

# Ćwiczenia - eksploracja

*Marcin K. Dyderski*

*17 kwietnia 2018*

## Eksploracja danych - zadania

1. Na stronie <https://github.com/mkdyderski/BSS> znajduje się plik 'sosny.csv'. Wczytaj zbiór danych i obejrzyj jakie są w nim zmienne i jakie mają rozkłady. Sprawdź poprawność typów danych, jeśli trzeba - popraw ręcznie w pliku źródłowym. Objaśnienie nazw:

Plot.ID - nazwa poletka type - leśne, porolne, przemysłowe N - szerokość geograficzna E - długość soil.type - typ gleby Age - wiek drzewostanu dens - zagęszczenie [szt./ha] G - pole powierzchni przekroju pierśnicowego [m<sup>2</sup>/ha] Hg - wysokość górna (średnia ważona g drzew) [m] V - miąższosć (zasobność, objętość drzew) [m<sup>3</sup>/ha] AB - biomasa części nadziemnej [t/ha] BR - biomasa gałęzi [t/ha] ST - biomasa pni [t/ha] FL - biomasa liści [t/ha]

Zbiór danych pochodzi z supplementary material z pracy Jagodziński et al. 2018 <https://www.sciencedirect.com/science/article/pii/S0378112717318261>

2. W zbiorze danych 'sosna' sprawdź które wartości komponentów biomasy (AB, FL, ST, BR) są odstające od trendu związanego z miąższoscią (V) i wysokością (Hg)
3. Wykonaj histogramy dla cech drzewostanów sosnowych
4. Sprawdź rozkłady zmiennych w czterech grupach gleb
5. Za pomocą boxplotów oceń w jakim przedziale wieku (używając funkcji cut() potnij wiek na cztery klasy), zagęszczenia i G jest największe FL i AB
6. Podaj średnie wartości kilku wybranych cech dla klas wieku i wykonaj wykres słupkowy (średnie + SE)