

Exploração dos dados da tese de doutorado da Vivian

Melina Leite

outubro de 2015

Tabela de dados

```
'data.frame': 27 obs. of 18 variables:
 $ pH : num 7.58 7.97 8.23 7.91 7.95 7.95 8.28 8.11 8.04 8.5 ...
 $ EC : num 15.4 17.4 25 30 19.4 29 28.8 23.6 35.8 64.4 ...
 $ OD : num 6.67 6.65 7.05 5.58 7.01 ...
 $ TDS : num 7.6 8.6 12.8 15 9.8 14.6 14.4 12 18 32.2 ...
 $ Sol : num 0.16 0.43 1.76 3.67 0.25 ...
 $ Turb : num 0.984 0.779 2.269 6.707 1.341 ...
 $ Temp : num 19.8 18.9 19.6 19.7 18.6 ...
 $ ColF : num 32 64 240 496 32 ...
 $ ColT : num 512 1072 3328 2368 720 ...
 $ PT : num 0 0 0 0.003 0 0 0.028 0.004 0.026 0.003 ...
 $ Nitri: num 0 0 0.001 0.003 0.001 0.003 0.004 0.001 0.012 0.004 ...
 $ Nitra: num 0.079 0.099 0.084 0.065 0.06 0.034 0.083 0.083 0.121 0.05 ...
 $ NumP : int 11 10 10 32 28 9 29 18 25 16 ...
 $ MPS : num 34.8 31.48 45.65 7.37 22.71 ...
 $ ED : num 1.66 1.36 1.85 3.5 3.71 ...
 $ TE : num 24706 20262 27498 52096 55280 ...
 $ MPE : num 2246 2026 2750 1628 1974 ...
 $ MSI : num 3.48 2.81 1.62 1.85 1.97 ...
```

pH	EC	OD	TDS
Min. :7.580	Min. :15.40	Min. :2.096	Min. : 7.60
1st Qu.:7.860	1st Qu.:27.10	1st Qu.:4.929	1st Qu.:13.60
Median :8.040	Median :36.00	Median :5.624	Median :18.00
Mean :8.070	Mean :41.26	Mean :5.633	Mean :20.62
3rd Qu.:8.265	3rd Qu.:52.20	3rd Qu.:6.648	3rd Qu.:26.00
Max. :8.620	Max. :97.20	Max. :9.066	Max. :48.80

Sol	Turb	Temp	ColF
Min. : 0.160	Min. : 0.7794	Min. :17.54	Min. : 32
1st Qu.: 1.602	1st Qu.: 4.4532	1st Qu.:18.93	1st Qu.: 80
Median : 3.934	Median : 10.1361	Median :19.80	Median : 176
Mean : 7.105	Mean : 15.5832	Mean :19.91	Mean : 1646
3rd Qu.: 6.270	3rd Qu.: 13.9608	3rd Qu.:20.87	3rd Qu.: 344
Max. :78.511	Max. :160.0335	Max. :22.10	Max. :19408

ColT	PT	Nitri	Nitra
Min. : 288	Min. :0.00000	Min. :0.000000	Min. :0.01500
1st Qu.: 1080	1st Qu.:0.00000	1st Qu.:0.001000	1st Qu.:0.05600
Median : 1664	Median :0.00600	Median :0.004000	Median :0.07900
Mean : 5441	Mean :0.01315	Mean :0.005963	Mean :0.08785
3rd Qu.: 3056	3rd Qu.:0.01700	3rd Qu.:0.006000	3rd Qu.:0.09400
Max. :60000	Max. :0.07800	Max. :0.067000	Max. :0.35800

NumP	MPS	ED	TE
Min. : 6.00	Min. : 7.014	Min. : 0.784	Min. : 11684
1st Qu.:12.00	1st Qu.:15.400	1st Qu.: 1.831	1st Qu.: 27245

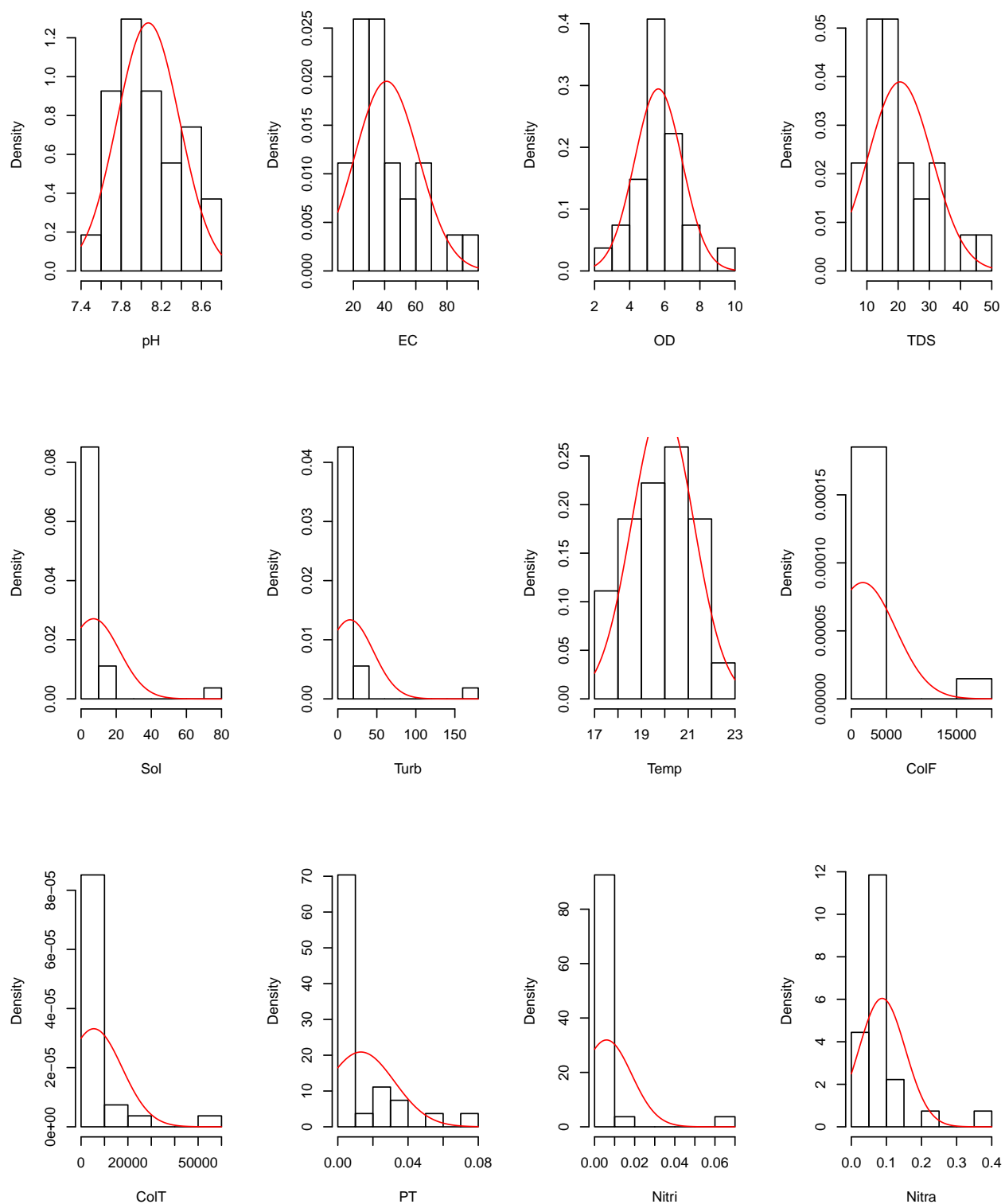
Median :22.00	Median :22.237	Median : 2.995	Median : 44640
Mean :24.48	Mean :23.088	Mean : 3.767	Mean : 56138
3rd Qu.:29.50	3rd Qu.:29.258	3rd Qu.: 3.769	3rd Qu.: 56180
Max. :92.00	Max. :45.650	Max. :15.889	Max. :236808
MPE	MSI		
Min. :1488	Min. :1.595		
1st Qu.:1881	1st Qu.:1.734		
Median :2294	Median :1.937		
Mean :2254	Mean :2.436		
3rd Qu.:2593	3rd Qu.:2.667		
Max. :3026	Max. :5.598		

Separando as variáveis de água e de paisagem para facilitar análises.

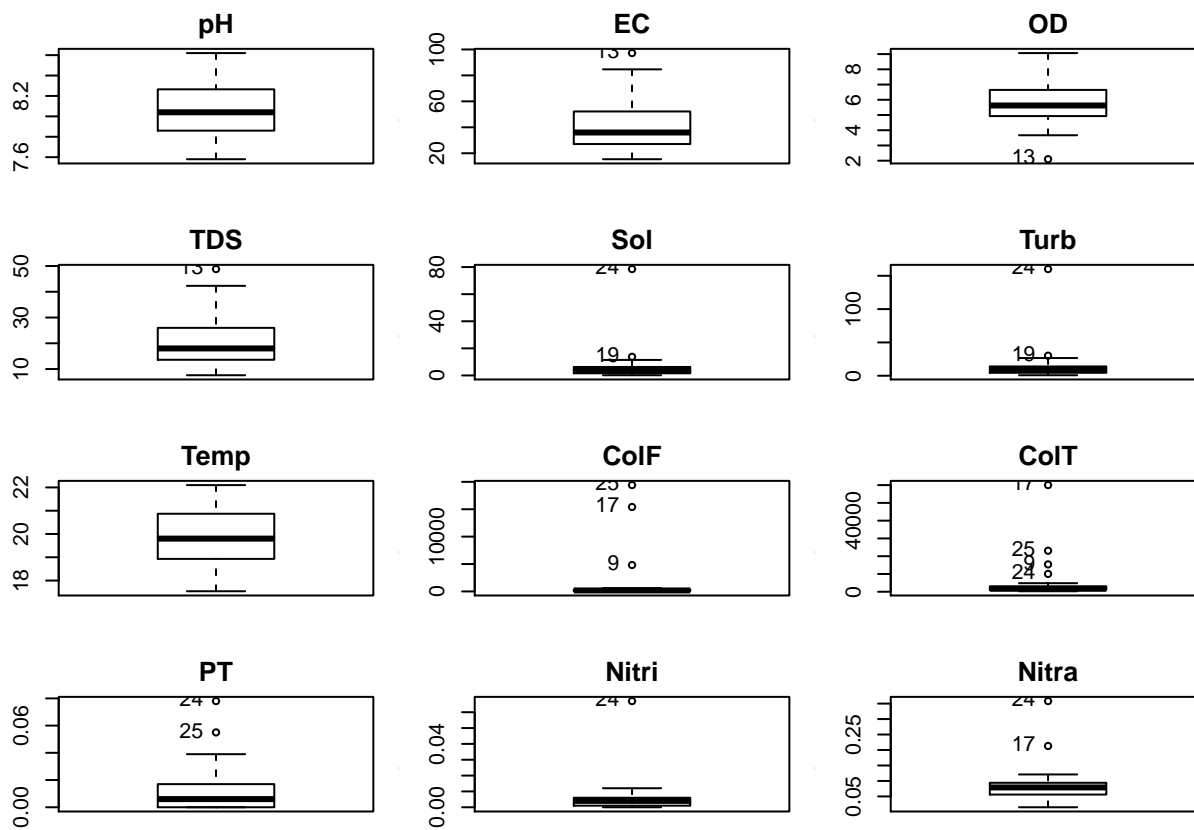
Gráficos exploratórios das variáveis brutas

variáveis da água

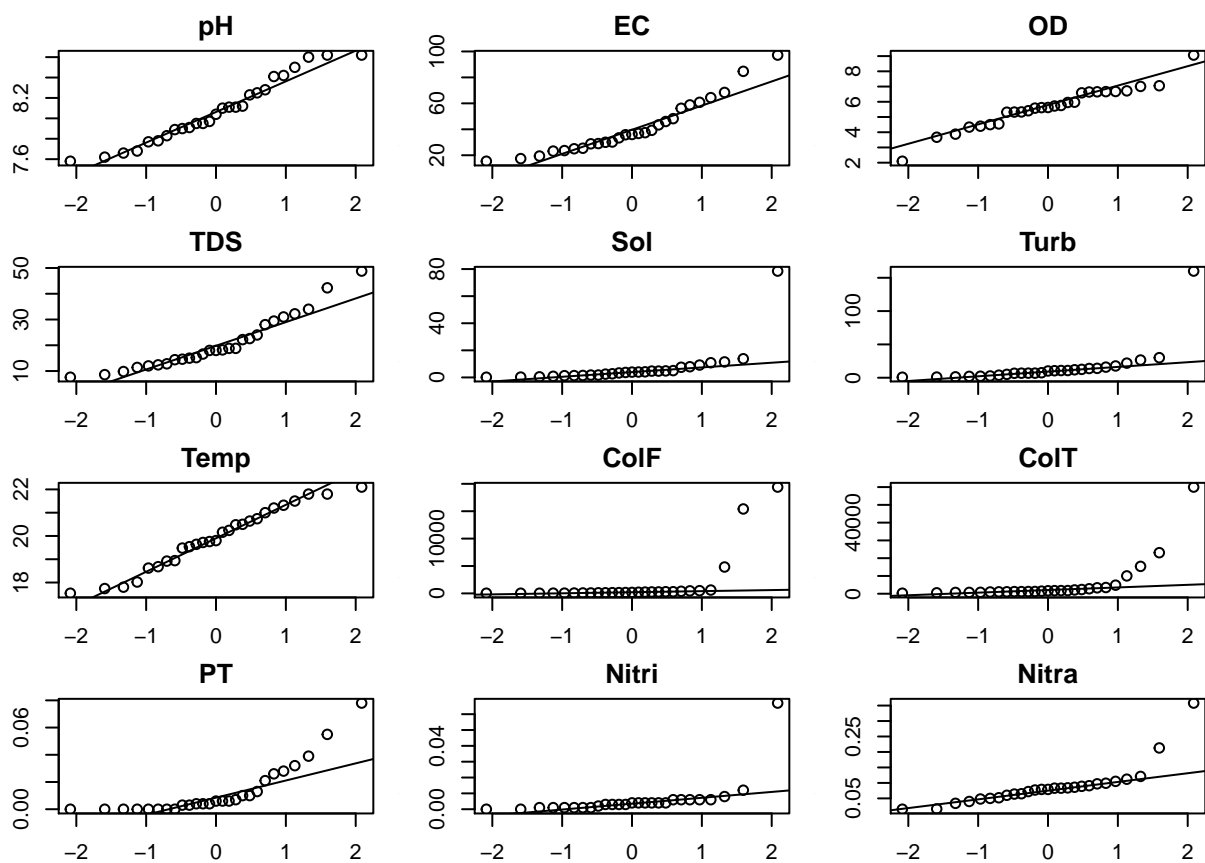
Histograma + density (como se fosse normal)



Boxplot identificando os outliers



QQplot para ver normalidade

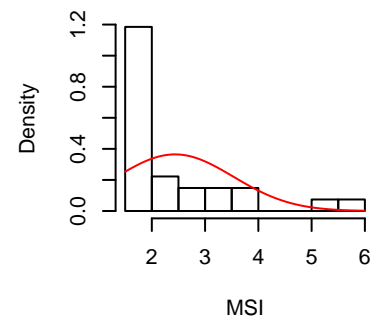
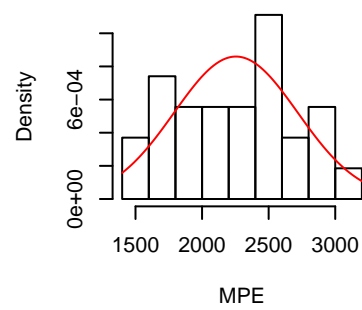
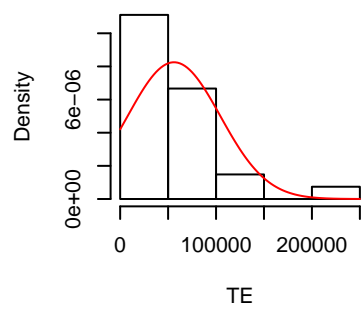
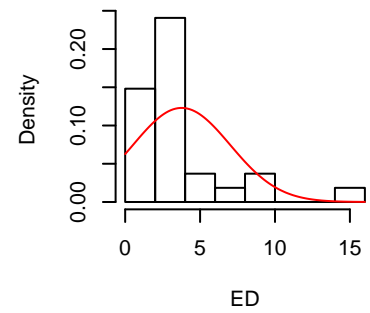
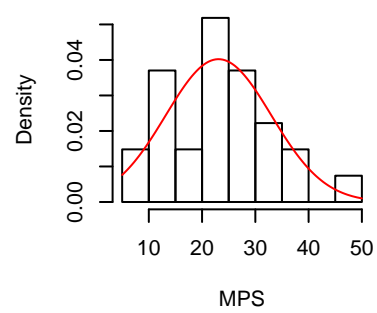
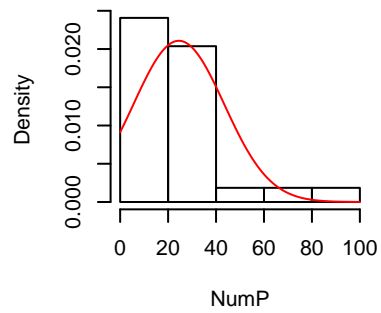


null device

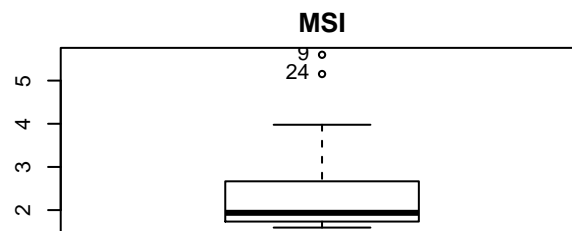
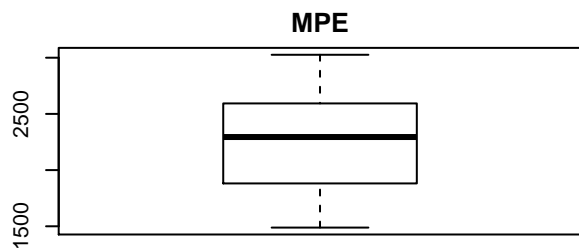
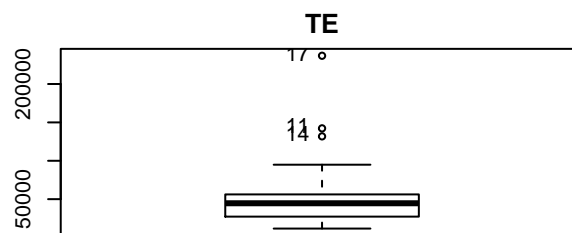
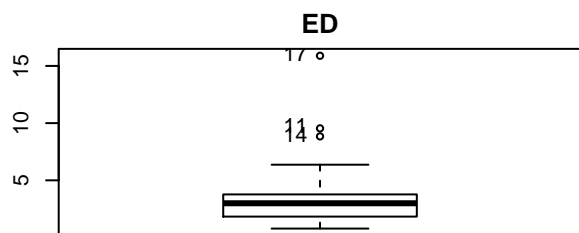
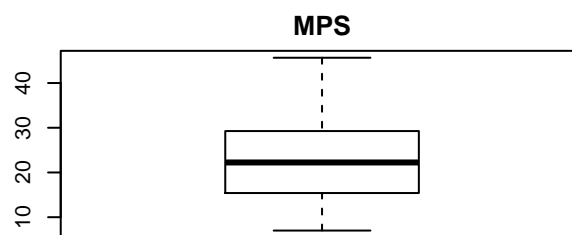
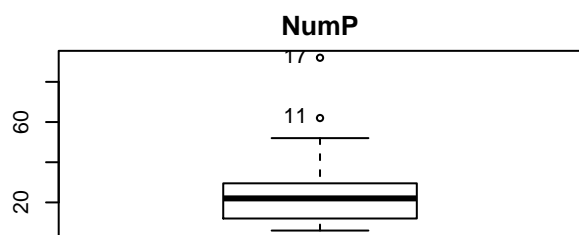
1

variáveis da paisagem

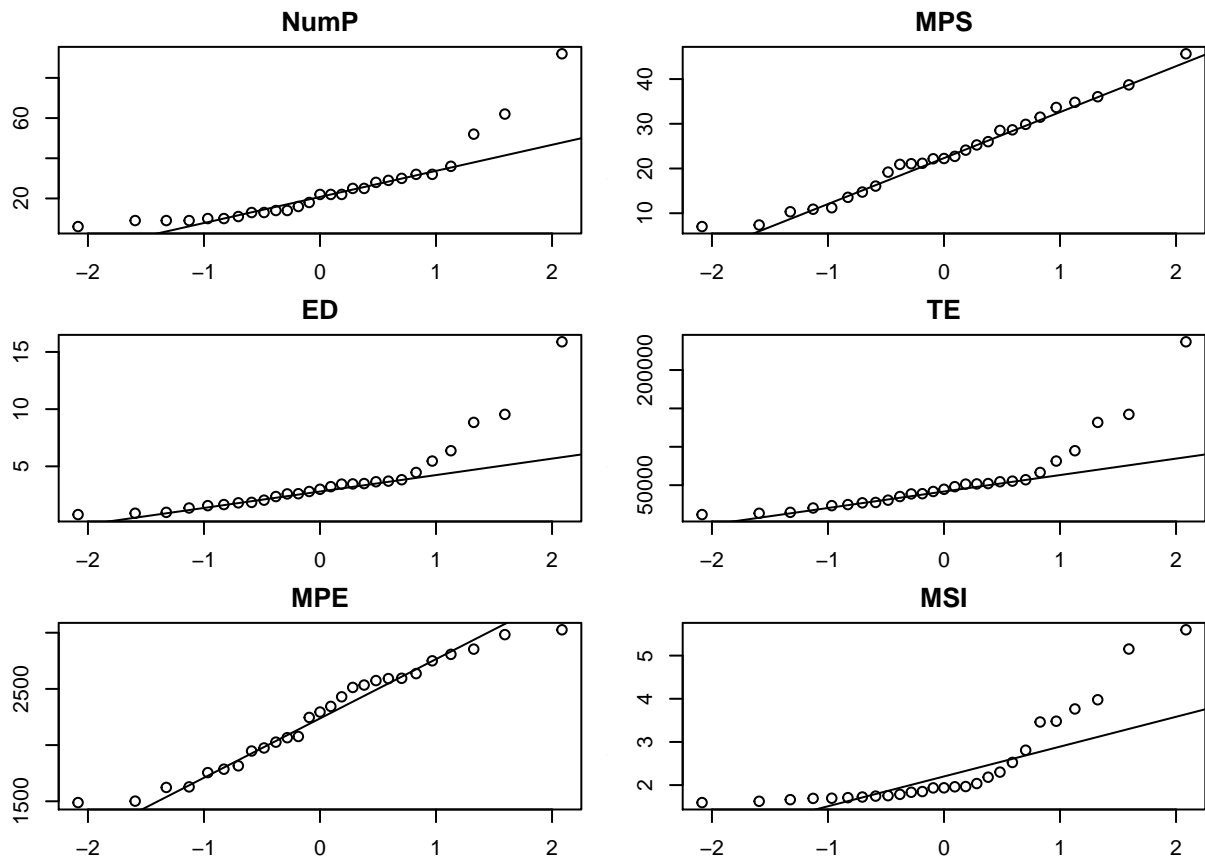
Histograma + density (como se fosse normal)



Boxplot identificando os outliers



QQplot para ver normalidade



Testes de premissas SEM

Valores de skewness e kurtosis

	skew	Pskew	kurt	Pkurt
pH	0.306	0.447	2.092	0.241
EC	1.090	0.015	3.639	0.215
OD	-0.185	0.644	4.062	0.110
TDS	1.099	0.014	3.661	0.208
Sol	4.449	0.000	22.133	0.000
Turb	4.419	0.000	21.950	0.000
Temp	-0.181	0.651	2.073	0.220
ColF	3.154	0.000	11.441	0.000
ColT	3.773	0.000	17.130	0.000
PT	1.982	0.000	6.520	0.004
Nitri	4.546	0.000	22.759	0.000
Nitra	2.793	0.000	11.890	0.000
NumP	2.035	0.000	7.425	0.002
MPS	0.243	0.545	2.473	0.764
ED	2.292	0.000	8.577	0.001
TE	2.292	0.000	8.576	0.001
MPE	-0.076	0.848	1.839	0.047
MSI	1.641	0.001	4.703	0.042

Dados com muitas variáveis com skewness e kurtosis significativos. Isso deve estar ocorrendo principalmente devido aos outliers.

Normalidade de cada variável separadamente

```
$`Shapiro-Wilk's Normality Test`  
  Variable Statistic    p-value Normality  
1    pH          0.9528    0.2508    YES  
2    EC          0.9036    0.0162    NO  
3    OD          0.9562    0.3021    YES  
4    TDS         0.9018    0.0147    NO  
5    Sol         0.3948    0.0000    NO  
6    Turb        0.4057    0.0000    NO  
7    Temp        0.9652    0.4816    YES  
8    ColF        0.3795    0.0000    NO  
9    ColT        0.4308    0.0000    NO  
10   PT          0.7154    0.0000    NO  
11   Nitri       0.3699    0.0000    NO  
12   Nitra       0.6853    0.0000    NO  
13   NumP        0.7776    0.0001    NO  
14   MPS         0.9776    0.8053    YES  
15   ED          0.7379    0.0000    NO  
16   TE          0.7379    0.0000    NO  
17   MPE         0.9559    0.2966    YES  
18   MSI         0.7335    0.0000    NO
```

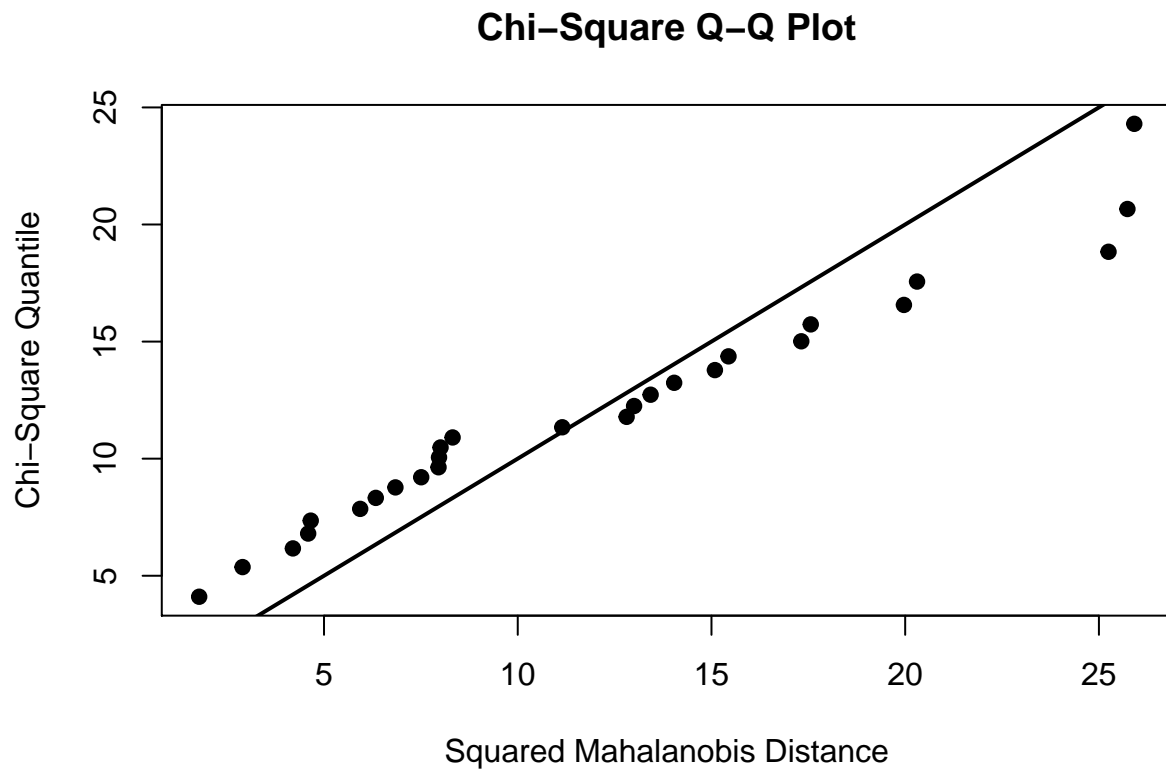
```
null device  
      1
```

Como era de se esperar pela visualização dos gráficos.

Teste de distribuição normal multivariada

Teste de Mardia recomendado no Shipley (2004) usando pacote [MVN](#). Como o teste só roda com até 15 variáveis, separei as de água e de paisagem:

água



Mardia's Multivariate Normality Test

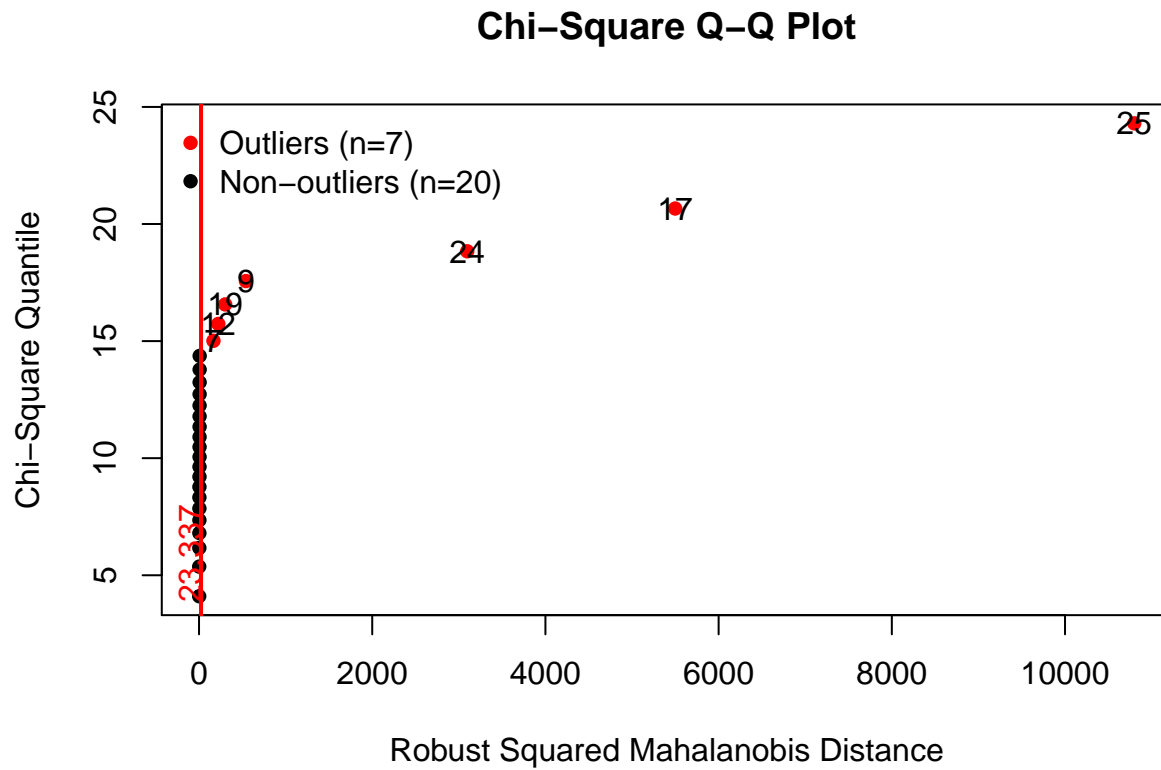
data : da[1:12]

g1p : 132.1651
chi.skew : 594.7431
p.value.skew : 2.299536e-13

g2p : 192.6751
z.kurtosis : 3.497368
p.value.kurt : 0.000469873

chi.small.skew : 671.9009
p.value.small : 1.338575e-20

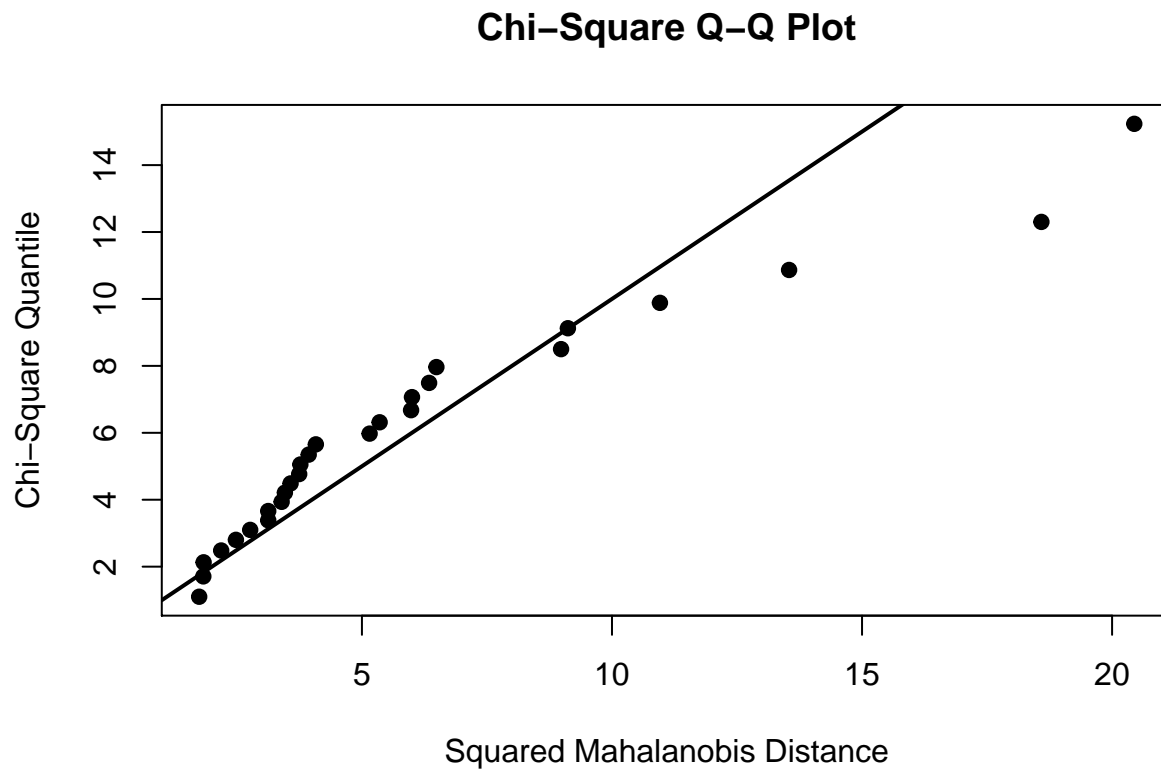
Result : Data are not multivariate normal.



```
$outlier
```

Observation	Mahalanobis Distance	Outlier
25	10799.844	TRUE
17	5498.571	TRUE
24	3097.240	TRUE
9	542.016	TRUE
19	302.117	TRUE
12	222.875	TRUE
7	166.787	TRUE
21	7.596	FALSE
11	7.371	FALSE
13	7.096	FALSE
15	6.994	FALSE
27	6.686	FALSE
20	6.636	FALSE
22	6.464	FALSE
26	6.432	FALSE
3	6.428	FALSE
14	6.281	FALSE
18	5.638	FALSE
6	5.273	FALSE
8	4.792	FALSE
16	4.375	FALSE
4	4.311	FALSE
10	4.296	FALSE
2	4.125	FALSE
1	3.529	FALSE
23	2.196	FALSE
5	1.882	FALSE

paisagem



Mardia's Multivariate Normality Test

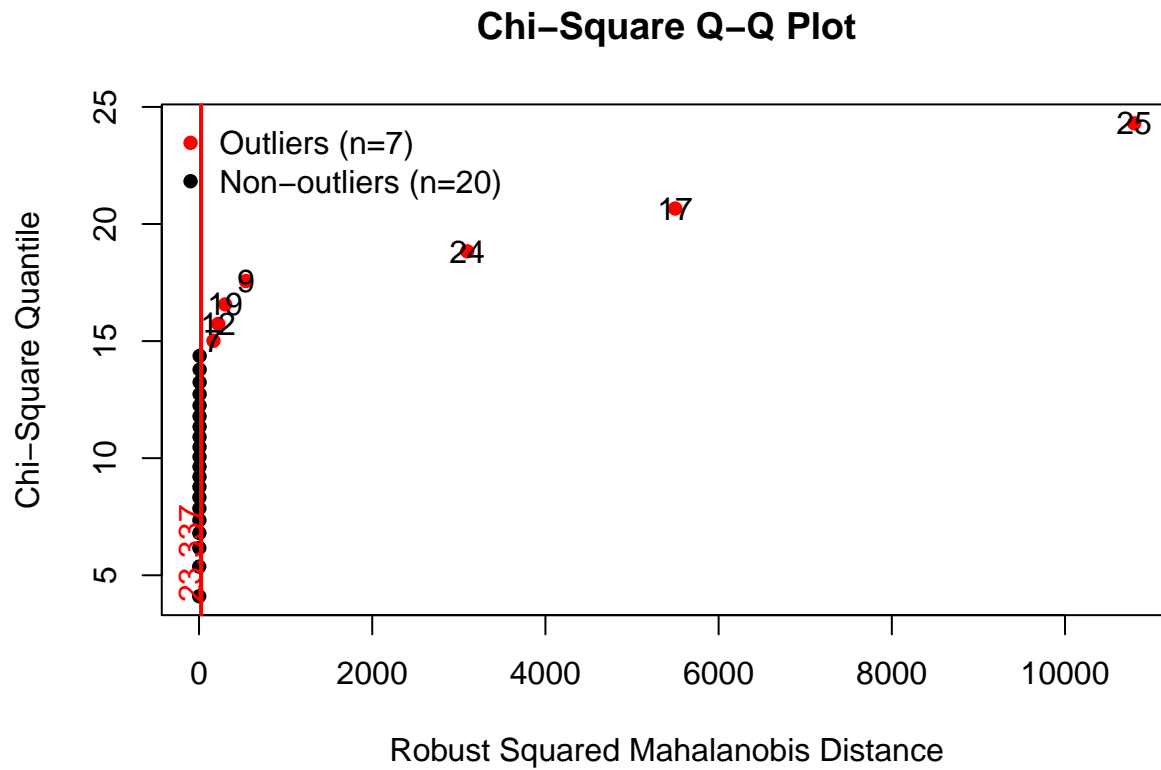
data : da[13:18]

g1p : 33.31189
chi.skew : 149.9035
p.value.skew : 1.660335e-10

g2p : 58.68711
z.kurtosis : 2.833847
p.value.kurt : 0.004599136

chi.small.skew : 171.8192
p.value.small : 1.078246e-13

Result : Data are not multivariate normal.



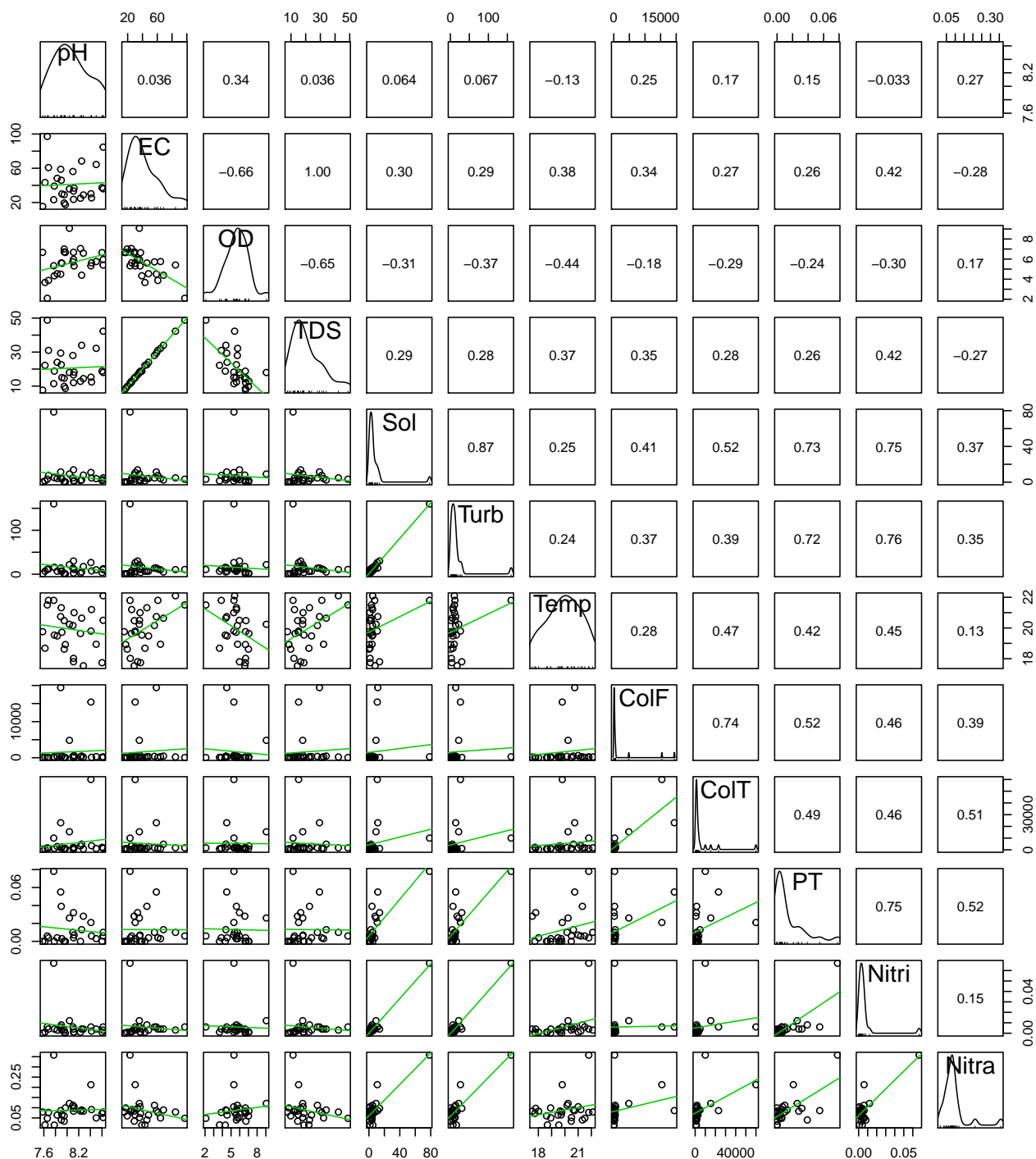
```
$outlier
```

Observation	Mahalanobis Distance	Outlier
25	10799.844	TRUE
17	5498.571	TRUE
24	3097.240	TRUE
9	542.016	TRUE
19	302.117	TRUE
12	222.875	TRUE
7	166.787	TRUE
21	7.596	FALSE
11	7.371	FALSE
13	7.096	FALSE
15	6.994	FALSE
27	6.686	FALSE
20	6.636	FALSE
22	6.464	FALSE
26	6.432	FALSE
3	6.428	FALSE
14	6.281	FALSE
18	5.638	FALSE
6	5.273	FALSE
8	4.792	FALSE
16	4.375	FALSE
4	4.311	FALSE
10	4.296	FALSE
2	4.125	FALSE
1	3.529	FALSE
23	2.196	FALSE
5	1.882	FALSE

Correlação entre as variáveis

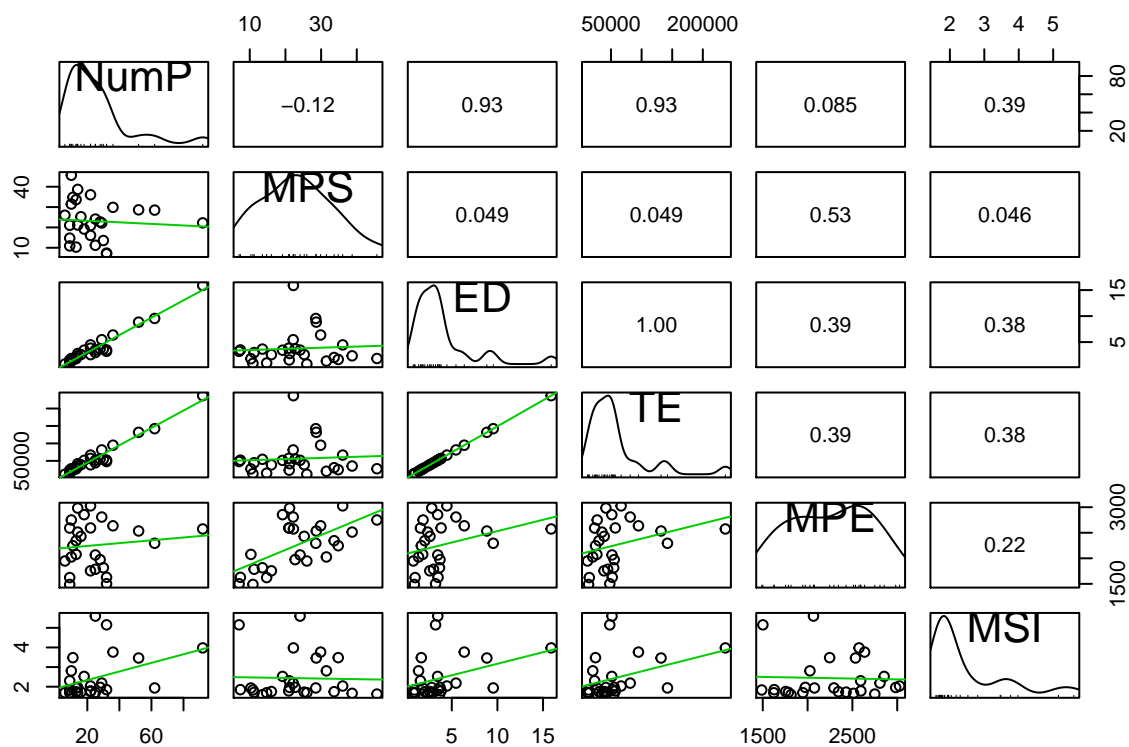
Lembrando que os valores destas correlações levam em consideração os outliers. e vendo as linhas verdes, eles estão puxando muito os valores das correlações.

Água



Comentários: correlação igual a 1 entre EC e TDS.

Paisagem



Comentários: correlação igual a 1 entre ED e TE.