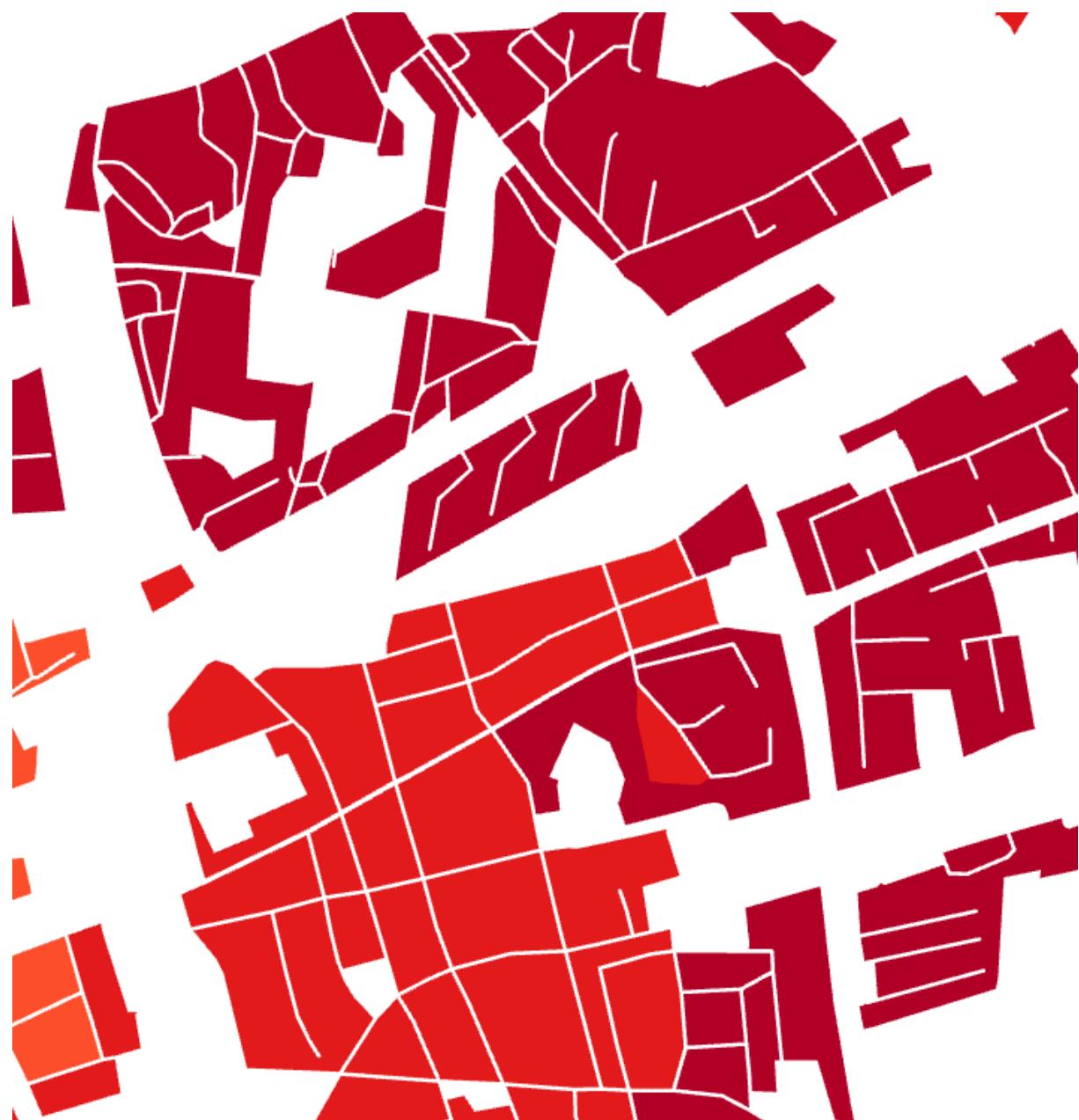


STO stron GIS



Jacek Urbański

Gdynia 2020

Spis treści.

Wstęp.

1. Współrzędne geograficzne. Lokalizacja.
2. Dane przestrzenne.
3. Współrzędne prostokątne. Odwzorowania.
4. Inżynieria danych przestrzennych.
 - 4.1 Tworzenie i pozyskiwanie danych przestrzennych
 - 4.2 Eksploracyjna analiza danych przestrzennych
 - 4.3 Podstawowe operacje przygotowania danych
 - 4.4 Obliczenia w tabeli atrybutowej (Python)
 - 4.5 Tworzenie i edycja geometrii danych wektorowych
 - 4.6 Georeferencja i rejestracja rastra
 - 4.7 Zmiana wartości w rastrze (Algebra Map, Raster Calculator)
5. Pozyskiwanie informacji przestrzennej (analiza danych).
 - 5.1 Pozyskiwanie atrybutów na podstawie relacji przestrzennych.
 - 5.2. Analiza odległości pomiędzy obiektami.
 - 5.3 Operacje na geometrii danych.
 - 5.4 Podstawowe funkcje i metody analizy rastrowej.
 - 5.5 Metody interpolacji danych punktowych.

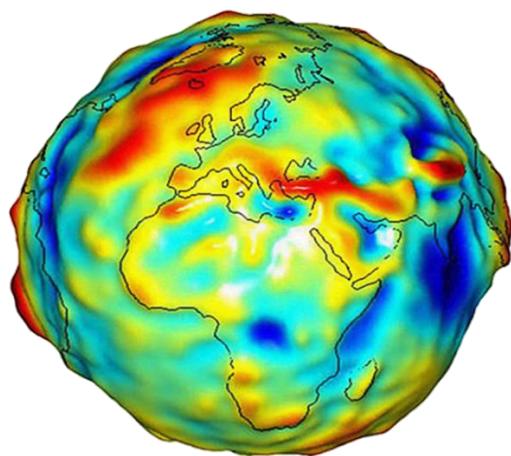
Wstęp.

Rozdział 1

Współrzędne geograficzne i lokalizacja

Unikalną cechą danych przestrzennych jest ich lokalizacja czyli wykorzystanie miejsca w którym znajdują się one na powierzchni Ziemi. Oznacza to, że każda obserwacja będzie musiała posiadać atrybuty opisujące jej położenie.

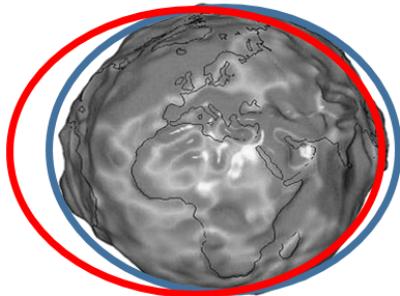
Ziemia ma kształt **geoidy**, bryły o nieregularnej powierzchni przypominającej lekko spłaszoną od dołu i góry kulę. Aby wyobrazić ją sobie należałooby przyjąć, że cała powierzchnia Ziemi została zalana przez ocean. Poziom wody w każdym miejscu tego oceanu ulegałby oczywiście ciągłym wahaniom. Jednak 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 biegunowych są natomiast rezultatem działania siły odśrodkowej będącej rezultatem ruchu wirowego Ziemi na około własnej osi, która to siła jest największa na równiku a najmniejsza na biegunach.



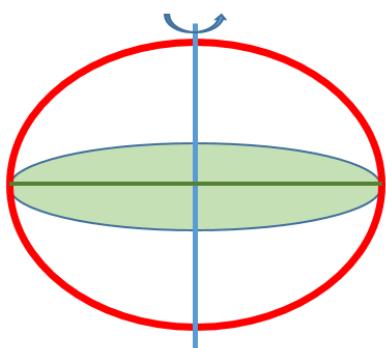
Rysunek 1.1 Kształt Ziemi. Nieregularna kulista bryła zwana geoidą.

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.



Modelem geoidy mogą być różne elipsoidy, mogą być także one różnie umieszczone względem geoidy.



Elipsoida jest bryłą, która powstaje w rezultacie obrotu elipsy wokół jej osi. Oś niebieska będzie modelem osi Ziemi (przechodzącej przez bieguny), oś zielona tworzy w wyniku obrotu koło (płaszczyznę równika).

Rysunek 1.2 Elipsoida jako model geoidy.

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

Tabela 1.1 Podstawowe elipsoidy – modele Ziemi.

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

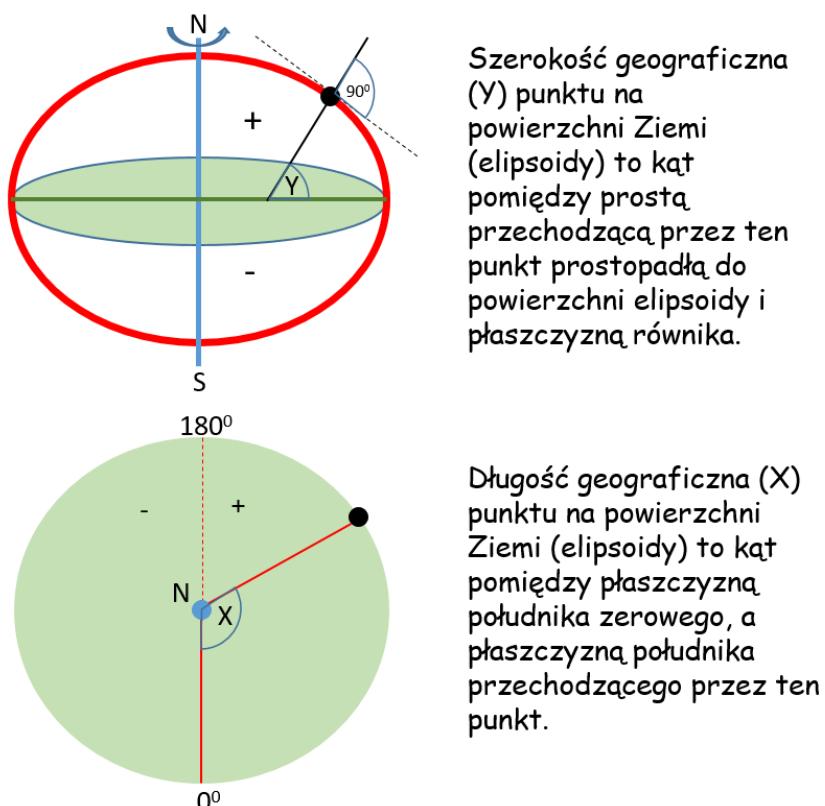
Są one prawie identyczne co sprawia, że często nie będzie istotne, która z nich jest aktualnie wykorzystywana. Oprócz wyboru elipsoidy istotne jest jej umieszczenie względem geoidy. Te dwa elementy rodzą elipsoidy i sposób jej umieszczenia względem geoidy definiują **poziomy układ odniesienia** zwany 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 układem pionowego odniesienia, wysokości są liczone od powierzchni elipsoidy.
ETRS89	GRS 80	Aktualny poziomy układ odniesienia dla Europy. Jest 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 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 zero. 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 spojrzymy na definicję długości geograficznej to najlepiej będzie spojrzeć na nasz model z góry, czyli z 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 przeciwniej 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 metrów.

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.

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 jako email.

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 (bezpłatny program 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, które pozwala na wykorzystywanie bardzo

słabych sygnałów w miejscach, gdzie jeszcze niedawno trudno było wykorzystywać GPS jak wąskie ulice otoczone wysokimi budynkami czy gęsty las.

Rozdział 2

Dane przestrzenne

Niezależnie od tego w jaki sposób wykorzystujemy GIS dane są jego niezbędnym elementem. Dziedziną i związanymi z nią zawodami, która jest postrzegana jako symbol współczesności 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. Dziedzina ta wykorzystuje ogromne ilości danych, które cały czas są tworzone za pomocą najróżnorodniejszych technik. Dane te aby mogły być przetwarzane przez uniwersalne narzędzia i metody muszą być przedstawiane w uniwersalny sposób. Jednym z powszechnie stosowanych rozwiązań jest najprostszy z możliwych sposobów – tekstowe pliki formatu **CSV** (comma-separated values – 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 polegającym na wprowadzeniu 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.

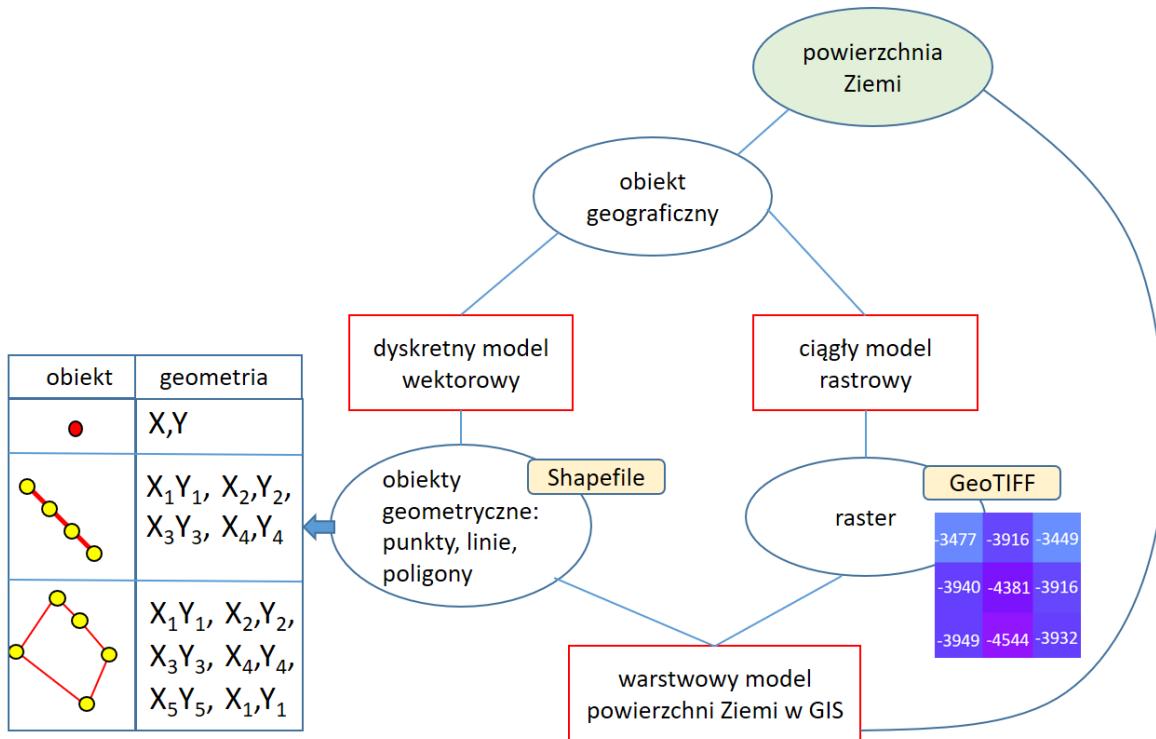
Id	Name	Alias	Type	Type name	Length	Precision
abc 5	kontynent		QString	text	0	0
123 4	litry_alkohol		double	double	0	0
123 3	wino_p		int	integer	0	0
123 2	wodka_p		int	integer	0	0
123 1	piwo_p		int	integer	0	0
abc 0	kraj		QString	text	0	0

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 oddzielna 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 ichłączenie. Nasze obserwacje dotyczą poszczególnych krajów, które mogą być traktowane jako obiekty geograficzne, oczywiście te obiekty w tej tablicy reprezentowane są wyłącznie przez nazwę.

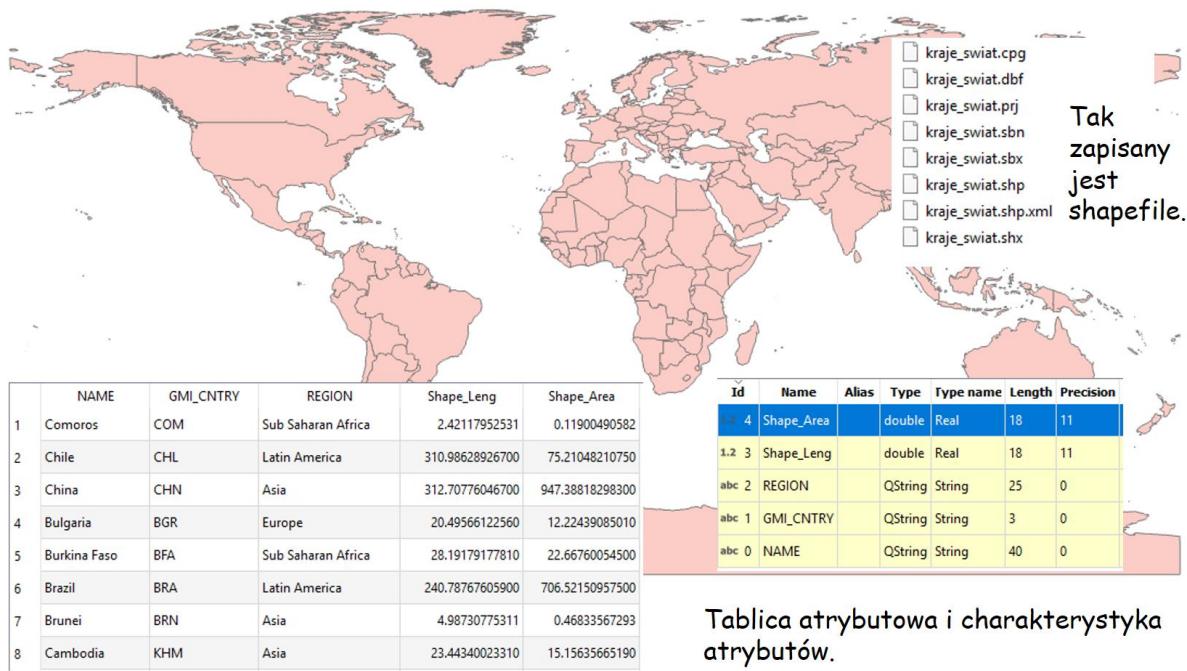
Aby obiekty geograficzne powierzchni Ziemi mogły być przedstawione w postaci danych, potrzebny jest model danych przestrzennych, który pozwoli na ich reprezentację w postaci cyfrowej.



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 różne klasy. Klasę obiektów dyskretnych, które zajmują określone miejsce w przestrzeni i często posiadają wyraźnie określone granice i klasę obiektów ciągłych jak na przykład rozkład opadu deszczu w ciągu roku czy wysokość terenu. Przykładem obiektów dyskretnych będą: studnia, źródło, rzeka, droga, jezioro, obszar powiatu. Ich dobrym reprezentantem będą punkt, linia oraz wielobok (zwany powszechnie w środowisku ludzi zajmujących się GIS poligonom). Geometrie tych obiektów tworzą przestrzenny element wektorowego modelu danych. **Dane wektorowe są pierwszym rodzajem danych przestrzennych wykorzystywany w GIS.** Geometria punktu na powierzchni Ziemi jest zapisywana jako dwie 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.

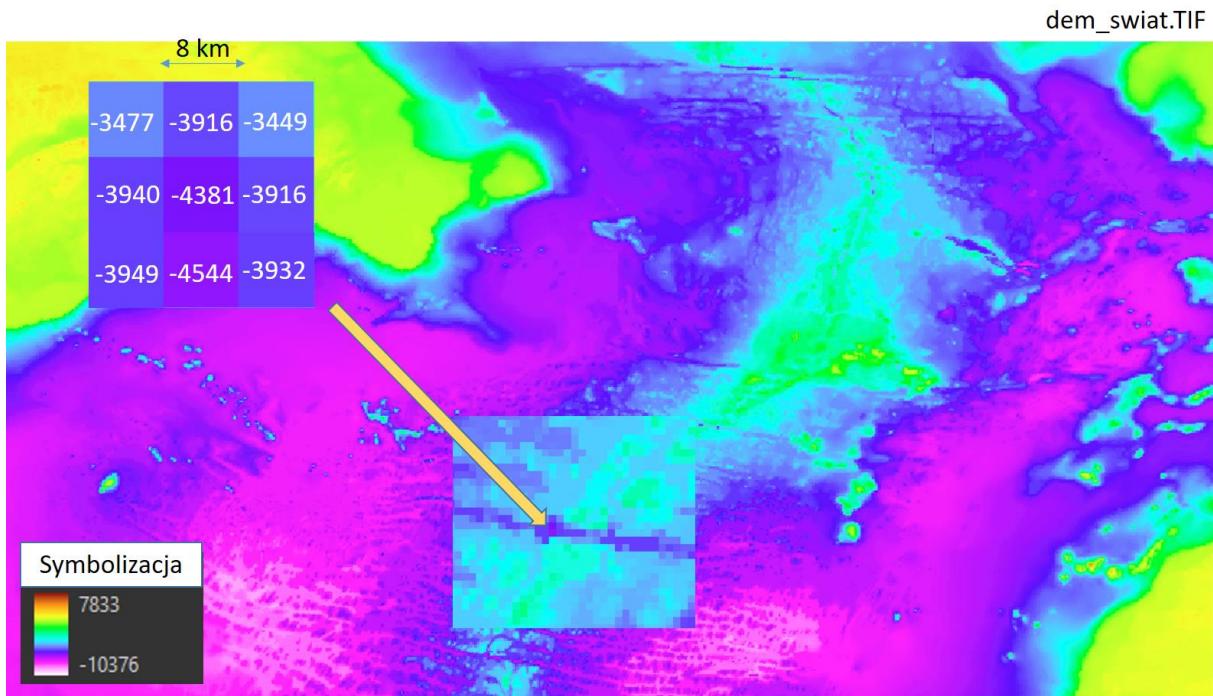


Rysunek 2.4 Wektorowa warstwa poligonów zapisana w formacie Shapefile.

Jest to warstwa poligonów. Warstwy (zbiory danych wektorowych) tworzy się oddziennie dla punktów, linii i poligonów. Pod względem logicznym 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**. Warstwa posiadająca obiekty geometryczne może być wyrysowana na obszarze roboczym danego programu GIS. Sposób rysowania poszczególnych obiektów nosi nazwę **symbolizacji**. Na rysunku 2.4 wszystkie poligony mają takie samo wypełnienie i obrzys (szara linia). Rodzaj symbolizacji może być zależny od wartości w wybranym polu (kolumnie) tablicy atrybutowej.

Obiekty o charakterze ciągłym są reprezentowane za pomocą modelu rasterowego. **Dane rasterowe są drugim podstawowym rodzajem danych przestrzennych używanym w GIS**. Model rasterowy 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, jest to tak zwana **rozdzielcość przestrzenna rastra**.

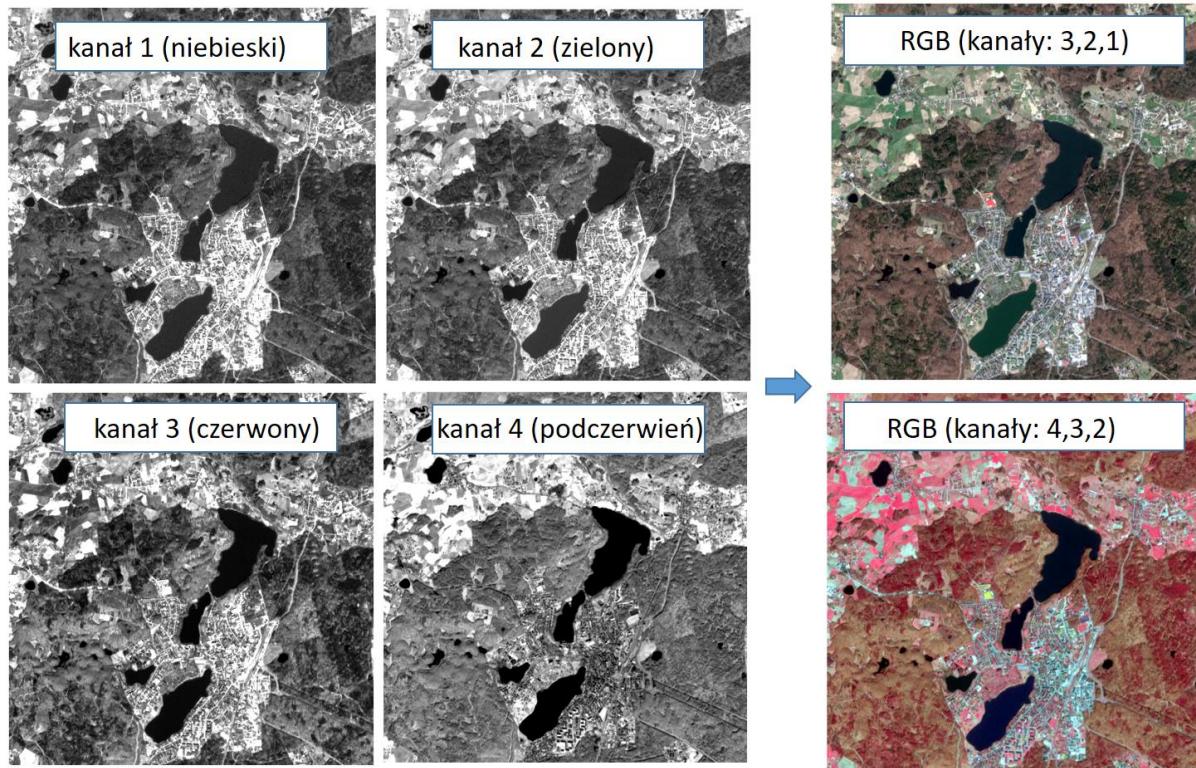


Rysunek 2.5 Rastrowa warstwa głębokości oceanu. Kolejne przybliżenia wycinków rastra pokazują jego strukturę składającą się z kwadratowych komórek.

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, a liczba 1 ląd). Komórki rastrowe mogą także zawierać liczby rzeczywiste. 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 wewnątrz 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 tworzonych 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**.



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

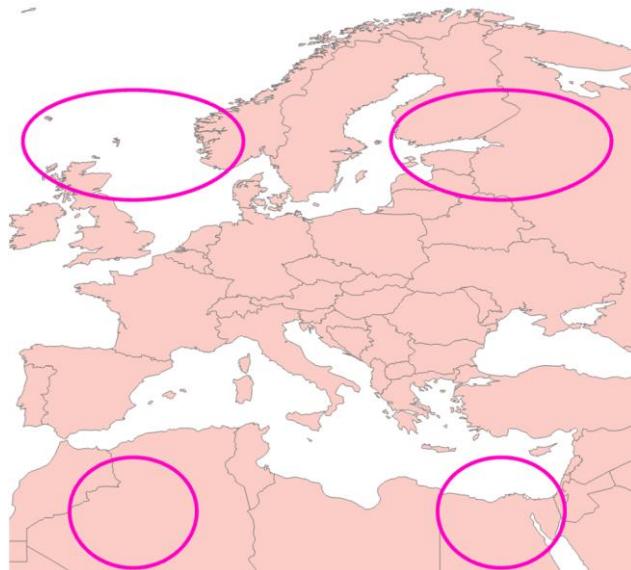
Na rysunku 2.6 z lewej strony pokazane są cztery kanały zdjęcia satelitarnego (są one wyświetlane za pomocą odcieni szarości). Połączono je w kompozyt czterokanałowy. Wizualizacja kompozytu polega na wykorzystaniu modelu kolorów **RGB** (czerwony, zielony, niebieski) do jednoczesnego wyświetlenia trzech 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 informacje niewidoczne w innych zestawach kanałów.

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

Rozdział 3

Współrzędne prostokątne, odwzorowania

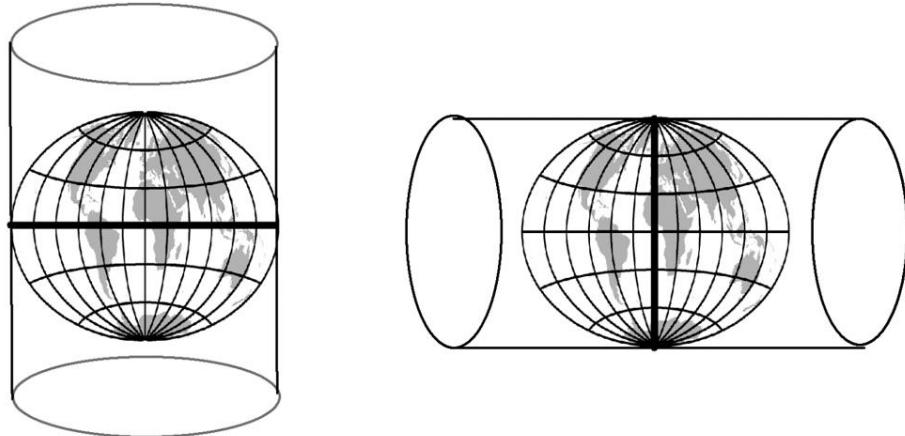
Współrzędne geograficzne służą nam znakomicie do określenia położenia punktów na powierzchni Ziemi (elipsoidy). Problemy pojawiają się, gdy chcemy zmienić ją na płaską powierzchnię ekranu lub mapy.



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

Mapa na Rysunku 3.1 traktuje stopnie szerokości i długości geograficznej jako jednostki długości układu współrzędnych prostokątnych XY aby stworzyć powyższy rysunek. Problem polega na tym, że jeden stopień długości geograficznej (w poziomie) zmienia swoją 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 (rozciągnięciu w poziomie wraz z postępem na północ). Wizualną miarą tego zniekształcenia są obrysów 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 naszej 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. Jedną z idei 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.

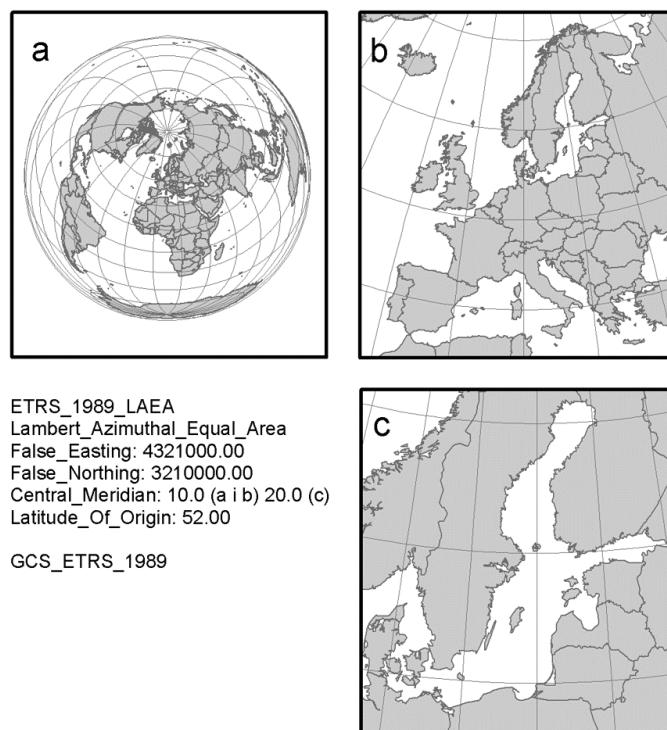
Każdemu odwzorowaniu odpowiada odwrotna funkcja odwzorowawcza, która pozwala na przeliczenie współrzędnych prostokątnych danego odwzorowania z powrotem do współrzędnych geograficznych (dla danego poziomego układu odniesienia). Aby zastosować dane odwzorowanie do konkretnego obszaru powierzchni Ziemi, 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 masie 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 (najlepiej robić to na bazie 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).

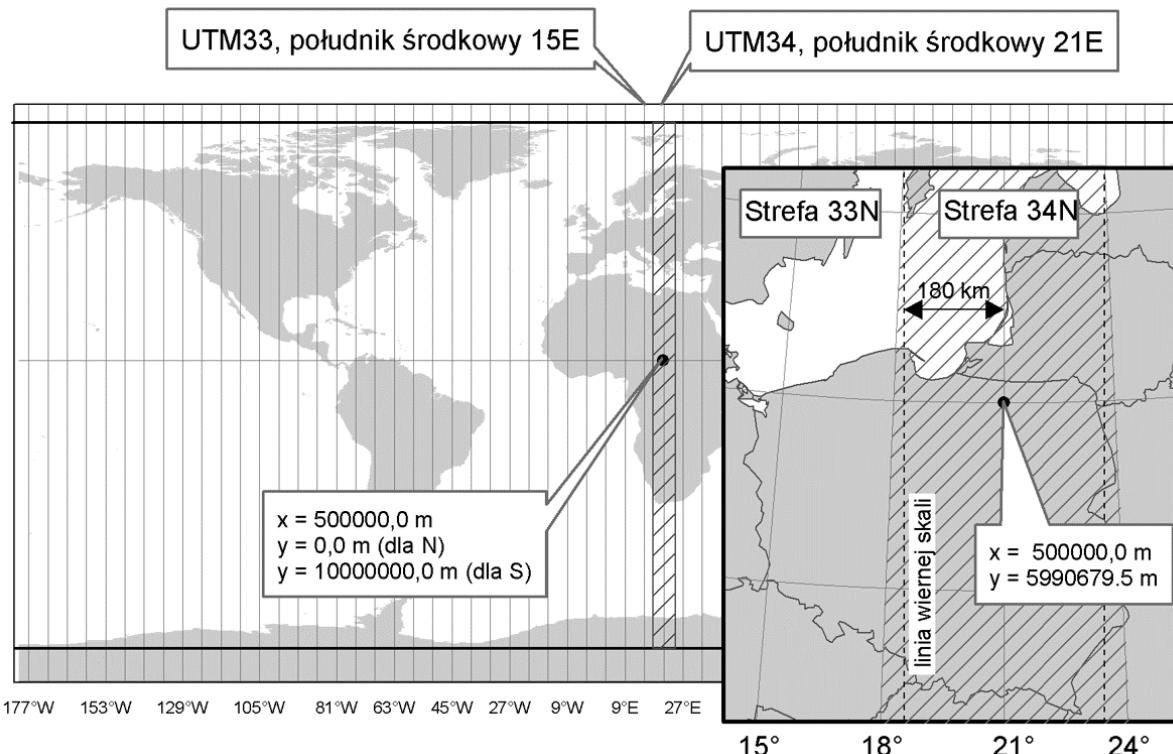
Pełen opis układu współrzędnych LAEA Europe będzie wyglądał w następujący sposób.

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), definiujący 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 1:500 000 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 i wzrastają w kierunku wschodnim). Początek każdego układu współrzędnych systemu UTM znajduje się na przecięciu południka danej strefy z równikiem. W tym miejscu X=500000 m

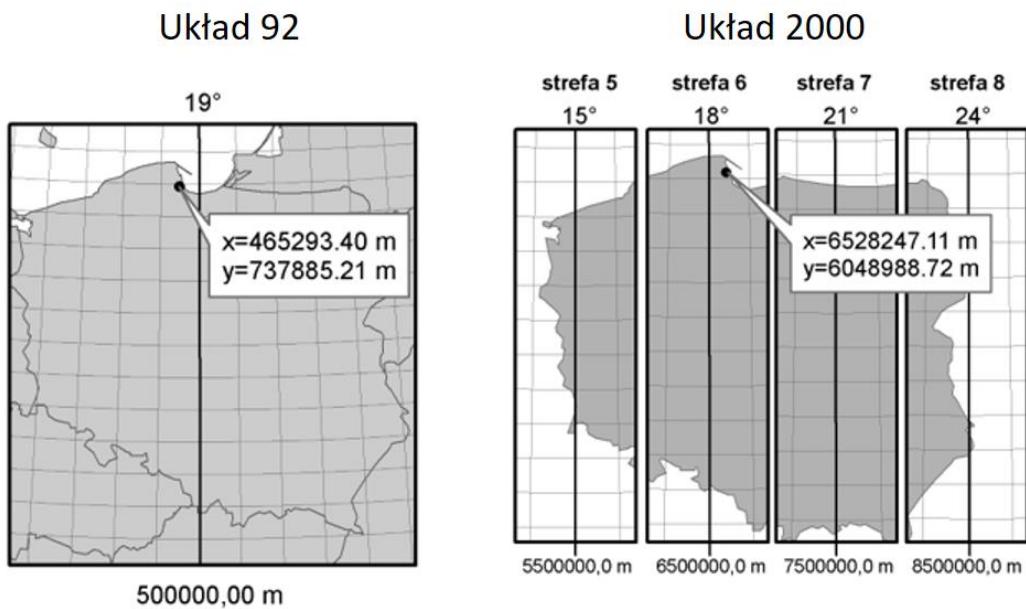
(rośnie na wschód), natomiast Y dla półkuli N = 0 m (rośnie na północ), a dla półkuli południowej N = 10 000 000 m (maleje na południe).



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 średzkowym jest 19° E, a początkowym równoleżnikiem jest równik. Punkt przecięcia równika z południkiem średzkowym (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.

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 cały, złożony często, 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 (papierowych) lub map są rastrami bez dowiązania do powierzchni Ziemi. Aby mogły być wykorzystywane w GIS musi być im nadana georeferencja.

Lipiec 2019

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 jest strona o adresie: <http://www.naturalearthdata.com/>

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*. Wiele instytucji, organizacji lub nieformalnych grup badaczy opublikowało tematyczne zbiory danych o charakterze globalnym, które ze względu na rozdzielcość przestrzenną mogą być wykorzystane do projektów lokalnych. Przykładem może być zbiór danych klimatycznych **WorldClim**:

WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

[Download](#) [Contact](#)

WorldClim

WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km². These data can be used for mapping and spatial modeling.

The current version is [Version 1.4](#).

For this version you can get data for past, current and future climates.

A preview of [Version 2](#) is also available (current climate only)

[Read more](#)

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.

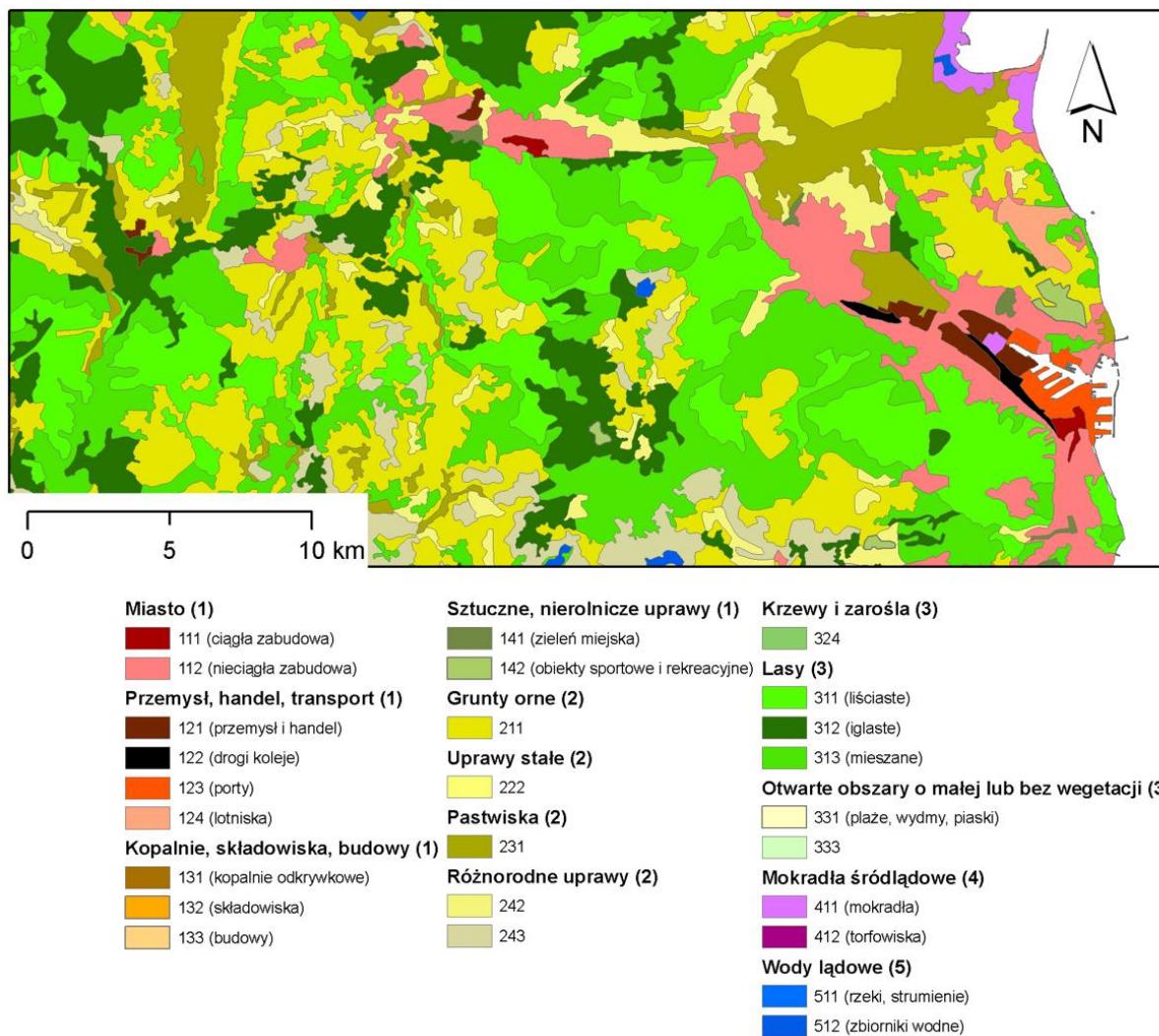
Rastrowe dane wysokościowe dla całego świata

Zupełnie inną genezę mają dane **OpenStreetMap**, są to dane

Szereg danych w formacie SHP można pobrać z : <http://download.geofabrik.de/>

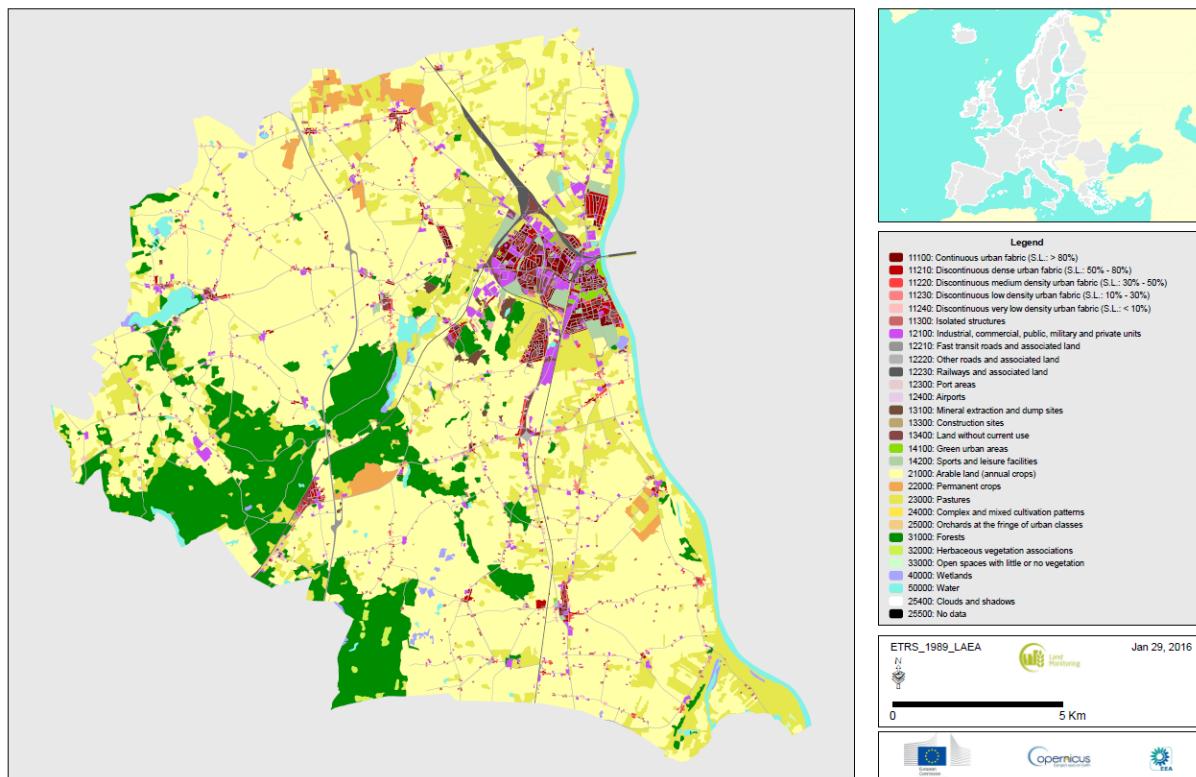
The screenshot shows the GEOFABRIK 'downloads' page for the Pomeranian Voivodeship. At the top, there's a link to 'Download OpenStreetMap data for this region'. Below it, a section titled 'Województwo pomorskie (Pomeranian Voivodeship)' provides a summary of the data: 'The OpenStreetMap data files provided on this server do **not** contain the user names, user IDs and changeset IDs of the OSM objects. These metadata fields contain personal information about the OpenStreetMap contributors and are subject to data protection regulations in the European Union. Please note that these regulations apply even to processing that happens outside the European Union because some OpenStreetMap contributors live in the European Union.' It also mentions that 'Extracts with full metadata are available to OpenStreetMap contributors only.' On the right, there's a map of the Pomeranian Voivodeship with several regions highlighted in orange, labeled with cities like Gdańsk, Szczecin, and Gdynia. Below the map, there's a section for 'Commonly Used Formats' with links to download .osm.pbf, .shp.zip, and other files. Another section for 'Other Formats and Auxiliary Files' lists more file types like .xml.gz and .index. A note at the bottom right explains that Geofabrik is a consulting and software development firm based in Karlsruhe, Germany, providing OpenStreetMap services.

Chyba 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 1:100 000 (ciągłe homogeniczne obszary o powierzchni ponad 25 ha i wymiarach ponad 100 m). Pokrycie terenu sklasyfikowane jest za pomocą trzy-poziomowej klasyfikacji hierarchicznej.



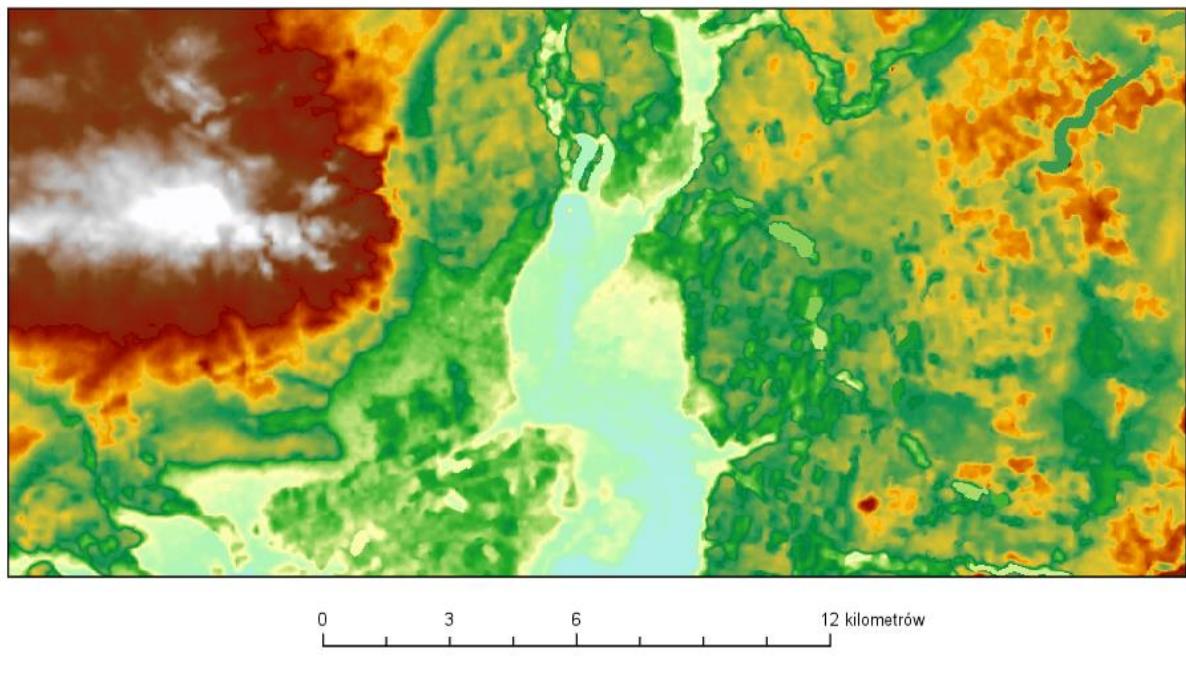
Rysunek 4.1 Przykład mapy z danych Corine 2000 wraz z hierarchicznymi klasami występującymi na mapie. W nawiasach identyfikatory klas pierwszego poziomu (1. Powierzchnie sztuczne; 2. Obszary rolnicze; 3. Lasy i obszary częściowo naturalne; 4. Mokradła i bagna; 5. Zbiorniki wodne). Klasy drugiego poziomu są opisane drukiem pogrubionym. Ich identyfikatory to dwie pierwsze cyfry klas trzeciego poziomu z nazwami w nawiasach.

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.



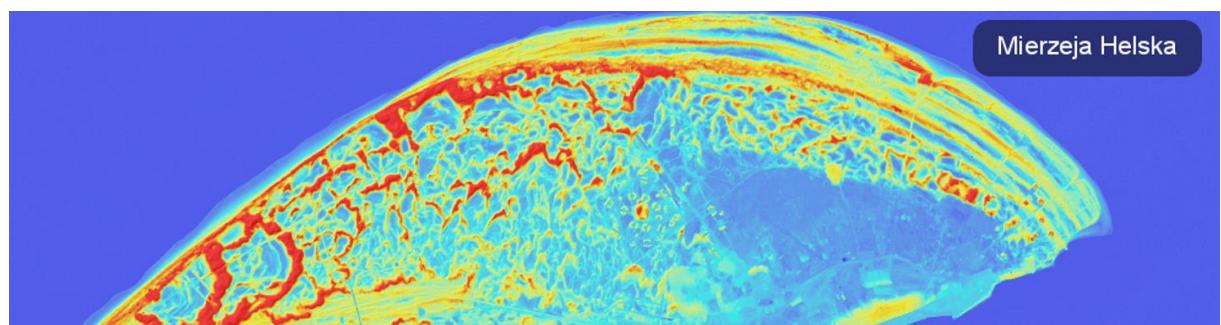
Rysunek 4.2 Przykład mapy z danych URBAN ATLAS. Użyta klasyfikacja dzieli zabudowę na pięć klas w zależności od gęstości zabudowy (cztery klasy) plus klasa reprezentująca pojedyncze zabudowania. Drogi przedstawione są za pomocą dwóch klas. Natomiast warstwa obejmuje tylko większe rzeki.

Drugą 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 (**DSM – Digital Surface Model**).



Rysunek 4.3 Przykład mapy z danych EU-DEM dla okolic Piły. Różnice wysokości na tej mapie wynoszą około 100 m.

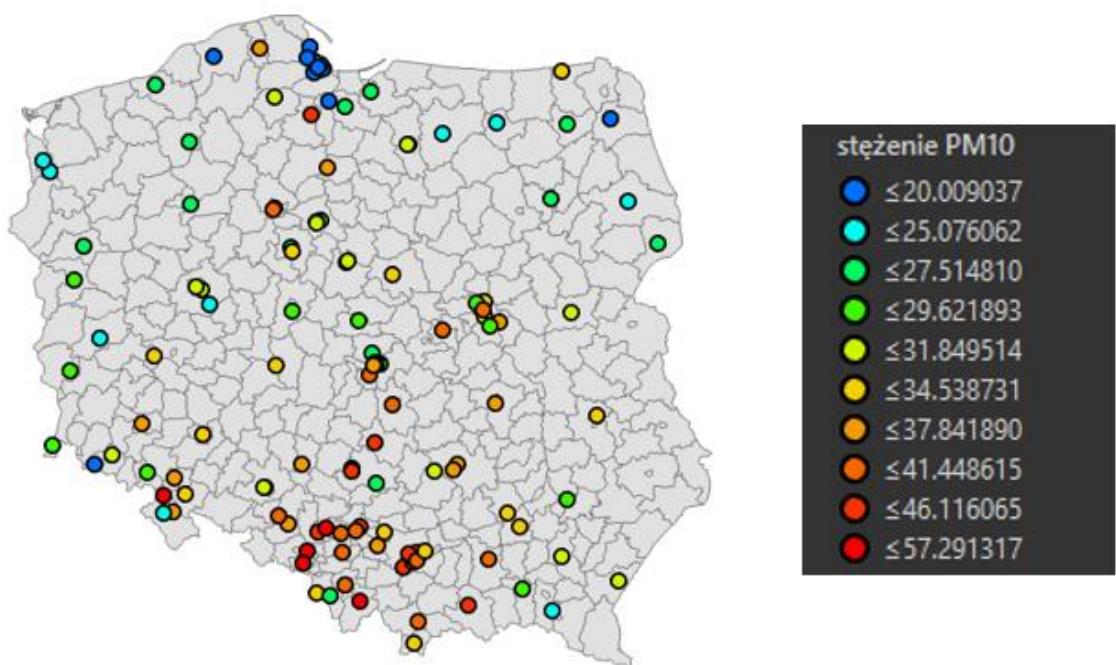
Dostępność przestrzennych danych w Polsce



4.2 Eksploracyjna analiza danych

Kiedy zaczynamy pracę z nową warstwą danych musimy się jej dokładnie przyjrzeć zanim użyjemy ją 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.



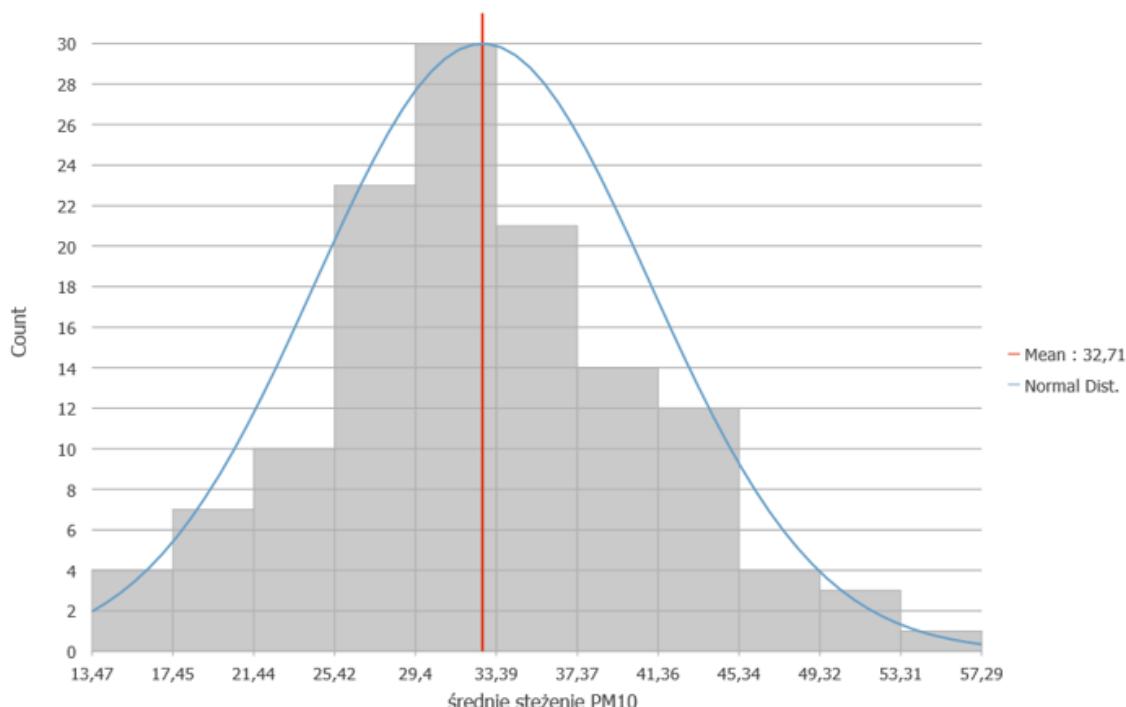
Rysunek 4.6 Symbolizacja punktowych danych wektorowych na podstawie ciągłego atrybutu (stężenie PM10)

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 być 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 wartości 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 nie są rozmieszczone 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. Jedną z podstawowych metod jest histogram i podstawowe statystyki.

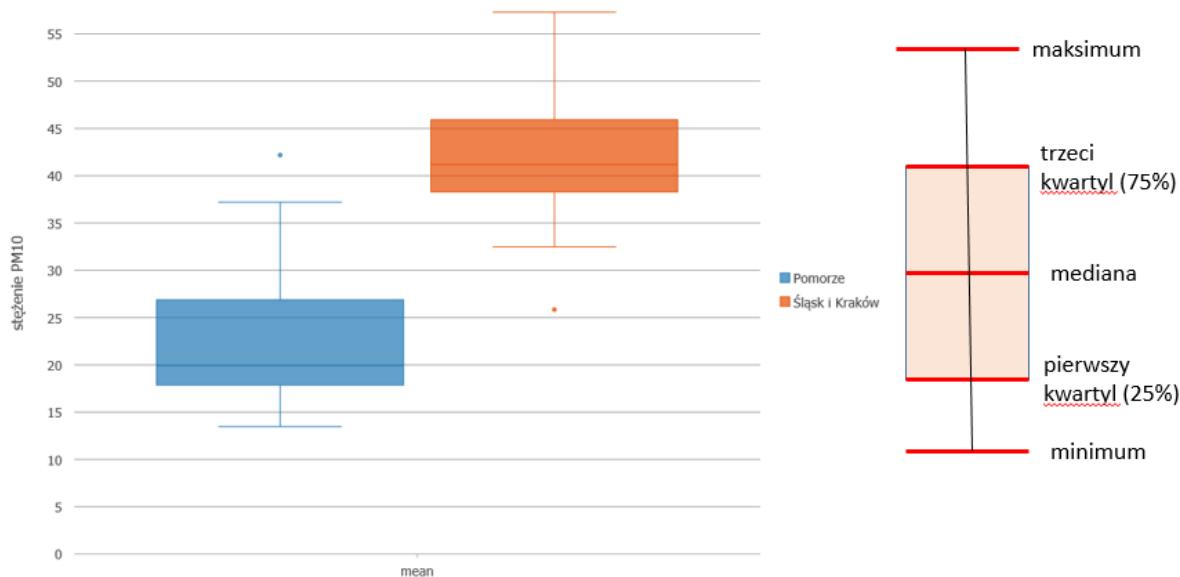


Rysunek 4.7 Histogram ciągłego atrybutu (stężenie PM10) z zaznaczoną średnią i wykresem rozkładu normalnego.

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 4.7 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*.



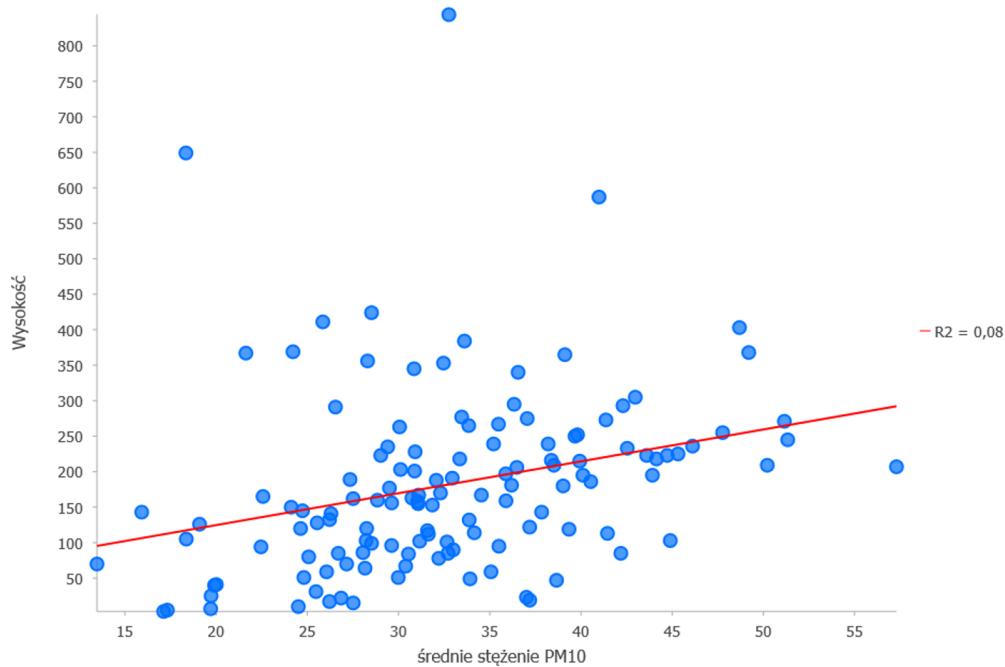
Rysunek 4.8 Wykres typu box-plot pokazuje różnice w stężeniu PM10 pomiędzy stacjami na Pomorzu a Śląskiem wraz z Krakowem

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 4.8 dane analizie poddano dwa podzbiorы 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).

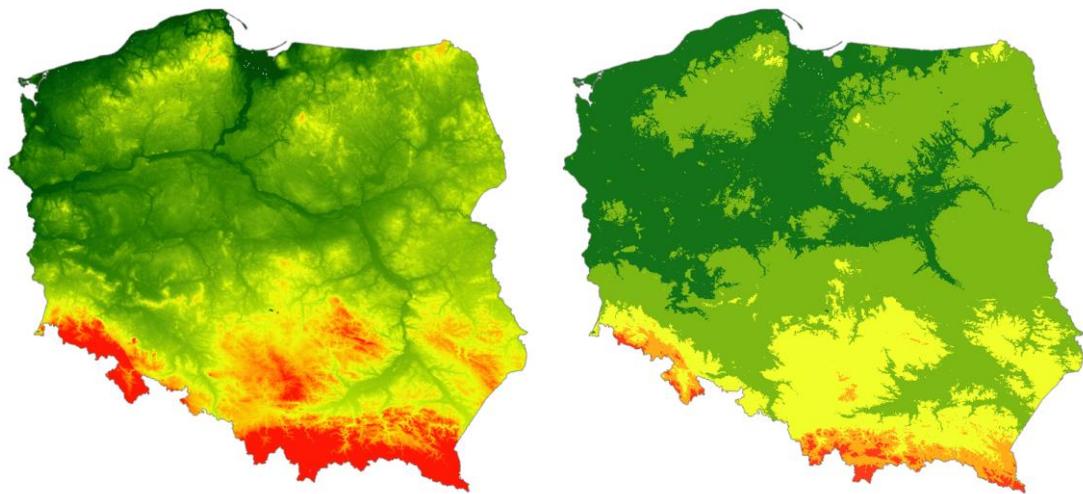


Rysunek 4.9 Wykres rozrzutu (scatter-plot) przedstawia relację między zmiennymi. Reprezentacją tej relacji jest często linia regresji liniowej oraz współczynnik R^2 opisujący zgodność punktów z tą linią.

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 4.9 jest,

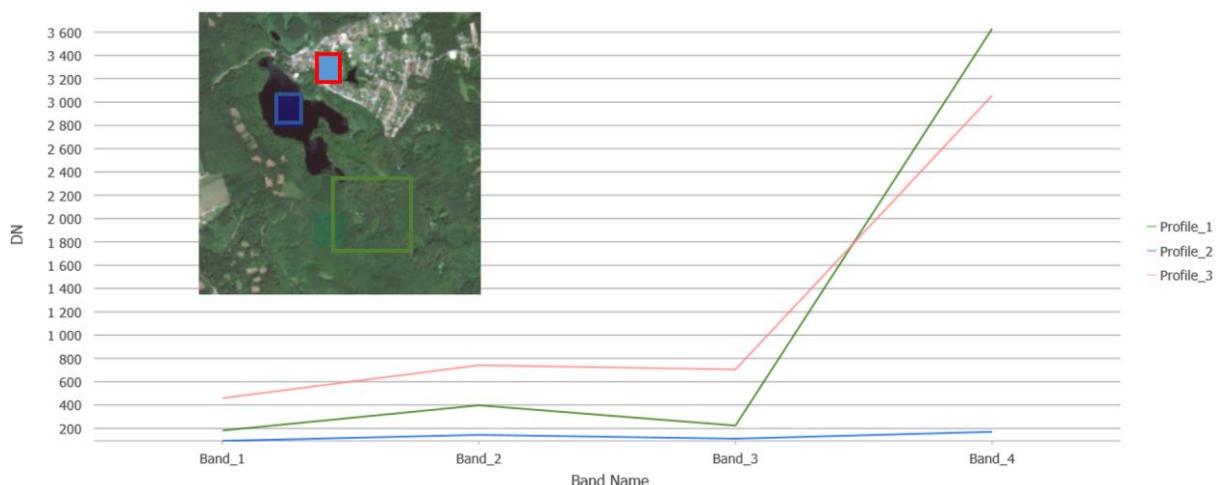
- Statystyka $R^2=0.08$ wykazuje brak relacji pomiędzy wartością stężenia i wysokością stacji pomiarowej

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 jego wartości od wartości minimalnej do maksymalnej (Stretch), druga przypisuje kolory do zdefiniowanych zakresów wartości (Classify).



Rysunek 4.10 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.



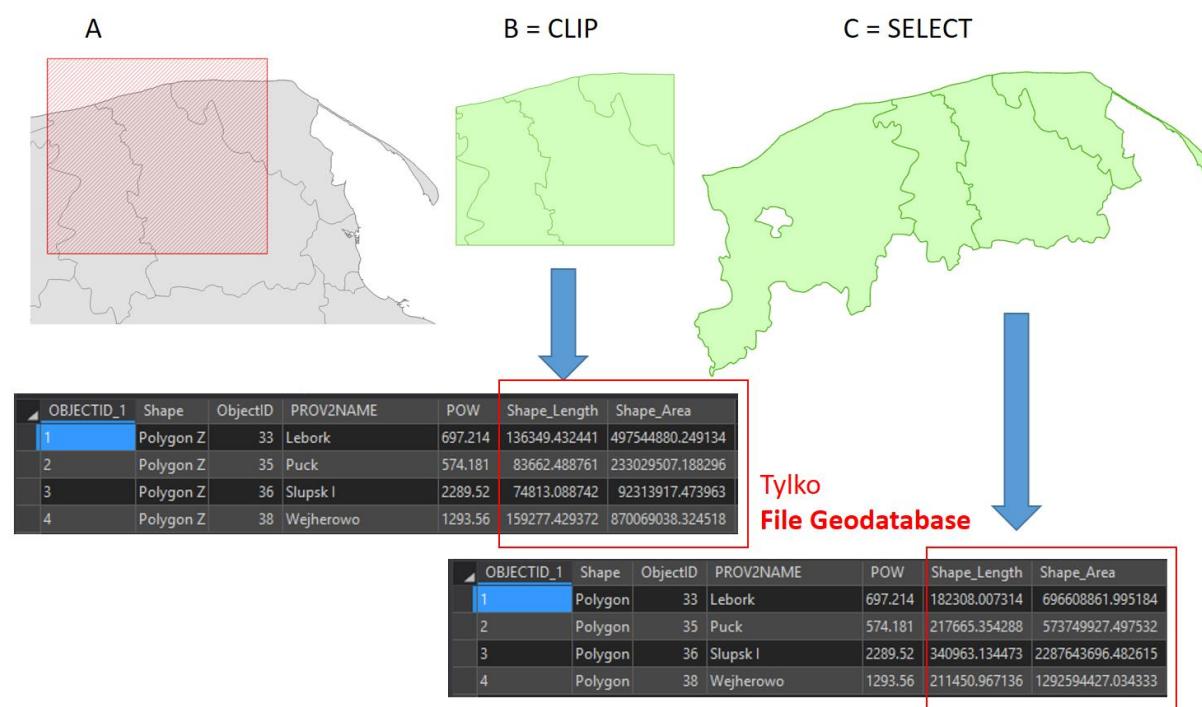
Rysunek 4.11 Profile spektralne (średnie) dla trzech typów pokrycia terenu (las, woda i zabudowania).

4.3 Podstawowe operacje przygotowania danych

Proces przygotowania danych zwykle zajmuje najwięcej czasu (nawet do 90%). Po zgromadzeniu wszystkich danych dla danego projektu następuje ich przetworzenie do postaci najbardziej dogodnej.

4.3.1 Tworzenie podzbiorów danych (CLIP, SELECT)

Zwykle 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 obszaru 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 tworzenia 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**.

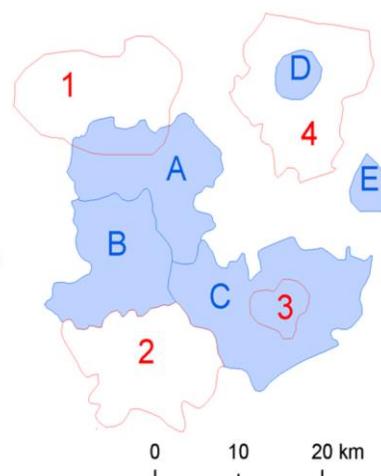
Istnieją dwie metody przeprowadzania selekcji, pierwsza na podstawie atrybutów danej warstwy i druga na podstawie relacji przestrzennej pomiędzy dwoma warstwami (może to też być relacja warstwy z samą sobą). Podstawą przeprowadzania selekcji atrybutowej jest tworzenie zapytań za pomocą języka SQL. SQL jest specjalnym językiem formowania zapytań do tablic baz danych.

Jego użycie w GIS jest uproszczone, może także być wykonywane w przyjaznym interfejsie. Podstawowe wyrażenia SQL będą miały postać 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 wybrane już rekordy). 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 obiekty C, D mają swój środek wewnątrz obiektów CYFR
share a line segment with	obiekty B, C mają wspólny odcinek (linii obrysu) 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
are within	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.

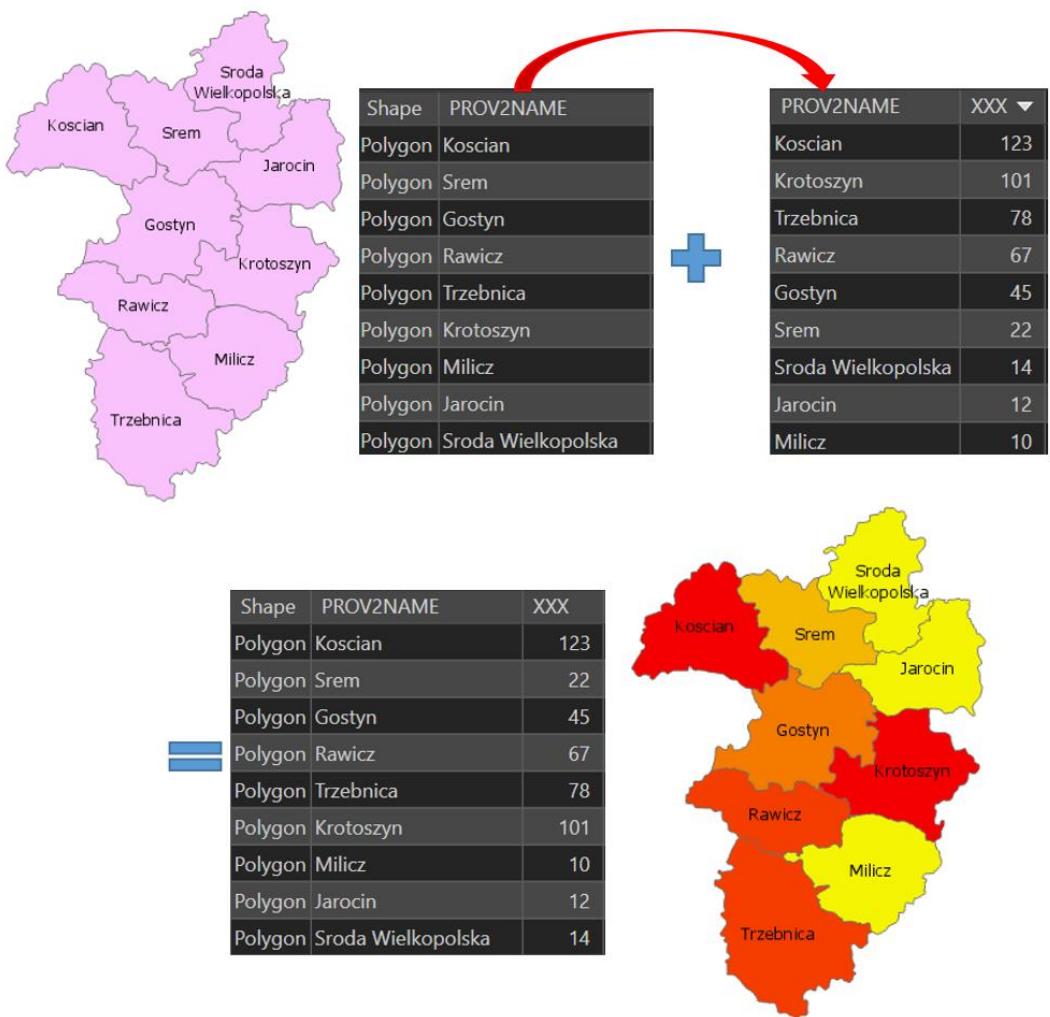
Po utworzeniu podzbioru rekordów należy uporządkować pola (kolumny) tablicy atrybutowej. Może to polegać na usunięciu niepotrzebnych pól lub utworzeniu nowych. Może także zaistnieć potrzeba zmiany nazwy lub typu danych pola. Najprościej jest utworzyć wtedy nowe pole odpowiedniego typu (z nową nazwą), 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 częste 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 dwoma 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 dzięki zgodności nazw w obu polach (PROV2NAME).

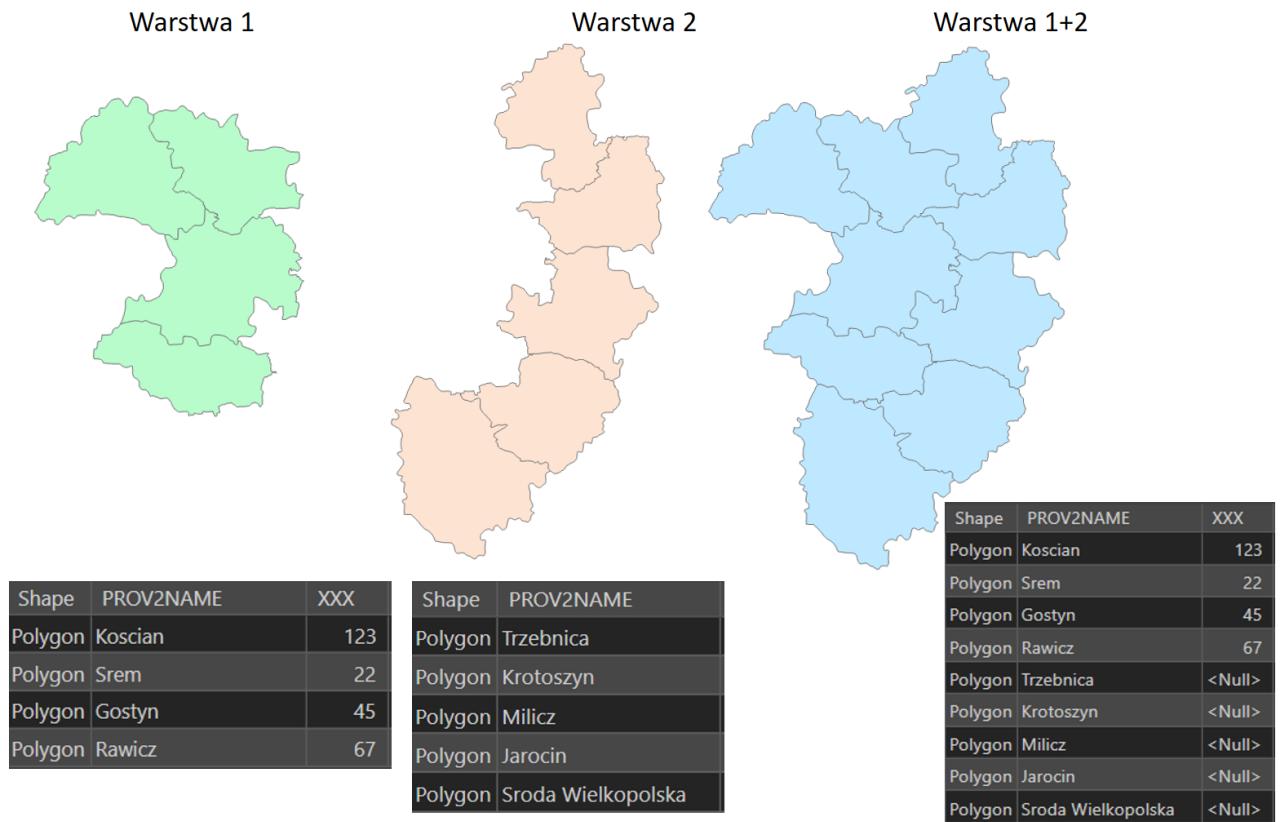
JOIN daje możliwość transferu danych w obie strony pomiędzy danymi przestrzennymi i wyłącznie atrybutowymi. Przykładowo możemy atrybutowe dane zawierające nazwy miast i liczbę ludności wzbogacić o atrybut odległości tych miast od najbliższego miasta wojewódzkiego wykorzystując warstwę przestrzenną miast 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).

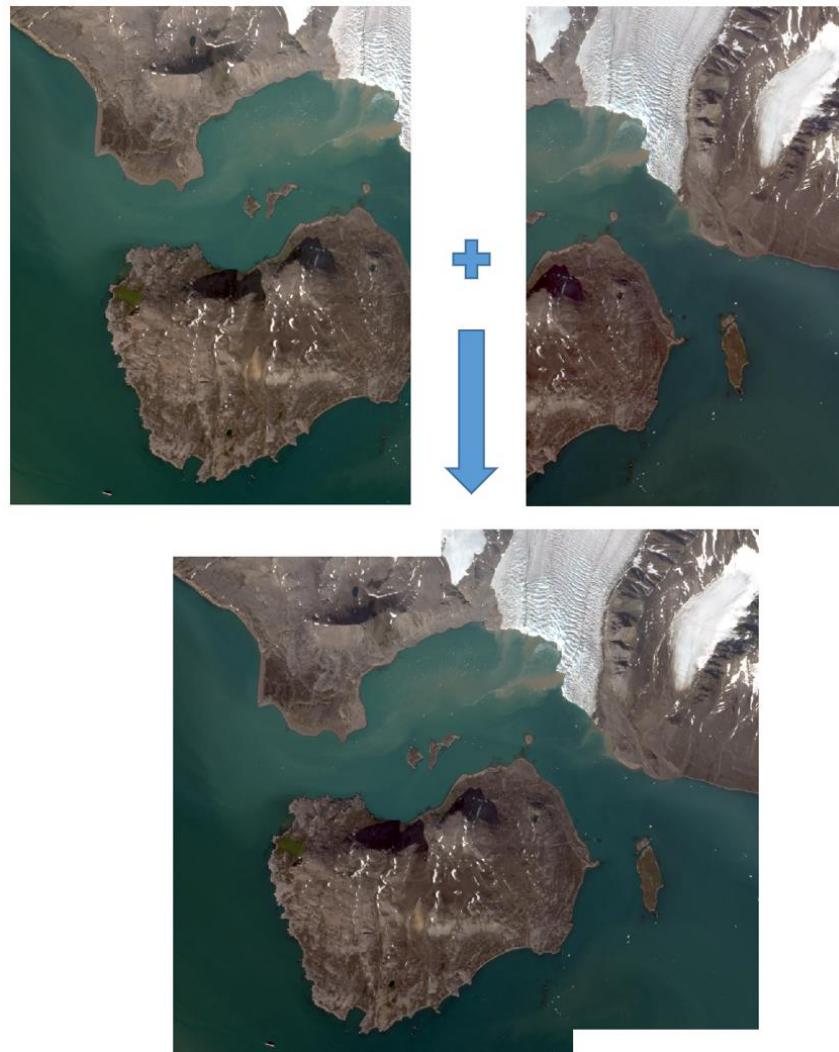
Drugi scenariusz polega na łączeniu kilku warstw w jedną. Do tego celu w GIS służy narzędzie **Merge**. Łączy ono warstwy wektorowe tego samego typu (punkty, linie lub poligony) w jedną warstwę. Nowa warstwa będzie zawierać wszystkie obiekty łączonych warstw. Atrybuty zostają przypisane do istniejących 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 pochodzącymi 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.



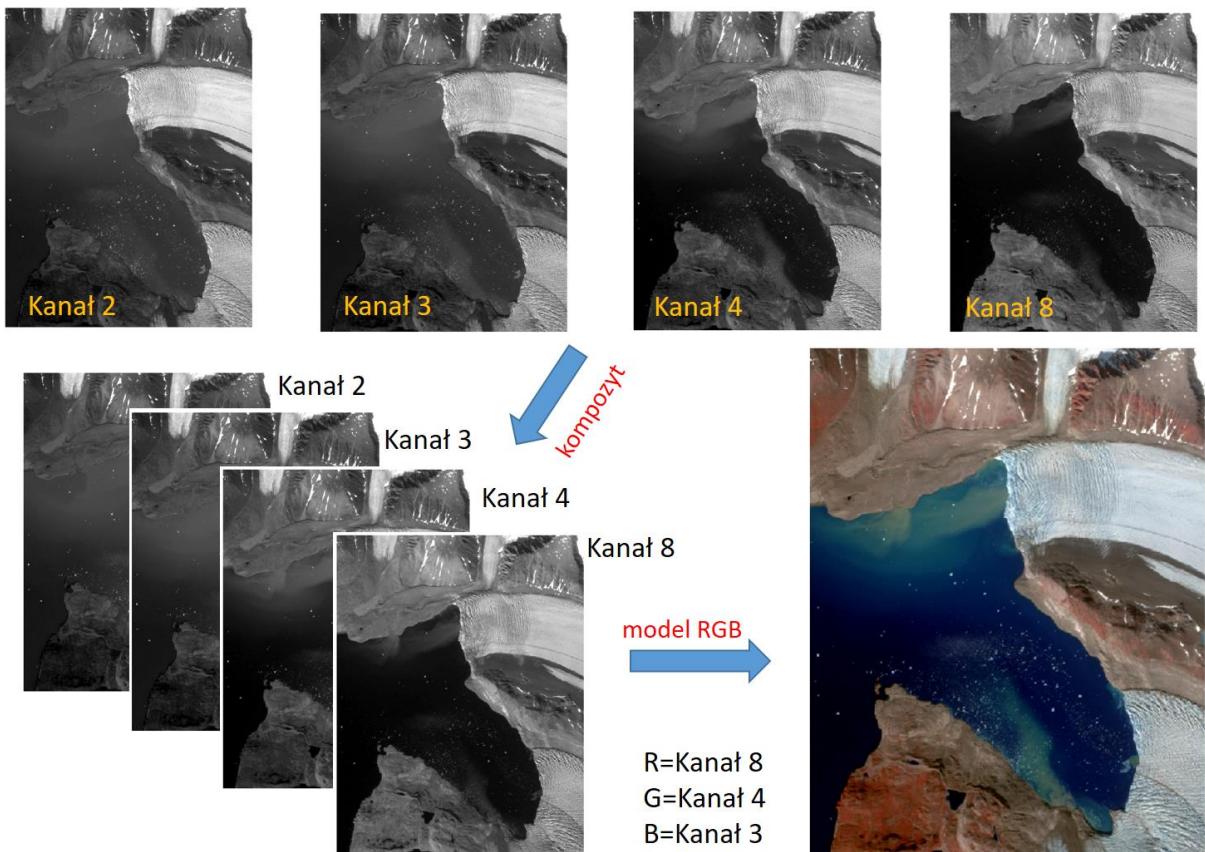
Rysunek 4.16 Łączenie kilku warstw za pomocą Merge.

Przygotowywanie danych rastrowych ze zdjęć satelitarnych wiąże się zwykle z użyciem dwóch narzędzi. Pierwszym z nich jest **Mosaic** pozwalający na łączenie ze sobą paru scen w jedną. Często mamy do czynienia z sytuacją, że nasz obszar zainteresowania obejmuje dwie pobraane 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 rastrów korzystając z ich referencji przestrzennej. Narzędzie to może być stosowane zarówno do rastrów jedno jak i wielokanałowych. Definiuje się także z którego rastrowa zostaną pobraane 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 połączyć te kanały 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-kanalowych ze zdjęcia Sentinel 2 w kompozyt cztero-kanalowy 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, dzięki czemu łatwo go opanować i używać.

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 do wykonania przez komputer. Python realizuje instrukcję po instrukcji, zamieniając je kolejno do postaci zrozumianej przez komputer (na tzw.

język maszynowy) i wykonując. Do pisania programów używa się specjalnego programu pełniącego rolę środowiska programowania zwanego IDE (od Interactive Development Environment). GIS zawiera szereg gotowych miejsc gdzie będziemy mogli programy w Pythonie tworzyć i wykonywać.

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

```
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.

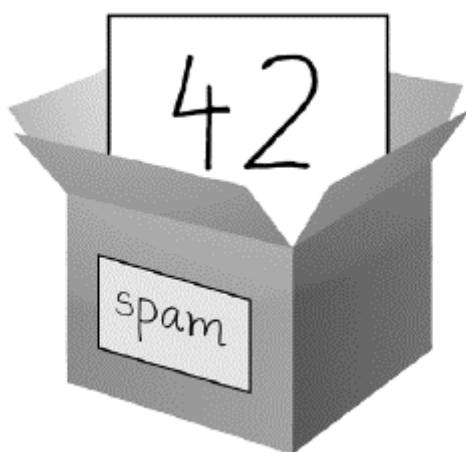
Każda wartość należy do określonego **typu danych**. Podstawowe typy danych w Pythonie to:

Typ danych	Przykład
Całkowite (integer)	-12, 0, 3, 234, 17
Rzeczywiste (float)	-1.30, 0.0, 12.01, 13.8
Tekstowe (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. Przyjęto, że 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:

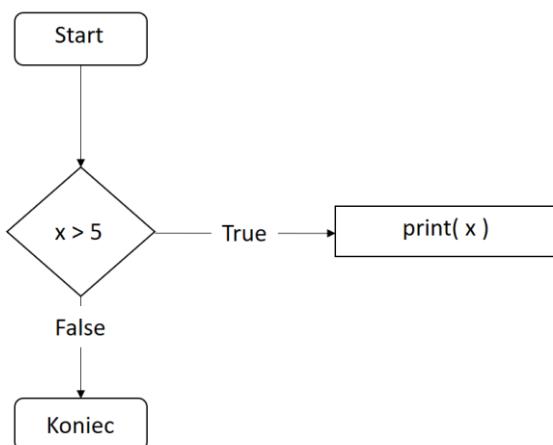
Operator	Znaczenie / wynik
==	równa się
!=	różne
<	mniejsze
>	większe
<=	mniejsze lub równe
>=	większe lub równe
True and True	True
True and False	False
False and True	False
False and False	False
True or True	True
True or False	True
False or True	True
False or False	False

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:

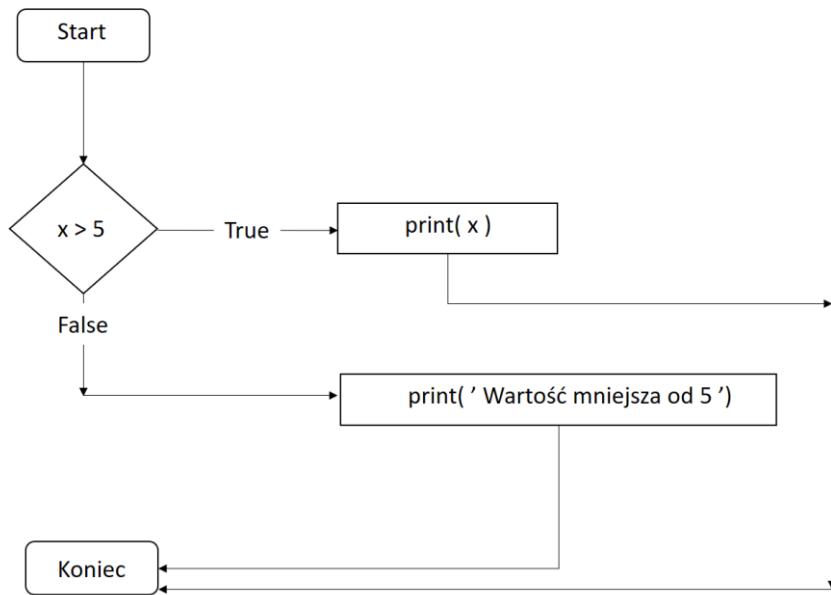


Rysunek 4.19 Realizacja bloku instrukcji IF.

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.

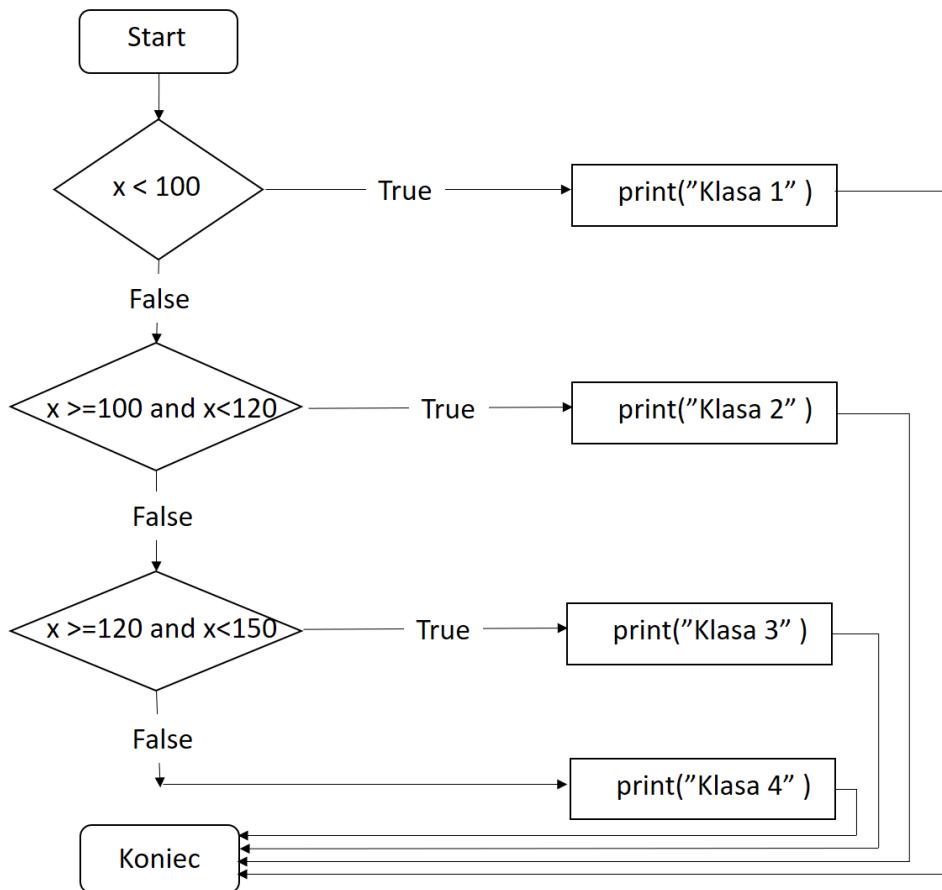
Kolejnym rozwinięciem tej instrukcji jest użycie **elif**.

```

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 zainportowania 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:

```

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, oraz przykładowe 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 wewnętrz 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,

```

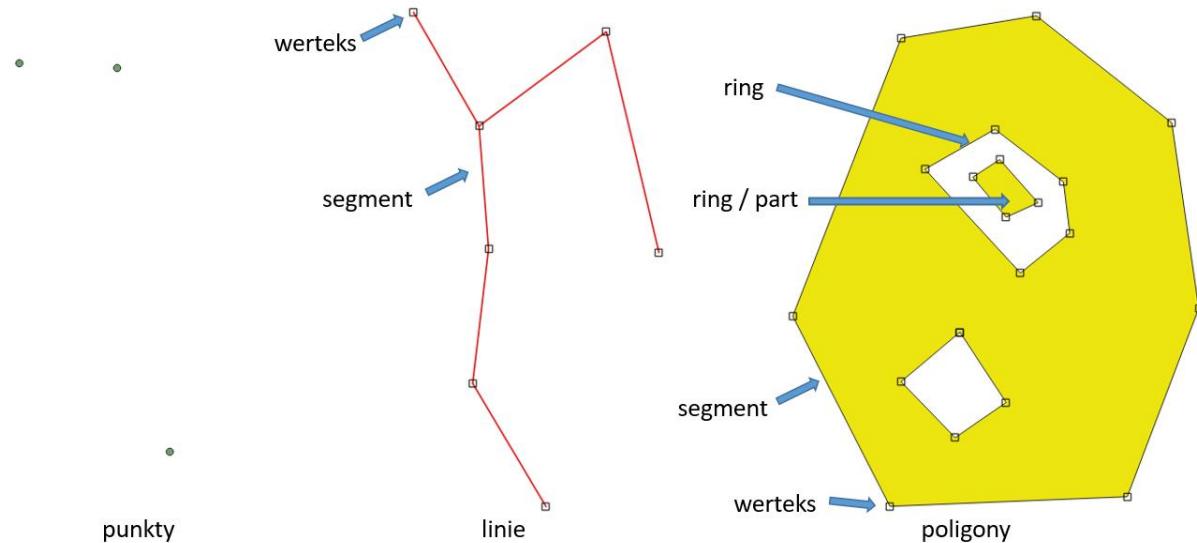
1
-1
0

```

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.

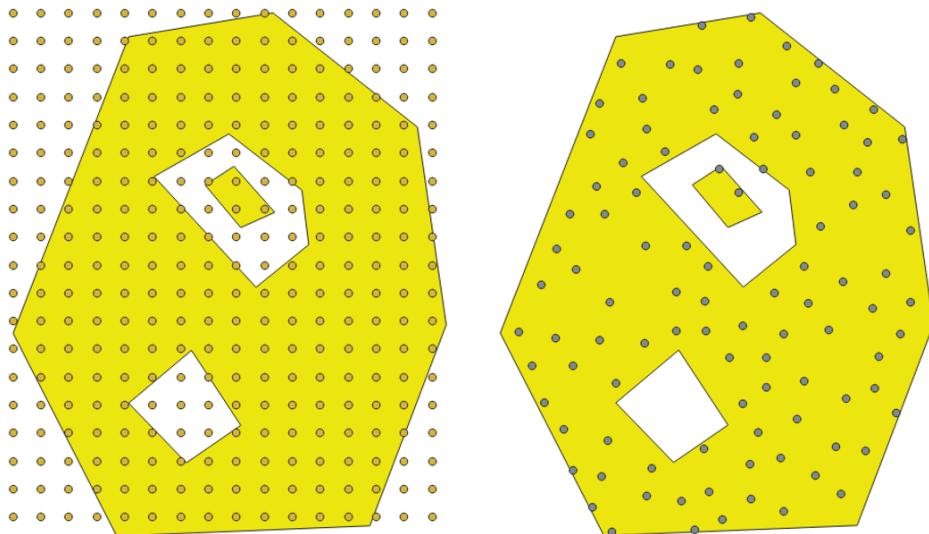
Geometria wszystkich trzech typów obiektów oparta jest na punktach o określonych współrzędnych. Najczęściej będziemy mieli do czynienia z danymi dwuwymiarowymi, w których każdy punkt (werteks) tej geometrii będzie miał dwie współrzędne **X** i **Y**. Coraz częściej wykorzystuje się jednak dane trójwymiarowe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



Rysunek 4.22 Podstawowe elementy konstrukcyjne danych wektorowych.

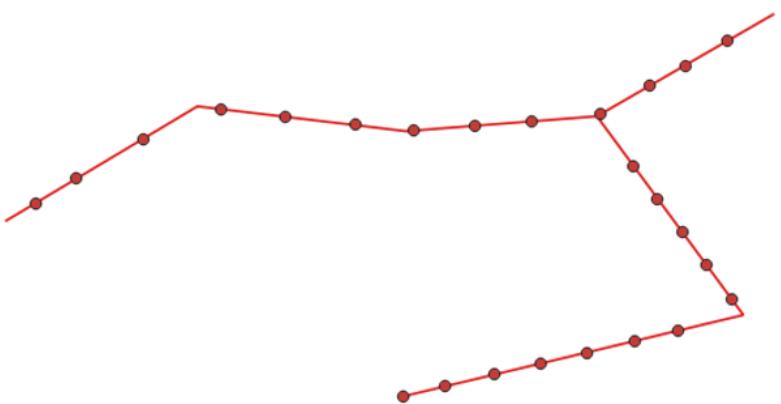
Tworzenie i modyfikowanie danych wektorowych opiera się o ich elementy konstrukcyjne. Zwykle wykonujemy te czynności w tzw. **trybie edycyjnym**, co umożliwia po ich zakończeniu albo zapisanie dokonanych zmian, albo powrót do stanu przed edycją. Podstawowymi elementami konstrukcyjnymi są **werteksy**, **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 obręb 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).



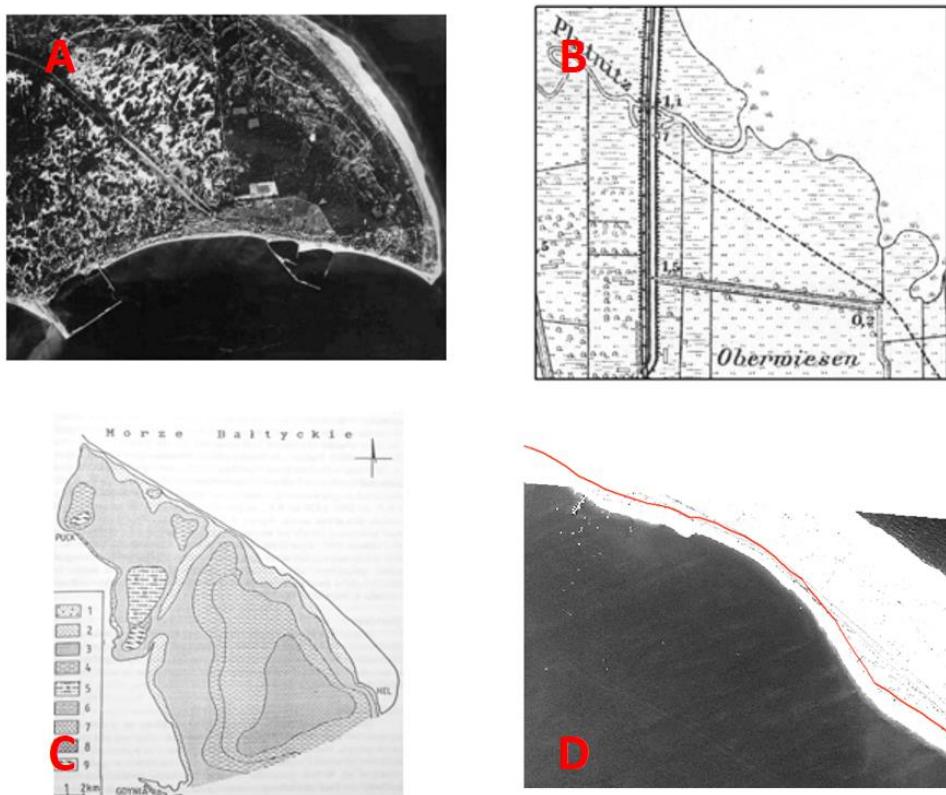
Rysunek 4.24 Punkty rozmieszczone losowo wzdłuż linii.

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.

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**) informacja znajduje się 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 4.25 Przykłady rastrów nie posiadających informacji o lokalizacji na powierzchni Ziemi.

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 4.25D 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,

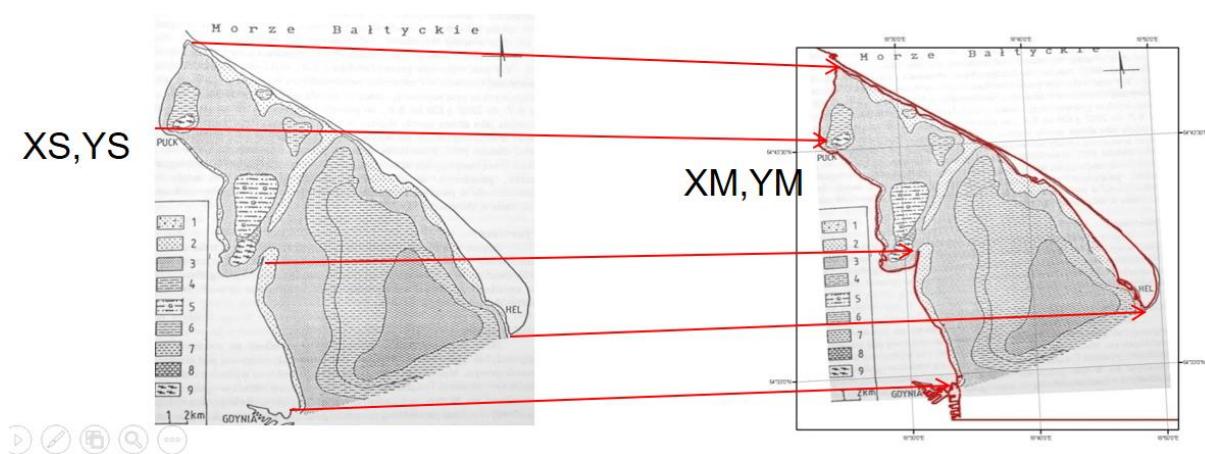
$$x_m = x_s + A_1 x_s + A_2 y_s + A_3 x_s y_s + A_4 x_s^2 + A_5 y_s^2 + A_6 x_s^2 y_s + A_7 x_s y_s^2 + A_8 x_s^3 + \dots$$

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

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>

Powyższy układ równań może być rozwiązywany (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 4.26 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 punktów. 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 otrzymanymi z wyrażenia transformacji geometrycznej (x_{st} , y_{st}) nosi nazwę rezydium,

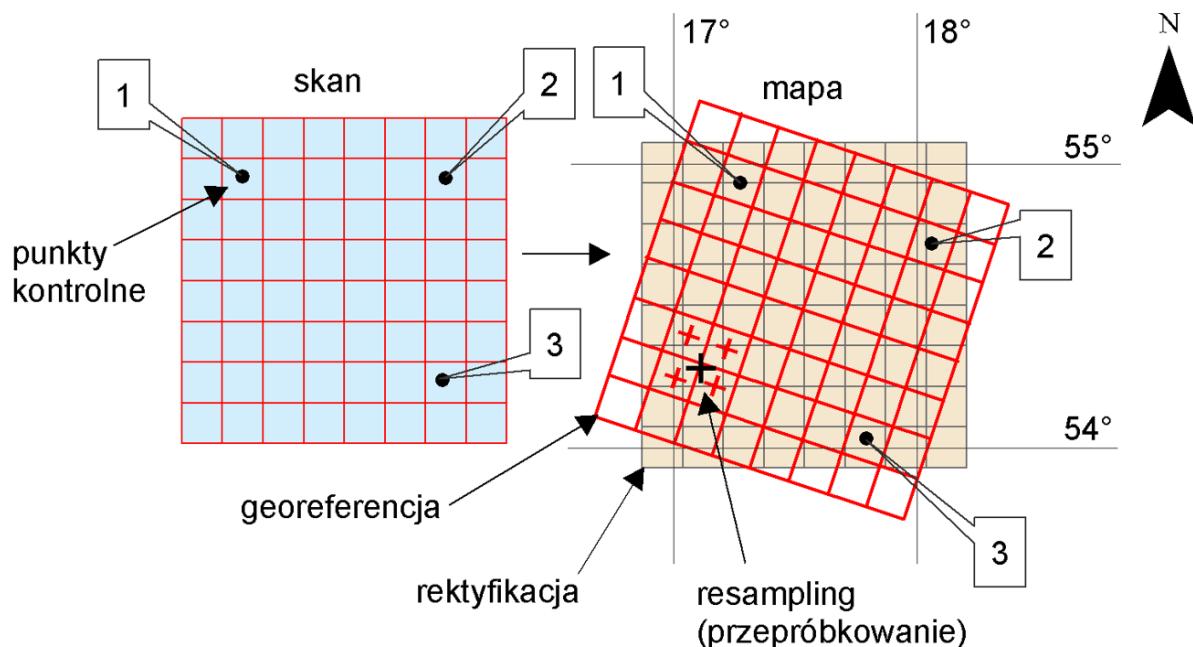
$$D = \left[(x_{st} - x_m)^2 + (y_{st} - y_m)^2 \right]^{1/2}$$

a dokładność całego procesu transformacji geometrycznej skanu opisuje błąd RMS,

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

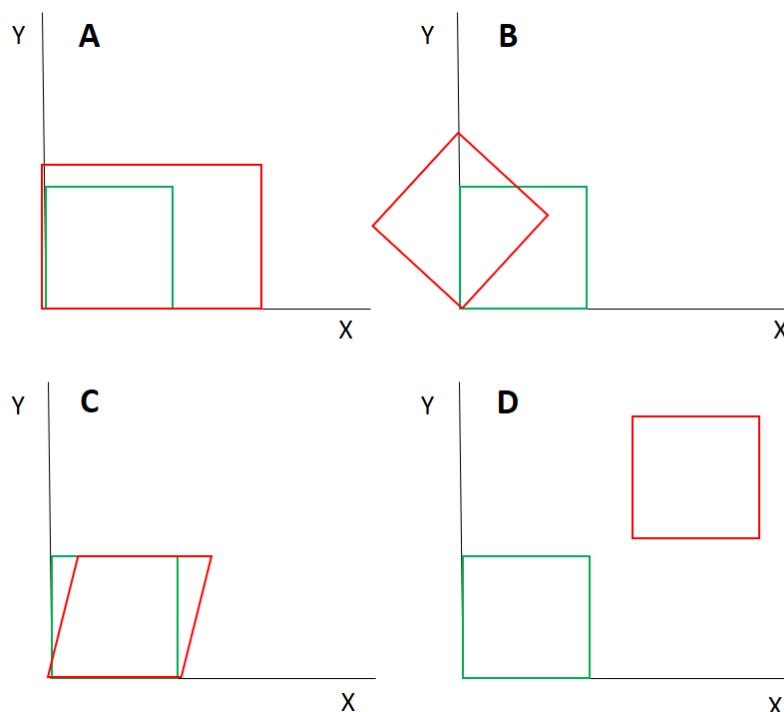
będący pierwiastkiem sumy kwadratu rezydiów podzielonej przez liczbę punktów kontrolnych (RMS ma taką samą jednostkę jak nasze dane).

Podsumowanie procesu georeferencji i rektyfikacji zostało przedstawione na Rysunku 4.27.



Rysunek 4.27 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 rastrowych 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 4.28. 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.

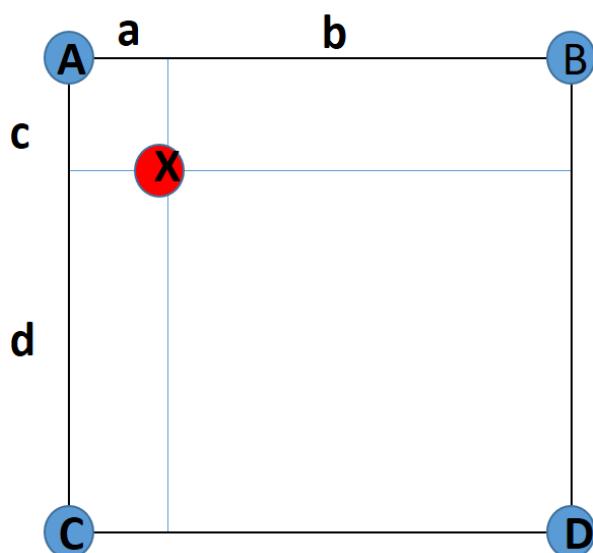


Rysunek 4.28 Cztery rodzaje przekształceń geometrycznych w transformacji afincznej:
A – niezależne skalowanie na osiach YX; B – obrót; C – pochylenie; D - przesuwanie.

Transformacja afinczna charakteryzuje się między innymi zachowaniem równoległości linii na mapie. Wracając do rysunku 4.27, 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 4.27 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 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 4.29.



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

Rysunek 4.29 Metoda interpolacji bilinearnej. Wartość X w środku komórki nowego rastra jest obliczana z czterech wartości ją otaczających.

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 na każdym elemencie rastra, lub w przypadku operacji na kilku rastrach na każdych odpowiadających sobie komórkach rastrowych. 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 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 opartego 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).

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.

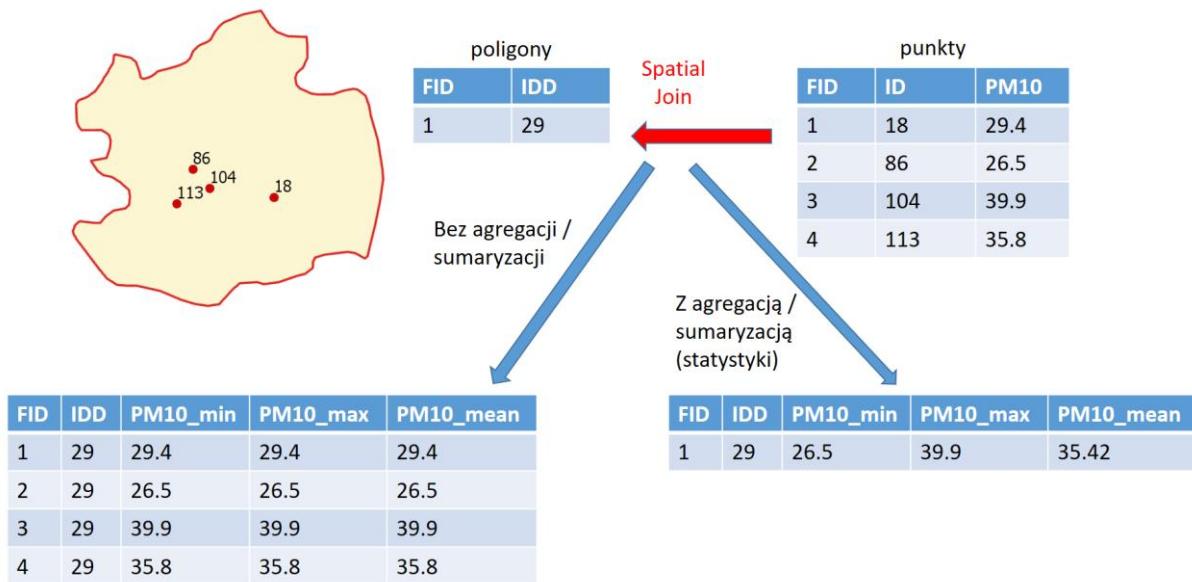
GIS jako narzędzie jest wykorzystywane w wielu dziedzinach (np. ekologii, kryminalistyce, gospodarce przestrzennej, hydrologii, geologii). W jednych korzysta się przeważnie z danych wektorowych w innych z rastrowych. W każdej z tych dziedzin dopracowano się specjalistycznych narzędzi analizy. Niemniej 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.

5.2 Analiza odległości pomiędzy obiektami

Założo