

Projekt zaliczeniowy: Analiza wybranych warunków aerosanitarnych Poznania w latach 2008-2015.



Wytyczne do projektu końcowego:

- **Podany poniżej schemat postępowania jest jedynie przykładem. Korzystając z informacji dostępnych w kursie pogramcydanych.icm.edu.pl; z podręcznika Przemysława Biecka lub Marka Gągolewskiego (dostępny za darmo przez stronę biblioteki uniwersyteckiej) można analogiczny efekt uzyskać za pomocą kilkunastu instrukcji programistycznych.**
- Przesłany projekt końcowy powinien mieć charakter raportu zawierającego zarówno kod programistyczny jak i komentarze do stosowanych poleceń wynikających z treści polecenia. W programie RStudio roboczą wersję raportu na podstawie przygotowanego kodu źródłowego można wygenerować np. w formacie MS Word wybierając z górnego menu opcję: `File -> Compile Notebook`. Przy pierwszym uruchomieniu program powinien nas poinformować o konieczności pobrania i zainstalowania dodatkowych komponentów.
- Stworzony roboczo dokument powinien stanowić szkielet, który następnie możemy edytować w dowolnym edytorze tekstowym (np. MS Word, Libre Office, Google Docs).
- Plik raportu powinien zawierać również jasno i klarownie opisane wyniki w odniesieniu do kolejnych podpunktów treści polecenia, tak aby nie było potrzeby uruchamiania kodu programistycznego w celu uzyskania wyników.
- W przypadku generowania wykresów, map lub innych form grafiki, również powinny one znaleźć się w przesłanym pliku raportu.
- Wyniki operacji kontrolnych (np. w pętlach programistycznych często stosowana kontrolnie instrukcja "print(i)" pozwalająca na określenie przebiegu pętli) - powinny zostać usunięte z finalnej wersji raportu jeśli nie wnoszą istotnych treści
- Wielkość grup projektowych: max. 2 osoby
- Projekt końcowy powinien zostać przesłany w formie pojedynczego pliku (PDF, DOC, DOCX, ODT) w nieprzekraczalnym terminie do 22.12.2016 r. na adres: nwp@amu.edu.pl
- Język programowania i sposób wykonania obliczeń oraz charakter projektu końcowego jest dowolny

-
- W razie wątpliwości przy korzystaniu z R w pierwszej kolejności użyj dostarczanego systemu pomocy (?nazwafunkcji) oraz literatury przedmiotu (Biecek 2013). Często odpowiedź można znaleźć w portalach tematycznych (np. Google, Stackoverflow).

Główne elementy oceny projektu końcowego:

- Umiejętność samodzielnego i prawidłowego rozwiązywania zdefiniowanego problemu badawczego
- Stosowanie instrukcji programistycznych pozwalających na automatyzację pracy (np. użycie pętli programistycznych: for i while, instrukcji sterujących if, else)
- Forma prezentacji i poprawność uzyskiwanych wyników obliczeń (etykiety osi, tytuł, itd.)

Cel projektu:

[1] Kompleksowa ocena warunków aerosanitarnych poprzez stworzenie kodu programistycznego umożliwiającego obliczenia średnich godzinowych, dobowych, miesięcznych i rocznych stężeń pyłów zawieszonych o średnicy aerodynamicznej poniżej 10 mikrometrów (PM_{10}). [2] Określenie rocznej liczby przekroczeń dopuszczalnych wartości średnich dobowych ($>50 \mu g/m^3$) w latach 2008-2015

Zadanie 1.

A: Wczytanie danych obserwacyjnych do środowiska RStudio

Dane do ćwiczenia zamieszczono na stronie www.openmeteo.pl/programowanie

A.1. Pobierz plik pm10.csv i zapisz go w specjalnie utworzonym na tą okoliczność katalogu. Pobrane archiwum zawiera wartości pomiarowe warunków aerosanitarnych w Poznaniu w latach 2008-2015. 2 ostatnie kolumny w pliku oznaczają godzinową koncentrację PM_{10} i $PM_{2.5}$. Dane pochodzą z tabeli generowanej na stronie Wojewódzkiego Inspektoratu Ochrony Środowiska w Poznaniu dla stacji pomiarowej Polanka: <http://powietrze.poznan.wios.gov.pl>

A.2. Otwórz plik i zapoznaj się z jego strukturą. (Na Windowsie znak nowej linii różni się od konwencji UNIXowej, więc prawdopodobnie plik może nie wyświetlać się poprawnie w notatniku. Warto skorzystać z opcji zawijania wierszy i zmiany rozmiaru okna w celu poprawnego

wyświetlania wartości w kolumnach.

A.3. Uruchom program RStudio. Wskaż ścieżkę do katalogu roboczego, w którym rozpakowałeś(aś) dane. Najszybciej katalog ten można wskazać poprzez wybranie z górnego menu opcji: `Session -> Set working directory -> Choose directory` (skrót `ctrl+shift+h`)

A.4. Aby wyświetlić czy katalog został poprawnie wskazany sprawdź wynik komendy:

```
getwd()
```

A.5. oraz wyświetl wszystkie zawarte w nim pliki komendą:

```
dir()
```

A.6. wczytaj plik za pomocą komendy `read.table` lub `read.csv`.

Argumentem, który będzie wymagał modyfikacji jest deklaracja dla brakujących wartości. W naszym pliku są one zapisane jako kreska (znak minusa). Aby komputer poprawnie je zinterpretował (nadał wartości 'NA') musimy ustawić wartość kolejnego argumentu jako:

```
na.strings="-" .
```

A.7. Jeśli udało Ci się stworzyć poprawną komendę do wczytywania plików w poprzednim punkcie możemy teraz wczytać zbiór danych do nowego obiektu, który nazwiemy '**dane**'

A.8. Nasza analiza powinna dotyczyć jedynie parametru PM_{10} , stąd w ostatecznym kształcie naszej bazy danych chcemy się pozbyć zbędnych kolumn. Stwórz nowy obiekt nazwany '**baza**', w którym będą przechowywane jedynie wartości dla: lat, miesięcy, dni, godzin oraz koncentracji PM_{10} .

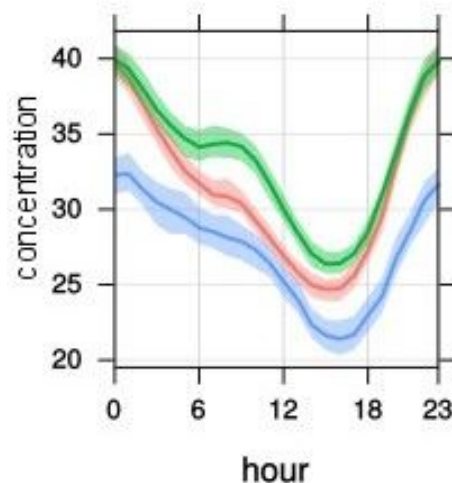
Zadanie 2.

Obliczanie średnich wartości stężeń zanieczyszczeń atmosferycznych PM_{10} w Poznaniu w latach 2008-2015.

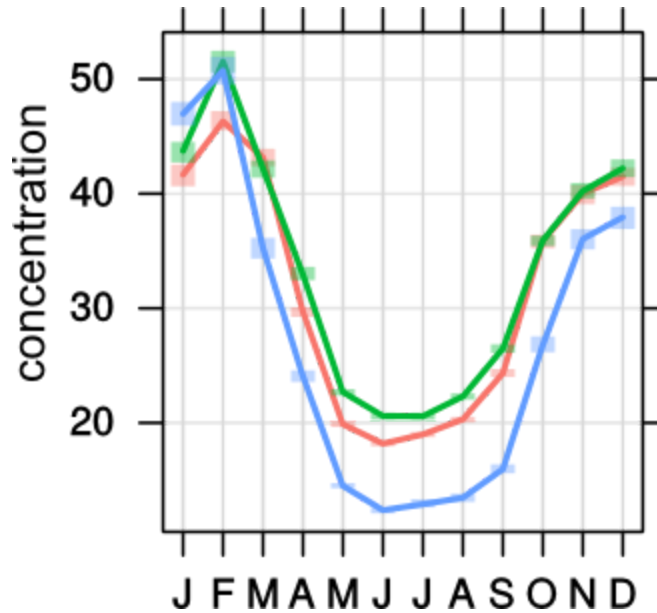
A: Wartości średnie dla poszczególnych godzin - określenie rytmu dobowego zmian koncentracji PM_{10}

A.1. Jednym ze sposobów obliczenia średnich wartości stężeń w poszczególnych godzinach w okresie 2008-2015 jest użycie instrukcji `which` do poprzednio stworzonego zbioru **'baza'**. Znajdź numery rzędów, w których przechowywane są wartości dla kolejnych godzin (01, 02, itd...). Następnie uzyskane indeksy podstaw jako numery elementów dla wartości z kolumny PM_{10} i oblicz ich wartości średnie.

A.2. Na podstawie uzyskanych wartości stwórz wykres zmienności dobowej stężeń PM_{10} w Poznaniu w latach 2008-2015. Powinien on wyglądać podobnie do wykresu załączonego poniżej (oczywiście linia powinna być jedna zamiast kilku). Pamiętaj aby zaznaczyć na osi wykresu, że w naukach atmosferycznych czas, którym się posługujemy to czas uniwersalny (UTC) a nie czas lokalny.



A.3. Wykonaj dodatkowy wykres analogicznie dla przebiegu wartości średnich miesięcznych. Wykres powinien przedstawiać stan średni w latach 2008-2015 (jak na przykładzie poniżej):



B: Wartości średnie miesięczne.

B.1. W celu otrzymania wartości średnich miesięcznych stężeń PM_{10} w latach 2008-2015 za każdym razem będziemy mieli do czynienia z różną liczbą dni. Często na automatycznych stacjach pomiarowych zdarzają się błędy w zapisie lub transmisji danych skutkujące np. brakami pomiarów lub utraceniu rekordów dla całych dni lub godzin. Z tego względu nie możemy przyjąć, że każde 24 kolejne rekordy stanowią nowy dzień a ilość dni w miesiącu jest zgodna z jego rzeczywistą długością (choć wyjątkowo w naszej bazie wszystko wskazuje na to, że moglibyśmy tak zrobić). Zatem aby policzyć wartości średnie miesięczne stężeń PM_{10} będziemy dość szeroko stosować instrukcje `which` (oraz dla chętnych `unique`).

Koncepcja, którą możemy w takim przypadku zastosować będzie bazować na zastosowaniu zagnieżdżonych pętli `for` z czego jedna będzie odnosiła się do kolejnych lat, a druga do miesiący, dokładnie tak, jakbyś chciał(a) wygenerować kolejno wartości:

- 2008 - 01
- 2008 - 02
- ...
- 2008 - 12
- 2009 - 01

2008 - 02

...

2015 - 12

Taki schemat można uzyskać np. w taki sposób

```
for (rok in 2008:2015){
  for (miesiac in 1:12) {

    # jakasinstrukcja1
    paste(rok, miesiac)
    ...
  } # zamkniecie petli dla miesiecy

} # zamkniecie petli dla lat
```

B.2. Jeśli pierwszą wygenerowaną wartością dla lat jest 2008, dla miesiący 1 (styczeń), to wówczas możemy sprawdzić jakie rzędy w naszej bazie danych zwracającej spełniają nam taki warunek logiczny.

B.3. Zanim przystąpimy do pisania całego rozwiązania w pętli, przetestujmy działanie funkcji `which` i sprawdźmy, czy taki sposób postępowania będzie działać poprawnie. Stwórz zatem zmienną `rok = 2008` i zmienną `miesiac=1`.

B.3.1. Za pomocą instrukcji `which` sprawdź jaki zakres rzędów (indeksy) odpowiadają temu okresowi (01-2008) i zapisz ten wynik (czyli numery wierszy) do zmiennej `'ind'`

B.3.2. Wykorzystaj uzyskany zakres rzędów (zmienna `'ind'`) i zastosuj go do wybrania interesujących nas wartości (nawiasy kwadratowe) z kolumny z wartościami PM_{10} .

B.3.3. Oblicz średnią arytmetyczną za pomocą funkcji `mean`. Zmodyfikuj argumenty funkcji tak, aby pomijane były przy obliczeniach wartości brakujące (NA). Skorzystaj z opcji pomocy (`?`) aby sprawdzić jak ten zapis powinien wyglądać. Uzyskany wynik zapisz do zmiennej tymczasowej `tmp`. Jako wynik kolejnej nowej zmiennej (`tmp2`) złącz po kolumnach (funkcja `cbind`) wartości dla roku, miesiąca i średniej koncentracji PM_{10} (czyli zmiennej `tmp`).

B.4. Sposób postępowania przedstawiony w punkcie B.3. zmodyfikuj w taki sposób aby wewnątrz wcześniej przygotowanego schematu dla zagnieżdżonych pętli, uzyskiwany wynik tymczasowy w obiekcie `tmp2` był dołączany w kolejnych rzędach do przygotowanego wcześniej

pustego obiektu **miesieczne**. Wskazówka: pusty obiekt można stworzyć przypisując mu wartość NULL

B.5. Przedstaw w formie wykresu zmienność uzyskanych średnich miesięcznych koncentracji PM_{10}

B.6. Wyświetl statystyki testowe uzyskanej bazy danych za pomocą funkcji `summary()`

=====

C: Wartości średnie dobowe

C.1. Zmodyfikuj schemat uzyskany w punkcie B. aby obliczyć (np. w pętlach `for`) wartości średnich dobowych stężeń PM_{10} w latach 2008-2015. Wynik zapisz do nowej tabeli i wyświetl jego podstawowe statystyki za pomocą funkcji `summary`.

C.2. Ile dni z przekroczeniem dobowych stężeń dla wartości progowej powyżej $50\mu g/m^3$ wystąpiło w poszczególnych latach?

Schemat postępowania w celu wykonania projektu końcowego jest dowolny, z użyciem dowolnego współczesnego i bezpłatnego języka programowania.