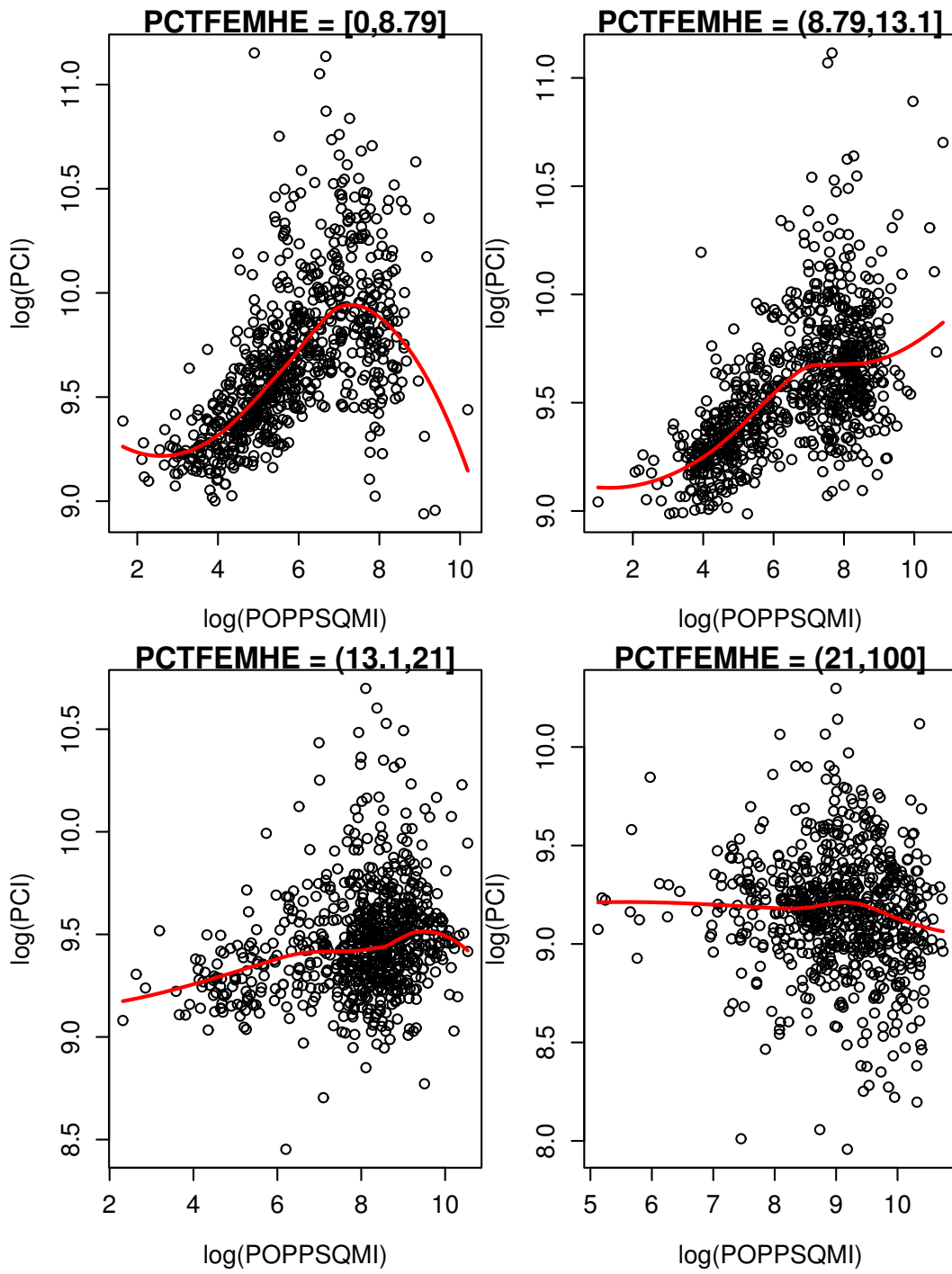




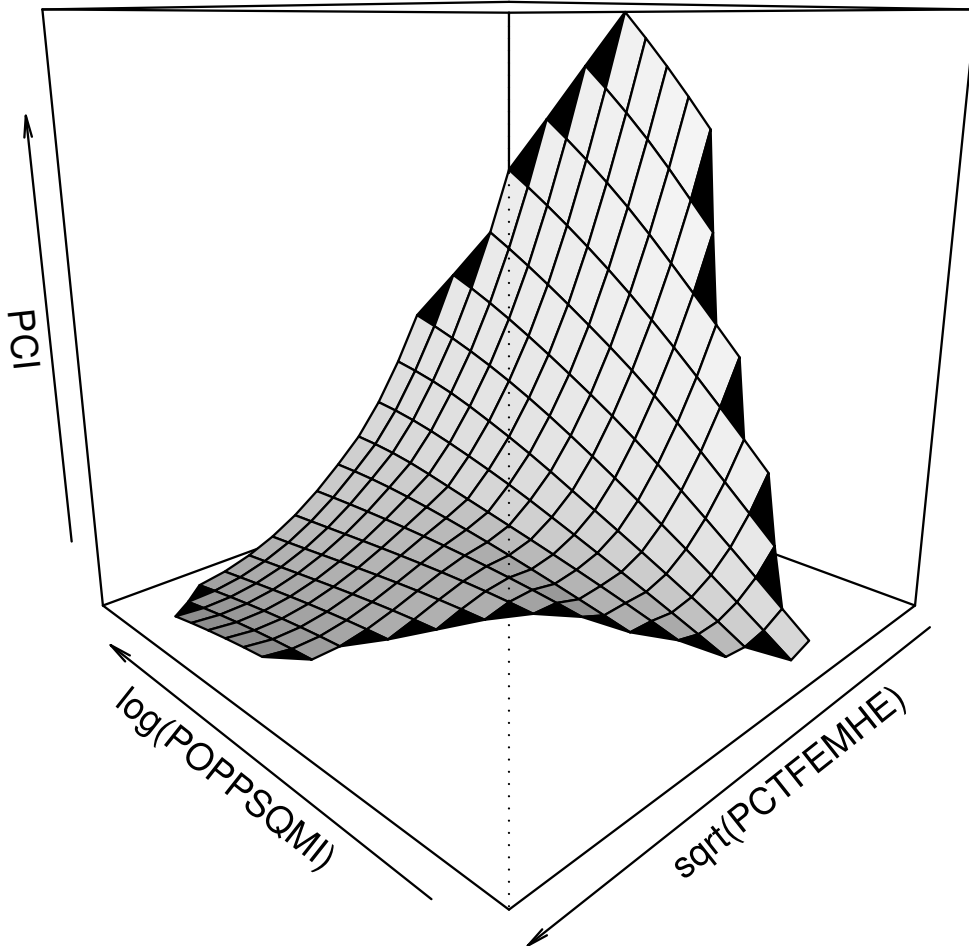




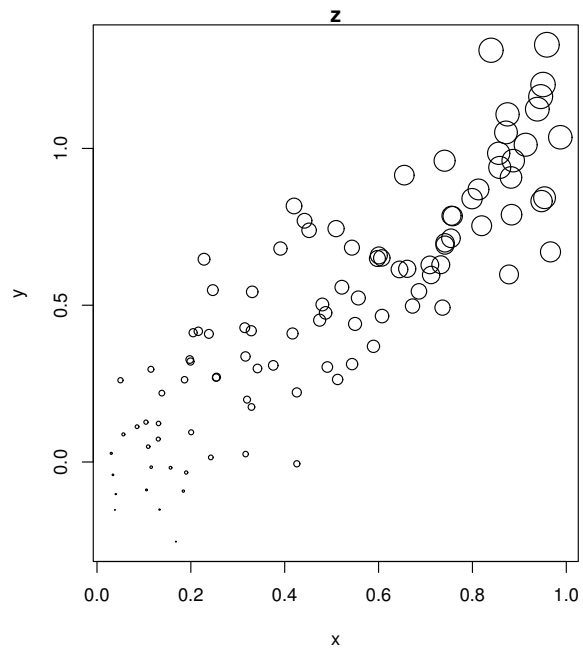
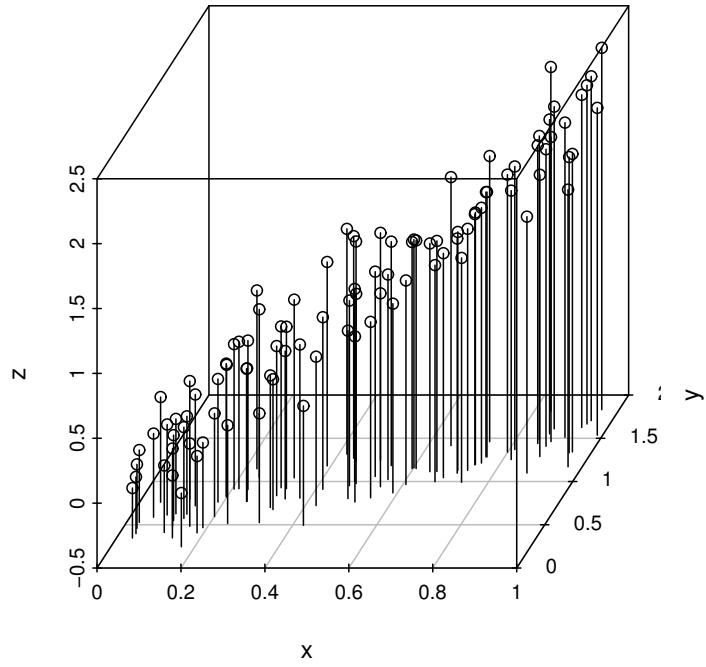
Slicing on PCTFEMHE:



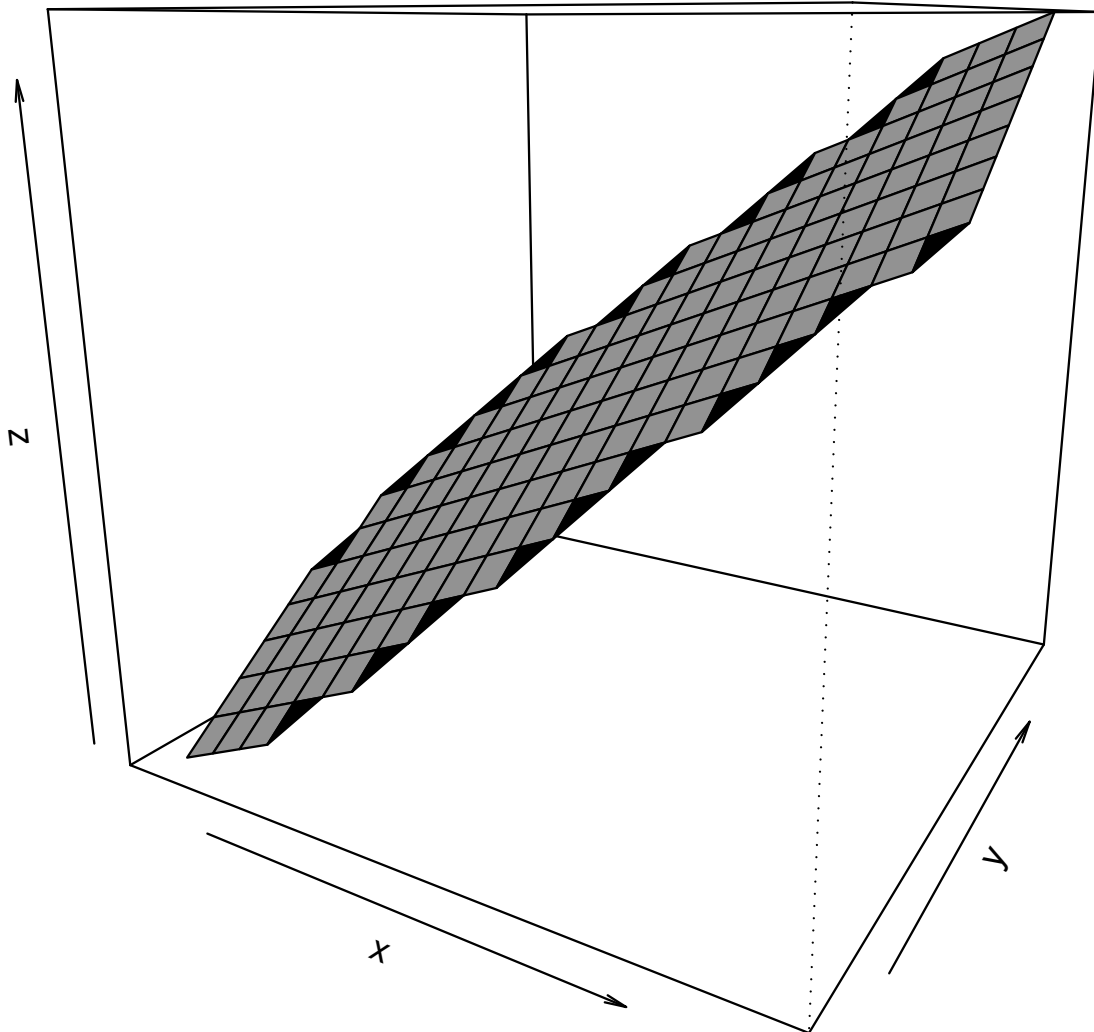
Prediction surface:



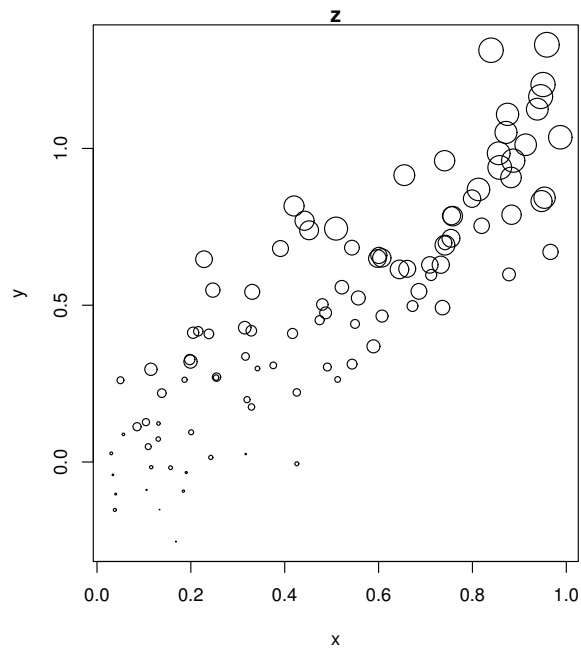
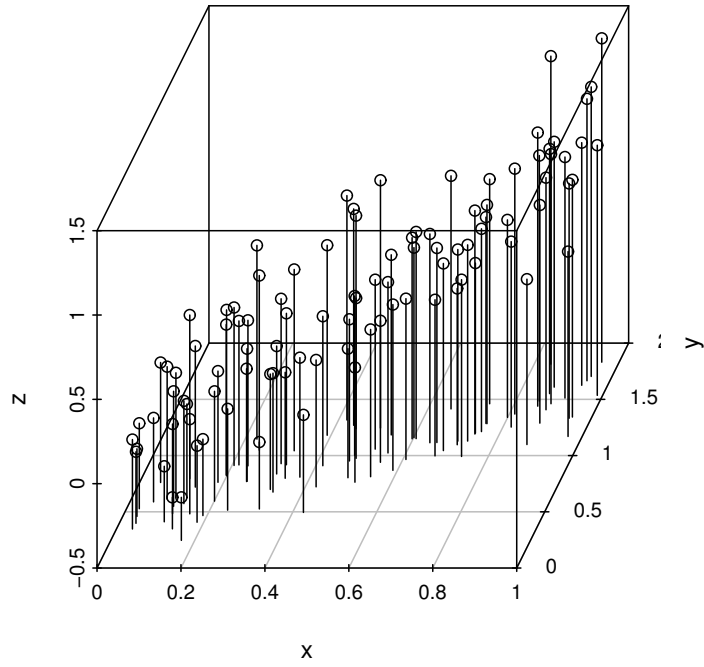
Data set 1:



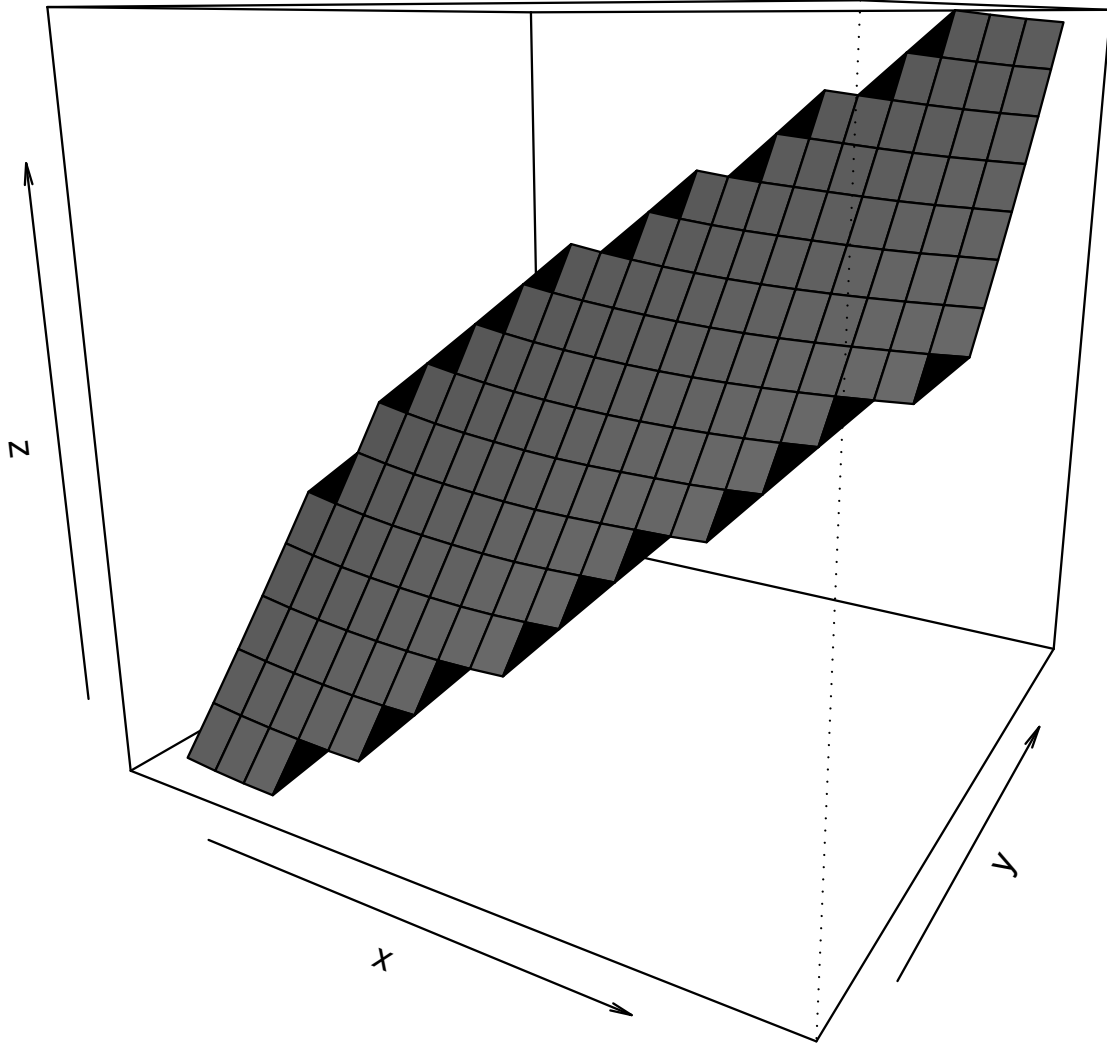
Prediction surface:



Data set 2:



Prediction surface:



Viewpoints

