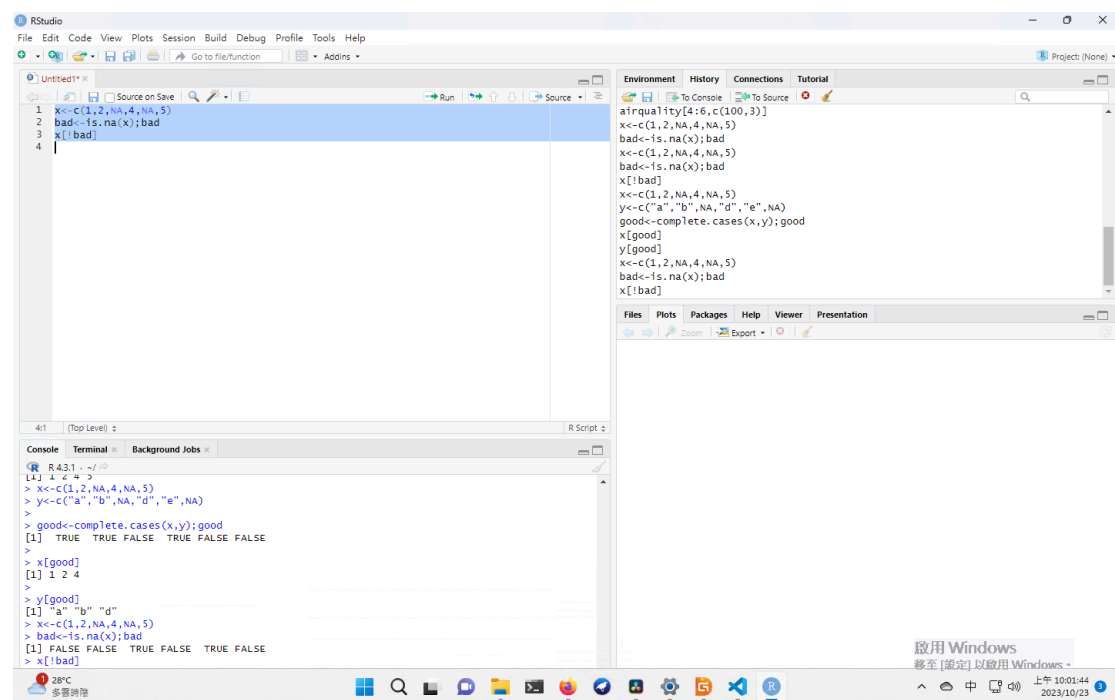


is.na()去除缺失值



The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 x<-c(1,2,NA,4,NA,5)
2 bad<-is.na(x);bad
3 x[!bad]
4
```

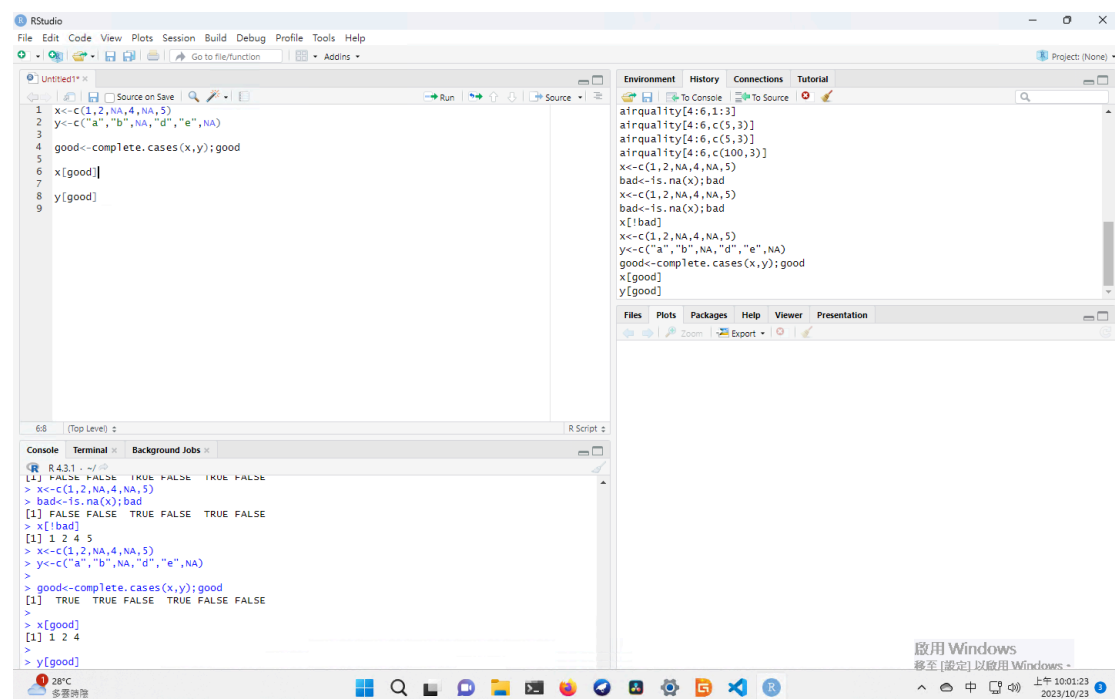
The console on the bottom left shows the execution results:

```
R 4.3.1 ~>
> x<-c(1,2,NA,4,NA,5)
> y<-c("a","b",NA,"d","e",NA)
>
> good<-complete.cases(x,y);good
[1] TRUE TRUE FALSE TRUE FALSE FALSE
> x[good]
[1] 1 2 4
> y[good]
[1] "a" "b" "d"
> x<-c(1,2,NA,4,NA,5)
> bad<-is.na(x);bad
[1] FALSE FALSE TRUE FALSE TRUE FALSE
> x[!bad]
[1] 1 2 4 5
```

The Environment pane on the right shows the objects created:

```
airquality[4:6,c(100,3)]
x<-c(1,2,NA,4,NA,5)
bad<-is.na(x);bad
x<-c(1,2,NA,4,NA,5)
bad<-is.na(x);bad
x[!bad]
x<-c(1,2,NA,4,NA,5)
y<-c("a","b",NA,"d","e",NA)
good<-complete.cases(x,y);good
x[good]
y[good]
x<-c(1,2,NA,4,NA,5)
bad<-is.na(x);bad
x[!bad]
```

complete.cases(x,y)去除缺失值



The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 x<-c(1,2,NA,4,NA,5)
2 y<-c("a","b",NA,"d","e",NA)
3
4 good<-complete.cases(x,y);good
5
6 x[good]
7
8 y[good]
9
```

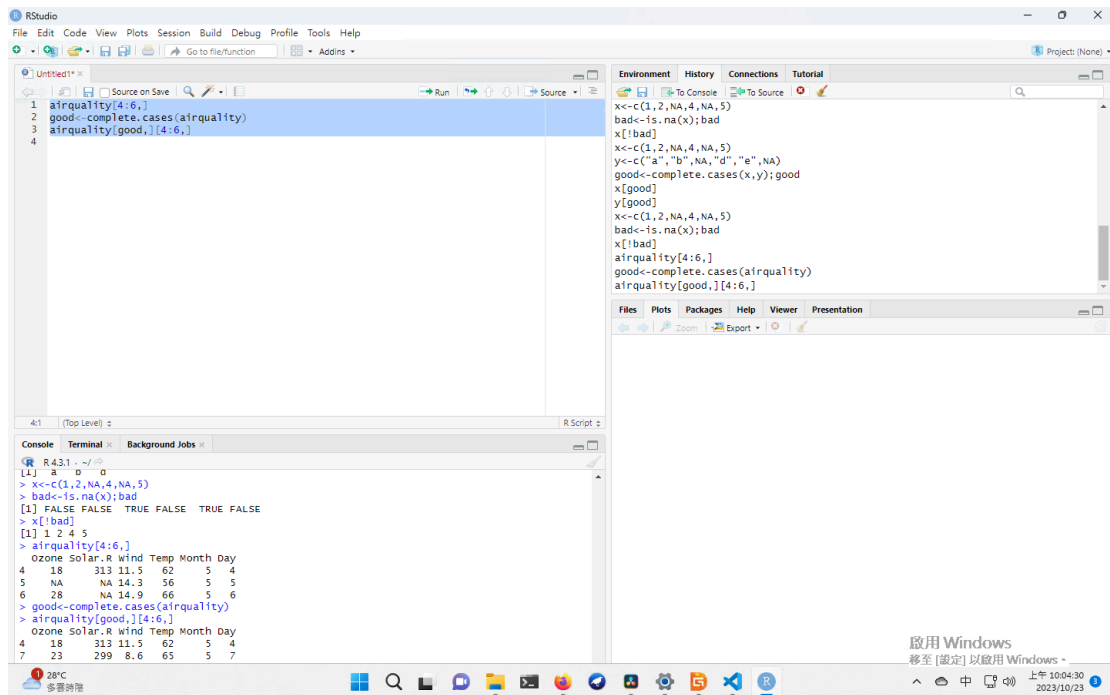
The console on the bottom left shows the execution results:

```
R 4.3.1 ~>
[1] FALSE FALSE TRUE FALSE TRUE FALSE
> x<-c(1,2,NA,4,NA,5)
> bad<-is.na(x);bad
[1] FALSE FALSE TRUE FALSE TRUE FALSE
> x[!bad]
[1] 1 2 4 5
> x<-c(1,2,NA,4,NA,5)
> y<-c("a","b",NA,"d","e",NA)
>
> good<-complete.cases(x,y);good
[1] TRUE TRUE FALSE TRUE FALSE FALSE
> x[good]
[1] 1 2 4
> y[good]
[1] "a" "b" "d"
```

The Environment pane on the right shows the objects created:

```
airquality[4:6,1:3]
airquality[4:6,c(5,3)]
airquality[4:6,c(100,3)]
x<-c(1,2,NA,4,NA,5)
bad<-is.na(x);bad
x<-c(1,2,NA,4,NA,5)
bad<-is.na(x);bad
x[!bad]
x<-c(1,2,NA,4,NA,5)
y<-c("a","b",NA,"d","e",NA)
good<-complete.cases(x,y);good
x[good]
y[good]
```

數據框去除缺失值



The screenshot shows the RStudio interface. The script editor contains the following code:

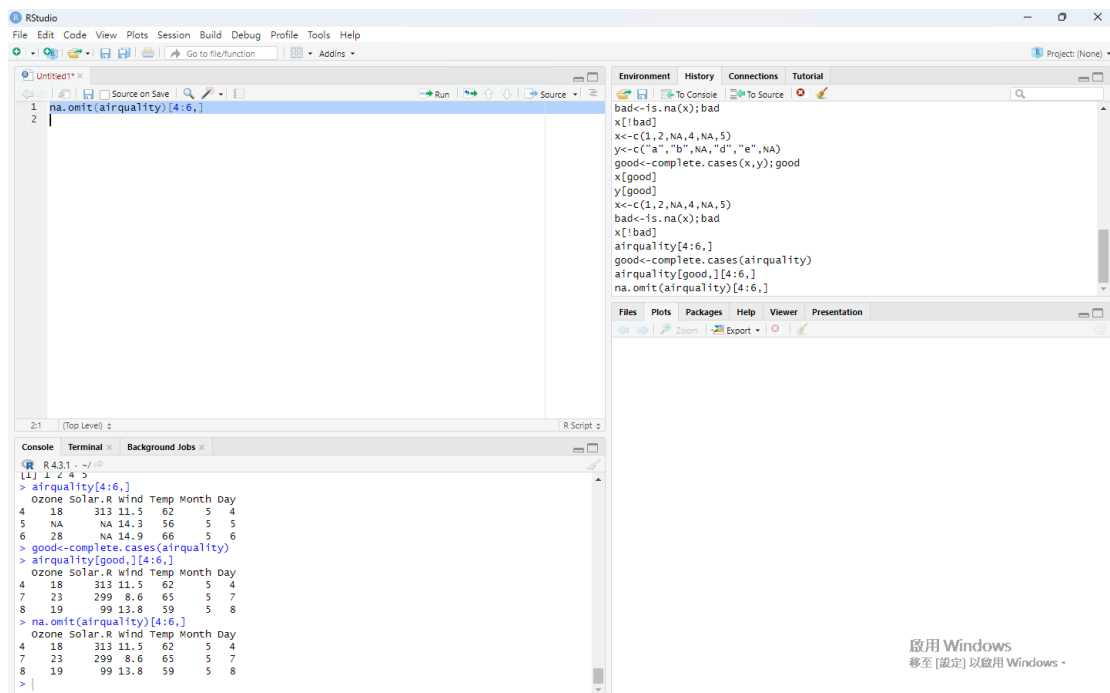
```
1 airquality[4:6,]
2 good<-complete.cases(airquality)
3 airquality[good,][4:6,]
4
```

The console output shows the execution of these commands:

```
R 4.3.1 ~\r
[1] 3 4 5
> x<-c(1,2,NA,4,NA,5)
> bad<-is.na(x);bad
[1] FALSE FALSE TRUE FALSE TRUE FALSE
> x[!bad]
[1] 1 2 4 5
> airquality[4:6,]
  Ozone Solar.R wind Temp Month Day
4    18    313  11.5   62     5    4
5     NA     NA  14.3   56     5    5
6    28     NA  14.9   66     5    6
> good<-complete.cases(airquality)
> airquality[good,][4:6,]
  Ozone Solar.R wind Temp Month Day
4    18    313  11.5   62     5    4
7    23    299   8.6   65     5    7
```

The Environment pane shows the objects created: `x`, `bad`, `good`, and `airquality[good,][4:6,]`.

另一種數據框去除缺失值



The screenshot shows the RStudio interface. The script editor contains the following code:

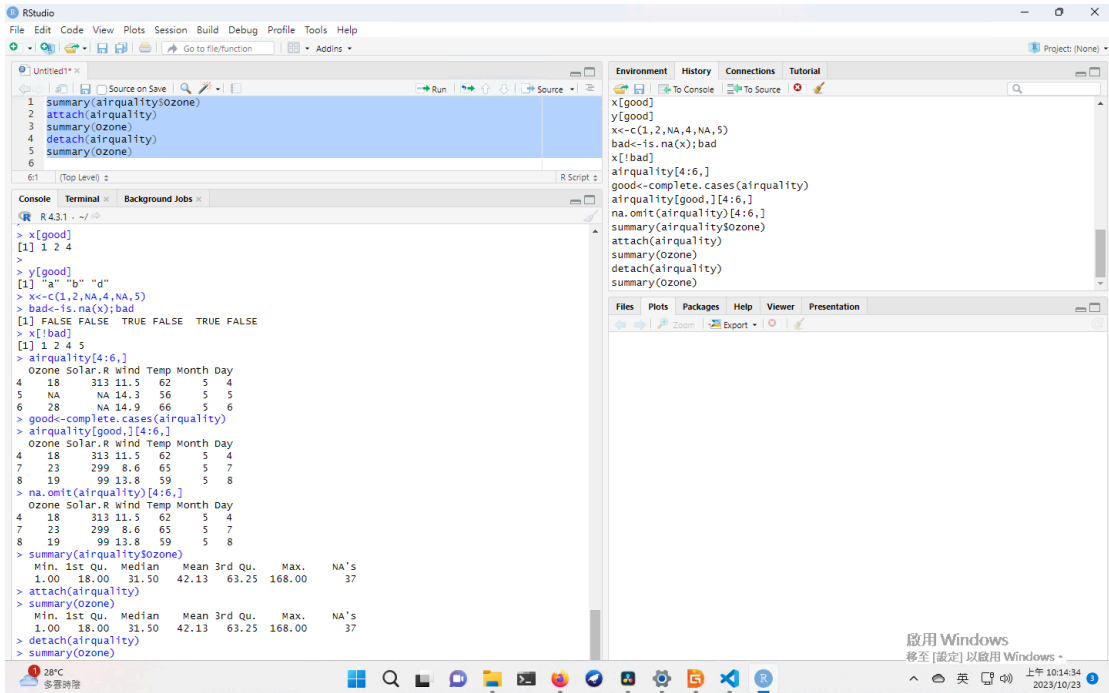
```
1 na.omit(airquality)[4:6,]
2
```

The console output shows the execution of these commands:

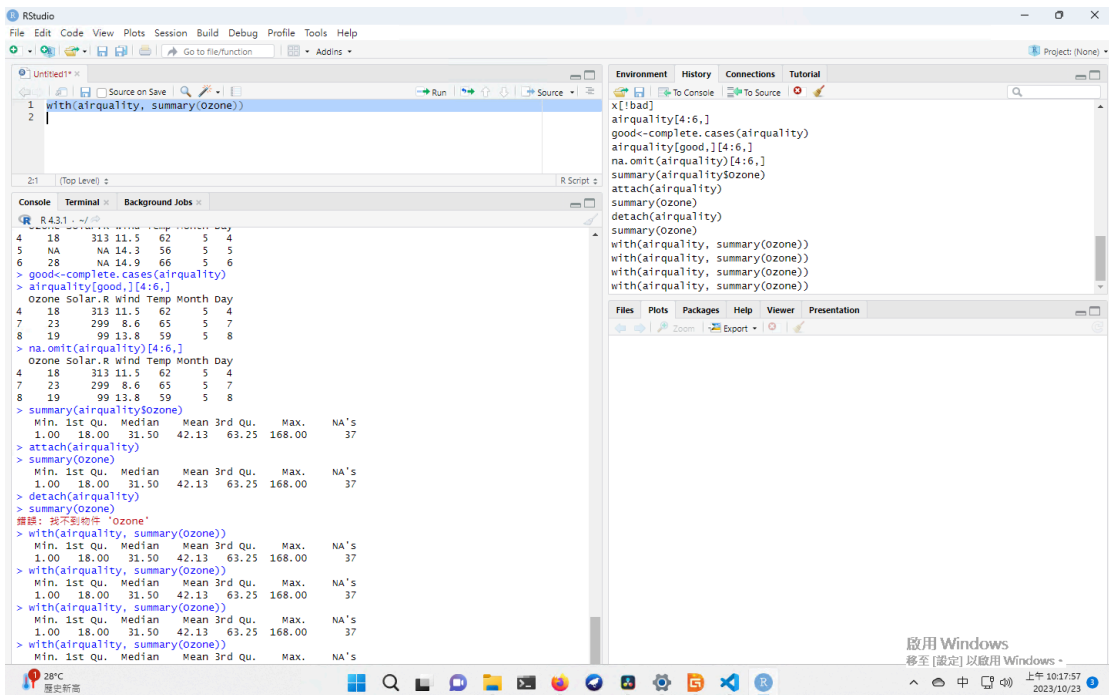
```
R 4.3.1 ~\r
[1] 1 2 4 5
> airquality[4:6,]
  Ozone Solar.R wind Temp Month Day
4    18    313  11.5   62     5    4
5     NA     NA  14.3   56     5    5
6    28     NA  14.9   66     5    6
> good<-complete.cases(airquality)
> airquality[good,][4:6,]
  Ozone Solar.R wind Temp Month Day
4    18    313  11.5   62     5    4
7    23    299   8.6   65     5    7
8    19     99  13.8   59     5    8
> na.omit(airquality)[4:6,]
  Ozone Solar.R wind Temp Month Day
4    18    313  11.5   62     5    4
7    23    299   8.6   65     5    7
8    19     99  13.8   59     5    8
>
```

The Environment pane shows the objects created: `bad`, `x`, `good`, and `na.omit(airquality)[4:6,]`.

- summary(Ozone)



with



練習

The screenshot shows the RStudio interface with a script editor on the left containing R code to load and summarize data. The console on the bottom left displays the output of the `summary(data[bad,])` command. The environment pane on the right shows the objects in the global environment.

```
1 setwd("C:/Users/eric9/downloads")
2 data = read.csv("german.csv", stringsAsFactors = TRUE)
3 summary(data)
4
5
6 summary(na.omit(data))
7
8
9 dataNA = complete.cases(data)
10 data1 = data[dataNA,]
11 summary(data1)
12
13
14
15 bad <- !is.na(data$duration)
16 summary(data[bad,])
17
```

Console Output:

```
> summary(data[bad,])
      ID      checking_status      duration      credit_history      purpose      credit_amount      savings_status      employment
Min.   : 1.0      A11:272      Min.   : 4.00      A30: 40      A43   :278      Min.   : 250      A61:601      A71: 62
1st Qu.:253.8     A12:269      1st Qu.:12.00     A31: 49      A40   :232      1st Qu.:1364     A62:103     A72:171
Median :502.5     A13: 63      Median :18.00     A32:527     A42   :180      Median :2317     A63: 63     A73:337
Mean   :502.5     A14:392      Mean   :20.86     A33: 88     A41   :103      Mean   :3267     A64: 48     A74:174
3rd Qu.:751.2     A15:107     3rd Qu.:24.00     A34:292     A49   : 96      3rd Qu.:3968     A65:181     A75:252
Max.   :1000.0     A16:177     Max.   :72.00     A35: 59     A46   : 48      Max.   :18424     A66: 48     A76:174
      installment_commitment      personal_status      other_parties      residence_since      property_magnitude      age      other_payment_plans
Min.   :1.000      A91: 50      Min.   :1.000      A101:903      Min.   :1.000      A121:282      Min.   :19.00      A141:139
1st Qu.:2.000      A92:308      1st Qu.:1.000      A102: 41      1st Qu.:2.000      A122:230      1st Qu.:27.00      A142: 47
Median :3.000      A93:546      Median :1.000      A103: 52      Median :3.000      A123:331      Median :33.00      A143:810
Mean   :2.974      A94: 92      Mean   :1.155      A104: 43      Mean   :2.84      A124:153      Mean   :35.55
3rd Qu.:4.000      A95:107     3rd Qu.:1.000      A105: 44      3rd Qu.:4.000      A125:153      3rd Qu.:42.00
Max.   :4.000      A96:107     Max.   :2.000      A106: 45      Max.   :4.000      A126:153      Max.   :75.00
      housing      existing_credits      job      num_dependents      own_telephone      foreign_worker      class
A151:177      Min.   :1.000      A171: 22      Min.   :1.000      A191:593      A201:959      Min.   :1.0
A152:712      1st Qu.:1.000      A172:199      1st Qu.:1.000      A192:403      A202: 37      1st Qu.:1.0
A153:107      Median :1.000      A173:627      Median :1.000      A193:403      A203: 37      Median :1.0
Mean   :1.408      A174:148      Mean   :1.155      A194:403      A204: 37      Mean   :1.3
3rd Qu.:2.000      A175:148      3rd Qu.:1.000      A195:403      A205: 37      3rd Qu.:1.3
Max.   :4.000      A176:148      Max.   :2.000      A196:403      A206: 37      Max.   :2.0
      NA's :2
```

Environment:

Object	Class	Attributes
data	data.frame	1000 obs. of 22 variables
data1	data.frame	994 obs. of 22 variables
bad	logical	[1:1000] TRUE TRUE ...
dataNA	logical	[1:1000] TRUE FALSE ...
x	integer	[1:4] 1 2 3 4
y	integer	[1:4] 6 7 8 9
z	integer	[1:2] 2 3

The screenshot shows the RStudio interface with a script editor on the left containing R code to filter and summarize data. The console on the bottom left displays the output of the `summary(data[bad,])` command. The environment pane on the right shows the objects in the global environment.

```
1 x <- 1:4; y <- 6:9; z <- 2:3
2
3 x+y
4
5 x+z
```

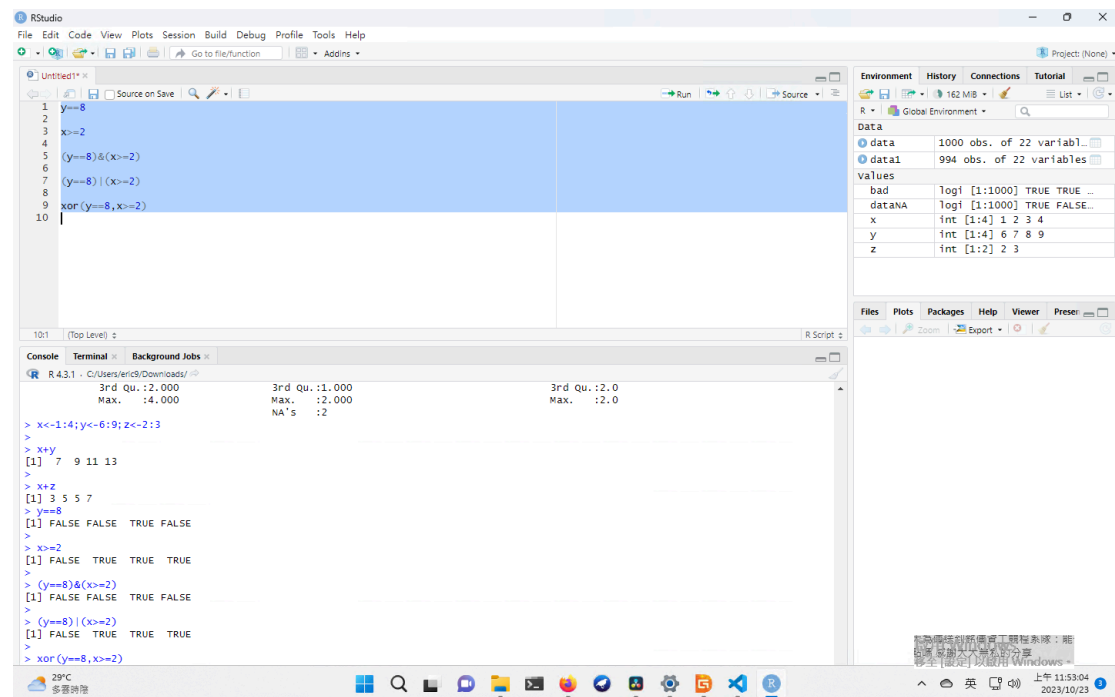
Console Output:

```
> summary(data[bad,])
      ID      checking_status      duration      credit_history      purpose      credit_amount      savings_status      employment
Min.   : 1.0      A11:272      Min.   : 4.00      A30: 40      A43   :278      Min.   : 250      A61:601      A71: 62
1st Qu.:253.8     A12:269      1st Qu.:12.00     A31: 49      A40   :232      1st Qu.:1364     A62:103     A72:171
Median :502.5     A13: 63      Median :18.00     A32:527     A42   :180      Median :2317     A63: 63     A73:337
Mean   :502.5     A14:392      Mean   :20.86     A33: 88     A41   :103      Mean   :3267     A64: 48     A74:174
3rd Qu.:751.2     A15:107     3rd Qu.:24.00     A34:292     A49   : 96      3rd Qu.:3968     A65:181     A75:252
Max.   :1000.0     A16:177     Max.   :72.00     A35: 59     A46   : 48      Max.   :18424     A66: 48     A76:174
      installment_commitment      personal_status      other_parties      residence_since      property_magnitude      age      other_payment_plans
Min.   :1.000      A91: 50      Min.   :1.000      A101:903      Min.   :1.000      A121:282      Min.   :19.00      A141:139
1st Qu.:2.000      A92:308      1st Qu.:1.000      A102: 41      1st Qu.:2.000      A122:230      1st Qu.:27.00      A142: 47
Median :3.000      A93:546      Median :1.000      A103: 52      Median :3.000      A123:331      Median :33.00      A143:810
Mean   :2.974      A94: 92      Mean   :1.155      A104: 43      Mean   :2.84      A124:153      Mean   :35.55
3rd Qu.:4.000      A95:107     3rd Qu.:1.000      A105: 44      3rd Qu.:4.000      A125:153      3rd Qu.:42.00
Max.   :4.000      A96:107     Max.   :2.000      A106: 45      Max.   :4.000      A126:153      Max.   :75.00
      housing      existing_credits      job      num_dependents      own_telephone      foreign_worker      class
A151:177      Min.   :1.000      A171: 22      Min.   :1.000      A191:593      A201:959      Min.   :1.0
A152:712      1st Qu.:1.000      A172:199      1st Qu.:1.000      A192:403      A202: 37      1st Qu.:1.0
A153:107      Median :1.000      A173:627      Median :1.000      A193:403      A203: 37      Median :1.0
Mean   :1.408      A174:148      Mean   :1.155      A194:403      A204: 37      Mean   :1.3
3rd Qu.:2.000      A175:148      3rd Qu.:1.000      A195:403      A205: 37      3rd Qu.:1.3
Max.   :4.000      A176:148      Max.   :2.000      A196:403      A206: 37      Max.   :2.0
      NA's :2
```

Environment:

Object	Class	Attributes
data	data.frame	1000 obs. of 22 variables
data1	data.frame	994 obs. of 22 variables
bad	logical	[1:1000] TRUE TRUE ...
dataNA	logical	[1:1000] TRUE FALSE ...
x	integer	[1:4] 1 2 3 4
y	integer	[1:4] 6 7 8 9
z	integer	[1:2] 2 3

向量化運算



RStudio interface showing a script with the following code:

```
1 y=8
2
3 x=2
4
5 (y==8)&(x==2)
6
7 (y==8)|(x==2)
8
9 xor(y==8,x==2)
10
```

The console output shows the results of the operations:

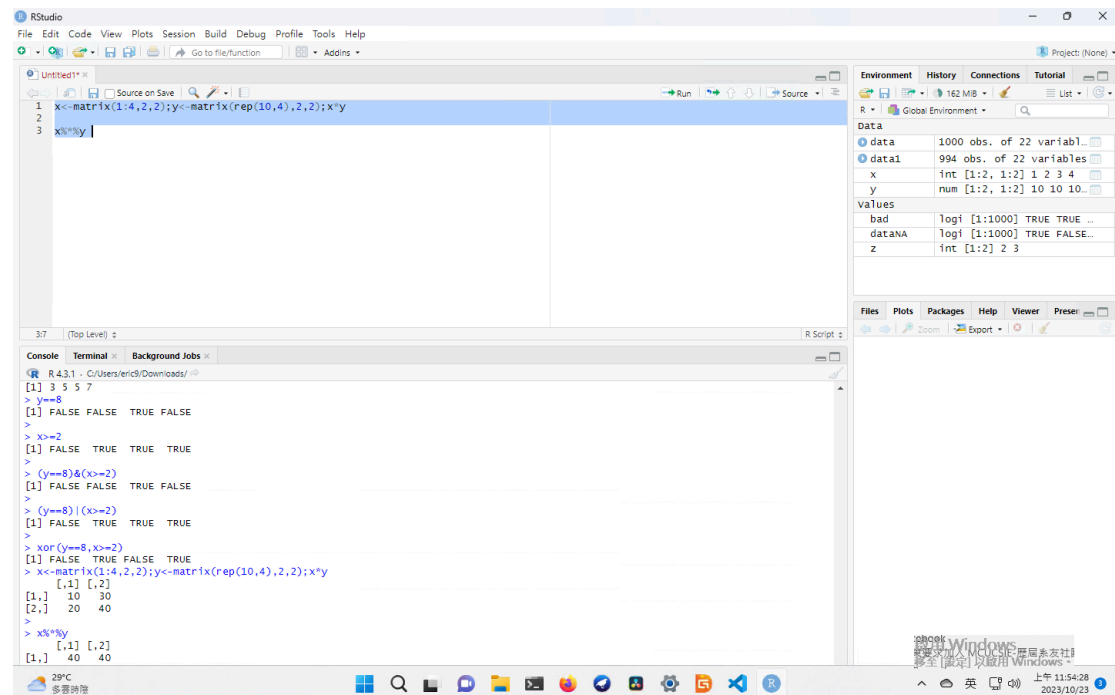
```
R 4.3.1 - C:/Users/eric/Downloads/
3rd Qu.: 2.000      3rd Qu.: 1.000      3rd Qu.: 2.0
Max.: 4.000      Max.: 2.000      Max.: 2.0
NA's      :2

> x<-1:4;y<-6:9;z<-2:3
>
> x+y
[1] 7 9 11 13
>
> x+z
[1] 3 5 5 7
> y==8
[1] FALSE FALSE TRUE FALSE
>
> x>=2
[1] FALSE TRUE TRUE TRUE
>
> (y==8)&(x>=2)
[1] FALSE FALSE TRUE FALSE
>
> (y==8)|(x>=2)
[1] FALSE TRUE TRUE TRUE
>
> xor(y==8,x>=2)
```

The Environment pane shows the following variables:

Variable	Value
data	1000 obs. of 22 variables
data1	994 obs. of 22 variables
bad	logi [1:1000] TRUE TRUE ..
dataNA	logi [1:1000] TRUE FALSE..
x	int [1:4] 1 2 3 4
y	int [1:4] 6 7 8 9
z	int [1:2] 2 3

x%*%y



RStudio interface showing a script with the following code:

```
1 x<-matrix(1:4,2,2);y<-matrix(rep(10,4),2,2);x*y
2
3 x%*%y
```

The console output shows the results of the operations:

```
R 4.3.1 - C:/Users/eric/Downloads/
[1] 3 5 5 7
> y==8
[1] FALSE FALSE TRUE FALSE
>
> x>=2
[1] FALSE TRUE TRUE TRUE
>
> (y==8)&(x>=2)
[1] FALSE FALSE TRUE FALSE
>
> (y==8)|(x>=2)
[1] FALSE TRUE TRUE TRUE
>
> xor(y==8,x>=2)
[1] FALSE TRUE FALSE TRUE
>
> x<-matrix(1:4,2,2);y<-matrix(rep(10,4),2,2);x*y
[1,] [,2]
[1,] 10 30
[2,] 20 40
>
> x%*%y
[1,] [,2]
[1,] 40 40
```

The Environment pane shows the following variables:

Variable	Value
data	1000 obs. of 22 variables
data1	994 obs. of 22 variables
x	int [1:2, 1:2] 1 2 3 4
y	num [1:2, 1:2] 10 10 10 ..
bad	logi [1:1000] TRUE TRUE ..
dataNA	logi [1:1000] TRUE FALSE..
z	int [1:2] 2 3

創建隨機數

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for generating random numbers and a matrix.
- Environment:** Shows the global environment with variables like 'data', 'x', 'y', 'bad', 'dataNA', and 'z'.
- Console:** Displays the output of the R code, including random numbers and matrix operations.

Source Editor Code:

```
1 rnorm(n, mean=0, sd=1)
2 sample(1:10, 4, replace=T)
3 sample(letters, 5, replace=F)
4 sample(1:10)
5 sample(1:10, replace=T)
6
```

Console Output:

```
R 4.3.1 - C:/Users/eric/Downloads/
[1] FALSE FALSE TRUE FALSE
> x>=2
[1] FALSE TRUE TRUE TRUE
> (y==8)&(x>=2)
[1] FALSE FALSE TRUE FALSE
> (y==8)|(x>=2)
[1] FALSE TRUE TRUE TRUE
> xor(y==8,x>=2)
[1] FALSE TRUE FALSE TRUE
> x<-matrix(1:4,2,2);y<-matrix(rep(10,4),2,2);x*y
      [,1] [,2]
[1,]   10   30
[2,]   20   40
> x%*%y
      [,1] [,2]
[1,]   40   40
[2,]   60   60
> rnorm(n, mean=0, sd=1)
```

Environment:

Variable	Value
data	1000 obs. of 22 variables
data1	994 obs. of 22 variables
x	int [1:2, 1:2] 1 2 3 4
y	num [1:2, 1:2] 10 10 10..
bad	logi [1:1000] TRUE TRUE ..
dataNA	logi [1:1000] TRUE FALSE..
z	int [1:2] 2 3

Console:

```
R 4.3.1 - C:/Users/eric/Downloads/
[1] FALSE FALSE TRUE FALSE
> x>=2
[1] FALSE TRUE TRUE TRUE
> (y==8)&(x>=2)
[1] FALSE FALSE TRUE FALSE
> (y==8)|(x>=2)
[1] FALSE TRUE TRUE TRUE
> xor(y==8,x>=2)
[1] FALSE TRUE FALSE TRUE
> x<-matrix(1:4,2,2);y<-matrix(rep(10,4),2,2);x*y
      [,1] [,2]
[1,]   10   30
[2,]   20   40
> x%*%y
      [,1] [,2]
[1,]   40   40
[2,]   60   60
> rnorm(n, mean=0, sd=1)
```