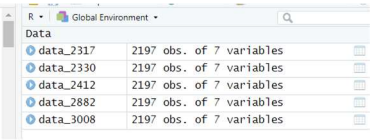
作業二 109304019 統計四周佳萱

1.

```
## Script for Exercise 2
## Fill "___" to complete the codes
## But don't forget to answer questions
rm(list = ls (all = TRUE))
##1. import data
data_2317<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2317.TW.csv", sep=",", header = T)
data_2330<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2330.TW.csv", sep=",", header = T)
data_2412<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2412.TW.csv", sep=",", header = T)
data_2882<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2882.TW.csv", sep=",", header = T)
data_3008<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/3008.TW.csv", sep=",", header = T)
```



2.

讀取 data_2317 的第二行和 data_2882 的第四行，我發現資料型態為 character。

```
> ## For example, choose the second column
> class(data_2317[,2])
[1] "character"
> ## Try another imported data, say 2882 and the fourth column
> class(data_2882[,4])
[1] "character"
```

3.

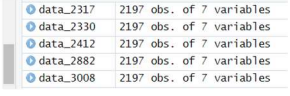
首先，讀取 data_2882 第二行，我發現缺失值會以”null”的形式表示。之後，我在計算 data_2882 的第二行有多少缺失值，發現有 5 個。最後，我分別計算 data_2317、data_2330、data_2412、data_2882、data_3008 整筆資料的缺失值，發現每個 data 均有 30 個缺失值。

```
> data_2882[,2]
   [1] "45.867981" "45.818977" "45.034908" "44.054825" "44.789886" "45.181923" "44.789886"
  [10] "44.348850" "44.054825" "43.907810" "43.809803" "44.299847" "44.691879" "45.720966"
  [19] "45.132915" "44.691879" "44.985905" "44.593872" "45.083912" "45.181923" "45.475948"
  [28] "45.622959" "46.358025" "46.554043" "46.701054" "46.946075" "47.142094" "47.240101"
  [37] "46.554043" "46.554043" "46.260014" "46.309017" "46.750057" "46.456032" "47.632133"
  [46] "47.583130" "47.142094" "47.387112" "48.563213" "48.857243" "49.200272" "48.514210"
  [55] "48.710228" "48.955250" "48.808235" "48.563213" "49.004253" "48.612221" "48.955250"
  [64] "49.494297" "49.690311" "49.396286" "49.788322" "49.200272" "49.396286" "49.788322"
  [73] "53.120609" "53.904678" "54.884766" "53.414635" "53.120609" "54.688747" "54.002686"
  [82] "54.198704" "52.924595" "54.394722" "54.100697" "52.924595" "53.414635" "53.610653"
  [91] "52.434551" "52.826584" "53.904678" "54.394722" "54.786755" "54.394722" "53.806671"
 [100] "52.728577" "50.964424" "50.964424" "50.964424" "51.552475" "51.748493" "51.748493"
 [109] "52.140526" "51.258450" "51.062431" "52.532558" "53.316628" "53.610653" "53.806671"
 [118] "51.454468" "51.650482" "51.160442" "50.768406" "50.964424" "50.768406" "48.906246"
 [127] "51.454468" "51.846500" "51.944508" "52.336544" "52.140526" "51.748493" "51.650482"
 [136] "50.866413" "49.984337" "48.857243" "49.004253" "49.102261" "49.886330" "49.396286"
 [145] "48.857243" "48.906246" "49.984337" "48.024166" "45.622959" "45.769974" "46.063999"
 [154] "42.780712" "42.241665" "40.869549" "40.379505" "42.143658" "43.809803" "44.299847"
 [163] "43.613785" "44.936901" "44.054825" "43.907810" "43.809803" "44.054825" "44.103828"
 [172] "44.348850" "44.593872" "46.063999" "46.652050" "45.867981" "45.818977" "45.573956"
 [181] "43.956814" "43.123741" "43.760799" "44.348850" "44.495861" "45.083912" "44.985905"
 [190] "45.279930" "45.279930" "45.279930" "45.377937" "45.573956" "45.720966" "45.769974"
 [199] "46.358025" "46.309017" "45.867981" "46.014996" "45.230927" "45.181923" "45.818977"
 [208] "48.416203" "47.926159" "46.554043" "45.818977" "45.524952" "44.887897" "44.740883"
 [217] "45.377937" "45.916985" "45.769974" "45.377937" "46.211010" "46.014996" "45.671963"
 [226] "45.279930" "44.691879" "44.446857" "44.691879" "44.201836" "42.927727" "41.359589"
 [235] "41.163574" "41.947639" "41.604610" "41.947639" "41.163574" "42.388680" "42.633701"
 [244] "43.074738" "43.123741" "43.074738" "42.143658" "43.319759" "41.163574" "40.624527"
 [253] "38.713360" "38.713360" "38.566345" "37.194229" "36.312153" "33.616917" "34.155964"
 [262] "33.518909" "33.812935" "34.008953" "33.665920" "33.469906" "33.665920" "null"
> ## Check how many missing data in these columns
> sum(data_2882[,2]=="null")
[1] 5
> sum(data_2317[,2:7]=="null")
[1] 30
> sum(data_2330[,2:7]=="null")
[1] 30
> sum(data_2412[,2:7]=="null")
[1] 30
> sum(data_2882[,2:7]=="null")
[1] 30
> sum(data_3008[,2:7]=="null")
[1] 30
```

4.

在使用 na.strings=”null”讀取資料以後，我發現每行資料型態變成 numeric。

```
## Set na.strings = "null" for the missing value
data_2317<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2317.TW.csv", sep=",", na.strings = "null", header = T)
data_2330<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2330.TW.csv", sep=",", na.strings = "null", header = T)
data_2412<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2412.TW.csv", sep=",", na.strings = "null", header = T)
data_2882<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/2882.TW.csv", sep=",", na.strings = "null", header = T)
data_3008<-read.table("C:/Users/chouchiahsuan/Desktop/財務資訊分析/HW2/3008.TW.csv", sep=",", na.strings = "null", header = T)
```



```
> ## Check class of one of the open, high, low, close again, for example
> class(data_2317[,2])
[1] "numeric"
> class(data_2317[,3])
[1] "numeric"
> class(data_2317[,4])
[1] "numeric"
> class(data_2317[,5])
[1] "numeric"
> class(data_2317[,6])
[1] "numeric"
> class(data_2317[,7])
[1] "integer"

> ## Try another imported data, say 2882 and the fourth column
> class(data_2882[,4])
[1] "numeric"
```

5.

Data_2317 每個變數的 summary statistics ：

```
> summary(data_2317)
      Date                Open            High            Low             Close           Adj.Close          Volume
 Min.   :2014-12-31   Min.   : 67.5   Min.   : 68.20   Min.   : 65.70   Min.   : 66.30   Min.   : 53.48   Min.   :        0
 1st Qu.:2017-03-30   1st Qu.: 90.0   1st Qu.: 90.77   1st Qu.: 89.00   1st Qu.: 89.79   1st Qu.: 67.80   1st Qu.: 20855576
 Median :2019-06-28   Median :101.9   Median :102.50   Median :101.00   Median :101.66   Median : 79.50   Median : 28469506
 Mean   :2019-06-30   Mean   :100.2   Mean   :100.98   Mean   : 99.36   Mean   :100.10   Mean   : 82.50   Mean   : 34962132
 3rd Qu.:2021-09-29   3rd Qu.:108.0   3rd Qu.:109.00   3rd Qu.:107.16   3rd Qu.:108.00   3rd Qu.: 97.78   3rd Qu.: 40454559
 Max.   :2023-12-28   Max.   :152.5   Max.   :153.12   Max.   :149.38   Max.   :151.88   Max.   :117.34   Max.   :316406354
                      NA's   :5       NA's   :5        NA's   :5        NA's   :5        NA's   :5        NA's   :5
```

Data_2330 每個變數的 summary statistics ：

```
> summary(data_2330)
      Date                Open            High            Low             Close           Adj.Close          Volume
 Min.   :2014-12-31   Min.   :118.0   Min.   :121.0   Min.   :112.5   Min.   :115.0   Min.   : 89.81   Min.   :        0
 1st Qu.:2017-03-30   1st Qu.:191.0   1st Qu.:192.0   1st Qu.:189.5   1st Qu.:191.5   1st Qu.:155.41   1st Qu.: 21605254
 Median :2019-06-28   Median :258.5   Median :260.0   Median :256.5   Median :258.8   Median :226.37   Median : 29203100
 Mean   :2019-06-30   Mean   :338.5   Mean   :340.8   Mean   :335.9   Mean   :338.4   Mean   :309.97   Mean   : 33518486
 3rd Qu.:2021-09-29   3rd Qu.:516.0   3rd Qu.:520.0   3rd Qu.:512.2   3rd Qu.:516.2   3rd Qu.:499.85   3rd Qu.: 40157952
 Max.   :2023-12-28   Max.   :685.0   Max.   :688.0   Max.   :678.0   Max.   :683.0   Max.   :651.42   Max.   :166307746
                      NA's   :5       NA's   :5       NA's   :5       NA's   :5       NA's   :5        NA's   :5
```

Data_2412 每個變數的 summary statistics ：

```
> summary(data_2412)
      Date                Open            High            Low             Close           Adj.Close          Volume
 Min.   :2014-12-31   Min.   : 92.5   Min.   : 92.9   Min.   : 92.1   Min.   : 92.5   Min.   : 63.41   Min.   :        0
 1st Qu.:2017-03-30   1st Qu.:106.5   1st Qu.:107.0   1st Qu.:106.0   1st Qu.:106.5   1st Qu.: 81.51   1st Qu.:  5524181
 Median :2019-06-28   Median :109.5   Median :110.0   Median :109.0   Median :110.0   Median : 92.49   Median :  7311065
 Mean   :2019-06-30   Mean   :110.1   Mean   :110.6   Mean   :109.5   Mean   :110.1   Mean   : 93.06   Mean   :  8420442
 3rd Qu.:2021-09-29   3rd Qu.:114.0   3rd Qu.:114.0   3rd Qu.:113.0   3rd Qu.:114.0   3rd Qu.:103.35   3rd Qu.: 10016514
 Max.   :2023-12-28   Max.   :132.0   Max.   :132.5   Max.   :131.5   Max.   :132.0   Max.   :123.05   Max.   :43766000
                      NA's   :5       NA's   :5       NA's   :5       NA's   :5       NA's   :5        NA's   :5
```

Data_2882 每個變數的 summary statistics ：

```
> summary(data_2882)
      Date                Open            High            Low             Close           Adj.Close          Volume
 Min.   :2014-12-31   Min.   :33.22   Min.   :33.42   Min.   :32.93   Min.   :32.93   Min.   :22.89   Min.   :        0
 1st Qu.:2017-03-30   1st Qu.:40.72   1st Qu.:40.87   1st Qu.:40.43   1st Qu.:40.70   1st Qu.:33.43   1st Qu.: 12400235
 Median :2019-06-28   Median :45.18   Median :45.52   Median :44.84   Median :45.18   Median :36.15   Median : 17434162
 Mean   :2019-06-30   Mean   :45.97   Mean   :46.29   Mean   :45.61   Mean   :45.95   Mean   :37.99   Mean   : 21414996
 3rd Qu.:2021-09-29   3rd Qu.:50.67   3rd Qu.:51.06   3rd Qu.:49.99   3rd Qu.:50.62   3rd Qu.:41.95   3rd Qu.: 25491879
 Max.   :2023-12-28   Max.   :66.26   Max.   :67.34   Max.   :65.47   Max.   :66.66   Max.   :60.70   Max.   :144866919
                      NA's   :5       NA's   :5       NA's   :5       NA's   :5       NA's   :5       NA's   :5
```

Data_3008 每個變數的 summary statistics ：

```
> summary(data_3008)
      Date                Open            High            Low             Close           Adj.Close          Volume
 Min.   :2014-12-31   Min.   :1540   Min.   :1585   Min.   :1525   Min.   :1565   Min.   :1461   Min.   :      0
 1st Qu.:2017-03-30   1st Qu.:2315   1st Qu.:2355   1st Qu.:2289   1st Qu.:2315   1st Qu.:2102   1st Qu.: 427532
 Median :2019-06-28   Median :3260   Median :3320   Median :3220   Median :3260   Median :2736   Median : 613272
 Mean   :2019-06-30   Mean   :3324   Mean   :3368   Mean   :3276   Mean   :3316   Mean   :2831   Mean   : 721635
 3rd Qu.:2021-09-29   3rd Qu.:4071   3rd Qu.:4135   3rd Qu.:4000   3rd Qu.:4060   3rd Qu.:3435   3rd Qu.: 886887
 Max.   :2023-12-28   Max.   :6000   Max.   :6075   Max.   :5955   Max.   :6000   Max.   :4973   Max.   :3866779
                      NA's   :5      NA's   :5      NA's   :5      NA's   :5      NA's   :5      NA's   :5
```

6.

每個資料的每個變數都有 5 個缺失值，不同變數的缺失值都出現在同個列，每筆股票資料總共有 30 個缺失值，缺失值都在日期 2016-01-30、2016-06-04、2016-09-10、2017-02-18、2017-06-03，並且都在禮拜六。

```
> which(is.na(data_2317[, 2]))
[1] 268 348 417 523 592
> which(is.na(data_2317[, 3]))
[1] 268 348 417 523 592
> which(is.na(data_2317[, 4]))
[1] 268 348 417 523 592
> which(is.na(data_2317[, 5]))
[1] 268 348 417 523 592
> which(is.na(data_2317[, 6]))
[1] 268 348 417 523 592
> which(is.na(data_2317[, 7]))
[1] 268 348 417 523 592
> na_2317<-which(is.na(data_2317[, 2]))            ## Construct an i
> data_2317[na_2317, 1]                            ## dates for the
[1] "2016-01-30" "2016-06-04" "2016-09-10" "2017-02-18" "2017-06-03"
> weekdays(data_2317[na_2317, 1])
[1] "星期六" "星期六" "星期六" "星期六" "星期六"
```

```
> which(is.na(data_2330[, 2]))
[1] 268 348 417 523 592
> which(is.na(data_2330[, 3]))
[1] 268 348 417 523 592
> which(is.na(data_2330[, 4]))
[1] 268 348 417 523 592
> which(is.na(data_2330[, 5]))
[1] 268 348 417 523 592
> which(is.na(data_2330[, 6]))
[1] 268 348 417 523 592
> which(is.na(data_2330[, 7]))
[1] 268 348 417 523 592
> na_2330<-which(is.na(data_2330[, 2]))
> data_2330[na_2330, 1]
[1] "2016-01-30" "2016-06-04" "2016-09-10" "2017-02-18" "2017-06-03"
> weekdays(data_2330[na_2330, 1])
[1] "星期六" "星期六" "星期六" "星期六" "星期六"
>
```

```
> which(is.na(data_2412[, 2]))
[1] 268 348 417 523 592
> which(is.na(data_2412[, 3]))
[1] 268 348 417 523 592
> which(is.na(data_2412[, 4]))
[1] 268 348 417 523 592
> which(is.na(data_2412[, 5]))
[1] 268 348 417 523 592
> which(is.na(data_2412[, 6]))
[1] 268 348 417 523 592
> which(is.na(data_2412[, 7]))
[1] 268 348 417 523 592
>
> na_2412<-which(is.na(data_2412[, 2]))
> data_2412[na_2412, 1]
[1] "2016-01-30" "2016-06-04" "2016-09-10" "2017-02-18" "2017-06-03"
> weekdays(data_2412[na_2412, 1])
[1] "星期六" "星期六" "星期六" "星期六" "星期六"
```

```
> which(is.na(data_2882[, 2]))
[1] 268 348 417 523 592
> which(is.na(data_2882[, 3]))
[1] 268 348 417 523 592
> which(is.na(data_2882[, 4]))
[1] 268 348 417 523 592
> which(is.na(data_2882[, 5]))
[1] 268 348 417 523 592
> which(is.na(data_2882[, 6]))
[1] 268 348 417 523 592
> which(is.na(data_2882[, 7]))
[1] 268 348 417 523 592
>
> na_2882<-which(is.na(data_2882[, 2]))
> data_2882[na_2882, 1]
[1] "2016-01-30" "2016-06-04" "2016-09-10" "2017-02-18" "2017-06-03"
> weekdays(data_2882[na_2882, 1])
[1] "星期六" "星期六" "星期六" "星期六" "星期六"
```

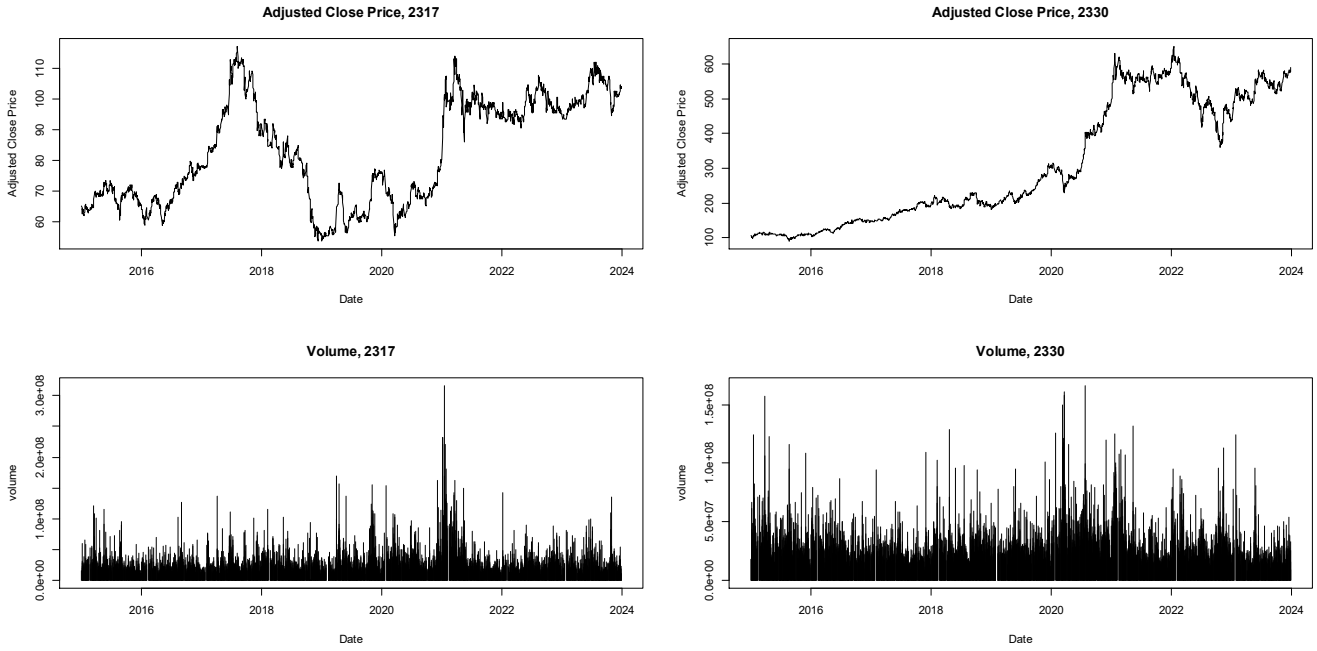
```
> which(is.na(data_3008[, 2]))
[1] 268 348 417 523 592
> which(is.na(data_3008[, 3]))
[1] 268 348 417 523 592
> which(is.na(data_3008[, 4]))
[1] 268 348 417 523 592
> which(is.na(data_3008[, 5]))
[1] 268 348 417 523 592
> which(is.na(data_3008[, 6]))
[1] 268 348 417 523 592
> which(is.na(data_3008[, 7]))
[1] 268 348 417 523 592
>
> na_3008<-which(is.na(data_3008[, 2]))
> data_3008[na_3008, 1]
[1] "2016-01-30" "2016-06-04" "2016-09-10" "2017-02-18" "2017-06-03"
> weekdays(data_3008[na_3008, 1])
[1] "星期六" "星期六" "星期六" "星期六" "星期六"
```

7.

```
##
##7. Eliminate NA's with complete.cases
data_2317<-data_2317[complete.cases(data_2317), ]
data_2330<-data_2330[complete.cases(data_2330), ]
data_2412<-data_2412[complete.cases(data_2412), ]
data_2882<-data_2882[complete.cases(data_2882), ]
data_3008<-data_3008[complete.cases(data_3008), ]
)
##
```

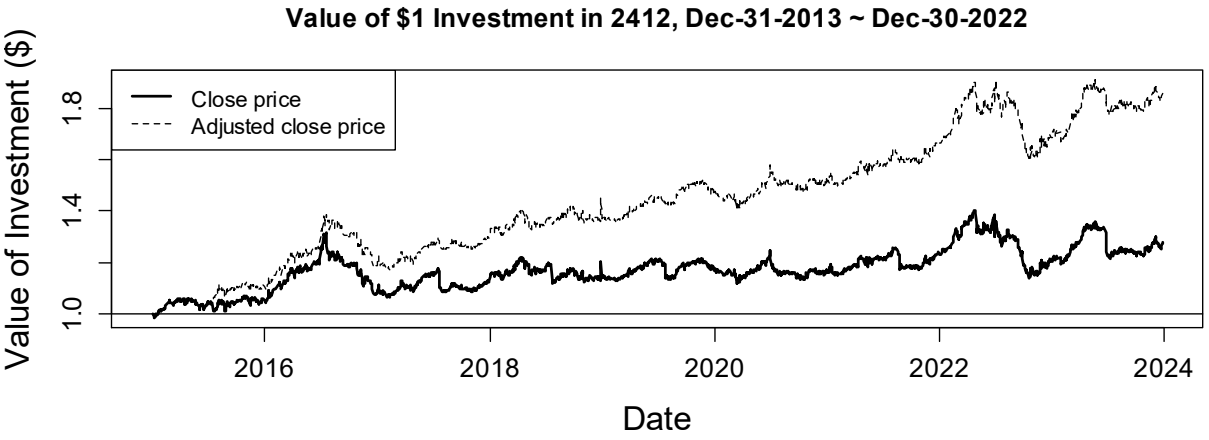
Data	
data_2317	2192 obs. of 7 variables
data_2330	2192 obs. of 7 variables
data_2412	2192 obs. of 7 variables
data_2882	2192 obs. of 7 variables
data_3008	2192 obs. of 7 variables
Values	

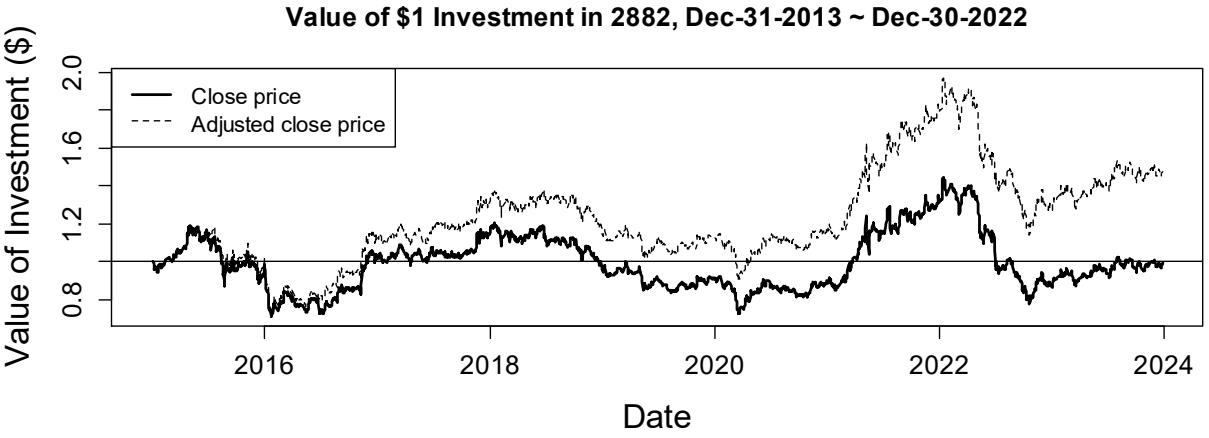
8.



9.

首先，我發現 2412 和 2882 的 normalized price 在 Adjusted close price 時較 closed price 高。Normalized price 在 2412 大致大於 1，且為上升的趨勢；在 2882 變動較大，呈上下震盪，有時大於 1，有時小於 1。





10.