

Klaster

w

Epidemiologii

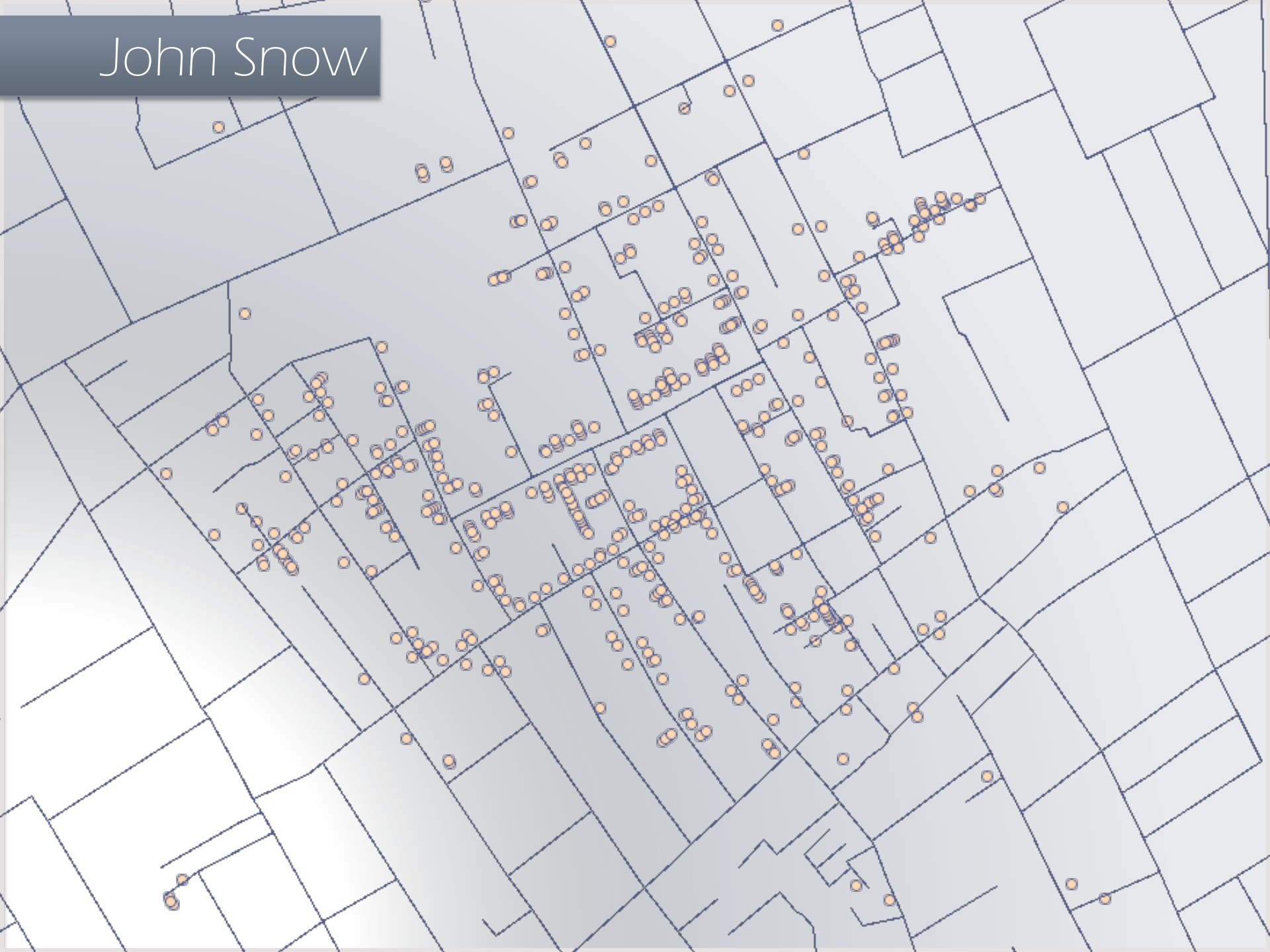
Analiza
przestrzenna

Epidemiologia

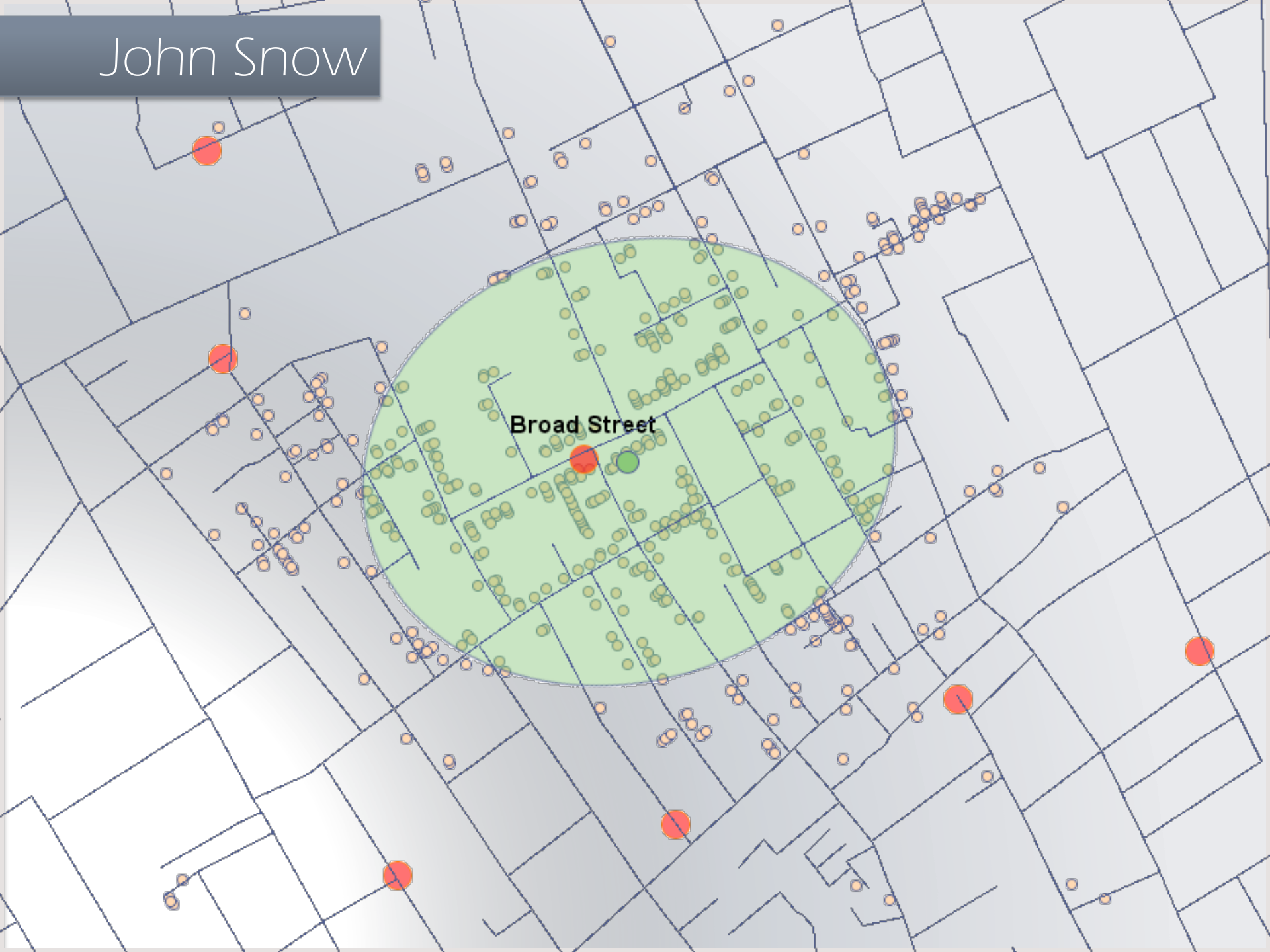
Dział medycyny zajmujący się badaniem **przyczyn** powstania, rozwoju, rozprzestrzeniania się **chorób**



John Snow



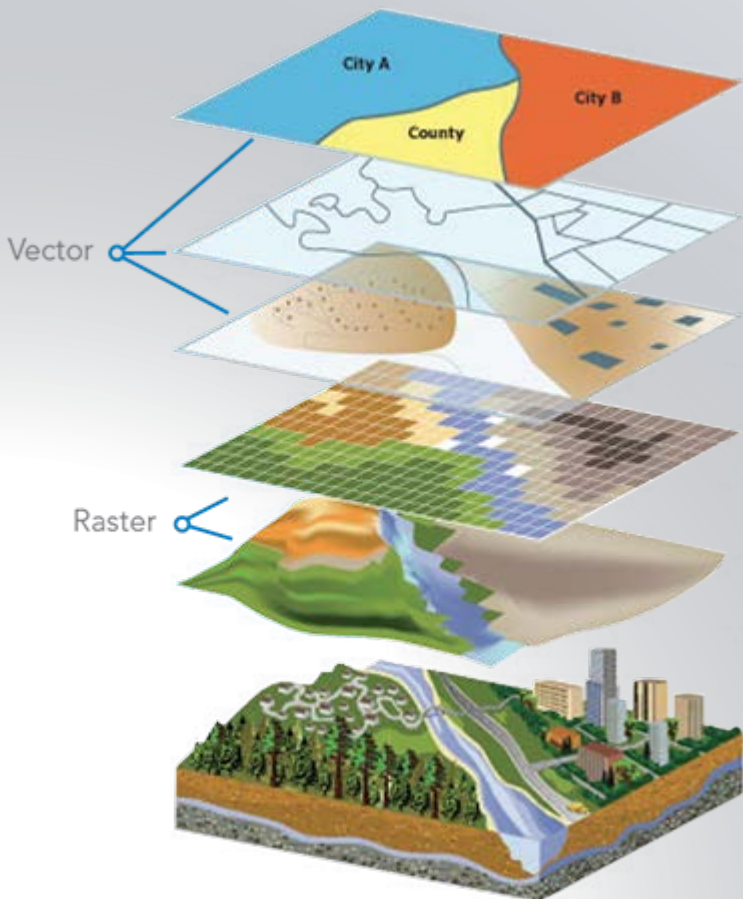
John Snow



John Snow



Analizy przestrzenne



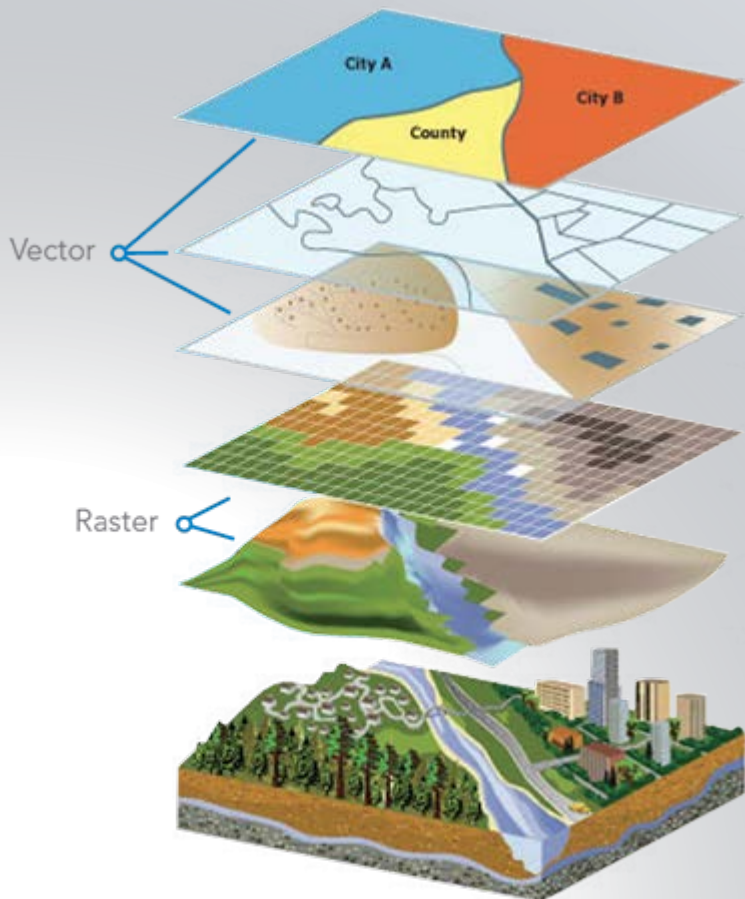
✓ związane z mapą

✓ złożone obliczeniowo

✓ rozwój uwarunkowany
rozwojem oprogramowania

Analizy przestrzenne

- ✓ 1854 John Snow
- ✓ 1963 -1971 GIS



Analizy przestrzenne



- ✓ 1854 John Snow
- ✓ 1963 -1971 GIS
- ✓ 1978 GPS

Analizy przestrzenne



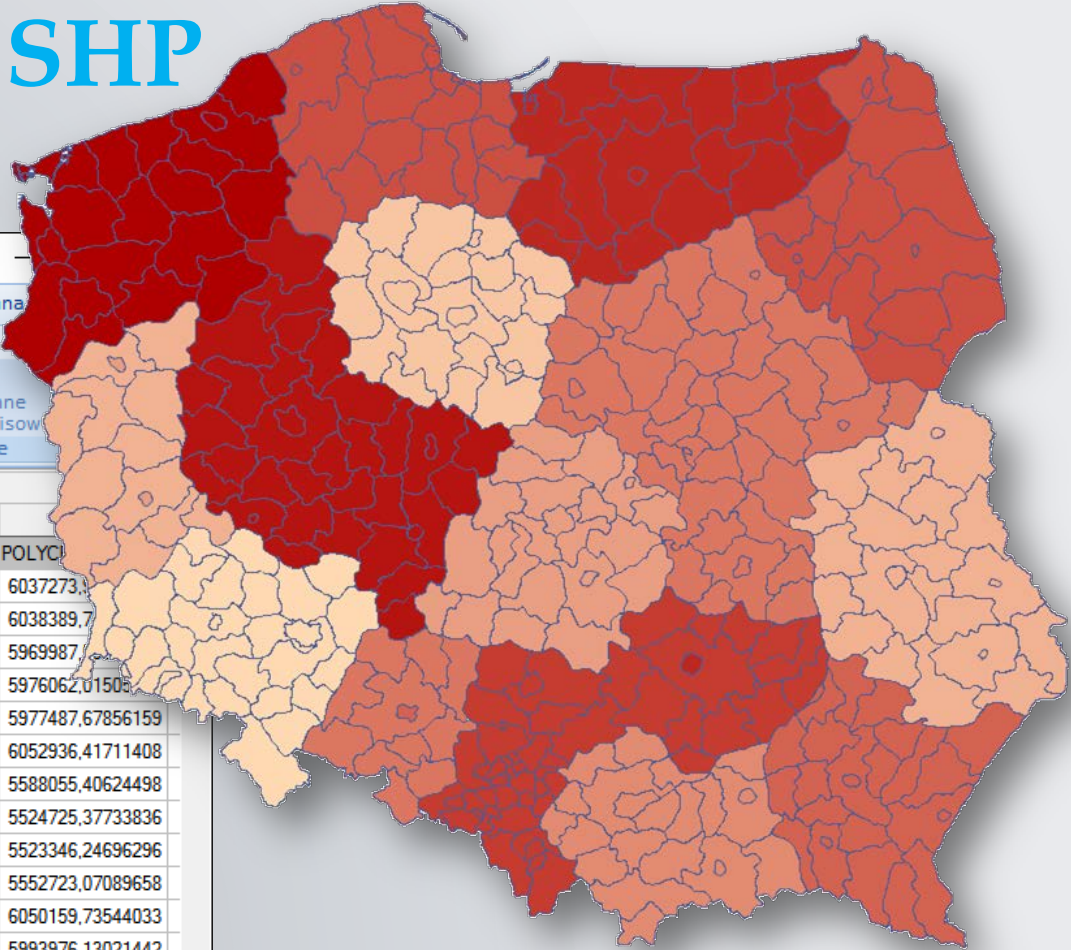
- ✓ 1854 John Snow
- ✓ 1963 -1971 GIS
- ✓ 1978 GPS

Dz. U. z **2014r.**
CODGiK

GIS

Analizy przestrzenne

DBF ↔ SHP



PQStat v.1.6.2 [C:\Program Files\PQStat\Dane\Maps\districts\districts.pqs]

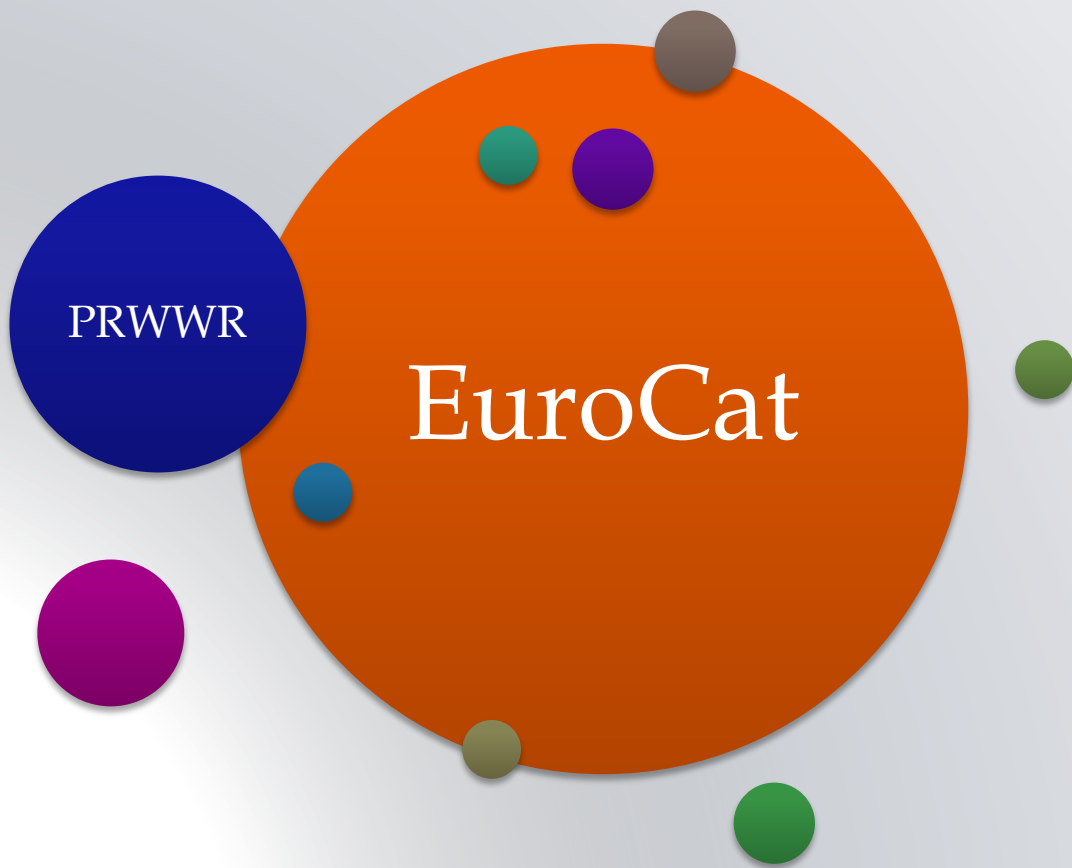
Plik Edycja Dane Statystyki Statystyki zaawansowane Analiza przestrzenna

Powtórzenie testu Historia Menadżer map Mapy Obliczenia geometryczne Macierz wag przestrzennych Narzędzia Wyglądanie zmiennych Przemienne statystyki opisowe Opisowe

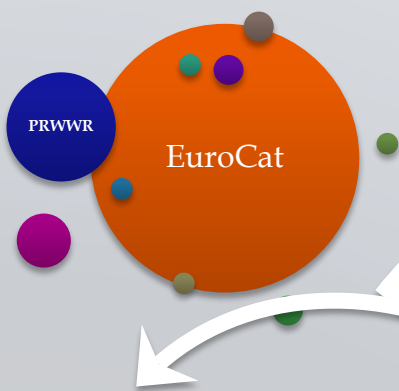
districts [SHP] Dane 1 Przemienne statystyki lokalne [WM] Dane 2

	5	6	7	
	TERYT_POW7	POWIAT	POLYCENTROID_X	POLYCENTROID_Y
ID1	2263000	Ślupsk	634864,662752021	6037273,...
ID2	2264000	Sopot	729960,54486002	6038389,7
ID3	2213000	starogardzki	725487,533870399	5969987,...
ID4	2216000	sztumski	777560,127603471	5976062,0150
ID5	2214000	tczewski	745583,374034472	5977487,67856159
ID6	2215000	wejherowski	698316,525602102	6052936,41711408
ID7	2401000	będziński	792703,839137373	5588055,40624498
ID8	2402000	bielski	791159,546197624	5524725,37733836
ID9	2461000	Bielsko-Biała	792102,215043912	5523346,24696296
ID10	2414000	bieruńsko-łódziński	797271,193302609	5552723,07089658
ID11	2208000	łębski	675254,758846684	6050159,73544033
ID12	2209000	malborski	762436,792482804	5993976,13021442
ID13	2210000	nowodworski	768483,311993157	6016773,65669093
ID14	2211000	nurki	711894,530071543	6069951,31665426

Moja praca

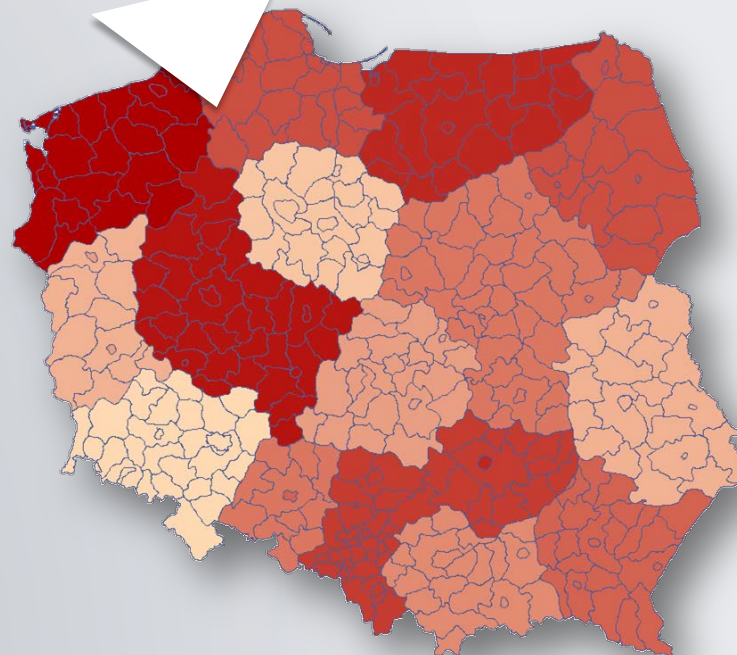


Moja praca



TERYT	Wady
3064011	739
3031055	5
3031054	3
3031042	9
3031082	9
3031025	2
3031072	3

TERYT	UŻ
3064011	548028
3031055	4896
3031054	3981
3031042	5658
3031082	9591
3031025	3011
3031072	4938



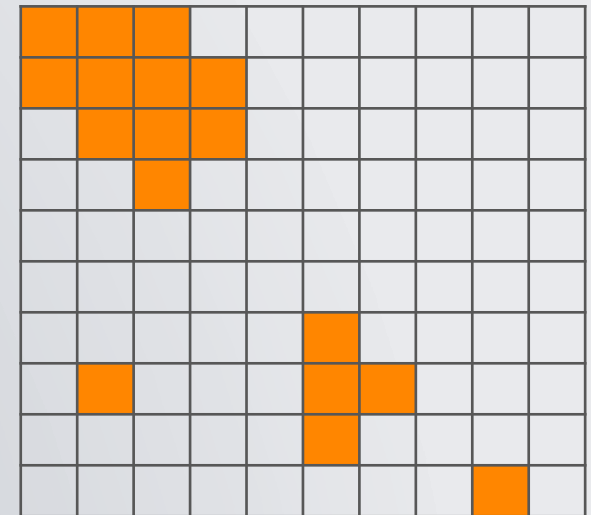
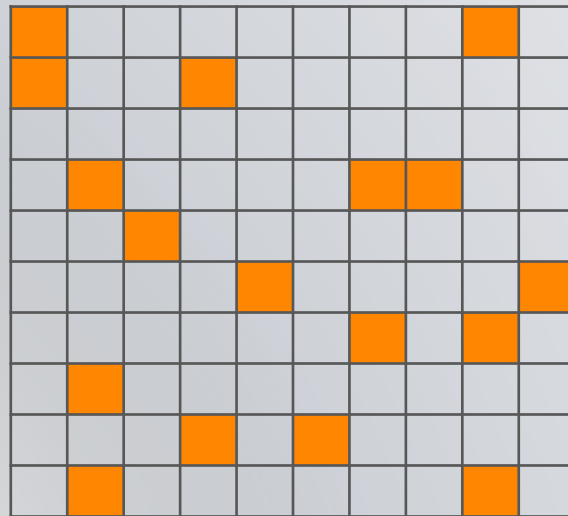
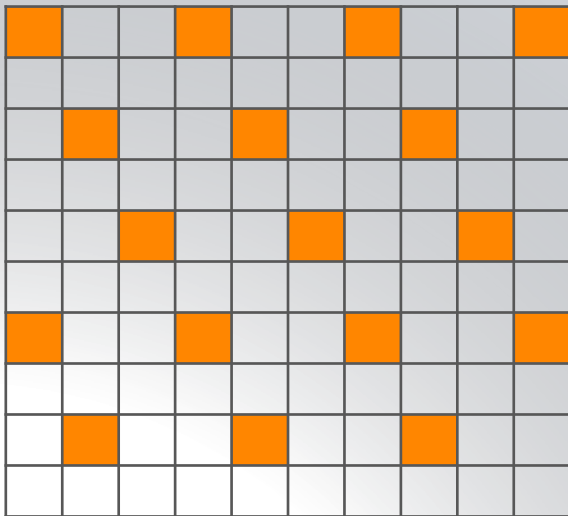
Klaster w epidemiologii

Wyszukiwanie nietypowych
SKUPISK
występowania chorób

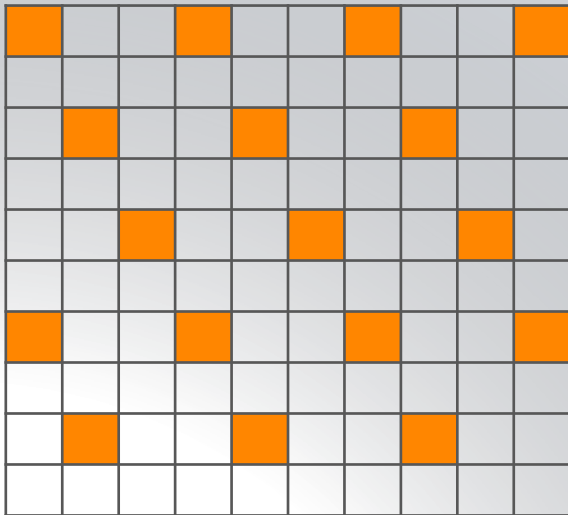
na określonym
obszarze
i w określonym
czasie



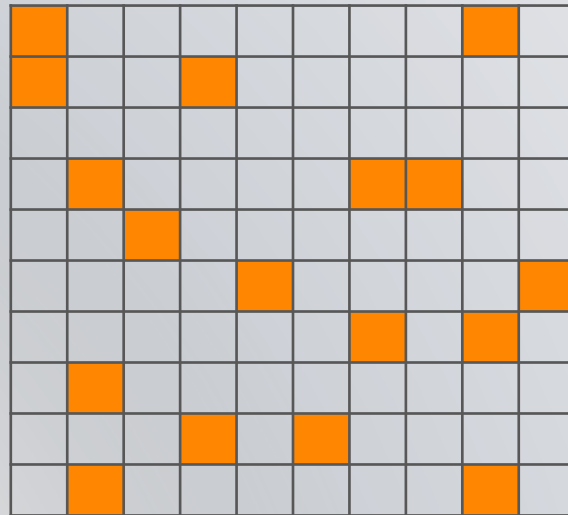
Autokorelacja



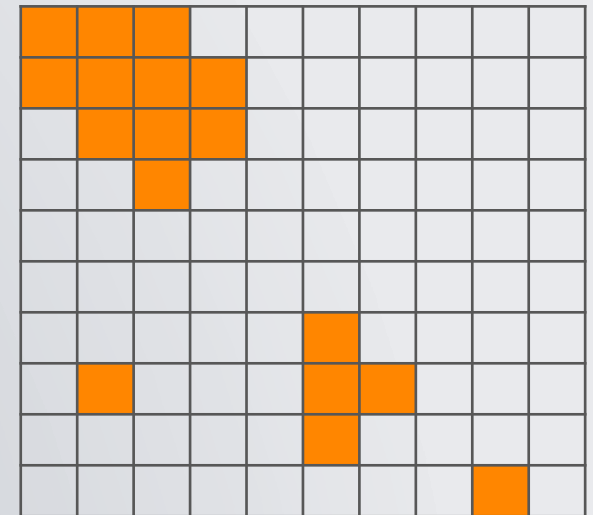
Autokorelacja



<0



losowość



>0

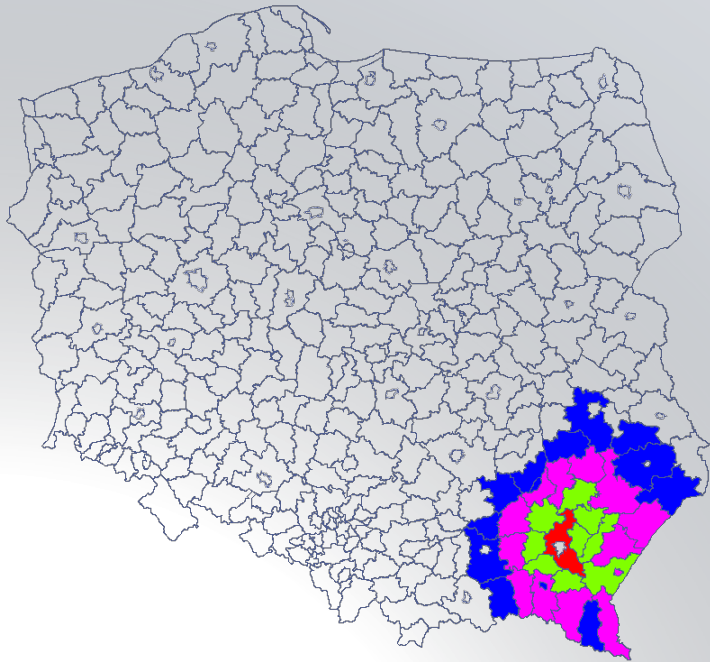
Sąsiedztwo

nr	1	2	3	4	5	6	7	8
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	1	1
6	0	1	1	0	1	0	1	0
7	0	1	0	0	1	1	0	1
8	0	0	0	0	1	0	1	0

nr obiektu

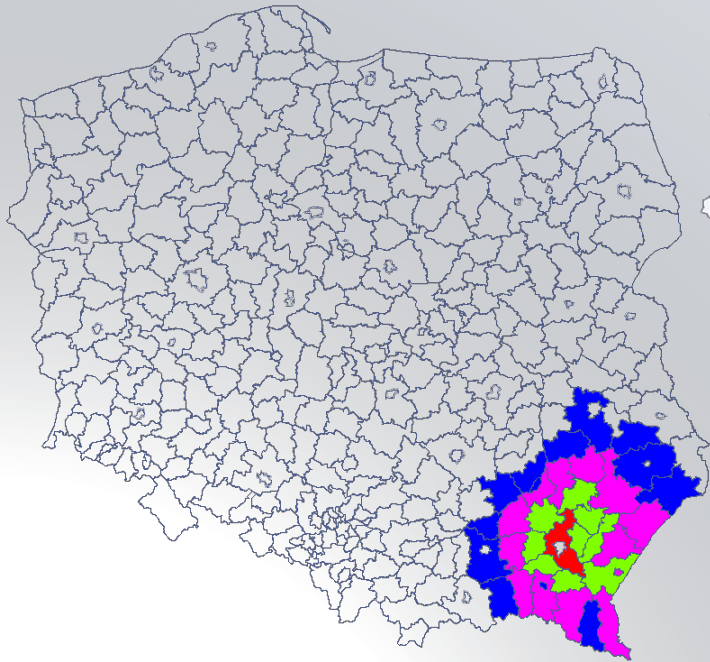


Sąsiedztwo

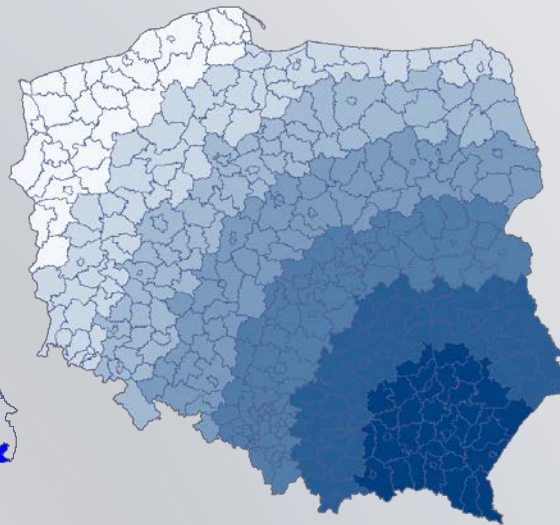


Queen

Sąsiedztwo

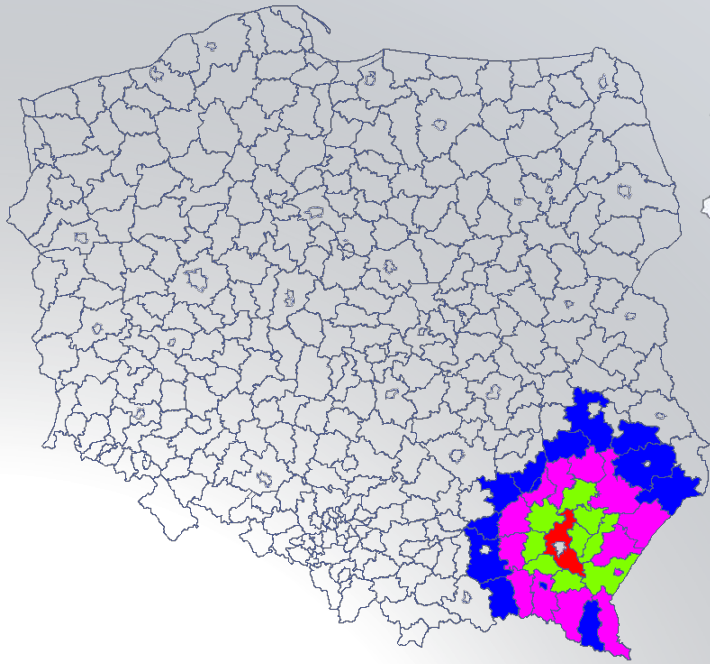


Queen

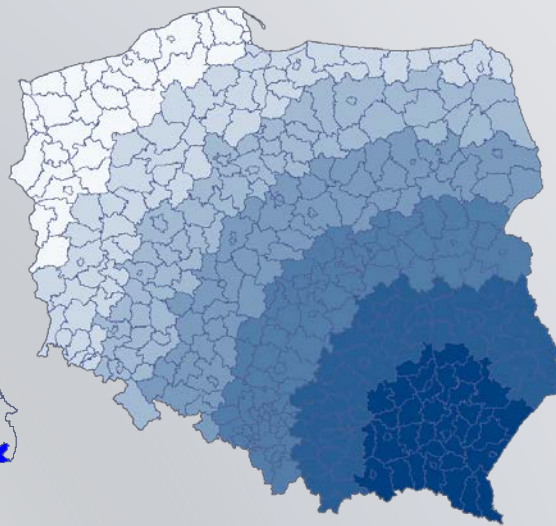


Euklidesowa

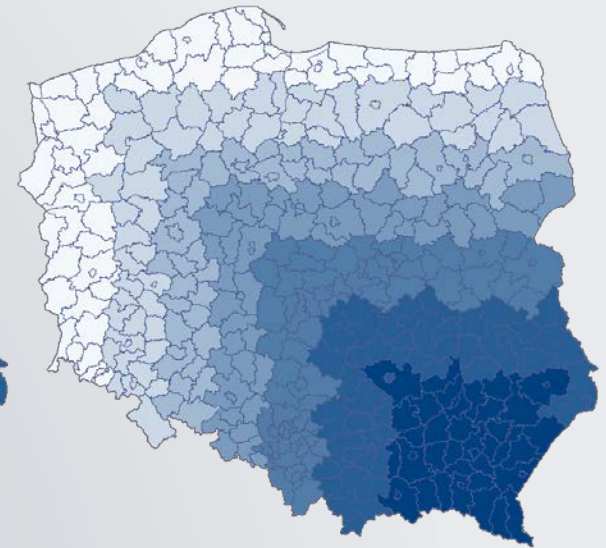
Sąsiedztwo



Queen



Euklidesowa



Czebyszewa

Definicja klasteru

KLASTER w rozumieniu wad wrodzonych, to skupisko przypadków wystąpienia wad wrodzonych w czasie i / lub przestrzeni, które wydaje się być **nieprzeciętne**

Definicja klasteru

KLASTER w rozumieniu wad wrodzonych, to skupisko przypadków wystąpienia wad wrodzonych w czasie i / lub przestrzeni, które wydaje się być **nieprzeciętne**



Nieprzeciętne?

Definicja klasteru



INNE niż zdefiniowana
PRZECIĘTNOŚĆ



Nieprzeciętne?

Definicja klasteru

Statystyka Lokalna Moran

<https://geodacenter.asu.edu/>

Częstość w obiekcie wyższa niż średnia:

- ✓ **wysokie** - wysokie
- ✓ **wysokie** - niskie

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x})}{\sigma^2}$$

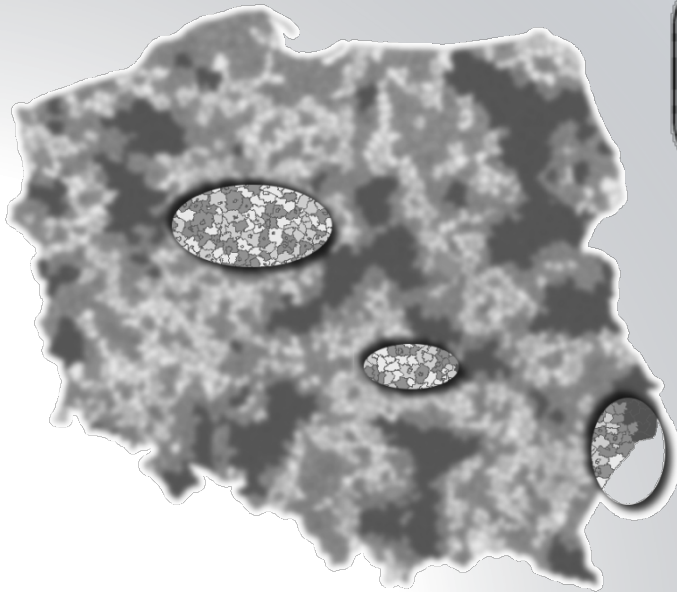
Nieprzeciętne?

Definicja klasteru

Statystyka Skanująca Kulldorffa

<http://www.satscan.org/>

Częstość wewnątrz okna wyższa niż na zewnątrz

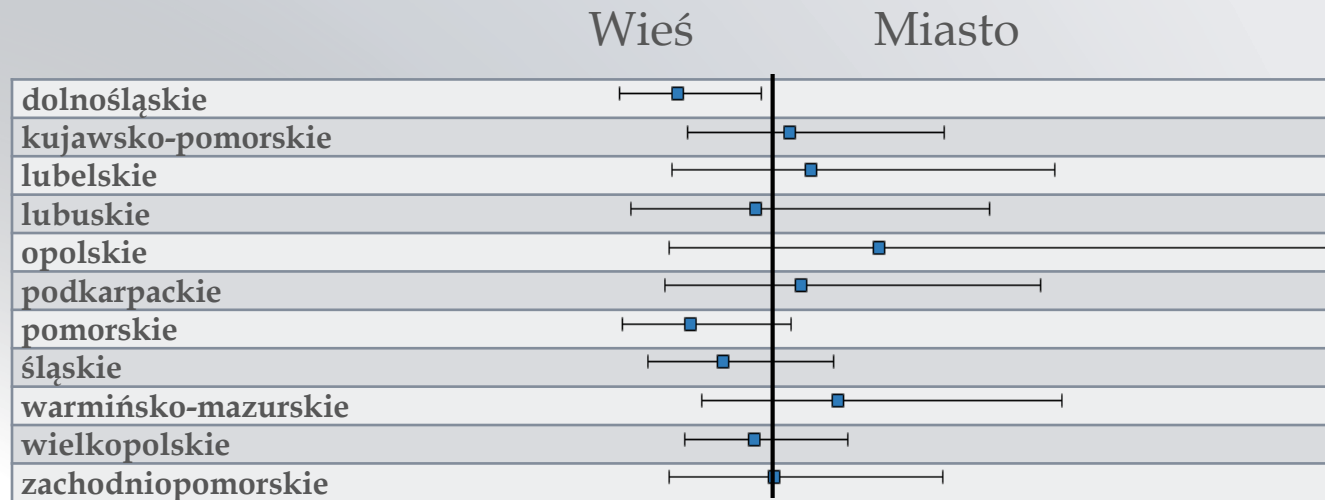


$$\left(\frac{c}{E[c]} \right)^c \left(\frac{C - c}{C - E[c]} \right)^{C-c} I_0$$

Nieprzeciętne?

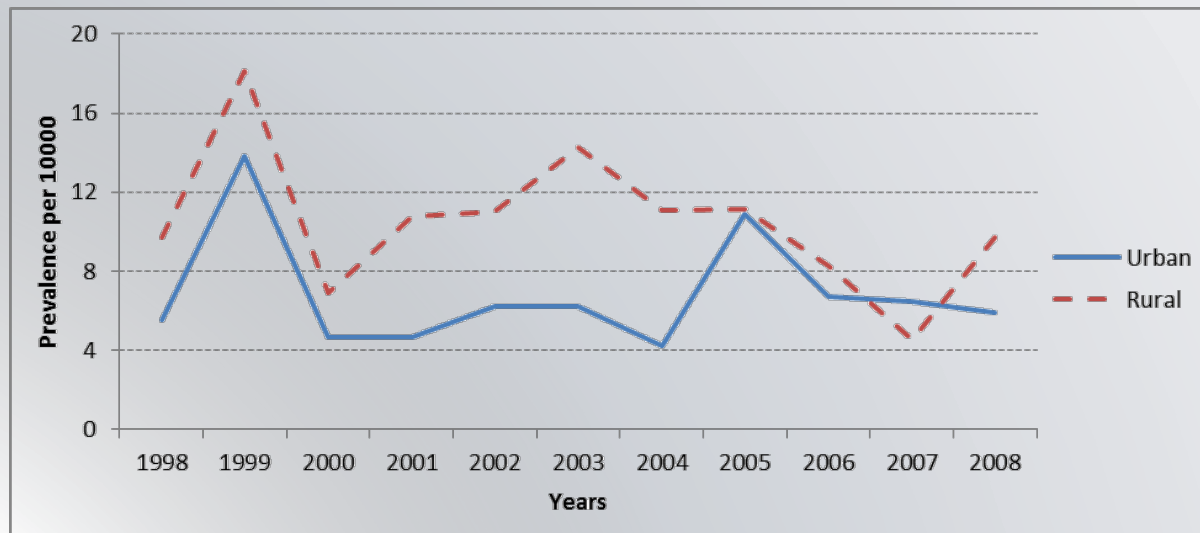
Wyniki analiz

Iloraz szans z 95% przedziałem ufności



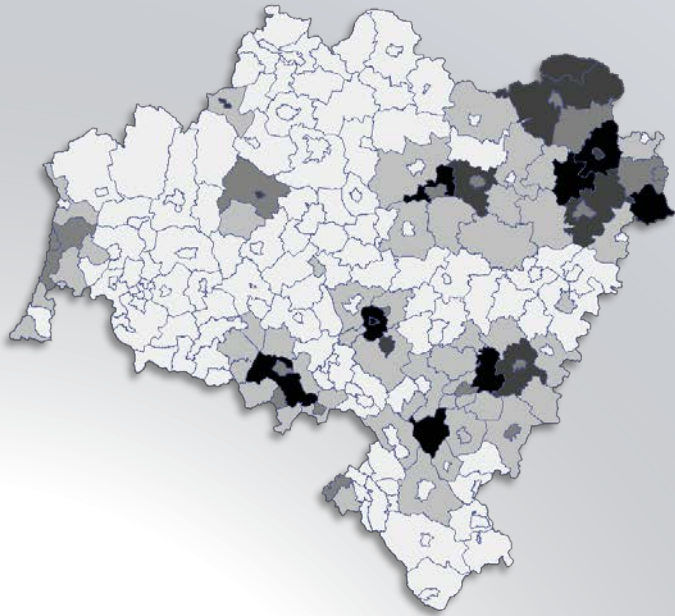
Rozszczep wargi i/lub podniebienia

Wyniki analiz



Rozszczep wargi i/lub podniebienia

Wyniki analiz



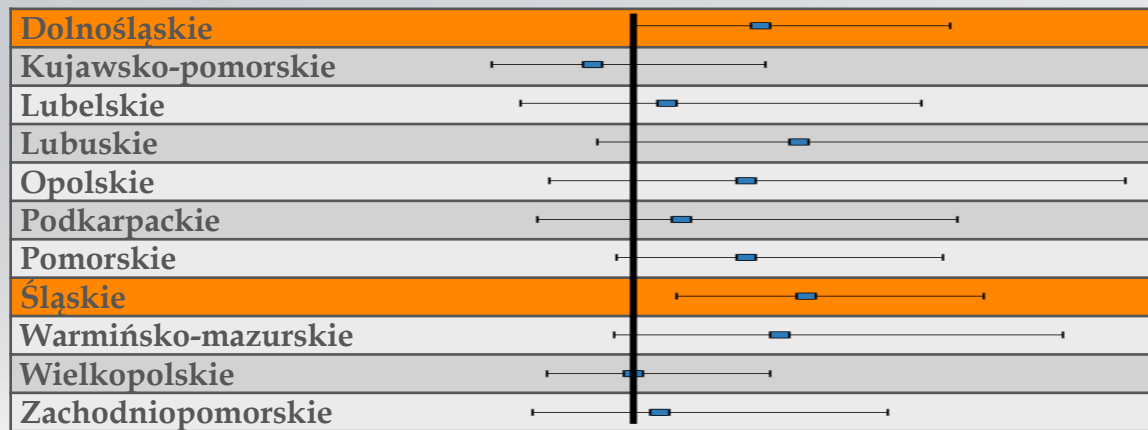
Rozszczep wargi i/lub podniebienia



Klaster: 24.05/10000
Poza kl.: 7.3 /10000
Wieś kl.: 32.5/10000
Miasto kl.: 20.4/10000

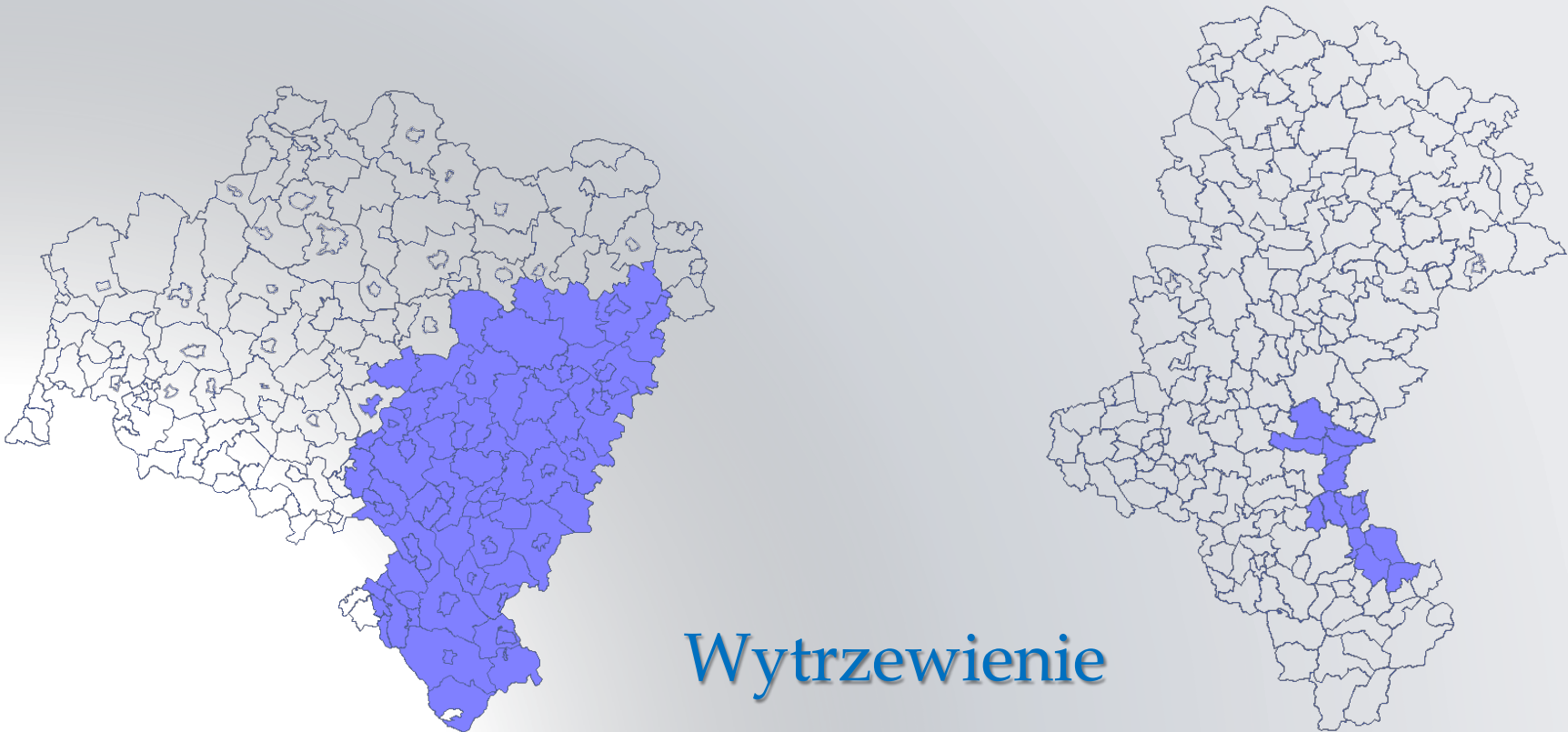
Klaster
1999 r

Wyniki analiz



Wytrzewienie

Wyniki analiz



Wytrzewienie

Wyniki analiz



Nie udało się znaleźć bezpośredniej przyczyny wewnętrznej typu:

palenie papierosów
wcześniejsze poronienia
niska masa urodzeniowa
wcześniejszy tydzień ciąży

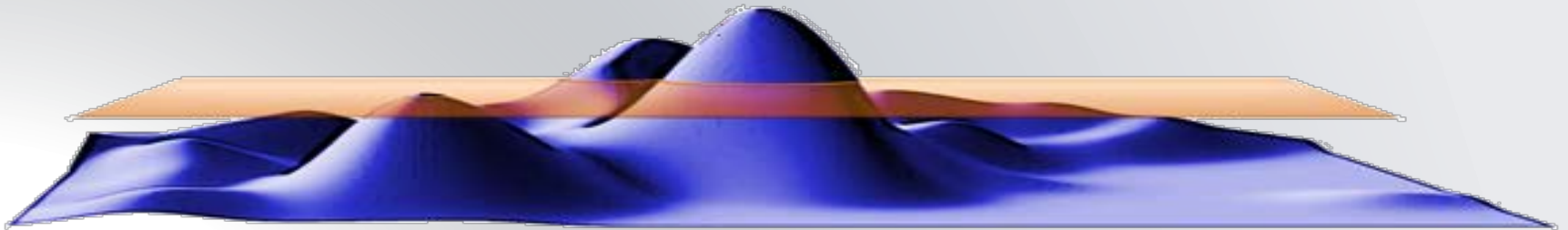
...

- ✓ 1997 „Powódź tysiąclecia”, która objęła również obszary klasteru
- ✓ Wysokie częstości obserwowane były na terenie ówczesnej Czechosłowacji
- ✓ Ok 85% składowisk odpadów na terenach wiejskich woj. dolnośląskiego
- ✓ Generowane odpady przemysłowe >27%
- ✓ Niskie odsetki korzystania z wód wodociągowych (brak: 30%) i kanalizacji (brak: 91%) na terenach wiejskich (woj. dolnośląskiego)
- ✓ Nie można wykluczyć aspektu genetycznego

Definicja klasteru

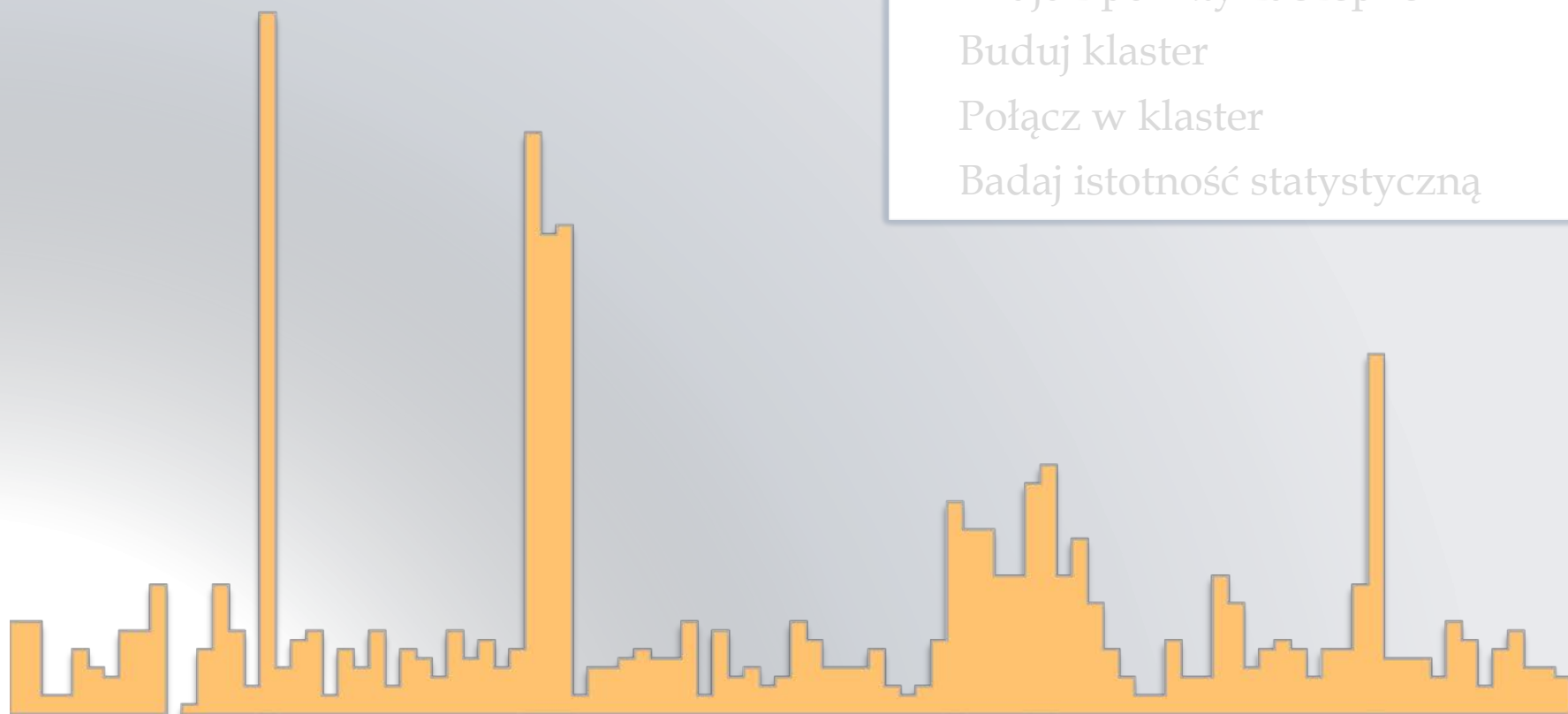
Statystyka CutLine

Częstość wewnątrz obiektu (grupy obiektów) wyższa niż
częstość **zadana**



Nieprzeciętne?

Algorytm



✓ Wyznacz współczynniki dla gmin

Podaj linię odcięcia CutLine

Wygładź powierzchnię

Znajdź punkty zaczepne

Buduj klaster

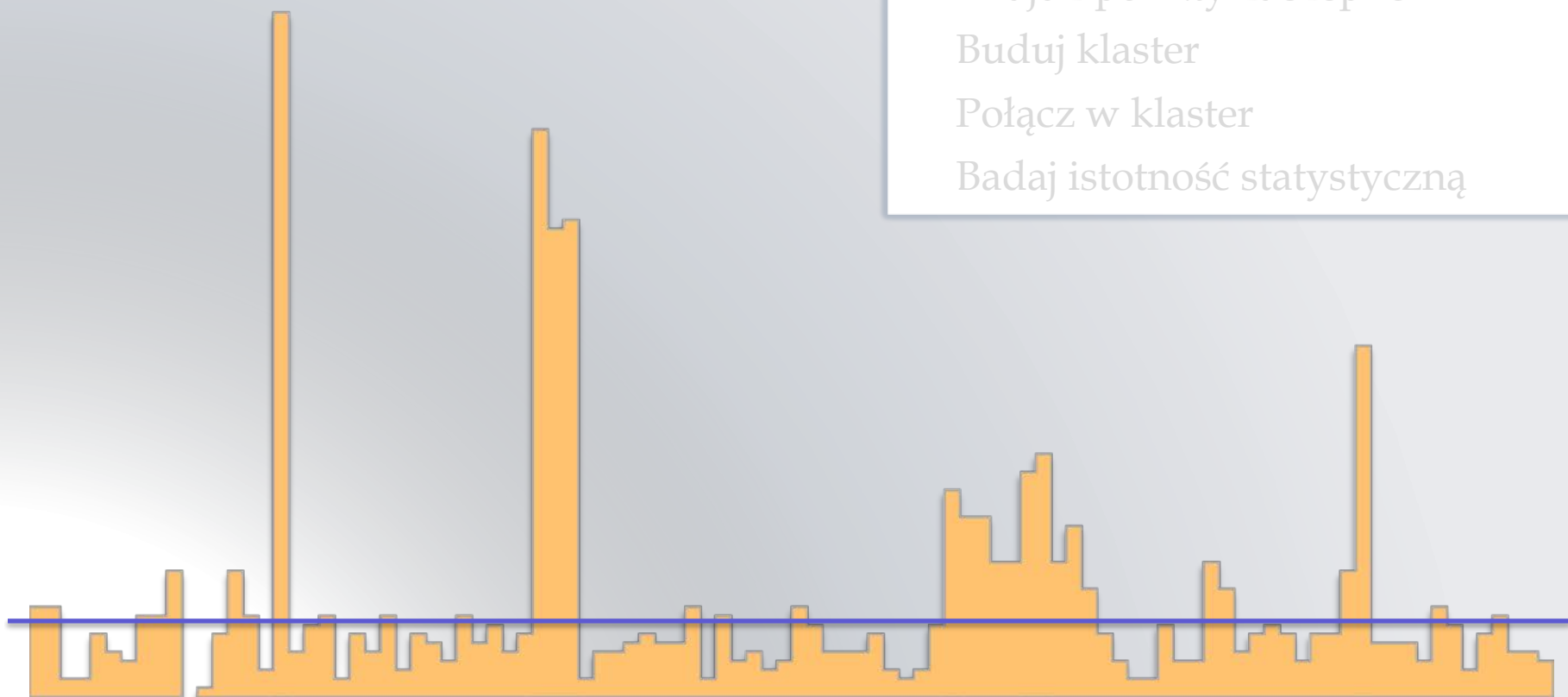
Połącz w klaster

Badaj istotność statystyczną

Statystyka CutLine

Algorytm

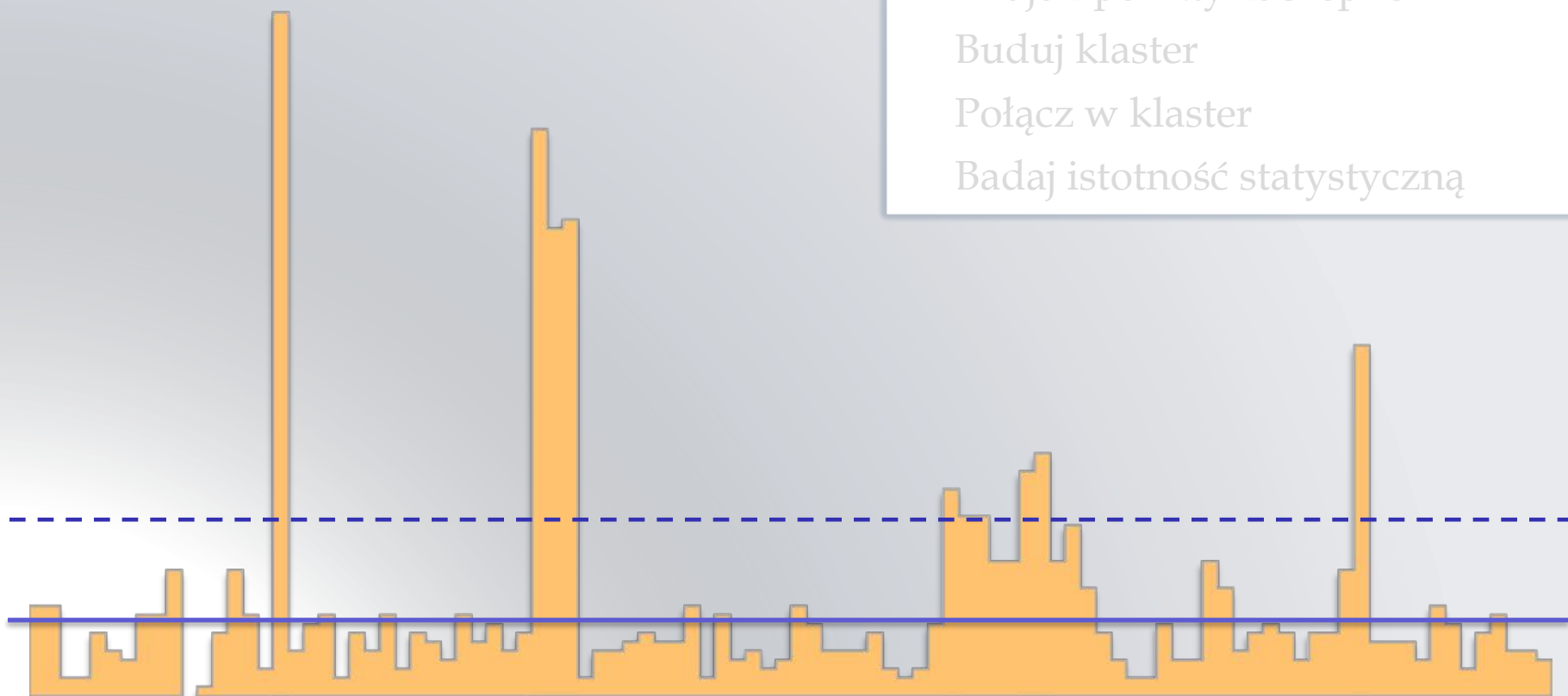
- ✓ Wyznacz współczynniki dla gmin
- ✓ **Podaj linię odcięcia CutLine**
- Wygładź powierzchnię
- Znajdź punkty zaczepne
- Buduj klaster
- Połącz w klaster
- Badaj istotność statystyczną



Statystyka CutLine

Algorytm

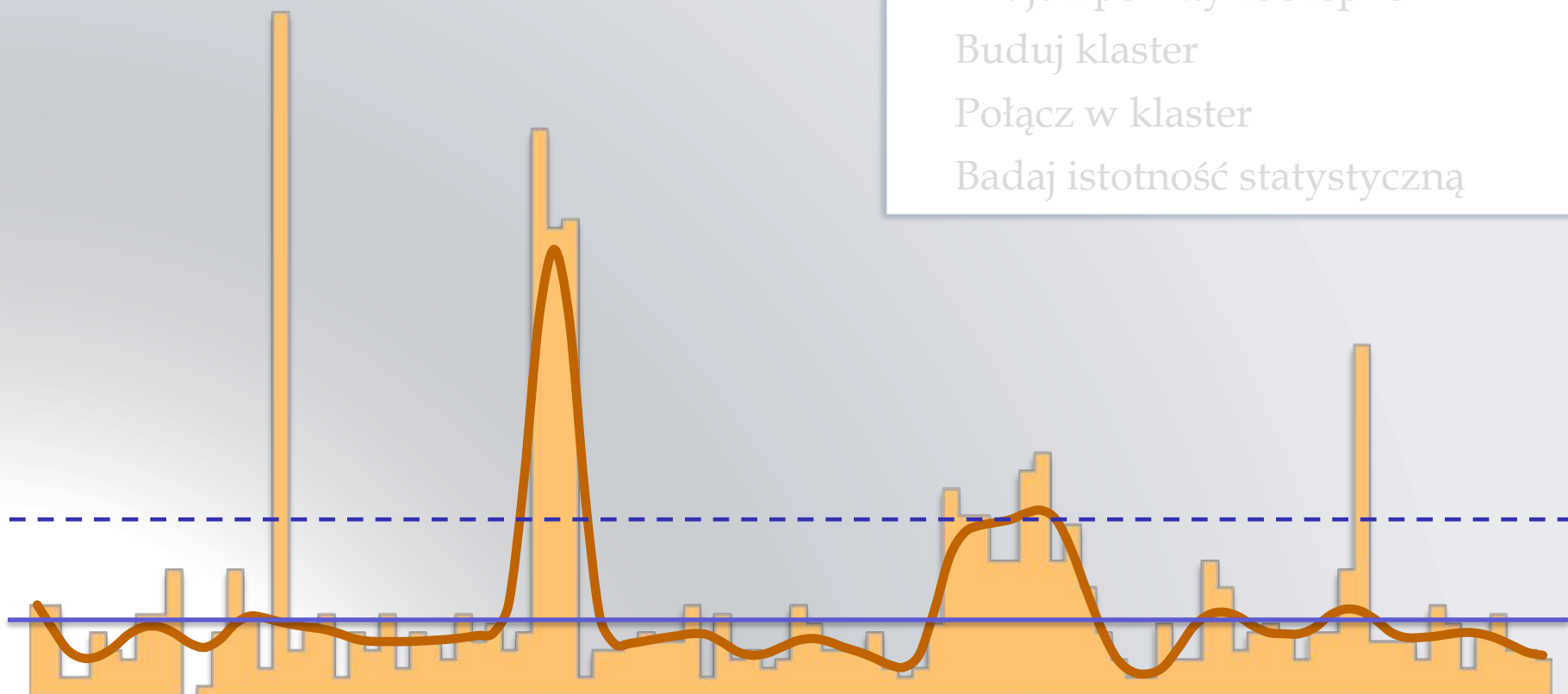
- ✓ Wyznacz współczynniki dla gmin
- ✓ **Podaj linię odcięcia CutLine**
- Wygładź powierzchnię
- Znajdź punkty zaczepne
- Buduj klaster
- Połącz w klaster
- Badaj istotność statystyczną



Statystyka CutLine

Algorytm

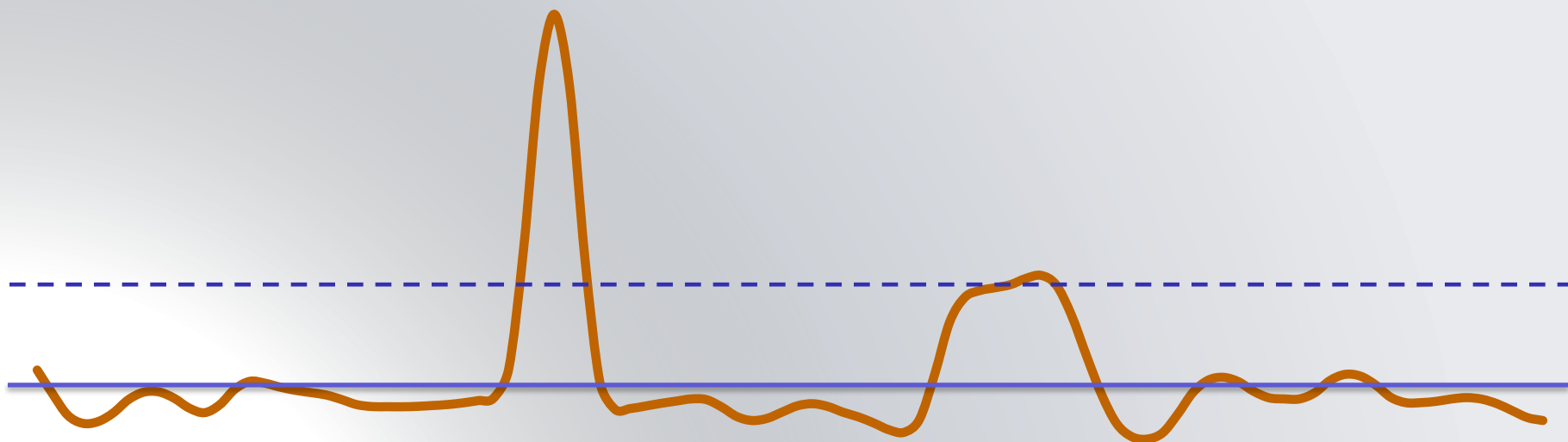
- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ **Wygładź powierzchnię**
 - Znajdź punkty zaczepne
 - Buduj klaster
 - Połącz w klaster
 - Badaj istotność statystyczną



Statystyka CutLine

Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ **Wygładź powierzchnię**
 - Znajdź punkty zaczepne
 - Buduj klaster
 - Połącz w klaster
 - Badaj istotność statystyczną



Statystyka CutLine

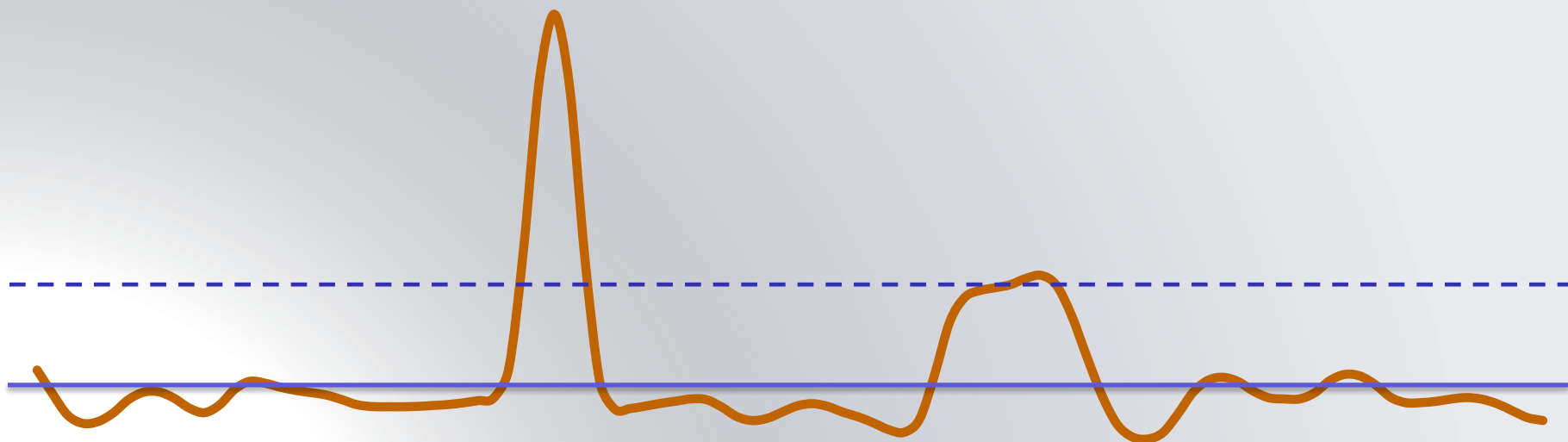
Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ **Znajdź punkty zaczepne**

Buduj klaster

Połącz w klaster

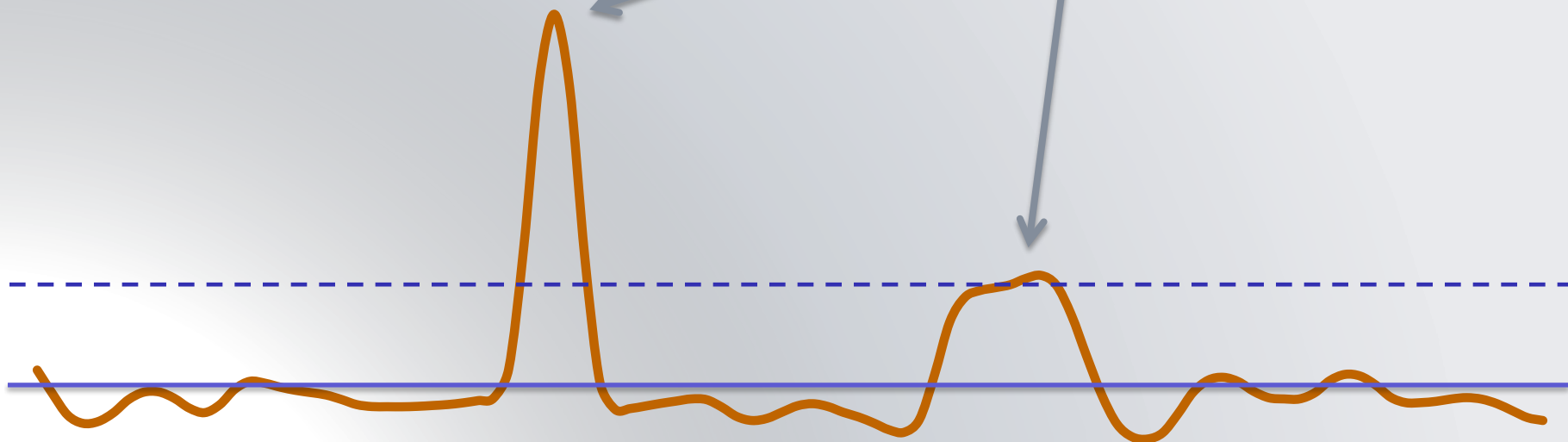
Badaj istotność statystyczną



Statystyka CutLine

Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ **Znajdź punkty zaczepne**
 - Buduj klaster
 - Połącz w klaster
 - Badaj istotność statystyczną



Statystyka CutLine

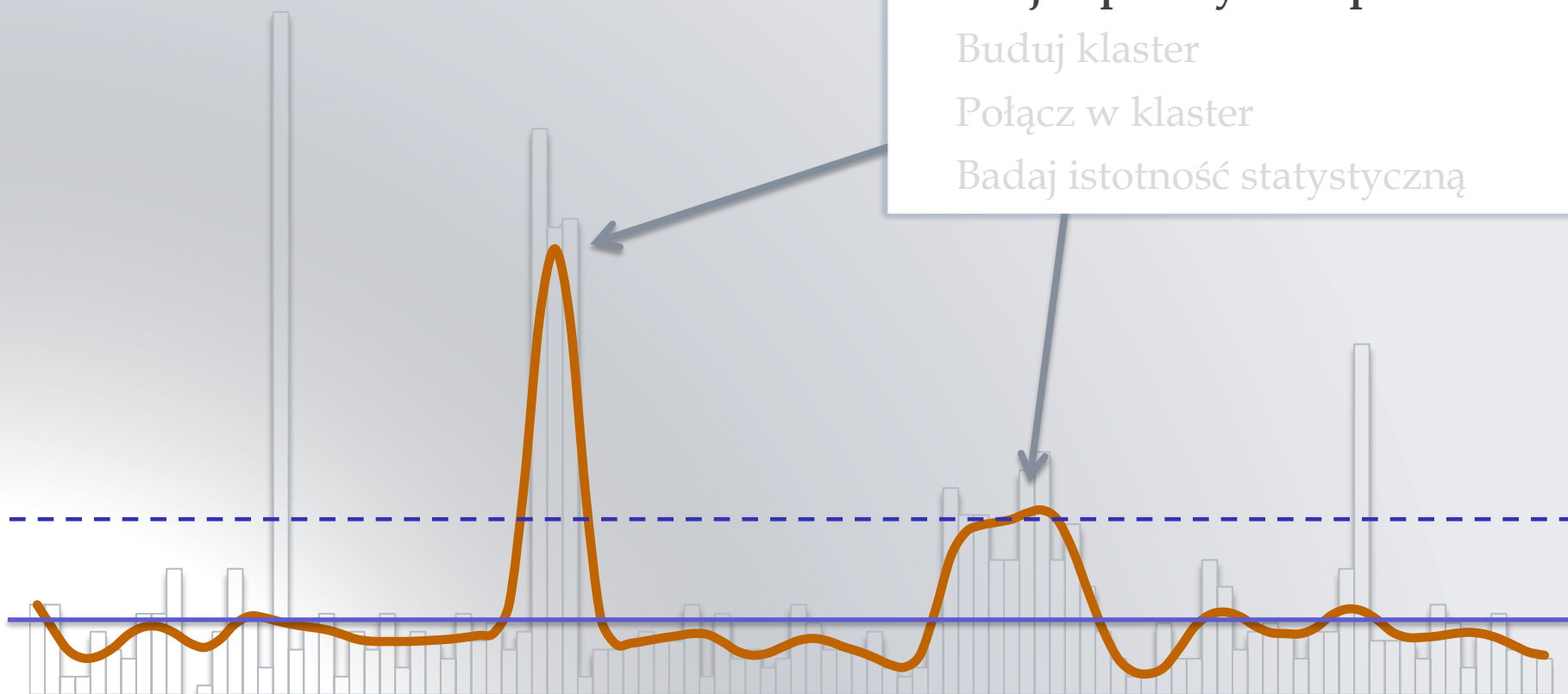
Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ **Znajdź punkty zaczepne**

Buduj klaster

Połącz w klaster

Badaj istotność statystyczną



Statystyka CutLine

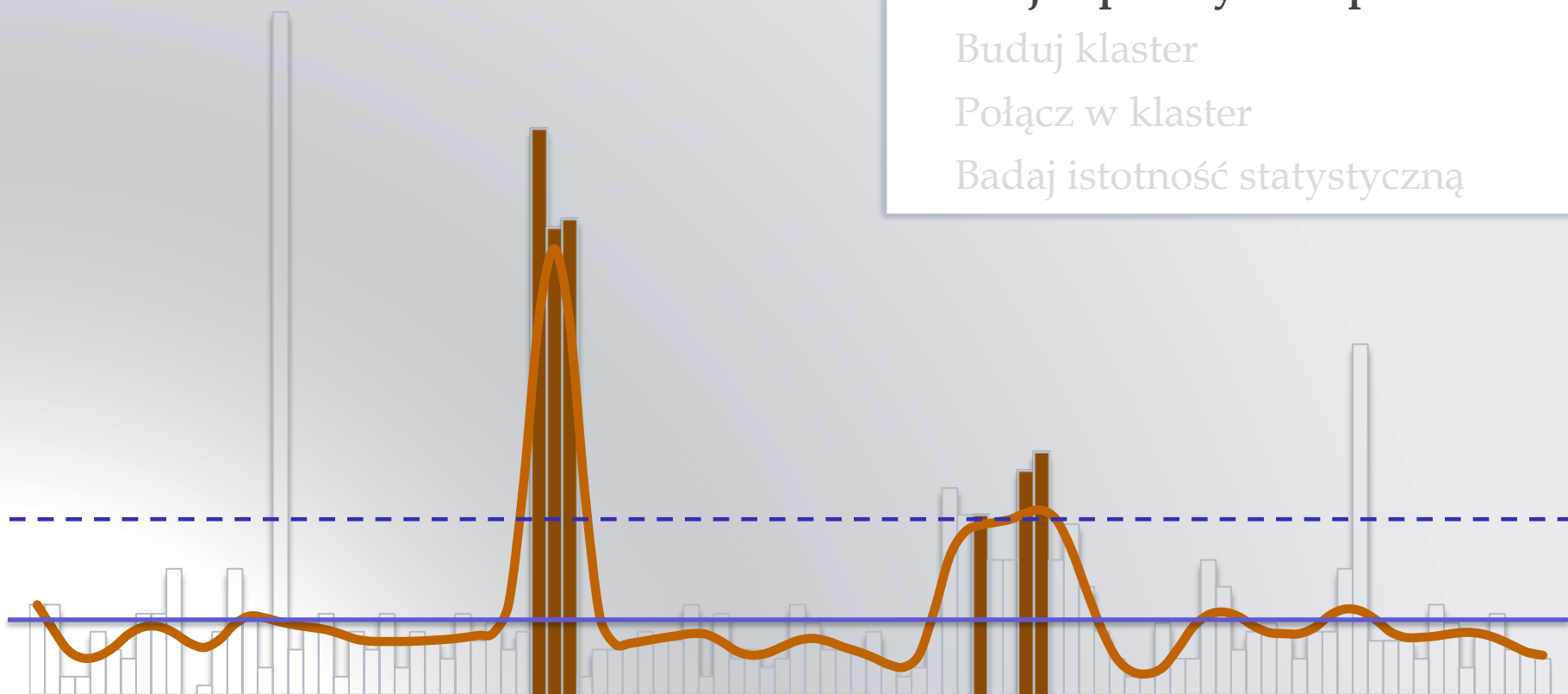
Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ **Znajdź punkty zaczepne**

Buduj klaster

Połącz w klaster

Badaj istotność statystyczną



Statystyka CutLine

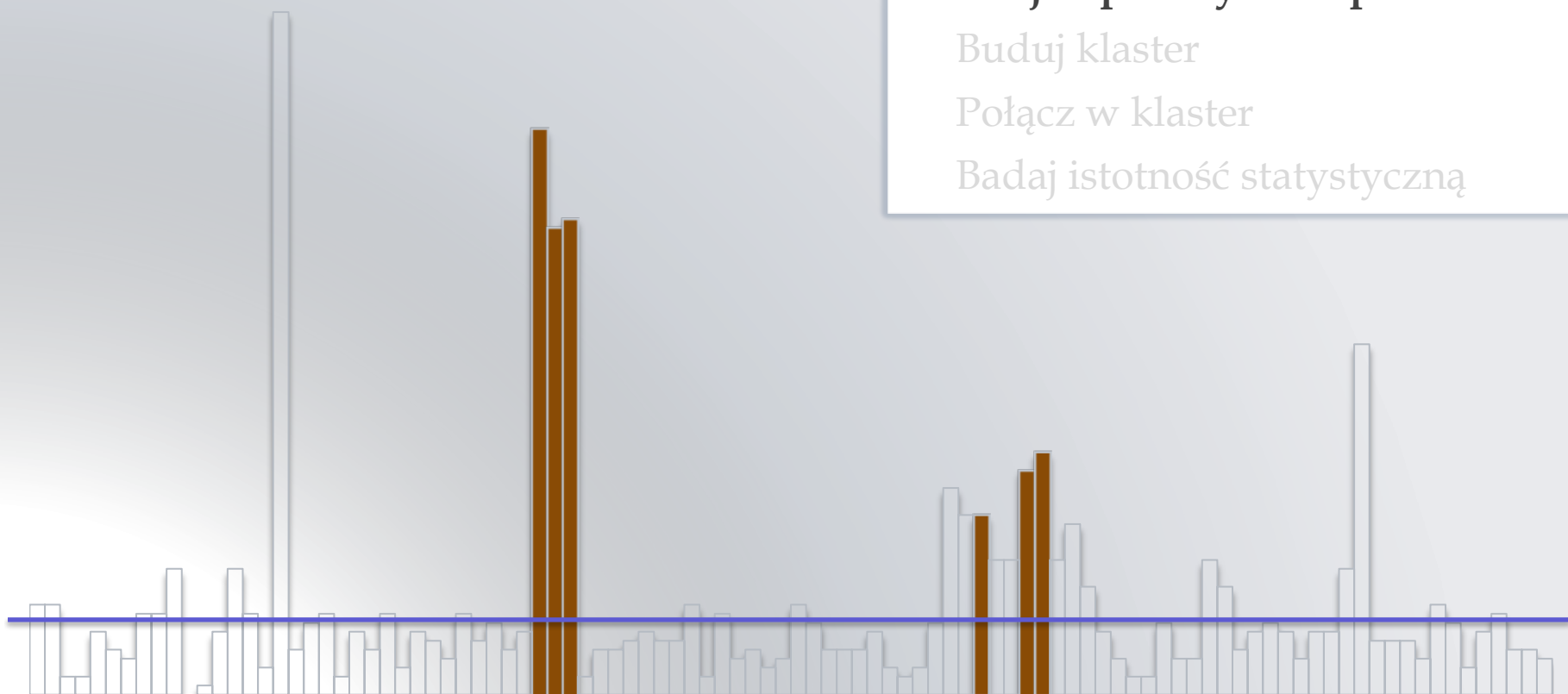
Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ **Znajdź punkty zaczepne**

Buduj klaster

Połącz w klaster

Badaj istotność statystyczną



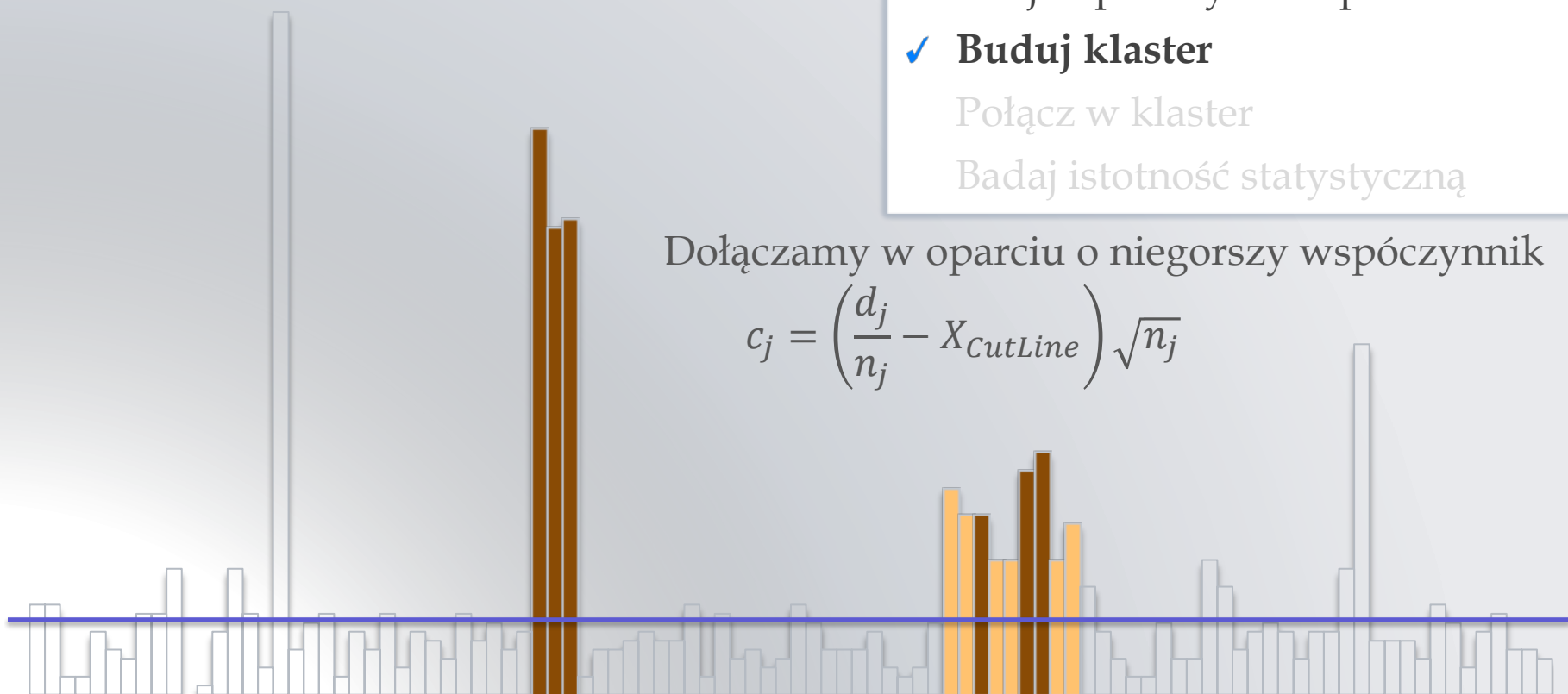
Statystyka CutLine

Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ Znajdź punkty zaczepne
- ✓ **Buduj klaster**
 - Połącz w klaster
 - Badaj istotność statystyczną

Dołączamy w oparciu o niegorszy współczynnik

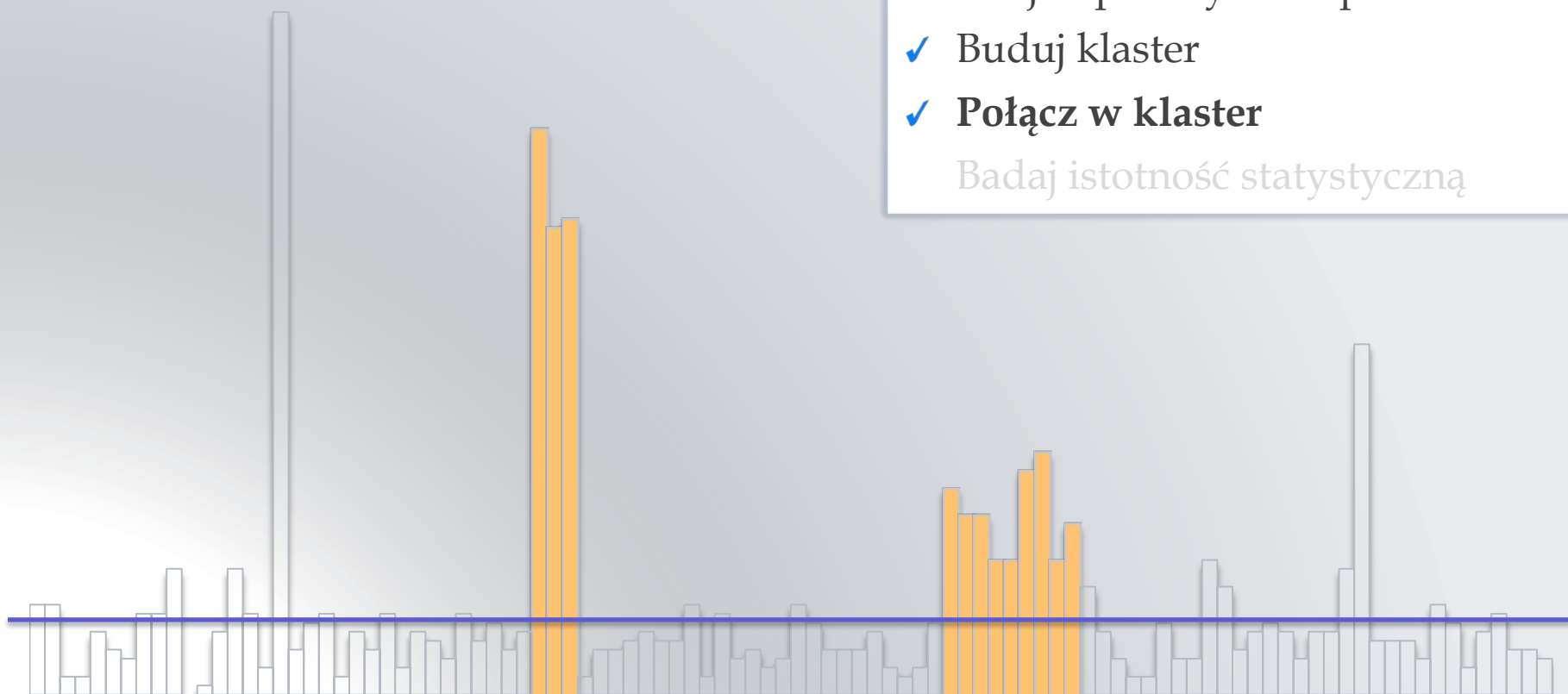
$$c_j = \left(\frac{d_j}{n_j} - X_{CutLine} \right) \sqrt{n_j}$$



Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ Znajdź punkty zaczepne
- ✓ Buduj klaster
- ✓ **Połącz w klaster**

Badaj istotność statystyczną

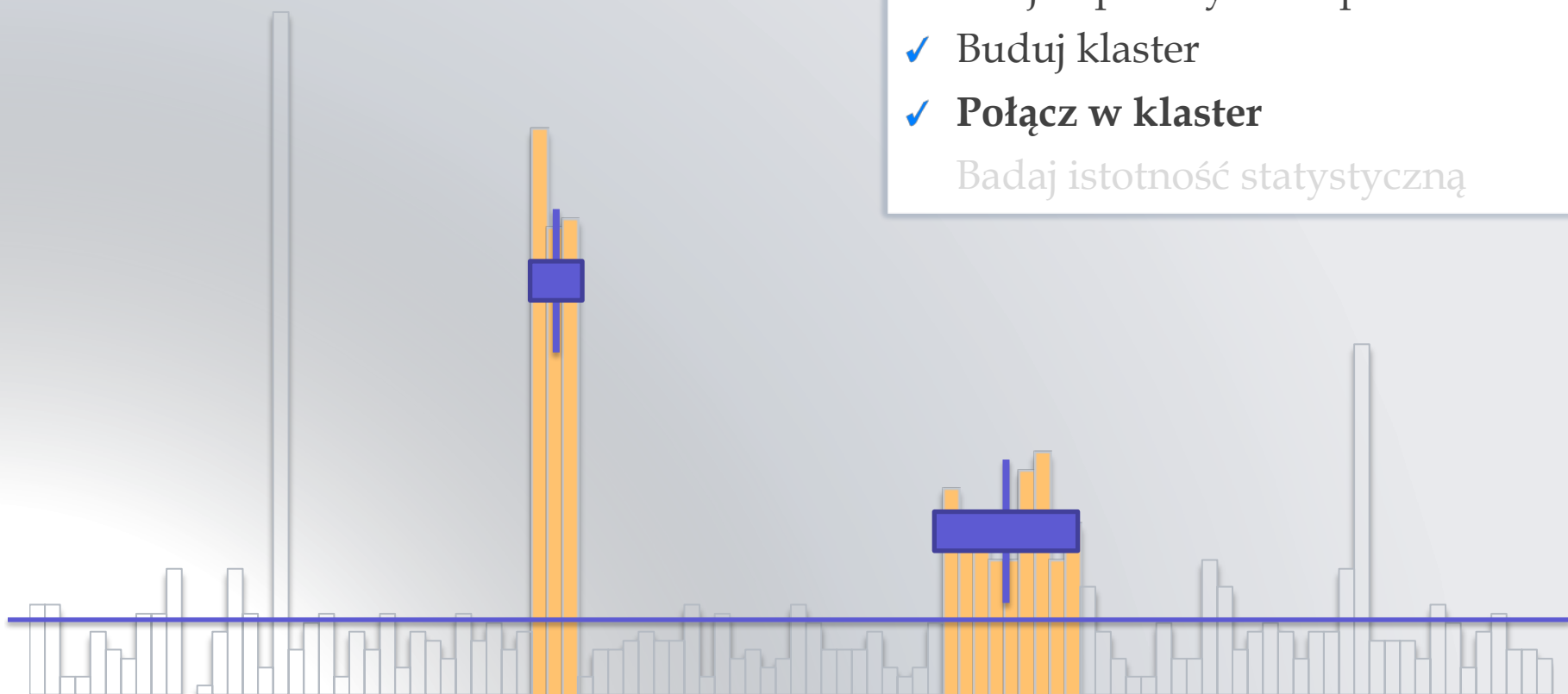


Statystyka CutLine

Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ Znajdź punkty zaczepne
- ✓ Buduj klaster
- ✓ **Połącz w klaster**

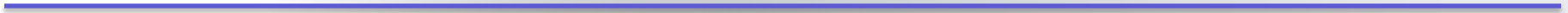
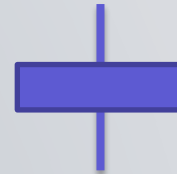
Badaj istotność statystyczną



Statystyka CutLine

Algorytm

- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ Znajdź punkty zaczepne
- ✓ Buduj klaster
- ✓ Połącz w klaster
- ✓ **Badaj istotność statystyczną**



Statystyka CutLine

Algorytm

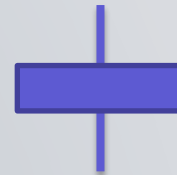
Binominal test
for one proportion



correction of
multiple
comparisons

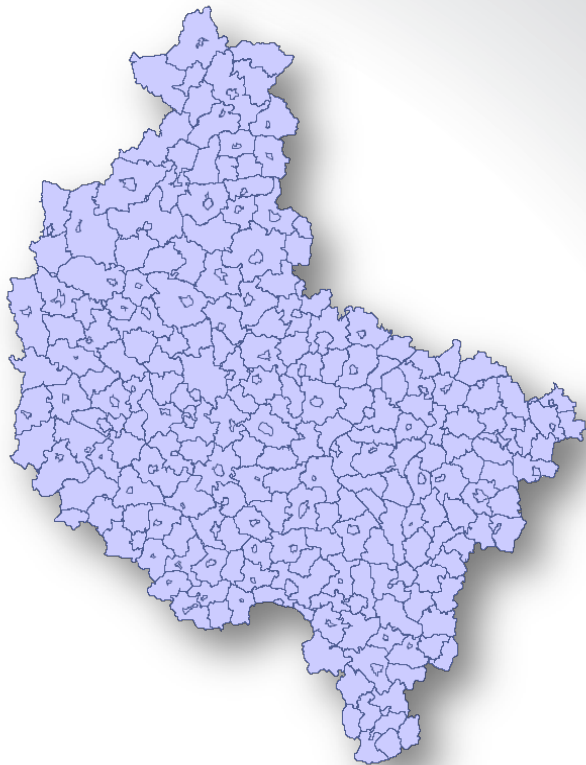


- ✓ Wyznacz współczynniki dla gmin
- ✓ Podaj linię odcięcia CutLine
- ✓ Wygładź powierzchnię
- ✓ Znajdź punkty zaczepne
- ✓ Buduj klaster
- ✓ Połącz w klaster
- ✓ **Badaj istotność statystyczną**



Statystyka CutLine

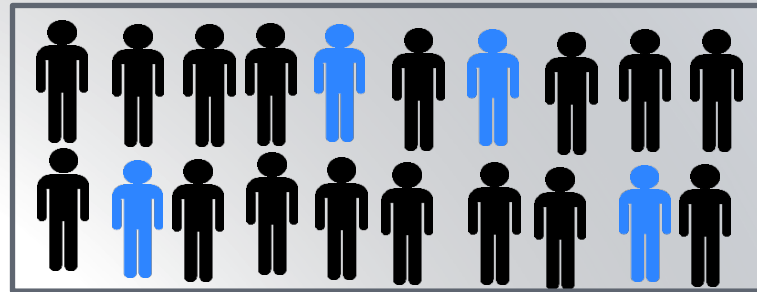
Symulacja



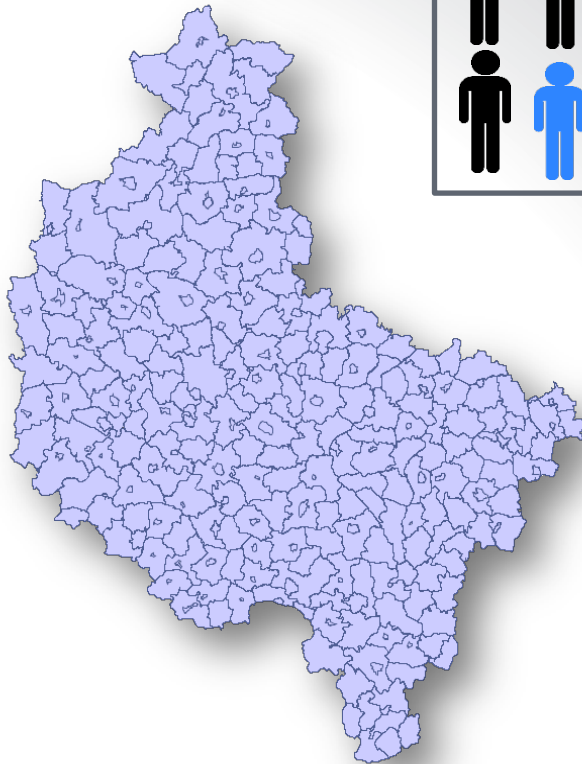
Statystyka CutLine

Symulacja

Populacja



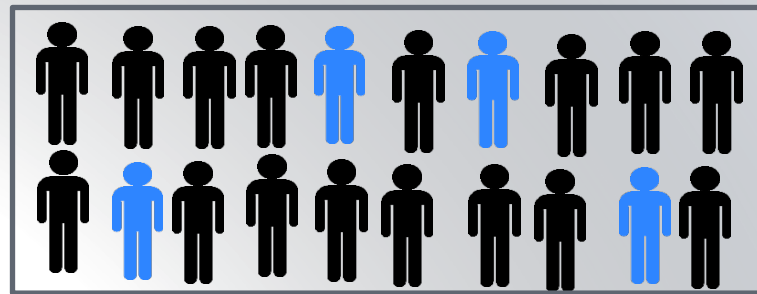
$N = 3467016$



Statystyka CutLine

Symulacja

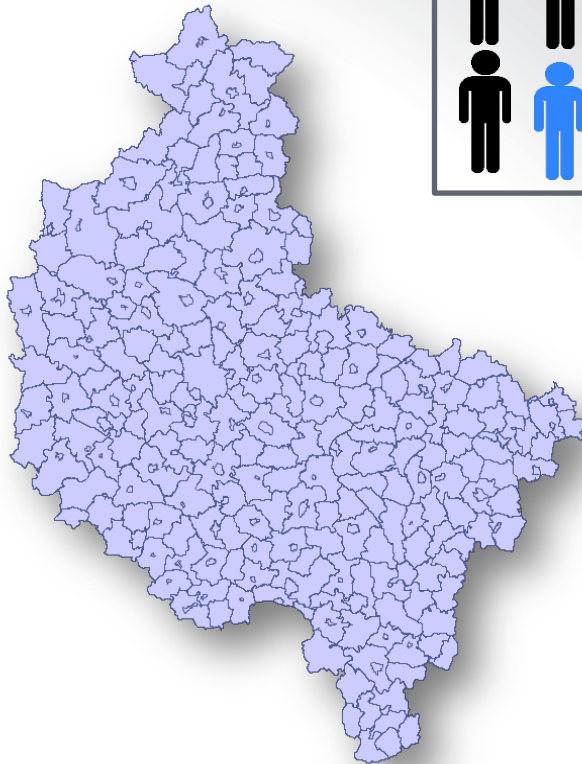
Populacja



$N = 3467016$

Chorzy

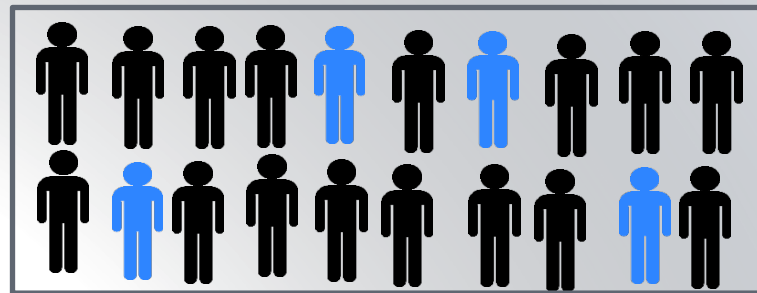
$D=3467$



Statystyka CutLine

Symulacja

Populacja



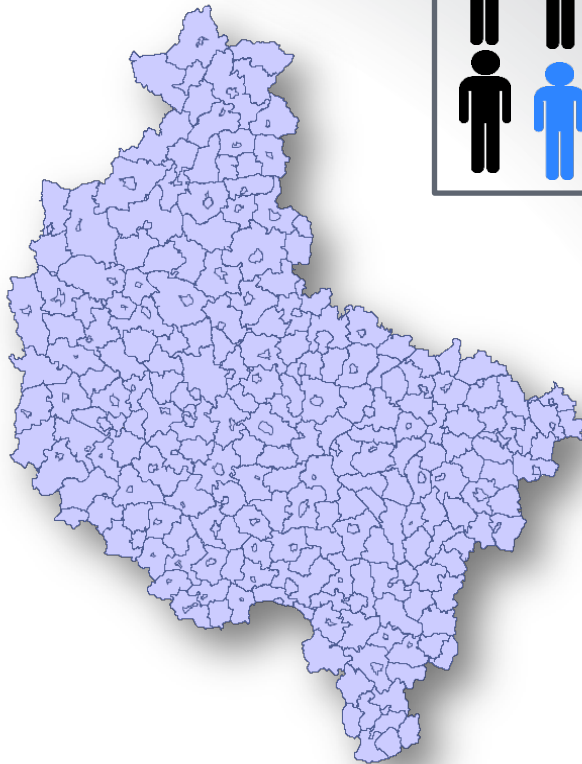
$$N = 3467016$$

Chorzy

$$D=3467 \longrightarrow$$

$$\text{prev} = \frac{D}{N}$$

0.001



Statystyka CutLine

Symulacja

brak klasterów

- $H_0: (d_1, d_2, \dots, d_m) \sim \text{multinomial} \left(\frac{d_1}{n_1}, \frac{d_2}{n_2}, \dots, \frac{d_m}{n_m} \right)$

występowanie klasterów

- $H_A: (d_1, d_2, \dots, d_m) \sim \text{multinomial} \left(r_1 \frac{d_1}{n_1}, r_2 \frac{d_2}{n_2}, \dots, r_m \frac{d_m}{n_m} \right)$

0.001

Statystyka CutLine

Symulacja

100 powtórzeń

RR=1

RR=1.5

RR=2.5

RR=4

RR=10

✓ CutLine

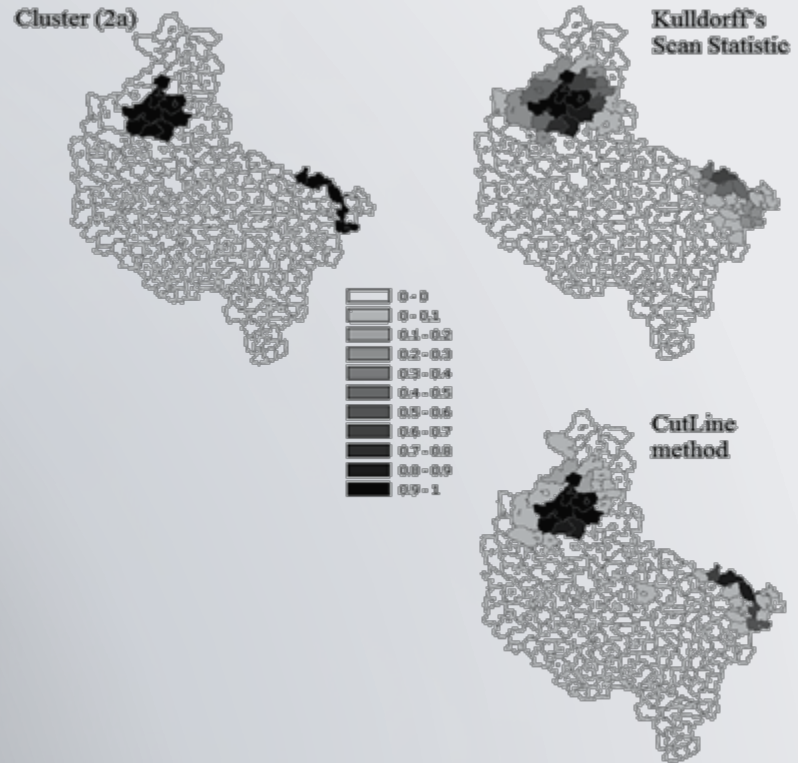
✓ Kulldorff's scan statistic

0.001

Statystyka CutLine

Symulacja

RR=2.5



Statystyka CutLine

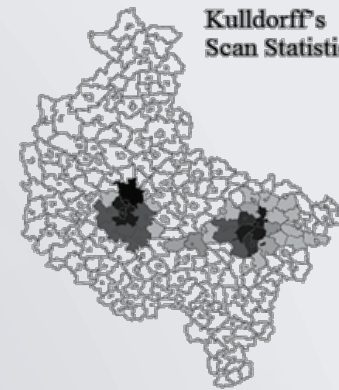
Symulacja

RR=2.5

Cluster (2b)



Kulldorff's
Scan Statistic



CutLine
method

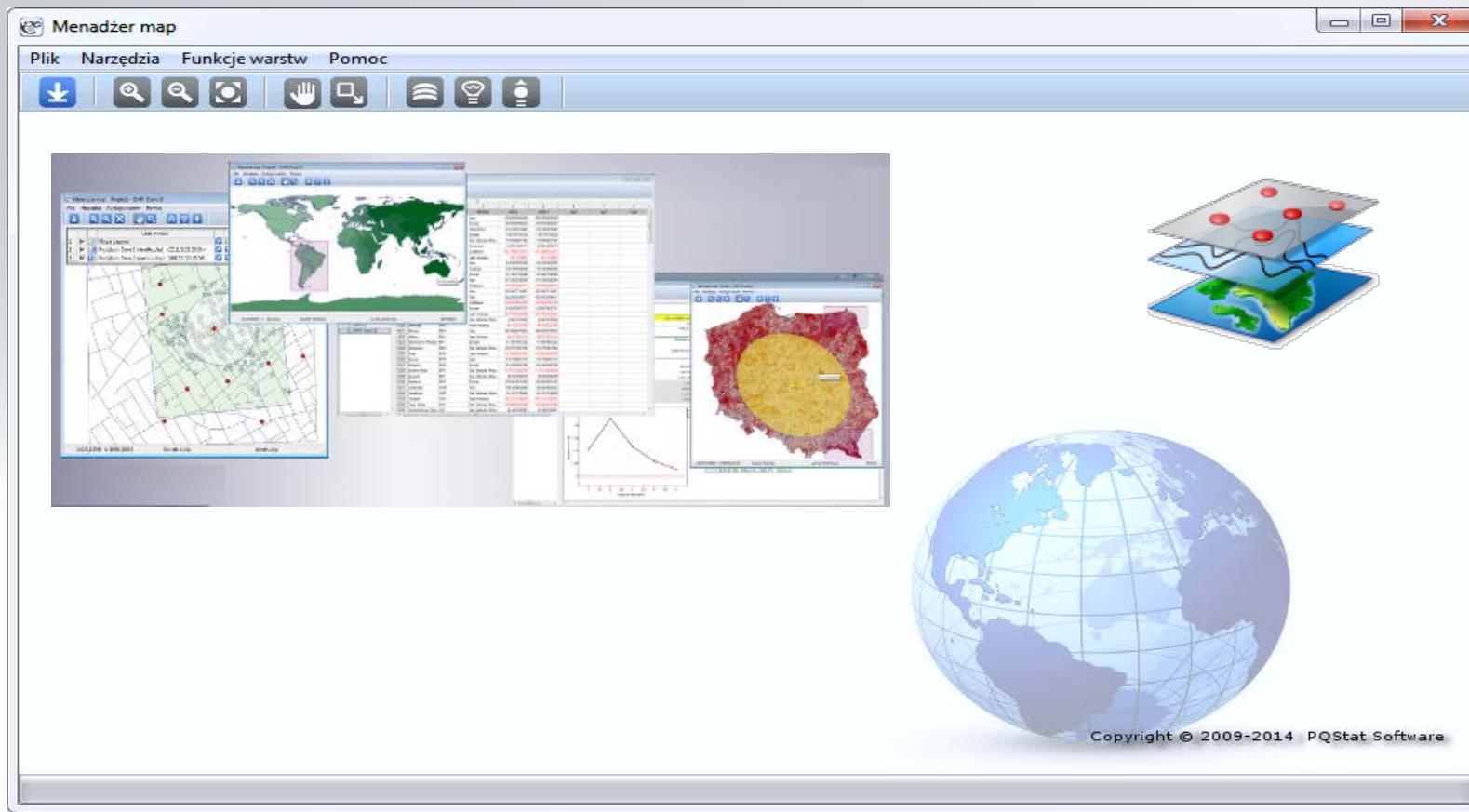


Symulacja

	Methods	Sensitivity	Specificity	PPV	NPV	Accuracy
(1) rozkład losowy						
RR=1	CutLine	-	0.999	-	-	0.999
	Kulldorff's scan statistic	-	0.995	-	-	0.995
(2a) dwa klastery						
RR=1.5	CutLine	0.217	0.996	0.753	0.954	0.951
	Kulldorff's scan statistic	0.488	0.970	0.495	0.969	0.942
RR=2.5	CutLine	0.811	0.997	0.935	0.989	0.986
	Kulldorff's scan statistic	0.713	0.969	0.582	0.982	0.954
RR=4	CutLine	0.979	0.988	0.829	0.999	0.987
	Kulldorff's scan statistic	0.488	0.970	0.495	0.969	0.942
RR=10	CutLine	1	0.985	0.802	1	0.986
	Kulldorff's scan statistic	0.979	0.960	0.595	0.999	0.961
(2b) klaster wzdłuż rzeki						
RR=1.5	CutLine	0.158	0.968	0.703	0.933	0.911
	Kulldorff's scan statistic	0.181	0.973	0.497	0.891	0.873
RR=2.5	CutLine	0.523	0.968	0.703	0.933	0.911
	Kulldorff's scan statistic	0.328	0.973	0.637	0.909	0.891
RR=4	CutLine	0.736	0.967	0.765	0.962	0.938
	Kulldorff's scan statistic	0.436	0.968	0.663	0.922	0.900
RR=10	CutLine	0.912	0.967	0.802	0.987	0.960
	Kulldorff's scan statistic	0.493	0.962	0.653	0.929	0.902

Statystyka CutLine

Metoda oprogramowana



Statystyka CutLine

Dziękuję

Barbara Więckowska
basia@ump.edu.pl