R

Yi-Ju Tseng

2017-03-25

Contents

R (API) (Shiny) R Hadoop Ecosystems

epub

R

```
install.packages("devtools")
devtools::install_github("yijutseng/DataAnalyticsWithRBook")
```

GitHub issue

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: macOS Sierra 10.12.3
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] backports_1.0.5 bookdown_0.3.9  magrittr_1.5    rprojroot_1.2
##  [5] tools_3.3.2     htmltools_0.3.5 rstudioapi_0.6  yaml_2.1.14
##  [9] Rcpp_0.12.9     stringi_1.1.2   rmarkdown_1.3   knitr_1.15.1
## [13] stringr_1.1.0   digest_0.6.12   evaluate_0.10
```

```
pkgInfo<-lapply(pkg, packageDescription, fields = c("Package", "Version"))
knitr::kable(data.frame(Package=sapply(pkgInfo, `[[`, 1),
  Version=sapply(pkgInfo, `[[`, 2)))
```

Package	Version
ggplot2	2.2.1
dplyr	0.5.0
lubridate	1.6.0
bit64	0.9-5
bookdown	0.3.9
knitr	1.15.1
rmarkdown	1.3
RCurl	1.95-4.8
data.table	1.10.0
stringr	1.1.0
reshape2	1.4.2
SportsAnalytics	0.2
readr	1.0.0
readxl	0.1.1
jsonlite	1.2
XML	3.98-1.5
Rfacebook	0.6.11
rvest	0.3.2
rgdal	1.2-5
rgeos	0.3-22
maptools	0.8-41
ggmap	2.6.1
choroplethr	3.5.3
choroplethrMaps	1.0.1
WDI	2.4
treemapify	0.2.2
shiny	1.0.0
plotly	4.5.6
ggvis	0.4.3
googleVis	0.6.2
rpart	4.1-10
rpart.plot	2.1.0
fields	8.10
arules	1.5-0
datasets	3.3.2
arulesViz	1.2-0
MASS	7.3-45
caret	6.0-73
purrr	0.2.2
treemap	2.4-1

Chapter 1

R 101

R

1.1 R

R 2000 R 1.0.0 R R **Packages** R (Standing on the shoulders of giants (Hal R. Varian, Google)) 2017 1 R 10,000 (R Studio) R Studio Quick list of useful R packages

Package

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

ggplot2 R Console

```
install.packages("ggplot2")
```

library( )

```
library(ggplot2)
```

1.2

R (1, 2,....) mean() :

```
mean(c(1,2,3,4,5,6)) ## 1~6
```

```
## [1] 3.5
```

?

```
?mean
```

seq() from, to, by

```
seq(from=1,to=9,by=2)#1~9  2
```

```
## [1] 1 3 5 7 9
```

```
seq(1,9,2)#
```

```
## [1] 1 3 5 7 9
```

```
seq(by=2,to=9,from=1)#
```

```
## [1] 1 3 5 7 9
```

1.3

R R R <- <- ( ) <- ->

```
a<-1
2->b
a
```

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

```
b
```

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

R = =

```
c=1
c
```

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

str() str()

```
d<-3
str(d)
```

```
##  num 3
```

:

- break, else, FALSE, for, function, if, Inf, NA, NaN, next, repeat, return, TRUE, while
- .
-

1.4

R (Console) > + R

1.5

R **(numeric)** **(character)** **(logic)** **(Date)**

1.5.1 numeric

```
num1<-100
num2<-1000.001
```

2^53 bit64 package (?) 2^63

```
print(2^53, digits=20)
```

```
## [1] 9007199254740992
```

```
print(2^53+1, digits=20) # +1   2^53
```

```
## [1] 9007199254740992
```

```
library(bit64) #  bit64 package
print(as.integer64(2)^53, digits=20)
```

```
## integer64
## [1] 9007199254740992
```

```
print(as.integer64(2)^53+1, digits=20)#  bit64
```

```
## integer64
## [1] 9007199254740993
```

1.5.2 character

"

```
char1<-"abcTest"
char2<-"100"
char3<-"200"
#char2+char3 #  Error message: non-numeric argument to binary operator
```

1.5.3 logic

TRUE T FALSE F

```
boolT<-TRUE
boolT1<-T
boolF<-FALSE
boolF1<-F
```

1.5.4 (Date)

Sys.Date()

```
dateBook<-Sys.Date()
dateBook
```

```
## [1] "2017-03-25"
```

lubridate(?) package / / ymd() y year m month d day / / mdy()

```
library(lubridate)
ymd('2012/3/3')
```

```
## [1] "2012-03-03"
```

```
mdy('3/3/2012')
```

```
## [1] "2012-03-03"
```

The Yhat Blog

1.6

1.6.1

R

- +
- -
- *
- /
- %%
- ^

```
num1<-1
num2<-100
num1+num2
```

```
## [1] 101
```

```
num1-num2
```

```
## [1] -99
```

```
num1*num2
```

```
## [1] 100
```

```
num1/num2
```

```
## [1] 0.01
```

```
100%%3 ##100 3
```

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

```
2^3 ##2 3
```

```
## [1] 8
```

1.6.2

- round()
- floor()
- ceiling()

```
num1<-1.568
num2<-2.121
round(num1,digits = 2) #
```

```
## [1] 1.6
```

```
round(num2,digits = 1) #
```

```
## [1] 2.1
```

```
floor(num1) ##1.568
```

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

```
ceiling(num2) ##2.121
```

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

1.6.3

R

- >
- <
- ==
- >=
- <=

```
num1<-1
num2<-100
num1>num2
```

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

```
num1<num2
```

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

```
char1<-"abcTest"
char2<-"defTest"
char1>char2
```

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

JAVA R & |

- &
- |

```
TRUE & TRUE
```

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

```
TRUE & FALSE
```

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

```
TRUE | TRUE
```

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

```
TRUE | FALSE
```

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

!

```
!TRUE
```

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

```
!FALSE
```

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

1.7

- Message
- Warning
- Error
- Condition

```
log(-1)
```

```
## Warning in log(-1): NaNs produced
```

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

```
mena(NA)
```

```
## Error in eval(expr, envir, enclos): could not find function "mena"
```

1:

```
# Error: could not find function "fetch_NBAPlayerStatistics"
#   "fetch_NBAPlayerStatistics" function
```

SportsAnalytics package

2:

```
# Error in library(knitr): there is no package called 'knitr'
#   "knitr" package
```

knitr package

1.8 Help

R R ? ?

```
?ggplot2
?ymd
```

Stack Overflow

/

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: macOS Sierra 10.12.3
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] lubridate_1.6.0 bit64_0.9-5     bit_1.1-12
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9     bookdown_0.3.9  digest_0.6.12   rprojroot_1.2
##  [5] backports_1.0.5 magrittr_1.5    evaluate_0.10   stringi_1.1.2
##  [9] rstudioapi_0.6  rmarkdown_1.3   tools_3.3.2     stringr_1.1.0
## [13] yaml_2.1.14     htmltools_0.3.5 knitr_1.15.1
```

Chapter 2

R

2.1 vector

c()

```
vec<-c('a','b','c','d','e')
```

a~e vec **(element)** a vec **1** b **2** vec **4**

```
vec[4] ## 4
```

```
## [1] "d"
```

```
vec[c(2,3)] ## 2 3
```

```
## [1] "b" "c"
```

vec

<-

```
vec[3]
```

```
## [1] "c"
```

```
vec[3]<-'z' ##     "z"
vec
```

```
## [1] "a" "b" "z" "d" "e"
```

2.1.1

1~20 :

```
1:20 ## c(1,2,...,19,20)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

seq()

```
seq(from=1,to=20,by=1) ##1~20  1
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

```
seq(from=1,to=50,by=2) ##1~50   2
```

```
##  [1]  1  3  5  7  9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
```

2.1.2

```
numvec<-1:10 ## c(1,2,3,4,5,6,7,8,9,10)
numvec+3 ##   +3
```

```
##  [1]  4  5  6  7  8  9 10 11 12 13
```

```
numvec*2 ##   *2
```

```
##  [1]  2  4  6  8 10 12 14 16 18 20
```

```
numvec1<-1:3 ## c(1,2,3)
numvec2<-4:6 ## c(4,5,6)
numvec1+numvec2
```

```
## [1] 5 7 9
```

```
numvec1*numvec2
```

```
## [1]  4 10 18
```

2.2 factor

factor(,levels=) levels

```
factor(c("  ","   ","   "),
       levels = c("  ","   ","   "))
```

```
## [1]
## Levels:
```

2.3 list

R **list** list()

```
listSample<-list(Students=c("Tom","Kobe","Emma","Amy"),Year=2017,
                 Score=c(60,50,80,40),School="CGU")
listSample
```

```
## $Students
## [1] "Tom"  "Kobe" "Emma" "Amy"
##
## $Year
## [1] 2017
##
## $Score
## [1] 60 50 80 40
```

```
##
## $School
## [1] "CGU"
```

2.3.1

$

```
listSample$Students ##   Students
```

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

[[ ]]

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

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

list

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

```
## $Students
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

2.3.2

```
listSample[[1]]
```

```
## [1] "Tom"  "Kobe" "Emma" "Amy"
```

```
listSample[[1]]<-c(" "," "," "," "," ") ## Students
listSample[[1]]
```

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

$ <-

```
listSample$Gender<-c("M","F","M","F","M") ## Gender
```

NULL

```
listSample$Score<-NULL ## Score
listSample
```

```
## $Students
## [1] " " " " " " " " " " " " " "
##
## $Year
## [1] 2017
##
## $School
## [1] "CGU"
##
## $Gender
## [1] "M" "F" "M" "F" "M"
```

2.4 matrix

```
a <- matrix(c(1:6), nrow=3, ncol=2) ## 3x2    1~6
a
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

2.5 data.frame

Column Row Excel data.frame()

```
StuDF <- data.frame(StuID=c(1,2,3,4,5), ##  =
                    name=c(" "," "," "," "," "),
                    score=c(80,60,90,70,50))
StuDF
```

```
##   StuID name score
## 1     1      80
## 2     2      60
## 3     3      90
## 4     4      70
## 5     5      50
```

StuID, name, score R V1 - Vn R StuID score name R 1 - n colnames() rownames()

```
colnames(StuDF) ##
```

```
## [1] "StuID" "name"  "score"
```

```
rownames(StuDF) ##
```

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

str()

```
str(StuDF)
```

```
## 'data.frame':    5 obs. of  3 variables:
##  $ StuID: num  1 2 3 4 5
##  $ name : Factor w/ 5 levels " "," ",..: 4 2 5 3 1
##  $ score: num  80 60 90 70 50
```

$

```
iris$Species ## iris   Species
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
```

```
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

$

```
iris$Species[2]<-"versicolor"
head(iris$Species)
```

```
## [1] setosa     versicolor setosa     setosa     setosa     setosa
## Levels: setosa versicolor virginica
```

NULL

```
iris$Species<-NULL ## Species
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1          5.1         3.5          1.4         0.2
## 2          4.9         3.0          1.4         0.2
## 3          4.7         3.2          1.3         0.2
## 4          4.6         3.1          1.5         0.2
## 5          5.0         3.6          1.4         0.2
## 6          5.4         3.9          1.7         0.4
```

2.6 data.table

data.table data.frame data.table (**?**) package `data.table` Chapter **??** Chapter **??** DataCamp

2.7

:

- `names()`
- `dimnames()`
- `length()`
- `dim()`
- `class()`
- `table()`
- `str()`

names()

```
head(islands) ##R
```

```
##       Africa   Antarctica         Asia    Australia Axel Heiberg       Baffin
##        11506         5500        16988         2968           16          184
```

```
head(names(islands)) ##
```

```
## [1] "Africa"       "Antarctica"   "Asia"         "Australia"    "Axel Heiberg"
## [6] "Baffin"
```

```
head(USArrests) ##R
```

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58   21
## Alaska       10.0     263       48   44
## Arizona       8.1     294       80   31
## Arkansas      8.8     190       50   20
## California    9.0     276       91   41
## Colorado      7.9     204       78   39
```

```
head(names(USArrests)) ##
```

```
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

dimnames()

```
dimnames(USArrests)
```

```
## [[1]]
##  [1] "Alabama"        "Alaska"         "Arizona"        "Arkansas"
##  [5] "California"     "Colorado"       "Connecticut"    "Delaware"
##  [9] "Florida"        "Georgia"        "Hawaii"         "Idaho"
## [13] "Illinois"       "Indiana"        "Iowa"           "Kansas"
## [17] "Kentucky"       "Louisiana"      "Maine"          "Maryland"
## [21] "Massachusetts"  "Michigan"       "Minnesota"      "Mississippi"
## [25] "Missouri"       "Montana"        "Nebraska"       "Nevada"
## [29] "New Hampshire"  "New Jersey"     "New Mexico"     "New York"
## [33] "North Carolina" "North Dakota"   "Ohio"           "Oklahoma"
## [37] "Oregon"         "Pennsylvania"   "Rhode Island"   "South Carolina"
## [41] "South Dakota"   "Tennessee"      "Texas"          "Utah"
## [45] "Vermont"        "Virginia"       "Washington"     "West Virginia"
## [49] "Wisconsin"      "Wyoming"
##
## [[2]]
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

length()

```
length(islands)
```

```
## [1] 48
```

```
length(USArrests)
```

```
## [1] 4
```

dim() dimnames()

```r
dim(USArrests)
```

```
## [1] 50  4
```

```
 class()
```

```r
class(1)
```

```
## [1] "numeric"
```

```r
class("Test")
```

```
## [1] "character"
```

```r
class(Sys.Date())
```

```
## [1] "Date"
```

```
 table()
```

```r
iris$Species ##
```

```
## NULL
```

```r
table(iris$Species) ##
```

```
## < table of extent 0 >
```

```
 str()
```

```r
str(iris)
```

```
## 'data.frame':    150 obs. of  4 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
```

```r
str(listSample)
```

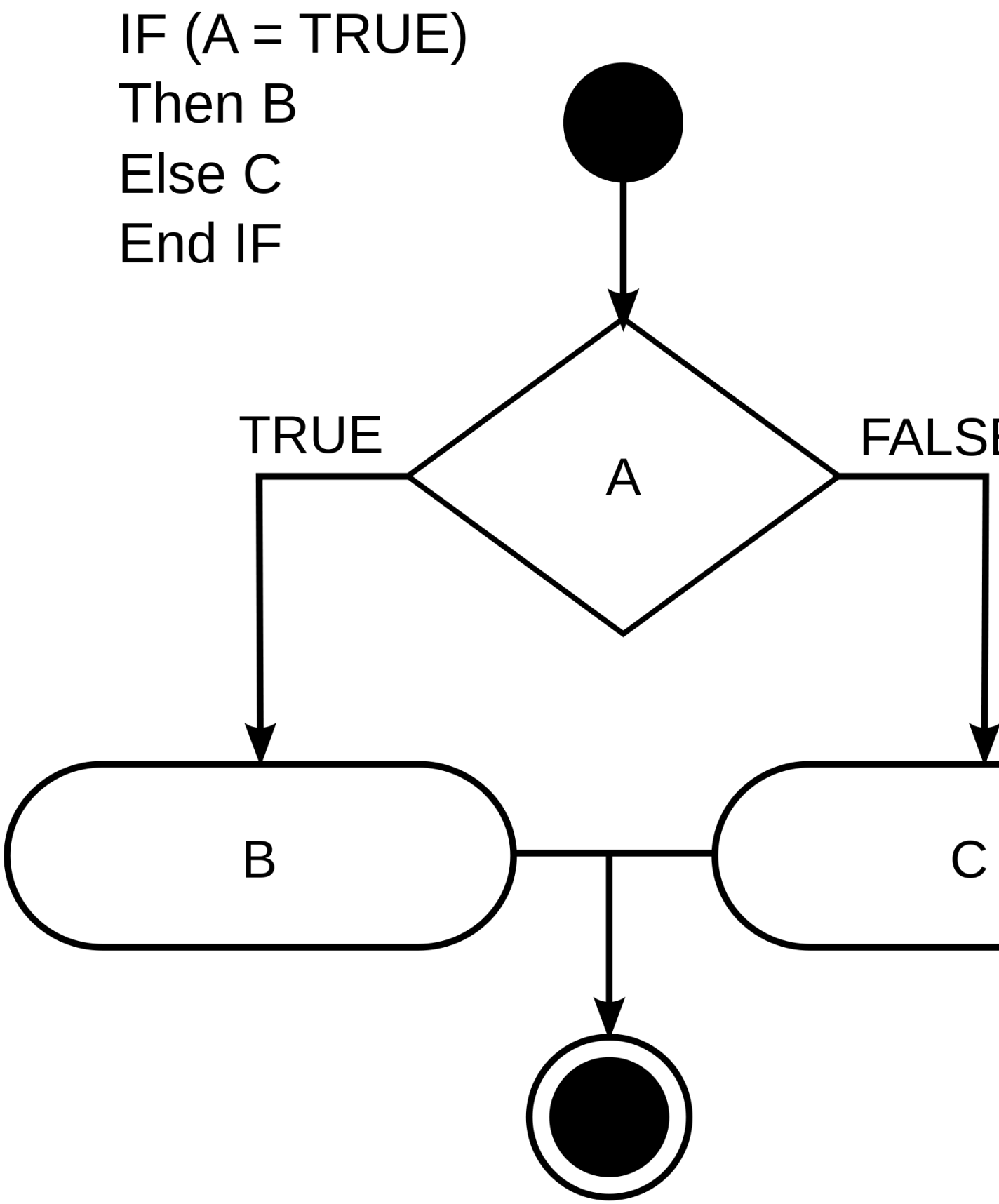
```
## List of 4
##  $ Students: chr [1:5] " " " " " " " " ...
##  $ Year    : num 2017
##  $ School  : chr "CGU"
##  $ Gender  : chr [1:5] "M" "F" "M" "F" ...
```

Chapter 3

3.1

3.1.1 if-else

if-else **if-else** **if** **(TRUE)** if **(FALSE)** **else** **else**


IF (A = TRUE)
Then B
Else C
End IF
TRUE
A
B
C

if else {} {} {}

60 60 :

```
score<-59
if(score>=60){
  print(" ")
}else{
  print("  ")
}
```

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

```
score<-80
if(score>=60){
  print(" ")
}else{
  print("  ")
}
```

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

3.1.2 if-else if-else

90 60 90 60 if else **else if** :

```
score<-95
if(score>=90){
  print(" ")
}else if(score>=60){
  print(" ")
}else{
  print("  ")
}
```

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

if-else if-else if else if 95 90 (if) 60 (else if)

3.1.3 if

if if if-else 60 60 60 :

```
CHscore<-95 ##
ENscore<-55 ##
if(CHscore>=60){
  if(ENscore>=60){
    print("  ")
  }else{
    print("     ")
  }
}else{
  if(ENscore>=60){
    print("     ")
  }else{
    print("   ")
```

```
  }
}
```

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

3.1.4 ifelse()

ifelse() if-else ifelse(, ,) :

```
score<-80
ifelse(score>=60," ","  ")
```

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

ifelse()

```
scoreVector<-c(30,90,50,60,80)
ifelse(scoreVector>=60," ", "  ")
```

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

3.2

3.2.1 for

R for for for (in){ } :

```
for (n in 1:10){ #n   1:10
  print(n)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
```

for if-else :

```
for (n in 1:10){
  if(n%%2==0){ #
    print(n)
  }else{
    print(" ") #   " "
  }
}
```

```
## [1] " "
## [1] 2
## [1] " "
## [1] 4
```

```
## [1] " "
## [1] 6
## [1] " "
## [1] 8
## [1] " "
## [1] 10
```

3.2.2 while

while while

```
x<-0
while(x<=5){
  print(x)
  x<-x+1
}
```

```
## [1] 0
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
```

3.2.3 break

break

```
for(n in 1:10){
  if(n==5){
    break ##  5
  }
  print(n)
}
```

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

3.2.4 next

next

```
for(n in 1:10){
  if(n==5){
    next ## 5
  }
  print(n)
}
```

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

```
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
```

3.3 purrr

better than sapply() and lapply()

Chapter 4

(Function)

:

-
-
-
-

```
 <-function( 1, 2,...){


}
```

:

-
-
-
- function

:

-
-

:

-
- R
-

:

-
-
-
-

Chapter 5

(Data) `values of qualitative or quantitative variables, belonging to a set of items.` Chapter ?? Chapter ?? Tidy data Chapter ?? Chapter ??

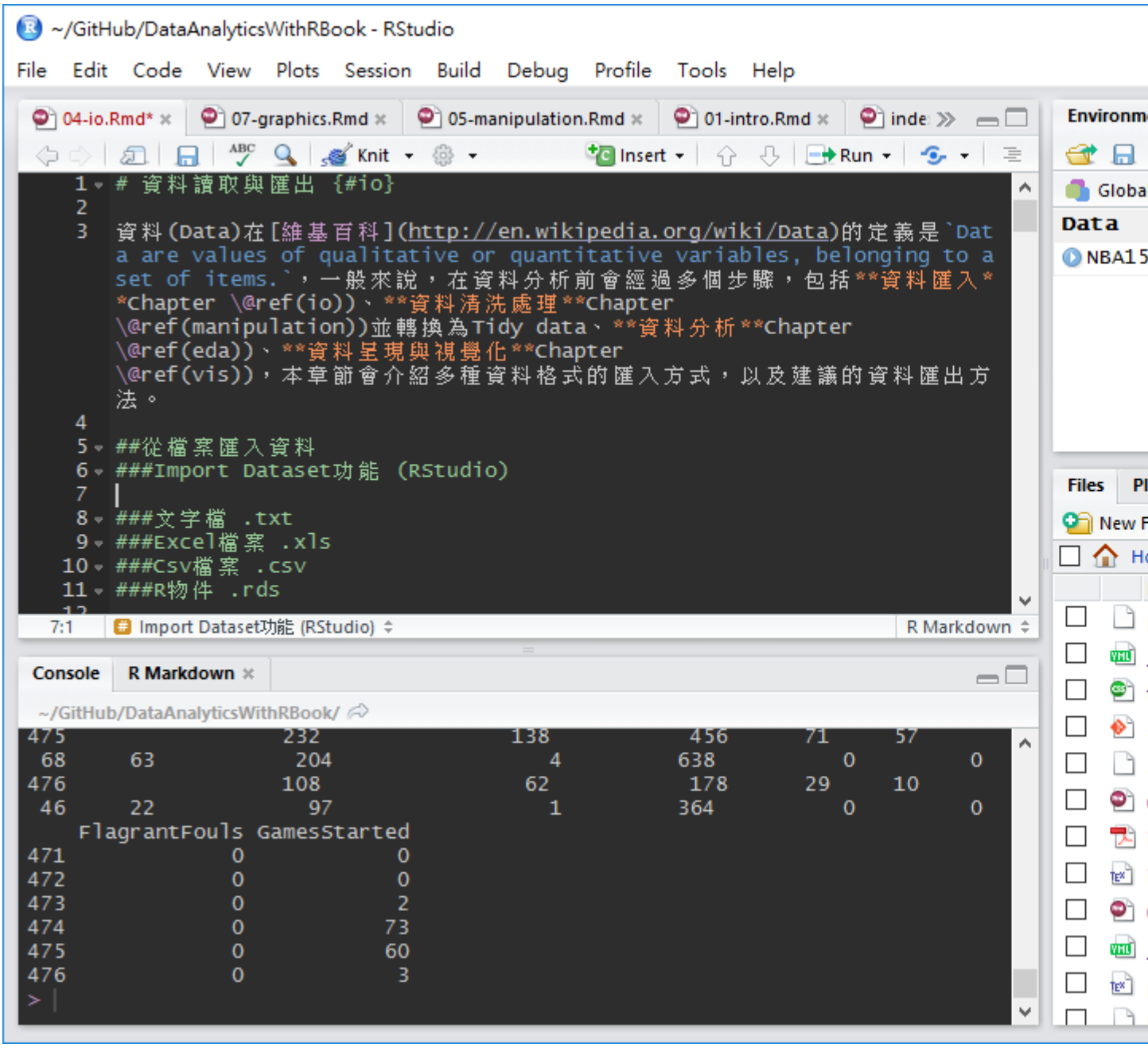
:

-
-
- Open Data (API)
- ( )
-

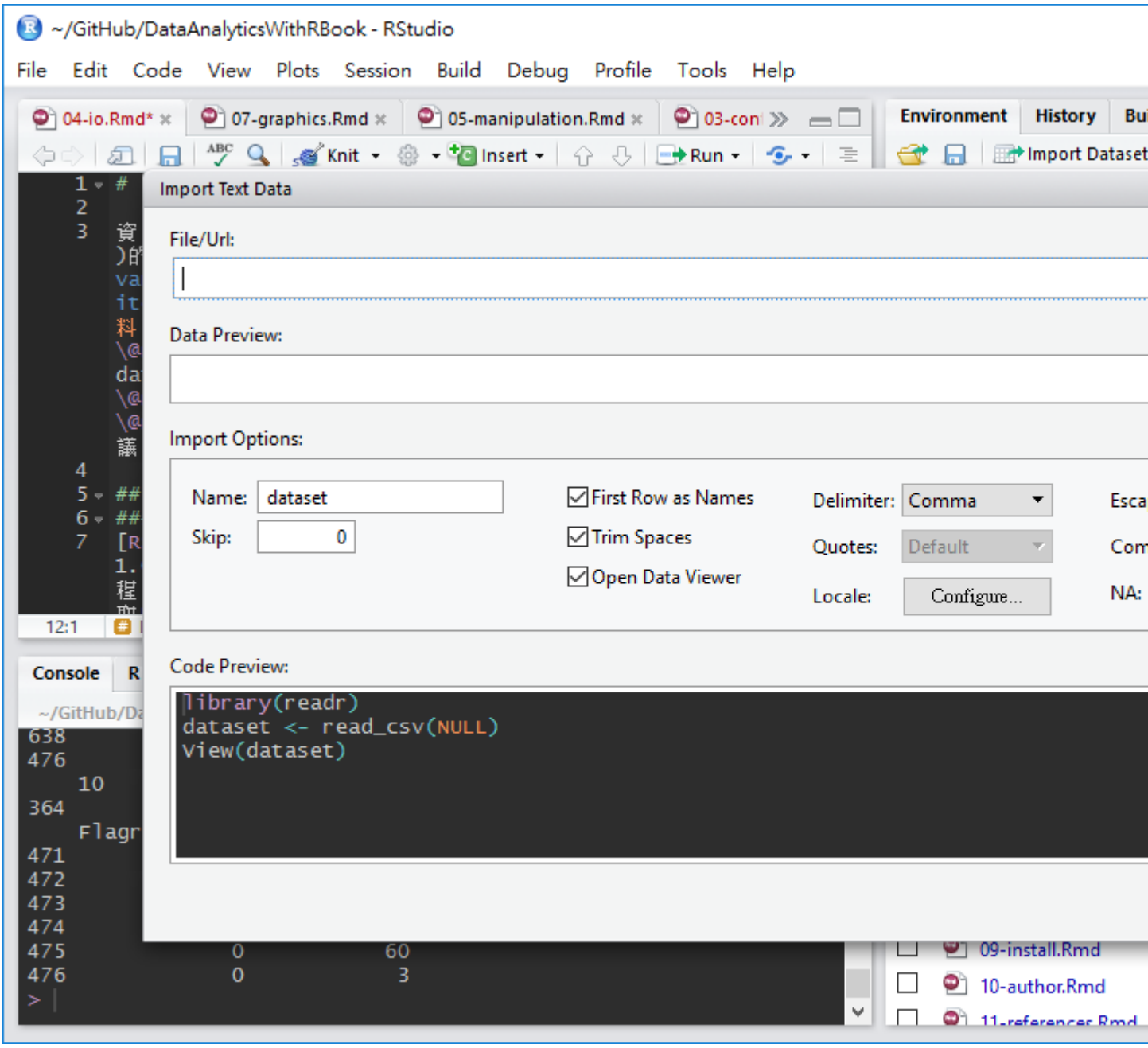
5.1

5.1.1 Import Dataset (RStudio)

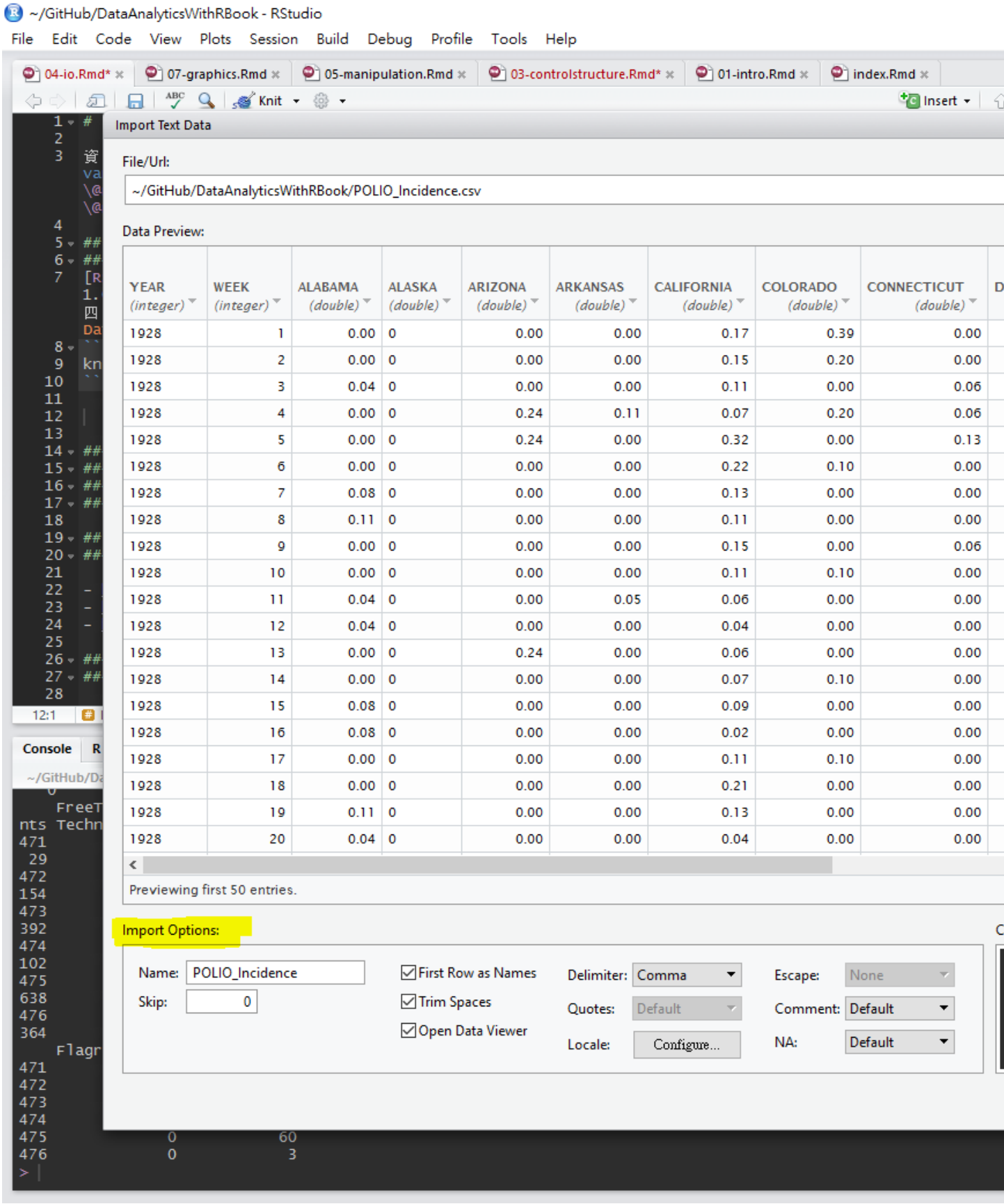
RStudio 1.0 `.csv` `Excel` `SAS` RStudio Environment **Import Dataset**



csv From CSV Browse



Import Options Delimiter First Row as Names


~/GitHub/DataAnalyticsWithRBook - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
04-io.Rmd*
07-graphics.Rmd
05-manipulation.Rmd
03-controlstructure.Rmd*
01-intro.Rmd
index.Rmd
Knit
Insert
Import Text Data
File/Url:
~/GitHub/DataAnalyticsWithRBook/POLIO_Incidence.csv
Data Preview:
YEAR (integer) | WEEK (integer) | ALABAMA (double) | ALASKA (double) | ARIZONA (double) | ARKANSAS (double) | CALIFORNIA (double) | COLORADO (double) | CONNECTICUT (double)
1928 1 0.00 0 0.00 0.00 0.17 0.39 0.00
1928 2 0.00 0 0.00 0.00 0.15 0.20 0.00
1928 3 0.04 0 0.00 0.00 0.11 0.00 0.06
1928 4 0.00 0 0.24 0.11 0.07 0.20 0.06
1928 5 0.00 0 0.24 0.00 0.32 0.00 0.13
1928 6 0.00 0 0.00 0.00 0.22 0.10 0.00
1928 7 0.08 0 0.00 0.00 0.13 0.00 0.00
1928 8 0.11 0 0.00 0.00 0.11 0.00 0.00
1928 9 0.00 0 0.00 0.00 0.15 0.00 0.06
1928 10 0.00 0 0.00 0.00 0.11 0.10 0.00
1928 11 0.04 0 0.00 0.05 0.06 0.00 0.00
1928 12 0.04 0 0.00 0.00 0.04 0.00 0.00
1928 13 0.00 0 0.24 0.00 0.06 0.00 0.00
1928 14 0.00 0 0.00 0.00 0.07 0.10 0.00
1928 15 0.08 0 0.00 0.00 0.09 0.00 0.00
1928 16 0.08 0 0.00 0.00 0.02 0.00 0.00
1928 17 0.00 0 0.00 0.00 0.11 0.10 0.00
1928 18 0.00 0 0.00 0.00 0.21 0.00 0.00
1928 19 0.11 0 0.00 0.00 0.13 0.00 0.00
1928 20 0.04 0 0.00 0.00 0.04 0.00 0.00
Previewing first 50 entries.
Import Options:
Name: POLIO_Incidence
Skip: 0
First Row as Names
Trim Spaces
Open Data Viewer
Delimiter: Comma
Quotes: Default
Locale: Configure...
Escape: None
Comment: Default
NA: Default
Console

tab .csv Delimiter Tab

Code Preview: R (.R)

5.1.2 .txt

readr (?) package read_delim() tab delim \t tab col_names TRU
X1, X2, X3 ...

(?read_delim)

- file,
- delim,
- quote,
- escape_backslash, FALSE /
- escape_double, TRUE quote
- col_names, T/F
- col_types,
- comment,
- skip,

```
library(readr)
dataset <- read_delim("    ", delim="\t")
```

5.1.3 CSV .csv

readr (?) package CSV ( ) read_csv()

```
library(readr)
dataset <- read_csv("    ")
```

5.1.4 Excel .xls

readxl (?) package Excel (xls, xlsx) read_excel() col_names sheet (sheet)

```
library(readxl)
dataset <- read_excel("    ")
```

5.1.5 R .rds

R R R R readRDS() (:A better way of saving and loading objects in R)

```
dataset <- readRDS("    ")
```

5.1.6 R .R

source, R Obejct or script, , ASCII (dump)

5.1.7 ( )

readLines,

5.1.8

R :

- MySQL `RMySQL`
- HDF5 `rhdf5`
- Weka `foreign`
- Stata `foreign`
- SPSS `Hmisc`
- SAS `Hmisc`
- GIS `rgdal`
- Images `jpeg`
- Music `tuneR`

5.1.9

R

- # Row
-
- Column
-

Column

5.2

5.2.1 Open Data

(Open data) Open data Open data Open data Data.gov Open data ( )

2011

-
- Data Taipei
- x
-

Open Data : `CSV`Chapter **??** `JSON`Chapter **??** `XML`Chapter **??** CSV Chapter **??**

5.2.2 API (Application programming interfaces)

Application programming interfaces (API) ( )

Open Data **API** API JSON Chapter **??** API **ID** **RID**

- **ID**:
- **RID**:

http://data.taipei/opendata/datalist/apiAccess?scope=resourceAquire&rid=f4a75ba9-7721-4363-884d-c3820b0b917c

5.2.3 JSON

JSON (**J**ava**s**cript **O**bject **N**otation) (Wiki) :

- from **a**pplication **p**rogramming **i**nterfaces (APIs)
- JavaScript Java Node.js
- NoSQL JSON **MongoDB**
-
  - Numbers (double)
  - Strings (double quoted)
  - Boolean (*true* or *false*)
  - Array (ordered, comma separated enclosed in square brackets)
  - Object (unorderd, comma separated collection of **key:value** pairs in curley brackets *{}*)

JSON

Open Data JSON API http://data.taipei/opendata/datalist/apiAccess?scope=resourceAquire&rid=f4
7721-4363-884d-c3820b0b917c

JSON R `jsonlite`(?) package Chapter ?? `fromJSON()` JSON API **httr** `httr`(?) package `GET()`

```
library(jsonlite)
PetData<-fromJSON("http://data.taipei/opendata/datalist/apiAccess?scope=resourceAquire&rid=f4a75ba9-772
str(PetData)
```

```
## List of 1
##  $ result:List of 5
##   ..$ offset : int 0
##   ..$ limit  : int 10000
##   ..$ count  : int 307
##   ..$ sort   : chr ""
##   ..$ results:'data.frame':  307 obs. of  20 variables:
##   .. ..$ _id            : chr [1:307] "1" "2" "3" "4" ...
##   .. ..$ Name           : chr [1:307] "" "" " " " " " ...
##   .. ..$ Sex            : chr [1:307] " " " " " " " " ...
##   .. ..$ Type           : chr [1:307] " " " " " " " " ...
##   .. ..$ Build          : chr [1:307] " " " " " " " " ...
##   .. ..$ Age            : chr [1:307] " " " " " " " " " ...
##   .. ..$ Variety        : chr [1:307] "  " "  " "  " "  " ...
##   .. ..$ Reason         : chr [1:307] "   " "   " "   " "   " ...
##   .. ..$ AcceptNum     : chr [1:307] "106032203" "106031503" "106031204" "106031003" ...
##   .. ..$ ChipNum        : chr [1:307] "" "" "" "" ...
##   .. ..$ IsSterilization: chr [1:307] "  " "  " "  " "  " ...
##   .. ..$ HairType       : chr [1:307] " " " " " " " " ...
##   .. ..$ Note        : chr [1:307] "   " "          " "  ~               ?      !!\n" "     \n       \n
##   .. ..$ Resettlement  : chr [1:307] "       106032203" "       106031503" "       106031204" "       106
##   .. ..$ Phone       : chr [1:307] "02-87913063" "02-8791-3063" "02-87913062" "02-87913062" ...
##   .. ..$ Email       : chr [1:307] "tcapoa8@mail.taipei.gov.tw" "tcapoa8@mail.taipei.gov.tw" "tcapoa8@ma
##   .. ..$ ChildreAnlong  : chr [1:307] "" "" "" "" ...
##   .. ..$ AnimalAnlong   : chr [1:307] "" "" "" "" ...
##   .. ..$ Bodyweight     : chr [1:307] "" "" "" "" ...
##   .. ..$ ImageName    : chr [1:307] "http://163.29.39.183/uploads/images/medium/9fe8fb19-ed6f-41a2-b25c
```

`fromJSON()` JSON `list` result (offset, limit, count, sort, results) results data.frame

```
head(PetData$result$results)
```

```
##   _id Name Sex Type Build  Age Variety   Reason AcceptNum ChipNum
```

```
## 1   1                          106032203
## 2   2                          106031503
## 3   3                         106031204
## 4   4                         106031003
## 5   5                         106031002
## 6   6                          106031021
##   IsSterilization HairType
## 1
## 2
## 3
## 4
## 5
## 6
##
## 1
## 2
## 3                                                    ~               ?      !!\n
## 4      \n       \n      \n          \n          \n      !
## 5                                                          \n ~                !\n
## 6
##                          Resettlement        Phone                      Email
## 1          106032203  02-87913063 tcapoa8@mail.taipei.gov.tw
## 2         106031503 02-8791-3063 tcapoa8@mail.taipei.gov.tw
## 3          106031204  02-87913062 tcapoa8@mail.taipei.gov.tw
## 4          106031003  02-87913062 tcapoa8@mail.taipei.gov.tw
## 5          106031002  02-87913062 tcapoa8@mail.taipei.gov.tw
## 6          106031021  02-87913062 tcapoa8@mail.taipei.gov.tw
##   ChildreAnlong AnimalAnlong Bodyweight
## 1
## 2
## 3
## 4
## 5
## 6
##                                                                      ImageName
## 1 http://163.29.39.183/uploads/images/medium/9fe8fb19-ed6f-41a2-b25c-719a80559085.jpg
## 2 http://163.29.39.183/uploads/images/medium/6b77ae0f-e893-4bca-af97-c6e9afd36d2e.jpg
## 3 http://163.29.39.183/uploads/images/medium/e0de6dfd-332e-459b-b157-36b006a1924c.jpg
## 4 http://163.29.39.183/uploads/images/medium/c5983ea5-b55a-4bd1-bb3b-2c03f8e496a2.jpg
## 5 http://163.29.39.183/uploads/images/medium/478ca5ce-4bc2-46c7-a22f-dd8c2bf7ed77.jpg
## 6 http://163.29.39.183/uploads/images/medium/c9318c07-988d-4ac4-b7ba-ddba0968759c.jpg
```

results 20

```
table(PetData$result$results$Reason)
```

```
##
##
##           26           93          121           44           23
```

JSON jsonlite package toJSON()

```
myjson <- toJSON(iris, pretty=TRUE)
str(myjson)
```

```
## Class 'json'  chr "[\n  {\n   \"Sepal.Length\": 5.1,\n   \"Sepal.Width\": 3.5,\n   \"Petal.Length\": 1.4,
```

5.2.4 XML

- Extensible markup language
-
- XML **Html**
- Components
  - Markup - labels that give the text structure
  - Content - the actual text of the document
- XML Wiki

Tags, elements and attributes

- Tags correspond to general labels
  - Start tags `<breakfast_menu>`, `<price>`
  - End tags `</breakfast_menu>`,`</price>`
  - Empty tags `<line-break />`
- Elements are specific examples of tags
  - `<name>Belgian Waffles</name>`
- Attributes are components of the label
  - `<book category="web">`

Open Data XML XML R `XML` (?) package `xmlParse()`

```
library(XML)
waterQ <- xmlParse("http://data.taipei/opendata/datalist/datasetMeta/download?id=961ca397-4a59-45e8-b31
```

`xpathSApply()` **XPath**

```
#  "code_name"
xpathSApply(waterQ,"//code_name",xmlValue)[1:10]
```

```
##  [1] "   "               "   "
##  [3] "                  " "   "
##  [5] "   "               "   "
##  [7] "  "                   "    "
##  [9] "   "                  "   "
```

```
#
xpathSApply(waterQ,"//longitude",xmlValue)[1:10]
```

```
##  [1] "121.56094" "121.54401" "121.55557" "121.53476" "121.54043" "121.55661"
##  [7] "121.55360" "121.53551" "121.59892" "121.60829"
```

XPath XML XML Path Language XML **XPath** W3C Schools Google :

```
// :    ,      //a
@ :   ,        //a/@href
```

5.2.5 Webscraping

API ptt **Webscraping** Webscraping Wiki R HTML XML `rve` package

IP

`readLines()`

```
con <- url("http://im.cgu.edu.tw/bin/home.php")
htmlCode <-readLines(con)
```

```
## Warning in readLines(con): incomplete final line found on 'http://im.cgu.edu.tw/
## bin/home.php'
```

```
close(con)
htmlCode[1:5]
```

```
## [1] "<!DOCTYPE html PUBLIC \"-//W3C//DTD XHTML 1.0 Transitional//EN\" \"http://www.w3.org/TR/xhtml1/DTD/
## [2] "<html xmlns=\"http://www.w3.org/1999/xhtml\" lang=\"zh-tw\">"
## [3] "<head>"
## [4] "<meta http-equiv=\"Content-Type\" content=\"text/html; charset=utf-8\" />"
## [5] "<meta http-equiv=\"X-UA-Compatible\" content=\"IE=EmulateIE7\" /><meta name=\"keywords\" content=\
```

XML (XML package) XML **XPath**

```
html <- htmlParse("http://im.cgu.edu.tw/bin/home.php")
xpathSApply(html, "//title", xmlValue)
```

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

```
xpathSApply(html, "//span[@class='ptname ']", xmlValue)
```

```
##  [1] "    "       "     "     "     "
##  [4] "    "         "    "         "     "
##  [7] "    "       "TA    "     "    "
## [10] "    "         "     "    "    "
## [13] "    "           "    /  "  "    /  "
## [16] "    "             "    /  "  "    "
## [19] "     " "    "
```

HTML XML rvest(?) package R

```
install.packages("rvest") ##
```

```
library(rvest) ##
```

rvest

- read_html(" ")
- html_nodes() (CSS xpath)
- html_text() /
- html_attr() url

```
YahooNewsurl="https://tw.news.yahoo.com/"
news_title = read_html(YahooNewsurl) %>% html_nodes(".tpl-title a") %>% html_text()
news_url = read_html(YahooNewsurl) %>% html_nodes(".tpl-title a") %>% html_attr("href")
Yahoo_news = data.frame(title = news_title, url=news_url)
head(Yahoo_news)
```

```
                              title                                          url
1         1 5        / 1 5      -    -091741737.html
2                                 /  -  - -    -081036215.html
3             1           /    -      -101500692.html
4                                  /    -    -072010033.html
5                                /    -     -044517088.html
6                  /     -     -160200179.html
```

html_nodes() html_text() html_attr() **Google Chrome** Yahoo HTML

```
<ul class="tpl-title yom-list list-style-none" id="yui_3_9_1_1_1486568229946_2408">
<li class="list-story first" id="yui_3_9_1_1_1486568229946_2407">
<div class="txt" id="yui_3_9_1_1_1486568229946_2406">
<a href="/ 1 5      -     -091741737.html" class="title " data-ylk="pkg:96a0ca11-47bc-3100-81ad-0a288707f15
<cite id="yui_3_9_1_1_1486568229946_2405">
<span class="provider" id="yui_3_9_1_1_1486568229946_2404">Yahoo    </span>
</cite></div></li>
....
```

ul css class tpl-title yom-list list-style-none classtpl-title CSS W3C Schools {target="_blank"}

:

- - r
- rvest GitHub
- R Bloggers
- Ptt
-

5.3 Facebook

Facebook Graph API API Facebook **Graph API** JSON Facebook Graph API Explorer **Graph API access token** ( ) Graph API Explorer **Get Token** ( ) (Application) **access token**

Facebook access token

5.3.1 Graph API in R

```
library(httr)
token<-"your token" # access token
FBData = GET(
    paste0("https://graph.facebook.com/v2.8/tsaiingwen?fields=posts%7Bmessage%7D&access_token=",
           token))
names(FBData)
```

```
## [1] "url"         "status_code" "headers"     "all_headers" "cookies"     "content"     "date"
## [8] "times"       "request"     "handle"
```

```
json1 = content(FBData)
names(json1)
```

```
## [1] "posts" "id"
```

```
names(json1$posts)
```

```
## [1] "data"   "paging"
```

```
head(json1$posts$data,3)
```

```
[[1]]
[[1]]$message
[1] "                                    \n\n     5+2                                    \n\n
```

```
[[1]]$id
[1] "46251501064_10154006497451065"


[[2]]
[[2]]$message
[1] "                              \n\n    5+2                                                    "

[[2]]$id
[1] "46251501064_10154006456601065"


[[3]]
[[3]]$message
[1] "                       \n\n                                  "

[[3]]$id
[1] "46251501064_10154001652641065"
```

```
json1$posts$data[[1]]$message
```

```
##[1] "                                    \n\n     5+2                                    \n\n
```

5.3.2 Rfacebook package

Graph API `Rfacebook`(?) package Facebook Rfacebook `tsaiingwen`

```
library(Rfacebook)
token<-"your token" # token
getPage("tsaiingwen", token,n = 5)
```

```
5 posts       from_id           from_name
1 46251501064    Tsai Ing-wen
2 46251501064    Tsai Ing-wen
3 46251501064    Tsai Ing-wen
4 46251501064    Tsai Ing-wen
5 46251501064    Tsai Ing-wen

1                                    \n\n     5+2                                    \n\n
2                                                                \n\n    5+2
3
4
5
              created_time  type
1 2017-02-07T08:02:45+0000 photo
2 2017-02-07T07:18:00+0000 photo
3 2017-02-05T07:12:52+0000 photo
4 2017-01-31T08:37:42+0000 photo
5 2017-01-30T11:41:07+0000 photo
                                                                                  link
1 https://www.facebook.com/tsaiingwen/photos/a.390960786064.163647.46251501064/10154006497206065/?type=3
2 https://www.facebook.com/tsaiingwen/photos/a.390960786064.163647.46251501064/10154006455396065/?type=3
3 https://www.facebook.com/tsaiingwen/photos/a.390960786064.163647.46251501064/10154001652641065/?type=3
4 https://www.facebook.com/tsaiingwen/photos/a.390960786064.163647.46251501064/10153989357181065/?type=3
5 https://www.facebook.com/tsaiingwen/photos/a.390960786064.163647.46251501064/10153987089121065/?type=3
```

```
                            id likes_count comments_count shares_count
1 46251501064_10154006497451065        2013            125           43
2 46251501064_10154006456601065        2217            163           57
3 46251501064_10154001652641065        9416            920          163
4 46251501064_10153989358051065       34116           1574          373
5 46251501064_10153987095776065       20592            665          269
```

30 **since until**

```
lastDate<-Sys.Date()
DateVector<-seq(as.Date("2017-01-01"),lastDate,by="5 days")
DateVectorStr<-as.character(DateVector)
DateVectorStr
```

```
## "2017-01-01" "2017-01-06" "2017-01-11" "2017-01-16" "2017-01-21" "2017-01-26" "2017-01-31" "2017-02-05"
```

since until

```
totalPage<-NULL
token<-'your token'
numberOfPost<-30
for(i in 1:(length(DateVectorStr)-1)){
    tempPage<-getPage("tsaiingwen", token,
                      since = DateVectorStr[i],until = DateVectorStr[i+1])
    totalPage<-rbind(totalPage,tempPage)
}
nrow(totalPage)
```

```
## 4 posts 8 posts 10 posts 3 posts 2 posts 14 posts 1 posts
## [1] 42
```

Rfacebook Packages

- getUsers()
- getPost()
- searchFacebook()
- Check https://github.com/pablobarbera/Rfacebook
- ?Rfacebook

5.4

R (Excel) tab (.txt) (.csv) R R (.rds)

5.4.1 .txt

`write.table()`

- `x` matrix data.frame
- `file`
- `append` T/F TRUE F (F)
- `quote` (T)
- `sep` ( )
- `eol`
- `na`
- `dec`

- `row.names` T/F row names
- `col.names` T/F column names
- `qmethod`
- `fileEncoding`

```
write.table(iris,file="iris.txt",sep=",",row.names = F,col.names = T)
```

5.4.2 CSV .csv

`write.table()` `write.csv()`

```
write.csv(iris,file="iris.csv",row.names = F)
```

5.4.3 R .rds

R R (.rds)

```
saveRDS(iris,"iris.rds")
```

Chapter 6

6.1 Tidy Data

Each column is a variable. Each row is an observation.

- Column Column Name
- Raw
-
- index
- One file, one table

6.2

Chapter **??** **(numeric)** **(character)** **(logic)** **(Date)**

6.2.1

`is.` TRUE

- `is.numeric( )`
- `is.character( )`
- `is.logical( )`

```
num<-100
cha<-'200'
boo<-T
is.numeric(num)
```

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

```
is.numeric(cha)
```

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

```
is.character(num)
```

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

```
is.character(cha)
```

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

```
is.logical(boo)
```

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

class( )

```
class(num)
```

```
## [1] "numeric"
```

```
class(cha)
```

```
## [1] "character"
```

```
class(boo)
```

```
## [1] "logical"
```

```
class(Sys.Date())
```

```
## [1] "Date"
```

6.2.2

as.

- as.numeric( )
- as.character( )
- as.logical( )

```
as.numeric(cha)
```

```
## [1] 200
```

```
as.numeric(boo)
```

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

```
as.character(num)
```

```
## [1] "100"
```

```
as.character(boo)
```

```
## [1] "TRUE"
```

NA Warning: NAs introduced by coercion Warning: NA

```
as.numeric("abc")
```

```
## Warning: NAs introduced by coercion
```

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

lubridate(?) package / / ymd() y year m month d day / / mdy()

```
library(lubridate)
ymd('2012/3/3')
```

```
## [1] "2012-03-03"
```

```
mdy('3/3/2012')
```

```
## [1] "2012-03-03"
```

6.3

6.3.1

- strsplit()
- substr()
- toupper() tolower()
- paste() paste0()
- gsub()
- str_trim() stringr(?) package

```
strsplit ("Hello World"," ")
```

```
## [[1]]
## [1] "Hello" "World"
```

```
toupper("Hello World")
```

```
## [1] "HELLO WORLD"
```

```
tolower("Hello World")
```

```
## [1] "hello world"
```

```
paste("Hello", "World", sep='')
```

```
## [1] "HelloWorld"
```

```
substr("Hello World", start=2,stop=4)
```

```
## [1] "ell"
```

```
gsub("o","0","Hello World")
```

```
## [1] "Hell0 W0rld"
```

```
library(stringr)
str_trim(" Hello World ")
```

```
## [1] "Hello World"
```

6.3.2

grep() grepl():

- (index) grep(,)
- (TRUE or FALSE) grepl(,)

```
grep("A",c("Alex","Tom","Amy","Joy","Emma")) ##      A   "A"
```

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

```
grepl("A",c("Alex","Tom","Amy","Joy","Emma")) ##      A      "A"
```

```
## [1]  TRUE FALSE  TRUE FALSE FALSE
```

```
grepl("a",c("Alex","Tom","Amy","Joy","Emma")) ##      a      "a"
```

```
## [1] FALSE FALSE FALSE FALSE  TRUE
```

6.4 Subset

6.4.1 ( )

{#vector} []

```
letters ##R
```

```
##  [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s"
## [20] "t" "u" "v" "w" "x" "y" "z"
```

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

```
## [1] "a"
```

```
letters[1:10] ## letters
```

```
##  [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j"
```

```
letters[c(1,3,5)] ## letters  1,3,5
```

```
## [1] "a" "c" "e"
```

```
letters[c(-1,-3,-5)] ## letters   1,3,5
```

```
##  [1] "b" "d" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s" "t" "u" "v"
## [20] "w" "x" "y" "z"
```

head() tail()

```
head(letters,5) ## letters
```

```
## [1] "a" "b" "c" "d" "e"
```

```
tail(letters,3) ## letters
```

```
## [1] "x" "y" "z"
```

6.4.2

data.frame (Row) (Column) [] , **Row, Column** , , (index) (TRUE/FALSE)

- : dataFrame[row index,column index]
- : dataFrame[c(T,F,T),c(T,F,T)]
- : dataFrame[row name,column name]

```
data(iris)
iris[1,2] ##  Row  Column
```

```
## [1] 3.5
```

```
iris[1:3,] ## 1~3 Row   Column
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
```

```
iris[,"Species"] ##  Row  Species Column
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

```
iris[1:10,c(T,F,T,F,T)] ## 1~10 Row 1,3,5 Column (TRUE)
```

```
##    Sepal.Length Petal.Length Species
## 1           5.1          1.4  setosa
## 2           4.9          1.4  setosa
## 3           4.7          1.3  setosa
## 4           4.6          1.5  setosa
## 5           5.0          1.4  setosa
## 6           5.4          1.7  setosa
## 7           4.6          1.4  setosa
## 8           5.0          1.5  setosa
## 9           4.4          1.4  setosa
## 10          4.9          1.5  setosa
```

$ Column

```
iris$Species ##  Row  Species Column
```

```
##   [1] setosa     setosa     setosa     setosa     setosa     setosa
##   [7] setosa     setosa     setosa     setosa     setosa     setosa
##  [13] setosa     setosa     setosa     setosa     setosa     setosa
##  [19] setosa     setosa     setosa     setosa     setosa     setosa
##  [25] setosa     setosa     setosa     setosa     setosa     setosa
##  [31] setosa     setosa     setosa     setosa     setosa     setosa
##  [37] setosa     setosa     setosa     setosa     setosa     setosa
##  [43] setosa     setosa     setosa     setosa     setosa     setosa
##  [49] setosa     setosa     versicolor versicolor versicolor versicolor
##  [55] versicolor versicolor versicolor versicolor versicolor versicolor
##  [61] versicolor versicolor versicolor versicolor versicolor versicolor
```

```
##  [67] versicolor versicolor versicolor versicolor versicolor versicolor
##  [73] versicolor versicolor versicolor versicolor versicolor versicolor
##  [79] versicolor versicolor versicolor versicolor versicolor versicolor
##  [85] versicolor versicolor versicolor versicolor versicolor versicolor
##  [91] versicolor versicolor versicolor versicolor versicolor versicolor
##  [97] versicolor versicolor versicolor versicolor virginica  virginica
## [103] virginica  virginica  virginica  virginica  virginica  virginica
## [109] virginica  virginica  virginica  virginica  virginica  virginica
## [115] virginica  virginica  virginica  virginica  virginica  virginica
## [121] virginica  virginica  virginica  virginica  virginica  virginica
## [127] virginica  virginica  virginica  virginica  virginica  virginica
## [133] virginica  virginica  virginica  virginica  virginica  virginica
## [139] virginica  virginica  virginica  virginica  virginica  virginica
## [145] virginica  virginica  virginica  virginica  virginica  virginica
## Levels: setosa versicolor virginica
```

Row `subset()` `subset( , )`

```
subset(iris,Species=="virginica") ##Species "virginica" Row   Column
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 101          6.3         3.3          6.0         2.5 virginica
## 102          5.8         2.7          5.1         1.9 virginica
## 103          7.1         3.0          5.9         2.1 virginica
## 104          6.3         2.9          5.6         1.8 virginica
## 105          6.5         3.0          5.8         2.2 virginica
## 106          7.6         3.0          6.6         2.1 virginica
## 107          4.9         2.5          4.5         1.7 virginica
## 108          7.3         2.9          6.3         1.8 virginica
## 109          6.7         2.5          5.8         1.8 virginica
## 110          7.2         3.6          6.1         2.5 virginica
## 111          6.5         3.2          5.1         2.0 virginica
## 112          6.4         2.7          5.3         1.9 virginica
## 113          6.8         3.0          5.5         2.1 virginica
## 114          5.7         2.5          5.0         2.0 virginica
## 115          5.8         2.8          5.1         2.4 virginica
## 116          6.4         3.2          5.3         2.3 virginica
## 117          6.5         3.0          5.5         1.8 virginica
## 118          7.7         3.8          6.7         2.2 virginica
## 119          7.7         2.6          6.9         2.3 virginica
## 120          6.0         2.2          5.0         1.5 virginica
## 121          6.9         3.2          5.7         2.3 virginica
## 122          5.6         2.8          4.9         2.0 virginica
## 123          7.7         2.8          6.7         2.0 virginica
## 124          6.3         2.7          4.9         1.8 virginica
## 125          6.7         3.3          5.7         2.1 virginica
## 126          7.2         3.2          6.0         1.8 virginica
## 127          6.2         2.8          4.8         1.8 virginica
## 128          6.1         3.0          4.9         1.8 virginica
## 129          6.4         2.8          5.6         2.1 virginica
## 130          7.2         3.0          5.8         1.6 virginica
## 131          7.4         2.8          6.1         1.9 virginica
## 132          7.9         3.8          6.4         2.0 virginica
## 133          6.4         2.8          5.6         2.2 virginica
## 134          6.3         2.8          5.1         1.5 virginica
```

```
## 135          6.1         2.6          5.6         1.4 virginica
## 136          7.7         3.0          6.1         2.3 virginica
## 137          6.3         3.4          5.6         2.4 virginica
## 138          6.4         3.1          5.5         1.8 virginica
## 139          6.0         3.0          4.8         1.8 virginica
## 140          6.9         3.1          5.4         2.1 virginica
## 141          6.7         3.1          5.6         2.4 virginica
## 142          6.9         3.1          5.1         2.3 virginica
## 143          5.8         2.7          5.1         1.9 virginica
## 144          6.8         3.2          5.9         2.3 virginica
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

Row `grepl()`

```
knitr::kable(iris[grepl("color",iris$Species),]) ##Species "color"
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor
56	5.7	2.8	4.5	1.3	versicolor
57	6.3	3.3	4.7	1.6	versicolor
58	4.9	2.4	3.3	1.0	versicolor
59	6.6	2.9	4.6	1.3	versicolor
60	5.2	2.7	3.9	1.4	versicolor
61	5.0	2.0	3.5	1.0	versicolor
62	5.9	3.0	4.2	1.5	versicolor
63	6.0	2.2	4.0	1.0	versicolor
64	6.1	2.9	4.7	1.4	versicolor
65	5.6	2.9	3.6	1.3	versicolor
66	6.7	3.1	4.4	1.4	versicolor
67	5.6	3.0	4.5	1.5	versicolor
68	5.8	2.7	4.1	1.0	versicolor
69	6.2	2.2	4.5	1.5	versicolor
70	5.6	2.5	3.9	1.1	versicolor
71	5.9	3.2	4.8	1.8	versicolor
72	6.1	2.8	4.0	1.3	versicolor
73	6.3	2.5	4.9	1.5	versicolor
74	6.1	2.8	4.7	1.2	versicolor
75	6.4	2.9	4.3	1.3	versicolor
76	6.6	3.0	4.4	1.4	versicolor
77	6.8	2.8	4.8	1.4	versicolor
78	6.7	3.0	5.0	1.7	versicolor
79	6.0	2.9	4.5	1.5	versicolor
80	5.7	2.6	3.5	1.0	versicolor
81	5.5	2.4	3.8	1.1	versicolor
82	5.5	2.4	3.7	1.0	versicolor
83	5.8	2.7	3.9	1.2	versicolor
84	6.0	2.7	5.1	1.6	versicolor
85	5.4	3.0	4.5	1.5	versicolor
86	6.0	3.4	4.5	1.6	versicolor
87	6.7	3.1	4.7	1.5	versicolor
88	6.3	2.3	4.4	1.3	versicolor
89	5.6	3.0	4.1	1.3	versicolor
90	5.5	2.5	4.0	1.3	versicolor
91	5.5	2.6	4.4	1.2	versicolor
92	6.1	3.0	4.6	1.4	versicolor
93	5.8	2.6	4.0	1.2	versicolor
94	5.0	2.3	3.3	1.0	versicolor
95	5.6	2.7	4.2	1.3	versicolor
96	5.7	3.0	4.2	1.2	versicolor
97	5.7	2.9	4.2	1.3	versicolor
98	6.2	2.9	4.3	1.3	versicolor
99	5.1	2.5	3.0	1.1	versicolor
100	5.7	2.8	4.1	1.3	versicolor

(Raw) `head()` `tail()`

```
head(iris,5) ## iris
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
```

```
tail(iris,3) ## iris
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

6.5

6.5.1 sort

sort()

```
head(islands) ##
```

```
##       Africa   Antarctica         Asia    Australia Axel Heiberg       Baffin
##        11506         5500        16988         2968           16          184
```

```
head(sort(islands)) ##
```

```
##       Vancouver          Hainan Prince of Wales           Timor          Kyushu
##              12              13              13              13              14
##          Taiwan
##              14
```

decreasing TRUE

```
head(sort(islands,decreasing = T)) ##
```

```
##          Asia        Africa North America South America    Antarctica
##         16988         11506          9390          6795          5500
##        Europe
##          3745
```

6.5.2 order

order() order() iris$Sepal.Length 14 iris$Sepal.Length 14

```
order(iris$Sepal.Length)
```

```
##   [1]  14   9  39  43  42   4   7  23  48   3  30  12  13  25  31  46   2  10
##  [19]  35  38  58 107   5   8  26  27  36  41  44  50  61  94   1  18  20  22
##  [37]  24  40  45  47  99  28  29  33  60  49   6  11  17  21  32  85  34  37
##  [55]  54  81  82  90  91  65  67  70  89  95 122  16  19  56  80  96  97 100
##  [73] 114  15  68  83  93 102 115 143  62  71 150  63  79  84  86 120 139  64
##  [91]  72  74  92 128 135  69  98 127 149  57  73  88 101 104 124 134 137 147
## [109]  52  75 112 116 129 133 138  55 105 111 117 148  59  76  66  78  87 109
## [127] 125 141 145 146  77 113 144  53 121 140 142  51 103 110 126 130 108 131
```

```
## [145] 106 118 119 123 136 132
```

```
iris$Sepal.Length[14]
```

```
## [1] 4.3
```

decreasing TRUE iris$Sepal.Length 132 iris$Sepal.Length 132

```
order(iris$Sepal.Length,decreasing = T)
```

```
##   [1] 132 118 119 123 136 106 131 108 110 126 130 103  51  53 121 140 142  77
##  [19] 113 144  66  78  87 109 125 141 145 146  59  76  55 105 111 117 148  52
##  [37]  75 112 116 129 133 138  57  73  88 101 104 124 134 137 147  69  98 127
##  [55] 149  64  72  74  92 128 135  63  79  84  86 120 139  62  71 150  15  68
##  [73]  83  93 102 115 143  16  19  56  80  96  97 100 114  65  67  70  89  95
##  [91] 122  34  37  54  81  82  90  91   6  11  17  21  32  85  49  28  29  33
## [109]  60   1  18  20  22  24  40  45  47  99   5   8  26  27  36  41  44  50
## [127]  61  94   2  10  35  38  58 107  12  13  25  31  46   3  30   4   7  23
## [145]  48  42   9  39  43  14
```

```
iris$Sepal.Length[132]
```

```
## [1] 7.9
```

order iris

```
head(iris) ##
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
head(iris[order(iris$Sepal.Length),]) ## Sepal.Length
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 14          4.3         3.0          1.1         0.1  setosa
## 9           4.4         2.9          1.4         0.2  setosa
## 39          4.4         3.0          1.3         0.2  setosa
## 43          4.4         3.2          1.3         0.2  setosa
## 42          4.5         2.3          1.3         0.3  setosa
## 4           4.6         3.1          1.5         0.2  setosa
```

```
head(iris[order(iris$Sepal.Length,decreasing = T),]) ##
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 132          7.9         3.8          6.4         2.0 virginica
## 118          7.7         3.8          6.7         2.2 virginica
## 119          7.7         2.6          6.9         2.3 virginica
## 123          7.7         2.8          6.7         2.0 virginica
## 136          7.7         3.0          6.1         2.3 virginica
## 106          7.6         3.0          6.6         2.1 virginica
```

6.6

- Row rbind()
- Column cbind()

rbind() cbind() :

```
rbind(c(1,2,3), #
      c(4,5,6)  #
      )
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6
```

:

```
irisAdd<-rbind(iris, #
      c(1,1,1,1,"versicolor")  #
      )
tail(irisAdd)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 146          6.7           3          5.2         2.3  virginica
## 147          6.3         2.5            5         1.9  virginica
## 148          6.5           3          5.2           2  virginica
## 149          6.2         3.4          5.4         2.3  virginica
## 150          5.9           3          5.1         1.8  virginica
## 151            1           1            1           1 versicolor
```

:

```
cbind(c(1,2,3), #
      c(4,5,6)  #
      )
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

:

```
irisAdd<-cbind(iris, #
      rep("Add",nrow(iris))  #
      )
tail(irisAdd)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
##     rep("Add", nrow(iris))
## 145                    Add
## 146                    Add
## 147                    Add
## 148                    Add
## 149                    Add
## 150                    Add
```

6.7

R reshape2(?) package

- melt(/ ,id.vars=)
- dcast(/ , ~)

airquality Ozone, Solar.R, Wind, Temp, Month, Day (Column) Month Day variable value

```
library(reshape2)
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
airqualityM<-melt(airquality,id.vars = c("Month","Day")) ##   "Month","Day"
head(airqualityM)
```

```
##   Month Day variable value
## 1     5   1    Ozone    41
## 2     5   2    Ozone    36
## 3     5   3    Ozone    12
## 4     5   4    Ozone    18
## 5     5   5    Ozone    NA
## 6     5   6    Ozone    28
```

airqualityM Month, Day, variable, value (Column) variable value :

```
library(reshape2)
##  "Month","Day"     variable
airqualityCast<-dcast(airqualityM, Month +Day~variable)
head(airqualityCast)
```

```
##   Month Day Ozone Solar.R Wind Temp
## 1     5   1    41     190  7.4   67
## 2     5   2    36     118  8.0   72
## 3     5   3    12     149 12.6   74
## 4     5   4    18     313 11.5   62
## 5     5   5    NA      NA 14.3   56
## 6     5   6    28      NA 14.9   66
```

6.8

(Missing Value) is.na() NA TRUE

```
naVec<-c("a","b",NA,"d","e")
is.na(naVec)
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE
```

```
naVec[!is.na(naVec)] ##   is.na()  FALSE
```

```
## [1] "a" "b" "d" "e"
```

complete.cases TRUE

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
complete.cases(airquality)
```

```
##   [1]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
##  [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [25] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [37] FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
##  [49]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [61] FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
##  [73]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
##  [85]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
##  [97] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
## [109]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
## [121]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [133]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [145]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
```

```
head(airquality[complete.cases(airquality),]) ##   complete.cases()  TRUE
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 7    23     299  8.6   65     5   7
## 8    19      99 13.8   59     5   8
```

skydome20 R –(10) (Impute Missing Value)

6.9 Case study

SportsAnalytics (**?**) package NBA

6.9.1

library() SportsAnalytics (Chapter **??**) fetch_NBAPlayerStatistics()

```
library(SportsAnalytics)
NBA1516<-fetch_NBAPlayerStatistics("15-16")
```

6.9.2

str() NBA1516

```
str(NBA1516)
```

```
## 'data.frame':    476 obs. of  25 variables:
##  $ League            : Factor w/ 1 level "NBA": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Name              : chr  "Quincy Acy" "Jordan Adams" "Steven Adams" "Arron Afflalo" ...
##  $ Team              : Factor w/ 31 levels "ATL","BOS","BRO",..: 27 15 22 20 19 13 28 26 12 15 ...
##  $ Position          : Factor w/ 5 levels "C","PF","PG",..: 4 5 1 5 1 1 2 2 2 5 ...
##  $ GamesPlayed       : int  59 2 80 71 59 60 74 9 79 64 ...
##  $ TotalMinutesPlayed: int  877 15 2019 2359 863 802 2260 37 1601 1622 ...
##  $ FieldGoalsMade    : int  119 2 261 354 150 134 536 5 191 215 ...
##  $ FieldGoalsAttempted: int  214 6 426 799 314 225 1045 10 370 469 ...
##  $ ThreesMade        : int  19 0 0 91 0 0 0 0 0 15 ...
##  $ ThreesAttempted   : int  49 1 0 238 1 0 16 0 0 42 ...
##  $ FreeThrowsMade    : int  50 3 114 110 52 60 259 0 46 90 ...
##  $ FreeThrowsAttempted: int  68 5 196 131 62 84 302 0 73 138 ...
##  $ OffensiveRebounds : int  65 0 218 23 75 86 175 2 162 104 ...
##  $ TotalRebounds     : int  188 2 531 266 269 288 631 6 424 297 ...
##  $ Assists           : int  27 3 61 145 32 50 110 0 76 70 ...
##  $ Steals            : int  29 3 42 25 19 47 38 1 26 109 ...
##  $ Turnovers         : int  27 2 84 82 54 64 99 1 69 78 ...
##  $ Blocks            : int  24 0 89 10 36 68 81 2 42 18 ...
##  $ PersonalFouls     : int  103 2 223 142 134 139 151 1 147 175 ...
##  $ Disqualifications : int  0 0 2 1 0 1 0 0 1 1 ...
##  $ TotalPoints       : int  307 7 636 909 352 328 1331 10 428 535 ...
##  $ Technicals        : int  3 0 2 1 2 0 0 0 0 1 ...
##  $ Ejections         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FlagrantFouls     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GamesStarted      : int  29 0 80 57 17 5 74 0 28 56 ...
```

NBA1516 476 (, obs) 25 (variables) ### head() tail()

```
head(NBA1516)
```

```
##   League          Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA    Quincy Acy  SAC       SF          59                877
## 2    NBA  Jordan Adams  MEM       SG           2                 15
## 3    NBA  Steven Adams  OKL        C          80               2019
## 4    NBA Arron Afflalo  NYK       SG          71               2359
## 5    NBA Alexis Ajinca  NOR        C          59                863
## 6    NBA  Cole Aldrich  LAC        C          60                802
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1            119                 214         19              49             50
## 2              2                   6          0               1              3
## 3            261                 426          0               0            114
## 4            354                 799         91             238            110
## 5            150                 314          0               1             52
## 6            134                 225          0               0             60
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                  68                65           188      27     29        27
## 2                   5                 0             2       3      3         2
## 3                 196               218           531      61     42        84
## 4                 131                23           266     145     25        82
## 5                  62                75           269      32     19        54
## 6                  84                86           288      50     47        64
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
```

```
## 1     24           103                0         307          3         0
## 2      0             2                0           7          0         0
## 3     89           223                2         636          2         0
## 4     10           142                1         909          1         0
## 5     36           134                0         352          2         0
## 6     68           139                1         328          0         0
##   FlagrantFouls GamesStarted
## 1             0           29
## 2             0            0
## 3             0           80
## 4             0           57
## 5             0           17
## 6             0            5
```

6.9.3

order() (decreasing = T) [,]

```
NBA1516Order<-NBA1516[order(NBA1516$GamesPlayed,decreasing = T),]
NBA1516Order[1:5,] ##  1~5  1~5
```

```
##     League            Name Team Position GamesPlayed TotalMinutesPlayed
## 11     NBA Al-farouq Aminu  POR       SF          82               2342
## 37     NBA     Will Barton  DEN       SG          82               2355
## 48     NBA Bismack Biyombo  TOR       PF          82               1810
## 62     NBA    Corey Brewer  HOU       SG          82               1670
## 118    NBA    Gorgui Dieng  MIN        C          82               2222
##     FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted
## 11             299                719        126             349
## 37             426                984        112             324
## 48             156                288          0               1
## 62             212                552         61             225
## 118            308                578          6              20
##     FreeThrowsMade FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists
## 11             115                156               98           498     138
## 37             216                268               60           477     204
## 48             142                226              182           655      29
## 62             105                140               42           199     109
## 118            205                248              156           584     143
##     Steals Turnovers Blocks PersonalFouls Disqualifications TotalPoints
## 11      72       120     53           171                 0         839
## 37      71       139     39           147                 0        1180
## 48      19        71    133           225                 2         454
## 62      84        78     19           168                 1         590
## 118     94       140     96           219                 0         827
##     Technicals Ejections FlagrantFouls GamesStarted
## 11           3         0             0           82
## 37           2         0             0            1
## 48           3         0             0           22
## 62           0         0             0           12
## 118          1         0             0           39
```

order() (decreasing = T) [,]

```
NBA1516OrderM<-NBA1516[order(NBA1516$TotalMinutesPlayed,decreasing = T),]
NBA1516OrderM[1:10,"Name"] ##   1~10    "Name"    Name
```

```
##  [1] "James Harden"     "Gordon Hayward"   "Kemba Walker"     "Trevor Ariza"
##  [5] "Khris Middleton"  "Kyle Lowry"       "Marcus Morris"    "Andrew Wiggins"
##  [9] "Paul George"      "Gi Antetokounmpo"
```

6.9.4

subset()

```
subset(NBA1516,Team=="BOS")
```

```
##     League            Name Team Position GamesPlayed TotalMinutesPlayed
## 60     NBA   Avery Bradley  BOS       PG          76                2536
## 89     NBA     Coty Clarke  BOS     <NA>           4                   8
## 102    NBA     Jae Crowder  BOS       SF          73                2310
## 213    NBA     R.j. Hunter  BOS       SG          36                 319
## 228    NBA   Jonas Jerebko  BOS       PF          78                1178
## 229    NBA    Amir Johnson  BOS       PF          79                1798
## 300    NBA   Jordan Mickey  BOS       PF          16                  59
## 340    NBA    Kelly Olynyk  BOS        C          69                1396
## 382    NBA    Terry Rozier  BOS       PG          39                 310
## 400    NBA    Marcus Smart  BOS       PG          61                1666
## 416    NBA Jared Sullinger  BOS       PF          81                1917
## 422    NBA   Isaiah Thomas  BOS       PG          82                2647
## 433    NBA     Evan Turner  BOS       SG          81                2270
## 471    NBA     James Young  BOS       SG          29                 200
## 476    NBA    Tyler Zeller  BOS        C          60                 714
##     FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted
## 60             456                1018        147             406
## 89               2                   4          2               2
## 102            359                 812        122             363
## 213             36                  98         19              63
## 228            118                 286         43             108
## 229            250                 427         10              43
## 300              8                  22          0               0
## 340            253                 556         85             210
## 382             29                 106          6              27
## 400            184                 529         61             241
## 416            351                 807         29             104
## 422            591                1382        167             465
## 433            343                 753         20              83
## 471             11                  36          6              26
## 476            138                 290          0               0
##     FreeThrowsMade FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists
## 60              96                 123                48           220     158
## 89               0                   0                 0             1       0
## 102            196                 239                70           373     135
## 213              6                   7                 2            37      13
## 228             61                  78                77           288      62
## 229             69                 121               178           505     137
## 300              5                  10                 6            13       1
## 340             96                 128                72           281     105
```

```
## 382              8                 10                 24             63      37
## 400            129                166                 76            255     186
## 416            103                161                194            673     187
## 422            474                544                 46            243     509
## 433            148                179                 50            397     359
## 471              1                  4                  4             26       9
## 476             88                108                 62            178      29
##     Steals Turnovers Blocks PersonalFouls Disqualifications TotalPoints
## 60     117       109     19           164                 2        1155
## 89       0         1      0             0                 0           6
## 102    126        83     35           198                 4        1036
## 213     14        11      4            29                 0          97
## 228     20        52     24           137                 2         340
## 229     52        94     83           214                 1         579
## 300      0         1     11             5                 0          21
## 340     52        74     33           163                 3         687
## 382      6        19      1            23                 0          72
## 400     91        80     18           183                 1         558
## 416     75       102     47           209                 2         834
## 422     91       220      9           167                 1        1823
## 433     80       169     28           139                 0         854
## 471      6         5      1            17                 0          29
## 476     10        46     22            97                 1         364
##     Technicals Ejections FlagrantFouls GamesStarted
## 60           0         0             0           72
## 89           0         0             0            0
## 102          3         0             0           73
## 213          0         0             0            0
## 228          1         0             0            0
## 229          0         0             0           76
## 300          0         0             0            0
## 340          1         0             0            8
## 382          0         0             0            0
## 400          2         0             0           10
## 416          2         0             0           73
## 422          9         0             0           79
## 433          2         0             0           12
## 471          0         0             0            0
## 476          0         0             0            3
```

6.9.5

grepl() “James”

```
NBA1516[grepl("James",NBA1516$Name),]
```

```
##     League           Name Team Position GamesPlayed TotalMinutesPlayed
## 15     NBA James Anderson  SAC       SG          51                721
## 132    NBA    James Ennis  NOR       SF          22                329
## 178    NBA   James Harden  HOU       SG          82               3121
## 222    NBA   Lebron James  CLE       SF          76               2710
## 231    NBA  James Johnson  TOR       PF          57                924
## 239    NBA    James Jones  CLE       SG          48                466
## 286    NBA   James Mcadoo  GSW       SG          41                265
```

```
## 471    NBA    James Young  BOS       SG          29                200
##     FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted
## 15              67                 178         23              86
## 132             54                 113         26              58
## 178            710                1617        236             656
## 222            737                1416         87             282
## 231            114                 240         20              66
## 239             59                 143         41             104
## 286             45                  84          1               2
## 471             11                  36          6              26
##     FreeThrowsMade FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists
## 15              22                  29                13            86      41
## 132             25                  34                21            42      21
## 178            720                 837                63           502     612
## 222            359                 491               111           565     512
## 231             39                  68                28           126      67
## 239             21                  26                 8            50      14
## 286             26                  49                30            58      17
## 471              1                   4                 4            26       9
##     Steals Turnovers Blocks PersonalFouls Disqualifications TotalPoints
## 15      21        42     14            54                 0         179
## 132     16        19      5            28                 1         159
## 178    138       374     51           229                 1        2376
## 222    104       249     49           143                 0        1920
## 231     29        54     33            84                 0         287
## 239     11        13     10            50                 0         180
## 286     10        16      8            39                 0         117
## 471      6         5      1            17                 0          29
##     Technicals Ejections FlagrantFouls GamesStarted
## 15           0         0             0           15
## 132          0         0             0            5
## 178          2         0             0           82
## 222          3         0             0           76
## 231          0         0             0           32
## 239          1         0             0            0
## 286          0         0             0            1
## 471          0         0             0            0
```

Chapter 7

7.1

(**E**xploratory **D**ata **A**nalysis)

/

:

- Graphical Quantitative
- Univariate Bivariate Multivariate

Ch **??**

- ( )
  - Mean `mean()`
  - Median `median()`
  - Mode R `table()`
-
  - Min `min()`
  - Max `max()`
  - Range `range()`
  - Quartiles `quantile()`
  - Variance `var()`
  - Standard deviation `sd()`

:

- Crosstabs `table()`, `ftable()`, `prop.table()`
- Covariance `cov()`
- Correlation `cor()`

R A B ... `data.ta`
packages packages

7.2 data.table

data.table data.frame data.table(**?**) package

```
install.packages("data.table") ##
```

```
library(data.table) ##
```

data.table fread() (Ch ??)

```
fread("  ")
```

data.frame data.table() data.frame data.table NBA Ch ?? SportsAnalytics

```
library(SportsAnalytics)
library(data.table)
NBA1516<-fetch_NBAPlayerStatistics("15-16")
NBA1516DT<-data.table(NBA1516)
class(NBA1516DT)
```

```
## [1] "data.table" "data.frame"
```

NBA1516DT data.table data.frame data.table data.frame data.table data.frame

data.table

DT[i,j,by=]

- i (Row)
- j (Column)
- by

, , i j DT[i,j]

7.2.1 i

i (Row) Ch ?? **(index)**

NBA James

```
NBA1516DT[grepl('James',Name)]
```

```
##    League           Name Team Position GamesPlayed TotalMinutesPlayed
## 1:    NBA James Anderson  SAC       SG          51                721
## 2:    NBA    James Ennis  NOR       SF          22                329
## 3:    NBA   James Harden  HOU       SG          82               3121
## 4:    NBA   Lebron James  CLE       SF          76               2710
## 5:    NBA  James Johnson  TOR       PF          57                924
## 6:    NBA    James Jones  CLE       SG          48                466
## 7:    NBA   James Mcadoo  GSW       SG          41                265
## 8:    NBA    James Young  BOS       SG          29                200
##    FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1:             67                 178         23              86             22
## 2:             54                 113         26              58             25
## 3:            710                1617        236             656            720
## 4:            737                1416         87             282            359
## 5:            114                 240         20              66             39
## 6:             59                 143         41             104             21
## 7:             45                  84          1               2             26
## 8:             11                  36          6              26              1
##    FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1:                  29                13            86      41     21        42
## 2:                  34                21            42      21     16        19
```

```
## 3:                837                63           502     612    138       374
## 4:                491               111           565     512    104       249
## 5:                 68                28           126      67     29        54
## 6:                 26                 8            50      14     11        13
## 7:                 49                30            58      17     10        16
## 8:                  4                 4            26       9      6         5
##    Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1:     14            54                 0         179          0         0
## 2:      5            28                 1         159          0         0
## 3:     51           229                 1        2376          2         0
## 4:     49           143                 0        1920          3         0
## 5:     33            84                 0         287          0         0
## 6:     10            50                 0         180          1         0
## 7:      8            39                 0         117          0         0
## 8:      1            17                 0          29          0         0
##    FlagrantFouls GamesStarted
## 1:             0           15
## 2:             0            5
## 3:             0           82
## 4:             0           76
## 5:             0           32
## 6:             0            0
## 7:             0            1
## 8:             0            0
```

"A"

```
NBA1516DT[grepl('A',Name)&Position=="C"]
```

```
##     League             Name Team Position GamesPlayed TotalMinutesPlayed
##  1:    NBA     Steven Adams  OKL        C          80               2019
##  2:    NBA    Alexis Ajinca  NOR        C          59                863
##  3:    NBA     Cole Aldrich  LAC        C          60                802
##  4:    NBA     Joel Anthony  DET        C          19                 95
##  5:    NBA        Omer Asik  NOR        C          68               1181
##  6:    NBA  Andrea Bargnani  BRO        C          46                634
##  7:    NBA     Andrew Bogut  GSW        C          70               1452
##  8:    NBA   Andre Drummond  DET        C          81               2664
##  9:    NBA     Al Jefferson  CHA        C          47               1096
## 10:    NBA         Alex Len  PHO        C          78               1820
## 11:    NBA Anderson Varejao  GSW        C          53                494
## 12:    NBA    Alan Williams  PHO        C          10                 67
##     FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted
##  1:            261                 426          0               0
##  2:            150                 314          0               1
##  3:            134                 225          0               0
##  4:              6                  10          0               0
##  5:            104                 196          0               0
##  6:            127                 278          3              15
##  7:            175                 279          1               1
##  8:            552                1061          2               6
##  9:            245                 505          0               0
## 10:            264                 623          1               7
## 11:             53                 124          0               1
## 12:             10                  24          0               0
```

```
##     FreeThrowsMade FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists
##  1:            114                196               218           531      61
##  2:             52                 62                75           269      32
##  3:             60                 84                86           288      50
##  4:              6                  8                 8            21       1
##  5:             61                112               119           413      26
##  6:             47                 57                28            97      18
##  7:             24                 50               121           491     162
##  8:            208                586               395          1198      67
##  9:             72                111                57           301      70
## 10:            174                239               178           594      97
## 11:             32                 50                37           141      35
## 12:              9                 14                14            38       5
##     Steals Turnovers Blocks PersonalFouls Disqualifications TotalPoints
##  1:     42        84     89           223                 2         636
##  2:     19        54     36           134                 0         352
##  3:     47        64     68           139                 1         328
##  4:      2         2     12            15                 0          18
##  5:     21        60     23           124                 0         269
##  6:      4        26      9            61                 0         304
##  7:     32        83    113           221                 4         375
##  8:    119       154    112           245                 2        1314
##  9:     30        34     41           117                 1         562
## 10:     38       145     62           230                 3         703
## 11:     16        22     10            70                 0         138
## 12:      4         6      5            15                 0          29
##     Technicals Ejections FlagrantFouls GamesStarted
##  1:          2         0             0           80
##  2:          2         0             0           17
##  3:          0         0             0            5
##  4:          0         0             0            0
##  5:          0         0             0           64
##  6:          0         0             0            0
##  7:          0         0             0           66
##  8:          7         0             0           81
##  9:          0         0             0           18
## 10:          1         0             0           46
## 11:          1         0             0            0
## 12:          0         0             0            0
```

70

```
NBA1516DT[GamesPlayed>70]
```

```
##      League             Name Team Position GamesPlayed TotalMinutesPlayed
##   1:    NBA     Steven Adams  OKL        C          80               2019
##   2:    NBA    Arron Afflalo  NYK       SG          71               2359
##   3:    NBA Lamarcu Aldridge  SAN       PF          74               2260
##   4:    NBA      Lavoy Allen  IND       PF          79               1601
##   5:    NBA  Al-farouq Aminu  POR       SF          82               2342
##  ---
## 172:    NBA Derrick Williams  NYK       PF          80               1442
## 173:    NBA  Marvin Williams  CHA       PF          81               2339
## 174:    NBA  Justise Winslow  MIA       SF          78               2237
## 175:    NBA   Thaddeus Young  BRO       SF          73               2413
```

```
## 176:    NBA      Cody Zeller  CHA       PF          73               1773
##      FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted
##   1:            261                 426          0               0
##   2:            354                 799         91             238
##   3:            536                1045          0              16
##   4:            191                 370          0               0
##   5:            299                 719        126             349
##  ---
## 172:            254                 565         44             150
## 173:            338                 747        152             379
## 174:            196                 463         32             116
## 175:            495                 963          7              30
## 176:            231                 437          1              10
##      FreeThrowsMade FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists
##   1:            114                 196               218           531      61
##   2:            110                 131                23           266     145
##   3:            259                 302               175           631     110
##   4:             46                  73               162           424      76
##   5:            115                 156                98           498     138
##  ---
## 172:            194                 256                47           296      75
## 173:            120                 144               127           520     110
## 174:             80                 117                81           403     117
## 175:            105                 163               177           661     135
## 176:            175                 232               138           456      71
##      Steals Turnovers Blocks PersonalFouls Disqualifications TotalPoints
##   1:     42        84     89           223                 2         636
##   2:     25        82     10           142                 1         909
##   3:     38        99     81           151                 0        1331
##   4:     26        69     42           147                 1         428
##   5:     72       120     53           171                 0         839
##  ---
## 172:     30        61      9            69                 0         746
## 173:     58        62     77           133                 1         948
## 174:     68        95     26           184                 0         504
## 175:    112       136     37           182                 3        1102
## 176:     57        68     63           204                 4         638
##      Technicals Ejections FlagrantFouls GamesStarted
##   1:          2         0             0           80
##   2:          1         0             0           57
##   3:          0         0             0           74
##   4:          0         0             0           28
##   5:          3         0             0           82
##  ---
## 172:          0         0             0            9
## 173:          0         0             0           81
## 174:          0         0             0            8
## 175:          1         0             0           73
## 176:          0         0             0           60
```

7.2.2 j

```
j
```

```
NBA1516DT[,mean(GamesPlayed)] ##     ,
```

```
## [1] 55
```

                        j .()

```
NBA1516DT[,.(mean(GamesPlayed),mean(PersonalFouls),mean(Steals))] ##     ,
```

```
##    V1  V2 V3
## 1: 55 105 41
```

            V1, V2, V3   =

```
NBA1516DT[,.(GamesPlayedMean=mean(GamesPlayed),
             PersonalFoulsMean=mean(PersonalFouls),
             StealsMean=mean(Steals))]
```

```
##    GamesPlayedMean PersonalFoulsMean StealsMean
## 1:              55               105         41
```

```
NBA1516DT[,.(GamesPlayedMax=max(GamesPlayed), #
             ThreesMadeMin=min(ThreesMade), #
             FieldGoalsMadeSD=sd(FieldGoalsMade))] #
```

```
##    GamesPlayedMax ThreesMadeMin FieldGoalsMadeSD
## 1:             82             0              166
```

             70

```
NBA1516DT[GamesPlayed>70,
          .(ThreesMadeMean=mean(ThreesMade), FieldGoalsMadeMean=mean(FieldGoalsMade))]
```

```
##    ThreesMadeMean FieldGoalsMadeMean
## 1:             76                335
```

7.2.3 by

  by   NBA   data.table .N   mean()   by=   (Team)

```
NBA1516DT[,.(.N,AssistsMean=mean(Assists)),
          by=Team]
```

```
##     Team  N AssistsMean
##  1:  SAC 15         134
##  2:  MEM 22          74
##  3:  OKL 14         126
##  4:  NYK 16         105
##  5:  NOR 21          87
##  6:  LAC 15         124
##  7:  SAN 16         130
##  8:  POR 15         116
##  9:  IND 15         125
## 10:  WAS 17         127
## 11:  DAL 15         124
## 12:  MIL 17         113
## 13:  DET 15         105
## 14:  ORL 15         128
## 15:  HOU 16         101
```

```
## 16:  LAL 15          99
## 17:  DEN 15         122
## 18:  CHI 15         121
## 19:  GSW 15         158
## 20:  BRO 16         100
## 21:  CHA 14         118
## 22:  ATL 15         142
## 23:  TOR 16          97
## 24:  MIN 14         129
## 25:  PHO 17          97
## 26:  UTA 17          94
## 27:  MIA 15         131
## 28:  BOS 15         128
## 29:  PHI 16         118
## 30:  CLE 16         117
## 31:  OKC  1         160
##     Team  N AssistsMean
```

.N data.table

NBA

```
NBA1516DT[Position=="C",
          .(.N,ThreesAttemptedMean=mean(ThreesAttempted)),
          by=Team]
```

```
##     Team N ThreesAttemptedMean
##  1:  OKL 3                7.00
##  2:  NOR 4                0.25
##  3:  LAC 2                0.50
##  4:  DET 2                3.00
##  5:  BRO 3               10.00
##  6:  LAL 3                0.67
##  7:  WAS 2                1.00
##  8:  GSW 4               16.00
##  9:  SAN 3               34.67
## 10:  HOU 2                3.50
## 11:  SAC 3               70.67
## 12:  PHO 3                2.67
## 13:  ORL 2                4.50
## 14:  MIN 3               36.00
## 15:  MEM 2                1.50
## 16:  UTA 3                1.33
## 17:  IND 2                0.50
## 18:  CHA 1                0.00
## 19:  DEN 2               43.50
## 20:  POR 2              116.00
## 21:  CLE 3                2.33
## 22:  NYK 1                1.00
## 23:  DAL 3                1.67
## 24:  MIL 2                0.50
## 25:  CHI 1                1.00
## 26:  PHI 2                4.00
## 27:  TOR 3                6.00
## 28:  BOS 2              105.00
## 29:  ATL 1                0.00
```

```
## 30:  MIA 1                0.00
##     Team N ThreesAttemptedMean
```

7.2.4

data.table

-
- The data.table R package cheat sheet
- A data.table R tutorial by DataCamp
- DataCamp

7.3 dplyr

dplyr(?) package Hadley Wickham dplyr package

```
install.packages("dplyr") ##
```

```
library(dplyr) ##
```

dplyr

- select(): (Column)
- filter(): (Row)
- mutate():
- summarise():
- group_by():
- arrange():
- rename():
- %>%: the "pipe" operator

NBA

```
library(SportsAnalytics)
NBA1516<-fetch_NBAPlayerStatistics("15-16")
```

7.3.1 select()

select() (Column) select(, 1, 2,...) 1 2 dplyr

- starts_with()
- ends_with()
- contains()
- matches()
- num_range()
- one_of()
- everything()

R ?select_helpers

Name ThreesMade ThreesAttempted FieldGoalsMade FieldGoalsAttempted

```
##
##NBA1516[,c("Name","ThreesMade","ThreesAttempted","FieldGoalsMade","FieldGoalsAttempted")]
select1<-select(NBA1516,Name,starts_with("Threes"),starts_with("FieldGoals"))
head(select1)
```

```
##            Name ThreesMade ThreesAttempted FieldGoalsMade FieldGoalsAttempted
## 1    Quincy Acy         19              49            119                  214
## 2  Jordan Adams          0               1              2                    6
## 3  Steven Adams          0               0            261                  426
## 4 Arron Afflalo         91             238            354                  799
## 5 Alexis Ajinca          0               1            150                  314
## 6  Cole Aldrich          0               0            134                  225
```

Name FreeThrowsAttempted :

```
##  NBA1516[,2:12]
select2<-select(NBA1516,Name:FreeThrowsAttempted)
head(select2)
```

```
##            Name Team Position GamesPlayed TotalMinutesPlayed FieldGoalsMade
## 1    Quincy Acy  SAC       SF          59                877            119
## 2  Jordan Adams  MEM       SG           2                 15              2
## 3  Steven Adams  OKL        C          80               2019            261
## 4 Arron Afflalo  NYK       SG          71               2359            354
## 5 Alexis Ajinca  NOR        C          59                863            150
## 6  Cole Aldrich  LAC        C          60                802            134
##   FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1                 214         19              49             50
## 2                   6          0               1              3
## 3                 426          0               0            114
## 4                 799         91             238            110
## 5                 314          0               1             52
## 6                 225          0               0             60
##   FreeThrowsAttempted
## 1                  68
## 2                   5
## 3                 196
## 4                 131
## 5                  62
## 6                  84
```

```
##  NBA1516[,c(2:4,612)]
select3<-select(NBA1516,Name:FreeThrowsAttempted,-GamesPlayed)
head(select3)
```

```
##            Name Team Position TotalMinutesPlayed FieldGoalsMade
## 1    Quincy Acy  SAC       SF                877            119
## 2  Jordan Adams  MEM       SG                 15              2
## 3  Steven Adams  OKL        C               2019            261
## 4 Arron Afflalo  NYK       SG               2359            354
## 5 Alexis Ajinca  NOR        C                863            150
## 6  Cole Aldrich  LAC        C                802            134
##   FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1                 214         19              49             50
## 2                   6          0               1              3
## 3                 426          0               0            114
## 4                 799         91             238            110
## 5                 314          0               1             52
## 6                 225          0               0             60
##   FreeThrowsAttempted
## 1                  68
```

```
## 2                   5
## 3                 196
## 4                 131
## 5                  62
## 6                  84
```

7.3.2 filter()

filter() (Row) filter(,) 2850

```
##  NBA1516[NBA1516$TotalMinutesPlayed>2850,]
filter1<-filter(NBA1516,TotalMinutesPlayed>2850)
filter1
```

```
##   League            Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA    Trevor Ariza  HOU       SF          81               2860
## 2    NBA    James Harden  HOU       SG          82               3121
## 3    NBA  Gordon Hayward  UTA       SG          80               2889
## 4    NBA      Kyle Lowry  TOR       PG          77               2853
## 5    NBA Khris Middleton  MIL       SF          79               2855
## 6    NBA   Marcus Morris  DET       SF          80               2852
## 7    NBA    Kemba Walker  CHA       PG          81               2885
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1            357                 858        185             497            126
## 2            710                1617        236             656            720
## 3            521                1202        143             410            393
## 4            512                1198        212             546            398
## 5            507                1144        143             362            277
## 6            410                 945        108             297            203
## 7            568                1332        182             490            371
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                 161                67           366     188    161       113
## 2                 837                63           502     612    138       374
## 3                 477                61           397     296     95       202
## 4                 491                55           365     494    158       225
## 5                 312                45           301     331    131       180
## 6                 271                91           404     201     67       140
## 7                 438                56           358     421    127       171
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1     26           177                 0        1025          2         0
## 2     51           229                 1        2376          2         0
## 3     27           183                 0        1578          0         0
## 4     34           211                 1        1634          9         0
## 5     19           204                 1        1434          5         0
## 6     23           170                 1        1131         11         0
## 7     39           111                 0        1689          5         0
##   FlagrantFouls GamesStarted
## 1             0           81
## 2             0           82
## 3             0           80
## 4             0           77
## 5             0           79
## 6             0           80
## 7             0           81
```

"BOS" "SAN"

```
##  NBA1516[NBA1516$Team %in% c("BOS","SAN"),]
filter2<-filter(NBA1516,Team %in% c("BOS","SAN"))
head(filter2)
```

```
##   League             Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA Lamarcu Aldridge  SAN       PF          74               2260
## 2    NBA    Kyle Anderson  SAN       SF          78               1247
## 3    NBA      Matt Bonner  SAN        C          30                210
## 4    NBA    Avery Bradley  BOS       PG          76               2536
## 5    NBA    Rasual Butler  SAN       SF          46                432
## 6    NBA      Coty Clarke  BOS     <NA>           4                  8
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1            536                1045          0              16            259
## 2            138                 296         12              37             62
## 3             29                  58         15              35              3
## 4            456                1018        147             406             96
## 5             49                 105         15              49             11
## 6              2                   4          2               2              0
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                 302               175           631     110     38        99
## 2                  83                25           245     123     60        59
## 3                   4                 3            27       9      6         3
## 4                 123                48           220     158    117       109
## 5                  16                 3            56      24     13         8
## 6                   0                 0             1       0      0         1
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1     81           151                 0        1331          0         0
## 2     29            97                 0         350          0         0
## 3      1            16                 0          76          0         0
## 4     19           164                 2        1155          0         0
## 5     23            11                 0         124          0         0
## 6      0             0                 0           6          0         0
##   FlagrantFouls GamesStarted
## 1             0           74
## 2             0           11
## 3             0            2
## 4             0           72
## 5             0            0
## 6             0            0
```

filter()

```
##
filter3<-filter(NBA1516,FieldGoalsMade/FieldGoalsAttempted>0.7)
filter3
```

```
##   League             Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA Th Antetokounmpo  NYK       SF           3                  7
## 2    NBA Rakeem Christmas  IND       PF           1                  6
## 3    NBA   Deandre Jordan  LAC        C          77               2600
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1              3                   4          0               1              0
## 2              2                   2          0               0              0
## 3            357                 507          0               1            266
```

```
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                   0                 0             1       0      0         0
## 2                   0                 1             1       0      0         0
## 3                 619               267          1059      90     52       107
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1      0             2                 0           6          0         0
## 2      0             1                 0           4          0         0
## 3    176           207                 1         980         10         0
##   FlagrantFouls GamesStarted
## 1             0            0
## 2             0            0
## 3             0           77
```

& |

```
##
filter4<-filter(NBA1516,FieldGoalsMade/FieldGoalsAttempted>0.7 & GamesPlayed>30)
filter4
```

```
##   League           Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA Deandre Jordan  LAC        C          77               2600
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1            357                 507          0               1            266
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                 619               267          1059      90     52       107
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1    176           207                 1         980         10         0
##   FlagrantFouls GamesStarted
## 1             0           77
```

7.3.3 mutate()

mutate() FieldGoalsRate FieldGoalsMade/FieldGoalsAttempted

```
mutate1<-mutate(NBA1516,FieldGoalsRate=FieldGoalsMade/FieldGoalsAttempted)
mutate1$FieldGoalsRate[1:10]
```

```
##  [1] 0.56 0.33 0.61 0.44 0.48 0.60 0.51 0.50 0.52 0.46
```

7.3.4 summarise()

summarise()

```
sum1<-summarise(NBA1516,
                nPlayer=n(),
                nTeam=n_distinct(Team),
                nPosition=n_distinct(Position))
sum1
```

```
##   nPlayer nTeam nPosition
## 1     476    31         6
```

filter() Ch ?? **2500**

```
filter1<-filter(NBA1516,TotalMinutesPlayed>2500)
sum2<-summarise(filter1,
                nPlayer=n(),
```

```
                meanFieldGoalsMade=mean(FieldGoalsMade),
                meanFieldGoalsAttempted=mean(FieldGoalsAttempted))
sum2
```

```
##   nPlayer meanFieldGoalsMade meanFieldGoalsAttempted
## 1      40                512                    1121
```

pipe %>% filter1

```
sum3<-filter(NBA1516,TotalMinutesPlayed>2500) %>%
  summarise(nPlayer=n(),meanFieldGoalsMade=mean(FieldGoalsMade),
                meanFieldGoalsAttempted=mean(FieldGoalsAttempted))
sum3
```

```
##   nPlayer meanFieldGoalsMade meanFieldGoalsAttempted
## 1      40                512                    1121
```

7.3.5 group_by()

group_by() summarise() Ch ?? Team

```
group1<-group_by(NBA1516,Team)%>%
  summarise(nPlayer=n(),meanFieldGoalsMade=mean(FieldGoalsMade),
                meanFieldGoalsAttempted=mean(FieldGoalsAttempted))
head(group1)
```

```
## # A tibble: 6 × 4
##     Team nPlayer meanFieldGoalsMade meanFieldGoalsAttempted
##   <fctr>   <int>              <dbl>                   <dbl>
## 1    ATL      15                215                     471
## 2    BOS      15                209                     475
## 3    BRO      16                181                     396
## 4    CHA      14                199                     451
## 5    CHI      15                209                     475
## 6    CLE      16                200                     433
```

Team Position

```
group2<-group_by(NBA1516,Team,Position)%>%
  summarise(nPlayer=n(),meanFieldGoalsMade=mean(FieldGoalsMade),
                meanFieldGoalsAttempted=mean(FieldGoalsAttempted))
head(group2)
```

```
## Source: local data frame [6 x 5]
## Groups: Team [2]
##
##     Team Position nPlayer meanFieldGoalsMade meanFieldGoalsAttempted
##   <fctr>   <fctr>   <int>              <dbl>                   <dbl>
## 1    ATL        C       1                 11                      19
## 2    ATL       PF       6                247                     516
## 3    ATL       PG       2                382                     884
## 4    ATL       SG       6                161                     364
## 5    BOS        C       2                196                     423
## 6    BOS       PF       4                182                     386
```

7.3.6 arrange()

```
arrange1<-arrange(NBA1516,TotalMinutesPlayed)
head(arrange1)
```

```
##   League             Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA     J.j. O'brien  UTA       SF           1                 2
## 2    NBA Rakeem Christmas  IND       PF           1                 6
## 3    NBA Th Antetokounmpo  NYK       SF           3                 7
## 4    NBA       Sam Dekker  HOU       SF           3                 7
## 5    NBA      Coty Clarke  BOS     <NA>           4                 8
## 6    NBA     Jordan Adams  MEM       SG           2                15
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
## 1              0                   1          0               0              0
## 2              2                   2          0               0              0
## 3              3                   4          0               1              0
## 4              0                   0          0               0              0
## 5              2                   4          2               2              0
## 6              2                   6          0               1              3
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                   0                 0             0       0      0         0
## 2                   0                 1             1       0      0         0
## 3                   0                 0             1       0      0         0
## 4                   0                 0             1       0      1         0
## 5                   0                 0             1       0      0         1
## 6                   5                 0             2       3      3         2
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1      0             0                 0           0          0         0
## 2      0             1                 0           4          0         0
## 3      0             2                 0           6          0         0
## 4      0             0                 0           0          0         0
## 5      0             0                 0           6          0         0
## 6      0             2                 0           7          0         0
##   FlagrantFouls GamesStarted
## 1             0            0
## 2             0            0
## 3             0            0
## 4             0            0
## 5             0            0
## 6             0            0
```

desc()

```
arrange2<-arrange(NBA1516,desc(TotalMinutesPlayed),desc(GamesPlayed))
head(arrange2)
```

```
##   League            Name Team Position GamesPlayed TotalMinutesPlayed
## 1    NBA    James Harden  HOU       SG          82               3121
## 2    NBA  Gordon Hayward  UTA       SG          80               2889
## 3    NBA    Kemba Walker  CHA       PG          81               2885
## 4    NBA    Trevor Ariza  HOU       SF          81               2860
## 5    NBA Khris Middleton  MIL       SF          79               2855
## 6    NBA      Kyle Lowry  TOR       PG          77               2853
##   FieldGoalsMade FieldGoalsAttempted ThreesMade ThreesAttempted FreeThrowsMade
```

```
## 1            710               1617        236             656            720
## 2            521               1202        143             410            393
## 3            568               1332        182             490            371
## 4            357                858        185             497            126
## 5            507               1144        143             362            277
## 6            512               1198        212             546            398
##   FreeThrowsAttempted OffensiveRebounds TotalRebounds Assists Steals Turnovers
## 1                 837                63           502     612    138       374
## 2                 477                61           397     296     95       202
## 3                 438                56           358     421    127       171
## 4                 161                67           366     188    161       113
## 5                 312                45           301     331    131       180
## 6                 491                55           365     494    158       225
##   Blocks PersonalFouls Disqualifications TotalPoints Technicals Ejections
## 1     51           229                 1        2376          2         0
## 2     27           183                 0        1578          0         0
## 3     39           111                 0        1689          5         0
## 4     26           177                 0        1025          2         0
## 5     19           204                 1        1434          5         0
## 6     34           211                 1        1634          9         0
##   FlagrantFouls GamesStarted
## 1             0           82
## 2             0           80
## 3             0           81
## 4             0           81
## 5             0           79
## 6             0           77
```

group_by() summarise() arrange() Team Position

```
arrange3<-group_by(NBA1516,Team,Position)%>%
  summarise(nPlayer=n(),meanFieldGoalsMade=mean(FieldGoalsMade),
                meanFieldGoalsAttempted=mean(FieldGoalsAttempted)) %>%
  arrange(desc(meanFieldGoalsMade))
head(arrange3)
```

```
## Source: local data frame [6 x 5]
## Groups: Team [6]
##
##     Team Position nPlayer meanFieldGoalsMade meanFieldGoalsAttempted
##   <fctr>   <fctr>   <int>              <dbl>                   <dbl>
## 1    GSW       PG       2                504                     988
## 2    CLE       SF       2                440                     864
## 3    ORL       SG       1                425                     969
## 4    MIA        C       1                412                     681
## 5    OKL       PG       2                385                     861
## 6    ATL       PG       2                382                     884
```

7.3.7 rename()

=

```
rename1<-rename(NBA1516,Po=Position)
rename1[1:5,1:5]
```

```
##   League          Name Team Po GamesPlayed
```

```
## 1    NBA    Quincy Acy  SAC SF          59
## 2    NBA  Jordan Adams  MEM SG           2
## 3    NBA  Steven Adams  OKL  C          80
## 4    NBA Arron Afflalo  NYK SG          71
## 5    NBA Alexis Ajinca  NOR  C          59
```

7.3.8

- Introduction to dplyr
- DataCamp Data Manipulation in R with dplyr

Chapter 8

8.1

Ch **??** Graphical Quantitative :

-
- (patterns)
-
-

(Exploratory graphs) (Final graphs) :

-
-
-
-

(Final graphs) :

-
  –
- causality , mechanism , explanation , systematic structure
  –
- Multivariate
  – Multivariate
  –
-
  –
  –
- xy
  –
-
  –

R (Base) `lattice` `ggplot2` `ggplot2`

8.2 ggplot2

`ggplot2` (**?**) Dr. Leland Wilkinson Grammar of Graphics

"In brief, the grammar tells us that a statistical graphic is a `mapping` from data to `aesthetic` attributes (colour, shape, size) of `geometric` objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system"

-from `ggplot2` book

`ggplot2` Package Hadley Wickham R Package R packages GitHub

- **Aesthetic attributes**
- **Geometric objects**

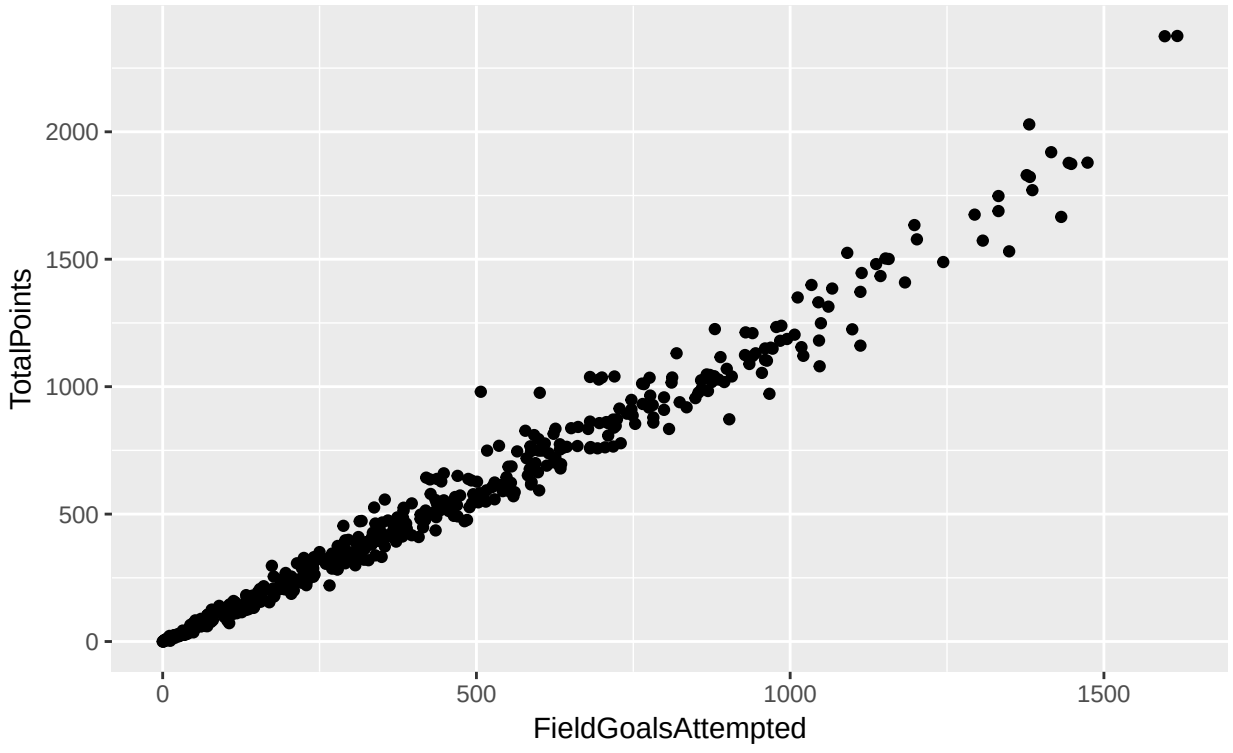
- **Facets** Faceting
- **Stats**
- **Scales** xy

ggplot2 `qplot()` ggplot2 Package plot() function

8.2.1 qplot()

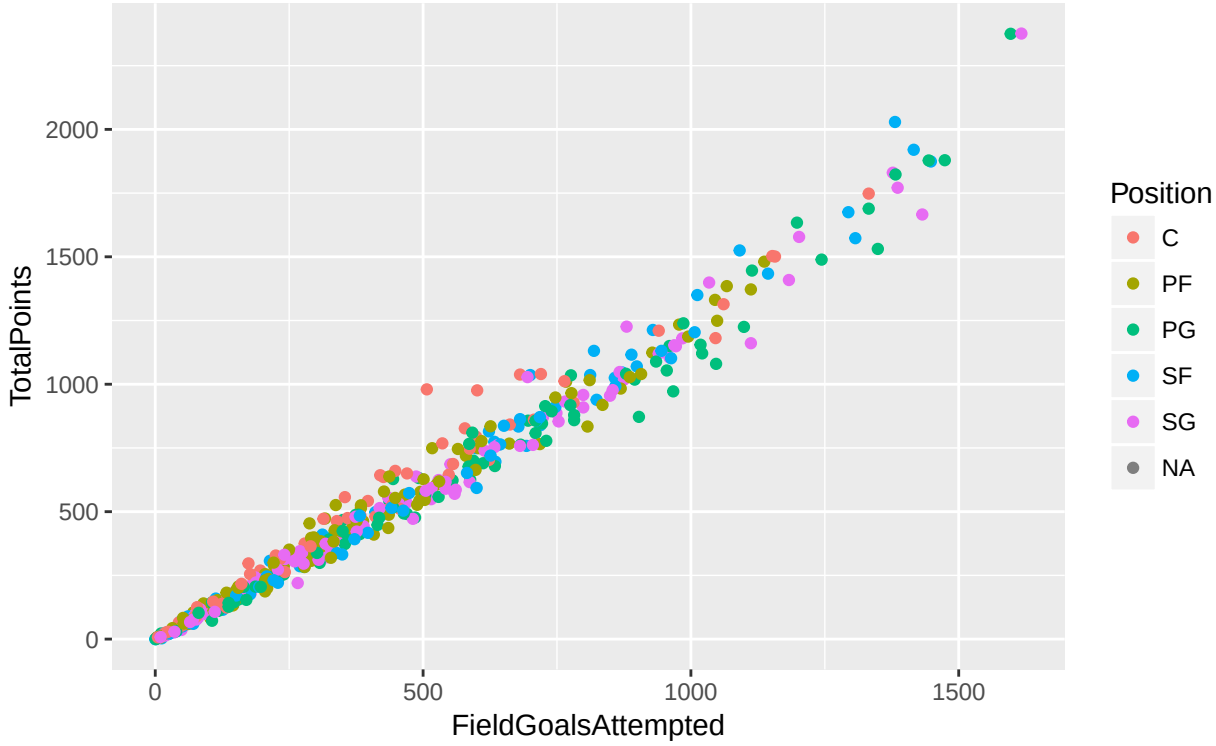
`qplot()` ggplot2 "Hello, world!" `qplot(x ,y ,data= )`

```
library(SportsAnalytics)
NBA1516<-fetch_NBAPlayerStatistics("15-16") ##
library(ggplot2) # ggplot2 package
qplot(FieldGoalsAttempted, TotalPoints, data = NBA1516)
```



Aesthetics `color=Position` Position

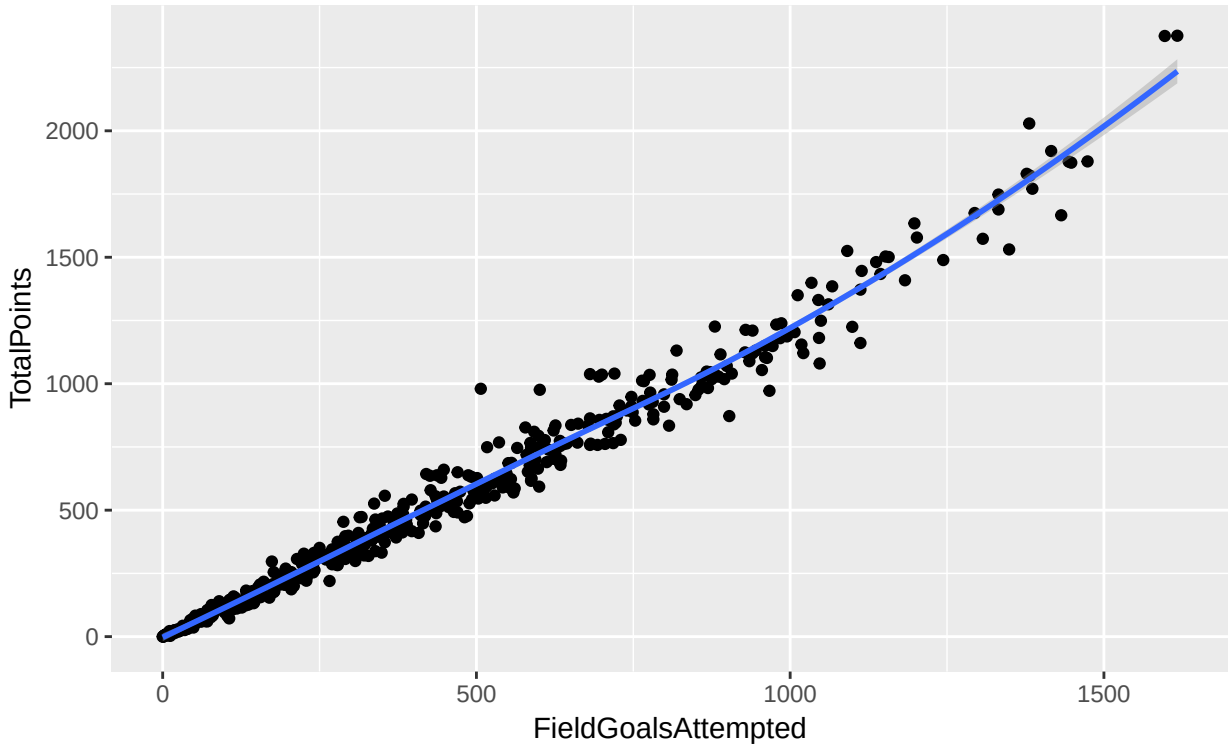
```
qplot(FieldGoalsAttempted, TotalPoints, data = NBA1516,color=Position)
```



Geometric `geom = c("point", "smooth")`

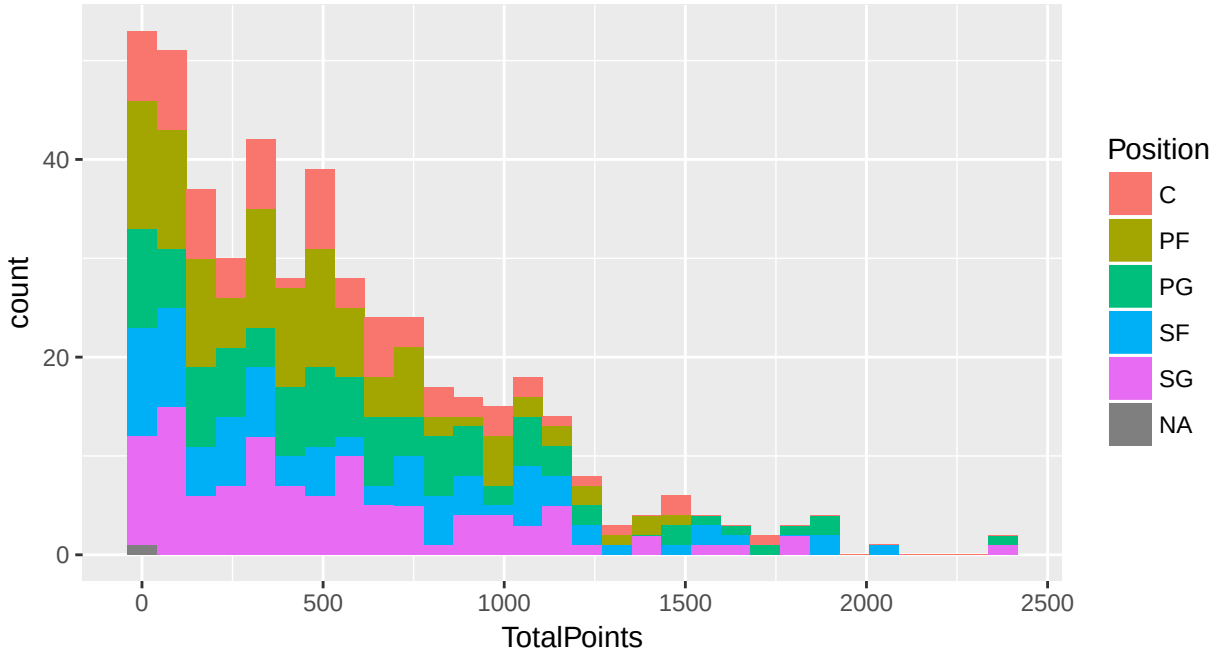
```
qplot(FieldGoalsAttempted, TotalPoints, data = NBA1516,
      geom = c("point", "smooth"))
```

```
## `geom_smooth()` using method = 'loess'
```



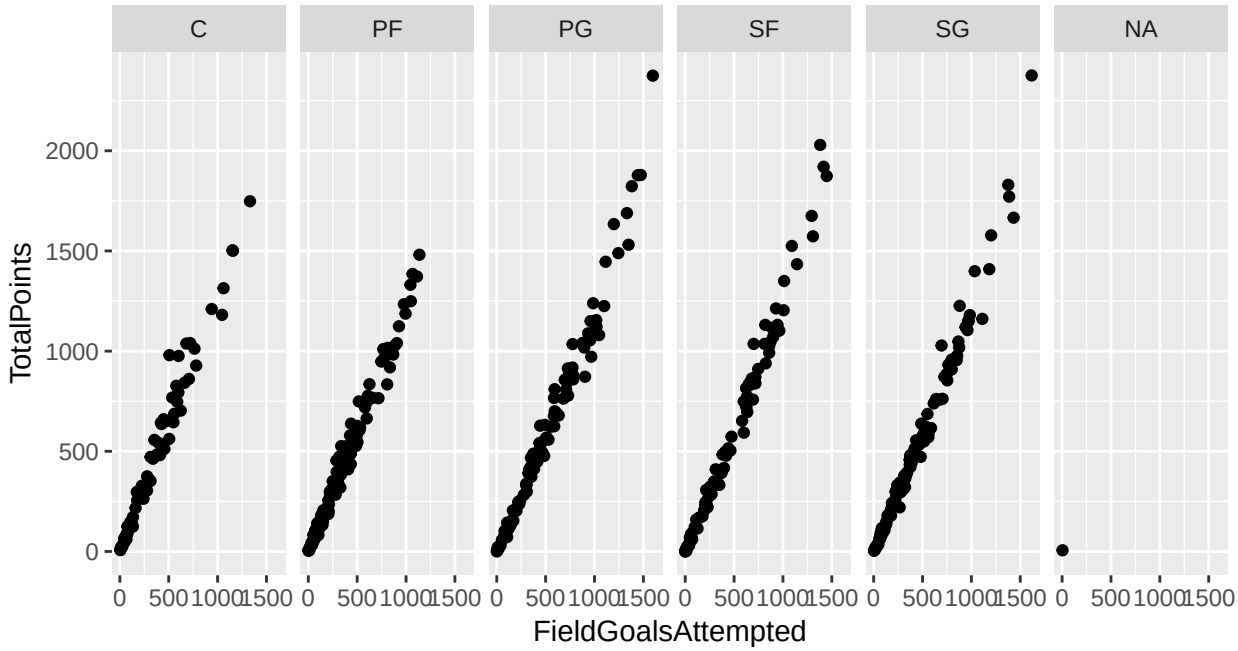
Histograms

```
# TotalPoints  / fill = Position    Position
qplot(TotalPoints, data = NBA1516, fill = Position)
```



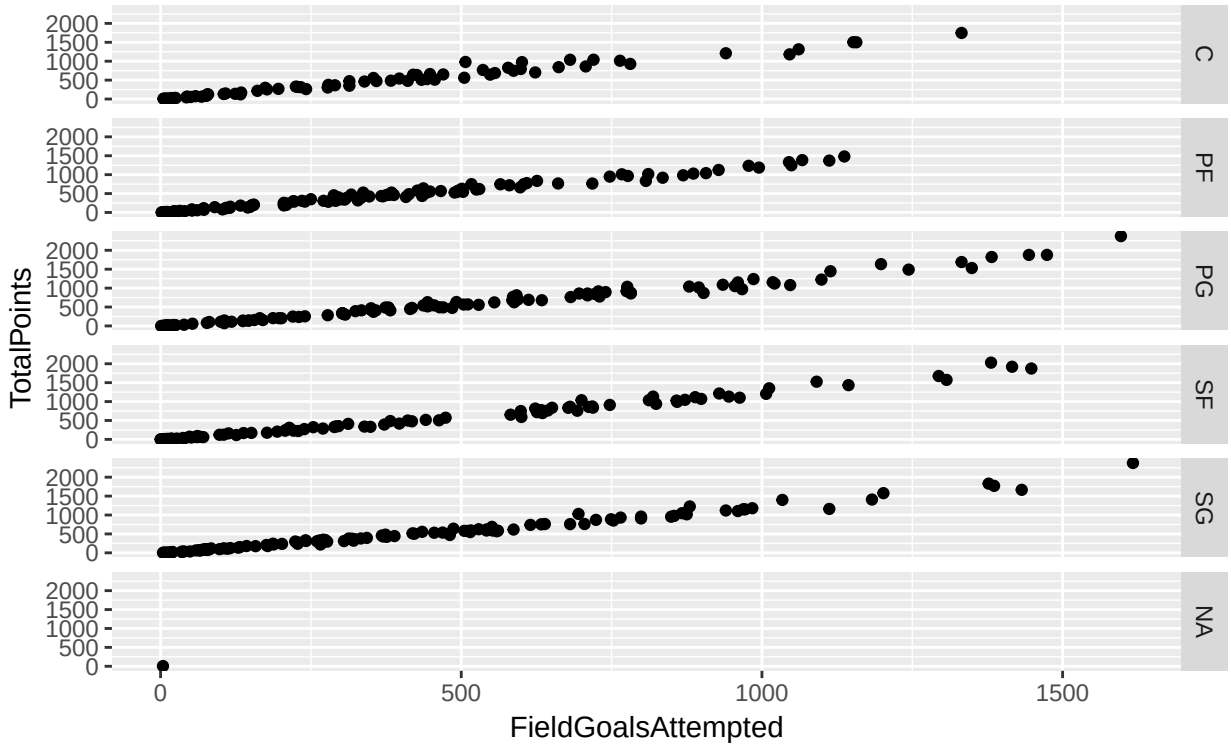
Facets facets = ~ Row Column

```
#facets = . ~ Position    Position  ( )
qplot(FieldGoalsAttempted, TotalPoints,
      data = NBA1516,
      facets = . ~ Position)
```



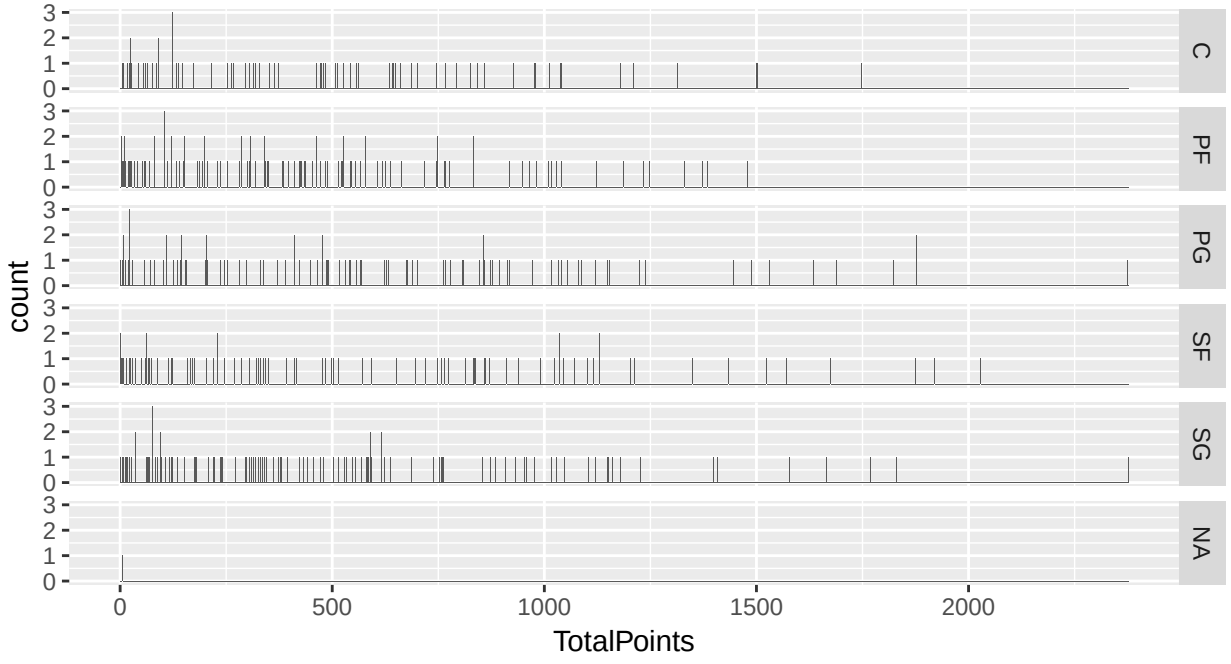

```
#facets = . ~ Position    Position  ( )
qplot(FieldGoalsAttempted, TotalPoints,
      data = NBA1516,
```

```
      facets = Position ~ .)
```

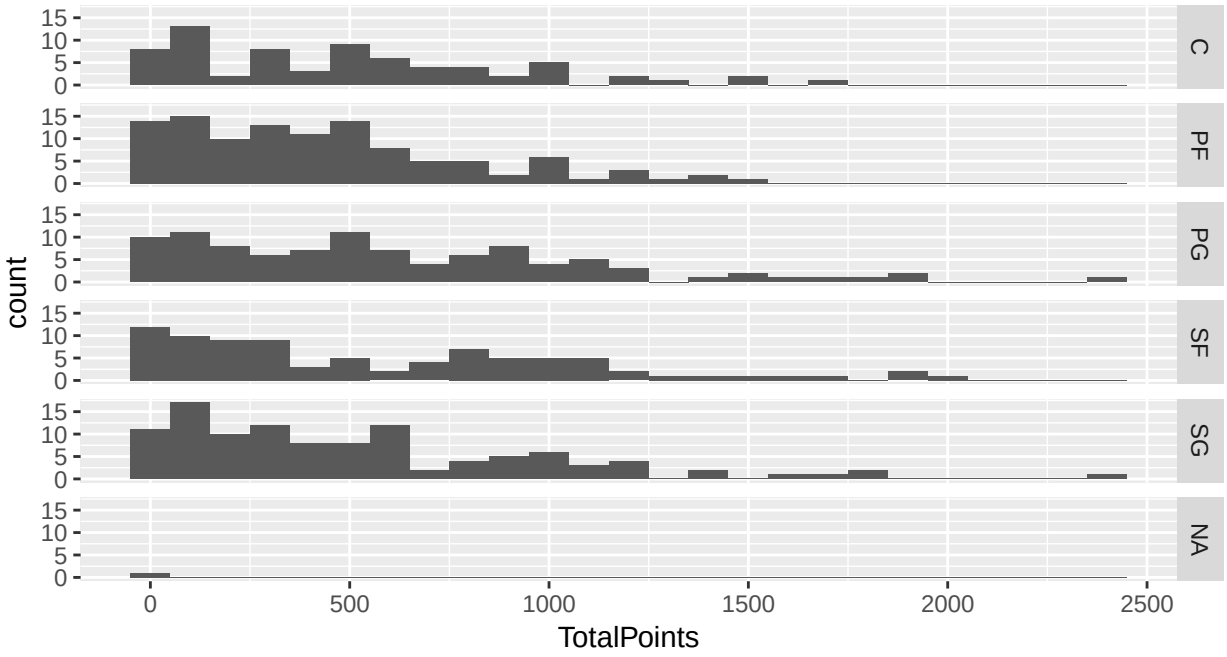


ggplot2 binwidth

```
#facets = . ~ Position    Position  ( )
##binwidth = 2  2
qplot(TotalPoints, data = NBA1516,
      facets = Position ~ ., binwidth = 2)
```



```
#facets = . ~ Position    Position   ( )
##binwidth = 100  100
qplot(TotalPoints, data = NBA1516,
      facets = Position ~ ., binwidth = 100)
```



qplot() ggplot()

8.2.2 ggplot()

ggplot2

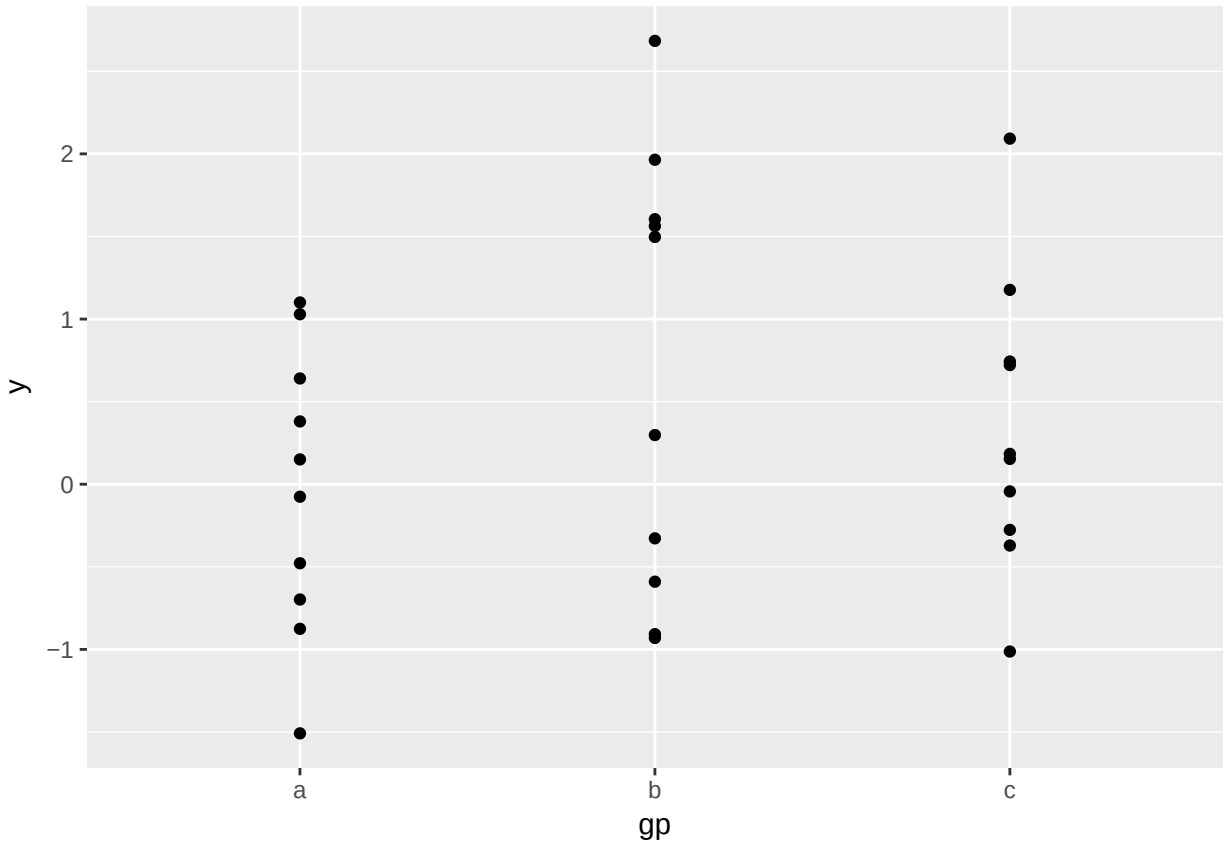
-
- **Aesthetic attributes** aes(x, y, ...)
- **Geometric objects** geom_point() geom_line() geom_point() geom_polygon() geom_errorbar()

```
library(ggplot2) ##   install.packages("ggplot2")
```

```
df <- data.frame(gp = factor(rep(letters[1:3], each = 10)),y = rnorm(30))
```

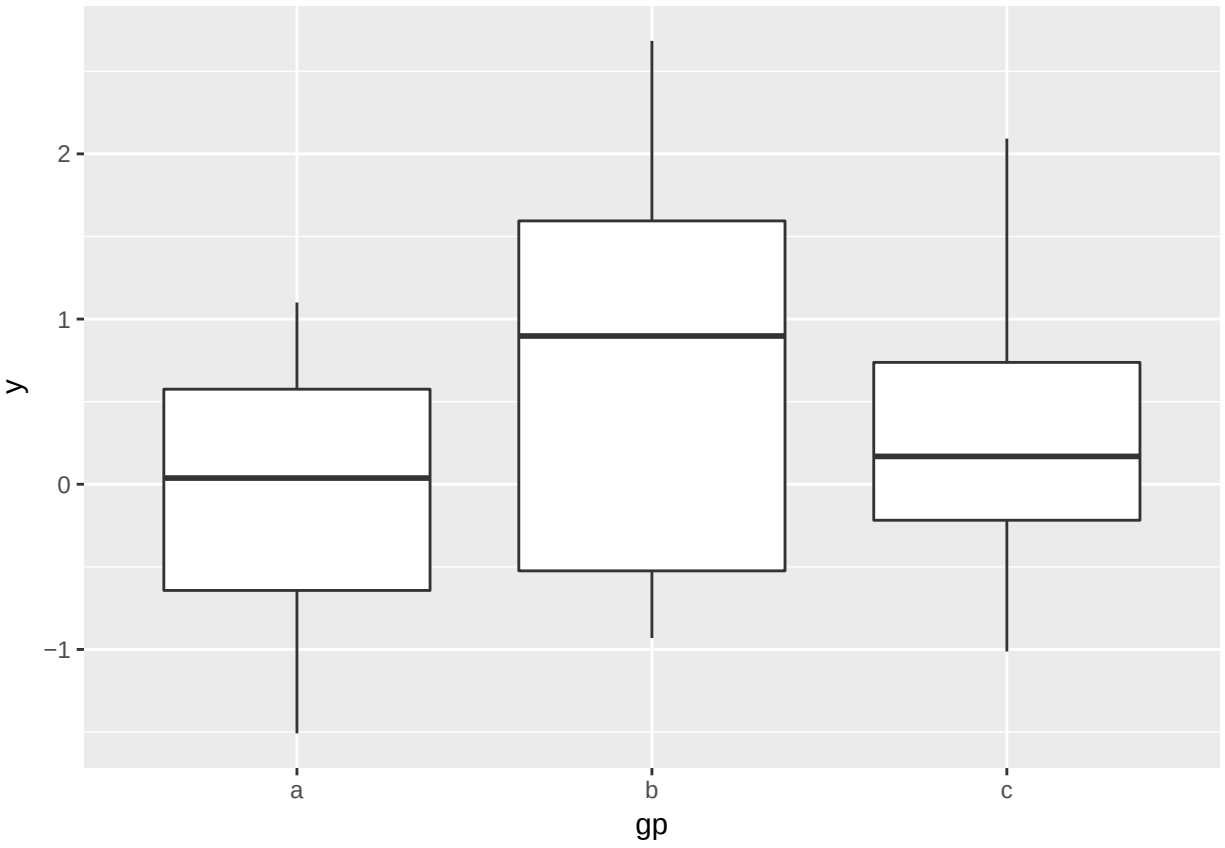
Aesthetic attributes Geometric objects

```
ggplot(df, aes(x = gp, y = y)) +geom_point()
```



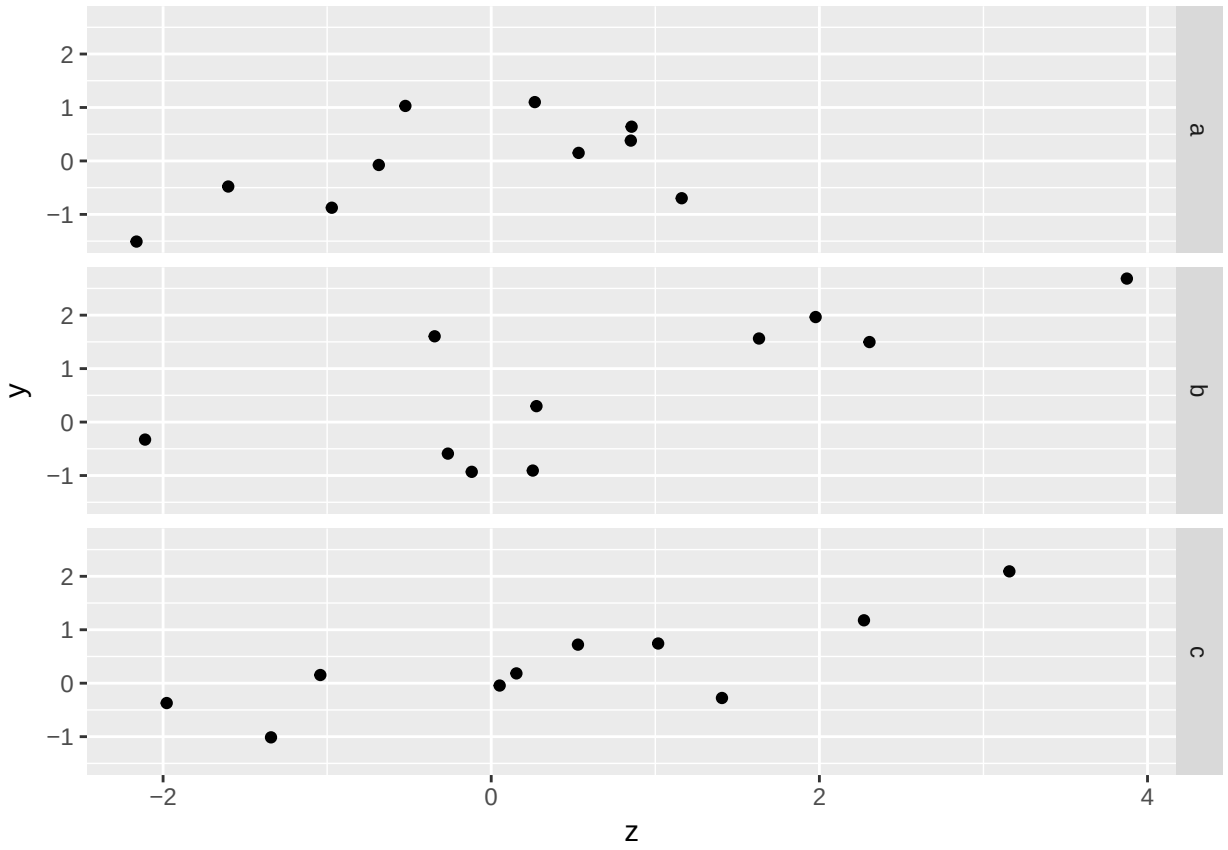
geom_boxpolt()

```
ggplot(df, aes(x = gp, y = y)) +geom_boxplot()
```

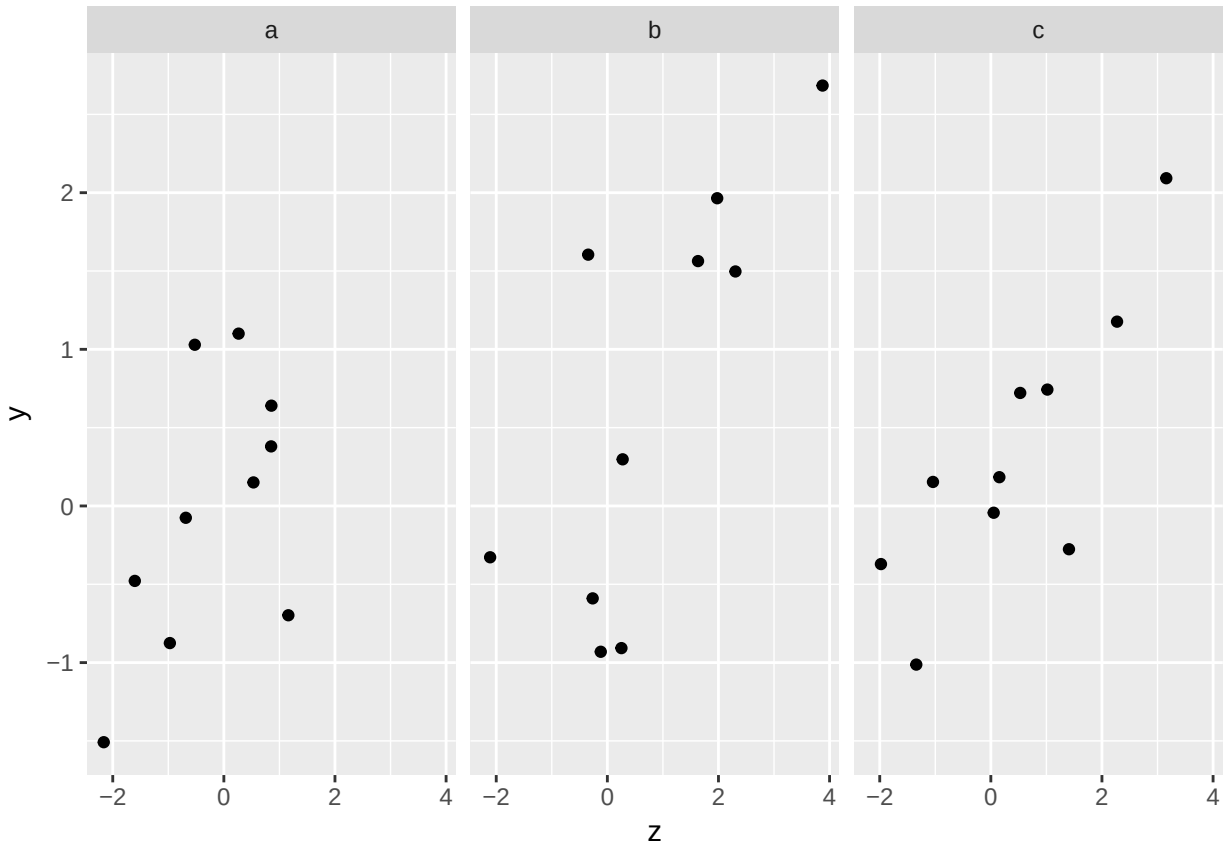


Faceting

```
df$z<-df$y+rnorm(30)
ggplot(df, aes(x = z, y = y)) +geom_point()+facet_grid(gp~.)
```



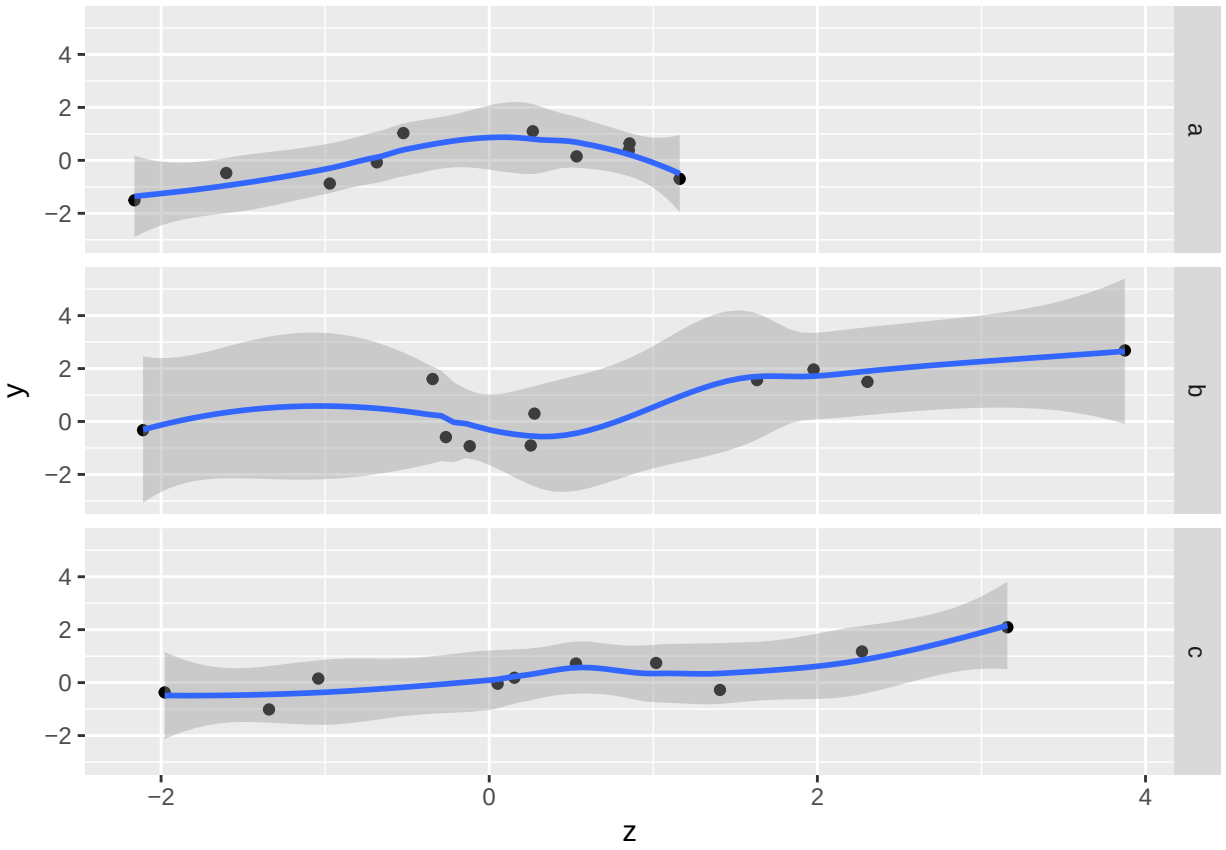

```
ggplot(df, aes(x = z, y = y)) +geom_point()+facet_grid(.~gp)
```



geom_smooth() xy

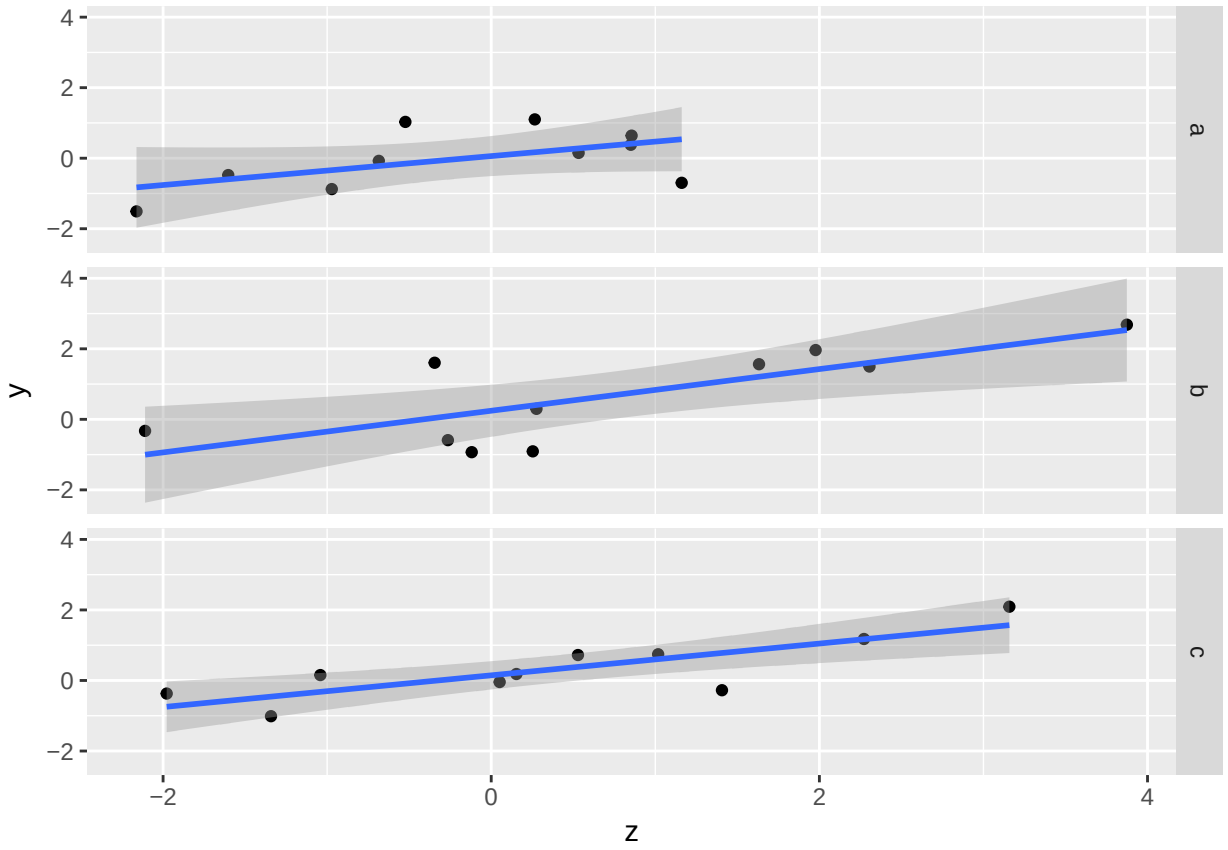
```
ggplot(df, aes(x = z, y = y)) +geom_point()+facet_grid(gp~.)+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess'
```



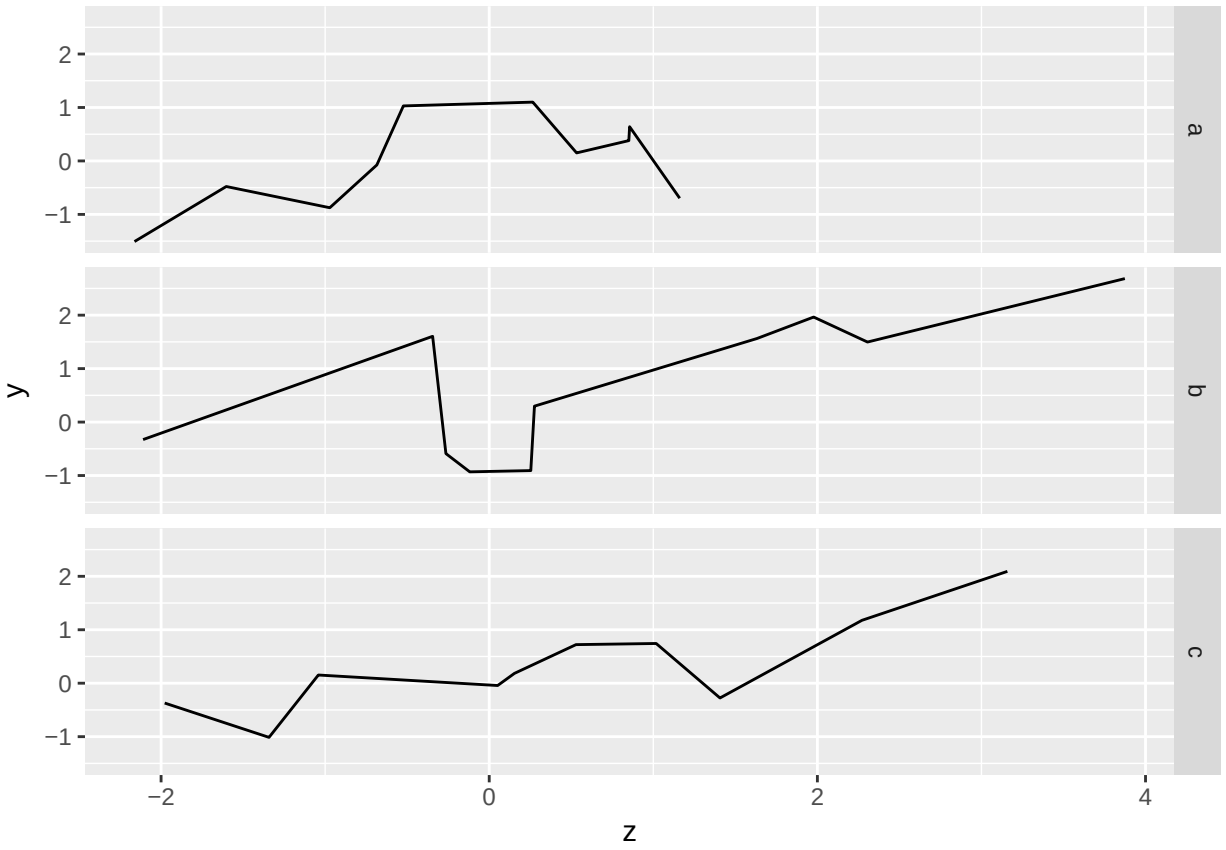
`geom_smooth()` xy linear regresion

```
ggplot(df, aes(x = z, y = y)) +geom_point()+facet_grid(gp~.)+geom_smooth(method='lm')
```

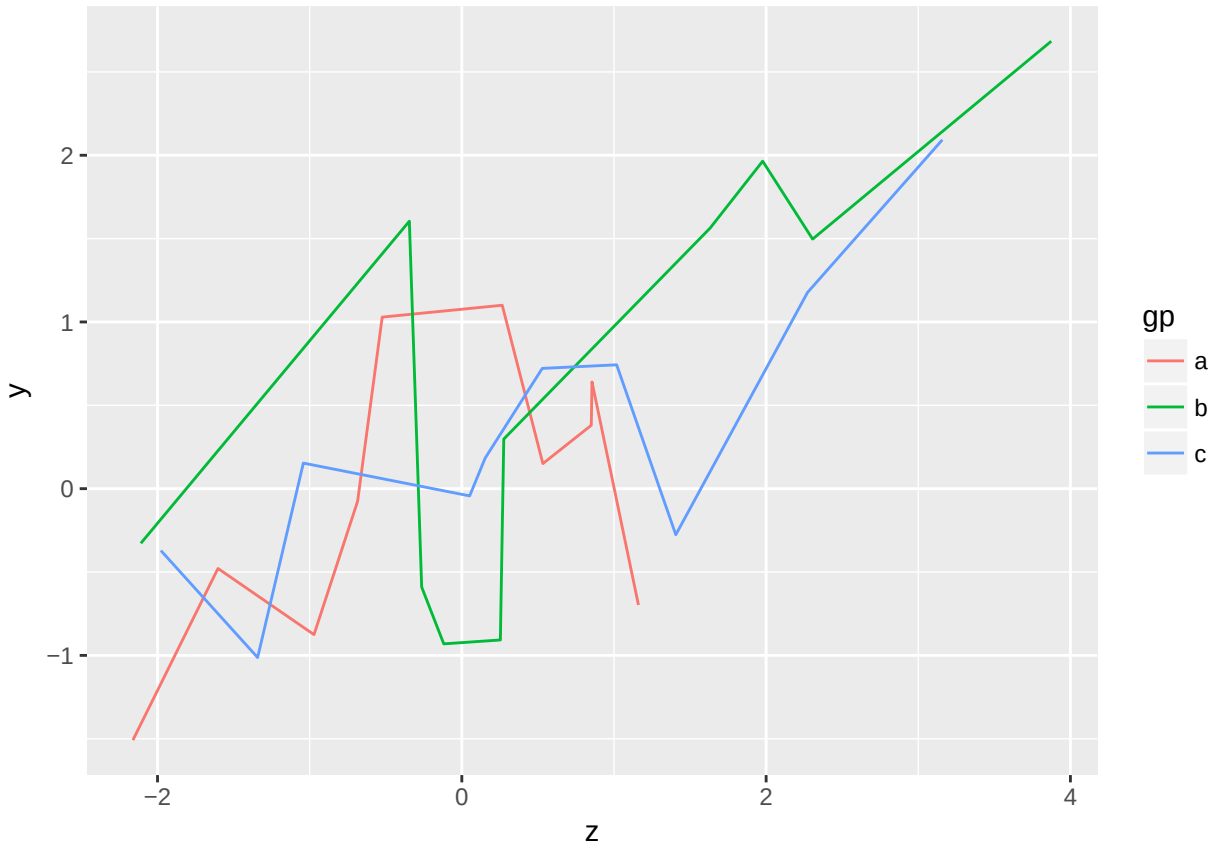


geom_line()

```
ggplot(df, aes(x = z, y = y)) +geom_line()+facet_grid(gp~.)
```




```
    aes(color='group name')
ggplot(df, aes(x = z, y = y, color=gp)) +geom_line()
```



ggplot2 a/b/c a/b/c 1,2,3 factor() ggplot factor

ggplot2

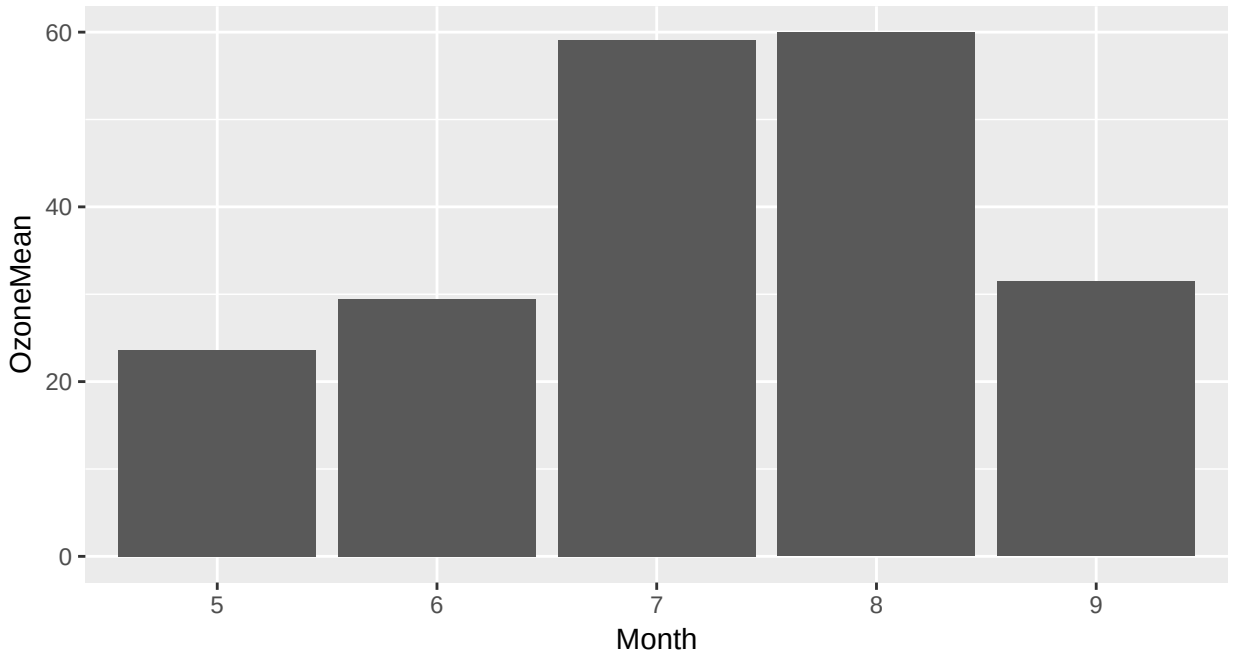
- xlab(), ylab(), labs(x=,y=), ggtitle()
- geom_*()
- theme()
- theme_gray():
- theme_bw():
-
- ggthemes packages Website
- xkcd packages Website

(Error bar) bar chart line chart

- Standard deviation (SD)
- Standard error (SE)
- Confidence interval (CI)

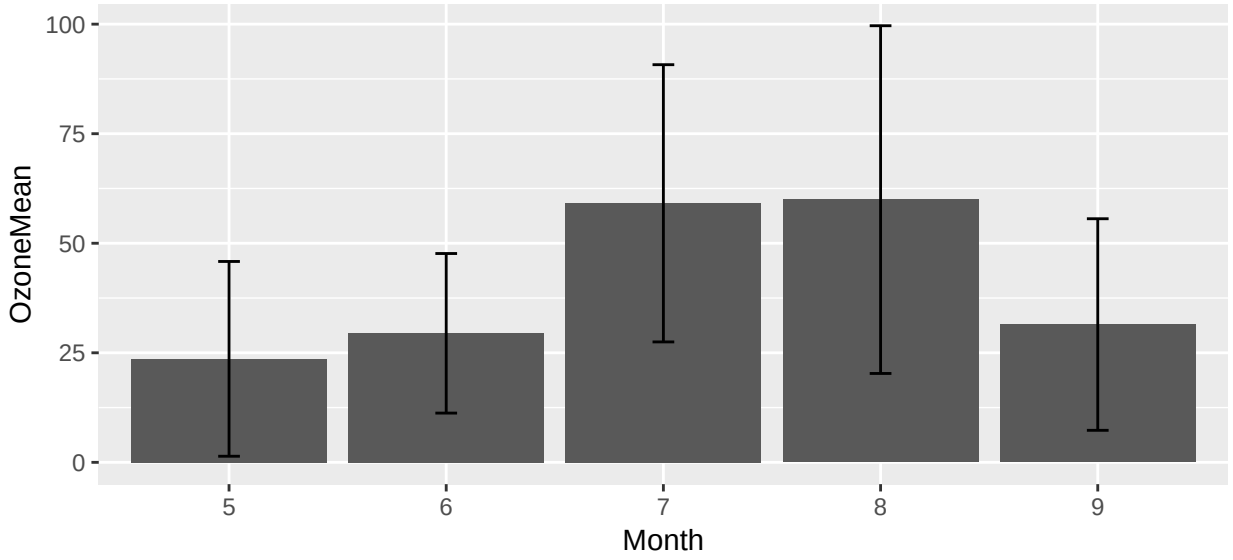
bar chart ggplot2 bar chart **Geometric objects** geom_bar

```
library(datasets)
library(data.table)
airquality$Month<-as.factor(airquality$Month) # Month
airquality.mean<-data.table(airquality)[,.(OzoneMean=mean(Ozone,na.rm = T)),by=Month] #  Ozone
ggplot()+geom_bar(data=airquality.mean,aes(x=Month,y=OzoneMean),
                  stat = "identity") #stat = "identity"
```



ggplot2 geom_errorbar()

```
library(datasets)
library(data.table)
airquality$Month<-as.factor(airquality$Month) # Month
airquality.stat<-data.table(airquality)[,.(OzoneMean=mean(Ozone,na.rm = T),OzoneSD=sd(Ozone,na.rm = T))
ggplot(data=airquality.stat)+ # airquality.eb
    geom_bar(aes(x=Month,y=OzoneMean),stat = "identity")+
    geom_errorbar( #ymin , ymax
        aes(x=Month,ymin=OzoneMean-OzoneSD,ymax=OzoneMean+OzoneSD), width=.1)
```



8.3 ggplot2+

8.3.1 Choropleth map

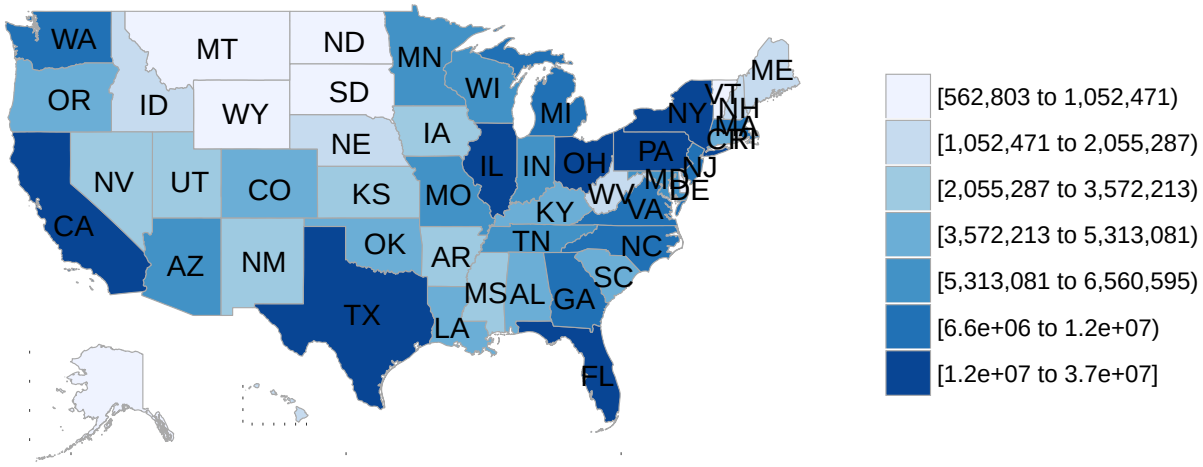
Choropleth map R choroplethr(?) package choroplethr package ggplot2 package choroplethrMaps package

```
install.packages(c("choroplethr","choroplethrMaps")) ##
```

```
library(choroplethr)
```

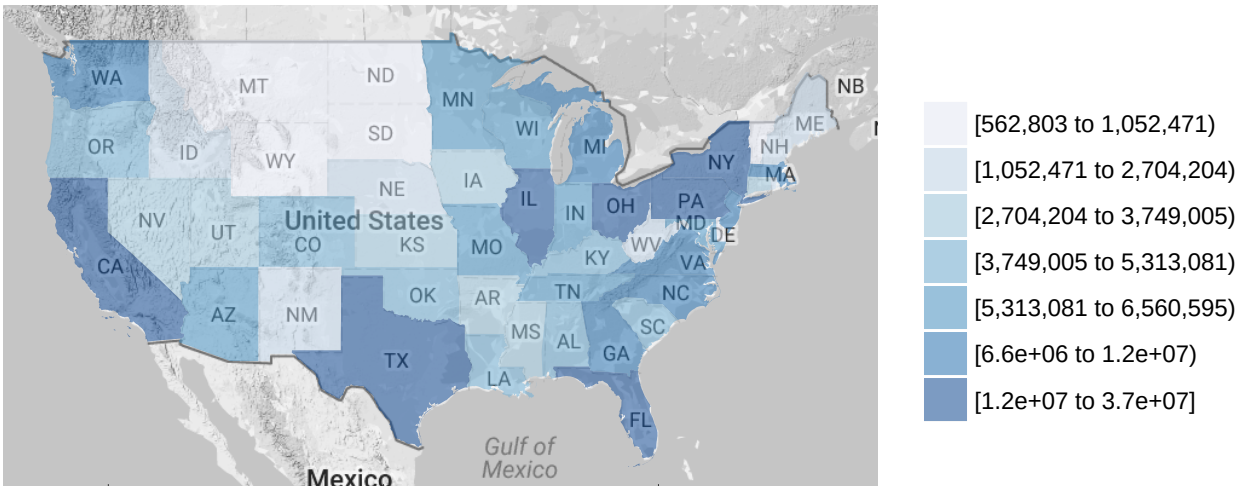
choroplethr(?) package state_choropleth()

```
data(df_pop_state) #
state_choropleth(df_pop_state) #
```



reference_map = TRUE google

```
data(continental_us_states)
state_choropleth(df_pop_state,reference_map = TRUE,
                 zoom= continental_us_states) #
```



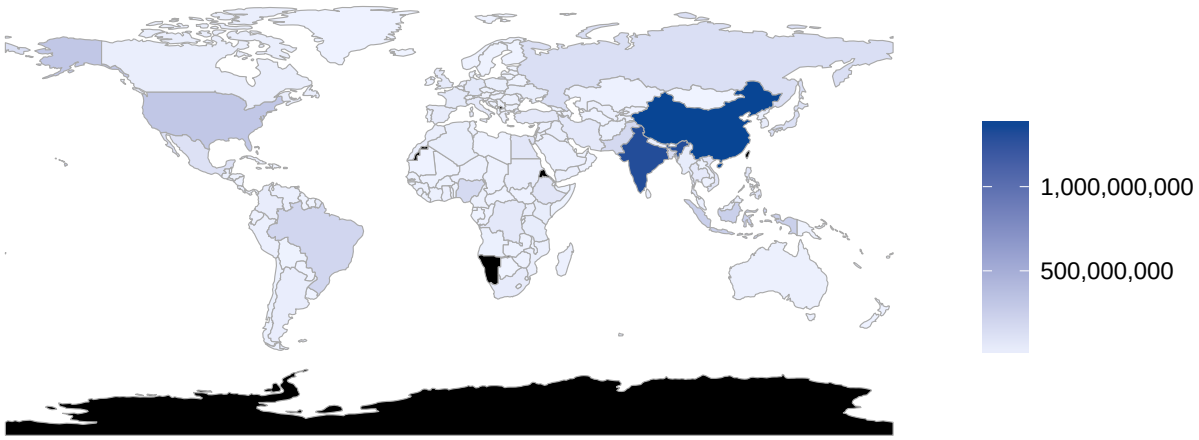
choroplethr package WDI: World Development Indicators SP.POP.TOTL World Development Indicators

WDI WDI(?) package

```
install.packages("WDI") ##
```

```
library(WDI)
choroplethr_wdi(code="SP.POP.TOTL", year=2014,
                title="2015 Population", num_colors=1)
```

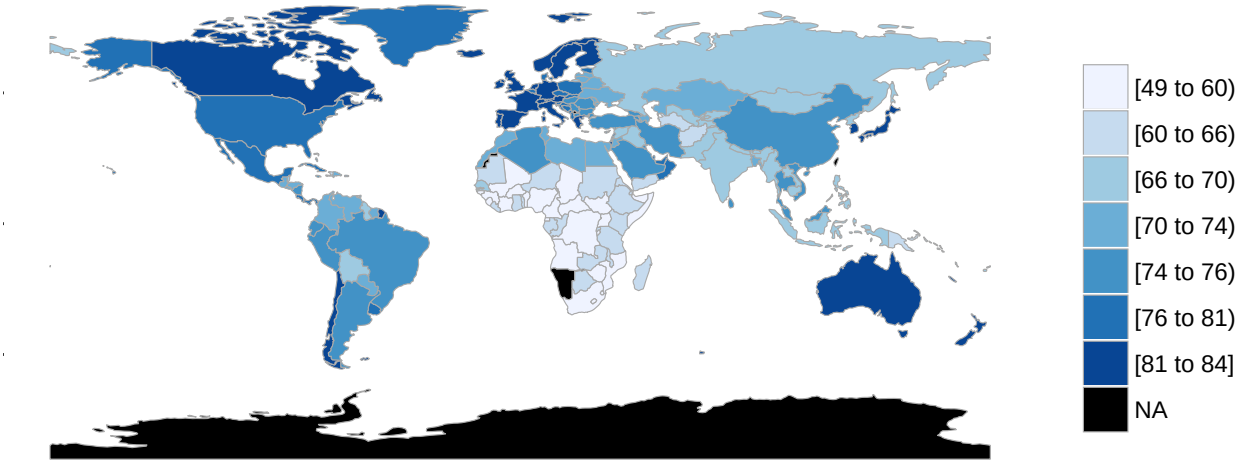
2015 Population



WDI SP.DYN.LE00.IN World Development Indicators

```
choroplethr_wdi(code="SP.DYN.LE00.IN", year=2014,
                title="2014 Life Expectancy")
```

2014 Life Expectancy



zoom country.regions

```
choroplethr_wdi(code="SP.POP.TOTL", year=2015,
                title="2015 Life Expectancy",
                zoom=c('taiwan','japan','south korea','philippines'))
```

8.3.2 ggmap()

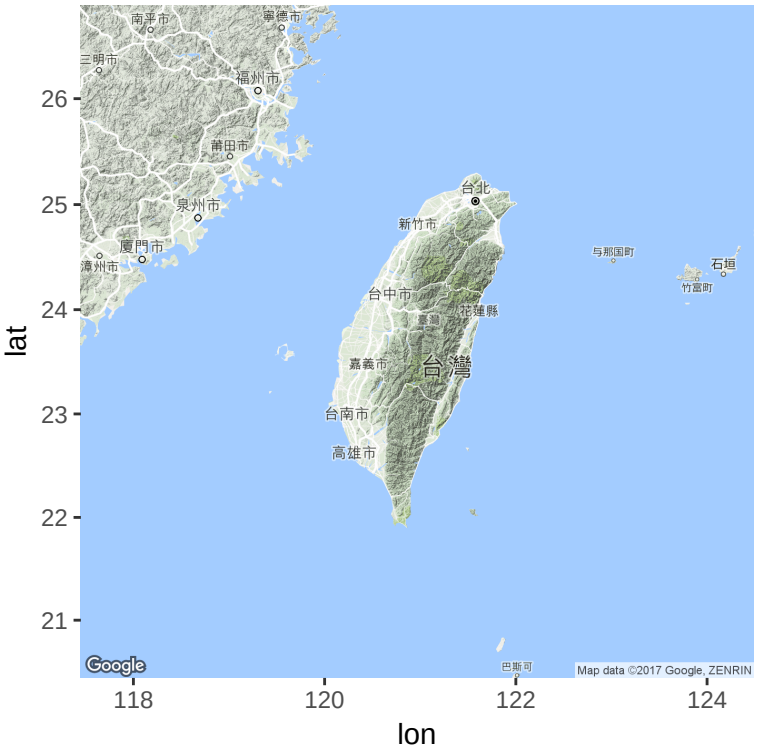
ggmap(?) package google map ggplot2

```
install.packages("ggmap", type = "source") ##
```

ggmap(?) package get_map() google map ggmap() get_map()

- location
- zoom
- language

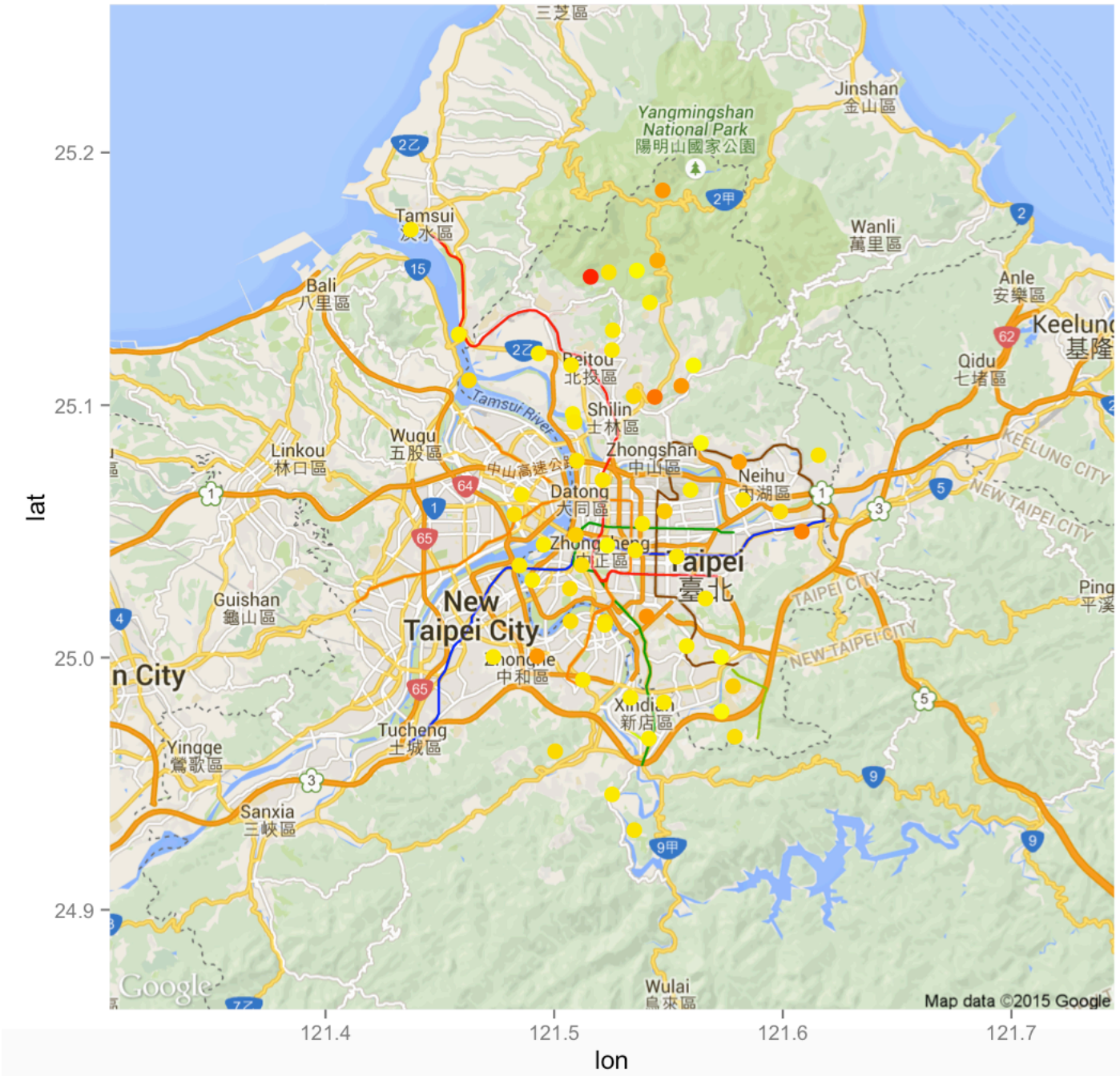
```
library(ggmap)
twmap <- get_map(location = 'Taiwan', zoom = 7,language = "zh-TW")
ggmap(twmap)
```



ggmap package Ch ?? Open data API http://data.taipei/opendata/datalist/apiAcc
7c56-42e0-8068-39242b8ec927

```
##
library(jsonlite)
library(RCurl)
WaterData<-fromJSON(getURL("http://data.taipei/opendata/datalist/apiAccess?scope=resourceAquire&rid=190
WaterDataFrame<-WaterData$result$results
WaterDataFrame$longitude<-as.numeric(WaterDataFrame$longitude)
WaterDataFrame$latitude<-as.numeric(WaterDataFrame$latitude)
WaterDataFrame$qua_cntu<-as.numeric(WaterDataFrame$qua_cntu)
## ggmap
library(ggmap)
TaipeiMap = get_map(location = c(121.43,24.93,121.62,25.19),
                    zoom = 11, maptype = 'roadmap')
TaipeiMap0 = ggmap(TaipeiMap)+
    geom_point(data=WaterDataFrame[WaterDataFrame$qua_cntu>=0,],
               aes(x=longitude, y=latitude,color=qua_cntu,size=3.5))+
    scale_color_continuous(low = "yellow",high = "red")+
    guides(size=FALSE)
```

```
TaipeiMap0
```



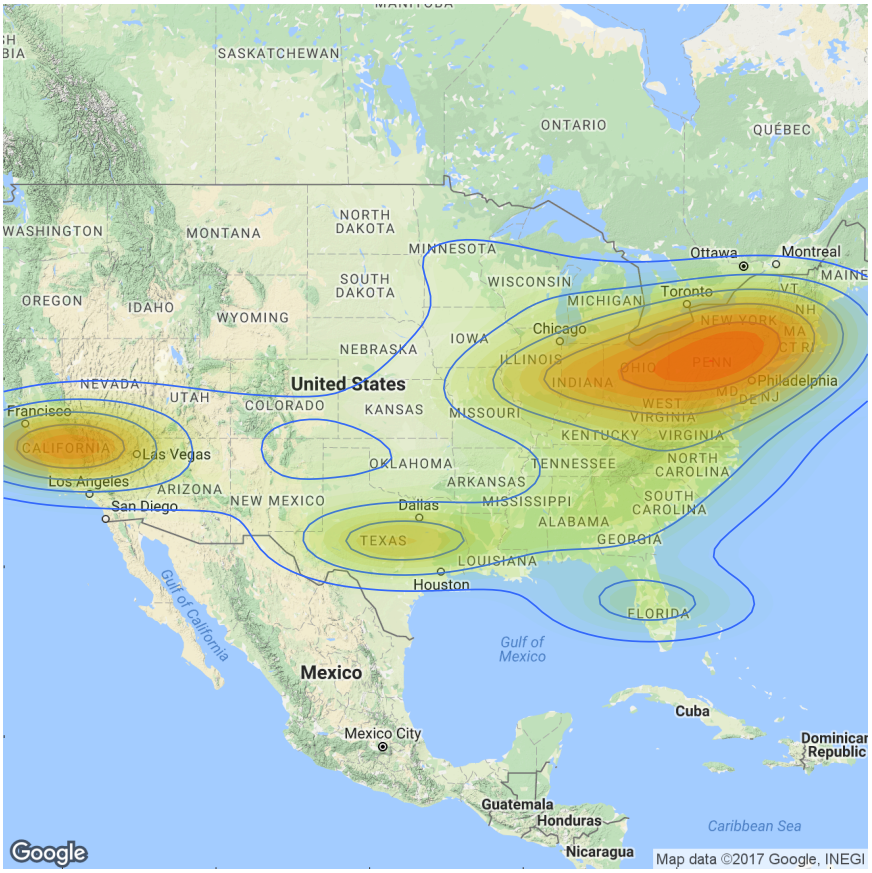
`ggmap` `maptype`

- terrain
- terrain-background
- satellite
- roadmap
- hybrid (google maps)
- watercolor
- toner (stamen maps)

extent

```
library(ggmap)
TaipeiMap = get_map(location = c(121.43,24.93,121.62,25.19),
                    zoom = 14, maptype = 'roadmap')
ggmap(TaipeiMap,extent = 'device') #extent = 'device'
```

8.3.3 Density Map



Density Map

ggplot2 + ggmap

```
#
StateCenter<-data.frame(
    region=tolower(state.name),lon=state.center$x,lat=state.center$y)
head(StateCenter,1)
```

```
##    region lon lat
## 1 alabama -87  33
```

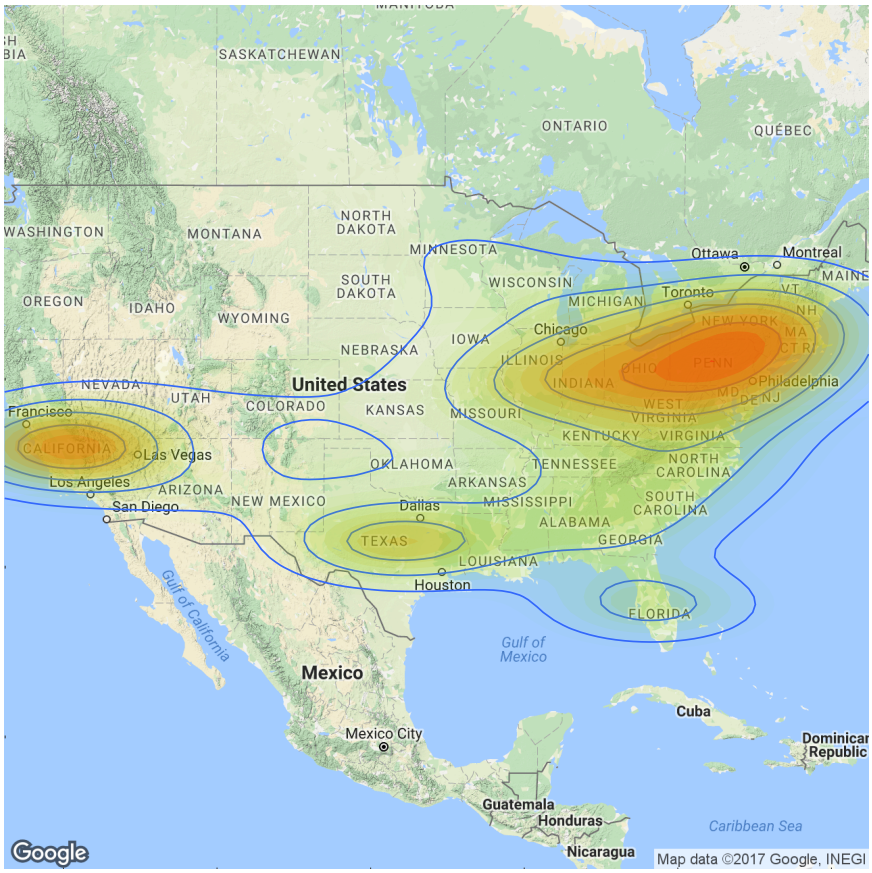
```
#
StatePop<-merge(df_pop_state,StateCenter,by="region")
head(StatePop,1)
```

```
##    region   value lon lat
## 1 alabama 4777326 -87  33
```

```
#
PopPoint<-NULL
for(i in 1:nrow(StatePop)){
    # 100  1
    for(j in 1:(StatePop[i,"value"]/1000000)){
        PopPoint<-rbind(PopPoint,StatePop[i,])
    }
}
head(PopPoint,3)
```

```
##    region   value lon lat
## 1 alabama 4777326 -87  33
## 2 alabama 4777326 -87  33
## 3 alabama 4777326 -87  33
```

```
USMap <- get_map(location = "United States", zoom = 4)
densityMap<-ggmap(USMap, extent = "device") +
    geom_density2d(data = PopPoint, aes(x = lon, y = lat), size = 0.3) +
    stat_density2d(data = PopPoint,
            aes(x = lon, y = lat, fill = ..level.., alpha = ..level..),
                size = 0.01, bins = 16, geom = "polygon") +
    scale_fill_gradient(low = "green", high = "red", guide = FALSE) +
    scale_alpha(range = c(0, 0.3), guide = FALSE)
densityMap
```



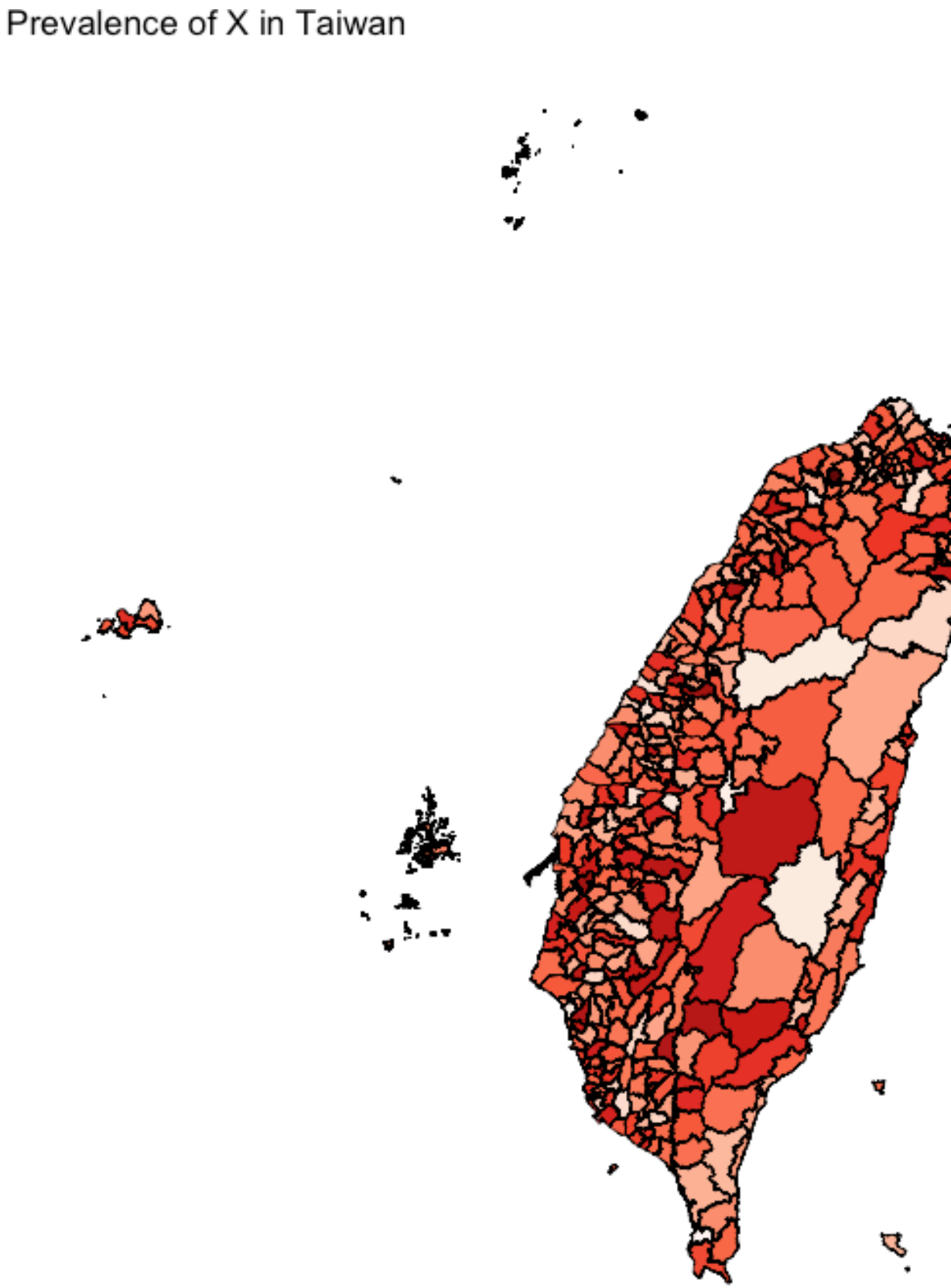
8.3.4

- ggmap package source code
- ggmap cheat sheet
- ggmap doc

8.4 Taiwan

Open Data / `shapefile` `.shp` `shapefile` `ggplot2`

-
- `rgdal`, `rgeos`,`maptools` package shapefile
- `ggplot2` & `RColorBrewer`


Prevalence of X in Taiwan