R

Last Update: 2021-10-19

Contents

Chapter 1

R

- R
- Word R Word

R

1.1

- R, RStudio
- RStudio
-
-
-
- t
-

Chapter 2 R RStudio

R

- 2018 RStudio
- 2018 R RStudio[ ] −tidyverse −

1.2

- 2021/10/19 YouTube
- 2020/09/03

1.3

-
-
  -
- tidyverse %>%
  - tidyverse

Chapter 2

R(Studio)

R(Studio) Windows

http://yukiyanai.github.io/jp/resources/

Windows

-
  - 1
  - 4 Rtools
  - RStudio
- →
- OneDrive →

  macOS

-
  - RStudio

Chapter 3

RStudio

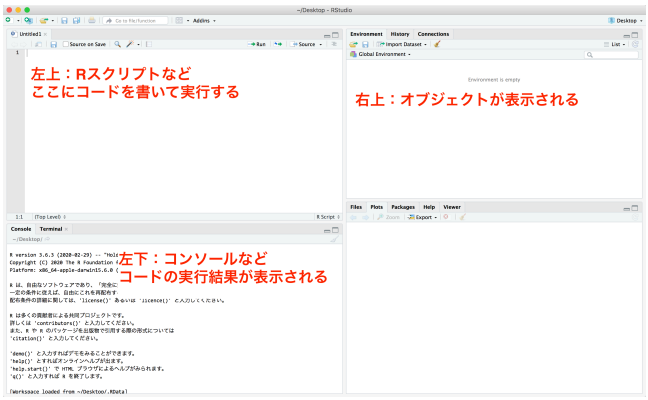
3.1 RStudio

RStudio Windows Mac
R OK

RStudio

3.2 RStudio

RStudio

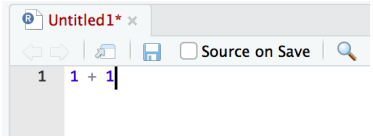


RStudio 4

- RStudio R Script
-
-
-

3.3 RStudio

R `1 + 1`

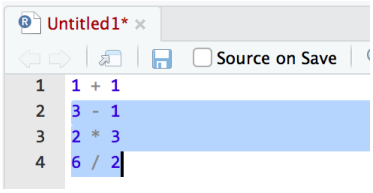


1 Windows Ctrl + Enter Mac Command + Enter Run

```
> 1 + 1
[1] 2
```

> 1 + 1 [1]

```
3 - 1
2 * 3
6 / 2
```



Ctrl + Enter (Command + Enter)

```
> 3 - 1
[1] 2
> 2 * 3
[1] 6
> 6 / 2
[1] 3
```

→

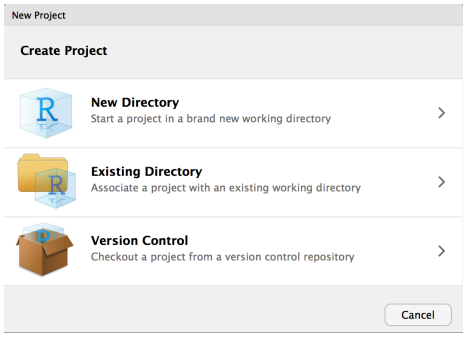
3.4

Windows Ctrl + S Mac Command + S

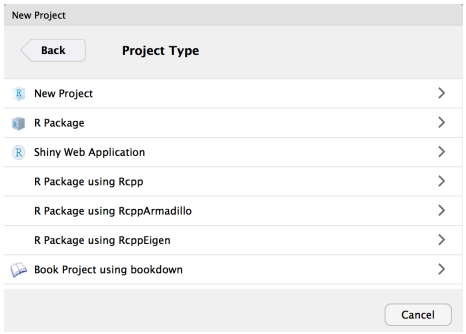
3.5

RStudio

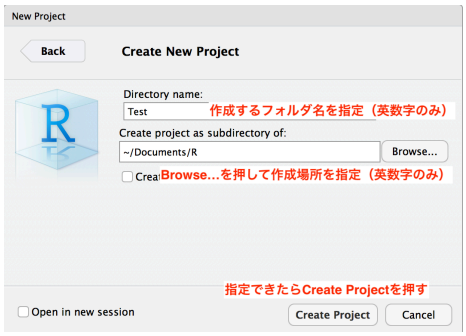
“New Directory”



“New Project”



Create Project



[.Rproj]

Test Test.Rproj

RStudio

- .Rproj
-
- RStudio

R R .R

Chapter 4

R tidyverse

```
install.packages("tidyverse")
library(tidyverse)
```

install.packages() OK library() RStudio R library()

mtcars mpg

```
mtcars %>% summarize(mean_mpg = mean(mpg))
```

```
##   mean_mpg
## 1 20.09062
```

R(Studio) Chapter 2

```
library(tidyverse)
```

Chapter 5

R

5.1

R object R

```
first_object <- 1
```

first_object 1 <-
OK

```
first_object
```

```
## [1] 1
```

[1] 1 1

() second_object 2

```
(second_object <- 2)
```

```
## [1] 2
```

””

```
first_string_object <- "Ritsumeikan University"
first_string_object
```

```
## [1] "Ritsumeikan University"
```

first_string_object "Ritsumeikan University"

1 c()

```
first_vector_object <- c(1, 2, 3, 4, 5)
first_vector_object
```

```
## [1] 1 2 3 4 5
```

first_vector_object 1 5

- third_object 10000
- my_name
- second_vector_object 1, 1, 2, 3, 5, 8

5.2

R +, -

```
1 + 1
```

```
## [1] 2
```

```
5 - 2
```

```
## [1] 3
```

* / ^ Excel

```
2 * 3
```

```
## [1] 6
```

```
10 / 2
```

```
## [1] 5
```

```
4 ^ 2
```

```
## [1] 16
```

'age' 10

```
age <- 20
age + 10
```

```
## [1] 30
```

1 + 1 one_plus_one

```
one_plus_one <- 1 + 1
one_plus_one
```

```
## [1] 2
```

one_plus_one 2

- a 3 b 4
- a, b 2 25

5.3

R function ( ) argument

5.3.1

R sqrt()

```
sqrt(4)
```

```
## [1] 2
```

sqrt() 4 4 2 log()

- a 3 b 4
- a, b 2 5
  – sqrt()

5.3.2

age_vector <- c(18, 21, 22, 23, 34)

min()

```
age_vector <- c(18, 21, 22, 23, 34)
min(age_vector)
```

```
## [1] 18
```

18 max()

mean() median() sd()

```
mean(age_vector)
```

```
## [1] 23.6
```

```
median(age_vector)
```

```
## [1] 22
```

```
sd(age_vector)
```

```
## [1] 6.107373
```

- income_vector 10, 100, 1000, 10000, 100000
- income_vector

5.4

R

age_vector gender_vector data.frame()

```
age <- c(18, 21, 22, 23, 34) #
gender <- c("female", "male", "male", "female", "female") #
first_dataframe <- data.frame(age, gender)
first_dataframe
```

```
##   age gender
## 1  18 female
## 2  21   male
## 3  22   male
## 4  23 female
## 5  34 female
```

1 18 2 21 ... Chapter @ref(#ImportData) Excel R

$ first_dataframe

```
first_dataframe$gender
```

```
## [1] "female" "male"   "male"   "female" "female"
```

```
mean(first_dataframe$age)
```

```
## [1] 23.6
```

- income 10, 100, 1000, 10000, 100000
- city "ibaraki", "takatsuki", "ibaraki", "takatsuki", "takatsuki"
- income city income_data
- income_data income

Chapter 6

RStudio

6.1

Chapter 3

CSV .csv Excel .xlsx, .xls

1

6.2

6.2.1 CSV .csv

CSV read_csv sotsuron.csv

```
data_original <- read_csv("sotsuron.csv")
```

csv data_original data_original

6.2.2 Excel .xlsx

Excel readxl

```
install.packages("readxl")
library(readxl)
```

read_excel

```
data_original <- read_excel("sotsuron.xlsx")
```

6.2.3 Qualtrics

Qualtrics qualtRics

```
install.packages("qualtRics")
library(qualtRics)
```

Qualtrics CSV CSV read_survey

```
data_original <- read_survey("sotsuron.csv")
```

Qualtrics

6.3

- “Data” data_original
  –
- head(data_original)
- str(data_original)

```
library(wooldridge)
```

Chapter 7

7.1

2

←

←

7.2

-
-

5 1 5

-
- 5
- 1, 2 3, 4, 5

7.3

wooldridge[1] saving

```
install.packages("wooldridge")
library(wooldridge)
```

data()

```
data("saving")
```

head()

```
head(saving)
```

```
##    sav   inc size educ age black  cons
## 1   30  1920    4    2  40     1  1890
## 2  874 12403    4    9  33     0 11529
## 3  370  6396    2   17  31     0  6026
## 4 1200  7005    3    9  50     0  5805
## 5  275  6990    4   12  28     0  6715
## 6 1400  6500    4   13  33     0  5100
```

1980

- sav:
- inc:
- size:
- educ:
- age:
- black:
- cons:

-
-
-

```
library(tidyverse)
library(wooldridge)
data("saving")
```

[1] Wooldridge "Introductory Econometrics: A Modern Approach"

Chapter 8

… 8

dplyr dplyr tidyverse tidyverse OK

8.1 %>%

%>% magrittr tidyverse

```
saving %>% head()
```

```
##    sav   inc size educ age black  cons
## 1   30  1920    4    2  40     1  1890
## 2  874 12403    4    9  33     0 11529
## 3  370  6396    2   17  31     0  6026
## 4 1200  7005    3    9  50     0  5805
## 5  275  6990    4   12  28     0  6715
## 6 1400  6500    4   13  33     0  5100
```

Chapter head(saving)

- : saving
- : %>%
- : head()

```
saving$sav %>% mean()
```

```
## [1] 1582.51
```

saving sav saving$sav mean() %>%

- 100
- saving inc

8.2

-
-
-
-
-

dplyr mutate()

8.2.1

mutate() sav inc saving_rate

```
saving_with_rate <-
  saving %>%
    mutate(saving_rate = sav / inc)

head(saving_with_rate)
```

```
##    sav   inc size educ age black  cons saving_rate
## 1   30  1920    4    2  40     1  1890  0.01562500
## 2  874 12403    4    9  33     0 11529  0.07046682
## 3  370  6396    2   17  31     0  6026  0.05784866
## 4 1200  7005    3    9  50     0  5805  0.17130621
## 5  275  6990    4   12  28     0  6715  0.03934192
## 6 1400  6500    4   13  33     0  5100  0.21538462
```

1 2 saving_with_rate 2 saving mutate()
3 mutate / saving_rate

head(saving_with_rate) 6

1 saving saving

- saving age age_squared

- saving inc inc_yen
  – 1 =140

8.2.2

mutate 1, 2, 3, 4, 5 5, 4, 3, 2, 1 5

- →(1, 2, 3, 4, 5) (-1, -2, -3, -4, -5)
- 6 →(-1, -2, -3, -4, -5) (5, 4, 3, 2, 1)

6

saving

```
data <- data.frame(Q1 = c(3, 2, 4, 1, 5)) #  Q1

data_gyakuten <-
  data %>%
    mutate(Q1_gyakuten = - Q1 + 6)

data_gyakuten
```

```
##   Q1 Q1_gyakuten
## 1  3           3
## 2  2           4
## 3  4           2
## 4  1           5
## 5  5           1
```

- 7 1 7

8.2.3

0 1 scale() mutate() educ educ_standardized

```
saving_standardized_educ <-
  saving %>%
    mutate(educ_standardized = scale(educ))

head(saving_standardized_educ)
```

```
##    sav   inc size educ age black  cons educ_standardized
## 1   30  1920    4    2  40     1  1890        -2.7886549
## 2  874 12403    4    9  33     0 11529        -0.7510156
## 3  370  6396    2   17  31     0  6026         1.5777150
## 4 1200  7005    3    9  50     0  5805        -0.7510156
## 5  275  6990    4   12  28     0  6715         0.1222584
```

```
## 6 1400  6500    4   13  33     0  5100         0.4133497
```

- `inc` `inc_standardized`

8.3

8.3.1

R `str()`

```
str(saving)
```

```
## 'data.frame':    100 obs. of  7 variables:
##  $ sav  : int  30 874 370 1200 275 1400 3159 1766 3984 1017 ...
##  $ inc  : int  1920 12403 6396 7005 6990 6500 26007 15363 14999 9185 ...
##  $ size : int  4 4 2 3 4 4 5 5 5 5 ...
##  $ educ : int  2 9 17 9 12 13 17 16 9 16 ...
##  $ age  : int  40 33 31 50 28 33 36 44 48 31 ...
##  $ black: int  1 0 0 0 0 0 0 0 1 0 ...
##  $ cons : int  1890 11529 6026 5805 6715 5100 22848 13597 11015 8168 ...
##  - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

$ `int` `int` (integer)

- int: integer,
- dbl: double,
- num: numeric,

8.3.2

fct (factor) `saving` `as_xxx()` xxx
`as_factor()` integer `black`

```
saving_with_factor <-
  saving %>%
    mutate(black_factor = as_factor(black))

str(saving_with_factor)
```

```
## 'data.frame':    100 obs. of  8 variables:
##  $ sav         : int  30 874 370 1200 275 1400 3159 1766 3984 1017 ...
##  $ inc         : int  1920 12403 6396 7005 6990 6500 26007 15363 14999 9185 ...
##  $ size        : int  4 4 2 3 4 4 5 5 5 5 ...
##  $ educ        : int  2 9 17 9 12 13 17 16 9 16 ...
##  $ age         : int  40 33 31 50 28 33 36 44 48 31 ...
##  $ black       : int  1 0 0 0 0 0 0 0 1 0 ...
##  $ cons        : int  1890 11529 6026 5805 6715 5100 22848 13597 11015 8168 ...
```

```
##  $ black_factor: Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 2 1 ...
##  - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

mutate black_factor str() Factor

as_factor

(chr, character) as_factor

- size

8.3.3

12 2 2 if_else
if_else() if_else(, ,)

```
table(saving$educ) #
```

```
##
##  2  3  4  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  1  1  1  1  4 10 11  9  4 32  2  4  1  9  6  2  1  1
```

```
saving_with_hsdummy <-
  saving %>%
    mutate(highschool = if_else(educ >= 12, 1, 0))

head(saving_with_hsdummy)
```

```
##    sav   inc size educ age black  cons highschool
## 1   30  1920    4    2  40     1  1890          0
## 2  874 12403    4    9  33     0 11529          0
## 3  370  6396    2   17  31     0  6026          1
## 4 1200  7005    3    9  50     0  5805          0
## 5  275  6990    4   12  28     0  6715          1
## 6 1400  6500    4   13  33     0  5100          1
```

```
table(saving_with_hsdummy$highschool) #
```

```
##
##  0  1
## 42 58
```

highschool 12 1 12 0 as_factor()

```
saving_with_hsdummy <-
  saving %>%
    mutate(highschool = if_else(educ >= 12, 1, 0),
           highschool = as_factor(highschool)) #highschool factor
```

```
head(saving_with_hsdummy)
```

```
##    sav   inc size educ age black  cons highschool
## 1   30  1920    4    2  40     1  1890          0
## 2  874 12403    4    9  33     0 11529          0
## 3  370  6396    2   17  31     0  6026          1
## 4 1200  7005    3    9  50     0  5805          0
## 5  275  6990    4   12  28     0  6715          1
## 6 1400  6500    4   13  33     0  5100          1
```

- 40 over40

8.3.4

2 if_else 3 case_when age case_when case_when(A ~ , B ~ ...

```
summary(saving$age) #
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   26.00   33.00   38.50   38.77   44.00   54.00
```

```
saving_with_age_category <-
  saving %>%
    mutate(age_category = case_when(age < 30 ~ "20s",
                                    age >= 30 & age < 40 ~ "30s",
                                    age >= 40 & age < 50 ~ "40s",
                                    age >= 50 ~ "50s"
                                    )
           )

head(saving_with_age_category)
```

```
##    sav   inc size educ age black  cons age_category
## 1   30  1920    4    2  40     1  1890          40s
## 2  874 12403    4    9  33     0 11529          30s
## 3  370  6396    2   17  31     0  6026          30s
## 4 1200  7005    3    9  50     0  5805          50s
## 5  275  6990    4   12  28     0  6715          20s
## 6 1400  6500    4   13  33     0  5100          30s
```

chr(character) as_factor

- 6000 “poor” 6000 12000 “middle” 12000 “rich” inc_category

8.4

dplyr select()

saving inc age

```
saving_selected <-
  saving %>%
    select(inc, age)

head(saving_selected)
```

```
##     inc age
## 1  1920  40
## 2 12403  33
## 3  6396  31
## 4  7005  50
## 5  6990  28
## 6  6500  33
```

2

- saving cons

```
saving_deleted <-
  saving %>%
    select(-cons)

head(saving_deleted)
```

```
##    sav   inc size educ age black
## 1   30  1920    4    2  40     1
## 2  874 12403    4    9  33     0
## 3  370  6396    2   17  31     0
## 4 1200  7005    3    9  50     0
## 5  275  6990    4   12  28     0
## 6 1400  6500    4   13  33     0
```

- sav size black 3
- educ age

8.5

dplyr arrange() saving inc

```
saving_arranged <-
  saving %>%
    arrange(inc)
```

```
head(saving_arranged)
```

```
##    sav  inc size educ age black cons
## 1    0  750    2    4  49     0  750
## 2   30 1920    4    2  40     1 1890
## 3   50 2340    2    6  46     1 2290
## 4 -112 2936    7   10  39     0 3048
## 5 2575 3941    4    9  34     0 1366
## 6 2483 4091    6    8  44     0 1608
```

desc()

```
saving_arranged_desc <-
  saving %>%
    arrange(desc(inc))

head(saving_arranged_desc)
```

```
##     sav   inc size educ age black  cons
## 1  1800 32080    2   16  54     0 30280
## 2 10668 30996    4   12  41     0 20328
## 3  4115 30610    4   16  44     0 26495
## 4  3159 26007    5   17  36     0 22848
## 5 -2749 24226    5   17  44     0 26975
## 6  5082 19362    3   11  48     0 14280
```

8.6

%>%

- saving_rate
- size
- inc

```
saving_handled <-
  saving %>%
    mutate(saving_rate = sav / inc) %>%
    select(-size) %>%
    arrange(desc(inc))

head(saving_handled)
```

```
##     sav   inc educ age black  cons saving_rate
## 1  1800 32080   16  54     0 30280  0.05610973
## 2 10668 30996   12  41     0 20328  0.34417344
## 3  4115 30610   16  44     0 26495  0.13443319
## 4  3159 26007   17  36     0 22848  0.12146730
```

```
## 5 -2749 24226   17  44     0 26975 -0.11347313
## 6  5082 19362   11  48     0 14280  0.26247289
```

```
       #
saving_handled
  saving %>%
    mutate(saving_rate = sav / inc) %>% #
    select(-size) %>% #
    arrange(desc(inc)) #
```

Chapter 9

summarytools

```
install.packages("summarytools")
library(summarytools)
```

Chapter Word Chapter ??Word)

9.1

summarytools descr() saving

```
saving %>%
  descr()
```

```
## Descriptive Statistics
## saving
## N: 100
##
##                       age    black        cons     educ        inc         sav     size
## ----------------- -------- -------- ----------- -------- ---------- ---------- --------
##              Mean    38.77     0.07     8358.73    11.58    9941.24    1582.51     4.35
##           Std.Dev     7.40     0.26     5729.53     3.44    5584.00    3284.90     1.49
##               Min    26.00     0.00   -13055.00     2.00     750.00   -5577.00     2.00
##                Q1    33.00     0.00     5726.00     9.00    6508.00     189.00     3.00
##            Median    38.50     0.00     7561.50    12.00    8776.50     982.00     4.00
##                Q3    44.00     0.00     9987.00    13.00   11965.00    1838.50     5.00
##               Max    54.00     1.00    30280.00    20.00   32080.00   25405.00    10.00
##               MAD     8.15     0.00     3092.70     2.97    3463.35    1235.75     1.48
##               IQR    11.00     0.00     4131.50     4.00    5393.00    1640.25     2.00
```

```
##                CV     0.19     3.66     0.69     0.30     0.56     2.08
##          Skewness     0.24     3.32     0.91     0.05     1.98     4.15
##       SE.Skewness     0.24     0.24     0.24     0.24     0.24     0.24
##          Kurtosis    -0.96     9.11     4.31     0.05     4.96    26.31
##           N.Valid   100.00   100.00   100.00   100.00   100.00   100.00
##         Pct.Valid   100.00   100.00   100.00   100.00   100.00   100.00
```

- stats
  - mean sd min max n.valid
- transpose TRUE
- heading FALSE

```
saving %>%
  descr(stats = c("mean", "sd", "min", "max", "n.valid"), transpose = TRUE, headings =
```

```
##
##                 Mean   Std.Dev          Min         Max   N.Valid
## ----------- --------- --------- ----------- ----------- ---------
##          age     38.77      7.40       26.00       54.00    100.00
##        black      0.07      0.26        0.00        1.00    100.00
##         cons   8358.73   5729.53   -13055.00   30280.00    100.00
##         educ     11.58      3.44        2.00       20.00    100.00
##          inc   9941.24   5584.00      750.00   32080.00    100.00
##          sav   1582.51   3284.90    -5577.00   25405.00    100.00
##         size      4.35      1.49        2.00       10.00    100.00
```

9.2

2

- descr()
  - black
- summarytools freq()

Chapter ??DataHandling) age_category

```
age_category <-
  saving %>%
    mutate(age_category = case_when(age < 30 ~ "20s",
                                    age >= 30 & age < 40 ~ "30s",
                                    age >= 40 & age < 50 ~ "40s",
                                    age >= 50 ~ "50s"
                                    )
           ) %>%
    select(age_category)
```

age_category freq()

```
age_category %>%
  freq()
```

```
## Frequencies
## age_category$age_category
## Type: Character
##
##               Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##         20s     12     12.00          12.00     12.00          12.00
##         30s     44     44.00          56.00     44.00          56.00
##         40s     31     31.00          87.00     31.00          87.00
##         50s     13     13.00         100.00     13.00         100.00
##        <NA>      0                                0.00         100.00
##       Total    100    100.00         100.00    100.00         100.00
```

```
age_category %>%
  freq(report.nas = FALSE, totals = FALSE, cumul = FALSE, headings = FALSE)
```

```
##
##             Freq       %
## --------- ------ -------
##       20s     12   12.00
##       30s     44   44.00
##       40s     31   31.00
##       50s     13   13.00
```

- report.nas = FALSE: NA
- totals = FALSE:
- cumul = FALSE:
- headings = FALSE:

Chapter 10

R ggplot2 ggplot2 tidyverse

ggplot2 ggplot2

```
     %>%
 ggplot(aes(x = x   , y = y   )) +
 geom_   ()
```

- 1
- 2 +
  - aes aesthetic
- 3
  - geom_bar:
  - geom_hitstogram:
  - geom_boxplot:
  - geom_point:
  - geom_smooth:

10.1 1

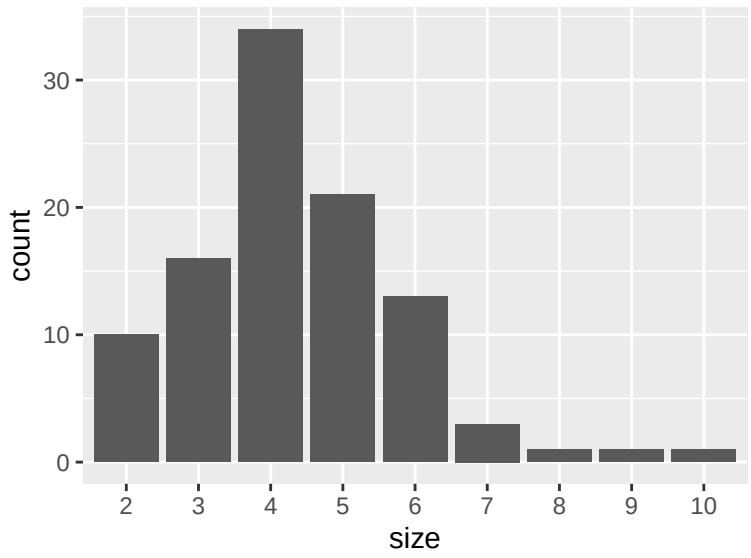
1

10.1.1

1 saving size

```
saving %>%
  mutate(size = as_factor(size)) %>% #size
```

```
  ggplot(aes(x = size)) +
  geom_bar()
```

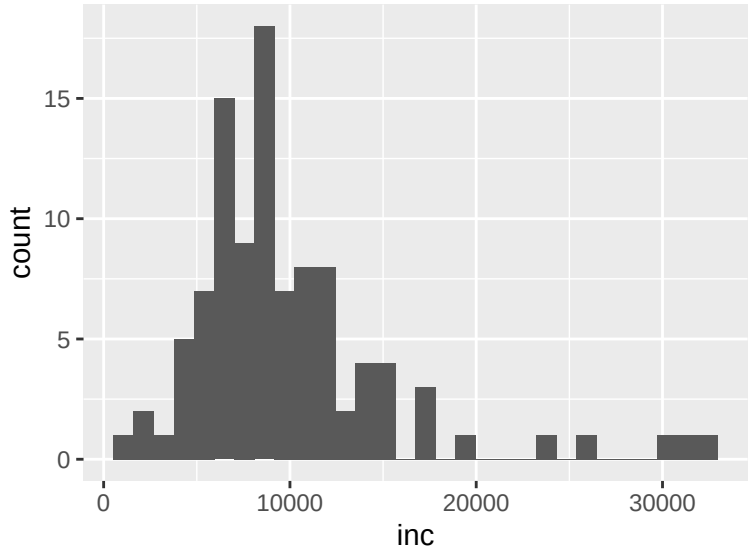


1 x y geom_bar() count 4 5 3 ...

10.1.2

1 saving inc

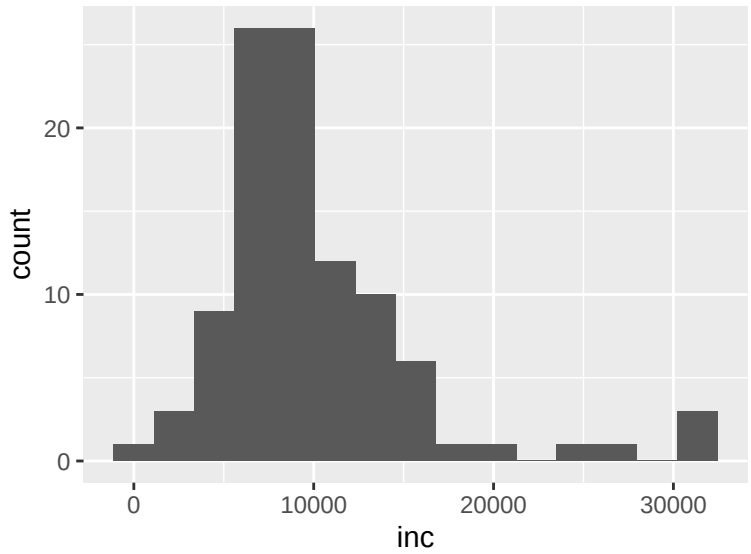
```
saving %>%
  ggplot(aes(x = inc)) +
  geom_histogram()
```



x geom_histogram bins 30

bins 15

```
saving %>%
  ggplot(aes(x = inc)) +
  geom_histogram(bins = 15) #   15
```



binwidth center boundary

10.2 2

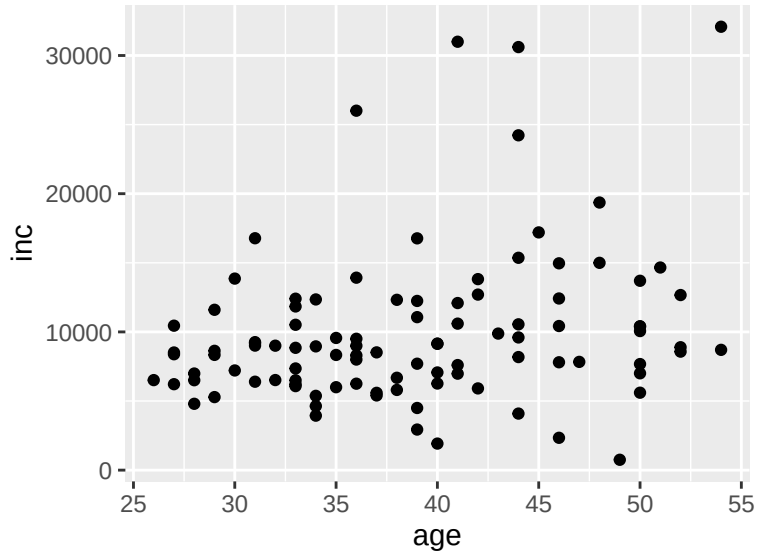
2

-
-
-

10.2.1

geom_point() age inc

```
saving %>%
  ggplot(aes(x = age, y = inc)) +
  geom_point()
```

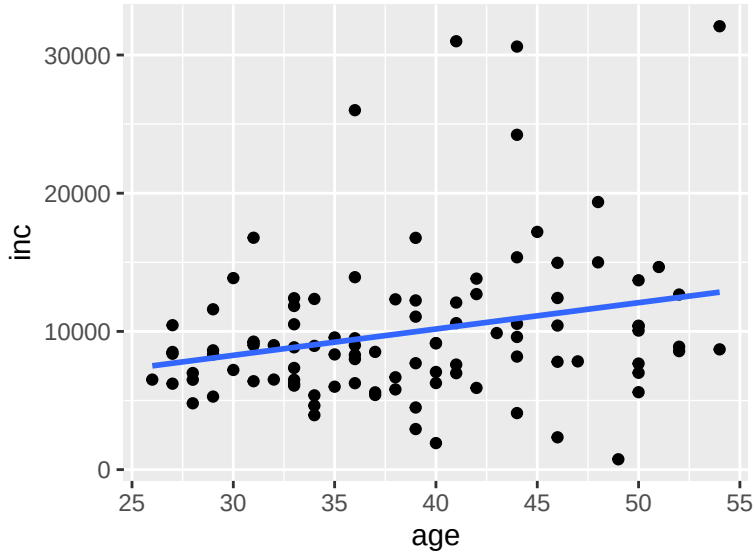


2 x y x y x y

geom_smooth

```
saving %>%
  ggplot(aes(x = age, y = inc)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

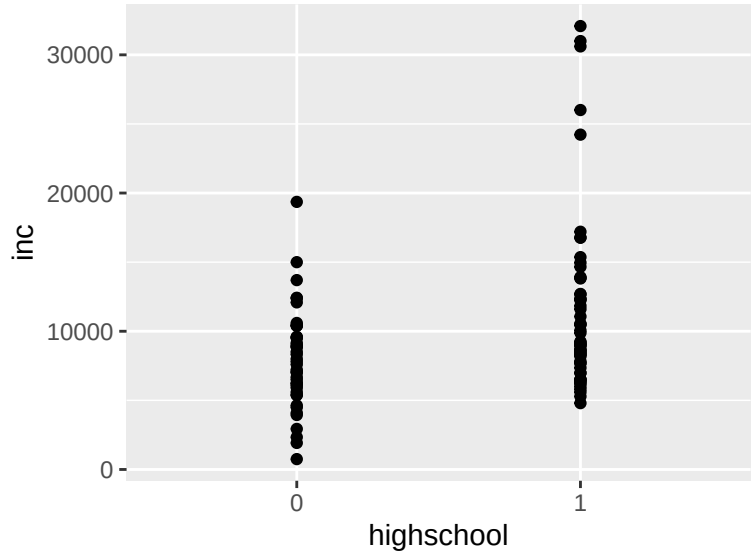
```
## `geom_smooth()` using formula 'y ~ x'
```



ggplot + geom_point() geom_smooth()
geom_smooth() method lm linear model
se FALSE

educ highschool inc

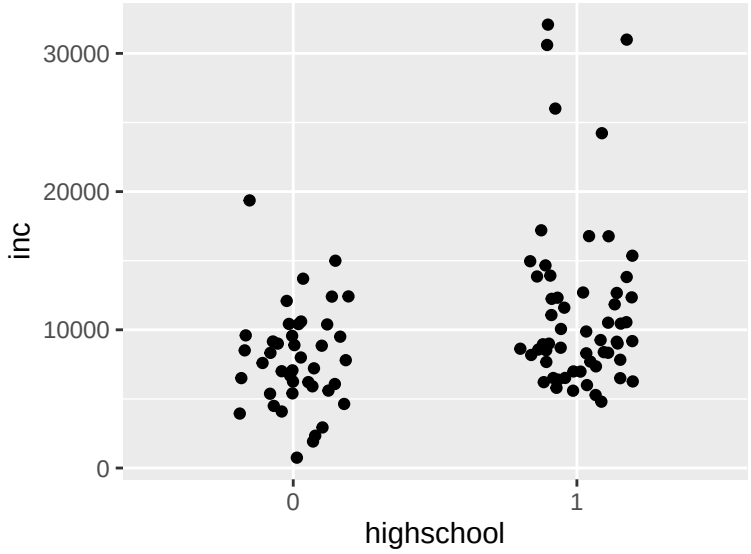
```
saving %>%
  mutate(highschool = if_else(educ >= 12, 1, 0),
         highschool = as_factor(highschool)) %>%
  ggplot(aes(x = highschool, y = inc)) +
  geom_point()
```



- geom_point() geom_jitter()
-

geom_jitter()

```
saving %>%
  mutate(highschool = if_else(educ >= 12, 1, 0),
         highschool = as_factor(highschool)) %>%
  ggplot(aes(x = highschool, y = inc)) +
  geom_jitter(width = 0.2)
```



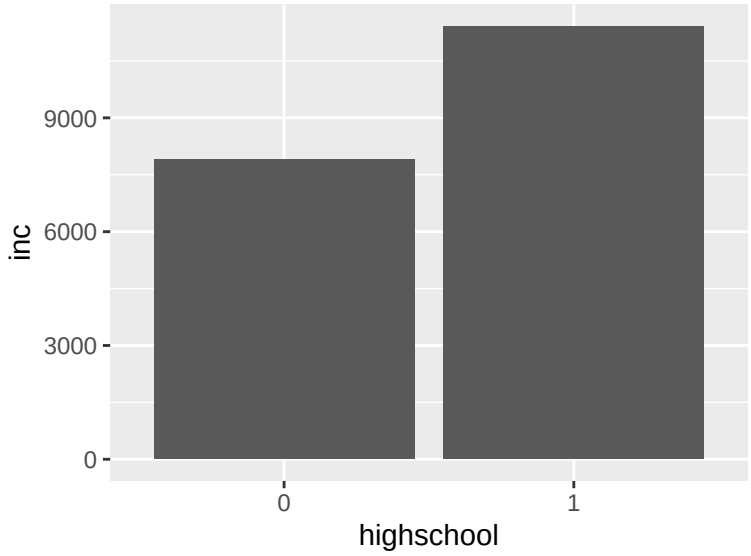
width = 0.2

10.2.2

geom_bar() stat_summary() stat_summary() fun.y mean geom

```
saving %>%
  mutate(highschool = if_else(educ >= 12, 1, 0),
         highschool = as_factor(highschool)) %>%
  ggplot(aes(x = highschool, y = inc)) +
  stat_summary(fun.y = "mean", geom = "bar")
```

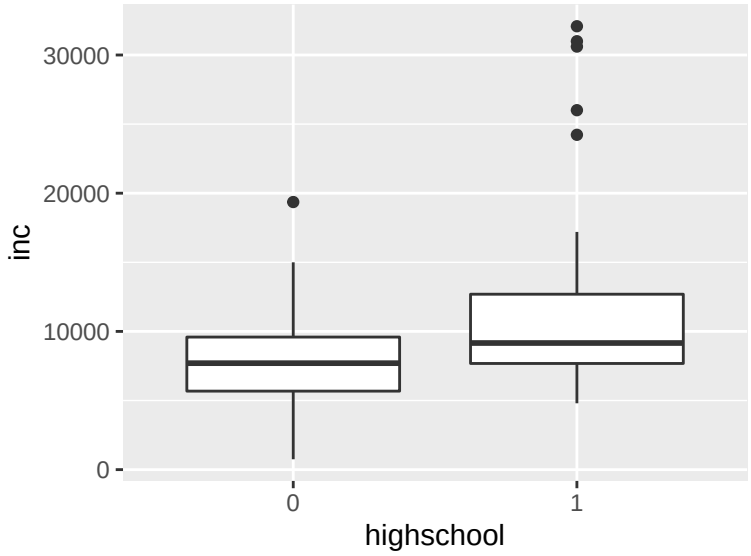
```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```



10.2.3

geom_boxplot

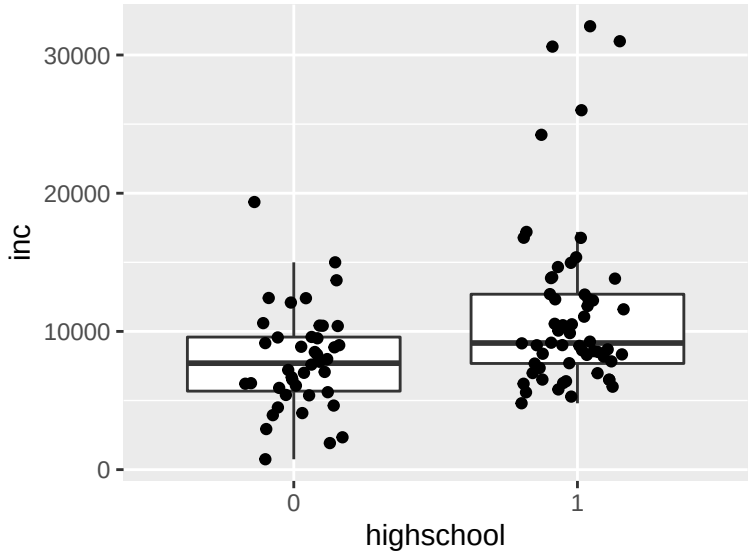
```
saving %>%
  mutate(highschool = if_else(educ >= 12, 1, 0),
         highschool = as_factor(highschool)) %>%
  ggplot(aes(x = highschool, y = inc)) +
  geom_boxplot()
```



- 25%
- 50%
- 75%
- 1.5×IQR -
-

geom_jitter() ggplot2 geom_xxxx +

```
saving %>%
  mutate(highschool = if_else(educ >= 12, 1, 0),
         highschool = as_factor(highschool)) %>%
  ggplot(aes(x = highschool, y = inc)) +
  geom_boxplot(outlier.shape = NA) + #
  geom_jitter(width = 0.2)
```


30000
20000
10000
0
inc
0
1
highschool

Chapter 11

t

2 t t t.test() t.test()

```
t.test(   ~   , data =     )
```

black inc

```
t.test(inc ~ black, data = saving)
```

```
##
##  Welch Two Sample t-test
##
## data:  inc by black
## t = 1.8562, df = 7.3906, p-value = 0.1036
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -890.170 7726.938
## sample estimates:
## mean in group 0 mean in group 1
##       10180.527        6762.143
```

t p 0.103

t.test() 2 Welch t var.equal = TRUE t Welch t

t broom tidy bloom

```
install.packages("broom")
```

tidy() t.test tidy()

```
library(broom)
t.test(inc ~ black, data = saving) %>%
  tidy()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##      <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl>
## 1    3418.    10181.     6762.      1.86   0.104      7.39    -890.     7727.
## # ... with 2 more variables: method <chr>, alternative <chr>
```

- estimate:
- estimate1: 1
- estimate2: 2
- statistic: t
- p.value: p
- parameter:
- conf.low, conf.high:
- method:
- alternative:

Chapter 12

Linear Regression

12.1

y x y x

$y = \alpha + \beta x$

y x β x y Coefficient x y
0 β x 1 y β

α β OLS R

12.2 R

R `y ~ x` Chapter 11 t `lm()` lm linear model

```
lm(  , data =     )
```

`inc` `educ` $inc = \alpha + \beta educ$

```
lm(inc ~ educ, data = saving)
```

```
##
## Call:
## lm(formula = inc ~ educ, data = saving)
##
## Coefficients:
## (Intercept)         educ
##      1342.7        742.5
```

Coefficients intercept α educ β β 1 742.5

`lm()` t p `lm()` `summary()`

```
lm(inc ~ educ, data = saving) %>%
  summary()
```

```
##
## Call:
## lm(formula = inc ~ educ, data = saving)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7570  -3297  -1288   1617  20743
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1342.7     1763.5   0.761    0.448
## educ           742.5      146.1   5.084 1.78e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4993 on 98 degrees of freedom
## Multiple R-squared:  0.2087, Adjusted R-squared:  0.2006
## F-statistic: 25.84 on 1 and 98 DF,  p-value: 1.777e-06
```

Coefficients

- `Estimate`:
- `Std. Error`:
- `t value`: t
- `Pr(>|t|)`: p

`educ` p 1.78e-06 $1.78 \times 10^{-6} = 1.78 \times 0.000001 = 0.00000178$

`educ` β

R^2

`summary()` Chapter 11 `broom` `tidy()`

```
lm(inc ~ educ, data = saving) %>%
  tidy()
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic    p.value
##   <chr>          <dbl>     <dbl>     <dbl>      <dbl>
## 1 (Intercept)    1343.     1764.     0.761 0.448
## 2 educ            743.      146.     5.08  0.00000178
```

- inc size

12.3

x_1, x_2

$y = \alpha + \beta_1 x_1 + \beta_2 x_2$

R y ~ x1 + x2 + inc educ size

```
lm(inc ~ educ + size, data = saving) %>%
  tidy()
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic    p.value
##   <chr>          <dbl>     <dbl>     <dbl>      <dbl>
## 1 (Intercept)    3027.     2283.      1.33 0.188
## 2 educ            743.      146.      5.10 0.00000171
## 3 size           -389.      335.     -1.16 0.249
```

3 4 … R y ~ x1 + x2 + x3 + x4 + ... +

age black

```
equation <- inc ~ educ + size + age + black
lm(equation, data = saving) %>%
  tidy()
```

```
## # A tibble: 5 x 5
##   term        estimate std.error statistic      p.value
##   <chr>          <dbl>     <dbl>     <dbl>        <dbl>
## 1 (Intercept)  -10005.    3934.     -2.54  0.0126
## 2 educ            857.     144.      5.97  0.0000000408
## 3 size           -101.     320.     -0.317 0.752
## 4 age             271.      66.3     4.09  0.0000917
## 5 black          -553.    1878.     -0.294 0.769
```

4

- sav educ size age

12.4

1 0

inc black

```
lm(inc ~ black, data = saving)
```

```
##
## Call:
## lm(formula = inc ~ black, data = saving)
##
## Coefficients:
## (Intercept)        black
##       10181        -3418
```

-3418 3418

12.5

-1 A, B, C 3

- B 1 0 B
- C 1 0 C

- B B A
- C C A

B C

R factor

age_category inc

```
saving_with_age_category <-
  saving %>%
    mutate(age_category = case_when(age < 30 ~ "20s",
                                    age >= 30 & age < 40 ~ "30s",
                                    age >= 40 & age < 50 ~ "40s",
                                    age >= 50 ~ "50s"
                                    )
           )

lm(inc ~ age_category, data = saving_with_age_category)
```

```
##
## Call:
## lm(formula = inc ~ age_category, data = saving_with_age_category)
##
## Coefficients:
##     (Intercept)  age_category30s  age_category40s  age_category50s
##            7685             1330             3761             3885
```

“20s” 20

- 30 20 1330
- 40 20 3761
- 50 20 3885

20 <30 <40 <50 40 50

12.6

R 2 :

R

```
lm(inc ~ educ + black + educ:black, data = saving) %>%
  tidy()
```

```
## # A tibble: 4 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   1595.      1926.    0.828  0.410
## 2 educ           727.       157.    4.63   0.0000115
## 3 black         -525.      5773.   -0.0909 0.928
## 4 educ:black     -63.1      615.   -0.103  0.918
```

educ 727 1 727

educ:black -63 727-63=664 1 664

p

*

```
lm(inc ~ educ*black, data = saving) %>%
  tidy()
```

```
## # A tibble: 4 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   1595.      1926.    0.828  0.410
## 2 educ           727.       157.    4.63   0.0000115
## 3 black         -525.      5773.   -0.0909 0.928
## 4 educ:black     -63.1      615.   -0.103  0.918
```

- inc age

12.7

`tidy()`

`stargazer`

2

1. `inc ~ educ`
2. `inc ~ educ + age`

`stargazer()`

```
regression1 <- lm(inc ~ educ, data = saving) #
regression2 <- lm(inc ~ educ + age, data = saving) #
stargazer(regression1, regression2, type = "html", out = "test.doc")
```

Dependent variable:

inc

(1)

(2)

educ

742.530***

869.852***

(146.062)

(137.566)

age

276.677***

(63.872)

Constant

1,342.745

-10,858.410***

(1,763.546)

(3,250.631)

Observations

100

100

R2

0.209

0.337

Adjusted R2

0.201

0.323

Residual Std. Error

4,992.593 (df = 98)

4,593.594 (df = 97)

F Statistic

25.844*** (df = 1; 98)

24.646*** (df = 2; 97)

Note:

p<0.1; ***p<0.05;*** p<0.01

1 2 `lm()` `regression1,regression2` `stargazer()`
`type` LaTeX HTML `out`

* *** 1% ** 5% * 10%

- `Observation`:
- R^2:
- `Adjusted` R^2:
- `Residual Std. Error`:
- `F Statistic`: F F

Chapter 13

Word

Word Word **R** **Word**

13.1

Word Excel

- R Excel
-
-
  – → →
- Word

R Word `flextable`