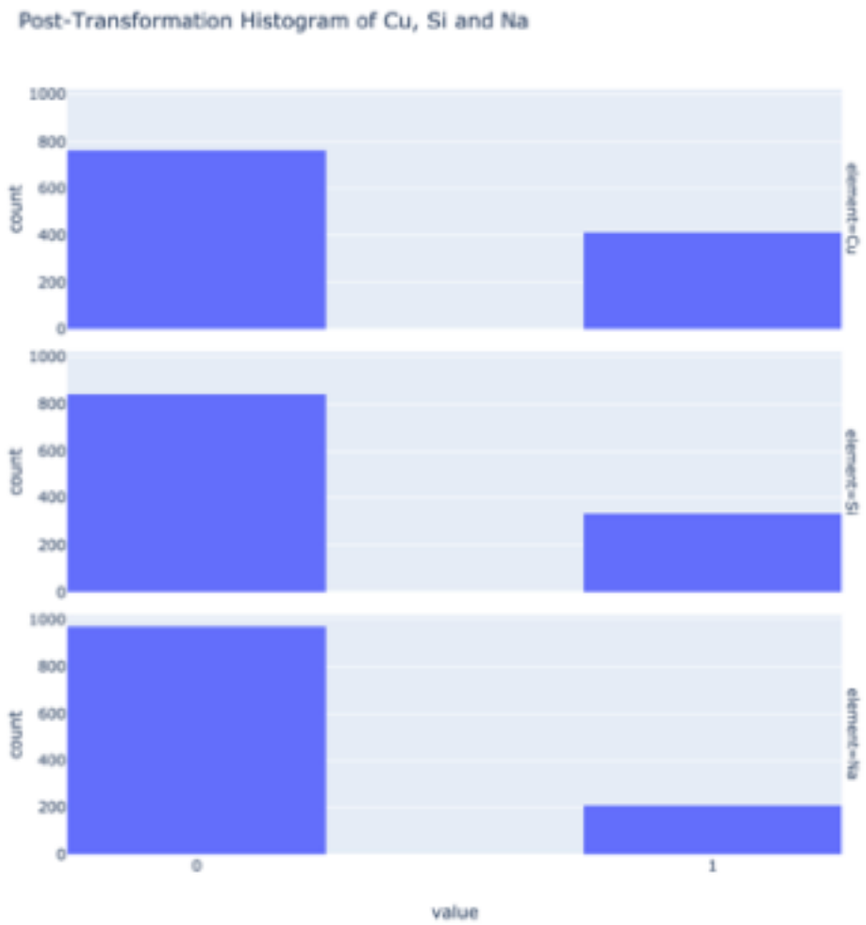
// transform skewed variables

before

after



Why aren’t they evenly distributed if you split them on the median?

…you SHAM

// fit the same logistic regression model

Binary LR model—engine.

Variable	Coefficient	Odds ratio	*p*-Value
Intercept	4.3712	NA	0.0000
Fe	−0.2520	0.7772	0.0000
Pb	−0.4496	0.6379	0.0005
OXI	−0.1033	0.9019	0.1310
Oilhours	0.0044	1.0044	0.0000
Cu	−0.7511	0.4719	0.0008
Si	−0.8569	0.4245	0.0002
Na	−0.7030	0.4950	0.0011
prevB	−1.4813	0.2274	0.0000
prevC	−2.0613	0.1273	0.0000
prevX	−2.0148	0.1333	0.0021

Phillips et al.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8974  -0.6020  -0.4018  -0.2943   2.6427

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.813e+00  2.819e-01 -13.525  < 2e-16 ***
Calc..Fluid.Hours 8.362e-05  3.635e-04   0.230   0.8181
Fe                1.115e-01  1.664e-02   6.704 2.03e-11 ***
Pb                2.611e-02  8.179e-02   0.319   0.7495
Cu_cat            4.673e-01  2.229e-01   2.096   0.0360 *
Si_cat            3.314e-01  2.125e-01   1.559   0.1189
Na_cat            1.574e-01  2.385e-01   0.660   0.5094
prev_evalb         1.329e+00  2.351e-01   5.656 1.55e-08 ***
prev_evalc        -3.317e-01  1.326e+00  -0.250   0.8025
prev_evalx         3.285e+00  1.433e+00   2.293   0.0219 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 851.33  on 939  degrees of freedom
Residual deviance: 723.67  on 930  degrees of freedom
AIC: 743.67

Number of Fisher Scoring iterations: 5
```

Troy et no-one.