

Sprawozdanie
Geostatystyka ćw 12
Modelowanie procesów niejednorodnych
Natalia Gadocha 304165
Geoinformatyka II

1

Sprawdzenie struktury i typu badanych danych

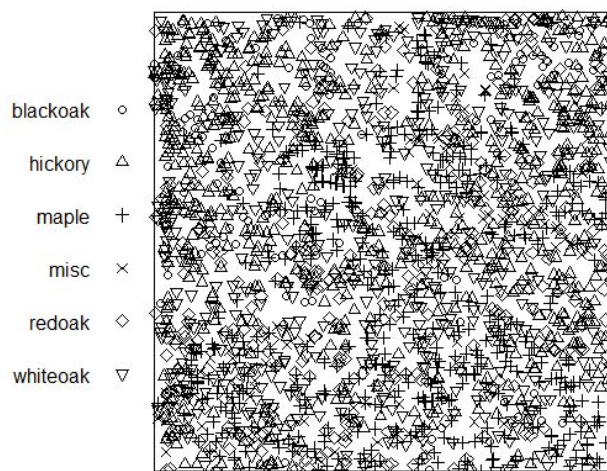
`str(lansing)`
`typeof(lansing)`

```
> str(lansing)
List of 6
 $ window :List of 4
  ..$ type : chr "rectangle"
  ..$ xrange: num [1:2] 0 1
  ..$ yrange: num [1:2] 0 1
  ..$ units :List of 3
  .. ..$ singular : chr "foot"
  .. ..$ plural : chr "feet"
  .. ..$ multiplier: num 924
  ..- attr(*, "class")= chr "unitname"
  ..- attr(*, "class")= chr "owin"
 $ n : int 2251
 $ x : num [1:2251] 0.078 0.076 0.051 0.015 0.03 0.102 0.135 0.121 0.04 0.065 ...
 $ y : num [1:2251] 0.091 0.266 0.225 0.366 0.426 0.474 0.498 0.489 0.596 0.608 ...
 $ markformat: chr "vector"
 $ marks : Factor w/ 6 levels "blackoak","hickory",...: 1 1 1 1 1 1 1 1 1 1 ...
 - attr(*, "class")= chr "ppp"
> typeof(lansing)
[1] "list"
>
```

Wszystkie dane razem prezentują się jak na poniżej. Nie jesteśmy w stanie rozróżnić danych. Dlatego też w tym celu konieczne było rozdzielenie poszczególnych warstw. Pozwoliło to na dokładniejszy ogląd badanych obiektów.

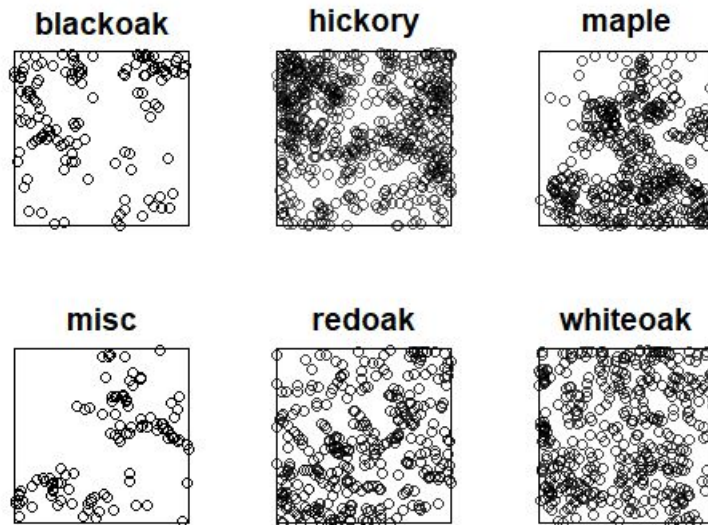
`plot(lansing)`

lansing



```
plot(split(lansing))
```

```
split(lansing)
```



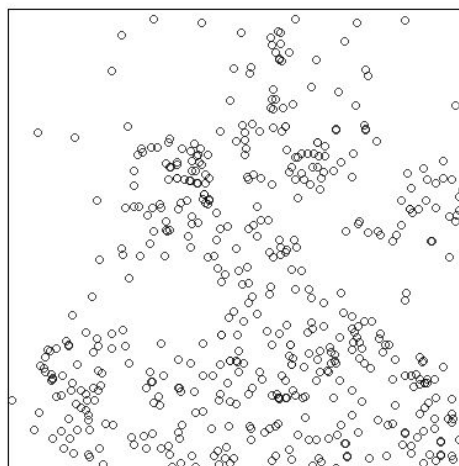
Segregacja danych była w tym przypadku możliwa, gdyż intensywność zależała w tym przypadku od znaku (rodzaju punktu). A w wyniku podziału otrzymaliśmy listę wzorców punktowych.

2

Danymi, nad którymi będzie przeprowadzana dokładniejsza analiza to dane dębu czarnego i klonu. Do dalszego korzystania podpisano je nowymi nazwami.

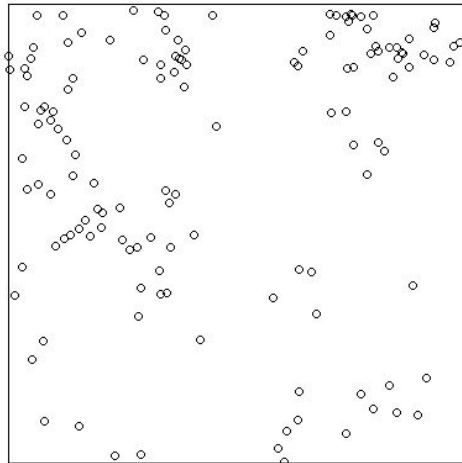
```
klon<-split(lansing)$maple  
plot(klon)
```

klon



```
dab<-split(lansing)$blackoak
plot(dab)
```

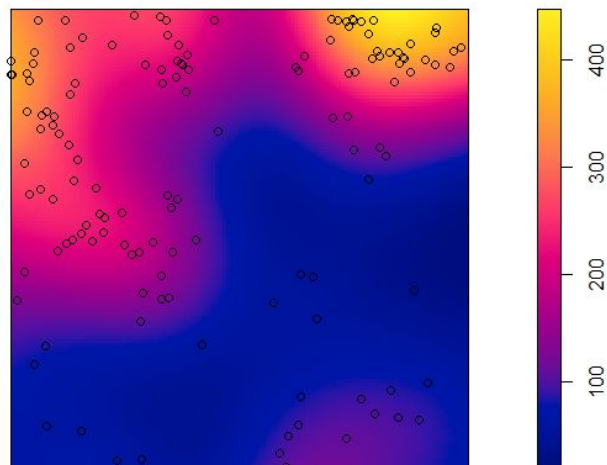
dab



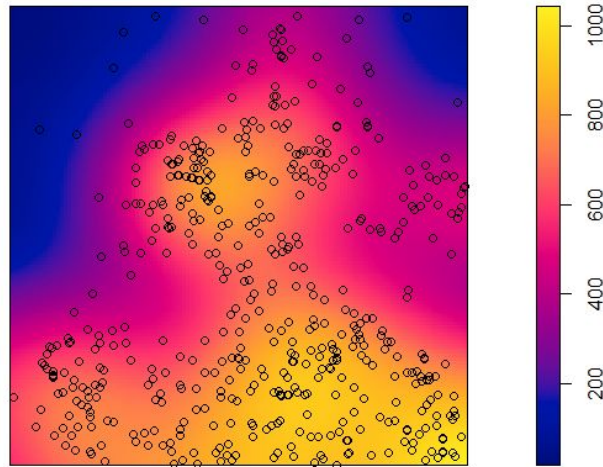
```
density(dab)
density(klon)
```

```
> #-----2
> #intensywnosc
> density(dab)
real-valued pixel image
128 x 128 pixel array (ny, nx)
enclosing rectangle: [0, 1] x [0, 1] units (one unit = 924 feet)
> density(klon)
real-valued pixel image
128 x 128 pixel array (ny, nx)
enclosing rectangle: [0, 1] x [0, 1] units (one unit = 924 feet)
>
```

```
plot(density(dab))
points(dab)
density(dab)
```



```
plot(density(klon))
points(klon)
density(klon)
```



Intensywność to średnia gęstość punktów (oczekiwana liczba punktów na jednostkę powierzchni). Na podstawie obu wykresów możemy dostrzec, iż zmiana intensywności dla klonu jest bardziej gwałtowna i większa w porównaniu z danymi dębu; bardziej obszerna.

3

Obliczanie intensywności rozkładów punktów dla obu zbiorów.

```
pD<-ppm(dab)
pK<-ppm(klon)
```

```
> pD<-ppm(dab)
> pD
stationary Poisson process
Intensity: 135
      Estimate      S.E. CI95.lo CI95.hi ztest   zval
log(lambda) 4.905275 0.0860663 4.736588 5.073962 *** 56.99414
> pK<-ppm(klon)
> pK
stationary Poisson process
Intensity: 514
      Estimate      S.E. CI95.lo CI95.hi ztest   zval
log(lambda) 6.242223 0.04410811 6.155773 6.328674 *** 141.521
>
```

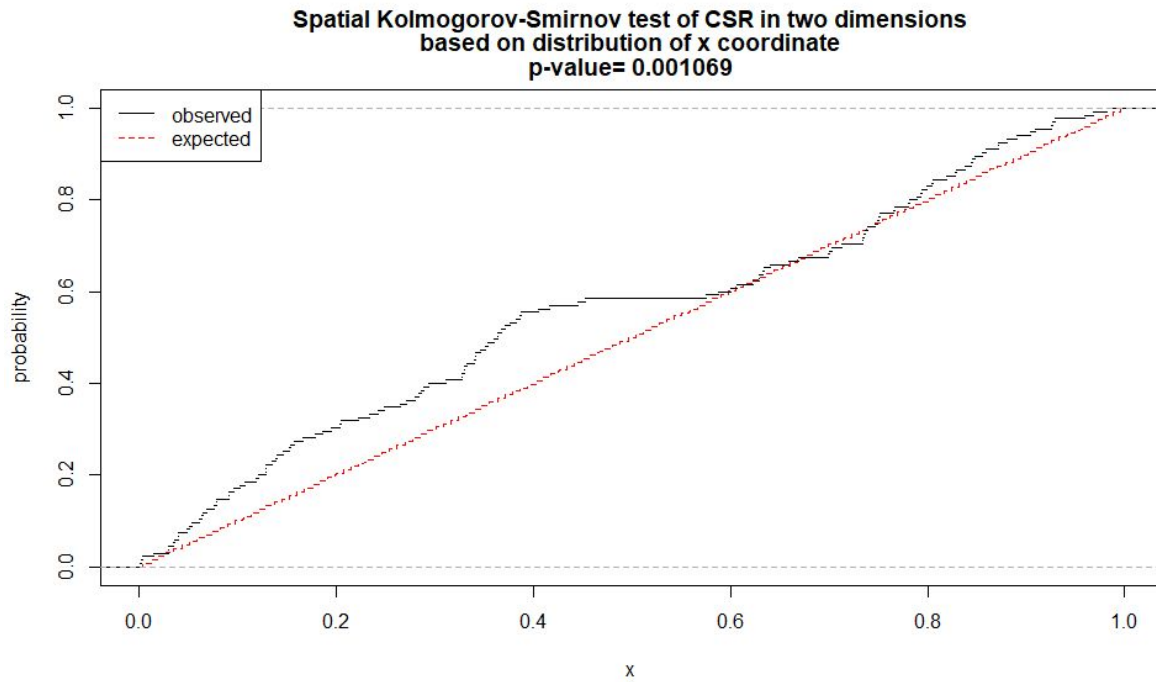
4

Testowanie hipotezy o losowości obu rozkładów

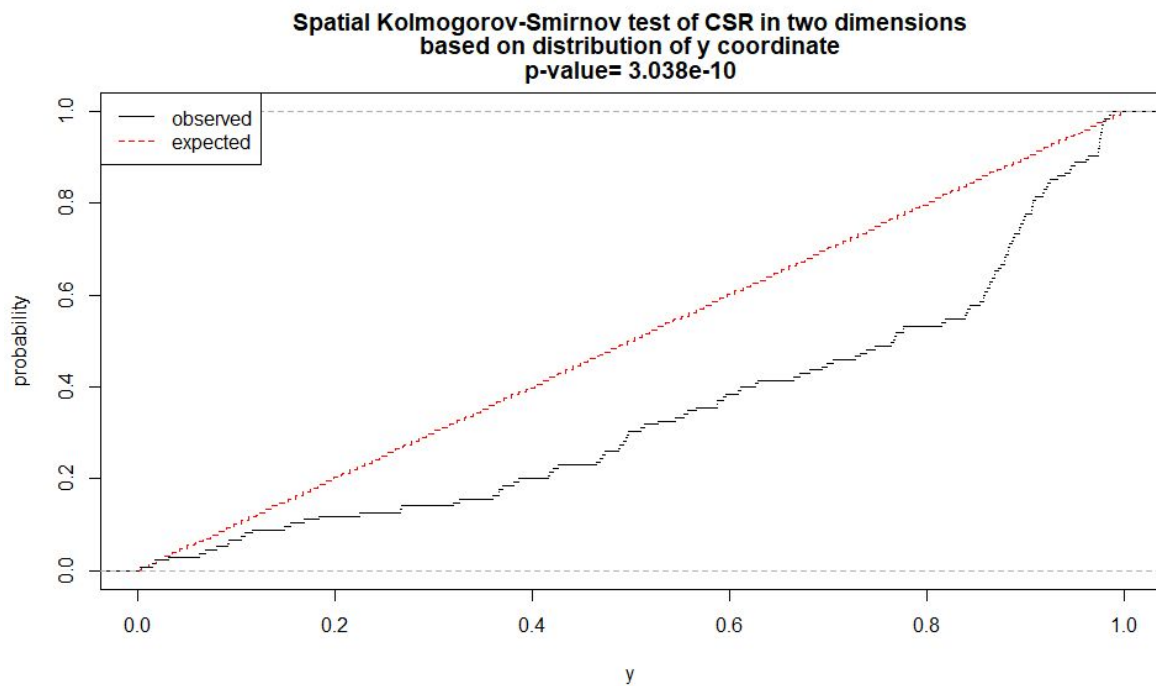
Test Kołmogorowa-Smirnowa służy nam do oceny zgodności rozkładu analizowanych zmiennych z rozkładem normalnym. Testuje on hipotezę zerową wskazującą na rozkład zbliżony do rozkładu normalnego. Używa się go również przed dokładniejszym badaniem danych, gdyż dzięki niemu możemy dowiedzieć się jakie kroki dokładniej należy wykonać w dalszej analizie.

dla debu:

```
ksDx <- cdf.test(dab, "x")  
plot(ksDx)
```

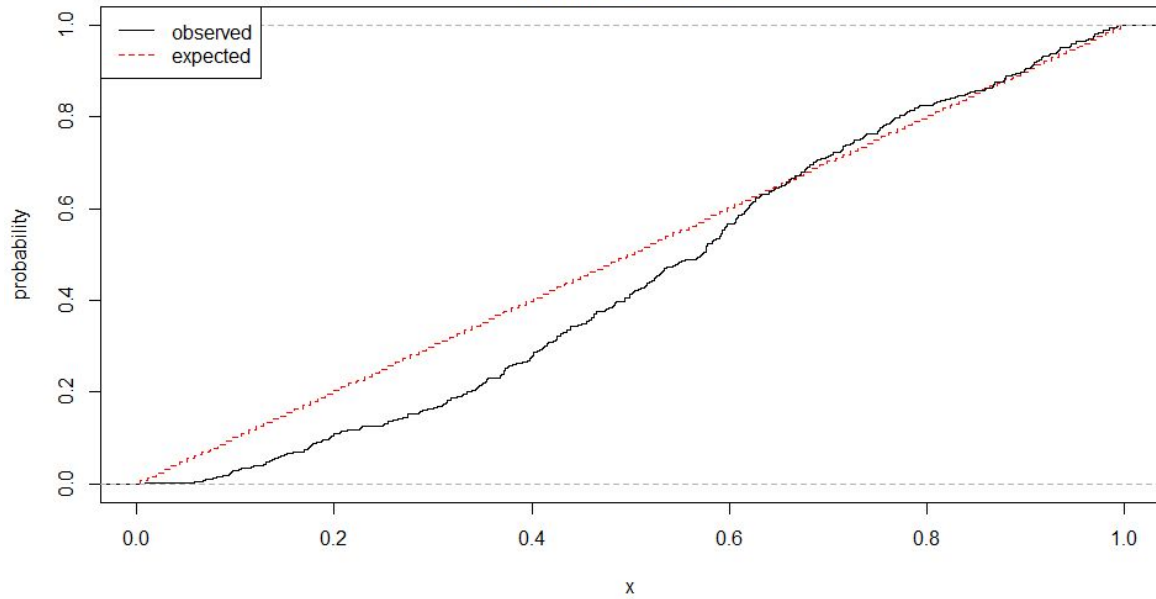


```
ksDy <- cdf.test(dab, "y")  
plot(ksDy)
```



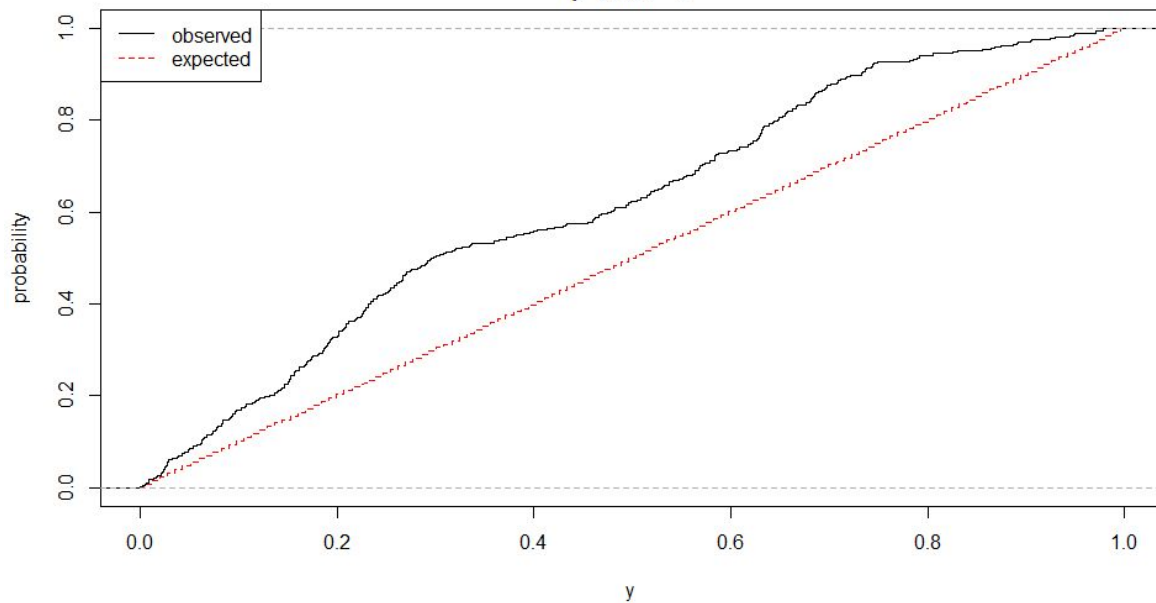
```
dla klonu:  
ksKx <- cdf.test(klon, "x")  
plot(ksKx)
```

**Spatial Kolmogorov-Smirnov test of CSR in two dimensions
based on distribution of x coordinate
p-value= 1.629e-09**



```
ksKy <- cdf.test(klon, "y")  
plot(ksKy)
```

**Spatial Kolmogorov-Smirnov test of CSR in two dimensions
based on distribution of y coordinate
p-value= 0**



ksDx, ksDy, ksKx, ksKy

```
> ksDx
      Spatial kolmogorov-Smirnov test of CSR in two dimensions
data: covariate 'x' evaluated at points of 'dab'
      and transformed to uniform distribution under CSR
D = 0.16705, p-value = 0.001069
alternative hypothesis: two-sided

> ksDy
      Spatial kolmogorov-Smirnov test of CSR in two dimensions
data: covariate 'y' evaluated at points of 'dab'
      and transformed to uniform distribution under CSR
D = 0.28937, p-value = 3.038e-10
alternative hypothesis: two-sided

> ksKx
      Spatial kolmogorov-Smirnov test of CSR in two dimensions
data: covariate 'x' evaluated at points of 'klon'
      and transformed to uniform distribution under CSR
D = 0.14268, p-value = 1.629e-09
alternative hypothesis: two-sided

> ksKy
      Spatial kolmogorov-Smirnov test of CSR in two dimensions
data: covariate 'y' evaluated at points of 'klon'
      and transformed to uniform distribution under CSR
D = 0.20635, p-value < 2.2e-16
alternative hypothesis: two-sided

>
```

Na podstawie uzyskanych wyników możemy stwierdzić, iż rozkład każdej badanej zmiennej nie jest zbliżony do rozkładu normalnego. Widzimy, że wartości p-value są w każdym przypadku jest mniejsze od wartości 0,05, co sprawia że prawdziwym jest odrzucenie hipotezy o jednorodnym rozkładzie punktów w badanym obszarze.

Ponadto o powyższych stwierdzeniach świadczą również wartości D. Są to maksymalne absolutne różnice bezwzględne pomiędzy dwiema dystrybuantami rozkładu. Ich wartość powinna być jak najbliższa zeru, a na naszych wynikach wartości są większe.

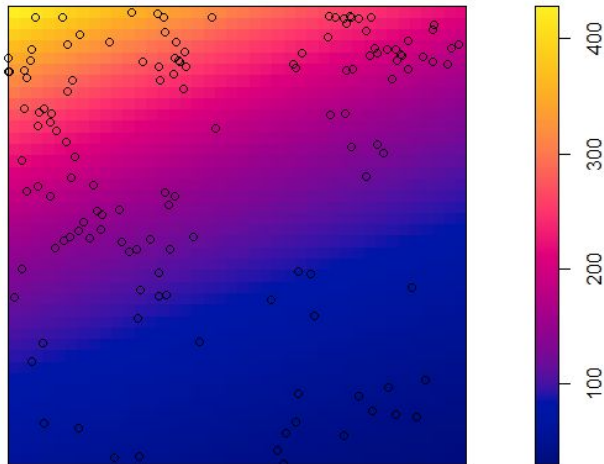
Różnice na powyższych wykresach są mocno uwypuklone. Myślę, że duży wpływ na owe różnice miało samo rozmieszczenie punktów, jak i ich liczba. Badania te są jednak początkiem do dalszych analiz, gdyż w tej chwili nie jesteśmy w stanie wyprowadzić z tego dokładniejszych i poprawnych całościowo wniosków.

5

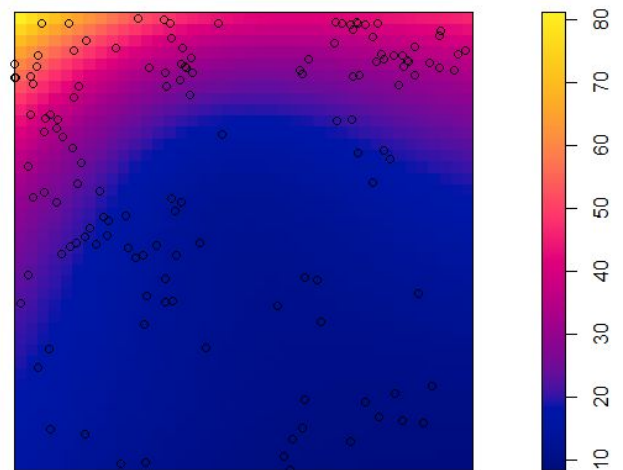
Wyznaczenie zmiennej w przestrzeni intensywności rozkładu
zmiana liniowa

```
PM <- ppm(dab ~ x + y)
plot(PM)
```

Fitted trend



Estimated se



Na powyższych modelach, widzimy jak kolory pokazują zmienną intensywność procesu punkтового najlepiej dopasowanego do modelu.

struktura naszej nowej badanej zmiennej
`str(PM)`

```
> #struktura
> str(PM)
List of 23
 $ method      : chr "mpl"
 $ fitter      : chr "glm"
 $ projected   : logi FALSE
 $ coef        : Named num [1:3] 4.043 -0.707 2.052
 .. attr(*, "names")= chr [1:3] "(Intercept)" "x" "y"
 $ trend       :class 'formula' language ~x + y
 .. attr(*, "Environment")=<environment: R_GlobalEnv>
 $ interaction: NULL
 $ fitin       :List of 5
 .. $ interaction:List of 13
 .. .. $ name      : chr "Poisson process"
 .. .. $ creator   : chr "Poisson"
 .. .. $ family    : NULL
 .. .. $ pot       : NULL
 .. .. $ par       : NULL
 .. .. $ parnames  : NULL
 .. .. $ init      :function (...)
 .. .. $ update    :function (...)
 .. .. $ print     :function (self)
 .. .. $ valid     :function (...)
 .. .. $ project   :function (...)
 .. .. $ irange    :function (...)
 .. .. $ version   : NULL
 .. .. attr(*, "class")= chr "interact"
 .. $ coefs       : num(0)
 .. $ vnames      : chr(0)
 .. $ isoffset    : logi(0)
 .. $ hasInf      : logi TRUE
 .. attr(*, "class")= chr [1:2] "fii" "list"
 .. attr(*, "units")=List of 3
 .. .. $ singular  : chr "foot"
 .. .. $ plural    : chr "feet"
 .. .. $ multiplier: num 924
 .. .. attr(*, "class")= chr "unitname"
 $ Q             :List of 4
 .. $ data       :List of 5
 .. .. $ window   :List of 4
 .. .. .. $ type   : chr "rectangle"
 .. .. .. $ xrange: num [1:2] 0 1
 .. .. .. $ yrange: num [1:2] 0 1
 .. .. .. $ units  :List of 3
 .. .. .. .. $ singular : chr "foot"
 .. .. .. .. $ plural   : chr "feet"
 .. .. .. .. $ multiplier: num 924
```


typ zmiennej
typeof(PM)

```
> #typ
> typeof(PM)      #list
[1] "list"
> |
```

Całościowe podsumowanie powstałej zmiennej
summary(PM)

```
~f ~r
> #i ogolne:
> summary(PM)
Point process model
Fitting method: maximum likelihood (Berman-Turner approximation)
Model was fitted using glm()
Algorithm converged
Call:
ppm.formula(q = dab ~ x + y)
Edge correction: "border"
[border correction distance r = 0 ]

-----
Quadrature scheme (Berman-Turner) = data + dummy + weights

Data pattern:
Planar point pattern: 135 points
Average intensity 135 points per square unit (one unit = 924 feet)
window: rectangle = [0, 1] x [0, 1] units
window area = 1 square unit
unit of length: 924 feet

Dummy quadrature points:
32 x 32 grid of dummy points, plus 4 corner points
dummy spacing: 0.03125 units (one unit = 924 feet)

Original dummy parameters: =
Planar point pattern: 1028 points
Average intensity 1030 points per square unit (one unit = 924 feet)
window: rectangle = [0, 1] x [0, 1] units
window area = 1 square unit
unit of length: 924 feet
Quadrature weights:
(counting weights based on 32 x 32 array of rectangular tiles)
All weights:
  range: [0.000195, 0.000977]    total: 1
weights on data points:
  range: [0.000195, 0.000488]    total: 0.0595
weights on dummy points:
  range: [0.000195, 0.000977]    total: 0.941

-----
FITTED MODEL:

Nonstationary Poisson process

---- Intensity: ----

Log intensity: ~x + y

Fitted trend coefficients:
(Intercept)      x      y
  4.0428137 -0.7065709  2.0515390

      Estimate      S.E.  CI95.lo  CI95.hi  Ztest    Zval
(Intercept)  4.0428137  0.2690954   3.515396  4.5702311 *** 15.023718
x            -0.7065709  0.3017766  -1.298042 -0.1150996 *  -2.341371
y             2.0515390  0.3295282   1.405676  2.6974024 ***  6.225685

----- gory details -----

Fitted regular parameters (theta):
(Intercept)      x      y
  4.0428137 -0.7065709  2.0515390

Fitted exp(theta):
(Intercept)      x      y
  56.986463   0.493333  7.779865
> |
```

Współczynniki wraz z ich parametrami statystycznymi
coef.ppm(PM)

```
> #współczynniki
> coef.ppm(PM)
(Intercept)          x          y
  4.0428137  -0.7065709   2.0515390
> |
```

Możemy zatem dostrzec, iż wyestymowana funkcja intensywności ma postać:

$$\lambda_{\theta}(x, y) = \exp(4.04281 - 0.70657x + 2.051539y)$$

6

Obliczenie macierzy kowariancji

a = vcov.ppm(PM)

wariacja całkowita, całej macierzy

var(a)

wariancje poszczególnych współczynników modelu trendu

diag(a)

diag(a)^2

współczynniki korelacji

cor(a)

```
> a = vcov.ppm(PM)
> #przekatna
> diag(a)
(Intercept)          x          y
  0.07241235  0.09106913  0.10858885
> #wariacja całej macierzy
> var(a)
(Intercept)          x          y
(Intercept)  0.005735522 -0.002949628 -0.006238402
x            -0.002949628  0.004523816  0.001121558
y            -0.006238402  0.001121558  0.008233463
> #obliczanie
> diag(a)^2
(Intercept)          x          y
  0.005243549  0.008293586  0.011791537
> #badanie korelacji
> cor(a)
(Intercept)          x          y
(Intercept)  1.0000000 -0.5790663 -0.9078113
x            -0.5790663  1.0000000  0.1837714
y            -0.9078113  0.1837714  1.0000000
> |
```

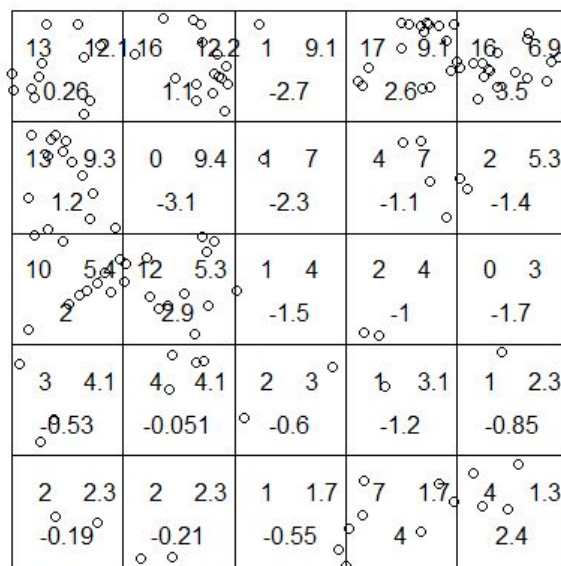
Testy statystyczne jakości dopasowania modelu liniowego do zmian intensywności

```
qt<-quadrat.test(PM)
```

```
> qt
Chi-squared test of fitted Poisson model 'PM' using quadrat counts
data: data from PM
x2 = 90.09, df = 22, p-value = 6.617e-10
alternative hypothesis: two.sided
Quadrats: 5 by 5 grid of tiles
> |
```

Ponieważ otrzymujemy wartość p-value mniejszą niż poziom istotności 0.05, bazując na wyniku tego testu moglibyśmy odrzucić hipotezę zerową.

```
plot(dab)
plot(qt, add=TRUE)
dab
```

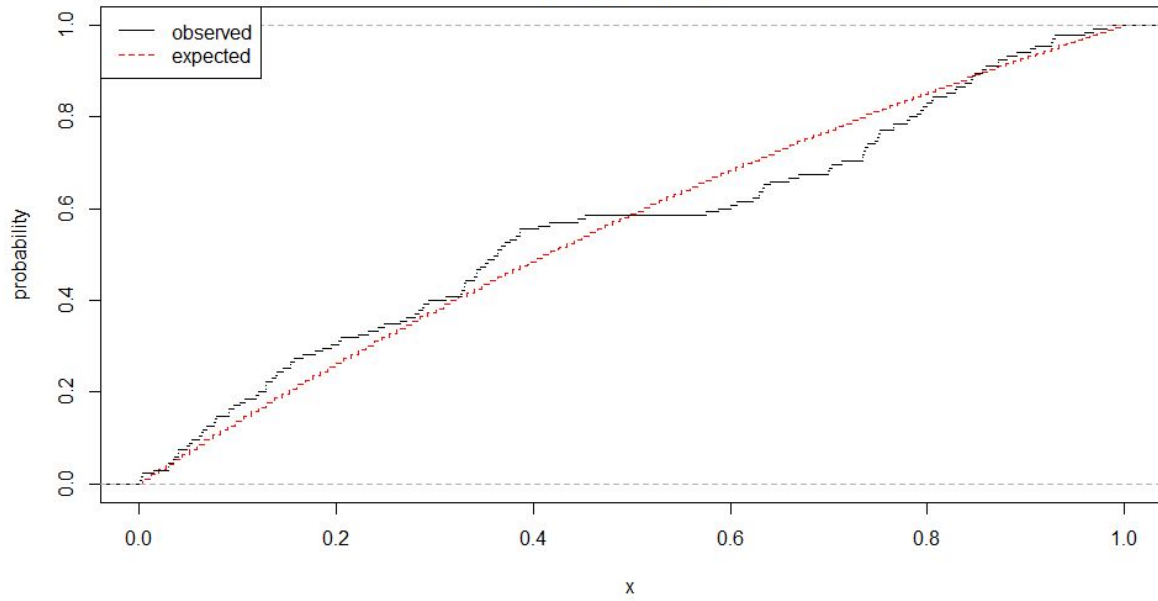


Test niezależności chi-kwadrat opiera się na porównywaniu ze sobą wartości obserwowanych z wartościami oczekiwanymi. W naszym przypadku wartości reszty r nie wskazują na bardzo duże odchyłki od modelu.

A ponieważ nasze p-value jest bliskie zera, odrzucając hipotezę zerową nie popełniamy błędu.

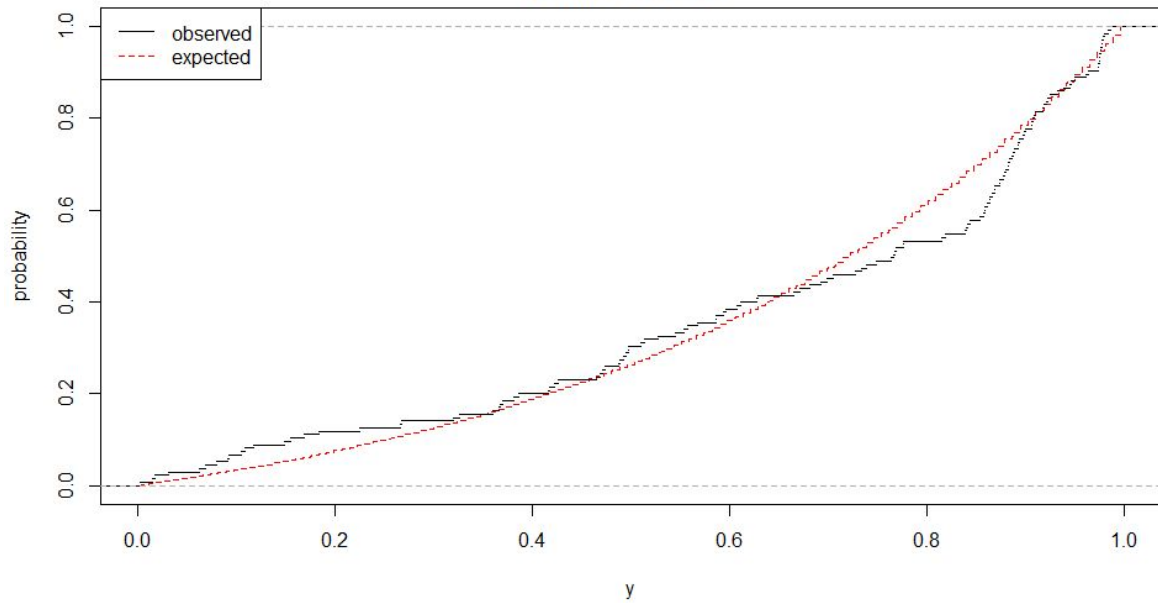
```
PctX <- cdf.test(PM, "x")  
plot(PctX)
```

**Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two dimensions
based on distribution of x coordinate
p-value= 0.1944**



```
PctY <- cdf.test(PM, "y")  
plot(PctY)
```

**Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two dimensions
based on distribution of y coordinate
p-value= 0.02546**



```

> PCTx
      Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two dimensions
data: covariate 'x' evaluated at points of 'dab'
      and transformed to uniform distribution under 'PM'
D = 0.092899, p-value = 0.1944
alternative hypothesis: two-sided

> PCTy
      Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two dimensions
data: covariate 'y' evaluated at points of 'dab'
      and transformed to uniform distribution under 'PM'
D = 0.12713, p-value = 0.02546
alternative hypothesis: two-sided

>

```

Na powyższych wykresach widzimy zmianę w porównaniu do wcześniej wykonanych wykresów dla testu K-S. Wartość D jest już bliższa wartości 0. Natomiast p-value dla zmiennej x jest większa od 0,05, jednakże dla wartości y nie spełnia tego warunku.

Myślę, że na wyniki jakie nam powstały duży wpływ miały wartości odstające, krytyczne, niepasujące do reszty, od których też zależy wcześniej wspomniana intensywność.

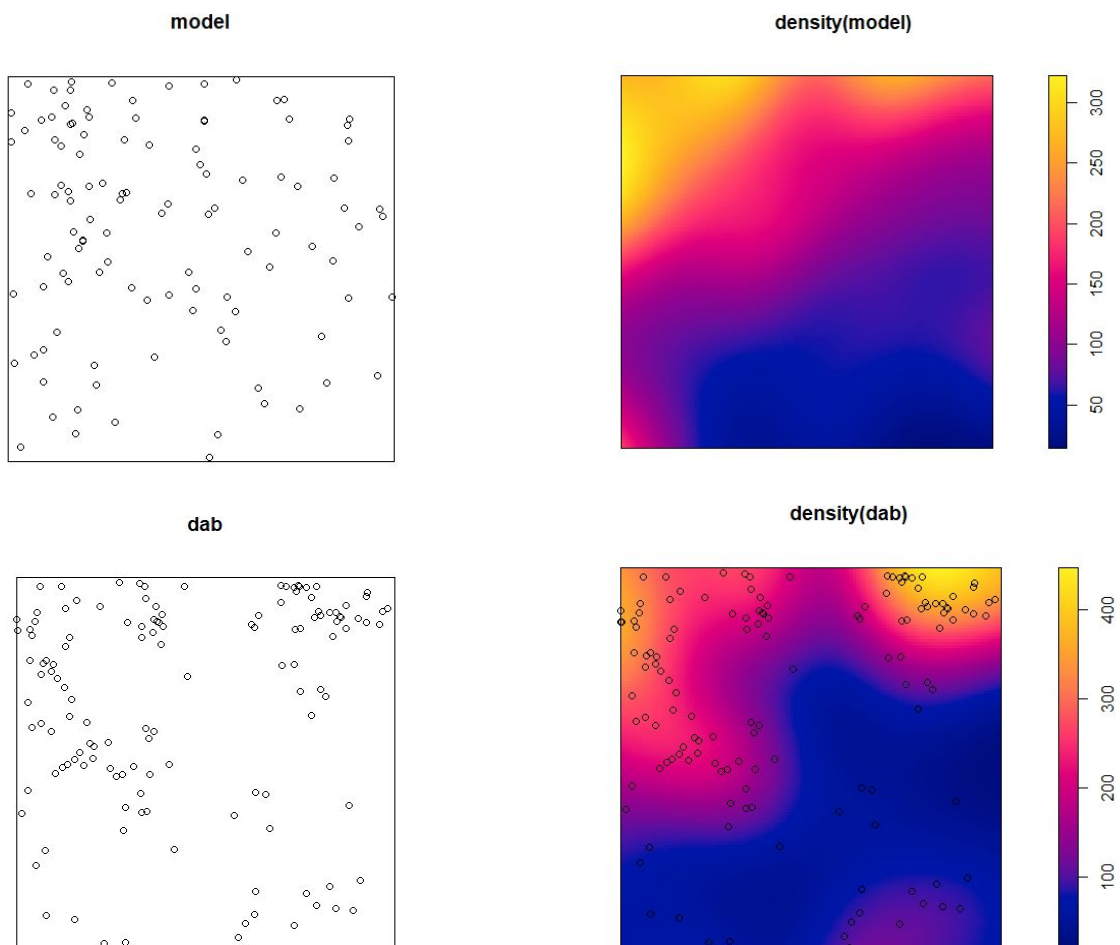
8

Generacja procesu losowego o zmiennej intensywności.

```

set.seed(123)
rm<- rmhmodel(PM)
model<-rmh(rm)
plot(model)
plot(density(model))

```

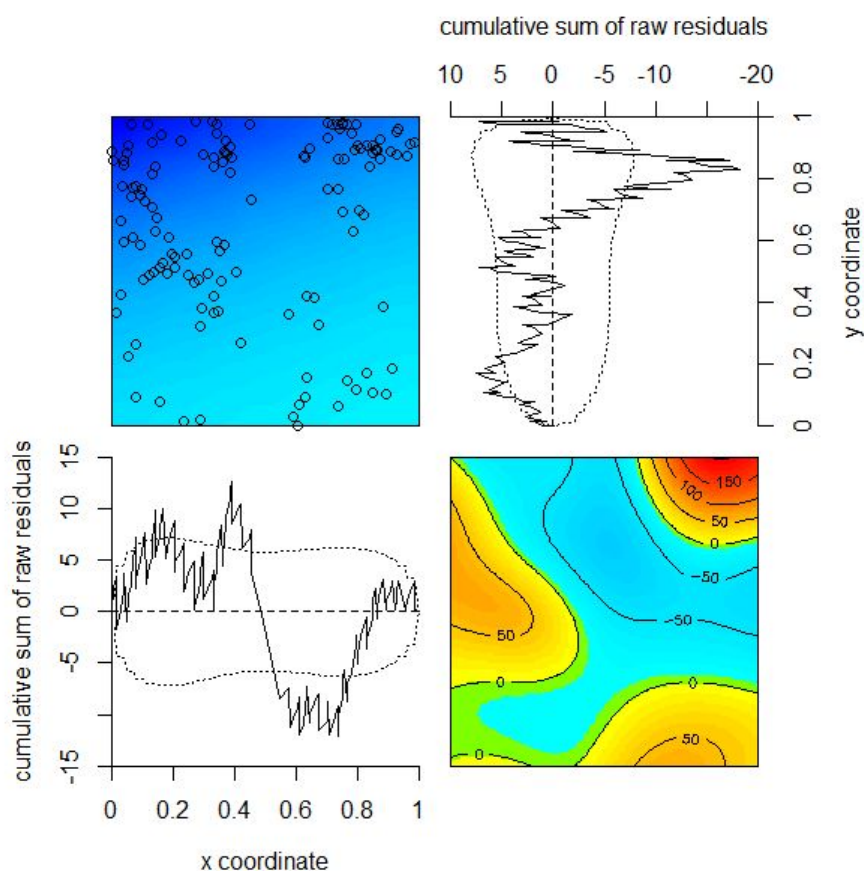


Na podstawie porównania powyższych modeli, uważam iż nowo powstały wykres funkcji intensywności jest dobrej jakości. Można dostrzec podobieństwa powstałe między nimi, zarówno na zwykłym wykresie, jak i na modelu intensywności. Na przykład większą intensywność widzimy w górnym lewym rogu wyznaczonych obszarów, a najmniejszą w prawej dolnej części.

9

Przeprowadzenie analizy modelu i jakości dopasowania na podstawie reszt.

diagnose.ppm(PM)



```
> diagnose.ppm(PM)
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 = -1.073e-06
area of entire window = 1
quadrature area = 1
range of smoothed field = [-87.84, 201.9]
```


Na powyższych wykresach widzimy sumaryczną analizę naszych wartości. W pierwszym kwadracie widać rozmieszczenie naszych punktów na tle odpowiadającej im intensywności, którą odpowiednio reprezentują dane barwy. Jesteśmy w stanie dostrzec, iż model ten jest podobny do wcześniej wygenerowanych modeli. Ma jednak mniej widoczne dokładniejsze granice zmieniających się przedziałów.

Dolny prawy obszar to miejsce, gdzie dane odchylenia zostały uśrednione. Jest on obrazem wygładzonego pola resztkowego. Dominującym uśrednieniem widocznym na środku obszaru są wartości mniejsze od zera. Natomiast prawy górny róg jest miejscem, gdzie wartości są najbardziej odstające. We wcześniej omawianym obszarze jest to miejsce, gdzie najwięcej wartości nachodzi na siebie. Podobna sytuacja, choć nie tak duża, znajduje się w lewej części badanego obszaru.

Dwa pozostałe wykresy przedstawiają zależności dla poszczególnych osi. Są to wykresy zmiennej zakłócającej dla współrzędnej y oraz x. Wykresy te jednak nie są takie same; są przekształcone. Brak dokładnego dopasowania, jak i wcześniej omówiony obraz wygładzonego pola resztkowego może sugerować występowania nadmiaru dodatnich reszt

Książka (link do niej), z której wykorzystałam dodatkowe informacje na temat owego projektu:

<https://darrylmcleod.com/wp-content/uploads/2016/06/Analysing-spatial-point-patterns-in-R.pdf>