我的第一份R作业

邱飞

3/9/2017

## 介绍：获取数据

这是我的第一份R语言作业，老师让我们对上证指数做一些简单的分析。

首先，是通过Quandl宏包来获取上证指数数据。

下面的api\_key是注册Quandl网站后，网站免费提供的。有了这个代码，可以保证日访问量超过500次。

library(Quandl)

## Loading required package: xts

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

sha = Quandl("GOOG/SHA\_000001", api\_key="9iGeZZoG6Vc46rfs1AgJ")

## 直观感受：数字

我们获得的上证指数数据已经保存在变量sha中，现在我们对它来做些简单的描述性分析。

### 看看数据的前几行长啥样

# the defualt is to see first 6 rows  
head(sha)

## Date Open High Low Close Volume  
## 1 2017-03-08 3240.53 3245.30 3230.61 3240.66 160731388  
## 2 2017-03-07 3233.09 3242.66 3226.82 3242.41 164064235  
## 3 2017-03-06 3217.33 3234.66 3215.07 3233.87 156092158  
## 4 2017-03-03 3219.20 3221.16 3206.61 3218.31 157082368  
## 5 2017-03-02 3250.52 3256.81 3228.66 3230.03 181215076  
## 6 2017-03-01 3240.07 3259.98 3237.87 3246.93 190677550

# if you want see first 10 rows  
head(sha,10)

## Date Open High Low Close Volume  
## 1 2017-03-08 3240.53 3245.30 3230.61 3240.66 160731388  
## 2 2017-03-07 3233.09 3242.66 3226.82 3242.41 164064235  
## 3 2017-03-06 3217.33 3234.66 3215.07 3233.87 156092158  
## 4 2017-03-03 3219.20 3221.16 3206.61 3218.31 157082368  
## 5 2017-03-02 3250.52 3256.81 3228.66 3230.03 181215076  
## 6 2017-03-01 3240.07 3259.98 3237.87 3246.93 190677550  
## 7 2017-02-28 3225.97 3242.68 3225.97 3241.73 151244318  
## 8 2017-02-27 3249.19 3251.65 3224.09 3228.66 182581071  
## 9 2017-02-24 3246.86 3253.96 3233.53 3253.43 186406362  
## 10 2017-02-23 3258.83 3264.08 3236.35 3251.38 209179800

如果你想让现实的结果好看一点，你可以用knitr宏包来美化一下输出结果。

library(knitr)  
simple\_head = head(sha)  
beautiful\_head = kable(simple\_head)  
  
beautiful\_head

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Open | High | Low | Close | Volume |
| 2017-03-08 | 3240.53 | 3245.30 | 3230.61 | 3240.66 | 160731388 |
| 2017-03-07 | 3233.09 | 3242.66 | 3226.82 | 3242.41 | 164064235 |
| 2017-03-06 | 3217.33 | 3234.66 | 3215.07 | 3233.87 | 156092158 |
| 2017-03-03 | 3219.20 | 3221.16 | 3206.61 | 3218.31 | 157082368 |
| 2017-03-02 | 3250.52 | 3256.81 | 3228.66 | 3230.03 | 181215076 |
| 2017-03-01 | 3240.07 | 3259.98 | 3237.87 | 3246.93 | 190677550 |

### 看看数据的结构

str(sha)

## 'data.frame': 665 obs. of 6 variables:  
## $ Date : Date, format: "2017-03-08" "2017-03-07" ...  
## $ Open : num 3241 3233 3217 3219 3251 ...  
## $ High : num 3245 3243 3235 3221 3257 ...  
## $ Low : num 3231 3227 3215 3207 3229 ...  
## $ Close : num 3241 3242 3234 3218 3230 ...  
## $ Volume: num 1.61e+08 1.64e+08 1.56e+08 1.57e+08 1.81e+08 ...  
## - attr(\*, "freq")= chr "daily"

### 看看数据的简单统计

summary(sha)

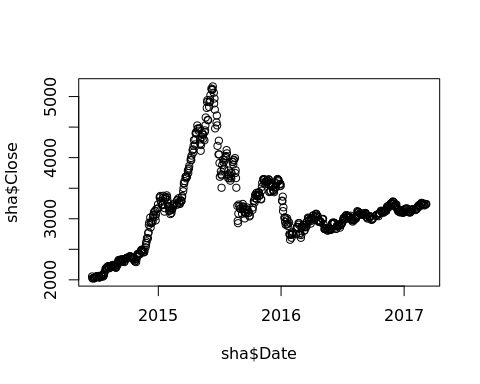
## Date Open High Low   
## Min. :2014-06-18 Min. :2013 Min. :2027 Min. :2011   
## 1st Qu.:2015-02-17 1st Qu.:2882 1st Qu.:2906 1st Qu.:2841   
## Median :2015-10-28 Median :3097 Median :3123 Median :3077   
## Mean :2015-10-25 Mean :3147 Mean :3181 Mean :3115   
## 3rd Qu.:2016-06-30 3rd Qu.:3359 3rd Qu.:3400 3rd Qu.:3322   
## Max. :2017-03-08 Max. :5174 Max. :5178 Max. :5103   
## Close Volume   
## Min. :2024 Min. : 13538758   
## 1st Qu.:2887 1st Qu.:163928625   
## Median :3103 Median :216955942   
## Mean :3154 Mean :278446697   
## 3rd Qu.:3373 3rd Qu.:357080811   
## Max. :5166 Max. :857132807

## 直观感受：图形

R语言一个强大之处在于它的图形表达能力，但是它的默认的图形输出实在不好看，所以需要我们用各种宏包做下美化。

### 默认图形画图

plot(sha$Date,sha$Close)



### ggplot画图

library(ggplot2)  
picture = ggplot(data = sha, aes(x = Date,y = Close))+  
 geom\_line()+  
 geom\_point()  
picture

