Modelling oak recruitment at abandoned croplands

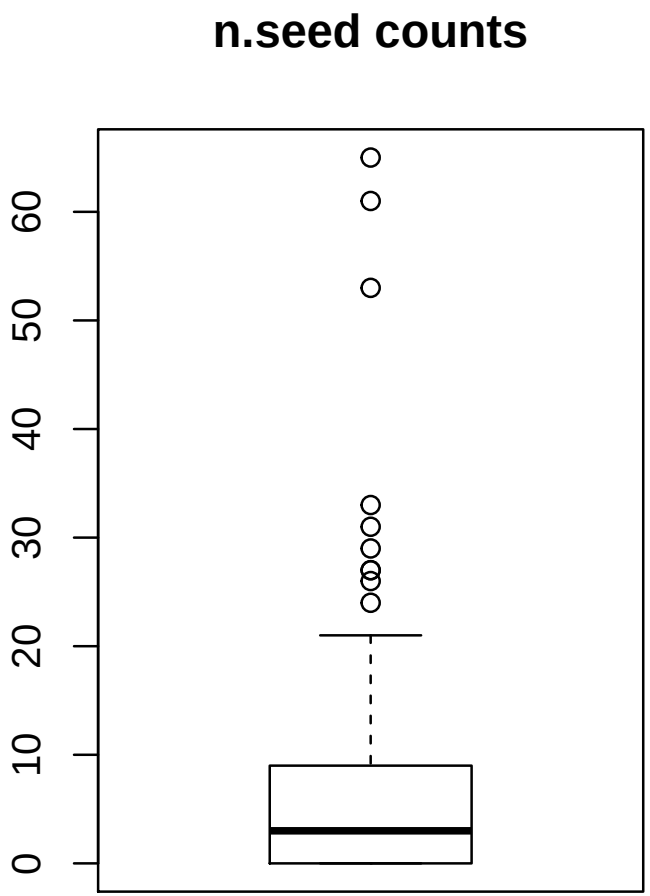
Antonio J. Perez-Luque (@ajpelu)

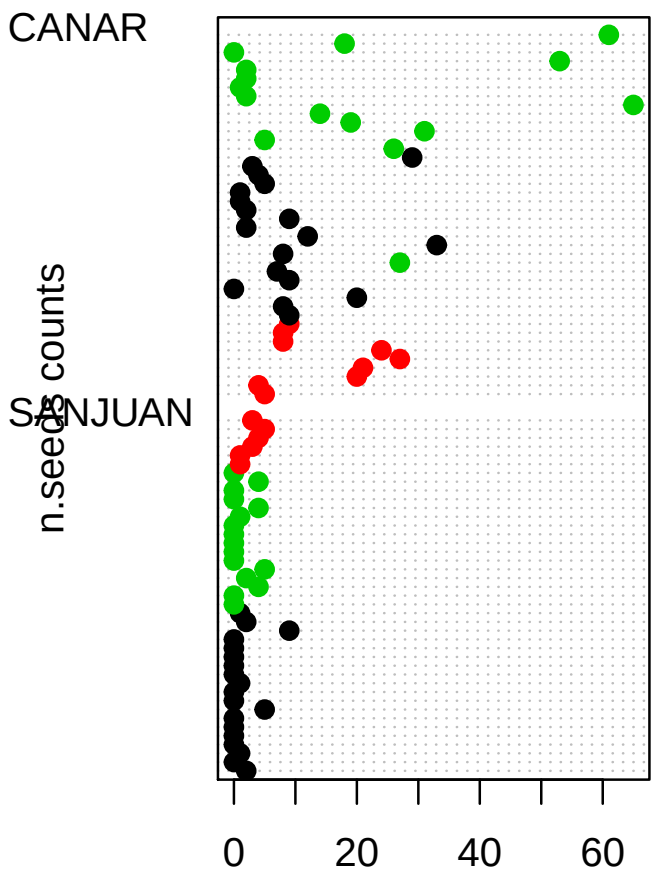
Aims

Model Oak seedlings (*n.seed*; heigth < 130 cm) as a function of:

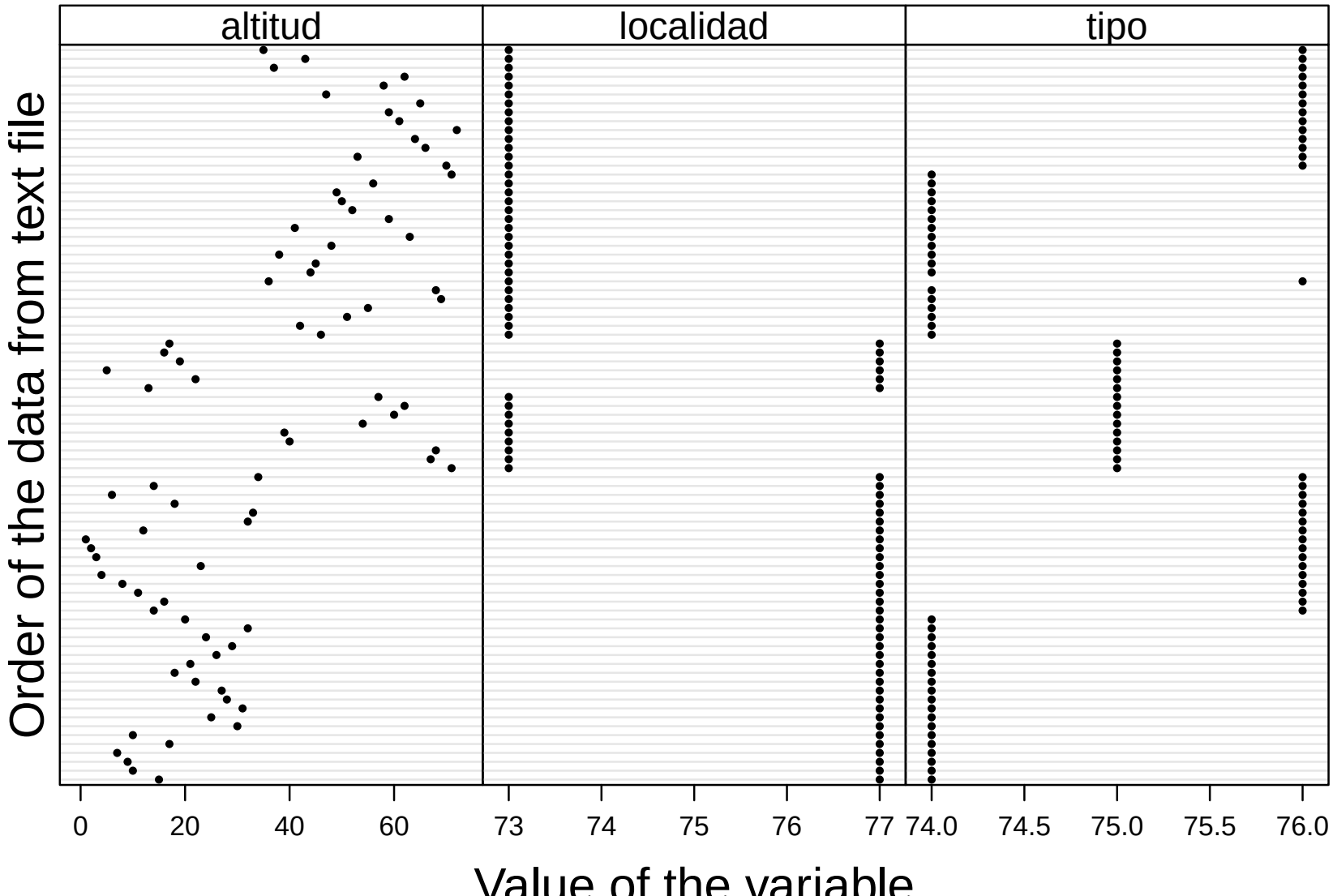
- locality
- type of ecosystem
- Elevation

Outliers

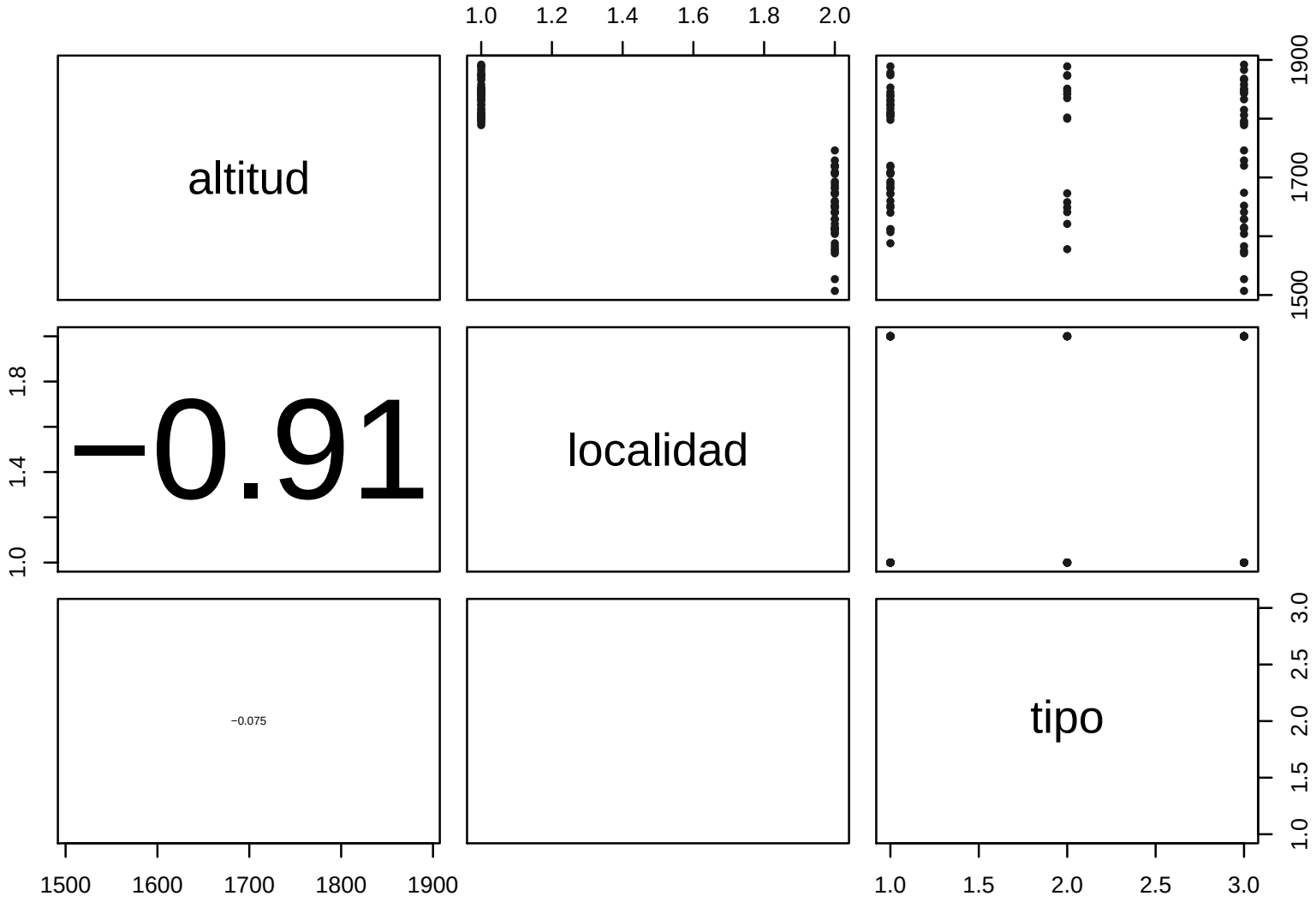

n.seed counts
0
10
20
30
40
50
60


CANAR
SANJUAN
n.seed counts
0
20
40
60
Range of data

Outliers


altitud
localidad
tipo
Order of the data from text file
0
20
40
60
73
74
75
76
77
74.0
74.5
75.0
75.5
76.0
Value of the variable

Collinearity of covariables



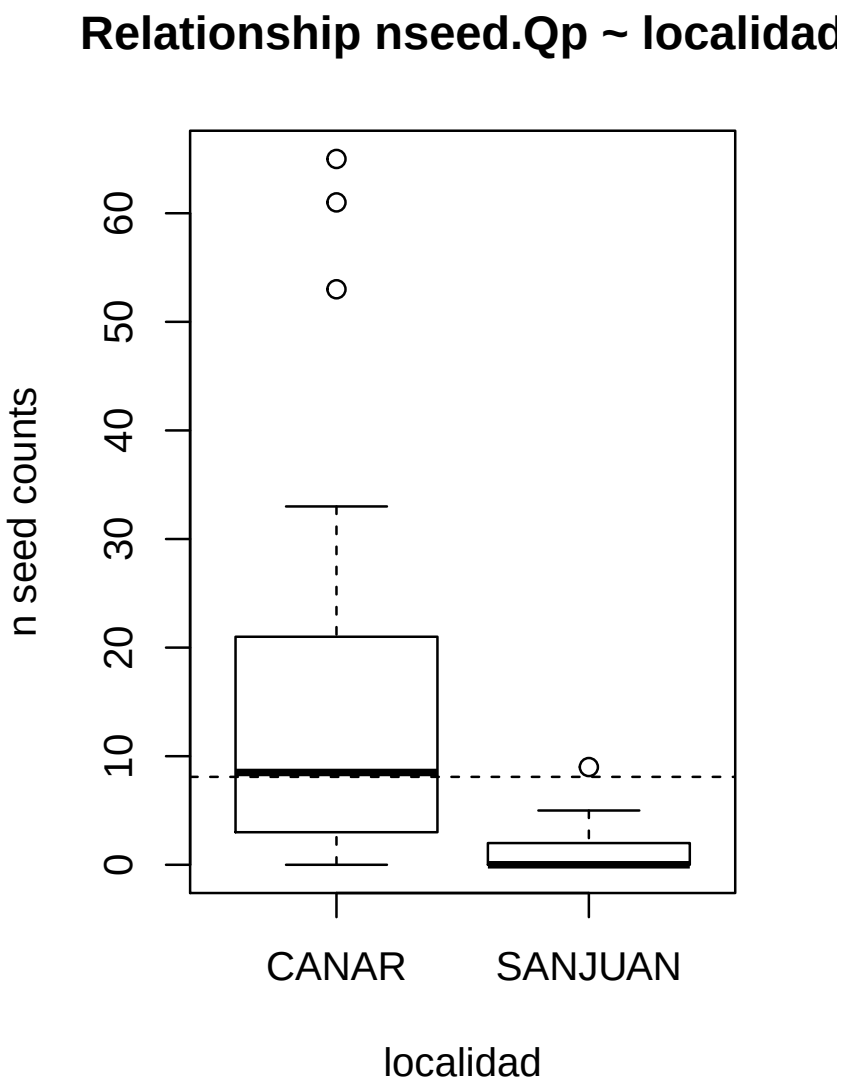
Data Exploration: VIF

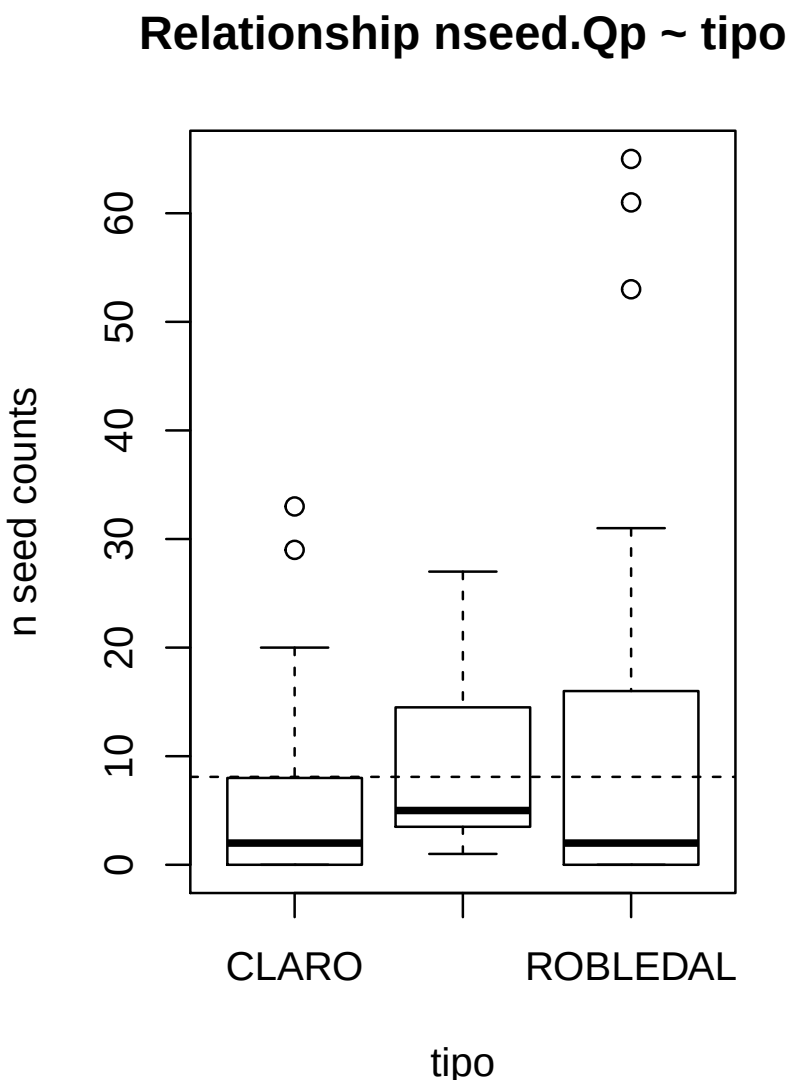
```
Variance inflation factors

              GVIF Df GVIF^(1/2Df)
altitud   6.060345  1     2.461777
localidad 6.026341  1     2.454861
tipo      1.043360  2     1.010668
```

```
Variance inflation factors

              GVIF Df GVIF^(1/2Df)
localidad 1.00786  1     1.003923
tipo      1.00786  2     1.001959
```

There is a high collinearity between ***altidud*** and ***localidad***. The variance inflation factors analysis reveals a medium value for ***altitud*** (vif=2.46).
We removed the ***altitud*** from analysis (according to Zuur et al, 2013).
We re-examined VIF and we found values close to 1 for the two variables.

Relationships Y vs X



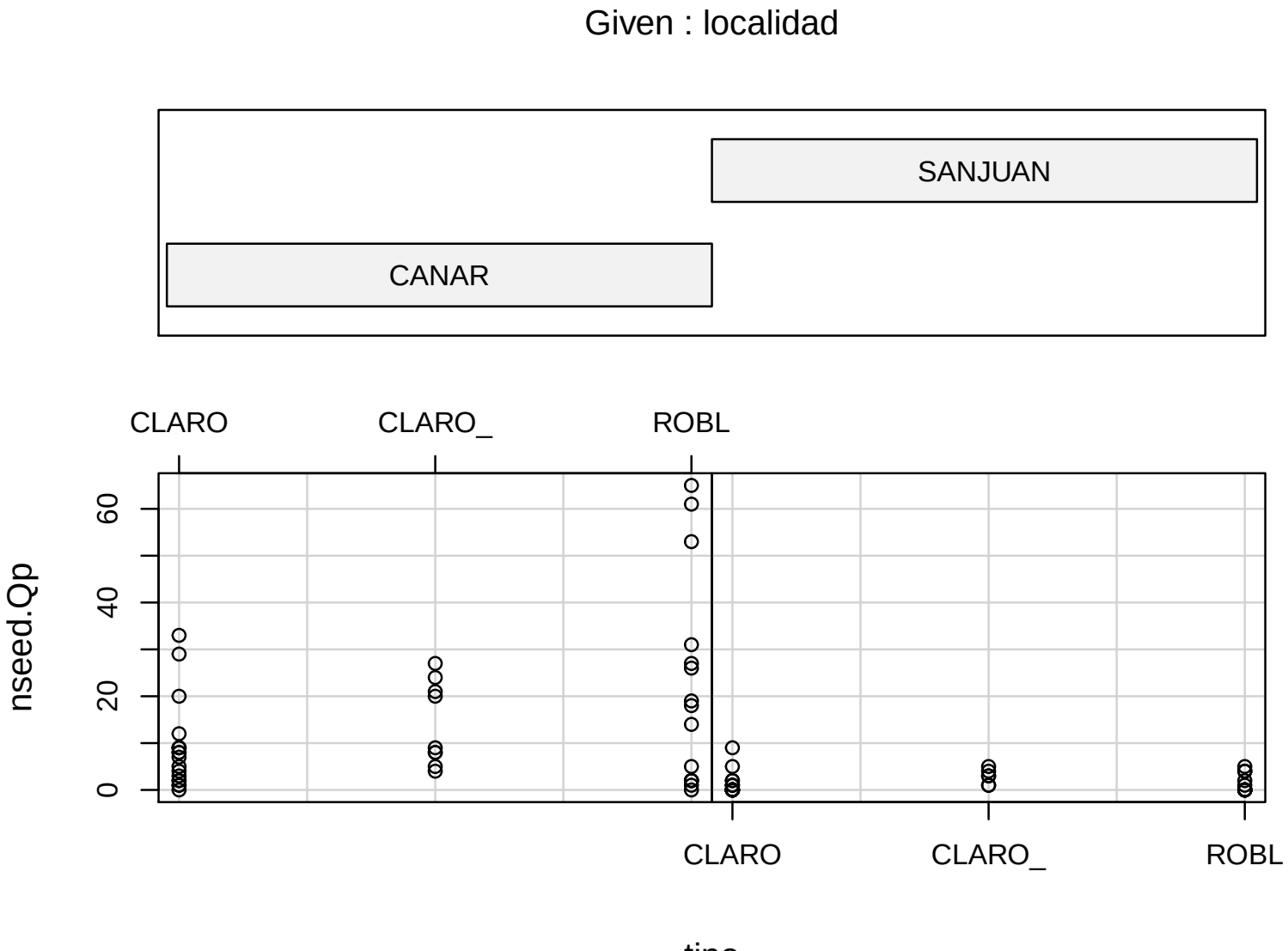


Another evidence of outliers for forest (`tipo='ROBLEDAL'`) of Canar locality (`localidad='CANAR'`).

Spatial/temporal aspects of sampling design

not relevant here

Interactions



Zero inflation Y

```
[1] 0.2891566
```

28.9156627% of data are zeros. We need to consider a model that accommodates this issue.

Summary from Data Exploration

- There are three outliers.
- The ***altitud*** variable has high collinearity and a elevated *vif*.
- We decided to remove the three outliers and do not consider the altitud variable.

Data Exploration

	nombre	localidad	tipo	subtipo	replicate	altitud	ns
75	P027	CANAR	ROBLEDAL	ROBLEDAL	R2	1849	
80	P034	CANAR	ROBLEDAL	ROBLEDAL	R3	1851	
83	P038	CANAR	ROBLEDAL	ROBLEDAL	R2	1789