Week 3 Assignment

Auriel M.V. Fournier
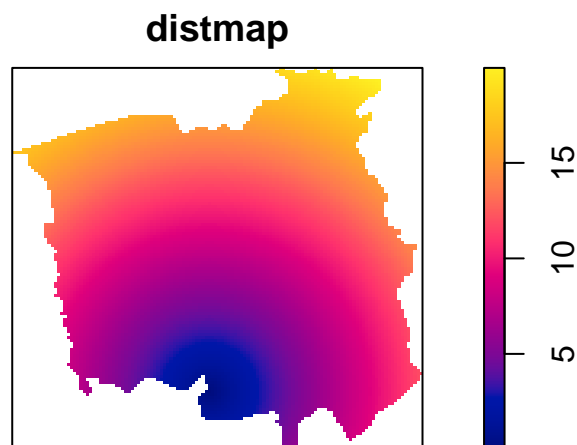
```r
library(spatstat)
```

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

```r
data(chorley)

ch <- unique.ppp(chorley)
ch_lung <- split(ch)$lung
ch_lar <- split(ch)$larynx
this.window <- ch$window
incin <- ppp(chorley.extra$incin$x, chorley.extra$incin$y, window=this.window)
distmap <- distmap(incin)

plot(distmap)
```
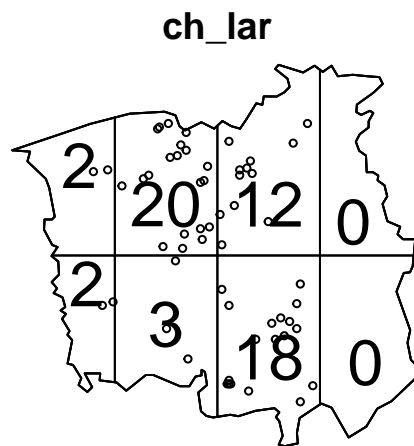

distmap

```r
lambda <- summary(ch)$intensity

q <- quadratcount(ch, nx=4,ny=2)
q_lung <- quadratcount(ch_lung, nx=4,ny=2)
q_lar <- quadratcount(ch_lar, nx=4,ny=2)

plot(ch_lar, cex=0.5)
plot(q_lar, add=T, cex=2)
```
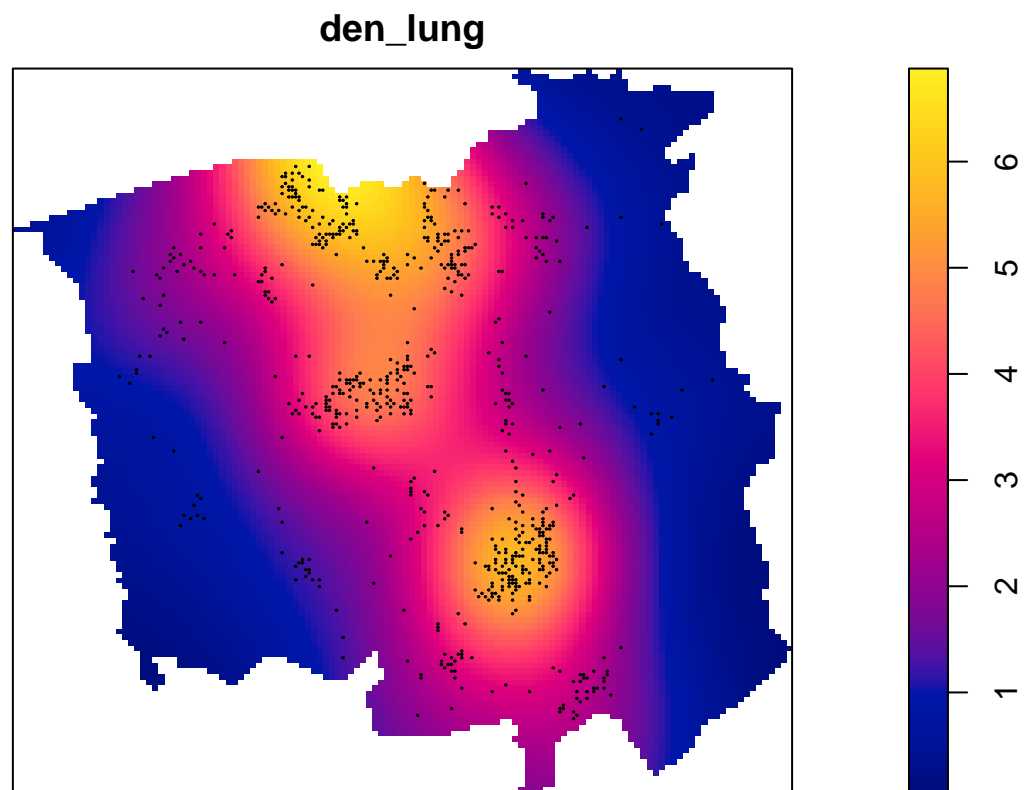
ch_lar


```r
#kde to create a surface predicting population over the
#entire surface using lung cancer occurances as a subsittute for population
den_lung <- density.ppp(ch_lung, sigma=2, kernel='gaussian')

plot(den_lung)
plot(ch_lung, add=T, cex=.1)
```
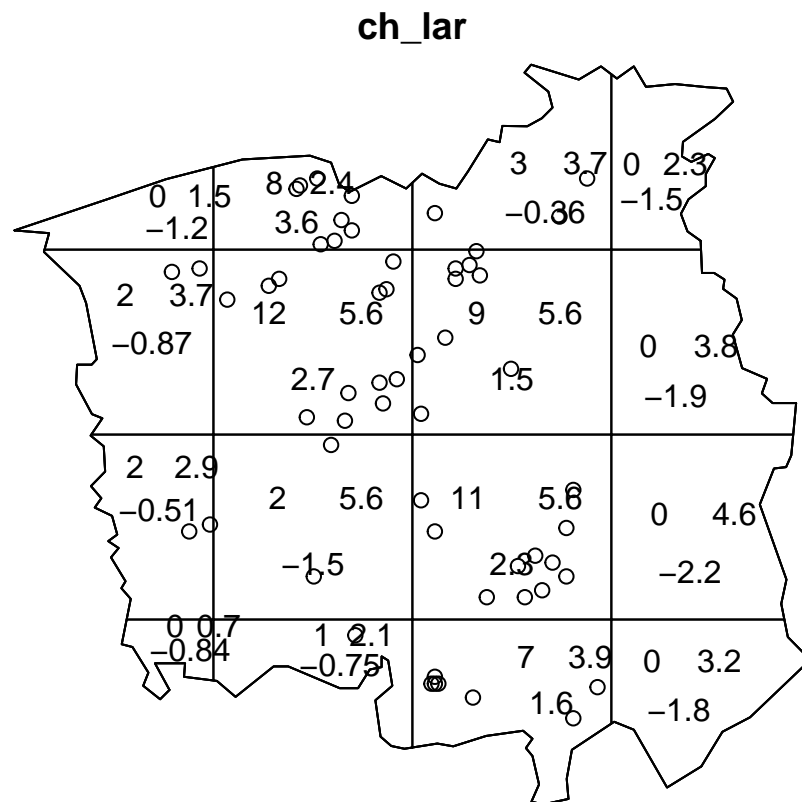
den_lung

```
csr.test <- quadrat.test(ch_lar, nx=4, ny=4, method="Chisq")
```

```
## Warning: Some expected counts are small; chi^2 approximation may be
## inaccurate
```

```
plot(ch_lar, color='blue')
plot(csr.test, add=T, color="red")
```

ch_lar

0 1.5 −1.2   8 ○2.4 3.6   3 3.7 −0.36   0 2.3 −1.5

2 3.7 −0.87   12 5.6 2.7   9 5.6 1.5   0 3.8 −1.9

2 2.9 −0.51   2 5.6 −1.5   11 5.6   0 4.6 −2.2

0 0.7 −0.84   1 2.1 −0.75   7 3.9 1.6   0 3.2 −1.8

```
csr.test
```

```
## 
##  Chi-squared test of CSR using quadrat counts
##  Pearson X2 statistic
## 
## data:  ch_lar
## X2 = 50.2628, df = 15, p-value = 2.181e-05
## alternative hypothesis: two.sided
## 
## Quadrats: 16 tiles (irregular windows)
```

So the p-value is really small, meaning that they are not Completely Spatially Random.

```
null <- ppm(ch_lar)
dist <- ppm(ch_lar ~distmap)
pop <- ppm(ch_lar ~den_lung)
pop_dist <- ppm(ch_lar ~den_lung+distmap)
pop.dist <- ppm(ch_lar ~den_lung*distmap)

aic <- matrix(ncol=2, nrow=5)
aic[,1] <- c("null","dist","pop","pop_dist","pop.dist")
aic[,2] <- c(AIC(null),AIC(dist),AIC(pop),AIC(pop_dist),AIC(pop.dist))

aic
```

```
##        [,1]       [,2]
## [1,] "null"     "310.941626252374"
## [2,] "dist"     "311.231896754453"
## [3,] "pop"      "250.170542454793"
## [4,] "pop_dist" "249.237138463892"
## [5,] "pop.dist" "250.562620405332"
```

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

```
## [1] "pop_dist"          "249.237138463892"
```

Based on the AIC of these models the model with 328.255 is the lowest and is for the pop_dist model.

First lets just look at the model output for pop_dist

```
pop_dist
```

```
## Nonstationary Poisson process
##
## Log intensity:  ~den_lung + distmap
##
## Fitted trend coefficients:
## (Intercept)    den_lung     distmap
## -3.19622651  0.63771937 -0.05733377
##
##                 Estimate        S.E.     CI95.lo      CI95.hi Ztest       Zval
## (Intercept) -3.19622651 0.42485713 -4.0289312 -2.363521828   *** -7.523062
## den_lung     0.63771937 0.08669755  0.4677953  0.807643456   ***  7.355679
## distmap     -0.05733377 0.03346583 -0.1229256  0.008258055       -1.713203
```

```
diagnose.ppm(pop_dist)
```

cumulative sum of raw residuals
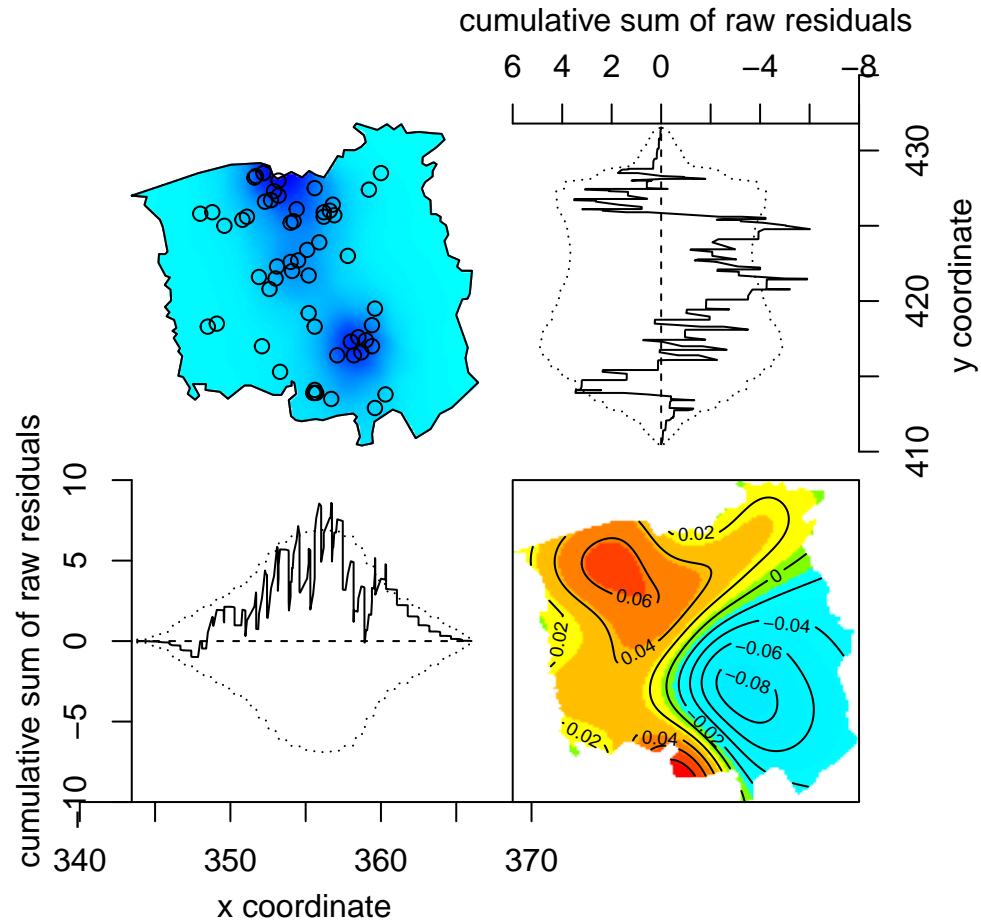
```
## Model diagnostics (raw residuals)
## Diagnostics available:
##   four-panel plot
##   mark plot
##   smoothed residual field
##   x cumulative residuals
##   y cumulative residuals
##   sum of all residuals
## sum of raw residuals in entire window = -1.455e-06
## area of entire window = 315.2
## quadrature area = 315.3
```

```
## range of smoothed field = [ -0.09258,0.0987 ]
```

So there is a positive response to den_lung (our proxy for population) so it looks like there is more occurances of larynx where there are more people. This is also how the standard error is distributed, meaning that our model doesn't fit very well.

the pop_dist model includes the distmap covariate, but it is not significant. This indicates that there is not a significant relationship between distnace from the incernartor and intensity of larynx cancer cases.

If we look at the model with only population we see that it has a slightly higher AIC, but it's with 2 Delta AIC, so that is another plausible model. It has a similar positive value for it's parameter estimate, a slightly lower value, but similar and still significant.

From this result I would say that larynx cancer rates are not related to distance from the incinerator, but are occuring at a certain rate within the population, perhaps related to some other factor that we do not have the data to understand right now.