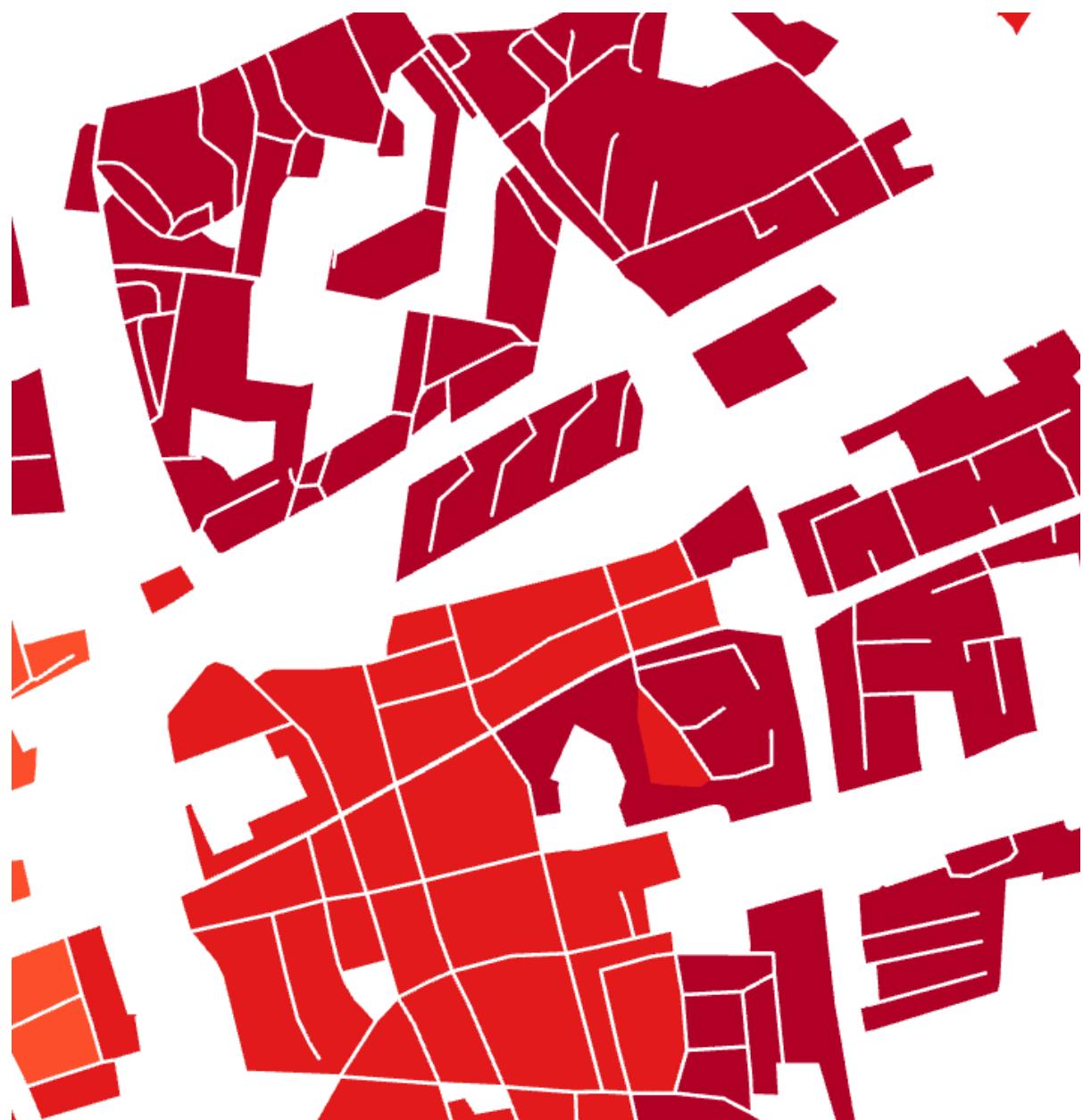


STO stron GIS



ArcMap, ArcGIS Pro, QGIS 3 – 150 zadań

Jacek Urbański

Gdynia 2022

Spis treści.

<u>Wstęp</u>	2		
1. <u>Współrzędne geograficzne. Lokalizacja</u>	3		
2. <u>Dane przestrzenne</u>	11		
3. <u>Współrzędne prostokątne. Odwzorowania</u>	18		
4. <u>Inżynieria danych przestrzennych</u>	25		
4.1 <u>Tworzenie i pozyskiwanie danych przestrzennych</u>	26		
4.2 <u>Eksploracyjna analiza danych przestrzennych</u>	35		
4.3 <u>Podstawowe operacje przygotowania danych</u>	42		
4.4 <u>Obliczenia w tabeli atrybutowej (Python)</u>	50		
4.5 <u>Tworzenie i edycja geometrii danych wektorowych</u>	60		
4.6 <u>Georeferencja i rejestracja rastra</u>	64		
4.7 <u>Zmiana wartości w rastrze (Algebra Map, Raster Calculator)</u>	69		
5. <u>Pozyskiwanie informacji przestrzennej (analiza danych)</u>	73		
5.1 <u>Pozyskiwanie atrybutów na podstawie relacji przestrzennych</u>	74		
5.2 <u>Analiza odległości pomiędzy obiektami</u>	78		
5.3 <u>Bufory i inne otoczenia obiektów</u>	80		
5.4 <u>Operacje na geometrii danych</u>	83		
5.5 <u>Podstawowe funkcje i metody analizy rastrowej</u>	93		
5.6 <u>Metody interpolacji danych punktowych</u>	110		
<u>Zadania z analizy wektorowej</u>	Narzędzia analizy wektorowej	<u>ArcGIS</u>	<u>QGIS</u>
<u>Zadania z analizy rastrowej</u>	Narzędzia analizy rastrowej	<u>ArcGIS</u>	<u>QGIS</u>
<u>Dane do zadań i inne</u>	<u>Definicje szerokości i długości geograficznej</u>		

Wstęp.

Geoprocessing (geoprzetwarzanie) jest dowolną operacją w GIS przeprowadzaną na danych. Zwykle operacja ta polega na zastosowaniu narzędzia GIS do którego wprowadzamy dane i w rezultacie działania tego narzędzia otrzymujemy dane wynikowe. Te dane wynikowe mogą stać się danymi wprowadzanymi do kolejnego narzędzia. Tworzymy w ten sposób logiczny ciąg przetwarzania danych, który realizuje rozwiązanie pewnego problemu o charakterze przestrzennym.

Cały proces można podzielić na cztery etapy. Etap pierwszy polega na określeniu celu naszego projektu. Zaczynamy od ogólnego zdefiniowania problemu, który chcemy rozwiązać. Następnie należy rozbić ten problem na ciąg zadań kończących się mierzalnym rezultatem. Dla każdego zadania określamy potrzebne dane.

Etap drugi polega na kompleksowym pozyskaniu i przetworzeniu danych do potrzebnej postaci. Ten etap nazywa się obecnie inżynierią danych, ewentualnie z dodaniem słowa ‘przestrzennych’. Dane muszą być wytworzony lub pozyskane. Powinny być przetworzone tak, aby pasowały do analizowanej przestrzeni i czasu. Zwykle organizuje się dane w bazach danych, które ze względu na przestrzenną cechę danych nazywa się geobazami. Dane muszą być przekształcane i dostosowane do wymogów projektu. Powinny być także sprawdzone i wstępnie zanalizowane.

Kolejny trzeci etap to analiza danych. Określa się metody analizy przestrzennej i ich sekwencję, aby zrealizować kolejne zadania. Przeprowadzamy geoprzetwarzanie danych, a następnie wykonujemy ewaluację i interpretację rezultatów. W zależności od potrzeb na tym etapie mogą być tworzone rozwiązania alternatywne.

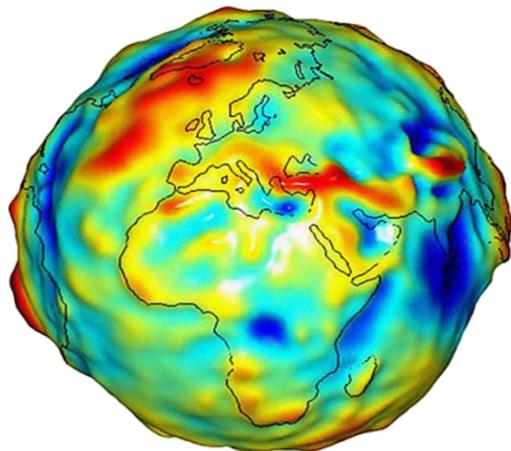
Ostatnim etapem jest tworzenie rezultatów analizy, w postaci dostosowanej do zadana i odbiorcy. Często rezultaty analiz w GIS mają charakter map, lub złożonych projektów obejmujących mapy, rysunki i tekst. Wszystkie etapy powinny być dokumentowane, aby cały proces mógł być powtórzony w przyszłości, a każdy kolejny krok szczegółowo opisany.

Rozdział 1

Współrzędne geograficzne i lokalizacja

Unikalną cechą **danych przestrzennych** jest ich **lokalizacja** czyli opis miejsca w którym zachodzi zjawisko, które jest opisane przez te dane. Oznacza to, że każda obserwacja zachodzącego na powierzchni Ziemi zjawiska będzie musiała zawierać informację o jego położeniu, jeśli to położenie ma dla nas znaczenie.

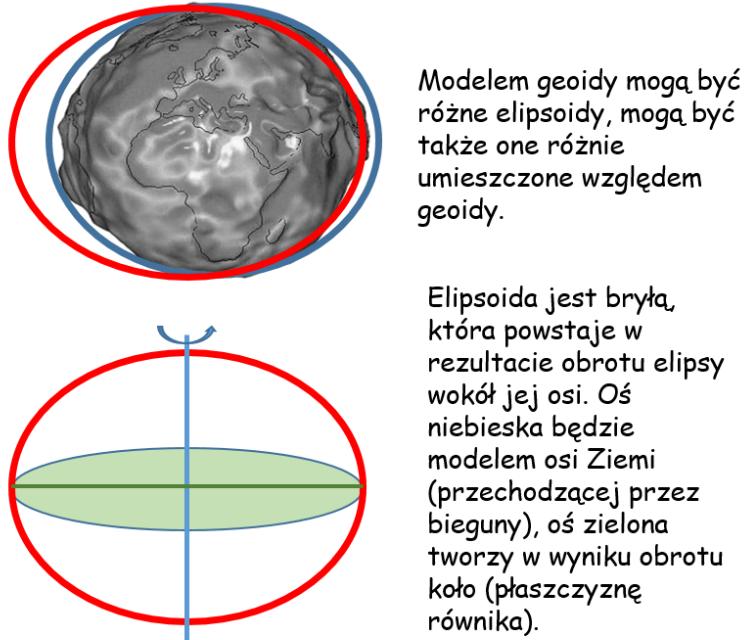
Ziemia ma kształt **geoidy**, bryły o nieregularnej powierzchni przypominającej lekko spłaszoną od dołu i góry kulę. Aby wyobrazić ją sobie należaałoby przyjąć, że cała powierzchnia Ziemi została zalana przez ocean. Poziom wody w każdym miejscu tego oceanu ulegałby oczywiście ciągłym wahaniom. Jednak gdybyśmy „zamrozili” powierzchnię wody w jej średnim położeniu w każdym miejscu kuli ziemskiej to otrzymalibyśmy właśnie geoidę, której powierzchnia pokryta byłaby wznieśniami i obniżniami o różnicach wysokości dochodzących do około stu metrów. Przyczyną tych nierówności powierzchni geoidy są różnice w gęstości skał tworzących wnętrze Ziemi. Spłaszczenia w okolicach biegunów są natomiast rezultatem działania siły odśrodkowej będącej rezultatem ruch wirowego Ziemi na około własnej osi, która to siła jest największa na równiku a najmniejsza na biegunach.



Rysunek 1.1 Kształt Ziemi. Nieregularna kulista bryła zwana geoidą spłaszcza na biegunach

Do określania położenia punktu w przestrzeni używamy **układów współrzędnych**. Na powierzchni nieregularnej geoidy nie jesteśmy w stanie

określić takiego układu, potrzebujemy do tego regularnej bryły będącej dobrym modelem geoidy. Taką bryłą jest **elipsoida**, która powstaje przez obrót elipsy dookoła jej osi.



Rysunek 1.2 Elipsoida jako model geoidy

Do tworzenia modelu geoidy używa się różnych elipsoid, które są jednoznacznie zdefiniowane przez wielkość swoich pól, a więc długość półosi wielkiej i spłaszczenie obliczane jako $\frac{(półos wielka - półos mała)}{półos wielka}$. Praktycznie istotne są tylko dwie z nich:

Tabela 1.1 Podstawowe elipsoidy – modele Ziemi.

Nazwa elipsoidy	Półos wielka	Spłaszczenie
WGS 1984 (WGS84)	6378137 m	1/298.257223563
GRS 80	6378137 m	1/298.2572

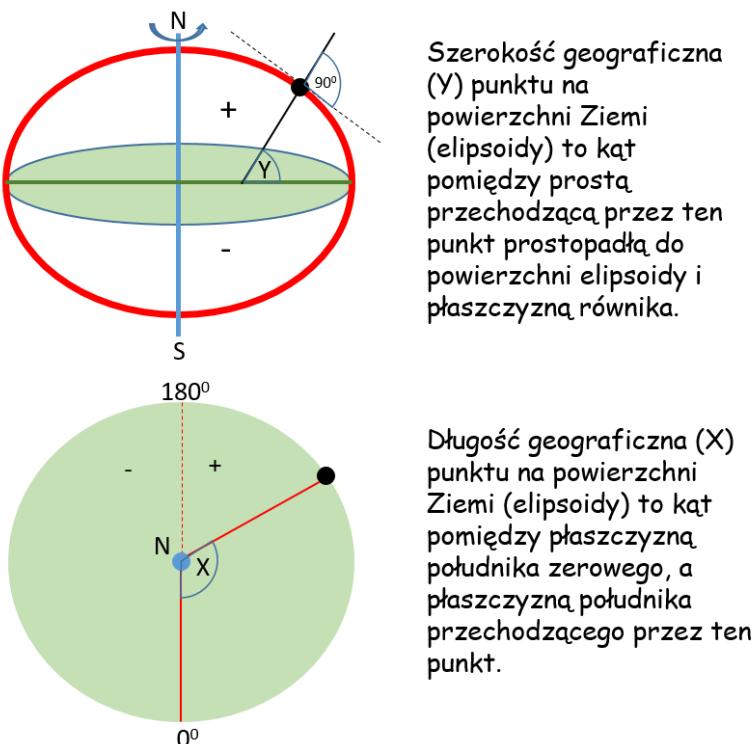
Są one prawie identyczne co sprawia, że często nie będzie istotne, która z nich jest aktualnie wykorzystywana. Oprócz wyboru elipsoidy istotne jest jej umieszczenie względem geoidy. Te dwa elementy rodzą elipsoidy i sposób jej umieszczenia względem geoidy definiują **poziomy układ odniesienia** zwany po angielsku **datum**. Podobnie jak w przypadku elipsoid praktycznie będziemy używali dwóch poziomych układów odniesienia.

Tabela 1.2 Podstawowe układy odniesienia (datum).

Nazwa układu odniesienia	Nazwa elipsoidy	Uwagi
WGS84	WGS84	Układ obejmujący całą Ziemię (globalny). Jest także podstawą pionowego układu odniesienia, wysokości są liczone od powierzchni elipsoidy.
ETRS89	GRS 80	Aktualny poziomy układ odniesienia dla Europy. Jest na stałe zakotwiczony do płyty kontynentalnej Europy i przesuwa się (około 2 cm na rok) względem układu WGS84. Różnica w porównaniu z WGS84 nie jest większa niż pół metra.

Jeżeli mamy określony poziomy układ odniesienia to lokalizację każdego punktu na powierzchni Ziemi (rzutowanej na elipsoidę) możemy określić używając współrzędnych **geograficznego układu współrzędnych** (ang. **GCS** – Geographic Coordination System). Geograficzny układ współrzędnych jest najważniejszą cechą danych geograficznych (posiadających lokalizację na powierzchni Ziemi). Wszystkie dane geograficzne mają określone położenie we współrzędnych geograficznych w danym poziomym układzie odniesienia.

Istnieją dwie współrzędne geograficzne: szerokość geograficzna (Y) i długość geograficzna (X), są to współrzędne kątowe, których jednostką jest **stopień**.



Rysunek 1.3 Współrzędne geograficzne

Szerokość i długość geograficzna zostały pokazane na Rysunku 1.3.

Co do szerokości to należy zwrócić uwagę na kilka istotnych faktów. Po pierwsze linia służąca do określenia szerokości, która przecina powierzchnię elipsoidy jest pod kątem prostym do powierzchni stycznej do płaszczyzny elipsoidy (obrazuje ją czarna przerywana linia). Po drugie kąty szerokości liczymy od płaszczyzny równika, na którym szerokość jest równa 0 stopni. Szerokości na półkuli północnej będą miały znak plus, a na półkuli południowej znak minus. Na biegunach północnym i południowym osiągną wartości 90.0 i -90.0 stopni (znaku plus zwyczajowo nie zapisujemy). Po trzecie warto zwrócić uwagę na fakt, że prosta przecinająca płaszczyznę równika, która określa szerokość geograficzną na elipsoidzie, nie przechodzi przez jej środek. Przechodziłaby tylko wtedy, gdyby modelem geoidy była kula.

Jeżeli przeanalizujemy definicję długości geograficznej to najlepiej będzie spojrzeć na nasz model z góry, czyli nad bieguna północnego. Zielone koło to płaszczyzna równika, a czerwone linie to połówki elips tworzących elipsoidę. Tworzą one południki. Czerwona linia oznaczona 0° to szczególny południk zwany południkiem zerowym (przebiega on przez obserwatorium w Greenwich koło Londynu). Ta połówka elipsy od bieguna północnego (N) po biegun południowy ograniczona osią Ziemi, tworzy płaszczyznę południka zerowego. Czerwona linia na rysunku to także rzut prostopadły południka na płaszczyznę równika. Określając kąt długości geograficznej punktu bierzemy zawsze pod uwagę południk zerowy i południk przecinający dany punkt. Kąt pomiędzy płaszczyznami tych południków, albo pomiędzy rzutami tych południków na płaszczyznę równika daje kąt długości geograficznej. Po przeciwej stronie południka zerowego jest południk 180° . Rozdziela on południki na przynależne do półkuli wschodniej i zachodniej. Południki na prawo od południka zero dają długości geograficzne wschodnie (dodatnie), a na lewo zachodnie (ujemne).

W GIS szerokość geograficzną będziemy oznaczać najczęściej literą Y, a długość geograficzną literą X. Wartości kątów będziemy zapisywać dwoma dziesiętnymi liczbami (np. X=18.3421, Y=54.2371). Ten sposób zapisu określa się skrótem DD.

Istnieją także inne sposoby zapisu współrzędnych geograficznych.

Tabela 1.3 Trzy rodzaje zapisu współrzędnych geograficznych.

Nazwa	Opis	Przykłady
DD	Stopnie dziesiętne (jedna liczba), szerokości południowe i długości zachodnie mają znak minus.	54.3241 -25.4 18.2345

DM	Stopnie i minuty dziesiętne (dwie liczby, pierwsza całkowita, druga dziesiętna). Po liczbach litery (N, S lub E, W).	54 19.446 N 25 24.0 W 18 14.07 E
DMS	Stopnie, minuty i sekundy (trzy liczby, dwie pierwsze całkowite). Po liczbach litery (N, S lub E, W).	54 19 26.76 N 25 24 0 W 18 14 4.2 E

Ze względu na to że istnieją trzy rodzaje zapisu powstaje problem przeliczania z jednego sposobu na drugi. Jest to proste jeśli zapamiętamy następującą zasadę.

Pamiętając, że 0.5 stopnia to 30 minut (analogicznie 0.5 minuty to 30 sekund) należy część dziesiętną pomnożyć przez 60. Natomiast ze względu na to, że 30 minut to 0.5 stopnia (a 30 sekund to 0.5 minuty) należy minuty lub sekundy podzielić przez 60. Czyli mnożymy albo dzielimy przez 60, tak aby relacja 0.5 do 30 lub 30 do 0.5 była zachowana.

Na przykład zamieniając 52.234 stopnia na stopnie i minuty część dziesiętną 0.234 musimy pomnożyć przez 60 co daje 14.04 czyli otrzymujemy $52^{\circ} 14.04'$.

Liczba cyfr po kropce dziesiętnej ma znaczenie, informuje nas o dokładności podanych wartości. Ze względu na to, że jedna minuta (na południku) odpowiada 1852 metrów (tak na marginesie jest to jednostka długości zwana milą morską) to:

0.1 minuty równa się 185.2 m

0.01 minuty równa się 18.52 m

0.001 minuty równa się 1.852 m

Czyli trzy miejsca po kropce w minutach dają nam dokładność rzędu 2 m. W przypadku stopni jeden stopień południka to około 111 000 m:

0.1 stopnia równa się 11100 m

0.01 stopnia równa się 1110 m

0.001 stopnia równa się 111 m

0.0001 stopnia równa się 11.1 m

0.00001 stopnia równa się 1.11 m

Czyli dopiero pięć miejsc po kropce daje nam dokładność pozycji rzędu jednego metra. Dlatego też sens ma zapis 54.00000 stopnia, oznacza to, że dokładność położenia jest rzędu metra.

Długości stopni i minut na równoleżnikach są zmienne i zmieniają się od równika, gdzie są w przybliżeniu równe długościom na południkach, w kierunku bieguna gdzie ich długość wynosi zero (uogólniając zmiana jest proporcjonalna do cosinusa szerokości geograficznej).

Trudno w to uwierzyć ale jeszcze trzydzieści lat temu praktycznie nie były powszechnie dostępne trzy technologie, bez których trudno sobie wyobrazić dzisiejszy świat. Są to internet, telefonia komórkowa i GPS. Pozwalają one praktycznie każdemu określić z dużą dokładnością aktualną własną lokalizację na powierzchni Ziemi, zapisanie tej lokalizacji i przesłanie jej innym.

Centryczna siatka nieba pokazuje satelity GPS wykryte przez telefon (GPS są reprezentowane przez dyski, a satelity GLONASS przez kwadraty, wielkość symbolu pokazuje moc sygnału). Kolor zielony pokazuje, że dany satelita jest używany do określenia lokalizacji.

Określany jest szacowany błąd określenia lokalizacji.

Wysokość jest podawana względem powierzchni elipsoidy lub średniego poziomu morza.



Igła kompasowa w centrum siatki pokazuje kierunek pola magnetycznego zmierzony przez telefon. Obliczona jest deklinacja magnetyczna. Podany jest azymut boku telefonu.

Zielone słupki obrazują moc sygnału, wykres jest podzielony na cztery sekcje (od lewej: GPS, GLONASS, BEIDOU i GALILEO).

DOP opisuje współczynnik geometrycznej dokładności zależny od geometrii konfiguracji satelitów.

Rysunek 1.4 GPS w smartfonie (na przykładzie bezpłatnego programu GPS Status & Toolbox)

Nasze smartfony są powszechnie wyposażone w różnego rodzaju czujniki w tym odbiorniki sygnału GPS. GPS jest to globalny nawigacyjny system satelitarny (GNSS – Global Navigation Satellite System). GPS jest zwykle używana jako nazwa systemu NAVSTAR GPS (NAVigation System with Time And Ranging Global Positioning System), który jest systemem armii USA. Ale nie jest to jedyny system obejmujący cały świat towarzyszą mu system rosyjski GLONASS, chiński BEIDOU i europejski GALILEO. Nasz smartfon potrafi już korzystać z więcej niż jednego systemu do jak najdokładniejszego określenia lokalizacji. Odbiór zakodowanych sygnałów satelitarnych z czterech satelitów pozwala na określenie naszego położenia w przestrzeni (szerokość i długość geograficzna i wysokość względem powierzchni elipsoidy). Większą dokładność uzyskujemy przy wykorzystaniu większej liczby satelitów. Coraz częściej smartfony

wykorzystują także technologię High Sensitivity GPS, która wykorzystuje specjalne przetwarzanie sygnału, pozwalające na wykorzystywanie bardzo słabych sygnałów w miejscowościach, gdzie jeszcze niedawno trudno było wykorzystywać GPS, jak wąskie ulice otoczone wysokimi budynkami czy gęsty las.

Pytania:

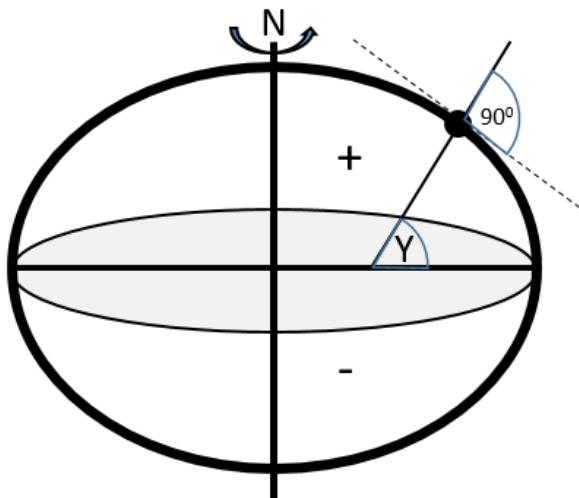
1. Dwie podstawowe cechy geoidy, to jej nieregularny kształt i spłaszczenie na biegunach. Co jest tego przyczyną.
2. Jakie dwa elementy definiują poziomy układ odniesienia, wymień dwa podstawowe układy używane obecnie.
3. Wykorzystując model elipsoidy zdefiniuj oś Ziemi, płaszczyznę równika i płaszczyznę południka.
4. Z jaką dokładnością w metrach (w przybliżeniu) wyznaczono położenie punktu jeśli pozycję określono jako: $60^{\circ} 22.0' N$ i $23^{\circ} 22.0' E$. Jaka jest dokładność w kierunku N-S i E-W.
5. Dla dowolnego miejsca wykonaj 15 pomiarów jego współrzędnych geograficznych za pomocą smartfona (odstęp czasu pomiędzy pomiarami powinien wynosić co najmniej 5 minut). Utwórz warstwę punktową z tych punktów. Co możesz powiedzieć o ich rozrzucie.
6. Obejrzyj film <https://www.youtube.com/watch?v=U3eX6QKS9kY> o tym jak działa GPS i napisz krótki tekst (max. 200 wyrazów) opisujący jego działanie.
7. Dlaczego na mapie wykonanej w układzie współrzędnych geograficznych kształt Polski będzie rozcięgnięty równoleżnikowo.



GCS – współrzędne geograficzne

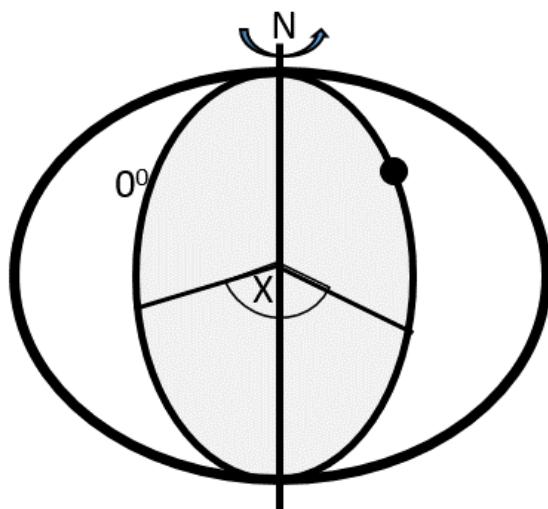
Szerokość geograficzna punktu na powierzchni Ziemi to kąt pomiędzy prostą prostopadłą do powierzchni Ziemi przechodzącą przez ten punkt i płaszczyzną równika.

Zakres: 0 – 90 N (+), 0 – 90 S (-) w stopniach



Długość geograficzna punktu na powierzchni Ziemi to kąt pomiędzy płaszczyzną południka zerowego i płaszczyzną południka na którym ten punkt leży.

Zakres: 0 – 180 E (+), 0 – 180 W (-) w stopniach



Rozdział 2

Dane przestrzenne

Niezależnie od tego w jaki sposób wykorzystujemy GIS dane są jego niezbędnym elementem. Dziedziną, która często jest postrzegana jako symbol nowoczesności jest obecnie ***data science***. Pozwala ona nie tylko pozyskiwać z danych informację ale wykorzystując techniki sztucznej inteligencji także wiedzę, czyli wykrywać jakie są mechanizmy zachodzenia różnorodnych złożonych procesów i stosować je w prognozach lub działaniach. Dziedzina ta wykorzystuje ogromne ilości danych, które cały czas są tworzone za pomocą najróżniejszych technik. Dane te aby mogły być przetwarzane przez uniwersalne narzędzia i metody *data science* muszą być przedstawiane w uniwersalny sposób. Jednym z powszechnie stosowanych rozwiązań jest najprostsze z możliwych – tekstowe pliki formatu **CSV** (comma-separated values – czyli wartości rozdzielone przecinkiem).

```
kraj,piwo_p,wodka_p,wino_p,litry_alkohol,kontynent
Afghanistan,0,0,0,0.0,Asia
Albania,89,132,54,4.9,Europe
Algeria,25,0,14,0.7,Africa
Andorra,245,138,312,12.4,Europe
Angola,217,57,45,5.9,Africa
Antigua & Barbuda,102,128,45,4.9,North America
Argentina,193,25,221,8.3,South America
Armenia,21,179,11,3.8,Europe
Australia,261,72,212,10.4,Oceania
Austria,279,75,191,9.7,Europe
```

Rysunek 2.1 Przykładowy plik tekstowy z danymi formatu **CSV**. Dane opisują różne rodzaje spożycia alkoholu w poszczególnych krajach. Każdy wiersz zawiera kolejno następujące dane: nazwę kraju, roczną liczbę porcji piwa na mieszkańca, porcji wódki, porcji wina, roczne spożycie litrów czystego alkoholu przez mieszkańca, nazwę kontynentu. Dodatkowo pierwsza linia zawiera nazwy poszczególnych zmiennych.

Każdy wiersz opisuje oddzielną obserwację z kompletem pozyskanych pomiarów (zmiennych, atrybutów). Te atrybuty każdej obserwacji mogą mieć charakter

(typ) tekstu, liczby całkowitej lub dziesiętnej (zamiast przecinka używamy kropki). Tak przygotowane dane mogą być w prosty sposób stablicowane co jest zwykle pierwszym krokiem wprowadzenia danych do środowiska analitycznego. Możemy je wprowadzić także do GIS i przedstawić w postaci tablicy. **Tablica danych tekstowych jest rodzajem danych stosowanych w GIS** (choć dane te nie mają charakteru przestrzennego).

	kraj	piwo_p	wodka_p	wino_p	litry_alkohol	kontynent
1	Colombia	159	76	3	4.2	South America
2	Chad	15	1	1	0.4	Africa
3	Chile	130	124	172	7.6	South America
4	Brunei	31	2	1	0.6	Asia
5	Bulgaria	231	252	94	10.3	Europe
6	Botswana	173	35	35	5.4	Africa
7	Brazil	245	145	16	7.2	South America

Atrybuty mają swój typ danych.

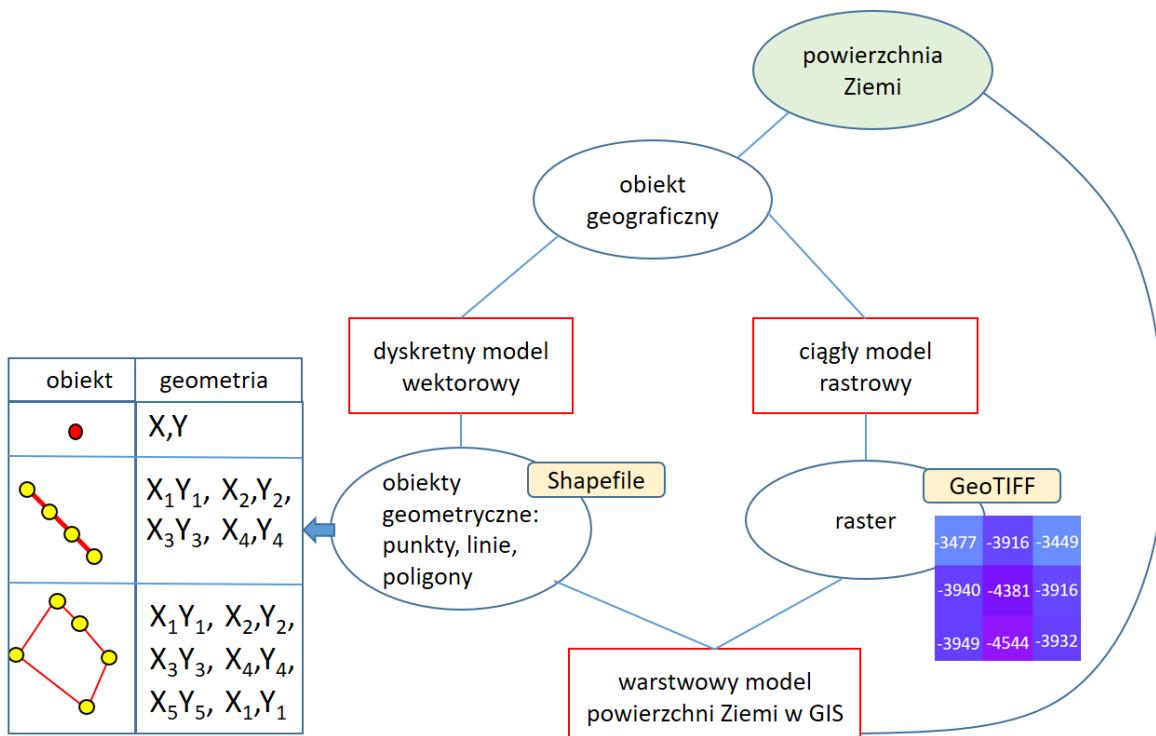
Id	Name	Alias	Type	Type name	Length	Precision
abc 5	kontynent		QString	text	0	0
1.2 4	litry_alkohol		double	double	0	0
123 3	wino_p		int	integer	0	0
123 2	wodka_p		int	integer	0	0
123 1	piwo_p		int	integer	0	0
abc 0	kraj		QString	text	0	0

Dane organizowane są w tablicach. Każda obserwacja tworzy oddzielny wiersz (**rekord**). Każdy pomiar (atrybut, zmienna) obserwacji jest zapisywany w oddzielnej kolumnie (**polu**).

Rysunek 2.2 Dane tekstowe w GIS w postaci tablicy

Struktura tych danych polegająca na specyficznej organizacji danych w tablicy jest metodą **data science** zapewniającą uniwersalne przetwarzanie danych. Tak zorganizowane i uporządkowane dane noszą nazwę **tidy data** (tidy można tłumaczyć jako zorganizowane lub uporządkowane). Mają one następujące wspólne cechy. Każda określona zmienna (atrybut) danej obserwacji jest zapisywana w kolumnie. Każda oddzielnna obserwacja szeregu zmiennych tworzy oddzielny wiersz. Dla każdego tematu obserwacji tworzona jest oddzielnna tablica. Jeżeli posiadamy parę tablic to powinny one zawierać kolumnę, która umożliwi ich łączenie. Nasze obserwacje dotyczą poszczególnych krajów, które mogą być traktowane jako obiekty geograficzne, oczywiście obiekty w tej tablicy reprezentowane są wyłącznie przez nazwę.

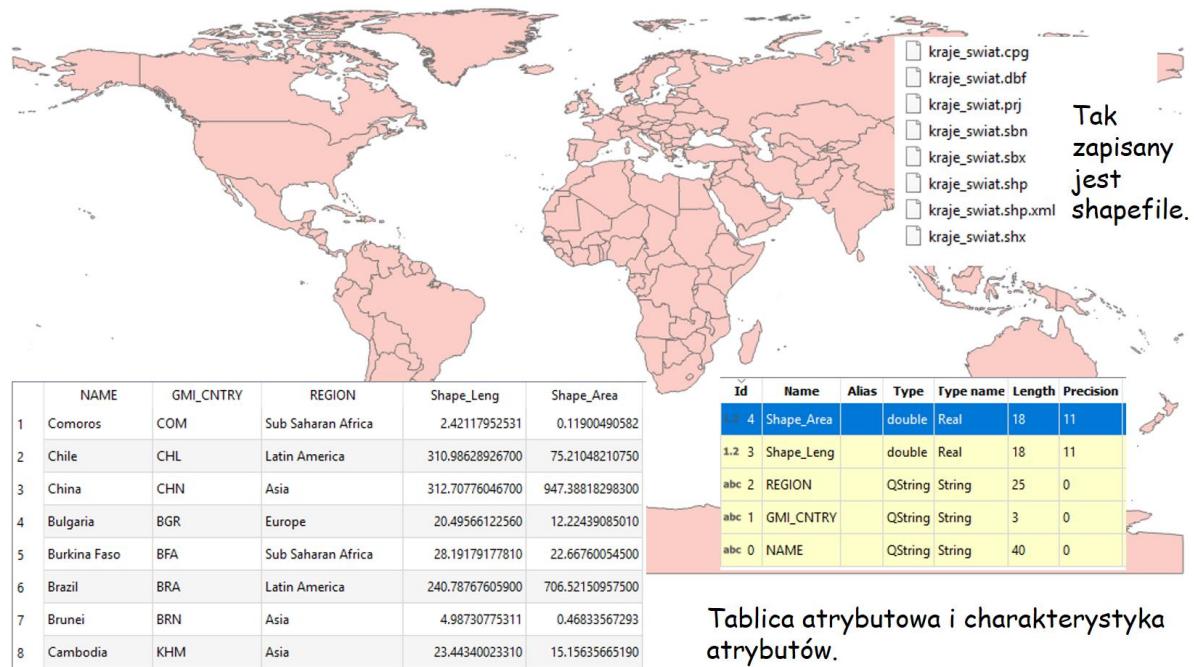
Aby obiekty geograficzne powierzchni Ziemi mogły być przedstawione w postaci danych, potrzebny jest **model danych przestrzennych**, który pozwoli na ich reprezentację w postaci cyfrowej.



Rysunek 2.3 Model danych przestrzennych stosowany w GIS

Powierzchnia Ziemi jest pokryta szeroko rozumianymi obiektami geograficznymi. Obiektami mogą być budynki, jeziora, drogi ale także teren i jego ukształtowanie, oraz ocean i jego temperatura powierzchniowa. Obiekty mogą być dzielone na klasy tematyczne (np. rzeki, jeziora, pokrycie terenu, wysokość terenu). Te różne rodzaje obiektów geograficznych można jednak podzielić na dwie podstawowe klasy. Klasę **obiektów dyskretnych**, które zajmują określone miejsce w przestrzeni i często posiadają wyraźnie określone granice i klasę **obiektów ciągłych lub powierzchniowych** jak na przykład rozkład opadu deszczu w ciągu roku czy wysokość terenu. Przykładem obiektów dyskretnych będą: punkty pomiarowe, studnie, źródła, rzeki, drogi, jeziora czy obszary powiatów. Ich dobrym reprezentantem będą punkt, linia oraz wielobok (zwany powszechnie w środowisku GIS poligонem). Geometrie tych obiektów tworzą przestrzenny element **wektorowego modelu danych**. Dane wektorowe są pierwszym rodzajem danych przestrzennych wykorzystywanym w GIS. Geometria punktu na powierzchni Ziemi jest zapisywana jako dwa jego współrzędne (np. długość i szerokość geograficzną). W przypadku konstrukcji geometrycznej linii potrzebujemy szeregu punktów (te konstrukcyjne punkty nazywamy

werteksami). Podobnie przy tworzeniu poligona (wieloboku) używamy werteksów i aby zamknąć ich ciąg powtarzamy pierwszy werteks. Zostało to przedstawione w tablicy obiektów i ich geometrii na rysunku 2.3. Każdy obiekt geometryczny wraz ze swoimi atrybutami tworzy jeden rekord (wiersz) w tablicy danych (tzw. **tablicy atrybutowej**). Na rysunku 2.4 przedstawiono zbiór danych przestrzennych tworzących warstwę krajów świata.



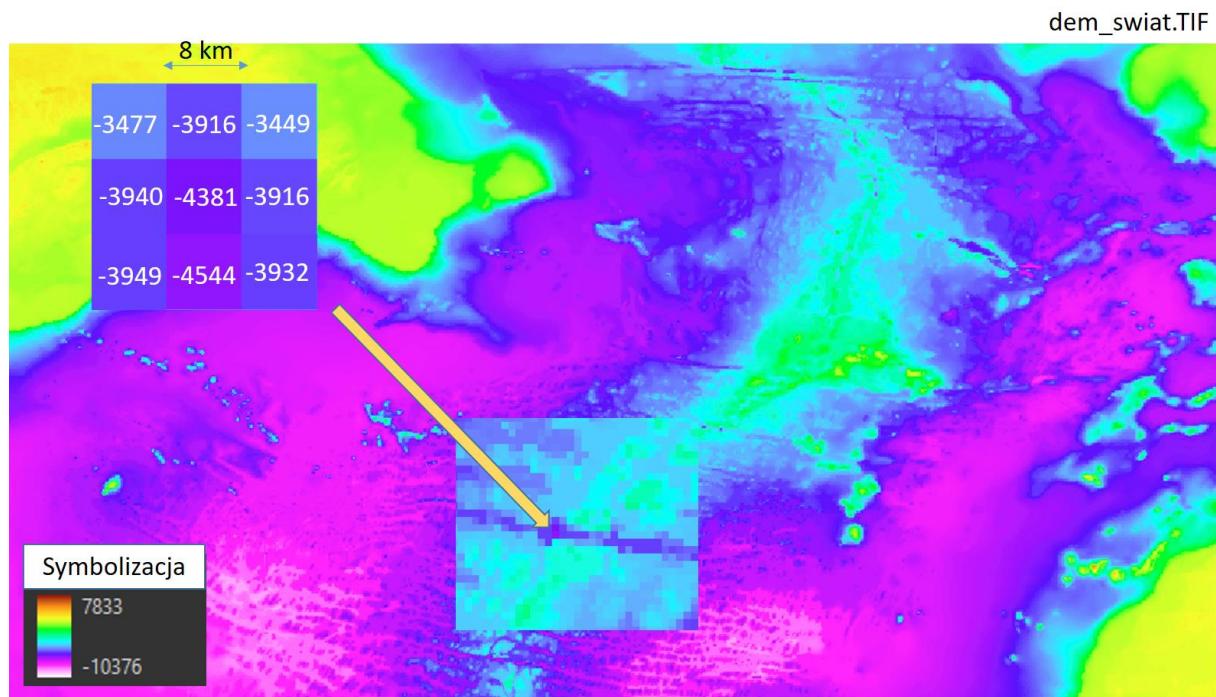
Tablica atrybutowa i charakterystyka atrybutów.

Rysunek 2.4 Wektorowa warstwa poligonów zapisana w formacie Shapefile

Jest to warstwa poligonów. Warstwy (zbiory danych wektorowych) tworzy się oddzielnie dla punktów, linii i poligonów. Pod względem logicznym (tak jak widzimy ją w GIS) każda warstwa składa się z rekordów zawierających geometrie obiektów i związane z nią atrybuty. Często geometria i atrybuty są fizyczne zapisywane oddzielnie. Na rysunku 2.4 przedstawiono najbardziej powszechny format zapisu danych wektorowych **Shapefile**. Warstwa poligonów wraz z atrybutami została zapisana w szeregu plików o wspólnej nazwie i różnych rozszerzeniach. Tablica z atrybutami, która może być traktowana jako samodzielna jest zapisywana z rozszerzeniem **.dbf** (jest to standard zapisu plików bazy danych przez dBASE). Warstwa posiadająca obiekty geometryczne może być wyświetlona na obszarze roboczym danego programu GIS. Sposób rysowania poszczególnych obiektów nosi nazwę **symbolizacji**. Na rysunku 2.4 wszystkie poligony mają takie samo wypełnienie i obrus (szara linia). Rodzaj symbolizacji może być zależny od wartości w wybranym polu (kolumnie) tablicy atrybutowej.

Obiekty o charakterze ciągłym są reprezentowane za pomocą **modelu rastrowego**. Dane rastrowe są drugim podstawowym rodzajem danych

przestrzennych używanym w GIS. Model rastrowy tworzy mozaikę kwadratów (komórek rastra) pokrywających powierzchnię Ziemi. Mozaika tworzy prostokąt składający się z wierszy i kolumn komórek. Stąd istotną charakterystyką rastra jest **liczba wierszy i kolumn**. Ponieważ komórki pokrywają powierzchnię Ziemi kolejną ważną cechą jest długość ich boku na powierzchni Ziemi, jest to tak zwana **rozdzielcość przestrzenna rastra** (wyrażamy ją w metrach).

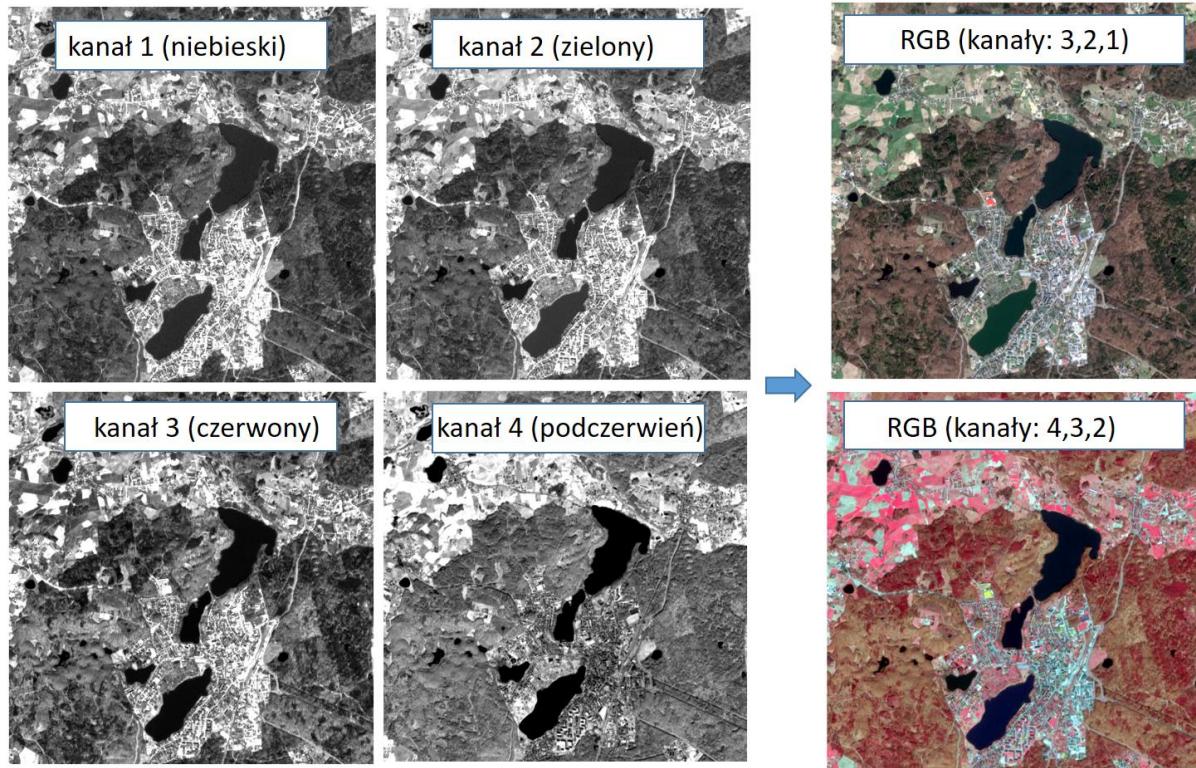


Rysunek 2.5 Rastrowa warstwa głębokości oceanu. Kolejne przybliżenia wycinków rastra pokazują jego strukturę składającą się z kwadratowych komórek

Każda komórka rastra zawiera jedną wartość. Może to być liczba całkowita tak jak na Rysunku 2.5, gdzie przedstawia ona głębokość w środku komórki, ale może także być identyfikatorem pewnej klasy (np. liczba 0 reprezentuje wodę a, liczba 1 ląd). Komórki rastrowe mogą także zawierać liczby rzeczywiste. Mówimy, że raster mogą być różnego **typu**. Podobnie jak w przypadku warstw wektorowych wizualizacja rastru wymaga **symbolizacji**, polega ona na przypisaniu określonego koloru lub odcienia szarości do konkretnych wartości lub ich przedziałów na rastrze. Istnieje wiele formatów zapisu danych rastrowych, jednym z najbardziej uniwersalnych jest **GeoTIFF**. W formacie tym zapisuje się rastry jako pojedynczy plik z rozszerzeniem **.tif**, informacja o położeniu w przestrzeni jest zawarta wewnętrz tego pliku.

Ważnym rodzajem danych wykorzystywanych w GIS są zdjęcia satelitarne. Są one pozyskiwane jako rastry o bardziej złożonej strukturze niż omawiane powyżej. Zdjęcia satelitarne to zestawy dokładnie pokrywających się rastrow

tworzących dla wybranych przedziałów promieniowania (najczęściej odbitego) dochodzącego z powierzchni Ziemi. Rastry te nazywamy **kanałami** zdjęcia satelitarnego. Może być ich parę, kilkanaście lub nawet kilkaset, ale są często zapisywane jako jeden plik formatu GeoTIFF w postaci tzw. **kompozytu**. O takich plikach mówimy, że są to **rastry wielokanałowe**.



Rysunek 2.6 Cztery kanały zdjęcia satelitarnego i symbolizacja ich kompozytu dla dwóch zestawów kanałów

Na rysunku 2.6 z lewej strony pokazane są cztery kanały zdjęcia satelitarnego (są one wyświetlane za pomocą odcieni szarości). Połączono je w kompozyt czterokanałowy. Wizualizacja kompozytu polega na wykorzystaniu modelu kolorów **RGB** (czerwony, zielony, niebieski) do jednoczesnego wyświetlenia trzech kanałów w tych kolorach, w rezultacie tworzy się bogata paleta różnych barw. Jeżeli do RGB przypiszemy odpowiadające im zdjęcia (zrobione w zakresie czerwonym, zielonym, niebieskim) otrzymamy obraz o kolorach zbliżonych do rzeczywistych (tak jak widzi to ludzkie oko). Jeżeli natomiast przypiszemy inny zestaw kanałów (kompozyt dolny) otrzymamy tzw. fałszywe kolory, które mogą nam dostarczyć informacji niewidocznych w innych zestawach kanałów.

W dostępnych zbiorach danych klimatycznych, meteorologicznych i oceanograficznych, stosuje się rastrowe dane **wielowymiarowe**. Wymiarem, może być czas, wysokość nad powierzchnią Ziemi lub głębokość. W rezultacie

powstaje struktura danych, w której mamy oddzielne rastry dla danych momentów w czasie i na przykład głębokości. Istnieje szereg formatów tych danych z których podstawowym jest **netCDF** (Network Common Data Format) z plikami o rozszerzeniu (.nc).

Warstwy wektorowe i rastrowe reprezentujące istotne dla danego zagadnienia obiekty geograficzne tworzą **warstwowy model powierzchni Ziemi**.

W praktyce warstwy w systemach GIS przechowujemy (zapisujemy i odczytujemy) w **Geobazach Danych**. Dwie podstawowe geobazy to File Geodatabase (**FileGDB**) i GeoPackage (**GPKG**). File Geodatabase jest domyślną geobazą ArcGIS Pro i pod względem fizycznym przyjmuje postać katalogu z nazwą zawierającą rozszerzenie **.gdb**. Z kolei GeoPackage jest domyślną geobazą programu QGIS i tworzy jeden plik. Zapisywanie danych w formacie SHP niesie ze sobą wiele niedogodności. Pojedyncza warstwa składa się z wielu plików, jej wielkość jest ograniczona do 2 GB, a jej nazwa nie może przekraczać 10 znaków. Geobazy nie posiadają takich ograniczeń. W ograniczonej funkcjonalności mogą być także wykorzystywane przez oba programy (ArcGIS Pro, QGIS).

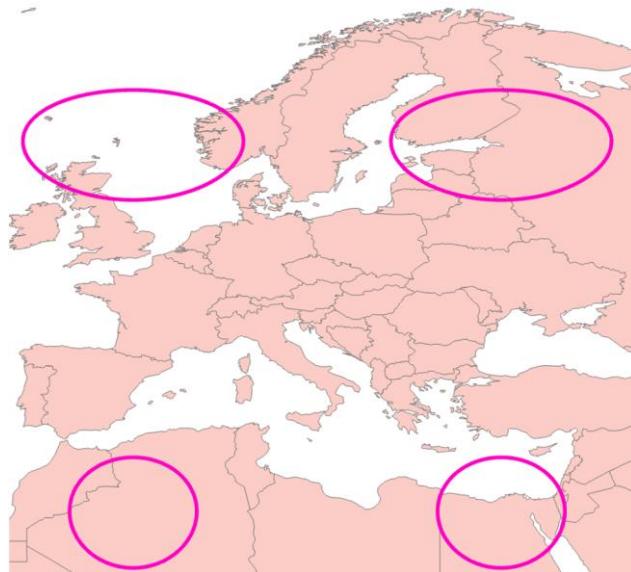
Pytania:

1. Utwórz plik tekstowy zawierający 10 nazw miast oraz ich położenie geograficzne (dane pozyskaj z Internetu) w postaci trzy kolumnowej tablicy (typu tidy data) z rozszerzeniem .csv. Pierwsza linia tablicy powinna zawierać nazwy kolumn.
2. Omów podstawowe cechy i charakterystyki modelu wektorowego i rastrowego.
3. Wyświetl cztery rodzaje warstw: punkty, linie, poligony oraz raster używając trzech różnych symbolizacji.
4. Wy tłumacz następujące pojęcia: warstwowy model powierzchni Ziemi, kompozyt, rozdzielcość przestrzenna rastra, typ rastra, liczba kanałów rastra.
5. Opisz dowolny raster podając następujące charakterystyki: liczba kanałów, typ danych, liczba wierszy i kolumn, rozdzielcość przestrzenna rastra.

Rozdział 3

Współrzędne prostokątne, odwzorowania

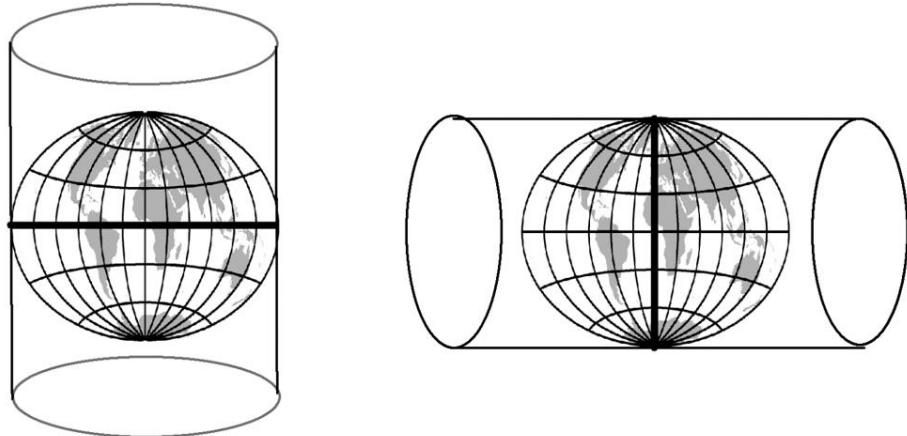
Współrzędne geograficzne służą nam znakomicie do określenia położenia punktów na powierzchni Ziemi (elipsoidy). Problemy pojawiają się, gdy chcemy zmienić ją na płaską powierzchnię ekranu lub mapy.



Rysunek 3.1 Mapa poligonów krajów Europy i północnej Afryki we współrzędnych geograficznych. Indykatory Tissota (czerwone elipsy) pokazują jej zniekształcenia

Mapa na Rysunku 3.1 traktuje stopnie szerokości i długości geograficznej jako jednostki długości układu współrzędnych prostokątnych XY aby stworzyć obrys krajów Europy i północnej Afryki. Problem polega na tym, że jeden stopień długości geograficznej (w poziomie) zmienia swą długość liniową (dystans w metrach na powierzchni Ziemi) wraz ze zmianą szerokości geograficznej. Długość liniowa jednego stopnia długości geograficznej na 60° szerokości geograficznej to połowa jego długości na równiku. Ze względu na to, że na naszej mapie jeden stopień długości ma stałą długość liniową geometria obiektów ulega zniekształceniu (zwiększającemu się rozciągnięciu w poziomie wraz z postępem na północ). Wizualną miarą tego zniekształcenia są obrys czerwonych elips zwane **indykatorami Tissota**. W rzeczywistości są to okręgi o promieniu 500 km. Ich zniekształcenie w stosunku do okręgu informuje nas o tym jakim deformacjom ulegają odległość, kierunek, powierzchnia i kształt na mapie.

Aby te zniekształcenia były jak najmniejsze stosuje się różne metody przeniesienia położenia punktów z powierzchni elipsoidy na powierzchnię płaską. Metody te to tzw. **odwzorowania**, funkcje przeliczające współrzędne geograficzne (w stopniach) na płaskie **współrzędne prostokątne** (w **metrach**) minimalizujące zniekształcenia. Ogólną (i przybliżoną) zasadą tych metod jest rzutowanie geometryczne powierzchni kuli na powierzchnię płaskie lub takie, które mogą być następnie rozwinięte do powierzchni płaskiej.



Rysunek 3.2 Rzutowanie powierzchni kuli na pobocznicę walca jest podstawą odwzorowania Merkatora. Z prawej strony pokazane jest poprzeczne odwzorowanie Merkatora, którego modyfikacje są podstawą wielu współczesnych układów współrzędnych i systemów układów współrzędnych stosowanych w geodezji i kartografii

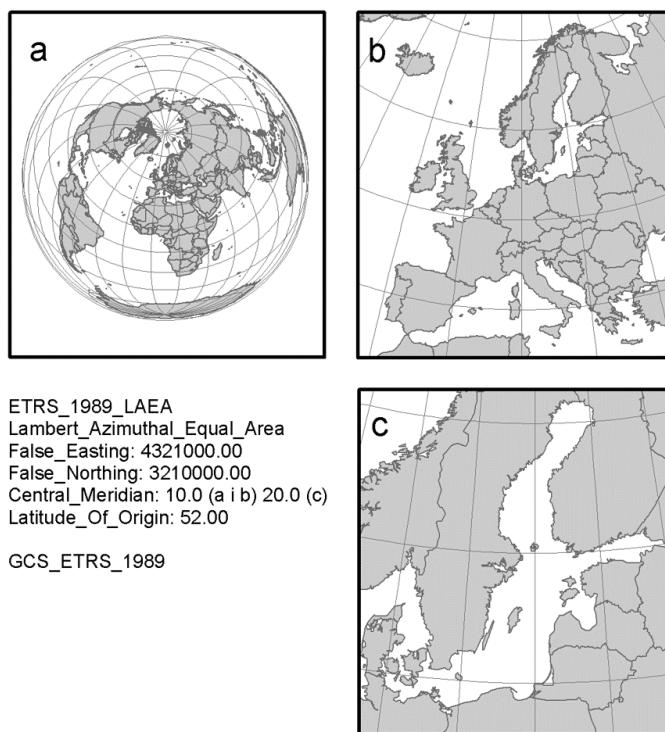
Każdemu odwzorowaniu odpowiada odwrotna funkcja odwzorowawcza, która pozwala na przeliczenie współrzędnych prostokątnych danego odwzorowania z powrotem do współrzędnych geograficznych (dla danego poziomego układu odniesienia). Aby zastosować dane odwzorowanie do konkretnego obszaru powierzchni Ziemi, musimy dostarczyć do niego szereg parametrów z których najważniejszymi są: południk środkowy (central meridian), równoleżnik standardowy (standard parallel), odcięta punktu głównego (false easting) i rzędna punktu głównego (false northing). Odwzorowania wraz z przyjętymi parametrami tworzą czasem oficjalne układy współrzędnych lub ich zestawy zwane systemem układów współrzędnych. W ogromnej liczbie odwzorowań, układów współrzędnych i systemów układów współrzędnych łatwo się jest pogubić. Na szczęście najczęściej używamy tylko kilka z nich.

Jeżeli chcemy wykonać mapę lub przygotować warstwę do analizy dużych obszarów np. mórz, dowolnego oceanu, kontynentu lub jego części to można

wykorzystać **wiernopowierzchniowe azymutalne odwzorowanie Lamberta** (Lambert Azimuthal Equal Area - LAEA). To odwzorowanie wraz ze zdefiniowanymi parametrami jest zalecany przez *Eurostat* i *European Environment Agency* układem współrzędnych dla Europy przy tworzeniu map w skalach mniejszych niż 1:500 000.

Coordinate Reference System	Authority ID
ETRS89 / LAEA Europe	EPSG:3035

Istniejące układy współrzędnych identyfikujemy za pomocą nazwy w tym przypadku będzie to LAEA Europe, ale istnieje też możliwość znalezienia jej za pomocą numeru **EPSG**. EPSG Geodetic Parameter Dataset został stworzony przez organizację o nazwie International Association of Oil & Gas Producers i stanowi z założenia kompletny zbiór układów współrzędnych. Możemy oczywiście dowolnie je modyfikować podstawiając własne parametry i tworzyć nowe układy współrzędnych (najwygodniej robić to na bazie już istniejących). Te identyfikatory występują często pod nazwą WKID (well-known ID). Mapa Europy będzie miała postać.



Rysunek 3.3 Trzy sposoby wykorzystania LAEA Europe. Dla połowy globu z Europą w centrum (a). Typowe przedstawienie Europy z południkiem środkowym 10° (b). Modyfikacja południka środkowego do 20° (c)

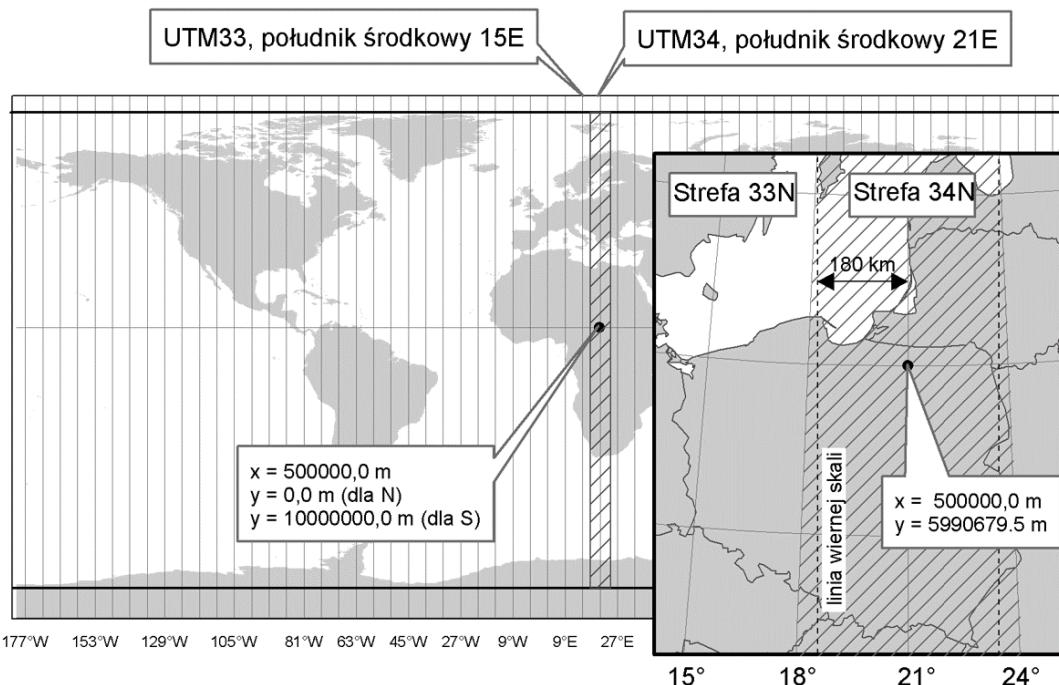
Pełen opis układu współrzędnych LAEA Europe będzie wyglądał w następujący sposób.

Coordinate System Details	
Projected Coordinate System	ETRS 1989 LAEA
Projection	Lambert Azimuthal Equal Area
WKID	3035
Authority	EPSG
Linear Unit	Meters (1.0)
False Easting	4321000.0
False Northing	3210000.0
Central Meridian	10.0
Latitude Of Origin	52.0
Geographic coordinate system	GCS ETRS 1989
WKID	4258
Authority	EPSG
Angular Unit	Degree (0.0174532925199433)
Prime Meridian	Greenwich (0.0)
Datum	D ETRS 1989
Spheroid	GRS 1980
Semimajor Axis	6378137.0
Seminor Axis	6356752.314140356
Inverse Flattening	298.257222101

Wszystkie dane geograficzne bazują na określonym poziomym układzie odniesienia (Datum), definiującym geograficzny układ współrzędnych. W dolnej części opisu znajduje się definicja tego układu GCS ETRS 1989 (Geographic Coordinate System + Datum). W górnej części znajduje się opis układu współrzędnych prostokątnych, który powstał w rezultacie odwzorowania z wymienionymi parametrami. Należy zwrócić uwagę, że oba układy posiadają swoje WKID.

Przy skalach większych niż 1:500000 powszechnie stosuje się **system współrzędnych UTM** (Universal Transverse Mercator) obejmujących prawie cały glob ziemski. Oparty on jest na poprzecznym odwzorowaniu Merkatora. W systemie tym Ziemia jest podzielona na 60 stref północnych i południowych (60 uzupełniających się układów współrzędnych) o szerokości 6 stopni długości geograficznej. Strefy te otaczają kolejne południki środkowe stref, które są prawie styczne z powierzchnią walca (Rysunek 3.2). Dwa południki styczne znajdują się po obu stronach południka środkowego w odległości około 180 km. Strefy nie pokrywają całej powierzchni Ziemi. Na północy sięgają 84° N a na południu 80° S. Strefy oznaczone są kolejnymi numerami i literą wskazującą półkulę (strefy 1N i 1S zaczynają się na 180° W a ich numeracja wzrasta w kierunku wschodnim). Początek każdego układu współrzędnych systemu UTM znajduje się na przecięciu południka środkowego danej strefy z równikiem. W

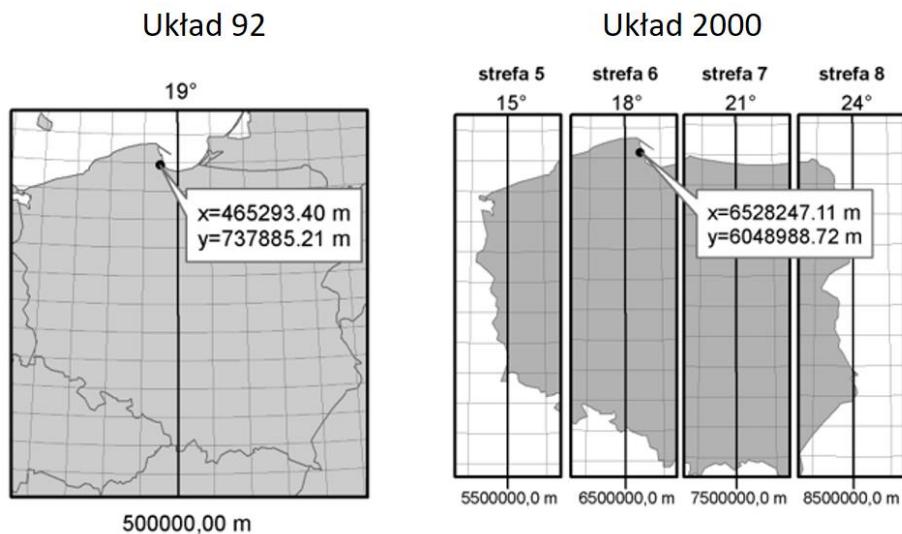
tym miejscu **X=500000 m** (rośnie na wschód), natomiast Y dla półkuli N = 0 m (rośnie na północ), a dla półkuli południowej N = 10 000 000 m (maleje na południe).



Rysunek 3.4 System współrzędnych UTM. Polskę obejmują dwie strefy UTM 33N i UTM 34N

Linie wiernej skali pokrywają się południkami stycznymi. Na granicach strefy błąd odległości jest rzędu 1 cm dla odległości 25 m, zniekształcenia kształtu i powierzchni są także niewielkie. Dopuszczalnym jest rozszerzanie danej strefy na strefy sąsiadujące. Oznacza to, że dla całego obszaru Polski możemy używać strefy 33N albo 34N. Strefy te mogą posiadać poziomy układ odniesienia WGS84 albo ETRS89, dla większości zastosowań różnice są nieistotne. Uzupełnieniem UTM dla dalekiej północy i południa jest układ współrzędnych UPS (Universal Polar Stereographic), który zapewnia podobną dokładność.

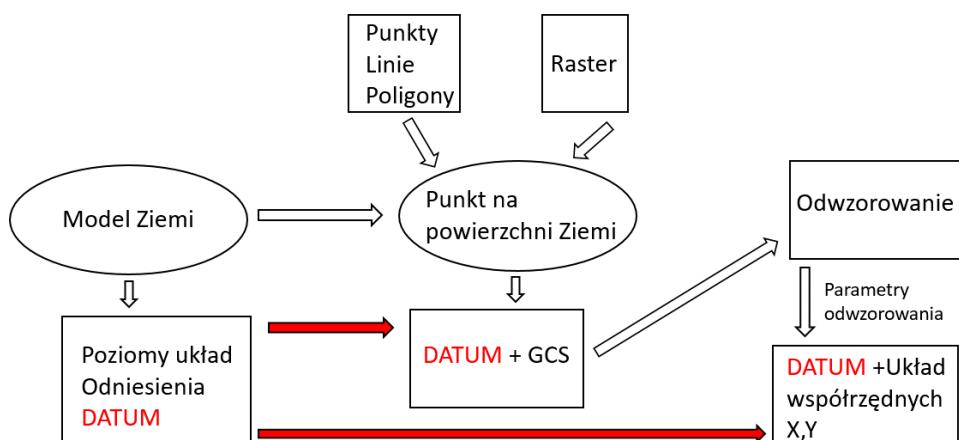
Praktycznie każdy kraj posiada urzędowy system odniesienia przestrzennego, w przypadku Polski jest to **Układ 1992** i **Układ 2000**. Ich poziomym układem odniesienia jest ETRS89. Oba układy wykorzystują odwzorowanie Gaussa-Krügera (pokrewne Mercatora). Wykazują one szereg podobieństw do systemu UTM. W układzie 1992 południkiem średnim jest 19° E, a początkowym równoleżnikiem jest równik. Punkt przecięcia równika z południkiem średnim (początek układu XY) ma współrzędne X= 500 000 m, Y= - 5 300 000 m.



Rysunek 3.5 Układy współrzędnych geodezyjnych obowiązujące w Polsce od 1 stycznia 2010 roku – układ 92 i układ 2000

Układ 2000 składa się czterech stref o numerach 5, 6, 7, 8 (środkowymi południkami są odpowiednio 15° E, 18° E, 21° E, 24° E). Współrzędna X w każdej ze stref jest poprzedzana numerem strefy. Pozostała jej część jest wyznaczana podobnie jak w UTM i Układzie 92. Współrzędna Y jest wyznaczana tak jak w UTM, dla półkuli północnej z zerową wartością na równiku.

Na schemacie poniżej przedstawiono w jaki sposób definiujemy lokalizację na powierzchni Ziemi. Podstawą jest model Ziemi zdefiniowany za pomocą poziomego układu odniesienia (DATUM). Na tym modelu lokalizacje przedstawiamy za pomocą punktów i pikseli (środki pikseli to też punkty). Położenie punktów jest jednoznacznie określone za pomocą DATUM i współrzędnych geograficznych. Współrzędne geograficzne za pomocą odwzorowania z parametrami są przeliczane na X,Y układu współrzędnych.



Pytania :

1. Dla dowolnej warstwy podaj:

1. Jaki jest poziomy układ odniesienia?
2. Jak nazywa się układ współrzędnych?
3. Jeśli jest to układ współrzędnych prostokątnych to:
 - a. Jakiego odwzorowania użyto?
 - b. Jaki jest południk środkowy?
 - c. Jakie współrzędne są na przecięciu osi X i Y?

2. Dlaczego na mapie wykonanej w układzie współrzędnych geograficznych kształt Polski będzie rozciągnięty równoleżnikowo.

3. Określ współrzędną X w układzie 1992 jeśli punkt leży 23500m na zachód od 19 stopnia długości geograficznej wschodniej.

Rozdział 4

Inżynieria danych przestrzennych

GIS jest narzędziem przeznaczonym do pracy z danymi przestrzennymi, i chociaż charakteryzuje się kompleksową funkcjonalnością, co oznacza, że może być wykorzystywany do realizacji najróżniejszych zadań, to zawsze podstawą każdej prowadzonej w nim operacji są dane o charakterze przestrzennym. Pojęcie inżynierii danych przestrzennych oznacza kompleksowy proces na końcu którego otrzymujemy dane potrzebne do wykonania konkretnego zadania. Może to być wykonanie mapy lub atlasu, tworzenie bazy danych do realizacji projektu, udostępnienie informacji przestrzennej w internecie, przeprowadzenie analizy opartej o metody statystki przestrzennej czy tworzenie modelu prognostycznego.

Pierwszym krokiem na tej drodze jest zawsze wytworzenie lub pozyskanie danych przestrzennych. Przy czym w praktyce tworzenie nowych danych jest działaniem epizodycznym, a pozyskiwanie powszechnym. Trudno wyobrazić sobie ten proces bez dostępu do danych za pomocą internetu. Praktyczna wiedza o tym jakie dane są dostępne i gdzie je można znaleźć, jest niezmiernie przydatna na tym etapie. Zasoby dostępnych danych ulegają cały czas zmianom i wymagają ciągłej kreatywności w ich odkrywaniu. Umiejętność znajdywania potrzebnych danych jest bardzo ważną i cenną kompetencją geoinformatyczną.

Pozyskane dane wymagają dokładnego zbadania zarówno pod względem ich przydatności, struktury jak i jakości. Dwie istotne metody służące do tego celu to wizualizacja i eksploracyjna (wstępna) analiza danych (EDA – Exploratory Data Analysis). Wizualizacja danych wektorowych i rastrowych na tle innych warstw z umiejętnie dobraną symbolizacją pozwala na odkrycie zarówno błędów jak i istniejących regularności w danych. Eksploracyjna analiza danych posługuje się szeregiem metod statystycznych, których rezultaty są zwykle przedstawiane w postaci różnego rodzajów wykresów (np. histogramu czy wykresu rozrzutu).

Przetworzenie danych do wymaganej postaci różni się w zależności od tego czy mamy do czynienia z danymi wektorowymi czy rastrowymi. Pracując z danymi wektorowymi najczęściej musimy utworzyć podzbiór potrzebnych obiektów i związanych z nimi atrybutów. Często istnieje także potrzeba dodania nowych atrybutów. Mogą być one obliczone z geometrii obiektu albo z atrybutów już

istniejących lub dodane z innych źródeł. Nowy zbiór danych tworzymy także w przypadku danych rastrowych. Może on być wynikiem połączenia kilku rastrów, wycięcia podzbioru z rastra, wybrania niektórych kanałów z kompozytu, zmiany typu danych lub przeliczenia wartości pikseli w rastrze. Nowe dane mogą być utworzone w innym układzie współrzędnych niż oryginalne dane wektorowe lub rastrowe. Specyfiką danych przestrzennych jest ich geometria. Punkty, linie i poligony mogą być tworzone i edytowane. Skany zdjęć analogowych (na papierze fotograficznym) lub map są rastrami bez dowiązania do powierzchni Ziemi. Aby mogły być wykorzystywane w GIS musi być im nadana georeferencja (czyli dowiązanie do powierzchni Ziemi)..

Luty 2022

4.1 Pozyskiwanie i tworzenie danych przestrzennych

Internet jest obecnie podstawowym źródłem danych dla GIS. Znajdujemy je nie tylko wchodząc na strony reprezentatywnych instytucji, ale także eksplorując różne strony w poszukiwaniu często bezcennych wskazówek, jak pozyskać potrzebne dane. Dane powinny spełniać nasze wymagania i być bezpłatne. Standardem staje się wytwarzanie danych za publiczne pieniądze czego konsekwencją jest publiczny do nich dostęp. Cieszy, że te zasady coraz częściej dotyczą także naszych krajowych zasobów. Podstawowym zakresem, który musimy wziąć pod uwagę jest skala przestrzenna danych, którą można sprecyzować trójwyrazowym hasłem – świat, Europa, Polska. Potrzebne dane nie zawsze będą dostępne w przestrzennych modelach danych. Często możemy je uzyskać jako dane tekstowe, ale z atrybutem, który będziemy mogli wykorzystać do połączenia tych danych z przestrzenną bazą. Przykładowo dane tabelaryczne z ludnością krajów świata, możemy łatwo połączyć z wektorowymi danymi obrysów krajów świata, jeżeli posiadają one atrybut z nazwą kraju.

Dane w skali całego świata reprezentują duży stopień generalizacji. Dobrym startowym miejscem danych wektorowych jest strona o adresie: <http://www.naturalearthdata.com/>



Free vector and raster map data at
1:10m, 1:50m, and 1:110m scales

 Search

Home Features Downloads Blog Issues About



Natural Earth is a public domain map dataset available at 1:10m, 1:50m, and 1:110 million scales. Featuring tightly integrated vector and raster data, with Natural Earth you can make a variety of visually pleasing, well-crafted maps with cartography or GIS software.

Natural Earth was built through a collaboration of many [volunteers](#) and is supported by [NACIS](#) (North American Cartographic Information Society), and is free for use in any type of project (see our [Terms of Use](#) page for more information). [Get the Data](#)

Dane są dostępne w trzech skalach dokładności. W każdej z nich zestawione są tematycznie w kategoriach: *Cultural, Physical i Raster*.

Stroną która zawiera bardzo różne dane łącznie ze zdjęciami satelitarnymi umożliwiającą ich pobranie jest strona <https://earthexplorer.usgs.gov/>.

Na tej stronie możemy znaleźć zdjęcia satelitarne o dużych rozdzielczościach (Landsat, Sentinel) oraz numeryczny model terenu dla dowolnego obszaru na Ziemi. Strona ta wymaga zalogowania się (utworzenia konta).

The screenshot shows the USGS EarthExplorer interface. At the top, there's a dark header with the USGS logo and the text "science for a changing world". Below it is a navigation bar with links for "Home", "Save Criteria", "Load Favorite", "Manage Criteria", "Search Criteria", "Data Sets", "Additional Criteria", and "Results".

The main area is divided into two sections. On the left is a sidebar titled "2. Select Your Data Set(s)" containing a list of data categories with checkboxes:

- Aerial Imagery
 - ASAS
 - Aerial Photo Mosaics
 - Aerial Photo Single Frames
 - Aircraft Scanners
 - Antarctic Flight Line Maps
 - Antarctic Single Frames
 - DOQ
 - High Resolution Orthoimagery
 - NAIP
 - NAPP
 - NHAP
 - PAO Image Gallery
 - SD NRCS Section Photos
 - SLAR
 - Space Acquired Photography
- AVHRR
- CEOS Legacy
- Commercial Satellites
- Declassified Data
- Digital Elevation

At the bottom of the sidebar, there's a link "GO TO SEARCH RESULTS".

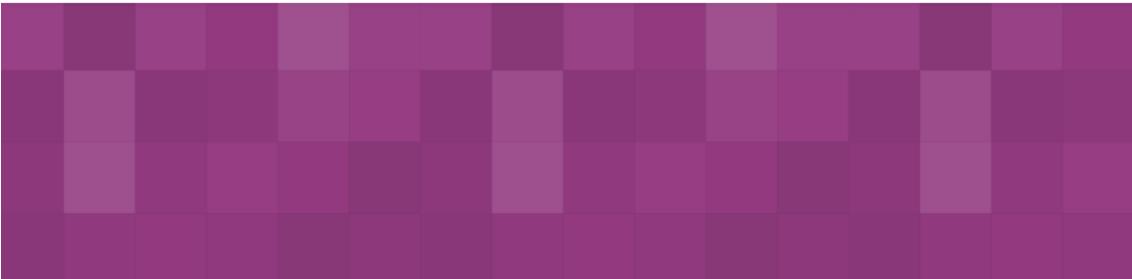
On the right is a map of a coastal region with a red circle overlaid on the water. The map includes labels for "Longyearbyen" and "Longyearbreen". Above the map, there's a section titled "Search Criteria Summary (Show)" with a "Close" button.

Wiele instytucji, organizacji lub nieformalnych grup badaczy opublikowało tematyczne zbiory danych o charakterze globalnym, które ze względu na rozdzielcość przestrenną mogą być wykorzystane do projektów lokalnych. Przykładem może być zbiór danych klimatycznych **WorldClim**: <https://www.worldclim.org/>



 WorldClim

Home



Global climate and weather data

Historical climate data
Historical monthly weather data
Future climate data

Welcome to the WorldClim data website.

Dane w formacie GeoTIFF są dostępne w różnych rozdzielcościach przestrzennych, aż do 1km. Dla różnych okresów czasu zawierają comiesięczne dane o minimalnej temperaturze, maksymalnej temperaturze, średniej temperaturze, opadach, promieniowaniu słonecznym, prędkości wiatru oraz prężności pary wodnej.

Ważną międzynarodową stroną zawierającą wiele różnorodnych danych związanych z środowiskiem morskim Europy jest serwis monitorowania środowiska morskiego Copernicus : <https://marine.copernicus.eu/pl>

   Copernicus Marine Service Usługi Możliwości Dostęp do danych



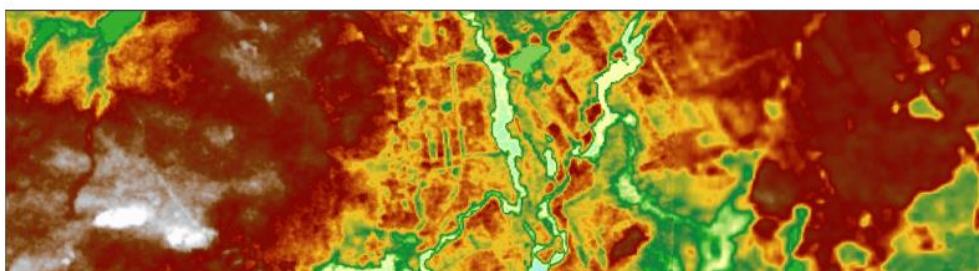
Serwis monitorowania środowiska morskiego Copernicus

Dostarczanie bezpłatnych i ogólnodostępnych danych oraz usług morskich w celu umożliwienia realizacji polityki morskiej, wspierania niebieskiego wzorstu, a także innowacji naukowych.

Pozwala on na pobranie różnorodnych danych oceanograficznych.

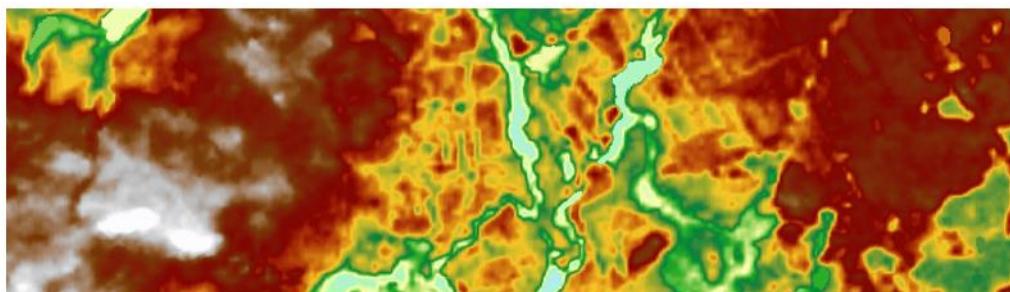
Podstawowym źródłem danych wysokościowych dla całego świata jest mapa cyfrowa **SRTM** (shuttle radar topographic mission). Mapa ta powstała w rezultacie misji wahadłowca Endeavor zorganizowanej w 2000 roku przez NASA. Od 2014 roku rezultaty tej misji w postaci map cyfrowych zostały udostępnione w rozdzielczości 1 sekundy (około 30 m) dla całego świata. Wycinki tej mapy można pobrać ze strony: <https://earthexplorer.usgs.gov/>.

Poniżej przedstawiono wycinek mapy SRTM.

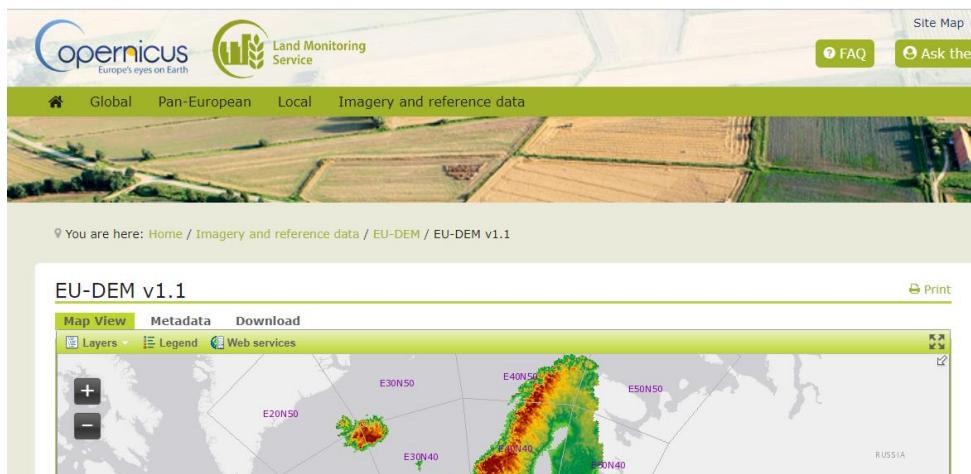


Ważną warstwą dla Europy jest raster wysokości terenu (**EU-DEM**) o rozdzielczości przestrzennej 30 m. Zrealizowany on został w ramach programu Copernicus i reprezentuje numeryczną mapę wysokości powierzchni terenu.

Poniżej przedstawiono wycinek mapy EU-DEM.



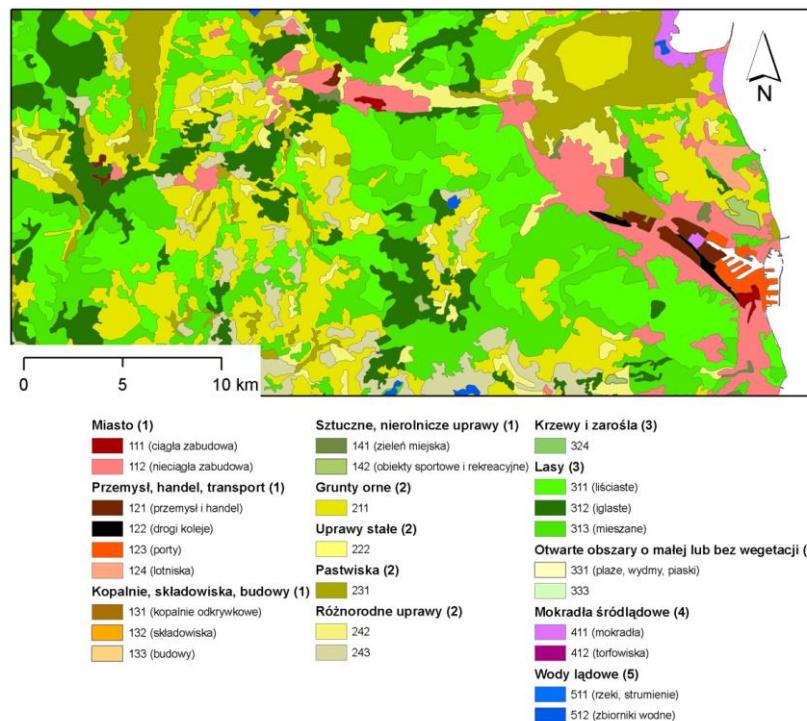
Mapy te można pobrać ze strony: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/view>



Zupełnie inną genezę mają dane **OpenStreetMap**, są to dane zawierające szereg wektorowych warstw pokrycia terenu takich jak ulice, obrysów budynków tworząc warstwy o bardzo dużej szczegółowości. Szereg danych w formacie SHP można pobrać z : <http://download.geofabrik.de/>



Chyba jednym z najczęściej używanym rodzajem danych dla Europy są wektorowe dane pokrycia terenu znane pod nazwą **Corine**. Kolejne warstwy tworzone w oparciu o zdjęcia satelitarne powstają co 6 lat od roku 2000 i odpowiadają pod względem szczegółowości skali 1:100000 (ciągłe homogeniczne obszary o powierzchni ponad 25 ha i wymiarach ponad 100 m). Pokrycie terenu sklasyfikowane jest za pomocą trzy-poziomowej klasyfikacji hierarchicznej.



Przykład mapy z danych Corine 2000 wraz z hierarchicznymi klasami występującymi na mapie przedstawiono powyżej. W nawiasach opisano identyfikatory klas pierwszego poziomu (1. Powierzchnie sztuczne; 2. Obszary rolnicze; 3. Lasy i obszary częściowo naturalne; 4. Mokradła i bagna; 5. Zbiorniki wodne). Klasy drugiego poziomu są opisane drukiem pogrubionym. Ich identyfikatory to dwie pierwsze cyfry klas trzeciego poziomu z nazwami w nawiasach. Pełna skala map Corine (w języku polskim) jest dostępna pod adresem <http://clc.gios.gov.pl/index.php/o-clc/legenda> a aktualną mapę można pobrać ze strony: <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018> jako raster lub wektor - ESRI Geodatabase lub GeoPackage (QGIS).

You are here: Home / Pan-European / CORINE Land Cover / CLC 2018

CLC 2018

Map View Metadata Download

Legend Web services

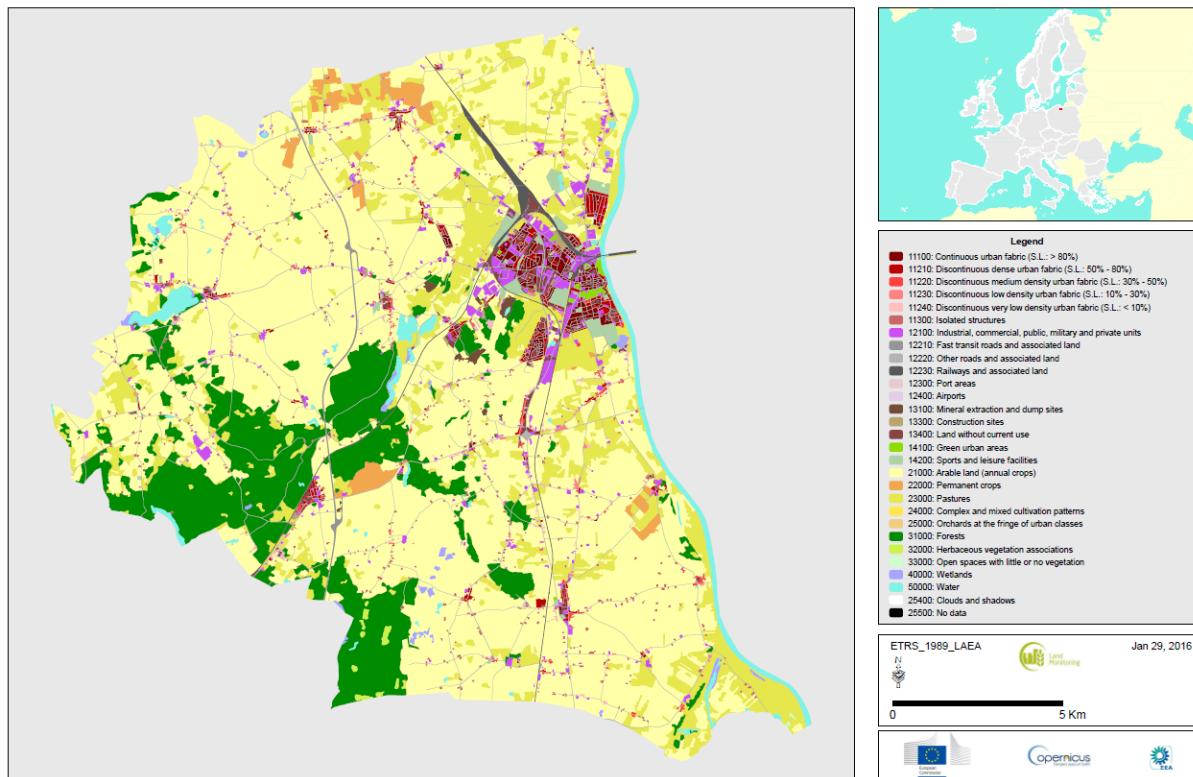
Print

User corner

- How to access our data
- Technical library
- Factsheets
- Use cases
- Looking for National projection & Expert products?

Dla obszarów metropolitalnych Europy w ramach projektu **URBAN ATLAS** (58 obszarów na terenie Polski) wykonano bardziej szczegółowe wektorowe warstwy pokrycia terenu na podstawie zdjęć satelitarnych o dużej rozdzielczości (wykorzystano głównie francuskie satelity SPOT). Podobnie jak w Corine warstwy te zbudowane są z poligonów. Mogą być pobrane z: <https://land.copernicus.eu/local/urban-atlas>

URBAN ATLAS 2012 - Tczew



Powyżej zamieszczono przykład mapy z danych URBAN ATLAS. Użyta klasyfikacja dzieli zabudowę na pięć klas w zależności od gęstości zabudowy (cztery klasy) plus klasa reprezentująca pojedyncze zabudowania. Drogi przedstawione są za pomocą dwóch klas. Natomiast warstwa rzek obejmuje tylko większe rzeki.

Dostępność polskich danych przestrzennych poprawiła się ostatnio dzięki wprowadzaniu w życie wymogów UE co do dostępu do danych publicznych. Wiele warstw do lokalnych analiz można też tworzyć z warstw globalnych lub Europejskich.

Podstawową stroną przeglądania i dostępu do danych przestrzennych jest **Geoportal** https://mapy.geoportal.gov.pl/imap/Imgp_2.html?gpm=gp0. Dane można pozyskać bezpłatnie lub odpłatnie. Dostęp do danych bezpłatnych można rozpocząć od adresu <https://www.gov.pl/web/gugik/dane-udostepniane-bez-platnie-do-pobrania-z-serwisu-wwwgeoportalgovpl>.

Dla dowolnego obszaru mogą być pobrane ortofotomapy o rozdzielczosci do 25 cm, NMT – numeryczny model terenu o rozdzielcości 1m i wiele innych danych.

Dane udostępniane bezpłatnie do pobrania z serwisu www.geoportal.gov.pl

I. Materiały centralnego zasobu geodezyjnego i kartograficznego udostępniane nieodpłatnie bez konieczności składania wniosku - zgodnie z art. 40a ust. 2 ustawy Prawo geodezyjne i kartograficzne.

1. [Państwowy Rejestr Granic i Powierzchni Jednostek Podziałów Terytorialnych Kraju \(PRG\)](#)
2. [Państwowy Rejestr Nazw Geograficznych \(PRNG\)](#)
3. [Państwowy Rejestr Podstawowych Osnów Geodezyjnych, Grawimetrycznych i Magnetycznych \(PRPOG\)](#)
4. [Baza Danych Obiektów Ogólnogeograficznych \(BDOO\)](#)
5. [Ortofotomapa \(ORTO\)](#)
6. [Numeryczny Model Terenu \(NMT\)](#)
7. [Numeryczny Model Pokrycia Terenu \(NMPT\)](#)
8. [Dane pomiarowe LIDAR](#)
9. [Baza Danych Obiektów Topograficznych \(BDOT10k\)](#)

Wiele instytucji udostępnia także dane związane ze swoją działalnością. Centralna Dyrekcja Ochrony Środowiska udostępnia pod adresem: <https://www.gov.pl/web/gdos/dostep-do-danych-geoprzestrzennych> dane w formacie wektorowym (SHP) związane z ochroną przyrody.

Do rozwiązania zadań zawartych w książce należy wykorzystać dane dydaktyczne z zasobów Pracowni GIS Instytutu Oceanografii UG.

Geobaza **Colorado2** zawiera szereg warstw dla stanu Colorado (między innymi związanych z wydobyciem surowców):

<https://cgis.oig.ug.edu.pl/ZADANIA/Colorado2.gdb.zip>

Geobaza **Model_Raster** zawiera szereg warstw przygotowanych do rozwiązywania zadań z analizy rastrowej:

https://cgis.oig.ug.edu.pl/ZADANIA/model_raster.gdb.zip

Geobaza **Ocean_ZG22** to baza danych przestrzennych zawierająca szereg warstw opisujących środowisko Zatoki Gdańskiej oraz pomiary temperatury i zasolenia prowadzone w czasie comiesięcznych rejsów w ostatnim 20-leciu XX wieku: https://cgis.oig.ug.edu.pl/geobazy/Ocean_ZG22

Geobaza **Pomorskie** zawiera wiele warstw dla województwa Pomorskiego:

https://cgis.oig.ug.edu.pl/ZADANIA/POMORSKIE_v1.gdb.zip

Geobaza **Sycylia** zawiera szereg warstw wektorowych i rastrowych dla Sycylii i okolicznych wysp (między innymi miesięczne warstwy opadów):

<https://cgis.oig.ug.edu.pl/ZADANIA/SICYLY.zip>

Geobaza **SST3** zawiera rastrowe warstwy temperatury powierzchni morza rejonu Australii oraz Azji Południowo Wschodniej :

<https://cgis.oig.ug.edu.pl/ZADANIA/SST3.gdb.zip>

Inne zbiory danych udostępnione z zasobów Pracowni GIS to:

Geobaza **GLOB2** zawiera warstwy granic krajów, oceanów i mórz oraz DEM dla całego świata z Natural Earth

<https://cgis.oig.ug.edu.pl/geobazy/GLOB2>

Geobaza **Pl_marine_waters_EG** zawiera szereg warstw opisujących dno w polskiej strefie ekonomicznej Bałtyku, pochodzących z projektu Atlasu siedlisk dna polskich obszarów morskich

https://cgis.oig.ug.edu.pl/geobazy/Pl_marine_waters

Geobaza **Polska_v3** zawiera podstawowe warstwy dla całego kraju: granice powiatów, rzeki, pokrycie terenu Corine i DEM o rozdzielczości 30m,

https://cgis.oig.ug.edu.pl/geobazy/POLSKA_v3

Geobaza **ZatokaGdanska07** zawiera podstawowe warstwy opisujące środowisko Zatoki Gdańskiej ze szczególnym uwzględnieniem jej dna (zawiera także szczegółowy opis warstw i projektu w ramach którego powstały),

<https://cgis.oig.ug.edu.pl/geobazy/ZatokaGdanska07>

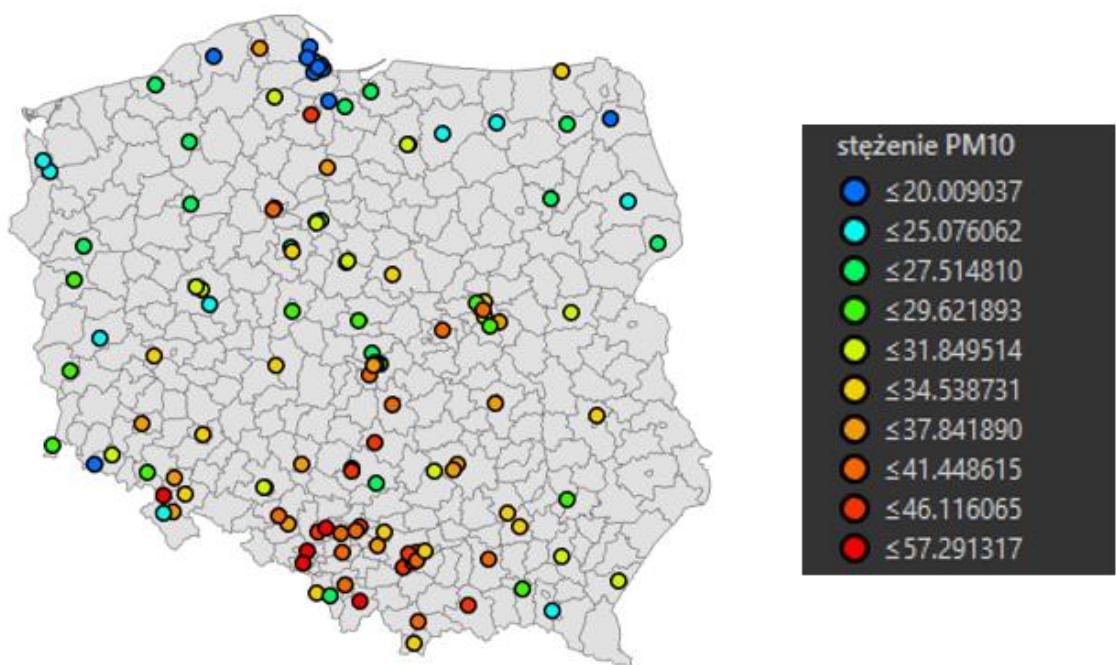
Dane **MPHP50** czyli Mapa Podziału Hydrograficznego Polski w skali 1: 50000 w formacie SHP była udostępniona publicznie, ale aktualnie jest trudna do znalezienia,

<https://cgis.oig.ug.edu.pl/geobazy/MPHP50>

4.2 Eksploracyjna analiza danych

Kiedy zaczynamy pracę z nową warstwą danych musimy się jej dokładnie przyjrzeć zanim użyjemy jej do tworzenia mapy lub wykorzystamy w procesie analizy. Dane przestrzenne dzięki swojej geometrii można oglądać co daje ogromne możliwości błyskawicznej oceny ich zawartości jak i wykrycia ewentualnych błędów. Proces ten nosi nazwę ESDA – *Exploratory Spatial Data Analysis* i wykorzystuje oprócz tradycyjnych metod analizy danych także ich wizualizację.

Unikalną cechą danych przestrzennych jest połączenie atrybutów (zmiennych) z geometrią danych.



Rysunek 4.6 Symbolizacja punktowych danych wektorowych na podstawie ciągłego atrybutu (stężenie PM10)

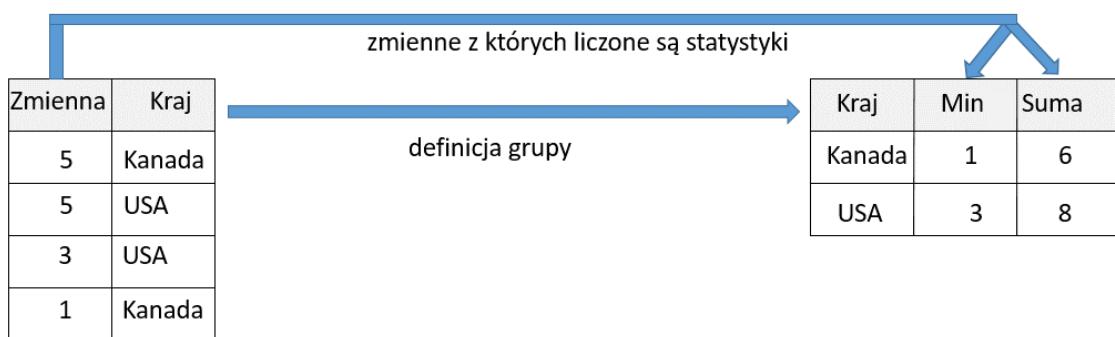
Do każdego obiektu geometrycznego (na Rysunku 4.6 są to punkty) przypisane są jego atrybuty, których wartości wykorzystywane są w procesie **symbolizacji**, polegającym na nadaniu geometrii specyficznego wyglądu. Istnieje szeroka gama potencjalnych możliwości. Na przykład geometria danych punktowych może być przedstawiana za pomocą różnych symboli (koła, kwadraty ale także obrys auta, domu czy drzewa), wielkość atrybutu może określać rozmiar symbolu, albo jego kolor. Często kolory są organizowane w palety reprezentując zmienność wartości (niebieski – małe wartości, zielony – większe wartości, czerwony największe wartości). Przedstawione na Rysunku 4.6 koła obrazują

stacje pomiarowe stężenia ważnego składnika smogu - pyłów PM10. Ich kolor reprezentuje wielkość atrybutu (wartości w kolumnie), którym są średnie stężenia w okresie obserwacji na danym punkcie. Za pomocą spojrzenia na dwie nałożone na siebie warstwy (warstwa poligonów – powiaty, stacje pomiarowe – punkty), możemy właściwie natychmiast zauważać że:

- Wartości stężenia nie są rozmiieszczone w przestrzeni losowo.
- Najniższe wartości występują na północy, a największe na południu.
- Zarówno na zachodzie, jak i wschodzie Polski występują niższe wartości stężeń niż w pasie centralnym.

Wartości atrybutu stanowią zbiór danych, który może być analizowany bez uwzględnienia geometrii.

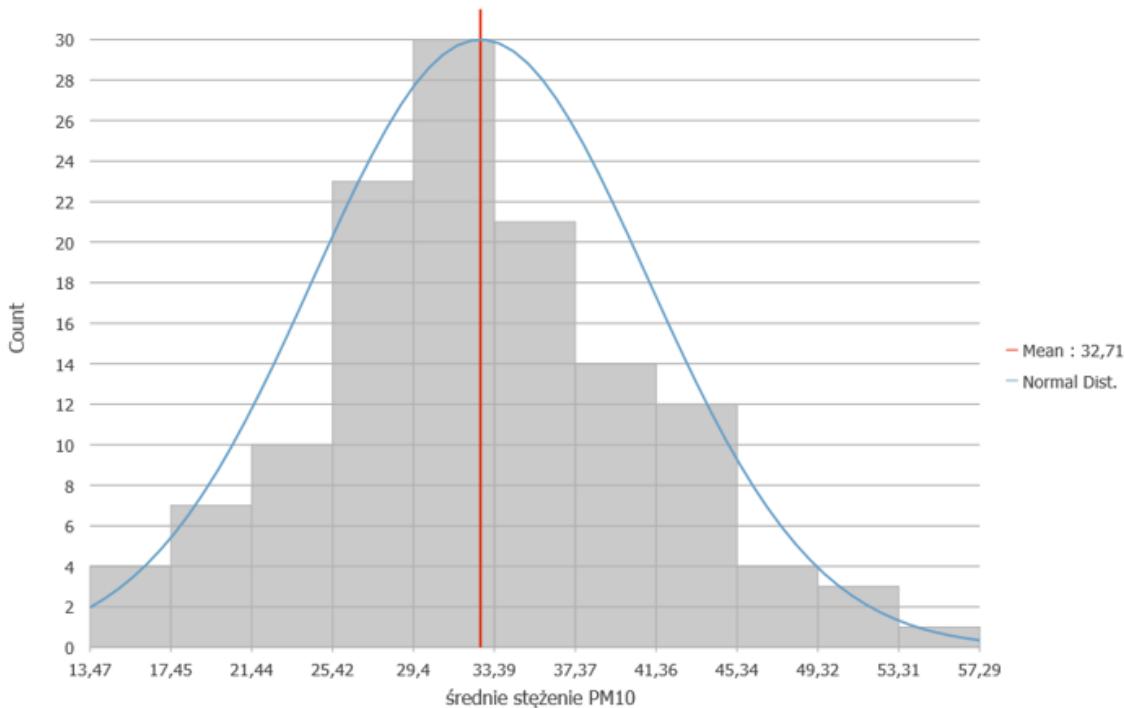
Stosowane są wtedy metody obliczeniowe i graficzne powszechnie używane przy analizie danych przedstawionych w postaci tablicy. Zbiór lub podzióbior danych atrybutowych zawarty w określonej kolumnie może być analizowany za pomocą podstawowych statystyk (np. wartość średnia, minimalna, maksymalna czy odchylenie standardowe). Istotnym sposobem analizy kolumny (pola) danych jest ich grupowanie (agregacja) w oparciu o inne pole. W rezultacie tworzona jest tabela z policzonymi statystykami oddzielnie dla każdej grupy (kategorii, przypadku). Dane z takiej tablicy mogą być następnie połączone z obiekta mi posiadającymi geometrię. Ten sposób operacji w GIS jest czasem nazywany **sumaryzacją**. Sposób jej wykonania został przedstawiony na rysunku 4.7.



Rysunek 4.7 Sumaryzacja w tablicy atrybutowej

Wartości z pola *Zmienna* są podzielone na podzbiory w zależności od wartości w polu *Kraj*. Pierwszy podzióbior dla wartości ‘Kanada’ wynosi [5, 1], a drugi dla wartości ‘USA’ [5,3]. Dla każdej wartości pola definiującego grupy zostają policzone wybrane statystyki z podzbiorów danych. Zostały one zestawione w tabeli po prawej stronie rysunku.

Jedną z podstawowych metod graficznych stosowanych w analizie danych jest histogram często uzupełniony przez podstawowe statystyki.

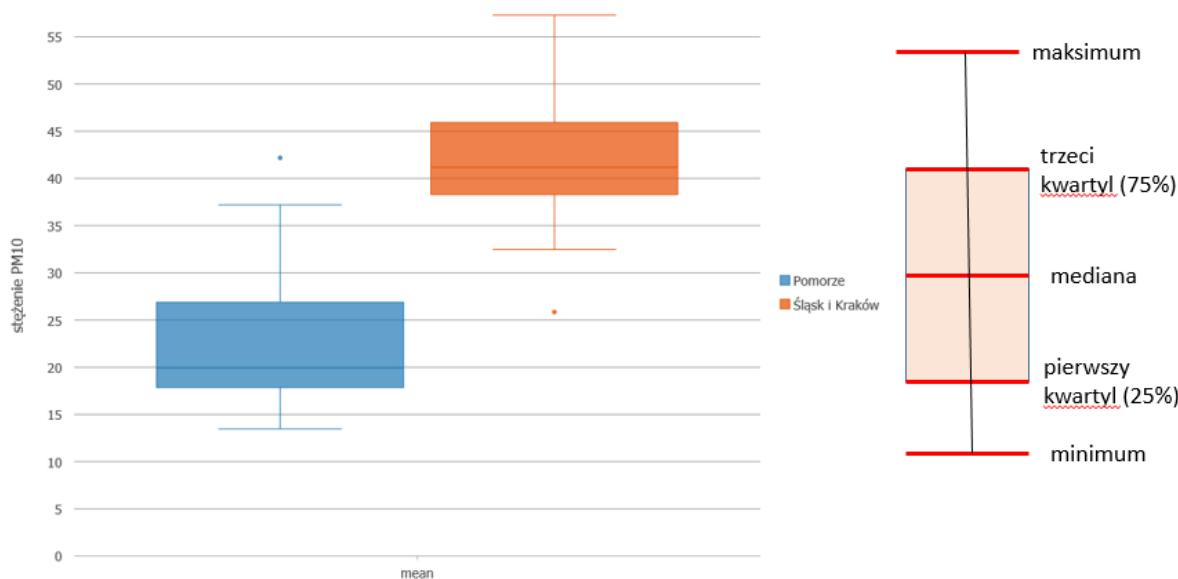


Rysunek 4.8 Histogram ciągłego atrybutu (stężenie PM10) z zaznaczoną średnią i wykresem rozkładu normalnego

W histogramie dane są zliczane w szeregu rosnących przedziałów wartości (bins). Każdy przedział jest reprezentowany następnie przez słupek histogramu, a jego wysokość odpowiada liczebności danych w przedziale. Dodatkowo możemy przedstawić podstawowe statystyki: wartość średnią, medianę, odchylenie standardowe, wartość minimalną i maksymalną. Z histogramu na Rysunku 4.8 wynika:

- Histogram ma jedno maksimum, czyli istnieje jeden przedział, od którego liczebność w innych systematycznie maleje.
- Wartość średnia wynosi 32.71, rozkład danych (reprezentowany przez kształt histogramu) jest symetryczny względem wartości średniej.
- Brak jest pojedynczych wartości wyraźnie różniących się (dużo większych lub mniejszych) od pozostałych.
- Rozkład danych jest zbliżony do rozkładu normalnego.

Drugim często używanym wykresem w analizie danych jest tzw. *box-plot*.

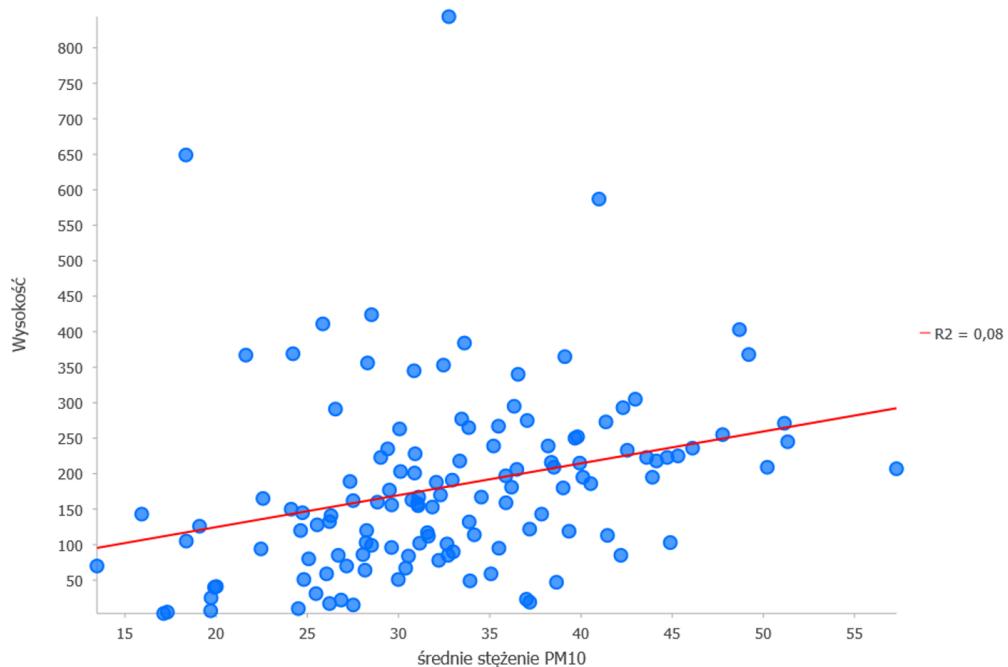


Rysunek 4.9 Wykres typu box-plot pokazuje różnice w stężeniu PM10 pomiędzy stacjami na Pomorzu a Śląskiem wraz z Krakowem

Wykres ten w wizualny sposób przedstawia szereg statystyk analizowanego zbioru danych. Składa się z pionowej linii z 5 poprzeczkami. Wartości statystyk przypisanych do tych linii znajdują się na pionowej osi wykresu. Od drugiej do trzeciej linii tworzy się prostokąt, sprawiający, że wykres jest bardziej czytelny. Krańcowe poprzeczki reprezentują wartość minimum i maksimum. Poprzeczka środkowa reprezentuje medianę (tyle samo jest wartości od niej mniejszych co większych). Druga poprzeczka pokazuje pierwszy kwartyl (25% wartości jest od niej mniejszych), a trzecia trzeci kwartyl (75% wartości jest od niej mniejszych). Na wykresie tym można zaznaczyć też wartości wyraźnie odróżniające się od innych za pomocą punktów. Często na jednym rysunku przedstawia się takie wykresy dla kilku lub kilkunastu podzbiorów analizowanych danych. Na Rysunku 4.9 analizie poddano dwa podzbiory danych, jeden dla Pomorza i drugi dla Śląska i okolic Krakowa. Wykres pokazuje nam następujące prawidłowości,

- Wartości stężenia PM10 są znacznie mniejsze na Pomorzu.
- Ponad 75% wartości na Pomorzu jest mniejsza od wartości minimalnej dla Śląska i Krakowa.
- Zakresy wartości (różnica między maksimum i minimum oraz pierwszym i trzecim kwartylem) są podobne dla obu podzbiorów.

Trzeci rodzaj wykresu często stosowany w analizie danych służy do badania relacji pomiędzy zmiennymi (atrybutami).

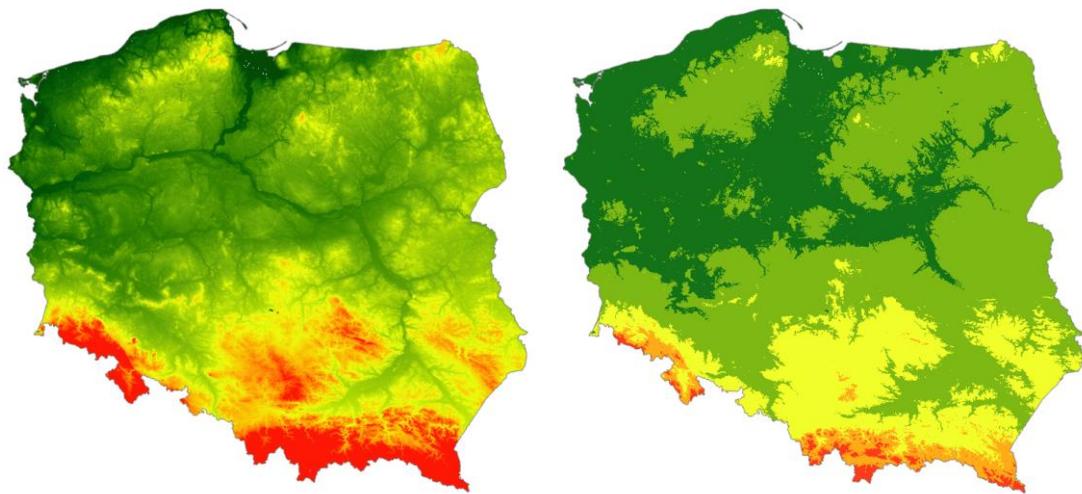


Rysunek 4.10 Wykres rozrzutu (scatter-plot) przedstawia relację między zmiennymi. Reprezentacją tej relacji jest często linia regresji liniowej oraz współczynnik R^2 opisujący zgodność punktów z tą linią

Wykres ten zwany jest wykresem rozrzutu (scatter-plot) i przedstawia chmurę punktów. Każdy punkt reprezentuje jeden rekord. A jego współrzędne odpowiadają dwóm atrybutom tego rekordu. Często dodatkowo na podstawie zbioru tych współrzędnych określana jest liniowa funkcja obrazująca związek między nimi. Obrazem tej funkcji jest linia. Często uzupełnia się ją statystyką zwaną R^2 , która pokazuje jak dobrze model (równanie liniowe $y = ax+b$) estymuje jeden atrybut za pomocą drugiego. Jeśli wszystkie punkty leżałyby na linii, istnieje doskonała relacja liniowa pomiędzy atrybutami, czyli jeden atrybut można obliczyć za pomocą drugiego atrybutu używając wyrażenia postaci $atrybut1=a * atrybut2 + b$, wtedy statystyka $R^2 = 1$. Przy całkowitym braku relacji $R^2=0$. W związku z tym jedynym wnioskiem płynącym z Rysunku 4.10 jest,

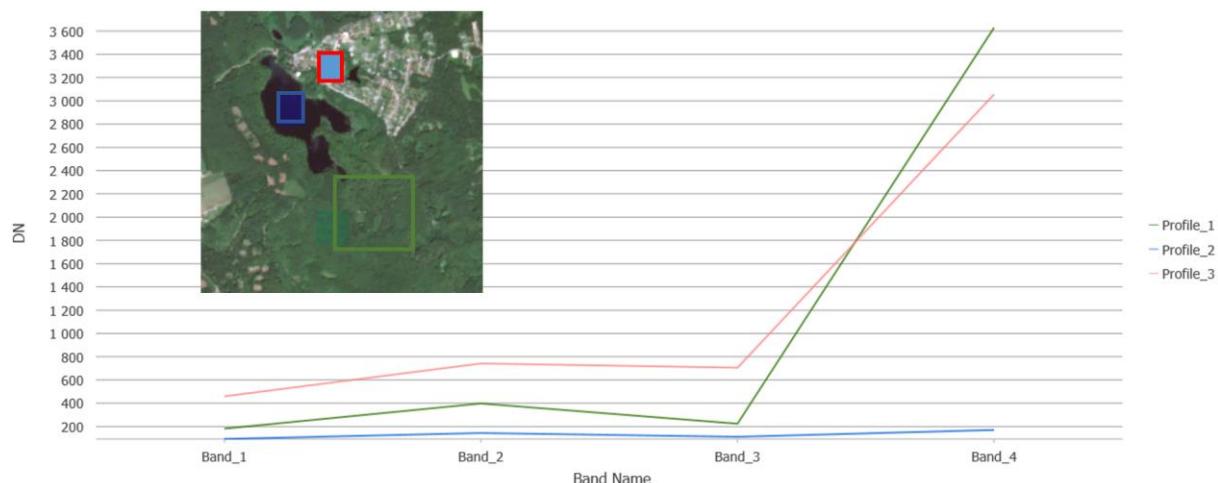
- Statystyka $R^2=0.08$ wykazuje brak relacji pomiędzy wartością stężenia i wysokością stacji pomiarowej.

Podobne metody wstępnej analizy danych stosujemy dla danych rastrowych. Istnieją dwa podstawowe sposoby symbolizacji danych rastrowych, które są używane do ich wizualizacji. Pierwsza wykorzystuje całe spektrum palety kolorów do rozciągnięcia wartości od wartości minimalnej do maksymalnej (Stretch), druga przypisuje kolory do zdefiniowanych zakresów wartości (Classify).



Rysunek 4.11 Symbolizacja danych rastrowych. Po lewej zastosowanie metody Stretch, po prawej metody Classify.

Charakterystyka spektralna obiektów zdjęć satelitarnych jest analizowana często za pomocą wykresów profili spektralnych, które przedstawiają statystyki wartości zdjęcia dla poszczególnych kanałów.



Rysunek 4.12 Profile spektralne (średnie) dla trzech typów pokrycia terenu (las, woda i zabudowania)

Pytania:

Dane: Ocean_ZG22

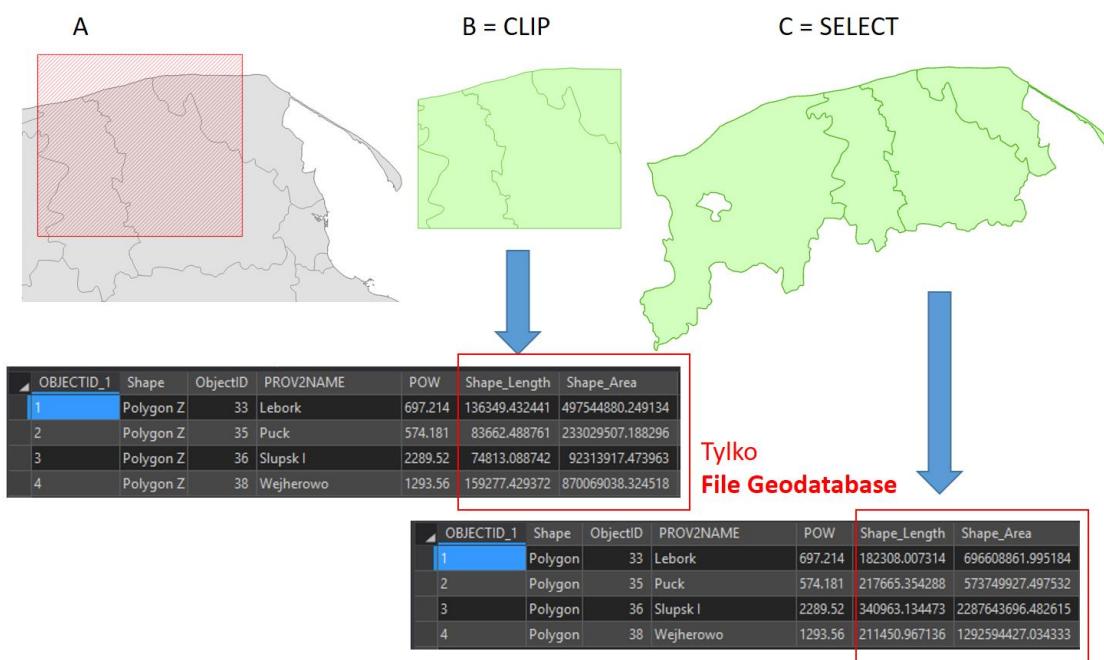
1. Wykonaj warstwę punktów pomiarowych na Zatoce Gdańskiej (z temperatura_zasolenie1981_1999). W warstwie powinny znajdować się punkty o głębokości maksymalnej większej od 0 na których pomiary wykonano co najmniej w trakcie 15 rejsów (czyli co najmniej 15 razy wykonano pomiar na 0 m). Warstwa powinna zawierać pola: Nr_punktu, Liczba_rejsów, Max_głębokość. Na jakich punktach wykonano najwięcej pomiarów (rejsów)? Wykonaj histogram maksymalnej głębokości punktów, ile posiada dominant (lokalnych maximów).
2. Porównaj temperaturę w powierzchniowej warstwie wody (0-5m) na obszarze Zatoki Puckiej Wewnętrznej (ZPW) i Zewnętrznej (ZPZ) w okresie zimy (styczeń – marzec) i lata (czerwiec – sierpień). Wykorzystaj wykres typu *Box Plot*.
3. Jak zmieniało się średnie zasolenie w warstwie poniżej 50 metrów w całej Zatoce Gdańskiej w czasie pomiarów. Wykorzystaj wykres liniowy.
4. Wykonaj mapę na której stacje pomiarowe będą symbolizowały wartości maksymalnego zasolenia zmierzonego na powierzchni (0-1m).
5. Jaka jest mediana zasolenia na powierzchni wody w Zatoce Gdańskiej (ze wszystkich punktów oraz z punktów odległych od brzegu co najmniej 2 kilometry).
6. Na podstawie analizy wykresu rozrzutu zależności zasolenia od głębokości wyznacz głębokości graniczne dla warstwy wody powierzchniowej i przydennej.
7. Wykonaj wykresy średnich letnich (czerwiec – sierpień) profili temperatury i zasolenia na punkcie 4. Wykorzystaj wykres liniowy.
8. Wykonaj wykres obrazujący możliwie najlepiej jak zmienia się rodzaj osadu w zależności od głębokości.
9. Utwórz warstwę punktów pomiarowych z polem zawierającym średnie zasolenie w warstwie przydennej ($\max \pm 5$ m). Nie bierz pod uwagę punktów w których było mniej niż 15 pomiarów. Wykonaj wykres zależności zasolenia od głębokości.

4.3 Podstawowe operacje przygotowania danych

Proces przygotowania danych zwykle zajmuje najwięcej czasu (nawet do 90%). Zwykle po zgromadzeniu wszystkich danych potrzebnych w naszym projekcie następuje ich przetworzenie do najbardziej dogodnej postaci.

4.3.1 Tworzenie podzbiorów danych (CLIP, SELECT)

Zazwyczaj pozyskane dane obejmują dużo większy przestrzennie obszar niż potrzebny w projekcie. Obszar projektu może być zdefiniowany w sposób naturalny (np. poligon zlewni), administracyjny (np. poligon gminy) lub za pomocą tzw. obszaru zainteresowania, który zwykle tworzy prostokąt (zwyczajowo nazywany **AOI** – Area of Interest). Utworzenie podzbioru przestrzennego danych, może odbywać się za pomocą **wycinania (CLIP)** geometrii za pomocą wieloboku definiującego obszar projektu **albo selekcji (SELECT)** obiektów.



Rysunek 4.12 Porównanie działania operacji CLIP i SELECT. CLIP wycina geometrię granicami poligona definiującego obszar zainteresowania. Tylko obiekty całkowicie znajdujące się wewnątrz i części obiektów przeciętych granicą, które znajdują się wewnątrz tworzą podzbiór. Każdy obiekt (cały, albo wycięta część) zachowuje oryginalne atrybuty. Aktualizacja atrybutów geometrii obiektów (długości, obwodu, powierzchni) następuje tylko w strukturze danych zwanych File Geodatabase. SELECT pozwala wyłącznie na wybranie podzbioru całych obiektów wraz z ich atrybutami

W przypadku CLIP mamy do czynienia z operacją, która tworzy nowy zbiór danych. Proces selekcji (SELECT) powoduje wyłącznie zaznaczenie części rekordów. Aby został utworzony nowe zbiór danych muszą być one wyeksportowane lub zapisane. Podczas ich tworzenia można dokonać także zmiany ich układu współrzędnych. W praktyce tworząc zestaw danych dla danego projektu należy dążyć do wspólnego układu współrzędnych.

Operacja CLIP jest także używana w danych rastrowych do ekstrakcji zawartości rastra dla pewnej maski, która może być poligonom (są różne narzędzia dla danych wektorowych i rastrowych). Dzięki temu możemy używać tego samego poligonu obszaru badań do tworzenia podzbioru danych rastrowych i wektorowych.



Rysunek 4.13 CLIP rastra pozwala na wycięcie dowolnego rastra za pomocą poligonu (wycięty raster jest zawsze prostokątem)

Proces określania podzbiorów rekordów za pomocą selekcji jest bardzo istotną operacją w GIS. Jeżeli wykonamy dowolną operację na zbiorze rekordów z których część jest wybrana (zaznaczona) za pomocą selekcji **to ta operacja zostanie wykonana tylko na zaznaczonych rekordach**.

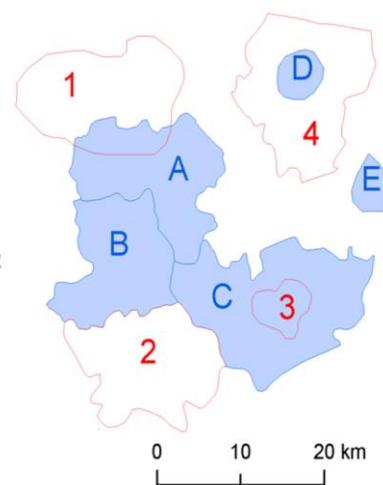
Istnieją dwie metody przeprowadzania selekcji, pierwsza na podstawie atrybutów danej warstwy i druga na podstawie relacji przestrzennej pomiędzy dwoma warstwami (może to też być relacja warstwy z samą sobą). Podstawą przeprowadzania selekcji atrybutowej jest tworzenie zapytań za pomocą języka **SQL**. SQL jest specjalnym językiem formowania zapytań do tabel baz danych. Jego użycie w GIS jest uproszczone, może także być wykonywane w przyjaznym interfejsie. Podstawowe wyrażenia SQL będą miały postać przedstawioną

poniżej i w zależności czy wynikiem operacji jest PRAWDA czy FAŁSZ zostaną wybrane (zaznaczone) lub nie.

Zadanie wyboru rekordów	Wyrażenie SQL
Wartość (w kolumnie POLE) jest równa 2	POLE=2
Wartość (w kolumnie POLE) jest różna od 2	POLE<>2
Wartość (w kolumnie POLE) jest równa lub większa od 2	POLE>=2
Wartość (w kolumnie POLE) zawiera (1,2,7)	POLE IN (1,2,7)
Wartość (w kolumnie POLE) nie zawiera wartości (1,2,7)	POLE NOT IN (1,2,7)
Wartość (w kolumnie POLE) jest równa NULL	POLE IS NULL
Tekst (w kolumnie POLE) zaczyna się od GD.....	POLE LIKE 'GD%'
Wartość (w kolumnie POLE) jest między 3 i 8	POLE >=3 AND POLE <=8
Wartość (w kolumnie POLE) jest 3 LUB jest większe od 8	POLE=3 OR POLE > 8

Dodatkowo istnieje możliwość budowania sekwencji operacji wyboru rekordów za pomocą operacji selekcji przy wybranych już rekordach. Można dodać do nich nowe rekordy (za pomocą nowego wyrażenia, które będzie dotyczyło tylko niewybranych jeszcze rekordów). Usunąć rekordy spełniające pewien warunek (za pomocą nowego wyrażenia, które będzie dotyczyło tylko wybranych już rekordów). Można także wybrać rekordy z podzbioru już wybranych rekordów. Daje to możliwość łączenia selekcji atrybutowej z drugim jej rodzajem, selekcją w oparciu o relacje przestrzenne.

intersect	obiekty A, B, C, D mają wspólny element z obiektami CYFR
are within a distance of	obiekty A, B, C, D, E są w zasięgu 3 km od obiektów CYFR
completely contain	obiekt C całkowicie zawiera obiekt(y) CYFR obiekt D jest całkowicie zawarty w obiekcie CYFR
share a line segment with	obiekty C, D mają swój środek wewnętrz obiektów CYFR
are within	obiekty B, C mają wspólny odcinek (linii obrys) z obiektami CYFR obiekty B, C stykają się z granicą obiektów CYFR obiekty A jest przecinany przez obrys obiektu(ów) CYFR obiekt C zawiera obiekt(y) CYFR obiekt D jest zawarty w obiekcie CYFR



Rysunek 4.14 Przykłady relacji przestrzennych pomiędzy obiektami

Selekcja za pomocą relacji przestrzennej pozwala na wybór geometrii (rekordów) jednego zbioru danych (warstwy), które pozostają w określonej relacji z

geometrią obiektów drugiego zbioru lub podzbioru (wybranych rekordów) drugiej warstwy .

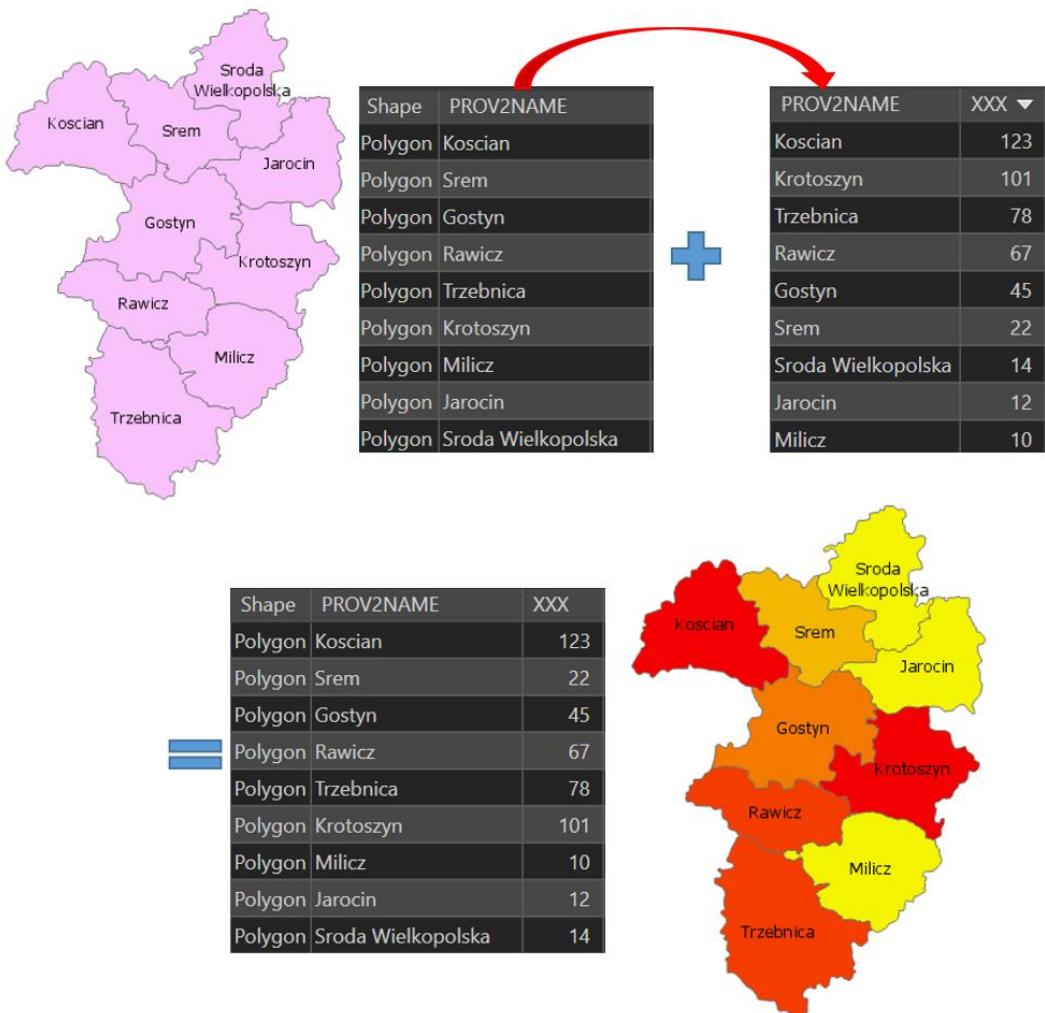
Po utworzeniu podzbioru rekordów należy uporządkować pola (kolumny) tablicy atrybutowej. Może to polegać na usunięciu niepotrzebnych pól lub utworzeniu nowych. Może także zaistnieć potrzeba zmiany nazwy lub typu danych pola. Najprościej jest utworzyć wtedy nowe pole odpowiedniego typu (z nową nazwą), przenieść zawartość ze starego pola do nowego i na końcu usunąć stare pole. GIS posiada do tego zadania odpowiednio przygotowane narzędzia. Tego typu operacja wiąże się z przypisywaniem lub obliczaniem wartości w polu, to zagadnienie zostanie omówione w jednym z kolejnych rozdziałów.

4.3.2 Łączenie danych (JOIN, MERGE, MOSAIC, COMPOSIT)

W trakcie przygotowywania danych wektorowych możliwe są dwa często spotykane scenariusze.

Pierwszy polega na dołączaniu pola lub pól atrybutowych z innego zbioru danych. Może to być zbiór danych przestrzennych albo tablica danych atrybutowych utworzona z danych tekstowych. Na Rysunku 4.15 przedstawiona jest warstwa powiatów (poligonów) z polem zawierającym ich nazwy oraz tablica danych tekstowych z dwiema kolumnami, nazwą powiatu i potrzebną zmienną liczbową. W GIS istnieje narzędzie **JOIN**, które na podstawie relacji między określonymi polami dwóch warstw (mogą one mieć różne nazwy) dokonuje przyłączenia dowolnego pola. W przykładzie poniżej taką rolę pełni pole z nazwą powiatu. Do warstwy powiatów dołączone zostaje pole XXX na podstawie zgodności nazw w obu polach (PROV2NAME).

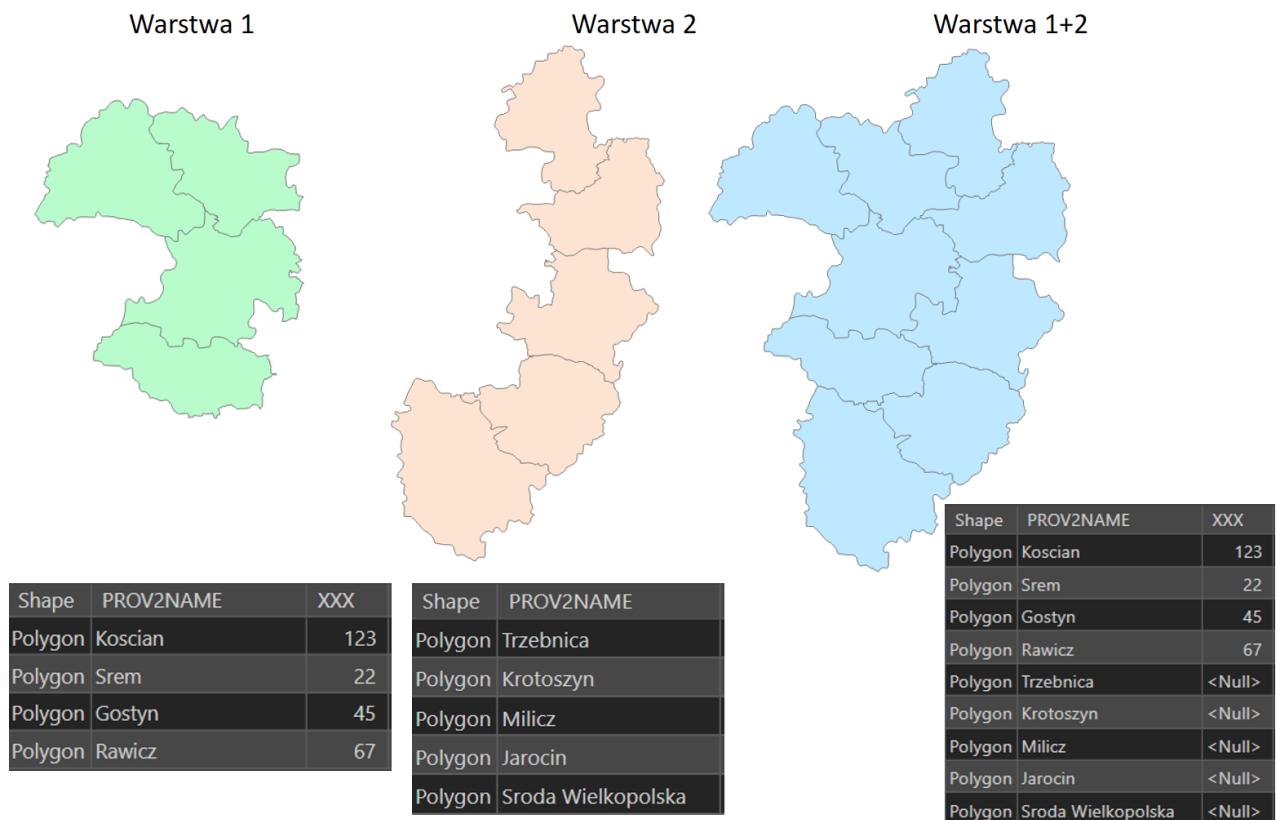
JOIN daje możliwość transferu danych w obie strony pomiędzy danymi przestrzennymi i wyłącznie atrybutowymi. Przykładowo możemy atrybutowe dane zawierające nazwy miast i liczbę ludności wzbogacić o atrybut odległości tych miast od najbliższego miasta wojewódzkiego wykorzystując warstwę przestrzenną miast na podstawie której przypiszemy każdemu miastu wymaganą odległość. Następnie do tablicy atrybutowej danych miasto i liczba ludności, dodamy pole z odlegością z tablicy atrybutowej warstwy miast, posługując się *kluczem* nazwy miasta.



Rysunek 4.15 Przyłączanie pola z jednej tablicy do drugiej za pomocą JOIN. Konieczne jest istnienie dwóch pól, które zawierają pasujące do siebie ‘klucze’, według których wartości trafiają do odpowiednich rekordów (w tym przypadku są to pola zawierające nazwę powiatów)

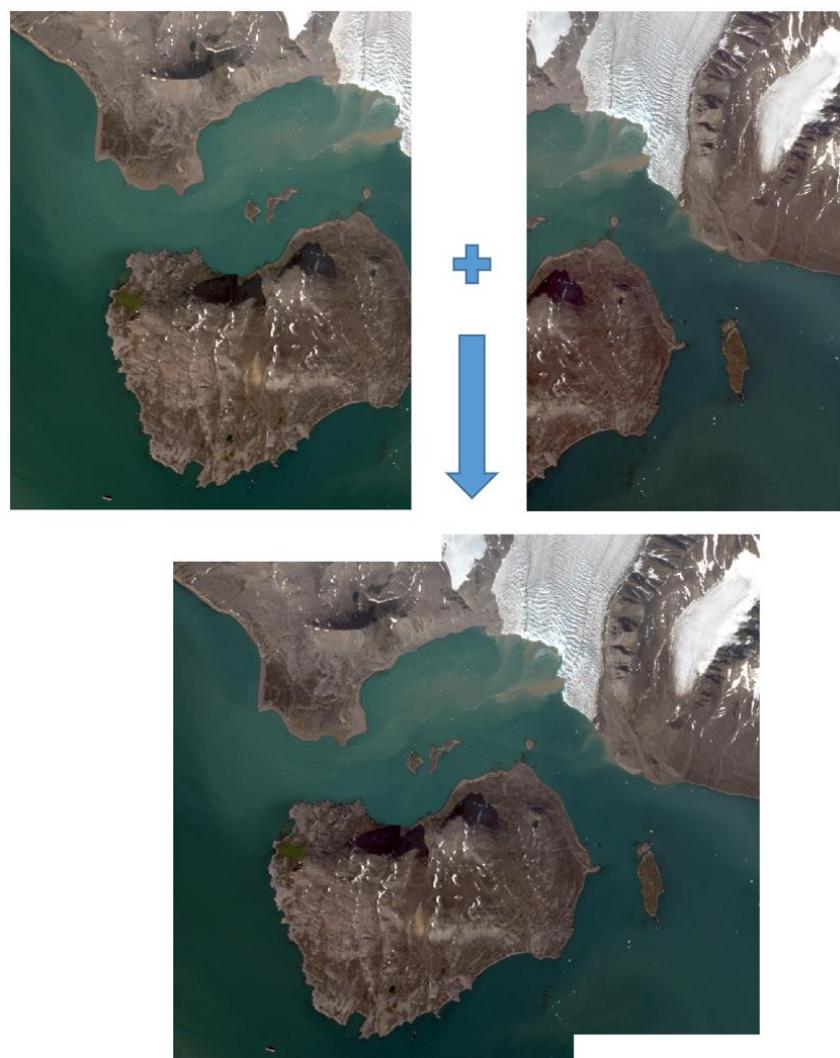
Drugi scenariusz polega na łączeniu kilku warstw w jedną. Do tego celu w GIS służy narzędzie **Merge**. Łączy ono warstwy wektorowe tego samego typu (punkty, linie lub poligony) w jedną warstwę. Nowa warstwa będzie zawierać wszystkie obiekty łączonych warstw. Atrybuty zostają przypisane do istniejących w obu warstwach pól. Brak wartości lub zero zostanie przypisane do obiektów, które nie posiadają danego pola. Niektóre implementacje funkcji Merge umożliwiają mapowanie pól. Można utworzyć nowe pole, w którym znajdą się atrybuty z różniącymi się nazwami pól. Na Rysunku 4.16 przedstawiono działanie Merge dla dwóch warstw poligonów. W warstwie wynikowej są wszystkie poligony obu warstw. Pole *PROV2NAME* występuje w obu warstwach. W warstwie wynikowej pojawia się z atrybutami pochodząymi z obu warstw. Pole

XXX występuje tylko w jednej warstwie. Obiekty z warstwy drugiej posiadają w tym polu w warstwie wynikowej wartość <Null> co oznacza brak wartości.



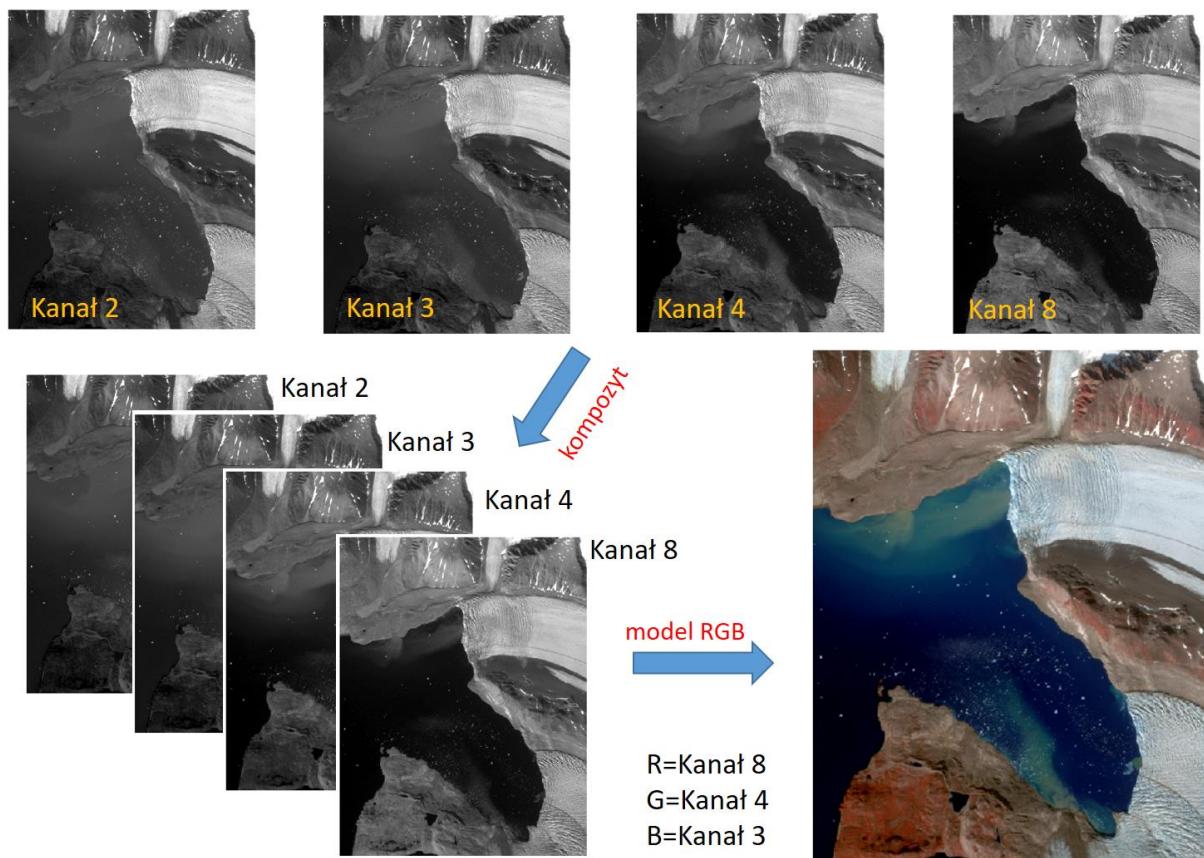
Rysunek 4.16 Łączenie kilku warstw za pomocą Merge

Przygotowywanie danych rastrowych ze zdjęć satelitarnych wiąże się zwykle z użyciem dwóch narzędzi. Pierwszym z nich jest **Mosaic** pozwalające na łączenie ze sobą paru scen w jedną. Często mamy do czynienia z sytuacją, w której nasz obszar zainteresowania obejmuje dwie pobrane sceny. Na Rysunku 4.17 mamy dwie oddzielne sceny (zdjęcia satelitarne), które w sumie obejmują wyspę, obiekt naszego zainteresowania. Potrzebujemy połączyć te zdjęcia, aby uzyskać jedno obejmujące całą wyspę. Narzędzie Mosaic służy do tego celu. Połączenie odbywa się na podstawie dopasowania rastrow korzystając z ich referencji przestrzennej. Narzędzie to może być stosowane zarówno do rastrow jedno jak i wielokanałowych. Definiuje się także z którego rastrow zostaną pobrane piksele dla wspólnego obszaru.



Rysunek 4.17 Łączenie zdjęć (rastrów) za pomocą MOSAIC

Dane satelitarne są najczęściej pozyskiwane w postaci oddzielnych kanałów danej sceny (zdjęcia satelitarnego). Do analizy eksploracyjnej takiego zdjęcia, jako wizualizacji, wykorzystania w produkcie kartograficznym a także w metodach analizy obrazu potrzebujemy połączenia tych kanałów w kompozyt. Do tego celu służy narzędzie **Composit**. W zależności od rodzaju narzędzia pozwala on na tworzenie wielokanałowego rastra składającego się z trzech lub większej liczby kanałów. Jeżeli zdjęcie składa się z kanałów o różnej rozdzielczości przestrzennej powinny być one sprowadzone wcześniej do jednej rozdzielczości.

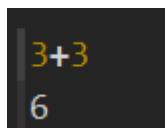


Rysunek 4.18 Łączenie rastrów jedno-kanałowych ze zdjęcia Sentinel 2 w kompozyt cztero-kanałowy za pomocą narzędzia COMPOSIT i jego przykładowa symbolizacja

4.4 Obliczenia w tabeli atrybutowej (język programowania Python)

Obliczenia w polach tablicy atrybutowej są realizowane w GIS za pomocą wyrażeń i skryptów (krótkich programów) w języku programowania Python. Python jest obecnie podstawowym językiem programowania służącym do przetwarzania danych (data science). Jego niewątpliwą zaletą jest prostota, co ułatwia jego opanowanie. Język programowania to język komunikacji użytkownika z komputerem. Tak jak każdy inny język używa się w nim słów i odpowiedniej składni, z których buduje się instrukcje, które mają być wykonane przez komputer. Python realizuje instrukcję po instrukcji, zamieniając je najpierw do postaci zrozumianej przez komputer (na tzw. język maszynowy), a następnie wykonując. Do pisania programów używa się specjalnego programu pełniącego rolę środowiska programowania zwanego IDE (od Interactive Development Environment). GIS zawiera szereg specjalnie przygotowanych miejsc gdzie będziemy mogli tworzyć i wykonywać programy napisane w Pythonie.

Najprostszym rodzajem instrukcji jest **wyrażenie** (expression). Przykładem takiego wyrażenia jest:

A screenshot of a dark-themed Python interpreter window. It shows the input line with the expression "3+3" and the output line with the result "6".

```
3+3
6
```

Podobnie jak w kalkulatorze wyrażenie **3+3** zostanie sprowadzone do pojedynczej wartości **6**. Wyrażenia składają się z **wartości** (3,3) i **operatorów** (+) i mogą być zawsze sprowadzone do pojedynczej wartości. W związku z tym wyrażenia mogą być używane w każdym miejscu naszego programu w którym oczekiwania jest wartość.

Najczęściej będziemy używali operatorów matematycznych. W tabeli zestawiono je w kolejności wykonywania.

Operator	Operacja	Przykład	Wynik
**	potęgowanie	2 ** 3	8
%	reszta z dzielenia	22 % 8	6
//	dzielenie całkowite	22 // 8	2
/	dzielenie	22 / 8	2.75
*	mnożenie	3 * 4	12
-	odejmowanie	7 - 5	2
+	dodawanie	2 + 3	5

Kolejność wykonywania może być zmieniona za pomocą wykorzystania nawiasów.

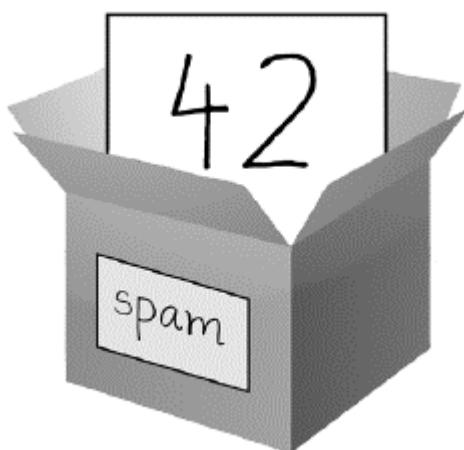
Każda wartość należy do określonego **typu danych**. Podstawowe typy danych w Pythonie to:

Typ danych	Przykład
Całkowity (integer)	-12, 0, 3, 234, 17
Rzeczywisty (float)	-1.30, 0.0, 12.01, 13.8
Tekstowy (string)	'klasa', 'A', 'koniec programu nr. 3',

Operatory mogą mieć różne działanie w zależności od typu danych. Przykład poniżej pokazuje jak działa operator dodawania w przypadku dwóch stringów (tekstów). Wyrażenie **str(3)** dokonuje **konwersji** (zamiany) liczby całkowitej **3** na tekst '3'. A operator + służy w tym przypadku do łączenia tekstów (concatenate).

```
'Punkt '+str(3)  
'Punkt 3'
```

Kolejnym podstawowym pojęciem w programowaniu jest **zmienna**. Można ją rozumieć jako pudełko z nazwą do którego możemy włożyć pojedynczą wartość, a następnie wywoływać tą wartość za pomocą nazwy zmiennej (pudełka).



Proces przypisania wartości do zmiennej odbywa się za pomocą **instrukcji przypisania** np. `spam = 42`. Oznacza to, że aktualnie zmienna *spam* ma wartość całkowitą 42.

Wykorzystanie zmiennej i instrukcji przypisania może mieć postać.

```
spam=42
spam=spam+10
spam+0
52
```

Najpierw przypisano 42 do zmiennej spam. Następnie do zmiennej *spam* przypisano wartość otrzymaną z dodania aktualnej wartości jaką posiada zmienna *spam* i liczby 10. Do nowej wartości *spam* dodano zero. Istnieją pewne restrykcje co do tworzenia nazw zmiennych. Przyjęto, że ich nazwy zaczynają się literą (bez polskich), a po nich mogą następować litery, cyfry oraz znak podkreślnika `_`. Nie można w nazwach używać takich znaków jak: `#`, `&`, `@` i `%` oraz spacji. Python jest językiem „case sensitive” oznacza to, że Temperatura i temperatura to dwie różne nazwy. Istnieje także pewien zbiór słów zarezerwowanych, które nie mogą być używane jako nazwy zmiennych.

Klasycznym pierwszym programem jest wydrukowanie tekstu „Witaj świecie!”. W Pythonie taki program będzie miał postać.

```
print('Witaj świecie!')
Witaj świecie!
```

Print to funkcja wykonująca pewną operację na argumencie znajdującym się wewnętrz nawiasów. Wewnątrz nawiasów znajduje się tekst (string) „Witaj świecie!”. Funkcja wydrukuje go na ekranie. Język Python zawiera wiele różnych funkcji. Na przykład funkcje **int()**, **float()**, **str()** służą do zamiany typu danych, a funkcja **round(x,2)** zaokrąglą wartość x do 2 miejsc po kropce.

Program (czasem nazywany skryptem) składać się może z szeregu linii. Kolejność ich wykonywania określają bloki instrukcji opisujące tzw. **flow control**. Istotnym ich elementem są zmienne typu logicznego (Boolowskie) przyjmujące tylko dwie możliwe wartości PRAWDĘ (True) albo FAŁSZ (False). Wartości logiczne są wynikiem działania operatorów porównania oraz operatorów logicznych:

Operator	Znaczenie / wynik
<code>==</code>	równa się
<code>!=</code>	różne
<code><</code>	mniejsze
<code>></code>	większe
<code><=</code>	mniejsze lub równe
<code>>=</code>	większe lub równe
<code>True and True</code>	True

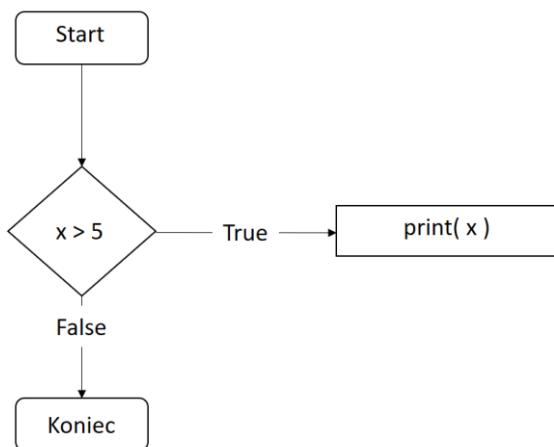
True and False	False
False and True	False
False and False	False
True or True	True
True or False	True
False or True	True
False or False	False

W naszych zastosowaniach podstawową konstrukcją określającą *flow control* będzie blok instrukcji oparty o **if**. W najprostszej wersji przyjmie on postać.

```
x=7
if x > 5:
    print(x)
7
```

Zmiennej x została przypisana wartość 7. Instrukcja **if** zawiera warunek ($x > 5$) jeśli warunek jest spełniony (True) wykonane zostaną linie znajdujące się pod **if** (blok linii zaznaczony wcięciem). W tym przykładzie warunek jest spełniony i na ekranie zostanie wypisana wartość x.

Realizacja kodu (flow control) może być przedstawiona w postaci graficznej jako:



Rysunek 4.19 Realizacja bloku instrukcji IF.

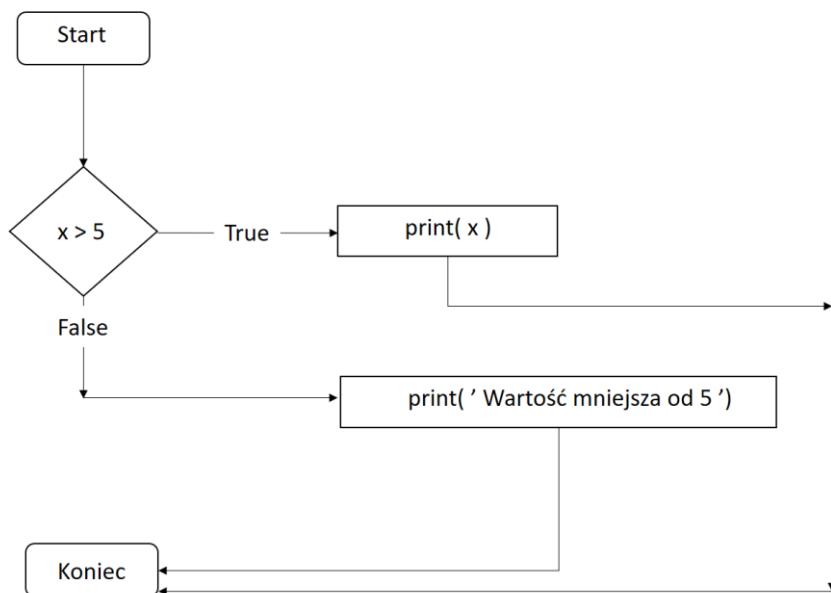
Rozwinięciem tej konstrukcji jest zastosowanie **else**, które oznacza w *przeciwnym razie*.

```

x=4
if x > 5:
    print(x)
else:
    print('Wartość mniejsza od 5')
Wartość mniejsza od 5

```

Realizacja bloku instrukcji if-else można przedstawić graficznie jako:



Rysunek 4.20 Realizacja bloku instrukcji IF-ELSE.

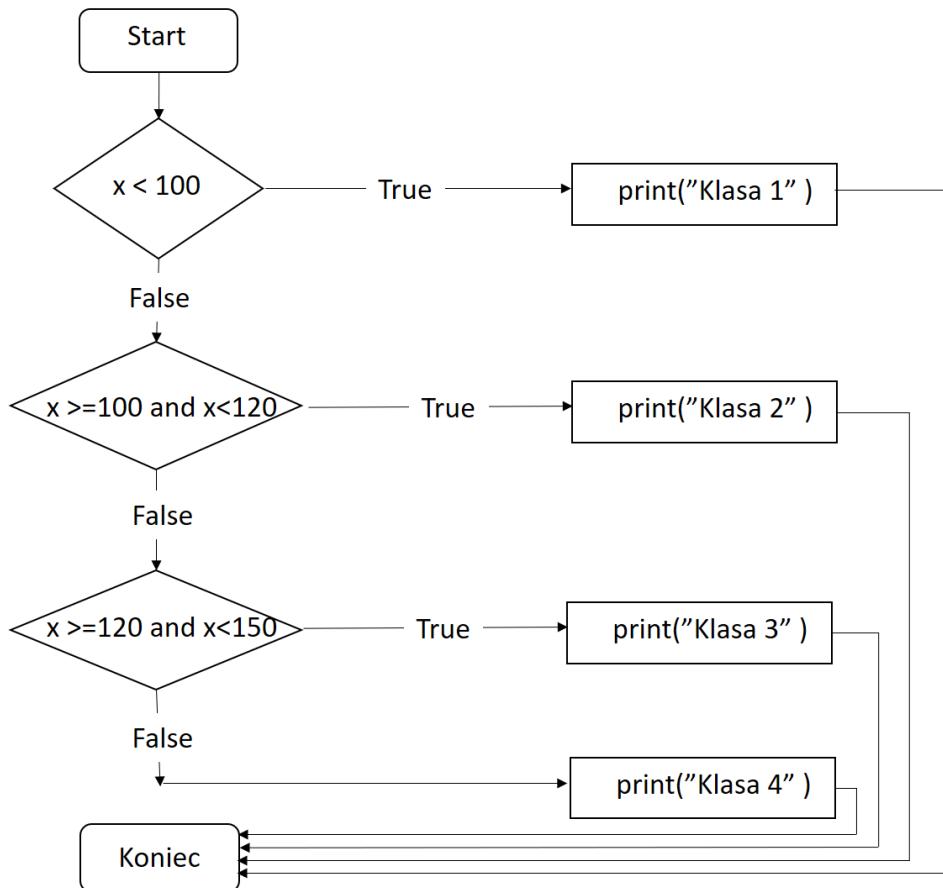
Kolejnym rozwinięciem tej instrukcji jest użycie **elif**.

```

x=122
if x < 100:
    print('Klasa 1')
elif x >= 100 and x < 120:
    print('Klasa 2')
elif x>= 120 and x <150:
    print('Klasa 3')
else:
    print('Klasa 4')
Klasa 3

```

Elif pozwala budować złożoną strukturę warunkową, której istotną zaletą jest zatrzymanie działania po napotkaniu prawdy. Graficzne przedstawienie tej konstrukcji będzie miało postać:



Rysunek 4.21 Realizacja bloku instrukcji IF-ELIF-ELSE.

Wszystkie programy w Pythonie mogą korzystać z zestawu podstawowych funkcji (built-in functions). Oprócz tego Python posiada szereg standardowych bibliotek (standard library), które zawierają najczęściej potrzebne funkcje. Na przykład biblioteka **math** zawiera funkcje matematyczne, a biblioteka **random** generatory liczb losowych. Są one często używane w GIS. Użycie funkcji zawartej w bibliotece wymaga zimportowania tej biblioteki, wykonuje się to za pomocą instrukcji:

Import <nazwa_biblioteki>

Ten sposób importu biblioteki wymaga poprzedzenia nazwy funkcji nazwą biblioteki i kropką:

<nazwa_biblioteki>.<nazwa_funkcji>()

Podstawowe funkcje biblioteki **math** zostały przedstawione poniżej:

Funkcje modułu math	Opis funkcji
ceil(x)	najmniejsza liczba całkowita nie mniejsza od x
floor(x)	największa liczba całkowita nie mniejsza od x
exp(x)	e do potęgi x
log(x)	logarytm naturalny z x
log10(x)	logarytm dziesiętny z x
pow(x,y)	wartość $x^{**}y$
sqrt(x)	pierwiastek kwadratowy z x
pi	stała 3.141599265359
e	stała 2.718281882846
sin(x)	sinus x (x w radianach)
cos(x)	cosinus x (x w radianach)
tan(x)	tangens z (w radianach)
acos(y)	zwraca arcus cosinus (w radianach)
atan(y)	zwraca arcus tangens (w radianach)
degrees(x)	zamienia radiany na stopnie
radians(x)	zamienia stopnie na radiany

Przykładowe zastosowanie funkcji matematycznych może wyglądać w następujący sposób. Najpierw wydrukowana zostanie wartość pierwiastka ze 100. Do zmiennej *fi* przypisano kąt w stopniach. Za pomocą funkcji *radians* zamieniono go na radiany i wydrukowano linię zawierającą wartość kąta w stopniach i jego cosinus zaokrąglony do czterech miejsc po kropce.

```
8 import math
9
10 print(math.sqrt(100))
11
12 fi=123.22 # stopnie
13 rfi=math.radians(fi)
14 print(fi,round(math.cos(rfi),4))
```

Drugą grupą funkcji są funkcje służące do generowania liczb pseudolosowych. Poniżej przedstawiono cztery najczęściej używane funkcje, oraz przykłady ich wykorzystanie.

Funkcje modułu random	Opis funkcji
randint(a,b)	losowa liczba całkowita od a do b
random()	losowa liczba rzeczywista od 0 do 1
gauss(a,b)	losowa liczba rozkładu normalnego o średniej a i odchyleniu standardowym b
uniform(a,b)	losowa liczba rzeczywista od a do b
seed(a)	tworzenie powtarzalnej serii liczb losowych

```
1 import random
2
3 print(random.randint(1,10))
4 print(random.random())
5 print(random.gauss(10,5))
6 print(random.uniform(10,20))
```

W Pythonie stworzono wiele bibliotek z najróżniejszymi funkcjami. Jest to ogromna wartość tego języka. Niemniej można także tworzyć w prosty sposób własne funkcje, co okazuje się bardzo przydatne w procesie programowania. Funkcja pełni rolę mini-programu w naszym programie, ale także pozwala na zawarcie wielu linii instrukcji w jednej linii (sprowadzonej do realizacji funkcji).

Przyjrzyjmy się przykładowej funkcji i metodzie jej wykorzystania.

```
1 -def hallo():
2     print('Pozdrowienia')
3     print('Dla Wszystkich')
4
5 hallo()
6 hallo()
7 hallo()
```

Pierwsza linia zawiera słowo **def**, które definiuje funkcję o nazwie **hallo()**. Pod funkcją znajduje się blok kodu (wcięte linie 2 i 3), jest to tzw. *body of function*. Ten kod będzie wykonany kiedy funkcja zostanie wykonana. Nie zostanie on wykonany przy definiowaniu funkcji. Linie 5,6 i 7 to trzykrotne wywołanie funkcji powodujące jej wykonanie. Kiedy wykonamy cały program (linie 1-7) rezultat będzie miał postać.

```
Pozdrowienia  
Dla Wszystkich  
Pozdrowienia  
Dla Wszystkich  
Pozdrowienia  
Dla Wszystkich
```

Podobnie jak w zdefiniowanych już funkcjach np. `round(2.341,1)` w nawiasach możemy umieszczać argumenty funkcji. Aby było to możliwe w procesie definiowania określamy parametry funkcji. Parametr to zmienna zawierająca argument funkcji. Nasza funkcja z parametrem przyjmie postać.

```
1 -def hallo(kto):  
2     print('Pozdrowienia')  
3     print('Dla Wszystkich '+kto)  
4  
5     hallo('białych')  
6     hallo('czarnych')  
7     hallo('żółtych')
```

Rezultat działania funkcji będzie miał postać,

```
Pozdrowienia  
Dla Wszystkich białych  
Pozdrowienia  
Dla Wszystkich czarnych  
Pozdrowienia  
Dla Wszystkich żółtych
```

Jedna istotna uwaga. Nasz parametr (zmienna) **kto** istnieje wyłącznie wewnątrz funkcji. Funkcje przedstawione do tej pory wykonywały pewną czynność (drukowanie), ale aby za pomocą funkcji przypisać wartość w tabeli atrybutowej musi ona zwracać pewną wartość, tak jak na przykład działa funkcja `round()`.

```
1 x=8.3478  
2 y=round(x,2)  
3 print(y)
```

Rezultat działania programu,

```
8.35
```

Aby funkcja zwracała wartość używamy słowa return przy jej definiowaniu. Przyjrzyjmy się funkcji służącej do klasyfikacji jednej zmiennej. Nasza funkcja w zależności od sytuacji zwróci wartość -1, 0 albo 1.

```
1 - def klasyfikacja(x):
2 -     if x>0:
3 -         klasa=1
4 -     elif x<0:
5 -         klasa=-1
6 -     else:
7 -         klasa=0
8 -     return klasa
9
10 print(klasyfikacja(123))
11 print(klasyfikacja(-123))
12 print(klasyfikacja(0))
```

Oczywiście wynikiem działania tego programu będzie,

```
1
-1
0
```

Pytania:

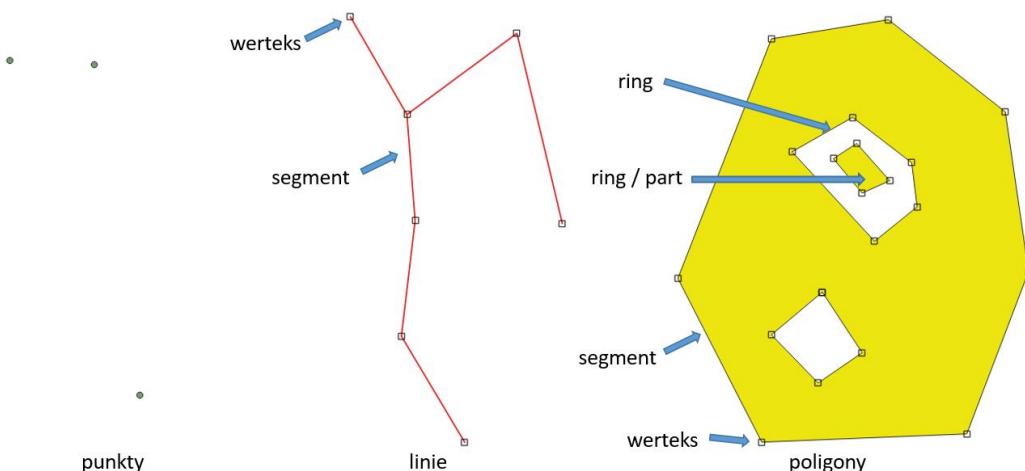
Dane: Pomorskie_v1.gdb

1. Do kopii warstwy jezior dodaj pole zawierające tekst z nazwą jeziora i jego powierzchnią w hektarach (np.' j.Białe powierzchnia 23.4 ha'). Wartość w hektarach powinna być zaokrąglona do jednego miejsca po kropce. Wykorzystaj wartości z pola jako etykiety.
2. Utwórz funkcje do zastosowania w narzędziu Field Calculator, która klasyfikuje wartości z dowolnego pole na pięć klas zwracając identyfikatory klas: 1,2,3,4,5. Funkcja powinna posiadać następujące parametry: nazwa pola oraz cztery wartości progowe oddzielające poszczególne klasy.

4.5 Tworzenie i edycja danych wektorowych

Tworzenie i edycja danych wektorowych obejmuje szereg typowych scenariuszy postępowania, których celem jest otrzymanie potrzebnych warstw danych, przygotowanych do analizy. Te scenariusze mogą być przedstawione jako sekwencja (workflow) stosowania różnego rodzaju metod i narzędzi służących do tworzenia, usuwania lub modyfikowania geometrii danych wektorowych czyli punktów, linii i poligonów. Jakkolwiek te narzędzia są specyficzne dla danego rodzaju oprogramowania to realizują działania prowadzące do podobnych rezultatów, niezależnie od platformy GIS.

Geometria wszystkich trzech typów obiektów oparta jest na punktach o określonych współrzędnych. Najczęściej będziemy mieli do czynienia z danymi dwuwymiarowymi, w których każdy punkt (werteks) tej geometrii będzie miał dwie współrzędne **X** i **Y**. Coraz częściej wykorzystuje się jednak dane trójwymiarowe o współrzędnych **X**, **Y** i **Z**.

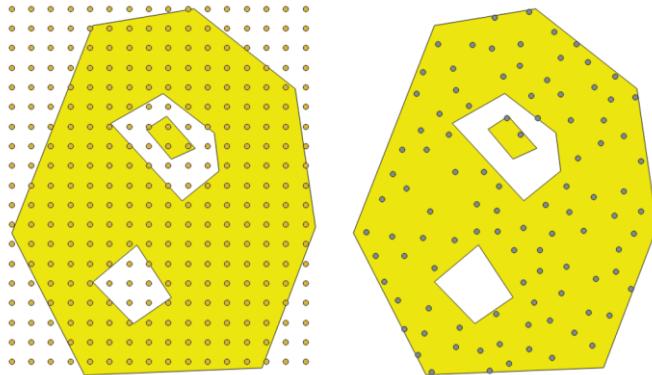


Rysunek 4.22 Podstawowe elementy konstrukcyjne danych wektorowych

Tworzenie i modyfikowanie danych wektorowych opiera się o ich elementy konstrukcyjne. Zwykle wykonujemy te czynności w tzw. **trybie edycyjnym**, co umożliwia po ich zakończeniu albo zapisanie dokonanych zmian, albo powrót do stanu przed edycją. Podstawowymi elementami konstrukcyjnymi są **werteksy**, **segmenty** i **ringi**. Werteksy czyli punkty o określonych współrzędnych są podstawowym elementem geometrycznym wszystkich konstrukcji. Połączone dwa werteksy tworzą segmenty (odcinki), które tworzą linie, a właściwie polilinie. Zamknięte łańcuchy segmentów tworzą ringi, które są podstawowym elementem konstrukcyjnym poligonów. Na rysunku 4.22 pierwszy ring tworzy

obrys poligonu. Następne dwa ringi tworzą dwie dziury w poligonie. Wyspa w jednej z dziur może być zależnie od implementacji kolejnym ringiem, albo kolejną częścią (part) obiektu.

Najprostszym rodzajem danych są punkty. Podstawowe operacje na punktach to ich tworzenie, przesuwane i usuwanie. Mogą istnieć różne odmiany tych trzech operacji, na przykład przesuwanie jako kopiowanie (przesunięty punkt jest kopią oryginalnego, który zostaje zachowany). W przypadku tworzenia warstw punktowych często istnieje potrzeba generowania punktów o położeniu regularnym lub losowym. Na Rysunku 4.23 przedstawiono oba sposoby. Po lewej stronie punkty zostały utworzone w sposób regularny na obszarze zdefiniowanym przez obrzędys warstwy poligonów (pojedynczy poligon). Podstawowym parametrem narzędzi do tego służących jest odległość pomiędzy punktami. Po prawej stronie przedstawiono punkty wygenerowane jako losowe wewnątrz poligonu. Dwa najważniejsze parametry w tym przypadku to liczba punktów i minimalna dozwolona odległość pomiędzy punktami.



Rysunek 4.23 Regularny i losowy sposób tworzenia punktów

Punkty mogą być także generowane losowo wzdłuż linii z zachowaniem minimalnej dozwolonej odległości (parametry to liczba punktów i dozwolona odległość).

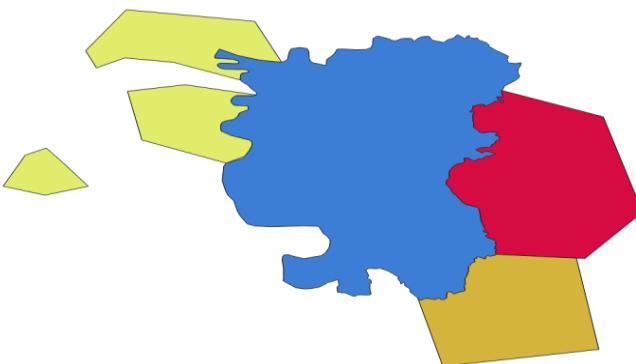


Rysunek 4.24 Punkty rozmieszczone losowo wzdłuż linii

Losowe generowanie punktów jest często wykorzystywane w planowaniu miejsc prowadzenia badań terenowych. Często dodatkowym warunkiem jest aby punkty (w odpowiedniej liczbie) znajdowały się w każdym rodzaju pokrycia terenu, reprezentowanego przez atrybuty warstwy poligonów. Tego typu losowanie nazywamy stratyfikowanym.

Narzędzia edycyjne do tworzenia i modyfikacji linii obejmują oprócz podstawowego narzędzia rysowania linii przez wprowadzanie jej kolejnych werteksów możliwość podziału linii (**Split**), łączenia linii (**Merge**) oraz zmianę jej kształtu (**Reshape**) za pomocą narysowania nowej linii przecinającej zmienianą. Zmiana kształtu będzie miała miejsce od pierwszego do ostatniego przecięcia linii. Werteksy mogą być usuwane, przesuwane i dodawane w dowolnym miejscu segmentu. Kolejność werteksów może też być odwracana (**Reverse**). Istnieje także możliwość prowadzenia nowej linii po istniejącej już linii czy obrysie poligonu (**Trace**).

Narzędzia do edycji poligonów są w części identyczne jak do edycji linii. Ze względu na bardziej złożoną geometrię składającą się z ‘rings’ zawierającą one dodatkowo narzędzia do ich tworzenia oraz metody pozwalające na modyfikację tak złożonej geometrii. Te zadania są w odmienny sposób rozwiązywane w różnych programach GIS. Bardzo potrzebnym narzędziem przy tworzeniu ciągłego pokrycia poligonami jest zapewnienie zgodności ich granic. Zostało to pokazane na Rysunku 4.25. Aby granice poligonów idealnie pasowały do siebie nawzajem tworząc ciągłą powierzchnię (bez luk i nakładania się poligonów) używa się narzędzia **Autocomplite** (ArcGIS Pro) lub opcji **Avoid Intersection** (QGIS). Polegają one na pozostawieniu części tworzonego poligonu, który nie przecina się z poligonami już istniejącymi.



Rysunek 4.25 Tworzenie poligonów o wspólnych granicach

Pytania: (obejmują także rozdział 4.6)

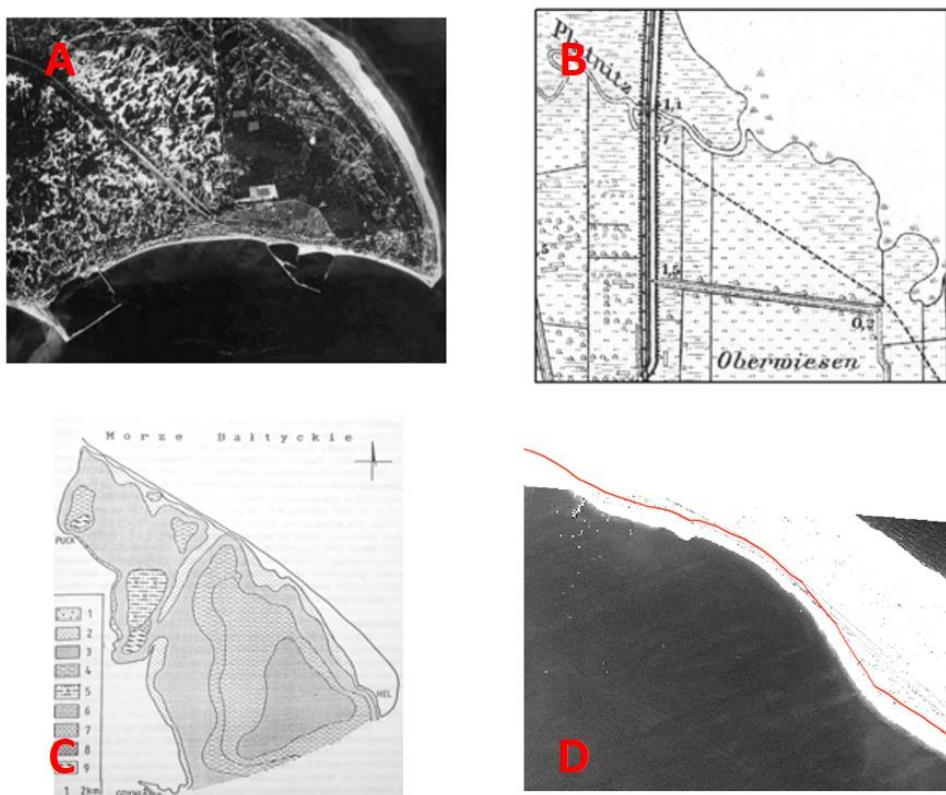
Dane: Sycylia_18.gdb (<https://cgis.oig.ug.edu.pl/ZADANIA/>)

1. Utwórz warstwy: obrys wyspy Pantelleria, pokrycie terenu wyspy Pantelleria, drogi wyspy Pantelleria. Do tablicy warstwy pokrycia dodaj pole z polskim opisem pokrycia (tablica corine-legend). [FILM z ROZWIĄZANIEM zad. 10]
2. Wykorzystując zrzut ekranowy Pantelleria_Scauri.png (katalog Sicilycommondata): [FILM z ROZWIĄZANIEM zad. 11]
 - a. Przeprowadź korektę linii brzegowej dla obszaru zdjęcia (ze szczególnym uwzględnieniem portu – prawdziwa linia brzegowa).
 - b. Przeprowadź korektę warstwy pokrycia terenu dostosowując ją do nowej linii brzegowej.
 - c. Wprowadź nowy poligon dla obszaru portu (nabrzeża plus baseny portowe)
3. Wykonaj to samo zadanie dla okolicy głównego portu wyspy, tworząc własny zrzut ekranowy. Dodatkowo zmodyfikuj pokrycie terenu dla zwartej zabudowy, tak aby opisujące ją wieloboki miały wyłącznie kąty zbliżone do prostych.
4. W strefie 2 Mm od brzegu wyspy Pantelleria rozmieść losowo 30 punktów pomiarowych (nie powinny one leżeć bliżej od siebie i brzegu niż 200 m). W tablicy atrybutowej punktów umieść pole z pozycją punktów w formacie ‘dd mm ss.ss N dd mm ss.ss E’. [FILM z ROZWIĄZANIEM zad. 13]
5. Korzystając z basemap – Open Street uzupełnij drogi w warstwie dróg wyspy (uwzględnij drogi o symbolizacji żółtych linii i pogrubionych białych linii).

4.6 Georeferencja i rejestracja rastra

Najczęściej pozyskane dane przestrzenne posiadają informację o układzie współrzędnych w którym określone są ich współrzędne (współrzędne punktów, lub środków komórek rastrowych). Ta informacja w przypadku formatu **SHP** danych wektorowych znajduje się w pliku o rozszerzeniu **.prj**. W formacie rastrowym **GeoTIFF** i **GeoJP2** (rastry o rozszerzeniu **.tif** i **.jp2**) znajduje się ona wewnątrz pliku rastrowego. Innym rozwiązaniem jest umieszczenie jej w pliku tekstowym zwanym **World File**. Jest to dodatkowy plik referencyjny z tą samą nazwą co raster i rozszerzeniem kończącym się dodatkową literą **w** (np. **raster.jpg** i tekstowy plik **raster.jpgw** lub **raster.tif** i **raster.tifw**).

W przypadku pozyskiwania danych rastrowych można spotkać się z sytuacją braku informacji o lokalizacji rastra na powierzchni Ziemi.



Rysunek 4.26 Przykłady rastrów nie posiadających informacji o lokalizacji na powierzchni Ziemi (A-zdjęcie lotnicze, B-stara mapa, C-skan mapy z książki), bądź posiadających błędную lokalizację (D-zdjęcie satelitarne)

Z taką sytuacją spotkamy się przy wykorzystywaniu starych zdjęć lotniczych na papierze fotograficznym po ich zeskanowaniu (Rysunek 4.25A), zeskanowaniu starych map (Rysunek 4.25B) lub map zawartych w książkach (Rysunek 4.25C)

oraz wykorzystaniu cyfrowych zdjęć satelitarnych z niedokładną lokalizacją (Rysunek 4.25D). Widoczna na Rysunku 4.26D czerwona linia będąca wektorową warstwą brzegu powinna zgadzać się ze skokiem wartości na zdjęciu pomiędzy morzem a lądem (zdjęcie w podczerwieni - morze ciemne, ląd jasny).

Podstawą nadania **georeferencji** warstwie rastrowej skanu jest określenie **geometrycznej transformacji** pozwalającej na przeliczenie współrzędnych lokalnych każdej komórki rastra na współrzędne odpowiadające jej położeniu na powierzchni Ziemi. Przez współrzędne lokalne rozumie się współrzędne prostokątne skanu w cm lub pikselach (lewy dolny róg skanu ma zwykle współrzędne 0,0). Transformacje te mogą mieć różną postać, ale reprezentatywnym ich przykładem może być wielomian pierwszego, drugiego lub trzeciego stopnia postaci,

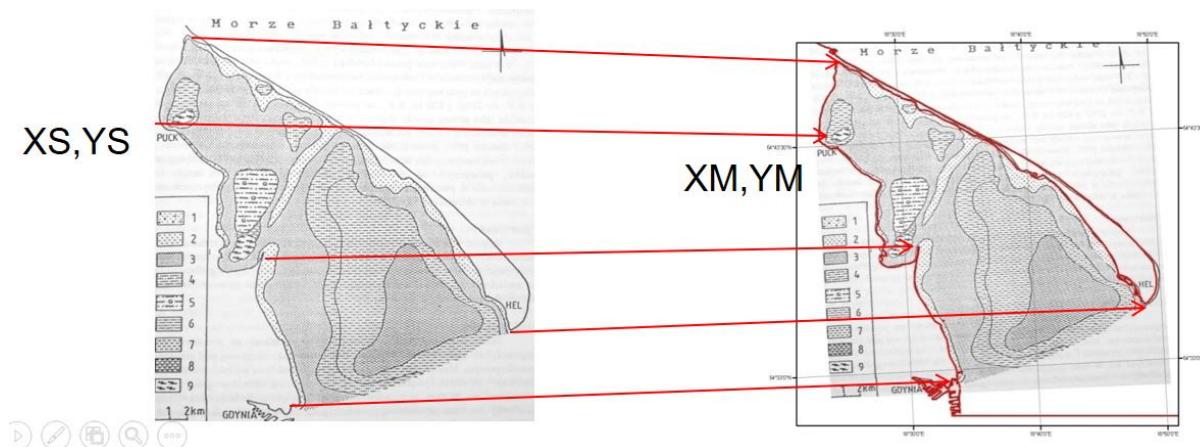
$$x_m = x_s + A_1 x_s + A_2 y_s + A_3 x_s y_s + A_4 x_s^2 + A_5 y_s^2 + A_6 x_s^2 y_s + A_7 x_s y_s^2 + A_8 x_s^3 + \dots$$

$$y_m = y_s + B_1 y_s + B_2 x_s + B_3 x_s y_s + B_4 x_s^2 + B_5 y_s^2 + B_6 x_s^2 y_s + B_7 x_s y_s^2 + B_8 x_s^3 + \dots$$

gdzie:

x_s	<i>współrzędna x skanu</i>	y_s	<i>współrzędna y skanu</i>
x_m	<i>współrzędna x mapy</i>	y_m	<i>współrzędna y mapy</i>

Powyższe równania mogą być rozwiązane (wyznaczenie nieznanych współczynników A_1, B_1, A_2, \dots) za pomocą par współrzędnych (skanu i mapy) dla szeregu odpowiadających sobie punktów na skanie i mapie (**punktów kontrolnych**).



Rysunek 4.27 Tworzenie punktów kontrolnych - zbioru par współrzędnych opisujących współrzędne skanu i mapy dla odpowiadających sobie punktów wskazujących na ten sam element terenu

Współrzędne par punktów kontrolnych są zwykle organizowane w tablicy, a odpowiednie narzędzia umożliwiają ich usuwanie lub dodawanie nowych. Na podstawie punktów kontrolnych określone jest uniwersalne (dla danego skanu) wyrażenie transformacji geometrycznej służące do wyznaczenia nowych współrzędnych dla każdej komórki rastrowej. Różnica pomiędzy współrzędnymi na mapie danego punktu kontrolnego (x_m , y_m), a współrzędnymi punktu skanu otrzymanego z wyrażenia transformacji geometrycznej (x_{st} , y_{st}) nosi nazwę rezydium,

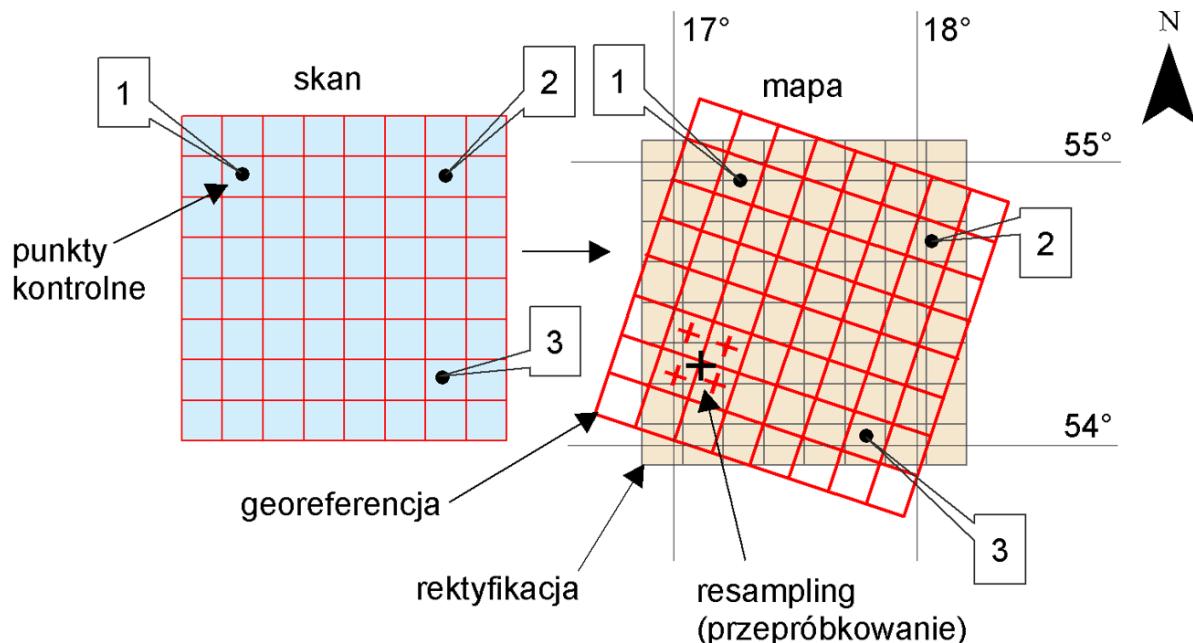
$$D = \left[(x_{st} - x_m)^2 + (y_{st} - y_m)^2 \right]^{1/2}$$

a dokładność całego procesu transformacji geometrycznej skanu opisuje błąd RMS,

$$\text{RMS} = \left\{ \sum \left[(x_{st} - x_m)^2 + (y_{st} - y_m)^2 \right] / n \right\}^{1/2}$$

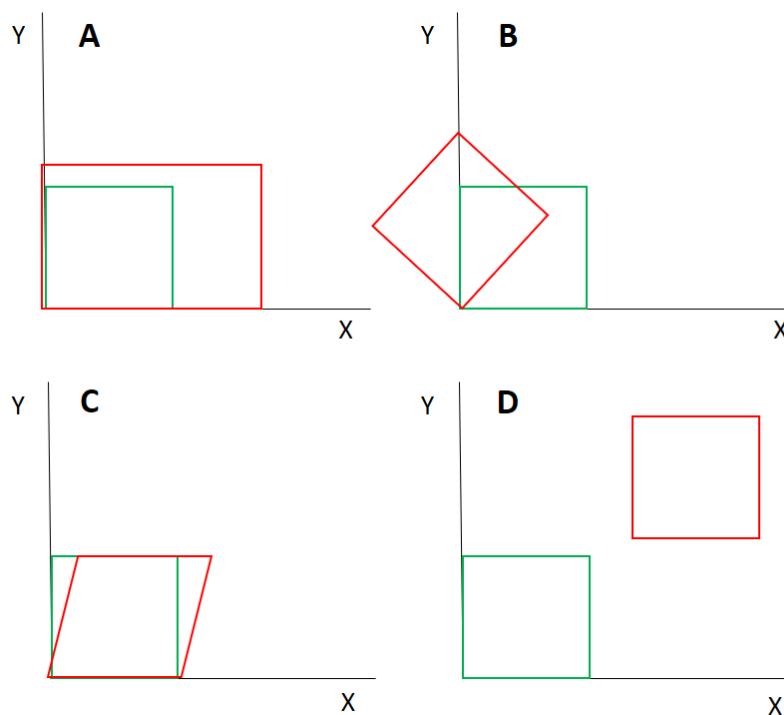
będący pierwiastkiem sumy kwadratu rezydiów podzielonej przez liczbę punktów kontrolnych (RMS ma taką samą jednostkę jak nasze dane).

Podsumowanie procesu georeferencji i rektyfikacji zostało przedstawione na Rysunku 4.28.



Rysunek 4.28 Podsumowanie procesu georeferencja i rektyfikacji

Pierwszy krok polega na wyznaczeniu zbioru punktów kontrolnych, a ściślej ich współrzędnych na skanie i mapie. Na podstawie tych punktów definiowana jest transformacja geometryczna, która służy do obliczenia położenia komórek rasterowych w przestrzeni mapy (czerwona siatka). Istnieją różne metody transformacji różniące się stopniem i sposobem przekształcania geometrii. Jedną z podstawowych metod jest transformacja afinczna (w postaci wielomianu pierwszego stopnia) pozwalająca na cztery rodzaje przekształceń geometrii przedstawionych na Rysunku 4.29. Innymi metodami stosowanymi w systemach GIS są wielomiany różnego stopnia, metoda splajnów, elastycznego dopasowania (Adjust) lub metody projekcji. Generalna zasada polega na wyborze metody, która daje najniższy błąd RMS.

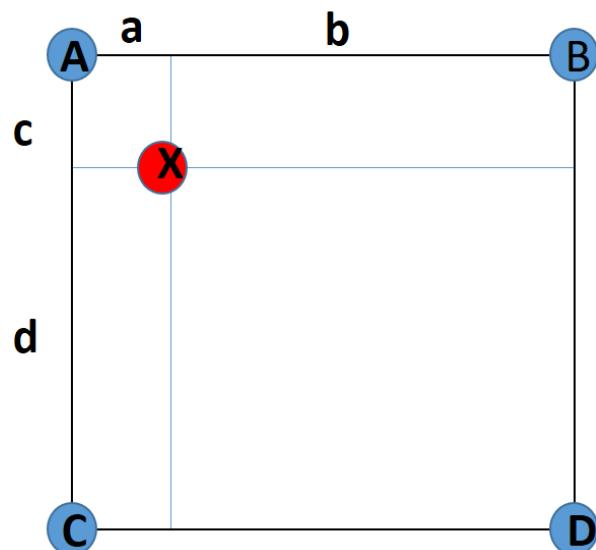


Rysunek 4.29 Cztery rodzaje przekształceń geometrycznych w transformacji afincznej:
A – niezależne skalowanie na osiach YX; B – obrót; C – pochylenie; D - przesuwanie

Transformacja afinczna charakteryzuje się między innymi zachowaniem równoległości linii na mapie. Wracając do rysunku 4.28, to zastosowanie transformacji pozwala na poprawne pod względem lokalizacji przedstawianie (wyświetlanie) rastra względem innych warstw (czerwona siatka rastra).

Aby utworzyć nowy raster, który będzie określony w danym układzie współrzędnych (boki każdej komórki rastra będą równoległe do osi współrzędnych) należy przeprowadzić **rejestrację** lub **rektyfikację** (praktycznie

oba terminy oznaczają to samo). Operacja ta polega na utworzeniu nowego rastra (zaznaczonego piaskowym kolorem), w którym wartości w komórkach (ich środkach geometrycznych) są obliczane na podstawie wartości komórek rastra po georeferencji. Na rysunku 4.28 nowa wartość komórki zlokalizowana w miejscu czarnego krzyżyka jest obliczana z otaczających ją komórek (czerwone krzyżyki). Obliczanie nowych wartości nosi nazwę przepróbkowania (resamplingu). W rastrach jakościowych stosuje się metodę najbliższego sąsiedztwa (Nearest Neighbor), natomiast w rastrach ilościowych najczęściej interpolację bilinearną - Bilinear Interpolation (w oparciu o cztery punkty) lub Cubic Convolution (w oparciu o szesnaście punktów). Zasada interpolacji bilinearnej została przedstawiona na Rysunku 4.30.



$$X = (D * ac + C * bc + B * ad + A * bd) / ((a+b)*(c+d))$$

Rysunek 4.30 Metoda interpolacji bilinearnej. Wartość X w środku komórki nowego rastra jest obliczana z czterech wartości ją otaczających. Wartości ac,bc,ad,bd oznaczają powierzchnię odpowiadających im prostokątów

4.7 Zmiana wartości w rastrze (Algebra Map, Raster Calculator)

Algebra map jest podstawową metodą wykonywania operacji na rastrach. Ogólna zasada polega na wykonaniu operacji algebraicznej oddzielnie w każdej komórce rastra, lub w przypadku operacji na kilku rastrach w każdych odpowiadających sobie komórkach rastrów. Na rysunku 4.30 przedstawiono podstawową zasadę tych dwóch operacji.

$$\begin{array}{|c|c|c|} \hline 2 & 11 & 2 \\ \hline 7 & 2 & 12 \\ \hline 3 & 6 & 1 \\ \hline \end{array} + 2 = \begin{array}{|c|c|c|} \hline 4 & 13 & 4 \\ \hline 9 & 4 & 14 \\ \hline 5 & 8 & 3 \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|} \hline 2 & 11 & 2 \\ \hline 7 & 2 & 12 \\ \hline 3 & 6 & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 4 & 13 & 4 \\ \hline 9 & 4 & 14 \\ \hline 5 & 8 & 3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 6 & 24 & 6 \\ \hline 16 & 6 & 26 \\ \hline 8 & 14 & 4 \\ \hline \end{array}$$

Rysunek 4.30 Podstawowe operacje algebry map wykorzystujące operator dodawania. Dodanie liczby i innego rastra

Trzecią możliwością jest zastosowanie funkcji, której parametrem jest raster (na przykład obliczenie pierwiastka kwadratowego). Wynikiem takiej operacji będzie raster z pierwiastkiem kwadratowym każdej komórki rastra. Znak plus nosi nazwę operatora (podobnie jak w Pythonie) i może być zastąpiony szeregiem operatorów arytmetycznych (+, -, *, /), porównania (==, >, <, <=, >=, !=) i logicznych (&, |, ~, ^). Operator & oznacza *and*, a operator | *or*. Za pomocą algebry map mogą być tworzone złożone wyrażenia co umożliwia dowolne przekształcanie rastrów.

Operacje algebry map są wykonywane w GIS w kalkulatorach rastrowych (**Raster Calculator**). Chociaż wszystkie kalkulatory realizują algebrę map ich działanie w poszczególnych programach różni się od siebie. Z punktu widzenia użytkownika istotna jest przede wszystkim składnia używana przy tworzeniu wyrażeń algebry map. Ogólna zasada jest taka, że w miejsce zmiennych wstawiamy nazwy rastrów. Oznacza to, że wyrażenie,

$$(\text{Raster1} + \text{Raster2} + \text{Raster3}) / 3$$

utworzy średni raster, w którym wartość każdej komórki zostanie obliczona zgodnie z tym wyrażeniem.

W algebrze map używa się wartości **NoData**, wystąpienie jej w dowolnej operacji algebry map daje w komórce wynik NoData. Jest to pokazane na Rysunku 4.31.

<table border="1"><tr><td>2</td><td>11</td><td>2</td></tr><tr><td>7</td><td>2</td><td>12</td></tr><tr><td>3</td><td></td><td></td></tr></table>	2	11	2	7	2	12	3			$+ 2 =$	<table border="1"><tr><td>4</td><td>13</td><td>4</td></tr><tr><td>9</td><td>4</td><td>14</td></tr><tr><td>5</td><td></td><td></td></tr></table>	4	13	4	9	4	14	5			 = NoData									
2	11	2																												
7	2	12																												
3																														
4	13	4																												
9	4	14																												
5																														
<table border="1"><tr><td>2</td><td>11</td><td>2</td></tr><tr><td>7</td><td></td><td>12</td></tr><tr><td></td><td>6</td><td>1</td></tr></table>	2	11	2	7		12		6	1	$+ \quad =$	<table border="1"><tr><td></td><td>13</td><td>4</td></tr><tr><td>9</td><td>4</td><td>14</td></tr><tr><td>5</td><td>8</td><td>3</td></tr></table>		13	4	9	4	14	5	8	3	<table border="1"><tr><td></td><td>24</td><td>6</td></tr><tr><td>16</td><td></td><td>26</td></tr><tr><td></td><td>14</td><td>4</td></tr></table>		24	6	16		26		14	4
2	11	2																												
7		12																												
	6	1																												
	13	4																												
9	4	14																												
5	8	3																												
	24	6																												
16		26																												
	14	4																												

Rysunek 4.31 Zasada operacji na wartościach NoData w algebrze map

Składania wyrażeń zamiast nazw rastrowów może wykorzystywać ich aliasy w postaci kolejnych liter alfabet: A, B, C, D, E... Operatory porównania mogą być zastąpione funkcjami **gt(x,y)**, **lt(x,y)** i **eq(x,y)**. które zwracają prawdę jeśli **x** jest odpowiednio większa, mniejsza lub równa **y**. Wartości NoData mogą być wprowadzane do wyrażeń algebry map w różny sposób (np. jako brak parametru ,0/0 lub zadeklarowany alias tej wartości). Każdy z dostępnych kalkulatorów rastrowych posiada specyficzne dostępne operatory, funkcje arytmetyczne, trygonometryczne oraz specjalne pozwalające na operacje między innymi na wartościach NoData. Dodatkowo kalkulatory zawierają funkcje do operacji kompleksowych realizowanych według schematu,

GDY (wyrażenie porównania lub logiczne , jeśli prawda to, jeśli fałsz to)

Funkcja tego typu w różnych kalkulatorach ma inną nazwę (**Con**, **ifelse**, **where**), ale zasada jej działania oparta na trzech argumentach jest taka sama. Pierwszy argument jest sprowadzony do wartości PRAWDA / FAŁSZ, jeśli jest to prawda funkcja zwraca wartość wynikającą z drugiego argumentu, a jeśli fałsz to z trzeciego. Argumentem tej funkcji może być ona sama, co daje duże możliwości realizacji obliczeń rożnego typu. Poniżej przedstawiono szereg przykładów zastosowania tego typu konstrukcji realizacji operacji algebry map.

GDY (Raster1 > 5, 1, 2)

2	11	2
7	2	12
3	6	1

Raster1



2	1	2
1	2	1
2	1	2

Jeżeli piksel w rastrze *Raster1* ma wartość większą od 5 przypisz odpowiadającej komórce w rastrze wynikowym wartość 1, jeśli jest mniejszy lub równy przypisz 2.

GDY (Raster1 > Raster2, 1, 0)

2	11	2
7	2	16
3	6	1

Raster1

22	11	1
17	4	14
1	3	0

Raster2



0	0	1
0	0	1
1	1	1

Jeżeli piksel w rastrze *Raster1* ma wartość większą od odpowiadającej komórki w rastrze *Raster2* to przypisz odpowiadającej komórce w rastrze wynikowym wartość 1, jeśli jest mniejszy lub równy przypisz 0.

GDY (Raster1 > 5, 1, Raster1)

2	11	2
7	2	12
3	6	1

Raster1

2	1	2
1	2	1
3	1	1

Jeżeli piksel w rastrze *Raster1* ma wartość większą od 5 przypisz odpowiadającej komórce w rastrze wynikowym wartość 1, jeśli jest mniejszy lub równy przypisz 2.

GDY ((Raster1 > 8) | (Raster2 < 10), **GDY**(Raster1 > 5, 1, 0), -1)

2	11	2
7	2	16
3	6	1

Raster1

22	11	1
17	4	14
1	3	0

Raster2



-1	1	0
-1	0	1
0	1	0

Jeżeli piksel w rastrze *Raster1* ma wartość większą od 8 lub w rastrze *Raster2* ma wartość mniejszą od 10 to przypisz odpowiadającej komórce w rastrze

wynikowym 1 albo 0. Przypisz 1 jeśli Raster 1 jest większy od 5, jeśli nie przypisz 0. Jeśli natomiast pierwszy warunek (operacja z lub) nie jest spełniony to przypisz odpowiadającej komórce w rastrze wynikowym -1.

GDY (Raster1 <= 5, 1, **GDY**((Raster1>5) & (Raster1<=10),2,3))

2	11	2
7	2	12
3	6	1

Raster1



1	3	1
2	1	3
2	2	1

Jeżeli piksel w rastrze *Raster1* ma wartość mniejszą lub równą 5 to przypisz odpowiadającej komórce w rastrze wynikowym wartość 1, jeżeli natomiast wartość mieści się w przedziale od 5 do 10 przypisz 2, jeśli jest większa od 10 przypisz 3. W ten sposób funkcja tego typu może być wykorzystywana w operacji reklasyfikacji rastra (będzie o tym mowa w kolejnych rozdziałach).

Pytania:

dane: SST3.gdb (<https://cgis.oig.ug.edu.pl/ZADANIA/>)

1. Przelicz rastry temperatury na wartości dziesiętne w stopniach Celsjusza.
2. Na podstawie pomiarów z lutego, maja i sierpnia oblicz średnią temperaturę powierzchni morza w każdej komórce rastra i wyraź ją w stopniach Celsjusza.
3. Przekształć rastry, tak by zamienić wartości 0 z lądu na "NoData" (reszta wartości ma pozostać taka sama).
4. Utwórz raster binarny (0,1) maskę ląd-morze, gdzie ląd będzie oznaczony jako "1", a woda jako "0".
5. Znajdź obszary w których zarówno w lutym jak i w sierpniu utrzymuje się temperatura powyżej 25 st. C.
6. Znajdź obszary, które różnią się temperaturą pomiędzy lutym i sierpniem o więcej niż 2 stopnie. Przypisz 1 gdy różnica ma znak plus, a 2 gdy minus
7. Przedstaw temperaturę powierzchni morza w trzech przedziałach: poniżej 10 °C klasa1; 10-20 °C klasa 2; powyżej 20 °C klasa3.
8. Wyznacz obszary gdzie temperatura wykazuje mniejszą zmienność niż 2 stopnie.

Rozdział 5

Pozyskiwanie informacji przestrzennej (analiza danych)

Podstawowym zadaniem analizy danych jest zrozumienia rzeczywistości lub próba odpowiedzi na określone pytania. Specyfika GIS polega na operowaniu danymi przestrzennymi. W rezultacie rozwinięte zostały specyficzne metody analizy uwzględniające lokalizację danych. Metody analizy danych realizowane w ramach *data science* osiągnęły w ostatnich latach ogromne możliwości rozwiązywania różnorodnych problemów. Niemniej w ich głównym nurcie element przestrzenny, może z wyjątkiem analizy obrazu, jest właściwie nieobecny. Podstawowym formatem danych, którymi operuje się w *data science* są ogromne zbiory danych tekstowych zorganizowane w formie tablic, w których wiersze opisują obserwacje, a kolumny cechy tych obserwacji. Możliwość pozyskania dodatkowych cech obserwacji wynikających z relacji przestrzennych ma potencjalnie ogromne znaczenie w inżynierii danych. Tego typu zadania mogą być realizowane przez różne narzędzia analizy przestrzennej, które umożliwiają dodawanie atrybutów do tablic atrybutowych z innych warstw danych na podstawie relacji przestrzennych. Podstawowym narzędziem służącym do tego celu jest SPATIAL JOIN.

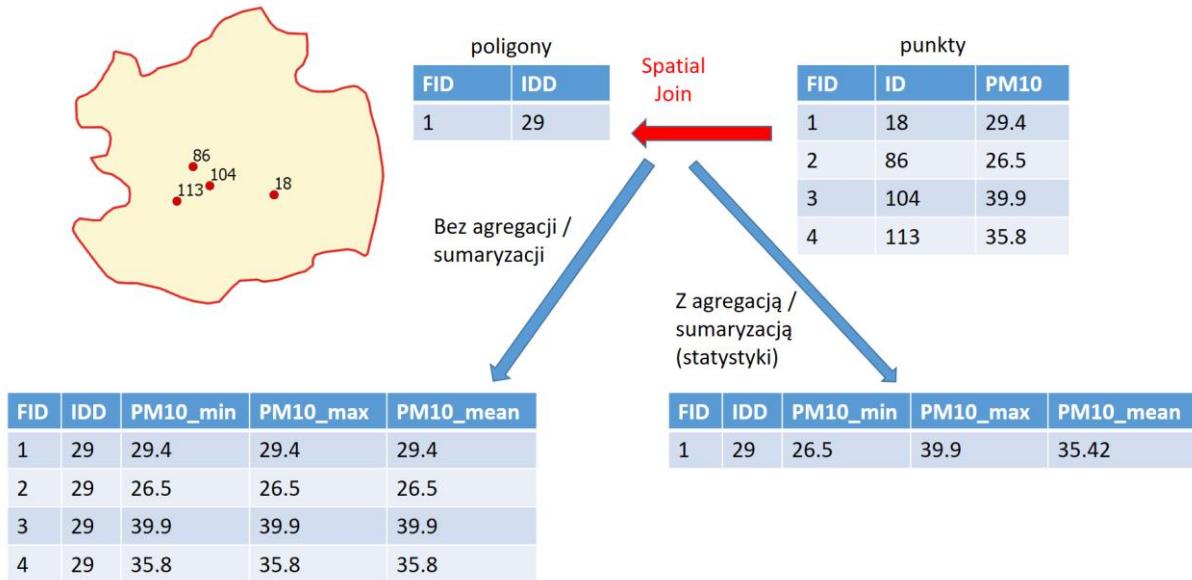
GIS jako narzędzie jest wykorzystywane w wielu dziedzinach (np. ekologii, kryminalistyce, gospodarce przestrzennej, hydrologii, geologii). W jednych korzysta się przeważnie z danych wektorowych w innych z rastrowych. W każdej z tych dziedzin dopracowano się specjalistycznych narzędzi analizy. Niemniej budowane są one na pewnym stałym zestawie analitycznych metod rastrowych i wektorowych, które stanowią rdzeń każdego systemu GIS.

5.1 Pozyskiwanie atrybutów na podstawie relacji przestrzennych

Założmy, że mamy dwie warstwy danych wektorowych posiadających atrybuty, warstwę poligonów z atrybutem będącym identyfikatorem poligonu i warstwę punktów z atrybutami opisującymi zanieczyszczenie powietrza i identyfikator punktu. Obie warstwy nie mają wspólnych atrybutów. Możemy jednak przypisać atrybuty punktów do poligonów na podstawie relacji przestrzennej pomiędzy obiekta mi obu warstw. Intuicyjnie najprostszą relacją będzie wzięcie pod uwagę atrybutów punktów, które znajdują się wewnątrz danego poligonu. Istota części relacji przestrzennych została przedstawiona na Rysunku 4.14 (przy omawianiu selekcji obiektów na podstawie relacji przestrzennych). W praktyce używa się następujących relacji przestrzennych, przy czym mogą się one odnosić do dowolnej kombinacji typów danych wektorowych (punktów, linii i poligonów).

Nazwa relacji dwóch obiektów dwóch warstw (do obiektu A dodajemy atrybuty obiektu B)	Opis
Intersect	Dowolny element jednego obiektu zasłania część drugiego obiektu. Istnieją wspólne punkty, linie lub powierzchnie tych obiektów. Mówimy także, że obiekty się przecinają.
Within a distance (DD)	Odległość pomiędzy obiektami jest mniejsza lub równa DD.
Contains	Obiekt A zawiera obiekt B .
Within	Obiekt A jest zawarty w obiekcie B .
Are identical / Equals	Obiekty są identyczne.
Boundary touches / Touches	Obiekty stykają się obrysami.
Share a line segment	Mają wspólny odcinek.
Crossed by the outline of / Crosses	Obrys obiektu przecina drugi obiekt
Closest	Najbliższy obiekt B od obiektu A .
Overlays	Jeden obiekt nachodzi na drugi obiekt.

Pozyskiwanie atrybutów na podstawie relacji przestrzennych odbywa się za pomocą narzędzia lub funkcji nazywanej **SPATIAL JOIN**. Istota jej działania została przedstawiona na Rysunku 5.1



Rysunek 5.1 Schemat działania pobierania atrybutów na podstawie relacji przestrzennej (*intersect*) za pomocą funkcji SPATIAL JOIN

Podstawą operacji jest zdefiniowanie warstwy do której chcemy pozyskiwać atrybuty oraz warstwy w której się one znajdują. W przykładzie powyżej warstwą do której chcemy dodać atrybuty (zmienną PM10 z warstwy punktów) jest warstwa poligonów. Interesują nas atrybuty punktów, które znajdują się wewnątrz poligonu. Taką relację przestrzenną opisuje między innymi *intersect*. Istnieją dwa sposoby przeprowadzenia tej operacji nazwane *one-to-one* i *one-to-many*. W pierwszym przypadku do poligona zostanie dołączony jeden atrybut. Jeśli wewnątrz poligonu znajduje się kilka punktów to nastąpi agregacja (sumaryzacja) atrybutów za pomocą wybranej opcji (może to być dowolna statystyka czyli np., wartość min, max czy średnia albo inna zdefiniowana wartość otrzymana z kompletu atrybutów). Funkcja SPATIAL JOIN zawiera możliwość definiowania sposobu agregacji w zależności od typu zmiennej. Jeżeli zostanie wybrana opcja *one-to-many*, dla każdego punktu zostanie utworzony oddzielny rekord zawierający jego wartości. Czyli w naszym przykładzie powstaną cztery rekordy z poligonom i atrybutami kolejnych punktów.

Pytania:

Dane: Sycylia_18.gdb (<https://cgis.oig.ug.edu.pl/ZADANIA/>)

1. Do kopii warstwy Sycylia_admin_pop dodaj dwa pola jedno typu całkowitego i drugie typu rzeczywistego.
2. W polu typu całkowitego wpisz 1 dla gmin (NAME_3) nadmorskich i 2 dla pozostałych.
3. W polu typu rzeczywistego wpisz powierzchnię gmin w km2.
4. Oblicz średnią, maksymalną i minimalną powierzchnię wszystkich gmin (w km2).
5. Oblicz średnią, maksymalną i minimalną powierzchnię gmin nadmorskich i pozostałych (w km2).
6. Oblicz ile osób mieszka w każdej prowincji (NAME_2) w miastach poniżej 2000. [FILM z ROZWIĄZANIEM zad. 6]
7. Utwórz warstwę miast powyżej 10tyś. odległych nie dalej niż 50 km od Palermo. [FILM z ROZWIĄZANIEM zad. 7]
8. Utwórz warstwę miast nie różniących się od miasta Scaccia pod względem liczby mieszkańców o więcej niż 10000 osób i odległych od Scaccia o mniej niż 100 km. [FILM z ROZWIĄZANIEM zad. 8]
9. Jaka jest powierzchnia w km2 różnych rodzajów pokrycia terenu na Sycylii (bez wysp), dodaj to tablicy z tymi obliczeniami pole z nazwą pokrycia terenu. Wykonaj mapę pokrycia terenu dla dowolnej prowincji (NAME_2) i utwórz odpowiednią legendę. Wykorzystaj warstwę pokrycia terenu z katalogu pomocniczego Sicilycommondata. [FILM z ROZWIĄZANIEM zad. 9]

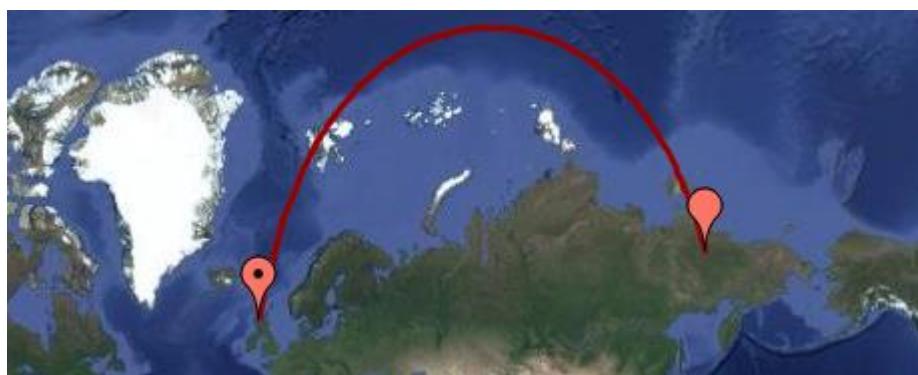
Dane: Pomorskie_v1.gdb (<https://cgis.oig.ug.edu.pl/ZADANIA/>)

10. Do kopii powiatów dodaj dwa pola jedno typu całkowitego i drugie typu rzeczywistego.
11. W polu typu całkowitego wpisz 1 dla powiatów nadmorskich i 2 dla pozostałych.
12. W polu typu rzeczywistego wpisz powierzchnię powiatu w km2.
13. Oblicz średnią, maksymalną i minimalną powierzchnię wszystkich powiatów (w km2).
14. Oblicz średnią, maksymalną i minimalną powierzchnię powiatów nadmorskich i pozostałych (w km2).
15. Oblicz ile osób mieszka w każdym powiecie w miastach poniżej 2000.

- 16.Utwórz warstwę punktów centralnych miast powyżej 2000 mieszkańców.
- 17.Utwórz warstwę miast powyżej 10tyś. odległych nie dalej niż 50 km od Gdańska.
- 18.Utwórz warstwę miast nie różniących się od Helu pod względem liczby mieszkańców o więcej niż 1000 osób i odległych od Helu o mniej niż 80 km.
- 19.Jaka jest powierzchnia w km² różnych rodzajów pokrycia terenu w województwie pomorskim, dodaj to tablicy z tymi obliczeniami pole z nazwą pokrycia terenu. Wykonaj mapę pokrycia terenu dla dowolnego powiatu utwórz odpowiednią legendę.
- 20.Dla każdej miejscowości przypisz nazwę i odległość najbliższego jeziora o powierzchni ponad 300 ha.
- 21.Dla każdego jeziora powyżej 50 ha przypisz liczbę mieszkańców w buforze 2 km.

5.2 Analiza odległości pomiędzy obiektami

Punkty na powierzchni Ziemi o określonej szerokości i długości geograficznej dają możliwość wykonania szeregu obliczeń związanych z odlegością. Obliczenia te mogą być prowadzone zarówno dla sferycznego modelu Ziemi (kula) jak i elipsoidy. Efekt nie uwzględniania spłaszczenia Ziemi powoduje błąd zwykle nie większy od 0.3% (czyli do 3 m na 1 km). Podstawową wielkością jest liczony po wielkim kole na powierzchni Ziemi (**ortodromie**) najkrótszy dystans pomiędzy dwoma punktami (**great-circle distance**). W operacji tej tworzy się fragmenty okręgu, który przebiega przez te punkty i posiada środek w środku sfery. W przybliżeniu odpowiada on trasie ptaka lecącego po najkrótszej drodze. Oczywiście ukształtowanie terenu nie jest brane pod uwagę.



Rysunek 5.1 Wyznaczanie najkrótszego dystansu pomiędzy dwoma punktami na sferze

Dystans ten jest liczony za pomocą wzoru wyprowadzonego z geometrii sferycznej (*spherical law of cosines*):

$$D = \arccos(\sin \varphi_1 \cdot \sin \varphi_2 + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \cos \Delta\lambda) \cdot R$$

gdzie: φ_1, φ_2 - szerokości geog., $\Delta\lambda$ – różnica dług.geog., R - promień Ziemi (średni = 6371 km)

albo za pomocą formuły o nazwie *haversine*. Pomocniczą wielkością, która może być wyznaczona jest początkowy azymut kursu (**bearing**) oraz punkt docelowy (we współrzędnych geograficznych), który może być policzony na podstawie początkowego azymutu, współrzędnych punktu startowego i odległości. Dystans może być liczony także wzdłuż **loksodromy** (rhumb line) czyli linii o stałym azymucie, przecinającej równoleżniki pod tym samym kątem. Dystans wzdłuż

loksodromy jest zwykle krótszy od dystansu liczonego wzdłuż ortodromy. Wyrażenia na obliczenia oraz on-line kalkulator można znaleźć na stronie:

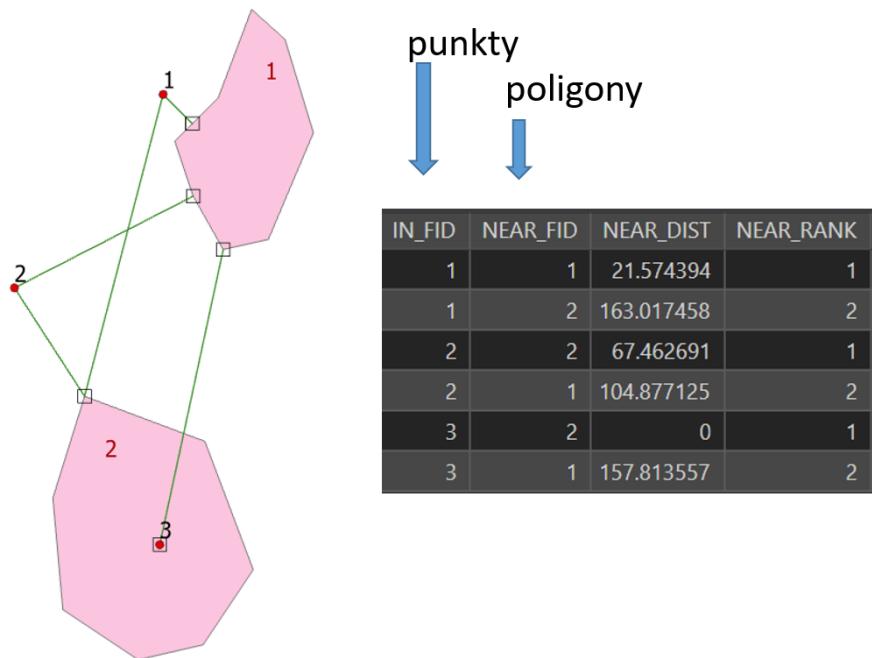
<https://www.movable-type.co.uk/scripts/latlong.html>

W GIS najczęściej używamy współrzędnych prostokątnych i dystans pomiędzy dwoma punktami liczony jest jako odległość euklidesowa równa długości odcinka łączącego te punkty:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

gdzie: x_1, y_1 i x_2, y_2 współrzędne punktów.

GIS oferuje szereg narzędzi służących do wyznaczania odległości pomiędzy obiekty. Podstawowym rezultatem tego typu operacji jest **macierz dystansu** (znaną też jako *near table*), która zawiera odległości pomiędzy obiekty dwu warstw. Na Rysunku 5.2 pokazany jest wynik działania takiej operacji w postaci tablicy.



Rysunek 5.2 Operacja analizy odległości pomiędzy obiekty warstwy punktów i poligonów

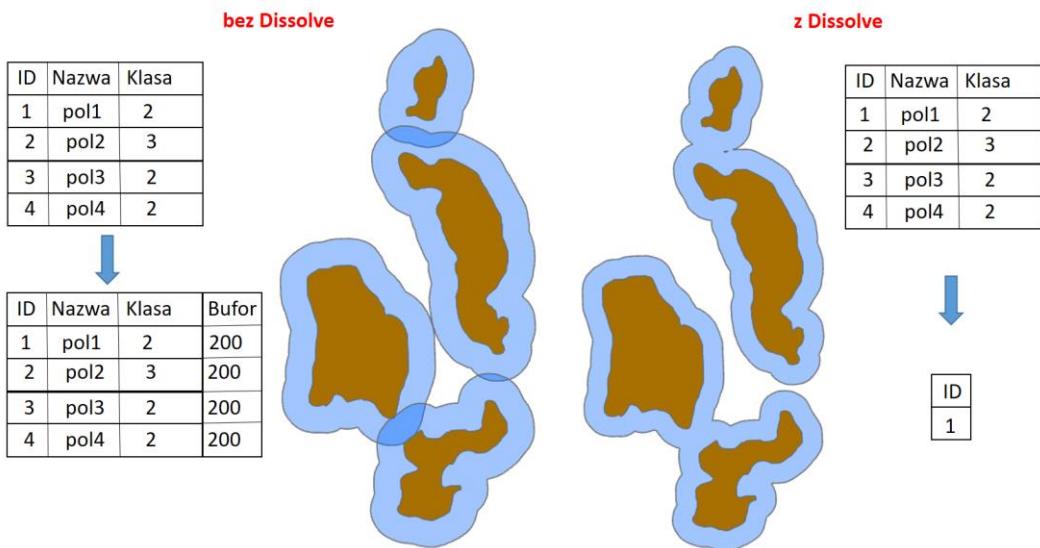
Warstwa punktów zawiera 3 obiekty, a warstwa poligonów 2 obiekty. Rezultatem jest $3 \times 2 = 6$ rekordów. W tablicy zawarta jest odległość od każdego punktu (IN_FID) do każdego poligona (NEAR_FID) oraz ranga tej odległości (najbliższy punkt ma rangę 1, następny 2 itd.). Procedura ta może być ograniczona tylko do najbliższych obiektów, wtedy tablica zawierałaby 3 rekordy. Dodatkową opcją może być wyznaczenie położenia najbliższego punktu

obiektu tworzącego odcinek o wyznaczonej odległości (na rysunku punkty te przedstawione zostały jako kwadraty, a odcinki kolorem zielonym). W tablicy mogą też być zamieszczone azymuty tych odcinków.

Istnieją dwa rodzaje narzędzi do obliczania dystansu, jednak różnią się one znacznie funkcjonalnością w ArcGIS i QGIS. Podstawową rolę pełnią narzędzia do tworzenia macierzy dystansu (**GENERATE NEAR TABLE** w ArcGIS i **DISTANCE MATRIX** w QGIS). Narzędzie w ArcGIS tworzy tablicę dystansów dla dowolnego typu obiektów, dodatkowo istnieje możliwość wykonania tzw. *self operation*, czyli warstwy z samą sobą. Narzędzie to posiada szereg dodatkowych opcji jak kolejność pod względem odległości, lokalizację najbliższego punktu oraz azymut odcinka odległości. Natomiast DISTANCE MATRIX w QGIS działa tylko na punktach i tworzy tablicę warstwy punktowej. Operacja typu *self* jest możliwa tylko za pomocą sklonowania warstwy punktów i wprowadzenia jej jako druga warstwa. Narzędzie nie posiada dodatkowych opcji. Za pomocą drugiego rodzaju narzędzi można pozyskać dla dowolnego obiektu warstwy informację o najbliższym obiekcie innej warstwy. Narzędzie **Near** w ArcGIS pozwala na przypisanie identyfikatora (FID) najbliższego obiektu wraz z odległością oraz opcjonalnie lokalizacji i azymutu. W QGIS określenie odległości do najbliższego obiektu jest możliwe za pomocą wtyczki **NNJoin**.

5.3 Bufory i inne otoczenia obiektów

W wielu praktycznych zagadnieniach istotną rolę pełni najbliższe otoczenie obiektów. Ogniska choroby otaczamy strefą buforową, która zabezpieczyć ma przed rozprzestrzenianiem się epidemii. Rzeki otaczamy buforem (riparian buffers), które tworzą pas lądu środkiem którego płynie rzeka. Pas ten ma istotne ekologiczne znaczenie w ochronie środowiska rzeki i zabezpiecza ją przed zanieczyszczeniami z obszarów przez które przepływa. Wielkie metropolie mogą być otoczone strefą oddziaływania gospodarczego lub kulturowego. W każdym z przypadków bufor to poligon otaczający obiekt, charakterystyczną cechą tego poligona jest jego rozmiar wyrażony w jednostce długości (metrach). Każdy punkt tego poligona leży nie dalej od obiektu niż wielkość tego zdefiniowanego rozmiaru. Czyli jeżeli otoczymy punkt buforem 300 metrowym to powstanie koło o promieniu 300 metrów.

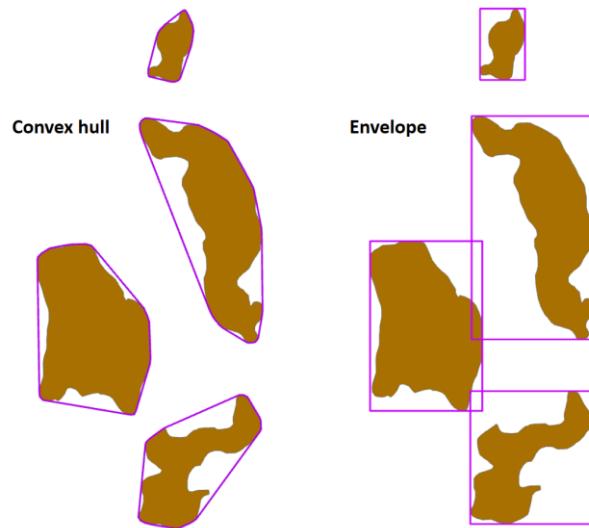


Rysunek 5.3 Dwa typy buforów i ich tablice atrybutowe. Oddzielne bufory dla każdego obiektu (bez użycia operacji Dissolve), oraz bufor będący wynikiem zlewania się nakładających się buforów (z użyciem operacji Dissolve)

Na rysunku 5.3 cztery poligony (wyspy) zostały otoczone buforami. Wyspy posiadają tabelę atrybutową z dwoma polami (Nazwa i Klasa). Buforowanie może być przeprowadzone na dwa sposoby. Pierwszy sposób (bez operacji **Dissolve**) tworzy oddzielny bufor dla każdego poligonu. Każdy bufor (poligon warstwy buforów) zachowuje atrybuty obiektu który otacza. Dodatkowo zostaje dodana kolumna z rozmiarem bufora (Bufor). Drugi sposób (z operacją **Dissolve**) powoduje zlewanie się pokrywających się poligonów. W rezultacie powstaje jeden złożony obiekt składający się z wielu poligonów. Zostaje on zapisany w jednym rekordzie. Warstwa, w której w pojedynczym rekordzie może znajdować się wiele obiektów posiada geometrię **multipart** (składającą się z wielu części). W każdym z GIS znajduje się narzędzie do zamiany geometrii z **multipart do singlepart** (składającą się z pojedynczych części, czyli każdy obiekt będzie zapisany w pojedynczym rekordzie).

Narzędzia przeprowadzające buforowanie mogą zawierać szereg dodatkowych opcji. Rozmiar bufora może być inny dla każdego obiektu. Zostaje on pobrany z pola warstwy buforowanej w którym zawarta jest informacja o rozmiarze bufora dla każdego obiektu. W przypadku buforowania poligonów bufor może zawierać buforowany poligon albo go tylko otaczać. Istnieje możliwość tworzenia buforów wewnętrznych (zwykle stosuje się do tego celu wprowadzanie rozmiaru ze znakiem minus). Na przykład dla jeziora można stworzyć bufor obejmujący jego strefę płytkowodną.

Buforowanie nie jest jedynym sposobem otaczania obiektów. Istnieje szereg geometrii które mogą je otaczać.



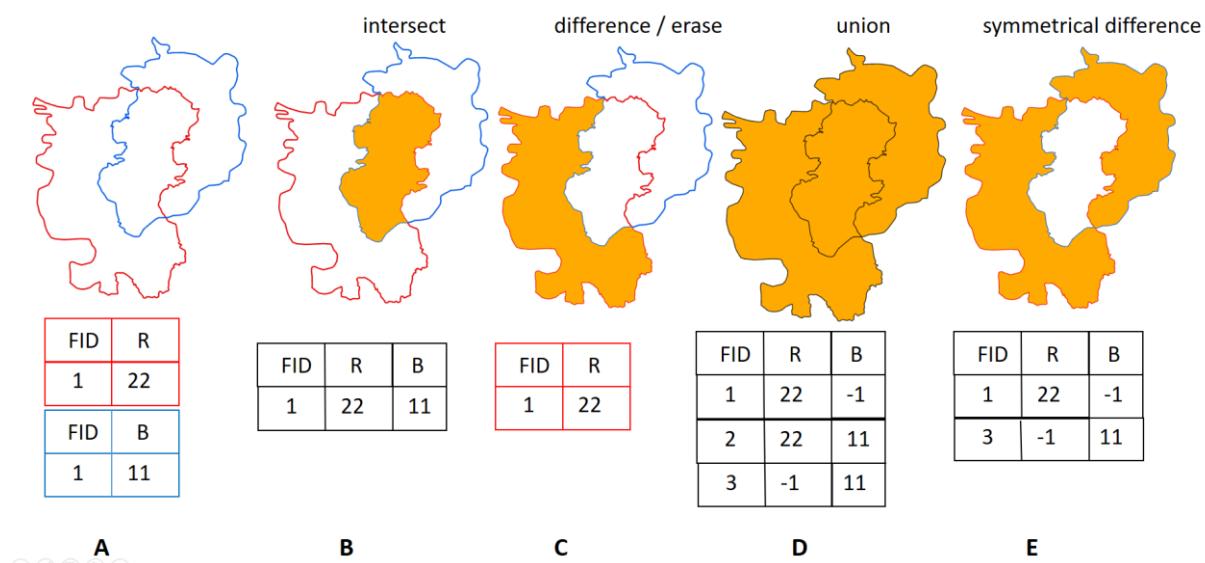
Rysunek 5.4 Otaczanie obiektów za pomocą Convex hull i Envelope

Na rysunku 5.4 przedstawiono dwie metody tworzenia warstwy poligonów otaczających obiekty. **Convex hull** (otoczka wypukła) otacza obiekt w taki sposób aby odcinki łączące dowolne dwa punkty obrysu obiektu znajdowały się wewnętrz otoczki. Convex hull jest jednym z najczęściej używanych zgeneralizowanych obrysów poligonów. Drugim często stosowanym obrysem obiektów jest **Envelope** tworzonym jako prostokąty zdefiniowane dla każdego obiektu (linii lub poligonów) których współrzędne określone są za pomocą maksymalnych i minimalnych wartości współrzędnych danego obiektu.

5.4 Operacje na geometrii danych

Podstawową cechą operacji na geometrii danych jest powstanie nowej geometrii. Tego typu operacje dają unikalne możliwości analityczne.

Jeżeli na dwóch przezroczystych foliach umieścimy warstwy poligonów a następnie nałożymy na siebie te folie to zobaczymy rezultat operacji zwanej **overlay** (nakładanie). Jest to najważniejsza operacja analizy wektorowej, pozwalająca na przeprowadzenie wielu przestrzennych analiz. Widok nałożonych warstw jest pokazany na rysunku 5.5 A.



Rysunek 5.5 Operacje nakładania dwóch warstw poligonów

Takie nałożenie warstw w GIS tworzy nową warstwę, zawierającą oddzielne poligony będące rezultatem ich wzajemnego przecinania się. Każdy z tych poligonów może pochodzić z jednej warstwy lub być pokryty przez dwie warstwy, czego rezultatem będą jego atrybuty. Istnieją różne operacje nakładania wynikające ze sposobu wyboru części obiektów, na podstawie ich wzajemnej relacji.

Podstawową operacją, która zachowuje wszystkie poligony będące rezultatem ich nałożenia i przecinania jest **UNION** (Rys. 5.5 D). W tablicy atrybutowej każdemu wyodrębnionemu poligoniowi przypisany jest rekord z kompletom pól z obu warstw. Poligon posiada atrybuty z poligonów, które go utworzyły. Jeśli jest pokryty tylko przez poligon jednej warstwy, wartość atrybutu w polach pochodzących z drugiej będzie przypisana jako NULL lub -1. Rezultaty wszystkich

pozostałych operacji mogą być pozyskane z UNION za pomocą selekcji atrybutowej odpowiedniego zbioru rekordów (obiektów).

Operacja **INTERSECT** (INTERSECTION Rys. 5.5 B) zawiera wyłącznie obiekty, które powstały z obiektów obu warstw. Będą to więc obiekty o rekordach bez wartości NULL lub -1. Jest to najprawdopodobniej najczęściej wykonywana operacja, gdyż bardzo często poszukujemy obszary spełniające jednocześnie dwa warunki. Istnieje podobieństwo pomiędzy tą operacją, a działaniem narzędzia CLIP. Obie operacje tworzą identyczną geometrię, ale w odróżnieniu do CLIP operacja INTERSECT zachowuje atrybuty z obu warstw.

Operacja **DIFFERENCE** lub **ERASE** (Rys. 5.5 C) zachowuje wszystkie obiekty pierwszej warstwy, które nie posiadają pokrycia w drugiej warstwie. Jest to operacja o odwrotnym działaniu niż CLIP. Przycina obiekty poligonami i pozostawia te, które znajdowały się poza wycinającymi poligonami (usuwa pozostałe – stąd ERASE).

Ostatnia z operacji **SYMMETRICAL DIFFERENCE** (Rys. 5.5 E) jest operacją odwrotną do INTERSECT pozostawia tylko obiekty, które nie są wspólne dla obu przecinających się warstw.

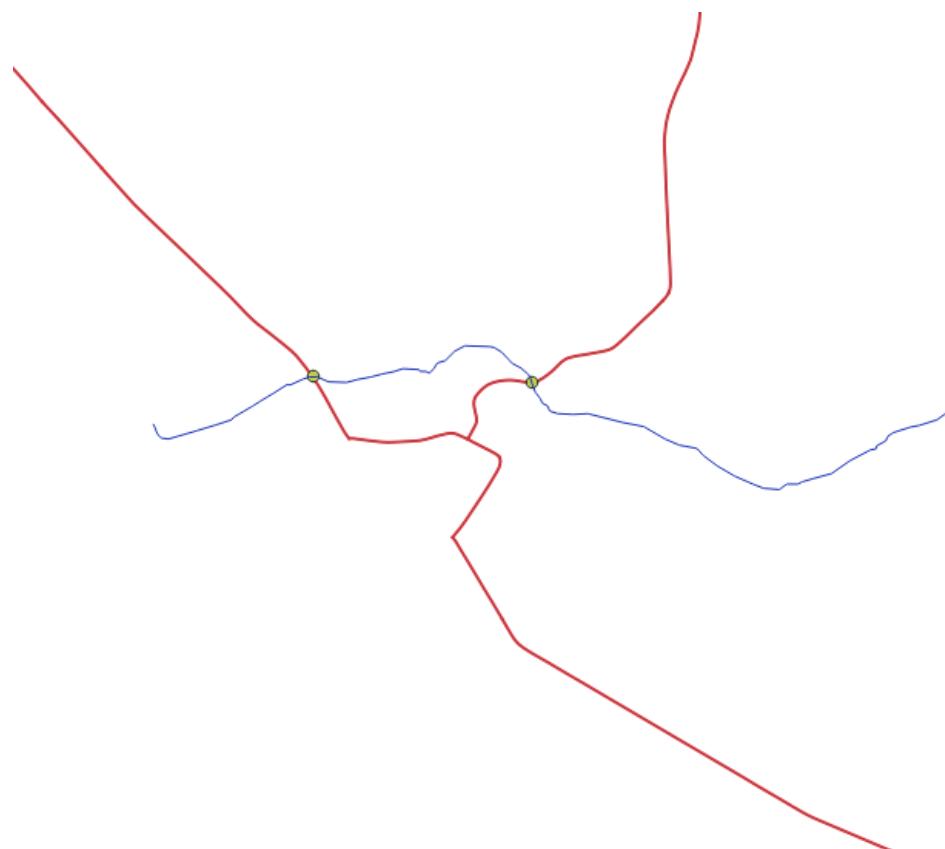
Jak wspomniano wyżej te operacje dotyczą warstw poligonów. Istnieje jednak często potrzeba wyznaczania obiektów będących rezultatem przecinania się warstw linii i poligonów. Przykładowo, mamy warstwę dróg (linie) i warstwę pokrycia terenu (poligony). Chcemy obliczyć sumaryczną długość dróg dla każdej klasy lub wyznaczyć najdłuższe odcinki dróg w każdym rodzaju pokrycia terenu. Drogi muszą być przycięte granicami poligonów i muszą otrzymać atrybut z poligonom w którym się znajdują. Zagadnieniem związanym z przecinaniem się warstw jest także otrzymanie warstwy punktów w miejscach przecinania się obiektów liniowych jednej warstwy lub dwu warstw. Ze względu na wagę tych metod, a jednocześnie specyficzne ich rozwiązanie w ArcGIS i QGIS zostaną one omówione oddziennie dla tych dwóch programów.

Operacja INTERSECT charakteryzuje się dodatkową funkcjonalnością. W praktyce możemy do niej wprowadzić warstwy różnych typów, a także tą samą warstwę dwukrotnie. W programie ArcGIS rezultatem może być warstwa dowolnego typu, ale nie wyższego pod względem wymiaru (poligony 2D, linie 1D, punkty 0D) niż warstwa o najniższym wymiarze. Daje to wiele kombinacji możliwych zastosowań.

W programie QGIS w narzędziu **INTERSECTION** geometria wyniku zależy od rodzaju i kolejności argumentów (warstw). Wspólna część poligonów i linii daje

puste poligony, a linii i poligonów linie. Kolejność nie ma znaczenia gdy argumentem są punkty, gdyż rezultatem jest zawsze warstwa punktów.

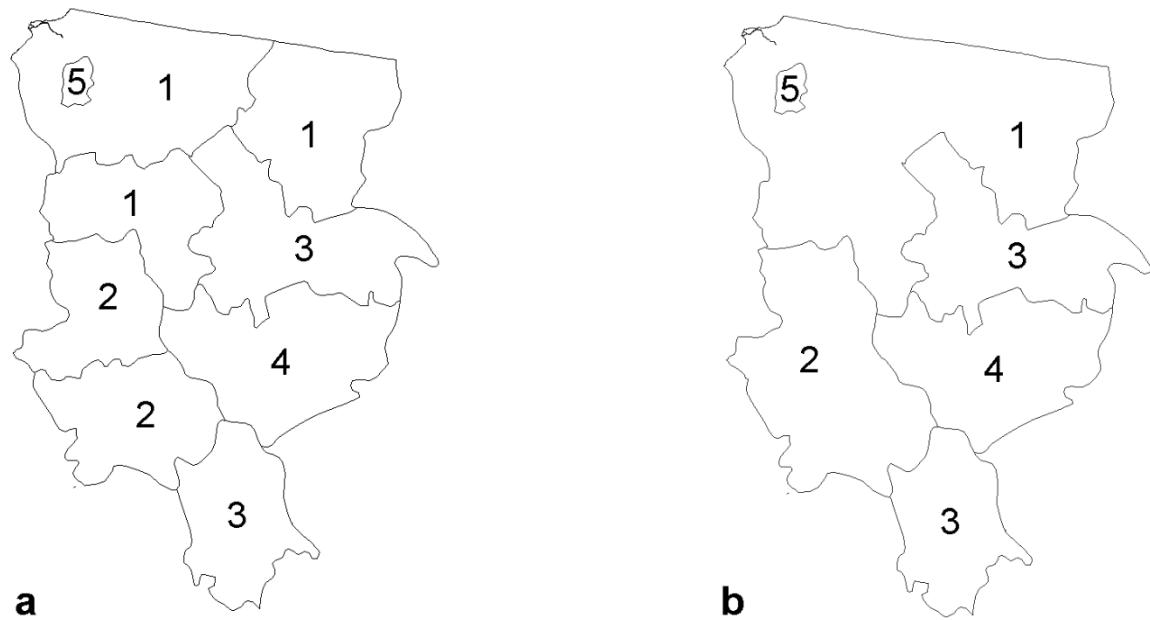
Warstwę punktów przecięcia linii z liniami lub obrysami poligonów można uzyskać w QGIS za pomocą narzędzia **LINE INTERSECTION**. Wymaga to jednak wcześniejszej ewentualnej konwersji warstwy poligonów do linii. Narzędzie to można także zastosować do znalezienia wszystkich przecięć linii w pojedynczej warstwie. Wymaga to wprowadzenia tej samej warstwy dwukrotnie.



Rysunek 5.6 Wyznaczenie punktów przecinania się linii za pomocą narzędzia **Line intersection**

Dodatkowo w QGIS istnieje możliwość rozdzielenia poligonów za pomocą linii za pomocą narzędzia **Split With Lines**.

Operacja lub narzędzie **Dissolve** daje możliwość łączenia obiektów o wspólnych granicach na podstawie identycznych wartości w wybranym polu. Dodatkowo istnieje możliwość agregacji atrybutów w łączonych obiektach. W QGIS istnieją dwa narzędzia DISSOLVE, podstawowe i drugie w skrzynce narzędziowej biblioteki GDAL.



Rysunek 5.7 Proste łączenie poligonów za pomocą Dissolve

Etykiety poligonów odpowiadają wartościom w wybranym polu tabeli atrybutowej. Poligony, które graniczą ze sobą i mają taki sam atrybut zostają zespولone w jeden poligon i rekord.

Podstawowe narzędzia analizy wektorowej [ArcMap, ArcGIS Pro]

----- *bez tworzenie nowej geometrii* -----

OPERACJE w TABLICY atrybutowej **Statistics, Calculate Field, Summary Statistics, Calculate Geometry Attributes, Sort**

SELEKCJA atrybutowa **Select Layer by Atribute, Select, Table Select**

Selekcja na podstawie relacji przestrzennych **Select Layer by Location**

POZYSKIWANIE ATRYBUTÓW przez łączenie tablic - **Join Field**

Pozyskiwanie atrybutów na podstawie relacji przestrzennych **Spatial Join**

----- *z tworzeniem nowej geometrii* -----

Bliskość (Proximity) **Buffer, Near, Generate Near Table**

Wycinanie (Extract) **Clip**

Nakładanie (Overlay) **Intersect, Union, Erase**

Agregacja **Dissolve**

Łączenie warstw **Merge**

Zmiana geometrii **Multipart to Single Part**

Konwersja wektorowa **Feature to ... Point, Line, Polygon**

Obrysy **Minimum Bounding Geometry**

Podstawowe narzędzia analizy wektorowej [QGIS 3]

----- *bez tworzenie nowej geometrii* -----

OPERACJE w TABLICY atrybutowej **Basic statistics for fields, Field Calculator, Statistics by Categories, Add Geometry Attributes**

SELEKCJA atrybutowa **Select by attribute, Select by expression**

Selekcja na podstawie relacji przestrzennych **Select by location**

POZYSKIWANIE ATRYBUTÓW przez łączenie tablic - **Join attributes by field value**

Pozyskiwanie atrybutów na podstawie relacji przestrzennych **Join attributes by location, Join attributes by location (summary)**

----- *z tworzeniem nowej geometrii* -----

Bliskość (Proximity) **Buffer, Fixed Distance Buffer,**

Variable Distance Buffer

Distance Matrix, NNJoin (wytyczka)

Wycinanie (Extract) **Clip**

Nakładanie (Overlay) **Intersection, Line intersection, Split With Lines**

Union, Difference

Agregacja **Dissolve**

Łączenie warstw **Merge Vector Layers**

Zmiana geometrii **Multipart to singleparts**

Konwersja wektorowa **Lines to Polygons, Polygons to Lines, Centroids**

Obrysy **Minimum Bounding Geometry**

Pytania:

dane: POMORSKIE_v1.gdb

1. Dla jeziora Rzuno wynacz w buforze 1 km powierzchnię w hektarach różnego rodzaju pokrycia terenu.
2. Rzekę Reknicę (ID_HYD_R=48686) otocz buforem 2 km. Jaką powierzchnię w tym buforze w ha zajmują lasy.
3. Jak jest łączna długość odcinków rzeki Reknicy, płynących przez las.
4. Utwórz warstwę mostów drogowych (punkty) na rzece Łebie.
5. Z warstw miast utwórz jedną warstwę poligonów, punkty należy zamienić na koła o promieniu w metrach równym liczba_mieszk /3.
6. Utwórz oddzielną warstwę pokrycia terenu okolic Kartuz o wymiarze 10 x 10 km, zastąp zabudowę w Kartuzach poligonem z warstwy miasta_pow2000_vmap
7. Do każdego miasta pon. 2000 w wybranym powiecie przypisz rodzaj pokrycia terenu o największej powierzchni w promieniu 2 km.

dane: Ocean_ZG22.gdb

8. Określ jakie punkty pomiarowe leżą najbliżej siebie i najdalej. Podaj te odległości. Utwórz podzbiór punktów pomiarowych, dla których do najbliższego punktu jest dalej niż 3000 m, uporządkuj je po odległościach od największej do najmniejszej.
9. Wyznacz na Zatoce Puckiej obszary, które leżą dalej niż 3 km od najbliższego punktu pomiarowego,
10. Jaka jest procentowa zawartość (żwiru, piasków średnio i gruboziarnistych) w ZP w trzech strefach licząc od brzegu 0-1000, 1000-2000, 2000-3000.
11. Jaki jest procentowy udział różnych rodzajów osadów na transekcje Cypel Redłowski – Cypel Helski
12. Utwórz mapę buforów w ujściach rzecznych Zatoki Gdańskiej (bez Zalewu Wiślanego), Wisła – 8 km, Wisła Śmiała i Martwa Wisła 3 km,, Reda 1 km, pozostałe 500m. Utwórz warstwę punktów leżących od tych buforów nie dalej niż połowa ich rozmiaru.
13. Wykonaj mapę linii brzegowej ZP.
14. Do każdego punktu przypisz rodzaj osadu o największej powierzchni w buforze 5 km.

15. Do każdego punktu przypisz nazwę rzeki do którego ujścia ma najbliższej.
16. Poprowadź linię od Gdyni do Cypla Hel. Przenieś na nią (w rzucie prostokątnym) kopie punktów do wykonania przekroju (utwórz nową warstwę). Oryginalne punkty nie powinny leżeć dalej niż 3 km od linii przekroju.
17. Wykorzystaj warstwę punktów przekroju i wykonaj (w układzie lokalnym) średni przekrój temperatury w miesiącach letnich (lipiec, sierpień).

dane: Sycylia_18.gdb

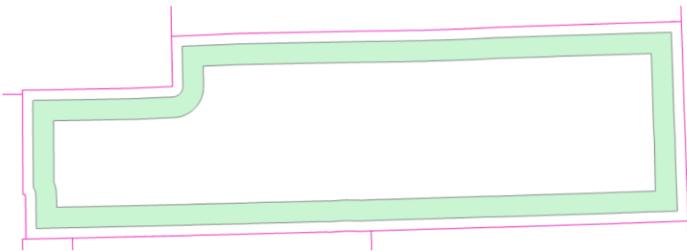
Zestaw C Analiza danych wektorowych (wyspa Pantelleria) (Rozwiążane <http://cgis.oig.ug.edu.pl/ZADANIA/>)

22. Wykonaj warstwę 2 kilometrowej strefy wód przybrzeżnych na około wyspy oraz kilometrowej strefy brzegowej lądu wyspy, umieść obie strefy w jednej warstwie. [FILM z ROZWIĄZANIEM zad. 15]
23. Wyznaczyć odcinki dróg, które leżą nie dalej niż 200 metrów od morza. [FILM z ROZWIĄZANIEM zad. 16]
24. Jaki jest procentowy udział różnych rodzajów pokrycia terenu, przez które przebiegają drogi. [FILM z ROZWIĄZANIEM zad. 17]
25. Utwórz oddzielną warstwę poligonów zabudowy posortowaną według odległości do lotniska. [FILM z ROZWIĄZANIEM zad. 18]
26. Sklasyfikuj linię brzegową na następujące klasy: 1 – porty; 2 – zabudowa; 3 – lasy; 4 – winnice; 5 – sucholubna roślinność; 6 – inne. [FILM z ROZWIĄZANIEM zad. 19]
27. Określ jaki procent linii brzegowej jeziora graniczy z winnicami, lasem i obszarami rolniczymi. [FILM z ROZWIĄZANIEM zad. 20]
28. Jaka jest średnia odległość od morza drogi przebiegającej przez las. [FILM z ROZWIĄZANIEM zad. 21]
29. Za pomocą kolejnych przybliżeń wyznacz obszar ciągły o powierzchni 5 km^2 leżący jak najdalej od brzegu morza. [FILM z ROZWIĄZANIEM zad. 22]
30. Pokryj wyspę Pantelleria siatką kwadratów o powierzchni 1 km^2 . Każdemu kwadratowi przypisz gęstość dróg w km/km^2 . [FILM z ROZWIĄZANIEM zad. 23]
31. Wyznacz obszary winnic, które spełniają następujące warunki: leżą nie dalej niż 4 km od morza, dalej niż 200 m od dróg i dalej niż 300 m od zabudowy. [FILM z ROZWIĄZANIEM zad. 24]

Dane: Colorado2.gdb

32. Przez ile obszarów spisowych (warstwa: demografia_census) przepływa rzeka Colorado (warstwa: Crzki, pole Name = Colorado River).
33. Jaka jest całkowita długość najdłuższej rzeki w warstwie: Crzeka, określonej przez jej nazwę (pole: Name). Długość podaj w pełnych kilometrach zgodnie z zasadami matematycznymi (np. 671). Rzeki bez nazw należy pominąć.
34. Ile mieszkańców mieszka w miastach (warstwa: MiastaPOP, pole: POP), które znajdują się w odległości nie większej niż 90 km od rzeki South Platte River (warstwa: Crzki). Przez odległość do miasta rozumie się odległość do jego najbliższego od rzeki punktu
35. Ile źródeł geotermalnych z warstwy Cwydobycie (pole: commod1 = 'Geothermal') posiada lokalizację w miejscu występowania piaskowców (ang. sandstone) - warstwa Cgeologia (pola ROCKTYPE1 lub ROCKTYPE2).
36. Jaka jest średnia odległość lotniska Denver International Airport (warstwa: lotniska) od pozostałych lotnisk. Podaj wynik w pełnych kilometrach zgodnie z zasadami matematycznymi.
37. W których trzech counties występuje najwięcej miejsc wydobycia (warstwa: Cwydobycie) rud złota (pole: ore) ? Złoto (Gold) może występować razem z innymi rudami (takie miejsce też się liczy). Która z sekwencji pokazuje prawdziwą kolejność (pierwsze, drugie, trzecie miejsce pod względem liczby miejsc wydobycia)
38. Który z parków narodowych (warstwa: ParkiN, pole: UNIT_TYPE='National Park) jest najbardziej odsunięty od większych miast (czyli miast, w których mieszka co najmniej 1 tys. mieszkańców)? (warstwa MiastaPOP, pole POP).
39. Jaka jest łączna długość systemu rzecznego rzeki South Platte River (warstwa: Crzki, pole: Name). System rzeczny tworzy rzeka główna wraz ze wszystkimi dopływami bezpośrednimi i pośrednimi, czyli wszystkie rzeki dorzecza South Platte River. Długość podaj w pełnych kilometrach bez jednostki np. 6231, wartość zaokrąglona zgodnie z zasadami matematycznym
40. Wewnątrz dowolnego county Z1_County (warstwa: ColoradoCounties, pole: NAME) wyznacz obszar badań w formie pierścienia o grubości 4 km oddalonego od granic county o 2 km (przykład na rysunku dla Kiowa county). Podaj powierzchnię piaskowców – sandstone (warstwa:

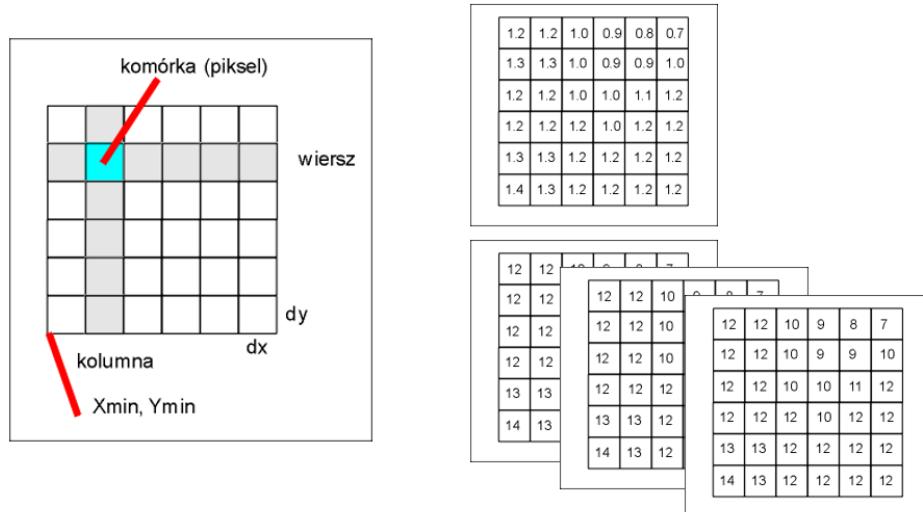
CGeologia, pole: ROCKTYPE1) w km² w obszarze badań (wartość powinna być zaokrąglona matematycznie do liczby całkowitej np. 102).



41. Wybrana rzeka (warstwa: Crzki, pole: Name) przepływa przez różnych rodzaju skały podłoża (warstwa: Cgeologia, pole: ROCKTYPE1). Podaj największą sumaryczną długość odcinków danej rzeki o tym samym podłożu. Wartość podaj w pełnych kilometrach. Przykład: Stwierdzono, że przez 122 km rzeka płynie sumarycznie przez piaskowiec – w trzech odcinkach oraz przez 82 km w dwóch odcinkach przez granit. Największa sumaryczna długość o tym samym podłożu to 122.
42. Jaki procent powierzchni wybranego county (warstwa: ColoradoCounties, pole: NAME) znajduje się dalej od najbliższego lotniska niż 30 km (lotnisko może być w innym county) (warstwa: lotniska). Procent podaj jako liczbę całkowitą (wartość powinna być zaokrąglona np. 22).
43. Wybierz z warstwy Cwydobycie rodzaj surowca. Określ rodzaj podstawowej skały podłoża (warstwa: Cgeologia, pole: ROCKTYPE1) w której istnieje największe prawdopodobieństwo znalezienia tego surowca (czyli jest on najczęściej znajdywany w tej skale), może on występować samodzielnie lub z innymi surowcami (warstwa: Cwydobycie, pola: commod1, commod2, commod3).
44. Każdemu miejscu wydobycia w wybranym county (warstwa: Cwydobycie) przypisz rodzaj skały (warstwa: Cgeologia, pole: ROCKTYPE1) , która najczęściej (ma największą powierzchnię) występuje w 5 km otoczeniu tego miejsca (otoczenie może wystawać poza granicę country). Jakiemu procentowi miejsc wydobycia została przypisana skała, o najczęstszym występowaniu. Procent podaj jako liczbę całkowitą (wartość powinna być zaokrąglona np. 22). Przykład: W county X są 4 miejsca wydobycia. Stwierdzono że największe powierzchnie w otoczeniu tych miejsc ma: 1 – granit; 2 – granit; 3 – granit; 4 – piaskowiec. Najczęściej przypisana skała to granit. Procent miejsc wydobycia, którym został przypisany granit to $3 / 4 * 100 = 75$.

5.5 Podstawowe funkcje i metody analizy rastrowej

Podstawowe cechy danych rastrowych, zostały podsumowane na rysunku poniżej.



Rysunek 5.9 Podstawowe cechy rastra

Warstwa rastrowa składa się z komórek rastrowych (pikseli) tworzących mozaikę w postaci siatki. Wielkość każdej komórki jest zdefiniowana długością jej boków na powierzchni Ziemi. Długość boku komórki wyrażona w metrach określa **rozdzielcość przestrzenną rastra**. Raster o rozdzielcości przestrzennej 10 metrów posiada komórki, które położone na powierzchni Ziemi miałyby rozmiar 10 na 10 metrów. Raster posiada określoną liczbę wierszy i kolumn. Boki komórek rastra są równoległe do osi układu współrzędnych w którym jest określone jego położenie. Raster jest dowiązany do lokalizacji na powierzchni Ziemi przez podanie współrzędnych jednego z jego rogów (najczęściej lewego dolnego rogu). W każdej komórce rastra znajduje się tylko jedna liczba (całkowita albo rzeczywista). To czy dany raster zawiera (jest typu) liczby całkowite czy rzeczywiste odgrywa znaczenie w procesie analitycznym i jego właściwościach. W rastrze typu całkowitego (integer) liczby opisują najczęściej dane jakościowe (pełnią rolę identyfikatorów np. opisuje kod osadu) lub przedstawiają dane otrzymane w rezultacie policzenia obiektów. Raster typu rzeczywistego zawiera najczęściej dane ilościowe (np. temperaturę czy wysokość terenu).

Tylko raster zawierający dane całkowite może posiadać tablicę atrybutową.

VAT – Value Attribute Table

The diagram shows a 7x9 raster grid on the left and a VAT table on the right. The VAT table has columns for ObjectID, Value, and Count.

ObjectID	Value	Count
	0	6
	1	5
	2	9
	3	6
	4	5
	5	7
	6	5
	7	8
	8	5
	9	8

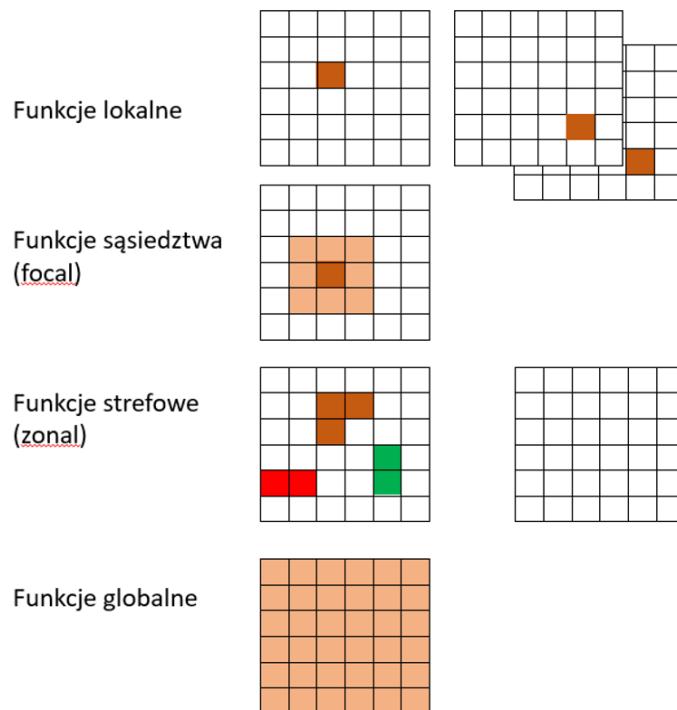
Rysunek 5.10 Raster zawierający dane całkowite może zawierać tablicę VAT

Tablica ta ma właściwości zwykłej tablicy atrybutowej i zawiera tyle rekordów ile jest w tablicy unikalnych wartości liczb całkowitych (pole Value). Tablica ta posiada dodatkowo pole Count, w którym określona jest liczliwość poszczególnych liczb całkowitych. Raster może posiadać w komórkach wartości NoData. W praktyce zawierają one zdefiniowaną wartość, która pełni rolę wartości NoData. Piksele z tą wartością nie są najczęściej wyświetlane i najczęściej nie uczestniczą w operacjach analitycznych (zwracają zwykle także wartość NoData).

Charakterystyczną cechą analizy rastrowej jest potrzeba zdefiniowania środowiska tej analizy. Podstawowymi elementami, które je określają (w zależności od oprogramowania może to być część wymienionych elementów) są: rozmiar komórki, maska, rozmiary rastra, zapewnienie zgodności z rastrem już istniejącym oraz układ współrzędny. **Rozmiar komórki** (cell size) rastra wynikowego domyślnie będzie odpowiadał największemu rozmiarowi komórki z rastrów biorących udział w analizie. Jeżeli chcemy aby raster wynikowy miał inny rozmiar komórki możemy to wymusić. Jest to potrzebne przykładowo przy tworzeniu kompozytu z jedno kanałowych zdjęć satelity Sentinel. W praktyce każdy kanał o rozdzielczości przestrzennej większej niż 10 metrów powinien zostać sprowadzony do tej rozdzielczości przed łączeniem ich w kompozyt. Jako **maska** (mask) może występować dowolny raster zawierający wartości NoData. W rastrze wynikowym wartości NoData znajdują się w tych samych miejscach co na masce. **Rozmiar rastra** (extent) jest określony zakresem współrzędnych na obu osiach. Aby obrysły komórek były zgodne (**snap raster**) należy wprowadzić

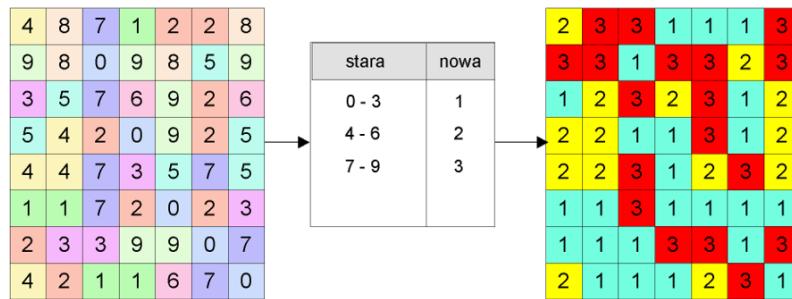
raster wzorcowy. Jest to istotne przy tworzeniu kompozytu z rastrów pochodzących oryginalnie z rastrów o różnej rozdzielczości. Ostatnim elementem środowiska jest układ współrzędnych rastra wynikowego (**output coordinates**).

Funkcje rastrowe mogą być podzielone na cztery główne rodzaje.



Rysunek 5.11 Rodzaje funkcji rastrowych: lokalne, sąsiedztwa, strefowe i globalne

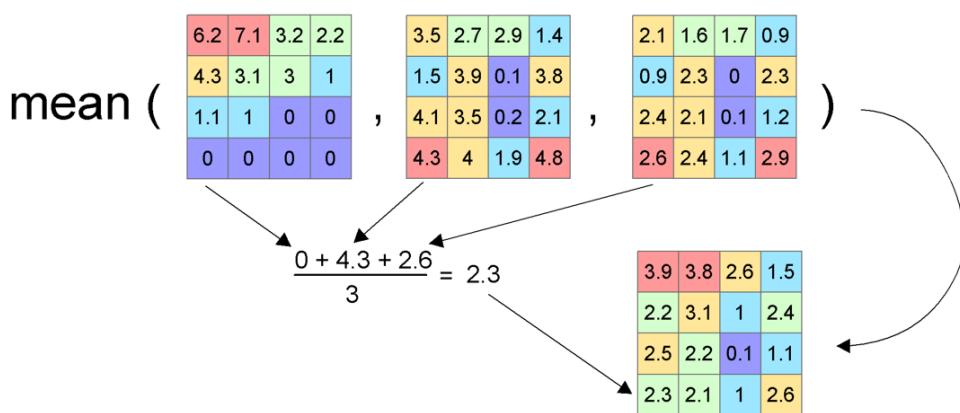
Funkcje lokalne charakteryzują się wykonywaniem niezależnych operacji oddzielnie w każdej komórce rastrowej, albo w odpowiadających sobie komórkach różnych rastrów. Podstawową grupą operacji są działania opisane wcześniej realizowane za pomocą kalkulatora rastrowego. Funkcją powszechnie znaną w analizie rastrowej jest funkcja **reklasyfikacji** (**RECLASSIFY** w ArcGIS Pro i **RECLASSIFY BY TABLE** w QGIS). Reklasyfikacja może być przeprowadzona za pomocą kalkulatora rastrowego. Zwykle GIS posiada specjalne narzędzie do przeprowadzenia tej operacji w sposób interaktywny. Reklasyfikacja polega na przypisaniu komórkom rastrowym rastra wynikowego nowych wartości na podstawie wartości w starym rastrze. Sposób tego przypisania jest przedstawiony najczęściej w tablicy zawierającej kolumnę z wartościami starego rastra lub zakresami wartości, oraz z przypisanymi im wartościami nowego rastra.



Rysunek 5.12 Reklasyfikacja

Na Rysunku 5.12 przedstawiono zasadę działania reklasyfikacji. Raster z lewej strony zawiera wartości od 1 do 9. Tablica reklasyfikacyjna przypisuje im wartości 1,2,3. Reklasyfikacja daje też możliwość przypisywania do rastra wartości NoData. Reklasyfikacja jest stosowana do tworzenia map tematycznych z map ilościowych przy wykorzystaniu założonego schematu klasyfikacyjnego. Przykładowo mając mapę wysokości terenu można ją sklasyfikować na mapę tematyczną zawierającą depresje (-4 – 0 m przypisz 1), niziny (0 – 200 m przypisz 2) i wyżyny (ponad 200 m przypisz 3). Tego typu mapy tematyczne mogą być przekształcane z wykorzystaniem innego schematu klasyfikacyjnego. Specjalnym rodzajem map są mapy zawierające tylko wartości 1 i 0 (lub 1 i NoData). Mogą być one wykorzystywane także przez inne funkcje analizy rastrowej. Taka mapa jak z prawej strony Rysunku 5.12 może pełnić rolę **mapy stref** wyznaczając obszary 3 klas. Reklasyfikacja daje możliwość tworzenia tego typu rastrów.

Drugą funkcją o lokalnym charakterze jest **statystyka komórek** (**CELL STATISTICS** w ArcGIS Pro i **r.series** w QGIS). Funkcja ta i związane z nią narzędzia służą do



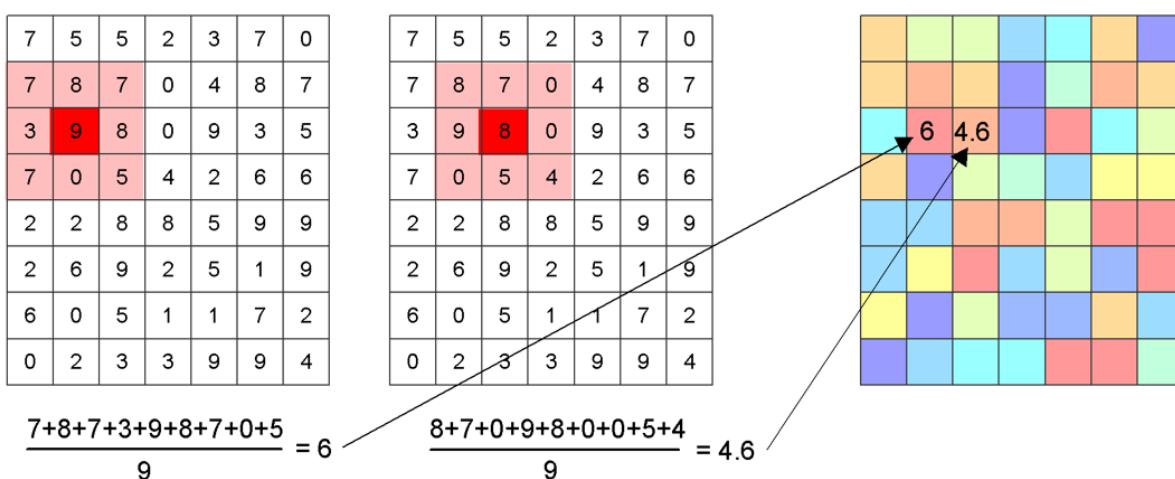
Rysunek 5.13 Obliczanie średniego rastra za pomocą

obliczania rastrów zawierających statystyki z odpowiadającymi sobie komórkami. W zależności od oprogramowania mogą być obliczane różne statystyki (średnia, mediana, odchylenie standardowe, zakres, maksimum, minimum, suma). Dla

wartości całkowitych dodatkowo mogą być wyznaczane: najczęstsza wartość (MAJORITY), najrzadsza wartość (MINORITY) oraz różnorodność rozumiana jako liczba występujących wartości (VARIETY / DIVERSITY).

Podstawową cechą **funkcji sąsiedztwa** jest to, że wartość komórki na mapie wynikowej zależy wyłącznie od sąsiednich komórek tej samej warstwy, czyli operacje sąsiedztwa są zawsze wykonywane na pojedynczej warstwie. Istnieją dwie podstawowe funkcje sąsiedztwa **obliczanie statystyk** (**FOCAL STATISTICS** w ArcGIS Pro i **r.neighbors** w QGIS) i **grupowanie** (**REGION GROUP** w ArcGIS Pro i **r.clump** w QGIS).

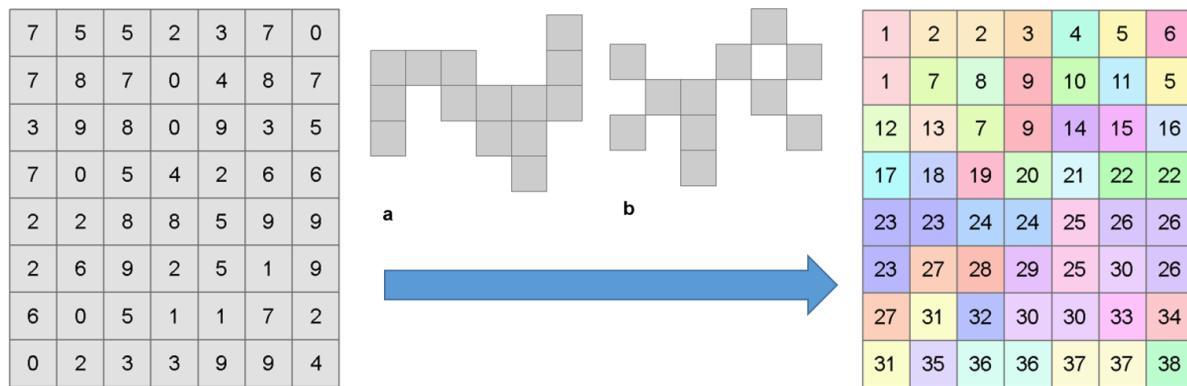
Obliczanie statystyk jest realizowane za pomocą metody **ruchomego okna** (istnieje także inna metoda obliczania statystyk za pomocą bloków komórek).



Rysunek 5.14 Zastosowanie ruchomego okna 3×3 do tworzenia rastra wartości średnich

Metoda ruchomego okna została pokazana na rysunku 5.14 przy tworzeniu rastra wartości średniej. Polega ona na przesuwaniu kwadratowego okna o boku z nieparzystą liczbą komórek piksel po pikselu. Nowa wartość jest wyznaczana na podstawie statystyki z wszystkich wartości w oknie i przypisana do komórki rastra wynikowego odpowiadającej komórce środkowej okna. Obliczać można podobne statystyki jak w funkcji Cell Statistics. Funkcja ta może być traktowana jako metoda filtracji rastra. Często stosuje się ją do filtracji zdjęć satelitarnych w celu usunięcia wadliwych pikseli (medianą jako statystyka).

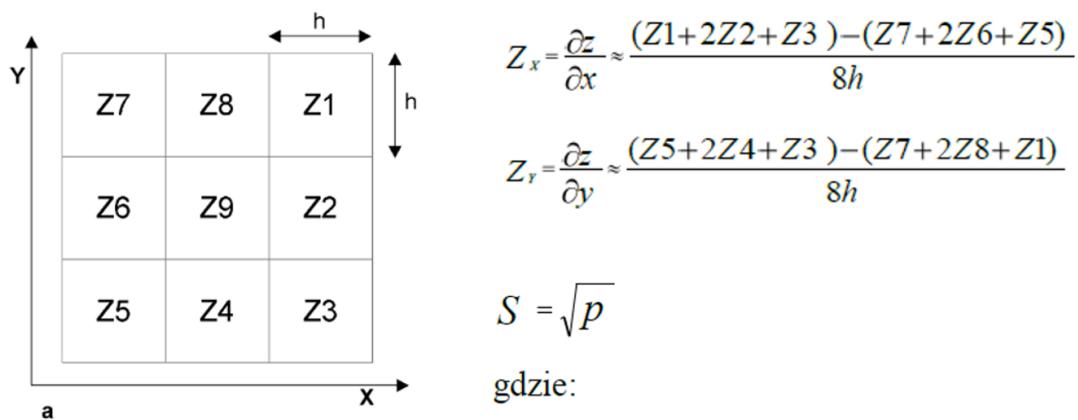
Funkcja grupowania polega na nadaniu unikalnego identyfikatora (liczby całkowitej) komórkom, które posiadają tą samą wartość i są ze sobą połączone. Połączenie może być zdefiniowane jako wspólny bok (max. czterech sąsiadów), albo wspólny bok lub róg (max. ośmioro sąsiadów).



Rysunek 5.15 Grupowanie może być zrealizowane metodą max. czterech sąsiadów (a) lub max. ośmioro sąsiadów (b)

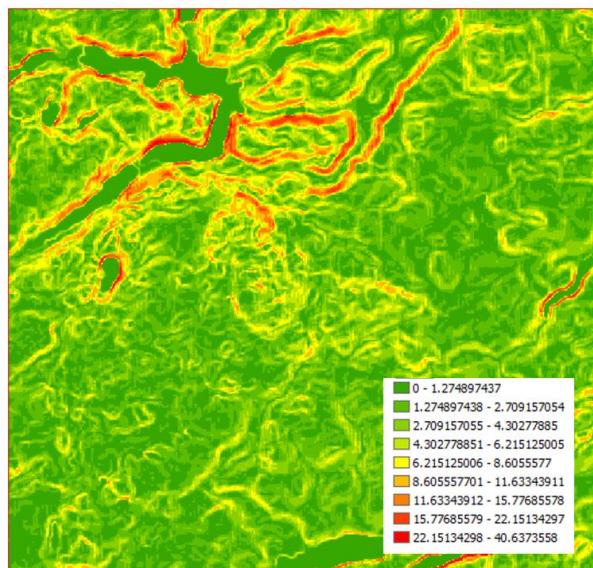
Na rysunku 5.15 pokazano jak działa grupowanie. Zostało ono przeprowadzone metodą ośmioro sąsiadów (za połączone uważa się komórki o tej samej wartości mające wspólny bok lub róg). Proces zaczyna się od lewej górnej komórki, której przypisano 1. Komórki połączone z nią o tej samej wartości otrzymają taki sam identyfikator tworząc grupę lub region. Następnie przechodzimy do kolejnej komórki w wierszu, która nie otrzymała identyfikatora. Cały proces jest realizowany dla wszystkich wierszy. Warstwa grup lub regionów może być traktowana jak warstwa stref.

Istnieją jeszcze dwie funkcje sąsiedztwa stosowane głównie do przetwarzania rastrowych warstw numerycznego modelu terenu (**DEM**). Są to funkcje do wyznaczenia nachylenia terenu (**SLOPE**) i kierunku nachylenia terenu zwanego także ekspozycją (**ASPECT**). **Nachylenie definiujemy jako maksymalne nachylenie terenu w danym miejscu.** Na rysunku 5.16 przedstawiono metodę obliczania nachylenia, która została zastosowana w ArcGIS Pro. Oprócz tej metody istnieje szereg innych, ale rezultaty w zasadzie powinny być podobne. Proces obliczania odbywa się za pomocą ruchomego okna o rozmiarach 3 x 3. Obliczany jest gradient wysokości wzduż osi X i osi Y. Następnie obliczany jest pierwiastek kwadratowy sumy kwadratów tych gradientów. Wynik odpowiada tangensowi kąta o maksymalnym nachyleniu. Aby obliczyć wartość tego kąta musimy obliczyć arcus tangens z tej wartości, Da nam to kąt wyrażony w radianach (kąt pełny w radianach wynosi $2 * \pi$). Zamiana na stopnie jest możliwa przez pomnożenie przez wartość 57.29578.



$$\text{Nachylenie w stopniach} = \arctan S * 57.29578$$

Rysunek 5.16 Metoda obliczania nachylenia terenu w modelu rastrowym

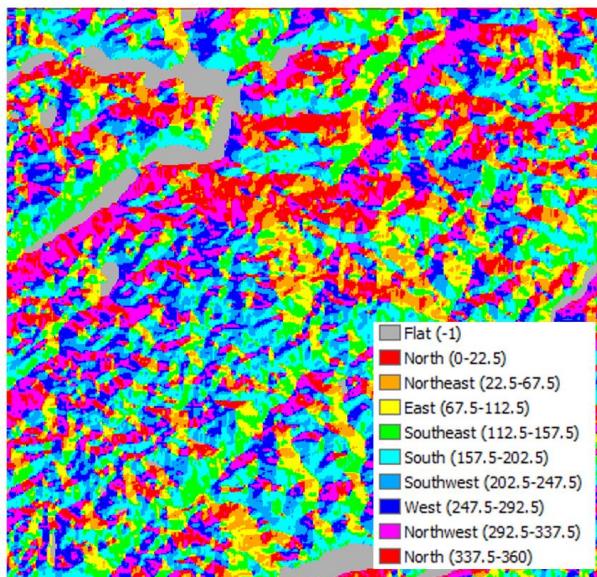


Rysunek 5.17 Rastrowa mapa nachylenia terenu (w stopniach)

Z nachyleniem terenu związana jest **ekspozycja** (elewacja), którą definiujemy najczęściej jako **azymut kąta o największym nachyleniu**. Wartość azymutu w stopniach (azymut to kąt liczyony od północy do danego kierunku w stronę zgodną z ruchem wskazówek zegara) można wyznaczyć z gradientów (liczonych analogicznie jak przy obliczaniu nachylenia) za pomocą następującego wyrażenia:

$$\Psi = 180 - \arctan \left(\frac{Z_y}{Z_x} \right) * 57.29578 + 90 \frac{Z_x}{|Z_x|}$$

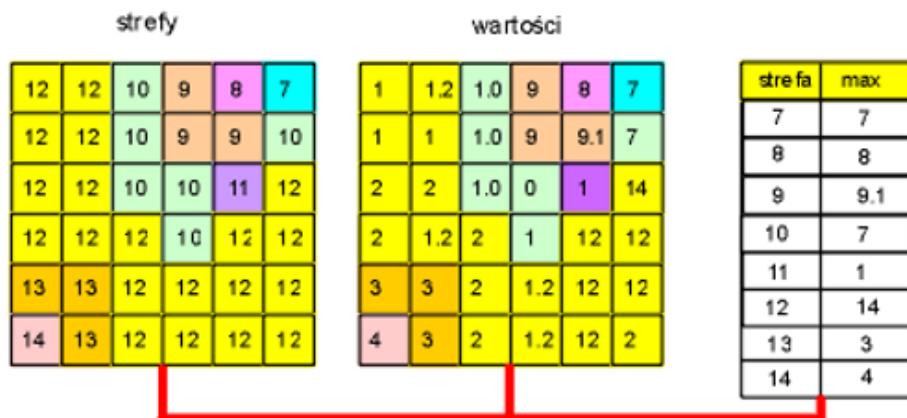
Poniżej przedstawiono przykładową mapę elewacji z symbolizacją do 9 kierunków.



Rysunek 5.18 Rastrowa mapa ekspozycji terenu

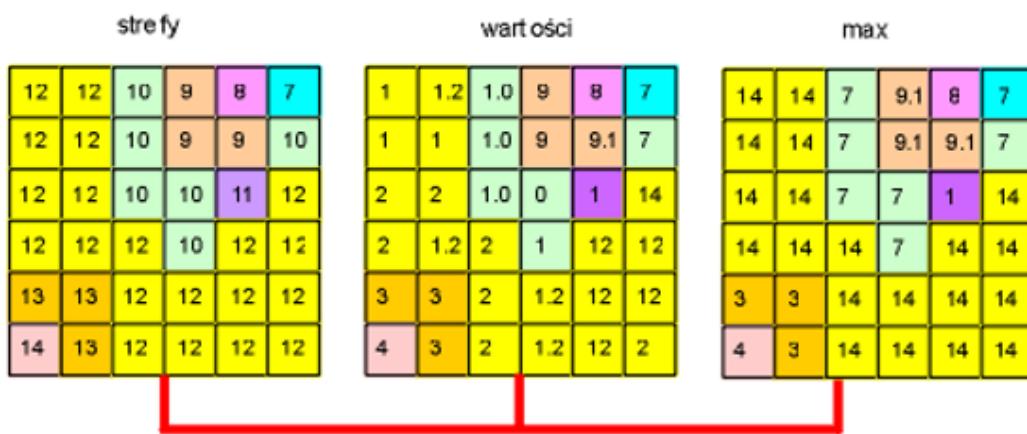
Funkcje **strefowe** wykorzystują rastry (lub wektorowe warstwy poligonów) stref, czyli warstwy wartości całkowitych, które definiują różne obszary lub obiekty. Warstwy stref mogą tworzyć tematyczną warstwę, na której liczbom całkowitym zostanie przypisana dowolna właściwość powierzchni Ziemi (np. woda, las, pola, powierzchnie nieprzepuszczalne) albo warstwę obiektów, na której każdy obiekt jest identyfikowany unikalną wartością rastra. Na przykład maska woda-ląd na której woda jest oznaczona jako 0 a ląd 1 wyznacza strefę lądu i jezior. Strefy te nie muszą zachowywać ciągłości (tak jak w powyższej masce). Taka warstwa może być utworzona za pomocą reklastryfikacji. Można sobie także wyobrazić warstwę jezior na której piksele każdego jeziora mają tą samą wartość całkowitą pełniącą rolę identyfikatora danego jeziora. Taka warstwa może być utworzona z maski woda-ląd za pomocą grupowania. Podstawowa grupa funkcji strefowych służy do wyznaczania statystyk zmiennej z warstwy rastrowej w strefach zdefiniowanych w warstwie stref. Statystyki są liczone z podzbioru pikseli rastra zmiennej, które znajdują się w danej strefie. Operacja jest przeprowadzana najczęściej na dwóch rastrach: rastrze stref i rastrze dowolnej zmiennej. Rezultatem takiej operacji może być tablica atrybutowa zawierająca w kolejnych rekordach identyfikatory stref i policzone dla nich statystyki (**ZONAL STATISTIC AS TABLE** w ArcGIS Pro i **Zonal Statistics** w QGIS). Zamiast warstwy rastrowej stref

może być użyta także warstwa poligonów (z identyfikatorami stref w dowolnym polu).



Rysunek 5.16 Funkcja strefowa obliczająca statystyki i zbierająca je w tablicy atrybutowej

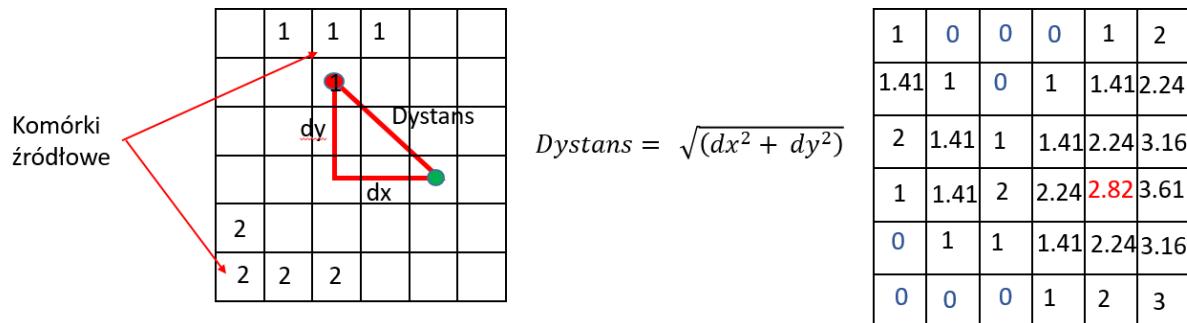
Funkcja strefowa obliczająca statystyki może też tworzyć nowy raster w którym każdej komórce zostanie przypisana statystyka zgodna z jej identyfikatorem. Tego typu funkcja (**ZONAL STATISTIC** w ArcGIS Pro i **r.stats.zonal** w QGIS) została przedstawiona na rysunku 5.17.



Rysunek 5.17 Funkcja strefowa obliczająca statystyki i wpisującą je w nowy raster

Oprócz strefowych funkcji służących do wyznaczania statystyk w strefach istnieją funkcje, które wyznaczają powierzchnie stref (zarówno jako tablice atrybutowe jak i nowy raster). Funkcje statystyczne pozwalają dodatkowo na wyznaczanie powierzchni (ArcGIS Pro w Zonal Statistic as Table) lub za pomocą „statystyki” liczebności (QGIS).

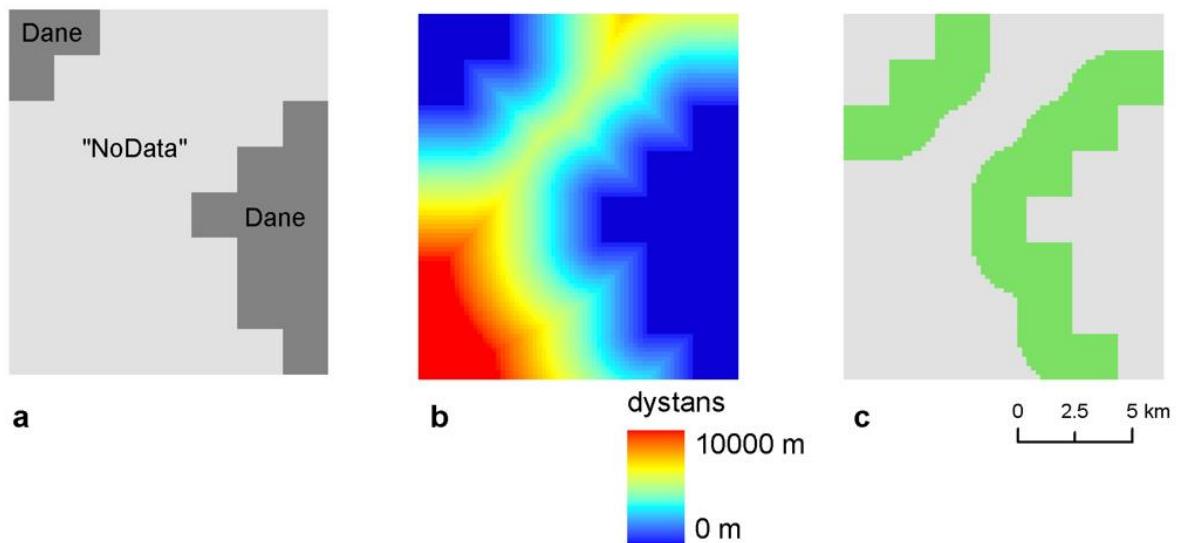
Najważniejszą funkcją globalną jest funkcja **dystansu** (**EUCLIDIAN DISTANCE** w ArcGIS Pro i **Proximity** w QGIS). Sposób działania funkcji został przedstawiony na rysunku 15.18.



Rysunek 5.18 Działanie rastrowej funkcji dystansu

Podstawą koncepcji tworzenia rastra dystansu jest podział komórek rastra będącego parametrem tej funkcji na komórki źródłowe (*source* w ArcGIS Pro lub *target* w QGIS) i pozostałe. Komórki źródłowe w ArcGIS Pro są zdefiniowane jako komórki posiadające dowolną wartość (czyli pozostałymi będą komórki *NoData*), natomiast komórkami źródłowymi w narzędziu Proximity (QGIS – biblioteka GDAL) są wszystkie komórki różne od zera. Dla każdej komórki rastra obliczana jest odległość (za pomocą wzoru na rysunku 5.18) od jej środka do środka najbliższej komórki źródłowej. Na rysunku raster z lewej strony zawiera komórki źródłowe z wartościami 1 i 2. Pozostałe komórki mają wartość *NoData*. Przyjęto, że bok komórki równa się 1m. Dla każdej komórki określono odległość (po prostej) od jej punktu środkowego do środka najbliższej komórki źródłowej. Komórki odpowiadające pod względem położenia komórkom źródłowym mają przypisaną wartość 0.

Reklasyfikacja rastrów dystansu pozwala na tworzenie warstw rastrowych buforów. Na rysunku 5.19 pokazano w jaki sposób wykonać maskę obszarów przybrzeżnych zdefiniowaną jako komórki z wodą znajdujące się nie dalej od brzegu niż pewna graniczna wartość. Warstwa (a) przedstawia ląd (dane) i wodę (*NoData*). Dla warstwy tej wykonano raster dystansu (b). Następnie zreklasyfikowano raster dystansu przypisując 1 do odległości większych od zera i mniejszych niż wartość graniczna. Gdybyśmy chcieli aby bufor objął także ląd warunek powinien także obejmować wartość zero.



Rysunek 5.19 Wykorzystanie funkcji dystansu i reklasyfikacji do tworzenia rastrowej strefy buforowej

Podstawowe narzędzia analizy rastrowej [ArcMap, ArcGIS Pro]

----- *funkcje lokalne* -----

REKLASYFIKACJA **Reclassify**

CROSSTABULACJA **Combine**

STATYSTYKI (rastry) **Cell Statistics**

ALGEBRA MAP **Raster Calculator**

----- *funkcje sąsiedztwa* -----

STATYSTYKI (raster) **Focal Statistics**

GRUPOWANIE **Region Group**

----- *funkcje strefowe* -----

STATYSTYKI STREF **Zonal Statistics As Table, Zonal Statistics**

GEOMETRIA STREF **Zonal Geometry As Table, Zonal Geometry**

----- *funkcje globalne* -----

DYSTANS **Euclidian Distance**

ALOKACJA **Euclidian Allocation**

----- *funkcje DEM* -----

NACHYLENIE TERENU **Slope**

EKSPOZYCJA TERENU **Aspect**

----- *rastry z punktów* -----

GĘSTOŚĆ **Kernel Density**

INTERPOLACJA **IDW**

Podstawowe narzędzia analizy rastrowej [QGIS 3]

funkcje lokalne

REKLASYFIKACJA **Reclassify By Table**

STATYSTYKI (rastry) **r.series**

ALGEBRA MAP **Raster Calculator**

funkcje sąsiedztwa

STATYSTYKI (raster) **r.neighbors**

GRUPOWANIE **r.clump**

funkcje strefowe

STATYSTYKI STREF **r.stats.zonal, Zonal Statistics**

funkcje globalne

DYSTANS **Proximity**

funkcje DEM

NACHYLENIE TERENU **Slope**

EKSPOZYCJA TERENU **Aspect**

rastry z punktów

GĘSTOŚĆ **Kernel Density Estimation**

INTERPOLACJA **IDW interpolation**

Pytania:

dane: Sycylia_18.gdb

(Rozwiążane <http://cgis.oig.ug.edu.pl/ZADANIA/>)

1. Utwórz warstwę, która zawiera maksymalne ilości opadu w lipcu w oknie w formie koła o promieniu 10 km. [FILM z ROZWIĄZANIEM zad. 25]
2. Obliczyć całkowitą ilość deszczu w miesiącach letnich (czerwiec, lipiec, sierpień). [FILM z ROZWIĄZANIEM zad. 26]
3. Wykonaj mapę tematyczną pokrycia terenu na 3 klasy (powierzchnie sztuczne, wykorzystane rolniczo i naturalne). [FILM z ROZWIĄZANIEM zad. 27]
4. Wykonaj mapę powierzchni naturalnych większych od 100 km^2 . [FILM z ROZWIĄZANIEM zad. 28]
5. Sklasyfikuj te powierzchnie pod względem średniego opadu. [FILM z ROZWIĄZANIEM zad. 29]
6. Wyznacz obszary naturalne leżące dalej od powierzchni sztucznych niż 5 km. [FILM z ROZWIĄZANIEM zad. 30]
7. Przeprowadzić reklasyfikację mapy rastrowej pokrycia terenu *Sycylia_pokrycie500* do następujących klas:
1 – Zabudowa; 2 – Winnice, sady, oliwki
Do wizualizacji wykorzystaj symbolizację zapisaną w katalogu – *SicilyCommondata* (*Corine_raster_labels.lyrx*). [FILM z ROZWIĄZANIEM zad. 31]
8. Utwórz rastrową warstwę sumarycznego rocznego opadu na Sycylii. Dla całej wyspy oblicz następujące statystyki rocznego opadu: min, max, średnia, odchylenie standardowe. [FILM z ROZWIĄZANIEM zad. 32]
9. Przeprowadź reklasyfikację sumarycznej warstwy rocznych opadów do następujących klas:
1 – opady roczne $<$ średnia – 1 odchylenie standardowe
2 – opady roczne \geq średnia – 1 odchylenie standardowe i $<$ średnia
3- opady roczne \geq średnia i $<$ średnia + 1 odchylenie standardowe
4 - opady roczne \geq średnia + 1 odchylenie standardowe [FILM z ROZWIĄZANIEM zad. 33]
10. Wykonać warstwę zawierającą cztery klasy:
1 – Zabudowa z opadami $<$ średnia – 1 odchylenie standardowe
2 – Zabudowa z opadami \geq średnia + 1 odchylenie standardowe
3 - Winnice, sady, oliwki z opadami $<$ średnia – 1 odchylenie standardowe

4 - Winnice, sady, oliwki z opadami < średnia + 1 odchylenie standardowe
[FILM z ROZWIĄZANIEM zad. 34]

11. Wyznacz obszary Sycylii leżące dalej niż 25 km od ciągłej zabudowy, biorąc pod uwagę tylko zabudowę która ma powierzchnię większe niż 5 km^2 (do określenia zabudowy wykorzystaj warstwę *Sycylia_pokrycie500* na której zabudowa ma wartości 11 oraz maskę *Sycylia_maska500*). [FILM z ROZWIĄZANIEM zad. 35]

12. Określić średnie wysokości obszarów nadmorskich Sycylii w strefach 0 – 5 km; 5 – 10 km; 10 – 15 km; 15 – 20 km; 20 – 25 km od brzegu (wykorzystaj warstwy *Sycylia_dem500* i *Sycylia_maska500*). [FILM z ROZWIĄZANIEM zad. 36]

13. W jakiej odległości od brzegu i na jakiej wysokości nad poziomem morza leży najbardziej odległy od morza punkt Sycylii (do tego zadania należy użyć warstw o rozdzielczości 25 m). [FILM z ROZWIĄZANIEM zad. 37]

14. Wykonaj warstwę poziomic od 2000 m (co 50m) dla Etny, zamień je na poligony i zaproponuj symbolizację dla ich wysokości (do tego zadania należy użyć warstw o rozdzielczości 25 m). [FILM z ROZWIĄZANIEM zad. 38]

15. Oblicz średnią głębokość i objętość wody (w km^3) na akwenie ograniczonym przez równoleżniki $38^\circ 9' \text{ N}$ i $38^\circ 17' \text{ N}$ oraz południki $15^\circ 33' \text{ E}$ i $15^\circ 43' \text{ E}$. Wykonaj mapę izobat dla tego akwenu. Wykorzystaj warstwę *Sycylia_baty*. [FILM z ROZWIĄZANIEM zad. 39]

dane: SST3.gdb

16. Użyj funkcji Focal Statistics (mean, okno 6x6), aby wygładzić mapę temperatury. Sprawdź efekt, przeprowadzając następnie reklasyfikację do trzech przedziałów (tak jak poprzednio poniżej 10, 10-20, powyżej 20 st. C). Porównaj wyniki z reklasyfikacją wykonaną na wynikowym rastrze z zad. 2. (Należy pamiętać, że wartość 0 na lądzie też jest brana do obliczeń i może zakłócać wyniki)dane: *sst_maj*

17. Dotnij raster wynikowy z zad. 8 do identycznej linii brzegowej co raster oryginalny (oba rastry mają się idealnie pokrywać).

18. Każdej wyspie (w obszarze *maska_wyspy*) nadaj osobny identyfikator.dane: *sst_maj*, *maska_wyspy*

19. Utwórz raster, gdzie Australia będzie miała wartość 1, a pozostałe obszary lądowe i morze - NoData.dane: *sst_maj*

20. Oblicz łączną powierzchnię wysp w obszarze *maska_wyspy*. Jaką procentową część powierzchni łącznej zajmuje każda z wysp?

21. Biorąc pod uwagę raster temperatur z maja, policz o ile średnia temperatura morza na obszarze wyznaczonym przez raster maska_wyspy, różni się od średniej temperatury morza z całego rastra.
22. Utwórz mapę odległości od brzegów.
23. Utwórz buffer rastrowy o promieniu 1000 km od brzegów Australii.
- 24.. Oblicz średnią temperaturę (oraz inne statystyki) w tak wyznaczonej strefie przybrzeżnej (1000 km od brzegów Australii).
25. Utwórz raster wskazujący, do którego lądu jest najbliżej z każdego miejsca na wodzie.

dane: model_raster.gdbZad.

26. Określ, o ile zmniejszy się powierzchnia zbiornika, jeżeli poziom wody opadnie o 1,9 m. dane: jeziorobat
27. Określ średnią temperaturę powierzchni morza na obszarach płytowych 0-10 m oraz głębokości powyżej 50 m. dane: temp_pow, batymetria
28. Utwórz wykres zależności minimalnej temperatury jezior w zależności od ich powierzchni jeziora. dane: temp_all, jezioraID
29. Przypisz jeziorom (jezioraID) odległość do brzegu morza.
30. Wyznacz miejsca, które leżą nie dalej niż 20 km od brzegu i mają temperaturę mniejszą niż 8. dane: temp_pow
31. Utwórz zgeneralizowaną rastrową maskę ląd/woda archipelagu wysp o powierzchni większej niż 0,5 km².
 - a) dane: wyspy2
 - b) dane: wyspy

Dane: Colorado2.gdb

32. Jaka jest średnia wysokość powierzchni terenu pokrytej przez dowolnego rodzaju zabudowę w określonym county (wartość wysokości podaj w pełnych metrach zaokrąglonych matematycznie). Warstwy rastrowe: coldem, pokrycie (klasy pokrycia znajdują się w tablicy atrybutowej)
33. W określonym county obliczyć średnią powierzchnię obiektów zdefiniowanych jako: zabudowa o gęstym i średnim zagęszczeniu otoczona buforem 500 metrów. Obiekty tworzą piksele łączące się bokami lub rogami.

Warstwa rastrowa: pokrycie (klasy pokrycia znajdują się w tablicy atrybutowej)

34. Jaką powierzchnię w określonym county zajmują obszary lodu i śniegu o pokryciu pow. 25% (typ pokrycia) przynależne do ciągłych płytów śnieżno-lodowych, o tym typie pokrycia, o łącznej powierzchni równej lub większej od 5 km^2 . Ciągłość jest typu 8 elementów (piksele łączące się bokami lub rogami), wartość powierzchni podaj w pełnych kilometrach kwadratowych zaokrąglonych matematycznie. Warstwa rastrowa: pokrycie (klasy pokrycia znajdują się w tablicy atrybutowej)

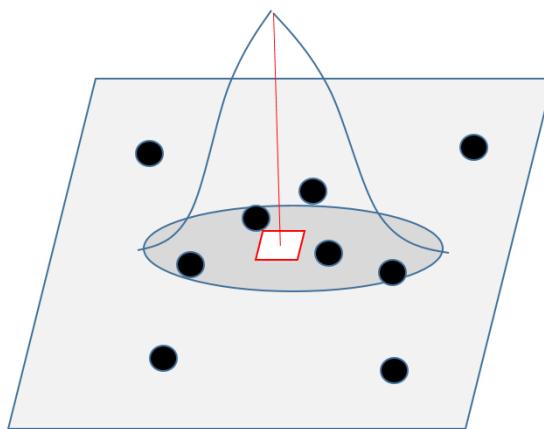
35. Jakie rodzaje pokrycia graniczą najczęściej w wybranym county z pastwiskami. Podaj kody pokrycia (pole Value, tablica warstwy pokrycie) dla trzech najczęściej występujących rodzajów. Podaj je w formacie: np. 42-43-95 (8 znaków bez przerw, przykładowy zapis oznacza, że najczęściej graniczą las iglasty, potem las mieszany, a na trzecim miejscu mokradła). Warstwa rastrowa: pokrycie (klasy pokrycia znajdują się w tablicy atrybutowej)

5.6 Metody interpolacji danych punktowych

W tym rozdziale omówione zostaną dwa sposoby zamiany danych punktowych na ciągłe. Dzięki temu otrzymujemy rastry, które mogą być wykorzystane w analizie rastrowej. Dane punktowe mogą mieć dwojaki sens, rejestracji występowania zjawiska albo punktowej wartości pewnej zmiennej. W tym pierwszym przypadku mogą to być lokalizacje dowolnego dyskretnego obiektu o charakterze naturalnym: obserwacji ornitologicznej, lokalizacji drzew lub źródeł oraz o charakterze społecznym: wystąpienia przestępstw lub lokalizacji sklepów. Z kolei wartość zmiennej może mieć charakter **jakościowy** (np. rodzaj pokrycia terenu, typ gleby lub osadu) albo **ilościowy** (np. liczba mieszkańców, zmierzona wartość zanieczyszczenia, wysokość terenu).

W rezultacie mamy dwa różne scenariusze analizy danych punktowych. Pierwszym jest tworzenie rastrowej warstwy **gęstości** występowania obiektów punktowych. Każdej komórce rastra zostaje przypisana miara opisująca zagęszczenie punktów w danym miejscu. Drugim jest prognoza wartości zmiennej (ilościowej lub jakościowej) w środkach komórek rastrowych na podstawie znajomości wartości tej zmiennej w otaczających punktach. Ten sposób prognozy (zwany mądrzej **estymacją**) nosi nazwę **interpolacji**.

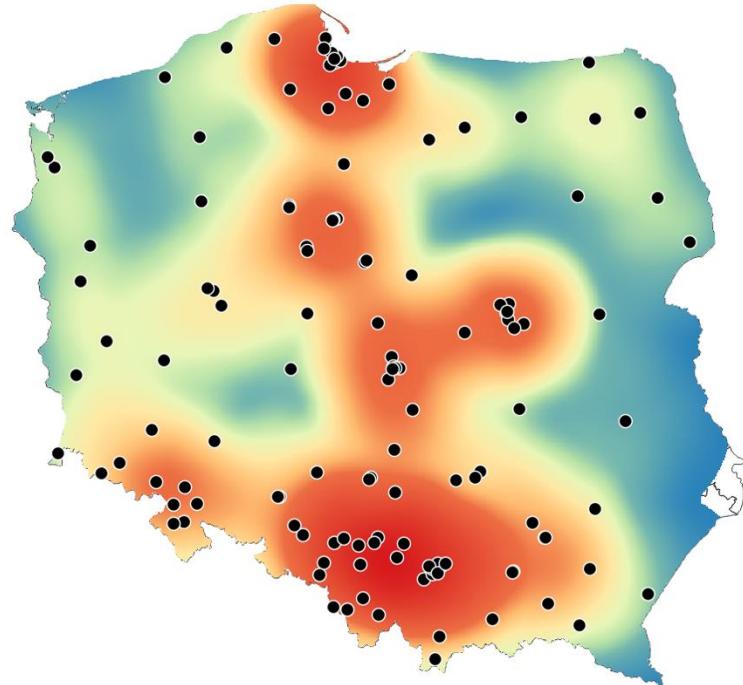
Mapy gęstości wykonuje się najczęściej stosując kernel (filtr) gęstości. Zasada działania tej techniki została przedstawiona na rysunku 5.20.



Rysunek 5.20 Zasada działania kernelu gęstości

Aby wyznaczyć wartość gęstości punktów w komórce rastra (czerwony kwadrat) bierze się pod uwagę wszystkie punkty w otoczeniu tej komórki (szara elipsa). Wartość obliczana jest z zastosowaniem wagi wyznaczanej przez funkcję opisaną

niebieską linią. Im punkt leży bliżej od komórki rastra tym większą ma wagę. W rezultacie otrzymuje się ciągłą powierzchnię, jak to pokazano na rysunku 21.



Rysunek 5.21 Mapa gęstości punktów pomiarowych

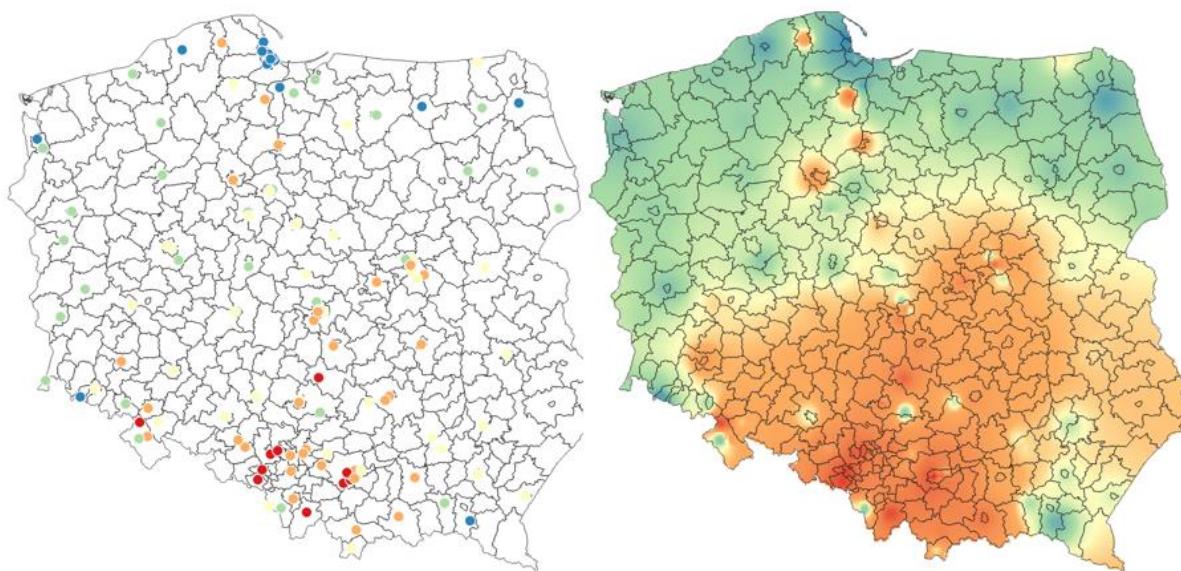
Mapy **gęstości** (zwane także Heat Map) wykonuje się narzędziem **KERNEL DENSITY** (ArcGIS Pro) lub **Kernel Density Estimation** (QGIS)

Metody **interpolacji** oparte są na założeniu lokalnej ciągłości danych przestrzennych. Ciągłość powoduje, że punkty leżące bliżej siebie mają bardziej zbliżone wartości atrybutów (zmiennych) niż punkty leżące daleko od siebie. Ta reguła wynika z charakteru powierzchni naturalnych lub antropogenicznych, które nie mają charakteru czysto losowego, ale powstają w rezultacie kompleksowych procesów fizycznych (np. dyfuzji zanieczyszczeń), biologicznych lub społecznych. Interpolacji dokonuje się na podstawie punktowych próbek zmiennej (warstwa punktowa). Jej celem jest określenie wartości tej zmiennej w miejscach niepróbkowanych. W praktyce miejscami niepróbkowanymi są środki komórek rastra. Istnieje wiele metod interpolacji. Wiele z nich oparta jest na metodzie ważonej ruchomej średniej. Polega ona na wyborze podzbioru punktów próbek (z wartościami zmiennej) w otoczeniu miejsca w którym ma być przeprowadzona interpolacja wartości. Następnie obliczona zostaje średnia ważona z wszystkich wartości z przypisaną wagą do każdego punktu. Waga zależy od położenia w przestrzeni punktu pomiarowego względem miejsca interpolacji.

Obliczanie odbywa się za pomocą wyrażenia:

$$f(x_i, y_j) = \frac{\sum(z(x_i, y_i) * \text{waga})}{\sum \text{waga}}$$

Podstawową metodą interpolacji obecną we wszystkich GIS jest metoda odwrotnych odległości (IDW – inverse distance weighted). W metodzie tej wagą jest odwrotność odległości. W ten sposób waga będzie tym większa im odległość pomiędzy miejscem interpolacji i wartości znanej jest mniejsza. Dodatkowo ten efekt może być zwiększyony przez podniesienie wagi do potęgi (najczęściej do kwadratu).



Rysunek 5.22 Z lewej strony dane punktowe z symbolizacją zmiennej ilościowej. Z prawej strony warstwa rastrowa tej zmiennej wykonana za pomocą interpolacji. Warstwa obrysu powiatów jest warstwą dodatkową

Interpolację metodą IDW przeprowadza się za pomocą **IDW interpolation** (QGIS) oraz **IDW** (ArcGIS Pro).

W przypadku danych jakościowych (atrybut stanowi klasa np. rodzaju gleby) korzysta się z tzw. interpolacji indykatorowej. Polega ona na stworzeniu rastrów prawdopodobieństwa wystąpienia każdej klasy jakościowej. Dla każdej klasy zamieniamy atrybuty na wartości 1,0 (indykatory). Jedynka oznacza wystąpienie danej klasy, zero brak. Interpolacja (np. IDW) daje warstwy prawdopodobieństwa wystąpienia danej klasy. Tworzy się je dla każdej klasy, a następnie dla każdej komórki rastra wynikowego wybiera klasę o największym prawdopodobieństwie.

Pytania:

Dane: Ocean_ZG22

1. Wykonaj warstwę punktów pomiarowych na Zatoce Gdańskiej (z temperatura_zasolenie1981_1999). W warstwie powinny znajdować się punkty o głębokości maksymalnej większej od 0 na których pomiary wykonano co najmniej w trakcie 15 rejsów (czyli co najmniej 15 razy wykonano pomiar na 0 m). Warstwa powinna zawierać pola: Nr_punktu, Liczba_rejsów, Max_głębokość.
2. Wykonaj mapę średniej temperatury wody w warstwie powierzchniowej na Zatoce Gdańskiej w okresie zimy (styczeń – luty) i lata (lipiec – sierpień).
3. Dla wybranego rejsu, któremu towarzyszył maksymalny rozpływ rzeki Wisły (najniższe średnie zasolenie w rejonie przyjściowym) wykonaj mapę zasolenia powierzchniowego.
4. Dla Zatoki Puckiej Zewnętrznej wykonaj mapę temperatury na 20 metrach w miesiącach (maj – czerwiec).