R로 타슈데이터 분석하기

R로 타슈 데이터를 분석해보자. 우선 R의 패키지 라이브러리를 읽어들이자. ggplot2는 그래프를 그리는 패키지이고, lubridate는 날짜관련 데이터를 다루는 패키지이다.

```
library("ggplot2")
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

각자 데이터가 있는 디렉토리를 설정하자.

```
setwd("/Users/youngseoklee/Dropbox/python-tashu")
```

csv파일로 저장된 타슈데이터를 tashu 변수에 읽어들이자.

```
tashu <- read.csv(file='2013_01.csv')
```

```
tashu[1:10,]
```

```
##    member rent_station   rent_date return_station return_date
## 1      No           43 2.01301e+13             34 2.01301e+13
## 2      No           97 2.01301e+13             NA 2.01301e+13
## 3      No            2 2.01301e+13             10 2.01301e+13
## 4      No          106 2.01301e+13            105 2.01301e+13
## 5     Yes            4 2.01301e+13              4 2.01301e+13
## 6      No           21 2.01301e+13            105 2.01301e+13
## 7      No           90 2.01301e+13             91 2.01301e+13
## 8      No           13 2.01301e+13             30 2.01301e+13
## 9     Yes            1 2.01301e+13              1 2.01301e+13
## 10    Yes            1 2.01301e+13              2 2.01301e+13
```

```
tashu$rent_date <- strptime(tashu$rent_date, "%Y%m%d%H%M%S")
tashu$return_date <- strptime(tashu$return_date, "%Y%m%d%H%M%S")
head(tashu)
```

```
##   member rent_station           rent_date return_station
## 1     No           43 2013-01-01 05:56:03             34
## 2     No           97 2013-01-01 06:04:00             NA
## 3     No            2 2013-01-01 06:04:06             10
## 4     No          106 2013-01-01 10:53:05            105
## 5    Yes            4 2013-01-01 11:22:23              4
## 6     No           21 2013-01-01 11:39:53            105
##           return_date
## 1 2013-01-01 06:02:17
## 2 2013-01-01 10:20:37
## 3 2013-01-01 06:18:59
## 4 2013-01-01 10:57:43
## 5 2013-01-01 12:17:53
## 6 2013-01-01 11:49:43
```

```
tashu$rent_daily <- format(tashu$rent_date, "%Y%m%d")
head(tashu)
```

```
##   member rent_station           rent_date return_station
## 1     No           43 2013-01-01 05:56:03             34
## 2     No           97 2013-01-01 06:04:00             NA
## 3     No            2 2013-01-01 06:04:06             10
## 4     No          106 2013-01-01 10:53:05            105
## 5    Yes            4 2013-01-01 11:22:23              4
## 6     No           21 2013-01-01 11:39:53            105
##           return_date rent_daily
## 1 2013-01-01 06:02:17   20130101
## 2 2013-01-01 10:20:37   20130101
## 3 2013-01-01 06:18:59   20130101
## 4 2013-01-01 10:57:43   20130101
## 5 2013-01-01 12:17:53   20130101
## 6 2013-01-01 11:49:43   20130101
```

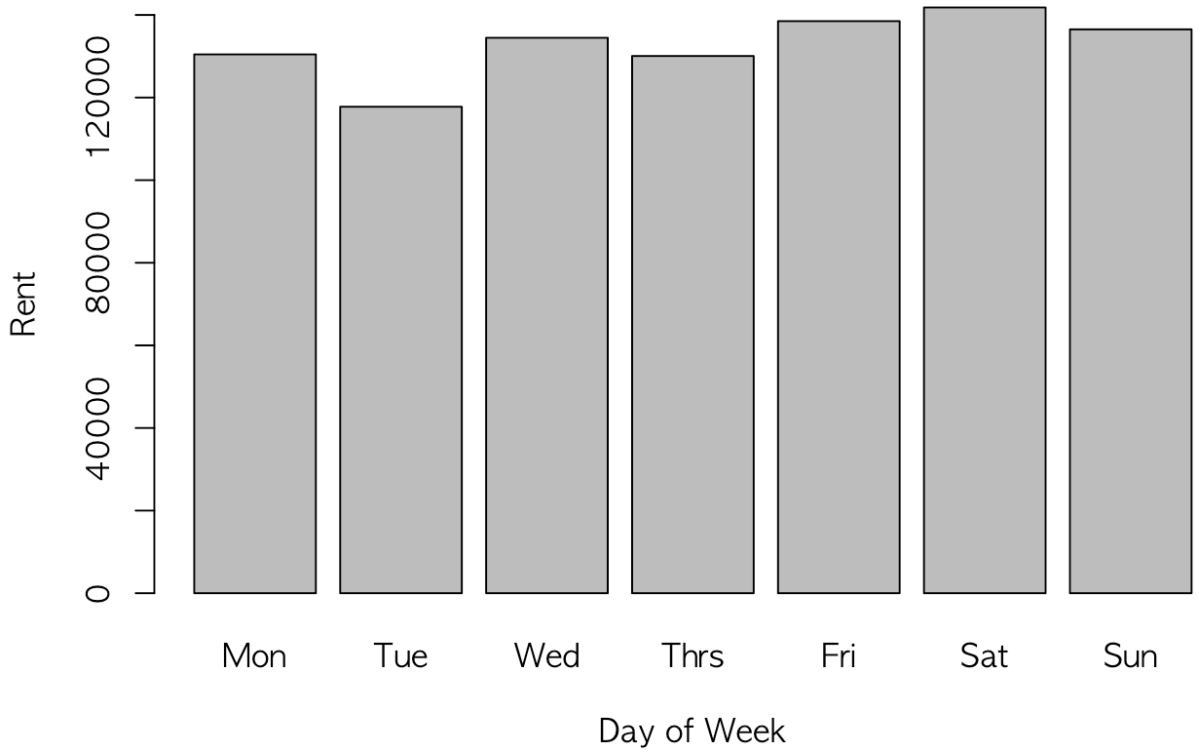
```
tashu$wday <- weekdays(tashu$rent_date)

wday_cnt <- data.frame(table(tashu$wday))
wday_cnt
```

```
##        Var1   Freq
## 1    Friday 138506
## 2    Monday 130451
## 3  Saturday 141804
## 4    Sunday 136509
## 5  Thursday 130066
## 6   Tuesday 117777
## 7 Wednesday 134470
```

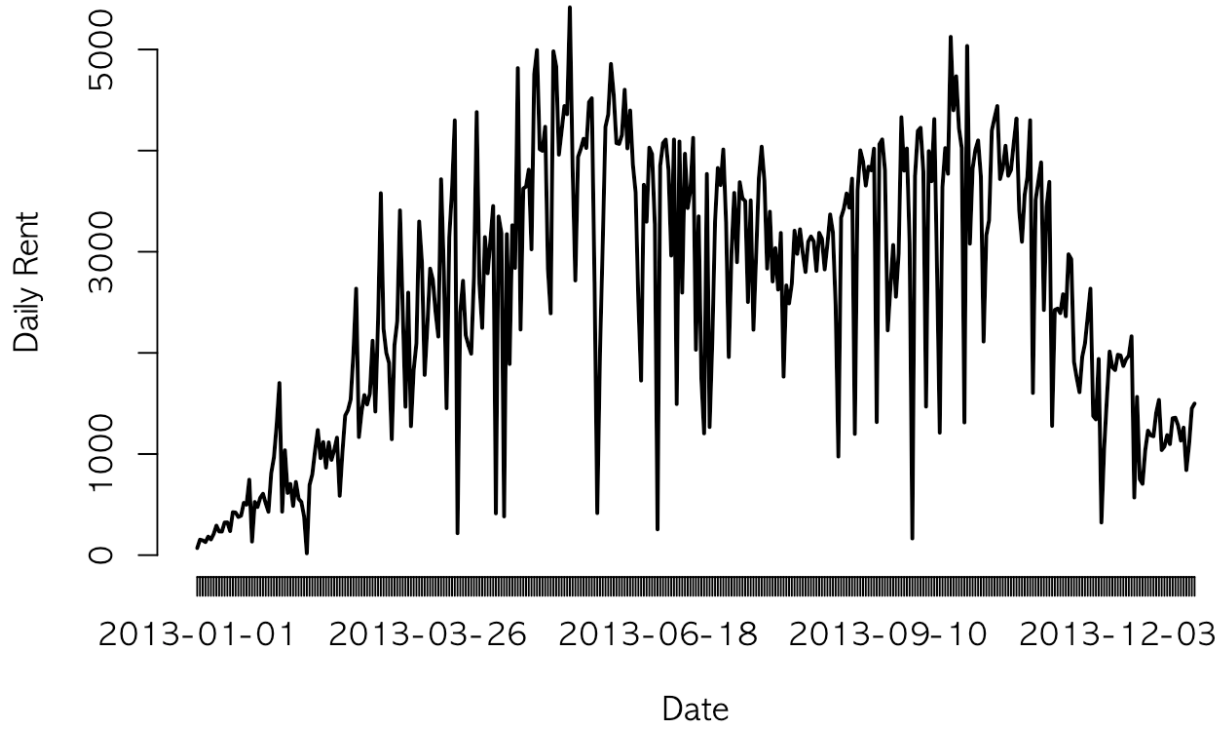
```
wday_cnt$Var1 <- factor(wday_cnt$Var1, levels= c("Monday", "Tuesday", "Wednesday", "T
hursday",
                                                 "Friday", "Saturday", "Sunday"))
wday_cnt2 <- wday_cnt[order(wday_cnt$Var1),]
wday_cnt2
```

```
##        Var1   Freq
## 2    Monday 130451
## 6   Tuesday 117777
## 7 Wednesday 134470
## 5  Thursday 130066
## 1    Friday 138506
## 3  Saturday 141804
## 4    Sunday 136509
```

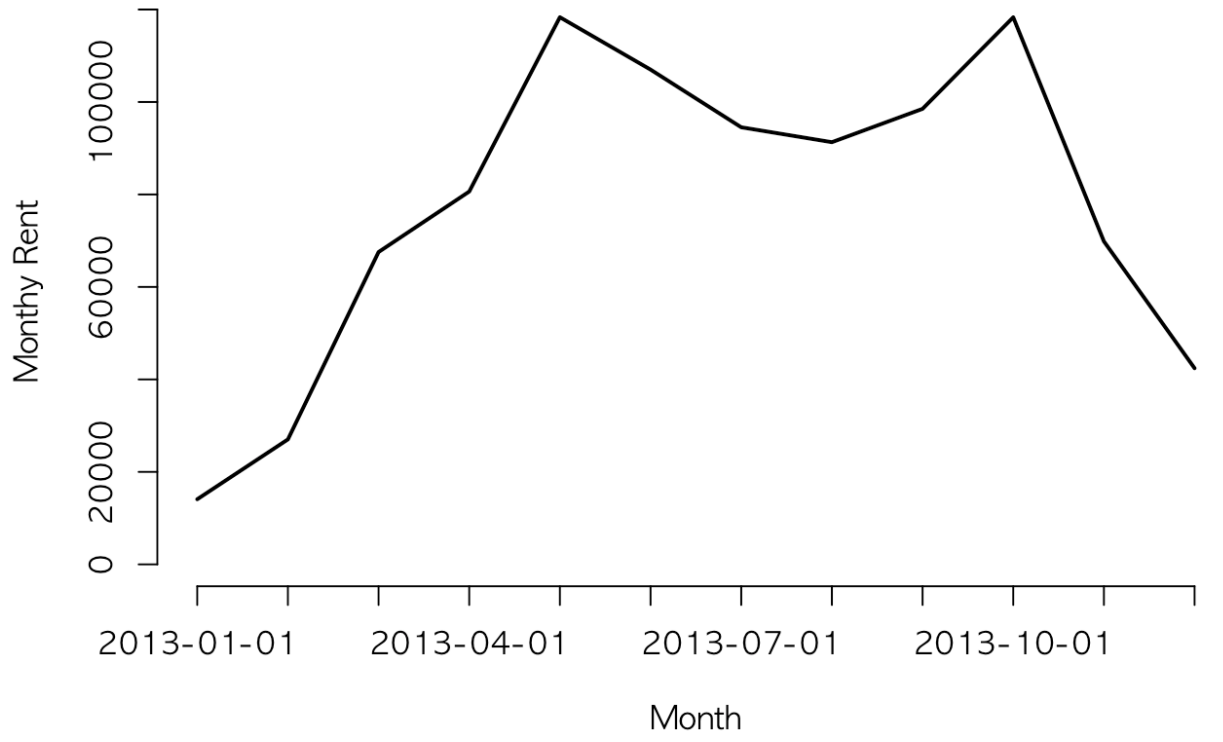
```
barplot(wday_cnt2$Freq,names.arg=c("Mon","Tue","Wed","Thrs","Fri","Sat","Sun"),
xlab="Day of Week", ylab="Rent")
```




```
ts_daily <- table(cut(tashu$rent_date, breaks="day"))
plot(ts_daily, type="l", xlab="Date", ylab="Daily Rent")
```




```
ts_monthly <- table(cut(tashu$rent_date, breaks="month"))
plot(ts_monthly, type="l", xlab="Month", ylab="Monthy Rent")
```



```
topstation <- data.frame(table(tashu$rent_station))
topstation2 <- topstation[order(-topstation$Freq),]
head(topstation2)
```

```
##    Var1  Freq
## 3     3 49416
## 56   56 30383
## 17   17 26262
## 31   31 25968
## 33   33 22792
## 32   32 22196
```