

Sprawozdanie  
Geostatystyka ćw 7  
KLASTERYZACJA DANYCH PRZESTRZENNYCH I CZASO-PRZESTRZENNYCH  
Natalia Gadocha 304165  
Geoinformatyka II

I Klasteryzacja przestrzenna

1

```
sasq <- readRDS("C:/Users/natal/Desktop/sasquatch.rds")
```

2

```
summary(sasq)
```

```
C:/Users/natal/Desktop/
> summary(sasq)
Marked planar point pattern: 423 points
Average intensity 2.097156e-09 points per square unit

*Pattern contains duplicated points*

coordinates are given to 1 decimal place
i.e. rounded to the nearest multiple of 0.1 units

Mark variables: date, year, month
Summary:
      date          year      month
Min.   :1990-05-03   Y2004   : 41   Sep    : 59
1st Qu.:2000-04-30   Y2000   : 36   Oct    : 56
Median :2003-11-05   Y2002   : 30   Aug    : 54
Mean   :2003-08-11   Y2005   : 30   Jul    : 50
3rd Qu.:2007-11-02   Y2001   : 26   Nov    : 43
Max.   :2016-04-05   Y2008   : 26   Jun    : 41
                        (other):234   (other):120

window: polygonal boundary
single connected closed polygon with 64 vertices
enclosing rectangle: [368187.8, 764535.6] x [4644873, 5434933] units
                    (396300 x 790100 units)
window area = 2.01702e+11 square units
Fraction of frame area: 0.644
>
```

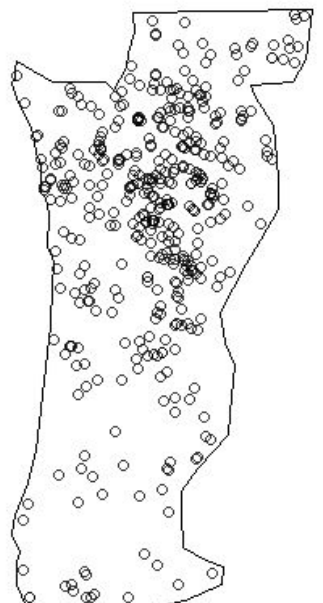
3

```
plot(unmark(sasq))
```

Czy punkty są rozmieszczone regularnie,  
równomiernie czy grupują się przestrzennie?

Powyższe punkty nie układają się regularnie. Można zauważyć, że najwięcej znajduje się w środkowej części wyznaczonego obszaru. W miejscu tym jest wiele punktów, które na siebie wzajemnie nachodzą. W dolnej części natomiast występują sporadycznie, jest ich tam niewielka ilość.

**unmark(sasq)**



Dla bardziej dokładnej i rzetelnej odpowiedzi przeprowadziłam jeszcze test zgodności chi-kwadrat. Użyłam do tego komendy:  
`quadrat.test(sasq, alternative = "clustered")` badając skupienie punktów. Jako też alternatywną przyjąłam, iż punkty są skupione. Wynik jaki otrzymałam to:

```
> quadrat.test(sasq, alternative = "clustered")  
  
Chi-squared test of CSR using quadrat counts  
  
data: sasq  
x2 = 261.32, df = 22, p-value < 2.2e-16  
alternative hypothesis: clustered  
  
Quadrats: 23 tiles (irregular windows)  
komunikat ostrzegawczy:  
Some expected counts are small; chi^2 approximation may be inaccurate
```

Można dostrzec, że wartość p-value jest bardzo niewielka, a więc możemy odrzucić hipotezę zerową.

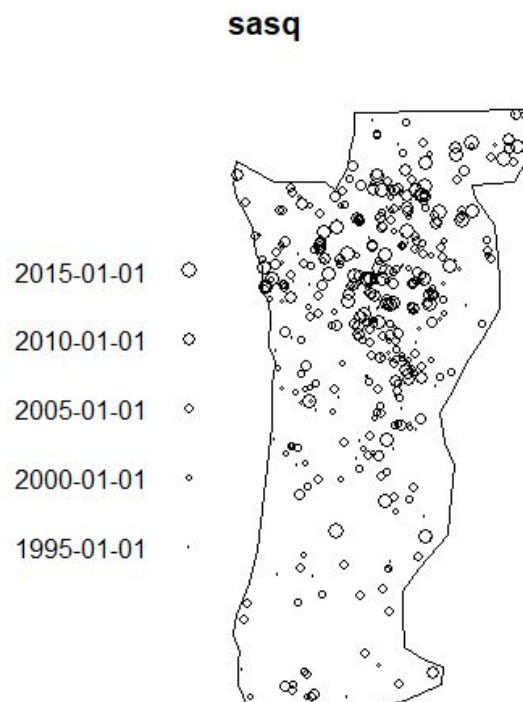
Nasze dane nie są jednorodnym procesem Poissona.

Możemy jednak po powyższych badaniach stwierdzić, że istnieje pewien rodzaj grupowania przestrzennego.

## II Klasteryzacja czasowa

**Czy liczba obserwacji wzrasta/ maleje? Czy zmienia się w ciągu roku, wykazując pewną sezonowość?**

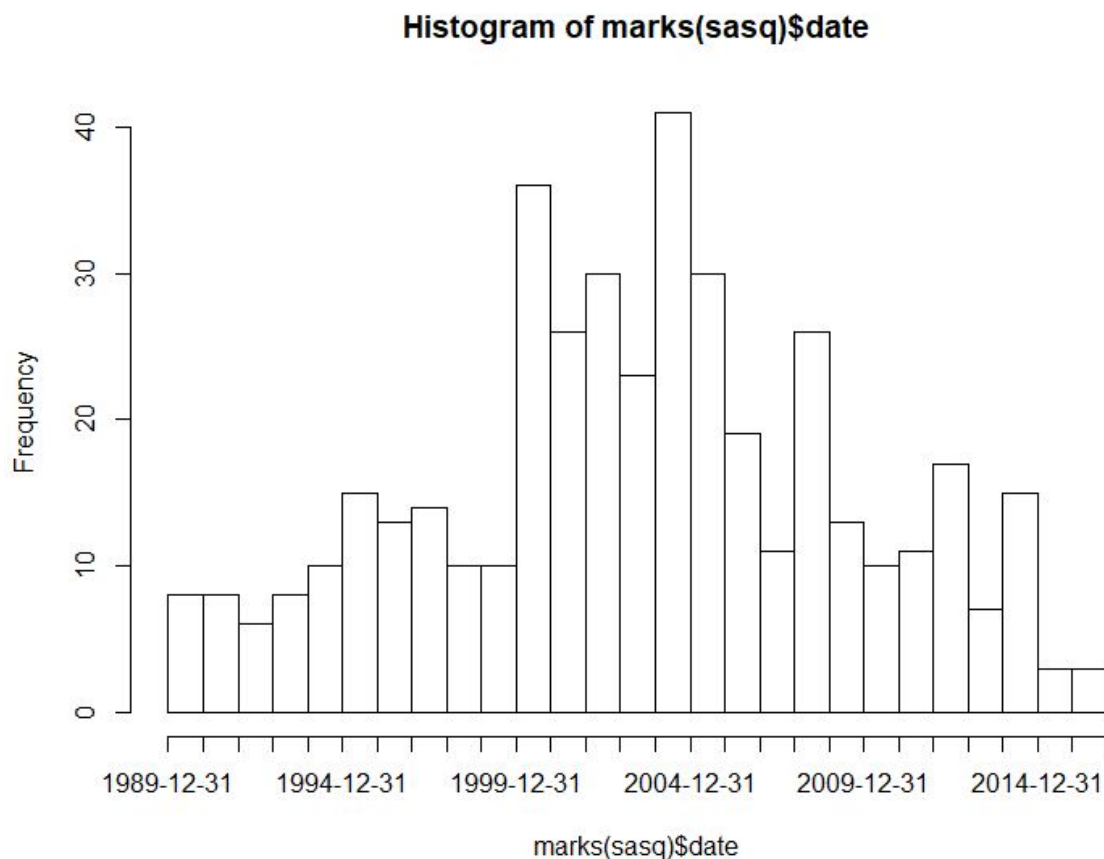
```
plot(sasq, which.marks = "date")
```



Na powyższym obrazie możemy prawdopodobnie dostrzec, iż w przeważającej części jest ilość okręgów odpowiadająca późniejszym latom. Znajdują się one głównie w środkowej części obszaru. Badanie to jest jednak zdecydowanie niewystarczające. Dokładność jego jest niewielka ze względu na mały rozmiar punktów dla wcześniejszych lat i nachodzenie się punktów na siebie.. Nie są one zatem dobrze widoczne, zwłaszcza jeżeli chodzi o środkową część.

4

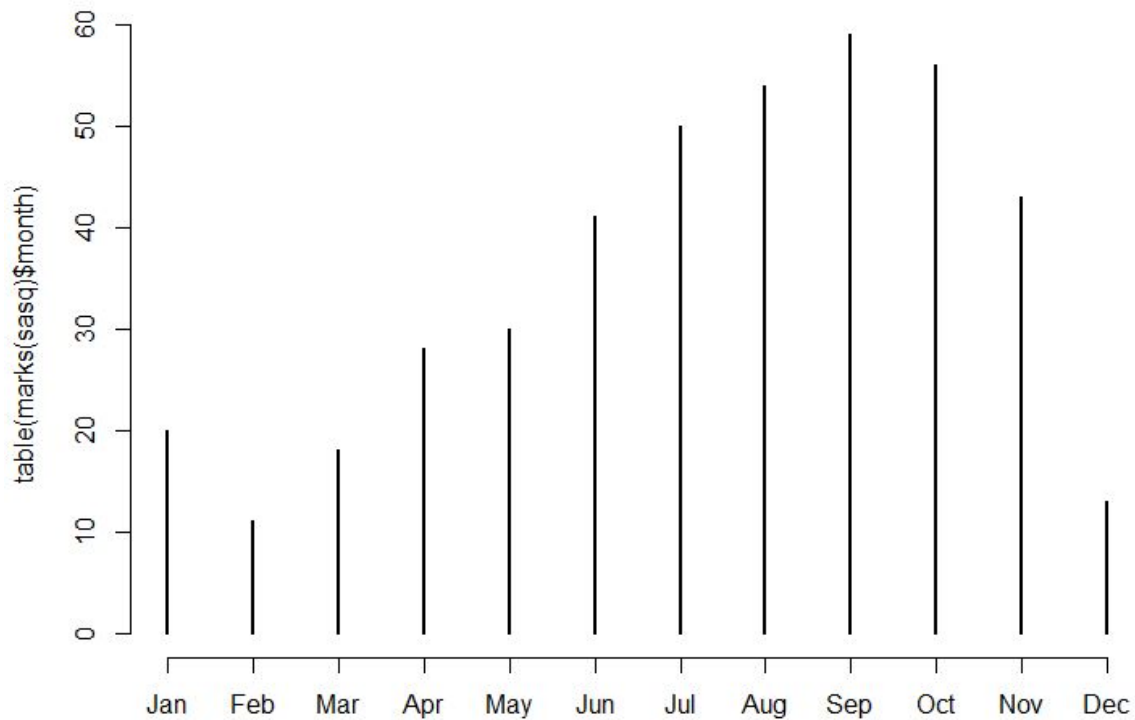
```
hist(marks(sasq)$date, "years", freq = TRUE)
```



Na powyższym histogramie widzimy, iż największa liczba pomiarów była jednak wykonana w latach 1999 - 2004. Najmniej pomiarów obserwujemy po 2014 roku. Jest ich zdecydowanie mniej niż na początku zbierania danych, czyli w okolicach początków lat dziewięćdziesiątych XX w. Ogólnie jednak widać, że ilość pomiarów początkowo lekko rośnie, później widać gwałtowne zwiększenie tych liczb, a następnie ponownie maleje.

5

```
plot(table(marks(sasq)$month))
```



Na powyższym wykresie można dostrzec, iż największa ilość obserwacji ma miejsce na przełomie lata i jesieni (sierpień - październik). Najmniejsza natomiast jest dla miesięcy zimowych (grudzień - marzec). Począwszy od lutego widzimy też stopniowy wzrost liczby pomiarów. Największą miesięczną różnicą jest ta pomiędzy listopadem i grudniem.

**sasq\_by\_month**

6, 7

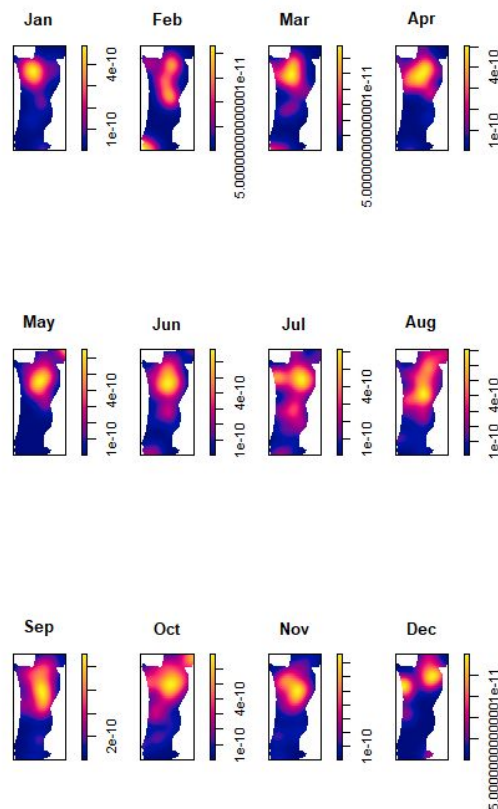
```
sasq_by_month <- split(sasq, "month", un = TRUE)
plot(sasq_by_month)
```



8

```
plot(density(sasq_by_month))
```

```
density(sasq_by_month)
```



### III Klasteryzacja czaso-przestrzenna

Czy liczba obserwacji wzrasta/ maleje? Czy zmienia się w ciągu roku, wykazując pewną sezonowość?

9

Przygotowane dane :

- Wyodrębnione współrzędne obserwacji w postaci macierzy:  

```
sasqXY <- as.matrix(coords(sasq))
```

Rozmiar powstałej macierzy:

```
dim(sasqXY)
```

```
> dim(sasqXY)
[1] 423  2
```

- Czas dla każdej obserwacji w postaci wektora

```
sasq_t <- marks(sasq)$date
```

- Macierz określająca współrzędne okna

```
sasq_Window <- as.matrix(as.data.frame(Window(sasq)))
```

Rozmiar powstałej macierzy

```
dim(sasq_Window)
```

```
> dim(sasq_Window)
[1] 64  2
> |
```

- dodanie 1 dnia przed obserwacjami i 1 dnia po obserwacjach

```
tplus <- range(sasq_t) + c(-1, 1)
```

- Wektor s skanujący obszar od 100 m do 20 km co 400m oraz wektor tm skanujący obszar od 1 tygodnia do 31 tygodni co 14 dni

```
s <- seq(100, 20000, by = 400)
```

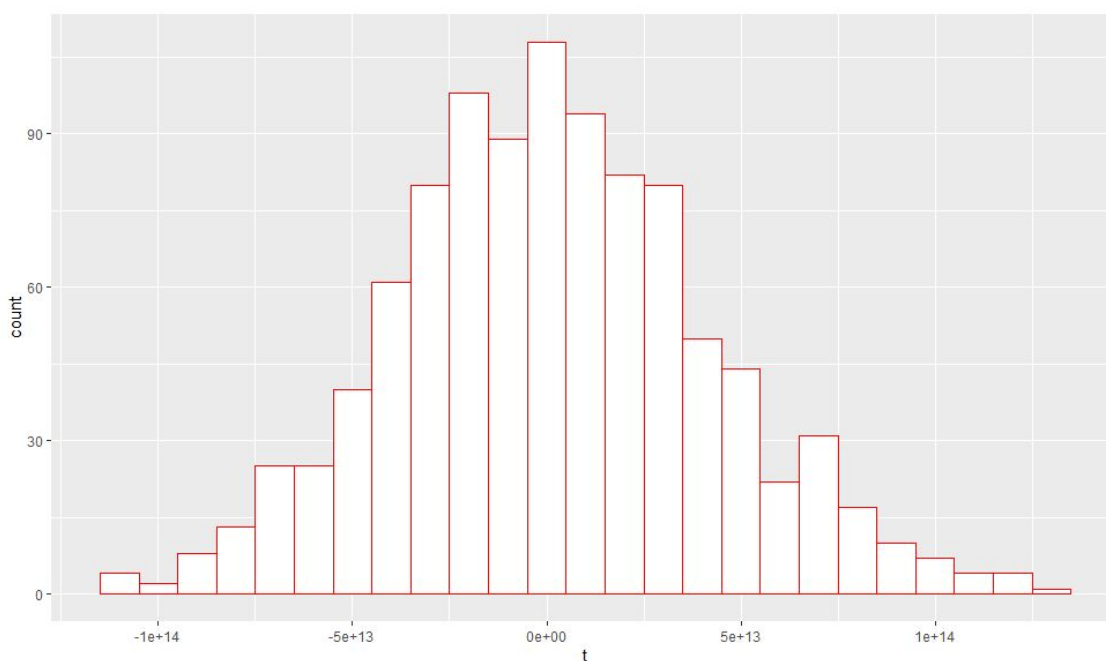
```
tm <- seq(7, 217, by = 14)
```

10

```
sasq_sym <- stmctest(sasqXY, sasq_t, sasq_Window, tplus, s, tm, nsim
= 999, quiet = TRUE)
```

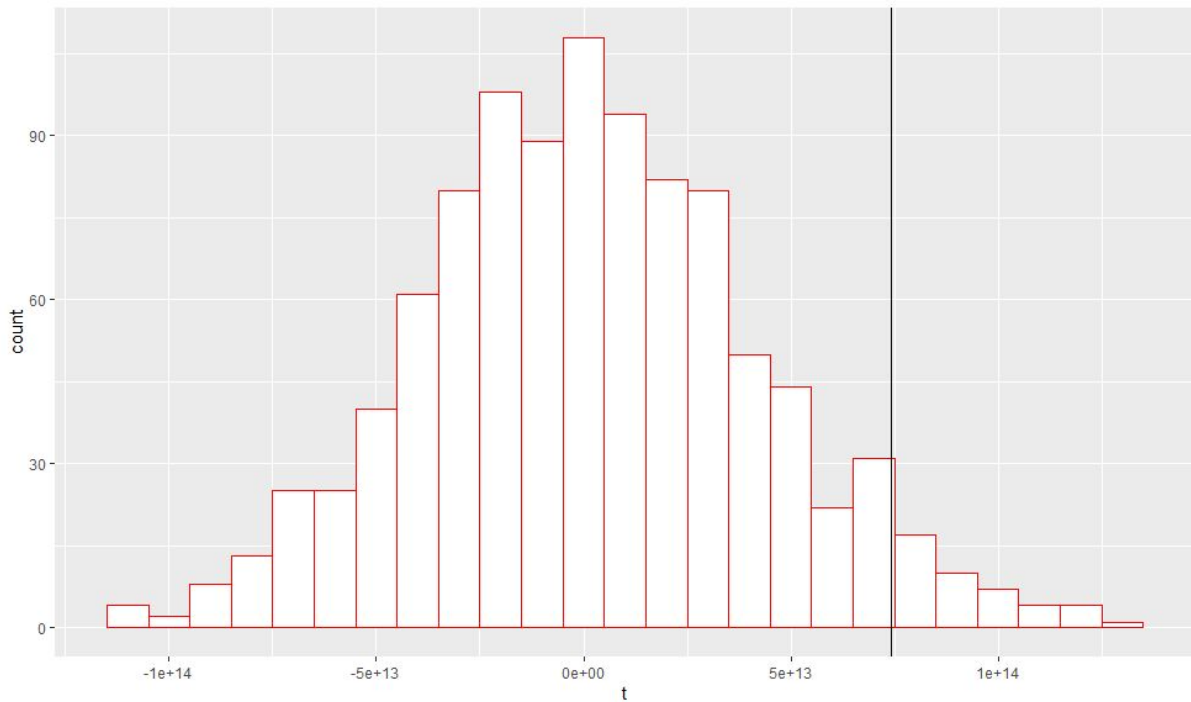
11

```
ggplot(data.frame(sasq_sym), aes(x = t)) + geom_histogram(binwidth
= 1e13, fill="white", col="red")
```



12

```
ggplot(data.frame(sasq_sym), aes(x = t)) + geom_histogram(binwidth
= 1e13, fill="white", col="red") + geom_vline(aes(xintercept = t0))
```



13

```
sum(sasq_sym$t > sasq_mc$t0) / 1000
```

```
> sum(sasq_sym$t > sasq_mc$t0) / 1000
[1] 0.051
>
```

Wynik interpretujemy przy pomocy testu Monte Carlo jak i wcześniej powstałych wykresów. Nasze uzyskane p wynosi 0,051, a więc możemy odrzucić hipotezę dotyczącą grupowania nieprzestrzenno - czasowego.