RDRP

lile
Thursday, April 16, 2015

loading and preprocessing the data

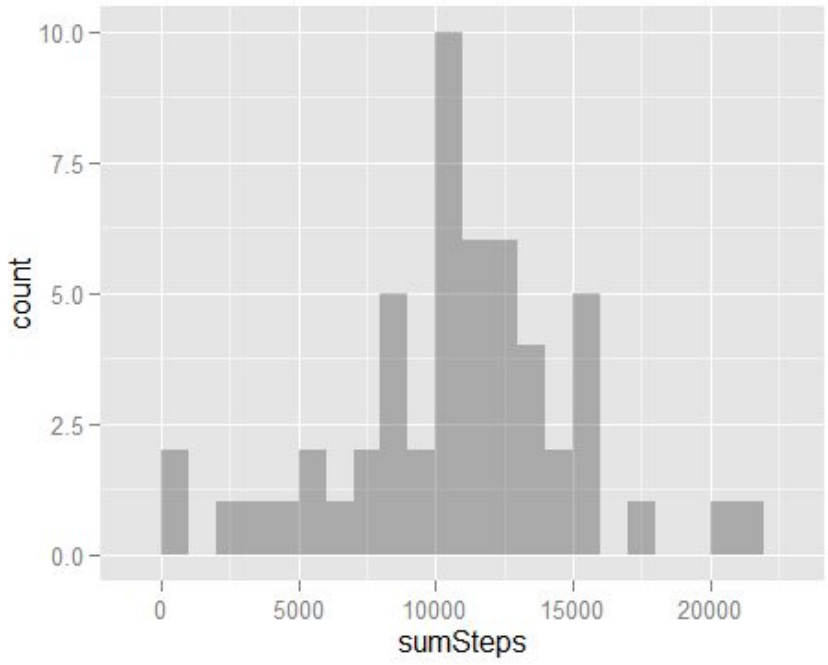
```
library(data.table)
setwd("E:\\XiGua YingShi\\可重复性报告")
data<-read.csv("activity.csv")
class(data)
```

```
## [1] "data.frame"
```

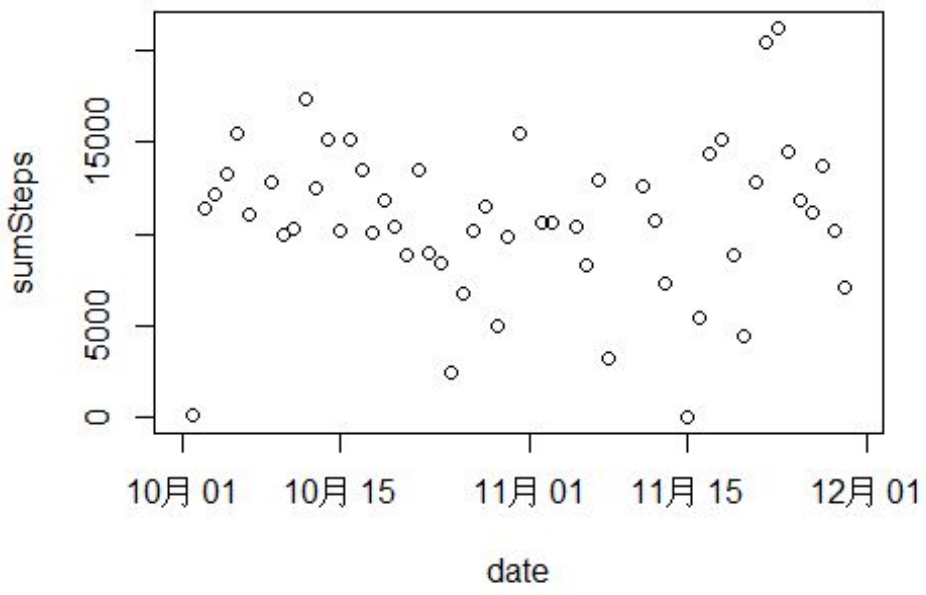
```
#head(data,3)
#data$steps
```

What is mean total number of steps taken per day

```
data<-data.table(data)
data<-data[,date:=as.Date(date)]
dtDaily<-data[,list(sumSteps=sum(steps)),date]
#head(dtDaily)
library(ggplot2)
ggplot(dtDaily,aes(x=sumSteps))+geom_histogram(alpha=1/3,binwidth=1000)
```



```
plot(dtDaily)
```




```
data<-data.table(data)
dtDaily2<-data[,list(meanSteps=mean(steps,na.rm=TRUE)),date]
#head(dtDaily2)
data<-data.table(data)
dtDaily3<-dtDaily[,list(medianSteps=median(sumSteps,na.rm=TRUE))]
#tab
dtIntervals<-data[,list(meanSteps=mean(steps,na.rm=TRUE)),interval]
```

what is the average daily activity pattern

```
ggplot(dtIntervals,aes(x=interval,y=meanSteps))+geom_line()
```

Imputing missing values

data<-data.table(data)

data<-data[,isStepMissing:=is.na(steps)]

tab<-data[,.N,isStepMissing]

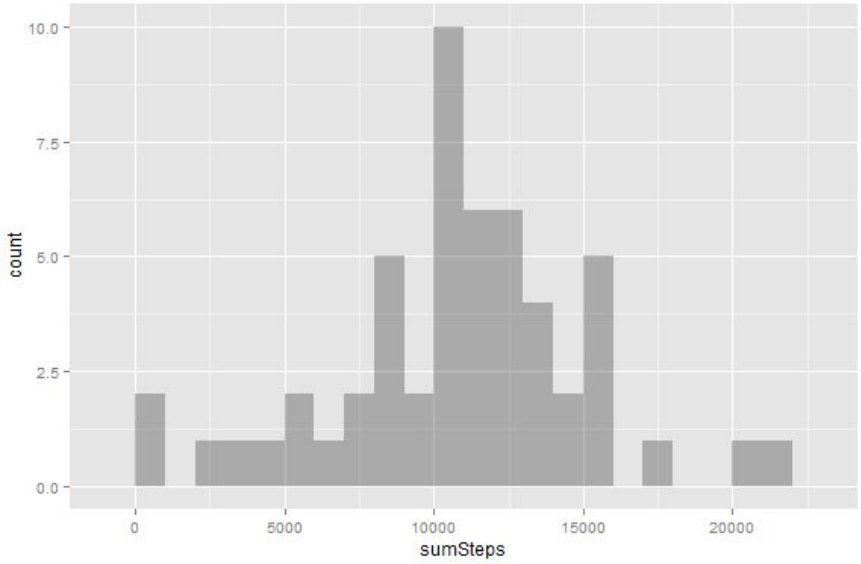
tab

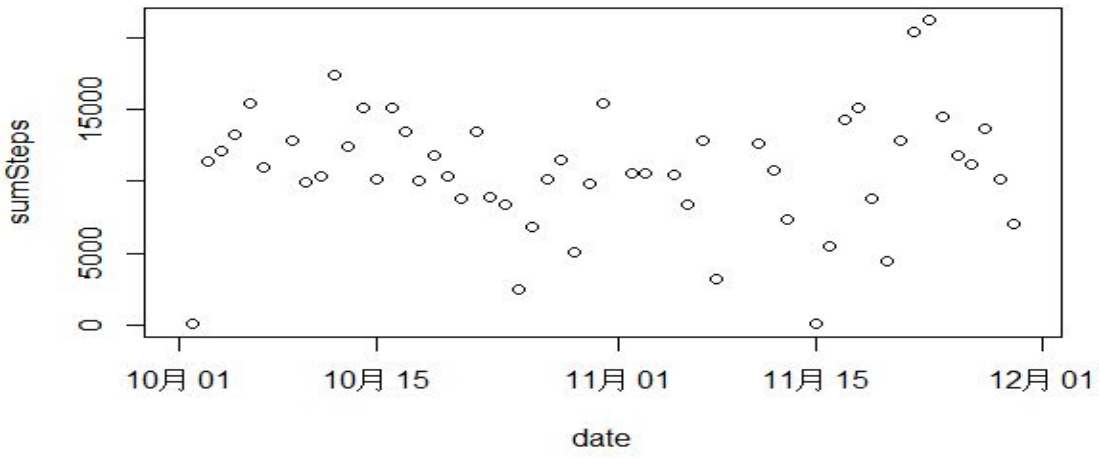
install.packages("VIM")

install.packages("robustbase")

library(VIM)

```
data<-data.frame(data)
dat<-irmi(data)
dat<-data.table(dat)
dtDaily<-dat[,list(sumSteps=sum(steps)),date]
head(dtDaily)
ggplot(dtDaily,aes(x=sumSteps))+geom_histogram(alpha=1/3,binwidth=1000)
plot(dtDaily)
```





Are there difference in activity patterns between weekdays and weekends

```
levels <- c("星期一", "星期二", "星期三", "星期四", "星期五", "星期六", "星期天")

newLevels <- c("Weekend", rep("Weekday", 5), "Weekend")

dat <- dat[, dayOfWeek := factor(weekdays(date), levels=levels)]

dat <- dat[, dayType := factor(newLevels[dayOfWeek])]

dat[, .N, list(dayType, dayOfWeek)]

dat<-dat[,list(meanSteps=mean(steps,na.rm=TRUE)),list(dayType,interval)]

dtIntervals <- dat[, list(meanSteps = mean(steps, na.rm=TRUE)), list(dayType,
interval)]

ggplot(dtIntervals,aes(x=interval,y=meanSteps,color=dayType))+geom_line()
```