

CSV データフレームの読み込み

```
myd <- read_csv("data/fake_data_01.csv")
is.data.frame(myd)
```

```
[1] TRUE
```

```
glimpse(myd)
```

Rows: 100

Columns: 6

```
$ id      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, ~
$ gender   <chr> "male", "male", "male", "male", "male", "male", "male", ~
$ age       <dbl> 52, 33, 22, 33, 26, 37, 50, 30, 62, 51, 55, 36, 66, 42, 36, 47, ~
$ height    <dbl> 174.0, 175.3, 175.0, 170.1, 167.4, 159.3, 173.3, 162.5, 160.2, ~
$ weight    <dbl> 63.1, 70.2, 82.6, 81.8, 51.2, 57.8, 68.6, 47.2, 68.2, 59.4, 66. ~
$ income    <dbl> 3475810, 457018, 1627793, 6070642, 1083052, 2984929, 1481061, 1~
```

```
names(myd)
```

```
[1] "id"      "gender"   "age"      "height"   "weight"   "income"
```

```
ncol(myd)
```

```
[1] 6
```

```
nrow(myd)
```

```
[1] 100
```

```
dim(myd)
```

```
[1] 100   6
```

```
head(myd)
```

```
# A tibble: 6 x 6
```

```
  id   gender   age  height  weight  income
```

```

<dbl> <chr> <dbl> <dbl> <dbl> <dbl>
1     1 male    52   174   63.1 3475810
2     2 male    33   175.  70.2 457018
3     3 male    22   175   82.6 1627793
4     4 male    33   170.  81.8 6070642
5     5 male    26   167.  51.2 1083052
6     6 male    37   159.  57.8 2984929

```

```
tail(myd)
```

```

# A tibble: 6 x 6
  id gender age height weight income
<dbl> <chr> <dbl> <dbl> <dbl> <dbl>
1 95 female 21   165.  56.3 1339138
2 96 female 65   161.  46.8 6127136
3 97 female 45   161.  48.7 1062663
4 98 female 53   166.  64.2 10154200
5 99 female 43   158.  48.5 8287163
6 100 female 48   154.  42    1125390

```

```
str(myd)
```

```

spc_tbl_ [100 x 6] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
$ id      : num [1:100] 1 2 3 4 5 6 7 8 9 10 ...
$ gender: chr [1:100] "male" "male" "male" "male" ...
$ age     : num [1:100] 52 33 22 33 26 37 50 30 62 51 ...
$ height: num [1:100] 174 175 175 170 167 ...
$ weight: num [1:100] 63.1 70.2 82.6 81.8 51.2 57.8 68.6 47.2 68.2 59.4 ...
$ income: num [1:100] 3475810 457018 1627793 6070642 1083052 ...
- attr(*, "spec")=
.. cols(
..   id = col_double(),
..   gender = col_character(),
..   age = col_double(),
..   height = col_double(),
..   weight = col_double(),
..   income = col_double()
.. )
- attr(*, "problems")=<externalptr>
```

データフレームの作成

```
df1 <- data.frame(x = 1:100,  
                   y = 100:1)  
is.data.frame(df1)
```

[1] TRUE

```
glimpse(df1)
```

Rows: 100
Columns: 2
\$ x <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 2~
\$ y <int> 100, 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84,~

```
df2 <- tibble(v1 = rnorm(100, mean = 0, sd = 5),  
              v2 = rnorm(100, mean = -4, sd = 5),  
              v3 = rnorm(100, mean = 0, sd = 1))  
is.data.frame(df2)
```

[1] TRUE

```
glimpse(df2)
```

Rows: 100
Columns: 3
\$ v1 <dbl> -2.26012024, 5.50264549, -3.10592118, -3.49378467, 4.21550806, -7.2~
\$ v2 <dbl> -1.8016196, 1.0491082, -7.8270854, -6.3953292, -2.9109673, -1.17651~
\$ v3 <dbl> 0.93645314, 0.41817297, -0.58227186, 0.95516310, -0.80602701, -1.69~

組み込みデータ

```
data(mtcars)  
glimpse(mtcars)
```

Rows: 32
Columns: 11

```
$ mpg <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19.2, 17.8,~  
$ cyl <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 4, 4, 4, 4, 8,~  
$ disp <dbl> 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 140.8, 16~  
$ hp <dbl> 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, 180, 180~  
$ drat <dbl> 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.92, 3.92,~  
$ wt <dbl> 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150, 3.~  
$ qsec <dbl> 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 22.90, 18~  
$ vs <dbl> 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0,~  
$ am <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0,~  
$ gear <dbl> 4, 4, 4, 3, 3, 3, 4, 4, 4, 3, 3, 3, 3, 3, 4, 4, 4, 3, 3,~  
$ carb <dbl> 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, 1, 1, 2,~
```

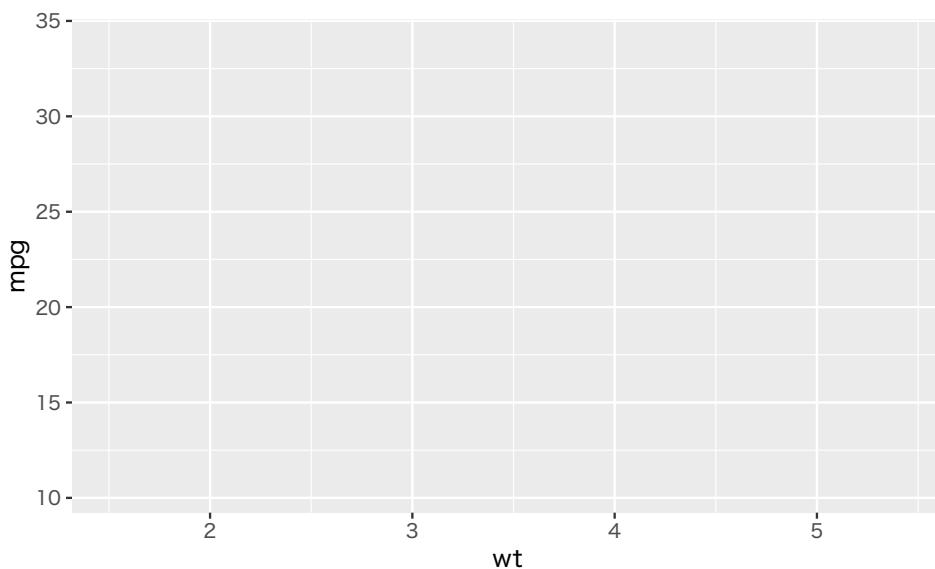
```
is.data.frame(mtcars)
```

```
[1] TRUE
```

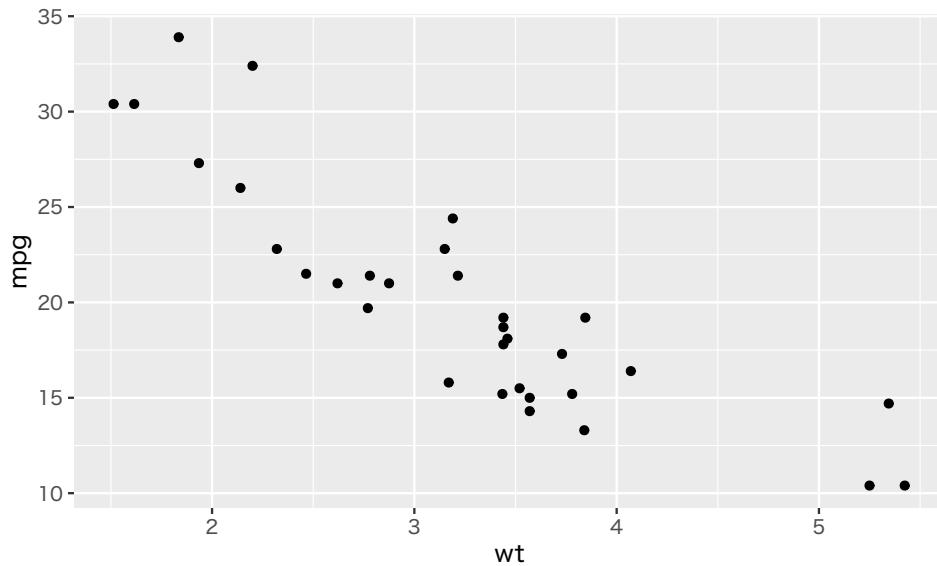
ggplot の基礎

散布図

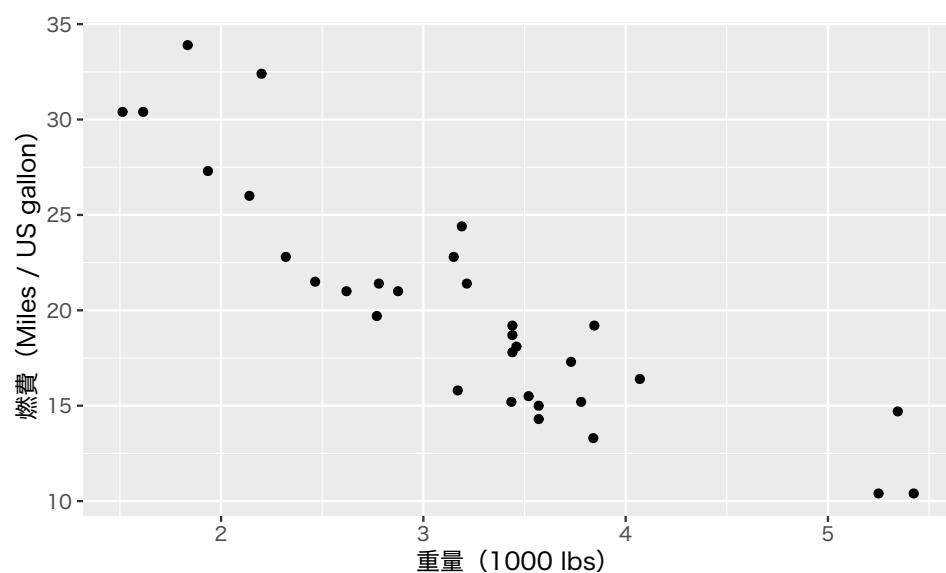
```
p1_1 <- ggplot(data = mtcars,  
                  mapping = aes(x = wt, y = mpg))  
plot(p1_1)
```



```
p1_2 <- p1_1 +  
  geom_point()  
plot(p1_2)
```



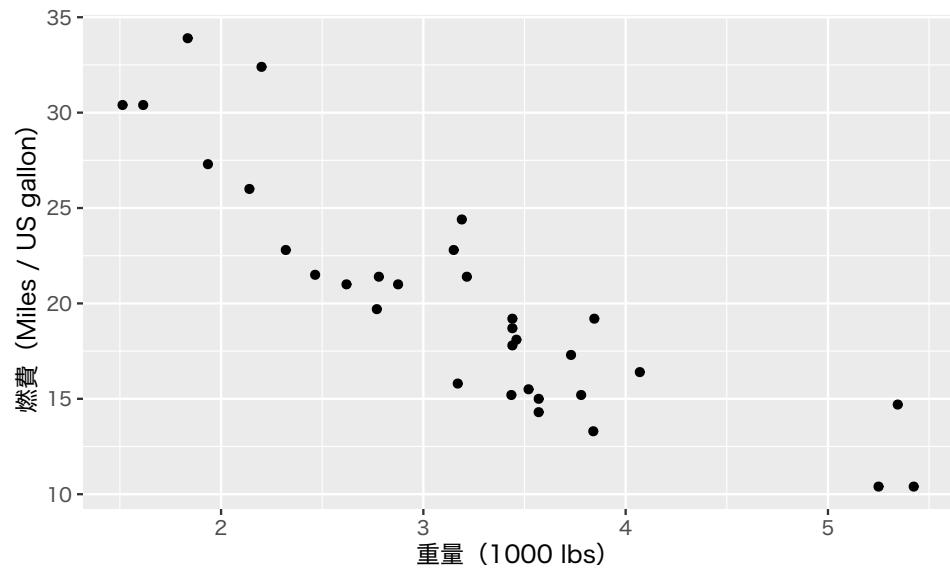
```
p1_3 <- p1_2 +  
  labs(x = '重量(1000 lbs)',  
        y = '燃費(Miles / US gallon)')  
plot(p1_3)
```



```

p1 <- ggplot(mtcars,
              aes(x = wt,
                  y = mpg)) +
  geom_point() +
  labs(x = '重量(1000 lbs)',
       y = '燃費(Miles / US gallon)')
plot(p1)

```

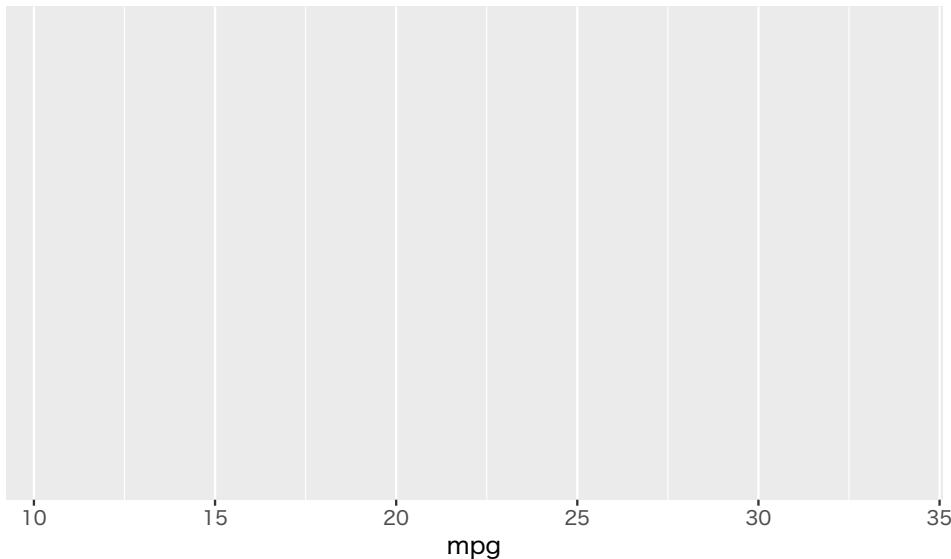


ヒストグラム

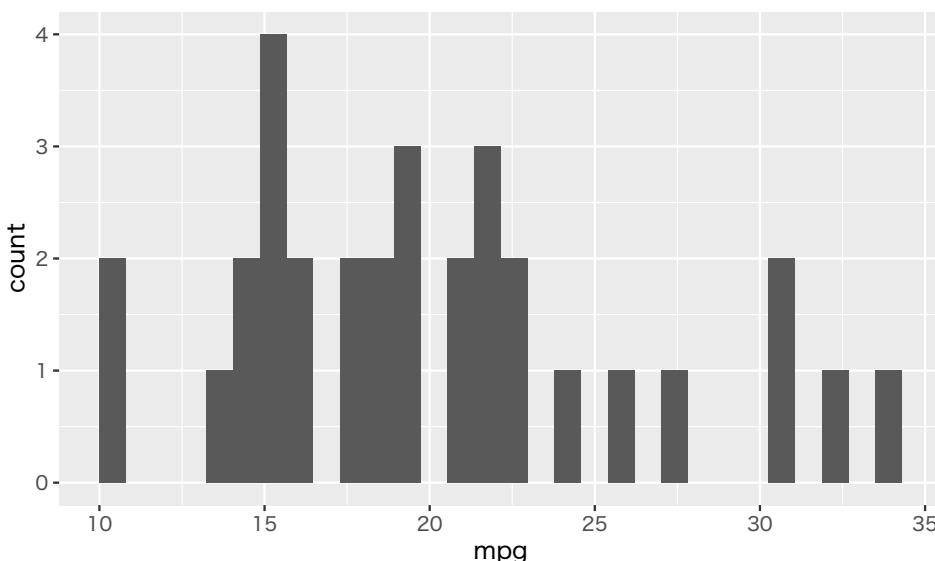
```

p2_1 <- ggplot(mtcars, aes(x = mpg))
plot(p2_1)

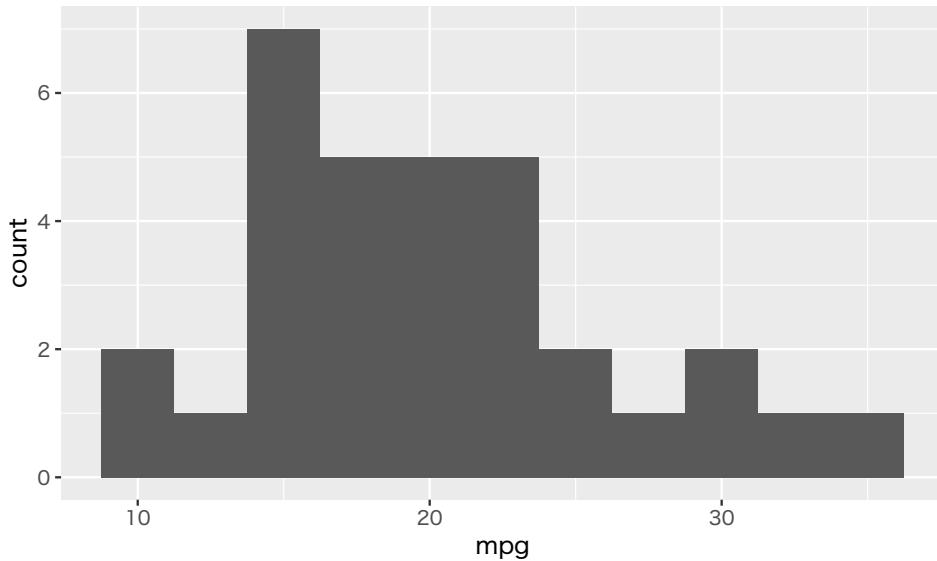
```



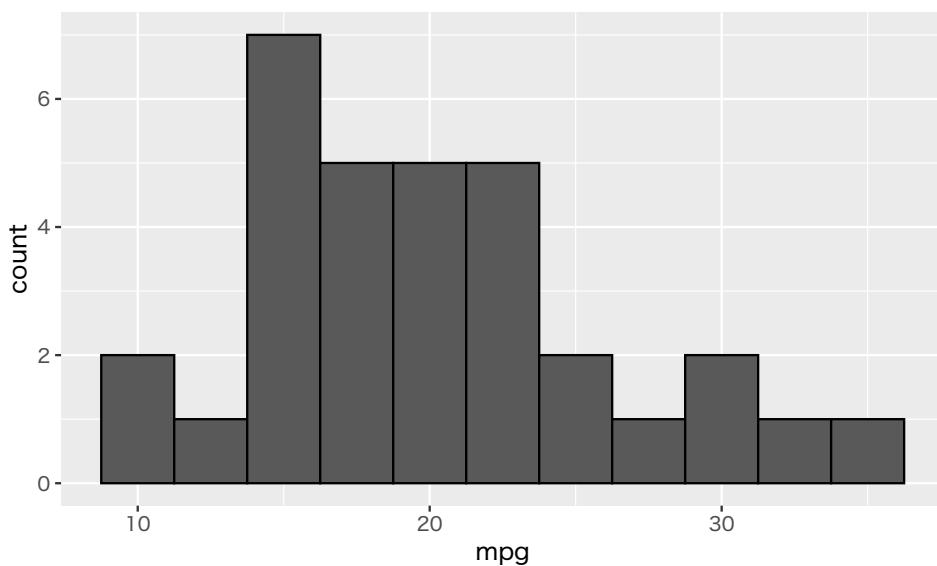
```
p2_2 <- p2_1 +  
  geom_histogram()  
plot(p2_2)
```



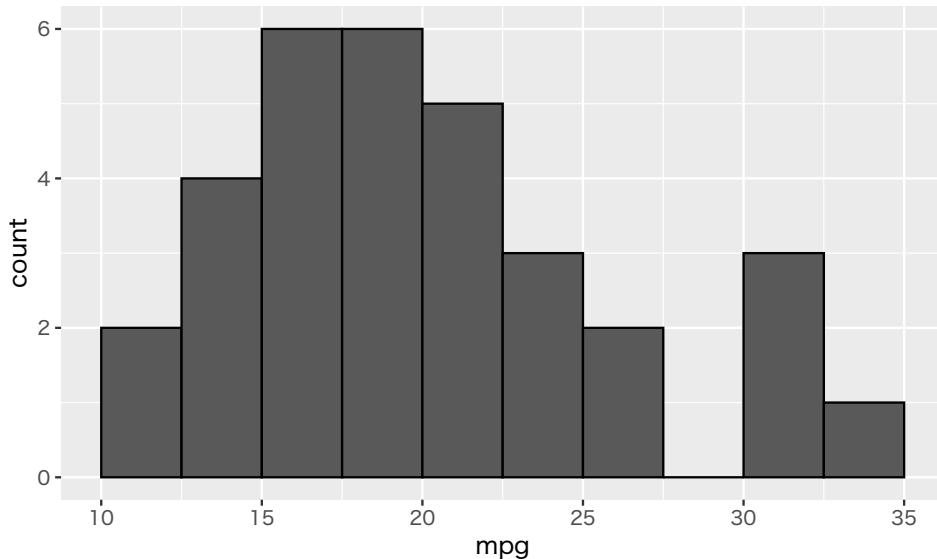
```
p2_3 <- p2_1 +  
  geom_histogram(binwidth = 2.5)  
plot(p2_3)
```



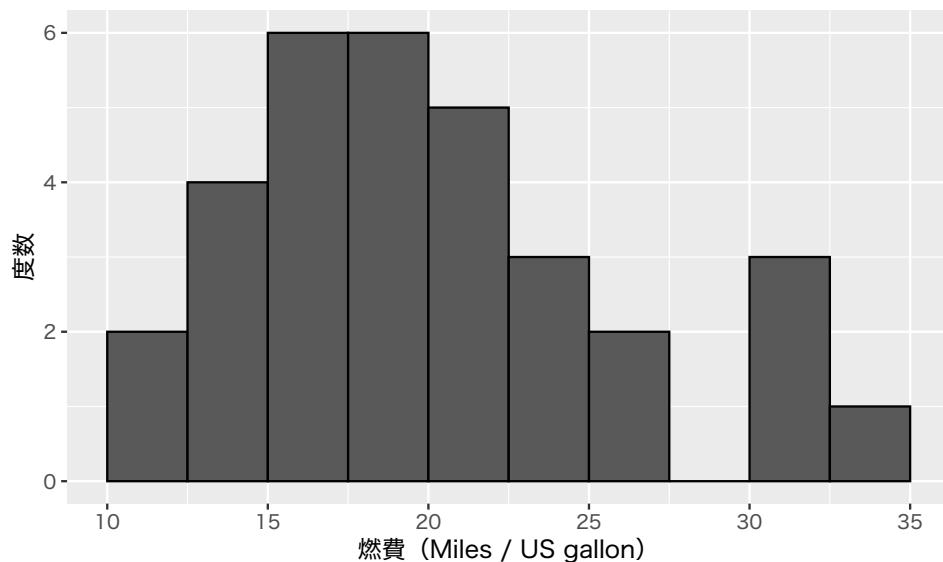
```
p2_4 <- p2_1 +  
  geom_histogram(binwidth = 2.5,  
                 color     = 'black')  
plot(p2_4)
```



```
p2_5 <- p2_1 +  
  geom_histogram(binwidth = 2.5,  
                 color     = 'black',  
                 boundary = 5)  
plot(p2_5)
```



```
p2_6 <- p2_5 +
  labs(x = '燃費(Miles / US gallon)' ,
       y = '度数')
plot(p2_6)
```



```
p2_dens <- ggplot(mtcars,
  aes(x = mpg)) +
  geom_histogram(binwidth = 2.5,
    color     = 'black',
    fill      = 'dodgerblue',
```

```
boundary = 5) +  
  labs(x = '燃費(Miles / US gallon)',  
        y = '確率密度')  
  plot(p2_dens)
```

