Exploratory Analysis of Meron Data

 # read in data

```
x   y   tree.64   alt   slope   ns.aspect   ew.aspect   stream.dist

road.dist   graze.type   graze.intensity   water.dist   tree.92
```

> # check frequencies of graze.type x graze.intensity

	graze.intensity			
graze.type	0	1	2	3
0	6	0	0	0
1	0	92	68	9
2	0	3	22	0

Only 6 pixels with no grazing, so perhaps should not consider this variable. For goats (type=2), confounding with intensity (3 vs 22). For cattle (type=1), no confounding with intensity (though should redefine levels as low and med/high). Try the following classification (others are possible):

- **no grazing**
- **goat grazing**
- **cattle grazing at low intensity**
- **cattle grazing at med/high intensity**

 # define appropriate indicator variables

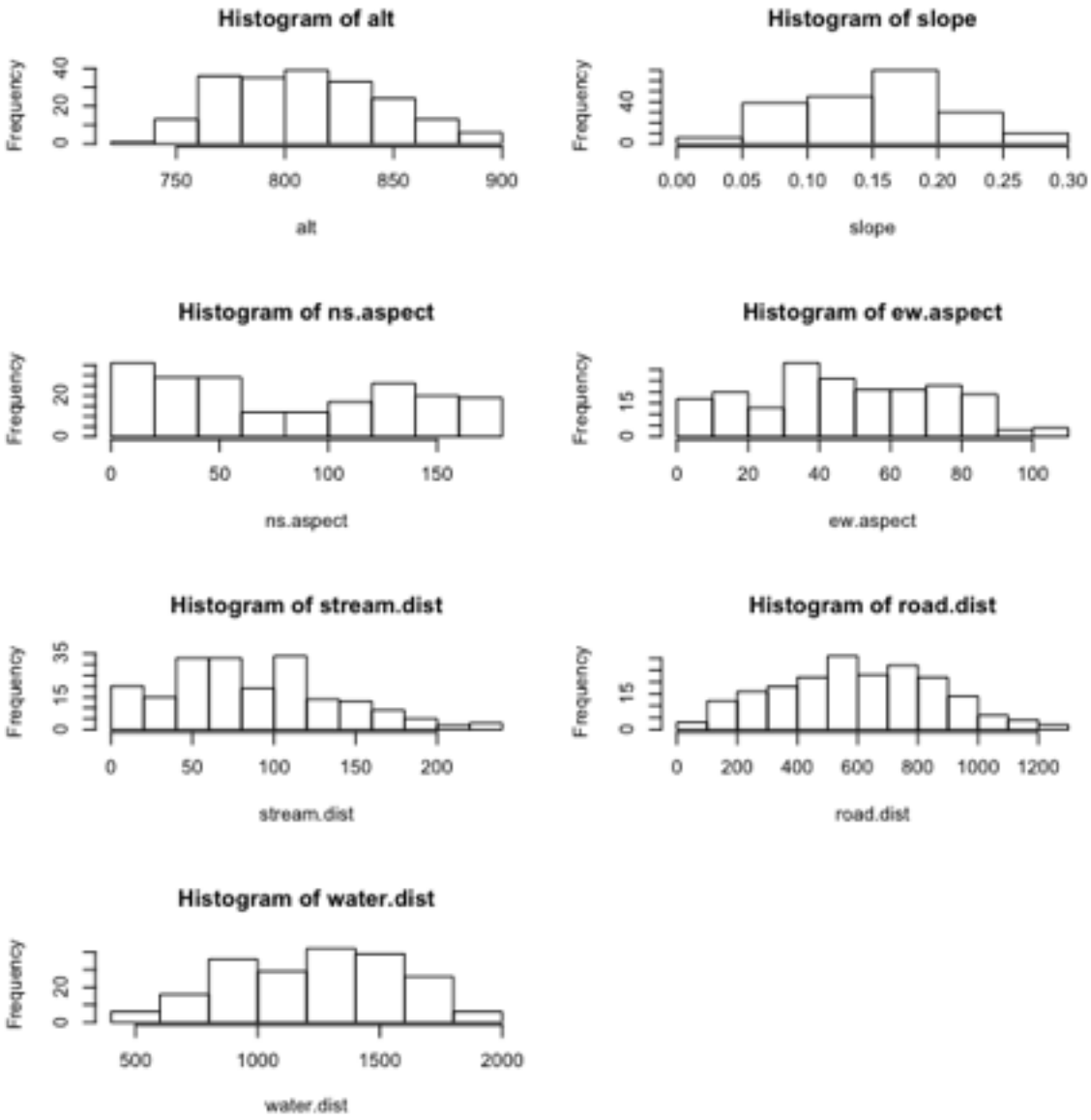
```
goats = 1 * (graze.type == 2) # so will be a numeric variable
cattle.low = (graze.type == 1)*(graze.intensity ==1)
cattle.med.high = (graze.type == 1)*(graze.intensity >= 2)
```

 # correlation matrix of continuous variables

	alt	slope	ns.aspect	ew.aspect	stream.dist	road.dist	water.dist
alt	1.00	-0.06	-0.01	-0.21	0.34	-0.73	0.02
slope	-0.06	1.00	-0.19	0.37	-0.25	0.03	0.22
ns.aspect	-0.01	-0.19	1.00	0.05	0.02	0.09	0.06
ew.aspect	-0.21	0.37	0.05	1.00	0.05	0.17	0.36
stream.dist	0.34	-0.25	0.02	0.05	1.00	0.03	0.26
road.dist	**-0.73**	0.03	0.09	0.17	0.03	1.00	0.20
water.dist	0.02	0.22	0.06	0.36	0.26	0.20	1.00

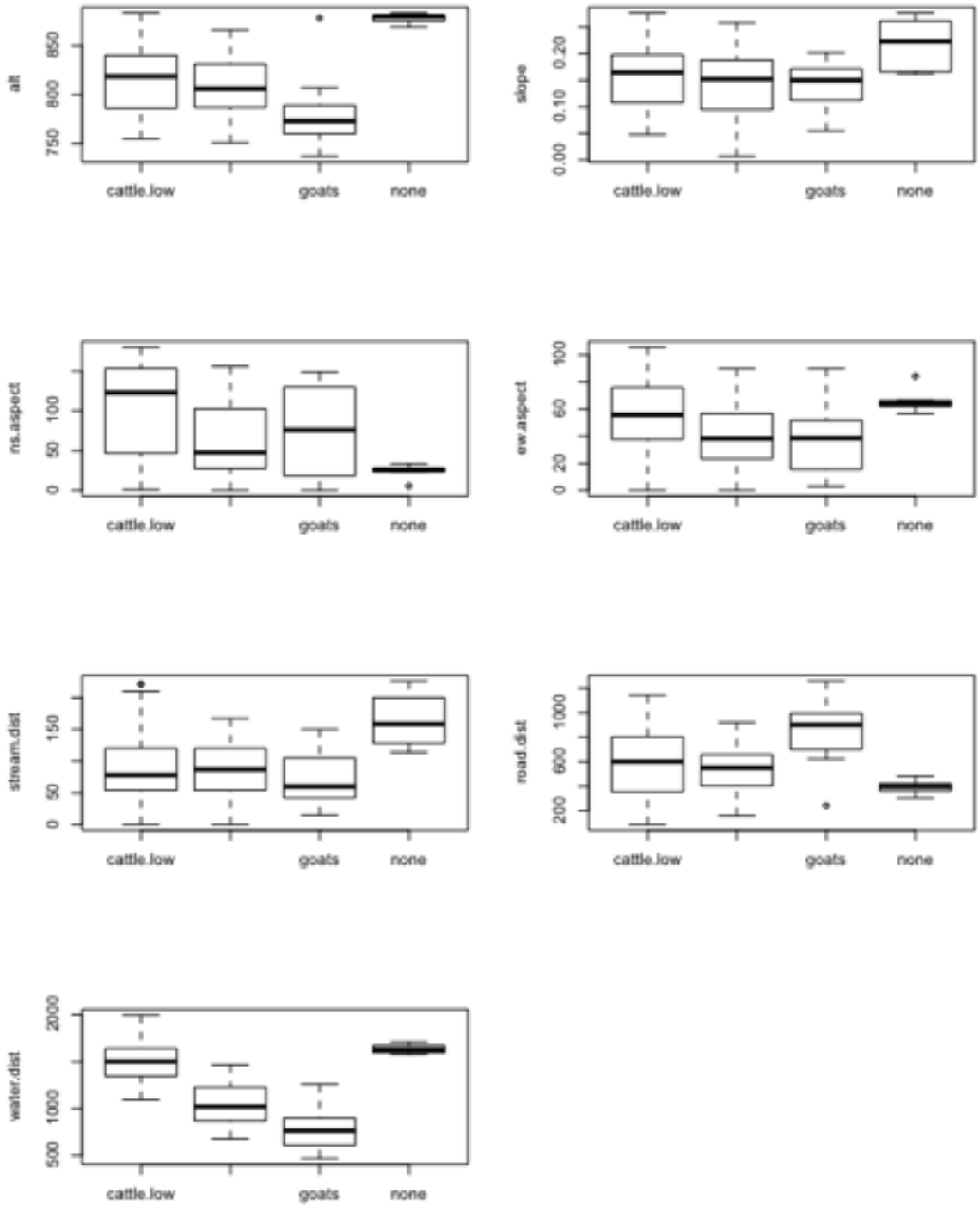
 # no problems

 # marginal distributions of continuous variables



 # shows nice distribution of values. note that the slope is at most 30%.

 # relation between the grazing variable and the continuous variables



 # No grazing when the altitude, slope or water distance are high; and when the ns.aspect or road distance are low. When there is, the type of grazing is not affected by the slope, aspect or stream distance. The goats are usually at a lower altitude and smaller water distance than the cattle.

relation between explanatory variables and response
relation between tree.92 and tree.64

```
          tree.92
tree.64    0   1
      0   68 132
```

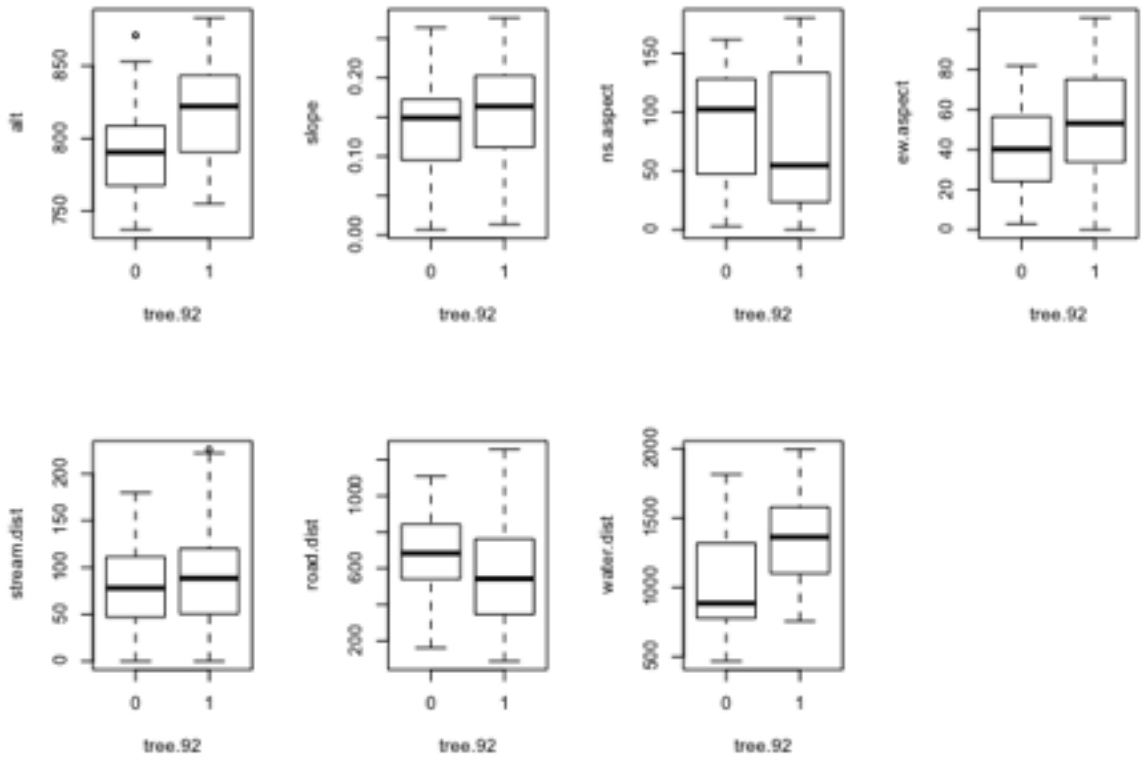
 # all of these pixels had no trees in 1964, so we can't use tree.64

> # relation between tree.92 and grazing

```
                   tree.92
grazing             0  1
  cattle.low        8 84
  cattle.med.high  40 37
  goats            20  5
  none              0  6
```

 # appears that goats are bad for the trees, and a little bit of cows are good (?)

 # relation between tree.92 and the continuous variables



More trees at a higher altitude and lower road distance (fewer people) and at a higher water.dist (fewer goats and cattle). Fewer trees with a high N/S aspect (N=0; more sun from the south), and more with a high E/W aspect (E=0; more moisture from the west).

Compute a linear regression (even though tree.92 is 0/1). Include slope x aspect interaction because the effect of the aspect depends on the slope.

```
Call:
lm(formula = tree.92 ~ alt + stream.dist + road.dist + water.dist +
    slope + ns.aspect + ew.aspect + ns.aspect * slope + ew.aspect *
    slope + cattle.low + cattle.med.high + goats)

Residuals:
     Min       1Q   Median       3Q      Max
-1.15400 -0.23830  0.02867  0.23995  0.81359

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -5.5060718  1.3355106  -4.123 5.62e-05 ***
alt              0.0066339  0.0014086   4.709 4.83e-06 ***
stream.dist     -0.0020502  0.0006666  -3.075  0.00242 **
road.dist        0.0001016  0.0001829   0.556  0.57900
water.dist       0.0003433  0.0001803   1.904  0.05850 .
slope            1.0618379  1.2894984   0.823  0.41130
ns.aspect        0.0025722  0.0018368   1.400  0.16305
ew.aspect        0.0010591  0.0031985   0.331  0.74092
cattle.low       0.3890631  0.1695149   2.295  0.02283 *
cattle.med.high  0.1834058  0.2066002   0.888  0.37582
goats            0.1478609  0.2694239   0.549  0.58379
slope:ns.aspect -0.0261511  0.0102421  -2.553  0.01147 *
slope:ew.aspect  0.0130690  0.0193914   0.674  0.50117
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3466 on 187 degrees of freedom
Multiple R-squared:  0.4996,    Adjusted R-squared:  0.4674
F-statistic: 15.56 on 12 and 187 DF,  p-value: < 2.2e-16
```

 # more trees for high alt, low stream.dist, high water.dist, cattle grazing at low intensity(!). tree.92 also depends on slope and aspect.

 # Remark: we should also try adding the coordinate variables x, y, x^2, y^2 and x*y