

Wykresy w środowisku R - pakiet ggplot2

Agnieszka Goroncy



UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU

Wydział Matematyki i Informatyki

Pakiet ggplot2 (Grammar of Graphics)

Grammar of graphics

Łatwiejsze wyobrażenie, rozumienie, komunikacja grafiki. W praktyce oznacza to budowanie niezależnych "klocków" i łączenie ich ze sobą w celu otrzymania dowolnej grafiki. Wykresy składają się z warstw, które nakłada się na obiekt bazowy.

Zalety ggplot2:

pozwała na otrzymanie efektownej i skomplikowanej grafiki, wykresy są łatwiejsze do wygenerowania, bardziej elastyczne, pakiet jest bardzo popularny (dużo użytkowników: fora, listy mailingowe itp.), możliwość wykorzystania gotowych "tematów".

Twórca pakietu: Hadley Wickham.

Instalacja i załadowanie do pamięci R:

```
> install.packages("ggplot2")  
> library(ggplot2)
```

Dokumentacja pakietu

Szybki wykres: `qplot()`

Dokumentacja

- odpowiednik `plot()`
- wygodny do tworzenia różnego rodzaju wykresów za pomocą jednej funkcji

Dane: **w formacie ramki danych** (`data.frame`)

Parametry:

```
qplot(x,y,data, color, shape, size, fill, alpha,  
geom, xlim, ylim, xlab, ylab, main, sub, ...)
```

qplot()

- `x`, `y` - zmienne umieszczone na osiach OX, OY; w wykresach jednej zmiennej można pominąć `y` (np. przy histogramach);
- `data` - nazwa ramki danych;
- `color`, `shape`, `size`, `fill` - pozwala przypisać wartości zmiennych kolorom, kształtom, wielkościom symboli na wykresie. Wypełnienie (`fill`) dotyczy wykresów gęstości oraz skrzynkowych;
- `alpha` - przejrzystość elementów, które się nakładają (wartość między 0 - przezroczysty a 1 - nieprzezroczysty);
- `geom` - definiuje obiekt geometryczny który określa typ wykresu (możliwe wartości: "point", "smooth", "boxplot", "line", "histogram", "density", "bar", "jitter");
- `main`, `sub` - definiuje tytuł i podtytuł wykresu;
- `xlab`, `ylob` - definiuje etykiety osi OX i OY;
- `xlim`, `ylim` - dwuwymiarowe wektory liczbowe definiujące minimalne i maksymalne wartości na osiach OX i OY;

qplot()

Przykłady:

```
> qplot(mpg, data=mtcars)
> qplot(mpg, drat, data=mtcars)
> qplot(mpg, drat, data=mtcars, size=gear)
> mtcars$gear <- factor(mtcars$gear, levels=c(3,4,5),
+ labels=c("3-biegowa", "4-biegowa", "5-biegowa"))
> qplot(mpg, data=mtcars, geom="density", fill=gear,
+ alpha=I(.5), main="Rozkład zużycia paliwa",
+ xlab="Liczba mil na galon", ylab="gęstość")
> qplot(gear, mpg, data=mtcars, geom=c("boxplot", "jitter"),
+ fill=gear, main="Zużycie paliwa w podziale na liczbę biegów",
+ xlab="", ylab="Liczba mil na galon")
```

Estetyka wykresów: kodowanie za pomocą aes()

Funkcja aes()...

... umożliwia doprecyzowanie estetyki wykresów, opisuje jak poszczególne zmienne mają odpowiadać poszczególnym estetycznym własnościom wykresów (np. wiek samochodu jest różnicowany na wykresie kolorem, natomiast model kształtem itp.).

```
aes(x, y, size, color, shape, group, linetype, slope,  
intercept,... )
```

ggplot()

Funkcja ggplot()...

... pozwala tworzyć obiekty bazowe, oraz nakładać na nie poszczególne warstwy z kolejnymi obiektami geometrycznymi (*geom_point*, *geom_bar*, *geom_jitter*, *geom_text*, *geom_step*, itp.).

Przykłady:

```
> mtcars$gear <- factor(mtcars$gear, levels=c(3,4,5),  
+ labels=c("3-biegowa", "4-biegowa", "5-biegowa"))  
> mtcars$cyl <- factor(mtcars$cyl, levels=c(4,6,8),  
+ labels=c("4-cyl", "6-cyl", "8-cyl"))  
> ggplot(mtcars, aes(x=cyl, fill=gear))+geom_bar()  
> ggplot(mtcars, aes(x=gear, y =wt, shape=cyl))  
+ +geom_point()
```

Obiekty *geom* - lista

- Odkrywać! Ujawniać! Objaśniać! Gramatyka języka wizualizacji danych, Przemysław Biecek
- Przewodnik po pakiecie R, Jak tworzyć wykresy z pakietem ggplot2?, Przemysław Biecek
- Introduction to R graphics with ggplot2 - Workshop Materials
- A grammar of graphics: past, present, and future, Hadley Wickham
- ggplot2: Hadley Wickham, Winston Chang
- R Graphics Cookbook, Winston Chang, O'Reilly Media, 2012
- Quick-R, Graphics with ggplot2