

Eksploracji danych

Lista 4

1. Dla danych `iris{datasets}` pogrupuj obiekty metodą k -średnich z $k = 3$. Porównaj z przypisanymi gatunkami, zrób macierz błędów (tablicę pomyłek) i wykres grup.
2. Dla danych `diamonds{ggplot2}[R]` wybierz losowo 1000 wierszy (dane mają 53940 wierszy) i znajdź optymalne k dla metody grupowania k -średnich wybierając jako atrybuty zmienne numeryczne, pogrupuj obiekty metodą k -średnich i zrób wykres grup.
3. Dla danych `votes.repub{cluster}[R]` (dane opisują ile procent danego stanu USA głosowało na republikanów w wybranych latach) znajdź optymalne k dla metody grupowania k -średnich, pogrupuj obiekty metodą k -średnich i zrób wykres grup. Wybierz kilka lat jako cechy objaśniające.
4. Dla danych `pluton{cluster}` (dane opisują skład izotopowy 45 partii plutonu) znajdź optymalne k dla metody grupowania k -średnich, pogrupuj obiekty metodą k -średnich i zrób wykres grup.