

Kiri Kuroda

Contents

		5
R RStudio		7
0.1 R		7
0.2 RStudio		7
0.3		7
0.4		7
0.5 R		7
		11
1		11
1.1		11
1.2 ggplot2		13
2		15
2.1		15
2.2		19
3		21
3.1		21
3.2 Cleveland		24
3.3		26
4		27
4.1		27
4.2		30
4.3		32
4.4		36
4.5		37
4.6		38
5		39
5.1		39
5.2		39
5.3		41
5.4		42
5.5		45
6		47
6.1		47
6.2		48

6.3	50
6.4	51
7		53
7.1	53
7.2	54
7.3	56
7.4	57
8		59
8.1	59
8.2	Small multiple	59
8.3	61
8.4	61
8.5	63
8.6	65
8.7	65
8.8	66
8.9	66
		69
9		69
9.1	tidyverse	70
9.2	70
9.3	73
9.4	74
9.5	75
9.6	76
9.7	78
9.8	ggplot2	78
		83
		83
		85

R RStudio

1

0.1 R

<https://cran.r-project.org> OS

0.2 RStudio

<https://www.rstudio.com> RStudio “Free”

0.3

R RStudio R RStudio iOS Android `install.packages(" ")` `install.packag`

```
#  
install.packages(c("tidyverse", "gcookbook", "GGally", "devtools"))
```

0.4

RStudio `library()`

```
library(tidyverse) #
```

0.5 R

..... https://kazutan.github.io/JSSP2018_spring/intro_rstudio.html R
Markdown

Chapter 1

Anscombe's quartet

1.1

4		I, II, III, IV	11	x	y	x	y
4	x y	mean; M	standard deviation; SD	correlation coefficient; r			
experiment							
mean_x							
sd_x							
mean_y							
sd_y							
pearson_r							
I							
9							
3.316625							
7.500909							
2.031568							
0.8164205							
II							
9							
3.316625							
7.500909							
2.031657							
0.8162365							
III							
9							

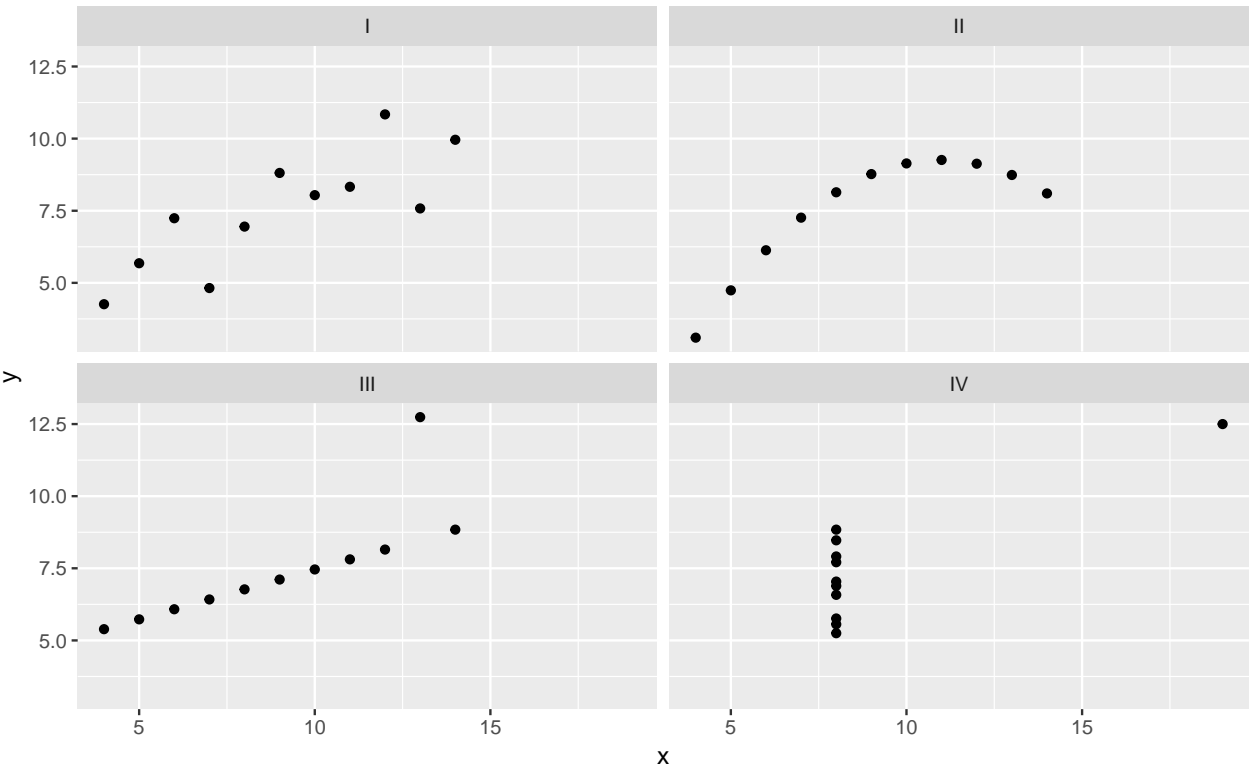


Figure 1.1: (ref:anscombe-scatter-plot)

3.316625

7.500000

2.030424

0.8162867

IV

9

3.316625

7.500909

2.030579

0.8165214

x y

.....

x y

(ref:anscombe-scatter-plot)

x y

- distribution
- association

1.2 ggplot2

R ggplot2 tidyverse 1 tidyverse ggplot2 R

ggplot2 Ex el

-
- ugly!
-

.....

R ggplot2

Chapter 2

2.1

```
library(ggplot2)
library(ggplot2)

# *
#
#
#

R      mtcars      mtcars      ?mtcars      str( )
str(mtcars) #

## 'data.frame':   32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110  93 110 175 105 245  62  95 123 ...
## $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num   16.5 17 18.6 19.4 17 ...
## $ vs  : num   0  0  1  1  0  1  0  1  1  1 ...
## $ am  : num   1  1  1  0  0  0  0  0  0  0 ...
## $ gear: num   4  4  4  3  3  3  3  4  4  4 ...
## $ carb: num   4  4  1  1  2  1  4  2  2  4 ...

11 variable 32 observation      32 x 11      head( ) 6
head(mtcars) # 6

##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 0  1    4    4
## Datsun 710      22.8   4  108  93 3.85 2.320 18.61 1  1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44 1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 0  0    3    2
## Valiant         18.1   6  225 105 2.76 3.460 20.22 1  0    3    1
```

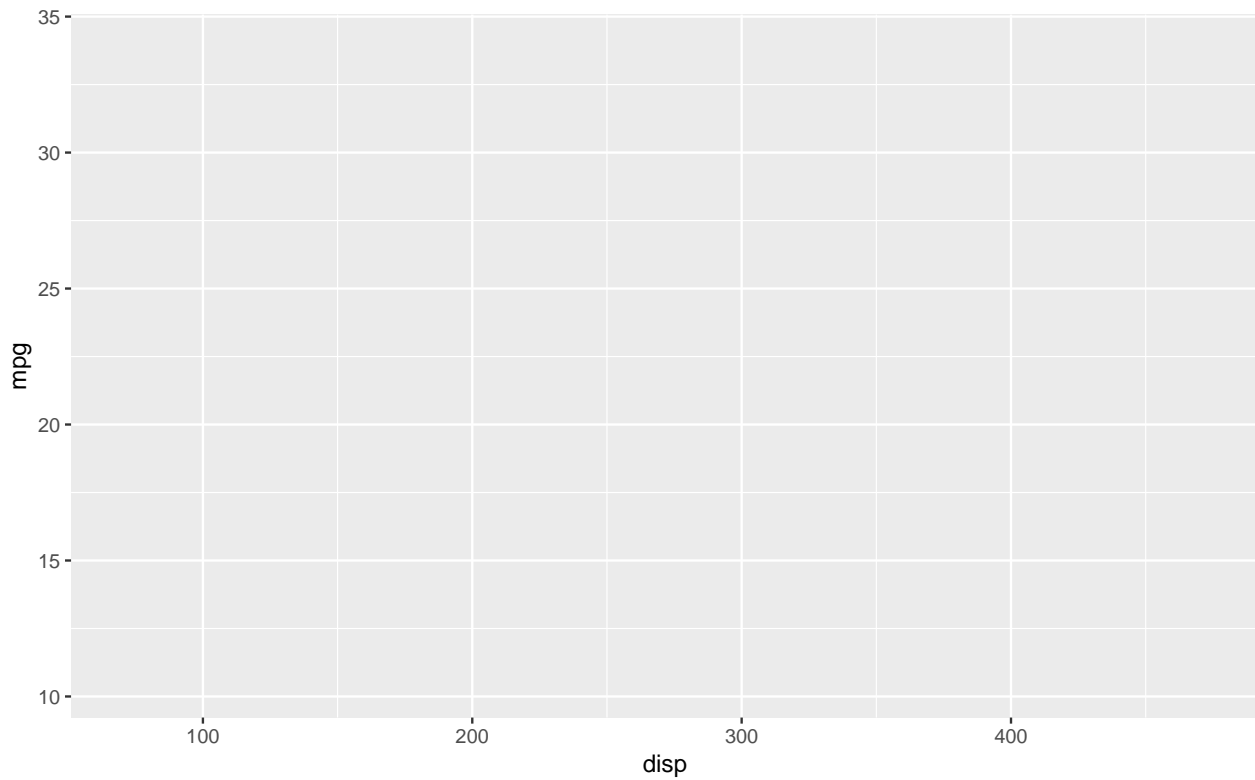


Figure 2.1:

```
11     2
```

- `dis`
- `mpg`

continuous variable quantitative variable

`ggplot2`

1. `x y`

```
ggplot2          ggplot(data = , aes(x = x , y = y ))
ggplot(data = mtcars, aes(x = disp, y = mpg))
```

1

2.

geometry `geom_xxxx()` `geom_point()`

+

```
ggplot(data = mtcars, aes(x = disp, y = mpg)) + #
  geom_point() # point
```

¹`ggplot2` layer

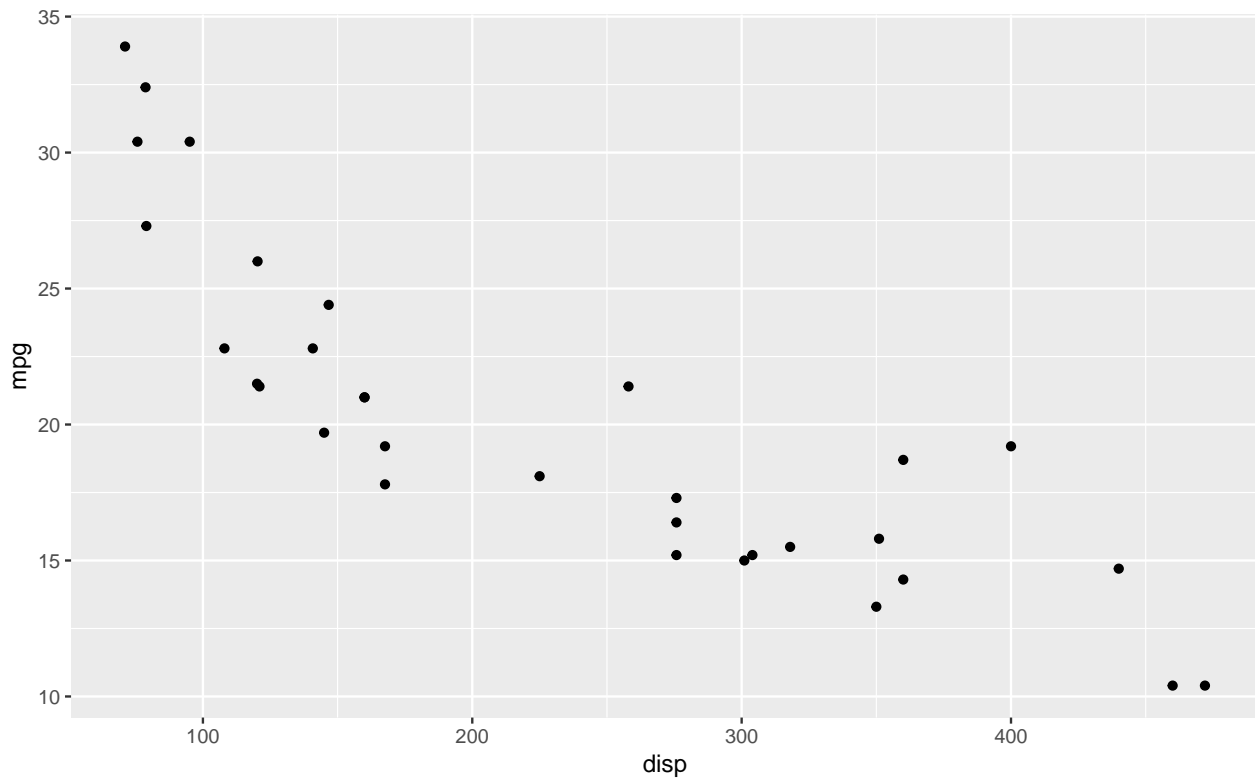


Figure 2.2: x y

x y x = disp y = mpg data = mtcars data, x,

```
ggplot(mtcars, aes(displ, mpg)) +  
  geom_point()
```

3.

OK 1

- cyl 4, 6, or 8

cyl OK

```
ggplot(mtcars, aes(displ, mpg, color = cyl)) + # color = year  
  geom_point()
```

aes() color = cyl cyl

able legend 1 4,5,6,7,8 4,6,8 3 discrete vari-

cyl factor()

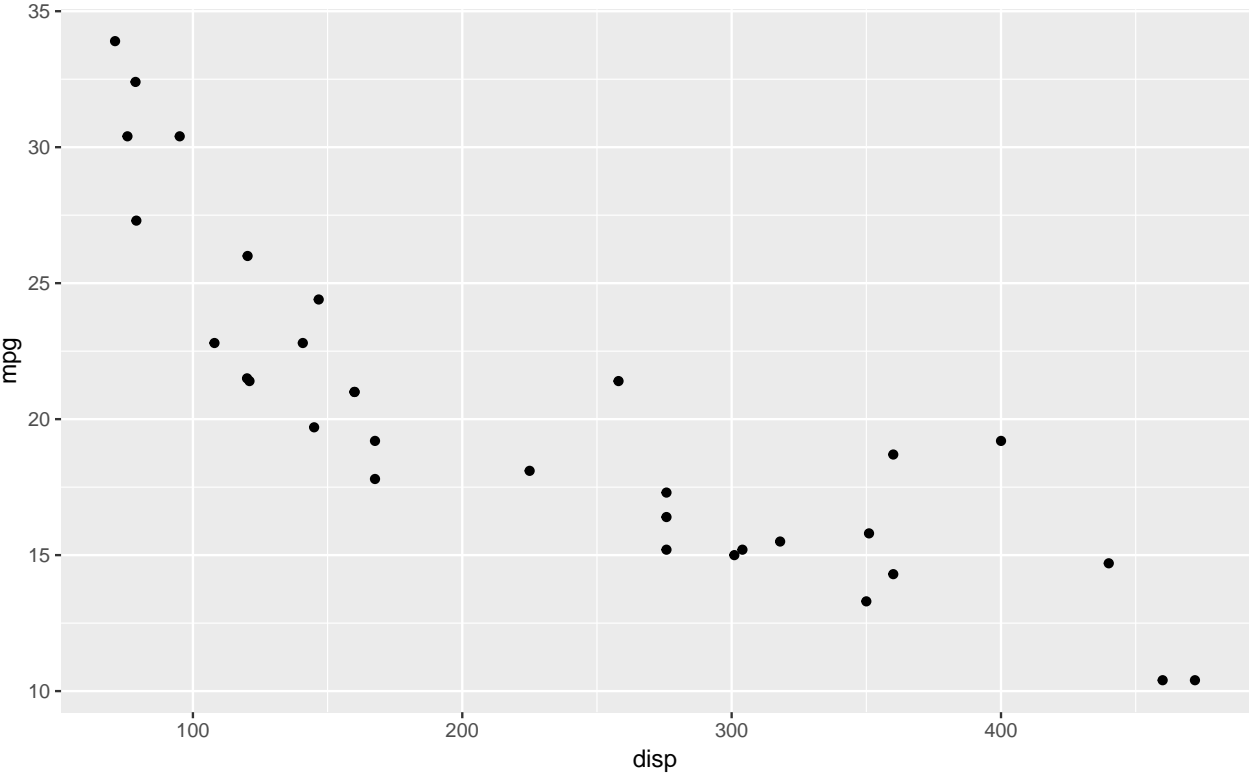


Figure 2.3: Figure 2.2

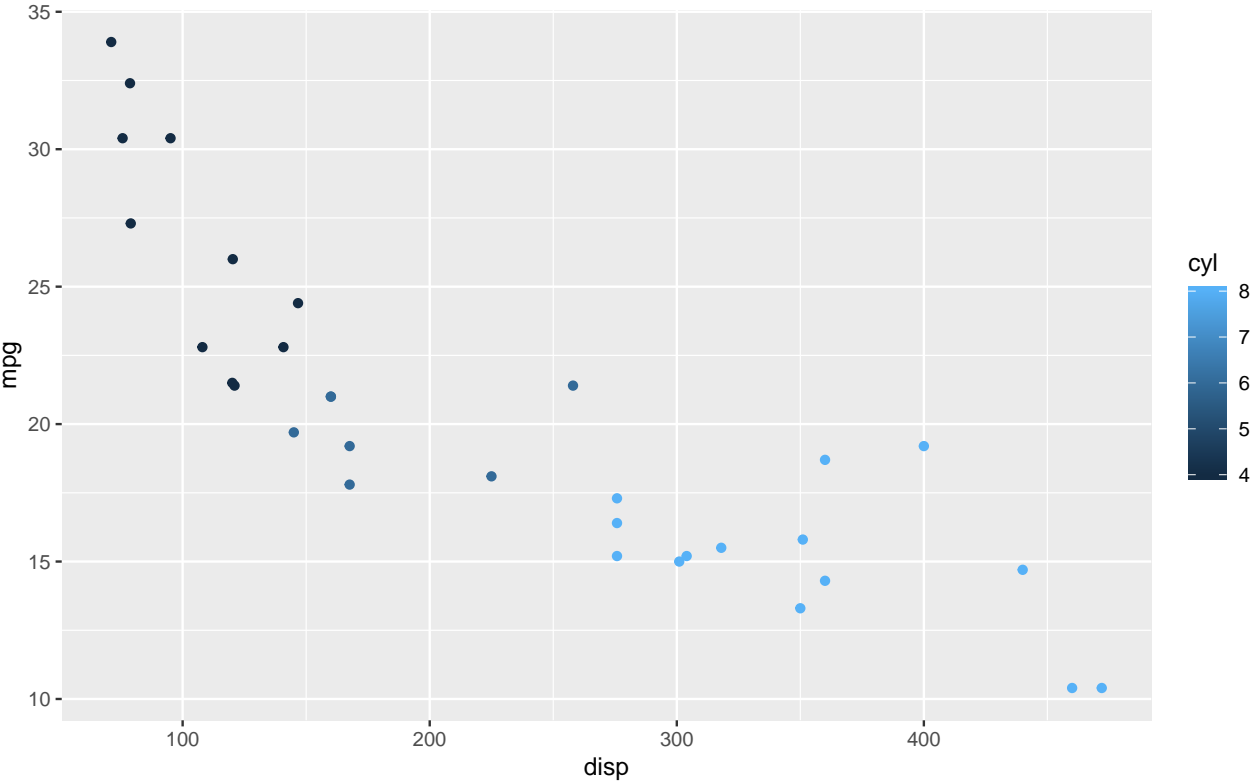


Figure 2.4:

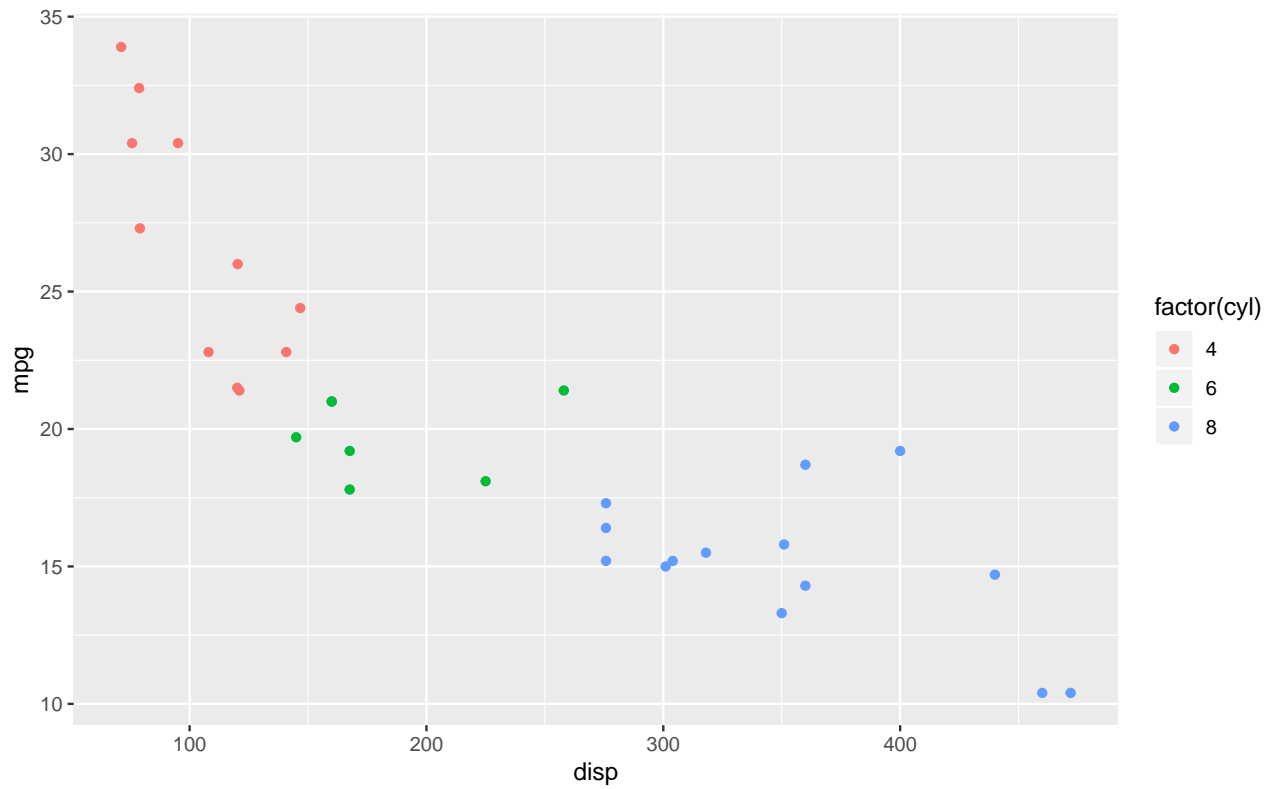


Figure 2.5:

```
ggplot(mtcars, aes(displacement, mpg, color = factor(cyl))) + # factor(cyl)
  geom_point()
```

R factor()

2.2

ggplot2

1. x y

- y x
-

2.

-
- x y 3
-

3.

- `x y`
-

<code>ggplot2</code>	<code>geom_xxxx()</code>																													
	<table><tr><td colspan="2"><code>geom</code></td></tr><tr><td colspan="2"><hr/></td></tr><tr><td colspan="2"><code>geom_bar()</code></td></tr><tr><td colspan="2"><code>geom_line()</code></td></tr><tr><td colspan="2"><code>geom_point()</code></td></tr><tr><td colspan="2"><code>geom_errorbar()</code></td></tr><tr><td colspan="2"><code>geom_pointrange()</code></td></tr><tr><td colspan="2"><code>geom_histogram()</code></td></tr><tr><td colspan="2"><code>geom_density()</code></td></tr><tr><td colspan="2"><code>geom_boxplot()</code></td></tr><tr><td colspan="2"><code>geom_violin()</code></td></tr><tr><td colspan="2"><code>geom_area()</code></td></tr><tr><td colspan="2"><code>geom_smooth()</code></td></tr><tr><td colspan="2"><code>geom_text()</code></td></tr><tr><td colspan="2"><hr/></td></tr></table>	<code>geom</code>		<hr/>		<code>geom_bar()</code>		<code>geom_line()</code>		<code>geom_point()</code>		<code>geom_errorbar()</code>		<code>geom_pointrange()</code>		<code>geom_histogram()</code>		<code>geom_density()</code>		<code>geom_boxplot()</code>		<code>geom_violin()</code>		<code>geom_area()</code>		<code>geom_smooth()</code>		<code>geom_text()</code>		<hr/>
<code>geom</code>																														
<hr/>																														
<code>geom_bar()</code>																														
<code>geom_line()</code>																														
<code>geom_point()</code>																														
<code>geom_errorbar()</code>																														
<code>geom_pointrange()</code>																														
<code>geom_histogram()</code>																														
<code>geom_density()</code>																														
<code>geom_boxplot()</code>																														
<code>geom_violin()</code>																														
<code>geom_area()</code>																														
<code>geom_smooth()</code>																														
<code>geom_text()</code>																														
<hr/>																														
<code>x y color</code>	<code>aes()</code>																													
	<table><tr><td colspan="2"><code>aes()</code></td></tr><tr><td colspan="2"><hr/></td></tr><tr><td><code>x</code></td><td><code>x</code></td></tr><tr><td><code>y</code></td><td><code>y</code></td></tr><tr><td colspan="2"><code>color</code></td></tr><tr><td colspan="2"><code>fill</code></td></tr><tr><td colspan="2"><code>linetype</code></td></tr><tr><td colspan="2"><code>size</code></td></tr><tr><td colspan="2"><code>shape</code></td></tr><tr><td colspan="2"><hr/></td></tr></table>	<code>aes()</code>		<hr/>		<code>x</code>	<code>x</code>	<code>y</code>	<code>y</code>	<code>color</code>		<code>fill</code>		<code>linetype</code>		<code>size</code>		<code>shape</code>		<hr/>										
<code>aes()</code>																														
<hr/>																														
<code>x</code>	<code>x</code>																													
<code>y</code>	<code>y</code>																													
<code>color</code>																														
<code>fill</code>																														
<code>linetype</code>																														
<code>size</code>																														
<code>shape</code>																														
<hr/>																														

Chapter 3

```
library(ggplot2) #
```

3.1

3.1.1

bar graph; bar chart; bar plot

gcookbook pg_mean gcookbook

```
library(gcookbook)
```

```
head(pg_mean) # 3
```

```
##   group weight
## 1  ctrl  5.032
## 2  trt1  4.661
## 3  trt2  5.526
```

2

- group 3
- weight

x y geom_bar()

```
ggplot(pg_mean, aes(group, weight)) + # x group y weight
geom_bar(stat = "identity") #      stat = "identity"
```

stat = "identity" geom_bar() stat = "count" ¹

stat = "identity" : stat_count() must not be used with a y aesthetic.

¹

² PC

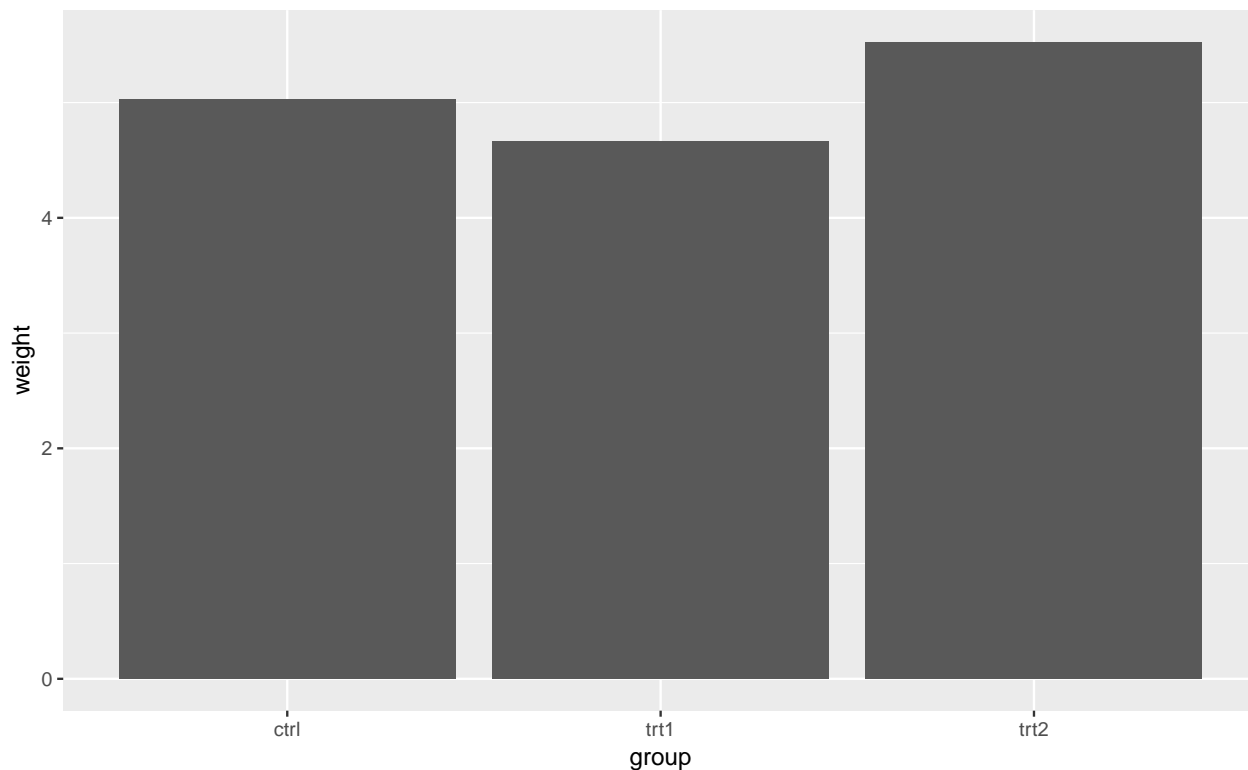


Figure 3.1: x y

3.1.2

```
1      gcookbook cabbage_exp
head(cabbage_exp) # 6
```

```
##   Cultivar Date Weight      sd  n      se
## 1      c39  d16   3.18 0.9566144 10 0.30250803
## 2      c39  d20   2.80 0.2788867 10 0.08819171
## 3      c39  d21   2.74 0.9834181 10 0.31098410
## 4      c52  d16   2.26 0.4452215 10 0.14079141
## 5      c52  d20   3.11 0.7908505 10 0.25008887
## 6      c52  d21   1.47 0.2110819 10 0.06674995
```

```
3
• Date      d16, d20, or d21
• Cultivar   c39 or c52
• Weight     kg
```

```
3      Date x Weight y Cultivar fill
ggplot(cabbage_exp, aes(Date, Weight, fill = Cultivar)) + # x Date y Weight fill = Cultivar
geom_bar(stat = "identity", position = "dodge") # position = "dodge"
```

```
2      fill      fill      Cultivar      4
```

```
3
4      color      geom_bar() color
```

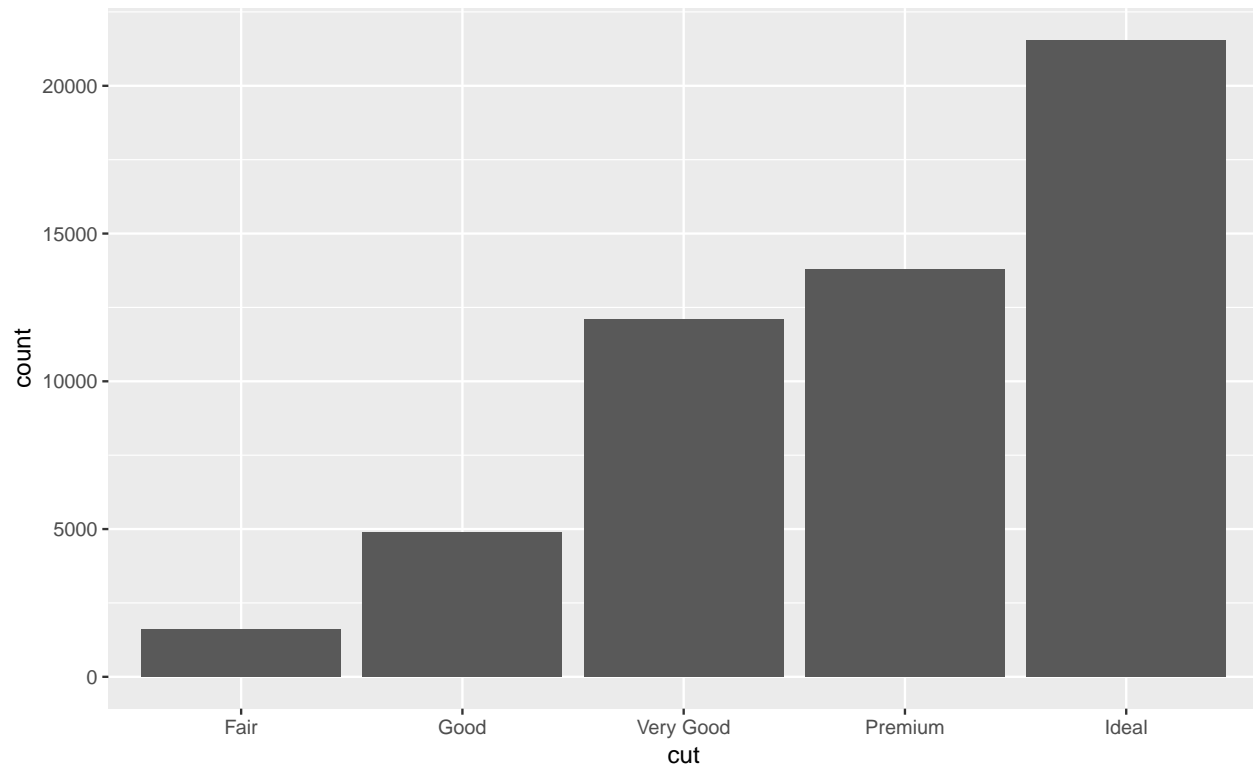



Figure 3.3:

3.2 Cleveland

gcookbook tophitters2001 2001 MLB 144

```
head(tophitters2001)
```

```
##      id   first last      name year stint team lg   g  ab   r
## 1 walkela01  Larry Walker  Larry Walker 2001     1  COL  NL 142 497 107
## 2 suzukic01  Ichiro Suzuki  Ichiro Suzuki 2001     1  SEA  AL 157 692 127
## 3 giambja01  Jason Giambi  Jason Giambi 2001     1  OAK  AL 154 520 109
## 4 alomaro01 Roberto Alomar  Roberto Alomar 2001     1  CLE  AL 157 575 113
## 5 heltoto01  Todd Helton   Todd Helton 2001     1  COL  NL 159 587 132
## 6 aloumo01  Moises Alou    Moises Alou 2001     1  HOU  NL 136 513   79
##      h 2b 3b hr rbi sb cs  bb  so ibb hbp sh sf gidp   avg
## 1 174 35  3 38 123 14  5  82 103  6 14  0  8    9 0.3501
## 2 242 34  8  8  69 56 14  30  53 10  8  4  4    3 0.3497
## 3 178 47  2 38 120  2  0 129  83 24 13  0  9   17 0.3423
## 4 193 34 12 20 100 30  6  80  71  5  4  9  9    9 0.3357
## 5 197 54  2 49 146  7  5  98 104 15  5  1  5   14 0.3356
## 6 170 31  1 27 108  5  1  57  57 14  3  0  8   18 0.3314
```

- name

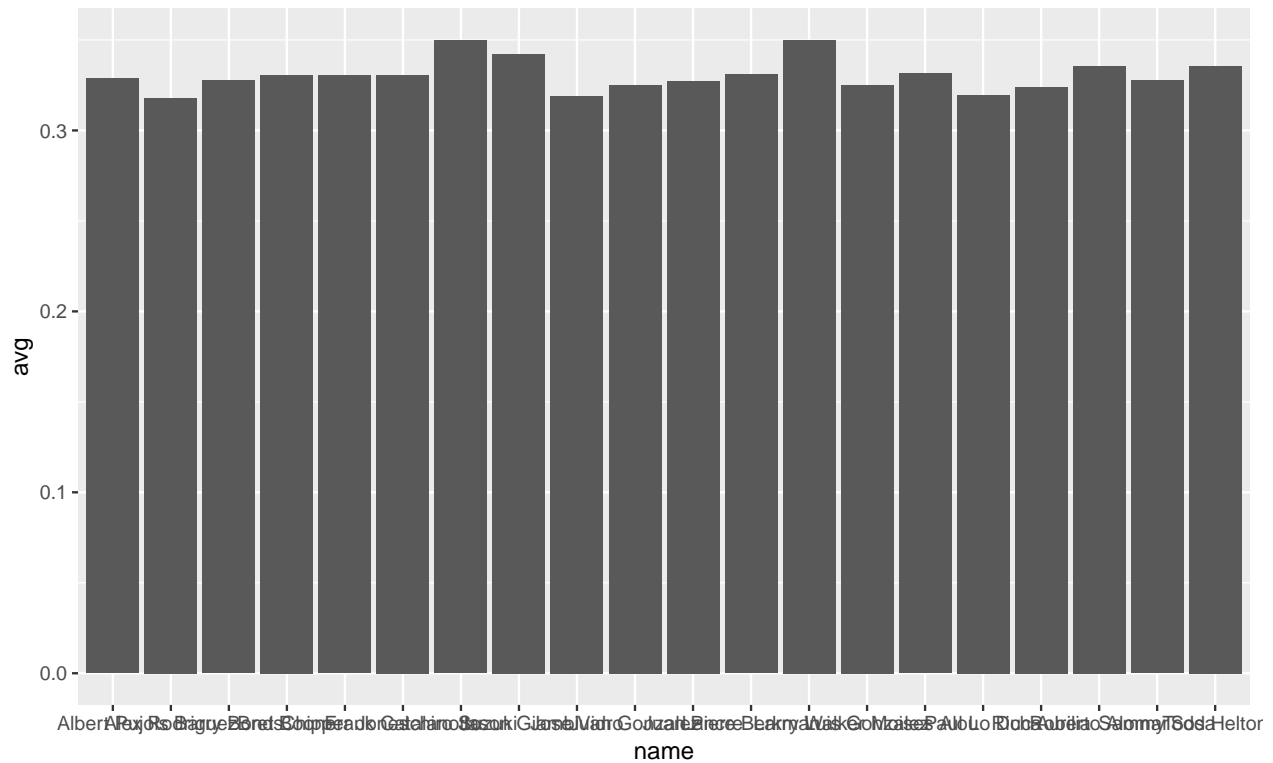


Figure 3.4: 2001 MLB top 20

- avg

```
144      20      20      top20hitters
top20hitters = tophitters2001[1:20,] # 20 top20hitters
```

```
top20hitters
```

```
ggplot(top20hitters, aes(name, avg)) + # x name y avg
  geom_bar(stat = "identity") #
```

```
2
```

```
1.
```

- y = 0 5 0.3 y = 0

```
2.
```

-

```
1 Cleveland Cleveland dot plot
```

```
ggplot(top20hitters, aes(avg, reorder(name, avg))) + # x y reorder()
  geom_point() # stat = "identity"
```

```
1. ggplot(top20hitter, aes(avg, reorder(name, avg)))
```

- x y

```
5 y = 0
```

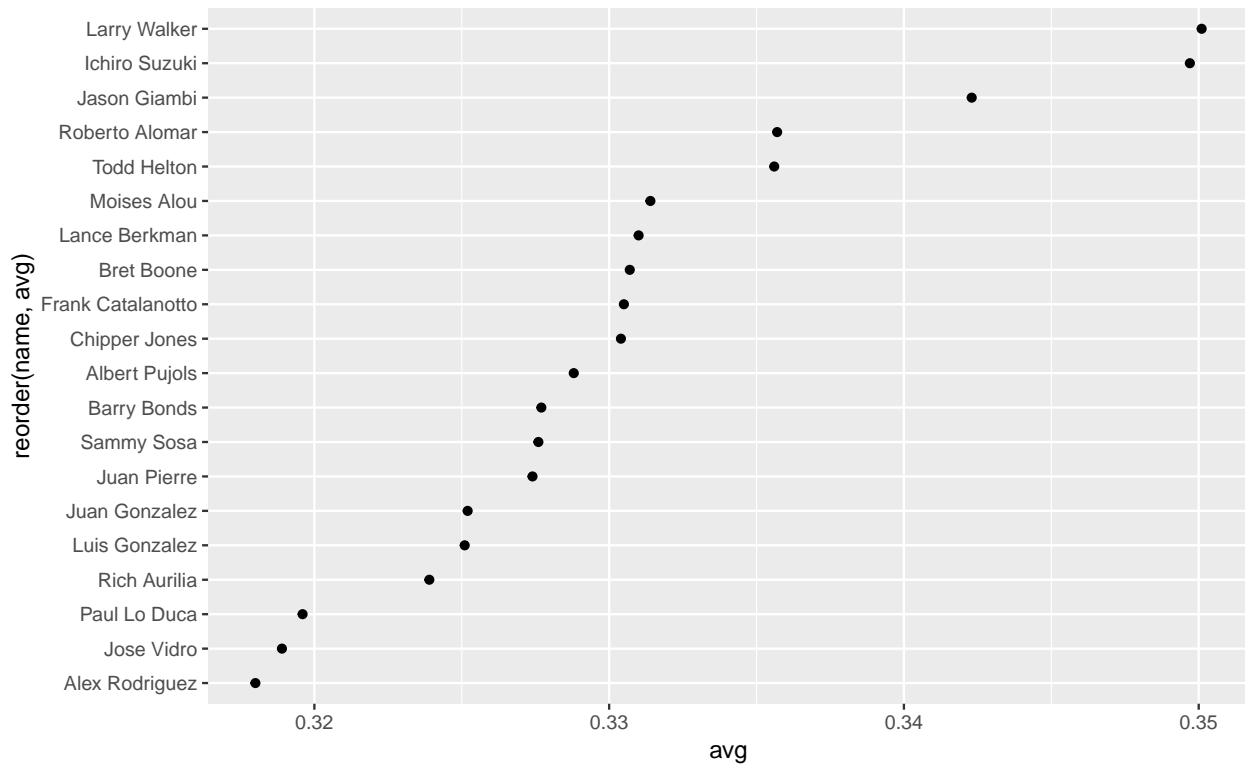


Figure 3.5: Cleveland 2001 MLB 20

```

• reorder(name, avg)          reorder(a, b) a b          reorder()
2. geom_point()
•
•   geom_point() stat = "identity"   geom_bar()   stat
Cleveland      6   Figure 3.4 Figure 3.5

```

3.3

- cabbage_exp x Cultivar y Weight Date
- diamonds clarity
- gcbookbook uspopchange ?uspopchange head(uspopchange)

Chapter 4

Chapter 1 summary statistics descriptive statistics distribution

```
library(ggplot2) #
```

4.1

4.1.1

 histogram x y

Chapter 1 gcookbook tophitters2001 2001 MLB 144 gcookbook

```
library(gcookbook)
```

```
head(tophitters2001) #                      6
```

```
##           id   first   last           name year stint team lg   g  ab   r
## 1 walkela01   Larry Walker   Larry Walker 2001     1  COL  NL 142 497 107
## 2 suzukic01  Ichiro Suzuki   Ichiro Suzuki 2001     1  SEA  AL 157 692 127
## 3 giambja01   Jason Giambi   Jason Giambi 2001     1  OAK  AL 154 520 109
## 4 alomaro01  Roberto Alomar   Roberto Alomar 2001     1  CLE  AL 157 575 113
## 5 heltoto01   Todd Helton     Todd Helton 2001     1  COL  NL 159 587 132
## 6 aloumo01   Moises Alou     Moises Alou 2001     1  HOU  NL 136 513   79
##           h 2b 3b hr rbi sb cs  bb  so ibb hbp sh sf gidp  avg
## 1 174 35   3 38 123 14   5  82 103   6  14   0   8    9 0.3501
## 2 242 34   8   8  69  56 14  30  53  10   8   4   4    3 0.3497
## 3 178 47   2 38 120   2   0 129  83  24  13   0   9   17 0.3423
## 4 193 34  12 20 100 30   6  80  71   5   4   9   9    9 0.3357
## 5 197 54   2 49 146   7   5  98 104  15   5   1   5   14 0.3356
## 6 170 31   1 27 108   5   1  57  57  14   3   0   8   18 0.3314
```

144 avg x

```
ggplot(tophitters2001, aes(avg)) + # x avg                      y
  geom_histogram() #
```

0.28 median mode

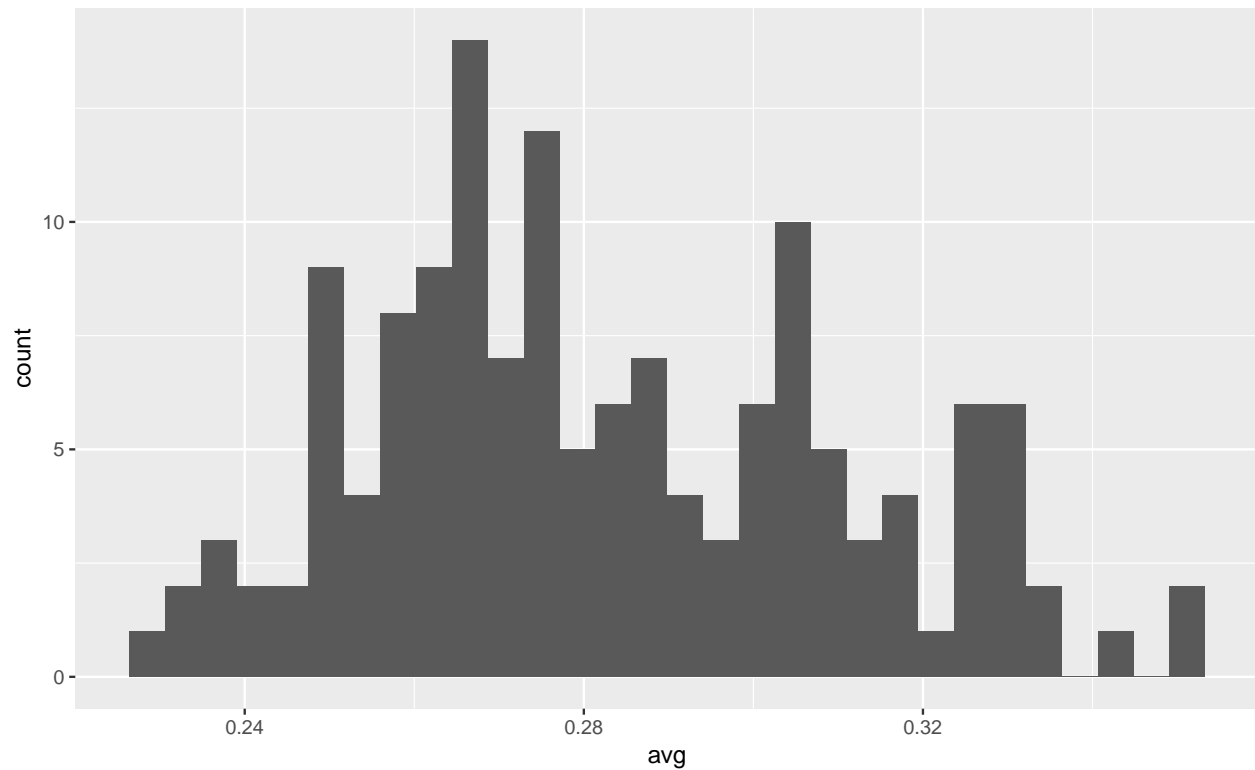


Figure 4.1: 2001 MLB 144

```
bin      1 1
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
0.001 0.01 0.1
```

```
ggplot(tophitters2001, aes(avg)) +
  geom_histogram(binwidth = 0.001) # binwidth = ...
```

```
ggplot(tophitters2001, aes(avg)) +
  geom_histogram(binwidth = 0.01) # binwidth = ...
```

```
ggplot(tophitters2001, aes(avg)) +
  geom_histogram(binwidth = 0.1) # binwidth = ...
```

4.1.2

```
tophitters2001
```

- lg AL or NL
- ```
fill lg
```

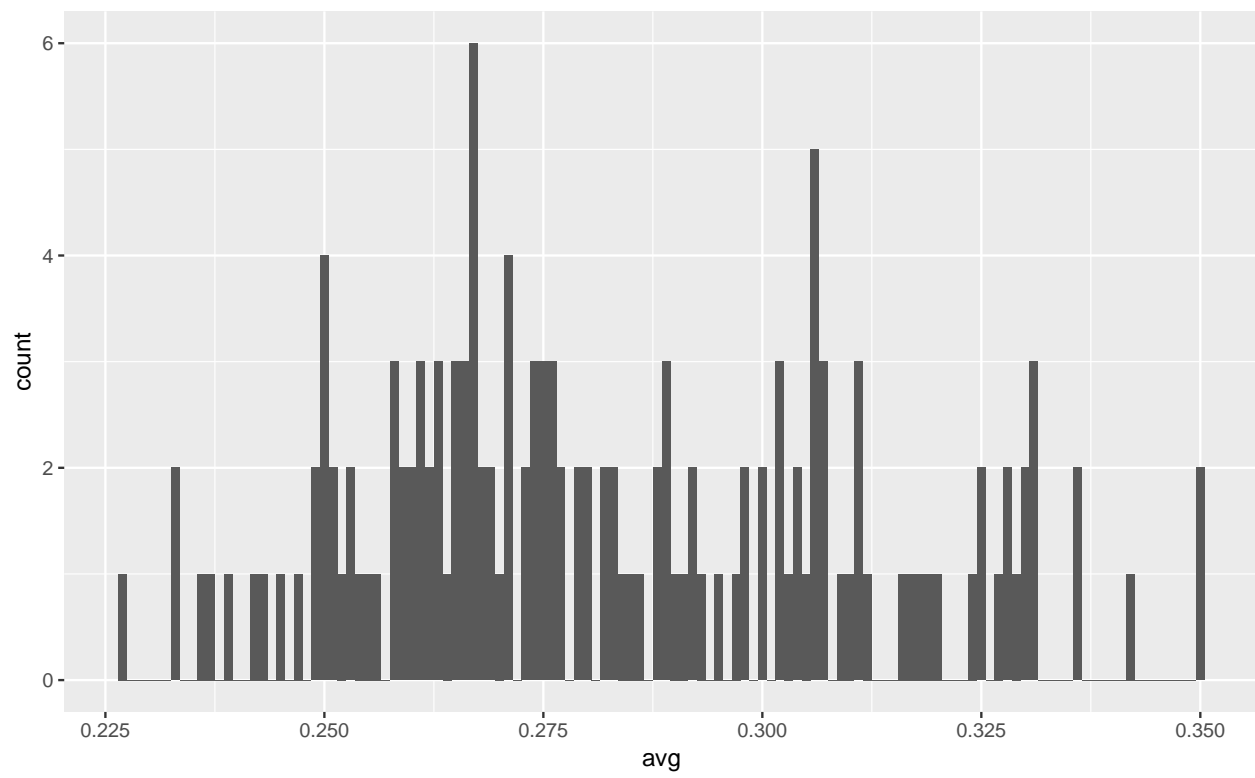


Figure 4.2:

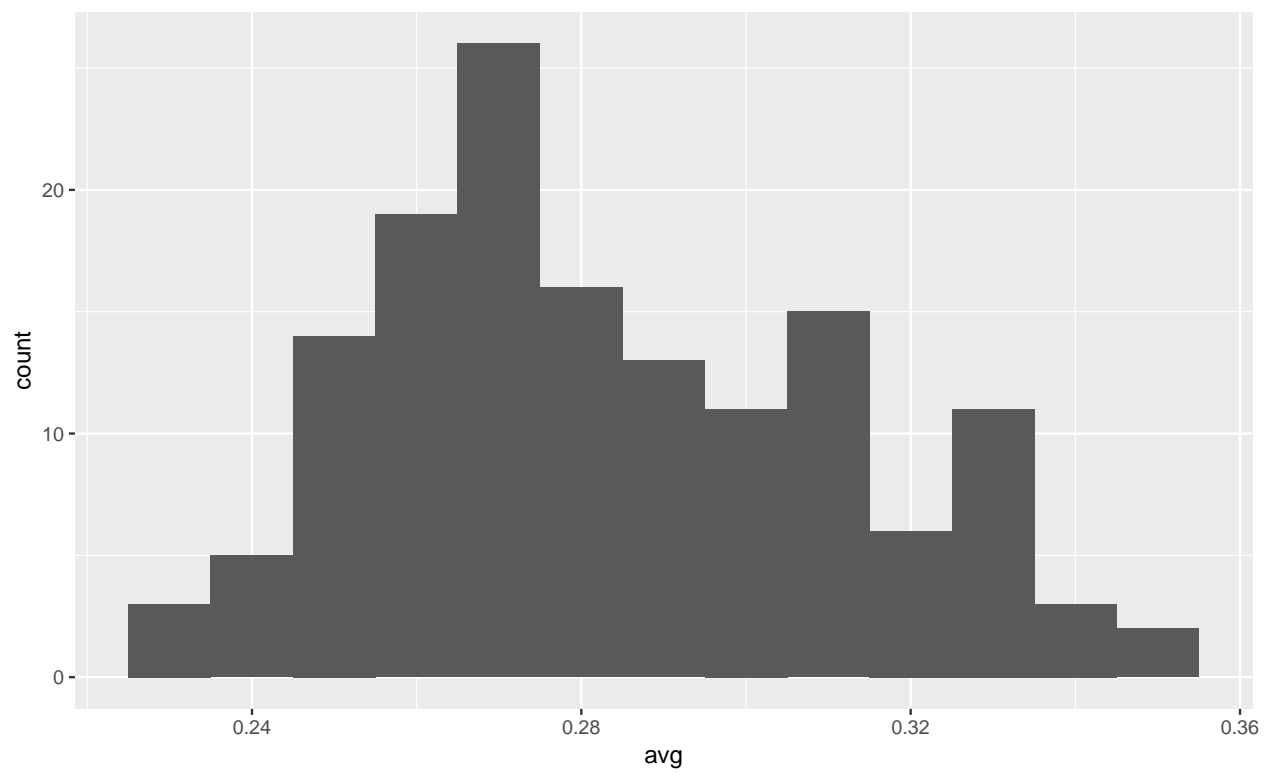


Figure 4.3:

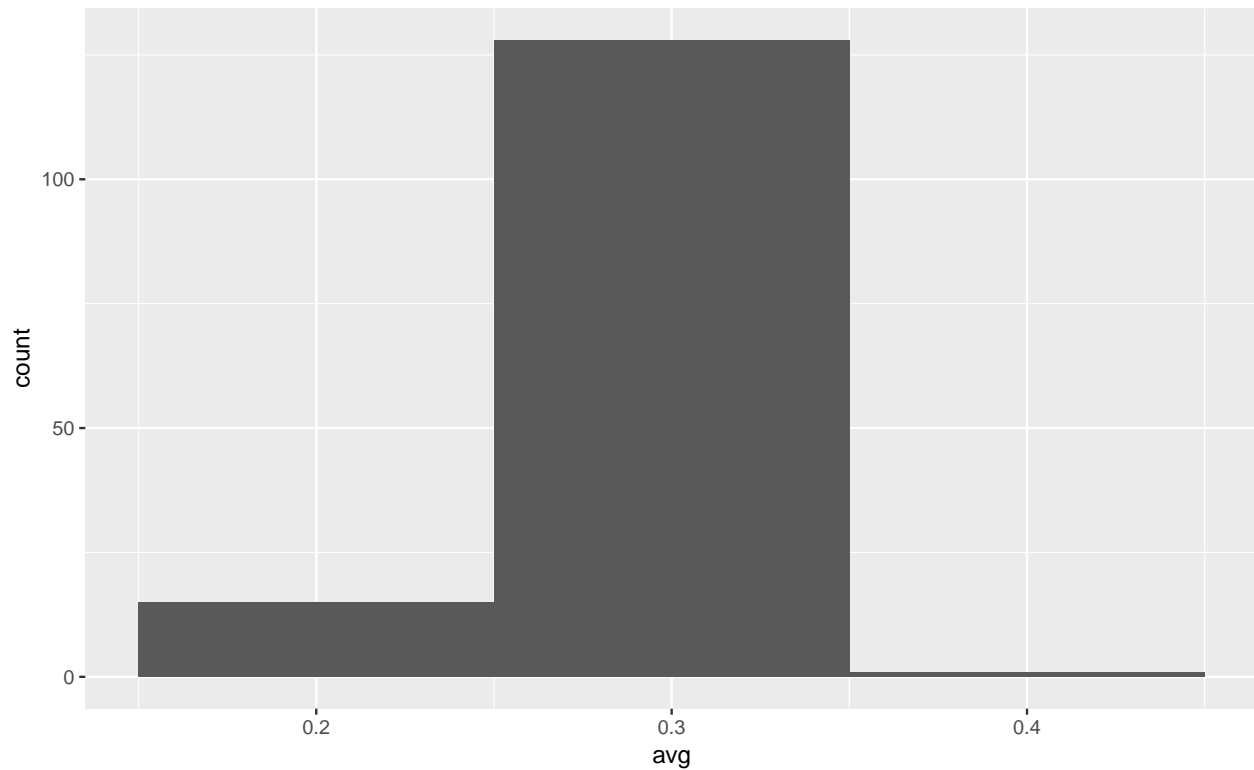


Figure 4.4:

```
ggplot(tophitters2001, aes(avg, fill = lg)) + # x avg fill
 geom_histogram(position = "identity") # position = "identity"
```

```
 position = "identity" geom_histogram() position = "stack"
= "identity"
```

```
alpha 0 1
```

```
ggplot(tophitters2001, aes(avg, fill = lg)) +
 geom_histogram(position = "identity", alpha = 0.7) # 0 1
```

## 4.2

density plot

raw data

kernel-density estimation

geom\_histogram()

geom\_density()

```
ggplot(tophitters2001, aes(avg, fill = lg)) + # x avg fill
 geom_density(alpha = 0.7) # position = "identity" geom_histogram(position = "identity")
```

```
y 5 10
```

```
1
```

```
geom_histogram() binwidth
```

```
geom_density() adjust
```

```
1, 4
```

```
(ref:density-plot-2)
```

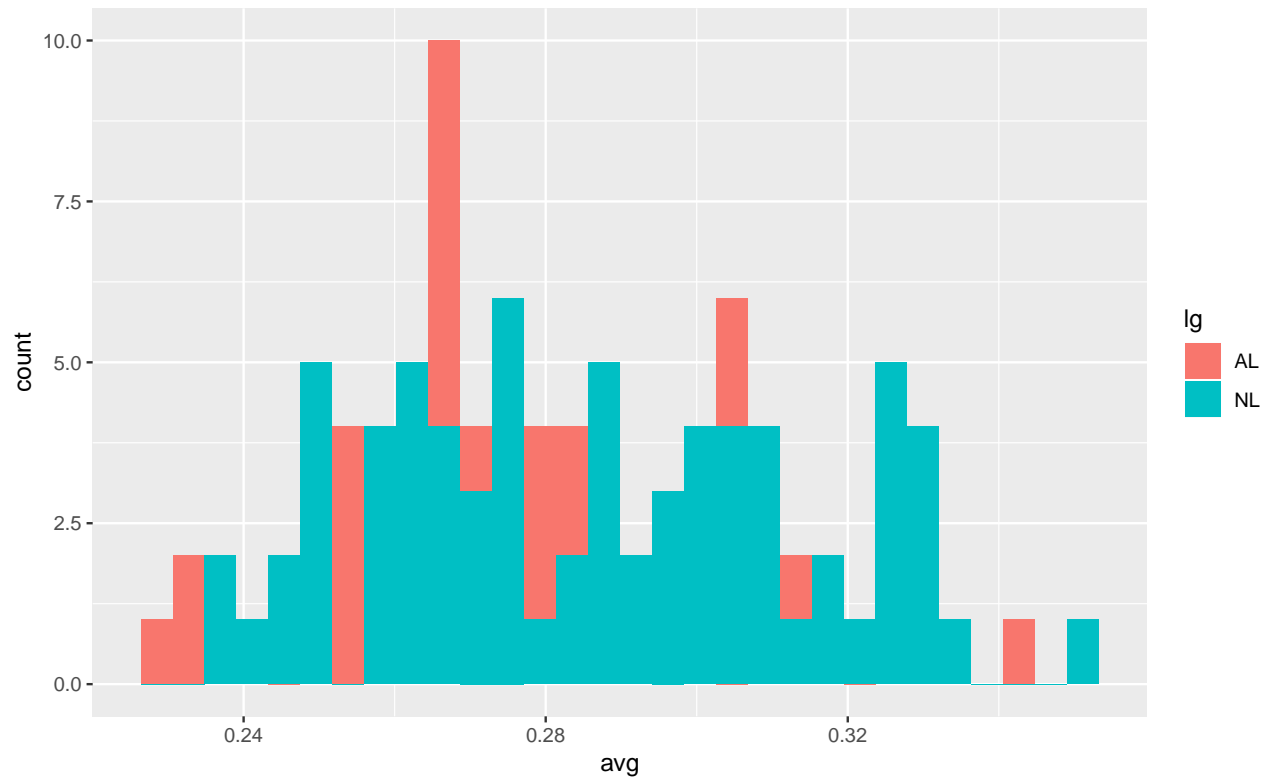


Figure 4.5:

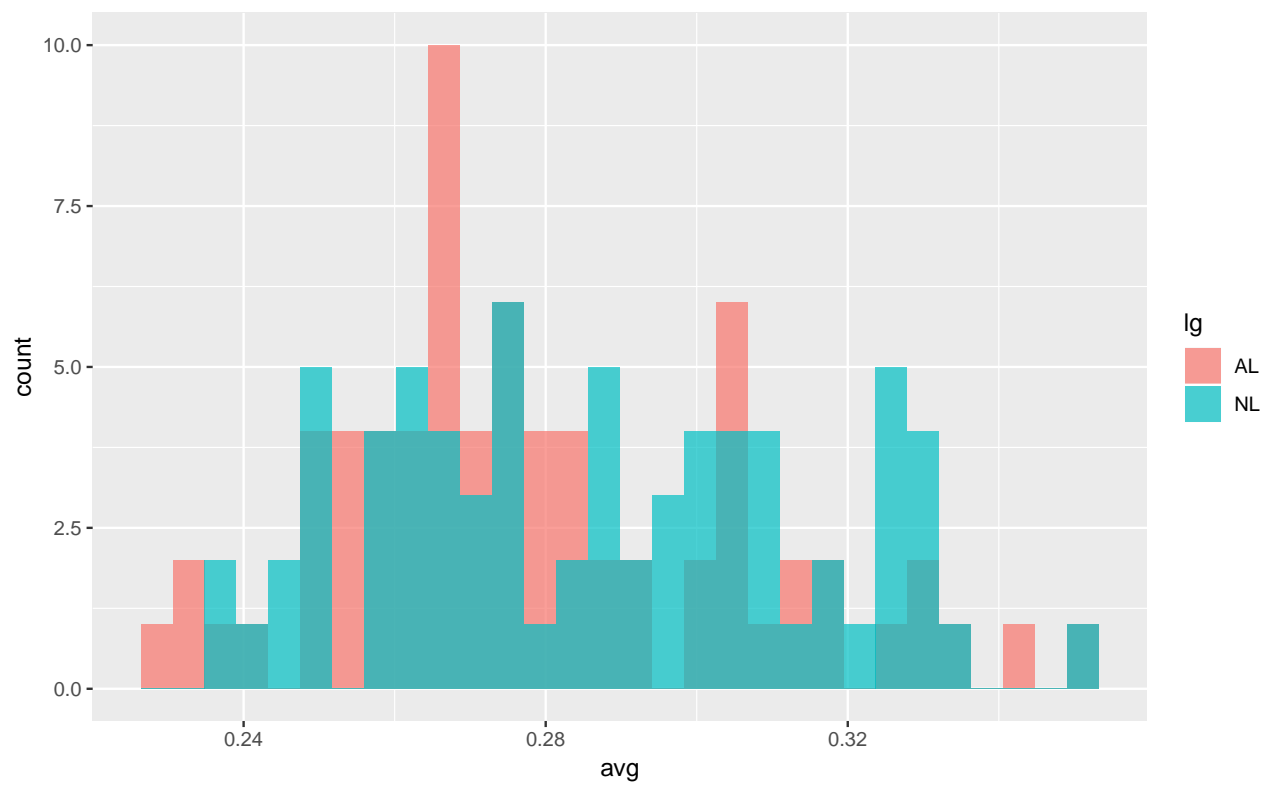


Figure 4.6: Figure 4.5

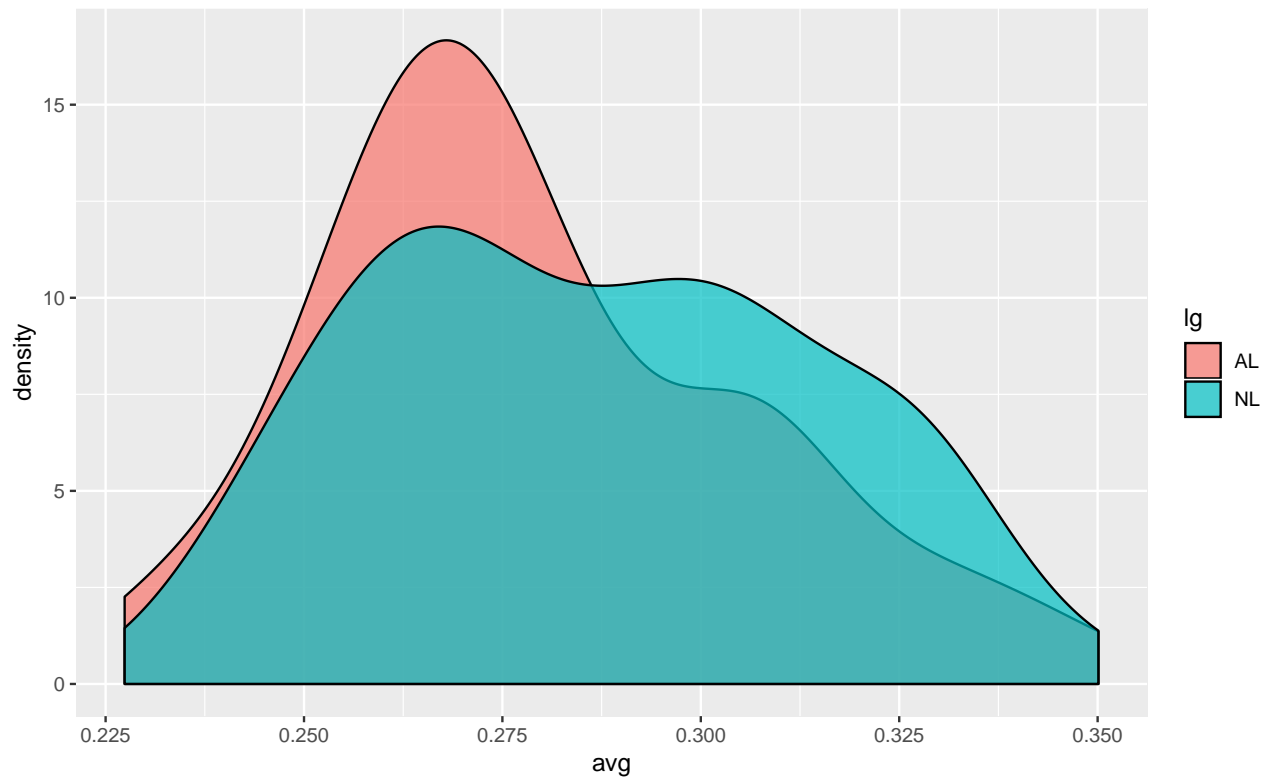


Figure 4.7:

```
ggplot(tophitters2001, aes(avg, fill = lg)) +
 geom_density(alpha = 0.7, adjust = 0.25) # adjust = ...
```

```
ggplot(tophitters2001, aes(avg, fill = lg)) +
 geom_density(alpha = 0.7, adjust = 1) # adjust = ...
```

(ref:density-plot-4)

```
ggplot(tophitters2001, aes(avg, fill = lg)) +
 geom_density(alpha = 0.7, adjust = 4) # adjust = ...
```

## Chapter 2 **ggplot2**

```
ggplot(tophitters2001, aes(avg, y = ..density..)) + # y = ..density..
 geom_histogram(binwidth = 0.01) + #
 geom_density(alpha = 0.7) #
```

y = ..density..      geom\_density()    y      geom\_histogram()      y      2      y

## 4.3

box plot; box-and-whisker plot

R      PlantGrowth



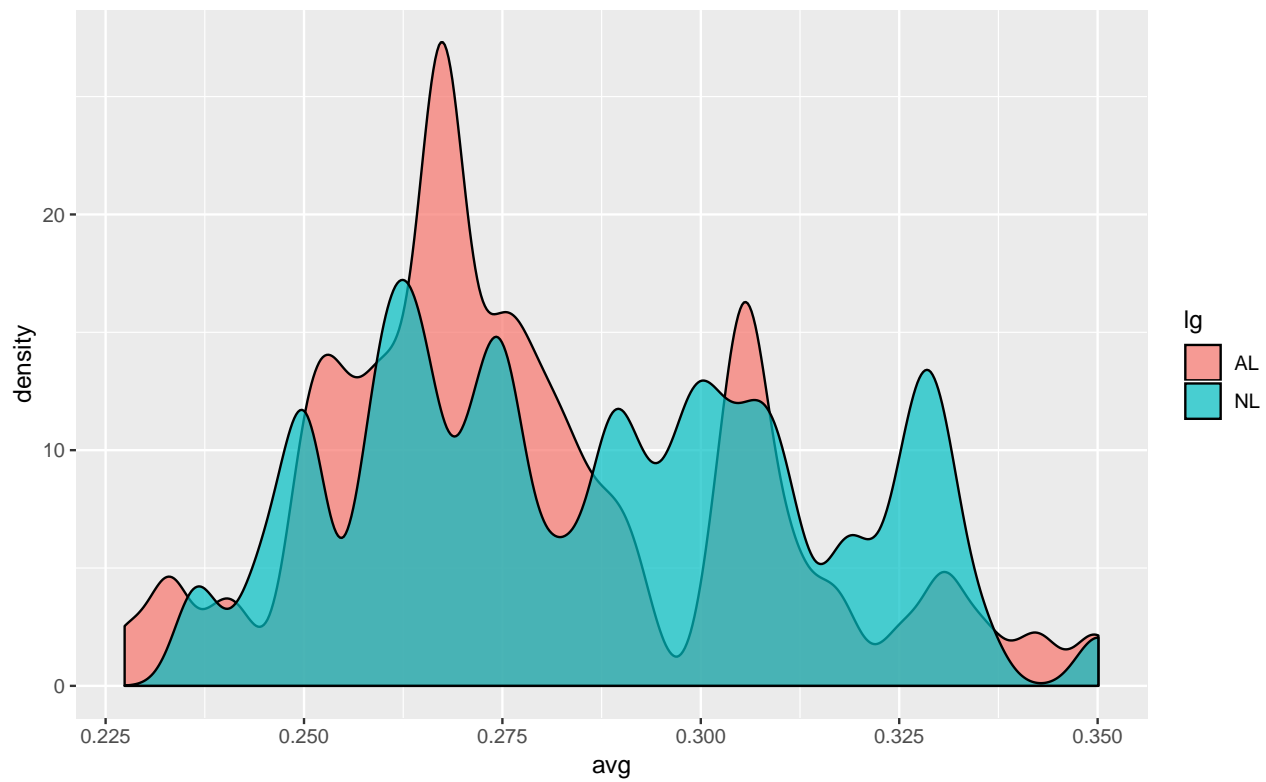


Figure 4.8: (ref:density-plot-2)

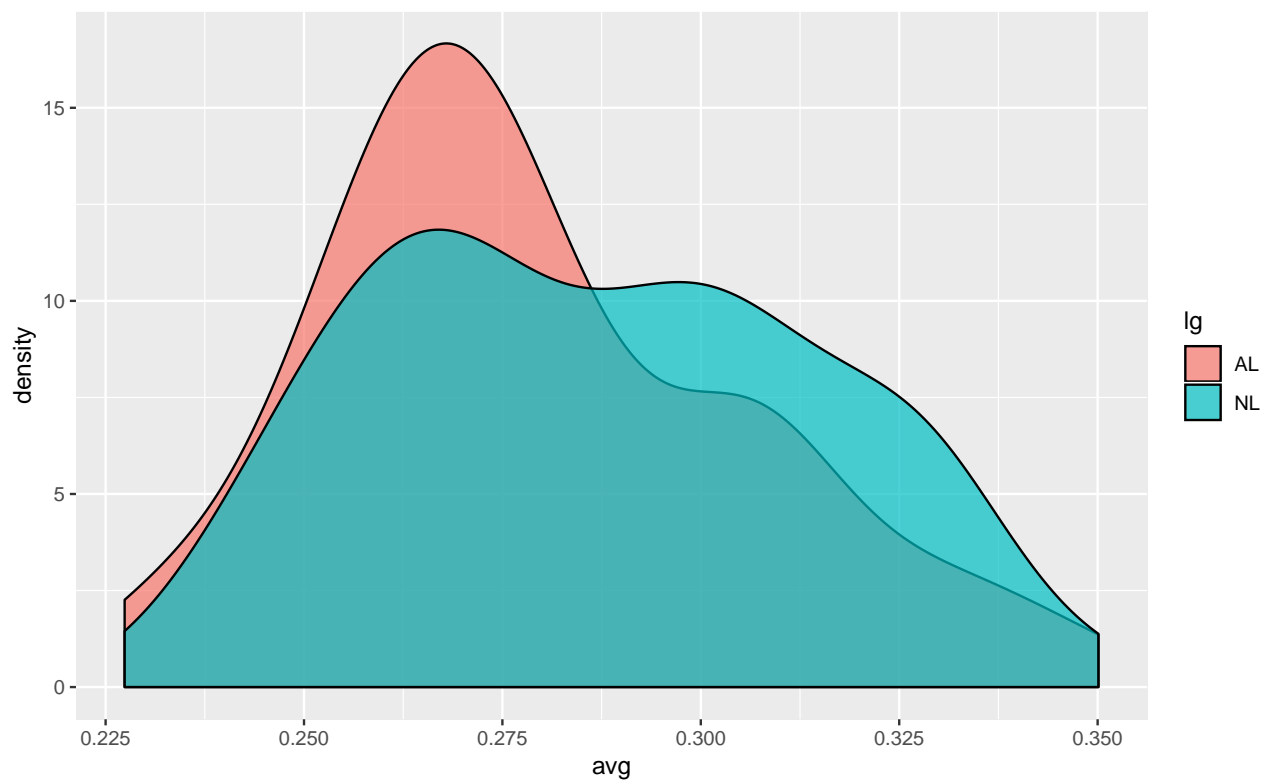


Figure 4.9:

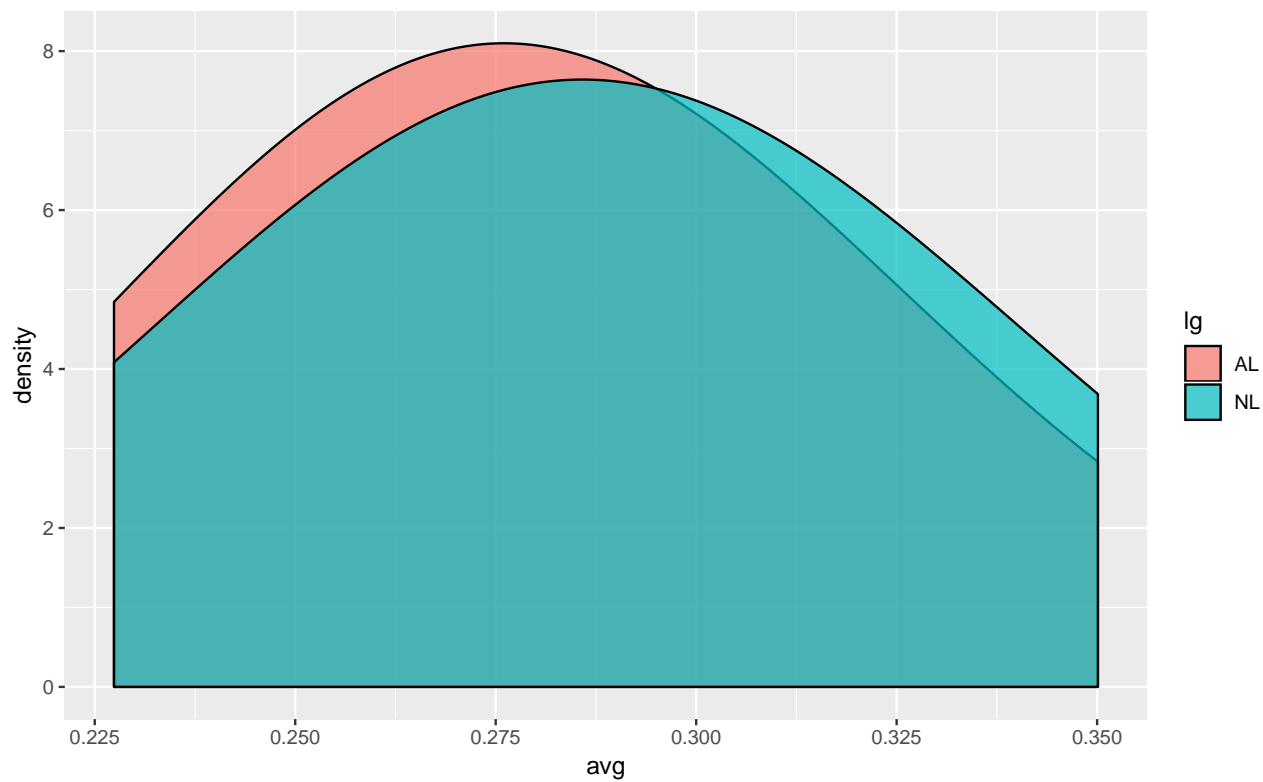


Figure 4.10: (ref:density-plot-4)

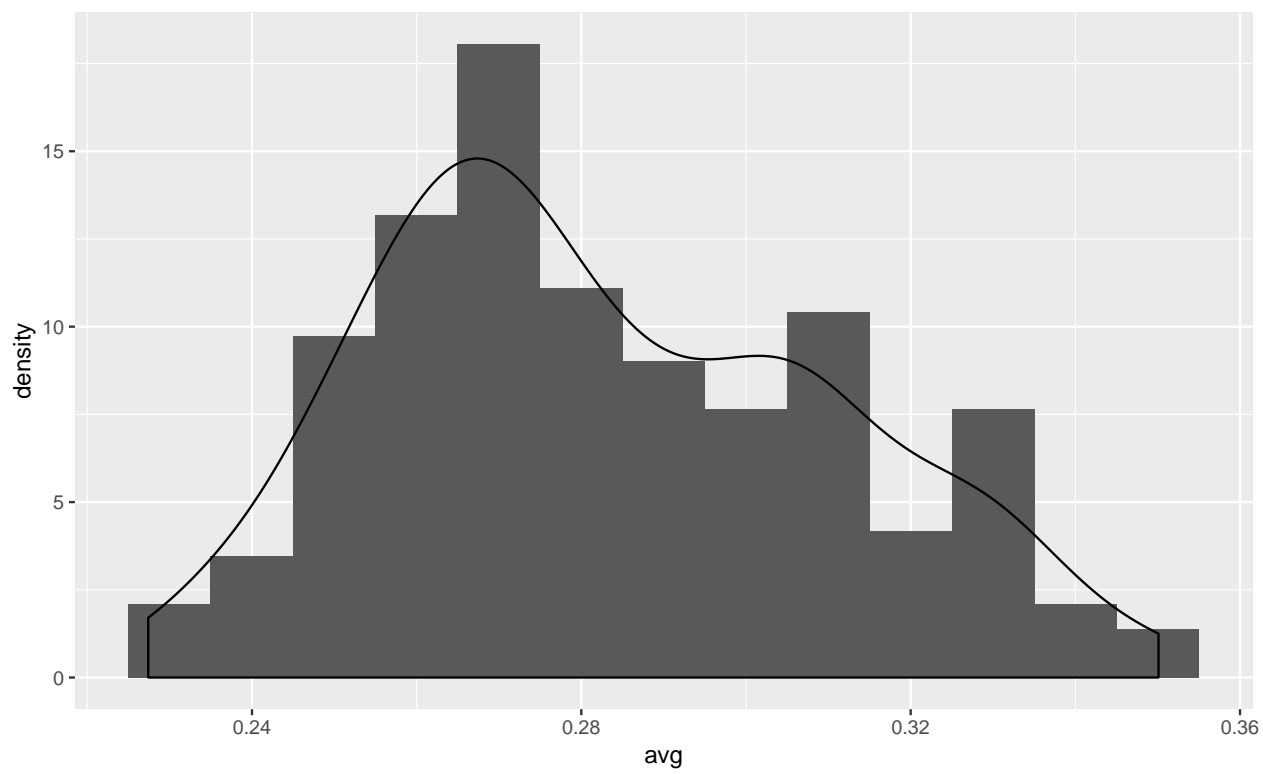


Figure 4.11:

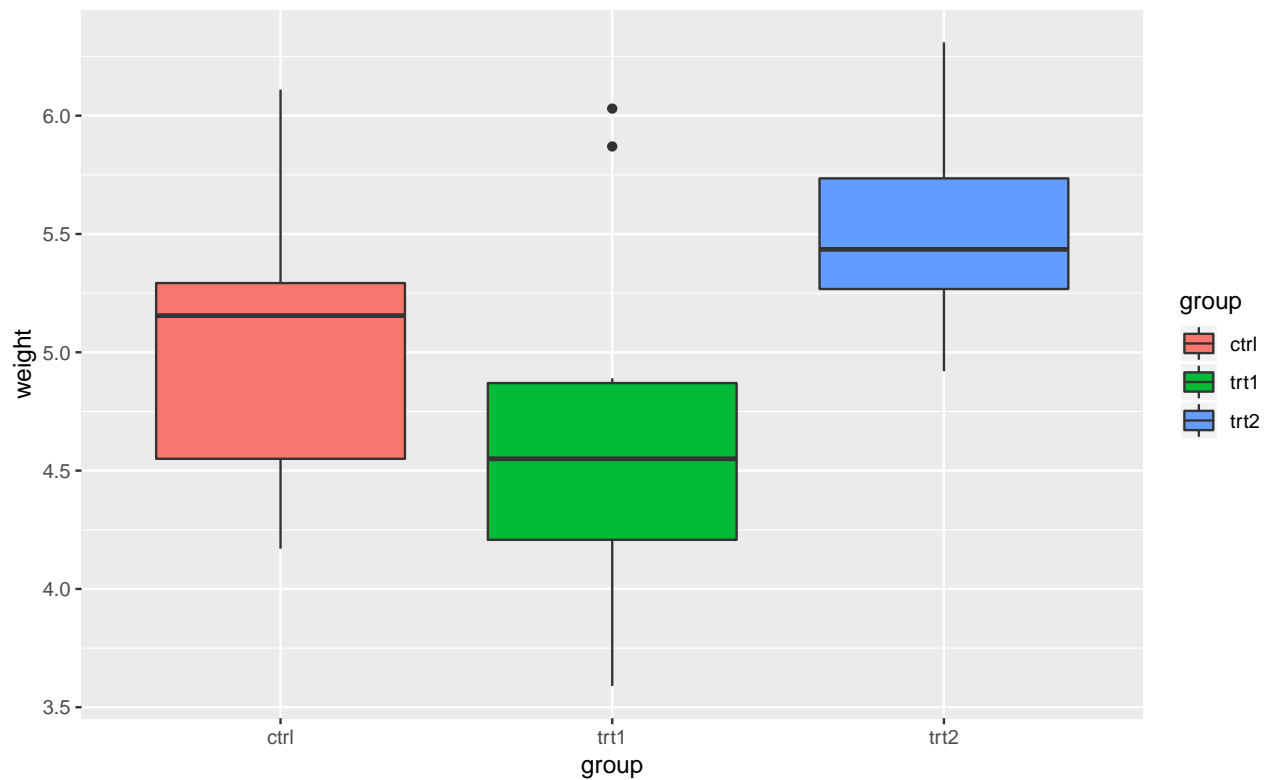


Figure 4.12: x y

```
head(PlantGrowth) # 6
```

```
weight group
1 4.17 ctrl
2 5.58 ctrl
3 5.18 ctrl
4 6.11 ctrl
5 4.50 ctrl
6 4.61 ctrl
```

```
2
```

- group ctrl, trt1, or trt2
- weight

```
3
```

```
fill
```

```
ggplot(PlantGrowth, aes(group, weight, fill = group)) + # x group y weight fill
 geom_boxplot() #
```

```
1
```

- interquartile range, IQR 25 75
- $\text{IQR} \times 1.5$

```
1
```

```
Tukey
```

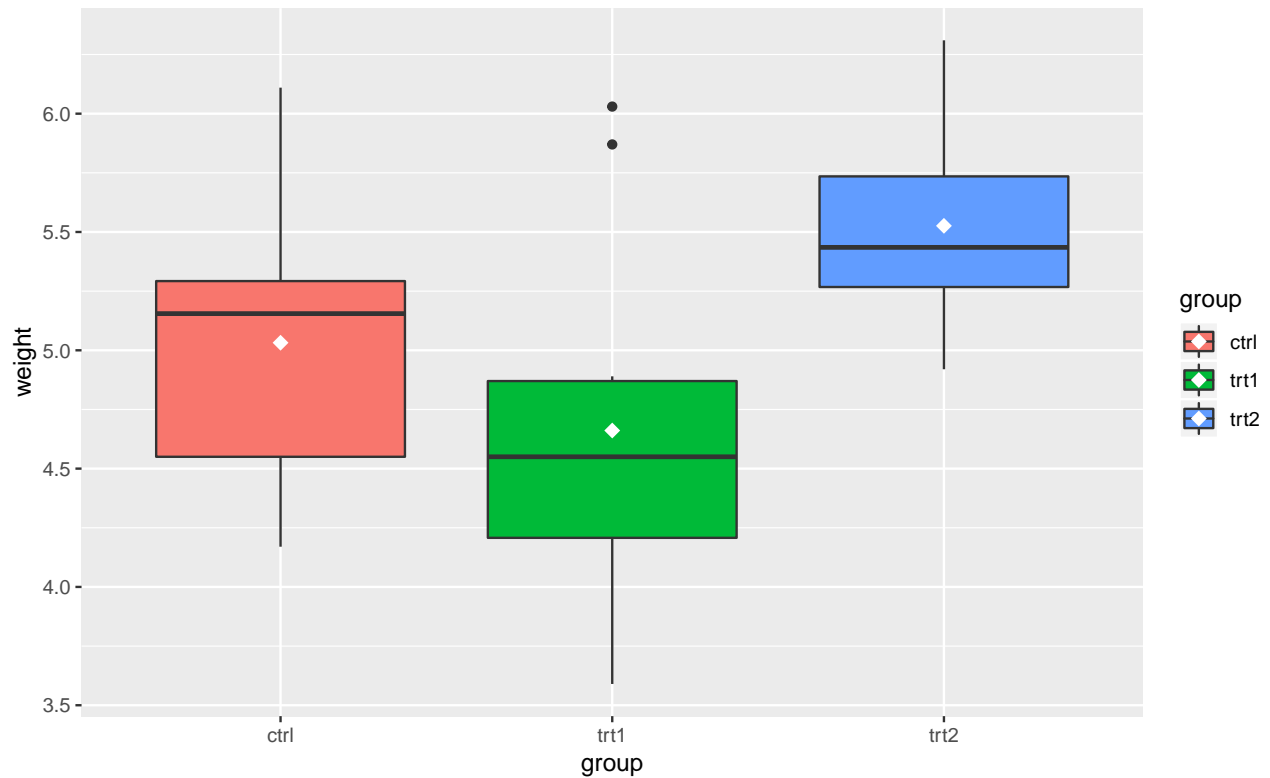


Figure 4.13:

- 50 percentile
- outlier  $IQR \times 1.5$

Wikipedia “Interquartile range” [https://en.wikipedia.org/wiki/Interquartile\\_range](https://en.wikipedia.org/wiki/Interquartile_range)

2

```
ggplot(PlantGrowth, aes(group, weight, fill = group)) +
 geom_boxplot() +
 stat_summary(geom = "point", fun.y = mean, color = "white", shape = "diamond", size = 3) #
```

```
stat_summary() y weight 3
```

## 4.4

violin plot 90°

geom\_violin()

(ref:violin-plot) x y

```
ggplot(PlantGrowth, aes(group, weight, fill = group)) + # x group y weight fill
 geom_violin() #
```

geom\_density() geom\_violin() adjust

2

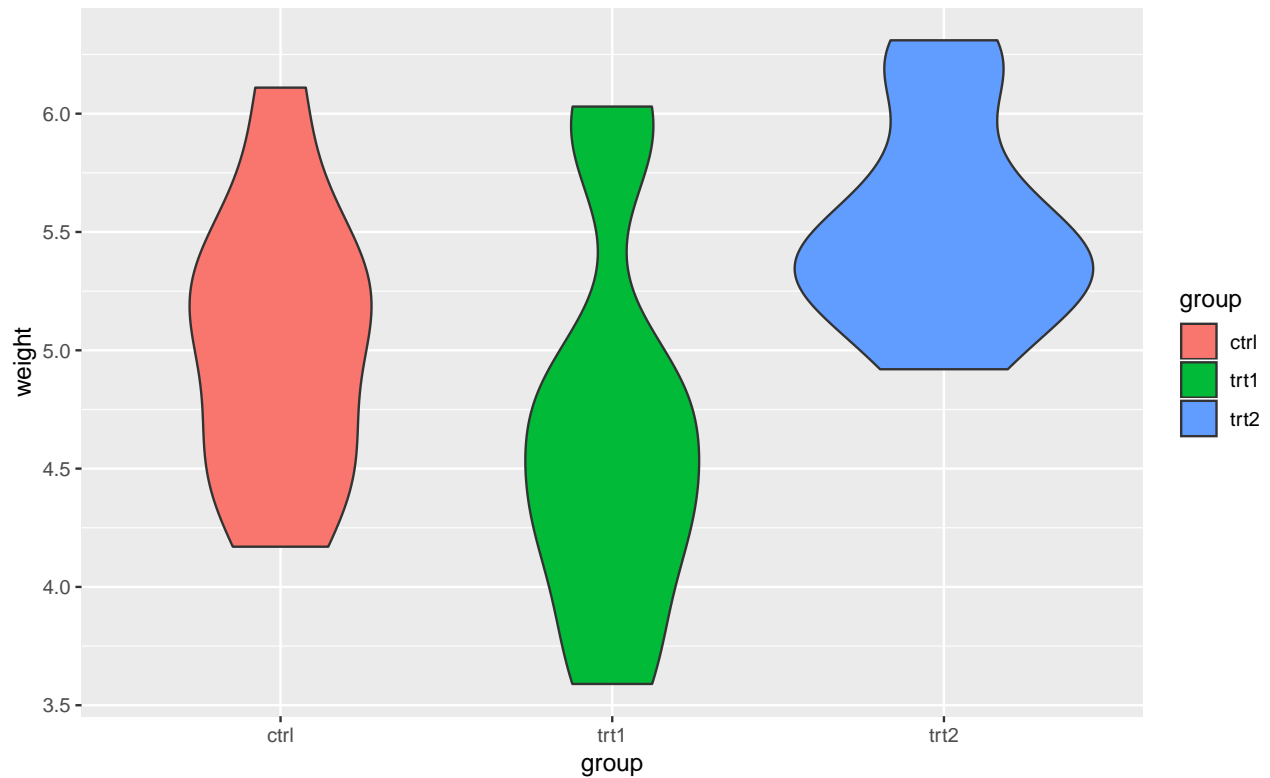


Figure 4.14: (ref:violin-plot)

## 4.5

strip plot

1 1

(ref:strip-plot)

```
ggplot(PlantGrowth, aes(group, weight, fill = group)) + # x group y weight fill
 geom_violin() + #
 geom_boxplot(outlier.shape = NA, width = 0.2) + #
 geom_point(position = position_jitter(width = 0.1, height = 0, seed = 1)) + #
 stat_summary(geom = "point", fun.y = mean, shape = "diamond", color = "white", size = 3) #
```

1 1

```
0. ggplot(PlantGrowth, aes(group, weight, fill = group))
```

```
 • x group y weight group
```

```
1. geom_violin()
```

```
 •
```

```
2. geom_boxplot(outlier.shape = NA, width = 0.2)
```

```
 •
```

```
 • outlier.shape = NA
```

```
 • width 1
```

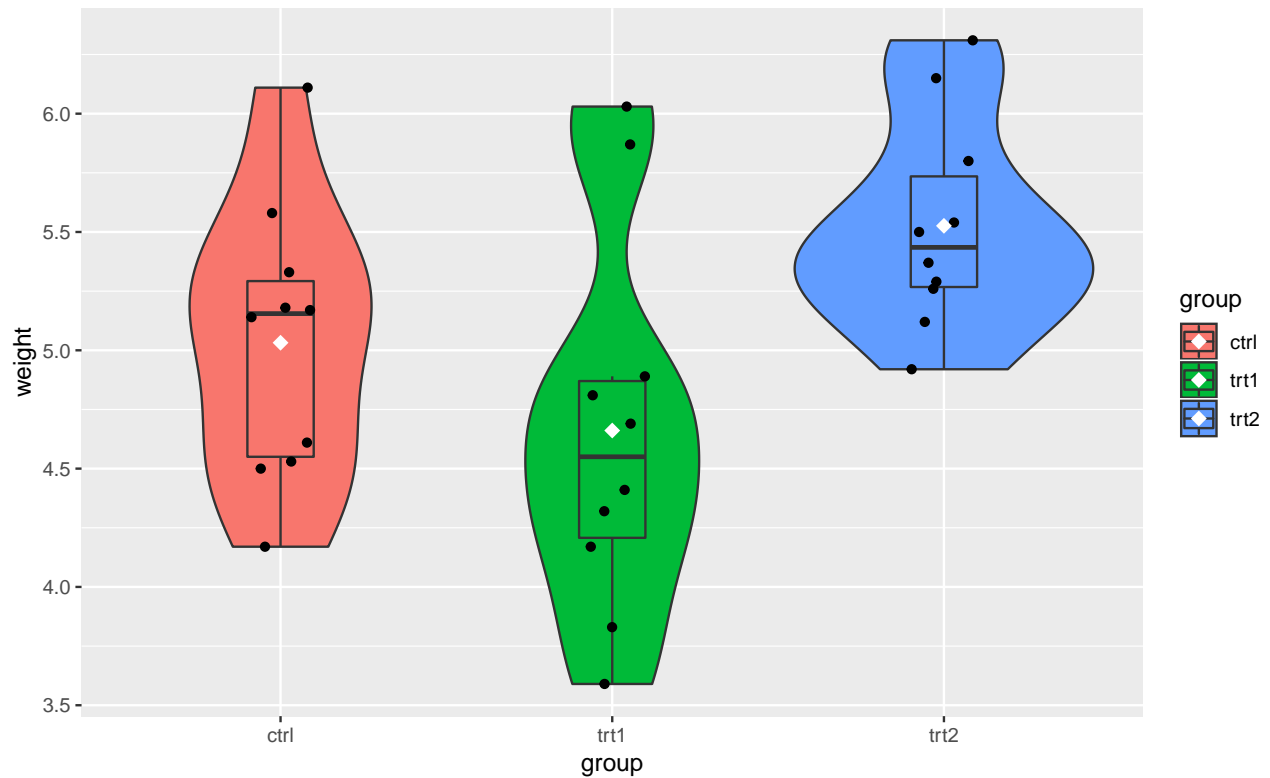


Figure 4.15: (ref:strip-plot)

```
3. geom_point(position = position_jitter(width = 0.1, seed = 1))
 •
 • position = position_jitter() jitter position
 • width x seed seed
 • y height 0
4. stat_summary(geom = "point", fun.y = mean, shape = "diamond", color = "white", size
 = 3)
 •
```

## 4.6

- R iris <sup>3</sup> Species Sepal.Length
- Species Sepal.Width

3

[https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set)

# Chapter 5

scatter plot

```
library(ggplot2) #
```

## 5.1

R faithful

1

```
head(faithful) # 6
```

```
eruptions waiting
1 3.600 79
2 1.800 54
3 3.333 74
4 2.283 62
5 4.533 85
6 2.883 55
```

- eruptions
- waiting

geom\_point()

(ref:scatter-plot-1) faithful 2

```
ggplot(faithful, aes(eruptions, waiting)) + # x eruptions y waiting
 geom_point() #
```

## 5.2

Chapter 2

R

mtcars

Chapter 2

geom\_point()

color

```
ggplot(mtcars, aes(wt, mpg, color = factor(cyl))) + # factor()
 geom_point()
```

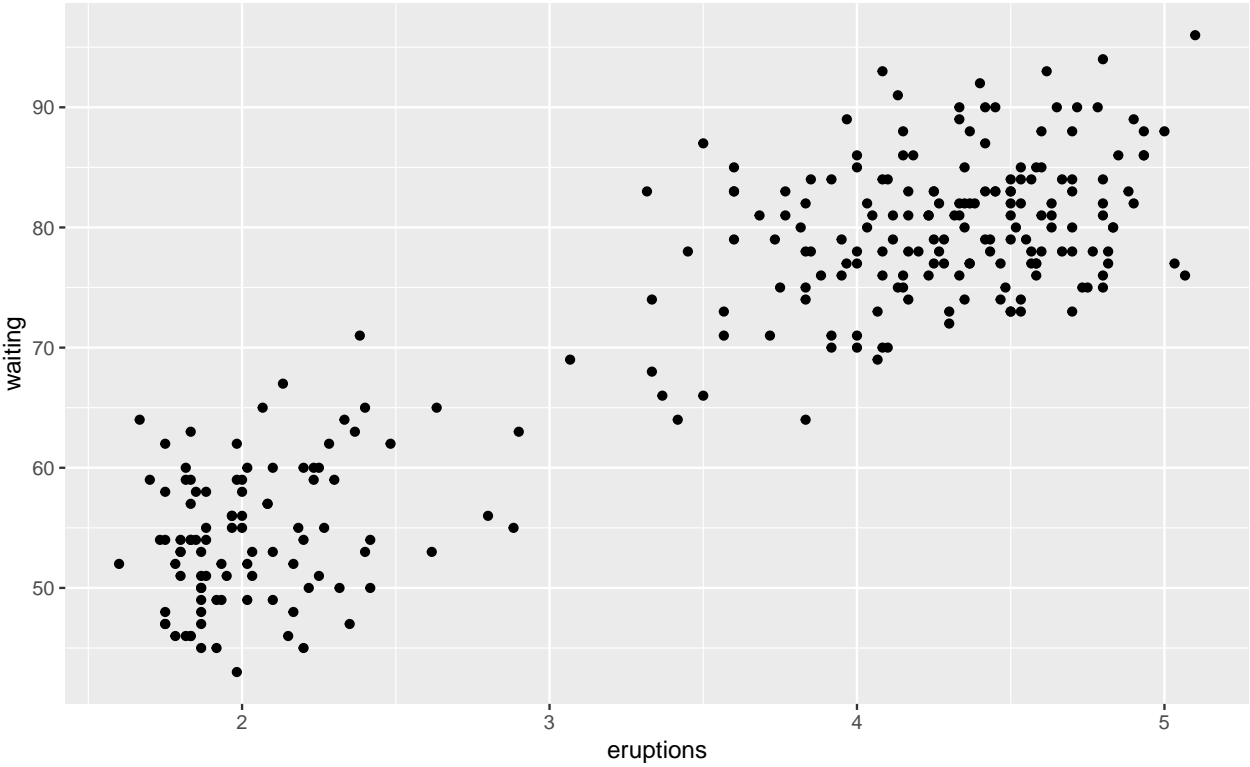


Figure 5.1: (ref:scatter-plot-1)

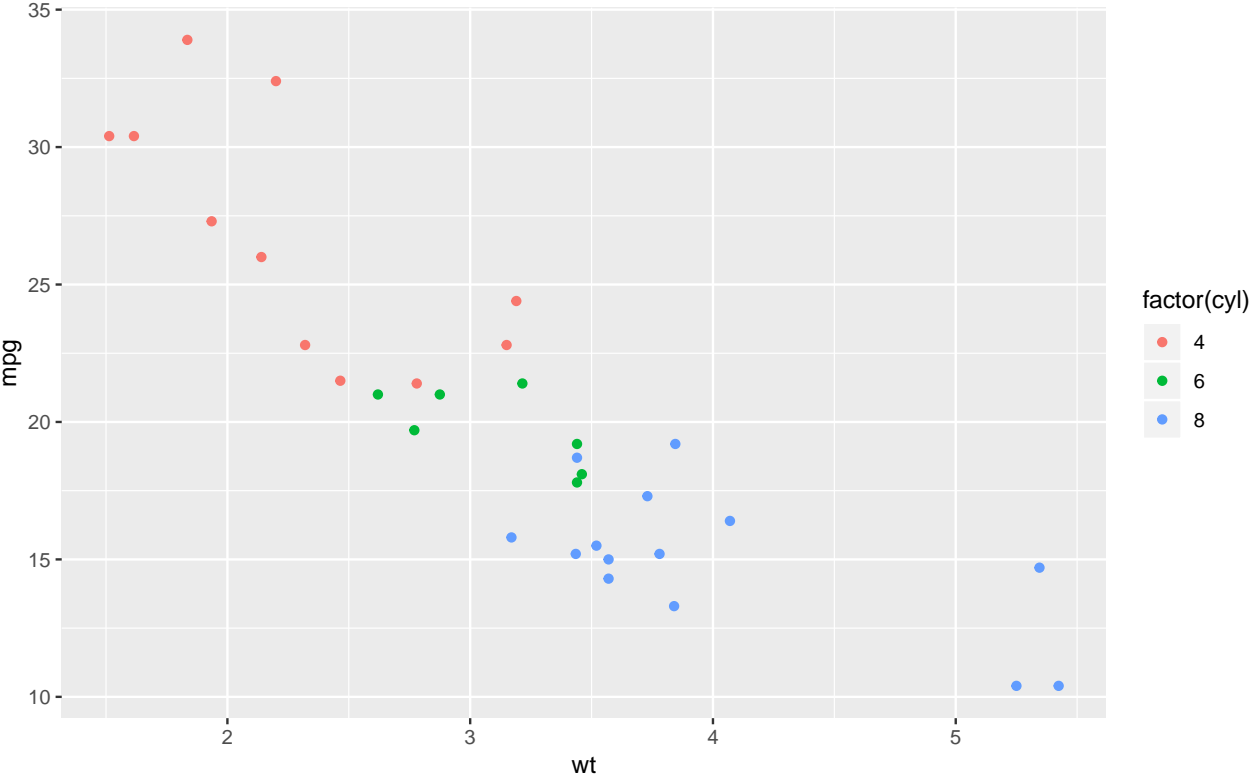


Figure 5.2:



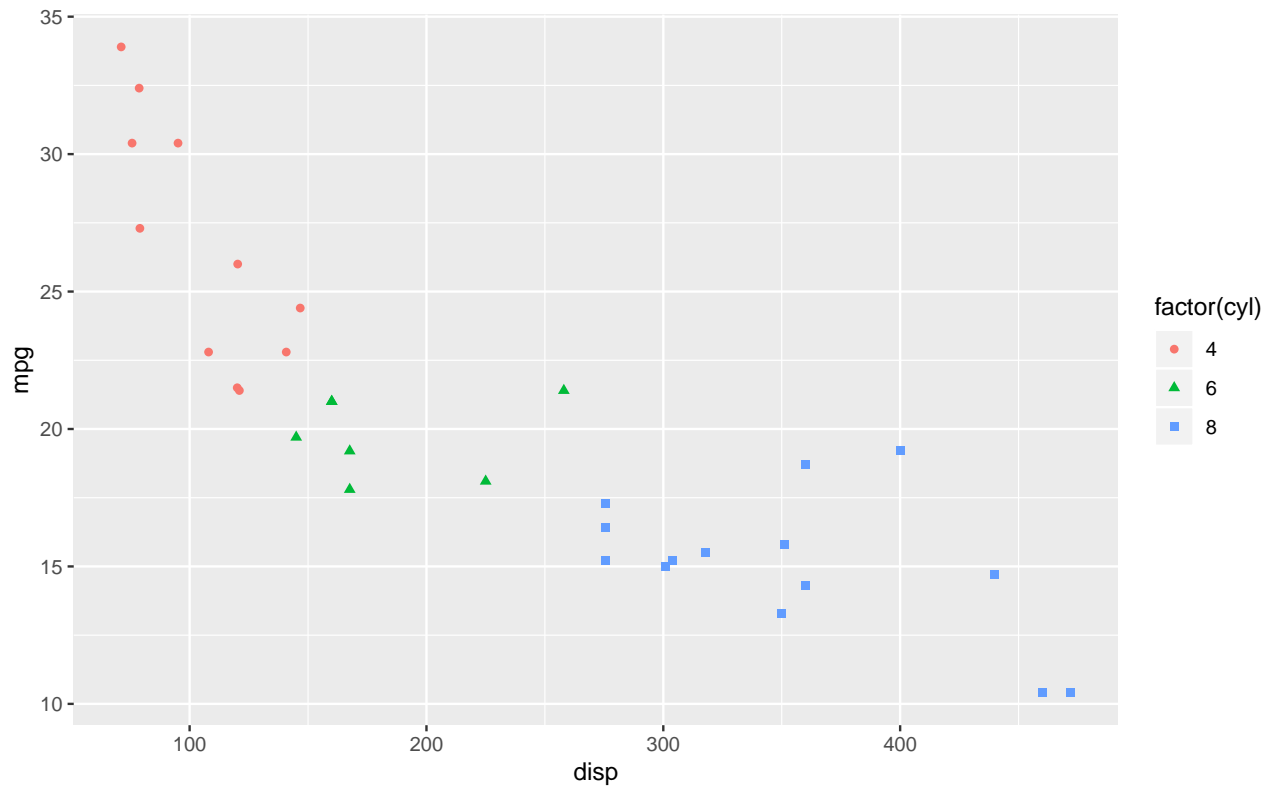


Figure 5.3: Figure 5.2

```

2 shape
ggplot(mtcars, aes(displ, mpg, color = factor(cyl), shape = factor(cyl))) + # shape = factor(year)
 geom_point()

```

cyl

## 5.3

```

faithful 2 1
3 GGally ggpairs() R iris GGally
library(GGally) #
ggpairs()
(ref:scatter-plot-matrix-1) ggpairs() iris
ggpairs(iris) #

```

Species ggplot2

<sup>1</sup><https://ja.wikipedia.org/wiki/>

<sup>2</sup>redundant coding

<sup>3</sup>

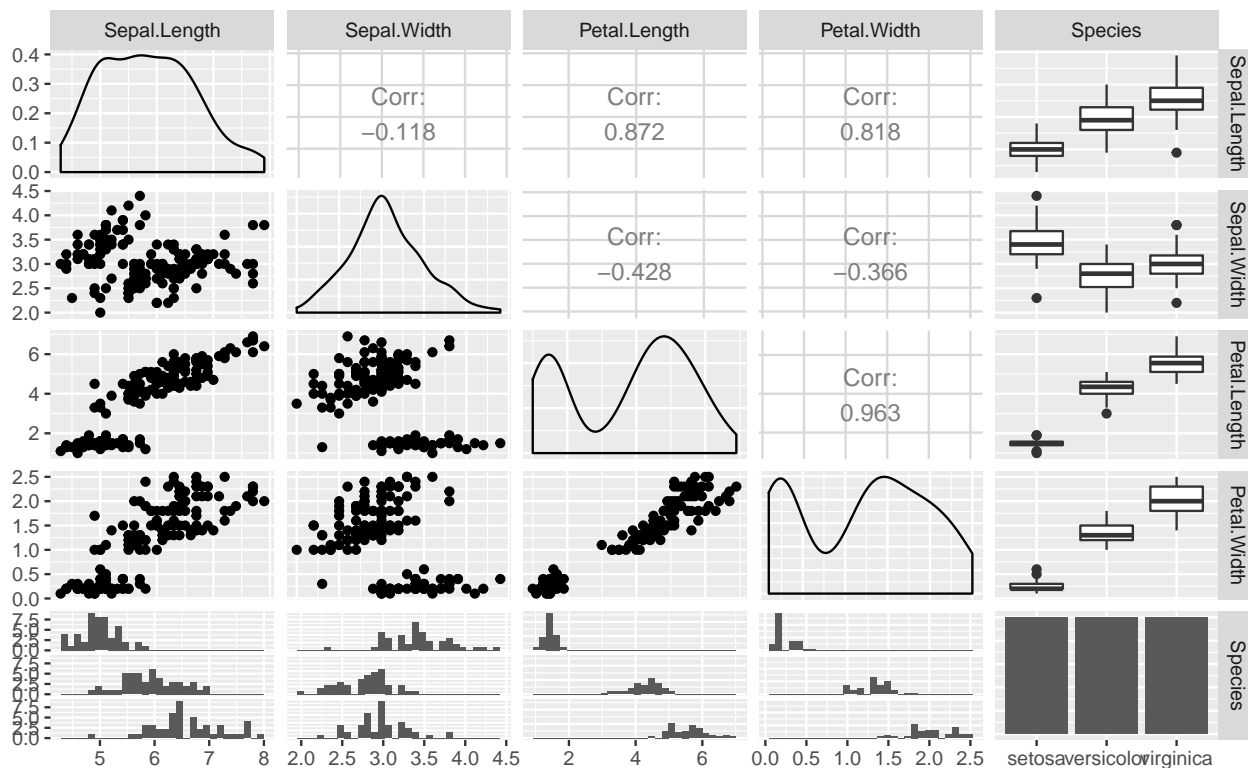


Figure 5.4: (ref:scatter-plot-matrix-1)

```
ggpairs(iris, aes(color = Species, alpha = 0.7)) # ggplot2
```

4

## 5.4

ggplot2 mpg

```
head(mpg) # 6
```

```
A tibble: 6 x 11
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto(~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manua~ f 21 29 p comp~
3 audi a4 2 2008 4 manua~ f 20 31 p comp~
4 audi a4 2 2008 4 auto(~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto(~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manua~ f 18 26 p comp~
```

- displ
- hwy

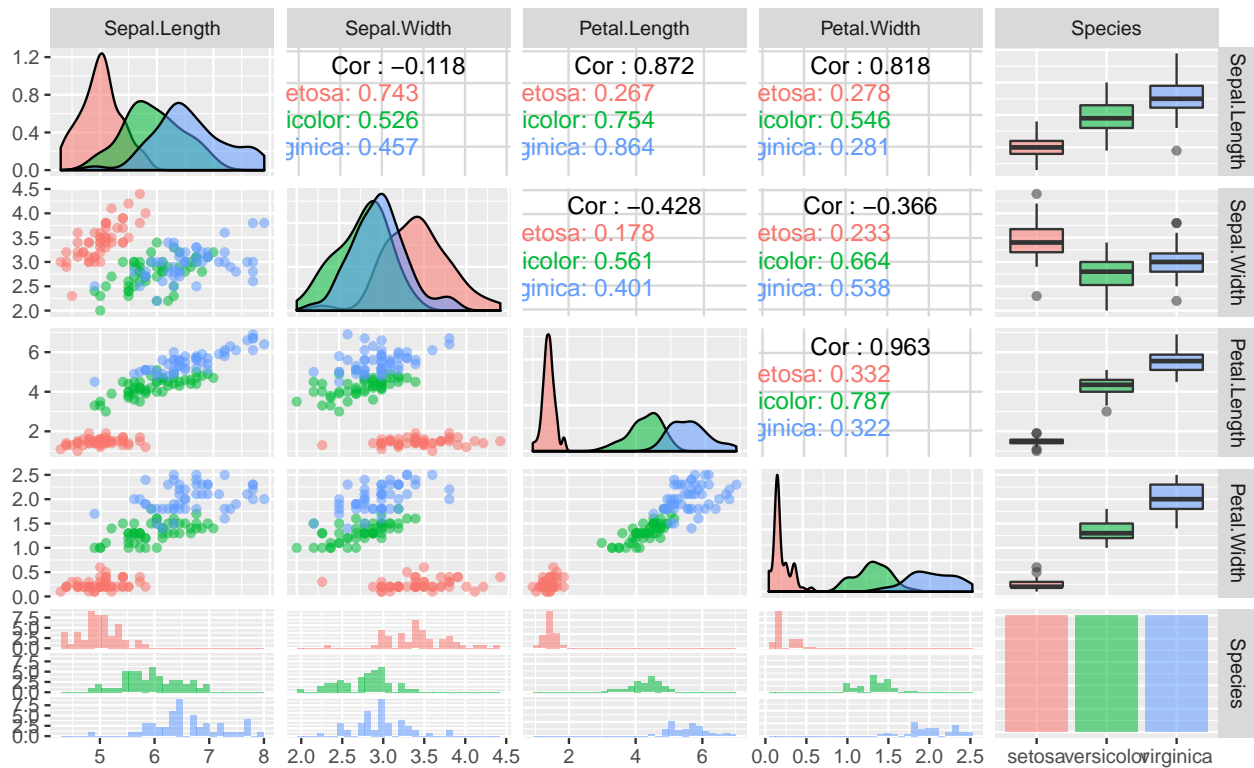


Figure 5.5: Figure 5.4

- cyl 4, 5, 6, or 8

```
2 x displ y hwy color factor(cyl)
```

```
ggplot(mpg, aes(displ, hwy, color = factor(cyl))) +
 geom_point()
```

```
mpg nrow()
nrow(mpg)
```

```
[1] 234
```

```
234 Figure 5.6
```

```
x y overplotting
```

```
7 Chapter 4 jittering
```

```
ggplot(mpg, aes(displ, hwy, color = factor(cyl))) +
 geom_point(position = position_jitter(width = 0.1, height = 0.4, seed = 1)) #
```

```
position = position_jitter(width = ..., height = ..., seed = ...) width height
```

```
ggplot(mpg, aes(displ, hwy, color = factor(cyl), shape = factor(cyl))) + # shape
 geom_point(position = position_jitter(width = 0.1, height = 0.4, seed = 1), alpha = 0.7) # alpha = 0.
```

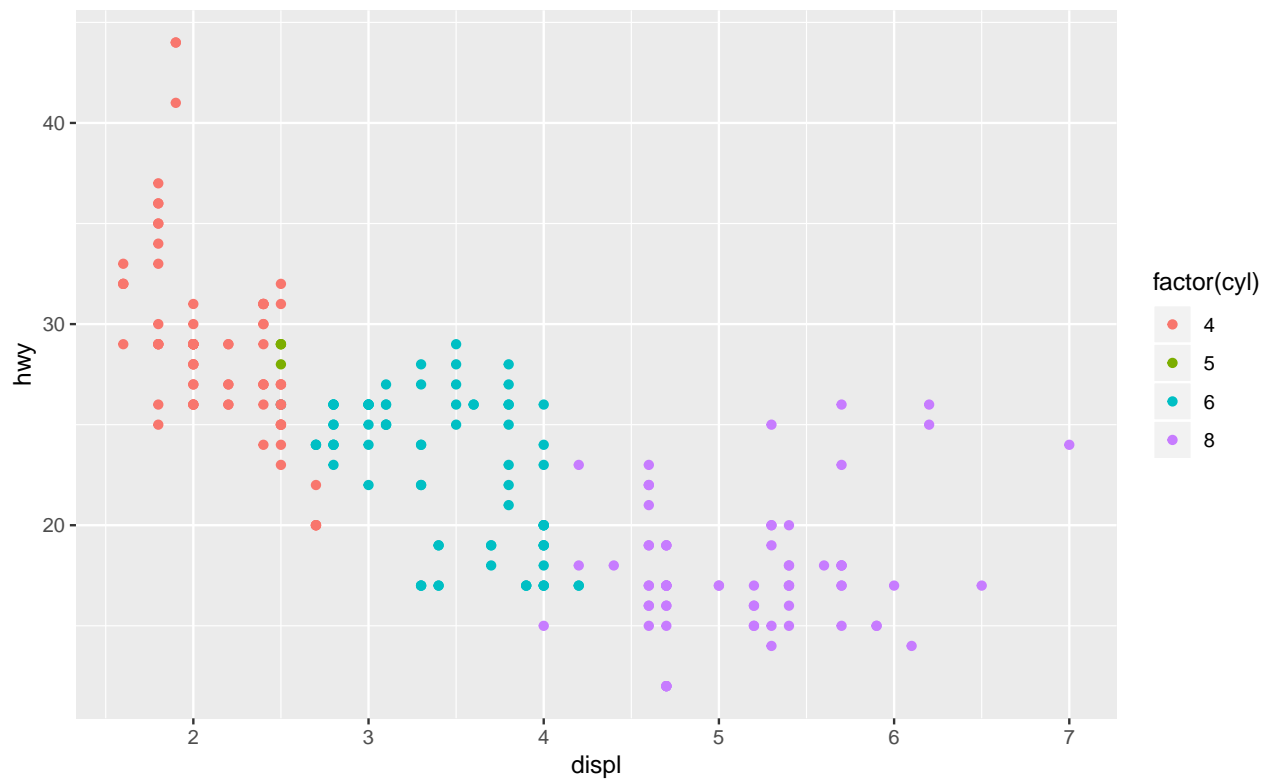


Figure 5.6:

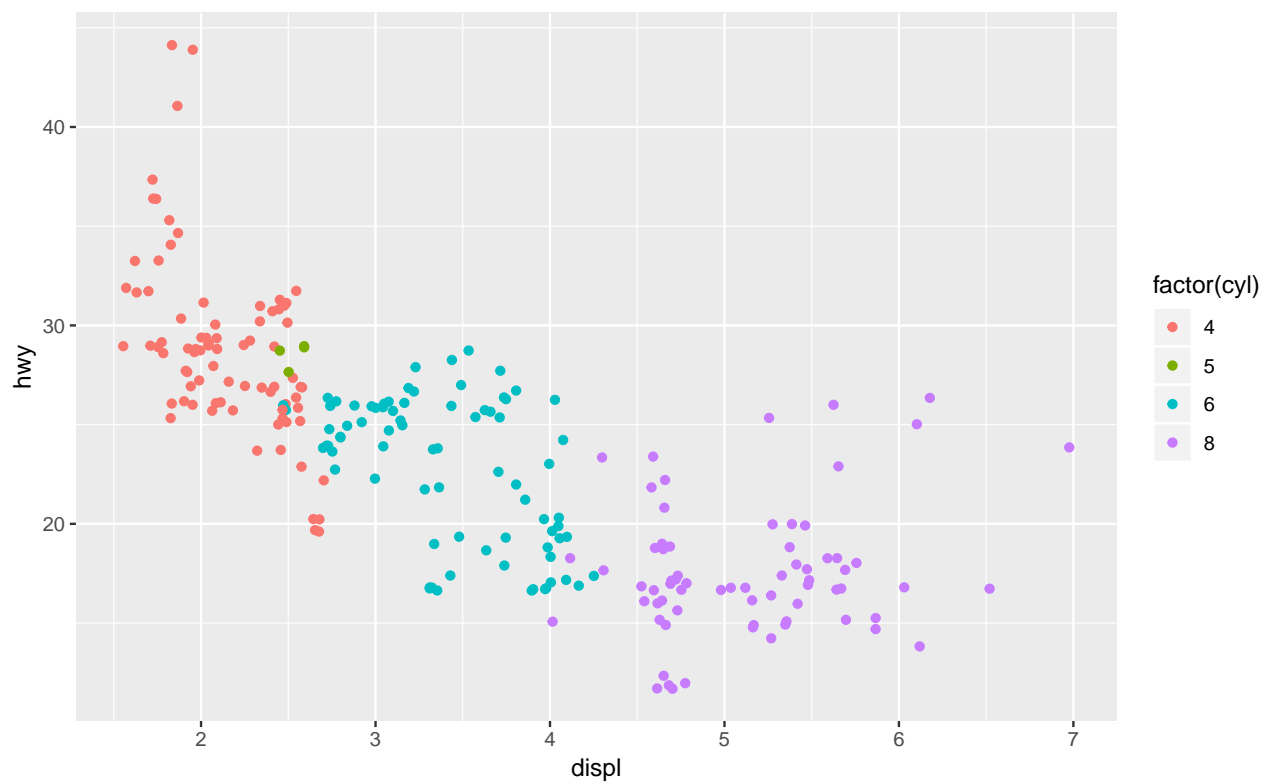


Figure 5.7:

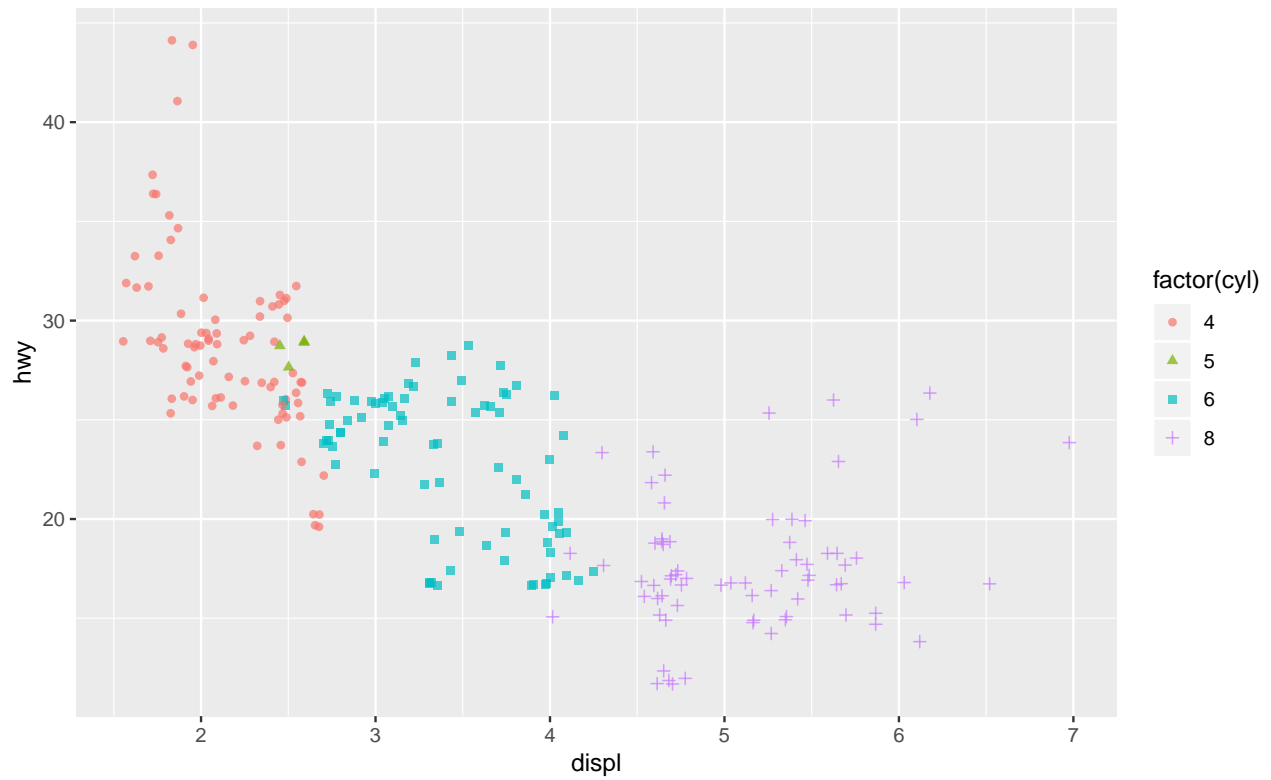


Figure 5.8: Figure 5.7

## 5.5

- R `trees` Girth Volume
- `gcookbook` heightweight



# Chapter 6

x line graph

```
library(ggplot2) #
```

## 6.1

gcookbook aapl Apple

```
library(gcookbook) #
```

```
head(aapl) # 6
```

```
date adj_price
1 1980-12-12 0.023268
2 1980-12-19 0.022863
3 1980-12-26 0.028731
4 1981-01-02 0.027921
5 1981-01-09 0.025797
6 1981-01-16 0.025089
```

- date
- adj\_price

```
geom_line()
```

```
ggplot(aapl, aes(date, adj_price)) + # x date y adj_price
 geom_line() #
```

R BOD

biochemical oxygen demand; BOD

<sup>1</sup>

```
head(BOD) # 6
```

```
Time demand
1 1 8.3
2 2 10.3
```

---

<sup>1</sup><https://ja.wikipedia.org/wiki/>

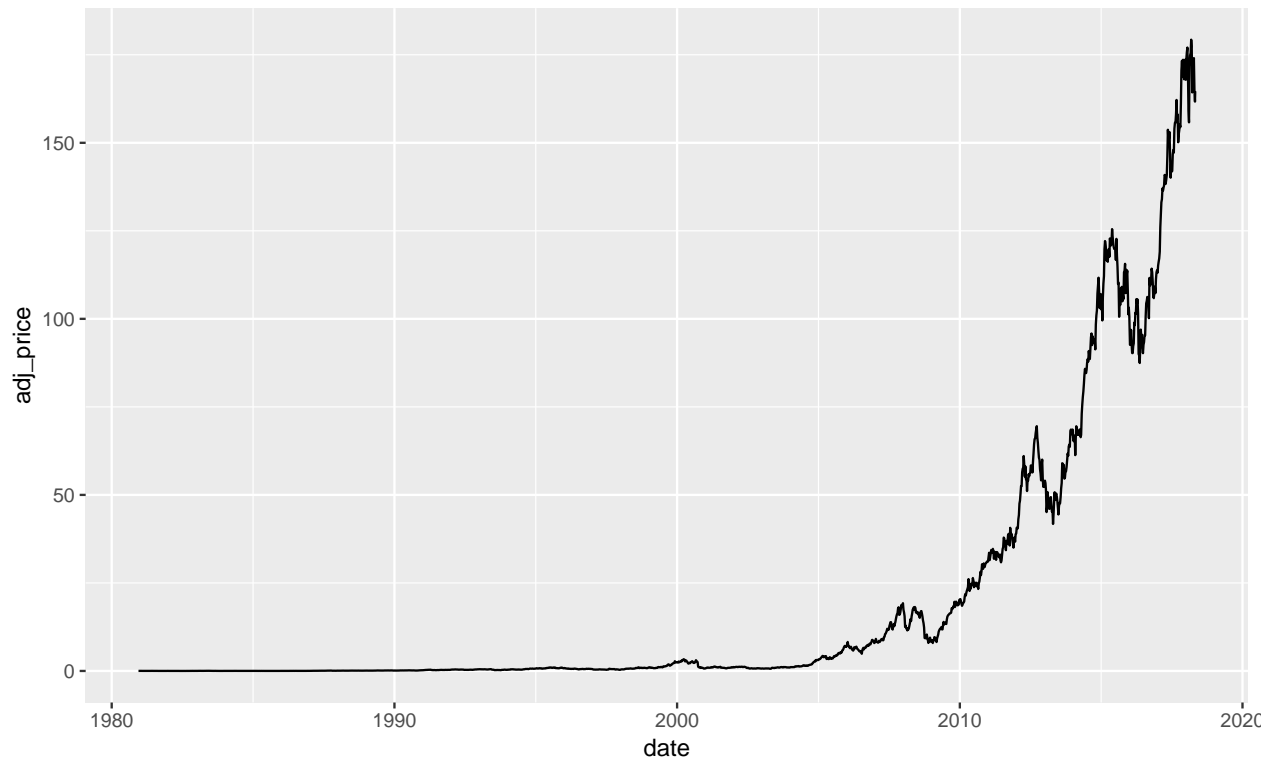


Figure 6.1: Apple

```
3 3 19.0
4 4 16.0
5 5 15.6
6 7 19.8
```

```
x Time y Demand BOD
```

```
ggplot(BOD, aes(Time, demand)) + # x Time y demand
 geom_line() + #
 geom_point() #
```

```
geom_point()
```

```
BOD
```

Figure 6.1

## 6.2

```
R ToothGrowth
```

```
head(ToothGrowth) # 6
```

```
len supp dose
1 4.2 VC 0.5
2 11.5 VC 0.5
3 7.3 VC 0.5
4 5.8 VC 0.5
5 6.4 VC 0.5
6 10.0 VC 0.5
```



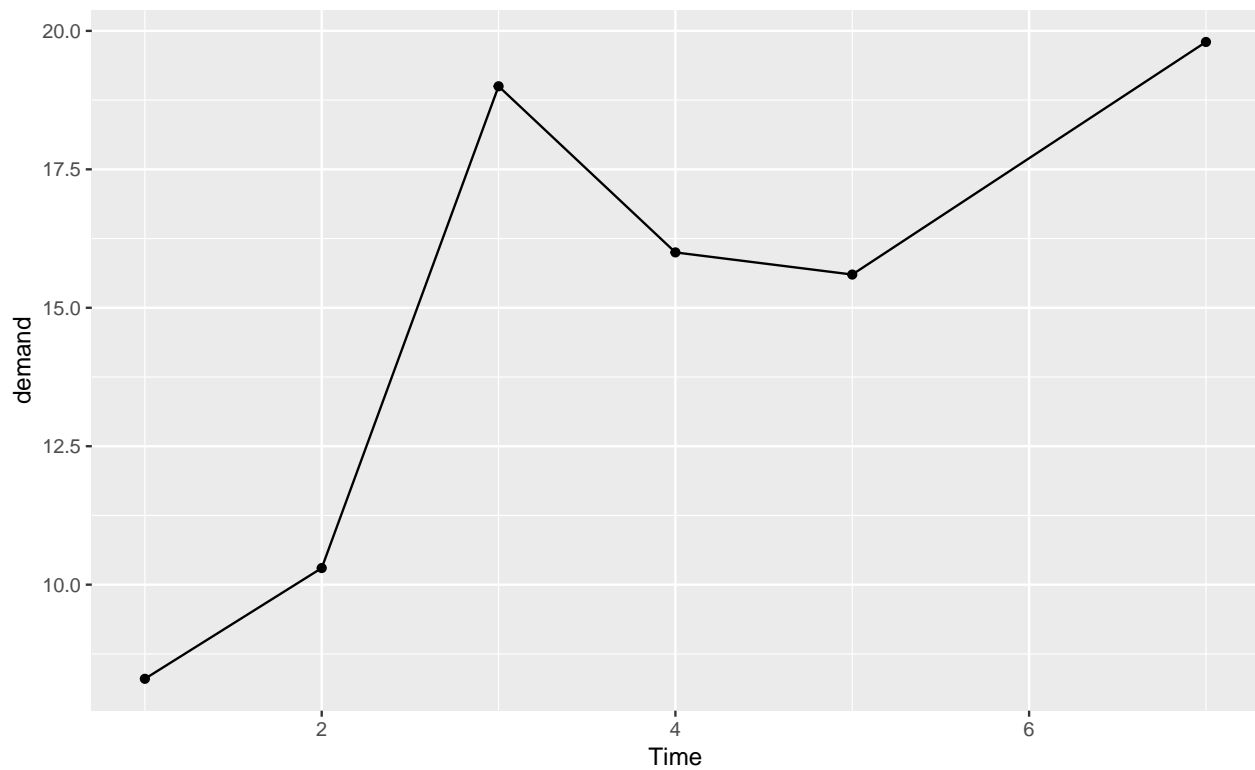


Figure 6.2: BOD

- len
- supp C VC or OJ
- dose C mg

```

 supp dose len ToothGrowth tg_mean
library(tidyverse) #

```

```

tg_mean = ToothGrowth %>% # ToothGrowth
 group_by(supp, dose) %>% # supp dose
 summarise(mean = mean(len)) # len

```

```
tg_mean
```

```

A tibble: 6 x 3
Groups: supp [?]
supp dose mean
<fct> <dbl> <dbl>
1 OJ 0.5 13.2
2 OJ 1 22.7
3 OJ 2 26.1
4 VC 0.5 7.98
5 VC 1 16.8
6 VC 2 26.1

```

```
mean len
```

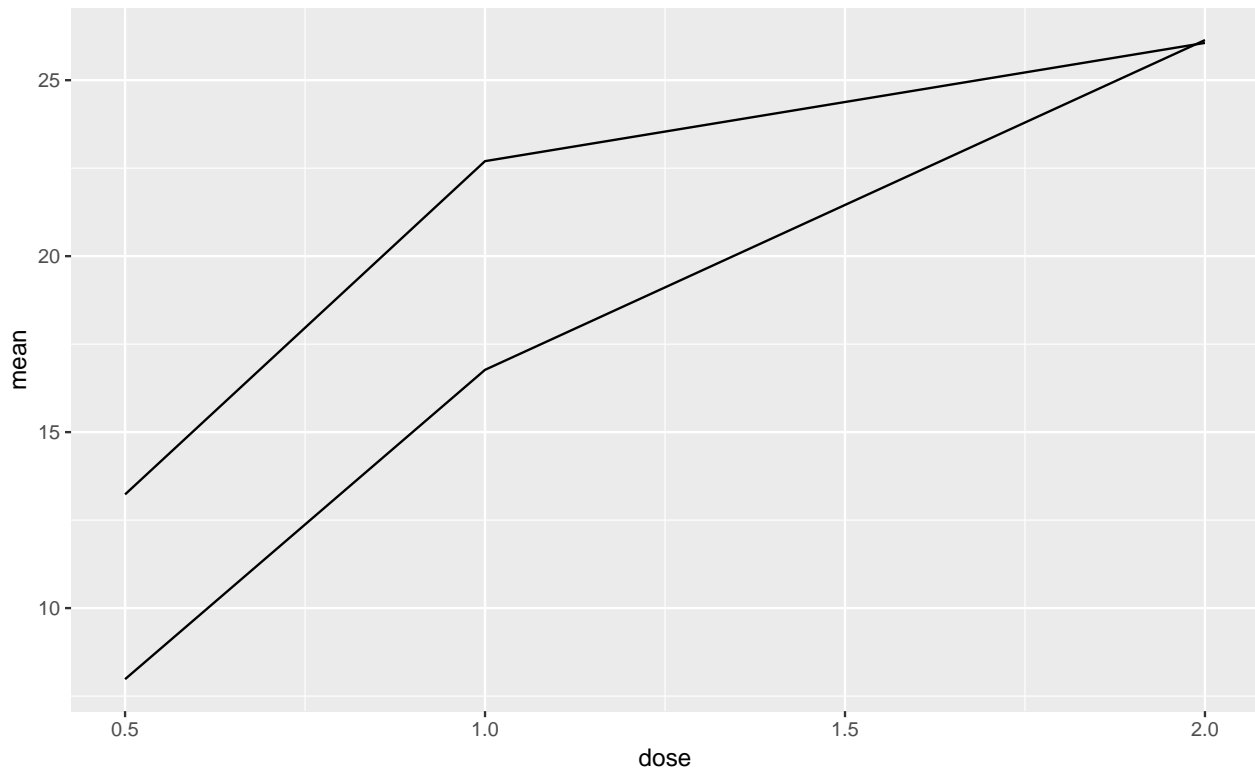


Figure 6.3: supp

```

 supp dose mean
ggplot(tg_mean, aes(dose, mean, group = supp)) + # x dose y mean group = supp
 geom_line()

 supp dose mean geom_line() group
 color linetype geom_line() size
 Chapter 5 shape geom_point() size
ggplot(tg_mean, aes(dose, mean, color = supp, group = supp, linetype = supp, shape = supp)) + # color linetype shape
 geom_line(size = 1.5) + # size
 geom_point(size = 5) # size

```

## 6.3

area graph; area chart

gcookbook uspopage 1900 2002

```
head(uspopage)
```

```
Year AgeGroup Thousands
1 1900 <5 9181
2 1900 5-14 16966
3 1900 15-24 14951
4 1900 25-34 12161
```

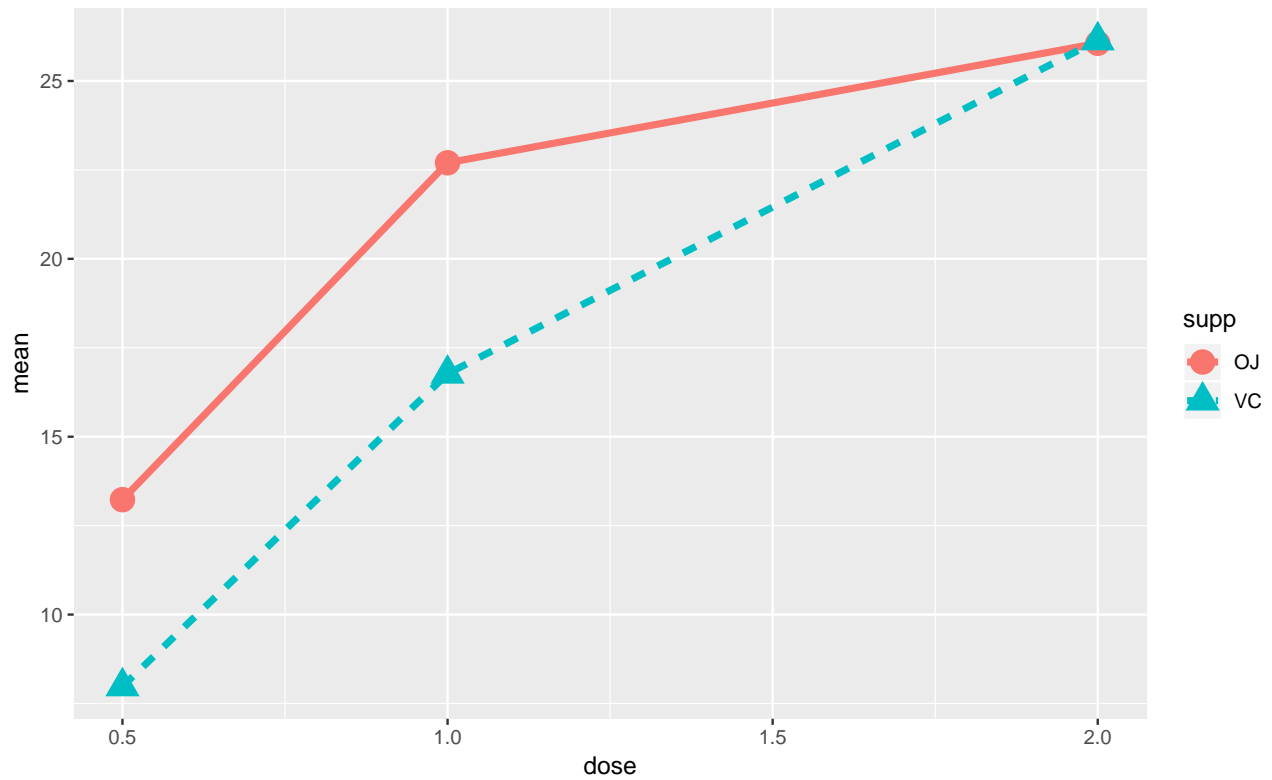


Figure 6.4: Figure 6.3

supp

```
5 1900 35-44 9273
6 1900 45-54 6437
```

- Year
- AgeGroup
- Thousands

```
geom_area() fill
ggplot(uspope, aes(Year, Thousands, fill = AgeGroup)) + # x Year y Thousands fill = AgeGroup
geom_area() #
```

## 6.4

- `ggplot2` economics
- `gcbook` worldpop `geom_line()` `geom_point()` `log10( )`

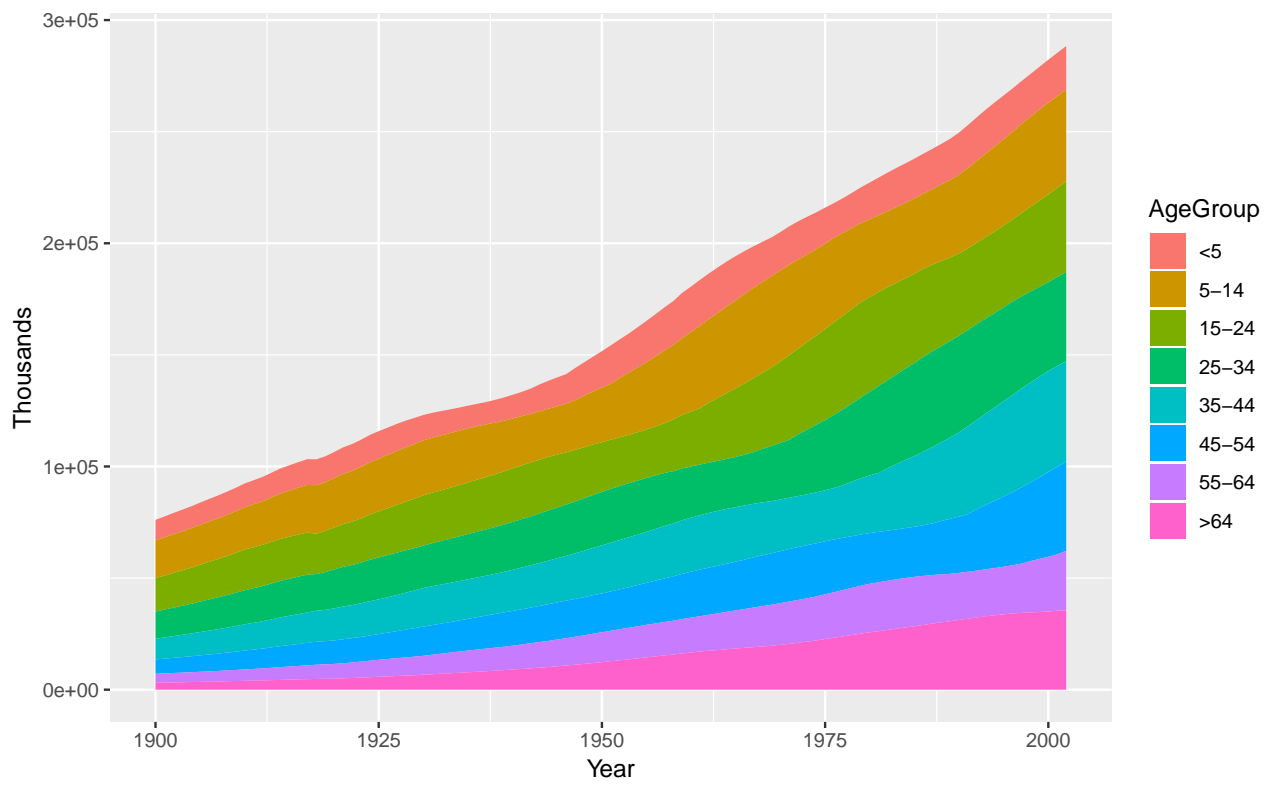


Figure 6.5:

# Chapter 7

```
library(ggplot2) #
```

## 7.1

error bar

standard error of the mean;  $SEM^1$

gcookbook cabbage\_exp

```
library(gcookbook) #
```

```
head(cabbage_exp) #6
```

| ##   | Cultivar | Date | Weight | sd        | n  | se         |
|------|----------|------|--------|-----------|----|------------|
| ## 1 | c39      | d16  | 3.18   | 0.9566144 | 10 | 0.30250803 |
| ## 2 | c39      | d20  | 2.80   | 0.2788867 | 10 | 0.08819171 |
| ## 3 | c39      | d21  | 2.74   | 0.9834181 | 10 | 0.31098410 |
| ## 4 | c52      | d16  | 2.26   | 0.4452215 | 10 | 0.14079141 |
| ## 5 | c52      | d20  | 3.11   | 0.7908505 | 10 | 0.25008887 |
| ## 6 | c52      | d21  | 1.47   | 0.2110819 | 10 | 0.06674995 |

- Cultivar      c39 or c52
- Date
- Weight
- se

(ref:errorbar-1)

```
ggplot(cabbage_exp, aes(Date, Weight, fill = Cultivar)) + # x Date y Weight Cultivar
 geom_bar(stat = "identity", position = "dodge") + #
 geom_errorbar(aes(ymin = Weight - se, ymax = Weight + se), position = position_dodge(width = 0.9), width = 0.5)
```

---

<sup>1</sup> structural equation modeling; SEM

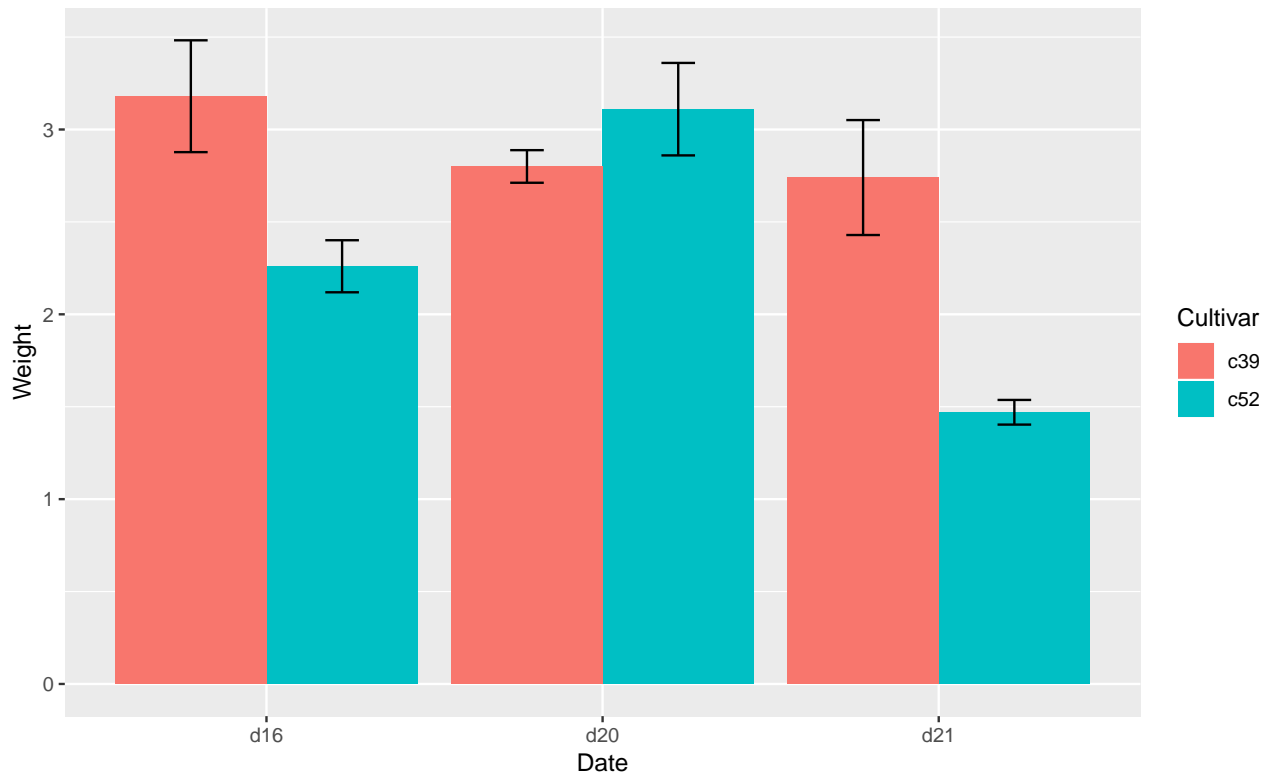


Figure 7.1: (ref:errorbar-1)

```
0. ggplot(cabbage_exp, aes(Date, Weight, fill = Cultivar))
 • x Date y Weight Cultivar
1. geom_bar(stat = "identity", position = "dodge")
 • stat = "identity" Cultivar position = "dodge"
 • Chapter 3
2. geom_errorbar(aes(ymin = Weight - se, ymax = Weight + se), position = position_dodge(width
 = 0.9), width = 0.2)
 • ymin ymax y
 • position = position_dodge(width = 0.9) 0.9 width 0.9 2
 • width
 95% 95% confidence interval; 95% CI
```

## 7.2

```
geom_errorbar() geom_ribbon() gcookbook climate
head(climate)

Source Year Anomaly1y Anomaly5y Anomaly10y Unc10y
 ^-----^-----^-----^-----^-----^-----^
 2 position = "dodge" position = position_dodge()
```

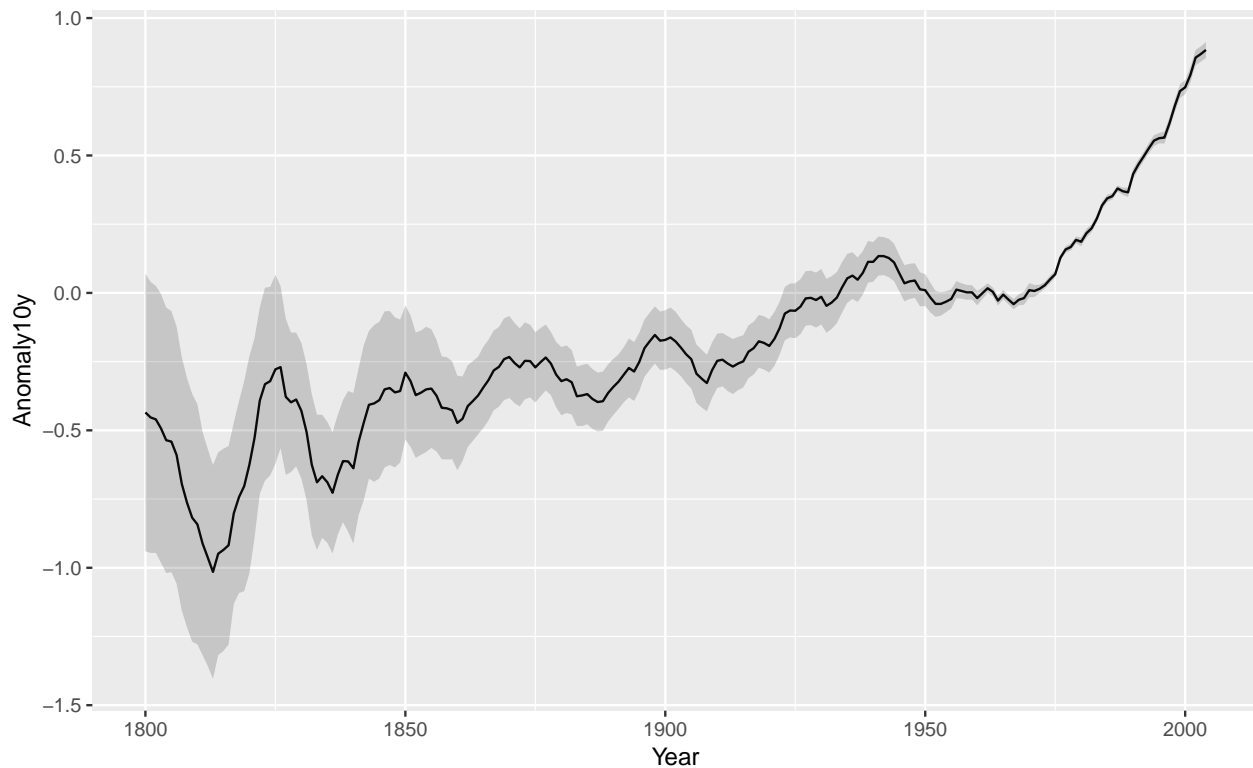


Figure 7.2: 95%

```
1 Berkeley 1800 NA NA -0.435 0.505
2 Berkeley 1801 NA NA -0.453 0.493
3 Berkeley 1802 NA NA -0.460 0.486
4 Berkeley 1803 NA NA -0.493 0.489
5 Berkeley 1804 NA NA -0.536 0.483
6 Berkeley 1805 NA NA -0.541 0.475
```

```
Source "Berkeley" clim
```

```
library(tidyverse) #
```

```
clim = climate %>% # climate
 filter(Source == "Berkeley") # Source "Berkeley"
```

- Year
- Anomaly10y 1951 1980
- Unc10y 95%

```
geom_ribbon() geom_errorbar() alpha
```

```
ggplot(clim, aes(Year, Anomaly10y)) +
 geom_line() +
 geom_ribbon(aes(ymin = Anomaly10y - Unc10y, ymax = Anomaly10y + Unc10y), alpha = 0.2) # alpha
```

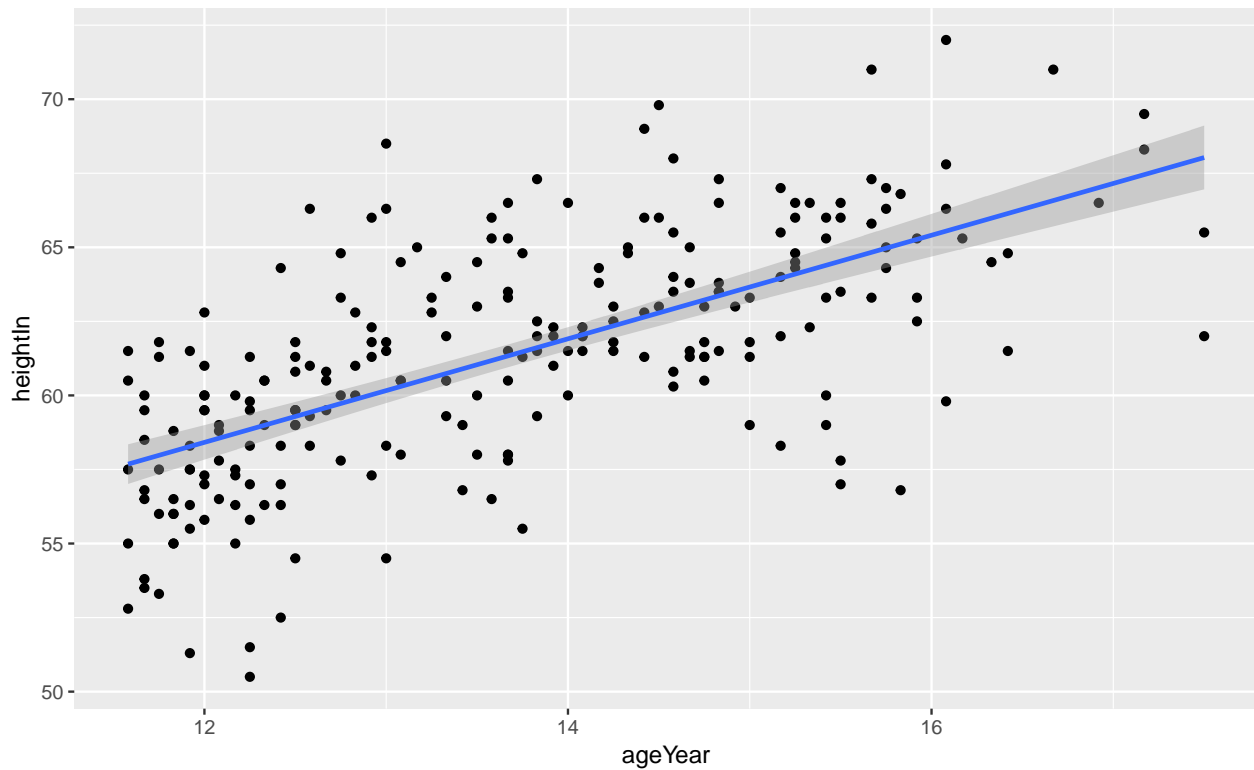


Figure 7.3: (ref:smooth-plot-1)

## 7.3

```

ggplot2 gcocbook heightweight
head(heightweight) # 6

sex ageYear ageMonth heightIn weightLb
1 f 11.92 143 56.3 85.0
2 f 12.92 155 62.3 105.0
3 f 12.75 153 63.3 108.0
4 f 13.42 161 59.0 92.0
5 f 15.92 191 62.5 112.5
6 f 14.25 171 62.5 112.0

geom_point() heightIn WeightLb geom_smooth() method = lm LOESS 3
R lm linear model

(ref:smooth-plot-1) 95%

ggplot(heightweight, aes(ageYear, heightIn)) +
 geom_point() + #
 geom_smooth(method = lm) # method = lm

fill

(ref:smooth-plot-2) 95%

```

<sup>3</sup>[https://en.wikipedia.org/wiki/Local\\_regression](https://en.wikipedia.org/wiki/Local_regression)



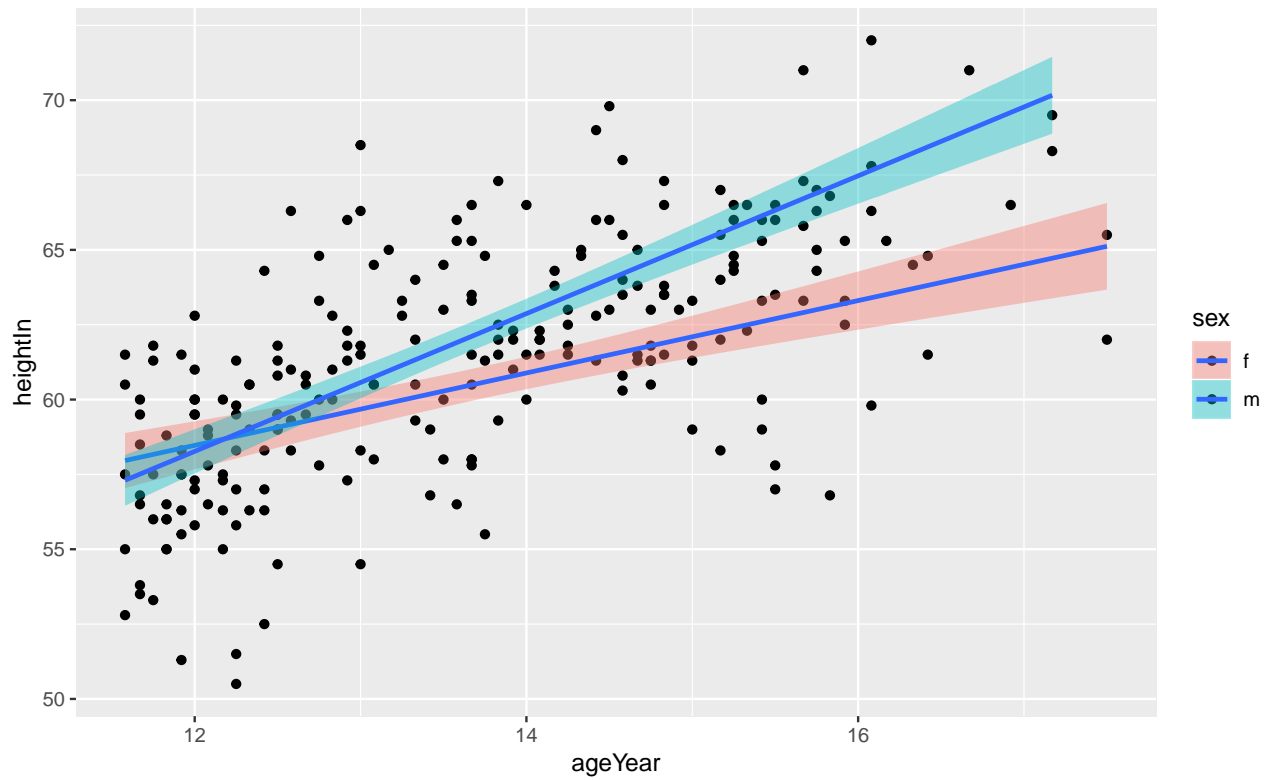


Figure 7.4: (ref:smooth-plot-2)

```
ggplot(heightweight, aes(ageYear, heightIn, fill = sex)) + # fill = sex
 geom_point() +
 geom_smooth(method = lm)
```

## 7.4

- Figure 7.1 cabbage\_exp
- R faithful



# Chapter 8

ggplot2

```
library(ggplot2) #
```

## 8.1

<sup>1</sup> Figure 3.5 Cleveland x y ggplot2  
coord\_flip() coordination flip  
Figure 3.3 diamonds cut coord\_flip()  
ggplot(diamonds, aes(cut)) + # x cut y  
geom\_bar() + # stat = "identity"  
coord\_flip() #

## 8.2 Small multiple

1 Small multiple<sup>2</sup> Small multiple facet\_grid() facet\_wrap()  
Chapter 5 mpg Figure 5.6 drv cyl facet\_grid( ~ )

(ref:facet-1) facet\_grid()

```
ggplot(mpg, aes(displ, hwy)) +
geom_point() +
facet_grid(drv ~ cyl) # drv cyl
```

---

```
facet_wrap() 2 drv cyl nrow ncol
ggplot(mpg, aes(displ, hwy)) +
geom_point() +
facet_wrap(drv ~ cyl) #
```

---

```
facet_grid() facet_wrap() 1 ?facet_grid() ?facet_wrap()
```

<sup>1</sup> .....

<sup>2</sup>[https://en.wikipedia.org/wiki/Small\\_multiple](https://en.wikipedia.org/wiki/Small_multiple)

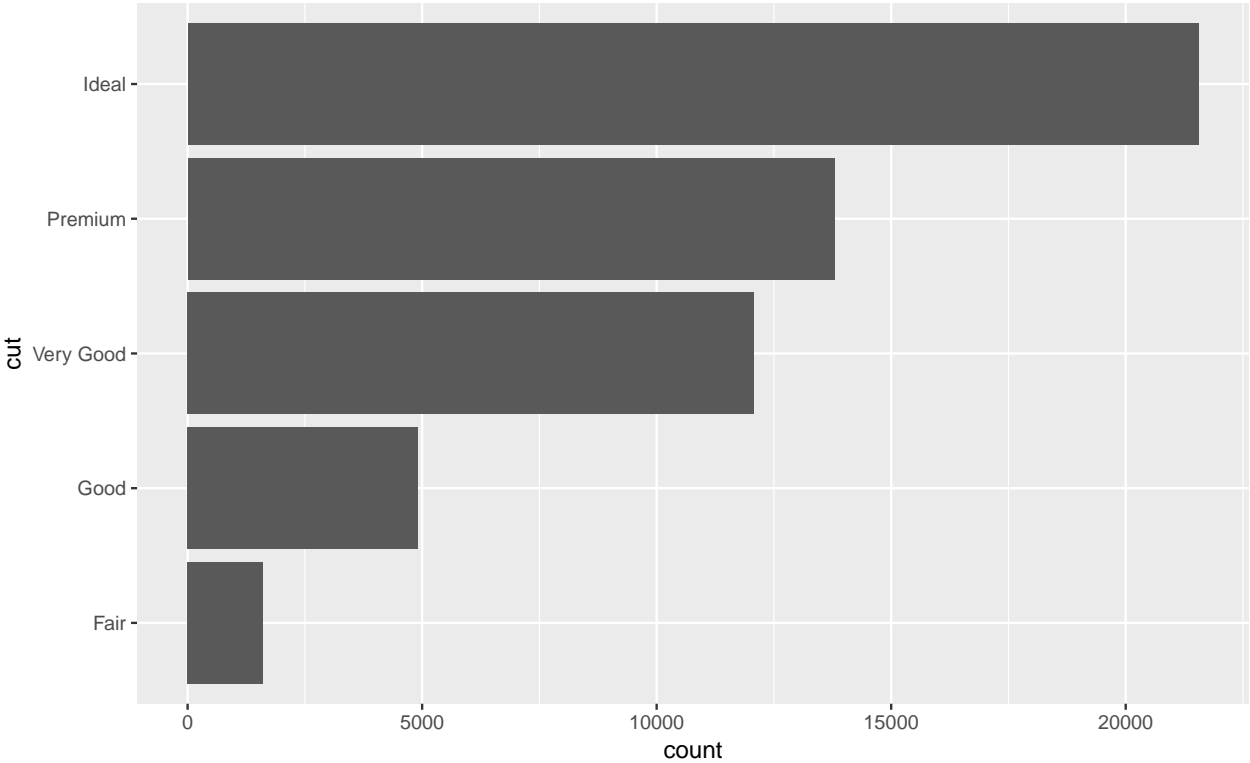


Figure 8.1: Figure 3.3

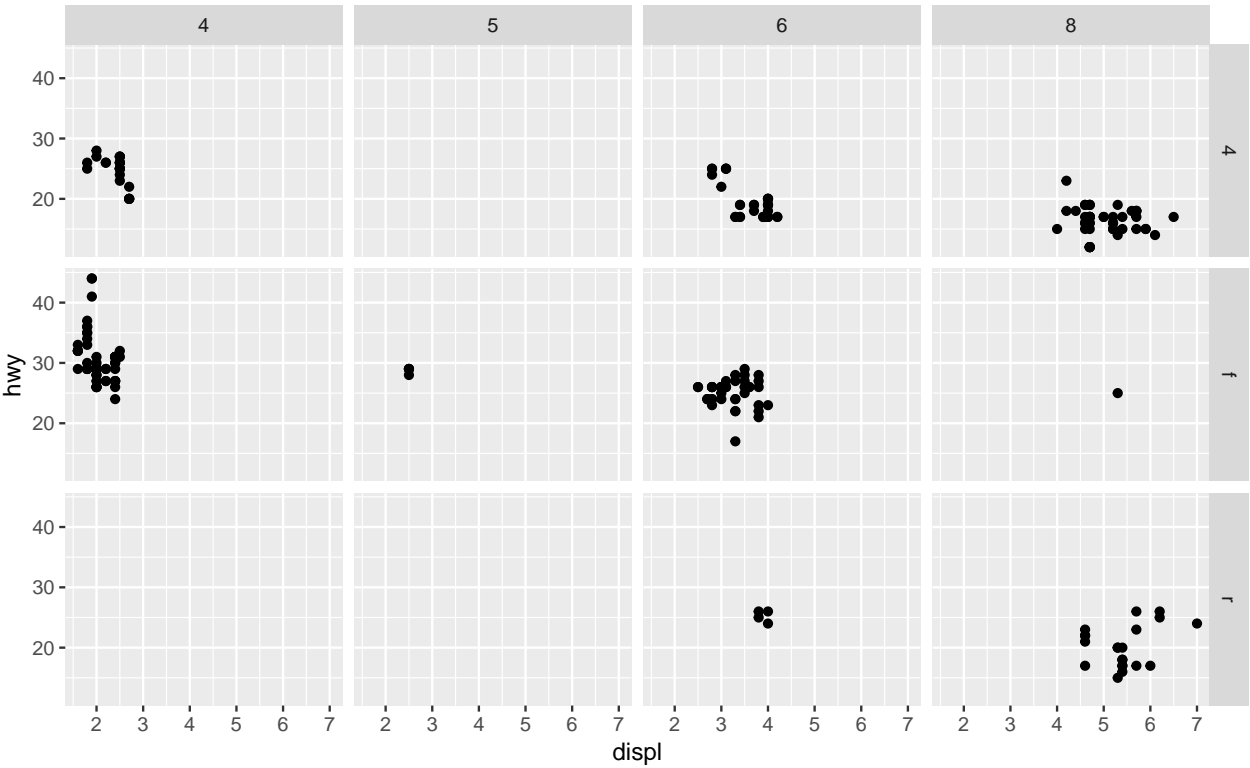
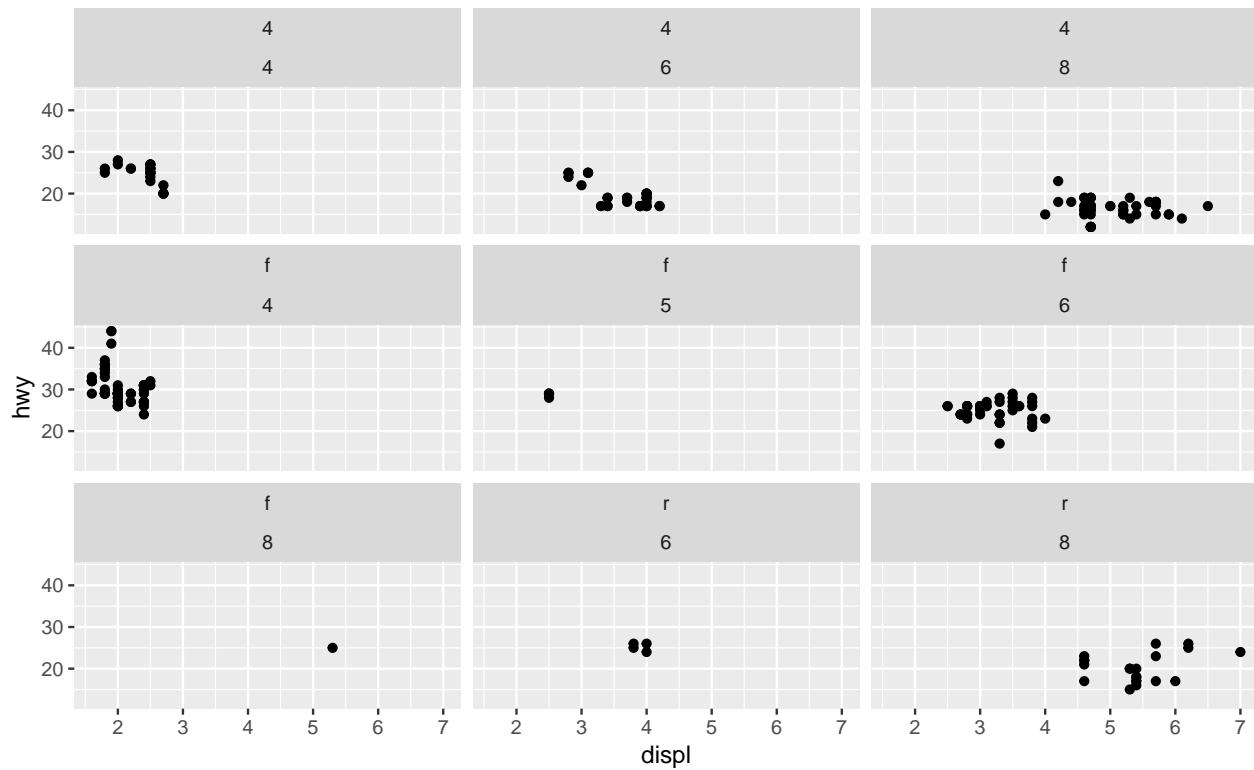


Figure 8.2: (ref:facet-1)

Figure 8.3: `facet_wrap()`      Figure 8.2

## 8.3

```
theme_xxxx() theme_gray() ggplot2
ggplot(mtcars, aes(wt, mpg)) +
 geom_point() +
 theme_gray()
```

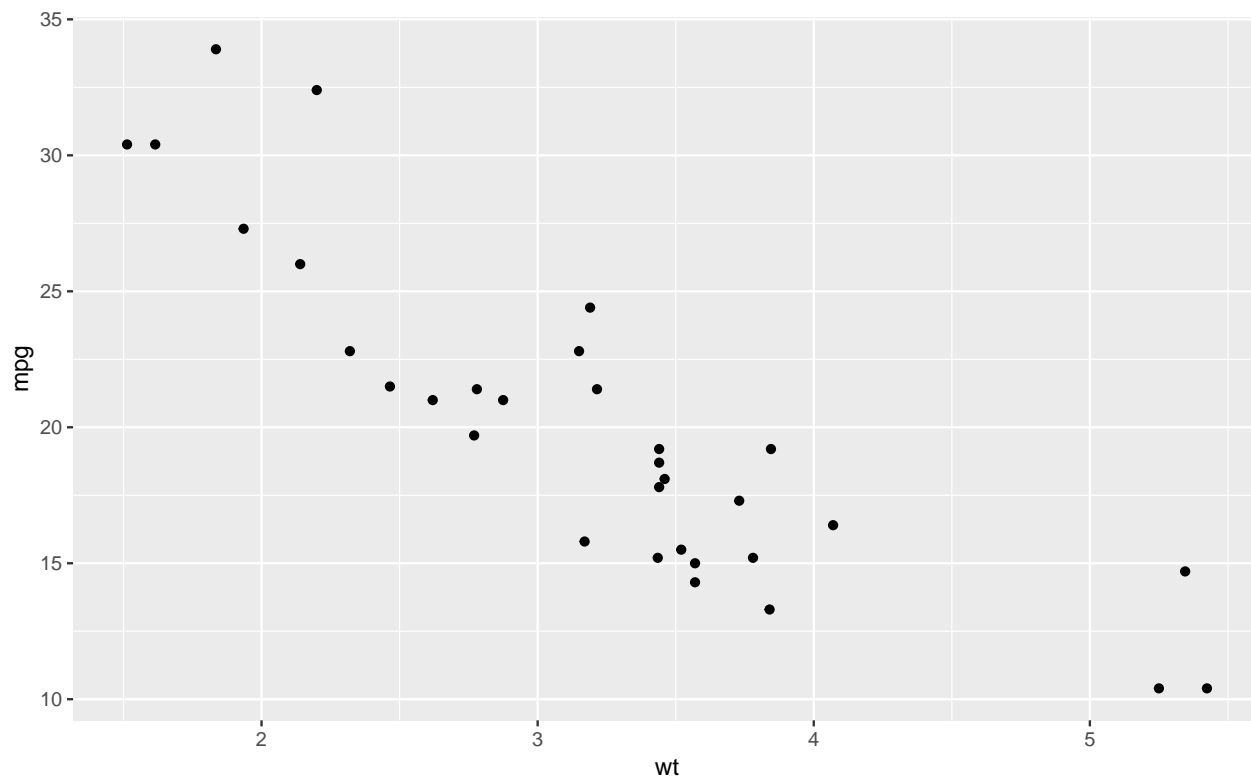
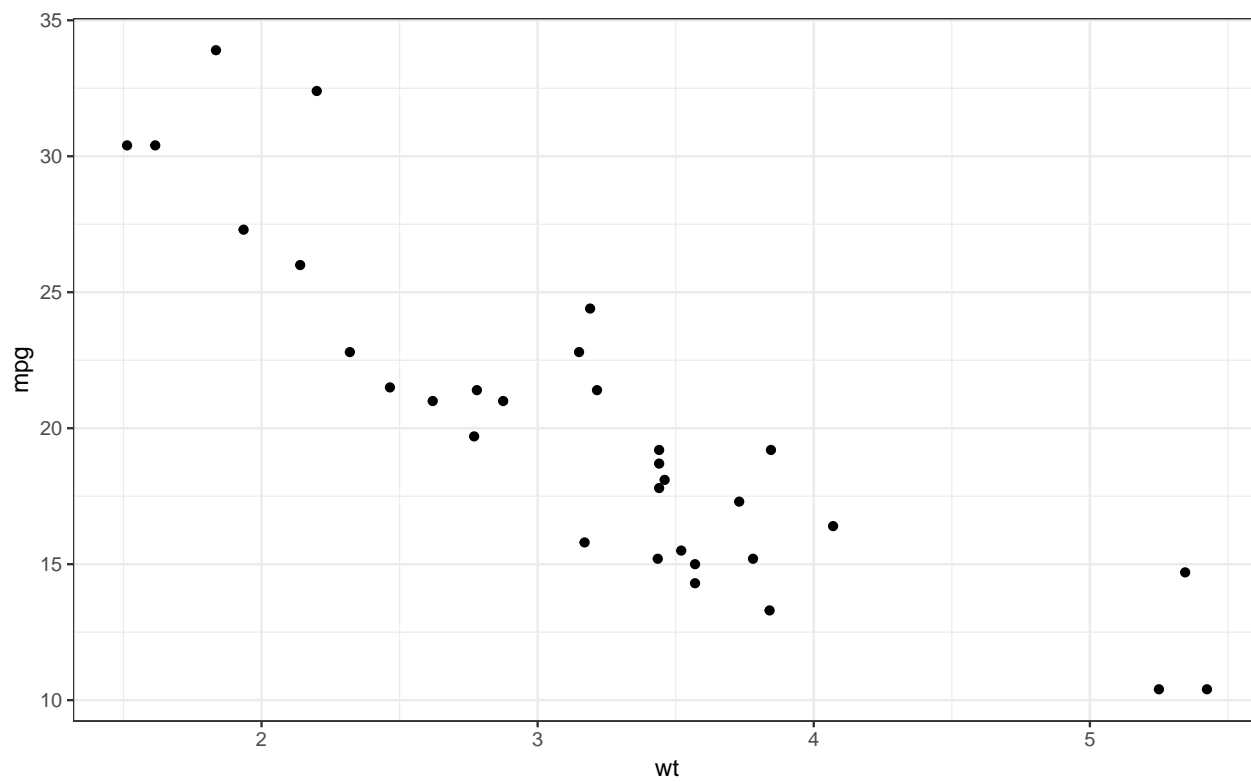
```
theme_bw() theme_classic()
ggplot(mtcars, aes(wt, mpg)) +
 geom_point() +
 theme_bw()
```

```
ggplot(mtcars, aes(wt, mpg)) +
 geom_point() +
 theme_classic()
```

<https://ggplot2.tidyverse.org/reference/ggtheme.html>

## 8.4

|                      |         |                   |                     |
|----------------------|---------|-------------------|---------------------|
| fill color           | ggplot2 |                   |                     |
| scale_color_manual() | color   | <sup>3</sup> fill | scale_fill_manual() |

Figure 8.4: `theme_gray()`Figure 8.5: `theme_bw()`

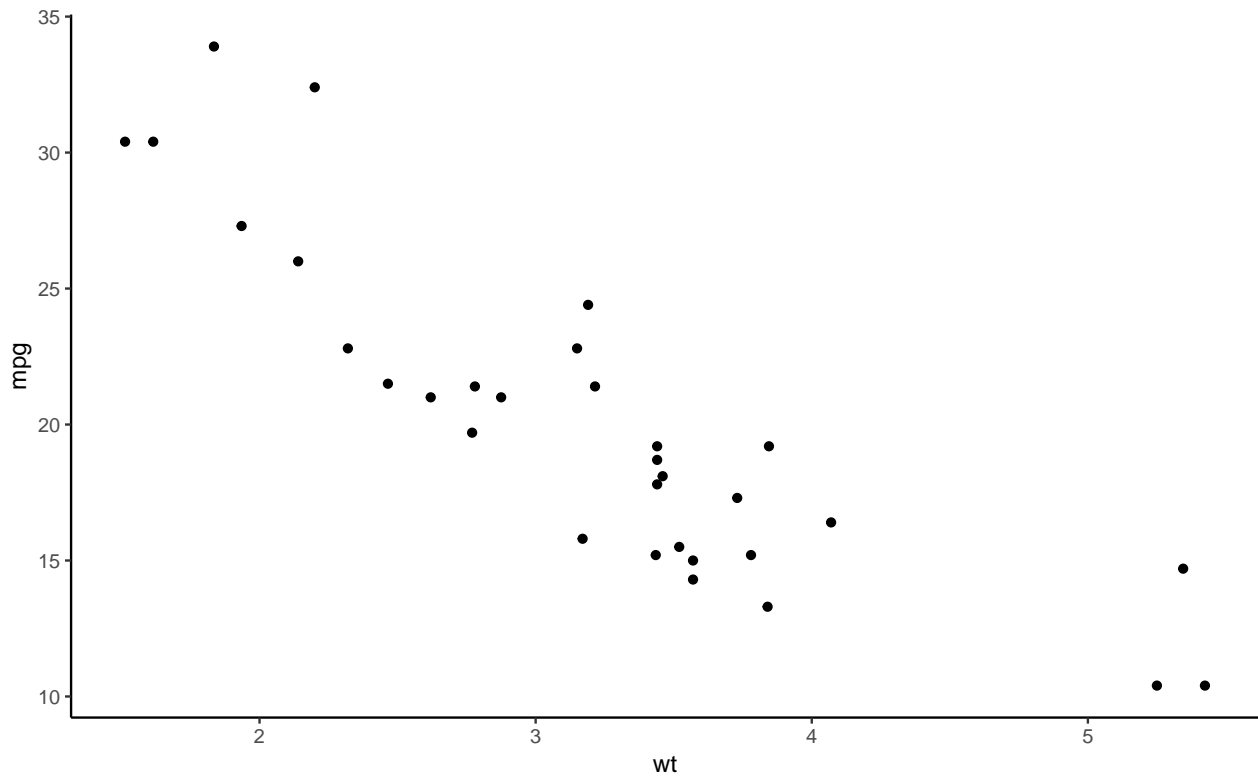


Figure 8.6: theme\_classic()

```
ggplot(mtcars, aes(wt, mpg, color = factor(cyl))) +
 geom_point() +
 scale_color_manual(values = c("red", "blue", "green"))
```

ggplot2

colorspace

4

## 8.5

data exploration

data presentation

5

labs()

Title Subtitle

(ref:label-1) labs()

```
ggplot(mtcars, aes(wt, mpg, color = factor(cyl))) +
 geom_point() +
 labs(x = "Weight (1,000 lbs)",
 y = "Miles/(US) gallon",
 title = "Title",
 subtitle = "Subtitle",
 caption = "Caption",
 tag = "Tag")
```

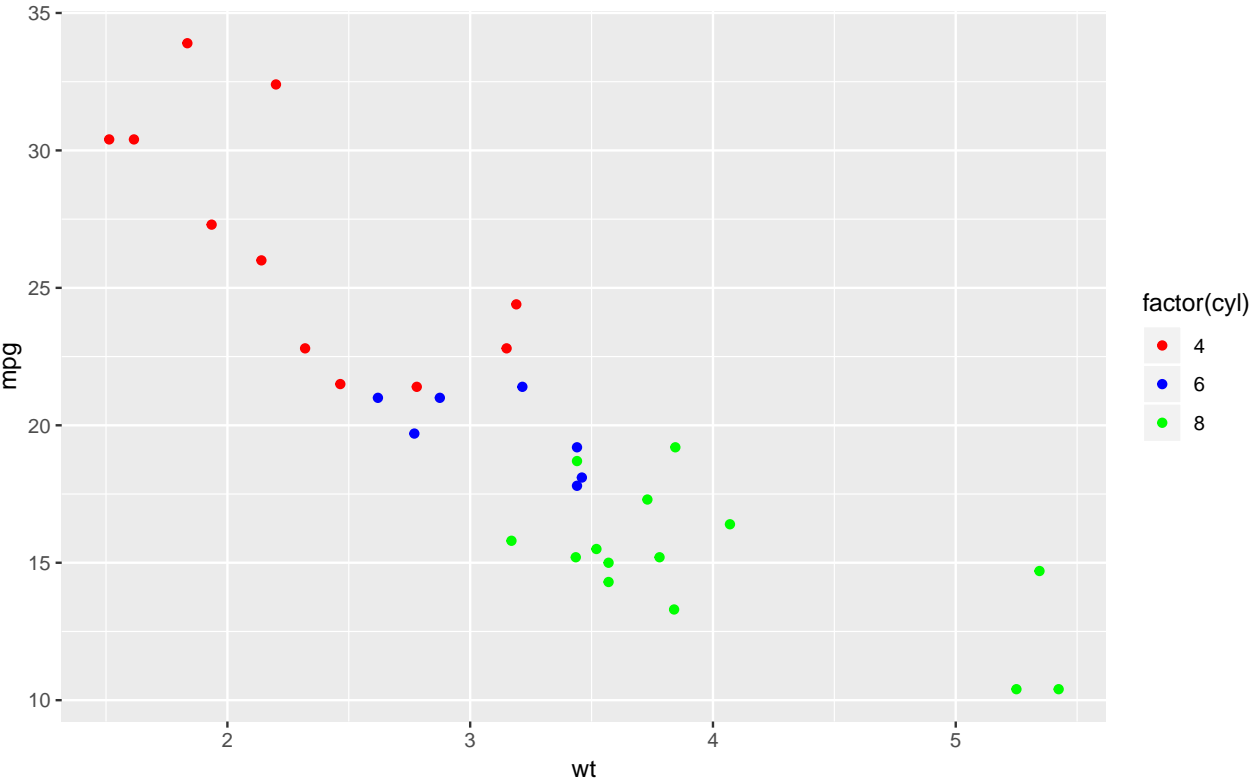


Figure 8.7: `scale_color_manual()`

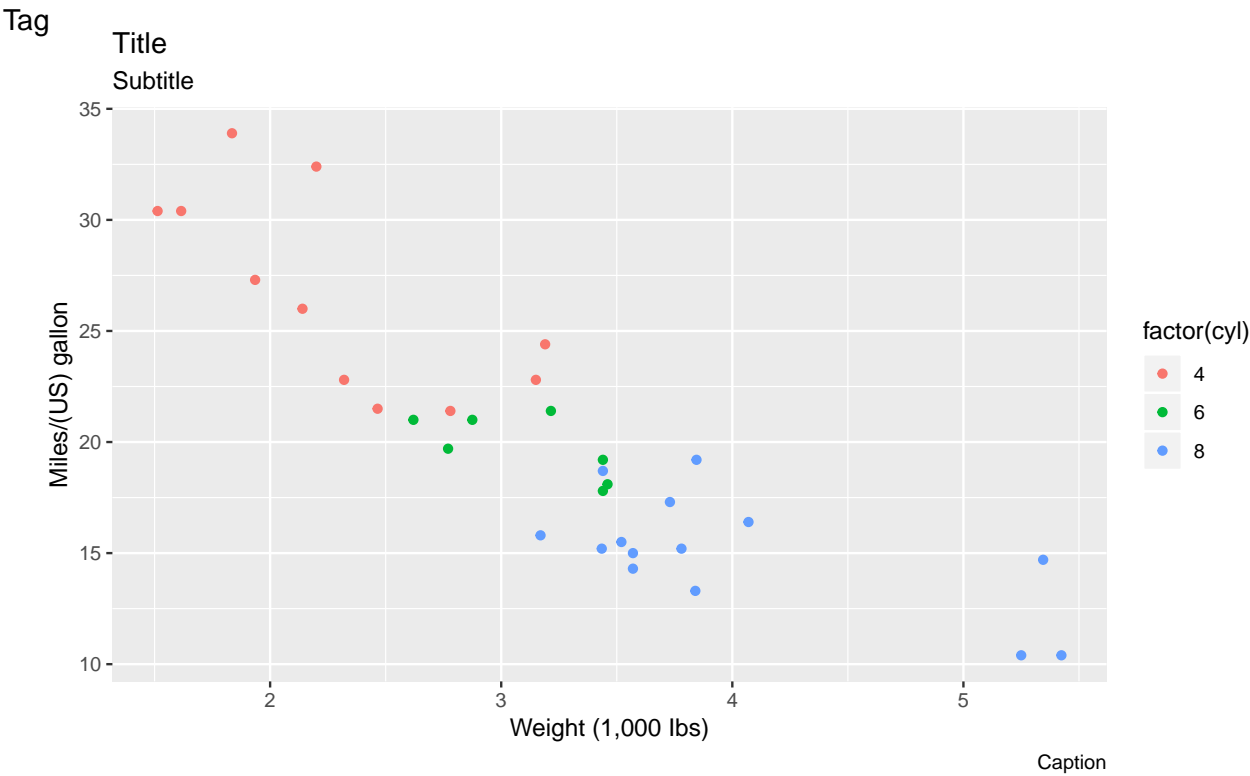


Figure 8.8: (ref:label-1)



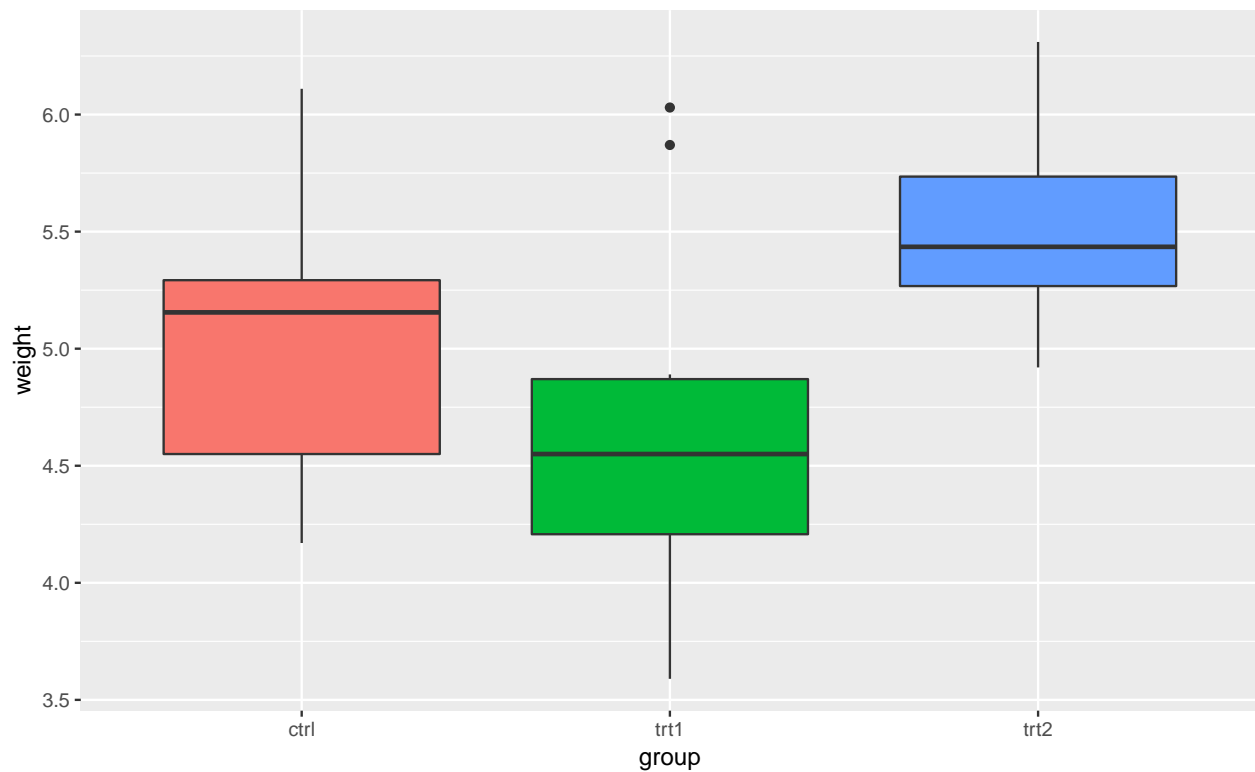


Figure 8.9:

## 8.6

Figure 4.12

```
ggplot(PlantGrowth, aes(group, weight, fill = group)) + # x group y weight fill
 geom_boxplot() + #
 guides(fill = FALSE) # fill
```

## 8.7

```
1 patchwork patchwork

library(devtools)
install_github("thomasp85/patchwork") #

library(patchwork) #

patchwork
a panel_a
panel_a = ggplot(mtcars, aes(wt, mpg, color = factor(cyl))) +
 geom_point()

b panel_b
panel_b = ggplot(mtcars, aes(wt, disp, color = factor(cyl))) +
 geom_point()
```

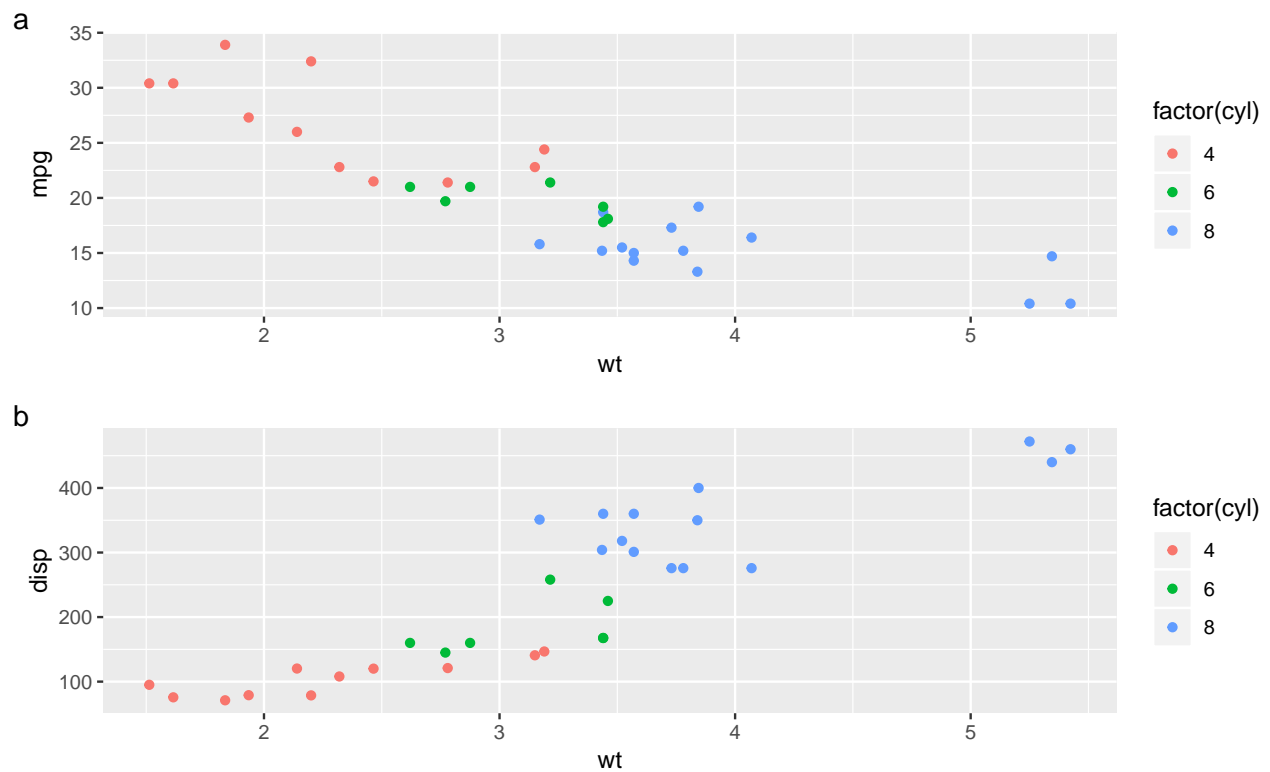


Figure 8.10: 2

```
2 1 panel_ab 1 ncol = 1
panel_ab = panel_a + panel_b +
 plot_layout(ncol = 1) +
 plot_annotation(tag_levels = "a") #
panel_ab
```

## 8.8

```
width = , height = , units = " ")
ggsave("panel_ab.pdf", plot = ,
ggsave("panel_ab.pdf", plot = panel_ab, width = 14, height = 14, units = "cm")
```

PDF <sup>6</sup> PDF Google

## 8.9

- gcookbook heightweight

---

<sup>6</sup>PDF





# Chapter 9

ggplot2                      ggplot2

mtcars faithful

```
head(mtcars) # 6
```

```
mpg cyl disp hp drat wt qsec vs am gear carb
Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4
Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4
Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1
Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1
Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2
Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1
```

```
head(faithful) # 6
```

```
eruptions waiting
1 3.600 79
2 1.800 54
3 3.333 74
4 2.283 62
5 4.533 85
6 2.883 55
```

---

1

3

- 1                      3
- 
- 2

tidyverse

## 9.1 tidyverse

```

tidyverse library(tidyverse) ggplot2 tidyverse
library(tidyverse) # ggplot2

```

```
ggplot2 mpg
```

## 9.2

### 9.2.1

```

mpg manufacturer
filter()

mpg %>%
 filter(manufacturer == "audi") # manufacturer "audi"

A tibble: 18 x 11
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~
3 audi a4 2 2008 4 manu~ f 20 31 p comp~
4 audi a4 2 2008 4 auto~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~
8 audi a4 q~ 1.8 1999 4 manu~ 4 18 26 p comp~
9 audi a4 q~ 1.8 1999 4 auto~ 4 16 25 p comp~
10 audi a4 q~ 2 2008 4 manu~ 4 20 28 p comp~
11 audi a4 q~ 2 2008 4 auto~ 4 19 27 p comp~
12 audi a4 q~ 2.8 1999 6 auto~ 4 15 25 p comp~
13 audi a4 q~ 2.8 1999 6 manu~ 4 17 25 p comp~
14 audi a4 q~ 3.1 2008 6 auto~ 4 17 25 p comp~
15 audi a4 q~ 3.1 2008 6 manu~ 4 15 25 p comp~
16 audi a6 q~ 2.8 1999 6 auto~ 4 15 24 p mids~
17 audi a6 q~ 3.1 2008 6 auto~ 4 17 25 p mids~
18 audi a6 q~ 4.2 2008 8 auto~ 4 16 23 p mids~

== == =

%>%

```

```

 manufacturer "audi" model "a4" &

mpg %>%
 filter(manufacturer == "audi" & model == "a4")

```

```

A tibble: 7 x 11
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto(~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manua~ f 21 29 p comp~

```

```
3 audi a4 2 2008 4 manua~ f 20 31 p comp~
4 audi a4 2 2008 4 auto(~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto(~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manua~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto(~ f 18 27 p comp~
```

```
manufacturer "audi" model "a4" |
```

```
mpg %>%
 filter(manufacturer == "audi" | model == "a4")
```

```
A tibble: 18 x 11
```

```
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~
3 audi a4 2 2008 4 manu~ f 20 31 p comp~
4 audi a4 2 2008 4 auto~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~
8 audi a4 q~ 1.8 1999 4 manu~ 4 18 26 p comp~
9 audi a4 q~ 1.8 1999 4 auto~ 4 16 25 p comp~
10 audi a4 q~ 2 2008 4 manu~ 4 20 28 p comp~
11 audi a4 q~ 2 2008 4 auto~ 4 19 27 p comp~
12 audi a4 q~ 2.8 1999 6 auto~ 4 15 25 p comp~
13 audi a4 q~ 2.8 1999 6 manu~ 4 17 25 p comp~
14 audi a4 q~ 3.1 2008 6 auto~ 4 17 25 p comp~
15 audi a4 q~ 3.1 2008 6 manu~ 4 15 25 p comp~
16 audi a6 q~ 2.8 1999 6 auto~ 4 15 24 p mids~
17 audi a6 q~ 3.1 2008 6 auto~ 4 17 25 p mids~
18 audi a6 q~ 4.2 2008 8 auto~ 4 16 23 p mids~
```

```
manufacturer "audi" model "a4" !() !() 1
```

```
mpg %>%
 filter(!(manufacturer == "audi" | model == "a4"))
```

```
A tibble: 216 x 11
```

```
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 chevrolet c150~ 5.3 2008 8 auto~ r 14 20 r suv
2 chevrolet c150~ 5.3 2008 8 auto~ r 11 15 e suv
3 chevrolet c150~ 5.3 2008 8 auto~ r 14 20 r suv
4 chevrolet c150~ 5.7 1999 8 auto~ r 13 17 r suv
5 chevrolet c150~ 6 2008 8 auto~ r 12 17 r suv
6 chevrolet corv~ 5.7 1999 8 manu~ r 16 26 p 2sea~
7 chevrolet corv~ 5.7 1999 8 auto~ r 15 23 p 2sea~
8 chevrolet corv~ 6.2 2008 8 manu~ r 16 26 p 2sea~
9 chevrolet corv~ 6.2 2008 8 auto~ r 15 25 p 2sea~
10 chevrolet corv~ 7 2008 8 manu~ r 15 24 p 2sea~
... with 206 more rows
```

---

<sup>1</sup> manufacturer "audi"      manufacturer != "audi"

## 9.2.2

```

 mpg manufacturer model trans drv
select() 1

mpg %>%
 select(manufacturer, model, trans, drv) # 4

A tibble: 234 x 4
manufacturer model trans drv
<chr> <chr> <chr> <chr>
1 audi a4 auto(l5) f
2 audi a4 manual(m5) f
3 audi a4 manual(m6) f
4 audi a4 auto(av) f
5 audi a4 auto(l5) f
6 audi a4 manual(m5) f
7 audi a4 auto(av) f
8 audi a4 quattro manual(m5) 4
9 audi a4 quattro auto(l5) 4
10 audi a4 quattro manual(m6) 4
... with 224 more rows

select() A G

mpg %>%
 select(manufacturer:drv) # manufacturer drv 7

A tibble: 234 x 7
manufacturer model displ year cyl trans drv
<chr> <chr> <dbl> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto(l5) f
2 audi a4 1.8 1999 4 manual(m5) f
3 audi a4 2 2008 4 manual(m6) f
4 audi a4 2 2008 4 auto(av) f
5 audi a4 2.8 1999 6 auto(l5) f
6 audi a4 2.8 1999 6 manual(m5) f
7 audi a4 3.1 2008 6 auto(av) f
8 audi a4 quattro 1.8 1999 4 manual(m5) 4
9 audi a4 quattro 1.8 1999 4 auto(l5) 4
10 audi a4 quattro 2 2008 4 manual(m6) 4
... with 224 more rows

select(-)

mpg %>%
 select(-manufacturer) # manufacturer

A tibble: 234 x 10
model displ year cyl trans drv cty hwy fl class
<chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 a4 1.8 1999 4 auto(l5) f 18 29 p compact
2 a4 1.8 1999 4 manual(m5) f 21 29 p compact
3 a4 2 2008 4 manual(m6) f 20 31 p compact
4 a4 2 2008 4 auto(av) f 21 30 p compact
5 a4 2.8 1999 6 auto(l5) f 16 26 p compact
6 a4 2.8 1999 6 manual(m5) f 18 26 p compact

```



```
7 a4 3.1 2008 6 auto(av) f 18 27 p compact
8 a4 quattro 1.8 1999 4 manual(m5) 4 18 26 p compact
9 a4 quattro 1.8 1999 4 auto(l5) 4 16 25 p compact
10 a4 quattro 2 2008 4 manual(m6) 4 20 28 p compact
... with 224 more rows
```

## 9.3

### 9.3.1

```
mutate(=)

mpg %>%
 mutate(one = 1) # one

A tibble: 234 x 12
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~
3 audi a4 2 2008 4 manu~ f 20 31 p comp~
4 audi a4 2 2008 4 auto~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~
8 audi a4 q~ 1.8 1999 4 manu~ 4 18 26 p comp~
9 audi a4 q~ 1.8 1999 4 auto~ 4 16 25 p comp~
10 audi a4 q~ 2 2008 4 manu~ 4 20 28 p comp~
... with 224 more rows, and 1 more variable: one <dbl>

one 1

mutate() if_else() displ 3 good 3 bad engine

mpg %>%
 mutate(engine = if_else(displ < 3, "good", "bad"))

A tibble: 234 x 12
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~
3 audi a4 2 2008 4 manu~ f 20 31 p comp~
4 audi a4 2 2008 4 auto~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~
8 audi a4 q~ 1.8 1999 4 manu~ 4 18 26 p comp~
9 audi a4 q~ 1.8 1999 4 auto~ 4 16 25 p comp~
10 audi a4 q~ 2 2008 4 manu~ 4 20 28 p comp~
... with 224 more rows, and 1 more variable: engine <chr>
```

## 9.3.2

```

 rename(=)

mpg %>%
 rename(nen = year) # year nen

A tibble: 234 x 11
manufacturer model displ nen cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto~ f 18 29 p comp~
2 audi a4 1.8 1999 4 manu~ f 21 29 p comp~
3 audi a4 2 2008 4 manu~ f 20 31 p comp~
4 audi a4 2 2008 4 auto~ f 21 30 p comp~
5 audi a4 2.8 1999 6 auto~ f 16 26 p comp~
6 audi a4 2.8 1999 6 manu~ f 18 26 p comp~
7 audi a4 3.1 2008 6 auto~ f 18 27 p comp~
8 audi a4 q~ 1.8 1999 4 manu~ 4 18 26 p comp~
9 audi a4 q~ 1.8 1999 4 auto~ 4 16 25 p comp~
10 audi a4 q~ 2 2008 4 manu~ 4 20 28 p comp~
... with 224 more rows

```

## 9.4

```

 group_by() summarise()

summarise(summarise(=) displ

mpg %>%
 summarise(mean_displ = mean(displ),
 sd_displ = sd(displ),
 min_displ = min(displ),
 max_displ = max(displ))

A tibble: 1 x 4
mean_displ sd_displ min_displ max_displ
<dbl> <dbl> <dbl> <dbl>
1 3.47 1.29 1.6 7

 summarise()

```

---

```

 group_by() year summarise()

mpg %>%
 group_by(year) %>%
 summarise(mean_displ = mean(displ),
 sd_displ = sd(displ),
 min_displ = min(displ),
 max_displ = max(displ))

A tibble: 2 x 5
year mean_displ sd_displ min_displ max_displ
<int> <dbl> <dbl> <dbl> <dbl>

```

```
1 1999 3.28 1.26 1.6 6.5
2 2008 3.66 1.30 1.8 7
1999 2008
```

## 9.5

mpg

2 5 3 2

2 commute

```
commute = data.frame(
 "name" = c("takashi", "takashi", "takashi", "takashi", "takashi", "hanako", "hanako", "hanako", "hanako", "hanako"),
 "day" = c(1, 2, 3, 4, 5, 1, 2, 3, 4, 5),
 "time" = c(10, 13, 12, 11, 14, 9, 15, 14, 10, 16)
)
```

commute

```
name day time
1 takashi 1 10
2 takashi 2 13
3 takashi 3 12
4 takashi 4 11
5 takashi 5 14
6 hanako 1 9
7 hanako 2 15
8 hanako 3 14
9 hanako 4 10
10 hanako 5 16
```

shoes

```
shoes = data.frame(
 "day" = c(1, 2, 3, 4, 5),
 "shoes" = c("tabi", "tabi", "bare", "tabi", "bare")
)
```

shoes

```
day shoes
1 1 tabi
2 2 tabi
3 3 bare
4 4 tabi
5 5 bare
```

inner\_join() inner\_join( , by = " ") day

```
commute %>%
 inner_join(shoes, by = "day")
```

```
name day time shoes
1 takashi 1 10 tabi
```

```
2 takashi 2 13 tabi
3 takashi 3 12 bare
4 takashi 4 11 tabi
5 takashi 5 14 bare
6 hanako 1 9 tabi
7 hanako 2 15 tabi
8 hanako 3 14 bare
9 hanako 4 10 tabi
10 hanako 5 16 bare
```

2

---

shoes

```
shoes = data.frame(
 "hinichi" = c(1,2,3,4,5), # "day" "hinichi"
 "shoes" = c("tabi","tabi","bare","tabi","bare")
)
```

shoes

```
hinichi shoes
1 1 tabi
2 2 tabi
3 3 bare
4 4 tabi
5 5 bare
```

```
commute %>%
 inner_join(shoes, by = c("day" = "hinichi")) # day hinichi
```

```
name day time shoes
1 takashi 1 10 tabi
2 takashi 2 13 tabi
3 takashi 3 12 bare
4 takashi 4 11 tabi
5 takashi 5 14 bare
6 hanako 1 9 tabi
7 hanako 2 15 tabi
8 hanako 3 14 bare
9 hanako 4 10 tabi
10 hanako 5 16 bare
```

left\_join() right\_join()

<https://dplyr.tidyverse.org/reference/join.html>

## 9.6

X, Y, Z 3

```
set.seed(1)
stocks = data.frame(
 time = as.Date('2009-01-01') + 0:9,
 X = rnorm(10, 0, 1),
```

```
Y = rnorm(10, 0, 2),
Z = rnorm(10, 0, 4)
)
```

```
stocks
```

```
time X Y Z
1 2009-01-01 -0.6264538 3.02356234 3.6759095
2 2009-01-02 0.1836433 0.77968647 3.1285452
3 2009-01-03 -0.8356286 -1.24248116 0.2982599
4 2009-01-04 1.5952808 -4.42939977 -7.9574068
5 2009-01-05 0.3295078 2.24986184 2.4793030
6 2009-01-06 -0.8204684 -0.08986722 -0.2245150
7 2009-01-07 0.4874291 -0.03238053 -0.6231820
8 2009-01-08 0.7383247 1.88767242 -5.8830095
9 2009-01-09 0.5757814 1.64244239 -1.9126002
10 2009-01-10 -0.3053884 1.18780264 1.6717662
```

```
3
```

```
gather() gather(key = " ", value = " ",)
```

```
stocks_long = stocks %>%
 gather(key = stock, value = price, X, Y, Z) # X:Z OK
```

```
stocks_long
```

```
time stock price
1 2009-01-01 X -0.62645381
2 2009-01-02 X 0.18364332
3 2009-01-03 X -0.83562861
4 2009-01-04 X 1.59528080
5 2009-01-05 X 0.32950777
6 2009-01-06 X -0.82046838
7 2009-01-07 X 0.48742905
8 2009-01-08 X 0.73832471
9 2009-01-09 X 0.57578135
10 2009-01-10 X -0.30538839
11 2009-01-01 Y 3.02356234
12 2009-01-02 Y 0.77968647
13 2009-01-03 Y -1.24248116
14 2009-01-04 Y -4.42939977
15 2009-01-05 Y 2.24986184
16 2009-01-06 Y -0.08986722
17 2009-01-07 Y -0.03238053
18 2009-01-08 Y 1.88767242
19 2009-01-09 Y 1.64244239
20 2009-01-10 Y 1.18780264
21 2009-01-01 Z 3.67590949
22 2009-01-02 Z 3.12854520
23 2009-01-03 Z 0.29825993
24 2009-01-04 Z -7.95740678
25 2009-01-05 Z 2.47930299
26 2009-01-06 Z -0.22451496
27 2009-01-07 Z -0.62318203
28 2009-01-08 Z -5.88300954
```

```
29 2009-01-09 Z -1.91260022
30 2009-01-10 Z 1.67176624
```

---

```
spread() spread(key = , value =)
stocks_long %>%
 spread(stock, price)
```

```
time X Y Z
1 2009-01-01 -0.6264538 3.02356234 3.6759095
2 2009-01-02 0.1836433 0.77968647 3.1285452
3 2009-01-03 -0.8356286 -1.24248116 0.2982599
4 2009-01-04 1.5952808 -4.42939977 -7.9574068
5 2009-01-05 0.3295078 2.24986184 2.4793030
6 2009-01-06 -0.8204684 -0.08986722 -0.2245150
7 2009-01-07 0.4874291 -0.03238053 -0.6231820
8 2009-01-08 0.7383247 1.88767242 -5.8830095
9 2009-01-09 0.5757814 1.64244239 -1.9126002
10 2009-01-10 -0.3053884 1.18780264 1.6717662
```

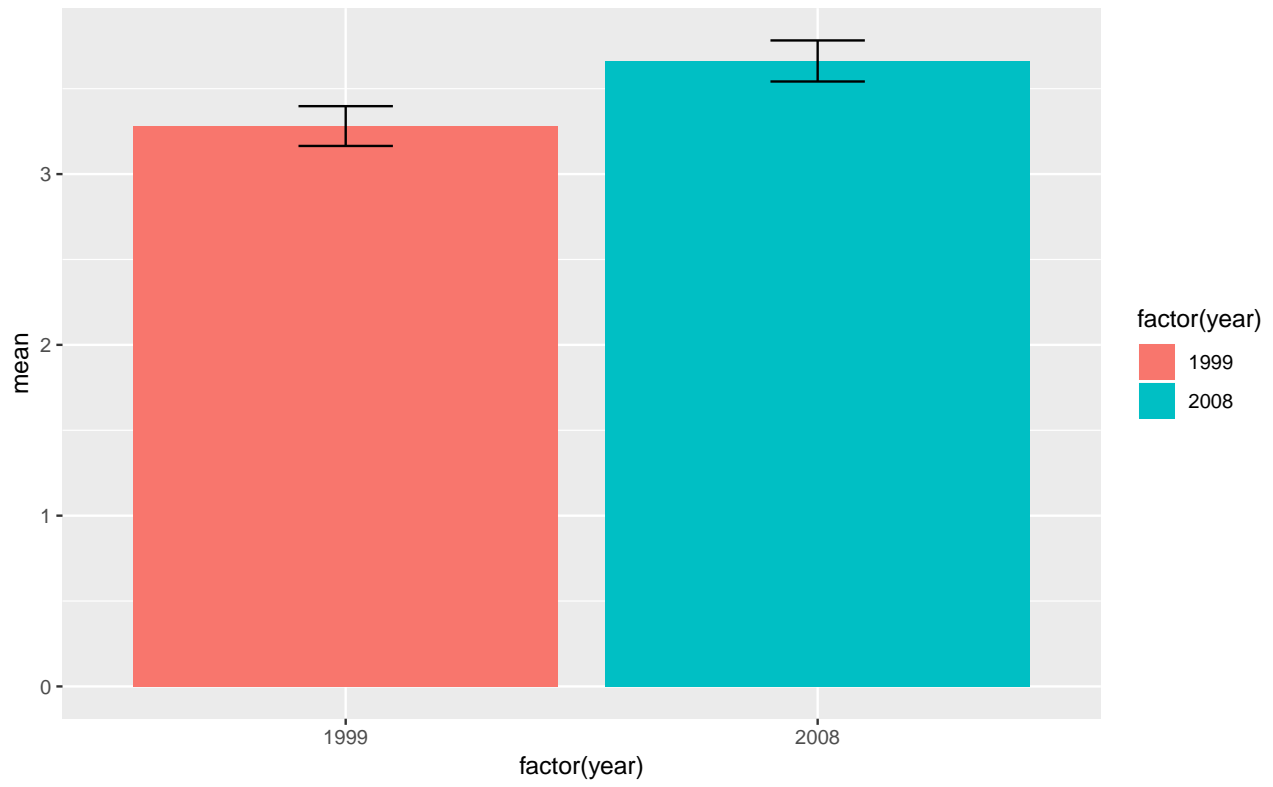
## 9.7

```
csv tsv xlsx read_csv() read_tsv() readxl read_excel()
```

2

## 9.8 ggplot2

```
%>%
mpg %>%
 group_by(year) %>% # year
 summarise(mean = mean(displ),
 se = sd(displ) / sqrt(n())) %>% # n()
 ggplot(aes(factor(year), mean, fill = factor(year))) + # data %>%
 geom_bar(stat = "identity") +
 geom_errorbar(aes(ymin = mean - se, ymax = mean + se), width = 0.2)
```











- Winston Chang. *Cookbook for R*. <http://www.cookbook-r.com/Graphs/>
  - ggplot2
  - 
  - *R ggplot2* <https://www.amazon.co.jp/R-ggplot2-Winston-Chang/dp/4873116538>
- Claus O. Wilke. *Fundamentals of Data Visualization*. <https://serialmentor.com/dataviz/>
  - 
  - 
  - GitHub
- Hadley Wickham. *Elegant Graphics for Data Analysis*. <https://github.com/hadley/ggplot2-book>
  - ggplot2
  - 
  - *ggplot2: Elegant Graphics for Data Analysis (Use R!)* [https://www.amazon.co.jp/ggplot2-Elegant-Graphics-Data-Analysis-Wickham/dp/331924275X/ref=pd\\_lpo\\_sbs\\_14\\_t\\_0?\\_encoding=UTF8&psc=1&refRID=2MNMAP5V2NFH89YZG5AR](https://www.amazon.co.jp/ggplot2-Elegant-Graphics-Data-Analysis-Wickham/dp/331924275X/ref=pd_lpo_sbs_14_t_0?_encoding=UTF8&psc=1&refRID=2MNMAP5V2NFH89YZG5AR)
- Kieran Healy. *Data Visualization: A Practical Introduction*. <https://socviz.co/index.html#preface>
  - ggplot2 ※
  - *Fundamentals of Data Visualization*
  - *Data Visualization: A Practical Introduction*. <https://www.amazon.co.jp/Data-Visualization-Introduction-Kieran-Healy/dp/0691181624>



## sessionInfo()

```
R version 3.5.2 (2018-12-20)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS Mojave 10.14.3
##
Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
locale:
[1] ja_JP.UTF-8/ja_JP.UTF-8/ja_JP.UTF-8/C/ja_JP.UTF-8/ja_JP.UTF-8
##
attached base packages:
[1] stats graphics grDevices utils datasets methods base
##
other attached packages:
[1] patchwork_0.0.1 GGally_1.4.0 gcookbook_2.0
[4] bindrcpp_0.2.2 formattable_0.2.0.1 forcats_0.3.0
[7] stringr_1.3.1 dplyr_0.7.8 purrr_0.3.0
[10] readr_1.3.1 tidyr_0.8.2 tibble_2.0.1
[13] ggplot2_3.1.0 tidyverse_1.2.1
##
loaded via a namespace (and not attached):
[1] tidyselect_0.2.5 xfun_0.4 reshape2_1.4.3
[4] haven_1.1.2 lattice_0.20-38 colorspace_1.4-0
[7] generics_0.0.2 htmltools_0.3.6 yaml_2.2.0
[10] utf8_1.1.4 rlang_0.3.1 pillar_1.3.1
[13] glue_1.3.0 withr_2.1.2 RColorBrewer_1.1-2
[16] modelr_0.1.2 readxl_1.1.0 bindr_0.1.1
[19] plyr_1.8.4 munsell_0.5.0 gtable_0.2.0
[22] cellranger_1.1.0 rvest_0.3.2 htmlwidgets_1.3
[25] evaluate_0.12 labeling_0.3 knitr_1.21
[28] fansi_0.4.0 highr_0.7 broom_0.5.1
[31] Rcpp_1.0.0 scales_1.0.0 backports_1.1.3
[34] jsonlite_1.6 hms_0.4.2 digest_0.6.18
[37] stringi_1.2.4 bookdown_0.9 grid_3.5.2
[40] cli_1.0.1 tools_3.5.2 magrittr_1.5
[43] lazyeval_0.2.1 crayon_1.3.4 pkgconfig_2.0.2
[46] xml2_1.2.0 lubridate_1.7.4 reshape_0.8.7
[49] assertthat_0.2.0 rmarkdown_1.11 httptr_1.4.0
[52] rstudioapi_0.9.0 R6_2.3.0 nlme_3.1-137
```

```
[55] compiler_3.5.2
```