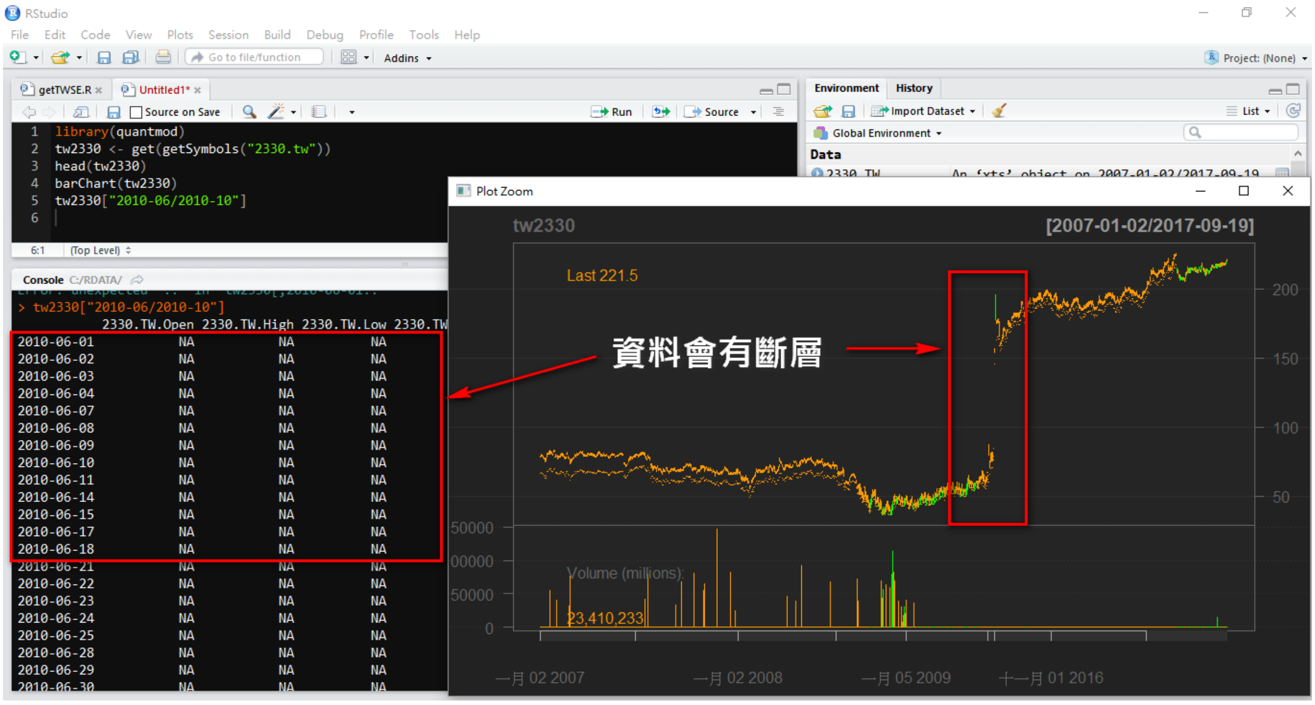
自己寫個 Function 抓證交所的股票資訊

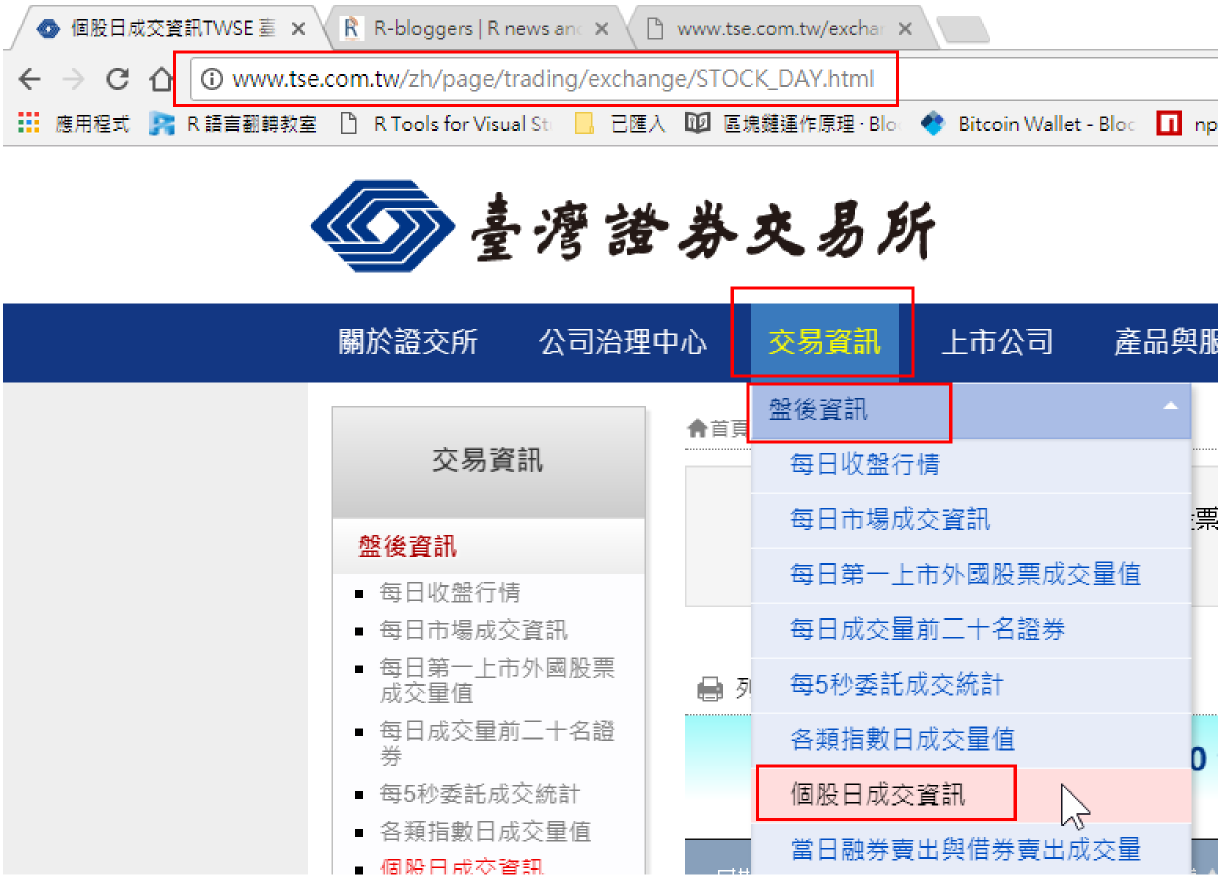
金融上的套件: quantmod，可在 Yahoo finance、Googlefinance 等網站，下載公開數據進行分析，但對於台股而言，雖然可在 Yahoo finance 抓取，但有時資料並不完整。
以台積電(2330)為例，透過 getSymbols(“2330.tw”) 抓取回來的資料，發現有缺漏。



```
library(quantmod)
tw2330 <- get(getSymbols("2330.tw"))
head(tw2330)
barChart(tw2330)
tw2330["2010-06/2010-10"]
```

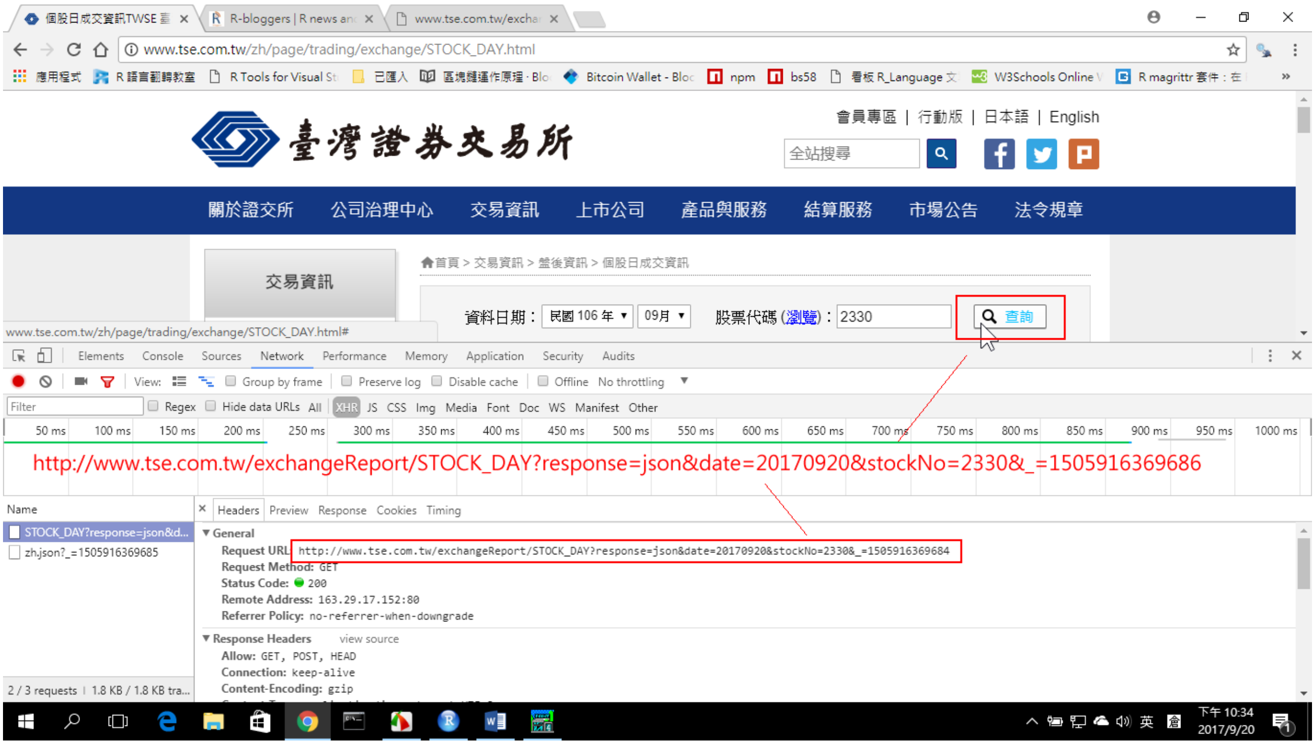
```
> tw2330["2010-06/2010-10"]
           2330.TW.Open 2330.TW.High 2330.TW.Low 2330.TW
2010-06-01           NA           NA          NA
2010-06-02           NA           NA          NA
2010-06-03           NA           NA          NA
2010-06-04           NA           NA          NA
2010-06-07           NA           NA          NA
2010-06-08           NA           NA          NA
2010-06-09           NA           NA          NA
2010-06-10           NA           NA          NA
2010-06-11           NA           NA          NA
2010-06-14           NA           NA          NA
2010-06-15           NA           NA          NA
2010-06-17           NA           NA          NA
2010-06-18           NA           NA          NA
2010-06-21           NA           NA          NA
2010-06-22           NA           NA          NA
2010-06-23           NA           NA          NA
2010-06-24           NA           NA          NA
2010-06-25           NA           NA          NA
2010-06-28           NA           NA          NA
2010-06-29           NA           NA          NA
2010-06-30           NA           NA          NA
```

因此改抓台灣的證券交易所，就是我們的目標。
網路上有許多類似的方法，可抓取證交所的股票資料，在這裏我也嚐試自己寫一個 Function
首先是利用 Google Chrome 開啟證交所網頁，點[交易資訊] -> [盤後資訊] -> [個股日成交資訊]



www.tse.com.tw/zh/page/trading/exchange/STOCK_DAY.html

按下 F12，可進入開發人員工具模式，有利我們觀查網頁的運作，以下圖為例
當按下[查詢]後，可發現網頁以 GET 方式，送出一串 URL，仔細觀查可發現 URL 中帶有個參數，其中 respons=json，代表回傳格式為 JSON，date=20170920，代表查詢日期，實測結果發現，日期的最後 2 碼並不影響查詢結果，即 date=20170901 與 date=20170920 查回來的結果是相同的，稍後的自訂函數中，天數統一用 01，stockNo=2330 為股票代號，最後一個參數 _=1505916369686，為時間戳記，隨時間變動。


www.tse.com.tw/zh/page/trading/exchange/STOCK_DAY.html
臺灣證券交易所
關於證交所 公司治理中心 交易資訊 上市公司 產品與服務 結算服務 市場公告 法令規章
交易資訊
首頁 > 交易資訊 > 盤後資訊 > 個股日成交資訊
資料日期：民國 106 年 09月
股票代碼 (瀏覽)：2330
查詢
Elements Console Sources Network Performance Memory Application Security Audits
Headers Preview Response Cookies Timing

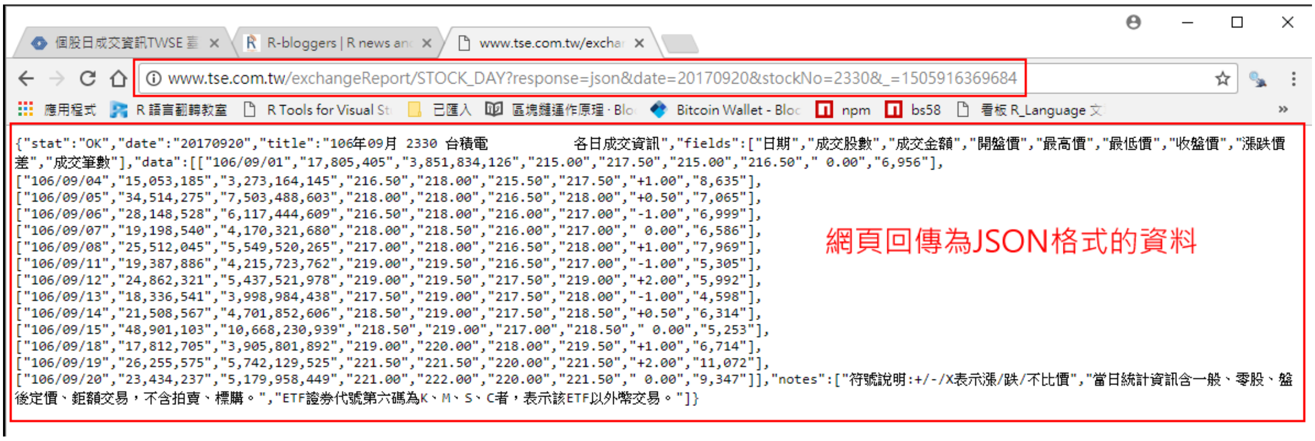
http://www.tse.com.tw/exchangeReport/STOCK_DAY?response=json&date=20170920&stockNo=2330&_=1505916369686

```
Name
STOCK_DAY?response=json&d...
zh.json?_=1505916369685

▼ General
  Request URL: http://www.tse.com.tw/exchangeReport/STOCK_DAY?response=json&date=20170920&stockNo=2330&_=1505916369684
  Request Method: GET
  Status Code: ● 200
  Remote Address: 163.29.17.152:80
  Referrer Policy: no-referrer-when-downgrade
▼ Response Headers    view source
  Allow: GET, POST, HEAD
  Connection: keep-alive
  Content-Encoding: gzip

2 / 3 requests | 1.8 KB / 1.8 KB tra...
```

回傳 JSON



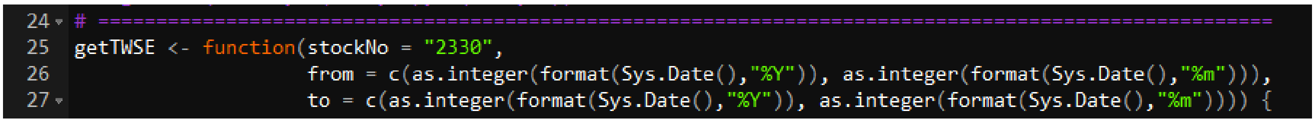
```
{"stat":"OK","date":"20170920","title":"106年09月 2330 台積電          各日成交資訊","fields":["日期","成交股數","成交金額","開盤價","最高價","最低價","收盤價","漲跌價差","成交筆數"],"data":[["106/09/01","17,805,405","3,851,834,126","215.00","217.50","215.00","216.50"," 0.00","6,956"],
["106/09/04","15,053,185","3,273,164,145","216.50","218.00","215.50","217.50","+1.00","8,635"],
["106/09/05","34,514,275","7,503,488,603","218.00","218.00","216.50","218.00","+0.50","7,065"],
["106/09/06","28,148,528","6,117,444,609","216.50","218.00","216.00","217.00","-1.00","6,999"],
["106/09/07","19,198,540","4,170,321,680","218.00","218.50","216.00","217.00"," 0.00","6,586"],
["106/09/08","25,512,045","5,549,520,265","217.00","218.00","216.50","218.00","+1.00","7,969"],
["106/09/11","19,387,886","4,215,723,762","219.00","219.50","216.50","217.00","-1.00","5,305"],
["106/09/12","24,862,321","5,437,521,978","219.00","219.50","217.50","219.00","+2.00","5,992"],
["106/09/13","18,336,541","3,998,984,438","217.50","219.00","217.50","218.00","-1.00","4,598"],
["106/09/14","21,508,567","4,701,852,606","218.50","219.00","217.50","218.50","+0.50","6,314"],
["106/09/15","48,901,103","10,668,230,939","218.50","219.00","217.00","218.50"," 0.00","5,253"],
["106/09/18","17,812,705","3,905,801,892","219.00","220.00","218.00","219.50","+1.00","6,714"],
["106/09/19","26,255,575","5,742,129,525","221.50","221.50","220.00","221.50","+2.00","11,072"],
["106/09/20","23,434,237","5,179,958,449","221.00","222.00","220.00","221.50"," 0.00","9,347"]],"notes":["符號說明:+/-/X表示漲/跌/不比價","當日統計資訊含一般、零股、盤後定價、鉅額交易，不含拍賣、標購。","ETF證券代號第六碼為K、M、S、C者，表示該ETF以外幣交易。"]}
```

網頁回傳為JSON格式的資料

有了上述的基本觀念，將有利於 Function 的寫作。

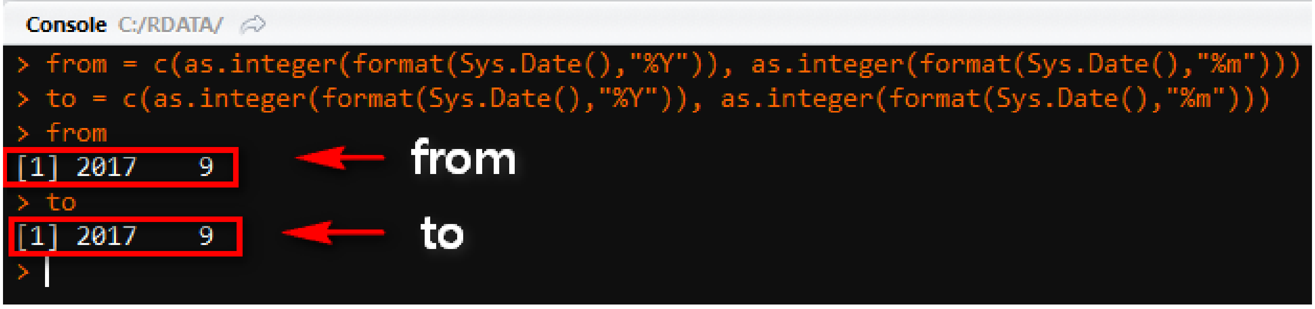
函數：getTWSE(stockNo, from = c(year, month), to = c(year, month))，預設有 3 個參數，皆帶有預設值；當呼叫函數而未帶參數時，系統會用預設值當查詢條件。

stockNo = “2330”，預設為台積電(2330)



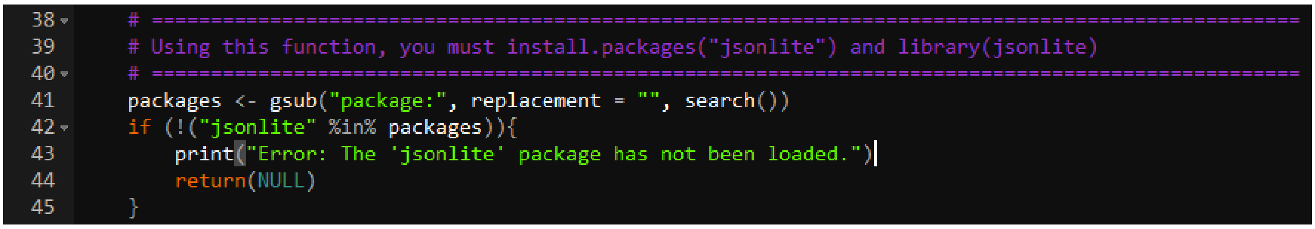
```r
# ==========================================================================================
getTWSE <- function(stockNo = "2330",
                    from = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m"))),
                    to = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m")))) {
```

From、to，若未輸入，預設為查詢當下年月，如下圖



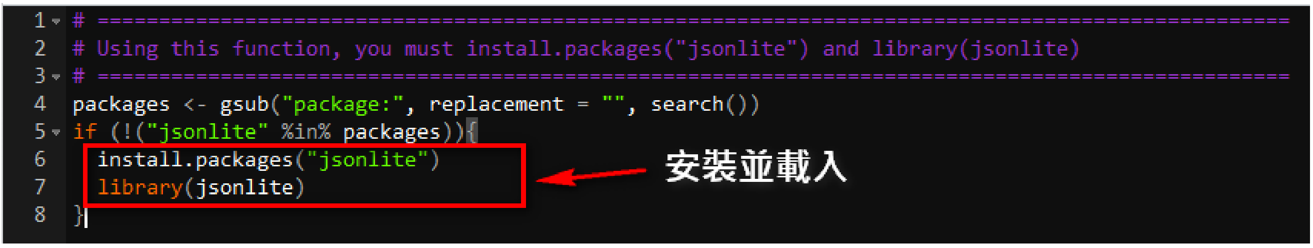
```
> from = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m")))
> to = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m")))
> from
[1] 2017    9
> to
[1] 2017    9
> 
```

行 38~45，主要是檢查 RStudio 是否已載入 jsonlite 套件，未載入則印出訊息並中斷 Function，也可以改寫成未載入，則自動安裝載入套件



```r
38      # ======================================================================================
39      # Using this function, you must install.packages("jsonlite") and library(jsonlite)
40      # ======================================================================================
41      packages <- gsub("package:", replacement = "", search())
42      if (!("jsonlite" %in% packages)){
43          print("Error: The 'jsonlite' package has not been loaded.")
44          return(NULL)
45      }
```

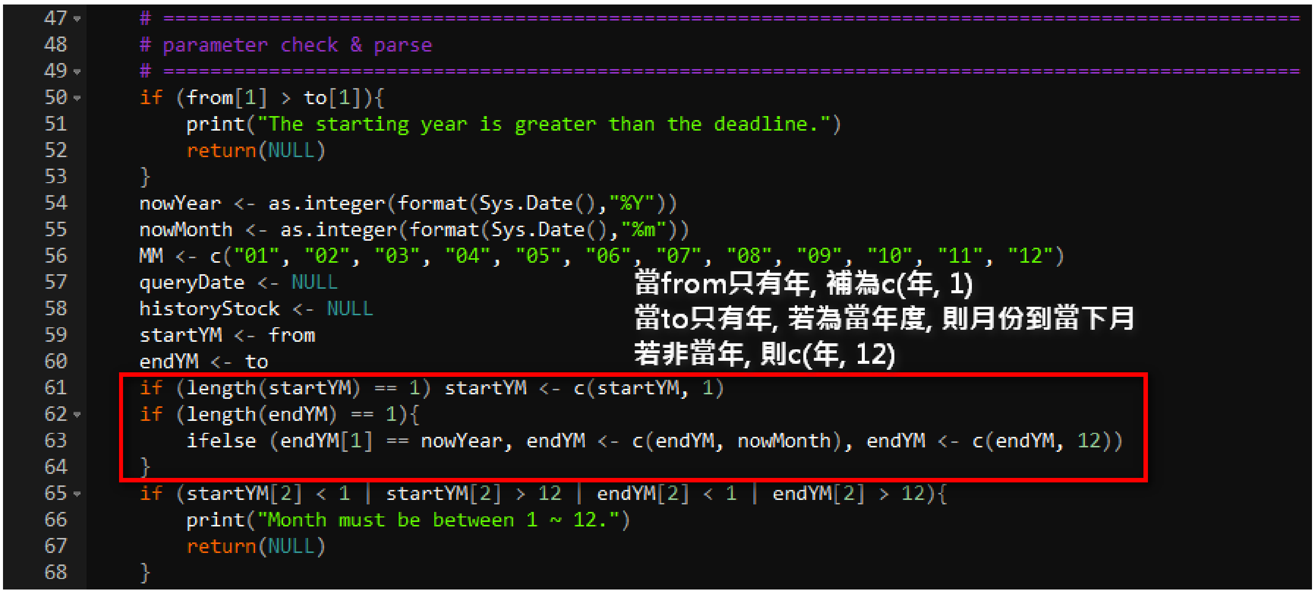
改為偵測到未載入則進行安裝及載入動作



```r
# ==========================================================================================
# Using this function, you must install.packages("jsonlite") and library(jsonlite)
# ==========================================================================================
packages <- gsub("package:", replacement = "", search())
if (!("jsonlite" %in% packages)){
  install.packages("jsonlite")
  library(jsonlite)
}
```

安裝並載入

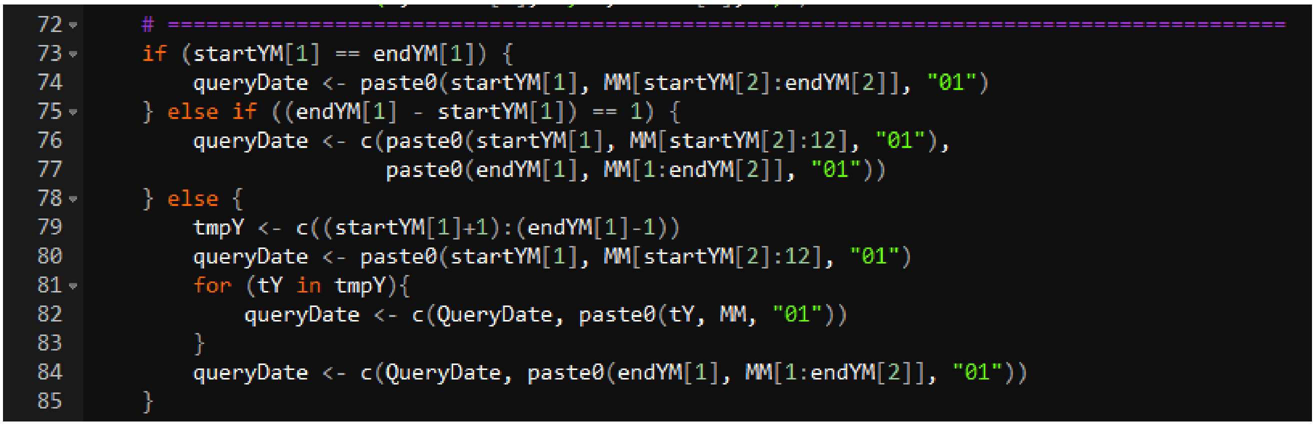
行 61~64，判斷 from 及 to 的內容，若僅有年度，則補上月份



```r
47      # ======================================================================================
48      # parameter check & parse
49      # ======================================================================================
50      if (from[1] > to[1]){
51          print("The starting year is greater than the deadline.")
52          return(NULL)
53      }
54      nowYear <- as.integer(format(Sys.Date(),"%Y"))
55      nowMonth <- as.integer(format(Sys.Date(),"%m"))
56      MM <- c("01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12")
57      queryDate <- NULL
58      historyStock <- NULL
59      startYM <- from
60      endYM <- to
61      if (length(startYM) == 1) startYM <- c(startYM, 1)
62      if (length(endYM) == 1){
63          ifelse (endYM[1] == nowYear, endYM <- c(endYM, nowMonth), endYM <- c(endYM, 12))
64      }
65      if (startYM[2] < 1 | startYM[2] > 12 | endYM[2] < 1 | endYM[2] > 12){
66          print("Month must be between 1 ~ 12.")
67          return(NULL)
68      }
```

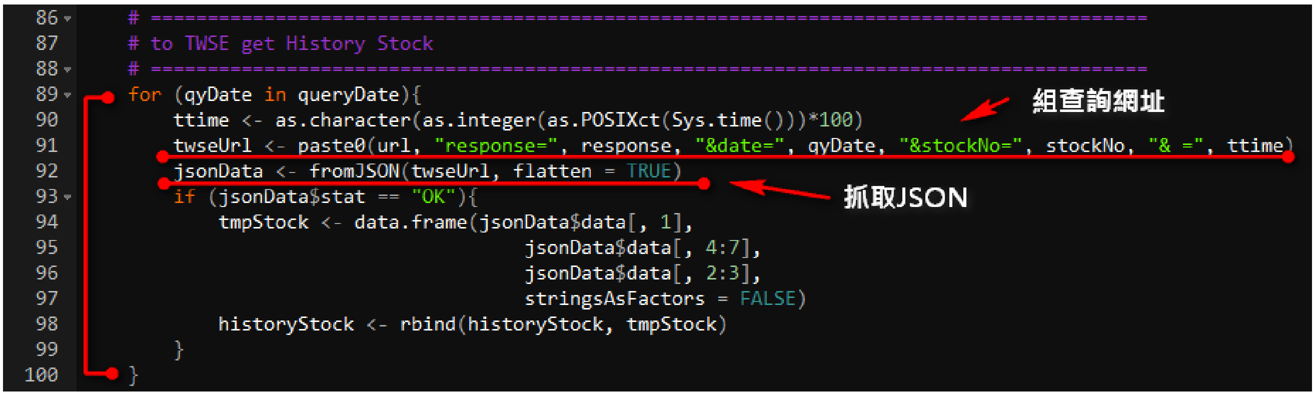
當from只有年, 補為c(年, 1)
當to只有年, 若為當年度, 則月份到當下月
若非當年, 則c(年, 12)

行 73~85，計算出查詢區間的明細，待會組 URL 時，參數 date 的內容



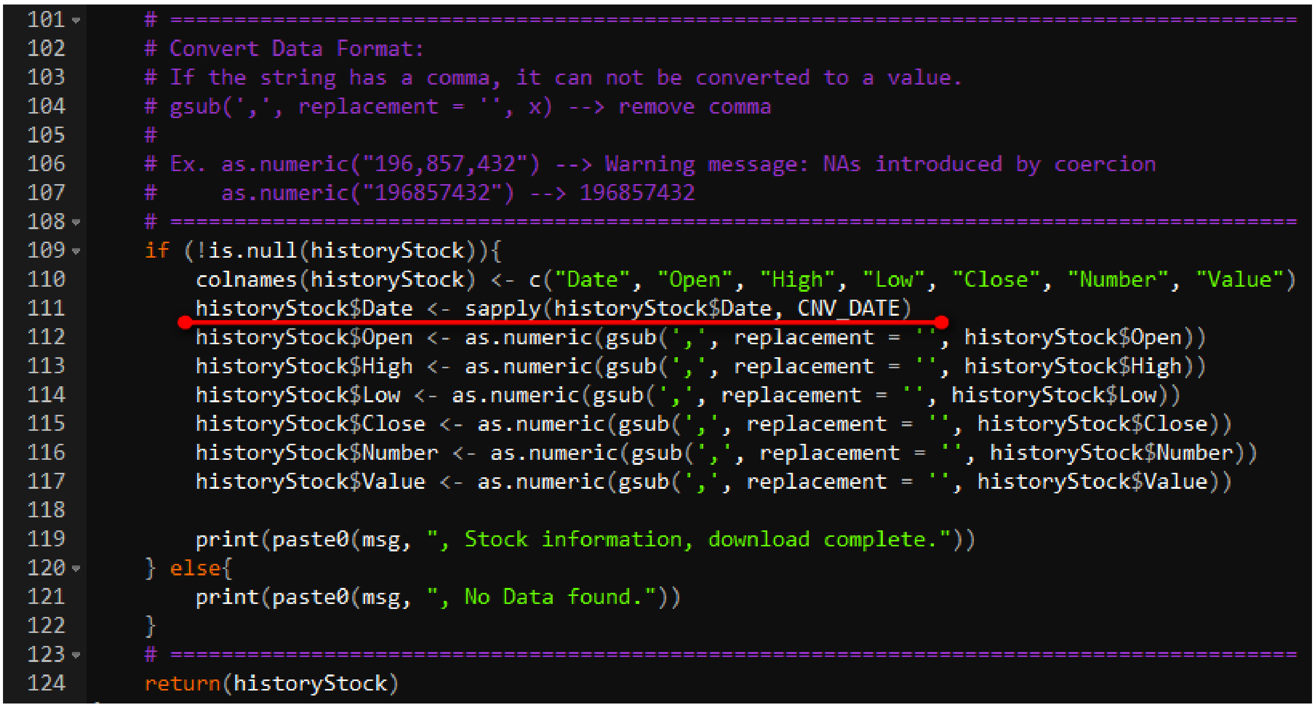
```
72    # ==========================================================================================
73    if (startYM[1] == endYM[1]) {
74        queryDate <- paste0(startYM[1], MM[startYM[2]:endYM[2]], "01")
75    } else if ((endYM[1] - startYM[1]) == 1) {
76        queryDate <- c(paste0(startYM[1], MM[startYM[2]:12], "01"),
77                       paste0(endYM[1], MM[1:endYM[2]], "01"))
78    } else {
79        tmpY <- c((startYM[1]+1):(endYM[1]-1))
80        queryDate <- paste0(startYM[1], MM[startYM[2]:12], "01")
81        for (tY in tmpY){
82            queryDate <- c(QueryDate, paste0(tY, MM, "01"))
83        }
84        queryDate <- c(QueryDate, paste0(endYM[1], MM[1:endYM[2]], "01"))
85    }
```

行 91，組查詢網址，傳入 4 個參數；行 92，真正到證交所抓資料

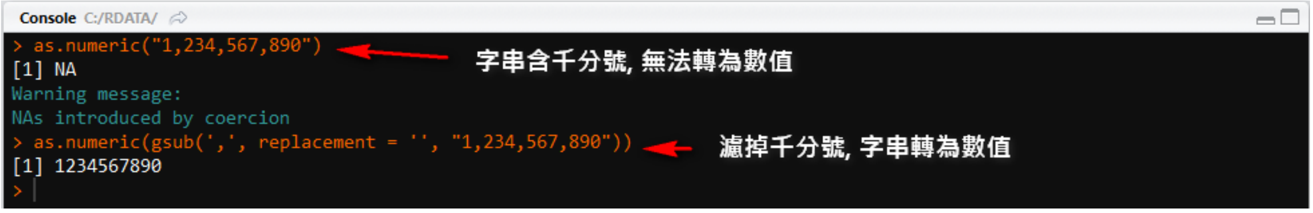


```
86    # ================================================================================
87    # to TWSE get History Stock
88    # ================================================================================
89    for (qyDate in queryDate){
90        ttime <- as.character(as.integer(as.POSIXct(Sys.time()))*100)
91        twseUrl <- paste0(url, "response=", response, "&date=", qyDate, "&stockNo=", stockNo, "& =", ttime)
92        jsonData <- fromJSON(twseUrl, flatten = TRUE)
93        if (jsonData$stat == "OK"){
94            tmpStock <- data.frame(jsonData$data[, 1],
95                                   jsonData$data[, 4:7],
96                                   jsonData$data[, 2:3],
97                                   stringsAsFactors = FALSE)
98            historyStock <- rbind(historyStock, tmpStock)
99        }
100   }
```

行 109~117，資料格式轉換，行 111，利用自訂函數 CNV_DATE()，透過 sapply 函數將民國年轉為西元年，行 112~117，則是將 string 轉為 number，因為含千分號(,)，故先要濾掉

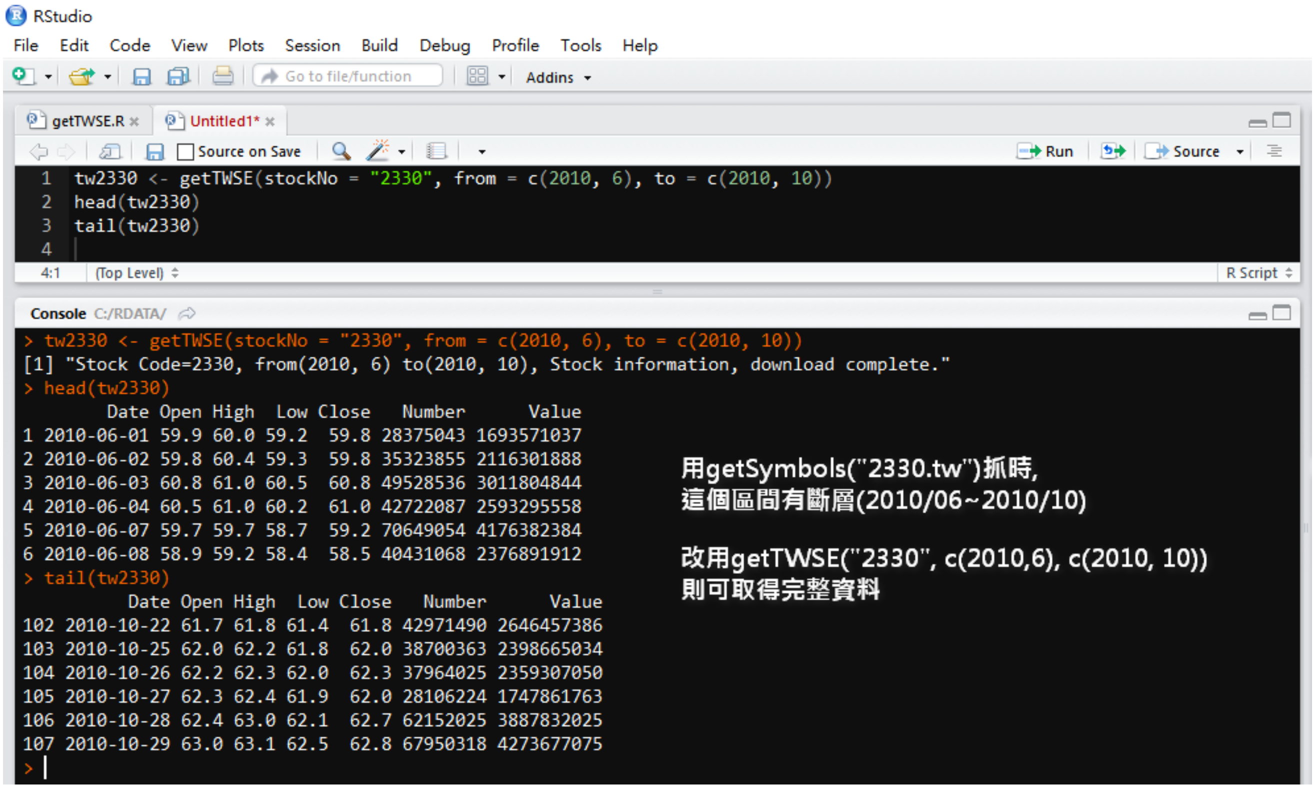


```
101   # ==========================================================================================
102   # Convert Data Format:
103   # If the string has a comma, it can not be converted to a value.
104   # gsub(',', replacement = '', x) --> remove comma
105   #
106   # Ex. as.numeric("196,857,432") --> Warning message: NAs introduced by coercion
107   #     as.numeric("196857432") --> 196857432
108   # ==========================================================================================
109   if (!is.null(historyStock)){
110       colnames(historyStock) <- c("Date", "Open", "High", "Low", "Close", "Number", "Value")
111       historyStock$Date <- sapply(historyStock$Date, CNV_DATE)
112       historyStock$Open <- as.numeric(gsub(',', replacement = '', historyStock$Open))
113       historyStock$High <- as.numeric(gsub(',', replacement = '', historyStock$High))
114       historyStock$Low <- as.numeric(gsub(',', replacement = '', historyStock$Low))
115       historyStock$Close <- as.numeric(gsub(',', replacement = '', historyStock$Close))
116       historyStock$Number <- as.numeric(gsub(',', replacement = '', historyStock$Number))
117       historyStock$Value <- as.numeric(gsub(',', replacement = '', historyStock$Value))
118
119       print(paste0(msg, ", Stock information, download complete."))
120   } else{
121       print(paste0(msg, ", No Data found."))
122   }
123   # ==========================================================================================
124   return(historyStock)
```



```
> as.numeric("1,234,567,890")
[1] NA
Warning message:
NAs introduced by coercion
> as.numeric(gsub(',', replacement = '', "1,234,567,890"))
[1] 1234567890
>
```

驗測結果如下圖，原來用 getSymbols 抓取為斷層的部份，改用 getTWSE 可正常取得資料。



```r
tw2330 <- getTWSE(stockNo = "2330", from = c(2010, 6), to = c(2010, 10))
head(tw2330)
tail(tw2330)
```

```
> tw2330 <- getTWSE(stockNo = "2330", from = c(2010, 6), to = c(2010, 10))
[1] "Stock Code=2330, from(2010, 6) to(2010, 10), Stock information, download complete."
> head(tw2330)
        Date Open High  Low Close   Number      Value
1 2010-06-01 59.9 60.0 59.2  59.8 28375043 1693571037
2 2010-06-02 59.8 60.4 59.3  59.8 35323855 2116301888
3 2010-06-03 60.8 61.0 60.5  60.8 49528536 3011804844
4 2010-06-04 60.5 61.0 60.2  61.0 42722087 2593295558
5 2010-06-07 59.7 59.7 58.7  59.2 70649054 4176382384
6 2010-06-08 58.9 59.2 58.4  58.5 40431068 2376891912
> tail(tw2330)
          Date Open High  Low Close   Number      Value
102 2010-10-22 61.7 61.8 61.4  61.8 42971490 2646457386
103 2010-10-25 62.0 62.2 61.8  62.0 38700363 2398665034
104 2010-10-26 62.2 62.3 62.0  62.3 37964025 2359307050
105 2010-10-27 62.3 62.4 61.9  62.0 28106224 1747861763
106 2010-10-28 62.4 63.0 62.1  62.7 62152025 3887832025
107 2010-10-29 63.0 63.1 62.5  62.8 67950318 4273677075
>
```

**用getSymbols("2330.tw")抓時,
這個區間有斷層(2010/06~2010/10)**

**改用getTWSE("2330", c(2010,6), c(2010, 10))
則可取得完整資料**

完整程式，請至 GitHub 下載

GitHub: https://github.com/dong945/R

完整程式碼

```
# +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# Author: TUNG-SHENG, CHEN
# DATE:       2017/09/20
# Package: jsonlite
# +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# install.packages("jsonlite")
library(jsonlite)
# ====================================================================================================
# Parameter:
#       stockNo: String, Stock code of TWSE, default: 2330 (TSMC)
#       from:         Vector, c(year, month), default: now
#       to:             Vector, c(year, month), default: now
# Examples:
#       ## return '2330', this month
#       getTWSE()
#
#       ## return '2330', from 2016/01 ~ now
#       getTWSE("2330", 2016)
#
#       ## return '2330', from 2015/01 ~ 2016/12
#       getTWSE("2330", 2015, 2016)
#
#       ## return '2330', from 2015/06 ~ 2017/03
#       getTWSE("2330", c(2105, 6), c(2017, 3))
# ====================================================================================================
getTWSE <- function(stockNo = "2330",
                        from = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m"))),
                        to = c(as.integer(format(Sys.Date(),"%Y")), as.integer(format(Sys.Date(),"%m")))) {
# ==============================================================================================================================
# Inside Function: Change Date to yyyy-mm-dd
# ==============================================================================================================================
CNV_DATE <- function(x){
      TMP <- strsplit(x, split = "/")
      paste(as.integer(TMP[[1]][1])+1911, TMP[[1]][2], TMP[[1]][3], sep = "-")
}
# ==============================================================================================================================
response <- "json"
url <- "http://www.tse.com.tw/exchangeReport/STOCK_DAY?"
# ==============================================================================================================================
# Using this function, you must install.packages("jsonlite") and library(jsonlite)
# ==============================================================================================================================
packages <- gsub("package:", replacement = "", search())
if (!("jsonlite" %in% packages)){
      print("Error: The 'jsonlite' package has not been loaded.")
      return(NULL)
}
# ==============================================================================================================================
# parameter check & parse
# ==============================================================================================================================
if (from[1] > to[1]){
      print("The starting year is greater than the deadline.")
      return(NULL)
}
```

```
nowYear <- as.integer(format(Sys.Date(),"%Y"))
nowMonth <- as.integer(format(Sys.Date(),"%m"))
MM <- c("01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12")
queryDate <- NULL
historyStock <- NULL
startYM <- from
endYM <- to
if (length(startYM) == 1) startYM <- c(startYM, 1)
if (length(endYM) == 1){
	ifelse (endYM[1] == nowYear, endYM <- c(endYM, nowMonth), endYM <- c(endYM, 12))
}
if (startYM[2] < 1 | startYM[2] > 12 | endYM[2] < 1 | endYM[2] > 12){
	print("Month must be between 1 ~ 12.")
	return(NULL)
}
msg <- paste0("Stock Code=", stockNo,
			", from(", startYM[1], ", ", startYM[2], ")",
			" to(", endYM[1], ", ", endYM[2], ")")
# ===============================================================================================================
if (startYM[1] == endYM[1]) {
	queryDate <- paste0(startYM[1], MM[startYM[2]:endYM[2]], "01")
} else if ((endYM[1] - startYM[1]) == 1) {
	queryDate <- c(paste0(startYM[1], MM[startYM[2]:12], "01"),
				paste0(endYM[1], MM[1:endYM[2]], "01"))
} else {
	tmpY <- c((startYM[1]+1):(endYM[1]-1))
	queryDate <- paste0(startYM[1], MM[startYM[2]:12], "01")
```

```
        for (tY in tmpY){
                queryDate <- c(QueryDate, paste0(tY, MM, "01"))
        }
        queryDate <- c(QueryDate, paste0(endYM[1], MM[1:endYM[2]], "01"))
}
# ==========================================================================================================
# to TWSE get History Stock
# ==========================================================================================================
for (qyDate in queryDate){
        ttime <- as.character(as.integer(as.POSIXct(Sys.time()))*100)
        twseUrl <- paste0(url, "response=", response, "&date=", qyDate, "&stockNo=", stockNo, "&_=", ttime)
        jsonData <- fromJSON(twseUrl, flatten = TRUE)
        if (jsonData$stat == "OK"){
                tmpStock <- data.frame(jsonData$data[, 1],
                                                        jsonData$data[, 4:7],
                                                        jsonData$data[, 2:3],
                                                        stringsAsFactors = FALSE)
                historyStock <- rbind(historyStock, tmpStock)
        }
}
# ==========================================================================================================
# Convert Data Format:
# If the string has a comma, it can not be converted to a value.
# gsub(',', replacement = '', x) --> remove comma
#
# Ex. as.numeric("196,857,432") --> Warning message: NAs introduced by coercion
#          as.numeric("196857432") --> 196857432
```

```
    # ==========================================================================================
    if (!is.null(historyStock)){
        colnames(historyStock) <- c("Date", "Open", "High", "Low", "Close", "Number", "Value")
        historyStock$Date <- sapply(historyStock$Date, CNV_DATE)
        historyStock$Open <- as.numeric(gsub(',', replacement = '', historyStock$Open))
        historyStock$High <- as.numeric(gsub(',', replacement = '', historyStock$High))
        historyStock$Low <- as.numeric(gsub(',', replacement = '', historyStock$Low))
        historyStock$Close <- as.numeric(gsub(',', replacement = '', historyStock$Close))
        historyStock$Number <- as.numeric(gsub(',', replacement = '', historyStock$Number))
        historyStock$Value <- as.numeric(gsub(',', replacement = '', historyStock$Value))

        print(paste0(msg, ", Stock information, download complete."))
    } else{
        print(paste0(msg, ", No Data found."))
    }
    # ==========================================================================================
    return(historyStock)
}
```