第一章：R 语言简介

梁龙跃

Email: lyliang@gzu.edu.cn
Tel: 18275242613

贵州大学 经济学院

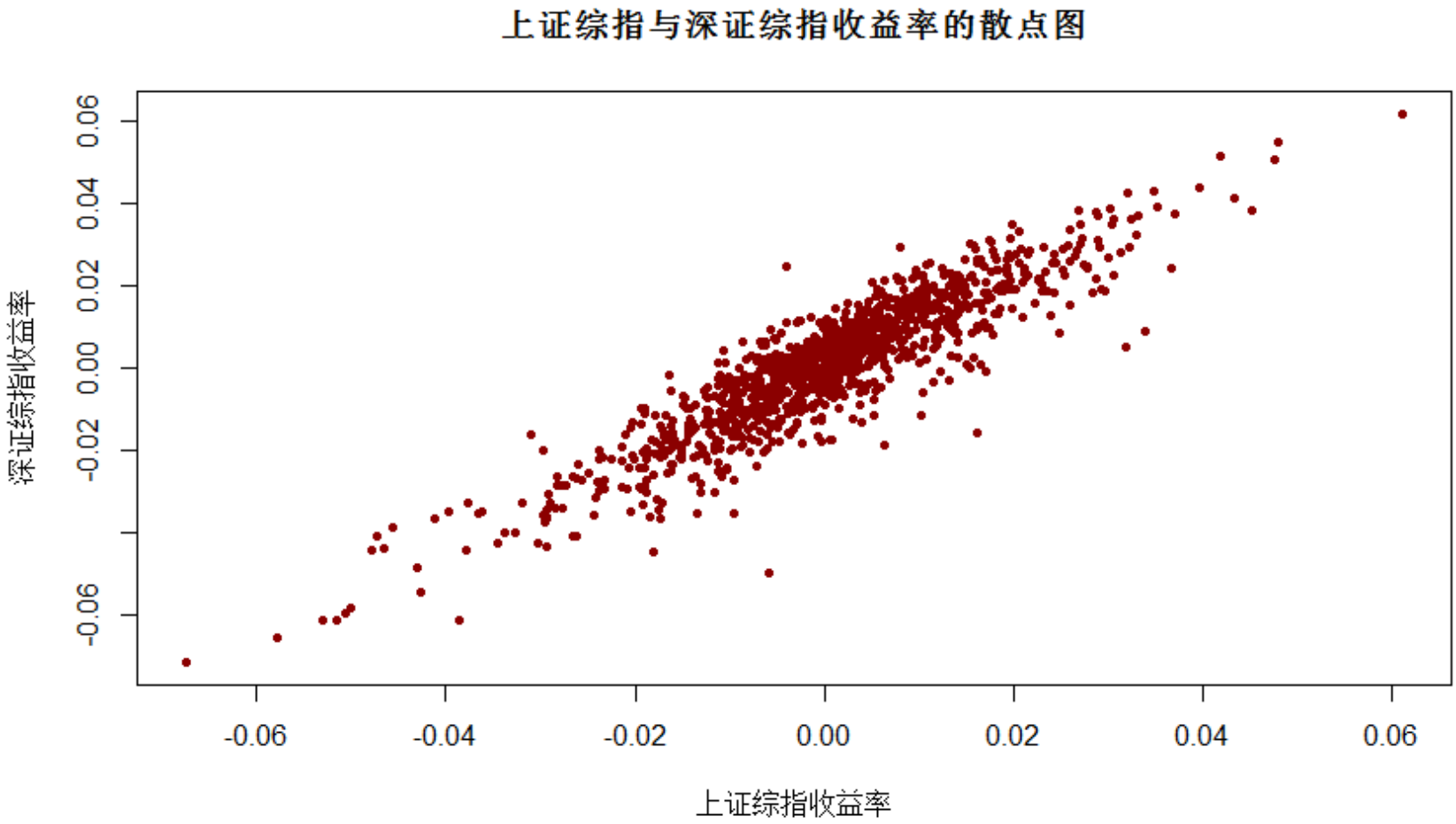
2018 年 3 月

本章内容

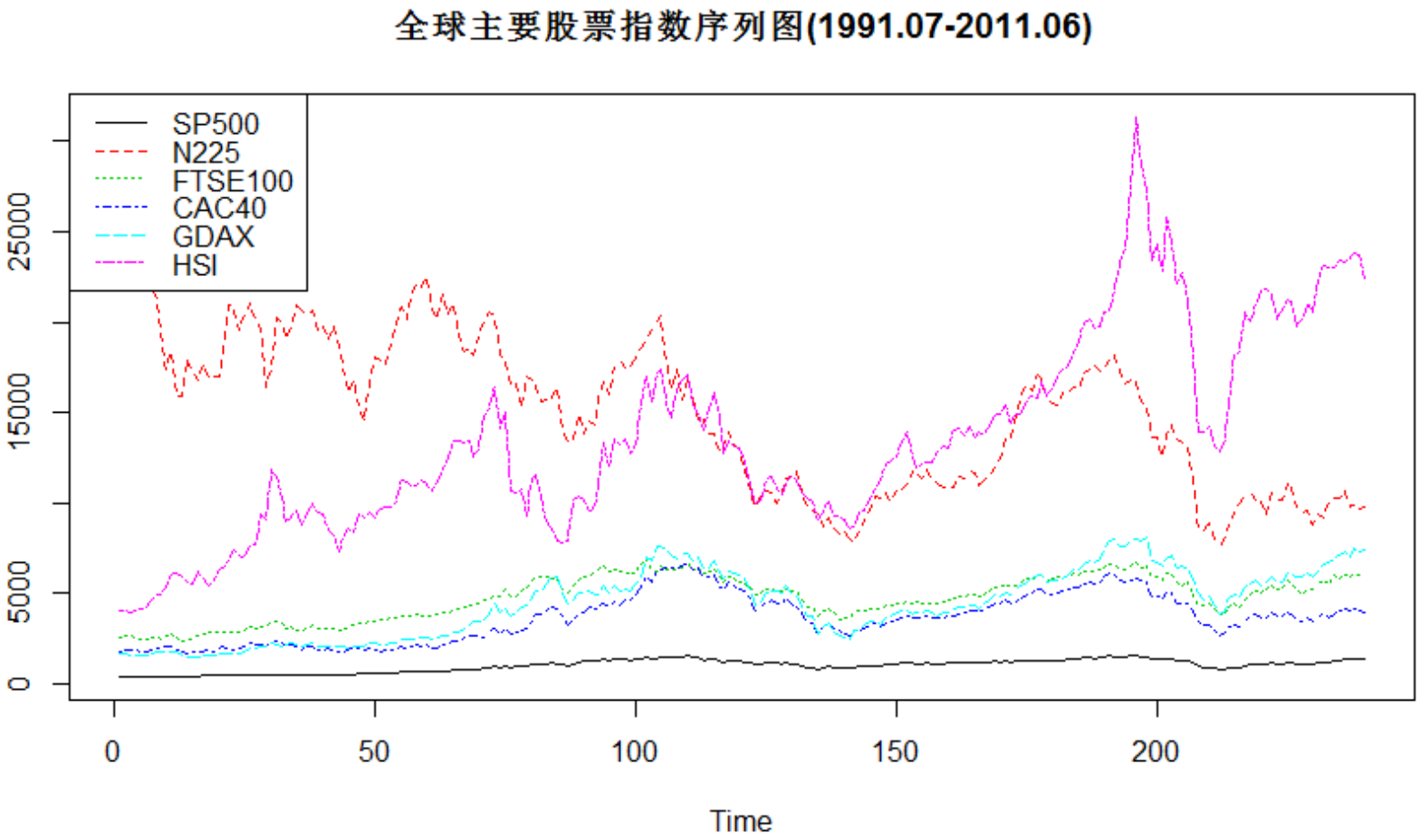
一些有趣的图


上证综指
6000
5000
4000
3000
2000
一月 04 2007
一月 05 2009
一月 04 2011
一月 04 2013
一月 02 2015

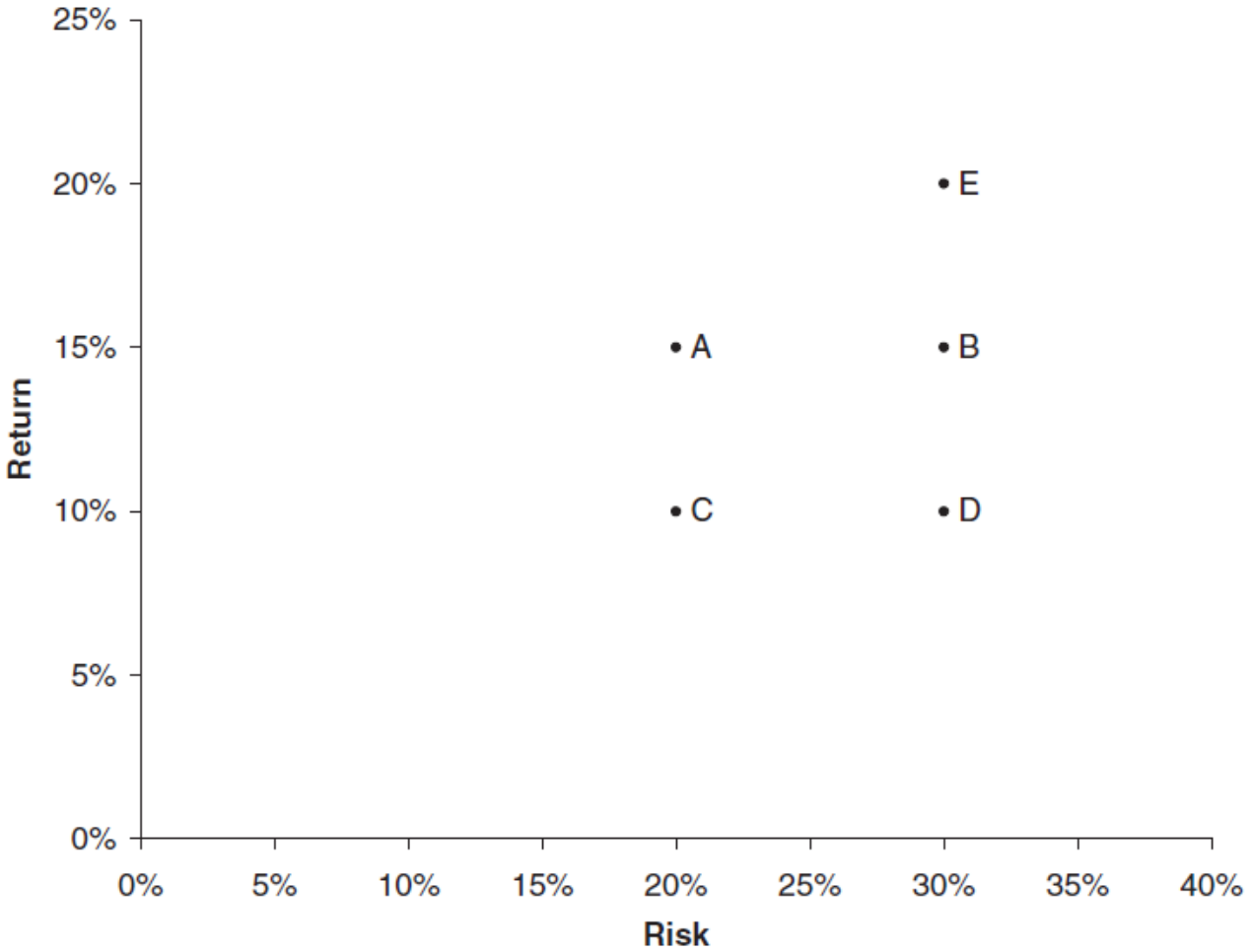
一些有趣的图


上证综指与深证综指收益率的散点图
深证综指收益率
-0.06
-0.02
0.00
0.02
0.04
0.06
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06
上证综指收益率

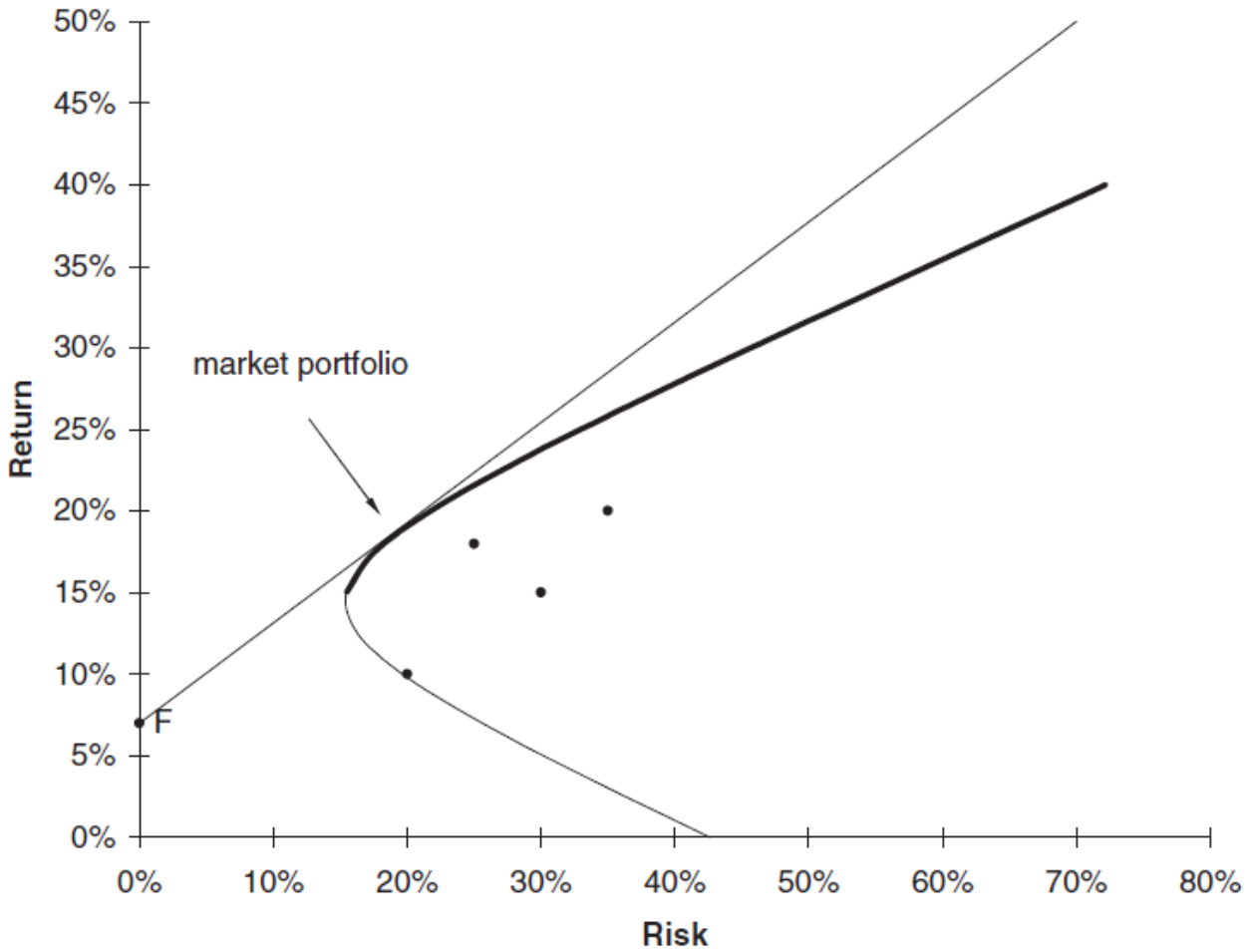
一些有趣的图


全球主要股票指数序列图(1991.07-2011.06)
SP500
N225
FTSE100
CAC40
GDAX
HSI
0
5000
15000
25000
0
50
100
150
200
Time

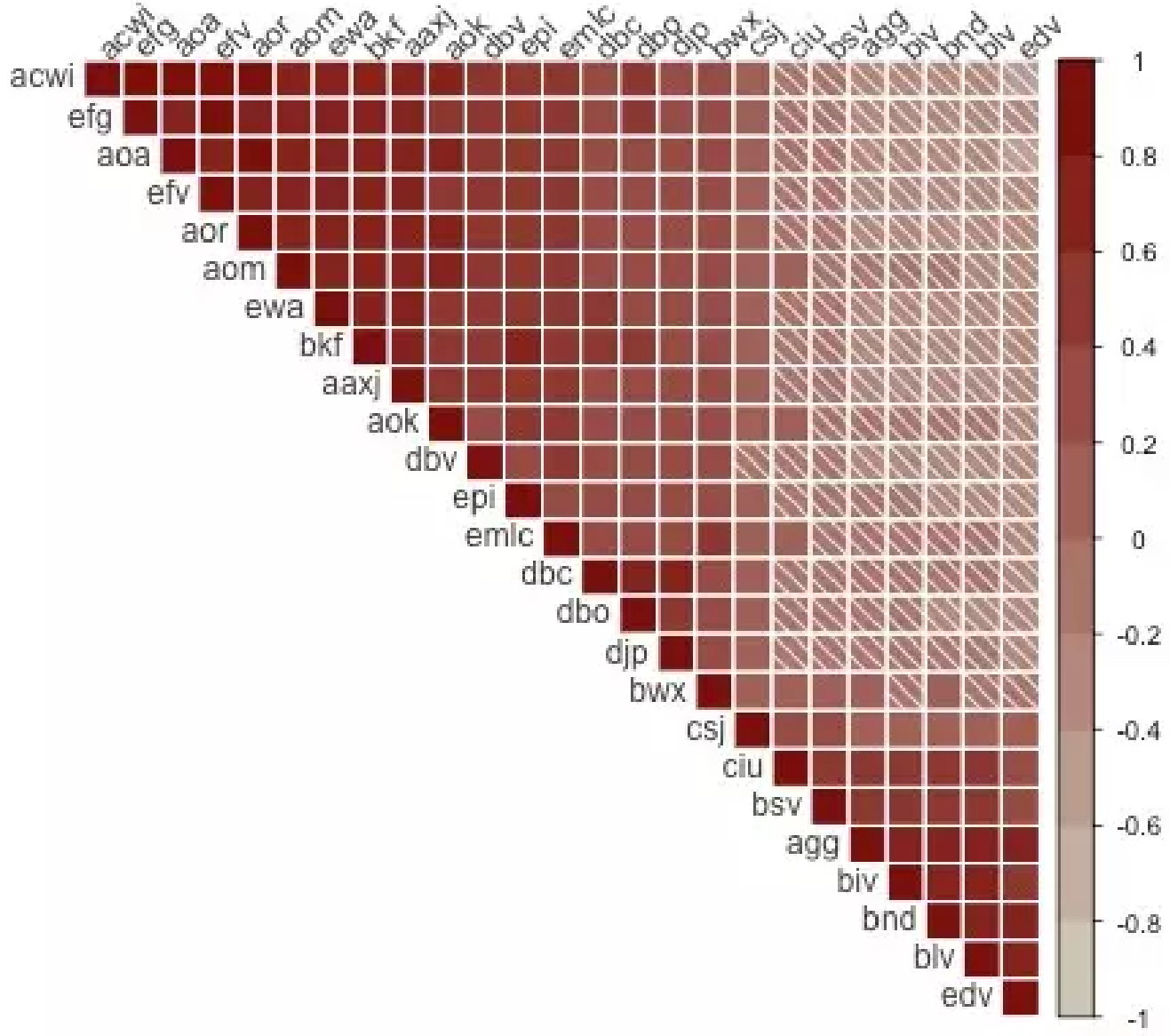
一些有趣的图



一些有趣的图


50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Return
market portfolio
F
0%
10%
20%
30%
40%
50%
60%
70%
80%
Risk

一些有趣的图


acwi
efg
aoa
efv
aor
aom
ewa
bkf
aaxj
aok
dbv
epi
emlc
dbc
dbo
djp
bwx
csj
ciu
bsv
agg
biv
bnd
blv
edv
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

一些有趣的图

06-16 年不同行业与沪深 300 的关联度走势图

本课程大纲

1. R 语言简介;
2. 现代投资组合理论和风险管理;
3. 金融时间序列
4. 风险价值;
5. 波动率;
6. 机器学习与风险管理;
7. 相关性与 Copula 函数;
8. 利率风险;
9. 信用风险;
10. 金融机构及其风险介绍;(由学生讨论讲授)

参考教材

1. 《风险管理与金融工具》, 赫尔, 机械工业出版社;
2. 《金融风险管理》, 克里斯托弗森, 中国人民大学出版社;
3. 《金融风险管理》, 王勇, 机械工业出版社;
4. 《金融风险管理师手册》, 乔瑞, 中国人民大学出版社;
5. 《R 软件及其在金融定量分析中的应用》, 许启发, 蒋翠侠, 清华大学出版社;
6. 《R 语言实战》, Kabacoff, 人民邮电出版社;
7. 《量化投资以R语言为工具》, 蔡立耑, 电子工业出版社;
8. 《金融数据分析导论基于 R 语言》, Tsay, 机械工业出版社.

R 简介

简单地说, R 语言就是一个用于数据统计处理的软件包, 它支持使用一种简单的语言(即所谓的 R 语言)来输入各种命令.

R 语言本来是由来自新西兰奥克兰大学的 Ross Ihaka 和 Robert Gentleman 开发(由于他们的名字以 R 开头所以该软件也因此称为 R), 现在由 "R 开发核心团队" 负责开发.

R 语言可以做什么

1, 高级计算器

计算

$$\frac{(1239+399.456i)^2\times(-243+0.45i)+(3989-21.5i)^3}{(341-0.0034i)^3\times(32.33-533i)+(3.211-65i)\times(543+12i)^5}$$

R 语言代码:

```
((1239+399.456i)^2*(-243+0.45i)+(3989-21.5i)^3)/
((341-0.0034i)^3*(32.33-533i)+(3.211-65i)*(543+12i)^5)
```

结果是: 4.89559e+14-3.03671e+15i.

R 语言可以做什么

1, 高级计算器

解线性方程组, 即从矩阵方程 $\mathbf{AX} = \mathbf{b}$ 中解出 $\mathbf{X}$, 演示如下:

```
A=matrix(rnorm(16),4,4);A   #构造矩阵 A 并显示
```

```
           [,1]       [,2]       [,3]        [,4]
[1,] -0.6078411  1.40734169 -1.1970641  0.02443725
[2,]  1.0762231  0.03665615 -1.4946442 -0.43005068
[3,] -0.5764258  1.92391863  0.7128978 -0.89508759
[4,]  1.0986264 -1.25501876 -0.5488986  1.27723800
```

```
b=c(1:4)    #构造向量 b
solve(A,b)    #解方程
```

解为:

```
[1]   4.999695  97.972865  64.588649 126.793903
```

R 语言可以做什么

2, 模拟抽签

- 抽签实际上是随机抽样. 比如有 100 个人, 盒子里装有写有 1-100 编号的纸条, 现有 10 个奖品, 需要抽签来确定谁得奖. 不放回地抽取 10 个纸条, 这 10 个纸条上的号码就是中奖人的编号(无重复号码). 如果一个人不限于得到一个奖品, 那就是有放回地抽取, 这次就可能有重复的号码了.
- 摇号买轿车的过程也是一个不放回的抽样过程.
- 计算机随机抽样简单, 方便, R 语言实现的代码如下:

```
sample(1:100,10)    #无放回抽样
```

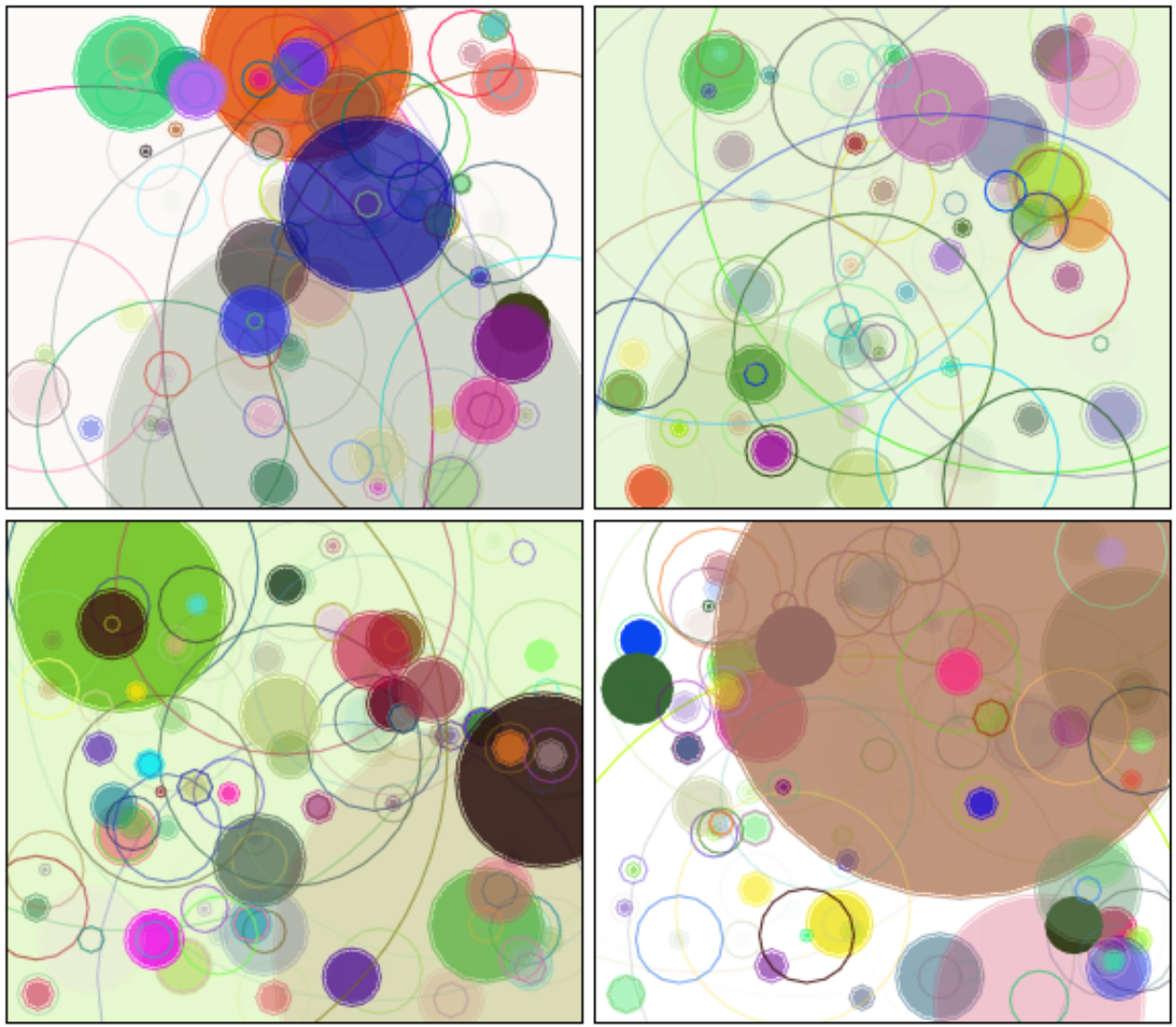
```
[1] 47 25 51 50 71 60 55 31 54 29
```

```
sample(1:100,10,rep=T)    #有放回抽样,rep为replace
```

```
[1] 95 51 20 73 78 41  7  3 73 60
```

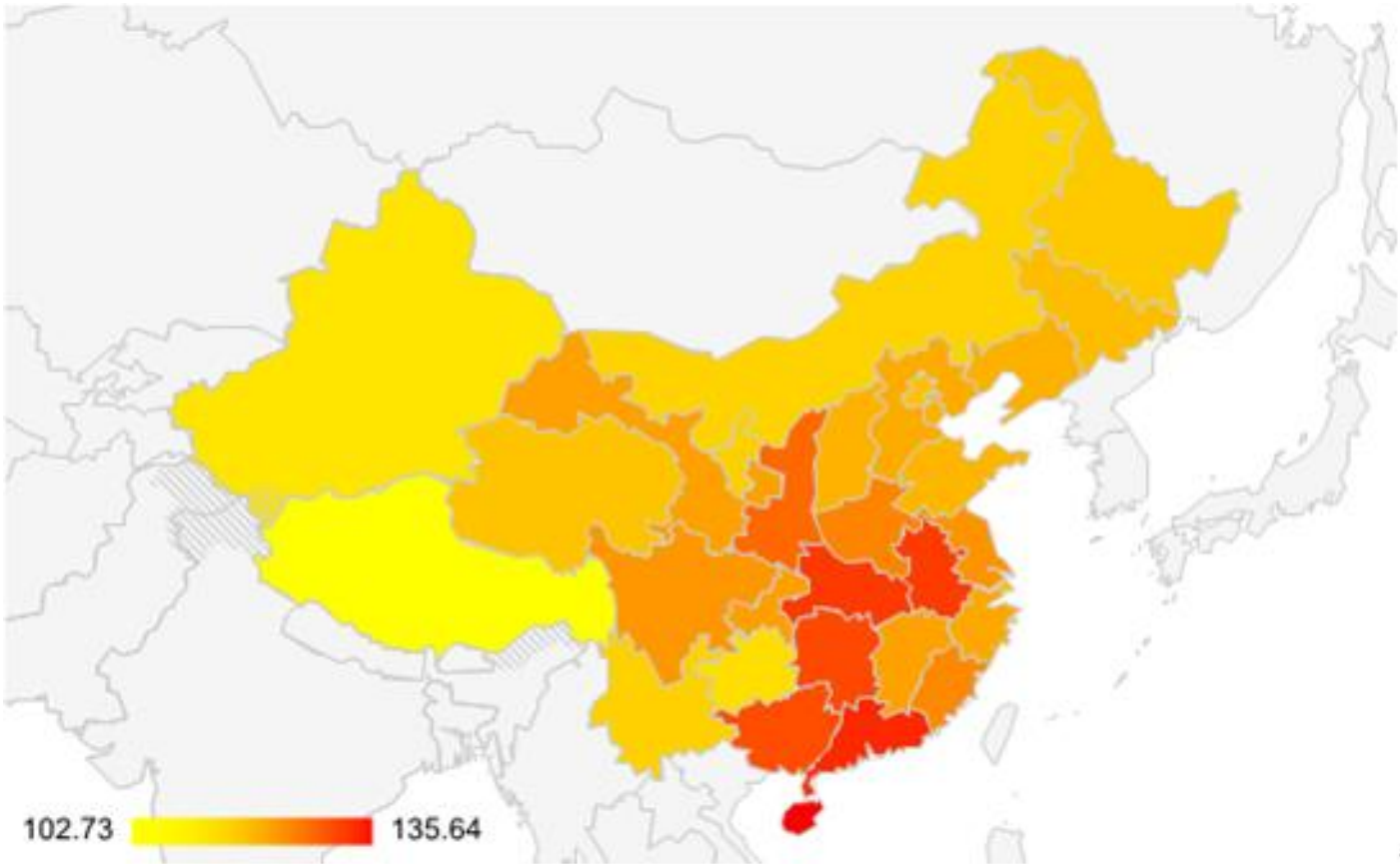
R 语言可以做什么

3, 绘制漂亮的图形



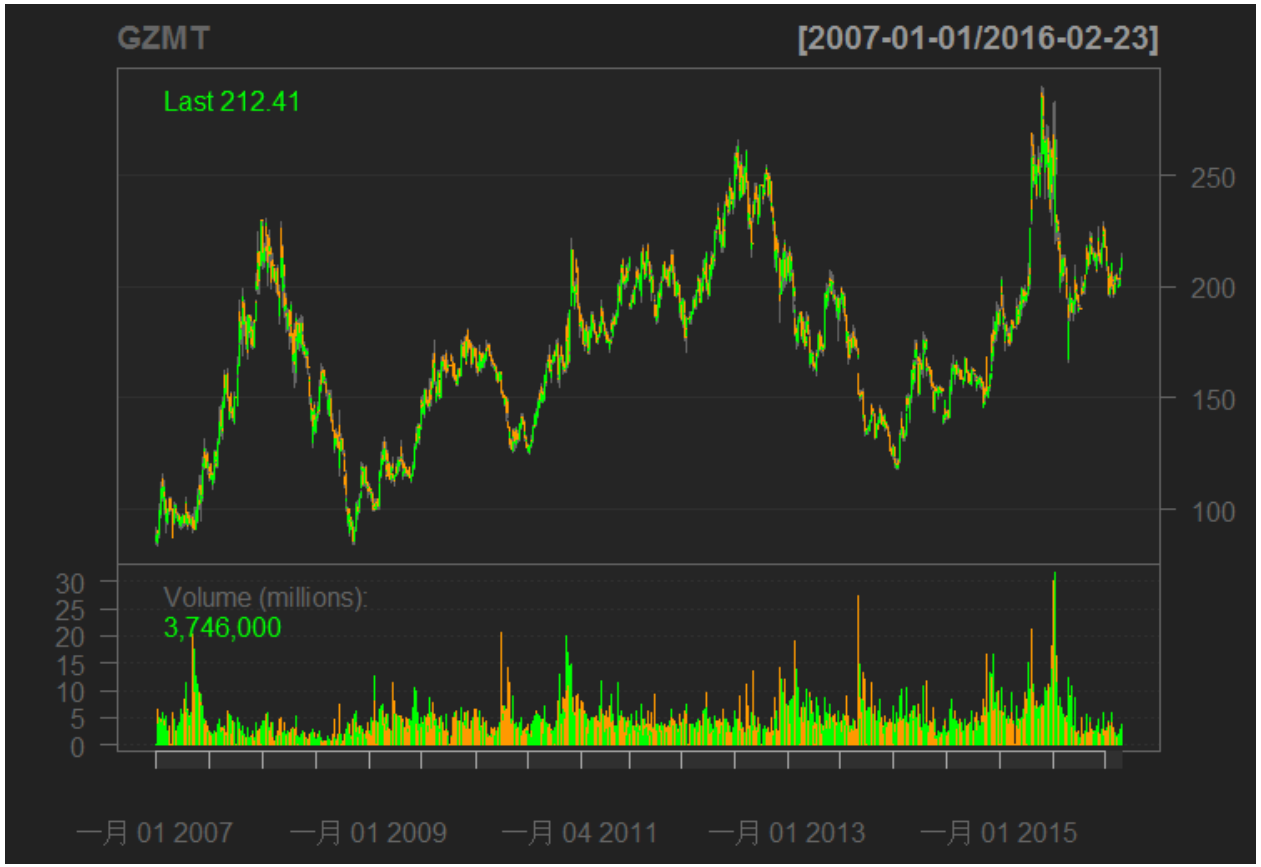
R 语言可以做什么

3, 绘制漂亮的图形



R 语言可以做什么

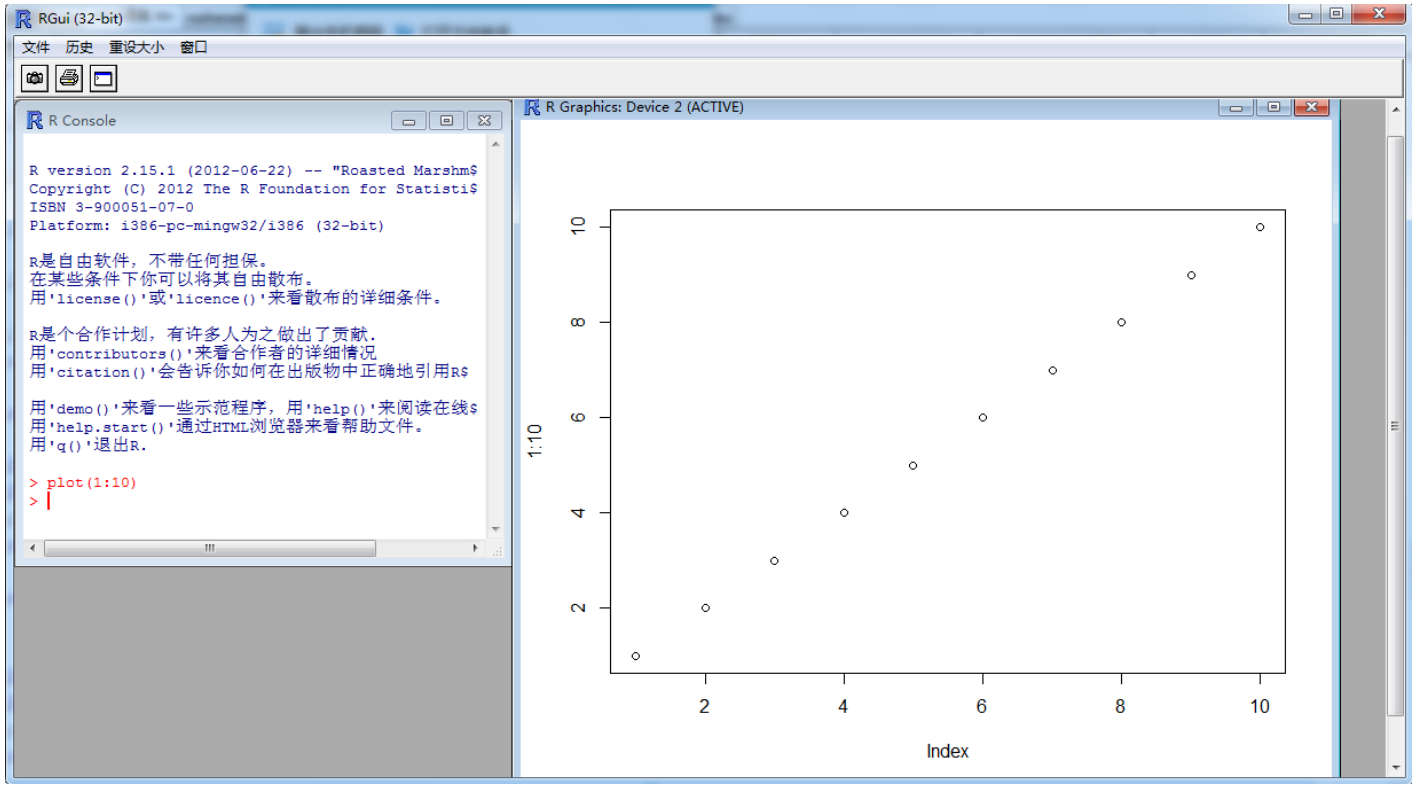
4, 在线获取股票数据



R 特点

- 易学
- 最热门
- 开放源代码
- 数据处理
- 内置大量函数
- 绘图精美
- 无数个用途各异的"包"
- 具有丰富的网上资源
- 免费

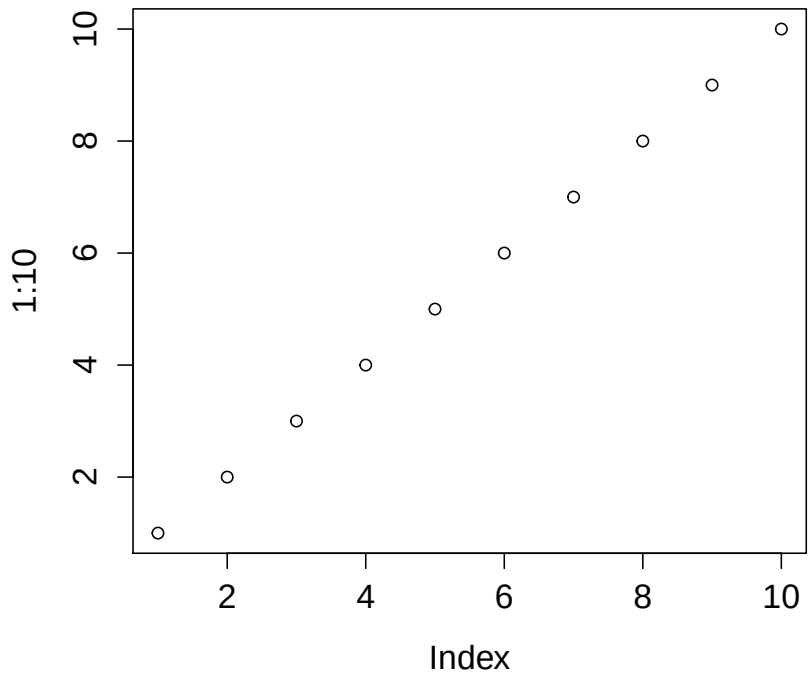
R 软件


RGui (32-bit)
文件 历史 重设大小 窗口
R Console
R version 2.15.1 (2012-06-22) -- "Roasted Marshm$
Copyright (C) 2012 The R Foundation for Statisti$
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)
R是自由软件，不带任何担保。
在某些条件下你可以将其自由散布。
用'license()'或'licence()'来看散布的详细条件。
R是个合作计划，有许多人为之做出了贡献.
用'contributors()'来看合作者的详细情况
用'citation()'会告诉你如何在出版物中正确地引用R$
用'demo()'来看一些示范程序，用'help()'来阅读在线$
用'help.start()'通过HTML浏览器来看帮助文件。
用'q()'退出R.
> plot(1:10)
>
R Graphics: Device 2 (ACTIVE)
1:10
Index
2
4
6
8
10

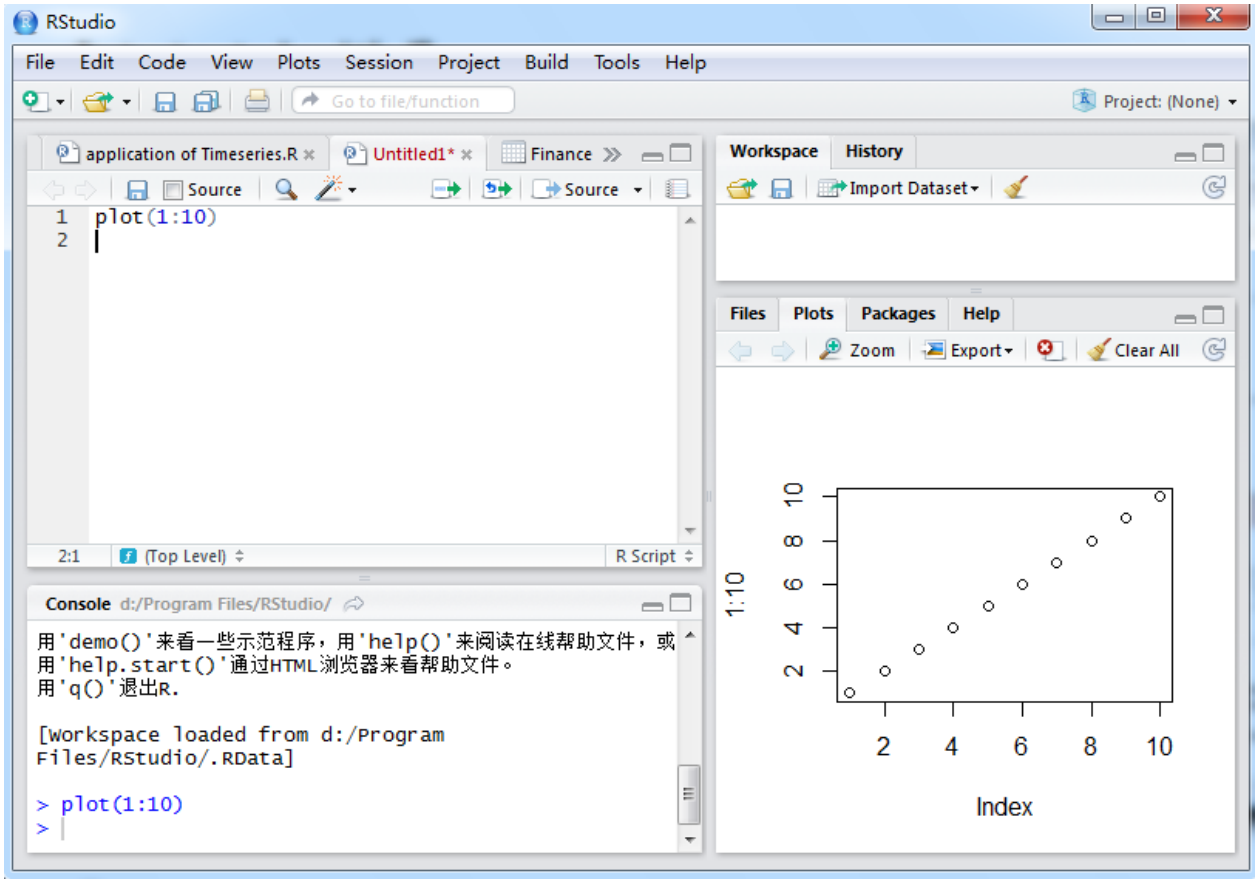
简单的例子

```
> plot(1:10)
```

上面的 ">" 符号, 是 R 的"提示符", 即所有命令都是在一个 ">" 符号后面输入的, 因此, 只有后面的 "plot(1:10)" 才是你输入的命令, 该命令的含义是: 作一个图, 里面包含从 1 到 10 共 10 个点.



Rstudio

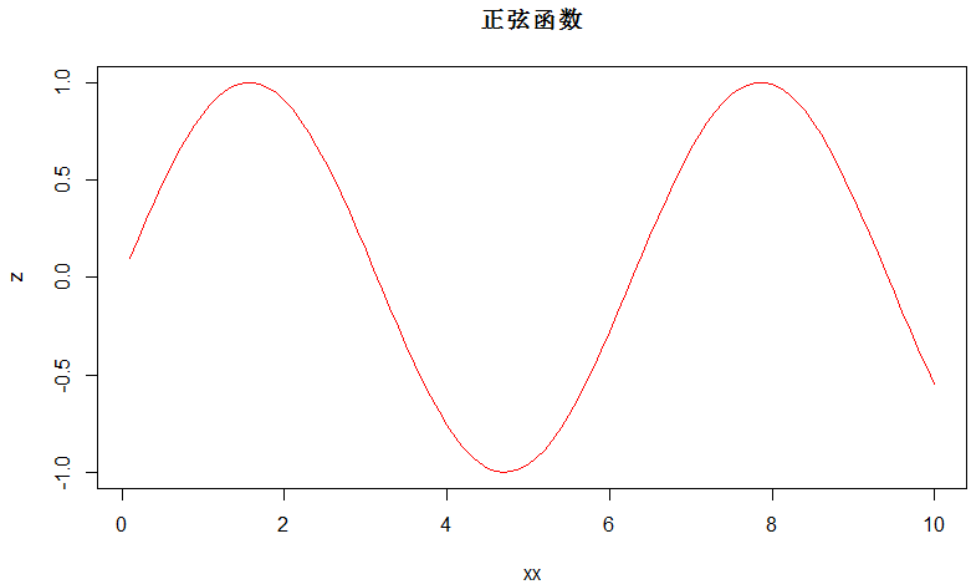


目前, R 语言软件中最好的编辑器是 Rstudio, 本课程的程序都在该平台上进行.

该编辑器可以在 Rstudio 网站(http://www.rstudio.org/) 下载.

简单的例子

```
x=1:100      #产生1-100个数,存储在x变量中
xx=x/10      #0-10以内的100个样本点
z=sin(xx)    #一次计算出100个数的正弦值
plot(xx,z,col="red",type="l")   #画图
```



R 语言快速入门

本课程主要讲解如何应用 R 语言进行量化投资分析, 所以在此重点介绍如何将收集到的数据读入 R 语言中, 以便进行数据分析. 下面介绍几种简单的读入和分析数据的方法.

一, 单变量数据分析

二, 多变量数据分析

三, 变量名的解析(绑定)

四, 软件包介绍及在线获取数据

一, 单变量数据分析

1, 输入变量数据

如果是单变量(向量)数据, 数据量也不是很大, 最简单直观的方法就是在 R 语言中直接输入数据. 如我们要了解 1993-2013 年我国税收收入(单位:百亿元)情况, 可用向量函数 c 命令直接输入."Ctrl+R" 执行语句.

```
X=c(42.55,51.27,60.38,69.10,82.34,92.63,
+ 106.83,125.82,153.01,176.36,200.17,241.66,
+ 287.79,348.04,456.22,542.24,595.22,732.11,
+ 897.38,1006.14,1105.31)
```

2, 显示变量数据

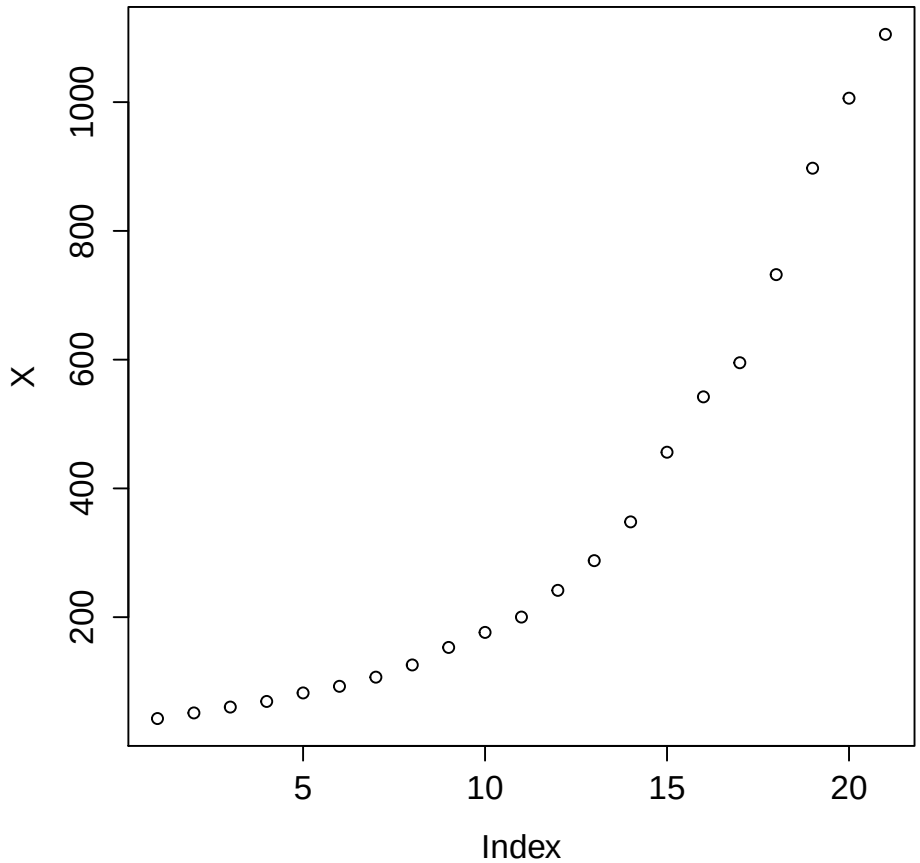
```
X   #R 语言是用变量名来显示数据的, 等价于 print(X)
```

```
 [1]   42.55   51.27   60.38   69.10   82.34   92.63  106.83  125.82  153.01  176.36
[11]  200.17  241.66  287.79  348.04  456.22  542.24  595.22  732.11  897.38 1006.14
[21] 1105.31
```

一，单变量数据分析

3，图示数据变量

`plot(X)`



一, 单变量数据分析

4, 定义数据为时间序列

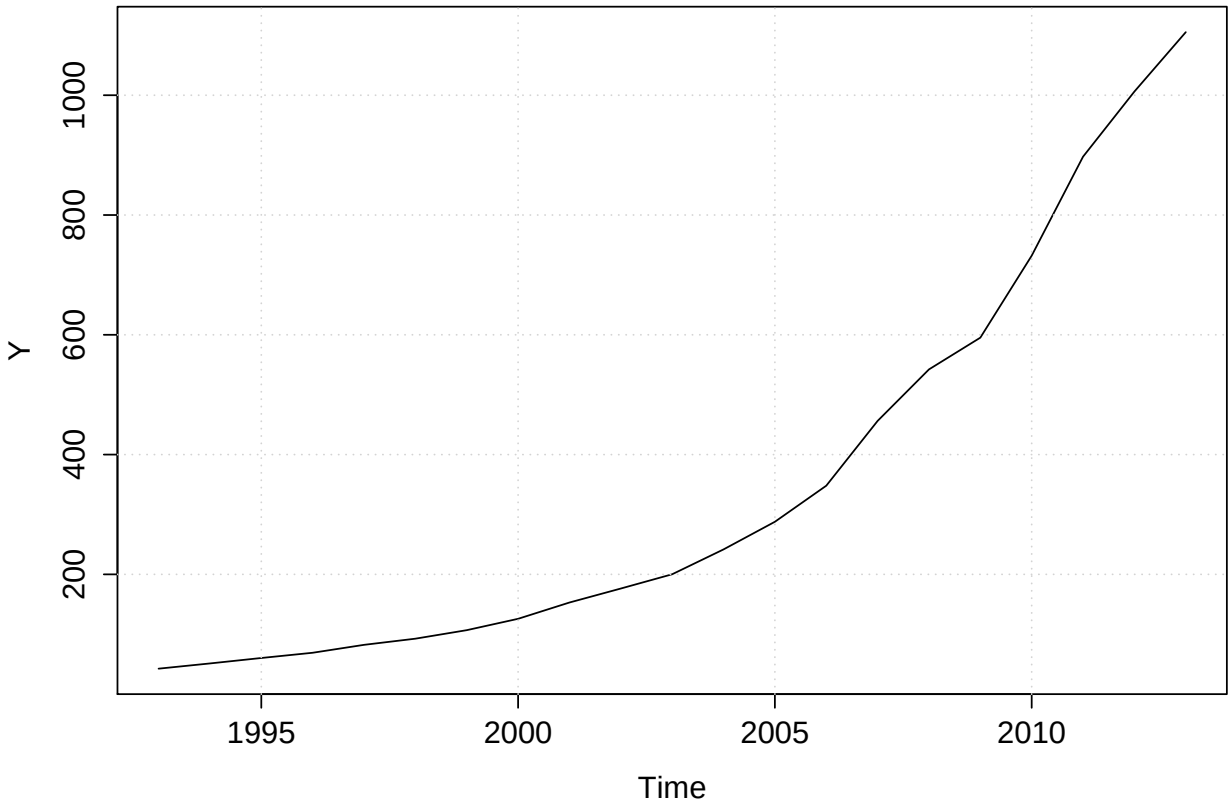
由于金融经济中的数据大多是以时间序列形式出现, 所以我们在分析前需将其转化为时间序列格式. R 语言中用于转化数据向量为时间序列格式的命令很简单, 用 ts 命令即可生产时间序列.

```
Y=ts(X,start=1993);Y   #Y=ts(X,start=1993,end=2013)
```

```
Time Series:
Start = 1993
End = 2013
Frequency = 1
 [1]   42.55   51.27   60.38   69.10   82.34   92.63  106.83  125.82  153.01  176.36
[11]  200.17  241.66  287.79  348.04  456.22  542.24  595.22  732.11  897.38 1006.14
[21] 1105.31
```

一，单变量数据分析

```
plot(Y);grid()      #grid 命令可给图形加网格线
```



二, 多变量数据分析

1, 读取多变量数据

对多变量来说, 大量数据对象常常是从外部文件读入, 而不是在 R 软件中直接键入. 外部的数据源很多, 可以是文本文件, 电子表格, 数据库等. 我们介绍一组中国宏观经济数据 "MEdata", MEdata 收集了 1978-2013 年我国 8 个宏观经济指标:

- TAX: 税收收入
- GDP: 国内生产总值
- IE: 进出口额
- EXP: 财政支出
- RS: 社会消费品零售总额
- COM: 城乡居民消费额
- INV: 全社会固定资产投资总额
- DEP: 城乡居民存款年底余额

二, 多变量数据分析

数据的前几行见下表:(单位: 百亿元. 这些数据来自 Wind 咨询)

	TAX	GDP	EXP	IE	RS	COM	INV	DEP
1978	5.1928	36.056	11.2209	3.550	15.586	17.591	8.008	2.106
1979	5.3782	40.926	12.8179	4.546	18.000	20.115	8.565	2.810
1980	5.7170	45.929	12.2883	5.700	21.400	23.312	9.109	3.958
1981	6.2989	50.088	11.3841	7.353	23.500	26.279	9.610	5.237
1982	7.0002	55.900	12.2998	7.713	25.700	29.029	12.304	6.754
1983	7.7559	62.162	14.0952	8.601	28.494	32.311	14.301	8.925

MEdata 数据可存为 Excel 文件格式, 也可以存为无格式的 txt 文本文件和 csv 逗号文件格式. 三种格式的数据可以通过联网下载:

- Excel 格式: http://eclab.jnu.edu.cn/stat/ME1978-2013.xls
- 文本格式: http://eclab.jnu.edu.cn/stat/MEdata.txt
- 逗号格式: http://eclab.jnu.edu.cn/stat/MEdata.csv

二, 多变量数据分析

- 从文本文件读取

  读入文本数据的命令是 read.table, 但它对外部文件常常有特定的格式要求: 第一行可以有该数据框的各变量名, 随后的行中的条目是各个变量的值. 如 "MEdata.txt" 文本数据,

```
setwd("E:/WorkShop/R")    #指定工作路径
MEdata=read.table("MEdata.txt")     #将数据存入 MEdata 变量
head(MEdata)     #查看所加载数据的前几行
#MEdata=read.table("http://eclab.jnu.edu.cn/stat/MEdata.txt",header=T)
```

- 读取电子表格数据

  虽然 R 语言可以直接读取 Excel 数据, 但最好一次只读 Excel 工作薄中的一个表格. 可以先把电子表格保存为 "csv" 格式, 再利用 "read.csv" 读入数据.

```
MEdata=read.csv("MEdata.csv")     # 本地目录下可不写路径
head(MEdata)     #查看所加载数据的前几行
#MEdata=read.csv("http://eclab.jnu.edu.cn/stat/MEdata.csv",header=T)
```

二, 多变量数据分析

2, 数据集(框)的基本信息

```
names(MEdata)     #显示变量名
[1] "TAX" "GDP" "EXP" "IE"  "RS"  "COM" "INV" "DEP"

nrow(MEdata)      #数据集行数
[1] 36

ncol(MEdata)      #数据集列数
[1] 8
```

二, 多变量数据分析

3, 数据的基本统计量

summary(MEdata)

```
      TAX                GDP                EXP                IE
 Min.   :   5.193   Min.   :  36.06   Min.   :  11.22   Min.   :   3.55
 1st Qu.:  21.279   1st Qu.: 118.35   1st Qu.:  22.48   1st Qu.:  29.58
 Median :  64.739   Median : 686.90   Median :  73.81   Median : 238.17
 Mean   : 211.788   Mean   :1281.44   Mean   : 257.73   Mean   : 611.19
 3rd Qu.: 253.189   3rd Qu.:1675.73   3rd Qu.: 298.48   3rd Qu.:1008.85
 Max.   :1105.307   Max.   :5866.73   Max.   :1402.12   Max.   :2582.53
       RS                 COM                INV                DEP
 Min.   :  15.59   Min.   :  17.59   Min.   :   8.008   Min.   :   2.106
 1st Qu.:  56.02   1st Qu.:  59.20   1st Qu.:  36.239   1st Qu.:  28.707
 Median : 259.87   Median : 311.63   Median : 214.664   Median : 340.916
 Mean   : 494.41   Mean   : 492.79   Mean   : 710.894   Mean   : 894.937
 3rd Qu.: 614.20   3rd Qu.: 671.54   3rd Qu.: 750.514   3rd Qu.:1249.293
 Max.   :2378.10   Max.   :2121.88   Max.   :4470.744   Max.   :4607.850
```

二, 多变量数据分析

还可以通过 fBasics 包中的 basicStats()函数得到更多的统计量：

```
install.packages("fBasics")    #在线安装
library(fBasics)
basicStats(MEdata)
basicStats(MEdata[,1:3])
```

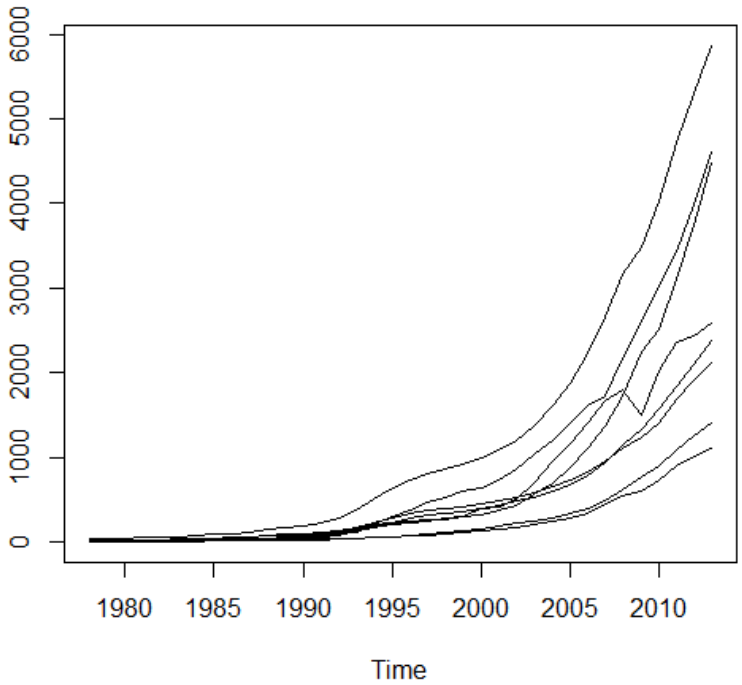
	TAX	GDP	EXP
nobs	36.000000	3.600000e+01	3.600000e+01
NAs	0.000000	0.000000e+00	0.000000e+00
Minimum	5.192800	3.605600e+01	1.122090e+01
Maximum	1105.307000	5.866730e+03	1.402121e+03
1. Quartile	21.279525	1.183517e+02	2.247863e+01
3. Quartile	253.188950	1.675733e+03	2.984774e+02
Mean	211.787925	1.281441e+03	2.577275e+02
Median	64.739300	6.869025e+02	7.380635e+01
Sum	7624.365300	4.613188e+04	9.278190e+03
SE Mean	50.671134	2.695533e+02	6.284107e+01
LCL Mean	108.920054	7.342189e+02	1.301533e+02
UCL Mean	314.655796	1.828663e+03	3.853017e+02
Variance	92432.297817	2.615722e+06	1.421640e+05
Stdev	304.026804	1.617320e+03	3.770464e+02
Skewness	1.635337	1.429076e+00	1.712274e+00
Kurtosis	1.527084	9.368660e-01	1.837013e+00

二, 多变量数据分析

4, 图形显示

在做时间序列图前, 最好用命令 ts 将数据变成时间序列格式, 否则做出的图通常不会显示时间坐标.

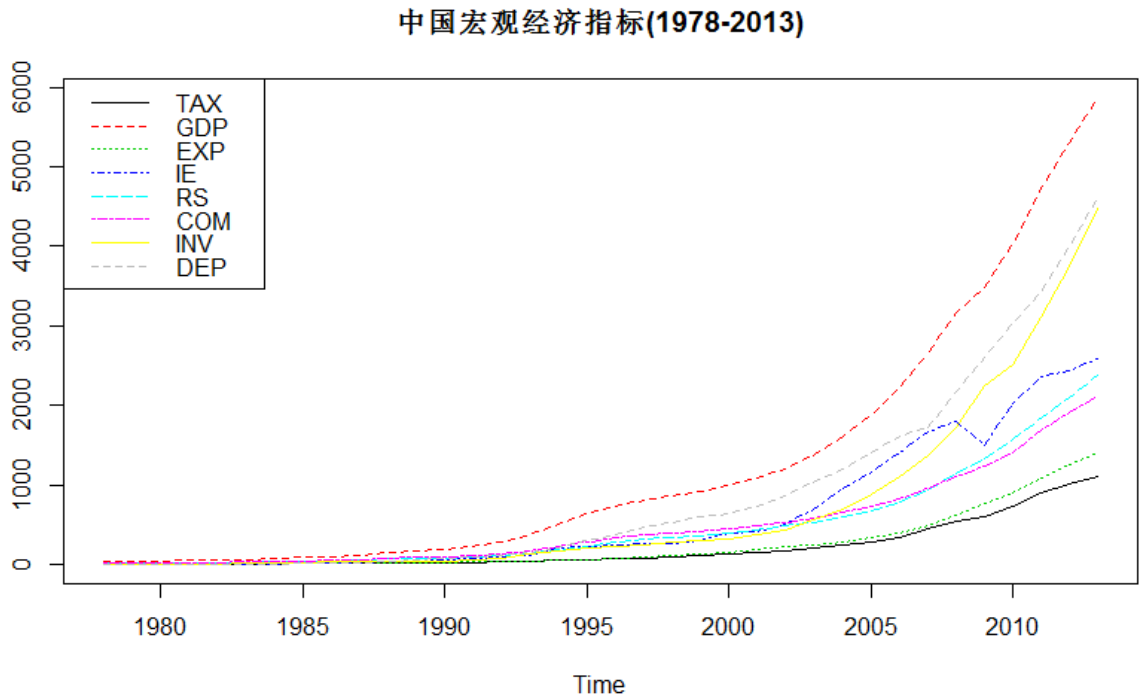
```
tsME=ts(MEdata,start=1978)      #变量序列化
ts.plot(tsME)      #等价于 plot(tsME,plot.type='single')
```



二, 多变量数据分析

我们可以进一步美化图形:

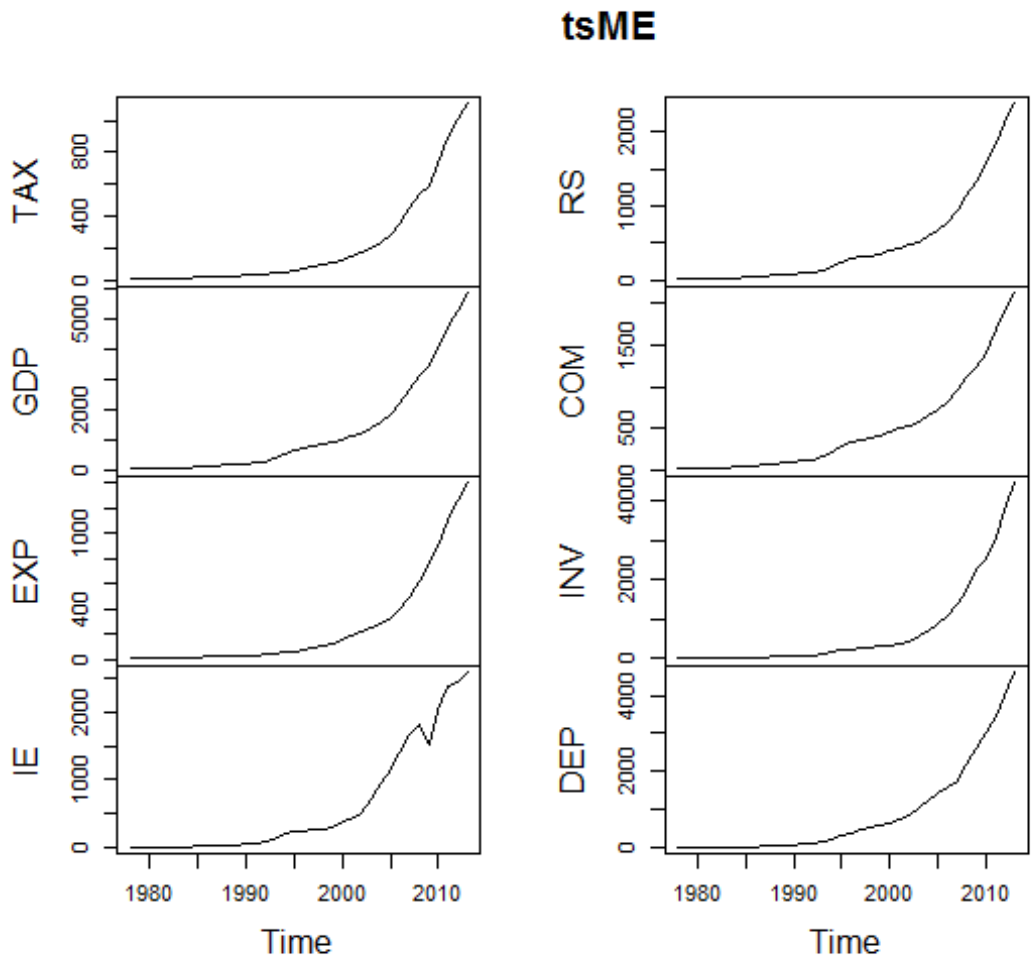
```
ts.plot(tsME, main="中国宏观经济指标(1978-2013)",
        lty=c(1:8),col=c(1:8))
legend("topleft",legend=names(MEdata),
        lty=c(1:8),col=c(1:8))
```



二, 多变量数据分析

也可以一起做出单变量序列图:

```
plot.ts(tsME)      #等价于 plot(tsME,plot.type='multiple')
```



二, 多变量数据分析

5, 拟合线性回归模型

```
LM=lm(TAX~GDP,data=MEdata); LM
```

```
Call:
lm(formula = TAX ~ GDP, data = MEdata)

Coefficients:
(Intercept)          GDP
   -28.0073       0.1871
```

二, 多变量数据分析

6, 线性回归模型检验

summary(LM)

```
Call:
lm(formula = TAX ~ GDP, data = MEdata)

Residuals:
    Min      1Q  Median      3Q     Max
-42.459 -29.555   7.861  24.982  43.489

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.007296   6.274253  -4.464 8.42e-05 ***
GDP           0.187129   0.003067  61.015  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29.35 on 34 degrees of freedom
Multiple R-squared: 0.9909, Adjusted R-squared: 0.9907
F-statistic:  3723 on 1 and 34 DF,  p-value: < 2.2e-16
```

二, 多变量数据分析

```
LM1=lm(TAX~.,data = MEdata)
summary(LM1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6766 -1.2977 -0.0722  2.0662  5.7986

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.965284   1.160703  -4.278 0.000199 ***
GDP         -0.014404   0.017557  -0.820 0.418889
EXP          0.702843   0.046283  15.186 4.80e-15 ***
IE           0.090401   0.008570  10.549 2.92e-11 ***
RS           0.211891   0.065535   3.233 0.003130 **
COM         -0.029802   0.036187  -0.824 0.417158
INV          0.003439   0.008662   0.397 0.694394
DEP         -0.104702   0.010208 -10.257 5.48e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.21 on 28 degrees of freedom
Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
F-statistic: 4.484e+04 on 7 and 28 DF,  p-value: < 2.2e-16
```

二, 多变量数据分析

```
step(LM1)  #逐步回归

Start:  AIC=90.93
TAX ~ GDP + EXP + IE + RS + COM + INV + DEP

       Df Sum of Sq     RSS     AIC
- INV   1      1.62  290.18  89.131
- GDP   1      6.94  295.49  89.784
- COM   1      6.99  295.54  89.791
<none>              288.55  90.929
- RS    1    107.73  396.29 100.350
- DEP   1   1084.20 1372.76 145.078
- IE    1   1146.75 1435.30 146.682
- EXP   1   2376.51 2665.07 168.961
```

二, 多变量数据分析

```
Step:  AIC=89.13
TAX ~ GDP + EXP + IE + RS + COM + DEP

       Df Sum of Sq     RSS     AIC
- GDP   1      6.52  296.70  87.931
- COM   1     11.98  302.16  88.587
<none>              290.18  89.131
- RS    1    122.19  412.37  99.782
- DEP   1   1085.40 1375.58 143.152
- IE    1   1166.70 1456.88 145.219
- EXP   1   2734.70 3024.88 171.520

Step:  AIC=87.93
TAX ~ EXP + IE + RS + COM + DEP

       Df Sum of Sq    RSS     AIC
- COM   1      13.9  310.6  87.583
<none>              296.7  87.931
- RS    1     144.9  441.6 100.244
- DEP   1    1079.5 1376.2 141.169
- EXP   1    2858.4 3155.1 171.038
- IE    1    4053.5 4350.2 182.601
```

二, 多变量数据分析

```
Step:  AIC=87.58
TAX ~ EXP + IE + RS + DEP

       Df Sum of Sq     RSS     AIC
<none>                 310.6  87.583
- RS    1     514.7   825.3 120.761
- DEP   1    1173.1  1483.7 141.876
- IE    1    5333.4  5644.0 189.974
- EXP   1   18993.6 19304.3 234.244

Call:
lm(formula = TAX ~ EXP + IE + RS + DEP, data = MEdata)

Coefficients:
(Intercept)          EXP           IE           RS          DEP
   -5.22672      0.76031      0.08137      0.13355     -0.10582
```

三, 变量名解析(绑定)

在进行数据分析时, 变量通常保存在数据框中, 例如上面的 MEdata, 要使用其中的变量通常需要 $ 命令, 例如需要使用 TAX, 需采用命令 MEdata$TAX. 如果直接使用 MEdata 的所有变量名, 可用 attach() 函数对变量进行解析.

```
TAX    #未解析前
```

错误: 找不到对象'TAX'

```
attach(MEdata)  #变量解析，MEdata中的变量名可单独使用
TAX   #解析后
```

```
 [1]    5.1928    5.3782    5.7170    6.2989    7.0002    7.7559    9.4735   20.4079
 [9]   20.9073   21.4036   23.9047   27.2740   28.2186   29.9017   32.9691   42.5530
[17]   51.2688   60.3804   69.0982   82.3404   92.6280  106.8258  125.8151  153.0138
[25]  176.3645  200.1731  241.6568  287.7854  348.0435  456.2197  542.2379  595.2159
[33]  732.1079  897.3839 1006.1428 1105.3070
```

在使用完这些变量后, 最好将这些变量去除绑定, 避免跟其他变量冲突. 变量名去绑定的命令为 detach(), 用 detach(MEdata) 后 MEdata 中的变量就不可单独使用了!

四, 软件包介绍及在线获取数据

R 语言力量源泉: 软件包. 截止到(2016.09.28)从 R 的官方主页: https://cran.r-project.org/ 可以看到, 可供使用的 R 软件包有 9249 个.

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

Available CRAN Packages By Name

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
ABCanalysis	Computed ABC Analysis
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests
abctools	Tools for ABC Analyses
abd	The Analysis of Biological Data
abf2	Load Gap-Free Axon ABF2 Files
ABHgenotypeR	Easy Visualization of ABH Genotypes
abind	Combine Multidimensional Arrays
abn	Modelling Multivariate Data with Additive Bayesian Networks
abodOutlier	Angle-Based Outlier Detection
AbsFilterGSEA	Improved False Positive Control of Gene-Permuting GSEA with Absolute Filtering
abundant	Abundant regression and high-dimensional principal fitted components
ACA	Abrupt Change-Point or Aberration Detection in Point Series

四, 软件包介绍及在线获取数据

R 中常用于量化投资的包有: xts, quantmod, TTR 等, xts(扩展的时间序列包)是常用于处理时间序列的包. 我们以上证综指收盘指数时序图为例简单介绍一下 xts 的常用函数.

```
Index<-read.table("TRD_Index.txt",header=TRUE)
tail(Index,3)
```

```
     Indexcd      Trddt Daywk Opnindex  Hiindex  Loindex Clsindex Retindex
4663  399903 2015/4/10     5 3922.880 4007.416 3903.789 4000.756 0.016780
4664  399903 2015/4/13     1 4057.985 4087.377 4031.636 4076.697 0.018982
4665  399903 2015/4/14     2 4089.519 4132.088 4049.741 4098.419 0.005328
```

四, 软件包介绍及在线获取数据

```
SHindex<-Index[Index$Indexcd==1,]
head(SHindex,3)
```

```
    Indexcd      Trddt Daywk Opnindex  Hiindex  Loindex Clsindex Retindex
309       1 2015/4/10     5 3947.492 4040.348 3929.319 4034.310 0.019400
310       1 2015/4/13     1 4072.723 4128.072 4057.293 4121.715 0.021665
311       1 2015/4/14     2 4125.782 4168.346 4091.257 4135.565 0.003360
```

- Indexcd：指数代码（Index code），000001 表示上证综合指数；
- Trddt：交易日期（Trading date），以 YYYY-MM-DD 表示；
- Daywk：星期，1= 星期一，2= 星期二，3= 星期三，4= 星期四，5= 星期五，6= 星期六，0= 星期天；
- Opnindex：开盘指数（Open index），每日交易中的第一条指数；
- Hiindex：最高指数（Highest index），每日交易中的最高一条指数；
- Loindex：最低指数（Lowest index），每日交易中的最低一条指数；
- Clsindex：收盘指数（Closed index），每日交易中的最后一条指数；
- Retindex：指数收益率（Index return）。

四, 软件包介绍及在线获取数据

```
#安装和加载 xts 包
install.packages("xts")  #在线安装
library(xts)  #加载xts包
Clsindex<-SHindex$Clsindex  #提取Clsindex
head(Clsindex)
```

```
[1] 2109.387 2083.136 2045.709
```

```
#将收盘指数转换成时间序列格式
Clsindex<-xts(Clsindex, order.by=as.Date(SHindex$Trddt))
```

```
                [,1]
2014-01-02 2109.387
2014-01-03 2083.136
2014-01-06 2045.709
2014-01-07 2047.317
2014-01-08 2044.340
2014-01-09 2027.622
```

四, 软件包介绍及在线获取数据

```
SHindex<-xts(SHindex[,-1],order.by=as.Date(SHindex$Trddt))
head(SHindex,3)
```

```
           Trddt      Daywk Opnindex   Hiindex    Loindex    Clsindex
2014-01-02 "2014/1/2" "4"   "2112.126" "2113.110" "2101.016" "2109.387"
2014-01-03 "2014/1/3" "5"   "2101.542" "2102.167" "2075.899" "2083.136"
2014-01-06 "2014/1/6" "1"   "2078.684" "2078.684" "2034.006" "2045.709"
           Retindex
2014-01-02 "-0.003115"
2014-01-03 "-0.012445"
2014-01-06 "-0.017967"
```

```
SHindex<-xts(SHindex[,-(1:2)],order.by=as.Date(SHindex$Trddt))
head(SHindex,3)
```

```
           Daywk Opnindex  Hiindex  Loindex Clsindex  Retindex
2014-01-02     4 2112.126 2113.110 2101.016 2109.387 -0.003115
2014-01-03     5 2101.542 2102.167 2075.899 2083.136 -0.012445
2014-01-06     1 2078.684 2078.684 2034.006 2045.709 -0.017967
```

四, 软件包介绍及在线获取数据

选取特定日期的时间序列数据.

```
#截取2014年10月8日到2014年11月1日的数据
SHindexPart<-SHindex["2014-10-08/2014-11-01"]
```

```
            Daywk Opnindex  Hiindex  Loindex Clsindex  Retindex
2014-10-08      3 2368.576 2382.794 2354.290 2382.794  0.008006
2014-10-09      4 2383.859 2391.348 2367.111 2389.371  0.002760
2014-10-10      5 2380.755 2386.277 2365.075 2374.540 -0.006207
```

```
#截取2015年数据
SHindex2015<-SHindex["2015"]
head(SHindex2015,2)
```

```
            Daywk Opnindex  Hiindex  Loindex Clsindex Retindex
2015-01-05      1 3258.627 3369.281 3253.883 3350.519 0.035813
2015-01-06      2 3330.799 3394.224 3303.184 3351.446 0.000277
```

四, 软件包介绍及在线获取数据

```
#截取2015年以后的数据
SHindexAfter2015<-SHindex["2015/"]
head(SHindexAfter2015,2)
```

```
           Daywk Opnindex  Hiindex  Loindex Clsindex Retindex
2015-01-05     1 3258.627 3369.281 3253.883 3350.519 0.035813
2015-01-06     2 3330.799 3394.224 3303.184 3351.446 0.000277
```

```
#截取2015年以前的数据
SHindexBefore2015<-SHindex["/2015-01-05"]
tail(SHindexBefore2015,2)
```

```
           Daywk Opnindex  Hiindex  Loindex Clsindex Retindex
2014-12-31     3 3172.597 3239.357 3157.259 3234.677 0.021752
2015-01-05     1 3258.627 3369.281 3253.883 3350.519 0.035813
```

四, 软件包介绍及在线获取数据

下面介绍常用的 quantmod 软件包. 安装和加载 quantmod 的过程如下:

```
install.packages("quantmod")    #在线安装
library(quantmod)    #加载软件包
```

quantmod 加载以后, 就可以使用里面的函数了. 例如 getSymbols():

```
getSymbols("MSFT")    #在线抓取MSFT股票数据
tail(MSFT)    #查看数据的最后几行
```

	MSFT.Open	MSFT.High	MSFT.Low	MSFT.Close	MSFT.Volume	MSFT.Adjusted
2016-09-16	57.63	57.63	56.75	57.25	44607000	57.25
2016-09-19	57.27	57.75	56.85	56.93	20937100	56.93
2016-09-20	57.35	57.35	56.75	56.81	17384000	56.81
2016-09-21	57.51	57.85	57.08	57.76	33707300	57.76
2016-09-22	57.92	58.00	57.63	57.82	19822200	57.82
2016-09-23	57.87	57.91	57.38	57.43	19825300	57.43

四, 软件包介绍及在线获取数据

```
getSymbols("MSFT",from="2010-1-1",to="2015-12-31")
tail(MSFT)    #查看数据的最后几行
```

	MSFT.Open	MSFT.High	MSFT.Low	MSFT.Close	MSFT.Volume	MSFT.Adjusted
2015-12-23	55.70	55.88	55.44	55.82	27279800	54.69622
2015-12-24	55.86	55.96	55.43	55.67	9570000	54.54924
2015-12-28	55.35	55.95	54.98	55.95	22458300	54.82361
2015-12-29	56.29	56.85	56.06	56.55	27731400	55.41152
2015-12-30	56.47	56.78	56.29	56.31	21704500	55.17636
2015-12-31	56.04	56.19	55.42	55.48	27334100	54.36307

四, 软件包介绍及在线获取数据

我们还可以使用 saveSymbols()将抓取的数据保存到本地硬盘上:

```
saveSymbols("MSFT",file.path="E://WorkShop//R")
```

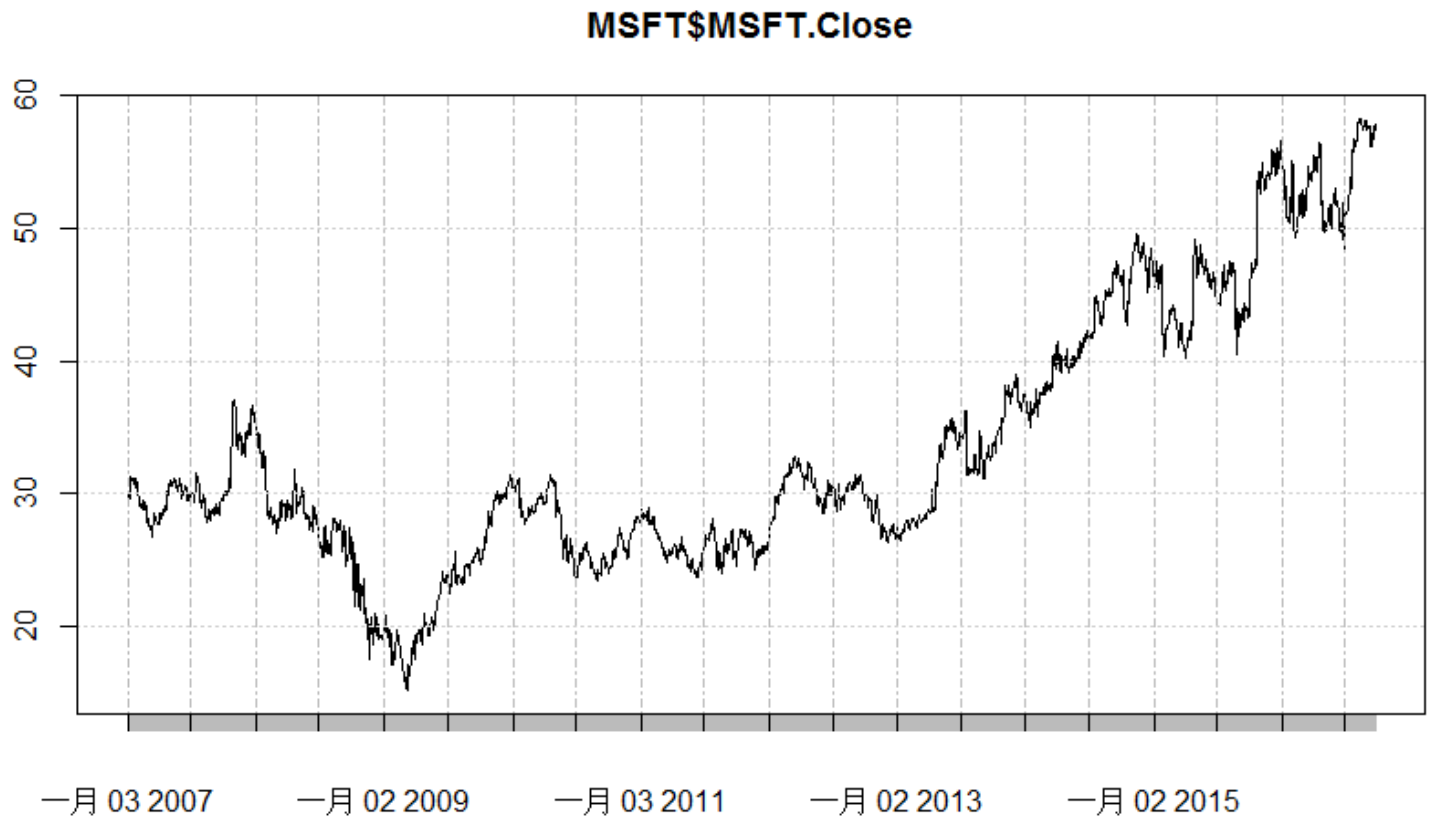
该命令将 MSFT 数据存盘为MSFT.RData 文件, 存入的路径为"E:/WorkShop/R"子目录中. 当需要应用时, 我们可以使用普通的文件加载方法也可以使用 getSymbols() 重新加载

```
getSymbols("MSFT",src="RData",dir="E://WorkShop//R",extension="RData")
```

四, 软件包介绍及在线获取数据

另一个简单的指令就可以绘制出 MSFT 的收盘价走势图

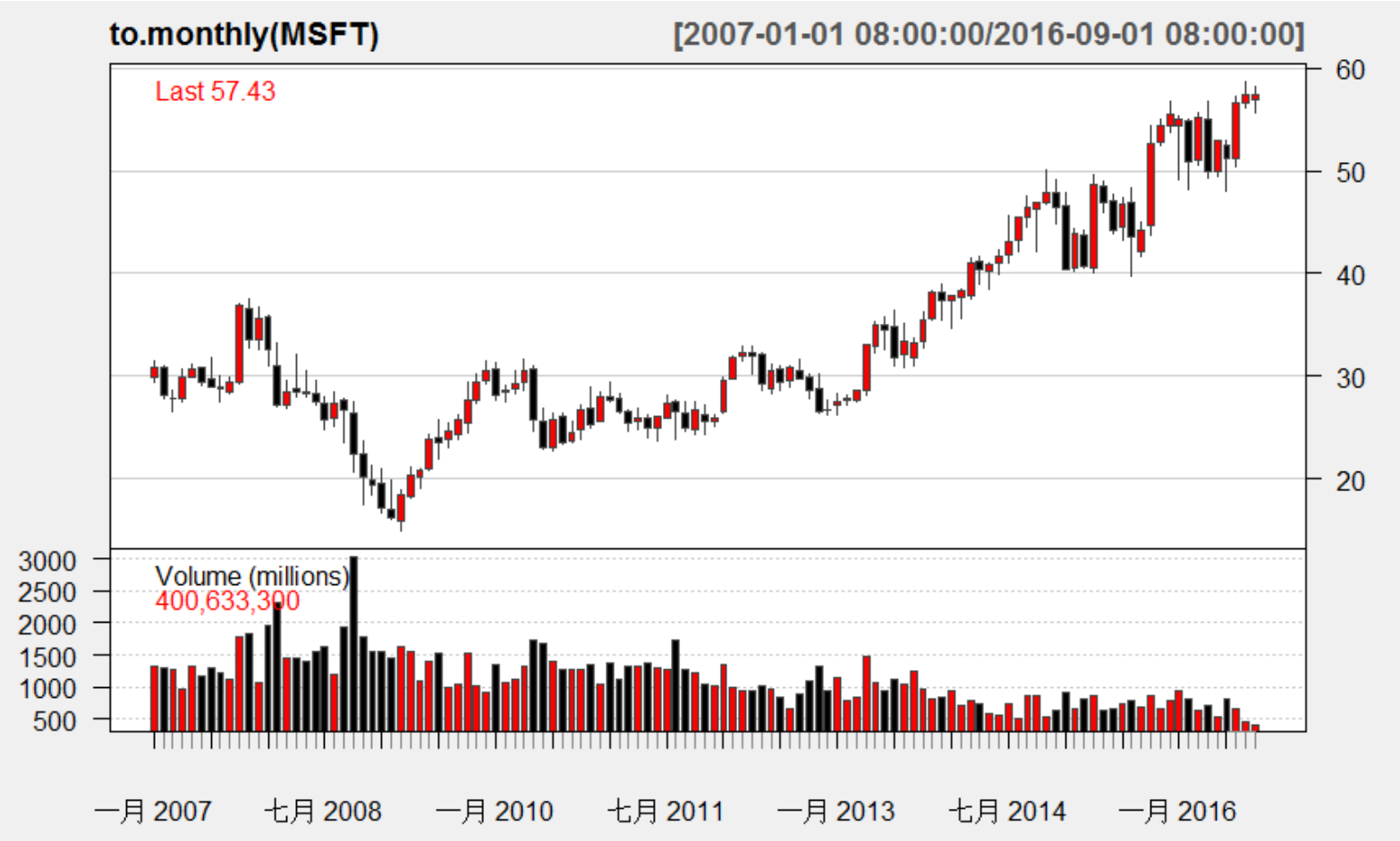
```
plot(MSFT$MSFT.Close)     #画出MSFT收盘价走势图
```



四, 软件包介绍及在线获取数据

股票数据可以非常容易地按传统的习惯可视化, 例如, K 线图:

```
candleChart(to.monthly(MSFT),theme="white",up.col="red",dn.col="black")
```



四, 软件包介绍及在线获取数据

国内股市的股票信息也可以使用这种形式获取. 雅虎上的部分中国股票代码(符号):

上海证交所		深圳证交所	
综合指数	^SSEC	营建指数	^SZCN
A 股指数	^SSEA	金融指数	^SZFI
B 股指数	^SSEB	IT 指数	^SZIT
工业指数	^SSEI	传媒业指数	^SZME
商业指数	^SSEM	电子工业指数	^SZMEL
地产指数	^SSEP	制造业指数	^SZMF
公用事业	^SSEU	食品指数	^SZMFB
30 指数	^SSE30	采矿指数	^SZMI
基金指数	^SSFD	机器指数	^SZMMC
		金属业指数	^SZMMT

四, 软件包介绍及在线获取数据

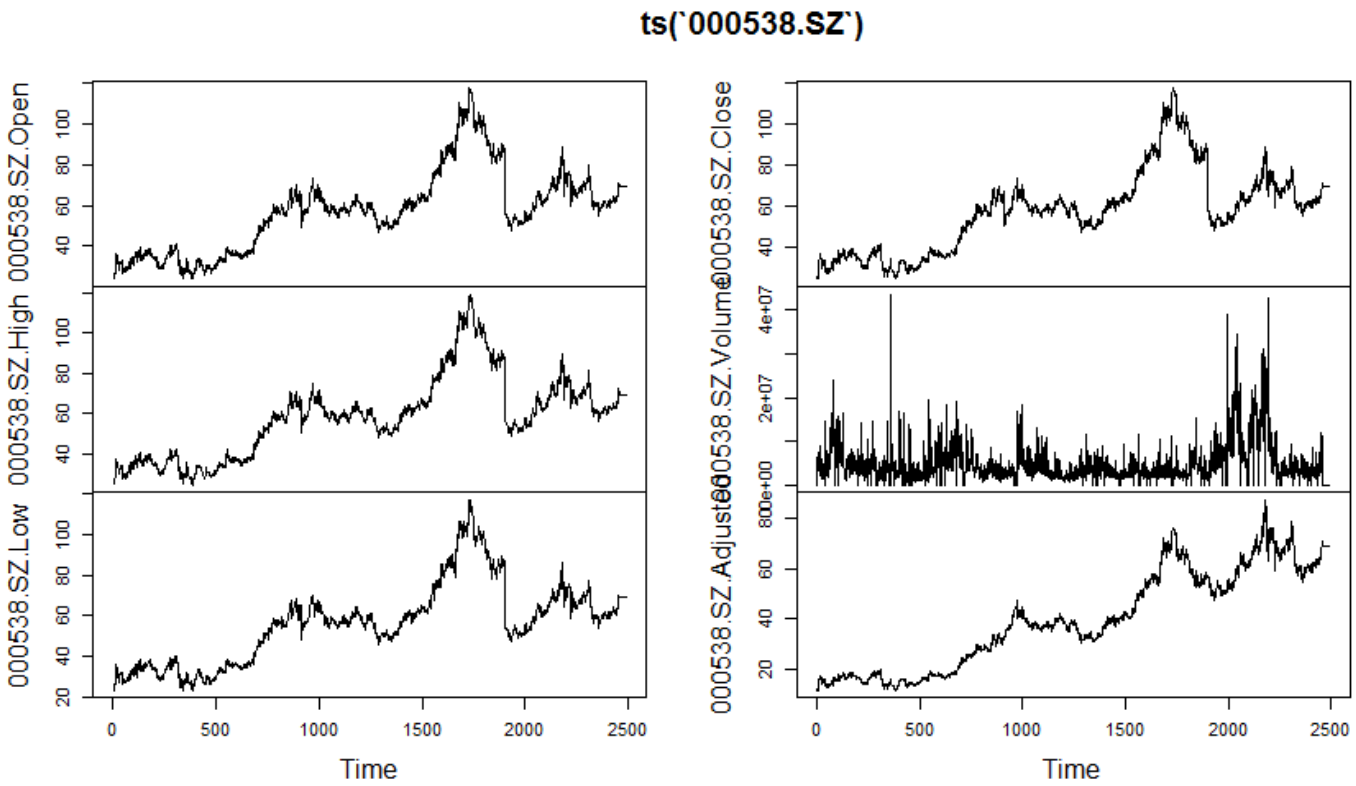
续表:

深圳证交所		石油化工指数	^SZMPC
深圳综合	^SZSC	药品业指数	^SZMPH
深圳成分	^SZSC1	造纸指数	^SZMPP
A 股指数	^SZSA	纺织业指数	^SZMTA
A 股成分	^SZSA1	木材业指数	^SZTF
B 股指数	^SZSB	地产指数	^SZRE
B 股成分	^SZSB1	服务业指数	^SZSO
基金指数	^SZSE	运输指数	^SZTP
农业指数	^SZAG	公用事业指数	^SZUT
集团指数	^SZCM	批发零售指数	^SZWR

四，软件包介绍及在线获取数据

读取股票市场的交易数据时，还可以使用股票的代码，例如，深圳证交所的云南白药代码为 000538，可以使用"000538.SZ"获取数据

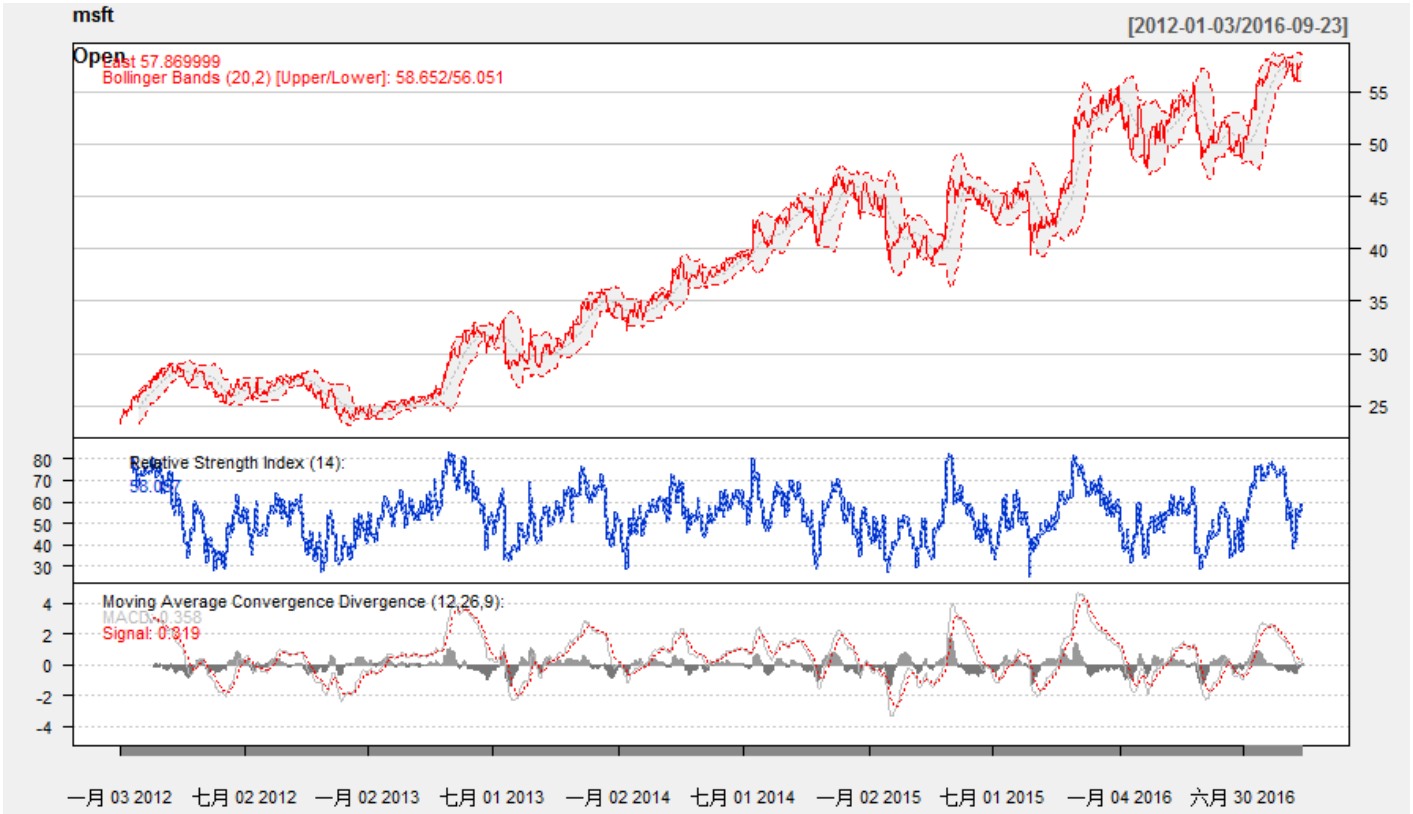
```
getSymbols("000538.SZ")
plot(ts(`000538.SZ`))  #云南白药的xts数据绘图
```



四, 软件包介绍及在线获取数据

quantmod 包还可以在已有图线上叠加技术指标图线.

```
msft=getYahooData("MSFT",start=20120101)
chartSeries(msft[,'Open'],theme="white",up.col="red",dn.col="black")
addBBands()  #增加布林带
addRSI()    #增加相对强弱指标
addMACD()   #增加移动平均线
```



四, 软件包介绍及在线获取数据

quantmod 包还可以在线获取公司的财务报表, 股息数据, 期权交易数据等. 常用的 quantmod 包函数有:

函数	作用	函数	作用
getSymbols()	从多种信息源里获得信息	getSymbols.csv()	从csv文件中读入数据
getDividends()	获取上市公司的股息数据	getSymbols.FRED()	从FRED中获取数据
getFinancials()	获取上市公司的财务报表	getSymbols.google()	从google中获取数据
getFX()	获取汇率数据	getSymbols.MySQL()	从MySQL中获取数据
getMetals()	获取重金属交易数据	getSymbols.oanda()	从oanda中获取数据
getSplits	获取上市公司的拆股数据	getSymbols.rda()	从R的二进制文件中获取数据
getOptionChain()	获取期权交易数据	getSymbols.SQLite()	从SQLite数据库中获取数据
getQuote	获取即时的网络报价	getSymbols.yahoo()	从雅虎网中获取数据

四, 软件包介绍及在线获取数据

除了 quantmod 软件包外, R 还有很多做量化分析的软件包, 例如:

- TTR 包. 主要提供各种技术指标的计算函数, 以及从美国股市和雅虎财经提取数据.
- PerformanceAnalytics 包. 主要用于绩效和风险分析的计量经济工具.
- quantstrast 包. 主要用于策略模型, 设定, 构建和回测检验量化金融交易和投资组合策略.
- blotter 包. 主要为交易系统和交易模拟定义金融工具, 组合与账户等事物基础框架.
- highfrequency 包. 主要用于处理高频数据.
- ……

可参考朱晓斌《量化投资以 R 为工具》,王乐等《金融时间序列预测- 基于 R 语言的应用实践》.