

Dzień 2 - Eksploracja - zadania

Marcin K. Dyderski, Patryk Czortek

2 kwietnia 2019

Zadania do wykonania

1. Z GitHuba pobierz dataset z cechami roślin, link: [https://github.com/mkdyderski/BSS/blob/BSS2019/datasety/vege_1517_traits.csv]. Możesz również ściągnąć go do R za pomocą funkcji `read.csv()`:

```
baza<-read.csv('https://raw.githubusercontent.com/mkdyderski/BSS/BSS2019/datasety/vege_1517_traits.csv'
               sep=';')
```

2. Sprawdź które kolumny zawierają zmienne liczbowe, a które tekstowe
3. Wykonaj histogram dla `seed_mass`
4. Wykonaj tabelę dwudzielną `growth_from` oraz strategię - czy widać jakieś zależności?
5. Wykonaj boxploty dla `seed_mass` w ramach strategii
6. Podaj średnie wartości kilku wybranych cech dla grup hg (historyczno-geograficznych) i wykonaj wykres słupkowy (średnie + SE)

Propozycje do pracy z własnym zbiorem danych

7. Wczytaj *własny zbiór danych* i poznaj jego strukturę i zakresy zmiennych - jakie rozkłady mają poszczególne zmienne? Czy są obserwacje odstające? Eksploracja danych pozwala wykryć wartości nielogiczne z biologicznego punktu widzenia biologicznego i naprawić je przed właściwymi analizami.
8. Wykonaj wykres punktowy pokazujący relacje pomiędzy cechami dla których zakładasz występowanie pewnych zależności, dodaj linię trendu i oceń czy jest w miarę sensownie dopasowana. W przypadku wątpliwości poproś prowadzących o odpowiedź odnośnie typu linii trendu. Możesz zestawzić np. sukces reprodukcyjny z cechami środowiska czy występowanie gatunków (0-1) z dostępnością zasobów.
9. Sprawdź czy zmienne liczbowe różnią się pomiędzy grupami (bez testów, na razie tylko wizualnie). Możesz np. sprawdzić bogactwo gatunkowe w różnych wariantach, wartości odbicia widma dla różnych gatunków, masę czy wielkość prób dla różnych terminów lub indeks heterotermii dla różnych osobników.
10. Sprawdź, czy proste przeliczenie pozwoli dać Ci kolejną cechę do analiz. Być może wystarczy podzielić masę przez objętość by zyskać gęstość? Albo określić udział gildii/grup funkcjonalnych organizmów? Spróbuj wykorzystać do tego pakiet `dplyr`.
11. Przygotuj zestawienie liczebności prób w różnych wariantach za pomocą wykresów słupkowych - takie aby móc łatwo opowiedzieć o układzie badań.