

Uniwersytet Ekonomiczny w Poznaniu
Instytut Informatyki i Ekonomii Ilościowej
Katedra Statystyki

Przetwarzanie danych w R

Pakiet dplyr

Materiały dydaktyczne
dr hab. Marcin Szymkowiak, prof. UEP

Spis treści

1.	Zadania	2
----	-------------------	---

1. Zadania

Zadanie 1. Zaimportuj plik gospodarstwa.xlsx. Zapisz go pod nazwą „dane”. Następnie wszystkie zadania rozwiąż korzystając z pakietu dplyr i porównaj z rozwiązaniami jakie można uzyskać bazując na podstawowych funkcjach języka R.

Zadanie 2. Utwórz zbiór, w którym są gospodarstwa ze wsi.

Zadanie 3. Utwórz zbiór, w którym są gospodarstwa o dochodach > 2000 zł.

Zadanie 4. Utwórz zbiór, który zawiera gospodarstwa z województwa wielkopolskiego oraz które zamieszkują wieś i mają dochody powyżej 3000 zł.

Zadanie 5. Wybierz gospodarstwa domowe z województwa dolnośląskiego i mazowieckiego z miast powyżej 500 tys. mieszkańców.

Zadanie 6. Losowo wybierz zbiór 30% gospodarstw domowych.

Zadanie 7. Losowo wybierz 100 gospodarstw.

Zadanie 8. Wybierz gospodarstwa domowe z wierszy 10–15.

Zadanie 9. Wybierz gospodarstwa domowe z danymi tylko dla kolumny woj i wydg ale w odniesieniu jedynie dla gospodarstw z województwa wielkopolskiego.

Zadanie 10. Wybierz wszystkie gospodarstwa, których dochód jest z przedziału 3000–4000 i zostaw tylko tę zmienną w zbiorze.

Zadanie 11. Wybierz wszystkie kolumny od klm do zut włącznie.

Zadanie 12. Wyświetl nazwy wszystkich zmiennych w zbiorze gospodarstwa.

Zadanie 13. Wybierz wszystkie kolumny, które zaczynają się na literę d.

Zadanie 14. Utwórz nową zmienną roznica=dochg-wydg. Pozostaw w zbiorze zmienne dochg, wydg i roznica.

Zadanie 15. Utwórz nowe zmienne $x=\ln(\text{dochg})$ oraz $y=\ln(\text{wydg})$. Pamiętaj, że \ln liczymy dla dodatnich wartości.

Zadanie 16. Dokonaj zamiany nazwy zmiennych dochg na dochod oraz wydg na wydatki.

Zadanie 17. Oblicz ile było gospodarstw domowych ze względu na poszczególne warianty zmiennej klm.

Zadanie 18. Oblicz ile było gospodarstw domowych w poszczególnych województwach.

Zadanie 19. Oblicz ile było gospodarstw domowych w poszczególnych województwach ze względu na warianty zmiennej d61 (Sytuacja materialna gospodarstwa).

Zadanie 20. Oblicz średnie wydatki wszystkich gospodarstw domowych.

Zadanie 21. Oblicz średnie wydatki i dochody wszystkich gospodarstw domowych.

Zadanie 22. Oblicz średnią, min, max, odchylenie standardowe oraz medianę wydatków.

Zadanie 23. Oblicz średnie wydatki w ramach klasy miejscowości.

Zadanie 24. Stwórz wykres słupkowy z liczebnościami dla klasy miejscowości.

Zadanie 25. Posortuj rosnąco zbiór gospodarstw domowych ze względu na wydatki.

Zadanie 26. Posortuj malejąco zbiór gospodarstw domowych ze względu na dochody.

Zadanie 27. Posortuj malejąco kod klasy miejscowości a w ramach tej klasy rosnąco wydatki.

Zadanie 28. Zrób wykres pudełkowy powierzchni mieszkań (zmienna d36) według klasy miejscowości.

Zadanie 29. Zrób wykres pudełkowy powierzchni mieszkań według źródła utrzymania dla gospodarstw z województwa mazowieckiego.

Zadanie 30. Utwórz zbiór wlkp. W zbiorze tym powinny się znaleźć gospodarstwa z województwa wielkopolskiego, których dochód jest z przedziału 2500–3000 zł. Dochód powinien być posortowany malejąco. Oprócz dochodu mają się znaleźć 4 inne zmienne: klm, woj, zut i wydg. Na tym zbiorze, dla zmiennych dochg i wydg, stwórz korelacyjny diagram rozrzutu.

Zadanie 31. Utwórz dwie nowe zmienne $roznica = dochg - wydg$ oraz $doch_os = dochg / los$ i pozostaw je jako jedyne w zbiorze danych.

Zadanie 32. Utwórz nową zmienną „klasa”, która przyjmuje dwie wartości: „wieś”, gdy gospodarstwo jest ze wsi oraz „miasto”, gdy gospodarstwo jest z miasta.

Zadanie 33. Utwórz nową zmienną „pow”, która przyjmuje trzy warianty: „do 60”, „(60-100>” oraz „powyżej 100” w zależności od wartości zmiennej d36 (powierzchnia mieszkań).