Modelling oak recruitment at abandoned croplands

Antonio J. Perez-Luque (@ajpelu)

```
library(lattice)
library(tidyverse)
library(here)
source(here::here("script/auxfx.R"))
```

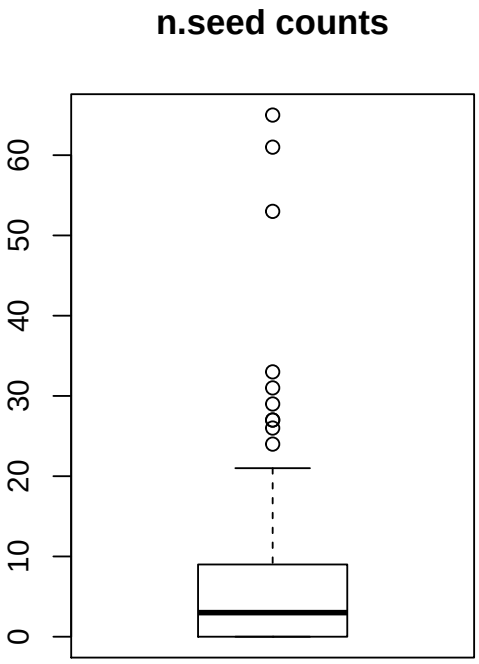
```
##  Read data
# Abandonded crop data
data_seed <- read.table(here::here("/datos/cul_abundance_seedling.txt"), head=TRUE, sep=";")

# Select only data of nseed.Qp
ab.s <- data_seed[, c('nombre','localidad','tipo','subtipo','replicate','altitud','nseed.Qp')]

# Set 0 to NA
ab.s[is.na(ab.s)] <- 0
```

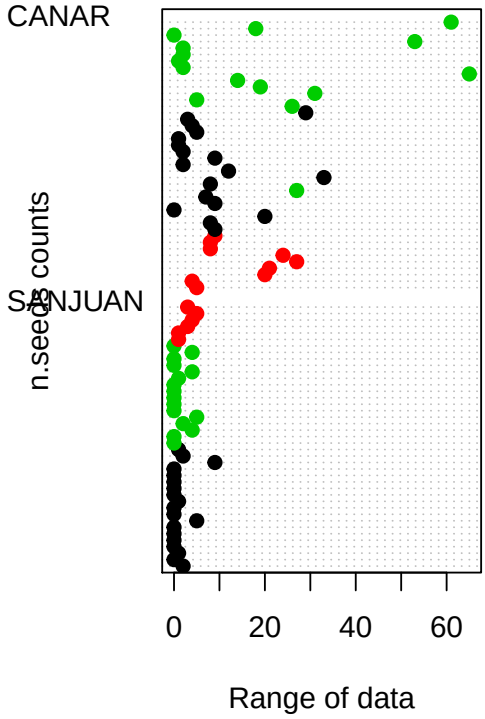
Aims

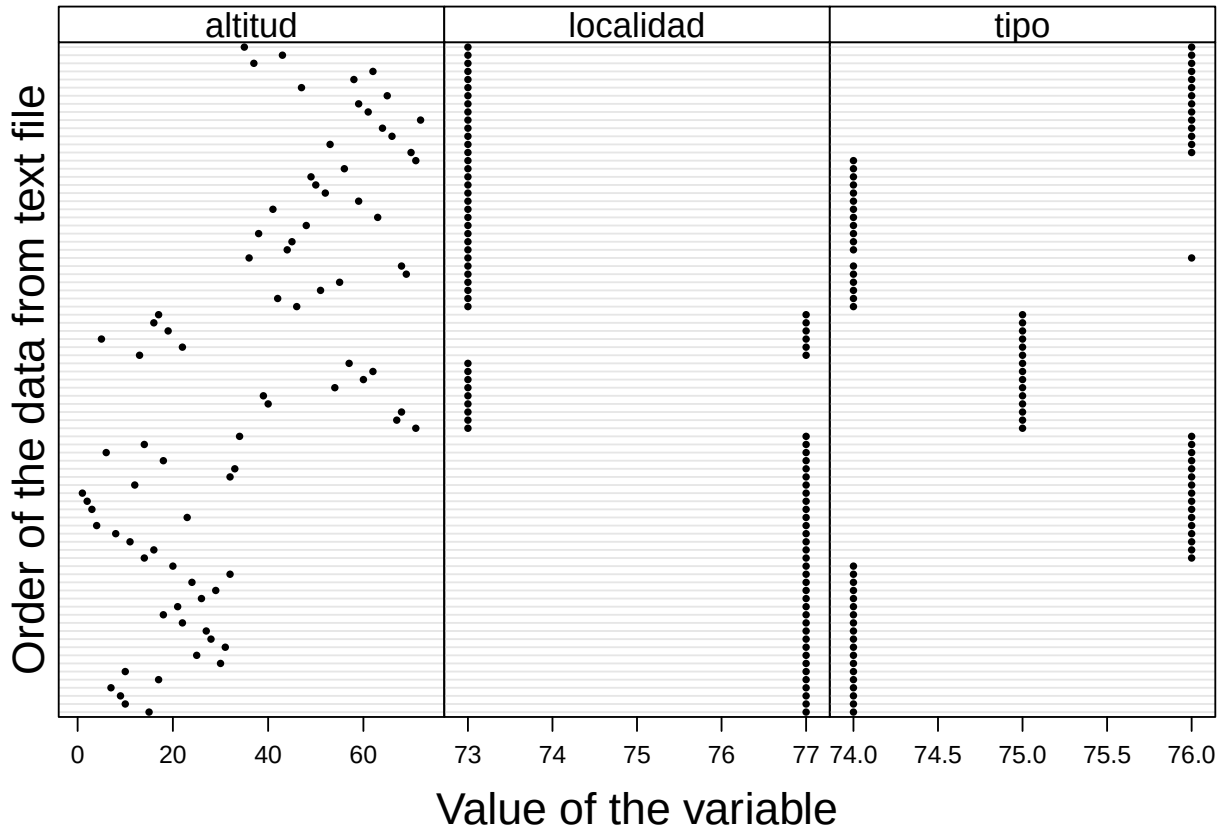
Model Oak seedlings (*n.seed*; heigth $<$ 130 cm) as a function of: - locality - type of ecosystem - Elevation

Data Exploration

Outliers


n.seed counts
0
10
20
30
40
50
60


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



There are some potential outliers, in Canar locality. There are three transects with more than 50 seedling.

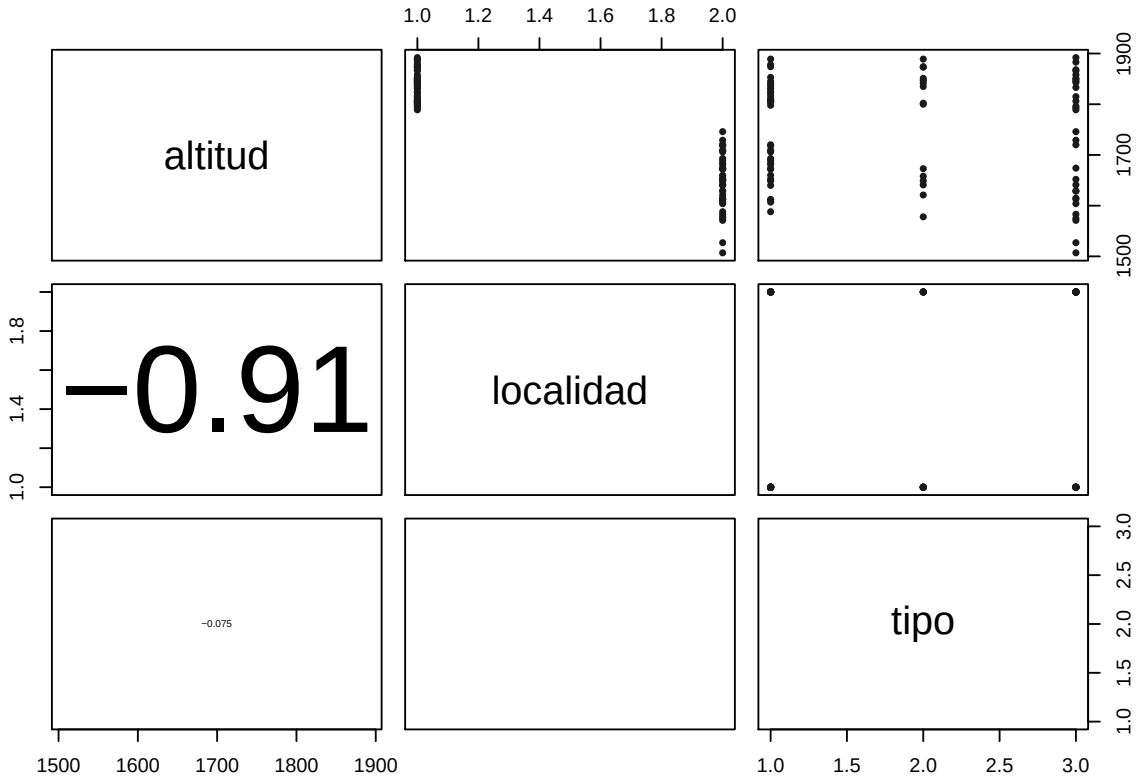
```
ab.s[ ab.s$nseed.Qp>50,]
```

```
##    nombre localidad     tipo  subtipo replicate altitud nseed.Qp
## 75   P027     CANAR ROBLEDAL ROBLEDAL        R2    1849       65
## 80   P034     CANAR ROBLEDAL ROBLEDAL        R3    1851       53
## 83   P038     CANAR ROBLEDAL ROBLEDAL        R2    1789       61
```

These data are from forest type.

Collinearity of covariables

```
## Collinearity
MyVar <- c('altitud', 'localidad', 'tipo')
Mypairs(ab.s[, MyVar])
```




```
### See VIF
# Todas las covariables
corvif(ab.s[, MyVar])
```

```
##
##
## 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
```

```
# Quitamos la de mayor valor VIF
MyVar2 <- c('localidad','tipo')
corvif(ab.s[, MyVar2])
```

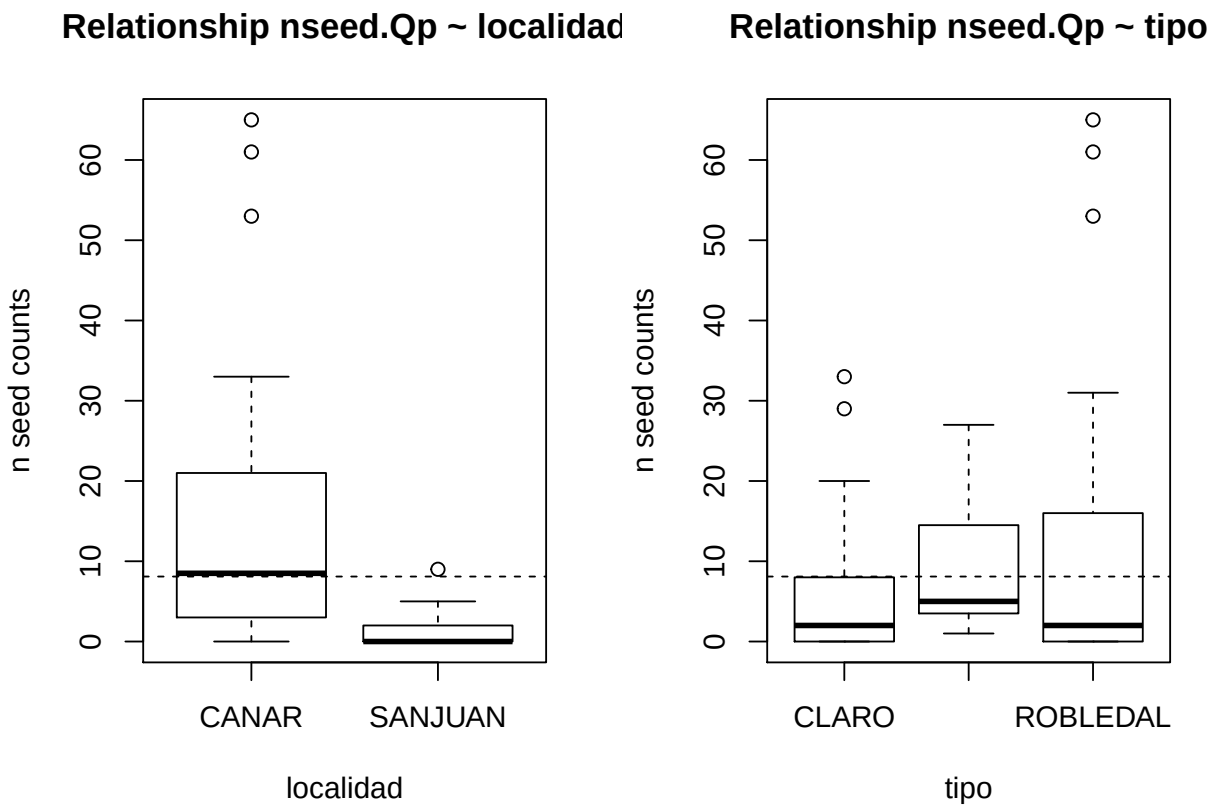
```
##
##
## Variance inflation factors
```

```
##
##                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

```
# If X is a factor....boxplots
par(mfrow = c(1, 2))
boxplot(nseed.Qp ~ localidad, data = ab.s, ylab='n seed counts', main='Relationship nseed.Qp ~
abline(h=mean(ab.s$nseed.Qp),lty = 2)
boxplot(nseed.Qp ~ tipo, data = ab.s, ylab='n seed counts', main='Relationship nseed.Qp ~ tipo
abline(h=mean(ab.s$nseed.Qp),lty = 2)
```

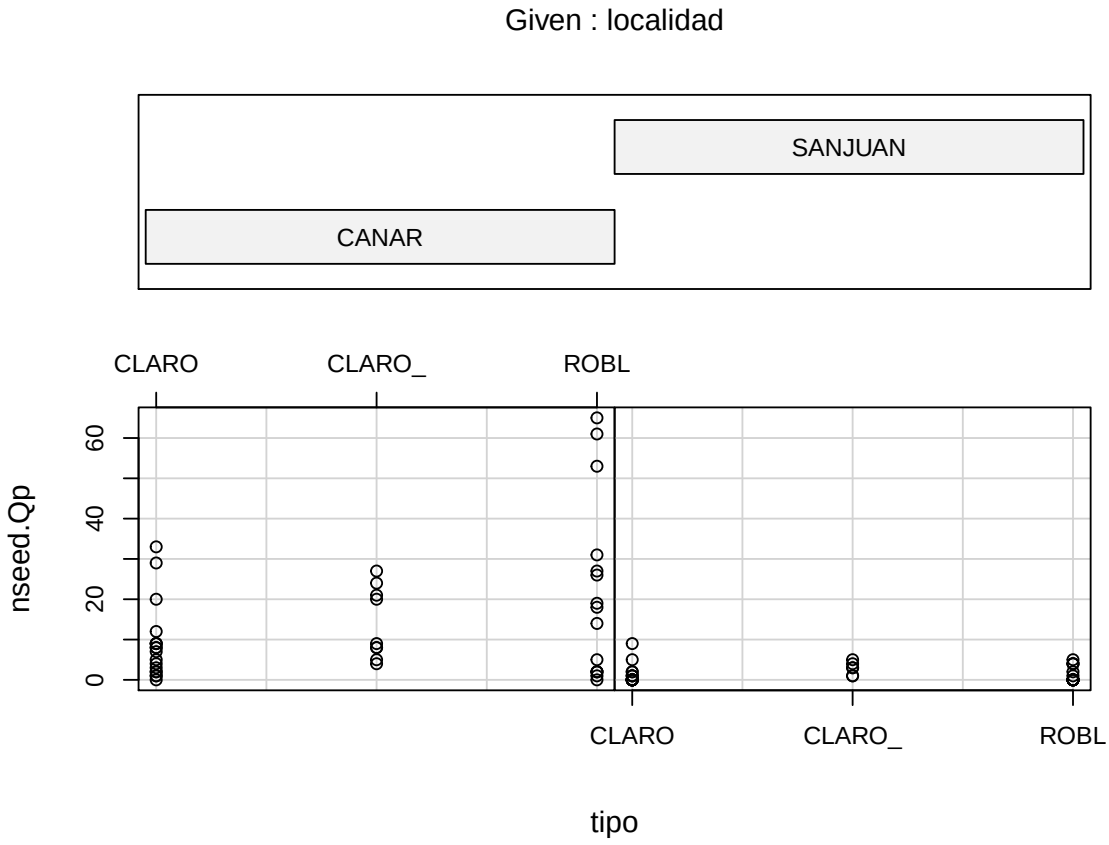


Another evidence of outliers for forest (`tipo='ROBLEDAL'`) of Canar locality (`localidad='CANAR'`).

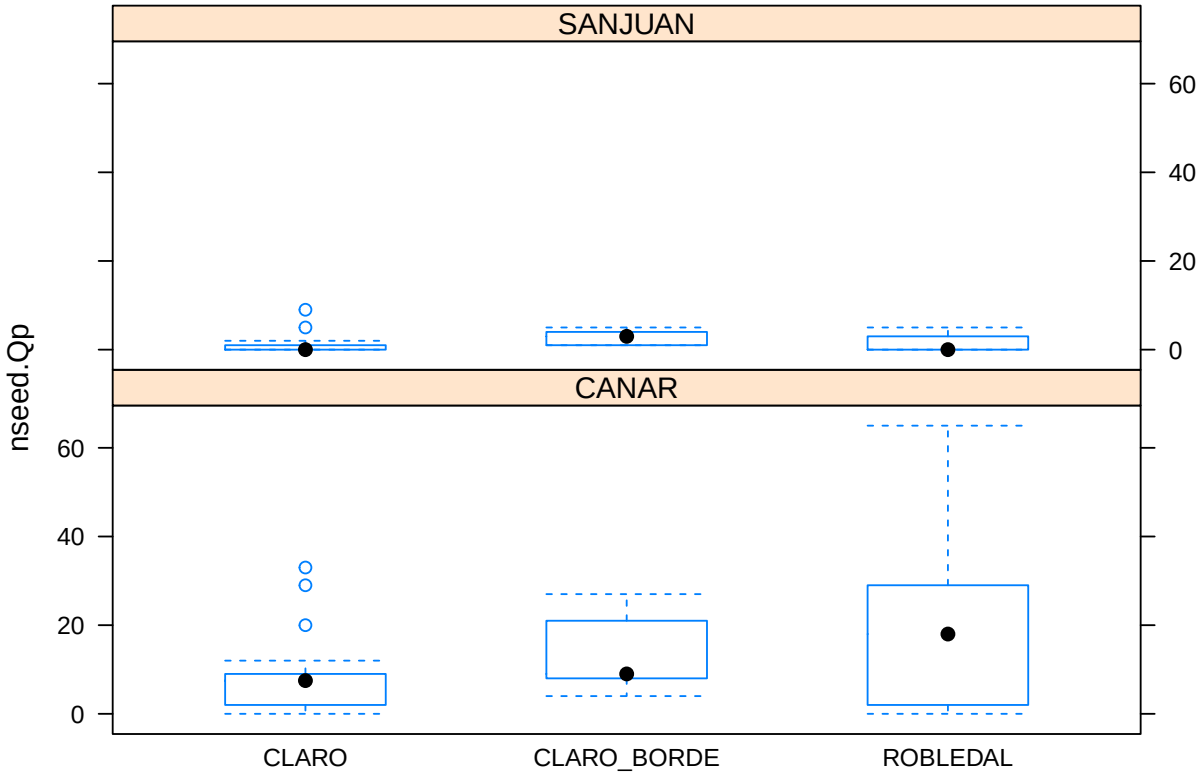
Spatial/temporal aspects of sampling design (not relevant here)

Interactions

```
# Interactions
coplot(nseed.Qp ~ tipo | localidad, rows=1, data = ab.s)
```




```
bwplot(nseed.Qp ~ factor(tipo) | factor(localidad), data = ab.s, col = 1, layout=c(1,2))
```



Zero inflation Y

```
# F. Zero inflation
sum(ab.s$nseed.Qp == 0) / length(ab.s$nseed.Qp)
```

```
## [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.

```
# Explore records with seelindg greather than 50
ab.s[ ab.s$nseed.Qp>50,]
```

```
##    nombre localidad     tipo  subtipo replicate altitud nseed.Qp
## 75   P027     CANAR ROBLEDAL ROBLEDAL        R2    1849       65
## 80   P034     CANAR ROBLEDAL ROBLEDAL        R3    1851       53
## 83   P038     CANAR ROBLEDAL ROBLEDAL        R2    1789       61
```

```r
# Remove data
ab.s1 <- ab.s[ab.s$nseed.Qp<50,]

# Explore again the interactions
bwplot(nseed.Qp ~ factor(tipo) | factor(localidad), data = ab.s1, col = 1, layout=c(1,2))
```