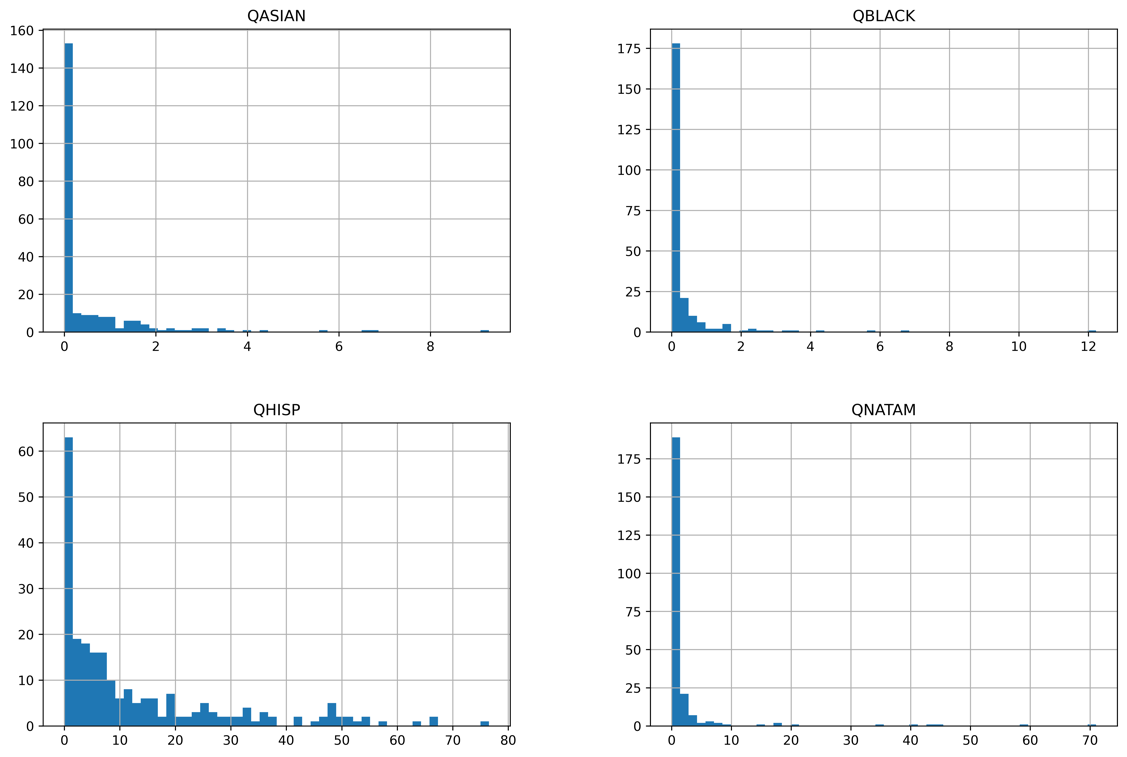
Data

1, social vulnerability data

to see the distribution of each variables, with the result of hist plot, we can see that the following 8 variables are heavy skewed, so we transfer these 8 variables to binary classification variables, 0 represent without, 1 represent with.



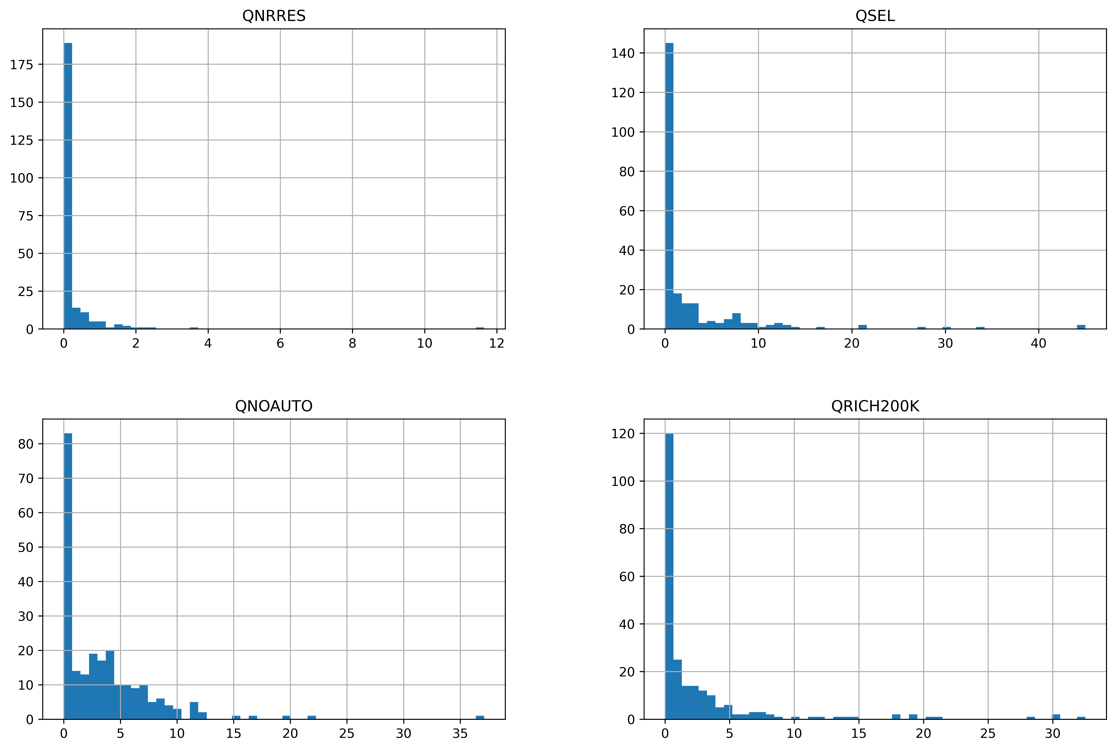
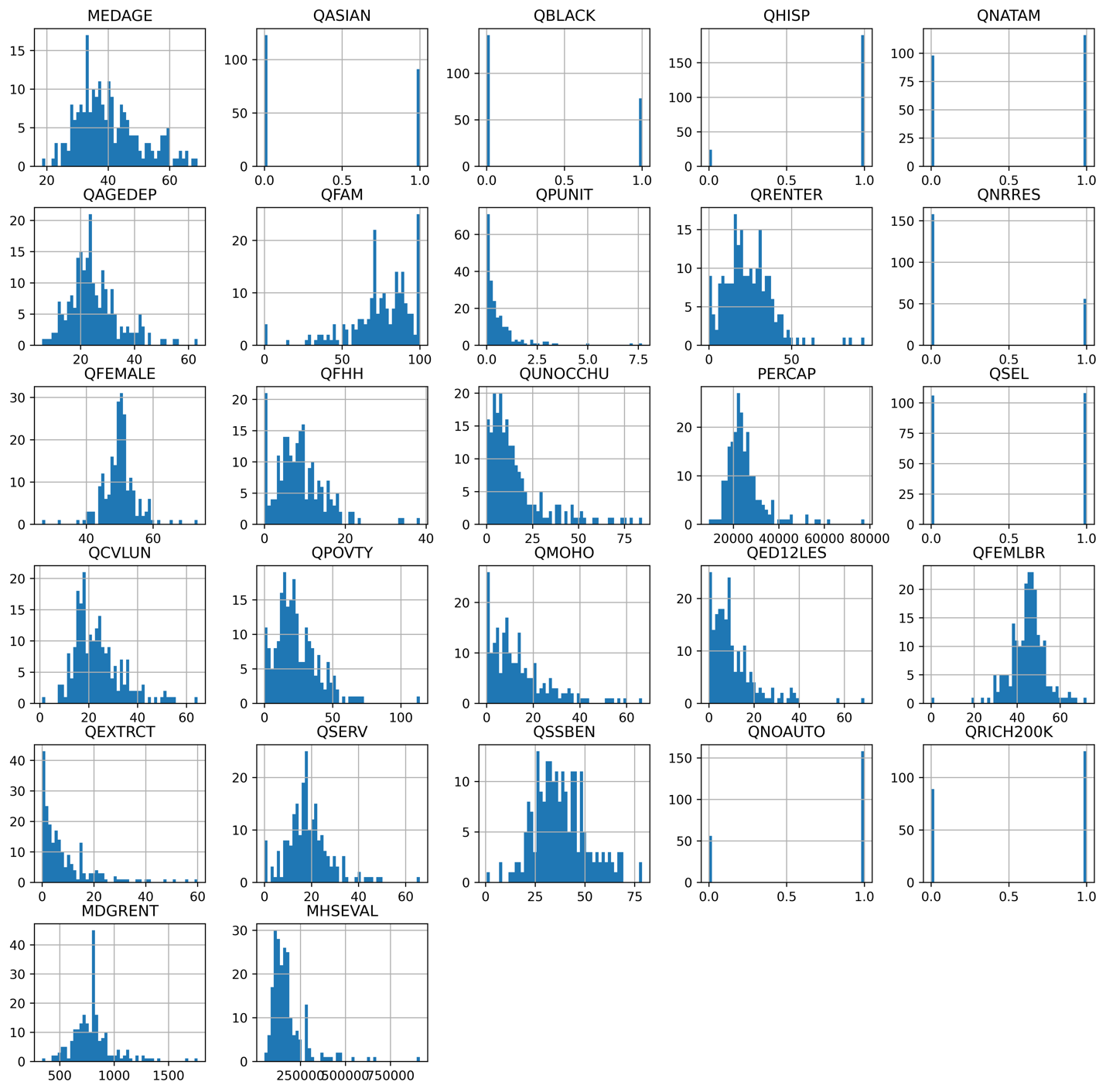


Figure 1: the distribution of four races and other four variables in communities(QASIAN, QBLACK, QHISP, QNATAM, QNRRES, QSEL, QNOAUTO, QRICH200K)

For that more than half of the percent of each race is race, we transfer these for variables into category variables, e.g., 0 represent no Asian, 1 represent with Asian.

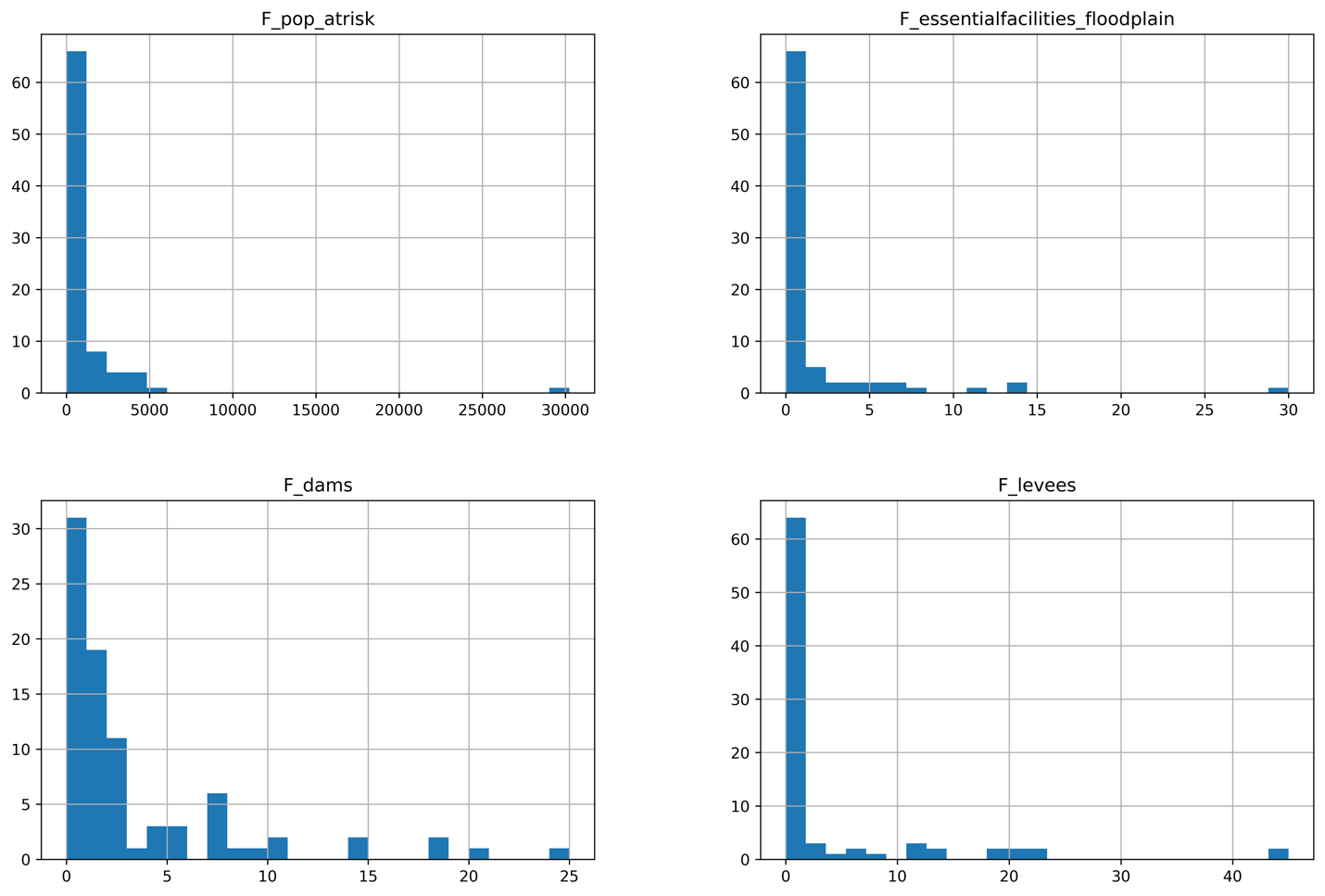
|  |  |  |
| --- | --- | --- |
| VARIABLE | 0 COUNTS | 1 COUNTS |
| QASIAN | 144 | 93 |
| QBLACK | 161 | 76 |
| QHISP | 42 | 195 |
| QNATAM | 119 | 118 |
| QNRRES | 178 | 59 |
| QSEL | 125 | 112 |
| QNOAUTO | 75 | 162 |
| QRICH200K | 105 | 135 |



As for the rest 19 variables, we do the standardization.

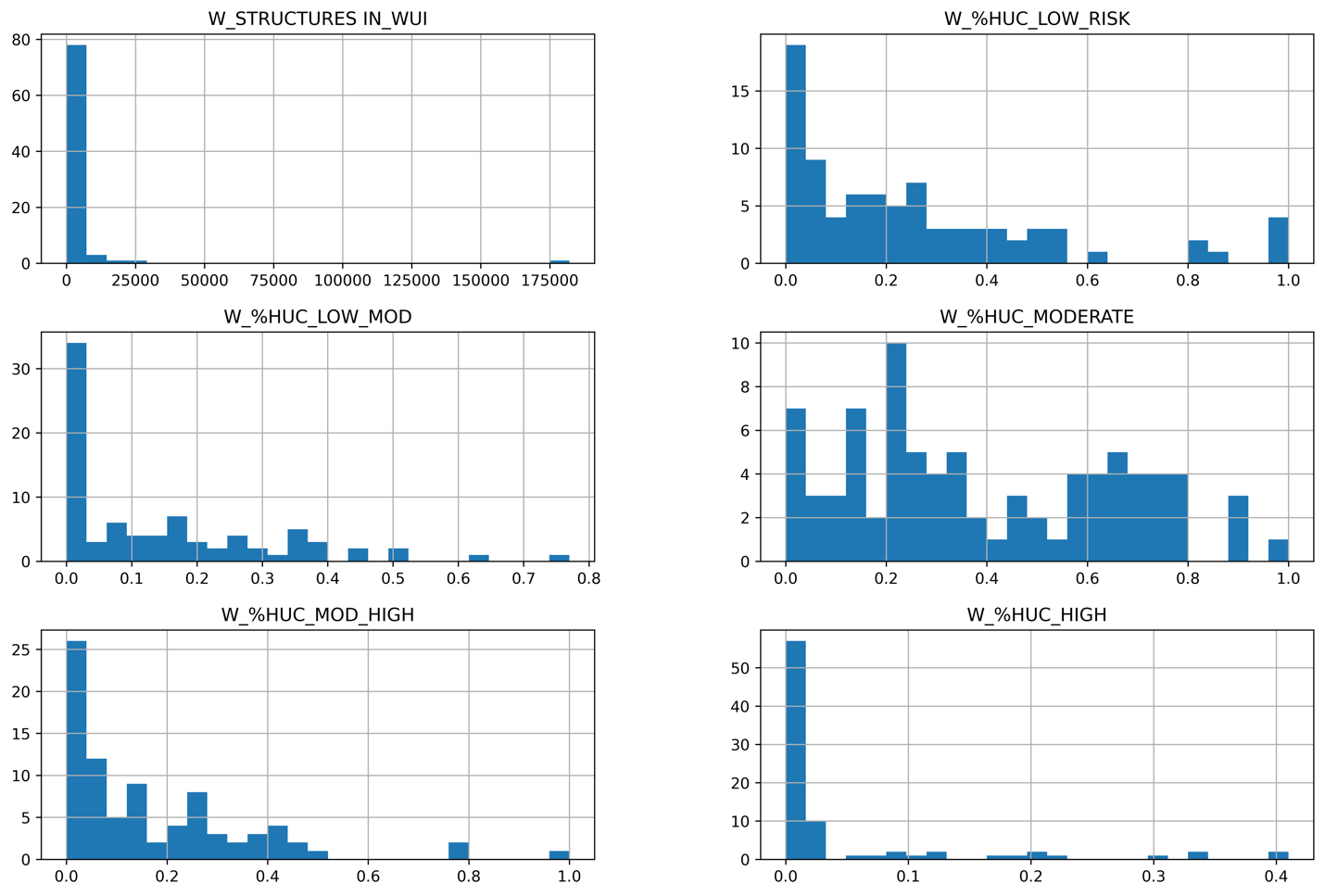
2, flood data

For flood hazards factors, we have four factors to be used in clustering. The distribution of these four factors are shown in the following figure. We also do the standardization for four factors.



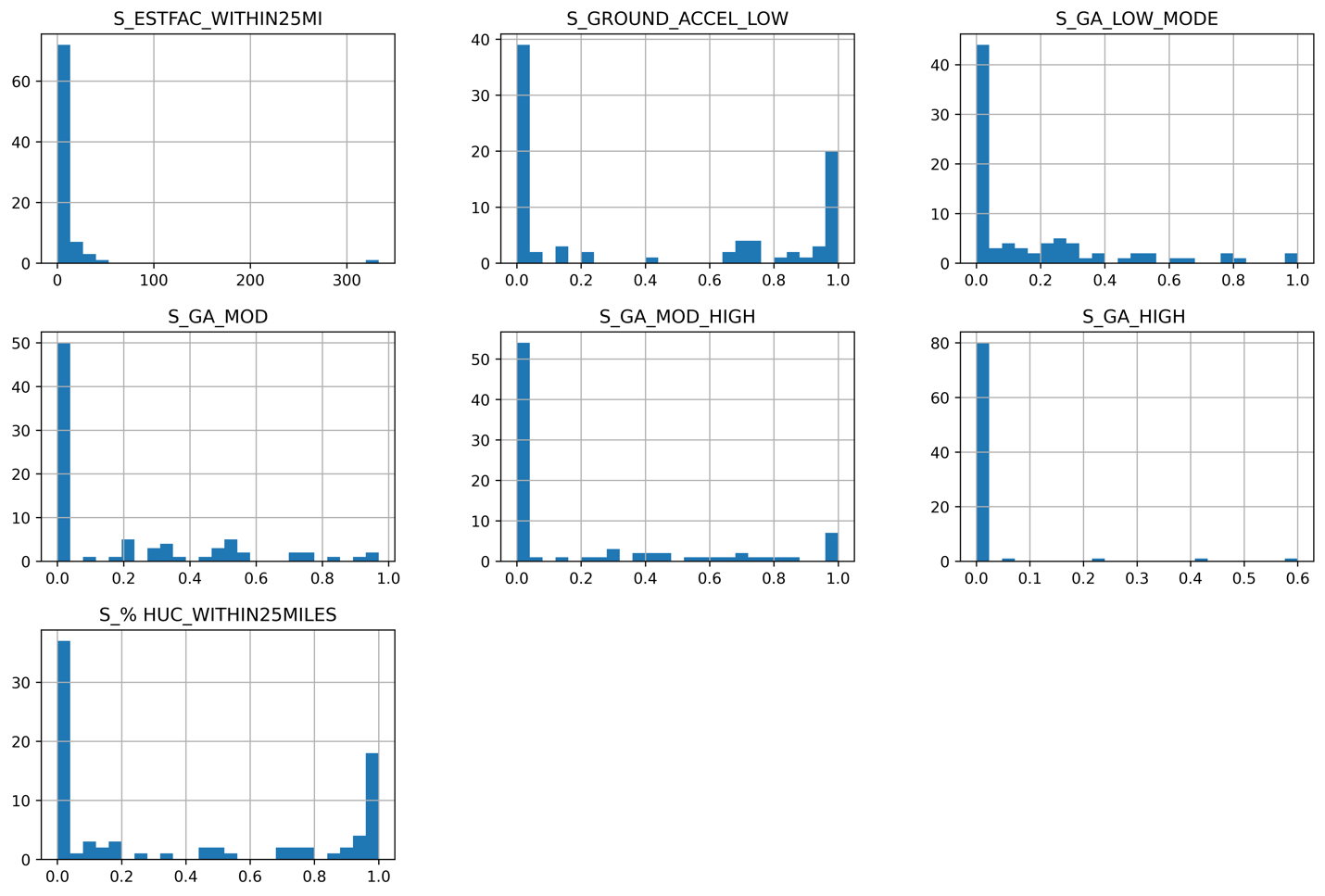
2, wildfire data

For wildfire hazards factors, we have six factors to be used in clustering. The distribution of these six factors are shown in the following figure. We also do the standardization for six factors.



2, seismic data

For seismic hazards factors, we have seven factors to be used in clustering. The distribution of these seven factors are shown in the following figure. We also do the standardization for seven factors.



Annotation:

Flood risk:

|  |  |
| --- | --- |
| value | level |
| 0 | Z(Zero) |
| 0-25% | L(low) |
| 25-50% | LM(low to median) |
| 50%-75 | MH(median to high) |
| 75-100% | H(High) |