

Warsztaty *R*STYLES

RS1 Przekształcanie danych w R i proste indeksy stylów odpowiedzi.

(Poziom podstawowy)

(Warsztaty realizowane w ramach projektu "Style odpowiedzi w danych samoopisowych: konsekwencje, sposoby zapobiegania, źródła" (2019/33/B/HS6 /00937) finansowanego przez Narodowe Centrum Nauki - NCN)

Planowany termin:

11-12 lutego 2021, godz. 10:00-14:00.

Cele:

- Zdobyć umiejętności sprawnego przekształcania danych w środowisku R z wykorzystaniem pakietów należących do tzw. [tidyverse](#).
- Zastosowanie zdobytych umiejętności do obliczenia prostych wskaźników opisujących style odpowiedzi respondentów w badaniach sondażowych, na przykładzie danych z badania PISA 2012.

Prowadzący:

- dr Tomasz Żółtak
- dr hab. Artur Pokropek, prof. IFiS PAN
- dr Marek Muszyński

Do uczestnictwa będą konieczne:

- dane z wynikami badania PISA 2012;
 - będziemy dystrybuować;
- znajomość podstaw R:
 - znajomość podstawowych struktur danych: wektorów (w tym *czynników* - ang. *factors*) i *ramek danych*;
 - umiejętność instalowania korzystania z dodatkowych pakietów (polecenie ``library()``);
- zainstalowane oprogramowanie:
 - R - do pobrania ze [strony](#);
 - RStudio - do pobrania ze [strony](#);
 - po zainstalowaniu proszę sprawdzić opcję kodowania zapisu polskich znaków zgodnie z instrukcją zamieszczoną na: [tej stronie](#);
 - dodatkowe pakiety R: *haven*, *dplyr*, *tidyr*:
``install.packages(c("haven", "dplyr", "tidyr", "forcats"))``

Ramowy plan:

Czasy podane przy poszczególnych zagadnieniach są orientacyjne.

11 lutego (czwartek)

1. Wprowadzenie - 5 min.
2. *tidyverse* jako dialekt R - 15 min.
3. Wczytywanie danych z plików SPSS i Staty przy pomocy pakietu *haven*; ograniczenia R w reprezentacji etykiet zmiennych i etykiet wartości - 15 min.
4. Zapis zbiorów do plików SPSS, Staty i R - 10 min.
5. Zapoznanie się ze strukturą i zawartością *ramki danych* - funkcje ``head()``, ``tail()``, ``names()``, ``str()``, ``summary()``, ``View()`` - 15 min.
6. Wybieranie wierszy *ramki danych* - funkcja ``filter()``, operatory logiczne i operator ``%in%``; konieczność przypisywania wyników przekształceń do obiektów - 15 min.
7. Sortowanie wierszy *ramki danych* - funkcja ``arrange()`` - 5 min.
8. Wybór wierszy *ramki danych* przy pomocy funkcji ``slice()`` i jej wariantów - 10 min.

Przerwa - 20 min.

9. Wybieranie kolumn *ramki danych* - funkcja ``select()`` i różne warianty składni, które można w niej stosować; zmiana nazw zmiennych: funkcja ``rename()`` - 25 min.
10. Przekształcanie danych z wykorzystaniem funkcji ``mutate()`` cz. I:
 - a. Podstawowa składnia; tworzenie nowych zmiennych i modyfikowanie istniejących; kolejność wykonywania się poleceń - 25 min.
 - b. Funkcje ``if_else()`` i ``case_when()`` - 20 min.
 - c. Dokonywanie tego samego przekształcenia na wielu zmiennych - składnia z wykorzystaniem funkcji ``across()`` - 20 min.

12 lutego (piątek)

11. Prezentacja: *Wprowadzenie do stylów odpowiedzi: źródła i konsekwencje* (+dyskusja) - 20 min. + 5 min.
12. Przekształcanie danych z wykorzystaniem funkcji ``mutate()`` cz. II:
 - a. Przekształcenia wykorzystujące wiele zmiennych jednocześnie - funkcje ``rowwise()`` i ``c_across()`` - 20 min.
 - b. Zamiana wartości na braki danych i braków danych na wartości - funkcje ``is.na()``, ``replace_na()`` i ``na_if()`` - 10 min.
13. Agregacja danych, cz. I:
 - a. Tworzenie zestawień zagregowanych parametrów przy pomocy funkcji ``summarise()`` - 20 min.
 - b. Agregacja w ramach grup - funkcja ``group_by()`` - 15 min.

Przerwa - 20 min.

14. Agregacja danych, cz. II:
 - a. Obliczanie parametrów w ramach grup bez zmiany struktury zbioru - użycie ``group_by()`` w połączeniu z ``mutate()`` - 25 min.

15. Łączenie ze sobą zbiorów danych:

- a. Łączenie zbiorów przy pomocy funkcji ``left_join()``, ``right_join()``, ``inner_join()`` i ``full_join()`` - 20 min.
- b. Problemy ze zduplikowanymi obserwacjami i usuwanie duplikatów - funkcje ``duplicated()`` i ``distinct()`` - 15 min.
- c. Filtrowanie wierszy przy pomocy funkcji ``semi_join()`` i ``anti_join()`` - 10 min.
- d. Funkcje ``bind_rows()`` i ``bind_cols()`` - 15 min.

16. Podsumowanie - 5 min.