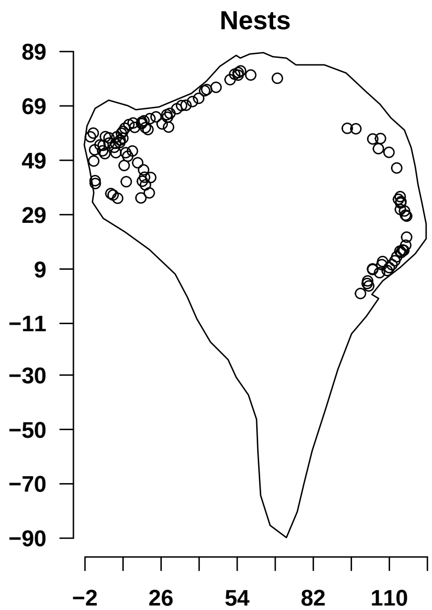
TASK 2

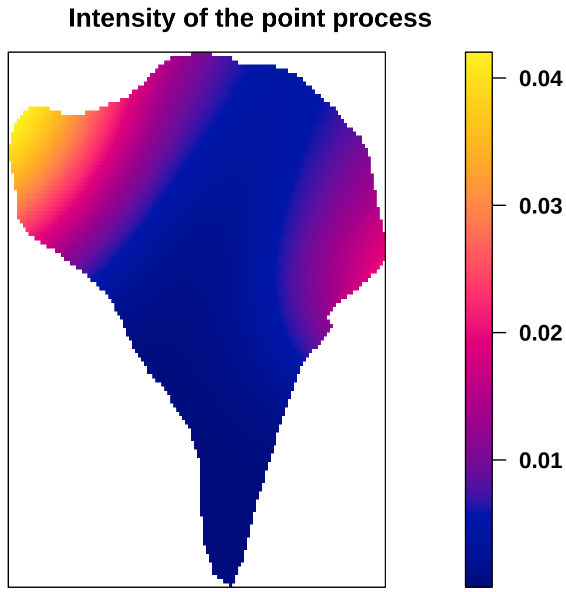
Task 2: Intensity and Randomness

The nest data from islet "nucli 84" is stored in nucli84.txt. Additionally, the coordinates of the islet are in poly84.txt.

1) Build a ppp object using the "nucli 84" data.



2) Draw a plot with the intensity of the point process computed by the non-parametric approach. Briefly comment the results.

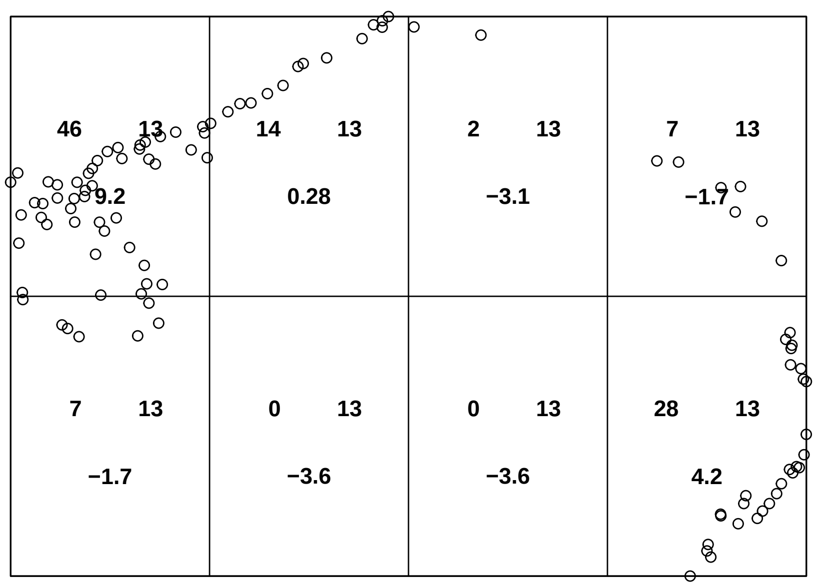


breve spiegazione

3) Assess the Completely Spatial Randomness hypothesis

- via Chi-square test:

We divide the region in 8 subareas with equal areas and under CSR we would expect more or less same number of nests in each subregion.



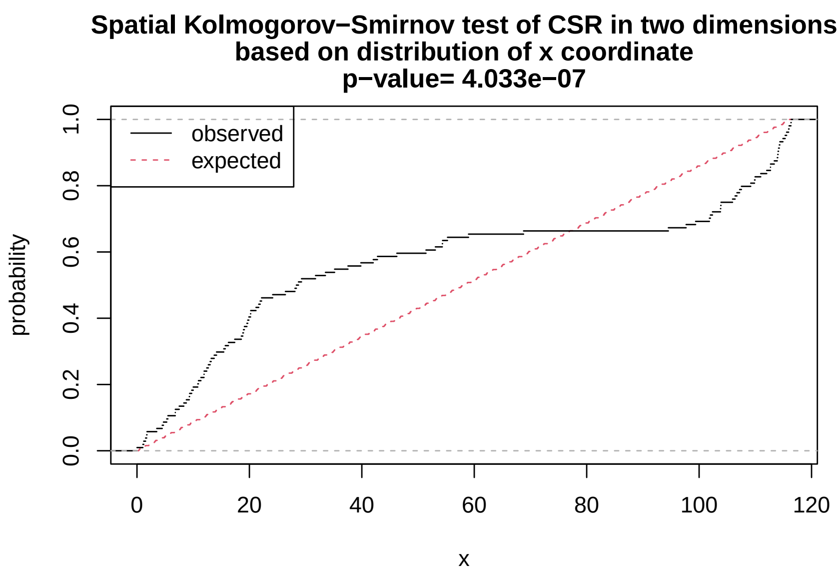
Nonetheless we notice that some subregions have more nests than other. Indeed after performing a Chi-square test we get:

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84
## X2 = 142, df = 7, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 4 by 2 grid of tiles
```

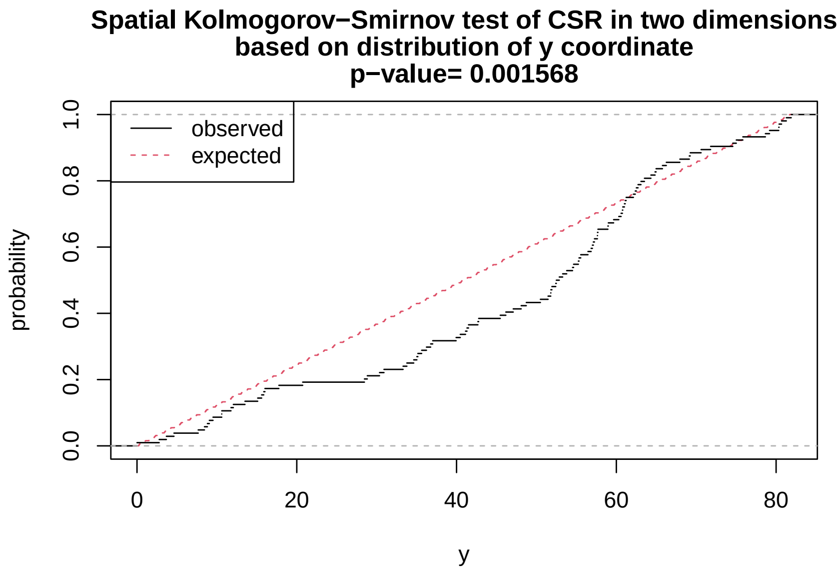
The p-value is extremely low ($p < 2.2e - 16$), so this test leads us to reject the null hypothesis of complete spatial randomness (CRS).

- via Kolmogorov-Smirnov test:

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'x' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.27225, p-value = 4.033e-07
## alternative hypothesis: two-sided
```



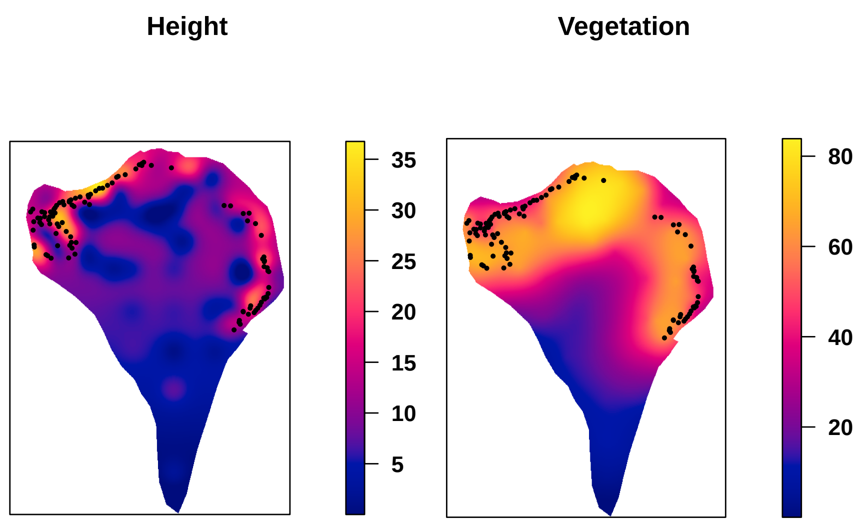

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'y' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.18542, p-value = 0.001568
## alternative hypothesis: two-sided
```



The observed distribution of Z at data points and the expected one are quite different, so we can reject the null hypothesis of CSR and state that there is a dependency between the intensity of the points and both the Cartesian coordinates.

4) Assess the relation between the intensity of the point process and the covariates height and vegetation.

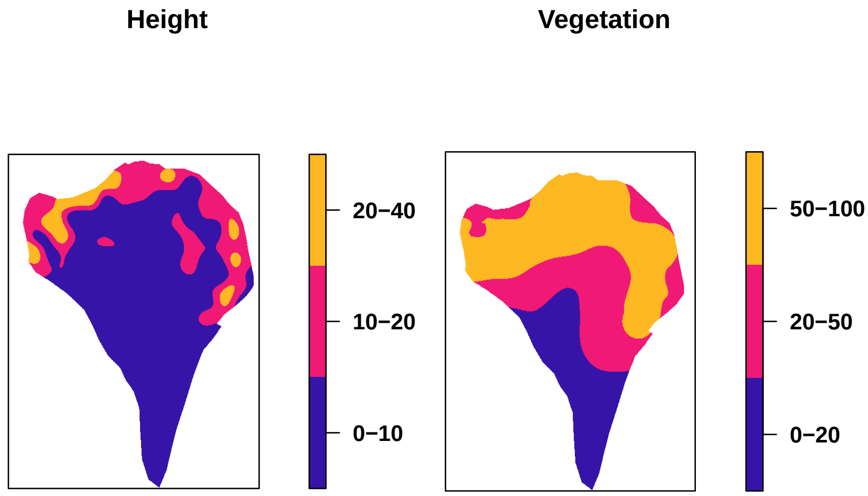
First we visualize the values of Height and Vegetation in the islet:



The nests seem to be concentrated where the height and the vegetation have an higher value.

Now we categorize the covariates using the suggested intervals:

- Height. [0,10], (10,20], (20,40]
- Vegetation. [0,20], (20,50], (50,100]

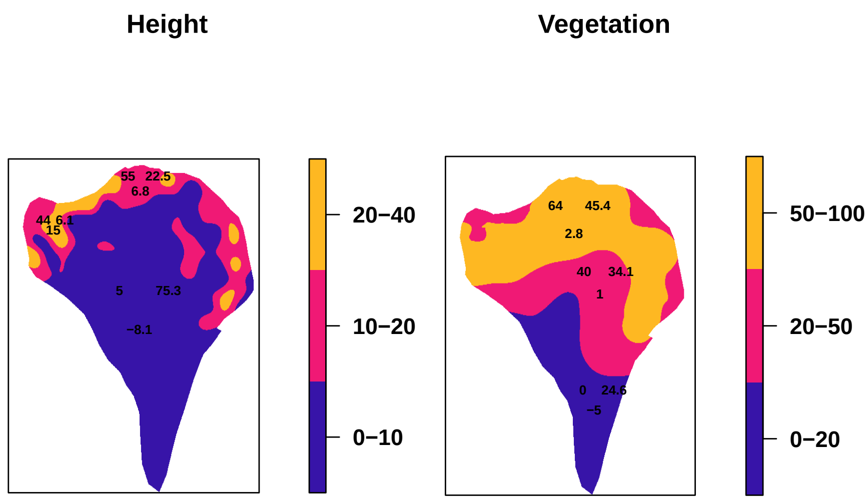


And to assess quantitatively this relation we use:

- a Chi-squared test:

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84
## X2 = 345.67, df = 2, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 3 tiles (levels of a pixel image)
```

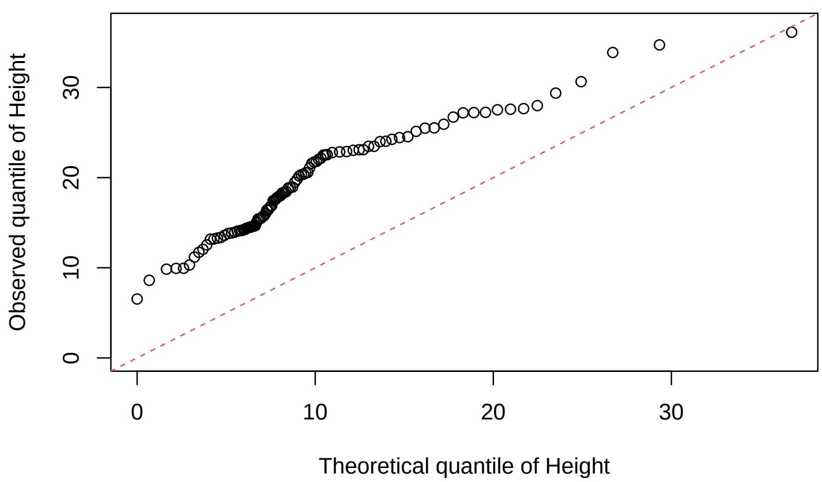
```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84p
## X2 = 33.266, df = 2, p-value = 1.195e-07
## alternative hypothesis: two.sided
##
## Quadrats: 3 tiles (levels of a pixel image)
```



The Chi-squared has $pvalue = 1.195 \cdot 10^{-7}$.

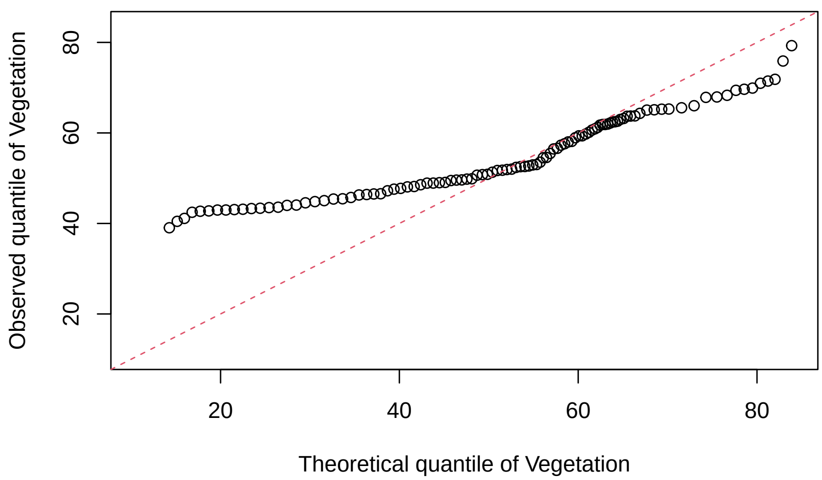
- a Kolmogorov-Smirnov test:

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'Height' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.68592, p-value < 2.2e-16
## alternative hypothesis: two-sided
```




```
##
```

```
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'Vegetation' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.28454, p-value = 9.713e-08
## alternative hypothesis: two-sided
```



The Kolmogorov test has very little p-values and the plots suggest a general difference of the observed quantiles from the theorical ones.

So, in the light of the above considerations, we can state there is in fact a relation between the covariates and the point process.

5) Fit an inhomogeneous Poisson model to data

We'll fit the model:

$$\lambda \sim x + y + \text{Height} + \text{Vegetation}$$

```
## Nonstationary Poisson process
##
## Log intensity:  ~x + y + Height + Vegetation
##
## Fitted trend coefficients:
##  (Intercept)            x            y       Height   Vegetation
## -6.584757676 -0.001911082 -0.014239000  0.159696511  0.013871300
##
##                 Estimate        S.E.       CI95.lo      CI95.hi Ztest
## (Intercept) -6.584757676 0.514867186 -7.5938788176 -5.575636535   ***
## x           -0.001911082 0.002851904 -0.0075007104  0.003678547
## y           -0.014239000 0.005477673 -0.0249750432 -0.003502958    **
## Height       0.159696511 0.012900281  0.1344124249  0.184980596   ***
## Vegetation   0.013871300 0.007465117 -0.0007600604  0.028502661
##                    Zval
```

```
## (Intercept) -12.7892355
## x            -0.6701073
## y            -2.5994613
## Height       12.3793050
## Vegetation    1.8581491
## Problem:
##  Values of the covariates 'Height', 'Vegetation' were NA or undefined at 14%
## (158 out of 1132) of the quadrature points
```

We'll do a step-wise feature selection, removing each time a variable which is not significant.

First we remove te x variable since it's the one with the value of the Z statistic closest to 0.

```
## Nonstationary Poisson process
##
## Log intensity:  ~y + Height + Vegetation
##
## Fitted trend coefficients:
## (Intercept)           y      Height  Vegetation
## -6.75181119 -0.01349451  0.16225587  0.01358246
##
##                Estimate        S.E.     CI95.lo      CI95.hi Ztest       Zval
## (Intercept) -6.75181119 0.457227288 -7.64796020 -5.855662168   *** -14.766860
## y           -0.01349451 0.005299778 -0.02388188 -0.003107134     *  -2.546240
## Height       0.16225587 0.012356964  0.13803667  0.186475077   ***  13.130723
## Vegetation   0.01358246 0.007508543 -0.00113401  0.028298938         1.808935
## Problem:
##  Values of the covariates 'Height', 'Vegetation' were NA or undefined at 14%
## (158 out of 1132) of the quadrature points
```

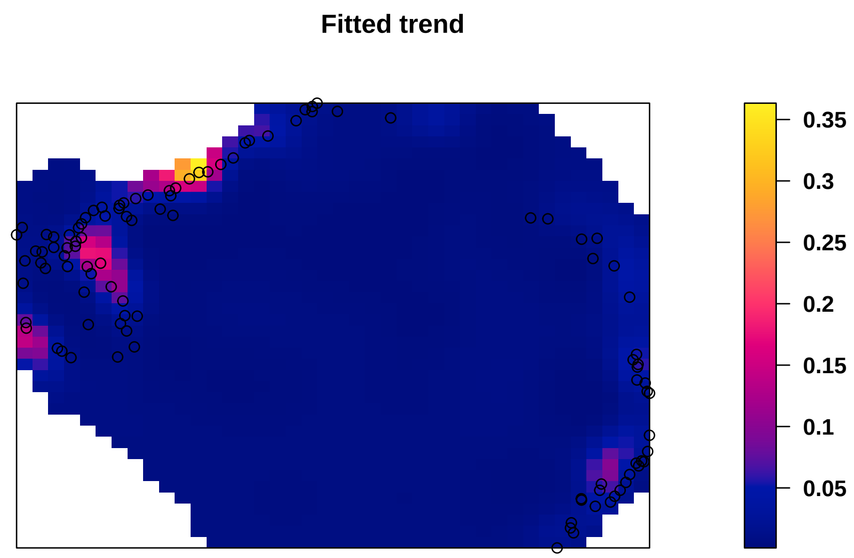
Then we can remove the "Vegetation" variable.

```
## Nonstationary Poisson process
##
## Log intensity:  ~y + Height
##
## Fitted trend coefficients:
## (Intercept)           y      Height
## -6.13014916 -0.01098515  0.16146681
##
##                Estimate        S.E.     CI95.lo       CI95.hi Ztest       Zval
## (Intercept) -6.13014916 0.261824769 -6.64331627 -5.6169820386   *** -23.413175
## y           -0.01098515 0.005302281 -0.02137743 -0.0005928702     *  -2.071778
## Height       0.16146681 0.012469824  0.13702640  0.1859072154   ***  12.948604
## Problem:
##  Values of the covariate 'Height' were NA or undefined at 14% (155 out of 1132)
## of the quadrature points
```

Now all the covariates are significant (they have the Z statistic far enough from zero and their confidence intervals don't contain 0), so we can keep the model:

$$\lambda \sim x + y + \text{Height} + \text{Vegetation}$$

and we plot the fitted trend:



Now we assess the goodness of fit of the chosen model:

```
##
##  Chi-squared test of fitted Poisson model 'chosen.model' using quadrat
##  counts
##
## data:  data from chosen.model
## X2 = 40.902, df = 5, p-value = 1.964e-07
## alternative hypothesis: two.sided
##
## Quadrats: 4 by 2 grid of tiles
```