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 wykres rozrzutu dla `canopy_height` i `SLA`, `SLA` i `seed_mass` oraz `seed_mass` i `canopy_height`. Dodaj linie trendu używając funkcji `geom_smooth(method='lm')` - sprawdź jak zmieni się wynik po dodaniu skal logarytmicznych.
5. Wykonaj boxploty dla `seed_mass` w ramach `strategy`. Dodaj skalę barwną z palety ColorBrewer używając `scale_fill_brewer()`
6. Podaj srednie wartosci kilku wybranych cech dla grup hg (historyczno-geograficznych) i wykonaj wykres słupkowy (srednie + SE)
7. Narysuj wykres na którym pokażesz zależnośc pomiędzy `SLA` i `canopy_height` a wielkość punktów (`aes(...size=...)`) zależeć będzie od `seed_mass`. Dopasuj skale i linię trendu.

Zadbaj o estetykę wszystkich wykresów - zmień tło, opisy osi i elementy graficzne. Skorzystaj z linków do dodatkowych materiałów. Wyobraź sobie, że robisz to do publikacji za duży impakt, którą recenzować będzie pedantyczny specjalista.

Propozycje do pracy z własnym zbiorem danych

8. 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.
9. 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 wątpliwosci poproś prowadzących o podpowiedź odnośnie typu linii trendu. Możesz zestawić np. sukces reprodukcyjny z cechami środowiska czy występowanie gatunków (0-1) z dostępnością zasobów. Dla zmiennych binarnych zamiast `+geom_smooth(method='lm')` użyj `+geom_smooth(method='glm',method.args=list(family='binomial'))`. Dla zmiennych o rozkładzie Poissona (np. liczba gatunków, liczba piskląt) użyj `+geom_smooth(method='glm',method.args=list(family`
10. Sprawdź czy zmienne liczbowe różnią sie 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.
11. 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.
12. Przygotuj zestawienie liczebności prób w różnych wariantach za pomocą wykresów słupkowych - takie aby móc łatwo opowiedzieć o układzie badań.