

Data Profiling Report

- Basic Statistics
 - Raw Counts
 - Percentages
- Data Structure
- Missing Data Profile
- Univariate Distribution
 - Histogram
 - Bar Chart (with frequency)
 - QQ Plot
 - QQ Plot (by critical_temp)
- Correlation Analysis
- Principal Component Analysis
- Bivariate Distribution
 - Boxplot (by critical_temp)
 - Scatterplot (by critical_temp)

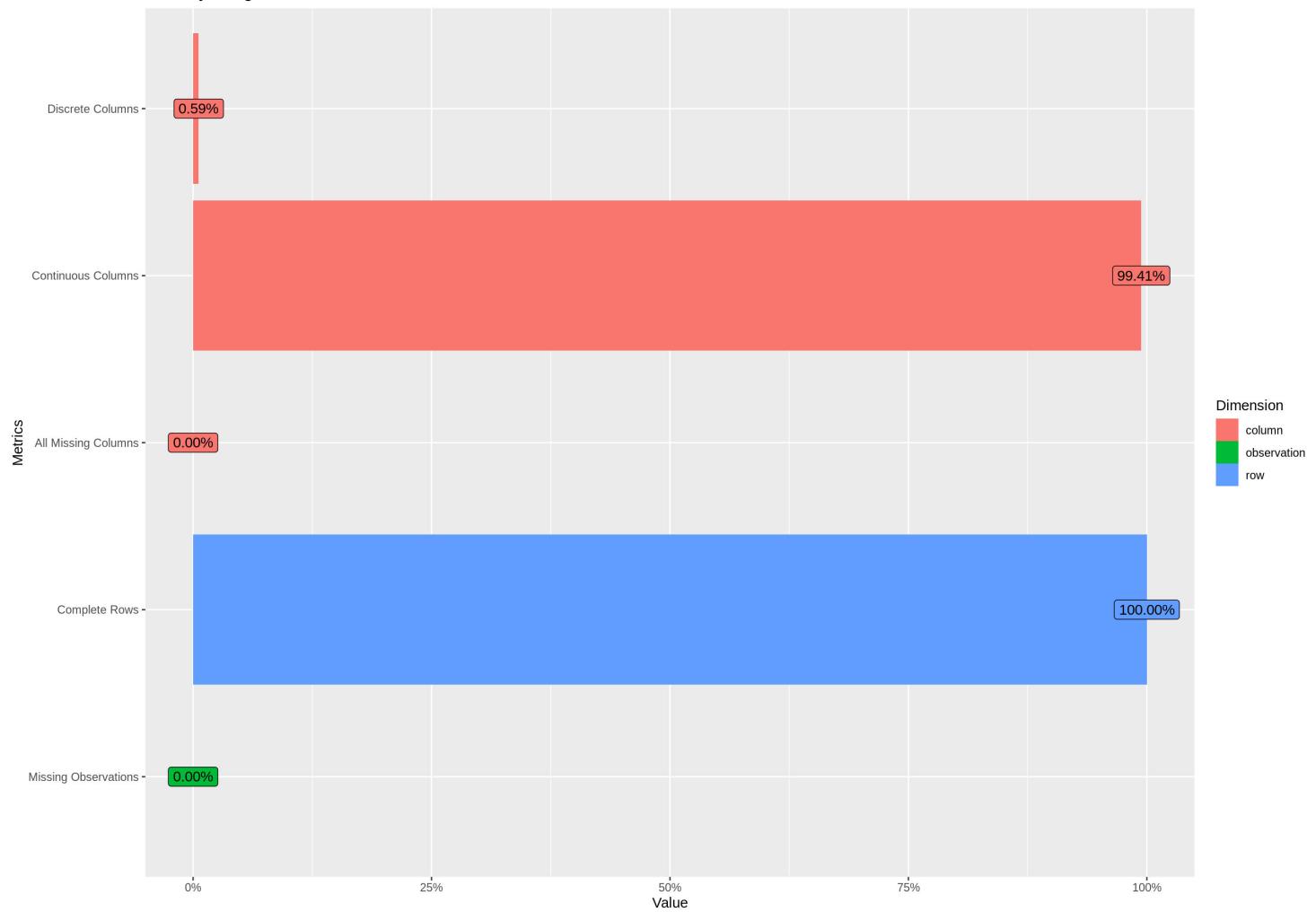
Basic Statistics

Raw Counts

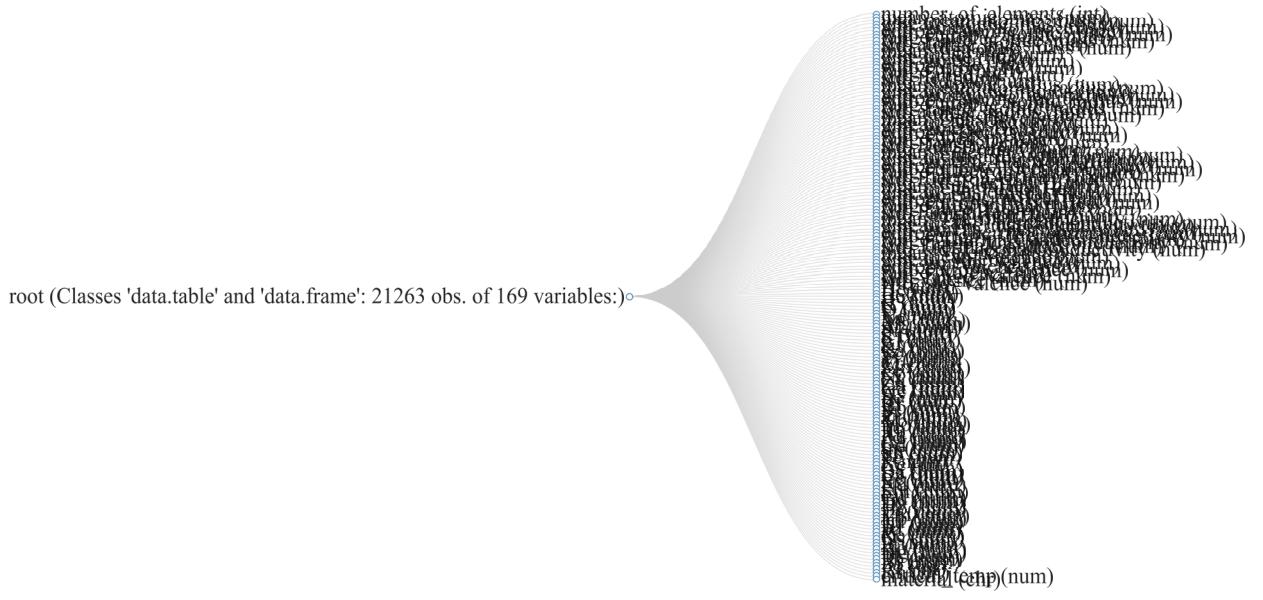
Name	Value
Rows	21,263
Columns	169
Discrete columns	1
Continuous columns	168
All missing columns	0
Missing observations	0
Complete Rows	21,263
Total observations	3,593,447
Memory allocation	27.6 Mb

Percentages

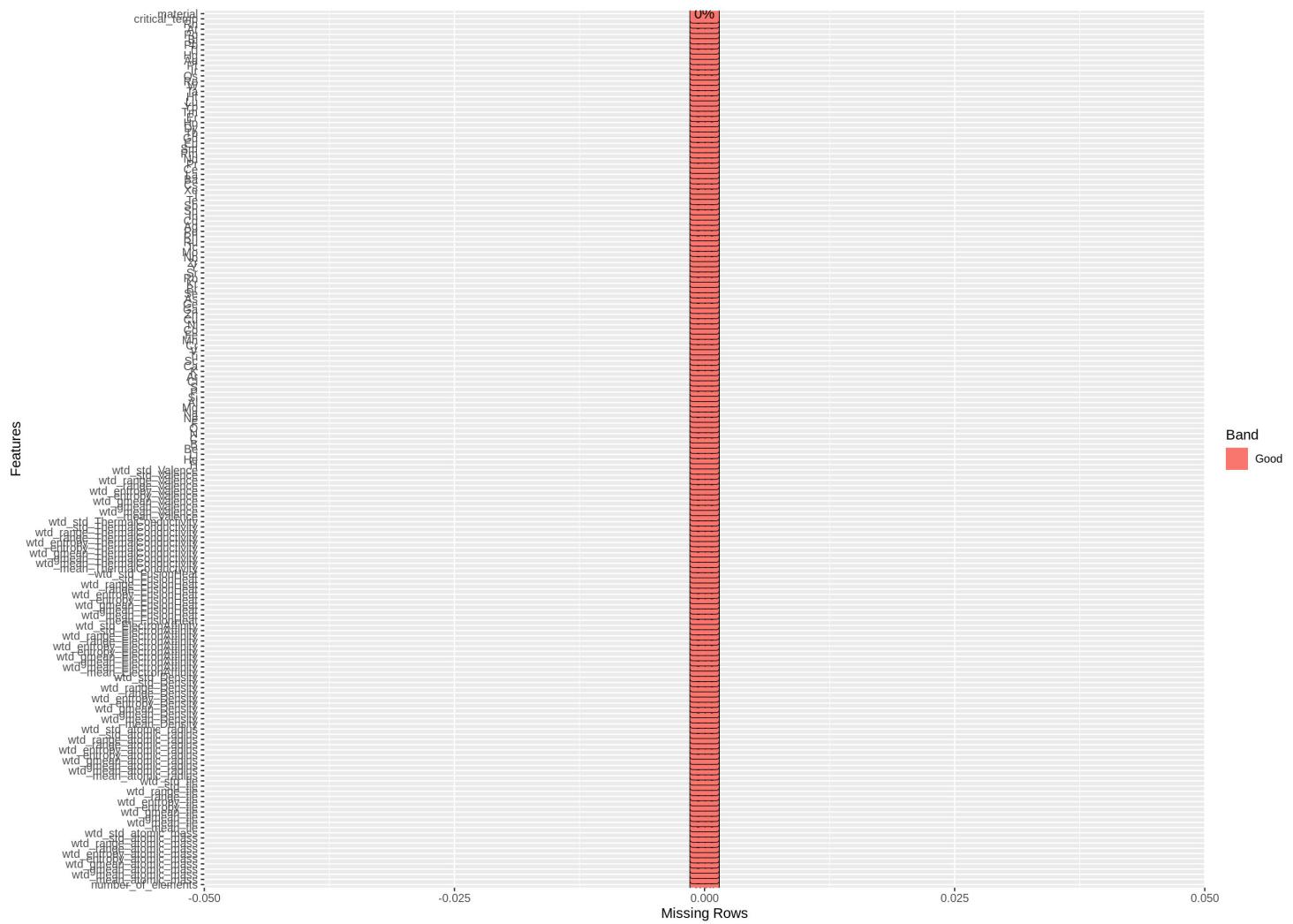
Memory Usage: 27.6 Mb



Data Structure

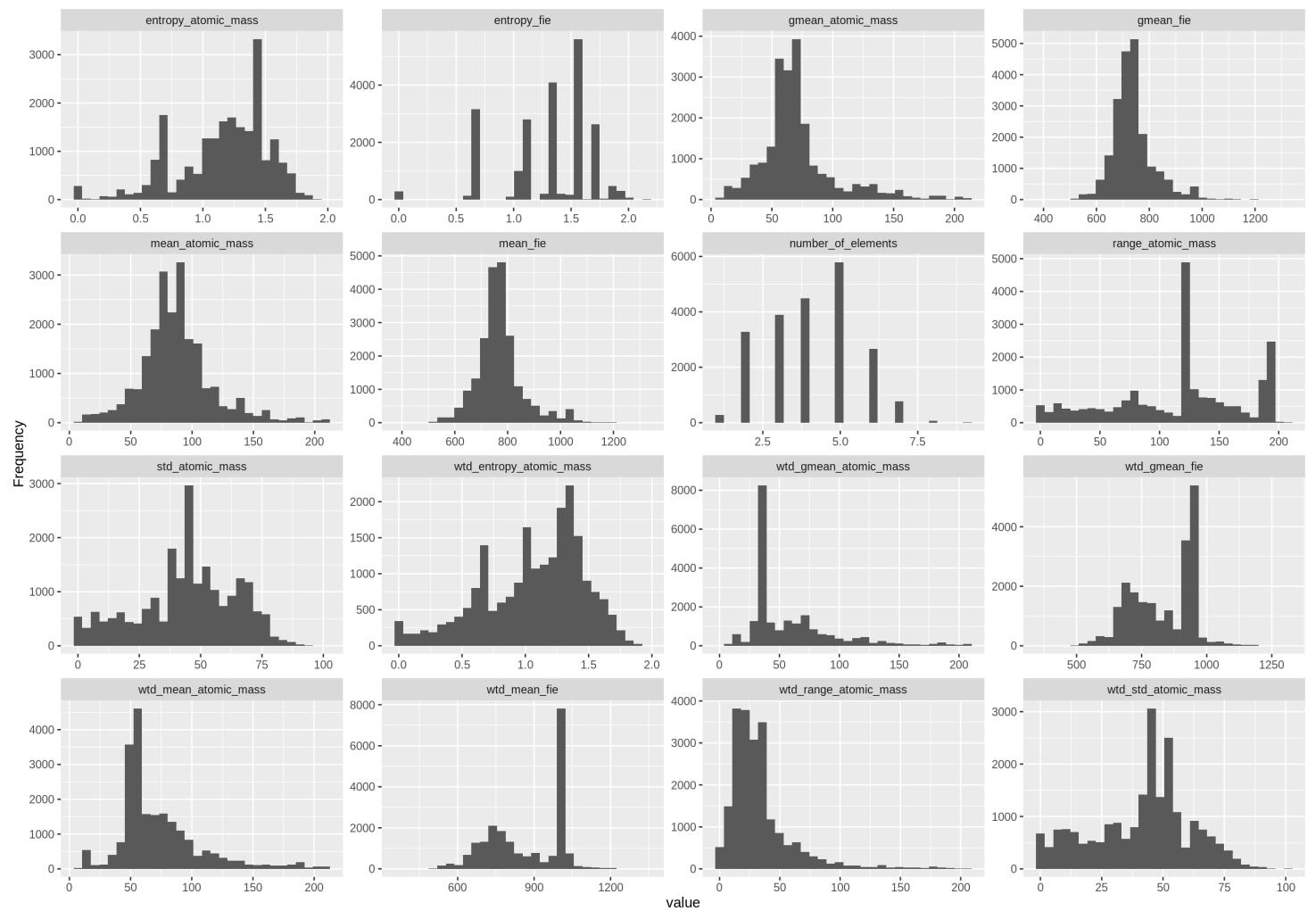


Missing Data Profile

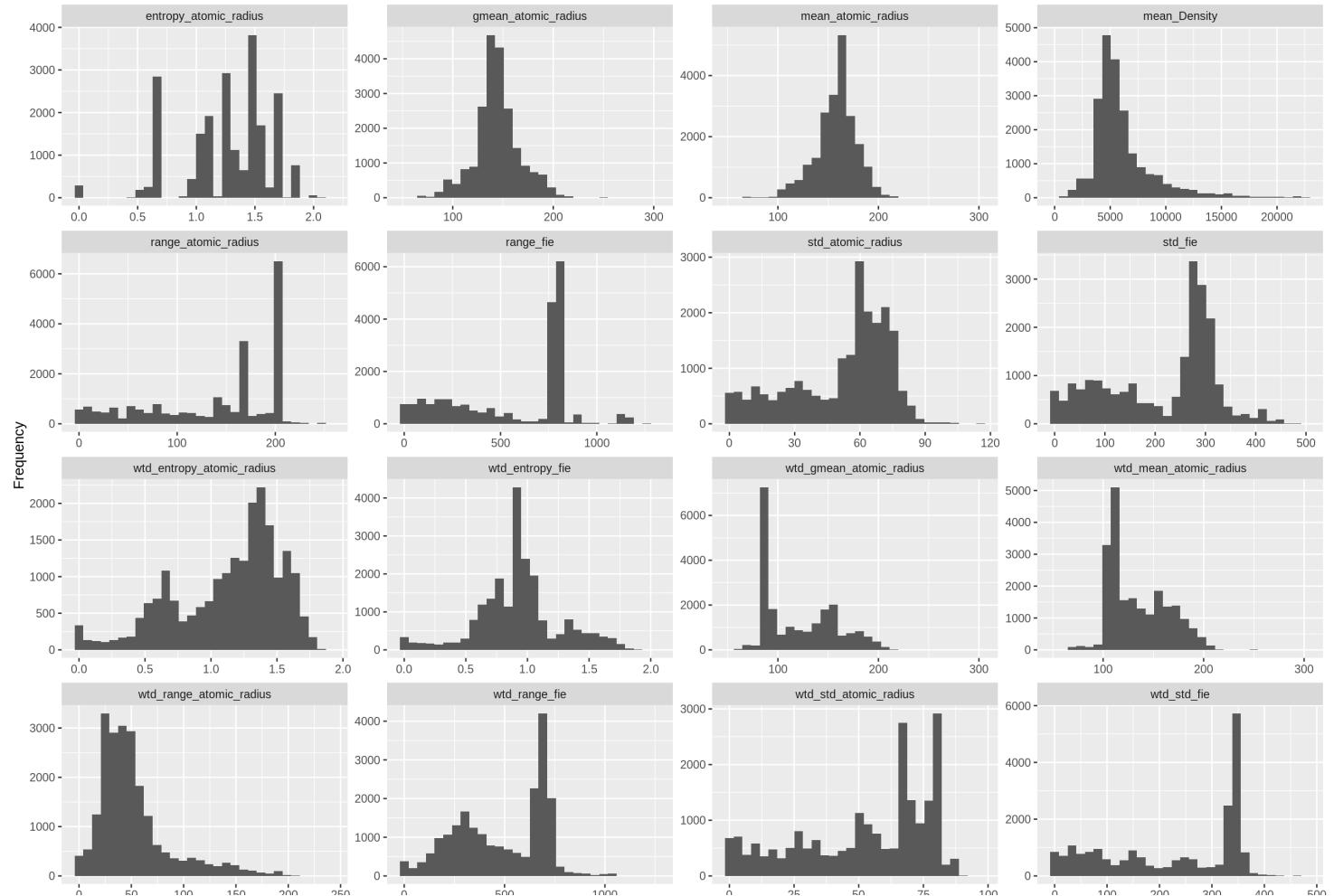


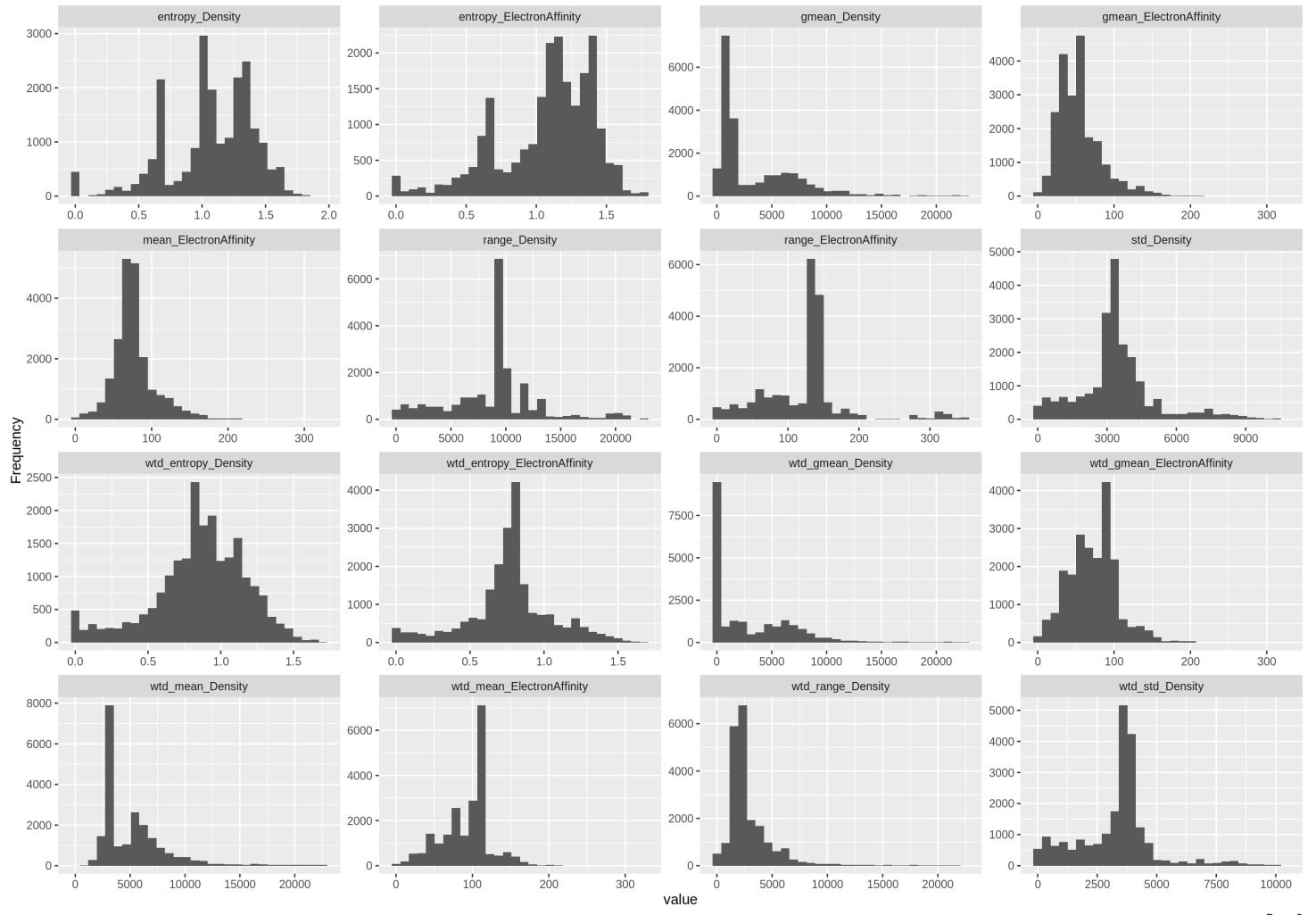
Univariate Distribution

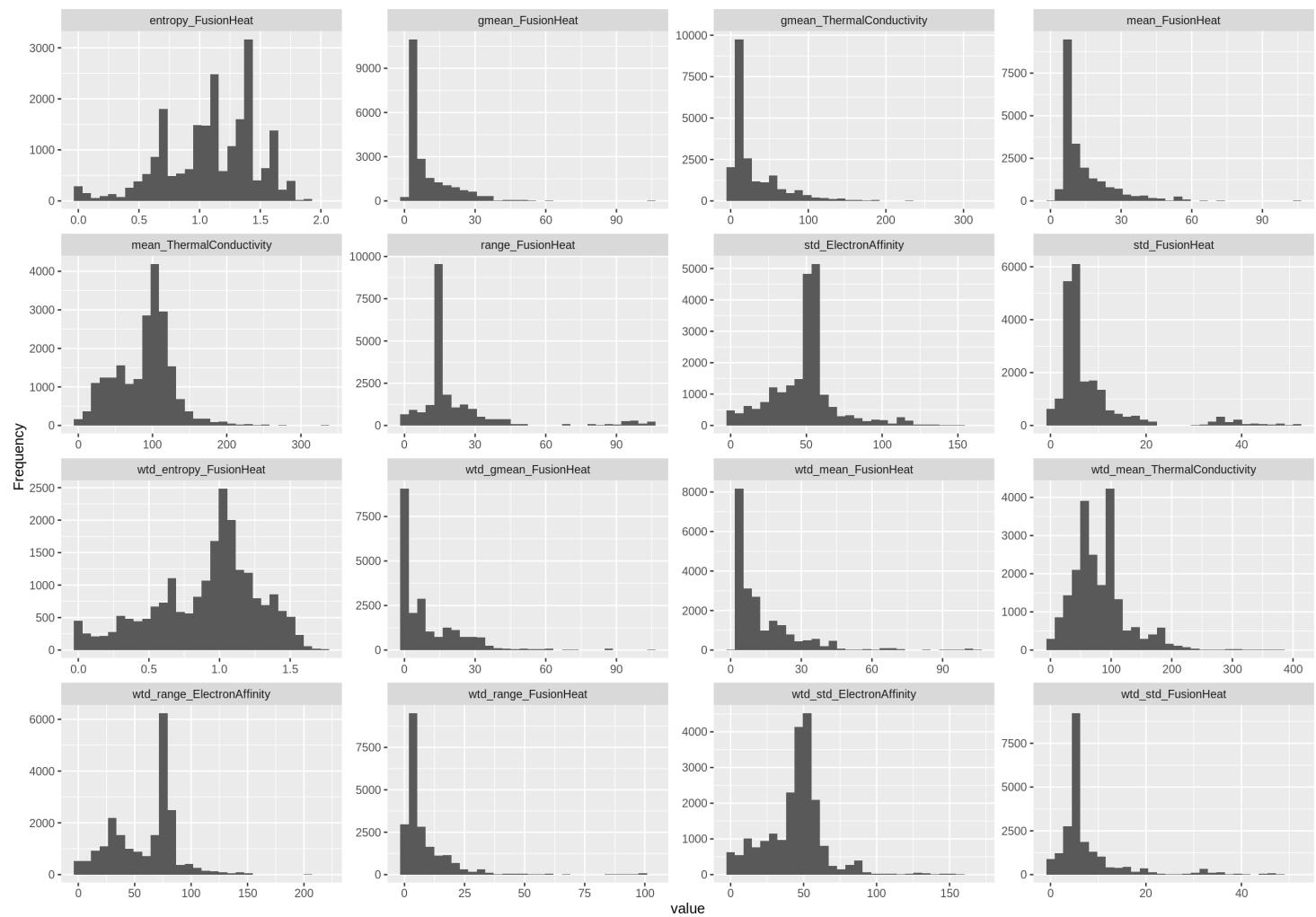
Histogram

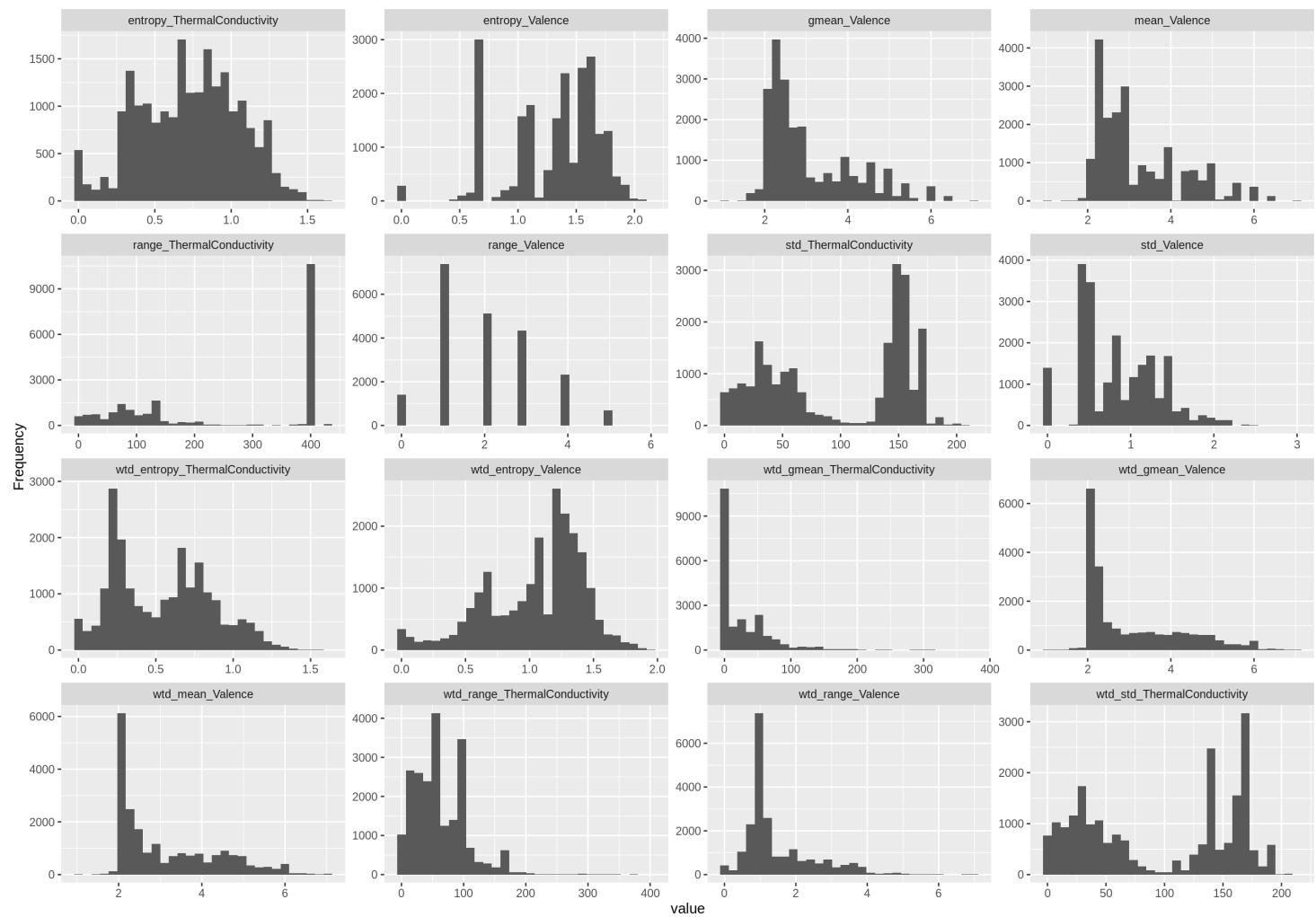


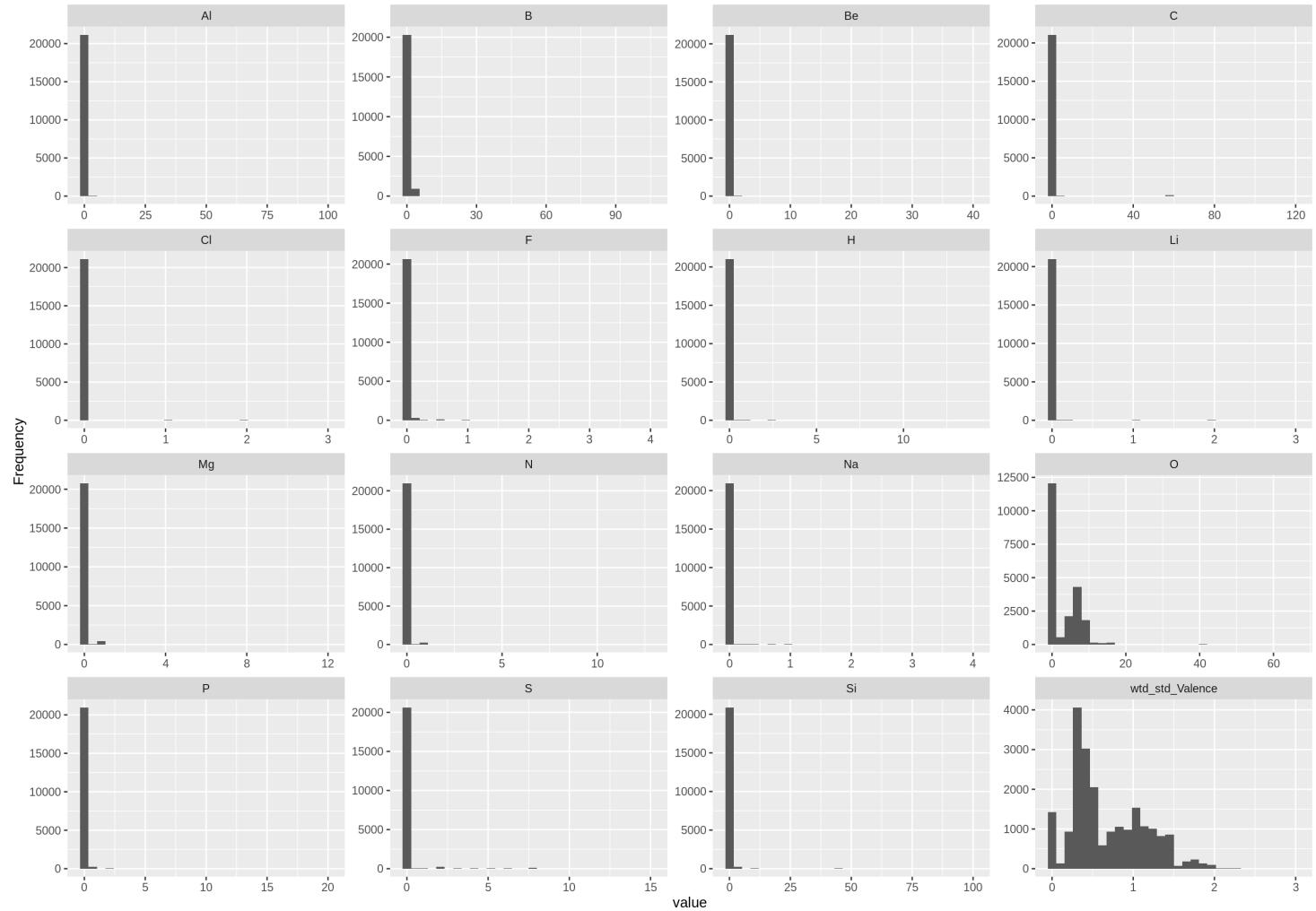
Page 1

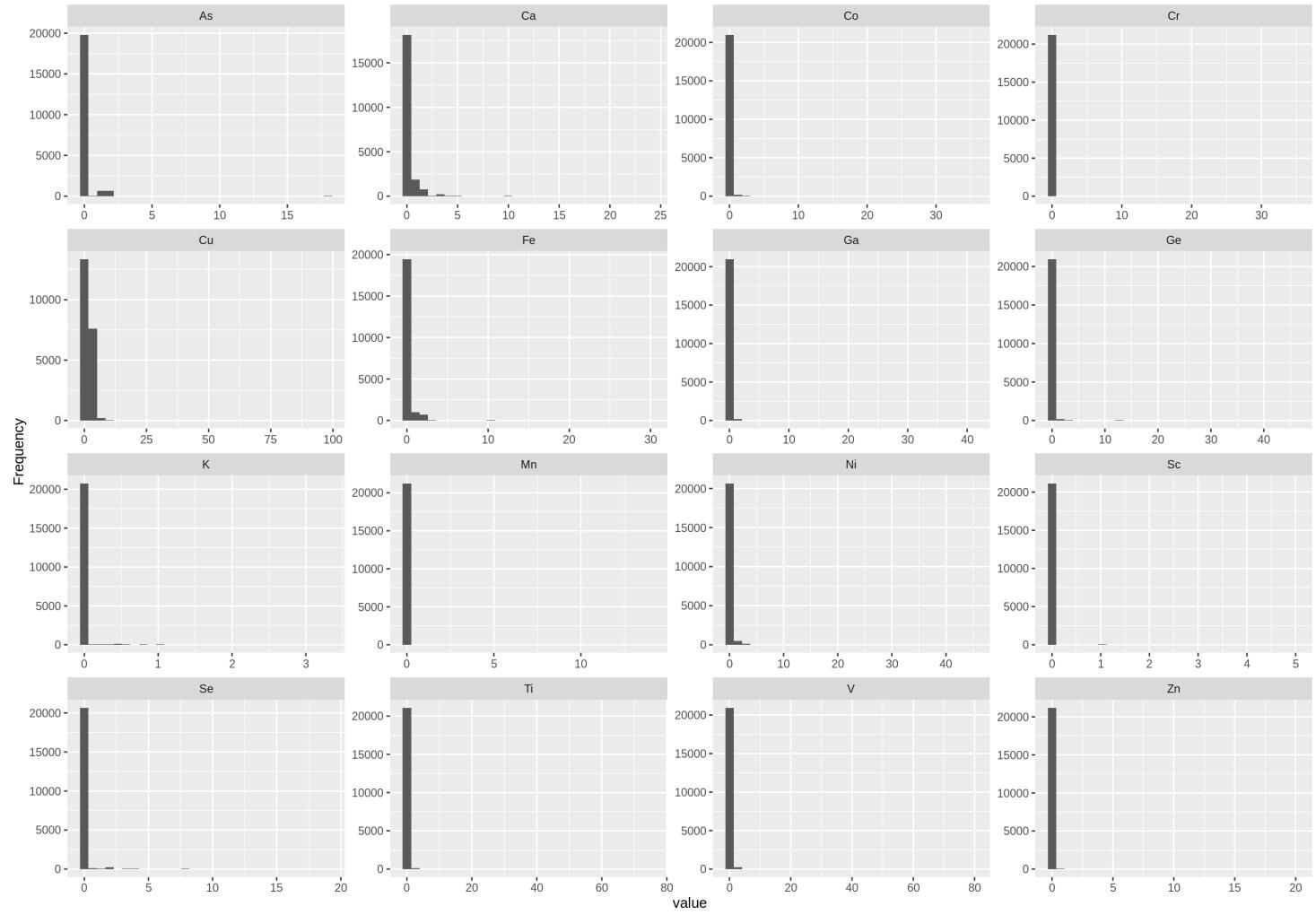


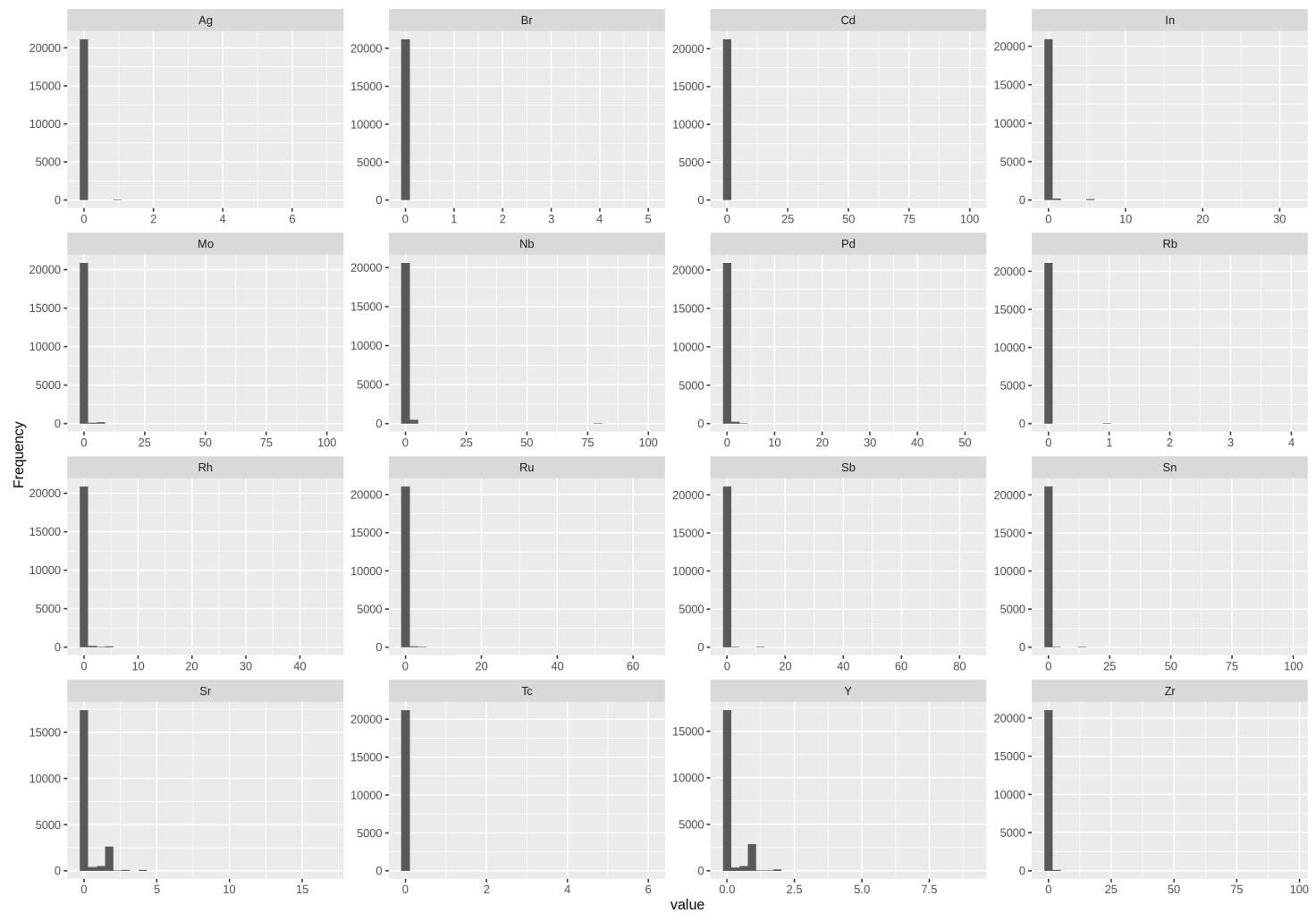


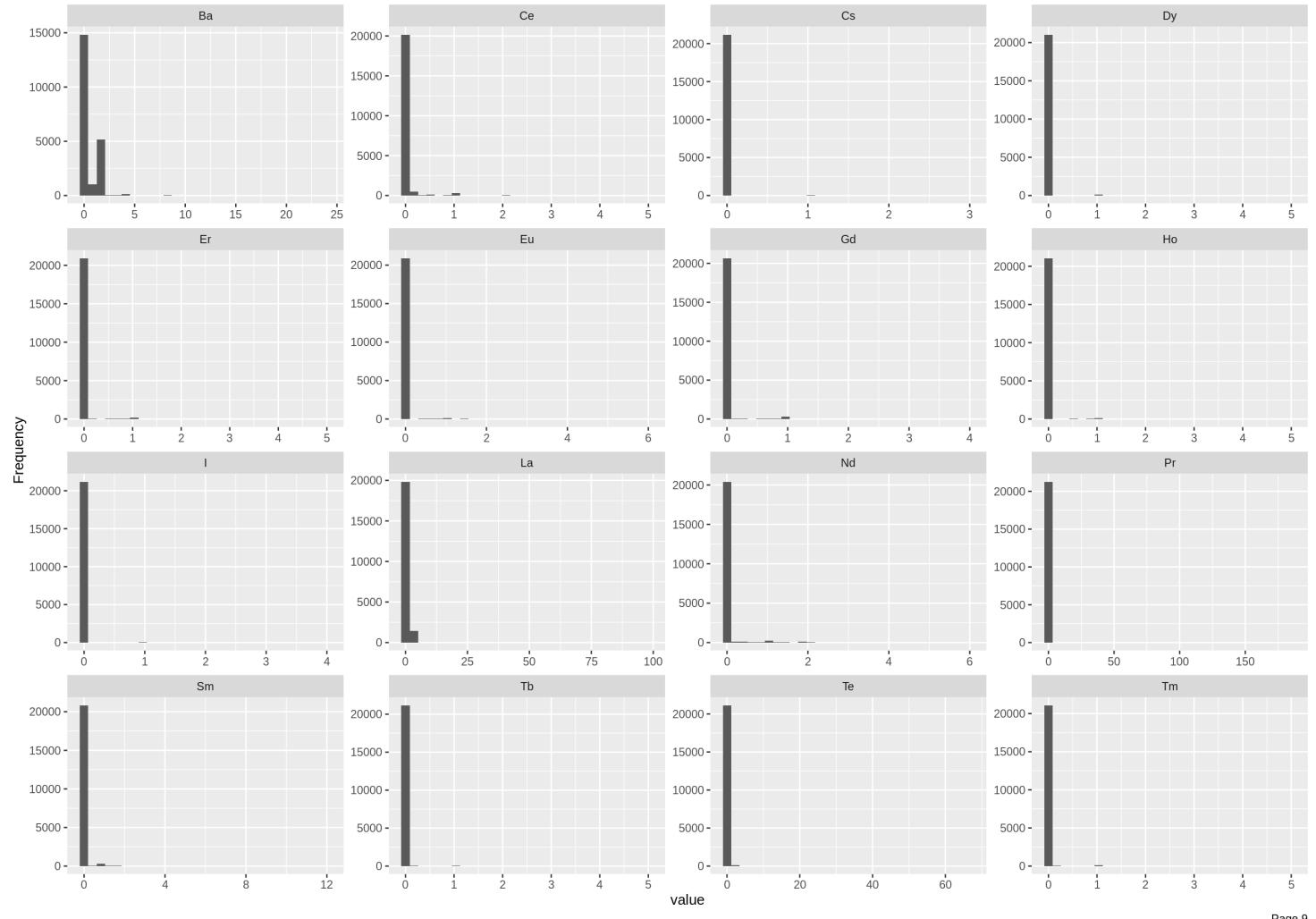


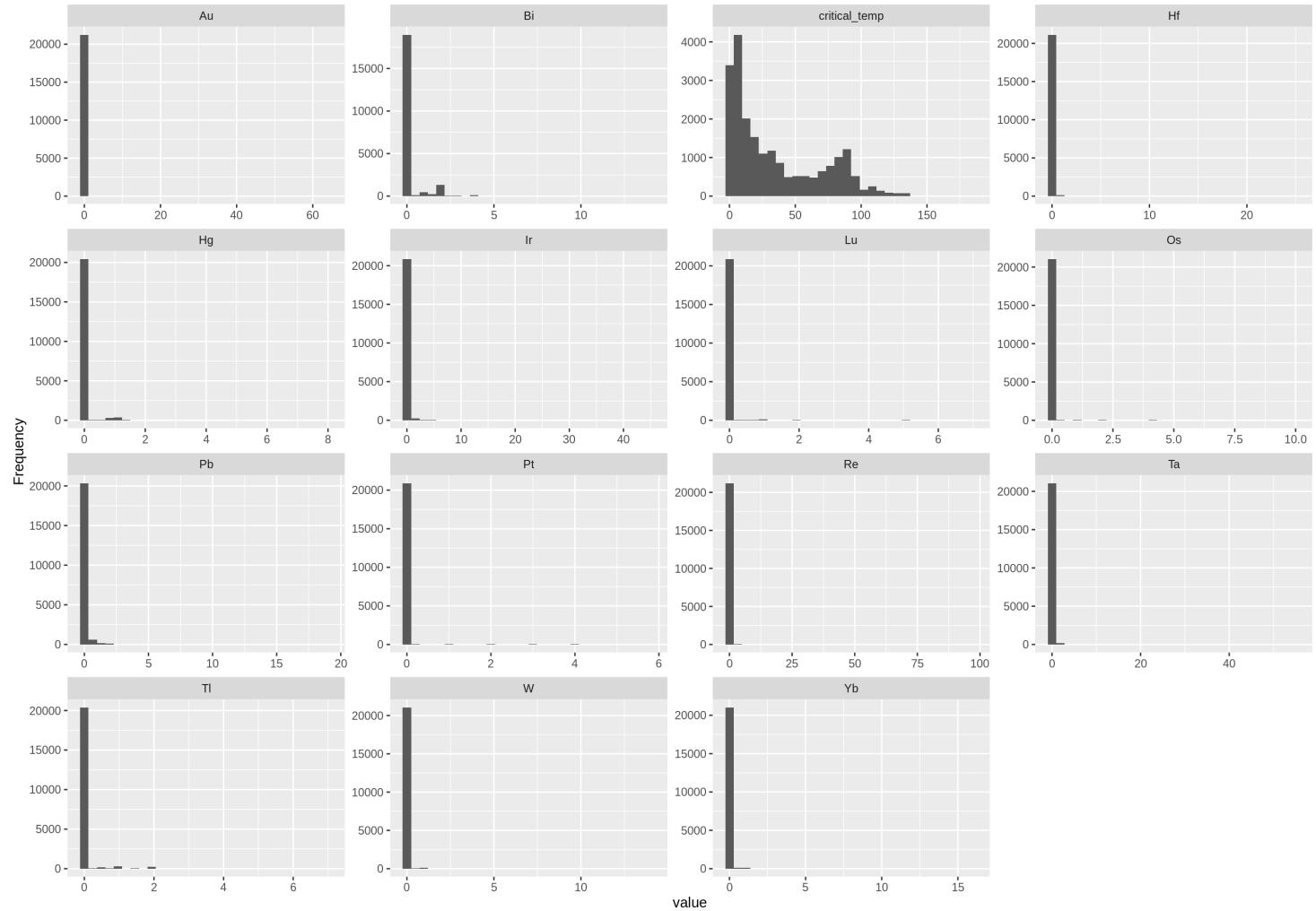






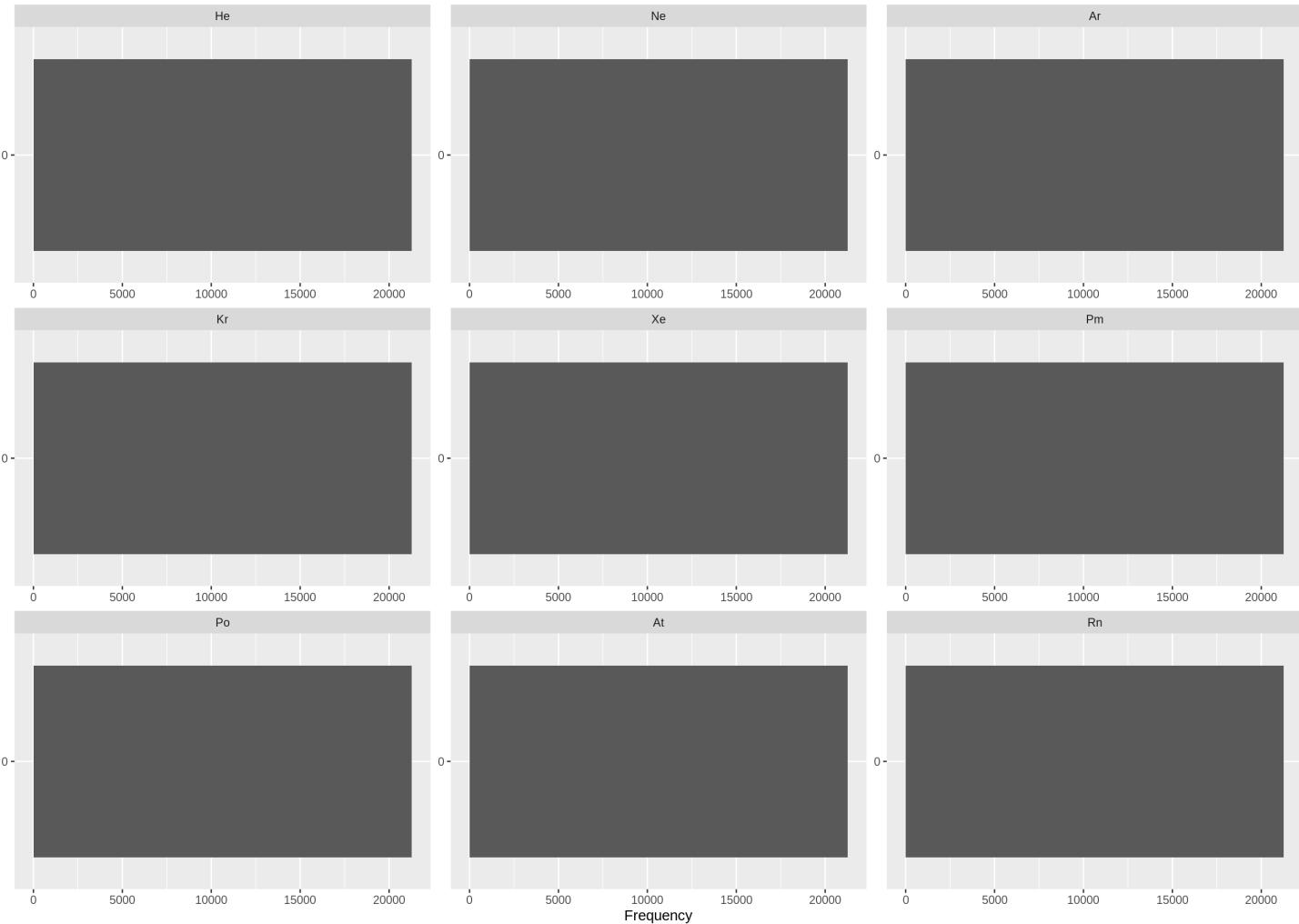




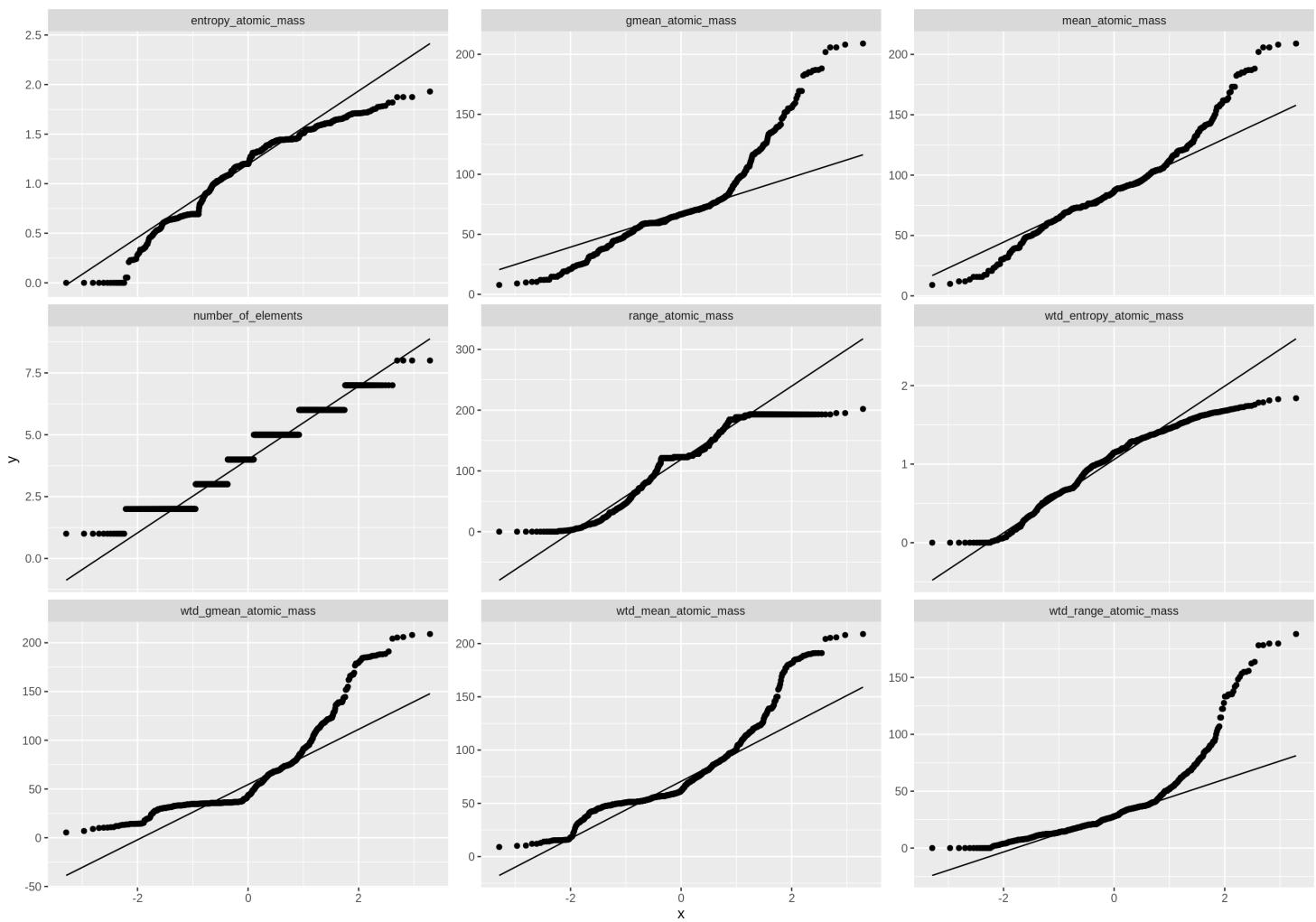


Bar Chart (with frequency)

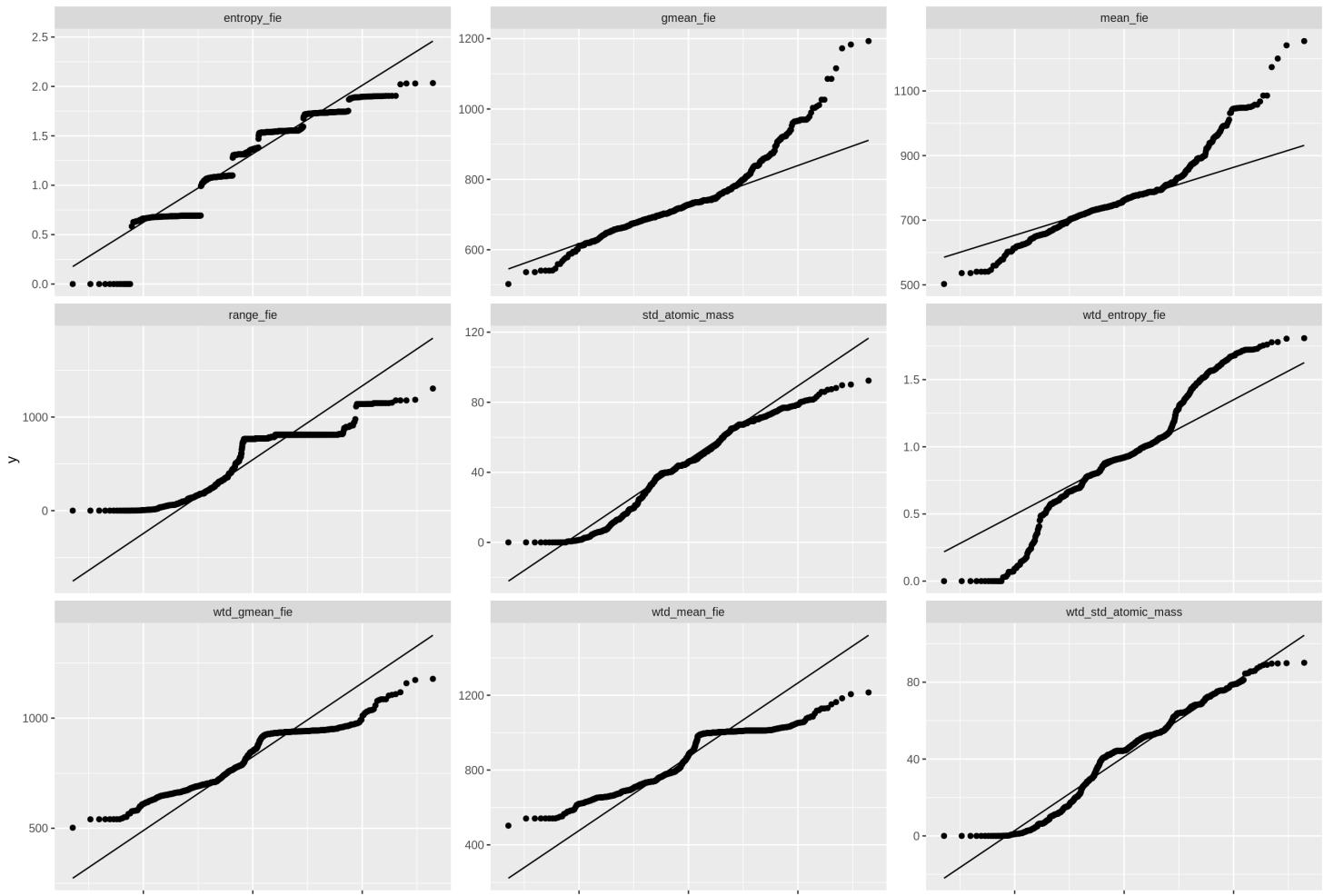
```
## 1 columns ignored with more than 50 categories.
## material: 15542 categories
```

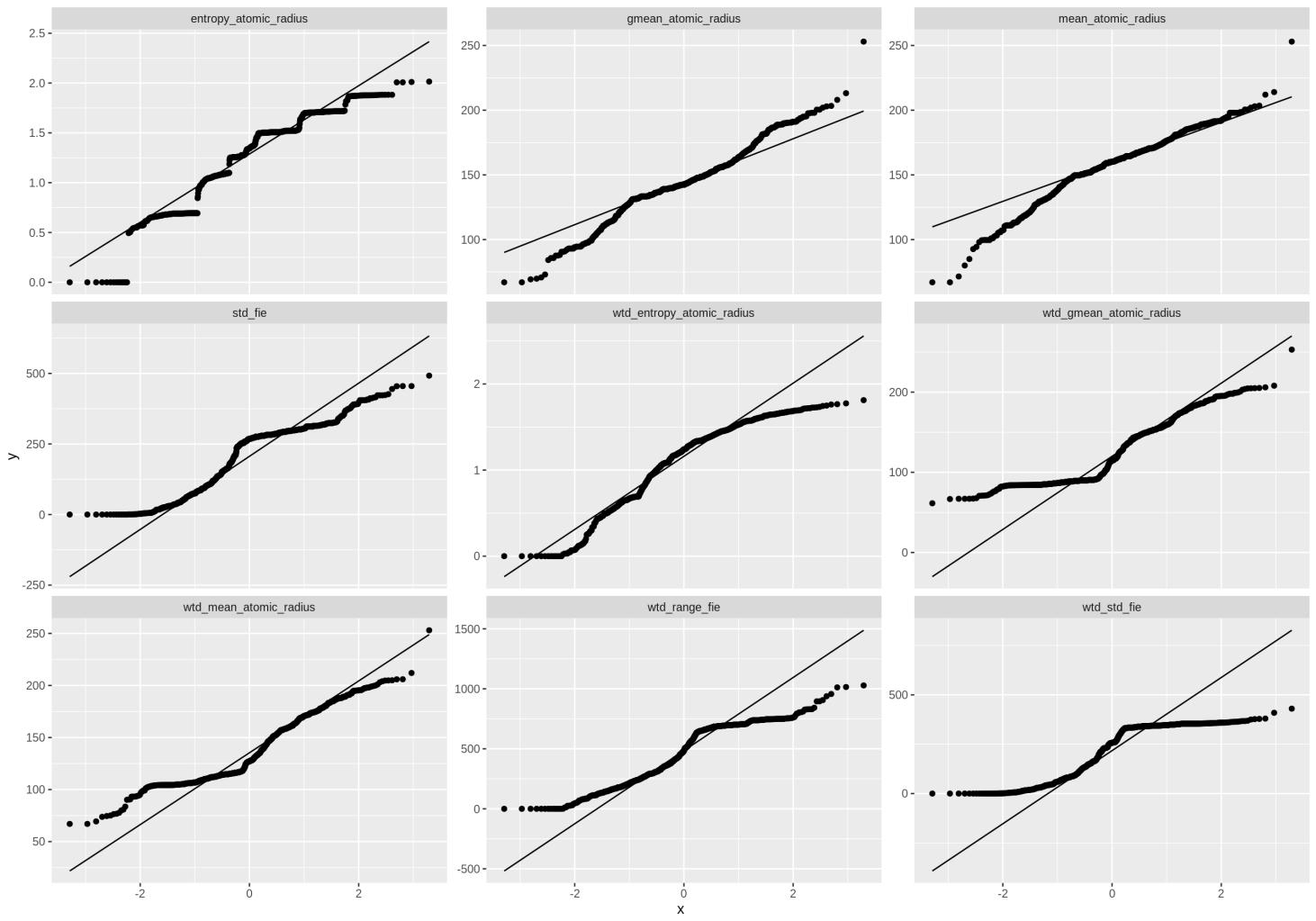


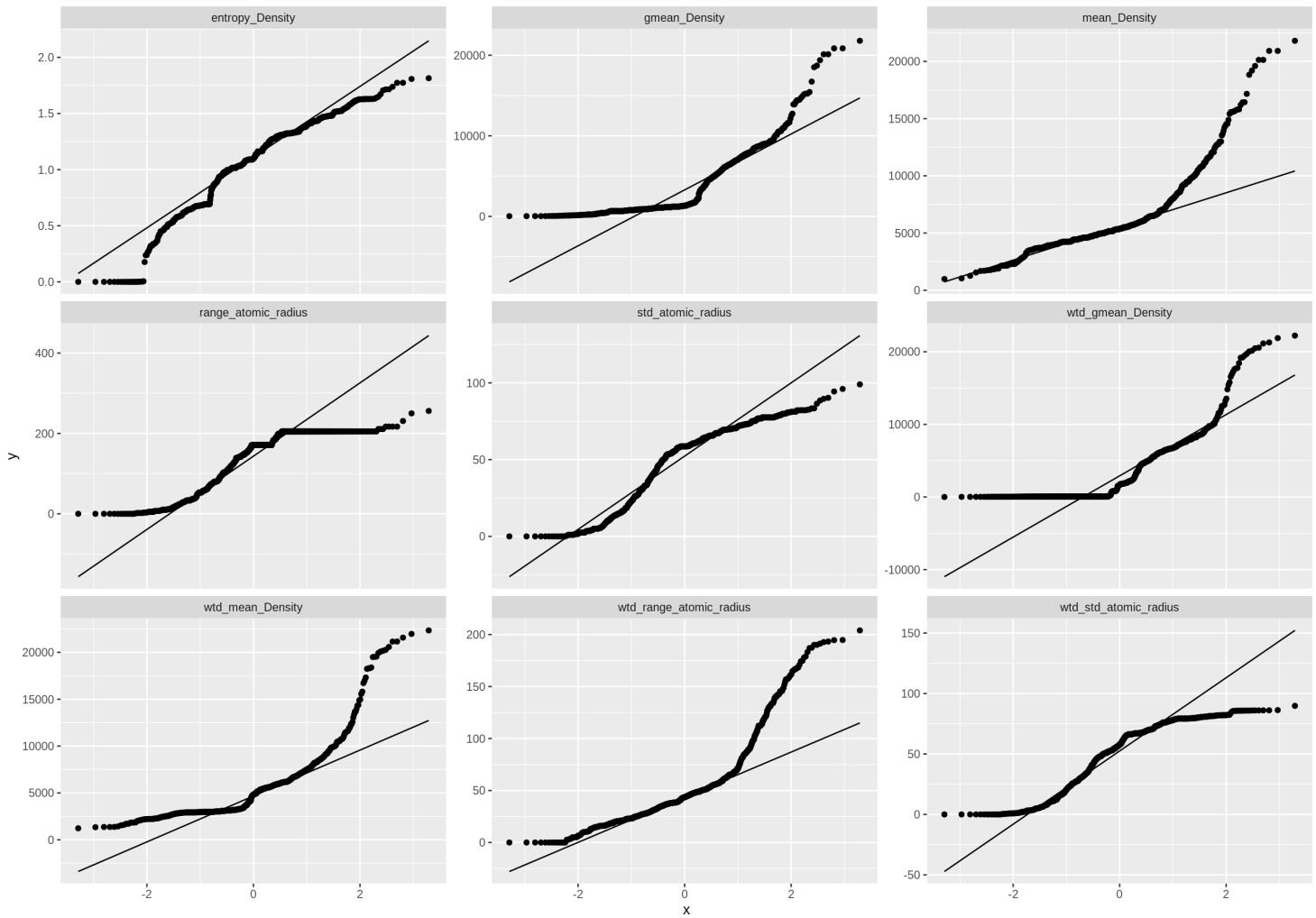
QQ Plot

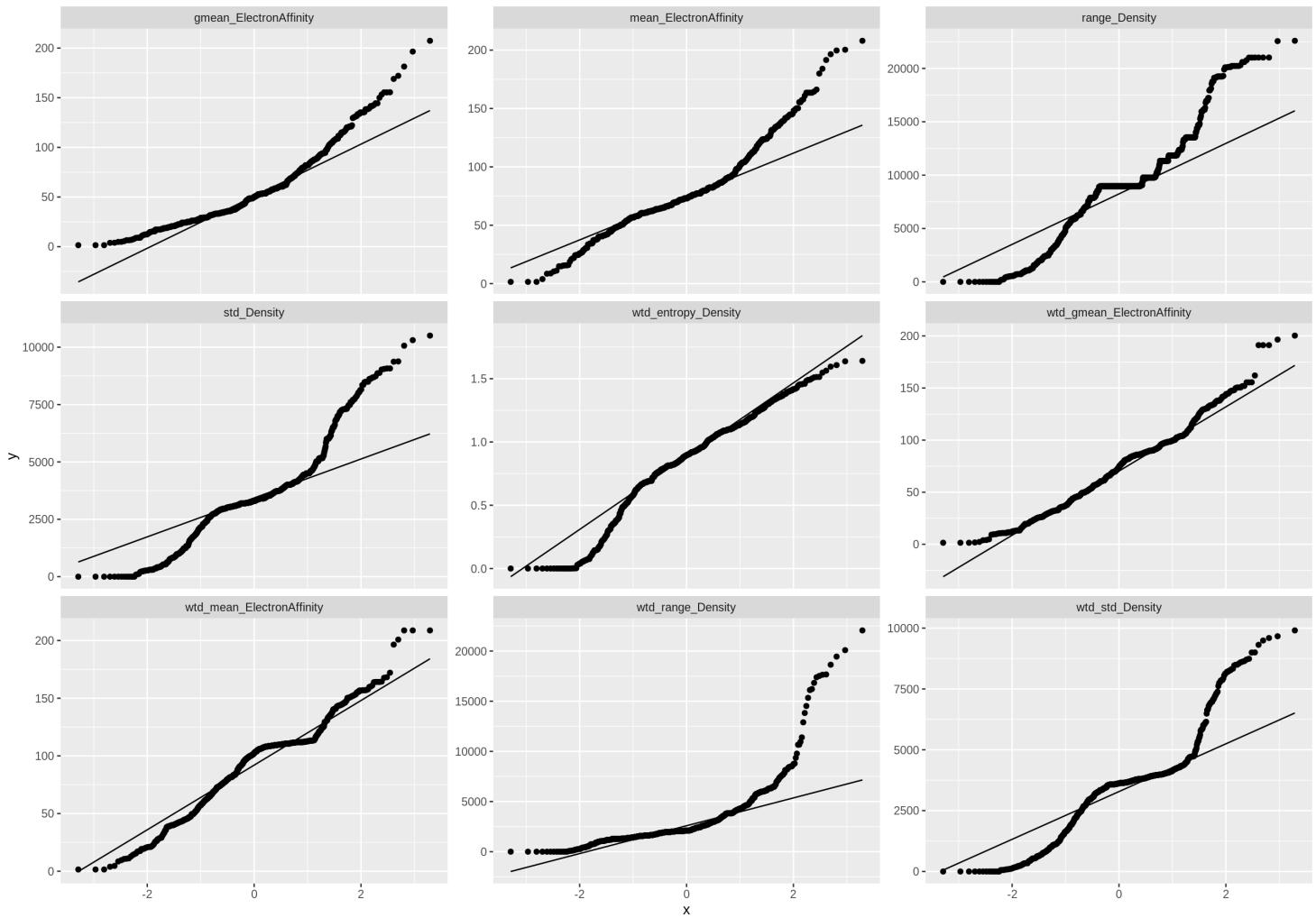


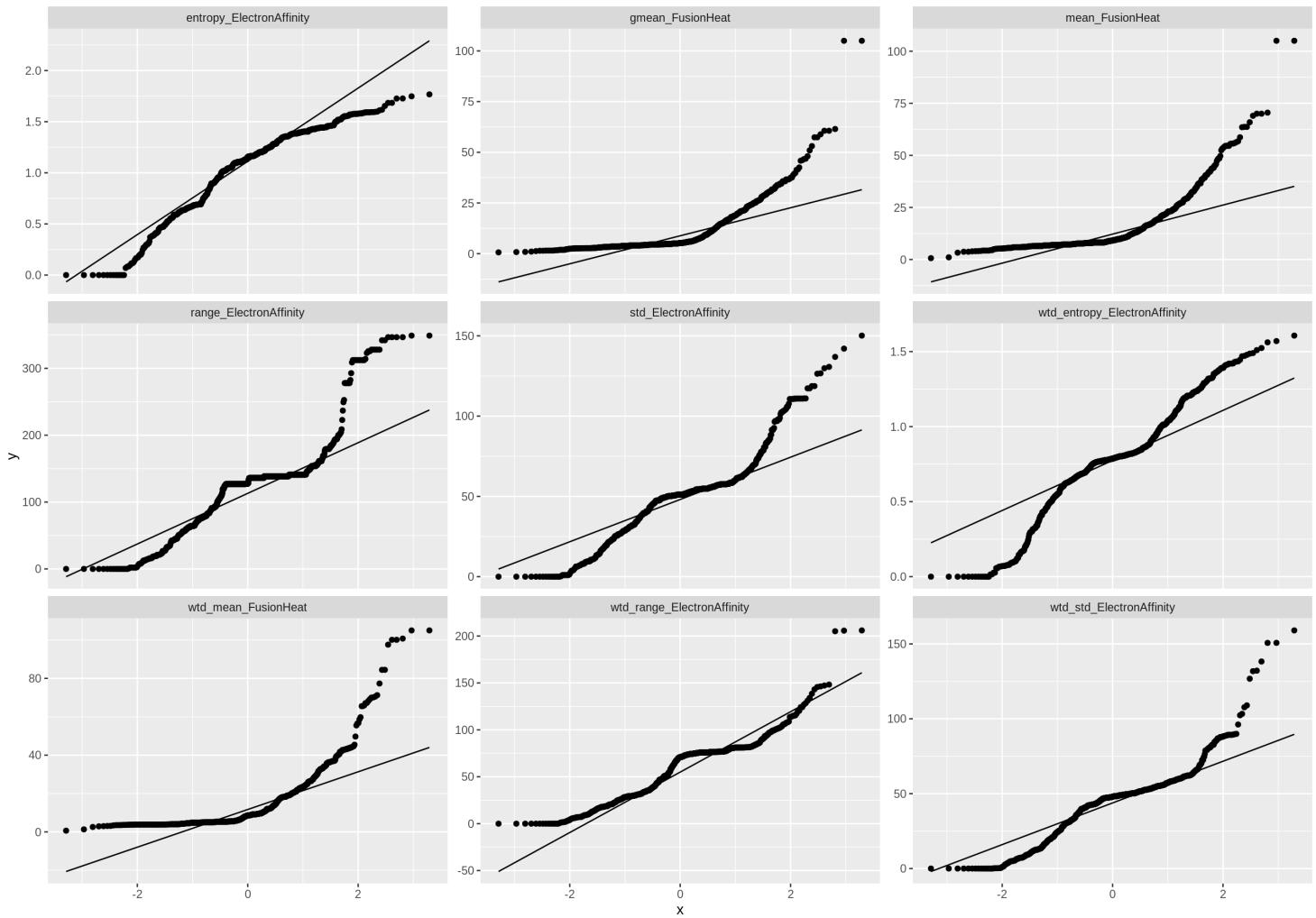
Page 1

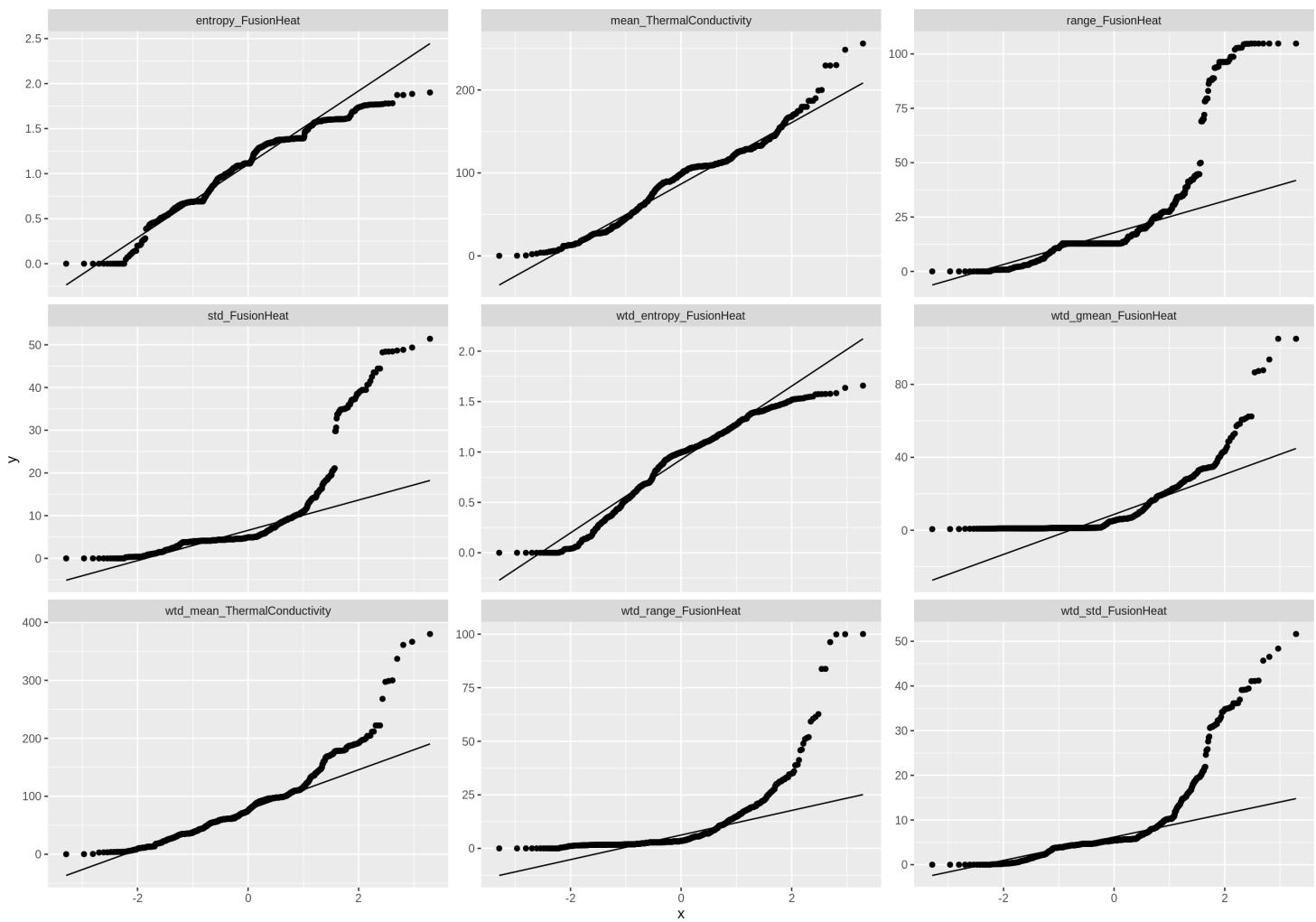


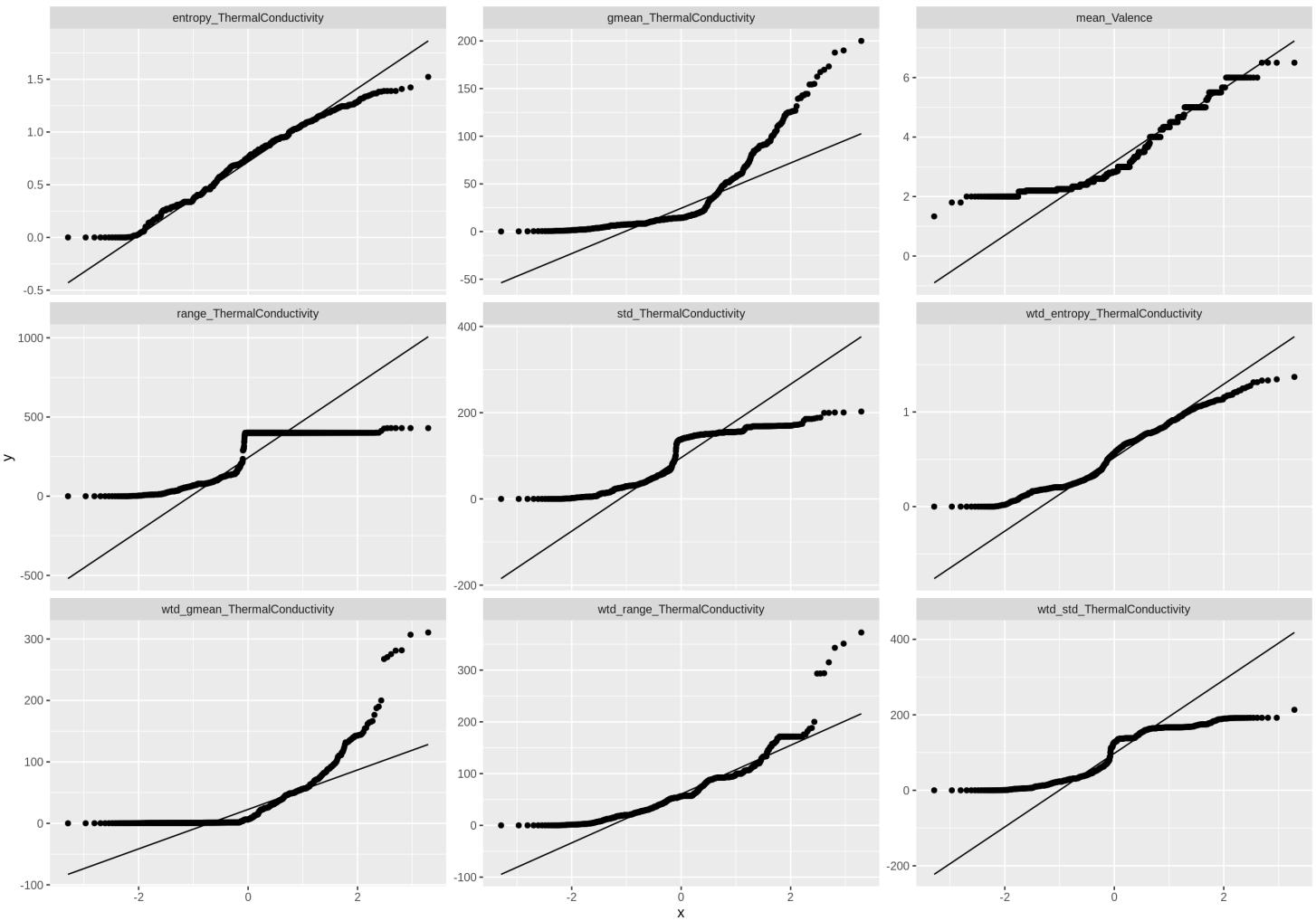


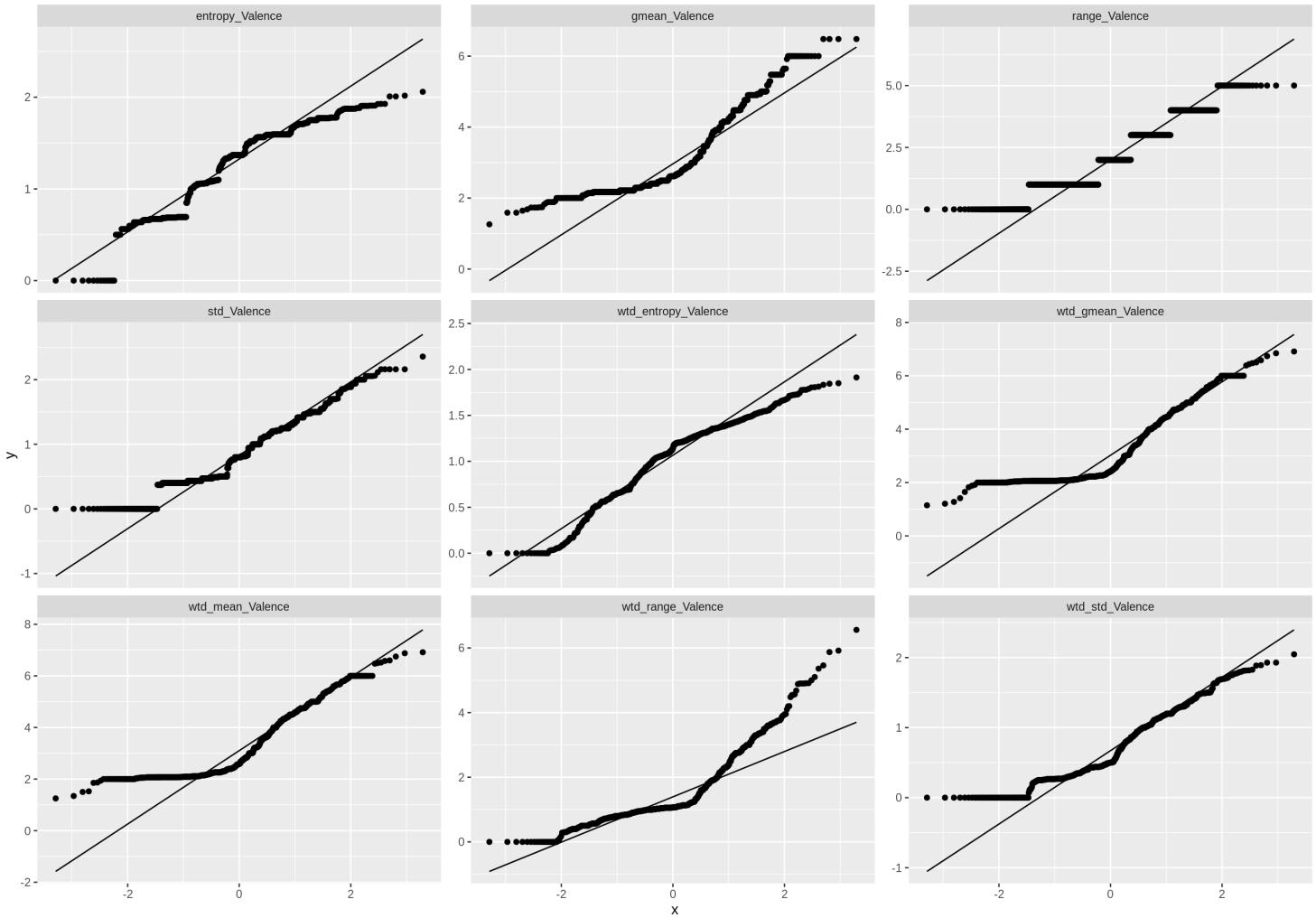


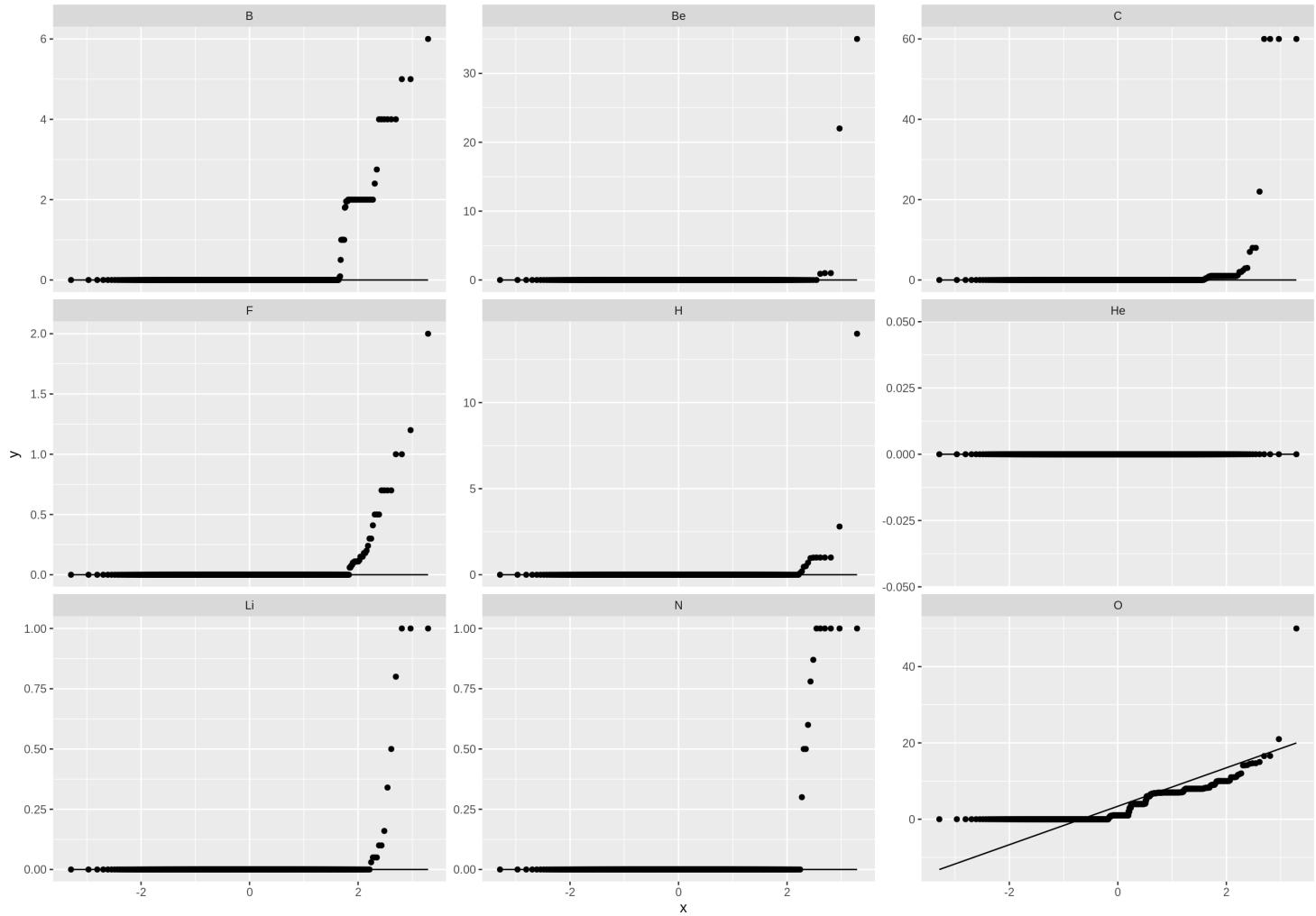


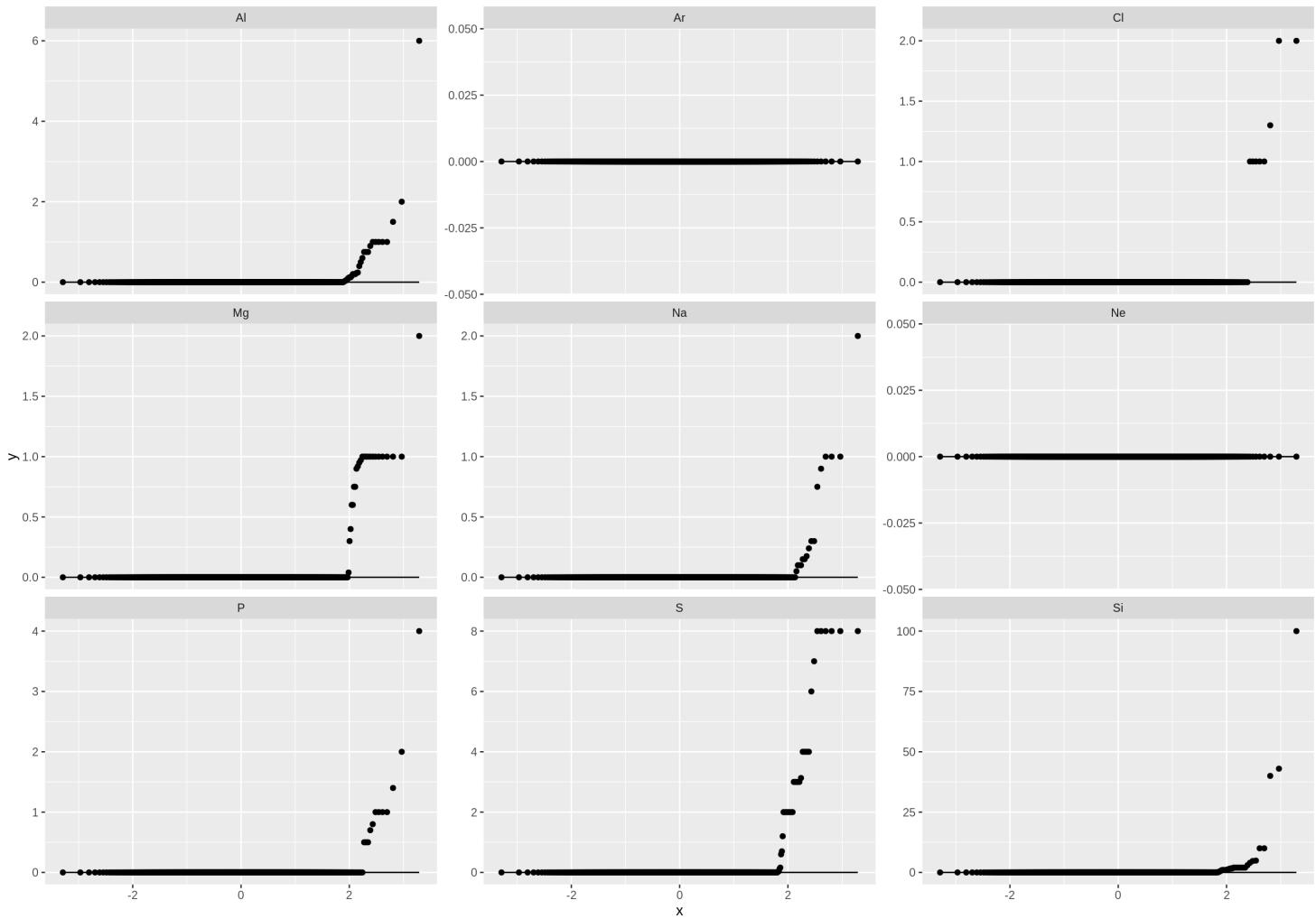


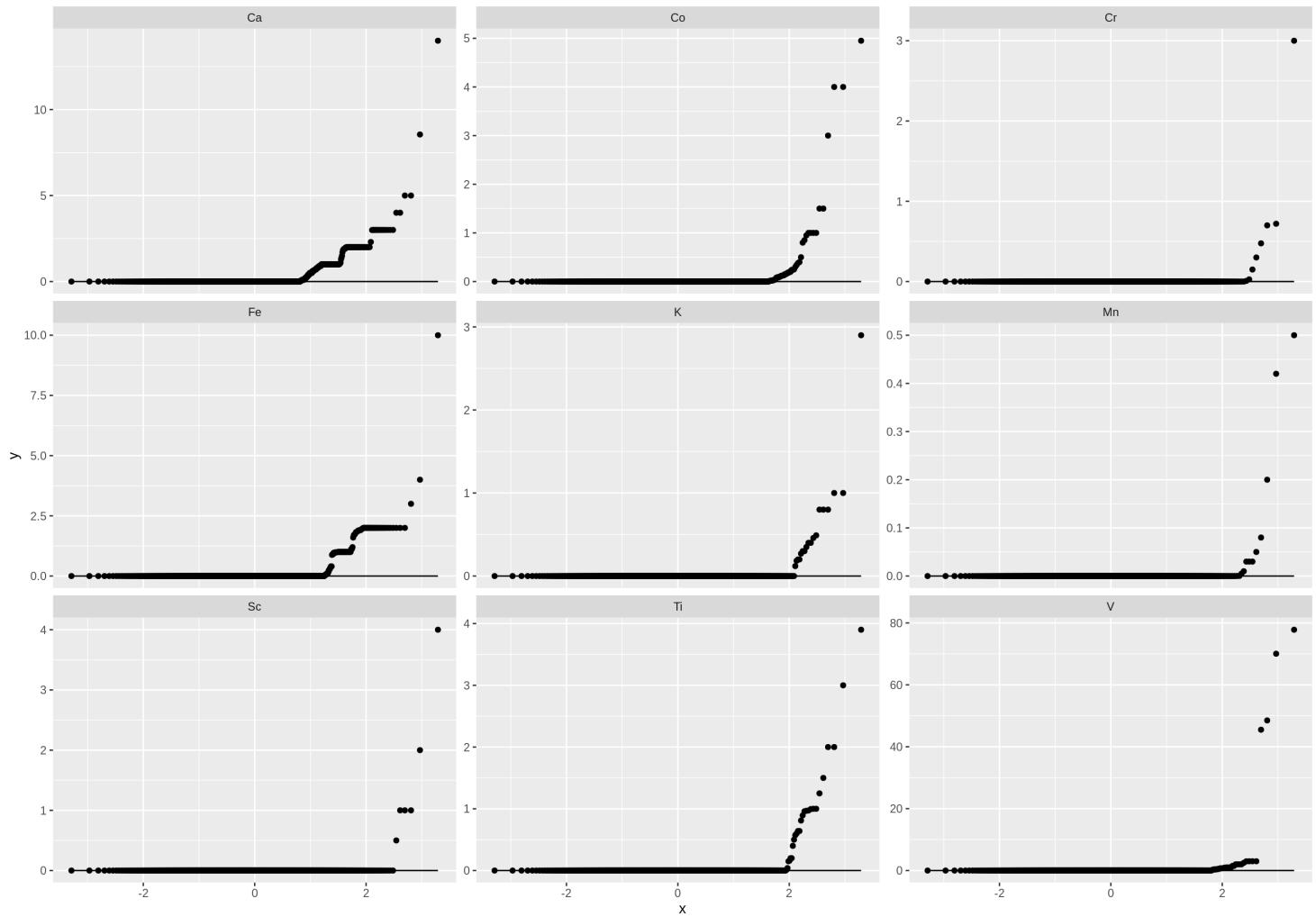


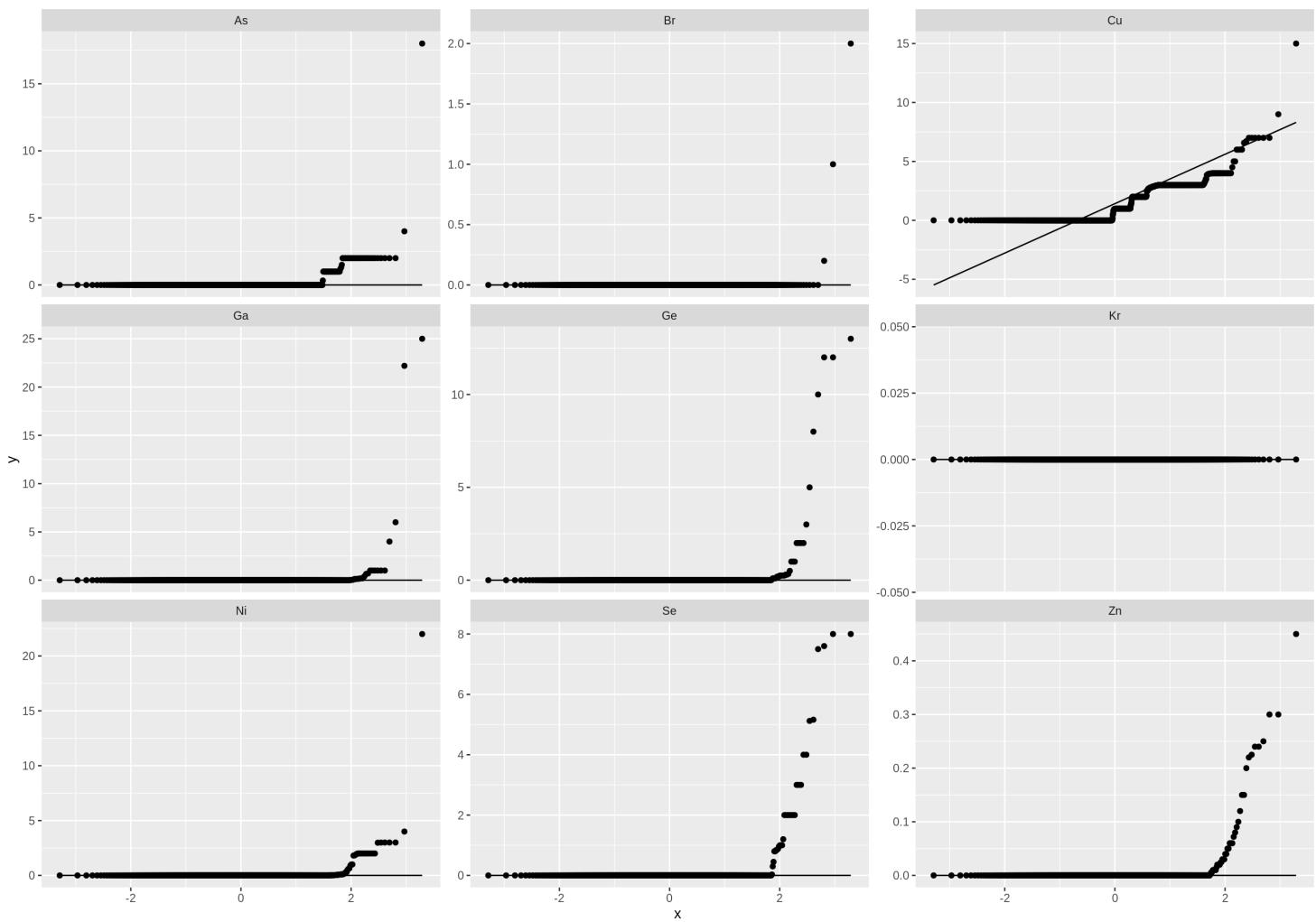


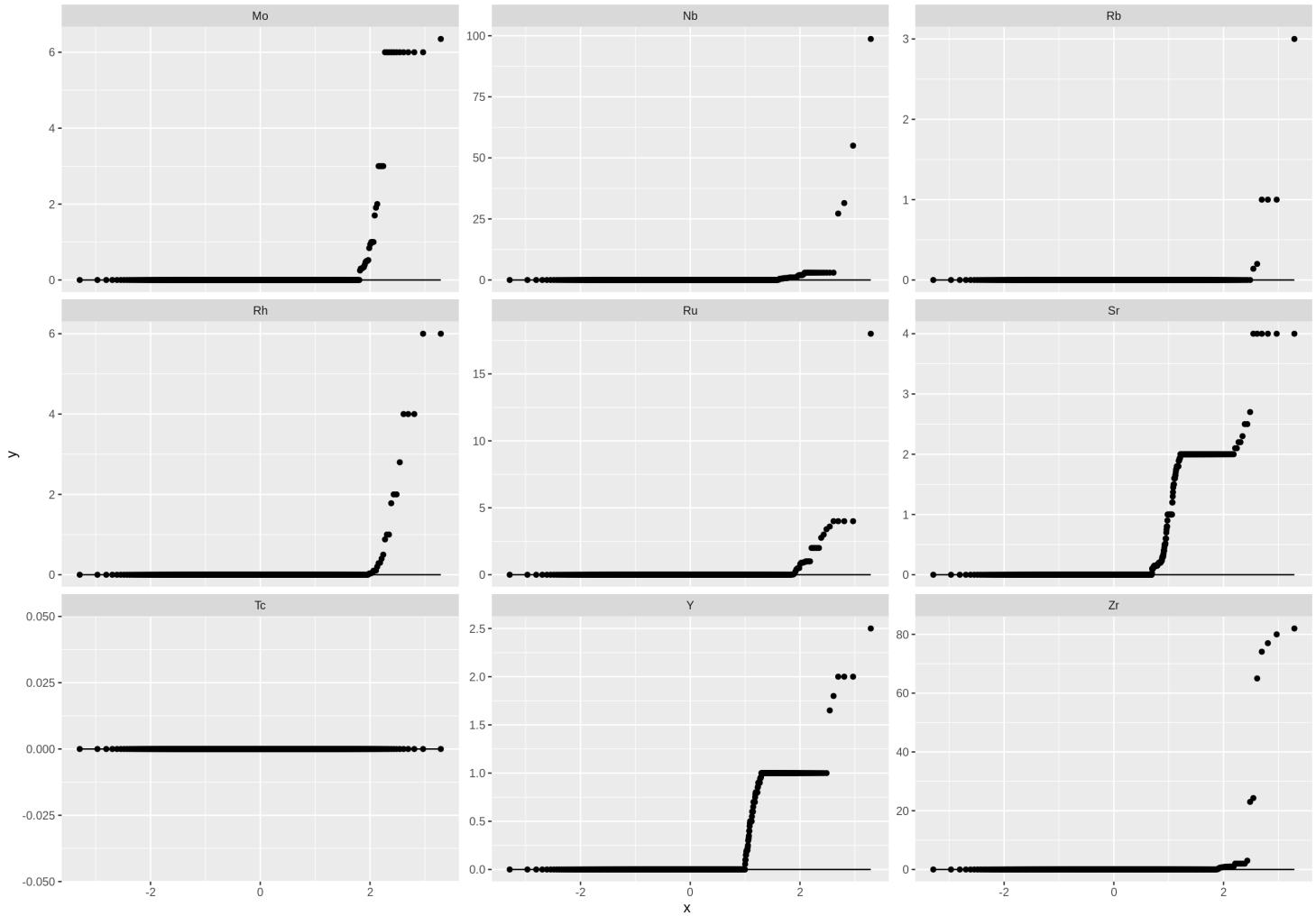


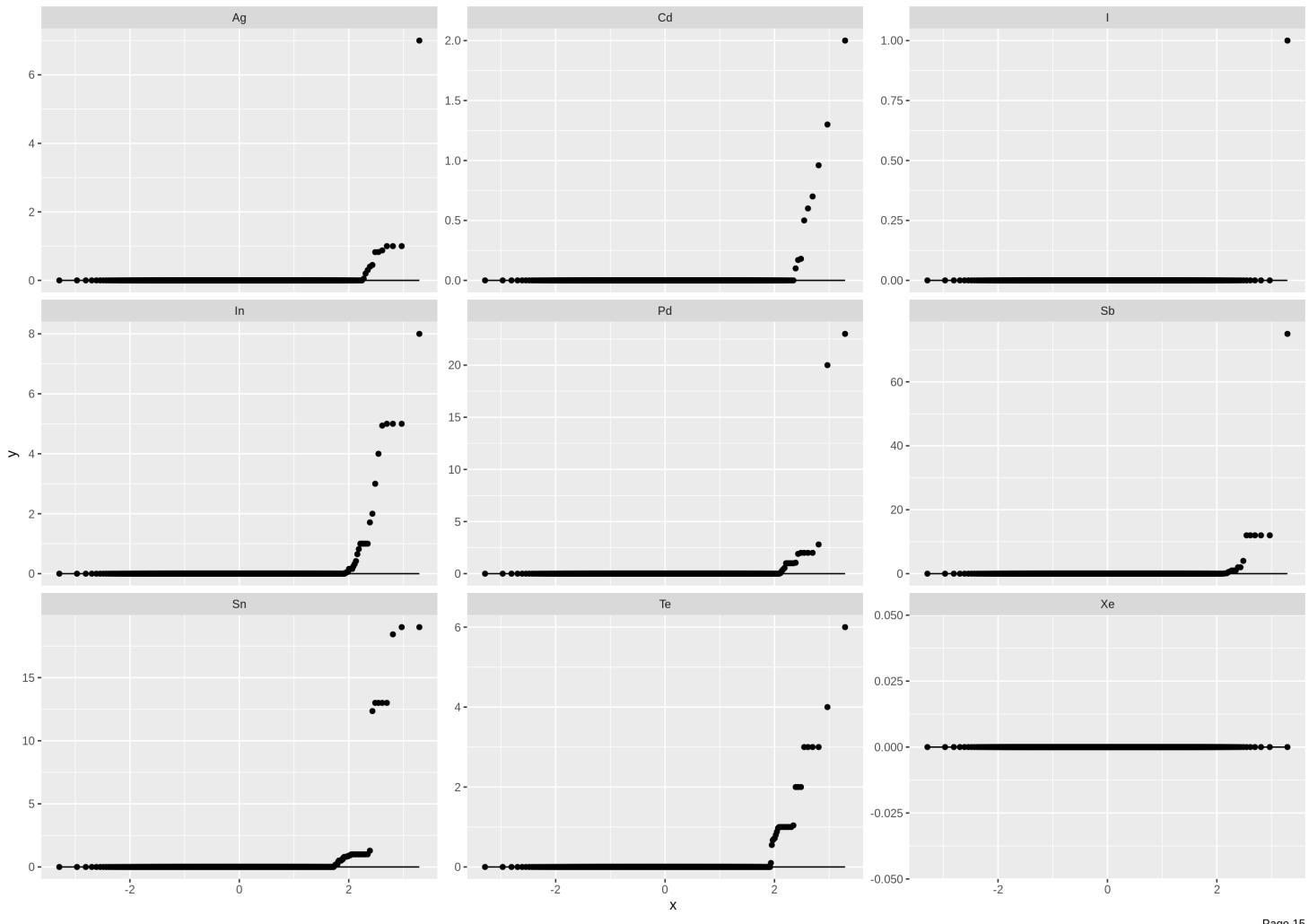


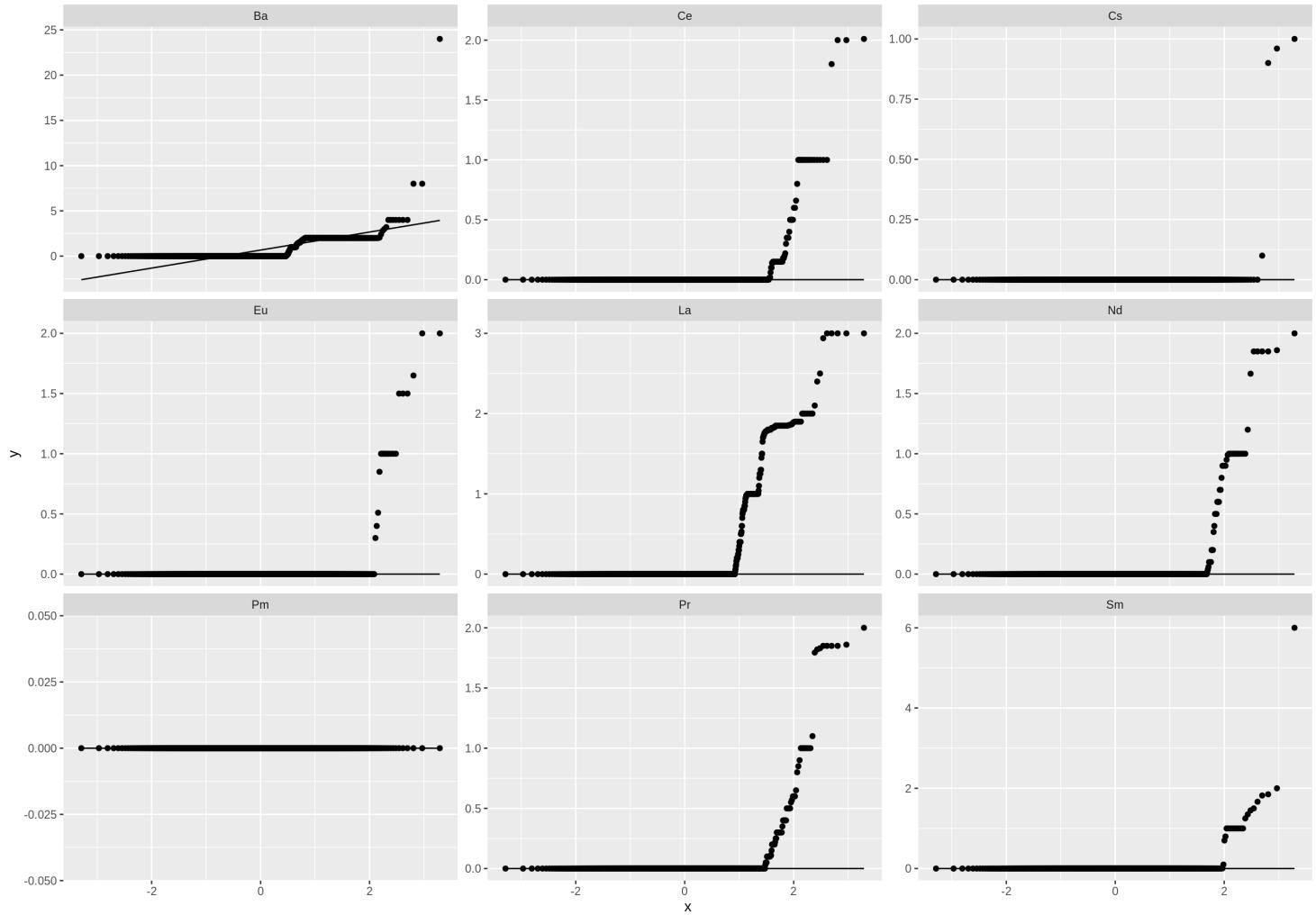


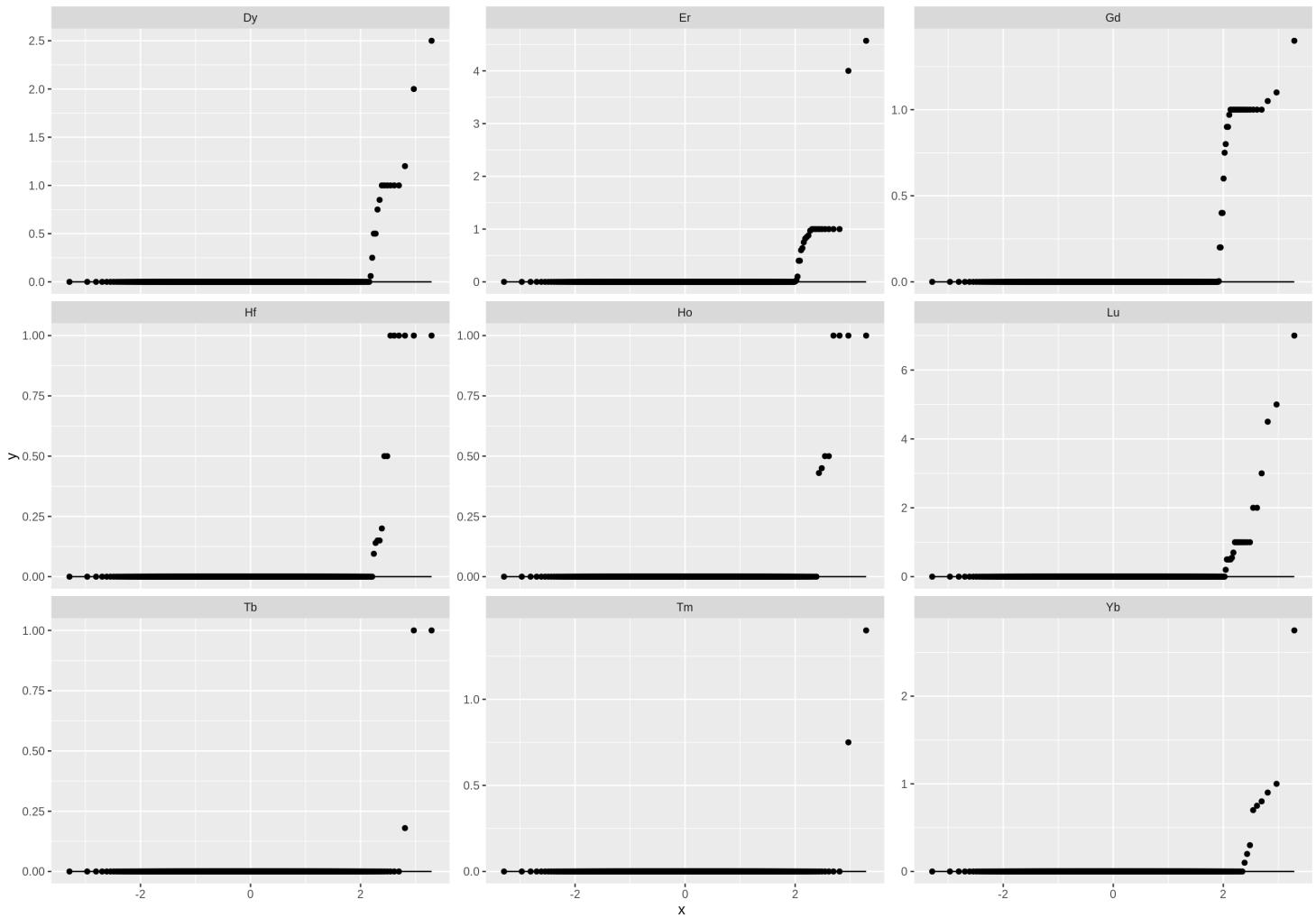


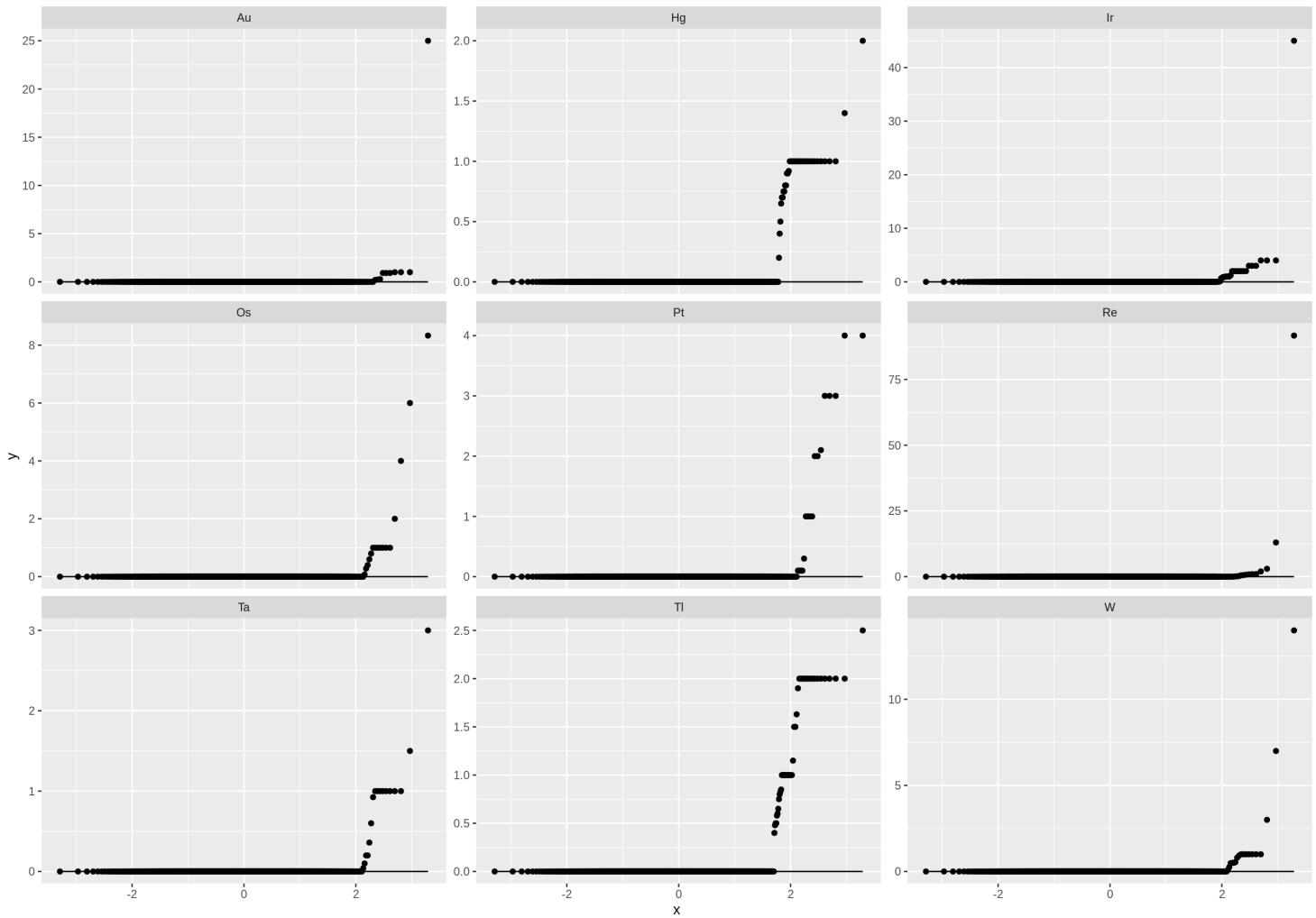


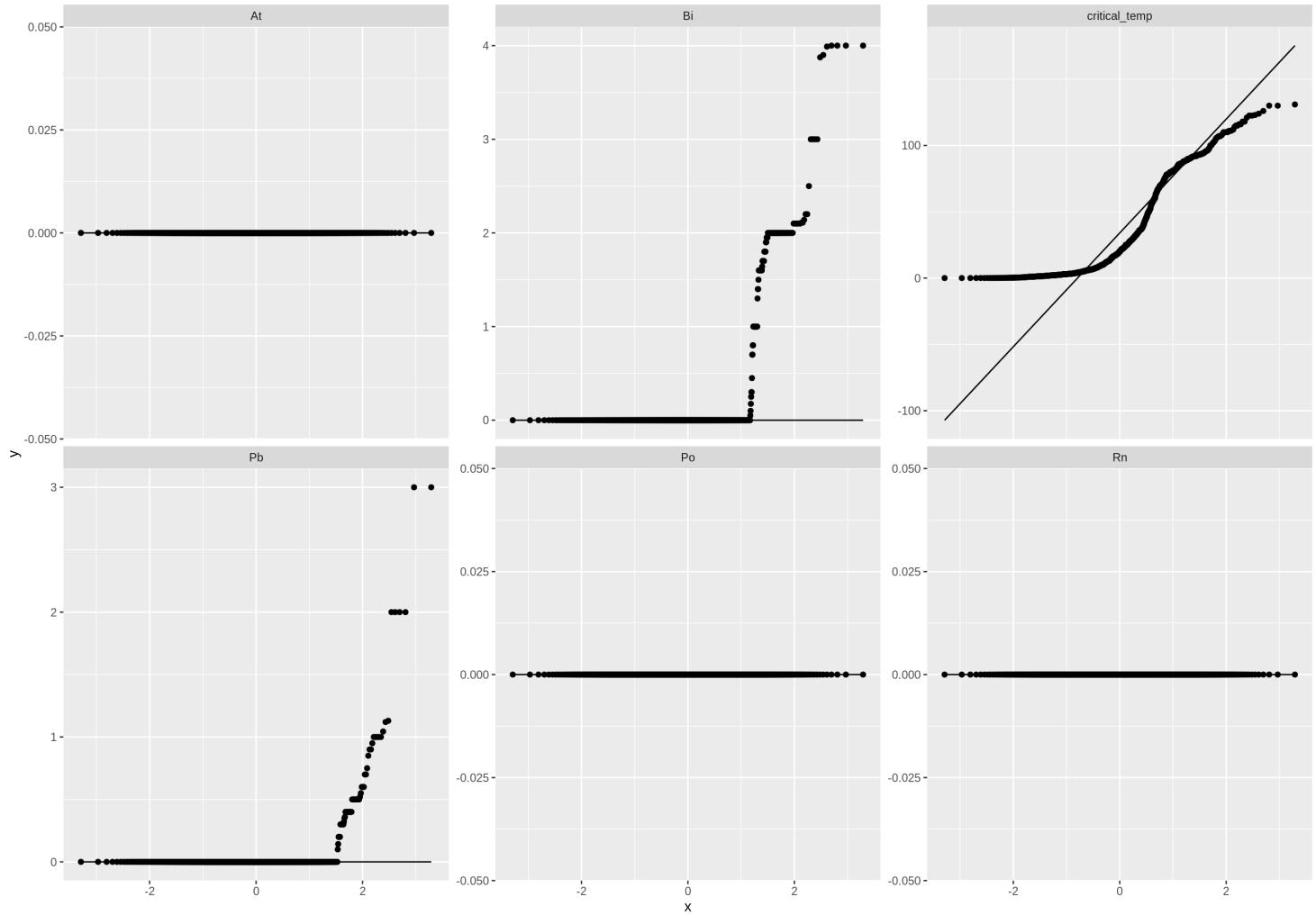




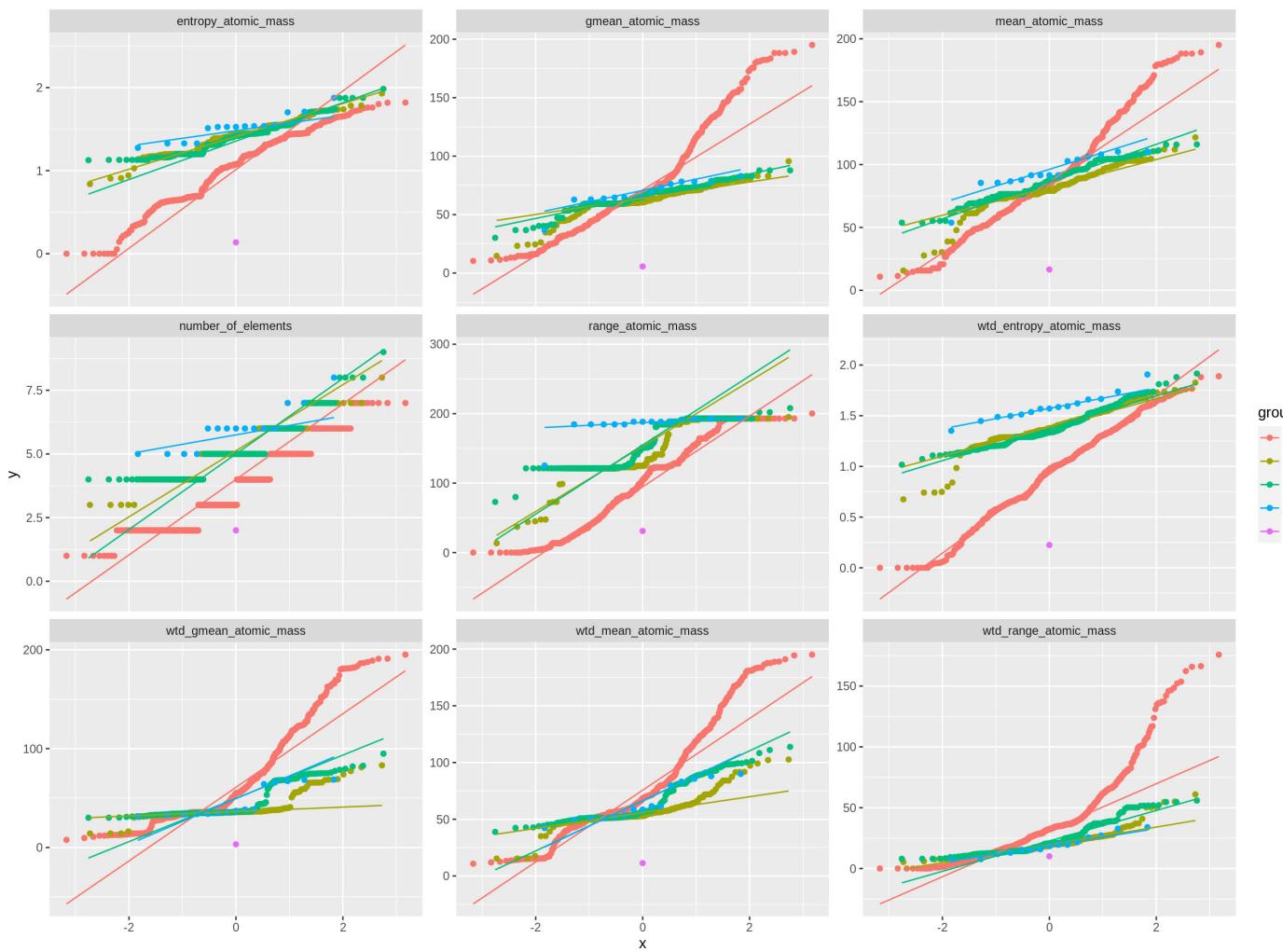




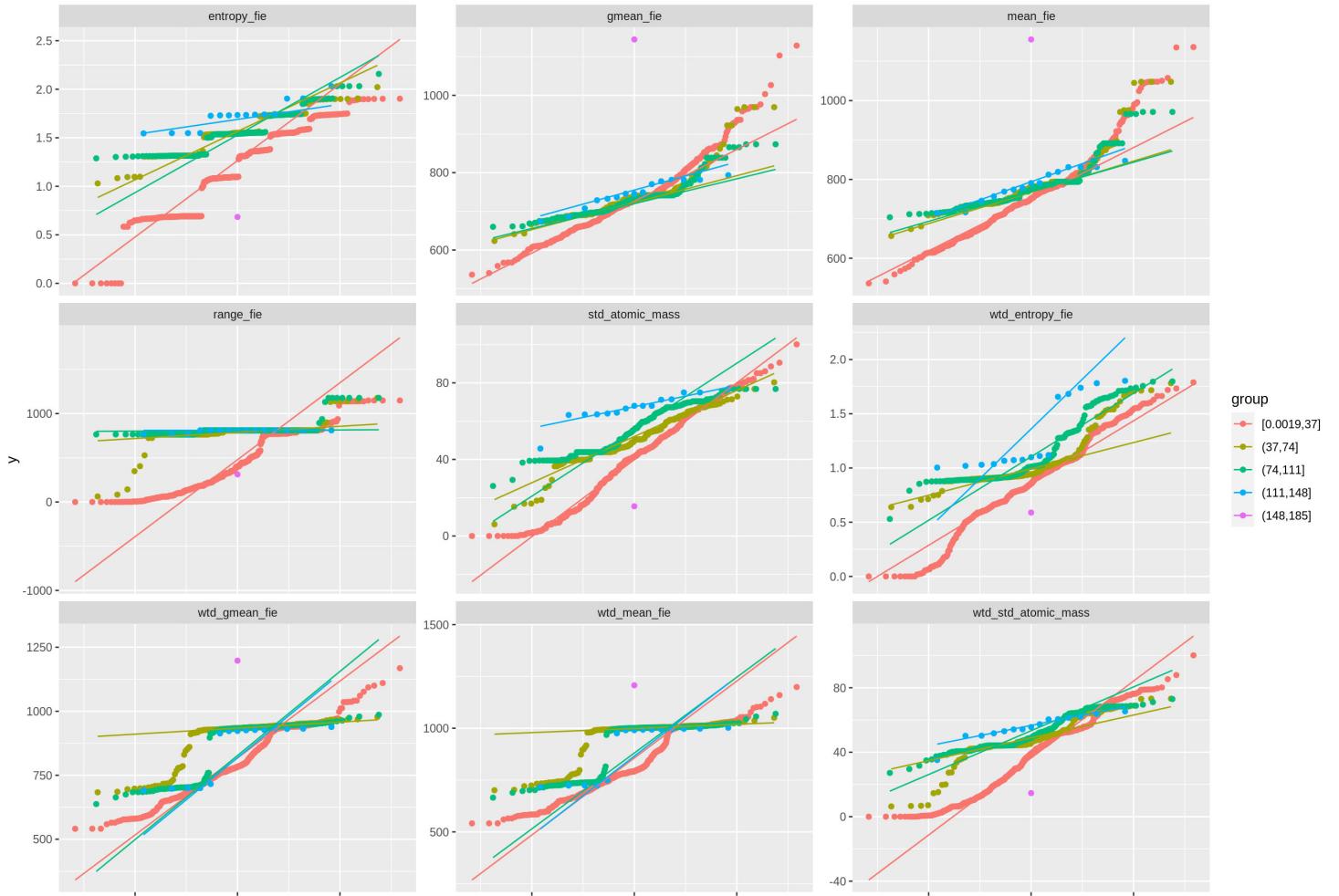


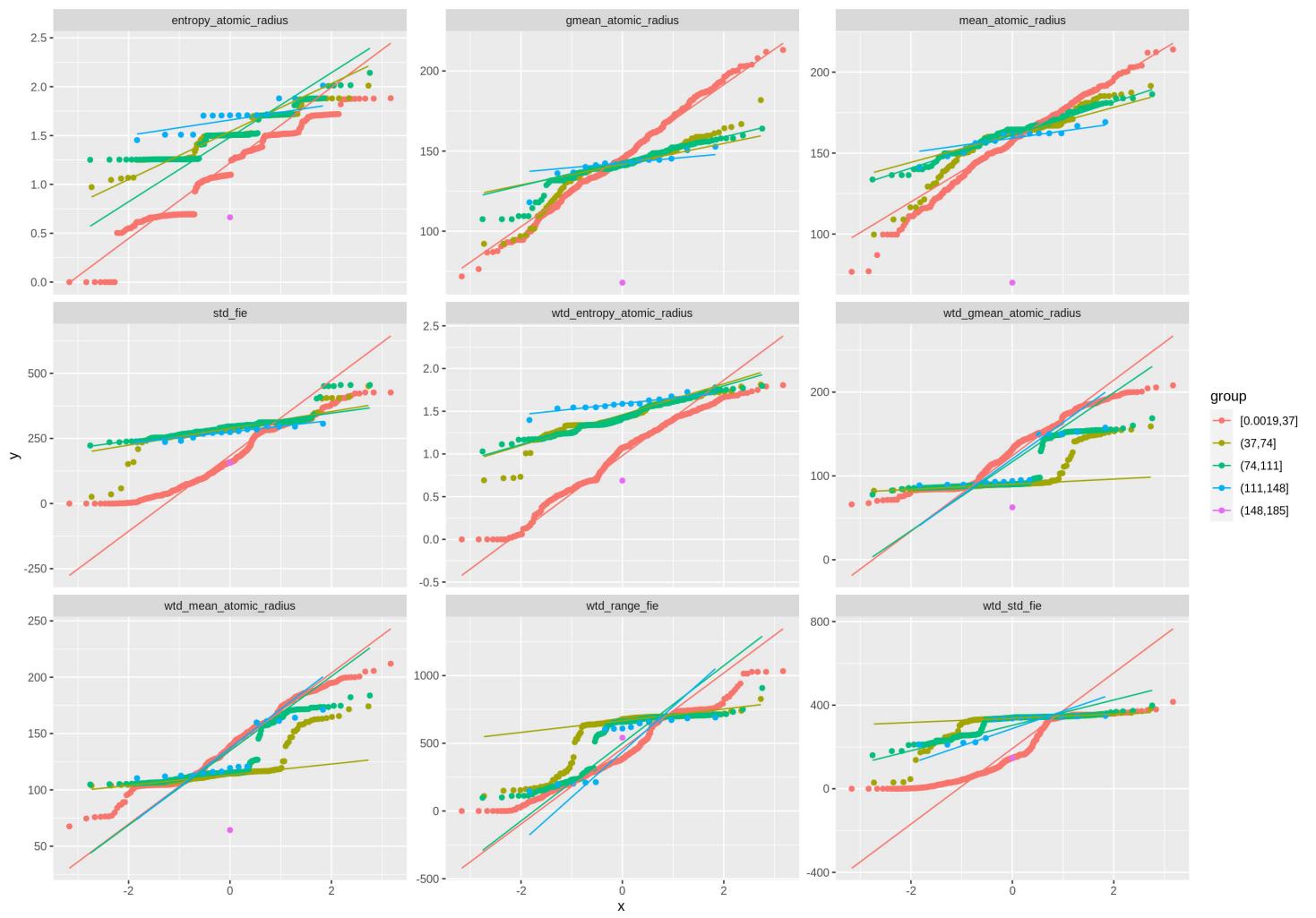


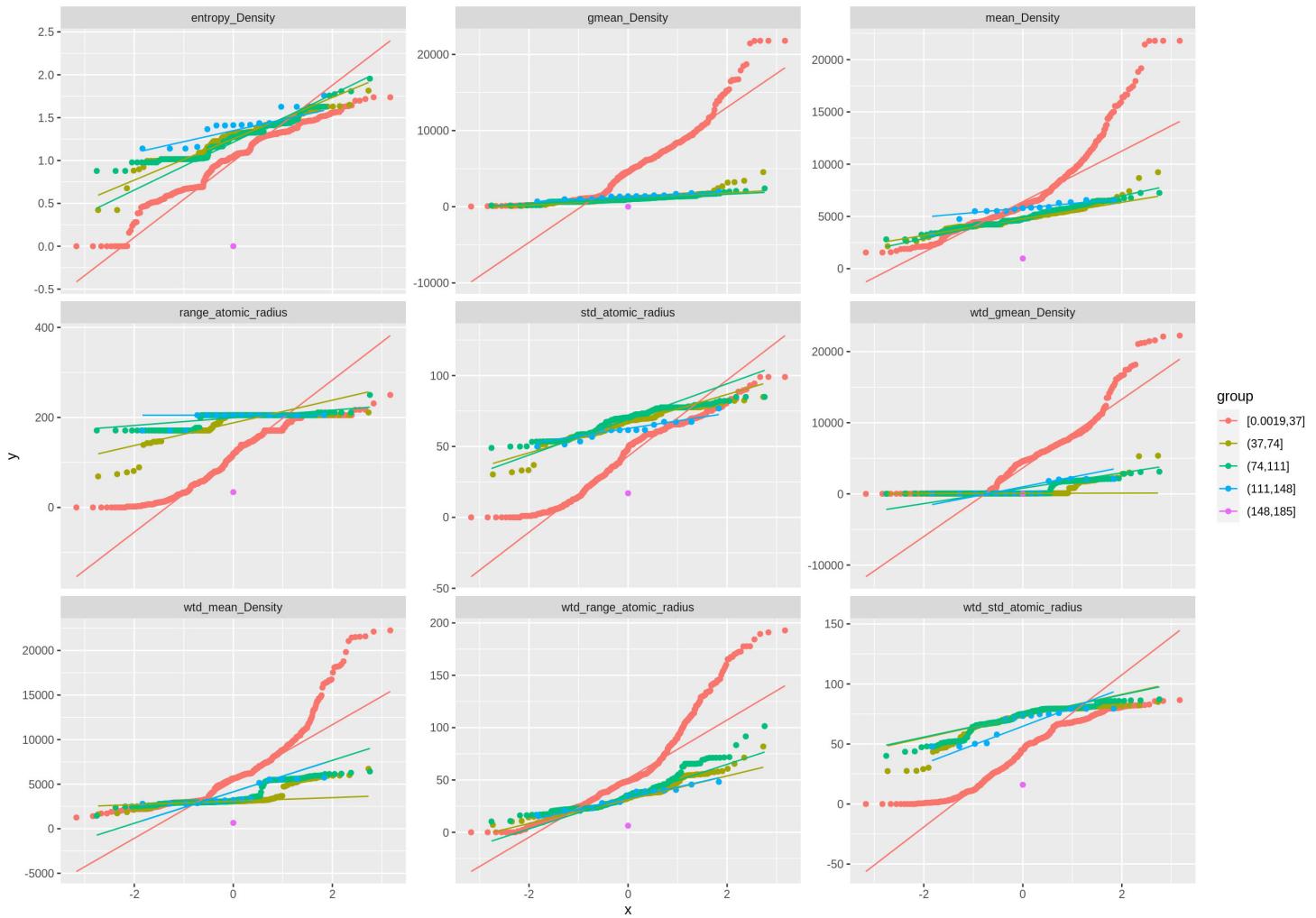
QQ Plot (by critical_temp)

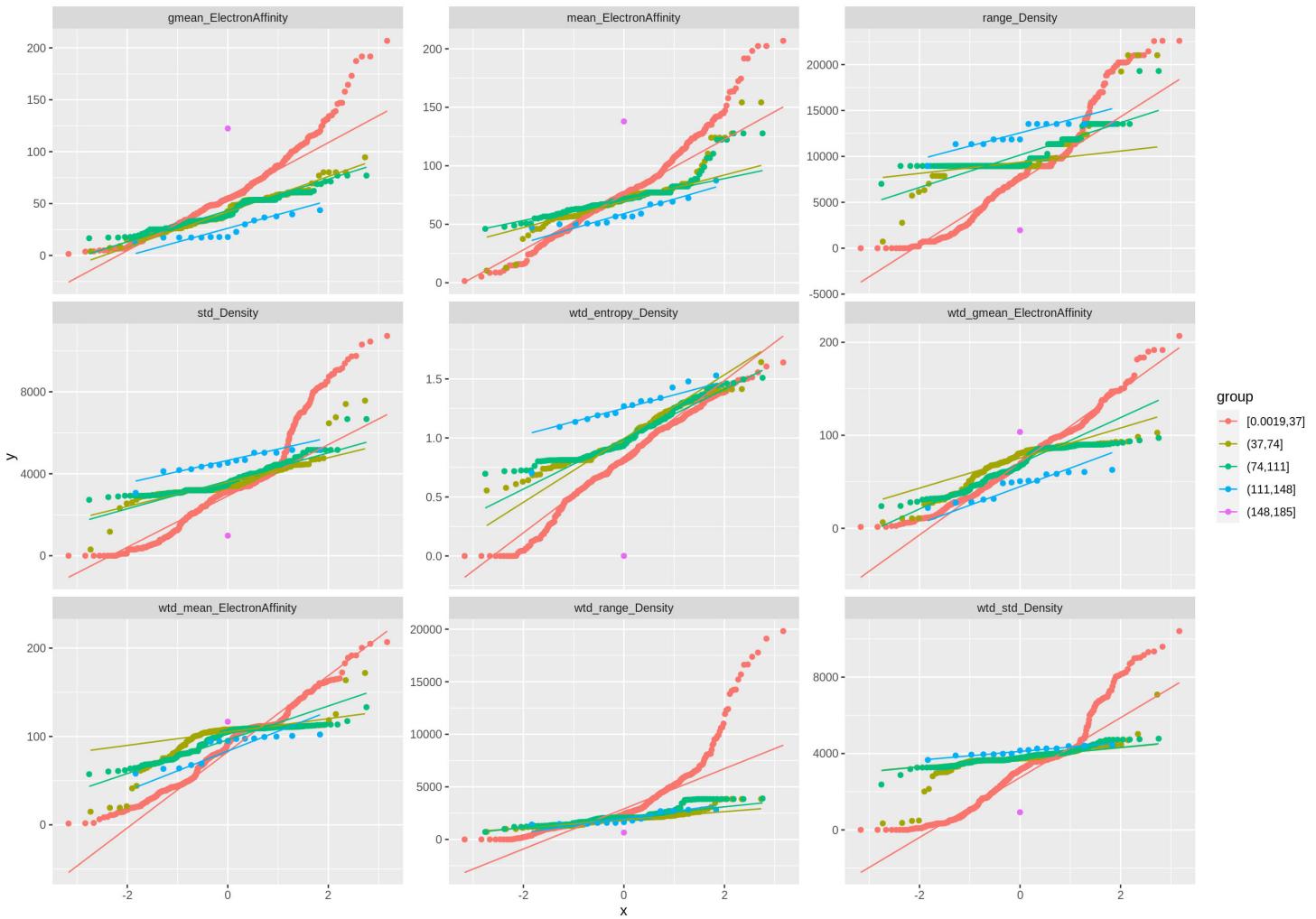


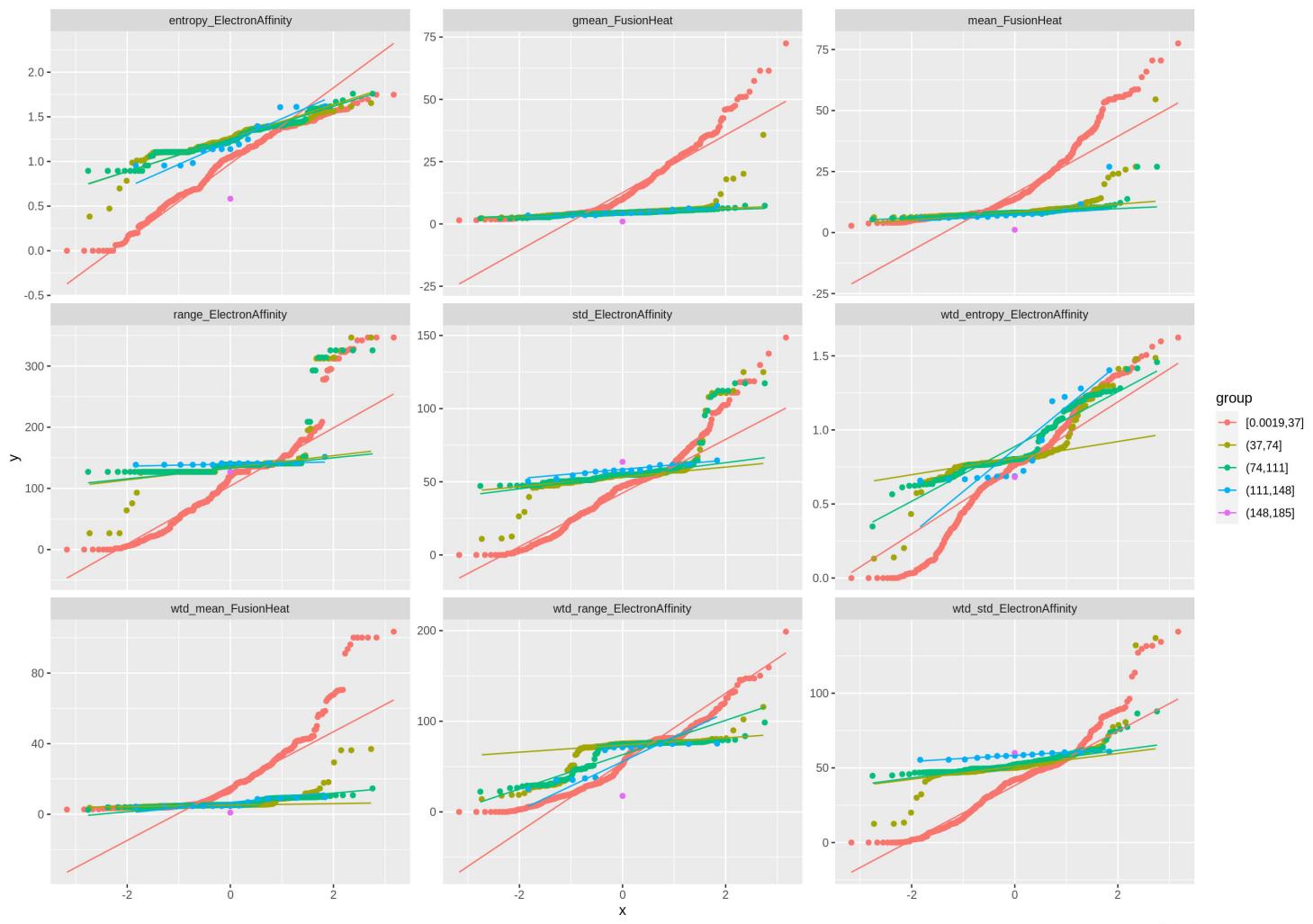
Page 1

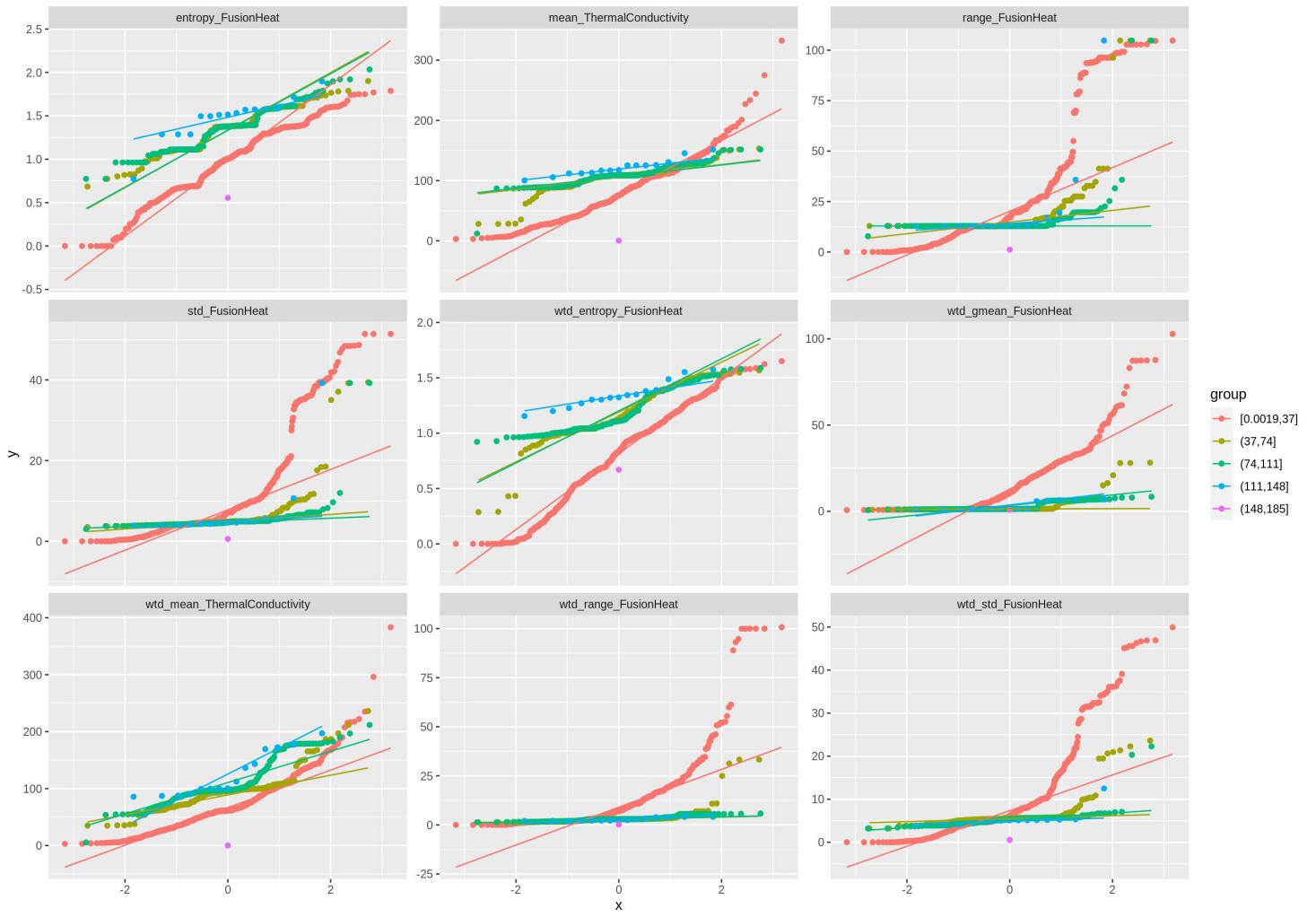


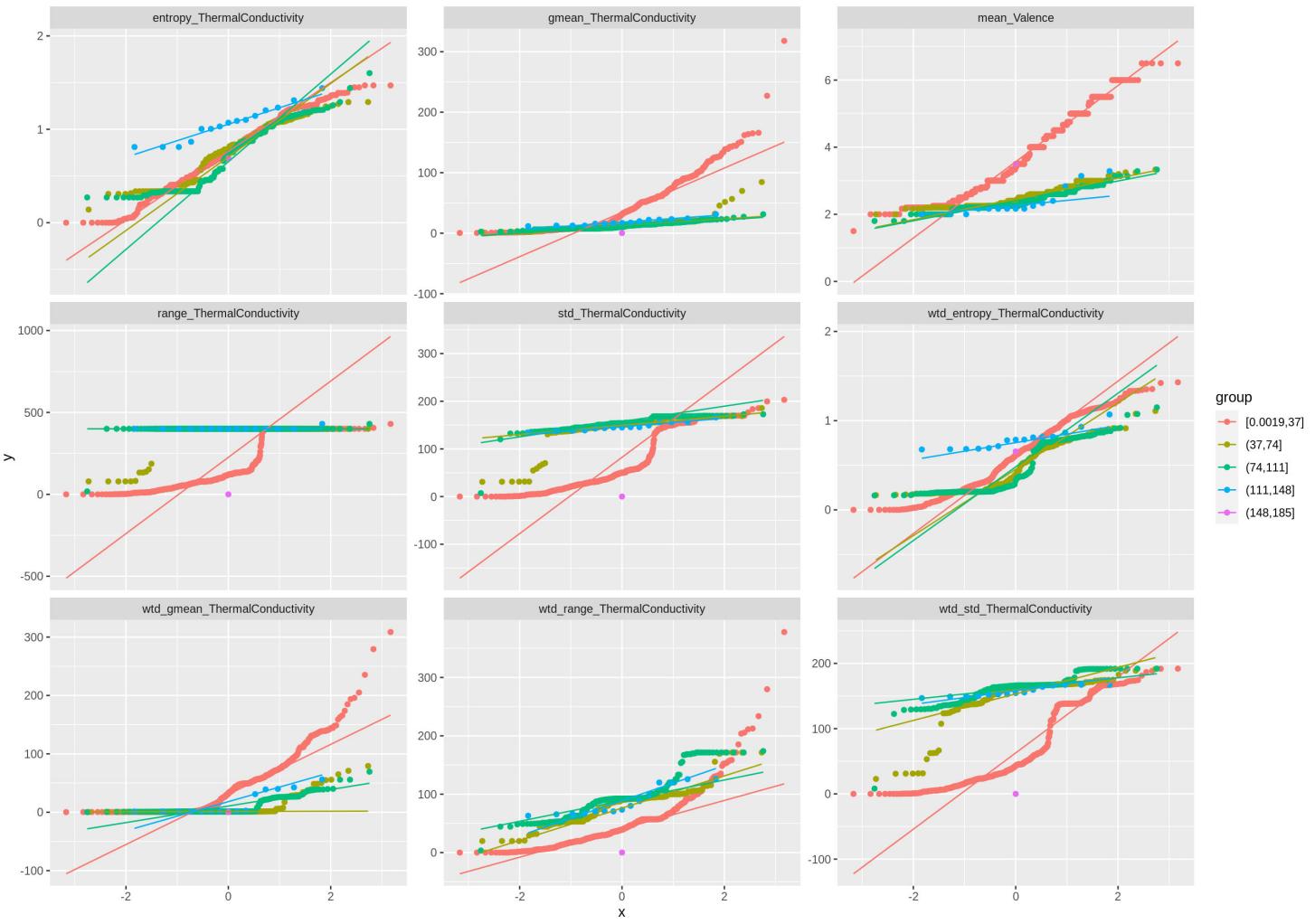


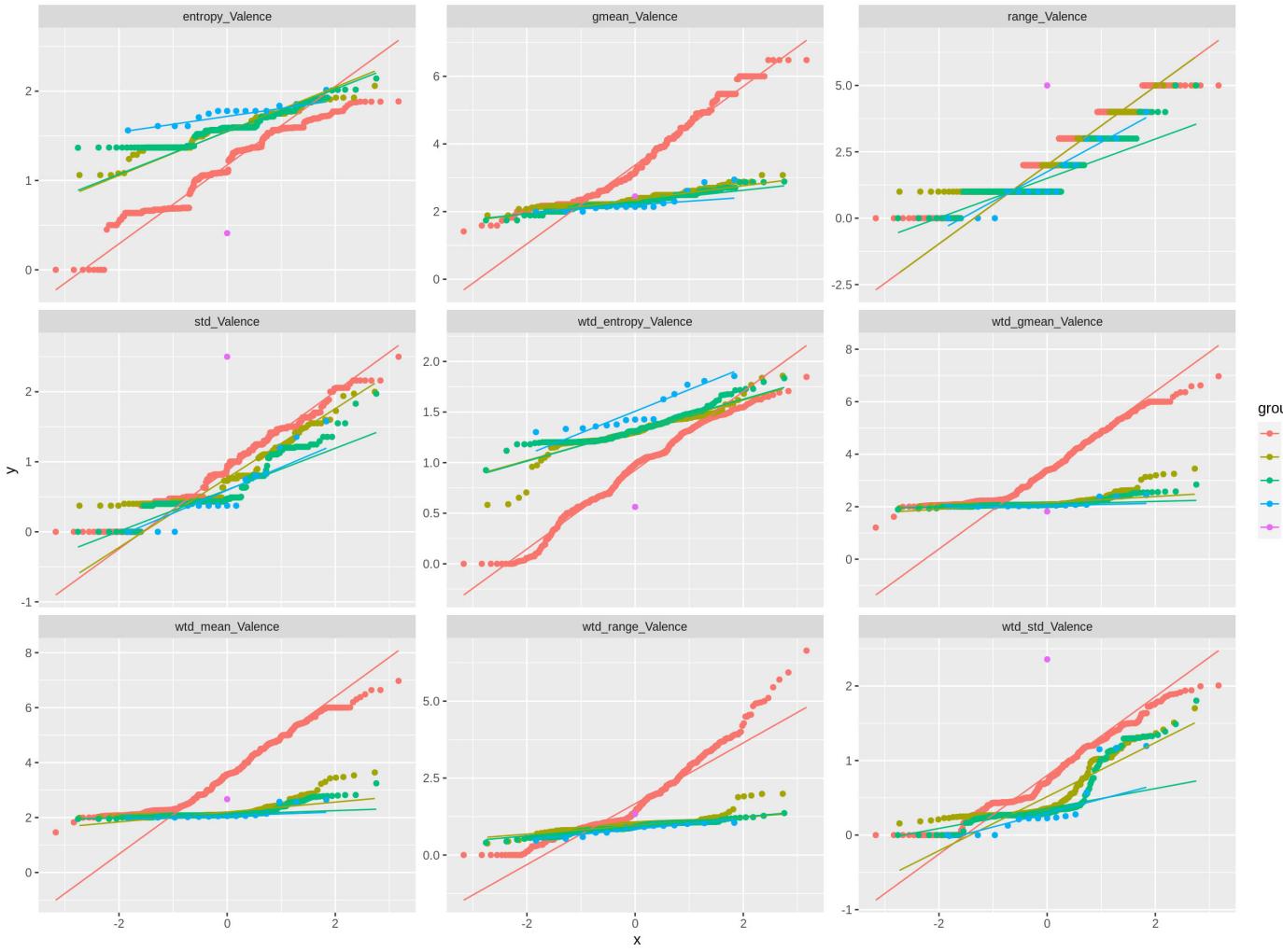


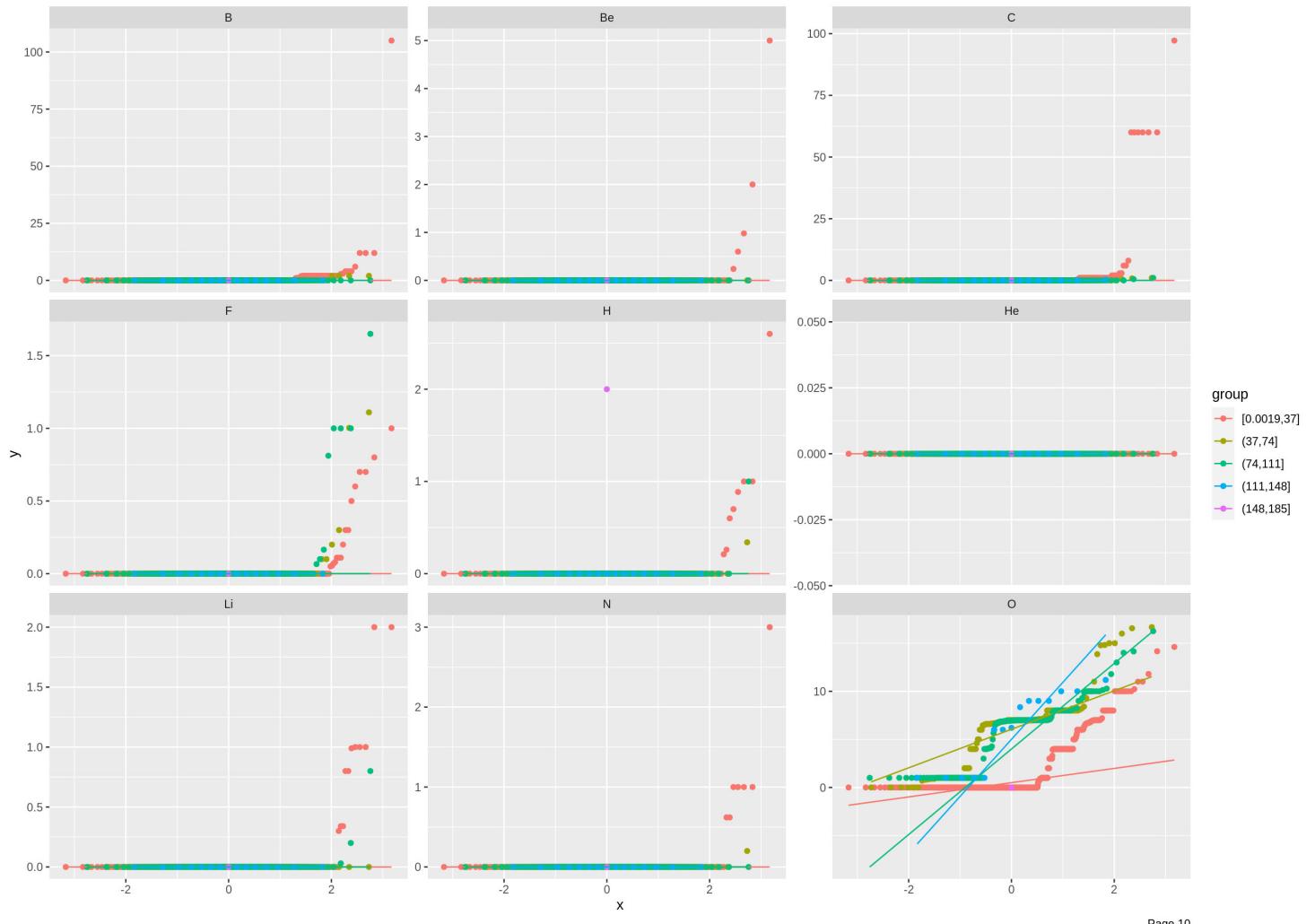


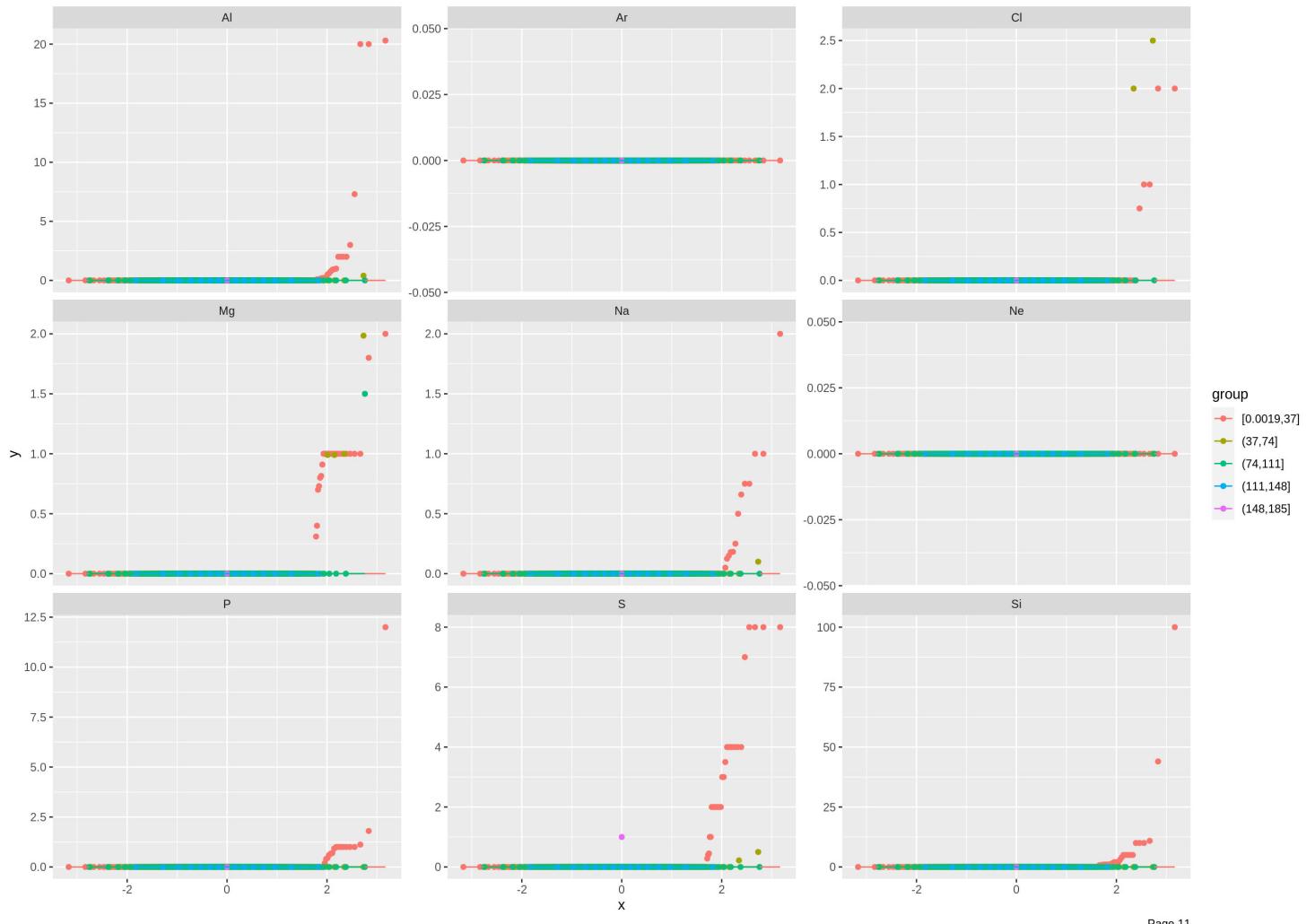


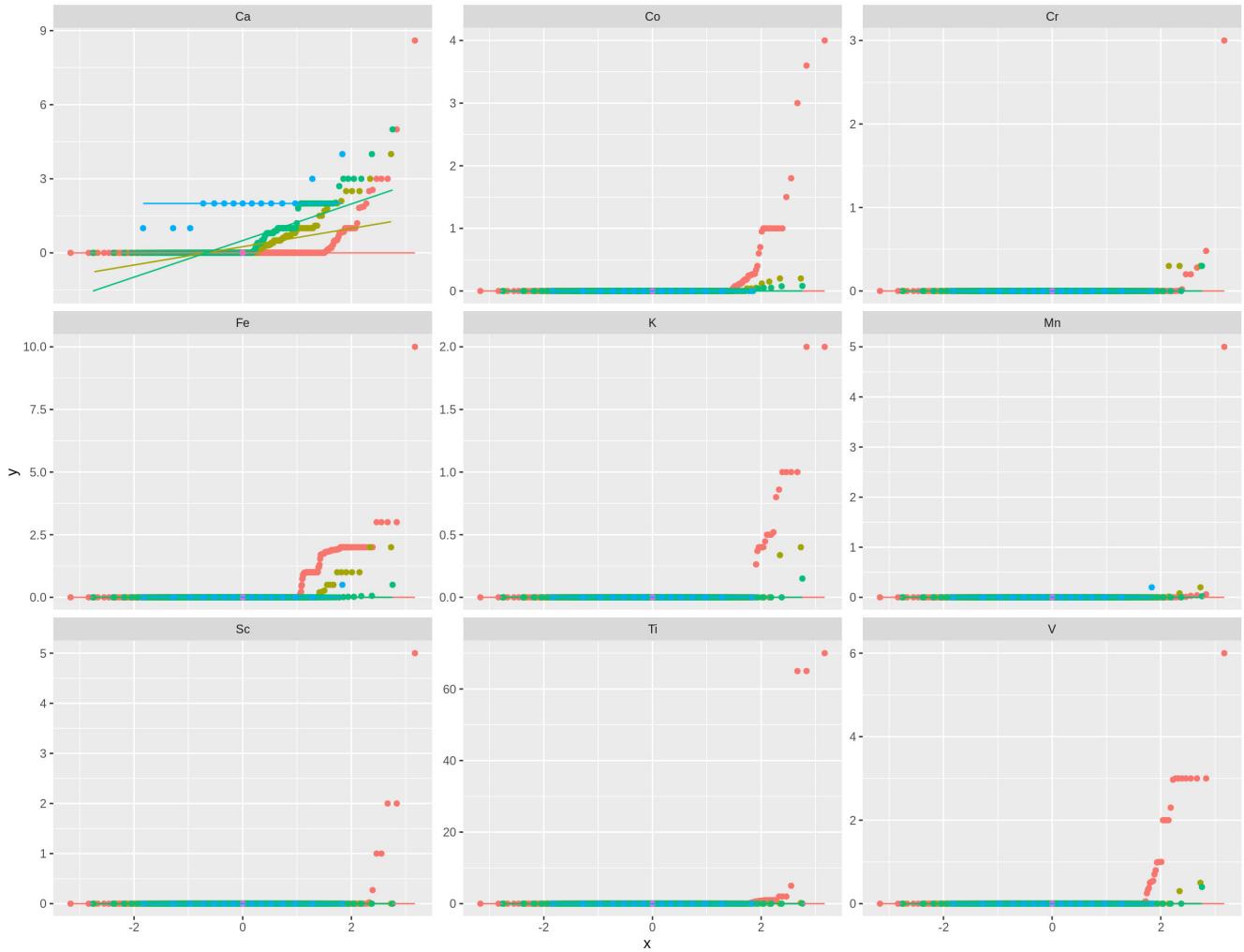


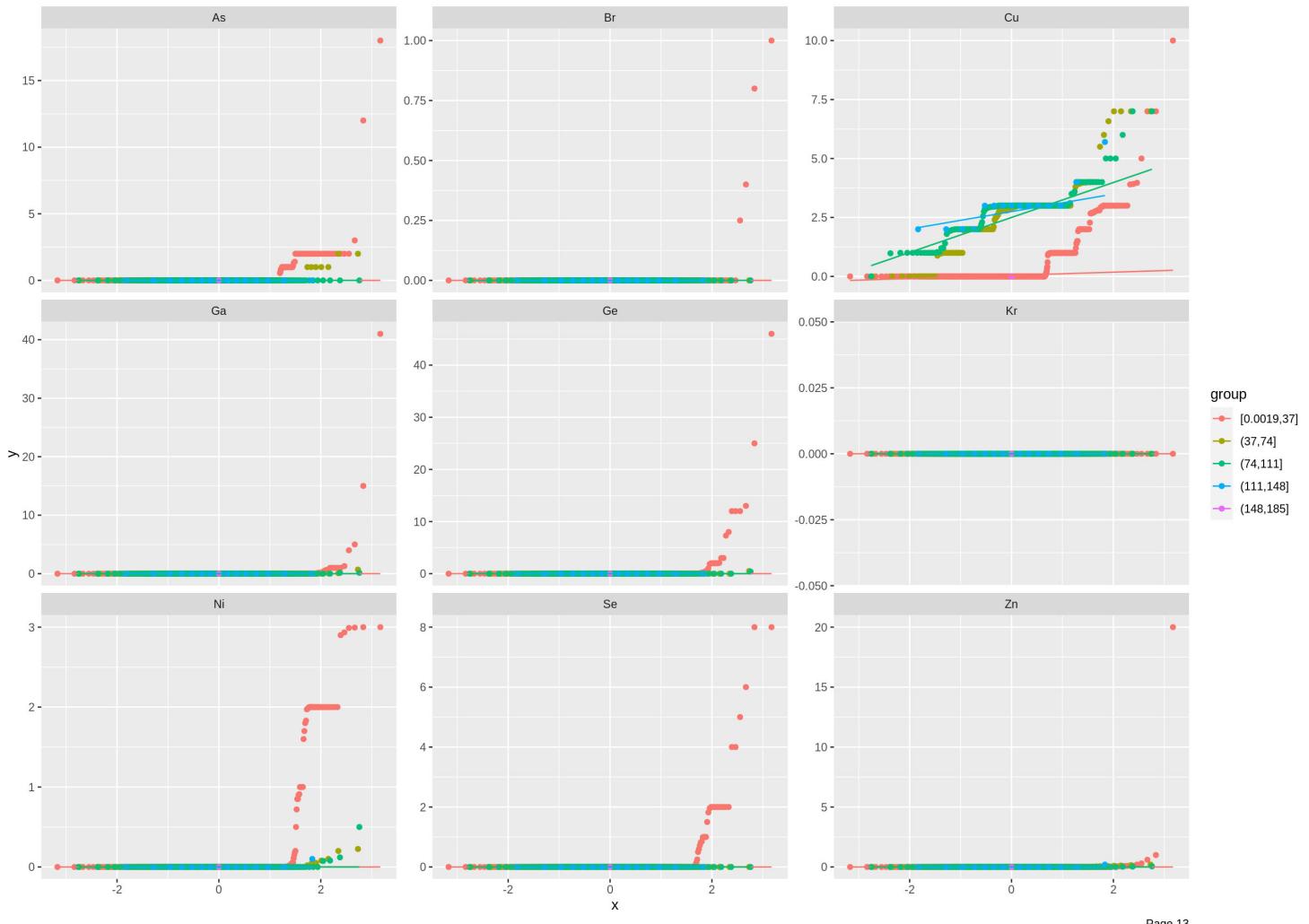


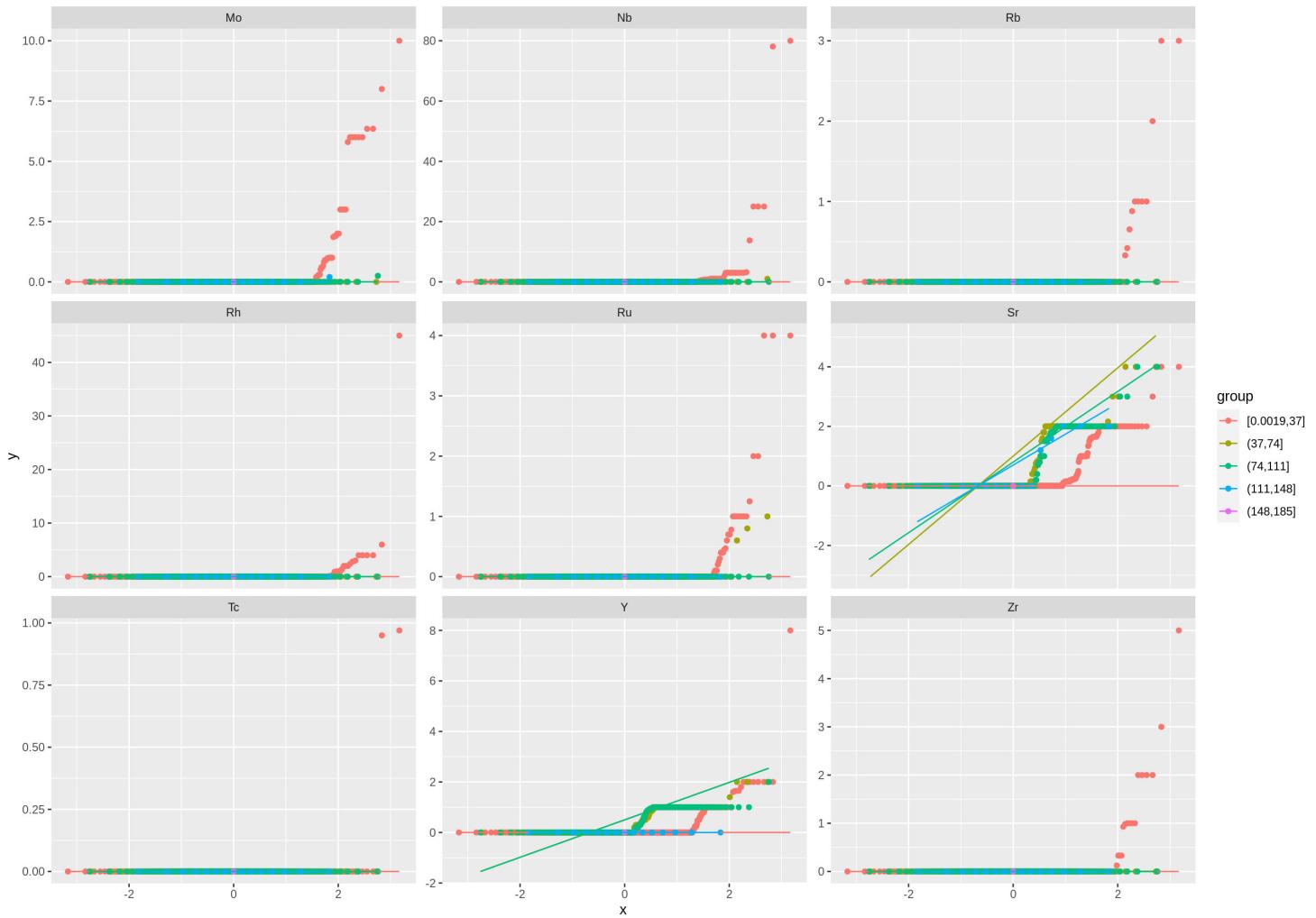


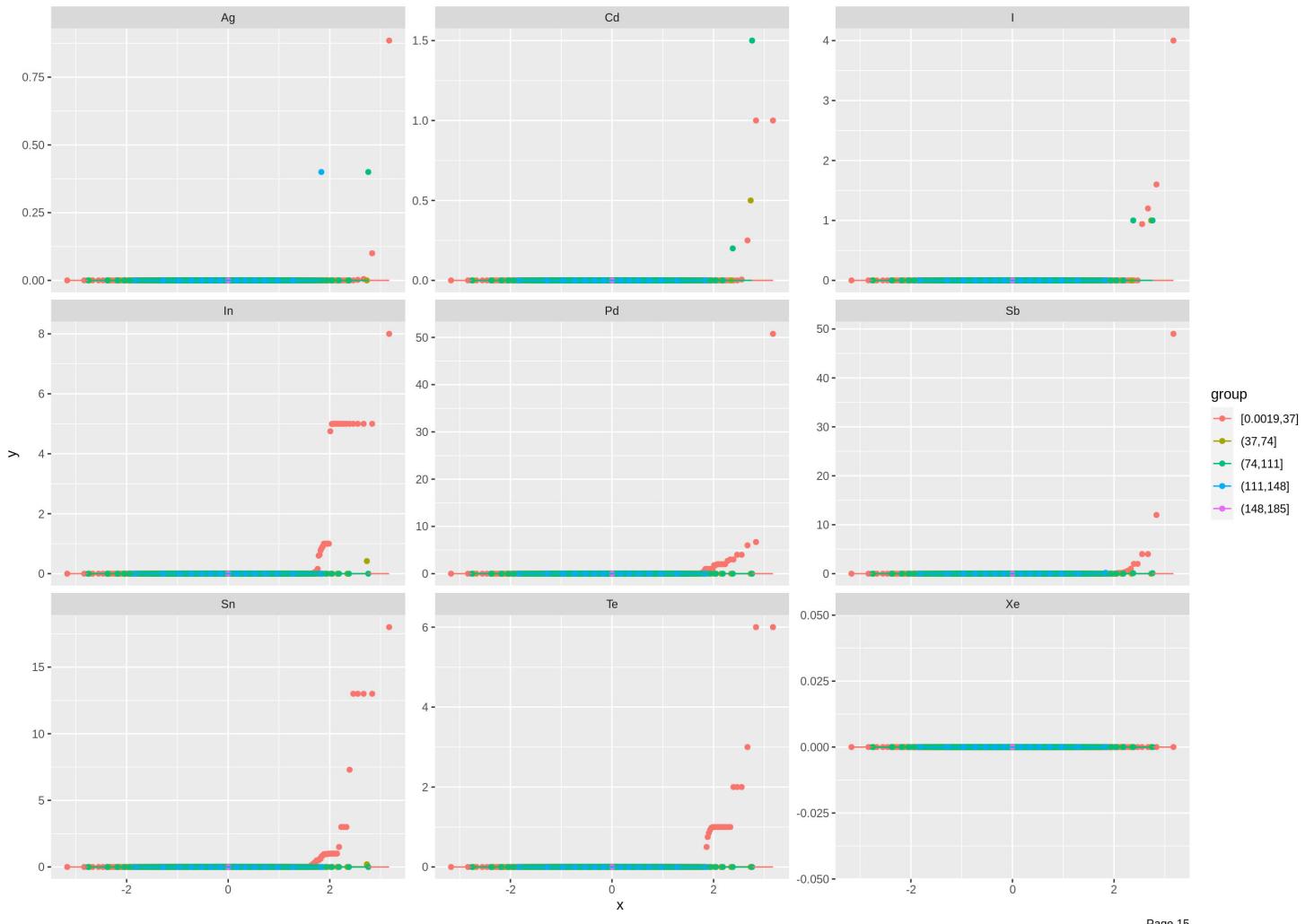


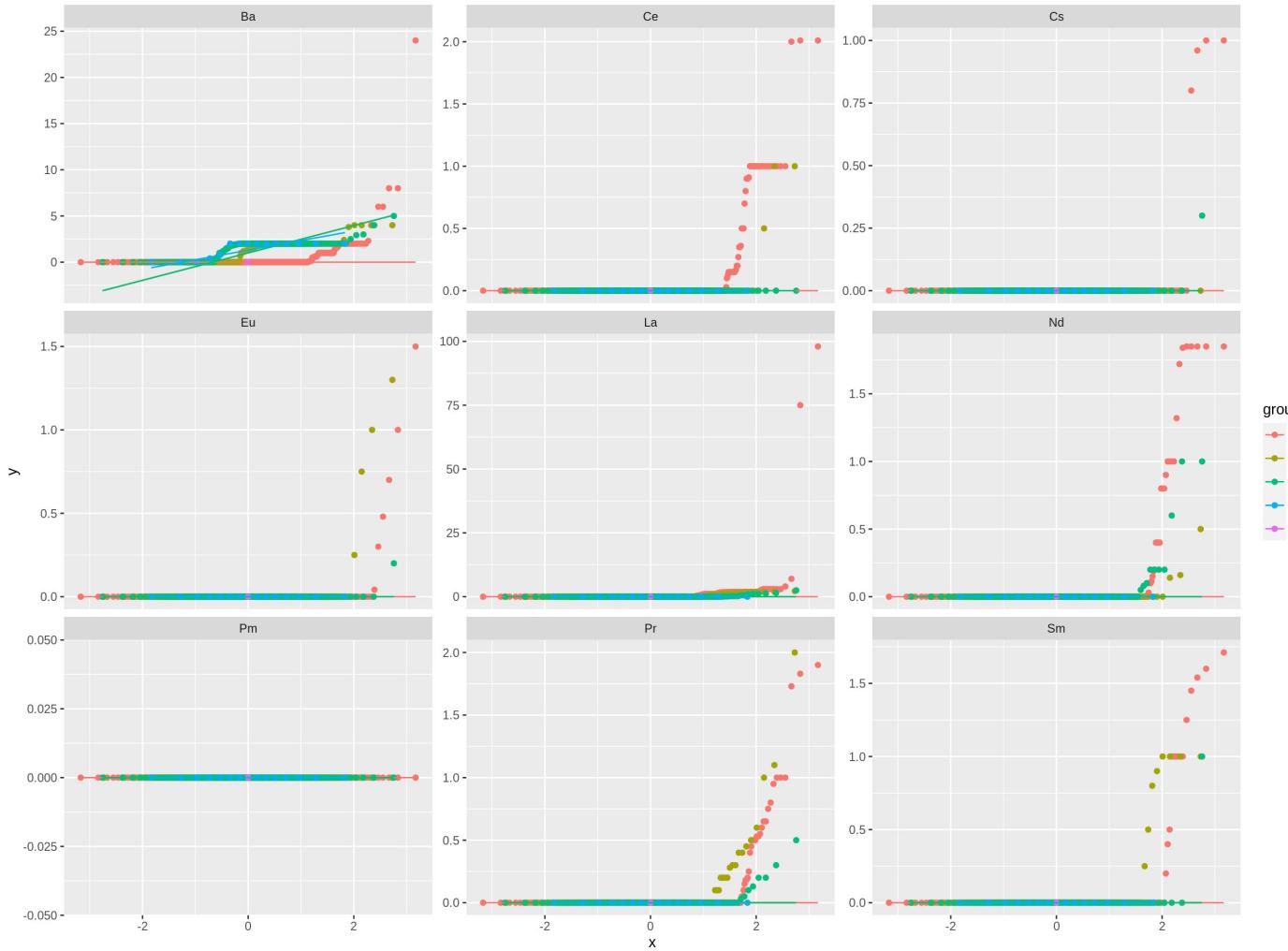


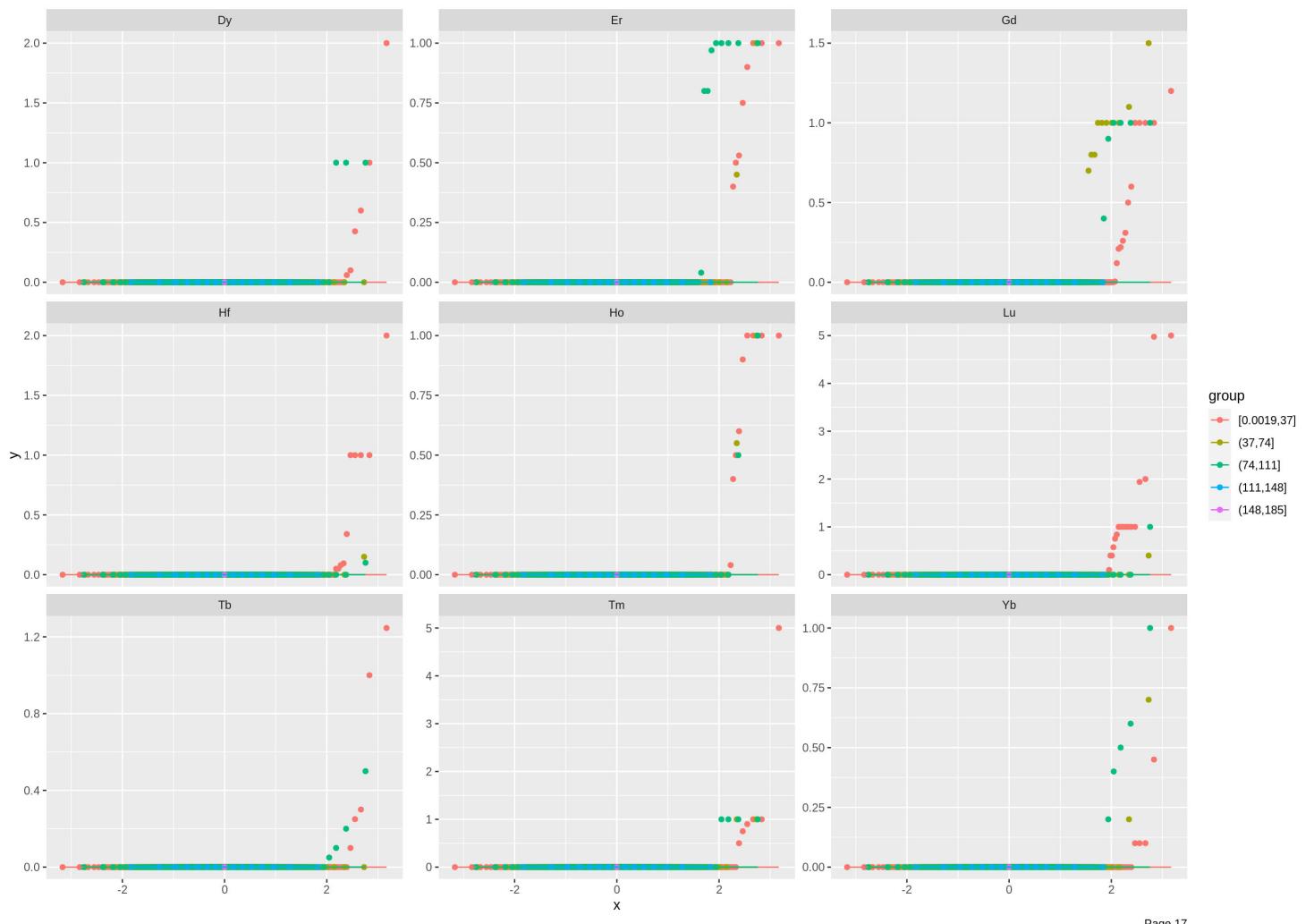


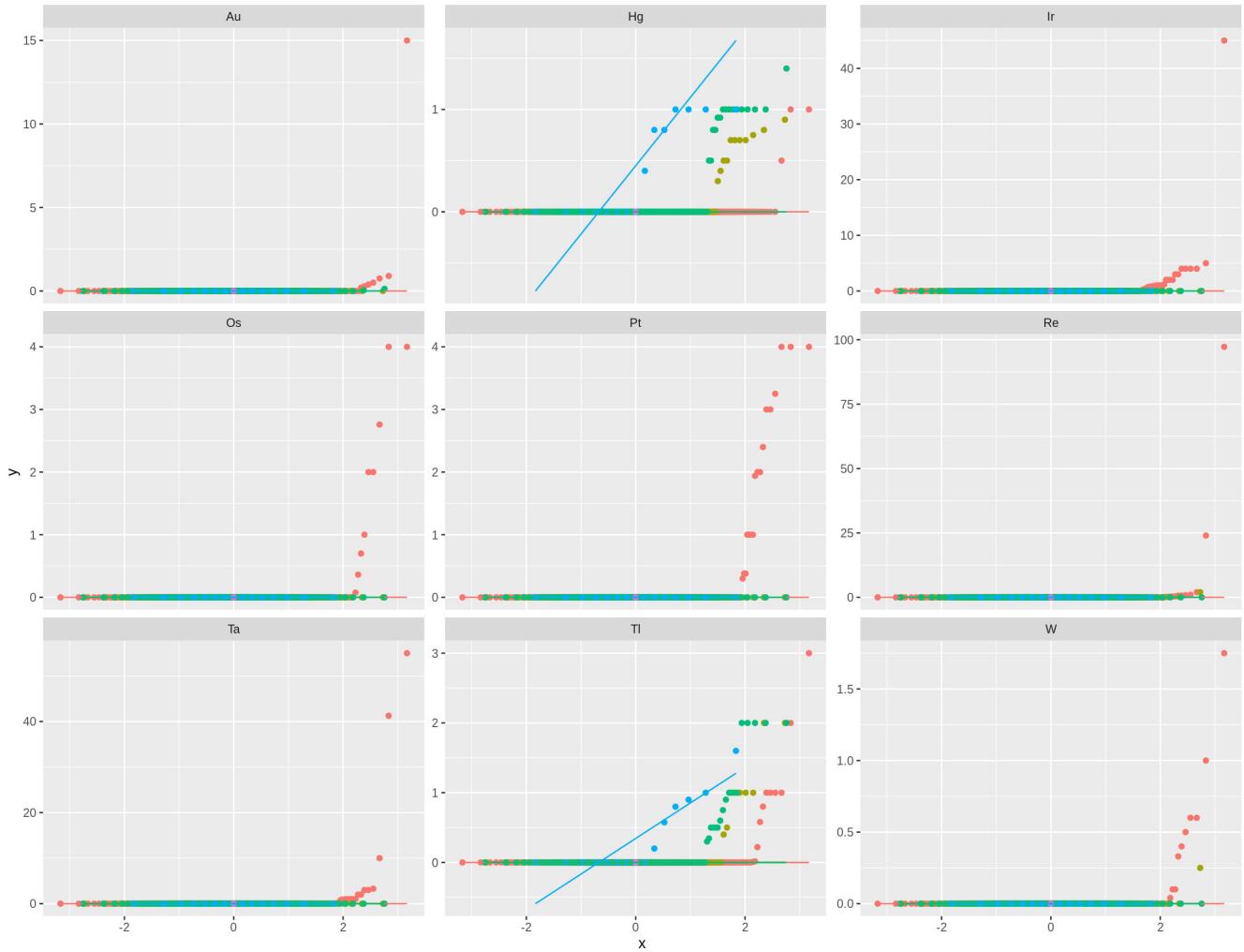


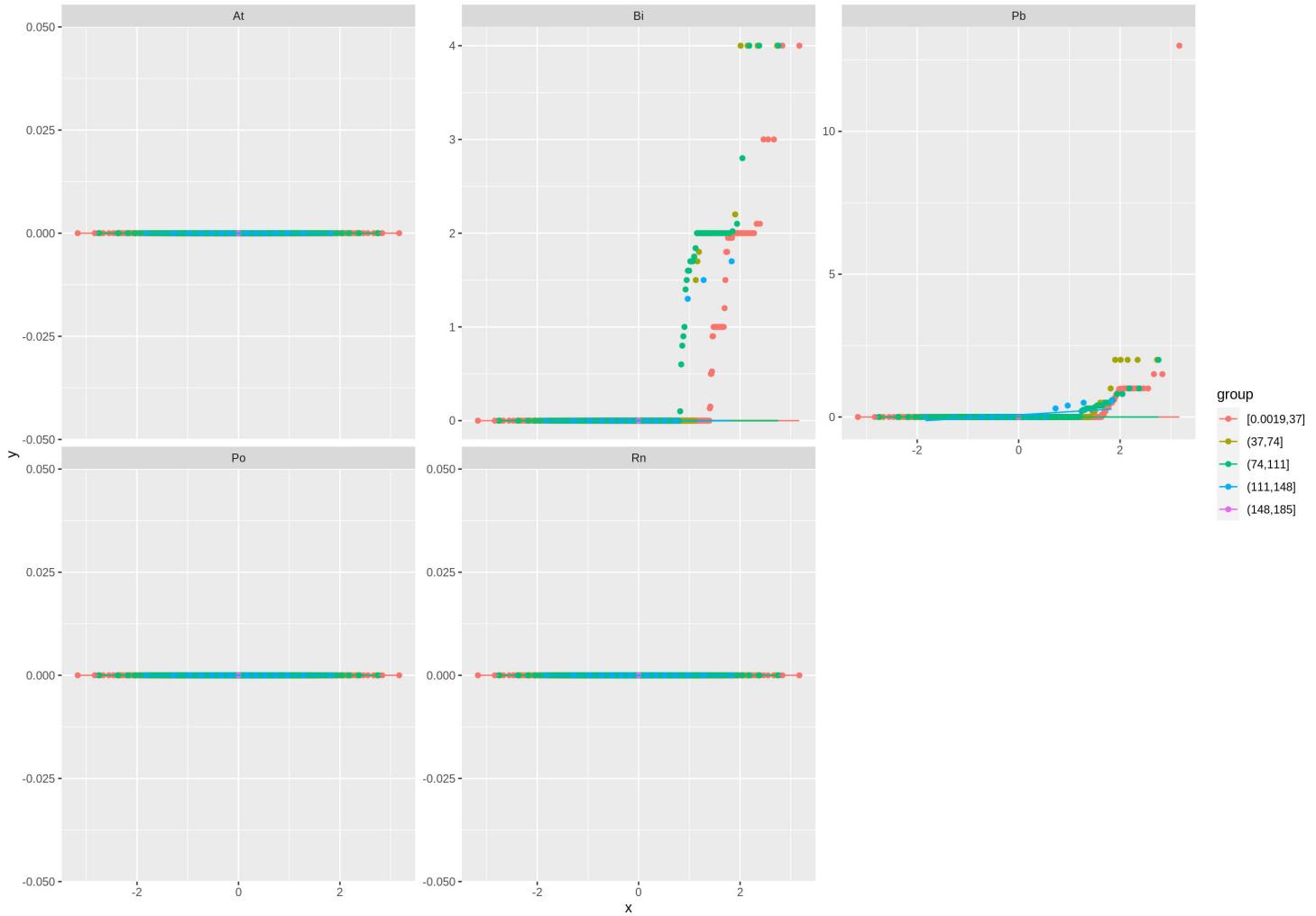










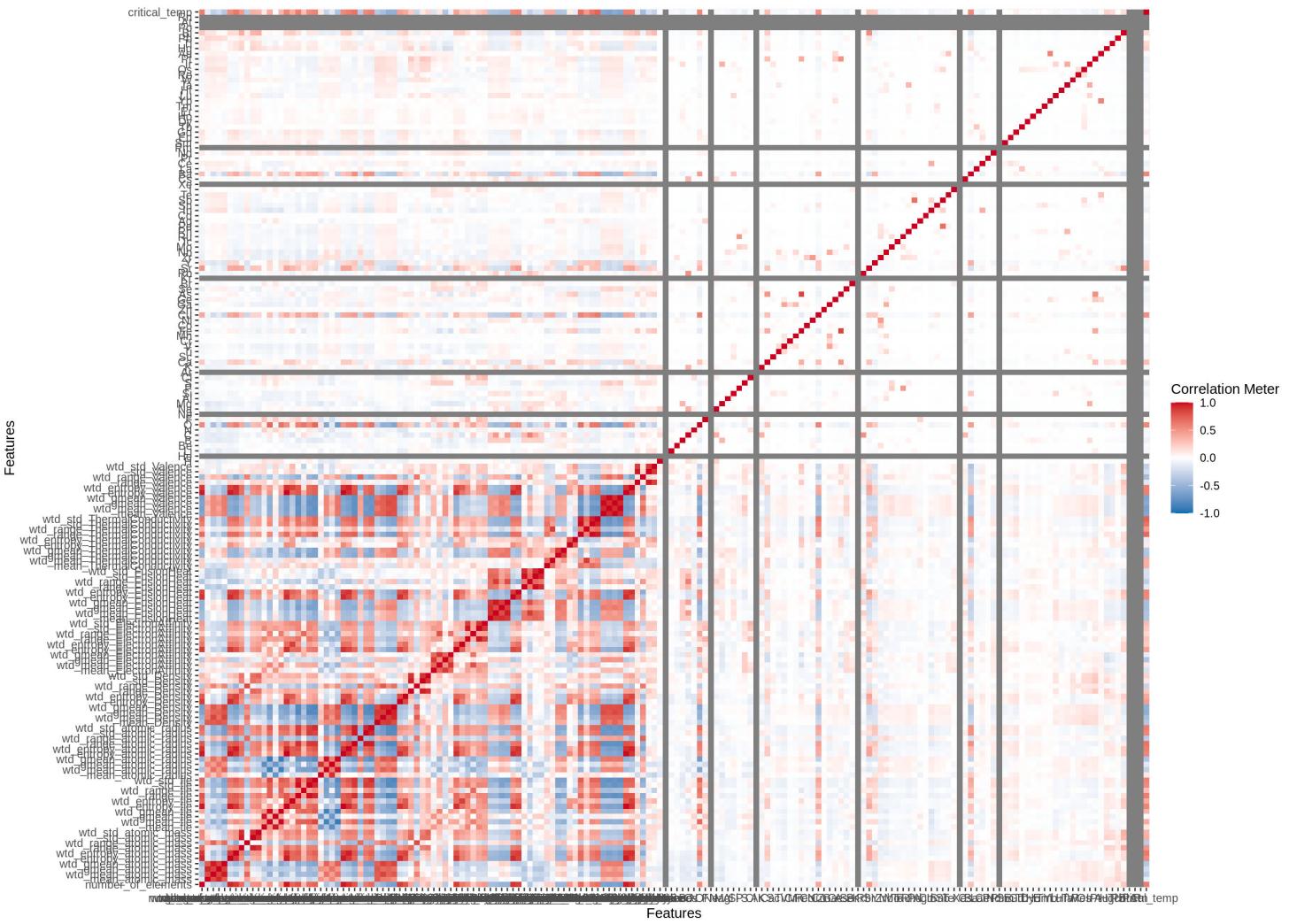


Page 19

Correlation Analysis

```
## Warning in dummify(data, maxcat = maxcat): Ignored all discrete features since
## `maxcat` set to 20 categories!
```

```
## Warning in cor(x = structure(list(number_of_elements = c(4L, 5L, 4L, 4L, : the
## standard deviation is zero
```



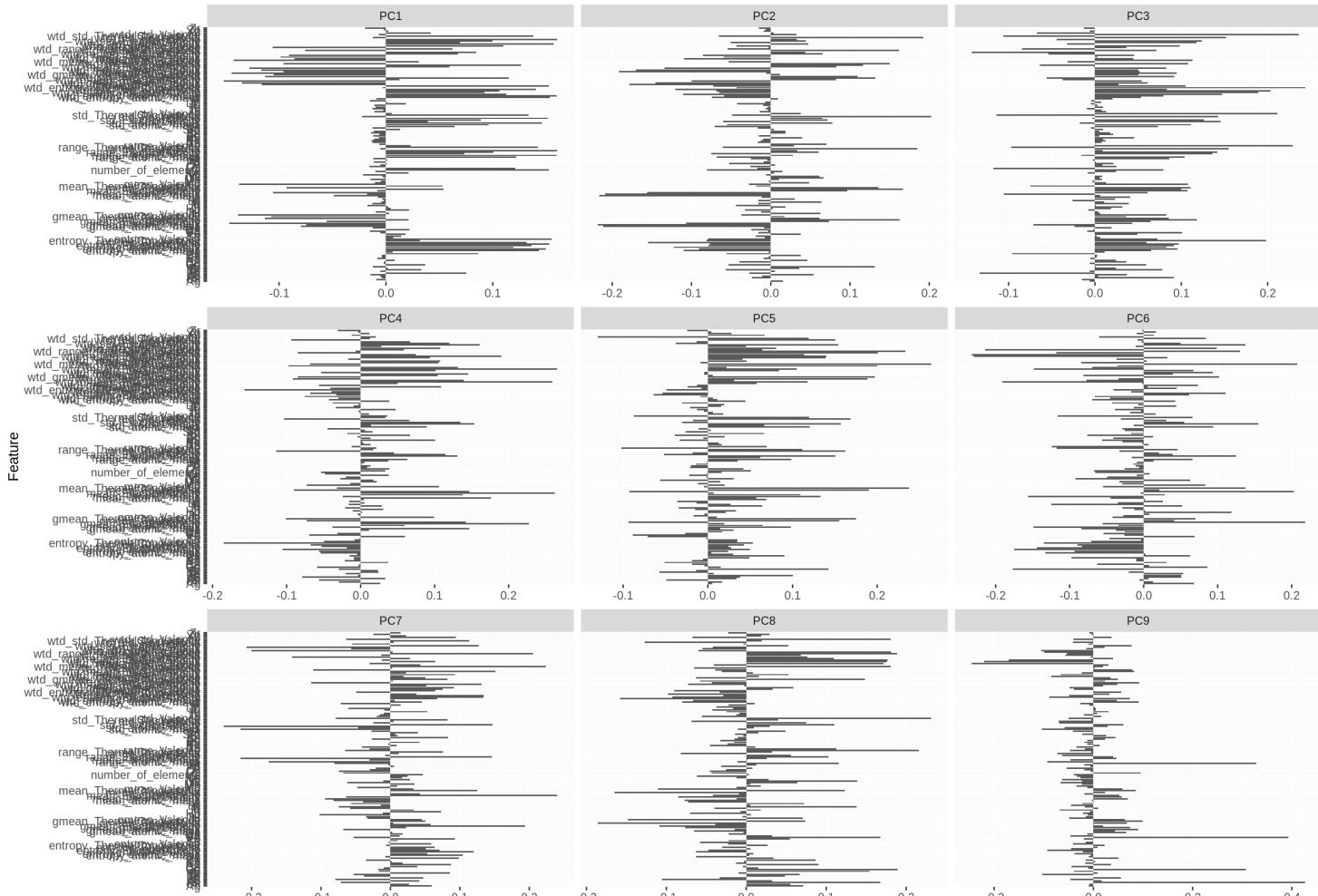
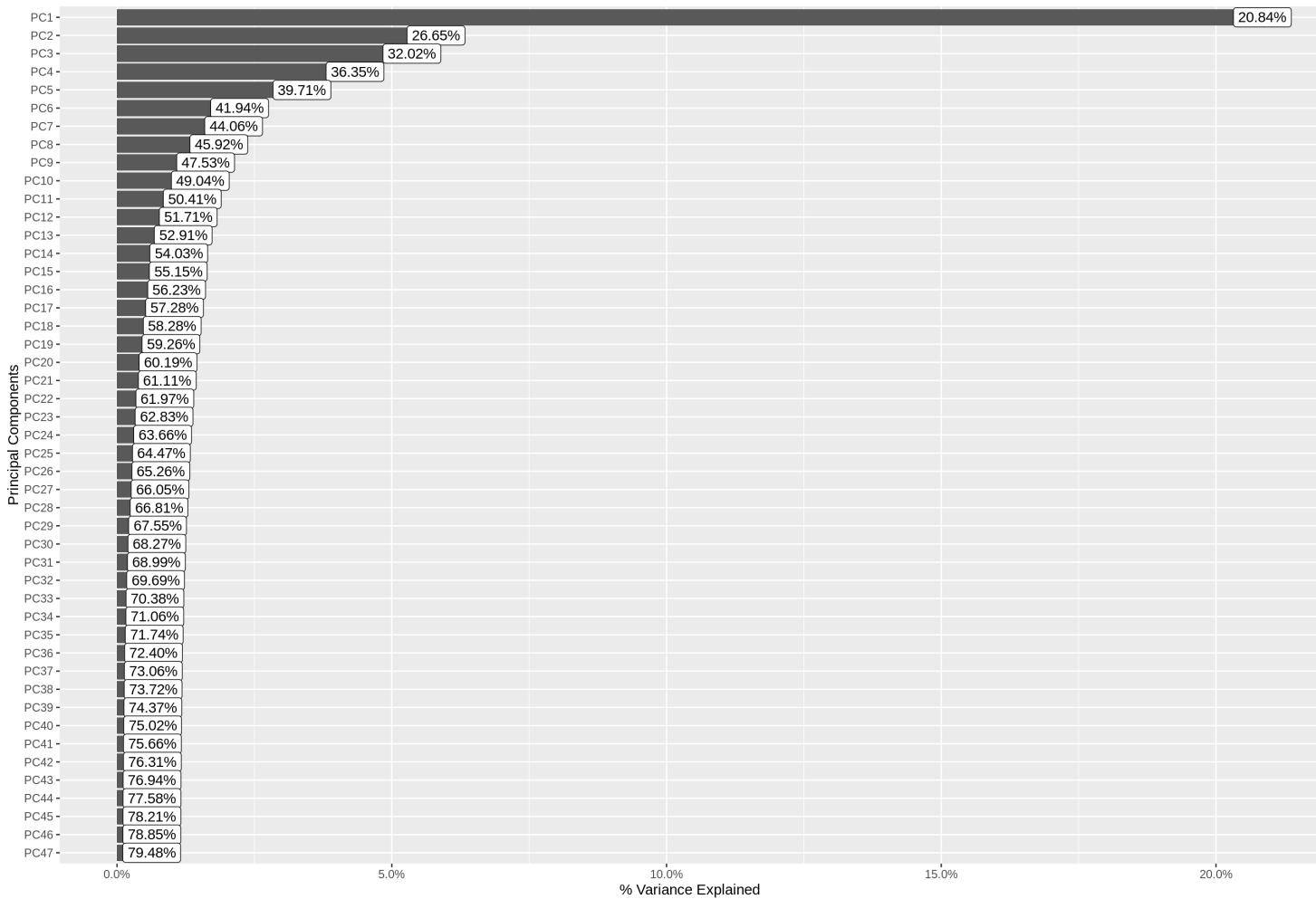
Principal Component Analysis

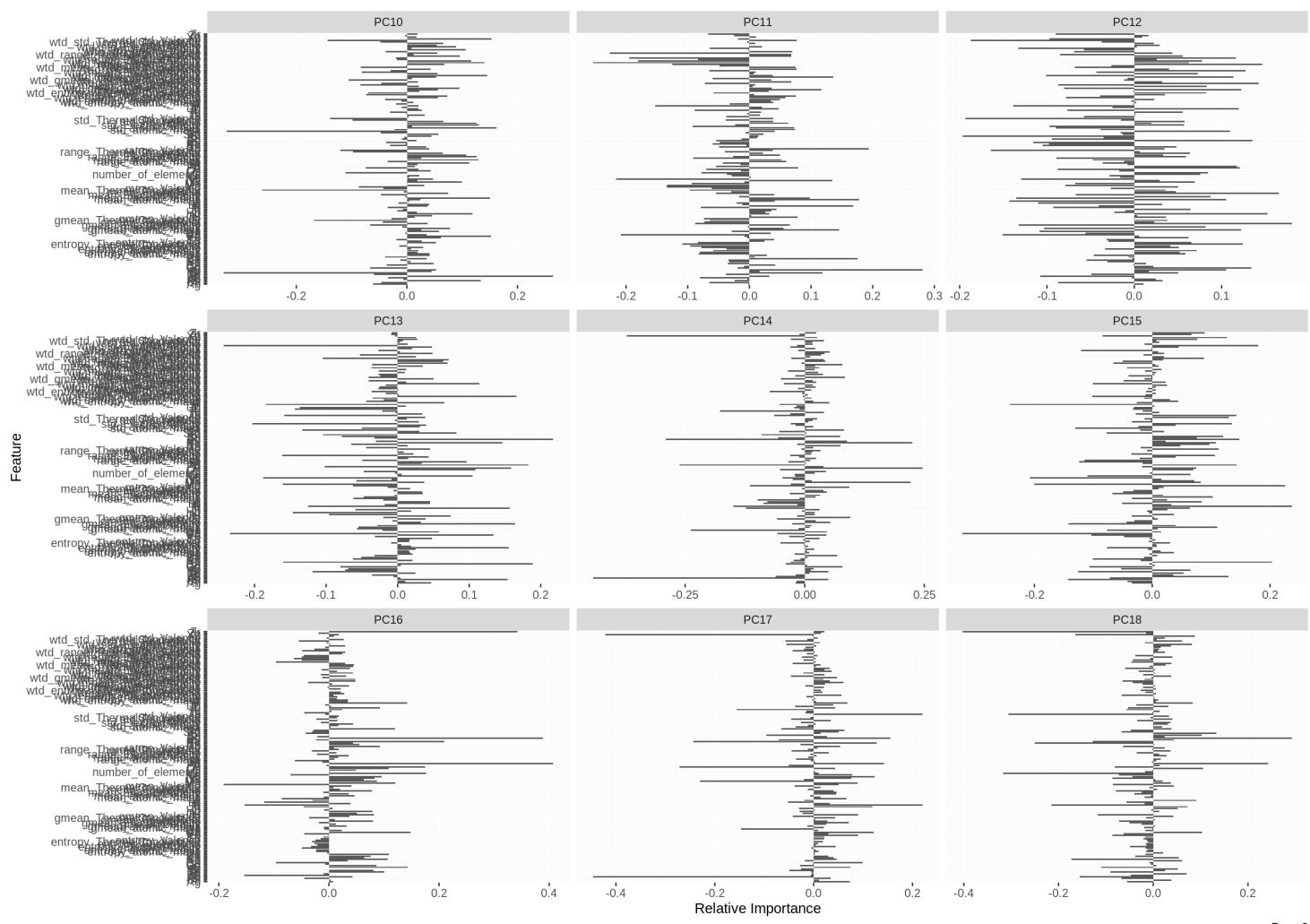
```

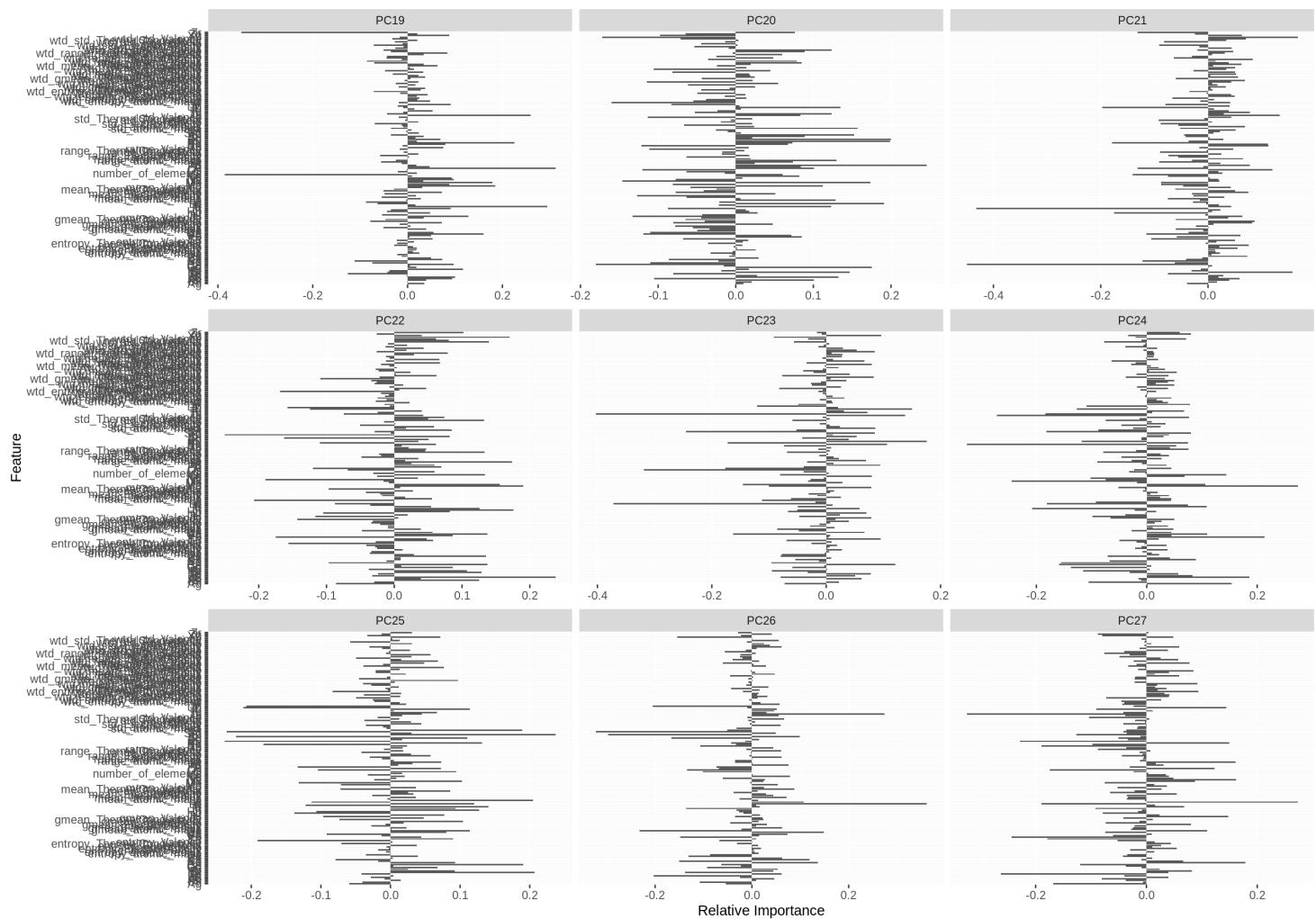
## Warning in (function (data, variance_cap = 0.8, maxcat = 50L, prcomp_args = list(scale. = TRUE), : The following
features are dropped due to zero variance:
## * He
## * Ne
## * Ar
## * Kr
## * Xe
## * Pm
## * Po
## * At
## * Rn

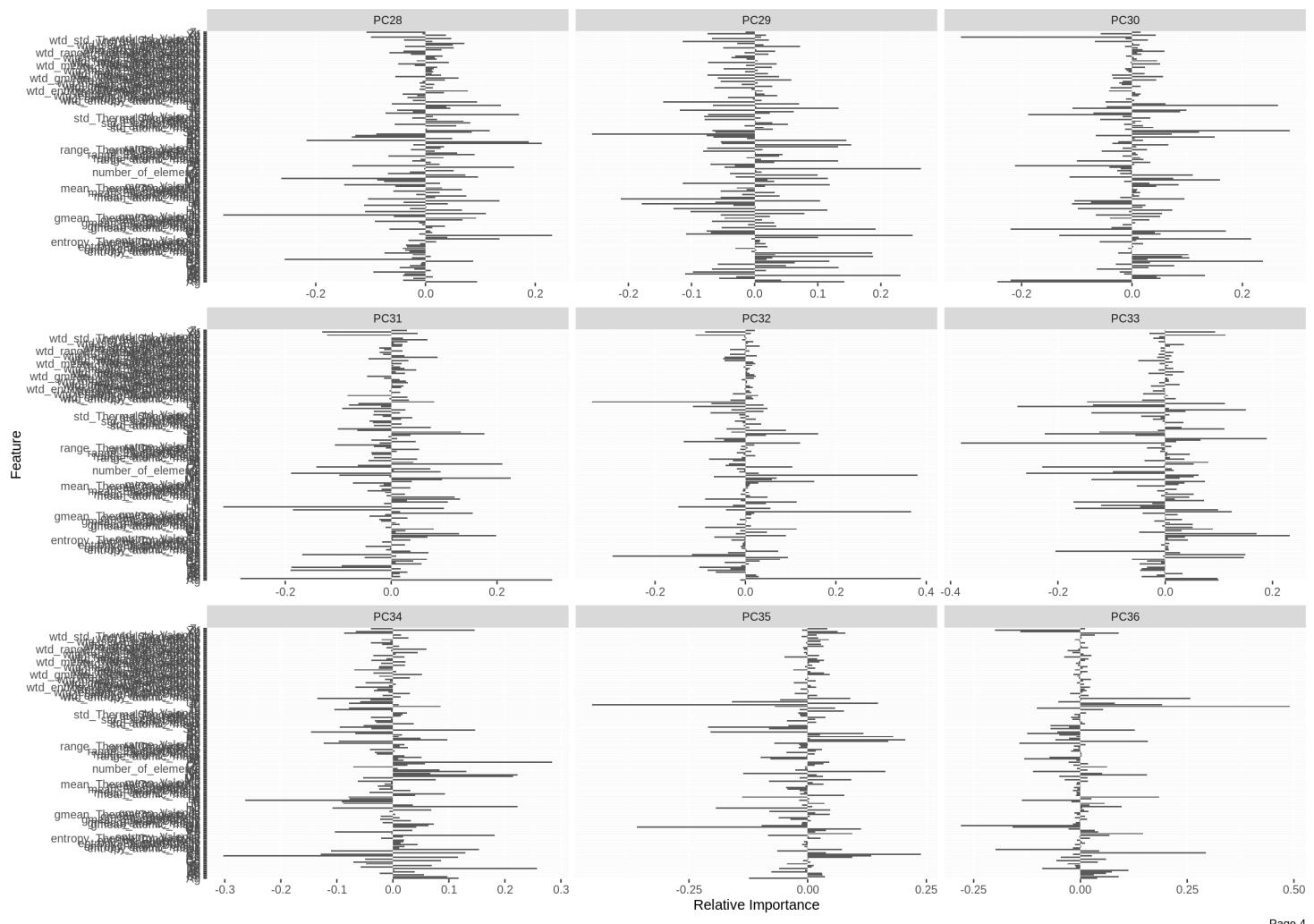
```

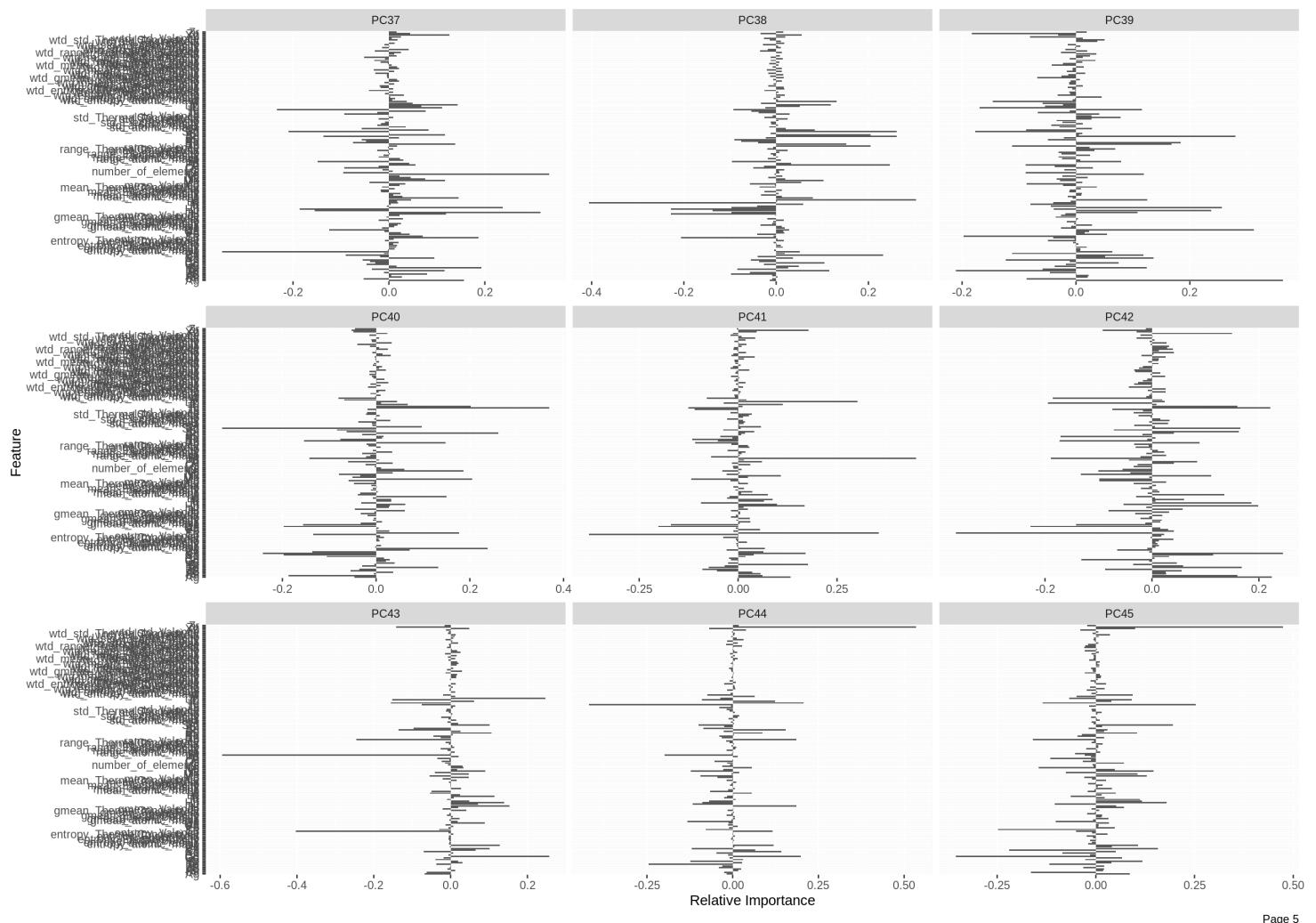

% Variance Explained By Principal Components (Note: Labels indicate cumulative % explained variance)

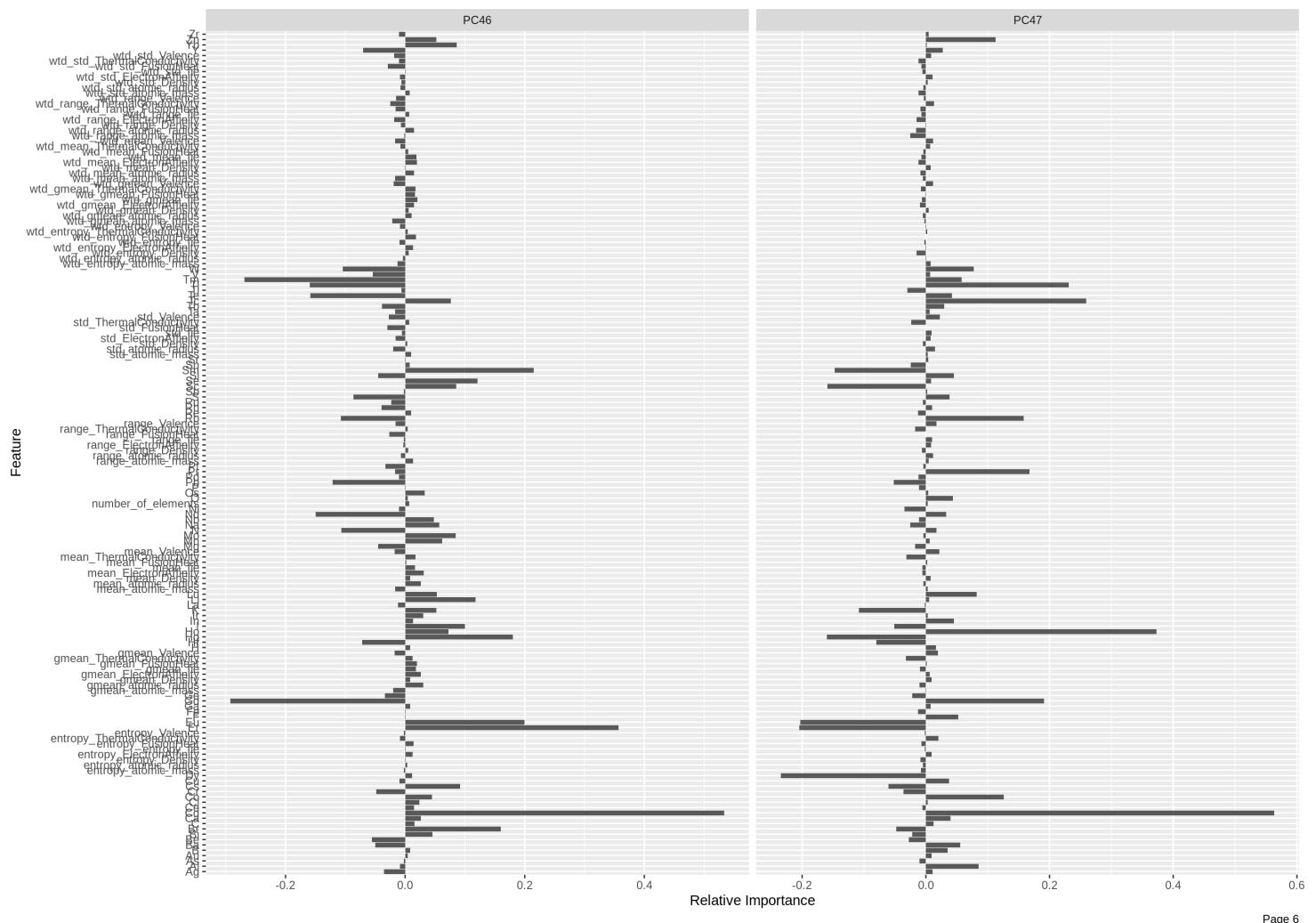






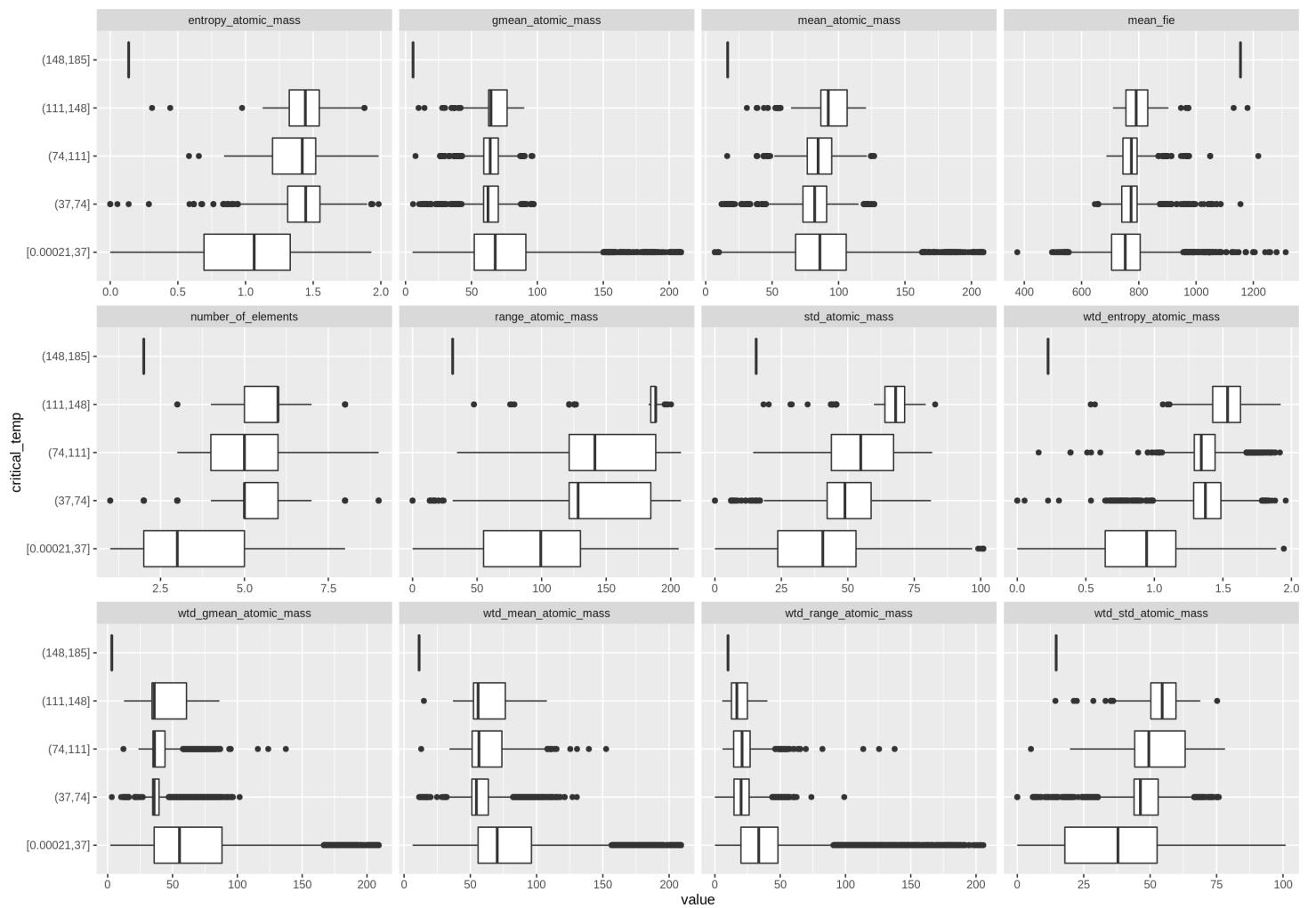




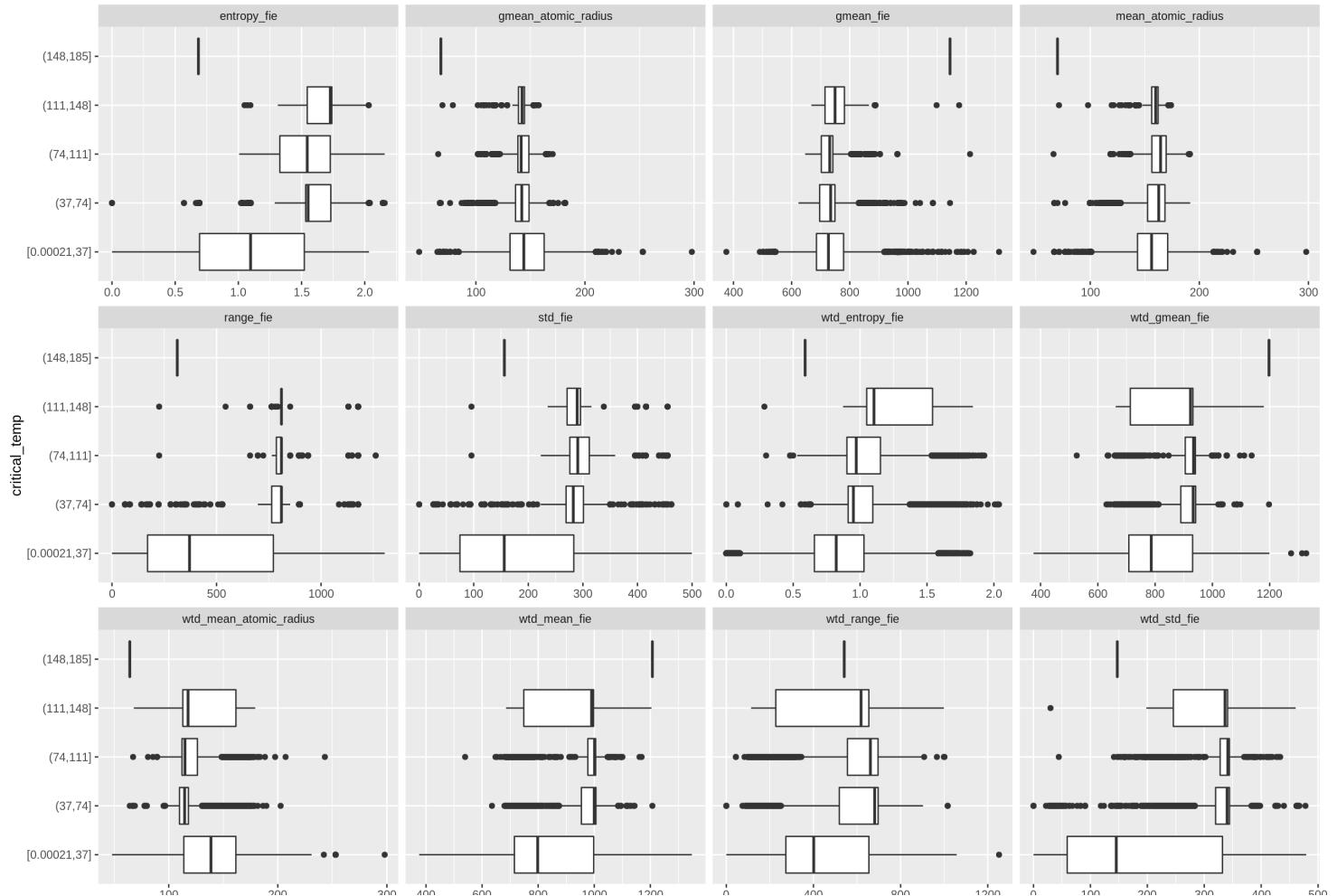


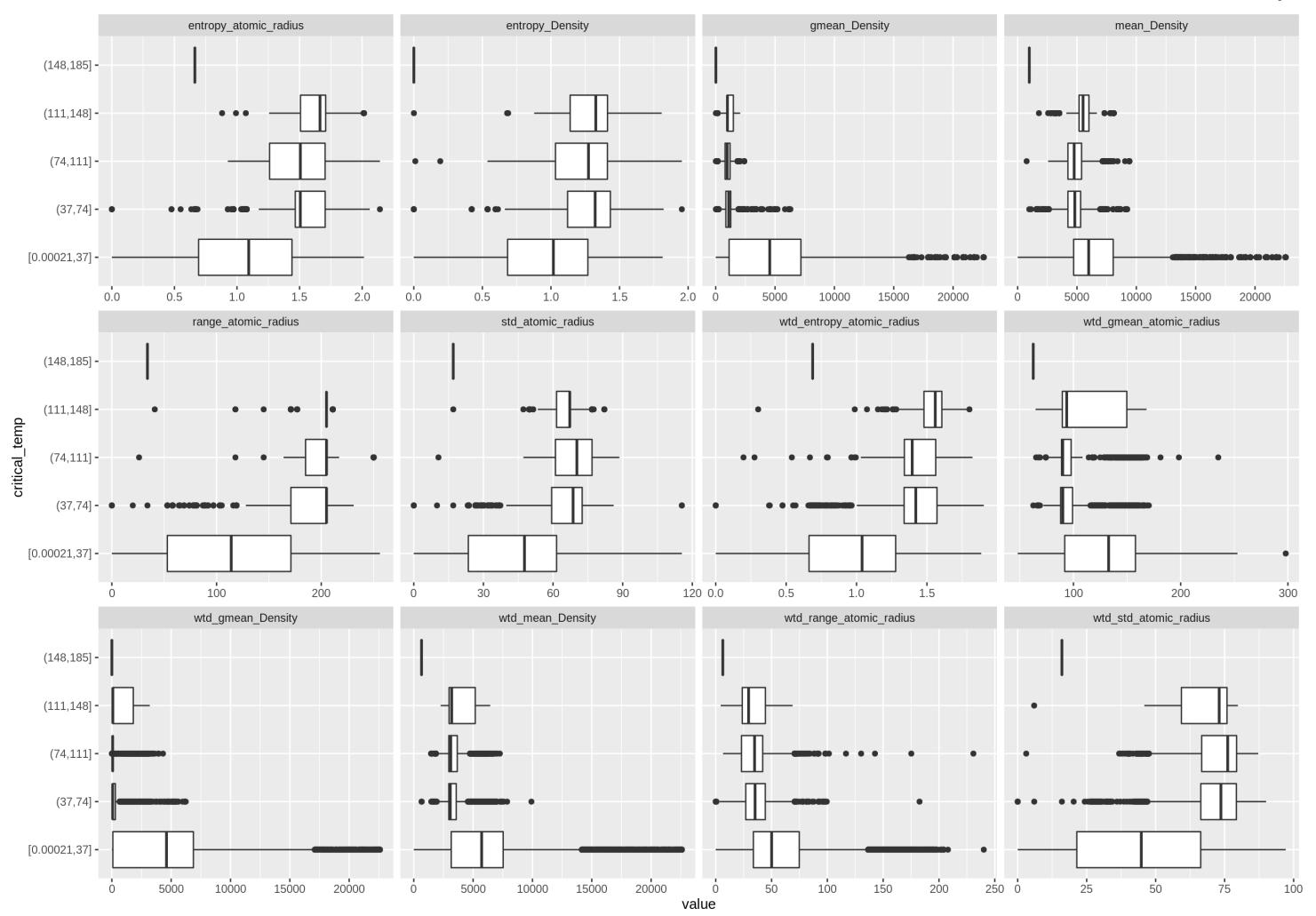
Bivariate Distribution

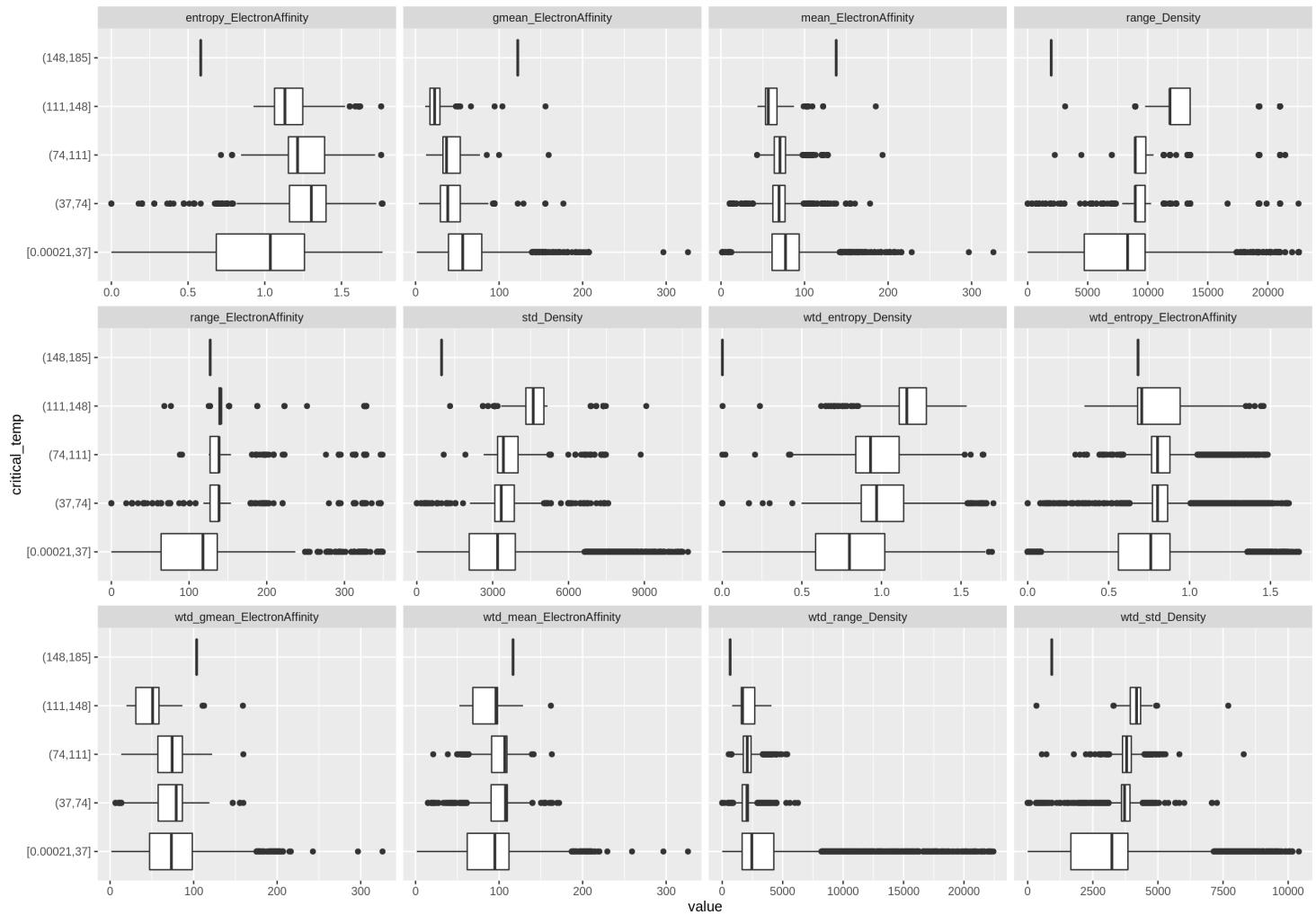
Boxplot (by critical_temp)

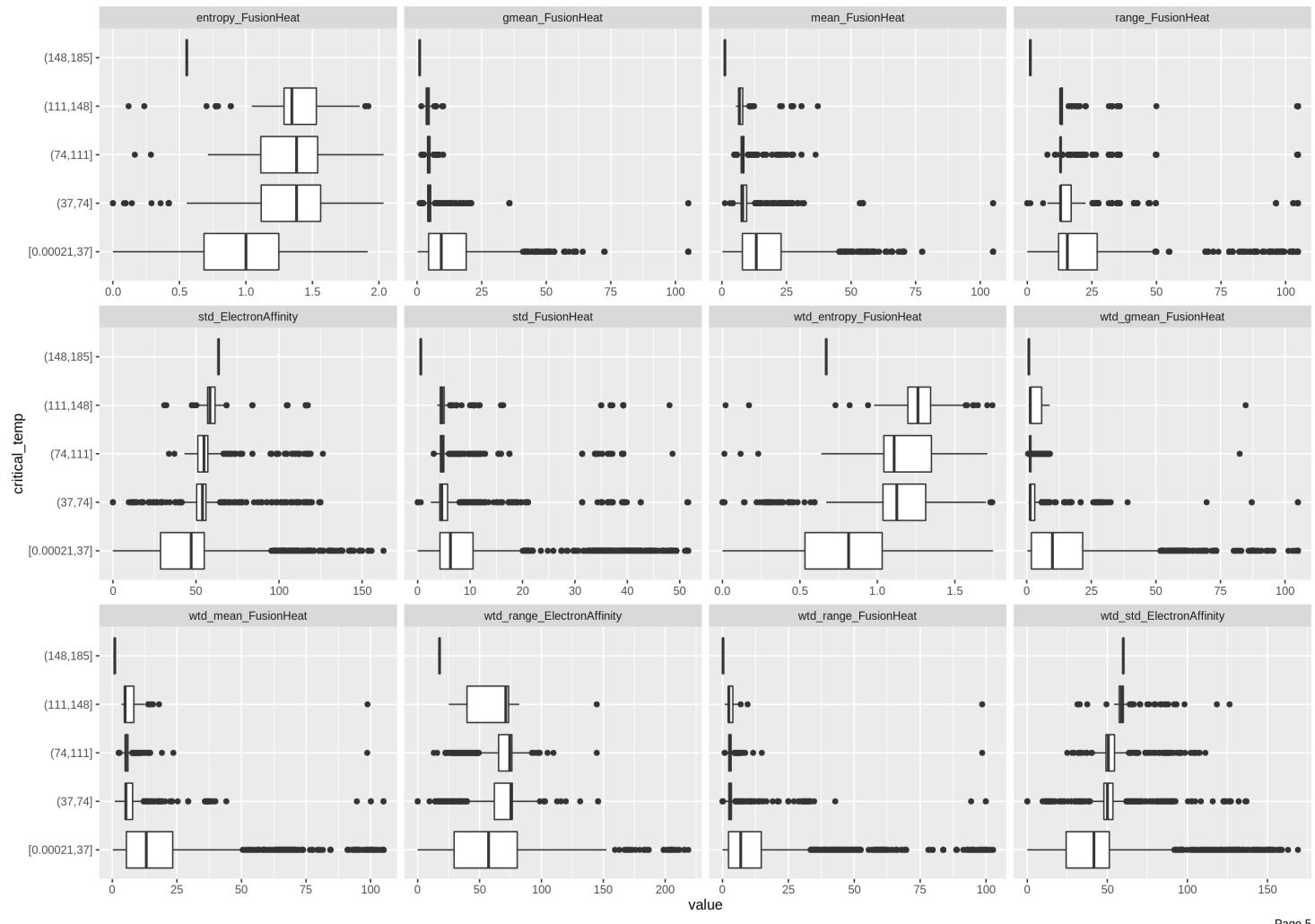


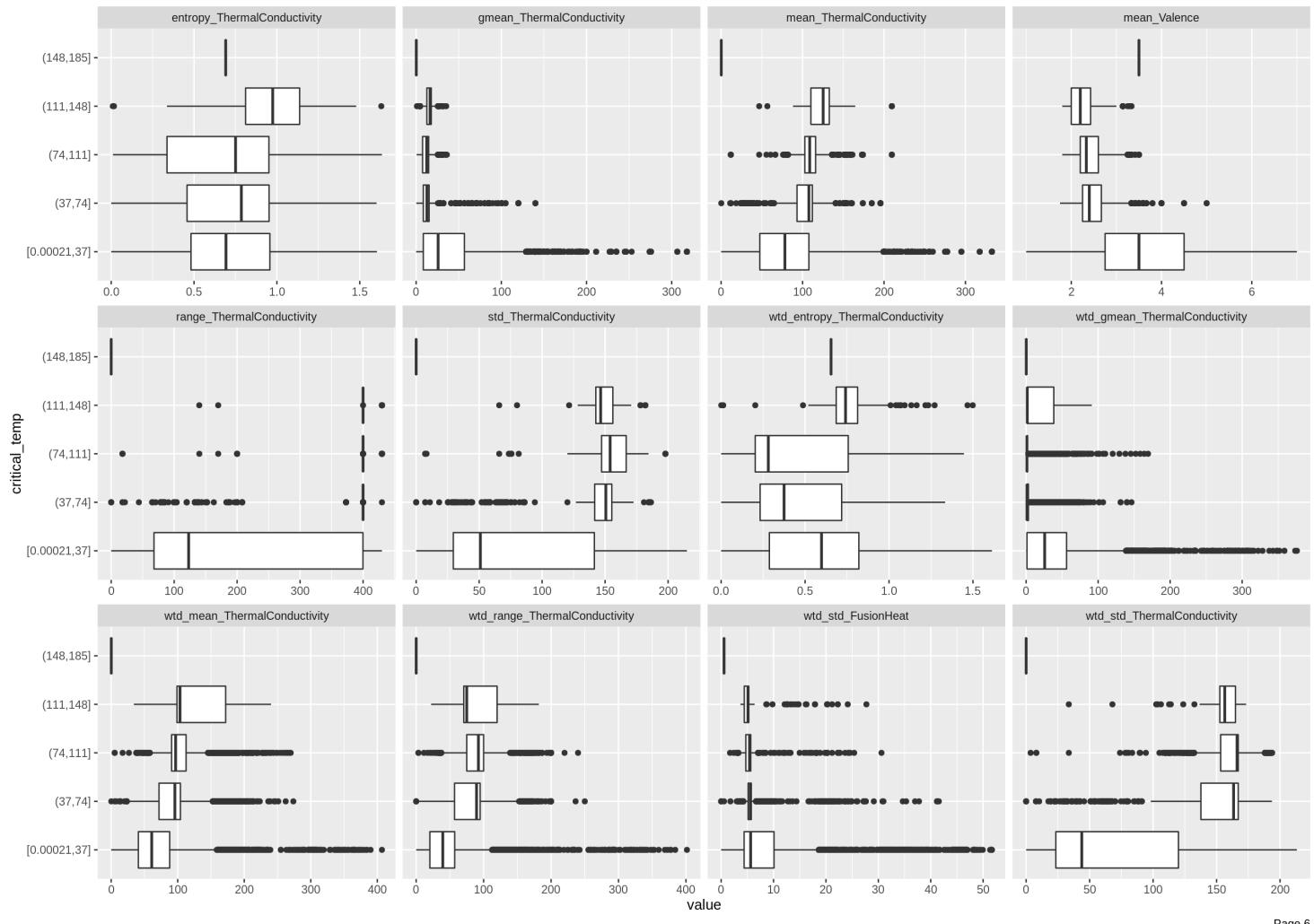
Page 1

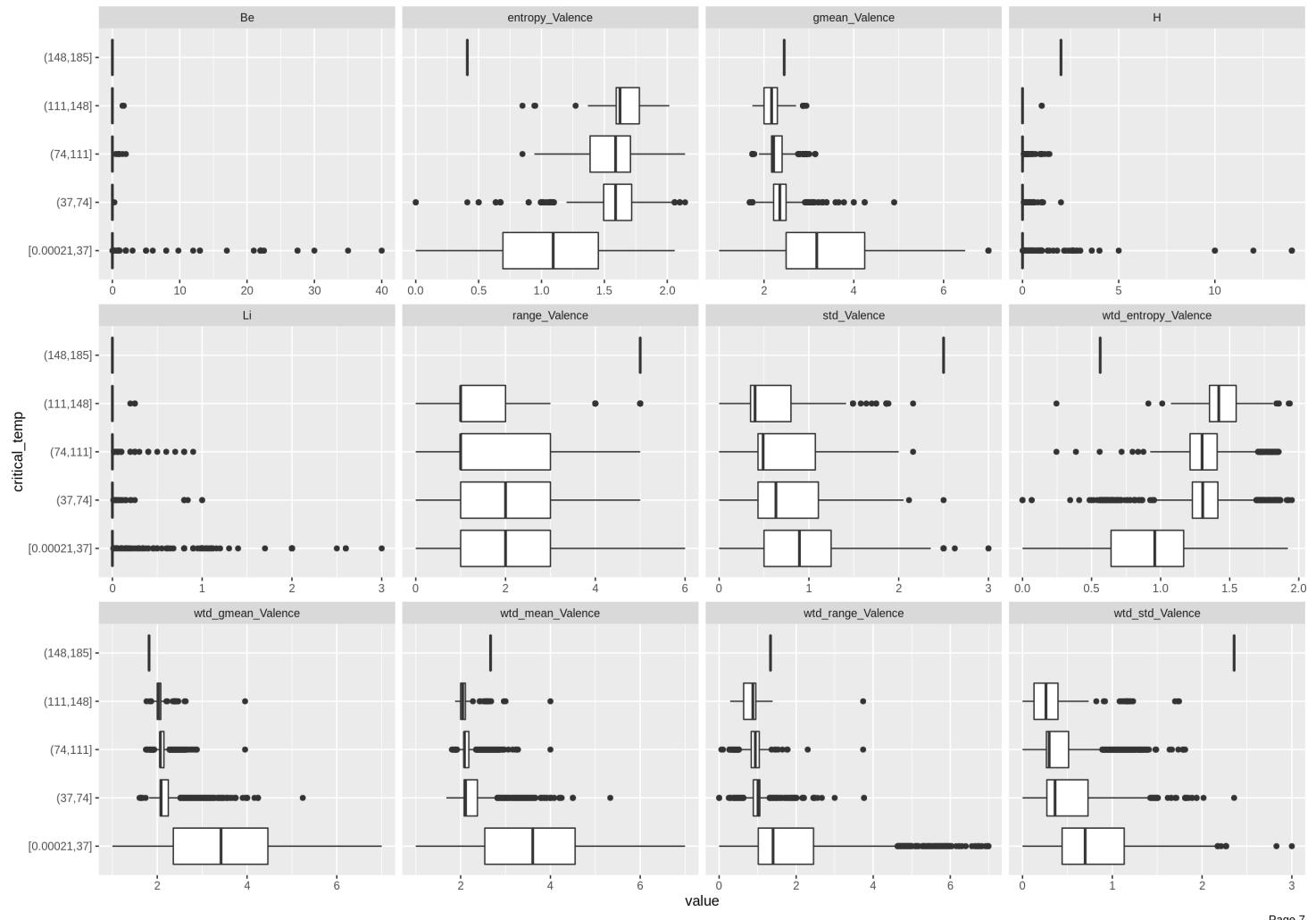


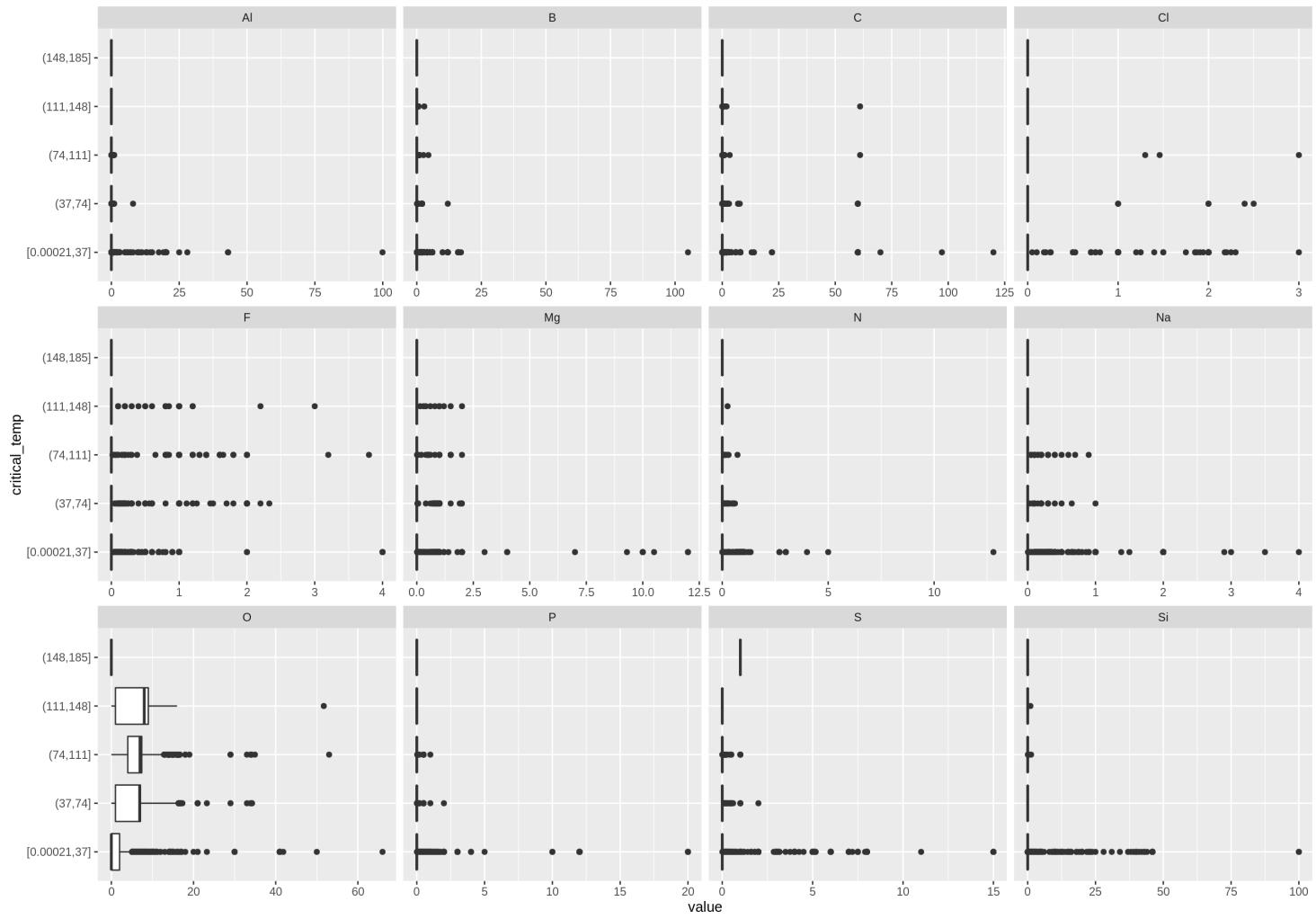


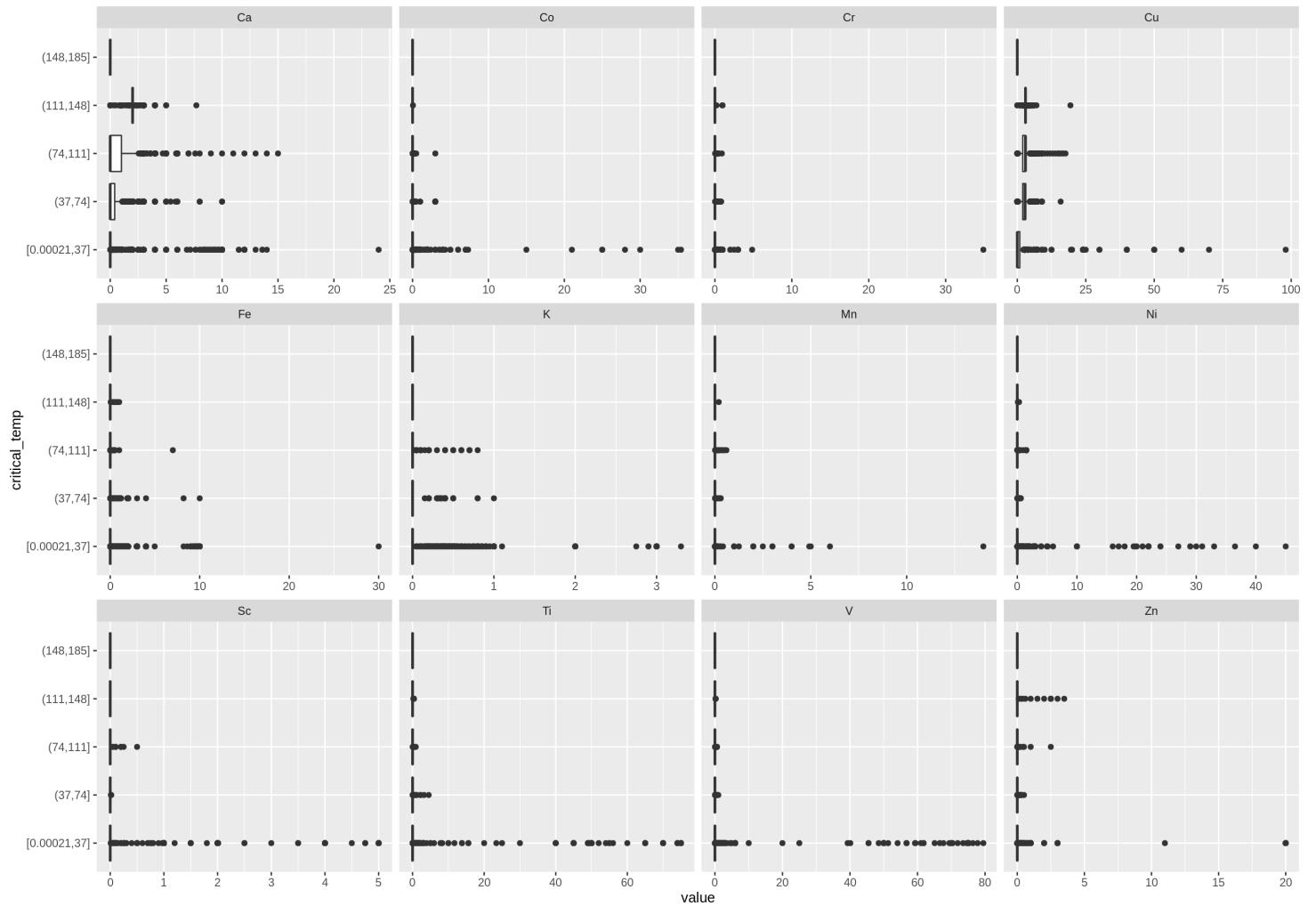


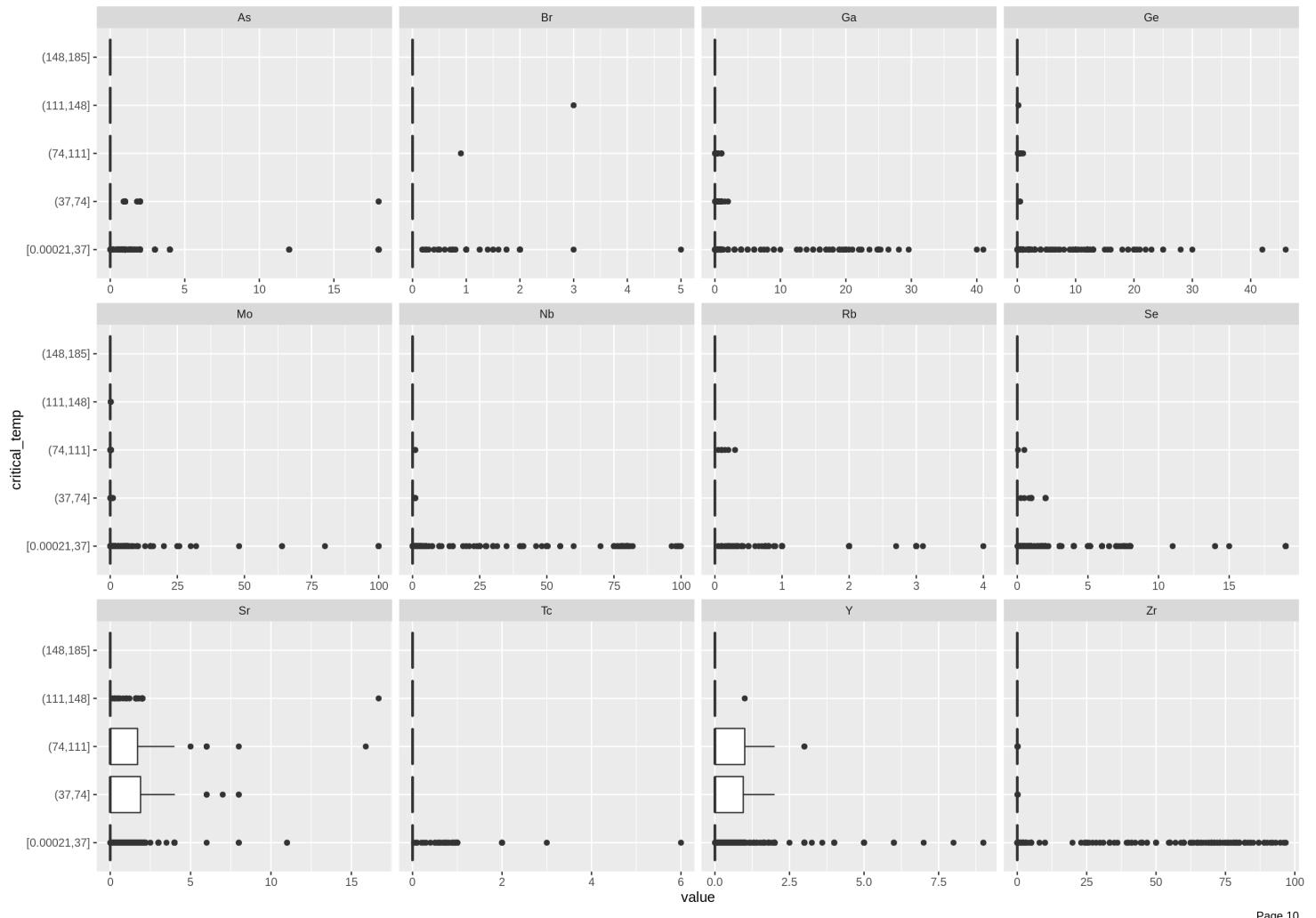


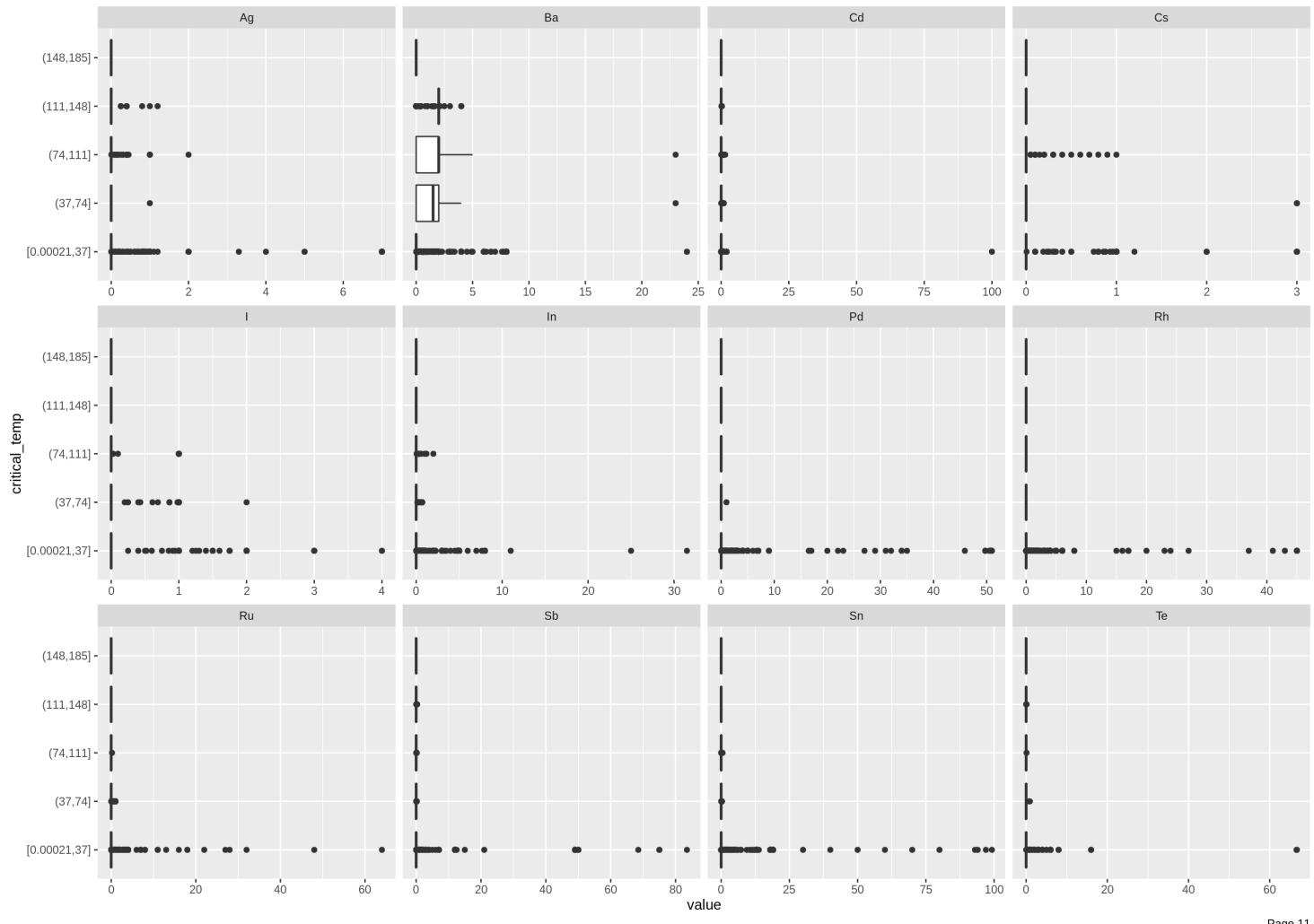


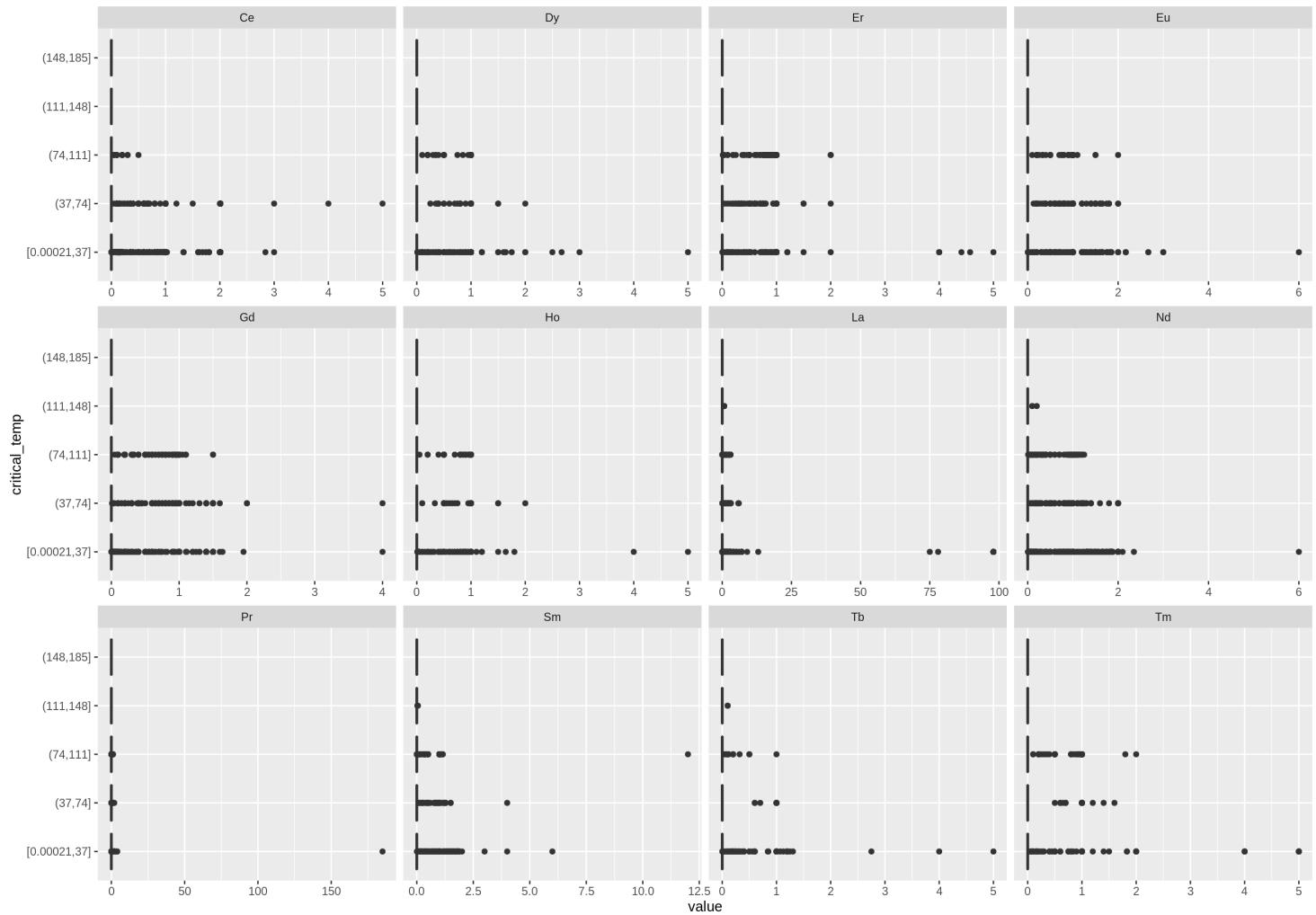


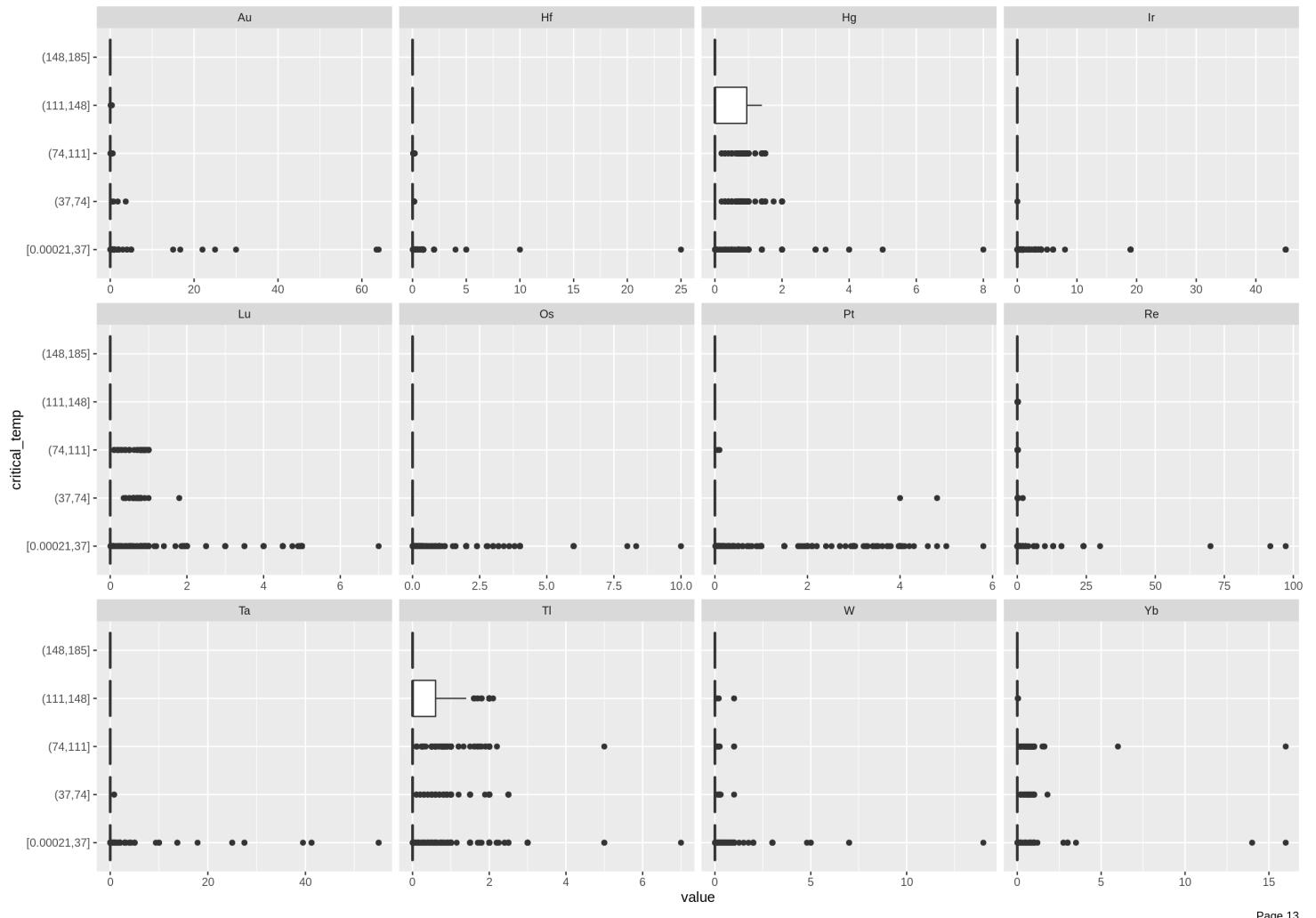


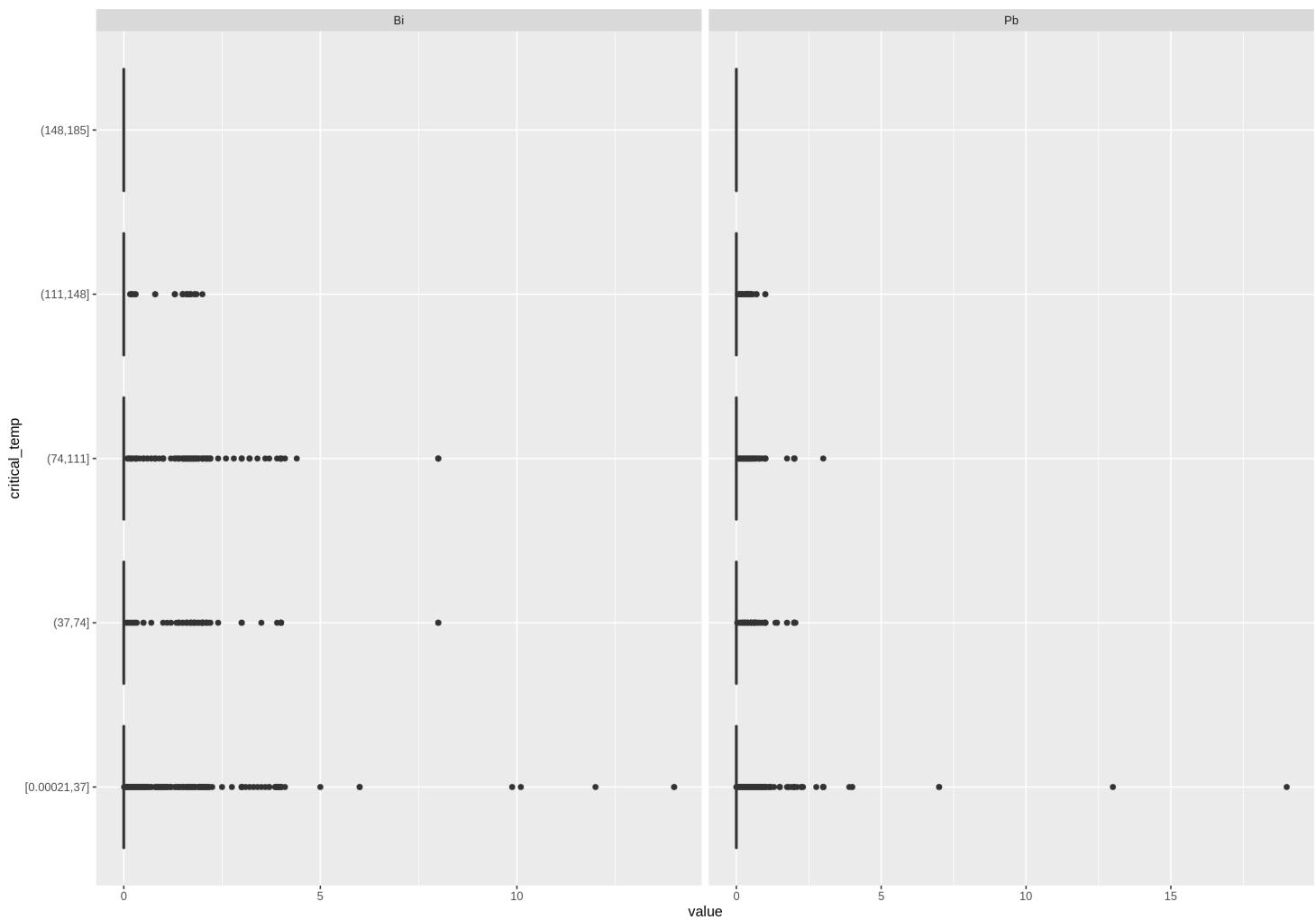




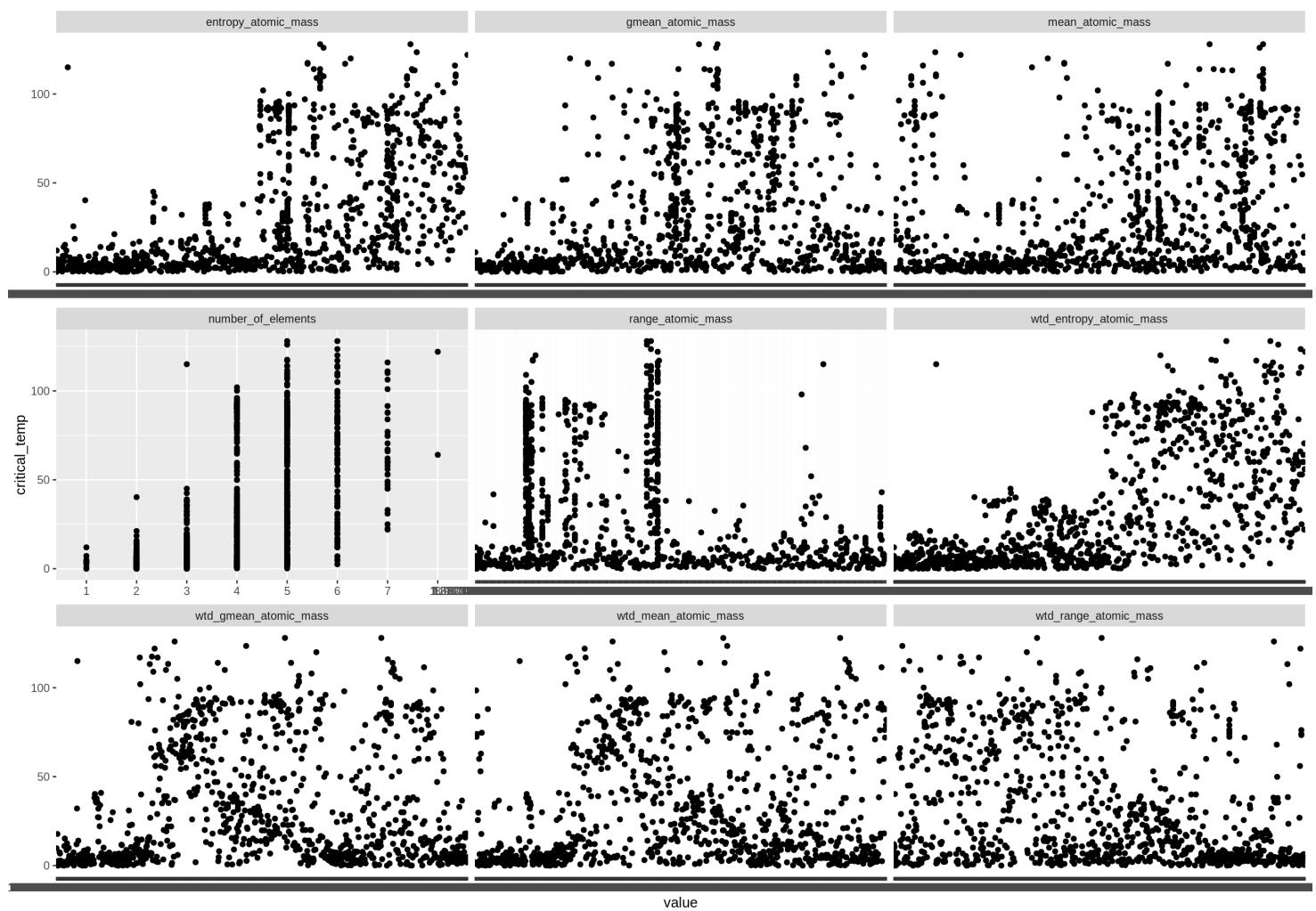








Scatterplot (by `critical_temp`)



Page 1

