

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 statystyki 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 znajdowania 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 rodzaju 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

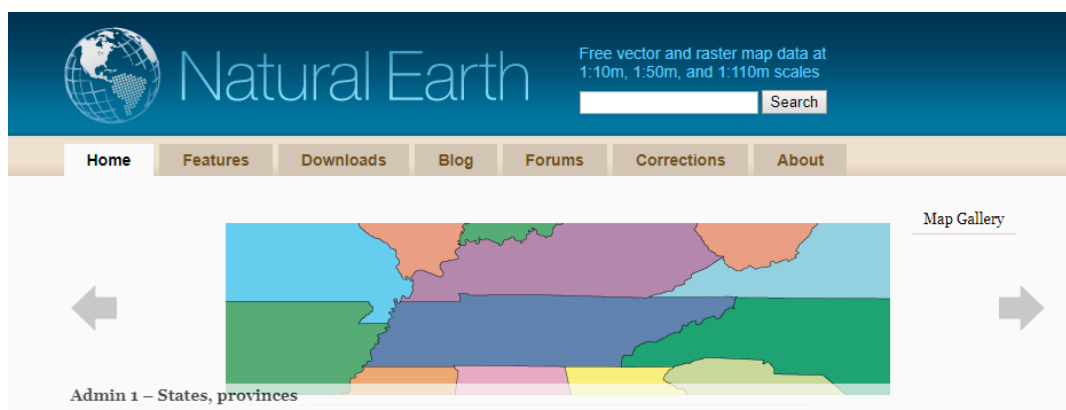
istniejących lub dodane z innych źródeł. Nowy zbiór danych tworzymy także w przypadku danych rastrowych. Może on być wynikiem połączenia kilku rastrów, wycięcia podzbioru z rastra, wybrania niektórych kanałów z kompozytu, zmiany typu danych lub przeliczenia wartości pikseli w rastrze. Nowe dane mogą 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.

Listopad 2019

4.1 Pozyskiwanie i tworzenie danych przestrzennych

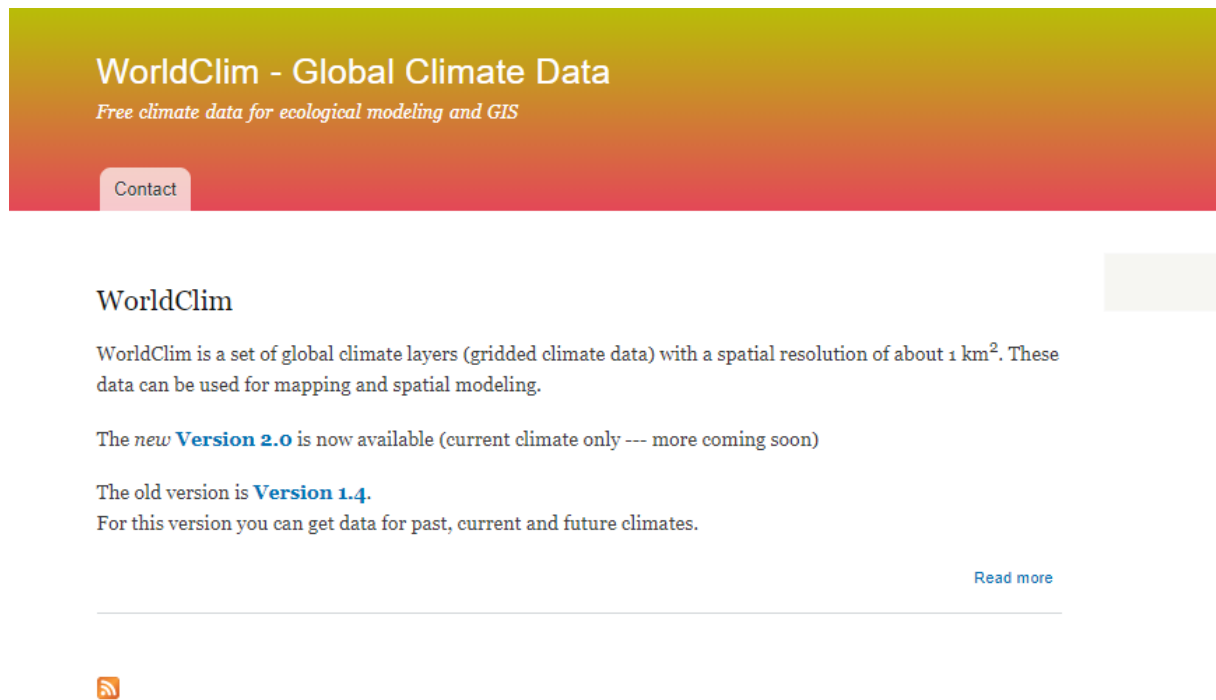
Internet jest obecnie podstawowym źródłem danych dla GIS. Znajdujemy je nie tylko wchodząc na strony reprezentatywnych instytucji, ale także eksplorując różne strony w poszukiwaniu często bezcennych wskazówek, jak pozyskać potrzebne dane. Dane powinny spełniać nasze wymagania i być bezpłatne. Standardem 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/>



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

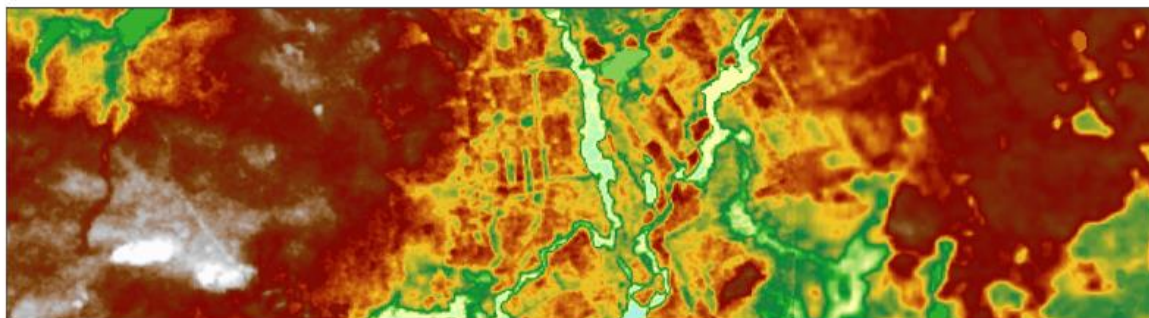
Wiele instytucji, organizacji lub nieformalnych grup badaczy opublikowało tematyczne zbiory danych o charakterze globalnym, które ze względu na rozdzielczość przestrzenną mogą być wykorzystane do projektów lokalnych. Przykładem może być zbiór danych klimatycznych **WorldClim**: <https://www.worldclim.org/>



Dane w formacie GeoTIFF są dostępne w różnych rozdzielczościach 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.

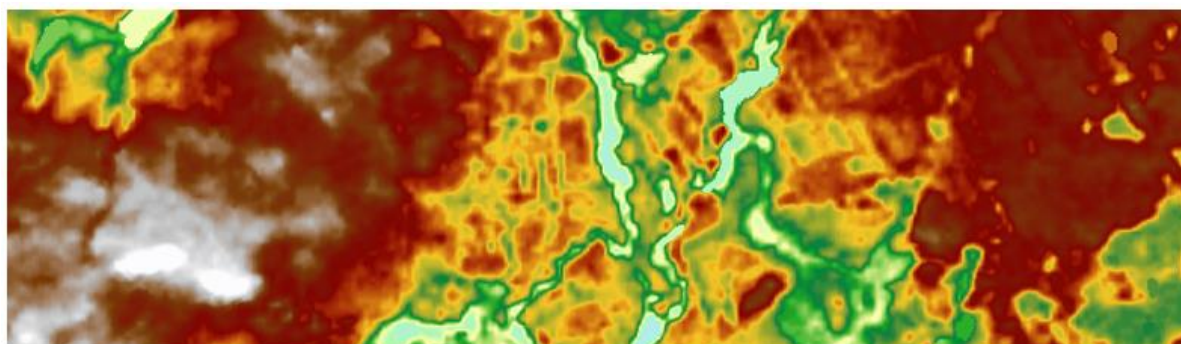
Podstawowym źródłem danych wysokościowych dla całego świata jest mapa cyfrowa **SRTM** (shuttle radar topographic mission). Mapa ta powstała w rezultacie misji wahadłowca Endeavor zorganizowanej w 2000 roku przez NASA. Od 2014 roku rezultaty tej misji w postaci map cyfrowych zostały udostępnione w rozdzielczości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

Poniżej przedstawiono wycinek mapy SRTM.



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

Poniżej przedstawiono wycinek mapy EU-DEM.



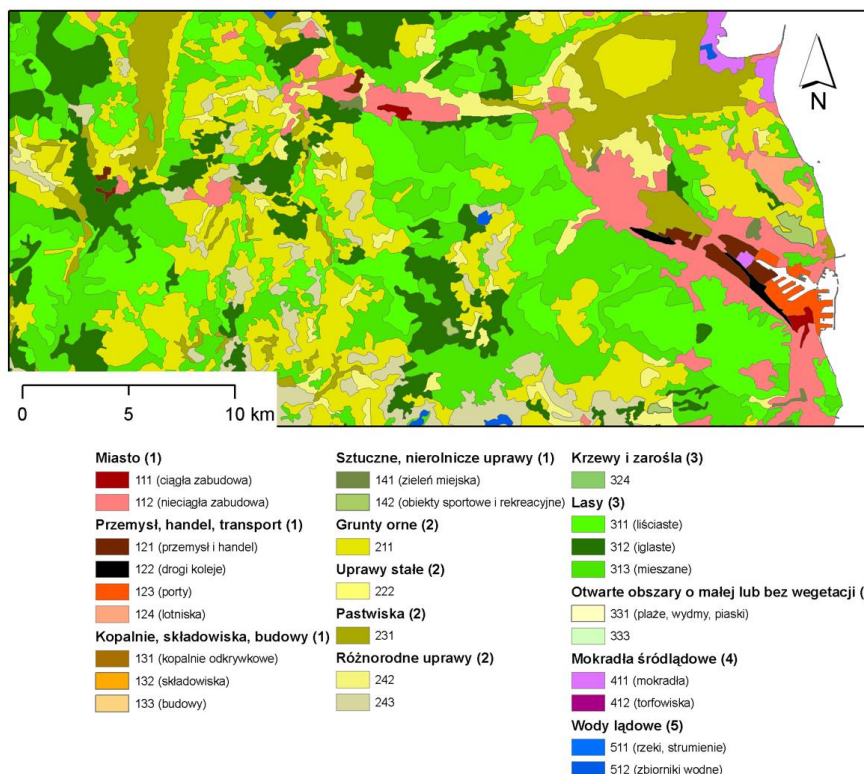
Mapy te można pobrać ze strony: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/view>

The screenshot shows the Copernicus Land Monitoring Service website. The header includes the Copernicus logo and the Land Monitoring Service logo. The main navigation bar has links for Global, Pan-European, Local, and Imagery and reference data. The page title is "EU-DEM v1.1". The map view shows a color-coded elevation map of Europe, with labels for various regions like E20N50, E30N50, E40N50, E50N50, E30N40, E40N40, and E50N40. The map is interactive, with zoom controls and a legend.

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



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.

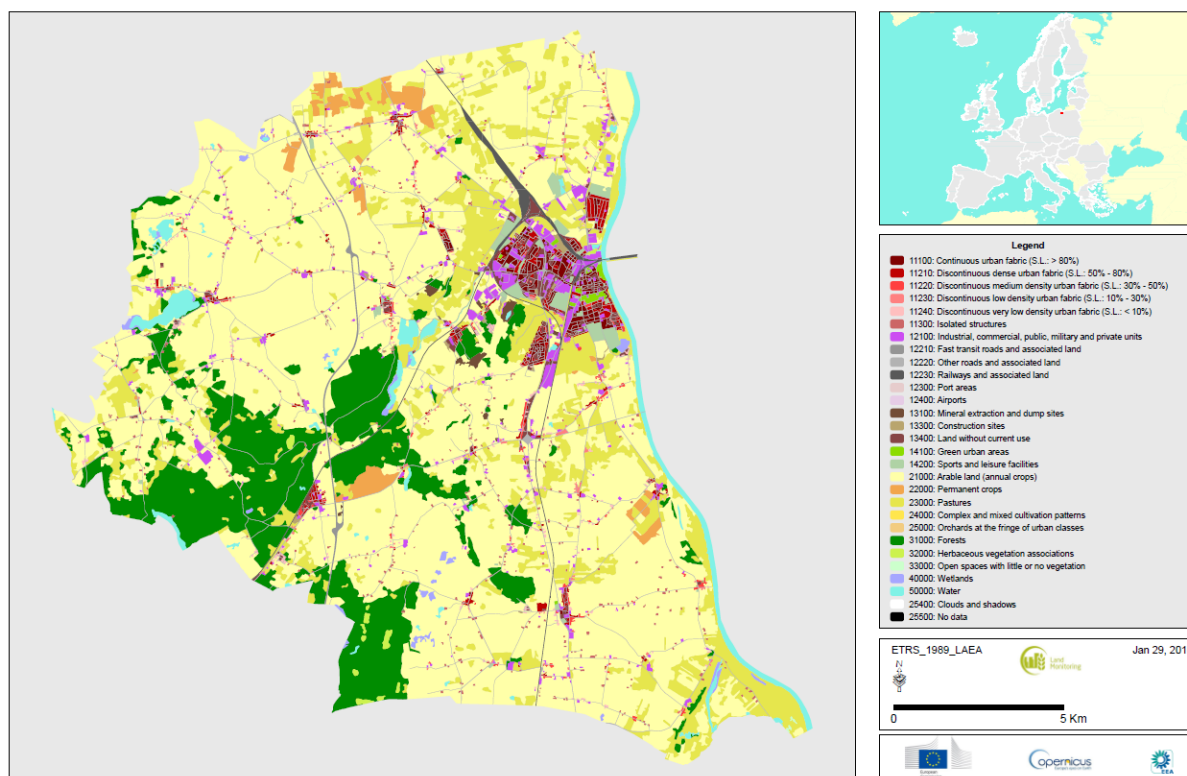


Przykład mapy z danych Corine 2000 wraz z hierarchicznymi klasami występującymi na mapie przedstawiono powyżej. W nawiasach opisano

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



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



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

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

Podstawową stroną przeglądania i dostępu do danych przestrzennych jest **Geoportal2** http://mapy.geoportal.gov.pl/imap/Imgp_2.html?gpmmap=gp0. Dane można pozyskać bezpłatnie lub odpłatnie. Dostęp do danych bezpłatnych można rozpocząć od adresu <http://www.gugik.gov.pl/pzgik/dane-bez-oplat>.



Dane mogą być też pobierane jako usługa **WMS** (Web Feature Service).

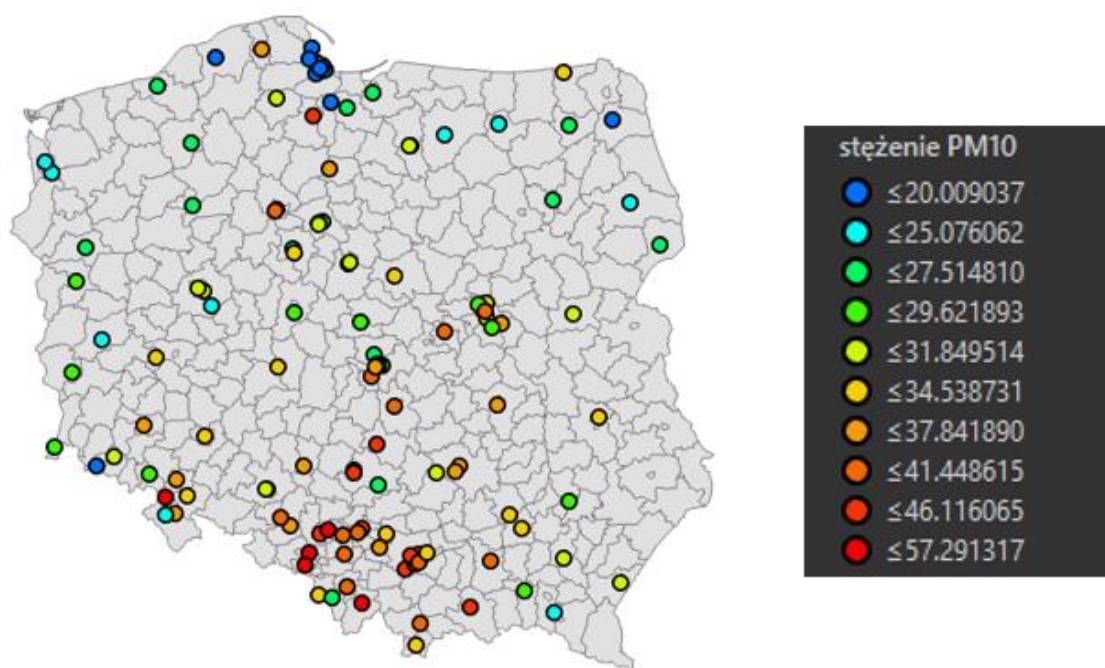
Wiele instytucji udostępnia także dane związane ze swoją działalnością. Centralna Dyrekcja Ochrony Środowiska udostępnia pod adresem: <http://www.gdos.gov.pl/dane-i-metadane> dane w formacie wektorowym (SHP) związane z ochroną przyrody.

Państwowe Gospodarstwo Wodne Wody Polskie udostępnia wektorową Komputerową Mapę Podziału Hydrologicznego Polski w skali 1:50 000 pod adresem <https://dane.gov.pl/dataset/869,komputerowa-mapa-podziau-hydrograficznego-polski> . Mapa obejmuje cały obszar Polski oraz położone poza granicami kraju części dorzecza Wisły i Odry. Ze strony tej można także pobrać dokumentację do MPHP.

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)

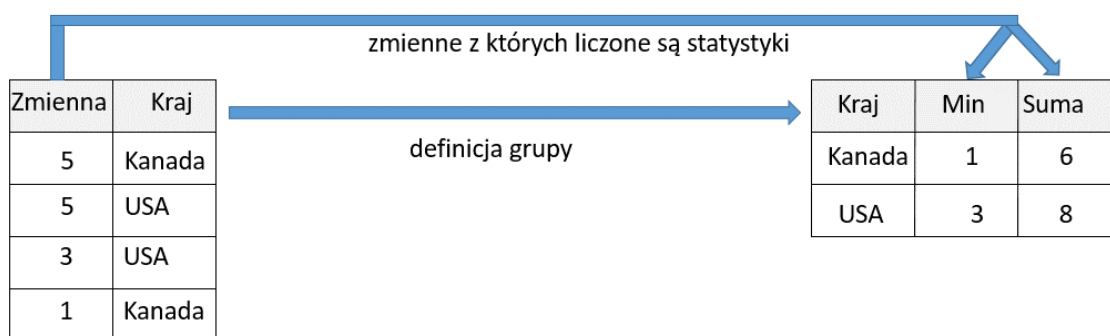
Do każdego obiektu geometrycznego (na Rysunku 4.6 są to punkty) przypisane są jego atrybuty, których wartości wykorzystywane są w procesie **symbolizacji**, polegającym na nadaniu geometrii specyficznego wyglądu. Istnieje szeroka gama potencjalnych możliwości. Na przykład geometria danych punktowych 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 zauważyć 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.

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

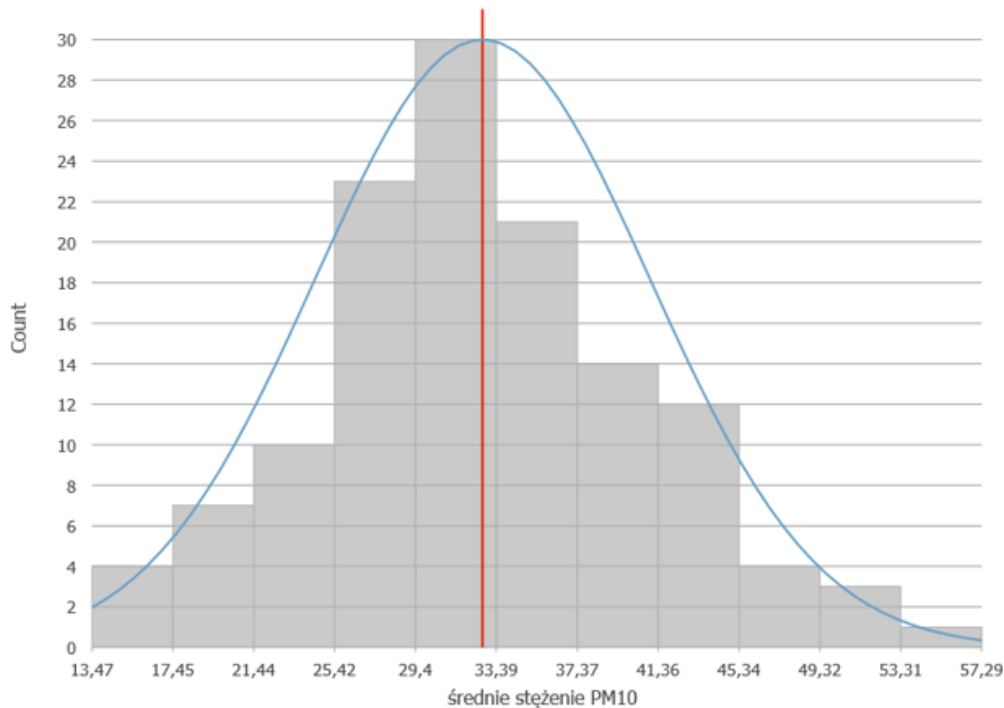


Rysunek 4.7 Sumaryzacja w tablicy atrybutowej

Wartości z pola *Zmienna* są podzielone na podzbiory w zależności od wartości w polu *Kraj*. Pierwszy podzbiór dla wartości 'Kanada' wynosi [5, 1], a drugi dla wartości 'USA' [5,3]. Dla każdej wartości pola definiującego grupy zostają

policzone wybrane statystyki z podzbiorów danych. Zostały one zestawione w tabeli po prawej stronie rysunku.

Jedną z podstawowych metod graficznych jest histogram często uzupełniony przez podstawowe statystyki.

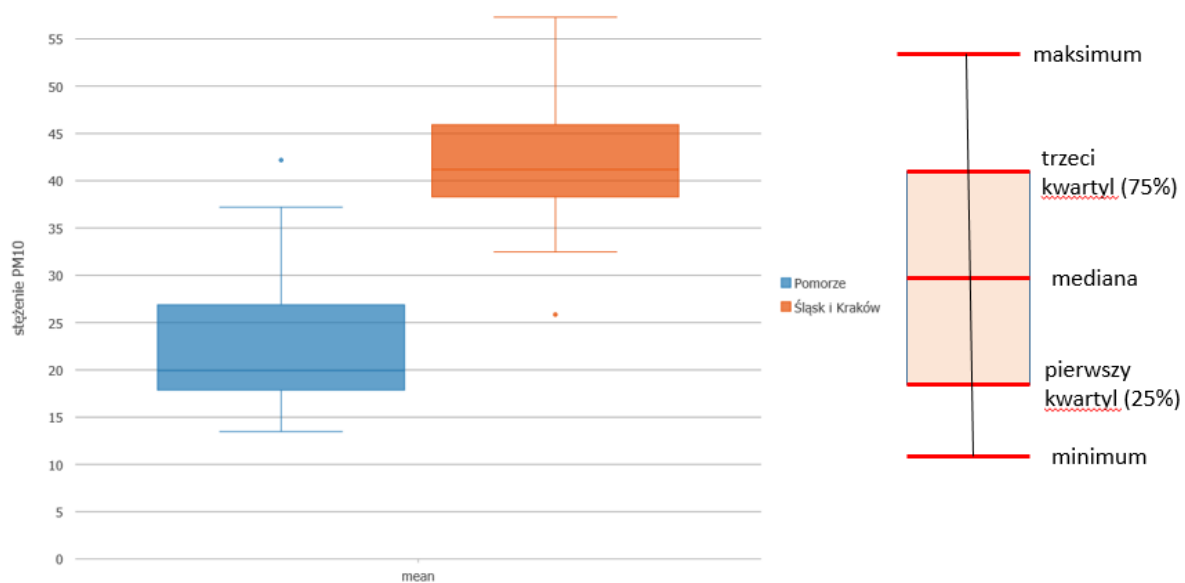


Rysunek 4.8 Histogram ciągłego atrybutu (stężenie PM10) z zaznaczoną średnią i wykresem rozkładu normalnego.

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

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

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

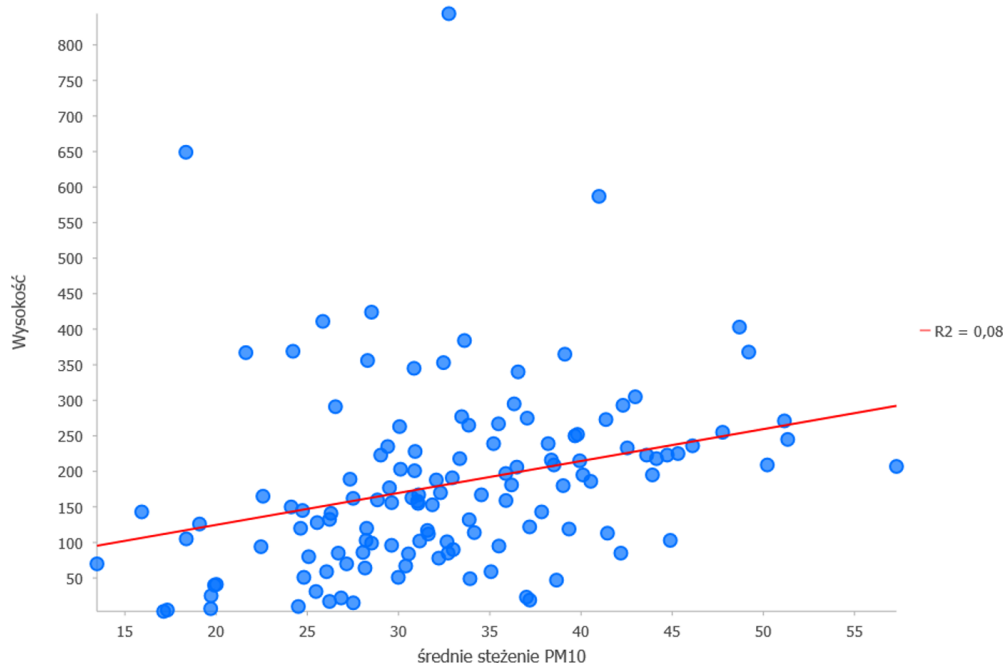


Rysunek 4.9 Wykres typu box-plot pokazuje różnice w stężeniu PM10 pomiędzy stacjami na Pomorzu a Śląskiem wraz z Krakowem

Wykres ten w wizualny sposób przedstawia szereg statystyk analizowanego zbioru danych. Składa się z pionowej linii z 5 poprzeczkami. Wartości statystyk przypisanych do tych linii znajdują się na pionowej osi wykresu. Od drugiej do trzeciej linii tworzy się prostokąt, sprawiający, że wykres jest bardziej czytelny. Krańcowe poprzeczki reprezentują wartość minimum i maksimum. Poprzeczka środkowa reprezentuje medianę (tyle samo jest wartości od niej mniejszych co większych). Druga poprzeczka pokazuje pierwszy kwartył (25% wartości jest od niej mniejszych), a trzecia trzeci kwartył (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 podzbiory danych, jeden dla Pomorza i drugi dla Śląska i okolic Krakowa. Wykres pokazuje nam następujące prawidłowości,

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

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

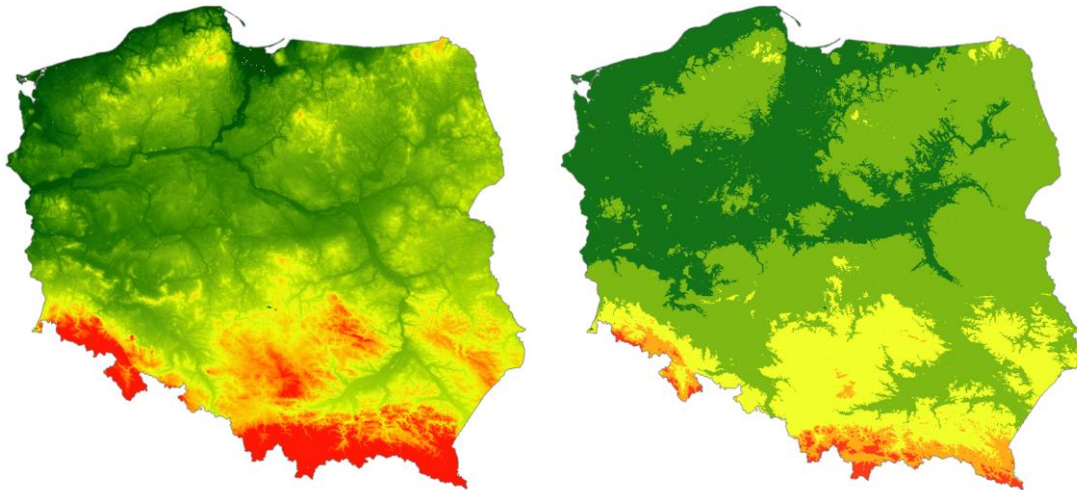


Rysunek 4.10 Wykres rozrzutu (scatter-plot) przedstawia relację między zmiennymi. Reprezentacją tej relacji jest często linia regresji liniowej oraz współczynnik R^2 opisujący zgodność punktów z tą linią.

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

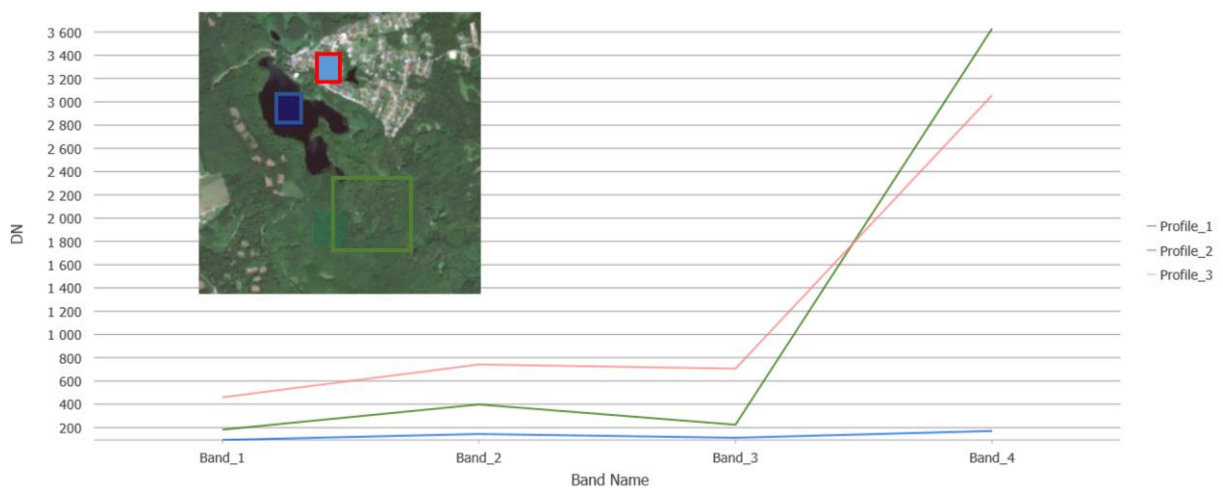
- Statystyka $R^2 = 0.08$ wykazuje brak relacji pomiędzy wartością stężenia i wysokością stacji pomiarowej

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



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

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



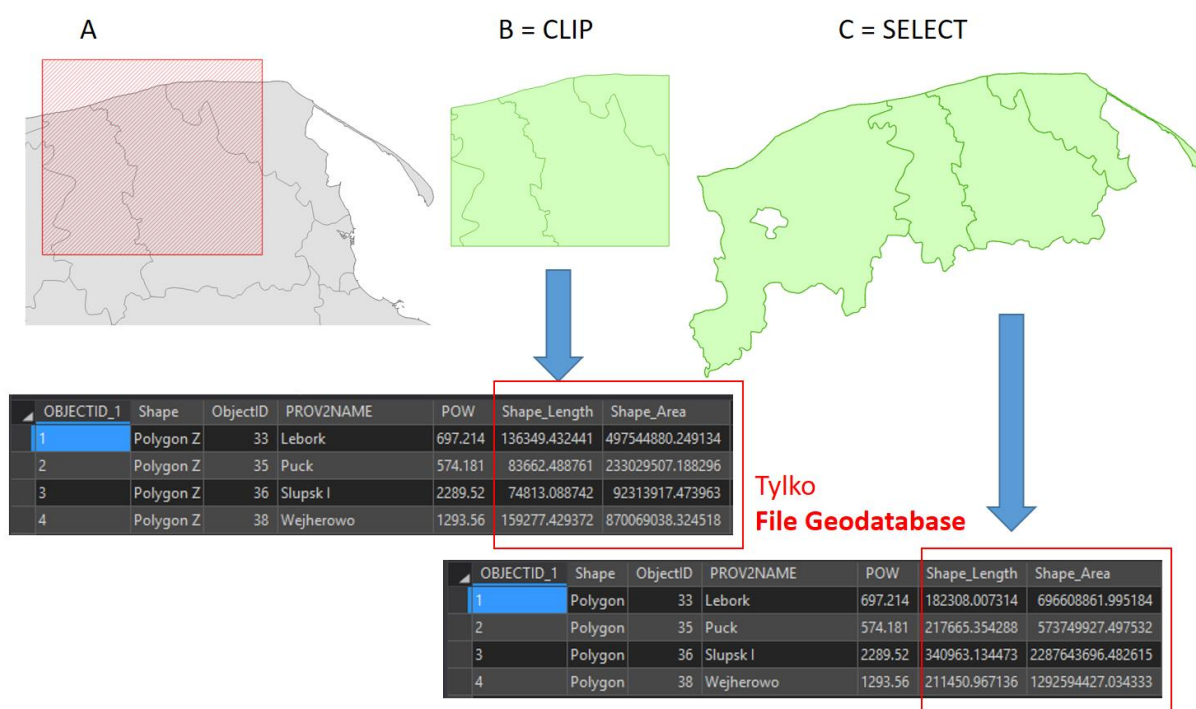
Rysunek 4.12 Profile spektralne (średnie) dla trzech typów pokrycia terenu (las, woda i zabudowania).

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 poligonu 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 nowy 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 poligonem (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).

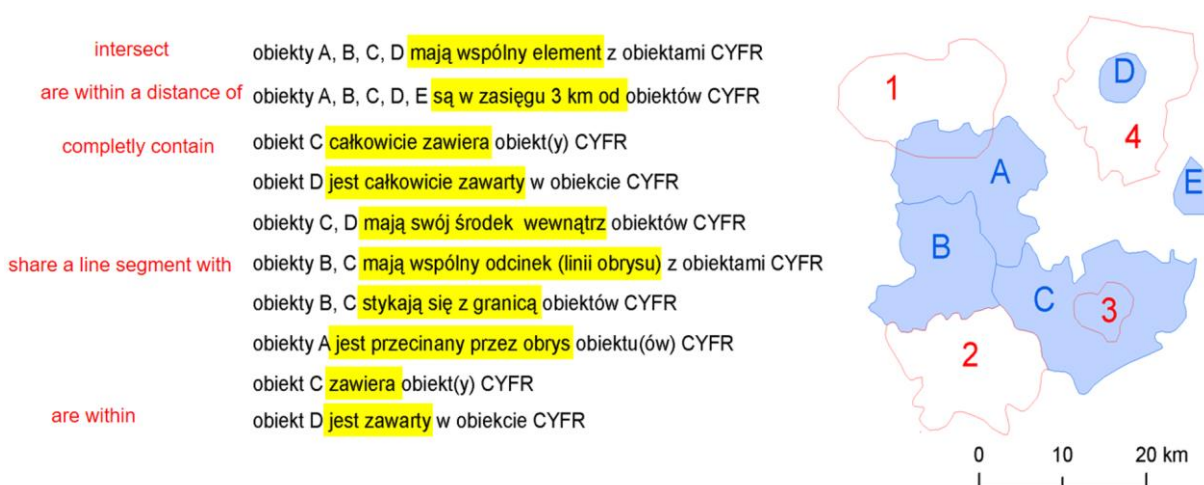
Proces tworzenia podzbiorów rekordów za pomocą selekcji jest bardzo istotną operacją w GIS. Jeżeli wykonamy dowolną operację na zbiorze rekordów z których część jest wybrana (zaznaczona) za pomocą selekcji **to ta operacja zostanie wykonana tylko na zaznaczonych rekordach**.

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.



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 .

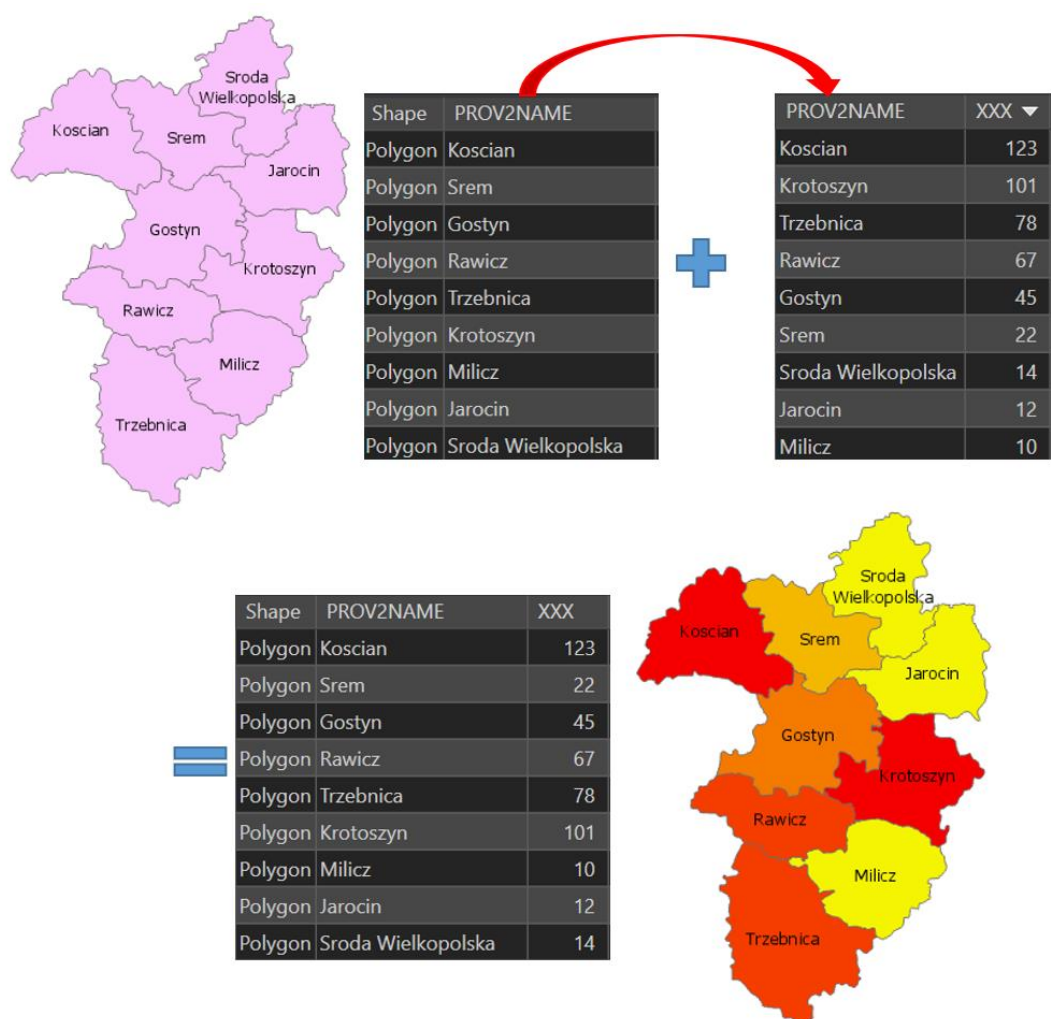
Po utworzeniu podzbioru rekordów należy uporządkować pola (kolumny) tablicy atrybutowej. Może to polegać na usunięciu niepotrzebnych pól lub utworzeniu nowych. Może także zaistnieć potrzeba zmiany nazwy lub typu danych pola. Najprościej jest utworzyć wtedy nowe pole odpowiedniego typu (z nową nazwą),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

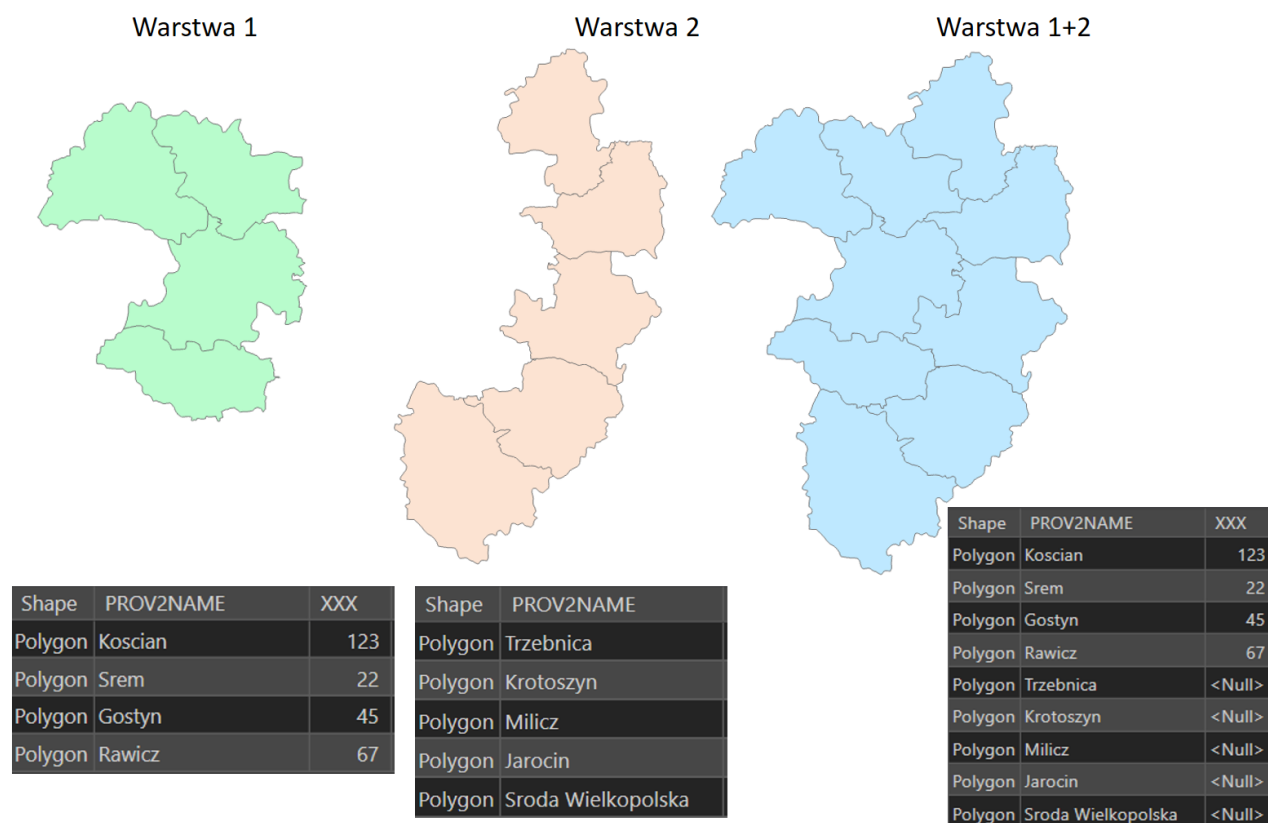
JOIN daje możliwość transferu danych w obie strony pomiędzy danymi przestrzennymi i wyłącznie atrybutowymi. Przykładowo możemy atrybutowe dane zawierające nazwy miast i liczbę ludności wzbogacić o atrybut odległości tych miast od najbliższego miasta wojewódzkiego wykorzystując warstwę przestrzenną miast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

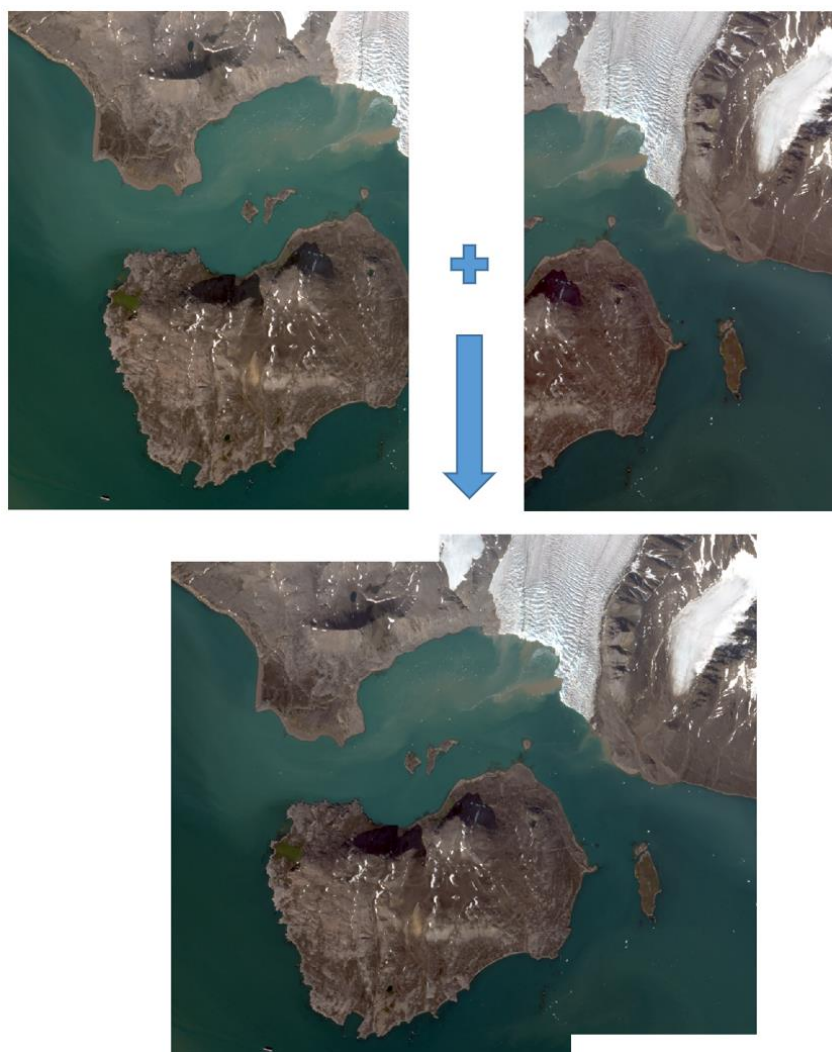
Drugi scenariusz polega na łączeniu kilku warstw w jedną. Do tego celu w GIS służy narzędzie **Merge**. Łączy ono warstwy wektorowe tego samego typu (punkty, linie lub poligony) w jedną warstwę. Nowa warstwa będzie zawierać wszystkie obiekty łączonych warstw. Atrybuty zostają przypisane do istniejących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



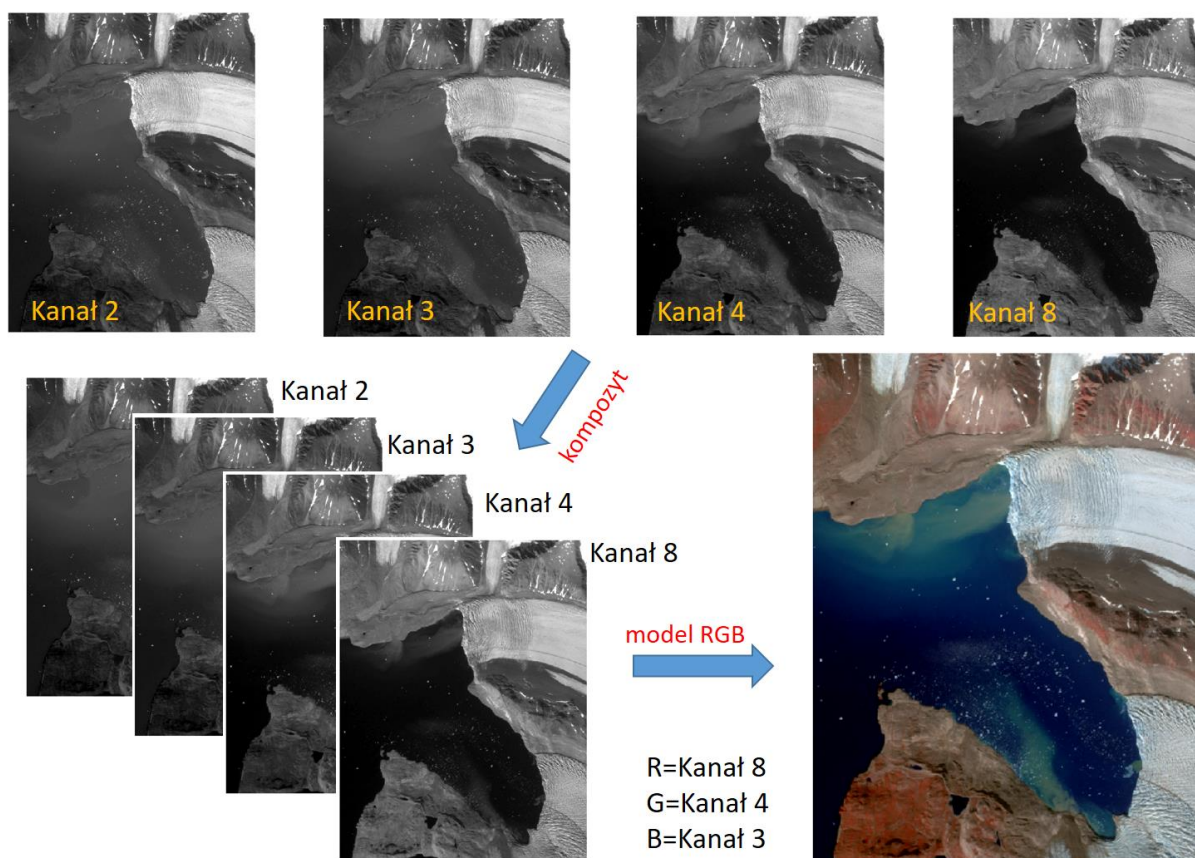
Rysunek 4.16 Łączenie kilku warstw za pomocą Merge.

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 pobrane sceny. Na Rysunku 4.17 mamy dwie oddzielne sceny (zdjęcia satelitarne), które w sumie obejmują wyspę, obiekt naszego zainteresowania. Potrzebujemy połączyć te zdjęcia, aby uzyskać jedno obejmujące całą wyspę. Narzędzie Mosaic służy do tego celu. Połączenie odbywa się na podstawie dopasowania rastrów korzystając z ich referencji przestrzennej. Narzędzie to może być stosowane zarówno do rastrów jedno jak i wielokanałowych. Definiuje się także z którego rastra zostaną pobrane piksele dla wspólnego obszaru.



Rysunek 4.17 Łączenie zdjęć (rastrów) za pomocą MOSAIC.

Dane satelitarne są najczęściej pozyskiwane w postaci oddzielnych kanałów danej sceny (zdjęcia satelitarne). Do analizy eksploracyjnej takiego zdjęcia, jego 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-kanałowych ze zdjęcia Sentinel 2 w kompozyt cztero-kanałowy za pomocą narzędzia COMPOSIT i jego przykładowa symbolizacja.

