# Point Pattern Exam Question 1

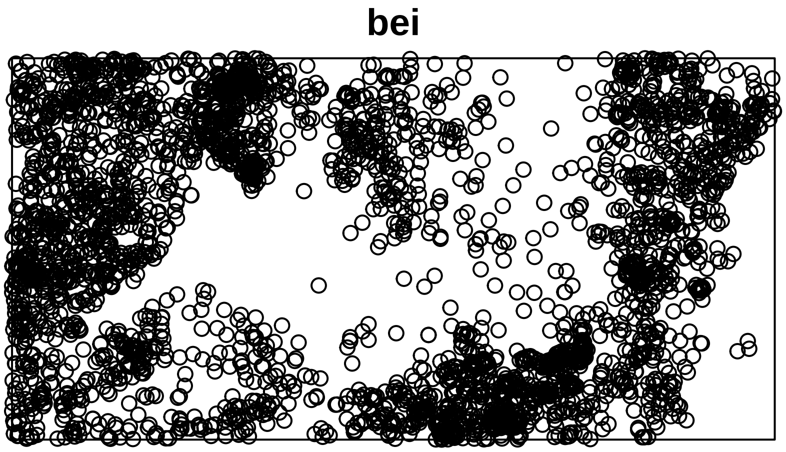
## Auriel MV Fournier

```
library(spatstat)
```

```
##
## spatstat 1.40-0       (nickname: 'Do The Maths')
## For an introduction to spatstat, type 'beginner'
```

First we read in and plot the data to see what we have.

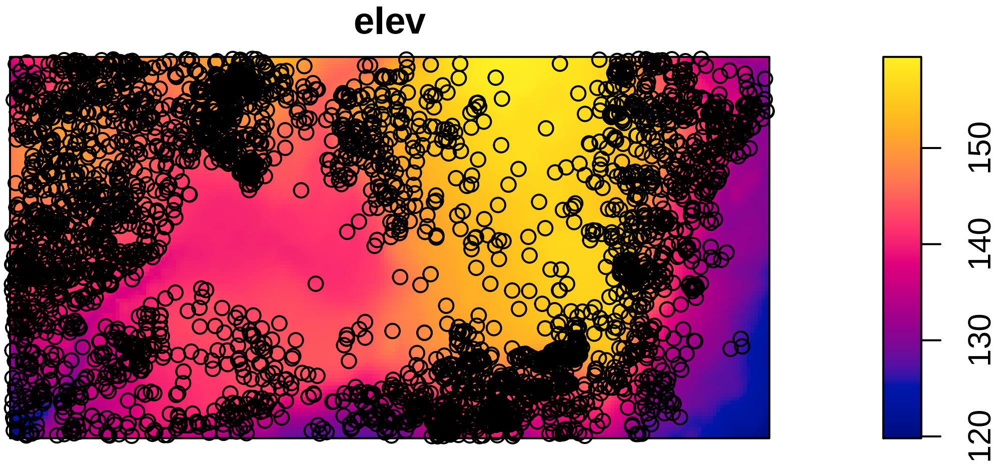
```
data(bei)

plot(bei)
```



Then we pull out the covariates and plot those as well.

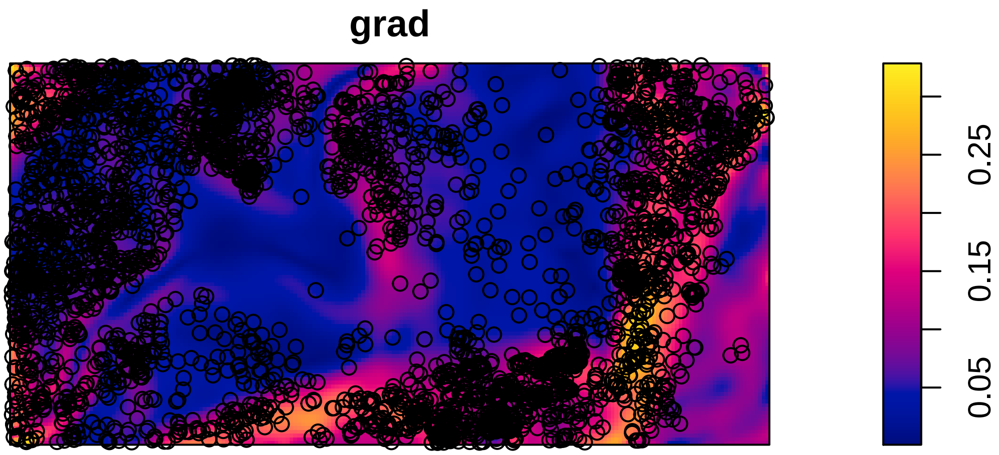
```
elev <- bei.extra$elev
grad <- bei.extra$grad
plot(elev)
points(bei, add=T)
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "add" is not a
## graphical parameter
```



```
plot(grad)
points(bei, add=T)
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "add" is not a
## graphical parameter
```

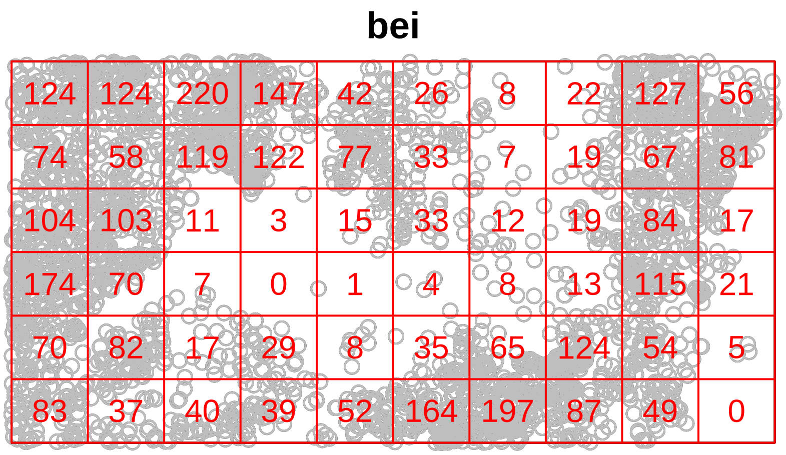


There appears to be quite a bit of variation in gradient (slope) but not as much in elevation. Based on the arrangement of the points vs the covariates I would guess that gradient will have a larger impact on intensity of points then elevation, though it almost appears in some places like the trees are on certain sides of the slope so I wonder if aspect might also play a role.

```
q <- quadratcount(bei, nx=10,ny=6)
plot(bei)
points(bei, col="gray",add=T)
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "add" is not a
## graphical parameter
```

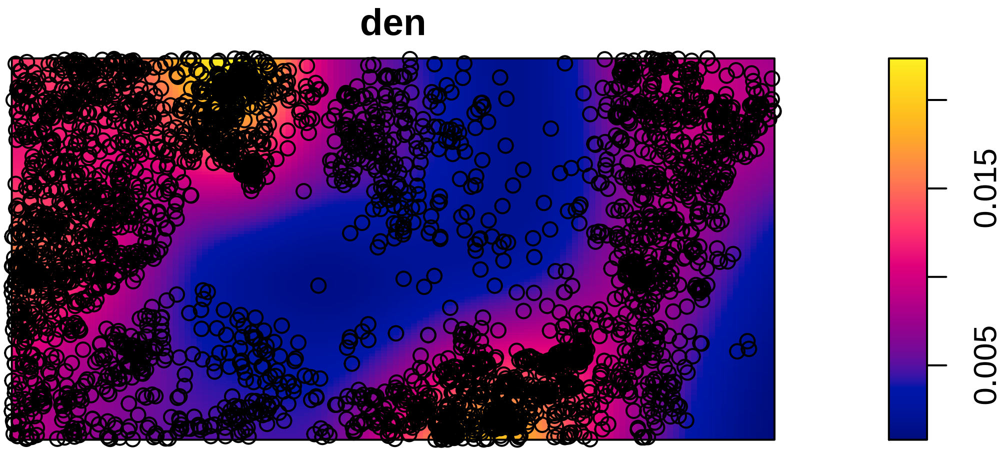
```
plot(q, add=T, col="red")
```



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 124 | 124 | 220 | 147 | 42 | 26 | 8 | 22 | 127 | 56 |
| 74 | 58 | 119 | 122 | 77 | 33 | 7 | 19 | 67 | 81 |
| 104 | 103 | 11 | 3 | 15 | 33 | 12 | 19 | 84 | 17 |
| 174 | 70 | 7 | 0 | 1 | 4 | 8 | 13 | 115 | 21 |
| 70 | 82 | 17 | 29 | 8 | 35 | 65 | 124 | 54 | 5 |
| 83 | 37 | 40 | 39 | 52 | 164 | 197 | 87 | 49 | 0 |

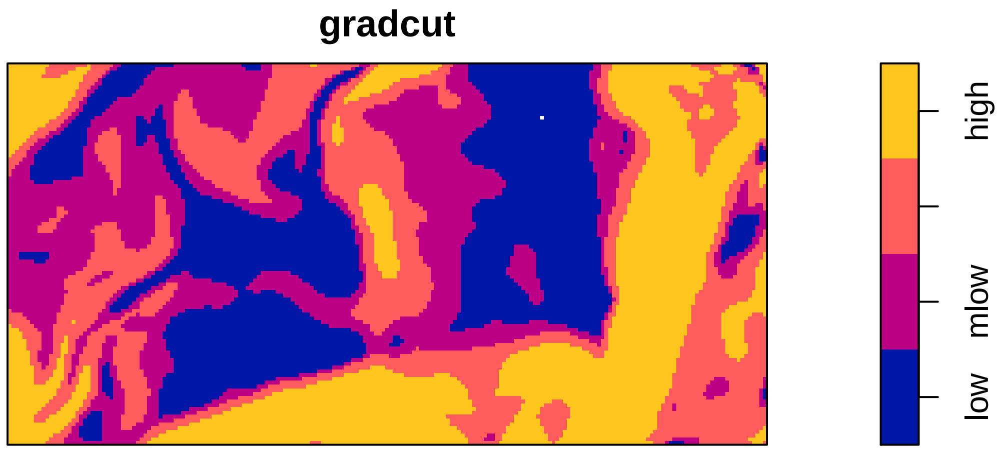
next step is looking at a surface of point density through the use of a kernel density estimator.

```
den <- density.ppp(bei, sigma=70, kernel="gaussian")
plot(den)
plot(bei,add=T)
```

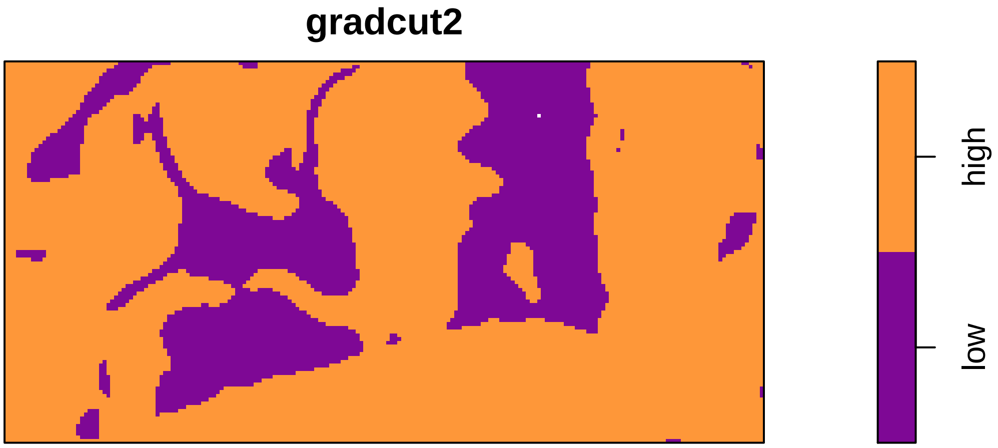


When you look at the points over the gradient covariate surface it appears like they are avoiding the areas with very low slope (at least in some areas). So lets try breaking up gradient into several categorical variables and see how those look.

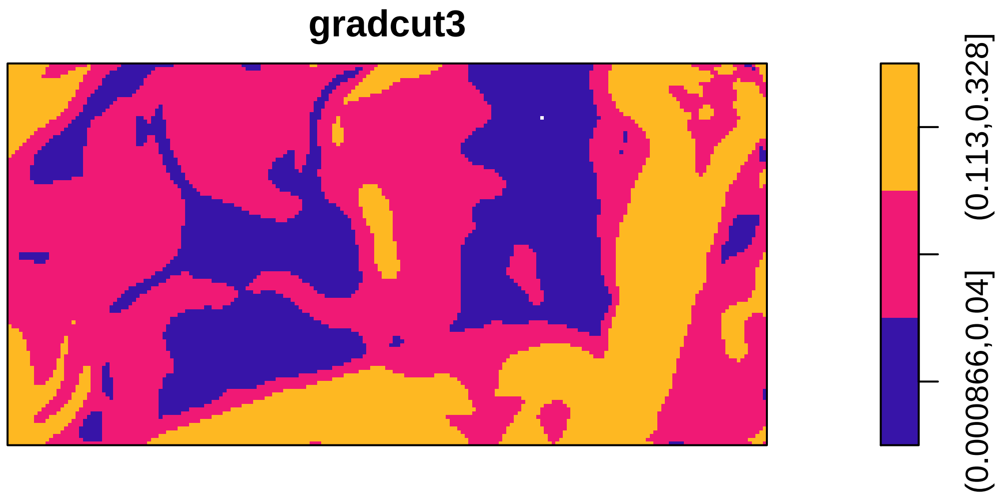
```
# four quantiles
b <- quantile(grad, probs=(0:4)/4)
# two quantiles, the lower 25% and everything else
b2 <- quantile(grad, probs=c(0,0.25,1))
# three quantiles, the lower 25%, the middle 50% and the upper 25%
b3 <- quantile(grad, probs=c(0,.25,.75,1))

gradcut <- cut(grad, breaks=b, labels=c('low','mlow','mhigh','high'))
gradcut2 <- cut(grad, breaks=b2, labels=c('low','high'))
gradcut3 <- cut(grad, breaks=b3)

plot(gradcut)
```



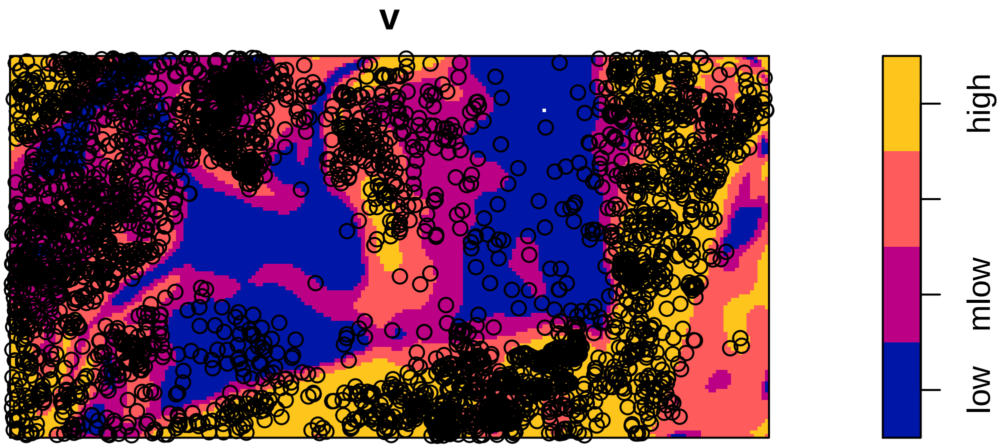
```
plot(gradcut2)
```



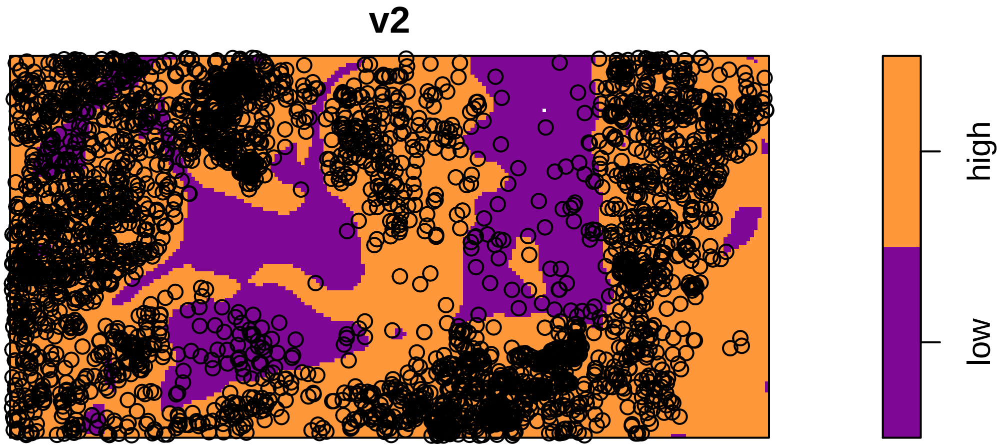
```
plot(gradcut3)
```



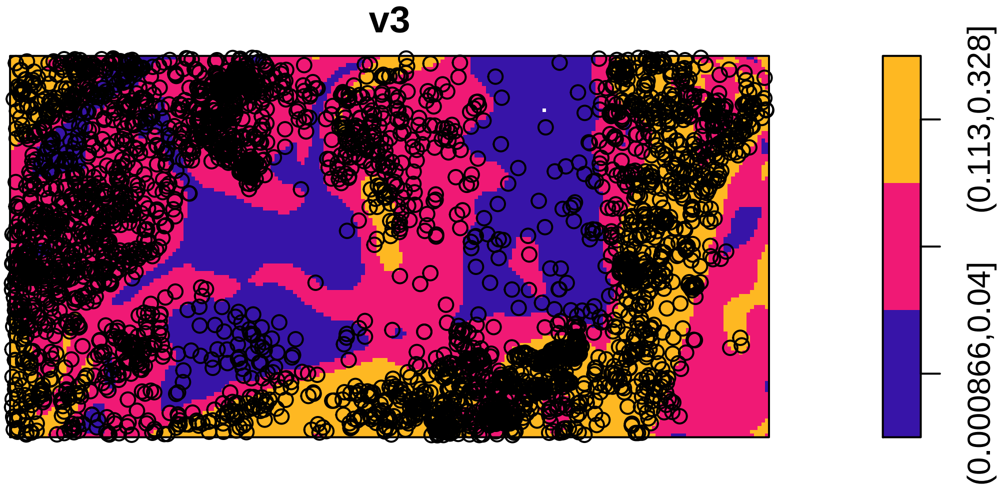
```
v <- tess(image=gradcut)
v2 <- tess(image=gradcut2)
v3 <- tess(image=gradcut3)

plot(v)
plot(bei, add=T)
```



```
plot(v2)
plot(bei, add=T)
```



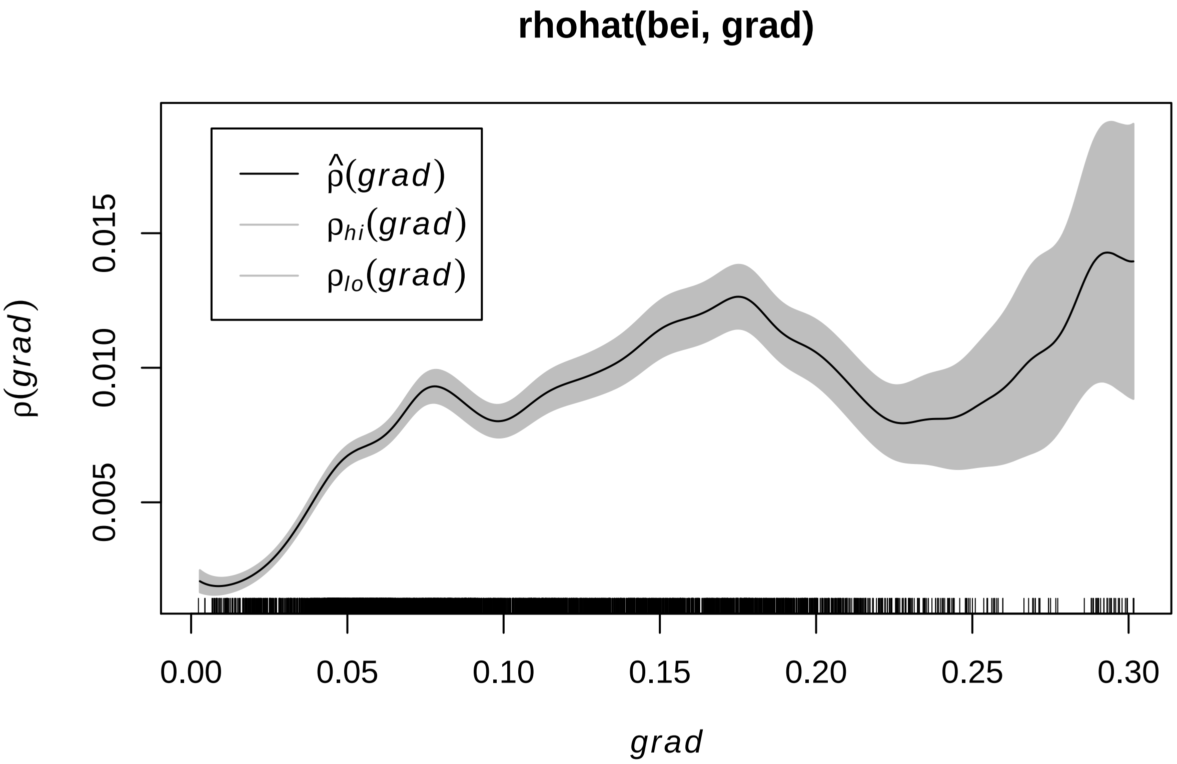
```
plot(v3)
plot(bei, add=T)
```



```
qb <- quadratcount(bei, tess=v)
qb2 <- quadratcount(bei,tess=v2)
qb3 <- quadratcount(bei,tess=v3)
```

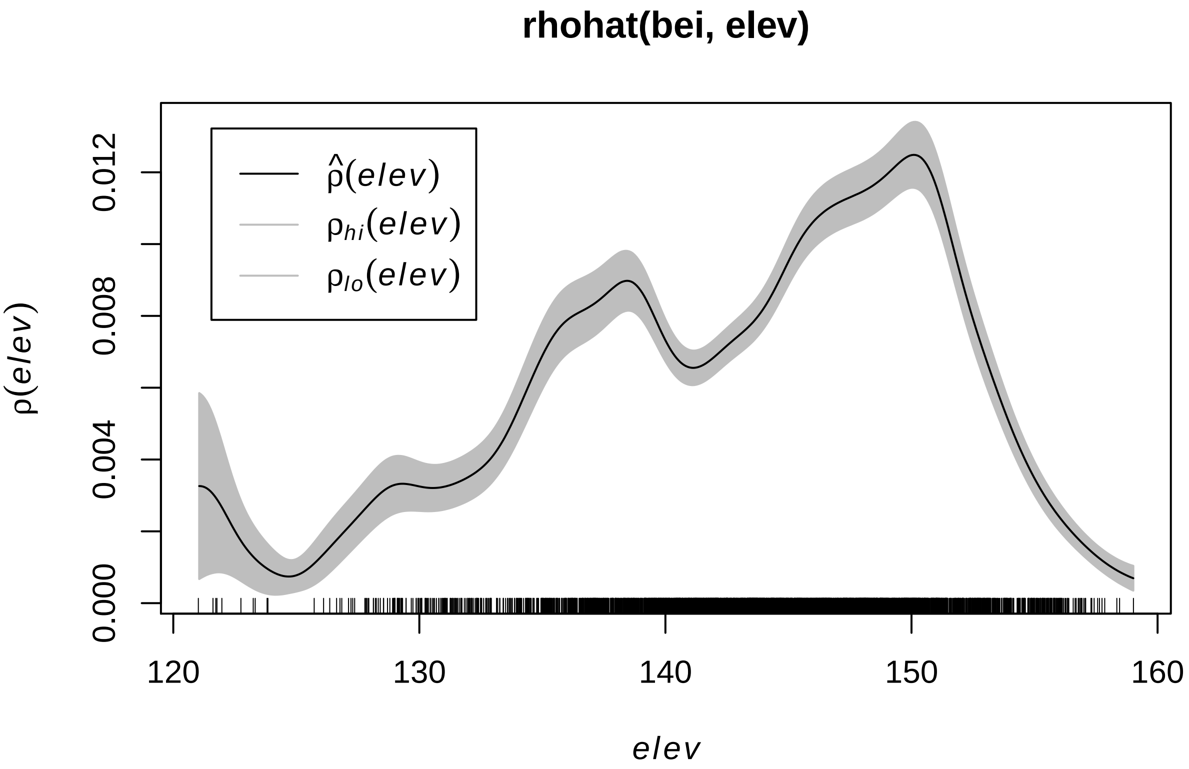
rhohat is useful for looking at the distribution of points across a continuous variable, the following four graphs examine it for gradient, elevation, gradient times elevation. I thought there might be an interaction between the two covariates since it could impact the soil or other environmental variables.

```
plot(rhohat(bei,grad))
```



The first rhohat graph shows that there are not very many trees at lower intensities, but many, with a tight envelope in the mid range. At higher gradients there is a increase in density, but the envelope is also wider, showing less certainty in this, so these values may not actually differ or might be lower then the mid range values.

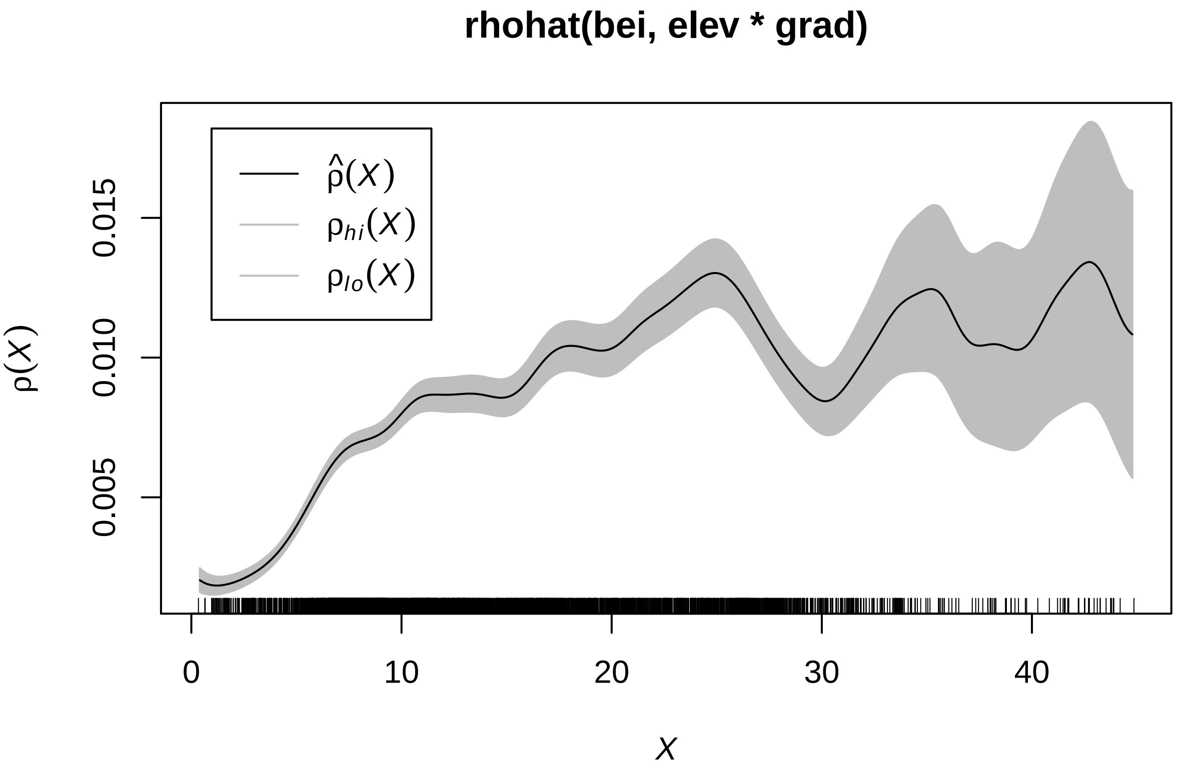
```
plot(rhohat(bei,elev))
```



The second rhohat graphs hows a more dramatic trend, with a low confidence beginning, but fairly tight error throughout the rest and two peaks, one at about 137(m) elevation and one around 150(m). It quickly drops after 150.

Based on both of these graphs I think there may be a relationship between the intensity of points and either of these covariates, but lets also look at the interaction between gradient and elevation.

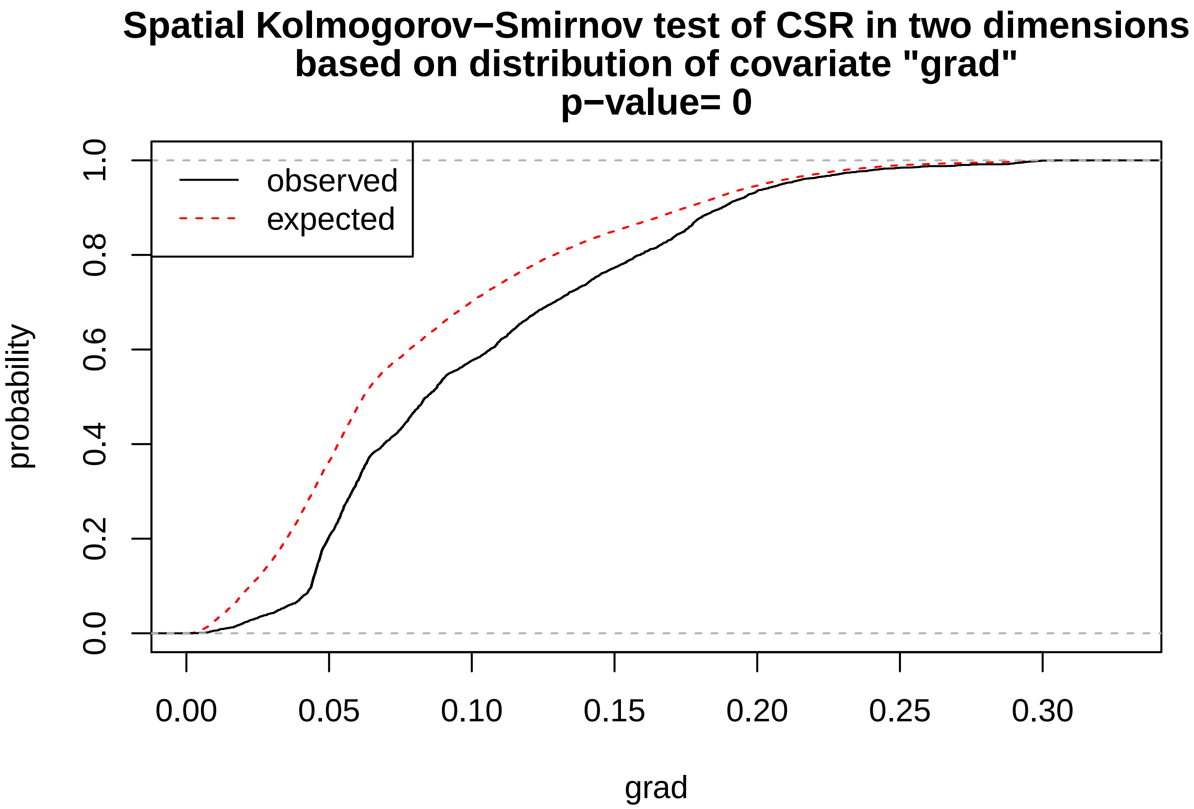
```
plot(rhohat(bei,elev*grad))
```



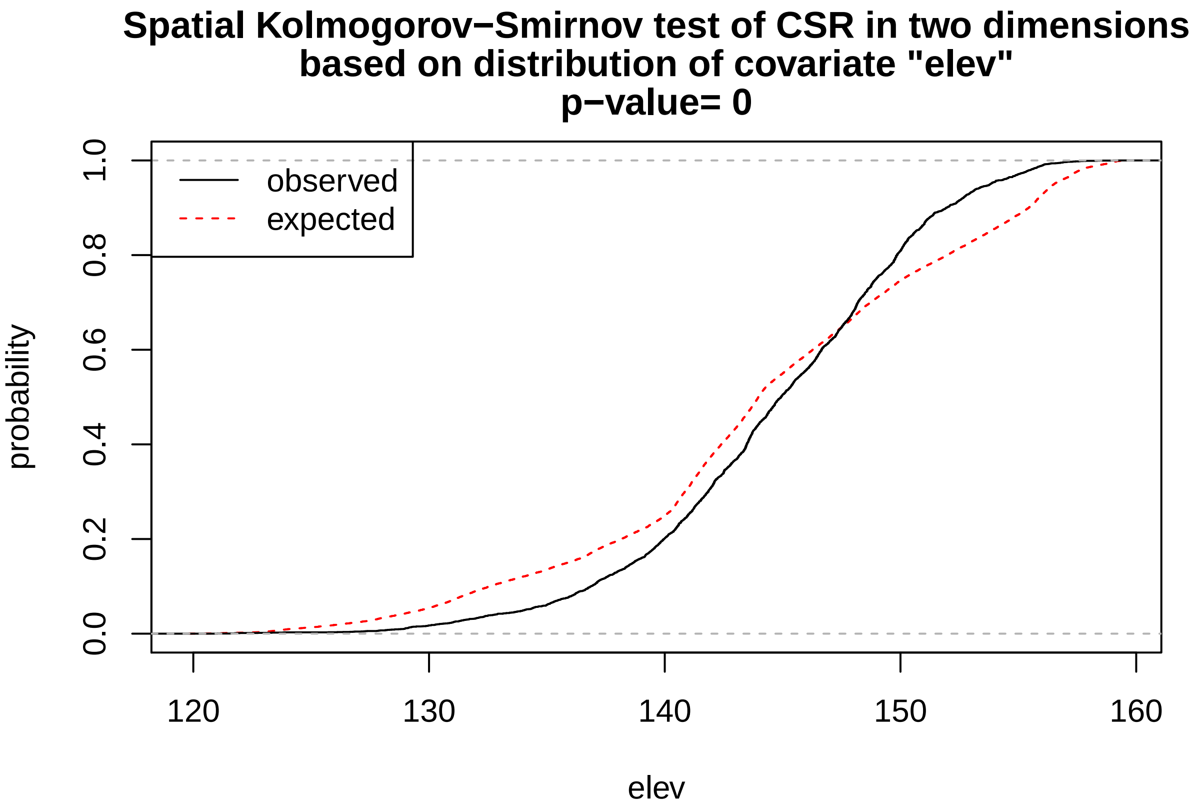
This third rhohat graph shows the combination of elevation and gradient, so the x axis is difficult to interpret, but with a higher value you would have higher gradient and elevation. So there is a fairly tight increasing relationship with gradient and elevation up till around 30, but then the error gets very large, and may be straight line or eve decreasing slightly. I think this gives us cause to include an interaction term in one of our models.

Before we get to actually doing a ppm() model we will do a K-S test to compare the curves of what the expected distribution of points would be for a given variable under CSR, and the curve of observed values. A small p-value indicates that the curves are significantly different.

```
plot(cdf.test(bei,grad,test="ks"))
```



```
plot(cdf.test(bei,elev,test="ks"))
```



For both KS tests the p-value is very small, suggesting that we can reject our null of complete spatial randomness and that the curves are significantly different.

The first curve (gradient) has the observed line under or at the expected line, suggesting there are less points at these values then would be expected.

the second curve (elevation) has values both above and below the line. When observed values are below the line there are fewer points then would be expected, when they are above, there are more then expected. Both K-S graphs show that either of these covariates could be having an impact on the intensity of points.

now that we have examined this in a variety of ways lets run the ppm() models.

We will run a null, five single covariate models and two models with multiple variables, including one with a gradient elevation interaction.

```
null <- ppm(bei)
grd <- ppm(bei ~ grad)
mv <- ppm(bei ~ v)
```

```
## Warning: Values of the covariate 'v' were NA or undefined at 0% (1 out of
## 20508) of the quadrature points. Occurred while executing: ppm.ppp(Q =
## bei, trend = ~v, data = NULL, interaction = NULL)
```

```
mv3 <- ppm(bei ~ v3)
```

```
## Warning: Values of the covariate 'v3' were NA or undefined at 0% (1 out of
## 20508) of the quadrature points. Occurred while executing: ppm.ppp(Q =
## bei, trend = ~v3, data = NULL, interaction = NULL)
```

```
mv2 <- ppm(bei ~ v2)
```

```
## Warning: Values of the covariate 'v2' were NA or undefined at 0% (1 out of
## 20508) of the quadrature points. Occurred while executing: ppm.ppp(Q =
## bei, trend = ~v2, data = NULL, interaction = NULL)
```

```
elv <- ppm(bei ~ elev)
grd.elv <- ppm(bei ~ elev * grad)
grd_elv <- ppm(bei ~ elev + grad)

aic <- matrix(ncol=2, nrow=8)
aic[,1] <- c("null","grd","mv","mv3","mv2","elv","grd.elv","grd_elv")
aic[,2] <- c(AIC(null),AIC(grd),AIC(mv),AIC(mv3),AIC(mv2),AIC(elv),AIC(grd.elv),AIC(grd_elv))

aic
```

```
##      [,1]      [,2]
## [1,] "null"    "42763.9195725373"
## [2,] "grd"     "42383.6647762241"
## [3,] "mv"      "41947.3838736782"
## [4,] "mv3"     "41946.2429944094"
## [5,] "mv2"     "42012.8445928219"
## [6,] "elv"     "42760.5117434741"
## [7,] "grd.elv" "42230.0652927119"
## [8,] "grd_elv" "42296.2095896515"
```

```
aic[aic[,2]==min(aic[,2]),]
```

```
## [1] "mv3"              "41946.2429944094"
```

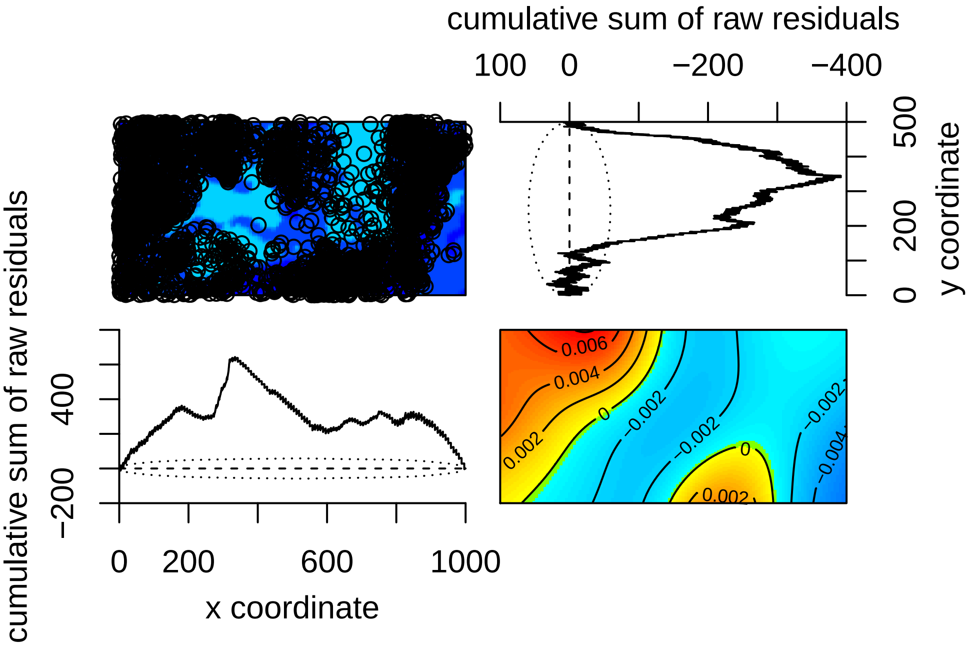
The model with the lowest AIC is the model with the mv3 covariate, or the categorical gradient variable with three categories (bottom 25%, 25-75% and upper 25%).

```
mv3
```

```
## Nonstationary Poisson process
##
## Log intensity:  ~v3
##
## Fitted trend coefficients:
##     (Intercept)  v3(0.04,0.113] v3(0.113,0.328]
##       -6.138162        1.309247        1.606064
##
##                   Estimate       S.E.   CI95.lo   CI95.hi Ztest       Zval
## (Intercept)     -6.138162 0.06074567 -6.257222 -6.019103   *** -101.04690
## v3(0.04,0.113]   1.309247 0.06470746  1.182422  1.436071   ***   20.23332
## v3(0.113,0.328]  1.606064 0.06668612  1.475362  1.736767   ***   24.08393
## Problem:
##  Values of the covariate 'v3' were NA or undefined at 0% (1 out of 20508)
## of the quadrature points
```

So this shows us that we have a significant positive relationship between two of the levels of our categorical surface, with one being a stronger relationship then the other. The intercept is also significant. The error is because for some reason one of the points is not on a part of the categorical surface that is defined, I tried to figure out why this is happening and haven't been able to fix it. Since it is 1 out of over 20000 points I decided it was excusable for the purposes of this exercise since it likely is not having an impact on our results.

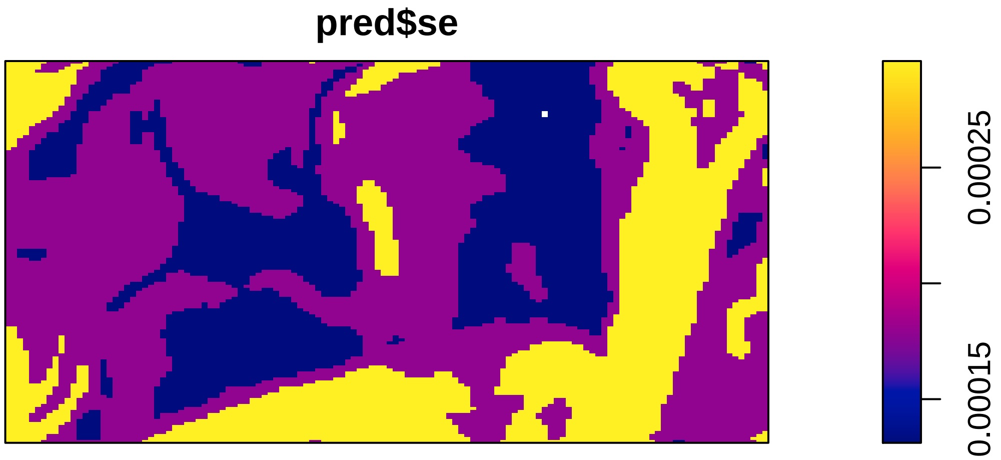
```
diagnose.ppm(mv3)
```



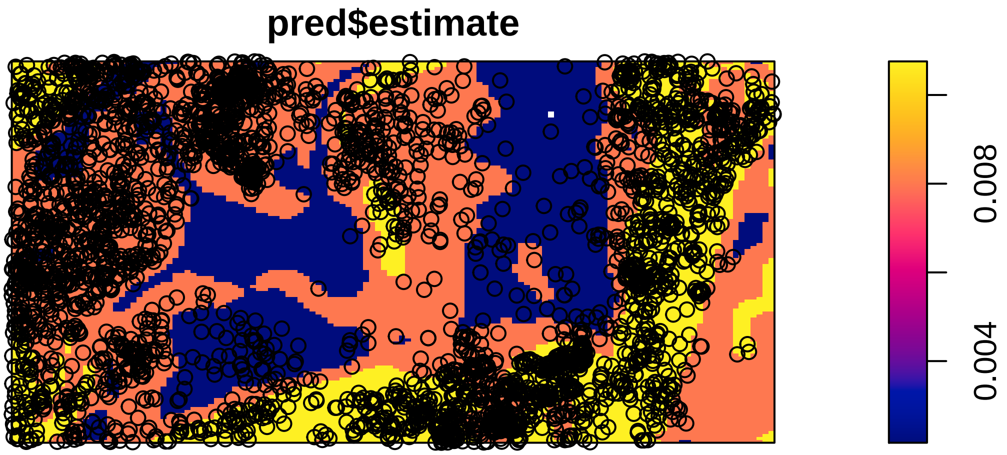
```
## Model diagnostics (raw residuals)
## Diagnostics available:
##  four-panel plot
##  mark plot
##  smoothed residual field
##   x cumulative residuals
##   y cumulative residuals
##   sum of all residuals
## sum of raw residuals in entire window = -3.117e-09
## area of entire window = 5e+05
## quadrature area = 5e+05
## range of smoothed field = [ -0.005887,0.008114 ]
```

These residuals are not normally distributed, suggesting that we have uneven error in our model or that we haven't taken everything into account. lets try predicting based on our model and see what kind of standard error we get.

```r
pred <- predict(mv3, se=T)

plot(pred$se)
```



```r
plot(pred$estimate)
plot(bei, add=T)
```



(note: on my home desktop I was unable to get the predict function to work, despite running the same version of R and all associated packages, it does run on my work desktop, I've got no explanation for this)

These graphs are a bit harder to interpret since they are categorical, but it does appear that the standard error varies with category, suggesting that perhaps this covariate really doesn't explain things so well. When we examine the prediction it looks like this might explain the variation in intensity, but I would want to examine other covariates, perhaps soil type or aspect of the ground to see if those explain the change in intensity.