



# 使用 R 語言進行 資料分析

講師：蔡芸琚 (Pecu)  
臉書社團：PecuClub

# 課程大綱

1. 相關套件介紹
2. 機器學習
  - 監督式學習之應用、非監督式學習之應用
  - 演化式學習之應用、混合式學習之應用、關聯性規則之應用
3. 社群網路分析與應用
4. 交易策略分析與應用
5. 資料視覺化分析應用

# 專案主題

交易策略分析

電影票房分析

商品銷售預測分析

會計資料分析

醫療病例分析

基因相關分析

生物資訊相關分析

數位人文資料探勘

# 暖身運動

安裝 R 語言

認識 R 語言

套件 Quantmod 介紹

大數據是什麼？

R 語言在大數據中的角色

# 安裝 R 語言

## Download R 3.3.1

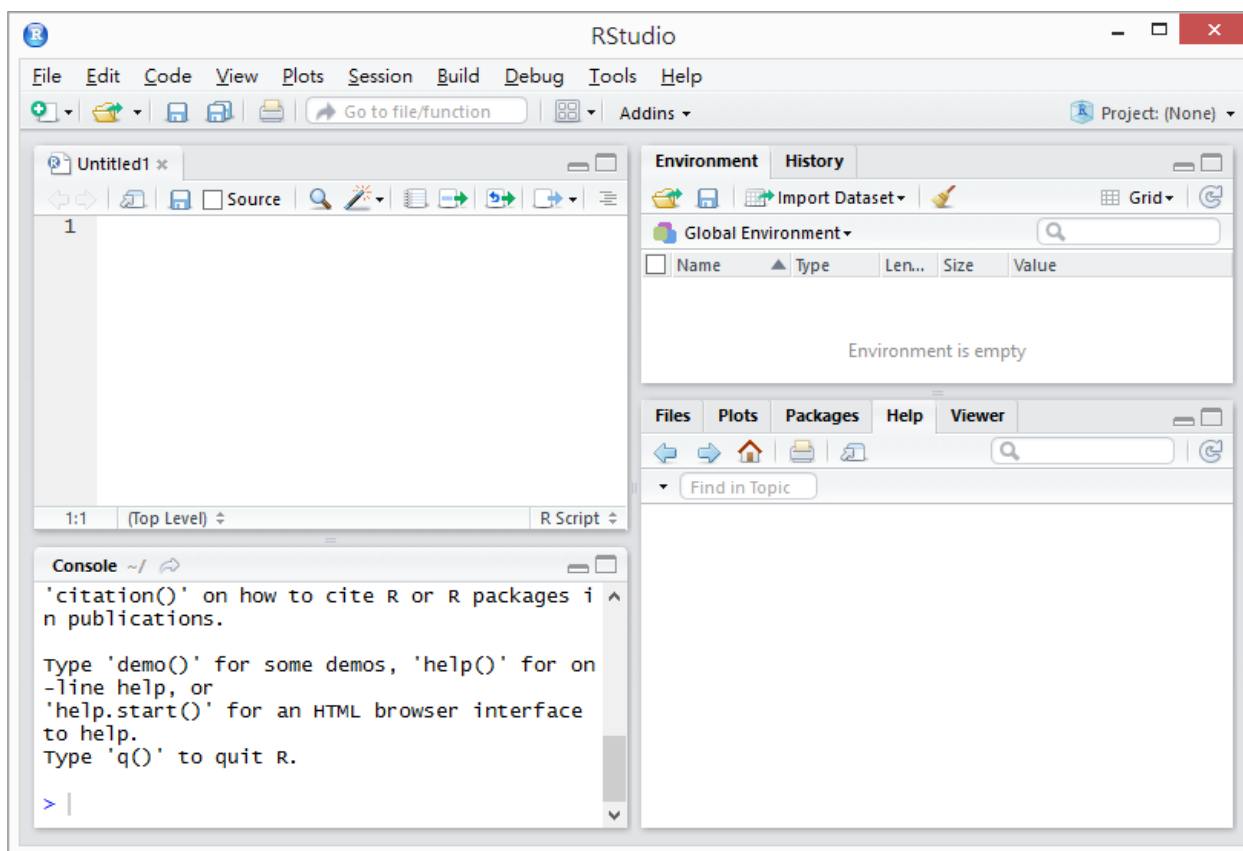
- For Windows  
<http://lib.stat.cmu.edu/R/CRAN/bin/windows/>
- For Mac OS X  
<http://lib.stat.cmu.edu/R/CRAN/bin/macosx/>
- For Linux  
<http://lib.stat.cmu.edu/R/CRAN/bin/linux/>

## Download Rstudio

- <https://www.rstudio.com/products/rstudio/download/>



# 安裝 R 語言



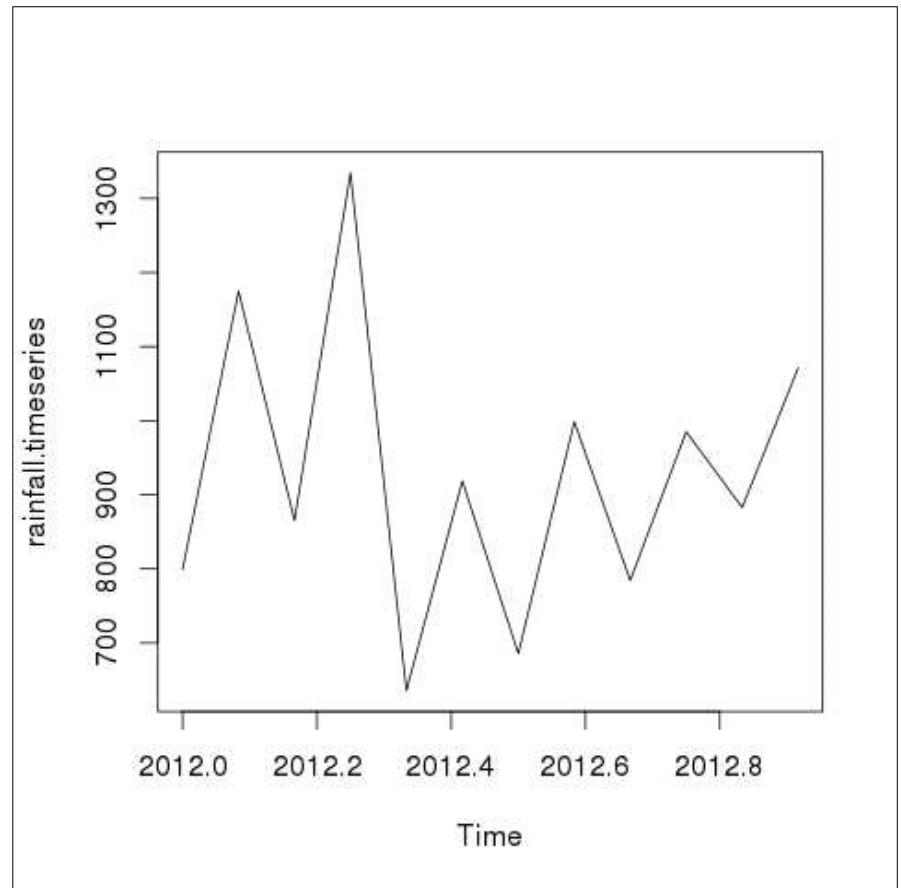
# 認識 R 語言

## R tutorials:

- <http://www.tutorialspoint.com/r/>
- <https://www.codeschool.com/courses/try-r>

## Time Series Analysis:

[http://www.tutorialspoint.com/r/r\\_time\\_series\\_analysis.htm](http://www.tutorialspoint.com/r/r_time_series_analysis.htm)



# 認識 R 語言

The screenshot displays the RStudio interface with three main panels:

- Script Editor:** Contains an R script for financial data analysis using the `quantmod` package. The script includes steps for loading the package, fetching AAPL stock data, calculating moving averages, and adding technical indicators.
- Console:** Shows the execution of the script commands. It displays the output of `library("quantmod")`, `getSymbols`, `barChart`, `chartSeries`, and `addTA`. The final command is `help addTA`, which results in an error: "Error: unexpected symbol in 'help addTA'".
- Help Window:** A search for `addTA` has been performed, displaying the documentation for the `addTA` function. The documentation includes a description of the function's purpose and its usage with various arguments.

```
1 rm(list=ls(all=TRUE))
2 library("quantmod")
3 getSymbols("AAPL",src="yahoo")
4 barChart(AAPL)
5 chartSeries(AAPL["2012-01::2012-06"],theme="white")
6 ma_20<-runMean(AAPL[,4],n=20)
7 addTA(ma_20,on=1,col="blue")
8 ma_60<-runMean(AAPL[,4],n=60)
9 addTA(ma_60,on=1,col="red")
```

**Console Output:**

```
> library("quantmod")
> getSymbols("AAPL",src="yahoo")
[1] "AAPL"
> barChart(AAPL)
> chartSeries(AAPL["2012-01::2012-06"],theme="white")
> ma_20<-runMean(AAPL[,4],n=20)
> addTA(ma_20,on=1,col="blue")
> ma_60<-runMean(AAPL[,4],n=60)
> addTA(ma_60,on=1,col="red")
> help addTA
Error: unexpected symbol in "help addTA"
> |
```

**Help Window Content:**

**Create A New TA Indicator For chartSeries**

**Description**

Functions to assist in the creation of indicators or content to be drawn on plots produced by chartSeries.

**Usage**

```
addTA(ta,
      order = NULL,
      on = NA,
      legend = "auto",
      yrange = NULL,
      ...)

newTA(FUN,
      preFUN,
      postFUN,
      on = NA,
      yrange = NULL,
      legend.name,
      fdots = TRUE,
      cdots = TRUE,
      data.at = 1,
```



# R 套件 QUANTMOD 介紹

Quantmod 套件: <http://www.quantmod.com/>

安裝 Quantmod 套件: `> install.packages('quantmod')`

取得 Apple 股價資料

```
> library("quantmod")  
> getSymbols("AAPL",src="yahoo")  
> barChart(AAPL)
```



# R 套件 QUANTMOD 介紹

畫出K線圖

```
> chartSeries(AAPL["2012-01::2012-06"],theme="white")
```

畫出 20 日平均線 (月線)

```
> ma_20<-runMean(AAPL[,4],n=20)
```

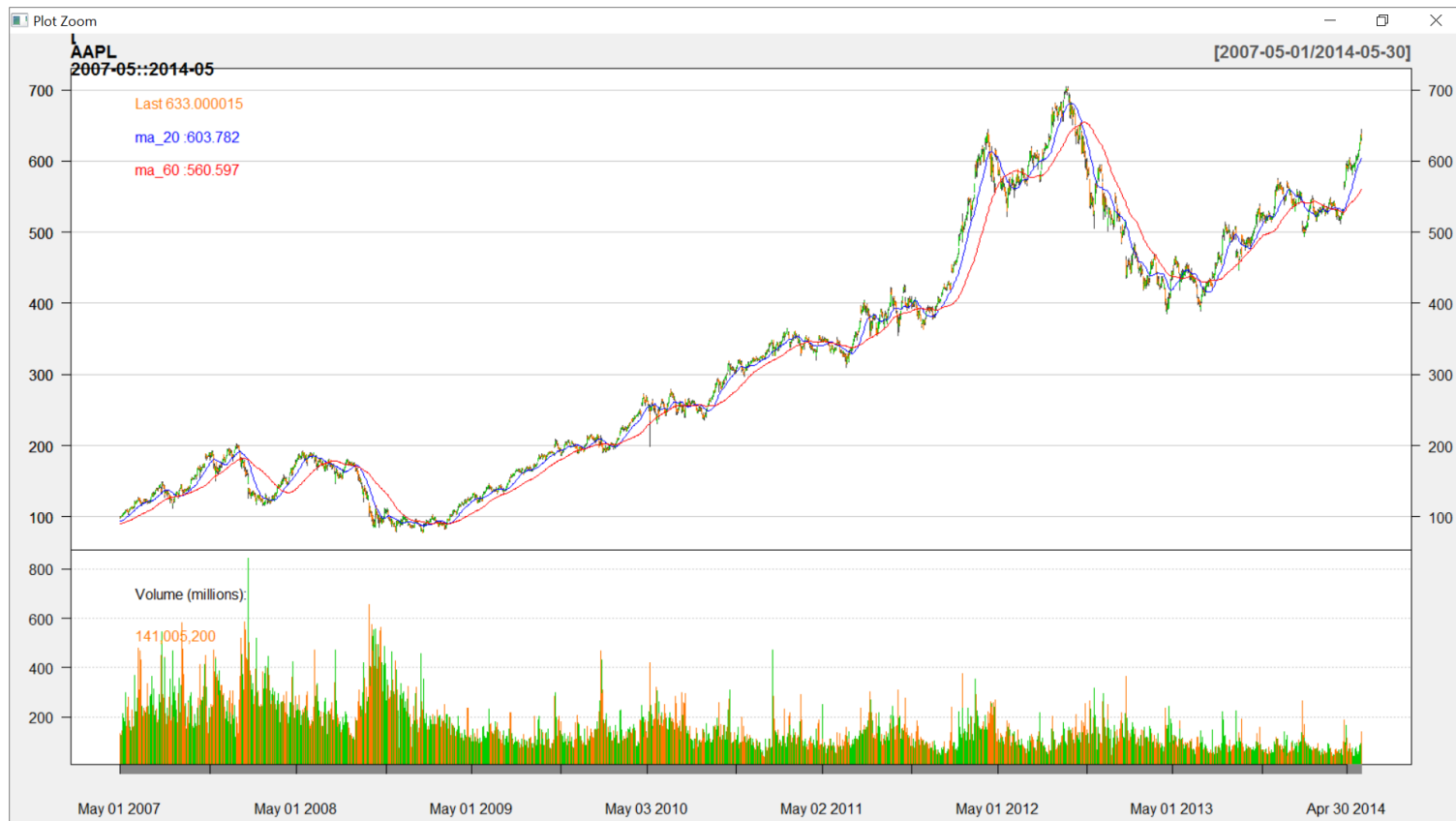
```
> addTA(ma_20,on=1,col="blue")
```

畫出 60 日平均線 (季線)

```
> ma_60<-runMean(AAPL[,4],n=60)
```

```
> addTA(ma_60,on=1,col="red")
```

# R 套件 QUANTMOD 介紹



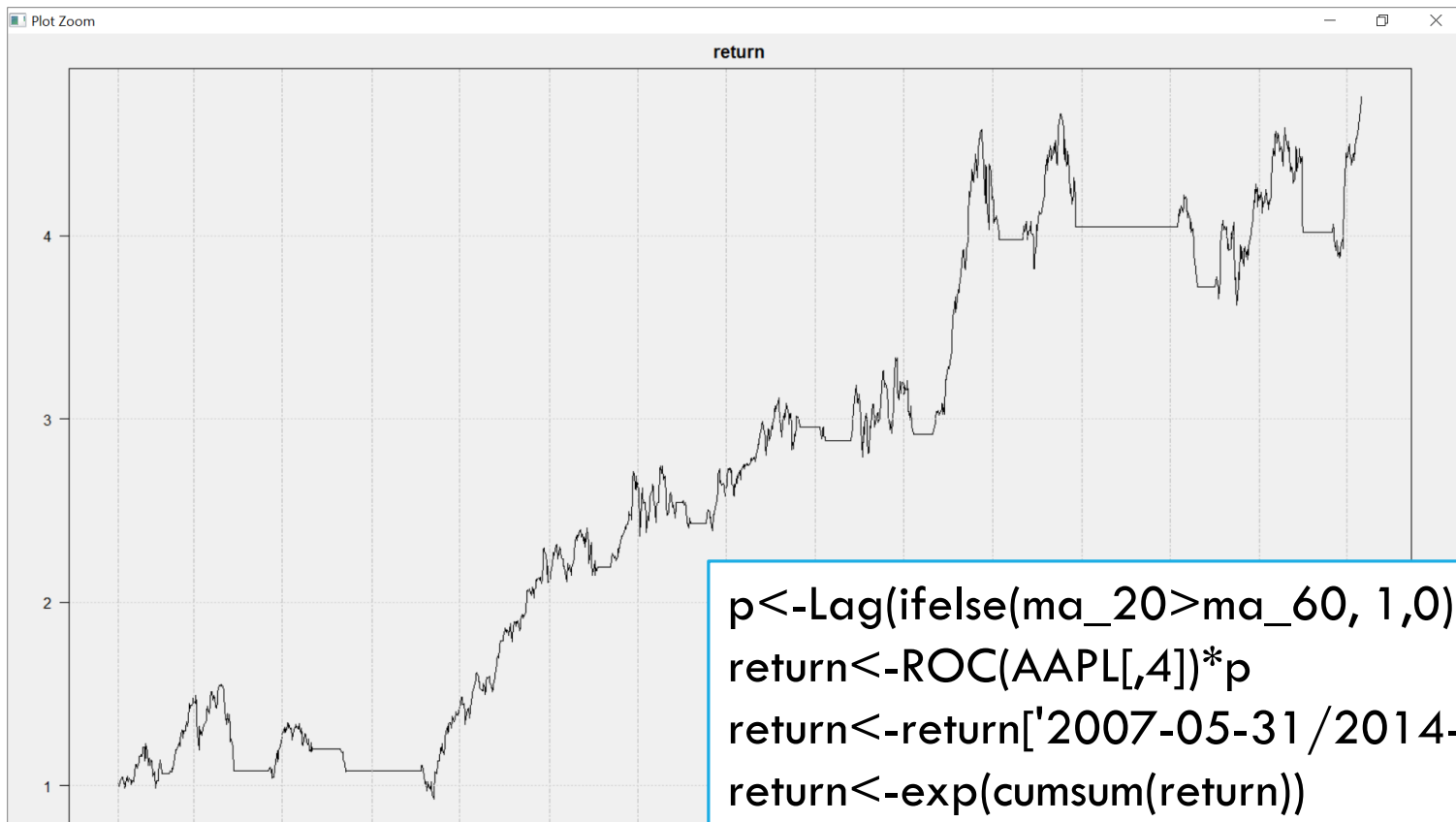
# R 套件 QUANTMOD 介紹

以均線策略回測為例

**當月線大於季線時，買進；當月線小於季線時，空手**

- 以日為單位  $t$ ；今天收盤價  $S(t)$ ；昨天收盤價  $S(t-1)$ 。
- 部位  $P(t)$ ：月線大於季線， $P(t)=1$ ；月線小於季線， $P(t)=0$ 。
- 對數報酬率： $\ln\left(\frac{S(t)}{S(t-1)}\right) \times P(t)$ 。
- 累計報酬率： $\exp\left(\sum_{t=1}^T \ln\left(\frac{S(t)}{S(t-1)}\right) \times P(t)\right)$ 。

# R 套件 QUANTMOD 介紹



```
p<-Lag(ifelse(ma_20>ma_60, 1,0))  
return<-ROC(AAPL[,4])*p  
return<-return['2007-05-31/2014-05-31']  
return<-exp(cumsum(return))  
plot(return)
```

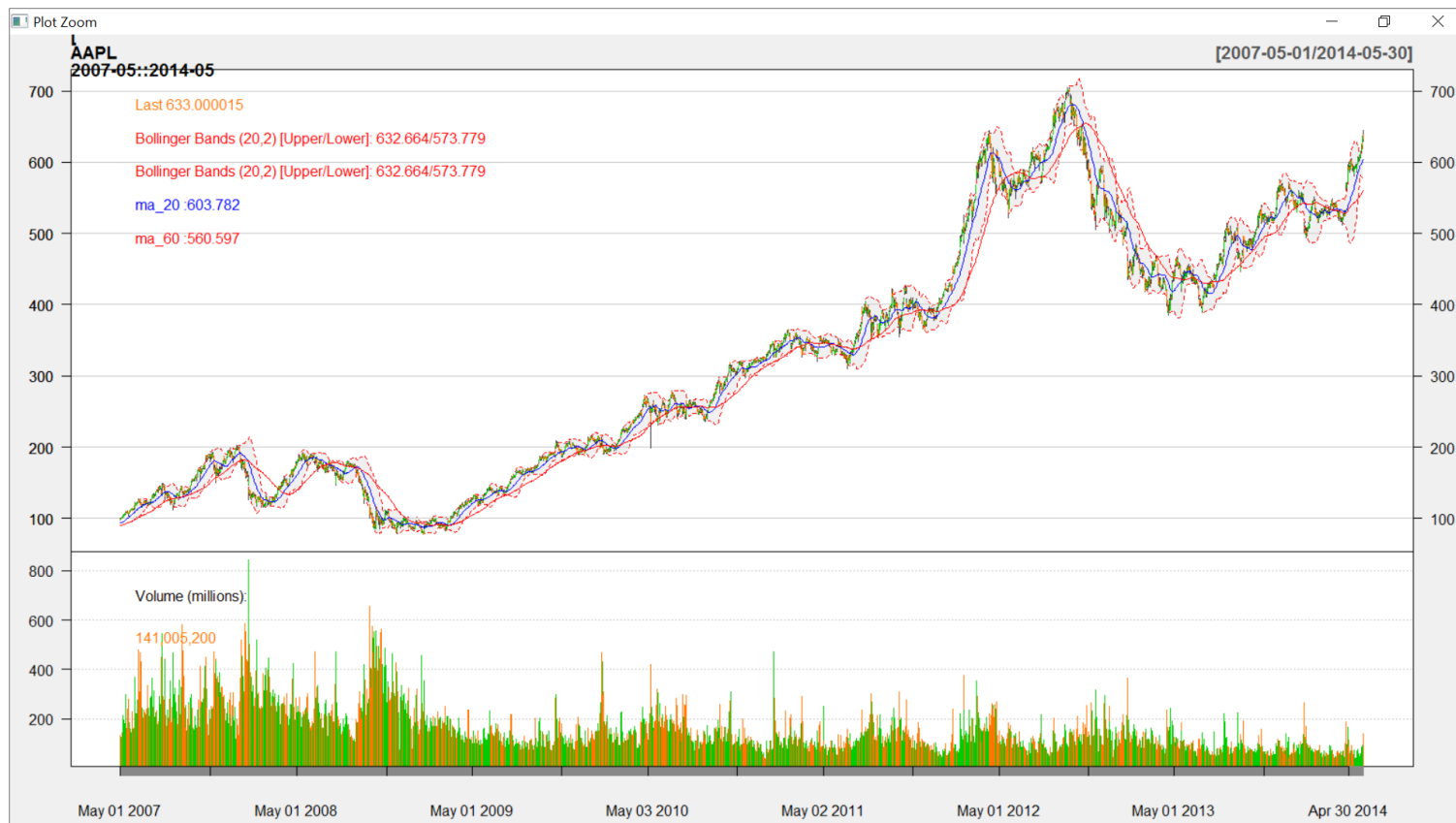
# R 套件 QUANTMOD 介紹

## 布林通道 (Bollinger Bands , Bbands)

> addBBands(n=20, sd=2)

- 結合移動平均和標準差，由上、中、下三條軌道線組成。
- 中軌為股價的平均成本； $N$  時間段的簡單移動平均線。
- 上軌為股價的壓力線；中軌 +  $K \times N$  時間段的標準差。
- 下軌為股價的支撐線；中軌 -  $K \times N$  時間段的標準差。
- 例如： $N=20$  為月均線， $K=2$  代表有 95% 的值分布在距離平均值正負 2 個標準差的範圍內。

# R 套件 QUANTMOD 介紹



# R 套件 QUANTMOD 介紹

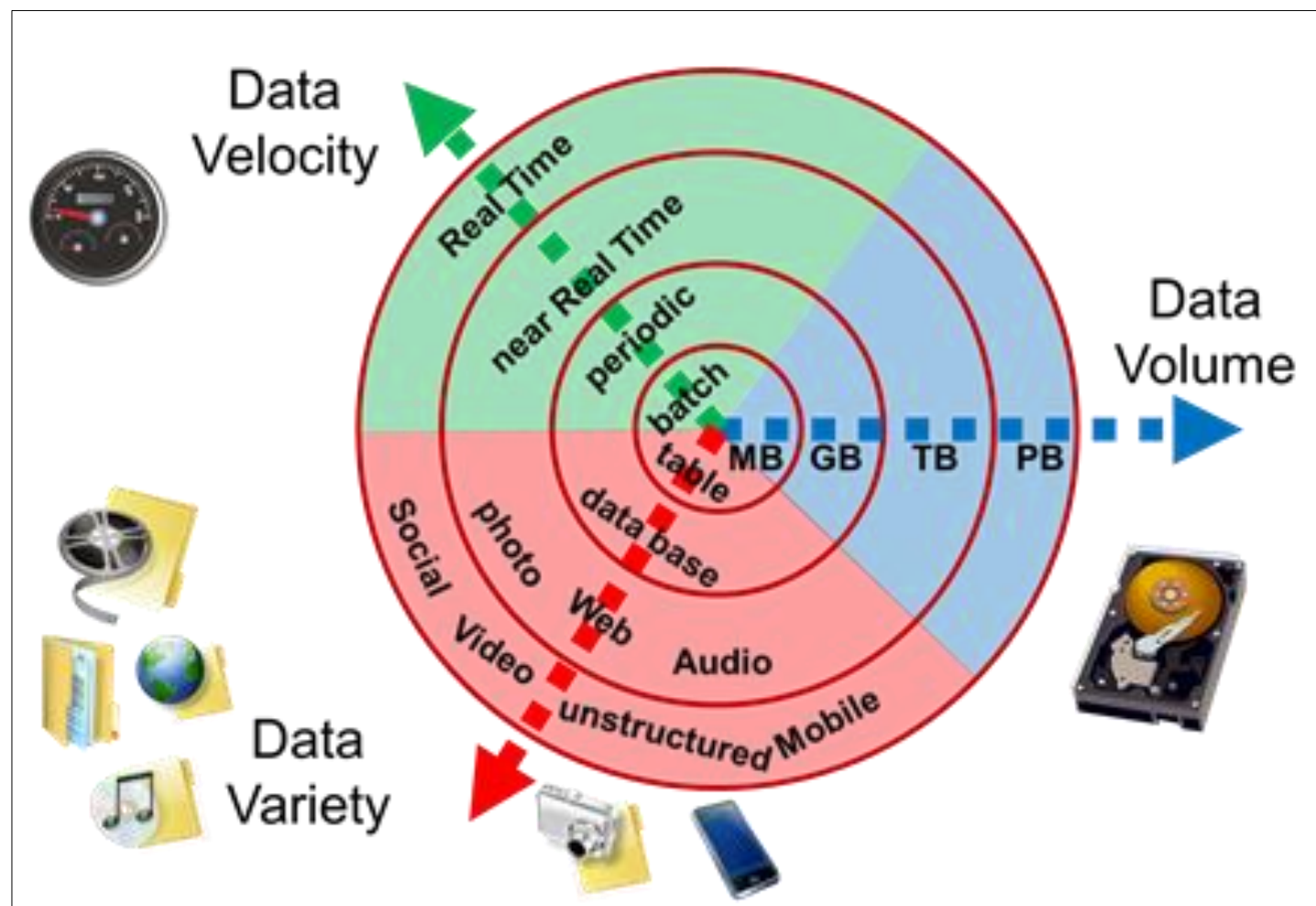
使用同一個交易區段 2007-05-31 ~ 2014-05-31

1. 每天開盤價交易 1 股，不停損，不停利。
2. 每天開盤價交易 1 股，不停損，漲 2% 停利。
3. 每天開盤價交易 1 股，跌 2% 停損，不停利。
4. 其他交易策略測試

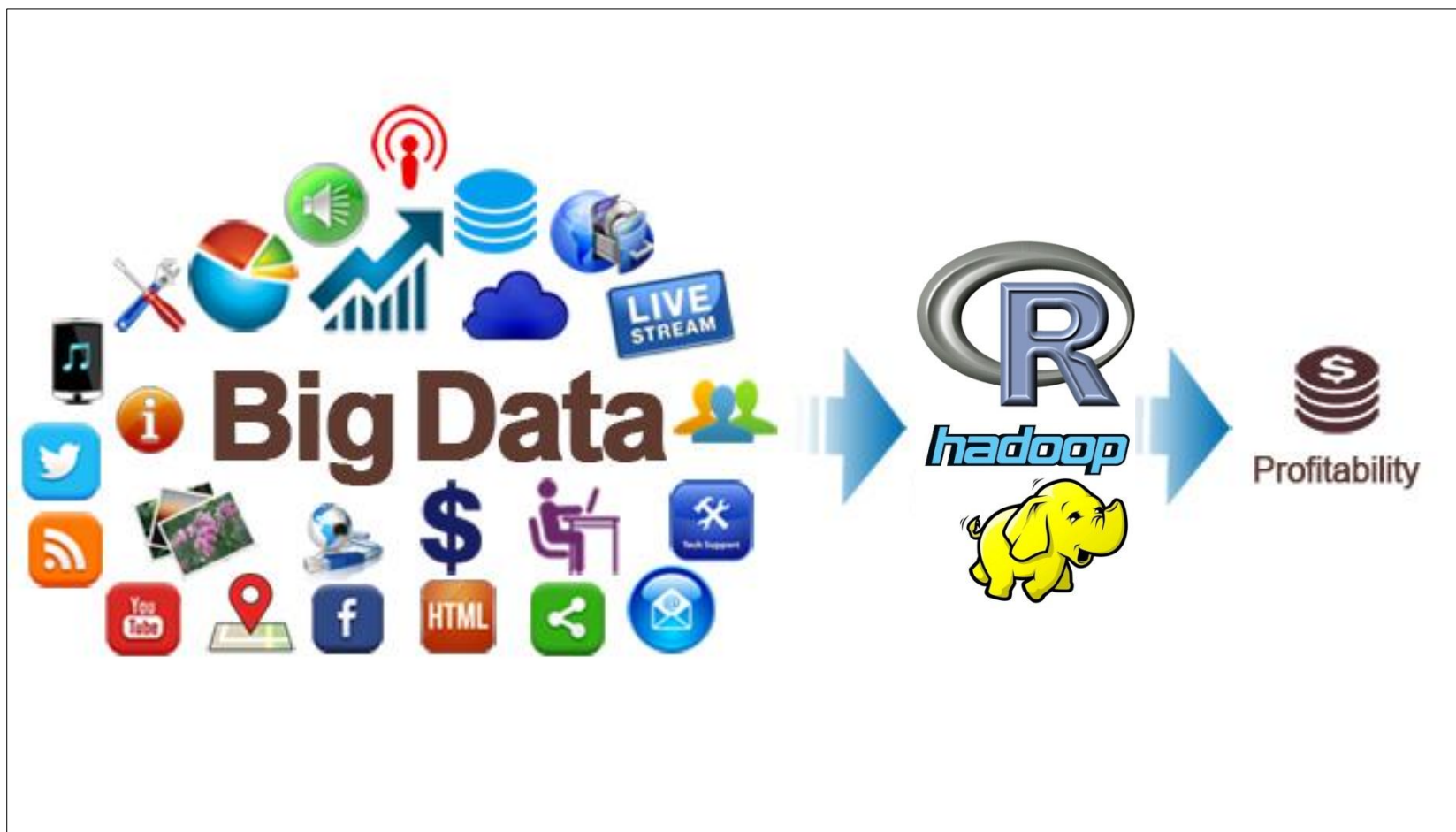
```
# first strategy
return<-ROC(AAPL[,4])
return<-return['2007-05-31/2014-05-31']
return<-exp(cumsum(return))
plot(return, main="First Strategy")
```



# 大數據是什麼？



# R 語言在大數據中的角色



Q&A

You have

Questions

We have

Answers