

Techniki wizualizacji danych - praca domowa 8

Autor: Jan Skwarek

Celem pracy domowej jest przygotowanie wstępnej analizy eksploracyjnej danych dla danego zbioru. Ja przeprowadzę analizę zbioru danych 'mushrooms.csv'.

Wstęp

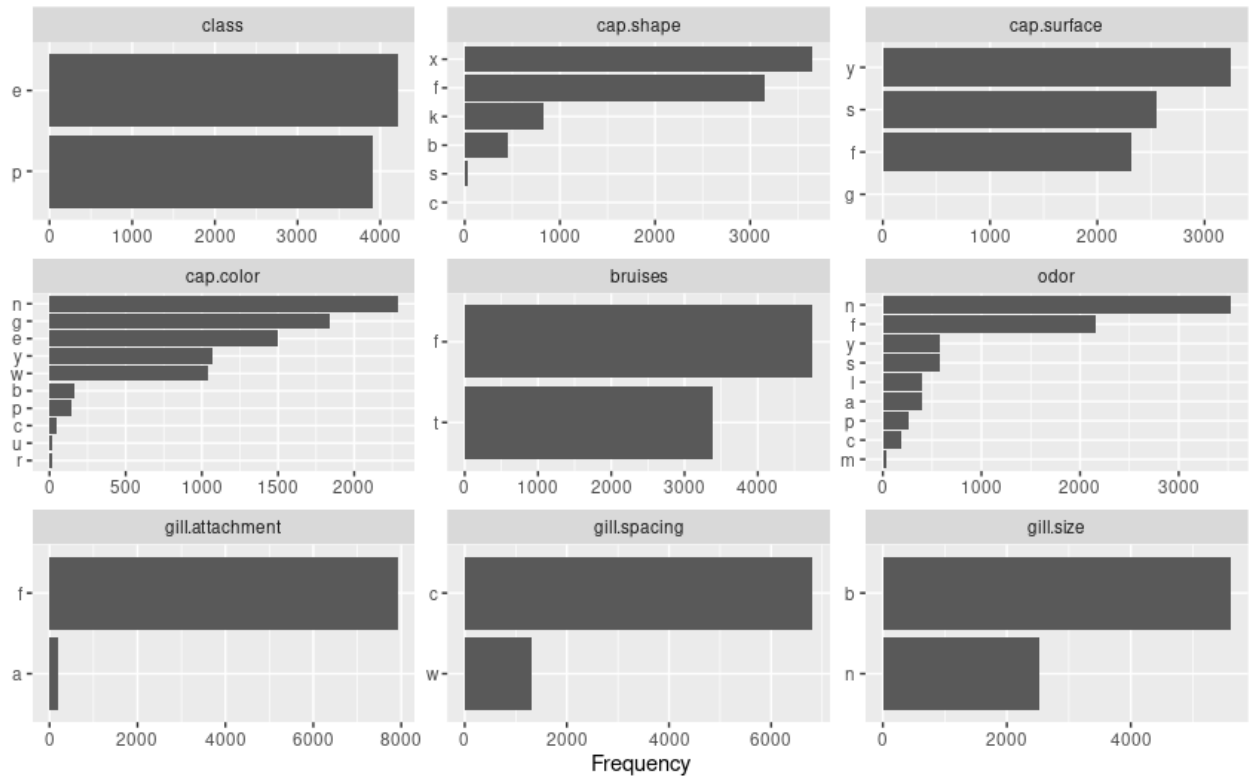
Zestaw 'mushrooms.csv' zawiera opisy hipotetycznych próbek odpowiadającym 23 gatunkom grzybów z rodzin Agaricus i Lepiota (grzyby pieczarkowate), zaczerpnięte z „The Audubon Society Field Guide to North American Mushrooms” (1981). Każdy gatunek jest identyfikowany jako zdecydowanie jadalny, zdecydowanie trujący lub o nieznanym jadalności (niepolecany). Ta ostatnia klasa została połączona z trującą. Zbiór danych zawiera następujące informacje:

- 1) klasa (jadalne bądź trujące)
- 2) kształt, powierzchnia (jej chropowatość), kolor kapelusza
- 3) obecność plamek na grzybie, tzw. „siniaków”
- 4) zapach
- 5) przywiązanie, odległości między nimi, wielkość, kolor blaszek
- 6) kształt, korzeń, powierzchnia nad pierścieniem, powierzchnia pod pierścieniem, kolor nad pierścieniem, kolor pod pierścieniem na trzonie
- 7) typ i kolor zasłonki
- 8) liczba, typ pierścienia
- 9) kolor zarodników
- 10) populacja
- 11) środowisko, w którym występują

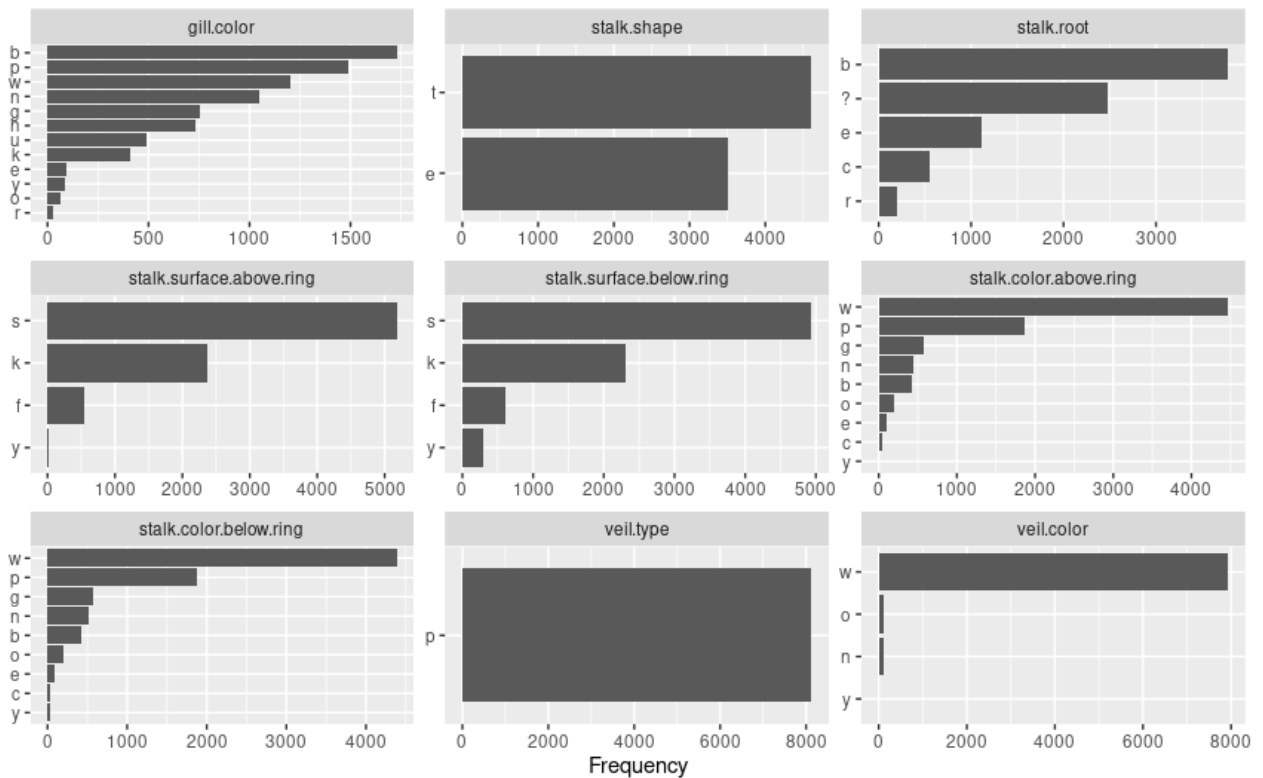
Spróbuję na podstawie wszystkich tych danych znaleźć zależności między poszczególnymi cechami wyglądu grzybów, a ich klasą (jadalnością). Takie badania mogą okazać się niezwykle pomocne dla wszystkich fanów zbierania grzybów.

Analiza danych

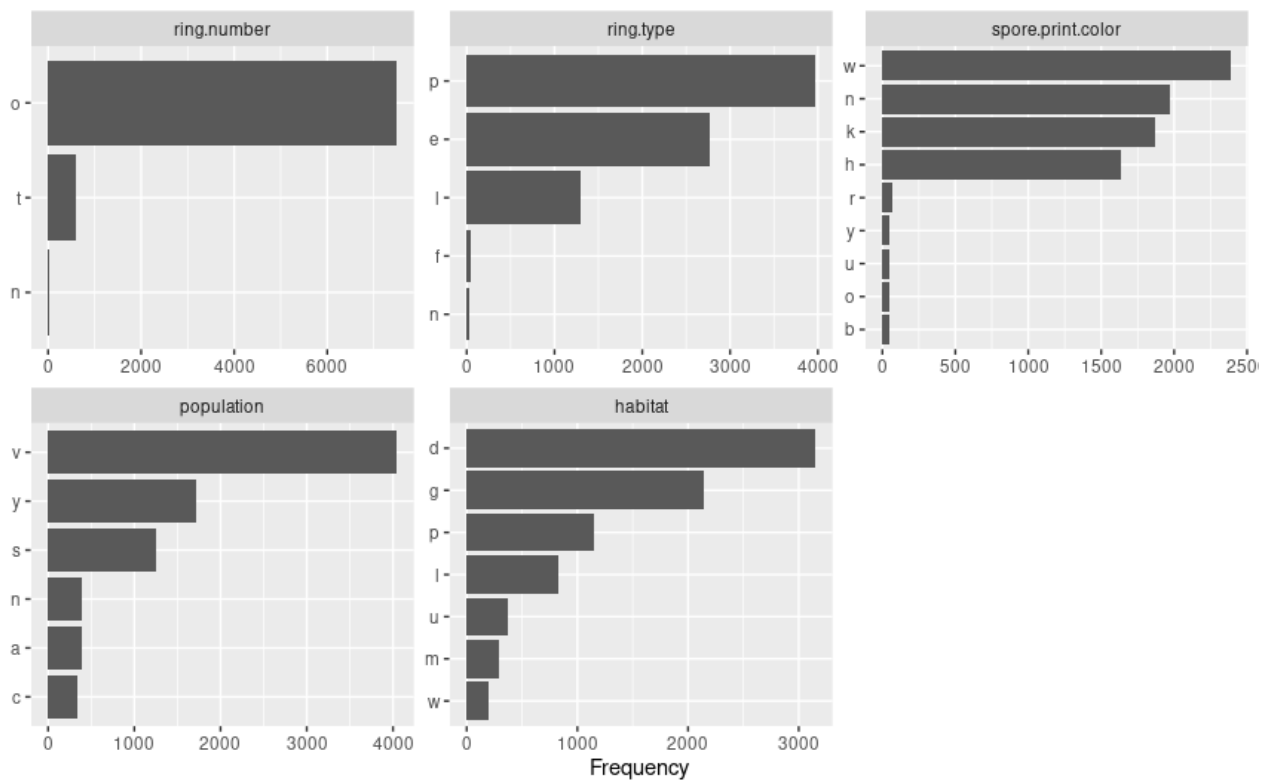
Spójrzmy najpierw na wykresy słupkowe wszystkich wartości:



Page 1



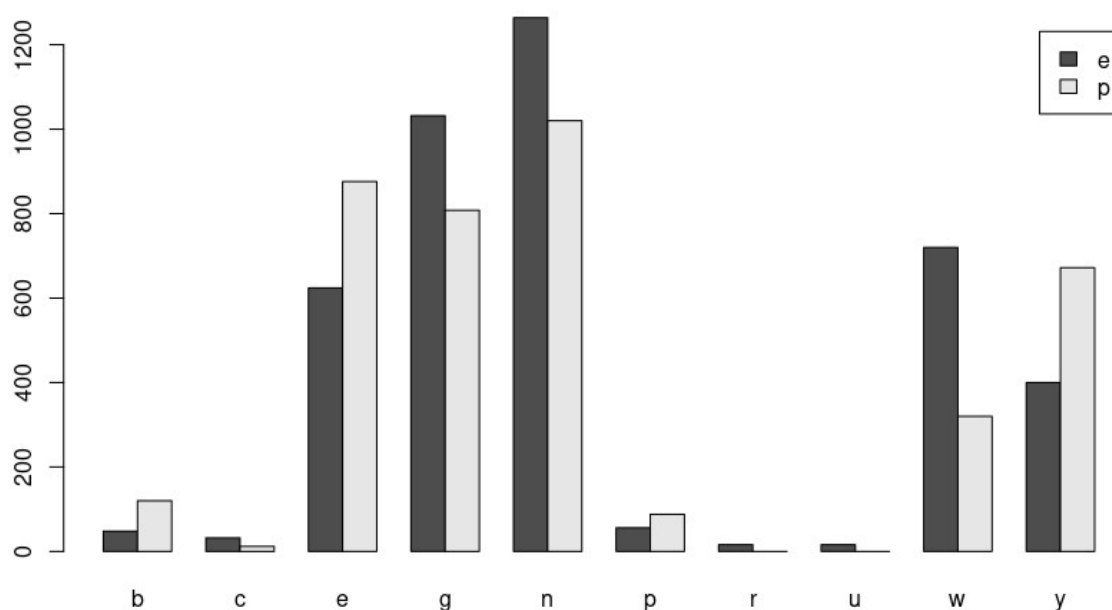
Page 2



Tak o to wygląda rozkład dla każdej z cech opisujących nasz zbiór danych. Poszukajmy teraz pewnych własności grzybów mogących pomóc nam określić ich jadalność.

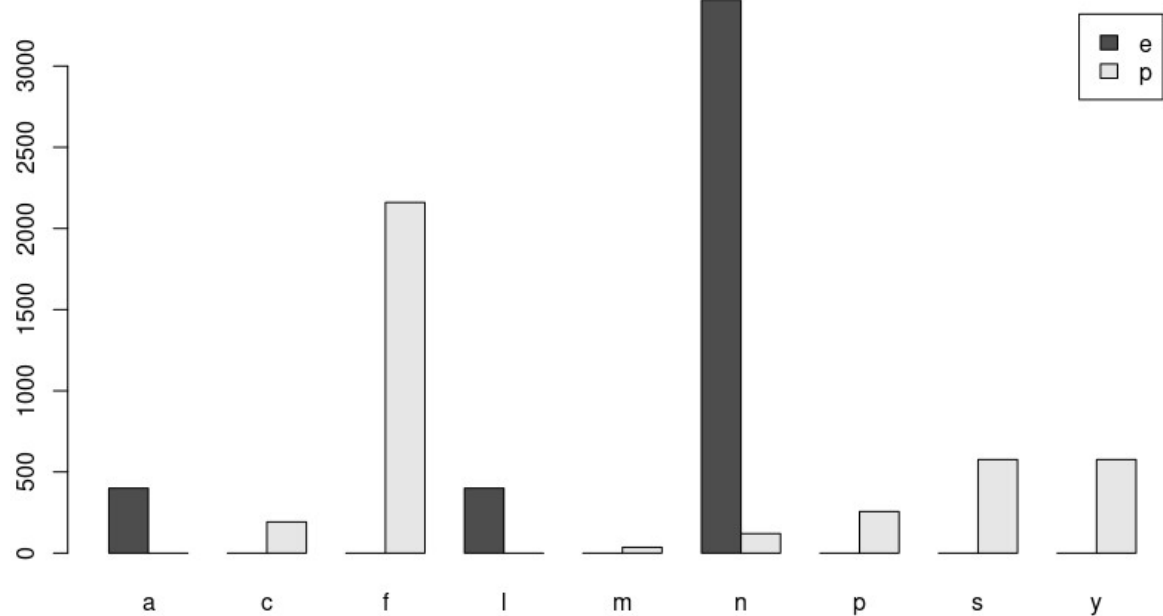
Pytanie 1: jakie są własności wspólne kapeluszy wszystkich grzybów jadalnych?

Na przedstawionym wykresie widać, że zdecydowana większość grzybów z białym kapeluszem jest jadalna. Może to się wydawać zaskakujące jeżeli nie pamiętamy, że badanie dotyczy jedynie grzybów z rodziny pieczarkowatych. Natomiast większość grzybów z brązowym kapeluszem jest trująca (ponownie znów nas zawiodła intuicja opierająca się na doświadczeniu w zbieraniu grzybów z pozostałych rodzin). Warto też podkreślić, że na przedstawionym wykresie nie ma pewnej jednej zasadniczej reguły. Nie jesteśmy w stanie podnieść grzyba i po kolorze jego kapelusza



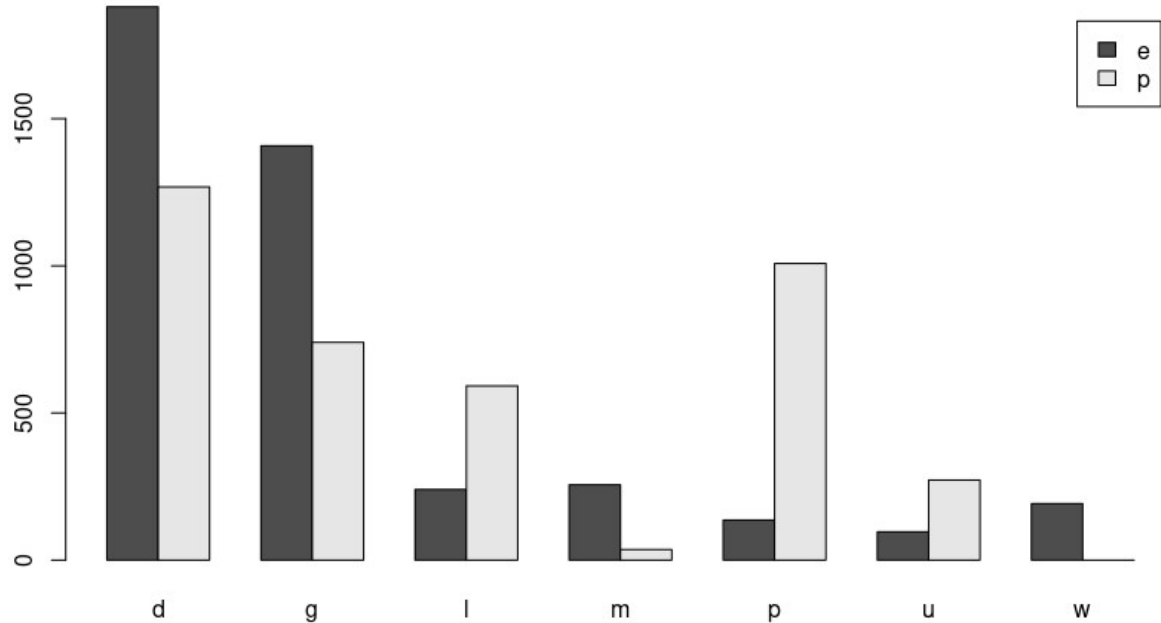
jednoznacznie (lub z dużym prawdopodobieństwem przekraczającym 90%) powiedzieć, że dany grzyb jest jadalny bądź trujący.

Pytanie 2: jak pachną grzyby jadalne?



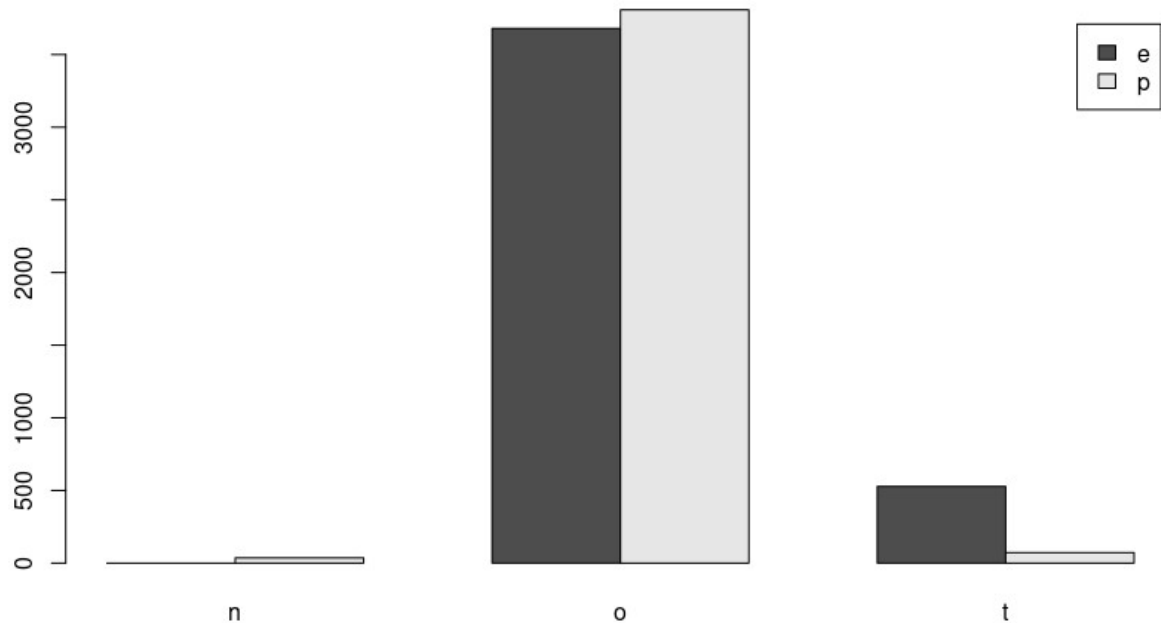
Zapach grzybów zdaje się być bardziej znaczący w ocenie jadalności danego grzyba. Możemy zaobserwować tutaj konkretne wzorce. Wszystkie grzyby o zapachu migdałowym, anyżowym oraz prawie wszystkie grzyby bez zapachu zdają się być jadalne. Grzyby o wszystkich pozostałych zapachach są natomiast w 100% trujące.

Pytanie 3: gdzie najczęściej występują grzyby jadalne?



Tutaj również możemy zaobserwować pewne ważne zależności. Wszystkie grzyby wyrzucone, ale zebrane okazywały się jadalne. Może to oznaczać to, że grzybiarze zbierający pieczarki, po zebraniu jadalnego gatunku, odkrywali że ten jest robaczywy, a następnie go wyrzucali. Większość grzybów na łąkach jest jadalna. Gdzie natomiast rosną grzyby trujące? Głównie przy drogach, na terenach zabudowanych oraz w liściach (aczkolwiek w przypadku liści jest to mniej więcej 70% - wcale nie tak dużo).

Pytanie 4: czy istnieje zależność między liczbą pierścieni a jadalnością grzyba?



Ciekawe wydaje się to, że w zbiorze danych wystąpiła praktycznie taka sama liczba grzybów jadalnych i trujących z jednym pierścieniem. Jest to zdecydowanie cecha wyglądu, którą nie mogą sugerować się grzybiarze. Co do innych ilości pierścieni – wszystkie grzyby bez pierścieni okazały się trujące, natomiast zdecydowana większość grzybów z dwoma pierścieniami to grzyby jadalne.

Podsumowanie

Ze zbioru danych 'mushrooms.csv' możemy dowiedzieć się wielu ciekawych rzeczy o grzybach. Mogą one zdecydowanie pomóc w określeniu tego, czy dany grzyb jest jadalny, czy może trujący. Analizę danych utrudniło trochę to, że nie ma ani jednej danej liczbowej w analizowanym zbiorze. Mimo to, analiza pozwoliła nam wyciągnąć następujące wnioski:

1. Grzyby z zielonym i fioletowym kapeluszem są trujące.
2. Większość (ale nie większość zdecydowana) grzybów z brązowym kapeluszem to grzyby trujące.
3. Większość (ale nie większość zdecydowana) grzybów z białym kapeluszem to grzyby jadalne.
4. Grzyby o zapachu migdałowym i anyżowym to grzyby jadalne.
5. Prawie wszystkie grzyby bez zapachu są jadalne.
6. Grzyby o zapachu innym niż wyżej wymienione to grzyby trujące.
7. Większość (ale nie większość zdecydowana) grzybów rosnących na łąkach to grzyby jadalne.
8. Większość grzybów rosnąca przy drogach i w terenach zabudowanych to grzyby trujące.
9. Grzyby bez pierścieni to grzyby trujące.
10. Większość grzybów z dwoma pierścieniami to grzyby jadalne.