## 05.2019 r.

**Przestrzenne bazy danych**

## Projekt

**Temat nr 6**

**Eksploracja przestrzenna danych z rejestrów pokładowych statków oceanicznych z okresu 1750 - 1850**

###### Autorzy: Jakub Gwiazda Miłosz Pluta

**Warszawa 2019**

**1. Charakterystyka danych źródłowych**

Dane źródłowe pochodzą z dzienników pokładowych statków oceanicznych, z lat 1750 – 1850. Wśród wielu motywów ponownego opracowania i analizy danych z dzienników można wyróżnić dwa najważniejsze. Pierwszym była chęć analizy danych historycznych przy pomocy nowoczesnych narzędzi informatycznych, co pozwoliłoby na lepsze zobrazowanie zmian klimatu na morzach i oceanach, jakie zachodziły ponad dwa wieki temu. Natomiast drugim było dostarczenie unikalnych i bezcennych danych dla społeczności naukowej, w celu umożliwienia przeprowadzenia dalszych badań.

Dane pochodzą z dzienników pokładowych statków należących do następujących państw:

* Anglia
* Holandia
* Francja
* Hiszpania
* Argentyna

Podstawowymi parametrami, mierzonymi przez nawigatorów były:

* Prędkość / kierunek wiatru
* Długość / szerokość geograficzna
* Czas
* Temperatura powietrza / wody

Oprócz powyższych były także odnotowywane, jeżeli zachodziły, inne zjawiska występujące na morzu np. liczbę błyskawic lub grzmotów.

Ze względu na brak standaryzacji pomiarów w badanym okresie, należało dane z dzienników odpowiednio przygotować przed wgraniem ich do bazy. W wielu przypadkach zdecydowano się na liczne uproszczenia, aby umożliwić także analizę niedokładnych pomiarów.

Baza danych **CLIWOC**, dostarcza wspomnianych danych w sposób uporządkowany, co ułatwia przeprowadzenie dalszej analizy. Rekordy są opisane szeregiem atrybutów. W tabeli przedstawiono dziesięć wybranych atrybutów wraz z krótkim opisem.

|  |  |
| --- | --- |
| YR | Rok(UTC) |
| MO | Miesiąc(UTC) |
| DY | Dzień(UTC) |
| HR | Godzina(UTC) |
| LAT | Szerokość geograficzna |
| LON | Długość geograficzna |
| C1 | Kod kraju macierzystego statku |
| W | Szybkość wiatru |
| AT | Temperatura powietrza |
| D | Kierunek wiatru |

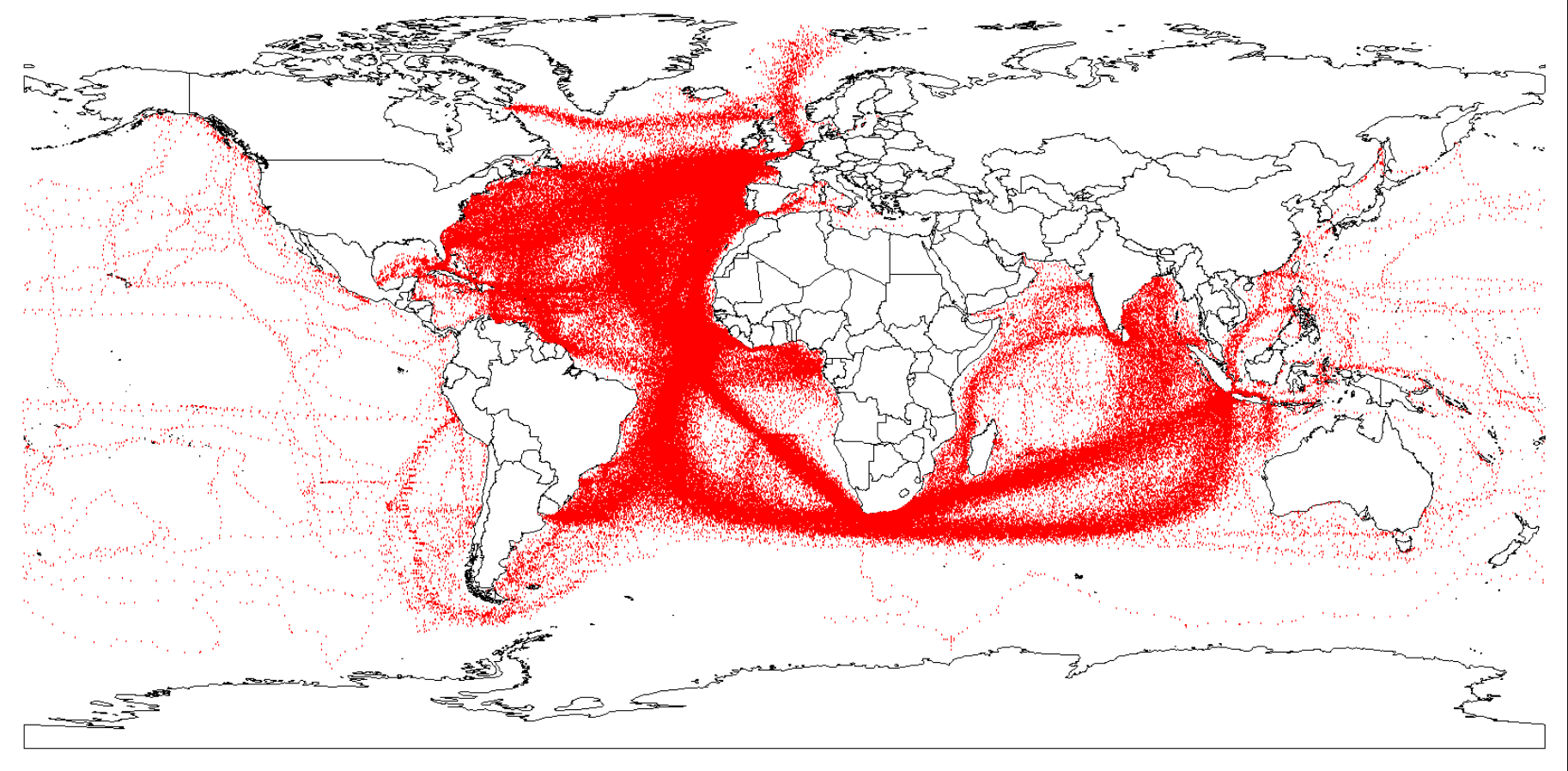
**Prezentacja danych źródłowych**

Pierwszym etapem wykonania projektu było wstępne przygotowanie danych źródłowych. W ramach tego etapu zaimplementowano:

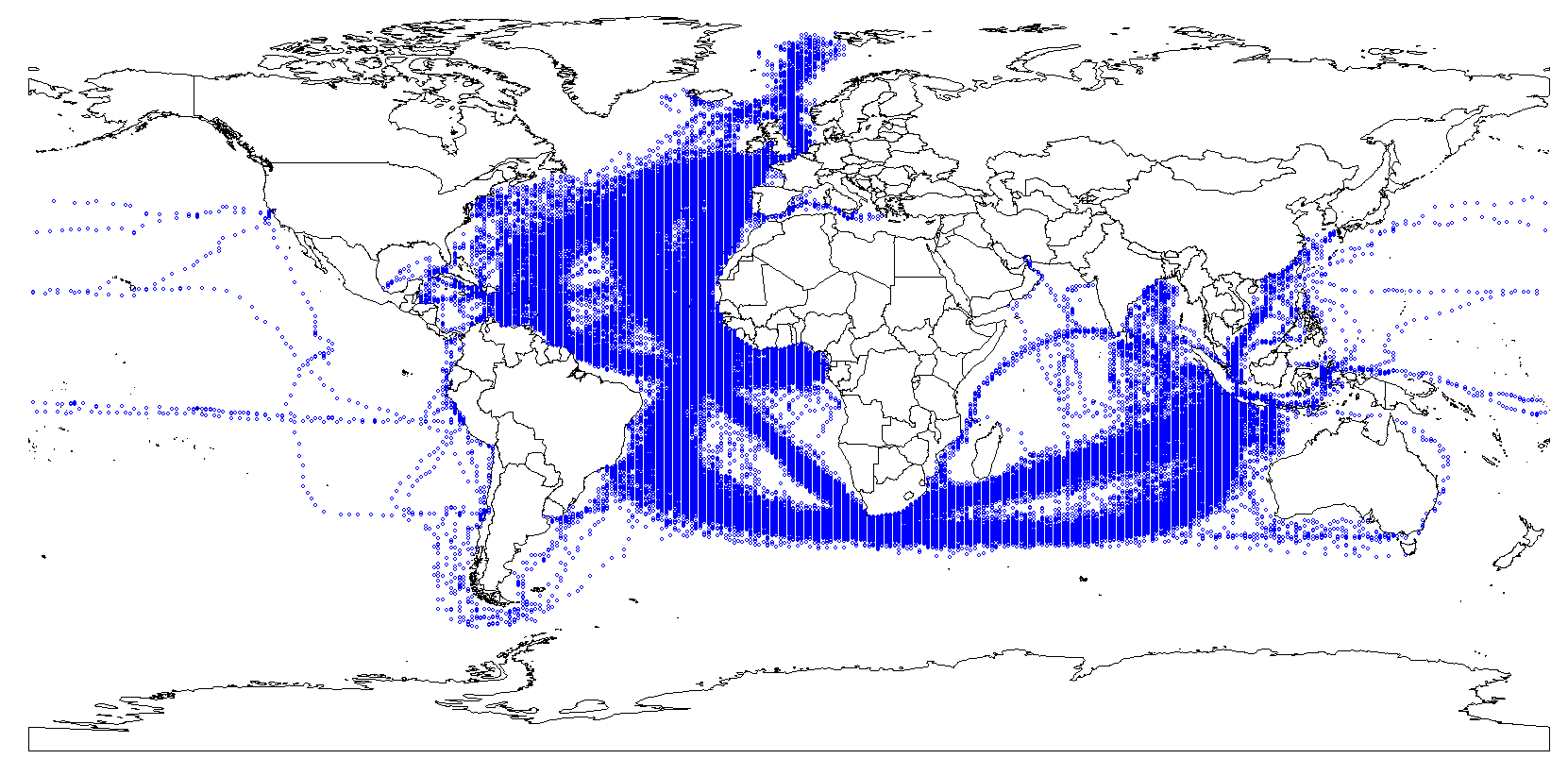
* Wczytywanie danych
* Normalizacje współrzędnych geograficznych(długość / szerokość)
* Wizualizacja znormalizowanych danych

Współrzędne geograficzne w kolumnach **LAT** oraz **LON** wymagały dostosowania do zakresów współrzędnych obsługiwanych przez biblioteki języka R. Po wykonaniu normalizacji zdecydowano się na wizualizację danych na mapie świata.

Przedstawienie wszystkich tras statków oceanicznych w latach 1750 – 1850



Przedstawienie tras oceanicznych dla statków z państwem macierzystym Holandia



**2. Cel badania**

Celem badania jest przeprowadzenie analizy danych źródłowych przy użyciu wybranych funkcji z języka R, służących do analizy danych przestrzennych. Poniżej przedstawiono eksperymenty, których wyniki zostaną przedstawione w kolejnym rozdziale:

* Ex 1 do uzupełnienia gdy będziemy wiedzieć jakie są exp
* Ex 2 do uzupełnienia gdy będziemy wiedzieć jakie są exp
* Ex 3 do uzupełnienia gdy będziemy wiedzieć jakie są exp
* Ex 4 do uzupełnienia gdy będziemy wiedzieć jakie są exp

Wynikiem wymienionych eksperymentów mogą być ciekawe wnioski na temat danych klimatycznych(np. prędkość wiatru), ale także na temat danych dotyczących ruchu statków po oceanach.

Każdy z eksperymentów będzie składał się z czterech części:

* Cel eksperymentu
* Opis wykorzystywanej funkcji z języka R
* Wizualizacja wyniku
* Analiza wyniku

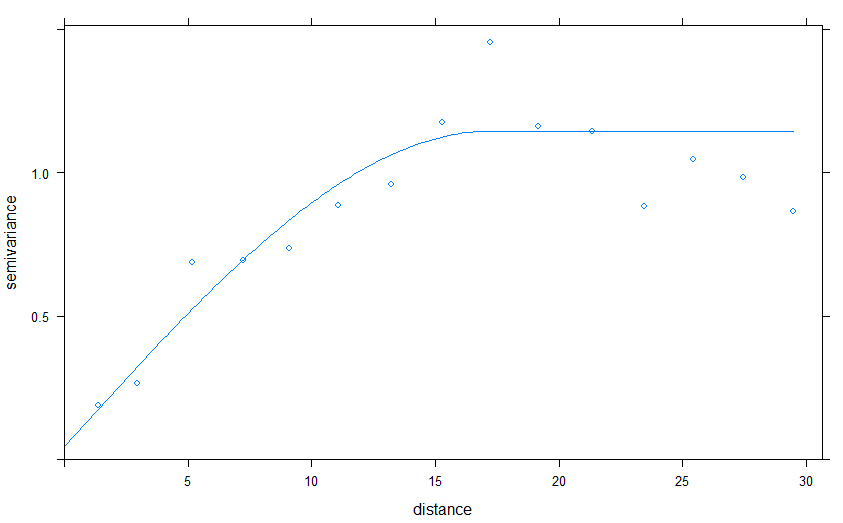
**3. Analiza przestrzenna danych**

// tu już same eksperymenty -> opis wybranej funkcji / wynik / analiza wyniku   
// pewnie min. 4 trzeba funkcje tak opracować

**3.1 Kriging – aktualnie opracowuje**

Celem eksperymentu jest predykcja wartości siły wiatru na podstawie wybranych danych początkowych(?statki hiszpańskie np.?)

Kriging opis – tutaj opisać dokładnie fun. Jak i metodę

Wizualizacja – model  


Wizualizacja heatmapa – todo

Wnioski - todo

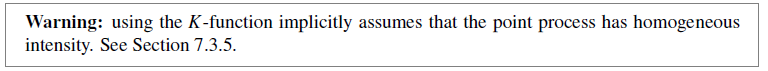
####################################################

**Uwagi – do rozkminy**

**Funkcje K,L,G z pakietu spatstat**

Powyższe funkcje służą do stwierdzenia czy wprowadzone dane(w postaci punktów o współrzędnych x,y) w jakiś sposób się grupują(klastrowanie) lub takie związki nie zachodzą.

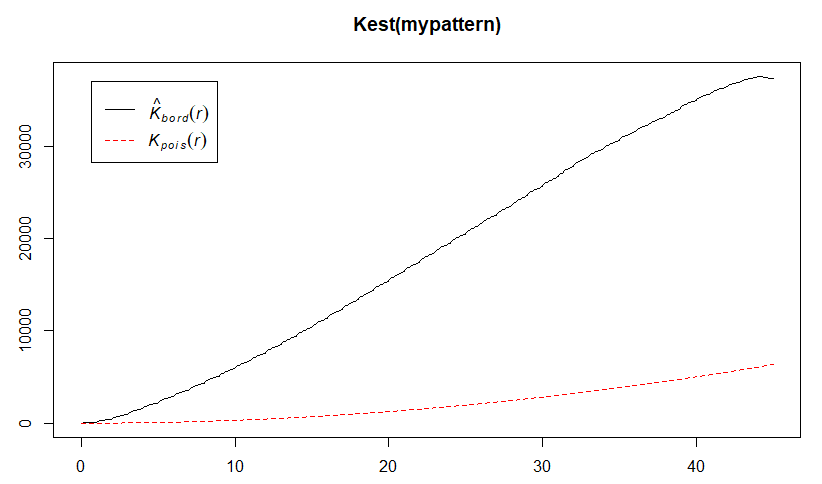
Str. 92 <https://training.fws.gov/courses/references/tutorials/geospatial/CSP7304/documents/PointPatterTutorial.pdf>  
  
U nas nie widzę dla tego zastosowania, ponieważ jedynymi współrzędnymi jakimi można określić położenie jest długość / szerokość. 1) jest podana w stopniach co już utrudnia przełożenie na liczenie odległości 2) dla nas zbędna jest informacja czy punkty gdzie znajdował się statek tworzą / nie tworzą klastrów

Kolejną przesłanką że powyższe funkcje nie mają zastosowania do naszych danych jest wpis z książki: ‘Spatial Points Patterns: Methodology and Applications with R’:   
  


Nasz zbiór punktów nie jest homogeniczny(lądy są puste od punktów) dlatego użycie tych funkcji byłoby zakłamane

Str 7 – zbiór homogeniczny wyjaśnienie  
<https://training.fws.gov/courses/references/tutorials/geospatial/CSP7304/documents/PointPatterTutorial.pdf>

Jeżeli chodzi o samą bibliotekę spatstat to można poszukać ewentualnie innych funkcji, które pozwalają na inne operacje(najpierw jednak bym analizował pozostałe wymienione z pdfa)

Próba rysowania wykresu nawet ‘na pałe’ wygląda tak jak poniżej, ale jest on przekłamany   


####################################################

**4. Porównanie wyników**

//Tu przede wszystkim trzeba się zorientować jak można porównać wyniki z tych różnych //funkcji

**5. Wnioski**