

Laboratorium 3 – Grafika w R

1. Wiadomości ogólne

R rozmieszcza automatycznie rysunki na obszarze strony (chyba format A4 dyfulth, możliwy do zmiany funkcją `par`). Na zewnątrz całej strony są marginesy zewnętrzne (górny, dolny, lewy, prawy). Wewnątrz marginesów zewnętrznych obszar rysunku. Wewnątrz obszaru rysunku obszar kreślenia.

Jeżeli na jednej stronie jest więcej rysunków, marginesy zewnętrzne są zawsze wspólne. Wewnątrz nich znajdują się oddzielne obszary rysunków, w których znajdują się indywidualne obszary kreślenia.

Zasadnicza część rysunku (np. wykresy) są wykonywane w obszarze kreślenia, osie i etykiety na marginesach.

Funkcja `windows()` daje podgląd okna graficznego.

2. Funkcje wysokiego poziomu

`plot(x, y)` rysuje wykres punktowy dla wektorów danych `x`, `y`. Jest funkcja "wirtualną" i jej działanie zależy od rodzaju danych.

Uwaga parametr `add=TRUE/FALSE` nie jest traktowany jako legalny w funkcji `plot`??

Nie umiem nałożyć wykresów.

3. Funkcje niskiego poziomu

Funkcje rysujące różne proste obiekty graficzne, np. `abline(a, b)` rysuje prostą $y=ax+b$.

Parametry graficzne: Ustawiane są głównie funkcją `par()`. Samo `par()` listuje aktualne ustawienia parametrów.

`par(mfrow=c(2, 3))` dzieli okno graficzne na 2 wiersze i 3 kolumny. Rysunki są wykonywane kolejno wierszami.

4. Proste prezentacje danych jednowymiarowych

Wykres słupkowy wykonuje funkcja `barplot(<dane>, horizon=TRUE/FALSE, las=1)` `Horizon` – określa kierunek rysowania słupków, `las` – kierunek opisu osi. W przykładzie zastosowano parametry `col` i `add=TRUE` dla narysowania dwóch wykresów słupkowych na tej samej osi. Słupki są rysowane dla każdego rodzaju (wartości) danych dyskretnych.

Czasem bezpieczniej jest przeczytać dane do tabeli i dalej traktować ją jako dane do wykresu.

Słupki są rysowane dla każdej wartości, przez co nadaje się ona najlepiej do prezentacji cech dyskretnych jakościowych.

Uwaga, funkcja `table(<wektor>)` zlicza liczbę wystąpień identycznych wartości współrzędnych i przygotowuje tablice do wykresu słupkowego

Wykres słupkowy - histogram

`histogram(<dane>, break, probability=FALSE, plot=TRUE)` rysuje ilości (częstości a nie częstości) występowania wartości ze zbioru skalarnych danych w

postaci wykresu słupkowego. Przedział w którym występują wszystkie dane jest dzielony na break przedziałów.

Inn a możliwością jest funkcja `hist()`. Przykład

```
windows()
hist(daneSoc$wiek, breaks = 15, main="Histogram wieku",
las=1, ylab="Liczba", xlab="Wiek")
```

`main`, `ylab`, `xlab` zawierają stringi opisu rysunku i osi odpowiednio.

Kolejny przykład poprawia wygląd

```
hist(daneSoc$wiek, breaks = 15, col="grey", border="white",
las=1, probability = TRUE, ylab="Czestosc", xlab="Wiek")
```

Oprócz tego użycie parametru `probability=TRUE` spowodowało przeliczenie ilości wystąpień na częstotliwości.

Wykres pudełkowy (boxplot) rysuje wartości statystyk dla wielu obserwacji. Standardowo, dla każdej grupy obserwacji prezentowany jest: zakres sensownych wartości, mediana oraz pierwszy i trzeci kwartył. Pomiędzy kwartynami rozpięte jest pudełko. Szerokość pudełka może być proporcjonalna do ilości danych. Pudełko „mieści” 50% danych.

Przykład:

```
boxplot(daneSoc$cisnienie.rozk, daneSoc$cisnienie.skur,
horizontal = TRUE, names = c("Skurczowe", "Rozkurczowe"))
```

Rysowane są pudełka dla dwóch zbiorów danych (1 i 2 parametr), Pudełka są rysowane poziomo (`horizontal=TRUE`) i nazwy zbiorów danych.

Pozostałe ważne parametry, to `range`, określająca szerokość przedziału, poza którym obserwacje są traktowane jako artefakty, `varwidth=TRUE/FALSE` określający, czy szerokość pudełek ma zależeć od ilości danych.

```
boxplot(wiek~wyksztalcenie, data = daneSoc, varwidth=TRUE,
col="lightgrey", ylab="Wiek", las=1)
```

Tutaj pierwszym argumentem jest formuła `wiek~wyksztalcenie`

Czasami, dla obserwacji granicznych (`qmin`) i maksymalnej (`qmax`) przyjętych jako sensowne należy przyjąć: `qmin = dół_pudełka - 1.2 długość_pudełka`, podobnie `qmax = góra_pudełka + 1.5 długość_pudełka`.

Wykres skrzypcowy (violinplot) (bardzo popularny!!) jest rozszerzeniem wykresu pudełkowego. Rysuje go funkcja `vioplot()` z pakietu o tej samej nazwie.

Przykład:

```
library("vioplot")
vioplot(wiek~wyksztalcenie, data = daneSoc)
```

rysuje wykresy wiolinowe dla tych samych danych co ostatni boxplot. Na osi wartości danych jest pokazany rozkład częstości. Wewnątrz narysowane jest pudełko i mediana.

Wykres jest atrakcyjny dla przypadku danych wielomodalnych, które koncentrują się wokół wielu wartości i mediana nie musi być najbardziej reprezentatywna.

Diagram lub wielobok liczebności jest aproksymacją funkcji gęstości próby. Można go wykreślić używając `plot(density(<dane>))`. Aproksymacja jest metodą „jądrową”, czyli używającą pewnej postaci rozkładu. Realizuje ją funkcja `density()`. Funkcja `ecdf()` estymuje dystrybuantę.

Przykład:

```
plot(density(daneSoc$wiek, bw=1.5), main="Rozklad
wieku", type="h")
```

Parametr `bw=1.5` zmienia tzw. „okno” interpolacji, które ma wpływ na stopień „wygładzenia” interpolacji wykresu. Parametr `n` to ilość punktów po której ma być zrobiona interpolacja (siatka). Parametr `kernel` daje możliwość wyboru typu kernela.

```
plot(ecdf(daneSoc$wiek), main="Dystrybuanta wieku")
```

Wykres kropkowy (scatterplot) pokazuje zależność pomiędzy dwiema zmiennymi losowymi. Do jego wykreślenia służy funkcja `scatterplot()` z biblioteki `car`. Są problemy z dostępnością funkcji `scatterplot` pomimo załadowania `carData`.

```
library("car") # wymaga pakietu carData są problemy??  
sp(cisnienie.rozkurczowe~cisnienie.skurczowe, data=daneSoc,  
smooth=FALSE, reg.line=FALSE, pch=19)
```

Diagram dla danych 2D.

Zadania

Interpretacja graficzna danych

Zadanie 1. Zrób wykres pudełkowy dla `acme$market` i `acme$acme`

Zadanie 2. Zrób wykres wiolinowy dla `acme$market` i `acme$acme`

Zadanie 3. Zrób wykres mozaikowy (punktowy) przy pomocy funkcji `plot` oraz funkcji `scatterplot` dla `acme$market` i `acme$acme`

Zadanie 4. Zrób we wspólnym układzie współrzędnych wykresy gęstości rozkładów: normalnego $N(0,3)$, $F(3,6)$ oraz chi-kwadrat dla $n = 3$ w przedziale $[0,10]$.