

Gramatyka grafiki

A. M. Machno

Po co robić wykresy?

Graficzna prezentacja danych ma dwa cele i w każdym z nich mamy całkiem różne wyzwania.

- ▶ **Wykresy odkrywcz**e (*exploratory graphs*): wykresy mające na celu poznanie zbioru danych przez badacza. Na tym etapie bardzo często powstają hipotezy badawcze, które mogą zostać zweryfikowane metodami statystycznymi.
- ▶ **Wykresy objaśniające** (*explanatory graphs*): wykresy mające na celu wizualizację otrzymanych wyników. Wykresy te są elementem komunikacji wyników.

Przykładowy wykres ggplot

```
ggplot(mtcars, aes(x = cyl, y = mpg)) + geom_point()
```



Elementy gramatyki grafiki

- ▶ **Dane:** zbiór danych, który chcemy przedstawić na wykresie.
- ▶ **Estetyka** (*aesthetics*): skale na które mapowane są dane.
- ▶ **Geometrie:** elementy graficzne użyte do przedstawienia danych.
- ▶ **Płaszczyzny** (*facets*): tworzy wielokrotność wykresu.
- ▶ **Statystyki:** reprezentacje danych pomagające w zrozumieniu lub ukazaniu prawidłowości.
- ▶ **Współrzędne:** powierzchnia na której mają być stworzone wykresy.
- ▶ **Motyw:** wszystkie elementy nie związane z danymi.

Estetyki

Przeważnie na jednym wykresie w skład estetyk wchodzi wiele elementów. Najbardziej popularnymi estetykami są:

- ▶ `x, y`: określa, które elementy danych mają konkretne współrzędne. Podstawowe estetyki, występujące prawie w każdym wykresie.
- ▶ `color, fill`: kolory odpowiednio brzegu oraz wnętrza geometrii. - `size`: wielkość.
- ▶ `labels`: określa, wartości tekstowe, które mogą być przekazane do geometrii tekstowych.
- ▶ `alpha`: przezroczystość geometrii. Wartość zero oznacza całkowicie przezroczysty obiekt, wartość jeden całkowicie wypełniony kolorem.
- ▶ `shape`: kształt.

Uwaga: w estetyce każda z powyższych charakterystyk musi być przypisana do elementu danych, wtedy tworzy estetykę.

Przykłady elementów gramatyki grafiki

- ▶ Przykładowymi geometriami są: punkty, linie, histogram, słupki, wykres pudełkowy.
- ▶ Przykładowymi płaszczyznami są: columny i wiersze.
- ▶ Przykładowymi statystykami są: wygładzenie, grupowanie, opisywanie, wnioskowanie.
- ▶ Przykładowymi współrzędnymi są: kartezjański, ustalony, biegunowy.

Pakiet ggplot2

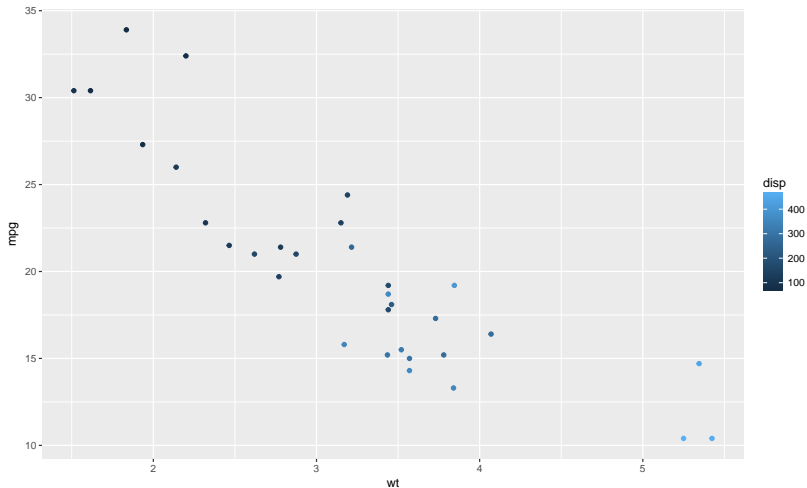
Podstawową funkcją pakietu ggplot2 jest `ggplot()`. Najprostrzym użyciem funkcji jest określenie zbioru danych oraz estetyk. W wyniku otrzymamy obiekt, który graficzny, ale nie posiadający części geometrycznej.

Dodatkowym elementem wymaganym, aby obiekt mógł być wyświetlony jest geometria. Dodanie kolejnego elementu do obiektu graficznego odbywa się poprzez dodanie wartości do istniejącego obiektu graficznego. W ten sposób można dodawać kolejne warstwy do obiektu graficznego. Wywołanie tegoż obiektu skutkuje stworzeniem grafiki.

Uwaga: aby obiekt graficzny mógł zostać wyświetlony wymagane są przynajmniej: **dane**, **estetyki** oraz **geometrie**.

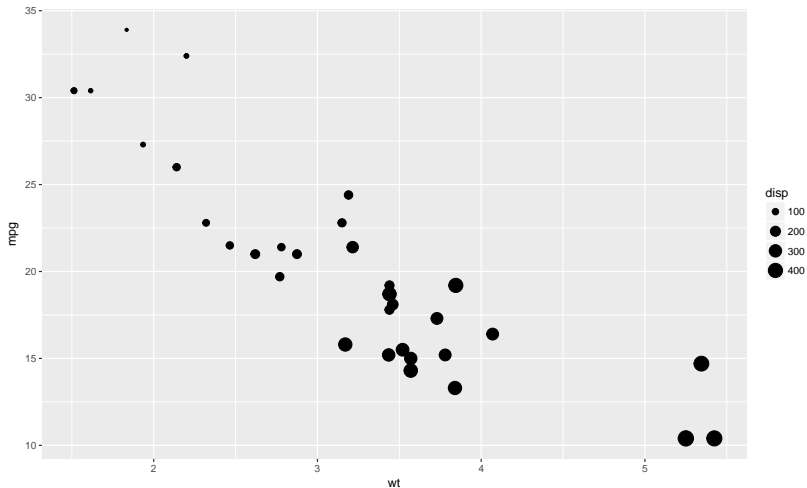
Estetyki x, y i color

```
ggplot(mtcars, aes(x = wt, y = mpg, color = disp)) +  
  geom_point()
```



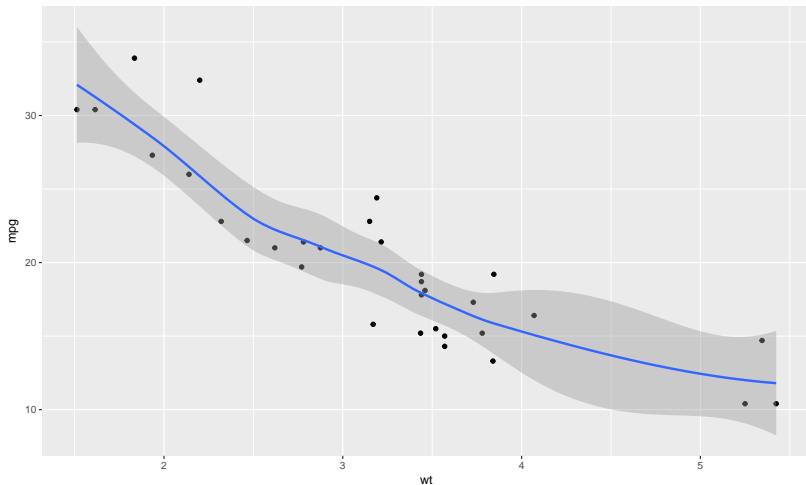
Estetyka size

```
ggplot(mtcars, aes(x = wt, y = mpg, size = disp)) +  
  geom_point()
```



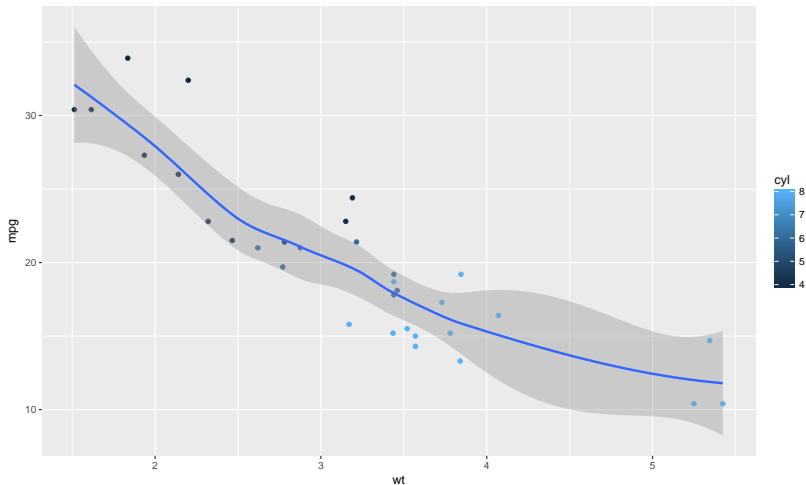
Dodatkowa geometria

```
ggplot(mtcars, aes(x = wt, y = mpg)) +  
  geom_point() + geom_smooth()
```



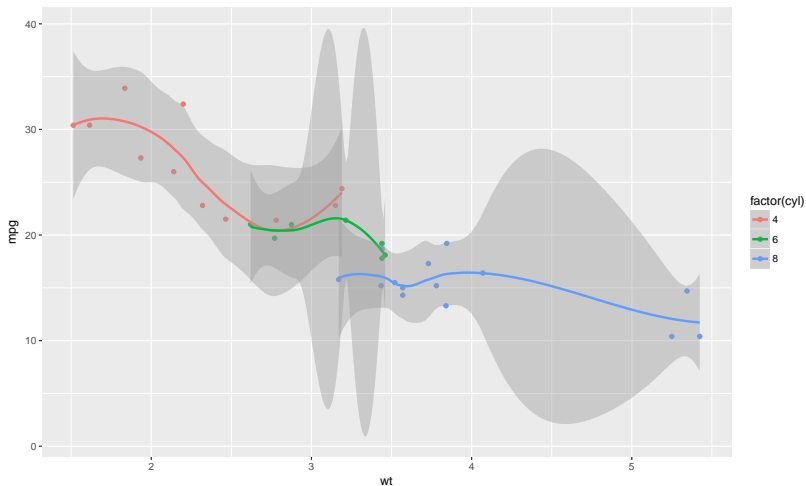
Estetyka color z dodatkową geometrią

```
ggplot(mtcars, aes(x = wt, y = mpg, color = cyl)) +  
  geom_point() + geom_smooth()
```



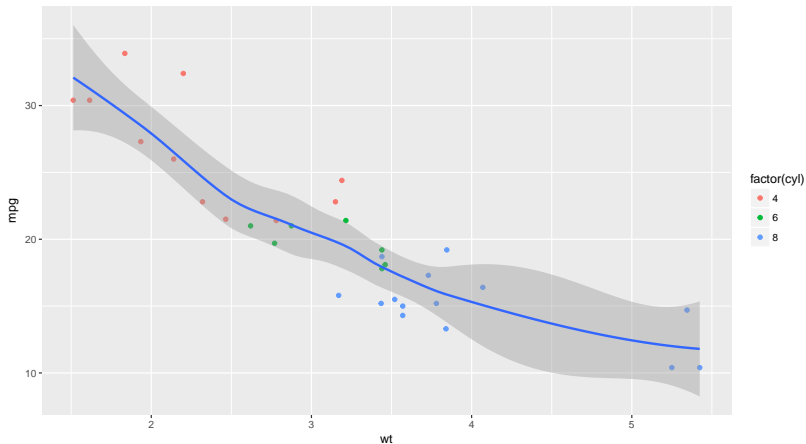
Estetyka color dla kategorii z dodatkową geometrią

```
ggplot(mtcars,
       aes(x = wt, y = mpg, color = factor(cyl))) +
  geom_point() + geom_smooth()
```



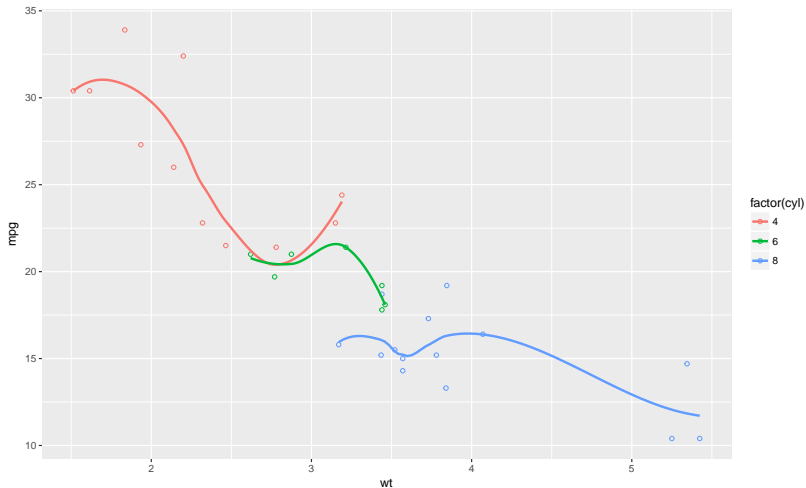
Różne estetyki dla różnych geometrii

```
ggplot(mtcars) +  
  geom_point(aes(x = wt, y = mpg,  
                 color = factor(cyl))) +  
  geom_smooth(aes(x = wt, y = mpg))
```



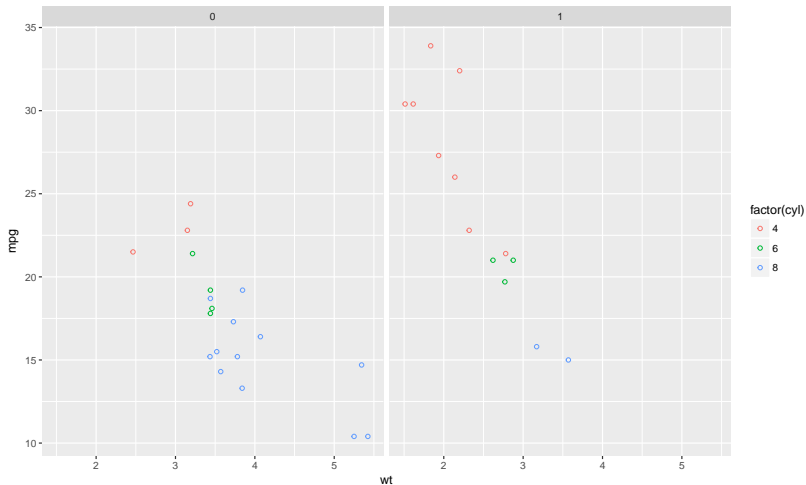
Atrybuty geometrii

```
ggplot(mtcars,  
  aes(x = wt, y = mpg, color = factor(cyl))) +  
  geom_point(shape = 21) + geom_smooth(se = FALSE)
```



Przykłady płaszczyzn

```
ggplot(mtcars,
       aes(x = wt, y = mpg, color = factor(cyl))) +
  geom_point(shape = 21) + facet_grid(.~factor(am))
```



ćwiczenia

1. Dla danych diamonds naszkicować wykres punktowy ceny od wagi.
2. Do wykresu z 1. dodać prostą obrazującą zależność liniową.
3. Dla wykresu z 1. określić kolor punktów w zależności od klarowności diamentu.
4. Do wykresu z 3. dodać wygładzone linie opisujące zależność bez zacieniowanych obszarów obisujących błędy estymacji tych krzywych dla każdej z klarowności. Niech klasa cięcia będzie reprezentowana przez kształt punktu.