

Course: R Language in Computational Probability and Statistics

Lecture 9: Low-level plotting functions

Lecturer: Nataliia Krushlova

Функції низького пріоритету додають об'єкти у вікно виводу, яке вже створено однією з функцій високого пріоритету. У таблиці наведено основні функції низького пріоритету і їх призначення.

Табл.1

Функція	Призначення
abline	Відображення прямих
arrows	Побудова стрілок
box	Відображення рамки навколо графіка
grid	Додавання координатної сітки
legend	Налаштування легенди графіку
lines	Зображення лінії, що з'єднує точки
mtext	Додавання текстової інформації до графіку
points	Зображення точок
polygon	Додавання багатокутника та заливка області
rect	Зображення прямокутника
segments	Точки з'єднуються відрізками

symbols	Побудова графічних примітивів: коло, термометр, квадрат, зірка
text	Відображення тексту на графіку
title	Відображення назви графіку
xspline	Інтерполяція сплайнами

Функція **abline()** має наступні параметри:

a,b – коефіцієнти рівняння $y=a+bx$;

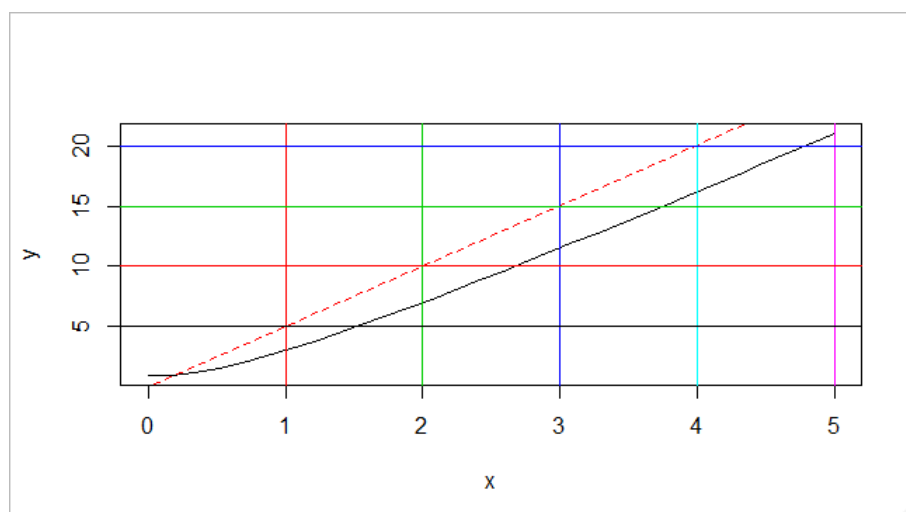
h - горизонтальна пряма $y=h$;

v - вертикальна пряма $x=v$.

Якщо не вказувати імена параметрів, то значення кутового коефіцієнту вказується другим параметром.

Приклад:

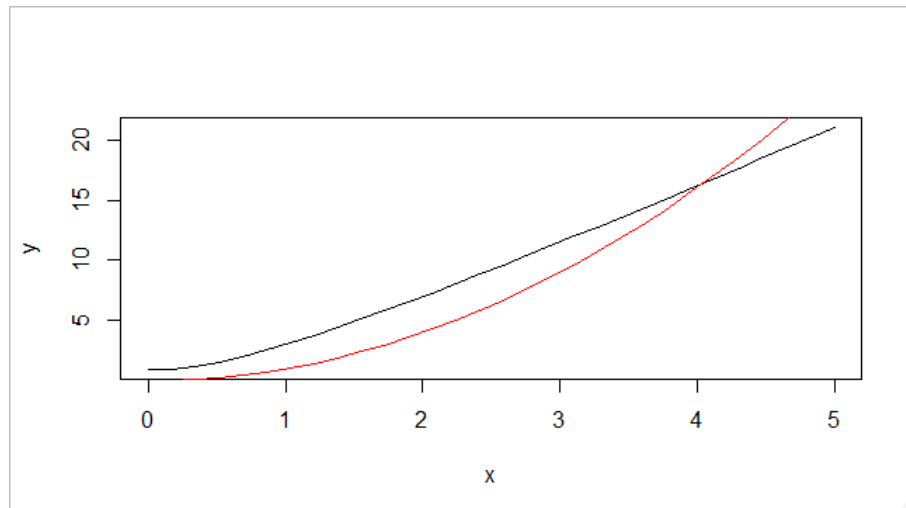
```
> x<-seq(0,5,length.out = 30)
> y<-(5*x^2+1)/(x+1)
> plot(x,y,type="l")
> abline(0,5,col="red",lty=2)
> abline(v=seq(0:5),col=c(10:16))
> abline(h=seq(5,20,by=5),col=c(17:20))
```



Основними параметрами функції **lines(x, y = NULL, type = "l" ...)** є x, y – координати точок, які з'єднує лінія.

Приклад.

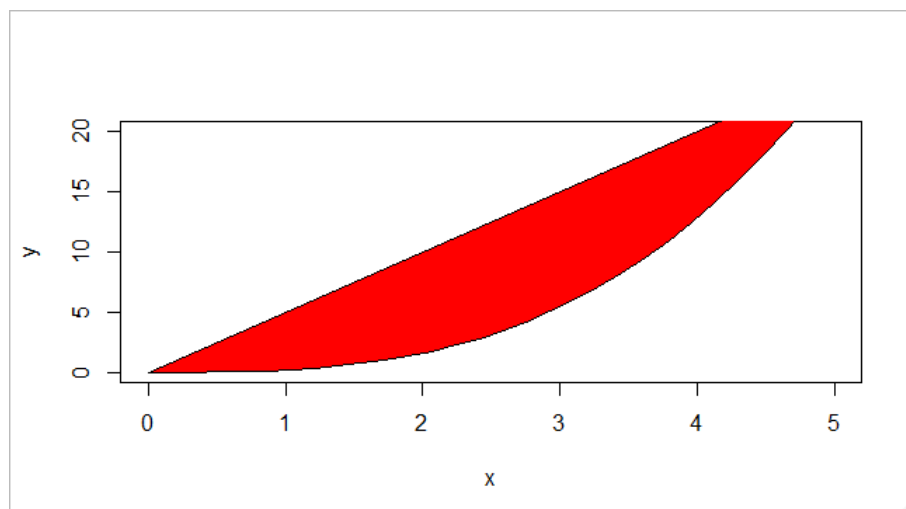
```
> x<-seq(0,5,length.out = 30)
> y<-(5*x^2+1)/(x+1)
> plot(x,y,type="l")
> z<-x^2
> lines(x,z,col="red")
```



Функція **polygon(x, y = NULL,...)** потрібна для заливки області між двома кривими. Параметри **x, y** – координати вершин багатокутника. Послідовність точок розбиття інтервалу по осі абсцис має бути зростаючою.

Приклад.

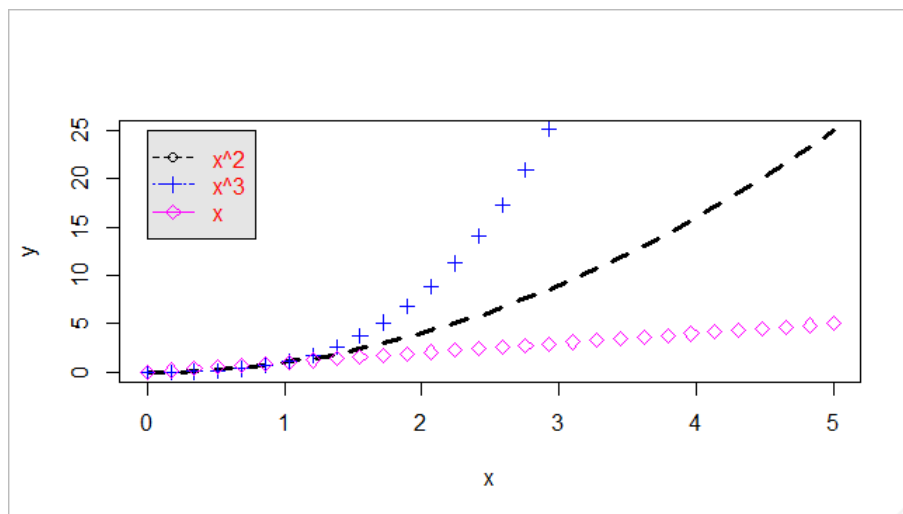
```
> plot(c(0,5),c(0,20),type="n",xlab="x",ylab="y")
> x<-seq(0,5,length.out = 30)
> polygon(x,x^3/5,col="red",border="black",lty="solid")
```



Функція **legend(x, y = NULL,...)** додає легенду до графіку. Параметри **x,y** – координати початку тексту на графіку.

Приклад.

```
> x<-seq(0,5,length.out = 30)
> plot(x,x^2,type="l",lwd=3,xlab="x",ylab="y",pch=1,col=1,lty=2)
> points(x, x^3, pch = 3, col = 4,lty=4)
> lines(x, x, type = "b", lty = 1, pch = 5, col = 6)
> legend(0, 25, c("x^2", "x^3", "x"), col = c(1,4,6), text.col = "red",
lty = c(2, 4, 1), pch = c(1, 3, 5),merge = TRUE, bg = "gray90")
```

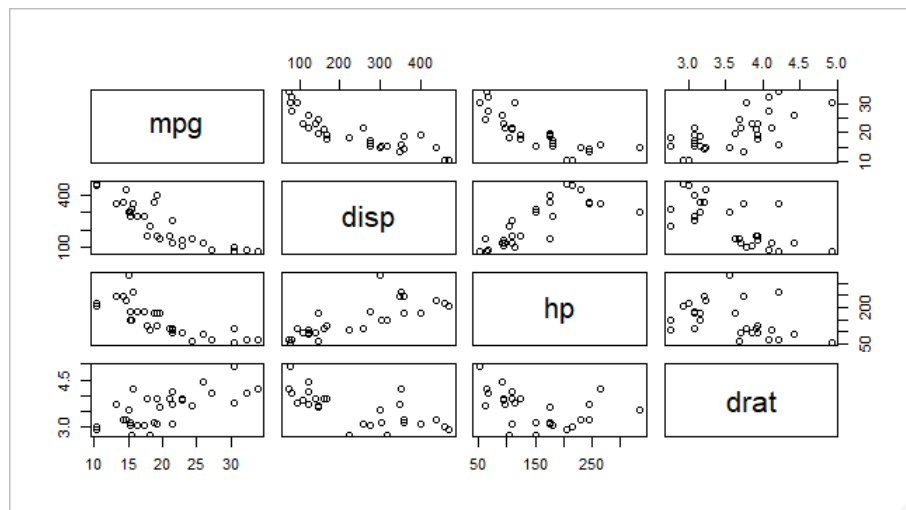


Матриця діаграм розсіювання

Матриця діаграм розсіювання – це діаграми розсіювання для кожної пари аргументів, що входять до набору даних. Створюється функцією **pairs()**.

Приклад.

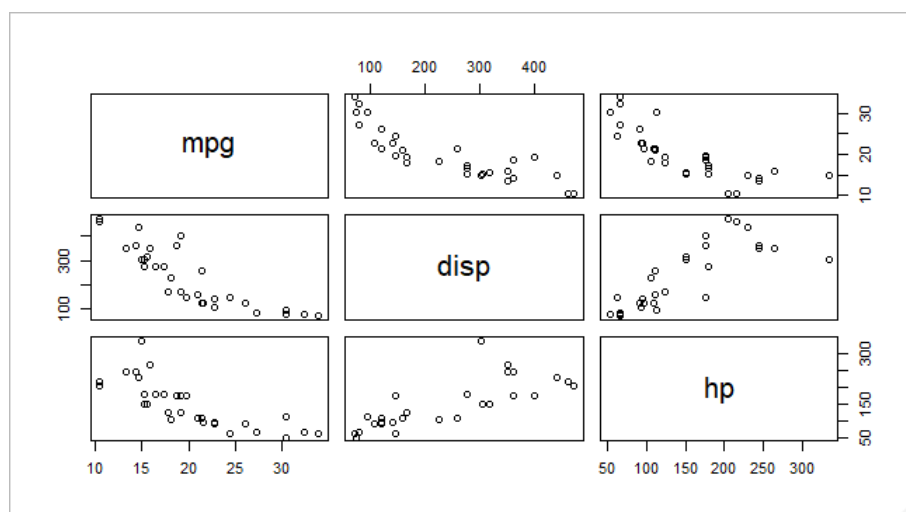
```
> pairs(mtcars[,c(1,3:5)])
```



Матрицю діаграм розсіювання для підмножини даних можна побудувати ще таким чином.

Приклад.

```
> pairs(~mpg+disp+hp,mtcars)
```

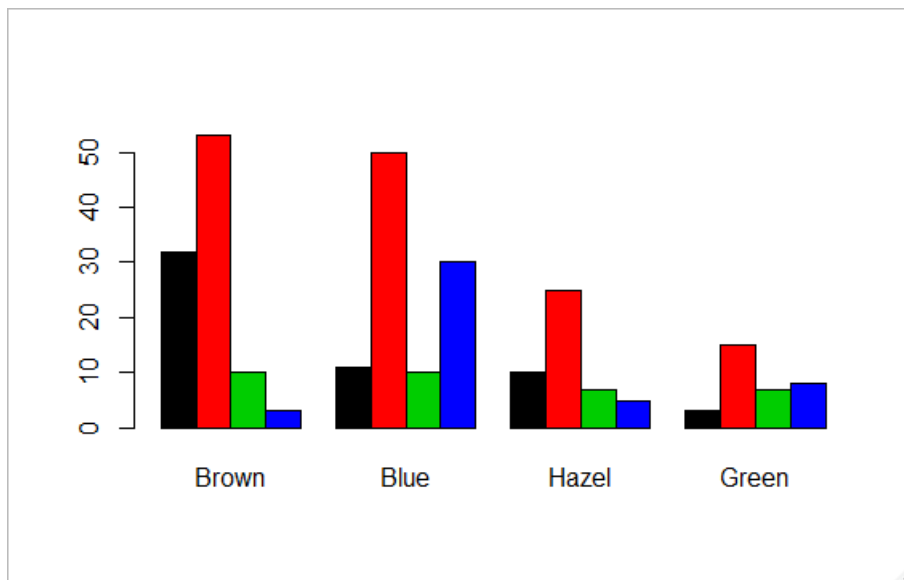


Стовпчикові діаграми

Стовпчикова діаграма – це спосіб зображення числових даних, в якому числу відповідає стовпець. Діаграми створює функція **barplot(height,...)**, де **height** - вектор або матриця чисел.

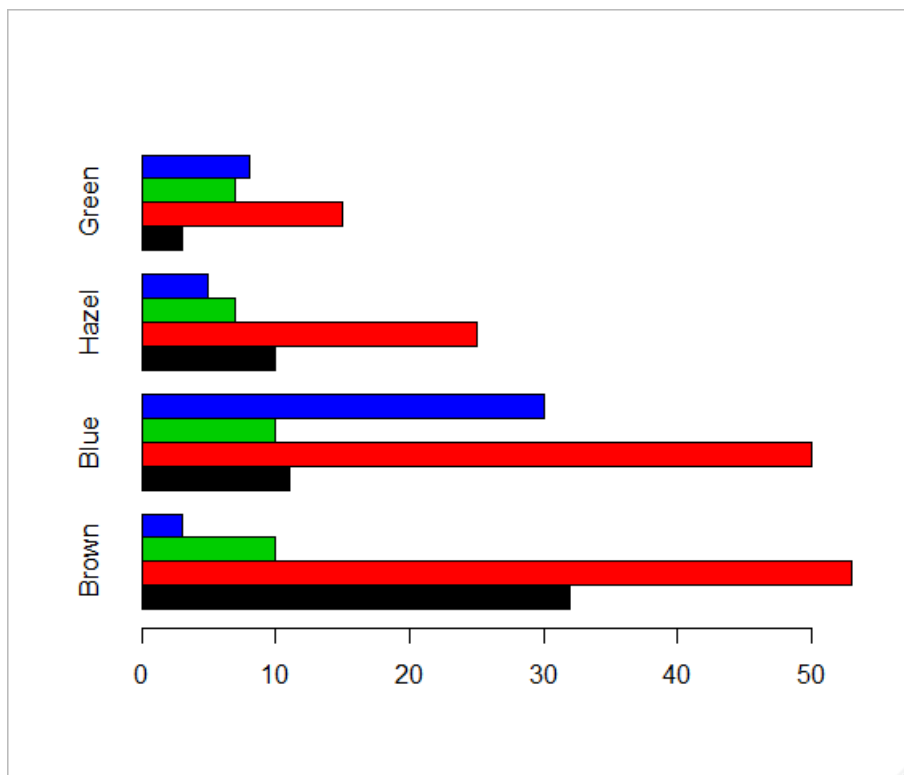
Приклад.

```
> new<-HairEyeColor
> barplot(new[,1],beside = T)
> barplot(new[,1],beside = T,col=c(1:4))
```



Приклад.

```
> barplot(new[,1],beside = T,col=c(1:4),horiz=T)
```

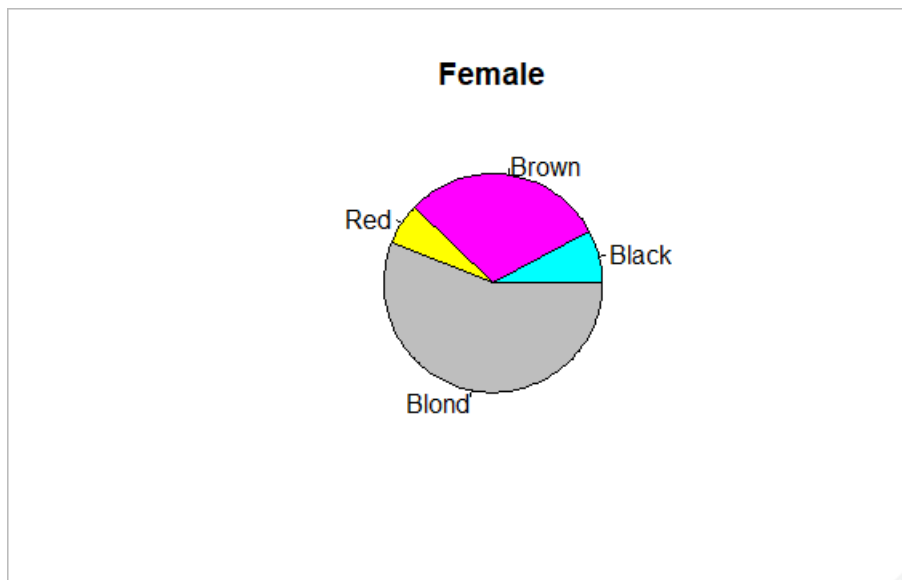


Кругові діаграми

Функція **pie(x, labels)** створює кругову діаграму. Параметри: **x**- числовий вектор; **labels** – символічний вектор, який містить імена секторів.

Приклад:

```
> pie(new[,2],main="Female",col=c(5:8))
```



Пакет ggplot2

Аналітиками даних більше використовується пакет **ggplot2** для візуалізації даних, аніж базова графіка. Розглянемо коротко деякі можливості цього пакету.

Графік в пакеті **ggplot2** представляє собою конструктор, що складається з окремих частин, які можна поєднувати між собою.

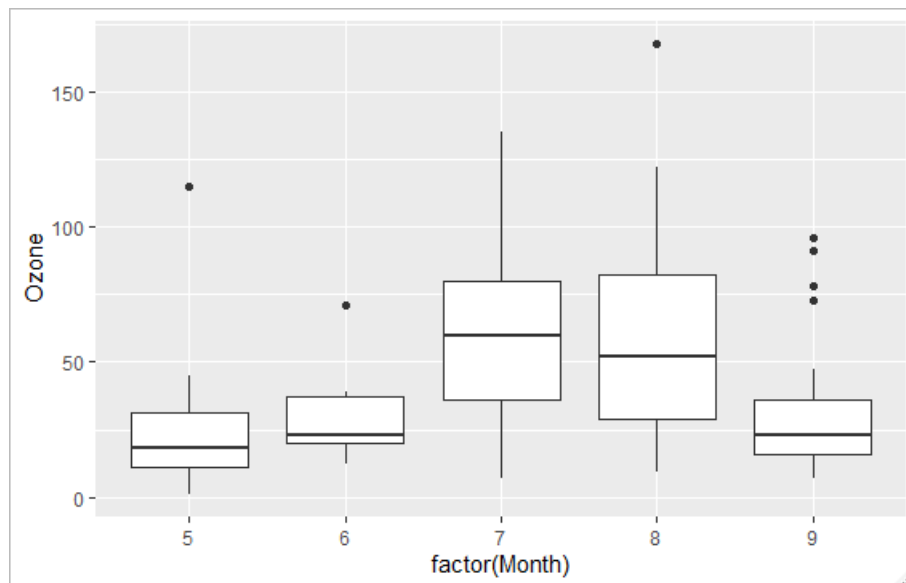
Основою є такі елементи (перші три обов'язкові):

1. **data** – фрейм, який містить дані для візуалізації.
2. **aes()** – задає зв'язок між даними; визначає змінні, що зображуються на осях; визначає, які змінні відповідають за колір і форму.
3. **geom_** - група функцій, які визначають вид зображення (точки, гістограми, лінії, текст).
4. **scale_** - група функцій, що співвідносять реальні дані з їх графічним представленням (колір, форма, розмір, діапазон осей координат).
5. **stat_** - група функцій, що додають на графік різні статистичні показники (середнє по групі, лінійна апроксимація і т. д.).
6. **coord_** - група функцій, яка задає систему координат (декартові, полярні, географічні).
7. **facet_** - група функцій, що дозволяють групувати графіки по заданому правилу.
8. **theme** - група функцій, що відповідають за оформлення графіка (розмір і колір шрифту координатних осей і поділок на них, фон графіка).

Роботу цього пакету розглянемо на декількох прикладах.

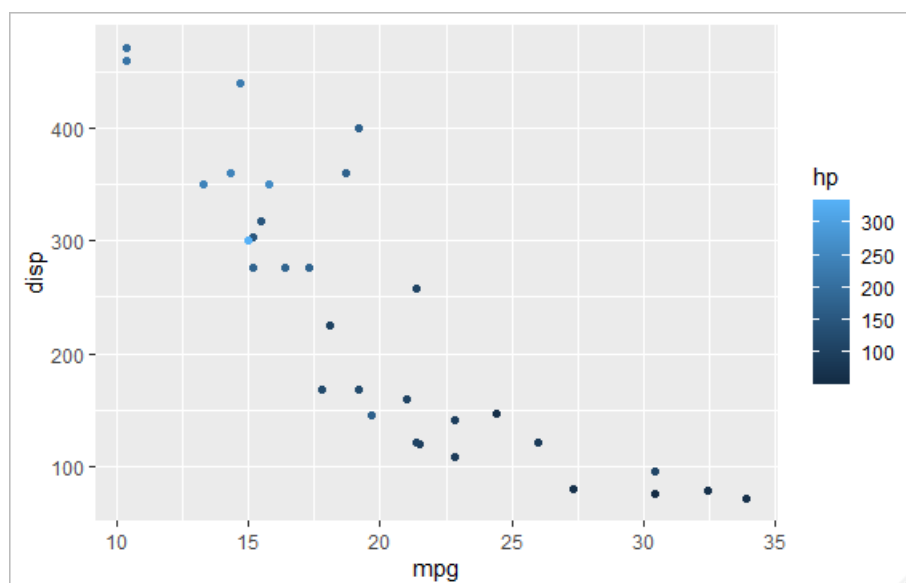
Приклад. Побудуємо `boxplot` для даних `airquality`. По осі X відкладемо номер місяця, по осі Y — значення змінної `Ozone`. Зверніть увагу, що `ggplot` очікує факторну змінну по осі X.

```
> library(ggplot2)
> ggplot(airquality,aes(factor(Month),Ozone))+
+ geom_boxplot()
```



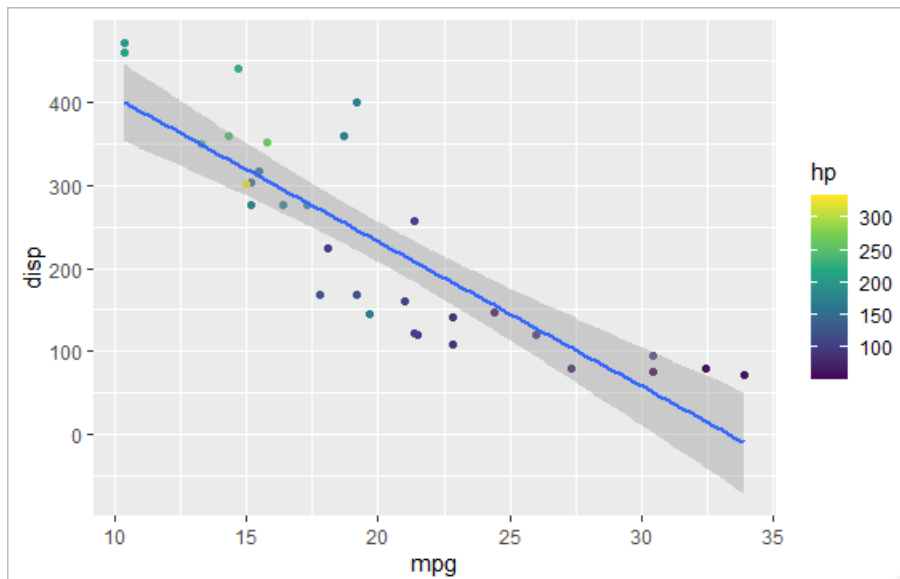
Приклад. Побудуємо точкову діаграму для даних `mtcars`: по осі X буде `mpg`, по осі Y - `disp`, а за колір буде відповідати змінна `hp`.

```
> ggplot(mtcars,aes(mpg,disp))+
+ geom_point(aes(colour = hp))
```



Приклад. До попереднього графіку додайте лінію регресії. Змініть кольорову схему точок.

```
> ggplot(mtcars,aes(mpg,disp))+  
+ geom_point(aes(colour = hp))+  
+ stat_smooth(method="lm")+  
+ scale_color_viridis_c()
```



Приклад. Побудуємо для даних **airquality** графік залежності змінної **Temp** (вісь Y) від змінної **Ozone** (вісь X), за колір відповідає змінна **Month**. Потрібно підписати осі і додати легенду:

-вісь X - "Озон".

-вісь Y - "Температура".

Назва легенди - "Місяць".

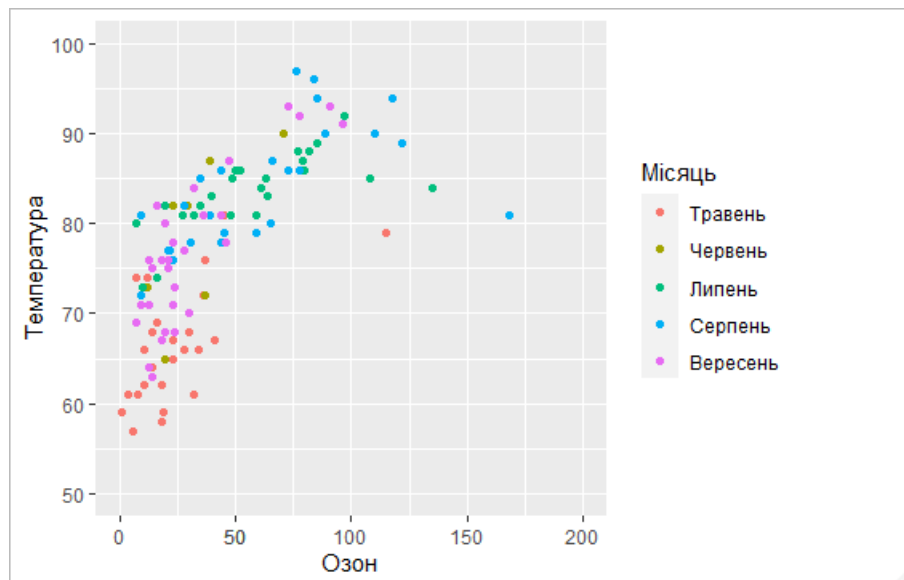
Підписи в легенді: "Травень", "Червень", "Липень", "Серпень", "Вересень".

Змінимо межі зображення по осям:

-значення по осі X повинні починатися з 0 і закінчуватись в 200 з кроком 50.

-значення по осі Y починаються з 50, закінчуються в 100 з кроком 10.

```
> ggplot(airquality,aes(Ozone,Temp,col=factor(Month)))+  
+ geom_point()+  
+ scale_color_discrete(name = "Місяць",  
+ labels = c("Травень", "Червень", "Липень", "Серпень",  
+ "Вересень"))+  
+ scale_x_continuous(name = "Озон",  
+ breaks = c(50, seq(0, 200, 50)),  
+ limits = c(0, 200))+  
+ scale_y_continuous(name = "Температура",  
+ breaks = c(10, seq(50, 100, 10)), limits = c(50, 100))
```



Функція `qplot`

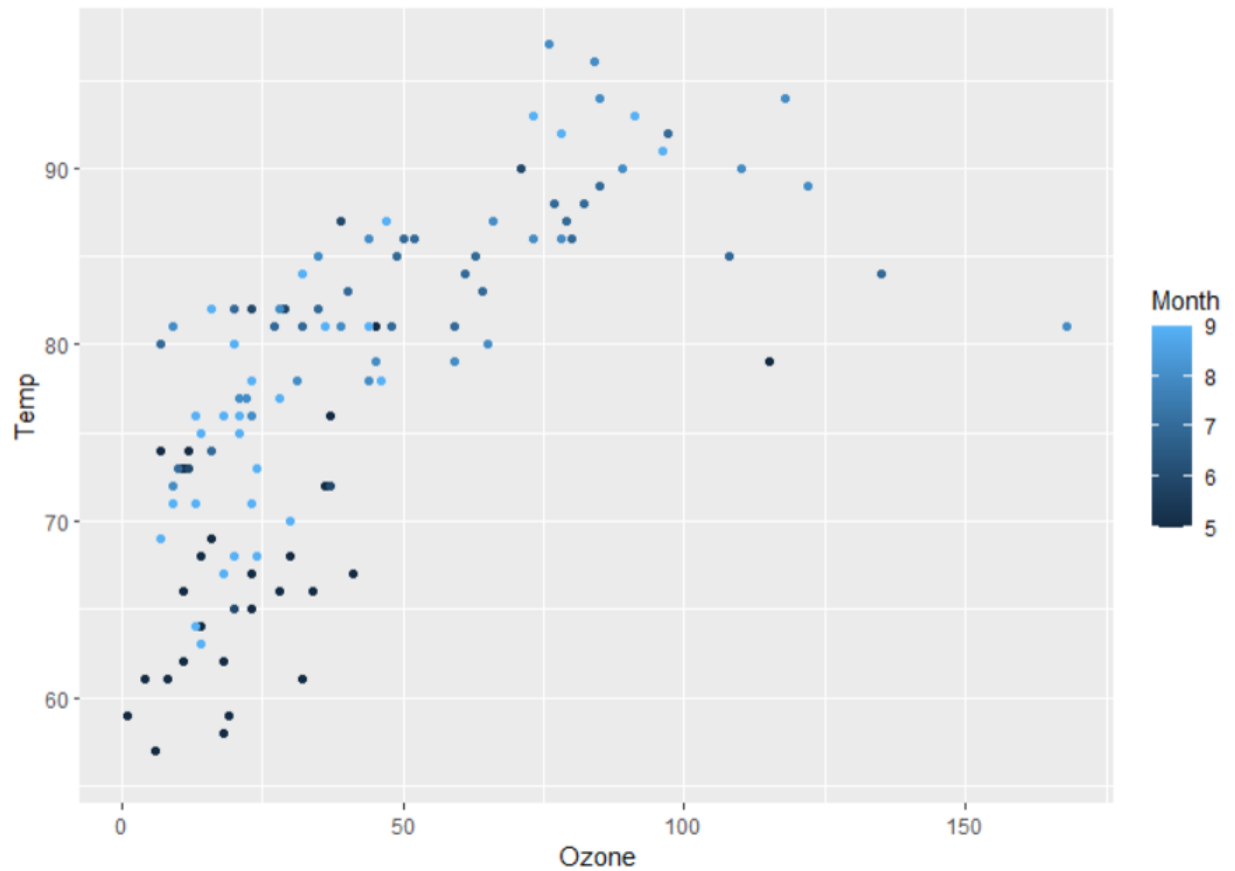
Аналог базової функції `plot` в `ggplot2` називається `qplot` і має синтаксис: `qplot(x, y, data=...)`.

Ця функція дозволяє будувати точкові діаграми, гістограми, діаграми «скриня з вусами». Параметри функції:

- **x,y** - це змінні, що відкладаються по осям. Другий аргумент можна не вказувати, якщо нам потрібні лінійні дані (наприклад, для гістограм).
- **data** – це набір даних, на основі якого будується графік. Вказавши його, ми можемо звертатись до змінних фрейму тільки за їх іменем.
- **color** відповідає за колір ліній. У випадку гістограм замість нього можна вказати **fill** = , що призведе до заповнення певним кольором.
- **shape** задає форму точок на графіку, **size** — їх розмір.
- **alpha** відповідає за прозорість частин графіку.
- **geom** задає формат графіку — **"point"**(точковий), **"line"**(лінійний), **"smooth"**(згладжений), **"boxplot"**(діаграма «скриня-вуса»), **"dotplot"**, **"histogram"**(різні форми гістограм), **"density"**(оцінка щільності на основі вибірки), **"jitter"**(графік, в якому додається шум, що розріджує точки), **"polygon"**(заповнений багатокутник).
- **facets** задає додаткові змінні.
- **xlab, ylab** задають назви осей.
- **xlim, ylim** — межі осей.
- **main, sub** — назва графіку.

Приклад.

```
> qplot(Ozone,Temp, data = airquality, color = Month)
```



Всі рисунки в лекції були створені в **RStudio** засобами базової графіки та пакету **ggplot2**.