

Praca domowa 6

Jakub Seliga

Opis pakietu

Narzędziem, jakie wziąłem w tej pracy domowej na tapet jest pakiet **hexbin** w R-ze (posiadający też analogiczny odpowiednik w Pythonie). Pozwala on tworzyć dwuwymiarowe heatmapy, na których zaznaczona jest gęstość występowania rekordów o danych wartościach. Na osiach znajdują się wartości dwóch badanych zmiennych, a właściwy wykres stanowią sześciokąty, pokolorowane zgodnie z ilością wystąpień rekordów o wartościach zmiennych w danym zakresie.

Wykresy utworzone za pomocą tego pakietu mogą stanowić czytelniejszą alternatywę głównie dla wykresów punktowych. Używane są przede wszystkim do zilustrowania zależności dwóch zmiennych oraz ilości rekordów o danych badanych parametrach.

W skład pakietu wchodzi tylko jedna funkcja **hexbin**, która tworzy obiekt, który można wywołać następnie jako argument funkcji **plot**. Funkcja ta przyjmuje za argumenty:

- **x**, **y** - dwa argumenty, wektory podające współrzędne dwuwymiarowych punktów danych do wykresu jako opcjonalne argumenty:

1. **xbins** - liczba sześciokątów na każdej z osi
2. **shape** - wartość współczynnika wysokość wykresu/szerokość wykresu
3. **xbnds**, **ybnds** - limity odpowiednio osi **x** i **y**
4. **xlab**, **ylab** - podpisy osi **x** i **y**
5. **IDs** - logiczne wskazanie, czy poszczególne „identyfikatory” komórek powinny zostać zwrócone

Domyślnie, powstała heatmapa tworzy się w kolorach czarno-białych, gdzie czarny oznacza większą ilość występowania. Tak jak w przykładzie poniżej, skalę kolorów można zmienić, np. za pomocą wspomnianego na zajęciach pakietu **RColorBrewer**.

Kod i wizualizacja

Dla przykładowej wizualizacji z użyciem tego pakietu wziąłem dane ze znanej z zajęć ramki `auta2012`. Aby wykres był czytelniejszy, brane pod uwagę są tylko auta o roku produkcji po II wojnie światowej.

```
library(PogromcyDanych)
library(hexbin)
library(RColorBrewer)

auta2012 <- auta2012 %>% filter(Rok.produkcji>1945)

hex <- hexbin(auta2012$Rok.produkcji, log(auta2012$Cena.w.PLN),
  xbins = 60,
  shape = 0.7,
  xlab = 'Rok produkcji',
  ylab = 'Logarytm z ceny w złotych')

my_colors = colorRampPalette(rev(brewer.pal(8, 'YlOrRd'))))

plot(hex, main="Auta2012: rok produkcji vs cena na skali logarytmicznej",
  colramp=my_colors,
  legend=F)
```

Auta2012: rok produkcji vs cena na skali logarytmicznej

