APCC 2016 Statistical Downscaling Training Program
Quick R start

August 2016

Dr. Moosup Kim

APCC
APEC CLIMATE CENTER

About Lecturer

- Name: Moosup Kim
- Research Fellow in APEC Climate Center
- Ph. D. in Statistics
- Mission: Statistical Downscaling
- Email: moosupkim@apcc21.org

Contents

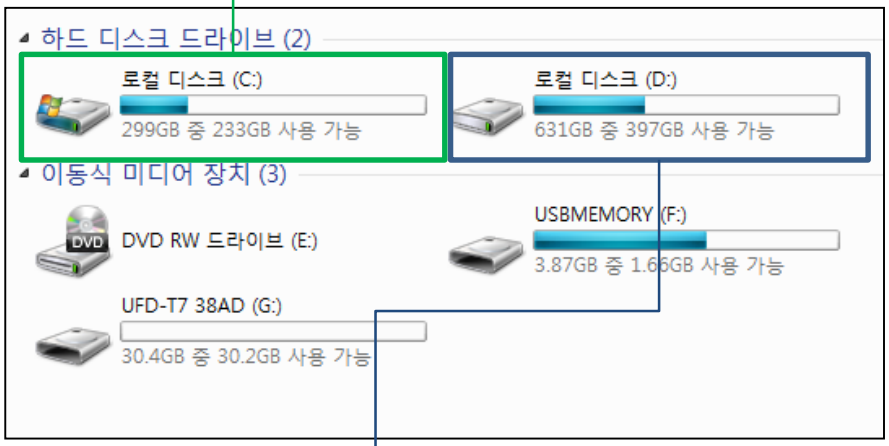
- Preparation
- Variables and Operations
- Data Analysis using R
- Reading your Observation Data

Contents

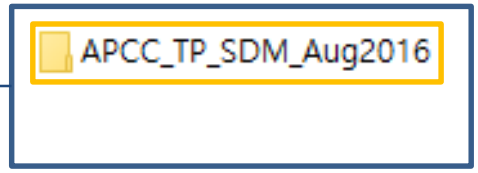
- Preparation
- Variables and Operations
- Data Analysis using R
- Reading your Observation Data

Preparation

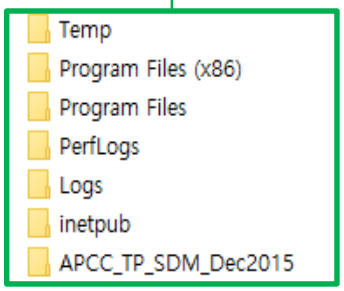
- Copy **D:\1Kim\APCC_TP_SDM_Aug2016** into **C:**.



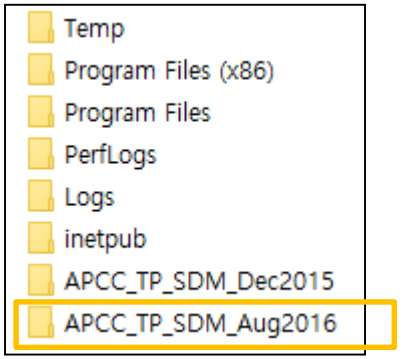
1. Open the D drive!



2. Copy folder!



3. Open C drive

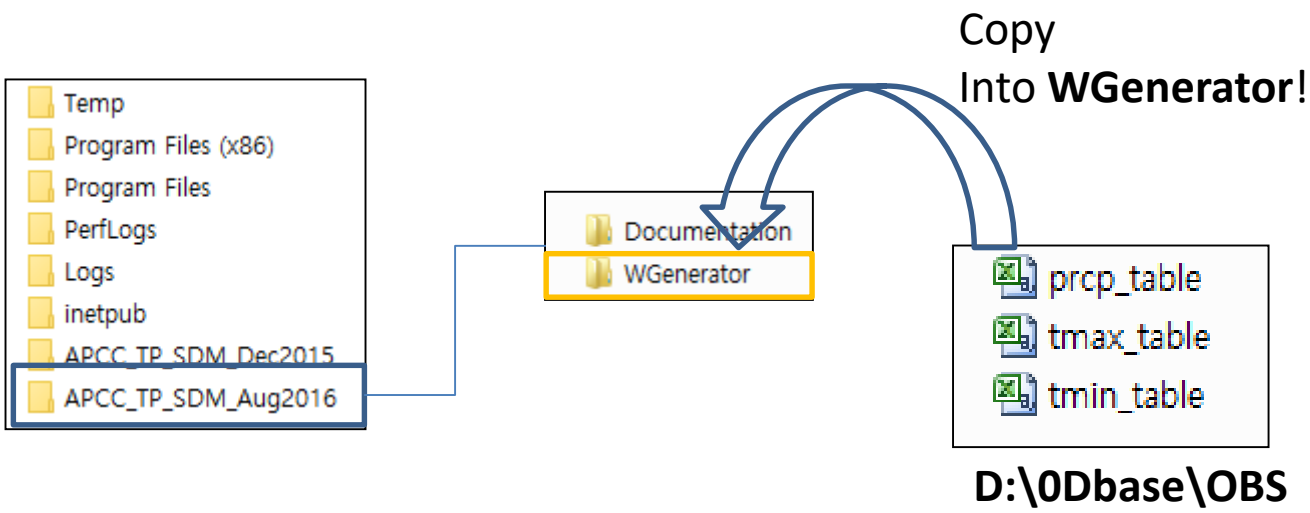


4. Paste the copied folder into C drive!

*Display and system language depend on your computer system

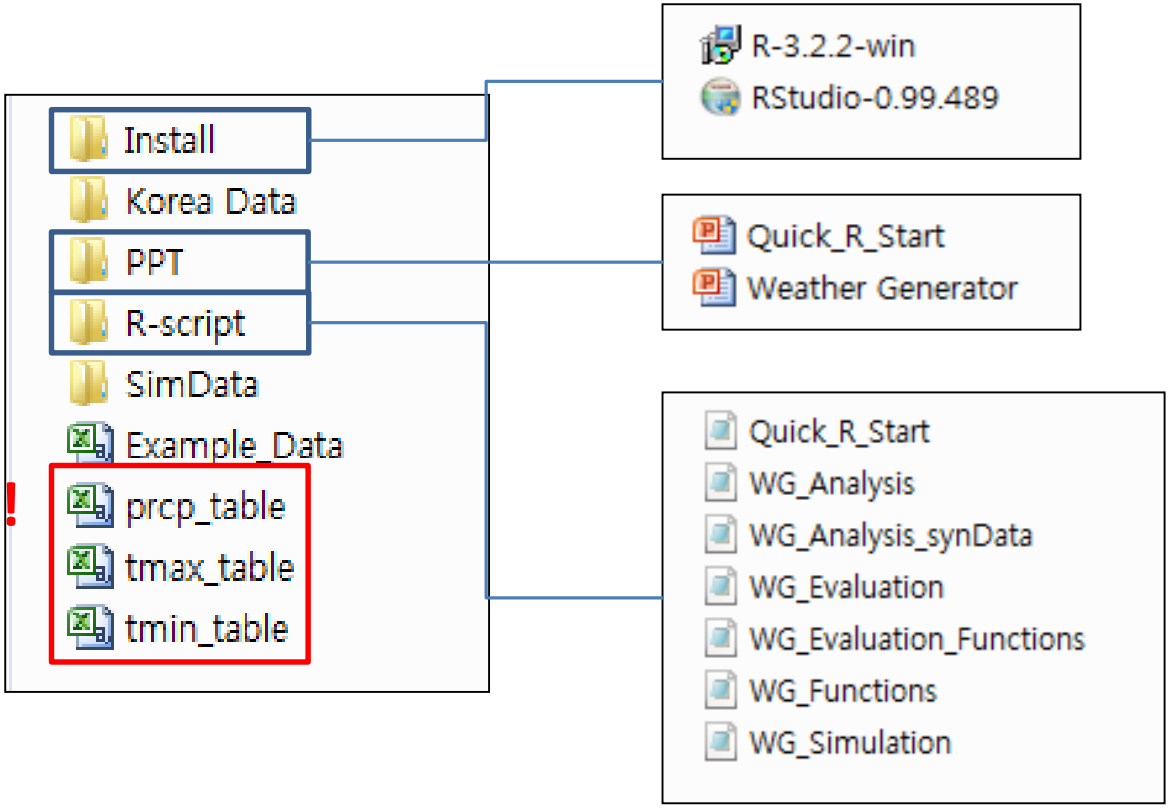
Preparation

- Copy your observation data into **C:\APCC_TP_SDM_Aug2016\WGenerator**
- At **D:\0Dbase\OBS**, **prcp_table, tmax_table, tmin_table**.



Preparation

- **APCC_TP_SDM_Aug2016/WGenerator**



Your data sets.
Check the file name!
(all small letters)
prcp_table
tmax_table
tmin_table

Preparation

- Check your data files!
  - Data file format is CSV.
  - First line shows names: YEAR, MONTH, DAY, rests are site names.

All capital letters!

Site names check!

	A	B	C	D	E	F	G	H	I
1	YEAR	MONTH	DAY	Andong	Bongwhoa	Tawbaekk			
2	1988	1	1	0	0	0			
3	1988	1	2	0	0	0			
4	1988	1	3	0	0	0			
5	1988	1	4	1	0.3	0.2			
6	1988	1	5	0	0	0			
7	1988	1	6	0	0	0			
8	1988	1	7	0	0	0			
9	1988	1	8	0.2	0.8	1.6			
10	1988	1	9	0	0	0			
11	1988	1	10	0	0	0			
12	1988	1	11	0	0	0			
13	1988	1	12	0	0	0			
14	1988	1	13	0	0	0			
15	1988	1	14	0.2	0	0			

Display depends on your data.

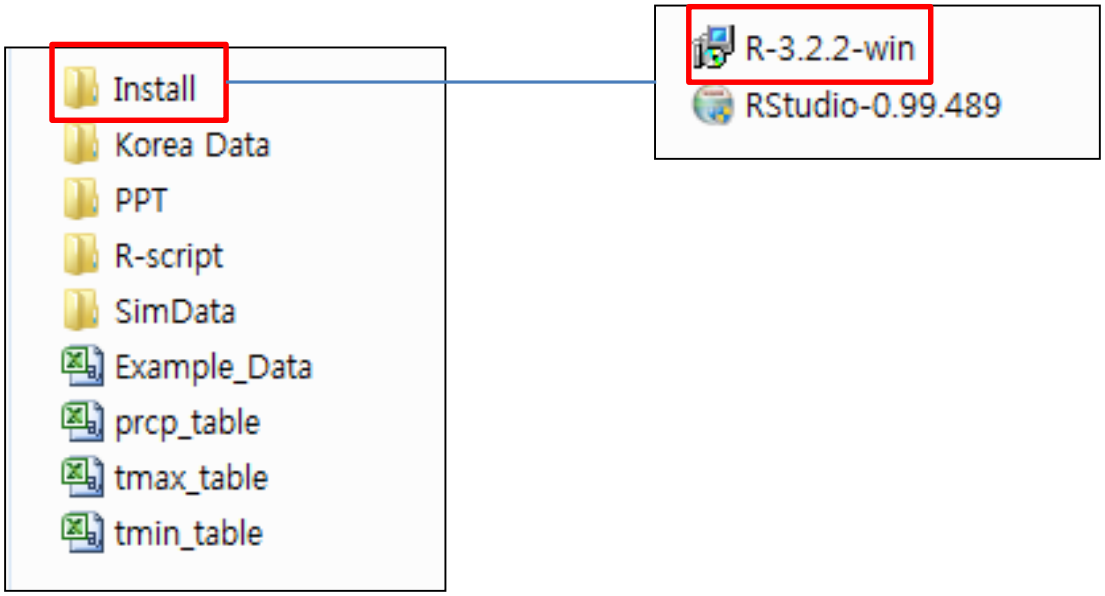
Check spelling and capital/small letters of the names!

Data have to be sorted in date (YEAR,MONTH,DAY)!!

No missing in YEAR, MONTH,DAY!!

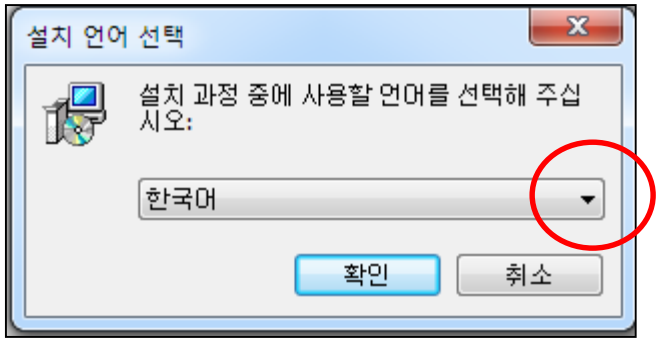
Preparation

- Install R
  - Open **Install** folder
  - Double click **R-3.2.2-win** (Start to install)



Preparation

- Install R
  - Select your language.

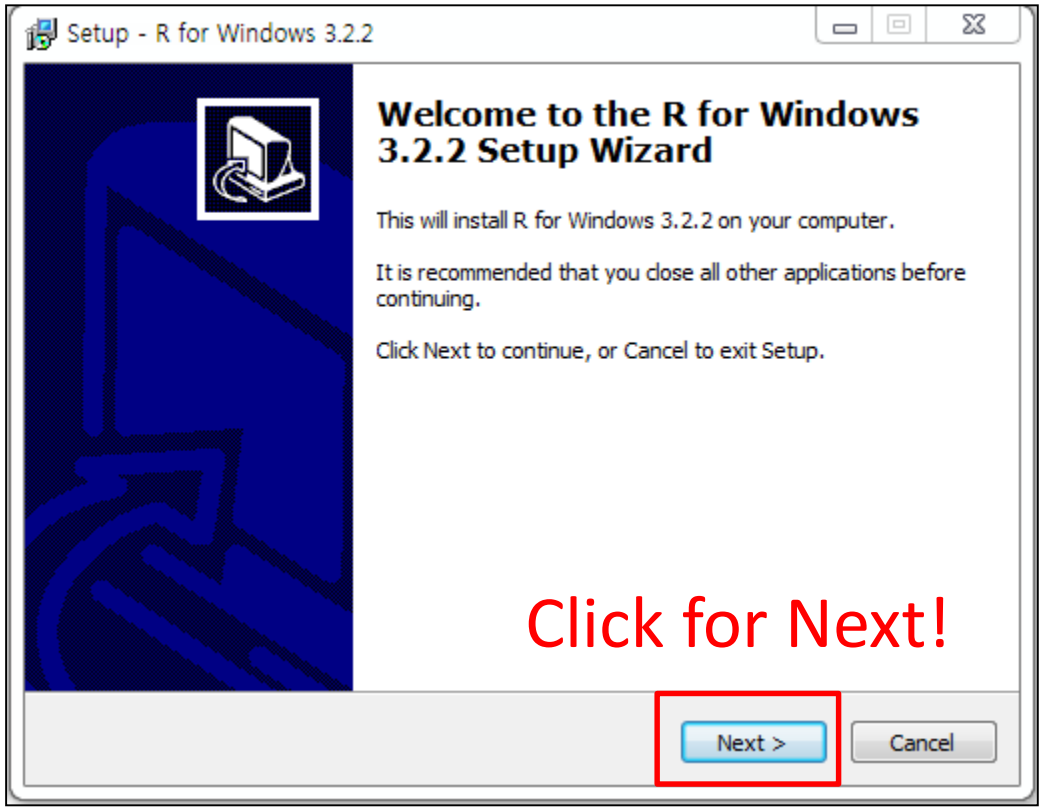


Click here!
Drop down menu appears.
Choose language!

Now, the window displays Korean.
The display language depends on your computer system!

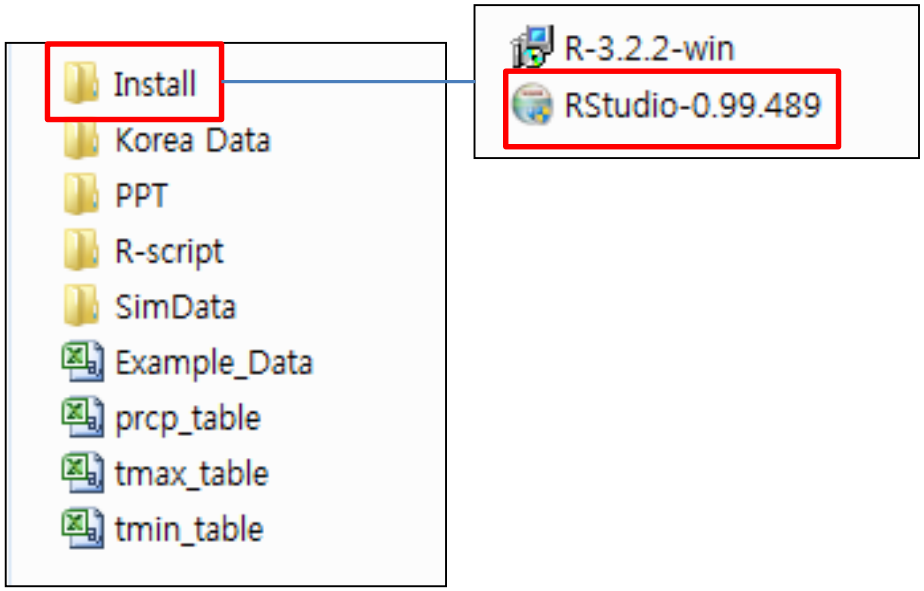
Preparation

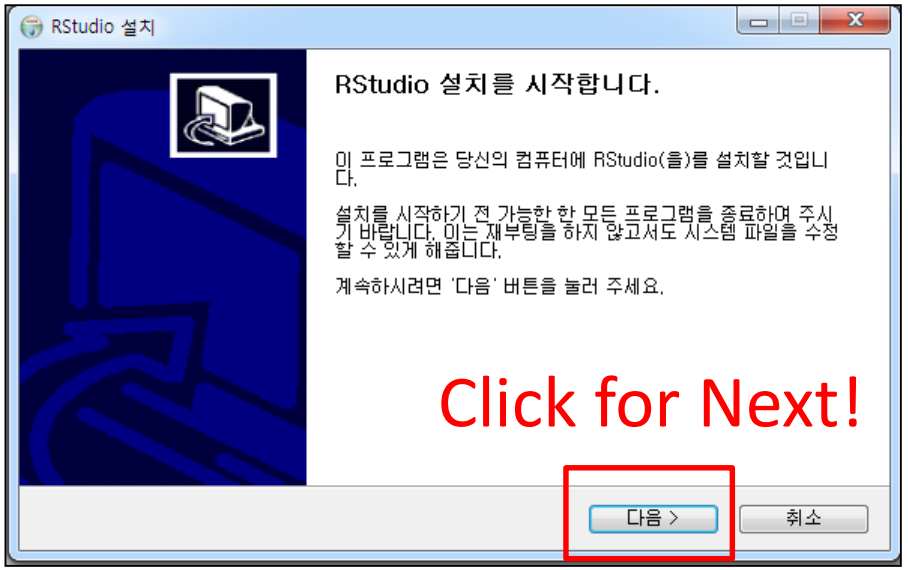
- Install R
  - Click **Next** button repeatedly by completing.



Preparation

- Install R Studio
  - Go to **Install** folder
  - Double-click **Rstudio-0.99.489**

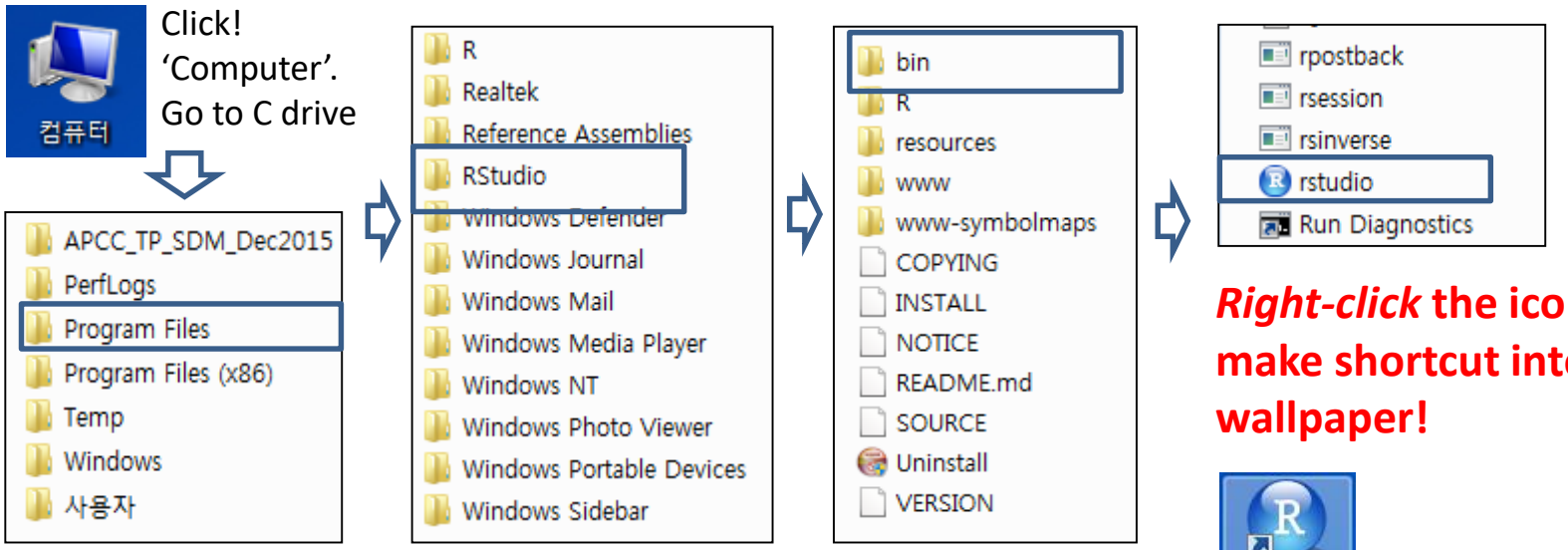




Now, the window displays Korean.
The display language depends on your computer system!

Preparation

- Make icon on wallpaper



* Displays depend on your computer system

Preparation

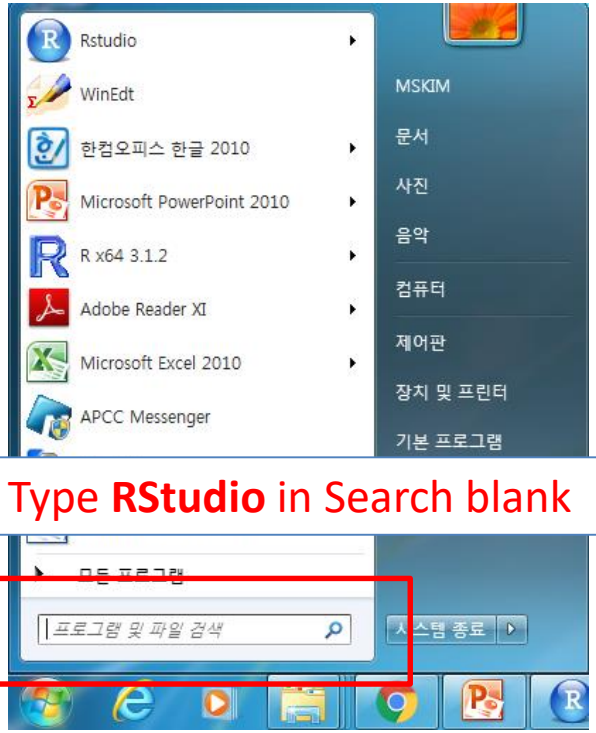
- Start R Studio by following a way:
  - Double-clicking in wallpaper
    Then, R Studio starts!

  - If the icon not appear, then
    Start > Search > Searching **RStudio** > Click!
    (See the next page for illustration!)

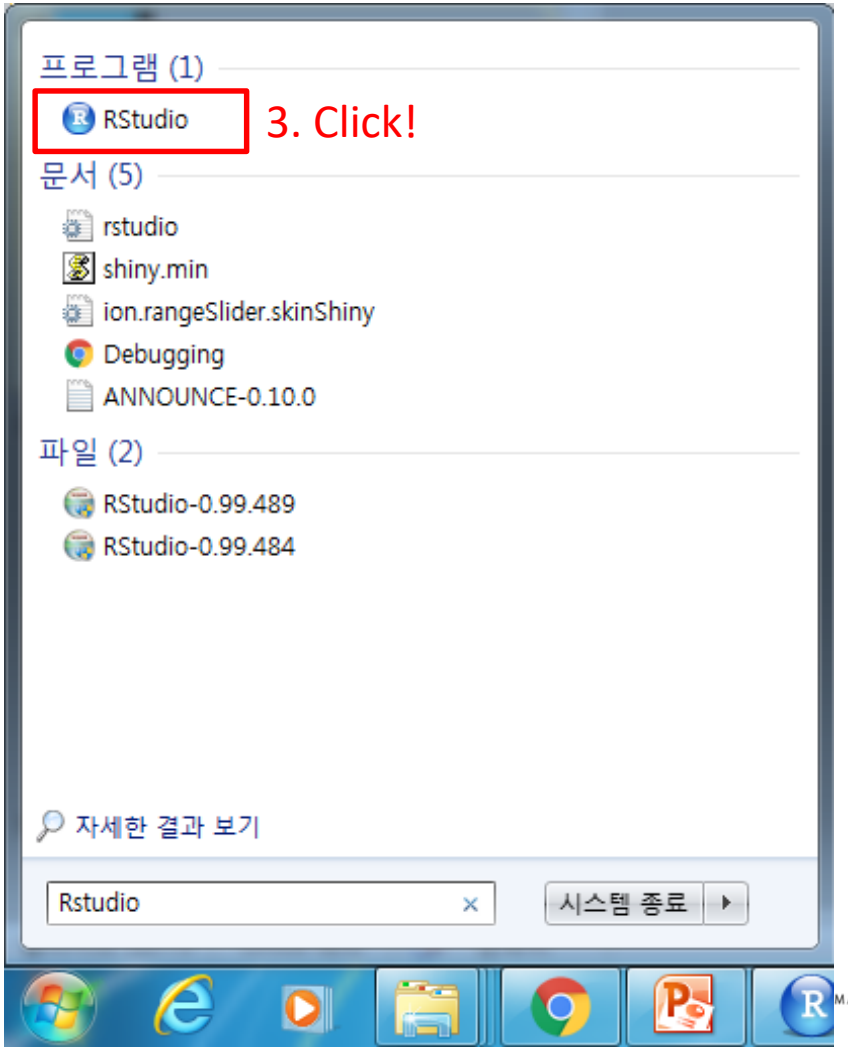
Preparation

- Start R Studio

1. Click **Start** Button

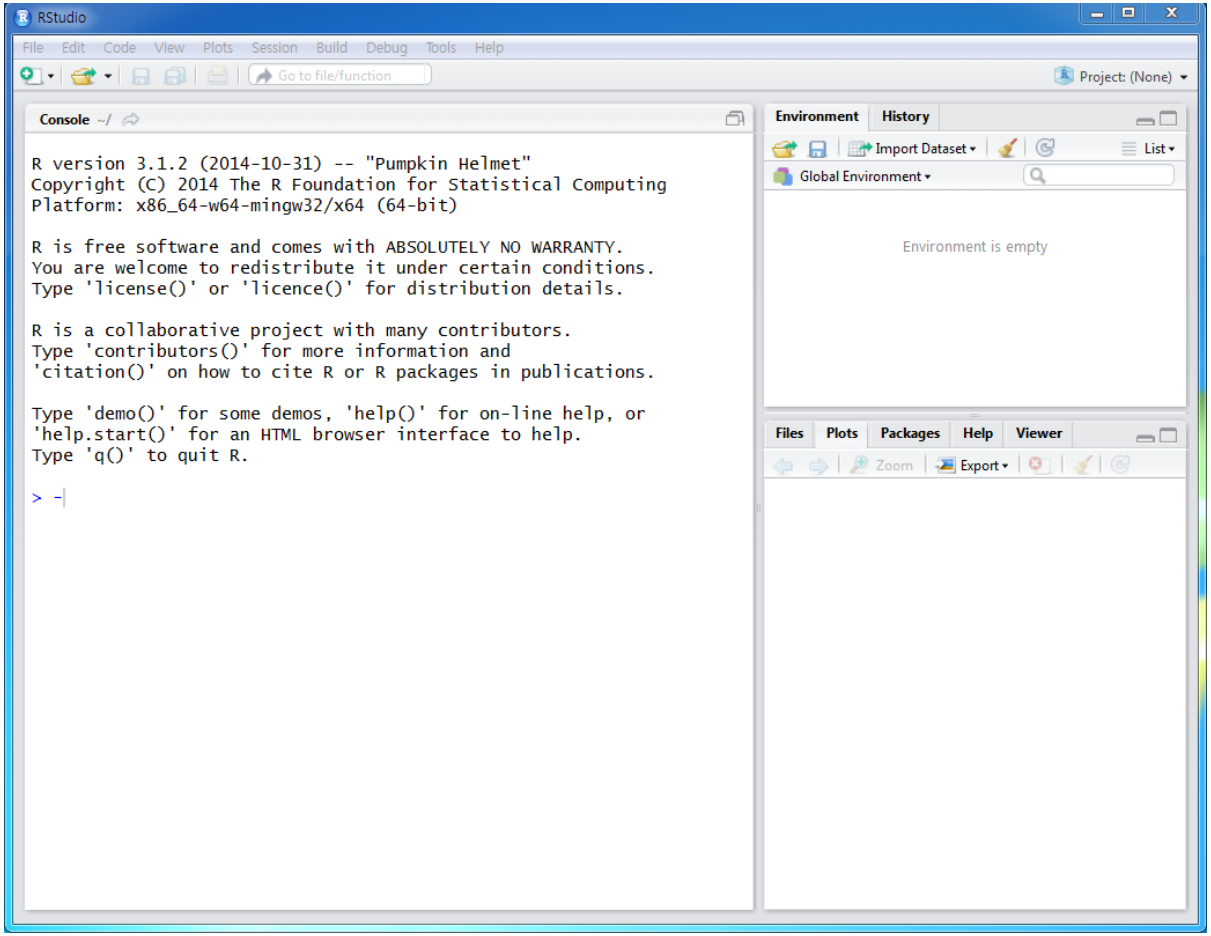


2. Type **RStudio** in Search blank



Preparation

- Start Rstudio! (Display can be slightly different.)



Preparation

- Open R-code file **Quick_R_Start**
  - File menu > Open File
  - Selecting the file in **C:\APCC_TP_SDM_Aug2016\WGenerator\R-script**
    Next page, the illustration is presented.

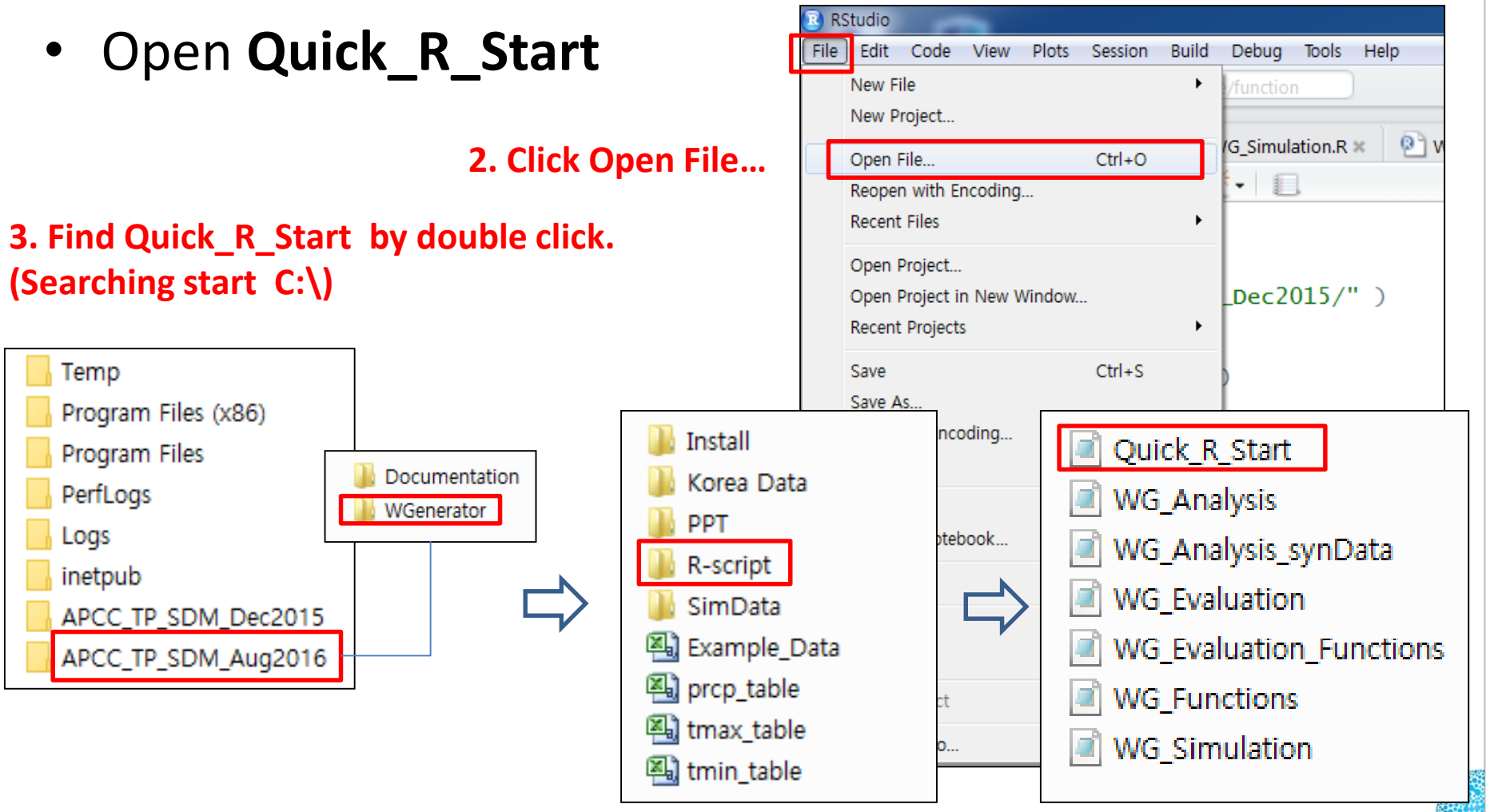
Preparation

- Open **Quick_R_Start**

1. Click File

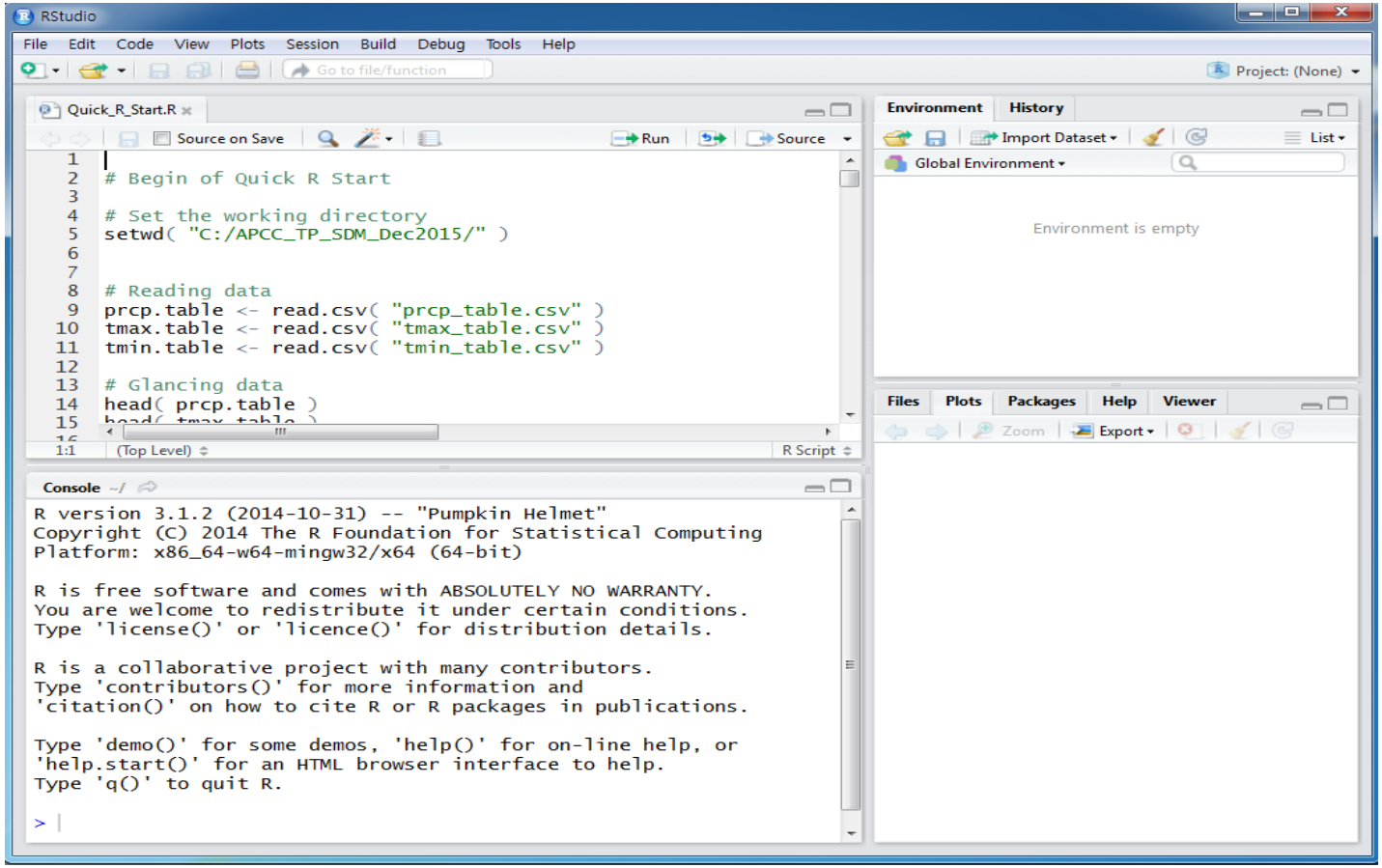
2. Click Open File…

3. Find Quick_R_Start by double click.
(Searching start C:\)



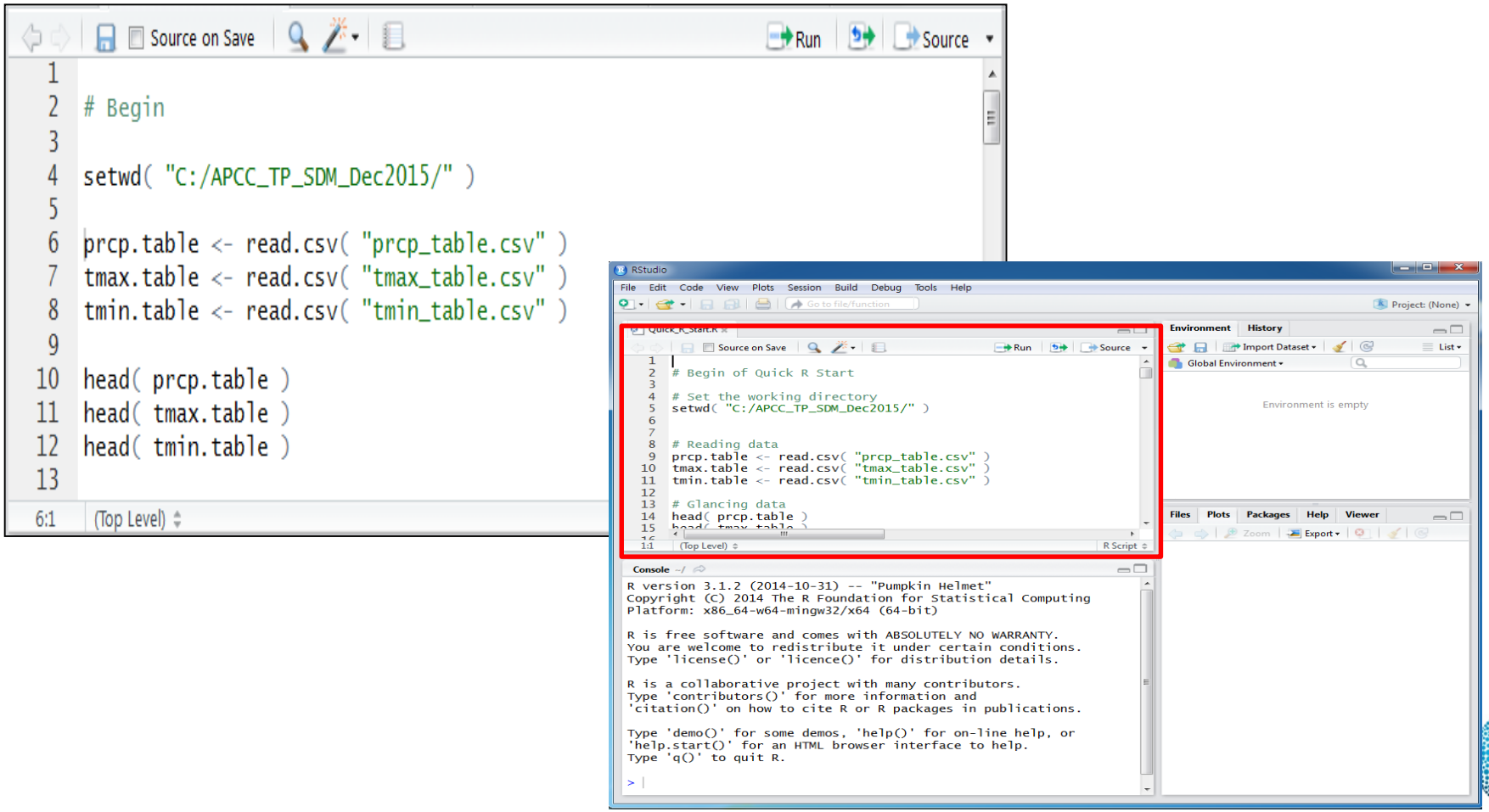
Preparation

- R studio has several windows



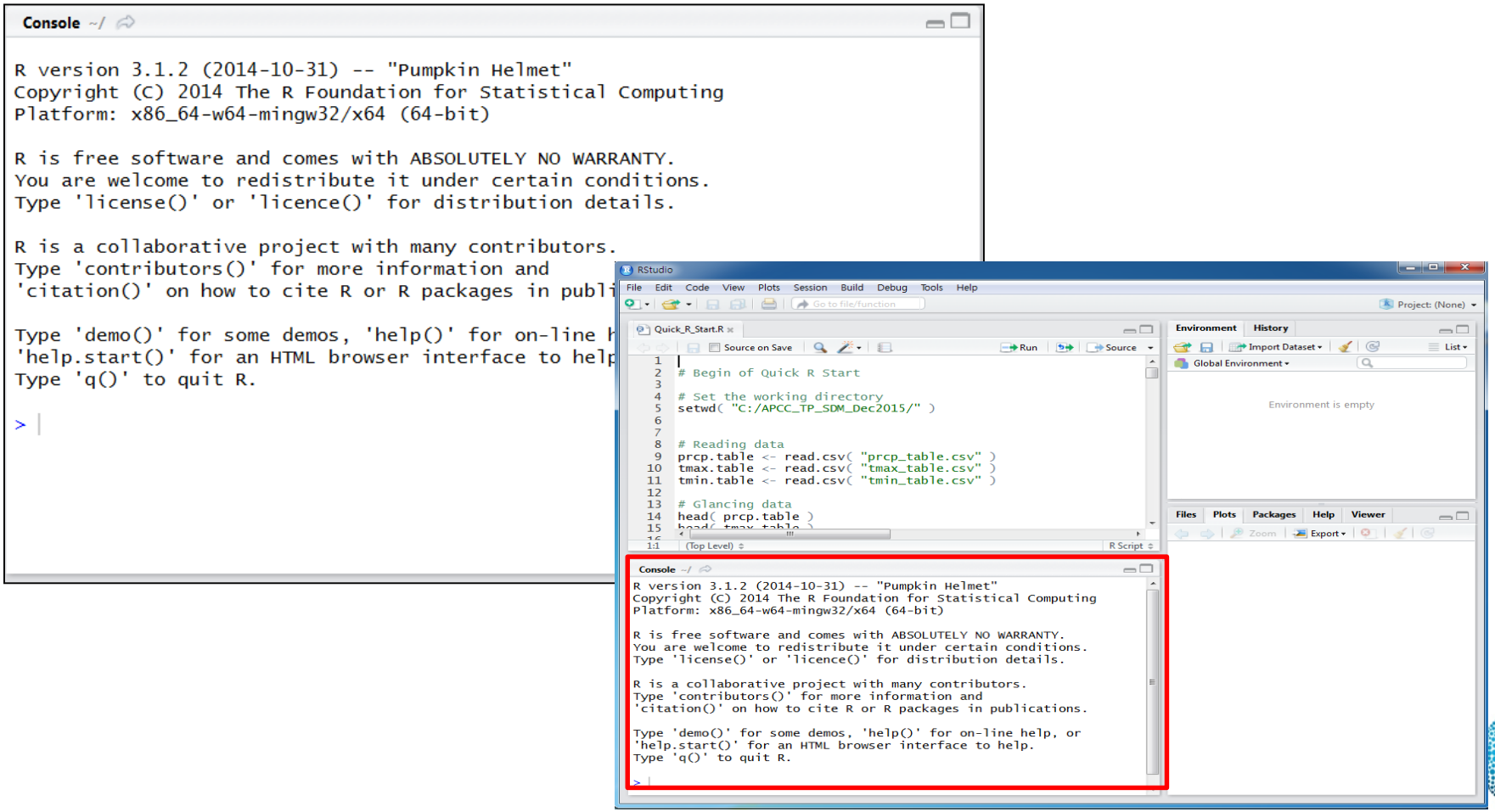
Preparation

- R-script window



Preparation

- Console



Contents

- Preparation and installation
- Variables and Operations
- Data Analysis using R
- Reading your Observation Data

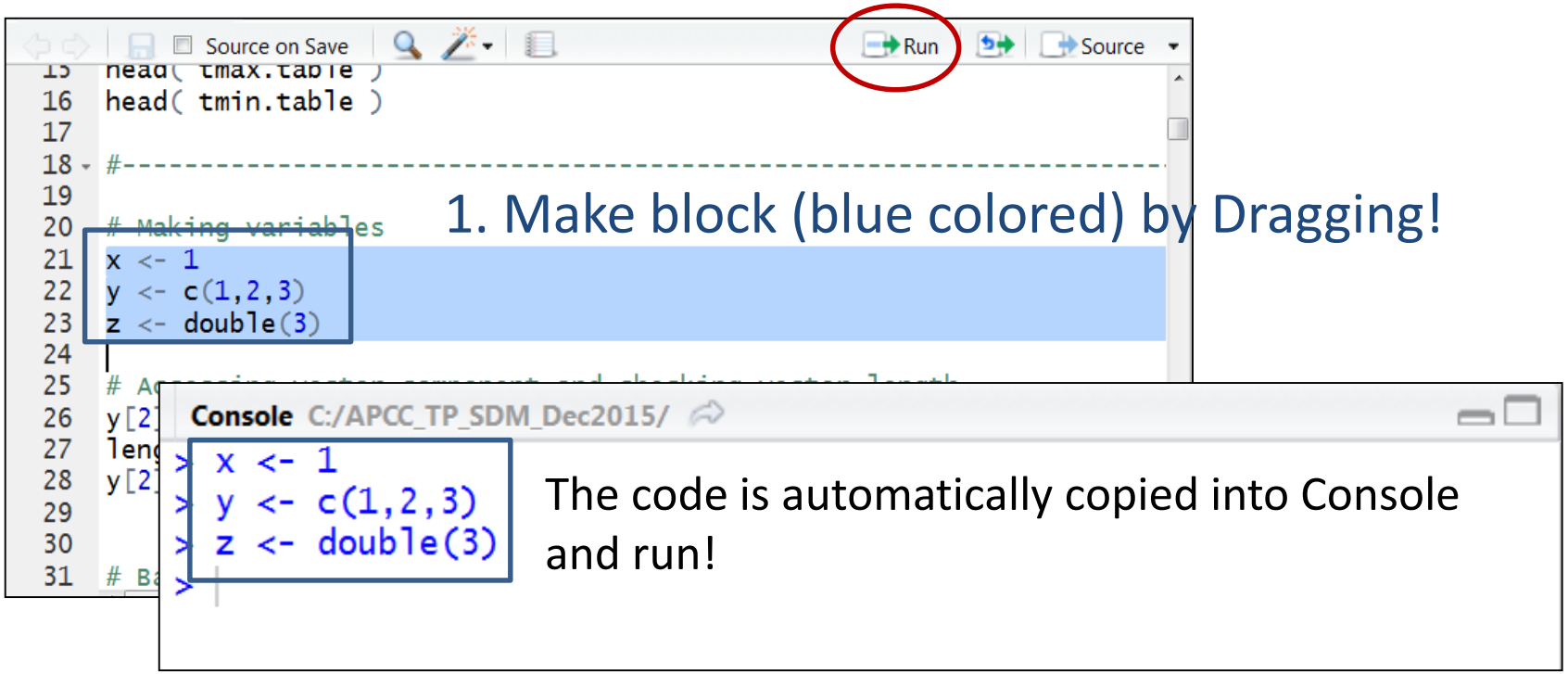
Variables and Operation

- Creating variable and vector

  - Variable ‘x’ is created with value 1.
  - Variable ‘y’ is a vector with initial values.
  - Variable ‘z’ is a vector with length 3 and default value 0.
  - Other characters and digits can be used for variable name. e.g., ‘x1’, ‘z2’, ‘ab1’, etc.

Variables and Operation

- How to run the code



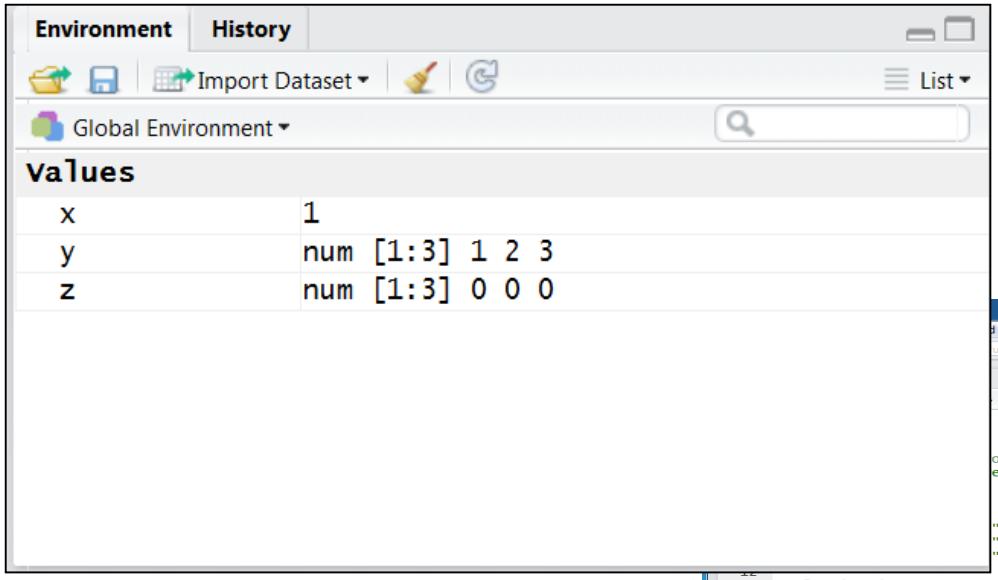
Variables and Operation

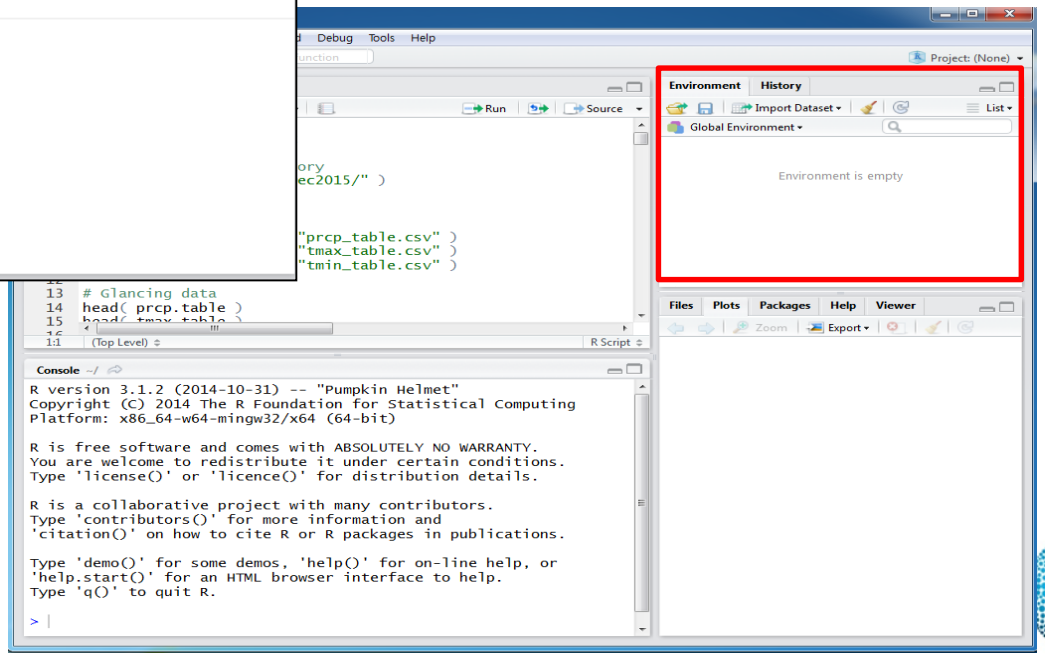
- Checking variables

- Type “x” on Console *directly* and push ‘Enter’!
  You can check the value of variable ‘x’.
- The others can be checked in the same way!

Variables and Operation

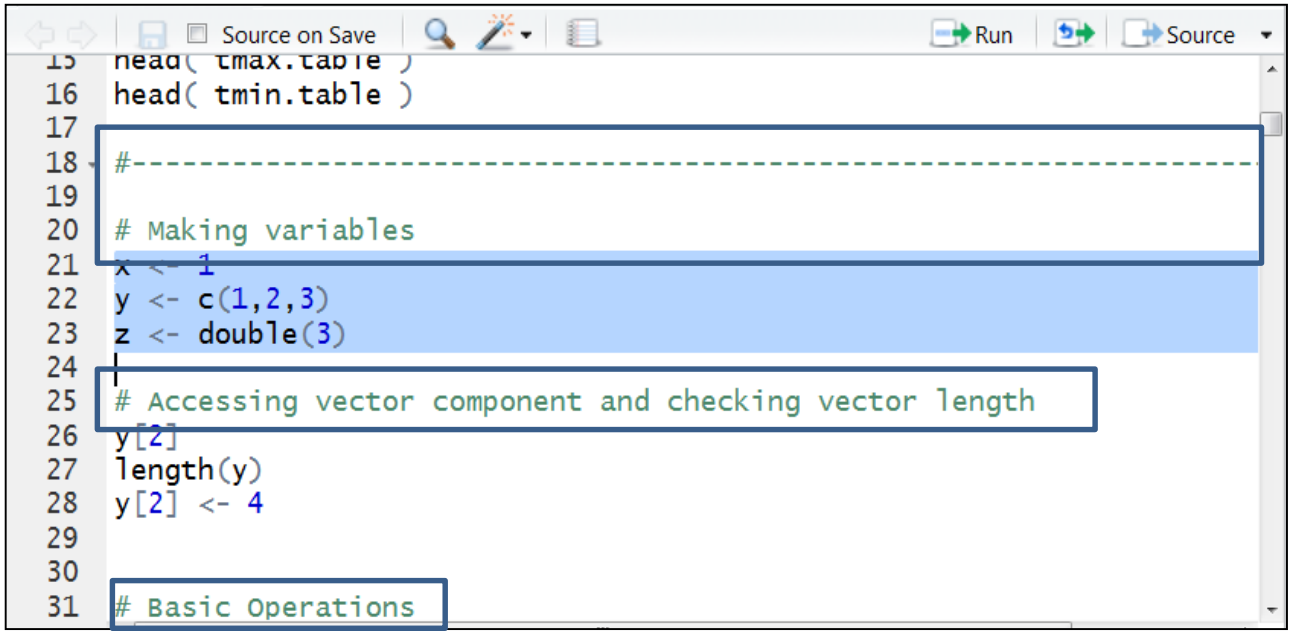
- You can check by seeing Environment window.





Variables and Operation

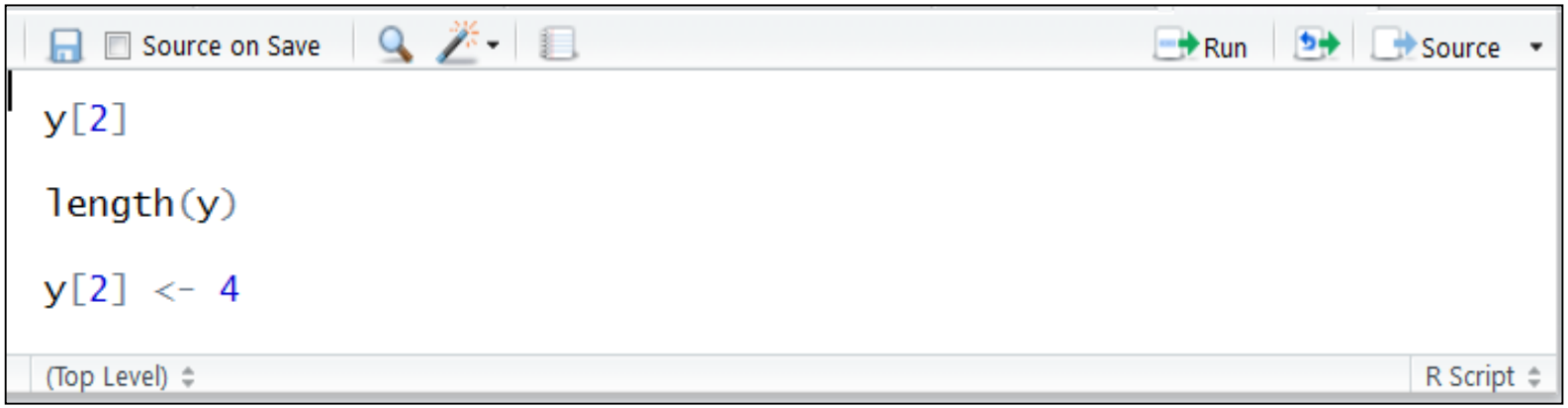
- # Comment



- The sentences with # head and green color are **comments** *which describes the code*.
- **They are not run when copied into Console**.

Variables and Operation

- Operation

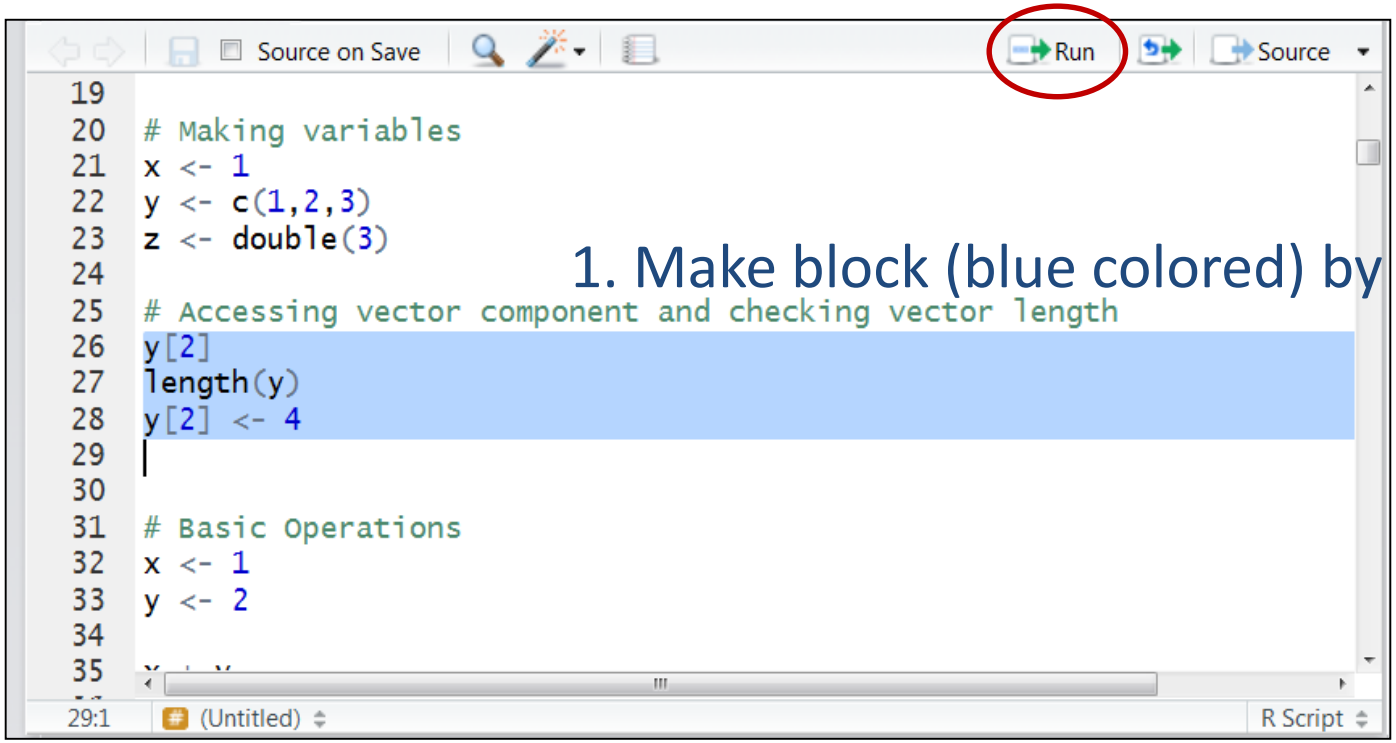


- Accessing the second entry of ‘y’.
- Length of vector ‘y’.
- Changing the value of y[2] to 4.

Variables and Operation

- Operation

2. Click!

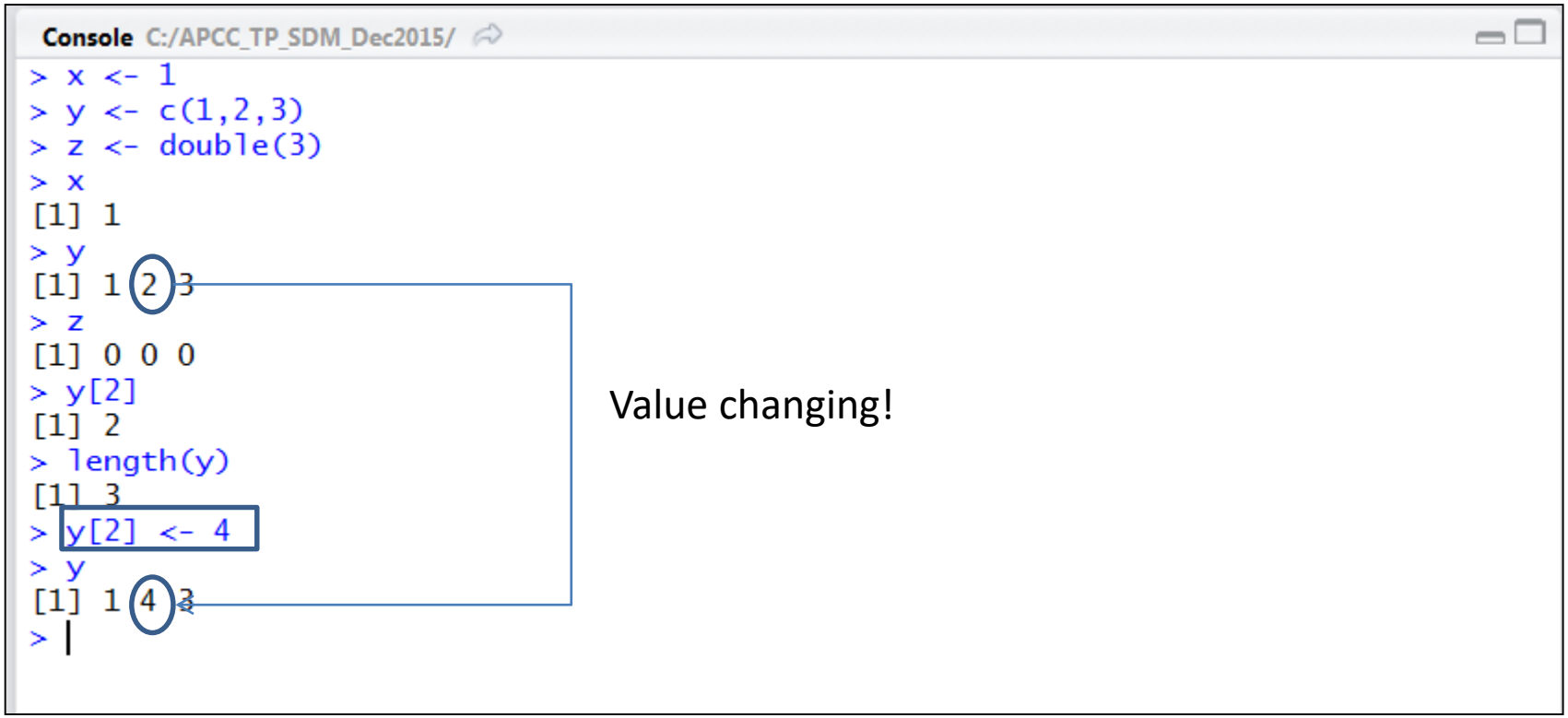


1. Make block (blue colored) by Dragging!

Remind: The sentences with # head and green color are **comments** which describes the code.
They are not run even though copied into Console.

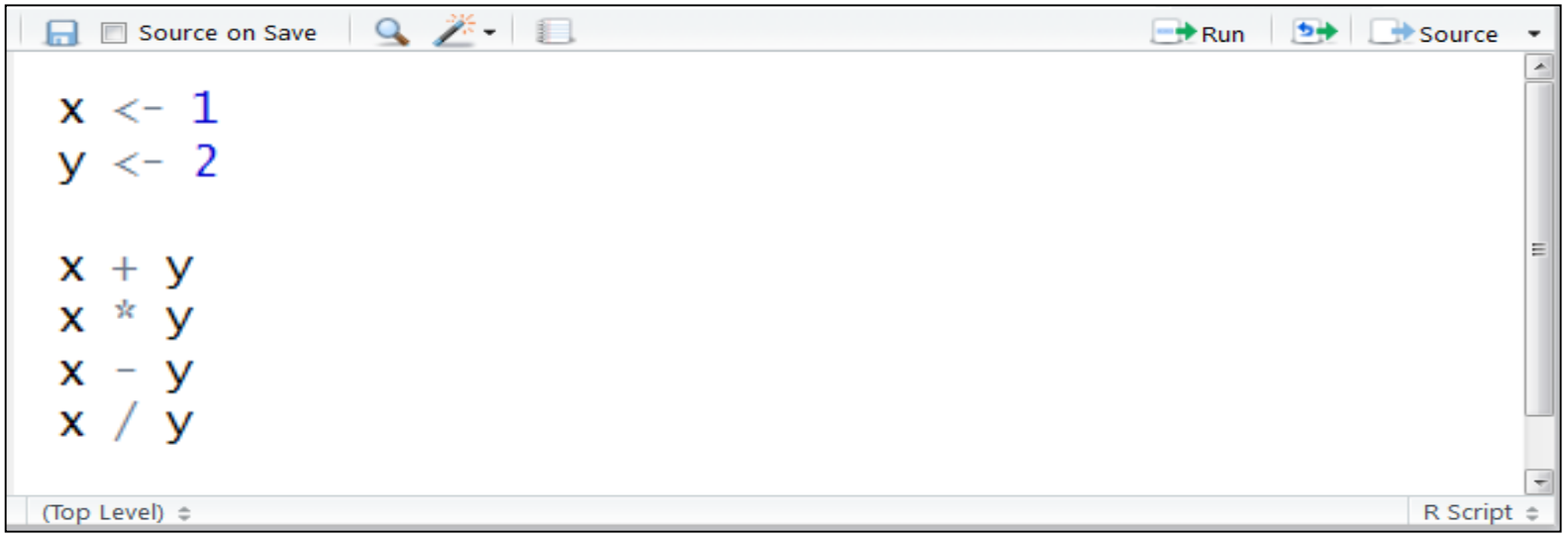
Variables and Operation

- Operation



Variables and Operation

- Operation



  - Two variables ‘x’ and ‘y’ with initial values
  - Addition (+), Multiplication (*), Subtraction (-), Division (/)

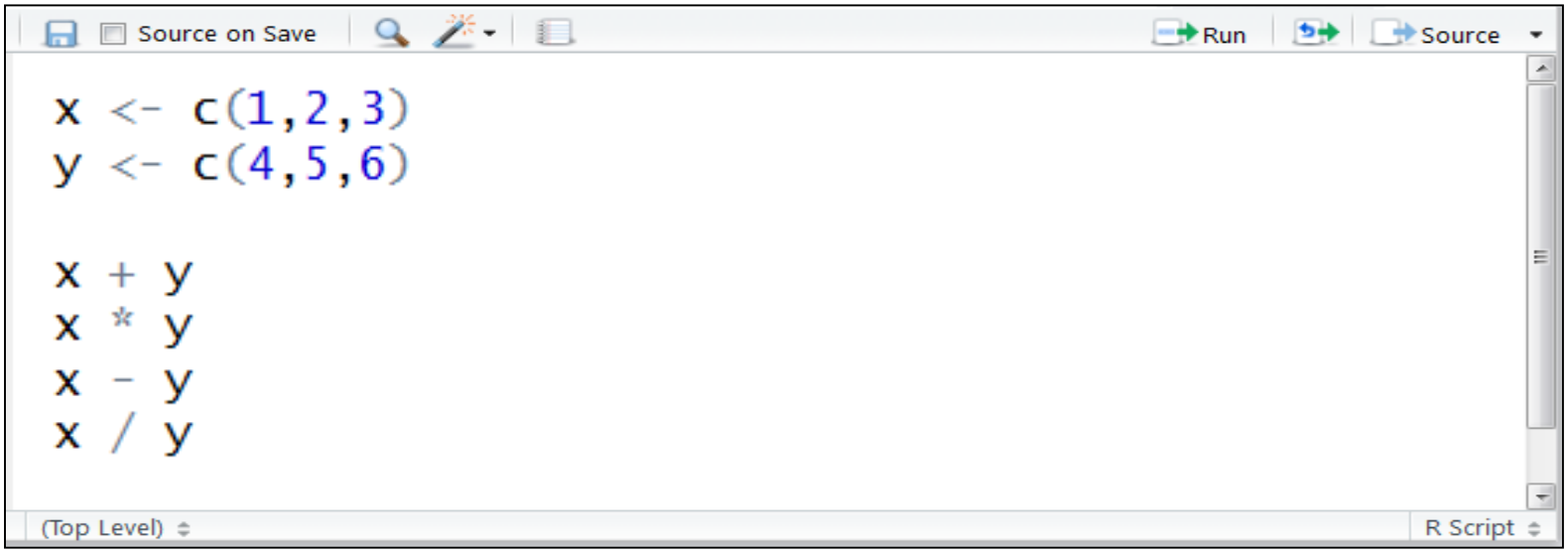
Variables and Operation

- Operation

```
Console C:/APCC_TP_SDM_Dec2015/
> x <- 1
> y <- 2
>
> x + y
[1] 3
> x * y
[1] 2
> x - y
[1] -1
> x / y
[1] 0.5
>
```

Variables and Operation

- Operation



- Two vectors ‘x’ and ‘y’.
- Operations are same, but every operation is made *component-wisely*!

Variables and Operation

- Operation

```
Console C:/APCC_TP_SDM_Dec2015/
> x <- c(1,2,3)
> y <- c(4,5,6)
>
> x + y
[1] 5 7 9
> x * y
[1]  4 10 18
> x - y
[1] -3 -3 -3
> x / y
[1] 0.25 0.40 0.50
>
```

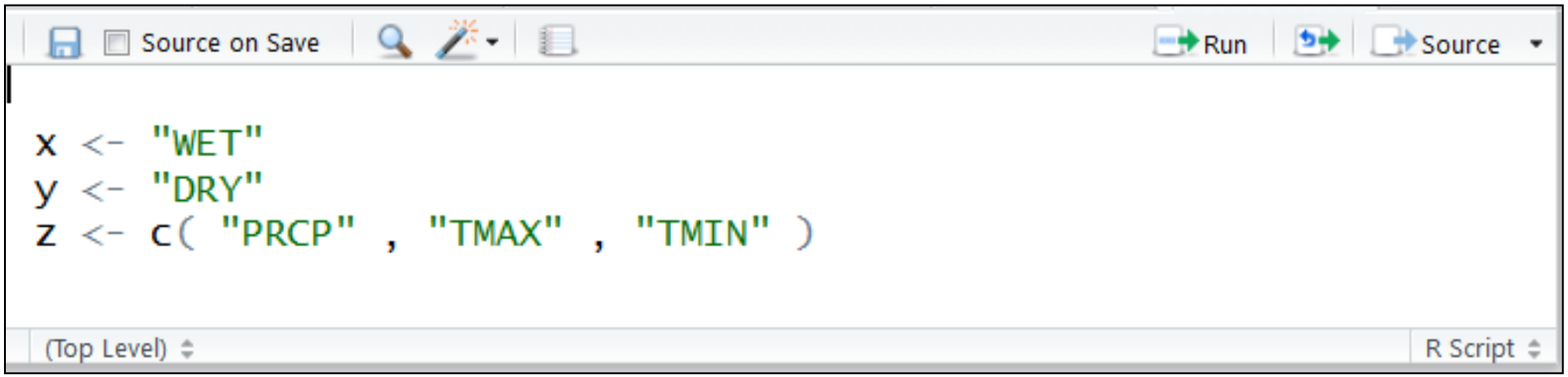
Note: operations are applied to every component, separately.

Variables and Operation

- Practice:
  - Create variables ‘a’, ‘b’ and vectors ‘z’, ‘w’ with length 5 and initial values you want.
  - Access components of vectors ‘z’ and ‘w’.
  - Operate the variables. For instance, z + w. Check the operation results.

Variables and Operation

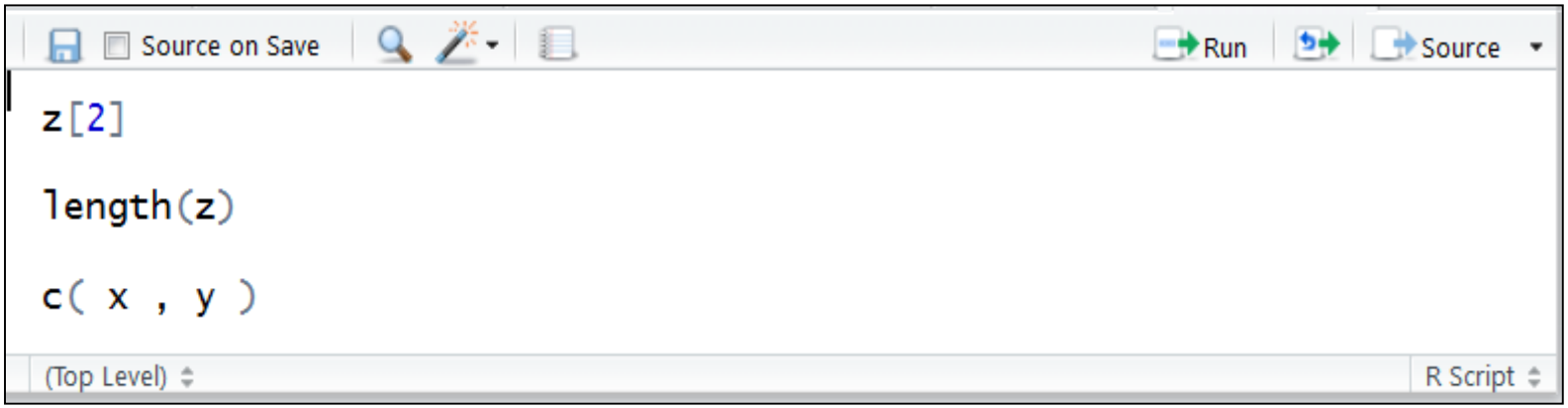
- Character variable and vector



- Variable ‘x’ is created with “WET” as initial.
- Variable ‘y’ is created with “DRY” as initial.
- Variable ‘z’ is a vector of length 3 with the initial characters.

Variables and Operation

- Character variable and vector



  - Accessing the second entry of vector ‘z’.
  - Length of vector ‘z’.
  - Connecting two character variables ‘x’ and ‘y’.

Variables and Operation

- Character variable and vector

```
Console C:/APCC_TP_SDM_Dec2015/
> x <- "WET"
> y <- "DRY"
> z <- c( "PRCP" , "TMAX" , "TMIN" )
>
> z[2]
[1] "TMAX"
>
> length(z)
[1] 3
>
> c( x , y )
[1] "WET" "DRY"
>
```

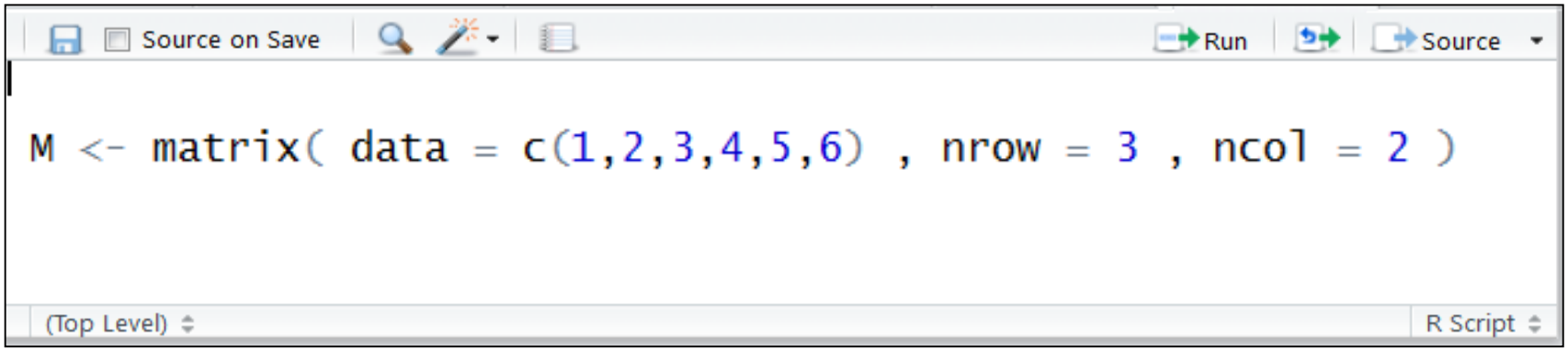
Note: “c(x,y)” connecting two variables ‘x’ and ‘y’ into a vector of length 2.

Variables and Operation

- Practice:
  - Create character variables ‘x’, ‘y’ and vectors ‘z’, ’w’ with length 3 and initial characters you want.
  - Access components of character vectors,
  - Connecting
    - ‘x’ and ‘y’,
    - ‘z’ and ‘w’,
    - ‘x’ and ‘z’.

Variables and Operation

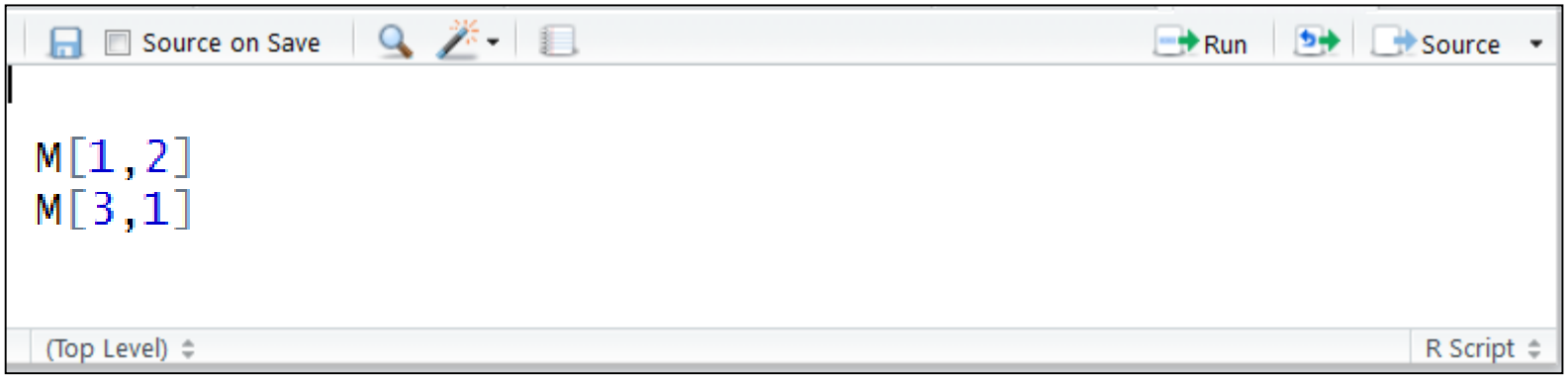
- Creating matrix and access



  - 'M' is a matrix of 3×2 with initial values.

Variables and Operation

- Creating matrix and access (cont.)



  - Accessing the entry of matrix ‘M’
    - The entry of the first row and second column.
    - The entry of the third row and first column.

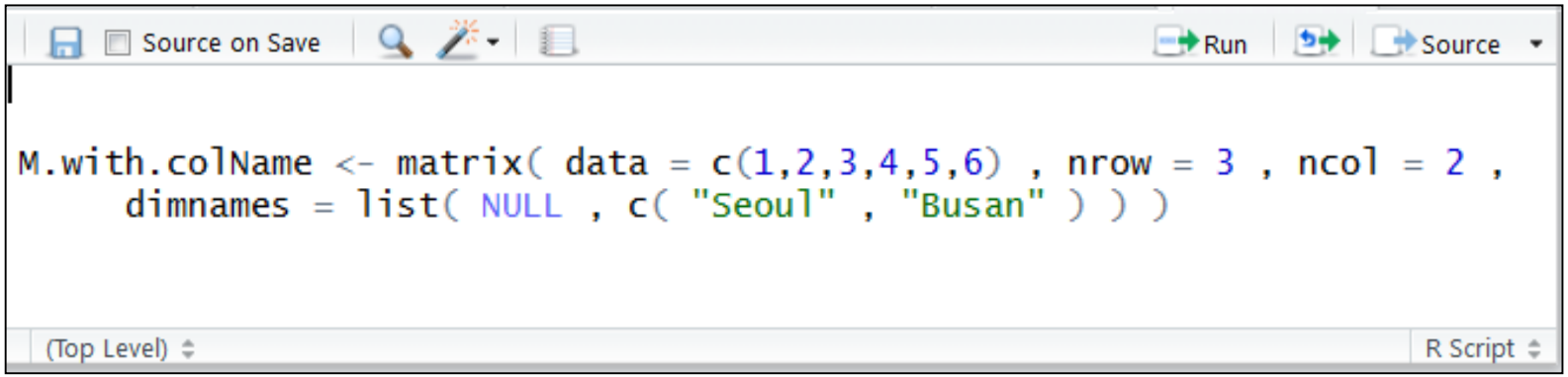
Variables and Operation

- Creating matrix and access (cont.)

```
Console C:/APCC_TP_SDM_Dec2015/
> M <- matrix( data = c(1,2,3,4,5,6) , nrow = 3 , ncol = 2 )
>
> M
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
>
> M[1,2]
[1] 4
> M[3,1]
[1] 3
>
```

Variables and Operation

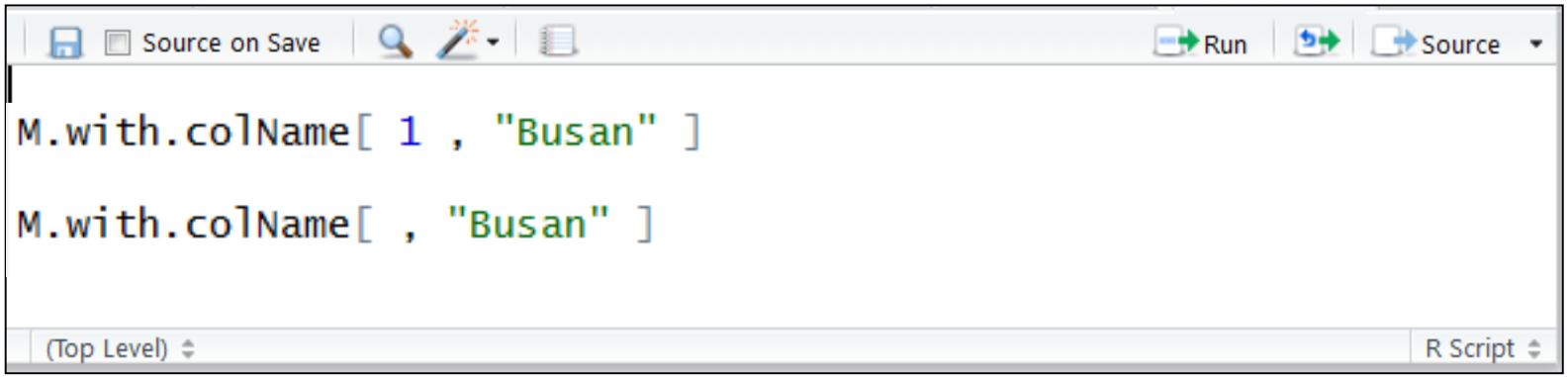
- Creating matrix with column names



  - 'M.with.colName' is created with column names "Seoul" and "Busan".

Variables and Operation

- Accessing matrix with column names



  - Accessing the entry of the first row and "Busan" column.
  - Accessing the entries of "Busan" column.

Variables and Operation

- Accessing matrix with column names

```
Console ~/
> M.with.colName <- matrix( data = c(1,2,3,4,5,6) , nrow = 3 , ncol = 2 ,
+       dimnames = list( NULL , c( "Seoul" , "Busan" ) ) )
>
> M.with.colName
     Seoul Busan
[1,]     1     4
[2,]     2     5
[3,]     3     6
>
> M.with.colName[ 1 , "Busan" ]
Busan
   4
>
> M.with.colName[ , "Busan" ]
[1] 4 5 6
>
```

Variables and Operation

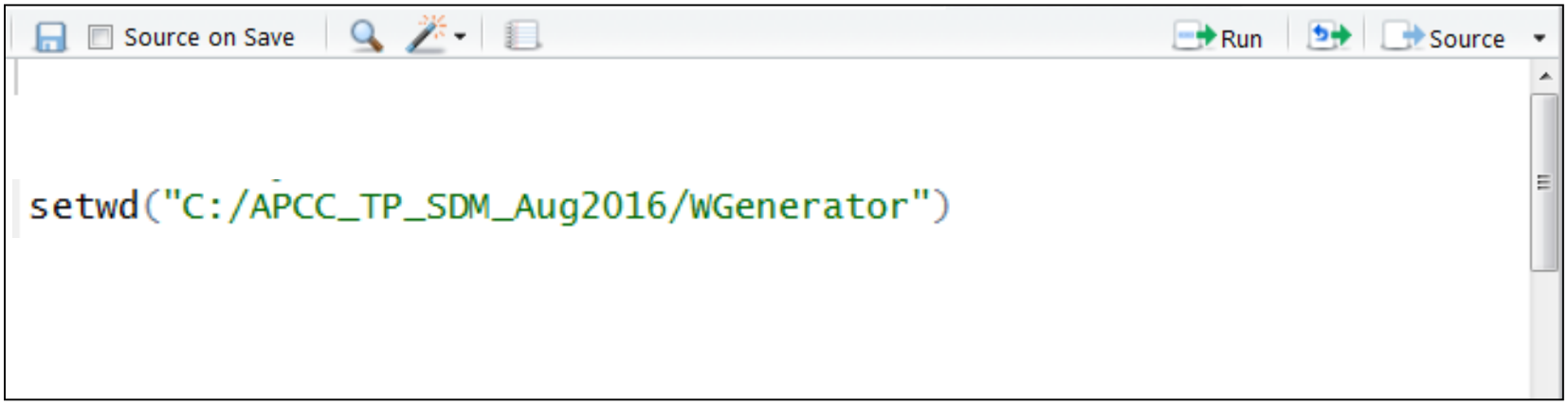
- Practice
  - Create a new matrix ‘Mat’ of another size.
    Note: *‘data vector length’ = nrow × ncol!*
  - Access entries of the created matrix ‘Mat’.
  - Access entries of “Seoul” column of ‘M.with.colName’.

Contents

- Preparation and installation
- Variables and Operations
- Data Analysis using R
- Reading your Observation Data

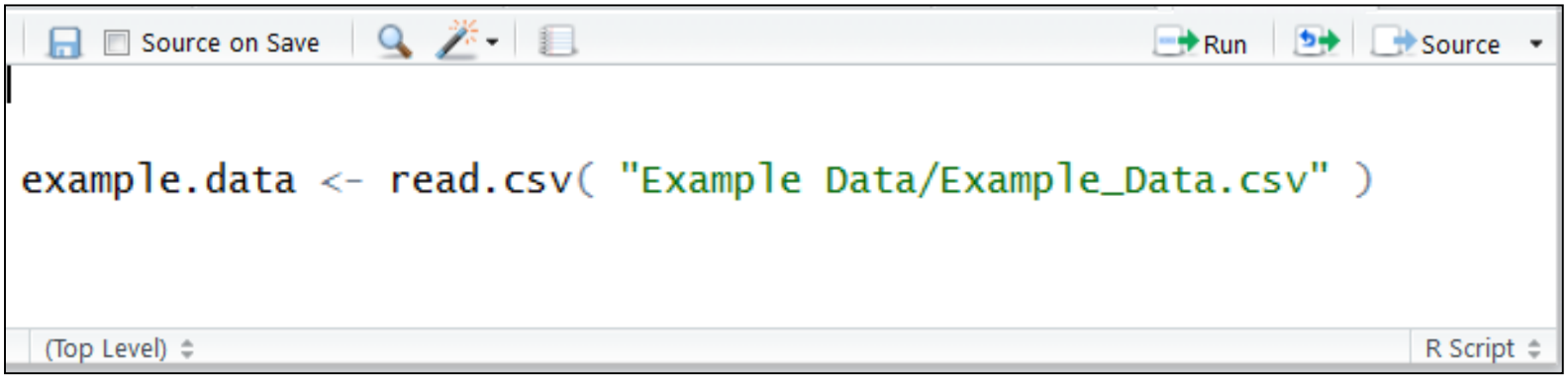
Data Analysis

- Set working directory.



Data Analysis

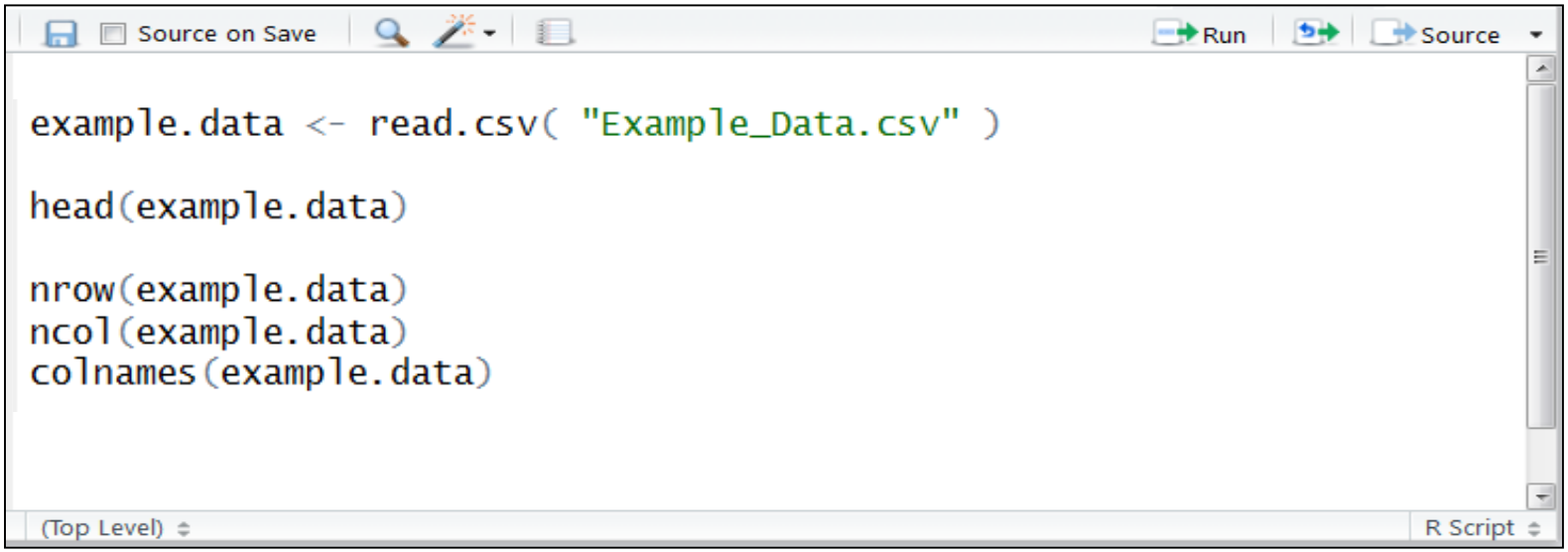
- Read data file **Example_Data.csv**



  - Variable **example.data** is a matrix.
  - It has 5 columns whose names are MONTH, DAY, PRCP, TMAX, TMIN.

Data Analysis

- Variable **example.data**



- nrow(): the number of rows
- ncol(): the number of columns
- colnames(): the names of columns

Data Analysis

- Variable **example.data**

Console C:/APCC_TP_SDM_Dec2015/

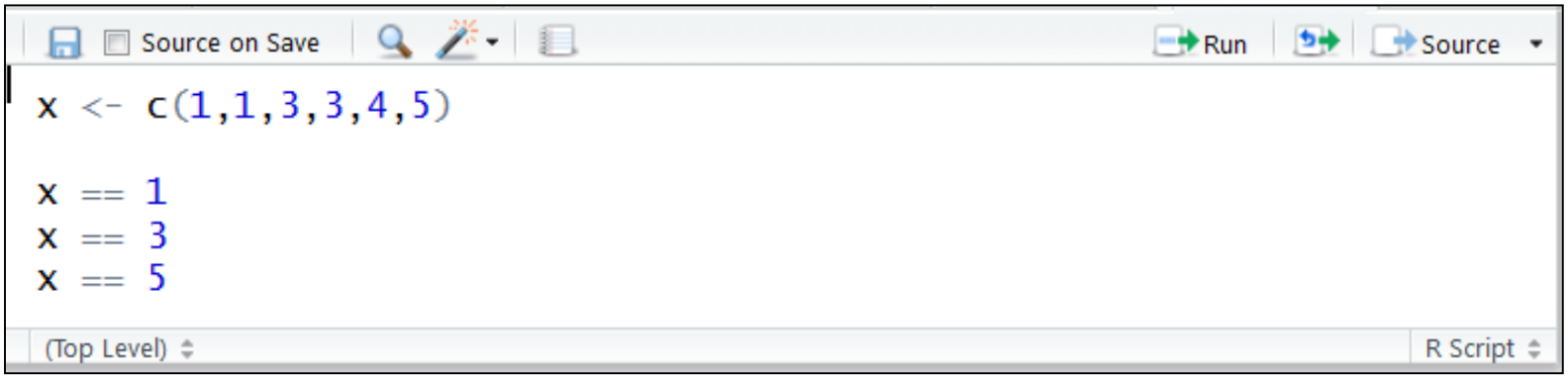
```
> example.data <- read.csv( "Example_Data.csv" )
> head(example.data)
  MONTH DAY PRCP TMAX TMIN
1     6   1    0 26.4 15.7
2     6   2    0 27.8 14.8
3     6   3    0 29.0 13.6
4     6   4    0 30.8 15.8
5     6   5    0 31.6 17.4
6     6   6    0 29.2 18.4
> nrow(example.data)
[1] 61
> ncol(example.data)
[1] 5
> colnames(example.data)
[1] "MONTH" "DAY"   "PRCP"  "TMAX"  "TMIN"
> |
```

Data Analysis

- Let's calculate
  - Averages of TMAX and TMIN for each month.
  - Accumulations of PRCP for each month.
  - Averages of TMAX and TMIN for wet/dry days.
    - wet day: PRCP > 0.
    - dry day : PRCP == 0

Data Analysis

- Choosing a part of vector by condition.



- Variable ‘x’ is a vector of length 6.
- ‘x == 1’ : every component is examined to be equal to 1.
  If true, the result is represented as TRUE, otherwise, FALSE.

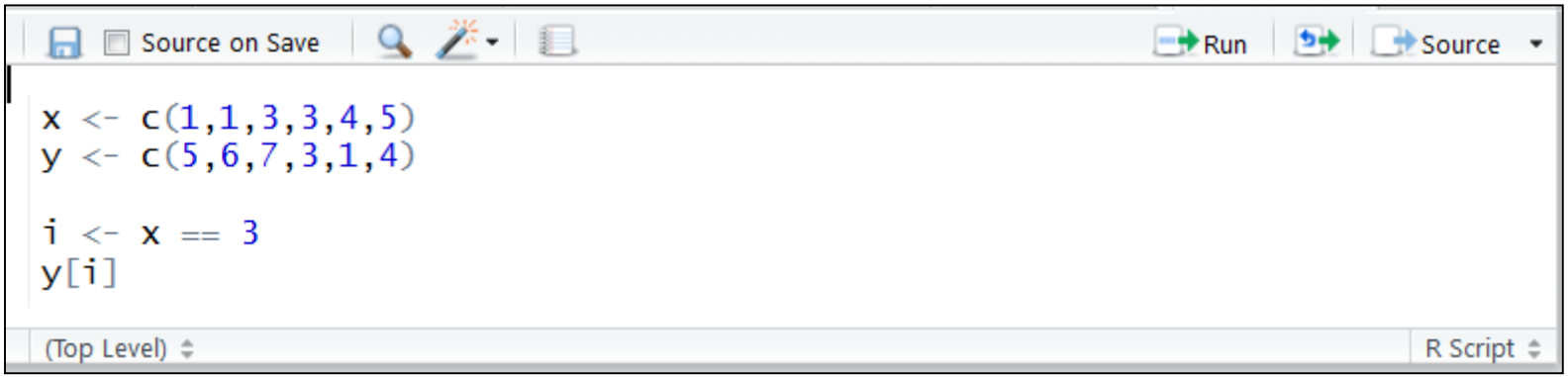
Data Analysis

- Choosing a part of vector by condition.

```
Console C:/APCC_TP_SDM_Dec2015/
> x <- c(1,1,3,3,4,5)
>
> x == 1
[1]  TRUE  TRUE FALSE FALSE FALSE FALSE
> x == 3
[1] FALSE FALSE  TRUE  TRUE FALSE FALSE
> x == 5
[1] FALSE FALSE FALSE FALSE FALSE  TRUE
>
```

Data Analysis

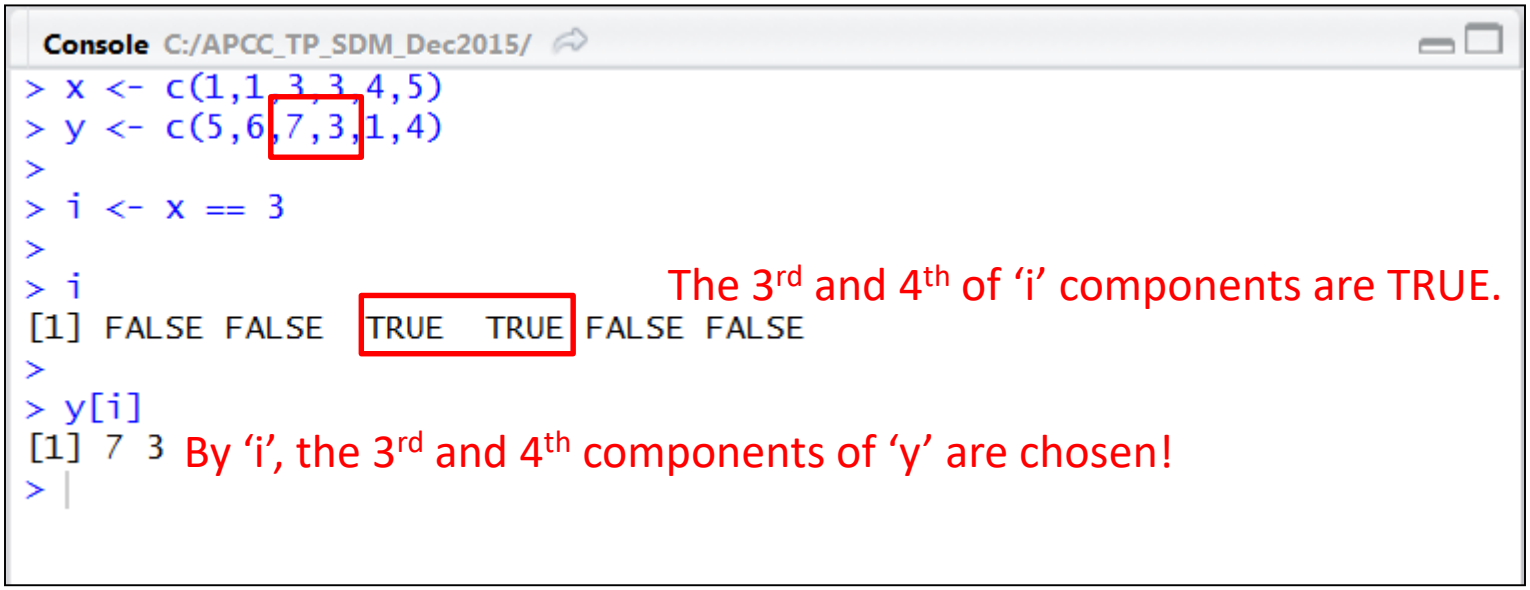
- Choosing a part of vector by condition.



  - Variable ‘i’ is a vector whose values are TRUE or FALSE.
  - ‘y[i]’ is a sub-vector of ‘y’ consisting of components with TRUE.

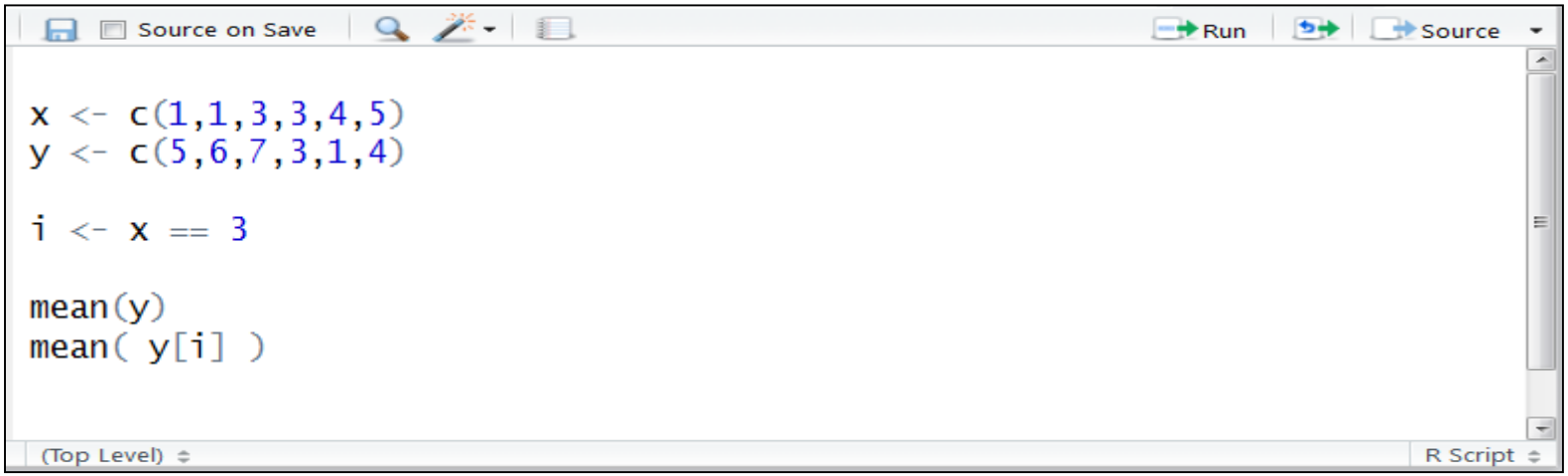
Data Analysis

- Choosing a part of vector by condition.



Data Analysis

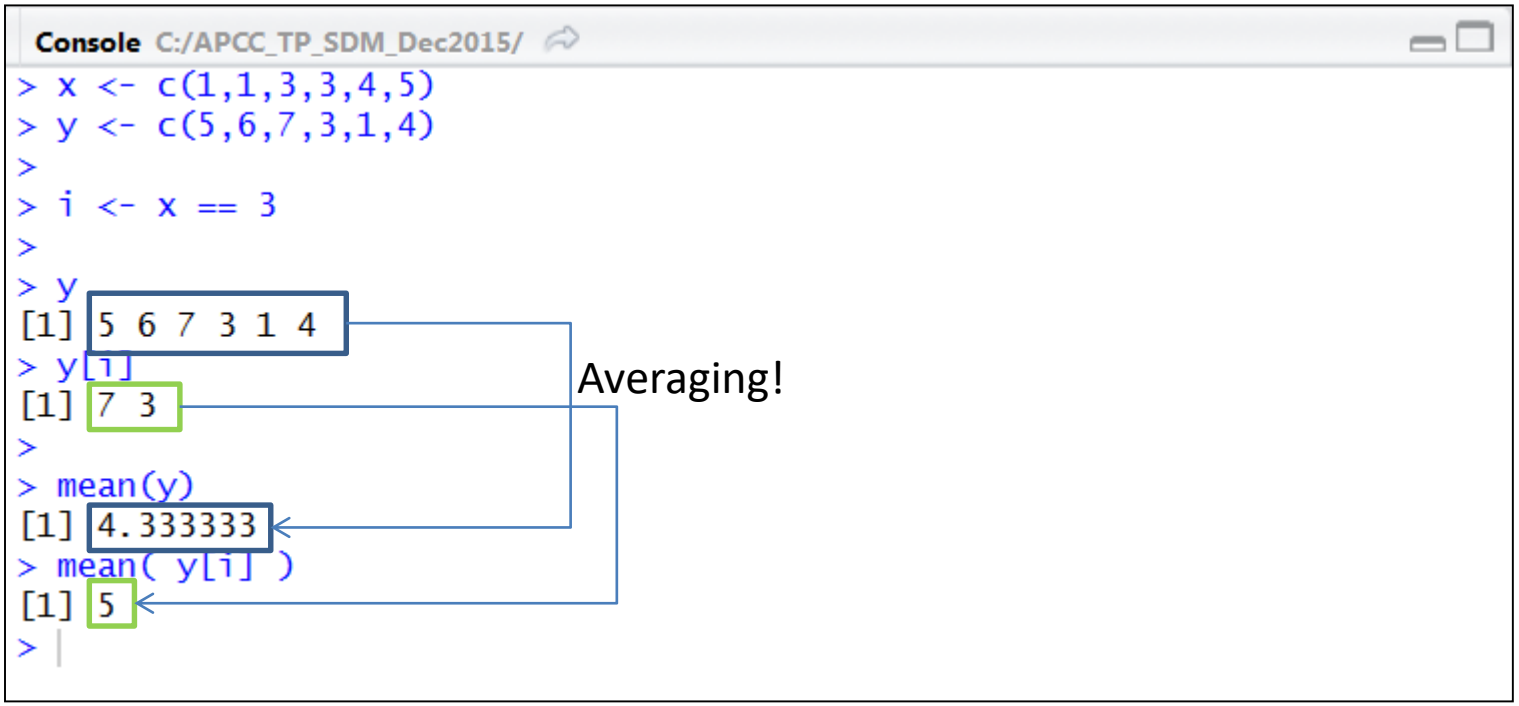
- Averaging



- Function ‘mean’ is used.
- mean(y): averaging all components
- mean(y[i]): averaging the chosen components

Data Analysis

- Averaging

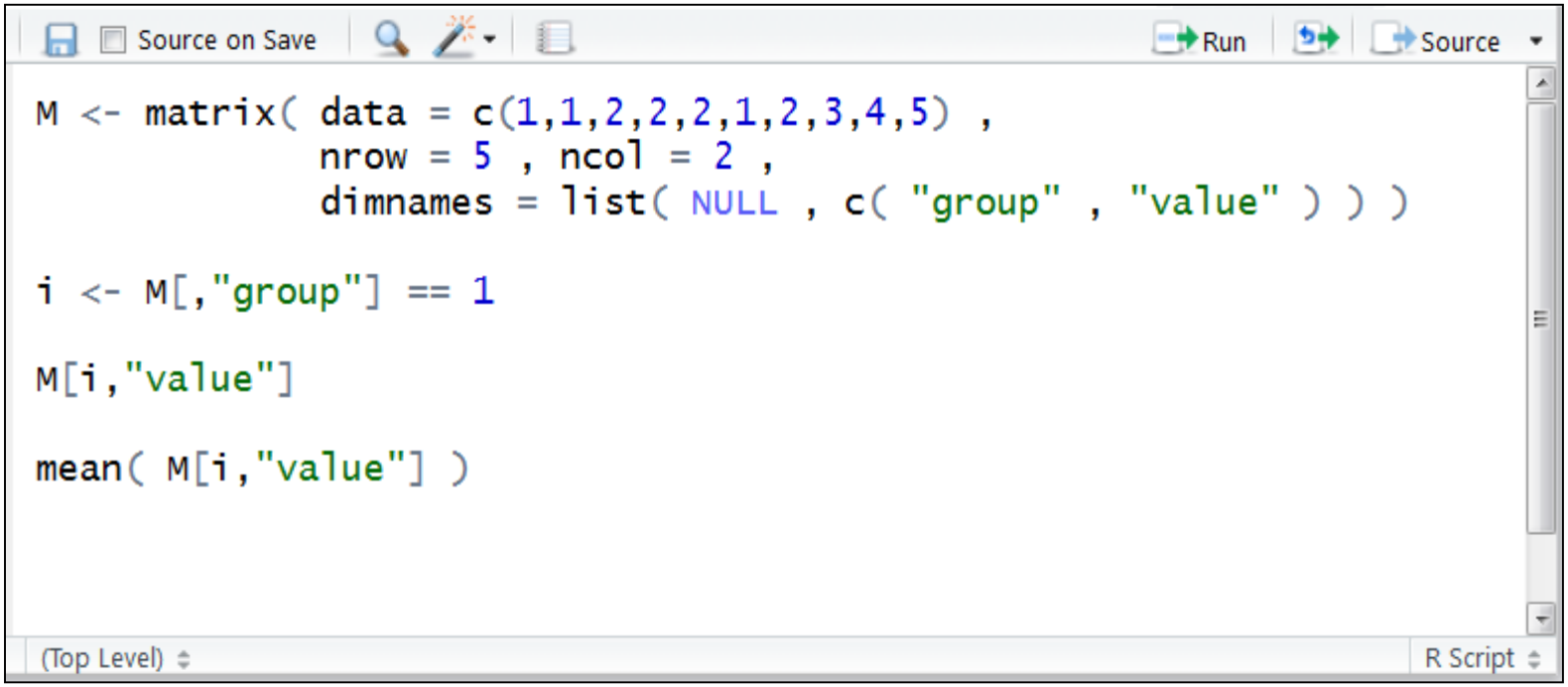


Data Analysis

- Exercise
  - Choose part of ‘y’ corresponding to x == 1
  - Average the part.

Data Analysis

- Choosing and averaging column of matrix



Data Analysis

- Choosing and averaging column of matrix

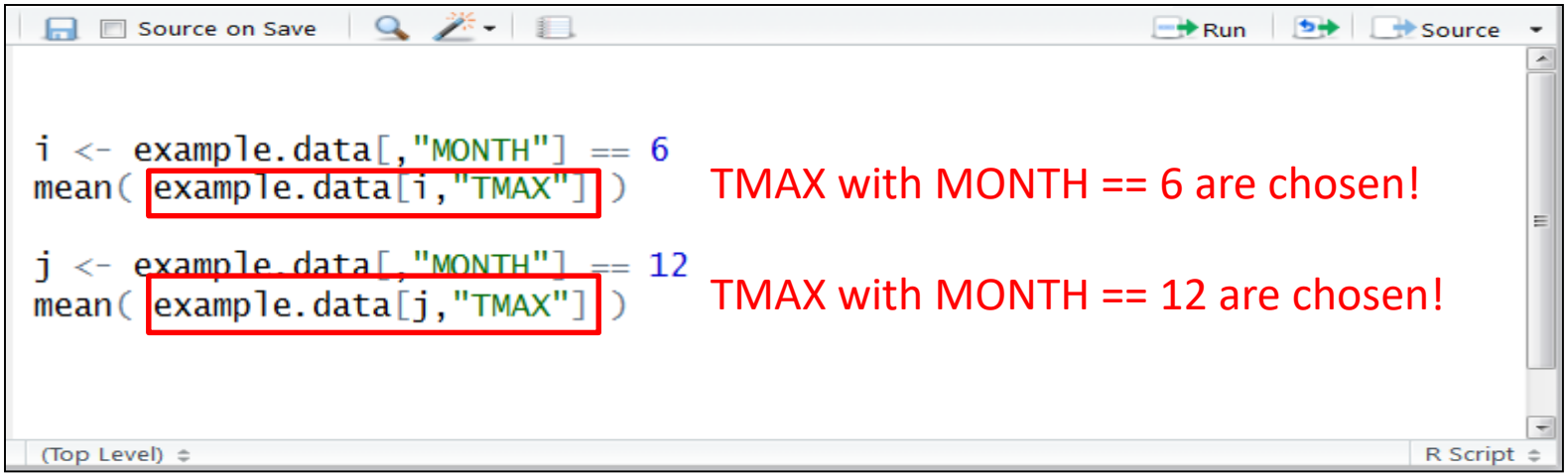
```
Console C:/APCC_TP_SDM_Dec2015/
> M <- matrix( data = c(1,1,2,2,2,1,2,3,4,5) ,
+              nrow = 5 , ncol = 2 ,
+              dimnames = list( NULL , c( "group" , "value" ) ) )
>
> M
     group value
[1,]     1     1
[2,]     1     2
[3,]     2     3
[4,]     2     4
[5,]     2     5
>
> i <- M[,"group"] == 1
>
> i
[1]  TRUE  TRUE FALSE FALSE FALSE
>
> M[i,"value"]
[1] 1 2
>
> mean( M[i,"value"] )
[1] 1.5
>
```

'value' with 'group' == 1 are chosen!

Data Analysis

- Return to **example.data**

- Calculating averages of TMAX for Jun. and Dec.



Data Analysis

- Calculating averages of TMAX for Jun. and Dec.

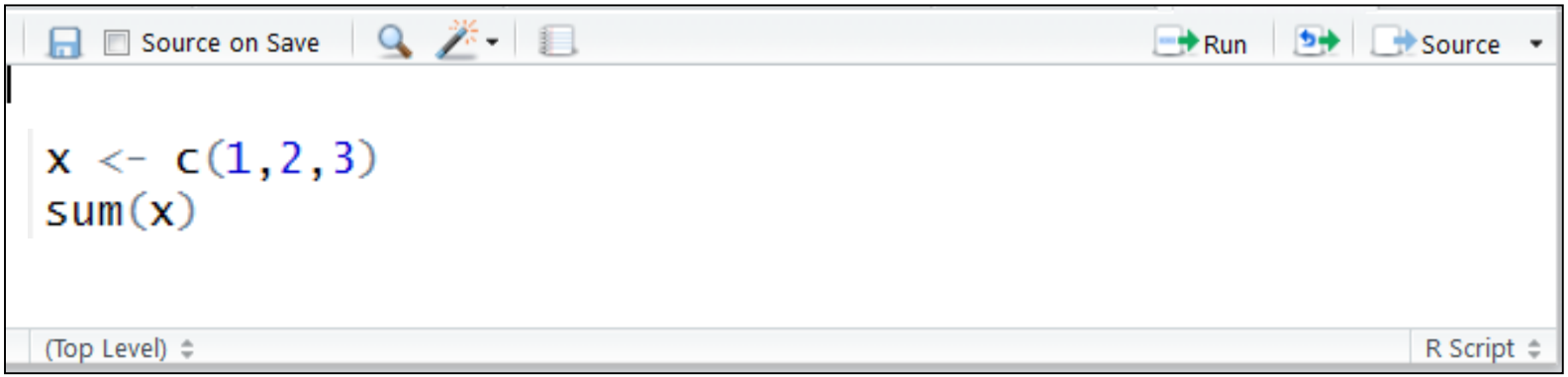
```
Console C:/APCC_TP_SDM_Dec2015/
>
> i <- example.data[,"MONTH"] == 6
> mean( example.data[i,"TMAX"] )
[1] 28.92667
>
> j <- example.data[,"MONTH"] == 12
> mean( example.data[j,"TMAX"] )
[1] 5.909677
>
```

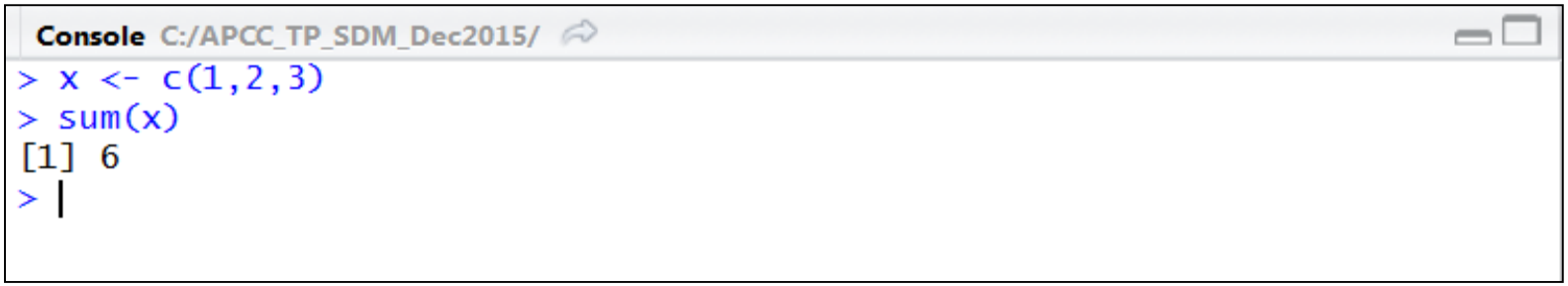
Data Analysis

- Exercise:
  - Calculate averages of TMIN for Jun. and Dec.

Data Analysis

- **sum()**





Data Analysis

- Calculating the accumulation of PRCP

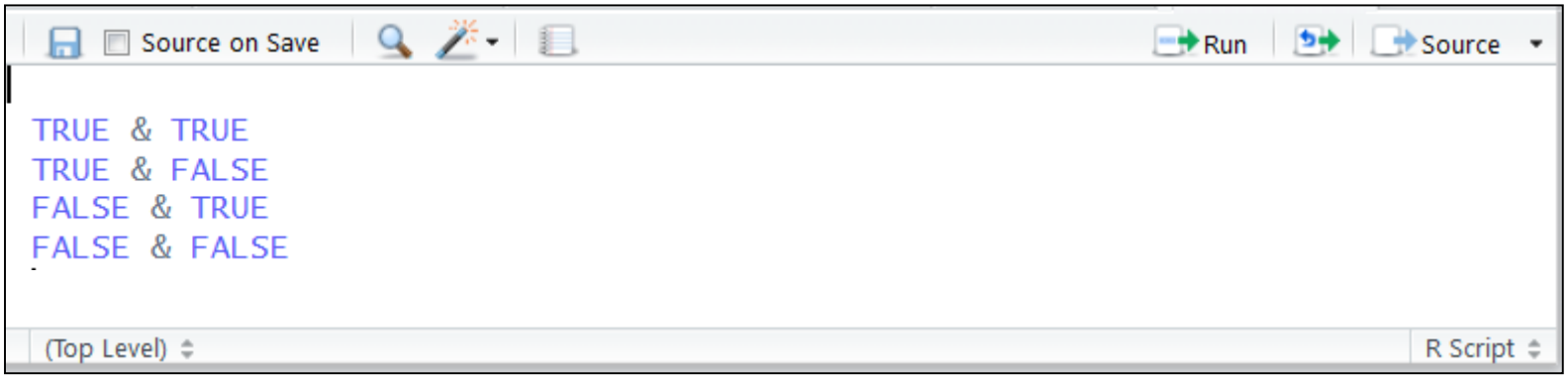
```
Source on Save                                              Run   Source
i <- example.data[,"MONTH"] == 6
sum( example.data[i,"PRCP"] )        PRCP with MONTH == 6 are chosen!

j <- example.data[,"MONTH"] == 12
sum( example.data[j,"PRCP"] )        PRCP with MONTH == 12 are chosen!
(Top Level)                                                 R Script
```

```
Console C:/APCC_TP_SDM_Dec2015/
> i <- example.data[,"MONTH"] == 6
> sum( example.data[i,"PRCP"] )
[1] 187.2
>
> j <- example.data[,"MONTH"] == 12
> sum( example.data[j,"PRCP"] )
[1] 12
>
```

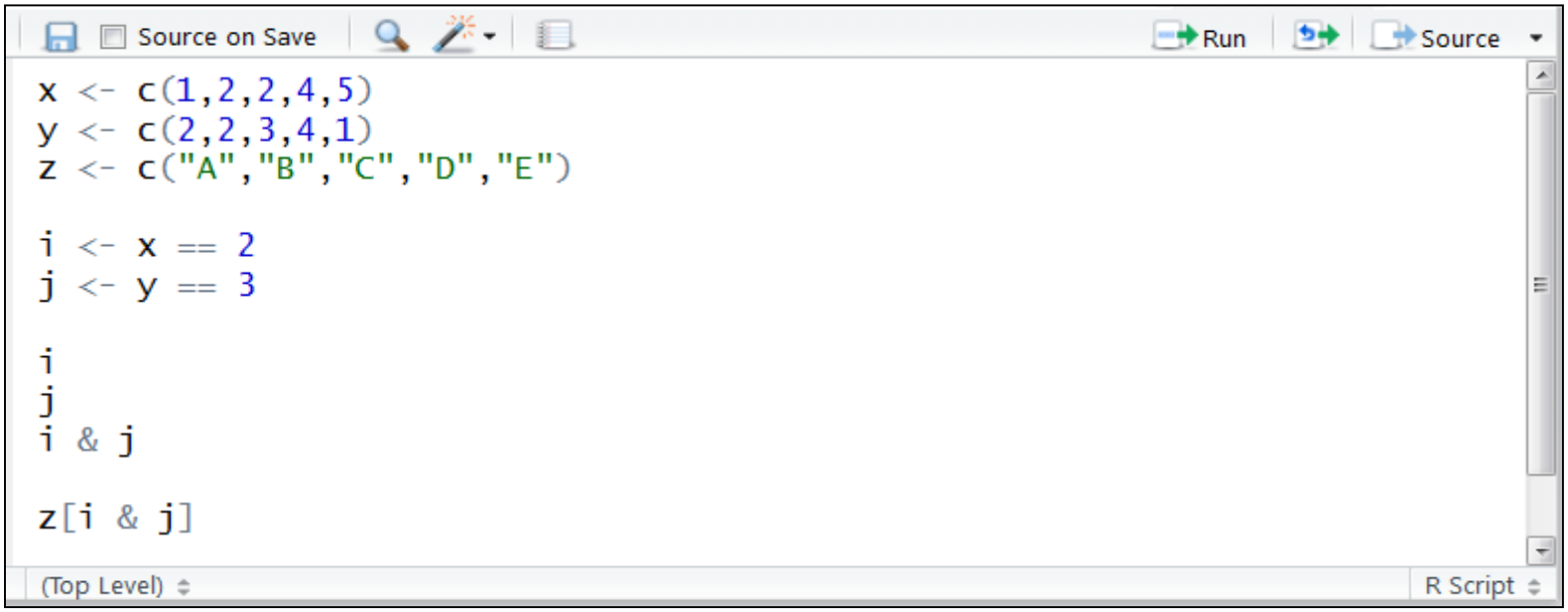
Data Analysis

- Operator '&'



Data Analysis

- Operator '&'



Data Analysis

- Operator '&'

```
Console C:/APCC_TP_SDM_Dec2015/
> x <- c(1,2,2,4,5)
> y <- c(2,2,3,4,1)
> z <- c("A","B","C","D","E")
>
> i <- x == 2
> j <- y == 3
>
> i
[1] FALSE  TRUE  TRUE FALSE FALSE
> j
[1] FALSE FALSE  TRUE FALSE FALSE
> i & j
[1] FALSE FALSE  TRUE FALSE FALSE
>
> z[i & j]
[1] "C"
>
```

Operation is done component-wisely!

The components of 'z' with x == 2 and y == 3 are chosen

Data Analysis

- Average of TMAX for wet day in June.

```
i <- example.data[,"MONTH"] == 6
j <- example.data[,"PRCP"] > 0
k <- i & j

mean( example.data[k,"TMAX"] )
```

TMAX with 'MONTH == 6 and PRCP > 0' are chosen

```
Console C:/APCC_TP_SDM_Dec2015/
> i <- example.data[,"MONTH"] == 6
> j <- example.data[,"PRCP"] > 0
> k <- i & j
>
> mean( example.data[k,"TMAX"] )
[1] 27.33
>
```

Data Analysis

- Exercise: Calculate the others:
  - Average of TMAX for dry day in June,
  - Average of TMAX for wet day in Dec.,
  - Average of TMAX for dry day in Dec.

  - The same averages of TMIN.

Data Analysis

- mean(): average
- sum(): total sum,
- sd(): standard deviation,
- max(): maximum,
- min(): minimum.

- Their usages are the same to mean().

Data Analysis

- Exercise: Calculate
  - Standard deviation of TMAX in June,
  - Standard deviation of TMAX in Dec.,
  - Minimum of TMIN in Dec.,
  - Maximum of PRCP in June.

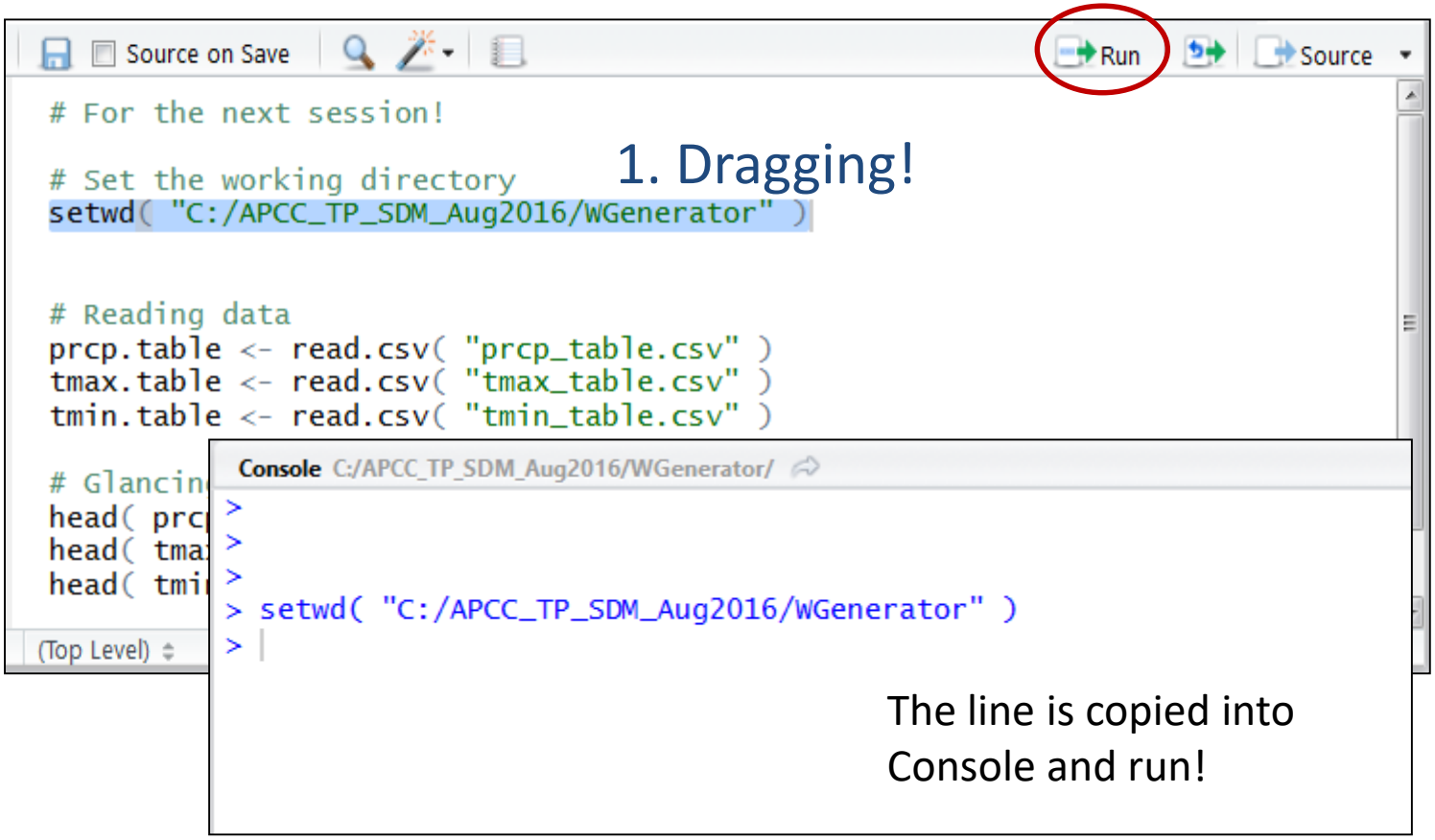
Contents

- Preparation and installation
- Variables and Operations
- Data Analysis using R
- Reading your Observation Data

Data Reading

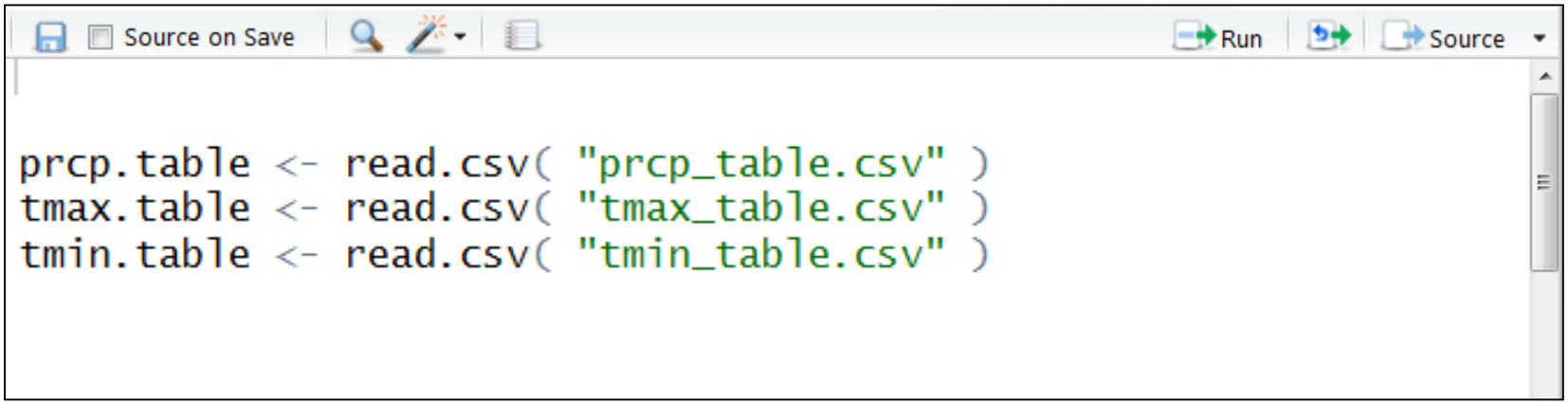
- First, set working directory.

2. Click!



Data Reading

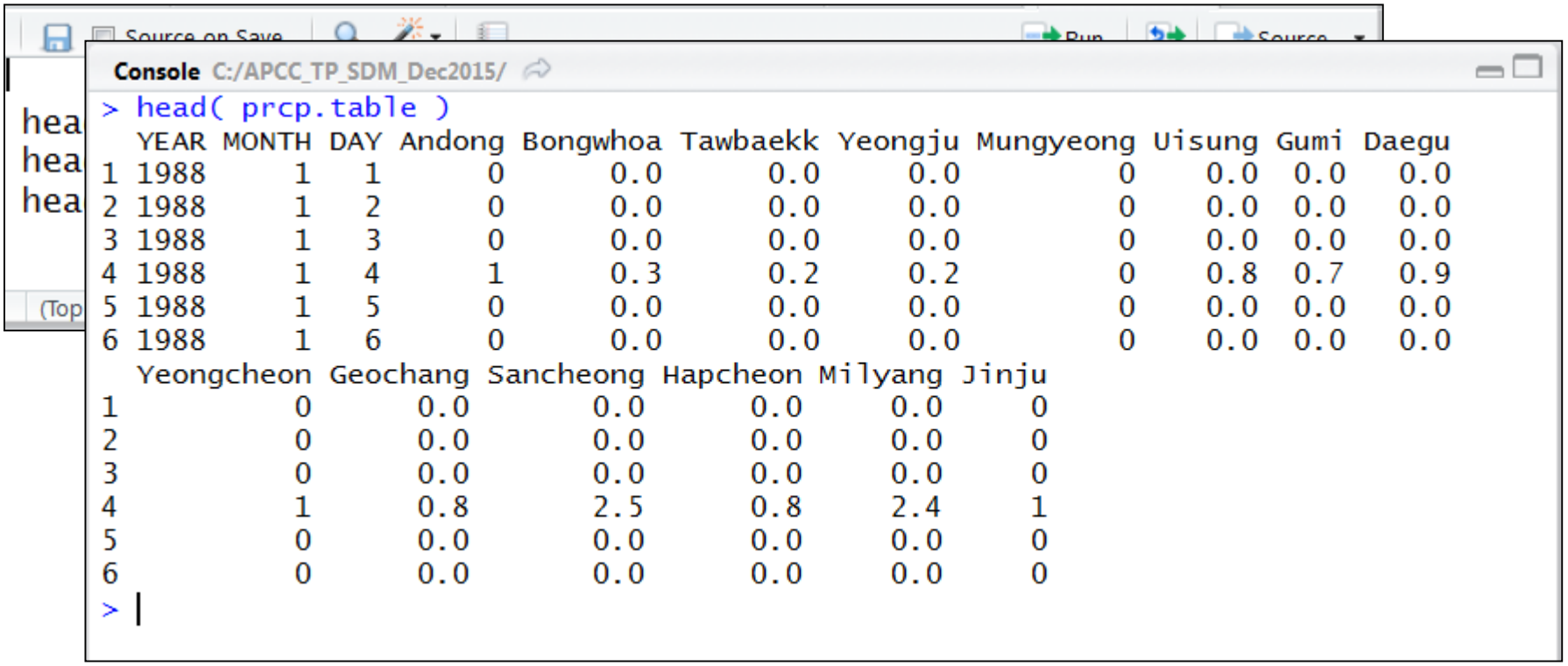
- Read data files.



  - The files have to be in the working directory!

Data Reading

- Glancing at the data.



* NOTE: The display depends on your data set.

The End of Session 1

Welcome any Questions.

APCC
APEC CLIMATE CENTER