

Kokoszka_Karolina_projekt

Karolina Kokoszka

2024-11-24

```
# wczytanie danych
osiedla <- st_read(paste0(base_path,"AnalizaDanychPrzestrzennych/cwiczenia/lab5_projekt/osiedla.shp"))

## Reading layer `osiedla' from data source
##   `C:\Users\Marcin\OneDrive - Hauraton GmbH & Co. KG\Dokumente\STUDIA\SEMESTR V\AnalizaDanychPrzestrzennych\osiedla.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 141 features and 30 fields
## Geometry type: POLYGON
## Dimension:      XY
## Bounding box:   xmin: 7413437 ymin: 5537344 xmax: 7443955 ymax: 5555031
## Projected CRS: ETRS89 / Poland CS2000 zone 7

punkty <- st_read(paste0(base_path,"AnalizaDanychPrzestrzennych/cwiczenia/lab6/zestaw3v2_XYTableToP_Project.shp"))

## Reading layer `zestaw3v2_XYTableToP_Project' from data source
##   `C:\Users\Marcin\OneDrive - Hauraton GmbH & Co. KG\Dokumente\STUDIA\SEMESTR V\AnalizaDanychPrzestrzennych\zestaw3v2_XYTableToP_Project.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 2000 features and 4 fields
## Geometry type: POINT
## Dimension:      XY
## Bounding box:   xmin: 7414276 ymin: 5538263 xmax: 7441201 ymax: 5553290
## Projected CRS: ETRS89 / Poland CS2000 zone 7

# wyodrębnienie współrzędnych punktów
pt_coords <- st_coordinates(punkty)
```

Funkcja do wyznaczania osiedli/dzielnic w których wyznaczono klastry

```
find_cluster_osiedla <- function(punkty, osiedla) {

  # przecięcie punktów klastrów z warstwą osiedli
  intersected_osiedla <- st_intersection(punkty, osiedla)

  # wyodrębnienie unikalnych nazw osiedli
  unique_osiedla <- intersected_osiedla %>%
    distinct(NAZWA_JEDN) %>%
```

```

pull(NAZWA_JEDN)

print("Osiedla, w których wyznaczono klastry:")
print(unique_osiedla)
}

```

ALGORYTM DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

To algorytm klasteryzacji oparty na gęstości, który identyfikuje klastry na podstawie zagęszczenia punktów w przestrzeni.

ZASADA DZIAŁANIA:

- >Obliczanie sąsiadów każdego punktu i identyfikowanie punktów rdzeniowych (punkty które w sąsiedztwie eps mają minPts punktów)
- >Tworzenie klastrów przez połączenie sąsiednich punktów punktu rdzeniowego (punkty w sąsiedztwie eps - punkty bezpośrednio osiągalne z rdzenia lub punkty graniczne)
- >Proces rozszerzania klastra - wykonanie dla każdego punktu nierdzeniowego:

- dodanie do sąsiedniego punktu rdzenia, jeśli to możliwe (spełnia kryterium)
- W przeciwnym razie dodanie do szumu (punkty odstające)

ZALETY:

- nie wymaga określenia liczby klastrów a priori

-może wykrywać klastry o dowolnych kształtach

- wyodrębnia szum i jest odporny na wartości odstające
- wymaga tylko dwóch parametrów i jest w większości przypadków niewrażliwy na kolejność punktów

WADY:

- wrażliwy na wybór parametrów eps i minPts (źle dobrane mogą prowadzić do błędów w klasteryzacji)
- nie klasteryzuje dobrze danych o dużych różnicach gęstości (nie da się wówczas odpowiednio dobrać kombinacji minPts- dla wszystkich klastrów)
- dla danych wielowymiarowych odległość euklidesowa, może być mało skuteczna (problem kłutwy wymiarowości)
- nie jest całkowicie deterministyczny (w przypadku punktów granicznych, które są osiągalne z więcej niż jednego klastra, ich przynależność zależy wówczas od kolejności przetwarzania danych)

Funkcja do generowania mapy DBSCAN

```

create_dbscan_map <- function(punkty,pt_coords, osiedla, eps, minPts) {
  # DBSCAN
  db <- dbscan(pt_coords, eps = eps , minPts = minPts)

  # dodanie wyników do zbioru danych (punkty)
  punkty$dbscan_cluster <- as.factor(db$cluster) # klastry jako kolumna w danych punktowych

  # usunięcie i wyodrębnienie szumu
  punkty_bez_szumu <- punkty %>% filter(dbscan_cluster != 0)
  szum <- punkty %>% filter(dbscan_cluster == 0)

  find_cluster_osiedla(punkty_bez_szumu,osiedla)

  # generowanie mapy
  ggplot() +
    geom_sf(data = osiedla, fill = "white", color = "black") +
    geom_sf(data = punkty_bez_szumu, aes(color = dbscan_cluster), size = 1.5) +
    labs(title = paste("DBSCAN: eps =", eps, ", minPts =", minPts, ", szum-półprzezroczysty szary"),
         color = "Cluster") +
    geom_sf(data = szum, col = "grey", size = 1.5, alpha = 0.3) +
    theme_minimal()
}

```

DBSCAN 1

```

eps1 <- 400 #promień otoczenia w którym sprawdzamy ilość sąsiadów
minPts1 <- 10 #minimalna ilość punktów które muszą tworzyć klaster
create_dbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, eps = eps1, minPts = minPts1)

```

```

## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries

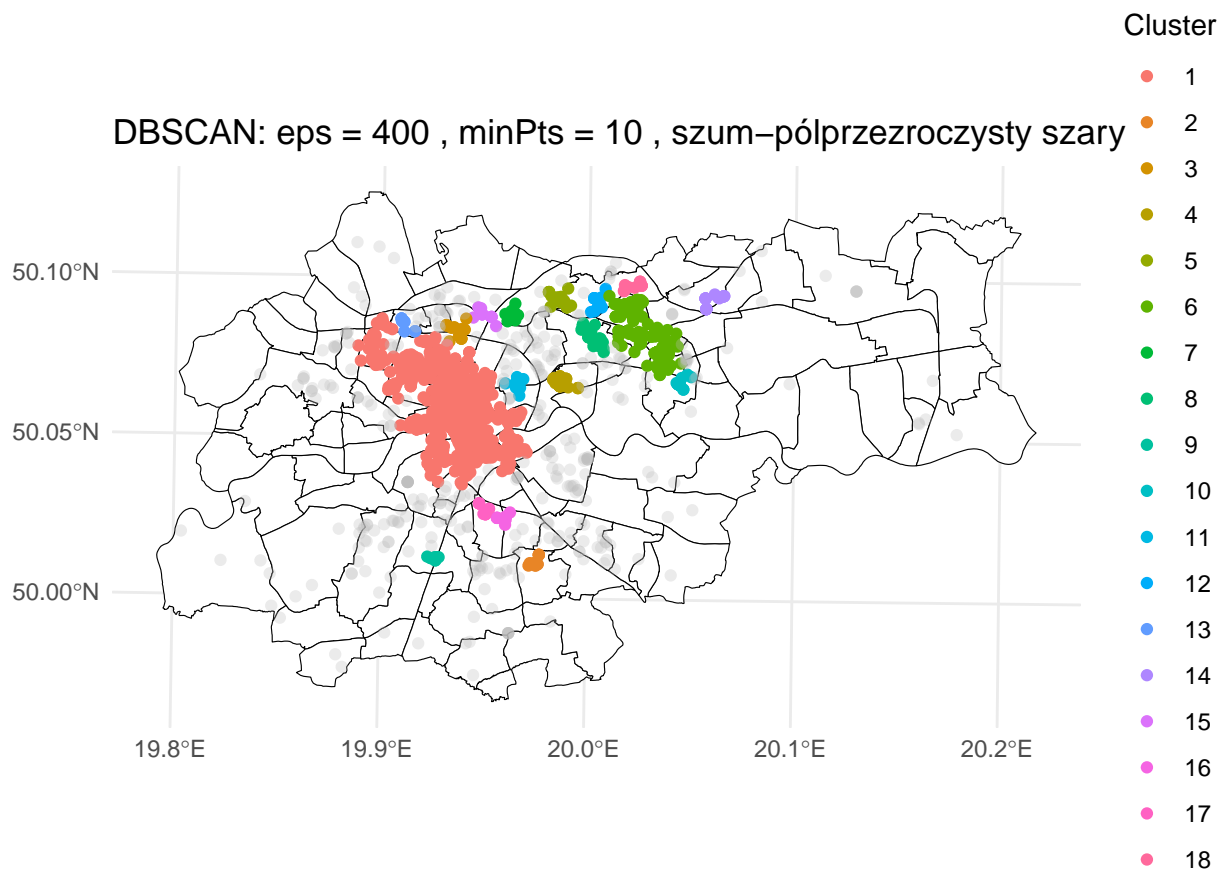
```

```

## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Piaski" "Wzg\xf3rza Krzes\xb3awickie"
## [3] "Krzes\xb3awice" "Bie\xflczyce Stare"
## [5] "Czy\xbfyny Lotnisko" "Mistrzejowice"
## [7] "Borek Fa\xb3\xeacki Wsch\xfd" "Borek Fa\xb3\xeacki Zach\xfd"
## [9] "Mateczny" "Wola Duchacka Zach\xfd"
## [11] "Wola Duchacka Wsch\xfd" "Zab\xb3ocie"
## [13] "Czy\xbfyny Park" "D\x9bie"
## [15] "Bie\xflczyce Nowe" "Pr\x9dnik Czerwony"
## [17] "Pr\x9dnik Bia\x3y Po\xbudnie" "Widok"
## [19] "Azory Zach\xfd" "Nowa Wie\x9c Po\xbudnie"
## [21] "Czarna Wie\x9c" "B\x3onia"
## [23] "Zakrz\x3wek" "Azory Wsch\xfd"
## [25] "Krowodrza P\x3noc" "Krowodrza Wsch\xfd"
## [27] "Warszawskie" "Olsza"
## [29] "Krowodrza Po\xbudnie" "Krowodrza - Nowa Wie\x9c"
## [31] "Osiedle Oficerskie" "Grzeg\xfrzki P\x3noc"
## [33] "Grzeg\xfrzki Wsch\xfd" "Kopiec Krakusa"

```

```
## [35] "Stare Podg\u00f3rze"          "Ludwin\u00f3w"
## [37] "D\u0142abniki"              "P\u00f3lsie Zwierzynieckie"
## [39] "Weso\u0142a Wsch\u00f3d"      "Grzeg\u00f3zki Zach\u00f3d"
## [41] "Kazimierz"                 "Stradom"
## [43] "Weso\u0142a Zach\u00f3d"        "Stare Miasto"
## [45] "Nowy \u015bwiat"             "Piasek P\u00f3\u0142udnie"
## [47] "Piasek P\u00f3\u0142noc"         "Kleparz"
## [49] "Zwierzyniec"              "Bronowice Ma\u0142e Wsch\u00f3d"
## [51] "Ma\u0142e B\u00f3nia"            "Osiedle Alberty\u0142skie"
## [53] "Centrum D - Handlowe"      "Sp\u00f3\u0142dzielcze - Kolorowe"
## [55] "Centrum C - Zgody"        "Krakowiak\u00f3w - G\u00f3rali"
## [57] "Sportowe - Zielone"       "Urocz\u0119"
## [59] "S\u0142oneczne"              "Centrum B - Szklane Domy"
## [61] "Stalowe - Wilowe - Wandy"  "Centrum A - Hutnicze Ogrodowe"
## [63] "Na Skarpie"
```



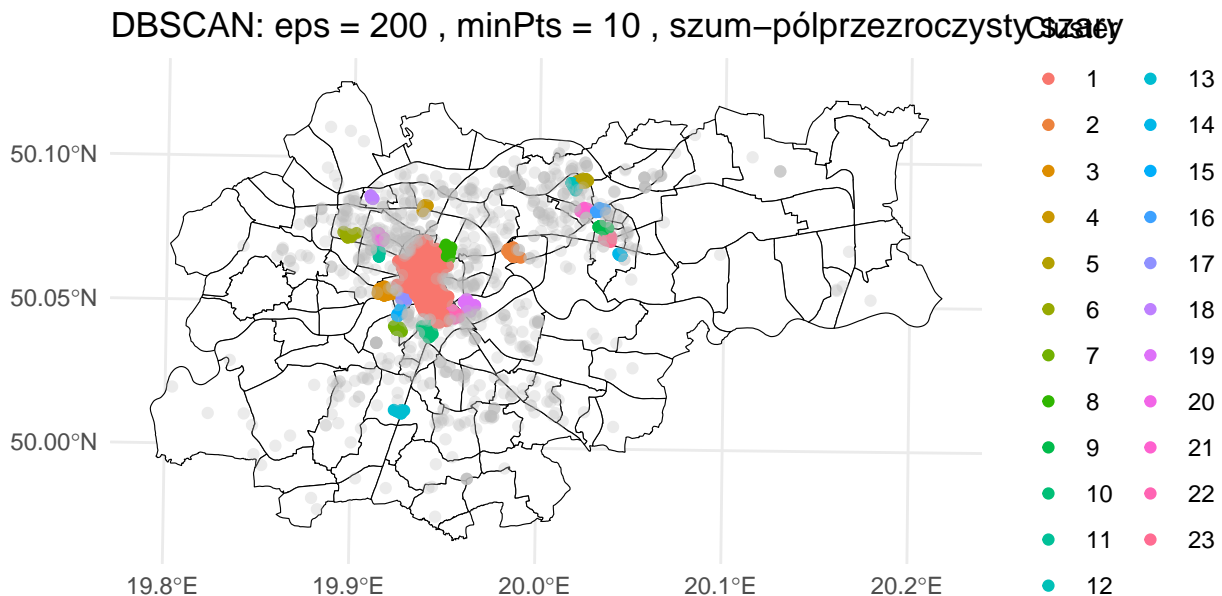
DBSCAN 2 - zmniejszone eps

```
eps2 <- 200
minPts2 <- 10
create_dbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, eps = eps2, minPts = minPts2)

## Warning: attribute variables are assumed to be spatially constant throughout
```

```
## all geometries
```

```
## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Bie\xf1czyce Stare" "Borek Fa\xb3\xeacki Wsch\xf3d"
## [3] "Borek Fa\xb3\xeacki Zach\xf3d" "Zab\xb3ocie"
## [5] "Czy\xbfyny Park" "D\xb9bie"
## [7] "Bie\xf1czyce Nowe" "Nowa Wie\x9c Po\xb3udnie"
## [9] "Czarna Wie\x9c" "B\xb3onia"
## [11] "Zakrz\xf3wek" "Azory Wsch\xf3d"
## [13] "Krowodrza Wsch\xf3d" "Warszawskie"
## [15] "Krowodrza Po\xb3udnie" "Krowodrza - Nowa Wie\x9c"
## [17] "Osiedle Oficerskie" "Stare Podg\xf3rze"
## [19] "Ludwin\xf3w" "D\xeabniki"
## [21] "P\xf3\xb3wsie Zwierzynieckie" "Weso\xb3a Wsch\xf3d"
## [23] "Kazimierz" "Stradom"
## [25] "Weso\xb3a Zach\xf3d" "Stare Miasto"
## [27] "Nowy \x8wiat" "Piasek Po\xb3udnie"
## [29] "Piasek P\xf3\xb3noc" "Kleparz"
## [31] "Zwierzyniec" "Bronowice Ma\xb3e Wsch\xf3d"
## [33] "Osiedle Alberty\xf1skie" "Centrum D - Handlowe"
## [35] "Centrum C - Zgody" "Krakowiak\xf3w - G\xf3rali"
## [37] "Centrum B - Szklane Domy" "Centrum A - Hutnicze Ogrodowe"
## [39] "Na Skarpie"
```



```
#zmniejszenie eps (zbyt mała wartość) spowodowało, że:
#-duża ilość punktów jest traktowana jako szum
#-klasteryzacja tworzy za dużo małych klastrów
```

DBSCAN 3 - zwiększone eps

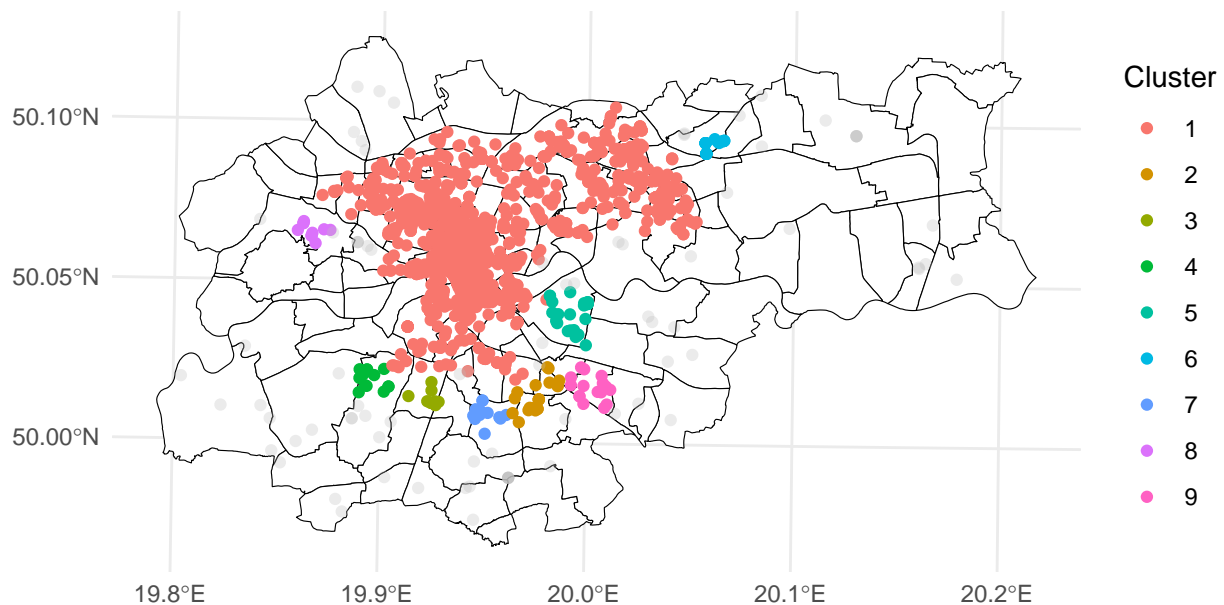
```
eps3 <- 700
minPts3 <- 10
create_dbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, eps = eps3, minPts = minPts3)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```
## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Batowice" "Zarzecze"
## [3] "R\xbf\xb9ka" "Piaski"
## [5] "Kurdwan\xf3w" "Wzg\xf3rza Krzes\xb3awickie"
## [7] "Krzes\xb3awice" "Nowy Prokocim"
## [9] "Stary Prokocim" "Bie\xf1czyce Stare"
## [11] "Czy\xbfyny Lotnisko" "Mistrzejowice"
## [13] "Czy\xbfyny \xa3\xeag" "Skotniki"
## [15] "Pychowice" "Kobierzyn"
## [17] "Borek Fa\xb3\xeacki Wsch\xf3d" "Borek Fa\xb3\xeacki Zach\xf3d"
## [19] "Mateczny" "Bonarka"
## [21] "Wola Duchacka Zach\xf3d" "Wola Duchacka Wsch\xf3d"
## [23] "Kabel" "Zab\xb3ocie"
## [25] "P\xb3asz\xf3w" "Czy\xbfyny Park"
## [27] "D\xb9bie" "Czy\xbfyny"
## [29] "Bie\xf1czyce Nowe" "Pr\xb9dnik Czerwony"
## [31] "Pr\xb9dnik Bia\xb3y Po\xb3udnie" "Bronowice Ma\xb3e Po\xb3udnie"
## [33] "Widok" "Azory Zach\xf3d"
## [35] "Nowa Wie\x9c Po\xb3udnie" "Czarna Wie\x9c"
## [37] "B\xb3onia" "Przegorza\xb3y Zwierzyniec"
## [39] "Zakrz\xf3wek" "Azory Wsch\xf3d"
## [41] "Krowodrza P\xf3\xb3noc" "Krowodrza Wsch\xf3d"
## [43] "Warszawskie" "Olsza"
## [45] "Rakowice" "Krowodrza Po\xb3udnie"
## [47] "Krowodrza - Nowa Wie\x9c" "Osiedle Oficerskie"
## [49] "Grzeg\xf3rzki P\xf3\xb3noc" "Grzeg\xf3rzki Wsch\xf3d"
## [51] "Kopiec Krakusa" "Stare Podg\xf3rze"
## [53] "Ludwin\xf3w" "D\xeabniki"
## [55] "P\xf3\xb3wsie Zwierzynieckie" "Weso\xb3a Wsch\xf3d"
## [57] "Grzeg\xf3rzki Zach\xf3d" "Kazimierz"
## [59] "Stradom" "Weso\xb3a Zach\xf3d"
## [61] "Stare Miasto" "Nowy \x8cwiat"
## [63] "Piasek Po\xb3udnie" "Piasek P\xf3\xb3noc"
## [65] "Kleparz" "Wola Justowska - Che\xb3m"
## [67] "Zwierzyniec" "Bronowice Ma\xb3e Wsch\xf3d"
## [69] "Cegielniana" "Ma\xb3e B\xb3onia"
## [71] "\xa3agiewniki" "Osiedle Alberty\xf1skie"
```

```
## [73] "Centrum D - Handlowe"      "Sp\xf3\xbdzielcze - Kolorowe"
## [75] "Centrum C - Zgody"        "Krakowiak\xfw - G\xfrali"
## [77] "Sportowe - Zielone"       "Urocze"
## [79] "S\xboneczne"              "Centrum B - Szklane Domy"
## [81] "Szkolne"                  "Stalowe - Willowe - Wandy"
## [83] "Centrum A - Hutnicze Ogrodowe" "Beszcz"
## [85] "Koz\xb3\xfwek"            "Na Skarpie"
```

DBSCAN: eps = 700 , minPts = 10 , szum-półprzezroczysty szary



```
# zwiększenie eps (zbyt duża wartość) spowodowało, że:
#-większość danych zostaje zgrupowana w jeden duży klaster
#-mała ilość punktów zakwalifikowanych jako szum
#-klasteryzacja przestaje być użyteczna, ponieważ większość punktów są traktowane jako jeden zbiór
```

DBSCAN 4 - zmniejszone minPts

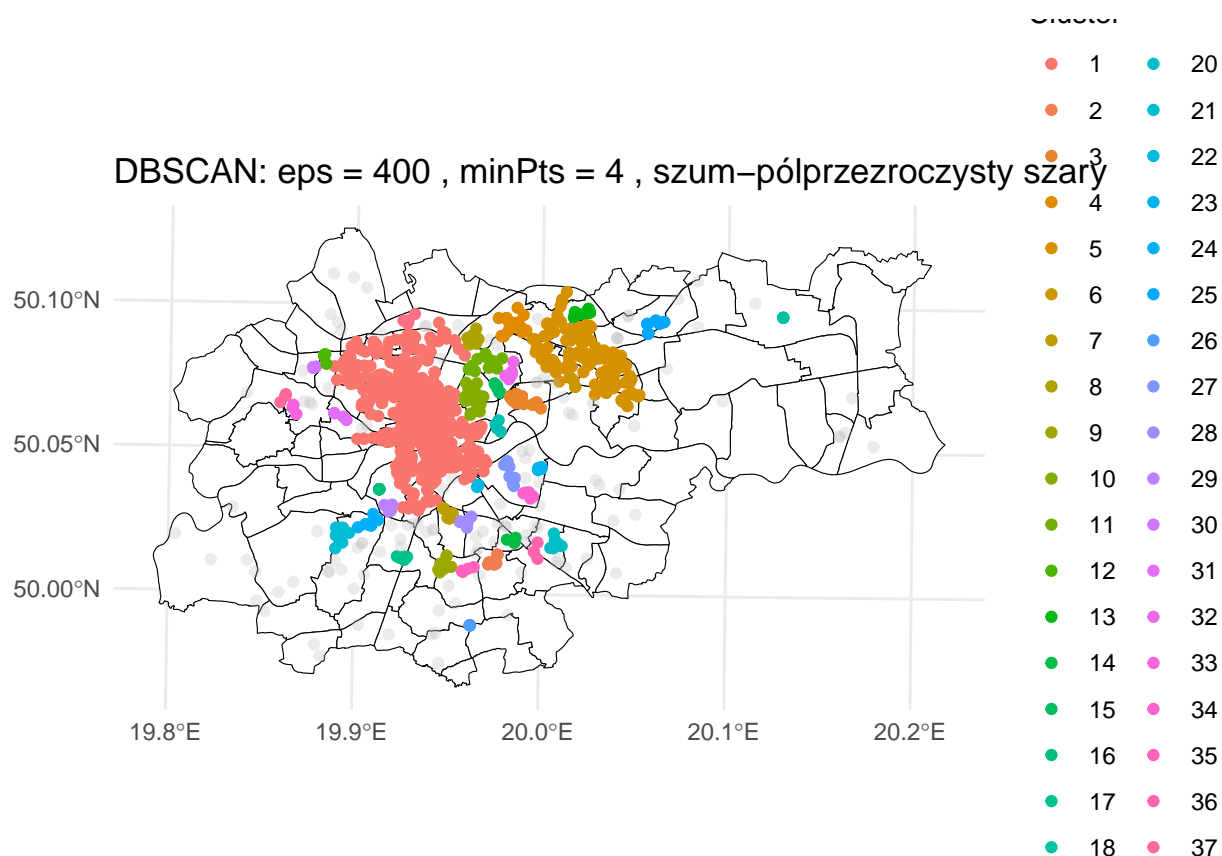
```
eps4 <- 400
minPts4 <- 4
create_dbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, eps = eps4, minPts = minPts4)

## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```

## [1] "Osiedla, w których wyznaczono klasty:"
## [1] "Batowice" "\xa3uczanowice - Wad\xf3w"
## [3] "R\xbf\xb9ka" "Piaski"
## [5] "Kurdwan\xf3w" "Wzg\xf3rza Krzes\xb3awickie"
## [7] "Krzes\xb3awice" "Nowy Prokocim"
## [9] "Bie\xf1czyce Stare" "Czy\xbfny Lotnisko"
## [11] "Mistrzejowice" "Czy\xbfny \xa3\xeag"
## [13] "Skotniki" "Pychowice"
## [15] "Kobierzyn" "Borek Fa\xb3\xeacki Wsch\xf3d"
## [17] "Borek Fa\xb3\xeacki Zach\xf3d" "Mateczny"
## [19] "Bonarka" "Wola Duchacka Zach\xf3d"
## [21] "Wola Duchacka Wsch\xf3d" "Kabel"
## [23] "Zab\xb3ocie" "P\xb3asz\xf3w"
## [25] "Czy\xbfny Park" "D\xb9bie"
## [27] "Czy\xbfny" "Bie\xf1czyce Nowe"
## [29] "Pr\xb9dnik Czerwony" "Pr\xb9dnik Bia\xb3y Po\xb3udnie"
## [31] "Bronowice Ma\xb3e Po\xb3udnie" "Widok"
## [33] "Azory Zach\xf3d" "Nowa Wie\x9c Po\xb3udnie"
## [35] "Czarna Wie\x9c" "B\xb3onia"
## [37] "Przegorza\xb3y Zwierzyniec" "Zakrz\xf3wek"
## [39] "Azory Wsch\xf3d" "Krowodrza P\xf3\xb3noc"
## [41] "Krowodrza Wsch\xf3d" "Warszawskie"
## [43] "Olsza" "Rakowice"
## [45] "Krowodrza Po\xb3udnie" "Krowodrza - Nowa Wie\x9c"
## [47] "Osiedle Oficerskie" "Grzeg\xf3rzki P\xf3\xb3noc"
## [49] "Grzeg\xf3rzki Wsch\xf3d" "Kopiec Krakusa"
## [51] "Stare Podg\xf3rze" "Ludwin\xf3w"
## [53] "D\xeabniki" "P\xf3\xb3wsie Zwierzynieckie"
## [55] "Weso\xb3a Wsch\xf3d" "Grzeg\xf3rzki Zach\xf3d"
## [57] "Kazimierz" "Stradom"
## [59] "Weso\xb3a Zach\xf3d" "Stare Miasto"
## [61] "Nowy \x8cwiat" "Piasek Po\xb3udnie"
## [63] "Piasek P\xf3\xb3noc" "Kleparz"
## [65] "Swoszowice" "Wola Justowska - Che\xb3m"
## [67] "Zwierzyniec" "Bronowice Ma\xb3e Wsch\xf3d"
## [69] "Cegielniana" "Ma\xb3e B\xb3onia"
## [71] "Osiedle Alberty\xf1skie" "Centrum D - Handlowe"
## [73] "Sp\xf3\xb3dzielcze - Kolorowe" "Centrum C - Zgody"
## [75] "Krakowiak\xf3w - G\xf3rali" "Sportowe - Zielone"
## [77] "Urocze" "S\xb3oneczne"
## [79] "Centrum B - Szklane Domy" "Szkolne"
## [81] "Stalowe - Willowe - Wandy" "Centrum A - Hutnicze Ogrodowe"
## [83] "Beszcz" "Koz\xb3\xf3wek"
## [85] "Na Skarpie"

```

zmniejszenie minPts (zbyt mała wartość) spowodowało, że:
 #-tworzy się wiele małych klastrów, z których część może być nieistotna i może być to szum
 #-punktów szumu może być zbyt mało

DBSCAN 5 - zwiększone minPts

```
eps5 <- 400
minPts5 <- 16
create_dbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, eps = eps5, minPts = minPts5)
```

Warning: attribute variables are assumed to be spatially constant throughout
 ## all geometries

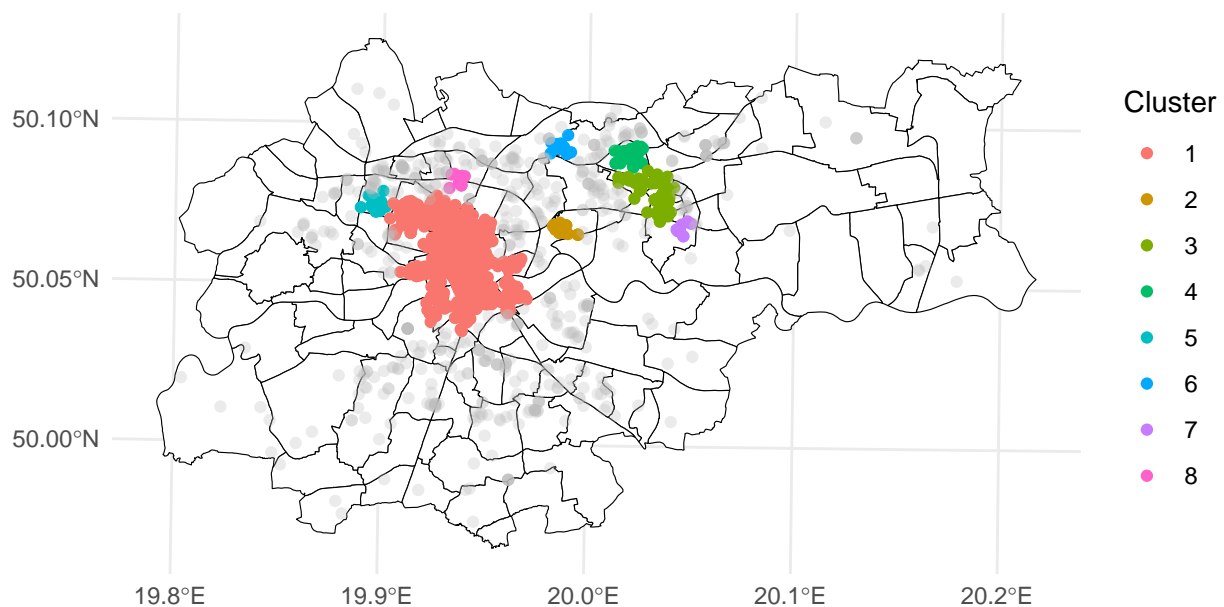
```
## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Bie\xf1czyce Stare" "Czy\xbfny Lotnisko"
## [3] "Mistrzejowice" "Mateczny"
## [5] "Zab\xb3ocie" "Czy\xbfny Park"
## [7] "D\xb9bie" "Bie\xf1czyce Nowe"
## [9] "Pr\xb9dnik Czerwony" "Nowa Wie\x9c Po\xb3udnie"
## [11] "Czarna Wie\x9c" "B\xb3onia"
## [13] "Zakrz\xf3wek" "Krowodrza P\xf3\xb3noc"
## [15] "Krowodrza Wsch\xf3d" "Warszawskie"
```

```

## [17] "Krowodrza Po\u03b3udnie"      "Krowodrza - Nowa Wie\u03c1c"
## [19] "Osiedle Oficerskie"          "Grzeg\u0333rzki Wsch\u0333d"
## [21] "Kopiec Krakusa"              "Stare Podg\u0333rze"
## [23] "Ludwin\u0333w"                "D\u0333abniki"
## [25] "P\u0333wsie Zwierzynieckie"   "Weso\u0333a Wsch\u0333d"
## [27] "Grzeg\u0333rzki Zach\u0333d"    "Kazimierz"
## [29] "Stradom"                     "Weso\u0333a Zach\u0333d"
## [31] "Stare Miasto"                "Nowy \u0333wiat"
## [33] "Piasek Po\u03b3udnie"           "Piasek P\u0333noc"
## [35] "Kleparz"                     "Zwierzyniec"
## [37] "Bronowice Ma\u0333e Wsch\u0333d"   "Ma\u0333e B\u0333onia"
## [39] "Osiedle Alberty\u0333skie"      "Centrum D - Handlowe"
## [41] "Sp\u0333dzielcze - Kolorowe"    "Centrum C - Zgody"
## [43] "Krakowiak\u0333w - G\u0333rali"   "Sportowe - Zielone"
## [45] "Urocze"                      "S\u0333oneczne"
## [47] "Centrum B - Szklane Domy"     "Stalowe - Willowe - Wandy"
## [49] "Centrum A - Hutnicze Ogrodowe" "Na Skarpie"

```

DBSCAN: eps = 400 , minPts = 16 , szum-p\u0333przezroczysty szary



zwi\u0333kszenie minPts (zbyt du\u0333a warto\u0333\u015b\u0107) spowodowa\u0333o, \u0333e:
#-wiele punkt\u0333w nie spe\u0333nia wymogu minimalnej liczby s\u0333siedz\u0333w, przez co powsta\u0333 zbyt du\u0333y szum
#-liczba klastr\u0333w jest zbyt ma\u0333a

ALGORYTM HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)

To algorytm będący rozwinięciem algorytmu DBSCAN

ZASADA DZIAŁANIA:

aspekt hierarchiczny -> zamiast jednej wartości eps, HDBSCAN tworzy hierarchię klastrow, zaczynając od małych gęstych obszarów i stopniowo je łącząc w większe

->Obliczanie odległości wzajemnej osiągalności między punktami (na podstawie odległości i odległości rdzenia)

->Na podstawie obliczonej odległości wzajemnej osiągalności, konstruowane jest drzewo rozpinające sąsiedztwa między punktami

->Wykorzystanie hierarchii klastrow do identyfikacji stabilnych klastrow o zmiennej gęstości, eliminacja słabo stabilnych klastrow

->Punkty należące do niestabilnych klastrow mogą być:

- przypisane do szumu, jeśli nie spełniają minimalnych kryteriów klasteryzacji
- przypisane do innych klastrow, jeśli są osiągalne z punktów rdzeniowych w stabilnym klastrze

ZALETY:

- nie jest konieczne wartości parametru eps (dostosowuje się do zmiennej gęstości danych)
- lepiej niż DBSCAN klasteryzuje dane o dużych różnicach gęstości
- wyodrębnia szum
- wykrywa bardziej stabilne i wiarygodne klastry dzięki hierarchicznemu podejściu (usuając niestabilne klastry minimalizuje ryzyko nadmiernego dopasowania)

WADY:

- jest bardziej złożony obliczeniowo niż DBSCAN
- może być bardziej wrażliwy na dane szumowe niż DBSCAN, jeśli parametry są źle dobrane
- niewłaściwy dobór parametru minPts może prowadzić do błędnych wyników klasteryzacji

Funkcja do generowania mapy HDBSCAN

```
create_hdbscan_map <- function(punkty, pt_coords, osiedla, minPts) {  
  # HDBSCAN  
  hdb <- hdbscan(pt_coords, minPts = minPts)  
  
  # dodanie wyników do zbioru danych (punkty)  
  punkty$hdbscan_cluster <- as.factor(hdb$cluster) # klastry jako kolumna w danych punktowych  
  
  # usunięcie i wyodrębnienie szumu
```

```

punkty_bez_szumu <- punkty %>% filter(hdbscan_cluster != 0)
szum <- punkty %>% filter(hdbscan_cluster == 0)

find_cluster_osiedla(punkty_bez_szumu,osiedla)

# generowanie mapy
ggplot() +
  geom_sf(data = osiedla, fill = "white", color = "black") +
  geom_sf(data = punkty_bez_szumu, aes(color = hdbscan_cluster), size = 1.5) +
  labs(title = paste("HDBSCAN: minPts =", minPts, ", szum-półprzezroczysty szary"),
        color = "Cluster") +
  geom_sf(data = szum, col = "grey", size = 1.5, alpha = 0.1) +
  theme_minimal()
}

```

HDBSCAN 1

```

minPts_h1 <- 30
create_hdbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, minPts = minPts_h1)

```

```

## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries

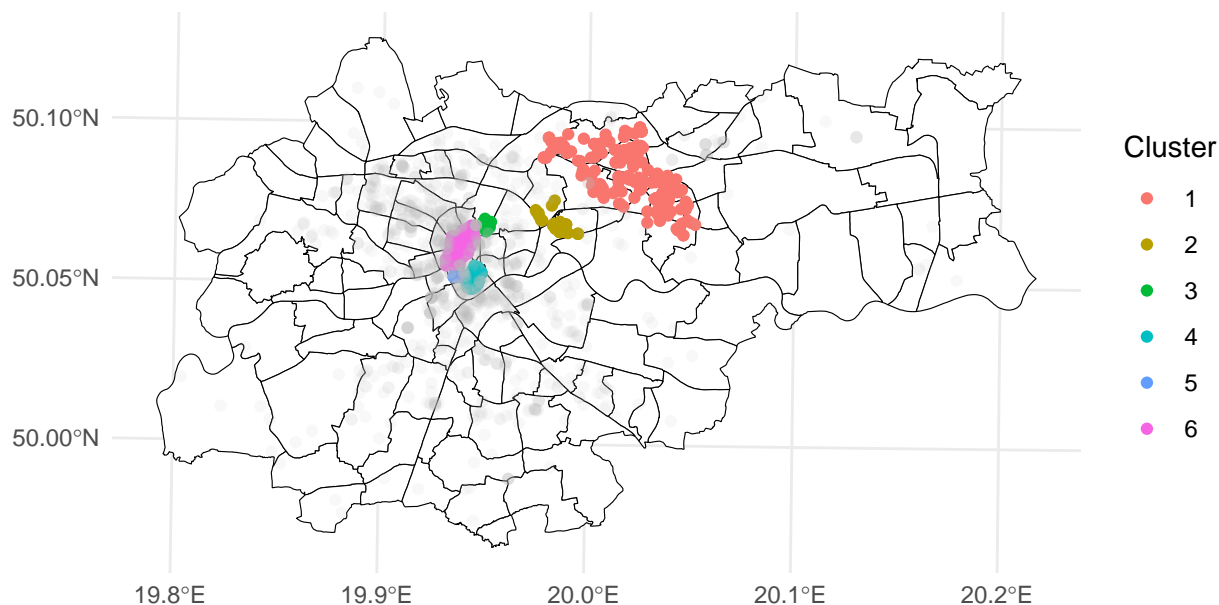
```

```

## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Krzes\&#3awice" "Bie\&#1czyce Stare"
## [3] "Czy\&#bfyn Lotnisko" "Mistrzejowice"
## [5] "Czy\&#bfyn Park" "D\&#9bie"
## [7] "Bie\&#1czyce Nowe" "Pr\&#9dnik Czerwony"
## [9] "Warszawskie" "Rakowice"
## [11] "Osiedle Oficerskie" "Kazimierz"
## [13] "Stradom" "Weso\&#3a Zach\&#3d"
## [15] "Stare Miasto" "Nowy \&#8wiat"
## [17] "Piasek P\&#3\&#3noc" "Kleparz"
## [19] "Osiedle Alberty\&#fiskie" "Centrum D - Handlowe"
## [21] "Sp\&#3\&#3dzielcze - Kolorowe" "Centrum C - Zgody"
## [23] "Krakowiak\&#3w - G\&#3rali" "Sportowe - Zielone"
## [25] "Uroczne" "S\&#3oneczne"
## [27] "Centrum B - Szklane Domy" "Szkolne"
## [29] "Stalowe - Willowe - Wandy" "Centrum A - Hutnicze Ogrodowe"
## [31] "Na Skarpie"

```

HDBSCAN: minPts = 30 , szum–półprzezroczysty szary



przy tak dobranej wartości minPts do klastrów zostały zakwalifikowane punkty znajdujące się w dużych s

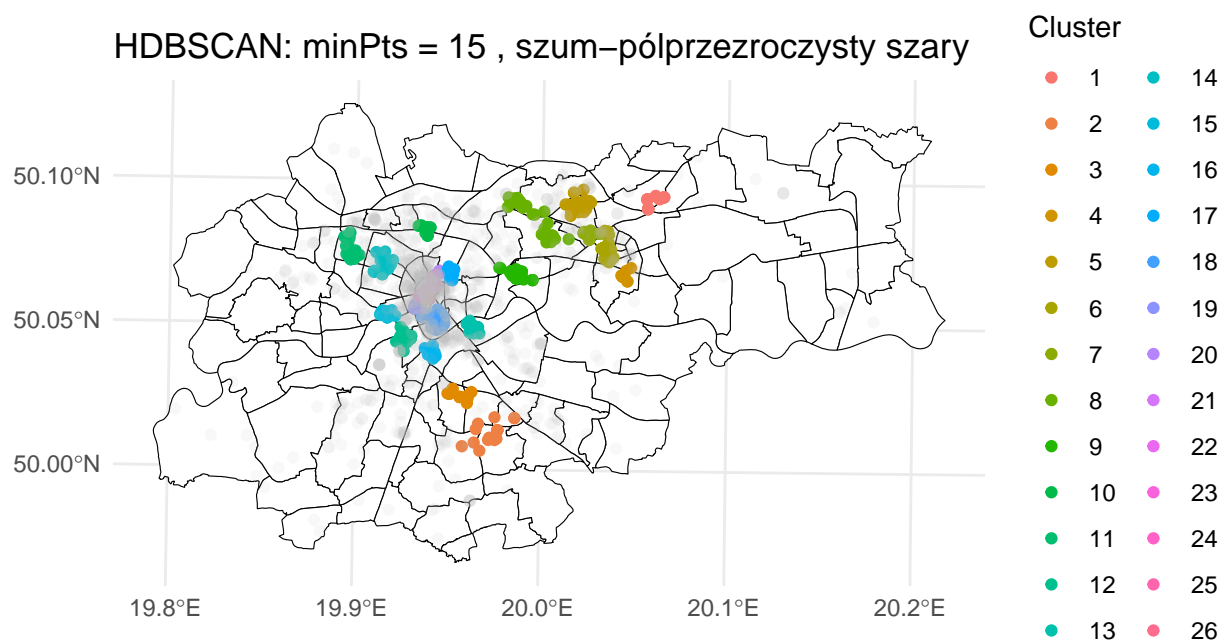
HDBSCAN 2

```
minPts_h2 <-15
create_hdbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, minPts = minPts_h2)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```
## [1] "Osiedla, w których wyznaczono klastry:"
## [1] "Piaski" "Kurdwan\xf3w"
## [3] "Wzg\xfrza Krzes\x3awickie" "Bie\xflczyce Stare"
## [5] "Czy\xbfyny Lotnisko" "Mistrzejowice"
## [7] "Wola Duchacka Zach\xfd" "Wola Duchacka Wsch\xfd"
## [9] "Zab\x3ocie" "Czy\xbfyny Park"
## [11] "D\x9bie" "Bie\xflczyce Nowe"
## [13] "Pr\x9dnik Czerwony" "Nowa Wie\x9c Po\x3udnie"
## [15] "Czarna Wie\x9c" "B\x3onia"
## [17] "Zakrz\x3wek" "Krowodrza P\xfd\x3noc"
## [19] "Krowodrza Wsch\xfd" "Warszawskie"
## [21] "Rakowice" "Krowodrza Po\x3udnie"
```

```
## [23] "Krowodrza - Nowa Wie\x9c"      "Osiedle Oficerskie"
## [25] "Stare Podg\xfc3rze"           "Ludwin\xfc3w"
## [27] "D\xecabniki"                  "P\xfc3\xb3wsie Zwierzynieckie"
## [29] "Kazimierz"                    "Stradom"
## [31] "Stare Miasto"                 "Nowy \x8cwiat"
## [33] "Kleparz"                      "Zwierzyniec"
## [35] "Bronowice Ma\xb3e Wsch\xfc3d"  "Osiedle Alberty\xfc1skie"
## [37] "Centrum D - Handlowe"         "Centrum C - Zgody"
## [39] "Krakowiak\xfc3w - G\xfc3rali"  "Urocze"
## [41] "Centrum B - Szklane Domy"     "Stalowe - Willowe - Wandy"
## [43] "Koz\xb3\xfc3wek"              "Na Skarpie"
```



przy tak dobranej wartości minPts zwiększyła się liczba klastrów (niektóre klastry zostały utworzone

HDBSCAN 3

```
minPts_h3 <- 6
create_hdbscan_map(punkty=punkty, pt_coords = pt_coords, osiedla = osiedla, minPts = minPts_h3)
```

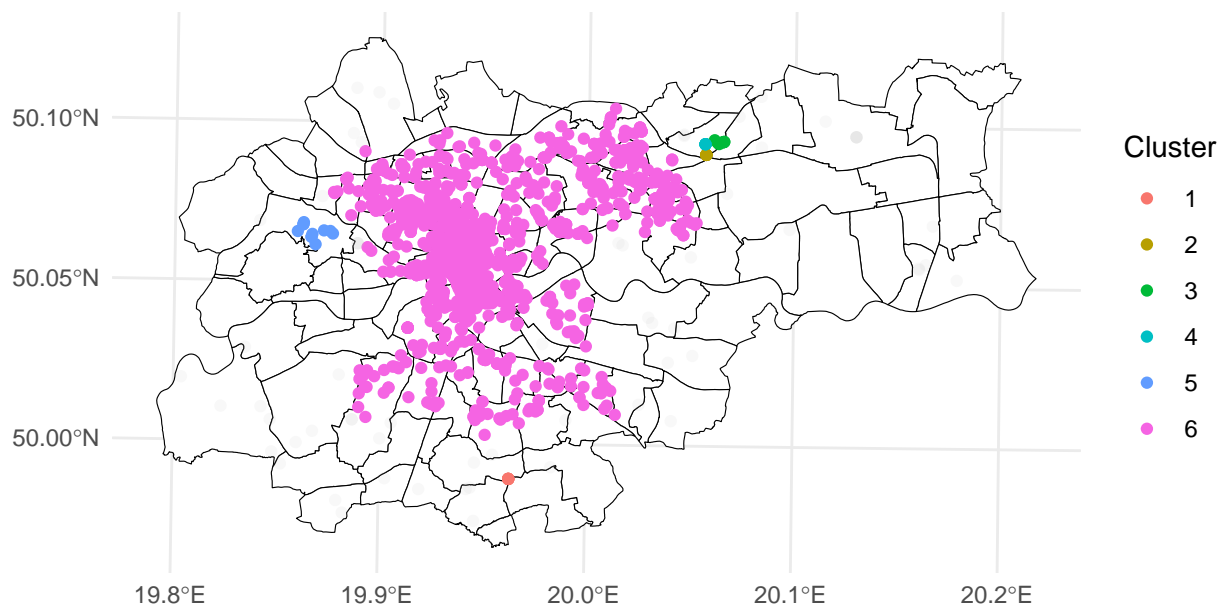
```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```

## [1] "Osiedla, w których wyznaczono klasty:"
## [1] "Bronowice Wielkie" "Batowice"
## [3] "Zarzecz" "Nowy Bie\xbf\fan\xf3w"
## [5] "R\xbf\9ka" "Piaski"
## [7] "Kurdwan\xf3w" "Wzg\xf3rza Krzes\3awickie"
## [9] "Krzes\3awice" "Nowy Prokocim"
## [11] "Stary Prokocim" "Bie\xf1czyce Stare"
## [13] "Czy\bfyny Lotnisko" "Mistrzejowice"
## [15] "Czy\bfyny \xa3\3eag" "Skotniki"
## [17] "Pychowice" "Kobierzyn"
## [19] "Borek Fa\3\3eacki Wsch\3d" "Borek Fa\3\3eacki Zach\3d"
## [21] "Mateczny" "Bonarka"
## [23] "Wola Duchacka Zach\3d" "Wola Duchacka Wsch\3d"
## [25] "Kabel" "Zab\3ocie"
## [27] "P\3asz\3w" "Czy\bfyny Park"
## [29] "D\9bie" "Czy\bfyny"
## [31] "Bie\xf1czyce Nowe" "Pr\9dnik Czerwony"
## [33] "Pr\9dnik Bia\3y Po\3udnie" "Bronowice Ma\3e Po\3udnie"
## [35] "Widok" "Azory Zach\3d"
## [37] "Nowa Wie\9c Po\3udnie" "Czarna Wie\9c"
## [39] "B\3onia" "Przegorza\3y Zwierzyniec"
## [41] "Zakrz\3wek" "Azory Wsch\3d"
## [43] "Krowodrza P\3\3noc" "Krowodrza Wsch\3d"
## [45] "Warszawskie" "Olsza"
## [47] "Rakowice" "Krowodrza Po\3udnie"
## [49] "Krowodrza - Nowa Wie\9c" "Osiedle Oficerskie"
## [51] "Grzeg\3rzki P\3\3noc" "Grzeg\3rzki Wsch\3d"
## [53] "Kopiec Krakusa" "Stare Podg\3rze"
## [55] "Ludwin\3w" "D\3eabniki"
## [57] "P\3\3wsie Zwierzynieckie" "Weso\3a Wsch\3d"
## [59] "Grzeg\3rzki Zach\3d" "Kazimierz"
## [61] "Stradom" "Weso\3a Zach\3d"
## [63] "Stare Miasto" "Nowy \8cwiat"
## [65] "Piasek Po\3udnie" "Piasek P\3\3noc"
## [67] "Kleparz" "Swoszowice"
## [69] "Wola Justowska - Che\3m" "Zwierzyniec"
## [71] "Bronowice Ma\3e Wsch\3d" "Cegielniana"
## [73] "Ma\3e B\3onia" "\xa3agiewniki"
## [75] "Osiedle Alberty\3fskie" "Centrum D - Handlowe"
## [77] "Sp\3\3dzielcze - Kolorowe" "Centrum C - Zgody"
## [79] "Krakowiak\3w - G\3rali" "Sportowe - Zielone"
## [81] "Urocze" "S\3oneczne"
## [83] "Centrum B - Szklane Domy" "Szkolne"
## [85] "Stalowe - Willowe - Wandy" "Centrum A - Hutnicze Ogrodowe"
## [87] "Beszcz" "Koz\3\3wek"
## [89] "Na Skarpie"

```

HDBSCAN: minPts = 6 , szum–półprzezroczysty szary



źle dobrane wartości minPts spowodowały, że większość danych zostaje zgrupowana w jeden duży klaster or

WNIOSKI

Jeśli skupiamy się na analizie obszarów o różnej gęstości (np. centrum miasta i obrzeża), HDBSCAN wydaje się być bardziej odpowiednim algorytmem dzięki zdolności adaptacji do lokalnej gęstości (natomiast HDBSCAN jest trudniejszy w diagnozowaniu problemów z parametrami). W przypadku DBSCAN należy eksperymentalnie dostosować eps i minPts w taki sposób, aby uwzględniały różnice w rozkładzie punktów między centrum a obrzeżami.

Na podstawie otrzymanych wyników, można stwierdzić, że najwięcej zarejestrowanych wykroczeń występuje w rejonie Starego Miasta i jego okolicach (dzielnice Krowdrza i Grzegórzki), a następnie okolice Mistrzejowic i Bieńczy.