Case Study 2, Part 3

Peter Claussen

10/3/2017

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(rsm)
library(ncf)
library(ape)
```

```
## Warning: package 'ape' was built under R version 3.3.2
```

```
##
## Attaching package: 'ape'
```

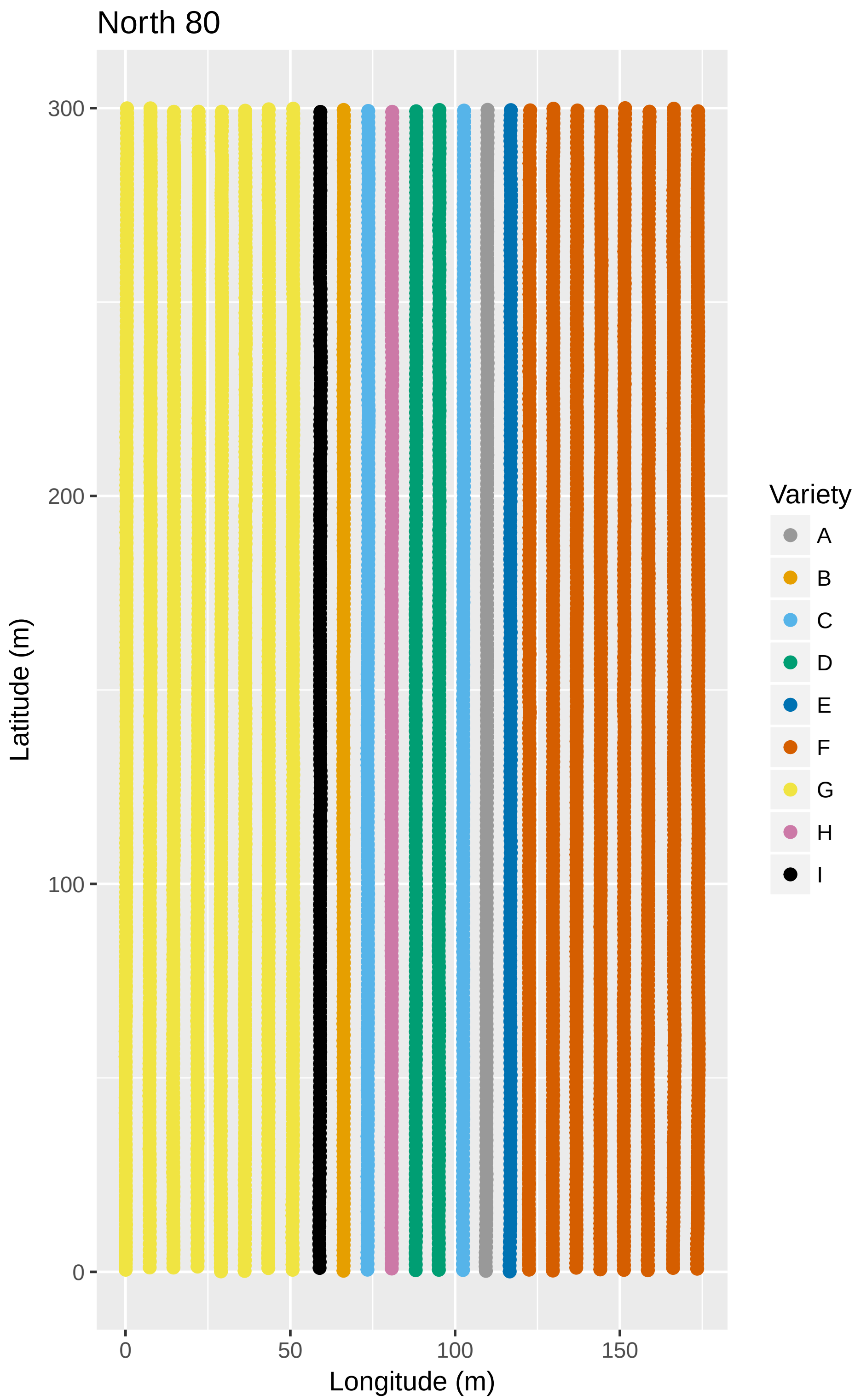
```
## The following object is masked from 'package:ncf':
##
##     mantel.test
```

```
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#0072B2", "#D55E00", "#F0E442","#CC79A7","#0
```

```
load(file="Strips.Rda")
```

North 80

```
ggplot(North80.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Product),size=2) +
scale_colour_manual(values=cbPalette) +
labs(colour = "Variety", x="Longitude (m)", y="Latitude (m)", title = "North 80")
```



We repeat the process from Part 2

Trend Surface

```
Yield7.lm <- lm(Yield ~ poly(Easting, Northing, degree=7), data=North80.dat)
Yield9.lm <- lm(Yield ~ poly(Easting, Northing, degree=9), data=North80.dat)
Yield11.lm <- lm(Yield ~ poly(Easting, Northing, degree=11), data=North80.dat)
Yield13.lm <- lm(Yield ~ poly(Easting, Northing, degree=13), data=North80.dat)
```

```
anova(Yield7.lm, Yield9.lm, Yield11.lm, Yield13.lm)
```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ poly(Easting, Northing, degree = 7)
## Model 2: Yield ~ poly(Easting, Northing, degree = 9)
## Model 3: Yield ~ poly(Easting, Northing, degree = 11)
## Model 4: Yield ~ poly(Easting, Northing, degree = 13)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   5461 6670821
## 2   5442 5692132 19    978690 64.751 < 2.2e-16 ***
## 3   5419 4874319 23    817813 44.697 < 2.2e-16 ***
## 4   5392 4289416 27    584902 27.232 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(Yield7.lm)
```

```
## [1] 54709.63
```

```
AIC(Yield9.lm)
```

```
## [1] 53875.49
```

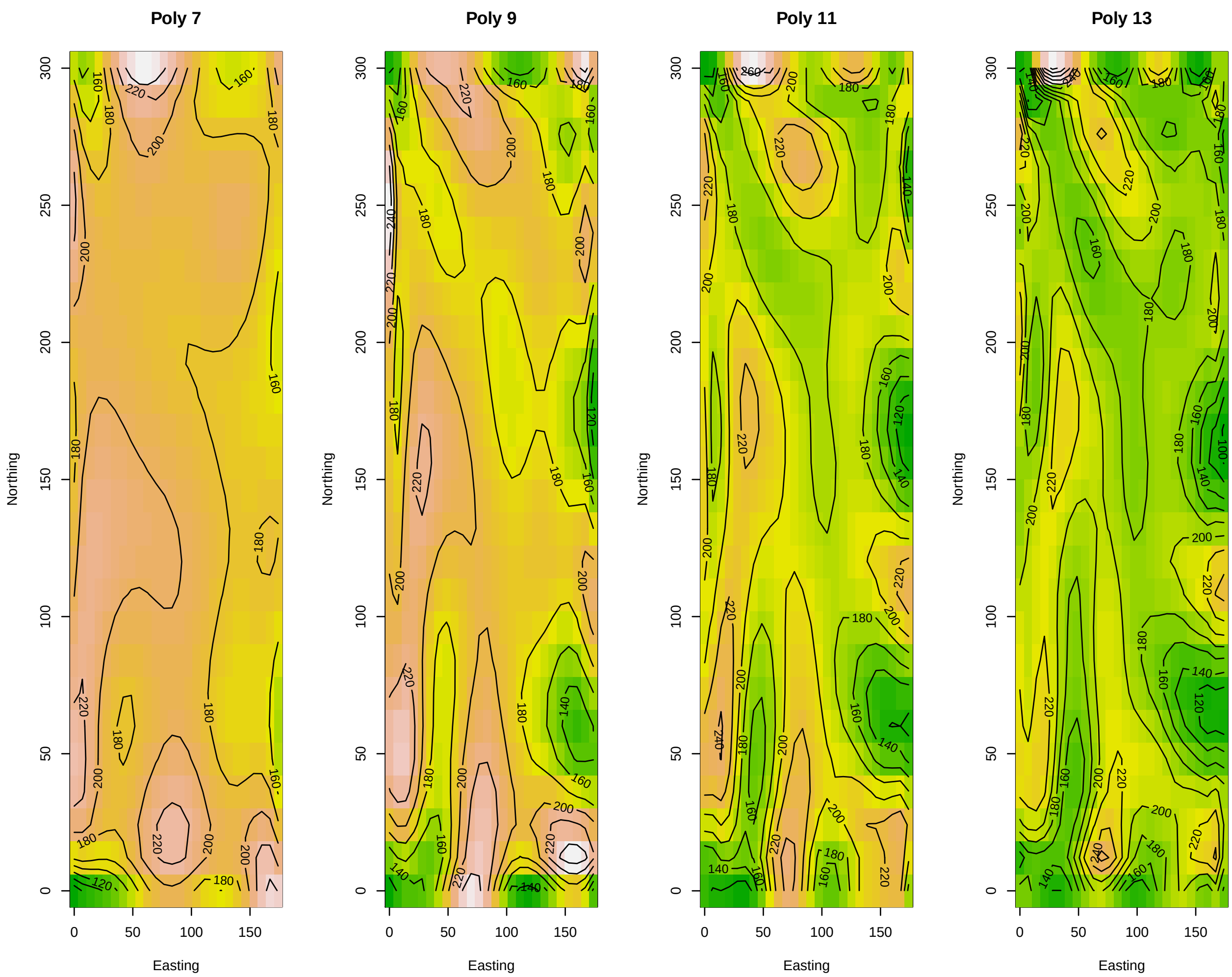
```
AIC(Yield11.lm)
```

```
## [1] 53068.88
```

```
AIC(Yield13.lm)
```

```
## [1] 52420.2
```

```
par(mfrow=c(1,4))
contour(Yield7.lm, Northing ~ Easting, image = TRUE,main="Poly 7")
contour(Yield9.lm, Northing ~ Easting, image = TRUE,main="Poly 9")
contour(Yield11.lm, Northing ~ Easting, image = TRUE,main="Poly 11")
contour(Yield13.lm, Northing ~ Easting, image = TRUE,main="Poly 13")
```




```r
par(mfrow=c(1,1))
```

I'm concerned that Poly 11 might be overfitting in the upper left corner, so we'll choose Poly 9 (note - we could continue with diagnostics of Local I to check that assertion).

```r
North80.dat$Yield9.resid <- Yield9.lm$residuals
North80.dat$Yield11.resid <- Yield11.lm$residuals
North80.dat$Yield13.resid <- Yield13.lm$residuals
```

Check for white noise

```r
Distance.mat <- as.matrix(dist(cbind(North80.dat$Easting, North80.dat$Northing)))
Distance.mat <- 1/Distance.mat
diag(Distance.mat) <- 0
```

```r
print(MoranYield <-Moran.I(North80.dat$Yield, Distance.mat))
```

```
## $observed
## [1] 0.08167636
##
## $expected
```

```
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371498
##
## $p.value
## [1] 0
```

```r
print(Moran9Resid <-Moran.I(North80.dat$Yield9.resid, Distance.mat))
```

```
## $observed
## [1] 0.04376936
##
## $expected
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371369
##
## $p.value
## [1] 0
```

```r
print(Moran11Resid <-Moran.I(North80.dat$Yield11.resid, Distance.mat))
```

```
## $observed
## [1] 0.03711791
##
## $expected
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371314
##
## $p.value
## [1] 0
```

```r
print(Moran13Resid <-Moran.I(North80.dat$Yield13.resid, Distance.mat))
```

```
## $observed
## [1] 0.0317406
##
## $expected
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371302
##
## $p.value
## [1] 0
```

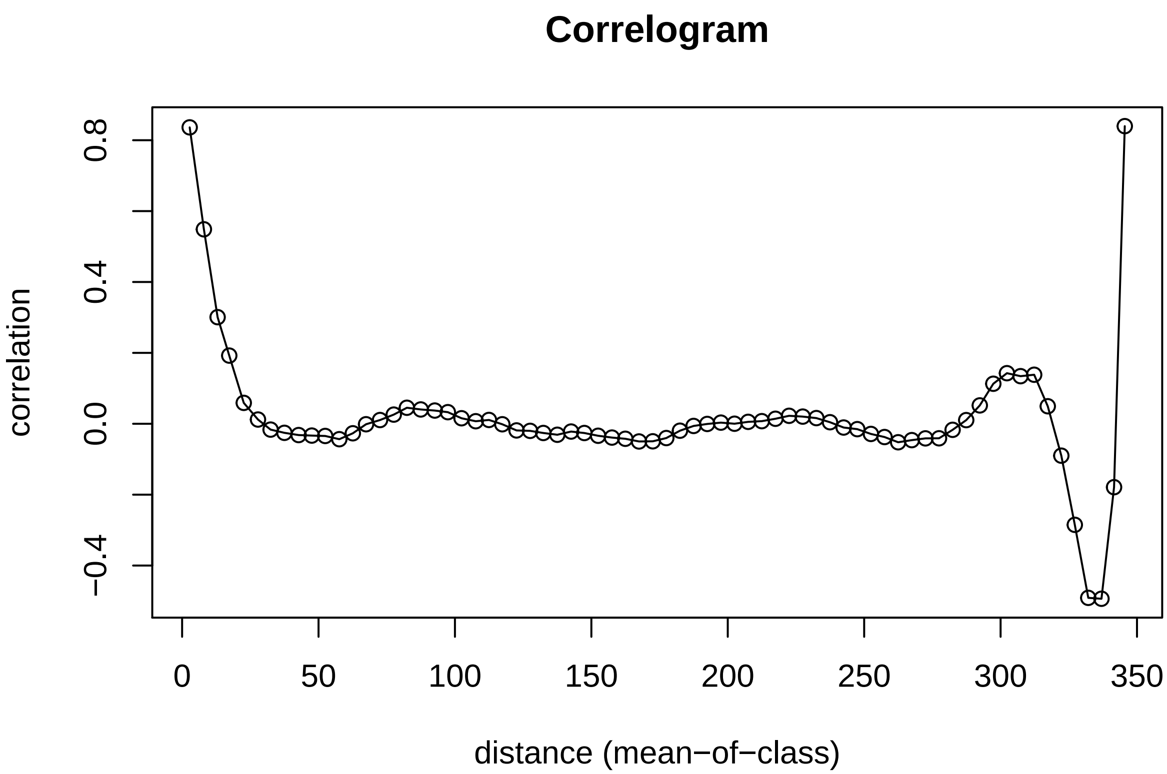
Local Correlation

Resampling can take a very long time, so I'll use a flag to control whether we resample or simply plot local measures. Resampling is needed for p-values; I'm not concerned about p-values for the points in the correlogram.

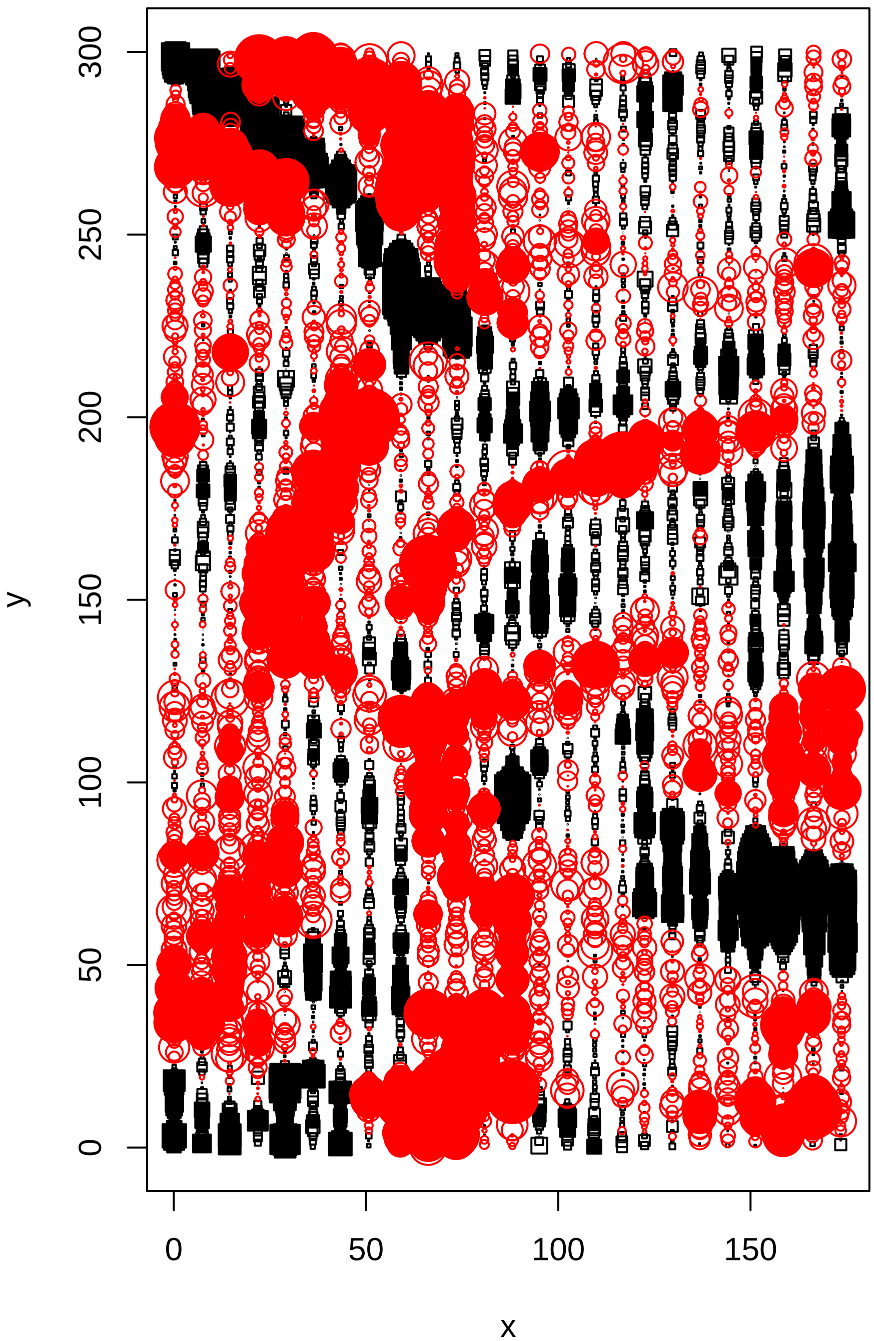
```r
resample = 100
```

Correlation in Yield

```r
Yield.clg <- correlog(North80.dat$Easting, North80.dat$Northing, North80.dat$Yield,
                      increment=5, resamp=0, quiet=TRUE)
plot(Yield.clg)
```

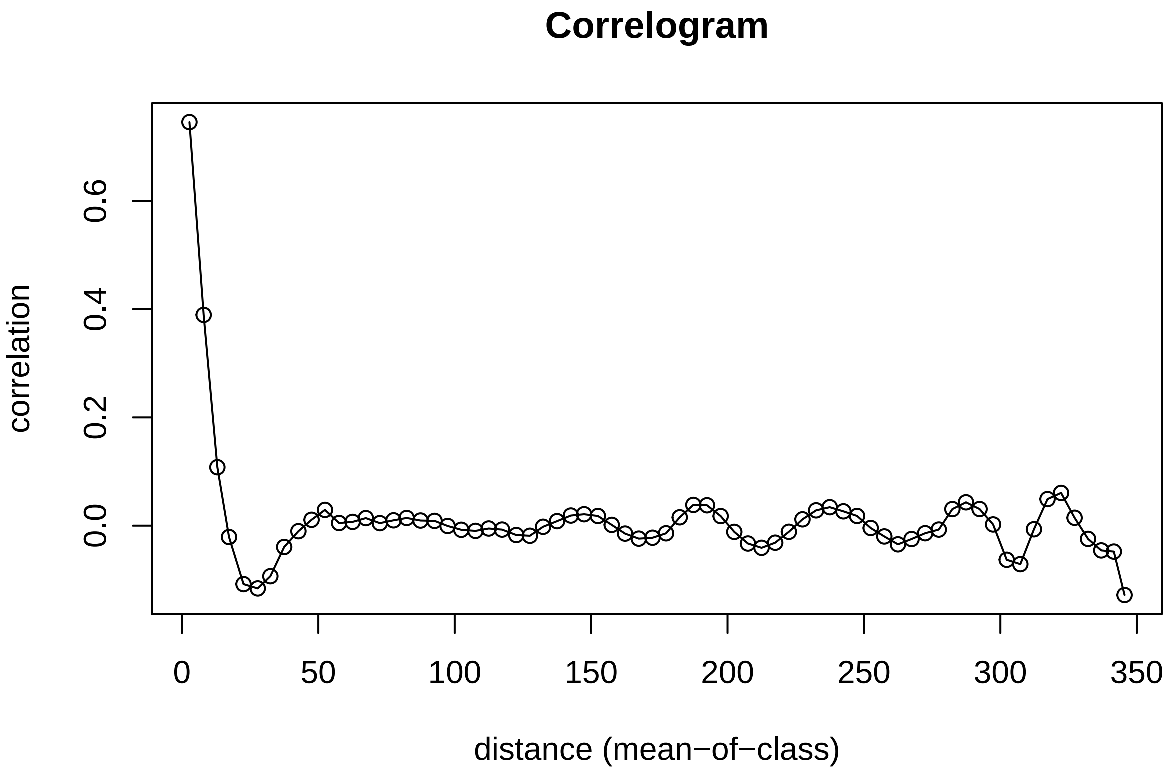



```r
Yield.lisa <- lisa(North80.dat$Easting, North80.dat$Northing, North80.dat$Yield,
                   neigh=5, resamp=resample, quiet=TRUE)
plot.lisa(Yield.lisa, negh.mean=FALSE)
```

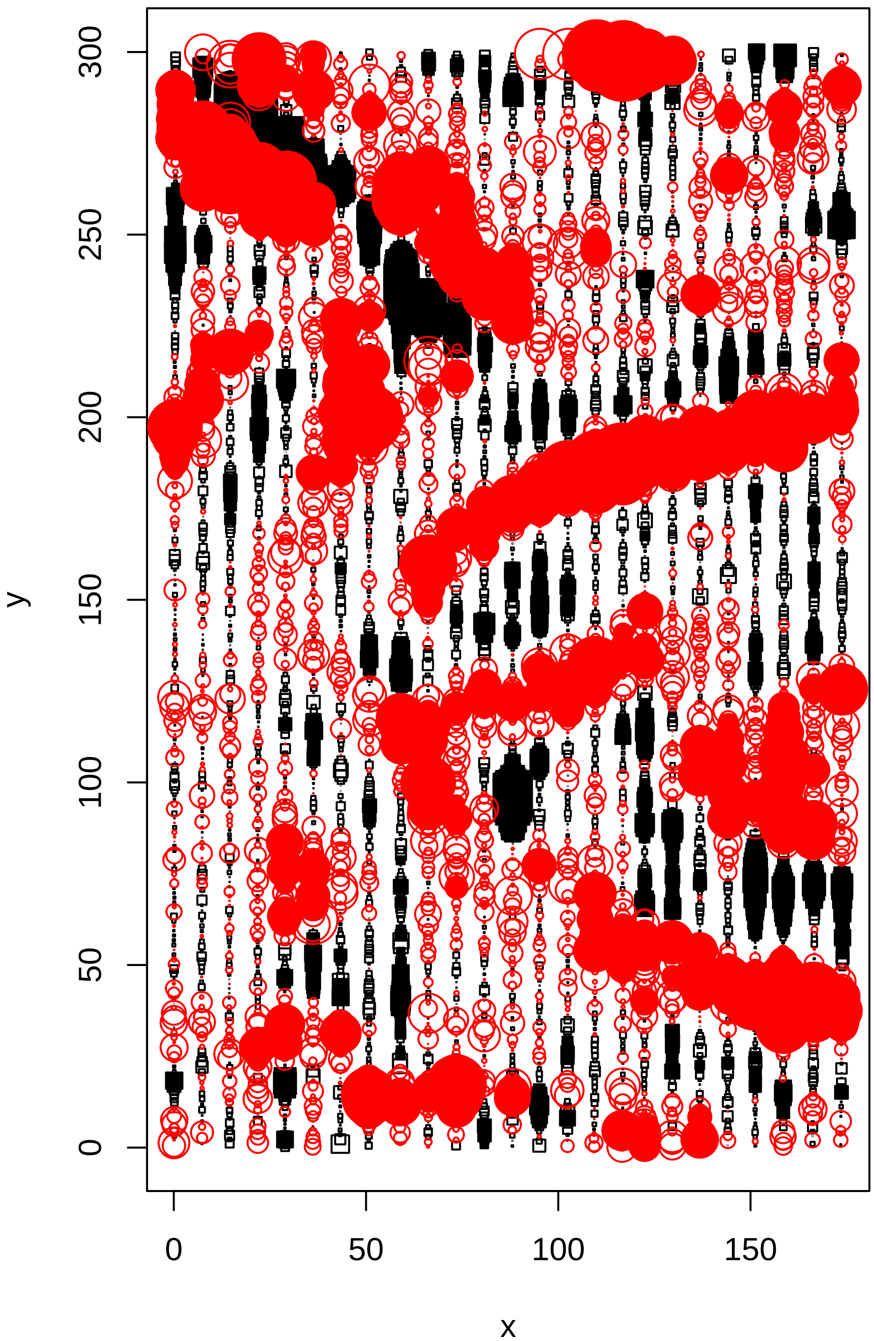


Correlation in Yield Residuals

```
Yield9.resid.clg <- correlog(North80.dat$Easting, North80.dat$Northing, North80.dat$Yield9.resid,
                             increment=5, resamp=0, quiet=TRUE)
plot(Yield9.resid.clg)
```

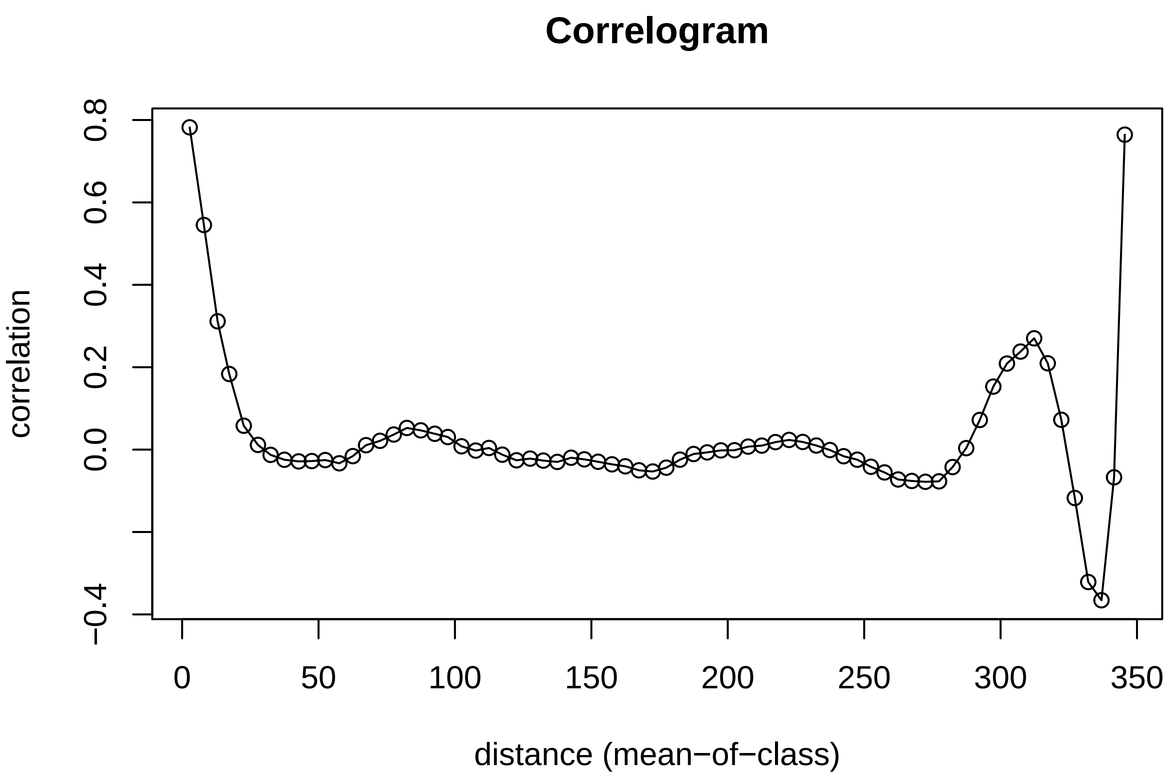



```r
Yield9.resid.lisa <- lisa(North80.dat$Easting, North80.dat$Northing, North80.dat$Yield9.resid,
                          neigh=5, resamp=resample, quiet=TRUE)
plot.lisa(Yield9.resid.lisa, negh.mean=FALSE)
```

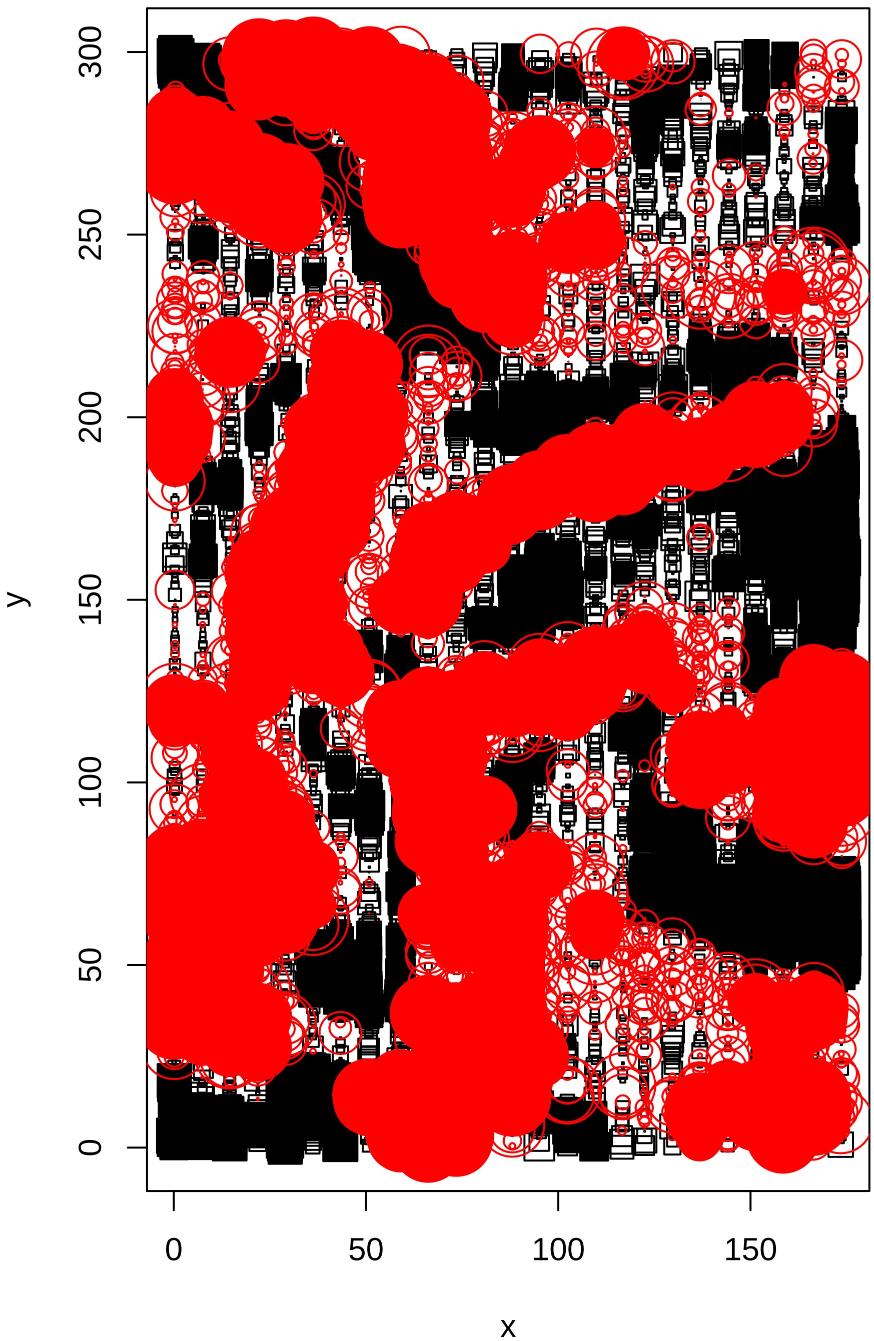


Correlation in Quantiles

```
Quantile.clg <- correlog(North80.dat$Easting, North80.dat$Northing, North80.dat$Quantile,
                         increment=5, resamp=0, quiet=TRUE)
plot(Quantile.clg)
```




```
Quantile.lisa <- lisa(North80.dat$Easting, North80.dat$Northing, North80.dat$Quantile,
                      neigh=5, resamp=resample, quiet=TRUE)
plot.lisa(Quantile.lisa, negh.mean=FALSE)
```



Variety G was planted in the first 8 passes from the left (west), then variety I. These where the two lowest yielding varieties, based on the trend adjusted yields. Unadjusted yield places variety G in the middle.

Variety F was planted in the rightmost 8 passes and is the lowest yielding variety based on unadjusted yields. In the adjusted yields, F ranges from mid-rank to the highest ranking variety (Poly 13). The large separation between F and G in the Poly 13 model suggests to me that this model over-fits the data.

One of the reasons I choose the Poly 9 model is that it produces the least separation among varieties. This is more plausible to me, since all varieties should be consider elite for the region. Perhaps I'm being too conservative in choosing a spatial model, but I'd like to look for stronger evidence that my spatial model is correct, before choosing a less conservative mean separation.

Variety A is identified as the highest yielding variety in the Poly 9 model. This is the second variety from the right, between C and E (adjacent to F), with C just along the 100m line. # Trend + Variety AOV

Does including variety (`Product`) in the model improve spatial correlation?

```
YieldVariety7.lm <- lm(Yield ~ poly(Easting, Northing, degree=7) + Product, data=North80.dat)
summary(aov(YieldVariety7.lm))
```

```
##                                       Df  Sum Sq Mean Sq F value Pr(>F)
## poly(Easting, Northing, degree = 7)   35 1593734   45535   38.13 <2e-16
## Product                                8  158753   19844   16.62 <2e-16
## Residuals                           5453 6512068    1194
##
## poly(Easting, Northing, degree = 7) ***
## Product                             ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
YieldVariety9.lm <- lm(Yield ~ poly(Easting, Northing, degree=9) + Product, data=North80.dat)
summary(aov(YieldVariety9.lm))
```

```
##                                       Df  Sum Sq Mean Sq F value Pr(>F)
## poly(Easting, Northing, degree = 9)   54 2572424   47637   46.37 <2e-16
## Product                                8  110078   13760   13.39 <2e-16
## Residuals                           5434 5582053    1027
##
## poly(Easting, Northing, degree = 9) ***
## Product                             ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
YieldVariety11.lm <- lm(Yield ~ poly(Easting, Northing, degree=11) + Product, data=North80.dat)
summary(aov(YieldVariety11.lm))
```

```
##                                        Df  Sum Sq Mean Sq F value Pr(>F)
## poly(Easting, Northing, degree = 11)   77 3390237   44029   49.94 <2e-16
## Product                                 8  103636   12955   14.69 <2e-16
## Residuals                            5411 4770682     882
##
## poly(Easting, Northing, degree = 11) ***
## Product                              ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
YieldVariety13.lm <- lm(Yield ~ poly(Easting, Northing, degree=13) + Product, data=North80.dat)
summary(aov(YieldVariety13.lm))
```

```
##                                        Df  Sum Sq Mean Sq F value Pr(>F)
## poly(Easting, Northing, degree = 13)  104 3975139   38222   49.19 <2e-16
## Product                                 8  105943   13243   17.04 <2e-16
## Residuals                            5384 4183473     777
##
## poly(Easting, Northing, degree = 13) ***
## Product                              ***
## Residuals
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

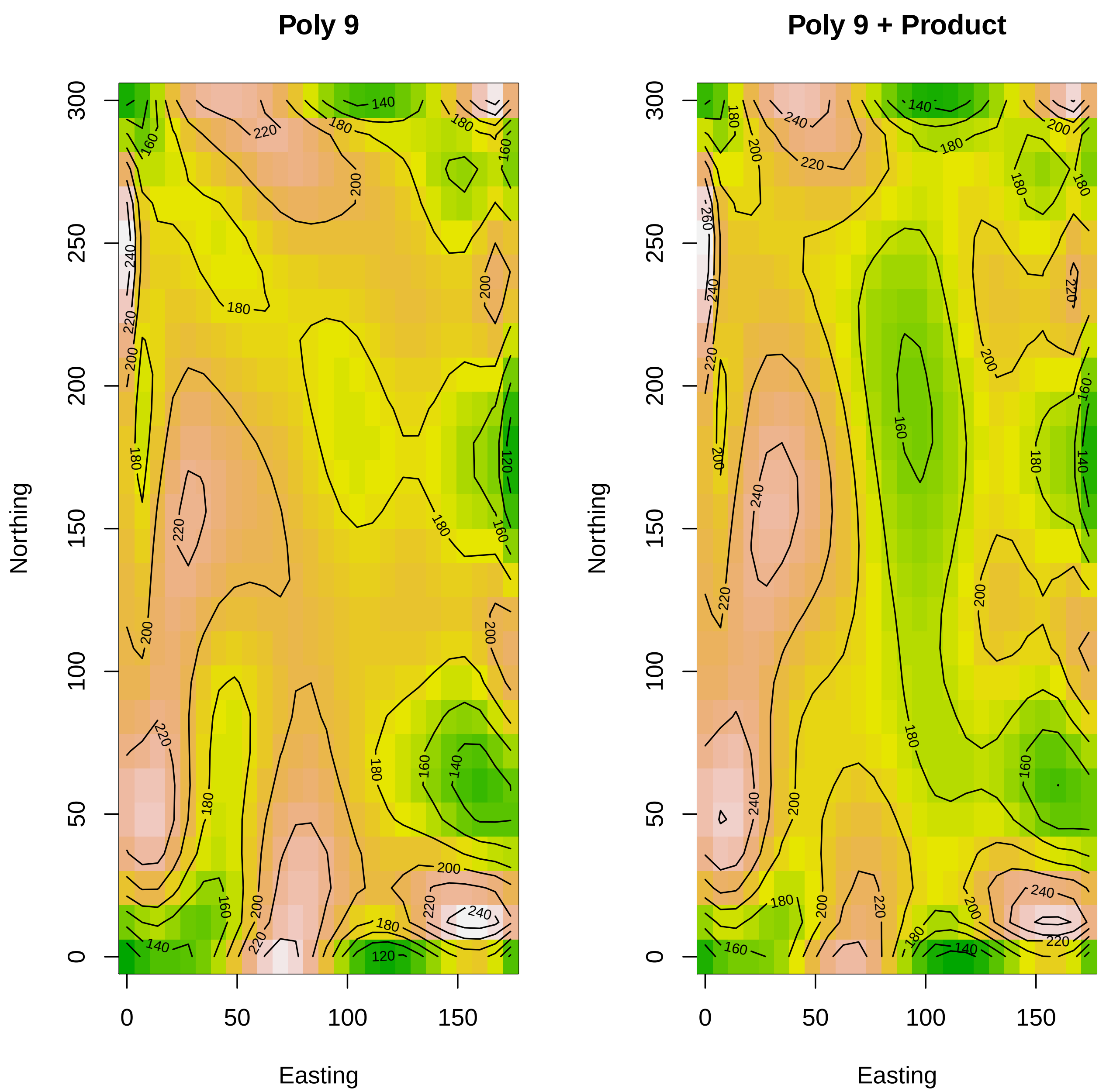
```r
anova(YieldVariety7.lm,YieldVariety9.lm,YieldVariety11.lm,YieldVariety13.lm)
```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ poly(Easting, Northing, degree = 7) + Product
## Model 2: Yield ~ poly(Easting, Northing, degree = 9) + Product
## Model 3: Yield ~ poly(Easting, Northing, degree = 11) + Product
## Model 4: Yield ~ poly(Easting, Northing, degree = 13) + Product
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   5453 6512068
## 2   5434 5582053 19    930014 62.995 < 2.2e-16 ***
## 3   5411 4770682 23    811371 45.400 < 2.2e-16 ***
## 4   5384 4183473 27    587209 27.990 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
North80.dat$YieldVariety9.resid <- YieldVariety9.lm$residuals
print(Moran9Variety <- Moran.I(North80.dat$YieldVariety9.resid, Distance.mat))
```

```
## $observed
## [1] 0.04391537
##
## $expected
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371369
##
## $p.value
## [1] 0
```

```r
par(mfrow=c(1,2))
contour(Yield9.lm, Northing ~ Easting, image = TRUE,main="Poly 9")
contour(YieldVariety9.lm, Northing ~ Easting, image = TRUE,main="Poly 9 + Product")
```



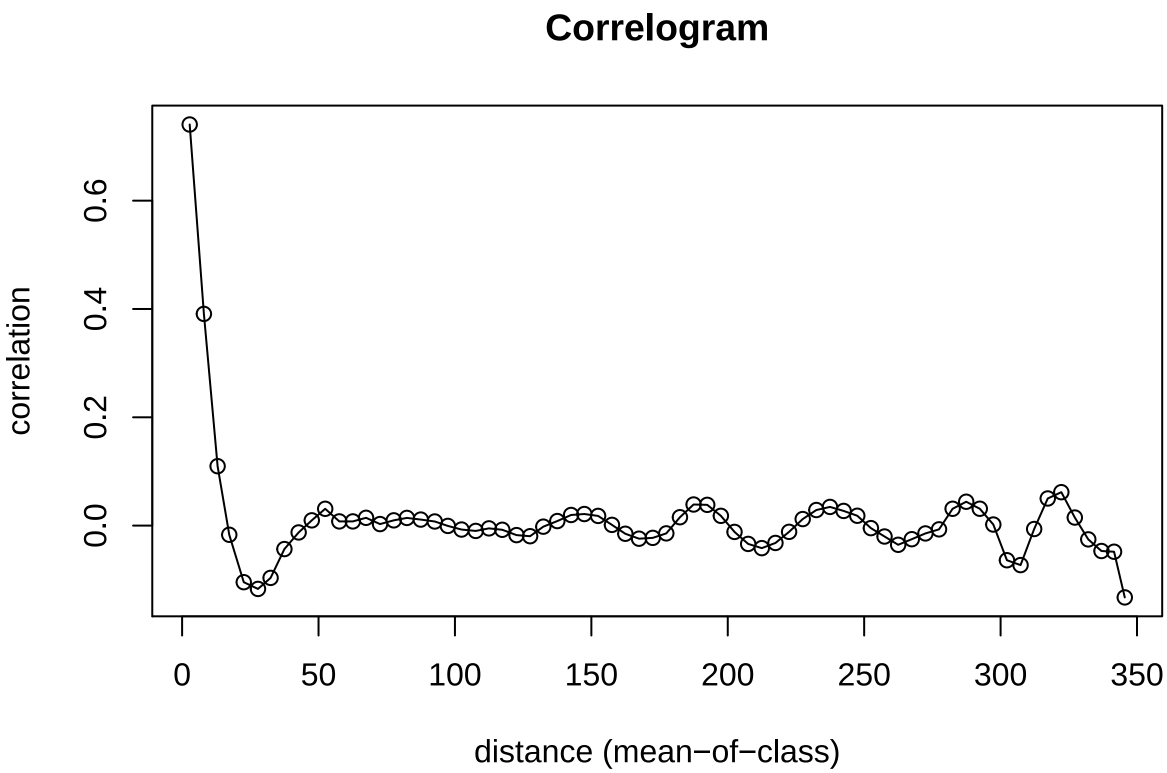

```
par(mfrow=c(1,1))
```

```
North80.dat$YieldVariety9.resid <- YieldVariety9.lm$residuals
print(Moran9Variety <-Moran.I(North80.dat$YieldVariety9.resid, Distance.mat))
```

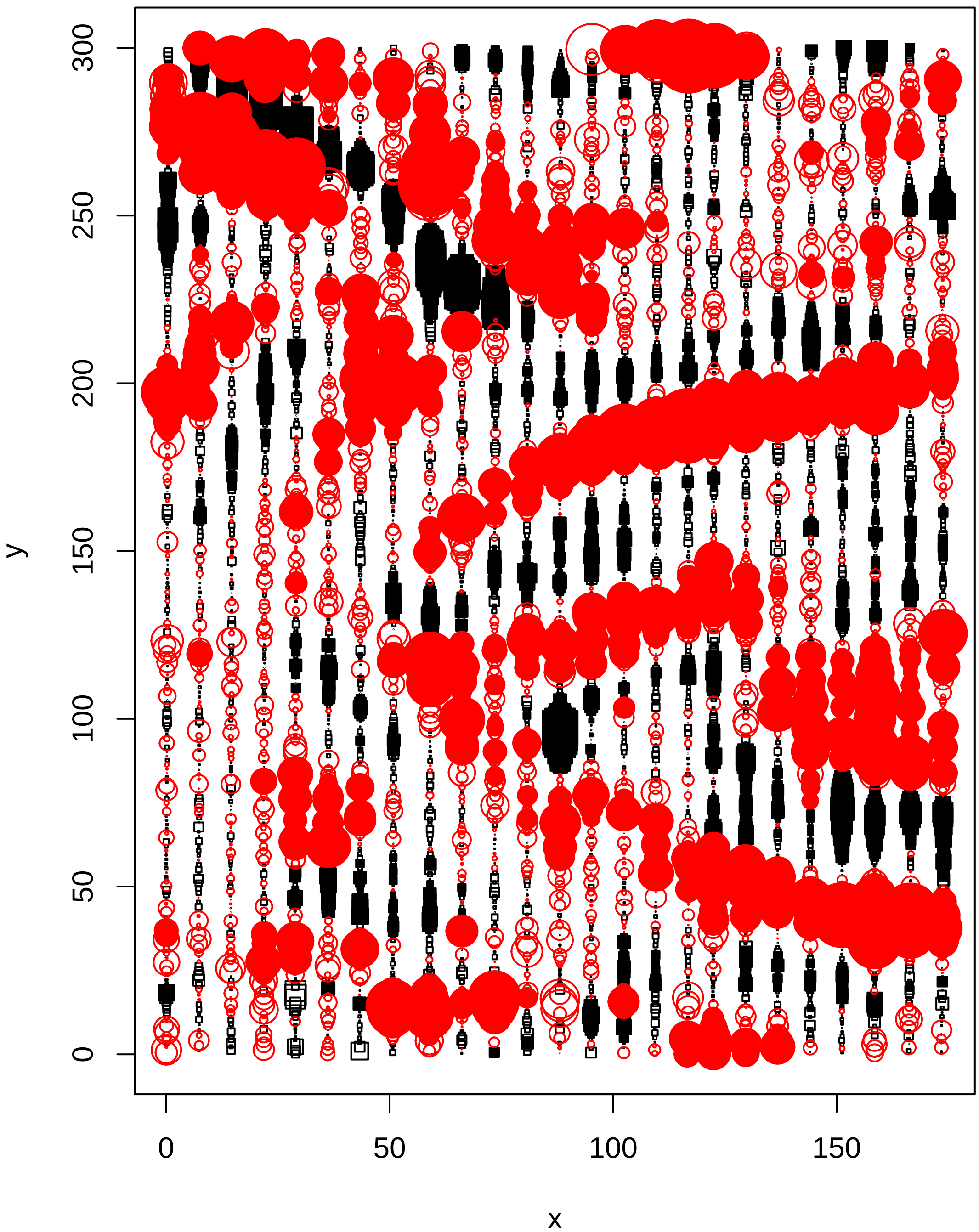
```
## $observed
## [1] 0.04391537
##
## $expected
## [1] -0.0001819505
##
## $sd
## [1] 0.0004371369
##
## $p.value
```

```
## [1] 0
```

```
YieldVariety9.resid.clg <- correlog(North80.dat$Easting, North80.dat$Northing, North80.dat$YieldVariety9
                                    increment=5, resamp=0, quiet=TRUE)
plot(YieldVariety9.resid.clg)
```




```
YieldVariety9.resid.lisa <- lisa(North80.dat$Easting, North80.dat$Northing, North80.dat$YieldVariety9.re
                                  neigh=10, resamp=resample, quiet=TRUE)
plot.lisa(YieldVariety9.resid.lisa, negh.mean=FALSE)
```



Trend AOV

```
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 3.3.2

## Loading required package: estimability

## Warning: package 'estimability' was built under R version 3.3.2
```

```r
Yield.lm <- lm(Yield ~ Product, data=North80.dat)
lsmeans(Yield.lm, cld ~ Product)
```

```
##  Product   lsmean        SE   df lower.CL upper.CL .group
##  F       178.3901 0.8897603 5488 176.6458 180.1344  1
##  I       182.3015 2.6352299 5488 177.1354 187.4676  12
##  E       186.8105 2.6611933 5488 181.5936 192.0275  123
##  D       189.2594 1.8411785 5488 185.6499 192.8688   23
##  A       191.3857 2.6100118 5488 186.2690 196.5023   234
##  G       193.2407 0.8890226 5488 191.4979 194.9836    34
##  C       199.0592 1.8544087 5488 195.4239 202.6946     4
##  H       200.6945 2.6038196 5488 195.5900 205.7990     45
##  B       210.2030 2.5915662 5488 205.1225 215.2835      5
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```

```r
lsmeans(YieldVariety7.lm, cld ~ Product)
```

```
##  Product   lsmean       SE   df lower.CL upper.CL .group
##  G       146.8787 7.535718 5453 132.1057 161.6517  1
##  I       152.1395 5.283167 5453 141.7824 162.4966  1
##  B       187.5922 4.161418 5453 179.4342 195.7503   2
##  D       190.7688 2.245838 5453 186.3660 195.1715   2
##  H       193.2256 2.864528 5453 187.6100 198.8412   2
##  F       193.8912 6.841786 5453 180.4785 207.3038   2
##  C       195.9567 2.403346 5453 191.2452 200.6682   2
##  E       199.5084 5.183623 5453 189.3465 209.6704   2
##  A       202.1578 4.152016 5453 194.0182 210.2974   2
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```

```r
lsmeans(YieldVariety9.lm, cld ~ Product)
```

```
##  Product   lsmean        SE   df lower.CL upper.CL .group
##  I       145.6522 10.080973 5434 125.8894 165.4149  1
##  G       148.7229 14.196358 5434 120.8924 176.5535  123
##  F       157.5931 12.308523 5434 133.4634 181.7227  12 4
##  E       173.9486  9.347166 5434 155.6244 192.2728    3 5
##  B       177.3052  6.964079 5434 163.6529 190.9576     45
##  H       183.7045  2.974938 5434 177.8724 189.5366   2345
##  D       183.7200  2.379733 5434 179.0548 188.3853   2345
##  C       185.5255  3.651721 5434 178.3667 192.6844     45
##  A       185.8440  6.548976 5434 173.0054 198.6826      5
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```

```r
lsmeans(YieldVariety11.lm, cld ~ Product)
```

```
##  Product   lsmean        SE   df  lower.CL upper.CL .group
##  I       136.2667 20.378380 5411  96.31690 176.2166  12
```

```
##  G       140.0045 26.028599 5411  88.97800 191.0311  1234
##  F       140.0584 23.472057 5411  94.04375 186.0731  1 3
##  E       158.7721 19.261794 5411 121.01125 196.5330   2 4
##  B       167.9891 14.119341 5411 140.30946 195.6686    34
##  A       174.4100 13.269188 5411 148.39703 200.4229    34
##  H       177.2601  3.528782 5411 170.34223 184.1779  1234
##  C       177.5440  7.160363 5411 163.50678 191.5812  1234
##  D       178.2062  2.622200 5411 173.06568 183.3468  1234
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```

```r
lsmeans(YieldVariety13.lm, cld ~ Product)
```

```
##  Product   lsmean        SE   df  lower.CL upper.CL .group
##  G       129.5061 30.576391 5384  69.56398 189.4482  12
##  I       144.5364 25.421403 5384  94.70021 194.3727  12
##  D       175.5466  2.921516 5384 169.81920 181.2739  1 3
##  H       191.5016  5.202121 5384 181.30336 201.6999   2 45
##  B       193.2737 18.938327 5384 156.14693 230.4005    34
##  C       205.8707 11.193625 5384 183.92662 227.8147    345
##  A       252.9536 23.626080 5384 206.63693 299.2703       6
##  E       285.3457 35.596837 5384 215.56151 355.1299      56
##  F       290.2032 41.909351 5384 208.04392 372.3625    3456
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 9 estimates
## significance level used: alpha = 0.05
```