

<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

Ściągamy bazę danych **UScereal.csv** (65 rzędów i 11 kolumn).

Baza zawiera dane dotyczące płatków śniadaniowych sprzedawanych w USA.

Opis kolumn znajduje się na poniższej stronie:

<https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/UScereal.html>

Zadanie jest wykonane poprawnie, kiedy jest opatrzone stosownym komentarzem, a kod działa w sposób poprawny.

max = 37 pkt.; dst.=22 pkt.

ZADANIA

1. Zmień nazwę kolumny "mfr" na "producent"; zmień nazwę "shelf" na "polka". (2 pkt.)
2. Dodaj do zbioru danych kolumnę "totalcarb", która będzie sumą węglowodanów "carbo" i cukrów "sugars". (2 pkt.)
3. Ile płatków w zbiorze to płatki z 3 półki? (np. subset) (3 pkt.)
4. Stwórz zbiór danych, w którym znajdą się tylko płatki "K" (Kellogg's). (3 pkt.)
5. Utwórz tabele częstości kaloryczności dla przedziałów (0-100], (100-200], (200-max) (np. cut). Który przedział jest najczęstszy? (4 pkt.)
6. Ile procent bazy składa się z płatków o najwyższej kaloryczności (powyżej 200)? (4 pkt.)
7. Utwórz osobny zbiór danych, w którym znajdują się wszystkie płatki, które mają mniej niż 150 kcal i więcej niż 9 jednostek błonnika. (4 pkt.)
8. Narysuj histogram dla zmiennej "calories", dodaj nazwy osi i tytuł, pokoloruj go (4 pkt.)
9. Narysuj wykres pudełkowy kaloryczności w zależności od półki na której stoją płatki. Na której półce stoją płatki najbardziej kaloryczne? (3 pkt.)
10. Czy w zbiorze znajdują się jakieś obserwacje odstające? (np. boxplot) (2 pkt.)
11. Wypisz i zinterpretuj najważniejsze statystyki opisowe dla zmiennej "calories". (4 pkt.)
12. Czy rozkład danych jest symetryczny? dlaczego tak? dlaczego nie? (2 pkt.)