

04. Grafika - zadania

4.1 Zainstaluj pakiet "nycflights13" (4,4 MB) poprzez polecenie `install.packages("nycflights13")`. Po jego zainstalowaniu mamy do dyspozycji 5 ramek danych:

- flights - informacje o wszystkich lotach z lotniska NYC w 2013 roku
- airports - nazwy lotnisk i ich lokacje (długość i szerokość geograficzna)
- planes - informacje o konstrukcji poszczególnych samolotów
- airlines - rozwinięcie dwuliterowych kodów na nazwy linii lotniczych
- weather - dane meteorologiczne w rozbiciu na godziny dla każdego lotniska

Zwizualizuj dane wypisane w kolejnych podpunktach. Pamiętaj, że w "prawdziwym życiu" często te wykresy będziesz prezentować komuś innemu - zadbaj o odpowiednie tytuły wykresów, nazwy osi i inne rzeczy poprawiające czytelność i zrozumienie wykresu.

- A. Jaki jest rozkład liczby silników w samolotach (kolumna engines w planes)
- B. Rozkład typów samolotów (kolumna type w planes)
- C. Wybierz jeden konkretny dzień i lotnisko, a następnie dla tego konkretnego dnia i lotniska zrób wykres zmiany ciśnienia w ciągu kolejnych godzin.
- D. Do wykresu z punktu C dodaj prędkość wiatru, a następnie zapisz wykres do pliku PDF
- E. Zagreguj temperaturę na wybranym lotnisku w poszczególnych dniach (np. używając średniej z 24 godzin) i przedstaw zmiany temperatury w kolejnych dniach (nie interesują nas zmiany między godzinami)
- F. Przedstaw rozkład temperatur za pomocą wykresu skrzynkowego dla każdego z trzech lotnisk (zarówno w rozkładzie godzinowym, jak i dniowym). Co możemy odczytać z wykresu skrzynkowego? Na którym lotnisku temperatura maksymalna była największa? Na którym lotnisku temperatura minimalna była najmniejsza? Gdzie średnia i mediana z temperatur jest największa? Chodzi nam o rozwiązanie, gdzie na jednym rysunku mamy naraz 3 wykresy skrzynkowe
- G. Przedstaw za pomocą punktów, na wykresie prostokątnym (w zwykłym układzie kartezjańskim) punkty oznaczające długość (longitude) i szerokość geograficzną (latitude) lotnisk. Gdzie są główne skupienia lotnisk? Czy umiesz rozszyfrować, gdzie są te miejsca? Wersją prymitywną (ale powinna zadziałać na laboratorium) jest otworenie Google Maps i wpisanie współrzędnych, np. 60.40338, -150.17403 (najpierw latitude, potem longitude). Liczba cyfr musi się zgadzać, separatory dziesiętne to kropki. Wersją bardziej R-ową, ale wymagającą zainstalowania dużego pakietu (i na laboratorium zapewne zakończy się zawieszeniem komputera), jest:

```
install.packages("rworldmap")  
library(rworldmap)
```

```
newmap <- getMap(resolution = "low")  
plot(newmap, xlim = c(-176, 174), ylim = c(19, 72), asp = 1) #należy dobrać xlim ylim  
points(airports$lon, airports$lat, col = "red", cex = .6)
```

4.2 Przy użyciu zbior danych Movie Lens przedstawionego przy poprzedniej liście zadań, należy zwizualizować dane wypisane w poniższych podpunktach. Pamiętaj, że w "prawdziwym życiu" często te wykresy będziesz prezentować komuś innemu - zadбай o odpowiednie tytuły wykresów, nazwy osi i inne rzeczy poprawiające czytelność i zrozumienie wykresu.

- A. ile filmów powstało w poszczególnych latach (w dwóch wersjach)
- B. liczba kobiet vs liczba mężczyzn
- C. częstość gatunków filmowych
- D. rozkład ocen kobiet i mężczyzn
- E. rozkład ocen dla poszczególnych gatunków
- F. rozkład liczby oddanych głosów na filmy

4.3 Korzystając ze znanego już zbioru Movie Lens oraz podanego niżej kodu, który zmienia datę w formacie timestamp na bardziej przyjazną człowiekowi:

```
oceny$rok <- as.POSIXlt(oceny$time, origin="1970-01-01")$year + 1900
oceny$mies<- as.POSIXlt(oceny$time, origin="1970-01-01")$mon + 1
oceny$data <- paste(oceny$rok, "-", sprintf("%02d", oceny$mies),
sep="")
```

Dzięki temu w zbiorze z ocenami otrzymamy rok oraz miesiąc wystawienia oceny. Na tej podstawie należy zagregować oceny w ramach miesiąca uśredniając je. Następnie dla wybranego przez siebie filmu (np Toy Story z id==1) narysować wykres średniej oceny filmu w miesiącu w zależności od czasu.

Wskazówki:

4.1

- A. Wykres słupkowy
- B. Wykres słupkowy
- C. Wykres liniowy
- D. Wykres liniowy
- E. Wykres liniowy
- F. Wykres skrzynkowy z wieloma grupami
- G. Wykres punktowy

4.2

- A. histogram oraz wykres liniowy (oś x - kolejne lata, oś y - liczba filmów)
- B. wykres słupkowy - dwa słupki odpowiadające liczbie użytkowników zadanej płci
- C. wykres słupkowy
- D. wykres skrzynkowy z dwiema grupami

- E. wykres skrzynkowy z wieloma grupami
- F. histogram liczby ocen - sprawdzamy ile ocen było dla każdego filmu i mając te wartości rysujemy histogram